Probabilités via l'intégrale de Riemann

Cours et exercices

Charles Suquet

Préparation à l'agrégation interne

Ce polycopié est extrait d'un ouvrage à paraître aux éditions Ellipses fin février 2013.

Avant-propos

Les probabilités et la statistique occupant une place grandissante dans tous les domaines de la science, leur enseignement s'est considérablement développé au cours des deux dernières décennies. L'évolution des programmes de mathématiques de l'enseignement secondaire illustre bien cette tendance. Désormais un bachelier scientifique est réputé avoir acquis une certaine familiarité avec la « loi normale » et les « intervalles de confiance », selon les programmes [4] en vigueur à partir de la rentrée 2012.

Bien entendu, les bases mathématiques donnant à ces notions une assise rigoureuse ne sont pas disponibles au niveau lycée. Elles relèvent d'un enseignement universitaire en 3^{e} année de licence, prolongé en 1re année de master. On assiste donc à une véritable discontinuité entre l'enseignement des probabilités au lycée et à l'université. Au niveau lycée, seules les probabilités sur un ensemble fini peuvent recevoir un traitement rigoureux. Ce simple constat ne remet pas en cause l'intérêt de l'élargissement culturel apporté par les évolutions récentes des programmes du secondaire. Les difficultés dans l'enseignement des probabilités sont davantage d'ordre conceptuel que technique et il est certainement profitable d'avoir vu fonctionner les notions probabilistes avant d'en faire une théorie rigoureuse.

La première difficulté sérieuse dans la théorie des probabilités est l'apparition de l'infini, qui s'invite très naturellement dans les problèmes les plus élémentaires de temps d'attente (« on lance un dé jusqu'à la première obtention d'un six »). On ne peut guère traiter cette situation avant de disposer au moins de l'outil des *séries*, généralement étudié en deuxième année de licence. En première année, tout ce que l'on peut faire (et ce ne serait déjà pas si mal) est de mettre en place le langage ensembliste et d'acquérir la notion de *dénombrabilité*. En deuxième année, la connaissance des séries permet de traiter proprement ce qu'il est convenu d'appeler les « probabilités discrètes [1] ». Par exemple, en se limitant aux variables aléatoires discrètes, il est tout à fait possible de traiter rigoureusement de l'espérance mathématique [14] et de ses propriétés, notamment la linéarité. À ce stade, les variables aléatoires réelles ou les « probabilités géométriques » sont rarement étudiées. Et il y a une bonne raison à

1. Il est arrivé à l'auteur d'avoir à enseigner les probabilités discrètes au premier semestre, à des étudiants n'ayant pas encore vu les séries en cours d'analyse. Il a pu constater à cette occasion que des questions probabilistes d'énoncé élémentaire étaient aussi une bonne motivation pédagogique pour introduire les séries.

cela, c'est que tout progrès au delà des probabilités discrètes nécessite des outils d'intégration.

La voie royale est alors l'introduction de l'intégrale de Lebesgue et de la théorie de la mesure. C'est réellement le socle mathématique de la théorie moderne des probabilités et c'est un passage obligé pour aborder des thématiques qui s'enseignent au delà du niveau licence : espérance conditionnelle, processus stochastiques, géométrie stochastique, statistique mathématique, etc. C'est donc tout naturellement que l'intégration au sens de Lebesgue et ses développements probabilistes figurent au programme de l'agrégation externe de mathématiques et des masters « recherche ».

L'expérience montre néanmoins que l'investissement requis est prohibitif pour bon nombre d'étudiants et, qu'après avoir passé beaucoup de temps à tenter d'assimiler la théorie de l'intégrale, ils se trouvent fort démunis face à un exercice de probabilités. Il est alors naturel de proposer un raccourci à ceux qui ne se destinent pas à devenir des probabilistes professionnels, mais ont quand même besoin d'acquérir une certaine familiarité avec la théorie des probabilités. C'est le cas notamment des enseignants du secondaire. Les programmes du CAPES ou de l'agrégation interne de mathématiques font ainsi le choix de n'utiliser que l'intégrale de Riemann et insèrent les probabilités dans ce cadre restreint. On n'y considère que deux types de variables aléatoires, discrètes ou à densité. L'espérance d'une variable aléatoire discrète est alors *définie* par une série et celle d'une variable à densité par une intégrale de Riemann généralisée. Même si cette approche pragmatique permet d'arriver rapidement à traiter des problèmes concrets de probabilités, elle se heurte aussi rapidement à ses limitations. Pour n'en mentionner qu'une, mais de taille, on ne dispose pas dans ce cadre de la propriété fondamentale d'additivité de l'espérance. En effet, la somme de deux variables aléatoires à densité n'est pas forcément à densité ou discrète et alors comment définir l'espérance de cette somme ?

Dans ce livre, issu d'un cours enseigné depuis 2004 en 3$^{\text{e}}$ année de licence et 1$^{\text{re}}$ année de master à l'Université Lille 1, nous avons fait le choix de respecter une approche élémentaire [2] des probabilités via l'intégrale de Riemann. Nous nous efforçons d'aller aussi loin que possible dans le cadre de cette contrainte et de voir où en sont les véritables limitations. On définit ainsi l'espérance d'une variable aléatoire réelle en toute généralité en n'utilisant que la fonction de répartition et une intégrale de Riemann généralisée. Ceci permet une approche *graphique* tout à fait naturelle de l'espérance. On établit ensuite rigoureusement les propriétés de cette espérance ainsi que les théorèmes d'interversion limite-espérance, classiques dans la théorie de Lebesgue, mais généralement admis dans le cadre de l'intégrale de Riemann.

Bref, tant qu'on ne s'intéresse qu'aux variables aléatoires réelles, les choses se passent assez bien. La situation se dégrade un peu quand on passe à l'étude des vecteurs aléatoires, en partie à cause de limitations de l'intégrale de Riemann multiple généralisée. Il n'y a néanmoins pas énormément de choses à admettre pour pouvoir franchir l'obstacle et dérouler ensuite les propriétés des suites indépendantes et arriver

2. En mathématiques, « élémentaire » n'équivaut pas à « facile », voir le célèbre exemple de la « preuve élémentaire » du théorème des nombres premiers.

aux grands théorèmes limite de la théorie des probabilités. Il s'agit de la loi des grands nombres et du théorème limite central (ou de normalité asymptotique) qui ouvrent les portes de la statistique.

Toutes les parties du livre n'ont pas été également enseignées. Certaines ont été écrites pour répondre à la curiosité et au souci de rigueur des lecteurs les plus « avancés ». Les débutants pourront sans dommage les sauter, quitte à y revenir ultérieurement. Il s'agit notamment du développement sur les familles sommables au chapitre 2, où l'on pourra se contenter en première lecture de regarder les séries doubles et de tester sa compréhension du sujet à l'aide de l'exercice 2.5. Les deux chapitres d'intégration sont là à titre d'outillage et pourront être passés rapidement en première lecture, pour peu que l'on ait déjà quelques notions sur l'intégrale de Riemann. Il faudra probablement y revenir ponctuellement, pour clarifier quelques utilisations de l'intégration en théorie des probabilités. L'exposition de celle-ci débute au chapitre 5 qui avec les deux suivants, constitue le noyau incompressible de l'ouvrage. Dans le chapitre 8 sur les vecteurs aléatoires, on pourra shunter la preuve de l'hérédité de l'indépendance (proposition 8.37). Enfin, au chapitre 10, la section 10.4 qui propose une démonstration du théorème limite central avec l'outillage requis, relève clairement du niveau master et ne s'adresse qu'aux lecteurs les plus acharnés.

Chaque chapitre contient une section d'exercices qui suit autant que possible l'ordre d'exposition du cours. Certains sont des applications directes du cours, la plupart d'entre eux ont fait partie de sujets d'examen avant d'être recyclés en travaux dirigés. Le chapitre 11 contient 47 solutions détaillées d'exercices et quelques indications de solutions.

L'auteur tient à remercier chaleureusement tous les collègues de l'équipe de Probabilités et Statistique du laboratoire P. Painlevé avec qui il a eu des échanges fructueux. Il voudrait exprimer sa profonde reconnaissance à Gwenaëlle Castellan, Myriam Fradon, Laurence Marsalle et Marie-Claude Viano pour un enrichissant travail en équipe pédagogique et leur importante contribution à la relecture du manuscrit.

Villeneuve d'Ascq, le 1er juin 2012.

Table des matières

1 Dénombrement **13**

1.1 Rappels ensemblistes . . . 13
1.1.1 Opérations ensemblistes . . . 13
1.1.2 Bijections . . . 15
1.2 Ensembles finis et dénombrement . . . 17
1.3 Dénombrabilité . . . 24
1.4 Exercices . . . 33

2 Sommes **37**

2.1 Rappels sur les séries . . . 37
2.1.1 Généralités . . . 37
2.1.2 Séries à termes positifs . . . 39
2.1.3 Séries à termes de signe non constant . . . 41
2.1.4 Opérations sur les séries . . . 42
2.2 Familles sommables . . . 47
2.3 Séries doubles . . . 60
2.4 Exercices . . . 64

3 Intégrale de Riemann sur $[a, b]$ **69**

3.1 Construction . . . 69
3.2 Riemann intégrabilité . . . 76
3.3 Propriétés de l'intégrale de Riemann . . . 83
3.3.1 Propriétés de l'ensemble $\mathcal{R}[a, b]$ . . . 83
3.3.2 Propriétés relatives à l'intervalle d'intégration . . . 88
3.4 Interversion limite intégrale . . . 95
3.5 Exercices . . . 97

4 Intégrale généralisée **99**

4.1 Construction . . . 99
4.2 Critère de Cauchy pour intégrales généralisées . . . 110
4.3 Intégrales généralisées de fonctions positives . . . 116
4.4 Divers . . . 121

4.4.1 Changements de variable . . . 121
4.4.2 Intégration par parties . . . 124
4.4.3 Comparaison des intégrales ordinaires et généralisées . . . 127
4.5 Un aperçu sur les intégrales multiples . . . 128
4.5.1 Intégration sur un pavé . . . 128
4.5.2 Intégration sur un ensemble borné . . . 131
4.5.3 Ensembles et fonctions admissibles . . . 135
4.5.4 Intégrales répétées . . . 141
4.5.5 Changement de variable . . . 145
4.5.6 Intégrale multiple généralisée . . . 146
4.6 Exercices . . . 152

5 Évènements et probabilités 155
5.1 Notion de mesure . . . 155
5.2 Modéliser l'aléatoire . . . 161
5.2.1 Notion d'expérience aléatoire . . . 161
5.2.2 Évènements . . . 162
5.2.3 Une question de dés . . . 163
5.3 La probabilité comme mesure . . . 166
5.4 Exemples . . . 172
5.5 Remarques sur le choix d'un modèle . . . 179
5.6 Probabilités conditionnelles . . . 181
5.6.1 Introduction . . . 181
5.6.2 Propriétés . . . 183
5.6.3 Quelques exemples . . . 186
5.7 Indépendance . . . 188
5.7.1 Indépendance de deux évènements . . . 188
5.7.2 Indépendance mutuelle . . . 191
5.7.3 Épreuves répétées . . . 192
5.7.4 L'algorithme du rejet . . . 193
5.8 Exercices . . . 195

6 Variables aléatoires 203
6.1 Introduction . . . 203
6.2 Généralités . . . 206
6.2.1 Variables aléatoires réelles . . . 206
6.2.2 Loi d'une variable aléatoire . . . 208
6.2.3 Fonction de répartition . . . 211
6.2.4 Lois à densité . . . 214
6.3 Lois discrètes classiques . . . 220
6.3.1 Lois de Bernoulli . . . 221
6.3.2 Loi uniforme sur un ensemble fini de réels . . . 221
6.3.3 Lois binomiales . . . 221
6.3.4 Lois hypergéométriques . . . 222

6.3.5 Lois géométriques . . . 224
6.3.6 Lois de Poisson . . . 225
6.3.7 Sur le caractère universel de la loi de Poisson . . . 230
6.4 Lois à densité classiques . . . 233
6.4.1 Lois uniformes . . . 233
6.4.2 Lois exponentielles . . . 234
6.4.3 Lois gaussiennes . . . 237
6.4.4 Lois de Cauchy . . . 239
6.5 Un exemple de loi continue sans densité . . . 239
6.6 Exercices . . . 243

7 Espérance **251**
7.1 Introduction . . . 251
7.2 Espérance d'une variable aléatoire positive . . . 254
7.3 Espérance d'une variable aléatoire réelle . . . 270
7.4 Moments . . . 276
7.5 Exercices . . . 283

8 Vecteurs aléatoires et indépendance **295**
8.1 Vecteurs aléatoires . . . 296
8.1.1 Généralités . . . 296
8.1.2 Covariance de deux variables aléatoires . . . 305
8.1.3 Espérance et covariance d'un vecteur aléatoire . . . 308
8.2 Indépendance de variables et vecteurs aléatoires . . . 312
8.2.1 Suites indépendantes . . . 312
8.2.2 Indépendance des composantes . . . 316
8.2.3 Indépendance et espérance de produits . . . 324
8.3 Exercices . . . 328

9 Théorèmes limite **333**
9.1 Convergences de suites de variables aléatoires . . . 333
9.1.1 Convergence presque sûre et en probabilité . . . 333
9.1.2 Convergence en moyenne d'ordre p . . . 342
9.1.3 Bilan sur les modes de convergence . . . 347
9.2 Loi des grands nombres . . . 348
9.2.1 Loi faible des grands nombres . . . 348
9.2.2 Loi forte des grands nombres . . . 349
9.2.3 L'aiguille de Buffon . . . 352
9.3 Exercices . . . 356

10 Théorème limite central **363**
10.1 Convergence en loi . . . 363
10.2 Normalité asymptotique . . . 369
10.2.1 Sommes de variables aléatoires i.i.d. . . . 370

10.2.2 Vitesse de convergence dans le TLC 372
10.2.3 Intervalle de confiance pour une probabilité inconnue 374
10.2.4 Généralisation du TLC 376
10.3 Théorème limite central vectoriel 379
10.3.1 Vecteurs aléatoires gaussiens 379
10.3.2 TLC vectoriel 383
10.4 Compléments sur la convergence en loi et le TLC 385
10.4.1 Outillage pour la convergence en loi 385
10.4.2 Démonstration du TLC 393
10.5 Exercices 401

11 Indications et solutions **413**
11.1 Exercices du chapitre 1 413
11.2 Exercices du chapitre 2 416
11.3 Exercices du chapitre 3 421
11.4 Exercices du chapitre 4 423
11.5 Exercices du chapitre 5 428
11.6 Exercices du chapitre 6 444
11.7 Exercices du chapitre 7 462
11.8 Exercices du chapitre 8 484
11.9 Exercices du chapitre 9 489
11.10 Exercices du chapitre 10 494

Tables de la loi normale **519**

Bibliographie **521**

Index **523**

Chapitre 1

Dénombrement

Compter des objets est certainement l'activité la plus élémentaire à la base des mathématiques. Pourtant, à y regarder de plus près, ce n'est pas si facile. Déjà pour un ensemble fini, la méthode qui consiste à regarder ses éléments l'un après l'autre et à les compter (donc à les numéroter) n'est applicable que pour de « petits » ensembles. Le plus souvent on s'en sort en faisant une représentation de l'ensemble à dénombrer à l'aide d'un autre ensemble plus familier. Cette représentation est ce que l'on appelle une bijection. Elle est d'ailleurs à la base du processus de comptage qui consiste simplement à mettre en bijection un ensemble avec un ensemble de nombres entiers. Cette notion de bijection permet d'étendre en un certain sens le dénombrement aux ensembles infinis.

Dans tout ce qui suit, la notation $[\![a, b]\!]$ pour $a, b \in \mathbb{N}$ désigne l'ensemble de tous les entiers compris au sens large entre a et b. On définit de même les intervalles d'entiers semi ouverts ou ouverts. L'écriture un peu abusive « $\forall i = 1, \ldots, n$ » signifie « $\forall i \in [\![1, n]\!]$ ».

1.1 Rappels ensemblistes

1.1.1 Opérations ensemblistes

Soit Ω un ensemble. On dit que l'ensemble A est un *sous-ensemble* ou une *partie* de Ω (notation $A \subset \Omega$) si tout élément de A est aussi un élément de Ω ($\forall \omega \in A$, $\omega \in \Omega$). On appelle $\mathcal{P}(\Omega)$ l'ensemble des parties de Ω, ce que l'on peut noter [1]

$$\mathcal{P}(\Omega) = \{A \ ; \ A \subset \Omega\}.$$

Ainsi les écritures $A \subset \Omega$ et $A \in \mathcal{P}(\Omega)$ sont deux façons de dire la même chose [2].

1. Dans toutes les écritures d'ensembles entre accolades, nous utilisons le point virgule au sens de « tel que ».

2. Noter cependant la différence de statut de A : dans la première écriture, A est considéré comme un ensemble, dans la deuxième comme un élément d'un ensemble d'un type un peu particulier.

Si A et B sont deux parties du même ensemble Ω, on dit que A est *incluse* dans B (notation $A \subset B$) si tout élément de A est aussi élément de B ($\forall \omega \in A,\ \omega \in B$), autrement dit, si l'appartenance à A *implique* l'appartenance à B :

$$A \subset B \quad \text{signifie} \quad \forall \omega \in \Omega, \quad (\omega \in A) \Rightarrow (\omega \in B).$$

Soit I un ensemble quelconque d'indices (fini ou infini) et $(A_i)_{i\in I}$ une famille de parties de Ω. On définit son intersection $\bigcap_{i\in I} A_i$ et sa réunion, $\bigcup_{i\in I} A_i$ par :

$$\bigcap_{i\in I} A_i := \{\omega \in \Omega\ ;\ \forall i \in I,\ \omega \in A_i\} \qquad \text{et} \qquad \bigcup_{i\in I} A_i = \{\omega \in \Omega\ ;\ \exists i \in I,\ \omega \in A_i\}. \quad (1.1)$$

Autrement dit, $\bigcap_{i\in I} A_i$ est l'ensemble des éléments appartenant à la fois à *tous* les A_i et $\bigcup_{i\in I} A_i$ est l'ensemble des éléments appartenant à *au moins l'un* des A_i.

Remarque 1.1. La réunion et l'intersection d'une famille de parties de Ω sont définies de façon globale, elles s'obtiennent d'un coup, *sans passage à la limite* quand I est infini et sans qu'un ordre éventuel sur l'ensemble d'indices I n'ait d'importance.

Réunion et intersection sont très utiles pour la traduction automatique des quantificateurs. Si I est un ensemble quelconque d'indices, (π_i) une propriété dépendant de l'indice i et A_i l'ensemble des $\omega \in \Omega$ vérifiant (π_i),

$$\begin{aligned} \{\omega \in \Omega\ ;\ \forall i \in I,\ \omega \text{ vérifie } (\pi_i)\} &= \bigcap_{i\in I} A_i, \\ \{\omega \in \Omega\ ;\ \exists i = i(\omega) \in I,\ \omega \text{ vérifie } (\pi_i)\} &= \bigcup_{i\in I} A_i. \end{aligned}$$

Ainsi le quantificateur $\forall$ *peut se traduire par une intersection* et le quantificateur $\exists$ *par une réunion*. Pour une importante application de cette remarque, voir page 334.

L'intersection et l'union sont distributives l'une par rapport à l'autre, c'est à dire

$$B \cup \Big(\bigcap_{i\in I} A_i\Big) = \bigcap_{i\in I} (A_i \cup B), \qquad B \cap \Big(\bigcup_{i\in I} A_i\Big) = \bigcup_{i\in I} (A_i \cap B).$$

Le *complémentaire* de A (dans Ω) est l'ensemble $A^c := \{\omega \in \Omega\ ;\ \omega \notin A\}$. L'opération passage au complémentaire (qui est une bijection de $\mathcal{P}(\Omega)$ dans lui-même) vérifie $(A^c)^c = A$, $\Omega^c = \varnothing$, $\varnothing^c = \Omega$ et échange réunions et intersections grâce aux très utiles formules :

$$\Big(\bigcap_{i\in I} A_i\Big)^c = \bigcup_{i\in I} A_i^c \qquad \Big(\bigcup_{i\in I} A_i\Big)^c = \bigcap_{i\in I} A_i^c.$$

On définit le *produit cartésien* de deux ensembles E et F, noté $E \times F$, par :

$$E \times F := \{(x, y)\ ;\ x \in E, y \in F\}.$$

Attention, dans cette écriture (x, y) ne désigne en aucune façon un ensemble mais un couple d'éléments (l'ordre d'écriture a une importance). Pour éviter toute confusion, on utilise des accolades pour la description des ensembles et des parenthèses pour les couples d'éléments.

On définit de manière analogue le produit cartésien d'une suite finie d'ensembles $E_1, \dots, E_n$ par [3]

$$E_1 \times \cdots \times E_n := \big\{(x_1, \dots, x_n) \ ; \ \forall i \in [\![1, n]\!], \ x_i \in E_i\big\}.$$

L'ensemble $E^2 := E \times E = \{(x_1, x_2) \ ; \ x_1 \in E, x_2 \in E\}$ peut être utilisé pour représenter l'ensemble de toutes les applications de $\{1, 2\}$ dans E, le couple (x_1, x_2) correspondant à l'application $f : \{1, 2\} \to E$ définie par $f(1) = x_1$ et $f(2) = x_2$. Il pourrait de la même façon, représenter les applications d'un ensemble à deux éléments dans E (remplacer les chiffres 1 et 2 par n'importe quelle paire de symboles distincts : 0 et 1, a et b, etc.).

Plus généralement, pour $n \geqslant 2$, E^n est l'ensemble des n-uplets ou listes de longueur n d'éléments de E. Dans un n-uplet $(x_1, \dots, x_n)$, il peut y avoir des répétitions. On peut aussi utiliser E^n pour représenter toutes les applications de l'ensemble $[\![1, n]\!]$ (ou de n'importe quel ensemble à n éléments) dans E.

Soit I un ensemble quelconque, fini ou infini. Par analogie avec ce qui précède, l'ensemble de toutes les applications $f : I \to E$ sera noté E^I. Par exemple avec $E = \{0, 1\}$ et $I = \mathbb{N}$, on obtient l'ensemble $\{0, 1\}^{\mathbb{N}}$ de toutes les suites de chiffres binaires indexées par $\mathbb{N}$: $\{0, 1\}^{\mathbb{N}} = \{u = (u_i)_{i \in \mathbb{N}} \ ; \ u_i = 0 \text{ ou } 1\}$. Avec $E = \mathbb{R}$ et $I = [0, 1]$, on obtient l'ensemble $\mathbb{R}^{[0,1]}$ des fonctions définies sur l'intervalle $[0, 1]$ et à valeurs dans $\mathbb{R}$.

1.1.2 Bijections

Définition 1.2 (injection)**.** *Une application $f : E \to F$ est dite* injective *si deux éléments distincts de E ont toujours des images distinctes dans F :*

$$\forall x \in E, \forall x' \in E, \quad (x \neq x') \Rightarrow (f(x) \neq f(x')).$$

Une formulation équivalente est :

$$\forall x \in E, \forall x' \in E, \quad (f(x) = f(x')) \Rightarrow (x = x').$$

Une application injective $f : E \to F$ est appelée injection *de E* dans *F.*

Définition 1.3 (surjection)**.** *Une application $f : E \to F$ est dite* surjective *si tout élément de l'ensemble d'arrivée a au moins un antécédent par f :*

$$\forall y \in F, \ \exists x \in E, \quad f(x) = y.$$

Une application surjective $f : E \to F$ est appelée surjection *de E* sur *F.*

3. Noter qu'ici le quantificateur « $\forall i$ » ne se traduit pas par une intersection. Ne confondez pas « $\forall i \in [\![1, n]\!], \ x \in E_i$ » qui traduit l'appartenance de x à $\bigcap_{i \in I} E_i$ avec « $\forall i \in [\![1, n]\!], \ x_i \in E_i$ ».

Définition 1.4 (bijection)**.** *Une application $f : E \to F$ est dite* bijective *si elle est à la fois injective et surjective, autrement dit si tout élément de l'ensemble d'arrivée F a un* unique *antécédent par f dans l'ensemble de départ E :*

$$\forall y \in F,\ \exists!\, x \in E, \quad f(x) = y.$$

Une application bijective $f : E \to F$ est appelée bijection *de E* sur *F.*

Remarque 1.5. Si $f : E \to F$ est une injection, en restreignant son ensemble d'arrivée à $f(E) := \{y \in F;\ \exists x \in E\ ;\ f(x) = y\}$ la nouvelle application $f : E \to f(E)$ est une bijection. En effet cette opération préserve clairement l'injectivité et rend f surjective.

Définition 1.6 (application réciproque)**.** *Soit $f : E \to F$ une bijection. Tout $y \in F$ admet un unique antécédent x par f dans E. En posant $f^{-1}(y) := x$, on définit une application $f^{-1} : F \to E$ appelée application réciproque de f ou inverse de f. Cette application f^{-1} est bijective.*

Justification. Pour vérifier l'injectivité de f^{-1}, soient y et y' deux éléments de F tels que $f^{-1}(y) = f^{-1}(y')$. Cela signifie qu'ils ont le même antécédent x par f, donc que $y = f(x)$ et $y' = f(x)$, d'où $y = y'$.

Pour la surjectivité, soit $x \in E$ quelconque. Posons $y = f(x)$. Alors x est antécédent de y par f, donc $f^{-1}(y) = x$ et ainsi y est antécédent de x par f^{-1}. Tout élément de E a donc un antécédent dans F par f^{-1}. Autrement dit, f^{-1} est surjective. □

Remarque 1.7. Ainsi l'existence d'une bijection $E \to F$ équivaut à celle d'une bijection $F \to E$. On dira que E « est en bijection avec » F s'il existe une bijection $E \to F$ (ou $F \to E$).

Proposition 1.8. *Soient $f : E \to F$ et $g : F \to G$ deux bijections. Alors $g \circ f$ est une bijection de E sur G. De plus $(g \circ f)^{-1} = f^{-1} \circ g^{-1}$.*

Preuve. Rappelons que $(g \circ f)(x) := g(f(x))$ pour tout $x \in E$. Pour vérifier l'injectivité, soient x et x' dans E tels que $(g \circ f)(x) = (g \circ f)(x')$. Cette égalité réécrite $g(f(x)) = g(f(x'))$ implique par injectivité de g l'égalité $f(x) = f(x')$, laquelle implique $x = x'$ par injectivité de f. Pour la surjectivité de $g \circ f$, soit $z \in G$ quelconque. Par surjectivité de g, z a au moins un antécédent y dans F avec $g(y) = z$. À son tour $y \in F$ a un antécédent $x \in E$ par la surjection f. Finalement $y = f(x)$ et $z = g(y)$, d'où $z = g(f(x)) = (g \circ f)(x)$, ce qui montre que z a pour antécédent x par $g \circ f$. Comme z était quelconque, la surjectivité de $g \circ f$ est établie. Ainsi $g \circ f$ est une bijection de E sur G. En conservant les notations, $(g \circ f)^{-1}(z) = x$. D'autre part $x = f^{-1}(y)$ et $y = g^{-1}(z)$, d'où $x = f^{-1}(g^{-1}(z)) = (f^{-1} \circ g^{-1})(z)$. Donc pour z quelconque dans G, $(g \circ f)^{-1}(z) = x = (f^{-1} \circ g^{-1})(z)$, d'où $(g \circ f)^{-1} = f^{-1} \circ g^{-1}$. □

1.2 Ensembles finis et dénombrement

Définition 1.9. *Un ensemble E est dit fini s'il est vide ou s'il est en bijection avec $[\![1, n]\!]$ pour un certain entier $n \geqslant 1$. Un tel n est alors unique et est appelé* cardinal *de E (notation $\operatorname{card} E$). Par convention le cardinal de l'ensemble vide est 0.*

La cohérence de la définition 1.9 repose sur le lemme suivant (pourquoi ?).

Lemme 1.10. *Si n et m sont deux entiers distincts, il n'existe pas de bijection entre $[\![1, n]\!]$ et $[\![1, m]\!]$.*

Preuve. On peut toujours supposer sans perte de généralité que $0 \leqslant n < m$. Dans le cas où $n = 0$, l'ensemble $[\![1, n]\!]$ est l'ensemble des entiers j tels que $1 \leqslant j \leqslant 0$, c'est donc l'ensemble vide. On prouve le résultat par récurrence sur n en adoptant comme hypothèse de récurrence :

$$(\mathcal{H}_n) \quad \forall m > n, \quad \text{il n'existe pas de bijection } [\![1, n]\!] \to [\![1, m]\!].$$

Initialisation. $(\mathcal{H}_0)$ est clairement vraie, car on ne peut définir aucune application sur l'ensemble vide donc *a fortiori* aucune bijection.

Induction. Montrons que si $(\mathcal{H}_n)$ est vérifiée pour un certain n, alors $(\mathcal{H}_{n+1})$ l'est aussi. Pour cela on raisonne par l'absurde : si $(\mathcal{H}_{n+1})$ n'était pas pas vérifiée, il existerait un entier $m > n + 1$ et une bijection $f : [\![1, n + 1]\!] \to [\![1, m]\!]$. Nous allons construire à partir de f une bijection $g : [\![1, n + 1]\!] \to [\![1, m]\!]$ telle que $g(n + 1) = m$. Notons $j = f(n + 1)$. Si $j = m$, il suffit de prendre $g = f$. Si $j \neq m$, considérons la transposition $\tau : [\![1, m]\!] \to [\![1, m]\!]$ qui échange j et m et laisse les autres éléments inchangés. C'est une bijection et l'application composée $g = \tau \circ f$ est une bijection $[\![1, n + 1]\!] \to [\![1, m]\!]$ vérifiant $g(n + 1) = m$.

La restriction $\tilde{g}$ de g à $[\![1, n]\!]$ est une bijection de $[\![1, n]\!]$ sur $\{1, \ldots, m - 1\}$ et comme $m > n+1$, on a bien $m-1 > n$, ce qui contredit $(\mathcal{H}_n)$. Nous venons d'établir l'implication $(\mathcal{H}_n) \Rightarrow (\mathcal{H}_{n+1})$, ce qui achève la récurrence. □

Remarque 1.11. Si l'ensemble F est en bijection avec un ensemble fini E, alors F est fini et a même cardinal que E. En effet en notant $n = \operatorname{card} E$, il existe une bijection $f : [\![1, n]\!] \to E$ et une bijection $g : E \to F$. La composée $g \circ f$ réalise alors une bijection de $[\![1, n]\!]$ sur F.

Proposition 1.12. *Soient E et F deux ensembles finis.*

$$\text{Si } E \cap F = \varnothing, \quad \operatorname{card}(E \cup F) = \operatorname{card} E + \operatorname{card} F. \tag{1.2}$$

Preuve. Dans le cas où l'un des deux ensembles est vide, (1.2) est triviale. On suppose désormais que $\operatorname{card} E = n \geqslant 1$ et $\operatorname{card} F = m \geqslant 1$. Il existe alors des bijections

$$f : [\![1, n]\!] \to E, \quad g : [\![1, m]\!] \to F.$$

On prouve (1.2) en construisant une bijection h de $[\![1, n+m]\!]$ sur $E \cup F$. La translation

$$t : [\![n + 1, n + m]\!] \to [\![1, m]\!], \quad i \mapsto i - n$$

est une bijection et l'application $g \circ t$ réalise une bijection de $[\![n+1, n+m]\!]$ sur F. Définissons alors h par

$$h(i) := \begin{cases} f(i) & \text{si } i \in [\![1, n]\!], \\ (g \circ t)(i) & \text{si } i \in [\![n+1, n+m]\!]. \end{cases}$$

Pour vérifier la surjectivité de h, soit z un élément quelconque de $E \cup F$. Si $z \in E$, alors il a un antécédent i dans $[\![1, n]\!]$ par la surjection f et comme $h(i) = f(i) = z$, i est aussi antécédent de z par h. Si $z \in F$, il a un antécédent j dans $[\![1, m]\!]$ par la surjection g et j a un antécédent i dans $[\![n+1, n+m]\!]$ par la surjection t. Alors $h(i) = g(t(i)) = g(j) = z$ donc i est antécédent de z par h.

Pour vérifier l'injectivité de h, notons i et i' deux éléments distincts de $[\![1, n+m]\!]$. S'ils sont l'un dans $[\![1, n]\!]$ et l'autre dans $[\![n+1, n+m]\!]$, leurs images $h(i)$ et $h(i')$ sont l'une dans E et l'autre dans F qui sont *disjoints* ($E \cap F = \varnothing$) donc $h(i) \neq h(i')$. Sinon i et i' sont tous deux dans $[\![1, n]\!]$ (resp. dans $[\![n+1, n+m]\!]$) et $h(i) \neq h(i')$ en raison de l'injectivité de f (resp. de $g \circ t$). □

Corollaire 1.13 (de la proposition 1.12)**.**

a) Si $E_1, \ldots, E_d$ sont d ensembles finis deux à deux disjoints, *$E_1 \cup \cdots \cup E_d$ est un ensemble fini et*

$$\operatorname{card}\left(\bigcup_{i=1}^{d} E_i\right) = \sum_{i=1}^{d} \operatorname{card} E_i.$$

b) Si E et F sont des ensembles finis quelconques (pas forcément disjoints), $E \cup F$ est fini et

$$\operatorname{card}(E \cup F) = \operatorname{card} E + \operatorname{card} F - \operatorname{card}(E \cap F).$$

Preuve. Laissée en exercice. □

Proposition 1.14. *Soient E et F deux ensembles finis. Leur produit cartésien a pour cardinal*

$$\operatorname{card}(E \times F) = \operatorname{card} E \times \operatorname{card} F. \tag{1.3}$$

Preuve. Le cas où l'un des deux ensembles E ou F est vide étant trivial, on suppose désormais qu'aucun des deux n'est vide. On fait une récurrence sur $n = \operatorname{card} E$, en prenant pour hypothèse de récurrence :

$(\mathcal{H}_n)$ si $\operatorname{card} E = n$, alors $\operatorname{card}(E \times F) = n \operatorname{card} F$ pour tout F fini non vide.

Initialisation. Si $\operatorname{card} E = 1$, E n'a qu'un élément x_1 et l'application $h : \{x_1\} \times F \to F$, $(x_1, y) \mapsto y$ est clairement une bijection donc $\operatorname{card}(\{x_1\} \times F) = \operatorname{card} F$ et $(\mathcal{H}_1)$ est vérifiée.

Induction. Supposons $(\mathcal{H}_n)$ vraie pour un certain n et soit E un ensemble de cardinal $n+1$. Il existe une bijection $f : [\![1, n+1]\!] \to E$ permettant de numéroter les éléments de E en posant pour tout $i \in [\![1, n+1]\!]$, $x_i := f(i)$. La restriction de f à $[\![1, n]\!]$ est

une bijection de $[\![1,n]\!]$ sur son image $E' = \{x_1, \dots, x_n\}$. Ainsi E' est de cardinal n et E est l'union de ses deux sous-ensembles disjoints E' et $\{x_{n+1}\}$. On en déduit immédiatement que $E \times F$ est l'union des deux produits cartésiens disjoints $E' \times F$ et $\{x_{n+1}\} \times F$. Par la proposition 1.12 on a alors

$$\operatorname{card}(E \times F) = \operatorname{card}(E' \times F) + \operatorname{card}(\{x_{n+1}\} \times F).$$

En utilisant $(\mathcal{H}_n)$ et $(\mathcal{H}_1)$, on obtient alors

$$\operatorname{card}(E \times F) = n \operatorname{card} F + \operatorname{card} F = (n+1) \operatorname{card} F,$$

donc $(\mathcal{H}_{n+1})$ est vérifiée. □

Remarque 1.15. On aurait pu aussi prouver (1.3) en construisant explicitement une bijection $E \times F \to [\![1, nm]\!]$ (avec $\operatorname{card} E = n$ et $\operatorname{card} F = m$). Voici une façon de la construire. On note $f : [\![1,n]\!] \to E$, $i \mapsto x_i := f(i)$ et $g : [\![1,m]\!] \to F$, $j \mapsto y_j := g(j)$ des numérotations bijectives de E et F. On définit alors $h : E \times F \to [\![1, nm]\!]$ en posant $h(x_i, y_j) := m(i-1)+j$, ou de manière plus formelle $h(x,y) := m(f^{-1}(x)-1)+g^{-1}(y)$ pour tout $(x,y) \in E \times F$. On laisse en exercice la vérification de la bijectivité de h. L'idée de sa construction est simplement de ranger les couples éléments de $E \times F$ sous la forme d'un tableau où le couple (x_i, y_j) se trouve à l'intersection de la ligne i et de la colonne j et de numéroter les éléments de ce tableau en balayant chaque ligne de gauche à droite de la ligne 1 jusqu'à la ligne n.

Corollaire 1.16 (de la proposition 1.14)**.** *Si $E_1, \dots, E_d$ sont d ensembles finis,*

$$\operatorname{card}(E_1 \times \cdots \times E_d) = \prod_{i=1}^{d} \operatorname{card} E_i.$$

Preuve. Une récurrence immédiate sur d fournit le résultat. □

Proposition 1.17 (nombre d'applications $E \to F$ et cardinal de $\mathcal{P}(E)$)**.**

(a) Si $\operatorname{card} E = n$ et $\operatorname{card} F = p$, l'ensemble F^E des applications de E dans F est fini et a pour cardinal p^n, autrement dit :

$$\operatorname{card}\left(F^E\right) = (\operatorname{card} F)^{\operatorname{card} E}.$$

(b) Comme $\mathcal{P}(E)$ est en bijection avec l'ensemble $\{0,1\}^E$ des applications de E dans $\{0,1\}$,

$$\operatorname{card} \mathcal{P}(E) = 2^n = 2^{\operatorname{card} E}.$$

Preuve. Le (a) se démontre facilement par récurrence sur le cardinal de E, en notant que si on ajoute un élément x_{n+1} à E, il y a p façons différentes de prolonger $f : E \to F$ en attribuant comme image à x_{n+1} l'un des éléments de F. La rédaction détaillée est laissée en exercice.

Une bijection naturelle entre $\mathcal{P}(E)$ et $\{0,1\}^E$ est l'application φ qui à toute partie A de E associe sa fonction *indicatrice* :

$$\varphi : \mathcal{P}(E) \to \{0,1\}^E \qquad A \mapsto \varphi(A) := \mathbf{1}_A.$$

Rappelons que l'indicatrice d'une partie A de E est l'application

$$\mathbf{1}_A : E \to \{0,1\} \qquad \omega \mapsto \mathbf{1}_A(\omega) := \begin{cases} 1 & \text{si } \omega \in A, \\ 0 & \text{si } \omega \notin A. \end{cases}$$

La vérification de la bijectivité de φ est laissée en exercice [4]. □

Définition 1.18 (arrangement)**.** *Si E est un ensemble de cardinal n et $k \in [\![1,n]\!]$, on appelle* arrangement *de k éléments de E tout k-uplet $(x_1, x_2, \ldots, x_k)$ d'éléments* tous distincts *de E.*

Proposition 1.19 (dénombrement des arrangements)**.** *Le nombre d'arrangements de k éléments de E ($1 \leqslant k \leqslant n = \operatorname{card} E$) est*

$$A_n^k = n(n-1)(n-2)\cdots(n-k+1) = \frac{n!}{(n-k)!}. \tag{1.4}$$

A_n^k est aussi le nombre d'injections d'un ensemble I de cardinal k, par exemple $[\![1,k]\!]$, dans E. En particulier pour $I = E$ (et donc $k = n$), on obtient le nombre de bijections de E dans lui même (appelées aussi permutations *de E) :*

$$\text{nombre de permutations de } E = A_n^n = n!$$

Preuve. On prouve (1.4) par récurrence finie sur k, le cas $k = 1$ étant évident. Supposons donc (1.4) vraie pour un $k < n$ et montrons qu'alors elle est aussi vraie pour $k+1$. Un arrangement $(x_1, \ldots, x_{k+1})$ de $k+1$ éléments de E est déterminé de manière unique par le choix de l'arrangement $(x_1, \ldots, x_k)$ de k éléments de E et le choix de l'élément $x_{k+1} \in E \backslash \{x_1, \ldots, x_k\}$. Cela laisse donc $n-k$ choix possibles pour x_{k+1}, d'où

$$A_n^{k+1} = A_n^k(n-k) = n(n-1)(n-2)\cdots(n-k+1)(n-k),$$

ce qui montre que (1.4) est vérifiée au rang $k+1$.

Rappelons qu'un k-uplet $(x_1, \ldots, x_k)$ de E^k peut être vu comme l'application $f : [\![1,k]\!] \to E$, $i \mapsto f(i) = x_i$. Cette représentation définit une bijection φ entre l'ensemble des applications $f : [\![1,k]\!] \to E$, autrement dit $E^{[\![1,k]\!]}$ et l'ensemble E^k des k-uplets de E. Dans la représentation ci-dessus, f est injective si et seulement si les x_i sont tous distincts. Donc l'image par φ de l'ensemble des applications injectives $[\![1,k]\!] \to E$ est exactement l'ensemble des k-uplets de E qui sont des arrangements. Ces deux ensembles ont donc même cardinal et A_n^k est bien le nombre d'injections de $[\![1,k]\!] \to E$. Soit maintenant I un ensemble de cardinal k et fixons une bijection

4. Ni la définition de φ, ni la preuve de sa bijectivité n'utilisent la finitude de E. Ainsi $\mathcal{P}(E)$ et $\{0,1\}^E$ sont en bijection quel que soit l'ensemble $E \neq \varnothing$, fini ou infini.

$h : [\![1,k]\!] \to I$. L'application $\psi : E^I \to E^{[\![1,k]\!]}$, $g \mapsto \psi(g) = g \circ h$ est une bijection (vérifiez) et comme $g \circ h$ est injective si et seulement si g est injective, la restriction de ψ à l'ensemble des injections de I dans E est une bijection de cet ensemble sur celui des injections de $[\![1,k]\!]$ dans E. Donc l'ensemble des injections de I dans E a lui aussi pour cardinal A_n^k. □

Définition 1.20 (combinaison)**.** *On appelle* combinaison *de k éléments de E toute partie de cardinal k de E.*

Une combinaison (non vide) a tous ses éléments distincts comme un arrangement, mais l'ordre d'écriture n'y a pas d'importance.

Proposition 1.21 (dénombrement des combinaisons)**.** *Le nombre de combinaisons de k éléments de E ($1 \leqslant k \leqslant n = \operatorname{card} E$) est*

$$C_n^k = \frac{n(n-1)(n-2)\cdots(n-k+1)}{k(k-1)\cdots 1} = \frac{n!}{k!(n-k)!}. \tag{1.5}$$

Preuve. Notons $\mathcal{A}_k(E)$ l'ensemble de tous les arrangements de k éléments de E. Il est alors clair que l'on a la décomposition en réunion d'ensembles disjoints

$$\mathcal{A}_k(E) = \bigcup_{B \subset E,\ \operatorname{card} B = k} \mathcal{A}_k(B). \tag{1.6}$$

Autrement dit on a partitionné $\mathcal{A}_k(E)$ en regroupant dans une même classe $\mathcal{A}_k(B)$ tous les arrangements formés à partir des éléments d'une même partie B de cardinal k. Il y a donc autant de classes distinctes dans cette décomposition que de parties B de cardinal k dans E, c'est-à-dire C_n^k classes. D'autre part chaque classe $\mathcal{A}_k(B)$ contient autant d'arrangements que de bijections $B \to B$ (ou permutations sur B), c'est-à-dire $k!$. Compte-tenu du corollaire 1.13 a), on déduit alors de (1.6) que :

$$\operatorname{card} \mathcal{A}_k(E) = \sum_{B \subset E,\ \operatorname{card} B = k} \operatorname{card} \mathcal{A}_k(B) = C_n^k k!.$$

Or $\operatorname{card} \mathcal{A}_k(E) = A_n^k$ par la définition des A_n^k, d'où $A_n^k = C_n^k k!$, ce qui donne (1.5). □

Les nombres C_n^k sont appelés aussi *coefficients binomiaux* en raison de leur rôle dans la formule du binôme de Newton, que nous retrouverons ci-dessous comme cas particulier de la formule du multinôme (corollaire 1.24). Avant de nous y attaquer, il n'est peut-être pas inutile de faire un rappel sur le développement d'un produit de sommes finies.

Proposition 1.22. *Soient $J_1, \ldots, J_n$ des ensembles finis non vides d'indices et pour $i \in [\![1,n]\!]$, $j \in J_i$, des nombres réels ou complexes $x_{i,j}$. En notant $K_n := J_1 \times \cdots \times J_n$,*

$$\prod_{i=1}^{n} \left(\sum_{j \in J_i} x_{i,j} \right) = \sum_{(j_1,\ldots,j_n) \in K_n} \left(\prod_{i=1}^{n} x_{i,j_i} \right) = \sum_{j_1 \in J_1} \cdots \sum_{j_n \in J_n} \left(\prod_{i=1}^{n} x_{i,j_i} \right). \tag{1.7}$$

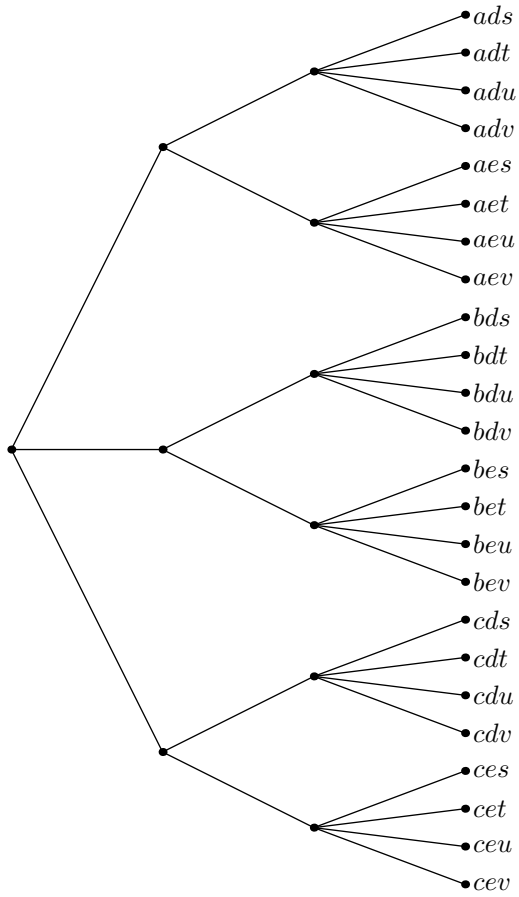


FIGURE 1.1 – Arbre de développement de $(a+b+c)(d+e)(s+t+u+v)$

Voici une traduction sans formules de cet énoncé : « le produit de n sommes finies est égal à la somme de tous les produits de n facteurs que l'on peut former en sélectionnant un facteur parmi les termes de chacune des n sommes ». La figure 1.1 illustre l'application de cette règle au développement de $(a+b+c)(d+e)(s+t+u+v)$. Ici $J_1 = \{1,2,3\}$, $x_{1,1} = a, \dots, x_{1,3} = c$, $J_2 = \{1,2\}$, $x_{2,1} = d$, $x_{2,2} = e$, $J_3 = \{1,\dots,4\}$, $x_{3,1} = s, \dots, x_{3,4} = v$. La proposition 1.22 se prouve facilement par récurrence sur n.

Proposition 1.23 (formule du multinôme)**.** *Pour tous entiers $d \geqslant 2$, $n \geqslant 2$ et tous nombres réels ou complexes $a_1, \dots, a_d$,*

$$\big(a_1 + \cdots + a_d\big)^n = \sum_{\substack{(k_1,\dots,k_d)\in\mathbb{N}^d \\ k_1+\cdots+k_d=n}} \frac{n!}{k_1!\dots k_d!}\, a_1^{k_1} \dots a_d^{k_d}. \tag{1.8}$$

Preuve. On commence par appliquer la formule (1.7) avec $J_1 = \cdots = J_n = \{1, \ldots, d\}$ pour obtenir :

$$\big(a_1+\cdots+a_d\big)^n = \underbrace{\big(a_1 + \cdots + a_d\big) \times \cdots \times \big(a_1 + \cdots + a_d\big)}_{n \text{ parenthèses}} = \sum_{(i_1,\ldots,i_n)\in\{1,\ldots,d\}^n} a_{i_1} \ldots a_{i_n}. \tag{1.9}$$

Parmi les n facteurs du produit $a_{i_1} \ldots a_{i_n}$, il peut y avoir des répétitions et il peut aussi manquer certains des a_i. En regroupant les facteurs identiques, on peut toujours écrire ce produit sous la forme $a_1^{k_1} \ldots a_d^{k_d}$, où pour $j = 1, \ldots, d$, on a noté k_j le nombre de facteurs égaux à a_j dans ce produit (éventuellement $k_j = 0$ si a_j ne figure pas dans le produit). Comme il y a n facteurs, les k_j vérifient $k_1 + \cdots + k_d = n$. Notons maintenant $M(n, k_1, \ldots, k_d)$, le nombre d'apparitions du produit $a_1^{k_1} \ldots a_d^{k_d}$ dans le développement de $(a_1 + \cdots + a_d)^n$. Avec cette notation, on peut réécrire (1.9) sous la forme

$$\big(a_1 + \cdots + a_d\big)^n = \sum_{\substack{(k_1,\ldots,k_d)\in\mathbb{N}^d \\ k_1+\cdots+k_d=n}} M(n, k_1, \ldots, k_d)\, a_1^{k_1} \ldots a_d^{k_d}. \tag{1.10}$$

Il ne nous reste plus qu'à faire un peu de dénombrement pour expliciter le coefficient $M(n, k_1, \ldots, k_d)$ pour $k_1, \ldots, k_d$ entiers fixés tels que $k_1 + \cdots + k_d = n$. Cela revient à compter de combien de façons on peut choisir les k_1 parenthèses fournissant a_1, les k_2 parenthèses fournissant a_2,.., les k_d parenthèses fournissant a_d, pour former le produit $a_1^{k_1} \ldots a_d^{k_d}$. Commençons par choisir les k_1 parenthèses dont la contribution est a_1, ceci peut se faire de $C_n^{k_1}$ façons. Parmi les $n - k_1$ parenthèses restantes, on choisit ensuite les k_2 parenthèses qui fournissent a_2, ce qui nous laisse $C_{n-k_1}^{k_2}$ possibilités. Et ainsi de suite, jusqu'au choix de k_d parenthèses parmi les $n - k_1 - \cdots - k_{d-1}$ dernières [5] pour obtenir a_d. On en déduit que

$$\begin{aligned} M(n, k_1, \ldots, k_d) &= C_n^{k_1} \times C_{n-k_1}^{k_2} \times C_{n-k_1-k_2}^{k_3} \times \cdots \times C_{n-k_1-\cdots-k_{d-1}}^{k_d} \\ &= \frac{n!}{k_1!(n-k_1)!} \times \frac{(n-k_1)!}{k_2!(n-k_1-k_2)!} \times \frac{(n-k_1-k_2)!}{k_3!(n-k_1-k_2-k_3)!} \times \cdots \\ &\quad \cdots \times \frac{(n-k_1-\cdots-k_{d-1})!}{k_d!(n-k_1-\cdots-k_d)!} \\ &= \frac{n!}{k_1!\ldots k_d!}, \end{aligned}$$

après simplifications et en notant que $(n - k_1 - \cdots - k_d)! = 0! = 1$. □

Les nombres $M(n, k_1, \ldots, k_d)$ sont appelés *coefficients multinomiaux*. On peut aussi les interpréter comme

a) le nombre de façons de répartir les éléments d'un ensemble de cardinal n en d sous-ensembles de cardinal respectif $k_1, \ldots, k_d$;

5. Comme $n - k_1 - \cdots - k_{d-1} = k_d$, il n'y a qu'un seul choix possible à ce stade : prendre toutes les parenthèses restantes.

b) le nombre d'applications f de $[\![1,n]\!]$ dans $[\![1,d]\!]$ telles que chaque i dans $[\![1,d]\!]$ ait exactement k_i antécédents dans $[\![1,n]\!]$ par f.

Notons en passant que d'après l'interprétation b), le nombre total d'applications de $[\![1,n]\!]$ dans $[\![1,d]\!]$, qui vaut d^n d'après la proposition 1.17 a), doit être égal à la somme de tous les $M(n,k_1,\dots,k_d)$, ce qui s'écrit :

$$d^n = \sum_{\substack{(k_1,\dots,k_d)\in\mathbb{N}^d \\ k_1+\cdots+k_d=n}} M(n,k_1,\dots,k_d).$$

C'est bien ce que donne la formule du multinôme en prenant tous les a_i égaux à 1 dans (1.8).

Corollaire 1.24 (formule du binôme)**.** *Pour tous réels ou complexes a et b et tout entier $n \geqslant 2$,*

$$(a+b)^n = \sum_{\substack{(k,l)\in\mathbb{N}^2 \\ k+l=n}} \frac{n!}{k!l!}\, a^k b^l = \sum_{k=0}^{n} C_n^k a^k b^{n-k}. \tag{1.11}$$

Preuve. Il suffit d'applique la formule du multinôme avec $d=2$, $a_1=a$, $a_2=b$. □

Remarque 1.25. Dans le calcul des $M(n,k_1,\dots,k_d)$ ci-dessus, il peut sembler que l'on ait avantagé les a_1 en commençant par les choisir en premier et pénalisé les a_d en les gardant pour la fin. Vous pourrez vous convaincre qu'il n'en est rien en refaisant ce calcul en plaçant les a_i dans l'ordre qui vous convient, par exemple en commençant par a_d, ou par a_3,... Pour un calcul alternatif où les a_i jouent un rôle symétrique, voir l'exercice 1.5.

1.3 Dénombrabilité

On peut comparer les ensembles finis par leur nombre d'éléments. Cette notion n'a plus de sens pour des ensembles infinis. Néanmoins on peut généraliser cette comparaison en disant que deux ensembles ont même cardinal s'ils sont en bijection. Ceci permet de comparer les ensembles infinis. Il convient de se méfier de l'intuition courante basée sur les ensembles finis. Par exemple si A et B sont finis et A est inclus strictement dans B, alors $\operatorname{card} A < \operatorname{card} B$ et il n'existe pas de bijection entre A et B. Ceci n'est plus vrai pour les ensembles infinis. Par exemple $\mathbb{N}^*$ est strictement inclus dans $\mathbb{N}$ mais est en bijection avec $\mathbb{N}$ par l'application $f : \mathbb{N} \to \mathbb{N}^*$, $n \mapsto n+1$, donc $\mathbb{N}^*$ et $\mathbb{N}$ ont même cardinal. Nous nous intéressons maintenant aux ensembles ayant même cardinal que $\mathbb{N}$.

Définition 1.26. *Un ensemble est dit* dénombrable *s'il est en bijection avec $\mathbb{N}$. Il est dit* au plus dénombrable *s'il est fini ou dénombrable.*

Exemple 1.27. L'ensemble $2\mathbb{N}$ des entiers pairs est dénombrable. Pour le voir, il suffit de considérer l'application $f : \mathbb{N} \to 2\mathbb{N}$, $n \mapsto 2n$ qui est clairement une bijection. On vérifie de même que l'ensemble des entiers impairs est dénombrable.

Exemple 1.28. L'ensemble $\mathbb{Z}$ est dénombrable. On peut en effet « numéroter » les entiers relatifs par les entiers naturels en s'inspirant du tableau suivant.

$\mathbb{Z}$	$\ldots$	-4	-3	-2	-1	0	$+1$	$+2$	$+3$	$+4$	$\ldots$
$\mathbb{N}$	$\ldots$	8	6	4	2	0	1	3	5	7	$\ldots$

Plus formellement, définissons l'application $f : \mathbb{N} \to \mathbb{Z}$ par

$$\forall n \in \mathbb{N}, \quad f(n) = \begin{cases} \frac{n+1}{2} & \text{si } n \text{ est impair,} \\ \frac{-n}{2} & \text{si } n \text{ est pair.} \end{cases}$$

Pour vérifier que f est une bijection, montrons que pour tout $k \in \mathbb{Z}$ donné, l'équation $f(n) = k$ a une solution unique. Si $k > 0$, il ne peut avoir d'antécédent pair par f puisque f envoie l'ensemble des entiers pairs dans $\mathbb{Z}^-$. L'équation $f(n) = k$ se réduit donc dans ce cas à $\frac{n+1}{2} = k$ qui a pour unique solution $n = 2k - 1$. Si $k \leqslant 0$, il ne peut avoir d'antécédent impair par f et l'équation $f(n) = k$ se réduit à $\frac{-n}{2} = k$ qui a pour unique solution $n = -2k$. Ainsi f est bien une bijection puisque tout élément k de $\mathbb{Z}$ a un antécédent n unique par f. La bijection inverse est donnée par

$$\forall k \in \mathbb{Z}, \quad f^{-1}(k) = \begin{cases} 2k - 1 & \text{si } k > 0, \\ -2k & \text{si } k \leqslant 0. \end{cases}$$

Exemple 1.29. L'ensemble $\mathbb{N}^2$ est dénombrable. Cet exemple relève de la proposition 1.34 ci-dessous, mais la vérification directe est instructive. Voici une façon de construire une bijection $f : \mathbb{N}^2 \to \mathbb{N}$. L'idée est de fabriquer une numérotation des couples de $\mathbb{N}^2$ par les entiers en s'inspirant du schéma de la figure 1.2.

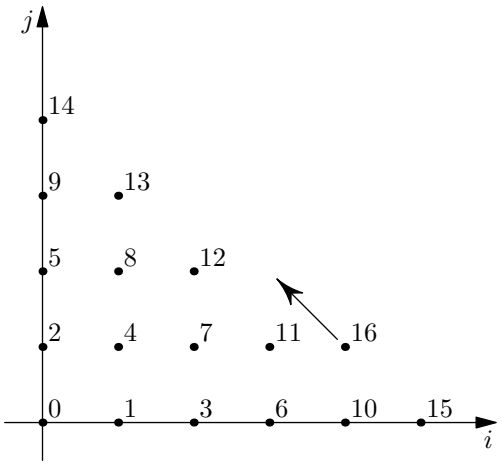


FIGURE 1.2 – Une numérotation des couples (i, j) de $\mathbb{N}^2$

Un peu de dénombrement nous conduit à proposer la définition [6]

$$\forall (i,j) \in \mathbb{N}^2, \quad f(i,j) := \frac{(i+j)(i+j+1)}{2} + j.$$

Preuve. La vérification de la bijectivité de f repose sur la remarque suivante. Définissons la suite $(u_k)_{k\in\mathbb{N}}$ par

$$\forall k \in \mathbb{N}, \quad u_k := \frac{k(k+1)}{2}.$$

Il s'agit clairement d'une suite *strictement croissante d'entiers*, de premier terme $u_0 = 0$. On a donc

$$\forall l \in \mathbb{N},\ \exists! k = k_l \in \mathbb{N}, \quad u_k \leqslant l < u_{k+1}. \tag{1.12}$$

De plus,

$$\text{si } l = f(i,j),\ k_l = i + j \text{ et } l = u_{k_l} + j. \tag{1.13}$$

Surjectivité de f. Soit l quelconque dans $\mathbb{N}$ et $k = k_l$ défini par (1.12). Posons $j = l - u_k$ et $i = k - j$. Alors

$$f(i,j) = \frac{(i+j)(i+j+1)}{2} + j = \frac{k(k+1)}{2} + j = u_k + l - u_k = l.$$

Le couple (i,j) ainsi défini à partir de l est donc antécédent de l par f et comme l était quelconque, f est surjective.

Injectivité de f. Soient (i,j) et (i',j') tels que $f(i,j) = f(i',j')$ et notons l cette valeur commune. D'après (1.13), on a $k_l = i + j = i' + j'$ et $l = u_{k_l} + j = u_{k_l} + j'$. On en déduit immédiatement que $j = j'$, puis que $i = i'$, d'où $(i,j) = (i',j')$, ce qui établit l'injectivité de f. □

Lemme 1.30. *Toute partie infinie de $\mathbb{N}$ est dénombrable.*

Preuve. Soit E une partie infinie de $\mathbb{N}$. On construit une bijection f de $\mathbb{N}$ sur E par récurrence, en utilisant le fait que toute partie non vide de $\mathbb{N}$ admet un plus petit élément. On initialise la récurrence en posant :

$$E_0 := E, \quad f(0) := \min E_0.$$

Ensuite, pour $n \geqslant 1$, si on a défini $f(k)$ et E_k pour $k = 0, \ldots, n-1$, on pose

$$E_n := E \backslash f(\{0, \ldots, n-1\}), \quad f(n) := \min E_n.$$

L'ensemble $f(\{0,\ldots,n-1\}) = \{f(0),\ldots,f(n-1)\}$ est fini donc E_n est non vide puisque E est infini. On peut donc bien construire ainsi de proche en proche tous les $f(n)$ pour $n \in \mathbb{N}$. De plus il est clair par construction que pour tout $n \geqslant 1$, $f(n-1) < f(n)$. L'application f est donc strictement croissante, ce qui entraîne

6. Justification laissée en exercice.

son injectivité. Pour voir qu'elle est surjective, soit m un élément quelconque de E. Comme m est un entier, il n'y a qu'un nombre fini d'entiers strictement inférieurs à m, donc *a fortiori* qu'un nombre fini n d'éléments de E inférieurs strictement à m (éventuellement aucun). Ainsi m est le $(n+1)$-ième plus petit élément de E, d'où $f(n) = m$ (comme on commence à 0, n est le $(n+1)$-ième plus petit entier de $\mathbb{N}$). Nous venons de montrer qu'un élément quelconque de E a au moins un antécédent par f, autrement dit que f est surjective. □

Proposition 1.31. *Toute partie infinie d'un ensemble dénombrable est elle-même dénombrable.*

Preuve. Soit A une partie infinie d'un ensemble dénombrable B. Il existe une bijection $g : B \to \mathbb{N}$. Sa restriction $\tilde{g}$ à A est une bijection de A sur $g(A)$. L'ensemble $g(A)$ est une partie infinie de $\mathbb{N}$, car si elle était finie, il en serait de même pour A. Par le lemme 1.30, il existe une bijection f de $g(A)$ sur $\mathbb{N}$. L'application $f \circ \tilde{g} : A \to \mathbb{N}$ est une bijection comme composée de deux bijections. Donc A est dénombrable. □

Remarque 1.32. La proposition 1.31 nous permet de caractériser les ensembles *au plus* dénombrables comme ceux qui *sont en bijection avec une partie de* $\mathbb{N}$, ou encore comme ceux qui *s'injectent dans* $\mathbb{N}$. De même, les ensembles dénombrables sont les ensembles *infinis* qui s'injectent dans $\mathbb{N}$.

Remarque 1.33. Il résulte immédiatement de la proposition 1.31 que si l'ensemble B contient une partie infinie A non dénombrable, B est lui même infini non dénombrable.

Proposition 1.34. *Le produit cartésien d'une suite* finie *d'ensembles dénombrables est dénombrable.*

Preuve. Notons $E_1, \ldots, E_n$ la suite finie d'ensembles dénombrables considérée. Pour $i \in [\![1, n]\!]$, nous disposons d'une bijection $E_i \to \mathbb{N}$, qui par composition avec la bijection $\mathbb{N} \to \mathbb{N}^*$, $n \mapsto n+1$ donne une bijection $f_i : E_i \to \mathbb{N}^*$. Comme $E = E_1 \times \cdots \times E_n$ est clairement un ensemble infini, il nous suffit de construire une *injection* de E dans $\mathbb{N}$. Pour cela il est commode d'utiliser les nombres premiers dont nous notons $(p_j)_{j \geqslant 1}$ la suite ordonnée : $p_1 = 2$, $p_2 = 3$, $p_3 = 5$, $p_4 = 7$, $p_5 = 11$, … Définissons $f : E \to \mathbb{N}$, par

$$\forall x = (x_1, \ldots, x_n) \in E, \quad f(x) := p_1^{f_1(x_1)} \ldots p_n^{f_n(x_n)} = \prod_{j=1}^{n} p_j^{f_j(x_j)}.$$

Remarquons que pour tout $x \in E$, $f(x) \geqslant 2^{f_1(x_1)} \geqslant 2$.

Pour vérifier l'injectivité de f, soient x et y dans E tels que $f(x) = f(y)$. En vertu de l'unicité de la décomposition en facteurs premiers d'un entier supérieur ou égal à 2, cette égalité équivaut à :

$$\forall i \in [\![1, n]\!], \quad f_i(x_i) = f_i(y_i).$$

Comme chaque f_i est injective, ceci implique $x_i = y_i$ pour tout i, d'où $x = y$. □

Corollaire 1.35. *Pour tout entier $d \geqslant 1$, $\mathbb{N}^d$, $\mathbb{Z}^d$ sont dénombrables. L'ensemble $\mathbb{Q}$ des nombres rationnels est dénombrable (de même que $\mathbb{Q}^d$, $d \geqslant 1$).*

Vérification. La dénombrabilité de $\mathbb{Q}$ s'obtient facilement en l'injectant dans le produit cartésien d'ensembles dénombrables $\mathbb{Z} \times \mathbb{N}^*$ *via* l'unicité de l'écriture en fraction irréductible (avec dénominateur positif) d'un rationnel. Les autres affirmations du corollaire découlent immédiatement de la Proposition 1.34. □

Voici un premier exemple d'ensemble infini non dénombrable.

Proposition 1.36. *L'ensemble $\{0,1\}^{\mathbb{N}}$ des suites infinies de 0 ou de 1 est infini mais* n'est pas *dénombrable.*

Il est clair que $\{0,1\}^{\mathbb{N}}$ est un ensemble infini, puisque qu'il contient un sous-ensemble en bijection avec $\mathbb{N}$, par exemple l'ensemble de suites $\left\{\left(\mathbf{1}_{\{n\}}(k)\right)_{k\in\mathbb{N}},\, n \in \mathbb{N}\right\}$.

Preuve. Supposons que $\{0,1\}^{\mathbb{N}}$ soit dénombrable, on peut alors numéroter ses éléments par les entiers, de sorte que $\{0,1\}^{\mathbb{N}} = \{x_n \ ; \ n \in \mathbb{N}\}$, chaque x_n étant une suite $(x_{n,k})_{k\in\mathbb{N}}$ de chiffres binaires. Construisons alors la suite $y = (y_k)_{k\in\mathbb{N}}$ de chiffres binaires en posant :

$$\forall k \in \mathbb{N}, \quad y_k = 1 - x_{k,k} = \begin{cases} 1 & \text{si } x_{k,k} = 0 \\ 0 & \text{si } x_{k,k} = 1. \end{cases}$$

Alors par construction, la suite y *diffère de chaque suite* x_n (au moins par son n^{e} terme). Or y est un élément de $\{0,1\}^{\mathbb{N}}$, donc la numérotation considérée ne peut être surjective. □

Corollaire 1.37. *$\mathcal{P}(\mathbb{N})$ n'est pas dénombrable. Le segment $[0,1]$ de $\mathbb{R}$ n'est pas dénombrable. $\mathbb{R}$ n'est pas dénombrable, $\mathbb{C}$ n'est pas dénombrable. Un intervalle de $\mathbb{R}$ est soit infini non dénombrable, soit réduit à un singleton, soit vide.*

Preuve. Comme $\mathcal{P}(\mathbb{N})$ est en bijection avec $\{0,1\}^{\mathbb{N}}$ par l'application $A \mapsto \mathbf{1}_A$, $\mathcal{P}(\mathbb{N})$ est infini non dénombrable. D'après la remarque 1.33, la non dénombrabilité de $\mathbb{R}$ ou de $\mathbb{C}$ résulte immédiatement de celle de $[0,1]$. Pour vérifier cette dernière, nous utilisons à nouveau la remarque 1.33 en construisant une partie de $[0,1]$ en bijection avec $\{0,1\}^{\mathbb{N}}$. La première idée qui vient à l'esprit pour une telle construction est d'utiliser le développement des nombres réels en base 2. Mais pour éviter les difficultés techniques liées à l'existence de développements propre et impropre pour les nombres de la forme $k2^{-n}$, nous utiliserons plutôt la base 3 en ne conservant que les chiffres binaires 0 et 1. Définissons donc [7]

$$f : \{0,1\}^{\mathbb{N}} \to [0,1], \quad u = (u_k)_{k\in\mathbb{N}} \longmapsto f(u) := \sum_{k=0}^{+\infty} \frac{u_k}{3^{k+1}}.$$

7. La lecture de ce qui suit requiert la connaissance des séries dont les principales propriétés sont rappelées section 2.1 ci-après, voir notamment les séries géométriques (exemple 2.8).

Comme les u_k ne peuvent prendre que les valeurs 0 ou 1, la série à termes positifs définissant $f(u)$ converge puisque son terme général vérifie l'encadrement $0 \leqslant u_k 3^{-k-1} \leqslant 3^{-k-1}$. Sa somme $f(u)$ vérifie donc

$$0 \leqslant f(u) \leqslant \sum_{k=0}^{+\infty} \frac{1}{3^{k+1}} = \frac{1}{3} \frac{1}{1-\frac{1}{3}} = \frac{1}{2}.$$

Ainsi f est bien une application de $\{0,1\}^{\mathbb{N}}$ dans $[0,1]$ et d'après la remarque 1.5, il suffit de vérifier son injectivité pour qu'elle réalise une bijection de $\{0,1\}^{\mathbb{N}}$ sur son image $A := \{f(u) \ ; \ u \in \{0,1\}^{\mathbb{N}}\}$. Par la proposition 1.36, on en déduira la non dénombrabilité de A.

Pour montrer l'injectivité de f, supposons qu'il existe deux suites $u \neq u'$ éléments de $\{0,1\}^{\mathbb{N}}$ telles que $f(u) = f(u')$. Comme $u \neq u'$, l'ensemble des entiers k tels que $u_k \neq u'_k$ est non vide et a donc un plus petit élément que nous notons j. On a ainsi $u_k = u'_k$ pour tout $k < j$ et $u_j \neq u'_j$. Quitte à permuter u et u', on ne perd pas de généralité en supposant que $u_j = 1$ et $u'_j = 0$. L'égalité $f(u) = f(u')$ implique alors :

$$\frac{1}{3^{j+1}} = \sum_{k=j+1}^{+\infty} \frac{u'_k - u_k}{3^{k+1}}.$$

Comme $u'_k - u_k$ ne peut prendre que les valeurs -1, 0 ou 1, il est majoré par 1, d'où :

$$\frac{1}{3^{j+1}} \leqslant \sum_{k=j+1}^{+\infty} \frac{1}{3^{k+1}} = \frac{1}{3^{j+2}} \frac{1}{1-\frac{1}{3}} = \frac{1}{2 \times 3^{j+1}},$$

ce qui est impossible. On en déduit que si $f(u) = f(u')$, nécessairement $u = u'$, ce qui établit l'injectivité de f.

Pour vérifier la non dénombrabilité d'un intervalle non vide et non réduit à un singleton de $\mathbb{R}$, il suffit de remarquer que cet intervalle a au moins deux éléments a et b et qu'il contient alors $[a,b]$. Il suffit maintenant de construire une bijection $[0,1] \to [a,b]$. L'application

$$f : [0,1] \to [a,b], \quad t \mapsto ta + (1-t)b,$$

fait l'affaire. □

Remarque 1.38. Il est facile de construire une bijection entre $]-1,1[$ et $\mathbb{R}$, par exemple $x \mapsto \tan(\pi x/2)$ et d'en déduire une bijection entre $\mathbb{R}$ et n'importe quel intervalle ouvert non vide. En fait on peut montrer que les ensembles suivants ont tous même cardinal : $\{0,1\}^{\mathbb{N}}$, $\mathbb{R}$, $\mathbb{C}$, tout intervalle non vide et non réduit à un singleton de $\mathbb{R}$. On dit qu'ils ont *la puissance du continu*. Sans aller jusqu'à démontrer complètement cette affirmation, nous nous contenterons de présenter ci-dessous une construction explicite d'une bijection entre $\{0,1\}^{\mathbb{N}^*}$ et $[0,1[$. Il est clair que $\{0,1\}^{\mathbb{N}^*}$ est lui même en bijection avec $\{0,1\}^{\mathbb{N}}$ (pourquoi ?).

Exemple 1.39 (une bijection entre $\{0,1\}^{\mathbb{N}^*}$ et $[0,1[$)**.** Commençons par un rappel sur le développement en base 2 des réels de $[0,1[$, c'est-à-dire l'existence pour chaque $x \in [0,1[$ d'une suite $(a_k)_{k\geqslant 1} \in \{0,1\}^{\mathbb{N}^*}$ telle que

$$x = \sum_{k=1}^{+\infty} \frac{a_k}{2^k}. \tag{1.14}$$

On appelle *nombre dyadique* de $]0,1[$, tout $x \in]0,1[$ de la forme $k2^{-n}$ avec $k \in \mathbb{N}$ et $n \in \mathbb{N}^*$. Un tel dyadique admet une écriture irréductible unique de la forme $l2^{-n}$ avec l impair. Notons Δ l'ensemble des dyadiques de $]0,1[$.

Si $x \in]0,1[\setminus \Delta$, il existe une unique suite $(a_k)_{k\geqslant 1} \in \{0,1\}^{\mathbb{N}^*}$ vérifiant (1.14). De plus cette suite $(a_k)_{k\geqslant 1}$ comporte à la fois une infinité de 0 *et* une infinité de 1.

Si $x \in \Delta$, il existe deux suites $(a_k)_{k\geqslant 1} \in \{0,1\}^{\mathbb{N}^*}$ et $(a'_k)_{k\geqslant 1} \in \{0,1\}^{\mathbb{N}^*}$ vérifiant (1.14). L'une $(a_k)_{k\geqslant 1}$ a tous ses termes nuls à partir d'un certain rang. C'est le *développement propre* du dyadique x. L'autre a tous ses termes égaux à 1 à partir d'un certain rang, c'est le *développement impropre* de x. Nous noterons $p(x)$ le développement propre de x et $i(x)$ son développement impropre. Par exemple le dyadique $\frac{3}{8}$ a pour développement propre $(0,1,1,0,0,0,0,\dots)$ car $\frac{3}{8} = \frac{1}{4} + \frac{1}{8}$. Son développement impropre est $(0,1,0,1,1,1,1,\dots)$ puisque $\frac{1}{8} = \sum_{k=4}^{+\infty} 2^{-k}$.

Définissons maintenant $f : [0,1[\to \{0,1\}^{\mathbb{N}^*}$ comme suit.

- Si $x \in]0,1[\setminus \Delta$, on prend pour $f(x)$ l'unique suite $(a_k)_{k\geqslant 1} \in \{0,1\}^{\mathbb{N}^*}$ vérifiant (1.14).
- Si $x \in \Delta$ et $x \neq 1/2$, il existe une écriture unique $x = l2^{-n}$ avec l *impair*. On pose alors

$$f\Big(\frac{l}{2^n}\Big) = \begin{cases} i\Big(\dfrac{l+1}{2^n}\Big) & \text{si } l = 1 \mod 4, \\ p\Big(\dfrac{l-1}{2^n}\Big) & \text{si } l = 3 \mod 4. \end{cases} \tag{1.15}$$

- Pour $f(0)$ on prend la suite ne comportant que des 0 et pour $f(1/2)$ la suite ne comportant que des 1.

Le mécanisme de construction de f sur les dyadiques autres que 0 et 1/2 peut être décrit de manière informelle à l'aide de l'arbre binaire de la figure 1.3. Chaque individu de cet arbre hérite un développement binaire de son ascendant direct et engendre lui même deux enfants et deux développements binaires. L'enfant de gauche hérite du développement impropre et celui de droite du développement propre. Pour voir comment cela conduit à la formule (1.15), on remarque qu'au niveau $n-1$, le dyadique de forme irréductible $j2^{-(n-1)}$ a pour enfant de gauche $(2j-1)2^{-n}$ et pour enfant de droite $(2j+1)2^{-n}$. Comme j est impair, $2j-1 = 1 \mod 4$ et $2j+1 = 3 \mod 4$. La valeur modulo 4 de l dans (1.15) permet ainsi de savoir si $l2^{-n}$ est un enfant de gauche ou de droite de son parent.

La compréhension de la définition de f demandant plus d'effort que la vérification de sa bijectivité, ce dernier point est laissé au lecteur.

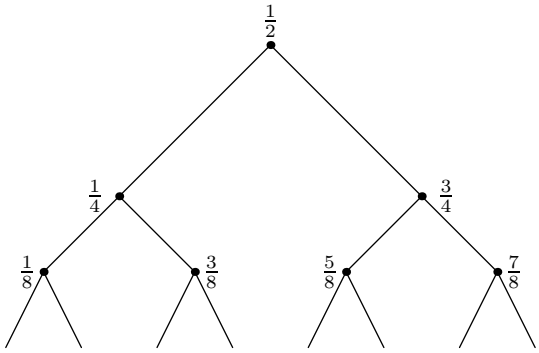


FIGURE 1.3 – Arbre binaire des dyadiques de $]0,1[$

Proposition 1.40. *Soit J un ensemble au plus dénombrable d'indices et pour tout $j \in J$, soit A_j un ensemble au plus dénombrable. Alors $A := \bigcup_{j\in J} A_j$ est au plus dénombrable.*

Preuve. Le cas $J = \varnothing$ est trivial puisqu'alors $A = \varnothing$. Si tous les A_j sont vides, A l'est aussi (quel que soit J). On suppose désormais que J n'est pas vide et qu'au moins un des A_j est non vide. On va montrer que A est au plus dénombrable en construisant une injection h de A dans $\mathbb{N}$.

On peut traduire les hypothèses en écrivant qu'il existe une injection [8] $f : J \to \mathbb{N}^*$ et que pour tout j tel que $A_j \neq \varnothing$, il existe une injection $g_j : A_j \to \mathbb{N}^*$. Admettons pour un instant que l'on peut construire une famille $(A'_j)_{j\in J}$ d'ensembles *deux à deux disjoints* tels que pour tout $j \in J$, $A'_j \subset A_j$ et que

$$A = \bigcup_{j\in J} A_j = \bigcup_{j\in J} A'_j.$$

On définit alors l'application $h : A \to \mathbb{N}$ comme suit. Si $x \in A$, il existe un *unique* $j \in J$ tel que $x \in A'_j$. On pose alors :

$$h(x) := p_{f(j)}^{g_j(x)},$$

où p_k désigne le k-ième nombre premier. Remarquons que $h(x) \geqslant 2$ puisque $f(j) \geqslant 1$, la suite (p_k) est croissante de premier terme $p_1 = 2$ et $g_j(x) \geqslant 1$.

Pour vérifier l'injectivité de h, soient x et y dans A tels que $h(x) = h(y)$. En notant l l'unique indice tel que $y \in A'_l$, cette égalité s'écrit

$$p_{f(j)}^{g_j(x)} = p_{f(l)}^{g_l(y)}.$$

8. Toute injection d'un ensemble dans $\mathbb{N}$ peut se transformer en injection du même ensemble dans $\mathbb{N}^*$ par composition avec la bijection $\mathbb{N} \to \mathbb{N}^*$, $n \mapsto n + 1$.

En raison de l'unicité de la décomposition d'un entier $n \geqslant 2$ en facteurs premiers, ceci entraîne $f(j) = f(l)$ et $g_j(x) = g_l(y)$, puis par injectivité de f, $j = l$ et $g_j(x) = g_j(y)$. Comme g_j est injective, on en déduit que $x = y$. L'injectivité de h est ainsi établie.

Il reste à justifier la construction de la famille $(A'_j)_{j\in J}$. On commence par transporter l'ordre de $\mathbb{N}$ sur J *via* l'application f en écrivant pour $j, l \in J$, que $j \leqslant l$ signifie que $f(j) \leqslant f(l)$. Ceci nous permet de noter les éléments de J en suivant cet ordre $J = \{j_0, j_1, \ldots\}$, où $j_0 \leqslant j_1 \leqslant \ldots$. Ensuite on construit A'_j par récurrence en posant

$$\begin{array}{ll} A'_{j_0} := A_{j_0}, & B_{j_0} := A_{j_0} \\ A'_{j_1} := A_{j_1} \cap (A \backslash B_{j_0}), & B_{j_1} := B_{j_0} \cup A_{j_1} \\ A'_{j_2} := A_{j_2} \cap (A \backslash B_{j_1}), & B_{j_2} := B_{j_1} \cup A_{j_2} \\ \ldots\ldots & \ldots\ldots \end{array}$$

Les détails de la vérification sont laissés au lecteur. □

Proposition 1.41 (dénombrabilité par image surjective)**.** *Soient A et B deux ensembles tels qu'il existe une surjection f de A sur B. Alors si A est dénombrable, B est au plus dénombrable. Si A est fini, B est fini et* $\operatorname{card} B \leqslant \operatorname{card} A$.

Il est possible que A soit infini et B fini, un exemple évident étant fourni par l'application nulle $f : x \mapsto 0$ de $A = \mathbb{N}$ dans $B = \{0\}$ qui est clairement une surjection.

Preuve. Supposons A dénombrable. Définissons sur A la relation d'équivalence $x \sim x'$ si $f(x) = f(x')$. Les classes d'équivalences pour cette relation réalisent une partition de A. Dans chaque classe d'équivalence, choisissons un représentant particulier [9]. Soit A_0 la partie de A formée de tous les représentants ainsi choisis. Si x et x' sont deux éléments distincts de A_0, ils sont dans deux classes c et c' disjointes, donc $f(x) \neq f(x')$. La restriction de f à A_0 est donc injective. Par ailleurs, puisque f est surjective, tout $y \in B$ a au moins un antécédent x dans A. Si c est la classe de x, il y a dans cette classe un (unique) élément de A_0 qui a lui aussi y pour image par f. Donc la restriction de f à A_0 reste surjective et c'est finalement une bijection de A_0 sur B. L'ensemble B est en bijection avec une partie A_0 de l'ensemble dénombrable A, il est donc au plus dénombrable.

Le cas A fini est une adaptation facile de ce qui précède. □

9. Par exemple en fixant une numérotation de A par les entiers $(k \mapsto x_k)$ et en décidant de prendre dans la classe c l'élément x_k d'indice k minimal. On évite ainsi l'invocation de l'axiome du choix...

1.4 Exercices

Ex. 1.1 Soient f et g deux applications de Ω dans $\mathbb{R}$. Vérifiez les affirmations suivantes.

1) $\big\{\omega \in \Omega\ ;\ \max\big(f(\omega), g(\omega)\big) \geqslant \varepsilon\big\} = \Big\{\omega \in \Omega\ ;\ f(\omega) \geqslant \varepsilon\Big\} \cup \Big\{\omega \in \Omega\ ;\ g(\omega) \geqslant \varepsilon\Big\}$.

2) $\big\{\omega \in \Omega\ ;\ \min\big(f(\omega), g(\omega)\big) \geqslant \varepsilon\big\} = \Big\{\omega \in \Omega\ ;\ f(\omega) \geqslant \varepsilon\Big\} \cap \Big\{\omega \in \Omega\ ;\ g(\omega) \geqslant \varepsilon\Big\}$.

3) $\big\{\omega \in \Omega\ ;\ f(\omega) + g(\omega) \geqslant \varepsilon\big\} \subset \Big\{\omega \in \Omega\ ;\ f(\omega) \geqslant \dfrac{\varepsilon}{2}\Big\} \cup \Big\{\omega \in \Omega\ ;\ g(\omega) \geqslant \dfrac{\varepsilon}{2}\Big\}$.

Ex. 1.2 *Réunions et intersections dénombrables d'intervalles*

Soient a, b deux réels ($a < b$). Chacun des ensembles E_i ci-dessous est un intervalle. Identifiez le et justifiez votre réponse.

$$E_1 := \bigcap_{n\geqslant 1}\Big[a, b + \frac{1}{n}\Big[, \qquad E_2 := \bigcap_{n\geqslant 1}\Big]a - \frac{1}{n}, b + 2^{-n}\Big],$$

$$E_3 := \bigcup_{n\geqslant 1}\Big[a + \frac{1}{n}, b - \frac{1}{n}\Big], \qquad E_4 := \bigcup_{n\geqslant 1}\Big]a + \frac{1}{n}, b + \frac{1}{n}\Big].$$

Ex. 1.3 Soit f une fonction $\Omega \to \mathbb{R}$. Justifiez l'égalité :

$$\big\{\omega \in \Omega\ ;\ f(\omega) > 0\big\} = \bigcup_{n\geqslant 1}\big\{\omega \in \Omega\ ;\ f(\omega) \geqslant 1/n\big\}.$$

Ex. 1.4 *Écriture ensembliste d'une convergence*

On considère une suite de fonctions $(f_n)_{n\geqslant 1} : \Omega \to \mathbb{R}, \quad \omega \mapsto f_n(\omega)$. On pose pour tout $\varepsilon > 0$ et tout $k \in \mathbb{N}^*$:

$$B_{\varepsilon,k} = \{\omega \in \Omega\ ;\ |f_k(\omega)| < \varepsilon\}.$$

1) On fixe $\varepsilon > 0$. Ecrire à l'aide d'opérations ensemblistes sur les $B_{\varepsilon,k}$, l'ensemble :

$$A_{\varepsilon,n} = \{\omega \in \Omega\ ;\ \forall k \geqslant n,\ |f_k(\omega)| < \varepsilon\}.$$

2) Même question pour :

$$A_{\varepsilon} = \{\omega \in \Omega\ ;\ \exists n(\omega),\ \forall k \geqslant n(\omega),\ |f_k(\omega)| < \varepsilon\}.$$

3) Montrer que l'ensemble :

$$A = \{\omega \in \Omega\ ;\ \lim_{k\to+\infty} f_k(\omega) = 0\}$$

peut s'écrire à l'aide d'opérations ensemblistes sur *une suite* d'ensembles du type $B_{\varepsilon,k}$.

Ex. 1.5 *Calcul des coefficients multinomiaux*

Cet exercice propose une méthode alternative à celle donnée dans la preuve de la proposition 1.23 pour le calcul du coefficient multinomial $M(n, k_1, \dots, k_d)$. On considère l'ensemble $\mathfrak{S}_n$ de toutes les permutations de $[\![1, n]\!]$. Son cardinal est $n!$. Fixons $k_1, \dots, k_d$ entiers de somme n. On pose $E_i =]\!]s_{i-1}, s_i]\!]$, avec $s_0 := 0$ et $s_i := \sum_{j=1}^{i} k_j$ pour $i = 1, \dots, d$. Ainsi les E_i sont deux à deux disjoints, $\operatorname{card} E_i = k_i$ et $\bigcup_{i=1}^{d} E_i = [\![1, n]\!]$.

1) Soit $(F_1, \dots, F_d)$ un d-uple de parties de $[\![1, n]\!]$, deux à deux disjointes, de réunion $[\![1, n]\!]$ et telles que $\operatorname{card} F_i = k_i$, pour tout $i = 1, \dots, d$. Combien y a-t-il de bijections $f \in \mathfrak{S}_n$ vérifiant $f(E_i) = F_i$ pour tout $i = 1, \dots, d$?

2) En déduire que $n! = M(n, k_1, \dots, k_d) k_1! \dots k_d!$.

Ex. 1.6 *Décompositions d'un entier en sommes*

Le but de cet exercice est d'établir les deux formules de dénombrement suivantes. Pour n, $r > 1$ entiers donnés, considérons l'équation à r inconnues :

$$x_1 + x_2 + \cdots + x_r = n. \tag{1.16}$$

Une solution de cette équation est un r-uplet $(x_1, \dots, x_r)$ dont la somme des composantes x_i $(1 \leqslant i \leqslant r)$ vaut n. Lorsque les x_i sont des réels, (1.16) a une infinité de solutions. On s'intéresse aux solutions dont *toutes* les composantes sont des entiers.

Proposition 1.42.

a) L'équation (1.16) admet exactement $C_{n+r-1}^{r-1} = C_{n+r-1}^{n}$ *solutions à composantes entières* positives ou nulles *;*

b) si $n \geqslant r > 1$*, l'équation (1.16) admet exactement* C_{n-1}^{r-1} *solutions à composantes entières* strictement positives*.*

Voici une traduction plus concrète du problème. On dispose de n jetons *identiques* et de r boîtes *numérotées* de 1 à r. Le cas a) donne le nombre de façons de répartir ces jetons dans les r boîtes (certaines pouvant rester vides). Le cas b) donne le nombre de répartitions pour lesquelles aucune boîte ne reste vide. Dans cette interprétation, x_i représente le nombre de jetons déposés dans la boîte n° i.

1) Démontrez la Proposition 1.42 b) en vous aidant du codage des répartitions possibles par des chaînes de caractères, illustré par l'exemple suivant avec $n = 9$ et $r = 4$. On représente d'abord chaque jeton par le caractère « O », *avec un espace* entre deux caractères consécutifs :

O O O O O O O O O.

Pour représenter les 4 boîtes, il suffit d'insérer 3 caractères « | », chacun dans un espace libre. Par exemple, la chaîne

O O|O O O|O|O O O

code la répartition avec 2 jetons dans la première boîte, 3 dans la deuxième, 1 dans la troisième et 3 dans la quatrième, autrement dit la solution $(x_1, x_2, x_3, x_4) = (2, 3, 1, 3)$.

2) Prouvez la Proposition 1.42 a), en utilisant le codage suivant illustré à nouveau dans le cas $n = 9$ et $r = 4$. Le caractère « | » représente à la fois le début et la fin d'une boîte (sauf s'il est le premier ou le dernier de la chaîne). Il nous en faut donc 5 pour représenter 4 boîtes. On représente encore chaque jeton par un « O », mais cette fois on ne laisse plus d'espace entre les caractères de la chaîne. L'exemple de la question précédente sera alors codé par la chaîne de 14 caractères sans espaces :

$$|\mathsf{OO}|\mathsf{OOO}|\mathsf{O}|\mathsf{OOO}|.$$

Ceci permet de représenter une boîte vide par deux caractères « | » consécutifs. Par exemple les chaînes

$$|\mathsf{OO}||\mathsf{OOOO}|\mathsf{OOO}|, \qquad |||\mathsf{OOOO}|\mathsf{OOOOO}|,$$

représentent respectivement les solutions $(2, 0, 4, 3)$ et $(0, 0, 4, 5)$.

3) Montrer que l'on peut passer du cas a) au cas b) par un simple changement d'inconnue et utiliser cette remarque pour contrôler les résultats précédents.

Solution p. 413

Ex. 1.7 *En attendant le métro*

Dans une station de métro est disposée une rangée de 16 sièges. Un sociologue observe la répartition des voyageurs assis sur cette rangée. Il représente son observation par une chaîne de caractères « O » (pour « occupé ») et « L » (pour « libre ») :

$$\mathsf{OLOLLOLOLLLOLOLO}.$$

Notre sociologue se demande si les voyageurs ont choisi leur place au hasard parmi les places disponibles, ou s'ils ont cherché à s'isoler. Si les voyageurs choisissent au hasard parmi les places disponibles, toutes les répartitions des 7 personnes sur les 16 places sont équiprobables. Pour se faire une idée, on compte les *séries*. Une série est ici un bloc de lettres identiques encadré (sauf s'il est en début ou en fin) par des lettres différentes. Dans l'observation réalisée, il y donc 13 séries :

$$\mathsf{O},\ \mathsf{L},\ \mathsf{O},\ \mathsf{LL},\ \mathsf{O},\ \mathsf{L},\ \mathsf{O},\ \mathsf{LLL},\ \mathsf{O},\ \mathsf{L},\ \mathsf{O},\ \mathsf{L},\ \mathsf{O},$$

dont 6 séries de L et 7 séries de O. Le nombre de séries de O est ici maximal puisqu'il n'y a que 7 personnes à répartir.

1) Une chaîne comportant m lettres L et n lettres O est construite au hasard. On se propose de calculer la probabilité qu'elle comporte exactement r séries de O. On pourra utiliser les résultats de l'exercice 1 comme suit. Pour $i = 1, \dots, r$, notons x_i le nombre de lettres O de la i-ème série de O (tous les x_i sont strictement positifs). Soit y_1 le nombre (éventuellement nul) de L avant la première série de O, y_{r+1} le nombre (éventuellement nul) de L après la dernière série de O. Pour $2 \leqslant i \leqslant r$, y_i désigne le nombre strictement positif de lettres L dans la série qui précède la i-ème série de O.

$$\underbrace{\mathsf{L}\dots\mathsf{L}}_{y_1}\ \underbrace{\mathsf{O}\dots\mathsf{O}}_{x_1}\ \underbrace{\mathsf{L}\dots\mathsf{L}}_{y_2}\ \underbrace{\mathsf{O}\dots\mathsf{O}}_{x_2}\ \dots\dots\ \underbrace{\mathsf{L}\dots\mathsf{L}}_{y_r}\ \underbrace{\mathsf{O}\dots\mathsf{O}}_{x_r}\ \underbrace{\mathsf{L}\dots\mathsf{L}}_{y_{r+1}}.$$

Les r-uplets $(x_1, \ldots, x_r)$ possibles sont les solutions à composantes strictement positives de l'équation $x_1 + \ldots x_r = n$. On se ramène à une condition du même type pour les y_i en faisant le changement de variable :

$$z_1 = y_1 + 1, \quad z_{r+1} = y_r + 1, \qquad z_i = y_i \ (2 \leqslant i \leqslant r).$$

Montrer que pour $1 \leqslant r \leqslant n$ et $1 \leqslant r \leqslant m + 1$, la probabilité p_r d'avoir r séries de O est

$$p_r = \frac{C_{m+1}^r C_{n-1}^{r-1}}{C_{m+n}^n}.$$

2) Pouvez vous aider le sociologue ?

Solution p. 415

Ex. 1.8 Montrer que l'ensemble des nombres irrationnels n'est pas dénombrable.

Ex. 1.9 *Dénombrabilité d'ensembles classiques*

1) Montrer que l'ensemble $\mathbb{D}$ des nombres décimaux (c'est-à-dire de la forme $k \times 10^{-n}$, $k \in \mathbb{Z}$, $n \in \mathbb{N}$) est dénombrable.

Les ensembles suivants sont-ils dénombrables ?

2) l'ensemble des polynômes à coefficients entiers.

3) l'ensemble des nombres algébriques (un nombre réel ou complexe est dit *algébrique* s'il est racine d'un polynôme à coefficients entiers).

Ex. 1.10 *Points de discontinuité d'une fonction monotone*

Montrer que l'ensemble des points de discontinuité d'une fonction monotone f est au plus dénombrable. *Indications :* on peut se ramener au cas où f est définie et croissante sur l'intervalle $[a, b]$; que peut-on dire de l'ensemble $D_\varepsilon(f)$ des points où f présente un saut d'amplitude supérieure ou égale à ε ?

Ex. 1.11 Montrez que toute famille d'intervalles ouverts de $\mathbb{R}$ non vides et deux à deux disjoints est au plus dénombrable.

Indication p. 416

Ex. 1.12 Existe-t-il des droites du plan $\mathbb{R}^2$ qui ne contiennent pas au moins deux points à coordonnées rationnelles ?

Indications p. 416

Chapitre 2

Sommes

Après le comptage, l'addition est une deuxième activité élémentaire à la base des mathématiques. Là aussi, les choses sont moins simples qu'il n'y paraît. L'extension de la notion de somme d'une suite finie de nombres à une suite infinie conduit naturellement à la notion de *série* que nous révisons ci-dessous. La théorie des probabilités utilise implicitement une notion plus générale, celle de *famille sommable*. Il s'agit de définir la somme, si elle existe, d'une famille de nombres indexée par un ensemble infini qui n'est pas forcément $\mathbb{N}$ ou $\mathbb{N}^*$. Nous présentons cette théorie dans la deuxième partie du chapitre. Ceci permet entre autres, de traiter proprement des *séries doubles*.

2.1 Rappels sur les séries

Dans cette section nous rappelons sans démonstrations les points essentiels de la théorie des séries numériques vue en deuxième année. Nous détaillerons seulement la question de la convergence commutative en raison de son rôle dans la théorie des familles sommables.

2.1.1 Généralités

Définition 2.1. *Soit $(u_k)_{k\in\mathbb{N}}$ une suite de nombres réels ou complexes. Pour tout $n \in \mathbb{N}$, on appelle somme partielle de rang n de cette suite, le nombre*

$$S_n := \sum_{k=0}^{n} u_k.$$

- *Si S_n tend vers une limite* finie *S quand n tend vers $+\infty$, on dit que la série de terme général u_k converge et a pour somme S, ce que l'on écrit*

$$S = \sum_{k=0}^{+\infty} u_k := \lim_{n\to+\infty} S_n.$$

Dans ce cas, $R_n := S - S_n$ est appelé reste de rang n de la série.

- *Si S_n n'a pas de limite finie (c'est-à-dire tend vers $-\infty$ ou $+\infty$ ou n'a pas de limite du tout), on dit que la série diverge.*

Remarque 2.2. Par un abus d'écriture courant, on désigne la série (convergente ou divergente) par la notation « $\sum_{k=0}^{+\infty} u_k$ », mais on ne peut faire intervenir cette expression dans des calculs que si elle représente vraiment un nombre, c'est-à-dire si la série converge.

Remarque 2.3. Si les u_k sont complexes, posons $x_k := \operatorname{Re} u_k$ et $y_k := \operatorname{Im} u_k$, $S'_n := \operatorname{Re} S_n$ et $S''_n := \operatorname{Im} S_n$. La suite S_n converge dans $\mathbb{C}$ vers S si et seulement si S'_n et S''_n convergent dans $\mathbb{R}$ vers respectivement $S' := \operatorname{Re} S$ et $S'' := \operatorname{Im} S$. On en déduit immédiatement que la série de terme général complexe $u_k = x_k + \mathrm{i}y_k$ converge si et seulement si les séries de terme général x_k et y_k convergent dans $\mathbb{R}$ et que dans ce cas

$$\sum_{k=0}^{+\infty} (x_k + \mathrm{i}y_k) = \sum_{k=0}^{+\infty} x_k + \mathrm{i} \sum_{k=0}^{+\infty} y_k.$$

Remarque 2.4. La définition 2.1 se généralise immédiatement au cas où les u_k sont éléments d'un espace vectoriel normé E, la convergence dans E de S_n vers le vecteur S signifiant alors $\lim_{n\to+\infty} \|S - S_n\| = 0$.

Proposition 2.5. *Si une série converge, son terme général u_n tend vers 0 quand n tend vers l'infini.*

Remarque 2.6. La réciproque de la proposition 2.5 est fausse. Un contre exemple bien connu est la *série harmonique* de terme général $u_k = 1/k$ pour $k \geqslant 1$ (et $u_0 = 0$).

Remarque 2.7. La contraposée de la proposition 2.5 s'énonce : si u_n ne tend pas vers 0 quand n tend vers l'infini, alors la série diverge. On parle dans ce cas de divergence grossière de la série.

Exemple 2.8 (séries géométriques)**.** Soit q un nombre réel ou complexe. La série géométrique standard $\sum_{k=0}^{+\infty} q^k$ converge si et seulement si $|q| < 1$. Dans ce cas, sa somme est donnée par

$$\sum_{k=0}^{+\infty} q^k = \frac{1}{1-q}.$$

Cette formule permet de calculer la somme de n'importe quelle série géométrique convergente (donc de raison q vérifiant $|q| < 1$). Il suffit de mettre en facteur le premier terme. Si $(u_k)_{k\in\mathbb{N}}$ est une suite géométrique de raison q (c'est-à-dire $u_{k+1} = qu_k$ pour tout k avec q ne dépendant pas de k), vérifiant $|q| < 1$, alors pour tout $j \in \mathbb{N}$,

$$\sum_{k=j}^{+\infty} u_k = \sum_{k=j}^{+\infty} u_j q^{k-j} = u_j \sum_{l=0}^{+\infty} q^l = \frac{u_j}{1-q}.$$

Théorème 2.9 (critère de Cauchy). *La série à termes réels ou complexes* $\sum_{k=0}^{+\infty} u_k$ *converge si et seulement si*

$$\forall \varepsilon > 0,\ \exists N \in \mathbb{N},\ \forall n, m \geqslant N, \quad |S_n - S_m| < \varepsilon.$$

Ce théorème n'est que l'application du critère de Cauchy à la suite des sommes partielles $S_n = \sum_{k=0}^{n} u_k$. Il se généralise immédiatement au cas où les termes u_k sont des éléments d'un espace vectoriel normé *complet* [1], en remplaçant $|S_n - S_m| < \varepsilon$ par $\|S_n - S_m\| < \varepsilon$.

Corollaire 2.10 (convergence absolue). *Si la série* $\sum_{k=0}^{+\infty} |u_k|$ *converge, alors la série* $\sum_{k=0}^{+\infty} u_k$ *converge aussi. On dit qu'elle est absolument convergente.*

Le résultat s'étend immédiatement aux séries à termes dans un espace vectoriel normé *complet* en remplaçant $|u_k|$ par $\|u_k\|$. On parle alors de convergence *normale*.

Remarque 2.11. La réciproque du corollaire 2.10 est fausse. Un contre exemple bien connu est la série alternée de terme général $u_k = (-1)^k/k$ $(k \geqslant 1)$ qui converge alors que la série des valeurs absolues est la série harmonique qui diverge.

2.1.2 Séries à termes positifs

Si tous les u_k sont positifs, $(S_n)_{n\in\mathbb{N}}$ est une suite *croissante* car pour tout $n \geqslant 1$, $S_n - S_{n-1} = u_n \geqslant 0$. Il n'y a alors que deux cas possibles :

- ou bien S_n a une limite finie S dans $\mathbb{R}_+$, la série converge et a pour somme S ;
- ou bien S_n tend vers $+\infty$ et la série diverge.

La divergence d'une série à termes positifs équivaut donc à la convergence vers $+\infty$ de la suite de ses sommes partielles. Il est commode dans ce cas de considérer que la série « converge dans $\overline{\mathbb{R}}_+$ » et que sa somme vaut $+\infty$. Ainsi lorsque les u_k sont tous positifs, l'écriture $\sum_{k=0}^{+\infty} u_k$ a toujours un sens comme représentant un élément S de $\overline{\mathbb{R}}_+$: S étant un réel positif si la série converge au sens de la définition 2.1, $S = +\infty$ sinon.

Il importe de bien comprendre que ceci est particulier aux séries à termes positifs. Une série à termes de signe quelconque [2] peut très bien diverger sans que la suite des sommes partielles tende vers $+\infty$ ou $-\infty$. Un exemple évident est la série de terme général $u_k = (-1)^k$.

Théorème 2.12 (comparaison). *On suppose qu'il existe un* $k_0 \in \mathbb{N}$ *tel que pour tout* $k \geqslant k_0$, $0 \leqslant u_k \leqslant v_k$. *Alors*

a) la convergence [3] *de* $\sum_{k=0}^{+\infty} v_k$ *implique celle de* $\sum_{k=0}^{+\infty} u_k$;

1. Un espace métrique est dit complet si toute suite de Cauchy dans cet espace converge. Les espaces $\mathbb{R}^d$ sont complets pour la distance associée à une norme (peu importe laquelle). Un espace vectoriel normé complet est appelé espace de Banach.

2. Plus précisément une série dont la suite des termes présente une infinité de changements de signe.

3. La convergence et la divergence dans cet énoncé sont entendues au sens de la définition 2.1.

b) la divergence de $\sum_{k=0}^{+\infty} u_k$ implique celle de $\sum_{k=0}^{+\infty} v_k$.

Parmi les applications du théorème de comparaison figure la comparaison à une série géométrique qui a donné naissance aux règles dites de D'Alembert et de Cauchy (basées respectivement sur l'étude du comportement asymptotique de u_{n+1}/u_n et de $u_n^{1/n}$). Ces règles n'ont d'autre utilité que de résoudre des exercices *ad hoc* et on peut facilement s'en passer.

Théorème 2.13 (règle des équivalents)**.** *On suppose que u_k et v_k sont équivalents*[4] *quand n tend vers l'infini et qu'à partir d'un rang k_0, $v_k \geqslant 0$. Alors les séries $\sum_{k=0}^{+\infty} u_k$ et $\sum_{k=0}^{+\infty} v_k$ sont de même nature (toutes deux convergentes ou toutes deux divergentes).*

Attention, l'hypothèse d'équivalence de u_k et v_k n'implique pas à elle seule que les deux séries soient de même nature si le signe de v_k n'est pas constant à partir d'un certain rang. Comme contre exemple on peut proposer $u_k = (-1)^k k^{-1/2}$ et $v_k = u_k + 1/k$ (exercice).

Théorème 2.14 (comparaison série-intégrale)**.** *Soit f continue sur $[k_0, +\infty[$, décroissante et positive sur cet intervalle. La série $\sum_{k=k_0}^{+\infty} f(k)$ converge si et seulement si $\int_{k_0}^{n} f(t)\,\mathrm{d}t$ a une limite finie quand n tend vers $+\infty$.*

La démonstration repose sur l'encadrement

$$\forall n > k_0, \qquad \sum_{k=k_0}^{n-1} f(k+1) \leqslant \int_{k_0}^{n} f(t)\,\mathrm{d}t \leqslant \sum_{k=k_0}^{n-1} f(k)$$

illustré par la figure 2.1. Cet encadrement a son intérêt propre pour contrôler le reste d'une série à l'aide d'une intégrale généralisée (ou *vice versa*).

En appliquant le théorème 2.14 avec $f(t) = t^{-\alpha}$, $\alpha > 0$ et $k_0 = 1$, on obtient la caractérisation de la convergence des séries de Riemann.

Corollaire 2.15 (séries de Riemann)**.**

$$\textit{La série } \sum_{k=1}^{+\infty} \frac{1}{k^\alpha} \textit{ est } \begin{cases} \textit{convergente pour } \alpha > 1, \\ \textit{divergente pour } 0 < \alpha \leqslant 1. \end{cases}$$

La série est aussi divergente pour $\alpha \leqslant 0$, puisqu'alors son terme général ne tend pas vers 0.

Corollaire 2.16 (séries de Bertrand)**.**

$$\textit{La série } \sum_{k=2}^{+\infty} \frac{1}{k(\ln k)^\beta} \textit{ est } \begin{cases} \textit{convergente pour } \beta > 1, \\ \textit{divergente pour } 0 < \beta \leqslant 1. \end{cases}$$

Remarque 2.17. Pour $\alpha \neq 1$ (ou pour $\alpha = 1$ et $\beta \leqslant 0$), la nature de la série de Bertrand $\sum_{k=2}^{+\infty} k^{-\alpha}(\ln k)^{-\beta}$ s'obtient directement par comparaison avec une série de Riemann.

4. Ce qui signifie que l'on peut écrire $u_k = c_k v_k$ avec $\lim_{k\to+\infty} c_k = 1$.

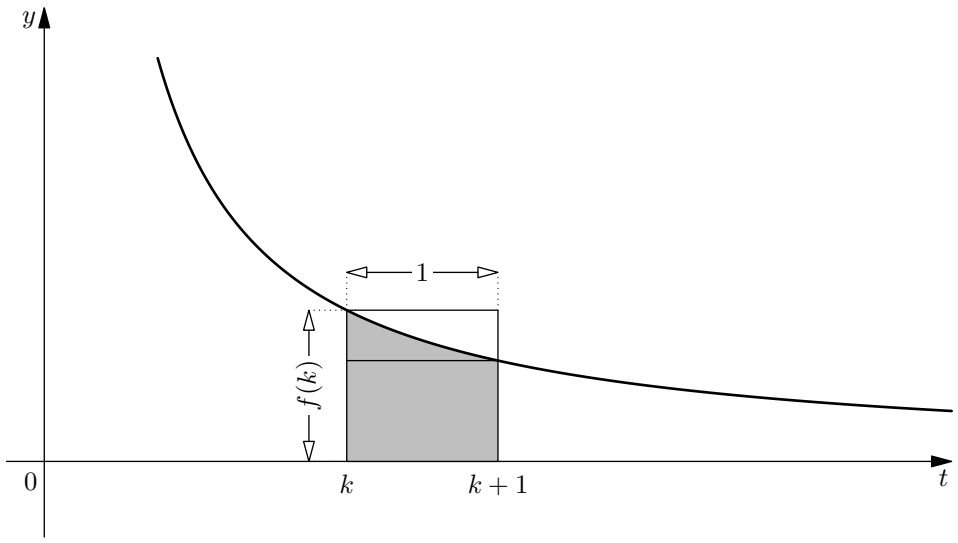


FIGURE 2.1 – Comparaison série-intégrale : $f(k+1) \leqslant \int_k^{k+1} f(t)\,\mathrm{d}t \leqslant f(k)$

2.1.3 Séries à termes de signe non constant

Après avoir vu que la convergence de $\sum_{k=0}^{+\infty} |u_k|$ implique celle de $\sum_{k=0}^{+\infty} u_k$, on peut se demander s'il existe des séries qui convergent sans converger absolument.

Remarquons d'abord que si le signe de u_k est constant à partir d'un certain rang k_0, l'étude de la convergence de $\sum_{k=0}^{+\infty} u_k$ se ramène à celle d'une série à termes positifs. En effet, $\sum_{k=0}^{+\infty} u_k$ et $\sum_{k=k_0}^{+\infty} u_k$ sont de même nature et si $u_k \leqslant 0$ pour $k \geqslant k_0$, il suffit de considérer $\sum_{k=k_0}^{+\infty} (-u_k)$. Ainsi pour une série à termes de signe constant à partir d'un certain rang, la convergence équivaut à la convergence absolue et la divergence ne peut se produire que si S_n tend vers $+\infty$ ou vers $-\infty$. Les seules séries susceptibles d'être convergentes sans l'être absolument sont donc celles dont le terme général *change de signe une infinité de fois.*

Par exemple, la série $\sum_{k=1}^{+\infty} (-1)^k/k$ converge mais pas absolument. C'est une application du théorème des séries alternées.

Théorème 2.18 (des séries alternées)**.** *Soit $(u_k)_{k\in\mathbb{N}}$ une suite alternée (pour tout k, u_k et u_{k+1} sont de signes contraires) telle que $|u_k|$ décroisse et tende vers 0. Alors*

a) la série $\sum_{k=0}^{+\infty} u_k$ converge,

b) pour tout $n \in \mathbb{N}$, les sommes partielles consécutives S_n et S_{n+1} encadrent[5] *la somme S et le reste $R_n = \sum_{k=n+1}^{+\infty} u_k$ vérifie*

$$|R_n| \leqslant |u_{n+1}|.$$

5. Attention, une fois sur deux on aura $S_{n+1} \leqslant S_n$.

Exemple 2.19. Pour $0 < \alpha \leqslant 1$, la série $\sum_{k=1}^{+\infty}(-1)^k k^{-\alpha}$ converge mais pas absolument. Pour $\alpha > 1$ elle est absolument convergente.

Exemple 2.20. La série $\sum_{k=2}^{+\infty}(-1)^k/\ln k$ converge mais pas absolument.

Remarque 2.21. Même si une série alternée est absolument convergente, le b) du théorème reste intéressant pour le calcul numérique de la somme S, cf. par exemple l'exercice 5.5.

2.1.4 Opérations sur les séries

La somme des séries de terme général u_k et v_k est la série de terme général (u_k+v_k). Le produit de la série de terme général u_k par la constante a est la série de terme général au_k. Le produit de deux séries sera étudié plus tard à l'aide des familles sommables.

Proposition 2.22. *Les affirmations suivantes sont vérifiées pour des séries à termes réels ou complexes ou dans un espace vectoriel normé.*

a) La somme de deux séries convergentes $\sum_{k=0}^{+\infty} u_k$ et $\sum_{k=0}^{+\infty} v_k$ est convergente et

$$\sum_{k=0}^{+\infty}(u_k + v_k) = \sum_{k=0}^{+\infty} u_k + \sum_{k=0}^{+\infty} v_k.$$

b) Le produit de la série convergente $\sum_{k=0}^{+\infty} u_k$ par la constante a est une série convergente et

$$\sum_{k=0}^{+\infty} au_k = a\sum_{k=0}^{+\infty} u_k.$$

c) La somme d'une série convergente et d'une série divergente est divergente.

d) La somme de deux séries divergentes peut être convergente ou divergente.

Dans une somme d'un *nombre fini* de termes, les propriétés de commutativité et d'associativité de l'addition permettent

– de changer l'ordre des termes sans changer la valeur de la somme,
– de faire des groupements de termes sans changer la valeur de la somme.

Que deviennent ces propriétés pour les séries considérées comme sommes d'une *infinité de termes* ? En gros la réponse est : si la série est absolument convergente, tout se passe bien, sinon on peut avoir des situations très pathologiques.

Le seul cas où l'on n'ait pas besoin de convergence absolue pour retrouver une situation conforme aux propriétés ci-dessus est celui de la sommation par paquets de taille finie fixe.

Proposition 2.23 (sommation par paquets de taille finie fixe)**.** *Soit $(u_n)_{n\in\mathbb{N}}$ une suite tendant vers 0. Fixons un entier $p \geqslant 2$ et définissons les « paquets de p termes consécutifs »*

$$v_j = \sum_{i=0}^{p-1} u_{jp+i}, \qquad j \in \mathbb{N}.$$

Alors les séries $\sum_{k=0}^{+\infty} u_k$ et $\sum_{j=0}^{+\infty} v_j$ sont de même nature et ont même somme lorsqu'elles convergent.

Remarque 2.24. Sans l'hypothèse de convergence vers 0 de la suite (u_n), le résultat devient grossièrement faux, contre exemple $u_n = (-1)^n$.

Remarque 2.25. Si $\sum_{k=0}^{+\infty} u_k$ est absolument convergente, la série de paquets $\sum_{j=0}^{+\infty} v_j$ l'est aussi. La réciproque est fausse, comme on peut le voir en groupant deux par deux les termes de la série $\sum_{k=0}^{+\infty}(-1)^k/(k+1)$, obtenant ainsi la série de terme général positif $v_j = (2j+1)^{-1}(2j+2)^{-1}$ équivalent à $(2j)^{-2}$.

Définition 2.26 (convergence commutative)**.** *La série $\sum_{k=0}^{+\infty} u_k$ est dite commutativement convergente et de somme S si pour toute bijection $f : \mathbb{N} \to \mathbb{N}$, la série $\sum_{k=0}^{+\infty} u_{f(k)}$ converge et a pour somme S. Dans ce cas nous pouvons utiliser la notation*

$$S = \sum_{k\in\mathbb{N}} u_k \quad \textit{au lieu de} \quad S = \sum_{k=0}^{+\infty} u_k,$$

puisque la somme de la série ne dépend pas de l'ordre dans lequel on effectue la sommation. L'écriture avec indexation par « $k \in \mathbb{N}$ » ne présuppose aucun ordre d'écriture des termes[6]*.*

Théorème 2.27. *La convergence absolue d'une série à termes réels ou complexes, implique sa convergence commutative.*

Nous allons démontrer le théorème en examinant successivement les cas des séries à termes positifs, réels et complexes.

Preuve du th. 2.27, cas d'une série à termes positifs. La convergence absolue se réduit alors à la convergence et signifie que la suite croissante des sommes partielles S_n a une limite finie S dans $\mathbb{R}_+$. Soit f une bijection $\mathbb{N} \to \mathbb{N}$. Rappelons que si A est une partie de $\mathbb{N}$, la notation $f(A)$ désigne l'ensemble des images des éléments de A par f : $f(A) = \{f(k)\ ;\ k \in A\}$. Définissons les deux suites d'entiers $(p_n)_{n\in\mathbb{N}}$ et $(q_n)_{n\in\mathbb{N}}$ comme suit. Pour p_n nous prenons le plus petit entier p tel que $f(\llbracket 0, p\rrbracket)$ recouvre $\llbracket 0, n\rrbracket$. L'existence de tels p et donc de p_n découle clairement de la surjectivité de f. Pour q_n on prend $\max f(\llbracket 0, p_n\rrbracket)$. Autrement dit p_n sert à recouvrir $\llbracket 0, n\rrbracket$ de la manière la plus économique par les premiers termes de la suite des $f(k)$ et q_n sert à « boucher les trous » de la suite d'entiers ainsi formée. On a ainsi

$$\forall n \in \mathbb{N}, \quad \llbracket 0, n\rrbracket \subset f(\llbracket 0, p_n\rrbracket) \subset \llbracket 0, q_n\rrbracket. \tag{2.1}$$

Clairement $n \leqslant p_n$ et $n \leqslant q_n$, donc p_n et q_n tendent vers l'infini avec n. De plus ces deux suites sont croissantes de par leur construction. Posons

$$S_n := \sum_{k=0}^{n} u_k, \quad T_n := \sum_{k=0}^{n} u_{f(k)}.$$

6. Par analogie avec l'écriture $\sum_{i\in I} u_i$, où I est un ensemble fini d'indices. Dans ce cas le résultat de l'addition des u_i pour $i \in I$ ne dépend pas de l'ordre des termes et d'ailleurs l'ensemble I n'a nul besoin ici d'être ordonné.

En raison de la positivité des u_k et de (2.1), nous disposons de l'encadrement

$$\forall n \in \mathbb{N}, \quad S_n \leqslant T_{p_n} \leqslant S_{q_n}. \tag{2.2}$$

Si S_n converge vers S, la sous-suite $(S_{q_n})_{n\in\mathbb{N}}$ converge aussi vers S et (2.2) implique alors la convergence de $(T_{p_n})_{n\in\mathbb{N}}$ vers S. En raison de la positivité des u_k, la suite $(T_n)_{n\in\mathbb{N}}$ est croissante. Nous venons de voir qu'elle a une sous-suite qui converge vers S, c'est donc *toute la suite* $(T_n)_{n\in\mathbb{N}}$ qui converge vers S. Autrement dit, $\sum_{k=0}^{+\infty} u_{f(k)}$ converge et a pour somme $S = \sum_{k=0}^{+\infty} u_k$. Comme la bijection f est quelconque, ceci achève la preuve du théorème 2.27 dans le cas d'une série à termes positifs. □

Remarque 2.28. Pour établir (2.2), nous n'avons pas utilisé l'hypothèse de convergence de la série de terme général u_k, mais seulement la positivité des u_k. Par conséquent (2.2) reste valable si la série à termes positifs $\sum_{k=0}^{+\infty} u_k$ diverge. Dans ce cas S_n tend vers $+\infty$, donc aussi T_{p_n} et la suite croissante $(T_n)_{n\in\mathbb{N}}$ ayant une sous-suite tendant vers $+\infty$, c'est toute la suite $(T_n)_{n\in\mathbb{N}}$ qui tend vers l'infini. Ainsi on a $\sum_{k=0}^{+\infty} u_{f(k)} = +\infty = \sum_{k=0}^{+\infty} u_k$.

Pour la commodité de référence et en raison du rôle essentiel joué par les séries à termes positifs dans cet ouvrage, nous rassemblons dans l'énoncé suivant les résultats obtenus dans la preuve du théorème 2.27 (cas $u_k \geqslant 0$) et dans la remarque 2.28 .

Proposition 2.29. *Si $(u_k)_{k\in\mathbb{N}}$ est une suite de réels positifs, on a pour toute bijection $f : \mathbb{N} \to \mathbb{N}$,*

$$\sum_{k=0}^{+\infty} u_k = \sum_{k=0}^{+\infty} u_{f(k)},$$

cette égalité ayant lieu dans $\overline{\mathbb{R}}_+$. Ainsi si les u_k sont positifs, la notation $\sum_{k\in\mathbb{N}} u_k$ est toujours légitime et représente la somme $S \in \overline{\mathbb{R}}_+$ de la série.

Dans la preuve du théorème 2.27, le cas d'une série à termes réels de signe quelconque va se ramener à celui d'une série à termes positifs en utilisant la décomposition d'un réel en partie positive et partie négative. Ceci requiert une digression préalable (de la définition 2.30 à la remarque 2.33).

Définition 2.30 (parties positive et négative)**.** *Pour tout x réel, sa partie positive notée x^+ et sa partie négative notée x^- sont* les réels positifs *définis par*

$$x^+ := \max(0, x), \quad x^- := \max(0, -x).$$

On a alors pour tout x réel

$$x = x^+ - x^-, \tag{2.3}$$
$$|x| = x^+ + x^-. \tag{2.4}$$

Par exemple $(3{,}12)^+ = 3{,}12$ et $(3{,}12)^- = 0$, $(-2{,}3)^+ = 0$ et $(-2{,}3)^- = 2{,}3$. Il importe de bien noter que la partie négative d'un réel est un nombre positif ou nul. Voyons d'abord comment intervient cette décomposition pour une série absolument convergente.

Lemme 2.31. *Si $\sum_{k=0}^{+\infty} u_k$ est une série à termes réels absolument convergente et de somme S, les séries à termes positifs $\sum_{k=0}^{+\infty} u_k^+$ et $\sum_{k=0}^{+\infty} u_k^-$ sont convergentes et on a*

$$\sum_{k=0}^{+\infty} u_k = \sum_{k=0}^{+\infty} (u_k^+ - u_k^-) = \sum_{k=0}^{+\infty} u_k^+ - \sum_{k=0}^{+\infty} u_k^-. \tag{2.5}$$

Preuve. Par (2.4), on a $0 \leqslant u_k^+ \leqslant |u_k|$ et $0 \leqslant u_k^- \leqslant |u_k|$. Les séries de terme général u_k^+ et u_k^- sont donc convergentes par le théorème de comparaison. La première égalité dans (2.5) résulte alors de (2.3) appliquée à $x = u_k$. La deuxième égalité s'obtient par la proposition 2.22 a) et b). □

Remarque 2.32. La réciproque est vraie : si les séries à termes positifs $\sum_{k=0}^{+\infty} u_k^+$ et $\sum_{k=0}^{+\infty} u_k^-$ sont convergentes, la série $\sum_{k=0}^{+\infty} u_k$ est absolument convergente et on a (2.5) grâce à la proposition 2.22. Ceci montre que l'on ne peut pas se passer de l'hypothèse de convergence absolue dans le lemme 2.31. Il est d'autre part facile d'exhiber un contre exemple : la série de terme général $u_k = (-1)^k/k$.

Remarque 2.33. L'égalité « $\sum_{k=0}^{+\infty} u_k^+ = \left(\sum_{k=0}^{+\infty} u_k\right)^+$ » est grossièrement fausse ! Elle n'est même pas vraie pour des sommes finies. Comparez $(a+b)^+$ et $a^+ + b^+$ pour $a = -1$ et $b = 2$ pour vous en convaincre.

Preuve du th. 2.27 dans le cas d'une série à termes réels. Soit $\sum_{k=0}^{+\infty} u_k$ une série absolument convergente à termes réels et f une bijection quelconque $\mathbb{N} \to \mathbb{N}$. On utilise la décomposition $u_{f(k)} = u_{f(k)}^+ - u_{f(k)}^-$. Par le lemme 2.31, la convergence absolue de la série de terme général u_k implique la convergence des séries à termes généraux positifs u_k^+ et u_k^-. La preuve du théorème 2.27 dans le cas des termes positifs nous fournit alors la convergence des séries $\sum_{k=0}^{+\infty} u_{f(k)}^+$ et $\sum_{k=0}^{+\infty} u_{f(k)}^-$ et l'égalité de leurs sommes avec $\sum_{k=0}^{+\infty} u_k^+$ et $\sum_{k=0}^{+\infty} u_k^-$ respectivement. Par la remarque 2.32, on en déduit la convergence absolue de $\sum_{k=0}^{+\infty} u_{f(k)}$. Compte-tenu du lemme 2.31, on a de plus

$$\sum_{k=0}^{+\infty} u_{f(k)} = \sum_{k=0}^{+\infty} u_{f(k)}^+ - \sum_{k=0}^{+\infty} u_{f(k)}^- = \sum_{k=0}^{+\infty} u_k^+ - \sum_{k=0}^{+\infty} u_k^- = \sum_{k=0}^{+\infty} (u_k^+ - u_k^-) = \sum_{k=0}^{+\infty} u_k.$$

□

Preuve du th. 2.27 dans le cas d'une série à termes complexes. Pour $u_k \in \mathbb{C}$, notons $x_k := \operatorname{Re} u_k$ et $y_k := \operatorname{Im} u_k$. Grâce aux inégalités $0 \leqslant |x_k| \leqslant |u_k|$ et $0 \leqslant |y_k| \leqslant |u_k|$, les séries de terme général réel x_k et y_k héritent de la convergence absolue de la série de terme général complexe u_k. On conclut alors en combinant la preuve du th. 2.27 dans le cas réel et la remarque 2.3. □

Remarque 2.34. L'argument utilisé ci-dessus dans le cas complexe se généralise aux séries à valeurs dans un espace vectoriel normé E de *dimension finie*, disons pour simplifier $E = \mathbb{R}^d$. En effet sur cet espace toutes les normes sont équivalentes et on

peut donc choisir la norme donnée par $\|x\| = \max_{1\leqslant i\leqslant d} |x_i|$ pour $x = (x_1, \dots, x_d)$. La convergence d'une suite dans $(E, \| \ \|)$ équivaut à la convergence composante par composante et la convergence de $\sum_{k=0}^{+\infty} \|u_k\|$ implique la convergence absolue des d séries de terme général $u_{k,i}$, en posant $u_k = (u_{k,1}, \dots, u_{k,d})$. L'adaptation de la preuve ci-dessus est alors immédiate. En revanche si E est de dimension infinie, ce raisonnement n'est plus valable.

Théorème 2.35. *La convergence commutative d'une série à termes réels ou complexes, implique sa convergence absolue.*

Preuve. Il est clair qu'il suffit de traiter le cas des séries à termes réels. Nous allons démontrer la *contraposée* de l'énoncé, c'est-à-dire que la négation de la conclusion implique la négation de l'hypothèse. Dans ce but nous supposons que la série de terme général u_k converge vers le réel S mais pas absolument, autrement dit que :

$$\sum_{k=0}^{+\infty} u_k = S \in \mathbb{R} \quad \text{et} \quad \sum_{k=0}^{+\infty} |u_k| = +\infty. \tag{2.6}$$

Nous allons construire une bijection $f : \mathbb{N} \to \mathbb{N}$ telle que $\sum_{k=0}^{+\infty} u_{f(k)}$ ne converge pas vers S.

On commence par noter que (2.6) implique que

$$\sum_{k=0}^{+\infty} u_k^+ = +\infty \quad \text{et} \quad \sum_{k=0}^{+\infty} u_k^- = +\infty. \tag{2.7}$$

En effet si ces deux sommes sont finies, la série $\sum_{k=0}^{+\infty} u_k$ est absolument convergente par la remarque 2.32. Si l'une est finie et l'autre infinie, la série $\sum_{k=0}^{+\infty} u_k$ diverge par la proposition 2.22 b) et c).

Nous allons montrer que l'on peut construire une bijection f pour laquelle les sommes partielles $\sum_{k=0}^{n} u_{f(k)}$ oscillent indéfiniment entre -1 et 1 et donc n'ont pas de limite. L'écriture explicite de la définition de f étant assez lourde, nous nous contenterons de donner l'idée de sa construction. Le point clé est le lemme suivant.

Lemme 2.36. *Soit $v = (v_k)_{k\in\mathbb{N}}$ une suite de réels positifs telle que $\sum_{k=0}^{+\infty} v_k = +\infty$. Pour tout intervalle borné I de $\mathbb{N}$, définissons*

$$T(I, v) := \sum_{k\in I} v_k,$$

avec la convention $T(\varnothing, v) := 0$. On a alors

$$\forall i \in \mathbb{N},\ \forall x \in \mathbb{R}_+,\ \exists! j = j(i, x, v) \in \mathbb{N}; \quad T([\![i, j[\![, v) \leqslant x < T([\![i, j]\!], v). \tag{2.8}$$

Preuve du lemme. Notons $t_{i,j} := T([\![i, j[\![, v)$. On remarque que pour tout $i \in \mathbb{N}$ fixé, la suite $(t_{i,j})_{j\geqslant i}$ est croissante à cause de la positivité des v_k et tend vers $+\infty$ quand j tend vers $+\infty$. Ce dernier point résulte de l'hypothèse $\sum_{k=0}^{+\infty} v_k = +\infty$ et du fait que

l'on ne change pas la divergence de cette série en supprimant ses i premiers termes. De plus on a $t_{i,i} = 0$ car $[\![i, i[\![= \varnothing$. Il est alors clair que l'on obtient (2.8) avec $j := \max\{l \geqslant i;\ t_{i,l} \leqslant x\}$. □

Partageons maintenant $\mathbb{N}$ en les deux sous-ensembles complémentaires :

$$A^+ := \{l \in \mathbb{N};\ u_l \geqslant 0\}, \quad A^- := \{l \in \mathbb{N};\ u_l < 0\}.$$

Il résulte immédiatement de (2.7) que ces deux ensembles sont infinis (donc dénombrables comme parties infinies de $\mathbb{N}$). Définissons alors la suites $v = (v_k)_{k\in\mathbb{N}}$ en réindexant la « suite » $(u_l)_{l\in A^+}$ par $\mathbb{N}$ (en respectant l'ordre des indices l). Définissons de même $w = (w_k)_{k\in\mathbb{N}}$ à partir de $(u_l)_{l\in A^-}$.

En appliquant le lemme 2.36 alternativement avec les suites v et $-w$, on peut construire f par récurrence de façon à ce que les sommes partielles $T_n = \sum_{k=0}^{n} u_{f(k)}$ se trouvent à gauche de -1 pour une infinité de n et à droite de $+1$ pour une infinité de n. Voici le début de la construction. On applique d'abord le lemme avec $i = 0$, $x = 1$ et v. On obtient alors $T_{n_1} \geqslant 1$ pour $n_1 = j(0, 1, v)$ donné par(2.8). Ensuite on revient en arrière en appliquant le lemme avec $-w$, $i = 0$ et $x = 1 + T_{n_1}$. Pour $n_2 = n_1 + j(0, x, -w)$, on a alors $T_{n_2} < -1$. Pour la troisième étape, on utilise le lemme avec v, $i = j(0, 1, v) + 1$, $x = -T_{n_2} + 1$ et pour $n_3 = n_2 + j(i, x, v)$ on obtient $T_{n_3} \geqslant 1$. Et ainsi de suite ... □

On peut adapter la démonstration ci-dessus pour construire d'autres bijections f de façon à faire osciller les sommes partielles entre a et b fixés, ou pour les faire converger vers n'importe quel réel fixé [7], pour les faire tendre vers $+\infty$, vers $-\infty$, etc.

Pour conclure cette section, retenons que pour les séries à termes réels ou complexes, *convergence commutative et convergence absolue sont équivalentes.*

2.2 Familles sommables

Dans cette section, nous allons généraliser la notion de série en essayant de donner un sens à une expression de la forme $\sum_{i\in I} u_i$, où I est un ensemble infini [8]. Dans le cadre de cet ouvrage, nous n'aurons besoin que du cas où les u_i sont réels ou complexes. Néanmoins, pour faciliter l'utilisation dans d'autres branches des mathématiques (par exemple analyse complexe, séries de Fourier, analyse fonctionnelle...), nous nous situerons d'emblée dans le cas où les u_i sont des éléments d'un espace vectoriel normé E. *Le lecteur que cela gênerait pourra facilement adapter les énoncés au cas des u_i réels ou complexes en remplaçant les normes par des valeurs absolues ou des modules.* Cette simplification des énoncés n'apporte d'ailleurs pas de simplification notable des démonstrations.

Partons de la définition de la convergence d'une série de vecteurs de terme général u_k dans un espace vectoriel normé E. Ici l'ensemble d'indexation est $\mathbb{N}$. On dit que

7. Il faudra utiliser alors la convergence vers 0 de u_n, ce dont nous n'avons pas eu besoin ci-dessus.

8. L'exposé présenté ici est essentiellement une adaptation du chapitre XIV *Séries dans les espaces vectoriels normés* de l'ouvrage de L. Schwartz [13].

cette série converge et a pour somme le vecteur S de E si la suite des sommes partielles $S_n := u_0 + u_1 + \cdots + u_n$ converge vers S, autrement dit si $\|S_n - S\|$ tend vers 0 quand n tend vers $+\infty$. Ceci s'écrit encore :

$$\forall \varepsilon > 0,\ \exists n_0 \in \mathbb{N},\ \forall n \geqslant n_0, \quad \|S_n - S\| < \varepsilon. \tag{2.9}$$

Si on veut généraliser ceci à une famille de vecteurs $(u_i)_{i\in I}$ où I est un ensemble infini quelconque, on se heurte immédiatement à une difficulté, c'est qu'une écriture comme « $\forall i \geqslant i_0$ », n'a en général pas de sens si I n'est pas muni d'une relation d'ordre. Essayons alors de traduire l'idée exprimée par (2.9) sans faire appel à la structure d'ordre de $\mathbb{N}$. On remarque pour cela que S_n réalise une approximation de S par la somme d'un *nombre fini* de termes de la famille $(u_k)_{k\in\mathbb{N}}$ avec une erreur $\|S_n - S\|$ inférieure à ε. Cette approximation peut être réalisée par une somme finie indexée par n'importe quel intervalle $K_n := [\![0, n]\!]$ pourvu que $n \geqslant n_0$. Pour se débarrasser de la relation d'ordre intervenant dans cette dernière écriture on la reformule en $K_n \supset K_{n_0}$. Réécrivons maintenant (2.9) à l'aide des ensembles emboîtés K_n.

$$\forall \varepsilon > 0,\ \exists K_{n_0} = [\![0, n_0]\!],\ \forall K_n \supset K_{n_0}, \quad \Big\| \sum_{k\in K_n} u_k - S \Big\| < \varepsilon. \tag{2.10}$$

Au risque d'insister lourdement, notons que dans l'écriture, « $\forall K_n \supset K_{n_0}$ », K_n désigne non pas n'importe quelle partie finie de $\mathbb{N}$, mais un ensemble fini de la forme $\{0, 1, 2, \ldots, n\}$. Avons-nous réussi ainsi à expurger (2.9) de la relation d'ordre sur $\mathbb{N}$? En fait non, nous avons seulement réussi à la cacher dans la définition des ensembles K_n d'entiers *consécutifs* entre 0 et n. Si on veut vraiment généraliser à un ensemble d'indexation I quelconque, on est donc condamné à renoncer à cette structure particulière des K_n et à ne retenir que leur finitude. On est ainsi amené à introduire une *nouvelle* notion de convergence pour la série de terme général u_k :

$$\forall \varepsilon > 0,\ \exists J \text{ fini } \subset \mathbb{N},\ \forall K \text{ fini tel que } J \subset K \subset \mathbb{N}, \quad \Big\| \sum_{k\in K} u_k - S \Big\| < \varepsilon. \tag{2.11}$$

Pourquoi avons-nous pris la précaution de qualifier cette convergence de *nouvelle* ? Parce que s'il est clair que (2.11) implique (2.10) et donc (2.9), rien ne nous permet d'affirmer que la réciproque soit vraie. Nous verrons d'ailleurs ci-dessous que cette réciproque est fausse pour les séries qui ne sont pas commutativement convergentes.

Après ce *briefing*, aventurons-nous dans l'exploration de la sommation de familles $(u_i)_{i\in I}$ indexées par un ensemble quelconque, les u_i étant éléments d'un espace vectoriel normé E. Pour K sous-ensemble *fini* de I, posons :

$$S_K := \sum_{i\in K} u_i.$$

Cette définition est cohérente puisque dans un espace vectoriel E la somme d'un nombre fini d'éléments de E est encore un élément de E et ne dépend pas de l'ordre d'écriture des termes. Cette notation n'est pas contradictoire avec celle utilisée pour

les sommes partielles S_n d'une série de terme général u_i, $i \in \mathbb{N}$, en considérant que S_n est une abréviation de $S_{[\![0,n]\!]}$ et qu'il ne faut pas confondre S_n avec $S_{\{n\}} = u_n$. Dans le cas particulier où $K = \varnothing$, on fait la convention $S_\varnothing := 0$ (vecteur nul de E). Dans tout ce qui suit, les lettres majuscules J, K désigneront toujours des parties finies de I et nous omettrons parfois d'écrire $K \subset I$ si le contexte lève toute ambiguïté. On pourrait d'ailleurs s'affranchir complètement de cette contrainte en définissant $S_K := \sum_{i \in K \cap I} u_i$ pour tout ensemble K fini (pas forcément inclus dans I).

Définition 2.37 (famille sommable)**.** *Soit E un espace vectoriel normé, I un ensemble d'indices et $(u_i)_{i\in I}$ une famille d'éléments de E. On dit que cette famille est sommable de somme $S \in E$ si*

$$\forall \varepsilon > 0,\ \exists J \text{ fini } \subset I,\ \forall K \text{ fini } \supset J, \quad \|S_K - S\| < \varepsilon. \tag{2.12}$$

On écrira alors $S = \sum_{i\in I} u_i$.

Insistons sur le fait que cette définition ne suppose *aucune structure d'ordre sur l'ensemble d'indices I* et donc aucun ordre privilégié d'écriture des termes. Si I est fini, toute famille $(u_i)_{i\in I}$ est trivialement sommable, il suffit de prendre $S = S_I$ et $J = I$ dans (2.12) pour avoir $\|S_K - S\| = 0$. La définition 2.37 n'a donc d'intérêt que pour I infini. Nous verrons bientôt qu'en fait, on peut se ramener à un I au plus dénombrable (prop. 2.44).

Remarque 2.38. La somme d'une famille sommable est unique. Vérifions en effet que si S et S' satisfont tous deux à (2.12) avec la même famille $\{u_i,\ i \in I\}$, alors $S = S'$. Bien sûr, l'ensemble fini J « associé » à S' et à ε par (2.12) n'a aucune raison d'être le même que pour S, nous le noterons donc plutôt J'. Pour tout $\varepsilon > 0$, l'ensemble fini $K = J \cup J'$ contient à la fois J et J', donc $\|S - S_K\| \leqslant \varepsilon$ et $\|S' - S_K\| \leqslant \varepsilon$. Par inégalité triangulaire, $\|S - S'\| \leqslant 2\varepsilon$. Dans cette dernière inégalité, le premier membre ne dépend plus de ε. L'inégalité étant vraie pour tout $\varepsilon > 0$, $\|S - S'\| = 0$ donc $S = S'$.

La sommabilité est préservée par combinaison linéaire, voici l'énoncé précis.

Proposition 2.39 (sommabilité et combinaison linéaire)**.** *Soient $(u_i)_{i\in I}$ et $(u'_i)_{i\in I}$ deux familles sommables dans le même espace vectoriel normé E, ayant même ensemble d'indexation I. Notons S et S' les sommes respectives. Alors pour tous scalaires a et b, la famille $(au_i + bu'_i)_{i\in I}$ est sommable dans E, de somme $aS + bS'$.*

Preuve. Il est clair qu'il suffit de traiter les deux cas particuliers de $(u_i + u'_i)_{i\in I}$ et de $(au_i)_{i\in I}$. Nous détaillons seulement le premier, laissant l'autre au lecteur. Par hypothèse de sommabilité des familles $(u_i)_{i\in I}$ et $(u'_i)_{i\in I}$, on a pour tout $\varepsilon > 0$, deux parties finies J et J' de I telles que pour tous K, K' finis tels que $J \subset K \subset I$ et $J' \subset K' \subset I$, $\|S - S_K\| < \varepsilon/2$ et $\|S' - S'_{K'}\| < \varepsilon/2$. Ces deux inégalités sont vraies en particulier pour $K = K' = L$ où L est n'importe quelle partie finie de I contenant $J \cup J'$. Par inégalité triangulaire, on a alors $\|(S + S') - (S_L + S'_L)\| < \varepsilon$. Par les

propriétés de l'addition dans E et la finitude de L, on a

$$S_L + S'_L = \sum_{i\in L} u_i + \sum_{i\in L} u'_i = \sum_{i\in L} (u_i + u'_i).$$

On a donc bien vérifié que $(u_i + u'_i)_{i\in I}$ est sommable et a pour somme $S + S'$. □

La sommabilité est préservée par permutation sur les indices.

Proposition 2.40 (invariance par permutation)**.** *Soit $(u_i)_{i\in I}$ une famille sommable de somme S. Alors pour toute bijection $\varphi : I \to I$, la famille $(u_{\varphi(i)})_{i\in I}$ est sommable de somme S.*

Preuve. En notant $v_i := u_{\varphi(i)}$, il s'agit de montrer que $(v_i)_{i\in I}$ est sommable de même somme S que $(u_i)_{i\in I}$. L'hypothèse de sommabilité de $(u_i)_{i\in I}$ s'écrit

$$\forall \varepsilon > 0,\ \exists J \text{ fini } \subset I,\ \forall K \text{ fini } \supset J, \quad \|S_K - S\| < \varepsilon \tag{2.13}$$

Posons $J' := \varphi^{-1}(J) = \{\varphi^{-1}(i);\ i \in J\}$. L'ensemble J' est fini car en bijection avec l'ensemble fini J par φ^{-1}. Pour tout K' fini contenant J', l'ensemble fini $\varphi(K')$ contient $\varphi(J')$ et ce dernier est égal à J. Donc $\|S_{\varphi(K')} - S\| < \varepsilon$ en appliquant (2.13) avec $\varphi(K')$ au lieu de K, et ceci est vrai pour tout K' fini contenant J'. D'autre part

$$S_{\varphi(K')} = \sum_{\ell\in\varphi(K')} u_\ell = \sum_{i\in K'} u_{\varphi(i)} = \sum_{i\in K'} v_i.$$

Nous avons donc montré que

$$\forall \varepsilon > 0,\ \exists J' \text{ fini } \subset I,\ \forall K' \text{ fini } \supset J', \quad \Big\| \sum_{i\in K'} v_i - S \Big\| < \varepsilon,$$

autrement dit que $(v_i)_{i\in I}$ est sommable de somme S. □

Remarque 2.41 (sommabilité d'une série)**.** Plaçons-nous un instant dans le cas particulier $I = \mathbb{N}$. Nous savions déjà, cf. l'implication (2.11) ⇒ (2.9) vue en introduction, que la sommabilité de $(u_k)_{k\in\mathbb{N}}$ implique la convergence de la série $\sum_{k=0}^{+\infty} u_k$. La proposition 2.40 nous apprend en plus que cette convergence est nécessairement commutative. Dans le cas où les u_k sont réels ou complexes, on peut donc dire que la sommabilité de $(u_k)_{k\in\mathbb{N}}$ implique la convergence commutative et absolue de la série $\sum_{k=0}^{+\infty} u_k$, cf. théorème 2.35. Ainsi comme nous l'avions annoncé en introduction, l'implication « (2.9) ⇒ (2.11) » est fausse. Par exemple la série de terme général $u_k = (-1)^k/(k+1)$ est convergente, mais la famille $(u_k)_{k\in\mathbb{N}}$ n'est pas sommable.

Il est facile de donner une caractérisation de la sommabilité pour les familles de réels positifs.

Proposition 2.42. *Soit $(u_i)_{i\in I}$ une famille d'éléments de $\mathbb{R}_+$.*
a) Si $(u_i)_{i\in I}$ est telle que

$$M := \sup_{K \text{ fini} \subset I} S_K < +\infty, \tag{2.14}$$

alors elle est sommable de somme M.
b) Réciproquement, si $(u_i)_{i\in I}$ est sommable de somme S, on a

$$S = \sup_{K \text{ fini} \subset I} S_K, \tag{2.15}$$

donc ce sup *est fini.*

Preuve du a). Soit $\varepsilon > 0$. Comme M est *fini*, $M-\varepsilon$ est strictement inférieur à M. Par minimalité du supremum M parmi tous les majorants de l'ensemble $\{S_K;\ K \text{ fini } \subset I\}$, il existe une partie finie J de I telle que

$$M - \varepsilon < S_J \leqslant M.$$

De plus pour toute partie finie K de I, contenant J,

$$M - \varepsilon < S_J \leqslant S_K \leqslant M. \tag{2.16}$$

En effet, $S_K - S_J = S_{K\setminus J}$ est une somme de réels positifs, donc positive, ce qui justifie la deuxième inégalité dans (2.16). La troisième découle de la définition de M. Comme (2.16) implique $|M - S_K| < \varepsilon$, nous avons ainsi établi que $\{u_i,\ i \in I\}$ est sommable et de somme M. □

Preuve du b). Supposons maintenant que $(u_i)_{i\in I}$ est sommable dans $\mathbb{R}_+$, de somme S. Notons M le supremum défini par (2.15), qu'il soit fini ou infini. On commence par remarquer que *pour toute partie finie J* de I, on a l'égalité suivante dans $\overline{\mathbb{R}}_+$:

$$M := \sup_{L \text{ fini} \subset I} S_L = \sup_{K \text{ fini},\ J\subset K\subset I} S_K. \tag{2.17}$$

En effet la positivité des u_i et l'inclusion de L fini dans l'ensemble fini $L\cup J$ contenant J, nous donnent pour tout L fini l'inégalité $S_L \leqslant S_{L\cup J}$. On en déduit l'inégalité des suprema

$$M \leqslant \sup_{K \text{ fini},\ J\subset K\subset I} S_K.$$

L'inégalité inverse étant évidente, (2.17) est vérifiée.

Soit $\varepsilon > 0$ quelconque. Il existe une partie finie $J = J(\varepsilon)$ de I telle que pour tout K fini vérifiant $J \subset K \subset I$, $S_K \in]S-\varepsilon, S+\varepsilon[$. Ceci implique

$$\sup_{K \text{ fini},\ J\subset K\subset I} S_K \in [S-\varepsilon, S+\varepsilon]$$

et compte-tenu de (2.17) appliqué avec $J = J(\varepsilon)$, $M \in [S-\varepsilon, S+\varepsilon]$. Ceci nous montre déjà que M est fini. On a de plus $|S - M| \leqslant \varepsilon$ pour tout $\varepsilon > 0$, d'où $M = S$. □

Remarque 2.43. Il résulte de la proposition 2.42 qu'une famille de réels positifs u_i est non sommable si et seulement si le supremum M des sommes S_K pour K partie finie de I vaut $+\infty$. La situation est analogue à celle des séries à termes positifs qui ne peuvent diverger que si la suite des sommes partielles tend vers $+\infty$. On peut donc toujours donner un sens à l'expression $\sum_{i\in I} u_i$, considérée comme élément de $\overline{\mathbb{R}}_+$, en posant

$$\sum_{i\in I} u_i := \sup_{K \text{ finie }, K\subset I} S_K = \begin{cases} S \in \mathbb{R}_+, & \text{si } (u_i)_{i\in I} \text{ est sommable de somme } S, \\ +\infty, & \text{sinon.} \end{cases} \tag{2.18}$$

Cette situation est particulière aux familles à termes positifs et c'est *le seul cas* où nous donnerons un sens à $\sum_{i\in I} u_i$ sans que la famille soit forcément sommable.

Proposition 2.44 (sommabilité et dénombrabilité)**.** *Si $(u_i)_{i\in I}$ est sommable dans E, l'ensemble d'indices $I' := \{i \in I;\ u_i \neq 0\}$ est au plus dénombrable.*

Preuve. Pour chaque $n \in \mathbb{N}^*$, appliquons (2.12) avec $\varepsilon = 1/n$ et choisissons l'un [9] des ensembles finis J donnés par (2.12), que nous noterons J_n. Posons

$$H := \bigcup_{n\in\mathbb{N}^*} J_n.$$

L'ensemble H est au plus dénombrable comme réunion dénombrable d'ensembles au plus dénombrables (prop. 1.40). Si $H = I$, alors I' qui est inclus dans H est donc au plus dénombrable. En dehors de ce cas trivial, $I\backslash H$ n'est pas vide. Soit i_0 un élément quelconque de $I\backslash H$. En appliquant (2.12) avec $\varepsilon = 1/n$, $J = J_n$ et chacun des deux ensembles finis $K = J_n$, $K' = J_n \cup \{i_0\}$, on obtient les deux inégalités :

$$\|S_{J_n} - S\| \leqslant \frac{1}{n} \quad \text{et} \quad \|u_{i_0} + S_{J_n} - S\| \leqslant \frac{1}{n},$$

vraies pour tout $n \in \mathbb{N}^*$. En écrivant $u_{i_0} = (u_{i_0} + S_{J_n} - S) + (S - S_{J_n})$, l'inégalité triangulaire nous donne :

$$\forall n \geqslant 1, \quad \|u_{i_0}\| \leqslant \frac{2}{n}.$$

Faisant tendre n vers l'infini (noter que i_0 est fixé et ne dépend pas de n), on en déduit $\|u_{i_0}\| = 0$, d'où $u_{i_0} = 0$. Ce raisonnement étant valable pour n'importe quel $i_0 \in I\backslash H$, on en déduit que I' est inclus dans H. Par conséquent, I' est au plus dénombrable. □

Nous allons voir maintenant que pour I dénombrable, sommabilité de $(u_i)_{i\in I}$ et convergence commutative de la série associée *via* une numérotation de I sont équivalentes.

9. La propriété (2.12) nous assure qu'il existe au moins un J pour ε donné, mais ne dit rien sur l'unicité.

Théorème 2.45. *Si I est dénombrable, les propriétés suivantes sont équivalentes :*

a) $(u_i)_{i\in I}$ est sommable de somme S ;

b) pour toute bijection $f : \mathbb{N} \to I$, la série $\sum_{k=0}^{+\infty} u_{f(k)}$ converge et a pour somme S.

Avant de démontrer ce théorème, remarquons que la propriété b) équivaut à la convergence commutative de la série *indexée par* $\mathbb{N}$ et de terme général $v_k := u_{f(k)}$ et ceci pour tout choix d'une bijection $f : \mathbb{N} \to I$. En effet, fixons une telle bijection et donc la suite $(v_k)_{k\in\mathbb{N}}$ correspondante. Soit σ une bijection quelconque de $\mathbb{N}$ sur $\mathbb{N}$. On a pour tout $k \in \mathbb{N}$, $v_{\sigma(k)} = u_{f(\sigma(k))} = u_{g(k)}$, où $g := f \circ \sigma$ est une bijection de $\mathbb{N}$ sur I. Par b), la série $\sum_{k=0}^{+\infty} u_{g(k)}$ converge et a pour somme S. Or cette série est exactement $\sum_{k=0}^{+\infty} v_{\sigma(k)}$. Par arbitrarité de la bijection σ on en déduit que $\sum_{k=0}^{+\infty} v_k$ est commutativement convergente.

Réciproquement, supposons qu'il existe au moins une bijection $h : \mathbb{N} \to I$ telle que la série $\sum_{k=0}^{+\infty} u_{h(k)}$ converge commutativement et ait pour somme S. Alors $(u_i)_{i\in I}$ vérifie la propriété b). En effet toute bijection $f : \mathbb{N} \to I$ peut s'écrire $f = (h\circ h^{-1})\circ f = h \circ (h^{-1} \circ f) = h \circ \sigma$, où $\sigma := h^{-1} \circ f$ est une bijection de $\mathbb{N}$ sur $\mathbb{N}$. Alors pour tout $k \in \mathbb{N}$, $u_{f(k)} = u_{h(\sigma(k))} = v_{\sigma(k)}$, où l'on a noté $v_k := u_{h(k)}$ le terme général de la série $\sum_{k=0}^{+\infty} u_{h(k)}$. La série $\sum_{k=0}^{+\infty} v_k$ étant commutativement convergente de somme S, on en déduit que la série de terme général $u_{f(k)} = v_{\sigma(k)}$ converge et a pour somme S.

Preuve de $a) \Rightarrow b)$. Avec les J_n choisis comme dans la preuve de la proposition 2.44, $\|S_K - S\| \leqslant 1/n$ pour toute partie finie K de I contenant J_n. Puisque f est une bijection de $\mathbb{N}$ *sur* I et J_n une partie finie de I, il existe pour chaque $n \in \mathbb{N}^*$ un entier $k_0(n)$ tel que

$$f\big([\![0, k_0(n)]\!]\big) \supset J_n.$$

Il suffit de prendre pour cela, $k_0(n) := \max\{f^{-1}(i);\, i \in J_n\}$. Autrement dit, on obtient l'intervalle $[\![0, k_0(n)]\!]$ en « bouchant les trous » de l'ensemble d'entiers $f^{-1}(J_n)$. Alors pour tout entier $l \geqslant k_0(n)$, $K := f\big([\![0, l]\!]\big)$ est une partie finie de I, contenant J_n et donc $\|S - S_K\| \leqslant 1/n$. Comme $S_K = \sum_{k=0}^{l} u_{f(k)}$, nous venons ainsi d'établir que :

$$\forall n \in \mathbb{N}^*,\ \exists k_0(n), \quad \forall l \geqslant k_0(n), \quad \Big\| \sum_{k=0}^{l} u_{f(k)} - S \Big\| \leqslant \frac{1}{n},$$

ce qui équivaut à la convergence vers S de la série de terme général $u_{f(k)}$ (c'est juste ce que l'on obtient en discrétisant le « $\forall \varepsilon > 0$ » en le remplaçant par $\varepsilon = 1/n$ dans la définition de cette convergence). □

Preuve de $b) \Rightarrow a)$. Nous allons prouver cette implication en démontrant sa contraposée : non $a) \Rightarrow$ non $b)$. La négation de $a)$ s'écrit

$$\exists \varepsilon > 0,\ \forall J \text{ partie finie de } I,\ \exists K \text{ fini}, \quad J \subset K \subset I \text{ et } \|S - S_K\| \geqslant \varepsilon. \tag{2.19}$$

Fixons une première bijection $\varphi : \mathbb{N} \to I$. Posons $k_0 := 0$ et $J_0 := \{\varphi(0)\}$. Par (2.19) il existe une partie finie K_0 de I, contenant J_0 et telle que $\|S - S_{K_0}\| \geqslant \varepsilon$. On peut maintenant choisir un entier $k_1 > k_0$ tel que $J_1 := \varphi\big([\![0, k_1]\!]\big)$ contienne *strictement* K_0, il suffit pour cela de prendre $k_1 = 1 + \max\{\varphi^{-1}(i);\ i \in K_0\}$. Une nouvelle invocation de (2.19) nous fournit une partie finie K_1 de I, contenant J_1 et telle que $\|S - S_{K_1}\| \geqslant \varepsilon$. Nous venons ainsi d'amorcer une récurrence qui basée sur l'utilisation itérée de (2.19), nous permet de construire une suite strictement croissante d'entiers (k_n), deux suites (J_n) et (K_n) de parties finies de I, vérifiant :

$$\forall n \in \mathbb{N}, \quad \varphi\big([\![0, k_n]\!]\big) = J_n \subset K_n \subsetneqq J_{n+1} \text{ et } \|S - S_{K_n}\| \geqslant \varepsilon. \tag{2.20}$$

Par cette construction, la suite (K_n) est strictement croissante pour l'inclusion. La réunion de cette suite est exactement I. En effet elle est évidemment incluse dans I puisque chaque K_n est une partie de I. Dans l'autre sens, si i est un élément quelconque de I, $\varphi^{-1}(i)$ est un entier qui est majoré par k_n pour n assez grand (la suite *strictement* croissante *d'entiers* (k_n) tend vers l'infini). Par construction de J_n, l'inégalité $\varphi^{-1}(i) \leqslant k_n$ implique l'appartenance de i à J_n, donc aussi à K_n. Ainsi un élément quelconque i de I appartient toujours à au moins un K_n, donc $I \subset \bigcup_{n\in\mathbb{N}} K_n$ et finalement $I = \bigcup_{n\in\mathbb{N}} K_n$. Posant maintenant

$$I_0 := K_0 \quad \text{et} \quad \forall n \in \mathbb{N}^*, \quad I_n := K_n \backslash K_{n-1},$$

il est clair qu'aucun des I_n n'est vide (par stricte croissance de (K_n)), qu'ils sont deux à deux disjoints (pour la même raison) et que leur réunion est I. La famille $\{I_n;\ n \in \mathbb{N}\}$ constitue donc une partition de I en sous-ensembles finis. Notons maintenant

$$\forall n \in \mathbb{N}, \quad m_n := \operatorname{card} K_n - 1.$$

Comme (K_n) est strictement croissante pour l'inclusion, la suite d'entiers (m_n) est strictement croissante. On obtient alors une partition $\{A_n;\ n \in \mathbb{N}\}$ de $\mathbb{N}$ en posant :

$$A_0 := [\![0, m_0]\!], \quad \text{et} \quad \forall n \in \mathbb{N}^*, \quad A_n :=]\!]m_{n-1}, m_n]\!].$$

Les A_n sont finis, $\operatorname{card} A_0 = m_0 + 1 = \operatorname{card} K_0 = \operatorname{card} I_0$ et pour $n \geqslant 1$,

$$\operatorname{card} A_n = m_n - m_{n-1} = \operatorname{card} K_n - \operatorname{card} K_{n-1} = \operatorname{card} I_n.$$

Dire qu'A_n et I_n ont même cardinal c'est dire précisément qu'il existe une bijection $f_n : A_n \to I_n$. En « recollant » ces bijections f_n entre ensembles finis, on construit une application $f : \mathbb{N} \to I$: tout $k \in \mathbb{N}$ appartient à un unique A_n (puisque les A_n partitionnent $\mathbb{N}$) et on pose alors

$$f(k) := f_n(k), \quad (k \in A_n).$$

Vérifions que f ainsi définie est une bijection. Tout $i \in I$ appartient à un I_n (puisque les I_n partitionnent I). Il a alors pour antécédent par f l'entier $f_n^{-1}(i) \in A_n$. Ceci montre que f est surjective. Pour vérifier l'injectivité, soient k et l deux entiers distincts. Ou

bien ils sont dans le même A_n et alors $f(k) = f_n(k) \neq f_n(l) = f(l)$ par injectivité de f_n. Ou bien $k \in A_n$ et $l \in A_{n'}$ avec $n \neq n'$. Alors $f(k) \in I_n$ et $f(l) \in I_{n'}$ et comme I_n et $I_{n'}$ sont disjoints, $f(k)$ et $f(l)$ sont forcément distincts. Nous avons finalement construit une suite strictement croissante d'entiers m_n (donc tendant vers l'infini) et une bijection $f : \mathbb{N} \to I$ telles que $f\big([\![0, m_n]\!]\big) = K_n$ pour tout $n \in \mathbb{N}$. Au vu de (2.20), nous avons ainsi

$$\forall n \in \mathbb{N}, \quad \Big\| \sum_{k=0}^{m_n} u_{f(k)} - S \Big\| \geqslant \varepsilon,$$

ce qui interdit la convergence vers S de la série de terme général $u_{f(k)}$. L'existence d'une bijection f pour laquelle la série de terme général $u_{f(k)}$ ne converge pas vers S est précisément la propriété « non b) ». □

Le prochain théorème est l'analogue du critère de Cauchy pour les familles sommables. Il est surtout important par ses corollaires. Il s'applique avec E espace vectoriel normé *complet* donc en particulier avec $E = \mathbb{R}$ ou $\mathbb{C}$.

Théorème 2.46 (critère de Cauchy pour la sommabilité)**.** *Soit E un espace vectoriel normé* complet*. La famille $(u_i)_{i\in I}$ d'éléments de E est sommable si et seulement si*

$$\forall \varepsilon > 0,\ \exists J \text{ fini } \subset I,\ \forall H \text{ fini } \subset I\backslash J, \quad \|S_H\| < \varepsilon. \tag{2.21}$$

Avant d'attaquer la preuve du théorème, faisons un peu de recherche en paternité en comparant (2.21) avec le critère de Cauchy classique pour une série. Supposons donc pour un instant que $I = \mathbb{N}$. Le critère de Cauchy classique pour la série de terme général u_k s'écrit :

$$\forall \varepsilon > 0,\ \exists n_0 \in \mathbb{N},\ \forall p, q \geqslant n_0, \quad \|S_p - S_q\| < \varepsilon. \tag{2.22}$$

On peut toujours par confort d'écriture supposer $p \leqslant q$. Définissons les ensembles $H_{p,q} :=]\!]p, q]\!]$. Avec cette notation, on peut réécrire (2.22) sous la forme équivalente

$$\forall \varepsilon > 0,\ \exists H_{0,n_0} \subset \mathbb{N},\ \forall H_{p,q} \subset \mathbb{N}\backslash H_{0,n_0}, \quad \big\|S_{H_{p,q}}\big\| < \varepsilon. \tag{2.23}$$

Cette écriture établit la filiation de (2.21) à partir de (2.22) et nous permet de voir que (2.21) implique (2.22). La réciproque est fausse en raison de la remarque 2.41 et du théorème 2.46. Par exemple la série de terme général $u_k = (-1)^k/(k+1)$ est convergente donc vérifie le critère de Cauchy (2.22). Pourtant elle n'est pas sommable, donc ne vérifie pas (2.21).

Preuve de « sommabilité ⇒ (2.21) ». Supposons d'abord que $(u_i)_{i\in I}$ soit sommable. Pour tout $\varepsilon > 0$, il existe une partie finie J de I telle que pour toute partie finie K de I contenant J, $\|S_K - S\| < \varepsilon/2$. En appliquant cette inégalité avec $K = J$, puis $K = J \cup H$, pour H partie finie quelconque de $I\backslash J$, on obtient :

$$\|S_J - S\| < \frac{\varepsilon}{2} \quad \text{et} \quad \|S_{J\cup H} - S\| < \frac{\varepsilon}{2}.$$

Les ensembles d'indices J et H étant disjoints et finis, $S_{J\cup H} = S_J + S_H$. En écrivant $S_H = (S_{J\cup H} - S) - (S_J - S)$, l'inégalité triangulaire nous donne $\|S_H\| < \varepsilon$. Ceci établit (2.21). Notons que nous n'avons pas utilisé la complétude de E dans cette partie. □

Preuve de « (2.21) ⇒ sommabilité ». Pour établir la réciproque, supposons que la famille $(u_i)_{i\in I}$ d'éléments de E vérifie (2.21).

Vérifions d'abord que l'ensemble d'indices $I' := \{i \in I;\, u_i \neq 0\}$ est au plus dénombrable [10]. En appliquant (2.21) avec $\varepsilon = 1/n$, pour tout $n \geqslant 1$, on obtient une suite de parties finies J_n de I telles que si $H \subset I\backslash J_n$ est fini, $\|S_H\| < 1/n$. Ceci est vrai en particulier avec $H = \{i\}$, pour $i \notin J_n$. Donc si $i \notin J_n$, $\|u_i\| < 1/n$. Ainsi si $i \notin \bigcup_{n\geqslant 1} J_n$, autrement dit si $i \in \bigcap_{n\geqslant 1}(I\backslash J_n)$, on a $\|u_i\| < 1/n$ pour *tout* $n \geqslant 1$, d'où $\|u_i\| = 0$, puis $u_i = 0$. On en déduit que I' est inclus dans $\bigcup_{n\geqslant 1} J_n$. Ce dernier ensemble est au plus dénombrable comme union dénombrable d'ensembles finis, donc I' est lui-même au plus dénombrable.

La sommabilité de $(u_i)_{i\in I}$ est immédiate dans le cas I' fini : on pose $J = I'$, $S := S_{I'}$ (bien défini comme somme d'un nombre fini de vecteurs de E). Pour tout K fini contenant I', $S_K = S_{I'} + S_{K\backslash I'} = S_{I'}$ puisque pour les $i \in K\backslash I'$, $u_i = 0$. On a donc $S_K = S$ pour tout $K \supset I'$, donc *a fortiori* $\|S_K - S\| < \varepsilon$ pour tout $\varepsilon > 0$ (ici on a même J indépendant de ε). Ainsi $(u_i)_{i\in I}$ est sommable de somme S.

Passons au cas I' infini (donc ici dénombrable). Fixons une bijection $f : \mathbb{N} \to I'$ et définissons pour $k \in \mathbb{N}$, $v_k := u_{f(k)}$. Notons $T_n = v_0 + \cdots + v_n$ la n-ième somme partielle de la série de vecteurs $\sum_{k=0}^{+\infty} v_k$. Vérifions que la suite $(T_n)_{n\in\mathbb{N}}$ satisfait au critère de Cauchy classique dans E. Pour $\varepsilon > 0$, choisissons l'un des ensembles J associés à ε par (2.21). Il est clair que l'on peut remplacer J par $J' := J \cap I'$ puisque pour les $i \notin I'$, $u_i = 0$. Comme J' est fini et f surjective, il existe un entier m_ε tel que $f\big([\![0, m_\varepsilon]\!]\big)$ recouvre J'. Alors pour $m_\varepsilon \leqslant m \leqslant n$, $H := f(]\!]m, n]\!])$ est fini et disjoint de J', donc par (2.21), $\|S_H\| < \varepsilon$, ce qui peut encore s'écrire $\|T_n - T_m\| < \varepsilon$. Ainsi la suite T_n est de Cauchy dans l'espace *complet* E. Elle est donc convergente. Notons S sa limite.

Montrons pour finir que $(u_i)_{i\in I}$ est sommable de somme S. En conservant les notations qui ont servi à vérifier que (T_n) est de Cauchy et en choisissant pour tout $n \geqslant m_\varepsilon$, $H_n := f(]\!]m_\varepsilon, n]\!])$, on a $\|T_n - T_{m_\varepsilon}\| < \varepsilon$. Compte-tenu de la convergence de T_n vers S, on en déduit en faisant tendre n vers l'infini que $\|S - T_{m_\varepsilon}\| \leqslant \varepsilon$. Notons $J'' := f([\![0, m_\varepsilon]\!])$ et rappelons que $J'' \supset J' = I' \cap J$, où J est associé à ε par (2.21). D'autre part $S_{J''} = T_{m_\varepsilon}$, d'où $\|S_{J''} - S\| \leqslant \varepsilon$. Soit K une partie finie de I, contenant J''. En écrivant que

$$S_K = S_{J''} + S_{K\backslash J''} = S + (S_{J''} - S) + S_{K\backslash J''},$$

on obtient par inégalité triangulaire

$$\|S_K - S\| \leqslant \|S_{J''} - S\| + \|S_{K\backslash J''}\| \leqslant \varepsilon + \|S_{K\backslash J''}\|. \tag{2.24}$$

10. Attention, ceci ressemble à la proposition 2.44, mais ici on suppose (2.21) vraie au lieu de la sommabilité de $(u_i)_{i\in I}$.

Puisque $K\backslash J''$ est fini et disjoint de J' (car $J' \subset J''$), on a $\|S_{K\backslash J''}\| < \varepsilon$. En reportant ceci dans (2.24), on voit que pour tout K fini contenant J'', $\|S_K - S\| < 2\varepsilon$. Ce raisonnement étant valable pour tout $\varepsilon > 0$ (avec J'' dépendant de ε), on conclut que $(u_i)_{i\in I}$ est sommable de somme S. □

Corollaire 2.47. *Soit $(u_i)_{i\in I}$ une famille d'éléments de l'espace vectoriel normé* complet *E. Si $(\|u_i\|)_{i\in I}$ est sommable dans $\mathbb{R}_+$, alors $(u_i)_{i\in I}$ est sommable dans E. On dit dans ce cas qu'elle est* normalement [11] *sommable ou si $E = \mathbb{R}$ ou $\mathbb{C}$,* absolument *sommable.*

Preuve. En appliquant la partie « sommabilité ⇒ (2.21) » du théorème 2.46 à la famille $(\|u_i\|)_{i\in I}$, ici l'espace vectoriel est $\mathbb{R}$, on obtient :

$$\forall \varepsilon > 0,\ \exists J \text{ fini } \subset I,\ \forall H \text{ fini } \subset I\backslash J,\quad \sum_{i\in H} \|u_i\| < \varepsilon.$$

Par inégalité triangulaire on en déduit :

$$\|S_H\| \leqslant \sum_{i\in H} \|u_i\| < \varepsilon.$$

La partie « (2.21) ⇒ sommabilité » du théorème 2.46 appliquée cette fois à la famille $(u_i)_{i\in I}$ dans l'espace E nous fournit la sommabilité de cette dernière. □

Nous sommes maintenant en mesure de faire un point définitif sur les différents modes de sommation étudiés pour une famille dans $\mathbb{R}$ ou $\mathbb{C}$.

Proposition 2.48. *Pour une famille $(u_i)_{i\in I}$ avec I dénombrable et les u_i dans $\mathbb{R}$ ou $\mathbb{C}$, les trois propriétés sommabilité, convergence commutative et sommabilité absolue sont équivalentes.*

Preuve. Cette équivalence résulte du théorème 2.45, du corollaire 2.47 et des théorèmes 2.27 et 2.35. □

Le prochain corollaire nous sera utile pour établir le théorème de sommation par paquets.

Corollaire 2.49 (du th. 2.46)**.** *Soit $(u_i)_{i\in I}$ une famille sommable d'éléments de l'espace vectoriel normé* complet *E.*

a) Pour toute partie L de I, $(u_i)_{i\in L}$ est sommable. On notera S_L sa somme.

b) Pour tout $\varepsilon > 0$, soit J une partie finie de I telle que pour toute partie K de I finie et contenant J, $\|S - S_K\| < \varepsilon$. Alors $\|S - S_L\| \leqslant \varepsilon$ pour toute partie $L \supset J$ de I.

11. Dans la théorie des séries de fonctions, on dit qu'une série $\sum_{n=0}^{+\infty} f_n$ de fonctions $f_n : D \to \mathbb{C}$ (ou $\mathbb{R}$) est normalement convergente sur D s'il existe une suite de réels positifs $(a_n)_{n\in\mathbb{N}}$ telle que $\sum_{n=0}^{+\infty} a_n < +\infty$ et pour tout $n \in \mathbb{N}$, tout $t \in D$, $|f_n(t)| \leqslant a_n$. En fait, ce n'est qu'un cas particulier de la convergence normale ci-dessus en considérant les f_n comme des vecteurs de l'espace vectoriel des fonctions bornées sur D, muni de la norme $\|f\|_\infty := \sup_{t\in D} |f(t)|$.

c) Si $L_1, \dots, L_n$ sont des parties deux à deux disjointes de I, de réunion L, alors $S_L = S_{L_1} + \cdots + S_{L_n}$.

Évidemment, ce corollaire n'a d'intérêt que si L est infini.

Preuve du a). Utilisons le critère de Cauchy (2.21).

$$\forall \varepsilon > 0,\ \exists J \text{ fini } \subset I,\ \forall K \text{ fini } \subset I \backslash J, \quad \|S_K\| < \varepsilon.$$

Ceci est vrai en particulier pour toute partie finie K de L disjointe de J, autrement dit pour toute partie finie K de L disjointe de $J' := J \cap L$. Nous avons donc

$$\forall \varepsilon > 0,\ \exists J' \text{ fini } \subset L,\ \forall K \text{ fini } \subset L \backslash J', \quad \|S_K\| < \varepsilon.$$

Ainsi $(u_i)_{i\in L}$ est sommable par le critère de Cauchy (2.21). □

Preuve du b). Soit $\varepsilon > 0$ et J une partie finie de I telle que $\|S - S_K\| < \varepsilon$ pour tout K fini tel que $J \subset K \subset I$. Soit L une partie de I contenant J. Par le a), $(u_i)_{i\in L}$ est sommable. Il existe donc pour tout $n \in \mathbb{N}^*$ une partie finie J_n de L telle que si K fini vérifie $J_n \subset K \subset L$, $\|S_L - S_K\| < 1/n$. En particulier pour $K = J \cup J_n$, on a simultanément $\|S - S_K\| < \varepsilon$ et $\|S_L - S_K\| < 1/n$. Alors par inégalité triangulaire $\|S - S_L\| < \varepsilon + \frac{1}{n}$. Cette inégalité étant vraie pour tout n et son premier membre ne dépendant pas de n, on en déduit en faisant tendre n vers $+\infty$ avec ε fixé, que $\|S - S_L\| \leqslant \varepsilon$. □

Preuve du c). Par le a), on sait que les n sous-familles $(u_i)_{i\in L_l}$, $l = 1, \dots, n$ sont sommables. Ainsi pour tout $\varepsilon > 0$, tout $l = 1, \dots, n$, il existe J_l partie finie de L_l telle que pour tout K_l fini vérifiant $J_l \subset K_l \subset L_l$, $\|S_{K_l} - S_{L_l}\| < \varepsilon/n$. Posons $J = J_1 \cup \cdots \cup J_n$ et soit K une partie finie quelconque de $L = L_1 \cup \cdots \cup L_n$, contenant J. En prenant $K_l := K \cap L_l$, les K_l sont des ensembles finis deux à deux disjoints et de réunion K, d'où $S_K = S_{K_1} + \cdots + S_{K_n}$. Alors par inégalité triangulaire :

$$\Big\| S_K - \sum_{l=1}^{n} S_{L_l} \Big\| = \Big\| \sum_{l=1}^{n} (S_{K_l} - S_{L_l}) \Big\| \leqslant \sum_{l=1}^{n} \|S_{K_l} - S_{L_l}\| < n\frac{\varepsilon}{n} = \varepsilon.$$

On a ainsi vérifié que $(u_i)_{i\in L}$ est sommable de somme $\sum_{l=1}^{n} S_{L_l}$. □

Théorème 2.50 (de sommation par paquets)**.** *Soit $(u_i)_{i\in I}$ une famille d'éléments de l'espace vectoriel normé* complet *E. On suppose que $I = \bigcup_{\alpha\in A} I_\alpha$, les I_α étant non vides et deux à deux disjoints. Si $(u_i)_{i\in I}$ est sommable de somme S, alors chacune des familles $(u_i)_{i\in I_\alpha}$ est sommable et en notant S_{I_α} sa somme, la famille $(S_{I_\alpha})_{\alpha\in A}$ est sommable de somme S. Autrement dit on a la formule de « sommation par paquets » :*

$$\sum_{i\in I} u_i = \sum_{\alpha\in A} \Big(\sum_{i\in I_\alpha} u_i \Big). \tag{2.25}$$

Remarquons que le cas où A est fini est déjà traité par le corollaire 2.49 c).

Preuve. Pour tout $\alpha \in A$, la sous-famille $(u_i)_{i\in I_\alpha}$ est sommable par le corollaire 2.49 a). Pour toute partie finie C de A, nous pouvons alors définir $T_C := \sum_{\alpha\in C} S_{I_\alpha}$, somme d'un nombre fini de vecteurs de E. De même, $I_C := \bigcup_{\alpha\in C} I_\alpha$ étant une partie de I, la famille $(u_i)_{i\in I_C}$ est sommable et nous notons sa somme S_{I_C}. Par le c) du corollaire 2.49, $T_C = S_{I_C}$.

Puisque $(u_i)_{i\in I}$ est sommable de somme S, on dispose pour tout $\varepsilon > 0$ d'un ensemble fini $J \subset I$ tel que pour tout K fini vérifiant $J \subset K \subset I$, $\|S - S_K\| < \varepsilon$. Comme J est fini et $I = \bigcup_{\alpha\in A} I_\alpha$, il existe une partie finie B de A telle que I_B contienne J. Pour tout C fini inclus dans A et contenant B, $I_C \supset I_B \supset J$, donc par le corollaire 2.49 b), même si I_C est infini, $\|S - S_{I_C}\| \leqslant \varepsilon$. Compte-tenu de l'égalité $S_{I_C} = T_C$, on a ainsi montré que

$$\forall \varepsilon > 0,\ \exists B \text{ fini } \subset A,\ \forall C \text{ fini tel que } B \subset C \subset A, \quad \|S - T_C\| \leqslant \varepsilon.$$

Autrement dit, $(S_{I_\alpha})_{\alpha\in A}$ est sommable de somme S. □

Remarque 2.51. Il importe de noter que la sommabilité de la famille $(S_{I_\alpha})_{\alpha\in A}$ n'implique pas celle de la famille $(u_i)_{i\in I}$. Voici un contre exemple avec $E = \mathbb{R}$: on prend $I = \mathbb{Z}$, $u_i = i$, $A = \mathbb{N}$ et pour tout $k \in A$, $I_k = \{-k, k\}$. Alors pour tout $k \in A$, $S_{I_k} = 0$ et la famille $(S_{I_k})_{k\in\mathbb{N}}$ est sommable de somme 0. En revanche, la famille $(u_i)_{i\in\mathbb{Z}}$ n'est pas sommable puisque $\sum_{i\in\mathbb{Z}} |i| = +\infty$.

Les familles de réels positifs constituent à nouveau un cas particulier pour la sommation par paquets (et une exception à la remarque 2.51). Il est naturel ici d'élargir un peu le problème en considérant les familles d'éléments de $\overline{\mathbb{R}}_+$. On prolonge l'addition de $\mathbb{R}_+$ à $\overline{\mathbb{R}}_+$ en posant

$$\forall x \in \mathbb{R}_+,\ x + (+\infty) = (+\infty) + x = +\infty \quad \text{et} \quad (+\infty) + (+\infty) = +\infty.$$

Avec cette convention, on peut toujours définir la somme de n'importe quelle famille d'éléments de $\overline{\mathbb{R}}_+$ en utilisant (2.18).

Théorème 2.52 (de sommation par paquets dans $\overline{\mathbb{R}}_+$)**.** *Toute famille $(u_i)_{i\in I}$ d'éléments $\overline{\mathbb{R}}_+$ vérifie la formule de sommation par paquets (2.25) interprétée comme égalité dans $\overline{\mathbb{R}}_+$.*

Preuve. Si $M := \sum_{i\in I} u_i < +\infty$, nécessairement tous les u_i sont finis et la famille $(u_i)_{i\in I}$ est sommable (prop 2.42). Il n'y a alors rien à démontrer puisqu'on est dans le champ d'application du théorème 2.50. La démonstration générale par les suprema que l'on présente maintenant englobe ce cas et a l'avantage d'éviter de discuter suivant la finitude ou non des u_i ou des S_{I_α}. Notons

$$M := \sup_{K \text{ fini } \subset I} S_K \quad \text{et} \quad M' := \sup_{B \text{ fini } \subset A} S_{I_B}.$$

Avec ces notations, l'égalité (2.25) s'écrit $M = M'$. Nous allons montrer que $M' \leqslant M$ et que $M \leqslant M'$. Pour la première inégalité, il suffit de remarquer que pour tout B fini inclus dans A,

$$S_{I_B} = \sup_{L \text{ fini } \subset I_B} S_L, \tag{2.26}$$

d'où en majorant S_L par M, $S_{I_B} \leqslant M$, puis en prenant le supremum pour tout B fini inclus dans A, $M' \leqslant M$. Dans l'autre sens, soit K une partie finie quelconque de I. Il existe une partie finie B de A telle que $I_B = \bigcup_{\alpha\in B} I_\alpha$ recouvre K. D'après (2.26) on a alors $S_K \leqslant S_{I_B}$. En majorant S_{I_B} par M', on voit que pour tout K fini inclus dans I, $S_K \leqslant M'$. En prenant le supremum sur K, on obtient $M \leqslant M'$. □

Théorème 2.53. *Soit $(u_i)_{i\in I}$ une famille d'éléments de l'espace vectoriel normé* complet *E. On suppose que $I = \bigcup_{\alpha\in A} I_\alpha$, les I_α étant non vides et deux à deux disjoints. On pose $T_{I_\alpha} := \sum_{i\in I_\alpha} \|u_i\|$. La famille $(u_i)_{i\in I}$ est normalement sommable dans E si et seulement si $\sum_{\alpha\in A} T_{I_\alpha} < +\infty$. Elle vérifie alors la formule de sommation par paquets* (2.25).

Preuve. La sommabilité dans $\mathbb{R}_+$ de $(\|u_i\|)_{i\in I}$ équivaut à $\sum_{\alpha\in A} T_{I_\alpha} < +\infty$ d'après le théorème 2.52 et la proposition 2.42. Si $(u_i)_{i\in I}$ est normalement sommable dans E (cf. corollaire 2.47) elle est sommable. Elle vérifie donc la formule de sommation par paquets par le théorème 2.50. □

2.3 Séries doubles

Nous examinons maintenant le cas particulier des familles $(u_i)_{i\in I}$ indexées par un produit cartésien $I = I' \times I''$. On parle dans ce cas de « série double ». Nous nous limiterons au cas où $I = \mathbb{N}^2$, mais il est facile d'adapter les énoncés ci-dessous à $I = I' \times I''$ avec I' et I'' dénombrables, une fois fixées des bijections $\mathbb{N} \to I$ et $\mathbb{N} \to I''$. L'indice i est désormais un couple d'entiers, $i = (k, l) \in \mathbb{N}^2$. Pour des raisons typographiques, nous écrirons $\{u_{k,l}\,;\,(k,l)\in\mathbb{N}^2\}$ de préférence à $(u_{(k,l)})_{(k,l)\in\mathbb{N}^2}$.

Définition 2.54 (série double)**.** *Soit $(u_{k,l})_{k,l\in\mathbb{N}}$ une suite double d'éléments de l'espace vectoriel normé E. On dit que la série double de terme général $u_{k,l}$ est convergente (resp. normalement convergente) si la famille $\{u_{k,l}\,;\,(k,l)\in\mathbb{N}^2\}$ est sommable (resp. normalement sommable). La somme S de cette famille est alors appelée somme de la série double et notée*

$$S = \sum_{(k,l)\in\mathbb{N}^2} u_{k,l} = \sum_{k,l\in\mathbb{N}} u_{k,l}.$$

Dans tout ce qui suit, nous donnerons les énoncés relatifs aux séries doubles à termes réels ou complexes en nous contentant d'indiquer sous forme de remarques ou de commentaires les modifications à apporter pour la généralisation aux espaces vectoriels normés. Pour $E = \mathbb{R}$ ou $\mathbb{C}$, la sommabilité normale s'appelle sommabilité absolue et équivaut à la sommabilité. Il importe de bien comprendre que contrairement au cas des séries simples, la définition 2.54 n'autorise pas l'existence d'une série double à termes dans $\mathbb{R}$ ou $\mathbb{C}$ qui soit convergente sans l'être absolument. Il peut arriver que pour une certaine bijection $f : \mathbb{N} \to \mathbb{N}^2$, la série simple $\sum_{j=0}^{+\infty} u_{f(j)}$ soit convergente sans l'être absolument, mais dans ce cas la série double $\sum_{k,l} u_{k,l}$ est divergente (cf. th. 2.45 et th. 2.35).

Voici un premier critère de convergence pour les séries doubles.

Proposition 2.55. *Pour que la série double de terme général réel ou complexe $u_{k,l}$ $(k,l \in \mathbb{N})$ soit convergente, il faut et il suffit qu'il existe une suite $(K_n)_{n\in\mathbb{N}}$ croissante pour l'inclusion*[12] *de parties finies de $\mathbb{N}^2$, telle que $\bigcup_{n\in\mathbb{N}} K_n = \mathbb{N}^2$ et que la suite de terme général*

$$T_n := \sum_{(k,l)\in K_n} |u_{k,l}|,$$

soit bornée.

Cet énoncé se généralise aux séries à termes dans E espace vectoriel normé *complet* en remplaçant « convergente » par « normalement convergente » et $|u_{k,l}|$ par $\|u_{k,l}\|$. Comme exemples de suites $(K_n)_{n\in\mathbb{N}}$, on peut prendre les « triangles » emboîtés $K_n = \{(k,l) \in \mathbb{N}^2\,;\, k+l \leqslant n\}$ ou les « carrés » emboîtés $K'_n = [\![0,n]\!]^2$.

Preuve. Vérifions d'abord que la condition est nécessaire en supposant que la série double de terme général $u_{k,l}$ est convergente. Alors la famille $\{u_{k,l}\,;\,(k,l)\in\mathbb{N}^2\}$ est absolument sommable et d'après la proposition 2.42 b) :

$$M := \sup_{K \text{ fini}\subset\mathbb{N}^2} \sum_{(k,l)\in K} |u_{k,l}| < +\infty. \tag{2.27}$$

Clairement $T_n \leqslant M$ pour tout n puisque T_n est la somme des termes indexés par le sous-ensemble fini K_n de $\mathbb{N}^2$. Donc la suite $(T_n)_{n\in\mathbb{N}}$ est bornée.

Réciproquement, supposons que la suite $(T_n)_{n\in\mathbb{N}}$ soit bornée par un réel positif M_1. Soit K une partie finie quelconque de $\mathbb{N}^2$. Il existe alors un entier n tel que K_n recouvre K. En effet posons pour tout $i = (k,l) \in K$, $m(i) = \min\{j \in \mathbb{N}\,;\, i \in K_j\}$. Comme $\bigcup_{j\in\mathbb{N}} K_j = \mathbb{N}^2$, il y a au moins un j tel que $i \in K_j$ et comme toute partie non vide de $\mathbb{N}$ a un plus petit élément, $m(i)$ est bien un entier[13]. En choisissant $n = \max\{m(i)\,;\, i \in K\}$, on voit grâce à la croissance pour l'inclusion de $(K_j)_{j\in\mathbb{N}}$, que K_n recouvre K. Alors

$$\sum_{(k,l)\in K} |u_{k,l}| \leqslant \sum_{(k,l)\in K_n} |u_{k,l}| = T_n \leqslant M_1$$

et l'ensemble fini K étant quelconque, ceci montre que (2.27) est vérifiée avec $M \leqslant M_1$. La sommabilité absolue de $\{u_{k,l};\,(k,l)\in\mathbb{N}^2\}$ en découle par la proposition 2.42 a). □

Le théorème suivant fournit à la fois un critère de convergence des série doubles et une méthode de calcul de la somme.

Théorème 2.56 (d'interversion des sommations)**.**

1. *Si les $v_{k,l}$ sont des éléments de $\overline{\mathbb{R}}_+$, les égalités*

$$\sum_{(k,l)\in\mathbb{N}^2} v_{k,l} = \sum_{k=0}^{+\infty}\left(\sum_{l=0}^{+\infty} v_{k,l}\right) = \sum_{l=0}^{+\infty}\left(\sum_{k=0}^{+\infty} v_{k,l}\right) \tag{2.28}$$

12. C'est-à-dire telle que pour tout $n \in \mathbb{N}$, $K_n \subset K_{n+1}$.
13. Rappelons que $\inf \varnothing = +\infty$.

sont toujours vérifiées dans $\overline{\mathbb{R}}_+$.

2. *Si les* $u_{k,l}$ *sont réels ou complexes, la série double de terme général* $u_{k,l}$ *converge si et seulement si l'une des deux conditions suivantes est réalisée :*

$$a)\ \sum_{k=0}^{+\infty}\left(\sum_{l=0}^{+\infty}|u_{k,l}|\right)<+\infty \quad ou \quad b)\ \sum_{l=0}^{+\infty}\left(\sum_{k=0}^{+\infty}|u_{k,l}|\right)<+\infty. \tag{2.29}$$

Dans ce cas, pour tout $k\in\mathbb{N}$, *la série simple* $\sum_{l=0}^{+\infty}u_{k,l}$ *est absolument convergente et il en va de même en échangeant les rôles de* k *et* l. *On a de plus la formule d'interversion des sommations :*

$$\sum_{(k,l)\in\mathbb{N}^2}u_{k,l}=\sum_{k=0}^{+\infty}\left(\sum_{l=0}^{+\infty}u_{k,l}\right)=\sum_{l=0}^{+\infty}\left(\sum_{k=0}^{+\infty}u_{k,l}\right). \tag{2.30}$$

Ce théorème s'étend au cas où les $u_{k,l}$ sont dans E espace vectoriel normé complet en remplaçant $|u_{k,l}|$ par $\|u_{k,l}\|$, la convergence de la série double par sa convergence normale et la convergence absolue des séries simples par leur convergence normale.

Preuve. Le cas des termes positifs est une application immédiate du théorème 2.52 en considérant les deux décompositions suivantes de $\mathbb{N}^2$, cf. figure 2.2 :

$$a)\ \mathbb{N}^2=\bigcup_{k\in\mathbb{N}}\big(\{k\}\times\mathbb{N}\big) \qquad b)\ \mathbb{N}^2=\bigcup_{l\in\mathbb{N}}\big(\mathbb{N}\times\{l\}\big). \tag{2.31}$$

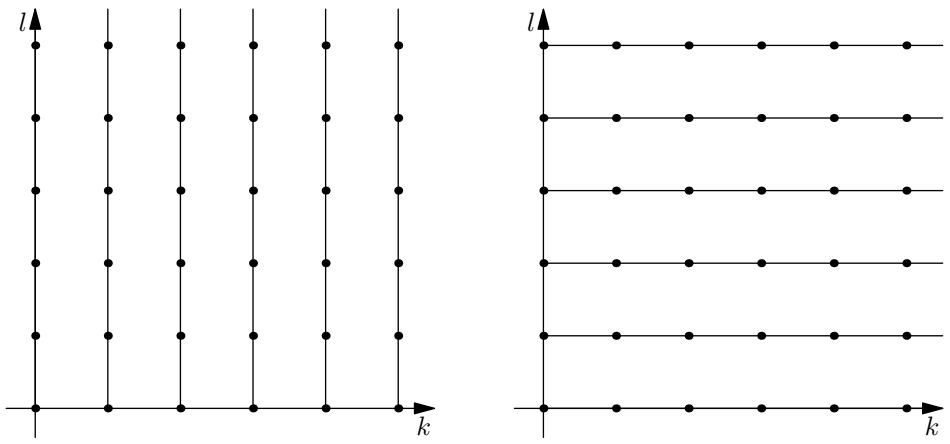


FIGURE 2.2 – Découpages a) et b) de $\mathbb{N}^2$

Dans le cas des termes réels ou complexes, la condition (2.29.a) utilisée avec la décomposition (2.31.a) équivaut à la sommabilité absolue de $\{u_{k,l}\,;\,(k,l)\in\mathbb{N}^2\}$ par le théorème 2.53. Il en va de même pour (2.29.b) utilisée avec la décomposition (2.31.b).

La sommabilité absolue de $\{u_{k,l};\ (k,l) \in \mathbb{N}^2\}$ implique *chacune* des conditions (2.29.a) et (2.29.b) par le théorème 2.53. La première des deux implique la convergence absolue de toutes les séries $\sum_{l=0}^{+\infty} u_{k,l}$, la deuxième celle des séries $\sum_{k=0}^{+\infty} u_{k,l}$.

Enfin lorsque $\{u_{k,l};\ (k,l) \in \mathbb{N}^2\}$ est sommable, on obtient (2.30) en appliquant la formule de sommation par paquets (th. 2.50) aux décompositions (2.31.a) et (2.31.b). □

Une application classique de la théorie des séries doubles est le produit de deux séries *absolument* convergentes.

Théorème 2.57 (produit de deux séries)**.** *Soient $\sum_{k=0}^{+\infty} u_k$ et $\sum_{l=0}^{+\infty} v_l$ deux séries à termes réels ou complexes,* absolument *convergentes. Alors la série double de terme général $u_k v_l$ est convergente et on a*

$$\sum_{(k,l)\in\mathbb{N}^2} u_k v_l = \left(\sum_{k=0}^{+\infty} u_k\right)\left(\sum_{l=0}^{+\infty} v_l\right) = \sum_{n=0}^{+\infty}\left(\sum_{k+l=n} u_k v_l\right). \tag{2.32}$$

La série simple de terme général $w_n = \sum_{k+l=n} u_k v_l$ est absolument convergente. On l'appelle série produit *de $\sum_{k=0}^{+\infty} u_k$ et $\sum_{l=0}^{+\infty} v_l$.*

Preuve. Commençons par vérifier la convergence de la série double. Posons

$$U_n := \sum_{k=0}^{n} u_k, \quad U := \sum_{k=0}^{+\infty} u_k, \quad U_n^{\mathrm{abs}} := \sum_{k=0}^{n} |u_k| \quad U^{\mathrm{abs}} := \sum_{k=0}^{+\infty} |u_k|$$

et définissons de même V_n, V, V_n^{abs} et V^{abs} en remplaçant u_k par v_l. Le produit

$$U_n^{\mathrm{abs}} V_n^{\mathrm{abs}} = \sum_{(k,l)\in[\![0,n]\!]^2} |u_k||v_l|$$

est borné par la constante $U^{\mathrm{abs}}V^{\mathrm{abs}}$ indépendante de n. La proposition 2.55 appliquée avec $K_n = [\![0,n]\!]^2$, nous assure de la convergence de la série double de terme général $u_k v_l$.

En appliquant la formule de sommation par paquets à cette série avec le découpage $\mathbb{N}^2 = \bigcup_{k\in\mathbb{N}}(\{k\}\times\mathbb{N})$, on obtient :

$$\sum_{(k,l)\in\mathbb{N}^2} u_k v_l = \sum_{k=0}^{+\infty}\left(\sum_{l=0}^{+\infty} u_k v_l\right),$$

où chacune des séries simples $\sum_{l=0}^{+\infty} u_k v_l$ est absolument convergente. Dans une telle série, on peut mettre en facteur la « constante » u_k, d'où

$$\sum_{(k,l)\in\mathbb{N}^2} u_k v_l = \sum_{k=0}^{+\infty} u_k\left(\sum_{l=0}^{+\infty} v_l\right).$$

Dans la série indexée par k, on peut mettre en facteur la « constante » $\left(\sum_{l=0}^{+\infty} v_l\right)$, d'où

$$\sum_{(k,l)\in\mathbb{N}^2} u_k v_l = \left(\sum_{l=0}^{+\infty} v_l\right) \sum_{k=0}^{+\infty} u_k,$$

ce qui achève la vérification de la première égalité dans (2.32).

La deuxième égalité dans (2.32) résulte de l'application du théorème de sommation par paquets à la famille sommable $\{u_k v_l;\ (k,l) \in \mathbb{N}^2\}$ avec le découpage $\mathbb{N}^2 = \bigcup_{n\in\mathbb{N}} \{(k,l) \in \mathbb{N}^2;\ k+l=n\}$.

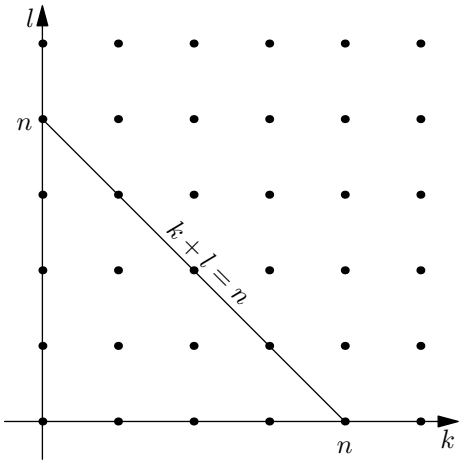


FIGURE 2.3 – Découpage $\mathbb{N}^2 = \bigcup_{n\in\mathbb{N}} \{(k,l) \in \mathbb{N}^2;\ k+l=n\}$

Le même découpage utilisé avec le théorème 2.53 nous fournit la convergence de la série à termes positifs

$$w_n^{\text{abs}} := \sum_{k+l=n} |u_k v_l|.$$

Cette dernière convergence entraîne par comparaison, la convergence absolue de la série de terme général w_n, puisque $|w_n| \leqslant w_n^{\text{abs}}$ par inégalité triangulaire.

□

2.4 Exercices

Ex. 2.1 *Série géométrique et découpage du carré*

La figure 2.4 page 65 propose une « preuve graphique » de la formule donnant la somme de la série géométrique $\sum_{k=1}^{+\infty} pq^{k-1}$, où $q \in]0,1[$ et $p = 1-q$. À vous de fournir les explications qui manquent pour rendre cette preuve convaincante.

Indications : commencez par noter sur la figure 2.5 page 66 l'aire de chacun des 3 rectangles issus du découpage du carré de côté c.

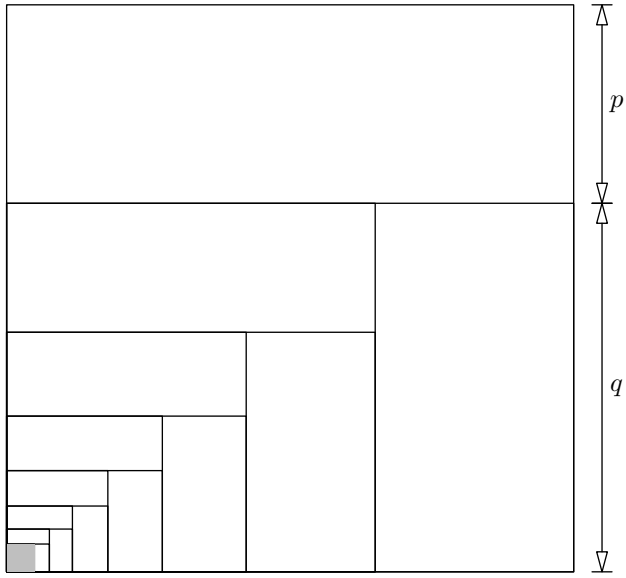


FIGURE 2.4 – Une méthode graphique pour calculer $\sum_{k=1}^{+\infty} pq^{k-1}$

Ex. 2.2 *Des trous dans la série harmonique*

On note C l'ensemble des entiers dont l'écriture en base 10 ne comporte aucun chiffre 7. On définit la suite $(u_n)_{n\in\mathbb{N}*}$ par

$$u_n = \begin{cases} \frac{1}{n} & \text{si } n \in C, \\ 0 & \text{si } n \notin C. \end{cases}$$

Que peut-on dire de la série de terme général u_n ? Que vaut $\sum_{n\in\mathbb{N}\setminus C} \frac{1}{n}$?

Indication p. 416

Ex. 2.3 *Développements d'un nombre décimal*

1) En calculant une somme de série, vérifier que le réel $0{,}379\,999\,999\,999\ldots$ est égal au nombre décimal 0,38.

2) L'écriture $0{,}379\,999\,999\,999\ldots$ porte le nom de « développement décimal illimité impropre » du nombre 0,38. On appelle « développement décimal illimité propre » d'un réel x, un développement qui ne comporte pas la répétition indéfinie du chiffre 9 à partir d'un certain rang. Par exemple $0{,}380\,000\,000\,000\ldots$ est le développement décimal illimité propre de 0,38. Y a-t-il des nombres réels qui n'ont pas de développement décimal illimité impropre ? Lesquels ?

Solution p. 416

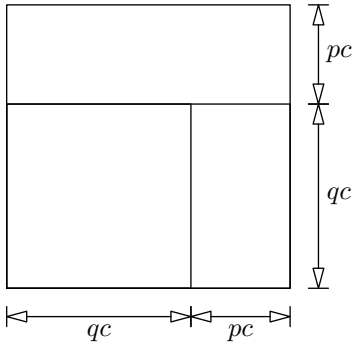


FIGURE 2.5 – Découpage du carré de côté c

Ex. 2.4 *Le théorème de convergence dominée pour les séries*

On considère une suite double de réels $\{u_{n,k}\ ;\ (n,k)\in\mathbb{N}^2\}$ et une suite de réels $(a_k)_{k\in\mathbb{N}}$ vérifiant les hypothèses suivantes.

a) $\forall (n,k)\in\mathbb{N}^2,\ |u_{n,k}|\leqslant a_k$.

b) $\forall k\in\mathbb{N},\ \lim\limits_{n\to+\infty} u_{n,k}=u_k\in\mathbb{R}$.

c) $\sum_{k=0}^{+\infty} a_k<+\infty$.

1) Montrer que les séries $\sum_{k=0}^{+\infty} u_{n,k}$ et $\sum_{k=0}^{+\infty} u_k$ convergent absolument dans $\mathbb{R}$.

2) Vérifier que pour tout entier N,

$$\Big|\sum_{k=0}^{+\infty} u_k-\sum_{k=0}^{+\infty} u_{n,k}\Big|\leqslant\Big|\sum_{k=0}^{N-1}(u_k-u_{n,k})\Big|+2\sum_{k=N}^{+\infty} a_k.$$

3) Prouver que

$$\lim_{n\to+\infty}\sum_{k=0}^{+\infty} u_{n,k}=\sum_{k=0}^{+\infty} u_k=\sum_{k=0}^{+\infty}\lim_{n\to+\infty} u_{n,k}.$$

4) Donner un exemple de suite double $\{u_{n,k}\ ;\ (n,k)\in\mathbb{N}^2\}$ vérifiant b) et telle que pour tout n la série $\sum_{k=0}^{+\infty} u_{n,k}$ converge absolument dans $\mathbb{R}$, mais pour laquelle la conclusion de la question précédente est en défaut.

Ex. 2.5 *Une somme doublement géométrique*

On définit la famille de réels $\mathcal{U}=\{u_{i,j}\ ;\ (i,j)\in\mathbb{N}^2\}$ par

$$\forall (i,j)\in\mathbb{N}^2,\quad u_{i,j}:=\frac{1}{18}\Big(\frac{5}{6}\Big)^i\Big(\frac{2}{3}\Big)^j.$$

1) Montrez que $\mathcal{U}$ est sommable et calculez sa somme.

2) On note L le sous ensemble de $\mathbb{N}^2$ défini par

$$L:=\big\{(i,j)\in\mathbb{N}^2;\ 0\leqslant i\leqslant j\big\}$$

et on s'intéresse à la famille $\mathcal{U}_L = \{u_{i,j}\ ;\ (i,j) \in L\}$.

a) Représentez graphiquement L.

b) Expliquez brièvement pourquoi $\mathcal{U}_L$ est sommable.

c) Calculez la somme de cette famille.

Solution p. 418

Ex. 2.6 La famille de réels $\{x^{-2}\ ;\ x \in \mathbb{Q} \cap [1, +\infty[\ \}$ est-elle sommable ?

Ex. 2.7 Soit $f : \mathbb{N} \to \mathbb{R}^+$, on pose pour i et j entiers, $u_{i,j} := f(i+j)$. Montrez que la famille $\{u_{i,j},\ (i,j) \in \mathbb{N}^2\}$ est sommable si et seulement si $\sum_{n=0}^{+\infty} nf(n) < +\infty$.

Ex. 2.8 Calculez en fournissant les justifications adéquates :

$$\sum_{(i,j)\in\mathbb{N}^2} \frac{1}{(1+i+j)!}.$$

Ex. 2.9 Soit $\{a_{n,m}\ ;\ (n,m) \in \mathbb{N}^2\}$ la famille de réels définie par

$$a_{n,m} = \frac{1}{n^2 - m^2} \quad \text{si} \quad n \neq m \quad \text{et} \quad a_{n,n} = 0.$$

1) La famille $\{a_{n,m}\ ;\ (n,m) \in \mathbb{N}^2\}$ est-elle sommable ?
2) Calculer

$$\sum_{n=0}^{+\infty} \sum_{m=0}^{+\infty} a_{n,m} \quad \text{et} \quad \sum_{m=0}^{+\infty} \sum_{n=0}^{+\infty} a_{n,m}.$$

Indications. Montrer que $a_{n,m} = \frac{1}{2n}(\frac{1}{n-m} + \frac{1}{n+m})$ (pour $n \neq 0$), en déduire que

$$\sum_{m=0}^{+\infty} a_{n,m} = \frac{1}{4n^2},$$

(toujours pour $n \neq 0$) et conclure.

Chapitre 3

Intégrale de Riemann sur $[a, b]$

3.1 Construction

Soit $[a, b]$ un intervalle *fermé borné* de $\mathbb{R}$. On appelle subdivision de $[a, b]$ toute suite finie du type $\Delta = \{x_0 = a < x_1 < \cdots < x_n = b\}$.

Définition 3.1 (sommes de Darboux)**.** *Soit $f : [a, b] \to \mathbb{R}$ une fonction* bornée *et Δ une subdivision de $[a, b]$. On leur associe les sommes de Darboux inférieure $S_\Delta(f)$ et supérieure $S^\Delta(f)$ définies par*

$$S_\Delta(f) := \sum_{k=1}^{n} (x_k - x_{k-1}) \inf_{[x_{k-1}, x_k]} f, \qquad S^\Delta(f) := \sum_{k=1}^{n} (x_k - x_{k-1}) \sup_{[x_{k-1}, x_k]} f. \quad (3.1)$$

Dans le cas trivial où $a = b$, on pose $S_\Delta(f) = S^\Delta(f) = 0$. Pour une illustration, voir les figures 3.1 et 3.2.

On dit que la subdivision Δ' est un raffinement de Δ si l'ensemble des valeurs de la suite finie Δ est inclus dans celui des valeurs de la suite Δ', ce que nous noterons avec un léger abus $\Delta \subset \Delta'$. Il est facile de vérifier que

$$\Delta \subset \Delta' \quad \Rightarrow \quad S_\Delta(f) \leqslant S_{\Delta'}(f) \text{ et } S^\Delta(f) \geqslant S^{\Delta'}(f). \quad (3.2)$$

Les figures 3.3 et 3.4 illustrent l'effet de l'insertion de deux nouveaux points dans la subdivision Δ des figures 3.1 et 3.2.

Les intégrales de Riemann *inférieure* $I_*(f)$ et *supérieure* $I^*(f)$ sont définies par

$$I_*(f) := \sup_\Delta S_\Delta(f), \quad I^*(f) := \inf_\Delta S^\Delta(f),$$

le supremum et l'infimum étant pris sur toutes les subdivisions Δ de $[a, b]$.

Il découle de (3.2) que pour toutes subdivisions Δ_1 et Δ_2 de $[a, b]$,

$$S_{\Delta_1}(f) \leqslant S_{\Delta_1 \cup \Delta_2}(f) \leqslant S^{\Delta_1 \cup \Delta_2}(f) \leqslant S^{\Delta_2}(f),$$

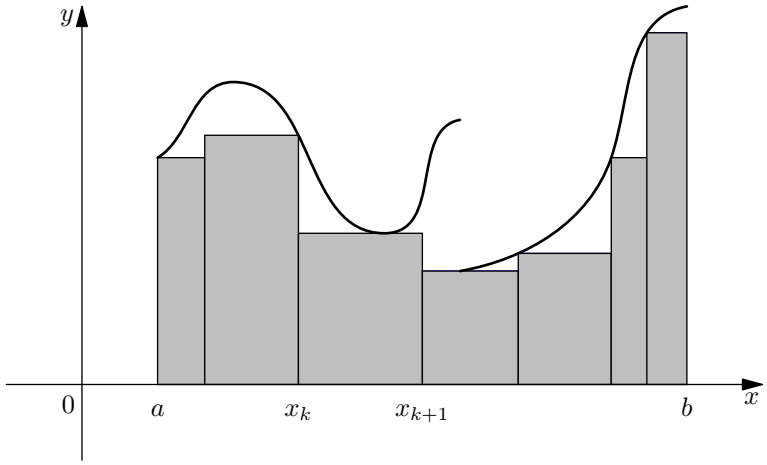


FIGURE 3.1 – $S_\Delta(f)$

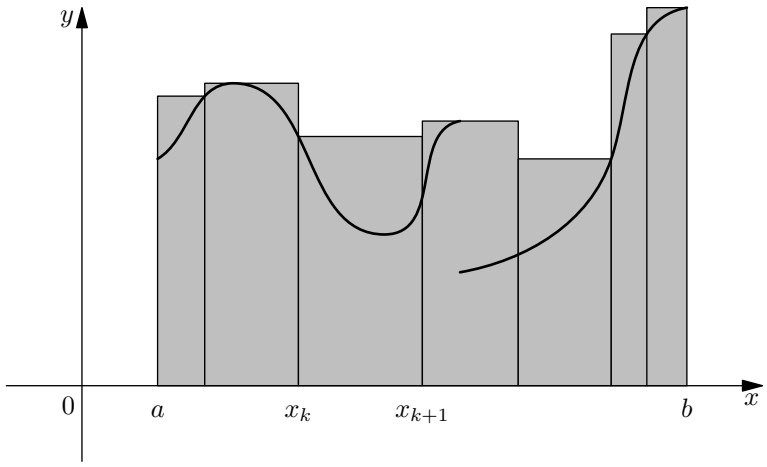


FIGURE 3.2 – $S^\Delta(f)$

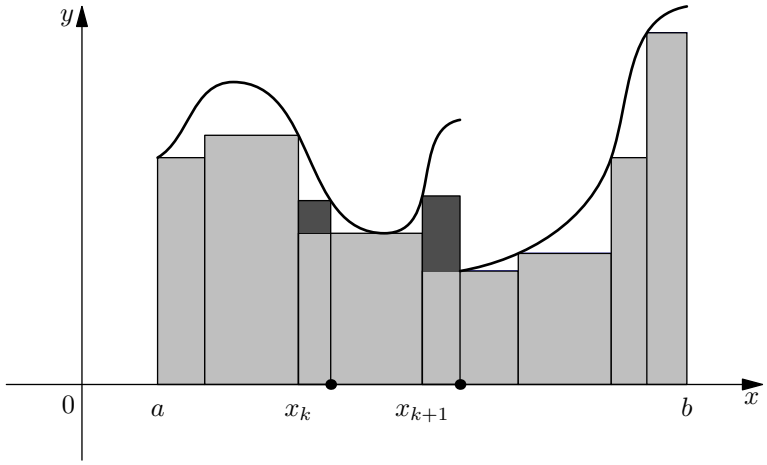


FIGURE 3.3 – Si $\Delta \subset \Delta'$, $S_\Delta(f) \leqslant S_{\Delta'}(f)$

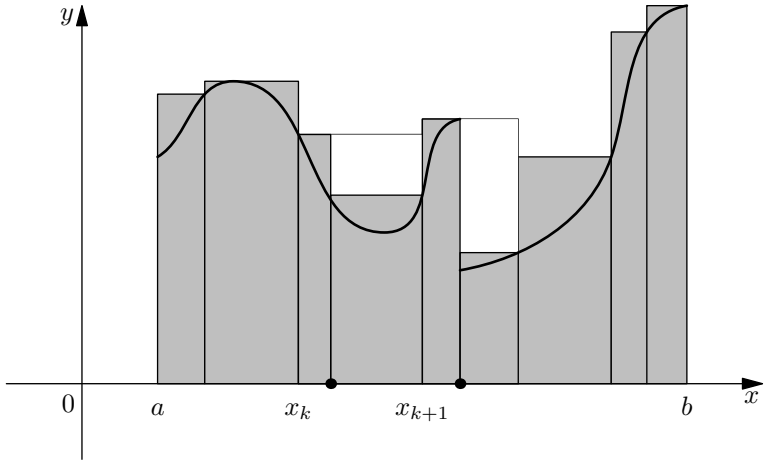


FIGURE 3.4 – Si $\Delta \subset \Delta'$, $S^\Delta(f) \geqslant S^{\Delta'}(f)$

donc en particulier $S_{\Delta_1}(f) \leqslant S^{\Delta_2}(f)$. En prenant successivement dans cette inégalité le sup sur tous les Δ_1, puis l'inf sur tous les Δ_2, on en déduit

$$I_*(f) \leqslant I^*(f), \tag{3.3}$$

inégalité vérifiée par toute fonction bornée $f : [a,b] \to \mathbb{R}$.

Définition 3.2. *On dit que f bornée $[a,b] \to \mathbb{R}$ est* Riemann intégrable *si, avec les notations ci-dessus, $I_*(f) = I^*(f)$. Dans ce cas, on définit son intégrale au sens de Riemann notée $\int_a^b f(x)\,\mathrm{d}x$ par*

$$\int_a^b f(x)\,\mathrm{d}x := I_*(f) = I^*(f).$$

Remarquons que f est toujours intégrable sur $[a,a]$ et que $\int_a^a f(x)\,\mathrm{d}x = 0$. Il est commode de donner aussi une définition de $\int_a^b f(x)\,\mathrm{d}x$ lorsque $b < a$. Cette définition peut se justifier en reprenant toute l'étude précédente avec des subdivisions Δ de $[b,a]$ par des suites finies décroissantes[1] $a = x_0 > x_1 > \cdots > x_{n-1} > x_n = b$. En conservant les mêmes définitions de $S^\Delta(f)$ et $S_\Delta(f)$, le seul changement par rapport aux subdivisions croissantes de $[b,a]$ est que les $(x_k - x_{k-1})$ sont négatifs ce qui implique une inversion des inégalités entre $S^\Delta(f)$ et $S_\Delta(f)$, on a maintenant $S^\Delta(f) \leqslant S_\Delta(f)$. Associons à chaque subdivision décroissante Δ de $[b,a]$, la subdivision *retournée* $\Delta' = \{x'_0, x'_1, \ldots, x'_n\}$ définie par $x'_0 = x_n$, $x'_1 = x_{n-1}, \ldots, x'_n = x_0$. Alors Δ' est une subdivision croissante de $[b,a]$ et $S^{\Delta'}(f) = -S^\Delta(f)$, $S_{\Delta'}(f) = -S_\Delta(f)$. On en déduit immédiatement que

$$I_*(f,a,b) := \sup_\Delta S^\Delta(f) = -\inf_{\Delta'} S^{\Delta'}(f) =: -I^*(f,b,a), \tag{3.4}$$

$$I^*(f,a,b) := \inf_\Delta S_\Delta(f) = -\sup_{\Delta'} S_{\Delta'}(f) =: -I_*(f,b,a), \tag{3.5}$$

les infima et suprema indexés par Δ s'entendant pour toute subdivision décroissante de a à b et ceux indexés par Δ' pour toute subdivision croissante de $[b,a]$. On définit alors l'intégrabilité de f de a à b par la condition $I_*(f,a,b) = I^*(f,a,b)$, dont on voit par (3.4) et (3.5) qu'elle équivaut à $I^*(f,b,a) = I_*(f,b,a)$, c'est-à-dire à l'intégrabilité de f sur $[b,a]$. En définissant enfin $\int_a^b f(x)\,\mathrm{d}x$ comme la valeur commune de $I_*(f,a,b)$ et $I^*(f,a,b)$, on obtient $\int_a^b f(x)\,\mathrm{d}x = -\int_b^a f(x)\,\mathrm{d}x$, cette dernière intégrale relevant de la définition 3.2. Tout ceci légitime la définition formelle suivante.

Définition 3.3. *Si $-\infty < b < a < +\infty$, on dit que f est Riemann intégrable de a à b si elle est Riemann intégrable sur $[b,a]$ et on pose dans ce cas :*

$$\int_a^b f(x)\,\mathrm{d}x := -\int_b^a f(x)\,\mathrm{d}x. \tag{3.6}$$

1. Rappelons que pour tous réels a et b, $[a,b]$ est défini comme l'ensemble des x réels tels que $a \leqslant x \leqslant b$. Ainsi pour $b < a$, $[a,b]$ est l'ensemble vide. C'est pour cela que l'on subdivise ici $[b,a]$ et non $[a,b]$. De même quand $b < a$ on parle d'intégrale de f de a à b mais pas d'intégrale sur $[a,b]$.

Remarque 3.4 (variable d'intégration)**.** Dans l'écriture $\int_b^a f(x)\,\mathrm{d}x$, la « variable d'intégration » x est « muette », on peut la remplacer par n'importe quelle autre lettre (sauf *ici* a, b ou f). Cette variable joue le même rôle que l'indice i de sommation dans $\sum_{i=1}^n u_i$ qui est lui aussi muet.

Remarque 3.5 (intégrale de Riemann et aire)**.** Soit f une fonction positive et Riemann intégrable sur $[a,b]$. On interprète classiquement $\int_a^b f(x)\,\mathrm{d}x$ comme *l'aire de l'hypographe* de f entre a et b, c'est-à-dire la région du plan délimitée par l'axe des abscisses, les droites verticales d'équation $x = a$ ou $x = b$ et le graphe[2] de f, la courbe d'équation $y = f(x)$, $x \in [a,b]$. Voici une justification informelle de cette affirmation, dont on pourra se contenter en première lecture. Reprenons la fonction f des figures 3.1 et 3.2. L'hypographe H de f est représenté figure 3.5.

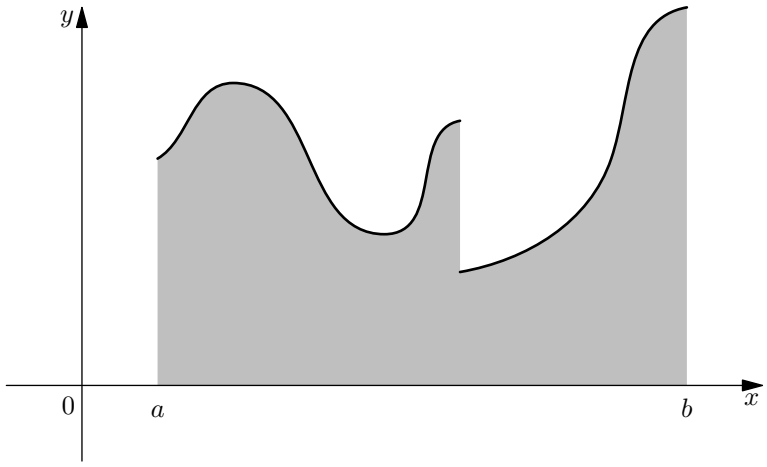


FIGURE 3.5 – Hypographe H de f entre a et b

On peut se convaincre « visuellement », cf. figure 3.1, que pour toute somme de Darboux inférieure l'aire des rectangles coloriés égale à $S_\Delta(f)$ est inférieure à l'aire de l'hypographe de f. De même cf. figure 3.2, pour toute somme de Darboux supérieure, l'aire des rectangles coloriés égale à $S^\Delta(f)$ est supérieure à l'aire de l'hypographe. L'aire de H est donc un majorant de toute S_Δ et un minorant de toute S^Δ. D'où

$$I_*(f) = \sup_\Delta S_\Delta(f) \leqslant \mathrm{aire}(H) \leqslant \inf_\Delta S_\Delta(f) = I^*(f).$$

La fonction f étant Riemman-intégrable, $I_*(f) = I^*(f) = \int_a^b f(x)\,\mathrm{d}x$, d'où $\mathrm{aire}(H) = \int_a^b f(x)\,\mathrm{d}x$.

Pour les lecteurs exigeants que la remarque 3.5 laisserait insatisfaits, nous proposons et démontrons ci-dessous un énoncé plus précis. Pour cela, il convient d'abord de

2. D'où le nom « hypographe », littéralement ce qui est « sous le graphe ».

s'interroger sur la définition mathématique de l'aire de H. Dans le cadre de cet ouvrage, nous admettons l'existence de la mesure de Lebesgue λ_2 sur $\mathbb{R}^2$, définie comme l'unique mesure μ sur la tribu borélienne de $\mathbb{R}^2$ vérifiant $\mu(]x_1,x_2]\times]y_1,y_2]) = (x_2 - x_1)(y_2 - y_1)$ pour tout pavé semi-ouvert $]x_1,x_2]\times]y_1,y_2]$ de $\mathbb{R}^2$, voir l'exemple 5.13 p. 160. C'est cette mesure de Lebesgue qui donne un sens mathématique précis à la notion d'aire. On se propose donc de montrer que $\lambda_2(H) = \int_a^b f(x)\,\mathrm{d}x$. Pour que cela ait un sens, encore faut-il que H soit un borélien de $\mathbb{R}^2$. Une condition suffisante pour que H soit un borélien de $\mathbb{R}^2$ est que la fonction f soit borélienne, c'est-à-dire mesurable (voir def. 6.1, p. 205) pour les tribus boréliennes de $[a,b]$ et de $\mathbb{R}$. La preuve de cette affirmation sort du programme de cet ouvrage. Signalons simplement que tous les exemples de fonctions Riemman intégrables donnés dans la suite de ce document — fonctions monotones, continues, réglées — sont boréliennes. Les seules propriétés de λ_2 utilisées dans ce qui suit sont la croissance et l'additivité finie — propriétés vérifiées par toute mesure, voir p. 155 — et le fait que les frontières des pavés sont de λ_2-mesure nulle [3], voir la proposition 5.14 v).

Proposition 3.6. *Soit $f : [a,b] \to \mathbb{R}$ une fonction positive et Riemann intégrable sur $[a,b]$. On suppose de plus que son hypographe entre a et b*

$$H := \{(x,y) \in \mathbb{R}^2;\ a \leqslant x \leqslant b \text{ et } 0 \leqslant y \leqslant f(x)\} \tag{3.7}$$

est un borélien de $\mathbb{R}^2$. Alors

$$\lambda_2(H) = \int_a^b f(x)\,\mathrm{d}x, \tag{3.8}$$

autrement dit, l'intégrale de f entre a et b est l'aire de son hypographe entre a et b.

Preuve. Soit $\Delta = \{x_0 = a < x_1 < \cdots < x_n = b\}$ une subdivision quelconque de $[a,b]$. Notons pour $k \in [\![1,n]\!]$,

$$m_k := \inf_{[x_{k-1},x_k]} f, \quad M_k := \sup_{[x_{k-1},x_k]} f.$$

Définissons les « rectangles » $R_{\Delta,k}$ par

$$R_{\Delta,1} := [a,x_1] \times [0,m_1], \quad R_{\Delta,k} :=]x_{k-1},x_k] \times [0,m_k] \quad \text{pour } k \in [\![2,n]\!],$$

et notons $R^{\Delta,k}$ les rectangles obtenus en remplaçant m_k par M_k, $k \in [\![1,n]\!]$. Posons enfin

$$R_\Delta := \bigcup_{k=1}^n R_{\Delta,k}, \quad R^\Delta := \bigcup_{k=1}^n R^{\Delta,k}.$$

Commençons par justifier la double inclusion :

$$\forall \Delta, \qquad R_\Delta \subset H \subset R^\Delta. \tag{3.9}$$

3. En réalité on a seulement besoin de savoir que si J est un segment vertical $\{a\} \times [y_1,y_2]$, $\lambda_2(J) = 0$ et de même pour un segment horizontal. Ceci se démontre facilement en exercice en utilisant la croissance de λ.

Soit (x', y') un élément quelconque de la *réunion* R_Δ. Il appartient donc à un $R_{\Delta,k'}$ d'où $x_{k'-1} \leqslant x' \leqslant x_{k'}$ et $0 \leqslant y' \leqslant m_{k'} \leqslant f(x')$ car $m_{k'}$ est l'infimum de f sur $[x_{k'-1}, x_{k'}]$. Le couple (x', y') vérifie ainsi les inégalités $a \leqslant x_{k'-1} \leqslant x \leqslant x_{k'} \leqslant b$ et $0 \leqslant y' \leqslant f(x)$, donc appartient à H. Ceci justifie la première inclusion dans (3.9). Soit maintenant (x'', y'') un élément quelconque de H, donc vérifiant $a \leqslant x'' \leqslant b$ et $0 \leqslant y'' \leqslant f(x'')$. La subdivision Δ induit la partition de $[a, b]$ en les intervalles $J_1 := [a, x_1]$, $J_k :=]x_{k-1}, x_k]$, $k = 2, \ldots, n$. Il existe donc un unique indice k'' entre 1 et n tel que $J_{k''}$ contienne x''. On a alors $0 \leqslant y'' \leqslant f(x'') \leqslant M_{k''}$ car $M_{k''}$ est le supremum de f sur $[x_{k''-1}, x_{k''}]$. Ainsi (x'', y'') appartient à $R^{\Delta,k''}$, donc aussi à R^Δ, ce qui justifie la deuxième inclusion dans (3.9).

Comme R_Δ, H et R^Δ sont des boréliens de $\mathbb{R}^2$, on déduit de (3.9) par croissance de λ_2 que

$$\forall \Delta, \qquad \lambda_2(R_\Delta) \leqslant \lambda_2(H) \leqslant \lambda_2(R^\Delta). \tag{3.10}$$

Calculons maintenant $\lambda_2(R_\Delta)$. Les $R_{\Delta,k}$ étant deux à deux disjoints, on a par additivité finie de λ_2 :

$$\lambda_2(R_\Delta) = \sum_{k=1}^{n} \lambda_2(R_{\Delta,k}). \tag{3.11}$$

Par la proposition 5.14 v) ou par la note 3 p. 74, $\lambda_2(R_{\Delta,k}) = (x_k - x_{k-1})m_k$ pour tout $k \in [\![1, n]\!]$, d'où en reportant dans (3.11), $\lambda_2(R_\Delta) = S_\Delta(f)$. De même il est clair que $\lambda_2(R^\Delta) = S^\Delta(f)$. On déduit alors de (3.10) que

$$\forall \Delta, \qquad S_\Delta(f) \leqslant \lambda_2(H) \leqslant S^\Delta(f). \tag{3.12}$$

La première inégalité dans (3.12) nous dit que le réel $\lambda_2(H)$ qui ne dépend pas de Δ majore toutes les sommes de Darboux inférieures $S_\Delta(f)$. Il majore donc aussi leur supremum $I_*(f)$. Par la deuxième inégalité, $\lambda_2(H)$ minore toutes les $S^\Delta(f)$, donc minore aussi leur infimum $I^*(f)$. Nous obtenons ainsi l'encadrement

$$I_*(f) \leqslant \lambda_2(H) \leqslant I^*(f).$$

Comme f est Riemann intégrable, $I_*(f) = I^*(f)$, donc $\lambda_2(H) = I_*(f) = I^*(f) = \int_a^b f(x)\,\mathrm{d}x$. □

Remarque 3.7. La mesure λ_2 étant invariante par la symétrie $(x, y) \mapsto (x, -y)$ cf. prop. 5.14 ii), on obtient immédiatement une version de la proposition 3.6 pour une fonction g négative sur $[a, b]$ en remplaçant H par

$$H' := \{(x, y) \in \mathbb{R}^2 \ ; \ a \leqslant x \leqslant b \text{ et } g(x) \leqslant y \leqslant 0\}.$$

En effet en posant $f = -g$, il vient

$$\lambda_2(H') = \lambda_2(H) = \int_a^b f(x)\,\mathrm{d}x = \int_a^b |g(x)|\,\mathrm{d}x. \tag{3.13}$$

3.2 Riemann intégrabilité

Dans cette section nous examinons la Riemann intégrabilité de certaines familles de fonctions. Les deux plus importantes en pratique sont celle des fonctions monotones et celle des fonctions continues. On généralise la Riemann intégrabilité des fonctions continues au cas des fonctions bornées continues sur $[a,b]$ sauf en un nombre fini de points, comme celle de la figure 3.1. Enfin nous établissons que toute limite uniforme sur $[a,b]$ d'une suite de fonctions Riemann intégrables sur $[a,b]$ est encore Riemann intégrable. Ceci nous donne notamment la Riemann intégrabilité de toutes les fonction *réglées*, c'est-à-dire limites uniformes de fonctions en escalier. Les fonctions en escalier sont les plus simples de toutes les fonctions Riemann intégrables et c'est par elles que nous commençons cette étude.

Définition 3.8 (fonction en escalier)**.** *Une application $f : [a,b] \to \mathbb{R}$ est appelée* fonction en escalier *sur $[a,b]$, s'il existe une subdivision $\Delta_0 = \{t_0 = a < t_1 < \cdots < t_j = b\}$ telle que f soit constante sur chaque intervalle* ouvert $]t_{i-1}, t_i[$, $i = 1, \ldots, j$.

Il est clair que Δ_0 n'est pas unique, en particulier pour tout raffinement Δ_0' de Δ_0, f est constante sur chacun des intervalles ouverts ayant pour extrémités deux points consécutifs de Δ_0'. Il y a donc une infinité de subdivisions Δ telles que f en escalier soit constante sur chacun des intervalles ouverts de Δ (c'est-à-dire les intervalles ayant pour extrémités deux points consécutifs de Δ). Nous appelerons *subdivision* associée à f en escalier, toute subdivision Δ telle que f soit constante sur chacun des intervalles ouverts de Δ. La moins fine des subdivisions associées à f en escalier est constituée des points a et b et des points de discontinuité de f dans $]a,b[$.

Proposition 3.9 (intégrabilité d'une fonction en escalier)**.** *Soit f une fonction en escalier sur $[a,b]$ et $\Delta_0 = \{t_0 = a < t_1 < \cdots < t_j = b\}$ une subdivision associée à f, la valeur constante de f sur $]t_{i-1}, t_i[$ étant notée c_i. Alors f est Riemann intégrable sur $[a,b]$ et on a*

$$\int_a^b f(x)\,\mathrm{d}x = \sum_{i=1}^{j} (t_i - t_{i-1})c_i. \tag{3.14}$$

Remarquons que comme $\int_a^b f(x)\,\mathrm{d}x$ ne dépend, lorsqu'elle existe, que de f, (3.14) implique que si $\Delta_1 = \{s_0 = a < s_1 < \cdots < s_l = b\}$ est une autre subdivision associée à f et en notant d_k la valeur constante de f sur $]s_{k-1}, s_k[$, $k = 1, \ldots, l$, on a

$$\sum_{i=1}^{j} (t_i - t_{i-1})c_i = \sum_{k=1}^{l} (s_k - s_{k-1})d_k.$$

Preuve. D'abord, f est bornée puisque $f([a,b]) = \{c_1, \ldots, c_j\} \cup \{f(t_0), \ldots, f(t_j)\}$ qui est fini (de cardinal *au plus* $2j+1$) donc borné dans $\mathbb{R}$. Pour chaque δ vérifiant

$$0 < \delta < \frac{1}{2} \min_{1 \leqslant i \leqslant j} (t_i - t_{i-1}), \tag{3.15}$$

notons Δ_δ la subdivision construite en adjoignant à Δ_0 les points $t_0+\delta, t_1-\delta, t_1+\delta, t_2-\delta, t_2+\delta, \ldots, t_{j-1}+\delta, t_j-\delta$. Notons en outre

$$m := \inf_{x\in[a,b]} f(x), \quad M := \sup_{x\in[a,b]} f(x),$$

$$m_i' := \inf_{|t_i-x|\leqslant\delta} f(x) = \min(c_i, c_{i+1}, f(t_i)), \quad M_i' := \sup_{|t_i-x|\leqslant\delta} f(x) = \max(c_i, c_{i+1}, f(t_i)),$$

avec l'adaptation évidente pour $i=j$. Bien sûr $M_i' \leqslant M$ et $m_i' \geqslant m$ pour tout i. Avec ces notations,

$$\begin{aligned} S^{\Delta_\delta}(f) &= \sum_{i=1}^{j}(t_i - t_{i-1} - 2\delta)c_i + \delta M_0' + \delta M_j' + 2\delta\sum_{i=1}^{j-1} M_i' \\ &\leqslant \sum_{i=1}^{j}(t_i - t_{i-1})c_i + 2j\delta(M-m). \end{aligned} \tag{3.16}$$

De même avec les m_i' à la place de M_i' on obtient

$$S_{\Delta_\delta}(f) \geqslant \sum_{i=1}^{j}(t_i - t_{i-1})c_i - 2j\delta(M-m). \tag{3.17}$$

Soit $\varepsilon>0$ quelconque, en choisissant $\delta=\delta(\varepsilon)$ vérifiant (3.15) et $2j\delta(M-m)<\varepsilon$, on dispose ainsi par (3.16) et (3.17) d'une subdivision $\Delta_{\delta(\varepsilon)}$ telle que

$$\sum_{i=1}^{j}(t_i - t_{i-1})c_i - \varepsilon < S_{\Delta_{\delta(\varepsilon)}} \leqslant S^{\Delta_{\delta(\varepsilon)}} < \sum_{i=1}^{j}(t_i - t_{i-1})c_i + \varepsilon.$$

On en déduit pour tout $\varepsilon>0$, l'encadrement :

$$\sum_{i=1}^{j}(t_i - t_{i-1})c_i - \varepsilon < I_*(f) \leqslant I^*(f) < \sum_{i=1}^{j}(t_i - t_{i-1})c_i + \varepsilon,$$

puis en faisant tendre ε vers 0 que

$$I_*(f) = I^*(f) = \sum_{i=1}^{j}(t_i - t_{i-1})c_i,$$

ce qui établit l'intégrabilité de f et (3.14). □

Proposition 3.10 (intégrabilité d'une fonction monotone)**.** *Si $f : [a,b] \to \mathbb{R}$ est monotone sur $[a,b]$, elle est Riemann intégrable sur $[a,b]$.*

Preuve. Supposons pour fixer les idées que f est décroissante, l'adaptation de ce qui suit au cas f croissante étant immédiate. Alors f est bornée puisque pour tout $x \in [a,b]$, $f(b) \leqslant f(x) \leqslant f(a)$. Pour $\Delta = \{x_0 = a < x_1 < \cdots < x_n = b\}$ subdivision quelconque de $[a,b]$, notons m_k et M_k les bornes inférieure et supérieure de f sur $[x_{k-1}, x_k]$ et remarquons que par décroissance de f, $m_k = f(x_k)$ et $M_k = f(x_{k-1})$. On a alors

$$\begin{aligned} 0 \leqslant S^\Delta(f) - S_\Delta(f) &= \sum_{k=1}^{n} (x_k - x_{k-1})(M_k - m_k) \\ &= \sum_{k=1}^{n} (x_k - x_{k-1})\big(f(x_{k-1}) - f(x_k)\big) \\ &\leqslant \max_{1\leqslant k\leqslant n} (x_k - x_{k-1}) \sum_{k=1}^{n} \big(f(x_{k-1}) - f(x_k)\big) \\ &= \max_{1\leqslant k\leqslant n} (x_k - x_{k-1})\big(f(a) - f(b)\big). \end{aligned}$$

Soit $\varepsilon > 0$ quelconque. En choisissant une subdivision Δ de « pas » au plus ε, c'est-à-dire telle que $\max_{1\leqslant k\leqslant n}(x_k - x_{k-1}) \leqslant \varepsilon$, on obtient :

$$0 \leqslant S^\Delta(f) - S_\Delta(f) \leqslant \varepsilon\big(f(a) - f(b)\big).$$

On en déduit que $I^*(f) - I_*(f) \leqslant \varepsilon\big(f(a) - f(b)\big)$ puis, ε étant quelconque, que $I^*(f) - I_*(f) = 0$. La fonction f est donc Riemann intégrable. □

Proposition 3.11 (intégrabilité d'une fonction continue)**.** *Si $f : [a,b] \to \mathbb{R}$ est continue, elle est Riemann intégrable sur $[a,b]$. De plus si F est une primitive de f sur $[a,b]$,*

$$\int_a^b f(x)\,\mathrm{d}x = F(b) - F(a). \tag{3.18}$$

Preuve. Sur le *compact* $[a,b]$, la fonction f est bornée et *uniformément* continue :

$$\forall \varepsilon > 0, \quad \exists \delta > 0, \quad \forall x, x' \in [a,b], \quad |x - x'| < \delta \Rightarrow |f(x) - f(x')| < \varepsilon.$$

Pour chaque $\varepsilon > 0$, il existe une subdivision $\Delta = \{x_0 = a < x_1 < \cdots < x_n = b\}$ (dépendant de δ) telle que $x_k - x_{k-1} < \delta$ pour tout $k \in [\![1, n]\!]$. Les bornes inférieure m_k et supérieure M_k de f sur le *compact* $[x_{k-1}, x_k]$ étant atteintes, $0 \leqslant M_k - m_k < \varepsilon$ pour tout k. Alors

$$0 \leqslant S^\Delta(f) - S_\Delta(f) = \sum_{k=1}^{n} (x_k - x_{k-1})(M_k - m_k) < \varepsilon \sum_{k=1}^{n} (x_k - x_{k-1}) = \varepsilon(b-a).$$

En raison de l'encadrement $S_\Delta(f) \leqslant I_*(f) \leqslant I^*(f) \leqslant S^\Delta(f)$, il est ainsi établi que

$$\forall \varepsilon > 0, \quad 0 \leqslant I^*(f) - I_*(f) \leqslant \varepsilon(b-a).$$

Comme $I^*(f) - I_*(f)$ ne dépend pas de ε, on en déduit que $I^*(f) - I_*(f) = 0$, d'où la Riemann intégrabilité de f sur $[a, b]$.

Rappelons que F est une primitive de f sur $[a, b]$ si elle est dérivable en tout point de $[a, b]$ (à droite en a et à gauche en b) et pour tout $x \in [a, b]$, $F'(x) = f(x)$. Soit $\Delta = \{x_0 = a < x_1 < \cdots < x_n = b\}$ une subdivision *quelconque* de $[a, b]$. Par le théorème des accroissements finis [4], il existe dans chaque $]x_{k-1}, x_k[$ un c_k tel que $F(x_k) - F(x_{k-1}) = (x_k - x_{k-1})F'(c_k) = (x_k - x_{k-1})f(c_k)$. En écrivant

$$F(b) - F(a) = \sum_{k=1}^{n} \big(F(x_k) - F(x_{k-1})\big) = \sum_{k=1}^{n} (x_k - x_{k-1}) f(c_k)$$

et en encadrant $f(c_k)$ entre les bornes inférieure et supérieure de f sur $[x_{k-1}, x_k]$, on en déduit

$$S_\Delta(f) \leqslant F(b) - F(a) \leqslant S^\Delta(f).$$

Cet encadrement est valide pour toute subdivision Δ et $F(b) - F(a)$ ne dépend pas de Δ. Par conséquent

$$I_*(f) \leqslant F(b) - F(a) \leqslant I^*(f)$$

et comme nous savons déjà que f est Riemann intégrable on en déduit $F(b) - F(a) = I_*(f) = I^*(f)$, ce qui établit (3.18). □

La preuve de (3.18) s'étend immédiatement au cas où F est dérivable sur $]a, b[$, continue à droite en a et à gauche en b et $F' = f$ sur $]a, b[$. D'autre part on peut utiliser l'intégrale de Riemann pour montrer que toute fonction continue sur $[a, b]$ admet des primitives sur $[a, b]$.

Proposition 3.12. *Toute fonction $f : [a, b] \to \mathbb{R}$, bornée sur $[a, b]$ et continue sur $[a, b]$, sauf en un nombre fini de points est Riemann intégrable sur $[a, b]$.*

Preuve. Nous nous contenterons de le montrer dans le cas où f présente un seul point de discontinuité $c \in]a, b[$, la généralisation ne coûtant qu'un alourdissement de notations. L'adaptation de ce qui suit au cas $c = a$ ou $c = b$ est aussi immédiate.

Fixons $\varepsilon > 0$ arbitraire et soit $\eta > 0$ assez petit pour que $[c - \eta, c + \eta] \subset]a, b[$ et dont le choix en fonction de ε sera précisé ultérieurement. Soit Δ une subdivision de $[a, b]$ ayant comme points *consécutifs* $c - \eta$ et $c + \eta$ ($x_{k_0} = c - \eta$ et $x_{k_0+1} = c + \eta$ pour un certain indice k_0). Cette subdivision Δ peut se construire comme réunion d'une subdivision quelconque Δ_1 de $[a, c - \eta]$ et d'une subdivision quelconque Δ_2 de $[c + \eta, b]$. Comme f est continue sur $[a, c - \eta]$ et $[c + \eta, b]$, elle est Riemann intégrable sur chacun de ces deux segments (prop. 3.11), ce qui nous autorise à choisir Δ_1 et Δ_2 telles que

$$0 \leqslant S^{\Delta_1}(f) - S_{\Delta_1}(f) \leqslant \frac{\varepsilon}{3}, \quad 0 \leqslant S^{\Delta_2}(f) - S_{\Delta_2}(f) \leqslant \frac{\varepsilon}{3}. \tag{3.19}$$

4. Appelé aussi *formule des accroissements finis* : si F est continue sur $[a, b]$ et dérivable sur $]a, b[$, il existe $c \in]a, b[$ tel que $F(b) - F(a) = F'(c)(b - a)$.

Notons m et M, m_η et M_η les bornes inférieure et supérieure de f sur respectivement $[a,b]$ et $[c-\eta, c+\eta]$. On a clairement $m \leqslant m_\eta \leqslant M_\eta \leqslant M$, d'où $M_\eta - m_\eta \leqslant M - m$, de sorte qu'en choisissant $\eta < \frac{\varepsilon}{6(M-m)}$, on ait

$$2\eta(M_\eta - m_\eta) < \frac{\varepsilon}{3}. \tag{3.20}$$

Avec le choix de Δ opéré ci-dessus,

$$\begin{aligned} S^{\Delta}(f) &= S^{\Delta_1}(f) + 2\eta M_\eta + S^{\Delta_2}(f) \\ S_{\Delta}(f) &= S_{\Delta_1}(f) + 2\eta m_\eta + S_{\Delta_2}(f), \end{aligned}$$

d'où, compte-tenu de (3.19) et (3.20),

$$0 \leqslant S^{\Delta}(f) - S_{\Delta}(f) \leqslant S^{\Delta_1}(f) - S_{\Delta_1}(f) + 2\eta(M_\eta - m_\eta) + S^{\Delta_2}(f) - S_{\Delta_2}(f) < \varepsilon.$$

On en déduit que $0 \leqslant I^*(f) - I_*(f) < \varepsilon$, puis par arbitrarité de ε que $I^*(f) = I_*(f)$, donc f est Riemann intégrable sur $[a,b]$. □

Nous allons maintenant établir que la Riemann intégrabilité se conserve par convergence uniforme sur $[a,b]$. Le lemme suivant nous sera utile.

Lemme 3.13. *Soit E une partie quelconque de $\mathbb{R}$. On suppose que chaque fonction f_n est définie et bornée sur E et que la suite $(f_n)_{n\geqslant 1}$ converge vers f uniformément sur E. Alors f est bornée sur E et*

$$\begin{aligned} m_n(E) := \inf_{x\in E} f_n(x) &\xrightarrow[n\to+\infty]{} \inf_{x\in E} f(x) =: m(E), \\ M_n(E) := \sup_{x\in E} f_n(x) &\xrightarrow[n\to+\infty]{} \sup_{x\in E} f(x) =: M(E). \end{aligned}$$

Plus précisément, si pour tout $n \geqslant n_\varepsilon$ et tout $x \in E$, $|f_n(x) - f(x)| < \varepsilon$, alors

$$\forall n \geqslant n_\varepsilon, \quad |m_n(E) - m(E)| \leqslant \varepsilon \quad \text{et} \quad |M_n(E) - M(E)| \leqslant \varepsilon. \tag{3.21}$$

Preuve. La convergence uniforme de (f_n) vers f sur E, s'écrit

$$\forall \varepsilon > 0,\ \exists n_\varepsilon \in \mathbb{N},\ \forall n \geqslant n_\varepsilon,\ \forall x \in E, \quad |f_n(x) - f(x)| < \varepsilon. \tag{3.22}$$

En réécrivant cette inégalité sous la forme $f_n(x) - \varepsilon < f(x) < f_n(x) + \varepsilon$ on en déduit :

$$\forall n \geqslant n_\varepsilon,\ \forall x \in E, \quad m_n(E) - \varepsilon < f(x) < M_n(E) + \varepsilon,$$

puis en prenant l'infimum et le supremum en $x \in E$ dans cette double inégalité [5] :

$$\forall n \geqslant n_\varepsilon, \qquad m_n(E) - \varepsilon \leqslant m(E) \quad \text{et} \quad M(E) \leqslant M_n(E) + \varepsilon. \tag{3.23}$$

5. Noter ici le passage des inégalités *strictes* aux inégalité larges.

En choisisssant un n particulier, par exemple $n = n_\varepsilon$, on en déduit que f est bornée sur E ($-\infty < m_n(E) - \varepsilon \leqslant m(E) \leqslant M(E) \leqslant M_n(E) + \varepsilon < +\infty$).

En réécrivant l'inégalité (3.22) sous la forme $f(x) - \varepsilon < f_n(x) < f(x) + \varepsilon$, on obtient de la même façon :

$$\forall n \geqslant n_\varepsilon, \qquad m(E) - \varepsilon \leqslant m_n(E) \quad \text{et} \quad M_n(E) \leqslant M(E) + \varepsilon. \tag{3.24}$$

En regroupant (3.23) et (3.24), on voit ainsi que pour tout $n \geqslant n_\varepsilon$,

$$m(E) - \varepsilon \leqslant m_n(E) \leqslant m(E) + \varepsilon \quad \text{et} \quad M(E) - \varepsilon \leqslant M_n(E) \leqslant M(E) + \varepsilon,$$

ce qui nous donne (3.21) et donc les convergences de $m_n(E)$ et $M_n(E)$ vers respectivement $m(E)$ et $M(E)$ puisque $\varepsilon > 0$ est ici arbitraire. □

Proposition 3.14. *Si f est limite uniforme sur $[a, b]$ d'une suite $(f_n)_{n\geqslant 1}$ de fonctions Riemann intégrables sur $[a, b]$, alors f est elle-même Riemann intégrable sur $[a, b]$.*

Preuve. D'abord, f est bornée sur $[a, b]$ comme limite uniforme d'une suite de fonctions bornées (lemme 3.13 avec $E = [a, b]$). On peut donc bien définir les sommes de Darboux $S_\Delta(f)$ et $S^\Delta(f)$ pour toute subdivision Δ de $[a, b]$.

Notons qu'il y a une difficulté supplémentaire dans cette démonstration par rapport aux preuves de la Riemann intégrabilité des fonctions monotones ou continues. Dans ces deux cas, f atteignait ses bornes inférieure et supérieure sur chaque intervalle de la subdivision, ce qui facilitait le traitement des sommes de Darboux. Ici, nous n'avons plus ce confort et c'est le lemme 3.13 qui arrange les choses.

Par convergence uniforme de f_n vers f sur $[a, b]$, pour tout $\varepsilon > 0$, il existe un entier n_ε tel que

$$\forall n \geqslant n_\varepsilon,\ \forall x \in [a, b], \quad |f_n(x) - f(x)| < \frac{\varepsilon}{b - a}. \tag{3.25}$$

Par le lemme 3.13, *avec le même n_ε* que pour (3.25), on a pour tout $n \geqslant n_\varepsilon$,

$$\forall E \subset [a, b], \quad |m_n(E) - m(E)| \leqslant \frac{\varepsilon}{b - a}, \quad |M_n(E) - M(E)| \leqslant \frac{\varepsilon}{b - a}. \tag{3.26}$$

Soit $\Delta = \{x_0 = a < x_1 < \cdots < x_j = b\}$ une subdivision quelconque de $[a, b]$. En appliquant (3.26) avec pour E chacun des intervalles $[x_{k-1}, x_k]$ de la subdivision, on vérifie immédiatement que :

$$\forall \Delta,\ \forall n \geqslant n_\varepsilon, \quad S_\Delta(f_n) - \varepsilon \leqslant S_\Delta(f) \leqslant S^\Delta(f) \leqslant S^\Delta(f_n) + \varepsilon. \tag{3.27}$$

La fonction f_{n_ε} étant par hypothèse Riemann intégrable sur $[a, b]$, il existe une subdivision Δ_ε telle que

$$S_{\Delta_\varepsilon}(f_{n_\varepsilon}) > S^{\Delta_\varepsilon}(f_{n_\varepsilon}) - \varepsilon. \tag{3.28}$$

En choisissant dans (3.27) $n = n_\varepsilon$ et $\Delta = \Delta_\varepsilon$ et en combinant l'encadrement ainsi obtenu avec l'inégalité (3.28), il vient :

$$S^{\Delta_\varepsilon}(f_{n_\varepsilon}) - 2\varepsilon < S_{\Delta_\varepsilon}(f) \leqslant S^{\Delta_\varepsilon}(f) \leqslant S^{\Delta_\varepsilon}(f_{n_\varepsilon}) + \varepsilon,$$

d'où l'on tire $0 \leqslant S^{\Delta_\varepsilon}(f) - S_{\Delta_\varepsilon}(f) < 3\varepsilon$, puis $0 \leqslant I^*(f) - I_*(f) < 3\varepsilon$. Par arbitrarité de ε, on en déduit $I_*(f) = I^*(f)$, ce qui établit la Riemann intégrabilité de f. □

Définition 3.15 (fonction réglée)**.** *On dit que f est réglée sur $[a,b]$ si elle est limite uniforme sur $[a,b]$ d'une suite de fonctions en escalier.*

Corollaire 3.16 (intégrabilité d'une fonction réglée)**.** *Toute fonction réglée $[a,b] \to \mathbb{R}$ est Riemann intégrable sur $[a,b]$.*

Preuve. C'est une conséquence immédiate des propositions 3.9 et 3.14. □

Pour finir cette section, nous donnons un exemple de fonction Riemann intégrable qui ne soit pas réglée (les fonctions monotones ou continues sont toutes réglées) et un exemple de fonction bornée et borélienne qui ne soit pas Riemann intégrable.

Exemple 3.17 (une fonction intégrable non réglée)**.** Soit $E := \{2^{-k};\ k \in \mathbb{N}^*\}$ et $f := \mathbf{1}_E$. La fonction f est bornée et Riemann intégrable sur $[0,1]$, mais pas réglée. Vérifions ces deux affirmations. Soit $\Delta_n := \{0,1\} \cup \{2^{-k} \pm 2^{-2n};\ 1 \leqslant k \leqslant n\}$. On voit immédiatement que pour tout $n \geqslant 2$:

$$0 = S_{\Delta_n}(f) \leqslant I_*(f) \leqslant I^*(f) \leqslant S^{\Delta_n}(f) = 2^{-n} - 2^{-2n} + 2n2^{-2n}.$$

En faisant tendre n vers $+\infty$, on en déduit que $I_*(f) = I^*(f) = 0$, ce qui prouve que f est Riemann intégrable sur $[0,1]$ et que $\int_0^1 f(x)\,\mathrm{d}x = 0$. Notons au passage qu'on a ainsi un exemple de fonction f positive d'intégrale de Riemann nulle sur $[0,1]$ sans que f soit identiquement nulle sur $[0,1]$. Cette situation ne pourrait pas se produire avec une f *continue* (exercice 3.3).

Supposons que $f = \mathbf{1}_E$ soit réglée. Ceci implique que pour tout $\varepsilon > 0$, il existe une fonction en escalier g telle que $|f(x) - g(x)| < \varepsilon$ pour tout $x \in [0,1]$. Prenons $\varepsilon = 1/3$, choisissons une telle g, notons $\Delta_0 = \{x_0 = 0 < x_1 < \cdots < x_j = 1\}$ une subdivision associée à g (cf. la définition 3.8) et c_1 la valeur constante de g sur $]0, x_1[$. Comme f prend au moins une fois la valeur 1 (en fait une infinité de fois) sur $]0, x_1[$, on a $|1 - c_1| < 1/3$ et de même f prenant au moins une fois la valeur 0 (en fait une infinité de fois) sur $]0, x_1[$, on a $|0 - c_1| < 1/3$. Ces deux inégalités sont incompatibles, donc f ne peut pas être réglée.

Exemple 3.18 (une fonction bornée non intégrable)**.** Soit $E := [0,1] \cap \mathbb{Q}$ et $f := \mathbf{1}_E$. La fonction f est bornée et borélienne (comme indicatrice d'un ensemble borélien de $\mathbb{R}$, cf. les définitions 5.5 et 6.2), mais n'est pas Riemann intégrable sur $[0,1]$. En effet en notant que dans tout intervalle ouvert non vide de $\mathbb{R}$ il y a au moins un rationnel et un irrationnel, on vérifie facilement que pour *toute* subdivision Δ de $[0,1]$,

$$S_\Delta(f) = 0 \quad \text{et} \quad S^\Delta(f) = 1.$$

On en déduit que $I_*(f) = 0$ et $I^*(f) = 1$, donc f n'est pas Riemann intégrable sur $[0,1]$.

3.3 Propriétés de l'intégrale de Riemann

Cette section regroupe les propriétés générales de l'intégrale de Riemann, à l'exception de celles relatives à l'interversion limite intégrale. Nous étudions d'abord les propriétés relatives aux fonctions à intégrer (les intégrandes), autrement dit la structure de l'ensemble $\mathcal{R}[a,b]$ des fonctions Riemann intégrables sur $[a,b]$. Nous verrons ensuite les propriétés concernant l'intervalle d'intégration.

3.3.1 Propriétés de l'ensemble $\mathcal{R}[a,b]$

Proposition 3.19 (additivité)**.** *Si f et g sont Riemann intégrables sur $[a,b]$, $f+g$ l'est aussi et*

$$\int_a^b (f+g)(x)\,\mathrm{d}x = \int_a^b f(x)\,\mathrm{d}x + \int_a^b g(x)\,\mathrm{d}x. \tag{3.29}$$

Preuve. Notons en préliminaire que si f et g sont bornées sur l'intervalle I, $f+g$ l'est aussi et on a

$$\inf_{x\in I} f(x) + \inf_{x\in I} g(x) \leqslant \inf_{x\in I}(f+g)(x), \quad \sup_{x\in I}(f+g)(x) \leqslant \sup_{x\in I} f(x) + \sup_{x\in I} g(x),$$

ces inégalités pouvant être strictes [6].

Fixons $\varepsilon > 0$ quelconque. La Riemann intégrabilité de f et de g nous fournit des subdivisions Δ_1 et Δ_2 telles que $\int_a^b f(x)\,\mathrm{d}x - \varepsilon < S_{\Delta_1}(f) \leqslant S^{\Delta_1}(f) < \int_a^b f(x)\,\mathrm{d}x + \varepsilon$ et $\int_a^b g(x)\,\mathrm{d}x - \varepsilon < S_{\Delta_2}(g) \leqslant S^{\Delta_2}(g) < \int_a^b g(x)\,\mathrm{d}x + \varepsilon$. Avec leur raffinement commun $\Delta := \Delta_1 \cup \Delta_2$, on a ainsi :

$$\int_a^b f(x)\,\mathrm{d}x - \varepsilon < S_\Delta(f) \leqslant S^\Delta(f) < \int_a^b f(x)\,\mathrm{d}x + \varepsilon, \tag{3.30}$$

$$\int_a^b g(x)\,\mathrm{d}x - \varepsilon < S_\Delta(g) \leqslant S^\Delta(g) < \int_a^b g(x)\,\mathrm{d}x + \varepsilon. \tag{3.31}$$

Notons x_i les points de Δ, m_i, m_i', m_i'', M_i, M_i', M_i'' les infima et suprema respectifs de $f+g$, f et g sur $[x_{i-1},x_i]$ pour $i \in [\![1,n]\!]$. Par la remarque faite en préliminaire,

$$\forall i \in [\![1,n]\!], \quad m_i' + m_i'' \leqslant m_i \leqslant M_i \leqslant M_i' + M_i''.$$

On en déduit que

$$S_\Delta(f) + S_\Delta(g) \leqslant S_\Delta(f+g) \leqslant S^\Delta(f+g) \leqslant S^\Delta(f) + S^\Delta(g). \tag{3.32}$$

En combinant (3.30), (3.31) et (3.32), on obtient :

$$\int_a^b f(x)\,\mathrm{d}x + \int_a^b g(x)\,\mathrm{d}x - 2\varepsilon < S_\Delta(f+g) \leqslant S^\Delta(f+g) < \int_a^b f(x)\,\mathrm{d}x + \int_a^b g(x)\,\mathrm{d}x + 2\varepsilon,$$

6. Par exemple $f : x \mapsto x$ et $g : x \mapsto 1-x$ sur $I = [0,1]$.

d'où

$$\int_a^b f(x)\,\mathrm{d}x + \int_a^b g(x)\,\mathrm{d}x - 2\varepsilon < I_*(f+g) \leqslant I^*(f+g) < \int_a^b f(x)\,\mathrm{d}x + \int_a^b g(x)\,\mathrm{d}x + 2\varepsilon.$$

Ce dernier encadrement étant vérifié pour tout $\varepsilon > 0$, on peut y faire tendre ε vers 0 pour obtenir finalement :

$$I_*(f+g) = I^*(f+g) = \int_a^b f(x)\,\mathrm{d}x + \int_a^b g(x)\,\mathrm{d}x,$$

ce qui établit l'intégrabilité de $f+g$ et (3.29). □

Proposition 3.20. *Si f est intégrable sur $[a,b]$ et $c \in \mathbb{R}$ est une constante, cf est intégrable sur $[a,b]$ et*

$$\int_a^b cf(x)\,\mathrm{d}x = c\int_a^b f(x)\,\mathrm{d}x. \tag{3.33}$$

Preuve. Le résultat est trivial si $c = 0$, puisqu'alors cf est la fonction identiquement nulle sur $[a,b]$, Riemann intégrable d'intégrale 0. Supposons $c \neq 0$. Si f est bornée sur l'ensemble E, cf est bornée sur E. On vérifie facilement que

$$\text{si } c > 0, \quad \inf_{x\in E}(cf)(x) = c\inf_{x\in E} f(x), \quad \sup_{x\in E}(cf)(x) = c\sup_{x\in E} f(x), \tag{3.34}$$

et que

$$\text{si } c < 0, \quad \inf_{x\in E}(cf)(x) = c\sup_{x\in E} f(x), \quad \sup_{x\in E}(cf)(x) = c\inf_{x\in E} f(x). \tag{3.35}$$

Pour $c > 0$, on déduit de (3.34) que pour toute subdivision Δ de $[a,b]$,

$$S_\Delta(cf) = cS_\Delta(f), \quad S^\Delta(cf) = cS^\Delta(f),$$

d'où $I_*(cf) = cI_*(f)$ et $I^*(cf) = cI^*(f)$. Ces deux égalités sont valables pour n'importe quelle fonction bornée f sur $[a,b]$. Ici f est de plus intégrable sur $[a,b]$, donc $I_*(f) = I^*(f) = \int_a^b f(x)\,\mathrm{d}x$, d'où $I_*(cf) = I^*(cf) = c\int_a^b f(x)\,\mathrm{d}x$, ce qui prouve l'intégrabilité de cf et établit (3.33). Pour $c < 0$, on déduit de (3.35) que pour toute subdivision Δ de $[a,b]$,

$$S_\Delta(cf) = cS^\Delta(f), \quad S^\Delta(cf) = cS_\Delta(f),$$

d'où $I_*(cf) = cI^*(f)$ et $I^*(cf) = cI_*(f)$. Par intégrabilité de f, on en déduit comme ci-dessus que $I_*(cf) = I^*(cf) = c\int_a^b f(x)\,\mathrm{d}x$, ce qui complète la preuve. □

On peut synthétiser les propositions 3.19 et 3.20 dans l'énoncé suivant.

Proposition 3.21 (linéarité)**.** *L'ensemble $\mathcal{R}[a,b]$ des applications $f : [a,b] \to \mathbb{R}$, Riemann intégrables sur $[a,b]$ est un $\mathbb{R}$-espace vectoriel et l'application $\Psi : \mathcal{R}[a,b] \to \mathbb{R}$, $f \mapsto \int_a^b f(x)\,\mathrm{d}x$ est une forme linéaire sur cet espace.*

Proposition 3.22 (croissance de l'intégrale)**.** *L'intégrale de Riemann sur* $[a,b]$ *possède les trois propriétés suivantes relativement à la relation d'ordre partiel* $\leqslant$ *définie sur* $\mathcal{R}[a,b]$ *par* $f \leqslant g$ *si pour tout* $x \in [a,b]$, $f(x) \leqslant g(x)$.

i) Positivité :

$$\text{si } f \in \mathcal{R}[a,b] \text{ et } f \geqslant 0 \text{ sur } [a,b], \quad \int_a^b f(x)\,\mathrm{d}x \geqslant 0. \tag{3.36}$$

ii) Croissance :

$$\text{si } f,g \in \mathcal{R}[a,b] \text{ et } f \leqslant g \text{ sur } [a,b], \quad \int_a^b f(x)\,\mathrm{d}x \leqslant \int_a^b g(x)\,\mathrm{d}x. \tag{3.37}$$

iii) Si $f \in \mathcal{R}[a,b]$, *l'application* $|f| : x \mapsto |f(x)|$ *est elle aussi Riemann intégrable sur* $[a,b]$ *et*

$$\Big|\int_a^b f(x)\,\mathrm{d}x\Big| \leqslant \int_a^b |f(x)|\,\mathrm{d}x. \tag{3.38}$$

Preuve. Rappelons que lorsqu'on parle d'intégrale sur $[a,b]$, on suppose implicitement $a \leqslant b$. Dans le cas $a > b$, on aurait $\int_a^b f(x)\,\mathrm{d}x \leqslant 0$ dans (3.36) et $\int_a^b g(x)\,\mathrm{d}x \leqslant \int_a^b f(x)\,\mathrm{d}x$ dans (3.37).

Pour prouver i), on remarque que si $f \geqslant 0$ sur $[a,b]$ et Δ est une subdivision (croissante) quelconque de $[a,b]$, alors $m_k \geqslant 0$ pour chaque intervalle $[x_{k-1},x_k]$ de Δ, donc $S_\Delta(f) \geqslant 0$ et par conséquent $I^*(f) \geqslant I_*(f) \geqslant 0$. Ce raisonnement est valable pour toute f positive bornée sur $[a,b]$. Comme ici f est de plus Riemann intégrable sur $[a,b]$, $I_*(f) = I^*(f) = \int_a^b f(x)\,\mathrm{d}x$ et (3.36) est vérifiée.

On vérifie ii) en notant que si f,g sont dans $\mathcal{R}[a,b]$, $h := g - f$ aussi (prop. 3.21) et $h \geqslant 0$. En utilisant i) et la proposition 3.21, on obtient

$$0 \leqslant \int_a^b h(x)\,\mathrm{d}x = \int_a^b g(x)\,\mathrm{d}x - \int_a^b f(x)\,\mathrm{d}x,$$

ce qui nous donne (3.37).

Admettons provisoirement que l'intégrabilité de f implique celle de $|f|$. En appliquant ii) avec l'encadrement $-|f| \leqslant f \leqslant |f|$, il vient [7] :

$$-\int_a^b |f(x)|\,\mathrm{d}x \leqslant \int_a^b f(x)\,\mathrm{d}x \leqslant \int_a^b |f(x)|\,\mathrm{d}x,$$

ce qui équivaut à (3.38).

Il reste à montrer que $|f|$ hérite de l'intégrabilité de f. Pour toute subdivision $\Delta = \{x_0 = a < x_1 < \cdots < x_n = b\}$, notons

$$m_k := \inf_{x\in[x_{k-1},x_k]} f(x), \quad m'_k := \inf_{x\in[x_{k-1},x_k]} |f(x)|,$$

7. En utilisant aussi $\int_a^b -|f(x)|\,\mathrm{d}x = -\int_a^b |f(x)|\,\mathrm{d}x$ par linéarité.

$$M_k := \sup_{x\in[x_{k-1},x_k]} f(x), \quad M'_k := \sup_{x\in[x_{k-1},x_k]} |f(x)|.$$

Par le lemme 3.23 ci-dessous, $0 \leqslant M'_k - m'_k \leqslant M_k - m_k$ pour tout $k \in [\![1,n]\!]$, d'où

$$0 \leqslant I^*(|f|) - I_*(|f|) \leqslant S^\Delta(|f|) - S_\Delta(|f|) \leqslant S^\Delta(f) - S_\Delta(f).$$

Comme f est Riemann intégrable sur $[a,b]$, on peut choisir pour tout ε une subdivision Δ telle que $S^\Delta(f) - S_\Delta(f) < \varepsilon$ et l'encadrement ci-dessus appliqué à cette subdivision nous donne $0 \leqslant I^*(|f|) - I_*(|f|) < \varepsilon$, d'où $I^*(|f|) = I_*(|f|)$ par arbitrarité de ε. □

Lemme 3.23. *Si f est bornée sur $E \subset \mathbb{R}$, en notant $m := \inf_E f$, $m' := \inf_E |f|$, $M := \sup_E f$, $M' := \sup_E |f|$, on a $0 \leqslant M' - m' \leqslant M - m$, l'inégalité pouvant être stricte.*

Preuve. Si $M' - m' = 0$, c'est trivial [8]. Supposons désormais que $M' - m' > 0$. Alors pour tout ε tel que $0 < \varepsilon < (M' - m')/2$, on peut trouver $x_1, x_2 \in E$, dépendants de ε, tels que :

$$|f(x_1)| < m' + \varepsilon < M' - \varepsilon < |f(x_2)|. \tag{3.39}$$

Par inégalité triangulaire, $|f(x_2)| - |f(x_1)| \leqslant |f(x_2) - f(x_1)|$. D'autre part $f(x_1)$ et $f(x_2)$ sont des réels du segment $[m,M]$, d'où $|f(x_2) - f(x_1)| \leqslant M - m$ et donc

$$|f(x_2)| - |f(x_1)| \leqslant M - m.$$

En combinant cette dernière inégalité avec (3.39), il vient :

$$M' - m' - 2\varepsilon < M - m.$$

Faisant tendre ε vers 0, on obtient l'inégalité *large* $M' - m' \leqslant M - m$. □

Remarque 3.24. La Riemann intégrabilité de $|f|$ n'implique pas celle de f. Voici un contre exemple avec $[a,b] = [0,1]$: $f(x) = 1$ si $x \in \mathbb{Q}$ et -1 si $x \notin \mathbb{Q}$. Alors, comme dans l'exemple 3.18, f n'est pas Riemann intégrable sur $[0,1]$, tandis que $|f|$ l'est (comme fonction constante).

Voici une conséquence immédiate de l'implication $f \in \mathcal{R}[a,b] \Rightarrow |f| \in \mathcal{R}[a,b]$ dans la proposition 3.22 iii).

Corollaire 3.25. *Si f est Riemann intégrable sur $[a,b]$, les fonctions $f^+ := \max(f,0)$ et $f^- := \max(-f,0) = -\min(f,0)$ le sont aussi. On a de plus*

$$\int_a^b f(x)\,\mathrm{d}x = \int_a^b f^+(x)\,\mathrm{d}x - \int_a^b f^-(x)\,\mathrm{d}x. \tag{3.40}$$

8. Cette situation peut se produire, par exemple $E = [-1,1]$ et $f(x) = \mathbf{1}_{[-1,0]}(x) - \mathbf{1}_{]0,1]}(x)$, on a dans ce cas $0 = M' - m' < M - m = 2$.

Preuve. On commence par remarquer que

$$f^+ = \frac{1}{2}(|f| + f), \quad f^- = \frac{1}{2}(|f| - f).$$

Par la proposition 3.22 iii) l'intégrabilité de f implique celle de $|f|$. L'ensemble $\mathcal{R}[a,b]$ étant un espace vectoriel (prop. 3.21), on en déduit la Riemann intégrabilité sur $[a,b]$ de f^+ et f^-. La linéarité de l'intégrale et l'égalité $f = f^+ - f^-$ nous donnent (3.40). □

Remarque 3.26 (semi norme sur $\mathcal{R}[a,b]$)**.** Grâce au point iii) de la proposition 3.22, on peut définir l'application

$$N : \mathcal{R}[a,b] \to \mathbb{R}_+, \quad f \mapsto N(f) := \int_a^b |f(x)|\,\mathrm{d}x.$$

Cette application est une *semi norme* sur $\mathcal{R}[a,b]$ car elle vérifie $N(cf) = |c|N(f)$ pour toute constante c et $N(f+g) \leqslant N(f) + N(g)$. Elle n'est pas une norme car on peut avoir $\int_a^b |f(x)|\,\mathrm{d}x = 0$ sans que f soit la fonction nulle sur $[a,b]$, voir l'exemple 3.17.

Proposition 3.27 (Intégrabilité d'un produit)**.** *Si f et g sont Riemann intégrables sur $[a,b]$, leur produit fg l'est aussi.*

Preuve. En écrivant $fg = (f^+ - f^-)(g^+ - g^-) = f^+g^+ - f^-g^+ - f^+g^- + f^-g^-$ et en utilisant le corollaire 3.25 et la linéarité de l'intégrale de Riemann, on voit qu'il suffit de traiter le cas où f et g sont toutes deux *positives* sur $[a,b]$.

Fixons $\varepsilon > 0$ quelconque. La Riemann intégrabilité de f et g nous fournissent des subdivisions Δ_1 et Δ_2 telles que $\int_a^b f(x)\,\mathrm{d}x - \varepsilon < S_{\Delta_1}(f) \leqslant S^{\Delta_1}(f) < \int_a^b f(x)\,\mathrm{d}x + \varepsilon$ et $\int_a^b g(x)\,\mathrm{d}x - \varepsilon < S_{\Delta_2}(g) \leqslant S^{\Delta_2}(g) < \int_a^b g(x)\,\mathrm{d}x + \varepsilon$. Avec leur raffinement commun $\Delta := \Delta_1 \cup \Delta_2$, on a ainsi :

$$0 \leqslant S^{\Delta}(f) - S_{\Delta}(f) \leqslant 2\varepsilon, \tag{3.41}$$
$$0 \leqslant S^{\Delta}(g) - S_{\Delta}(g) \leqslant 2\varepsilon. \tag{3.42}$$

Rappelons ici que f et g Riemann intégrables sur $[a,b]$ sont *ipso facto bornées* sur cet intervalle, donc fg est aussi bornée sur $[a,b]$. Notons x_i les points de Δ, m_i, m_i', m_i'', M_i, M_i', M_i'' les infima et suprema respectifs de fg, f et g sur $[x_{i-1}, x_i]$ pour $i \in [\![1,n]\!]$. Par positivité, pour tout $i \in [\![1,n]\!]$,

$$\forall x \in [x_{i-1}, x_i], \quad 0 \leqslant m_i' m_i'' \leqslant f(x)g(x) \leqslant M_i' M_i'',$$

d'où

$$0 \leqslant m_i' m_i'' \leqslant m_i \leqslant M_i \leqslant M_i' M_i''.$$

Notons c un majorant commun sur $[a,b]$ aux fonctions positives bornées f et g . Alors pour tout $i \in [\![1,n]\!]$,

$$\begin{aligned} 0 \leqslant M_i - m_i \leqslant M_i' M_i'' - m_i' m_i'' &= (M_i' - m_i')M_i'' + m_i'(M_i'' - m_i'') \\ &\leqslant c(M_i' - m_i') + c(M_i'' - m_i''). \end{aligned}$$

En reportant cette majoration dans le calcul de $S^\Delta(fg)-S_\Delta(fg)$ et en tenant compte de (3.41) et (3.42), on obtient

$$0 \leqslant S^\Delta(fg) - S_\Delta(fg) \leqslant 4c\varepsilon.$$

Comme ε était arbitraire, on en déduit la Riemann intégrabilité de fg. □

Contrairement à ce qui se passe pour la Riemann intégrabilité d'une somme $f+g$, il n'y a pas de formule permettant de calculer $\int_a^b f(x)g(x)\,\mathrm{d}x$ en fonction de $\int_a^b f(x)\,\mathrm{d}x$ et $\int_a^b g(x)\,\mathrm{d}x$. La formule « $\int_a^b f(x)g(x)\,\mathrm{d}x = \int_a^b f(x)\,\mathrm{d}x \int_a^b g(x)\,\mathrm{d}x$ » est *grossièrement fausse*. Voici un contre exemple élémentaire avec des fonctions en escalier. Prenons $a=0$, $b=2$, $f=\mathbf{1}_{[0,1]}$, $g=\mathbf{1}_{]1,2]}$. Alors fg est la fonction nulle sur $[0,2]$ et donc $\int_0^2 f(x)g(x)\,\mathrm{d}x = 0$, alors que $\int_0^2 f(x)\,\mathrm{d}x \int_0^2 g(x)\,\mathrm{d}x = 1$ parce que $\int_0^2 f(x)\,\mathrm{d}x$ et $\int_0^2 g(x)\,\mathrm{d}x$ valent chacune 1.

Proposition 3.28 (inégalité de Cauchy-Schwarz dans $\mathcal{R}[a,b]$)**.** *Si f et g sont Riemann intégrables sur $[a,b]$, on a l'inégalité*

$$\Big|\int_a^b f(x)g(x)\,\mathrm{d}x\Big| \leqslant \left\{\int_a^b f(x)^2\,\mathrm{d}x\right\}^{1/2} \left\{\int_a^b g(x)^2\,\mathrm{d}x\right\}^{1/2}. \tag{3.43}$$

Preuve. Soit t un réel quelconque. Par les propositions 3.21 et 3.27, les fonctions f^2, g^2, fg et $(tf+g)^2$ héritent de la Riemann intégrabilité de f et g. Posons

$$P(t) := \int_a^b \big(tf(x)+g(x)\big)^2\,\mathrm{d}x.$$

Il est clair que $P(t)$ est positif ou nul pour tout t réel. Or en développant le carré $(tf+g)^2$ et en utilisant la linéarité de l'intégrale, on obtient :

$$P(t) = \left\{\int_a^b f(x)^2\,\mathrm{d}x\right\} t^2 + 2\left\{\int_a^b f(x)g(x)\,\mathrm{d}x\right\} t + \int_a^b g(x)^2\,\mathrm{d}x.$$

On reconnaît là un trinôme du second degré At^2+Bt+C dont les coefficients A, B, C sont des intégrales. Ce trinôme ne peut avoir de signe constant, celui de $A = \int_a^b f(x)^2\,\mathrm{d}x \geqslant 0$, que si son discriminant $\Delta = B^2-4AC$ est négatif ou nul. Remplaçant A, B et C par leurs expressions sous forme d'intégrales, on en déduit (3.43). □

3.3.2 Propriétés relatives à l'intervalle d'intégration

L'intégrale de Riemann se laisse volontiers découper en morceaux. Voici les énoncés précis dont la vérification est laissée au lecteur.

Proposition 3.29 (additivité relative aux intervalles)**.** *Soit $f : [a, b] \to \mathbb{R}$ et $c \in]a, b[$. Pour que f soit Riemann intégrable sur $[a, b]$, il faut et il suffit qu'elle soit Riemann intégrable sur $[a, c]$ et sur $[c, b]$. Alors*

$$\int_a^b f(x)\,\mathrm{d}x = \int_a^c f(x)\,\mathrm{d}x + \int_c^b f(x)\,\mathrm{d}x. \tag{3.44}$$

En combinant la proposition 3.29 avec la définition 3.3, on obtient la formule classique suivante.

Proposition 3.30 (relation de Chasles)**.** *Pour tous réels a, b, c, on a*

$$\int_a^b f(x)\,\mathrm{d}x = \int_a^c f(x)\,\mathrm{d}x + \int_c^b f(x)\,\mathrm{d}x,$$

pourvu que f soit Riemann intégrable sur $[\min(a, b, c), \max(a, b, c)]$.

Une autre application de l'additivité relative aux intervalles est la généralisation de la formule (3.13) pour le calcul de l'aire du domaine H délimité par le graphe de f, l'axe des abscisses et les deux droites verticales d'équations $x = a$ et $x = b$, voir figure 3.6.

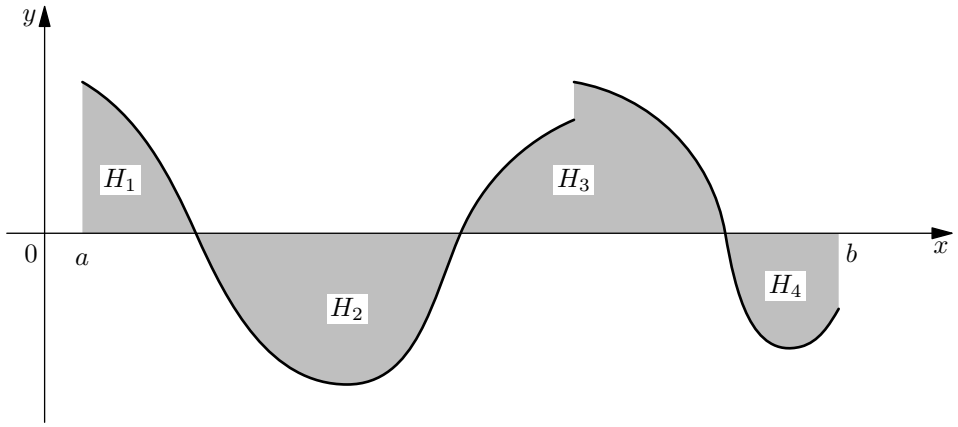


FIGURE 3.6 – Domaine H délimité par f entre a et b

Plus précisément, H est défini par

$$H := \{(x, y) \in \mathbb{R}^2;\ a \leqslant x \leqslant b \text{ et } -f^-(x) \leqslant y \leqslant f^+(x)\}, \tag{3.45}$$

en notant que $-f^-(x) = f(x)$ ou 0 selon que $f(x) < 0$ ou non et que $f^+(x) = f(x)$ ou 0 selon que $f(x) > 0$ ou non. En combinant la proposition 3.6, la remarque 3.7, l'additivité relative aux intervalles de l'intégrale de Riemann et l'additivité finie de λ_2, on obtient le résultat suivant pour les fonctions n'ayant qu'un nombre fini de changements de signe sur $[a, b]$.

Proposition 3.31. *Soit f une fonction borélienne sur $[a,b]$ et Riemann intégrable sur $[a,b]$. On suppose qu'il existe une subdivision $x_0 = a < x_1 < \cdots < x_n = b$ telle que le signe de f soit constant sur chacun des $[x_{i-1}, x_i]$, $1 \leqslant i \leqslant n$. Alors l'aire du domaine H défini par (3.45) est donnée par*

$$\lambda_2(H) = \int_a^b |f(x)|\,\mathrm{d}x. \tag{3.46}$$

Dans cet énoncé, « signe constant » s'entend au sens large : ou bien $f(x) \geqslant 0$ pour tout $x \in [x_{i-1}, x_i]$ ou bien $f(x) \leqslant 0$ pour tout $x \in [x_{i-1}, x_i]$. L'hypothèse f borélienne assure que sa restriction à chacun des $[x_{i-1}, x_i]$ est encore borélienne (pour les tribus adéquates) et donc que chaque $H_i := \{(x,y) \in \mathbb{R}^2;\ x_{i-1} \leqslant x \leqslant x_i \text{ et } -f^-(x) \leqslant y \leqslant f^+(x)\}$ est un borélien de $\mathbb{R}^2$ (admis). Bien entendu en écrivant cet énoncé, on a en tête le cas où le signe de f change à la traversée de chaque x_i, $0 < i < n$, mais la formule (3.46) reste évidemment vraie sans cette hypothèse. Si f a le même signe sur deux intervalles consécutifs, on peut trouver une subdivision « plus économique » en les fusionnant.

Nous pouvons maintenant donner une interprétation géométrique de l'intégrale de Riemann, au moins pour les fonctions f vérifiant les hypothèses de la proposition 3.31. Pour cela on appelle « aire algébrique », la somme des $\lambda_2(H_i)$, *chacun étant compté avec le signe de f* sur l'intervalle correspondant. L'aire algébrique du domaine H représenté à la figure 3.6 vaut ainsi $\lambda_2(H_1) - \lambda_2(H_2) + \lambda_2(H_3) - \lambda_2(H_4)$. Plus formellement, posons

$$s_i := \begin{cases} +1 & \text{si } f(t) > 0 \text{ pour au moins un } t \in]x_{i-1}, x_i[, \\ -1 & \text{si } f(t) < 0 \text{ pour au moins un } t \in]x_{i-1}, x_i[, \\ 0 & \text{si } f(t) = 0 \text{ pour tout } t \in]x_{i-1}, x_i[. \end{cases}$$

L'argumentation esquissée pour la proposition 3.31 nous donne

$$\text{aire algébrique}(H) := \sum_{i=1}^{n} s_i \lambda_2(H_i) = \int_a^b f(x)\,\mathrm{d}x. \tag{3.47}$$

Nous regroupons dans le théorème suivant les propriétés de « l'intégrale indéfinie », c'est à dire de la *fonction* $x \mapsto \int_a^x f(t)\,\mathrm{d}t$.

Théorème 3.32. *Soit f Riemann intégrable sur $[a,b]$. Alors elle est aussi Riemann intégrable sur $[a,x]$ pour tout $x \in [a,b]$, ce qui permet de définir l'application*

$$F : [a,b] \to \mathbb{R}, \quad x \mapsto F(x) := \int_a^x f(t)\,\mathrm{d}t.$$

i) F est continue sur $[a,b]$ et même lipschitzienne.

ii) Si f a une limite ℓ au point c de $[a,b]$ (resp. une limite à droite, resp. à gauche), F est dérivable au point c (resp. à gauche, resp. à droite) et $F'(c) = \ell$.

iii) Si f est continue sur $[a,b]$, F est dérivable sur $[a,b]$ et a pour fonction dérivée f. C'est l'unique primitive de f sur $[a,b]$ qui s'annule au point a.

Preuve de i). Notons $C := \sup_{t\in[a,b]} |f(t)|$. En utilisant la relation de Chasles et la proposition 3.22, on a pour tous $a \leqslant x \leqslant y \leqslant b$,

$$|F(y) - F(x)| = \Big|\int_x^y f(t)\,\mathrm{d}t\Big| \leqslant \int_x^y |f(t)|\,\mathrm{d}t \leqslant \int_x^y C\,\mathrm{d}t = C|y-x|.$$

Ceci montre que F est C-lipschitzienne sur $[a,b]$, donc *a fortiori* continue. □

Preuve de ii). Commençons par noter que pour tout $x \neq c$ dans $[a,b]$,

$$\frac{F(x)-F(c)}{x-c} = \frac{1}{x-c}\int_c^x f(t)\,\mathrm{d}t, \quad \text{et} \quad \ell = \frac{1}{x-c}\int_c^x \ell\,\mathrm{d}t. \tag{3.48}$$

Comme f a pour limite ℓ au point c, on a :

$$\forall \varepsilon > 0,\ \exists \delta > 0,\ 0 \leqslant |t-c| \leqslant |x-c| < \delta \Rightarrow |f(t)-\ell| < \varepsilon. \tag{3.49}$$

En combinant (3.48) et (3.49), on voit que pour tout $x \in [a,b]$ vérifiant $|x-c| < \delta$,

$$\Big|\frac{F(x)-F(c)}{x-c} - \ell\Big| = \Big|\frac{1}{x-c}\int_c^x (f(t)-\ell)\,\mathrm{d}t\Big| \leqslant \frac{1}{|x-c|}\int_{\min(c,x)}^{\max(c,x)} |f(t)-\ell|\,\mathrm{d}t \leqslant \varepsilon,$$

en notant que si $t \in [c,x]$ ou $t \in [x,c]$, $|t-c| \leqslant |x-c| < \delta$. Ceci montre que F est dérivable au point c, de nombre dérivé $F'(c) = \ell$. L'adaptation au cas d'une limite à droite ou à gauche (avec dérivée à droite ou à gauche) est immédiate. □

Preuve de iii). Si f est continue sur $[a,b]$, elle a pour limite $f(c)$ en tout point c de $[a,b]$ et donc d'après ii), F est dérivable sur $[a,b]$ et $F'(c) = f(c)$. Cette dernière égalité ayant lieu maintenant pour tout $c \in [a,b]$, on a $F' = f$, autrement dit F est *une primitive* de f sur $[a,b]$. On sait que toutes les primitives de f sur *l'intervalle* $[a,b]$ diffèrent entre elles d'une constante [9]. Il y en a donc une seule qui s'annule au point a, c'est F. □

Proposition 3.33 (changement de variable)**.**

i) Translation. Soit $c \in \mathbb{R}$. Pour toute f Riemann intégrable sur $[a+c, b+c]$, l'application $g : [a,b] \to \mathbb{R}$, $x \mapsto f(x+c)$ est Riemann intégrable sur $[a,b]$ et

$$\int_a^b f(x+c)\,\mathrm{d}x = \int_{a+c}^{b+c} f(y)\,\mathrm{d}y. \tag{3.50}$$

9. C'est une conséquence de la formule des accroissements finis (cf. p.79) : si une fonction continue sur $[a,b]$ a une dérivée nulle sur $]a,b[$, elle est constante sur $[a,b]$ et on applique ceci à la différence de deux primitives quelconques de f sur $[a,b]$.

ii) Changement d'échelle. Soit $c \in \mathbb{R}^*$. Pour toute f Riemann intégrable sur l'intervalle fermé d'extrémités[10] *ac et bc, l'application $h : [a,b] \to \mathbb{R}$, $x \mapsto f(cx)$ est Riemann intégrable sur $[a,b]$ et*

$$\int_a^b f(cx)\,\mathrm{d}x = \frac{1}{c}\int_{ac}^{bc} f(y)\,\mathrm{d}y. \tag{3.51}$$

iii) Classique. Soit $\varphi : [a,b] \to \mathbb{R}$, une fonction ayant une dérivée continue sur $[a,b]$ (autrement dit $\varphi \in C^1[a,b]$). Pour toute fonction f continue *sur l'intervalle fermé borné $\varphi([a,b])$, on a*

$$\int_a^b f\big(\varphi(x)\big)\varphi'(x)\,\mathrm{d}x = \int_{\varphi(a)}^{\varphi(b)} f(y)\,\mathrm{d}y. \tag{3.52}$$

Bien sûr i) et ii) sont contenus dans iii) si f est *continue*, mais l'intérêt de ces deux énoncés séparés est qu'ils sont valables avec *n'importe quelle* fonction f Riemann intégrable.

Preuve de i). À chaque subdivision $\Delta = \{x_0 = a < x_1 < \cdots < x_n = b\}$ de $[a,b]$, associons la subdivision translatée $\Delta' = \{y_0 = a + c < \cdots < y_k = x_k + c < \cdots < y_n = b + c\}$. Comme f est bornée sur $[a+c, b+c]$, g est bornée sur $[a,b]$ avec mêmes bornes. De plus en notant m_k, m'_k les infima respectifs de f sur $[x_{k-1}, x_k]$ et de g sur $[y_{k-1}, y_k]$, et en définissant de même M_k et M'_k pour les suprema, $m_k = m'_k$ et $M_k = M'_k$ pour $k \in [\![1,n]\!]$. Par conséquent $S_\Delta(g) = S_{\Delta'}(f)$ et $S^\Delta(g) = S^{\Delta'}(f)$, pour toute subdivision Δ de $[a,b]$. Comme la transformation $\Delta \mapsto \Delta'$ réalise une bijection entre l'ensemble des subdivisions de $[a,b]$ et l'ensemble des subdivisions de $[a+c, b+c]$, on en déduit que

$$I^*(g) = \inf_{\Delta} S^\Delta(g) = \inf_{\Delta'} S^{\Delta'}(f) = I^*(f) \tag{3.53}$$

et de même

$$I_*(g) = \sup_{\Delta} S_\Delta(g) = \sup_{\Delta'} S_{\Delta'}(f) = I_*(f). \tag{3.54}$$

De plus, comme f est Riemann intégrable sur $[a+c, b+c]$, $I^*(f) = I^*(f)$. Compte-tenu de (3.53)–(3.54), on en déduit $I^*(g) = I_*(g) = I^*(f) = \int_{a+c}^{b+c} f(y)\,\mathrm{d}y$, ce qui nous donne la Riemann intégrabilité de g sur $[a,b]$ et l'égalité (3.50). □

Preuve de ii). La méthode étant essentiellement la même que pour le i), nous nous contenterons d'indiquer les adaptations nécessaires. Si $c > 0$, on associe à $\Delta = \{x_0 = a < x_1 < \cdots < x_n = b\}$ la subdivision Δ' dont les points sont les $y_k = cx_k$. Alors Δ' est une subdivision croissante de $[ac, bc]$ et comme $y_k - y_{k-1} = c(x_k - x_{k-1})$, on voit que $S_\Delta(h) = \frac{1}{c} S_{\Delta'}(f)$ et $S^\Delta(h) = \frac{1}{c} S^{\Delta'}(f)$. On en déduit comme ci-dessus l'intégrabilité de h et (3.51).

10. Il s'agit de $[ac, bc]$ si $c > 0$ et de $[bc, ac]$ si $c < 0$.

Si $c < 0$, $bc \leqslant ac$ et on prend pour subdivision croissante Δ' associée à Δ la subdivision de $[bc, ac]$ ayant pour points les $y_k = cx_{n-k}$. Alors pour $k \in [\![1, n]\!]$, $y_k - y_{k-1} = c(x_{n-k} - x_{n-k+1}) = -c(x_{n-k+1} - x_{n-k})$. On en déduit que $S_\Delta(h) = \frac{-1}{c} S_{\Delta'}(f)$ et $S^\Delta(h) = \frac{-1}{c} S^{\Delta'}(f)$ puis, que h est Riemann intégrable et

$$\int_a^b h(x)\,\mathrm{d}x = \frac{-1}{c}\int_{bc}^{ac} f(y)\,\mathrm{d}y = \frac{1}{c}\int_{ac}^{bc} f(y)\,\mathrm{d}y,$$

en utilisant la définition 3.3. □

Preuve de iii). D'abord, φ étant continue, l'image J de $[a, b]$ par φ est un *intervalle* (théorème des valeurs intermédiaires) et comme $[a, b]$ est *compact*, $J = \varphi([a, b])$ est aussi compact (l'image d'un compact par une application continue est un compact). Ainsi J est un intervalle compact, donc un intervalle fermé borné. Cet intervalle contient évidemment l'intervalle I d'extrémités $\varphi(a)$ et $\varphi(b)$ (pas forcément dans cet ordre), l'inclusion pouvant être stricte. La fonction f étant continue sur J l'est aussi par restriction sur I et l'intégrale au second membre de (3.52) est donc bien définie. L'intégrale du premier membre l'est tout autant puisque $(f \circ \varphi)\varphi'$ est continue sur $[a, b]$.

Introduisons les fonctions F, G, H suivantes :

$$F : J \to \mathbb{R}, s \mapsto \int_{\varphi(a)}^{s} f(y)\,\mathrm{d}y, \quad G : [a, b] \to \mathbb{R}, s \mapsto \int_a^s f\big(\varphi(x)\big)\varphi'(x)\,\mathrm{d}x, \quad H := F \circ \varphi.$$

Par le théorème 3.32, F est dérivable sur J et $F' = f$. De même G est dérivable sur $[a, b]$ et $G' = (f \circ \varphi)\varphi'$. D'autre part H est dérivable comme fonction composée et

$$H' = (F' \circ \varphi)\varphi' = (f \circ \varphi)\varphi' = G'.$$

Les fonctions H et G ont ainsi même dérivée sur $[a, b]$, leur différence est donc constante sur $[a, b]$. Or $H(a) = 0$ et $G(a) = 0$, donc $H = G$. En particulier, $H(b) = G(b)$, ce qui établit (3.52). □

Voici maintenant une version moins « classique » du changement de variable.

Proposition 3.34 (changement de variable)**.** *Si φ est C^1 et* monotone *sur $[a, b]$ et f est* Riemann intégrable *sur $\varphi([a, b])$,*

$$\int_a^b f\big(\varphi(x)\big)\varphi'(x)\,\mathrm{d}x = \int_{\varphi(a)}^{\varphi(b)} f(y)\,\mathrm{d}y. \tag{3.55}$$

La différence avec le changement de variable « classique » (Proposition 3.33 iii) est qu'on ne suppose plus f *continue*. On paie cette plus grande généralité pour f avec une hypothèse plus restrictive de monotonie pour φ. Comme cas particuliers, on retrouve les points i) et ii) de la proposition 3.33.

Preuve. Notons d'abord que l'existence de l'intégrale au second membre de (3.55) fait partie des hypothèses, mais que celle de l'intégrale du premier membre reste à établir. Pour alléger les écritures, convenons que les notations I, I^*, I_*, S_Δ, S^Δ, sont relatives aux bornes $\varphi(a)$ et $\varphi(b)$ lorsqu'elles s'appliquent à f et relatives aux bornes a et b lorsqu'elles s'appliquent à $(f\circ\varphi)\varphi'$. Nous détaillons la preuve dans le cas où φ est croissante, laissant au lecteur l'adaptation au cas où φ décroît.

Il suffit de démontrer (3.55) pour f positive sur $[\varphi(a),\varphi(b)]$. En effet (3.55) est clairement vérifiée si on remplace f par une fonction constante c et comme f est bornée inférieurement sur $[\varphi(a),\varphi(b)]$ car Riemann intégrable, on peut toujours trouver c telle que $g=f+c$ soit une fonction Riemann intégrable positive. Si on arrive à démontrer (3.55) pour g, alors on en déduit facilement la même égalité pour f par soustraction. Désormais, nous supposons donc sans perte de généralité que f est positive.

Fixons un $\varepsilon>0$ arbitraire. Comme f est Riemann intégrable sur $[\varphi(a),\varphi(b)]$, il existe une subdivision $\Delta=\{y_0=\varphi(a)<y_1<\cdots<y_n=\varphi(b)\}$ telle que

$$I(f)-\varepsilon<S_\Delta(f)=\sum_{k=1}^n (y_k-y_{k-1})m_k\leqslant S^\Delta(f)=\sum_{k=1}^n (y_k-y_{k-1})M_k<I(f)+\varepsilon.$$

Rappelons que $m_k=\inf\{f(y)\,;\,y\in[y_{k-1},y_k]\}$ et $M_k=\sup\{f(y)\,;\,y\in[y_{k-1},y_k]\}$. Comme φ est croissante et continue, l'image de $[a,b]$ par φ est le segment $[\varphi(a),\varphi(b)]$ et φ est une surjection [11] de $[a,b]$ sur $[\varphi(a),\varphi(b)]$. On peut donc trouver pour chaque y_k au moins un antécédent x_k par φ et quelque soit le choix de cet antécédent, on aura nécessairement $x_{k-1}<x_k$ car y_{k-1} est *strictement* inférieur à y_k et φ est croissante au sens large.

La fonction φ' étant continue sur le compact $[a,b]$ est *uniformément* continue sur cet intervalle. On peut donc subdiviser chaque $[x_{k-1},x_k]$ en $x_{k,0}=x_{k-1}<\cdots<x_{k,j_k}=x_k$ de sorte qu'en notant pour $j=0,\ldots,j_k$, $m'_{k,j}=\inf\{\varphi'(x)\,;\,x\in[x_{k,j-1},x_{k,j}]\}$ et $M'_{k,j}=\sup\{\varphi'(x)\,;\,x\in[x_{k,j-1},x_{k,j}]\}$, on ait :

$$M'_{k,j}-m'_{k,j}<\frac{\varepsilon}{M(b-a)}=:\varepsilon',\quad\text{où } M:=\sup_{[\varphi(a),\varphi(b)]} f.$$

En retenant de cette inégalité que $m'_{k,j}>M'_{k,j}-\varepsilon'$, on en déduit la minoration suivante :

$$\begin{aligned} y_k-y_{k-1}=\int_{x_{k-1}}^{x_k}\varphi'(x)\,\mathrm{d}x&=\sum_{j=1}^{j_k}\int_{x_{k,j-1}}^{x_{k,j}}\varphi'(x)\,\mathrm{d}x\\ &\geqslant\sum_{j=1}^{j_k} m'_{k,j}(x_{k,j}-x_{k,j-1})\\ &\geqslant\sum_{j=1}^{j_k}(M'_{k,j}-\varepsilon')(x_{k,j}-x_{k,j-1}),\end{aligned}$$

11. Elle n'est pas forcément injective car on n'a pas supposé la croissance stricte.

d'où en remarquant que M_k hérite de la positivité de f,

$$M_k(y_k - y_{k-1}) \geqslant \sum_{j=1}^{j_k} M_k M'_{k,j}(x_{k,j} - x_{k,j-1}) - \varepsilon' M_k(x_k - x_{k-1}).$$

En notant $M_{k,j}$ le supremum de $f \circ \varphi$ sur $[x_{k,j-1}, x_{k,j}]$, on a

$$M_k M'_{k,j} \geqslant M_{k,j} M'_{k,j} \geqslant \sup_{[x_{k,j-1}, x_{k,j}]} (f \circ \varphi)\varphi'.$$

Finalement en désignant par Δ' la subdivision de $[a, b]$ formée par tous les $x_{k,j}$, $k = 1, \ldots, n$, $j = 0, \ldots, j_k$, on aboutit à

$$S^{\Delta}(f) \geqslant S^{\Delta'}\big((f \circ \varphi)\varphi'\big) - \varepsilon \geqslant I^*\big((f \circ \varphi)\varphi'\big) - \varepsilon.$$

Comme ce dernier minorant ne dépend pas de Δ, on en déduit que

$$I(f) \geqslant I^*\big((f \circ \varphi)\varphi'\big) - \varepsilon.$$

Enfin, comme ε est arbitraire, il vient $I(f) \geqslant I^*\big((f \circ \varphi)\varphi'\big)$.

On laisse le soin au lecteur d'adapter l'argumentation ci-dessus pour montrer de même que $I(f) \leqslant I_*\big((f \circ \varphi)\varphi'\big) + \varepsilon$, puis $I(f) \leqslant I_*\big((f \circ \varphi)\varphi'\big)$. En rassemblant ces inégalités et en se souvenant que $I_* \leqslant I^*$, on en déduit que

$$I(f) = I^*\big((f \circ \varphi)\varphi'\big) = I_*\big((f \circ \varphi)\varphi'\big),$$

ce qui prouve (3.55) en établissant au passage la Riemann intégrabilité de $(f \circ \varphi)\varphi'$. □

3.4 Interversion limite intégrale

Théorème 3.35. *Soit $(f_n)_{n \geqslant 1}$ une suite de fonctions Riemann intégrables sur $[a, b]$. On suppose que cette suite converge uniformément vers f sur $[a, b]$. Alors f est Riemann intégrable sur $[a, b]$ et*

$$\int_a^b f_n(t)\,\mathrm{d}t \xrightarrow[n \to +\infty]{} \int_a^b f(t)\,\mathrm{d}t. \tag{3.56}$$

Preuve. Dans la preuve de ce théorème, la partie difficile est d'établir la Riemann intégrabilité de f, mais nous l'avons déjà vue par la proposition 3.14. Une fois que l'on sait que f est Riemann intégrable, on peut écrire :

$$\forall n \in \mathbb{N}^*, \quad \left| \int_a^b f(t)\,\mathrm{d}t - \int_a^b f_n(t)\,\mathrm{d}t \right| \leqslant \int_a^b |f(t) - f_n(t)|\,\mathrm{d}t. \tag{3.57}$$

Par convergence uniforme, pour tout $\varepsilon > 0$, il existe un entier n_ε tel que

$$\forall n \geqslant n_\varepsilon,\ \forall t \in [a, b], \quad |f(t) - f_n(t)| = |f - f_n|(t) < \frac{\varepsilon}{b - a}.$$

En reportant cette inégalité dans (3.57), en utilisant la linéarité de l'intégrale et la proposition 3.22 iii), on en déduit que pour tout $n \geqslant n_\varepsilon$,

$$\Big|\int_a^b f(t)\,\mathrm{d}t - \int_a^b f_n(t)\,\mathrm{d}t\Big| = \Big|\int_a^b (f-f_n)(t)\,\mathrm{d}t\Big| \leqslant \int_a^b |f-f_n|(t)\,\mathrm{d}t \leqslant \int_a^b \frac{\varepsilon}{b-a}\,\mathrm{d}t = \varepsilon.$$

Ceci étant valable pour tout $\varepsilon > 0$, la convergence (3.56) est établie. □

Pour finir, voici un théorème non classique [12] dont la preuve pourrait être vue comme un simple exercice mais qui jouera un rôle essentiel dans la suite de cet ouvrage, cf. la preuve du théorème de B. Levi (th. 7.23).

Théorème 3.36. *Soit $(f_n)_{n\geqslant 1}$ une suite de* fonctions toutes décroissantes sur $[a,b]$. *On suppose de plus que :*

$$\forall t \in [a,b], \quad f_n(t) \xrightarrow[n\to+\infty]{} f(t) \in \mathbb{R}.$$

Alors la fonction f ainsi définie est Riemann intégrable sur $[a,b]$ et

$$\int_a^b f_n(t)\,\mathrm{d}t \xrightarrow[n\to+\infty]{} \int_a^b f(t)\,\mathrm{d}t.$$

Il est clair que le théorème reste vrai si toutes les f_n d'indice $n \geqslant n_0$ sont décroissantes ou si elles sont croissantes pour tout $n \geqslant n_0$. Attention à ne pas confondre « suite de *fonctions décroissantes* sur $[a,b]$ » avec « *suite décroissante* de fonctions définies sur $[a,b]$ ». Ici on est dans le premier cas et on ne suppose rien sur le sens de variation des suites de réels $\big(f_n(t)\big)_{n\geqslant 1}$, $t \in [a,b]$.

Preuve. Remarquons d'abord que la fonction limite f est décroissante sur $[a,b]$ comme limite d'une suite de fonctions décroissantes puisque le passage à la limite conserve les inégalités larges. Par la proposition 3.10, f est donc elle aussi Riemann intégrable.

Cette Riemann intégrabilité de f nous assure de l'existence, pour $\varepsilon > 0$ arbitraire fixé, d'une subdivision $\Delta = \{t_0 = a < t_1 < \cdots < t_j = b\}$ telle que

$$S^\Delta(f) - \varepsilon < \int_a^b f(t)\,\mathrm{d}t < S_\Delta(f) + \varepsilon. \tag{3.58}$$

Notons que par *décroissance* de f et de f_n, ces fonctions *atteignent* sur $[t_{k-1}, t_k]$ leur supremum au point t_{k-1} et leur infimum [13] au point t_k. On peut donc expliciter comme suit pour f et f_n les sommes de Darboux supérieures

$$S^\Delta(f) = \sum_{k=1}^{j} f(t_{k-1})(t_k - t_{k-1}), \quad S^\Delta(f_n) = \sum_{k=1}^{j} f_n(t_{k-1})(t_k - t_{k-1}),$$

12. L'auteur n'en a pas trouvé trace dans la littérature et a du l'établir pour les besoins de cet ouvrage. Il serait évidemment reconnaissant à tout lecteur qui lui indiquerait une référence adéquate.

13. Donc le supremum et l'infimum sont ici respectivement un maximum et un minimum.

et les sommes de Darboux inférieures :

$$S_\Delta(f) = \sum_{k=1}^{j} f(t_k)(t_k - t_{k-1}), \quad S_\Delta(f_n) = \sum_{k=1}^{j} f_n(t_k)(t_k - t_{k-1}).$$

Par convergence simple de f_n vers f sur $[a,b]$, on peut trouver n_ε tel que

$$\forall n \geqslant n_\varepsilon,\ \forall k \in [\![0,j]\!], \quad |f(t_k) - f_n(t_k)| < \frac{\varepsilon}{b-a}. \tag{3.59}$$

En effet on n'a qu'un nombre fini $j+1$ d'écarts à contrôler (rappelons que pour l'instant ε est fixé et donc j aussi) et par convergence de $f_n(t_k)$ vers $f(t_k)$, on trouve pour chaque $k \in [\![0,j]\!]$ un rang $n_{\varepsilon,k}$ à partir duquel l'inégalité ci-dessus est toujours réalisée. On prend alors $n_\varepsilon = \max_{0\leqslant k\leqslant j} n_{\varepsilon,k}$. En utilisant (3.59) et la positivité des $(t_k - t_{k-1})$, on en déduit immédiatement que

$$S^\Delta(f) > S^\Delta(f_n) - \varepsilon, \quad S_\Delta(f) < S_\Delta(f_n) + \varepsilon. \tag{3.60}$$

En reportant ces inégalités dans (3.58), on obtient

$$S^\Delta(f_n) - 2\varepsilon < \int_a^b f(t)\,\mathrm{d}t < S_\Delta(f_n) + 2\varepsilon,$$

puis

$$\int_a^b f_n(t)\,\mathrm{d}t - 2\varepsilon < \int_a^b f(t)\,\mathrm{d}t < \int_a^b f_n(t)\,\mathrm{d}t + 2\varepsilon.$$

Autrement dit, nous avons trouvé un entier n_ε tel que

$$\forall n \geqslant n_\varepsilon, \quad \left|\int_a^b f_n(t)\,\mathrm{d}t - \int_a^b f(t)\,\mathrm{d}t\right| < 2\varepsilon.$$

Comme ε était arbitraire, ceci exprime précisément la convergence quand n tend vers l'infini, de $\int_a^b f_n(t)\,\mathrm{d}t$ vers $\int_a^b f(t)\,\mathrm{d}t$. □

3.5 Exercices

Ex. 3.1 La fonction indicatrice de l'ensemble $\mathbb{D}$ des nombres décimaux est-elle Riemann-intégrable sur $[0,1]$?

Ex. 3.2 Montrez que si f et g sont Riemann intégrables sur $[a,b]$, les fonctions

$$\min(f,g) : x \mapsto \min\big(f(x), g(x)\big) \quad \text{et} \quad \max(f,g) : x \mapsto \max\big(f(x), g(x)\big)$$

le sont aussi.

Indication p. 421

Ex. 3.3 Montrez que si f est continue et positive sur $[a,b]$ et telle que $\int_a^b f(t)\,\mathrm{d}t = 0$, alors $f(x) = 0$ pour tout $x \in [a,b]$. Le résultat subsiste-t-il si on remplace « continue » par « Riemann intégrable » ?

Indication p. 421

Ex. 3.4 Soit f une fonction positive et continue sur $[a,b]$. Notons $M := \sup_{[a,b]} f$. Montrez que

$$\lim_{n\to+\infty} \left(\int_a^b f(t)^n \,\mathrm{d}t \right)^{1/n} = M.$$

Solution p. 421

Ex. 3.5 *Sommes de Riemann*

Soit $f : [a,b] \to \mathbb{R}$ continue sur $[a,b]$. On suppose que pour tout $n \geqslant 1$, la suite finie $(t_{n,k})_{k\in[\![1,n]\!]}$ vérifie :

$$\forall k \in [\![1,n]\!], \quad a + \frac{k-1}{n}(b-a) \leqslant t_{n,k} \leqslant a + \frac{k}{n}(b-a).$$

1) Montrez que

$$\frac{1}{n}\sum_{k=1}^{n} f(t_{n,k}) \xrightarrow[n\to+\infty]{} \frac{1}{b-a}\int_a^b f(t)\,\mathrm{d}t.$$

2) Montrez que si f est lipschitzienne, la vitesse de convergence est en $O(1/n)$.

3) Montrez que

$$\lim_{n\to+\infty} \sum_{k=1}^{n} \frac{1}{n+k} = \ln 2.$$

Solution p. 421

Ex. 3.6 *Riemann intégrabilité d'une fonction admissible*

Un sous-ensemble D de $\mathbb{R}$ est dit $\mathcal{R}$-négligeable si pour tout $\varepsilon > 0$, on peut le recouvrir par une union finie d'intervalles ouverts dont la somme des longueurs est inférieure à ε. On dit que $f : [a,b] \to \mathbb{R}$ est *admissible* si elle est bornée sur $[a,b]$ et continue sauf peut-être sur une partie $\mathcal{R}$-négligeable de $[a,b]$. Montrez que toute fonction admissible est Riemann intégrable [14]. Ceci sera généralisé au chapitre 4, cf. proposition 4.71.

Ex. 3.7 *Riemann intégrabilité par prolongement*

Soit $f : [a,b[\to \mathbb{R}$, Riemann intégrable sur tout $[a,c] \subset [a,b[$ et *bornée* sur $[a,b[$, a et b étant deux réels. Montrez qu'en prolongeant f par n'importe quelle valeur réelle en b, la nouvelle fonction $\tilde{f}$ ainsi obtenue est Riemann intégrable sur $[a,b]$.

Solution p. 423

14. En élargissant la notion d'ensemble négligeable en remplaçant ci-dessus « union finie » par « union dénombrable », on obtient *toutes* les fonctions Riemann intégrables sur $[a,b]$, en vertu du théorème de Lebesgue de caractérisation de la Riemann intégrabilité, cf. Deheuvels [3] pp. 42–44.

Chapitre 4

Intégrale généralisée

L'intégrale de Riemann étudiée dans le chapitre 3 concerne des fonctions f définies en tout point d'un intervalle *fermé borné* $[a, b]$ et *bornées* sur cet intervalle. Il est utile de généraliser cette notion au cas de fonctions définies sur un intervalle quelconque, sauf peut-être en un nombre fini de points, et pas forcément bornées. Dans le cadre de cet ouvrage, les principales applications de cette notion d'intégrale généralisée concernent les lois à densité, l'espérance et les moments de variables aléatoires.

4.1 Construction

Commençons par examiner cette généralisation de l'intégrale de Riemann sur quelques exemples simples.

Exemple 4.1. *Peut-on définir* $\int_0^{+\infty} \mathrm{e}^{-x}\,\mathrm{d}x$ *?*

Ici l'intervalle d'intégration est $[0, +\infty[$ et l'intégrande $f : x \mapsto \mathrm{e}^{-x}$ est définie et continue sur cet intervalle, donc en particulier Riemann intégrable sur tout sous-intervalle fermé borné de la forme $[0, b]$. L'intégrale $\int_0^b \mathrm{e}^{-x}\,\mathrm{d}x$ a bien un sens. Elle se calcule d'ailleurs immédiatement par primitivation de f et vaut $\left[-\mathrm{e}^{-x}\right]_0^b = 1 - \mathrm{e}^{-b}$. Cette valeur a pour limite 1 quand b tend vers l'infini et il est naturel de prendre cette limite comme définition de l'intégrale généralisée $\int_0^{+\infty} \mathrm{e}^{-x}\,\mathrm{d}x$. L'interprétation géométrique de ce résultat est que l'aire de l'hypographe $H := \{(x, y) \in \mathbb{R}^2;\ x \in \mathbb{R}_+,\ 0 \leqslant y \leqslant \mathrm{e}^{-x}\}$ vaut 1, cf. figure 4.1. Pour le justifier, on remarque que H est la réunion de la suite croissante pour l'inclusion $(H_n)_{n\geqslant 1}$, où H_n est l'hypographe de f entre 0 et n. Par continuité séquentielle croissante de la mesure de Lebesgue λ_2, cf. proposition 5.16 page 167, on a $\lambda_2(H) = \lim_{n\to+\infty} \lambda_2(H_n) = \lim_{n\to+\infty} \int_0^n \mathrm{e}^{-x}\,\mathrm{d}x = 1$.

Exemple 4.2. *Peut-on définir* $\displaystyle\int_0^{+\infty} \frac{x}{1+x^2}\,\mathrm{d}x$ *?*

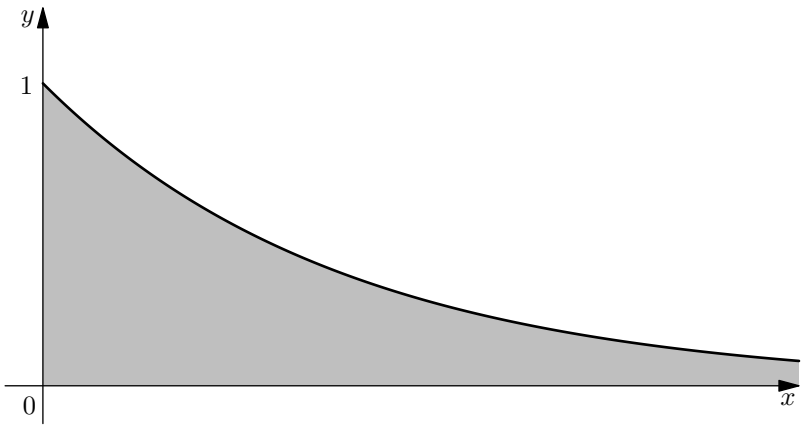


FIGURE 4.1 – Hypographe H de $f : x \mapsto \mathrm{e}^{-x}$ entre 0 et $+\infty$

Comme dans l'exemple 4.1, l'intégrale sur $[0, b]$ a un sens pour tout $b \in \mathbb{R}_+$. Elle se calcule par changement de variable $u = x^2$ et primitivation :

$$\int_0^b \frac{x}{1+x^2}\,\mathrm{d}x = \frac{1}{2}\int_0^{b^2} \frac{\mathrm{d}u}{1+u} = \frac{1}{2}\big[\ln(1+u)\big]_0^{b^2} = \frac{1}{2}\ln(1+b^2).$$

Cette quantité tend vers $+\infty$ lorsque b tend vers $+\infty$. On conviendra donc que

$$\int_0^{+\infty} \frac{x}{1+x^2}\,\mathrm{d}x = +\infty.$$

L'interprétation géométrique est que l'aire de l'hypographe de $x \mapsto x(1+x^2)^{-1}$ entre 0 et $+\infty$ est infinie. La justification repose sur la continuité séquentielle croissante de λ_2 comme pour l'exemple 4.1.

Exemple 4.3. *Peut-on définir l'intégrale* $\int_0^{+\infty} \cos x\,\mathrm{d}x$ *?*

La fonction cosinus est continue sur $\mathbb{R}$, donc Riemann intégrable sur tout intervalle $[0, b]$. Le calcul par primitivation de cette intégrale donne

$$\forall b > 0, \quad \int_0^b \cos x\,\mathrm{d}x = \big[\sin x\big]_0^b = \sin b.$$

Lorsque b tend vers $+\infty$, $\sin b$ n'a pas de limite, même dans $\overline{\mathbb{R}}$, on ne peut donc pas définir $\int_0^{+\infty} \cos x\,\mathrm{d}x$. Géométriquement, le domaine H délimité par le demi-axe des abscisses positives, l'axe des ordonnées et la courbe $y = \cos x$, $x \geqslant 0$, n'a pas d'aire algébrique (figure 4.2).

Exemple 4.4. *Peut-on définir les intégrales* $\displaystyle\int_0^1 \frac{\mathrm{d}t}{t^\alpha}$, $\alpha > 0$ *?*

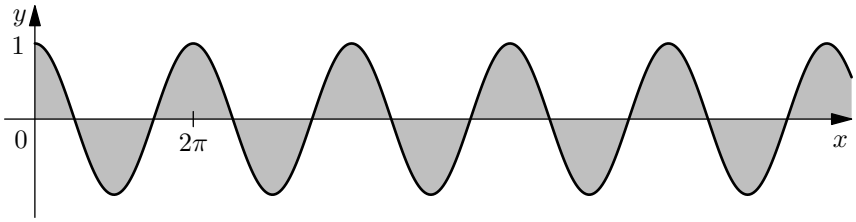


FIGURE 4.2 – Domaine H délimité par le graphe de $f : x \mapsto \cos x$ entre 0 et $+\infty$

Notons en préalable que si $\alpha \leqslant 0$, la réponse est immédiate puisque sur $[0,1]$ la fonction $f : t \mapsto t^{-\alpha}$ est continue donc Riemann intégrable. En revanche si $\alpha > 0$, f n'est pas définie en 0, est continue sur $]0,1]$ et tend vers $+\infty$ à droite en 0.

Puisque f est continue sur $]0,1]$, elle est Riemann intégrable sur tout intervalle $[a,1]$ pour $a > 0$. On va donc regarder la convergence éventuelle de $I(a) := \int_a^1 f(t)\,\mathrm{d}t$ lorsque a tend vers 0 par valeurs supérieures. L'intégrale $I(a)$ se calcule par primitivation. Si $\alpha \neq 1$, on obtient

$$\forall a > 0, \quad I(a) = \int_a^1 \frac{\mathrm{d}t}{t^\alpha} = \left[\frac{t^{-\alpha+1}}{-\alpha+1}\right]_a^1 = \frac{1-a^{-\alpha+1}}{-\alpha+1}$$

Quand a tend vers 0 par la droite, $I(a)$ tend vers une limite finie $1/(1-\alpha)$ si $-\alpha+1 > 0$, c'est-à-dire si $\alpha < 1$. En revanche si $-\alpha+1 < 0$, $I(a)$ tend vers $+\infty$, compte tenu du signe négatif du dénominateur constant $-\alpha+1$. Dans le cas particulier $\alpha = 1$, une primitive de f est la fonction logarithme néperien, d'où

$$\forall a > 0, \quad I(a) = \int_a^1 \frac{\mathrm{d}t}{t} = \big[\ln t\big]_a^1 = -\ln a,$$

ce qui tend vers $+\infty$ quand a tend vers 0 par la droite.

Finalement nous pouvons écrire

$$\int_0^1 \frac{\mathrm{d}t}{t^\alpha} = \begin{cases} \frac{1}{1-\alpha} & \text{si } \alpha < 1, \\ +\infty & \text{si } \alpha \geqslant 1. \end{cases}$$

À nouveau on peut interpréter géométriquement ce résultat. Soit $H_\alpha := \{(t,y) \in \mathbb{R}^2;\ 0 < t \leqslant 1 \text{ et } 0 \leqslant y \leqslant t^{-\alpha}\}$ l'hypographe de f entre 0 et 1, cf. figure 4.3. Si $\alpha < 1$, son aire est finie et vaut $(1-\alpha)^{-1}$. Si $\alpha \geqslant 1$, son aire est infinie. Pour la justification, on peut utiliser la suite, croissante pour l'inclusion, des hypographes de f entre $1/n$ et 1.

Exemple 4.5. *Peut-on définir* $\displaystyle\int_{-1}^1 \frac{\mathrm{d}t}{t}$ *?*

L'intégrande $f : t \mapsto t^{-1}$ est définie et continue sur l'intervalle « troué » $[-1,1]\backslash\{0\}$. Elle est donc Riemann intégrable sur chacun des intervalles fermés bornés $[-1,a]$ et

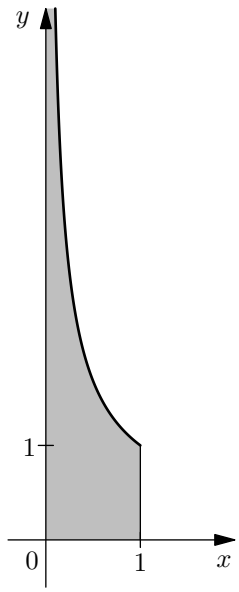


FIGURE 4.3 – Hypographe H_α de $f : t \mapsto t^{-\alpha}$ entre 0 et 1

$[b, +1]$ pour $-1 < a < 0$ et $0 < b < 1$. Ceci nous amène à étudier la limite quand a et b tendent vers respectivement 0 à gauche ou à droite de

$$I(a,b) := \int_{-1}^{a} \frac{\mathrm{d}t}{t} + \int_{b}^{1} \frac{\mathrm{d}t}{t} = \big[\ln|t|\big]_{-1}^{a} + \big[\ln|t|\big]_{b}^{1} = \ln|a| - \ln b.$$

Quand a et b tendent vers $0-$ et $0+$ respectivement, $\ln|a|$ et $\ln b$ tendent tous deux vers $+\infty$, donc leur différence $I(a,b)$ n'a pas de limite. Si vous n'en êtes pas convaincu, regardez ce problème de convergence avec les suites $a_n := 1/n$, $b_n = 1/n^2$, puis avec les suites $a'_n = -1/n^2$ et $b'_n = 1/n$. Dans le premier cas $I(a_n, b_n) = \ln n$ tend vers $+\infty$, tandis que dans le second, $I(a'_n, b'_n) = -\ln n$ tend vers $-\infty$. Ceci interdit à la *fonction de deux variables* $(a,b) \mapsto I(a,b)$ d'avoir une limite en $(0,0)$. On ne peut donc pas définir l'intégrale généralisée $\int_{-1}^{1} \frac{\mathrm{d}t}{t}$, *même comme élément de* $\overline{\mathbb{R}}$. Soit H le domaine délimité par le graphe de f et l'axe des abscisses entre -1 et 1 :

$$H := \{(t,y) \in \mathbb{R}^2;\ t \in [-1,1]\backslash\{0\},\ -(1/t)^- \leqslant y \leqslant (1/t)^+\},$$

voir figure 4.4.

Le fait que l'on ne puisse définir $\int_{-1}^{1} \frac{\mathrm{d}t}{t}$ signifie que H *n'a pas* d'aire *algébrique*[1].

Attention, il peut paraître tentant au vu de l'imparité de f de définir $\int_{-1}^{1} \frac{\mathrm{d}t}{t}$ comme valant 0 et de dire que l'aire algébrique de H est nulle. Il faut absolument résister à

1. En revanche il a une aire $\lambda_2(H) = +\infty$, ce qui correspond au fait que $\int_{-1}^{1} \frac{\mathrm{d}t}{|t|} = +\infty$ (exercice).

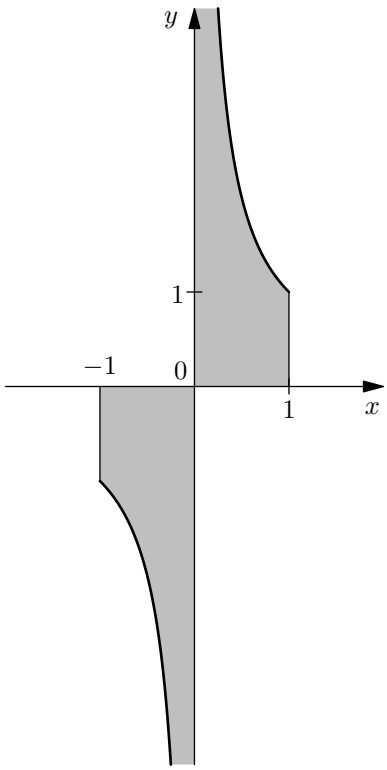


FIGURE 4.4 – $H := \{(t,y) \in \mathbb{R}^2;\ t \in [-1,1]\backslash\{0\},\ -(1/t)^- \leqslant y \leqslant (1/t)^+\}$

cette tentation. En effet, ceci reviendrait à dire que $I(a,b)$ tend vers 0 quand (a,b) tend vers $(0,0)$ simplement parce que $I(a,-a) = 0$ pour tout $a \in [-1,0[$.

Après ces exemples introductifs, nous allons formaliser la définition de l'intégrale de Riemann généralisée. Il est commode de désigner les ensembles d'intégration considérés sous le nom [2] d'« intervalle troué ».

Définition 4.6 (intervalle troué)**.** *Soit I un intervalle quelconque de $\mathbb{R}$ et $T = \{t_1, \ldots, t_d\}$ une partie de cardinal d de I, l'indexation des t_i vérifiant :*

$$-\infty \leqslant t_0 := \inf I < t_1 < \cdots < t_d < t_{d+1} := \sup I \leqslant +\infty.$$

On appelle intervalle troué *I_T l'ensemble $I\backslash T$. On a alors*

$$I_T = \bigcup_{i=0}^{d} I_i, \tag{4.1}$$

2. Non standard.

avec $I_i :=]t_i, t_{i+1}[$ *pour* $1 \leqslant i < d$, I_0 *a pour bornes* t_0 *et* t_1 *et est ouvert à droite,* I_d *a pour bornes* t_d *et* t_{d+1} *et est ouvert à gauche. Nous engloberons dans cette définition et ces notations le cas particulier* $d = 0$ *où il n'y a pas de trous, la réunion ci-dessus se réduisant à* $I_\varnothing = I = I_0$.

Les ensembles d'intégration utilisés dans les exemples 4.1–4.5 sont ainsi des intervalles troués avec $I = [0, +\infty[$ et $d = 0$ pour les exemples 4.1, 4.2 et 4.3, $I =]0, 1]$ et $d = 0$ pour l'exemple 4.4, $I = [-1, 1]$, $d = 1$, $t_1 = 0$ pour l'exemple 4.5.

Définition 4.7 (fonction localement Riemann intégrable)**.** *Soient* I_T *un intervalle troué et* f *une fonction* $I_T \to \mathbb{R}$. *On dit que* f *est* localement *Riemann intégrable sur* I_T *si elle est Riemann intégrable sur tout intervalle fermé borné* $[\alpha, \beta]$ *inclus dans* I_T, *donc nécessairement inclus dans l'un des intervalles* I_i *de la décomposition (4.1).*

Définition 4.8 (intégrale généralisée)**.** *Soient* I *un intervalle de* $\mathbb{R}$, *de bornes* $a, b \in \overline{\mathbb{R}}$, $T = \{t_1, \dots, t_d\}$ *un ensemble de trous dans* I *et* f *localement Riemann intégrable sur l'intervalle troué* I_T. *On dit que l'intégrale généralisée de* f *entre* a *et* b *converge si on peut trouver une suite finie* $c_0, c_1, \dots, c_d$ *avec* $c_i \in I_i$ *pour* $0 \leqslant i \leqslant d$ *telle que chacune des limites suivantes existe et soit finie :*

$$\forall i \in [\![0, d]\!], \quad \lim_{\substack{x_i \to t_{i+1}, \\ x_i < t_{i+1}}} \int_{c_i}^{x_i} f(t)\,\mathrm{d}t =: \ell_i, \quad \lim_{\substack{x'_i \to t_i, \\ x'_i > t_i}} \int_{x'_i}^{c_i} f(t)\,\mathrm{d}t =: \ell'_i. \tag{4.2}$$

On dit alors que l'intégrale généralisée $\int_a^b f(t)\,\mathrm{d}t$ converge *et on définit* $\int_a^b f(t)\,\mathrm{d}t$ *comme le réel*

$$\int_a^b f(t)\,\mathrm{d}t := \sum_{i=0}^{d} (\ell'_i + \ell_i). \tag{4.3}$$

Si l'une au moins des conditions (4.2) n'est pas vérifiée, c'est-à-dire s'il n'y a pas de limite ou une limite infinie, on dit que l'intégrale généralisée $\int_a^b f(t)\,\mathrm{d}t$ diverge. *Dans ce cas l'écriture* $\int_a^b f(t)\,\mathrm{d}t$ *ne représente pas un nombre réel.*

L'utilisation du symbole $\int_a^b f(t)\,\mathrm{d}t$ est analogue à celle de $\sum_{k=0}^{+\infty} u_k$ qui désigne à la fois une série et lorsqu'elle converge dans $\mathbb{R}$, sa somme qui est le réel limite de la suite des sommes partielles $S_n = \sum_{k=0}^{n} u_k$. La série peut diverger parce que la suite $(S_n)_{n\geqslant 1}$ tend vers $+\infty$ (resp. $-\infty$), auquel cas on s'autorise l'écriture $\sum_{k=0}^{+\infty} u_k = +\infty$ (resp. $= -\infty$). Mais elle peut aussi diverger parce que $(S_n)_{n\geqslant 1}$ *n'a pas* de limite, *même infinie*. Dans ce cas l'écriture $\sum_{k=0}^{+\infty} u_k$ est *purement formelle* et ne représente pas un élément de $\overline{\mathbb{R}}$. Pour l'intégrale généralisée divergente $\int_a^b f(t)\,\mathrm{d}t$, on s'autorisera l'écriture $\int_a^b f(t)\,\mathrm{d}t = +\infty$ si certains des ℓ_i, ou ℓ'_i valent $+\infty$, les autres étant *finis*. De même on écrira $\int_a^b f(t)\,\mathrm{d}t = -\infty$ si certains des ℓ_i, ou ℓ'_i valent $-\infty$, les autres étant *finis*. Dans tous les autres cas de divergence [3], le symbole $\int_a^b f(t)\,\mathrm{d}t$ est seulement une

3. C'est-à-dire si l'une au moins des intégrales de (4.2) n'a pas de limite même dans $\overline{\mathbb{R}}$ ou si elles ont toutes une limite dans $\overline{\mathbb{R}}$, mais avec au moins une des limites valant $-\infty$ *et* au moins une valant $+\infty$.

écriture formelle et ne représente pas un élément de $\overline{\mathbb{R}}$.

Le lecteur attentif n'aura pas manqué de noter que la définition 4.8 pose un problème de cohérence car l'existence et la valeur de $\int_a^b f(t)\,\mathrm{d}t$ semblent dépendre du choix de la suite $c_0, c_1, \ldots, c_d$. Le lemme suivant répond à cette légitime inquiétude.

Lemme 4.9. *Avec les notations de la définition 4.8, on suppose que la suite finie $c_0, c_1, \ldots, c_d$ vérifie (4.2). Soit $\tilde{c}_0, \tilde{c}_1, \ldots, \tilde{c}_d$ une suite telle que $\tilde{c}_i \in I_i$ pour $i \in [\![0, d]\!]$. Alors on a*

$$\forall i \in [\![0, d]\!], \quad \lim_{\substack{x_i \to t_{i+1}, \\ x_i < t_{i+1}}} \int_{\tilde{c}_i}^{x_i} f(t)\,\mathrm{d}t = \ell_i + \int_{\tilde{c}_i}^{c_i} f(t)\,\mathrm{d}t =: \tilde{\ell}_i \tag{4.4}$$

et

$$\forall i \in [\![0, d]\!], \quad \lim_{\substack{x'_i \to t_i, \\ x'_i > t_i}} \int_{x'_i}^{\tilde{c}_i} f(t)\,\mathrm{d}t = \ell'_i + \int_{c_i}^{\tilde{c}_i} f(t)\,\mathrm{d}t =: \tilde{\ell}'_i. \tag{4.5}$$

Une conséquence immédiate de ce lemme est que

$$\sum_{i=0}^{d} (\ell'_i + \ell_i) = \sum_{i=0}^{d} (\tilde{\ell}'_i + \tilde{\ell}_i)$$

puisque $\ell'_i + \ell_i = \tilde{\ell}'_i + \tilde{\ell}_i$, la somme de $\int_{\tilde{c}_i}^{c_i}$ et $\int_{c_i}^{\tilde{c}_i}$ s'annihilant par la relation de Chasles. Ainsi ni la convergence de l'intégrale généralisée de f entre a et b ni la définition de sa valeur par (4.3) ne dépendent du choix de la suite $c_0, c_1, \ldots, c_d$.

Preuve du lemme 4.9. Puisque l'on fait tendre x_i vers t_{i+1} par valeurs inférieures, on peut toujours supposer que $\max(c_i, \tilde{c}_i) < x_i < t_{i+1}$. La fonction f est alors Riemann intégrable sur $[\min(c_i, \tilde{c}_i), x_i]$ et la relation de Chasles combinée avec (4.2) nous donne

$$\int_{\tilde{c}_i}^{x_i} f(t)\,\mathrm{d}t = \int_{\tilde{c}_i}^{c_i} f(t)\,\mathrm{d}t + \int_{c_i}^{x_i} f(t)\,\mathrm{d}t \xrightarrow[\substack{x_i \to t_{i+1}, \\ x_i < t_{i+1}}]{} \int_{\tilde{c}_i}^{c_i} f(t)\,\mathrm{d}t + \ell_i.$$

La vérification de (4.5) est analogue. □

Remarque 4.10. Dans le cas où l'intervalle I de la définition 4.8 est fermé en b (donc $b \in I$), la condition

$$\lim_{\substack{x_d \to b, \\ x_d < b}} \int_{c_d}^{x_d} f(t)\,\mathrm{d}t =: \ell_d \in \mathbb{R}$$

est automatiquement vérifiée. En effet, $[c_d, b]$ est inclus dans I_T, donc f est Riemann intégrable sur $[c_d, b]$ et la fonction $x_d \mapsto \int_{c_d}^{x_d} f(t)\,\mathrm{d}t$ est continue sur cet intervalle, cf. théorème 3.32 i), donc continue à gauche au point b, ce qui nous donne l'existence de la limite finie ℓ_d et sa valeur $\ell_d = \int_{c_d}^{b} f(t)\,\mathrm{d}t$. En pratique on s'abstiendra de revérifier

l'existence de cette limite. Par exemple si f est localement Riemann intégrable sur $]c,b]$, il faut seulement regarder ce qui se passe au voisinage de c.

Bien entendu si $a \in I$, on a une situation analogue, à savoir la convergence automatique quand $x_0' \to a+$ de $\int_{x_0'}^{c_0} f(t)\,\mathrm{d}t$ vers l'intégrale de Riemann ordinaire $\int_a^{c_0} f(t)\,\mathrm{d}t$.

Nous définissons aussi les intégrales généralisées $\int_a^b$ avec $a > b$ en cohérence avec la définition 3.3.

Définition 4.11. *Si $a > b$ et si l'intégrale généralisée $\int_b^a f(t)\,\mathrm{d}t$ converge, on pose*

$$\int_a^b f(t)\,\mathrm{d}t := -\int_b^a f(t)\,\mathrm{d}t. \tag{4.6}$$

Remarque 4.12 (réduction du problème)**.** Une fois payé notre tribut au formalisme avec la définition 4.8, il convient de se simplifier la vie en notant que l'étude de la convergence d'une intégrale généralisée se ramène à l'étude de limites du type $\int_{x'}^{c} f(t)\,\mathrm{d}t$ quand $x' \to a+$ avec f localement Riemann intégrable sur $]a,c]$ ou $\int_c^x f(t)\,\mathrm{d}t$ quand $x \to b-$ avec f localement Riemann intégrable sur $[c,b[$. Nous nous contenterons la plupart du temps, d'énoncer des résultats relatifs au deuxième type, en laissant au lecteur le soin d'écrire leur adaptation immédiate au premier type et de recoller les morceaux par (4.3).

Définition 4.13 (notation $\mathcal{R}^{\mathrm{loc}}[a,b[$)**.** *Soient a et b tels que $-\infty < a < b \leqslant +\infty$. Nous notons $\mathcal{R}^{\mathrm{loc}}[a,b[$ l'ensemble des fonctions localement Riemann intégrables sur $[a,b[$.*

Pour étudier la convergence d'une intégrale généralisée, on a souvent recours à une technique de comparaison avec une intégrale de référence. Les intégrales généralisées des fonctions puissances $t \mapsto t^{-\alpha}$ au voisinage de 0 ou de $+\infty$ sont les intégrales de référence les plus utilisées.

Proposition 4.14. *Soit α un réel.*

a) $\displaystyle\int_0^1 \frac{\mathrm{d}t}{t^\alpha}$ *converge si $\alpha < 1$ et diverge si $\alpha \geqslant 1$.*

b) $\displaystyle\int_1^{+\infty} \frac{\mathrm{d}t}{t^\alpha}$ *converge si $\alpha > 1$ et diverge si $\alpha \leqslant 1$.*

Preuve. Le a) a déjà été traité à l'exemple 4.4 en rappelant que si $\alpha \leqslant 0$, on a affaire à l'intégrale de Riemann ordinaire d'une fonction continue sur $[0,1]$. Pour le b), on note que $f : t \mapsto t^{-\alpha}$ est continue sur $[1,+\infty[$ donc localement Riemann intégrable sur cet intervalle, donc Riemann intégrable sur tout intervalle $[1,x]$ pour $x \geqslant 1$. Par primitivation on a

$$\int_1^x \frac{\mathrm{d}t}{t^\alpha} = \begin{cases} \frac{x^{-\alpha+1}}{-\alpha+1} - \frac{1}{1-\alpha} & \text{si } \alpha \neq 1, \\ \ln x & \text{si } \alpha \neq 1. \end{cases}$$

On en déduit que pour $\alpha \leqslant 1$, $\int_1^x t^{-\alpha}\,\mathrm{d}t$ tend vers $+\infty$ quand x tend vers $+\infty$ et que pour $\alpha > 1$, $\int_1^x t^{-\alpha}\,\mathrm{d}t$ tend vers la limite finie $1/(\alpha-1)$ quand x tend vers $+\infty$. □

Corollaire 4.15. *Si $-\infty < a < b < +\infty$ et si α est un réel,*

$$\int_a^b \frac{\mathrm{d}t}{(t-a)^\alpha} \qquad \textit{converge si et seulement si } \alpha < 1, \tag{4.7}$$

$$\int_a^b \frac{\mathrm{d}t}{(b-t)^\alpha} \qquad \textit{converge si et seulement si } \alpha < 1. \tag{4.8}$$

Preuve. C'est une adaptation immédiate de la preuve du a) ci-dessus, voir exemple 4.4, *via* les changements de variable $u = t - a$ et $v = b - t$ respectivement. □

Dans la situation de la proposition 4.14, l'intégrale diverge lorsque f tend trop vite vers $+\infty$ en 0 dans le cas a) et tend trop lentement vers 0 en $+\infty$ dans le cas b) [4]. Dans le cas b), on peut penser à l'analogie avec la série de terme général $k^{-\alpha}$, cf. théorème 2.14 p. 40. Pour autant il faut se garder de tirer des conclusions hâtives de cette analogie. Si la convergence d'une série implique toujours la convergence vers 0 de son terme général, la situation est plus compliquée pour les intégrales de la forme $\int_a^{+\infty} f(t)\,\mathrm{d}t$.

Remarque 4.16. La convergence de l'intégrale $\int_a^{+\infty} f(t)\,\mathrm{d}t$ *n'implique pas* que $f(t)$ ait pour limite 0 en $+\infty$. Voici un contre exemple. Prenons $a = 0$ et pour f la fonction continue affine par morceaux dont l'hypographe se réduit (en dehors des segments où f est nulle) à la réunion de la suite $(T_n)_{n\geqslant 1}$ des triangles isocèles de sommet principal $(n, 2^n)$ et de base $\left[n - 4^{-n}, n + 4^{-n}\right]$, cf. figure [5] 4.5. La fonction f vérifie les deux propriétés suivantes.

1. $\int_a^{+\infty} f(t)\,\mathrm{d}t$ converge (et vaut 1).
2. Pour tout $n \in \mathbb{N}^*$, $f(n) = 2^n$ et $f(n + 1/2) = 0$. Par conséquent f n'a pas de limite en $+\infty$.

Preuve. Comme fonction continue, f appartient à $\mathcal{R}^{\mathrm{loc}}[0, +\infty[$ donc est Riemann intégrable sur tout intervalle $[0, x]$, $x \in \mathbb{R}_+$. Les triangles T_n sont deux à deux disjoints et l'aire $\lambda_2(T_n)$ se calcule par la formule classique demi-produit de la base par la hauteur [6], d'où

$$\lambda_2(T_n) = 4^{-n} \times 2^n = 2^{-n}.$$

4. On pourrait d'ailleurs déduire le b) du a) par un argument géométrique sur les hypographes en notant que les fonctions $f : t \mapsto t^{-\alpha}$ et $g : t \mapsto t^{-1/\alpha}$ définies sur $]0, +\infty[$ sont réciproques l'une de l'autre, donc que leurs graphes en repère orthonormé se correspondent par la symétrie relativement à la première bissectrice du repère. Par conservation de λ_2, on en déduit que l'hypographe de g entre 1 et $+\infty$ a même aire que l'hypographe de f entre 0 et 1 privé du carré unité (faites le dessin !).

5. Pour des raisons de lisibilité on a muni l'axe des ordonnées d'une échelle logarithmique et on a fortement exagéré la base de chaque triangle isocèle.

6. Si vous êtres sceptiques vous pouvez toujours chercher une expression analytique pour la restriction de f à $[n - 4^{-n}, n + 4^n]$ et l'intégrer sur ce segment pour voir si vous trouvez le même résultat.

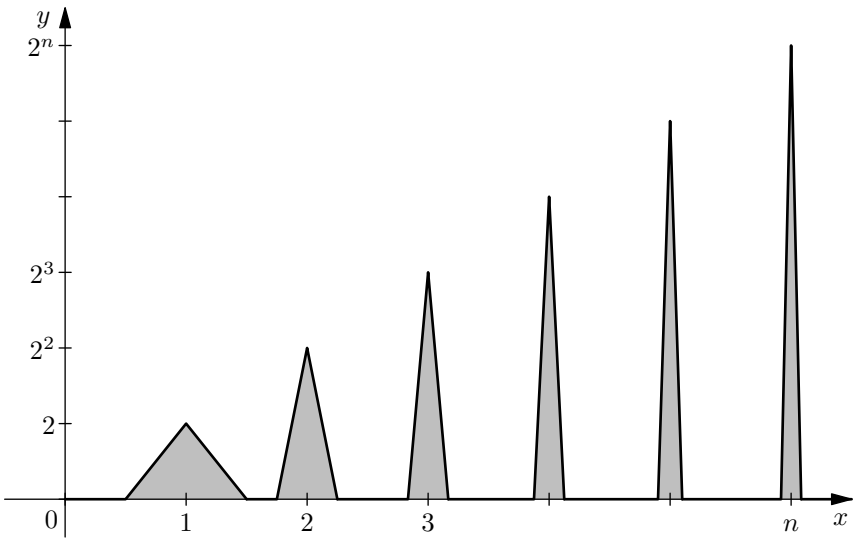


FIGURE 4.5 – $\int_0^{+\infty} f(t)\,\mathrm{d}t$ peut converger sans que f ait une limite en $+\infty$

On en déduit immédiatement que la série de terme général $\lambda_2(T_k)$ est géométrique convergente. Le calcul de sa somme partielle S_n est bien connu :

$$S_n := \sum_{k=1}^{n} \lambda_2(T_k) = \sum_{k=1}^{n} 2^{-k} = 2^{-1} \sum_{j=0}^{n-1} 2^{-j} = 2^{-1} \frac{1-2^{-n}}{1-2^{-1}} = 1 - 2^{-n}.$$

Nous allons montrer la convergence de $\int_0^{+\infty} f(t)\,\mathrm{d}t$ en comparant $\int_0^x f(t)\,\mathrm{d}t$ et S_n pour $n = [x]$, la partie entière de x. En effet, pour n fixé, $g_n : x \mapsto \int_0^x f(t)\,\mathrm{d}t - S_n$ est une fonction croissante puisque f est positive. Sur l'intervalle $[n, n+1]$, cette fonction a pour minimum $g_n(n) = -\frac{1}{2}\lambda_2(T_n)$ et pour maximum $g_n(n+1) = \frac{1}{2}\lambda_2(T_{n+1})$. Comme $\lambda_2(T_n) > \lambda_2(T_{n+1})$, on en déduit

$$\forall x \in [n, n+1], \quad \left| \int_0^x f(t)\,\mathrm{d}t - S_n \right| \leqslant \frac{1}{2}\lambda_2(T_n) < 2^{-n}.$$

Compte-tenu du calcul de S_n rappelé ci-dessus, ceci nous permet d'écrire :

$$\forall x \geqslant 1, \quad \left| \int_0^x f(t)\,\mathrm{d}t - \left(1 - 2^{-[x]}\right) \right| < 2^{-[x]}.$$

En faisant tendre x vers $+\infty$, on en déduit que $\int_0^x f(t)\,\mathrm{d}t$ tend vers 1. Donc $\int_0^{+\infty} f(t)\,\mathrm{d}t$ converge et vaut 1. Le point 2 est évident. □

Ce genre de pathologie n'est pas réservé aux intégrales généralisées de la forme $\int_0^{+\infty} f(t)\,\mathrm{d}t$. À titre d'exercice, on vous laisse le soin de construire une fonction f

continue sur $[0,1[$ telle que $\int_0^1 f(t)\,\mathrm{d}t$ converge et qu'il existe 3 suites $(u_n)_{n\geqslant 1}$, $(v_n)_{n\geqslant 1}$ et $(w_n)_{n\geqslant 1}$ convergentes vers 1 dans $[0,1[$ telles que $f(u_n)$ tende vers $+\infty$, $f(v_n)$ tende vers 0 et $f(w_n)$ tende vers $-\infty$. Voici une suggestion parmi les multiples solutions possibles. Découper $[0,1[$ en trois segments de même longueur $[0,1/3[$, $[1/3,2/3[$ et $[2/3,1[$. Sur les deux premiers prendre pour graphe de f les triangles isocèles de base ces segments et de « hauteurs » respectives $+2$ et -2. Itérer ce partage en trois sur $[2/3,1[$ avec sur les deux premiers segments du partage des triangles isocèles de hauteur $+4$ et -4 et ainsi de suite jusqu'à l'infini.

Après ces contre exemples, voyons ce que l'on peut dire dans des situations moins pathologiques sur la relation entre convergence de $\int_a^{+\infty} f(t)\,\mathrm{d}t$ et comportement de f au voisinage de $+\infty$.

Proposition 4.17. *Soit $f \in \mathcal{R}^{\mathrm{loc}}[a,+\infty[$.*

i) On suppose qu'il existe $A \in [a,+\infty[$ et $m > 0$ tels que $f(t) \geqslant m$ pour tout $t \geqslant A$. Alors $\int_a^{+\infty} f(t)\,\mathrm{d}t$ diverge.

ii) Cette divergence a lieu aussi s'il existe $m' < 0$ et $A \geqslant a$ tels que $f(t) \leqslant m'$ pour tout $t \geqslant A$.

iii) En conséquence, si f a une limite non nulle *$\ell \in \overline{\mathbb{R}}$ en $+\infty$, $\int_a^{+\infty} f(t)\,\mathrm{d}t$ diverge.*

Preuve. Pour i), il suffit de remarquer que pour tout $x \geqslant A$,

$$\int_a^x f(t)\,\mathrm{d}t = \int_a^A f(t)\,\mathrm{d}t + \int_A^x f(t)\,\mathrm{d}t \geqslant \int_a^A f(t)\,\mathrm{d}t + m(x-A),$$

par croissance de l'intégrale de Riemann sur $[A,x]$, voir la proposition 3.22 ii) et la figure 4.6. En faisant tendre x vers $+\infty$, on en déduit que $\int_a^x f(t)\,\mathrm{d}t$ tend vers $+\infty$, d'où la divergence de $\int_a^{+\infty} f(t)\,\mathrm{d}t$.

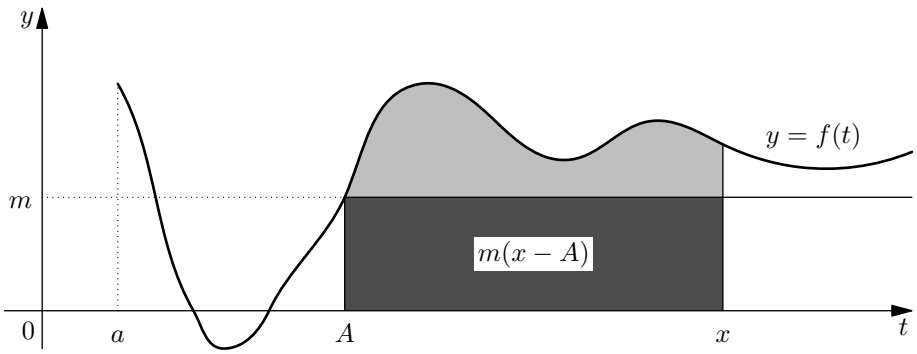


FIGURE 4.6 – $\int_A^x f(t)\,\mathrm{d}t \geqslant m(x-A)$

De même pour ii), on obtient la minoration $\int_a^x f(t)\,\mathrm{d}t \leqslant \int_a^A f(t)\,\mathrm{d}t + m'(x-A)$ et donc $\int_a^x f(t)\,\mathrm{d}t$ tend vers $-\infty$ quand x tend vers $+\infty$.

Supposons maintenant que f ait une limite *non nulle* $\ell \in \overline{\mathbb{R}}$ en $+\infty$. On peut distinguer 4 cas.

Cas 1. $\ell \in]0, +\infty[$, alors il existe un $A \geqslant a$ tel que pour tout $t \geqslant A$, $f(t) > \frac{\ell}{2}$. On applique i) avec $m = \frac{\ell}{2} > 0$.

Cas 2. $\ell = +\infty$, alors il existe un $A \geqslant a$ tel que pour tout $t \geqslant A$, $f(t) \geqslant 1$, on pourrait bien sûr remplacer ce minorant 1 par n'importe quel réel $B > 0$ choisi à l'avance[7]. On applique i) avec $m = 1$.

Cas 3. $\ell \in]-\infty, 0[$, alors il existe un $A \geqslant a$ tel que pour tout $t \geqslant A$, $f(t) < \frac{\ell}{2}$. On applique ii) avec $m' = \frac{\ell}{2} < 0$.

Cas 4. $\ell = -\infty$, alors il existe un $A \geqslant a$ tel que pour tout $t \geqslant A$, $f(t) \leqslant -1$. On applique ii) avec $m = -1$.

□

Corollaire 4.18. *Soit a un réel et f une fonction positive et décroissante sur $[a, +\infty[$. Si $\int_a^{+\infty} f(t)\,\mathrm{d}t$ converge, alors f tend vers 0 en $+\infty$.*

Preuve. Une fonction décroissante sur $[A, +\infty[$ est localement Riemann intégrable sur cet intervalle, cf. proposition 3.10. D'autre part comme fonction monotone, elle admet toujours en $+\infty$ une limite $\ell \in \overline{\mathbb{R}}$. Par décroissance et positivité de f, cette limite est nécessairement dans $[0, +\infty[$. Si $\ell > 0$, alors par le cas 1 ci-dessus, $\int_a^{+\infty} f(t)\,\mathrm{d}t$ diverge, ce qui contredit l'hypothèse de convergence de cette intégrale. Donc $\ell = 0$. □

4.2 Critère de Cauchy pour intégrales généralisées

L'intérêt du critère de Cauchy — dans un espace complet — est de permettre d'établir l'*existence* d'une limite sans connaître *a priori* sa valeur. Nous allons voir une version de ce critère pour la convergence des intégrales généralisées. Auparavant, il n'est peut-être pas superflu de rappeler quelques versions du critère de Cauchy pour l'existence de limites de suites ou de fonctions.

Proposition 4.19 (critères de Cauchy)**.**

1. *La suite de réels $(u_n)_{n\geqslant 1}$ converge dans $\mathbb{R}$ si et seulement si*

$$\forall \varepsilon > 0,\ \exists N \in \mathbb{N}^*, \quad \forall n, p \geqslant N, \quad |u_n - u_p| < \varepsilon. \tag{4.9}$$

2. *Soit F une fonction à valeurs réelles ou complexes, définie sur $D \subset \mathbb{R}$ et a un point adhérent à D. Alors F a une limite* finie *au point a si et seulement si :*

$$\forall \varepsilon > 0,\ \exists \delta > 0, \quad \forall x, x' \in D \cap]a-\delta, a+\delta[\setminus \{a\}, \quad |F(x) - F(x')| < \varepsilon. \tag{4.10}$$

3. *Soit F une fonction à valeurs réelles ou complexes, définie sur $D \subset \mathbb{R}$ et $a \in \mathbb{R}$ tel que pour tout réel $\eta > 0$, $D \cap]a, a+\eta[$ soit non vide. Alors F a une limite à droite* finie *au point a si et seulement si :*

$$\forall \varepsilon > 0,\ \exists \delta > 0, \quad \forall x, x' \in D \cap]a, a+\delta[, \quad |F(x) - F(x')| < \varepsilon. \tag{4.11}$$

7. Mais pas par $\frac{\ell}{2}$ qui a ici le mauvais goût de valoir $+\infty$!

4. *Soit F une fonction à valeurs réelles ou complexes, définie sur $D \subset \mathbb{R}$ et $a \in \mathbb{R}$ tel que pour tout réel $\eta > 0$, $D \cap]a-\eta, a[$ soit non vide. Alors F a une limite à gauche* finie *au point a si et seulement si :*

$$\forall \varepsilon > 0,\ \exists \delta > 0, \quad \forall x, x' \in D \cap]a-\delta, a[, \quad |F(x) - F(x')| < \varepsilon. \tag{4.12}$$

5. *Soit F une fonction à valeurs réelles ou complexes, définie sur un intervalle $[a, +\infty[$, $a \in \mathbb{R}$. Alors F a une limite* finie *en $+\infty$ si et seulement si :*

$$\forall \varepsilon > 0,\ \exists A > 0, \quad \forall x, x' \geqslant A, \quad |F(x) - F(x')| < \varepsilon. \tag{4.13}$$

De même que l'application du critère de Cauchy (4.9) à la suite des sommes partielles d'une série conduit au critère de Cauchy pour les séries, cf. théorème 2.9, les critères (4.10)–(4.13) nous fournissent des critères de Cauchy pour la convergence d'intégrales généralisées. Nous les énoncerons seulement pour la convergence en b des intégrales $\int_a^b f(t)\,\mathrm{d}t$ en utilisant (4.12) ou (4.13) selon que b est fini ou non. Au lecteur de compléter.

Théorème 4.20 (critère de Cauchy pour les intégrales)**.**

1. *Soit $f \in \mathcal{R}^{\mathrm{loc}}[a, +\infty[$. L'intégrale $\int_a^{+\infty} f(t)\,\mathrm{d}t$ converge si et seulement si :*

$$\forall \varepsilon > 0,\ \exists A > 0, \quad \forall x, x' \geqslant A, \quad \left| \int_x^{x'} f(t)\,\mathrm{d}t \right| < \varepsilon. \tag{4.14}$$

2. *Soit $f \in \mathcal{R}^{\mathrm{loc}}[a, b[$, avec $b < +\infty$. Alors $\int_a^b f(t)\,\mathrm{d}t$ converge si et seulement si :*

$$\forall \varepsilon > 0,\ \exists \delta \in]0, b-a[, \quad \forall x, x' \in]b-\delta, b[, \quad \left| \int_x^{x'} f(t)\,\mathrm{d}t \right| < \varepsilon. \tag{4.15}$$

Preuve. Il suffit d'appliquer le critère de Cauchy (4.12) ou (4.13) à la fonction $F : [a, b[\to \mathbb{R}$, $x \mapsto F(x) := \int_a^x f(t)\,\mathrm{d}t$ qui est bien définie sur $[a, b[$ puisque f est Riemann intégrable sur $[a, x]$ pour tout $x \in [a, b[$. □

Le corollaire 4.21 ci-dessous et l'exemple qui suit paraîtront évidents, voire superflus, si l'on a traité l'exercice 3.7 qui montre que dans la situation du corollaire 4.21, l'intégrale généralisée $\int_a^b f(t)\,\mathrm{d}t$ coïncide en fait avec l'intégrale de Riemann ordinaire d'un prolongement de f sur $[a, b]$.

Corollaire 4.21. *Soit $f \in \mathcal{R}^{\mathrm{loc}}[a, b[$ avec b fini.*

1. *Si f est* bornée sur $[a, b[$, *alors $\int_a^b f(t)\,\mathrm{d}t$ converge.*
2. *Si f a une limite à gauche finie en b, alors $\int_a^b f(t)\,\mathrm{d}t$ converge.*

Preuve. Vérifions le point 1. Par hypothèse, il existe $M \in \mathbb{R}_+$ tel que pour tout $t \in [a,b[$, $|f(t)| \leqslant M$. D'autre part f étant localement Riemann intégrable sur $[a,b[$ est Riemann intégrable sur tout segment $[x,x'] \subset [a,b[$. Par croissance de l'intégrale de Riemann, cf. proposition 3.22, on en déduit :

$$\forall x, x' \in [a,b[\text{ avec } x < x', \qquad \left| \int_x^{x'} f(t)\,\mathrm{d}t \right| \leqslant \int_x^{x'} |f(t)|\,\mathrm{d}t \leqslant M(x'-x). \tag{4.16}$$

Cette inégalité nous permet de vérifier le critère de Cauchy (4.15). En effet, soit $\varepsilon > 0$ arbitraire. Posons $\delta := \min(\varepsilon/M, b-a)$. Pour tous $x, x' \in]b-\delta, b[$, on a clairement $M(x'-x) < \varepsilon$, donc compte-tenu de (4.16), $\left|\int_x^{x'} f(t)\,\mathrm{d}t\right| < \varepsilon$.

Pour le point 2, il suffit de noter que si f a une limite finie ℓ à gauche en b, alors sur un intervalle $]b-\delta', b[$ suffisamment petit, on a $|f(t)| \leqslant |\ell| + 1$. Comme f est aussi bornée sur $[a, b-\delta']$ car Riemann intégrable sur ce segment, elle est bornée sur $[a,b[$, réunion des deux intervalles et on conclut en appliquant le point 1. □

Exemple 4.22. *L'intégrale généralisée* $\displaystyle\int_{-1/\pi}^{0} \sin\left(\frac{1}{t}\right) \mathrm{d}t$ *converge.*

En effet, $f : t \mapsto \sin(1/t)$ est définie et continue sur $[-1/\pi, 0[$, donc f appartient à $\mathcal{R}^{\mathrm{loc}}[-1/\pi, 0[$. Elle est bornée sur cet intervalle puisque $|\sin(1/t)| \leqslant 1$ pour tout $t \in \mathbb{R}^*$. Le point 1 du corollaire 4.21 donne la convergence de $\int_{-1/\pi}^{0} f(t)\,\mathrm{d}t$. Notons au passage que f n'a pas de limite à gauche en zéro, car elle oscille une infinité de fois entre les valeurs -1 et 1 sur tout voisinage à gauche de 0, aussi petit soit-il, cf. figure 4.7.

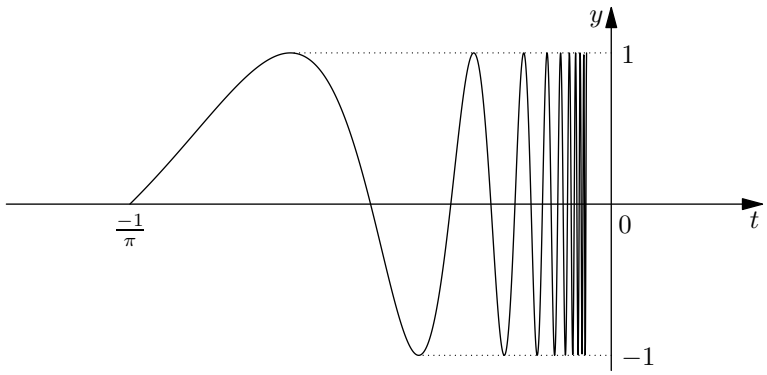


FIGURE 4.7 – Graphe de $t \mapsto \sin(1/t)$ pour $t \in \left[\frac{-1}{\pi}, \frac{-2}{39\pi}\right]$

Corollaire 4.23. *Soit* $f \in \mathcal{R}^{\mathrm{loc}}[a,b[$. *Si* $\int_a^b |f(t)|\,\mathrm{d}t$ *converge, alors* $\int_a^b f(t)\,\mathrm{d}t$ *converge.*

Preuve. Par le théorème 4.20, la convergence de $\int_a^b |f(t)|\,\mathrm{d}t$ implique le critère de Cauchy (4.14) si $b \in \mathbb{R}$ ou (4.15) si $b = +\infty$, avec $|f|$ à la place de f. L'inégalité

$$\left| \int_x^{x'} f(t)\,\mathrm{d}t \right| \leqslant \int_x^{x'} |f(t)|\,\mathrm{d}t$$

montre que le critère de Cauchy correspondant est aussi vérifié par f, d'où la convergence de $\int_a^b f(t)\,\mathrm{d}t$ par une nouvelle invocation du théorème 4.20. □

Remarque 4.24. La réciproque du corollaire 4.23 est fausse. Nous verrons un peu plus tard un contre exemple : $\int_1^{+\infty} \frac{\sin t}{t}\,\mathrm{d}t$ converge, mais $\int_1^{+\infty} \left|\frac{\sin t}{t}\right|\,\mathrm{d}t$ diverge.

Définition 4.25 (convergence absolue)**.** *Soit $f \in \mathcal{R}^{\mathrm{loc}}[a,b[$. Si $\int_a^b |f(t)|\,\mathrm{d}t$ converge, on dit que $\int_a^b f(t)\,\mathrm{d}t$ est* absolument convergente*.*

Une intégrale absolument convergente est toujours convergente (cor. 4.23), la réciproque est fausse (rem. 4.24).

Le théorème suivant permet entre autres de montrer la convergence d'intégrales de la forme $\int_a^{+\infty} f(t)\sin t\,\mathrm{d}t$ avec f positive décroissante et tendant vers 0 en $+\infty$. Sa preuve combine le critère de Cauchy et la deuxième formule de la moyenne que nous n'avons pas vue. Nous admettrons ce théorème.

Théorème 4.26 (critère d'Abel)**.** *Soient $f, g \in \mathcal{R}^{\mathrm{loc}}[a,+\infty[$ et vérifiant*

i) f est positive et décroissante sur $[a,+\infty[$ et a pour limite 0 en $+\infty$.

ii) Il existe une constante M telle que

$$\forall x, y \in [a,+\infty[, \quad \left| \int_x^y g(t)\,\mathrm{d}t \right| \leqslant M. \tag{4.17}$$

Alors $\int_a^{+\infty} f(t)g(t)\,\mathrm{d}t$ converge.

Si on prend en particulier $g(t) = \sin t$, il est facile de vérifier (4.17). En effet

$$\int_x^y \sin t\,\mathrm{d}t = \big[-\cos t\big]_x^y = \cos x - \cos y \quad \text{et} \quad |\cos x - \cos y| \leqslant 2.$$

Il en va de même avec $g(t) = \cos t$, ou $g(t) = \sin(ct)$, ou $g(t) = \cos(ct)$.

Énonçons séparément ce cas particulier du théorème 4.26 avant d'en proposer une démonstration directe.

Proposition 4.27. *Si f est positive et décroissante sur $[a,+\infty[$ et a pour limite 0 en $+\infty$, les intégrales généralisées $\int_a^{+\infty} f(t)\sin t\,\mathrm{d}t$ et $\int_a^{+\infty} f(t)\cos t\,\mathrm{d}t$ convergent.*

Preuve. Nous montrerons simplement la convergence de $\int_a^{+\infty} f(t)\sin t\,\mathrm{d}t$, l'adaptation de la méthode au cas de $\int_a^{+\infty} f(t)\cos t\,\mathrm{d}t$ étant immédiate. Remarquons d'abord que l'intégrande $h : t \mapsto f(t)\sin t$ est *localement* Riemann intégrable sur $[a,+\infty[$. En effet,

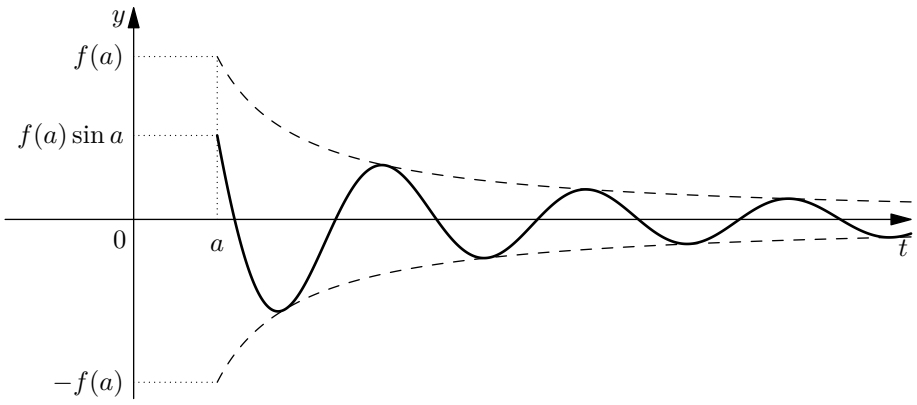


FIGURE 4.8 – Graphe de $t \mapsto f(t)\sin(t)$ avec $f \downarrow 0$, graphes de f et $-f$ en tirets

les restrictions des fonctions f et sin au segment $[\alpha, \beta] \subset [a, +\infty[$ sont respectivement monotone et continue donc Riemann intégrables. Leur produit h est donc aussi Riemann intégrable sur $[\alpha, \beta]$, cf. prop. 3.27. Ceci étant vrai pour tout $[\alpha, \beta] \subset [a, +\infty[$, h est dans $\mathcal{R}^{\mathrm{loc}}[a, +\infty[$.

Notons $I_k := [k\pi, (k+1)\pi]$, $k \in \mathbb{N}^*$. Pour k assez grand, disons $k \geqslant k_0$, $I_k \subset [a, +\infty[$. Aux bornes de I_k, $\sin t$ s'annule et pour tout t intérieur à I_k, $\sin t$ a même signe que $(-1)^k$. On a donc pour tout $t \in I_k$, $\sin t = (-1)^k |\sin t|$ d'où

$$\int_{k\pi}^{(k+1)\pi} f(t)\sin t \,\mathrm{d}t = (-1)^k \int_{k\pi}^{(k+1)\pi} f(t)|\sin t| \,\mathrm{d}t =: (-1)^k v_k. \tag{4.18}$$

On vérifie maintenant que la convergence de $\int_a^x h(t)\,\mathrm{d}t$ quand x tend vers $+\infty$, se réduit à la convergence de la série de terme général $(-1)^k v_k$. Pour $x > k_0\pi$, il existe un unique entier n tel que $(n+1)\pi \leqslant x < (n+2)\pi$. Cet entier dépendant de x tend évidemment vers $+\infty$ avec x. On peut alors écrire

$$\int_a^x h(t)\,\mathrm{d}t = \int_a^{k_0\pi} h(t)\,\mathrm{d}t + \int_{k_0\pi}^{(n+1)\pi} h(t)\,\mathrm{d}t + \int_{(n+1)\pi}^x h(t)\,\mathrm{d}t. \tag{4.19}$$

Traitons d'abord le terme « résiduel »

$$\varepsilon(x) := \int_{(n+1)\pi}^x h(t)\,\mathrm{d}t,$$

en notant que par la majoration $|\sin t| \leqslant 1$ et la décroissance de f,

$$|\varepsilon(x)| \leqslant \int_{(n+1)\pi}^x |h(t)|\,\mathrm{d}t \leqslant \int_{(n+1)\pi}^x f(t)\,\mathrm{d}t \leqslant \big(x-(n+1)\pi\big) f\big((n+1)\pi\big) \leqslant \pi f\big((n+1)\pi\big).$$

Or f tend vers 0 en $+\infty$ et $n = n(x)$ tend vers l'infini avec x, donc $\varepsilon(x)$ tend vers 0 en $+\infty$. En notant C la *constante* $\int_a^{k_0\pi} h(t)\,\mathrm{d}t$, en utilisant la relation de Chasles et (4.18), nous pouvons ainsi réécrire (4.19) sous la forme

$$\int_a^x h(t)\,\mathrm{d}t = C + \sum_{k=k_0}^{n} (-1)^k v_k + \varepsilon(x), \quad \lim_{x\to+\infty} \varepsilon(x) = 0.$$

Il est alors clair que la convergence en $+\infty$ de $\int_a^x h(t)\,\mathrm{d}t$ vers une limite finie équivaut à la convergence de la *série* de terme général $(-1)^k v_k$.

La convergence de cette série résultera du théorème des séries *alternées* (th. 2.18) si l'on montre que la suite $(v_k)_{k\geqslant k_0}$ tend vers 0 en décroissant.

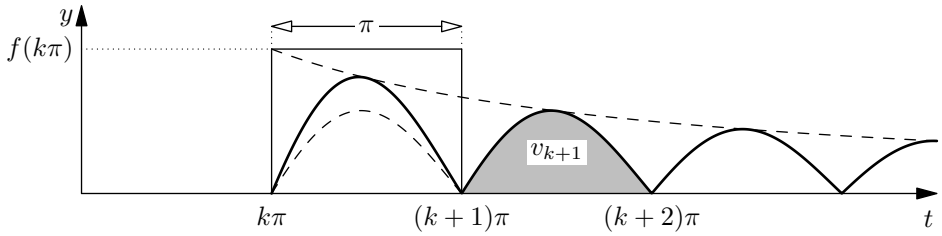


FIGURE 4.9 – $v_k \downarrow 0$ car $\pi f(k\pi) \geqslant v_k \geqslant v_{k+1}$

Les mêmes majorations que celles utilisées pour $\varepsilon(x)$ nous donnent

$$v_k = \int_{k\pi}^{(k+1)\pi} f(t)|\sin t|\,\mathrm{d}t \leqslant \pi f(k\pi) \xrightarrow[k\to+\infty]{} 0.$$

Pour vérifier la décroissance de $(v_k)_{k\geqslant k_0}$, comparons v_k et v_{k+1} grâce au changement de variable $s = t + \pi$ qui transforme l'intervalle I_{k+1} en I_k :

$$v_{k+1} = \int_{(k+1)\pi}^{(k+2)\pi} f(s)|\sin s|\,\mathrm{d}s = \int_{k\pi}^{(k+1)\pi} f(t+\pi)|\sin t|\,\mathrm{d}t.$$

Par décroissance de f on a pour tout $t \in I_k$, $f(t)|\sin t| \geqslant f(t+\pi)|\sin t|$ d'où par intégration de cette inégalité sur I_k, $v_k \geqslant v_{k+1}$. □

Exemple 4.28. *L'intégrale* $\displaystyle\int_1^{+\infty} \frac{\sin t}{t}\,\mathrm{d}t$ *est convergente, mais pas absolument.*

La convergence résulte de la proposition 4.27 avec $f(t) = 1/t$. Vérifions que l'intégrale n'est pas absolument convergente. Pour $x \geqslant \pi$, il existe un unique entier n tel que $(n+1)\pi \leqslant x < (n+2)\pi$ et cet entier dépendant de x tend vers $+\infty$ avec x. En posant $|h(t)| := t^{-1}|\sin t|$ et $v_k := \int_{k\pi}^{(k+1)\pi} |h(t)|\,\mathrm{d}t$, on a

$$\int_1^x |h(t)|\,\mathrm{d}t = \int_1^{\pi} |h(t)|\,\mathrm{d}t + \sum_{k=1}^{n} v_k + \int_{(n+1)\pi}^{x} |h(t)|\,\mathrm{d}t \geqslant \sum_{k=1}^{n} v_k.$$

Pour prouver que $\int_1^x |h(t)|\,\mathrm{d}t$ tend vers $+\infty$ avec x, il suffit donc de montrer la divergence de la série de terme général *positif* v_k. Ceci résulte de la minoration suivante qui utilise la décroissance de $t \mapsto 1/t$ et la π-périodicité de $|\sin t|$:

$$v_k = \int_{k\pi}^{(k+1)\pi} \frac{|\sin t|}{t}\,\mathrm{d}t \geqslant \int_{k\pi}^{(k+1)\pi} \frac{|\sin t|}{(k+1)\pi}\,\mathrm{d}t = \frac{1}{(k+1)\pi}\int_0^{\pi} \sin t\,\mathrm{d}t = \frac{2}{(k+1)\pi}.$$

Comme la série de terme général positif $u_k := \frac{2}{(k+1)\pi}$ diverge (cor. 2.15), on en déduit que

$$\sum_{k=1}^{n} v_k \geqslant \sum_{k=1}^{n} u_k \xrightarrow[n\to+\infty]{} +\infty.$$

Ainsi $\int_1^x |h(t)|\,\mathrm{d}t$ tend bien vers $+\infty$ avec x.

Nous avons choisi pour cet exemple d'intégrer entre 1 et $+\infty$, mais le résultat reste valable en intégrant la même fonction entre 0 et $+\infty$. La justification de la convergence en 0 est laissée en exercice.

4.3 Intégrales généralisées de fonctions positives

La fonction f définie sur $[a,b[$ est dite positive sur $[a,b[$ si pour tout $t \in [a,b[$, $f(t) \geqslant 0$. Nous énoncerons tous les résultats de cette section avec des fonctions positives sur $[a,b[$, mais il est clair qu'ils s'étendent au cas plus général des fonctions f définies sur $[a,b[$ et positives *au voisinage de* b, c'est-à-dire sur un intervalle $[c,b[$ pour un certain $c \in [a,b[$. Ils s'étendent aussi au cas des fonctions de signe constant au voisinage de b, modulo une adaptation laissée au lecteur.

Soit $f \in \mathcal{R}^{\mathrm{loc}}[a,b[$ et positive sur $[a,b[$. Alors la fonction F définie par

$$F(x) := \int_a^x f(t)\,\mathrm{d}t, \quad x \in [a,b[,$$

est *croissante* sur $[a,b[$. En effet si $a \leqslant x' \leqslant x'' < b$,

$$F(x'') - F(x') = \int_{x'}^{x''} f(t)\,\mathrm{d}t \geqslant 0, \tag{4.20}$$

par positivité de f sur $[x',x'']$. Il n'y a donc *que deux possibilités* pour le comportement de $F(x)$ quand x tend vers b à gauche.

1. La fonction croissante F est *majorée* sur $[a,b[$: il existe $M \in \mathbb{R}_+$ tel que pour tout $x \in [a,b[$, $F(x) \leqslant M < +\infty$. Alors F a une limite finie ℓ à gauche en b ($\ell \leqslant M$), autrement dit l'intégrale généralisée $\int_a^b f(t)\,\mathrm{d}t$ *converge* et $\int_a^b f(t)\,\mathrm{d}t = \ell$.
2. La fonction croissante F *n'est pas majorée* sur $[a,b[$. Alors F tend vers $+\infty$ à gauche en b, l'intégrale généralisée $\int_a^b f(t)\,\mathrm{d}t$ *diverge* et $\int_a^b f(t)\,\mathrm{d}t = +\infty$.

Rappelons ici qu'il existe des intégrales divergentes auxquelles on ne peut attribuer aucune valeur, pas même infinie, voir l'exemple 4.3. Comme nous venons de le voir, ce type de divergence ne peut se produire lorsque f est de signe constant.

Théorème 4.29 (comparaison)**.** *Soient* $f, g \in \mathcal{R}^{\mathrm{loc}}[a,b[$ *telles que* $0 \leqslant f \leqslant g$ *sur* $[a,b[$*. Alors*

$$\int_a^b g(t)\,\mathrm{d}t \text{ converge} \implies \int_a^b f(t)\,\mathrm{d}t \text{ converge}, \tag{4.21}$$

$$\int_a^b f(t)\,\mathrm{d}t \text{ diverge} \implies \int_a^b g(t)\,\mathrm{d}t \text{ diverge}. \tag{4.22}$$

Preuve. Soit x quelconque dans $[a,b[$. Alors f et g sont Riemann intégrables sur $[a,x]$ et en intégrant sur cet intervalle l'inégalité $f \leqslant g$, on voit que

$$\forall x \in [a,b[, \quad F(x) := \int_a^x f(t)\,\mathrm{d}t \leqslant \int_a^x g(t)\,\mathrm{d}t =: G(x).$$

Comme f et g sont positives, les fonctions F et G sont croissantes d'après (4.20).

Si $\int_a^x g(t)\,\mathrm{d}t$ converge, cela signifie que G a une limite à gauche finie L en b. La fonction *croissante* G est donc majorée sur $[a,b[$ par L et F l'est aussi puisque $F \leqslant G$ sur $[a,b[$. Étant croissante sur $[a,b[$ et majorée par $L < +\infty$ sur cet intervalle, F a aussi une limite à gauche finie $\ell \leqslant L$ en b. Autrement dit, $\int_a^b f(t)\,\mathrm{d}t$ converge. Ceci établit l'implication (4.21).

Si $\int_a^x f(t)\,\mathrm{d}t$ diverge, le point 2 de l'alternative ci-dessus entre en vigueur [8]. Autrement dit, $F(x)$ tend vers $+\infty$ à gauche en b. Il en va de même pour G puisque $F \leqslant G$. Donc $\int_a^b g(t)\,\mathrm{d}t = +\infty$ et cette intégrale diverge. L'implication (4.21) est ainsi vérifiée. □

Exemple 4.30. *Les intégrales « gaussiennes »* $\displaystyle\int_{-\infty}^{+\infty} \mathrm{e}^{-ct^2}\,\mathrm{d}t$*,* $c > 0$*, convergent.*

Par réduction du problème, cf. remarque 4.12, on se ramène à l'étude séparée de $\int_{-\infty}^0$ et de $\int_0^{+\infty}$. Par parité de l'intégrande $f : t \mapsto \exp(-ct^2)$, il est clair qu'il suffit d'étudier la convergence de l'intégrale généralisée de f sur $[0,+\infty[$. Notons au passage que f est continue sur $\mathbb{R}$, donc clairement membre de $\mathcal{R}^{\mathrm{loc}}]-\infty,0]$ et de $\mathcal{R}^{\mathrm{loc}}[0,+\infty[$. Puisque f est en particulier Riemann intégrable sur $[0,1]$, le découpage $\int_0^x f(t)\,\mathrm{d}t = \int_0^1 f(t)\,\mathrm{d}t + \int_1^x f(t)\,\mathrm{d}t$ nous ramène finalement à l'étude de la convergence de $\int_1^{+\infty} f(t)\,\mathrm{d}t$.

Cette dernière réduction est motivée par l'inégalité $t^2 \geqslant t$ pour $t \geqslant 1$. Par positivité de c et croissance de la fonction exponentielle, on en déduit $-ct^2 \leqslant -ct$ et $f(t) = \exp(-ct^2) \leqslant \exp(-ct) =: g(t)$. Nous pouvons alors appliquer l'implication (4.21) pour conclure à la convergence de $\int_1^{+\infty} f(t)\,\mathrm{d}t$. En effet

$$\int_1^x \exp(-ct)\,\mathrm{d}t = \left[\frac{\exp(-ct)}{-c}\right]_1^x = \frac{\exp(-c)}{c} - \frac{\exp(-cx)}{c} \xrightarrow[x\to+\infty]{} \frac{\exp(-c)}{c},$$

8. Parce que f est *positive* sur $[a,b[$.

donc $\int_1^{+\infty} g(t)\,\mathrm{d}t$ converge.

Corollaire 4.31. *Soient $f, g \in \mathcal{R}^{\mathrm{loc}}[a,b[$. On suppose que $|f| \leqslant g$ sur $[c,b[$ pour un $c \in [a,b[$ et que $\int_c^b g(t)\,\mathrm{d}t$ converge. Alors l'intégrale généralisée $\int_a^b f(t)\,\mathrm{d}t$ est absolument convergente.*

Preuve. Les fonctions f et g (donc aussi leur valeur absolue) sont Riemann intégrables sur $[a,c]$. Le découpage $\int_a^x = \int_a^c + \int_c^x$ montre alors que la convergence absolue de $\int_a^b f(t)\,\mathrm{d}t$ équivaut à celle de $\int_c^b f(t)\,\mathrm{d}t$. Cette dernière convergence découle immédiatement de celle de $\int_c^b g(t)\,\mathrm{d}t$ par (4.21) appliqué sur $[c,b[$ au lieu de $[a,b[$, avec $|f|$ à la place de f. □

Exemple 4.32. *L'intégrale* $\displaystyle\int_1^{+\infty} \frac{\sin(t\cos t)}{t^2}\,\mathrm{d}t$ *converge absolument.*

C'est une application immédiate du corollaire 4.31 avec $a = c = 1$ et $g(t) = t^{-2}$.

Théorème 4.33 (intégrandes équivalentes)**.** *Soient $f, g \in \mathcal{R}^{\mathrm{loc}}[a,b[$, positives au voisinage à gauche de b. Si elles sont équivalentes en $b-$, $\int_a^b f(t)\,\mathrm{d}t$ et $\int_a^b g(t)\,\mathrm{d}t$ sont de même nature.*

Preuve. Rappelons que « f équivalente à g en $b-$ » noté encore $f \underset{b-}{\sim} g$ signifie qu'il existe un réel $c \in [a,b[$ et une fonction h définie que $[c,b[$ telle que

$$\forall t \in [c,b[, \quad f(t) = g(t)h(t) \quad \text{et} \quad \lim_{t\to b-} h(t) = 1.$$

Cette limite à gauche de h en b nous permet de trouver un $d \in [c,b[$ tel que l'encadrement $1/2 \leqslant h(t) \leqslant 3/2$ soit vérifié [9] pour tout $t \in [d,b[$. Par positivité de f et g au voisinage de b et quitte à remplacer d par $d' \in [d,b[$, on se ramène au cas où g est positive sur $[d,b[$. On a alors

$$\forall t \in [d,b[, \quad 0 \leqslant \frac{g(t)}{2} \leqslant f(t) = g(t)h(t) \leqslant \frac{3g(t)}{2}, \tag{4.23}$$

d'où l'on tire

$$\forall x \in [d,b[, \qquad \frac{1}{2}\int_d^x g(t)\,\mathrm{d}t \leqslant \int_d^x f(t)\,\mathrm{d}t \leqslant \frac{3}{2}\int_d^x g(t)\,\mathrm{d}t. \tag{4.24}$$

Supposons que $\int_a^b g(t)\,\mathrm{d}t$ diverge. Comme $\int_a^d g(t)\,\mathrm{d}t$ est une constante finie, $\int_d^b g(t)\,\mathrm{d}t$ diverge aussi. Par *positivité* de g sur $[d,b[$, $\int_d^x g(t)\,\mathrm{d}t$ tend alors vers $+\infty$ quand x tend vers b à gauche. Il en va de même pour $\int_d^x f(t)\,\mathrm{d}t$ à cause de la première inégalité dans (4.24). Par addition de la constante $\int_a^d f(t)\,\mathrm{d}t$ on voit finalement que $\int_a^x f(t)\,\mathrm{d}t$ tend vers $+\infty$ en $b-$. Ainsi la divergence de $\int_a^b g(t)\,\mathrm{d}t$ implique celle de $\int_a^b f(t)\,\mathrm{d}t$.

9. Appliquer la définition de la limite avec $\varepsilon = 1/2$.

Si $\int_a^b g(t)\,\mathrm{d}t$ converge, l'ensemble $\{\int_d^x g(t)\,\mathrm{d}t;\ x \in [d,b[\,\}$ est majoré et la deuxième inégalité dans (4.24), montre qu'il en va de même pour $\{\int_d^x f(t)\,\mathrm{d}t;\ x \in [d,b[\,\}$. On en déduit facilement que $\int_a^b f(t)\,\mathrm{d}t$ converge en utilisant la *positivité* sur $[d,b[$ de f qui résulte de (4.23). □

Corollaire 4.34. *Soient $f, g \in \mathcal{R}^{\mathrm{loc}}[a,b[$ telles que g soit* strictement positive *sur un voisinage à gauche de b et que*

$$\lim_{t\to b-} \frac{f(t)}{g(t)} = K, \quad K \in]0,+\infty[. \tag{4.25}$$

Alors $\int_a^b f(t)\,\mathrm{d}t$ converge si et seulement si $\int_a^b g(t)\,\mathrm{d}t$ converge.

Preuve. La positivité stricte de g sur un voisinage à gauche de b et (4.25) impliquent que f et Kg sont toutes deux positives sur un même voisinage $[c,b[$ de b et que f et Kg sont équivalentes en $b-$. On conclut en appliquant le théorème 4.33 à f et Kg. □

Exemple 4.35. $I := \displaystyle\int_0^1 t^\alpha (1-t)^\beta\,\mathrm{d}t$ *converge si et seulement si* $\alpha > -1$ et $\beta > -1$.

L'intégrande $f : t \mapsto t^\alpha(1-t)^\beta$ est toujours continue au moins sur $]0,1[$, donc f est localement Riemann intégrable sur $]0,1[$.

Pour $\alpha \geqslant 0$ et $\beta \geqslant 0$, f est continue sur $[0,1]$ donc Riemann intégrable sur $[0,1]$. I est alors une intégrale ordinaire. Si $\alpha < 0$ ou $\beta < 0$, I est une intégrale généralisée. Pour étudier sa convergence, on regarde séparément $\int_0^{1/2}$ et $\int_{1/2}^1$. Notons que f est strictement positive sur $]0,1[$, ce qui nous permet d'utiliser le théorème 4.33. On voit ainsi que

$$t^\alpha(1-t)^\beta \underset{0+}{\sim} t^\alpha \quad \text{et} \quad t^\alpha(1-t)^\beta \underset{1-}{\sim} (1-t)^\beta.$$

Compte-tenu du corollaire 4.15, on en déduit que I converge si et seulement si $\alpha > -1$ et $\beta > -1$.

Exemple 4.36. $I := \displaystyle\int_0^{+\infty} \frac{\mathrm{d}t}{1+t^3}$ *converge.*

On a envie de dire que c'est une application immédiate du théorème 4.33 puisque $f(t) := (1+t)^{-3} \simeq t^{-3} =: g(t)$ en $+\infty$. Mais alors on introduit artificiellement un problème en zéro pour g et $\int_0^{+\infty} g(t)\,\mathrm{d}t$ diverge (à cause de la borne 0). En y regardant de plus près, on voit que les hypothèses du théorème 4.33 ne sont pas toutes vérifiées puisque g est localement intégrable sur $]0,+\infty[$, *mais pas* sur $[0,+\infty[$. Notons que f elle, est bien dans $\mathcal{R}^{\mathrm{loc}}[0,+\infty[$ comme fonction continue sur $[0,+\infty[$. On se sort de ce mauvais pas en remarquant que I et $\int_1^{+\infty} f(t)\,\mathrm{d}t$ sont de même nature et en appliquant le théorème 4.33 sur l'intervalle $[1,+\infty[$ avec les restrictions de f et g à cet intervalle. En effet $\int_1^{+\infty} t^{-3}\,\mathrm{d}t$ converge par la proposition 4.14 a).

On aurait pu aussi prouver par cette méthode la convergence de $\int_0^{+\infty} \frac{\mathrm{d}t}{1+t^2}$, mais dans ce cas, il y a bien plus simple puisque $\int_0^x \frac{\mathrm{d}t}{1+t^2} = \arctan(x)$ qui tend vers $\pi/2$ lorsque x tend vers $+\infty$. Donc on voit directement que $\int_0^{+\infty} \frac{\mathrm{d}t}{1+t^2}$ converge et vaut $\pi/2$.

Exemple 4.37. $I := \displaystyle\int_0^{+\infty} \frac{3t+2}{5t^3+t^2+4}\,\mathrm{d}t$ *converge.*

L'intégrande $f : t \mapsto (3t+2)(5t^3+t^2+4)^{-1}$ est positive et continue sur $[0,+\infty[$ comme quotient de deux fonctions continues car le dénominateur qui est minoré par 4 sur cet intervalle ne s'y annule pas. Ainsi f appartient à $\mathcal{R}^{\mathrm{loc}}[0,+\infty[$. D'autre part en $+\infty$, $f(t)$ est équivalent à $\frac{3}{5}t^{-2} =: g(t)$. On est confronté au même piège qu'à l'exemple 4.36 et il faut éviter d'introduire artificiellement un problème en 0 à cause de la divergence de $\int_0^{\varepsilon} g(t)\,\mathrm{d}t$. Là encore il suffit de se ramener à $\int_1^{+\infty} f(t)\,\mathrm{d}t$ qui converge par comparaison avec $\int_1^{+\infty} t^{-2}\,\mathrm{d}t$.

Exemple 4.38. $I := \displaystyle\int_0^{\frac{\pi}{2}} \frac{\mathrm{d}t}{(\cos t)^{\alpha}}$ *converge si et seulement si* $\alpha < 1$.

Sur $[0,\pi/2]$ la fonction continue cosinus ne s'annule qu'au point $\pi/2$ et est positive ailleurs. Pour $\alpha \leqslant 0$, l'intégrande $f : t \mapsto (\cos t)^{-\alpha}$ est continue sur $[0,\pi/2]$ et I est une intégrale de Riemann ordinaire. Pour $\alpha > 0$, f est continue sur $[0,\pi/2[$ et tend vers $+\infty$ à gauche en $\pi/2$. Dans ce cas $f \in \mathcal{R}^{\mathrm{loc}}[0,\pi/2[$ et I est une véritable intégrale généralisée. Un développement limité à l'ordre 1 du cosinus au point $\pi/2$ s'écrit :

$$\cos t = \cos\frac{\pi}{2} + \Big(-\sin\frac{\pi}{2}\Big)\Big(t-\frac{\pi}{2}\Big) + \Big(t-\frac{\pi}{2}\Big)\varepsilon\Big(t-\frac{\pi}{2}\Big), \quad \varepsilon(u) \xrightarrow[u\to 0]{} 0,$$

d'où

$$\cos t = \Big(\frac{\pi}{2}-t\Big)\Big(1-\varepsilon\Big(t-\frac{\pi}{2}\Big)\Big),$$

autrement dit, $\cos t$ a pour équivalent $\pi/2 - t$ en $\pi/2$. On en déduit que

$$f(t) \underset{\frac{\pi}{2}-}{\sim} \Big(\frac{\pi}{2}-t\Big)^{-\alpha} =: g(t).$$

La fonction g étant comme f, continue et positive sur $[0,\pi/2[$, le théorème 4.33 s'applique et nous dit que I et $\int_0^{\pi/2}(\pi/2-t)^{-\alpha}\,\mathrm{d}t$ sont de même nature. Cette dernière intégrale converge si et seulement si $\alpha < 1$ par le corollaire 4.15.

Remarque 4.39. Le théorème 4.33 s'adapte immédiatement au cas où f et g sont *toutes deux négatives* au voisinage à gauche de b. En revanche, le théorème n'est plus valable si f n'a pas un *signe constant* au voisinage à gauche de b, même si f et g sont de même signe au voisinage de $b-$. Le contre exemple suivant devrait vous en convaincre.

Exemple 4.40 (à méditer)**.** *Définissons* $f, g : [1,+\infty[\to \mathbb{R}$ *par*

$$f(t) := \frac{\sin t}{\sqrt{t}}, \qquad g(t) := \frac{\sin t}{\sqrt{t}} + \frac{\sin^2 t}{t}. \tag{4.26}$$

Alors $f(t) \sim g(t)$ *en* $+\infty$*,* f *change de signe une infinité de fois au voisinage de* $+\infty$*,* f *et* g *sont de même signe au voisinage de* $+\infty$*. L'intégrale généralisée* $\int_1^{+\infty} f(t)\,\mathrm{d}t$ *converge mais* $\int_1^{+\infty} g(t)\,\mathrm{d}t$ *diverge.*

Justifications. Les fonctions f et g sont continues sur $[1,+\infty[$ donc dans $\mathcal{R}^{\mathrm{loc}}[1,+\infty[$. On note d'abord que pour tout $t\in[1,+\infty[$, $g(t)=f(t)h(t)$ avec

$$g(t)=f(t)h(t)\quad\text{avec}\quad h(t)=1+\frac{\sin t}{\sqrt{t}}\xrightarrow[t\to+\infty]{}1. \tag{4.27}$$

Ceci établit l'équivalence de f et g en $+\infty$.

La fonction f ayant le signe du sinus change de signe une infinité de fois au voisinage de $+\infty$. Il en va de même pour g à cause de (4.27) puisque $h(t)$ est strictement positif[10] sur $[1,+\infty[$. De plus, f et g ont même signe et mêmes zéros sur tout l'intervalle $[1,+\infty[$.

L'intégrale généralisée $\int_1^{+\infty} f(t)\,\mathrm{d}t$ converge par le théorème d'Abel ou la proposition 4.27.

Supposons que $\int_1^{+\infty} g(t)\,\mathrm{d}t$ converge, alors nécessairement $\int_1^{+\infty} t^{-1}\sin^2 t\,\mathrm{d}t$ doit converger. En effet

$$\int_1^x \frac{\sin^2 t}{t}\,\mathrm{d}t=\int_1^x g(t)\,\mathrm{d}t-\int_1^x f(t)\,\mathrm{d}t$$

et le second membre doit avoir une limite finie quand x tend vers $+\infty$ en raison de la convergence des deux intégrales $\int_1^{+\infty} f(t)\,\mathrm{d}t$ et $\int_1^{+\infty} g(t)\,\mathrm{d}t$. Nous allons montrer que l'on aboutit à une *contradiction* en vérifiant *directement* que $\int_1^{+\infty} t^{-1}\sin^2 t\,\mathrm{d}t$ *diverge*. En effet, l'identité $\sin^2 t=\frac{1}{2}(1-\cos 2t)$ nous donne

$$\int_1^x \frac{\sin^2 t}{t}\,\mathrm{d}t=\int_1^x\Big(\frac{1}{2t}-\frac{\cos 2t}{2t}\Big)\,\mathrm{d}t=\frac{1}{2}\ln x-\frac{1}{2}\int_1^x\frac{\cos 2t}{t}\,\mathrm{d}t.$$

Quand x tend vers $+\infty$, $\frac{1}{2}\ln x$ tend vers $+\infty$, tandis que $\frac{1}{2}\int_1^x \frac{\cos 2t}{t}\,\mathrm{d}t$ tend vers une limite finie car $\frac{1}{2}\int_1^{+\infty} \frac{\cos 2t}{t}\,\mathrm{d}t$ converge grâce au théorème d'Abel ou à la proposition 4.27 (poser $s=2t$). Donc $\int_1^x t^{-1}\sin^2 t\,\mathrm{d}t$ tend vers $+\infty$ avec x, autrement dit $\int_1^{+\infty} t^{-1}\sin^2 t\,\mathrm{d}t$ diverge, ce qui établit la contradiction annoncée et impose la divergence de $\int_1^{+\infty} g(t)\,\mathrm{d}t$. □

4.4 Divers

4.4.1 Changements de variable

Nous examinons l'extension des formules de changement de variable au cas des intégrales généralisées. *Grosso modo* tout se passe bien lorsque l'on utilise un changement de variable *monotone*. Si ce n'est pas le cas, il convient d'être prudent et de revenir à la définition de l'intégrale généralisée $\int_a^b=\lim_{x\to b-}\int_a^x$ pour appliquer le changement de variable aux intégrales de Riemann ordinaires $\int_a^x$ avant de faire tendre x vers $b-$.

10. Il suffirait que h soit strictement positive sur un voisinage $[c,+\infty[$ de $+\infty$, ce qui découle du fait que h a une limite strictement positive en $+\infty$. Mais ici il est plus simple de remarquer que $t^{-1/2}\sin t>-1$ pour $t>1$ et que $h(1)=1+\sin 1>0$.

Proposition 4.41 (translation et changement d'échelle)**.**

i) Translation. Soient $c \in \mathbb{R}$ et f localement Riemann intégrable sur $[a+c, b+c[$, avec $b+c := +\infty$ si $b = +\infty$. Alors l'application $g : [a,b[\to \mathbb{R}$, $t \mapsto f(t+c)$ est localement Riemann intégrable sur $[a,b[$. Les intégrales $\int_a^b g(t)\,\mathrm{d}t$ et $\int_{a+c}^{b+c} f(s)\,\mathrm{d}s$ sont de même nature. Si l'une des deux converge on a

$$\int_a^b f(t+c)\,\mathrm{d}t = \int_{a+c}^{b+c} f(s)\,\mathrm{d}s. \tag{4.28}$$

Cette égalité reste vraie dans $\overline{\mathbb{R}}_+$ sans hypothèse de convergence si f ou g est positive sur son intervalle d'intégration.

ii) Changement d'échelle. Soient $c \in \mathbb{R}^*$ et f localement Riemann intégrable sur l'intervalle d'extrémités[11] *ac et bc, semi-fermé en ac, avec dans le cas où $b = +\infty$, $bc := +\infty$ si $c > 0$, $bc := -\infty$ si $c < 0$. Alors l'application $h : [a,b] \to \mathbb{R}$, $t \mapsto f(ct)$ est localement Riemann intégrable sur $[a,b[$. Les intégrales $\int_a^b h(t)\,\mathrm{d}t$ et $\int_{ac}^{bc} f(s)\,\mathrm{d}s$ sont de même nature. Si l'une des deux converge on a*

$$\int_a^b f(ct)\,\mathrm{d}t = \frac{1}{c}\int_{ac}^{bc} f(s)\,\mathrm{d}s. \tag{4.29}$$

Cette égalité reste vraie dans $\overline{\mathbb{R}}_+$ sans hypothèse de convergence si f ou h est positive sur son intervalle d'intégration.

Remarquons qu'il n'y a ici aucune hypothèse de continuité sur f pour ces formules de changement de variable par translation ou changement d'échelle dans les intégrales généralisées. On peut donc les appliquer notamment avec des fonctions décroissantes positives qui peuvent avoir une infinité de discontinuités, mais sont toujours localement Riemann intégrables.

Preuve de i). Puisque $f \in \mathcal{R}^{\mathrm{loc}}[a+c, b+c[$, elle est Riemann intégrable sur le segment $[a+c, x+c]$ pour tout $x \in [a,b[$. Alors par la proposition 3.33 i), g est Riemann intégrable sur $[a,x]$ et ceci valant pour tout $x \in [a,b[$, g est bien dans $\mathcal{R}^{\mathrm{loc}}[a,b[$. De plus on a par la formule de changement de variable (3.50) :

$$\forall x \in [a,b[, \quad \int_a^x g(t)\,\mathrm{d}t = \int_a^x f(t+c)\,\mathrm{d}t = \int_{a+c}^{x+c} f(s)\,\mathrm{d}s.$$

En faisant tendre x vers $b-$, on en déduit que $\int_a^b g(t)\,\mathrm{d}t$ et $\int_{a+c}^{b+c} f(s)\,\mathrm{d}s$ sont de même nature. Si l'une des deux intégrales généralisées converge, cela signifie que l'intégrale de Riemann ordinaire correspondante ci-dessus a une limite dans $\mathbb{R}$ quand x tend vers $b-$. En raison de l'égalité, il en va de même pour l'autre intégrale et les limites sont égales, ce qui nous donne (4.28). Si f ou g est positive, ces deux intégrales dépendant de x ont toujours une limite dans $\overline{\mathbb{R}}_+$ et les limites sont égales. □

11. Il s'agit de $[ac, bc[$ si $c > 0$ et de $]bc, ac]$ si $c < 0$.

Preuve de ii). La preuve est analogue à celle de i) à quelques alourdissements d'écriture près que l'auteur abandonne lâchement au lecteur. □

Voici maintenant une extension partielle [12] aux intégrales généralisées du changement de variable « classique » de la proposition 3.33 iii).

Proposition 4.42 (changement de variable C^1 monotone)**.** *Soit $\varphi : [a,b[\to \mathbb{R}$, une fonction* monotone *ayant une dérivée continue sur $[a,b[$. On suppose de plus que φ est* strictement *monotone au voisinage à gauche de b. Pour toute fonction f* continue *sur l'intervalle $\varphi([a,b[)$, les deux intégrales généralisées ci-dessous sont de même nature et si l'une converge on a*

$$\int_a^b f\big(\varphi(t)\big)\varphi'(t)\,\mathrm{d}t = \int_{\varphi(a)}^{\varphi(b-)} f(s)\,\mathrm{d}s, \tag{4.30}$$

où $\varphi(b-) \in \overline{\mathbb{R}}$ désigne la limite à gauche de φ en b. Cette égalité demeure vraie dans $\overline{\mathbb{R}}$ sans condition de convergence si f est de signe constant sur l'intervalle $\varphi([a,b[)$.

Notons $h := (f\circ\varphi)\varphi'$. La condition sur la stricte monotonie de φ à gauche de b est là pour écarter un cas artificiel où l'intégrale $\int_a^b h(t)\,\mathrm{d}t$ est une intégrale de Riemann ordinaire d'un fonction continue sur un intervalle $[a,c]$. En effet si φ est constante sur $[c,b[$ pour un $c \in [a,b[$, on voit que $\int_a^x h(t)\,\mathrm{d}t = \int_a^c h(t)\,\mathrm{d}t$ pour tout $x \in [c,b[$ en raison de la nullité de φ' sur $[c,b[$. En faisant tendre x vers $b-$, cette égalité nous donne $\int_a^b h(t)\,\mathrm{d}t = \int_a^c h(t)\,\mathrm{d}t$. Comme h est continue sur $[a,c]$, le changement de variable classique prop. 3.33 iii) nous donne $\int_a^c h(t)\,\mathrm{d}t = \int_{\varphi(a)}^{\varphi(c)} f(s)\,\mathrm{d}s$ et comme $\varphi(b-) = \varphi(c)$, on obtient bien (4.30). On voit ainsi que dans ce cas les deux intégrales dans (4.30) sont de fausses intégrales généralisées puisqu'elles peuvent s'écrire comme intégrales de fonctions continues sur $[a,c]$.

Preuve de la proposition 4.42. La fonction φ étant monotone admet une limite à gauche finie ou infinie en b que nous notons $\varphi(b-)$. En raison de la continuité et de la monotonie de φ sur $[a,b[$, stricte au voisinage de $b-$, $\varphi([a,b[)$ est l' intervalle de bornes $\varphi(a)$ et $\varphi(b-)$, fermé en $\varphi(a)$ et ouvert en $\varphi(b-)$. Nous traitons le cas où φ est décroissante, l'adaptation au cas où elle est croissante étant immédiate. On a alors $\varphi([a,b[) =]\varphi(b-),\varphi(a)]$. Par application de la prop. 3.33 iii), on a

$$\forall x \in [a,b[,\quad \int_a^x f\big(\varphi(t)\big)\varphi'(t)\,\mathrm{d}t = \int_{\varphi(a)}^{\varphi(x)} f(s)\,\mathrm{d}s = -\int_{\varphi(x)}^{\varphi(a)} f(s)\,\mathrm{d}s. \tag{4.31}$$

Les fonctions h et f étant continues l'une sur $[a,b[$ et l'autre sur $]\varphi(b-),\varphi(a)]$ appartiennent respectivement à $\mathcal{R}^{\mathrm{loc}}[a,b[$ et $\mathcal{R}^{\mathrm{loc}}]\varphi(b-),\varphi(a)]$. En faisant tendre x vers $b-$ dans (4.31), et en notant que par continuité et décroissance de φ, $\varphi(x)$ tend alors vers $\varphi(b-)$ par la droite, on voit que les intégrales généralisées $\int_a^b h(t)\,\mathrm{d}t$ et $\int_{\varphi(a)}^{\varphi(b-)} f(s)\,\mathrm{d}s$

12. On notera l'hypothèse plus restrictive sur le changement de variable φ.

sont de même nature. Si l'une des deux converge, on en déduit en se souvenant de la définition 4.11 :

$$\int_a^b f\big(\varphi(t)\big)\varphi'(t)\,\mathrm{d}t = -\int_{\varphi(b-)}^{\varphi(a)} f(s)\,\mathrm{d}s = \int_{\varphi(a)}^{\varphi(b-)} f(s)\,\mathrm{d}s,$$

ce qui nous donne (4.30). D'autre part si f est positive, h est négative car φ est décroissante donc $\varphi' \leqslant 0$. Alors dans (4.31) les intégrales $\int_a^x h(t)\,\mathrm{d}t$ et $-\int_{\varphi(x)}^{\varphi(a)} f(s)\,\mathrm{d}s$ sont des fonctions négatives et décroissantes de x donc convergent dans $\overline{\mathbb{R}}$ quand x tend vers $b-$, soit vers un réel négatif, soit vers $-\infty$ et l'égalité (4.31) se conserve par passage à la limite. Si f est négative, h est positive car $\varphi' \leqslant 0$. Alors les intégrales de (4.31) sont des fonctions positives et croissantes de la variable x et elles restent égales à la limite (dans $\overline{\mathbb{R}}_+$) quand x tend vers $b-$. □

Exemple 4.43. *L'intégrale généralisée* $I := \displaystyle\int_0^{+\infty} \sin(t^2)\,\mathrm{d}t$ *converge.*

Ce résultat semble surprenant de prime abord, surtout au vu de l'exemple 4.3. Par le changement de variable croissant et C^1, $\varphi : t \mapsto t^2$, l'intervalle $[0, +\infty[$ a pour image $[0, +\infty[$ et I est de même nature que

$$J := \int_0^{+\infty} \frac{\sin s}{2\sqrt{s}}\,\mathrm{d}s.$$

L'intégrale J converge par la proposition 4.27, donc I converge. De plus on a alors $I = J$ par l'égalité (4.30).

Exemple 4.44 (un changement de variable illicite)**.** Voici un exemple où un changement de variable C^1 non monotone appliqué sans précaution conduit à une erreur. Dans l'intégrale généralisée $I := \int_0^{2\pi} \frac{\sin t}{\cos t}\,\mathrm{d}t$, on pose $s = \cos t$. On obtient alors l'intégrale de Riemann ordinaire $J := \int_1^1 \frac{-\,\mathrm{d}s}{s} = 0$. L'égalité $I = J$ ne peut être valide ici car I diverge. En effet la fonction tangente est localement Riemann intégrable sur l'intervalle troué $[0, 2\pi]\backslash\{\pi/2, 3\pi/2\}$ et nous devons considérer séparément chacune des intégrales $\int_0^{\pi/2}$, $\int_{\pi/2}^{3\pi/2}$ et $\int_{3\pi/2}^{2\pi}$. L'intégrale $\int_0^{\pi/2} \tan t\,\mathrm{d}t$ diverge car la fonction tangente est positive sur $[0, \pi/2[$ et équivalente à $1/\cos t$ au voisinage à gauche de $\pi/2$. Or on sait par l'exemple 4.38 que $\int_0^{\pi/2} \frac{\mathrm{d}t}{\cos t}$ diverge [13].

4.4.2 Intégration par parties

Il n'y a *pas d'extension automatique de la règle d'intégration par parties (i.p.p.)* des intégrales de Riemann ordinaires (calculables par primitivation) aux intégrales généralisées. Lorsque l'on effectue une intégration par parties sur une intégrale généralisée, tout peut arriver :

13. On pourrait contester cet exemple, car dans le contexte de la proposition 4.42, avant d'envisager un changement de variable dans l'intégrale I, il convient de vérifier que l'intégrande est localement intégrable sur $[0, 2\pi[$, ce qui n'est pas le cas ici.

1. transformation d'une intégrale absolument convergente en intégrale absolument convergente ;
2. transformation d'une intégrale absolument convergente en intégrale convergente mais pas absolument et *vice versa* ;
3. transformation d'une intégrale absolument convergente en intégrale divergente.

En pratique pour effectuer une intégration par parties sur $\int_a^b f(t)\,\mathrm{d}t$ avec $f \in \mathcal{R}^{\mathrm{loc}}[a,b[$ (en fait avec $f \in C[a,b[$), on l'effectue d'abord sur l'intégrale de Riemann ordinaire $\int_a^x f(t)\,\mathrm{d}t$ pour x quelconque dans $[a,b[$ avant de regarder ce qui se passe lorsque l'on fait tendre x vers $b-$.

Voici quelques exemples illustrant les différentes situations possibles.

Exemple 4.45. *Calcul de* $I := \displaystyle\int_0^{+\infty} t\mathrm{e}^{-t}\,\mathrm{d}t$ *par i.p.p.*

L'intégrande est une fonction positive et continue sur $[0,+\infty[$. Comme $t\mathrm{e}^{-t/2}$ tend vers 0 en $+\infty$, cette quantité est majorée pour $t \geqslant t_0$ par une constante M. Alors pour $t \geqslant t_0$, $t\mathrm{e}^{-t} \leqslant M\mathrm{e}^{-t/2} =: g(t)$ et comme $\int_0^{+\infty} g(t)\,\mathrm{d}t$ converge (évident par primitivation), le théorème de comparaison (th. 4.29) nous donne la convergence de I.

On effectue l'i.p.p. sur $I(x) := \int_0^x t\mathrm{e}^{-t}\,\mathrm{d}t$ en posant

$$u(t) = t,\ v'(t) = \mathrm{e}^{-t}, \qquad u'(t) = 1,\ v(t) = -\mathrm{e}^{-t},$$

d'où

$$\forall x \in [0,+\infty[, \quad I(x) = \left[-t\mathrm{e}^{-t}\right]_0^x - \int_0^x (-\mathrm{e}^{-t})\,\mathrm{d}t = x\mathrm{e}^{-x} + \int_0^x \mathrm{e}^{-t}\,\mathrm{d}t.$$

Quand x tend vers $+\infty$, $x\mathrm{e}^{-x}$ tend vers 0 et $\int_0^x \mathrm{e}^{-t}\,\mathrm{d}t$ tend vers l'intégrale généralisée convergente $\int_0^{+\infty} \mathrm{e}^{-t}\,\mathrm{d}t = 1$, cf.exemple 4.1. Ici l'i.p.p. permet de retrouver la convergence de I et de calculer sa valeur : $I = 1$.

Exemple 4.46. *Intégration par parties de* $I := \displaystyle\int_\pi^{+\infty} \frac{\sin^2 t}{t^2}\,\mathrm{d}t$.

Ici l'intégrande f est localement Riemann intégrable sur $[\pi,+\infty[$, comme fonction continue sur cet intervalle. La convergence de I résulte de l'inégalité $0 \leqslant f(t) \leqslant t^{-2}$ et de la convergence de $\int_\pi^{+\infty} t^{-2}\,\mathrm{d}t$, par le théorème de comparaison. Ainsi I est une intégrale généralisée absolument convergente.

On effectue l'i.p.p. sur $I(x) := \int_0^x f(t)\,\mathrm{d}t$ en posant

$$u(t) = \sin^2 t,\ v'(t) = t^{-2}, \qquad u'(t) = 2\sin t\cos t = \sin(2t),\ v(t) = -t^{-1},$$

ce qui nous donne

$$\forall x \geqslant \pi, \quad I(x) = \left[-\frac{\sin^2 t}{t}\right]_\pi^x + \int_\pi^x \frac{\sin(2t)}{t}\,\mathrm{d}t = -\frac{\sin^2 x}{x} + \int_{2\pi}^{2x} \frac{\sin s}{s}\,\mathrm{d}s.$$

Faisons tendre x vers $+\infty$, alors au premier membre $I(x)$ tend vers I puisque l'on sait déjà que I converge. Au second membre $-x^{-1}\sin^2 x$ tend vers 0. On en déduit que $J(x) := \int_{2\pi}^{2x} s^{-1}\sin s\,\mathrm{d}s$ tend vers une limite finie égale à I. Ceci prouve que l'intégrale généralisée

$$J := \int_{2\pi}^{+\infty} \frac{\sin s}{s}\,\mathrm{d}s$$

converge et est égale à I. On sait par ailleurs que J n'est pas absolument convergente, voir l'exemple 4.28. Ici l'intégration par parties a transformé une intégrale absolument convergente I en une intégrale J convergente mais pas absolument. Une autre i.p.p. partant de J donnerait

$$J = \frac{1}{2\pi} - \int_{2\pi}^{+\infty} \frac{\cos s}{s^2}\,\mathrm{d}s,$$

nous fournissant un exemple de transformation d'une intégrale convergente mais pas absolument, en intégrale absolument convergente.

Exemple 4.47. *Tentative d'i.p.p. sur* $I := \displaystyle\int_0^{\pi/2} \frac{\sin t}{t^{3/2}}\,\mathrm{d}t$

L'intégrande f est dans $\mathcal{R}^{\mathrm{loc}}]0,\pi/2]$ et positive. L'intégrale converge en 0 grâce à la majoration $0 \leqslant \sin t \leqslant t$ d'où $f(t) \leqslant t^{-1/2}$ valable pour tout $t \in]0,\pi/2]$. L'intégrale généralisée I est donc absolument convergente. Intégrons par parties $I(x) := \int_x^{\pi/2} f(t)\,\mathrm{d}t$ en posant :

$$u(t) = t^{-3/2},\ v'(t) = \sin t, \qquad u'(t) = -\frac{3}{2}t^{-5/2},\ v(t) = -\cos t,$$

d'où

$$I(x) = \left[\frac{-\cos t}{t^{3/2}}\right]_x^{\pi/2} - \frac{3}{2}\int_x^{\pi/2} \frac{\cos t}{t^{5/2}}\,\mathrm{d}t = \frac{\cos x}{x^{3/2}} - \frac{3}{2}\int_x^{\pi/2} \frac{\cos t}{t^{5/2}}\,\mathrm{d}t.$$

Si on fait tendre x vers 0+, on obtient une forme indéterminée du type « $\infty - \infty$ », car l'intégrale généralisée

$$J := \int_0^{\pi/2} \frac{\cos t}{t^{5/2}}\,\mathrm{d}t$$

est divergente et vaut $+\infty$ (justifiez !). On a là un exemple d'une intégration par parties sur une intégrale absolument convergente qui fait apparaître une intégrale divergente.

Ceci dit l'i.p.p. ci-dessus n'est pas complètement inutile. Elle permet en effet de donner une « vitesse de divergence » de J. En effet puisque $I(x)$ a une limite finie I en 0+, on en déduit que

$$J(x) = \int_x^{\pi/2} \frac{\cos t}{t^{5/2}}\,\mathrm{d}t \underset{0+}{\sim} \frac{2\cos x}{3x^{3/2}} \underset{0+}{\sim} \frac{2}{3x^{3/2}}.$$

4.4.3 Comparaison des intégrales ordinaires et généralisées

L'intégrale généralisée hérite des propriétés suivantes de l'intégrale de Riemann ordinaire, à condition de remplacer l'hypothèse « $f, g \in \mathcal{R}[a,b]$ » par « $f, g \in \mathcal{R}^{\mathrm{loc}}[a,b[$ *et* les intégrales généralisées $\int_a^b f(t)\,\mathrm{d}t$ et $\int_a^b g(t)\,\mathrm{d}t$ convergent ». Nous laissons le soin au lecteur de vérifier ces propriétés en les appliquant d'abord à $\int_a^x$ avant de faire tendre x vers $b-$.

- Additivité, voir prop. 3.19.
- Linéarité, voir prop. 3.21.
- Positivité et croissance, voir les points i) et ii) de la proposition 3.22.
- L'additivité relative aux intervalles, cf. prop. 3.29 en notant que la convergence de $\int_a^b f(t)\,\mathrm{d}t$ implique celle de $\int_a^c f(t)\,\mathrm{d}t$ et de $\int_c^b f(t)\,\mathrm{d}t$, pour tout $c \in]a,b[$.
- La relation de Chasles prop. 3.30, à condition que chacune des trois intégrales concernées soit convergente. En effet en prenant c extérieur à $[a,b[$, on risque de faire apparaître une intégrale divergente.

Voyons maintenant les propriétés qui ne passent pas de l'intégrale ordinaire à l'intégrale généralisée. Il s'agit essentiellement de ce qui concerne la valeur absolue et le produit.

Rappelons d'abord que la Riemann intégrabilité de f implique celle de $|f|$ et que la réciproque est fausse, cf. remarque 3.24. On en déduit immédiatement que si f est *localement* Riemann intégrable sur $[a,b[$, $|f|$ l'est aussi. En revanche pour $f \in \mathcal{R}^{\mathrm{loc}}[a,b[$, la convergence de l'intégrale généralisée $\int_a^b f(t)\,\mathrm{d}t$ *n'implique pas* celle de $\int_a^b |f|(t)\,\mathrm{d}t$, voir l'exemple 4.28. Dans le même ordre d'idées, la Riemann intégrabilité locale de f sur $[a,b[$ implique celle de f^+ et de f^-, mais la convergence de $\int_a^b f(t)\,\mathrm{d}t$ *n'implique pas* celle de $\int_a^b f^+(t)\,\mathrm{d}t$ et $\int_a^b f^-(t)\,\mathrm{d}t$. L'exemple 4.28 avec $f(t) = t^{-1}\sin t$ sert aussi de contre exemple ici (vérification laissée en exercice). En revanche si $\int_a^b f(t)\,\mathrm{d}t$ est *absolument* convergente, $\int_a^b f^+(t)\,\mathrm{d}t$ et $\int_a^b f^-(t)\,\mathrm{d}t$ sont convergentes et on a

$$\int_a^b |f|(t)\,\mathrm{d}t = \int_a^b f^+(t)\,\mathrm{d}t + \int_a^b f^-(t)\,\mathrm{d}t, \quad \int_a^b f(t)\,\mathrm{d}t = \int_a^b f^+(t)\,\mathrm{d}t - \int_a^b f^-(t)\,\mathrm{d}t,$$

ainsi que

$$\Big|\int_a^b f(t)\,\mathrm{d}t\Big| \leqslant \int_a^b |f(t)|\,\mathrm{d}t < +\infty. \tag{4.32}$$

Regardons maintenant le produit. Si $f, g \in \mathcal{R}^{\mathrm{loc}}[a,b[$, il découle immédiatement de la proposition 3.27 que leur produit fg est lui aussi dans $\mathcal{R}^{\mathrm{loc}}[a,b[$. En revanche, la convergence, même absolue, des intégrales généralisées $\int_a^b f(t)\,\mathrm{d}t$ et $\int_a^b g(t)\,\mathrm{d}t$ *n'implique pas* celle de $\int_a^b f(t)g(t)\,\mathrm{d}t$. Un contre exemple immédiat est avec $[a,b[= [0,1[$, $f(t) = g(t) = (1-t)^{-1/2}$. On peut néanmoins obtenir la convergence de $\int_a^b f(t)g(t)\,\mathrm{d}t$ à partir de celle des intégrales généralisées non pas de f et g mais de f^2 et g^2.

Proposition 4.48 (inégalité de Cauchy-Schwarz)**.** *Si $f, g \in \mathcal{R}^{\mathrm{loc}}[a,b[$ sont telles que $\int_a^b f(t)^2\,\mathrm{d}t$ et $\int_a^b g(t)^2\,\mathrm{d}t$ convergent, alors $\int_a^b f(t)g(t)\,\mathrm{d}t$ converge absolument et vérifie :*

$$\left|\int_a^b f(t)g(t)\,\mathrm{d}t\right| \leqslant \left\{\int_a^b f(t)^2\,\mathrm{d}t\right\}^{1/2} \left\{\int_a^b g(t)^2\,\mathrm{d}t\right\}^{1/2}. \tag{4.33}$$

Preuve. D'abord, puisque f, g sont dans $\mathcal{R}^{\mathrm{loc}}[a,b[$, il en va de même pour $|f|$, $|g|$ et leur produit $|f| \cdot |g| = |fg|$. En appliquant l'inégalité de Cauchy-Schwarz pour les intégrales ordinaires, on en déduit que

$$\forall x \in [a,b[, \quad \int_a^x |f(t)g(t)|\,\mathrm{d}t \leqslant \left\{\int_a^x f(t)^2\,\mathrm{d}t\right\}^{1/2} \left\{\int_a^x g(t)^2\,\mathrm{d}t\right\}^{1/2}.$$

Toutes ces intégrales sont des fonctions croissantes de x. En faisant tendre x vers $b-$, elles convergent toutes dans $\overline{\mathbb{R}}_+$. Compte-tenu de l'hypothèse convergence de $\int_a^b f(t)^2\,\mathrm{d}t$ et $\int_a^b g(t)^2\,\mathrm{d}t$, on en déduit que

$$\int_a^b |f(t)g(t)|\,\mathrm{d}t \leqslant \left\{\int_a^b f(t)^2\,\mathrm{d}t\right\}^{1/2} \left\{\int_a^b g(t)^2\,\mathrm{d}t\right\}^{1/2} < +\infty.$$

Ceci montre que $\int_a^b f(t)g(t)\,\mathrm{d}t$ converge absolument et on conclut en appliquant l'inégalité (4.32) à la fonction fg. □

4.5 Un aperçu sur les intégrales multiples

Nous abordons maintenant l'extension de la notion d'intégrale de Riemann aux fonctions à valeurs réelles de plusieurs variables. La théorie est plus complexe qu'en dimension 1. Jusqu'ici nous n'avons intégré que sur un intervalle de $\mathbb{R}$. Le passage à la dimension $d > 1$ enrichit considérablement la famille des ensembles sur lesquels on souhaite pouvoir intégrer. La généralisation naturelle de la notion d'intervalle est celle de pavé, nous commencerons donc par examiner l'intégration d'une fonction bornée sur un pavé. De là, nous passerons à l'intégration d'une fonction bornée sur certains ensembles bornés de $\mathbb{R}^d$ appelés « admissibles ». Nous verrons ensuite comment calculer pratiquement ces intégrales, soit par intégrations répétées permettant de se ramener à la dimension 1, soit par changement de variable. Enfin nous verrons comment définir la notion d'intégrale généralisée pour des fonctions non bornées ou définies sur des domaines non bornés et quelles sont les propriétés essentielles d'une telle intégrale.

4.5.1 Intégration sur un pavé

Définition 4.49 (pavé)**.** *On appelle pavé fermé de $\mathbb{R}^d$, tout ensemble P de la forme*

$$P = [a_1,b_1] \times \cdots \times [a_d,b_d] = \prod_{i=1}^{d} [a_i,b_i], \tag{4.34}$$

où pour tout $i \in [\![1,d]\!]$, $-\infty < a_i \leqslant b_i < +\infty$. *En remplaçant ci-dessus chaque* $[a_i,b_i]$ *par* $]a_i,b_i[$, *on obtient un pavé ouvert (l'intérieur*[14] *de* P*) que nous noterons* P° .

Dans la suite, le mot « pavé » désignera implicitement un pavé fermé.

Définition 4.50 (volume d'un pavé)**.** *On appelle volume du pavé fermé* P *de* $\mathbb{R}^d$, *défini par* (4.34) *le réel*

$$\mathrm{vol}(P) := \prod_{i=1}^{d}(b_i - a_i).$$

Nous employons le mot volume pour simplifier. On pourrait parler de longueur pour $d = 1$, d'aire pour $d = 2$, de volume (au sens traditionnel) pour $d = 3$ et d'hypervolume pour $d \geqslant 4$.

Définition 4.51 (subdivision d'un pavé)**.** *On appelle subdivision du pavé fermé* P *de* $\mathbb{R}^d$, *défini par* (4.34), *toute partie finie* Δ *de* $\mathbb{R}^d$ *de la forme*

$$\Delta = \Delta_1 \times \cdots \times \Delta_d = \{(c_{1,j_1},\ldots,c_{d,j_d})\ ;\ j_i \in [\![0,n_i]\!],\ i \in [\![1,d]\!]\},$$

où pour tout $i \in [\![1,d]\!]$, $\Delta_i := \{c_{i,0},\ldots,c_{i,n_i}\}$ *est une subdivision de l'intervalle* $[a_i,b_i]$, *appelée* i^e *projection de* Δ. *Les pavés*

$$Q = \prod_{i=1}^{d}[c_{i,j_i-1},c_{i,j_i}], \quad j_i \in [\![1,n_i]\!], \quad i \in [\![1,d]\!],$$

sont appelés cellules de la subdivision. Nous notons $C(\Delta)$ *l'ensemble des cellules de* Δ.

Par un simple calcul algébrique, on vérifie la formule d'additivité du volume pour les subdivisions.

Proposition 4.52. *Pour toute subdivision* Δ *du pavé* $P = [a_1,b_1] \times \cdots \times [a_d,b_d]$,

$$\mathrm{vol}(P) = \sum_{Q \in C(\Delta)} \mathrm{vol}(Q). \tag{4.35}$$

Définition 4.53 (sommes de Darboux)**.** *Soit* P *un pavé de* $\mathbb{R}^d$, $f : P \to \mathbb{R}$ *une fonction bornée sur* P *et* Δ *une subdivision de* P. *On définit les sommes de Darboux inférieure* $S_\Delta(f)$ *et supérieure* $S^\Delta(f)$ *par*

$$S_\Delta(f) := \sum_{Q \in C(\Delta)} \mathrm{vol}(Q) \inf_Q f, \qquad S^\Delta(f) := \sum_{Q \in C(\Delta)} \mathrm{vol}(Q) \sup_Q f.$$

Lorsque $d = 1$, les cellules sont des intervalles de $[a_1,b_1]$ et on retrouve la définition 3.1.

14. Voir page 138 la définition générale de l'intérieur d'une partie A de $\mathbb{R}^d$.

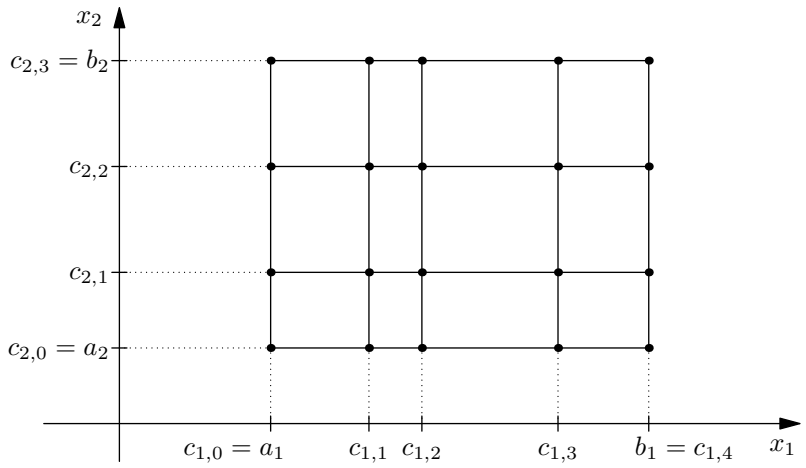


FIGURE 4.10 – Une subdivision du pavé $P = [a_1, b_1] \times [a_2, b_2]$

Définition 4.54 (intégrales inférieure et supérieure)**.** *Si P est un pavé de $\mathbb{R}^d$ et $f : P \to \mathbb{R}$ une fonction bornée sur P, on appelle respectivement intégrale de Riemann inférieure ou supérieure de f sur P les réels :*

$$I_*(P, f) := \sup_{\Delta} S_\Delta(f), \qquad I^*(P, f) := \inf_{\Delta} S^\Delta(f),$$

le supremum et l'infimum étant pris sur toutes les subdivisions Δ de P.

Définition 4.55 (raffinement d'une subdivision)**.** *Si $\Delta = \Delta_1 \times \cdots \times \Delta_d$ et $\Delta' = \Delta'_1 \times \cdots \times \Delta'_d$ sont deux subdivisions du pavé P, on dit que Δ' est un raffinement de Δ si Δ est inclus dans Δ' ou de manière équivalente si pour tout $i \in [\![1, d]\!]$, Δ'_i est un raffinement de Δ_i.*

On vérifie facilement que si Δ' est un raffinement de Δ, toute cellule de Δ' est incluse dans une seule cellule de Δ et que toute cellule de Δ est la réunion des cellules de Δ' qu'elle contient. Compte-tenu de la proposition 4.52, on en déduit que

$$\Delta \subset \Delta' \qquad \Longrightarrow \qquad S_\Delta(f) \leqslant S_{\Delta'}(f) \quad \text{et} \quad S^{\Delta'}(f) \leqslant S^\Delta(f).$$

Si Δ et Δ' sont deux subdivisions quelconques de P, l'ensemble $\Delta \cup \Delta'$ n'est en général pas une subdivision [15] de P, mais il existe une plus petite subdivision de P qui raffine à la fois Δ et Δ', c'est la subdivision Δ'' telle que pour chaque i, $\Delta''_i = \Delta_i \cup \Delta'_i$. On peut alors utiliser Δ'' pour montrer que $S_\Delta(f) \leqslant S^{\Delta'}(f)$, comme à la page 72 et en déduire que pour toute fonction bornée $f : P \to \mathbb{R}$, $I_*(P, f) \leqslant I^*(P, f)$.

15. Faites le dessin avec $P = [0, 3]^2$, $\Delta_1 = \Delta_2 = \{0, 1, 3\}$ et $\Delta'_1 = \Delta'_2 = \{0, 2, 3\}$ si cela ne vous paraît pas évident.

Définition 4.56 (intégrale de Riemann sur un pavé)**.** *Soit P un pavé de $\mathbb{R}^d$. On dit que f bornée $P \to \mathbb{R}$ est* Riemann intégrable *si avec les notations ci-dessus, $I_*(P,f) = I^*(P,f)$. Dans ce cas on définit son intégrale au sens de Riemann sur P, notée $\int_P f(x)\,\mathrm{d}x$ ou $\int_P f(x_1,\dots,x_d)\,\mathrm{d}x_1 \dots \mathrm{d}x_d$, par*

$$\int_P f(x)\,\mathrm{d}x := I_*(P,f) = I^*(P,f).$$

On note $\mathcal{R}(P)$ la famille de toutes les fonctions Riemann intégrables sur P.

Dans tout le reste de ce chapitre, sauf contre-indication explicite, les expressions abrégées « intégrable » et « intégrale » signifieront respectivement Riemann intégrable et intégrale au sens de Riemann.

Voici une première famille de fonctions Riemann intégrables sur P.

Proposition 4.57 (intégrabilité d'une fonction continue)**.** *Si P est un pavé fermé de $\mathbb{R}^d$, toute fonction $f : P \to \mathbb{R}$ continue sur P est Riemann intégrable sur P.*

Si f est continue sur le pavé fermé P qui est un compact, elle y est bornée et uniformément continue. On peut alors facilement adapter la preuve de la proposition 3.11 pour établir la Riemann intégrabilité de f sur P.

4.5.2 Intégration sur un ensemble borné

Pour définir l'intégrabilité et le cas échéant, l'intégrale d'une fonction bornée f sur une partie bornée A de $\mathbb{R}^d$, une idée naturelle est d'inclure A dans un pavé P et de considérer la fonction f_P égale à f sur A et nulle sur $P\backslash A$. On aimerait pouvoir dire que f est intégrable sur A si f_P est intégrable sur P et dans ce cas, définir $\int_A f(x)\,\mathrm{d}x := \int_P f_P(x)\,\mathrm{d}x$. Pour assurer la cohérence de cette définition, il nous faut en préliminaire établir quelques lemmes.

Définition 4.58. *On appelle pas de la subdivision Δ, notation* $\mathrm{pas}(\Delta)$*, la longueur maximale des côtés de ses cellules. Avec les notations de la définition 4.51,*

$$\mathrm{pas}(\Delta) = \max_{i\in[\![1,d]\!]} \max_{j_i\in[\![1,n_i]\!]} |c_{i,j_i} - c_{i,j_i-1}|.$$

Lemme 4.59. *Soit P un pavé fermé contenant le pavé fermé R. Notons $\mathcal{S}(P,R)$ l'ensemble de toutes les subdivisions de P contenant une subdivision de R. Pour toute subdivision $\Delta \in \mathcal{S}(P,R)$, notons $V(R,\Delta)$ l'ensemble des cellules Q de Δ telles que $Q \cap R \neq \varnothing$ et $Q \cap R^\circ = \varnothing$. Alors pour tout $\varepsilon > 0$, on peut trouver $\delta > 0$ tel que*

$$\forall \Delta \in \mathcal{S}(P,R), \qquad \mathrm{pas}(\Delta) < \delta \quad \Rightarrow \quad \sum_{Q\in V(R,\Delta)} \mathrm{vol}(Q) < \varepsilon.$$

Preuve. Notons $R = [a'_1, b'_1] \times \cdots \times [a'_d, b'_d]$. Si $\Delta \in \mathcal{S}(P,R)$, on peut classer les cellules Q de Δ en trois catégories :

– celles qui intersectent R°, ce sont les celulles de la subdivision de R contenue dans Δ, notée $\Delta \cap R$;
– celles de $V(R, \Delta)$;
– celles qui ont une intersection vide avec R.

On remarque que la réunion des cellules des deux premières catégories est un pavé R' et que les sommets de ces cellules forment une subdivision de R'. Par additivité du volume pour les subdivisions (proposition 4.52),

$$\mathrm{vol}(R') = \sum_{Q \in C(\Delta \cap R)} \mathrm{vol}(Q) + \sum_{Q \in V(R,\Delta)} \mathrm{vol}(Q).$$

D'autre part en notant $\delta' = \mathrm{pas}(\Delta)$, il est clair que $R' \subset R^{\delta'} := \prod_{i=1}^d [a'_i - \delta', b'_i + \delta']$. En particulier $\mathrm{vol}(R') \leqslant \mathrm{vol}(R^{\delta'})$, d'où

$$\sum_{Q \in V(R,\Delta)} \mathrm{vol}(Q) \leqslant \mathrm{vol}(R^{\delta'}) - \sum_{Q \in C(\Delta \cap R)} \mathrm{vol}(Q) = \prod_{i=1}^d (b'_i - a'_i + 2\delta') - \prod_{i=1}^d (b'_i - a'_i).$$

Comme ce majorant tend vers 0 avec δ', le lemme est prouvé. □

Lemme 4.60. *Soit P un pavé fermé contenant le pavé fermé R. Soit $g : P \to \mathbb{R}$, Riemann intégrable sur P et nulle sur $P \backslash R$. Alors g est Riemann intégrable sur R et $\int_R g(x)\,\mathrm{d}x = \int_P g(x)\,\mathrm{d}x$.*

Preuve. Notons $I = \int_P g(x)\,\mathrm{d}x$. Soit $\varepsilon > 0$ arbitraire et δ choisi en fonction de ε comme dans le lemme 4.59. Puisque g est Riemann intégrable sur P, on peut trouver, quitte à raffiner, une subdivision Δ de P contenant une subdivision de R, vérifiant $\mathrm{pas}(\Delta) < \delta$ et

$$I - \varepsilon \leqslant S_\Delta(g) \leqslant S^\Delta(g) \leqslant I + \varepsilon. \tag{4.36}$$

Pour chaque cellule Q de Δ, notons $m_Q := \inf_Q g$ et $M_Q := \sup_Q g$. L'ensemble $\Delta \cap R$ est une subdivision de R. Comparons les sommes de Darboux $S_{\Delta\cap R}(g)$ et $S_\Delta(g)$.

$$S_\Delta(g) = \sum_{Q \in C(\Delta \cap R)} m_Q \,\mathrm{vol}(Q) + \sum_{Q \in V(R,\Delta)} m_Q \,\mathrm{vol}(Q) + \sum_{\substack{Q \in C(\Delta) \\ Q \cap R = \varnothing}} m_Q \,\mathrm{vol}(Q).$$

La dernière de ces trois sommes est nulle puisque g est nulle sur chacune de ses cellules. La première est exactement $S_{\Delta\cap R}(g)$. La deuxième est majorée en valeur absolue par $M\varepsilon$ où $M = \sup_P |g|$, grâce au lemme 4.59. Donc $|S_{\Delta\cap R}(g) - S_\Delta(g)| < M\varepsilon$. On vérifie de même que $|S^{\Delta\cap R}(g) - S^\Delta(g)| < M\varepsilon$. En reportant ceci dans (4.36), on obtient :

$$I - (1+M)\varepsilon \leqslant S_{\Delta\cap R}(g) \leqslant S^{\Delta\cap R}(g) \leqslant I + (1+M)\varepsilon,$$

d'où

$$I - (1+M)\varepsilon \leqslant I_*(R,g) \leqslant I^*(R,g) \leqslant I + (1+M)\varepsilon.$$

Comme $\varepsilon > 0$ est arbitraire, on en déduit que $I_*(R,g) = I^*(R,g) = I$, donc g est Riemann intégrable sur R et $\int_R g(x)\,\mathrm{d}x = I = \int_P g(x)\,\mathrm{d}x$. □

Nous laissons au lecteur le soin d'adapter les arguments de la preuve ci-dessus pour vérifier le lemme suivant.

Lemme 4.61. *Soit P un pavé fermé contenant le pavé fermé R. Soit $g : P \to \mathbb{R}$, Riemann intégrable sur R et nulle sur $P\backslash R$. Alors g est Riemann intégrable sur P et $\int_R g(x)\,\mathrm{d}x = \int_P g(x)\,\mathrm{d}x$.*

Nous pouvons maintenant établir le lemme qui légitime la définition de l'intégrale sur une partie bornée de $\mathbb{R}^d$.

Lemme 4.62. *Soit A une partie bornée de $\mathbb{R}^d$ et $f : A \to \mathbb{R}$. Pour tout pavé P contenant A, notons $f_P : P \to \mathbb{R}$ la fonction égale à f sur A et à 0 sur $P\backslash A$. Si f_P est Riemann intégrable sur P pour un pavé P contenant A, alors pour tout pavé P' contenant A, $f_{P'}$ est Riemann intégrable sur P' et $\int_P f_P(x)\,\mathrm{d}x = \int_{P'} f_{P'}(x)\,\mathrm{d}x$.*

Preuve. On remarque que $A \subset R = P \cap P'$, que R est un pavé et que f_P comme $f_{P'}$ sont nulles en dehors de R. Il suffit alors d'appliquer le lemme 4.60 avec le pavé $R \subset P$, puis le lemme 4.61 avec le même $R \subset P'$. □

Définition 4.63 (intégrale de Riemann sur une partie bornée de $\mathbb{R}^d$)**.** *Soit A une partie bornée de $\mathbb{R}^d$, P un pavé contenant A et f une fonction réelle définie sur A. On note $f_P : P \to \mathbb{R}$ la fonction égale à f sur A et à 0 sur $P\backslash A$. On dit que f est Riemann intégrable sur A si f_P est Riemann intégrable sur le pavé P et dans ce cas, on appelle intégrale de f sur A le réel*

$$\int_A f(x)\,\mathrm{d}x := \int_P f_P(x)\,\mathrm{d}x.$$

L'ensemble des fonctions Riemann intégrables sur A est noté $\mathcal{R}(A)$.

Puisque les intégrales de Riemann sur un ensemble borné peuvent se ramener à une intégrale de Riemann sur un pavé, elles héritent des propriétés de l'intégrale de Riemann sur un pavé. Pour établir ces dernières, il suffit essentiellement d'adapter les démonstrations faites pour $d = 1$ (intégrale de Riemann sur le segment $[a, b]$) en remplaçant dans les sommes de Darboux les intervalles de subdivision unidimensionnelle par les cellules de subdivision de pavé [16]. Nous nous contenterons donc d'énoncer les résultats ainsi obtenus.

Proposition 4.64 (propriétés de l'intégrale de Riemann sur A borné)**.**

1. Linéarité. *L'ensemble $\mathcal{R}(A)$ des applications $f : A \to \mathbb{R}$ Riemann intégrables sur A est un $\mathbb{R}$-espace vectoriel et l'application $\Psi : \mathcal{R}(A) \to \mathbb{R}$, $f \mapsto \int_A f(x)\,\mathrm{d}x$ est une forme linéaire sur cet espace.*
2. Croissance de l'intégrale. *L'intégrale de Riemann sur A possède les quatre propriétés suivantes relativement à la relation d'ordre partiel $\leqslant$ définie sur $\mathcal{R}(A)$ par $f \leqslant g$ si pour tout $x \in A$, $f(x) \leqslant g(x)$.*

16. Cette méthode fonctionne bien pour les propriétés de l'intégrale sur $[a, b]$ où la structure d'ordre de $[a, b]$ n'est pas essentielle.

a) Positivité :

$$\text{si } f \in \mathcal{R}(A) \text{ et } f \geqslant 0 \text{ sur } A, \quad \int_A f(x)\,\mathrm{d}x \geqslant 0. \tag{4.37}$$

b) Croissance :

$$\text{si } f, g \in \mathcal{R}(A) \text{ et } f \leqslant g \text{ sur } A, \quad \int_A f(x)\,\mathrm{d}x \leqslant \int_A g(x)\,\mathrm{d}x. \tag{4.38}$$

c) Si $f \in \mathcal{R}(A)$, l'application $|f| : x \mapsto |f(x)|$ est elle aussi Riemann intégrable sur A et

$$\Big|\int_A f(x)\,\mathrm{d}x\Big| \leqslant \int_A |f(x)|\,\mathrm{d}x. \tag{4.39}$$

d) Si f est Riemann intégrable sur A, les fonctions $f^+ := \max(f, 0)$ et $f^- := \max(-f, 0) = -\min(f, 0)$ le sont aussi. On a de plus

$$\int_A f(x)\,\mathrm{d}x = \int_A f^+(x)\,\mathrm{d}x - \int_A f^-(x)\,\mathrm{d}x. \tag{4.40}$$

3. Produit de deux fonctions de $\mathcal{R}(A)$.

a) Si f et g sont Riemann intégrables sur A, leur produit fg l'est aussi.

b) Inégalité de Cauchy-Schwarz dans $\mathcal{R}(A)$. Pour toutes $f, g \in \mathcal{R}(A)$,

$$\Big|\int_A f(x)g(x)\,\mathrm{d}x\Big| \leqslant \left\{\int_A f(x)^2\,\mathrm{d}x\right\}^{1/2} \left\{\int_A g(x)^2\,\mathrm{d}x\right\}^{1/2}. \tag{4.41}$$

Proposition 4.65 (additivité par rapport aux ensembles d'intégration)**.** *Soient A et B deux parties bornées* disjointes *de $\mathbb{R}^d$ et $f : A \cup B \to \mathbb{R}$. Si la restriction de f à chacun des ensembles A et B est Riemann intégrable sur cet ensemble, alors f est Riemann intégrable sur $A \cup B$ et*

$$\int_{A\cup B} f(x)\,\mathrm{d}x = \int_A f(x)\,\mathrm{d}x + \int_B f(x)\,\mathrm{d}x. \tag{4.42}$$

Preuve. Soit P un pavé contenant l'ensemble borné $A \cup B$. Notons g la fonction égale à f sur A et à 0 sur $P\backslash A$, h la fonction égale à f sur B et nulle sur $P\backslash B$. Puisque la restriction de f à A est intégrable sur A, g est intégrable sur P et $\int_P g(x)\,\mathrm{d}x = \int_A f(x)\,\mathrm{d}x$. De même h est intégrable sur P et $\int_P h(x)\,\mathrm{d}x = \int_A f(x)\,\mathrm{d}x$. Par la proposition 4.64.1, $g + h$ est intégrable sur P et $\int_P((g(x) + h(x))\,\mathrm{d}x = \int_P g(x)\,\mathrm{d}x + \int_P h(x)\,\mathrm{d}x$. Comme $A \cap B = \varnothing$, $g + h$ est égale à f sur $A \cup B$ et à zéro sur $P\backslash(A \cup B)$. On en déduit que f est intégrable sur $A \cup B$ et

$$\int_{A\cup B} f(x)\,\mathrm{d}x = \int_P (g+h)(x)\,\mathrm{d}x = \int_P g(x)\,\mathrm{d}x + \int_P h(x)\,\mathrm{d}x = \int_A f(x)\,\mathrm{d}x + \int_B f(x)\,\mathrm{d}x,$$

ce qui établit (4.42). □

Remarque 4.66. La « réciproque » de la proposition 4.65 *est fausse* : l'intégrabilité de f sur $A \cup B$ n'implique pas son intégrabilité sur A et sur B. Plus généralement, l'intégrabilité de f sur une partie bornée E de $\mathbb{R}^d$ n'implique pas son intégrabilité sur tout sous-ensemble de E. Pour s'en convaincre, voir l'exemple 4.68 ci-dessous.

Proposition 4.67. *On suppose que f est Riemann intégrable sur A partie bornée de $\mathbb{R}^d$ et aussi sur $B \subset A$. Alors f est Riemann intégrable sur $A\backslash B$ et*

$$\int_{A\backslash B} f(x)\,\mathrm{d}x = \int_A f(x)\,\mathrm{d}x - \int_B f(x)\,\mathrm{d}x. \tag{4.43}$$

Preuve. Soit P un pavé contenant A et donc aussi B et $A\backslash B$. Pour tout $C \subset A$, notons $f_{P,C}$ la fonction égale à f sur C et à 0 sur $P\backslash C$. Alors $f_{P,A\backslash B} = f_{P,A} - f_{P,B}$. Par hypothèse, f est intégrable sur A et sur B, donc par définition de l'intégrabilité sur un ensemble borné, $f_{P,A}$ et $f_{P,B}$ sont intégrables sur le pavé P. Comme $\mathcal{R}(P)$ est un espace vectoriel, $f_{P,A\backslash B} = f_{P,A} - f_{P,B}$ est elle aussi intégrable sur P, ce qui justifie l'intégrabilité de f sur $A\backslash B$. De plus, par linéarité de l'intégrale sur P,

$$\begin{aligned}\int_P f_{P,A\backslash B}(x)\,\mathrm{d}x = \int_P (f_{P,A} - f_{P,B})(x)\,\mathrm{d}x &= \int_P f_{P,A}(x)\,\mathrm{d}x - \int_P f_{P,B}(x)\,\mathrm{d}x \\ &= \int_A f(x)\,\mathrm{d}x - \int_B f(x)\,\mathrm{d}x,\end{aligned}$$

ce qui établit (4.43). □

4.5.3 Ensembles et fonctions admissibles

La définition 4.63 et les propositions 4.64 et 4.65 peuvent donner une impression de simplicité et de généralité trompeuses. Certes, on a réussi à passer de la notion d'intégrale sur un segment $[a, b]$ à un pavé de $\mathbb{R}^d$, puis à un ensemble borné quelconque de $\mathbb{R}^d$ en conservant une bonne partie des propriétés de l'intégrale. Il y a néanmoins un *hic*, c'est que pour une partie bornée A quelconque de $\mathbb{R}^d$, il n'est pas clair qu'il y ait beaucoup de fonctions intégrables sur A (il y en a toujours au moins une, c'est la fonction nulle). D'autre part, même pour un pavé P, nous savons seulement pour l'instant que les fonctions continues sont dans $\mathcal{R}(P)$. Voici un exemple révélateur d'un ensemble borné où $\mathcal{R}(A)$ est assez pauvre.

Exemple 4.68. Dans $\mathbb{R}^2$, prenons $A = [0,1]^2 \cap \mathbb{Q}^2$ et soit $f : A \to \mathbb{R}$, $x \mapsto 1$. Alors A est borné, f est constante sur A (donc continue pour la topologie trace de $\mathbb{R}^2$). Pourtant f *n'est pas* Riemann intégrable sur A. Pour le voir, il suffit de prendre le pavé $P = [0,1]^2$ qui contient A et, avec les notations de la définition 4.63, de constater que pour toute subdivision Δ de P, $S_\Delta(f_P) = 0$ et $S^\Delta(f_P) = 1$, car dans chaque cellule Q de Δ, f_P prend les valeurs 0 et 1 et aucune autre. Donc $I_*(P, f_P) = 0 < I^*(f_P) = 1$ et d'après le lemme 4.62 et la définition 4.63, f n'est pas intégrable sur A. Par le même raisonnement, on peut voir que si g est la restriction à A d'une fonction continue et strictement positive sur $[0,1]^2$, g n'est pas Riemann intégrable sur A.

Dans ce qui suit, nous allons trouver d'autres fonctions Riemann intégrables sur un pavé que les fonctions continues et donner une condition topologique sur l'ensemble borné A pour éliminer des pathologies comme celle de l'exemple 4.68. Notons au passage que dans cet exemple, la *frontière* de A (cf. déf. 4.73) est $\partial A = \overline{A}\backslash A^\circ = [0,1]^2\backslash\varnothing = [0,1]^2$.

Notre première étape sera de montrer qu'une fonction bornée sur P et dont l'ensemble des discontinuités est « $\mathcal{R}$-négligeable » est encore Riemann intégrable sur P.

Définition 4.69 (ensemble $\mathcal{R}$-négligeable)**.** *Une partie D de $\mathbb{R}^d$ est dite négligeable au sens de Riemann ou $\mathcal{R}$-négligeable si, pour tout $\varepsilon > 0$, on peut trouver une suite finie $R_1, \ldots, R_m$ de pavés fermés telle que*

$$D \subset \bigcup_{j=1}^{m} R_j^\circ \quad \textit{et} \quad \sum_{j=1}^{m} \operatorname{vol}(R_j) < \varepsilon.$$

Il est clair d'après cette définition, qu'une partie $\mathcal{R}$-négligeable est bornée, que sa fermeture est aussi $\mathcal{R}$-négligeable et compacte. Une union finie d'ensembles $\mathcal{R}$-négligeables est aussi $\mathcal{R}$-négligeable.

Définition 4.70 (fonction admissible sur un pavé)**.** *Une fonction à valeurs réelles définie sur un pavé fermé P de $\mathbb{R}^d$ est dite* admissible *sur P, si elle est bornée sur P et si l'ensemble de ses points de discontinuité est une partie $\mathcal{R}$-négligeable de P.*

Proposition 4.71 (intégrabilité d'une fonction admissible)**.** *Toute fonction admissible sur le pavé P est Riemann intégrable sur P. Si f et g sont admissibles sur P et égales partout, sauf sur un sous-ensemble $\mathcal{R}$-négligeable de P, alors $\int_P f(x)\,\mathrm{d}x = \int_P g(x)\,\mathrm{d}x$.*

Lemme 4.72. *Soit R un pavé contenu dans le pavé P. Pour tout $\varepsilon > 0$, il existe un $\delta > 0$ tel que pour toute subdivision Δ de P de pas inférieur à δ,*

$$\sum_{\substack{Q\in C(\Delta)\\ Q\cap R\neq\varnothing}} \operatorname{vol}(Q) \leqslant \operatorname{vol}(R) + \varepsilon.$$

Preuve. Avant de se lancer dans la preuve formelle, il n'est pas inutile de se faire une intuition avec la figure 4.11. Sur cette figure, la zone grisée représente la réunion S des cellules de Δ qui intersectent R. On « voit » qu'en resserrant les mailles du filet (c.-à-d. en réduisant le pas de Δ), on peut rendre l'aire de S aussi proche que l'on veut de celle de R. Le rectangle en noir qui entoure la zone grisée est R^δ, dilaté d'ordre δ de R, voir ci-dessous.

Notons $P = [a_1, b_1] \times \cdots \times [a_d, b_d]$ et $R = [a'_1, b'_1] \times \cdots \times [a'_d, b'_d]$. Soit Δ une subdivision quelconque de P, de pas inférieur à δ, le choix de δ en fonction de ε sera précisé ultérieurement. Comme $R \subset P$, $[a'_i, b'_i] \subset [a_i, b_i]$ pour tout i. On peut donc encadrer a'_i entre deux termes consécutifs de la subdivision Δ_i et faire de même pour b'_i, ce que nous écrirons

$$c_{i,j'_i} \leqslant a'_i < c_{i,j'_i+1}, \qquad c_{i,j''_i-1} < b'_i \leqslant c_{i,j''_i},$$

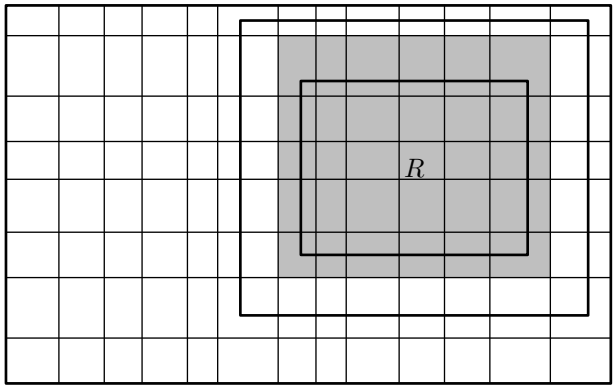


FIGURE 4.11 – Cellules de Δ intersectant R (en grisé), elles-mêmes encadrées par R^δ

avec $j'_i < j''_i$. Alors la réunion des cellules Q de Δ qui intersectent R est exactement le pavé $S = \prod_{i=1}^d [c_{i,j'_i}, c_{i,j''_i}]$. D'autre part, puisque le pas de Δ est inférieur à δ, $a'_i - \delta < c_{i,j'_i}$ et $c_{i,j''_i} < b'_i + \delta$. On a donc les inclusions de pavés :

$$R \subset S \subset R^\delta := \prod_{i=1}^d [a'_i - \delta, b'_i + \delta].$$

En notant que le volume d'un pavé fermé est évidemment croissant pour l'inclusion, on en déduit que

$$\sum_{\substack{Q \in C(\Delta) \\ Q \cap R \neq \varnothing}} \mathrm{vol}(Q) = \mathrm{vol}(S) \leqslant \mathrm{vol}(R^\delta) = \prod_{i=1}^d (b'_i - a'_i + 2\delta).$$

Comme ce produit est une fonction continue de δ, on peut choisir δ suffisamment petit pour le rendre inférieur à $\prod_{i=1}^d (b'_i - a'_i) + \varepsilon$ (et ce choix ne dépend que de R et de ε). Le lemme est démontré. □

Preuve de la proposition 4.71. Soit f une fonction admissible sur le pavé fermé P et D l'ensemble de ses points de discontinuité. Fixons un $\varepsilon > 0$ arbitraire. Puisque D est $\mathcal{R}$-négligeable, il existe une suite finie $R_1, \ldots, R_m$ de pavés fermés tels que

$$D \subset V := \bigcup_{j=1}^m R_j^\circ \quad \text{et} \quad \sum_{j=1}^m \mathrm{vol}(R_j) < \varepsilon.$$

L'ensemble V est ouvert, donc l'intersection $K = P \cap V^c$ est fermée, bornée car incluse dans P, donc compacte. Comme f est continue en tout point de K, elle y est uniformément continue. Pour traduire cette continuité uniforme, il est commode d'utiliser la norme $\|x\| = \|(x_1, \ldots, x_d)\| = \max_{1 \leqslant i \leqslant d} |x_i|$. Il existe alors $\delta > 0$ tel que

$$\forall x, y \in K, \quad \|x - y\| < \delta \Rightarrow |f(x) - f(y)| < \varepsilon. \tag{4.44}$$

Rappelons d'autre part, qu'en tant que fonction admissible sur P, f est bornée sur P. Nous noterons $\|f\|_\infty := \sup_{x\in P} |f(x)|$. En appliquant le lemme 4.72 à chacun des pavés R_j, et quitte à diminuer δ, on peut trouver une subdivision Δ de pas inférieur à δ telle que si $Q_1, \dots, Q_n$ sont les cellules de Δ intersectant au moins l'un des R_j,

$$\mathrm{vol}(Q_1) + \cdots + \mathrm{vol}(Q_n) < 2\varepsilon. \tag{4.45}$$

Pour majorer la différence $S^\Delta(f) - S_\Delta(f)$, notons C'_Δ l'ensemble des cellules de Δ n'intersectant aucun R_j. Leur réunion est incluse dans K, on peut donc utiliser (4.44) sur chaque $Q \in C'_\Delta$ et en particulier $\sup_Q f - \inf_Q f < \varepsilon$. En effet, avec notre choix de norme sur $\mathbb{R}^d$, si x et y sont dans la même cellule de Δ, $\|x - y\| \leqslant \mathrm{pas}(\Delta) < \delta$. Pour les cellules $Q_1, \dots, Q_n$, on utilisera plutôt (4.45) combinée avec la bornitude de f. Ceci nous donne :

$$\begin{aligned} S^\Delta(f) - S_\Delta(f) &= \sum_{k=1}^{n} (\sup_{Q_k} f - \inf_{Q_k} f)\,\mathrm{vol}(Q_k) + \sum_{Q\in C'_\Delta} (\sup_{Q} f - \inf_{Q} f)\,\mathrm{vol}(Q) \\ &\leqslant 4\|f\|_\infty \varepsilon + \mathrm{vol}(P)\varepsilon. \end{aligned}$$

Comme $\varepsilon > 0$ était arbitraire, l'intégrabilité de f est établie.

Supposons maintenant que f et g admissibles sur P sont égales partout, sauf sur un ensemble $\mathcal{R}$-négligeable D_0. Notons respectivement D_1 et D_2 les ensembles de points de discontinuité de f et g. Alors $D = D_0 \cup D_1 \cup D_2$ est $\mathcal{R}$ négligeable. En choisissant Δ comme ci-dessus, les sommes de Darboux $S_\Delta(f)$ et $S_\Delta(g)$ diffèrent seulement par les contributions de $Q_1, \dots, Q_m$, donc $|S_\Delta(f) - S_\Delta(g)| \leqslant 4(\|f\|_\infty + \|g\|_\infty)\varepsilon$. On en déduit que $|I_*(P,f) - I_*(P,g)| \leqslant 4(\|f\|_\infty + \|g\|_\infty)\varepsilon$, d'où $I_*(P,f) = I_*(P,g)$ par arbitrarité de ε. Comme f et g sont toutes deux Riemann intégrables sur P d'après la première partie de la démonstration, $\int_P f(x)\,\mathrm{d}x = \int_P g(x)\,\mathrm{d}x$. □

Avant de définir les ensembles admissibles, il convient de faire quelques rappels sur la notion de *frontière*.

Définition 4.73 (intérieur, extérieur, frontière)**.** *Soit A une partie de $\mathbb{R}^d$. L'*intérieur *de A, noté A°, est la réunion de tous les ouverts inclus dans A. La* fermeture *de A, notée $\overline{A}$, est l'intersection de tous les fermés contenant A. L'*extérieur *de A, noté A^{ext}, est l'intérieur de son complémentaire A^c dans $\mathbb{R}^d$. La frontière de A, notée ∂A, est le complémentaire dans $\mathbb{R}^d$ de $A^\circ \cup A^{\mathrm{ext}}$.*

On vérifie facilement que $x \in \partial A$ si et seulement si tout voisinage de x contient à la fois des points de A et de son complémentaire, cf. par exemple [13] th. (T.2, II,6,1). Notons aussi quelques propriétés de la frontière qui nous seront utiles (exercice 4.46).

$$\begin{array}{llll} a) & \partial A = \overline{A}\setminus A^\circ & b) & \partial(A\cup B) \subset (\partial A)\cup(\partial B) \\ c) & \partial(A\cap B) \subset (\partial A)\cup(\partial B) & d) & \partial(A\setminus B) \subset (\partial A)\cup(\partial B). \end{array} \tag{4.46}$$

Définition 4.74 (ensemble admissible)**.** *Un sous-ensemble A de $\mathbb{R}^d$ est dit admissible s'il est borné et si sa frontière ∂A est $\mathcal{R}$-négligeable.*

La frontière d'un pavé étant $\mathcal{R}$-négligeable (exercice 4.6), tous les pavés sont admissibles. La frontière de $A = [0,1]^2 \cap \mathbb{Q}^2$ n'est pas $\mathcal{R}$-négligeable (cf. exemple 4.68).

La définition 4.74 nous permet d'énoncer le corollaire suivant de la proposition 4.71.

Corollaire 4.75. *Si A est une partie admissible de $\mathbb{R}^d$, toute fonction $f : A \to \mathbb{R}$, continue et bornée sur A est Riemann intégrable sur A.*

Preuve. Notons $\tilde{f}$ la fonction égale à 0 sur $\mathbb{R}^d \backslash A$ et à f sur A. Cette fonction bornée est continue sur *l'ouvert* A° et constante valant 0 sur *l'ouvert* A^{ext}, donc aussi continue sur cet ouvert [17]. Elle est donc continue sur la réunion de ces deux ouverts et l'ensemble de ses discontinuités est inclus dans ∂A, donc $\mathcal{R}$-négligeable. Par conséquent la restriction de $\tilde{f}$ à tout pavé P contenant A est admissible sur P. Cette restriction coïncide sur P avec la fonction f_P de la définition 4.63. Par la proposition 4.71, f_P est intégrable sur P, donc f est intégrable sur A par la définition 4.63. □

Le corollaire 4.75 s'applique en particulier à la fonction indicatrice d'un ensemble admissible B. Cette remarque nous fournit deux autres corollaires de la proposition 4.71.

Corollaire 4.76. *Soit A une partie bornée de $\mathbb{R}^d$ et $f : A \to \mathbb{R}$ Riemann intégrable sur A. Alors f est aussi Riemann intégrable sur toute partie admissible B de A, et*

$$\int_B f(x)\,\mathrm{d}x = \int_A f(x)\mathbf{1}_B(x)\,\mathrm{d}x.$$

Preuve. Soit P un pavé contenant A et donc aussi B. Pour $C \subset A$, notons $f_{P,C}$ la fonction égale à f sur C et nulle sur $P \backslash C$. La fonction $f_{P,B}$ est Riemann intégrable sur P car $f_{P,B} = f_{P,A}\mathbf{1}_B$ est le produit de deux fonctions Riemann intégrables sur P. Donc, f est Riemann intégrable sur B et

$$\int_B f(x)\,\mathrm{d}x = \int_P f_{P,B}(x)\,\mathrm{d}x = \int_P f_{P,A}(x)\mathbf{1}_B(x)\,\mathrm{d}x = \int_A f(x)\mathbf{1}_B(x)\,\mathrm{d}x,$$

car les fonctions $f_{P,A}\mathbf{1}_B$ et $f\mathbf{1}_B$ sont égales sur A et $f_{P,A}\mathbf{1}_B$ est nulle sur $P \backslash A$. □

Corollaire 4.77. *Soit A une partie admissible de $\mathbb{R}^d$ et $f : A \to \mathbb{R}$ Riemann intégrable sur A. Alors f est aussi Riemann intégrable sur l'ouvert A° et*

$$\int_{A^\circ} f(x)\,\mathrm{d}x = \int_A f(x)\,\mathrm{d}x.$$

Preuve. On commence par remarquer que $\partial(A^\circ) \subset \partial A$ (l'inclusion pouvant être stricte [18]), par exemple en appliquant (4.46-a) en notant que $(A^\circ)^\circ = A^\circ$ et $\overline{A^\circ} \subset \overline{A}$.

17. La fonction f est continue en tout point de A pour *la topologie trace* de $\mathbb{R}^d$. Elle ne conserve pas forcément cette continuité par le prolongement $\tilde{f}$. Mais si x est un point de l'intérieur de A, il existe une boule ouverte $B(x,r)$ de $\mathbb{R}^d$ incluse dans A et la continuité de f au point x équivaut à celle de $\tilde{f}$.

18. C'est le cas dans l'exemple 4.68 où $A^\circ = \varnothing$ donc $\partial(A^\circ) = \varnothing$, tandis que $\partial A = [0,1]^2$.

Par conséquent, A° hérite de l'admissibilité de A. Le corollaire 4.76 avec $B = A^\circ$ nous dit alors que f est Riemann intégrable sur A° et que

$$\int_{A^\circ} f(x)\,\mathrm{d}x = \int_A f(x)\mathbf{1}_{A^\circ}(x)\,\mathrm{d}x.$$

On note alors que sur A, f et $f\mathbf{1}_{A^\circ}$ sont bornées et ne diffèrent que sur l'ensemble $A\backslash A^\circ$ qui est $\mathcal{R}$-négligeable en raison de l'inclusion $A\backslash A^\circ \subset \overline{A}\backslash A^\circ = \partial A$. La proposition 4.71 combinée avec la définition de l'intégrale sur A nous permet alors de voir que $\int_A f(x)\mathbf{1}_{A^\circ}(x)\,\mathrm{d}x = \int_A f(x)\,\mathrm{d}x$, ce qui conclut la preuve. □

Pour pouvoir utiliser en pratique la propriété d'admissibilité, nous avons beoin de conditions suffisantes pour qu'un ensemble soit $\mathcal{R}$-négligeable. Rappelons qu'une application $\varphi : D \to \mathbb{R}^d$ est lipschitzienne sur une partie D de $\mathbb{R}^m$, s'il existe une constante c telle que

$$\forall x, y \in D, \quad \|\varphi(x) - \varphi(y)\| \leqslant c\|x - y\|. \tag{4.47}$$

Puisque toutes les normes sur $\mathbb{R}^d$ sont équivalentes, on est libre du choix des normes dans (4.47), quitte à modifier la valeur de la constante c. Nous choisissons par commodité $\|x\| = \|(x_1, \ldots, x_m)\| = \max_{1\leqslant i\leqslant m} |x_i|$ et nous faisons de même pour la norme sur $\mathbb{R}^d$.

Proposition 4.78. *Soit D une partie $\mathcal{R}$-négligeable de $\mathbb{R}^m$ et $\varphi : D \to \mathbb{R}^d$ une application lipschitzienne sur D. Alors $\varphi(D)$ est $\mathcal{R}$-négligeable.*

Preuve. L'ensemble D peut être recouvert par un nombre fini n de pavés ouverts R_j° dont la somme des volumes est inférieure à ε. Il est facile de voir que, quitte à modifier n, on peut imposer à ces pavés d'être des cubes, c.-à-d. des pavés dont tous les côtés ont même longueur, soit $R_j = \prod_{i=1}^m [a_{j,i}, b_{j,i}]$, avec $b_{j,i} - a_{j,i} = r_j > 0$ pour tout $i \in [\![1, m]\!]$. Grâce à (4.47), chaque $j \in [\![1, n]\!]$, $\varphi(D \cap R_j)$ est inclus dans un cube R'_j de $\mathbb{R}^d$ dont les côtés ont une longueur $r'_j \leqslant cr_j$, d'où $\mathrm{vol}(R'_j) \leqslant c^d \mathrm{vol}(R_j)$. Par conséquent, $\varphi(D)$ est recouvert par n cubes R'_j dont la somme des volumes est majorée par $c^d\varepsilon$. On en déduit que $\varphi(D)$ est $\mathcal{R}$-négligeable. □

Proposition 4.79. *Pour $1 \leqslant m < d$, soient B une partie bornée de $\mathbb{R}^m$ et $\varphi : B \to \mathbb{R}^d$ lipschitzienne sur B. Alors $\varphi(B)$ est une partie $\mathcal{R}$-négligeable de $\mathbb{R}^d$.*

Preuve. Puisque B est borné dans $\mathbb{R}^m$, il est contenu dans un pavé ouvert R de $\mathbb{R}^m$. Alors pour tout $\varepsilon > 0$, on peut trouver $\delta > 0$ tel que $R \times [0, \delta]^{d-m}$ soit de volume inférieur à ε. Ceci montre que B considéré comme sous-ensemble de $\mathbb{R}^d$, plus précisément $\tilde{B} := B \times \{0\}^{d-m}$, est une partie $\mathcal{R}$-négligeable de $\mathbb{R}^d$. L'application φ se prolonge naturellement en une application lipschitzienne $\tilde{\varphi} : \tilde{B} \to \mathbb{R}^d$ en posant pour tout $x \in B$, $\tilde{\varphi}(x_1, \ldots, x_m, 0, \ldots, 0) := \varphi(x)$. La proposition 4.78 nous dit alors que $\tilde{\varphi}(\tilde{B}) = \varphi(B)$ est une partie $\mathcal{R}$-négligeable de $\mathbb{R}^d$. □

Les propositions 4.78 et 4.79 donnent des conditions suffisantes très pratiques pour vérifier qu'une partie bornée de $\mathbb{R}^d$ est $\mathcal{R}$-négligeable ou admissible.

Par exemple toute courbe dans $\mathbb{R}^d$ ($d > 1$) qui admet un paramétrage $\varphi : [0,1] \to \mathbb{R}^d$ de classe C^1 est $\mathcal{R}$-négligeable dans $\mathbb{R}^d$. Ainsi toute partie bornée de $\mathbb{R}^2$ dont la frontière est une réunion finie de courbes de ce type est admissible. De même, toute partie bornée de $\mathbb{R}^3$ dont la frontière est une réunion finie de surfaces compactes ayant un paramétrage C^1 est admissible. Toute partie bornée de $\mathbb{R}^d$ dont la frontière est incluse dans une réunion finie d'hyperplans est admissible, etc.

4.5.4 Intégrales répétées

La réduction du calcul d'une intégrale sur une partie de $\mathbb{R}^d$ à des intégrations successives relativement à une variable réelle est un outil important pour le calcul d'une telle intégrale. Le théorème de Fubini, qui relève de l'intégration au sens de Lebesgue, permet de traiter cette question de façon satisfaisante. Nous nous contenterons ici d'une version élémentaire, adaptée à l'intégrale de Riemann multiple.

Soient A et B deux pavés fermés dans $\mathbb{R}^{d_1}$ et $\mathbb{R}^{d_2}$ respectivement. Alors $A \times B$ est un pavé fermé de $\mathbb{R}^{d_1+d_2}$. Pour $f : A \times B \to \mathbb{R}$, nous noterons, un peu abusivement, f_x et f_y les applications partielles $f(x, \,.\,)$ et $f(\,.\,, y)$ respectivement :

$$f_x : B \to \mathbb{R}, \quad y \mapsto f_x(y) = f(x,y), \qquad f_y : A \to \mathbb{R}, \quad x \mapsto f_y(x) = f(x,y).$$

Supposons que f soit Riemann intégrable sur $A \times B$. Nous aurons besoin, dans ce qui suit, d'intégrer f_x sur B. Cette fonction n'est pas forcément intégrable sur B pour toutes les valeurs de $x \in A$. Nous supposerons qu'elle l'est, sauf peut-être pour les x d'une partie $\mathcal{R}$-négligeable de A. Lorsque f_x est intégrable sur B, nous noterons pour alléger les écritures :

$$I_B f_x := \int_B f(x,y)\,\mathrm{d}y = \int_B f_x(y)\,\mathrm{d}y.$$

Pour les $x \in A$ pour lesquels f_x n'est pas intégrable sur B, nous attribuerons une valeur arbitraire à $I_B f_x$, mais de façon à ne pas empêcher la fonction $x \mapsto I_B f_x$ d'être bornée sur A, donc le plus simple sera de poser dans ce cas $I_B f_x = 0$. Nous notons $I_B f$ la fonction $x \mapsto I_B f_x$ ainsi définie sur A.

En échangeant les rôles de A et de B dans ce qui précède, on définit de même $I_A f : y \mapsto I_A f_y = \int_A f(x,y)\,\mathrm{d}x$.

Théorème 4.80 (Fubini, produit de pavés)**.** *Soient A et B deux pavés fermés dans $\mathbb{R}^{d_1}$ et $\mathbb{R}^{d_2}$ respectivement. Soit f Riemann intégrable sur $A \times B$. On suppose que $\{x \in A \;;\; f_x \notin \mathcal{R}(B)\}$ est une partie $\mathcal{R}$-négligeable de A. Alors la fonction $I_B f$ est intégrable sur A et*

$$\int_{A\times B} f(z)\,\mathrm{d}z = \int_A \left\{ \int_B f(x,y)\,\mathrm{d}y \right\} \mathrm{d}x. \tag{4.48}$$

De même si $\{y \in B\ ;\ f_y \notin \mathcal{R}(A)\}$ *est une partie* $\mathcal{R}$*-négligeable de* B*, la fonction* $I_A f$ *est intégrable sur* B *et*

$$\int_{A\times B} f(z)\,\mathrm{d}z = \int_B \left\{ \int_A f(x,y)\,\mathrm{d}x \right\} \mathrm{d}y. \tag{4.49}$$

Preuve. Nous nous contenterons de prouver (4.48), la preuve de (4.49) s'en déduit en échangeant les rôles de A et B. On commence par remarquer que toute subdivision Δ de $A \times B$ peut s'écrire comme le produit $\Delta_A \times \Delta_B$ d'une subdivision de A par une subdivision de B. En effet, en notant $d = d_1 + d_2$, $A = \prod_{i=1}^{d_1}[a_i, b_i]$ et $B = \prod_{i=d_1+1}^{d}[a_i, b_i]$, on peut écrire [19] :

$$A \times B = \prod_{i=1}^{d}[a_i, b_i].$$

Une subdivision quelconque Δ du pavé $A \times B$ de $\mathbb{R}^d$ est par définition de la forme $\Delta_1 \times \cdots \times \Delta_d$, où chaque Δ_i est une subdivision de l'intervalle $[a_i, b_i]$ correspondant. Alors

$$\Delta = \underbrace{(\Delta_1 \times \cdots \times \Delta_{d_1})}_{=:\Delta_A} \times \underbrace{(\Delta_{d_1+1} \times \cdots \times \Delta_d)}_{=:\Delta_B}.$$

On vérifie de même que les cellules de Δ sont les produits $Q_A \times Q_B$ d'une cellule de A par une cellule de B.

Avec les notations ci-dessus, nous prétendons que l'encadrement :

$$S_\Delta(f) \leqslant S_{\Delta_A}(I_B f) \leqslant S^{\Delta_A}(I_B f) \leqslant S^\Delta(f), \tag{4.50}$$

est vérifié pour toute subdivision Δ de $A \times B$. Nous détaillons la justification de la première inégalité dans (4.50), laissant la troisième au lecteur.

$$\begin{aligned}
S_\Delta(f) = \sum_{Q\in C(\Delta)} \inf_Q f\,\mathrm{vol}(Q) &= \sum_{\substack{Q_A\in C(\Delta_A)\\ Q_B\in C(\Delta_B)}} \inf_{Q_A\times Q_B} f\ \mathrm{vol}(Q_A)\,\mathrm{vol}(Q_B) \\
&\leqslant \sum_{Q_A\in C(\Delta_A)} \sum_{Q_B\in C(\Delta_B)} \inf_{x\in Q_A} \inf_{Q_B}(f_x)\,\mathrm{vol}(Q_B)\,\mathrm{vol}(Q_A) \\
&\leqslant \sum_{Q_A\in C(\Delta_A)} \inf_{x\in Q_A} \left(\sum_{Q_B\in C(\Delta_B)} \inf_{Q_B}(f_x)\,\mathrm{vol}(Q_B) \right) \mathrm{vol}(Q_A) \\
&= \sum_{Q_A\in C(\Delta_A)} \inf_{x\in Q_A} (I_B f)\,\mathrm{vol}(Q_A) = S_{\Delta_A}(I_B f).
\end{aligned}$$

19. Avec l'abus usuel qui identifie $((x_1, \ldots, x_{d_1}), (x_{d_1+1}, \ldots, x_d))$ avec $(x_1, \ldots, x_d)$.

Puisque f est Riemann intégrable sur $A\times B$, il existe pour tout $\varepsilon>0$, une subdivision Δ de $A\times B$ telle que $\int_{A\times B} f(z)\,\mathrm{d}z-\varepsilon < S_\Delta(f)\leqslant S^\Delta(f) < \int_{A\times B} f(z)\,\mathrm{d}z+\varepsilon$. Compte-tenu de (4.50), on en déduit que

$$\int_{A\times B} f(z)\,\mathrm{d}z-\varepsilon < S_{\Delta_A}(I_Bf)\leqslant S^{\Delta_A}(I_Bf) < \int_{A\times B} f(z)\,\mathrm{d}z+\varepsilon,$$

ce qui établit l'intégrabilité sur A de I_Bf et l'égalité (4.48) par arbitrarité de ε. □

Une récurrence permet d'étendre le théorème 4.80 au cas d'une intégrale sur un produit cartésien de n pavés. En particulier, lorsque tous ces pavés sont des intervalles fermés bornés de $\mathbb{R}$, on obtient le corollaire suivant qui justifie *a posteriori* l'appelation « intégrale multiple » pour une intégrale sur un pavé de $\mathbb{R}^d$.

Corollaire 4.81. *Pour toute $f : P\to\mathbb{R}$ Riemann intégrable sur le pavé $P=\prod_{i=1}^d[a_i,b_i]$,*

$$\int_P f(x)\,\mathrm{d}x=\int_{a_1}^{b_1}\left\{\int_{a_2}^{b_2}\left\{\dots\left\{\int_{a_d}^{b_d} f(x_1,\dots,x_d)\,\mathrm{d}x_d\right\}\dots\right\}\,\mathrm{d}x_2\right\}\,\mathrm{d}x_1, \tag{4.51}$$

pourvu que les intégrales successives apparaissant au second membre aient toutes un sens. Ceci justifie la notation

$$\int_P f(x)\,\mathrm{d}x=\int_{a_1}^{b_1}\int_{a_2}^{b_2}\dots\int_{a_d}^{b_d} f(x_1,\dots,x_d)\,\mathrm{d}x_1\dots\,\mathrm{d}x_d.$$

Dans le théorème 4.80, l'hypothèse « $\{x\in A\ ;\ f_x\notin\mathcal{R}(B)\}$ est une partie $\mathcal{R}$-négligeable de A », même si elle ne pose pas de gros problèmes dans les cas usuels, est quand même un inconvénient. Dans le « vrai » théorème de Fubini relatif à l'intégrale de Lebesgue, la négligeabilité (au sens de Lebesgue) de l'ensemble des x pour lesquels f_x n'est pas intégrable sur B est une conséquence de l'hypothèse d'intégrabilité de f sur $A\times B$. Dans le cadre qui est le nôtre, il y a quand même un cas particulier important où l'intégrabilité de f_x sur B sera automatique pour tout $x\in A$. C'est celui de l'intégration des fonctions f « à variables séparées » sur $A\times B$, appelées encore produit tensoriel $f=g\otimes h$.

Théorème 4.82. *Soient A et B deux parties bornées de $\mathbb{R}^{d_1}$ et $\mathbb{R}^{d_2}$ respectivement. Soient $g:A\to\mathbb{R}$, $h:B\to\mathbb{R}$ et f définie par*

$$f=g\otimes h : A\times B\to\mathbb{R},\quad (x,y)\mapsto f(x,y)=g(x)h(y).$$

Si g et h sont Riemann intégrables sur respectivement A et B, alors f est Riemann intégrable sur $A\times B$ et

$$\int_{A\times B} f(z)\,\mathrm{d}z=\int_{A\times B} g(x)h(y)\,\mathrm{d}x\,\mathrm{d}y=\left\{\int_A g(x)\,\mathrm{d}x\right\}\left\{\int_B h(y)\,\mathrm{d}y\right\}. \tag{4.52}$$

Notons qu'ici l'intégrabilité de f sur $A \times B$ n'est pas dans les hypothèses.

Preuve. Nous nous contenterons de détailler la preuve dans le cas particulier où g et h sont positives. L'extension au cas général se fait par linéarité de l'intégrale en utilisant les décompositions $g = g^+ - g^-$ et $h = h^+ - h^-$.

Nous traitons d'abord le cas particulier où A et B sont des pavés fermés. En conservant les notations de la preuve du théorème 4.80, le point clé est de noter les égalités suivantes.

$$\inf_{Q_A \times Q_B} (g \otimes h) = \Big(\inf_{Q_A} g\Big)\Big(\inf_{Q_B} h\Big), \quad \sup_{Q_A \times Q_B} (g \otimes h) = \Big(\sup_{Q_A} g\Big)\Big(\sup_{Q_B} h\Big). \tag{4.53}$$

Détaillons la vérification de l'inégalité pour les infima, celle pour les suprema étant analogue. Notons pour alléger $m_{A,B}$, m_A et m_B les trois infima apparaissant successivement dans cette formule. D'abord, pour tout $(x, y) \in Q_A \times Q_B$, $g(x) \geqslant m_A \geqslant 0$ et $h(y) \geqslant m_B \geqslant 0$. Comme tout est positif, la multiplication membre à membre de ces deux encadrements nous donne $g(x)h(y) \geqslant m_A m_B$ et ceci valant pour tout (x, y) dans $Q_A \times Q_B$, on en déduit que $m_{A,B} \geqslant m_A m_B$. Pour prouver l'inégalité dans l'autre sens, soit $\varepsilon > 0$ arbitraire. Par définition de m_A et m_B, il existe au moins un couple $(x, y) \in Q_A \times Q_B$ tel que $g(x) < m_A + \varepsilon$ et $h(y) < m_B + \varepsilon$. On en déduit que $m_{A,B} \leqslant f(x, y) = g(x)h(y) < (m_A + \varepsilon)(m_B + \varepsilon)$, puis par arbitrarité de ε, que $m_{A,B} \leqslant m_A m_B$.

En rappelant que $\Delta = \Delta_A \times \Delta_B$, les égalités (4.53) nous permettent de voir facilement que

$$S_\Delta(g \otimes h) = S_{\Delta_A}(g) S_{\Delta_B}(h), \qquad S^\Delta(g \otimes h) = S^{\Delta_A}(g) S^{\Delta_B}(h).$$

Par hypothèse, g est intégrable sur A et h l'est sur B. Notons $I_A g = \int_A g(x)\,\mathrm{d}x \geqslant 0$ et $I_B h = \int_B h(y)\,\mathrm{d}y \geqslant 0$. Ces deux intégrabilités nous assurent pour tout $\varepsilon > 0$ de l'existence de subdivisions Δ_A et Δ_B dépendant de ε, telles que $I_A g - \varepsilon < S_{\Delta_A}(g) \leqslant S^{\Delta_A}(g) < I_A g + \varepsilon$ et $I_B h - \varepsilon < S_{\Delta_B}(h) \leqslant S^{\Delta_B}(h) < I_B h + \varepsilon$. Pour $\Delta = \Delta_A \times \Delta_B$, on en déduit

$$I^*(A \times B, g \otimes h) \leqslant S^\Delta(g \otimes h) = S^{\Delta_A}(g) S^{\Delta_B}(h) < I_A g I_B h + \varepsilon(I_A g + I_B h + \varepsilon).$$

Comme $\varepsilon > 0$ est arbitraire, il en résulte que $I^*(A \times B, g \otimes h) \leqslant I_A g I_B h$. Pour obtenir la minoration $I_*(A \times B, g \otimes h) \geqslant I_A g I_B h$, il faut prendre garde au fait que $I_A g - \varepsilon$ ou $I_B h - \varepsilon$ peut être négatif. Si cela se produit, minorons $S_{\Delta_A}(g) S_{\Delta_B}(h)$ par 0 et sinon par $(I_A g - \varepsilon)(I_B h - \varepsilon)$. On réunifie les deux cas en écrivant que

$$\max\big(0; I_A g I_B h + \varepsilon(I_A g + I_B h - \varepsilon)\big) \leqslant S_{\Delta_A}(g) S_{\Delta_B}(h) = S_\Delta(g \otimes h) \leqslant I_*(A \times B, g \otimes h),$$

d'où en faisant tendre ε vers 0, $I_A g I_B h \leqslant I_*(A \times B, g \otimes h)$. Ceci achève la preuve de l'intégrabilité de $g \otimes h$ sur $A \times B$ et de l'égalité (4.52).

Passons au cas général où A et B sont seulement des parties bornées de $\mathbb{R}^{d_1}$ et $\mathbb{R}^{d_2}$. Soit P un pavé de $\mathbb{R}^{d_1}$ contenant A et R un pavé de $\mathbb{R}^{d_2}$ contenant B. Notons

g_P la fonction égale à g sur P et nulle sur $P\backslash A$ et notons h_R la fonction égale à h sur B et nulle sur $R\backslash B$. Alors

$$\int_A g(x)\,\mathrm{d}x = \int_P g_P(x)\,\mathrm{d}x \quad \text{et} \quad \int_B h(y)\,\mathrm{d}y = \int_R h_R(y)\,\mathrm{d}y. \tag{4.54}$$

En appliquant la partie déjà démontrée du théorème avec le produit $P \times R$, on voit que $g_P \otimes h_R$ est Riemann intégrable sur $P \times R$ et que

$$\int_{P\times R} (g_P \otimes h_R)(x,y)\,\mathrm{d}x\,\mathrm{d}y = \left\{\int_P g_P(x)\,\mathrm{d}x\right\}\left\{\int_R h_R(y)\,\mathrm{d}y\right\}. \tag{4.55}$$

Or il est immédiat de vérifier que $g_P \otimes h_R$ est égale à $g \otimes h$ sur $A \times B$ et est nulle sur $(P \times R)\backslash(A \times B)$. On a donc trouvé un pavé $S = P \times R$ dans $\mathbb{R}^{d_1+d_2}$, contenant $A \times B$ tel que la fonction $f_S = g_P \otimes h_R$, qui est égale à $f = g \otimes h$ sur $A \times B$ et est nulle sur $S\backslash(A \times B)$, soit intégrable sur S. Ceci montre que $f = g \otimes h$ est Riemann intégrable sur $A \times B$ et que

$$\int_{A\times B} f(z)\,\mathrm{d}z = \int_S f_S(z)\,\mathrm{d}z = \int_{P\times R} (g_P \otimes h_R)(x,y)\,\mathrm{d}x\,\mathrm{d}y. \tag{4.56}$$

La formule (4.52) résulte alors des égalités (4.54)–(4.56). □

4.5.5 Changement de variable

Pour le calcul d'intégrales multiples, on dispose d'une formule de changement de variable qui peut être vue comme une généralisation à la dimension d de la proposition 3.34, en remplaçant l'hypothèse de monotonie du changement de variable φ par l'inversibilité C^1 et la dérivée de φ par la valeur absolue de son déterminant jacobien. La preuve de cette formule est trop longue pour trouver place dans cet ouvrage, le lecteur intéressé pourra en trouver une version détaillée dans [9].

Soit U un ouvert de $\mathbb{R}^d$ et $\varphi : U \to \mathbb{R}^d$ une application continûment différentiable sur U (autrement dit de classe C^1 sur U). Pour $i \in [\![1, d]\!]$, notons π_i la projection canonique $\mathbb{R}^d \to \mathbb{R}$, $(y_1, \ldots, y_d) \mapsto y_i$ et notons $\varphi_i = \pi_i \circ \varphi$ l'application coordonnée correspondante de φ. On peut ainsi représenter φ par le vecteur $(\varphi_1, \ldots, \varphi_d)$ d'applications $\mathbb{R}^d \to \mathbb{R}$,

$$\varphi : x = (x_1, \ldots, x_d) \longmapsto y = \varphi(x) = (y_1, \ldots, y_d) = (\varphi_1(x), \ldots, \varphi_d(x)).$$

La matrice jacobienne $J_\varphi(x)$ de φ au point x est la matrice de la différentielle de φ au point x (qui est une application linéaire). On peut expliciter la fonction J_φ à l'aide des dérivéees partielles :

$$J_\varphi = \begin{pmatrix} \dfrac{\partial\varphi_1}{\partial x_1} & \cdots & \dfrac{\partial\varphi_1}{\partial x_d} \\ \vdots & & \vdots \\ \dfrac{\partial\varphi_d}{\partial x_1} & \cdots & \dfrac{\partial\varphi_d}{\partial x_d} \end{pmatrix}$$

On appelle *jacobien* de φ au point x, le déterminant

$$\mathrm{Jac}(\varphi)(x) := \det(J_\varphi(x)).$$

Nous dirons que φ est C^1-inversible sur l'ouvert $V \subset U$ si l'image de V par φ est un ouvert W et si la restriction de φ à V est une bijection dont l'inverse $\varphi^{-1} : W \to V$ est encore C^1. On dit aussi dans ce cas que φ est un C^1-difféomorphisme de V sur W.

Théorème 4.83 (changement de variable)**.** *Soit U un ouvert de $\mathbb{R}^d$ et $\varphi : U \to \mathbb{R}^d$ une application C^1. Soit A une partie admissible de $\mathbb{R}^d$ dont la fermeture $\overline{A}$ est contenue dans U. On suppose de plus que φ est C^1-inversible sur l'intérieur de A. Alors pour toute $f : A \to \mathbb{R}$ Riemann intégrable sur $\varphi(A)$, $f \circ \varphi$ est Riemann intégrable sur A et*

$$\int_{\varphi(A)} f(y)\,\mathrm{d}y = \int_A f(\varphi(x))\,|\mathrm{Jac}(\varphi)(x)|\,\mathrm{d}x. \tag{4.57}$$

4.5.6 Intégrale multiple généralisée

Jusqu'ici, nous nous sommes restreints à l'étude de l'intégration d'une fonction f définie et bornée sur une partie bornée A de $\mathbb{R}^d$. Nous examinons maintenant la situation des intégrales multiples *généralisées*, pour lesquelles f ou A n'est pas bornée. Pour définir ces intégrales, on va les considérer comme limites d'intégrales $\int_{H_n} f(x)\,\mathrm{d}x$ où $(H_n)_{n\geqslant 1}$ est une suite de parties bornées de A, convergeant vers A en un certain sens et telles que f soit intégrable sur chaque H_n. Une extension naturelle de ce que nous avons déjà fait en dimension 1 (définition 4.8) serait de prendre pour H_n une réunion finie de pavés fermés. En fait, nous aurons plus de souplesse en travaillant avec des compacts de $\mathbb{R}^d$ plus généraux.

Définition 4.84. *Soit A une partie de $\mathbb{R}^d$ et $(H_n)_{n\geqslant 1}$ une suite croissante, pour l'inclusion, de parties de A. On dit que $(H_n)_{n\geqslant 1}$ épuise A (ou est exhaustive) si, pour tout compact K de $\mathbb{R}^d$ inclus dans A, il existe un entier n_0 tel que H_{n_0} contienne K.*

Définition 4.85. *Soit A une partie de $\mathbb{R}^d$. Une application $f : A \to \mathbb{R}$ est localement Riemann intégrable sur A, s'il existe une suite croissante de compacts de $\mathbb{R}^d$ épuisant A, sur chacun desquels f soit Riemann intégrable.*

Pour légitimer notre définition de l'intégrale multiple généralisée, nous aurons besoin du résultat suivant.

Proposition 4.86. *Soit A une partie de $\mathbb{R}^d$ et f une application $A \to \mathbb{R}$. On suppose qu'il existe une suite croissante $(K_n)_{n\geqslant 1}$ de compacts de $\mathbb{R}^d$ épuisant A, telle que f soit Riemann intégrable sur chaque K_n et que*

$$\lim_{n\to+\infty} \int_{K_n} |f(x)|\,\mathrm{d}x =: J \in \mathbb{R}_+. \tag{4.58}$$

Alors

$$\lim_{n\to+\infty}\int_{K_n} f(x)\,\mathrm{d}x =: I \in \mathbb{R}. \tag{4.59}$$

De plus, pour toute autre suite croissante $(K'_n)_{n\geqslant 1}$ de compacts de $\mathbb{R}^d$ épuisant A, telle que f soit Riemann intégrable sur chaque K'_n, $\int_{K'_n}|f(x)|\,\mathrm{d}x$ et $\int_{K'_n} f(x)\,\mathrm{d}x$ convergent vers respectivement J et I.

Preuve. Notons $I_n = \int_{K_n} f(x)\,\mathrm{d}x$ et $J_n := \int_{K_n}|f(x)|\,\mathrm{d}x$. Les hypothèses faites assurent l'existence de ces deux intégrales comme intégrales de Riemann sur l'ensemble borné K_n, cf. prop. 4.64-2 c). Notons au passage que $(J_n)_{n\geqslant 1}$ est croissante dans $\mathbb{R}_+$, donc convergente dans $\overline{\mathbb{R}}_+$. Pour que (4.58) soit vérifiée, il suffit donc que la suite $(J_n)_{n\geqslant 1}$ soit bornée dans $\mathbb{R}_+$. Pour établir (4.59), nous allons voir que $(I_n)_{n\geqslant 1}$ hérite du critère de Cauchy vérifié par la suite $(J_n)_{n\geqslant 1}$ convergente dans $\mathbb{R}$. Pour cela, nous allons montrer que pour tous entiers $m > n$, $|I_m - I_n| \leqslant |J_m - J_n| = J_m - J_n$. En effet, en utilisant la proposition 4.67 d'abord avec f, puis avec $|f|$, on obtient :

$$|I_m - I_n| = \left|\int_{K_m\setminus K_n} f(x)\,\mathrm{d}x\right| \leqslant \int_{K_m\setminus K_n}|f(x)|\,\mathrm{d}x = J_m - J_n.$$

Notons $I'_n = \int_{K'_n} f(x)\,\mathrm{d}x$ et $J'_n := \int_{K'_n}|f(x)|\,\mathrm{d}x$. Par croissance, $(J'_n)_{n\geqslant 1}$ converge au moins dans $\overline{\mathbb{R}}_+$. Notons J' sa limite éventuellement infinie. Par exhaustivité des deux suites $(K_n)_{n\geqslant 1}$ et $(K'_n)_{n\geqslant 1}$, on peut construire par récurrence une suite strictement croissante d'entiers $(n_j)_{j\geqslant 1}$ telle que :

$$\forall j \geqslant 1, \quad K_{n_{2j-1}} \subset K'_{n_{2j}} \subset K_{n_{2j+1}}.$$

En faisant tendre j vers l'infini dans l'encadrement $J_{n_{2j-1}} \leqslant J'_{n_{2j}} \leqslant J_{n_{2j+1}}$, on obtient $J \leqslant J' \leqslant J$, ce qui établit la convergence de *toute* la suite *croissante* $(J'_n)_{n\geqslant 1}$ vers le réel J. De là on déduit comme ci-dessus que la suite $(I'_n)_{n\geqslant 1}$ est de Cauchy dans $\mathbb{R}$ donc convergente vers $I' \in \mathbb{R}$. Enfin, on note que $|I'_{n_{2j}} - I_{n_{2j-1}}| \leqslant J'_{n_{2j}} - J_{n_{2j-1}}$ et comme ce majorant tend vers 0 quand j tend vers l'infini, on en déduit que $I' = I$. □

Remarque 4.87. Si on laisse tomber l'hypothèse (4.58), alors le comportement asymptotique de la suite des intégrales $\int_{K_n} f(x)\,\mathrm{d}x$ peut dépendre fortement du choix de la suite exhaustive $(K_n)_{n\geqslant 1}$, voir l'exercice 4.8.

Nous pouvons maintenant donner une première définition d'intégrale multiple généralisée.

Définition 4.88 (intégrale multiple généralisée)**.** *Soit A une partie de $\mathbb{R}^d$ et f une application $A \to \mathbb{R}$ telles que A ou f ne soit pas bornée. On suppose qu'il existe une suite croissante $(K_n)_{n\geqslant 1}$ de compacts de $\mathbb{R}^d$ épuisant A, telle que f soit Riemann intégrable sur chaque K_n et que*

$$\sup_{n\geqslant 1}\int_{K_n}|f(x)|\,\mathrm{d}x < +\infty. \tag{4.60}$$

On dit alors que les intégrales généralisées de $|f|$ et f sur A sont convergentes et on les définit par

$$\int_A |f(x)|\,\mathrm{d}x := \lim_{n\to+\infty} \int_{K_n} |f(x)|\,\mathrm{d}x, \quad \int_A f(x)\,\mathrm{d}x := \lim_{n\to+\infty} \int_{K_n} f(x)\,\mathrm{d}x. \tag{4.61}$$

Il importe de noter que la convergence d'intégrale généralisée ainsi définie est une convergence *absolue*. Il n'est pas question ici de définir d'intégrale généralisée $\int_A f(x)\,\mathrm{d}x$ convergente sans que $\int_A |f(x)|\,\mathrm{d}x$ ne le soit également. Ceci contraste avec ce que nous avons vu précédemment pour l'intégrale généralisée « classique » dans le cas $d = 1$. Ce contraste entre la dimension $d = 1$, où la structure d'ordre de l'ensemble de sommation joue un rôle essentiel, et la dimension $d \geqslant 2$ est analogue à ce que nous avons vu pour les séries et les séries doubles.

L'intégrale multiple généralisée définie ci-dessus hérite, par passage à la limite, de toutes les propriétés énoncées à la proposition 4.64, à l'exception de celle relative au produit : la convergence de $\int_A f(x)\,\mathrm{d}x$ et $\int_A g(x)\,\mathrm{d}x$ n'implique plus celle de $\int_A f(x)g(x)\,\mathrm{d}x$. Néanmoins cette dernière converge pourvu que $\int_A f(x)^2\,\mathrm{d}x$ et $\int_A g(x)^2\,\mathrm{d}x$ convergent et dans ce cas, l'inégalité de Cauchy-Schwarz est vérifiée.

On peut se satisfaire de la définition 4.88 lorsque A est le seul ensemble sur lequel on a besoin d'intégrer. Insistons sur le fait que, dans cette définition, f est *définie en tout point de A*. Il convient aussi de remarquer que si f vérifie la définition 4.88, elle doit être bornée sur tout compact C inclus dans A puisque C est inclus dans un K_n sur lequel f est Riemann intégrable, donc bornée. Les situations typiques relevant de cette définition se produisent lorsque f tend vers l'infini en des points de la frontière de A ou lorsque A n'est pas borné (les deux aspects n'étant pas incompatibles). Les ennuis apparaissent lorsque l'on essaie de prolonger f en dehors de A, en notant $\tilde{f}$ la fonction égale à f sur A et à 0 en dehors de A. Si f tend vers l'infini en certains points de la frontière de A, en prenant pour compact C une boule fermée centrée en l'un de ces points, $\tilde{f}$ ne sera pas bornée sur C et donc $\int_{\mathbb{R}^d} \tilde{f}(x)\,\mathrm{d}x$ ne pourra pas être convergente au sens de la définition 4.88. Ceci se produit par exemple dans $\mathbb{R}^2$ avec en prenant pour A le disque unité privé de son centre et $f : A \to \mathbb{R}_+$, $(s,t) \mapsto f(s,t) = (s^2+t^2)^{-1/2}$ (exercice).

Or il est courant, pour construire une densité de loi d'un vecteur aléatoire sur $\mathbb{R}^d$, de partir d'une fonction f positive dont l'intégrale généralisée sur une certaine partie A de $\mathbb{R}^d$ converge et vaut 1, de prolonger f à $\mathbb{R}^d$ par $\tilde{f}$ et d'écrire que $\int_{\mathbb{R}^d} \tilde{f}(x)\,\mathrm{d}x = 1$. Comme la définition 4.88 ne nous permet pas de faire cela lorsque f n'est pas bornée sur A, nous allons la compléter comme suit, ce procédé étant sans objet lorsque f vérifie la définition 4.88 avec $A = \mathbb{R}^d$.

Définition 4.89 (intégrale multiple généralisée sur $\mathbb{R}^d$)**.** *Soit f une application $\mathbb{R}^d \to \mathbb{R}$, non bornée sur $\mathbb{R}^d$. Pour tout entier $N \geqslant 1$, on note f_N la fonction définie sur $\mathbb{R}^d$ par*

$$f_N(x) = \begin{cases} -N & \text{si } f(x) \leqslant -N, \\ f(x) & \text{si } |f(x)| < N, \\ N & \text{si } f(x) \geqslant N. \end{cases} \tag{4.62}$$

On suppose de plus que pour tout $N \geqslant 1$, l'intégrale généralisée $\int_{\mathbb{R}^d} f_N(x)\,\mathrm{d}x$ converge au sens de la définition 4.88 et que

$$\sup_{N\geqslant 1} \int_{\mathbb{R}^d} |f_N(x)|\,\mathrm{d}x < +\infty. \tag{4.63}$$

Alors la suite $(\int_{\mathbb{R}^d} f_N(x)\,\mathrm{d}x)_{N\geqslant 1}$ converge dans $\mathbb{R}$ lorsque N tend vers l'infini et sa limite est encore appelée intégrale généralisée de f sur $\mathbb{R}^d$ et notée $\int_{\mathbb{R}^d} f(x)\,\mathrm{d}x$.

La convergence dans $\mathbb{R}$ de $(\int_{\mathbb{R}^d} f_N(x)\,\mathrm{d}x)_{N\geqslant 1}$ se vérifie facilement pour f positive car alors $(f_N)_{N\geqslant 1}$ est croissante, ce qui implique la croissance de $(\int_{\mathbb{R}^d} f_N(x)\,\mathrm{d}x)_{N\geqslant 1}$. Cette suite croissante de réels est majorée en raison de (4.63), donc elle converge dans $\mathbb{R}$. Le cas général se ramène au cas où f est positive en décomposant $f = f^+ - f^-$. Ce raisonnement utilise implicitement les propriétés de croissance et de linéarité de l'intégrale généralisée au sens de la définition 4.88, héritées de la proposition 4.64.

Pour pouvoir exploiter la définition 4.89 afin d'obtenir des intégrales généralisées sur $\mathbb{R}^d$, il nous reste à donner des conditions pratiques permettant de passer de l'intégrale généralisée $\int_A f(x)\,\mathrm{d}x$ à $\int_{\mathbb{R}^d} \tilde{f}(x)\,\mathrm{d}x$.

Proposition 4.90. *Soit A une partie de $\mathbb{R}^d$ et $f : A \to \mathbb{R}^d$ bornée telle que l'intégrale généralisée $\int_A f(x)\,\mathrm{d}x$ converge au sens de la définition 4.88. On suppose de plus que la frontière de A est localement $\mathcal{R}$-négligeable, autrement dit que $P \cap \partial A$ est $\mathcal{R}$-négligeable pour tout pavé fermé P de $\mathbb{R}^d$. On note $\tilde{f}$ la fonction $\mathbb{R}^d \to \mathbb{R}$, égale à f sur A et nulle sur $\mathbb{R}^d \backslash A$. Alors $\int_{\mathbb{R}^d} \tilde{f}(x)\,\mathrm{d}x$ converge au sens de la définition 4.88 et*

$$\int_{\mathbb{R}^d} \tilde{f}(x)\,\mathrm{d}x = \int_A f(x)\,\mathrm{d}x.$$

Preuve. Remarquons en préliminaire que si $P \cap \partial A$ est $\mathcal{R}$-négligeable, alors $\partial(P \cap A)$ l'est aussi. En effet, d'une part comme P est fermé contenant $A \cap P$, il contient aussi $\partial(A \cap P)$. D'autre part, $\partial(A \cap P) \subset (\partial A) \cup (\partial P)$ d'après (4.46-c). On en déduit que

$$\partial(A \cap P) \;\subset\; P \cap \big((\partial A) \cup (\partial P)\big) = (P \cap \partial A) \cup (P \cap \partial P) = (P \cap \partial A) \cup \partial P.$$

Il suffit prouver la proposition dans le cas où f est positive. Nous notons $M := \sup_{x\in\mathbb{R}^d} f(x) < +\infty$. Par hypothèse, il existe une suite croissante $(K_n)_{n\geqslant 1}$ de compacts de $\mathbb{R}^d$ épuisant A telle que chaque K_n soit inclus dans A, que f soit Riemann intégrable sur chaque K_n et

$$\int_A f(x)\,\mathrm{d}x = \lim_{n\to\infty} \int_{K_n} f(x)\,\mathrm{d}x. \tag{4.64}$$

Pour prouver la proposition, on commence par montrer que $\tilde{f}$ est Riemann intégrable sur tout pavé P et que $\int_P \tilde{f}(x)\,\mathrm{d}x \leqslant \int_A f(x)\,\mathrm{d}x$.

Le cas où $P \subset A^{\mathrm{ext}}$ est trivial puisqu'alors $\tilde{f}$ est nulle sur P. Si $P \subset A^\circ$, alors P est un compact inclus dans A et il existe par exhaustivité un K_n compact tel que

$P \subset K_n \subset A$ et f soit Riemann intégrable sur K_n. Comme P est admisible, f est aussi Riemann intégrable sur P d'après le corollaire 4.76 appliqué avec P et K_n et

$$\int_P \tilde{f}(x)\,\mathrm{d}x = \int_P f(x)\,\mathrm{d}x \leqslant \int_{K_n} f(x)\,\mathrm{d}x \leqslant \int_A f(x)\,\mathrm{d}x.$$

On suppose dans la suite que P a une intersection non vide avec ∂A. Fixons ε arbitraire. Alors il existe une famille finie de pavés ouverts $\{R_i^\circ\ ;\ i \in [\![1,k]\!]\}$, telle que $H = \bigcup_{i=1}^k R_i^\circ$ recouvre $P \cap \partial A$ et $\sum_{i=1}^k \mathrm{vol}(R_i^\circ) < \varepsilon$.

L'ensemble $B = P \cap H^c \cap A$ est *fermé*, éventuellement vide. En effet si $(v_n)_{n\geqslant 1}$ est une suite de points de B convergente vers un $v \in \mathbb{R}^d$, alors v appartient au fermé $P \cap H^c \cap \overline{A}$. Mais comme H recouvre $P \cap \partial A$, v ne peut appartenir à $\partial A = \overline{A}\backslash A^\circ$. Donc v est dans A°, d'où $v \in P \cap H^c \cap A$. Comme B est aussi borné, c'est un compact de $\mathbb{R}^d$ inclus dans A. Alors par exhaustivité, $B \subset K_n \subset A$ pour un certain n et f est Riemann intégrable sur K_n. On remarque alors que B est admissible car

$$\partial B = \partial\big(H^c \cap (P \cap A)\big) \ \subset\ \big(\partial(H^c)\big) \cup \partial(P \cap A) = (\partial H) \cup \partial(P \cap A)$$

et $\partial(P \cap A)$ est $\mathcal{R}$-négligeable comme nous l'avons noté en préliminaire. D'après le corollaire 4.76 appliqué avec B et K_n, f est Riemann intégrable sur B et

$$\int_B f(x)\,\mathrm{d}x = \int_{K_n} f(x)\mathbf{1}_B(x)\,\mathrm{d}x.$$

Par conséquent, la fonction $f_{P,B}$, égale à f sur B et nulle sur $P\backslash B$, est Riemann intégrable sur le pavé P et $\int_P f_{P,B}(x)\,\mathrm{d}x = \int_B f(x)\,\mathrm{d}x$. Il existe donc une subdivision Δ de P telle que

$$\int_B f(x)\,\mathrm{d}x - \varepsilon < S_\Delta(f_{P,B}) \leqslant S^\Delta(f_{P,B}) < \int_B f(x)\,\mathrm{d}x + \varepsilon.$$

Quitte à raffiner Δ, on peut toujours s'arranger pour que la somme des volumes des cellules de Δ nécessaires pour recouvrir $P \cap H$ ait un volume inférieur à 2ε, cf. le lemme 4.72. La contribution totale de ces cellules à $S_\Delta(\tilde{f})$ et à $S^\Delta(\tilde{f})$ est alors majorée par $2M\varepsilon$. En recollant les morceaux, on obtient :

$$\int_B f(x)\,\mathrm{d}x - (2M+1)\varepsilon < I_*(\tilde{f}) \leqslant I^*(\tilde{f}) < \int_B f(x)\,\mathrm{d}x + (2M+1)\varepsilon.$$

On en déduit que $0 \leqslant I^*(\tilde{f}) - I_*(\tilde{f}) \leqslant (4M+2)\varepsilon$, puis par arbitrarité de ε que $I_*(\tilde{f}) = I^*(\tilde{f})$, ce qui établit l'intégrabilité de $\tilde{f}$ sur P.

De plus $\int_P \tilde{f}(x)\,\mathrm{d}x = I^*(\tilde{f}) < \int_B f(x)\,\mathrm{d}x + (2M+1)\varepsilon$, où B dépend de ε, mais $\int_B f(x)\,\mathrm{d}x \leqslant \int_{K_n} f(x)\,\mathrm{d}x \leqslant \int_A f(x)\,\mathrm{d}x$. Il en résulte donc par arbitrarité de ε que $\int_P \tilde{f}(x)\,\mathrm{d}x \leqslant \int_A f(x)\,\mathrm{d}x$.

Ayant examiné tous les cas possibles pour P, nous avons ainsi établi que $\tilde{f}$ est Riemann intégrable sur tout pavé fermé P de $\mathbb{R}^d$ et que $\int_P \tilde{f}(x)\,\mathrm{d}x \leqslant \int_A f(x)\,\mathrm{d}x$. Il en résulte que

$$\sup_P \int_P \tilde{f}(x)\,\mathrm{d}x \leqslant \int_A f(x)\,\mathrm{d}x,$$

le supremum étant pris sur tous les pavés fermés P de $\mathbb{R}^d$. En prenant une suite croissante de pavés $(P_n)_{n\geqslant 1}$ épuisant $\mathbb{R}^d$, on en déduit que l'intégrale généralisée $\int_{\mathbb{R}^d} \tilde{f}(x)\,\mathrm{d}x$ converge dans $\mathbb{R}_+$ et que

$$\int_{\mathbb{R}^d} \tilde{f}(x)\,\mathrm{d}x \leqslant \int_A f(x)\,\mathrm{d}x.$$

Pour prouver l'inégalité dans l'autre sens, soit à nouveau $\varepsilon > 0$ arbitraire. La convergence (4.64) nous fournit un K_{n_0} tel que $\int_{K_{n_0}} f(x)\,\mathrm{d}x > \int_A f(x)\,\mathrm{d}x - \varepsilon$. Alors pour P pavé fermé contenant K_{n_0}, on a

$$\int_{\mathbb{R}^d} \tilde{f}(x)\,\mathrm{d}x \geqslant \int_P \tilde{f}(x)\,\mathrm{d}x \geqslant \int_{K_{n_0}} \tilde{f}(x)\,\mathrm{d}x = \int_{K_{n_0}} f(x)\,\mathrm{d}x > \int_A f(x)\,\mathrm{d}x - \varepsilon,$$

d'où par arbitrarité de ε, $\int_{\mathbb{R}^d} \tilde{f}(x)\,\mathrm{d}x \geqslant \int_A f(x)\,\mathrm{d}x$. □

L'extension du théorème de Fubini 4.80 aux intégrales généralisées est assez problématique. En revanche le théorème 4.82 se généralise facilement.

Théorème 4.91. *Soient A et B deux parties de $\mathbb{R}^{d_1}$ et $\mathbb{R}^{d_2}$ respectivement. Soient $g : A \to \mathbb{R}$, $h : B \to \mathbb{R}$ et f définie par*

$$f = g \otimes h : A \times B \to \mathbb{R}, \quad (x, y) \mapsto f(x, y) = g(x)h(y).$$

Si g et h ont des intégrales généralisées convergentes sur respectivement A et B, alors f a une intégrale généralisée convergente sur $A \times B$ et

$$\int_{A\times B} f(z)\,\mathrm{d}z = \int_{A\times B} g(x)h(y)\,\mathrm{d}x\,\mathrm{d}y = \left\{\int_A g(x)\,\mathrm{d}x\right\}\left\{\int_B h(y)\,\mathrm{d}y\right\}. \tag{4.65}$$

Preuve. Il suffit d'examiner le cas où g et h, donc aussi f, sont positives. Soient $(K_n)_{n\geqslant 1}$ et $(K'_n)_{n\geqslant 1}$ les suites croissantes exhaustives de compacts associées à A et à B respectivement par la définition de la convergence des intégrales généralisées de g et de h. On obtient le résultat en appliquant le théorème 4.82 sur $K_n \times K'_n$ et en faisant tendre n vers l'infini. Le point clé est de remarquer que chaque $K_n \times K'_n$ est un compact pour la topologie produit dans $\mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$, cet espace étant assimilé à $\mathbb{R}^{d_1+d_2}$ et que la suite croissante $(K_n \times K'_n)_{n\geqslant 1}$ épuise $A \times B$. Vérifions cette dernière affirmation. Soit C un compact de $\mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$, inclus dans $A \times B$. Les projections canoniques $\pi_1 : \mathbb{R}^{d_1} \times \mathbb{R}^{d_2} \to \mathbb{R}^{d_1}$ et $\pi_2 : \mathbb{R}^{d_1} \times \mathbb{R}^{d_2} \to \mathbb{R}^{d_2}$ étant *continues*, $\pi_1(C)$ et $\pi_2(C)$ sont compacts inclus respectivement dans A et B. On a donc par exhaustivité $\pi_1(C) \subset K_{n_1}$ et $\pi_2(C) \subset K'_{n_2}$ pour certains indices n_1 et n_2 et par croissance, $\pi_1(C) \times \pi_2(C) \subset K_n \times K'_n$ pour $n = \max(n_1, n_2)$. On conclut en notant que C est inclus dans $\pi_1(C) \times \pi_2(C)$. □

4.6 Exercices

Ex. 4.1 *Fonction Gamma*

On définit la fonction Gamma par :

$$\Gamma : a \mapsto \int_0^{+\infty} t^{a-1}\mathrm{e}^{-t}\,\mathrm{d}t.$$

1) Précisez l'ensemble de définition de Γ. Calculez $\Gamma(1)$.

2) Montrez que pour tout réel $a > 0$, $\Gamma(a+1) = a\Gamma(a)$. En déduire que pour tout $n \in \mathbb{N}^*$, $\Gamma(n) = (n-1)!$.

3) Sachant que $\int_{-\infty}^{+\infty} \mathrm{e}^{-x^2}\,\mathrm{d}x = \pi^{1/2}$, vérifiez que $\Gamma(1/2) = \pi^{1/2}$. Calculez $\Gamma(6{,}5)$.

4) Exprimez à l'aide de la fonction Γ les moments d'ordre pair de la loi $\mathfrak{N}(0,1)$:

$$m_{2k} = \frac{2}{\sqrt{2\pi}} \int_0^{+\infty} x^{2k}\mathrm{e}^{-x^2/2}\,\mathrm{d}x, \quad k \in \mathbb{N}^*.$$

Calculez les explicitement en fonction de k.

Ex. 4.2 *Loi de Poisson et fonction Gamma incomplète*

Soit $\alpha > 0$ un réel, on définit :

$$p(k,\alpha) := \frac{\mathrm{e}^{-\alpha}\alpha^k}{k!}, \quad k \in \mathbb{N}.$$

Autrement dit, $p(k,\alpha)$ est la probabilité qu'une variable aléatoire suivant la loi de Poisson de paramètre α, prenne la valeur k, cf. chapitre 6. Montrez que

$$\forall n \in \mathbb{N}, \quad \sum_{k=0}^{n} p(k,\alpha) = \frac{1}{n!}\int_\alpha^{+\infty} \mathrm{e}^{-t}t^n\,\mathrm{d}t.$$

Ex. 4.3 *La queue de la loi normale*

Pour tout réel x, on note

$$I(x) := \int_x^{+\infty} \exp\left(\frac{-t^2}{2}\right)\mathrm{d}t.$$

Cette intégrale est égale à $\sqrt{2\pi}(1-\Phi(x))$, où Φ est la fonction de répartition de la loi normale standard $\mathfrak{N}(0,1)$, cf. chapitre 6. On se propose de trouver un encadrement de $I(x)$ pour $x > 0$.

1) Justifiez l'existence de $I(x)$.

2) Calculez pour tout entier k la dérivée de la fonction $t \mapsto t^{-2k-1}\exp(-t^2/2)$.

3) Utilisez ce calcul de dérivée pour établir pour $x > 0$ les formules :

$$I(x) = \frac{1}{x}\exp\left(\frac{-x^2}{2}\right) - \int_x^{+\infty} \frac{1}{t^2}\exp\left(\frac{-t^2}{2}\right)\mathrm{d}t$$

$$I(x) = \left(\frac{1}{x} - \frac{1}{x^3}\right)\exp\left(\frac{-x^2}{2}\right) + \int_x^{+\infty} \frac{3}{t^4}\exp\left(\frac{-t^2}{2}\right)\mathrm{d}t.$$

4) En déduire que

$$\forall x > 0, \quad \left(\frac{1}{x} - \frac{1}{x^3}\right) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) < 1 - \Phi(x) < \frac{1}{x} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right). \quad (4.66)$$

Ex. 4.4 *Comment construire la table des valeurs de Φ ?*

Utiliser (4.66) pour donner une méthode pratique de construction d'une table des valeurs de Φ avec une précision au moins égale à 10^{-d}. *Indication.* On considérera le développement en série de Taylor de la densité gaussienne standard sur un intervalle compact $[-a, +a]$ convenablement choisi et on remarquera qu'il s'agit d'une série alternée. Il est donc facile d'encadrer les sommes partielles et de majorer l'erreur commise en remplaçant la série par l'une de ses sommes partielles. On est ainsi ramené à un calcul de polynôme.

Ex. 4.5 *Propriétés de la frontière*

Démontrez les propriétés (4.46).

Solution p. 423

Ex. 4.6 *Frontière d'un pavé*

Soit $P = [a_1, b_1] \times \cdots \times [a_d, b_d]$ un pavé fermé.

1) Montrez que $\partial P = \{x \in P \ ; \ \exists i_0 \in [\![1, d]\!], \ x_{i_0} = a_{i_0} \text{ ou } b_{i_0}\}$.

2) En déduire que ∂P est $\mathcal{R}$-négligeable.

Ex. 4.7 *Admissibilité*

1) Soit $f : [a, b] \to \mathbb{R}$ une fonction continue. Montrez que son graphe $C = \{(x, y) \in \mathbb{R}^2 \ ; \ y = f(x)\}$ est $\mathcal{R}$-négligeable dans $\mathbb{R}^2$.

2) Soient f_1 et f_2 deux fonctions continues $[a, b] \to \mathbb{R}$ telles que $f_1 \leqslant f_2$. Soit $A := \{(x, y) \in \mathbb{R}^2 \ ; \ f_1(x) \leqslant y \leqslant f_2(x)\}$. Montrez que A est admissible.

Indications p. 424

Ex. 4.8 Soit $g : \mathbb{R} \to \mathbb{R}$ définie par $g(s) = s$ si $|s| < 1$ et $g(s) = 1/s$ si $|s| \geqslant 1$. On définit $f : \mathbb{R}^2 \to \mathbb{R}$, $x = (s, t) \mapsto f(x) = g(s)g(t)$.

1) Calculez, pour $a < -1$ et $b > 1$, l'intégrale $\int_{[a,b]^2} f(x)\,\mathrm{d}x$.

2) Trouvez une suite croissante de compacts K_n épuisant $\mathbb{R}^2$ telle que la condition (4.58) ne soit pas vérifiée. Peut-elle l'être avec une autre suite convenablement choisie ?

3) Montrez que pour chaque $\ell \in \overline{\mathbb{R}}$, on peut construire une suite croissante de compacts K_n épuisant $\mathbb{R}^2$, telle que f soit Riemann intégrable sur chaque K_n et que $\int_{K_n} f(x)\,\mathrm{d}x$ converge vers ℓ. Montrez aussi que si l'on se donne un ensemble fini de réels $\ell_1, \ldots \ell_k$, on peut aussi construire $(K_n)_{n\geqslant 1}$ telle que chaque ℓ_j soit limite d'une sous-suite de $(\int_{K_n} f(x)\,\mathrm{d}x)_{n\geqslant 1}$.

Ex. 4.9 *Fonction Bêta*

La fonction Bêta [20] est définie par

$$B :]0, +\infty[^2 \longrightarrow \mathbb{R}, \quad (a, b) \longmapsto B(a, b) := \int_0^1 t^{a-1}(1-t)^{b-1}\,\mathrm{d}t.$$

20. La lettre Bêta majuscule de l'alphabet grec s'écrit B.

1) Prouvez que

$$B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}.$$

2) En déduire la valeur de $B(m,n)$ pour m et n entiers.

Indication p. 424

Ex. 4.10 *Compatibilité des définitions de l'intégrale généralisée*

Soit f telle que $I = \int_{-\infty}^{+\infty} f(x)\,\mathrm{d}x$ converge *absolument* au sens des définitions 4.8 et 4.25. Montrez qu'alors f est aussi intégrable au sens de la définition 4.88 (appliquée avec $d = 1$) si f est bornée sur $\mathbb{R}$, ou de la définition 4.89 sinon et que $I = \int_{\mathbb{R}} f(x)\,\mathrm{d}x$.

Solution p. 424

Ex. 4.11 *Espérance et fonction de survie*

Le titre de cet exercice s'éclairera au chapitre 7. Soit $f : \mathbb{R}_+ \to \mathbb{R}_+$ telle que $\int_0^{+\infty} f(x)\,\mathrm{d}x$ converge dans $\mathbb{R}_+$. On définit $G : \mathbb{R}_+ \to \mathbb{R}_+$, $t \mapsto G(t) := \int_t^{+\infty} f(x)\,\mathrm{d}x$. Alors G est décroissante sur $\mathbb{R}_+$, donc en particulier localement Riemann intégrable sur $\mathbb{R}_+$. En fait, G est aussi continue (cf. la preuve du th. 3.32 et de la prop. 6.26).

1) On suppose que f est Riemann intégrable sur $[a,b] \subset \mathbb{R}_+$ et on note $h : [a,b]^2 \to \mathbb{R}_+$, $(x,t) \mapsto h(x,t) := f(x)\mathbf{1}_{[0,x]}(t)$. Vérifiez que h est Riemann intégrable sur le pavé $[a,b]^2$ et montrez en calculant de deux façons $\int_{[a,b]^2} h(x,t)\,\mathrm{d}x\,\mathrm{d}t$ que

$$\int_a^b x f(x)\,\mathrm{d}x = \int_a^b G(t)\,\mathrm{d}t + aG(a) - bG(b).$$

Dans le cas où f est continue, voyez-vous un autre moyen d'obtenir cette formule ?

2) Étendez cette formule au cas où $\int_a^b f(x)\,\mathrm{d}x$ est généralisée en a ou en b, mais Riemann intégrable sur tout $[a',b'] \subset]a,b[$.

3) Montrez que si $\int_0^{+\infty} G(t)\,\mathrm{d}t < +\infty$ ou si $\int_0^{+\infty} x f(x)\,\mathrm{d}x < +\infty$, $nG(n)$ tend vers zéro quand n tend vers l'infini.

4) Montrez que

$$\int_0^{+\infty} x f(x)\,\mathrm{d}x = \int_0^{+\infty} G(t)\,\mathrm{d}t, \quad \text{égalité dans } \overline{\mathbb{R}}_+.$$

Solution p. 426

Chapitre 5

Évènements et probabilités

Ce chapitre débute (enfin ?) l'étude de la théorie des probabilités. On y verra les notions classiques d'espace probabilisé, de probabilité conditionnelle et d'indépendance d'évènements. En préambule, on met en place des rudiments de théorie de la mesure (essentiellement du vocabulaire) en vue de pouvoir disposer d'emblée d'exemples de probabilités sortant du cadre des probabilités discrètes. Nous pourrons ainsi modéliser l'expérience aléatoire consistant à choisir un point *au hasard* sur le segment $[0,1]$. Le langage de la mesure est présenté dans la première section. La deuxième section contient une approche informelle de la modélisation de l'aléatoire avant de présenter à la section 3 les axiomes de la théorie des probabilités et les propriétés générales des probabilités. On examine ensuite les notions de conditionnement et d'indépendance.

5.1 Notion de mesure

Soit Ω un ensemble, on note $\mathcal{P}(\Omega)$ l'ensemble des parties de Ω. Une mesure sur Ω est une *fonction d'ensembles* m qui, à certaines parties A de Ω, associe un réel positif $m(A)$ appelé mesure de A. La théorie que nous allons présenter fournit un cadre mathématique commun pour des notions comme le dénombrement, la mesure des grandeurs géométriques (longueur, aire, volume), physiques (masse) et les probabilités. À partir du concept de mesure, on peut bâtir une intégrale [1] sur l'ensemble Ω permettant d'unifier les notions d'intégrale simple ou multiple au sens classique, de série absolument convergente et d'espérance mathématique d'une variable aléatoire. Il est commode d'élargir d'emblée l'ensemble d'arrivée de m à $\overline{\mathbb{R}}_+$ au lieu de $\mathbb{R}_+$, par exemple pour pouvoir dire que l'aire d'un quart de plan est $+\infty$.

Intuitivement, en pensant par exemple à la mesure des aires, les propriétés minimales que l'on puisse exiger de m sont

a) la *croissance* : si $A \subset B$, $m(A) \leqslant m(B)$,

1. Appelée « intégrale de Lebesgue ». Sa construction n'est pas au programme de cet ouvrage.

b) l'*additivité* : si $A \cap B = \varnothing$, $m(A \cup B) = m(A) + m(B)$, sous réserve que $m(A)$, $m(B)$ et $m(A \cup B)$ soient définies.

Il est facile de voir que la croissance est une conséquence de l'additivité en écrivant si $A \subset B$, $B = A \cup (B \backslash A)$ et $m(B) = m(A) + m(B \backslash A) \geqslant m(A)$, avec les mêmes réserves d'existence. L'additivité s'étend, par une récurrence immédiate, aux suites finies $A_1, \ldots, A_n$, donnant *l'additivité finie* : $m(A_1 \cup \cdots \cup A_n) = m(A_1) + \cdots + m(A_n)$ si les A_i sont deux à deux disjoints. En revanche elle ne s'étend pas automatiquement aux suites infinies d'ensembles deux à deux disjoints.

Pour avoir une théorie assez riche, on doit pouvoir effectuer certains passages à la limite, par exemple pour pouvoir mesurer l'aire d'un disque dans le plan en le « pavant » par des carreaux à côtés parallèles aux axes. La propriété correspondante est appelée *σ-additivité* :

$$\text{si les } A_n \text{ sont deux à deux disjoints,} \quad m\Big(\bigcup_{n\in\mathbb{N}} A_n\Big) = \sum_{n=0}^{+\infty} m(A_n),$$

pourvu que toutes ces quantités soient définies.

Pourquoi ces clauses restrictives sur l'existence des $m(A)$? Il se trouve que dans la plupart des cas intéressants (sauf lorsque Ω est au plus dénombrable), on ne sait pas définir $m(A)$ pour tout $A \in \mathcal{P}(\Omega)$. Souvent, pour construire une mesure m, on commence par attribuer une valeur $m(A)$ à chaque A dans une famille $\mathcal{C}$ bien particulière de parties de Ω. Par exemple si $\Omega = \mathbb{R}$, on peut prendre pour $\mathcal{C}$ la famille des intervalles $]a, b]$ et « décider » que $m(]a, b]) := b - a$, choisissant ainsi de mesurer $]a, b]$ par sa longueur [2]. De même si $\Omega = \mathbb{R}^2$, on peut partir de la famille $\mathcal{C}$ des « rectangles » $R =]a, b] \times]c, d]$ et décider que $m(R) := (b - a)(d - c)$. Dans un deuxième temps, on essaie de prolonger m à une famille $\mathcal{F}$ de parties de Ω plus grande que $\mathcal{C}$, tout en préservant la σ-additivité. Il se trouve qu'il n'est pas toujours possible de réaliser ce prolongement jusqu'à prendre $\mathcal{F} = \mathcal{P}(\Omega)$. Certaines parties de $\mathbb{R}$ sont d'une trop grande complexité pour qu'on puisse leur attribuer une « longueur ». Ainsi la fonction d'ensembles m a un *ensemble de définition* qui est une sous-famille de $\mathcal{P}(\Omega)$. Cet ensemble de définition de m est ce que l'on appelle une *tribu* de parties de Ω.

Il est temps maintenant de formaliser les définitions suivantes.

Définition 5.1 (tribu)**.** *Une famille $\mathcal{F}$ de parties de Ω est appelée* tribu *(ou σ-algèbre) sur Ω si elle*

a) possède l'ensemble vide : $\varnothing \in \mathcal{F}$;

b) est stable par passage au complémentaire : $\forall A \in \mathcal{F}$, $A^c \in \mathcal{F}$;

c) est stable par union dénombrable : $(\forall i \in \mathbb{N}^*,\ A_i \in \mathcal{F}) \Rightarrow \bigcup_{i\in\mathbb{N}^*} A_i \in \mathcal{F}$.

Le couple $(\Omega, \mathcal{F})$ est appelé espace mesurable.

On vérifie à partir de cette définition qu'une tribu est stable par unions finies (prendre tous les A_i vides à partir d'un certain rang), intersections finies et par intersections dénombrables (combiner b) et c)).

2. Il y a bien d'autres choix possibles, on pourrait poser plus généralement $m(]a, b]) := F(b) - F(a)$ où F est une fonction croissante $\mathbb{R} \to \mathbb{R}$, continue à droite.

Définition 5.2 (mesure)**.** *Soit $\mathcal{F}$ une* tribu *sur Ω. On appelle* mesure positive *sur $(\Omega, \mathcal{F})$ une application*

$$m : \mathcal{F} \longrightarrow [0, +\infty],$$

vérifiant

a) $m(\varnothing) = 0$ *;*

b) m *est σ-additive : pour toute suite $(A_i)_{i\in\mathbb{N}^*}$ d'éléments de $\mathcal{F}$ deux à deux disjoints,*

$$m\Big(\bigcup_{i\in\mathbb{N}^*} A_i\Big) = \sum_{i=1}^{+\infty} m(A_i). \tag{5.1}$$

Dans la définition de la σ-additivité, ou de la stabilité par union dénombrable, on aurait évidemment pu tout aussi bien indexer la suite (A_i) par $\mathbb{N}$ au lieu de $\mathbb{N}^*$.

Remarque 5.3. La réunion des A_i est invariante par permutation sur les indices et si chaque A_i est à son tour union dénombrable d'ensembles $B_{i,j} \in \mathcal{F}$ ($j \in \mathbb{N}^*$) deux à deux disjoints, on a clairement

$$\bigcup_{i\in\mathbb{N}^*} A_i = \bigcup_{i\in\mathbb{N}^*} \bigcup_{j\in\mathbb{N}^*} B_{i,j} = \bigcup_{(i,j)\in\mathbb{N}^*\times\mathbb{N}^*} B_{i,j}.$$

Sans les propriétés de convergence commutative et de sommation par paquets dans $\overline{\mathbb{R}}_+$ vues au chapitre 2, la définition de la σ-additivité serait incohérente.

Voyons maintenant des exemples de tribus qui nous seront utiles. Les trois plus simples sont les suivants.

- La tribu triviale sur Ω est $\mathcal{F} = \{\Omega, \varnothing\}$.
- $\mathcal{P}(\Omega)$ est une tribu.
- Si A est une partie de Ω, alors $\mathcal{F} := \{\Omega, \varnothing, A, A^c\}$ est une tribu. C'est la *plus petite* tribu possédant A comme élément, au sens où toute tribu $\mathcal{G}$ telle que $A \in \mathcal{G}$ contient $\mathcal{F}$. On dit que $\mathcal{F}$ est la tribu *engendrée* par A.

Cette notion de tribu engendrée se généralise en remarquant que si $(\mathcal{G}_i)_{i\in I}$ est une famille quelconque de tribus sur Ω, $\mathcal{G} := \bigcap_{i\in I} \mathcal{G}_i$ est une tribu sur Ω (vérification immédiate à partir de la définition 5.1).

Définition 5.4 (tribu engendrée)**.** *Soit $\mathcal{C}$ une famille de parties d'un ensemble Ω. On appelle* tribu engendrée *par $\mathcal{C}$, et on note $\sigma(\mathcal{C})$, la plus petite tribu contenant $\mathcal{C}$. C'est l'intersection de toutes les tribus sur Ω contenant $\mathcal{C}$.*

Définition 5.5 (tribu borélienne)**.** *On appelle* tribu borélienne *sur $\mathbb{R}^d$ la tribu engendrée par la famille $\mathcal{O}$ des ensembles* ouverts [3] *de $\mathbb{R}^d$. On la notera $\mathrm{Bor}(\mathbb{R}^d)$. Ainsi $\mathrm{Bor}(\mathbb{R}^d) = \sigma(\mathcal{O})$. Les sous-ensembles de $\mathbb{R}^d$ qui sont éléments de sa tribu borélienne sont appelés boréliens de $\mathbb{R}^d$ ou boréliens tout court quand il n'y a pas d'ambiguïté.*

3. Un ensemble ouvert de $\mathbb{R}^d$ est une réunion (quelconque) de *pavés* ouverts $\prod_{k=1}^d]a_k, b_k[$. Un *fermé* est le complémentaire d'un ouvert.

Remarque 5.6. On peut démontrer que $\mathrm{Bor}(\mathbb{R})$ est aussi engendrée par les fermés de $\mathbb{R}$, ou par les intervalles ouverts, ou les intervalles fermés, ou les semi-ouverts, ou les intervalles (ouverts ou fermés) à extrémités rationnelles, ou les intervalles $]-\infty, a]$, ou les intervalles $[a, +\infty[$. De même, $\mathrm{Bor}(\mathbb{R}^d)$ est engendrée par les pavés ouverts ou par les pavés de la forme $\prod_{k=1}^{d}]a_k, b_k]$.

Voyons maintenant des exemples importants de mesures.

Exemple 5.7 (masse de Dirac)**.** Soit x_0 un élément fixé de Ω. On appelle *masse de Dirac* au point x_0 ou *mesure de Dirac* au point x_0, la mesure δ_{x_0} sur $(\Omega, \mathcal{P}(\Omega))$ définie par

$$\forall A \in \mathcal{P}(\Omega), \quad \delta_{x_0}(A) := \begin{cases} 1 & \text{si } x_0 \in A, \\ 0 & \text{si } x_0 \notin A. \end{cases}$$

Par restriction, δ_{x_0} est aussi une mesure sur $(\Omega, \mathcal{F})$ pour toute tribu $\mathcal{F}$ sur Ω.

Vérification. Il est clair que $\delta_{x_0}(\varnothing) = 0$. Pour montrer la σ-additivité, soit $(A_n)_{n\in\mathbb{N}}$ une suite d'éléments de $\mathcal{P}(\Omega)$, deux à deux disjoints. Nous distinguons deux cas.

a) $x_0 \notin \bigcup_{n\in\mathbb{N}} A_n$. Alors $\delta_{x_0}\left(\bigcup_{n\in\mathbb{N}} A_n\right) = 0$. D'autre part, x_0 ne peut appartenir à aucun des A_n, donc pour tout $n \in \mathbb{N}$, $\delta_{x_0}(A_n) = 0$ et $\sum_{n\in\mathbb{N}} \delta_{x_0}(A_n) = 0$.

b) $x_0 \in \bigcup_{n\in\mathbb{N}} A_n$. Alors $\delta_{x_0}\left(\bigcup_{n\in\mathbb{N}} A_n\right) = 1$. D'autre part, comme les A_n sont deux à deux disjoints, x_0 doit appartenir à *un seul* des A_n, disons A_{n_0}. Ainsi, $\delta_{x_0}(A_{n_0}) = 1$ et pour tout $n \neq n_0$, $\delta_{x_0}(A_n) = 0$, d'où $\sum_{n\in\mathbb{N}} \delta_{x_0}(A_n) = 1$.

Comme $\delta_{x_0}\left(\bigcup_{n\in\mathbb{N}} A_n\right) = \sum_{n\in\mathbb{N}} \delta_{x_0}(A_n)$ dans les deux cas, δ_{x_0} est bien σ-additive. C'est une mesure sur $(\Omega, \mathcal{P}(\Omega))$. □

Exemple 5.8 (série de mesures finies)**.** Si les $(\mu_k)_{k\in\mathbb{N}}$ sont des mesures finies[4] sur $(\Omega, \mathcal{F})$ et si $(a_k)_{k\in\mathbb{N}}$ est une suite de réels positifs, la fonction d'ensembles $\mu = \sum_{k\in\mathbb{N}} a_k\mu_k$ définie sur $\mathcal{F}$ par

$$\mu : \mathcal{F} \to \overline{\mathbb{R}}_+, \quad A \mapsto \mu(A) := \sum_{k\in\mathbb{N}} a_k\mu_k(A)$$

est une mesure sur $(\Omega, \mathcal{F})$. Le résultat se généralise à $\mu = \sum_{i\in I} a_i\mu_i$, avec I au plus dénombrable.

Vérification. Pour tout k, $a_k\mu_k(\varnothing) = 0$ donc $\mu(\varnothing) = \sum_{k\in\mathbb{N}} a_k\mu_k(\varnothing) = 0$. Soit $(A_n)_{n\in\mathbb{N}}$ une suite d'éléments de $\mathcal{F}$, deux à deux disjoints. En utilisant la σ-additivité de chaque μ_k et l'interversion des sommations pour les séries doubles à termes positifs,

4. La mesure μ_k est *finie* si $\mu_k(A) < +\infty$ pour tout $A \in \mathcal{F}$. Elle est donc aussi *bornée* puisque $\mu_k(A) \leqslant \mu_k(\Omega) < +\infty$. L'hypothèse « μ_k finie » nous évite ici la gestion du conflit entre $a_k = 0$ et $\mu_k(A) = +\infty$.

on peut écrire :

$$\begin{aligned}\mu\Big(\bigcup_{n\in\mathbb{N}} A_n\Big) = \sum_{k\in\mathbb{N}} a_k\mu_k\Big(\bigcup_{n\in\mathbb{N}} A_n\Big) = \sum_{k\in\mathbb{N}} a_k\left\{\sum_{n\in\mathbb{N}}\mu_k(A_n)\right\} &= \sum_{k\in\mathbb{N}}\left\{\sum_{n\in\mathbb{N}} a_k\mu_k(A_n)\right\}\\ &= \sum_{n\in\mathbb{N}}\left\{\sum_{k\in\mathbb{N}} a_k\mu_k(A_n)\right\}\\ &= \sum_{n\in\mathbb{N}}\mu(A_n),\end{aligned}$$

ce qui établit la σ-additivité de μ. □

Exemple 5.9 (mesure ponctuelle)**.** Soient I un ensemble au plus dénombrable, $(x_i)_{i\in I}$ une famille dans Ω et $(a_i)_{i\in I}$ une famille de réels positifs. Alors $\mu := \sum_{i\in I} a_i\delta_{x_i}$ est une mesure sur $(\Omega, \mathcal{P}(\Omega))$, donc aussi par restriction sur tout espace mesurable $(\Omega, \mathcal{F})$. C'est un cas particulier de l'exemple précédent. Les mesures de ce type sont appelées *mesures ponctuelles*.

Remarque 5.10. Si $\Omega = \{\omega_i;\ i \in I\}$ est au plus dénombrable, toute mesure sur $(\Omega, \mathcal{P}(\Omega))$ qui est *finie sur les singletons* est une mesure ponctuelle.

Vérification. La tribu $\mathcal{P}(\Omega)$ possède les singletons [5]. Soit μ une mesure sur $(\Omega, \mathcal{P}(\Omega))$. On peut définir :

$$\forall i \in I, \quad a_i := \mu(\{\omega_i\}),$$

en notant que puisque μ est finie sur les singletons, $0 \leqslant a_i < +\infty$. Considérons alors la mesure ponctuelle

$$\nu := \sum_{i\in I} a_i\delta_{\omega_i}.$$

Tout $A \in \mathcal{P}(\Omega)$ est fini ou dénombrable et s'écrit comme union disjointe finie ou dénombrable de singletons :

$$A = \bigcup_{\omega_i\in A}\{\omega_i\},$$

d'où par σ-additivité (ou par additivité quand A est fini, cf. prop. 5.16 ci-dessous) :

$$\mu(A) = \sum_{\omega_i\in A}\mu(\{\omega_i\}) = \sum_{\omega_i\in A} a_i = \sum_{i\in I} a_i\delta_{\omega_i}(A) = \nu(A).$$

Ainsi $\mu(A) = \nu(A)$ pour tout $A \in \mathcal{P}(\Omega)$, donc $\mu = \nu$ et μ est une mesure ponctuelle. □

Exemple 5.11 (mesure de comptage)**.** Si $\Omega = \{\omega_i;\ i \in I\}$ est au plus dénombrable, l'application $\mu : \mathcal{P}(\Omega) \to \overline{\mathbb{R}}_+$ définie par

$$\forall A \in \mathcal{P}(\Omega), \quad \mu(A) = \begin{cases}\operatorname{card} A & \text{si } A \text{ est fini,}\\ +\infty & \text{sinon,}\end{cases}$$

5. Quand Ω est fini ou dénombrable, c'est même la seule tribu ayant cette propriété.

est une mesure, appelée *mesure de comptage*. Pour le voir, il suffit de remarquer que pour tout A, $\mu(A) = \nu(A)$, où ν est définie comme la mesure ponctuelle $\nu = \sum_{i\in I} \delta_{\omega_i}$.

Exemple 5.12. Soit μ une mesure sur $(\Omega, \mathcal{F})$ et $B \in \mathcal{F}$. La fonction d'ensembles $\nu = \mu(\,.\,\cap B)$ définie sur $\mathcal{F}$ par

$$\forall A \in \mathcal{F}, \quad \nu(A) := \mu(A \cap B)$$

est une mesure sur $(\Omega, \mathcal{F})$. Si de plus $0 < \mu(B) < +\infty$, la fonction d'ensembles μ_B définie sur $\mathcal{F}$ par

$$\forall A \in \mathcal{F}, \quad \mu_B(A) := \frac{\mu(A \cap B)}{\mu(B)}$$

est une mesure sur $(\Omega, \mathcal{F})$, vérifiant $\mu_B(\Omega) = 1$, autrement dit une *probabilité*[6]. Quand $\mu = P$ est déjà une probabilité sur $(\Omega, \mathcal{F})$, la mesure P_B est appelée *probabilité conditionnelle* ; notation : $P_B(A) =: P(A \mid B)$.

Vérification. On se contente de vérifier que ν est σ-additive, tout le reste étant évident. Soit $(A_i)_{i\in\mathbb{N}}$ une suite d'éléments de $\mathcal{F}$, deux à deux disjoints. Pour $i \neq j$, $A_i \cap A_j = \varnothing$ et comme $(A_i \cap B) \cap (A_j \cap B) \subset A_i \cap A_j$, les $A_i \cap B$ sont aussi deux à deux disjoints. Comme

$$\Big(\bigcup_{i\in\mathbb{N}} A_i\Big) \cap B = \bigcup_{i\in\mathbb{N}} (A_i \cap B),$$

la σ-additivité de ν découle alors clairement de celle de μ. □

Exemple 5.13 (mesure de Lebesgue λ_d)**.** Nous admettons qu'il existe une unique mesure μ sur $(\mathbb{R}^d, \mathrm{Bor}(\mathbb{R}^d))$ telle que pour tout pavé $\prod_{i=1}^d]a_i, b_i]$ non vide[7],

$$\mu\left(\prod_{i=1}^d]a_i, b_i]\right) = \prod_{i=1}^d (b_i - a_i).$$

On l'appelle *mesure de Lebesgue* sur $\mathbb{R}^d$ et on la note λ_d. Pour $d = 1$, on étend ainsi la notion de longueur des intervalles à tous les ensembles membres de $\mathrm{Bor}(\mathbb{R})$, pour $d = 2$ on étend de même la notion d'aire des rectangles à tous les ensembles boréliens du plan $\mathbb{R}^2$, pour $d = 3$, on étend la notion de volume. Pour $d > 3$, on continuera à parler de volume ou d'hypervolume. Nous admettrons que les formules classiques de calcul d'aire ou de volume se réécrivent à l'aide de λ_d. Par exemple si D est un disque de rayon r de $\mathbb{R}^2$, $\lambda_2(D) = \pi r^2$. Plus généralement, voici les principales propriétés de la mesure de Lebesgue.

Proposition 5.14. *La mesure de Lebesgue λ_d sur $(\mathbb{R}^d, \mathrm{Bor}(\mathbb{R}^d))$ a les propriétés suivantes.*

i) λ_d est invariante par translations : si $h : x \mapsto x + v$ est une translation de $\mathbb{R}^d$, alors pour tout $B \in \mathrm{Bor}(\mathbb{R}^d)$, $h(B) \in \mathrm{Bor}(\mathbb{R}^d)$ et $\lambda_d(B) = \lambda_d\big(h(B)\big)$.

6. En anticipant un peu sur la suite du chapitre.
7. Ce qui suppose implicitement que $a_i < b_i$ pour chaque $i \in [\![1, d]\!]$.

ii) λ_d *est invariante par toute isométrie euclidienne de* $\mathbb{R}^d$ *: symétrie, rotation, etc.*

iii) Si h *est l'homothétie* $x \mapsto cx$ *dans* $\mathbb{R}^d$*, pour tout borélien* B*,* $\lambda_d\big(h(B)\big) = |c|^d \lambda_d(B)$*.*

iv) λ_d *ne charge pas les points :* $\lambda_d(\{x\}) = 0$ *pour tout* $x \in \mathbb{R}^d$*. Si* $A \subset \mathbb{R}^d$ *est fini ou dénombrable,* $\lambda_d(A) = 0$*.*

v) $\lambda_d\left(\prod_{i=1}^{d} [a_i, b_i]\right) = \lambda_d\left(\prod_{i=1}^{d}]a_i, b_i[\right)$ *et cette égalité implique, bien sûr, l'égalité des mesures des* 4^d *pavés obtenus en jouant sur l'ouverture ou la fermeture des extrémités* a_i*,* b_i *des intervalles.*

vi) Si E *est un sous-espace affine de* $\mathbb{R}^d$ *et* $E \neq \mathbb{R}^d$*,* $\lambda_d(E) = 0$*.*

Un bon exercice consiste à établir la formule $\lambda_2(D) = \pi r^2$ en utilisant le calcul de l'aire de l'hypographe de la fonction $f : [-1, 1] \to \mathbb{R}$, $x \mapsto \sqrt{1 - x^2}$ par une intégrale de Riemann, cf. proposition 3.6 p. 74 et certaines des propriétés de λ_2 énoncées ci-dessus.

5.2 Modéliser l'aléatoire

5.2.1 Notion d'expérience aléatoire

La théorie des probabilités fournit des modèles mathématiques permettant l'étude d'expériences dont le résultat ne peut être prévu avec une totale certitude. Le tableau 5.1 en donne quelques exemples.

Expérience	**Résultat observable**
Lancer d'un dé	Un entier $k \in \{1, \ldots, 6\}$
Prélèvement de n objets en sortie d'une chaîne de production	Nombre d'objets défectueux dans l'échantillon
Questionnaire à 100 questions binaires	Suite ω de 100 réponses $\omega \in \{\text{oui}, \text{non}\}^{100}$
Lancer d'une pièce jusqu'à la première obtention de pile	Un entier $k \in \mathbb{N}$: le temps d'attente du premier succès
Mise en service d'une ampoule	Durée de vie $T \in \mathbb{R}_+$
Lancer d'une fléchette sur une cible	Point d'impact
Mouvement d'un grain de pollen dans un liquide	Une fonction continue : la trajectoire
Mélange de deux gaz	Répartition spatiale de deux types de molécules

TABLE 5.1 – Quelques expériences aléatoires typiques

Bien que le résultat précis de chacune de ces expériences soit imprévisible, l'observation et l'intuition nous amènent à penser que ces phénomènes obéissent à certaines lois. Par exemple si on jette 6 000 fois un dé, on s'attend à ce que le nombre d'apparitions de la face « 3 » soit *voisin* de 1 000. Si on met en service 100 ampoules, leurs durées de vie observées seront *concentrées* autour d'une certaine valeur moyenne.

La théorie des probabilités permet de donner un sens précis à ces considérations un peu vagues. La *statistique* permet de confronter les modèles probabilistes avec la réalité observée afin de les valider ou de les invalider. Par exemple si quelqu'un a 60 bonnes réponses sur 100 à un questionnaire, est-il légitime de considérer qu'il a « mieux fait » que le hasard ? Sur les n objets prélevés en sortie de chaîne, k sont défectueux. Peut-on en déduire quelque chose sur la qualité de la production globale ?

5.2.2 Évènements

La théorie moderne des probabilités utilise le langage des ensembles pour modéliser une expérience aléatoire. Nous noterons Ω un ensemble dont les éléments représentent tous les résultats possibles ou *évènements élémentaires* d'une expérience aléatoire donnée. Les *évènements* (ou évènements composés) seront représentés par des parties (sous-ensembles) de Ω.

Il n'est pas toujours facile de trouver un ensemble Ω permettant de modéliser l'expérience aléatoire. Voici une règle pratique pour y arriver : les évènements élémentaires sont ceux qui contiennent *l'information maximale* qu'il est possible d'obtenir de l'expérience. Par exemple si on jette un dé, l'évènement A : « obtention d'un chiffre pair » n'est pas élémentaire. Il est composé des trois évènements élémentaires 2, 4, 6 : $A = \{2, 4, 6\}$. Ici $\Omega = \{1, 2, 3, 4, 5, 6\}$. De même si on lance trois fois une pièce de monnaie, les évènements élémentaires sont des triplets comme (p,f,p) indiquant le résultat précis de chacun des trois lancers. Ici $\Omega = \{\mathtt{f}, \mathtt{p}\}^3$. L'évènement B « obtention de pile au deuxième des trois lancers » est composé : $B = \{(\mathtt{f}, \mathtt{p}, \mathtt{f}); (\mathtt{f}, \mathtt{p}, \mathtt{p}); (\mathtt{p}, \mathtt{p}, \mathtt{f}); (\mathtt{p}, \mathtt{p}, \mathtt{p})\}$.

Avec ce mode de représentation, les opérations logiques sur les évènements : « et », « ou », « négation » se traduisent par des opérations ensemblistes : intersection, réunion, passage au complémentaire. Le tableau 5.2 page 163 présente la correspondance entre les deux langages.

Les opérations logiques sur les évènements peuvent bien sûr faire intervenir plus de deux évènements. Ainsi, si $A_1, \ldots, A_n$ sont des évènements,

$$\bigcup_{i=1}^{n} A_i = A_1 \cup A_2 \cup \cdots \cup A_n$$

est l'ensemble des ω qui sont dans l'un au moins des A_i. C'est donc l'évènement « réalisation de l'un au moins des A_i $(1 \leqslant i \leqslant n)$ ». De même :

$$\bigcap_{i=1}^{n} A_i = A_1 \cap A_2 \cdots \cap A_n$$

est l'ensemble des ω qui sont dans tous les A_i. C'est donc l'évènement « réalisation de chacun des A_i $(1 \leqslant i \leqslant n)$ ». Ceci s'étend aux réunions et intersections d'une suite

infinie d'évènements :

$$\begin{aligned} \bigcup_{i\in\mathbb{N}^*} A_i &= \{\text{réalisation de l'un au moins des } A_i,\ i\in\mathbb{N}^*\},\\ \bigcap_{i\in\mathbb{N}^*} A_i &= \{\text{réalisation de tous les } A_i,\ i\in\mathbb{N}^*\}. \end{aligned}$$

Ces opérations logiques sur des suites d'évènements sont très utiles pour analyser des évènements complexes à l'aide d'évènements plus simples et, comme nous le verrons plus tard, calculer ainsi des probabilités.

Notations	Vocabulaire ensembliste	Vocabulaire probabiliste
$\varnothing$	ensemble vide	évènement impossible
Ω	ensemble plein	évènement certain
ω	élément de Ω	évènement élémentaire
A	sous-ensemble de Ω	évènement
$\omega \in A$	ω appartient à A	Le résultat ω est une des réalisations possibles de A
$A \subset B$	A inclus dans B	A implique B
$A \cup B$	réunion de A et B	A ou B
$A \cap B$	intersection de A et B	A et B
A^c	complémentaire de A dans Ω	évènement contraire de A
$A \cap B = \varnothing$	A et B sont disjoints	A et B sont incompatibles

TABLE 5.2 – Langage ensembliste - langage probabiliste

5.2.3 Une question de dés

Pour finir cette introduction informelle, nous allons discuter un problème d'énoncé très simple pour voir comment les notions de tribu et de mesure s'imposent naturellement dès que l'on veut évaluer une probabilité dans une expérience où apparaît l'infini. Voici la question. On effectue des lancers répétés d'une paire de dés et on observe pour chaque lancer, la *somme* des points indiqués par les deux dés. On se propose d'attribuer une probabilité à l'évènement E défini ainsi : *dans la suite des résultats observés, la première obtention d'un* 9 *a lieu avant la première obtention d'un* 7. On suppose ici déjà connue la définition d'une probabilité sur un espace Ω fini : à savoir une application *additive*[8] $P : \mathcal{P}(\Omega) \to \mathbb{R}_+$ telle que $P(\Omega) = 1$.

Commençons par modéliser un lancer. Disons que l'on a un dé bleu et un dé rouge. L'information maximale que l'on puisse envisager ici est de savoir le résultat du dé

8. Si Ω est fini, la σ-additivité se réduit à l'additivité car dans toute suite (A_n) de parties de Ω deux à deux disjointes, seul un nombre fini d'entre elles ne sont pas vides.

bleu et celui du rouge. On prendra donc comme espace Ω_1, le carré cartésien $[\![1,6]\!]^2$ en convenant de représenter par le couple $(i,j) \in \Omega_1$ le résultat du dé bleu (première composante) et celui du rouge (deuxième composante). Notons pour $i \in \mathbb{N}^*$,

$$\begin{aligned} F_i &:= \{\text{obtention de la somme 9 au } i^{\text{e}} \text{ lancer}\},\\ G_i &:= \{\text{obtention de la somme 7 au } i^{\text{e}} \text{ lancer}\},\\ H_i &:= \{\text{obtention d'une autre somme que 7 ou 9 au } i^{\text{e}} \text{ lancer}\}. \end{aligned}$$

Pour l'instant, nous n'envisageons qu'un lancer et nous attribuerons par symétrie la même probabilité à tous les évènements élémentaires en prenant $P_1(\{(i,j)\}) = 1/36$ pour tout évènement élémentaire (i,j). En remarquant que F_1 est constitué de 4 évènements élémentaires : $(3,6)$, $(4,5)$, $(5,4)$ et $(6,3)$, on en déduit que $P_1(F_1) = 4/36 = 1/9$. De même, G_1 étant constitué de 6 évènements élémentaires, $P_1(G_1) = 6/36 = 1/6$, cf. figure 5.1. On en déduit que $P_1(H_1) = 1 - 10/36 = 13/18$.

	1	2	3	4	5	6
1						7
2					7	
3				7		9
4			7		9	
5		7		9		
6	7		9			

FIGURE 5.1 – Évènements $F1$ et G_1

Pour modéliser les n premiers lancers, on voit que l'information maximale que l'on puisse envisager est la connaissance de la suite finie des n couples d'entiers représentant les résultats des n lancers. Ceci conduit à prendre comme espace $\Omega_n = \Omega_1 \times \cdots \times \Omega_1 = \Omega_1^n$. Cet Ω_n est un ensemble fini de cardinal 36^n. Si nous convenons là encore d'attribuer à tous les évènements élémentaires la même probabilité (donc maintenant 36^{-n}), on obtient la probabilité P_n définie sur $(\Omega_n, \mathcal{P}(\Omega_n))$ par

$$\forall A \in \mathcal{P}(\Omega_n), \quad P_n(A) = \frac{\operatorname{card} A}{\operatorname{card} \Omega_n}.$$

Regardons le cas particulier où A est de la forme $A_1 \times A_2 \times \cdots \times A_n$, chaque A_i étant une partie de $[\![1,6]\!]^2$. Cela signifie que la réalisation de A_i ne dépend que du résultat du i^{e} lancer. La formule sur le cardinal d'un produit cartésien nous donne alors :

$$\begin{aligned} P_n(A_1 \times \cdots \times A_n) = \frac{\operatorname{card}(A_1 \times \cdots \times A_n)}{\operatorname{card} \Omega_n} &= \frac{(\operatorname{card} A_1) \times \cdots \times (\operatorname{card} A_n)}{(\operatorname{card} \Omega_1)^n}\\ &= \frac{\operatorname{card} A_1}{36} \times \cdots \times \frac{\operatorname{card} A_n}{36}\\ &= P_1(A_1) \times \cdots \times P_1(A_n). \end{aligned} \tag{5.2}$$

Appliquons ceci à l'évènement E_n défini par

$$E_n := \{\text{les } n-1 \text{ premiers lancers ne donnent ni 7 ni 9 et le } n^{\text{e}} \text{ donne 9}\}. \tag{5.3}$$

On peut représenter[9] E_n dans Ω_n en écrivant $E_n = H_1 \times H_2 \times \cdots \times H_{n-1} \times F_n$, donc

$$P_n(E_n) = \frac{1}{9}\Big(\frac{13}{18}\Big)^{n-1},$$

formule valable pour tout $n \geqslant 1$, le cas $n = 1$ se réduisant à $E_1 = F_1$ et $P_1(F_1) = 1/9$.

Pour $1 \leqslant k < n$, E_k est représenté dans Ω_n par $H_1 \times H_2 \times \cdots \times H_{k-1} \times F_k \times \Omega_1^{n-k}$. La formule (5.2) nous donne alors :

$$\forall n \geqslant k, \quad P_n(E_k) = P_1(H_1) \times \cdots \times P_1(H_{k-1}) \times P_1(F_k) \times 1^{n-m} = P_k(E_k), \tag{5.4}$$

où dans l'écriture « $P_k(E_k)$ », on interprète E_k comme une partie de Ω_k, explicitement $H_1 \times H_2 \times \cdots \times H_{k-1} \times F_k$. Ainsi (5.4) exprime une sorte de *compatibilité ascendante* des modèles $(\Omega_n, \mathcal{P}(\Omega_n), P_n)$. Si on note E'_n l'évènement *au cours des n premiers lancers, la première obtention de la somme* 9 *a lieu avant la première obtention de la somme* 7, clairement $E'_n = \bigcup_{k=1}^n E_k$ et grâce à cette compatibilité, pour tout $j \geqslant n$,

$$P_j(E'_n) = P_j\Big(\bigcup_{k=1}^n E_k\Big) = \sum_{k=1}^n P_j(E_k) = \sum_{k=1}^n P_k(E_k) = \sum_{k=1}^n \frac{1}{9}\Big(\frac{13}{18}\Big)^{k-1}, \tag{5.5}$$

en notant que les E_k sont deux à deux disjoints et en utilisant l'additivité de P_j.

Pour attribuer une probabilité à E, on ne peut malheureusement pas se contenter de (5.5), même avec n « grand ». En effet on ne peut pas exclure que la question de la priorité entre le 9 et le 7 ne soit tranchée qu'au delà du n^{e} lancer, voire jamais. Ceci conduit à prendre pour Ω l'ensemble $\Omega_1^{\mathbb{N}^*}$ des suites infinies de couples $(i, j) \in [\![1, 6]\!]^2$. On peut représenter E_k dans Ω par $E_k = H_1 \times H_2 \times \cdots \times H_{k-1} \times F_k \times \Omega_1^{]\!]k, +\infty[\![}$. Au vu de la formule de compatibilité (5.4), il est naturel d'attribuer une probabilité à E_k considéré comme évènement de Ω en *définissant*

$$P(E_k) := P_k(E_k) = \frac{1}{9}\Big(\frac{13}{18}\Big)^{k-1}.$$

Finalement, pour pouvoir attribuer une probabilité à E, on remarque que E est l'union des E_k pour $k \in \mathbb{N}^*$, ces ensembles étant deux à deux disjoints. On a donc bien envie d'écrire

$$P(E) = P\Big(\bigcup_{k\in\mathbb{N}^*} E_k\Big) = \sum_{k\in\mathbb{N}^*} P(E_k) = \sum_{k=1}^{+\infty} \frac{1}{9}\Big(\frac{13}{18}\Big)^{k-1} = \frac{1}{9}\,\frac{1}{1-\frac{13}{18}} = \frac{2}{5}.$$

9. L'évènement E_n est *défini* par la *phrase* entre les accolades dans (5.3). Son *écriture ensembliste* dépend de l'espace Ω considéré. D'habitude, on ne fait pas cette distinction parce qu'on travaille avec un seul Ω, mais ici elle s'impose.

La propriété qui nous manque pour justifier cette écriture est précisément la σ-additivité.

Une autre façon d'obtenir le même résultat est de passer à la limite quand n tend vers l'infini dans (5.5), après avoir réécrit $P_j(E'_n)$ sous la forme $P(E'_n)$. Remarquons que la suite d'ensembles E'_n est croissante pour l'inclusion ($E'_n \subset E'_{n+1}$) et que $\bigcup_{n\geqslant 1} E'_n = E$. La propriété qui nous permettrait de faire cela s'appelle continuité séquentielle croissante de P (voir prop. 5.16-6.a).

On voit ainsi qu'il est souhaitable de définir la probabilité P sur Ω comme une *mesure*. Ceci pose la question de la tribu $\mathcal{F}$. En fait tout ce que nous savons faire ici, c'est définir la probabilité d'évènements comme les E_k, de la forme $A \times \Omega_1^{\mathbb{N}_k}$, avec $A \subset \Omega_k$. Ces évènements sont ceux dont la réalisation ne dépend que du résultat des k premières épreuves, pour un certain k. On peut dire aussi qu'il s'agit des évènements dont la réalisation ne dépend que d'un nombre fini d'épreuves. Notons $\mathcal{C}$ la famille de ces évènements. Un évènement comme E n'appartient pas à $\mathcal{C}$, mais à la tribu engendrée par $\mathcal{C}$, $\mathcal{F} = \sigma(\mathcal{C})$. Cette tribu est plus petite [10] que $\mathcal{P}(\Omega)$.

Enfin, soit H l'évènement *aucun lancer ne produit la somme* 7 *ou la somme* 9. Avec

$$K_n := H_1 \times \cdots \times H_n \times \Omega_1^{\mathbb{N}_n},$$

il est clair que

$$H = H_1^{\mathbb{N}^*} = \bigcap_{n\geqslant 1} K_n.$$

L'ensemble H apparaît ainsi comme intersection *dénombrable* d'ensembles $K_n \in \mathcal{C}$, ce qui nous assure de l'appartenance de H à la tribu $\mathcal{F}$. On a de plus $H \subset K_n$ pour tout $n \geqslant 1$, donc par croissance de P pour l'inclusion (c'est une conséquence immédiate de l'additivité, cf. prop. 5.16-4.),

$$\forall n \geqslant 1, \quad P(H) \leqslant \left(\frac{13}{18}\right)^n.$$

Cette inégalité étant vérifiée pour tout n, on peut faire tendre n vers l'infini pour obtenir $P(H) = 0$. On a ainsi un exemple d'évènement non vide et de probabilité nulle (H est l'ensemble des suites infinies de couples éléments de H_1, il est infini non dénombrable).

5.3 La probabilité comme mesure

La *probabilité* P, telle que nous allons la définir ci-dessous, est une fonction qui à un évènement, associe un nombre compris entre 0 et 1 et censé mesurer les chances de réalisation de cet évènement. Pour des raisons sortant du cadre de cet ouvrage, il n'est pas toujours possible d'attribuer ainsi de manière cohérente une probabilité à chaque partie de Ω. En d'autres termes, P ne peut pas être considérée comme une *application* de l'ensemble $\mathcal{P}(\Omega)$ de toutes les parties de Ω dans $[0,1]$ mais comme une

10. Nous l'admettrons, mais si vous n'êtes pas convaincu, essayez de montrer que $\mathcal{F} = \mathcal{P}(\Omega)$ ...

fonction ayant pour domaine de définition une tribu $\mathcal{F}$ généralement plus petite que $\mathcal{P}(\Omega)$. La tribu $\mathcal{F}$ est aussi appelée famille des évènements observables [11].

Définition 5.15. *Soit Ω un ensemble et $\mathcal{F}$ une tribu sur Ω. On appelle probabilité sur $(\Omega, \mathcal{F})$ toute application P de $\mathcal{F}$ dans $[0,1]$ vérifiant :*

(i) $P(\Omega) = 1$.

(ii) Pour toute suite $(A_j)_{j\geqslant 1}$ d'évènements de $\mathcal{F}$ deux à deux disjoints (incompatibles) :

$$P\Big(\bigcup_{j\in\mathbb{N}^*} A_j\Big) = \sum_{j=1}^{+\infty} P(A_j).$$

Le triplet $(\Omega, \mathcal{F}, P)$ s'appelle espace probabilisé.

Définir une probabilité sur $(\Omega, \mathcal{F})$ c'est en quelque sorte attribuer une « masse » à chaque évènement observable, avec par convention une masse totale égale à 1 pour l'évènement certain Ω. Une formulation équivalente à la définition 5.15 est : *P est une mesure sur $(\Omega, \mathcal{F})$ telle que $P(\Omega) = 1$.*

Proposition 5.16 (propriétés générales d'une probabilité)**.**
Toute probabilité P sur $(\Omega, \mathcal{F})$ vérifie les propriétés suivantes :

1. $P(\varnothing) = 0$.
2. *Additivité.*
   a) Si $A \cap B = \varnothing$, $P(A \cup B) = P(A) + P(B)$.
   b) Si les A_i $(1 \leqslant i \leqslant n)$ sont deux à deux disjoints : $P\big(\bigcup_{i=1}^n A_i\big) = \sum_{i=1}^n P(A_i)$.
3. $\forall A \in \mathcal{F},\ P(A^c) = 1 - P(A)$.
4. $\forall A \in \mathcal{F},\ \forall B \in \mathcal{F},\ A \subset B \ \Rightarrow\ P(A) \leqslant P(B)$.
5. $\forall A \in \mathcal{F},\ \forall B \in \mathcal{F},\ P(A \cup B) = P(A) + P(B) - P(A \cap B)$.
6. *Continuité monotone séquentielle.*
   a) Si $(B_n)_{n\geqslant 0}$ est une suite croissante d'évènements de $\mathcal{F}$ convergente [12] vers $B \in \mathcal{F}$, alors $P(B) = \lim_{n\to+\infty} P(B_n)$, en notation abrégée :

$$B_n \uparrow B \Rightarrow P(B_n) \uparrow P(B) \quad (n \to +\infty).$$

   b) Si $(C_n)_{n\geqslant 0}$ est une suite décroissante d'évènements de $\mathcal{F}$ convergente [13] vers $C \in \mathcal{F}$, alors $P(C) = \lim_{n\to+\infty} P(C_n)$ en notation abrégée :

$$C_n \downarrow C \Rightarrow P(C_n) \downarrow P(C) \quad (n \to +\infty).$$

11. La définition générale d'une tribu $\mathcal{F}$ ne suppose pas que tous les singletons $\{\omega\}$ soient des éléments de $\mathcal{F}$. Donc un « évènement élémentaire » n'est pas toujours un évènement observable. Néanmoins dans la plupart des exemples que nous étudierons, la tribu possèdera les singletons.
12. Ce qui signifie : $\forall n \geqslant 0$, $B_n \subset B_{n+1}$ et $B = \bigcup_{n\geqslant 0} B_n$.
13. Ce qui signifie : $\forall n \geqslant 0$, $C_{n+1} \subset C_n$ et $C = \bigcap_{n\geqslant 0} C_n$.

7. *Sous-additivité et sous-σ-additivité.*
   a) $\forall A \in \mathcal{F}, \forall B \in \mathcal{F}, \quad P(A \cup B) \leqslant P(A) + P(B)$.
   b) $\forall A_1, \ldots, A_n \in \mathcal{F}, \quad P\big(\bigcup_{i=1}^{n} A_i\big) \leqslant \sum_{i=1}^{n} P(A_i)$.
   c) $\forall A_1, \ldots, A_n, \ldots \in \mathcal{F}$, $P\big(\bigcup_{i\in\mathbb{N}^*} A_i\big) \leqslant \sum_{i=1}^{+\infty} P(A_i)$.

Preuve. Soit P une fonction d'ensembles $\mathcal{F} \to [0,1]$ satisfaisant aux conditions (i) et (ii) de la définition 5.15, il s'agit de démontrer que P vérifie les propriétés 1 à 7.

Preuve de 1. Comme $P(A_j) \geqslant 0$ pour tout $A_j \in \mathcal{F}$, on a toujours

$$\sum_{j\in\mathbb{N}^*} P(A_j) \geqslant P(A_1) + P(A_2),$$

le premier membre pouvant être égal à $+\infty$. En choisissant $A_j = \varnothing$ pour tout $j \in \mathbb{N}^*$ et en utilisant la σ-additivité (ii), on en déduit :

$$P(\varnothing) = P\Big(\bigcup_{j\in\mathbb{N}^*} A_j\Big) = \sum_{j=1}^{+\infty} P(A_j) \geqslant P(\varnothing) + P(\varnothing).$$

Par conséquent, $P(\varnothing) \geqslant 2P(\varnothing)$ et comme $P(\varnothing) \geqslant 0$, ceci entraîne $P(\varnothing) = 0$. □

Preuve de 2. Soient $A_1, \ldots, A_n$, n évènements de $\mathcal{F}$ deux à deux disjoints. Pour $j > n$, posons $A_j = \varnothing$. On a ainsi une suite infinie $(A_j)_{j\geqslant 1}$ d'évènements deux à deux disjoints. En utilisant la σ-additivité, on obtient alors :

$$P\Big(\bigcup_{j=1}^{n} A_j\Big) = P\Big(\bigcup_{j\in\mathbb{N}^*} A_j\Big) = \sum_{j=1}^{n} P(A_j) + \sum_{j=n+1}^{+\infty} P(A_j).$$

D'après 1, la somme pour $j \geqslant n+1$ vaut 0, ceci prouve 2 b). Bien sûr, 2 a) n'est que le cas particulier $n = 2$. □

Preuve de 3. Prendre $B = A^c$ dans 2 a) et utiliser (i). □

Preuve de 4. Si $A \subset B$, alors $B = A \cup (B \cap A^c)$ et cette réunion est disjointe. D'après 2 a) on a $P(B) = P(A) + P(B \cap A^c)$ et comme $P(B \cap A^c) \geqslant 0$, on en déduit $P(B) \geqslant P(A)$. □

Preuve de 5. On a les décompositions suivantes en unions disjointes :

$$\begin{aligned} A \cup B &= (A \cap B^c) \cup (A \cap B) \cup (A^c \cap B), \\ A &= (A \cap B^c) \cup (A \cap B), \\ B &= (A \cap B) \cup (A^c \cap B). \end{aligned}$$

En utilisant la décomposition de $A \cup B$ et l'additivité on obtient :

$$\begin{aligned} P(A \cup B) &= P(A \cap B^c) + P(A \cap B) + P(A^c \cap B) \\ &= \big[P(A \cap B^c) + P(A \cap B)\big] + \big[P(A \cap B) + P(A^c \cap B)\big] - P(A \cap B). \end{aligned}$$

Par additivité, la somme du premier crochet vaut $P(A)$, celle du deuxième vaut $P(B)$, d'où le résultat. □

Preuve de 6. Il suffit de prouver 6 a), la propriété 6 b) s'en déduit en appliquant 6 a) à la suite d'évènements $B_n = C_n^c$. Admettons, pour l'instant, que pour tout $n \geqslant 1$, B_n vérifie la décomposition suivante en union disjointe (cf. figure 5.2)

$$B_n = B_0 \cup \left(\bigcup_{i=1}^{n} (B_i \backslash B_{i-1}) \right).$$

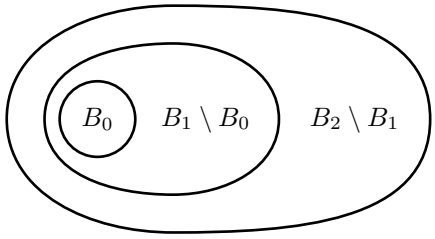


FIGURE 5.2 – Décomposition de $B_0 \cup B_1 \cup B_2$ en union disjointe

En écrivant la réunion infinie des B_n à l'aide de cette décomposition et en « effaçant » toutes les répétitions des $B_i \backslash B_{i-1}$, on en déduit immédiatement que B vérifie la décomposition en union disjointe :

$$B = B_0 \cup \left(\bigcup_{i \in \mathbb{N}^*} (B_i \backslash B_{i-1}) \right).$$

Passant aux probabilités, ces deux décompositions nous donnent :

$$\begin{aligned} P(B_n) &= P(B_0) + \sum_{i=1}^{n} P(B_i \backslash B_{i-1}), \\ P(B) &= P(B_0) + \sum_{i=1}^{+\infty} P(B_i \backslash B_{i-1}). \end{aligned}$$

Comme cette série converge, sa somme est la limite de la suite de ses sommes partielles de rang n, ce qui s'écrit :

$$P(B) = \lim_{n \to +\infty} \left\{ P(B_0) + \sum_{i=1}^{n} P(B_i \backslash B_{i-1}) \right\} = \lim_{n \to +\infty} P(B_n).$$

Ainsi pour compléter la preuve, il ne reste plus qu'à justifier la décomposition de B_n. Posons :

$$D_n = B_0 \cup \left(\bigcup_{i=1}^{n} (B_i \backslash B_{i-1}) \right).$$

Pour montrer que $B_n = D_n$, il suffit de montrer que $D_n \subset B_n$ et $B_n \subset D_n$. La première inclusion est évidente car $B_i \backslash B_{i-1} \subset B_i \subset B_n$, pour $i \leqslant n$. Pour prouver l'inclusion inverse, on note ω un élément *quelconque* de B_n et on montre que ω appartient à D_n. Soit $i_0 = i_0(\omega)$ le plus petit des indices i tels que $\omega \in B_i$. Comme cet ensemble d'indices contient au moins n, on a $0 \leqslant i_0 \leqslant n$. Si $i_0 = 0$, $\omega \in B_0$ et comme $B_0 \subset D_n$, $\omega \in D_n$. Si $i_0 \geqslant 1$, par la définition même de i_0, on a $\omega \in B_{i_0}$ et $\omega \notin B_{i_0-1}$, donc $\omega \in B_{i_0} \backslash B_{i_0-1}$ et comme $i_0 \leqslant n$, $B_{i_0} \backslash B_{i_0-1} \subset D_n$ donc $\omega \in D_n$. Le raisonnement précédent étant valable pour tout ω de B_n, on en déduit $B_n \subset D_n$. □

Preuve de 7 a). D'après 5 :

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \leqslant P(A) + P(B),$$

car $P(A \cap B) \geqslant 0$. □

Preuve de 7 b). On remarque que pour tout $n \geqslant 1$, $\bigcup_{i=1}^n A_i = \bigcup_{i=1}^n B_i$, où les B_i sont des évènements *deux à deux disjoints* définis comme suit :

$$B_1 = A_1,\ B_2 = A_2 \cap B_1^c,\ B_3 = A_3 \cap (B_1 \cup B_2)^c, \ldots, B_n = A_n \cap (B_1 \cup \cdots \cup B_{n-1})^c.$$

Par additivité :

$$P\Big(\bigcup_{i=1}^n A_i\Big) = P\Big(\bigcup_{i=1}^n B_i\Big) = \sum_{i=1}^n P(B_i).$$

Par construction pour tout i, $B_i \subset A_i$, d'où $P(B_i) \leqslant P(A_i)$ et finalement,

$$P\Big(\bigcup_{i=1}^n A_i\Big) = \sum_{i=1}^n P(B_i) \leqslant \sum_{i=1}^n P(A_i).$$

□

Preuve de 7 c). Posons pour tout $n \geqslant 1$,

$$D_n = \bigcup_{i=1}^n A_i, \qquad D = \bigcup_{n \geqslant 1} D_n = \bigcup_{i \in \mathbb{N}^*} A_i.$$

La suite $(D_n)_{n \geqslant 1}$ est *croissante* et a pour limite D. Donc d'après 6 a), $P(D_n) \uparrow P(D)$ $(n \to +\infty)$. D'après 7 b) on a :

$$\forall n \geqslant 1, \quad P(D_n) \leqslant \sum_{i=1}^n P(A_i).$$

Les deux membres de cette inégalité étant les termes généraux de deux suites croissantes de réels positifs, on obtient en passant à la limite quand n tend vers l'infini :

$$P\Big(\bigcup_{i \in \mathbb{N}^*} A_i\Big) = P(D) \leqslant \sum_{i=1}^{+\infty} P(A_i),$$

ce qui prouve 7 c). Remarquons que les sommes partielles de la série convergent dans $\overline{\mathbb{R}}_+$. Bien sûr, l'inégalité obtenue n'a d'intérêt que lorsque la série de terme général $P(A_i)$ converge et a une somme strictement inférieure à 1. □

La vérification de la proposition 5.16 est maintenant complète. □

Remarque 5.17 (propriétés d'une mesure)**.** Les propriétés d'une probabilité énoncées par la proposition 5.16 s'étendent à une mesure positive quelconque, avec les exceptions suivantes. La propriété 3 est valable seulement pour $\mu(\Omega)$ fini en remplaçant 1 par $\mu(\Omega)$. La propriété 5 est vraie à condition que $\mu(A \cap B)$ soit fini. Pour la continuité séquentielle décroissante 6 b), il faut rajouter l'hypothèse $\mu(C_0) < +\infty$.

Le calcul de probabilités de réunions ou d'intersections est une question cruciale. La propriété 5 montre qu'en général on ne peut pas calculer $P(A \cup B)$ à partir de la *seule connaissance* de $P(A)$ et $P(B)$ et qu'on se heurte à la même difficulté pour $P(A \cap B)$. Le calcul des probabilités d'intersections sera discuté plus tard, à propos du conditionnement. Pour les probabilités de réunions, on peut se demander comment se généralise la propriété 5 lorsqu'on réunit plus de deux évènements. Il est facile de vérifier (faites-le !) que :

$$P(A\cup B\cup C) = P(A)+P(B)+P(C)-P(A\cap B)-P(A\cap C)-P(B\cap C)+P(A\cap B\cap C).$$

Le cas général est donné par la formule de Poincaré qui exprime $P(A_1 \cup \cdots \cup A_n)$ à l'aide des probabilités de *toutes les intersections* des A_i : 2 à 2, 3 à 3, etc.

Proposition 5.18 (formule de Poincaré)**.**
Pour tout entier $n \geqslant 2$ et tous évènements $A_1, \ldots, A_n$:

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i) + \sum_{k=2}^{n} (-1)^{k+1} \sum_{1\leqslant i_1<i_2<\ldots<i_k\leqslant n} P(A_{i_1} \cap \cdots \cap A_{i_k}). \quad (5.6)$$

Preuve. On raisonne par récurrence [14]. La formule est vraie pour $n = 2$, car dans ce cas elle se réduit à

$$P(A \cup B) = P(A) + P(B) - P(A \cap B). \quad (5.7)$$

Supposons la formule de Poincaré vraie au rang n (plus précisément on suppose que pour *toute* suite de n évènements $A_1, \ldots, A_n$, l'égalité (5.6) est vérifiée). Pour en déduire qu'elle est alors vraie au rang $n + 1$, il nous faut calculer $P\big(\bigcup_{i=1}^{n+1} A_i\big)$. On commence par appliquer (5.7) avec $A = \bigcup_{i=1}^{n} A_i$ et $B = A_{n+1}$. On obtient ainsi :

$$\begin{aligned} P\left(\bigcup_{i=1}^{n+1} A_i\right) &= P\left(\bigcup_{i=1}^{n} A_i\right) + P(A_{n+1}) - P\left(\left(\bigcup_{i=1}^{n} A_i\right) \cap A_{n+1}\right) \\ &= P\left(\bigcup_{i=1}^{n} A_i\right) + P(A_{n+1}) - P\left(\bigcup_{i=1}^{n} (A_i \cap A_{n+1})\right). \end{aligned}$$

14. Il y a une autre méthode plus élégante utilisant l'écriture d'une probabilité comme espérance d'une fonction indicatrice, voir l'exercice 7.4.

On applique maintenant l'hypothèse de récurrence (formule de Poincaré (5.6)) d'abord avec les n évènements $A_1, \dots, A_n$ puis avec les n évènements $A'_1, \dots, A'_n$, où l'on a posé $A'_i := A_i \cap A_{n+1}$. Il vient :

$$\begin{aligned} P\left(\bigcup_{i=1}^{n+1} A_i\right) &= \sum_{i=1}^{n} P(A_i) + \sum_{k=2}^{n} (-1)^{k+1} \sum_{1 \leqslant i_1 < i_2 < \ldots < i_k \leqslant n} P(A_{i_1} \cap \cdots \cap A_{i_k}) \\ &\quad + P(A_{n+1}) \\ &\quad - \sum_{i=1}^{n} P(A'_i) - \sum_{j=2}^{n} (-1)^{j+1} \sum_{1 \leqslant i_1 < i_2 < \ldots < i_j \leqslant n} P(A'_{i_1} \cap \cdots \cap A'_{i_j}) \\ &= \sum_{i=1}^{n+1} P(A_i) \qquad (5.8) \\ &\quad + \sum_{k=2}^{n} (-1)^{k+1} \sum_{1 \leqslant i_1 < i_2 < \ldots < i_k \leqslant n} P(A_{i_1} \cap \cdots \cap A_{i_k}) \qquad (5.9) \\ &\quad + (-1)^{2+1} \sum_{i=1}^{n} P(A_i \cap A_{n+1}) \qquad (5.10) \\ &\quad + \sum_{j=2}^{n} (-1)^{(j+1)+1} \sum_{1 \leqslant i_1 < i_2 < \ldots < i_j \leqslant n} P(A_{i_1} \cap \cdots \cap A_{i_j} \cap A_{n+1}) \qquad (5.11) \end{aligned}$$

Comparons ce résultat avec ce que l'on espère trouver, c'est-à-dire avec

$$\sum_{i=1}^{n+1} P(A_i) + \underbrace{\sum_{k=2}^{n+1} (-1)^{k+1} \sum_{1 \leqslant i_1 < i_2 < \ldots < i_k \leqslant n+1} P(A_{i_1} \cap \cdots \cap A_{i_k})}_{=: T_{n+1}}.$$

Cela revient à vérifier que T_{n+1} est égal à la somme des lignes (5.9) à (5.11) ci-dessus. Partageons T_{n+1} en deux blocs comme suit. Le premier bloc regroupe tous les termes tels que $i_k < n+1$ (et donc $i_k \leqslant n$ et $k \leqslant n$). On le retrouve exactement à la ligne (5.9). Le deuxième bloc regroupe tous les termes pour lesquels $i_k = n+1$. Dans ce bloc, la somme des termes pour lesquels $k = 2$ se retrouve ligne (5.10). Il reste alors la somme des termes pour lesquels $3 \leqslant k \leqslant n+1$ et $i_k = n+1$ (donc $i_{k-1} \leqslant n$). Cette somme est exactement le contenu de la ligne (5.11), comme on peut le voir en faisant le changement d'indice $k = j+1$ dans (5.11). Ceci achève la récurrence. □

5.4 Exemples

Nous examinons maintenant quelques exemples d'espaces probabilisés et de calcul de probabilités d'évènements.

Exemple 5.19. On effectue une partie de pile ou face en trois coups. Quelle est la probabilité d'obtenir pile aux premier et troisième lancers ?

On peut modéliser cette expérience en prenant $\Omega = \{\mathtt{f}, \mathtt{p}\}^3$ et pour famille d'évènements observables $\mathcal{F} = \mathcal{P}(\Omega)$ l'ensemble de toutes les parties [15] de Ω. La pièce étant supposée symétrique, nous n'avons *a priori* pas de raison de supposer que l'un des 8 triplets de résultats possibles soit favorisé ou défavorisé par rapport aux autres. Nous choisirons donc P de sorte que tous les évènements élémentaires aient même probabilité (hypothèse d'équiprobabilité), soit :

$$\forall \omega \in \Omega, \quad P(\{\omega\}) = \frac{1}{\operatorname{Card} \Omega} = \frac{1}{2^3}.$$

L'évènement B dont on veut calculer la probabilité s'écrit :

$$B = \{(\mathtt{p},\mathtt{f},\mathtt{p}); (\mathtt{p},\mathtt{p},\mathtt{p})\}.$$

D'où :

$$P(B) = \frac{1}{8} + \frac{1}{8} = \frac{1}{4}.$$

Exemple 5.20. On fait remplir un questionnaire comportant 20 questions binaires. Quelle est la probabilité qu'un candidat répondant au hasard obtienne au moins 16 bonnes réponses ?
On choisit ici :

$$\Omega = \{\text{oui}, \text{non}\}^{20}, \qquad \mathcal{F} = \mathcal{P}(\Omega).$$

Si le candidat répond complètement au hasard, on peut considérer que chacune des 2^{20} grilles de réponses possibles a la même probabilité d'apparaître (hypothèse d'équiprobabilité sur Ω). Pour tout $B \subset \Omega$, on a alors :

$$P(B) = \frac{\operatorname{Card} B}{\operatorname{Card} \Omega}.$$

En particulier pour $B = \{\text{obtention d'au moins 16 bonnes réponses}\}$,

$$P(B) = \frac{C_{20}^{16} + C_{20}^{17} + C_{20}^{18} + C_{20}^{19} + C_{20}^{20}}{2^{20}} = \frac{6196}{2^{20}} \simeq 0,006.$$

Exemple 5.21 (contrôle de production)**.** On prélève au hasard un échantillon de k pièces dans une production totale de N pièces comprenant en tout n pièces défectueuses. Le prélèvement est *sans remise* (donc $k \leqslant N$). Cherchons la probabilité de :

$$A_j = \{\text{il y a exactement } j \text{ pièces défectueuses dans l'échantillon}\}.$$

On prend pour Ω l'ensemble de toutes les parties à k éléments d'un ensemble à N éléments (ensemble de tous les échantillons possibles de taille k), $\mathcal{F} = \mathcal{P}(\Omega)$ et P l'équiprobabilité sur Ω. Il suffit alors de dénombrer tous les échantillons ayant exactement j pièces défectueuses. Un tel échantillon se construit en prenant j pièces dans le sous-ensemble des défectueuses (C_n^j choix possibles) et en complétant par $k-j$

15. Lorsque Ω est fini, il est toujours possible de faire ce choix.

pièces prises dans le sous-ensemble des non-défectueuses (C_{N-n}^{k-j} choix possibles). On en déduit :

$$P(A_j) = \frac{C_n^j C_{N-n}^{k-j}}{C_N^k} \quad \text{si} \begin{cases} 0 \leqslant j \leqslant n, \\ 0 \leqslant j \leqslant k, \\ k-j \leqslant N-n. \end{cases}$$

Si l'une de ces trois conditions n'est pas vérifiée, $P(A_j) = 0$.

Remarque 5.22. Lorsque Ω est fini, la façon la plus simple de construire une probabilité sur $(\Omega, \mathcal{P}(\Omega))$ est de choisir $P(\{\omega\}) = 1/\operatorname{card}\Omega$. On parle alors d'*équiprobabilité* ou de *probabilité uniforme* sur $(\Omega, \mathcal{P}(\Omega))$. C'est la modélisation qui s'impose naturellement lorsqu'on n'a pas de raison de penser *a priori* qu'un résultat élémentaire de l'expérience soit favorisé ou défavorisé par rapport aux autres. La situation est radicalement différente lorsque Ω est infini *dénombrable*. Sur un tel ensemble, *il ne peut pas y avoir d'équiprobabilité*. Imaginons que l'on veuille tirer une boule *au hasard* dans une urne contenant une infinité de boules numérotées de manière bijective par les entiers naturels. Soit $\{\omega_i\}$ l'évènement *tirage de la boule numérotée* i ($i \in \mathbb{N}$) et p_i sa probabilité. Par σ-additivité, les p_i vérifient nécessairement :

$$\sum_{i\in\mathbb{N}} p_i = 1.$$

Mais si les p_i sont égaux, tous les termes de la série ci-dessus valent p_0. Sa somme est alors $+\infty$ si $p_0 > 0$ ou 0 si $p_0 = 0$, il y a donc une contradiction.

Voici maintenant une caractérisation de toutes les probabilités sur les espaces au plus dénombrables.

Proposition 5.23. *Soit $\Omega = \{\omega_i \ ; \ i \in I\}$ un ensemble au plus dénombrable. La donnée d'une probabilité sur $(\Omega, \mathcal{P}(\Omega))$ équivaut à la donnée d'une famille $(p_i)_{i\in I}$ dans $\mathbb{R}_+$ telle que :*

$$\sum_{i\in I} p_i = 1$$

et des égalités

$$P(\{\omega_i\}) = p_i, \quad i \in I.$$

La probabilité P s'écrit alors $P = \sum_{i\in I} p_i \delta_{\omega_i}$, où δ_ω désigne la masse de Dirac (ou mesure de Dirac) au point ω, définie sur $\mathcal{P}(\Omega)$ par $\delta_\omega(A) = 1$ si $\omega \in A$ et $\delta_\omega(A) = 0$ si $\omega \notin A$.

Preuve. En préliminaire, notons les propriétés suivantes des mesures de Dirac sur Ω.

$$\forall \omega \in \Omega, \quad \delta_\omega(\Omega) = 1. \tag{5.12}$$

$$\forall i, j \in I, \quad \delta_{\omega_i}(\{\omega_j\}) = \begin{cases} 0 & \text{si } i \neq j, \\ 1 & \text{si } i = j. \end{cases} \tag{5.13}$$

Soit P une probabilité sur $(\Omega, \mathcal{P}(\Omega))$. Comme la tribu $\mathcal{P}(\Omega)$ possède les singletons, $P(\{\omega\})$ est défini et fini (puisque majoré par 1). La mesure P est donc finie sur les singletons et d'après la remarque 5.10, c'est une mesure ponctuelle. Cela signifie qu'il existe $J \subset I$ (donc au plus dénombrable) et une famille $\{p_i)_{i\in J}\}$ de $\mathbb{R}_+$ telle que $P = \sum_{i\in J} p_i \delta_{\omega_i}$. On peut compléter cette écriture en posant $p_i := 0$ pour $i \in I \backslash J$, pour obtenir

$$P = \sum_{i\in I} p_i \delta_{\omega_i}. \tag{5.14}$$

En écrivant que $P(\Omega) = 1$ et en utilisant (5.12), il vient $\sum_{i\in I} p_i = 1$. D'autre part on voit grâce à (5.13) que pour tout $i \in I$, $P(\{\omega_i\}) = p_i$.

Réciproquement, donnons nous une famille $(p_i)_{i\in I}$ de réels positifs de somme 1 et définissons la mesure P par (5.14). Grâce à (5.12), il est clair que $P(\Omega) = \sum_{i\in I} p_i = 1$, donc P est une probabilité. De plus pour tout $i \in I$, $P(\{\omega_i\}) = p_i$ en utilisant (5.13). □

Exemple 5.24 (une probabilité définie sur $\big(\mathbb{N}, \mathcal{P}(\mathbb{N})\big)$)**.**
Soit a un réel strictement positif fixé. On pose :

$$\forall k \in \mathbb{N}, \qquad p_k = \frac{\mathrm{e}^{-a} a^k}{k!}.$$

On remarque que p_k est le terme général positif d'une série convergente :

$$\sum_{k=0}^{+\infty} \frac{\mathrm{e}^{-a} a^k}{k!} = \mathrm{e}^{-a} \sum_{k=0}^{+\infty} \frac{a^k}{k!} = \mathrm{e}^{-a}\mathrm{e}^{a} = 1.$$

Pour tout $A \subset \mathbb{N}$, on définit :

$$P(A) = \sum_{k\in A} p_k = \sum_{k\in\mathbb{N}} p_k \delta_k(A).$$

D'après la proposition 5.23, P est une probabilité sur $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$. On l'appelle *loi de Poisson de paramètre a*. Calculons par exemple $P(2\mathbb{N})$ où $2\mathbb{N}$ désigne l'ensemble des entiers pairs.

$$P(2\mathbb{N}) = \sum_{k\in 2\mathbb{N}} p_k = \sum_{l=0}^{+\infty} \frac{\mathrm{e}^{-a} a^{2l}}{(2l)!} = \mathrm{e}^{-a} \operatorname{ch} a = \frac{1}{2}(1 + \mathrm{e}^{-2a}).$$

Une conséquence de ce résultat est : si l'on tire un nombre entier au hasard suivant une loi de Poisson, la probabilité qu'il soit pair est strictement supérieure à $\frac{1}{2}$.

Exemple 5.25 (loi uniforme sur un segment)**.** Prenons $\Omega = \mathbb{R}$, $\mathcal{F} = \mathrm{Bor}(\mathbb{R})$ et rappelons que λ_1 désigne la mesure de Lebesgue sur $\mathbb{R}$ (cf. exemple 5.13). Soit $[a, b]$ un segment fixé de $\mathbb{R}$. On définit une probabilité P sur $(\mathbb{R}, \mathrm{Bor}(\mathbb{R}))$ en posant :

$$\forall A \in \mathrm{Bor}(\mathbb{R}), \quad P(A) = \frac{\lambda_1(A \cap [a, b])}{\lambda_1([a, b])} = \frac{\lambda_1(A \cap [a, b])}{b - a}. \tag{5.15}$$

D'après l'exemple 5.12, P est une mesure et comme $P(\mathbb{R}) = 1$, c'est une probabilité. On l'appelle *loi uniforme sur* $[a, b]$. Remarquons que pour cette probabilité, tout singleton est de probabilité nulle ($\forall x \in \mathbb{R}$, $P(\{x\}) = 0$), ce qui résulte de la propriété analogue de λ_1. On voit sur cet exemple que la probabilité d'une union infinie *non dénombrable* d'évènements deux à deux disjoints n'est pas forcément égale à la somme de la famille correspondante de probabilités d'évènements. En effet,

$$1 = P([a,b]) = P\Big(\bigcup_{x\in[a,b]} \{x\}\Big) \neq \sum_{x\in[a,b]} P(\{x\}) = 0.$$

Exemple 5.26 (lois uniformes dans $\mathbb{R}^d$)**.** On peut généraliser l'exemple précédent en prenant au lieu d'un segment un borélien B de $\mathbb{R}^d$, *i.e.* $B \in \mathrm{Bor}(\mathbb{R}^d)$, tel que $0 < \lambda_d(B) < +\infty$. On définit alors une probabilité P sur $(\mathbb{R}^d, \mathrm{Bor}(\mathbb{R}^d))$, en posant :

$$\forall A \in \mathrm{Bor}(\mathbb{R}^d), \quad P(A) = \frac{\lambda_d(A \cap B)}{\lambda_d(B)}. \tag{5.16}$$

Cette probabilité est appelée *loi uniforme sur* B.

Nous allons donner maintenant une caractérisation de *toutes les probabilités* P sur $(\mathbb{R}, \mathrm{Bor}(\mathbb{R}))$. Comme le montre l'exemple de la loi uniforme sur un segment, il est illusoire d'espérer caractériser ces probabilités par les $P(\{x\})$. La situation est donc radicalement différente du cas d'un espace Ω au plus dénombrable. Au lieu des $P(\{x\})$, nous allons utiliser les $P(]a, b])$, ou les $P(]-\infty, x])$. Nous aurons besoin pour cela de la notion de fonction de répartition.

Définition 5.27 (fonction de répartition)**.** *Soit P une probabilité sur $(\mathbb{R}, \mathrm{Bor}(\mathbb{R}))$. On appelle fonction de répartition de P, l'application*

$$F : \mathbb{R} \to [0,1], \quad x \mapsto F(x) := P(]-\infty, x]).$$

Voici les propriétés générales des fonctions de répartition.

Proposition 5.28. *La fonction de répartition F d'une probabilité P sur $(\mathbb{R}, \mathrm{Bor}(\mathbb{R}))$ a les propriétés suivantes.*

a) F est croissante sur $\mathbb{R}$.

b) F a pour limite 0 en $-\infty$ et 1 en $+\infty$.

c) F est continue à droite sur $\mathbb{R}$ et a une limite à gauche en tout point $x \in \mathbb{R}$.

d) En notant $F(x-) := \lim_{\varepsilon \downarrow 0} F(x-\varepsilon)$ la limite à gauche[16] *au point x, on a*

$$\forall x \in \mathbb{R}, \quad P(\{x\}) = F(x) - F(x-). \tag{5.17}$$

De plus l'ensemble des $x \in \mathbb{R}$ tels que $F(x) \neq F(x-)$ est au plus dénombrable.

16. Cette notation est un peu abusive, puisqu'il ne s'agit pas forcément d'une valeur prise par la fonction F. Attention à ne pas confondre le « $x-$ » dans $F(x-)$ avec le « x^- » partie négative de x.

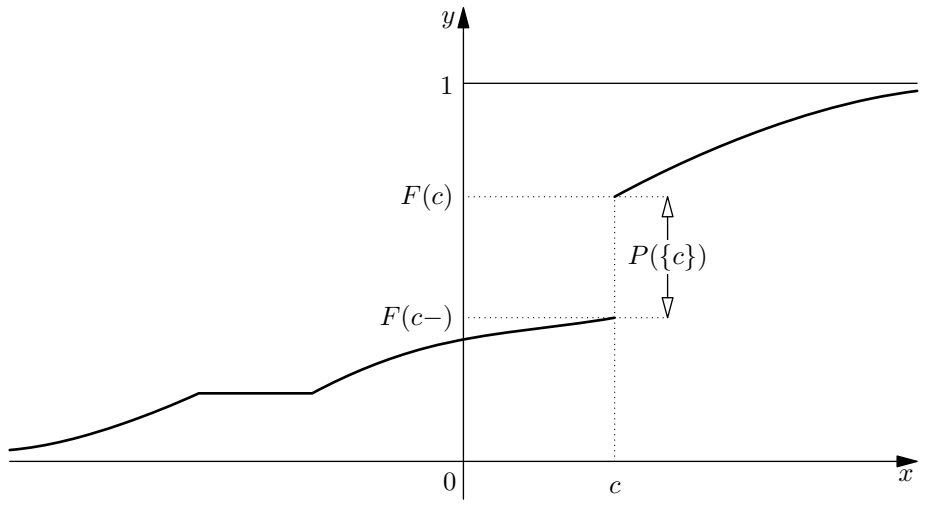


FIGURE 5.3 – Une fonction de répartition avec une discontinuité au point $x = c$

e) Si deux probabilités P_1 et P_2 sur $(\mathbb{R}, \mathrm{Bor}(\mathbb{R}))$ ont même fonction de répartition F, elles sont égales, autrement dit $P_1(B) = P_2(B)$ pour tout $B \in \mathrm{Bor}(\mathbb{R})$.

Preuve.

Preuve du a). La croissance de F sur $\mathbb{R}$ est une conséquence immédiate de la croissance de P pour l'inclusion (prop. 5.16 4). En effet si $x \leqslant x'$, $]-\infty, x] \subset]-\infty, x']$, d'où

$$F(x) = P(]-\infty, x]) \leqslant P(]-\infty, x']) = F(x').$$

Ainsi F est croissante sur $\mathbb{R}$. Il en résulte qu'elle possède une limite à gauche et une limite à droite en tout point x de $\mathbb{R}$. □

Preuve du b). Comme F est croissante, elle admet des limites en $-\infty$ et $+\infty$. On identifie la limite en $-\infty$ (dont on connaît l'existence) grâce à une suite particulière :

$$\lim_{x \to -\infty} F(x) = \lim_{n \to +\infty} F(-n) = \lim_{n \to +\infty} P(]-\infty, -n]).$$

On utilise la continuité séquentielle décroissante de P (proposition 5.16 7 b) avec $B_n =]-\infty, -n]$, en vérifiant que $\bigcap_{n \in \mathbb{N}^*}]-\infty, -n] = \varnothing$. En effet, soit x un élément de cette intersection. Alors $x \leqslant -n$ pour tout $n \in \mathbb{N}^*$ donc en passant à la limite, $x = -\infty$, ce qui est impossible puisque x est un réel. Donc cette intersection est vide et $\lim_{n \to +\infty} P(]-\infty, -n]) = P(\varnothing) = 0$. La preuve de $\lim_{x \to +\infty} F(x) = 1$ est analogue et est laissée au lecteur. □

Preuve du c). Soit $x \in \mathbb{R}$ fixé. Comme on est assuré de l'existence de la limite à droite de F en ce point, pour montrer que cette limite vaut $F(x)$ et établir la continuité à

droite de F en x, il suffit de vérifier que : $F(x) = \lim_{n\to+\infty} F(x+1/n)$. Comme $F(x+1/n) = P(]-\infty, x+1/n])$, ceci résulte de la continuité séquentielle décroissante de P appliquée aux évènements $B_n =]-\infty, x+1/n]$ en remarquant que :

$$\bigcap_{n\in\mathbb{N}^*} \Big]-\infty, x+\frac{1}{n}\Big] =]-\infty, x]. \tag{5.18}$$

En effet, pour tout $y \in]-\infty, x]$, on a $y \leqslant x \leqslant x+1/n$ pour tout $n \in \mathbb{N}^*$ et donc $y \in \bigcap_{n\in\mathbb{N}^*} B_n$ d'où $]-\infty, x]$ est inclus dans $\bigcap_{n\in\mathbb{N}^*} B_n$. Réciproquement tout y dans cette intersection vérifie : $\forall n \in \mathbb{N}^*$, $y \leqslant x+1/n$. Le passage à la limite quand n tend vers l'infini conservant cette inégalité large, on en déduit $y \leqslant x$ soit encore $y \in]-\infty, x]$. Ainsi $\bigcap_{n\in\mathbb{N}^*} B_n$ est incluse dans $]-\infty, x]$, ce qui achève la vérification de (5.18). □

Preuve du d). On remarque que pour tout $x \in \mathbb{R}$ et tout $n \in \mathbb{N}^*$,

$$P\Big(\Big]x-\frac{1}{n}, x\Big]\Big) = P(]-\infty, x]) - P\Big(\Big]-\infty, x-\frac{1}{n}\Big]\Big) = F(x) - F\Big(x-\frac{1}{n}\Big).$$

On vérifie d'autre part que :

$$\{x\} = \bigcap_{n\in\mathbb{N}^*} \Big]x-\frac{1}{n}, x\Big]. \tag{5.19}$$

En effet, le premier membre de (5.19) est clairement inclus dans le second. Réciproquement, si y est un élément quelconque de cette intersection, pour tout $n \in \mathbb{N}^*$, $x-1/n < y \leqslant x$, donc en passant à la limite quand n tend vers l'infini, $x \leqslant y \leqslant x$ d'où $y = x$. Ceci termine la justification de (5.19). Comme on sait que F a une limite à gauche au point x, on en déduit par continuité de P pour la suite décroissante d'évènements $B_n =]x-1/n, x]$:

$$P(\{x\}) = \lim_{n\to+\infty} P\Big(\Big]x-\frac{1}{n}, x\Big]\Big) = \lim_{n\to+\infty} \big(F(x) - F(x-1/n)\big) = F(x) - F(x-).$$

Enfin, considérons la famille de réels *positifs* $\big\{P(\{x\}),\ x \in \mathbb{R}\big\}$. Pour toute partie finie K de $\mathbb{R}$,

$$S_K := \sum_{x\in K} P(\{x\}) = P(K) \leqslant 1.$$

Les sommes finies S_K étant ainsi uniformément bornées par 1, la famille est sommable par la proposition 2.42 a). On sait qu'alors l'ensemble des éléments non nuls de cette famille est au plus dénombrable, cf. proposition 2.44. □

Nous admettrons le e), dont la preuve sort du programme de cet ouvrage. □

Remarque 5.29. Les probabilités d'intervalles s'expriment facilement à l'aide de la fonction de répartition. Il faut prendre garde aux extrémités. En général on a la

égalités suivantes pour une probabilité P sur $(\mathbb{R}, \mathrm{Bor}(\mathbb{R}))$, de fonction de répartition F, avec a et b réels quelconques tels que $a < b$,

$$\begin{aligned} P(\,]a,b]) &= F(b) - F(a), && (5.20)\\ P([a,b]) &= F(b) - F(a-), && (5.21)\\ P([a,b[\,) &= F(b-) - F(a-), && (5.22)\\ P(\,]a,b[\,) &= F(b-) - F(a), && (5.23)\\ P(\,]-\infty,a[\,) &= F(a-), && (5.24)\\ P(\,]b,+\infty[\,) &= 1 - F(b), && (5.25)\\ P([b,+\infty[\,) &= 1 - F(b-). && (5.26) \end{aligned}$$

Ces formules n'ont pas à être apprises par coeur. Elles se retrouvent avec un peu de réflexion. Elles se simplifient lorsque F est continue en a ou b. En particulier si F est continue aux points a et b, les 4 intervalles d'extrémités a et b ont même probabilité.

Nous admettrons le théorème suivant qui permet de construire toutes les probabilités sur $(\mathbb{R}, \mathrm{Bor}(\mathbb{R}))$.

Théorème 5.30. *Soit F une fonction croissante et continue à droite sur $\mathbb{R}$, ayant pour limite 0 en $-\infty$ et 1 en $+\infty$. Il existe une unique probabilité P sur $(\mathbb{R}, \mathrm{Bor}(\mathbb{R}))$ telle que*

$$\forall a, b \in \mathbb{R},\ \textit{avec } a \leqslant b, \quad P(\,]a,b]) = F(b) - F(a).$$

Alors F est la fonction de répartition de P.

5.5 Remarques sur le choix d'un modèle

Envisageons l'expérience « on jette deux dés ... ». Un évènement élémentaire pour cette expérience est la sortie de deux nombres entiers de $[\![1,6]\!]$, distincts ou non .

Une première modélisation consiste à choisir $\Omega = \{1,2,3,4,5,6\}^2$, à prendre comme ensemble d'évènements observables $\mathcal{F} = \mathcal{P}(\Omega)$ et à attribuer à chaque évènement élémentaire $\{(i,j)\}$ la probabilité $1/36$ (modèle (1)). Il est commode de représenter Ω sous la forme d'un tableau à 36 cases correspondant chacune à un évènement élémentaire. On peut alors définir la probabilité comme un rapport d'aires (modèle (1′)) : $P(B) = \dfrac{\text{Aire de } B}{\text{Aire de } \Omega}$. Ce modèle (1) ou (1′) est accepté d'autant plus facilement qu'on aura précisé que les deux dés sont distinguables (par exemple un dé rouge et un vert, ou deux dés blancs lancés chacun sur une table différente). On peut ainsi distinguer l'évènement $A' = \{(2,3)\}$ de l'évènement $A'' = \{(3,2)\}$. Aux questions : quelle est la probabilité de l'évènement $A :=$

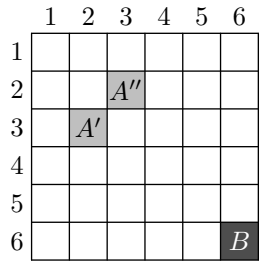


FIGURE 5.4 – Modèle (1′)

{obtenir un 2 et un 3}, quelle est celle de l'évènement $B := \{\text{obtenir un double 6}\}$? On répondra *naturellement* :

$$P(A) = P(\{A' \cup A''\}) = \frac{1}{36} + \frac{1}{36} = \frac{1}{18} \text{ et } P(B) = P(\{(6,6)\}) = \frac{1}{36}.$$

Supposons maintenant que les deux dés ne sont plus distinguables, par exemple jet de deux dés blancs sur une même table ou jet de deux dés de couleurs différentes, l'observation étant notée sans tenir compte de la couleur. Voici une modélisation possible. On ne considère comme famille $\tilde{\mathcal{F}}$ d'évènements observables que les parties *symétriques* de Ω (c'est-à-dire invariantes par $(i,j) \mapsto (j,i)$). Cela revient à remplacer $\Omega = [\![1,6]\!]^2$ par l'ensemble $\tilde{\Omega}$, dont les éléments (et donc les évènements élémentaires) sont de deux types : les doubles et les paires de deux chiffres distincts. Alors $\tilde{\mathcal{F}} = \mathcal{P}(\tilde{\Omega})$ est constituée de 2^{21} évènements, ce modèle est moins riche en information que le précédent (2^{36} évènements). On peut *raisonnablement* penser que la couleur n'influe pas sur les résultats et attribuer pour rester cohérent avec notre première modélisation la probabilité 1/36 aux *doubles* et 1/18 aux paires de deux chiffres distincts. On voit ainsi un exemple de situation pourtant très simple où *il n'y a pas* équiprobabilité des évènements élémentaires (modèle (2)). Remarquons là aussi que l'on peut donner une représentation géométrique de ce modèle en repliant le carré le long de sa diagonale $i = j$. Les évènements élémentaires sont maintenant représentés par un carré ou un triangle et leur probabilité est définie comme un rapport d'aires (modèle (2′)).

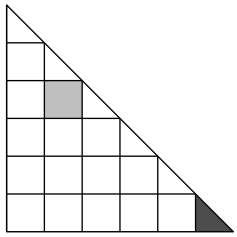


FIGURE 5.5 – Modèle (2′)

Un troisième modèle peut être proposé et il sera souvent adopté implicitement par des débutants à qui on aura dit : « on jette deux dés de même couleur ». On considère 21 évènements élémentaires : les 6 doubles et les 15 paires à deux chiffres distincts [17].

$$\Omega = \{\underbrace{1,\ 2,\ldots,6}_{\text{doubles}},\ \underbrace{12,\ldots,16,\ 23,\ldots 26,\ 34,\ldots,\ 56}_{\text{distincts}}\}$$

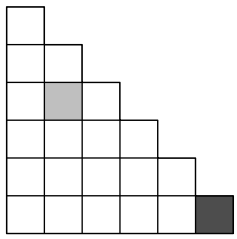


FIGURE 5.6 – Modèle (3′)

On prend comme ensemble d'évènements observables $\mathcal{F} = \mathcal{P}(\Omega)$ ensemble des parties de Ω. On définit la probabilité P par l'hypothèse d'équiprobabilité (modèle (3)). On obtient une représentation géométrique équivalente à l'aide de la figure 5.6.

Les évènements élémentaires sont représentés par un carré et la probabilité est définie comme un rapport d'aires (modèle (3′)). Avec ce modèle, la probabilité d'obtenir un double six est la même que celle d'obtenir un deux et un trois et vaut $\frac{1}{21}$.

17. La paire « 1 et 2 » est notée « 12 »,...

On peut imaginer toute une liste d'excellents arguments qui militent en faveur des modèles (1) et (2) contre le modèle (3), par exemple : « on jette deux dés de couleurs différentes et on les filme avec deux caméras, une couleur et l'autre noir et blanc... », « il y a deux façons d'obtenir 2 et 3 et une seule de faire un double 6 ».

Tous ces arguments relèvent d'une analyse *a priori* de l'expérience et pas de la théorie mathématique des probabilités. Chacun des modèles présentés ci-dessus a sa cohérence comme objet mathématique. La question pertinente est : *parmi ces modèles, lequel (lesquels ?) représente(nt) le mieux la réalité ?*

Pour un physicien, ce qui fait la valeur d'un modèle est sa capacité à permettre la prévision du résultat d'expériences. Ainsi certaines expériences sont « expliquées » par la théorie de la relativité alors que leur résultat ne s'accorde pas avec la théorie de la mécanique classique.

Si l'on se place dans le cadre des modèles (1) et (2), la loi forte des grands nombres nous dit que si l'on jette les dés une infinité de fois, la fréquence d'apparition du double six va converger [18] vers $\frac{1}{36}$ tandis qu'avec le modèle (3) cette fréquence convergera vers $\frac{1}{21}$. La réalisation d'un grand nombre de lancers permet donc ici de constater que les modèles (1) et (2) rendent mieux compte de la réalité que le (3). Une question importante est alors : que faut-il entendre par *grand nombre* ? Cette question sera discutée ultérieurement.

5.6 Probabilités conditionnelles

5.6.1 Introduction

Comment doit-on modifier la probabilité que l'on attribue à un évènement lorsque l'on dispose d'une *information supplémentaire* ? Le concept de probabilité conditionnelle permet de répondre à cette question.

Par exemple, soit E l'ensemble des licenciés d'une fédération d'escrime, F le sous-ensemble des escrimeuses, H celui des escrimeurs, G l'ensemble des membres de la fédération qui sont gauchers et D celui des droitiers. On choisit un individu au hasard dans E, chaque licencié a donc une probabilité $1/\operatorname{card} E$ d'être choisi, ce qui revient à prendre comme espace probabilisé $\Omega = E$, $\mathcal{F} = \mathcal{P}(E)$ et P la probabilité uniforme (ou équiprobabilité) sur E. Si cet individu est un homme, quelle est la probabilité qu'il soit gaucher ? Sans condition de sexe, la probabilité que l'individu choisi au hasard soit gaucher est naturellement $\operatorname{card} G/\operatorname{card} E$. Comment tenir compte de l'information supplémentaire apportée ici par la réalisation de l'évènement H ? On considère que tout se passe comme si d'emblée, on avait limité le tirage au sort aux éléments du sous-ensemble H. Ceci conduit à attribuer *dans ce nouveau contexte*, la probabilité $\operatorname{card}(G \cap H)/\operatorname{card} H$ au choix d'un individu gaucher. Cette probabilité que l'individu

18. En un sens qui sera précisé dans le chapitre sur la loi des grands nombres.

choisi soit gaucher *sachant* que c'est un homme peut s'écrire :

$$\frac{\operatorname{card}(G \cap H)}{\operatorname{card} H} = \frac{\dfrac{\operatorname{card}(G \cap H)}{\operatorname{card} E}}{\dfrac{\operatorname{card} H}{\operatorname{card} E}} = \frac{P(G \cap H)}{P(H)}.$$

Par analogie, nous donnons dans le cas général la définition formelle suivante.

Définition 5.31 (probabilité conditionnelle)**.** *Soit H un évènement de probabilité* non nulle*. Pour tout évènement A, on définit :*

$$P(A \mid H) = \frac{P(A \cap H)}{P(H)},$$

appelée probabilité conditionnelle *de l'évènement A* sous l'hypothèse *H.*

Remarquons que pour l'instant, il ne s'agit que d'un jeu d'écriture. On a simplement défini un réel $P(A \mid H)$ pour que :

$$P(A \cap H) = P(A \mid H)P(H).$$

Ce qui fait l'intérêt du concept de probabilité conditionnelle, c'est qu'il est souvent bien plus facile d'attribuer *directement* une valeur à $P(A \mid H)$ en tenant compte des conditions expérimentales (liées à l'*information* H) et d'en déduire ensuite la valeur de $P(A \cap H)$. Le raisonnement implicite alors utilisé est : tout espace probabilisé modélisant correctement la réalité expérimentale devrait fournir cette valeur pour $P(A \mid H)$.

Exemple 5.32. Une urne contient r boules rouges et v boules vertes. On en tire deux l'une après l'autre, sans remise. Quelle est la probabilité d'obtenir deux rouges ?

Notons H et A les évènements :

$$H = \{\text{rouge au } 1^{\text{er}} \text{ tirage}\}, \quad A = \{\text{rouge au } 2^{\text{e}} \text{ tirage}\}.$$

Un espace probabilisé $(\Omega, \mathcal{F}, P)$ modélisant correctement cette expérience devrait vérifier :

$$\begin{aligned} P(H) &= \frac{r}{r+v}, \\ P(A \mid H) &= \frac{r-1}{r+v-1}. \end{aligned}$$

En effet, si H est réalisé, le deuxième tirage a lieu dans une urne contenant $r+v-1$ boules dont $r-1$ rouges. On en déduit :

$$P(\text{deux rouges}) = P(A \cap H) = P(A \mid H)P(H) = \frac{r-1}{r+v-1} \times \frac{r}{r+v}.$$

On aurait pu arriver au même résultat en prenant pour Ω l'ensemble des arrangements de deux boules parmi $r+v$, muni de l'équiprobabilité et en faisant du dénombrement :

$$\operatorname{card}\Omega = A^2_{r+v} = (r+v)(r+v-1), \quad \operatorname{card}(A\cap H) = A^2_r = r(r-1).$$

d'où :

$$P(A\cap H) = \frac{r(r-1)}{(r+v)(r+v-1)}.$$

Notons d'ailleurs que $\operatorname{card} H = r(r+v-1)$ d'où

$$P(H) = \frac{\operatorname{card} H}{\operatorname{card}\Omega} = \frac{r(r+v-1)}{(r+v)(r+v-1)} = \frac{r}{r+v}.$$

En appliquant la définition formelle de $P(A \mid H)$ on retrouve :

$$P(A \mid H) = \frac{P(A\cap H)}{P(H)} = \frac{r(r-1)}{(r+v)(r+v-1)} \times \frac{r+v}{r} = \frac{r-1}{r+v-1},$$

ce qui est bien la valeur que nous avions attribuée *a priori* en analysant les conditions expérimentales.

Remarque 5.33. Il importe de bien comprendre que l'écriture « $A \mid H$ » ne désigne pas un nouvel évènement [19] différent de A. Quand on écrit $P(A \mid H)$, ce que l'on a modifié, ce n'est pas l'évènement A, mais la valeur numérique qui lui était attribuée par la fonction d'ensembles P. Il serait donc en fait plus correct d'écrire $P_H(A)$ que $P(A \mid H)$. On conservera néanmoins cette dernière notation essentiellement pour des raisons typographiques : $P(A \mid H_1 \cap H_2 \cap H_3)$ est plus lisible que $P_{H_1\cap H_2\cap H_3}(A)$.

5.6.2 Propriétés

Proposition 5.34. *Soit $(\Omega, \mathcal{F}, P)$ un espace probabilisé et H un évènement* fixé *tel que $P(H) \neq 0$. Alors la fonction d'ensembles $P(\,.\mid H)$ définie par :*

$$P(\,.\mid H): \quad \mathcal{F} \to [0,1] \quad B \mapsto P(B \mid H)$$

est une nouvelle probabilité sur $(\Omega, \mathcal{F})$.

La preuve a déjà été donnée à l'occasion de l'exemple 5.12. Une conséquence immédiate est que la fonction d'ensembles $P(\,.\mid H)$ vérifie toutes les propriétés de la proposition 5.16.

Corollaire 5.35. *La fonction d'ensembles $P(\,.\mid H)$ vérifie :*

1. *$P(\varnothing \mid H) = 0$, $P(\Omega \mid H) = 1$ et si $A \supset H$, $P(A \mid H) = 1$.*

19. En fait cette écriture prise isolément (sans le P) *n'a pas de sens* et ne devrait *jamais* être utilisée. Le symbole $\mid$ ne représente pas une opération sur les évènements qui l'entourent.

2. *Si les A_i sont deux à deux disjoints :*

$$P\Big(\bigcup_{i=1}^{n} A_i \mid H\Big) = \sum_{i=1}^{n} P(A_i \mid H).$$

3. *Pour tout $B \in \mathcal{F}$, $P(B^c \mid H) = 1 - P(B \mid H)$.*
4. *Pour tous $A \in \mathcal{F}$ et $B \in \mathcal{F}$, si $A \subset B$, $P(A \mid H) \leqslant P(B \mid H)$.*
5. *Pour tous $A \in \mathcal{F}$ et $B \in \mathcal{F}$,*

$$P(A \cup B \mid H) = P(A \mid H) + P(B \mid H) - P(A \cap B \mid H).$$

6. *Pour toute suite $(A_i)_{i\geqslant 1}$ d'évènements :*

$$P\Big(\bigcup_{i\in\mathbb{N}^*} A_i \mid H\Big) \leqslant \sum_{i=1}^{+\infty} P(A_i \mid H).$$

7. *Si $B_n \uparrow B$, $P(B_n \mid H) \uparrow P(B \mid H)$, $(n \to +\infty)$.*
8. *Si $C_n \downarrow C$, $P(C_n \mid H) \downarrow P(C \mid H)$, $(n \to +\infty)$.*

Nous n'avons vu jusqu'ici aucune formule permettant de calculer la probabilité d'une intersection d'évènements à l'aide des probabilités de ces évènements. Une telle formule *n'existe pas dans le cas général*. Les probabilités conditionnelles fournissent une méthode générale tout à fait naturelle pour calculer une probabilité d'intersection.

Proposition 5.36 (règle des conditionnements successifs)**.**
Si $A_1, \ldots, A_n$ sont n évènements tels que $P(A_1 \cap \ldots \cap A_{n-1}) \neq 0$, on a :

$$\begin{aligned} P(A_1 \cap \ldots \cap A_n) \quad &= \quad P(A_1)P(A_2 \mid A_1)P(A_3 \mid A_1 \cap A_2) \times \cdots \\ &\qquad \cdots \times P(A_n \mid A_1 \cap \ldots \cap A_{n-1}). \end{aligned}$$

Preuve. Pour $1 \leqslant i \leqslant n-1$, on a $\bigcap_{j=1}^{n-1} A_j \subset \bigcap_{j=1}^{i} A_j$ d'où :

$$0 < P\Big(\bigcap_{j=1}^{n-1} A_j\Big) \leqslant P\Big(\bigcap_{j=1}^{i} A_j\Big).$$

Donc $P\big(\bigcap_{j=1}^{i} A_j\big)$ n'est nul pour aucun $i \leqslant n-1$ et on peut conditionner par l'évènement $\bigcap_{j=1}^{i} A_j$. Ceci légitime le calcul suivant :

$$\begin{aligned} & P(A_1)P(A_2 \mid A_1)P(A_3 \mid A_1 \cap A_2) \times \cdots \times P(A_n \mid A_1 \cap \ldots \cap A_{n-1}) \\ = \quad & P(A_1) \times \frac{P(A_1 \cap A_2)}{P(A_1)} \times \frac{P(A_1 \cap A_2 \cap A_3)}{P(A_1 \cap A_2)} \times \cdots \times \frac{P(A_1 \cap \ldots \cap A_n)}{P(A_1 \cap \ldots \cap A_{n-1})} \\ = \quad & P(A_1 \cap \ldots \cap A_n), \end{aligned}$$

après simplifications en chaîne de toutes ces fractions. □

Les probabilités conditionnelles permettent aussi de calculer la probabilité d'un évènement en conditionnant par *tous les cas possibles*. Du point de vue ensembliste, ces *cas possibles* correspondent à une *partition* de Ω.

Définition 5.37 (partition)**.** *On dit qu'une famille $(H_i)_{i\in I}$ de parties de Ω est une partition de Ω si elle vérifie les trois conditions :*
- $\forall i \in I$, $H_i \neq \varnothing$.
- $\Omega = \bigcup_{i\in I} H_i$.
- *Les H_i sont deux à deux disjoints ($i \neq j \Rightarrow H_i \cap H_j = \varnothing$).*

Proposition 5.38 (conditionnement par les cas possibles [20])**.**
(i) Si H est tel que $P(H) \neq 0$ et $P(H^c) \neq 0$, on a

$$\forall A \in \mathcal{F}, \quad P(A) = P(A \mid H)P(H) + P(A \mid H^c)P(H^c).$$

(ii) Si $H_1, \ldots, H_n$ est une partition finie de Ω en évènements de probabilité non nulle,

$$\forall A \in \mathcal{F}, \quad P(A) = \sum_{i=1}^{n} P(A \mid H_i)P(H_i).$$

(iii) Si $(H_i)_{i\in\mathbb{N}}$ est une partition de Ω telle qu'aucun $P(H_i)$ ne soit nul,

$$\forall A \in \mathcal{F}, \quad P(A) = \sum_{i=0}^{+\infty} P(A \mid H_i)P(H_i).$$

Preuve. Il suffit de vérifier (*iii*), les deux premières propriétés se démontrant de façon analogue. Comme $(H_i)_{i\in\mathbb{N}}$ est une partition de Ω,

$$A = A \cap \Omega = A \cap \Big(\bigcup_{i\in\mathbb{N}} H_i \Big) = \bigcup_{i\in\mathbb{N}} (A \cap H_i)$$

et cette réunion est disjointe car les H_i étant deux à deux disjoints, il en est de même pour les $A \cap H_i$. Donc par σ-additivité :

$$P(A) = \sum_{i=0}^{+\infty} P(A \cap H_i) = \sum_{i=0}^{+\infty} P(A \mid H_i)P(H_i).$$

□

Lorsqu'on a une partition de Ω en n hypothèses ou cas possibles H_i et que l'on sait calculer les $P(H_i)$ et les $P(A \mid H_i)$, on peut se poser le problème inverse : calculer $P(H_j \mid A)$ à l'aide des quantités précédentes. La solution est donnée par la formule suivante quelquefois appelée (abusivement) formule des probabilités des causes.

20. Ou formule des probabilités totales.

Proposition 5.39 (formule de Bayes)**.**
Soit A un évènement de probabilité non nulle. Si les évènements H_i $(1 \leqslant i \leqslant n)$ forment une partition de Ω et si aucun $P(H_i)$ n'est nul, on a pour tout $j \in [\![1, n]\!]$:

$$P(H_j \mid A) = \frac{P(A \mid H_j)P(H_j)}{\sum_{i=1}^{n} P(A \mid H_i)P(H_i)}.$$

Preuve. On note que par définition des probabilités conditionnelles :

$$P(H_j \mid A) = \frac{P(A \cap H_j)}{P(A)} = \frac{P(A \mid H_j)P(H_j)}{P(A)}.$$

Il ne reste plus qu'à développer $P(A)$ en conditionnant par la partition $(H_i, 1 \leqslant i \leqslant n)$ comme à la proposition 5.38. □

La même formule se généralise au cas d'une partition dénombrable. Ces formules sont plus faciles à retrouver qu'à mémoriser.

5.6.3 Quelques exemples

Exemple 5.40. On considère deux urnes U_1 et U_2. L'urne U_1 contient r_1 boules rouges et v_1 boules vertes. L'urne U_2 contient r_2 boules rouges et v_2 boules vertes. On lance un dé. S'il indique le chiffre 1, on choisit l'urne U_1, sinon on choisit U_2. Dans chaque cas on effectue deux tirages avec remise dans l'urne choisie. Quelle est la probabilité d'obtenir une rouge au premier tirage ? deux rouges en tout ?
Adoptons les notations d'évènements suivantes :

$$R = \{\text{rouge au } 1^{\text{er}} \text{ tirage}\}, \quad R' = \{\text{rouge au } 2^{\text{e}} \text{ tirage}\},$$

$$H_1 = \{\text{choix de l'urne } U_1\}, \quad H_2 = H_1^{\text{c}} = \{\text{choix de l'urne } U_2\}.$$

Il est facile de calculer directement $P(R \mid H_i)$ et $P(R \cap R' \mid H_i)$ pour $i = 1, 2$. En effet, une fois l'urne U_i choisie, on a un problème classique de tirages avec remise dans la même urne que l'on peut traiter (par exemple) par le dénombrement. On a ainsi :

$$P(R \mid H_i) = \frac{r_i}{r_i + v_i}, \quad P(R \cap R' \mid H_i) = \Big(\frac{r_i}{r_i + v_i}\Big)^2.$$

La formule de conditionnement par la partition $\{H_1, H_2\}$ donne :

$$\begin{aligned} P(R) &= P(R \mid H_1)P(H_1) + P(R \mid H_2)P(H_2) \\ &= \frac{1}{6}\,\frac{r_1}{r_1 + v_1} + \frac{5}{6}\,\frac{r_2}{r_2 + v_2} \end{aligned}$$

et

$$\begin{aligned} P(R \cap R') &= P(R \cap R' \mid H_1)P(H_1) + P(R \cap R' \mid H_2)P(H_2) \\ &= \frac{1}{6}\Big(\frac{r_1}{r_1 + v_1}\Big)^2 + \frac{5}{6}\Big(\frac{r_2}{r_2 + v_2}\Big)^2. \end{aligned}$$

Exemple 5.41. Un questionnaire à choix multiples propose m réponses pour chaque question. Soit $p \in]0,1[$ la probabilité qu'un étudiant connaisse la réponse à une question donnée. S'il ignore la réponse, il choisit au hasard l'une des réponses proposées. Quelle est pour le correcteur, la probabilité qu'un étudiant connaisse vraiment la bonne réponse lorsqu'il l'a donnée ?

Notons :

$$\begin{aligned} B &= \{\text{l'étudiant donne la bonne réponse}\}, \\ C &= \{\text{l'étudiant connaît la bonne réponse}\}. \end{aligned}$$

On cherche $P(C \mid B)$. Avec ces notations, les données de l'énoncé se traduisent par :

$$P(C) = p, \quad P(C^c) = 1 - p, \qquad P(B \mid C) = 1, \quad P(B \mid C^c) = \frac{1}{m}.$$

On en déduit :

$$\begin{aligned} P(C \mid B) &= \frac{P(B \cap C)}{P(B)} \\ &= \frac{P(B \mid C)P(C)}{P(B \mid C)P(C) + P(B \mid C^c)P(C^c)} \\ &= \frac{1 \times p}{1 \times p + \frac{1}{m}(1-p)} = \frac{mp}{1 + (m-1)p}. \end{aligned}$$

Notons $P(C \mid B) = f(m,p)$. Pour p fixé, $f(\,.\,,p)$ croît sur $\mathbb{N}^*$ depuis $f(1,p) = p$ jusqu'à $\lim_{m\to+\infty} f(m,p) = 1$. D'autre part pour m fixé, $f(m,\,.\,)$ croît sur $]0,1[$ avec $\lim_{p\to 0} f(m,p) = 0$ et $\lim_{p\to 1} f(m,p) = 1$. On a pour $0 < p < 1$:

$$\frac{P(C \mid B)}{p} = \frac{m}{1 + (m-1)p} \geqslant 1, \quad \text{d'où} \quad P(C \mid B) > p.$$

Si $p = 0$ (resp. $p = 1$), on ne peut plus conditionner par C (resp. par C^c) et dans le calcul ci-dessus, seule l'égalité $P(C \mid B) = P(B \cap C)/P(B)$ reste valable. Pour $p = 0$, $P(B \cap C) \leqslant P(C) = 0$, d'où $P(C \mid B) = 0$. Pour $p = 1$, $P(B \cap C^c) \leqslant P(C^c) = 0$ d'où $P(B \cap C^c) = 0$ et $P(B \cap C) = P(B) - P(B \cap C^c) = P(B)$ d'où $P(C \mid B) = 1$. Tous ces résultats sont conformes à l'intuition.

Exemple 5.42. Un test sanguin a une probabilité de 0,95 de détecter un certain virus lorsque celui-ci est effectivement présent. Il donne néanmoins un *faux résultat positif* pour 1% des personnes non infectées. Si 0,5% de la population est porteuse du virus, quelle est la probabilité qu'une personne ait le virus sachant qu'elle a un test positif ?

Notons :

$$\begin{aligned} V &= \{\text{la personne testée a le virus}\}, \\ T &= \{\text{la personne testée a un test positif}\}. \end{aligned}$$

On cherche $P(V \mid T)$. Or on sait que $P(V) = 0{,}005$, $P(T \mid V) = 0{,}95$ et $P(T \mid V^c) = 0{,}01$. On en déduit :

$$P(V \mid T) = \frac{P(T \cap V)}{P(T)} = \frac{P(T \mid V)P(V)}{P(T \mid V)P(V) + P(T \mid V^c)P(V^c)} = \frac{0{,}95 \times 0{,}005}{0{,}95 \times 0{,}005 + 0{,}01 \times 0{,}995} \simeq 0{,}323.$$

On voit ainsi que contrairement à ce que l'on aurait pu croire, le test n'est pas fiable : si la personne présente un test positif, il y a approximativement deux chances sur trois qu'elle ne soit pas porteuse du virus !

Exemple 5.43. Revenons sur le problème de dés étudié à la sous-section 5.2.3, à savoir l'attribution d'une probabilité à l'évènement *E première obtention de la somme 9 avant celle de la somme 7*. Les probabilités conditionnelles permettent d'en proposer une solution simple, sans calcul de série. En rappelant que F_1, G_1, H_1 désignent respectivement l'évènement *obtention au premier lancer d'une somme 9 (resp. 7, resp. ni 7 ni 9)*, la formule des probabilités totales nous donne :

$$P(E) = P(E \mid F_1)P(F_1) + P(E \mid G_1)P(G_1) + P(E \mid H_1)P(H_1). \tag{5.27}$$

On a clairement $P(E \mid F_1) = 1$ et $P(E \mid G_1) = 0$. Pour attribuer une valeur à $P(E \mid H_1)$, on considère que l'obtention d'une somme autre que 7 ou 9 au premier lancer ne devrait pas influer sur l'apparition ultérieure du 7 ou du 9. Pour traduire cette idée, on pose $P(E \mid H_1) = P(E)$. En reportant ces valeurs dans (5.27), il vient

$$P(E) = P(F_1) + P(E)P(H_1) = \frac{1}{9} + \frac{13}{18}P(E), \quad \text{d'où} \quad \Big(1 - \frac{13}{18}\Big)P(E) = \frac{1}{9},$$

ce qui se résout en

$$P(E) = \frac{1}{9} \times \frac{18}{5} = \frac{2}{5}.$$

On retrouve ainsi la valeur obtenue à la sous-section 5.2.3, lorsque nous avons esquissé la construction d'un espace $(\Omega, \mathcal{F}, P)$ modélisant cette expérience aléatoire. Le point clé dans le raisonnement ci-dessus est l'égalité $P(E \mid H_1) = P(E)$, qui exprime *l'indépendance* des évènements H_1 et E, une notion que nous allons étudier maintenant.

5.7 Indépendance

5.7.1 Indépendance de deux évènements

Soient A et B deux évènements de probabilité non nulle. Il arrive que la connaissance de la réalisation de A ne modifie pas notre information sur celle de B, autrement dit que $P(B \mid A) = P(B)$. C'est le cas par exemple lorsque l'on fait deux tirages avec remise et que la réalisation de A ne dépend que du résultat du premier tirage, celle de B que du deuxième. Symétriquement on aura dans cet exemple $P(A \mid B) = P(A)$. Cette remarque se généralise :

Proposition 5.44. *Si A et B sont des évènements de probabilité non nulle, les trois égalités suivantes sont équivalentes :*
(i) $P(B \mid A) = P(B)$,
(ii) $P(A \mid B) = P(A)$,
(iii) $P(A \cap B) = P(A)P(B)$.

Preuve. Comme $P(A) \neq 0$ et $P(B) \neq 0$, on a la chaîne d'équivalences :

$$\frac{P(A \cap B)}{P(A)} = P(B) \Leftrightarrow P(A \cap B) = P(A)P(B) \Leftrightarrow \frac{P(A \cap B)}{P(B)} = P(A).$$

□

D'autre part la relation *(iii)* est toujours vérifiée dans le cas dégénéré où $P(A) = 0$ ou $P(B) = 0$. En effet, on a alors à la fois $P(A)P(B) = 0$ et $0 \leqslant P(A \cap B) \leqslant \min\big(P(A), P(B)\big) = 0$ d'où $P(A \cap B) = 0$. Ainsi la relation *(iii)* est un peu plus générale que *(i)* et *(ii)*. Elle a aussi sur les deux autres l'avantage de la symétrie d'écriture. C'est elle que l'on retient pour définir l'indépendance.

Définition 5.45. *Soit $(\Omega, \mathcal{F}, P)$ un espace probabilisé. Deux évènements A et B de cet espace sont dits indépendants lorsque :*

$$P(A \cap B) = P(A)P(B).$$

Exemple 5.46. On jette deux fois le même dé. Les évènements

$$\begin{aligned} A &= \{\text{obtention d'un chiffre pair au premier lancer}\}, \\ B &= \{\text{obtention du 1 au deuxième lancer}\}, \end{aligned}$$

sont indépendants.

En effet, en prenant $\Omega = [\![1, 6]\!]^2$, $\mathcal{F} = \mathcal{P}(\Omega)$ et P l'équiprobabilité, on vérifie que :

$$P(A) = \frac{3 \times 6}{36} = \frac{1}{2}, \quad P(B) = \frac{6 \times 1}{36} = \frac{1}{6},$$

$$P(A \cap B) = \frac{3 \times 1}{36} = \frac{1}{12}, \quad P(A)P(B) = \frac{1}{2} \times \frac{1}{6} = \frac{1}{12}.$$

Remarques 5.47.
- Si A est un évènement tel que $P(A) = 0$ ou $P(A) = 1$, alors il est indépendant de tout évènement, *y compris de lui-même* (c'est le cas en particulier pour Ω et $\varnothing$).
- Deux évènements *incompatibles* A et B avec $P(A) > 0$ et $P(B) > 0$ ne sont *jamais indépendants*. En effet $A \cap B = \varnothing$ implique $P(A \cap B) = 0$ or $P(A)P(B) \neq 0$.
- L'indépendance de deux évènements A et B n'est pas une propriété intrinsèque aux évènements, elle est toujours relative au modèle $(\Omega, \mathcal{F}, P)$ que l'on a choisi. Voici un exemple pour l'illustrer.

Exemple 5.48. Une urne contient 12 boules numérotées de 1 à 12. On en tire une au hasard et on considère les évènements :

$$A = \{\text{tirage d'un nombre pair}\}, \quad B = \{\text{tirage d'un multiple de 3}\}.$$

L'espace probabilisé qui s'impose naturellement ici est $\Omega = [\![1, 12]\!]$ muni de l'équiprobabilité P. Les évènements A et B s'écrivent :

$$A = \{2, 4, 6, 8, 10, 12\}, \quad B = \{3, 6, 9, 12\}, \quad A \cap B = \{6, 12\}.$$

Dans ce modèle,

$$P(A) = \frac{6}{12} = \frac{1}{2}, \quad P(B) = \frac{4}{12} = \frac{1}{3},$$
$$P(A \cap B) = \frac{2}{12} = \frac{1}{6}, \quad P(A)P(B) = \frac{1}{2} \times \frac{1}{3} = \frac{1}{6},$$

donc A et B sont indépendants.

On *rajoute maintenant dans l'urne une boule numérotée treize* et on recommence l'expérience. Les évènements A et B restent les mêmes, mais *le modèle a changé*. On a maintenant l'équiprobabilité P' sur $\Omega' = [\![1, 13]\!]$ et

$$P'(A) = \frac{6}{13}, \quad P'(B) = \frac{4}{13}, \quad P'(A \cap B) = \frac{2}{13},$$

mais

$$P'(A)P'(B) = \frac{6}{13} \times \frac{4}{13} = \frac{24}{169} \neq \frac{2}{13},$$

donc A et B ne sont plus indépendants. Un peu de réflexion permet de relier ces résultats calculatoires avec la notion intuitive d'indépendance présentée en introduction. Dans le premier cas, la proportion des multiples de trois parmi les pairs est la même que parmi les impairs. Le fait de savoir que la boule tirée est paire ne modifie en rien notre information sur B. En revanche dans le deuxième cas, l'ajout de la treizième boule modifie la proportion des multiples de trois : elle est plus élevée chez les pairs que chez les impairs. Donc le fait de savoir que la boule tirée est paire augmente un peu la probabilité que nous pouvons attribuer à B.

Proposition 5.49. *Si A et B sont indépendants, il en est de même pour les paires d'évènements A et B^c, A^c et B, A^c et B^c.*

Preuve. Par hypothèse, $P(A \cap B) = P(A)P(B)$. En considérant la réunion disjointe $A = (A \cap B) \cup (A \cap B^c)$, nous avons : $P(A \cap B^c) = P(A) - P(A \cap B)$, d'où :

$$P(A \cap B^c) = P(A) - P(A)P(B) = P(A)\big(1 - P(B)\big) = P(A)P(B^c).$$

Donc A et B^c sont indépendants. L'échange des rôles de A et B dans ce raisonnement donne l'indépendance de A^c et B. En réutilisant le premier résultat avec A^c à la place de A, on obtient alors celle de A^c et B^c. □

5.7.2 Indépendance mutuelle

On se propose de généraliser la notion d'indépendance à plus de deux évènements. Examinons d'abord la situation suivante.

Exemple 5.50. Une urne contient quatre jetons : un bleu, un blanc, un rouge et un bleu-blanc-rouge. On en tire un au hasard. Considérons les trois évènements

$$\begin{aligned} A &= \{\text{le jeton tiré contient du bleu}\},\\ B &= \{\text{le jeton tiré contient du blanc}\},\\ C &= \{\text{le jeton tiré contient du rouge}\}. \end{aligned}$$

Il est clair que $P(A) = P(B) = P(C) = 2/4 = 1/2$. D'autre part :

$$P(A \cap B) = P(\text{tricolore}) = \frac{1}{4} = P(A)P(B)$$

et de même $P(B \cap C) = 1/4 = P(B)P(C)$, $P(C \cap A) = 1/4 = P(C)P(A)$. Ainsi les évènements A, B, C sont *deux à deux indépendants*.

D'autre part $P(A \mid B \cap C) = 1$ car $B \cap C = \{\text{tricolore}\}$. Donc la connaissance de la réalisation *simultanée* de B *et* C modifie notre information sur A. La notion d'indépendance deux à deux n'est donc pas suffisante pour traduire l'idée intuitive d'indépendance de plusieurs évènements. Ceci motive la définition suivante.

Définition 5.51. *Trois évènements A, B, C sont dits mutuellement indépendants lorsqu'ils vérifient les quatre conditions :*

$$\begin{aligned} P(A \cap B) &= P(A)P(B),\\ P(B \cap C) &= P(B)P(C),\\ P(C \cap A) &= P(C)P(A),\\ P(A \cap B \cap C) &= P(A)P(B)P(C). \end{aligned}$$

Avec cette définition de l'indépendance des évènements A, B et C on a bien[21] $P(A \mid B) = P(A)$, $P(A \mid B \cap C) = P(A)$, ainsi que toutes les égalités qui s'en déduisent par permutation sur les lettres A, B, C. On peut généraliser cette définition comme suit.

Définition 5.52. *Les n évènements $A_1, \ldots, A_n$ sont dits mutuellement indépendants si pour toute sous-famille $A_{i_1}, \ldots, A_{i_k}$ avec $1 \leqslant i_1 < \ldots < i_k \leqslant n$,*

$$P(A_{i_1} \cap \ldots \cap A_{i_k}) = P(A_{i_1}) \times \cdots \times P(A_{i_k}). \tag{5.28}$$

L'indépendance mutuelle implique évidemment l'indépendance deux à deux et la réciproque est fausse comme le montre l'exemple 5.50. Dans toute la suite, lorsque nous parlerons d'une famille de plusieurs évènements indépendants sans autre précision, nous sous-entendrons systématiquement *mutuellement* indépendants.

21. Lorsque les probabilités conditionnelles existent.

Proposition 5.53. *Si $\{A_1, \ldots, A_n\}$ est une famille de n évènements indépendants, toute famille obtenue en remplaçant certains des A_i par leur complémentaire est encore indépendante.*

Preuve. Supposons la proposition démontrée dans le cas où l'on a remplacé *un seul* A_i par son complémentaire. Le cas général s'en déduit en utilisant cette propriété autant de fois qu'il y a de A_i changés en leur complémentaire. Dans le cas d'un seul A_i remplacé, on ne perd pas de généralité en supposant qu'il s'agit de A_1 (il suffit de changer l'indexation des évènements, ce qui n'affecte pas leur indépendance mutuelle). Il nous reste alors à vérifier (5.28) dans le cas où $i_1 = 1$ avec $A_{i_1}^c$ à la place de A_{i_1} (dans le cas $i_1 > 1$, l'égalité ne fait intervenir que des éléments de la famille initiale et il n'y a donc rien à vérifier). Posons $B = A_{i_2} \cap \cdots \cap A_{i_k}$. L'hypothèse (5.28) appliquée à la famille $A_{i_2} \ldots, A_{i_k}$ nous donne $P(B) = P(A_{i_2}) \times \cdots \times P(A_{i_k})$. La même hypothèse appliquée à $A_{i_1} \ldots, A_{i_k}$ nous donne alors l'indépendance de A_{i_1} et de B. Par la proposition 5.49, on en déduit :

$$P(A_{i_1}^c \cap B) = P(A_{i_1}^c) \times P(B) = P(A_{i_1}^c) \times P(A_{i_2}) \times \cdots \times P(A_{i_k}),$$

ce qui achève la preuve. □

Définition 5.54 (indépendance d'une suite d'évènements)**.** *Une suite infinie d'évènements est dite indépendante si toute sous-suite finie est formée d'évènements mutuellement indépendants.*

Remarque 5.55. Compte-tenu de la proposition 5.53, on voit immédiatement que si $(A_i)_{i\in\mathbb{N}}$ est une suite indépendante d'évènements, toute suite formée en remplaçant certains des A_i (éventuellement tous) par leur complémentaire est encore indépendante.

5.7.3 Épreuves répétées

Considérons une suite d'*épreuves* réalisées dans les mêmes conditions expérimentales, par exemple tirages avec remise dans la même urne, lancers successifs d'un dé, etc. Il est alors raisonnable de supposer que les résultats de tout sous-ensemble fini d'épreuves n'ont aucune influence sur ceux des autres épreuves.

Définition 5.56. *On dit que les épreuves sont indépendantes si toute suite $(A_i)_{i\geqslant 1}$ telle que la réalisation de chaque A_i est déterminée uniquement par le résultat de la i^e épreuve est une suite indépendante d'évènements.*

Exemple 5.57. On réalise une suite d'épreuves indépendantes. Chaque épreuve résulte en un succès avec probabilité $p \in]0, 1[$ ou en un échec avec probabilité $q = 1 - p$. Quelle est la probabilité des évènements suivants ?
a) $A = \{$Au moins un succès au cours des n premières épreuves$\}$.
b) $B = \{$Exactement k succès au cours des n premières épreuves$\}$.
c) $C = \{$Toutes les épreuves donnent un succès$\}$.

Notons pour tout $i \geqslant 1$: $R_i = \{\text{succès à la } i^{\text{e}} \text{ épreuve}\}$, R_i^c est alors l'échec à la i^{e} épreuve.
a) $A = \bigcup_{i=1}^n R_i$, d'où $A^c = \bigcap_{i=1}^n R_i^c$. Les R_i^c étant indépendants, on en déduit :

$$P(A^c) = \prod_{i=1}^{n} P(R_i^c) = (1-p)^n = q^n,$$

d'où $P(A) = 1 - q^n$.
b) Traitons d'abord le cas $0 < k < n$. L'évènement B est la réunion disjointe de tous les évènements du type :

$$B_I = \Big(\bigcap_{i\in I} R_i\Big) \cap \Big(\bigcap_{j\in J} R_j^c\Big),$$

où I est une partie de cardinal k de $[\![1, n]\!]$ et J son complémentaire dans $[\![1, n]\!]$. L'ensemble d'indices I représente un choix possible des k épreuves donnant un succès, les autres épreuves indexées par J donnant alors un échec. En considérant tous les choix possibles de l'ensemble I (il y en a C_n^k), on obtient une partition de B par les B_I. Par indépendance des épreuves, pour tout I on a :

$$P(B_I) = \prod_{i\in I} P(R_i) \times \prod_{j\in J} P(R_j^c) = p^k q^{n-k}.$$

On voit ainsi que $P(B_I)$ ne dépend pas de I. On en déduit :

$$P(B) = \sum_{\substack{I\subset[\![1,n]\!] \\ \operatorname{card} I = k}} P(B_I) = C_n^k p^k q^{n-k}.$$

La vérification de la validité de la formule $P(B) = C_n^k p^k q^{n-k}$ dans les cas $k = 0$ et $k = n$ est laissée au lecteur.
c) Pour $n \geqslant 1$, soit $C_n = \{\text{succès aux } n \text{ premières épreuves}\}$. Clairement C est inclus dans C_n donc $0 \leqslant P(C) \leqslant P(C_n)$. En utilisant l'indépendance des R_i on obtient :

$$P(C_n) = P\Big(\bigcap_{i=1}^{n} R_i\Big) = \prod_{i=1}^{n} P(R_i) = p^n.$$

donc pour tout $n \geqslant 1$, $0 \leqslant P(C) \leqslant p^n$. En faisant tendre n vers $+\infty$, on en déduit $P(C) = 0$.

5.7.4 L'algorithme du rejet

Nous allons voir maintenant comment utiliser des épreuves répétées pour « fabriquer » des probabilités conditionnelles. La méthode exposée ci-dessous est le coeur de l'algorithme du rejet utilisé pour la simulation de variables ou de vecteurs aléatoires. Pour en comprendre le principe, il n'est pas nécessaire de connaître en détail ces notions qui seront étudiées dans les chapitres suivants.

Considérons une suite d'épreuves répétées où la i^{e} épreuve consiste à générer un point aléatoire M_i du plan [22]. Formellement, on a une suite $(M_i)_{i\geqslant 1}$ de points aléatoires telle que pour tout borélien A, les évènements $\{M_i \in A\} = \{\omega \in \Omega;\ M_i(\omega) \in A\}$ sont indépendants et de même probabilité ne dépendant que de A. On fixe un borélien B de $\mathbb{R}^2$ tel que $P(M_1 \in B) > 0$ et on définit l'indice $T(\omega)$ comme le premier indice i tel que $M_i(\omega)$ soit dans B, voir figure 5.7. Si $M_i(\omega)$ n'appartient à B pour aucun $i \in \mathbb{N}^*$, on pose $T(\omega) = +\infty$. On note M_T le point aléatoire ainsi obtenu (c'est le premier M_i à tomber dans B, mais attention son numéro est lui même aléatoire et peut changer avec ω), en prenant par exemple $M_T(\omega) = (0, 0)$ dans le cas particulier où $T(\omega) = +\infty$. On se propose de calculer $P(M_T \in A)$ pour A borélien quelconque de $\mathbb{R}^2$.

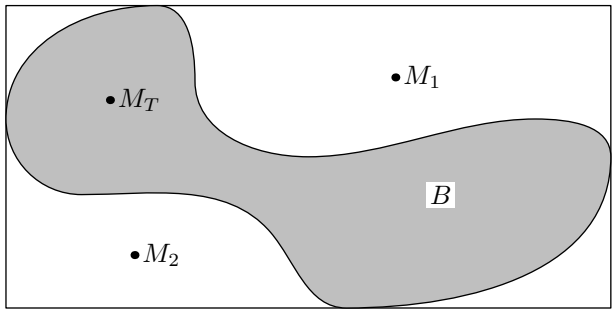


FIGURE 5.7 – Simulation par rejet, ici $T(\omega) = 3$

Nous supposons implicitement ici que les écritures $\{M_i \in A\}$, $\{M_T \in A\}$, $\{T = k\}$ représentent effectivement des évènements [23].

Comme les évènements $\{T = k\}$, $k \in \overline{\mathbb{N}}^*$ réalisent une partition dénombrable de Ω, on a la décomposition en union dénombrable disjointe :

$$\forall A \in \mathrm{Bor}(\mathbb{R}^2), \quad \{M_T \in A\} = \bigcup_{k\in\overline{\mathbb{N}}^*} \{M_T \in A \text{ et } T = k\}. \tag{5.29}$$

Pour tout $k \in \mathbb{N}^*$, l'évènement $\{M_T \in A \text{ et } T = k\}$ peut s'écrire comme suit :

$$\{M_T \in A \text{ et } T = k\} = \{M_k \in A \text{ et } T = k\} = \underbrace{\bigcap_{i<k} \{M_i \notin B\} \cap \{M_k \in B\}}_{\{T=k\}} \cap \{M_k \in A\}$$

Ceci nous permet d'exploiter l'indépendance des épreuves répétées pour obtenir :

$$P(M_T \in A \text{ et } T = k) = (1 - P(M_1 \in B))^{k-1} P(M_k \in A \cap B).$$

22. Tout ce qui suit est valable aussi avec des points aléatoires de $\mathbb{R}^d$ au lieu de $\mathbb{R}^2$.

23. On peut justifier ces suppositions avec la notion de mesurabilité qui sera vue lors de l'étude des variables aléatoires (section 6.1) ou des vecteurs aléatoires (sous-section 8.1.1).

En notant $p = P(M_1 \in B)$, $q = 1 - p$ et en appliquant cette formule avec $A = \mathbb{R}^2$, on obtient :

$$P(M_T \in \mathbb{R}^2 \text{ et } T = k) = P(T = k) = q^{k-1}p.$$

En rappelant que p est strictement positif, d'où $0 < q < 1$ et en sommant pour tous les $k \in \mathbb{N}^*$, il vient :

$$P(T \in \mathbb{N}^*) = \sum_{k\in\mathbb{N}^*} P(T = k) = \sum_{k\in\mathbb{N}^*} q^{k-1}p = \frac{p}{1-q} = 1.$$

Par conséquent, $P(T = +\infty)$ est nulle, tout comme $P(M_T \in A \text{ et } T = +\infty)$ qu'elle majore. Donc pour calculer $P(M_T \in A)$ par σ-additivité à partir de la décomposition (5.29), on ne perd rien en laissant tomber $P(M_T \in A \text{ et } T = +\infty)$. Ainsi :

$$\begin{aligned} P(M_T \in A) = \sum_{k\in\mathbb{N}^*} P(M_T \in A \text{ et } T = k) &= \sum_{k\in\mathbb{N}^*} q^{k-1} P(M_1 \in A \cap B) \\ &= \frac{P(M_1 \in A \cap B)}{1-q} \\ &= \frac{P(M_1 \in A \cap B)}{P(M_1 \in B)}. \end{aligned}$$

Finalement, pour tout ensemble borélien A du plan $\mathbb{R}^2$:

$$P(M_T \in A) = P(M_1 \in A \mid M_1 \in B).$$

Voici une première application. Supposons que les M_i soient choisis au hasard suivant la loi uniforme sur un rectangle C (c'est très facile à simuler informatiquement) contenant B, comme celui de la figure 5.7, autrement dit que pour tout entier i,

$$P(M_i \in A) = \frac{\lambda_2(A \cap C)}{\lambda_2(C)}.$$

Alors pour tout borélien A,

$$P(M_T \in A) = \frac{\dfrac{\lambda_2(A \cap B \cap C)}{\lambda_2(C)}}{\dfrac{\lambda_2(B \cap C)}{\lambda_2(C)}} = \frac{\lambda_2(A \cap B)}{\lambda_2(B)},$$

ce qui montre que le point aléatoire M_T suit la loi uniforme sur B.

5.8 Exercices

Ex. 5.1 Sur un espace probabilisé $(\Omega, \mathcal{F}, P)$, on considère une suite $(A_n)_{n\geqslant 1}$ d'évènements de probabilité 1. Que peut-on dire de $P\big(\bigcap_{n\geqslant 1} A_n\big)$?

Indication p. 428

Ex. 5.2 *Conditionnements multiples*

Soit $(\Omega, \mathcal{F}, P)$ un espace probabilisé et $H \in \mathcal{F}$ un évènement tel que $P(H) \neq 0$. On note P_H la fonction d'ensembles définie par

$$\forall A \in \mathcal{F}, \quad P_H(A) := P(A \mid H).$$

On sait que P_H est une nouvelle probabilité sur $(\Omega, \mathcal{F})$, on peut donc l'utiliser pour construire de nouvelles probabilités conditionnelles. On définit ainsi :

$$P(A \mid H_2 \mid H_1) := P_{H_1}(A \mid H_2).$$

1) Quelle condition doivent vérifier H_1 et H_2 pour que cette définition soit légitime ?

2) Exprimer $P(A \mid H_2 \mid H_1)$ sous la forme d'une probabilité conditionnelle relative à un seul évènement.

3) En déduire que $P(A \mid H_2 \mid H_1) = P(A \mid H_1 \mid H_2)$.

4) Généraliser la question 2) au cas de n évènements $H_1, \ldots, H_n$ et justifier à l'aide de ces nouvelles notations l'appellation « règle des conditionnements successifs »pour la proposition 5.36.

Ex. 5.3 *Une inégalité injustement méconnue*

Sur l'espace probabilisé $(\Omega, \mathcal{F}, P)$, on note A un évènement quelconque et B un évènement tel que $0 < P(B) < 1$.

1) Montrez que

$$|P(A \cap B) - P(A)P(B)| \leqslant \frac{1}{4}\big|P(A \mid B) - P(A \mid B^c)\big|. \tag{5.30}$$

Indication : commencez par exprimer $P(A \mid B) - P(A \mid B^c)$ en fonction des seules probabilités $P(A \cap B)$, $P(A)$, $P(B)$.

2) Que donne l'inégalité (5.30) lorsque $A \subset B$?

3) Dans quels cas (5.30) est-elle une égalité ?

Ex. 5.4 Soit P une mesure de probabilité sur $(\mathbb{R}, \mathrm{Bor}(\mathbb{R}))$ ne prenant que les valeurs 0 et 1. Démontrer que c'est une mesure de Dirac.

Indication p. 429

Ex. 5.5 *La piscine (permutations aléatoires)*

Dans le vestiaire d'une piscine, chaque nageur range ses vêtements sur un cintre. Il le dépose au guichet où un employé équipe le cintre d'un bracelet rouge numéroté et remet au nageur un bracelet jaune portant le même numéro. Ainsi, à la fin de la séance, le nageur peut récupérer ses affaires en échange du bracelet. Avant l'ouverture au public, les bracelets sont rangés sur un tableau à N crochets supportant chacun un bracelet rouge et un jaune de même numéro.

Deux gamins turbulents s'introduisent dans le vestiaire avant l'ouverture. En se battant ils renversent le tableau portant les bracelets. Pour ne pas être découverts, ils les remettent en place en prenant bien soin de placer sur chaque crochet un bracelet jaune et un rouge, mais sans tenir compte des numéros.

A l'ouverture, N nageurs se présentent et devant l'affluence, l'employé remet à chacun son bracelet jaune sans prendre le temps de vérifier les numéros. On se propose de calculer les probabilités des évènements :

$$E_{N,k} = \{\text{exactement } k \text{ nageurs retrouvent leurs affaires}\}.$$

On choisit comme espace probabilisé Ω_N ensemble des toutes les permutations (ou bijections) sur $[\![1, N]\!]$ muni de l'équiprobabilité P_N. On notera :

$$B_i = \{\text{le } i^{\text{e}} \text{ nageur retrouve ses affaires}\}.$$

1) Pour $j \leqslant N$ et $1 \leqslant i_1 < i_2 < \cdots < i_j \leqslant N$, calculer $P_N(B_{i_1} \cap \cdots \cap B_{i_j})$. En déduire :

$$\sum_{1 \leqslant i_1 < i_2 < \cdots < i_j \leqslant N} P_N(B_{i_1} \cap \cdots \cap B_{i_j}) = \frac{1}{j!}.$$

2) Calculer $P_N(E_{N,0})$.

3) On fixe k nageurs retrouvant leurs affaires. Exprimer à l'aide de $P_{N-k}(E_{N-k,0})$ le *nombre* de permutations des $N-k$ autres nageurs telles qu'aucun d'eux ne retrouve ses affaires. En déduire la valeur de $P_N(E_{N,k})$.

4) Pour k fixé, calculer : $p_k = \lim\limits_{N \to \infty} P_N(E_{N,k})$. Montrer que la suite $(p_k)_{k \in \mathbb{N}}$ définit une probabilité P sur $\mathbb{N}$ (il s'agit d'une loi de Poisson).

5) Montrer que :

$$\forall k \in [\![0, N]\!], \qquad |P_N(E_{N,k}) - p_k| < \frac{1}{k!\,(N+1-k)!}.$$

En déduire que

$$\forall n \in [\![0, N]\!], \qquad \sum_{j=0}^{n} |P_N(E_{N,j}) - p_j| \leqslant \frac{\mathrm{e}}{(N+1-n)!}.$$

6) *Application* : après avoir vérifié que : $0{,}999\,4 < \sum_{j=0}^{5} p_j < 0{,}999\,5$, donner pour $N \geqslant 12$ quelconque des valeurs numériques approchées à 10^{-4} près des $P_N(E_{N,j})$ pour $j \in [\![1, 5]\!]$ (ces valeurs ne dépendent pas de N). Même question pour la probabilité qu'au moins 6 nageurs retrouvent leurs affaires.

Solution p. 429

Ex. 5.6 *Pile ou face triphasé*

Trois personnes nommées A, B, C lancent à tour de rôle la même pièce de monnaie (ABCABCABC…). La probabilité d'obtenir pile lors d'un lancer est $p \in]0, 1[$. Le gagnant est le premier qui obtient pile (la partie *s'arrête* alors).

1) On note A_n (resp. B_n, C_n) l'évènement *A gagne la partie lors du n^e lancer* (resp. *B*, *C*). Calculer $P(A_1)$, $P(B_2)$, $P(C_3)$. Les évènements A_1 et B_2 sont-ils indépendants ?

2) En discutant suivant les valeurs de n, calculer $P(A_n)$, $P(B_n)$, $P(C_n)$.

3) Calculer la probabilité de gagner de chacun des trois joueurs.

4) Quelle est la probabilité qu'il y ait un vainqueur ? Conclure.

Solution p. 434

Ex. 5.7 *Triangle aléatoire de pixels*

Le but de cet exercice est de montrer que si on choisit trois points au hasard sur un écran numérique, la probabilité qu'ils forment un vrai triangle est très proche de 1. Sur un écran numérique, chaque point est en fait un pixel et il n'y a qu'un nombre fini de pixels. Il s'agit donc d'un problème de probabilités discrètes.

1) Soit $(\Omega, \mathcal{F}, P)$ un espace probabilisé et $H_1, \dots, H_n$ une partition finie de Ω en évènements de probabilité non nulle. Montrez que

$$\forall E \in \mathcal{F}, \quad P(E) \leqslant \max_{1 \leqslant i \leqslant n} P(E \mid H_i).$$

Dans la suite, nous assimilerons l'écran numérique au carré cartésien :

$$T := [\![0, N[\![^2 = \{z = (x, y) \in \mathbb{N}^2 \ ; \ 0 \leqslant x < N, \ 0 \leqslant y < N\},$$

où N est un entier. Pour toute paire de points z_1 et z_2 distincts de T, on note Δ_{z_1, z_2} l'ensemble des points $z \in T$ qui sont alignés avec z_1 et z_2, cf. la figure 5.8.

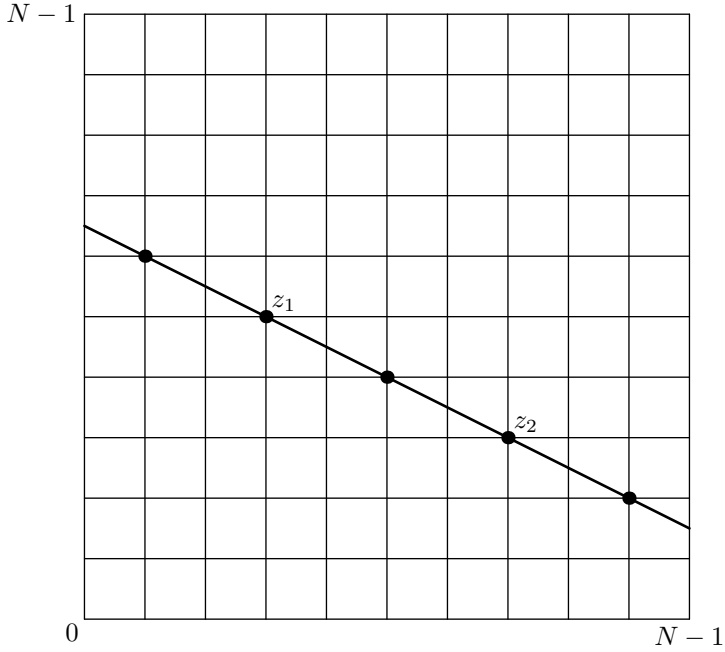


FIGURE 5.8 – Δ_{z_1, z_2}, ici $\operatorname{card} \Delta_{z_1, z_2} = 5$

On choisit sur T trois points aléatoires A, B, C indépendants et de même loi uniforme sur T. Cela signifie que pour toute partie S de T,

$$P(A \in S) = \frac{\operatorname{card} S}{\operatorname{card} T} = \frac{\operatorname{card} S}{N^2},$$

et de même pour B et C et que pour toutes parties S_1, S_2, S_3 de T,

$$P(A \in S_1, B \in S_2, C \in S_3) = P(A \in S_1)P(B \in S_2)P(C \in S_3).$$

On peut traiter tout l'exercice sans expliciter l'espace probabilisé $(\Omega, \mathcal{F}, P)$ modélisant cette expérience.

2) Calculez $P(A = B)$.

3) On définit les évènements :

$$F := \{A, B \text{ et } C \text{ forment un } \textit{vrai} \text{ triangle}\}, \quad E := \{A \neq B \text{ et } C \in \Delta_{A,B}\}.$$

Justifiez la décomposition :

$$F^c = \{A = B\} \cup E.$$

4) Expliquez pourquoi pour toute paire de points distincts z_1 et z_2 de T,

$$P(C \in \Delta_{z_1,z_2}) \leqslant \frac{1}{N}.$$

5) En utilisant *tout* ce qui précède, montrez proprement [24] que

$$P(F) \geqslant 1 - \frac{1}{N} - \frac{1}{N^2}.$$

Ex. 5.8 *Jour de chance*

Un site de jeux propose le jeu suivant. Chaque internaute désireux de jouer s'inscrit en misant un euro et en indiquant son jour de naissance (lundi, mardi, ...). Le jeu commence systématiquement un lundi et une fois par jour, les organisateurs filment et mettent en ligne un seul lancer d'un dé *équilibré*, jusqu'au jour de la première apparition du 6. Le jeu s'arrête alors et les candidats nés ce même jour de la semaine se partagent les mises [25]. Par exemple si la première apparition du 6 a lieu un jeudi, les gagnants sont tous les joueurs nés un jeudi.

1) Calculez la probabilité de gagner pour un joueur né un dimanche.

2) Le jeu est-il équitable (en admettant que les jours de naissance de l'ensemble des joueurs se répartissent uniformément sur les 7 jours de la semaine) ?

Ex. 5.9 *Probabilité de gagner un jeu sur son service au tennis*

On considère le modèle simplifié suivant du jeu de tennis [26] : le joueur A est au service et affronte B. Un « jeu » est constitué d'un certain nombre d'« échanges » et à la fin de chaque « échange », celui des deux joueurs qui a gagné l'échange marque un point. Le joueur qui gagne le « jeu » est le premier à totaliser au moins 4 points *avec* au moins deux points d'avance sur son adversaire. A peut donc gagner le jeu sur

24. Cela va sans dire, mais mieux en le disant.

25. Le site étant financé par la publicité, les organisateurs n'effectuent pas de prélèvement sur les mises.

26. Il n'est pas nécessaire de connaître les règles classiques du tennis pour faire cet exercice. Elles sont rappelées dans l'énoncé sous une forme permettant de simplifier les écritures.

le score de 4 à 0, 4 à 1, 4 à 2, 5 à 3, 6 à 4, ... On suppose que tous les échanges sont indépendants et que lors de chaque échange, la probabilité pour A de gagner le point reste constante et vaut p. Notre objectif est de calculer en fonction de p la probabilité que A remporte le jeu. On introduit les notations d'évènements suivantes.

$$\begin{aligned} G &= \{A \text{ gagne le jeu}\}, \\ A_n &= \{\ A \text{ gagne le } n^{\text{e}} \text{ échange }\}, \\ B_n &= \{\ B \text{ gagne le } n^{\text{e}} \text{ échange }\}, \\ E_{i,j} &= \{\text{Au bout de } i+j \text{ échanges, } A \text{ a } i \text{ points et } B \text{ en a } j\}. \end{aligned}$$

1) Expliquez la décomposition

$$G = E_{4,0} \cup E_{4,1} \cup E_{4,2} \cup \Big(\bigcup_{k=3}^{+\infty} E_{k+2,k} \Big).$$

2) Calculer $P(E_{4,0})$, $P(E_{4,1})$ et $P(E_{4,2})$. *Indication :* on remarquera que $E_{4,1} = A_5 \cap E_{3,1}$ et $E_{4,2} = A_6 \cap E_{3,2}$ et on utilisera l'indépendance des échanges.

3) Expliquer pourquoi pour tout $k \geqslant 3$, on a :

$$E_{k+2,k} = E_{3,3} \cap \Big(\bigcap_{j=4}^{k} C_j \Big) \cap A_{2k+1} \cap A_{2k+2},$$

où l'on a posé $C_j := (A_{2j-1} \cap B_{2j}) \cup (B_{2j-1} \cap A_{2j})$.

4) Calculer $P(E_{3,3})$, puis $r = P(C_j)$. En déduire $P(E_{k+2,k})$.

5) Montrer que les $E_{k+2,k}$ sont deux à deux disjoints et calculer $P\big(\bigcup_{k\geqslant 3} E_{k+2,k} \big)$.

6) En déduire $P(G)$.

7) L'évènement $N = \{$le jeu continue indéfiniment sans vainqueur$\}$ s'écrit :

$$N = E_{3,3} \cap \Big(\bigcap_{j=4}^{+\infty} C_j \Big).$$

Montrer que sa probabilité est nulle.

8) Compte tenu du résultat de la question précédente, pouvez vous donner sans calcul la valeur de $P(G)$ lorsque $p = 1/2$? Utilisez cette réponse pour tester la formule trouvée à la question 6.

Solution p. 436

Ex. 5.10 *Inégalité et ball trap accéléré*

On rappelle que si A est un évènement, la variable aléatoire indicatrice de A est la fonction définie sur Ω par :

$$\mathbf{1}_A(\omega) = \begin{cases} 1 & \text{si } \omega \in A, \\ 0 & \text{sinon.} \end{cases}$$

Autrement dit, $\mathbf{1}_A$ vaut 1 si A est réalisé et 0 sinon.

Soit $(A_i)_{i\geqslant 1}$ une suite d'évènements *indépendants*. On note $p_i = P(A_i)$ et on suppose que :

$$a := \sum_{i=1}^{+\infty} p_i < +\infty.$$

Le but de cet exercice est de prouver l'inégalité

$$\forall n, k \in \mathbb{N}^*, \quad P\left(\sum_{i=1}^{n} \mathbf{1}_{A_i} \geqslant k\right) \leqslant \frac{a^k}{k!}.$$

La dernière question propose une application de cette inégalité [27].

1) Que peut-on dire du cas $k > n$? On suppose dans la suite $k \leqslant n$.

2) On note $B_{n,k} := \{\omega \in \Omega;\ \sum_{i=1}^{n} \mathbf{1}_{A_i}(\omega) \geqslant k\}$. Justifier l'inclusion

$$B_{n,k} \subset \bigcup_{\substack{F\subset[\![1,n]\!]\\ \text{card } F=k}} \bigcap_{i\in F} A_i.$$

3) En déduire que

$$P(B_{n,k}) \leqslant \sum_{\substack{F\subset[\![1,n]\!]\\ \text{card } F=k}} \prod_{i\in F} p_i.$$

4) On note $a_n = \sum_{i=1}^{n} p_i$. Montrer que

$$a_n^k \geqslant k! \sum_{\substack{F\subset[\![1,n]\!]\\ \text{card } F=k}} \prod_{i\in F} p_i.$$

Indication : On remarquera que

$$a_n^k = \sum_{(i_1,\dots,i_k)\in[\![1,n]\!]^k} p_{i_1}\cdots p_{i_k}.$$

5) Conclure.

6) *Application à un problème de tir.* Dans un stand de tir, une cible mobile traverse le champ visuel d'un tireur (une traversée par épreuve). À chaque épreuve, le tireur tire un coup sur la cible. D'une épreuve à la suivante, la vitesse de la cible augmente de 20%. On suppose que pour un tireur donné, la probabilité de toucher la cible est *inversement proportionnelle* à la vitesse de la cible. Elle vaut ainsi $p \in]0,1[$ pour le premier tir, $\frac{5}{6}p$ pour le second (pourquoi ?), etc. Les tirs sont supposés indépendants, le tireur dispose d'autant de cartouches qu'il le souhaite et le défi qu'il doit relever est celui de toucher au moins 20 fois la cible. En utilisant le résultat démontré ci-dessus, majorer sa probabilité de réussir (indépendamment de p).

Solution p. 441

27. Cette inégalité peut se trouver dans le livre *Probability in Banach spaces* de M. Ledoux et M. Talagrand [10]. L'usage qui y en est fait est assez différent de celui proposé ici à la dernière question.

Chapitre 6

Variables aléatoires

6.1 Introduction

DANS de nombreux jeux, on fait intervenir le hasard en observant la somme des points marqués par deux dés. Considérons le jet d'un dé bleu et d'un dé rouge et notons S la somme des points obtenus. On modélise cette expérience en prenant l'équiprobabilité sur $\Omega = [\![1,6]\!]^2$. Un évènement élémentaire ω est ici un couple (b,r) où b désigne le résultat du dé bleu et r celui du rouge et $S(\omega) = b + r$. Il est commode de décrire la situation par un tableau à 36 cases en écrivant la valeur de $S(\omega)$ dans la case représentant $\omega = (b,r)$ à l'intersection de la ligne b et de la colonne r.

	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

FIGURE 6.1 – Somme des points de deux dés

On a ainsi défini une application S de Ω dans l'ensemble des sommes de points possibles : $[\![2,12]\!]$. On dit que S est une *variable aléatoire* sur Ω. En fait, l'observation qui nous intéresse dans cette expérience, ce n'est pas ω, mais seulement $S(\omega)$. Ce que l'on aimerait connaître, c'est la probabilité que la somme des points prenne une valeur donnée, soit $P(S = k)$ pour $k \in [\![2,12]\!]$. Ici la notation « $P(S = k)$ » est un abus d'écriture commode pour désigner $P\big(\{\omega \in \Omega \ ; \ S(\omega) = k\}\big)$. En utilisant l'équiprobabilité sur Ω et la figure 6.1, nous obtenons le tableau 6.1.

k	2	3	4	5	6	7	8	9	10	11	12
$P(S=k)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

TABLE 6.1 – Probabilités des valeurs de la somme des points

Cela revient à considérer un nouvel ensemble d'évènements élémentaires :

$$\Omega' = S(\Omega) = \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$$

et à le munir de la probabilité P_S définie par le tableau 6.1, plus précisément :

$$P_S := \sum_{k\in\Omega'} P(S=k)\delta_k. \tag{6.1}$$

Cette nouvelle probabilité s'appelle *loi de la variable aléatoire* S. En d'autre termes, nous avons réalisé *via* S un *transfert* de l'espace probabilisé $\big(\Omega, \mathcal{P}(\Omega), P\big)$ sur l'espace probabilisé $(\Omega', \mathcal{P}(\Omega'), P_S)$. Si $B \in \mathcal{P}(\Omega')$, en notant $\{S \in B\} := \{\omega \in \Omega;\ S(\omega) \in B\}$, on a ainsi puisque B est l'union finie de ses singletons,

$$P(S \in B) = \sum_{k\in B} P(S=k) = \sum_{k\in\Omega'} P(S=k)\delta_k(B) = P_S(B). \tag{6.2}$$

Remarquons maintenant que l'on peut facilement agrandir l'ensemble d'arrivée de S en remplaçant Ω' par $\Omega'' = \mathbb{R}$. On peut munir $\mathbb{R}$ au choix de la tribu $\mathcal{P}(\mathbb{R})$ ou $\mathrm{Bor}(\mathbb{R})$. En définissant encore P_S par (6.1), le calcul (6.2) s'adapte facilement pour tout $B \in \mathcal{P}(\mathbb{R})$, donc *a fortiori* pour tout B borélien :

$$P(S \in B) = P(S \in B \cap \Omega') = \sum_{k\in B\cap\Omega'} P(S=k) = \sum_{k\in\Omega'} P(S=k)\delta_k(B) = P_S(B). \tag{6.3}$$

Il y a de bonnes raisons à cet agrandissement de l'ensemble d'arrivée, ne serait ce que pour pouvoir faire des opérations sur les variables aléatoires. Imaginons par exemple que l'on jette n fois la paire de dés et que l'on s'intéresse à la moyenne arithmétique M_n des sommes obtenues. Il ne serait guère commode de travailler avec un espace d'arrivée $\Omega'_n := M_n(\Omega'^n)$, surtout si l'on s'intéresse au comportement de M_n pour n tendant vers l'infini. Nous prendrons donc désormais comme ensemble d'arrivée $\mathbb{R}$ pour les variables aléatoires que nous étudierons.

Si $X(\omega)$ est une grandeur physique (masse, température, pression, longueur, etc.) mesurée à partir du résultat ω d'une expérience, il n'est en général pas possible de la déterminer avec une précision absolue. Tout ce que l'on peut dire est que X appartient à un certain intervalle, dont la longueur dépend de la précision de l'instrument de mesure utilisé. Les quantités pertinentes pour identifier la *loi* de X sont alors les $P(X \in I)$ pour I intervalle [1] de $\mathbb{R}$, plutôt que les $P(X = x)$ qui pourraient être nulles

1. Ceci est en accord avec la remarque faite p. 176 à propos de la caractérisation des probabilités sur $\big(\mathbb{R}, \mathrm{Bor}(\mathbb{R})\big)$.

pour tout x réel. Nous définirons la loi de X en posant $P_X(B) = P(X \in B)$ pour B borélien quelconque de $\mathbb{R}$, *sous réserve que cela ait un sens*. Voyons cela de plus près.

Soit X une application $\Omega \to \mathbb{R}$. Pour tout $B \subset \mathbb{R}$, notons

$$\{X \in B\} := \{\omega \in \Omega;\ X(\omega) \in B\} =: X^{-1}(B).$$

Cette écriture « X^{-1} » ne suppose aucunement la bijectivité de X. Il s'agit seulement d'une notation commode pour « l'ensemble des antécédents des éléments de B par l'application X ». On aimerait pouvoir transporter par X la probabilité P, mesure définie sur $(\Omega, \mathcal{F})$ en une probabilité P_X, définie sur $\big(\mathbb{R}, \mathrm{Bor}(\mathbb{R})\big)$ en posant :

$$\forall B \in \mathrm{Bor}(\mathbb{R}), \quad P_X(B) := P\big(X^{-1}(B)\big) = P(X \in B).$$

Pour que cette écriture ait un sens, encore faut-il que $X^{-1}(B)$ soit un élément de la tribu $\mathcal{F}$ sur laquelle est définie P. Nous supposerons donc que X vérifie la condition :

$$\forall B \in \mathrm{Bor}(\mathbb{R}), \quad X^{-1}(B) \in \mathcal{F}. \tag{6.4}$$

On dit alors que X est *mesurable* pour les tribus $\mathcal{F}$ et $\mathrm{Bor}(\mathbb{R})$. Nous réserverons le nom de *variable aléatoire* aux applications $X : \Omega \to \mathbb{R}$ vérifiant (6.4). La mesurabilité est d'ailleurs une notion plus générale définie comme suit.

Définition 6.1 (mesurabilité)**.** *Soit $h : \Omega_1 \to \Omega_2$, où Ω_1 et Ω_2 sont munis respectivement des tribus $\mathcal{F}_1$ et $\mathcal{F}_2$. On dit que h est* mesurable [2] *$\mathcal{F}_1$ - $\mathcal{F}_2$ si pour tout $B \in \mathcal{F}_2$, $h^{-1}(B) \in \mathcal{F}_1$.*

Définition 6.2 (application borélienne)**.** *Soit $h : \mathbb{R}^{d_1} \to \mathbb{R}^{d_2}$ et $\mathcal{F}_i = \mathrm{Bor}(\mathbb{R}^{d_i})$ pour $i = 1, 2$. On dit que h est une application* borélienne *si elle est mesurable $\mathcal{F}_1$ - $\mathcal{F}_2$.*

On démontre en théorie de la mesure le résultat suivant.

Proposition 6.3. *Soient Ω_1 et Ω_2 deux ensembles munis respectivement des tribus $\mathcal{F}_1$ et $\mathcal{F}_2$ et $\mathcal{C}$ une famille de parties de Ω_2 engendrant $\mathcal{F}_2$ ($\sigma(\mathcal{C}) = \mathcal{F}_2$). L'application $h : \Omega_1 \to \Omega_2$ est mesurable $\mathcal{F}_1$ - $\mathcal{F}_2$ si pour tout $B \in \mathcal{C}$, $h^{-1}(B) \in \mathcal{F}_1$.*

En particulier en prenant $\Omega_2 = \mathbb{R}$, $\mathcal{F}_2 = \mathrm{Bor}(\mathbb{R})$ et $\mathcal{C}$ la famille des intervalles $]a, b]$, on voit que pour que $X : \Omega \to \mathbb{R}$ soit une variable aléatoire, il suffit que

$$\forall a, b \in \mathbb{R}, \quad X^{-1}(]a, b]) \in \mathcal{F}.$$

La mesurabilité $\mathcal{F}$ - $\mathrm{Bor}(\mathbb{R})$ est conservée par toutes les opérations usuelles de l'algèbre (somme, produit, combinaisons linéaires, composition,...) et de l'analyse [3], pourvu que les familles d'applications concernées soient au plus dénombrables (inf, sup, lim inf, lim sup, limite si elle existe d'une suite de fonctions, série de fonctions). Bref, il s'agit d'une notion très riche. Tellement riche en fait, que nous ne rencontrerons jamais dans cet ouvrage d'application $X : \Omega \to \mathbb{R}$ qui ne soit pas une variable aléatoire. Ceci explique que dans les ouvrages de probabilités élémentaires, on appelle variable aléatoire n'importe quelle application $\Omega \to \mathbb{R}$.

2. Ce langage est trompeur : la mesurabilité ne fait intervenir aucune mesure, elle concerne seulement h et les tribus.

3. Pour les énoncés précis, voir [15, Chap. 2].

6.2 Généralités

6.2.1 Variables aléatoires réelles

Avant de formaliser la définition d'une variable aléatoire, commençons par donner deux propriétés utiles des inverses ensemblistes.

Proposition 6.4. *Soit $h : \Omega_1 \to \Omega_2$ une application. Pour tout $B \subset \Omega_2$, notons $h^{-1}(B) := \{\omega \in \Omega_1;\ h(\omega) \in B\}$. L'inverse ensembliste h^{-1} ainsi défini commute avec les unions et les intersections quelconques. Autrement dit, si $(B_i)_{i\in I}$ est une famille quelconque de parties de Ω_2,*

$$h^{-1}\Big(\bigcup_{i\in I} B_i\Big) = \bigcup_{i\in I} h^{-1}(B_i), \tag{6.5}$$

$$h^{-1}\Big(\bigcap_{i\in I} B_i\Big) = \bigcap_{i\in I} h^{-1}(B_i). \tag{6.6}$$

L'inverse ensembliste commute aussi avec le passage au complémentaire au sens suivant :

$$\forall B \subset \Omega_2, \quad h^{-1}(\Omega_2\backslash B) = \Omega_1\backslash h^{-1}(B). \tag{6.7}$$

Preuve. L'égalité d'ensembles (6.5) se vérifie par la chaîne d'équivalences logiques suivantes qui montre que l'appartenance au premier membre de (6.5) équivaut à l'appartenance à son deuxième membre :

$$\begin{aligned}
\omega \in h^{-1}\Big(\bigcup_{i\in I} B_i\Big) &\Leftrightarrow h(\omega) \in \bigcup_{i\in I} B_i \\
&\Leftrightarrow \exists i \in I\ ;\ h(\omega) \in B_i \\
&\Leftrightarrow \exists i \in I\ ;\ \omega \in h^{-1}(B_i) \\
&\Leftrightarrow \omega \in \bigcup_{i\in I} h^{-1}(B_i).
\end{aligned}$$

On procède de même pour vérifier (6.6) :

$$\begin{aligned}
\omega \in h^{-1}\Big(\bigcap_{i\in I} B_i\Big) &\Leftrightarrow h(\omega) \in \bigcap_{i\in I} B_i \\
&\Leftrightarrow \forall i \in I,\ h(\omega) \in B_i \\
&\Leftrightarrow \forall i \in I,\ \omega \in h^{-1}(B_i) \\
&\Leftrightarrow \omega \in \bigcap_{i\in I} h^{-1}(B_i).
\end{aligned}$$

Voici la vérification de (6.7) :

$$\begin{aligned}
h^{-1}(\Omega_2\backslash B) &= \{\omega \in \Omega_1\ ;\ h(\omega) \in (\Omega_2\backslash B)\} \\
&= \{\omega \in \Omega_1\ ;\ h(\omega) \notin B\} \\
&= \{\omega \in \Omega_1\ ;\ \omega \notin h^{-1}(B)\} \\
&= \Omega_1\backslash h^{-1}(B).
\end{aligned}$$

□

Définition 6.5 (variable aléatoire réelle)**.** *Soit* $(\Omega, \mathcal{F}, P)$ *un espace probabilisé. On appelle* variable aléatoire réelle *sur* $(\Omega, \mathcal{F})$*, ou plus simplement* variable aléatoire*, toute application* X *:*

$$X : \Omega \longrightarrow \mathbb{R} \quad \omega \mapsto X(\omega),$$

mesurable $\mathcal{F}$ - $\mathrm{Bor}(\mathbb{R})$*, c'est-à-dire vérifiant :*

$$\forall B \in \mathrm{Bor}(\mathbb{R}), \quad X^{-1}(B) \in \mathcal{F}.$$

En raison de la proposition 6.3, il suffit que la condition ci-dessus soit vérifiée pour $B =]a, b]$ avec a et b réels quelconques, pour que X soit une variable aléatoire.

Remarque 6.6. Il importe de noter que la mesure de probabilité P sur $(\Omega, \mathcal{F})$ ne joue *aucun rôle* dans la définition de la notion de variable aléatoire. C'est pour cela que nous parlons de « variable aléatoire sur $(\Omega, \mathcal{F})$ » plutôt que sur $(\Omega, \mathcal{F}, P)$.

Définition 6.7 (variable aléatoire discrète)**.** *On appelle* variable aléatoire discrète *sur* $(\Omega, \mathcal{F})$*, toute application* $X : \Omega \to \mathbb{R}$ *vérifiant les deux conditions suivantes.*

(i) L'ensemble des images $X(\Omega) = \{X(\omega),\ \omega \in \Omega\}$ *est une partie au plus dénombrable de* $\mathbb{R}$*. On peut donc numéroter ses éléments par des indices entiers*[4]

$$X(\Omega) = \{x_0, x_1, \ldots, x_k, \ldots\}.$$

(ii) X *est mesurable* $\mathcal{F}$ - $\mathcal{P}(\mathbb{R})$*, ce qui équivaut ici à*

$$\forall x \in X(\Omega), \quad X^{-1}(\{x\}) \in \mathcal{F}. \tag{6.8}$$

Remarquons que si X est mesurable $\mathcal{F}$ - $\mathcal{P}(\mathbb{R})$, elle est *a fortiori* mesurable $\mathcal{F}$ - $\mathrm{Bor}(\mathbb{R})$ puisque la condition $X^{-1}(B) \in \mathcal{F}$ doit être satisfaite dans ce cas seulement pour les B appartenant à la tribu $\mathrm{Bor}(\mathbb{R})$ qui est une sous-famille de $\mathcal{P}(\mathbb{R})$. Par conséquent, toute variable aléatoire discrète est aussi une variable aléatoire réelle.

L'équivalence, pour $X(\Omega)$ au plus dénombrable, entre la mesurabilité $\mathcal{F}$ - $\mathcal{P}(\mathbb{R})$ et la condition (6.8), se justifie comme suit. D'abord il est clair que la mesurabilité $\mathcal{F}$ - $\mathcal{P}(\mathbb{R})$ implique (6.8). Pour la réciproque, on suppose (6.8) vérifiée, on prend $B \in \mathcal{P}(\mathbb{R})$ quelconque et on montre qu'alors $X^{-1}(B) \in \mathcal{F}$. Notons $B' := B \cap X(\Omega)$ et $B'' := B \cap \big(\mathbb{R} \backslash X(\Omega)\big)$. Remarquons que $B' \subset X(\Omega)$ est au plus dénombrable et que $X^{-1}(B'') = \varnothing$. Il suffit alors d'écrire en utilisant (6.5) :

$$\begin{aligned} X^{-1}(B) = X^{-1}(B' \cup B'') &= X^{-1}(B') \cup X^{-1}(B'') \\ &= X^{-1}(B') \\ &= X^{-1}\Big(\bigcup_{x \in B'} \{x\}\Big) \\ &= \bigcup_{x \in B'} X^{-1}(\{x\}), \end{aligned} \tag{6.9}$$

pour faire apparaître $X^{-1}(B)$ comme union au plus dénombrable d'éléments de la tribu $\mathcal{F}$, ce qui implique son appartenance à $\mathcal{F}$.

4. Pour tous les exemples classiques que nous rencontrerons, il est possible de les numéroter de manière croissante : $x_0 < x_1 < x_2 \ldots$. Mais ce n'est pas toujours le cas, car $X(\Omega)$ peut être par exemple, l'ensemble des décimaux (ou des rationnels) de $[0, 1]$, cf. l'exercice 6.17.

6.2.2 Loi d'une variable aléatoire

Une variable aléatoire X permet de transporter la probabilité P définie sur $(\Omega, \mathcal{F})$ en une probabilité P_X définie sur $\big(\mathbb{R}, \mathrm{Bor}(\mathbb{R})\big)$.

Proposition 6.8. *Soit X une variable aléatoire sur $(\Omega, \mathcal{F})$ et P une probabilité sur $(\Omega, \mathcal{F})$. La fonction d'ensembles $P_X = P \circ X^{-1}$ définie sur* $\mathrm{Bor}(\mathbb{R})$ *par*

$$\forall B \in \mathrm{Bor}(\mathbb{R}), \quad P_X(B) := P\big(X^{-1}(B)\big) = P(X \in B) \tag{6.10}$$

est une probabilité sur $\big(\mathbb{R}, \mathrm{Bor}(\mathbb{R})\big)$.

Preuve. D'abord P_X est bien définie comme application $\mathrm{Bor}(\mathbb{R}) \to [0,1]$ en raison de la mesurabilité $\mathcal{F}$ - $\mathrm{Bor}(\mathbb{R})$ de la variable aléatoire X. Il est clair que $P_X(\mathbb{R}) = P(\Omega) = 1$. Montrons la σ-additivité de P_X. Soit $(B_k)_{k\geqslant 1}$ une suite quelconque de boréliens de $\mathbb{R}$ deux à deux disjoints. Par (6.6), les $X^{-1}(B_k)$ sont des éléments deux à deux disjoints de la tribu d'évènements $\mathcal{F}$. En combinant (6.5) avec la σ-additivité de la probabilité P, on en déduit :

$$\begin{aligned} P_X\Big(\bigcup_{k\geqslant 1} B_k\Big) = P\Big(X^{-1}\Big(\bigcup_{k\geqslant 1} B_k\Big)\Big) = P\Big(\bigcup_{k\geqslant 1} X^{-1}(B_k)\Big) &= \sum_{k\geqslant 1} P\big(X^{-1}(B_k)\big) \\ &= \sum_{k\geqslant 1} P_X(B_k). \end{aligned}$$

La fonction d'ensembles P_X est donc σ-additive, c'est une probabilité sur $\big(\mathbb{R}, \mathrm{Bor}(\mathbb{R})\big)$. □

Définition 6.9 (loi d'une variable aléatoire)**.** *Soient $(\Omega, \mathcal{F}, P)$ un espace probabilisé et X une variable aléatoire sur $(\Omega, \mathcal{F})$. On appelle* loi de X sous P*, ou plus simplement loi de X, la probabilité P_X sur $\big(\mathbb{R}, \mathrm{Bor}(\mathbb{R})\big)$ définie par (6.10).*

Si μ est une mesure de probabilité sur $\big(\mathbb{R}, \mathrm{Bor}(\mathbb{R})\big)$, on dit que X « suit » la loi μ si $P_X = P \circ X^{-1} = \mu$, autrement dit, si la loi de X sous P est la mesure μ.

Remarque 6.10. Dans les problèmes usuels de probabilités, on travaille souvent avec un seul $(\Omega, \mathcal{F}, P)$ et on se contente alors de l'appelation *loi de X*. Il n'en va pas de même en statistique où l'on met généralement en concurrence *plusieurs* modèles $(\Omega, \mathcal{F}, P_\theta)$, où θ est un paramètre inconnu et où on se propose de choisir un de ces modèles au vu des valeurs $X(\omega)$ observées. C'est là que l'appelation loi de X *sous* P_θ s'impose. Pour donner un exemple simple, considérons le problème du sondage d'un échantillon de 500 personnes avant le second tour d'une élection présidentielle opposant le candidat A au candidat B. Ici θ est la proportion *inconnue* d'électeurs votant A dans la population totale. Si X est le nombre de personnes interrogées favorables à A, la loi de X sous P_θ est la loi binomiale [5] $\mathrm{Bin}(500, \theta)$.

5. En fait c'est une loi hypergéométrique (tirages sans remise), mais en raison du théorème 6.37, on peut la remplacer en pratique par une binomiale.

Une autre situation où il est naturel de considérer plusieurs lois pour une même variable aléatoire est celle du *conditionnement*. Rappelons que si $(\Omega, \mathcal{F}, P)$ est un espace probabilisé et $H \in \mathcal{F}$ un évènement tel que $P(H) > 0$, on peut définir sur $\mathcal{F}$ une nouvelle *mesure* de probabilité $P_H = P(\,.\mid H)$ par

$$\forall A \in \mathcal{F}, \quad P_H(A) := P(A \mid H) = \frac{P(A \cap H)}{P(H)}.$$

Définition 6.11 (loi conditionnelle)**.** *Soient $(\Omega, \mathcal{F}, P)$ un espace probabilisé, $H \in \mathcal{F}$ tel que $P(H) > 0$, X une variable aléatoire sur $(\Omega, \mathcal{F})$. On appelle* loi conditionnelle de X sachant H*, la loi de X sous P_H. En la notant $P_{X|H}$, on a donc*

$$\forall B \in \mathrm{Bor}(\mathbb{R}), \quad P_{X|H}(B) = P_H\big(X^{-1}(B)\big) = P(X \in B \mid H).$$

Il importe de ne pas se laisser induire en erreur par la notation $P_{X|H}$, elle ne concerne pas une nouvelle variable aléatoire « $X \mid H$ » mais bien toujours *la même* variable aléatoire X. Ce qui a changé, c'est la probabilité dont on munit $(\Omega, \mathcal{F})$ et sous laquelle on considère la loi de X.

Dans le cas d'une variable aléatoire discrète X, il est facile de donner une formule explicite pour la loi P_X.

Proposition 6.12. *Soient $(\Omega, \mathcal{F}, P)$ un espace probabilisé et X une variable aléatoire* discrète *sur $(\Omega, \mathcal{F})$. La loi de X sous P est la probabilité*

$$P_X = \sum_{x \in X(\Omega)} P(X = x)\delta_x, \tag{6.11}$$

que l'on peut considérer comme probabilité sur $\big(X(\Omega), \mathcal{P}(X(\Omega))\big)$, ou sur $\big(\mathbb{R}, \mathrm{Bor}(\mathbb{R})\big)$ ou sur $\big(\mathbb{R}, \mathcal{P}(\mathbb{R})\big)$.

Preuve. Vérifions l'égalité de mesures (6.11) sur $\big(\mathbb{R}, \mathcal{P}(\mathbb{R})\big)$, le résultat analogue pour $\big(\mathbb{R}, \mathrm{Bor}(\mathbb{R})\big)$ et $\big(X(\Omega), \mathcal{P}(X(\Omega))\big)$ s'en déduisant immédiatement par restriction. Par une adaptation immédiate de la preuve de la proposition 6.8, P_X est aussi une probabilité[6] sur $\big(\mathbb{R}, \mathcal{P}(\mathbb{R})\big)$. En utilisant la décomposition (6.9) on a pour tout $B \subset \mathbb{R}$,

$$P_X(B) = P\Big(\bigcup_{x \in B \cap X(\Omega)} X^{-1}(\{x\})\Big) = \sum_{x \in B \cap X(\Omega)} P\big(X^{-1}(\{x\})\big) = \sum_{x \in X(\Omega)} P(X = x)\delta_x(B).$$

L'égalité de mesures (6.11) sur $\big(\mathbb{R}, \mathcal{P}(\mathbb{R})\big)$ est ainsi vérifiée. □

Définition 6.13 (loi discrète sur $\mathbb{R}$)**.** *On appelle loi discrète sur $\mathbb{R}$, toute mesure ponctuelle μ sur $(\mathbb{R}, \mathcal{P}(\mathbb{R}))$ qui est aussi une probabilité. Une telle loi admet donc une représentation sous la forme*

$$\mu = \sum_{i \in I} p_i \delta_{x_i},$$

où I est un ensemble d'indices au plus dénombrable, $(x_i)_{i \in I}$ une famille de nombres réels et $(p_i)_{i \in I}$ une famille sommable de réels positifs, de somme $\sum_{i \in I} p_i = 1$.

6. Rappelons qu'une v.a. discrète X sur $(\Omega, \mathcal{F})$ est mesurable $\mathcal{F}$-$\mathcal{P}(\mathbb{R})$ et pas seulement $\mathcal{F}$-$\mathrm{Bor}(\mathbb{R})$.

Il est clair que par restriction, μ est aussi une probabilité sur $(\mathbb{R}, \mathrm{Bor}(\mathbb{R}))$.

Remarque 6.14. Deux variables aléatoires peuvent *avoir même loi sans être égales*. Par exemple considérons le jet de deux dés, l'un bleu et l'autre rouge. Notons X le nombre de points indiqué par le dé bleu et Y celui du rouge. Les variables aléatoires X et Y sont définies sur le même espace probabilisé $\Omega = \{1,2,3,4,5,6\}^2$ muni de l'équiprobabilité. On a $X(\Omega) = Y(\Omega) = [\![1,6]\!]$ et :

$$\forall k \in [\![1,6]\!], \quad P(X=k) = \frac{1}{6}, \quad P(Y=k) = \frac{1}{6}.$$

Donc X et Y ont même loi : $P_X = P_Y = \sum_{k=1}^{6} \frac{1}{6}\delta_k$. Pour autant, on n'a pas l'égalité des variables aléatoires X et Y qui signifierait $X(\omega) = Y(\omega)$ *pour tout* $\omega \in \Omega$ (égalité de deux applications). Autrement dit, en lançant deux dés on obtiendrait *à coup sûr* un double. En revanche nous pouvons considérer *l'évènement* $\{X = Y\}$ dont la réalisation n'est pas certaine et calculer sa probabilité :

$$P(X=Y) = P\Big(\bigcup_{k=1}^{6} \{(X,Y) = (k,k)\}\Big) = \frac{6}{36} = \frac{1}{6}.$$

On en déduit : $P(X \neq Y) = 5/6$.

Remarque 6.15. Deux variables aléatoires peuvent avoir même loi en étant définies sur des espaces probabilisés différents $(\Omega, \mathcal{F}, P)$ et $(\Omega', \mathcal{F}', P')$. Prenons par exemple pour X les points du dé bleu comme ci-dessus et posons $Z = 0$ si X est pair, $Z = 1$ si X est impair. La variable aléatoire Z est définie sur $\Omega = \{1,2,3,4,5,6\}^2$ muni de la tribu $\mathcal{P}(\Omega)$ et de l'équiprobabilité P sur Ω. Sa loi est $P_Z = \frac{1}{2}(\delta_0 + \delta_1)$. Prenons maintenant $\Omega' = \{-1, 1\}$, muni de la tribu $\mathcal{P}(\Omega')$ et de l'équiprobabilité P' sur Ω' et posons pour $\omega' \in \Omega'$, $Z'(\omega') := (1+\omega')/2$. Alors la loi de Z' est aussi $P'_{Z'} = \frac{1}{2}(\delta_0 + \delta_1) = P_Z$. Remarquons que si X et Y sont définies sur des espaces probabilisés différents, *il n'y a pas* d'évènement $\{X = Y\}$, pas plus que de variable aléatoire « $X + Y$ ». Essayez d'en écrire la définition explicite pour vous en convaincre.

La remarque suivante peut être sautée en première lecture.

Remarque 6.16. Si X est une v.a. discrète sur $(\Omega, \mathcal{F})$, alors pour *toute* probabilité P sur $(\Omega, \mathcal{F})$, la loi de X sous P est une loi discrète au sens de la définition 6.13. C'est une conséquence de la proposition 6.12. Mais il peut aussi exister sur $(\Omega, \mathcal{F})$ une variable aléatoire réelle Y et une probabilité P telles que Y ne soit pas discrète ($Y(\Omega)$ étant un ensemble infini non dénombrable) mais que la loi de Y sous P soit une loi discrète. Voici un exemple simple. On prend $\Omega = \mathbb{R}$, $\mathcal{F} = \mathcal{P}(\mathbb{R})$ et $Y : \omega \mapsto \omega$ l'application identité sur $\mathbb{R}$. Alors Y n'est pas une v.a. discrète puisque $Y(\Omega) = \mathbb{R}$. Notons au passage que Y est mesurable relativement à n'importe quelle tribu sur l'ensemble d'arrivée puisque l'ensemble de départ Ω est muni ici de la tribu $\mathcal{P}(\Omega)$. En particulier, Y est bien une variable aléatoire réelle. Munissons maintenant $(\Omega, \mathcal{F}) = (\mathbb{R}, \mathcal{P}(\mathbb{R}))$ de la mesure de probabilité $P = \frac{1}{2}\delta_0 + \frac{1}{2}\delta_1$ et cherchons la loi de Y sous P. Comme Y est l'identité sur $\mathbb{R}$, on a l'égalité $Y^{-1}(B) = B$ pour tout borélien B. Par conséquent

$P_Y(B) = P(Y^{-1}(B)) = P(B)$ pour tout $B \in \mathrm{Bor}(\mathbb{R})$. On a donc $P_Y = P = \frac{1}{2}\delta_0 + \frac{1}{2}\delta_1$, ce qui montre que la loi de Y sous P est la loi discrète $\frac{1}{2}\delta_0 + \frac{1}{2}\delta_1$. Bien sûr on a un résultat analogue en remplaçant P par n'importe quelle loi discrète Q sur $\mathbb{R}$, au sens de la définition 6.13.

Remarque 6.17. Pour toute probabilité Q sur $\big(\mathbb{R}, \mathrm{Bor}(\mathbb{R})\big)$, il existe au moins un espace probabilisé $(\Omega, \mathcal{F}, P)$ et une variable aléatoire réelle X sur $(\Omega, \mathcal{F})$ dont la loi sous P soit égale à Q. Il suffit de prendre $\Omega = \mathbb{R}$, $\mathcal{F} = \mathrm{Bor}(\mathbb{R})$ et pour X l'application identité de $\mathbb{R} \to \mathbb{R}$. En prenant $P = Q$, on a clairement $P_X = Q$. Bien entendu, il y a une infinité d'autres solutions à ce problème.

Il y a donc identité entre les mesures de probabilité sur $\big(\mathbb{R}, \mathrm{Bor}(\mathbb{R})\big)$ et les lois des variables aléatoires réelles. Comme nous savons caractériser une probabilité sur $\big(\mathbb{R}, \mathrm{Bor}(\mathbb{R})\big)$ par sa fonction de répartition (théorème 5.30), ceci va nous permettre de classifier les lois des variables aléatoires réelles.

6.2.3 Fonction de répartition

Définition 6.18 (f.d.r. d'une variable aléatoire)**.** *Soient $(\Omega, \mathcal{F}, P)$ un espace probabilisé et X une variable aléatoire sur $(\Omega, \mathcal{F})$. On appelle fonction de répartition (f.d.r.) de X, la fonction F_X définie sur $\mathbb{R}$ par :*

$$\forall x \in \mathbb{R}, \quad F_X(x) = P_X(\,]-\infty, x]) = P(X \leqslant x).$$

La fonction F_X est la fonction de répartition de la probabilité P_X, au sens de la définition 5.27. Elle ne dépend donc que de la loi [7] de X. Deux variables aléatoires de même loi ont même fonction de répartition. La proposition suivante donnant les propriétés générales des fonctions de répartition des variables aléatoires n'est qu'une simple réécriture de la proposition 5.28.

Proposition 6.19. *La fonction de répartition F_X d'une variable aléatoire X est croissante sur $\mathbb{R}$, avec limite 0 en $-\infty$ et 1 en $+\infty$. Elle est continue à droite et limitée à gauche en tout point de $\mathbb{R}$ et vérifie :*

$$\forall x \in \mathbb{R}, \quad P(X = x) = F(x) - F(x-). \tag{6.12}$$

La fonction de répartition d'une variable aléatoire caractérise sa loi, autrement dit : $F_X = F_Y$ si et seulement si les variables aléatoires X et Y ont même loi.

On peut aussi traduire la remarque 5.29 pour obtenir les formules suivantes de

7. Il serait plus correct, mais plus long, de parler de f.d.r. de la loi de X ou même de f.d.r. de la loi de X sous P.

calcul à l'aide de F_X des $P(X \in I)$ pour I intervalle de $\mathbb{R}$:

$$\begin{aligned} P(a < X \leqslant b) &= F_X(b) - F_X(a), && (6.13)\\ P(a \leqslant X \leqslant b) &= F_X(b) - F_X(a-), && (6.14)\\ P(a \leqslant X < b) &= F_X(b-) - F_X(a-), && (6.15)\\ P(a < X < b) &= F_X(b-) - F_X(a), && (6.16)\\ P(X \leqslant a) &= F_X(a), && (6.17)\\ P(X < a) &= F_X(a-), && (6.18)\\ P(X > b) &= 1 - F_X(b), && (6.19)\\ P(X \geqslant b) &= 1 - F_X(b-). && (6.20) \end{aligned}$$

Dans le cas particulier des variables aléatoires discrètes, on peut donner une formule explicite de calcul de la fonction de répartition.

Proposition 6.20 (f.d.r. d'une variable aléatoire discrète)**.** *Soient $(\Omega, \mathcal{F}, P)$ un espace probabilisé et X une variable aléatoire discrète sur $(\Omega, \mathcal{F})$. Fixons une numérotation de l'ensemble au plus dénombrable $X(\Omega)$ par les entiers : $X(\Omega) = \{x_0, x_1, \ldots, x_k, \ldots\}$ et notons $p_k := P(X = x_k)$. La fonction de répartition F_X vérifie alors :*

$$\forall x \in \mathbb{R}, \quad F_X(x) = \sum_{x_k \in X(\Omega)} p_k \mathbf{1}_{[x_k, +\infty[}(x), \tag{6.21}$$

ce qui s'écrit aussi

$$\forall x \in \mathbb{R}, \quad F_X(x) = \sum_{\substack{x_k \in X(\Omega) \\ x_k \leqslant x}} P(X = x_k). \tag{6.22}$$

De plus, si on peut numéroter les éléments de $X(\Omega)$ de manière croissante ($x_k < x_{k+1}$ pour tout k), la fonction F_X est constante sur chaque intervalle $[x_n, x_{n+1}[$ et vaut sur cet intervalle $F_X(x) = \sum_{k \leqslant n} p_k$.

Cette proposition se généralise au cas des variables aléatoires *de loi discrète*[8], en remplaçant $X(\Omega)$ par $X_P(\Omega) := \{x \in \mathbb{R} \,;\, P(X = x) > 0\}$.

Preuve. En utilisant (6.11), on obtient en effet pour tout $x \in \mathbb{R}$,

$$F_X(x) = P_X(]-\infty, x]) = \sum_{x_k \in X(\Omega)} p_k \delta_{x_k}(]-\infty, x]) = \sum_{x_k \in X(\Omega)} p_k \mathbf{1}_{[x_k, +\infty[}(x), \tag{6.23}$$

en notant que

$$\delta_{x_k}(]-\infty, x]) = \begin{cases} 1 & \text{si } x_k \in]-\infty, x] \\ 0 & \text{si } x_k \notin]-\infty, x] \end{cases} = \begin{cases} 1 & \text{si } x_k \leqslant x \\ 0 & \text{si } x_k > x \end{cases} = \mathbf{1}_{[x_k, +\infty[}(x).$$

8. Voir la remarque 6.16.

Si la suite (x_k) est strictement croissante, le réel x appartient à un seul intervalle $[x_n, x_{n+1}[$. Pour $k \leqslant n$, on a alors $x_k \leqslant x_n \leqslant x$ et $\mathbf{1}_{[x_k,+\infty[}(x) = 1$, tandis que si $k > n$, $x_k \geqslant x_{n+1} > x$, donc $\mathbf{1}_{[x_k,+\infty[}(x) = 0$. On a donc $F_X(x) = \sum_{k\leqslant n} p_k$ et ceci étant valable pour tout $x \in [x_n, x_{n+1}[$, la fonction F_X est constante sur cet intervalle. □

À titre d'exemple, la figure 6.2 donne la représentation graphique de F_S où S est la variable aléatoire somme des points de deux dés.

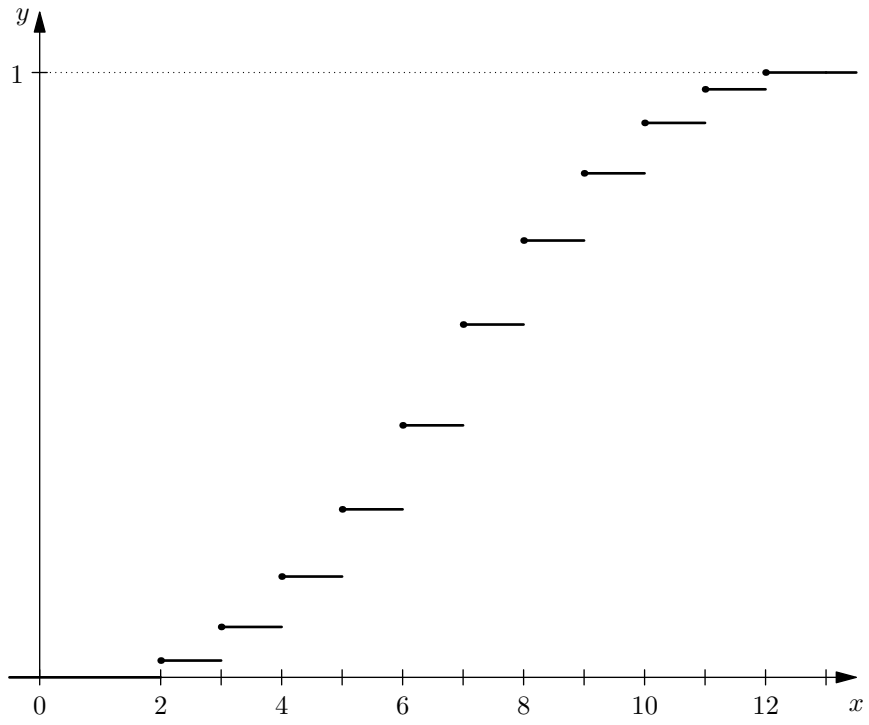


FIGURE 6.2 – f.d.r. de S somme des points de deux dés

Si l'on veut esquisser une classification sommaire des lois des variables aléatoires, on peut commencer par les partager entre les lois à f.d.r. continue sur $\mathbb{R}$ et les lois à f.d.r. non continue [9] sur $\mathbb{R}$. On parle plus simplement de *lois continues* ou encore *lois diffuses* dans le premier cas et de lois non continues ou non diffuses dans le deuxième. Dans la famille des lois non continues, nous connaissons déjà la sous-famille des lois discrètes. Dans la famille des lois continues, une importante sous-famille est celle des *lois à densité* que nous allons examiner maintenant.

9. Rappelons que l'ensemble des points de discontinuité d'une f.d.r. quelconque est au plus dénombrable.

6.2.4 Lois à densité

La loi d'une variable aléatoire X est à densité f si pour tout intervalle de $\mathbb{R}$, la probabilité d'appartenance de X à cet intervalle peut s'écrire comme l'intégrale de f sur cet intervalle. L'apparente simplicité de cette définition informelle est trompeuse. Dans le cadre de cet ouvrage, nous ne pouvons utiliser que l'intégration au sens de Riemann et il se trouve que cette notion n'est pas totalement satisfaisante pour les besoins de la théorie des probabilités. L'intégrale de Lebesgue donnerait une notion plus générale de densité permettant, entre autres, de caractériser les lois à densité comme celles dont la f.d.r. est *absolument continue* [10]. La définition plus restrictive que nous donnons ci-dessous est néanmoins suffisante pour la plupart des cas pratiques.

Définition 6.21 (densité de probabilité)**.** *On appelle densité de probabilité sur* $\mathbb{R}$ *toute fonction* f *vérifiant*

a) f *est définie et positive sur* $\mathbb{R}\backslash K$*, où* K *est une partie finie (éventuellement vide) de* $\mathbb{R}$ *;*

b) f *est Riemann intégrable sur tout intervalle* $[a,b] \subset \mathbb{R}\backslash K$ *;*

c) l'intégrale généralisée de f *sur* $]-\infty,+\infty[$ *converge et*

$$\int_{-\infty}^{+\infty} f(t)\,\mathrm{d}t = 1.$$

Si f est une fonction positive définie seulement sur un intervalle $]a,b[$ de $\mathbb{R}$ et telle que $\int_a^b f(t)\,\mathrm{d}t = 1$, on peut en faire une densité en la prolongeant à tout $\mathbb{R}$ en posant $f(t) := 0$ pour $t \notin]a,b[$. Voici quatre exemples simples de densités :

$$f_1(t) := \frac{1}{b-a}\mathbf{1}_{[a,b]}(t); \qquad f_2(t) := \frac{1}{2\sqrt{t}}\mathbf{1}_{]0,1]}(t);$$

$$f_3(t) := \mathrm{e}^{-t}\mathbf{1}_{[0,+\infty[}(t); \qquad f_4(t) := \frac{1}{\pi(1+t^2)}.$$

Remarque 6.22 (usage des indicatrices dans les formules explicites)**.** La définition de f_2 repose sur un *abus d'écriture* d'usage courant. En effet il y a en toute rigueur un problème pour calculer $f_2(t)$ lorsque $t \leqslant 0$, puisqu'alors il nous faut former le produit de l'expression $\frac{1}{2\sqrt{t}}$ non définie (du moins en tant que nombre réel) par 0. La convention adoptée est que si la formule de calcul d'une fonction contient le produit d'une indicatrice par une expression non définie lorsque cette indicatrice est nulle, le produit vaut 0 dans ce cas. Ceci permet de considérer que la « définition » de f_2 comme ci-dessus est un raccourci d'écriture commode pour :

$$f_2(t) := \begin{cases} \frac{1}{2\sqrt{t}} & \text{si } t \in]0,1], \\ 0 & \text{si } t \notin]0,1]. \end{cases}$$

10. Voir la définition 6.30, p. 219.

Définition 6.23. *Soient $(\Omega, \mathcal{F}, P)$ un espace probabilisé et X une variable aléatoire réelle sur $(\Omega, \mathcal{F})$. La loi de X sous P a pour densité f si :*

$$\forall a \in \mathbb{R},\ \forall b \geqslant a, \qquad P(X \in]a, b]) = \int_a^b f(t)\,\mathrm{d}t. \tag{6.24}$$

On dit aussi par abus de langage que X a pour densité f (lorsqu'il n'y a pas ambiguïté sur P).

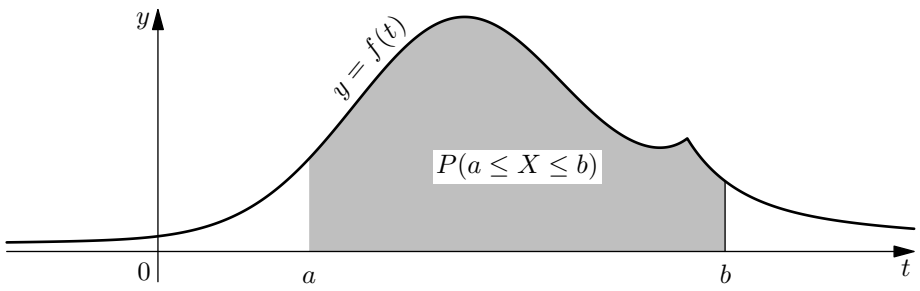


FIGURE 6.3 – $P(a < X \leqslant b) = \int_a^b f(t)\,\mathrm{d}t$ pour X de densité f

Remarque 6.24. Il est clair d'après cette définition que si Y est une autre variable aléatoire ayant même loi que X (donc mêmes probabilités d'appartenance aux intervalles), elle a aussi la densité f. D'autre part, il n'y a pas unicité de la densité d'une variable aléatoire. Par exemple $g_1 = \mathbf{1}_{[0,1]}$ et $g_2 = \mathbf{1}_{]0,1[}$ sont deux densités de probabilité qui donnent les mêmes intégrales : $\int_a^b g_1(t)\,\mathrm{d}t = \int_a^b g_2(t)\,\mathrm{d}t$ pour toute paire de réels a et b. Ces deux fonctions peuvent chacune être prise comme densité de la loi uniforme sur $[0, 1]$. Nous y reviendrons ci-dessous.

L'exemple ci-dessus montre qu'il ne suffit pas de vérifier que deux variables aléatoires ont des densités qui diffèrent en un point pour en déduire qu'elles n'ont pas même loi. Le lemme suivant donne une condition suffisante pratique pour que deux variables à densité n'aient pas même loi.

Lemme 6.25. *Soient X et Y deux variables aléatoires admettant respectivement pour densité les fonctions f et g. On suppose qu'il existe un réel t_0 tel que $f(t_0) \neq g(t_0)$ et que de plus, f et g sont* toutes deux continues au point t_0. *Alors X et Y n'ont pas même loi.*

Preuve. On peut supposer sans perte de généralité que $f(t_0) < g(t_0)$. On va exploiter la continuité de f et g au point t_0 pour construire un intervalle $[a, b]$ voisinage de t_0 tel que $\int_a^b f(t)\,\mathrm{d}t < \int_a^b g(t)\,\mathrm{d}t$. Comme ces deux intégrales sont égales respectivement à $P(X \in]a, b])$ et $P(Y \in]a, b])$, ceci impliquera que X et Y n'ont pas même loi.

Fixons $\varepsilon > 0$ tel que $f(t_0) + \varepsilon < g(t_0) - \varepsilon$, par exemple $\varepsilon = (g(t_0) - f(t_0))/3$. Par continuité de f et g au point t_0, il existe deux réels $\delta_1 > 0$ et $\delta_2 > 0$ tels que :

$$\forall t \in]t_0 - \delta_1, t_0 + \delta_1[, \quad f(t) < f(t_0) + \varepsilon \quad \text{et} \quad \forall t \in]t_0 - \delta_2, t_0 + \delta_2[, \quad g(t) > g(t_0) - \varepsilon.$$

L'intersection des deux intervalles ouverts $]t_0 - \delta_1, t_0 + \delta_1[$ et $]t_0 - \delta_2, t_0 + \delta_2[$ contient un intervalle $[a, b]$ avec $a < b$, on peut prendre par exemple $a = t_0 - \frac{1}{2}\min(\delta_1, \delta_2)$ et $b = t_0 + \frac{1}{2}\min(\delta_1, \delta_2)$. On dispose alors des inégalités suivantes :

$$\forall t \in [a, b], \quad f(t) < f(t_0) + \varepsilon < g(t_0) - \varepsilon < g(t),$$

qui par intégration sur $[a, b]$ nous donnent

$$\int_a^b f(t)\,\mathrm{d}t \leqslant (b-a)(f(t_0) + \varepsilon) < (b-a)(g(t_0) - \varepsilon) \leqslant \int_a^b g(t)\,\mathrm{d}t,$$

d'où $\int_a^b f(t)\,\mathrm{d}t \neq \int_a^b g(t)\,\mathrm{d}t$, ce qui achève la preuve [11]. □

Examinons maintenant les relations entre densité (lorsqu'elle existe) et fonction de répartition (qui elle, existe toujours) d'une variable aléatoire.

Proposition 6.26. *Si la variable aléatoire X a pour densité f, sa fonction de répartition F vérifie :*

a) $\forall x \in \mathbb{R}$, $F(x) = \int_{-\infty}^{x} f(t)\,\mathrm{d}t$;

b) F *est continue sur* $\mathbb{R}$;

c) si f *est continue au point* x_0, *alors* F *est dérivable en ce point et* $F'(x_0) = f(x_0)$.

Corollaire 6.27. *Si la variable aléatoire X a pour densité f, les égalités suivantes sont vérifiées pour tous $a, b \in \mathbb{R}$ tels que $a < b$:*

$$P(a < X < b) = P(a < X \leqslant b) = P(a \leqslant X < b) = P(a \leqslant X \leqslant b) = \int_a^b f(t)\,\mathrm{d}t,$$

$$P(X < a) = P(x \leqslant a) = \int_{-\infty}^{a} f(t)\,\mathrm{d}t,$$

$$P(X > b) = P(X \geqslant b) = \int_b^{+\infty} f(t)\,\mathrm{d}t.$$

Preuve de la proposition 6.26 et du corollaire 6.27.

Preuve de a). Puisque X a pour densité f, on a pour tous réels $a < b$,

$$P(X \in]a, b]) = F(b) - F(a) = \int_a^b f(t)\,\mathrm{d}t. \tag{6.25}$$

11. Parmi les inégalités de la ligne précédente, *seule la deuxième est stricte* et ce n'est pas une erreur typographique. Voyez-vous pourquoi ?

Il suffit d'appliquer (6.25) avec $b = x$ fixé et $a = -n$ pour chaque $n \in \mathbb{N}$ tel que $-n < x$. La suite d'évènements

$$A_n := \{X \in \left]-n, x\right]\}, \quad n > -x,$$

est croissante pour l'inclusion et a pour réunion $A = \{X \in \left]-\infty, x\right]\}$. Par continuité monotone séquentielle (cf. proposition 5.16), on a $P(A_n) \uparrow P(A)$, d'où

$$F(x) = P(A) = \lim_{n\to+\infty} P(X \in \left]-n, x\right]) = \lim_{n\to+\infty} \int_{-n}^{x} f(t)\,\mathrm{d}t = \int_{-\infty}^{x} f(t)\,\mathrm{d}t,$$

en notant que l'intégrale généralisée de la densité f converge en $-\infty$. □

Preuve de b). Fixons $x_0 \in \mathbb{R}$ quelconque. On sait déjà que F est continue à droite en tout point comme *toute* fonction de répartition. Il suffit donc de montrer la continuité à gauche en x_0. D'après le point b) de la définition 6.21, il existe $a < x_0$ tel que f soit définie et Riemann intégrable sur tout intervalle $[a, a'] \subset [a, x_0[$. On a alors

$$\lim_{x\uparrow x_0} \int_a^x f(t)\,\mathrm{d}t = \int_a^{x_0} f(t)\,\mathrm{d}t,$$

où la deuxième intégrale est soit une intégrale de Riemann ordinaire soit une intégrale généralisée convergente. Cette relation peut aussi s'écrire à l'aide de F :

$$\lim_{x\uparrow x_0} \big(F(x) - F(a)\big) = F(x_0) - F(a).$$

On en déduit par addition de $F(a)$ que $F(x)$ tend vers $F(x_0)$ quand x tend vers x_0 par valeurs inférieures. □

Preuve de c). Puisque f est continue en x_0, elle est définie sur tout un voisinage de x_0 et donc sur tout un intervalle $]a, b[$ contenant x_0. La continuité de f en x_0 peut alors s'écrire :

$$\forall \varepsilon > 0, \exists\, \left]x_0 - \delta, x_0 + \delta\right[\subset]a, b[\,; \quad \forall t \in \left]x_0 - \delta, x_0 + \delta\right[, \quad |f(t) - f(x_0)| < \varepsilon. \tag{6.26}$$

Pour tout h tel que $0 < |h| < \delta$, on a alors $F(x_0 + h) - F(x_0) = \int_{x_0}^{x_0+h} f(t)\,\mathrm{d}t$ d'où

$$|F(x_0 + h) - F(x_0) - hf(x_0)| = \left| \int_{x_0}^{x_0+h} \Big(f(t) - f(x_0)\Big)\,\mathrm{d}t \right| \leqslant h\varepsilon.$$

En divisant par h on voit que F a bien une dérivée en x_0 et que celle-ci vaut $f(x_0)$. □

La proposition 6.26 est maintenant complètement démontrée. Pour vérifier le corollaire 6.27, il suffit de combiner les relations générales (6.13)–(6.20) avec (6.24) et les points a) et b) de la proposition 6.26. □

Remarques 6.28.

1. Pour toute densité f (au sens de la définition 6.21), il existe une variable aléatoire X ayant f pour densité : il suffit d'appliquer le théorème 5.30 pour obtenir l'existence d'une mesure de probabilité μ sur $\big(\mathbb{R}, \mathrm{Bor}(\mathbb{R})\big)$ ayant pour f.d.r F définie par a). La remarque 6.17 nous assure de l'existence d'une variable aléatoire X de loi μ, donc de f.d.r. F. La preuve du b) ci-dessus nous montre que F est continue sur $\mathbb{R}$. En particulier pour toute paire de réels $a \leqslant b$, on a $P(a < X \leqslant b) = F(b) - F(a) = \int_{-\infty}^{b} f(t)\,\mathrm{d}t - \int_{-\infty}^{a} f(t)\,\mathrm{d}t = \int_a^b f(t)\,\mathrm{d}t$.
2. D'après b) toute variable aléatoire à densité a une fonction de répartition continue. La réciproque *est fausse* : il existe des lois à fonction de répartition continue sans densité, voir la section 6.5 pour un exemple.
3. Par ailleurs si X a une densité, sa fonction de répartition n'est pas forcément dérivable en tout point. Par exemple la densité f_2 ci-dessus a pour fonction de répartition associée $F_2(x) = \sqrt{x}\mathbf{1}_{]0,1]}(x) + \mathbf{1}_{]1,+\infty[}(x)$ (cette écriture condensée signifie que $F_2(x)$ est nul sur $\mathbb{R}^-$, vaut $\sqrt{x}$ entre 0 et 1 et reste constant égal à 1 sur $]1, +\infty[$). F_2 est dérivable en tout point sauf en 0 et en 1.

La proposition suivante donne une règle pratique permettant de trouver la densité (lorsqu'elle existe !) à partir de la fonction de répartition dans les cas les plus courants.

Proposition 6.29. *On suppose que la fonction de répartition F de X est C^1 par morceaux au sens suivant : F est continue sur $\mathbb{R}$ et dérivable sur $\mathbb{R}$ privé (éventuellement) d'un ensemble fini de points $a_1 < \ldots < a_n$. Sur chacun des intervalles ouverts $]-\infty, a_1[$, $]a_i, a_{i+1}[$ $(1 \leqslant i < n)$, $]a_n, +\infty[$, la dérivée f de F est continue. Alors X a pour densité f.*

Preuve. Il est commode de poser $a_0 := -\infty$ et $a_{n+1} = +\infty$. Sur chacun des intervalles ouverts I découpés par les a_i, F est dérivable et sa dérivée f est continue. On sait alors que f a une infinité de primitives sur I et que si l'on fixe un α dans I, toute primitive H de f sur I est de la forme $H(x) = \int_\alpha^x f(t)\,\mathrm{d}t + C$, avec C constante. Comme F est l'une des primitives de f sur I, en prenant $H = F$ et en faisant $x = \alpha$, on voit que la constante C vaut $F(\alpha)$. On a donc pour α et x quelconques dans I, $F(x) - F(\alpha) = \int_\alpha^x f(t)\,\mathrm{d}t$. Fixons α et prenons $x \geqslant \alpha$. Faisons tendre x vers la borne supérieure a_i de I. Comme F est continue (ou dans le cas $a_i = +\infty$, F a une limite 1), l'intégrale généralisée $\int_\alpha^{a_i} f(t)\,\mathrm{d}t$ converge et vaut $F(a_i) - F(\alpha)$ (ou $1 - F(\alpha)$ quand $a_i = +\infty$). De même en faisant tendre α vers a_{i-1} on voit que l'intégrale généralisée $\int_{a_{i-1}}^{a_i} f(t)\,\mathrm{d}t$ converge et vaut $F(a_i) - F(a_{i-1})$ (ou $F(a_i)$ quand $a_{i-1} = -\infty$). Finalement soient a et $b > a$ quelconques dans $\mathbb{R}$. Si a et b sont dans le même intervalle I on a directement $F(b) - F(a) = \int_a^b f(t)\,\mathrm{d}t$. Sinon on note $(a_i)_{i_0 \leqslant i \leqslant i_1}$ l'ensemble de tous les a_i qui sont dans $[a, b]$ et on écrit

$$F(b) - F(a) = F(a_{i_0}) - F(a) + \sum_{i=i_0}^{i_1-1} \big(F(a_{i+1}) - F(a_i)\big) + F(b) - F(a_{i_1}) = \int_a^b f(t)\,\mathrm{d}t,$$

en utilisant la relation de Chasles pour les intégrales généralisées. On a donc toujours $P(X \in]a,b]) = F(b) - F(a) = \int_a^b f(t)\,\mathrm{d}t$, ce qui montre que X a pour densité f. □

L'idée sous-jacente à la proposition 6.29 est que si la f.d.r. F d'une variable aléatoire X est suffisamment régulière, alors la loi de X a une densité. On a vu d'autre part, proposition 6.26 b), que si la loi de X est à densité, sa f.d.r. F est continue. La question qui surgit alors naturellement est : peut-on caractériser l'existence d'une densité par la régularité de la fonction de répartition ? La réponse est oui et la notion de régularité de la f.d.r. qui équivaut à l'existence d'une densité est celle d'absolue continuité. La preuve de cette équivalence est hors de portée de cet ouvrage. Nous allons néanmoins introduire cette notion et montrer que toute loi à densité a une f.d.r. absolument continue. Cela nous sera utile pour montrer l'existence de lois à f.d.r. continue mais sans densité.

Définition 6.30. *La fonction $F : \mathbb{R} \to \mathbb{R}$ est dite* absolument continue *sur $\mathbb{R}$, si pour tout $\varepsilon > 0$, il existe $\delta > 0$ tel que pour toute famille finie d'intervalles $[a_k, b_k] \subset \mathbb{R}$, $1 \leqslant k \leqslant n$ tels que les $]a_k, b_k[$ soient 2 à 2 disjoints,*

$$\sum_{k=1}^{n} (b_k - a_k) < \delta \implies \sum_{k=1}^{n} |F(b_k) - F(a_k)| < \varepsilon. \tag{6.27}$$

Notons que si F est croissante, les valeurs absolues sont superflues dans la dernière inégalité ci-dessus. La continuité absolue sur $\mathbb{R}$ implique évidemment la continuité uniforme sur $\mathbb{R}$. Par ailleurs toute f.d.r. continue sur $\mathbb{R}$ est *uniformément* continue sur $\mathbb{R}$ (exercice 6.15). Nous verrons ultérieurement qu'il existe des f.d.r. continues sur $\mathbb{R}$ mais non absolument continues (cf. p. 239), donc que la continuité absolue est une propriété strictement plus forte que la continuité uniforme.

Proposition 6.31. *La fonction de répartition F d'une loi à densité f est absolument continue sur $\mathbb{R}$.*

Preuve. D'après la définition 6.21, f est Riemann intégrable sur tout sous intervalle fermé borné inclus dans $\mathbb{R}\backslash K$, où K est fini éventuellement vide, f est positive et d'intégrale généralisée sur $\mathbb{R}$ convergente. Notons $d = \operatorname{card} K$, $d = 0$ si K est vide, $s_1 < \cdots < s_d$ les éléments de K et posons $s_0 = -\infty$, $s_{d+1} = +\infty$.

Soit $\varepsilon > 0$ quelconque. En raison de la convergence de l'intégrale généralisée sur $\mathbb{R}$ de f, on peut construire deux suites finies $(r_i)_{0\leqslant i\leqslant d+1}$ et $(t_i)_{0\leqslant i\leqslant d+1}$ telles que $r_i < s_i < t_i < r_{i+1}$ pour $1 \leqslant i \leqslant d$, $r_0 = -\infty$, $t_0 > -\infty$, $r_{d+1} < +\infty$, $t_{d+1} = +\infty$ et

$$\forall i \in [\![0, d+1]\!], \quad F(t_i) - F(r_i) = \int_{r_i}^{t_i} f(u)\,\mathrm{d}u < \frac{\varepsilon}{2(d+2)}, \tag{6.28}$$

avec la convention $F(-\infty) = 0$ et $F(+\infty) = 1$.

Puisque f est Riemann intégrable sur chacun des intervalles fermés bornés $[t_i, r_{i+1}]$, $i = 0, \ldots, d$, elle y est bornée par une constante $M_i = M_i(\varepsilon)$, dépendant de ε

puisque les bornes t_i et r_{i+1} dépendent de ε, de sorte qu'en posant $M = M(\varepsilon) = \max_{0\leqslant i\leqslant d} M_i(\varepsilon)$, on dispose de la majoration :

$$\forall u \in \bigcup_{i=0}^{d} [t_i, r_{i+1}], \quad 0 \leqslant f(u) \leqslant M.$$

Soit $\{[a_k, b_k],\ k = 1, \ldots, n\}$ une famille d'intervalles comme dans la définition 6.30. On va vérifier l'implication (6.27) en prenant $\delta = \varepsilon/(2M)$. Pour cela on remarque que si l'on introduit une subdivision finie $a_k = c_{k,0} < c_{k,1} < \cdots < c_{k,j_k} = b_k$ de l'intervalle $[a_k, b_k]$, en raison de la croissance de F, on a

$$|F(b_k) - F(a_k)| = \sum_{j=1}^{j_k} |F(c_{k,j}) - F(c_{k,j-1})|.$$

Évidemment cette subdivision de certains des intervalles $[a_k, b_k]$ ne modifie pas la longueur totale $\sum_{k=1}^{n}(b_k - a_k)$ de la réunion de tous les intervalles ainsi obtenus.

On procède alors de la façon suivante. Si l'intervalle $]a_k, b_k[$ n'intersecte aucun des $]r_i, t_i[$ on ne le subdivise pas et on utilise la majoration :

$$F(b_k) - F(a_k) = \int_{a_k}^{b_k} f(u)\,\mathrm{d}u \leqslant M(b_k - a_k).$$

Si $]a_k, b_k[$ a une intersection non vide avec un ou plusieurs des $]r_i, t_i[$, on utilise la subdivision générée par cette intersection. Pour les intervalles $]c_{k,j-1}, c_{k,j}[$ de cette subdivision n'intersectant aucun des $]r_i, t_i[$, on majore f comme ci-dessus par M, pour obtenir $|F(c_{k,j}) - F(c_{k,j-1})| \leqslant M(c_{k,j} - c_{k,j-1})$, pour les autres, forcément inclus dans l'un des $]r_i, t_i[$, on se contente provisoirement d'écrire $|F(c_{k,j}) - F(c_{k,j-1})| = \int_{c_{k,j-1}}^{c_{k,j}} f(u)\,\mathrm{d}u$.

En recollant tous les morceaux et en tenant compte de (6.28) et du choix de δ, on obtient bien la conclusion souhaitée. □

6.3 Lois discrètes classiques

Dans toute la suite du chapitre, on fixe un espace probabilisé $(\Omega, \mathcal{F}, P)$, on désigne par X une variable aléatoire sur $(\Omega, \mathcal{F})$ et par P_X sa loi sous P. Cette clause sera implicite chaque fois que nous écrirons dans les définitions « La variable aléatoire X suit la loi ... si ... ». Pour X de loi discrète, nous utiliserons la notation :

$$X_P(\Omega) := \{x \in \mathbb{R};\ P(X = x) > 0\}. \tag{6.29}$$

Bien sûr, $X_P(\Omega)$ est toujours inclus dans l'ensemble des valeurs « possibles » $X(\Omega)$ et dans la plupart des situations pratiques d'utilisation des variables aléatoires, ces deux ensembles sont égaux [12].

12. Pour comprendre l'utilité de ce distinguo entre $X(\Omega)$ et $X_P(\Omega)$, (re)lisez la remarque 6.16.

6.3.1 Lois de Bernoulli

Définition 6.32. *La variable aléatoire X suit la loi de Bernoulli de paramètre p ($p \in [0,1]$) si $X_P(\Omega) = \{0,1\}$ avec :*

$$P(X=1) = p, \qquad P(X=0) = 1-p = q.$$

On notera $X \sim \mathrm{Bern}(p)$.

Si A est un évènement de probabilité p, son indicatrice définie par :

$$\mathbf{1}_A(\omega) = \begin{cases} 1 & \text{si} \quad \omega \in A \\ 0 & \text{si} \quad \omega \notin A \end{cases} = \begin{cases} 1 & \text{si} \quad A \text{ est réalisé} \\ 0 & \text{si} \quad A \text{ n'est pas réalisé} \end{cases}$$

est une variable aléatoire suivant la loi de Bernoulli de paramètre p. Réciproquement, si X est une v.a. de Bernoulli, on peut toujours écrire que $X = \mathbf{1}_A$ *presque-sûrement*, c'est-à-dire $P(X = \mathbf{1}_A) = 1$, en définissant $A = \{\omega \in \Omega,\ X(\omega) = 1\}$.

6.3.2 Loi uniforme sur un ensemble fini de réels

Définition 6.33. *La variable aléatoire X suit la loi uniforme sur $\{x_1, \ldots, x_n\} \subset \mathbb{R}$ si P_X est l'équiprobabilité sur cet ensemble. Notation : $X \sim \mathrm{Unif}\{x_1, \ldots, x_n\}$.*

Autrement dit, $X_P(\Omega) = \{x_1, \ldots, x_n\}$ et

$$\forall k \in [\![1, n]\!], \qquad P(X = x_k) = \frac{1}{n}.$$

D'où

$$P_X = \frac{1}{n} \sum_{k=1}^{n} \delta_{x_k}.$$

Par exemple, le nombre de points indiqué par un dé équilibré suit la loi uniforme sur $[\![1, 6]\!]$.

6.3.3 Lois binomiales

Définition 6.34. *La variable aléatoire X suit la loi binomiale de paramètres n et p, $n \in \mathbb{N}^*$ et $p \in [0,1]$, notation $X \sim \mathrm{Bin}(n,p)$, si $X_P(\Omega) = [\![0, n]\!]$ et*

$$\forall k \in [\![0, n]\!], \qquad P(X = k) = C_n^k p^k (1-p)^{n-k}.$$

La formule ci-dessus définit bien une loi de probabilité puisque les $C_n^k p^k (1-p)^{n-k}$ sont positifs et :

$$\sum_{k=0}^{n} C_n^k p^k (1-p)^{n-k} = \big(p + (1-p)\big)^n = 1^n = 1,$$

en appliquant la formule du binôme de Newton (d'où le nom de la loi). La loi binomiale $\mathrm{Bin}(n,p)$ est la loi du nombre de succès obtenus en une suite de n épreuves répétées indépendantes avec pour chaque épreuve une probabilité de succès p. Ceci a été démontré dans l'exemple 5.57.

De même, soit $A_1,\dots,A_n$ une famille d'évènements mutuellement indépendants tous de même probabilité p et notons X_i la variable de Bernoulli indicatrice de A_i :

$$X_i(\omega)=\begin{cases}1 & \text{si} \quad \omega\in A_i,\\ 0 & \text{si} \quad \omega\in A_i^c.\end{cases}$$

Alors la variable aléatoire $S_n=\sum_{i=1}^{n} X_i$ suit la loi binomiale $\mathrm{Bin}(n,p)$.

6.3.4 Lois hypergéométriques

Alors que la loi binomiale intervient dans les tirages avec remise, la loi hypergéométrique correspond aux tirages sans remise.

Exemple 6.35. Dans une production totale de N objets dont M sont défectueux, on prélève au hasard un échantillon de n objets (tirage sans remise). Soit X le nombre aléatoire d'objets défectueux dans l'échantillon. Quelle est sa loi ?

On peut prendre comme espace Ω l'ensemble de tous les échantillons possibles (toutes les parties à n éléments d'un ensemble de cardinal N) muni de l'équiprobabilité. Chaque échantillon a ainsi une probabilité $1/C_N^n$ d'être choisi. Les échantillons (évènements élémentaires) réalisant l'évènement $\{X=k\}$ sont ceux qui contiennent k objets défectueux et $n-k$ objets non défectueux. Ceci n'est réalisable que si $0\leqslant k\leqslant M$ et $0\leqslant n-k\leqslant N-M$. Dénombrons ces échantillons. On les forme en choisissant k objets défectueux dans une sous-population de taille M et en complétant par $n-k$ objets non défectueux choisis dans une sous-population de taille $N-M$. Il y en a donc $C_M^k\times C_{N-M}^{n-k}$. Finalement :

$$P(X=k)=\frac{C_M^k\times C_{N-M}^{n-k}}{C_N^n} \quad \text{si} \quad \begin{cases}0\leqslant k\leqslant M,\\ 0\leqslant n-k\leqslant N-M.\end{cases} \tag{6.30}$$

Définition 6.36. *La loi définie par* (6.30) *s'appelle loi hypergéométrique de paramètres N, M et n. Notation : $X\sim\mathrm{Hypg}(N,M,n)$. Le paramètre N est l'effectif de la population totale, M celui de la sous-population à laquelle on s'intéresse et n la taille de l'échantillon observé.*

Pour une taille d'échantillon n *fixée*, plus N et M sont grands, moins les tirages sans remise diffèrent des tirages avec remise. Plus précisément, la loi hypergéométrique converge vers la loi binomiale au sens suivant.

Théorème 6.37 (convergence de l'hypergéométrique vers la binomiale)**.** *On suppose que quand N tend vers $+\infty$, $M = M(N)$ tend vers $+\infty$ en vérifiant la condition :*

$$\lim_{N\to+\infty} \frac{M}{N} = p \quad \textit{avec} \quad 0 < p < 1. \tag{6.31}$$

Alors, n restant fixé, la loi hypergéométrique $\mathrm{Hypg}(N, M, n)$ *converge vers la loi binomiale* $\mathrm{Bin}(n, p)$*, ce qui signifie que si $(X_N)_{N\geqslant 1}$ est une suite de v.a. avec $X_N \sim \mathrm{Hypg}(N, M, n)$ et Y est une v.a. de loi binomiale* $\mathrm{Bin}(n, p)$*,alors :*

$$\forall k \in [\![0, n]\!], \qquad \lim_{N\to+\infty} P(X_N = k) = P(Y = k), \tag{6.32}$$

autrement dit :

$$\forall k \in [\![0, n]\!], \qquad \lim_{N\to+\infty} \frac{C_M^k \times C_{N-M}^{n-k}}{C_N^n} = C_n^k p^k (1-p)^{n-k}. \tag{6.33}$$

Preuve. Notons d'abord que comme p est *strictement* positif, l'hypothèse (6.31) implique que M tend vers $+\infty$ avec N ; il en va de même pour $N - M$ puisque $p < 1$. Ensuite,

$$\begin{aligned}\frac{C_M^k \times C_{N-M}^{n-k}}{C_N^n} &= \frac{M!}{k!(M-k)!} \times \frac{(N-M)!}{(n-k)!\big((N-M)-(n-k)\big)!} \times \frac{n!(N-n)!}{N!} \\ &= C_n^k \frac{M!}{(M-k)!} \times \frac{(N-M)!}{\big((N-M)-(n-k)\big)!} \times \frac{(N-n)!}{N!}. \end{aligned} \tag{6.34}$$

Comme k est *fixé* et M tend vers $+\infty$, la première fraction dans (6.34) est le produit de k facteurs M, $(M-1)$, …, $(M-k+1)$ tous équivalents [13] à M d'où :

$$\frac{M!}{(M-k)!} \sim M^k, \qquad N \to +\infty. \tag{6.35}$$

Par le même argument avec $n-k$ et $N-M$ au lieu de k et M :

$$\frac{(N-M)!}{\big((N-M)-(n-k)\big)!} \sim (N-M)^{n-k}, \qquad N \to +\infty. \tag{6.36}$$

Enfin :

$$\frac{(N-n)!}{N!} \sim \frac{1}{N^n}, \qquad N \to +\infty. \tag{6.37}$$

En reportant ces équivalents dans (6.34), on voit que lorsque N tend vers $+\infty$:

$$P(X_N = k) \sim C_n^k \frac{M^k (N-M)^{n-k}}{N^n} = C_n^k \Big(\frac{M}{N}\Big)^k \Big(\frac{N-M}{N}\Big)^{n-k}, \tag{6.38}$$

d'où : $\lim_{N\to+\infty} P(X_N = k) = C_n^k p^k (1-p)^{n-k}$. □

13. Rappelons que deux suites (u_N) et (v_N) sont dites équivalentes lorsque $u_N = v_N(1+\varepsilon_N)$ avec ε_N tendant vers 0 quand N tend vers $+\infty$ (notation : $u_N \sim v_N$).

6.3.5 Lois géométriques

Exemple 6.38 (un problème de temps d'attente)**.**
Considérons une suite infinie d'épreuves répétées indépendantes avec même probabilité de succès $p \in]0,1[$. Soit X le numéro (aléatoire) de la première épreuve où l'on obtient un succès. Si l'on n'obtient jamais de succès, on conviendra que $X = +\infty$. Calculer $P(X = k)$ pour tout $k \in \mathbb{N}^*$. En déduire les valeurs de $P(X \in \mathbb{N}^*)$ et $P(X = +\infty)$.
En notant $R_i = \{\text{succès à la } i\text{-ème épreuve}\}$, on a :

$$\begin{aligned} \{X = k\} &= \{\text{échec aux } (k-1) \text{ premières et succès à la } k\text{-ième}\} \\ &= \Big(\bigcap_{i=1}^{k-1} R_i^c\Big) \cap R_k. \end{aligned}$$

D'où par indépendance des épreuves :

$$P(X = k) = \Big(\prod_{i=1}^{k-1} P(R_i^c)\Big) \times P(R_k) = (1-p)^{k-1}p.$$

Posons $q = 1 - p$ et notons que $q \in]0,1[$. La décomposition de l'évènement $\{X \in \mathbb{N}^*\}$ en la réunion disjointe des $\{X = k\}$ $(k \in \mathbb{N}^*)$ nous donne par σ-additivité :

$$\begin{aligned} P(X \in \mathbb{N}^*) = \sum_{k\in\mathbb{N}^*} P(X = k) &= \sum_{k\in\mathbb{N}^*} q^{k-1}p \\ &= p\sum_{l\in\mathbb{N}} q^l \qquad (l = k-1) \\ &= p\frac{1}{1-q} = 1. \end{aligned}$$

Ainsi avec probabilité 1, le premier succès a lieu au bout d'un nombre *fini* d'épreuves [14]. Remarquons qu'on aurait pu arriver au même résultat en montrant que $P(X = +\infty) = 0$ par la méthode utilisée à l'exemple 5.57 c) en échangeant les rôles de succès et échec. En toute rigueur, X n'est pas une variable aléatoire discrète au sens de la définition 6.7 puisque $X(\Omega)$ est est une partie dénombrable de $\overline{\mathbb{R}}$ au lieu de $\mathbb{R}$. Néanmoins $X' := X\mathbf{1}_{\{X<+\infty\}}$ est une variable aléatoire discrète et ce qui précède montre que X' a même loi [15] que X. Cette loi est celle du temps d'attente du premier succès dans une suite d'épreuves répétées indépendantes, on l'appelle *loi géométrique de paramètre p.*

Définition 6.39. *Une variable aléatoire X suit la loi géométrique de paramètre $p \in]0,1[$, si $X_P(\Omega) = \mathbb{N}^*$ et :*

$$\forall k \in \mathbb{N}^*, \quad P(X = k) = (1-p)^{k-1}p.$$

Notation : $X \sim \mathrm{Geom}(p)$.

14. Mais pas borné par un nombre fixé choisi avant le début des épreuves...

15. Pour être tout à fait rigoureux, il faudrait avoir défini les v.a. à valeurs dans $\overline{\mathbb{R}}$ et leurs lois, ce qui nous aurait fait sortir du cadre de cet ouvrage...

Lorsque X suit une loi géométrique, les probabilités $P(X > n)$ ont une expression particulièrement simple en fonction de $q = 1 - p$. Calculons les de deux façons.

Première méthode. On calcule le reste d'une série géométrique :

$$\begin{aligned} P(X > n) &= \sum_{k=n+1}^{+\infty} q^{k-1}p = \sum_{l=n}^{+\infty} q^l p \\ &= pq^n \sum_{l=n}^{+\infty} q^{l-n} = pq^n \sum_{j=0}^{+\infty} q^j \\ &= \frac{pq^n}{1-q} = q^n. \end{aligned}$$

Deuxième méthode. On se place dans la situation de l'exemple 6.38. L'évènement $\{X > n\}$ se réalise si et seulement si les n premières épreuves donnent un échec, d'où $\{X > n\} = \bigcap_{i=1}^{n} R_i^c$. En utilisant l'indépendance des R_i on en déduit :

$$P(X > n) = \prod_{i=1}^{n} P(R_i^c) = q^n.$$

6.3.6 Lois de Poisson

Définition 6.40. *On dit que la variable aléatoire discrète X suit la loi de Poisson de paramètre $\alpha > 0$ si $X_P(\Omega) = \mathbb{N}$ et*

$$\forall k \in \mathbb{N}, \quad P(X = k) = \frac{\mathrm{e}^{-\alpha}\alpha^k}{k!}.$$

Notation : $X \sim \mathrm{Pois}(\alpha)$.

On sait que la fonction exponentielle a un développement en série entière avec rayon de convergence infini. En particulier :

$$\forall \alpha > 0, \quad \mathrm{e}^{\alpha} = \sum_{k=0}^{+\infty} \frac{\alpha^k}{k!}.$$

On a donc bien :

$$\sum_{k=0}^{+\infty} P(X = k) = \mathrm{e}^{-\alpha} \sum_{k=0}^{+\infty} \frac{\alpha^k}{k!} = \mathrm{e}^{-\alpha}\mathrm{e}^{\alpha} = 1.$$

Une des raisons de l'importance de cette loi est le théorème de convergence de la loi binomiale vers la loi de Poisson.

Théorème 6.41. *Si $(p_n)_{n\geqslant 1}$ est une suite de réels de $[0,1]$ vérifiant*

$$np_n \xrightarrow[n\to+\infty]{} \alpha \in]0, +\infty[, \tag{6.39}$$

alors pour tout $k \in \mathbb{N}$,

$$C_n^k p_n^k (1-p_n)^{n-k} \xrightarrow[n\to+\infty]{} \frac{\mathrm{e}^{-\alpha}\alpha^k}{k!}.$$

Preuve. L'hypothèse (6.39) peut s'écrire sous la forme plus maniable : $np_n = \alpha u_n$ avec u_n tendant vers 1 quand n tend vers $+\infty$. Ainsi $p_n = \alpha u_n/n$ et

$$C_n^k p_n^k (1-p_n)^{n-k} = \frac{n!}{(n-k)!}\frac{1}{k!}\Big(\frac{\alpha}{n}\Big)^k u_n^k \Big(1-\frac{\alpha u_n}{n}\Big)^{n-k}. \tag{6.40}$$

Pour obtenir la limite de cette expression lorsque n tend vers $+\infty$, k restant *fixé*, on remarque successivement que :

$$\lim_{n\to+\infty} \frac{n!}{(n-k)!}\frac{1}{n^k} = 1, \tag{6.41}$$

$$\lim_{n\to+\infty} u_n^k = 1, \tag{6.42}$$

$$\lim_{n\to+\infty} \Big(1-\frac{\alpha u_n}{n}\Big)^{n-k} = \mathrm{e}^{-\alpha}. \tag{6.43}$$

Pour justifier (6.43), on écrit :

$$\Big(1-\frac{\alpha u_n}{n}\Big)^{n-k} = \exp\Big[(n-k)\ln\Big(1-\frac{\alpha u_n}{n}\Big)\Big], \tag{6.44}$$

puis comme $\alpha u_n/n$ tend vers 0 :

$$(n-k)\ln\Big(1-\frac{\alpha u_n}{n}\Big) \sim n\Big(-\frac{\alpha u_n}{n}\Big) \sim -\alpha, \quad (n\to+\infty).$$

Par continuité de la fonction exponentielle, la limite du second membre de (6.44) est donc bien $\mathrm{e}^{-\alpha}$, ce qui prouve (6.43). On obtient alors la conclusion du théorème en passant à la limite dans (6.40). □

Le théorème 6.41 sert de justification théorique à la règle pratique suivante : lorsque n est « grand » et np « petit », on peut remplacer la loi $\mathrm{Bin}(n,p)$ par la loi $\mathrm{Pois}(\alpha)$ où $\alpha = np$. En général on considère que n de l'ordre de quelques centaines et np de l'ordre de quelques unités donnent une bonne approximation. Sous cette forme, cette règle relève plus de la cuisine que des mathématiques. Il est possible par des techniques élémentaires de contrôler l'erreur commise en utilisant cette approximation. Nous nous contenterons ici d'un exemple classique et d'une comparaison graphique pour illustrer la qualité de cette approximation.

Exemple 6.42. Le président d'un bureau de vote est né un 1er avril. Il décide de noter le nombre X de personnes ayant leur anniversaire le même jour que lui parmi les 500 premiers électeurs qui se présentent.

La situation peut être assimilée à une suite d'épreuves répétées indépendantes et X est une variable aléatoire binomiale de paramètres $n = 500$ et $p = 1/365$ (en négligeant la question des années bissextiles sinon on prendrait $p = 4/(3 \times 365 + 366)$, ce qui ne changerait pas grand chose numériquement). Ainsi :

$$P(X = k) = C_{500}^{k} \Big(\frac{1}{365}\Big)^{k} \Big(\frac{364}{365}\Big)^{500-k}.$$

La règle énoncée ci-dessus nous conduit à approximer la loi de X par une loi de Poisson de paramètre :

$$\alpha = np = 500 \times \frac{1}{365}.$$

Voici une comparaison numérique pour les petites valeurs de k :

k	0	1	2	3	4	5
$P(X = k)$	0,253 7	0,348 4	0,238 8	0,108 9	0,037 2	0,010 1
$\dfrac{\mathrm{e}^{-\alpha}\alpha^{k}}{k!}$	0,254 1	0,348 1	0,238 5	0,108 9	0,037 3	0,010 2

Remarquons que la probabilité d'observer plus de 5 anniversaires un 1er avril, calculée par la loi exacte de X ou par son approximation poissonienne est inférieure à 0,003.

Comparaison graphique :

Les *diagrammes en bâtons* ci-dessous représentent la loi binomiale $\mathrm{Bin}(n, p)$ et la loi de Poisson approximante $\mathrm{Pois}(\alpha)$ avec $\alpha = np$. Les segments verticaux (les bâtons) du diagramme représentant la loi d'une variable discrète X (à valeurs dans $\mathbb{N}$) ont une hauteur égale à $P(X = k)$ avec une extrémité inférieure au point d'abscisse k de l'axe horizontal. Pour la lisibilité, on a légèrement décalé vers la gauche les bâtons de la loi de Poisson (en bleu) et vers la droite ceux de la loi binomiale(en rouge). Bien que le diagramme en bâtons de la loi binomiale $\mathrm{Bin}(n, p)$ soit constitué théoriquement de $n+1$ bâtons (et que celui de la loi de Poisson en ait une infinité), seul un petit nombre de bâtons est visible sur les graphiques, les autres correspondant à des probabilités trop petites [16]. L'échelle verticale de chaque figure a été choisie de façon adaptative de façon que l'avant dernière graduation verticale donne la valeur de la plus grande probabilité binomiale. On constate que pour $n = 200$ (figure 6.7), la différence entre les deux diagrammes n'est pratiquement plus discernable *visuellement*.

16. En fait, on s'est contenté d'afficher les probabilités correspondant à k inférieur ou égal à la partie entière supérieure de $2\alpha + 4$. On peut vérifier que la somme des probabilités ainsi négligées est inférieure à 1%, pour chacune des deux lois.

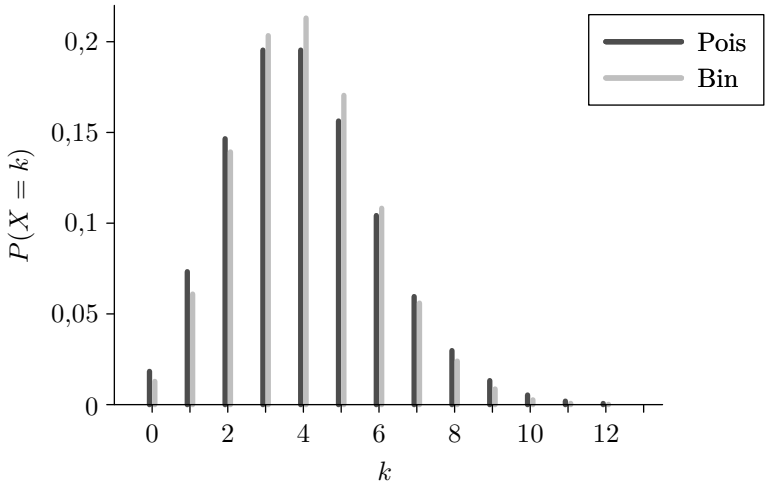


FIGURE 6.4 – Lois Bin(25; 0,16) et Pois(4)

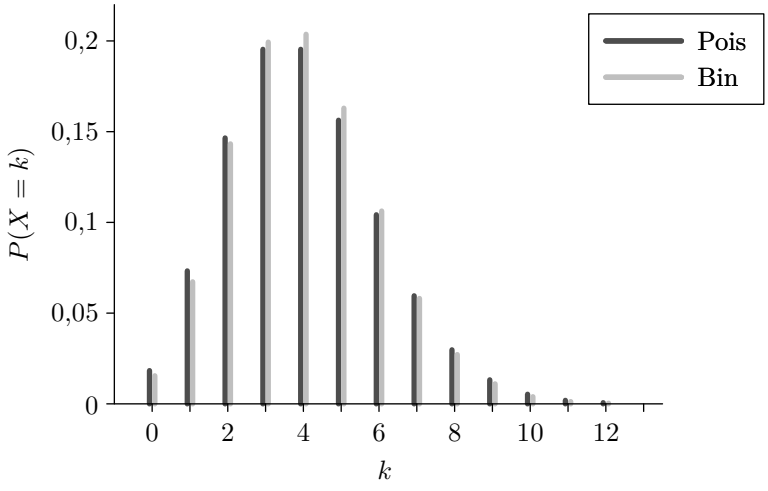


FIGURE 6.5 – Lois Bin(50; 0,08) et Pois(4)

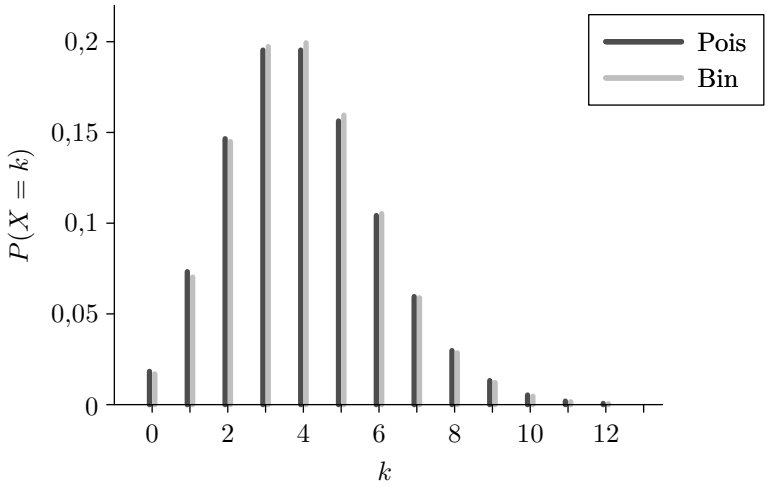


FIGURE 6.6 – Lois Bin(100; 0,04) et Pois(4)

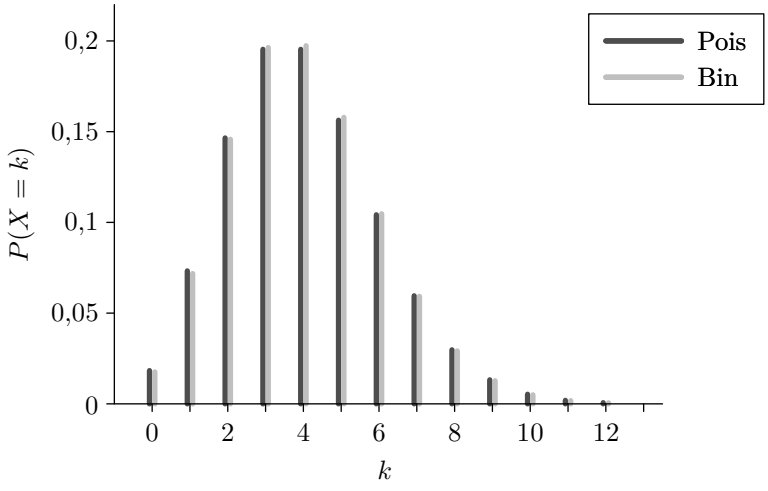


FIGURE 6.7 – Lois Bin(200; 0,02) et Pois(4)

6.3.7 Sur le caractère universel de la loi de Poisson

L'étude qui suit a pour but de mieux faire saisir l'importance de la loi de Poisson, en justifiant au passage le bien fondé de l'hypothèse (6.39) du théorème de convergence de la loi binomiale vers la loi de Poisson.

Considérons un phénomène se traduisant par des observations (ou réalisations) aléatoires pendant un intervalle de temps $[0,1[$ (exemples : désintégrations d'atomes, accidents d'avion, faux numéros de téléphone sur un standard, éruptions volcaniques, naissances de triplés, ...). On suppose que le phénomène vérifie les hypothèses suivantes :

(a) Les observations dans des intervalles de temps disjoints sont indépendantes.

(b) Pour tout réel t tel que $0 \leqslant t < t+T \leqslant 1$ la loi du nombre (aléatoire) d'observations dans l'intervalle $[t, t+T[$ ne dépend que de la durée T de cet intervalle.

Partageons l'intervalle de temps $[0,1[$ en n intervalles disjoints

$$I_{n,k} = \Big[\frac{k}{n}, \frac{k+1}{n}\Big[, \quad 0 \leqslant k < n.$$

Notons :

$$\begin{aligned} p_n &= P\big(\text{avoir } \textit{exactement} \text{ une observation dans } I_{n,k}\,\big) \\ r_n &= P\big(\text{avoir } \textit{au moins} \text{ une observation dans } I_{n,k}\,\big). \end{aligned}$$

D'après (b), p_n et r_n ne dépendent pas de k. En écrivant de deux façons la probabilité de n'avoir aucune observation dans $[0,1[$, on obtient :

$$\begin{aligned} 1 - r_1 &= P\Big(\bigcap_{k=0}^{n-1} \Big\{\text{aucune observation dans } \Big[\frac{k}{n}, \frac{k+1}{n}\Big[\Big\}\Big) \\ &= (1-r_n)^n \qquad \text{en utilisant } (a) \text{ et } (b). \end{aligned}$$

D'où $1 - r_n = (1-r_1)^{1/n}$. Un développement limité à l'ordre 1 de cette expression nous permet d'écrire :

$$(1-r_1)^{1/n} = \exp\Big(\frac{1}{n}\ln(1-r_1)\Big) = 1 + \frac{1}{n}\ln(1-r_1) + \frac{\delta_n}{n}, \quad \text{où } \lim_{n\to+\infty} \delta_n = 0.$$

Nous noterons désormais :

$$-\ln(1-r_1) = \alpha, \qquad \alpha \in \,]0, +\infty[. \tag{6.45}$$

Il vient $r_n = \frac{\alpha}{n} - \frac{\delta_n}{n}$ d'où $\lim_{n\to+\infty} n r_n = \alpha$.

Pour le type de phénomène que nous envisageons, il est vraisemblable que l'on arrive à isoler les observations lorsque les intervalles de la subdivision sont assez petits : asymptotiquement, la probabilité d'avoir plus d'une observation dans $[k/n, (k+1)/n[$ est négligeable devant celle d'en avoir exactement une. Plus précisément, nous rajoutons à notre modèle l'hypothèse :

(c) $\varepsilon_n = \dfrac{r_n - p_n}{p_n} \longrightarrow 0$, quand $n \to +\infty$.

D'après (c), r_n/p_n converge vers 1, d'où $\lim_{n\to+\infty} np_n = \alpha$.

Cherchons maintenant la probabilité d'avoir exactement l observations dans $[0,1[$. Cette probabilité peut se décomposer en :

$$P(l \text{ observations dans } [0,1[) = P(A_n) + P(B_n), \tag{6.46}$$

où

$$\begin{aligned} A_n &= \big\{l \text{ observations avec au plus une dans chaque } I_{n,k}\big\}, \\ B_n &= \big\{l \text{ observations avec au moins un } I_{n,k} \text{ en contenant plusieurs}\big\}. \end{aligned}$$

Calcul de $P(A_n)$: Notons

$$\begin{aligned} D_i &= \Big\{\text{exactement une observation dans } \Big[\frac{i}{n}, \frac{i+1}{n}\Big[\Big\}, \\ E_i &= \Big\{\text{aucune observation dans } \Big[\frac{i}{n}, \frac{i+1}{n}\Big[\Big\}, \quad 0 \leqslant i < n. \end{aligned}$$

L'évènement A_n est la réunion disjointe de tous les évènements du type :

$$\Big(\bigcap_{i\in I} D_i\Big) \cap \Big(\bigcap_{j\in J} E_j\Big),$$

où $I \subset [\![1,n]\!]$, $\operatorname{card} I = l$ et $J = [\![1,n]\!]\backslash I$. D'après l'hypothèse d'indépendance (a), la probabilité de chacun de ces évènements est $p_n^l(1-r_n)^{n-l}$ d'où :

$$P(A_n) = C_n^l p_n^l (1-r_n)^{n-l}.$$

Pour trouver la limite de $P(A_n)$ lorsque n tend vers l'infini, l restant fixé, il suffit d'adapter la preuve du théorème 6.41 : ici nous avons à trouver la limite de $(1-r_n)^{n-l}$ au lieu de $(1-p_n)^{n-l}$. Or

$$(n-l)\ln(1-r_n) \sim -nr_n \sim -\alpha,$$

d'où $\lim_{n\to+\infty}(1-r_n)^{n-l} = \mathrm{e}^{-\alpha}$. On en déduit :

$$\lim_{n\to+\infty} P(A_n) = \frac{\mathrm{e}^{-\alpha}\alpha^l}{l!}. \tag{6.47}$$

Majoration de $P(B_n)$: Le calcul de $P(B_n)$ étant trop compliqué, nous nous contenterons d'une majoration. La réalisation de l'évènement B_n implique l'existence d'au moins deux observations dans au moins l'un des intervalles de longueur $1/n$. Autrement dit :

$$B_n \subset \bigcup_{k=0}^{n-1} \Big\{\text{au moins deux observations dans } \Big[\frac{k}{n}, \frac{k+1}{n}\Big[\Big\}.$$

Par conséquent

$$\begin{aligned} P(B_n) &\leqslant P\Bigg(\bigcup_{k=0}^{n-1}\Big\{\text{au moins deux observations dans } \Big[\frac{k}{n},\frac{k+1}{n}\Big[\Big\}\Bigg) \\ &\leqslant \sum_{k=0}^{n-1}(r_n-p_n) = n(r_n-p_n) = np_n\varepsilon_n. \end{aligned}$$

D'après (c) et la convergence de np_n vers α, $np_n\varepsilon_n$ tend vers 0 quand n tend vers l'infini. Il en est donc de même pour $P(B_n)$.

Pour conclure, on remarque que (6.46) est vérifiée pour tout entier $n \geqslant 1$ et que le premier membre de cette égalité ne dépend pas de n. Cette égalité reste donc vraie à la limite :

$$P\big(l \text{ observations dans } [0,1[\big) = \lim_{n\to+\infty}\big(P(A_n)+P(B_n)\big) = \frac{\mathrm{e}^{-\alpha}\alpha^l}{l!},$$

d'après (6.47) et la majoration de $P(B_n)$. Ce résultat étant valable pour tout entier l, nous avons donc démontré le théorème suivant.

Théorème 6.43. *Soit un phénomène donnant lieu à des observations aléatoires vérifiant les hypothèses suivantes.*

- (a) *Les observations dans des intervalles de temps disjoints sont indépendantes*
- (b) *Pour tout réel t tel que $0 \leqslant t < t+T \leqslant 1$ la loi du nombre (aléatoire) d'observations dans l'intervalle $[t,t+T[$ ne dépend que de la durée T de cet intervalle.*
- (c) *En notant p_n la probabilité d'avoir exactement une observation dans un intervalle de temps de durée $1/n$ et r_n celle d'en avoir au moins une, $\varepsilon_n = (r_n-p_n)/p_n$ tend vers 0 quand n tend vers l'infini..*

Alors le nombre aléatoire d'observations dans l'intervalle $[0,1[$ suit la loi de Poisson de paramètre α défini par $\alpha = -\ln(1-r_1)$.

Remarque 6.44. L'examen attentif de la démonstration ci-dessus montre que la structure d'ordre de l'intervalle $[0,1[$ n'y joue aucun rôle. L'important est la possibilité de réaliser une partition de $[0,1[$ en intervalles de même longueur tendant vers 0. Par conséquent en remplaçant la longueur par l'aire ou le volume, il est possible d'obtenir une version spatiale en dimension 2 ou 3 du théorème 6.43. Ceci permet de comprendre pourquoi la loi de Poisson fournit une bonne modélisation par exemple du nombre d'erreurs typographiques dans une page imprimée, du nombre d'impacts de météorites sur un territoire donné, du nombre d'accidents sur une portion d'autoroute pendant une période donnée, du nombre de raisins dans une portion de cake, du nombre d'étoiles dans une région de l'univers, ...

6.4 Lois à densité classiques

6.4.1 Lois uniformes

Définition 6.45. *La variable aléatoire réelle X suit la loi uniforme sur l'intervalle $[a,b]$ $(-\infty < a < b < +\infty)$ si*

$$\forall B \in \mathrm{Bor}(\mathbb{R}), \quad P(X \in B) = P_X(B) = \frac{\lambda_1([a,b] \cap B)}{\lambda_1([a,b])}, \tag{6.48}$$

où λ_1 désigne la mesure de Lebesgue sur $\mathbb{R}$ (en particulier $\lambda_1([a,b]) = b-a$). Notation : $X \sim \mathrm{Unif}[a,b]$.

Calculons la fonction de répartition F en prenant $B =]-\infty, x]$ pour x quelconque dans (6.48).

$$F(x) = P_X(]-\infty,x]) = \frac{\lambda_1([a,b]\cap]-\infty,x])}{\lambda_1([a,b])} = \begin{cases} 0 & \text{si} \quad -\infty < x < a; \\ \dfrac{x-a}{b-a} & \text{si} \quad a \leqslant x < b; \\ 1 & \text{si} \quad b \leqslant x < +\infty. \end{cases}$$

La fonction de répartition F est affine par morceaux, donc aussi C^1 par morceaux au sens de la proposition 6.29, avec dérivabilité sur $\mathbb{R}\backslash\{a,b\}$ (figure 6.8). La loi a donc

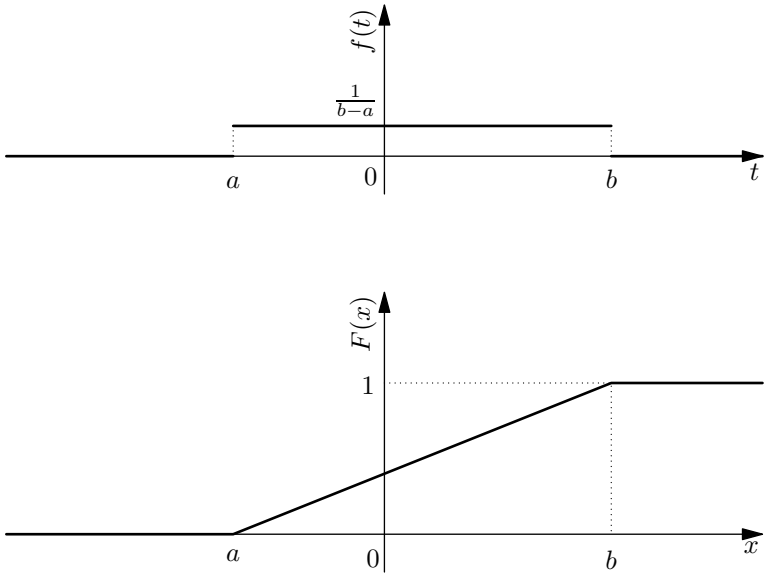


FIGURE 6.8 – Densité f et f.d.r. F de la loi $\mathrm{Unif}[a,b]$

une densité f qui s'obtient par dérivation de F, ce qui nous donne $f(t) = 0$ si $t < a$,

$f(t) = \frac{1}{b-a}$ si $a < t < b$ et $f(t) = 0$ si $t > b$. On complète la définition de f en la prolongeant en a et b, par exemple en posant $f(a) = f(b) = \frac{1}{b-a}$. La loi uniforme sur $[a, b]$ admet donc pour densité

$$f = \frac{1}{b-a}\mathbf{1}_{[a,b]}.$$

Dans les calculs faisant intervenir la loi uniforme sur $[a, b]$, il est vivement conseillé d'utiliser chaque fois que c'est possible la formule (6.48) de préférence aux calculs d'intégrales de f.

Remarque 6.46. Comme $\lambda_1(\{a\}) = \lambda_1(\{b\}) = 0$, la loi uniforme sur $[a, b]$ est aussi la loi uniforme sur $]a, b]$, $[a, b[$ ou $]a, b[$.

Une des raisons de l'importance de la loi uniforme sur $[0, 1]$ est le théorème suivant.

Théorème 6.47. *Si X est une variable aléatoire réelle de fonction de répartition continue strictement croissante F et si U est une variable aléatoire de loi uniforme sur $[0, 1]$, alors la variable aléatoire $Y := F^{-1}(U)$ a même loi que X.*

Rappelons qu'avoir même loi que X ne signifie aucunement être égale à X. Ce théorème permet de réduire la simulation informatique de la loi de X à celle de U. Nous verrons ultérieurement que ce résultat s'étend à toutes les fonctions de répartition, sans hypothèse de continuité ni de croissance stricte, *via* une redéfinition de F^{-1}.

Preuve. Comme F est continue strictement croissante, c'est une *bijection* de $\mathbb{R}$ sur son image $]0, 1[$ (en raison de la stricte monotonie de F, les bornes 0 et 1 ne sont pas atteintes). Par conséquent $F^{-1} :]0, 1[\to \mathbb{R}$ est bien définie et vérifie :

$$\forall u \in]0, 1[, \forall x \in \mathbb{R}, \quad F^{-1}(u) \leqslant x \text{ si et seulement si } u \leqslant F(x).$$

Comme $P(0 < U < 1) = 1$, on en déduit que les évènements $\{F^{-1}(U) \leqslant x\}$ et $\{U \leqslant F(x)\}$ ont même probabilité. Pour obtenir la fonction de répartition de Y, on remarque alors que pour tout $x \in \mathbb{R}$,

$$P(Y \leqslant x) = P(F^{-1}(U) \leqslant x) = P(U \leqslant F(x)) = \frac{\lambda_1([0, F(x)])}{\lambda_1([0, 1])} = F(x).$$

Ainsi Y a pour fonction de répartition F, donc a même loi que X. □

6.4.2 Lois exponentielles

Définition 6.48. *Soit a un réel strictement positif. La variable aléatoire réelle X suit la loi exponentielle de paramètre a si elle admet pour densité*

$$f(t) = a\mathrm{e}^{-at}\mathbf{1}_{[0,+\infty[}(t).$$

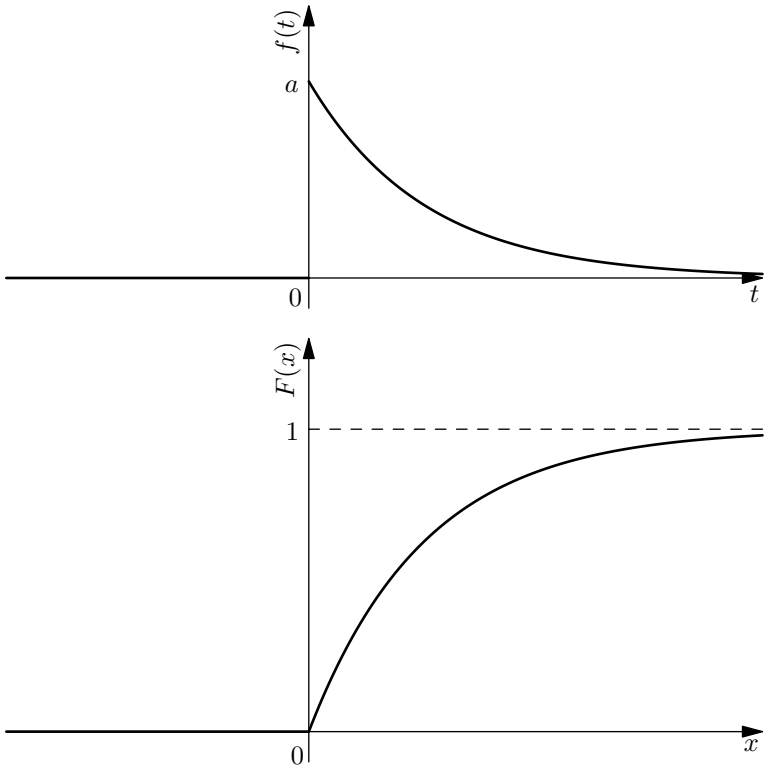


FIGURE 6.9 – Densité et f.d.r. de la loi Exp(a)

En pratique, plutôt que de travailler avec la fonction de répartition d'une loi exponentielle, il est plus commode d'utiliser la *fonction de survie* G :

$$G(x) = P(X > x) = 1 - F(x) = \begin{cases} 1 & \text{si } x \leqslant 0, \\ \mathrm{e}^{-ax} & \text{si } x > 0. \end{cases}$$

Les lois exponentielles sont souvent choisies pour modéliser des temps d'attente : temps d'attente à partir de maintenant du prochain tremblement de terre, du prochain faux numéro sur une ligne téléphonique, de la prochaine désintégration d'un atome de radium, etc.

La raison de ce choix est la propriété *d'absence de mémoire* en temps continu qui caractérise la famille des lois exponentielles.

Théorème 6.49 (absence de mémoire)**.**

i) Si la variable aléatoire X suit une loi exponentielle, elle vérifie la propriété d'absence de mémoire :

$$\forall s \in \mathbb{R}^+, \forall t \in \mathbb{R}^+, \quad P(X > t + s \mid X > t) = P(X > s). \tag{6.49}$$

ii) Réciproquement si une variable aléatoire X vérifie (6.49), elle suit une loi exponentielle.

Comme la fonction de survie caractérise la loi, (6.49) signifie que la loi de $(X-t)$ conditionnelle à $\{X > t\}$ est la même que la loi de X.

En préliminaire à la preuve du théorème, remarquons que la probabilité conditionnelle dans (6.49) s'exprime commodément à l'aide de la fonction de survie G,définie par $G(x) := P(X > x)$. En effet, s étant positif, l'inégalité $t+s \geqslant t$ entraîne l'inclusion $\{X > t+s\} \subset \{X > t\}$ et l'égalité d'évènements : $\{X > t+s\} \cap \{X > t\} = \{X > t+s\}$. On en déduit

$$P(X > t+s \mid X > t) = \frac{P(X > t+s)}{P(X > t)} = \frac{G(t+s)}{G(t)}. \tag{6.50}$$

Preuve de i). Si X suit la loi exponentielle de paramètre a, $G(x) = \mathrm{e}^{-ax}$ pour tout x positif et (6.50) se traduit alors par :

$$P(X > t+s \mid X > t) = \frac{\mathrm{e}^{-a(t+s)}}{\mathrm{e}^{-at}} = \mathrm{e}^{-as} = P(X > s).$$

Ainsi X de loi $\mathrm{Exp}(a)$ vérifie la propriété d'absence de mémoire (6.49). □

Preuve de ii). Soit X une variable aléatoire dont la loi vérifie (6.49) et G sa fonction de survie. Comme $G = 1 - F$ (où F désigne la fonction de répartition de X), G est décroissante et continue à droite et tend vers 0 en $+\infty$. De plus l'écriture de (6.49) suppose implicitement que $G(t) > 0$ pour tout $t \geqslant 0$ car sinon $P(\,.\mid X > t)$ ne serait pas définie. Grâce à (6.50), on voit que la propriété d'absence de mémoire (6.49) équivaut à

$$\forall s \in \mathbb{R}^+, \forall t \in \mathbb{R}^+, \quad \frac{G(t+s)}{G(t)} = G(s).$$

La fonction de survie G doit donc être une solution décroissante, continue à droite, tendant vers 0 en $+\infty$ et telle que $0 < G(t) \leqslant 1$ de l'équation fonctionnelle [17] :

$$\forall s \in \mathbb{R}^+, \forall t \in \mathbb{R}^+, \quad G(t+s) = G(t)G(s). \tag{6.51}$$

En faisant $s = t = 0$ dans (6.51), on obtient $G(0) = G(0)^2$ et comme $G(0) > 0$, on a

$$G(0) = 1. \tag{6.52}$$

En faisant $s = t$ dans (6.51), on obtient $G(2t) = G(t)^2$, puis de proche en proche

$$\forall n \in \mathbb{N}^*, \forall t \geqslant 0, \quad G(nt) = G(t)^n. \tag{6.53}$$

17. Une équation fonctionnelle est une équation dont l'inconnue est ... une fonction ! Les équations différentielles sont des exemples bien connus d'équations fonctionnelles.

En particulier pour $t = 1/d$, $d \in \mathbb{N}^*$:

$$\forall n \in \mathbb{N}^*, \forall d \in \mathbb{N}^*, \quad G\Big(\frac{n}{d}\Big) = G\Big(\frac{1}{d}\Big)^n. \tag{6.54}$$

Lorsque $n = d$, (6.54) donne $G(1) = G(1/d)^d$ d'où

$$\forall d \in \mathbb{N}^*, \quad G\Big(\frac{1}{d}\Big) = G(1)^{1/d}. \tag{6.55}$$

Nous connaissons maintenant G sur l'ensemble des rationnels positifs puisque (6.52), (6.53), (6.54) et (6.55) nous donnent

$$\forall r \in \mathbb{Q}^+, \quad G(r) = G(1)^r. \tag{6.56}$$

Soit $x \in \mathbb{R}^+ \backslash \mathbb{Q}^+$, x est limite d'une suite décroissante (r_n) de rationnels. Comme G est continue à droite, $G(r_n)$ converge vers $G(x)$. D'autre part l'application $y \mapsto G(1)^y$ est continue sur $\mathbb{R}$. Ainsi en appliquant (6.56) à r_n et en faisant tendre n vers l'infini on obtient

$$\forall x \in \mathbb{R}^+, \quad G(x) = G(1)^x. \tag{6.57}$$

A priori la constante $G(1)$ est dans $]0, 1]$. On peut écarter la valeur $G(1) = 1$ car sinon d'après (6.57), la limite en $+\infty$ de G serait 1 alors qu'elle vaut 0.

Finalement, puisque $0 < G(1) < 1$, on peut poser $G(1) = \mathrm{e}^{-a}$ pour un réel $a > 0$ (cela revient à prendre $a = -\ln G(1)$). On peut alors réécrire (6.57) sous la forme

$$\forall x \in \mathbb{R}^+, \quad G(x) = \mathrm{e}^{-ax}.$$

La fonction de survie G est donc la même que celle de la loi exponentielle de paramètre a, donc X suit cette loi (puisque la fonction de survie caractérise la loi au même titre que la fonction de répartition). □

6.4.3 Lois gaussiennes

Ces lois jouent un rôle capital dans l'étude des lois limites de sommes de variables aléatoires indépendantes. Par exemple (théorème de de Moivre Laplace) si S_n suit la loi $\mathrm{Bin}(n, p)$, alors pour tout $x \in \mathbb{R}$, $P\big(S_n - np \leqslant x\sqrt{np(1-p)}\big)$ converge quand n tend vers l'infini vers $\Phi(x)$, où Φ est la f.d.r. de la loi gaussienne $\mathfrak{N}(0, 1)$.

Définition 6.50. *On dit que la variable aléatoire X suit la loi gaussienne ou normale $\mathfrak{N}(m, \sigma)$ si elle a pour densité la fonction :*

$$f_{m,\sigma} : \mathbb{R} \longrightarrow \mathbb{R}^+ \quad t \longmapsto \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(t-m)^2}{2\sigma^2}\right).$$

La loi $\mathfrak{N}(0, 1)$ est appelée loi normale standard.

Tous les calculs de probabilités concernant une variable aléatoire de loi $\mathfrak{N}(m, \sigma)$ peuvent se ramener à des calculs sur une variable de loi normale standard.

Proposition 6.51. *Si la variable aléatoire X suit la loi $\mathfrak{N}(m,\sigma)$, $Y := (X-m)/\sigma$ suit la loi $\mathfrak{N}(0,1)$. Autrement dit, toute v.a. gaussienne X de loi $\mathfrak{N}(m,\sigma)$ peut s'écrire $X = \sigma Y + m$ avec Y de loi $\mathfrak{N}(0,1)$.*

Preuve. On calcule $P(a < Y \leqslant b)$ pour a et b réels quelconques ($a < b$).

$$\begin{aligned} P\Big(a < \frac{X-m}{\sigma} \leqslant b\Big) &= P(\sigma a + m < X \leqslant \sigma b + m) \\ &= \int_{\sigma a+m}^{\sigma b+m} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right) \mathrm{d}x. \end{aligned}$$

Il suffit alors de faire le changement de variable $y = (x-m)/\sigma$ pour obtenir

$$\forall a \in \mathbb{R}, \forall b > a, \quad P(a < Y \leqslant b) = \int_a^b \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right) \mathrm{d}y.$$

Donc Y a bien pour densité $f_{0,1}$. □

Remarque 6.52. En adaptant le calcul fait dans la preuve de la proposition 6.51, on voit facilement que la famille des lois gaussiennes est *stable par transformations affines* : si X a pour loi $\mathfrak{N}(m,\sigma)$, alors pour tout $(\alpha,\beta) \in \mathbb{R}^* \times \mathbb{R}$, la v.a. $\alpha X + \beta$ est encore gaussienne, de loi $\mathfrak{N}(\alpha m + \beta, |\alpha|\sigma)$.

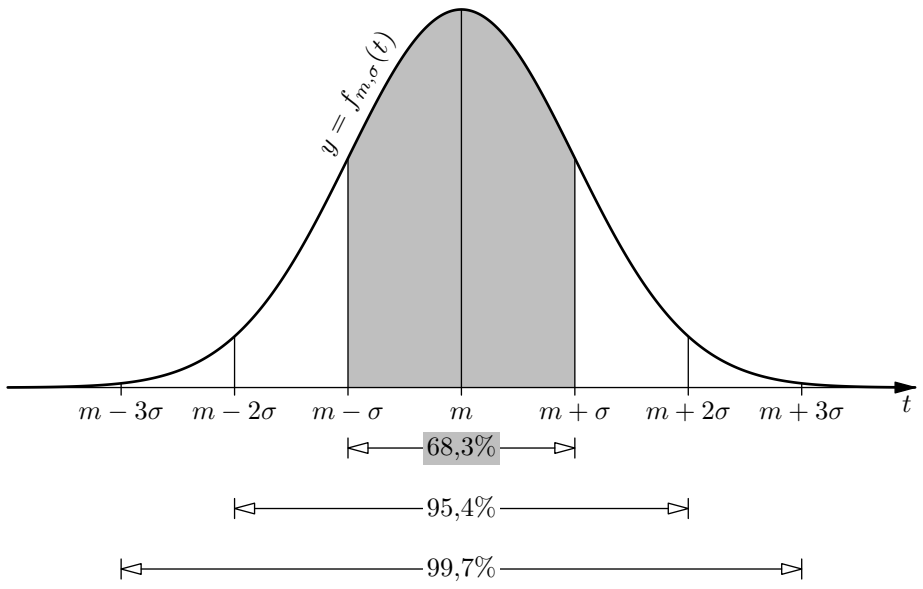


FIGURE 6.10 – Concentration de la loi $\mathfrak{N}(m,\sigma)$ autour de m

La figure 6.10 illustre la signification du paramètre de position m et du paramètre de dispersion σ pour la loi gaussienne $\mathfrak{N}(m, \sigma)$.
Cette concentration de pratiquement toute la probabilité dans l'intervalle $[m-3\sigma, m+3\sigma]$ permet l'utilisation des lois gaussiennes pour modéliser des grandeurs aléatoires qui *a priori* prennent leurs valeurs seulement dans un petit intervalle de $\mathbb{R}^+$: taille, poids, ..., même si théoriquement une variable gaussienne peut prendre toute valeur entre $-\infty$ et $+\infty$.

Il n'existe pas d'expression d'une primitive de la densité gaussienne $f_{m,\sigma}$ à l'aide des fonctions usuelles. Les valeurs de la fonction de répartition Φ de $\mathfrak{N}(0,1)$ sont tabulées, cf. page 520. D'après la proposition 6.51, ceci suffit pour calculer numériquement n'importe quelle f.d.r. de loi gaussienne.

6.4.4 Lois de Cauchy

Définition 6.53. *La variable aléatoire X suit la loi de Cauchy (ou loi de Cauchy de paramètres 0 et 1) si elle admet pour densité :*

$$f(t) = \frac{1}{\pi(1+t^2)}.$$

Notation : $X \sim \mathrm{Cau}(0,1)$.

Cette loi est *symétrique*, ce qui signifie que X et $-X$ ont même loi, ceci résultant ici de la parité de f. La fonction de répartition F est donnée par :

$$F(x) = \int_{-\infty}^{x} \frac{\mathrm{d}t}{\pi(1+t^2)} = \frac{1}{\pi}\Big(\frac{\pi}{2} + \arctan x\Big),$$

où $\arctan x$ est l'unique réel $y \in \left]-\pi/2, \pi/2\right[$ tel que $\tan y = x$.

Si $Y = a + bX$, avec X de loi $\mathrm{Cau}(0,1)$, $a \in \mathbb{R}$ et $b \in \mathbb{R}_+^*$, on dit encore que Y suit une loi de Cauchy, de paramètres (a,b), notation $Y \sim \mathrm{Cau}(a,b)$. La densité est alors

$$f_{a,b}(t) = \frac{1}{\pi b} \frac{1}{1 + \left(\frac{t-a}{b}\right)^2}.$$

6.5 Un exemple de loi continue sans densité

Voici un exemple de variable aléatoire ayant une fonction de répartition continue mais pas de densité.

On génère les chiffres décimaux d'un nombre réel X de $[0,1]$ en effectuant une suite infinie de tirages avec remise d'une boule dans une urne qui en contient 9 numérotées de 0 à 8. Autrement dit, X peut s'écrire :

$$X = \sum_{j=1}^{+\infty} \frac{X_j}{10^j}, \tag{6.58}$$

les variables aléatoires X_j étant indépendantes et de même loi uniforme sur $[\![0,8]\!]$. L'indépendance des variables aléatoires sera étudiée au chapitre 8. Pour comprendre ce qui suit, il suffit de raisonner en termes d'épreuves répétées pour disposer de l'indépendance de la famille d'évènements $(\{X_j = i_j\})_{j\in\mathbb{N}^*}$ pour toute suite d'entiers $(i_j)_{j\in\mathbb{N}^*}$.

Notons d'abord que la série de variables aléatoires (6.58) converge sur tout Ω. En effet, chaque $X_j(\omega)$ étant dans $[\![0,8]\!]$, la série $\sum_{j=1}^{+\infty} X_j(\omega)10^{-j}$ est à termes positifs et son terme général est majoré par 8×10^{-j} qui est le terme général d'une série géométrique de raison $1/10$ convergente.

Montrons que la fonction de répartition F de X est continue sur $\mathbb{R}$. Comme X est positive, $F(t) = 0$ pour tout $t < 0$. D'autre part, $X(\omega) \leqslant \sum_{j=1}^{+\infty} \frac{8}{10^j} = \frac{8}{9}$ pour tout $\omega\in\Omega$, d'où $F(\frac{8}{9}) = P(X \leqslant \frac{8}{9}) = P(\Omega) = 1$, donc F est constante égale à 1 sur $[\frac{8}{9}, +\infty[$. Pour prouver que F est continue sur $\mathbb{R}$, il nous suffit donc de vérifier que pour tout réel $x\in[0,\frac{8}{9}]$, $P(X = x) = 0$, autrement dit, que f n'a aucun saut dans cet intervalle, cf. prop. 6.19. Soit donc x quelconque dans cet intervalle et $(x_j)_{j\geqslant 1}$ la suite des chiffres de son développement décimal illimité propre (cf. exercice 2.3). Notons que si x admet un développement décimal illimité impropre (c'est-à-dire ne comportant que des 9 à partir d'un certain rang), ce développement ne peut être généré par la suite $(X_j(\omega))_{j\geqslant 1}$ puisqu'il n'y a aucune boule numérotée 9 dans l'urne. Nous avons ainsi pour tout $\omega\in\Omega$, équivalence entre « $X(\omega) = x$ » et « pour tout $j\geqslant 1$, $X_j(\omega) = x_j$ », ce qui se traduit par l'égalité d'évènements :

$$\{X = x\} = \bigcap_{j\in\mathbb{N}^*} \{X_j = x_j\}.$$

Notons $A_n = \bigcap_{j=1}^n \{X_j = x_j\}$. Pour tout $n\geqslant 1$, $\{X = x\}$ est inclus dans A_n, d'où $0 \leqslant P(X = x) \leqslant P(A_n)$. Or par indépendance des X_j,

$$P(A_n) = \prod_{j=1}^{n} P(X_j = x_j) \leqslant 9^{-n},$$

car $P(X_j = x_j)$ vaut $1/9$ si $x_j\in[\![0,8]\!]$ et 0 si $x_j = 9$. Par conséquent, $P(A_n)$ tend vers zéro quand n tend vers l'infini et en passant à la limite dans l'encadrement $0 \leqslant P(X = x) \leqslant P(A_n)$, on voit que $P(X = x) = 0$. Ceci établit la continuité de F.

Pour prouver que la loi de X n'a pas de densité, nous allons exploiter la proposition 6.31 en vérifiant que F n'est pas *absolument continue*. En prenant la négation de la propriété définissant l'absolue continuité, on voit qu'il s'agit de montrer qu'il existe un $\varepsilon > 0$ tels que *pour tout* $\delta > 0$, il existe une famille finie d'intervalles $[a_k, b_k]\subset\mathbb{R}$ $(k = 1,\dots,n)$ tels que les $]a_k, b_k[$ soient 2 à 2 disjoints, vérifiant :

$$\sum_{k=1}^{n} (b_k - a_k) < \delta \quad \text{et} \quad \sum_{k=1}^{n} |F(b_k) - F(a_k)| \geqslant \varepsilon. \tag{6.59}$$

Notons d'emblée qu'on ne peut pas espérer trouver un tel ε supérieur à 1 puisque $\sum_{k=1}^n |F(b_k) - F(a_k)| = \sum_{k=1}^n (F(b_k) - F(a_k)) = P\left(X\in\bigcup_{1\leqslant k\leqslant n}[a_k, b_k]\right)$ donc cette

somme d'accroissements de F ne peut excéder 1. Fixons un $\varepsilon \in]0,1[$, avec lequel nous allons vérifier (6.59).

Le point clé est de remarquer qu'il existe une famille dénombrable d'intervalles ouverts disjoints dans lesquels X est « interdit de séjour » et que la somme des longueurs de ces intervalles vaut 1. L'interdiction de séjour dans l'intervalle $]a,b[$ signifie ici précisément que $X(\omega)$ n'appartient à $]a,b[$ pour *aucun* $\omega \in \Omega$. Ceci implique notamment que $P(X \in]a,b[) = 0$. Or F étant continue, $P(X \in]a,b[) = F(b) - F(a)$ d'où $F(b) = F(a)$ et F est constante sur $[a,b]$. Tout intervalle $[a,b]$ de ce type correspond donc à un « palier horizontal » pour la représentation graphique de F. On « devine » que les intervalles suivants vont interdire le séjour de X : $]0{,}9;1[$, $]0{,}09;0{,}1[$, $]0{,}19;0{,}2[$, $]0{,}29;0{,}3[$, …, $]0{,}89;0{,}9[$, $]0{,}009;0{,}01[$, …, $]0{,}889;0{,}89[$, … Plus précisément, rappelons que la suite $(X_j(\omega))_{j\geqslant 1}$ ne peut reproduire aucun développement décimal illimité impropre. Les $x \in [0,1[$, pour lesquels il n'existe aucun $\omega \in \Omega$ tel que $X(\omega) = x$ sont ceux dont le développement décimal illimité propre comporte au moins un chiffre 9. Parmi eux, il y a tous les nombres décimaux dont la forme réduite ($m \times 10^{-k}$ avec m non divisible par 10) s'écrit avec une seule apparition de 9, pour le dernier chiffre, par exemple $x = 0{,}359 = 359 \times 10^{-3}$. Si d est l'un quelconque de ces décimaux, on peut donc l'écrire

$$d = c_1 10^{-1} + c_2 10^{-2} + \cdots + c_{k-1} 10^{-k+1} + 9 \times 10^{-k}, \quad c_i \in [\![0,8]\!], 1 \leqslant i < k. \quad (6.60)$$

Remarquons que

$$d + 10^{-k} = c_1 10^{-1} + c_2 10^{-2} + \cdots + c_{k-2} 10^{-k+2} + (1 + c_{k-1}) 10^{-k+1},$$

qui est encore une écriture décimale réduite puisque c_{k-1} étant inférieur ou égal à 8, il n'y a pas de propagation vers la gauche de la retenue en deça du rang $k-1$. Pour chaque $k \in \mathbb{N}^*$, notons E_k l'ensemble des décimaux d de la forme (6.60). Comme il y a 9 choix possibles (de 0 à 8) pour chacun des chiffres décimaux $c_1, \ldots c_{k-1}$, il est clair que $\operatorname{card} E_k = 9^{k-1}$. Notons $E = \bigcup_{k\geqslant 1} E_k$. Si d appartient à E, il appartient à *un seul* des E_k (parce que l'écriture (6.60) est une forme réduite).

On définit maintenant l'intervalle

$$I(d) =]d, d + 10^{-k}[.$$

Alors $\{I(d)\ ;\ d \in E\}$ est famille dénombrable d'intervalles deux à deux *disjoints*. Vérifions ce point clé. La dénombrabilité est claire puisque l'indexation par $d \in E$ est bijective et E est une partie infinie de l'ensemble dénombrable $\mathbb{Q}$. Définissons pour chaque $k \in \mathbb{N}^*$, la famille des intervalles :

$$J_{k,m} = \left]\frac{m}{10^k}, \frac{m+1}{10^k}\right[, \quad 0 \leqslant m < 10^k.$$

Les 10^k intervalles $J_{k,m}$ sont deux à deux disjoints. Parmi eux, les seuls intervalles de la forme $I(d)$ sont tous les $I(d)$, $d \in E_k$. Ceci nous montre déjà que si d et $d' \neq d$ sont dans le même E_k, $I(d)$ et $I(d')$ sont disjoints. Il ne reste donc qu'à vérifier que si d

et d' sont respectivement dans E_k et E_ℓ avec $\ell \neq k$, les intervalles $I(d)$ et $I(d')$ sont disjoints. Quitte à échanger les rôles de d et d', on peut toujours supposer que $k < \ell$. Conservant l'écriture (6.60) pour d, écrivons la décomposition analogue pour d' :

$$d' = c'_1 10^{-1} + \cdots + c'_k 10^{-k} + \cdots + 9 \times 10^{-\ell}, \quad c'_i \in [\![0, 8]\!], 1 \leqslant i < \ell.$$

De cette décomposition on déduit facilement que :

$$I(d') \subset J_{k,m'}, \quad m' = c'_1 10^1 + \cdots + c'_k 10^k.$$

Par (6.60), $I(d) = J_{k,m}$ avec $m = c_1 10^1 + \cdots + c_{k-1} 10^{k-1} + 9 \times 10^k$. Comme c'_k est différent de 9, $m \neq m'$ et ainsi $J_{k,m'}$ et $J_{k,m}$ sont disjoints d'où $I(d') \cap I(d) = \varnothing$.

Calculons la somme des longueurs de tous les $I(d)$:

$$\sum_{d \in E} \lambda_1(I(d)) = \sum_{k=1}^{+\infty} 9^{k-1} 10^{-k} = \frac{1}{10} \sum_{j=0}^{+\infty} \left(\frac{9}{10}\right)^j = \frac{1}{10} \frac{1}{1 - \frac{9}{10}} = 1. \tag{6.61}$$

Ce résultat qui peut sembler paradoxal à première vue [18] signifie que les intervalles d'interdiction de séjour occupent *presque* toute la place sur $[0, 1]$. Comme F est constante sur chacun d'entre eux, il ne reste plus à F qu'un espace de longueur nulle pour réaliser sa croissance de 0 à 1. C'est cette idée qui va nous permettre de mettre en défaut la propriété d'absolue continuité.

Rappelons que nous avons choisi de fixer un $\varepsilon \in]0, 1[$ et de vérifier (6.59) avec cet ε. Soit $\delta \in]0, 1[$ quelconque (on peut clairement se ramener à $\delta \in]0, 1[$ pour vérifier (6.59)). En raison de la convergence vers 1 de la série (6.61), il existe un entier ℓ dépendant de δ tel que $\sum_{i=1}^{\ell} 9^{i-1} 10^{-i} > 1 - \delta$. En notant

$$D_\ell = \bigcup_{i=1}^{\ell} \bigcup_{d \in E_i} I(d),$$

on a donc $\lambda_1(D_\ell) > 1 - \delta$. D_ℓ est une réunion finie d'intervalles ouverts deux à deux disjoints de $]0, 1[$. Son complémentaire dans $[0, 1]$ peut s'écrire comme une réunion finie d'intervalles fermés $[a_k, b_k]$, $1 \leqslant k \leqslant n$ (on bouche les trous). La somme des accroissements de F sur chacun des $I(d)$ et sur les $[a_k, b_k]$ vaut 1 car $F(1) - F(0) = 1$. L'accroissement de F sur chacun des $I(d)$ est nul puisque chaque $I(d)$ est un intervalle d'interdiction de séjour pour X. On voit ainsi que

$$\sum_{k=1}^{n} |F(b_k) - F(a_k)| = 1 > \varepsilon.$$

18. Attention à ne pas en conclure que $\bigcup_{d \in E} I(d) =]0, 1[$. Cette union est un ouvert dont le complémentaire dans $[0, 1]$ est un fermé *d'intérieur vide* car sinon cet intérieur contiendrait au moins un intervalle ouvert de longueur non nulle ce qui contredirait le fait que l'union des $I(d)$ et $[0, 1]$ ont même longueur.

D'autre part, par construction,

$$\lambda_1\left(\bigcup_{k=1}^{n}[a_k,b_k]\right)<\delta,$$

puisque son complémentaire dans $[0,1]$ est D_ℓ, réunion finie d'intervalles deux à deux disjoints, de longueur totale strictement supérieure à $1-\delta$. On a ainsi vérifié (6.59) et F n'est donc pas absolument continue.

6.6 Exercices

Ex. 6.1 *Interprétation du graphique d'une f.d.r.*

La variable aléatoire X a pour fonction de répartition F représentée figure 6.11.

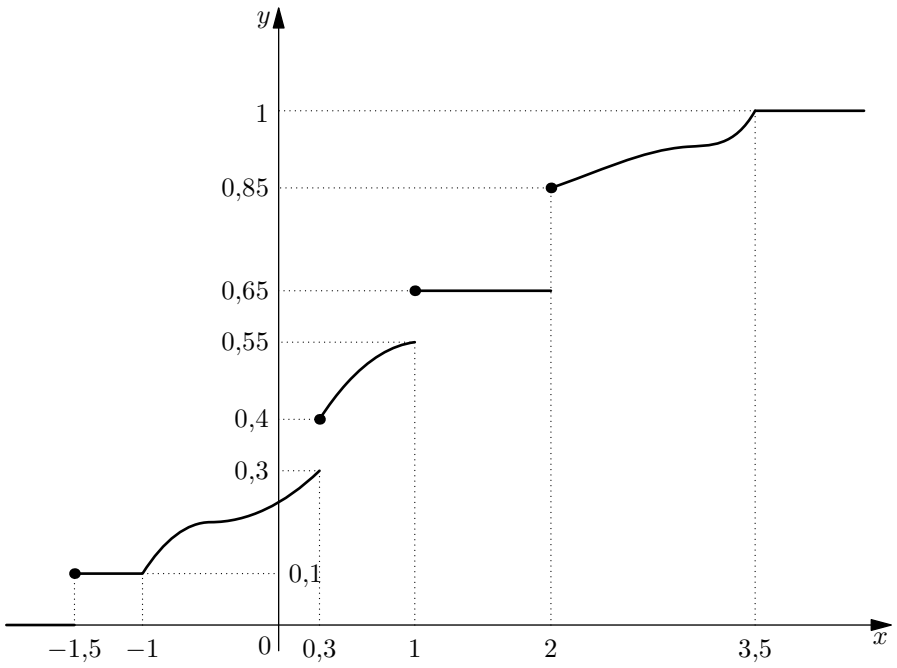


FIGURE 6.11 – Fonction de répartition F de la v.a. X

1) En exploitant les informations fournies par ce graphique, donner les valeurs des probabilités suivantes.

$$P(X\leqslant -1),\quad P(X=-1{,}5),\quad P(X=0{,}4),\quad P(X\geqslant 0{,}3),$$
$$P(X>2),\quad P(X\in[1;1{,}5]),\quad P(X\in]1;2]),\quad P(|X|>1).$$

2) La variable aléatoire X est-elle à densité ?

3) Calculer la somme des sauts de F. La variable aléatoire X est-elle discrète ?

Ex. 6.2 *Erreurs typographiques* ↓ 7.9

Un polycopié de 140 pages contient avant la première lecture, un nombre $n \geqslant 100$ d'erreurs typographiques réparties au hasard (on suppose que n est inférieur au nombre total de caractères que peut contenir une page).

1) Quelle est la loi exacte de la variable aléatoire Y_0 égale au nombre d'erreurs de la page 13 avant la première lecture ? Par quelle loi peut-on l'approximer ? On supposera dans les deux questions suivantes que Y_0 suit cette loi approchée.

2) On effectue des lectures successives de cette page. A chaque lecture les erreurs subsistantes peuvent être détectées (et donc corrigées) indépendamment les unes des autres, chacune avec une probabilité 1/3. On note Y_i le nombre d'erreurs restant sur la page 13 après la i-ème lecture. Calculer pour j et k entiers les probabilités conditionnelles $P(Y_1 = j \mid Y_0 = k)$ (distinguer selon que $j \leqslant k$ ou $j > k$).

3) En déduire la loi de Y_1, puis celle de Y_i.

Ex. 6.3 *Assurances maritimes et concentration du capital*

Une compagnie d'assurances assure une flotte de 500 navires de pêche valant chacun 1 million d'euros. Le risque assuré est la perte totale du navire qui est un événement de probabilité 0,001 pour une année. Les naufrages des différents navires sont considérés comme indépendants. On note X la variable aléatoire égale au nombre de navires perdus en une année.

1) Trouver la loi exacte de X.

2) Evaluer $P(X = 10)$ en utilisant une approximation de la loi de X.

3) La compagnie rembourse le 31 décembre, sur ses réserves, la valeur des bateaux assurés ayant fait naufrage dans l'année. À combien doivent s'élever ces réserves financières pour qu'elle puisse effectuer la totalité de ces remboursements avec une probabilité supérieure à 0,999 ? *Indication :* en utilisant l'approximation poissonnienne de la loi binomiale, montrer qu'il suffit que ces réserves représentent la valeur d'un très petit nombre de navires.

4) La compagnie fusionne avec une autre compagnie identique (500 navires assurés). Reprendre brièvement la question 3) et commenter le résultat obtenu.

Ex. 6.4 *Code de la Route I*

Pour un examen du type Code de la Route, les candidats doivent remplir un questionnaire de 40 questions en choisissant pour chacune d'elles l'une des 4 réponses proposées, dont une seule est exacte. Un candidat totalement ignorant décide de tenter sa chance en cochant complètement au hasard une réponse pour chaque question.

1) Quelle est la loi du nombre S de bonnes réponses du candidat ? Justifiez votre réponse.

2) Calculer $P(S \geqslant 36)$.

Solution p. 444

Ex. 6.5 *Code de la Route II*

Le modèle précédent est trop simpliste, en voici un plus réaliste. Le candidat n'est pas complètement ignorant et il arrive qu'il connaîsse la réponse à certaines questions. Il répond alors à coup sûr. S'il ignore la réponse, il choisit au hasard entre les 4 réponses proposées. On suppose toutes les questions indépendantes et que pour chacune de ces questions, la probabilité que le candidat connaisse la vraie réponse est p. Ce paramètre p mesure donc le vrai niveau du candidat. Malheureusement il n'est pas connu de l'examinateur. On définit les variables aléatoires suivantes :

S nombre de réponses connues du candidat ;

T nombre de bonnes réponses obtenues au hasard ;

U nombre total de bonnes réponses ($U = S + T$) ;

V nombre de mauvaises réponses ($S + T + V = 40$).

1) Quelle est la loi de S ?

2) On note B_i l'évènement « le candidat donne la bonne réponse à la $i^{\text{ième}}$ question ». Calculer $P(B_i)$ en fonction de p (on conditionnera par les deux cas possibles).

3) En déduire la loi de U, puis celle de V.

4) Utiliser une méthode analogue pour trouver la loi de T.

Solution p. 445

Ex. 6.6 *Épreuves répétées à trois issues*

On considère une suite de n épreuves répétées indépendantes avec pour chaque épreuve trois résultats possibles : A avec probabilité a, B avec probabilité b ou C avec probabilité c ($a + b + c = 1$). On notera respectivement A_i, B_i et C_i les événements obtention du résultat A (respectivement, B, C) à la i^{e} épreuve.

1) Dans cette question, $n = 5$. Quelle est la probabilité d'obtenir dans cet ordre 2 A suivis d'un B et de 2 C ? Quelle est celle d'obtenir (sans condition d'ordre) 2 A, 1 B et 2 C ?

2) Généraliser en montrant que la probabilité d'obtenir au cours des n épreuves (et sans condition d'ordre) i résultats A, j résultats B et k résultats C ($i + j + k = n$) vaut :

$$\frac{n!}{i!\,j!\,k!} a^i b^j c^k.$$

Solution p. 446

Ex. 6.7 *Code de la Route III*

On reprend les notations et les hypothèses de l'exercice 6.5. Ce que peut réellement observer l'examinateur, c'est la valeur prise par U (ou ce qui revient au même par V). Les quantités intéressantes pour tirer des conclusions sur le niveau réel du candidat sont les $P(S = i \mid U = m)$ pour $i \leqslant m \leqslant 40$. Par exemple si le candidat a obtenu 38 bonnes réponses, on aimerait évaluer $P(S \geqslant 36 \mid U = 38)$.

1) Expliquer pourquoi $P(S = i, U = m) = P(S = i, T = m - i, V = 40 - m)$, pour $i \leqslant m \leqslant 40$.

2) En utilisant le résultat de l'exercice 6.6, en déduire en fonction de i, m et p, l'expression de $P(S = i \mid U = m)$.

3) En déduire l'expression de $P(S \geqslant 36 \mid U = 38)$.

Solution p. 447

Ex. 6.8 *Loi triangulaire*

Soit f la fonction affine par morceaux, nulle en dehors de $]0, 2[$, représentée par la figure 6.12.

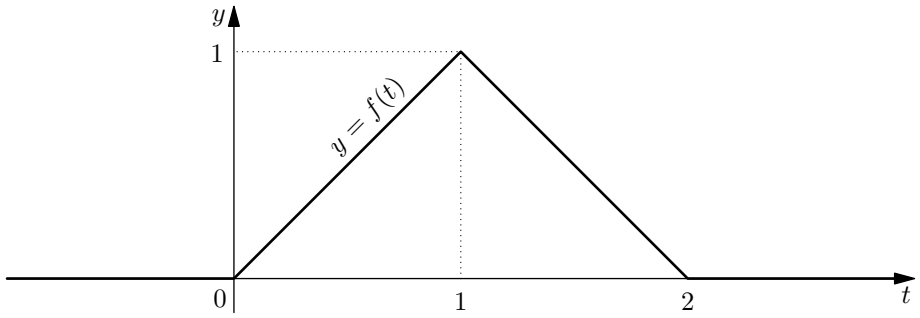


FIGURE 6.12 – Densité triangulaire f

1) Vérifiez que f est une densité de probabilité sur $\mathbb{R}$.

2) Soit Z une variable aléatoire positive de densité f (on ne vous demande pas d'en justifier l'existence) et F sa fonction de répartition. Donnez *sans calculer d'intégrale* les valeurs de $F(0)$, $F(1/2)$, $F(1)$ et $F(2)$. Expliquez comment vous les obtenez.

3) Calculez $\mathbf{E}Z$ en utilisant f, cf. le chapitre 7. Le résultat était-il prévisible ?

Solution p. 449

Ex. 6.9 *Une variable aléatoire tronquée*

Sur l'espace probabilisé $(\Omega, \mathcal{F}, P)$, X est une variable aléatoire de loi exponentielle. Soient a et b deux réels fixés $0 < a < b$. On définit sur le même espace $(\Omega, \mathcal{F}, P)$, la variable aléatoire Y par

$$\forall \omega \in \Omega, \quad Y(\omega) := \begin{cases} 0 & \text{si } X(\omega) \leqslant a, \\ X(\omega) & \text{si } a < X(\omega) \leqslant b, \\ b & \text{si } X(\omega) > b. \end{cases}$$

1) Dessinez l'allure de la représentation graphique de la fonction de répartition de Y en justifiant brièvement votre dessin.

2) La variable aléatoire Y est-elle à densité ? Si oui, calculez cette densité.

Solution p. 449

Ex. 6.10 *Un problème de tige brisée*

Dans tout l'exercice, X désigne une variable aléatoire à valeurs dans $]0, 1[$ et de loi uniforme sur $]0, 1[$. On pose

$$Z := \frac{1 - X}{X}.$$

1) Calculez explicitement la fonction de répartition de la variable aléatoire positive Z.

2) La loi de Z est-elle à densité ? Si oui, calculez la.

3) Expliquez si possible sans calcul pourquoi Z et $1/Z$ ont *même loi*.

4) On brise une tige de longueur 1 en choisissant au hasard le point de rupture suivant une loi uniforme sur $]0,1[$. On notera X la longueur du morceau gauche. Quelle est la probabilité que l'un des deux morceaux soit plus de deux fois plus long que l'autre ?

Ex. 6.11 *Un peu de trigonométrie aléatoire*

On définit une variable aléatoire X grâce à la construction représentée à la figure 6.13. L'angle $\left(\overrightarrow{AO}, \overrightarrow{AM}\right)$ a pour mesure en radians U, *variable aléatoire* de loi uniforme sur $]-\pi/2, \pi/2[$. La distance AO vaut 1 et X est l'abscisse du point M sur la droite de repère $(O, \vec{i}\,)$: $\overrightarrow{OM} = X\vec{i}$. Les angles sont orientés dans le sens trigonométrique, la figure est faite pour U positif. Pour $U = 0$, M coïncide avec le point O et pour $-\pi/2 < U < 0$, M est « à gauche » de O.

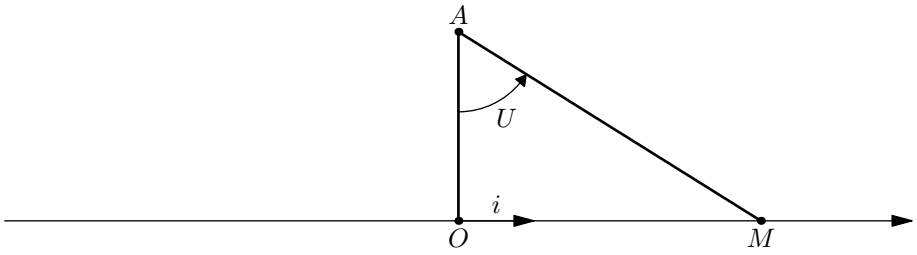


FIGURE 6.13 – Construction de M

1) Exprimez X en fonction de U.

2) Pour x réel, calculez $P(X \leqslant x)$. On obtient ainsi la fonction de répartition de la variable aléatoire réelle X.

3) Expliquez pourquoi la loi de X est à densité et calculez cette densité. Que reconnaissez vous ainsi ?

4) Quelle est la valeur de $P(|X| \leqslant 1)$? Cette question peut se résoudre avec ou sans l'aide des précédentes.

Solution p. 451

Ex. 6.12 On note $[x]$ la partie entière du réel x. Soit f la fonction :

$$f : x \longmapsto \frac{1}{(1+x)\ln 2}\mathbf{1}_{]0,1[}(x).$$

1) Vérifier que f est une densité de probabilité. Expliciter la fonction de répartition associée.

2) Soit X une variable aléatoire réelle dont la loi a pour densité f. Calculer :

$$P\left(k \leqslant \frac{1}{X} \leqslant k+t\right) \qquad k \in \mathbb{N}^*,\ t \in]0,1[$$

3) Soit $Y = \frac{1}{X} - \left[\frac{1}{X}\right]$. Montrer que Y est une variable aléatoire réelle de même loi que X. *Indication* : calculer sa fonction de répartition.

Solution p. 453

Ex. 6.13 *Une somme de deux v.a. à densité qui n'est ni discrète ni à densité*

Soit X une variable aléatoire de loi uniforme sur $[0,1]$ et Y la variable aléatoire définie sur le même espace Ω par

$$Y(\omega) := \begin{cases} X(\omega) & \text{si} \quad X(\omega) \in [0,1/4] \cup [3/4,1]; \\ 1 - X(\omega) & \text{si} \quad X(\omega) \in]1/4,3/4[. \end{cases}$$

1) Quelle est la loi de Y ?

2) Trouver la fonction de répartition de la variable aléatoire $Z := X + Y$ et vérifier que Z n'est ni discrète ni à densité.

Ex. 6.14 *Somme d'une v.a. entière et d'une v.a. uniforme sur* $]0,1[$

Soient N et U deux variables aléatoires définies sur le même espace probabilisé, N étant à valeurs dans $\mathbb{N}$ et U de loi uniforme sur $]0,1[$. On note $X := N + U$.

1) Montrez que la fonction de répartition de X est continue sur $\mathbb{R}$.

2) On suppose de plus que N et U sont indépendantes [19]. Montrez qu'alors la fonction de répartition de X est affine par morceaux. La loi de X est-elle à densité ? Si oui, exprimez cette densité à l'aide de la loi de N.

Ex. 6.15 *Continuité uniforme d'une f.d.r. continue*

Montrez que si F est une fonction de répartition continue sur $\mathbb{R}$, elle est aussi *uniformément* continue sur $\mathbb{R}$.

Ex. 6.16 *Une v.a. série*

Soit $(\Omega, \mathcal{F}, P)$ un espace probabilisé sur lequel existe une suite $(A_k)_{k \geqslant 1}$ d'évènements indépendants et de même probabilité $p \in]0,1[$. On pose

$$X := \sum_{k=1}^{+\infty} \frac{\mathbf{1}_{A_k}}{k^2}.$$

1) Expliquez pourquoi cette formule définit une variable aléatoire positive et bornée.

2) On se propose dans tout ce qui suit de montrer que X a une fonction de répartition continue, ce qui revient à montrer que $P(X = x)$ vaut zéro pour tout $x \in \mathbb{R}$ (pourquoi ?). On introduit à cet effet l'ensemble :

$$D := \left\{ x \in \mathbb{R};\ \exists (c_k)_{k \geqslant 1} \in \{0,1\}^{\mathbb{N}^*}, \quad x = \sum_{k=1}^{+\infty} \frac{c_k}{k^2} \right\},$$

19. L'indépendance de deux variables aléatoires est définie p. 312. Pour résoudre cette question, il suffit de savoir que l'indépendance de N et U implique l'indépendance des deux évènements $\{N = j\}$ et $\{U \leqslant t\}$ pour tous $j \in \mathbb{N}$ et $t \in \mathbb{R}$.

autrement dit, D est l'ensemble des valeurs prises par $X(\omega)$ pour $\omega \in \Omega$. Que vaut $P(X = x)$ pour $x \notin D$?

3) En utilisant une comparaison série intégrale, montrez l'inégalité *stricte* :

$$\forall n \geqslant 1, \quad \sum_{k=n+1}^{+\infty} \frac{1}{k^2} < \frac{1}{n}. \tag{6.62}$$

Il est vivement conseillé d'illustrer votre réponse par un dessin.

4) L'inégalité (6.62) permet de démontrer *l'unicité* de la suite binaire $(c_k)_{k\geqslant 1}$ dans la représentation d'un $x \in D$. Pour le voir, raisonnez par l'absurde en supposant qu'il existe deux suites binaires $(b_k)_{k\geqslant 1}$ et $(c_k)_{k\geqslant 1}$ distinctes telles que

$$\sum_{k=1}^{+\infty} \frac{b_k}{k^2} = \sum_{k=1}^{+\infty} \frac{c_k}{k^2},$$

notez n le plus petit entier i tel que $b_i \neq c_i$ et remarquez que pour tout k, $|b_k - c_k|$ a un majorant simple, ...

5) En utilisant l'unicité établie ci-dessus, montrez que pour $x \in D$, $\{X = x\}$ peut s'écrire $\bigcap_{k\geqslant 1} B_k$, les évènements B_k étant indépendants. Trouvez une constante $r \in]0, 1[$ telle que $P(B_k) \leqslant r$ pour tout $k \geqslant 1$. Déduisez-en *proprement* que $P(X = x) = 0$.

Ex. 6.17 *Une loi discrète pathologique* ↓ 7.12

Le but de cet exercice est de construire et d'étudier une v.a. discrète ayant pour ensemble $X(\Omega)$ de valeurs possibles l'ensemble $\mathbb{D}$ des nombres *décimaux* de $[0, 1[$. Un nombre décimal peut se représenter par une suite *finie* de *chiffres décimaux*. Cette représentation n'est pas unique, par exemple :

$$0{,}375 = 0{,}375\,0 = 0{,}375\,00 = 0{,}375\,000 = \ldots$$

On peut aussi l'écrire sous la forme $k \times 10^{-n}$ avec $k \in \mathbb{N}$ et $n \in \mathbb{N}^*$. Si k n'est pas divisible par 10, nous dirons que l'on a la forme réduite (pour l'exemple précédent, $k = 375$ et $n = 3$). Nous construisons X par la procédure suivante.

On dispose de deux urnes. La première contient des boules rouges et des vertes. On note p la proportion de boules rouges ($0 < p < 1$) et q celle des vertes. La deuxième urne contient 10 boules numérotées de 0 à 9. On effectue des tirages avec remise dans la première urne jusqu'à la première apparition d'une rouge. On note N le nombre (aléatoire) de tirages nécessaires. Une fois connue la valeur de N, on effectue N tirages avec remise d'une boule dans la deuxième urne. En notant Y_j le *chiffre* sorti lors du j-ème tirage dans la deuxième urne ($i \leqslant N$), on forme le nombre décimal :

$$X(\omega) = 0, Y_1(\omega)Y_2(\omega)\ldots Y_{N(\omega)}(\omega) = \sum_{j=1}^{N(\omega)} \frac{Y_j(\omega)}{10^j}.$$

1) Quelle est la loi de N ?

2) Soit n fixé. Lorsque l'on effectue n tirages dans la deuxième urne, quelle est la probabilité d'obtenir une suite de n chiffres décimaux choisie à l'avance ?

3) Calculer $P(X = 0{,}375)$. Attention, il y a une infinité de façons d'obtenir cette valeur !

4) Soit $d \in \mathbb{D}$ et $k \times 10^{-n}$ sa forme réduite. Calculer $P(X = d)$.

5) Vérifier que $\mathbb{D}$ est dénombrable en décrivant un procédé permettant de *numéroter* bijectivement les éléments de $\mathbb{D}$ par ceux de $\mathbb{N}$.

6) Montrer qu'il n'existe pas de numérotation *croissante* des éléments de $X(\Omega)$ (c.-à-d. vérifiant $\forall k \in \mathbb{N}$, $x_k < x_{k+1}$).

7) Soit F la fonction de répartition de X. Montrer que F n'est constante dans aucun intervalle de $[0, 1[$ aussi petit soit-il. Ce n'est donc pas une fonction en escaliers (elle est constante sur $]-\infty, 0[$ et sur $[1, +\infty[$).

8) Montrer que F n'est continue sur aucun intervalle de $[0, 1[$, aussi petit soit-il. Cependant F est continue en tout réel non décimal de $[0, 1[$ (il y en a une infinité non dénombrable).

9) Détailler le calcul de $P(X \leqslant 0{,}375 \mid N = j)$ en distinguant les deux cas $j < 3$ et $j \geqslant 3$.

10) Généraliser en montrant que :

$$\forall d \in \mathbb{D}, \qquad P(X \leqslant d \mid N = j) = \frac{\lfloor 10^j d \rfloor + 1}{10^j}.$$

11) Calculer $F(k/10)$ pour $k \in [\![0, 9]\!]$ et $F(l/100)$ pour $l \in [\![0, 99]\!]$.

12) Peut-on donner une formule générale pour $F(d)$, $d \in \mathbb{D}$?

Solution p. 455

Chapitre 7

Espérance

7.1 Introduction

L'ESPÉRANCE d'une variable aléatoire est, lorsqu'elle existe, la *moyenne des valeurs de cette variable, pondérées par leurs probabilités de réalisation.* On voit bien comment traduire cette définition informelle dans le cas d'une variable aléatoire discrète X en posant :

$$\mathbf{E}X := \sum_{x \in X(\Omega)} xP(X = x). \tag{7.1}$$

Cette formule n'a de sens que si la famille de réels $\{xP(X = x) \ ; \ x \in X(\Omega)\}$ est sommable, ce qui se traduit par la condition suivante pour l'existence de l'espérance de la v.a. discrète X :

$$\sum_{x \in X(\Omega)} |x|P(X = x) < +\infty. \tag{7.2}$$

Tant que l'on reste dans le cadre des variables aléatoires discrètes, cette définition est satisfaisante et permet d'établir toutes les propriétés de l'espérance [14, Chap. 5]. En bonne place parmi ces propriétés, figure *l'additivité* de l'espérance : si X et Y définies sur le même $(\Omega, \mathcal{F}, P)$ ont une espérance, il en va de même pour $X + Y$ et

$$\mathbf{E}(X + Y) = \mathbf{E}X + \mathbf{E}Y. \tag{7.3}$$

Essayons de traduire la définition informelle ci-dessus dans le cas d'une variable aléatoire à densité f. Partant de (7.1), on remplace $P(X = x)$ par $P(X \in [x, x + \mathrm{d}x])$, probabilité « valant [1] $f(x)\,\mathrm{d}x$ » et on remplace la somme (ou série) par une intégrale, ce qui conduit à :

$$\mathbf{E}X := \int_{-\infty}^{+\infty} xf(x)\,\mathrm{d}x, \tag{7.4}$$

1. Nous ne prétendons pas donner un sens rigoureux à cette probabilité d'appartenance à un « intervalle infinitésimal », il s'agit juste d'une approche intuitive.

la condition d'existence de l'espérance étant tout simplement la convergence absolue de cette intégrale généralisée, ce qui vu la positivité de f, se traduit par

$$\int_{-\infty}^{+\infty} |x| f(x)\,\mathrm{d}x < +\infty. \tag{7.5}$$

Cette définition malgré son analogie formelle avec (7.1) est loin d'offrir la même souplesse pour établir les propriétés de l'espérance. Par exemple la preuve de l'additivité est *complètement hors de portée*. En effet, si X et Y sont à densité, $X+Y$ peut n'être ni discrète ni à densité [2], cf. l'exercice 6.13 pour un exemple, et alors le premier membre de (7.3) n'est même pas défini pour la v.a. $Z = X + Y$.

La solution donnée à ce problème par la théorie moderne des probabilités est la définition dans le cas général, de l'espérance de X comme une intégrale abstraite sur Ω, relativement à la mesure P :

$$\mathbf{E}X := \int_\Omega X(\omega)\,\mathrm{d}P(\omega), \quad \text{si} \int_\Omega |X(\omega)|\,\mathrm{d}P(\omega) < +\infty. \tag{7.6}$$

On peut donner une première idée de ce qu'est cette intégrale abstraite en considérant le cas d'une variable aléatoire X telle que $X(\Omega) = \{x_1, \ldots, x_n\}$. Alors en notant $A_k := \{X = x_k\} = \{\omega \in \Omega\,;\ X(\omega) = x_k\}$,

$$\int_\Omega X(\omega)\,\mathrm{d}P(\omega) = \sum_{k=1}^{n} x_k P(A_k), \tag{7.7}$$

ce qui traduit bien la définition informelle de $\mathbf{E}X$ comme la moyenne des valeurs de X pondérées par leurs probabilités de réalisation. Le passage au cas d'une variable aléatoire X quelconque revient précisément à construire une intégrale au sens de Lebesgue sur $(\Omega, \mathcal{F}, P)$ et cette théorie sort du cadre de ce livre.

Il nous faut donc trouver une autre définition de $\mathbf{E}X$. Cette définition doit permettre un traitement unifié de toutes les lois [3]. Rappelons qu'il existe des lois qui ne sont ni discrètes ni à densité et que la description la plus générale des lois de variables aléatoires réelles est donnée par leur fonction de répartition, cf. le théorème 5.30 et la remarque 6.17. Il est donc naturel de chercher à définir $\mathbf{E}X$ *à partir de la fonction de répartition* $F : t \mapsto P(X \leqslant t)$. Nous allons motiver cette définition en nous restreignant au cas des variables aléatoires positives et en partant du cas simple où X est discrète avec $X(\Omega) = \{x_1, \ldots, x_n\}$ partie *finie* de $\mathbb{R}_+$. Dans ce cas, la définition informelle de $\mathbf{E}X$ se traduit par la formule $\mathbf{E}X = \sum_{k=1}^{n} x_k P(X = x_k)$. Les figures 7.1 et 7.2 nous montrent comment exprimer cette moyenne pondérée à l'aide de F. Rappelons que dans ce cas, F présente en chaque x_k un saut d'amplitude $P(X = x_k)$. L'interprétation graphique en terme d'aires donnée par la figure 7.2 nous permet d'écrire $\mathbf{E}X$ comme l'intégrale de Riemann ordinaire : $\mathbf{E}X = \int_0^{x_n} (1 - F(t))\,\mathrm{d}t$ et aussi comme la fausse intégrale généralisée $\int_0^{+\infty} (1 - F(t))\,\mathrm{d}t$.

2. Alors que la somme de deux variables aléatoires discrètes est *toujours* une variable aléatoire discrète.

3. La définition informelle de $\mathbf{E}X$ nous fait pressentir que $\mathbf{E}X$ ne doit dépendre que de la loi de X, ce qui est bien le cas dans les formules (7.1) et (7.4).

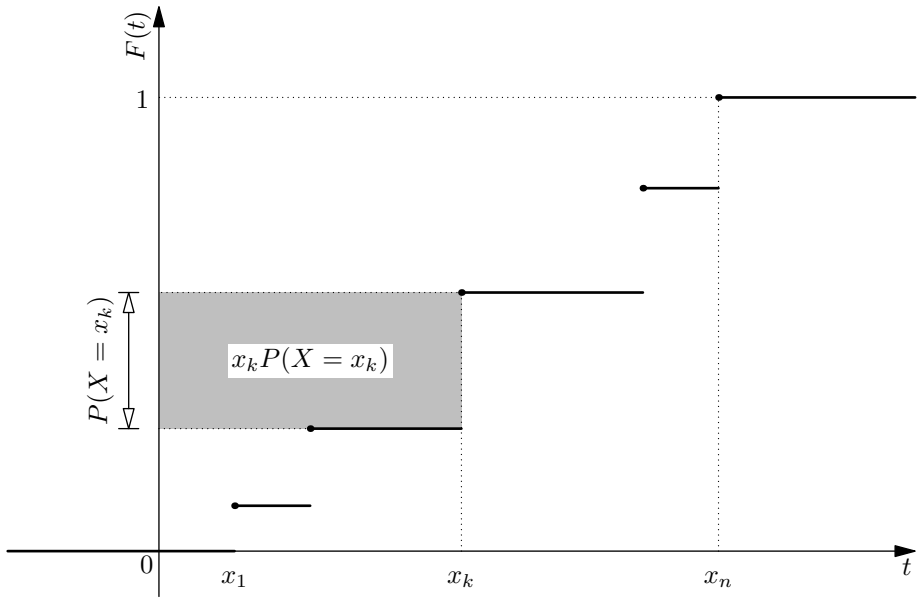


FIGURE 7.1 – Interprétation graphique des $x_k P(X = x_k)$, pour $x_k \geqslant 0$

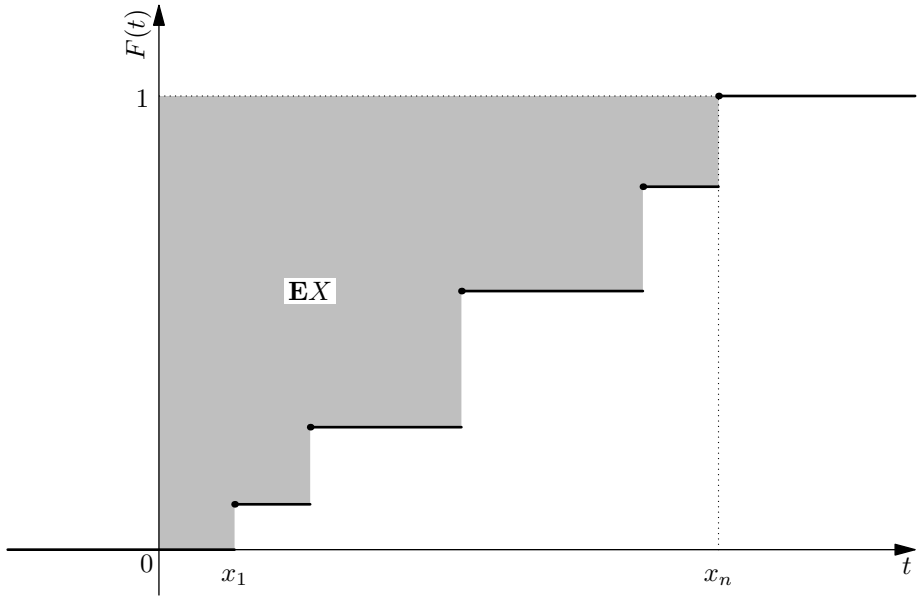


FIGURE 7.2 – Interprétation graphique de $\mathbf{E}X = \sum_{k=1}^{n} x_k P(X = x_k)$, les $x_k \geqslant 0$.

Si on passe maintenant au cas d'une variable aléatoire positive quelconque, il paraît alors naturel de considérer que $\mathbf{E}X$ est l'aire (éventuellement infinie) délimitée par le segment vertical $t = 0$, $y \in [0, 1]$, la demi droite « asymptote » $y = 1$, $t \geqslant 0$ et le graphe de F, ce qui nous conduit à la formule

$$\mathbf{E}X := \int_0^{+\infty} (1 - F(t))\,\mathrm{d}t = \int_0^{+\infty} P(X > t)\,\mathrm{d}t, \quad \text{pour toute v.a. positive } X.$$

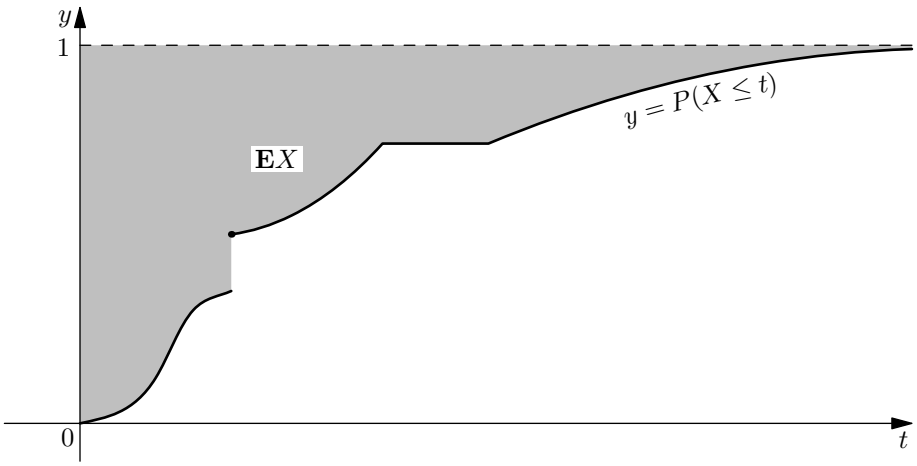


FIGURE 7.3 – Interprétation graphique de $\mathbf{E}X$ *via* la f.d.r. de X v.a. positive.

Nous verrons que cette définition permet d'établir en toute généralité les propriétés de l'espérance. Bien sûr, nous devrons retrouver à partir de cette définition, les formules (7.1) et (7.4) pour X discrète ou à densité.

7.2 Espérance d'une variable aléatoire positive

Dans toute la suite de ce chapitre, on fixe un espace probabilisé $(\Omega, \mathcal{F}, P)$. Toutes les variables aléatoires considérées seront, sauf mention explicite du contraire, définies sur cet espace et leur loi sera la loi sous P.

Définition 7.1 (espérance d'une v.a. positive)**.** *Soit X une variable aléatoire positive*[4] *sur $(\Omega, \mathcal{F})$. On appelle* espérance *de X (ou espérance de X sous P) la quantité*

$$\mathbf{E}X := \int_0^{+\infty} P(X > t)\,\mathrm{d}t, \tag{7.8}$$

qui est un élément de $\overline{\mathbb{R}}_+$.

4. C'est-à-dire une application $\Omega \to \mathbb{R}_+$, mesurable $\mathcal{F}$ - $\mathrm{Bor}(\mathbb{R}_+)$.

Pour justifier l'existence de $\mathbf{E}X$, on commence par noter que l'application $G : \mathbb{R}_+ \to [0,1]$, $t \mapsto G(t) := P(X > t)$ est *décroissante* sur $\mathbb{R}_+$, donc *Riemann intégrable* sur $[0,b]$ pour tout $b \in \mathbb{R}_+^*$, cf. proposition 3.10. L'intégrale $\int_0^b G(t)\,\mathrm{d}t = \int_0^b P(X > t)\,\mathrm{d}t$ existe donc bien et est un réel positif pour tout b. Comme c'est une fonction croissante de sa borne supérieure b, elle converge dans $\overline{\mathbb{R}}_+$ quand b tend vers $+\infty$.

Dans cette section, nous utiliserons l'interprétation graphique de $\mathbf{E}X$ *via* la fonction de survie $t \mapsto P(X > t)$, cf. figure 7.4, plutôt que *via* la f.d.r. $F : t \mapsto P(X \leqslant t)$. On passe évidemment d'une représentation à l'autre en effectuant une symétrie orthogonale par rapport à la droite $y = 1/2$, puisque $G = 1 - F$. Cette symétrie conserve les aires, cf. prop. 5.14.

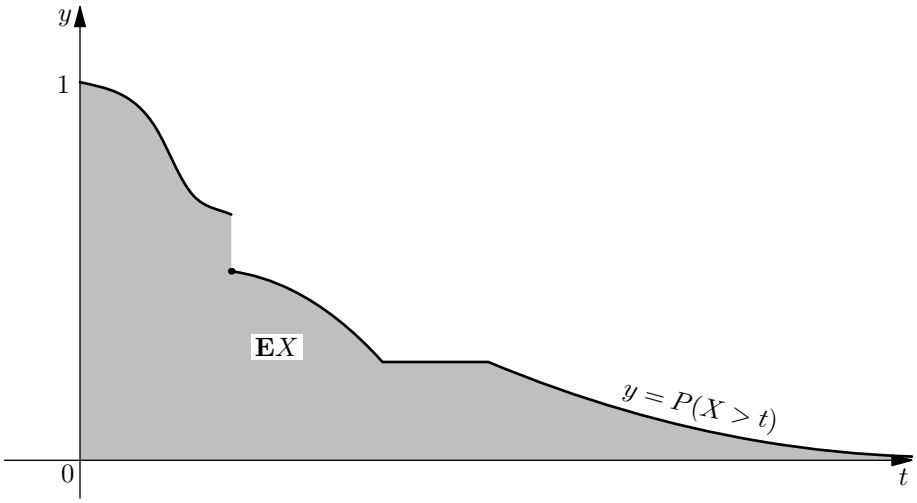


FIGURE 7.4 – Interprétation graphique de $\mathbf{E}X$ *via* la fonction de survie de X v.a. positive.

Remarque 7.2. $\mathbf{E}X$ ne dépend que de la loi de X, il serait donc plus correct de parler de l'espérance de la loi de X sous P au lieu de l'espérance de X. L'usage donne néanmoins la préférence à cette dernière appellation *quand il n'y a pas d'ambiguïté sur P.*

Remarque 7.3 (espérance d'une v.a. presque sûrement positive)**.** Dans les exercices, la variable aléatoire X n'est pas toujours donnée explicitement, il arrive assez souvent que l'on ne connaisse que sa loi P_X. Si $P_X(\mathbb{R}_+) = 1$, on s'autorisera une généralisation de la définition 7.1 en considérant que la formule (7.8) reste valable. Il s'agit bien d'une généralisation car on peut avoir $P(X \geqslant 0) = 1$ sans que $X(\omega)$ soit positif ou nul pour tout ω, par exemple si $\Omega = \mathbb{R}$, $\mathcal{F} = \mathrm{Bor}(\mathbb{R})$, P est la loi uniforme sur $[0,1]$ et $X : \omega \mapsto \omega$ est l'identité sur $\mathbb{R}$. On a alors $X(\omega) < 0$ pour une infinité non dénombrable de ω et $P(X \geqslant 0) = 1$. Cette généralisation est cohérente avec la

remarque 7.2 car si X est une variable aléatoire réelle définie sur un espace probabilisé $(\Omega, \mathcal{F}, P)$ et telle que $P(X \geqslant 0) = 1$, on peut toujours trouver un espace probabilisé $(\Omega', \mathcal{F}', P')$ et une variable aléatoire positive X' définie sur cet espace tels que X et X' aient même loi. Il suffit de prendre $\Omega' = \mathbb{R}_+$, $\mathcal{F}' = \mathrm{Bor}(\mathbb{R}_+)$, $P' = P_X$ et $X' : \mathbb{R}_+ \to \mathbb{R}$, $\omega \mapsto \omega$ égale à l'identité sur $\mathbb{R}_+$. On a alors pour tout borélien B de $\mathbb{R}$, $P'_{X'}(B) = P'(X' \in B) = P'(\{\omega \in \Omega' \ ; \ X'(\omega) \in B\} = P'(B \cap \mathbb{R}_+) = P_X(B \cap \mathbb{R}_+) = P_X(B)$ car $P_X(B \cap \mathbb{R}_-^*) \leqslant P_X(\mathbb{R}_-^*) = 0$. Ceci montre que X et X' ont même loi [5].

Définition 7.4 (intégrabilité d'une v.a. positive)**.** *On dit que la variable aléatoire positive X est* intégrable *si*

$$\int_0^{+\infty} P(X > t)\,\mathrm{d}t < +\infty. \tag{7.9}$$

Exemple 7.5. Si la variable aléatoire positive X est bornée, c.-à-d. s'il existe une constante c telle que pour tout $\omega \in \Omega$, $0 \leqslant X(\omega) \leqslant c$, alors elle est intégrable. En effet pour $t \geqslant c$, $P(X > t) = 0$, ce qui réduit l'intégrale généralisée définissant $\mathbf{E}X$ à une intégrale de Riemann ordinaire $\int_0^c P(X > t)\,\mathrm{d}t$ donc finie (et majorée par c).

Plus généralement, si la loi de X, v.a. positive, vérifie $P(X > t) \leqslant Ct^{-\alpha}$ pour un certain $\alpha > 1$ et tout $t \geqslant t_0 > 0$, ou si $P(X > t) \leqslant t^{-1}(\ln t)^{-\beta}$ pour un $\beta > 1$ et tout $t \geqslant t_0 > 0$, alors X est intégrable. Réciproquement, l'intégrabilité de X nous donne un renseignement sur la vitesse de convergence [6] vers 0 de $P(X > t)$ quand t tend vers $+\infty$. C'est l'inégalité de Markov que nous verrons ci-dessous (proposition 7.16).

Voyons maintenant quelques exemples simples de calcul d'espérance de variables aléatoires positives.

Exemple 7.6 (espérance d'une constante positive)**.** Si la variable aléatoire X est une constante positive c, c.-à-d. $X(\omega) = c$ pour tout $\omega \in \Omega$, alors $\mathbf{E}X = c$. En effet :

$$P(X > t) = \begin{cases} 1 & \text{si } t < c \\ 0 & \text{si } t \geqslant c \end{cases} = \mathbf{1}_{]-\infty, c[}(t),$$

d'où

$$\mathbf{E}X = \int_0^{+\infty} \mathbf{1}_{]-\infty, c[}(t)\,\mathrm{d}t = \int_0^c \mathbf{1}_{[0, c[}(t)\,\mathrm{d}t = \int_0^c 1\,\mathrm{d}t = c.$$

L'exemple suivant est d'une grande importance car il permet d'écrire toute probabilité d'évènement comme une espérance. Nous le formulons sous forme de proposition.

Proposition 7.7 (espérance d'une indicatrice d'évènement)**.** *Pour tout évènement $A \in \mathcal{F}$,*

$$\mathbf{E}\big(\mathbf{1}_A\big) = P(A). \tag{7.10}$$

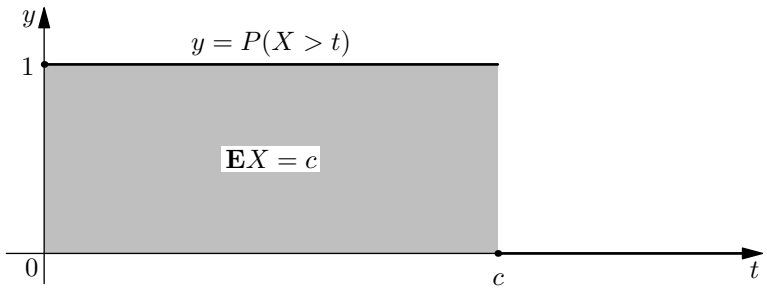


FIGURE 7.5 – Espérance de la v.a. constante $X = c$

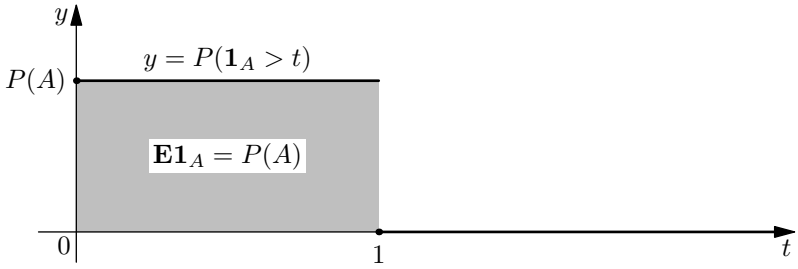


FIGURE 7.6 – Espérance de la v.a. indicatrice $X = \mathbf{1}_A$

Preuve. La variable aléatoire positive $\mathbf{1}_A$ prend la valeur 1 sur l'évènement A et 0 sur A^c, elle suit la loi de Bernoulli de paramètre $p = P(A)$. L'évènement $\{\mathbf{1}_A > t\}$ est donc égal à A si $0 \leqslant t < 1$ et à l'ensemble vide si $t \geqslant 1$. On en déduit que

$$P(\mathbf{1}_A > t) = \begin{cases} P(A) & \text{si } 0 \leqslant t < 1, \\ 0 & \text{si } t \geqslant 1. \end{cases}$$

Par conséquent,

$$\mathbf{E}\big(\mathbf{1}_A\big) = \int_0^1 P(A)\,\mathrm{d}t = P(A).$$

□

Dans ce chapitre, les variables aléatoires discrètes X ne prenant qu'un nombre fini de valeurs jouent un rôle important car elles vont nous permettre d'établir par passage à la limite les principales propriétés de l'espérance. Il est commode de les dénommer comme suit.

5. La tribu borélienne de $\mathbb{R}_+$ est la plus petite tribu contenant tous les ouverts de $\mathbb{R}_+$. On peut vérifier qu'un sous-ensemble B' de $\mathbb{R}_+$ est un borélien de $\mathbb{R}_+$ si et seulement s'il s'écrit $B \cap \mathbb{R}_+$ où B est un borélien de $\mathbb{R}$.

6. Pour n'importe quelle variable aléatoire X, $P(X > t)$ tend vers 0 quand t tend vers $+\infty$, car $P(X > t) = 1 - F(t)$, où F est la f.d.r. de X qui tend toujours vers 1 en $+\infty$.

Définition 7.8 (variable aléatoire simple)**.** *On dit que la variable aléatoire réelle X définie sur $(\Omega, \mathcal{F})$ est* simple *ou* étagée *si $X(\Omega)$ est fini. En notant $X(\Omega) = \{x_1, \dots, x_n\}$, les x_i étant tous distincts, X admet la décomposition*

$$X = \sum_{k=1}^{n} x_k \mathbf{1}_{A_k}, \quad \text{où } A_k := \{X = x_k\},\ k \in [\![1, n]\!], \tag{7.11}$$

les évènements A_k formant une partition de Ω.

Proposition 7.9 (espérance d'une v.a. positive simple)**.** *Si X est une variable aléatoire positive simple avec $X(\Omega) = \{x_1, \dots, x_n\}$,*

$$\mathbf{E}X = \sum_{k=1}^{n} x_k P(X = x_k). \tag{7.12}$$

On retrouve ainsi la formule (7.1) de l'introduction dans le cas particulier où $X(\Omega)$ est fini ; voir aussi (7.7).

Preuve. Notons en préliminaire qu'il nous faut résister ici à la tentation de dire « c'est immédiat en utilisant la décomposition (7.11), la proposition 7.7 et la linéarité de l'espérance », car *nous n'avons pas encore prouvé* que l'espérance est linéaire. En fait la proposition 7.9 est l'un des ingrédients de la preuve de la linéarité de l'espérance. Il nous faut donc vérifier (7.12) par un *calcul direct* basé sur la définition 7.1.

Quitte à réindexer, on peut toujours supposer que les x_k sont rangés par ordre croissant. Notons $p_i := P(X = x_i)$ et $s_k := \sum_{1 \leqslant i \leqslant k} p_i$. La fonction de répartition F peut alors s'écrire (cf. la proposition 6.20) :

$$F(t) = \sum_{k=1}^{n-1} s_k \mathbf{1}_{[x_k, x_{k+1}[}(t) + s_n \mathbf{1}_{[x_n, +\infty[}(t). \tag{7.13}$$

Notons que pour $t \geqslant x_n$, $F(t) = s_n = 1$, donc $P(X > t) = 1 - F(t) = 0$. Ainsi $\mathbf{E}X = \int_0^{+\infty} P(X > t)\,\mathrm{d}t = \int_0^{x_n} P(X > t)\,\mathrm{d}t$. On peut alors calculer $\mathbf{E}X$ en utilisant la décomposition (7.13) et les propriétés de l'intégrale de Riemann sur l'intervalle fermé borné $[0, x_n]$:

$$\mathbf{E}X = \int_0^{x_n} (1 - F(t))\,\mathrm{d}t = x_n - \int_0^{x_n} F(t)\,\mathrm{d}t = x_n - \sum_{k=1}^{n-1} \int_{x_k}^{x_{k+1}} s_k\,\mathrm{d}t,$$

d'où

$$\begin{aligned}
\mathbf{E}X = x_n - \sum_{k=1}^{n-1} (x_{k+1} - x_k) s_k &= x_n - \sum_{j=2}^{n} x_j s_{j-1} + \sum_{j=1}^{n-1} x_j s_j \\
&= x_n - x_n s_{n-1} + \sum_{j=2}^{n-1} x_j (s_j - s_{j-1}) + x_1 s_1 \\
&= x_n p_n + \sum_{j=2}^{n-1} x_j p_j + x_1 p_1 = \sum_{j=1}^{n} p_k x_k.
\end{aligned}$$

Ceci établit (7.12). □

Proposition 7.10 (espérance d'une v.a. positive à densité)**.** *Si la variable aléatoire positive X a pour densité f,*

$$\mathbf{E}X = \int_0^{+\infty} xf(x)\,\mathrm{d}x. \tag{7.14}$$

Dans cette formule, $\mathbf{E}X$ peut prendre la valeur $+\infty$ si l'intégrale généralisée diverge.

Preuve. Si X admet pour densité f, $P(X > t) = \int_t^{+\infty} f(x)\,\mathrm{d}x$ pour tout t. En reportant cette égalité dans la définition de $\mathbf{E}X$, on obtient :

$$\int_0^{+\infty} P(X > t)\,\mathrm{d}t = \int_0^{+\infty} \left\{ \int_t^{+\infty} f(x)\,\mathrm{d}x \right\} \mathrm{d}t = \int_0^{+\infty} \left\{ \int_0^{+\infty} f(x)\mathbf{1}_{[t,+\infty[}(x)\,\mathrm{d}x \right\} \mathrm{d}t.$$

Notons que pour $t \geqslant 0$, $\mathbf{1}_{[t,+\infty[}(x) = \mathbf{1}_{[0,x]}(t)$. L'intégrande $(x,t) \mapsto \mathbf{1}_{[0,x]}(t)f(x)$ étant *positive*, le théorème de Fubini-Tonelli légitime l'interversion des intégrations [7], ce qui donne :

$$\mathbf{E}X = \int_0^{+\infty} \left\{ \int_0^{+\infty} f(x)\mathbf{1}_{[0,x]}(t)\,\mathrm{d}t \right\} \mathrm{d}x = \int_0^{+\infty} f(x) \left\{ \int_0^{+\infty} \mathbf{1}_{[0,x]}(t)\,\mathrm{d}t \right\} \mathrm{d}x.$$

Comme pour $x \geqslant 0$, $\int_0^{+\infty} \mathbf{1}_{[0,x]}(t)\,\mathrm{d}t = \int_0^x \mathrm{d}t = x$, on en déduit (7.14). □

Remarque 7.11. Notons que dans la démonstration ci-dessus, nous n'avons utilisé à aucun moment la positivité de la variable aléatoire X. On peut donc appliquer ce calcul à toute variable aléatoire réelle X ayant une densité f pour obtenir :

$$\int_0^{+\infty} P(X > t)\,\mathrm{d}t = \int_0^{+\infty} xf(x)\,\mathrm{d}x \quad \text{(égalité dans } \overline{\mathbb{R}}_+\text{)}. \tag{7.15}$$

Attention à ne pas écrire $\mathbf{E}X$ au premier membre de (7.15), cette quantité n'étant pour l'instant définie que pour X *positive*. La vraie formule pour $\mathbf{E}X$ lorsque la v.a. réelle X est à densité est donnée à la proposition 7.32.

Proposition 7.12. *Si X est une variable aléatoire positive et c une constante réelle strictement positive, on a*

$$\mathbf{E}(cX) = c\mathbf{E}X.$$

Cette égalité reste vraie pour $c = 0$ si X est de plus intégrable*.*

Preuve. Puisque X est une variable aléatoire positive et c une constante positive, $cX : \omega \mapsto (cX)(\omega) := cX(\omega)$ est une variable aléatoire positive. En lui appliquant la définition 7.1, on obtient :

$$\mathbf{E}(cX) = \int_0^{+\infty} P(cX > t)\,\mathrm{d}t = \int_0^{+\infty} P\Big(X > \frac{t}{c}\Big)\,\mathrm{d}t.$$

7. Même si les intégrales valent $+\infty$. Pour une preuve ne reposant pas sur un théorème admis, voir l'exercice 4.11.

Dans cette intégrale généralisée d'une fonction *positive* localement intégrable sur $[0,+\infty[$, on peut effectuer le changement de variable $s = t/c$, cf. proposition 4.41-ii), qui nous donne :

$$\mathbf{E}(cX) = \int_0^{+\infty} P\Big(X > \frac{t}{c}\Big)\,\mathrm{d}t = c\int_0^{+\infty} P(X > s)\,\mathrm{d}s = c\mathbf{E}X.$$

Dans le cas particulier $c = 0$, cette méthode n'est plus valable (on ne peut déjà plus écrire « $P(cX > t) = P(X > t/c)$ ») mais la formule est vraie trivialement, à condition que $\mathbf{E}X$ soit *finie*, puisqu'alors $\mathbf{E}(0 \times X) = \mathbf{E}(0) = 0$ et $0 \times \mathbf{E}X = 0$. □

Proposition 7.13 (croissance de l'espérance)**.** *Si X et Y sont deux variables aléatoires positives définies sur le même $(\Omega, \mathcal{F}, P)$ et si $X \leqslant Y$ c.-à-d. $X(\omega) \leqslant Y(\omega)$ pour tout $\omega \in \Omega$, alors $\mathbf{E}X \leqslant \mathbf{E}Y$.*

Preuve. Si $X(\omega) > t$, alors comme $Y(\omega) \geqslant X(\omega)$, on a aussi $Y(\omega) > t$. Ceci justifie l'inclusion d'évènements $\{X > t\} \subset \{Y > t\}$, puis l'inégalité $P(X > t) \leqslant P(Y > t)$. Cette dernière inégalité étant vérifiée pour tout t, on peut l'intégrer entre 0 et $+\infty$, pour obtenir [8] :

$$\mathbf{E}X = \int_0^{+\infty} P(X > t)\,\mathrm{d}t \leqslant \int_0^{+\infty} P(Y > t)\,\mathrm{d}t = \mathbf{E}Y.$$

□

La proposition suivante nous donne un peu de confort pour l'expression des espérances de variables aléatoires positives et nous sera utile pour l'inégalité de Markov.

Proposition 7.14. *Pour toute variable aléatoire positive X,*

$$\int_0^{+\infty} P(X > t)\,\mathrm{d}t = \int_0^{+\infty} P(X \geqslant t)\,\mathrm{d}t \quad \textit{(égalité dans } \overline{\mathbb{R}}_+\textit{)}. \tag{7.16}$$

Avant d'en donner la preuve, notons que (7.16) n'a rien d'évident car $P(X > t)$ et $P(X \geqslant t)$ peuvent différer pour certaines valeurs de t (au plus pour une infinité dénombrable de valeurs de t).

Preuve. Notons respectivement I et J le premier et le deuxième membre de (7.16). On prouve leur égalité en montrant l'inégalité dans les deux sens. L'inégalité $I \leqslant J$ s'obtient par intégration de l'inégalité $P(X > t) \leqslant P(X \geqslant t)$ vraie pour tout t.

Pour montrer que $J \leqslant I$, fixons $\varepsilon > 0$ quelconque. L'intégrande dans J est une fonction *positive* localement Riemann intégrable sur $[0,+\infty[$. On peut donc effectuer

8. La croissance de l'intégrale de Riemann (cf. prop. 3.22 ii)) passe aux intégrales généralisées de fonctions *positives*. En effet, si f et g sont positives et localement Riemann intégrables sur $[0,+\infty[$ et telles que $f \leqslant g$ sur $[0,+\infty[$, alors $\int_0^x f(t)\,\mathrm{d}t \leqslant \int_0^x g(t)\,\mathrm{d}t$ pour tout $x \geqslant 0$ et cette inégalité entre deux fonctions *croissantes* de x se conserve dans $\overline{\mathbb{R}}_+$ par passage à la limite quand x tend vers $+\infty$.

dans J le changement de variable « translation » $t = s + \varepsilon$, cf. proposition 4.41-i), qui nous donne :

$$J = \int_{-\varepsilon}^{+\infty} P(X \geqslant s + \varepsilon)\,\mathrm{d}s = \int_{-\varepsilon}^{0} P(X \geqslant s + \varepsilon)\,\mathrm{d}s + \int_{0}^{+\infty} P(X \geqslant s + \varepsilon)\,\mathrm{d}s.$$

En majorant $P(X \geqslant s + \varepsilon)$ par 1 sur $[-\varepsilon, 0]$ et par $P(X > s)$ sur $[0, +\infty[$, on en déduit que

$$J \leqslant \varepsilon + \int_{0}^{+\infty} P(X > s)\,\mathrm{d}s = \varepsilon + I.$$

L'inégalité $J \leqslant I + \varepsilon$ étant ainsi vérifiée pour tout $\varepsilon > 0$, on en déduit en faisant tendre ε vers 0 que $J \leqslant I$. □

Remarque 7.15. Dans la démonstration ci-dessus, la positivité de X ne joue aucun rôle. Donc (7.16) reste valable pour n'importe quelle variable aléatoire réelle X. Une adaptation facile de la preuve ci-dessus montre que l'on a aussi

$$\int_{-\infty}^{0} P(X < t)\,\mathrm{d}t = \int_{-\infty}^{0} P(X \leqslant t)\,\mathrm{d}t. \tag{7.17}$$

Proposition 7.16 (inégalité de Markov)**.** *Si X est une variable aléatoire positive,*

$$\forall x > 0, \quad P(X \geqslant x) \leqslant \frac{\mathbf{E}X}{x}. \tag{7.18}$$

Remarques 7.17. Cette inégalité n'a d'intérêt que lorsque le second membre est inférieur à 1, c'est-à-dire lorsque $\mathbf{E}X < +\infty$ *et* $x > \mathbf{E}X$.

D'autre part il peut sembler un peu incongru de vouloir contrôler $P(X \geqslant x)$ à l'aide de $\mathbf{E}X$, puisque le calcul de cette espérance par la définition 7.1 présuppose la connaissance des $P(X > t)$ pour $t \geqslant 0$ (dont on déduit facilement les $P(X \geqslant t)$). Il se trouve qu'il arrive souvent en pratique que l'on sache calculer $\mathbf{E}X$ sans connaître, ou sans avoir besoin de calculer, la loi de X. C'est le cas par exemple quand X est une somme finie de variables aléatoires d'espérances connues. On peut aussi savoir majorer $\mathbf{E}X$ sans connaître la loi de X. Dans ces situations, l'inégalité de Markov est très utile. Pour ne citer qu'un exemple, l'inégalité de Markov est l'un des outils pour établir des « lois des grands nombres ».

Voici maintenant 3 preuves de l'inégalité de Markov, libre au lecteur de choisir celle qu'il préfère.

Preuve n° 1. C'est la preuve « muette » donnée par la figure 7.7.

□

Preuve n° 2. Cette preuve ne fait que traduire explicitement la preuve graphique n° 1. Fixons $x > 0$, la quantité $P(X \geqslant x)$ devenant ainsi une constante. À partir de cette constante, définissons la fonction $h : [0, +\infty[\to \mathbb{R}_+,\ t \mapsto P(X \geqslant x)\mathbf{1}_{[0,x]}(t)$. Par

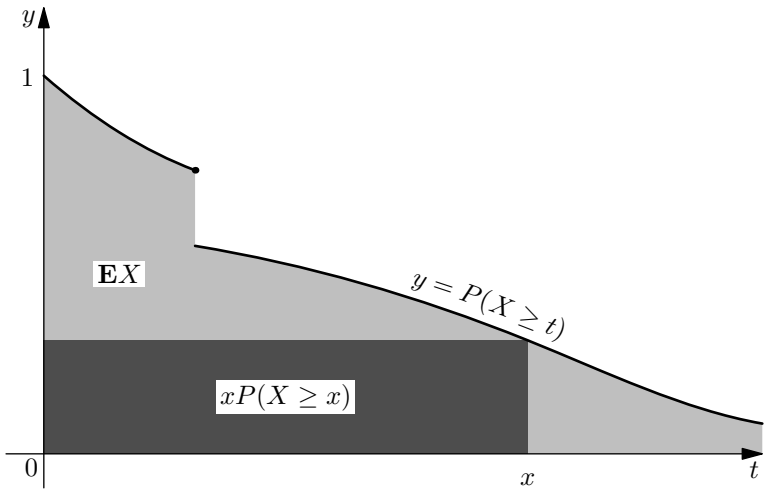


FIGURE 7.7 – Inégalité de Markov : $xP(X \geqslant x) \leqslant \int_0^{+\infty} P(X \geqslant t)\,\mathrm{d}t = \mathbf{E}X$.

décroissance de la fonction $t \mapsto P(X \geqslant t)$, on a $h(t) \leqslant P(X \geqslant t)$ pour tout $t \in [0, x]$. D'autre part cette inégalité est aussi vérifiée pour tout $t > x$ car alors $h(t) = 0$. En intégrant sur $[0, +\infty[$ l'inégalité $h(t) \leqslant P(X \geqslant t)$, on obtient compte-tenu de (7.16) :

$$\int_0^{+\infty} h(t)\,\mathrm{d}t \leqslant \int_0^{+\infty} P(X \geqslant t)\,\mathrm{d}t = \mathbf{E}X.$$

D'autre part, puisque h est nulle sur $]x, +\infty[$,

$$\int_0^{+\infty} h(t)\,\mathrm{d}t = \int_0^{x} P(X \geqslant x)\,\mathrm{d}t = xP(X \geqslant x).$$

Par conséquent

$$xP(X \geqslant x) \leqslant \mathbf{E}X,$$

ce qui nous donne (7.18) puisque $x > 0$. □

Preuve n° 3. Cette preuve plus abstraite exploite les propriétés déjà connues de l'espérance des v.a. positives. On fixe x qui joue donc le rôle d'une constante dans toute la preuve. On part de l'inégalité entre v.a. positives : $x\mathbf{1}_{\{X \geqslant x\}} \leqslant X$ (vérifiez) dont on déduit par croissance de $\mathbf{E}$ (proposition 7.13) :

$$\mathbf{E}\big(x\mathbf{1}_{\{X \geqslant x\}}\big) \leqslant \mathbf{E}X,$$

puis grâce aux propositions 7.12 et 7.7, $xP(X \geqslant x) \leqslant \mathbf{E}X$. On conclut en divisant par $x > 0$. □

Corollaire 7.18. *Si X est une variable aléatoire positive, on a l'équivalence*

$$\mathbf{E}X = 0 \Leftrightarrow P(X = 0) = 1,$$

autrement dit, $\mathbf{E}X$ *est nulle si et seulement si* X *est* presque sûrement *nulle.*

Preuve. Supposons d'abord que $\mathbf{E}X = 0$. La v.a. X étant positive, l'égalité $P(X = 0) = 1$ équivaut à $P(X > 0) = 0$. Introduisons la suite des évènements $A_n := \{X \geqslant 1/n\}$, $n \in \mathbb{N}^*$. Cette suite est croissante de réunion $A := \{X > 0\}$. Par continuité séquentielle croissante de P, $P(A) = \lim_{n\to+\infty} P(A_n)$. Or l'inégalité de Markov appliquée avec $x = 1/n$ nous montre que pour tout $n \in \mathbb{N}^*$, $0 \leqslant P(A_n) \leqslant n\mathbf{E}X = 0$. Ainsi $P(A_n) = 0$ pour tout n et $P(A) = 0$ comme limite de la suite nulle.

Réciproquement, si $P(X = 0) = 1$, alors pour tout $t \geqslant 0$, $P(X > t) = 0$, d'où $\mathbf{E}X = \int_0^{+\infty} 0\,\mathrm{d}t = 0$. □

Proposition 7.19 (inégalité de Markov raffinée)**.** *Si X est une variable aléatoire positive* intégrable *(*$\mathbf{E}X < +\infty$*),*

$$\lim_{x\to+\infty} xP(X \geqslant x) = 0. \tag{7.19}$$

Preuve. Voir la figure 7.8 et sa légende. □

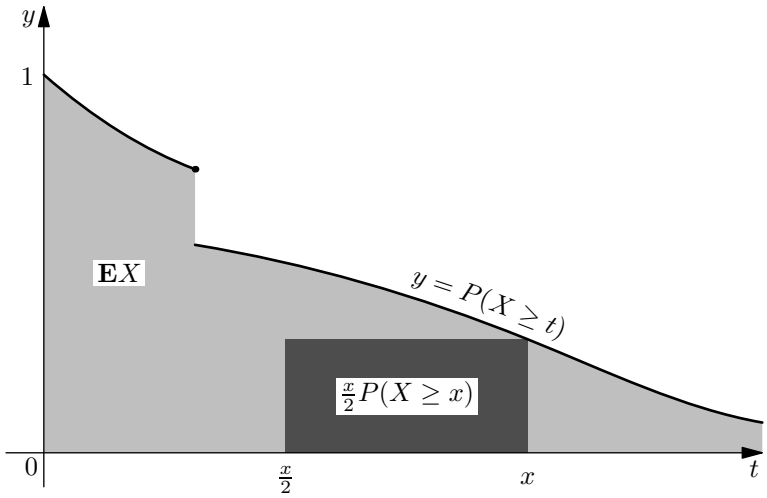


FIGURE 7.8 – $\frac{x}{2}P(X \geqslant x) \leqslant \int_{x/2}^{+\infty} P(X \geqslant t)\,\mathrm{d}t = o(1)$ si $\mathbf{E}X < +\infty$.

Remarque 7.20. La réciproque de la proposition 7.19 est fausse. Il est possible que $xP(X \geqslant x) = o(1)$ sans que X soit intégrable, par exemple avec X de fonction de survie $G(t) = P(X > t) = 1$ si $t \leqslant \mathrm{e}$ et $1/(t \ln t)$ si $t > \mathrm{e}$.

Théorème 7.21 (approximation par suite croissante de v.a. simples)**.** *Toute variable aléatoire positive X sur $(\Omega, \mathcal{F})$ est limite simple sur Ω d'une suite* croissante $(X_n)_{n\geqslant 1}$ *de variables aléatoires positives simples.*

Preuve. L'idée est d'utiliser pour construire X_n, les *valeurs approchées dyadiques par défaut* de X au *niveau de résolution* n. On peut procéder comme suit en définissant pour $n \in \mathbb{N}$, les ensembles

$$\begin{aligned} A_{n,k} &:= X^{-1}\big([k2^{-n}, (k+1)2^{-n}[\big), \quad 0 \leqslant k \leqslant n2^n - 1; \\ A_{n,n2^n} &:= X^{-1}\big([n, +\infty[\big). \end{aligned}$$

On prend alors

$$X_n := \sum_{k=0}^{n2^n} \frac{k}{2^n} \mathbf{1}_{A_{n,k}}.$$

Autrement dit,

$$X_n(\omega) = \begin{cases} n & \text{si } n \leqslant X(\omega), \\ k2^{-n} & \text{pour l'unique entier } k \text{ tel que } k2^{-n} \leqslant X(\omega) < (k+1)2^{-n} \text{ sinon.} \end{cases}$$

Comme X est mesurable, les $A_{n,k}$ sont dans $\mathcal{F}$, ce qui entraîne la mesurabilité de X_n (combinaison linéaires d'indicatrices d'éléments de la tribu $\mathcal{F}$). Comme $X_n(\Omega)$ est une partie finie de $\mathbb{R}_+$, X_n est simple positive.

Il reste à vérifier que pour tout $\omega \in \Omega$, la suite de *réels* $\big(X_n(\omega)\big)_{n\geqslant 1}$ est croissante et converge dans $\mathbb{R}_+$ vers $X(\omega)$. On note $[x]$ la partie entière du réel x, unique entier m tel que $m \leqslant x < m+1$. On voit que $X_n(\omega) = n$ pour $n \leqslant [X(\omega)]$ et que pour $n > [X(\omega)]$,

$$X_n(\omega) = \frac{k(n,\omega)}{2^n} = \max\Big\{ \frac{l}{2^n} \,;\, \frac{l}{2^n} \leqslant X(\omega),\ l \in \mathbb{N} \Big\} = 2^{-n}[2^n X(\omega)]. \tag{7.20}$$

La suite *finie* $\big(X_n(\omega)\big)_{n\leqslant [X(\omega)]}$ est clairement croissante. Passons à $\big(X_n(\omega)\big)_{n>[X(\omega)]}$. D'après (7.20), on a

$$X_n(\omega) = \frac{k(n,\omega)}{2^n} = \frac{2k(n,\omega)}{2^{n+1}} \leqslant X(\omega) \quad \Rightarrow \quad \frac{2k(n,\omega)}{2^{n+1}} \leqslant \frac{k(n+1,\omega)}{2^{n+1}} = X_{n+1}(\omega),$$

d'où la croissance de la suite $\big(X_n(\omega)\big)_{n>[X(\omega)]}$. Pour établir définitivement la croissance de *toute* la suite $\big(X_n(\omega)\big)_{n\geqslant 1}$, il ne reste plus qu'à examiner le point de raccord des deux sous-suites, donc à comparer $X_n(\omega)$ et $X_{n+1}(\omega)$ pour $n = [X(\omega)]$. Il suffit de remarquer que $X_n(\omega) = n = (n2^{n+1})2^{-n-1} \leqslant X(\omega)$ et comme $X_{n+1}(\omega)$ est donné par (7.20), on a $(n2^{n+1})2^{-n-1} \leqslant k(n+1,\omega)2^{-n-1} = X_{n+1}(\omega)$.

La convergence est immédiate, puisque pour $n > [X(\omega)]$, on a d'après (7.20)

$$X_n(\omega) \leqslant X(\omega) < X_n(\omega) + \frac{1}{2^n},$$

d'où $0 \leqslant X(\omega) - X_n(\omega) < 2^{-n}$. Le théorème est démontré. On peut remarquer en bonus que la convergence est *uniforme* sur Ω si X est bornée c'est-à-dire $M := \sup_{\omega\in\Omega} X(\omega) < +\infty$. En effet pour $n > M$ (constante indépendante de ω), on a pour tout $\omega \in \Omega$, $0 \leqslant X(\omega) - X_n(\omega) < 2^{-n}$. □

Le lemme suivant est un outil pour l'interversion limite-espérance dans le cas d'une suite croissante de variables aléatoires (théorème de B. Levi).

Lemme 7.22. *Si $(X_n)_{n\geqslant 1}$ est une suite croissante de variables aléatoires positives ayant pour limite sur tout Ω la variable aléatoire positive X, on a*

$$\forall t \geqslant 0, \quad P(X_n > t) \uparrow P(X > t). \tag{7.21}$$

Preuve. Pour t fixé, la suite des évènements $A_n := \{X_n > t\}$ est *croissante pour l'inclusion* (puisque si $\omega \in A_n$, $X_{n+1}(\omega) \geqslant X_n(\omega) > t$, donc $\omega \in A_{n+1}$). Par continuité séquentielle croissante de P, $P(A_n)$ tend en croissant vers $P(A)$, où $A := \bigcup_{n\geqslant 1} A_n$. On prouve alors la convergence (7.21) en vérifiant que $A = \{X > t\}$. Pour cela on montre l'inclusion dans les deux sens.

Soit $\omega \in A$ quelconque. Cela signifie qu'il existe un $n_0 = n_0(\omega)$ tel que $\omega \in A_{n_0}$, c.-à-d. $X_{n_0}(\omega) > t$. Alors $X_n(\omega) \geqslant X_{n_0}(\omega)$ pour tout $n \geqslant n_0$, par croissance de la suite $(X_n(\omega))_{n\geqslant 1}$, d'où en faisant tendre n vers l'infini, $X(\omega) \geqslant X_{n_0}(\omega) > t$. Donc en particulier, $X(\omega) > t$, d'où $\omega \in \{X > t\}$. Comme ω était quelconque, ceci établit l'inclusion $A \subset \{X > t\}$.

Pour l'inclusion inverse, soit ω quelconque dans $\{X > t\}$. Alors $X(\omega) > t$ et comme cette inégalité est *stricte*[9] et $X_n(\omega)$ tend vers $X(\omega)$ quand n tend vers l'infini, on peut trouver un $n_0 = n_0(\omega)$ tel que pour tout $n \geqslant n_0$, $X_n(\omega) > t$. Ainsi $\omega \in A_n$ pour tout $n \geqslant n_0$, donc *a fortiori* $\omega \in A$. L'inclusion $\{X > t\} \subset A$ est ainsi établie et ceci termine la preuve du lemme. □

Théorème 7.23 (de Beppo Levi)**.** *Soit $(X_n)_{n\geqslant 1}$ une suite de variables aléatoires positives qui converge en croissant vers la variable aléatoire positive X, c'est-à-dire $X_n \leqslant X_{n+1}$ pour tout n et $\lim_{n\to+\infty} X_n(\omega) = X(\omega)$ pour tout $\omega \in \Omega$. Alors la suite $(\mathbf{E}X_n)_{n\geqslant 1}$ converge en croissant vers $\mathbf{E}X$ dans $\overline{\mathbb{R}}_+$.*

Preuve. La croissance de la suite $(\mathbf{E}X_n)_{n\geqslant 1}$ dans $\overline{\mathbb{R}}_+$ est évidente par croissance de l'espérance (proposition 7.13). Pour étudier sa convergence, nous allons distinguer deux cas selon l'intégrabilité de X.

Cas 1, $\mathbf{E}X < +\infty$. Fixons $\varepsilon > 0$ quelconque. Par convergence de l'intégrale de Riemann généralisée $\int_0^{+\infty} P(X > t)\,\mathrm{d}t$, il existe un réel $b = b(\varepsilon) > 0$ tel que :

$$\int_b^{+\infty} P(X > t)\,\mathrm{d}t < \varepsilon. \tag{7.22}$$

9. C'est là qu'on voit pourquoi le lemme ne marcherait pas avec $P(X_n \geqslant t)$ et $P(X \geqslant t)$. Voici d'ailleurs un contre exemple élémentaire. On prend $X_n = 1 - 1/n$, v.a. constante et $X = 1$. Alors avec $t = 1$, on obtient $P(X_n \geqslant 1) = 0$ pour tout n, tandis que $P(X \geqslant 1) = 1$, ce qui empêche la convergence.

De cette égalité et de la croissance de (X_n) qui implique que $X_n \leqslant X$ et donc que pour tout t, $P(X_n > t) \leqslant P(X > t)$, on déduit :

$$\forall n \geqslant 1, \quad \int_b^{+\infty} P(X_n > t)\,\mathrm{d}t < \varepsilon. \tag{7.23}$$

Posons $G_n(t) := P(X_n > t)$ et $G(t) := P(X > t)$. Chaque G_n est une fonction *décroissante* sur $[0, b]$ et par le lemme 7.22, $G_n(t)$ converge vers $G(t)$ pour tout t dans $[0, b]$. La suite (G_n) satisfait ainsi les hypothèses du théorème 3.36 d'interversion limite et intégrale de Riemann pour une suite de *fonctions décroissantes*, donc :

$$\int_0^b P(X_n > t)\,\mathrm{d}t \xrightarrow[n\to+\infty]{} \int_0^b P(X > t)\,\mathrm{d}t. \tag{7.24}$$

En combinant (7.22), (7.23) et (7.24), on voit qu'il existe un entier $n_0(\varepsilon)$ tel que

$$\forall n \geqslant n_0(\varepsilon), \quad \int_0^{+\infty} P(X > t)\,\mathrm{d}t - 2\varepsilon \leqslant \int_0^{+\infty} P(X_n > t)\,\mathrm{d}t \leqslant \int_0^{+\infty} P(X > t)\,\mathrm{d}t,$$

ce qui établit la convergence de $\mathbf{E}X_n$ vers $\mathbf{E}X$ par arbitrarité de ε. □

Cas 2, $\mathbf{E}X = +\infty$. Il s'agit cette fois de montrer que $\mathbf{E}X_n$ tend vers $+\infty$. La divergence de l'intégrale généralisée $\int_0^{+\infty} P(X > t)\,\mathrm{d}t$ signifie ici que $\int_0^x P(X > t)\,\mathrm{d}t$ tend vers $+\infty$ quand x tend vers $+\infty$. Donc si on fixe $A > 0$ arbitraire, on peut trouver un $b \in]0, +\infty[$ tel que

$$\int_0^b P(X > t)\,\mathrm{d}t > A. \tag{7.25}$$

En appliquant comme ci-dessus le théorème d'interversion limite intégrale pour la suite de fonctions décroissantes $(G_n)_{n\geqslant 1}$, on obtient la convergence de $\int_0^b P(X_n > t)\,\mathrm{d}t$ vers $\int_0^b P(X > t)\,\mathrm{d}t$ quand n tend vers $+\infty$. Il existe donc un entier $n_0(A)$ tel que :

$$\forall n \geqslant n_0(A), \quad \int_0^b P(X_n > t)\,\mathrm{d}t > \frac{A}{2}. \tag{7.26}$$

Ceci implique évidemment que

$$\forall n \geqslant n_0(A), \quad \mathbf{E}X_n = \int_0^{+\infty} P(X_n > t)\,\mathrm{d}t > \frac{A}{2},$$

et comme A est arbitraire, on a ainsi prouvé que $\mathbf{E}X_n$ tend vers $+\infty$. □

La démonstration du théorème de B. Levi est achevée. □

Corollaire 7.24 (additivité de l'espérance)**.** *Si X et Y sont deux variables aléatoires positives définies sur le même $(\Omega, \mathcal{F}, P)$,*

$$\mathbf{E}(X + Y) = \mathbf{E}X + \mathbf{E}Y \quad \textit{(égalité dans } \overline{\mathbb{R}}_+\textit{)}. \tag{7.27}$$

Preuve. On montre (7.27) par un passage à la limite à partir de l'additivité de l'espérance des variables simples. Cette propriété est établie par le lemme 7.25 ci-dessous.

Par le théorème 7.21, il existe deux suites (X_n) et (Y_n) de variables aléatoires simples positives telles que $X_n \uparrow X$ et $Y_n \uparrow Y$. Alors en posant $Z_n := X_n + Y_n$, on a aussi $Z_n \uparrow X + Y$. Par le lemme 7.25 ci-dessous,

$$\forall n \in \mathbb{N}, \quad \mathbf{E}Z_n = \mathbf{E}(X_n + Y_n) = \mathbf{E}X_n + \mathbf{E}Y_n. \tag{7.28}$$

En appliquant le théorème de Beppo Levi à chacune des suites (Z_n), (X_n) et (Y_n), on peut passer à la limite quand n tend vers $+\infty$ dans (7.28) pour obtenir (7.27). □

Lemme 7.25. *Si X et Y sont deux variables aléatoires positives simples, il en est de même pour $Z := X + Y$ et $\mathbf{E}Z = \mathbf{E}X + \mathbf{E}Y$.*

Preuve. Par hypothèse $X(\Omega)$ et $Y(\Omega)$ sont des sous-ensembles finis de $\mathbb{R}_+$. Notons $X(\Omega) = \{x_1, \ldots, x_l\}$ et $Y(\Omega) = \{y_1, \ldots, y_m\}$. Les seules valeurs possibles pour $Z(\omega)$ sont parmi les réels de la forme $x_i + y_j$ avec $(i, j) \in [\![1, l]\!] \times [\![1, m]\!]$. Donc $Z(\Omega)$ est fini ($\operatorname{card} Z(\Omega) = n \leqslant lm$) et Z est une v.a. positive simple. Notons $Z(\Omega) = \{z_1, \ldots, z_n\}$.

On voit facilement, en partitionnant Ω suivant toutes les valeurs possibles de Y, resp. de X, que

$$P(X = x_i) = \sum_{y_j \in Y(\Omega)} P(X = x_i \text{ et } Y = y_j), \tag{7.29}$$

$$P(Y = y_j) = \sum_{x_i \in X(\Omega)} P(X = x_i \text{ et } Y = y_j). \tag{7.30}$$

Pour $z_k \in Z(\Omega)$, la décomposition en union finie d'évènements 2 à 2 disjoints [10]

$$\{Z = z_k\} = \bigcup_{x_i + y_j = z_k} \{X = x_i \text{ et } Y = y_j\},$$

nous donne

$$P(Z = z_k) = \sum_{x_i + y_j = z_k} P(X = x_i \text{ et } Y = y_j). \tag{7.31}$$

On obtient alors la conclusion souhaitée par le calcul suivant dans lequel tous les Σ sont des sommes d'un nombre fini de termes et où on utilise (7.29)–(7.31) et la

10. Pour $z_k \in Z(\Omega)$ et certains couples (x_i, y_j) tels que $x_i + y_j = z_k$, il se peut qu'il n'existe aucun $\omega \in \Omega$ tel que $X(\omega) = x_i$ et $Y(\omega) = y_j$, l'évènement $\{X = x_i \text{ et } Y = y_j\}$ est alors vide.

proposition 7.9 appliquée d'abord à Z, puis à X et Y.

$$\begin{aligned}
\mathbf{E}Z &= \sum_{z_k\in Z(\Omega)} z_k P(Z=z_k)\\
&= \sum_{z_k\in Z(\Omega)} z_k \sum_{x_i+y_j=z_k} P(X=x_i \text{ et } Y=y_j)\\
&= \sum_{z_k\in Z(\Omega)} \sum_{x_i+y_j=z_k} (x_i+y_j)P(X=x_i \text{ et } Y=y_j)\\
&= \sum_{x_i\in X(\Omega), y_j\in Y(\Omega)} (x_i+y_j)P(X=x_i \text{ et } Y=y_j)\\
&= \sum_{x_i\in X(\Omega), y_j\in Y(\Omega)} x_i P(X=x_i \text{ et } Y=y_j) + \sum_{x_i\in X(\Omega), y_j\in Y(\Omega)} y_j P(X=x_i \text{ et } Y=y_j)\\
&= \sum_{x_i\in X(\Omega)} x_i \sum_{y_j\in Y(\Omega)} P(X=x_i \text{ et } Y=y_j) + \sum_{y_j\in Y(\Omega)} y_j \sum_{x_i\in X(\Omega)} P(X=x_i \text{ et } Y=y_j)\\
&= \sum_{x_i\in X(\Omega)} x_i P(X=x_i) + \sum_{y_j\in Y(\Omega)} y_j P(Y=y_j)\\
&= \mathbf{E}X + \mathbf{E}Y.
\end{aligned}$$

□

Corollaire 7.26 (interversion série-espérance)**.** *Soit $(X_k)_{k\in\mathbb{N}}$, une suite de variables aléatoires* positives *définies sur $(\Omega,\mathcal{F})$. On suppose que pour tout $\omega\in\Omega$, la série $\sum_{k=0}^{+\infty} X_k(\omega)$ converge*[11] *dans $\mathbb{R}_+$. Alors l'application*

$$S:\Omega\to\mathbb{R}_+, \quad \omega\longmapsto S(\omega):=\sum_{k=0}^{+\infty} X_k(\omega),$$

est une variable aléatoire positive sur $(\Omega,\mathcal{F})$ et

$$\mathbf{E}S = \mathbf{E}\left(\sum_{k=0}^{+\infty} X_k\right) = \sum_{k=0}^{+\infty} \mathbf{E}X_k, \tag{7.32}$$

cette égalité ayant lieu dans $\overline{\mathbb{R}}_+$.

Preuve. L'application $S:\Omega\to\mathbb{R}_+$ est bien définie, grâce à l'hypothèse de convergence simple de la série sur tout Ω. Posons

$$\forall n\in\mathbb{N}, \quad S_n := \sum_{k=0}^{n} X_k.$$

11. Nous donnons ici une « version bridée » de l'interversion série-espérance. On pourrait se passer de cette hypothèse de convergence dans $\mathbb{R}_+$, en considérant que cette série converge toujours dans $\overline{\mathbb{R}}_+$. Il faudrait alors considérer qu'une variable aléatoire positive est une application à valeurs dans $\overline{\mathbb{R}}_+$, ce que nous avons évité de faire dans cet ouvrage...

Les S_n sont mesurables $\mathcal{F}$ - Bor($\mathbb{R}_+$) comme sommes finies d'applications mesurables (les X_k). Comme S est limite simple sur tout Ω d'une suite d'applications mesurables, elle est mesurable pour les tribus $\mathcal{F}$ - Bor($\mathbb{R}_+$), c'est donc bien une *variable aléatoire* positive.

Par additivité de l'espérance (corollaire 7.24, qui s'étend par une récurrence immédiate à toute somme d'un nombre *fini* de variables aléatoires positives [12]) on a

$$\forall n \in \mathbb{N}, \quad \mathbf{E}S_n = \sum_{k=0}^{n} \mathbf{E}X_k \quad \text{(égalité dans } \overline{\mathbb{R}}_+\text{)}. \tag{7.33}$$

En raison de la positivité des X_k, la convergence de S_n vers S est *croissante* ($S_n \uparrow S$), donc par le théorème de Beppo Levi, $\mathbf{E}S_n$ converge vers $\mathbf{E}S$ quand n tend vers $+\infty$. En faisant tendre n vers $+\infty$ dans (7.33), on obtient alors la conclusion souhaitée (noter que le second membre de (7.33) converge dans $\overline{\mathbb{R}}_+$ vers le second membre de (7.32)). □

Corollaire 7.27 (espérance d'une v.a. discrète positive)**.** *Pour toute variable aléatoire discrète positive X,*

$$\mathbf{E}X = \sum_{x \in X(\Omega)} xP(X = x), \qquad \textit{(égalité dans } \overline{\mathbb{R}}_+\textit{)}. \tag{7.34}$$

Preuve. Puisque X est discrète positive, $X(\Omega)$ est une partie finie ou dénombrable de $\mathbb{R}_+$. Le cas où $X(\Omega)$ est finie est celui des variables aléatoires simples, déjà traité ci-dessus (proposition 7.9). Supposons désormais $X(\Omega)$ dénombrable et indexons ses éléments par les entiers : $X(\Omega) = \{x_k \ ; \ k \in \mathbb{N}\}$. On peut alors représenter X comme la somme d'une série de variables aléatoires positives en écrivant (la justification suit) :

$$X = \sum_{k=0}^{+\infty} x_k \mathbf{1}_{A_k}, \quad (A_k := \{X = x_k\}, \ \forall k \in \mathbb{N}). \tag{7.35}$$

En effet les évènements A_k réalisent une partition de Ω, donc pour tout $\omega \in \Omega$, il existe un unique indice $j = j(\omega)$ tel que $\omega \in A_j$ et pour cet ω, $X(\omega) = x_j$. On a alors

$$\sum_{k=0}^{+\infty} x_k \mathbf{1}_{A_k}(\omega) = x_j \mathbf{1}_{A_j}(\omega) = x_j = X(\omega),$$

car dans cette série de réels positifs, il y a au plus un terme non nul [13], celui d'indice j (pour $k \neq j$, $\omega \notin A_k$, d'où $\mathbf{1}_{A_k}(\omega) = 0$). Ceci prouve que la série de variables aléatoires positives (7.35) converge sur tout Ω vers la variable aléatoire positive X. Les hypothèses du corollaire 7.26 étant ainsi vérifiées, on a

$$\mathbf{E}X = \sum_{k=0}^{+\infty} \mathbf{E}(x_k \mathbf{1}_{A_k}) = \sum_{k=0}^{+\infty} x_k \mathbf{E}(\mathbf{1}_{A_k}) = \sum_{k=0}^{+\infty} x_k P(A_k) = \sum_{k=0}^{+\infty} x_k P(X = x_k),$$

en utilisant aussi les propositions 7.12 et 7.7. La formule (7.34) est ainsi démontrée. □

12. Mais cette récurrence ne permet pas de traiter le cas d'une *série*.
13. Il se peut qu'ils soient tous nuls si $x_j = 0$.

7.3 Espérance d'une variable aléatoire réelle

L'extension de la notion d'espérance au cas des variables aléatoires réelles repose sur la décomposition $X = X^+ - X^-$ et la formule « $\mathbf{E}X := \mathbf{E}(X^+) - \mathbf{E}(X^-)$ », sous réserve que cette soustraction ait lieu dans $\mathbb{R}$. Avant d'expliquer cela, donnons quelques précisions sur les variables aléatoires positives X^+ et X^-, définies par

$$X^+ := \max(0, X), \quad X^- := \max(0, -X).$$

Notons que X^+ *et* X^- sont des variables aléatoires [14] positives. En se rappelant la définition 2.30, on a donc :

$$X^+ : \Omega \to \mathbb{R}_+,\ \omega \mapsto X^+(\omega) = (X(\omega))^+, \quad X^- : \Omega \to \mathbb{R}_+,\ \omega \mapsto X^-(\omega) = (X(\omega))^-.$$

Au risque d'insister lourdement, notons encore que pour tout $\omega \in \Omega$,

$$X^+(\omega) = \begin{cases} |X(\omega)| & \text{si } X(\omega) \geqslant 0, \\ 0 & \text{si } X(\omega) \leqslant 0, \end{cases} \quad X^-(\omega) = \begin{cases} 0 & \text{si } X(\omega) \geqslant 0, \\ |X(\omega)| & \text{si } X(\omega) \leqslant 0. \end{cases}$$

On en déduit les égalités

$$X = X^+ - X^-, \qquad |X| = X^+ + X^-.$$

Pour définir $\mathbf{E}X$ comme la différence $\mathbf{E}(X^+) - \mathbf{E}(X^-)$, il est clair qu'il nous faut interdire que ces deux espérances de variables aléatoires positives vaillent *simultanément* $+\infty$. On pourrait autoriser l'une des deux à valoir $+\infty$ à condition que l'autre soit finie, $\mathbf{E}X$ prenant alors la valeur $+\infty$ si $\mathbf{E}(X^+) = +\infty$ et $\mathbf{E}(X^-) < +\infty$, ou dans le cas inverse, $\mathbf{E}X = -\infty$. Nous ne retiendrons pas cette option, car le problème de « $+\infty - \infty$ » réapparaîtrait de toutes façons pour la somme de deux variables aléatoires, ce qui empêcherait l'additivité de l'espérance pour les variables aléatoires réelles. Ainsi nous ne parlerons de l'espérance (sous-entendu dans $\mathbb{R}$) de la variable aléatoire réelle X que si $\mathbf{E}(X^+) < +\infty$ *et* $\mathbf{E}(X^-) < +\infty$, ou ce qui est équivalent si $\mathbf{E}|X| < +\infty$. Les propriétés de l'espérance établies dans cette section ne concernent que les variables aléatoires remplissant cette condition [15]. On dit que ces variables aléatoires réelles sont *intégrables*.

Définition 7.28 (v.a. réelle intégrable)**.** *La variable aléatoire réelle X est dite* intégrable *si $\mathbf{E}|X| < +\infty$, autrement dit si :*

$$\int_0^{+\infty} P(|X| > t)\,\mathrm{d}t < +\infty.$$

14. Elles héritent de la mesurabilité de X.

15. À titre exceptionnel, dans certaines situations, il peut être commode de dire qu'une variable aléatoire réelle a une espérance *infinie* si *une seule* des deux quantités $\mathbf{E}(X^+)$ et $\mathbf{E}(X^-)$ est infinie, par exemple dans la situation de la remarque 7.3, pour rester cohérent avec la définition de l'espérance d'une v.a. *positive* comme élément de $\overline{\mathbb{R}}_+$. Il convient alors de manipuler ces « espérances infinies » avec les plus grandes précautions. En tout état de cause, on ne parle *jamais* d'espérance de X, ni finie, ni infinie, lorsque $\mathbf{E}X^+$ et $\mathbf{E}X^-$ valent tous deux $+\infty$.

Définition 7.29 (espérance d'une v.a. réelle)**.** *Si X est une variable aléatoire réelle intégrable, on appelle* espérance de X*, le réel* $\mathbf{E}X$ *défini par*

$$\mathbf{E}X := \mathbf{E}(X^+) - \mathbf{E}(X^-). \tag{7.36}$$

Le second membre de (7.36) a bien un sens et représente un nombre réel, puisqu'en raison des inégalités $0 \leqslant X^+ \leqslant |X|$ et $0 \leqslant X^- \leqslant |X|$, l'intégrabilité de X implique la finitude de $\mathbf{E}(X^+)$ et de $\mathbf{E}(X^-)$.

Remarque 7.30. Une fois le réel $\mathbf{E}X$ défini comme ci-dessus, on peut parler de sa partie positive $(\mathbf{E}X)^+$ et de sa partie négative $(\mathbf{E}X)^-$. Il faut prendre garde à ne pas les confondre avec $\mathbf{E}(X^+)$ et $\mathbf{E}(X^-)$. Pour voir un exemple simple où ces quantités diffèrent, prenons X variable aléatoire de Rademacher : $X(\Omega) = \{-1, 1\}$ et $P(X = -1) = P(X = 1) = 1/2$. Comme X est bornée, elle est intégrable, donc $\mathbf{E}X$ existe. Comme la loi de X est symétrique, $\mathbf{E}X = 0$, donc $(\mathbf{E}X)^+ = 0^+ = 0$ et $(\mathbf{E}X)^- = 0^- = 0$. D'autre part, on vérifie que X^+ et X^- sont des variables de Bernoulli de paramètre $1/2$, donc $\mathbf{E}(X^+) = \mathbf{E}(X^-) = 1/2$.

La formule (7.36) ne semble pas très commode pour le calcul explicite de $\mathbf{E}X$. On aimerait pouvoir exprimer $\mathbf{E}X$, directement à l'aide de la loi de X, par exemple de sa fonction de répartition, sans passer par la loi de X^+ et celle de X^-. C'est l'objet de la proposition suivante.

Proposition 7.31 (calcul de $\mathbf{E}X$ à l'aide de la f.d.r.)**.** *Soit X une variable aléatoire réelle intégrable de fonction de répartition F. Son espérance vérifie*

$$\mathbf{E}X = \int_0^{+\infty} P(X > t)\,\mathrm{d}t - \int_{-\infty}^0 P(X < t)\,\mathrm{d}t = \int_0^{+\infty} \big(1 - F(t)\big)\,\mathrm{d}t - \int_{-\infty}^0 F(t)\,\mathrm{d}t. \tag{7.37}$$

Preuve. Compte-tenu des remarques faites au début de cette section, on vérifie facilement pour tout $t \geqslant 0$, les équivalences logiques suivantes :

$$X^+(\omega) > t \quad \Leftrightarrow \quad X(\omega) > t, \tag{7.38}$$

$$X^-(\omega) > t \quad \Leftrightarrow \quad X(\omega) < -t. \tag{7.39}$$

Laissant (7.38) au lecteur, détaillons la deuxième. Comme $t \geqslant 0$, l'inégalité $X^-(\omega) > t$ implique que $X^-(\omega)$ est strictement positif, donc que $X(\omega)$ est strictement négatif et dans ce cas, $X^-(\omega) = -X(\omega)$, d'où $-X(\omega) > t$ et donc $X(\omega) < -t$. Ceci justifie l'implication $\Rightarrow$ dans (7.39). Réciproquement, si $X(\omega) < -t$, $X(\omega)$ est strictement négatif, donc $X(\omega) = -X^-(\omega)$, d'où $-X^-(\omega) < -t$ et donc $X^-(\omega) > t$.

Les équivalences (7.38) et (7.39) donnent les égalités d'évènements $\{X^+ > t\} = \{X > t\}$ et $\{X^- > t\} = \{X < -t\}$, valables pour tout [16] $t \geqslant 0$, d'où :

$$\forall t \geqslant 0, \quad P(X^+ > t) = P(X > t), \quad P(X^- > t) = P(X < -t). \tag{7.40}$$

16. Elles ne seraient pas valables pour tout t réel. La positivité de t est utilisée dans la vérification de (7.38) et (7.39).

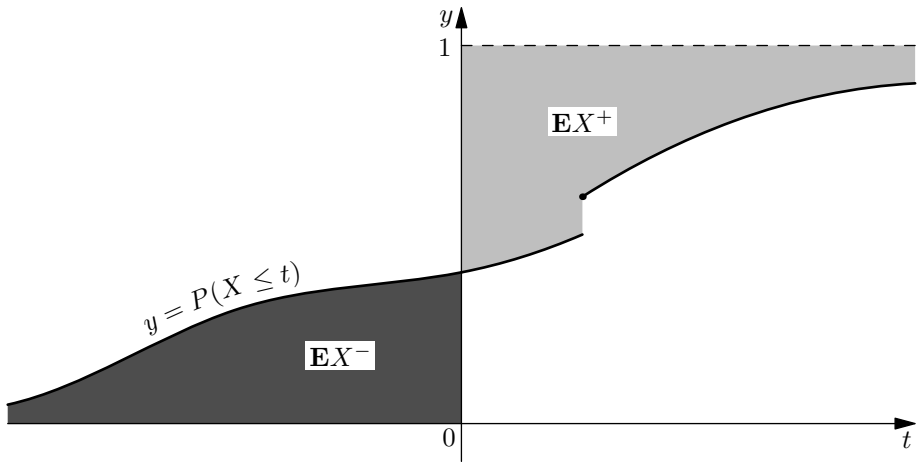


FIGURE 7.9 – Espérance d'une v.a. réelle $\mathbf{E}X := \mathbf{E}(X^+) - \mathbf{E}(X^-)$

En reportant ces égalités dans la définition de $\mathbf{E}X$, on justifie le passage à (7.41) dans le calcul suivant :

$$\begin{aligned}
\mathbf{E}X = \mathbf{E}(X^+) - \mathbf{E}(X^-) &= \int_0^{+\infty} P(X^+ > t)\,\mathrm{d}t - \int_0^{+\infty} P(X^- > t)\,\mathrm{d}t \\
&= \int_0^{+\infty} P(X > t)\,\mathrm{d}t - \int_0^{+\infty} P(X < -t)\,\mathrm{d}t \qquad (7.41) \\
&= \int_0^{+\infty} P(X > t)\,\mathrm{d}t - \int_{-\infty}^{0} P(X < s)\,\mathrm{d}s \qquad (7.42) \\
&= \int_0^{+\infty} \big(1 - F(t)\big)\,\mathrm{d}t - \int_{-\infty}^{0} F(t)\,\mathrm{d}t. \qquad (7.43)
\end{aligned}$$

Le passage de (7.41) à (7.42) résulte bien sûr du changement de variable $s = -t$, cf. la proposition 4.41 ii), et justifie la première égalité dans (7.37). Le passage de (7.42) à (7.43) résulte de la remarque 7.15 et du remplacement de la « variable muette s » par t. □

Proposition 7.32 (espérance d'une v.a. réelle à densité)**.** *Soit X une variable aléatoire réelle de densité f. Alors X est intégrable si et seulement si*

$$\int_{-\infty}^{+\infty} |x| f(x)\,\mathrm{d}x < +\infty \qquad (7.44)$$

et dans ce cas,

$$\mathbf{E}X = \int_{-\infty}^{+\infty} x f(x)\,\mathrm{d}x. \qquad (7.45)$$

Preuve. Nous savons déjà par (7.40) et la remarque 7.11 que

$$\int_0^{+\infty} P(X^+ > t)\,\mathrm{d}t = \int_0^{+\infty} P(X > t)\,\mathrm{d}t = \int_0^{+\infty} xf(x)\,\mathrm{d}x = \int_0^{+\infty} |x|f(x)\,\mathrm{d}x, \quad (7.46)$$

sans aucune condition d'intégrabilité sur X.

Cherchons une formule analogue pour $\int_0^{+\infty} P(X^- > t)\,\mathrm{d}t$. En utilisant (7.40), le passage de (7.41) à (7.42) et la remarque 7.15, on peut démarrer avec :

$$\int_0^{+\infty} P(X^- > t)\,\mathrm{d}t = \int_{-\infty}^0 P(X < t)\,\mathrm{d}t = \int_{-\infty}^0 P(X \leqslant t)\,\mathrm{d}t.$$

On procède alors comme dans la preuve de la proposition 7.10 en injectant dans cette formule l'égalité $P(X \leqslant t) = \int_{-\infty}^t f(x)\,\mathrm{d}x$ et en justifiant l'interversion des intégrations par le théorème de Fubini-Tonelli ou par une simple adaptation de l'exercice 4.11.

$$\begin{aligned}
\int_{-\infty}^0 P(X \leqslant t)\,\mathrm{d}t &= \int_{-\infty}^0 \left\{ \int_{-\infty}^t f(x)\,\mathrm{d}x \right\} \mathrm{d}t \\
&= \int_{-\infty}^0 \left\{ \int_{-\infty}^0 \mathbf{1}_{]-\infty,t]}(x) f(x)\,\mathrm{d}x \right\} \mathrm{d}t \\
&= \int_{-\infty}^0 \left\{ \int_{-\infty}^0 \mathbf{1}_{[x,0]}(t) f(x)\,\mathrm{d}t \right\} \mathrm{d}x \\
&= \int_{-\infty}^0 \left\{ \int_{-\infty}^0 \mathbf{1}_{[x,0]}(t)\,\mathrm{d}t \right\} f(x)\,\mathrm{d}x.
\end{aligned}$$

Comme pour $x \leqslant 0$, $\int_{-\infty}^0 \mathbf{1}_{[x,0]}(t)\,\mathrm{d}t = \int_x^0 \mathrm{d}t = -x$, on obtient finalement :

$$\int_0^{+\infty} P(X^- > t)\,\mathrm{d}t = \int_{-\infty}^0 (-x)f(x)\,\mathrm{d}x = \int_{-\infty}^0 |x|f(x)\,\mathrm{d}x. \quad (7.47)$$

En appliquant l'additivité de l'espérance des v.a. positives à l'égalité $|X| = X^+ + X^-$ et en rassemblant (7.46) et (7.47), on voit que

$$\mathbf{E}|X| = \mathbf{E}(X^+) + \mathbf{E}(X^-) = \int_0^{+\infty} |x|f(x)\,\mathrm{d}x + \int_{-\infty}^0 |x|f(x)\,\mathrm{d}x = \int_{-\infty}^{+\infty} |x|f(x)\,\mathrm{d}x,$$

ce qui établit la condition nécessaire et suffisante d'intégrabilité (7.44). Si cette condition est réalisée, $\mathbf{E}(X^+)$ et $\mathbf{E}(X^-)$ sont finis et

$$\mathbf{E}X = \mathbf{E}(X^+) - \mathbf{E}(X^-) = \int_0^{+\infty} xf(x)\,\mathrm{d}x - \int_{-\infty}^0 (-x)f(x)\,\mathrm{d}x = \int_{-\infty}^{+\infty} xf(x)\,\mathrm{d}x,$$

ce qui établit (7.45). □

Proposition 7.33 (espérance d'une v.a. discrète)**.** *La variable aléatoire discrète X est intégrable si et seulement si*

$$\sum_{x\in X(\Omega)} |x|P(X=x) < +\infty \tag{7.48}$$

et dans ce cas,

$$\mathbf{E}X = \sum_{x\in X(\Omega)} xP(X=x), \tag{7.49}$$

cette somme désignant une série absolument convergente dans le cas où $X(\Omega)$ est infini[17]*.*

Preuve. X étant discrète, $X(\Omega)$ est au plus dénombrable. Notons $B := X(\Omega)\cap]0,+\infty[$ et $B' := X(\Omega)\cap]-\infty,0[$. Alors $X^+(\Omega)$ est égal à B si tous les $x \in X(\Omega)$ sont strictement positifs ou à $B\cup\{0\}$ si l'un au moins d'entre eux est négatif ou nul[18]. Quoiqu'il en soit, on a toujours en utilisant le corollaire 7.27 :

$$\mathbf{E}(X^+) = \sum_{x\in X^+(\Omega)} xP(X^+=x) = \sum_{x\in X^+(\Omega)} xP(X=x) = \sum_{x\in B} xP(X=x), \tag{7.50}$$

puisque pour $x=0$ le terme éventuel $xP(X=x)$ est nul. De même, $X^-(\Omega)$ est égal à $\{-x\ ;\ x\in B'\} =: -B'$ si tous les $x\in X(\Omega)$ sont strictement négatifs ou à $(-B')\cup\{0\}$ si l'un au moins d'entre eux est positif ou nul. En appliquant le corollaire 7.27 à la v.a. positive discrète X^-,

$$\mathbf{E}(X^-) = \sum_{y\in X^-(\Omega)} yP(X^-=y) = \sum_{x\in B'} (-x)P(X=x) = \sum_{x\in B'} |x|P(X=x). \tag{7.51}$$

En rassemblant (7.50) et (7.51), on obtient (noter que $B\cap B' = \varnothing$) :

$$\begin{aligned}\mathbf{E}|X| &= \sum_{x\in B} xP(X=x) + \sum_{x\in B'} |x|P(X=x)\\ &= \sum_{x\in B\cup B'} |x|P(X=x) = \sum_{x\in X(\Omega)} |x|P(X=x),\end{aligned}$$

ce qui nous donne la CNS d'intégrabilité (7.48). Si cette condition est réalisée, $\mathbf{E}(X^+)$ et $\mathbf{E}(X^-)$ sont finis et

$$\begin{aligned}\mathbf{E}X = \mathbf{E}(X^+) - \mathbf{E}(X^-) &= \sum_{x\in B} xP(X=x) - \sum_{x\in B'} (-x)P(X=x)\\ &= \sum_{x\in B\cup B'} xP(X=x)\\ &= \sum_{x\in X(\Omega)} xP(X=x),\end{aligned}$$

ce qui établit (7.49). □

17. $X(\Omega)$ est alors forcément dénombrable puisque X est discrète.
18. Il en résulte que $X^+(\Omega)$ est au plus dénombrable et donc que la v.a. positive X^+ est discrète.

Proposition 7.34 (Linéarité de l'espérance)**.**

a) L'espérance des v.a. réelles intégrables est additive *: si X et Y v.a. réelles définies sur le même $(\Omega, \mathcal{F}, P)$ sont intégrables, alors $X+Y$ l'est aussi et*

$$\mathbf{E}(X+Y) = \mathbf{E}X + \mathbf{E}Y. \tag{7.52}$$

b) Si X est intégrable, cX l'est aussi pour toute constante réelle c et

$$\mathbf{E}(cX) = c\mathbf{E}X. \tag{7.53}$$

Preuve du a). Posons $Z := X + Y$. Alors Z est mesurable et compte-tenu de la croissance et de l'additivité de l'intégrale pour les v.a. positives, l'inégalité $|Z| \leqslant |X| + |Y|$ nous donne :

$$\mathbf{E}|Z| \leqslant \mathbf{E}(|X| + |Y|) = \mathbf{E}|X| + \mathbf{E}|Y| < +\infty.$$

Ainsi Z est intégrable et $\mathbf{E}Z$ bien définie.

L'égalité $Z^+ - Z^- = X^+ - X^- + Y^+ - Y^-$, peut se réécrire comme égalité entre sommes de v.a. positives :

$$Z^+ + X^- + Y^- = Z^- + X^+ + Y^+.$$

Par additivité de l'espérance pour les v.a. positives, on en déduit

$$\mathbf{E}(Z^+) + (\mathbf{E}X^-) + \mathbf{E}(Y^-) = \mathbf{E}(Z^-) + \mathbf{E}(X^+) + \mathbf{E}(Y^+).$$

La finitude de ces 6 espérances (due à l'intégrabilité de X, Y et Z) nous permet alors d'écrire :

$$\mathbf{E}(Z^+) - \mathbf{E}(Z^-) = \mathbf{E}(X^+) - \mathbf{E}(X^-) + \mathbf{E}(Y^+) - \mathbf{E}(Y^-),$$

ce qui donne bien $\mathbf{E}(X+Y) = \mathbf{E}X + \mathbf{E}Y$. □

Preuve du b). L'intégrabilité de cX résulte de l'intégrabilité de X et de la proposition 7.12 appliquée avec la constante positive $|c|$. En effet

$$\mathbf{E}|cX| = \mathbf{E}(|c||X|) = |c|\mathbf{E}|X| < +\infty.$$

Si $c > 0$, on voit que $(cX)^+ = cX^+$ et $(cX)^- = cX^-$. En utilisant la proposition 7.12, on en déduit :

$$\mathbf{E}(cX) = \mathbf{E}\big((cX)^+\big) - \mathbf{E}\big((cX)^-\big) = \mathbf{E}(cX^+) - \mathbf{E}(cX^-) = c\mathbf{E}(X^+) - c\mathbf{E}(X^-) = c\mathbf{E}X.$$

Si $c < 0$, $(cX)^+ = (-c)X^-$ et $(cX)^- = (-c)X^+$, d'où en utilisant la proposition 7.12 avec la constante positive $(-c)$:

$$\begin{aligned} \mathbf{E}(cX) = \mathbf{E}\big((cX)^+\big) - \mathbf{E}\big((cX)^-\big) &= \mathbf{E}\big((-c)X^-\big) - \mathbf{E}\big((-c)X^+\big) \\ &= (-c)\mathbf{E}(X^-) - (-c)\mathbf{E}(X^+) = c\mathbf{E}X. \end{aligned}$$

Le cas $c = 0$ est évident directement. □

Proposition 7.35 (espérance et ordre)**.**

a) L'espérance des v.a. réelles intégrables est croissante *: si X et Y v.a. réelles définies sur le même $(\Omega, \mathcal{F}, P)$ sont intégrables et vérifient $X \leqslant Y$, c'est-à-dire pour tout $\omega \in \Omega$, $X(\omega) \leqslant Y(\omega)$, alors $\mathbf{E}X \leqslant \mathbf{E}Y$.*

b) Si X est intégrable, $|X|$ l'est aussi et

$$\left|\mathbf{E}X\right| \leqslant \mathbf{E}|X|. \tag{7.54}$$

Preuve. Pour le a), il suffit de noter que si $X \leqslant Y$, $Y - X$ est une v.a. *positive* et donc $\mathbf{E}(Y - X) \geqslant 0$. Comme X et Y sont intégrables, par linéarité de l'espérance, $\mathbf{E}(Y - X) = \mathbf{E}Y - \mathbf{E}X$. Ainsi la positivité de $Y - X$ et la linéarité de $\mathbf{E}$ impliquent $\mathbf{E}Y - \mathbf{E}X \geqslant 0$ c'est-à-dire $\mathbf{E}Y \geqslant \mathbf{E}X$.

Une fois la croissance de $\mathbf{E}$ ainsi établie, on en déduit le b) en partant de l'encadrement $-|X| \leqslant X \leqslant |X|$ qui implique $\mathbf{E}(-|X|) \leqslant \mathbf{E}X \leqslant \mathbf{E}|X|$. Par linéarité $\mathbf{E}(-|X|) = -\mathbf{E}|X|$, ce qui permet de réécrire l'encadrement ci-dessus sous la forme :

$$-\mathbf{E}|X| \leqslant \mathbf{E}X \leqslant \mathbf{E}|X|,$$

qui équivaut à (7.54). □

7.4 Moments

On étudie dans cette section les $\mathbf{E}h(X)$, où h est une fonction réelle. Pour que l'expression $\mathbf{E}h(X)$ ait un sens, il est nécessaire que $Y := h(X)$ soit une variable aléatoire réelle. Cette condition sera réalisée si $h : \mathbb{R} \to \mathbb{R}$ est *borélienne*, c'est-à-dire si $B' = h^{-1}(B) \in \mathrm{Bor}(\mathbb{R})$ pour tout $B \in \mathrm{Bor}(\mathbb{R})$. Alors en effet,

$$Y^{-1}(B) = \{\omega \in \Omega\ ;\ h(X(\omega)) \in B\} = \{\omega \in \Omega\ ;\ X(\omega) \in h^{-1}(B)\} = X^{-1}(B') \in \mathcal{F},$$

en raison de la mesurabilité $\mathcal{F} - \mathrm{Bor}(\mathbb{R})$ de la variable aléatoire X. Comme le borélien B ci-dessus est quelconque, ceci montre que Y est elle aussi mesurable $\mathcal{F} - \mathrm{Bor}(\mathbb{R})$ et est donc bien une variable aléatoire sur $(\Omega, \mathcal{F})$. Ainsi si h est borélienne, $\mathbf{E}|h(X)|$ existe toujours comme élément de $\overline{\mathbb{R}}_+$ et si $\mathbf{E}|h(X)| < +\infty$, $\mathbf{E}h(X)$ existe (et $\mathbf{E}h(X) \in \mathbb{R}$). On pourra désigner $\mathbf{E}|h(X)|$ et $\mathbf{E}h(X)$ respectivement par l'appellation h-moment absolu de X et h-moment de X. Bien entendu si h est borélienne positive, h-moment absolu et h-moment sont confondus et ce dernier existe toujours dans $\overline{\mathbb{R}}_+$. Nous utiliserons aussi l'appelation générique de *moments fonctionnels* pour désigner les h-moments [19].

Le cas le plus utile est celui où h est une fonction puissance, $h(x) = x^r$, on parle alors de moment d'ordre r de X.

Définition 7.36. *Soit r un réel positif. On appelle* moment absolu d'ordre r *de la variable aléatoire réelle X la quantité $\mathbf{E}(|X|^r)$ (élément de $\overline{\mathbb{R}}_+$). Si r est entier et*

19. Ces appellations ne sont pas standard, nous les adoptons par confort de rédaction.

X^r intégrable (donc si le moment absolu d'ordre r de X est fini*), on appelle* moment d'ordre r *de X le réel $\mathbf{E}(X^r)$. On notera $\mathbf{E}|X|^r$ pour $\mathbf{E}(|X|^r)$ et $\mathbf{E}X^r$ pour $\mathbf{E}(X^r)$ en prenant garde de ne pas confondre ces quantités avec $(\mathbf{E}|X|)^r$ et $(\mathbf{E}X)^r$ respectivement.*

Remarquons qu'on ne définit pas le moment d'ordre r *non entier* pour X v.a. réelle, même si $\mathbf{E}|X|^r < +\infty$. En effet dans ce cas, X^r n'est pas définie sur l'évènement $\{X < 0\}$ et si $P(X < 0) \neq 0$, X^r ne peut être égale presque sûrement à une v.a. définie sur tout Ω. Bien entendu, si X est une v.a. positive, $\mathbf{E}X^r$ existe toujours dans $\overline{\mathbb{R}}_+$.

Proposition 7.37. *Si la variable aléatoire X a un moment absolu d'ordre r fini, elle a aussi un moment absolu d'ordre p fini pour tout $p \in [0, r]$.*

Preuve. Si $0 \leqslant p \leqslant r$, on a les inégalités suivantes entre variables aléatoires positives :

$$|X|^p \leqslant \mathbf{1}_{\{|X|\leqslant 1\}} + |X|^r \mathbf{1}_{\{|X|>1\}} \leqslant 1 + |X|^r,$$

d'où

$$\mathbf{E}(|X|^p) \leqslant \mathbf{E}(1) + \mathbf{E}(|X|^r) = 1 + \mathbf{E}(|X|^r) < +\infty.$$

par croissance de l'espérance des v.a. positives. □

L'existence d'un moment absolu d'ordre r fini donne un renseignement sur la vitesse de convergence vers 0 de $P(|X| \geqslant t)$ quand t tend vers $+\infty$. On a alors $P(|X| \geqslant t) = O(t^{-r})$ par le corollaire suivant de l'inégalité de Markov.

Proposition 7.38 (inégalité de Markov avec moment)**.** *Pour toute variable aléatoire réelle X, pour tout réel $r > 0$,*

$$\forall t > 0, \quad P(|X| \geqslant t) \leqslant \frac{\mathbf{E}|X|^r}{t^r}. \tag{7.55}$$

Bien entendu, cette inégalité n'a d'intérêt que si $\mathbf{E}|X|^r < +\infty$ et $t^{-r}\mathbf{E}|X|^r < 1$.

Preuve. Il suffit de noter que $P(|X| \geqslant t) \leqslant P(|X|^r \geqslant t^r) \leqslant t^{-r}\mathbf{E}|X|^r$, en utilisant la croissance de l'application $\mathbb{R}_+ \to \mathbb{R}_+$, $x \mapsto x^r$ et l'inégalité de Markov pour la v.a. positive $|X|^r$. □

Proposition 7.39 (moments d'une v.a. discrète)**.** *Si X est une variable aléatoire discrète, pour tout réel $r \geqslant 0$,*

$$\mathbf{E}|X|^r = \sum_{x \in X(\Omega)} |x|^r P(X = x). \tag{7.56}$$

De plus si r est entier et $\mathbf{E}|X|^r < +\infty$,

$$\mathbf{E}X^r = \sum_{x \in X(\Omega)} x^r P(X = x). \tag{7.57}$$

Cette proposition n'est qu'une application de la formule de calcul de $\mathbf{E}h(X)$ pour X discrète qui est établie dans toute sa généralité à la proposition 7.41 ci-dessous.

Proposition 7.40 (moments d'une v.a. à densité)**.** *Si X est une variable aléatoire réelle à densité f, pour tout réel $r \geqslant 0$,*

$$\mathbf{E}|X|^r = \int_{-\infty}^{+\infty} |x|^r f(x)\,\mathrm{d}x. \tag{7.58}$$

Si de plus r est entier et $\mathbf{E}|X|^r < +\infty$,

$$\mathbf{E}X^r = \int_{-\infty}^{+\infty} x^r f(x)\,\mathrm{d}x. \tag{7.59}$$

Là aussi, il s'agit d'un cas particulier d'une formule générale pour $\mathbf{E}h(X)$ lorsque X est à densité, donnée ci-dessous. La démonstration de cette formule générale étant assez ardue, nous invitons le lecteur à démontrer directement en exercice la proposition 7.40. La preuve est grandement facilitée par la monotonie de $h : x \mapsto |x|^r$ sur chacun des ensembles $\mathbb{R}_-$ et $\mathbb{R}_+$.

Proposition 7.41 (calcul de $\mathbf{E}h(X)$, X discrète)**.** *Si X est une variable aléatoire discrète et $h : \mathbb{R} \to \mathbb{R}$ une application borélienne, c.-à-d. mesurable* $\mathrm{Bor}(\mathbb{R})$–$\mathrm{Bor}(\mathbb{R})$*,*

$$\mathbf{E}|h(X)| = \sum_{x \in X(\Omega)} |h(x)| P(X = x). \tag{7.60}$$

De plus, si $\mathbf{E}|h(X)| < +\infty$,

$$\mathbf{E}h(X) = \sum_{x \in X(\Omega)} h(x) P(X = x). \tag{7.61}$$

Preuve. Posons $Y = |h(X)|$. L'ensemble $X(\Omega) = \{x_0, x_1, \ldots\}$ est au plus dénombrable. L'application $|h|$ n'étant pas supposée injective, il peut y avoir des répétitions dans la suite des $|h(x_k)|$. L'ensemble $Y(\Omega)$ qui peut s'écrire en effaçant toutes les répétitions de cette suite est lui-même au plus dénombrable. La variable aléatoire $Y = |h(X)|$ étant discrète positive, la formule établie au corollaire 7.27 nous donne :

$$\mathbf{E}|h(X)| = \mathbf{E}Y = \sum_{y \in Y(\Omega)} y P(Y = y). \tag{7.62}$$

Pour chaque réel y de $Y(\Omega)$, notons B_y l'ensemble de ses antécédents par $|h|$:

$$B_y = \{x \in X(\Omega)\ ;\ |h(x)| = y\}, \qquad y \in Y(\Omega).$$

Ce sous-ensemble de $X(\Omega)$ contient au moins un élément et est au plus dénombrable. On peut alors décomposer l'évènement $\{Y = y\}$ en réunion disjointe d'une famille au plus dénombrable :

$$\{Y = y\} = \bigcup_{x \in B_y} \{X = x\}.$$

Le terme général de la somme (7.62) peut donc s'écrire :

$$yP(Y=y) = y\sum_{x\in B_y} P(X=x) = \sum_{x\in B_y} yP(X=x) = \sum_{x\in B_y} |h(x)|P(X=x).$$

Comme les B_y forment une partition de $X(\Omega)$, la propriété de sommation par paquets des familles à termes *positifs* nous donne :

$$\sum_{x\in X(\Omega)} |h(x)|P(X=x) = \sum_{y\in Y(\Omega)}\sum_{x\in B_y} |h(x)|P(X=x) = \sum_{y\in Y(\Omega)} yP(Y=y),$$

ce qui, compte-tenu de (7.62), établit la formule (7.60) pour $\mathbf{E}|h(X)|$.

Supposons maintenant que $\mathbf{E}|h(X)| < +\infty$. Grâce à (7.60), la famille de réels $\{h(x)P(X=x)\,;\, x\in X(\Omega)\}$ est *sommable*. On peut alors reprendre le calcul fait ci-dessus en remplaçant $|h|$ par h et utiliser la propriété de sommation par paquets des familles sommables pour aboutir à la formule (7.61) de calcul de $\mathbf{E}h(X)$. □

Proposition 7.42 (calcul de $\mathbf{E}h(X)$, X à densité)**.** *Si X est une variable aléatoire de densité f et $h:\mathbb{R}\to\mathbb{R}$ une application réglée sur tout intervalle fermé borné de $\mathbb{R}$,*

$$\mathbf{E}|h(X)| = \int_{-\infty}^{+\infty} |h(x)|f(x)\,\mathrm{d}x. \tag{7.63}$$

De plus, si $\mathbf{E}|h(X)| < +\infty$,

$$\mathbf{E}h(X) = \int_{-\infty}^{+\infty} h(x)f(x)\,\mathrm{d}x. \tag{7.64}$$

Une application h est réglée sur $[a,b]$ si elle est limite uniforme sur $[a,b]$ d'une suite de fonctions en escaliers [20]. On démontre en analyse que h est réglée sur $[a,b]$ si et seulement si elle admet en tout point de $]a,b[$ une limite à gauche et une limite à droite (finies) ainsi qu'une limite finie à droite en a et à gauche en b. La classe des fonctions réglées, sans être aussi grande que celle des fonctions boréliennes, devrait donc suffire à nos besoins. Elle contient en particulier les fonctions continues et les fonctions monotones par morceaux.

Schéma de la preuve. On procède en 5 étapes.

1. On examine d'abord le cas de h en escaliers et positive sur $[a,b]$, nulle en dehors de $[a,b]$. La fonction h peut alors s'écrire $h=\sum_{i=1}^m y_i\mathbf{1}_{A_i}$, où les y_i sont des réels positifs et les A_i sont des intervalles formant une partition de $[a,b]$. Par linéarité, on se ramène au cas encore plus simple où $h=\mathbf{1}_{A_i}$, lequel est évident, puisque $\int_{-\infty}^{+\infty}\mathbf{1}_{A_i}(x)f(x)\,\mathrm{d}x = P(X\in A_i) = \mathbf{E}\mathbf{1}_{A_i}(X)$.

20. On en déduit ici la mesurabilité de h en montrant qu'elle est limite simple *sur* $\mathbb{R}$ d'une suite de fonctions en escalier, donc boréliennes. Ainsi $h(X)$ est bien une variable aléatoire car mesurable $\mathcal{F}$ – Bor($\mathbb{R}$) par composition.

2. On traite ensuite le cas où h est réglée et positive sur un intervalle *fixé* $[a,b]$, et nulle en dehors de cet intervalle. Elle est donc limite uniforme sur $[a,b]$ d'une suite $(g_k)_{k\geqslant 1}$ de fonctions en escaliers positives sur $[a,b]$ et nulles en dehors de $[a,b]$. Comme les g_k et h sont nulles en dehors de $[a,b]$, cette convergence est *uniforme* sur $\mathbb{R}$ et on en déduit (exercice) que $\mathbf{E}h(X) = \lim_{k\to\infty} \mathbf{E}g_k(X)$. Notons au passage que h est bornée, donc que la variable aléatoire positive $h(X)$ a une espérance finie. D'autre part la convergence uniforme sur $[a,b]$ de g_k vers h permet de légitimer l'interversion limite intégrale $\int_a^b h(x)f(x)\,\mathrm{d}x = \int_a^b \lim_{k\to+\infty} g_k(x)f(x)\,\mathrm{d}x = \lim_{k\to+\infty}\int_a^b g_k(x)f(x)\,\mathrm{d}x = \lim_{k\to+\infty}\mathbf{E}g_k(X)$ et on obtient ainsi (7.64) pour h réglée positive sur $[a,b]$ et nulle en dehors [21].
3. Pour h positive réglée sur tout $[a,b]\subset\mathbb{R}$, on pose $h_n := h\mathbf{1}_{[-n,n]}$ et on note que $h_n \uparrow h$. On étend alors le résultat du cas 2 (valable pour chaque h_n) en utilisant le théorème de Beppo Levi (Th. 7.23).
4. En appliquant ce qui précède avec $|h|$ au lieu de h, on obtient (7.63) pour h réglée sur tout $[a,b]\subset\mathbb{R}$ sans condition de positivité.
5. On obtient enfin (7.64) en écrivant $h = h^+ - h^-$, en utilisant le cas 3 et en recollant les morceaux, sans problème puisqu'ici les intégrales généralisées et les espérances concernées sont toutes finies.

□

Nous présentons maintenant une formule injustement méconnue, permettant de calculer des moments fonctionnels à partir de la fonction de survie. Son intérêt est de permettre un tel calcul pour des variables aléatoires qui ne sont ni discrètes ni à densité.

Proposition 7.43. *Soient X une variable aléatoire positive et g une application continue strictement croissante $\mathbb{R}_+\to\mathbb{R}$, de classe C^1 sur $\mathbb{R}_+^*$. Alors*

$$\mathbf{E}g(X) = g(0) + \int_0^{+\infty} P(X>s)g'(s)\,\mathrm{d}s. \tag{7.65}$$

La preuve de cette formule utilise le lemme suivant, valide sous des hypothèses plus générales pour g.

Lemme 7.44. *Soit $g:[0,+\infty[\to\mathbb{R}$, continue et strictement croissante. Notons ℓ la limite de g en $+\infty$. Alors pour toute variable aléatoire positive X,*

$$\forall t\in[\,g(0),\ell[,\quad P\big(g(X)>t\big) = P\big(X>g^{-1}(t)\big), \tag{7.66}$$

Preuve du lemme. En raison de sa croissance, g a une limite ℓ en $+\infty$, qui est soit un réel, soit $+\infty$. En raison de sa continuité et de sa croissance stricte, g réalise une bijection de $[0,+\infty[$ sur $[g(0),\ell[$ et la réciproque g^{-1} de cette bijection est définie et

21. Attention, l'interversion limite intégrale ne s'obtient pas directement par convergence uniforme de g_kf vers hf, car f n'est pas forcément définie sur tout $[a,b]$ ni bornée.

strictement croissante sur $[g(0), \ell[$. Pour prouver (7.66), nous allons établir l'égalité des évènements $A := \{g(X) > t\}$ et $B := \{X > g^{-1}(t)\}$ en vérifiant l'inclusion dans les deux sens.

Pour tout $\omega \in A$, $g(X(\omega)) > t$ et, comme les deux membres de cette inégalité appartiennent à $[g(0), \ell[$ qui est l'ensemble de définition de l'application strictement croissante g^{-1}, on en déduit que $g^{-1}\big(g(X(\omega))\big) > g^{-1}(t)$, ce qui s'écrit aussi $X(\omega) > g^{-1}(t)$. Ainsi tout ω dans A appartient aussi à B, autrement dit, $A \subset B$.

En échangeant les rôles de g et g^{-1}, le même argument montre que $B \subset A$. Finalement $A = B$ et $P(A) = P(B)$. □

Preuve de la proposition 7.43. Par croissance de g, la fonction $g - g(0)$ est positive et il est clair que si (7.65) est vraie pour $g - g(0)$, elle l'est aussi pour g. On ne perd donc pas de généralité en supposant désormais que g est positive.

Pour établir (7.65), on définit la variable aléatoire positive $Y = g(X)$ et on part de la définition $\mathbf{E}Y = \int_0^{+\infty} P(Y > t)\,\mathrm{d}t$ en découpant $\int_0^{+\infty} = \int_0^{g(0)} + \int_{g(0)}^{\ell}$ si $\ell = +\infty$, ou $\int_0^{+\infty} = \int_0^{g(0)} + \int_{g(0)}^{\ell} + \int_{\ell}^{+\infty}$ si $\ell < +\infty$.

Remarquons d'abord que si $\ell < +\infty$, l'application strictement croissante g ne peut prendre que des valeurs strictement inférieures à ℓ et par conséquent pour tout $t \geqslant \ell$, $\{g(X) > t\} = \varnothing$, d'où $P(Y > t) = P(\varnothing) = 0$ et $\int_{\ell}^{+\infty} P(Y > t)\,\mathrm{d}t = 0$. Que ℓ soit fini ou non, on a donc dans tous les cas :

$$\mathbf{E}Y = \int_0^{g(0)} P(Y > t)\,\mathrm{d}t + \int_{g(0)}^{\ell} P(Y > t)\,\mathrm{d}t.$$

Examinons la première intégrale. Par croissance de g, $Y(\omega) = g(X(\omega)) \geqslant g(0)$ pour tout $\omega \in \Omega$ et donc pour tout $t \in [0, g(0)[$, $1 \geqslant P(Y > t) \geqslant P(Y \geqslant g(0)) = 1$. La fonction $t \mapsto P(Y > t)$ est donc constante égale à 1 sur $[0, g(0)[$ et sa valeur en $g(0)$ est un réel[22] de $[0, 1]$. L'intégrale de Riemann $\int_0^{g(0)} P(Y > t)\,\mathrm{d}t$ vaut donc $\int_0^{g(0)} 1\,\mathrm{d}t = g(0)$. À ce stade nous avons donc établi la formule :

$$\mathbf{E}g(X) = g(0) + \int_{g(0)}^{\ell} P(Y > t)\,\mathrm{d}t.$$

En utilisant (7.66) on peut écrire cette dernière intégrale sous la forme :

$$\int_{g(0)}^{\ell} P(Y > t)\,\mathrm{d}t = \int_{g(0)}^{\ell} P\big(X > g^{-1}(t)\big)\,\mathrm{d}t.$$

On peut alors appliquer le changement de variable C^1 et croissant[23] (cf. prop. 3.34)

22. Cette valeur peut être strictement inférieure à 1 si $P(X = 0) \neq 0$. Mais en tout état de cause, elle n'a aucune influence sur la valeur de l'intégrale de Riemann $\int_0^{g(0)} P(Y > t)\,\mathrm{d}t$.

23. Pour le justifier proprement, on peut d'abord faire le changement de variable sur $\int_{g(0)}^{g(b)}$, puis faire tendre b vers $+\infty$.

$t = g(s)$ ou $s = g^{-1}(t)$ qui donne :

$$\int_{g(0)}^{\ell} P(Y > t)\,\mathrm{d}t = \int_0^{+\infty} P(X > s)g'(s)\,\mathrm{d}s$$

et achève la vérification de (7.65). □

Le h-moment $\mathbf{E}h(X)$ pour $h : x \mapsto (x - \mathbf{E}X)^2$ occupe une place particulière dans la théorie des probabilités.

Définition 7.45 (variance et écart type)**.** *Si X est de carré intégrable ($\mathbf{E}X^2 < +\infty$), on appelle* variance *de X le réel positif noté* $\operatorname{Var} X$ *défini par*

$$\operatorname{Var} X := \mathbf{E}(X - \mathbf{E}X)^2. \tag{7.67}$$

On appelle alors écart type *de X le réel* $\sigma(X) := (\operatorname{Var} X)^{1/2}$.

Remarquons que si $\mathbf{E}X^2$ est fini, $\mathbf{E}|X|$ l'est aussi (proposition 7.37), donc $\mathbf{E}X$ est bien défini. De plus $(X - \mathbf{E}X)^2 = X^2 - 2(\mathbf{E}X)X + (\mathbf{E}X)^2$ apparaît alors comme une combinaison linéaire de trois variables [24] intégrables, donc est aussi intégrable. Ainsi la v.a. positive $(X - \mathbf{E}X)^2$ est intégrable et $\mathbf{E}(X - \mathbf{E}X)^2$ est bien un réel positif, ce qui justifie la définition 7.45. Notons aussi que si X représente une grandeur physique, X, $\mathbf{E}X$ et $\sigma(X)$ ont la même unité, mais pas $\operatorname{Var} X$.

Lorsqu'elle existe, la variance de X est une façon de mesurer la *dispersion* de la loi de X autour de l'espérance. Les raisons de l'importance de la variance apparaîtront ultérieurement dans cet ouvrage (inégalité de Tchebycheff, théorème limite central). L'application des propositions 7.41 et 7.42 nous donne (sous réserve d'intégrabilité de X^2) les formules respectives :

$$\operatorname{Var} X = \sum_{x \in X(\Omega)} (x - \mathbf{E}X)^2 P(X = x) \quad (\text{cas } X \text{ discrète}), \tag{7.68}$$

$$\operatorname{Var} X = \int_{-\infty}^{+\infty} (x - \mathbf{E}X)^2 f(x)\,\mathrm{d}x \quad (\text{cas } X \text{ à densité } f). \tag{7.69}$$

Dans la pratique, ces formules sont rarement utilisées, on leur préfère la formule suivante qui simplifie les calculs.

Proposition 7.46 (formule de Koenig pour la variance)**.** *Si la variable aléatoire X est de carré intégrable,*

$$\operatorname{Var} X = \mathbf{E}X^2 - (\mathbf{E}X)^2. \tag{7.70}$$

Preuve. Rappelons que nous notons $\mathbf{E}X^2$ pour $\mathbf{E}(X^2)$ et que le second membre de la formule ci-dessus n'est donc généralement pas nul. On pose $c = \mathbf{E}X$.

$$\begin{aligned} \operatorname{Var} X &= \mathbf{E}(X - c)^2 = \mathbf{E}(X^2 - 2cX + c^2) \\ &= \mathbf{E}X^2 - 2c\mathbf{E}X + \mathbf{E}c^2 \\ &= \mathbf{E}X^2 - 2c^2 + c^2 = \mathbf{E}X^2 - c^2, \end{aligned}$$

24. À savoir X^2, X et la v.a. constante $(\mathbf{E}X)^2$.

en utilisant la linéarité de l'espérance et l'espérance d'une constante. □

Proposition 7.47 (translation et changement d'échelle)**.** *Si* $\mathbf{E}X^2 < +\infty$*,*

$$\forall a \in \mathbb{R},\ \forall b \in \mathbb{R}, \quad \operatorname{Var}(aX+b) = a^2 \operatorname{Var} X, \quad \sigma(aX+b) = |a|\sigma(X). \tag{7.71}$$

Preuve. En utilisant la définition 7.45, la linéarité de l'espérance et le fait que l'espérance de la constante b est b :

$$\begin{aligned} \operatorname{Var}(aX+b) &= \mathbf{E}\big(aX+b-\mathbf{E}(aX+b)\big)^2 = \mathbf{E}(aX+b-a\mathbf{E}X-b)^2 \\ &= \mathbf{E}\big(a(X-\mathbf{E}X)\big)^2 = \mathbf{E}\big(a^2(X-\mathbf{E}X)^2\big) \\ &= a^2\mathbf{E}\big((X-\mathbf{E}X)^2\big) = a^2 \operatorname{Var} X, \end{aligned}$$

ce qui nous donne les formules (7.71). □

Il est clair, d'après la définition de la variance, que la variance d'une constante est nulle. La réciproque est *presque* vraie :

Proposition 7.48 (nullité de la variance et constance p.s.)**.**

$$\operatorname{Var} X = 0 \Leftrightarrow X = \mathbf{E}X \text{ p.s.} \Leftrightarrow X \textit{ est presque sûrement constante.} \tag{7.72}$$

Preuve. Les implications de droite à gauche dans (7.72) sont déjà acquises. On sait en effet que l'espérance d'une constante est cette constante et que si la v.a. $Y := (X-\mathbf{E}X)^2$ vaut 0 avec probabilité 1, son espérance est nulle [25].

Pour la première implication de gauche à droite, il suffit d'appliquer le corollaire 7.18 à la v.a. positive Y. La deuxième implication est triviale. □

7.5 Exercices

Dans certains des exercices de ce chapitre, on fait référence à une suite $X_1, \ldots, X_n$ de variable aléatoires indépendantes. Cette notion sera étudiée en détail au chapitre 8. Pour traiter les exercices, tout ce que l'on a besoin de savoir à ce sujet est que l'indépendance des X_i implique (en fait équivaut à) l'indépendance mutuelle des n évènements $\{X_1 \in I_1\}$, …, $\{X_n \in I_n\}$ pour tous intervalles $I_1, \ldots, I_n$ de $\mathbb{R}$.

Ex. 7.1 *Intégrabilité*

Soit X une variable aléatoire réelle de fonction de répartition F. Démontrez que X est intégrable si et seulement si l'intégrale généralisée $\int_{-\infty}^{\infty} F(t)(1-F(t))\,\mathrm{d}t$ converge dans $\mathbb{R}_+$.

Ex. 7.2 *Contrôleur contre fraudeur*

Une compagnie de métro pratique les tarifs suivants. Le ticket donnant droit à un trajet coûte 1 €; les amendes sont fixées à 20 € pour la première infraction constatée,

25. En effet, pour tout $t \geqslant 0$, $P(Y > t) = 0$, donc Y étant positive, $\mathbf{E}Y = \int_0^{+\infty} P(Y > t)\,\mathrm{d}t = 0$.

40 € pour la deuxième et 400 € pour la troisième. La probabilité p pour un voyageur d'être contrôlé au cours d'un trajet est supposée constante et connue de la seule compagnie. Un fraudeur décide de prendre systématiquement le métro sans payer jusqu'à la deuxième amende et d'arrêter alors de frauder. On note T le nombre de trajets effectués jusqu'à la deuxième amende (T est le numéro du trajet où le fraudeur est contrôlé pour la deuxième fois). On note $q = 1 - p$ la probabilité de faire un trajet sans contrôle.

1) Montrer que la loi de T est donnée par

$$P(T = k) = (k-1)p^2q^{k-2}, \quad k \geqslant 2.$$

2) Pour $n \in \mathbb{N}^*$, calculer $P(T > n)$. *Indication* : on pourra commencer par chercher une formule explicite pour la somme de la série entière

$$f(x) := \sum_{k=n+1}^{+\infty} x^{k-1},$$

puis pour sa dérivée terme à terme.

3) Calculer numériquement $P(T > 60)$ (pourquoi s'intéresse-t-on à cette quantité ?) lorsque $p = 1/10$ et lorsque $p = 1/20$.

4) Calculez en fonction de p la quantité

$$\mathbf{E}T := \sum_{k=2}^{+\infty} kP(T = k).$$

Cette quantité est l'espérance de T et peut s'interpréter comme la valeur moyenne du temps d'attente.

5) D'un point de vue purement financier (et donc hors de toute considération de moralité), quel conseil donneriez vous au fraudeur ? N.B. ceci est une question ouverte et il y a plusieurs réponses valables possibles.

Solution p. 462

Ex. 7.3 *Franchise et plafond*

En cas d'incendie d'un certain type de logement, la loi du coût X des « dommages aux biens » a une fonction de survie G de la forme

$$\forall t \geqslant 0, \quad G(t) := P(X > t) = \frac{1}{(1+t)^a},$$

où a est un paramètre positif.

Une compagnie d'assurances propose la couverture de ce risque « dommages aux biens », par un contrat qui prévoit une franchise m et un plafond M. La franchise est destinée à éviter les déclarations de petits sinistres et ainsi à économiser sur les frais de gestion. Le plafond M limite la responsabilité de la compagnie. Notons Y le

remboursement perçu par l'assuré en cas d'incendie. La règle franchise-plafond nous permet d'écrire [26]

$$\forall \omega \in \Omega, \quad Y(\omega) = \begin{cases} 0 & \text{si } X(\omega) \leqslant m, \\ X(\omega) & \text{si } m < X(\omega) \leqslant M, \\ M & \text{si } X(\omega) > M. \end{cases}$$

On note $H : [0, +\infty[\to [0, 1]$, $t \mapsto H(t) := P(Y > t)$ la fonction de survie de Y. Sa représentation graphique est esquissée à la figure 7.10.

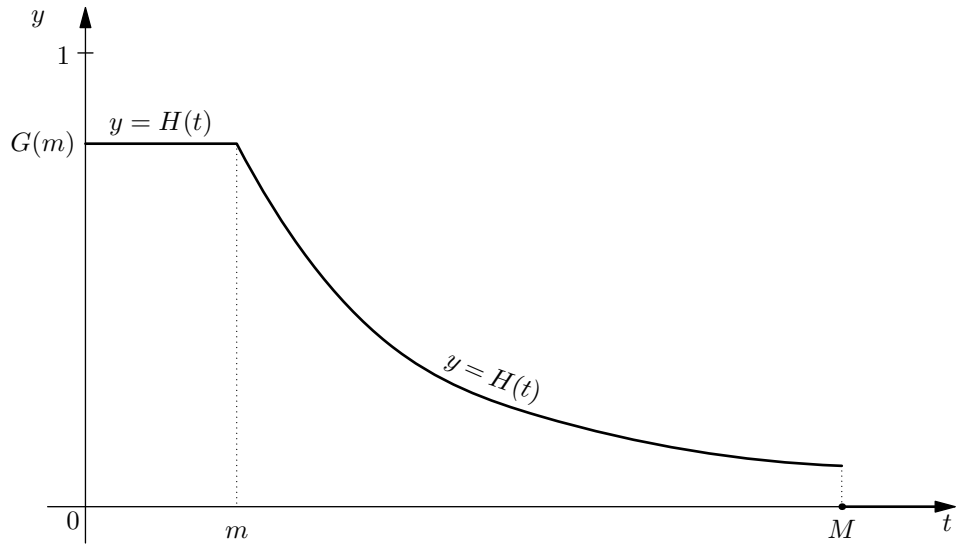


FIGURE 7.10 – Fonction de survie H de la v.a. positive Y

1) Justifiez cette représentation graphique en exprimant $H(t)$ à l'aide de G dans les trois cas $0 \leqslant t < m$, $m \leqslant t < M$ et $t \geqslant M$.

2) La loi de la variable aléatoire positive Y est-elle à densité ?

3) Calculer $\mathbf{E}Y$ en fonction de m, M et a.

Ex. 7.4 *La formule de Poincaré*

La formule de Poincaré (prop. 5.18) a été démontrée par récurrence. Le but de cet exercice est d'en proposer une autre démonstration basée sur les propriétés de l'espérance.

1) Nous aurons besoin de la formule algébrique suivante où $x_1, \ldots, x_n$ sont des

26. Pour simplifier, on suppose que le seul but de la franchise est d'éviter les déclarations de petits sinistres. Donc quand l'assuré est remboursé, on ne déduit pas m de son remboursement.

nombres réels quelconques ($n \geqslant 2$).

$$\prod_{i=1}^{n}(1-x_i) = \sum_{I\subset[\![1,n]\!]} (-1)^{\operatorname{card} I} \prod_{i\in I} x_i. \tag{7.73}$$

Dans cette formule la collection de tous les sous-ensembles I de $[\![1,n]\!]$ qui indexe le $\sum$ comprend en particulier $I = \varnothing$ pour lequel on pose par convention $\prod_{i\in\varnothing} x_i := 1$.

a) Vérifiez directement cette formule dans le cas $n = 3$ en développant le produit $(1-x_1)(1-x_2)(1-x_3)$.

b) Démontrez la formule (7.73).

2) Dans toute la suite, on travaille sur un espace probabilisé $(\Omega, \mathcal{F}, P)$ fixé. Soient $B_1, \ldots, B_k$, k sous-ensembles de Ω. Vérifiez l'égalité

$$\mathbf{1}_B = \prod_{j=1}^{k} \mathbf{1}_{B_j}, \quad \text{où } B := \bigcap_{j=1}^{k} B_j. \tag{7.74}$$

3) Montrez que pour tout entier $n > 1$ et toute suite finie $A_1, \ldots, A_n$ dans $\mathcal{F}$,

$$P\left(\bigcup_{i=1}^{n} A_i\right) = 1 - \mathbf{E}\left(\prod_{i=1}^{n}(1-\mathbf{1}_{A_i})\right). \tag{7.75}$$

4) En utilisant tout ce qui précède, établir la *formule de Poincaré*, à savoir :

$$\forall A_1, \ldots, A_n \in \mathcal{F}, \qquad P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{\substack{I\subset[\![1,n]\!]\\ I\neq\varnothing}} (-1)^{1+\operatorname{card} I} P\Big(\bigcap_{i\in I} A_i\Big). \tag{7.76}$$

Indiquez clairement quelles propriétés de l'espérance mathématique sont utilisées dans cette démonstration.

Ex. 7.5 Soit X une variable aléatoire discrète à valeurs dans $\mathbb{N}^*$. Pour tout $k \in \mathbb{N}^*$, on note $p_k = P(X = k)$. On suppose que X a une espérance mathématique $\mathbf{E}X$ finie et que la suite $(p_k)_{k\geqslant 1}$ est *décroissante* sur $\mathbb{N}^*$.

1) Démontrer l'inégalité :

$$\forall k \in \mathbb{N}^*, \quad P(X = k) < \frac{2\mathbf{E}X}{k^2}. \tag{7.77}$$

Indication : Considérer la somme partielle de rang k de la série définissant $\mathbf{E}X$.

2) Est-il possible qu'il existe une constante $c > 0$ et un entier k_0 tels que :

$$\forall k \geqslant k_0, \quad P(X = k) \geqslant \frac{c\mathbf{E}X}{k^2}? \tag{7.78}$$

3) Donner un exemple où (7.77) est en défaut si on supprime l'hypothèse de décroissance de $(p_k)_{k\geqslant 1}$.

Solution p. 465

Ex. 7.6 *Espérance du min et du max de deux v.a. uniformes*

Soient X et Y deux variables aléatoires *positives* définies sur le même espace probabilisé $(\Omega, \mathcal{F}, P)$, toutes deux de loi uniforme sur $[0,1]$. Sauf dans la dernière question, on ne suppose pas X et Y indépendantes. On s'intéresse à l'espérance des variables aléatoires $\min(X,Y)$ et $\max(X,Y)$.

1) Calculez $\mathbf{E}X$ et $\mathbf{E}(X^2)$.

2) Écrivez plus simplement la somme $\min(X,Y) + \max(X,Y)$ et en déduire une relation entre $\mathbf{E}\min(X,Y)$ et $\mathbf{E}\max(X,Y)$. En déduire l'expression de $\mathbf{E}|X-Y|$ en fonction de $\mathbf{E}\min(X,Y)$, puis en fonction de $\mathbf{E}\max(X,Y)$.

3) Montrez (en deux lignes maximum) que l'on a toujours

$$\mathbf{E}\min(X,Y) \leqslant \frac{1}{2} \leqslant \mathbf{E}\max(X,Y).$$

Est-il possible avec un choix convenable de X et Y que ces deux inégalités soient des égalités ?

4) On suppose maintenant que X et Y sont indépendantes, ce qui implique notamment que pour tous intervalles I et J de $\mathbb{R}$ les évènements $\{X \in I\}$ et $\{Y \in J\}$ sont indépendants. Calculez $\mathbf{E}\min(X,Y)$, $\mathbf{E}\max(X,Y)$ et en déduire $\mathbf{E}|X-Y|$.
Indication. Commencez par calculer $P(\min(X,Y) > t)$ ou $P(\max(X,Y) \leqslant t)$.

Ex. 7.7 *Espérance des statistiques d'ordre*

1) *Question préliminaire.* Pour tout $n \in \mathbb{N}^*$ et tout $j \in [\![0,n]\!]$, on pose

$$I_{j,n} := \int_0^1 t^j (1-t)^{n-j}\, \mathrm{d}t.$$

Calculez $I_{n,n}$ et justifiez l'égalité $I_{j,n} = \dfrac{n-j}{j+1} I_{j+1,n}$ pour $j \in [\![0,n[\![$. En déduire que

$$I_{j,n} = \frac{j!(n-j)!}{(n+1)!}.$$

Dans toute la suite de l'exercice, on note $X_1, \ldots, X_n$ une suite de variables aléatoires *positives*, définies sur le même espace probabilisé $(\Omega, \mathcal{F}, P)$, *indépendantes et de même loi*, donc de même fonction de répartition notée F. Le but de ce problème est de calculer l'espérance des variables aléatoires $X_{n:1}, \ldots X_{n:n}$ obtenues par réarrangement croissant de $X_1, \ldots, X_n$ et appelées *statistiques d'ordre*. Plus précisément, pour tout $\omega \in \Omega$, on a *l'égalité des ensembles* $\{X_{n:k}(\omega)\ ;\ k \in [\![1,n]\!]\}$ et $\{X_k(\omega)\ ;\ k \in [\![1,n]\!]\}$ et

$$\forall k \in [\![0,n[\![, \quad X_{n:k}(\omega) \leqslant X_{n:k+1}(\omega).$$

En particulier, $X_{n:1} = \min(X_1, \ldots, X_n)$ et $X_{n:n} = \max(X_1, \ldots, X_n)$. Pour illustrer cette définition, on montre ci-dessous les valeurs des statistiques d'ordre pour $n = 5$ déterminées à partir de valeurs prises par la suite des $X_i(\omega)$ pour ω et ω' différents.

k	1	2	3	4	5
$X_k(\omega)$	2,7	0,3	6	1,2	0,9
$X_{5:k}(\omega)$	0,3	0,9	1,2	2,7	6

k	1	2	3	4	5
$X_k(\omega')$	4,1	0,7	8,3	0,7	0,2
$X_{5:k}(\omega')$	0,2	0,7	0,7	4,1	8,3

Noter que la présence d'*ex-æquo* ne pose pas de problème.

2) Soit I une partie non vide de $[\![1,n]\!]$. En exploitant les hypothèses faites sur les X_i, exprimez à l'aide de $F(x)$, n et $\operatorname{card} I$, la probabilité

$$p_I(x) := P(\forall i \in I,\ X_i \leqslant x \text{ et } \forall i \in [\![1,n]\!]\backslash I,\ X_i > x).$$

3) On note F_k la fonction de répartition de la variable aléatoire positive $X_{n:k}$. Montrez que

$$\forall x \geqslant 0, \quad F_k(x) = P\big(X_{n:k} \leqslant x\big) = \sum_{j=k}^{n} C_n^j F(x)^j \big(1-F(x)\big)^{n-j}. \tag{7.79}$$

Indication : commencez par remarquer que « $X_{n:k} \leqslant x$ » est équivalent à « au moins k parmi les X_i sont inférieurs ou égaux à x ».

4) Expliquez comment déduire de (7.79) la formule

$$\forall x \geqslant 0, \quad 1 - F_k(x) = \sum_{j=0}^{k-1} C_n^j F(x)^j \big(1-F(x)\big)^{n-j}. \tag{7.80}$$

En déduire une formule générale exprimant $\mathbf{E}X_{n:k}$ à l'aide d'intégrales où intervient la fonction F.

5) On suppose dans cette question que les X_i suivent la loi uniforme sur $[0,1]$.

a) Explicitez alors $F(x)$ et achevez le calcul de $\mathbf{E}X_{n:k}$.

b) On pose de plus $X_{n:0} := 0$, $X_{n:n+1} := 1$ et $Y_{n,k} := X_{n:k+1} - X_{n:k}$. Les variables aléatoires $Y_{n,k}$, $k \in [\![0,n]\!]$, sont appelées les *espacements*. Que valent les $\mathbf{E}Y_{n,k}$?

6) On suppose dans cette question que les X_i suivent la loi exponentielle de paramètre 1. On a donc pour tout $x \geqslant 0$, $P(X_i > x) = \mathrm{e}^{-x}$. Calculez les $\mathbf{E}X_{n:k}$.

7) Montrez que si la variable aléatoire positive X_1 est *intégrable*, alors chacune des v.a. positives $X_{n:k}$ l'est aussi.

8) On suppose dans cette question que F est C^1 sur $[0,+\infty[$, avec $F(0) = 0$ et on note f sa dérivée. Montrez qu'alors $X_{n:k}$ admet pour densité la fonction f_k nulle sur $]-\infty,0[$ et vérifiant

$$\forall x > 0, \quad f_k(x) = \frac{n!}{(k-1)!(n-k)!} F(x)^{k-1} \big(1-F(x)\big)^{n-k} f(x).$$

Indication : commencez par vérifier les formules

$$jC_n^j = nC_{n-1}^{j-1}, \qquad (n-j)C_n^j = nC_{n-1}^j.$$

Solution p. 466

Ex. 7.8 *Loi logistique*

1) Soit Y une variable aléatoire de loi exponentielle de paramètre 1, sa fonction de survie est donc donnée par

$$P(Y > x) = \begin{cases} 1 & \text{si } x < 0, \\ \mathrm{e}^{-x} & \text{si } x \geqslant 0. \end{cases}$$

Expliquer pourquoi Y est intégrable et calculer $\mathbf{E}Y$. Calculer et représenter graphiquement la fonction de répartition de la variable aléatoire $Y' := -Y$.

2) Dans toute la suite de l'exercice, on note U une variable aléatoire réelle à valeurs dans $]0,1[$, de loi uniforme sur $]0,1[$. On pose

$$Z := \ln U$$

Montrer que Z et $-Y$ ont même loi. *Indication :* on pourra calculer la fonction de répartition de Z. En déduire la valeur de $\mathbf{E}Z$.

3) On définit la variable aléatoire réelle X par

$$X := \ln\Big(\frac{1-U}{U}\Big)$$

Calculer sa fonction de survie $G : \mathbb{R} \to [0,1]$, $x \mapsto G(x) := P(X > x)$. En déduire sa fonction de répartition F et l'existence d'une densité f que l'on calculera.

4) Justifier l'existence de $\mathbf{E}X$ et la calculer. *Indication :* il y a moyen de traiter cette question *sans écrire aucune intégrale*, en notant que $X = \ln(1-U) - \ln U$ et en cherchant la loi de $1-U$.

Solution p. 470

Ex. 7.9 *Erreurs typographiques (suite)* ↑ 6.2

1) En fait le nombre total d'erreurs dans le polycopié étant inconnu, on le modélise par une variable aléatoire N. Comme N est au plus égale au nombre maximum de caractères que peuvent contenir les 140 pages du polycopié, N est bornée et son espérance existe. On la note μ. Il sera commode d'écrire toutes les formules comme si l'ensemble des valeurs possibles de N était $N(\Omega) = \mathbb{N}$, ce qui revient à écrire des séries dans lesquelles $P(N = n) = 0$ pour n assez grand. On définit la variable aléatoire Z_i ($i \geqslant 1$) nombre total d'erreurs subsistant dans l'ensemble du polycopié après la i^{e} lecture (l'indépendance des détections et leurs probabilités sont les mêmes que dans l'exercice 6.2). Comme la loi de N est inconnue, nous ne pouvons plus calculer les lois des Z_i. Calculer $P(Z_1 = k \mid N = n)$. Montrer que $\mathbf{E}Z_1 = \mu/3$.

2) Exprimer $\mathbf{E}Z_i$ en fonction de μ.

Ex. 7.10 *Une somme doublement aléatoire*

Sur le même espace probabilisé $(\Omega, \mathcal{F}, P)$, on suppose définies

– une variable aléatoire N à valeurs dans $\mathbb{N}$,

– une suite de variables aléatoires positives $(X_i)_{i \geqslant 1}$, ayant toutes même loi.

On pose alors :

$$S_0 := 0, \quad S_n = \sum_{i=1}^{n} X_i \quad (n \geqslant 1).$$

On définit la variable aléatoire T par

$$T := \sum_{i=1}^{N} X_i, \tag{7.81}$$

autrement dit,

$$\forall \omega \in \Omega, \quad T(\omega) = S_{N(\omega)}(\omega).$$

Ce modèle est d'un usage courant. Par exemple si une compagnie d'assurances s'intéresse au risque « dommages au véhicule » pour une certaine catégorie de véhicules, dans un département donné, N représente le nombre de sinistres déclarés au cours d'une période de temps donnée et X_i le remboursement payé par la compagnie pour le i^{e} sinistre déclaré pendant cette période. Alors T est le montant total des remboursements effectués pour la période. On peut imaginer facilement d'autres applications comme le cumul des hauteurs de pluie sur une année (ou des heures d'ensoleillement), le total des retraits effectués sur un distributeur automatique bancaire en une journée, etc.

Le but de cet exercice est de *calculer l'espérance de T en fonction des espérances des X_i et de N* sans supposer connues les lois des X_i ou de N.

On suppose de plus que N est *indépendante de la suite* $(X_i)_{i\geqslant 1}$, ce qui implique notamment (on ne vous demande pas de le justifier) que pour tout $j \in \mathbb{N}$ et tous boréliens B et B' de $\mathbb{R}$, les évènements $\{S_j \in B\}$ et $\{N \in B'\}$ sont indépendants.

1) Trouvez l'erreur dans le raisonnement suivant qui n'aurait pas manqué de fleurir dans de nombreuses copies [27] si l'énoncé s'arrêtait ici. *Par additivité de l'espérance des variables aléatoires positives,*

$$\mathbf{E}T = \mathbf{E}\left(\sum_{i=1}^{N} X_i\right) = \sum_{i=1}^{N} \mathbf{E}X_i = N\mathbf{E}X_1.$$

2) Soit $A \in \mathcal{F}$ un évènement et Y une variable aléatoire sur $(\Omega, \mathcal{F}, P)$ tels que pour tout $t \geqslant 0$, les évènements $\{Y > t\}$ et A soient indépendants. Montrez que

$$\forall t \geqslant 0, \quad P(Y\mathbf{1}_A > t) = P(Y > t)P(A). \tag{7.82}$$

Indication : commencez par calculer $P(\{Y\mathbf{1}_A > t\} \cap A^c)$.

3) Déduire de ce qui précède que

$$\forall j \in \mathbb{N}, \quad \mathbf{E}\big(S_j \mathbf{1}_{\{N=j\}}\big) = P(N = j) j \mathbf{E}X_1. \tag{7.83}$$

27. Pas la vôtre bien sûr, mais celles de vos voisins...

4) Pour tout entier $n \geqslant 1$, on pose $T_n := T\mathbf{1}_{\{N\leqslant n\}}$. Justifiez l'égalité

$$T_n = \sum_{j=0}^{n} S_j \mathbf{1}_{\{N=j\}}. \tag{7.84}$$

En déduire que

$$\forall n \in \mathbb{N}^*, \quad \mathbf{E}T_n = \mathbf{E}X_1 \sum_{j=0}^{n} jP(N=j). \tag{7.85}$$

5) Vérifiez que la suite de variables aléatoires positives $(T_n)_{n\geqslant 1}$ converge en croissant vers T et en déduire que

$$\mathbf{E}T = (\mathbf{E}N)(\mathbf{E}X_1). \tag{7.86}$$

6) Retrouvez (7.86) à partir de (7.83) en établissant directement la formule

$$T = \sum_{j\in\mathbb{N}} S_j \mathbf{1}_{\{N=j\}} \tag{7.87}$$

(on comparera les valeurs prises par les deux membres en un ω quelconque) et en utilisant un résultat du cours.

Solution p. 473

Ex. 7.11 *Deux problèmes de minimisation*

Le but de cet exercice est de trouver une constante c qui soit « la plus proche » d'une variable aléatoire réelle X, au sens d'une certaine distance. La première question traite le cas de la distance L^2, les questions suivantes celui de la distance L^1. La résolution de cet exercice ne suppose aucune connaissance sur ces distances et n'utilise que les propriétés de la fonction de répartition et de l'espérance.

1) Soit X une variable aléatoire réelle de carré intégrable. On pose pour tout $a \in \mathbb{R}$,

$$g(a) := \mathbf{E}\big((X-a)^2\big).$$

Montrez que g est bien définie sur $\mathbb{R}$ et y admet un minimum unique atteint en un point a_0 que vous calculerez. Que reconnaissez vous en $g(a_0)$?

2) Dans cette question et les suivantes, X est une variable aléatoire intégrable dont la fonction de répartition sera notée F. On pose pour tout $a \in \mathbb{R}$,

$$h(a) := \mathbf{E}|X-a|$$

et on cherche quelle(s) valeur(s) de a minimise(nt) $h(a)$. Expliquez pourquoi h est bien définie sur tout $\mathbb{R}$.

3) Donnez une représentation graphique de $\mathbf{E}|X-a|$ en utilisant le graphe de F.

4) En revenant à la définition de l'espérance de la variable aléatoire positive $|X-a|$, établissez la formule suivante :

$$\forall a \in \mathbb{R}, \qquad \mathbf{E}|X-a| = \mathbf{E}|X| + 2\int_0^a \Big(F(s) - \frac{1}{2}\Big)\,\mathrm{d}s. \tag{7.88}$$

On pourra commencer par examiner le cas $a > 0$.

5) L'inverse généralisée de la fonction de répartition F est définie sur $]0,1[$ par :

$$\forall u \in]0,1[, \quad F^{-1}(u) := \inf\{t \in \mathbb{R} \,;\, F(t) \geqslant u\}.$$

On dit d'autre part que le réel m est *une médiane* de X, s'il vérifie les deux conditions $P(X \geqslant m) \geqslant 1/2$ et $P(X \leqslant m) \geqslant 1/2$. Vérifiez que $F^{-1}(1/2)$ est une médiane de X et que c'est la plus petite de toutes les médianes dans le cas où il y en aurait plusieurs. Pourriez vous proposer un exemple d'un tel cas en donnant F par sa représentation graphique ?

6) En revenant à (7.88), montrez que h atteint un minimum absolu au point $a = F^{-1}(1/2)$. Ce minimum peut-il être atteint ailleurs ?

7) Question bonus : proposez une résolution graphique de ce problème.

Ex. 7.12 *Espérance d'une loi discrète pathologique* ↑ 6.17

Calculer l'espérance de la variable aléatoire discrète X définie à l'exercice 6.17. Le résultat dépend de p. On vérifiera que dans tous les cas :

$$\frac{9}{20} < \mathbf{E}X < \frac{1}{2}.$$

Commenter ce résultat. Comment interpréter les cas limite $p = 1$ et $p = 0$?

Solution p. 477

Ex. 7.13 *Transformée de Laplace*

La transformée de Laplace d'une variable aléatoire positive X est la fonction L_X définie sur $\mathbb{R}_+$ par :

$$\forall t \in \mathbb{R}_+, \quad L_X(t) = \mathbf{E}\exp(-tX).$$

Clairement L_X ne dépend que de la loi de X. On parlera donc aussi bien de la transformée de Laplace d'une variable que de sa loi.

1) Expliquez pourquoi cette définition a toujours un sens et calculez la transformée de Laplace d'une loi de Bernoulli de paramètre p.

2) Vérifiez que L_X est décroissante sur $\mathbb{R}_+$.

3) On rappelle que f est convexe sur $\mathbb{R}_+$ si pour tout $u \in [0,1]$, et tous $s, t \in \mathbb{R}_+$, $f(us + (1-u)t) \leqslant uf(s) + (1-u)f(t)$. Vérifiez que L_X est convexe sur $\mathbb{R}_+$.

4) Cette convexité de L_X sur $\mathbb{R}_+$ implique sa continuité sur $]0,+\infty[$. Pour montrer la continuité de L_X au point 0, il suffit, compte-tenu de la monotonie de L_X, de montrer que $L_X(1/n)$ tend vers $L_X(0) = 1$ quand n tend vers $+\infty$. Justifiez cette convergence, par exemple à l'aide du théorème de B. Levi.

5) Vérifiez que la suite de v.a. $(Y_n)_{n\geqslant 1}$ définies par $Y_n = \exp(-nX)$ tend en décroissant vers $\mathbf{1}_{\{X=0\}}$. Utilisez ceci pour trouver la limite de $L_X(n)$ quand n tend vers l'infini.

6) Que peut-on dire de L_{X+Y} lorsque X et Y sont des variables aléatoires positives indépendantes ?

Solution p. 479

Ex. 7.14 *Inégalité de Jensen*

Soit φ une fonction convexe sur $\mathbb{R}$. Si X est une variable aléatoire telle que X et $\varphi(X)$ soient intégrables, alors

$$\varphi(\mathbf{E}X) \leqslant \mathbf{E}\varphi(X). \tag{7.89}$$

Le but de cet exercice est de prouver cette inégalité de Jensen.

1) Montrez que pour tout réel $x_0 \in \mathbb{R}$, il existe un réel c tel que pour tout $x \in \mathbb{R}$, $\varphi(x) \geqslant \varphi(x_0) + c(x - x_0)$. Proposez une interprétation géométrique de cette inégalité.

2) En prenant $x = X(\omega)$ et $x_0 = \mathbf{E}X$ dans l'inégalité ci-dessus, prouvez (7.89).

3) Généralisez au cas où φ est définie et convexe sur un intervalle *ouvert* I de $\mathbb{R}$. Que peut-on dire si I n'est pas ouvert ?

Ex. 7.15 *Moments fonctionnels et décroissance lente*

Dans cet exercice, on utilise la formule (7.65) de calcul de moments fonctionnels pour apporter une réponse partielle au problème suivant. Grâce à l'inégalité de Markov, on sait que plus la v.a. positive X a des moments d'ordre élevé, plus vite sa fonction de survie G tend vers 0 en $+\infty$. On se demande *a contrario* si, quelle que soit la vitesse de convergence vers 0 de G, il est possible de trouver un moment fonctionnel $\mathbf{E}h(X)$ fini avec h tendant vers $+\infty$ en $+\infty$. Les deux questions suivantes sont consacrées à l'étude d'un exemple où G tend « lentement » vers 0, afin de se faire une intuition.

1) Soit a un réel. Montrez que l'intégrale généralisée

$$I(a) := \int_2^{+\infty} \frac{\mathrm{d}t}{t(\ln t)^a}$$

converge si et seulement si $a > 1$.

2) Soit X une variable aléatoire positive de fonction de survie G définie par

$$\forall t \in \mathbb{R}, \qquad G(t) = \begin{cases} 1 & \text{si } t < 0, \\ \dfrac{1}{\ln(\mathrm{e} + t)} & \text{si } t \geqslant 0. \end{cases}$$

Dans cette formule, la constante e est la base du logarithme néperien ($\ln(\mathrm{e}) = 1$).

a) Montrez que pour tout $\varepsilon > 0$, $\mathbf{E}X^{\varepsilon} = +\infty$.

b) Montrez que pour tout $b \in]0, 1[$, $\mathbf{E}(\ln(1 + X))^b < +\infty$.

3) On suppose plus généralement que X est une variable aléatoire positive dont la fonction de survie G est C^1 sur $]0, +\infty[$ et strictement décroissante sur $[0, +\infty[$. Ceci implique en particulier que $G(t)$ est strictement positif pour tout $t \in [0, +\infty[$ (pourquoi ?). Montrez l'existence d'une fonction positive h, strictement croissante sur $[0, +\infty[$, tendant vers $+\infty$ en $+\infty$ et telle que $\mathbf{E}h(X) < +\infty$. *Indication* : essayez avec $h : t \mapsto G(t)^{-1/2}$. La variable aléatoire X considérée dans cette question est-elle nécessairement à densité ?

Solution p. 481

Chapitre 8

Vecteurs aléatoires et indépendance

L'INFORMATION pertinente résultant d'une expérience aléatoire ne se résume pas toujours à la valeur prise par une seule variable aléatoire réelle. On a souvent besoin de connaître les valeurs d'une suite finie de variables aléatoires. Par exemple au jeu de 421, on lance trois dés et on a besoin de connaître les points affichés par chacun des dés, le résultat sera donc décrit par un vecteur $(X_1(\omega), X_2(\omega), X_3(\omega))$. Si on tire sur une cible, le résultat sera décrit par les coordonnées $(X(\omega), Y(\omega))$ du point d'impact. Si on étudie le fonctionnement d'un guichet en observant les n premiers clients, le résultat sera décrit par la suite des $X_1, Y_1, Z_1, X_2, Y_2, Z_2, \ldots X_n, Y_n$ où X_i est le temps d'attente au guichet du i^{e} client, Y_i son temps de service et Z_i le temps s'écoulant entre le départ du i^{e} client et l'arrivée du $(i+1)^{\text{e}}$. Ces suites finies de variables aléatoires sont appelées des *vecteurs aléatoires*. De même qu'une variable aléatoire peut être vue comme un procédé de choix d'un nombre réel au hasard, un vecteur aléatoire de dimension d, $X = (X_1, \ldots, X_d)$ est un procédé de choix au hasard d'un point de $\mathbb{R}^d$. Ses composantes $X_1, \ldots, X_d$ sont alors autant de variables aléatoires réelles. Arrivé là, le lecteur peut se demander légitimement s'il y a un intérêt à consacrer tout un chapitre aux vecteurs aléatoires puisque ces objets ne sont que des suites finies de variables aléatoires et que ces dernières sont maintenant bien connues. L'intérêt de cette étude repose sur la remarque informelle suivante à laquelle nous donnerons bientôt un sens mathématique précis : la *connaissance probabiliste globale du vecteur* $X = (X_1, \ldots, X_d)$ *apporte davantage d'information que la connaissance probabiliste individuelle de chacune de ses composantes* X_i. Au premier abord, cette idée peut paraître choquante car une lecture rapide de la phrase précédente laisse croire que la connaissance de $X(\omega)$ apporte quelque chose de plus que celle de tous les $X_i(\omega)$, $i = 1, \ldots, d$, ce qui n'est évidemment pas vrai. La clé de l'énigme est dans l'expression « connaissance *probabiliste* » que nous remplacerons bientôt par connaissance de la *loi*, dès que nous aurons défini la loi d'un vecteur

aléatoire. En attendant voici une image qui peut nous aider à comprendre ce dont il s'agit. Considérons un ensemble de 10 coureurs de fond, chacun muni d'un dossard numéroté de 1 à 10. Si on les rassemble sur une même piste pour une épreuve de 5 000 mètres, on peut représenter le résultat de la course par le vecteur $(X_1, \dots, X_{10})$, où X_i désigne le temps mis par le coureur numéroté i pour parcourir les 5 000 mètres. Tout amateur d'athlétisme sait bien que cette expérience n'est pas équivalente à faire courir isolément un 5 000 mètres à chacun des 10 coureurs sur des stades séparés. La différence ici vient de la compétition, de la tactique de course, etc. En revanche, dans d'autres situations, le comportement global du vecteur des d composantes se réduit au comportement individuel de chacune d'elles. On admet généralement que c'est le cas lorsqu'on lance trois dés en considérant qu'il revient au même de les lancer ensemble sur la même table ou séparément sur trois tables. On parle alors d'indépendance des composantes. Cette notion d'indépendance des composantes X_i du vecteur aléatoire X est reliée à celle d'une suite d'évènements A_i, où la réalisation ou non de A_i ne dépend que des valeurs de X_i. L'étude des suites finies de variables aléatoires indépendantes prend donc place naturellement dans ce chapitre comme le cas particulier des vecteurs aléatoires à composantes indépendantes.

Sauf mention explicite du contraire, toutes les variables aléatoires et tous les vecteurs aléatoires considérés dans ce chapitre seront définis sur le même espace probabilisé $(\Omega, \mathcal{F}, P)$.

8.1 Vecteurs aléatoires

8.1.1 Généralités

Définition 8.1 (vecteur aléatoire de $\mathbb{R}^d$)**.** *On dit que l'application $X : \Omega \to \mathbb{R}^d$ est un vecteur aléatoire sur $(\Omega, \mathcal{F})$ si c'est une application mesurable $\mathcal{F}$ – $\mathrm{Bor}(\mathbb{R}^d)$.*

Proposition 8.2. *Soit X un vecteur aléatoire sur $(\Omega, \mathcal{F})$ et P une probabilité sur $(\Omega, \mathcal{F})$. La fonction d'ensembles $P_X = P \circ X^{-1}$ définie sur $\mathrm{Bor}(\mathbb{R}^d)$ par*

$$\forall B \in \mathrm{Bor}(\mathbb{R}^d), \quad P_X(B) := P\big(X^{-1}(B)\big) = P(X \in B) \tag{8.1}$$

est une probabilité sur $\big(\mathbb{R}^d, \mathrm{Bor}(\mathbb{R}^d)\big)$.

La preuve est exactement la même que celle de la proposition 6.8, en remplaçant les boréliens de $\mathbb{R}$ par ceux de $\mathbb{R}^d$. Comme pour les variables aléatoires, cette proposition légitime la définition de la loi de X sous P.

Définition 8.3 (loi d'un vecteur aléatoire)**.** *Soient $(\Omega, \mathcal{F}, P)$ un espace probabilisé et $X : \Omega \to \mathbb{R}^d$ un vecteur aléatoire sur $(\Omega, \mathcal{F})$. On appelle* loi de X sous P*, ou plus simplement loi de X, la probabilité P_X sur $\big(\mathbb{R}^d, \mathrm{Bor}(\mathbb{R}^d)\big)$ définie par (8.1).*

Proposition 8.4 (lois marginales)**.** *Si $X = (X_1, \dots, X_d)$ est un vecteur aléatoire sur $(\Omega, \mathcal{F})$, chacune de ses composantes X_i $(1 \leqslant i \leqslant d)$ est une variable aléatoire réelle sur $(\Omega, \mathcal{F})$. La loi de X_i est appelée i^e* loi marginale *de X et est donnée par :*

$$\forall B_i \in \mathrm{Bor}(\mathbb{R}), \quad P_{X_i}(B_i) = P(X_i \in B_i) = P(X \in \mathbb{R}^{i-1} \times B_i \times \mathbb{R}^{d-i}). \tag{8.2}$$

Preuve. La mesurabilité $\mathcal{F}$ – Bor($\mathbb{R}$) de X_i s'obtient par composition à partir de la mesurabilité $\mathcal{F}$ – Bor($\mathbb{R}^d$) de X. En effet $X_i = \pi_i \circ X$, où $\pi_i : (x_1, \ldots, x_d) \mapsto x_i$ est la i^{e} projection canonique de $\mathbb{R}^d$ sur $\mathbb{R}$; comme π_i est continue, elle est borélienne, c'est-à-dire ici mesurable Bor($\mathbb{R}^d$) – Bor($\mathbb{R}$). Donc X_i est bien une variable aléatoire. Pour vérifier (8.2), il suffit de remarquer que l'équivalence

$$X_i \in B_i \;\Leftrightarrow\; X \in \mathbb{R}^{i-1} \times B_i \times \mathbb{R}^{d-i}$$

entraîne l'égalité des évènements correspondants et de leur probabilité. □

Exemple 8.5. Soit $X = (X_1, X_2)$ un vecteur aléatoire de loi uniforme sur $[0,1]^2$. Alors $P(X \in B) = \lambda_2(B \cap [0,1]^2)$ pour tout borélien B de $\mathbb{R}^2$ et on vérifie (exercice) que les lois marginales P_{X_1} et P_{X_2} sont égales à la loi uniforme sur $[0,1]$.

Remarque 8.6 (d'importance capitale)**.** Une conséquence de la proposition 8.4 est que la connaissance de la loi du vecteur aléatoire X détermine complètement celle de ses lois marginales. La *réciproque est fausse*. On peut même affirmer sans hésiter qu'il est impossible de comprendre la notion de vecteur aléatoire tant que l'on n'a pas assimilé ce fait. La comparaison de l'exemple 8.7 ci-dessous avec l'exemple 8.5 permet de voir que la connaissance des lois marginales d'un vecteur ne détermine pas la loi du vecteur.

Exemple 8.7. Prenons une variable aléatoire réelle Y_1 de loi uniforme sur $[0,1]$ et posons $Y_2 := Y_1$ et $Y := (Y_1, Y_2)$. Le vecteur aléatoire Y a par construction les mêmes lois marginales que le vecteur X de l'exemple 8.5. Notons $\Delta := \{(s,t) \in [0,1]^2 ;\, s = t\}$ la première diagonale du carré unité. Il est clair par construction que $P(Y \in \Delta) = 1$.

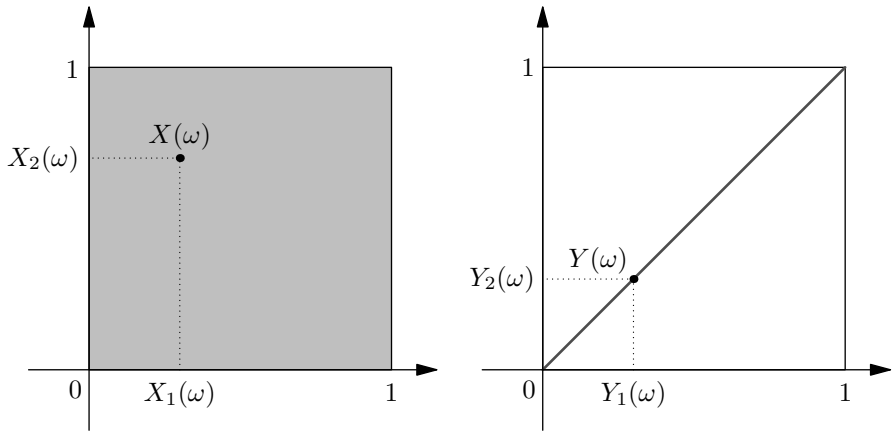


FIGURE 8.1 – Ensembles de probabilité 1 pour les lois P_X et P_Y

D'un autre côté, $P(X \in \Delta) = 0$, car le segment Δ est de λ_2 mesure nulle. Ceci empêche que X et Y aient même loi. La figure 8.2 propose une illustration en affichant

le résultat de la simulation du choix « au hasard » de 50 points suivant la loi P_X puis suivant la loi P_Y.

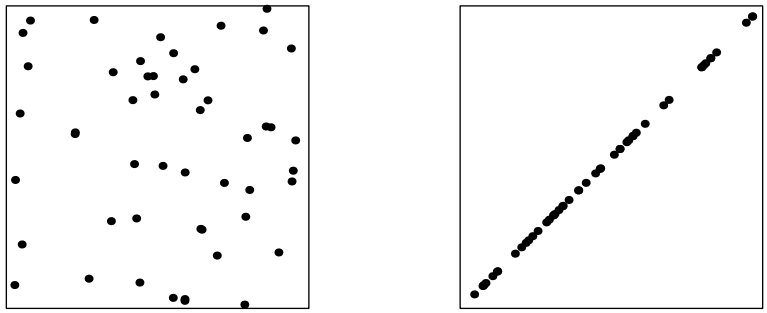


FIGURE 8.2 – 50 points choisis au hasard suivant la loi P_X, puis suivant la loi P_Y

Définition 8.8 (vecteur aléatoire discret)**.** *Le vecteur aléatoire X de $\mathbb{R}^d$ est dit* discret *si $X(\Omega)$ est une partie au plus dénombrable de $\mathbb{R}^d$.*

Il est clair que la loi de X s'écrit alors

$$P_X = \sum_{x \in X(\Omega)} P(X = x)\delta_x.$$

Les variables aléatoires marginales X_i de $X = (X_1, \dots, X_d)$ sont alors des variables aléatoires discrètes. En effet, soit π_i, la restriction à $X(\Omega)$ de la projection canonique sur la i-ième composante de $\mathbb{R}^d$. Cette application réalise une surjection de $X(\Omega)$ sur $X_i(\Omega)$. Par la proposition 1.41, on en déduit que $X_i(\Omega)$ est au plus dénombrable.

Exemple 8.9 (lois multinomiales)**.** Le vecteur aléatoire N suit la loi multinomiale de paramètres n et $(p_1, \dots, p_d)$ où $n \in \mathbb{N}^*$ et les p_i sont strictement positifs et de somme 1 si pour tout d-uple $(j_1, j_2, \dots, j_d)$ d'entiers tels que $j_1 + j_2 + \cdots + j_d = n$,

$$P\big\{N = (j_1, j_2, \dots, j_d)\big\} = \frac{n!}{j_1!\, j_2! \dots j_d!} p_1^{j_1} p_2^{j_2} \dots p_d^{j_d}.$$

Ici l'ensemble $N(\Omega) = \{(j_1, j_2, \dots, j_d) \in \mathbb{N}^d;\ j_1 + j_2 + \cdots + j_d = n\}$ est fini et on vérifie grâce à la formule du multinôme, cf. prop. 1.23, que

$$\sum_{x \in N(\Omega)} P(N = x) = \sum_{j_1 + \cdots + j_d = n} \frac{n!}{j_1!\, j_2! \dots j_d!} p_1^{j_1} p_2^{j_2} \dots p_d^{j_d} = \big(p_1 + \cdots + p_d\big)^n = 1^n = 1.$$

La loi multinomiale est celle du vecteur des résultats d'une suite d'épreuves répétées indépendantes ayant chacune d issues possibles de probabilités respectives $p_1, \dots, p_d$. On pourra justifier cette affirmation en exercice. Par exemple considérons 20 tirages

d'une boule avec remise dans une urne contenant 1 boule bleue, 3 jaunes, 4 rouges et 2 vertes. Notons $N = (N_1, N_2, N_3, N_4)$ où N_i est le nombre de boules de la couleur i en numérotant les couleurs par ordre alphabétique (b,j,r,v). On a $(p_1, p_2, p_3, p_4) = (\frac{1}{10}, \frac{3}{10}\frac{4}{10}, \frac{2}{10})$. La probabilité d'obtenir en 20 tirages 3 bleues, 5 jaunes, 10 rouges et 2 vertes est

$$P\big(N = (3,5,10,2)\big) = \frac{20!}{3!\,5!\,10!\,2!}\Big(\frac{1}{10}\Big)^3\Big(\frac{3}{10}\Big)^5\Big(\frac{4}{10}\Big)^{10}\Big(\frac{2}{10}\Big)^2 \simeq 0{,}004\,745.$$

Il est facile de voir que les lois des vecteurs aléatoires discrets sur $\mathbb{R}^d$ sont exactement les mesures de probabilité ponctuelles sur $\mathbb{R}^d$ (au sens de l'exemple 5.9). Autrement dit, une mesure μ sur $(\mathbb{R}^d, \mathrm{Bor}(\mathbb{R}^d))$ est la loi d'un vecteur aléatoire discret si et seulement si elle peut s'écrire $\mu := \sum_{i\in I} p_i\delta_{x_i}$ où $\{x_i,\ i \in I\}$ est une famille de vecteurs de $\mathbb{R}^d$ et $\{p_i,\ i \in I\}$ est une famille *sommable* de réels *strictement* positifs de somme 1.

Certaines lois de vecteurs aléatoires ont une « densité » par rapport à la mesure de Lebesgue de $\mathbb{R}^d$. Nous donnons ci-dessous une définition volontairement restrictive des densités sur $\mathbb{R}^d$, restant dans le cadre de l'intégration au sens de Riemann. Cette définition n'est ni la seule possible ni la plus générale. Elle devrait largement suffire pour couvrir les exemples courants.

Définition 8.10 (densité de probabilité sur $\mathbb{R}^d$)**.** *On appelle densité de probabilité sur $\mathbb{R}^d$ toute fonction borélienne positive définie sur $\mathbb{R}^d$, ayant une intégrale de Riemann généralisée sur $\mathbb{R}^d$ convergente au sens des définitions 4.88 ou 4.89 et vérifiant :*

$$\int_{\mathbb{R}^d} f(t)\,\mathrm{d}t = \int_{\mathbb{R}^d} f(t_1,\dots,t_d)\,\mathrm{d}t_1\dots\,\mathrm{d}t_d = 1.$$

Définition 8.11 (vecteur aléatoire à densité)**.** *Soit f une densité de probabilité sur $\mathbb{R}^d$. On dit que le vecteur aléatoire X de $\mathbb{R}^d$ a pour densité f si pour tout pavé fermé borné $C = [a_1, b_1] \times \cdots \times [a_d, b_d]$,*

$$P(X \in C) = \int_C f(t_1,\dots,t_d)\,\mathrm{d}t_1\dots\,\mathrm{d}t_d.$$

Comme la classe des pavés fermés bornés engendre la tribu borélienne de $\mathbb{R}^d$, on peut montrer que si deux vecteurs aléatoires X et Y ont même densité f, ils ont même loi. Par ailleurs, si un vecteur aléatoire X admet une densité f, celle-ci n'est pas unique (si on la modifie en un point, on ne changera pas la valeur des ses intégrales sur les pavés fermés bornés).

Voici un premier exemple de vecteur aléatoire à densité. D'autres seront vus ultérieurement.

Exemple 8.12 (densité de la loi uniforme sur un borélien de $\mathbb{R}^d$)**.** Soit B un borélien de $\mathbb{R}^d$ tel que $0 < \lambda_d(B) < +\infty$. Si le vecteur aléatoire X de $\mathbb{R}^d$ suit la loi uniforme sur B, cf. exemple 5.26, il admet pour densité la fonction

$$f = \frac{1}{\lambda_d(B)}\mathbf{1}_B.$$

En effet, pour tout pavé fermé borné $C = [a_1, b_1] \times \cdots \times [a_d, b_d]$, on a

$$\begin{aligned} P(X \in C) = \frac{\lambda_d(B \cap C)}{\lambda_d(B)} &= \frac{1}{\lambda_d(B)} \int_{\mathbb{R}^d} \mathbf{1}_{B\cap C}(t_1, \ldots, t_d)\, \mathrm{d}t_1 \ldots \mathrm{d}t_d \\ &= \frac{1}{\lambda_d(B)} \int_{\mathbb{R}^d} \mathbf{1}_B(t_1, \ldots, t_d)\mathbf{1}_C(t_1, \ldots, t_d)\, \mathrm{d}t_1 \ldots \mathrm{d}t_d \\ &= \frac{1}{\lambda_d(B)} \int_C \mathbf{1}_B(t_1, \ldots, t_d)\, \mathrm{d}t_1 \ldots \mathrm{d}t_d \\ &= \int_C f(t_1, \ldots, t_d)\, \mathrm{d}t_1 \ldots \mathrm{d}t_d. \end{aligned}$$

Le calcul ci-dessus utilise l'identité $\mathbf{1}_{B\cap C} = \mathbf{1}_B\mathbf{1}_C$ et l'expression intégrale de la mesure d'un borélien A par $\lambda_d(A) = \int_A \mathrm{d}t_1 \ldots \mathrm{d}t_d = \int_{\mathbb{R}^d} \mathbf{1}_A(t_1, \ldots, t_d)\, \mathrm{d}t_1 \ldots \mathrm{d}t_d$. Nous touchons ici aux limitations de cet ouvrage sur la théorie de l'intégrale. Dans notre cadre, nous ne devrions écrire cette relation que si $\mathbf{1}_A$ est localement Riemann intégrable sur $\mathbb{R}^d$. En pratique nous n'utiliserons la loi uniforme que sur des boréliens assez simples pour lesquels ces conditions seront remplies.

Comme dans le cas $d = 1$, on dispose d'une condition suffisante pratique pour vérifier que deux vecteurs aléatoires à densité n'ont pas même loi.

Lemme 8.13. *Soient X et Y deux vecteurs aléatoires de $\mathbb{R}^d$ admettant respectivement pour densité les fonctions f et g. On suppose qu'il existe $t_0 \in \mathbb{R}^d$ tel que $f(t_0) \neq g(t_0)$ et que de plus, f et g sont* toutes deux continues au point t_0. *Alors X et Y n'ont pas même loi.*

La preuve est essentiellement la même que celle du lemme 6.25, en remplaçant les intervalles ouverts par des pavés ouverts et sera donc omise.

Proposition 8.14 (densités marginales)**.** *Si le vecteur aléatoire X de $\mathbb{R}^d$ est à densité f, ses lois marginales sont aussi à densité et pour $i \in [\![1, d]\!]$, une densité de X_i est donnée par*

$$x_i \longmapsto f_{X_i}(x_i) = \int_{\mathbb{R}^{i-1}\times\mathbb{R}^{d-i}} f(t_1, \ldots, t_{i-1}, x_i, t_{i+1}, \ldots, t_d)\, \mathrm{d}t_1 \ldots \mathrm{d}t_{i-1}\, \mathrm{d}t_{i+1} \ldots \mathrm{d}t_d.$$

L'outillage d'intégration développé dans cet ouvrage est insuffisant pour démontrer proprement cette proposition. La difficulté vient de ce que pour certaines valeurs de x_i, l'intégrale généralisée ci-dessus n'est pas forcément définie et que l'ensemble de ces x_i peut ne pas être $\mathcal{R}$-négligeable (il est négligeable au sens de Lebesgue). Le théorème de Fubini-Tonnelli permet d'avoir une solution correcte dans le cadre de l'intégrale de Lebesgue. En pratique nous ne rencontrerons pas de difficulté à ce sujet. Si on fait abstraction de ces difficultés, la proposition se « démontre » simplement.

Simulacre de preuve. On calcule $P(X_i \in [a_i, b_i])$ en utilisant (8.2) :

$$\begin{aligned} P(X_i \in [a_i, b_i]) &= \int_{\mathbb{R}^{i-1}\times[a_i,b_i]\times\mathbb{R}^{d-i}} f(t_1, \dots, t_d)\, \mathrm{d}t_1 \dots \mathrm{d}t_d \\ &= \int_{[a_i,b_i]} \left\{ \int_{\mathbb{R}^{i-1}\times\mathbb{R}^{d-i}} f(t_1, \dots, t_d)\, \mathrm{d}t_1 \dots \mathrm{d}t_{i-1}\, \mathrm{d}t_{i+1} \dots \mathrm{d}t_d \right\} \mathrm{d}t_i, \end{aligned}$$

en admettant la dernière égalité. Comme a_i et b_i sont quelconques, on en déduit que la fonction entre accolades ci-dessus est une densité de X_i. □

Exemple 8.15 (loi uniforme sur un disque)**.** Soit D le disque unité de $\mathbb{R}^2$ et $X = (X_1, X_2)$ un vecteur aléatoire suivant la loi uniforme sur D. D'après l'exemple 8.12, nous savons qu'il admet pour densité $f = \pi^{-1}\mathbf{1}_D$. Calculons la densité marginale f_{X_1} fournie par la proposition 8.14.

$$f_{X_1}(x_1) = \int_{\mathbb{R}} \frac{1}{\pi}\mathbf{1}_D(x_1, t_2)\, \mathrm{d}t_2 = \frac{1}{\pi}\int_{-\infty}^{+\infty} \mathbf{1}_D(x_1, t_2)\, \mathrm{d}t_2. \tag{8.3}$$

Notons ici que l'appartenance à D du point (t_1, t_2) de $\mathbb{R}^2$ est caractérisée par l'inégalité $t_1^2 + t_2^2 \leqslant 1$. On a donc

$$\mathbf{1}_D(x_1, t_2) = \begin{cases} 1 & \text{si } x_1^2 + t_2^2 \leqslant 1 \\ 0 & \text{sinon} \end{cases} = \begin{cases} 1 & \text{si } x_1^2 \leqslant 1 \text{ et } t_2^2 \leqslant 1 - x_1^2 \\ 0 & \text{sinon} \end{cases}$$

D'où

$$\mathbf{1}_D(x_1, t_2) = \begin{cases} 1 & \text{si } |x_1| \leqslant 1 \text{ et } |t_2| \leqslant \sqrt{1 - x_1^2}, \\ 0 & \text{sinon,} \end{cases}$$

ce qui peut aussi s'écrire

$$\mathbf{1}_D(x_1, t_2) = \mathbf{1}_{[-1,1]}(x_1)\mathbf{1}_{\left[-(1-x_1^2)^{1/2}, (1-x_1^2)^{1/2}\right]}(t_2).$$

Géométriquement, ce petit calcul revient à chercher l'intersection du disque D avec la droite verticale $t_1 = x_1$, où x_1 joue le rôle d'une constante, et à projeter cette intersection sur le deuxième axe, cf figure 8.3. En reportant cette expression de $\mathbf{1}_D(x_1, t_2)$ dans (8.14), on obtient

$$\begin{aligned} f_{X_1}(x_1) &= \frac{1}{\pi}\int_{-\infty}^{+\infty} \mathbf{1}_{[-1,1]}(x_1)\mathbf{1}_{\left[-(1-x_1^2)^{1/2}, (1-x_1^2)^{1/2}\right]}(t_2)\, \mathrm{d}t_2 \\ &= \frac{1}{\pi}\mathbf{1}_{[-1,1]}(x_1)\int_{-\infty}^{+\infty} \mathbf{1}_{\left[-(1-x_1^2)^{1/2}, (1-x_1^2)^{1/2}\right]}(t_2)\, \mathrm{d}t_2 \\ &= \frac{1}{\pi}\mathbf{1}_{[-1,1]}(x_1)\int_{-(1-x_1^2)^{1/2}}^{(1-x_1^2)^{1/2}} \mathrm{d}t_2. \end{aligned}$$

Cette dernière intégrale n'est rien d'autre que la longueur du segment défini par ses bornes (rappelons que x_1 est considéré comme une constante dans tout ce calcul), ou celle du segment AB sur la figure 8.3. On aboutit ainsi à

$$f_{X_1}(x_1) = \frac{2}{\pi}(1-x_1^2)^{1/2}\mathbf{1}_{[-1,1]}(x_1).$$

Par raison de symétrie il est clair que X_1 et X_2 ont même loi et donc même densité, d'où

$$f_{X_2}(x_2) = \frac{2}{\pi}(1-x_2^2)^{1/2}\mathbf{1}_{[-1,1]}(x_2).$$

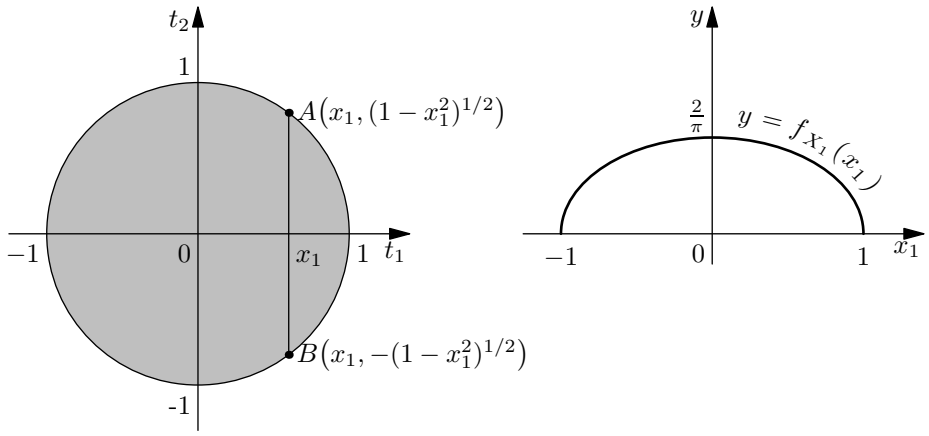


FIGURE 8.3 – Densité marginale f_{X_1} de la loi uniforme sur le disque unité D

La construction du graphe de f_{X_1} est très simple : on l'obtient en appliquant au demi cercle unité supérieur d'équation $y = \sqrt{1-x_1^2}$ la transformation $(x_1, y) \mapsto (x_1, \frac{2}{\pi}y)$. On obtient ainsi une moitié d'ellipse. Un coup d'oeil sur ce graphe devrait vous convaincre que la loi de X_1, bien qu'ayant toute sa masse portée par le segment $[-1,1]$, *n'est pas* la loi uniforme sur $[-1,1]$.

On peut définir la fonction de répartition F d'un vecteur aléatoire par

$$\forall x = (x_1,\dots,x_d), \quad F(x) = P\big(X \in]-\infty, x_1] \times \cdots \times]-\infty, x_d]\big).$$

Comme en dimension 1, la fonction de répartition caractérise la loi. Ceci est lié au fait que la tribu borélienne de $\mathbb{R}^d$ est engendrée par la classe des ensembles de la forme $]-\infty, x_1] \times \cdots \times]-\infty, x_d]$. Néanmoins le rôle des f.d.r. en dimension $d > 1$ est bien moindre qu'en dimension 1. On préfère caractériser la loi d'un vecteur aléatoire par une collection de h-moments $\mathbf{E}h(X)$, $h : \mathbb{R}^d \to \mathbb{R}$, au sens suivant.

Proposition 8.16 (caractérisation par les moments fonctionnels)**.** *La loi d'un vecteur aléatoire X de $\mathbb{R}^d$ est caractérisée par la famille de moments fonctionnels $\{\mathbf{E}h(X)\ ;\ h \in \mathcal{H}\}$, où $\mathcal{H}$ est une classe « suffisamment riche » de fonctions boréliennes $\mathbb{R}^d \to \mathbb{R}$. Autrement dit, les deux vecteurs aléatoires X et Y ont même loi si et seulement si $\mathbf{E}h(X) = \mathbf{E}h(Y)$ pour toute $h \in \mathcal{H}$. Comme famille $\mathcal{H}$ « suffisamment riche », on peut prendre :*

- *l'ensemble des fonctions boréliennes positives $\mathbb{R}^d \to \mathbb{R}_+$,*
- *l'espace $C^{\mathrm{b}}(\mathbb{R}^d)$ des fonctions continues bornées $\mathbb{R}^d \to \mathbb{R}$,*
- *l'espace $C^{\mathrm{c}}(\mathbb{R}^d)$ des fonctions continues à support compact $\mathbb{R}^d \to \mathbb{R}$,*
- *le sous-ensemble $C^{\mathrm{c}}_+(\mathbb{R}^d)$ des fonctions positives de $C^{\mathrm{c}}(\mathbb{R}^d)$.*

Preuve. Comme $C^{\mathrm{c}}_+(\mathbb{R}^d)$ est inclus dans chacune des 3 autres familles proposées, il suffit de faire la preuve pour $\mathcal{H} = C^{\mathrm{c}}_+(\mathbb{R}^d)$. La partie évidente est que l'égalité des lois des vecteurs aléatoires X et Y implique l'égalité des lois des variables aléatoires réelles bornées $h(X)$ et $h(Y)$ et donc l'égalité de leur espérance. Pour montrer la réciproque, on suppose désormais que

$$\forall h \in C^{\mathrm{c}}_+(\mathbb{R}^d), \qquad \mathbf{E}h(X) = \mathbf{E}h(Y). \tag{8.4}$$

Comme la loi d'un vecteur aléatoire Z de $\mathbb{R}^d$ est caractérisée par les $P(Z \in C)$, où C décrit la famille des pavés ouverts de $\mathbb{R}^d$, il nous suffit de prouver que $P(X \in C) = P(Y \in C)$ pour tout pavé ouvert C.

Soit donc

$$C := \prod_{j=1}^{d}]a_j, b_j[,$$

un pavé ouvert non vide de $\mathbb{R}^d$ (donc $a_j < b_j$ pour tout j). Définissons pour tout $n \geqslant n_0$, tel que $2/n_0 < \min_{1 \leqslant j \leqslant d}(b_j - a_j)$, la fonction continue à support compact h_n par

$$h_n(x) = h_n(x_1, \dots, x_d) := h_{n,1}(x_1) \dots h_{n,d}(x_d),$$

où $h_{n,j}$ est l'unique fonction continue $\mathbb{R} \to [0,1]$, valant 1 sur $[a_j + 1/n, b_j - 1/n]$, 0 hors de $]a_j, b_j[$ et affine sur chacun des intervalles $[a_j, a_j + 1/n]$ et $[b_j - 1/n, b_j]$, voir la figure 8.4.

La fonction h_n est continue, à support compact (son support est exactement le pavé fermé $\overline{C}$). On remarque que pour tout $j \in [\![1, d]\!]$, la suite de fonctions positives $(h_{n,j})_{n \geqslant n_0}$ est croissante et converge en tout point de $\mathbb{R}$ vers l'indicatrice $\mathbf{1}_{]a_j, b_j[}$. On en déduit immédiatement que pour tout $x \in \mathbb{R}^d$, la suite de réels positifs $(h_n(x))_{n \geqslant n_0}$ converge en croissant vers $\mathbf{1}_C(x)$. Par conséquent $(h_n(X))_{n \geqslant n_0}$ et $(h_n(Y))_{n \geqslant n_0}$ sont deux suites croissantes de variables aléatoires positives qui convergent (sur tout Ω) respectivement vers $\mathbf{1}_C(X)$ et $\mathbf{1}_C(Y)$. Comme par l'hypothèse (8.4), $\mathbf{E}h_n(X) = \mathbf{E}h_n(Y)$ pour tout $n \geqslant n_0$, on en déduit en faisant tendre n vers $+\infty$ et en appliquant le théorème de Beppo Levi que $\mathbf{E}\mathbf{1}_C(X) = \mathbf{E}\mathbf{1}_C(Y)$, ce qui s'écrit aussi $P(X \in C) = P(Y \in C)$. □

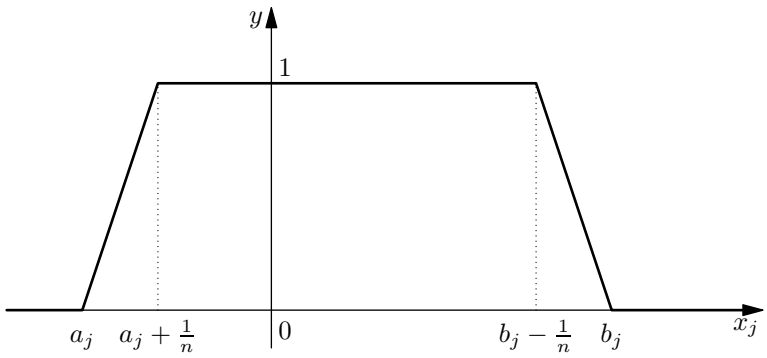


FIGURE 8.4 – Fonction trapèze $h_{n,j}$

Il importe de savoir calculer les $\mathbf{E}h(X)$ quand on connaît la loi de X. Les formules sont analogues à celles déjà données en dimension 1.

Proposition 8.17 (calcul de $\mathbf{E}h(X)$)**.**

a) Si le vecteur aléatoire X est discret et si h est une fonction borélienne $\mathbb{R}^d \to \mathbb{R}$,

$$\mathbf{E}|h(X)| = \sum_{x \in X(\Omega)} |h(x)| P(X = x).$$

Si de plus $\mathbf{E}|h(X)| < +\infty$,

$$\mathbf{E}h(X) = \sum_{x \in X(\Omega)} h(x) P(X = x).$$

b) Si X est à densité f et $h : \mathbb{R}^d \to \mathbb{R}$ est continue bornée sur $\mathbb{R}^d$,

$$\mathbf{E}h(X) = \int_{\mathbb{R}^d} h(x) f(x)\,\mathrm{d}x = \int_{\mathbb{R}^d} h(x_1, \ldots, x_d) f(x_1, \ldots, x_d)\,\mathrm{d}x_1 \ldots\, \mathrm{d}x_d.$$

Cette formule se généralise au cas où h est continue non bornée sur $\mathbb{R}^d$, sous réserve que $\int_{\mathbb{R}^d} |h(x)| f(x)\,\mathrm{d}x < +\infty$.

Nous donnons maintenant une formule de calcul de densité du vecteur aléatoire $g(X)$ où X est un vecteur aléatoire à densité.

Proposition 8.18 (densité d'un vecteur aléatoire image)**.** *Soit $X = (X_1, \ldots, X_d)$ un vecteur aléatoire de $\mathbb{R}^d$ ayant une densité f_X. On suppose de plus que D est un ouvert de $\mathbb{R}^d$ tel que $P(X \in D) = 1$ et que $g : D \to \mathbb{R}^d$ est C^1, injective avec un déterminant jacobien $\mathrm{Jac}(g)$ qui ne s'annule en aucun point de D. Alors g réalise une bijection C^1 d'inverse C^1 (autrement dit un C^1-difféomorphisme) entre D et son image $D' := g(D)$. Le vecteur aléatoire $Y := g(X)$ admet pour densité*

$$f_Y(y) = f_X\big(g^{-1}(y)\big)\,|\,\mathrm{Jac}(g^{-1})(y)|\mathbf{1}_{D'}(y).$$

Nous ne démontrerons pas cette proposition. Le fait que g soit un C^1-difféomorphisme découle d'un théorème d'analyse (théorème d'inversion globale). Rappelons que le jacobien de g est donné par

$$\mathrm{Jac}(g) = \det\Big(\Big[\frac{\partial g_i}{\partial x_j}\Big]_{1\leqslant i,j\leqslant d}\Big),$$

où les g_i sont les applications coordonnées de $g = (g_1, \dots, g_d)$, c.-à-d. $g_i(x_1, \dots, x_d) = \pi_i\big(g(x_1, \dots, x_d)\big)$, π_i étant la projection canonique sur la i^{e} coordonnée dans $\mathbb{R}^d$. Pour calculer $\mathrm{Jac}(g^{-1})(y)$, on peut soit utiliser la formule ci-dessus en remplaçant g par g^{-1} et les x_j par les y_j, soit utiliser la relation

$$\mathrm{Jac}(g^{-1})(y) = \frac{1}{\mathrm{Jac}(g)(g^{-1}(y))}.$$

Un cas particulier facile et important est celui où g est une application linéaire bijective de $\mathbb{R}^d$ sur $\mathbb{R}^d$. Dans ce cas, soit $A = [a_{i,j}]$ la matrice $d \times d$ de g relativement à la base canonique de $\mathbb{R}^d$. Cette matrice est inversible, son déterminant est donc non nul. La i^{e} composante de $g(x)$ est ici $g_i(x_1, \dots, x_d) = \sum_{j=1}^d a_{i,j}x_j$, d'où

$$\forall (i,j) \in [\![1,d]\!]^2, \quad \frac{\partial g_i}{\partial x_j} = a_{i,j}.$$

On en déduit que le jacobien de g comme celui de g^{-1} sont constants et valent

$$\mathrm{Jac}(g) = \det A = \det(g), \quad \mathrm{Jac}(g^{-1}) = \frac{1}{\det A} = \frac{1}{\det g}.$$

Nous obtenons ainsi le corollaire suivant de la proposition 8.18, avec ici $D = \mathbb{R}^d$ et $D' = g(D) = \mathbb{R}^d$.

Corollaire 8.19 (changement de variable linéaire)**.** *Soit $X = (X_1, \dots, X_d)$ un vecteur aléatoire de $\mathbb{R}^d$ ayant une densité f_X et $g : \mathbb{R}^d \to \mathbb{R}^d$ une application linéaire bijective. Alors le vecteur aléatoire $Y := g(X)$ admet pour densité*

$$f_Y(y) = \frac{1}{|\det g|} f_X\big(g^{-1}(y)\big).$$

8.1.2 Covariance de deux variables aléatoires

Le calcul de la variance de la somme de deux variables aléatoires de carré intégrable X et Y ne peut se faire en général qu'à partir de la loi du couple (X, Y). Ceci nous amène à introduire la covariance de (X, Y).

Proposition 8.20 (inégalité de Cauchy-Schwarz)**.** *Si les variables aléatoires réelles X et Y ont des moments d'ordre 2, alors la variable aléatoire XY est intégrable et*

$$|\mathbf{E}(XY)| \leqslant (\mathbf{E}X^2)^{1/2}(\mathbf{E}Y^2)^{1/2}. \tag{8.5}$$

Preuve. L'intégrabilité de la v.a. XY résulte de l'inégalité élémentaire $|XY| \leqslant X^2 + Y^2$. On remarque alors que la fonction trinômiale suivante de la variable réelle t,

$$g : t \longmapsto t^2 \mathbf{E}Y^2 + 2t\mathbf{E}(XY) + \mathbf{E}X^2 = \mathbf{E}(X + tY)^2,$$

est définie sur $\mathbb{R}$ et positive sur $\mathbb{R}$. Ceci n'est possible que si son discriminant est négatif, ce qui s'écrit

$$\Delta' = \big(\mathbf{E}(XY)\big)^2 - (\mathbf{E}X^2)(\mathbf{E}Y^2) \leqslant 0,$$

d'où l'inégalité (8.5). □

Définition 8.21 (covariance)**.** *Si les variables aléatoires réelles X et Y ont des moments d'ordre 2, on appelle covariance du couple aléatoire (X, Y) la quantité :*

$$\mathrm{Cov}(X, Y) = \mathbf{E}\big[(X - \mathbf{E}X)(Y - \mathbf{E}Y)\big].$$

Remarquons que $\mathrm{Cov}(X, X) = \mathrm{Var}\, X$.

Proposition 8.22 (propriétés de la covariance)**.** *Les propriétés suivantes sont vérifiées pour tout couple (X, Y) de v.a. réelles ayant des moments d'ordre 2.*

(i) $\mathrm{Cov}(X, Y) = \mathrm{Cov}(Y, X)$.

(ii) *Pour tous réels a, b, c, d :* $\mathrm{Cov}(aX + b, cY + d) = ac\,\mathrm{Cov}(X, Y)$.

(iii) $|\,\mathrm{Cov}(X, Y)| \leqslant \sigma(X)\sigma(Y)$.

La vérification est laissée au lecteur.

Définition 8.23 (coefficient de corrélation)**.** *Si X et Y sont des variables aléatoires réelles non constantes ayant des moments d'ordre 2, on appelle coefficient de corrélation entre X et Y la quantité :*

$$\rho(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sigma(X)\sigma(Y)}.$$

D'après *(iii)* on a toujours $-1 \leqslant \rho(X, Y) \leqslant 1$. D'autre part il résulte facilement du cas d'égalité dans l'inégalité de Cauchy-Schwarz que $|\rho|$ est maximal lorsque Y est une fonction affine de X : $Y = aX + b$. Quand $\rho = 0$, ce qui arrive en particulier lorsque X et Y sont indépendantes (cf. la section 8.2), on dit que X et Y sont *non corrélées*.

Proposition 8.24 (formule de Koenig pour la covariance)**.** *Si la covariance de X et Y existe, elle peut se calculer par :*

$$\mathrm{Cov}(X, Y) = \mathbf{E}(XY) - \mathbf{E}X\mathbf{E}Y.$$

Preuve. La vérification est analogue à celle de la formule de Koenig pour la variance (qui n'est que le cas particulier $Y = X$) et est laissée en exercice. □

Remarque 8.25 (calcul explicite de la covariance)**.** Les formules de calcul des h-moments appliquées au vecteur aléatoire (X, Y) (prop. 8.17) et aux variables aléatoires réelles X et Y (prop. 7.41 et 7.42) nous donnent pour la covariance (lorsqu'elle existe) les formules explicites suivantes.

Si (X, Y) est discret,

$$\operatorname{Cov}(X,Y) = \sum_{\substack{x\in X(\Omega)\\ y\in Y(\Omega)}} xyP(X=x, Y=y) - \sum_{x\in X(\Omega)} xP(X=x) \sum_{y\in Y(\Omega)} yP(Y=y). \tag{8.6}$$

Si (X, Y) est à densité f, en notant f_X et f_Y les densités marginales,

$$\operatorname{Cov}(X,Y) = \int_{\mathbb{R}^2} xyf(x,y)\,\mathrm{d}x\,\mathrm{d}y - \int_{-\infty}^{+\infty} xf_X(x)\,\mathrm{d}x \int_{-\infty}^{+\infty} yf_Y(y)\,\mathrm{d}y. \tag{8.7}$$

Proposition 8.26 (variance d'une somme)**.** *Si $X_1, \ldots, X_n$ sont des variables aléatoires réelles ayant des moments d'ordre* 2*,*

$$\operatorname{Var}\left(\sum_{i=1}^n X_i\right) = \sum_{i,j=1}^n \operatorname{Cov}(X_i, X_j) \tag{8.8}$$

$$= \sum_{i=1}^n \operatorname{Var} X_i + \sum_{\substack{i,j=1\\ i\neq j}}^n \operatorname{Cov}(X_i, X_j). \tag{8.9}$$

Dans le cas $n = 2$ (8.9) s'écrit :

$$\operatorname{Var}(X+Y) = \operatorname{Var} X + \operatorname{Var} Y + 2\operatorname{Cov}(X,Y). \tag{8.10}$$

Preuve. Pour $n \geqslant 2$ quelconque, l'identité algébrique :

$$\left(\sum_{i=1}^n Y_i\right)^2 = \sum_{i,j=1}^n Y_i Y_j,$$

utilisée avec $Y_i = X_i - \mathbf{E}X_i$ et la linéarité de l'espérance nous donnent :

$$\begin{aligned}
\operatorname{Var}\left(\sum_{i=1}^n X_i\right) &= \mathbf{E}\left\{\sum_{i=1}^n X_i - \mathbf{E}\left(\sum_{i=1}^n X_i\right)\right\}^2 \\
&= \mathbf{E}\left\{\sum_{i=1}^n X_i - \sum_{i=1}^n \mathbf{E}X_i\right\}^2 \\
&= \mathbf{E}\left\{\sum_{i=1}^n Y_i\right\}^2 \\
&= \sum_{i,j=1}^n \mathbf{E}(Y_iY_j) = \sum_{i,j=1}^n \operatorname{Cov}(X_i, X_j).
\end{aligned}$$

□

8.1.3 Espérance et covariance d'un vecteur aléatoire

En dimension 1, l'espérance et la variance d'une variable aléatoire, lorsqu'elles existent, permettent de se faire une idée de la localisation de la loi et de sa dispersion. Elles jouent un rôle important notamment dans le théorème limite central, cf. page 371. Nous allons étendre ces notions au cas des vecteurs aléatoires.

Définition 8.27 (espérance d'un vecteur aléatoire)**.** *Soit $X = (X_1, \dots, X_d)$ un vecteur aléatoire de $\mathbb{R}^d$. On dit qu'il est intégrable si la variable aléatoire positive $\|X\|$ est intégrable ($\mathbf{E}\|X\| < +\infty$), ce qui équivaut à l'intégrabilité de chacune des composantes ($\mathbf{E}|X_i| < +\infty$ pour tout $i \in [\![1, d]\!]$). Dans ce cas on appelle* espérance de X *ou* vecteur d'espérances *de X le vecteur*

$$\mathbf{E}X := (\mathbf{E}X_1, \dots, \mathbf{E}X_d). \tag{8.11}$$

C'est délibérément que nous n'avons pas précisé le choix de la norme dans cette définition. En effet toutes les normes sur $\mathbb{R}^d$ sont équivalentes. Si donc $\|\ \|$ est une norme sur $\mathbb{R}^d$ elle est équivalente en particulier à la norme $\|\ \|_1$ définie par $\|x\|_1 := |x_1| + \cdots + |x_d|$. On en déduit l'existence de deux constantes a et b strictement positives telles que

$$a\big(|X_1| + \cdots + |X_d|\big) \leqslant \|X\| \leqslant b\big(|X_1| + \cdots + |X_d|\big). \tag{8.12}$$

De la première inégalité on tire $|X_i| \leqslant a^{-1}\|X\|$, ce qui montre que l'intégrabilité de $\|X\|$ implique celle de chaque X_i. De la seconde inégalité on déduit que si les v.a. X_i sont toutes intégrables, les $|X_i|$ le sont aussi, ainsi que leur somme finie indexée par i, d'où l'intégrabilité de $\|X\|$. Nous venons de vérifier que l'intégrabilité de $\|X\|$ équivaut à celle de toutes les X_i, ce qui montre aussi que l'intégrabilité de X ne dépend pas du choix de la norme.

Une propriété importante de l'espérance des variables aléatoires est la *linéarité*. Sa généralisation aux *vecteurs* aléatoires est immédiate : si X et Y sont des vecteurs aléatoires intégrables de $\mathbb{R}^d$, a et b des scalaires quelconques, le vecteur aléatoire $aX + bY$ est intégrable et $\mathbf{E}(aX + bY) = a\mathbf{E}X + b\mathbf{E}Y$. Pour le vérifier, il suffit d'appliquer composante par composante, la linéarité de l'espérance des variables aléatoires réelles.

Le résultat suivant nous dit *grosso modo* que l'espérance commute avec les applications linéaires.

Proposition 8.28 (espérance et applications linéaires)**.** *Soit $X = (X_1, \dots, X_d)$ un vecteur aléatoire intégrable de $\mathbb{R}^d$.*

a) Pour toute forme linéaire $u : \mathbb{R}^d \to \mathbb{R}$, la v.a. réelle $u(X)$ est intégrable et

$$\mathbf{E}u(X) = u(\mathbf{E}X). \tag{8.13}$$

b) $\mathbf{E}X$ est le seul vecteur z de $\mathbb{R}^d$ vérifiant $\mathbf{E}u(X) = u(z)$ pour toute forme linéaire u sur $\mathbb{R}^d$.

c) Si $A : \mathbb{R}^d \to \mathbb{R}^j$, $x \mapsto A(x)$ est une application linéaire, le vecteur aléatoire AX de $\mathbb{R}^j$ est intégrable et

$$\mathbf{E}A(X) = A(\mathbf{E}X). \tag{8.14}$$

Preuve. Soit $u : \mathbb{R}^d \to \mathbb{R}$, $x \mapsto a_1x_1 + \cdots + a_dx_d$ une forme linéaire quelconque. L'intégrabilité de la v.a. $u(X)$ résulte immédiatement [1] de l'intégrabilité des X_i puisque $|u(X)| \leqslant |a_1||X_1| + \cdots + |a_d||X_d|$. Par linéarité de l'espérance des v.a. réelles, on a

$$\mathbf{E}u(X) = \mathbf{E}(a_1X_1 + \cdots + a_dX_d) = a_1\mathbf{E}X_1 + \cdots + a_d\mathbf{E}X_d = u(\mathbf{E}X),$$

ce qui justifie (8.13).

D'après le a), il existe au moins un vecteur (non aléatoire) z de $\mathbb{R}^d$ tel que pour toute u, $\mathbf{E}u(X) = u(z)$, c'est $z = \mathbf{E}X$. Si $z' \in \mathbb{R}^d$ a la même propriété, nécessairement $u(z) = u(z')$ pour toute forme linéaire u sur $\mathbb{R}^d$. Ceci s'écrit encore $u(z - z') = 0$ pour toute u, donc en particulier pour les d formes coordonnées $(x_1, \ldots, x_d) \mapsto x_i$. Le vecteur $z - z'$ a ainsi toutes ses coordonnées nulles, donc $z - z' = 0$ et $z = z'$.

Soit $A : \mathbb{R}^d \to \mathbb{R}^j$, $x \mapsto A(x)$ une application linéaire. On sait que

$$\forall x \in \mathbb{R}^d, \quad \|A(x)\| \leqslant \|A\|\,\|x\|,$$

où $\|A\|$ désigne la « norme opérateur » de l'application linéaire A, définie par

$$\|A\| := \sup_{\|x\|=1} \frac{\|A(x)\|}{\|x\|}.$$

On en déduit immédiatement que si X est un vecteur aléatoire intégrable de $\mathbb{R}^d$, le vecteur aléatoire $Y = A(x)$ de $\mathbb{R}^j$ est intégrable puisque

$$\mathbf{E}\|Y\| = \mathbf{E}\|A(X)\| \leqslant \mathbf{E}(\|A\|\,\|X\|) = \|A\|\,\mathbf{E}\|X\| < +\infty.$$

Ceci justifie l'existence de $\mathbf{E}Y$. Soit maintenant v une forme linéaire quelconque sur $\mathbb{R}^j$. D'après le a) appliqué avec $\mathbb{R}^j$ et son dual, on a

$$v(\mathbf{E}Y) = \mathbf{E}v(Y) = \mathbf{E}v\big(A(x)\big) = \mathbf{E}u(X) \quad \text{avec } u := v \circ A.$$

Il est clair que u est une forme linéaire sur $\mathbb{R}^d$, donc d'après le a) appliqué cette fois avec $\mathbb{R}^d$

$$\mathbf{E}u(X) = u(\mathbf{E}X) = (v \circ A)(\mathbf{E}X) = v\big(A(\mathbf{E}X)\big).$$

Nous venons ainsi de vérifier que pour *toute* forme linéaire v sur $\mathbb{R}^j$, $v(\mathbf{E}Y) = v\big(A(\mathbf{E}X)\big)$. On en déduit que $\mathbf{E}Y = A(\mathbf{E}X)$, ce qui établit (8.14). □

Définition 8.29 (matrice de covariance)**.** *Soit $X = (X_1, \ldots, X_d)$ un vecteur aléatoire de carré intégrable, c.-à-d. $\mathbf{E}\|X\|^2 < +\infty$. On appelle* matrice de covariance *de X la matrice carrée K de terme général*

$$K_{i,j} := \mathrm{Cov}(X_i, X_j), \quad (i,j) \in [\![1,d]\!]^2.$$

En utilisant (8.12) comme ci-dessus on a $|X_i| \leqslant a^{-1}\|X\|$ ce qui montre que les composantes X_i d'un vecteur de carré intégrable sont des v.a. de carré intégrable. Ceci justifie l'existence des $\mathrm{Cov}(X_i, X_j)$. La connaissance de la matrice de covariance K de X permet le calcul de $\mathrm{Var}\,u(X)$ pour toute forme linéaire u sur $\mathbb{R}^d$.

1. On pourrait aussi remarquer que $|u(X)| \leqslant \|u\|\,\|X\|$, d'où $\mathbf{E}|u(X)| \leqslant \|u\|\mathbf{E}\|X\| < +\infty$.

Proposition 8.30. *Soit $X = (X_1, \dots, X_d)$ un vecteur aléatoire de carré intégrable et $u : x = (x_1, \dots, x_d) \longmapsto a_1 x_1 + \cdots + a_d x_d$ une forme linéaire sur $\mathbb{R}^d$. Alors la variable aléatoire réelle $u(X)$ est de carré intégrable et*

$$\operatorname{Var} u(X) = \sum_{i,j=1}^{d} a_i a_j \operatorname{Cov}(X_i, X_j). \tag{8.15}$$

Preuve. L'inégalité $|u(X)| \leqslant \|u\| \, \|X\|$ nous donne $\mathbf{E}|u(X)|^2 \leqslant \|u\|^2 \mathbf{E}\|X\|^2 < +\infty$, donc $u(X)$ est de carré intégrable et sa variance est définie. En appliquant la formule générale pour la variance d'une somme et la bilinéarité de la covariance (propositions 8.22 et 8.26), on obtient :

$$\operatorname{Var} u(X) = \operatorname{Var}\left(\sum_{i=1}^{d} a_i X_i \right) = \sum_{i,j=1}^{d} \operatorname{Cov}(a_i X_i, a_j X_j) = \sum_{i,j=1}^{d} a_i a_j \operatorname{Cov}(X_i, X_j).$$

□

Outre son utilité pour le calcul de la variance d'une combinaison linéaire des composantes de X, la matrice de covariance K est porteuse d'informations sur la loi de X. Elle permet notamment de savoir quel sous-espace affine de $\mathbb{R}^d$ est « l'habitat naturel » de X. En particulier, si le déterminant de K est nul, la loi de X est portée par un sous-espace affine de dimension strictement inférieure à d. Et dans ce cas, on peut montrer que la loi de X ne peut avoir de densité au sens de la définition 8.11.

Proposition 8.31 (support d'un vecteur aléatoire de carré intégrable)**.** *Soit X un vecteur aléatoire de carré intégrable, $\mathbf{E}X$ son vecteur d'espérances et K sa matrice de covariance. Alors la loi de X est portée par le sous-espace affine $\mathbf{E}X + \operatorname{Im}(K)$, où $\operatorname{Im}(K)$ désigne le s.e.v. image de $\mathbb{R}^d$ par l'application linéaire de matrice K dans la base canonique. Autrement dit,*

$$P\Big(X - \mathbf{E}X \in \operatorname{Im}(K)\Big) = 1. \tag{8.16}$$

Bien entendu, ce résultat n'a d'intérêt que si $\operatorname{Im}(K)$ est un sous-espace vectoriel strict de $\mathbb{R}^d$, autrement dit si l'application linéaire associée à K n'est pas bijective.

Pour démontrer cette proposition, il est commode d'introduire l'opérateur de covariance de X comme suit [2]. Dans ce qui suit, nous poserons $E = \mathbb{R}^d$, nous noterons E' son dual et E'' son bidual. Tous ces espaces peuvent être assimilés à $\mathbb{R}^d$, mais il importe ici de distinguer leurs rôles respectifs par des notations distinctes. Supposons que pour tout $u \in E'$, $u(X)$ soit de carré intégrable et définissons

$$B : E' \times E' \to \mathbb{R}, \quad (u, v) \mapsto B(u, v) := \operatorname{Cov}\big(u(X), v(X)\big).$$

B est clairement une forme bilinéaire sur $E' \times E'$ et en utilisant l'identification de E'' avec E, on a pour chaque $v \in E'$ un unique élément $y = \mathbf{K}v \in E$ tel que $u(y) =$

2. Les lecteurs novices pourront sauter sans problème ce passage et admettre la proposition.

$B(u,v)$. L'opérateur $\mathbf{K} : E' \to E$ est linéaire, autoadjoint ($u(\mathbf{K}v) = v(\mathbf{K}u)$) et positif ($u(\mathbf{K}u) \geqslant 0$). On l'appelle *l'opérateur de covariance de X*. Il est défini de manière intrinsèque par

$$\forall u, v \in E', \quad u(\mathbf{K}v) = \mathrm{Cov}\big(u(X), v(X)\big). \tag{8.17}$$

On peut donner une représentation matricielle de $\mathbf{K}$ en munissant l'espace d'arrivée $E = \mathbb{R}^d$ de sa base canonique $(e_1, \dots, e_d)$ et l'espace de départ E' de la base duale $(e'_1, \dots, e'_d)$ des formes coordonnées $(e'_i(x) = x_i)$. En choisissant $u = e'_i$ et $v = e'_j$ dans (8.17), on voit que la matrice de $\mathbf{K}$ a pour terme général

$$\mathbf{K}_{i,j} = \mathrm{Cov}\big(e'_i(X), e'_j(X)\big) = \mathrm{Cov}\big(X_i, X_j\big), \quad 1 \leqslant i, j \leqslant d. \tag{8.18}$$

C'est donc la matrice de covariance K du vecteur X.

Preuve de la proposition 8.31. Définissons $H := \{u \in E';\ u(Ku) = 0\}$. Comme l'appartenance de u à H équivaut à $\mathrm{Var}\, u(X) = 0$, il est facile de vérifier que H est un sous-espace vectoriel de E'. Considérons alors $H^0 := \{x \in E\ ;\ \forall u \in H, u(x) = 0\}$. Il est clair que H^0 est un s.e.v. de E. La loi de X est supportée par le sous-espace affine $\mathbf{E}X + H^0$, ce qui signifie que $P(X - \mathbf{E}X \in H^0) = 1$. Pour le vérifier, notons d_0 la dimension de H et choisissons une base quelconque $(f_i)_{1 \leqslant i \leqslant d_0}$ de H. On a alors

$$x \notin H^0 \Leftrightarrow \exists i \in \{1, \dots, d_0\}, f_i(x) \neq 0,$$

d'où

$$P(X - \mathbf{E}X \notin H^0) = P\Big(\bigcup_{i=1}^{d} \{f_i(X - \mathbf{E}X) \neq 0\}\Big) \leqslant \sum_{i=1}^{d} P\big(f_i(X - \mathbf{E}X) \neq 0\big). \tag{8.19}$$

Or par définition de H, $\mathrm{Var}\, f_i(X - \mathbf{E}X) = 0$ pour tout $i \in [\![1, d_0]\!]$, ce qui implique que la variable aléatoire réelle $f_i(X - \mathbf{E}X)$ qui est d'espérance nulle, vaut 0 avec probabilité 1, cf. prop. 7.48. Le membre de droite de (8.19) est donc nul et on a bien $X \in \mathbf{E}X + H^0$ presque sûrement. □

Pour l'étude des vecteurs aléatoires et de leur convergence en loi, les formes linéaires sur $\mathbb{R}^d$ sont un outil très pratique car elles permettent de se ramener à la dimension 1. Nous admettrons le résultat suivant.

Proposition 8.32. *Soient X et Y deux vecteurs aléatoires de $\mathbb{R}^d$ tels que pour* toute *forme linéaire $u : \mathbb{R}^d \to \mathbb{R}$, les variables aléatoires réelles $u(X)$ et $u(Y)$ aient même loi. Alors les vecteurs aléatoires X et Y ont même loi.*

Remarque 8.33. Attention à l'hypothèse « pour toute u », on ne peut pas l'affaiblir en la remplaçant par « pour toute u d'une base du dual de $\mathbb{R}^d$ ». Sinon en prenant les formes coordonnées $u : x = (x_1, \dots, x_d) \mapsto x_i$, on en déduirait que si pour tout $i \in [\![1, d]\!]$, X_i a même loi que Y_i, alors les vecteurs $X = (X_1, \dots, X_d)$ et $Y = (Y_1, \dots, Y_d)$ ont même loi. Or on sait, cf. remarque 8.6, que ceci est faux.

8.2 Indépendance de variables et vecteurs aléatoires

8.2.1 Suites indépendantes

Définition 8.34 (indépendance de n variables aléatoires)**.** *Les variables aléatoires réelles $X_1, \dots, X_n$ définies sur le même $(\Omega, \mathcal{F}, P)$ sont dites indépendantes si :*

$$\forall B_1, \dots, B_n \in \mathrm{Bor}(\mathbb{R}), \quad P\big(X_1 \in B_1, \dots, X_n \in B_n\big) = P(X_1 \in B_1) \dots P(X_n \in B_n). \tag{8.20}$$

Remarques 8.35.

1. La condition (8.20) équivaut à *l'indépendance mutuelle* des évènements $\{X_1 \in B_1\}, \dots, \{X_n \in B_n\}$ *pour tout choix des boréliens B_i*. En effet la liberté de choisir $B_i = \mathbb{R}$ permet « d'effacer » pour ce choix le rôle de X_i et ainsi d'obtenir à partir de (8.20) une égalité de même type pour l'intersection de toute sous-famille finie des évènements considérés. La rédaction détaillée de la justification de cette équivalence est laissée en exercice. Le même argument montre que l'indépendance de $X_1, \dots, X_n$ entraîne celle des $X_{i_1}, \dots, X_{i_k}$ pour tout choix d'indices $1 \leqslant i_1 < i_2 < \dots < i_k \leqslant n$.
2. L'indépendance des X_i, $1 \leqslant i \leqslant n$ est une propriété plus forte que leur indépendance deux à deux.
3. Pour vérifier l'indépendance de $X_1, \dots, X_n$, il suffit de tester l'égalité dans (8.20) pour des B_i quelconques pris dans l'une des sous-familles suivantes de $\mathrm{Bor}(\mathbb{R})$:
   – les intervalles $]-\infty, b_i]$;
   – les intervalles $]a_i, b_i]$;
   – les intervalles $]a_i, b_i[$;
   – les intervalles $[a_i, b_i]$;
   – les intervalles $]b_i, +\infty[$.
   Cette réduction repose essentiellement sur le fait que chacune de ces familles d'intervalles engendre $\mathrm{Bor}(\mathbb{R})$ et est stable par intersections finies (en autorisant $b_i < a_i$ pour avoir l'ensemble vide dans chaque famille). Nous nous contenterons de cette justification pour ne pas trop sortir du cadre de cet ouvrage.
4. Clairement l'indépendance n'est pas affectée par une permutation sur les indices i des variables aléatoires. On peut donc parler d'indépendance d'une famille finie de variables aléatoires sans se préoccuper de l'ordre d'indexation.

La définition de l'indépendance se généralise comme suit à une famille finie de vecteurs[3] aléatoires.

Définition 8.36 (indépendance de n vecteurs aléatoires)**.** *Pour $i = 1, \dots, n$, soient $X_i : \Omega \to \mathbb{R}^{d_i}$ des vecteurs aléatoires sur le même espace probabilisé $(\Omega, \mathcal{F}, P)$. On dit qu'ils sont indépendants si l'égalité*

$$P\big(X_1 \in B_1, \dots, X_n \in B_n\big) = P(X_1 \in B_1) \dots P(X_n \in B_n) \tag{8.21}$$

3. Il va sans dire, mais mieux en le disant, que l'indépendance des vecteurs dont il est question ici n'a rien à voir avec l'indépendance linéaire !

est vérifiée pour tout choix des boréliens $B_i \in \mathrm{Bor}(\mathbb{R}^{d_i})$, $i = 1, \dots, n$.

Notons qu'avec cette définition, la collection de vecteurs aléatoires considérée peut être complètement hétéroclite quant aux dimensions, y compris avec $d_i = 1$ pour certaines valeurs de i (les X_i correspondants étant alors des variables aléatoires réelles).

Une propriété bien commode de l'indépendance des v.a. est *l'hérédité*. Avant de l'énoncer formellement, voyons sa signification sur un exemple. Supposons que les v.a. réelles $X_1, \dots, X_5$ soient indépendantes. Alors les trois variables aléatoires Y_1, Y_2, Y_3 suivantes sont indépendantes

$$Y_1 := X_1 + X_2, \quad Y_2 := X_3 \sin X_4, \quad Y_3 := \exp(X_5^2 - X_5).$$

Il en va de même pour les vecteurs aléatoires Z_1, Z_2, Z_3

$$Z_1 := (X_1, X_2), \quad Z_2 := (X_3 + \cos X_4, X_4^2), \quad Z_3 := (X_5^3, 2X_5, X_5^2).$$

Nous énonçons le résultat à partir d'une suite de variables aléatoires indépendantes pour ne pas trop alourdir les notations, mais il se généralise à une famille finie hétéroclite de vecteurs aléatoires indépendants au sens de la définition 8.36.

Proposition 8.37 (hérédité de l'indépendance)**.** *Soient* $X_1, \dots, X_n$ *des variables aléatoires réelles indépendantes. Découpons* $[\![1, n]\!]$ *en* k *blocs disjoints, en notant* m_j *le cardinal du* j^e *bloc et* $n_j = \sum_{1 \leqslant l \leqslant j} m_l$*, avec* $n_0 := 0$*. Pour* $j \in [\![1, k]\!]$*, associons au* j^e *bloc une fonction borélienne* $h_j : \mathbb{R}^{m_j} \to \mathbb{R}^{d_j}$*. Posons enfin* $Z_j := h_j(X_{n_{j-1}+1}, \dots, X_{n_j})$*. Alors la suite finie* $(Z_j)_{1 \leqslant j \leqslant k}$ *de vecteurs aléatoires est indépendante (au sens de la définition 8.36).*

En prenant pour chaque j, $d_j = m_j$ et h_j égale à l'identité $\mathbb{R}^{m_j} \to \mathbb{R}^{m_j}$, on voit que la proposition 8.37 contient en particulier le résultat suivant.

Corollaire 8.38 (indépendance de blocs disjoints)**.** *Si* $X_1, \dots, X_n$ *sont des variables aléatoires réelles indépendantes, les « blocs disjoints »* $Y_j := (X_{n_{j-1}+1}, \dots, X_{n_j})$*, où* $0 = n_0 < n_1 < n_2 < \cdots < n_k = n$*, sont des vecteurs aléatoires indépendants.*

Remarque 8.39. Compte-tenu de la remarque 8.35-4, on généralise facilement la proposition 8.37 et le corollaire 8.38 au cas où les « blocs » ne sont plus forcément indexés par des entiers consécutifs, pourvu que les blocs d'indices correspondants restent deux à deux disjoints. Ainsi par exemple si $X_1, \dots, X_7$ sont indépendantes, (X_3, X_1), (X_2, X_5, X_7) et (X_6, X_4) sont des vecteurs aléatoires indépendants.

La démonstration détaillée de la proposition 8.37 sortirait du programme de ce cours où sont seulement exigibles l'énoncé et son utilisation pratique. Ce qui suit s'adresse aux lecteurs plus avancés ou désireux d'en savoir plus. Le lecteur débutant peut sauter ce passage et aller directement à la définition 8.41.

La preuve informelle[4] de la proposition 8.37 que nous décrivons maintenant distingue deux idées dans cette proposition. La première est l'hérédité pour les blocs

4. Nous ne prétendons pas donner une preuve complète, mais plutôt pointer les difficultés et indiquer l'outillage mathématique utilisé pour les surmonter.

(corollaire 8.38) et utilise la construction de « probabilités produits ». La deuxième est l'hérédité pour les images des blocs par les transformations boréliennes h_j et repose sur la notion d'indépendance de sous-tribus de $\mathcal{F}$.

Preuve informelle de l'hérédité pour les blocs. Partons donc de l'égalité (8.20) qui exprime notre hypothèse d'indépendance des X_i. Avec les notations pour les blocs introduites dans l'énoncé de la proposition 8.37, identifions $\mathbb{R}^n$ avec $\mathbb{R}^{m_1} \times \cdots \times \mathbb{R}^{m_k}$. Pour $j = 1, \ldots, k$, notons $\mathcal{C}_j$ la famille des produits cartésiens $B'_j = B_{n_{j-1}+1} \times \cdots \times B_{n_j}$ de m_j boréliens quelconques de $\mathbb{R}$. En rappelant que $Y_j := (X_{n_{j-1}+1}, \ldots, X_{n_j})$, ceci nous permet de réécrire (8.20) sous la forme suivante vérifiée pour tous choix des $B'_j \in \mathcal{C}_j$.

$$P(Y_1 \in B'_1, \ldots, Y_k \in B'_k) = \prod_{j=1}^{k} \prod_{n_{j-1}<i\leqslant n_j} P(X_i \in B_i). \tag{8.22}$$

En particulier si on choisit ci-dessus pour un j fixé, $B_i = \mathbb{R}$ pour tous les i hors de $]n_{j-1}, n_j]$, chaque B'_l pour $l \neq j$ est égal à $\mathbb{R}^{m_l}$ et on obtient après effacement des conditions inutiles du type $Y_l \in \mathbb{R}^{m_l}$ ou $X_i \in \mathbb{R}$,

$$\forall B'_j \in \mathcal{C}_j, \quad P(Y_j \in B'_j) = \prod_{n_{j-1}<i\leqslant n_j} P(X_i \in B_i). \tag{8.23}$$

Ceci étant vrai pour j quelconque, on obtient en reportant dans (8.22) :

$$\forall B'_1 \in \mathcal{C}_1, \ldots, \forall B'_k \in \mathcal{C}_k, \quad P(Y_1 \in B'_1, \ldots, Y_k \in B'_k) = \prod_{j=1}^{k} P(Y_j \in B'_j). \tag{8.24}$$

Pour en déduire que la suite finie de vecteurs aléatoires $Y_1, \ldots, Y_k$ est indépendante au sens de la définition 8.36, il faudrait pouvoir étendre (8.24) à tous les $B'_j \in \mathrm{Bor}(\mathbb{R}^{m_j})$, ce qui est loin d'être trivial. Notons que si $m_j \geqslant 2$, $\mathcal{C}_j$ n'est pas une tribu et est strictement incluse dans $\mathrm{Bor}(\mathbb{R}^{m_j})$. Par exemple le disque unité ouvert de $\mathbb{R}^2$ est un borélien de $\mathbb{R}^2$ puisque c'est un ouvert et il est impossible de l'écrire comme produit cartésien de deux sous-ensembles de $\mathbb{R}$ (exercice !). Notons $\mu := P_{Y_1,\ldots,Y_k}$, la loi du vecteur aléatoire hétéroclite $(Y_1, \ldots, Y_k)$ de $\mathbb{R}^{m_1} \times \cdots \times \mathbb{R}^{m_k}$, définie *via* l'indentification de cet espace avec $\mathbb{R}^n$ par :

$$\forall D \in \mathrm{Bor}(\mathbb{R}^n), \quad \mu(D) = P\big((Y_1, \ldots, Y_k) \in D\big).$$

Définissons la tribu produit $\mathrm{Bor}(\mathbb{R}^{m_1}) \otimes \cdots \otimes \mathrm{Bor}(\mathbb{R}^{m_k})$ comme la tribu sur $\mathbb{R}^{m_1} \times \cdots \times \mathbb{R}^{m_k}$ engendrée par la famille des produits cartésiens $D_1 \times \cdots \times D_k$, où chaque D_j est un borélien quelconque de $\mathbb{R}^{m_j}$. On peut montrer que cette tribu coïncide avec la tribu borélienne de $\mathbb{R}^{m_1} \times \cdots \times \mathbb{R}^{m_k}$, donc de $\mathbb{R}^n$. On montre aussi que l'on peut définir sur cette tribu produit la *probabilité produit* $\nu = P_{Y_1} \otimes \cdots \otimes P_{Y_k}$ comme l'unique mesure vérifiant

$$\forall D_1 \in \mathrm{Bor}(\mathbb{R}^{m_1}), \ldots, \forall D_k \in \mathrm{Bor}(\mathbb{R}^{m_k}), \quad \nu(D_1 \times \cdots \times D_k) = P_{Y_1}(D_1) \ldots P_{Y_k}(D_k).$$

Notons enfin $\mathcal{C}$ la famille des sous-ensembles de $\mathbb{R}^{m_1} \times \cdots \times \mathbb{R}^{m_k}$ de la forme $B'_1 \times \cdots \times B'_k$, où $B'_j \in \mathcal{C}_j$, $j = 1, \ldots, k$. Muni de tout cet attirail, on peut réécrire (8.24) sous la forme :

$$\forall C \in \mathcal{C}, \quad \mu(C) = \nu(C).$$

Finalement, on vérifie que $\mathcal{C}$ est une famille stable par intersections finies, qui engendre la tribu borélienne de $\mathbb{R}^{m_1} \times \cdots \times \mathbb{R}^{m_k}$. Comme les mesures de probabilité μ et ν coïncident sur $\mathcal{C}$, elles coïncident sur la tribu engendrée [5], autrement dit $\mu(D) = \nu(D)$ pour tout $D \in \mathrm{Bor}(\mathbb{R}^{m_1} \times \cdots \times \mathbb{R}^{m_k})$. Comme cette tribu contient tous les produits cartésiens $D_1 \times \cdots \times D_k$ de boréliens quelconques des $\mathbb{R}^{m_j}$, on en déduit l'extension de (8.24) avec les D_j au lieu des B'_j, obtenant ainsi l'indépendance de la suite de vecteurs aléatoires $Y_1, \ldots, Y_k$. □

Preuve informelle de l'hérédité par transformations boréliennes. Soient Ω et Ω' deux ensembles quelconques, $\mathcal{F}'$ une tribu sur Ω' et $f : \Omega \to \Omega'$ quelconque. Notons $f^{-1}(\mathcal{F}')$ la famille $\{f^{-1}(B);\, B \in \mathcal{F}'\}$, en rappelant que si $B \subset \Omega'$, $f^{-1}(B) = \{\omega \in \Omega;\, f(\omega) \in B\}$. Il est facile de voir que $f^{-1}(\mathcal{F}')$ est une tribu sur Ω. Cela résulte du fait que la réunion et le passage au complémentaire commutent avec l'inverse ensembliste (cf. proposition 6.4) et que $\mathcal{F}'$ est une tribu sur Ω'. Cette tribu $f^{-1}(\mathcal{F}')$ est appelée tribu engendrée par f (relativement à $\mathcal{F}'$) et notée aussi $\sigma(f)$.

Si Ω est muni lui-même d'une tribu $\mathcal{F}$, la condition $f^{-1}(\mathcal{F}') \subset \mathcal{F}$ équivaut tout simplement à la mesurabilité de f pour les tribus $\mathcal{F}$ – $\mathcal{F}'$. En particulier si X est une application $\Omega \to \mathbb{R}^d$, dire que X est un vecteur aléatoire sur $(\Omega, \mathcal{F})$ s'écrit $X^{-1}(\mathrm{Bor}(\mathbb{R}^d)) \subset \mathcal{F}$.

Arrivés là, il nous faut donner une définition qui jette un nouvel éclairage sur l'indépendance des variables ou vecteurs aléatoires.

Définition 8.40 (indépendance de n sous-tribus de $\mathcal{F}$)**.** *Soient $(\Omega, \mathcal{F}, P)$ un espace probabilisé et $\mathcal{F}_1, \ldots, \mathcal{F}_n$ des sous-tribus [6] de $\mathcal{F}$. On dit qu'elles sont indépendantes si*

$$\forall A_1 \in \mathcal{F}_1, \ldots, A_n \in \mathcal{F}_n, \quad P\Big(\bigcap_{i=1}^{n} A_i\Big) = \prod_{i=1}^{n} P(A_i). \tag{8.25}$$

Grâce à cette définition, on peut voir que l'indépendance des variables aléatoires $X_1, \ldots, X_n$ est exactement l'indépendance des tribus engendrées [7] $\sigma(X_1), \ldots, \sigma(X_n)$. On peut réécrire de la même façon l'indépendance des vecteurs aléatoires vue à la définition 8.36, la seule adaptation étant que $\sigma(X_i) = X_i^{-1}\big(\mathrm{Bor}(\mathbb{R}^{d_i})\big)$ au lieu de $X_i^{-1}\big(\mathrm{Bor}(\mathbb{R})\big)$.

Il est clair d'après la définition 8.40 que si les tribus $\mathcal{F}_1, \ldots, \mathcal{F}_n$ sont des sous-tribus de $\mathcal{F}$ indépendantes et si pour chaque i, $\mathcal{G}_i$ est une sous-tribu de $\mathcal{F}_i$, alors $\mathcal{G}_1, \ldots, \mathcal{G}_n$ sont aussi des sous-tribus indépendantes de $\mathcal{F}$. En d'autres termes, l'indépendance des tribus est héritée par leurs sous-tribus.

5. Par le théorème d'unicité des mesures [15, Chap. 1, Th. 1.34], appliqué dans le cas de mesures finies.

6. Une sous-tribu de $\mathcal{F}$ est une sous-famille de $\mathcal{F}$ qui possède encore la structure de tribu.

7. Qui sont bien des sous-tribus de $\mathcal{F}$, à cause de la mesurabilité des X_i.

Disposant maintenant de tous les ingrédients, nous pouvons prouver la proposition 8.37 comme suit. D'abord l'indépendance de la suite $X_1, \dots, X_n$ passe à celle des vecteurs blocs $Y_j = (X_{n_{j-1}+1}, \dots, X_{n_j})$, $j = 1, \dots, k$, comme nous l'avons établi ci-dessus. Cette indépendance des Y_j équivaut à celle des sous-tribus $\mathcal{F}_j := \sigma(Y_j)$ de $\mathcal{F}$. Notre but est d'établir l'indépendance des vecteurs images $Z_j = h_j(Y_j)$, qui équivaut à l'indépendance des tribus $\mathcal{G}_j := \sigma(Z_j)$. Celle-ci découle par héritage de l'indépendance des tribus $\sigma(Y_j)$, $j = 1, \dots, k$ en raison des inclusions $\sigma(Z_j) \subset \sigma(Y_j)$ que l'on peut justifier comme suit :

$$\begin{aligned} \sigma(Z_j) &= \big\{ Z_j^{-1}(B) \,;\, B \in \mathrm{Bor}(\mathbb{R}^{d_j}) \big\} \\ &= \big\{ (h_j \circ Y_j)^{-1}(B) \,;\, B \in \mathrm{Bor}(\mathbb{R}^{d_j}) \big\} \\ &= \big\{ Y_j^{-1}\big(h_j^{-1}(B)\big) \,;\, B \in \mathrm{Bor}(\mathbb{R}^{d_j}) \big\} \end{aligned}$$

et $h_j : \mathbb{R}^{m_j} \to \mathbb{R}^{d_j}$ étant borélienne, $h_j^{-1}(B) \in \mathrm{Bor}(\mathbb{R}^{m_j})$ pour tout $B \in \mathrm{Bor}(\mathbb{R}^{d_j})$, d'où

$$\sigma(Z_j) \subset \big\{ Y_j^{-1}(C) \,;\, C \in \mathrm{Bor}(\mathbb{R}^{m_j}) \big\} = \sigma(Y_j).$$

□

Définition 8.41 (indépendance d'une suite infinie de v.a.)**.** *Soit $(X_i)_{i\in\mathbb{N}}$, une suite de variables aléatoires définies sur le même espace probabilisé $(\Omega, \mathcal{F}, P)$. On dit qu'elles sont indépendantes si toute sous-suite finie est indépendante au sens de la définition 8.34, c'est-à-dire si pour tout ensemble* fini *$\varnothing \neq K \subset \mathbb{N}$, et toute famille $(B_i)_{i\in K}$ de boréliens de $\mathbb{R}$,*

$$P\Big(\bigcap_{i\in K} \{X_i \in B_i\} \Big) = \prod_{i\in K} P(X_i \in B_i). \tag{8.26}$$

Cette définition se généralise au prix d'alourdissements d'écriture à une suite de vecteurs aléatoires.

8.2.2 Indépendance des composantes

Nous examinons maintenant la caractérisation de l'indépendance des composantes d'un vecteur aléatoire, en nous restreignant aux deux cas particuliers importants des vecteurs aléatoires discrets ou à densité.

Proposition 8.42 (vecteur aléatoire discret à composantes indépendantes)**.** *Soit $X = (X_1, \dots, X_d)$ un vecteur aléatoire discret. Ses composantes X_i sont indépendantes si et seulement si pour tout $(x_1, \dots, x_d) \in X(\Omega)$,*

$$P(X_1 = x_1, \dots, X_d = x_d) = P(X_1 = x_1) \dots P(X_d = x_d). \tag{8.27}$$

Preuve. L'indépendance des X_i implique clairement (8.27) en prenant $B_i = \{x_i\}$, pour $i \in [\![1, d]\!]$ dans la définition 8.34. Pour la réciproque, on suppose que le vecteur aléatoire discret X vérifie (8.27) et il s'agit de montrer que $P_X(B) = P_{X_1}(B_1) \dots P_{X_d}(B_d)$, pour tout $B := B_1 \times \cdots \times B_d$ où les B_i sont des boréliens quelconques de $\mathbb{R}$. Comme

X est discret, chaque $X_i(\Omega)$ est au plus dénombrable (cf. le commentaire après la définition 8.8) donc $E := X_1(\Omega) \times \cdots \times X_d(\Omega)$ est au plus dénombrable. De plus E contient $X(\Omega)$ et si $x \in E \backslash X(\Omega)$, $P(X = x) = 0$. Ces remarques nous permettent de démarrer le calcul de $P_X(B)$ ainsi :

$$P_X(B) = \left(\sum_{x \in X(\Omega)} P(X = x)\delta_x \right)(B) = \sum_{x \in X(\Omega)} P(X = x)\delta_x(B) = \sum_{x \in C} P(X = x),$$

en notant $C = E \cap B = C_1 \times \cdots \times C_d$, où $C_i := \big(X_i(\Omega) \cap B_i\big)$ (vérifiez !). En utilisant l'hypothèse (8.27) et le produit de d « séries » à termes positifs [8] :

$$\begin{aligned}
P_{(X_1,\ldots,X_d)}(B) &= \sum_{(x_1,\ldots,x_d) \in C} P_{(X_1,\ldots,X_d)}\big(\{(x_1,\ldots,x_d)\}\big) \\
&= \sum_{(x_1,\ldots,x_d) \in C} P(X_1 = x_1, \ldots, X_d = x_d) \\
&= \sum_{(x_1,\ldots,x_d) \in C_1 \times \cdots \times C_d} P(X_1 = x_1) \ldots P(X_d = x_d) \qquad (8.28) \\
&= \left(\sum_{x_1 \in C_1} P(X_1 = x_1) \right) \times \cdots \times \left(\sum_{x_d \in C_d} P(X_d = x_d) \right) \qquad (8.29) \\
&= P(X_1 \in B_1) \ldots P(X_d \in B_d).
\end{aligned}$$

Les boréliens B_i étant quelconques, ceci établit l'indépendance des variables aléatoires $X_1, \ldots, X_d$. □

Passons au cas d'un vecteur aléatoire à densité, pour lequel il est commode d'introduire la notation suivante. Si $f_1, \ldots, f_d$ sont des applications $\mathbb{R} \to \mathbb{R}$, on définit l'application « produit tensoriel » des f_i par

$$f_1 \otimes \cdots \otimes f_d : \mathbb{R}^d \to \mathbb{R}, \quad (x_1, \ldots, x_d) \longmapsto f_1(x_1) \ldots f_d(x_d). \qquad (8.30)$$

On prendra bien garde de ne pas confondre ce produit tensoriel avec le produit ordinaire. Par exemple si f est l'identité sur $\mathbb{R}$, $f \otimes f$ est l'application $(s, t) \mapsto st$, tandis que $ff = f^2 : \mathbb{R} \to \mathbb{R}$, $s \mapsto s^2$.

Proposition 8.43 (vecteur à densité et indépendance des composantes)**.** *Soit $X = (X_1, \ldots, X_d)$ un vecteur aléatoire de $\mathbb{R}^d$.*

a) Si les composantes de X sont indépendantes et si chaque X_i ($i \in [\![1, d]\!]$) a une densité f_{X_i}, alors X admet pour densité la fonction $f_{X_1} \otimes \cdots \otimes f_{X_d}$.

b) Si X admet une densité de la forme $f = f_1 \otimes \cdots \otimes f_d$, où les f_i sont des fonctions $\mathbb{R} \backslash K_i \to \mathbb{R}^+$ (les K_i étant finis éventuellement vides), alors les v.a. réelles X_i sont indépendantes et chaque X_i admet une densité $f_{X_i} = c_i f_i$ avec des constantes $c_i \in]0, +\infty[$ dont le produit $c_1 \ldots c_d$ vaut 1.

8. Certaines d'entre elles peuvent n'être que des sommes finies. La formule utilisée pour passer de (8.28) à (8.29) repose sur le théorème de sommation par paquets des familles à termes positifs et sa vérification est essentiellement la même que pour le produit de deux séries.

Preuve du a). Calculons $P_X(C)$ pour $C := [a_1, b_1] \times \cdots \times [a_d, b_d]$ pavé fermé borné quelconque de $\mathbb{R}^d$. On commence par écrire l'évènement $\{X \in C\}$ comme une intersection d'évènements en notant que

$$\{X \in C\} = \big\{(X_1, \ldots, X_d) \in [a_1, b_1] \times \cdots \times [a_d, b_d]\big\} = \big\{\forall i \in [\![1, d]\!],\ X_i \in [a_i, b_i]\big\},$$

d'où, les X_i étant indépendantes et à densité,

$$P(X \in C) = P\Big(\bigcap_{i=1}^{d} \{X_i \in [a_i, b_i]\}\Big) = \prod_{i=1}^{d} P\big(X_i \in [a_i, b_i]\big) = \prod_{i=1}^{d} \int_{a_i}^{b_i} f_{X_i}(t_i)\,\mathrm{d}t_i.$$

Par le théorème 4.91, ce produit d'intégrales de fonctions positives peut s'écrire comme une intégrale multiple :

$$\prod_{i=1}^{d} \int_{a_i}^{b_i} f_{X_i}(t_i)\,\mathrm{d}t_i = \int_{[a_1,b_1]\times\cdots\times[a_d,b_d]} f_{X_1}(t_1) \ldots f_{X_d}(t_d)\,\mathrm{d}t_1 \ldots\, \mathrm{d}t_d.$$

Finalement, nous obtenons ainsi

$$P(X \in C) = \int_C f_{X_1} \otimes \cdots \otimes f_{X_d}(t_1, \ldots, t_d)\,\mathrm{d}t_1 \ldots\, \mathrm{d}t_d,$$

pour tout pavé fermé borné C. Pour en déduire en vertu de la définition 8.10 que le vecteur aléatoire X admet bien pour densité $f := f_{X_1} \otimes \cdots \otimes f_{X_d}$, il nous reste seulement à vérifier que f ainsi construite est bien une densité de probabilité sur $\mathbb{R}^d$. Ceci résulte immédiatement du théorème 4.91. □

Preuve du b). Puisque X admet pour densité $f = f_1 \otimes \cdots \otimes f_d$, on sait (proposition 8.14) que ses lois marginales sont toutes à densité et qu'une densité de X_i s'obtient en intégrant f par rapport à toutes les variables d'indice différent de i. Ceci nous donne pour $f_{X_i}(x_i)$ l'intégrale :

$$\int_{\mathbb{R}^{i-1}\times\mathbb{R}^{d-i}} f_1(t_1) \ldots f_{i-1}(t_{i-1}) f_i(x_i) f_{i+1}(t_{i+1}) \ldots f_d(t_d)\,\mathrm{d}t_1 \ldots\, \mathrm{d}t_{i-1}\,\mathrm{d}t_{i+1} \ldots \mathrm{d}t_d.$$

Dans cette intégration avec x_i fixé et relativement aux variables muettes t_j, $j \neq i$, $f_i(x_i)$ est une constante que l'on peut sortir de l'intégrale multiple. On obtient ainsi la relation $f_{X_i}(x_i) = c_i f_i(x_i)$, où c_i désigne l'intégrale multiple relative aux t_j ($j \neq i$) et ne dépend pas de x_i. Ce calcul étant valable pour tout réel x_i, on a bien $f_{X_i} = c_i f_i$, avec une constante c_i dans $\overline{\mathbb{R}}_+$ (pour l'instant). En fait $c_i \neq 0$, car sinon f_{X_i} serait identiquement nulle sur son ensemble de définition, ce qui l'empêcherait d'être une densité car cela impliquerait la nullité de $\int_{-\infty}^{+\infty} f_{X_i}(t)\,\mathrm{d}t$.

D'après la remarque 8.35-3, pour établir l'indépendance des $X_1, \ldots, X_d$, il suffit de tester l'égalité (8.20) pour des boréliens B_i de la forme $[a_i, b_i]$. Notons $C :=$

$[a_1, b_1] \times \cdots \times [a_d, b_d]$. Compte-tenu de la forme particulière de f, on a ici en appliquant à nouveau le théorème 4.91,

$$P(X \in C) = \int_C f_1(t_1) \dots f_d(t_d)\, \mathrm{d}t_1 \dots \mathrm{d}t_d = \prod_{i=1}^{d} \int_{a_i}^{b_i} f_i(t_i)\, \mathrm{d}t_i. \tag{8.31}$$

D'autre part, en utilisant les relations $f_{X_i} = c_i f_i$, on obtient :

$$\prod_{i=1}^{d} P(X_i \in [a_i, b_i]) = \prod_{i=1}^{d} \int_{a_i}^{b_i} c_i f_i(t_i)\, \mathrm{d}t_i = (c_1 \dots c_d) \prod_{i=1}^{d} \int_{a_i}^{b_i} f_i(t_i)\, \mathrm{d}t_i. \tag{8.32}$$

Par comparaison de (8.31) et (8.32), on voit alors que pour tout pavé fermé borné $C = [a_1, b_1] \times \cdots \times [a_d, b_d]$,

$$\prod_{i=1}^{d} P(X_i \in [a_i, b_i]) = (c_1 \dots c_d) P(X \in C). \tag{8.33}$$

En prenant une suite $C_n = [a_{1,n}, b_{1,n}] \times \cdots \times [a_{d,n}, b_{d,n}]$ croissante pour l'inclusion et de réunion $\mathbb{R}^d$, on obtient par continuité séquentielle croissante de P_X et des P_{X_i} :

$$\prod_{i=1}^{d} P(X_i \in \mathbb{R}) = (c_1 \dots c_d) P(X \in \mathbb{R}^d),$$

d'où

$$c_1 \dots c_d = 1. \tag{8.34}$$

En réinjectant cette valeur dans (8.33), on conclut à l'indépendance des X_i. Notons en passant que (8.34) interdit que l'un des c_i vaille $+\infty$ (on a déjà vu qu'aucun ne peut être nul). □

Exemple 8.44. Supposons que les variables aléatoires réelles $X_1, \dots, X_d$ soient indépendantes et que pour $i \in [\![1, d]\!]$, X_i soit gaussienne de loi $\mathfrak{N}(0, \sigma_i)$, avec $\sigma_i > 0$. La proposition 8.43 a) nous dit alors que le vecteur $X = (X_1, \dots, X_d)$ admet pour densité :

$$f(t_1, \dots, t_d) = \frac{1}{(2\pi)^{d/2}} \frac{1}{\sigma_1 \dots \sigma_d} \exp\left(- \frac{t_1^2}{2\sigma_1^2} - \cdots - \frac{t_d^2}{2\sigma_d^2} \right). \tag{8.35}$$

Le vecteur X suit une loi gaussienne sur $\mathbb{R}^d$. Il s'agit ici d'un cas particulier de *vecteur gaussien*. Il existe des lois gaussiennes sur $\mathbb{R}^d$ pour lesquelles les composantes ne sont pas indépendantes. Par ailleurs, il existe des lois gaussiennes sur $\mathbb{R}^d$ n'ayant pas de densité [9], cf. la remarque 10.32.

9. Dans ce cas, il existe un sous-espace affine de $\mathbb{R}^d$, strictement inclus dans $\mathbb{R}^d$ et de probabilité 1 pour cette loi gaussienne.

Remarque 8.45. L'exemple 8.44 ne contredit pas la remarque 8.6 qui affirme que la seule connaissance des lois marginales ne suffit pas à reconstruire la loi du vecteur. En effet ici outre la connaissance des lois marginales, on dispose d'une information supplémentaire essentielle, l'indépendance des composantes. C'est d'ailleurs un exemple d'un principe tout à fait général illustré par les propositions 8.42 et 8.43 : si on connaît les lois marginales *et si on sait* que les composantes sont indépendantes, alors on peut reconstruire la loi du vecteur aléatoire [10].

Exemple 8.46. Soit $X = (X_1, X_2)$ un vecteur aléatoire de densité f donnée par

$$f(s,t) := \frac{1}{(2\pi^3)^{1/2}} \frac{1}{(1+s^2)(t\mathrm{e}^t)^{1/2}} \mathbf{1}_{]0,+\infty[}(t),$$

avec la convention habituelle pour l'usage des indicatrices dans les formules explicites (cf. remarque 6.22). La densité est bien le produit d'une fonction de la seule variable s par une fonction de la seule variable t. On peut prendre par exemple $f_1(s) := (1+s^2)^{-1}$ et $f_2(t) := (2\pi^3)^{-1/2}(t\mathrm{e}^t)^{-1/2}\mathbf{1}_{]0,+\infty[}(t)$. On en déduit par la proposition 8.43 b) que les v.a. X_1 et X_2 sont indépendantes et à densité. On sait de plus qu'alors X_i admet une densité $f_{X_i} = c_i f_i$ $(i = 1, 2)$. On détermine c_1 en écrivant :

$$1 = \int_{-\infty}^{+\infty} f_{X_1}(s)\,\mathrm{d}s = \int_{-\infty}^{+\infty} \frac{c_1}{1+s^2}\,\mathrm{d}s = c_1\pi.$$

Donc $c_1 = 1/\pi$ et comme $c_1 c_2 = 1$, $c_2 = \pi$. Les densités marginales sont donc finalement :

$$f_{X_1} : s \longmapsto \frac{1}{\pi(1+s^2)}, \quad f_{X_2} : t \longmapsto \frac{1}{\sqrt{2\pi t}}\mathrm{e}^{-t/2}\mathbf{1}_{]0,+\infty[}(t).$$

On voit que X_1 suit la loi de Cauchy standard $\mathrm{Cau}(0,1)$. Quant à la loi de X_2, c'est une loi $\chi^2(1)$, c.-à-d. la loi de Z^2 où Z est gaussienne $\mathfrak{N}(0,1)$. On vous laisse le soin de vérifier cette dernière affirmation en calculant la loi de Z^2. Ceci vous permettra de vérifier *a posteriori* que f était bien une densité de probabilité sur $\mathbb{R}^2$ grâce à la proposition 8.43 a).

De même que la seule connaissance des lois des v.a. X et Y ne suffit pas en général pour déterminer la loi du vecteur aléatoire (X, Y), elle ne suffit pas non plus pour connaître la loi de la v.a. $X + Y$. Voici un exemple presque trivial pour s'en convaincre. On prend d'abord X de loi $\mathrm{Bern}(1/2)$, donc $P(X = 0) = P(X = 1) = 1/2$ et $Y := 1 - X$. Alors Y suit aussi la loi $\mathrm{Bern}(1/2)$ car $P(Y = 0) = P(X = 1) = 1/2$ et $P(Y = 1) = P(X = 0) = 1/2$. La somme $X + Y$ est la variable constante 1, sa loi est donc la masse de Dirac δ_1. Prenons maintenant $Y' := X$, alors Y' suit encore la loi $\mathrm{Bern}(1/2)$ et $X + Y' = 2X$ a pour loi $\frac{1}{2}\delta_0 + \frac{1}{2}\delta_2$.

10. Pour les lecteurs n'ayant pas sauté la démonstration informelle de l'hérédité de l'indépendance, signalons que les composantes de X sont indépendantes si et seulement si $P_X = P_{X_1} \otimes \cdots \otimes P_{X_d}$, probabilité produit des lois marginales.

Supposons maintenant que les variables aléatoires réelles X et Y sont indépendantes. Comme $X+Y = s(X,Y)$, où s est la fonction borélienne $s : \mathbb{R}^2 \to \mathbb{R}$, $(x,y) \mapsto x+y$, on devrait pouvoir déterminer la loi de $X+Y$, à partir des lois de X et de Y puisque l'indépendance permet de reconstruire la loi de (X,Y) à partir de ses lois marginales. Nous allons traiter ce problème dans les deux cas particuliers où X et Y sont toutes deux discrètes ou toutes deux à densité. On pourra voir en exercice un autre cas particulier, X discrète et Y à densité.

Proposition 8.47 (loi de la somme de deux v.a. discrètes)**.** *Si X et Y sont deux v.a. discrètes définies sur le même espace probabilisé $(\Omega, \mathcal{F}, P)$, la loi de la v.a. discrète $Z := X+Y$ est donnée par :*

$$\forall z \in Z(\Omega), \quad P(Z=z) = \sum_{x \in X(\Omega)} P(X=x, Y=z-x) \tag{8.36}$$

$$= \sum_{y \in Y(\Omega)} P(X=z-y, Y=y). \tag{8.37}$$

Si de plus X et Y sont indépendantes, P_Z peut se calculer explicitement à partir des lois P_X et P_Y par les formules :

$$\forall z \in Z(\Omega), \quad P(Z=z) = \sum_{x \in X(\Omega)} P(X=x)P(Y=z-x) \tag{8.38}$$

$$= \sum_{y \in Y(\Omega)} P(X=z-y)P(Y=y). \tag{8.39}$$

Preuve. Vérifions d'abord que Z est discrète, c'est-à-dire que $Z(\Omega)$ est au plus dénombrable. Pour cela, on note que l'ensemble $A := \big\{ \big(X(\omega), Y(\omega)\big) \ ; \ \omega \in \Omega \big\}$ est au plus dénombrable parce qu'inclus dans $X(\Omega) \times Y(\Omega)$, produit cartésien de deux ensembles au plus dénombrables. L'application $s : A \to Z(\Omega)$, $(x,y) \mapsto x+y$ est *surjective*, donc $Z(\Omega)$ est au plus dénombrable, cf. proposition 1.41. Nous reviendrons ci-dessous (remarque 8.48) sur la description de $Z(\Omega)$ que nous n'avons pas besoin d'expliciter davantage à ce stade.

Justifions (8.36). On commence par découper Ω suivant les valeurs de X en union au plus dénombrable d'évènements deux à deux disjoints :

$$\Omega = \bigcup_{x \in X(\Omega)} \{X=x\}.$$

On en déduit que pour tout $z \in Z(\Omega)$,

$$\{Z=z\} = \Big(\bigcup_{x \in X(\Omega)} \{X=x\} \Big) \cap \{Z=z\} = \bigcup_{x \in X(\Omega)} \big(\{X=x\} \cap \{Z=z\} \big)$$

$$= \bigcup_{x \in X(\Omega)} \{X=x, Y=z-x\}.$$

Cette dernière union est au plus dénombrable et les évènements concernés deux à deux disjoints (certains pouvant être vides). Par σ-additivité de P on en déduit (8.36). Cette

formule montre qu'il est toujours possible de calculer la loi de $X+Y$ si l'on connaît la loi $P_{(X,Y)}$ du couple (X,Y). Si de plus X et Y sont indépendantes, alors

$$P(X=x, Y=z-x) = P(X=x)P(Y=z-x),$$

d'où (8.38).

Les formules (8.37) et (8.39) s'obtiennent en échangeant les rôles joués par X et Y dans la preuve ci-dessus. □

Remarque 8.48. Si l'on connaît seulement la loi de X et celle de Y, on ne peut même pas déterminer $Z(\Omega)$. Pour vous en convaincre, revoyez le contre exemple avec les lois Bern(1/2) donné ci-dessus où $Z(\Omega)$ peut être selon le cas $\{1\}$ ou $\{0,2\}$. D'une manière générale, il est clair que l'on a toujours

$$Z(\Omega) \subset \{x+y\ ;\ x \in X(\Omega),\ y \in Y(\Omega)\}, \tag{8.40}$$

l'inclusion pouvant être stricte. Cette inclusion est une égalité dans le cas où X et Y sont indépendantes *et* vérifient la condition supplémentaire suivante

$$\forall x \in X(\Omega), \quad P(X=x) \neq 0, \quad \forall y \in Y(\Omega), \quad P(Y=y) \neq 0. \tag{8.41}$$

Pour le voir, montrons que si (x,y) est un couple quelconque de $X(\Omega) \times Y(\Omega)$, il existe au moins un $\omega \in \Omega$ tel que $X(\omega)=x$ et $Y(\omega)=y$, donc pour cet ω, $Z(\omega) = X(\omega)+Y(\omega) = x+y$. En effet par indépendance de X et Y et la condition (8.41),

$$P(X=x, Y=y) = P(X=x)P(Y=y) \neq 0,$$

donc l'évènement $\{\omega \in \Omega;\ X(\omega)=x, Y(\omega)=y\}$ n'est pas vide.

Exemple 8.49 (somme de deux v.a. de Poisson indépendantes)**.** Si X et Y sont indépendantes et suivent la loi de Poisson de paramètre α et β respectivement, $Z := X+Y$ suit la loi de Poisson de paramètre $\gamma = \alpha+\beta$.

Vérification. Ici $X(\Omega) = Y(\Omega) = \mathbb{N}$ et

$$\forall i \in \mathbb{N}, \quad P(X=i) = \frac{\mathrm{e}^{-\alpha}\alpha^i}{i!}, \quad \forall j \in \mathbb{N}, \quad P(Y=j) = \frac{\mathrm{e}^{-\beta}\beta^j}{j!},$$

donc (8.41) est vérifiée et $Z(\Omega) = \{k = i+j;\ i \in \mathbb{N},\ j \in \mathbb{N}\} = \mathbb{N}$. Appliquons maintenant (8.38) en notant que pour $i > k$, $k-i \notin Y(\Omega)$, donc $P(Y=k-i)=0$:

$$\begin{aligned}
\forall k \in \mathbb{N}, \quad P(Z=k) = \sum_{i\in\mathbb{N}} P(X=i)P(Y=k-i) &= \sum_{i=0}^{k} \frac{\mathrm{e}^{-\alpha}\alpha^i}{i!}\frac{\mathrm{e}^{-\beta}\beta^{k-i}}{(k-i)!} \\
&= \mathrm{e}^{-(\alpha+\beta)}\frac{1}{k!}\sum_{i=0}^{k}\frac{k!}{i!(k-i)!}\alpha^i\beta^{k-i} \\
&= \frac{\mathrm{e}^{-(\alpha+\beta)}(\alpha+\beta)^k}{k!},
\end{aligned}$$

ce qui montre que Z suit la loi Pois(γ) pour $\gamma = \alpha+\beta$. □

Proposition 8.50 (somme de deux v.a. à densité indépendantes)**.** *Si les variables aléatoires réelles indépendantes X et Y admettent pour densités respectives f et g, leur somme $S = X + Y$ admet pour densité le produit de convolution $f * g$ défini sur $\mathbb{R}$ par*

$$(f * g)(s) = \int_{-\infty}^{+\infty} f(s-t)g(t)\,\mathrm{d}t = \int_{-\infty}^{+\infty} f(t)g(s-t)\,\mathrm{d}t. \tag{8.42}$$

Preuve. Notons d'abord que les deux intégrales généralisées figurant dans (8.42) existent toujours comme éléments de $\overline{\mathbb{R}}_+$ et que l'on passe de la première à la deuxième par le changement de variable $t \mapsto s - t$ (pour s fixé).

Calculons $\mathbf{E}h(S)$ pour h continue bornée quelconque à l'aide de la loi du couple (X, Y). On sait par la proposition 8.43 a) que cette loi $P_{(X,Y)}$ admet pour densité la fonction $p = f \otimes g$. En utilisant la proposition 8.17 appliquée à la fonction composée $(x, y) \mapsto h(x + y)$, on obtient :

$$\mathbf{E}h(S) = \mathbf{E}h(X+Y) = \int_{\mathbb{R}^2} h(x+y)p(x,y)\,\mathrm{d}x\,\mathrm{d}y = \int_{\mathbb{R}^2} h(x+y)f(x)g(y)\,\mathrm{d}x\,\mathrm{d}y.$$

En effectuant dans cette dernière intégrale double le changement de variable linéaire :

$$(s,t) = (x+y, y), \quad \text{d'inverse} \quad (x,y) = (s-t, t),$$

on obtient (noter que le déterminant du changement de variable vaut ici 1) :

$$\mathbf{E}h(S) = \int_{\mathbb{R}^2} h(s)f(s-t)g(t)\,\mathrm{d}s\,\mathrm{d}t = \int_{-\infty}^{+\infty} h(s)\left\{\int_{-\infty}^{+\infty} f(s-t)g(t)\,\mathrm{d}t\right\}\mathrm{d}s.$$

Ainsi pour toute $h : \mathbb{R} \to \mathbb{R}$ continue bornée sur $\mathbb{R}$, on a

$$\mathbf{E}h(S) = \int_{-\infty}^{+\infty} h(s)(f * g)(s)\,\mathrm{d}s. \tag{8.43}$$

En particulier pour $h = 1$ constante sur $\mathbb{R}$, (8.43) s'écrit $1 = \int_{-\infty}^{+\infty}(f * g)(s)\,\mathrm{d}s$, ce qui nous montre [11] que $f * g$ est une densité de probabilité sur $\mathbb{R}$. Soit Z une variable aléatoire de densité $f * g$. Alors par la proposition 7.42,

$$\mathbf{E}h(Z) = \int_{-\infty}^{+\infty} h(s)(f * g)(s)\,\mathrm{d}s = \mathbf{E}h(S)$$

pour toute h continue bornée sur $\mathbb{R}$. Par la proposition 8.16 appliquée en dimension 1, on en déduit que S et Z ont même loi, donc que S admet pour densité $f * g$. □

Exemple 8.51 (somme de deux v.a. exponentielles indépendantes)**.** Soient X et Y deux variables aléatoires indépendantes suivant des lois exponentielles de paramètres

11. En toute rigueur, il faudrait aussi vérifier que $f * g$ est bien Riemann intégrable sur tout intervalle fermé borné de son ensemble de définition c.-à-d. de l'ensemble $\{s \in \mathbb{R}\ ;\ (f * g)(s) < +\infty\}$.

respectifs a et b. Calculons la densité $h = f * g$ de la loi de $X + Y$, avec ici $f(t) = a\,\mathrm{e}^{-at}\mathbf{1}_{\mathbb{R}_+}(t)$ et $g(t) = b\,\mathrm{e}^{-bt}\mathbf{1}_{\mathbb{R}_+}(t)$.

$$\begin{aligned} h(s) &= \int_{-\infty}^{+\infty} a\mathrm{e}^{-a(s-t)} b\mathrm{e}^{-bt}\mathbf{1}_{\mathbb{R}_+}(s-t)\mathbf{1}_{\mathbb{R}_+}(t)\,\mathrm{d}t \\ &= ab\mathrm{e}^{-as}\int_{-\infty}^{+\infty} \mathrm{e}^{(a-b)t}\mathbf{1}_{\mathbb{R}_+}(s-t)\mathbf{1}_{\mathbb{R}_+}(t)\,\mathrm{d}t. \end{aligned}$$

On peut réécrire le produit d'indicatrices comme suit :

$$\mathbf{1}_{\mathbb{R}_+}(s-t)\mathbf{1}_{\mathbb{R}_+}(t) = \begin{cases} 1 & \text{si } t \geqslant 0 \text{ et } s-t \geqslant 0 \\ 0 & \text{sinon} \end{cases} = \begin{cases} \mathbf{1}_{[0,s]}(t) & \text{si } s \geqslant 0 \\ 0 & \text{sinon} \end{cases} = \mathbf{1}_{\mathbb{R}_+}(s)\mathbf{1}_{[0,s]}(t).$$

En reportant ceci dans l'intégrale ci-dessus, il vient

$$h(s) = ab\,\mathrm{e}^{-as}\mathbf{1}_{\mathbb{R}_+}(s)\int_{-\infty}^{+\infty} \mathrm{e}^{(a-b)t}\mathbf{1}_{[0,s]}(t)\,\mathrm{d}t = ab\,\mathrm{e}^{-as}\mathbf{1}_{\mathbb{R}_+}(s)\int_0^s \mathrm{e}^{(a-b)t}\,\mathrm{d}t.$$

Cette dernière intégrale est une intégrale de Riemann ordinaire qui se calcule par primitivation en distinguant les cas $a \neq b$ et $a = b$.

Si $a \neq b$, on obtient ainsi

$$\begin{aligned} h(s) = ab\mathrm{e}^{-as}\mathbf{1}_{\mathbb{R}_+}(s)\left[\frac{\mathrm{e}^{(a-b)t}}{a-b}\right]_0^s &= \frac{ab\,\mathrm{e}^{-as}}{a-b}\big(\mathrm{e}^{(a-b)\,s} - 1\big)\mathbf{1}_{\mathbb{R}_+}(s) \\ &= \frac{ab}{a-b}\big(\mathrm{e}^{-bs} - \mathrm{e}^{-as}\big)\mathbf{1}_{\mathbb{R}_+}(s). \end{aligned}$$

À titre de précaution, on peut vérifier que cette densité h est bien positive en notant que pour $s \geqslant 0$, $\big(\mathrm{e}^{-bs} - \mathrm{e}^{-as}\big)/(b-a)$ est le taux d'accroissement entre a et b de la fonction *décroissante* $u \mapsto \mathrm{e}^{-su}$, donc que ce quotient est négatif.

Si $a = b$, on obtient

$$h(s) = a^2\,\mathrm{e}^{-as}\mathbf{1}_{\mathbb{R}_+}(s)\int_0^s \mathrm{d}t = a^2 s\,\mathrm{e}^{-as}\mathbf{1}_{\mathbb{R}_+}(s).$$

8.2.3 Indépendance et espérance de produits

Une propriété essentielle de l'indépendance des suites de variables aléatoires est que l'espérance d'un produit est égale au produit des espérances. La formulation précise, d'une portée encore plus générale, est la suivante.

Théorème 8.52. *Pour $i \in [\![1, n]\!]$, soient $X_i : \Omega \to \mathbb{R}^{d_i}$ des vecteurs aléatoires indépendants et $h_i : \mathbb{R}^{d_i} \to \mathbb{R}$, des fonctions boréliennes telles que les variables aléatoires réelles $h_i(X_i)$ soient intégrables. Alors le produit $h_1(X_1)\dots h_n(X_n)$ est intégrable et*

$$\mathbf{E}\big(h_1(X_1)\dots h_n(X_n)\big) = \prod_{i=1}^n \mathbf{E}h_i(X_i). \tag{8.44}$$

En particulier si pour tout i, $d_i = 1$ et h_i est l'identité sur $\mathbb{R}$, on obtient l'interversion espérance produit.

Corollaire 8.53. *Si $X_1, \ldots, X_n$ sont des variables aléatoires réelles indépendantes et intégrables, leur produit $X_1 \ldots X_n$ est intégrable et*

$$\mathbf{E}(X_1 \ldots X_n) = (\mathbf{E}X_1) \ldots (\mathbf{E}X_n). \tag{8.45}$$

Preuve. Par hérédité de l'indépendance, plus précisément par la version de la proposition 8.37 où la suite de départ $X_1, \ldots, X_n$ est une famille hétéroclite de vecteurs aléatoires indépendants, le problème se réduit à la preuve de l'intégrabilité de $Y = Y_1 \ldots Y_n$ et de l'égalité

$$\mathbf{E}(Y_1 \ldots Y_n) = (\mathbf{E}Y_1) \ldots (\mathbf{E}Y_n), \tag{8.46}$$

pour *toute suite finie* $Y_1, \ldots, Y_n$ ($n \geqslant 2$) de v.a. intégrables et indépendantes.

Bien qu'il soit possible de traiter le cas n quelconque d'un coup, nous allons réduire le problème au cas $n = 2$, pour alléger les notations. Admettons donc pour un instant que nous ayons établi (8.46) dans le cas $n = 2$. On fait alors une récurrence finie sur $2 \leqslant k < n$ en prenant pour hypothèse que (8.46) est vérifiée pour les produits de k facteurs. Alors l'indépendance des Y_i ($1 \leqslant i \leqslant n$) implique celle de $(Y_1, \ldots, Y_k)$ et de Y_{k+1}, par hérédité pour les blocs disjoints cf. corollaire 8.38 et remarque 8.35-1. Par l'hypothèse de récurrence, $Y_1 \ldots Y_k$ est intégrable et comme Y_{k+1} l'est aussi, leur produit l'est encore par le cas $n = 2$. On a donc

$$\begin{aligned}
\mathbf{E}\big(Y_1 \ldots Y_k Y_{k+1}\big) &= \mathbf{E}\big((Y_1 \ldots Y_k)Y_{k+1}\big) \\
&= \big(\mathbf{E}(Y_1 \ldots Y_k)\big)\big(\mathbf{E}Y_{k+1}\big) \quad \text{par le cas } n = 2 \\
&= (\mathbf{E}Y_1) \ldots (\mathbf{E}Y_k)(\mathbf{E}Y_{k+1}) \quad \text{par l'hypothèse de récurrence.}
\end{aligned}$$

Ainsi la validité de (8.46) au rang k implique sa validité au rang $k+1$, ce qui achève la partie « hérédité » de notre récurrence finie.

Il nous reste maintenant à prouver que pour toutes v.a. réelles Y_1 et Y_2 indépendantes et intégrables,

$$Y_1Y_2 \text{ est intégrable et } \quad \mathbf{E}(Y_1Y_2) = (\mathbf{E}Y_1)(\mathbf{E}Y_2). \tag{8.47}$$

Nous allons traiter successivement les cas suivants.

1. $Y_i = \mathbf{1}_{A_i}$, $i = 1, 2$ où les $A_i \in \mathcal{F}$ sont des évènements ;
2. Y_1 et Y_2 variables aléatoires simples ;
3. Y_1 et Y_2 variables aléatoires positives ;
4. Y_1 et Y_2 variables aléatoires réelles intégrables.

Cas 1. Si $Y_i = \mathbf{1}_{A_i}$ pour un $A_i \in \mathcal{F}$, les v.a. Y_i et leur produit sont intégrables car bornées. L'indépendance des v.a. Y_1 et Y_2 implique [12] celle des *évènements* A_1 et A_2 car $A_i = \big\{\mathbf{1}_{A_i} \in \{1\}\big\}$. D'autre part, on vérifie facilement (faites-le) que

$$\mathbf{1}_{A_1}\mathbf{1}_{A_2} = \mathbf{1}_{A_1 \cap A_2}.$$

12. En fait *équivaut* à ...(exercice).

Grâce à la formule $\mathbf{E}(\mathbf{1}_A) = P(A)$ et à l'indépendance de A_1 et A_2, on en déduit

$$\mathbf{E}(Y_1Y_2) = \mathbf{E}\big(\mathbf{1}_{A_1\cap A_2}\big) = P(A_1\cap A_2) = P(A_1)P(A_2) = \mathbf{E}\big(\mathbf{1}_{A_1}\big)\mathbf{E}\big(\mathbf{1}_{A_2}\big) = (\mathbf{E}Y_1)(\mathbf{E}Y_2),$$

ce qui achève la vérification de (8.47) dans le cas des variables aléatoires indicatrices. □

Cas 2. Supposons maintenant que Y_1 et Y_2 sont deux variables aléatoires simples indépendantes. Elles sont alors bornées donc intégrables ainsi que leur produit. Notons $Y_1(\Omega) = \{y_{1,1}, \dots, y_{1,l}\}$, les $y_{1,j}$ étant tous distincts et faisons de même pour $Y_2(\Omega) = \{y_{2,1}, \dots, y_{2,m}\}$. On a alors les décompositions

$$Y_1 = \sum_{j=1}^{l} y_{1,j}\mathbf{1}_{A_{1,j}}, \quad Y_2 = \sum_{k=1}^{m} y_{2,k}\mathbf{1}_{A_{2,k}},$$

avec $A_{1,j} = \{Y_1 = y_{1,j}\} = \{Y_1 \in \{y_{1,j}\}\}$ et $A_{2,k} = \{Y_2 = y_{2,k}\} = \{Y_2 \in \{y_{2,k}\}\}$. On en déduit que les deux évènements $A_{1,j}$ et $A_{2,k}$ héritent de l'indépendance de Y_1 et Y_2. En utilisant la linéarité de l'espérance, cette indépendance d'évènements et le cas 1, on obtient :

$$\begin{aligned}
\mathbf{E}(Y_1Y_2) = \mathbf{E}\bigg(\sum_{j=1}^{l}\sum_{k=1}^{m} y_{1,j}y_{2,k}\mathbf{1}_{A_{1,j}}\mathbf{1}_{A_{2,k}}\bigg) &= \sum_{j=1}^{l}\sum_{k=1}^{m} y_{1,j}y_{2,k}\mathbf{E}\big(\mathbf{1}_{A_{1,j}}\mathbf{1}_{A_{2,k}}\big) \\
&= \sum_{j=1}^{l}\sum_{k=1}^{m} y_{1,j}y_{2,k}\mathbf{E}\big(\mathbf{1}_{A_{1,j}}\big)\mathbf{E}\big(\mathbf{1}_{A_{2,k}}\big) \\
&= \bigg(\sum_{j=1}^{l} y_{1,j}\mathbf{E}\big(\mathbf{1}_{A_{1,j}}\big)\bigg)\bigg(\sum_{k=1}^{m} y_{2,k}\mathbf{E}\big(\mathbf{1}_{A_{2,k}}\big)\bigg) \\
&= \mathbf{E}\bigg(\sum_{j=1}^{l} y_{1,j}\mathbf{1}_{A_{1,j}}\bigg)\mathbf{E}\bigg(\sum_{k=1}^{m} y_{2,k}\mathbf{1}_{A_{2,k}}\bigg) \\
&= \mathbf{E}Y_1\,\mathbf{E}Y_2.
\end{aligned}$$

Ainsi (8.47) est vérifiée dans le cas des variables aléatoires simples. □

Cas 3. Soient Y_1 et Y_2 deux variables aléatoires positives indépendantes (on n'a pas besoin de les supposer intégrables ici). On sait (cf. théorème 7.21) que la v.a. positive Y_i est limite d'une suite croissante de v.a. positives simples : $Y_{i,n} \uparrow Y_i$, $i = 1, 2$. Plus précisément, on peut prendre (revoyez la preuve du théorème 7.21 et vérifiez la formule proposée ci-après)

$$Y_{i,n} = h_n(Y_i), \quad h_n : \mathbb{R}_+ \to \mathbb{R}_+,\ x \longmapsto \min\big(n, 2^{-n}[2^n x]\big).$$

La fonction h_n est croissante donc borélienne. La proposition 8.37 nous dit alors que pour chaque n, les v.a. $Y_{1,n}$ et $Y_{2,n}$ héritent de l'indépendance de Y_1 et Y_2. Comme ce sont des v.a. simples, le cas 2 nous donne :

$$\forall n \in \mathbb{N}^*, \quad \mathbf{E}\big(Y_{1,n}Y_{2,n}\big) = \big(\mathbf{E}Y_{1,n}\big)\big(\mathbf{E}Y_{2,n}\big). \tag{8.48}$$

En notant que $Y_{1,n}Y_{2,n}$ converge en croissant vers Y_1Y_2, un passage à la limite dans (8.48) et une triple application du théorème de Beppo Levi nous donnent $\mathbf{E}(Y_1Y_2) = \mathbf{E}Y_1\,\mathbf{E}Y_2$. □

Cas 4. Si Y_1 et Y_2 sont deux variables aléatoires réelles indépendantes et intégrables, $|Y_1|$ et $|Y_2|$ sont des v.a. positives indépendantes et par le cas 3,

$$\mathbf{E}|Y_1Y_2| = \mathbf{E}(|Y_1||Y_2|) = (\mathbf{E}|Y_1|)(\mathbf{E}|Y_2|) < +\infty,$$

ce qui établit l'intégrabilité de Y_1Y_2. Cette intégrabilité implique aussi celle des produits de v.a. positives $Y_1^+Y_2^+$, $Y_1^+Y_2^-$, $Y_1^-Y_2^+$ et $Y_1^-Y_2^-$. D'autre part les facteurs de chacun de ces produits héritent leur indépendance de celle de Y_1 et Y_2, puisque $Y_i^+ = \max(Y_i, 0)$ et $Y_i^- = \max(-Y_i, 0)$. En utilisant le cas 3 on obtient alors (notez que toutes les espérances intervenant dans ce calcul sont finies) :

$$\begin{aligned}
\mathbf{E}(Y_1Y_2) &= \mathbf{E}\big((Y_1^+ - Y_1^-)(Y_2^+ - Y_2^-)\big) \\
&= \mathbf{E}(Y_1^+Y_2^+) - \mathbf{E}(Y_1^+Y_2^-) - \mathbf{E}(Y_1^-Y_2^+) + \mathbf{E}(Y_1^-Y_2^-) \\
&= \mathbf{E}Y_1^+\mathbf{E}Y_2^+ - \mathbf{E}Y_1^+\mathbf{E}Y_2^- - \mathbf{E}Y_1^-\mathbf{E}Y_2^+ + \mathbf{E}Y_1^-\mathbf{E}Y_2^- \\
&= \big(\mathbf{E}Y_1^+ - \mathbf{E}Y_1^-\big)\big(\mathbf{E}Y_2^+ - \mathbf{E}Y_2^-\big) \\
&= (\mathbf{E}Y_1)(\mathbf{E}Y_2),
\end{aligned}$$

ce qui achève la vérification de (8.47). □

La preuve du théorème 8.52 est maintenant complète. □

Comme sous-produit de la démonstration ci-dessus et notamment du cas 3, on a le résultat suivant.

Proposition 8.54. *Le théorème 8.52 s'applique sans condition d'intégrabilité des $Y_i = h_i(X_i)$ si les fonctions boréliennes h_i sont positives. Le corollaire 8.53 est vrai sans condition d'intégrabilité pour des variables aléatoires positives.*

Une application importante du théorème 8.52 est que l'indépendance de deux v.a. implique la nullité de leur covariance lorsqu'elle existe.

Proposition 8.55 (indépendance et covariance)**.**

a) Si X et Y sont des v.a. réelles de carré intégrable et indépendantes, leur covariance est nulle.

b) Si $X_1, \ldots, X_n$ sont des v.a. réelles de carré intégrable et deux à deux indépendantes, en notant $S_n := \sum_{k=1}^n X_k$, on a

$$\operatorname{Var} S_n = \sum_{k=1}^n \operatorname{Var} X_k. \tag{8.49}$$

Preuve. Le a) est une conséquence immédiate du corollaire 8.53 et de la formule de Koenig puisque $\mathrm{Cov}(X,Y) = \mathbf{E}(XY) - (\mathbf{E}X)(\mathbf{E}Y) = (\mathbf{E}X)(\mathbf{E}Y) - (\mathbf{E}X)(\mathbf{E}Y)$. On peut le voir aussi à partir de la définition de la covariance puisque $X - \mathbf{E}X$ et $Y - \mathbf{E}Y$ héritent de l'indépendance de X et Y et sont deux v.a. d'espérance nulle. Notons aussi que pour des v.a. indépendantes, on n'a pas besoin de supposer que X et Y soient de *carré intégrable* pour définir leur covariance, leur intégrabilité suffit.

Le b) est une conséquence immédiate du a) *via* la formule 8.9. Notons qu'il suffit d'avoir ici l'indépendance deux à deux, plus faible que l'indépendance mutuelle [13]. □

Remarque 8.56. La nullité de la covariance de deux v.a. réelles X et Y *n'implique pas* leur indépendance. Voici un contre exemple. Prenons X de loi uniforme sur $[-1,+1]$ et $Y := X^2$. On a alors en notant que X a pour densité $\frac{1}{2}\mathbf{1}_{[-1,1]}$,

$$\mathbf{E}X = \frac{1}{2}\int_{-1}^{+1} x\,\mathrm{d}x = 0 \quad \text{et} \quad \mathbf{E}(XY) = \mathbf{E}(X^3) = \frac{1}{2}\int_{-1}^{+1} x^3\,\mathrm{d}x = 0,$$

d'où $\mathrm{Cov}(X,Y) = \mathbf{E}(XY) - (\mathbf{E}X)(\mathbf{E}Y) = 0$. Il est clair intuitivement que X et Y ne sont pas indépendantes puisque Y est une fonction déterministe de X. Pour vérifier cette non-indépendance par le calcul, on peut remarquer que d'une part :

$$\begin{aligned} P\big(X \in [0,1/2] \text{ et } Y \in [0,1/4]\big) &= P\big(X \in [0,1/2] \text{ et } X \in [-1/2,1/2]\big) \\ &= P\big(X \in [0,1/2]\big) = \frac{1}{4} \end{aligned}$$

et d'autre part :

$$P\big(X \in [0,1/2]\big)P\big(Y \in [0,1/4]\big) = \frac{1}{4}P\big(X \in [-1/2,1/2]\big) = \frac{1}{8}.$$

8.3 Exercices

Ex. 8.1 On suppose que le couple de variables aléatoires discrètes (X,Y) a une loi donnée par :

$$\forall (i,j) \in \mathbb{N}^2, \quad P(X=i, Y=j) = \frac{\alpha}{(1+i+j)!},$$

où α est une constante strictement positive qui sera précisée ultérieurement.

1) Expliquer sans calcul pourquoi les marginales X et Y ont même loi.
2) On pose $S = X + Y$. Montrer que :

$$\forall k \in \mathbb{N}, \quad P(S=k) = \frac{\alpha}{k!}.$$

13. En fait il suffit d'avoir la non-corrélation deux à deux, c'est-à-dire $\mathrm{Cov}(X_i, X_j) = 0$ pour $i \neq j$, ce qui est encore plus faible.

3) En déduire la valeur de α et reconnaître la loi de S.

4) Calculer $P(X = 0)$. Les variables aléatoires X et Y sont-elles indépendantes ?

5) Donner sans calcul l'espérance de S et en déduire la valeur de $\mathbf{E}X$. La même méthode peut-elle servir pour obtenir $\operatorname{Var} X$?

6) Calculer $P(X = Y)$ et en déduire sans calcul $P(X > Y)$.

Ex. 8.2 *Usagers et usagères à la poste*

Le nombre de personnes entrant dans un bureau de poste un jour donné est une variable de Poisson S de paramètre α. Chaque nouvel arrivant est un arrivant avec probabilité p et une arrivante avec probabilité q. On se propose de montrer que le nombre X d'hommes et le nombre Y de femmes entrant dans le bureau de poste au cours d'une journée sont deux variables de Poisson indépendantes de paramètres respectifs $p\alpha$ et $q\alpha$.

1) Calculer $P(X = i, Y = j \mid S = n)$ pour tout entier n et tout couple (i, j) tel que $i + j = n$.

2) Quelle est la loi du vecteur (X, Y) ?

3) En déduire les lois de X et Y.

Solution p. 484

Ex. 8.3 *Lois de Poisson multidimensionnelles*

Soit $\alpha = (\alpha_1, \ldots, \alpha_d)$ un vecteur de $\mathbb{R}^d$ à composantes toutes strictement positives. On dit que $X = (X_1, \ldots, X_d)$ suit la loi de Poisson de paramètre α si

$$\forall (k_1, \ldots, k_d) \in \mathbb{N}^d, \quad P(X_1 = k_1, \ldots, X_d = k_d) = \mathrm{e}^{-(\alpha_1 + \cdots + \alpha_d)} \frac{\alpha_1^{k_1} \ldots \alpha_d^{k_d}}{k_1! \ldots k_d!} \tag{8.50}$$

1) Vérifiez que (8.50) définit bien une loi de probabilité discrète sur $\mathbb{R}^d$.

2) Que peut-on dire des lois marginales de X ?

3) Soit I un sous-ensemble de $[\![1, d]\!]$ ayant au moins deux éléments. On pose $X_I = \sum_{i \in I} X_i$. Quelle est la loi de X_I ?

Ex. 8.4 On considère la fonction h définie sur $\mathbb{R}^2$ par :

$$h(x, y) = \frac{1}{4}(1 + \sin \pi x \sin \pi y) \mathbf{1}_{[-1,+1]^2}(x, y).$$

1) Vérifier que h est une densité de probabilité.

2) Soit (X, Y) un vecteur aléatoire dont la loi a pour densité h. Montrer que X et Y suivent la loi uniforme sur $[-1, +1]$. Calculer $\operatorname{Cov}(X, Y)$. Les variables aléatoires X et Y sont elles indépendantes ?

Solution p. 485

Ex. 8.5 Montrez que si le vecteur aléatoire $X = (X_1, \ldots, X_d)$ est à densité, toute combinaison linéaire $a_1 X_1 + \cdots + a_d X_d$ avec au moins un des a_i non nul est une variable aléatoire à densité.

Ex. 8.6 *Tir à l'arc* ↓ 9.10

Dans un stand de tir à l'arc, un archer vise une cible placée sur une grande palissade (assimilée à un plan vertical). On munit ce plan d'un repère orthonormé de centre celui de la cible, l'axe des abscisses étant horizontal et celui des ordonnées vertical. On modélise le point d'impact de la flèche par le vecteur aléatoire gaussien (X, Y) de densité :

$$f : (x, y) \longmapsto \frac{1}{2\pi\sigma^2} \exp\left(\frac{-(x^2+y^2)}{2\sigma^2}\right).$$

Intuitivement le paramètre σ mesure l'adresse du tireur. Pour ε positif fixé, plus σ est petit, plus grande sera la probabilité que la flèche atteigne un point à distance inférieure à ε du centre.

1) Montrer sans calcul que X et Y sont indépendantes et de même loi que l'on précisera.

2) On se propose dans cette question de déterminer la loi du vecteur image de (X, Y) par un passage en coordonnées polaires. On note $\Delta := \mathbb{R}_- \times \{0\}$ et $\varphi : \mathbb{R}^2 \backslash \Delta \to]0, +\infty[\times]-\pi, \pi[$ le passage des coordonnées cartésiennes aux coordonées polaires pour le plan privé de la demi-droite Δ. C'est un C^1-difféomorphisme dont l'inverse φ^{-1} est donné par :

$$\forall (r, u) \in]0, +\infty[\times]-\pi, \pi[, \quad (x, y) = \varphi^{-1}(r, u) = (r\cos u, r \sin u).$$

L'évènement $\Omega' := \big\{\omega \in \Omega;\ \big(X(\omega), Y(\omega)\big) \in \mathbb{R}^2 \backslash \Delta\big\}$ est de probabilité 1 puisque $P\big((X, Y) \in \Delta\big) = \int_\Delta f(x, y)\,\mathrm{d}x\,\mathrm{d}y = 0$. On pose pour tout $\omega \in \Omega'$, $\big(R(\omega), U(\omega)\big)\big) = \varphi\big(X(\omega), Y(\omega)\big)$ et on *admet* qu'il est possible de prolonger (R, U) à tout Ω de façon à en faire une application mesurable, donc un *vecteur aléatoire*. Comme $P(\Omega \backslash \Omega') = 0$, la façon dont on effectue ce prolongement, pourvu qu'il soit mesurable, n'influence pas la loi de (R, U). Déterminer la loi du vecteur aléatoire (R, U) par sa densité. En déduire que R et U sont indépendantes et donner la loi de R et celle de U.

Solution p. 486

Ex. 8.7 *Une inégalité de Markov à deux variables*

Dans tout l'exercice, X et Y sont deux variables aléatoires positives définies sur l'espace probabilisé $(\Omega, \mathcal{F}, P)$.

1) Montrez que

$$\forall s > 0,\ \forall t > 0, \qquad P(X > s \text{ et } Y > t) \leqslant \frac{1}{st}\mathbf{E}(XY). \tag{8.51}$$

Indication : on pourra utiliser après l'avoir soigneusement justifiée, l'inégalité

$$st\mathbf{1}_{\{X>s \text{ et } Y>t\}} \leqslant XY. \tag{8.52}$$

2) On suppose de plus que $c := \mathbf{E}(XY) < +\infty$. Pour quelles valeurs de (s, t), l'inégalité (8.51) a-t-elle un intérêt ? Représentez graphiquement l'ensemble de ces valeurs de (s, t).

3) On suppose maintenant que X et Y sont indépendantes. Expliquez comment on peut dans ce cas retrouver (8.51) en utilisant l'inégalité de Markov classique.

Ex. 8.8 *Covariances d'une loi multinomiale*[14]

On considère une suite de n épreuves répétées indépendantes, chaque épreuve ayant k résultats possibles $r_1, \dots, r_k$. On note p_i la probabilité de réalisation du résultat r_i lors d'une épreuve donnée. Par exemple si on lance n fois un dé équilibré, $k = 6$ et $p_i = 1/6$ pour $1 \leqslant i \leqslant 6$. Pour $1 \leqslant i \leqslant k$, notons X_i le nombre de réalisations du résultat r_i au cours des n épreuves.

1) Expliquer sans calcul pourquoi $\mathrm{Var}(X_1 + \cdots + X_k) = 0$.

2) Quelle est la loi de X_i ? Que vaut sa variance ?

3) Pour $i \neq j$, donner la loi et la variance de $X_i + X_j$.

4) En déduire $\mathrm{Cov}(X_i, X_j)$.

5) Contrôler ce résultat en développant $\mathrm{Var}(X_1 + \cdots + X_k)$ et en utilisant la première question.

Indications p. 487

Ex. 8.9 *Une somme mixte*

On définit sur le même espace probabilisé une variable aléatoire N à valeurs dans $\mathbb{N}$ et une variable aléatoire U de loi uniforme sur $[0, 1]$. On pose $X = N + U$.

1) Montrez que la fonction de répartition F de X est continue sur $\mathbb{R}$.

2) On suppose désormais que N et U sont *indépendantes*. Montrez que F est affine par morceaux. Donnez une méthode simple pour tracer le graphe de F à partir de celui de la f.d.r de N.

3) Montrez que X est à densité et donnez une expression de cette densité à l'aide des $p_k = P(N = k)$.

Ex. 8.10 *Produit d'une v.a. discrète par une uniforme*

Soient X et Y deux variables aléatoires définies sur le même espace probabilisé $(\Omega, \mathcal{F}, P)$ et *indépendantes* : X est une variable aléatoire discrète à valeurs dans $\mathbb{N}$ dont la loi est donnée par $P(X = k) = p_k$, $k \in \mathbb{N}$, et Y est de loi uniforme sur $[0, 1]$. Déterminer la fonction de répartition F de la variable aléatoire XY. Que peut on dire de F ? La loi de XY est-elle à densité ?

Ex. 8.11 *La méthode de Box Muller*

Il s'agit d'une méthode de simulation de variables gaussiennes $\mathfrak{N}(0, 1)$ à partir d'un générateur de variables de loi uniforme sur $]0, 1[$. Elle repose sur le résultat suivant. Si U_1 et U_2 sont indépendantes et de loi uniforme sur $]0, 1[$, les variables aléatoires

$$X := (-2 \ln U_1)^{1/2} \cos(2\pi U_2) \quad \text{et} \quad Y := (-2 \ln U_1)^{1/2} \sin(2\pi U_2)$$

sont indépendantes et de même loi $\mathfrak{N}(0, 1)$.

1) Quelle est la loi du vecteur (U_1, U_2) ?

2) En déduire la loi de (X, Y) par changement de variable et conclure.

Solution p. 487

14. Il n'est pas nécessaire de connaître la loi multinomiale pour pouvoir faire cet exercice.

Ex. 8.12 *Loi uniforme sur l'hypographe d'une densité*

Soit f une densité de probabilité sur $\mathbb{R}$ et G son hypographe :

$$G := \big\{(x,y) \in \mathbb{R}\times\mathbb{R}\ ;\ 0 \leqslant y \leqslant f(x)\big\}.$$

Soit $M = (Z,Y)$ un vecteur aléatoire de $\mathbb{R}\times\mathbb{R}$ de loi *uniforme* sur G. Vérifiez que la loi de la variable aléatoire Z a pour densité f.

Solution p. 488

Ex. 8.13 *Simulation d'une loi uniforme sur un hypographe*

Soit X une variable aléatoire réelle de densité g. Posons

$$M := \big(X, cg(X)U\big),$$

où U est une variable aléatoire de loi uniforme sur $[0,1]$, *indépendante* de X et $c > 0$ une constante. On note H l'hypographe de cg :

$$H := \big\{(x,y) \in \mathbb{R}\times\mathbb{R}\ ;\ 0 \leqslant y \leqslant cg(x)\big\}.$$

Montrez que M suit la loi uniforme sur H. *Indication.* On pourra utiliser la caractérisation de la loi du vecteur alétoire M par les moments fonctionnels $\mathbf{E}h(M)$ où $h : \mathbb{R}^2 \to \mathbb{R}_+$ est continue à support compact, cf. prop. 8.16.

Solution p. 489

Ex. 8.14 *Simulation d'une v.a. à densité par rejet* ↑ 8.12, ↑ 8.13

On voudrait simuler une variable aléatoire Z, de densité f. On suppose que l'on sait simuler X de densité g et trouver une constante c telle que $f \leqslant cg$ (nécessairement $c \geqslant 1$, pourquoi ?). Les X_i sont des v.a. indépendantes de même loi ayant pour densité g, les U_i des variables aléatoires indépendantes de même loi uniforme sur $[0,1]$ et les suites $(X_i)_{i\geqslant 1}$ et $(U_i)_{i\geqslant 1}$ sont indépendantes. On génère la suite des $M_i := \big(X_i, cg(X_i)U_i\big)$ en s'arrêtant au premier indice i_0 (aléatoire) tel que $cg(X_{i_0})U_{i_0} \leqslant f(X_{i_0})$. On pose alors $Z = X_{i_0}$ et Z a pour densité f. Justifiez cet algorithme.

Chapitre 9

Théorèmes limite

Nous commençons dans ce chapitre l'étude du comportement asymptotique de suites de variables aléatoires. Après avoir vu les différents modes de convergence de ces suites, nous abordons la loi des grands nombres. Ce résultat essentiel nous dit que les moyennes arithmétiques d'une suite de v.a. X_i indépendantes et de même loi ayant une espérance, convergent en un certain sens, vers cette espérance :

$$M_n := \frac{S_n}{n} = \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow[n\to+\infty]{} \mathbf{E}X_1. \tag{9.1}$$

Cette convergence est très utile en statistique pour *estimer* des paramètres d'une loi inconnue, sur la base de l'observation d'un *échantillon* $X_1, \ldots, X_n$ de grande taille. Le prolongement naturel de ce chapitre est l'étude du théorème limite central qui donne une sorte de vitesse de convergence pour la loi des grands nombres, permettant notamment de construire des « intervalles de confiance » pour l'estimation d'un paramètre.

9.1 Convergences de suites de variables aléatoires

9.1.1 Convergence presque sûre et en probabilité

Quand on envisage la question de la convergence d'une suite de v.a. (Y_n) vers une v.a. Y, la première notion de convergence qui vienne à l'esprit est la convergence simple sur tout Ω, au sens de l'analyse [1] :

$$\forall \omega \in \Omega, \quad Y_n(\omega) \xrightarrow[n\to+\infty]{} Y(\omega).$$

On voit immédiatement que cette notion n'est pas satisfaisante pour la convergence de la suite (M_n) donnée en (9.1). En effet considérons le modèle probabiliste infini

1. Ceci suppose que les Y_n et Y sont définies sur le même Ω.

le plus simple possible, à savoir le jeu de pile ou face infini. On peut prendre ici $\Omega = \{\mathsf{f},\mathsf{p}\}^{\mathbb{N}^*}$ et ω est une suite infinie $\omega = (u_i)_{i\geqslant 1}$, avec $u_i \in \{\mathsf{f},\mathsf{p}\}$ pour tout i. En prenant pour X_i l'indicatrice de l'évènement obtention de pile au i^{e} lancer, M_n est la fréquence d'apparition de pile au cours des n premiers lancers. Si la pièce est équilibrée, on s'attend à ce que M_n converge vers $1/2$. Or il est clair qu'il y a une infinité d'évènements élémentaires ω pour lesquels $M_n(\omega)$ ne converge pas vers $1/2$. On peut même construire facilement une infinité de ω pour lesquels $M_n(\omega)$ n'a *aucune* limite [2]. Ce simple exemple montre que la notion de convergence simple n'est pas pertinente en théorie des probabilités. Pour dépasser ce problème, on introduit la notion de convergence presque sûre, qui est la convergence simple sur un sous-ensemble Ω' de probabilité 1 de Ω.

Définition 9.1 (convergence presque sûre)**.** *Soient $(Y_n)_{n\geqslant 1}$ et Y des variables aléatoires définies sur le même espace probabilisé $(\Omega,\mathcal{F},P)$. Notons*

$$\Omega' := \big\{\omega\in\Omega\,;\ \lim_{n\to+\infty} Y_n(\omega) = Y(\omega)\big\}. \tag{9.2}$$

Si $P(\Omega')=1$, on dit qu'Y_n converge presque sûrement *vers Y, notation $Y_n \xrightarrow[n\to+\infty]{\text{p.s.}} Y$.*

Cette définition soulève immédiatement une question. Pour que l'égalité $P(\Omega')=1$ ait un sens, encore faut-il que Ω' appartienne à la tribu $\mathcal{F}$ sur laquelle est définie la fonction d'ensembles P. Pour établir l'appartenance à $\mathcal{F}$ de Ω', il est naturel de s'appuyer sur la mesurabilité des Y_n et de Y dont héritent les v.a. positives $|Y_n - Y|$. Pour cela, on commence par écrire avec des quantificateurs l'appartenance à Ω' :

$$\omega\in\Omega' \quad\Leftrightarrow\quad \forall\varepsilon>0,\ \exists j = j(\omega,\varepsilon)\in\mathbb{N},\ \forall k\geqslant j,\ |Y_k(\omega)-Y(\omega)|<\varepsilon. \tag{9.3}$$

En utilisant la traduction automatique des quantificateurs en opérations ensemblistes, cf. p. 14, on en déduit que

$$\Omega' = \bigcap_{\varepsilon>0}\ \bigcup_{j\in\mathbb{N}}\ \bigcap_{k\geqslant j}\ \{|Y_k - Y|<\varepsilon\}. \tag{9.4}$$

En lisant (9.4) de droite à gauche, on obtient les appartenances successives à $\mathcal{F}$, d'abord des ensembles $\{|Y_k-Y|<\varepsilon\}$ en raison de la mesurabilité des v.a. $|Y_n-Y|$, puis de l'intersection dénombrable sur k, puis de l'union dénombrable sur j. Arrivés là, on est coincé car la dernière intersection sur ε a pour ensemble d'indexation $]0,+\infty[$ qui est infini *non*-dénombrable. On ne peut donc pas en déduire l'appartenance à $\mathcal{F}$ de Ω'. Pour franchir cet obstacle, il suffit de revenir à (9.3) et d'écrire une version « discrétisée » de la convergence de $Y_n(\omega)$ vers $Y(\omega)$. Pour cela on choisit une suite de réels strictement positifs $(\varepsilon_i)_{i\geqslant 1}$, tendant vers 0 et on écrit que

$$\omega\in\Omega' \quad\Leftrightarrow\quad \forall i\geqslant 1,\ \exists j = j(\omega,i)\in\mathbb{N},\ \forall k\geqslant j,\ |Y_k(\omega)-Y(\omega)|<\varepsilon_i. \tag{9.5}$$

2. Pour approfondir cette question, voir la section « 6.5 Discussion » dans [14].

La traduction automatique des quantificateurs nous donne maintenant

$$\Omega' = \bigcap_{i\in\mathbb{N}^*} \bigcup_{j\in\mathbb{N}} \bigcap_{k\geqslant j} \{|Y_k - Y| < \varepsilon_i\}. \tag{9.6}$$

Sous cette forme, il est maintenant clair que Ω' appartient à $\mathcal{F}$ et ceci légitime la définition 9.1. Nous avons établi au passage que

$$Y_n \xrightarrow[n\to+\infty]{\text{p.s.}} Y \quad \Leftrightarrow \quad P\Big(\bigcap_{i\in\mathbb{N}^*} \bigcup_{j\in\mathbb{N}} \bigcap_{k\geqslant j} \{|Y_k - Y| < \varepsilon_i\}\Big) = 1, \tag{9.7}$$

pour une suite de réels $\varepsilon_i > 0$, tendant vers 0.

Pour l'instant, (9.7) n'est pas directement exploitable comme méthode pratique pour montrer une convergence presque sûre, mais on peut progresser dans cette direction en « faisant sortir le $\forall i$ » de la probabilité. Cette opération est légitimée par le lemme suivant.

Lemme 9.2. *Si $(B_i)_{i\geqslant 1}$ est une suite d'évènements, on a l'équivalence*

$$P\Big(\bigcap_{i\geqslant 1} B_i\Big) = 1 \quad \Leftrightarrow \quad \forall i \geqslant 1,\ P(B_i) = 1. \tag{9.8}$$

Preuve. L'implication « $\Rightarrow$ » est évidente car $\bigcap_{i\geqslant 1} B_i \subset B_j$ pour tout $j \geqslant 1$, d'où

$$1 = P\Big(\bigcap_{i\geqslant 1} B_i\Big) \leqslant P(B_j) \leqslant 1.$$

Pour la réciproque, on montre que si tous les $P(B_i)$ valent 1, alors $P\big((\bigcap_{i\geqslant 1} B_i)^c\big) = 0$. Ceci s'obtient par sous-σ-additivité de P :

$$P\Big(\Big(\bigcap_{i\geqslant 1} B_i\Big)^c\Big) = P\Big(\bigcup_{i\geqslant 1} B_i^c\Big) \leqslant \sum_{i=1}^{+\infty} P(B_i^c) = 0,$$

puisque tous les termes de cette série sont nuls. □

En appliquant le lemme 9.2 aux évènements $B_i := \bigcup_{j\in\mathbb{N}} \bigcap_{k\geqslant j} \{|Y_k - Y| < \varepsilon_i\}$, nous obtenons à partir de (9.7) *l'équivalence* entre la convergence presque sûre de Y_n vers Y et la condition

$$\forall i \in \mathbb{N}^*,\ P\Big(\bigcup_{j\in\mathbb{N}} \bigcap_{k\geqslant j} \{|Y_k - Y| < \varepsilon_i\}\Big) = 1, \tag{9.9}$$

Maintenant que nous avons réussi à faire sortir le « $\forall\varepsilon$ » de la probabilité, on peut laisser tomber la discrétisation pour aboutir à la caractérisation suivante de la convergence presque sûre.

Proposition 9.3 (une c.n.s. de convergence p.s.)**.** *La suite de variables aléatoires $(Y_n)_{n\geqslant 1}$ converge presque sûrement vers la variable aléatoire Y si et seulement si*

$$\forall \varepsilon > 0, \quad P\Big(\bigcup_{j\in\mathbb{N}} \bigcap_{k\geqslant j} \{|Y_k - Y| < \varepsilon\} \Big) = 1, \tag{9.10}$$

ou de manière équivalente,

$$\forall \varepsilon > 0, \quad P\Big(\bigcap_{j\in\mathbb{N}} \bigcup_{k\geqslant j} \{|Y_k - Y| \geqslant \varepsilon\} \Big) = 0. \tag{9.11}$$

Preuve. L'équivalence entre (9.10) et (9.11) est évidente par passage au complémentaire. Vérifions l'équivalence entre (9.10) et la c.n.s. (9.9) de convergence presque sûre. Il est clair que (9.10) implique (9.9). Pour la réciproque, il suffit de remarquer que la fonction $t \mapsto P\big(\bigcup_{j\in\mathbb{N}} \bigcap_{k\geqslant j} \{|Y_k - Y| < t\}\big)$ est *croissante*. En effet si $s < t$, l'inclusion d'évènements $\{|Y_k - Y| < s\} \subset \{|Y_k - Y| < t\}$ se propage à l'intersection sur $k \geqslant j$, puis à l'union sur j. Supposons (9.9) vraie et fixons $\varepsilon > 0$ quelconque. Comme la suite $(\varepsilon_i)_{i\geqslant 1}$ tend vers 0, on peut trouver un i_0 tel que $\varepsilon_{i_0} < \varepsilon$ et alors

$$1 = P\Big(\bigcup_{j\in\mathbb{N}} \bigcap_{k\geqslant j} \{|Y_k - Y| < \varepsilon_{i_0}\} \Big) \leqslant P\Big(\bigcup_{j\in\mathbb{N}} \bigcap_{k\geqslant j} \{|Y_k - Y| < \varepsilon\} \Big) \leqslant 1,$$

d'où (9.10). □

Intéressons nous maintenant à (9.11). En notant $A_k := \{|Y_k - Y| \geqslant \varepsilon\}$, on aimerait trouver une condition qui nous assure que $P(\bigcap_{j\in\mathbb{N}} \bigcup_{k\geqslant j} A_k) = 0$. En pratique, on a souvent une majoration des probabilités $P(A_k)$ *via* une inégalité du type inégalité de Markov et il est donc naturel de chercher une condition portant sur les $P(A_k)$. Ce problème a un intérêt propre, indépendant de la définition des A_k, provenant de l'interprétation suivante de l'évènement $\bigcap_{j\in\mathbb{N}} \bigcup_{k\geqslant j} A_k$:

$$\begin{aligned} \bigcap_{j\in\mathbb{N}} \bigcup_{k\geqslant j} A_k &= \{\omega \in \Omega \,;\, \omega \text{ appartient à une infinité de } A_k\} \\ &= \{\omega \in \Omega \,;\, \omega \text{ réalise une infinité de } A_k\} \\ &= \{\text{réalisation d'une infinité de } A_k\}. \end{aligned}$$

L'étude de la probabilité de cet évènement conduit aux deux lemmes de Borel Cantelli.

Lemme 9.4 (Borel Cantelli I)**.** *Soit $(A_n)_{n\in\mathbb{N}}$ une suite d'évènements telle que*

$$\sum_{n=0}^{+\infty} P(A_n) < +\infty. \tag{9.12}$$

Alors

$$P\big(\textit{réalisation d'une infinité de } A_n\big) = 0.$$

Preuve. Introduisons la notation [3] $A_\infty :=$ {réalisation d'une infinité de A_n} et posons $C_j := \bigcup_{k\geqslant j} A_k$. La suite $(C_j)_{j\in\mathbb{N}}$ est décroissante pour l'inclusion et d'intersection A_∞. Par continuité séquentielle décroissante de P, on a donc

$$P(C_j) \xrightarrow[j\to+\infty]{} P(A_\infty). \tag{9.13}$$

Par sous-σ-additivité de P, on a d'autre part

$$\forall j \in \mathbb{N}, \quad 0 \leqslant P(C_j) \leqslant \sum_{k\geqslant j} P(A_k) =: r(j). \tag{9.14}$$

Grâce à l'hypothèse (9.12), $r(j)$ est le reste d'une série *convergente*, donc tend vers 0 quand j tend vers $+\infty$. Combinée avec la majoration (9.14), cette convergence donne

$$P(C_j) \xrightarrow[j\to+\infty]{} 0. \tag{9.15}$$

La conclusion $P(A_\infty) = 0$ découle alors de (9.13) et (9.15). □

Le lemme de Borel Cantelli I est d'une portée très générale, puisqu'on obtient la conclusion $P(A_\infty) = 0$ sous la seule hypothèse de convergence de la série $\sum_{n\in\mathbb{N}} P(A_n)$, *sans rien supposer sur la structure de dépendance* de la suite (A_n). Pour les suites d'évènements indépendants, on a le résultat complémentaire suivant.

Lemme 9.5 (Borel Cantelli II)**.** *Soit $(A_n)_{n\in\mathbb{N}}$ une suite d'évènements* indépendants *telle que*

$$\sum_{n=0}^{+\infty} P(A_n) = +\infty. \tag{9.16}$$

Alors

$$P\big(\textit{réalisation d'une infinité de } A_n\big) = 1.$$

Preuve. Notons encore $A_\infty :=$ {réalisation d'une infinité de A_n} et posons

$$C_{j,l} := \bigcup_{j\leqslant k\leqslant l} A_k, \qquad C_j := \bigcup_{k\geqslant j} A_k,$$

d'où

$$A_\infty = \bigcap_{j\in\mathbb{N}} C_j.$$

Les A_k^c héritant de l'indépendance des A_k, cf. remarque 5.55, on a

$$P(C_{j,l}) = 1 - P(C_{j,l}^c) = 1 - P\Big(\bigcap_{k=j}^{l} A_k^c\Big) = 1 - \prod_{k=j}^{l}\big(1 - P(A_k)\big).$$

3. Cette écriture est commode, mais non standard, il convient donc d'en expliciter la définition en cas de réutilisation. On trouve dans la littérature la notation $\limsup_{n\to+\infty} A_n$ pour l'évènement « réalisation d'une infinité de A_n ». Cette notation est proscrite de cet ouvrage par choix pédagogique.

On utilise alors l'inégalité de convexité [4] $\mathrm{e}^{-x} \geqslant 1-x$ avec $x = P(A_k)$ pour obtenir la minoration

$$1 \geqslant P(C_{j,l}) \geqslant 1 - \prod_{k=j}^{l} \exp\bigl(-P(A_k)\bigr) = 1 - \exp\Bigl(-\sum_{k=j}^{l} P(A_k)\Bigr). \tag{9.17}$$

En laissant j fixe et faisant tendre l vers l'infini dans (9.17), on en déduit grâce à l'hypothèse (9.16) que $P(C_{j,l})$ tend vers 1. D'autre part C_j est limite croissante des $C_{j,l}$ pour l'inclusion, donc par continuité croissante séquentielle de P,

$$P(C_j) = \lim_{l \to +\infty} P(C_{j,l}) = 1.$$

Cette égalité étant vraie pour tout $j \in \mathbb{N}$, on en déduit par le lemme 9.2 que

$$P\Bigl(\bigcap_{j\in\mathbb{N}} C_j\Bigr) = 1.$$

Comme l'intersection de tous les C_j est l'évènement A_∞, le lemme est démontré. □

Après cette longue, mais utile, digression sur les lemmes de Borel Cantelli, revenons à notre quête d'une condition pratique de convergence presque sûre. Le premier lemme de Borel Cantelli nous permet d'aboutir au résultat suivant.

Proposition 9.6 (condition suffisante de convergence p.s.)**.** *Soient Y et $(Y_n)_{n\geqslant 1}$ des variables aléatoires réelles définies sur le même espace probabilisé $(\Omega, \mathcal{F}, P)$ et vérifiant*

$$\forall \varepsilon > 0, \quad \sum_{n=1}^{+\infty} P(|Y_n - Y| \geqslant \varepsilon) < +\infty. \tag{9.18}$$

Alors Y_n converge presque sûrement vers Y.

On dit d'une suite $(Y_n)_{n\geqslant 1}$ vérifiant (9.18) qu'elle converge *presque complètement* vers Y.

Preuve. Fixons $\varepsilon > 0$ quelconque et posons $A_n := \{|Y_n - Y| \geqslant \varepsilon\}$. Par le premier lemme de Borel-Cantelli, on déduit de (9.18) que la probabilité de réalisation d'une infinité de A_n est nulle, ce qui s'écrit encore

$$P\Bigl(\bigcap_{j\in\mathbb{N}} \bigcup_{k\geqslant j} \{|Y_k - Y| \geqslant \varepsilon\}\Bigr) = 0.$$

Ceci étant vérifié pour tout $\varepsilon > 0$, la c.n.s. (9.11) de la proposition 9.3 nous donne la convergence presque sûre de Y_n vers Y. □

4. La représentation graphique de la fonction convexe $x \mapsto \mathrm{e}^{-x}$ est toujours au-dessus de sa tangente à l'origine d'où $\mathrm{e}^{-x} \geqslant 1-x$ pour tout $x \in \mathbb{R}$.

Nous introduisons maintenant un nouveau mode de convergence de Y_n vers Y, la convergence en probabilité. Cette notion nous sera utile pour la loi faible des grands nombres.

Définition 9.7 (Convergence en probabilité)**.** *La suite Y_n converge en probabilité vers Y (notation $Y_n \xrightarrow[n\to+\infty]{\Pr} Y$) si*

$$\forall \varepsilon > 0, \quad P\big(|Y_n - Y| \geqslant \varepsilon\big) \xrightarrow[n\to+\infty]{} 0.$$

La convergence ci-dessus devant avoir lieu pour *tout* $\varepsilon > 0$, il est clair qu'on obtient une définition équivalente en remplaçant $P\big(|Y_n - Y| \geqslant \varepsilon\big)$ par $P\big(|Y_n - Y| > \varepsilon\big)$.

Notons la différence de point de vue par rapport à la convergence presque sûre. Dans la convergence presque-sûre, on reste proche de la notion de convergence simple de l'analyse. Il s'agit de la convergence de la suite de réels $(Y_n(\omega))_{n\geqslant 1}$, pour tous les ω d'un même évènement de probabilité 1. La convergence en probabilité ne concerne pas le comportement asymptotique individuel de chaque suite $(Y_n(\omega))_{n\geqslant 1}$, mais plutôt celui de la suite d'évènements $D_{n,\varepsilon} := \{|Y_n - Y| < \varepsilon\}$ dont la probabilité $P(D_{n,\varepsilon})$ doit tendre vers 1, pour tout ε. En pratique, établir la convergence en probabilité de Y_n vers Y est souvent un travail préliminaire pour prouver la convergence presque-sûre de Y_n vers Y. En effet si on utilise la condition suffisante de convergence p.s. (9.18), pour que la série converge, il faut déjà que son terme général tende vers 0 et cette convergence vers 0 (pour tout ε) est précisément la convergence en probabilité de Y_n vers Y. Bien sûr, comme (9.18) n'est qu'une condition suffisante, cette remarque ne nous permet pas d'affirmer que la convergence en probabilité est une notion plus faible que la convergence presque-sûre. Ce qui suit montre qu'il en est pourtant bien ainsi.

Proposition 9.8. *La convergence presque-sûre implique la convergence en probabilité.*

Preuve. Fixons $\varepsilon > 0$. L'hypothèse de convergence presque sûre de Y_n vers Y signifie que l'événement

$$\Omega' := \{\omega \in \Omega \,;\, \lim_{n\to+\infty} Y_n(\omega) = Y(\omega)\}$$

a pour probabilité 1. Définissons

$$\Omega'_\varepsilon := \{\omega \in \Omega \,;\, \exists k_0 = k_0(\omega),\ \forall n \geqslant k_0,\ |Y_n(\omega) - Y(\omega)| < \varepsilon\}.$$

C'est bien un évènement, c'est-à-dire $\Omega'_\varepsilon \in \mathcal{F}$, puisqu'il s'écrit

$$\Omega'_\varepsilon = \bigcup_{k\in\mathbb{N}*} \bigcap_{n\geqslant k} \{|Y_n - Y| < \varepsilon\}.$$

De plus Ω'_ε contient Ω', donc $P(\Omega'_\varepsilon) = 1$. Pour tout $k \geqslant 1$, notons

$$A_k := \{\omega \in \Omega \,;\, \forall n \geqslant k,\ |Y_n(\omega) - Y(\omega)| < \varepsilon\} = \bigcap_{n\geqslant k} \{|Y_n - Y| < \varepsilon\}.$$

La suite $(A_k)_{k\geqslant 1}$ est clairement croissante pour l'inclusion et sa réunion est Ω'_ε. Par continuité séquentielle croissante de P, on a donc $P(A_k) \uparrow P(\Omega'_\varepsilon) = 1$ $(k \to +\infty)$. Par conséquent,

$$\forall \delta > 0,\ \exists k_1, \quad P(A_{k_1}) > 1 - \delta.$$

Pour tout $n \geqslant k_1$, l'évènement $\{|Y_n - Y| < \varepsilon\}$ contient A_{k_1}, d'où

$$\forall n \geqslant k_1, \quad P(|Y_n - Y| < \varepsilon) > 1 - \delta.$$

En passant à l'évènement complémentaire, on obtient finalement

$$\forall \delta > 0, \exists k_1 \in \mathbb{N}, \forall n \geqslant k_1, \quad P(|Y_n - Y| \geqslant \varepsilon) < \delta.$$

Ceci établit la convergence de $P(|Y_n - Y| \geqslant \varepsilon)$ vers 0. Comme ε était quelconque, il y a bien convergence en probabilité de Y_n vers Y. □

Remarque 9.9. La convergence en probabilité n'implique pas la convergence presque-sûre. Voici un exemple de suite $(Y_n)_{n\geqslant 1}$ de variables aléatoires qui converge en probabilité vers 0 mais ne converge pas au sens presque-sûr. Dans une urne contenant 9 boules numérotées de 1 à 9, on effectue une suite infinie de tirages d'une boule avec remise. On note X_i le numéro sorti au i^{e} tirage. Pour définir Y_n, on considère l'écriture décimale unique de n, où les $c_i \in [\![0, 9]\!]$ sont ses chiffres décimaux :

$$n = c_j 10^j + c_{j-1} 10^{j-1} + \cdots + c_1 10 + c_0, \quad j \in \mathbb{N},\ c_j \neq 0.$$

On pose alors :

$$Y_n = \begin{cases} 1 & \text{si pour tout } i \in [\![1, j+1]\!],\ X_i = c_{i-1}, \\ 0 & \text{sinon.} \end{cases}$$

On peut prendre comme espace Ω, $[\![1, 9]\!]^{\mathbb{N}^*}$. Par exemple si les 4 premières composantes de ω sont 8, 2, 5, 9, c'est-à-dire si les 4 premiers tirages ont donné dans l'ordre ces numéros, alors pour $n \in [\![1, 9\,999]\!]$, tous les $Y_n(\omega)$ sont nuls [5] sauf $Y_8(\omega)$, $Y_{28}(\omega)$, $Y_{528}(\omega)$ et $Y_{9\,528}(\omega)$ qui valent 1. Notons que si l'écriture décimale de n contient au moins un chiffre 0, $Y_n(\omega)$ est nul pour tout $\omega \in \Omega$.

Vérifions maintenant la convergence en probabilité de $(Y_n)_{n\geqslant 1}$ vers 0. Clairement si $\varepsilon \geqslant 1$, $P(|Y_n| > \varepsilon) = P(\varnothing) = 0$. Pour $0 < \varepsilon < 1$, on remarque que :

$$P(|Y_n| > \varepsilon) = P(Y_n = 1) = P\left(\bigcap_{1\leqslant i\leqslant j+1} \{X_i = c_{i-1}\}\right) = \prod_{1\leqslant i\leqslant j+1} P(X_i = c_{i-1}),$$

par indépendance des tirages. Par conséquent,

$$P(|Y_n| > \varepsilon) = \begin{cases} 0 & \text{si l'un au moins des } c_i \text{ est nul,} \\ 9^{-j-1} & \text{sinon.} \end{cases}$$

5. En fait c'est vrai au moins pour $n \in [\![1, 19\,527]\!]$.

Comme le nombre $j+1$ de chiffres décimaux de n tend vers l'infini avec n, on a bien vérifié que $P(Y_n > \varepsilon)$ tend vers 0.

Montrons maintenant que la suite de réels $((Y_n(\omega))_{n\geqslant 1}$ ne converge pour *aucun* $\omega \in \Omega$, ce qui interdit évidemment la convergence p.s. On note d'abord que cette suite dont les termes ne peuvent valoir que 0 ou 1, comporte une infinité de zéros, par exemple tous les termes dont l'indice n s'écrit avec un chiffre 0. On en déduit que la limite inférieure de la suite est 0 (ou plus simplement qu'il y a une sous-suite qui converge vers 0). D'autre part, pour tout $\omega \in \Omega$, il y a une sous-suite infinie de $((Y_n(\omega))_{n\geqslant 1}$ dont tous les termes valent 1, cette sous-suite correspond aux entiers n dont les chiffres décimaux sont donnés (de droite à gauche) par les composantes de ω, comme illustré ci-dessus. Par conséquent la limite supérieure $((Y_n(\omega))_{n\geqslant 1}$ vaut 1 et cette suite ne peut avoir de limite.

Remarque 9.10. Il est possible d'obtenir la convergence presque sûre à partir de la convergence en probabilité, à condition d'avoir une *bonne vitesse* de convergence en probabilité. Le sens précis de cette affirmation est donné par la proposition 9.6.

La proposition 9.6 permet aussi, même sans bonne vitesse de convergence en probabilité, d'obtenir de la convergence p.s. pour une *sous-suite*.

Proposition 9.11 (convergence p.s. d'une sous-suite)**.** *Si la suite $(Y_n)_{n\geqslant 1}$ converge en probabilité vers Y, on peut en extraire une sous-suite $(Y_{n_i})_{i\geqslant 1}$ qui converge presque sûrement vers Y.*

Preuve. Notons $\varepsilon_i = 2^{-i}$. La convergence en probabilité de Y_n vers Y implique pour tout $i \geqslant 1$, la convergence de $P(|Y_n - Y| > \varepsilon_i)$ vers 0 quand n tend vers l'infini. On en déduit l'existence d'une suite strictement croissante d'indices n_i telle que

$$\forall i \geqslant 1, \quad P\big(|Y_{n_i} - Y| \geqslant \varepsilon_i\big) \leqslant \frac{1}{i^2}.$$

Vérifions maintenant que

$$\forall \varepsilon > 0, \quad \sum_{i=1}^{+\infty} P(|Y_{n_i} - Y| \geqslant \varepsilon) < +\infty.$$

En effet la convergence vers 0 de ε_i nous assure de l'existence d'un $i_0 = i_0(\varepsilon)$ tel que pour tout $i \geqslant i_0$, $\varepsilon_i < \varepsilon$. Pour $i \geqslant i_0$, on a donc $\{|Y_{n_i} - Y| \geqslant \varepsilon\} \subset \{|Y_{n_i} - Y| \geqslant \varepsilon_i\}$ et cette inclusion d'évènements nous permet de majorer le terme général de la série ci-dessus par i^{-2} à partir du rang i_0. Ainsi la suite (Y_{n_i}) converge presque complètement vers Y donc aussi presque sûrement. □

Remarque 9.12. Dans la preuve ci-dessus, on peut prendre pour $(\varepsilon_i)_{i\geqslant 1}$ n'importe quelle suite de réels strictement positifs tendant vers 0 et on peut remplacer i^{-2} par u_i, terme général strictement positif d'une série convergente.

9.1.2 Convergence en moyenne d'ordre p

Nous introduisons maintenant un nouveau mode de convergence, utile notamment dans les problèmes d'interversion limite espérance.

Définition 9.13. *Soit $p \geqslant 1$ un réel et $(Y_n)_{n\geqslant 1}$ une suite de v.a. ayant un moment absolu d'ordre p fini. On dit que cette suite converge en moyenne d'ordre p (ou au sens L^p) vers la v.a. Y si*

$$\lim_{n\to+\infty} \mathbf{E}\big(|Y_n - Y|^p\big) = 0. \tag{9.19}$$

Notation : $Y_n \xrightarrow[n\to+\infty]{\mathrm{L}^p} Y$.

Remarque 9.14. Si Y_n converge vers Y en moyenne d'ordre p, Y a nécessairement un moment absolu d'ordre p fini. Pour le voir, on utilise la *convexité* de la fonction $\varphi : \mathbb{R}_+ \to \mathbb{R}_+$, $x \mapsto x^p$, pour $p \geqslant 1$. Ceci signifie que la « corde » entre deux points de son graphe est « au dessus » de l'arc de courbe correspondant, ou encore que « l'image du barycentre » est majorée par « le barycentre des images », voir la figure 9.1. Cette convexité implique notamment que pour tous $a, b \geqslant 0$, $\varphi\big(\frac{a+b}{2}\big) \leqslant \frac{1}{2}\varphi(a) + \frac{1}{2}\varphi(b)$, ce qui s'écrit encore

$$\forall a, b \geqslant 0, \quad (a+b)^p \leqslant 2^{p-1}(a^p + b^p). \tag{9.20}$$

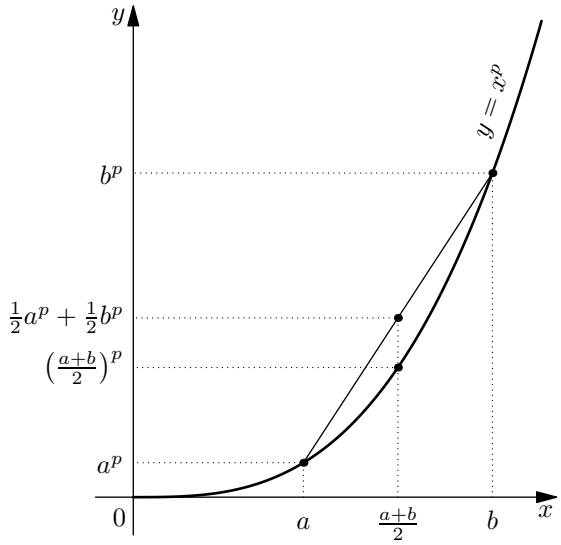


FIGURE 9.1 – Convexité de $x \mapsto x^p$ et inégalité $\big(\frac{a+b}{2}\big)^p \leqslant \frac{1}{2}(a^p + b^p)$

Par croissance de φ, inégalité triangulaire et (9.20) appliquée avec $a = |Y_n(\omega)|$ et

$b = |Y(\omega) - Y_n(\omega)|$, on voit que pour tout $\omega \in \Omega$,

$$|Y(\omega)|^p \leqslant \big(|Y_n(\omega)| + |Y(\omega) - Y_n(\omega)|\big)^p \leqslant 2^{p-1}|Y_n(\omega)|^p + 2^{p-1}|Y(\omega) - Y_n(\omega)|^p.$$

Par croissance de l'espérance, on en déduit :

$$\mathbf{E}|Y|^p \leqslant 2^{p-1}\mathbf{E}|Y_n|^p + 2^{p-1}\mathbf{E}|Y - Y_n|^p.$$

Ce majorant est fini car $\mathbf{E}|Y_n|^p$ est fini par hypothèse et $\mathbf{E}|Y - Y_n|^p$ tend vers 0 donc est fini au moins pour n assez grand.

Proposition 9.15. *La convergence en moyenne d'ordre p implique la convergence en probabilité.*

Preuve. C'est une conséquence immédiate de l'inégalité de Markov avec moment, cf. la proposition 7.38, puisque

$$\forall \varepsilon > 0, \quad P\big(|Y_n - Y| \geqslant \varepsilon\big) \leqslant \frac{\mathbf{E}\big(|Y_n - Y|^p\big)}{\varepsilon^p}.$$

□

Remarque 9.16. En revanche, il n'y a aucune implication entre convergence p.s. et convergence en moyenne d'ordre p. Pour voir que la convergence L^p n'implique pas la convergence p.s., le contre exemple déjà vu à la remarque 9.9 fait l'affaire car on vérifie (exercice) que la suite Y_n de cet exemple converge au sens L^p vers 0 (pour tout $p \geqslant 1$). Voici maintenant un contre exemple montrant que la convergence p.s. n'implique pas la convergence L^p. On prend U de loi uniforme sur $[0,1]$ et on pose $Y_n := n^2\mathbf{1}_{[0,1/n]}(U)$. On voit facilement que

$$P\Big(\big\{\omega \in \Omega\,;\ \lim_{n\to+\infty} Y_n(\omega) = 0\big\}\Big) = P\big(U \in]0,1]\big) = 1,$$

d'où la convergence presque-sûre de Y_n vers la v.a. constante 0. D'autre part, Y_n est une v.a. discrète ne prenant que les valeurs 0 ou n^2 et on a immédiatement $\mathbf{E}|Y_n|^p = n^{2p-1}$. Le fait que ce moment absolu d'ordre p tende vers $+\infty$ avec n interdit la convergence L^p de la suite (Y_n). Pour s'en convaincre, supposons que Y_n converge vers Y au sens L^p. Alors $\mathbf{E}|Y|^p$ doit être fini et par l'inégalité de convexité (9.20), on obtient

$$\mathbf{E}|Y_n|^p \leqslant 2^{p-1}\mathbf{E}|Y|^p + 2^{p-1}\mathbf{E}|Y_n - Y|^p,$$

ce qui est impossible puisque le premier membre tend vers l'infini avec n tandis que le second membre reste borné.

Il y a une hiérarchie entre les convergences L^p pour diverses valeurs de p et c'est la même que celle établie à la proposition 7.37 pour l'existence des moments absolus d'ordre p.

Proposition 9.17 (hiérarchie des convergences L^p)**.** *Si* $1 \leqslant p < r < +\infty$ *et si* $Y_n \xrightarrow[n\to+\infty]{\mathrm{L}^r} Y$*, alors* $Y_n \xrightarrow[n\to+\infty]{\mathrm{L}^p} Y$*.*

Preuve. Posons $Z_n := |Y_n - Y|$. Il s'agit de montrer que la convergence vers 0 de $\mathbf{E}(Z_n^r)$ implique celle de $\mathbf{E}(Z_n^p)$ vers 0. Fixons ε arbitraire dans $]0,1[$. On commence par le découpage

$$\mathbf{E}(Z_n^p) = \int_0^{+\infty} P(Z_n^p > t)\,\mathrm{d}t = \int_0^1 P(Z_n^p > t)\,\mathrm{d}t + \int_1^{+\infty} P(Z_n^p > t)\,\mathrm{d}t. \tag{9.21}$$

Pour tout réel $x \geqslant 1$, on a $x^p \leqslant x^r$ puisque $p < r$. On en déduit que l'inclusion d'évènements $\{Z_n^p > t\} \subset \{Z_n^r > t\}$ est vraie pour tout $t \geqslant 1$, d'où

$$\int_1^{+\infty} P(Z_n^p > t)\,\mathrm{d}t \leqslant \int_1^{+\infty} P(Z_n^r > t)\,\mathrm{d}t \leqslant \mathbf{E}(Z_n^r). \tag{9.22}$$

Comme ce dernier majorant tend vers 0 quand n tend vers $+\infty$, on peut trouver un entier $n_1 = n_1(\varepsilon)$ tel que

$$\forall n \geqslant n_1, \quad \int_1^{+\infty} P(Z_n^p > t) < \varepsilon. \tag{9.23}$$

Rappelant que $\varepsilon < 1$, on a d'autre part

$$\begin{aligned}
\int_0^1 P(Z_n^p > t)\,\mathrm{d}t &= \int_0^{\varepsilon} P(Z_n^p > t)\,\mathrm{d}t + \int_{\varepsilon}^1 P(Z_n^p > t)\,\mathrm{d}t \\
&\leqslant \int_0^{\varepsilon} \mathrm{d}t + \int_{\varepsilon}^1 P(Z_n^p > t)\,\mathrm{d}t \\
&= \varepsilon + \int_{\varepsilon}^1 P(Z_n > t^{1/p})\,\mathrm{d}t \\
&\leqslant \varepsilon + (1-\varepsilon)P(Z_n > \varepsilon^{1/p}),
\end{aligned}$$

d'où finalement

$$\forall n \in \mathbb{N}^*, \quad \int_0^1 P(Z_n^p > t)\,\mathrm{d}t \leqslant \varepsilon + P(Z_n > \varepsilon^{1/p}). \tag{9.24}$$

Par la proposition 9.15, la convergence de Z_n vers 0 au sens L^r implique sa convergence en probabilité [6]. Par conséquent $P(Z_n > \varepsilon^{1/p})$ converge vers 0 quand n tend vers $+\infty$, ce qui nous assure de l'existence d'un entier $n_2 = n_2(\varepsilon)$ tel que

$$\forall n \geqslant n_2, \quad \int_0^1 P(Z_n^p > t)\,\mathrm{d}t \leqslant 2\varepsilon. \tag{9.25}$$

6. Remarquer que $\mathbf{E}(Z_n^r) = \mathbf{E}|Z_n - 0|^r$ et donc que la convergence L^r de Z_n vers 0 équivaut à la convergence vers 0 de $\mathbf{E}(Z_n^r)$. Cette simplification est particulière à la limite 0, on n'a bien sûr pas équivalence en général entre convergence de Z_n vers Z au sens L^r et convergence de $\mathbf{E}|Z_n|^r$ vers $\mathbf{E}|Z|^r$.

On déduit de (9.21), (9.23) et (9.25) que

$$\forall n \geqslant n_0 := \max(n_1, n_2), \quad \mathbf{E}(Z_n^p) < 3\varepsilon.$$

Comme $\varepsilon \in]0,1[$ était arbitraire, $\mathbf{E}(Z_n^p)$ tend bien vers 0 quand n tend vers $+\infty$. □

Remarque 9.18. La démonstration classique de la proposition 9.17 est plus rapide que celle présentée ci-dessus puisqu'elle est une conséquence immédiate de l'inégalité

$$(\mathbf{E}|X|^p)^{1/p} \leqslant (\mathbf{E}|X|^r)^{1/r}, \quad \forall 1 \leqslant p < r < +\infty.$$

On pourra éventuellement voir cette inégalité en exercice. La preuve proposée ci-dessus, certes moins élégante, a néanmoins l'avantage de donner comme sous-produit le résultat suivant qui a son intérêt propre.

Proposition 9.19. *Si la suite de variables aléatoires Y_n est bornée par une constante positive c (ou plus largement si pour tout n, $P(|Y_n| \leqslant c) = 1$) et converge en probabilité vers Y, alors Y_n converge au sens L^p vers Y, pour tout $p \geqslant 1$. En particulier pour $p = 1$, on en déduit l'interversion limite espérance :* $\lim_{n\to+\infty} \mathbf{E}Y_n = \mathbf{E}Y$.

L'argument de la preuve est essentiellement le même que ci-dessus, *modulo* quelques adaptations mineures que l'on vous laisse le soin de rédiger en exercice.

Nous allons maintenant établir le célèbre théorème de *convergence dominée* qui est très utile pour l'interversion limite espérance.

Théorème 9.20 (convergence dominée)**.** *On suppose que les variables aléatoires réelles Y_n ($n \geqslant 1$), Y et Z définies sur le même espace probabilisé $(\Omega, \mathcal{F}, P)$ vérifient*

a) Y_n converge presque-sûrement vers Y ;

b) pour tout $n \geqslant 1$, $|Y_n| \leqslant Z$ p.s. ;

c) Z est intégrable.

Dans ces conditions,

1. *les Y_n et Y sont intégrables ;*
2. *Y_n converge vers Y au sens L^1 : $\mathbf{E}|Y_n - Y| \to 0$;*
3. *on a l'interversion limite espérance :* $\lim_{n\to+\infty} \mathbf{E}Y_n = \mathbf{E}Y$.

Preuve. En utilisant a), b) et le lemme 9.2, on vérifie facilement l'existence d'un évènement $\Omega' \in \mathcal{F}$ de probabilité 1 tel que

$$\forall \omega \in \Omega', \forall n \geqslant 1, \quad |Y_n(\omega)| \leqslant Z(\omega) \text{ et } |Y(\omega)| \leqslant Z(\omega). \tag{9.26}$$

Remarquons que pour tout évènement A, $P(A \cap \Omega'^c) \leqslant P(\Omega'^c)$, d'où :

$$\text{quand } P(\Omega') = 1, \quad \forall A \in \mathcal{F}, \quad P(A \cap \Omega'^c) = 0 \quad \text{et} \quad P(A \cap \Omega') = P(A). \tag{9.27}$$

On en déduit de (9.26) et (9.27) que

$$\forall t \geqslant 0, \quad P(|Y_n| > t) = P(\{|Y_n| > t\} \cap \Omega') \leqslant P(\{Z > t\} \cap \Omega') = P(Z > t)$$

et de même avec Y à la place de Y_n. En intégrant sur $\mathbb{R}_+$ relativement à t, on obtient grâce à c) les inégalités

$$\mathbf{E}|Y_n| \leqslant \mathbf{E}Z < +\infty, \quad \mathbf{E}|Y| \leqslant \mathbf{E}Z < +\infty,$$

qui établissent l'intégrabilité des Y_n et de Y.

Une fois établie cette intégrabilité et donc l'existence de $\mathbf{E}Y_n$ et $\mathbf{E}Y$, l'interversion limite espérance sera une conséquence immédiate de la convergence L^1 puisque :

$$|\mathbf{E}Y_n - \mathbf{E}Y| = |\mathbf{E}(Y_n - Y)| \leqslant \mathbf{E}|Y_n - Y|.$$

Il nous reste donc à prouver la convergence L^1, autrement dit, en posant $Z_n := |Y_n - Y|$, la convergence vers 0 de $\mathbf{E}Z_n$.

Pour cela on utilise comme dans la preuve de la proposition 9.17 un découpage en trois de l'intégrale définissant $\mathbf{E}Z_n$ en écrivant pour $\varepsilon > 0$ arbitraire et $b = b(\varepsilon)$ convenablement choisi, $\int_0^{+\infty} = \int_0^{\varepsilon} + \int_{\varepsilon}^{b} + \int_b^{+\infty}$. Voyons d'abord le choix de b. Par inégalité triangulaire, $Z_n \leqslant 2Z$ sur Ω', ce qui grâce à (9.27) nous donne pour tout $t > 0$, $P(Z_n > t) \leqslant P(2Z > t)$. La variable aléatoire Z étant intégrable par hypothèse c), il en est de même pour $2Z$, ce qui implique la convergence dans $\mathbb{R}_+$ de l'intégrale de Riemann généralisée $\int_0^{+\infty} P(2Z > t)\,\mathrm{d}t$. On peut donc choisir b assez grand (et supérieur à ε) pour que $\int_b^{+\infty} P(2Z > t)\,\mathrm{d}t < \varepsilon$. On a alors

$$\forall n \in \mathbb{N}^*, \quad \int_b^{+\infty} P(Z_n > t)\,\mathrm{d}t \leqslant \int_b^{+\infty} P(2Z > t)\,\mathrm{d}t < \varepsilon. \tag{9.28}$$

Ensuite on contrôle l'intégrale $\int_0^b P(Z_n > t)\,\mathrm{d}t$ en écrivant

$$\int_0^b P(Z_n > t)\,\mathrm{d}t \leqslant \int_0^{\varepsilon} \mathrm{d}t + \int_{\varepsilon}^{b} P(Z_n > t)\,\mathrm{d}t \leqslant \varepsilon + (b-\varepsilon)P(Z_n > \varepsilon) \leqslant \varepsilon + bP(Z_n > \varepsilon).$$

Par l'hypothèse a), Z_n converge presque-sûrement vers 0, donc converge aussi en probabilité vers 0. On peut donc trouver un entier $n_0 = n_0(\varepsilon)$ tel que $bP(Z_n > \varepsilon) < \varepsilon$ pour tout [7] $n \geqslant n_0$. On a alors

$$\forall n \geqslant n_0, \quad \int_0^b P(Z_n > t)\,\mathrm{d}t < 2\varepsilon. \tag{9.29}$$

En recollant les morceaux à partir de (9.28) et (9.29), on obtient

$$\forall n \geqslant n_0, \quad \mathbf{E}Z_n = \int_0^b P(Z_n > t)\,\mathrm{d}t + \int_b^{+\infty} P(Z_n > t)\,\mathrm{d}t < 3\varepsilon.$$

Comme $\varepsilon > 0$ était arbitraire, ceci établit la convergence vers 0 de $\mathbf{E}Z_n$ et achève la preuve. □

7. Noter que b dépend de ε et peut très bien tendre vers $+\infty$ quand ε tend vers 0, mais que l'on travaille avec un ε arbitraire fixé, donc aussi avec b fixé.

Dans l'esprit de la proposition 9.19, on peut généraliser le théorème de convergence dominée en remplaçant la convergence p.s. par la convergence en probabilité.

Théorème 9.21 (convergence dominée en probabilité)**.** *Le théorème 9.20 de convergence dominée reste valide en remplaçant l'hypothèse a) de convergence presque sûre par :*

a′) Y_n converge en probabilité vers Y.

Preuve. L'examen de la preuve du théorème 9.20 montre que la convergence p.s. de Y_n vers Y est utilisée à deux endroits. D'abord dans (9.26) pour obtenir l'inégalité $|Y(\omega)| \leqslant Z(\omega)$ sur Ω' de probabilité 1. Cette inégalité résulte du passage à la limite dans l'inégalité $|Y_n(\omega)| \leqslant Z(\omega)$, valide sur Ω'. Si l'on remplace la convergence p.s. de Y_n vers Z par de la convergence en probabilité, on obtiendra la même inégalité en extrayant de (Y_n) une sous-suite convergeant p.s. vers Y, cf. la proposition 9.11. La deuxième utilisation de la convergence p.s. est pour obtenir $bP(Z_n > \varepsilon) < \varepsilon$ grâce à la convergence en probabilité de Z_n vers 0. Or cette dernière équivaut à l'hypothèse a'). □

9.1.3 Bilan sur les modes de convergence

Arrivé à ce stade, il est bon de faire le point sur les différents modes de convergence étudiés. Le diagramme de la figure 9.2 résume les relations entre ces modes de convergence. Les flèches en trait plein représentent des implications (si la suite converge selon le mode de la case de départ, alors elle converge aussi selon celui de la case d'arrivée). Les flèches en tirets signalent l'existence d'une sous-suite convergente selon le mode de la case d'arrivée.

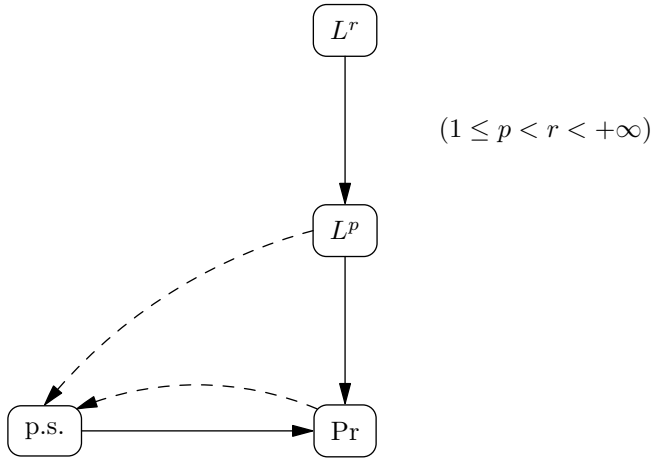


FIGURE 9.2 – Diagramme des convergences des suites de v.a.

Signalons deux propriétés communes aux trois modes de convergence étudiés jusqu'ici (vérification laissée au lecteur).

D'abord la limite pour ces modes de convergence n'est pas à strictement parler unique. Elle l'est *modulo* l'égalité presque sûre. Notons (m), (m') l'un des trois modes de convergence (p.s., en probabilité ou L^p). On peut vérifier que si Y_n converge vers Y au sens (m) et vers Y' au sens (m) alors $Y = Y'$ presque-sûrement (la réciproque est évidente). D'autre part si Y_n converge au sens (m) vers Y et au sens (m') vers Y', ces deux convergences impliquent la convergence en probabilité de Y_n vers Y et vers Y', donc l'égalité p.s. de Y et Y'.

Chacun des trois modes de convergence est compatible avec la structure d'espace vectoriel. Si $X_n \xrightarrow{(m)} X$ et $Y_n \xrightarrow{(m)} Y$, $X_n + Y_n \xrightarrow{(m)} X + Y$ et $aX_n \xrightarrow{(m)} aX$ pour tout $a \in \mathbb{R}$.

9.2 Loi des grands nombres

Nous abordons maintenant les lois des grands nombres. Il s'agit d'étudier la convergence des moyennes arithmétiques S_n/n construites à partir d'une suite de variables aléatoires $(X_i)_{i\geqslant 1}$ en posant $S_n := X_1 + \cdots + X_n$. Si S_n/n converge en probabilité, on parle d'une loi faible des grands nombres, tandis que si elle converge presque-sûrement on parle d'une loi forte.

9.2.1 Loi faible des grands nombres

Rappelons que si X est de carré intégrable ($\mathbf{E}X^2 < +\infty$), sa variance est définie par

$$\operatorname{Var} X := \mathbf{E}(X - \mathbf{E}X)^2.$$

Les X_k sont dites deux à deux *non-corrélées* si $\operatorname{Cov}(X_i, X_j) = 0$ pour tous i, j distincts. Ceci se produit en particulier lorsque les X_k sont deux à deux indépendantes, cf. prop. 8.55. Pour des X_k deux à deux non-corrélées, on déduit de (8.9) l'égalité

$$\operatorname{Var} S_n = \sum_{k=1}^{n} \operatorname{Var} X_k. \tag{9.30}$$

Proposition 9.22 (inégalité de Bienaymé-Tchebycheff)**.** *Si les X_k sont de carré intégrable et deux à deux non-corrélées,*

$$\forall t > 0, \quad P\Big(\Big| \sum_{k=1}^{n} (X_k - \mathbf{E}X_k) \Big| \geqslant t \Big) \leqslant \frac{1}{t^2} \sum_{k=1}^{n} \operatorname{Var} X_k. \tag{9.31}$$

Preuve. Il suffit d'écrire :

$$\begin{aligned} P\big(|S_n - \mathbf{E}S_n| \geqslant t\big) = P\big(|S_n - \mathbf{E}S_n|^2 \geqslant t^2\big) &\leqslant \frac{1}{t^2}\mathbf{E}(S_n - \mathbf{E}S_n)^2 \qquad (9.32)\\ &= \frac{1}{t^2} \operatorname{Var} S_n, \end{aligned}$$

où l'inégalité dans (9.32) est l'inégalité de Markov appliquée à la variable aléatoire positive $|S_n - \mathbf{E}S_n|^2$. On conclut avec (9.30). □

Théorème 9.23 (loi faible des grands nombres)**.** *Si les X_k sont de même loi, de carré intégrable et deux à deux non-corrélées, on a la convergence en probabilité :*

$$\frac{1}{n}\sum_{k=1}^{n} X_k \xrightarrow[n\to+\infty]{\Pr} \mathbf{E}X_1. \tag{9.33}$$

Preuve. Comme les X_k ont même loi, on a pour tout k les égalités $\mathbf{E}X_k = \mathbf{E}X_1$ et $\operatorname{Var} X_k = \operatorname{Var} X_1$. Comme elles sont aussi deux à deux non-corrélées, (9.30) nous donne $\operatorname{Var} S_n = n \operatorname{Var} X_1$. Par linéarité de l'espérance on a aussi $\mathbf{E}S_n = n\mathbf{E}X_1$. L'inégalité de Bienaymé-Tchebycheff nous dit alors que :

$$\forall t > 0, \quad P\big(|S_n - \mathbf{E}S_n| \geqslant t\big) = P\big(|S_n - n\mathbf{E}X_1| \geqslant t\big) \leqslant \frac{n \operatorname{Var} X_1}{t^2}.$$

Posant $t = n\varepsilon$, on en déduit :

$$\forall \varepsilon > 0, \forall n \in \mathbb{N}^*,\ P\big(|S_n - n\mathbf{E}X_1| \geqslant n\varepsilon\big) = P\Big(\Big|\frac{S_n}{n} - \mathbf{E}X_1\Big| \geqslant \varepsilon\Big) \leqslant \frac{n \operatorname{Var} X_1}{n^2\varepsilon^2} = \frac{\operatorname{Var} X_1}{n\varepsilon^2}.$$

Pour tout $\varepsilon > 0$ fixé, on a ainsi

$$P\Big(\Big|\frac{S_n}{n} - \mathbf{E}X_1\Big| \geqslant \varepsilon\Big) \leqslant \frac{\operatorname{Var} X_1}{n\varepsilon^2} \xrightarrow[n\to+\infty]{} 0,$$

ce qui établit la convergence en probabilité de la suite de variables aléatoires S_n/n vers la variable aléatoire constante $\mathbf{E}X_1$. □

9.2.2 Loi forte des grands nombres

Dans le cadre de cet ouvrage, nous limiterons notre étude des lois fortes des grands nombres au cas où les X_k sont i.i.d. (indépendantes identiquement distribuées), c'est-à-dire indépendantes et de même loi.

Théorème 9.24 (loi forte des grands nombres de Khintchine)**.** *On suppose les X_k indépendantes,* de même loi *et $\mathbf{E}|X_1| < +\infty$. Alors*

$$\frac{1}{n}\sum_{k=1}^{n} X_k \xrightarrow[n\to+\infty]{\text{p.s.}} \mathbf{E}X_1. \tag{9.34}$$

Ce résultat est le meilleur possible en raison du théorème suivant.

Théorème 9.25 (réciproque de la LFGN de Khintchine)**.** *Soit $(X_k)_{k\geqslant 1}$ une suite de variables aléatoires indépendantes et de même loi telle que S_n/n converge presque sûrement. Alors $\mathbf{E}|X_1| < +\infty$ et la limite* p.s. *de S_n/n est la constante $\mathbf{E}X_1$.*

La démonstration de ces deux théorèmes sort du cadre de cet ouvrage. Nous nous contenterons de prouver (9.34) sous l'hypothèse plus restrictive $\mathbf{E}(X_1^2) < +\infty$.

Preuve de (9.34) pour des X_k i.i.d. de carré intégrable. Puisque les X_k ont même loi, elles sont intégrables comme X_1 et de même espérance. On en déduit que S_n et S_n/n sont intégrables et que par linéarité de l'espérance,

$$\mathbf{E}\Big(\frac{S_n}{n}\Big) = \frac{1}{n}\mathbf{E}S_n = \frac{1}{n}(n\mathbf{E}X_1) = \mathbf{E}X_1.$$

Posons $X'_k := X_k - \mathbf{E}X_k$ et $S'_n := \sum_{k=1}^n X'_k = S_n - n\mathbf{E}X_1$. Alors $S'_n/n = S_n/n - \mathbf{E}X_1$, donc la convergence p.s. de S_n/n vers $\mathbf{E}X_1$ équivaut à la convergence p.s. de S'_n/n vers 0 (noter aussi que les X'_k sont i.i.d., propriété héritée des X_k). On ne perd donc pas de généralité en supposant désormais pour le confort d'écriture que $\mathbf{E}X_1 = 0$. On a alors $\operatorname{Var} X_1 = \mathbf{E}(X_1^2) =: \sigma^2$ et il s'agit maintenant de prouver que

$$M_n := \frac{1}{n}S_n \xrightarrow[n\to+\infty]{\text{p.s.}} 0. \tag{9.35}$$

Montrons dans un premier temps que la *sous-suite* $(M_{n^2})_{n\geqslant 1}$ converge p.s. vers 0. En effet l'inégalité de Bienaymé-Tchebycheff nous donne pour tout $\varepsilon > 0$,

$$P(|M_{n^2}| \geqslant \varepsilon) = P(|S_{n^2}| \geqslant n^2\varepsilon) \leqslant \frac{\operatorname{Var}(S_{n^2})}{\varepsilon^2 n^4} = \frac{n^2 \operatorname{Var} X_1}{\varepsilon^2 n^4} = \frac{\sigma^2}{\varepsilon^2 n^2}.$$

On en déduit que

$$\forall \varepsilon > 0, \quad \sum_{n=1}^{+\infty} P(|M_{n^2}| \geqslant \varepsilon) \leqslant \frac{\sigma^2}{\varepsilon^2}\sum_{n=1}^{+\infty}\frac{1}{n^2} < +\infty,$$

ce qui implique, par la proposition 9.6 :

$$M_{n^2} \xrightarrow[n\to+\infty]{\text{p.s.}} 0. \tag{9.36}$$

Pour tout $n \in \mathbb{N}^*$, notons $r(n)$ la partie entière de $n^{1/2}$, ce qui nous donne l'encadrement

$$r(n)^2 \leqslant n < \big(r(n)+1\big)^2. \tag{9.37}$$

La suite d'entiers $(r(n))_{n\geqslant 1}$ tend vers l'infini comme $\sqrt{n}$, mais avec des blocs de valeurs répétées de plus en plus longs : $r(n) = 4$ pour $n \in [\![16, 24]\!]$, $r(n) = 5$ pour $n \in [\![25, 35]\!]$, etc. Pour cette raison $(M_{r(n)^2})_{n\geqslant 1}$ n'est pas à proprement parler une sous-suite de $(M_n)_{n\geqslant 1}$. Nous allons néanmoins voir que l'on peut raccrocher le comportement asymptotique de $(M_n)_{n\geqslant 1}$ à celui de $(M_{r(n)^2})_{n\geqslant 1}$ en commençant par écrire :

$$\begin{aligned} M_n = \frac{1}{n}\sum_{1\leqslant j\leqslant r(n)^2} X_j + \frac{1}{n}\sum_{r(n)^2<j\leqslant n} X_j &= \frac{r(n)^2}{n}\frac{1}{r(n)^2}\sum_{1\leqslant j\leqslant r(n)^2} X_j + T_n \\ &= \frac{r(n)^2}{n}M_{r(n)^2} + T_n, \end{aligned}$$

où l'on a posé

$$T_n := \frac{1}{n} \sum_{r(n)^2 < j \leqslant n} X_j.$$

Par indépendance des X_j, on a

$$\operatorname{Var} T_n = \frac{1}{n^2}\big(n - r(n)^2\big) \operatorname{Var} X_1. \tag{9.38}$$

En utilisant (9.37),

$$\frac{1}{n^2}\big(n - r(n)^2\big) \leqslant \frac{\big(r(n)+1\big)^2 - r(n)^2}{n^2} = \frac{2r(n)+1}{n^2} \leqslant \frac{2n^{1/2}+1}{n^2} \leqslant \frac{3}{n^{3/2}}.$$

En reportant cette majoration dans (9.38) et en utilisant l'inégalité de Markov avec moment d'ordre 2, on obtient en notant que $\mathbf{E}T_n = 0$:

$$\forall \varepsilon > 0, \quad P(|T_n| > \varepsilon) \leqslant \frac{\operatorname{Var} T_n}{\varepsilon^2} \leqslant \frac{3\sigma^2}{\varepsilon^2} \frac{1}{n^{3/2}}.$$

Ce majorant étant le terme général d'une série convergente, on en déduit par une nouvelle application de la proposition 9.6 :

$$T_n \xrightarrow[n\to+\infty]{\text{p.s.}} 0. \tag{9.39}$$

Pour conclure, on écrit

$$M_n = T_n + \frac{r(n)^2}{n} M_{r(n)^2}.$$

Comme $r(n)^2/n$ est toujours dans $[0,1]$, on déduit de (9.36) et (9.39) que M_n converge presque sûrement vers zéro [8]. □

L'application la plus simple et aussi une des plus importantes du théorème 9.24 est la convergence des fréquences de succès dans une suite d'épreuves répétées de Bernoulli indépendantes. Ce résultat explique *a posteriori* l'approche « fréquentiste » dans la définition d'une probabilité. En effet si $(X_k)_{k\geqslant 1}$ est une suite de variables aléatoires de Bernoulli indépendantes et de même paramètre p, le théorème 9.24 nous donne la convergence presque sûre de $n^{-1}S_n$ vers $\mathbf{E}X_1 = p$. Soit maintenant $A \in \mathcal{F}$ un évènement et $(A_k)_{k\geqslant 1}$ une suite d'évènements indépendants de même probabilité que A. En prenant $X_k = \mathbf{1}_{A_k}$ et en notant que $\mathbf{E}\mathbf{1}_{A_k} = P(A_k) = P(A)$, on obtient

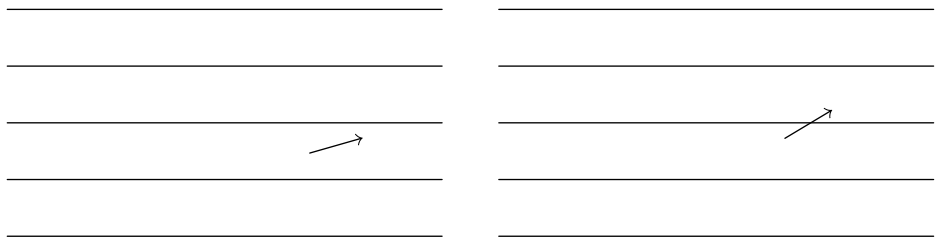
$$\frac{1}{n}\sum_{k=1}^{n} \mathbf{1}_{A_k} \xrightarrow[n\to+\infty]{\text{p.s.}} P(A).$$

Par exemple si A est l'évènement « obtention du cinq lors du lancer d'un dé équilibré », ceci nous dit que la *fréquence* d'obtention du cinq en n lancers converge presque sûrement vers $1/6$ lorsque n tend vers l'infini.

8. Noter que les suites $(M_{r(n)^2})_{n\geqslant 1}$ et $(M_{n^2})_{n\geqslant 1}$ ne sont pas les mêmes. La première a des séquences de termes consécutifs répétés de plus en plus longues et c'est en effaçant ces répétitions que l'on retrouve la deuxième. On voit ainsi que ces deux suites ont même limite. Au lecteur d'écrire proprement la justification de ce point.

9.2.3 L'aiguille de Buffon

À titre d'illustration historique de la loi forte des grands nombres, nous présentons maintenant une méthode expérimentale pour obtenir une approximation numérique du nombre π dont l'intérêt est essentiellement d'ordre culturel [9]. Cette méthode a été proposée en 1 777 par le célèbre naturaliste Buffon [10]. On trace sur une surface plane horizontale des droites parallèles équidistantes, séparées par une distance a (on peut par exemple utiliser les rainures d'un parquet). On laisse tomber sur cette surface une aiguille de longueur $\ell \leqslant a$ et une fois l'aiguille immobilisée, on observe si elle coupe l'une des droites du réseau.

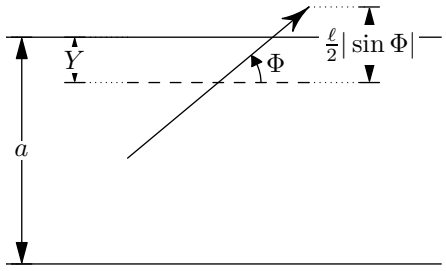


« Échec » « Succès »

On répète l'expérience en notant la fréquence des intersections. Lorsque le nombre d'expériences augmente indéfiniment, cette fréquence converge, selon Buffon, vers $p = \frac{2\ell}{\pi a}$, permettant ainsi d'obtenir une estimation expérimentale du nombre π.

Cherchons une modélisation de cette expérience. On note Y la distance du milieu de l'aiguille à la droite du réseau la plus proche. Y prend ses valeurs dans $[0, \frac{a}{2}]$. On note Φ une mesure de l'angle entre les droites du réseau (toutes orientées dans le même sens) et l'aiguille orientée du chas vers la pointe. Φ prend ses valeurs dans $[0, 2\pi]$ (par exemple) [11].

Y et Φ sont des variables aléatoires. La connaissance du couple $(Y(\omega), \Phi(\omega))$ suffit pour savoir s'il y a ou non intersection.



Nous ferons les hypothèses suivantes sur les variables aléatoires Y et Φ :

(H_1) Y suit la loi uniforme sur $[0, \frac{a}{2}]$.

9. C'est aussi un excellent problème de révision.

10. Georges Louis Leclerc, comte de Buffon (1707–1788), auteur de l'*Histoire naturelle*, organisateur du Jardin des Plantes de Paris.

11. On pourrait aussi utiliser les angles de droites, Φ serait alors à valeurs dans un intervalle de longueur π.

(H_2) Φ suit la loi uniforme sur $[0, 2\pi]$.

(H_3) Y et Φ sont indépendantes.

On note E l'événement « l'aiguille coupe l'une des droites du réseau ». La longueur de la projection de la demi-aiguille sur une droite orthogonale au réseau est $Z = \frac{\ell}{2}|\sin\Phi|$. Il y a donc intersection si et seulement si la distance Y du centre de l'aiguille à la droite du réseau la plus proche est inférieure ou égale à Z. Ceci nous permet d'écrire l'évènement E sous la forme :

$$E = \left\{ Y \leqslant \frac{\ell}{2}|\sin\Phi| \right\}.$$

Comme Y et Φ sont indépendantes, la loi du couple est le produit des lois marginales : $P_{(Y,\Phi)} = P_Y \otimes P_\Phi$. Comme ces lois marginales sont à densité sur $\mathbb{R}$, on en déduit que $P_{(Y,\Phi)}$ est à densité $f_Y \otimes f_\Phi$ sur $\mathbb{R}^2$, d'où

$$f_{(Y,\Phi)}(y,t) = f_Y(y) f_\Phi(t) = \frac{2}{a}\mathbf{1}_{[0,a/2]}(y)\frac{1}{2\pi}\mathbf{1}_{[0,2\pi]}(t) = \frac{1}{a\pi}\mathbf{1}_{[0,a/2]\times[0,2\pi]}(y,t).$$

On voit ainsi que le couple (Y, Φ) suit la loi uniforme sur le rectangle $[0, a/2] \times [0, 2\pi]$.

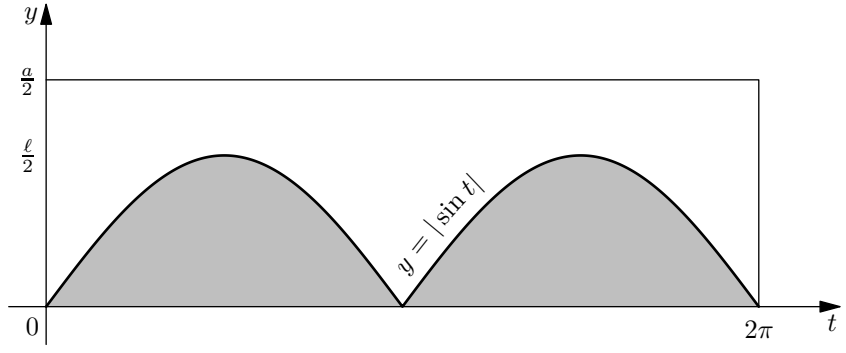
Notons D le borélien de $\mathbb{R}^2$ défini par

$$D := \left\{ (t,y) \in [0,2\pi] \times [0,a/2] \ ; \ y \leqslant \frac{\ell}{2}|\sin t| \right\}.$$

Comme l'évènement E s'écrit aussi $\{(\Phi, Y) \in D\}$, on peut calculer $P(E)$ en utilisant la loi du couple (Φ, Y) qui est la loi uniforme sur $[0, 2\pi] \times [0, a/2]$:

$$\begin{aligned} P(E) = P\big((\Phi,Y) \in D\big) &= P_{(\Phi,Y)}(D) \\ &= \frac{1}{a\pi}\lambda_2\big(D \cap [0,2\pi] \times [0,a/2]\big) \\ &= \frac{1}{a\pi}\lambda_2(D), \end{aligned}$$

en remarquant que $D \subset [0, 2\pi] \times [0, a/2]$. Le calcul de $P(E)$ se réduit ainsi à celui de l'aire de l'hypographe de la fonction $g : t \mapsto \frac{\ell}{2}|\sin t|\mathbf{1}_{[0,2\pi]}(t)$.



Par conséquent

$$\lambda_2(D) = \int_0^{2\pi} \frac{\ell}{2} |\sin t| \, \mathrm{d}t = \ell \int_0^{\pi} \sin t \, \mathrm{d}t = 2\ell.$$

Finalement,

$$P(E) = \frac{2\ell}{a\pi}.$$

On effectue une suite de lancers de l'aiguille et on note E_i l'évènement « lors du i^{e} lancer, l'aiguille intersecte une des droites du réseau ». On pose $X_i = \mathbf{1}_{E_i}$ et

$$F_n := \frac{1}{n} \sum_{i=1}^{n} X_i.$$

Les X_i sont des variables aléatoires de Bernoulli, de paramètre $p := P(E_i) = P(E)$. Elles sont clairement intégrables, puisque bornées. Les E_i forment une suite d'évènements mutuellement indépendants et de même probabilité p. Il en résulte que les X_i forment une suite de variables aléatoires indépendantes et de même loi $\mathrm{Bern}(p)$. Par la loi forte des grands nombres pour des variables i.i.d. et intégrables,

$$F_n := \frac{1}{n} \sum_{i=1}^{n} X_i \xrightarrow[n\to+\infty]{\text{p.s.}} \mathbf{E}X_1 = P(E).$$

Compte-tenu du calcul de $P(E)$, on peut réécrire ce résultat sous la forme :

$$\frac{2\ell}{aF_n} \xrightarrow[n\to+\infty]{\text{p.s.}} \pi.$$

L'interprétation physique est la suivante. Si on réalise *une* série de lancers avec n grand, la valeur $F_n(\omega)$ observée nous fournira l'approximation

$$\frac{2\ell}{aF_n(\omega)} \simeq \pi.$$

On considère ainsi qu'il est physiquement impossible d'observer un ω n'appartenant pas à l'évènement $\{F_n \text{ converge vers } p\}$ qui est de probabilité 1.

Le document de la page 355 représente les résultats de 1 200 lancers réalisés avec une allumette et un réseau tracé sur une feuille de format A4. On a ici $\ell = a = 4{,}5$ cm et $p = \frac{2}{\pi} \simeq 0{,}637$. Les lancers sont regroupés par dizaine. Les bits 0 ou 1 sont les valeurs observées pour $X_i(\omega)$. Après chaque dizaine on a noté le nombre d'intersections observées sur la dizaine et le nombre d'intersections cumulé depuis le début des lancers. La table 9.1 présente les fréquences observées F_{10k} pour $k \in [\![1, 120]\!]$.

Cette expérience permet de proposer l'estimation :

$$\pi \simeq \frac{2}{0,627} \simeq 3{,}189\,8.$$

Si vous avez la patience et le loisir de réaliser votre propre expérience, vous trouverez probablement une valeur légèrement différente... Bien entendu, cette méthode pour calculer π n'est pas très performante. On peut montrer que sa vitesse de convergence est en $O(n^{-1/2})$. Son intérêt est essentiellement d'ordre culturel et historique.

Résultats de 1200 lancers

0111110001	6	6	1111111011	9	69	1011111110	8	131
0101111011	7	13	1110100111	7	76	1001110111	7	138
0011100110	5	18	0111100010	5	81	1101101110	7	145
1110011101	7	25	1110111100	7	88	1001011110	6	151
1110000011	5	30	0000100101	3	91	0111110001	6	157
0001111001	5	35	0001111011	6	97	1101001001	5	162
1101000101	5	40	1110011001	6	103	1101001100	5	167
1101111000	6	46	1011111110	8	111	1110100101	6	173
1111011111	9	55	1101010101	6	117	0110001100	4	177
1000100111	5	60	1100100111	6	123	1011010110	6	183
1011101011	7	190	0111110011	7	251	1101100110	6	315
1101100010	5	195	0110110111	7	258	1011011101	7	322
1100000111	5	200	1011110010	6	264	1100101111	7	329
0001111111	7	207	0000111001	4	268	1111010010	6	335
0010001101	4	211	0111100101	6	274	1110001111	7	342
0101001011	5	216	1111111101	9	283	0101001111	6	348
0100111011	6	222	1101010101	6	289	1001100101	5	353
0111111101	8	230	0111111110	8	297	1101111001	7	360
1111111010	8	238	0011101000	4	301	1010111111	8	368
1110001011	6	244	0111111101	8	309	0111101000	5	373
1001110011	6	379	1011011100	6	444	0111111110	8	518
1111011111	9	388	1011111010	7	451	0100010010	3	521
0110011111	7	395	1011110111	8	459	0101101011	6	527
1000011101	5	400	0011001111	6	465	0011111011	7	534
1111100111	8	408	1000001101	4	469	1100100101	5	539
1010010001	4	412	1011011111	8	477	1010110110	6	545
0100001110	4	416	1111111111	10	487	1111101101	8	553
1111011010	7	423	1011111101	8	495	1110111101	8	561
1010111110	7	430	1011111101	8	503	1110110110	7	568
1101011111	8	438	1110011101	7	510	0111001101	6	574
1000010110	4	578	1101001001	5	634	1100110011	6	690
0110110110	6	584	1110111111	9	643	1010111101	7	697
1000101100	4	588	0110110110	6	649	1010111110	7	704
1011011101	7	595	1000001000	2	651	1111111111	10	714
1100110110	6	601	1000011000	3	654	0011011110	6	720
1111011010	7	608	0111001011	6	660	0111111110	8	728
0001000110	3	611	1100010100	4	664	1001111101	7	735
0001110001	4	615	0101110110	6	670	0100111101	6	741
1111100111	8	623	1111111001	8	678	0100110101	5	746
0010111110	6	629	0000111111	6	684	1110101010	6	752

$10k$	0	10	20	30	40	50	60	70	80	90
		0,600	0,650	0,600	0,625	0,600	0,583	0,571	0,575	0,611
100	0,600	0,627	0,633	0,623	0,629	0,607	0,606	0,606	0,617	0,616
200	0,615	0,624	0,627	0,630	0,629	0,628	0,623	0,619	0,618	0,610
300	0,610	0,613	0,609	0,606	0,609	0,603	0,600	0,600	0,605	0,610
400	0,610	0,612	0,614	0,614	0,609	0,609	0,615	0,615	0,619	0,614
500	0,618	0,618	0,619	0,621	0,620	0,622	0,621	0,619	0,621	0,624
600	0,622	0,621	0,626	0,627	0,625	0,628	0,624	0,621	0,622	0,623
700	0,626	0,625	0,626	0,629	0,628	0,625	0,628	0,632	0,635	0,637
800	0,637	0,640	0,635	0,635	0,636	0,634	0,634	0,636	0,637	0,638
900	0,638	0,635	0,635	0,632	0,633	0,633	0,633	0,630	0,628	0,629
1000	0,629	0,628	0,630	0,630	0,626	0,623	0,623	0,621	0,620	0,622
1100	0,622	0,622	0,622	0,623	0,626	0,626	0,628	0,628	0,628	0,627
1200	0,627									

TABLE 9.1 – Tableau des fréquences observées

9.3 Exercices

Ex. 9.1 *Unicité de la limite en probabilité*

On veut vérifier que la limite en probabilité est unique modulo l'égalité presque-sûre. Pour cela on supposera que Y_n converge en probabilité vers Y *et* vers Y'.

1) Justifier pour tout $\varepsilon > 0$ l'inégalité :

$$P(|Y - Y'| > \varepsilon) \leqslant P\Big(|Y - Y_n| > \frac{\varepsilon}{2}\Big) + P\Big(|Y_n - Y'| > \frac{\varepsilon}{2}\Big).$$

2) En déduire la valeur de $P(|Y - Y'| > \varepsilon)$ et conclure.

Ex. 9.2 *Convergence en probabilité et opérations*

Soient $(X_n)_{n\geqslant 1}$, $(Y_n)_{n\geqslant 1}$, X et Y des variables aléatoires définies sur le même espace probabilisé $(\Omega, \mathcal{F}, P)$. On suppose que X_n et Y_n convergent en probabilité vers X et Y respectivement.

1) Montrer que $X_n + Y_n$ converge en probabilité vers $X + Y$.

2) Montrer que X_nY_n converge en probabilité vers XY. *Indication* : commencer par montrer que pour tout $\delta > 0$, il existe un réel c et un entier n_0 tels que pour tout $n \geqslant n_0$, $P(|Y_n| > c) < \delta$.

Ex. 9.3 *Convergence en probabilité et bornitude stochastique*

Montrez que si Y_n converge en probabilité vers une variable aléatoire Y, alors la suite $(Y_n)_{n\geqslant 1}$ est *stochastiquement bornée*, c'est-à-dire :

$$\forall \varepsilon > 0,\ \exists c > 0,\ \forall n \geqslant 1, \quad P(|Y_n| > c) < \varepsilon.$$

Ex. 9.4 *Convergences p.s. ou en probabilité et image continue*

Soient $(Y_n)_{n\geqslant 1}$ et Y des variables aléatoires définies sur le même espace probabilisé.

1) On suppose que Y_n converge en probabilité vers une constante c et que g est une fonction définie au voisinage de c et continue au point c. Montrez que $g(Y_n)$ converge en probabilité vers $g(c)$. On commencera par préciser la définition de $g(Y_n)$.

2) Même question pour la convergence presque-sûre.

3) On suppose maintenant que Y_n converge en probabilité vers Y et que g est continue sur $\mathbb{R}$. Montrez que $g(Y_n)$ converge en probabilité vers $g(Y)$. *Indication* : on pourra commencer par examiner le cas plus simple où g est *uniformément* continue sur $\mathbb{R}$. Pour passer au cas général, on pourra utiliser la bornitude stochastique.

4) Même question pour la convergence presque-sûre.

Ex. 9.5 *Convergence en probabilité de gaussiennes*

Soit $(Y_n)_{n\geqslant 1}$ une suite de variables aléatoires définies sur le même espace probabilisé, gaussiennes de loi $\mathfrak{N}(c, \sigma_n)$, où la constante c ne dépend pas de n. Montrez que Y_n converge en probabilité vers c si et seulement si σ_n tend vers 0.

Ex. 9.6 *Une sournoise question de sous-suites*

Trouvez l'erreur dans le raisonnement suivant. La remarque 9.9 donne un exemple de suite $(Y_n)_{n\geqslant 1}$ qui converge en probabilité vers 0 et telle que pour tout $\omega \in \Omega$, $\liminf Y_n(\omega) = 0$ et $\limsup Y_n(\omega) = 1$. Or la convergence en probabilité passe aux sous-suites et la convergence presque-sûre implique la convergence en probabilité. On obtient alors une contradiction pour la sous-suite de (Y_n) qui converge p.s. vers 1 avec l'unicité de la limite en probabilité modulo l'égalité p.s.

Ex. 9.7 *Hasard et compensations exactes*

On considère une épreuve ayant r issues élémentaires équiprobables ($r = 2$ pour le lancer d'une pièce, $r = 6$ pour celui d'un dé, ...). On répète cette épreuve dans des conditions identiques. On note A_n l'événement : *au cours des nr premières épreuves, chacune des r issues distinctes se produit exactement n fois*. On dira que A_n est une *compensation exacte*.

1) Calculer $p_n = P(A_n)$.

2) Donner un équivalent de p_n quand n tend vers l'infini en utilisant la formule de Stirling.

3) En déduire que si $r \geqslant 4$, presque sûrement il n'y aura plus jamais de compensation exacte au delà d'un certain nombre d'épreuves.

Ex. 9.8 *Sur le max de gaussiennes*

1) Soit X une variable aléatoire gaussienne de loi $\mathfrak{N}(0, 1)$. Montrez que

$$\forall t > 0, \quad P(|X| \geqslant t) \leqslant \exp\Big(\frac{-t^2}{2}\Big). \tag{9.40}$$

Indication : partir de $P(X \geqslant t) = \int_t^{+\infty} (2\pi)^{-1/2} \exp(-x^2/2)\,\mathrm{d}x$ et ramener cette intégrale à $\int_0^{+\infty}$ par changement de variable.

2) Soit $(X_k)_{k\geqslant 1}$ une suite de variables aléatoires gaussiennes de même loi $\mathfrak{N}(0, 1)$.

Utilisez (9.40) pour majorer $P(\max_{k\leqslant n}|X_k|\geqslant t)$ pour $t>0$ et en déduire que :

$$\forall s>\sqrt{2},\quad P\Big(\max_{k\leqslant n}|X_k|\geqslant s(\ln n)^{1/2}\Big)\xrightarrow[n\to+\infty]{}0. \tag{9.41}$$

3) Proposez une constante c, la plus petite que vous puissiez trouver parmi les décimaux à trois chiffres après la virgule, telle que

$$P\Big(\max_{k\leqslant 10^6}|X_k|<c\Big)\geqslant 0{,}999\,999.$$

4) On suppose en plus que les X_k sont indépendantes. Montrez que :

$$\forall t>0,\quad P\Big(\max_{k\leqslant n}|X_k|\geqslant t\Big)\xrightarrow[n\to+\infty]{}1.$$

Donnez un exemple simple où les X_k ne sont pas indépendantes et où cette convergence vers 1 n'a pas lieu.

Ex. 9.9 Soit $(X_k)_{k\geqslant 1}$ une suite de variables aléatoires définies sur le même espace probabilisé, indépendantes et de même loi. Montrez que

$$Y_n=\frac{1}{n}\sum_{k=1}^{n}\frac{X_k}{1+|X_k|}$$

converge presque-sûrement quand n tend vers l'infini vers une constante $c\in[-1,1]$.

Ex. 9.10 *Tir à l'arc (suite)* ↑ 8.6

Les notations sont celles de l'exercice 8.6. Calculer $\mathbf{E}R$ et exprimer σ en fonction de $\mathbf{E}R$. En déduire une méthode pour *estimer* σ à partir de l'observation d'un grand nombre n de tirs du même archer réalisés dans les mêmes conditions.

Solution p. 489

Ex. 9.11 *Escargot aléatoire*

Soit $(X_i)_{i\in\mathbb{N}*}$ une suite de variables aléatoires, définies sur le même espace probabilisé $(\Omega,\mathcal{F},P)$, indépendantes et de même loi, de carré intégrable ($\mathbf{E}X_1^2<+\infty$). On définit par récurrence la suite de variables aléatoires positives $(R_n)_{n\geqslant 1}$ par

$$R_1=|X_1|,\quad R_n=\sqrt{R_{n-1}^2+X_n^2},\quad n\geqslant 2.$$

1) Montrer que la suite $(n^{-1/2}R_n)_{n\geqslant 1}$ converge presque sûrement vers une limite constante τ que l'on exprimera en fonction de $\mathbf{E}X_1^2$.

2) Pour illustrer graphiquement cette convergence, on prend les X_i de même loi uniforme sur $[0,1]$ et on construit dans un repère orthonormé du plan la suite $(M_n)_{n\geqslant 1}$ de points aléatoires par la récurrence suivante. On pose d'abord $M_1=(X_1,0)$, puis connaissant M_{n-1} on obtient M_n comme l'unique point tel que $M_{n-1}M_n=X_n$ et que l'angle $(\overrightarrow{M_{n-1}M_n},\overrightarrow{M_{n-1}O})$ ait pour mesure $+\pi/2$. En d'autre termes, on construit M_n tel que le triangle $OM_{n-1}M_n$ soit rectangle en M_{n-1}, de côtés de l'angle droit

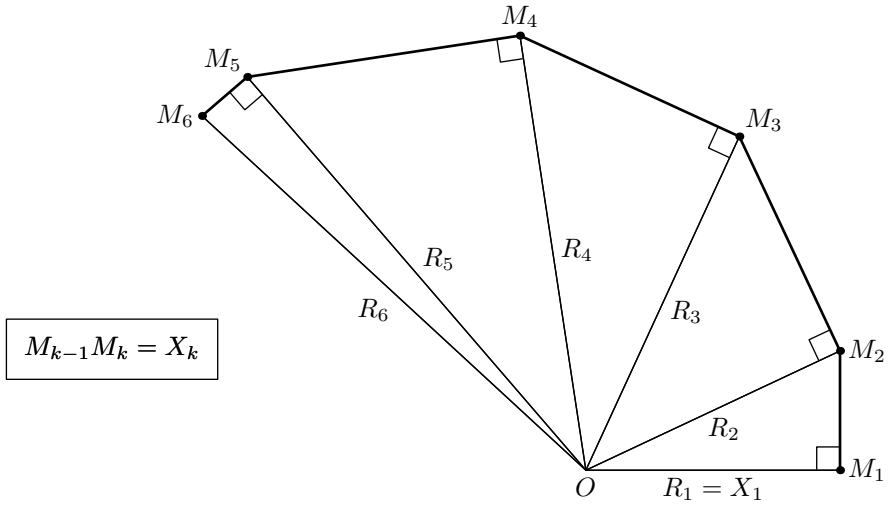


FIGURE 9.3 – Construction de l'escargot aléatoire $\mathcal{E}_6$

$OM_{n-1} = R_{n-1}$ et $M_{n-1}M_n = X_n$ et en tournant toujours dans le sens trigonométrique. On trace ainsi la ligne polygonale $\mathcal{E}_n$ de sommets $M_1, M_2, \ldots, M_n$ (l'escargot aléatoire, cf. figure 9.3).

On fixe un $\varepsilon > 0$. Montrer que presque sûrement, $\mathcal{E}_n$ est inclus dans le disque de centre O et de rayon $(1+\varepsilon)\sqrt{n/3}$ pour tout n assez grand (cf. figure 9.4 p. 360).

Solution p. 490

Ex. 9.12 *Étendue d'un échantillon*

Si $X_1, \ldots, X_n$ est un échantillon, c'est-à-dire une suite finie de variables aléatoires indépendantes et de même loi. On appelle *étendue* de l'échantillon la variable aléatoire :

$$T_n := \max_{1 \leqslant i \leqslant n} X_i - \min_{1 \leqslant i \leqslant n} X_i.$$

Le but de cet exercice est de démontrer que si $\mathbf{E}|X_1| < +\infty$, $n^{-1}T_n$ converge vers 0 en probabilité. Autrement dit, l'étendue est un $o(n)$ en probabilité.

1) Dans cette question, on suppose les X_i de *même loi*, mais pas forcément indépendantes. Montrez que si $\mathbf{E}|X_1| < +\infty$,

$$\frac{1}{n} \max_{1 \leqslant i \leqslant n} |X_i| \xrightarrow[n \to +\infty]{\Pr} 0. \tag{9.42}$$

Indications : posez $t = n\varepsilon$, décomposez l'évènement $\{\max_{1 \leqslant i \leqslant n} |X_i| \geqslant t\}$ en une union d'évènements dont chacun ne fait intervenir qu'une seule des variables X_k, majorez sa probabilité et utilisez l'inégalité de Markov raffinée.

2) Proposez une majoration simple de l'étendue faisant intervenir $\max_{1 \leqslant i \leqslant n} |X_i|$ et concluez.

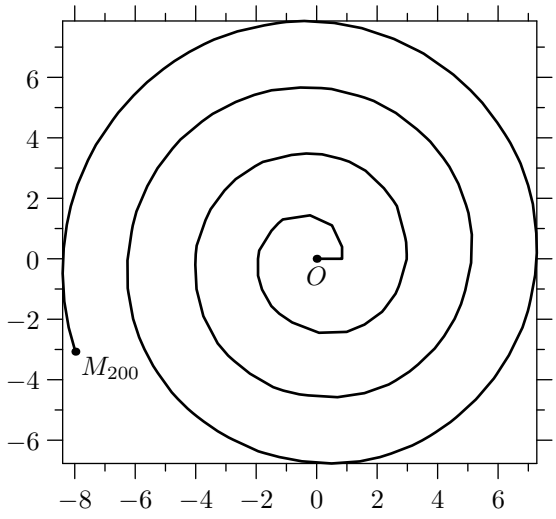


FIGURE 9.4 – Escargot aléatoire $\mathcal{E}_{200}$ (simulation)

Ex. 9.13 *Les ex-aequo d'un échantillon*

1) On se propose de montrer que si X et Y sont deux variables aléatoires réelles *indépendantes* de fonctions de répartition respectives F et G et si F est *continue*, alors $P(X = Y) = 0$. Voici quelques indications. On note pour $M \in \mathbb{N}^*$,

$$\Delta_M := \big\{(x,y) \in]-M,M]^2 \ ; \ x = y\big\} \quad \text{et} \quad \Delta := \{(x,y) \in \mathbb{R}^2 \ ; \ x = y\}.$$

Il s'agit de prouver que $P_{(X,Y)}(\Delta) = 0$ (pourquoi ?). En remarquant [12] que pour tout $k \in \mathbb{N}^*$,

$$\Delta_M \subset \bigcup_{-Mk \leqslant j < Mk} \Big]\frac{j}{k}, \frac{j+1}{k}\Big]^2,$$

utiliser la continuité *uniforme* de F sur le *compact* $[-M,M]$ pour montrer que

$$\forall \varepsilon > 0, \quad P_{(X,Y)}(\Delta_M) \leqslant \varepsilon\big(G(M) - G(-M)\big).$$

En déduire que $P_{(X,Y)}(\Delta_M) = 0$, puis que $P_{(X,Y)}(\Delta) = 0$.

2) Montrez que si $X_1, \ldots, X_n$ est un échantillon de la loi de f.d.r. F continue, autrement dit si les X_i sont indépendantes et de même loi de f.d.r. F continue, alors presque sûrement il n'y a pas d'*ex-aequo* dans l'échantillon (on observe donc n valeurs distinctes).

3) On suppose que la f.d.r. F est discontinue au point x_0 et on note

$$N_n(x_0) := \operatorname{card}\{i \in [\![1,n]\!] \ ; \ X_i = x_0\} = \sum_{i=1}^{n} \mathbf{1}_{\{X_i = x_0\}}.$$

12. Dessin impératif !

Montrez que $N_n(x_0)$ tend presque-sûrement vers l'infini avec n. Vous devriez pouvoir donner un résultat plus précis que cette affirmation.

Ex. 9.14 *Une série de produits d'uniformes*

Toutes les variables aléatoires considérées dans cet exercice sont définies sur le même espace probabilisé $(\Omega, \mathcal{F}, P)$.

1) Soit U une variable aléatoire de loi uniforme sur $[0,1]$ et $X := \ln U$. Trouver la fonction de répartition et la densité de la variable aléatoire X. Calculer $\mathbf{E}X$.

2) Soit $(U_k)_{k\in\mathbb{N}*}$ une suite de variables aléatoires réelles indépendantes et de même loi uniforme sur $[0,1]$. On note

$$T_1 := U_1,\ T_2 := \sqrt{U_1U_2}, \ldots\ldots,\ T_n := (U_1U_2\ldots U_n)^{1/n}.$$

En utilisant la loi forte des grands nombres, montrer que T_n converge presque sûrement vers $\exp(\mathbf{E}X)$.

3) En déduire que la série $\sum_{n=1}^{+\infty} U_1U_2\ldots U_n$ converge presque sûrement. *Indication* : on pourra comparer avec une série géométrique en remarquant que $\mathbf{E}X < 0$.

4) Soit S la somme de cette série. Calculer $\mathbf{E}S$ (en justifiant votre réponse).

Ex. 9.15 *Convergences p.s. de produits*

1) Soit U une variable aléatoire de loi uniforme sur $]0,1[$. On supposera pour simplifier que U est à valeurs dans $]0,1[$: pour tout $\omega \in \Omega$, $0 < U(\omega) < 1$. On pose

$$X := -\ln U.$$

Calculez la fonction de survie G de X, donnée par

$$\forall x \in \mathbb{R}, \quad G(x) = P(X > x).$$

En déduire que X a pour espérance 1. Quel nom porte la loi de X ?

2) Soit $(U_i)_{i\geqslant 1}$ une suite de variables aléatoires définies sur le même espace probabilisé $(\Omega, \mathcal{F}, P)$, à valeurs dans $]0,1[$, indépendantes et de même loi uniforme sur $]0,1[$. Pour tout $n \geqslant 1$, on note M_n la *moyenne géométrique* de $U_1, \ldots, U_n$, définie par

$$M_n := \left(\prod_{i=1}^{n} U_i\right)^{1/n}.$$

Montrez que M_n converge presque-sûrement vers e^{-1} quand n tend vers l'infini.

Dans la suite de l'exercice, on propose deux méthodes pour montrer la convergence p.s. vers 0 *de*

$$T_n := \prod_{i=1}^{n} U_i = M_n^n.$$

3) Grâce à la question 2, montrez que presque-sûrement à partir d'un certain rang (aléatoire), $M_n < 2/\mathrm{e}$. En déduire que T_n converge p.s. vers 0.

4) Voici une deuxième méthode qui n'utilise pas la question 2.

a) Montrez par un argument d'analyse élémentaire que pour tout $\omega \in \Omega$, la suite de nombres réels $(T_n(\omega))_{n\geqslant 1}$ converge dans $[0,1[$. Expliquez en une ligne pourquoi ceci permet de définir une *variable aléatoire* T en posant

$$\forall \omega \in \Omega, \quad T(\omega) := \lim_{n\to+\infty} T_n(\omega).$$

b) Calculez $\mathbf{E}T_n$ et donnez sa limite quand n tend vers l'infini. En déduire la convergence en probabilité de T_n vers 0.

c) Conclure en utilisant l'unicité de la limite en probabilité modulo l'égalité presque-sûre (on ne vous demande pas de justifier cette unicité).

Ex. 9.16 *Une convergence L^1*

Soit $(X_k)_{k\geqslant 1}$ une suite de variables aléatoires réelles définies sur le même espace probabilisé, indépendantes et de même loi et telle que $\mathbf{E}|X_1| < +\infty$. Pour tout $n \geqslant 1$, on pose $S_n := \sum_{k=1}^n X_k$. On note $h : \mathbb{R} \to \mathbb{R}$ une fonction continue bornée. Montrez que

$$\lim_{n\to+\infty} \mathbf{E}\left|h\big(n^{-1}S_n\big) - h\big(\mathbf{E}X_1\big)\right| = 0.$$

Indication p. 492

Ex. 9.17 *Loi de Cauchy et LFGN*

Le but de cet exercice est d'étudier la convergence des moyennes arithmétiques d'une suite de v.a. indépendantes et de même loi de Cauchy. Toutes les variables aléatoires intervenant dans l'énoncé sont supposées définies sur le même espace probabilisé $(\Omega, \mathcal{F}, P)$.

1) Soit X une variable aléatoire réelle suivant la loi de Cauchy de densité :

$$f : \mathbb{R} \to \mathbb{R}_+, \quad t \mapsto \frac{1}{\pi(1+t^2)}.$$

Pour tout entier $n \geqslant 1$, exprimez $P(X > n)$ à l'aide d'une intégrale. Déduisez-en la minoration :

$$\forall n \geqslant 1, \quad P(X > n) \geqslant \frac{1}{2\pi n}.$$

2) Soit $(X_k)_{k\geqslant 1}$ une suite de variables aléatoires indépendantes et de même loi que X. On pose

$$A := \{\omega \in \Omega \,;\, X_k(\omega) > k \text{ pour une infinité d'indices } k\}.$$

Expliquez pourquoi $P(A) = 1$.

3) Déduire de ce qui précède que la suite $(X_k)_{k\geqslant 1}$ ne vérifie pas la loi forte des grands nombres. *Indication* : en posant comme d'habitude $S_n := X_1 + \cdots + X_n$, exprimez $\frac{X_n}{n}$ en fonction de $\frac{S_n}{n}$ et $\frac{S_{n-1}}{n-1}$ et raisonnez par l'absurde en supposant que $\frac{S_n}{n}$ converge p.s. vers une certaine variable aléatoire Y (pas forcément constante).

4) Ce résultat est-il contradictoire avec la loi forte des grands nombres (th. 9.24) ?

Solution p. 492

Ex. 9.18 *Loi forte des grands nombres dans $\mathbb{R}^d$*

Tout est dans le titre, il ne reste plus qu'à l'énoncer et à la démontrer.

Chapitre 10

Théorème limite central

Le théorème limite central nous dit qu'une somme d'un grand nombre de variables aléatoires indépendantes, de carré intégrable, convenablement normalisée, se comporte asymptotiquement « en loi » comme une variable aléatoire gaussienne. Il explique l'importance centrale des lois gaussiennes dans la théorie des probabilités et la statistique. Il complète la loi des grands nombres en donnant une sorte de vitesse de convergence, permettant notamment de construire des « intervalles de confiance » pour l'estimation d'un paramètre.

Pour donner un sens mathématique précis à cette notion de « comportement asymptotique en loi », il nous faut d'abord introduire la convergence en loi.

10.1 Convergence en loi

Nous admettrons l'équivalence des deux définitions suivantes.

Définition 10.1 (convergence en loi)**.** *Notons F_n et F les fonctions de répartition respectives des variables aléatoires réelles Y_n $(n \geqslant 1)$ et Y. On dit que la suite $(Y_n)_{n\geqslant 1}$* converge en loi *vers Y si*

$$\text{pour tout } x \text{ point de continuité de } F, \quad F_n(x) \xrightarrow[n\to+\infty]{} F(x). \tag{10.1}$$

Rappelons que x est point de continuité de la fonction de répartition F si et seulement si $F(x-) = F(x)$ ou encore $P(Y = x) = 0$.

Définition 10.2 (convergence en loi)**.** *On dit que la suite $(Y_n)_{n\geqslant 1}$ de variables aléatoires réelles* converge en loi *vers la variable aléatoire réelle Y si*

$$\text{pour toute } h \text{ continue bornée } \mathbb{R} \to \mathbb{R}, \quad \mathbf{E}h(Y_n) \xrightarrow[n\to+\infty]{} \mathbf{E}h(Y). \tag{10.2}$$

Remarquons que si h est continue bornée, les $h(Y_n)$ et $h(Y)$ sont des v.a. bornées, donc intégrables. Nous noterons la convergence en loi de Y_n vers Y par

$$Y_n \xrightarrow[n\to+\infty]{\text{loi}} Y.$$

La définition 10.1 est la plus concrète, surtout lorsque F est continue sur tout $\mathbb{R}$, cas souvent rencontré en pratique. En effet dans ce cas, la convergence en loi équivaut à la convergence simple sur $\mathbb{R}$ des fonctions de répartition et nous donne, pour tous réels $a < b$, la convergence des $P(Y_n \in I(a,b))$ vers les $P(Y \in I(a,b))$, où $I(a,b)$ désigne n'importe lequel des 4 intervalles d'extrémités a et b.

La définition 10.2 est souvent plus commode pour établir les propriétés de la convergence en loi et a l'intérêt d'une généralisation immédiate aux vecteurs aléatoires de $\mathbb{R}^d$.

Définition 10.3 (convergence en loi de vecteurs aléatoires)**.** *On dit que la suite $(Y_n)_{n\geqslant 1}$ de vecteurs aléatoires de $\mathbb{R}^d$* converge en loi *vers le vecteur aléatoire Y de $\mathbb{R}^d$ si*

$$\text{pour toute } h \text{ continue bornée } \mathbb{R}^d \to \mathbb{R}, \quad \mathbf{E}h(Y_n) \xrightarrow[n\to+\infty]{} \mathbf{E}h(Y). \tag{10.3}$$

Remarques 10.4 (les pièges de la convergence en loi)**.** Pointons d'emblée des différences importantes entre la convergence en loi et les autres modes de convergence vus jusqu'ici.

1. Il n'est pas nécessaire, pour la convergence en loi de Y_n vers Y, que ces variables aléatoires soient définies *sur le même* $(\Omega, \mathcal{F}, P)$.
2. Il n'y a pas unicité de la v.a. limite en loi. Si $(Y_n)_{n\geqslant 1}$ converge en loi vers Y, elle converge aussi en loi vers *n'importe quelle variable aléatoire Z* ayant *même loi* que Y (éventuellement définie sur un autre espace probabilisé). Ceci se voit facilement sur chacune des deux définitions de la convergence en loi [1]. Réciproquement si Y_n converge en loi vers Y et aussi vers Z, alors Y et Z ont même loi. En effet en utilisant la définition 10.2 et l'unicité de la limite d'une suite convergente de réels, on voit que $\mathbf{E}h(Y) = \mathbf{E}h(Z)$ pour toute $h : \mathbb{R} \to \mathbb{R}$ continue bornée. Par la caractérisation des lois par leurs h-moments, cf. la proposition 8.16 avec $d = 1$, on en déduit que Y et Z ont même loi. En résumé, s'il n'y a pas unicité de la v.a. limite en loi, il y a unicité de sa loi, que l'on appelera *loi limite* [2].
3. La convergence en loi n'est pas compatible avec l'addition. Si X_n converge en loi vers X et si Y_n converge en loi vers Y, il est faux en général que $X_n + Y_n$ converge

1. Cette non-unicité de la limite est bien plus générale que pour les autres modes de convergence vus jusqu'ici où l'on avait convergence vers n'importe quelle Z égale p.s. à Y. Bien sûr, si Y et Z sont définies sur le même espace et sont égales p.s., elles ont même loi, mais la réciproque est grossièrement fausse. Quand on lance deux dés, on n'est pas sûr d'obtenir un double !

2. Ceci incite à voir la convergence en loi de Y_n vers Y comme la *convergence de la loi P_{Y_n}* vers la loi P_Y. On pourrait d'ailleurs, en sortant nettement du programme de cet ouvrage, donner un sens mathématique précis à cette convergence, appelée *convergence étroite des mesures de probabilité* en notant que $\mathbf{E}h(Y_n)$ ne dépend que de h et de P_{Y_n}.

en loi vers $X+Y$. En effet si c'était le cas, comme X_n converge en loi vers n'importe quel X' ayant même loi que X, X_n+Y_n devrait converger aussi en loi vers $X'+Y$. Le *hic* c'est que $X+Y$ n'a pas forcément même loi que $X'+Y$.

Après ces mises en garde, voyons un exemple assez typique où la convergence en loi est le concept pertinent pour décrire le comportement asymptotique d'une suite de variables aléatoires.

Exemple 10.5 (une loi limite de records)**.** Soit $(X_k)_{k\geqslant 1}$ une suite de variables aléatoires indépendantes et de même loi avec fonction de répartition commune F. Définissons la suite de variables aléatoires « records » $(M_n)_{n\geqslant 1}$ par :

$$M_n := \max_{1\leqslant k\leqslant n} X_k, \qquad n\in\mathbb{N}^*. \tag{10.4}$$

Connaissant F, il est facile d'obtenir la fonction de répartition G_n de M_n :

$$G_n(x) = P(M_n\leqslant x) = P\big(\forall k\in[\![1,n]\!],\, X_k\leqslant x\big) = P\Big(\bigcap_{k=1}^{n}\{X_k\leqslant x\}\Big).$$

En utilisant l'indépendance des X_k, puis le fait qu'elles ont même loi, on en déduit :

$$G_n(x) = \prod_{k=1}^{n} P(X_k\leqslant x) = \big(F(x)\big)^n. \tag{10.5}$$

Supposons désormais que les X_k ont pour loi commune la loi exponentielle de paramètre a, alors

$$F(x) = 1-\mathrm{e}^{-ax} \text{ si } x\geqslant 0, \quad F(x)=0 \text{ si } x<0\,;$$
$$G_n(x) = \big(1-\mathrm{e}^{-ax}\big)^n \text{ si } x\geqslant 0, \quad G_n(x)=0 \text{ si } x<0.$$

Donc pour x réel fixé, $\lim_{n\to+\infty} G_n(x)=0$. La signification intuitive de ce résultat est que le record M_n finira par dépasser n'importe quel niveau x fixé pour n assez grand [3]. Afin de préciser cette idée, on cherche une suite non aléatoire tendant vers $+\infty$ à la même vitesse que M_n. On peut vérifier que $\mathbf{E}M_n = \frac{1}{a}\big(1+\frac{1}{2}+\frac{1}{3}+\cdots+\frac{1}{n}\big)$, donc $\mathbf{E}M_n\sim a^{-1}\ln n$, cf. par exemple le corrigé de l'exercice 7.7. Ceci nous amène à étudier le comportement asymptotique de $P(M_n - a^{-1}\ln n\leqslant x)$:

$$P\left(M_n-\frac{\ln n}{a}\leqslant x\right) = G_n\left(x+\frac{\ln n}{a}\right) = \big(1-\mathrm{e}^{-ax-\ln n}\big)^n = \left(1-\frac{\mathrm{e}^{-ax}}{n}\right)^n. \tag{10.6}$$

On en déduit que

$$\lim_{n\to+\infty} P\left(M_n-\frac{\ln n}{a}\leqslant x\right) = \exp\big(-\mathrm{e}^{-ax}\big). \tag{10.7}$$

3. N'appliquez pas cette remarque au sport, même avec dopage. Cette « convergence en probabilité vers l'infini » de M_n n'est possible que parce que chaque X_k peut elle-même prendre une valeur supérieure à x avec une probabilité non nulle. Si on prend pour X_k des variables de loi uniforme sur $[0,1]$, la suite des records restera bornée par 1.

Le calcul (10.6) est valable pour $\ln n \geqslant -ax$, donc pour tout $n \in \mathbb{N}^*$ et tout $x \geqslant 0$. Pour $x < 0$ fixé, $\ln n \geqslant -ax$ pour $n \geqslant n_0(x)$ donc (10.7) est valable pour tout x réel. On peut donc dire qu'*asymptotiquement*, M_n est de l'ordre de grandeur de $a^{-1} \ln n$ et que la dispersion aléatoire de M_n autour de cette valeur est donnée par la loi de fonction de répartition :

$$H(x) = \exp\bigl(-\mathrm{e}^{-ax}\bigr), \qquad x \in \mathbb{R}. \tag{10.8}$$

On vérifie immédiatement que H est continue sur $\mathbb{R}$, croissante (comme composée de deux fonctions décroissantes) avec pour limites 0 en $-\infty$ et 1 en $+\infty$. C'est donc bien une fonction de répartition. La loi de f.d.r. H est une loi de Gumbel.

D'après la définition 10.1, on peut reformuler la conclusion en disant que la suite de variables aléatoires $M_n - a^{-1} \ln n$ converge en loi vers une v.a. suivant la loi de Gumbel de f.d.r. H donnée par (10.8).

Une propriété bien commode de la convergence en loi est sa conservation par image continue.

Proposition 10.6 (convergence en loi par image continue)**.** *Si Y_n converge en loi vers Y, alors pour toute f continue $\mathbb{R} \to \mathbb{R}$, $f(Y_n)$ converge en loi vers $f(Y)$.*

Noter que l'on ne suppose pas f *bornée* sur $\mathbb{R}$.

Preuve. D'après la définition 10.2, il nous faut vérifier que pour toute fonction *continue bornée* $g : \mathbb{R} \to \mathbb{R}$, $\mathbf{E}g(f(Y_n))$ tend vers $\mathbf{E}g(f(Y))$ quand n tend vers l'infini. Or la fonction $g \circ f$ est continue sur $\mathbb{R}$ par composition et bornée sur $\mathbb{R}$ par $\sup_{t\in\mathbb{R}} |g(t)|$. On sait par hypothèse que $\mathbf{E}h(Y_n)$ converge vers $\mathbf{E}h(Y)$ pour toute h continue bornée sur $\mathbb{R}$. En appliquant ceci avec $h = g \circ f$, on obtient la conclusion souhaitée. □

La preuve ci-dessus se généralise immédiatement aux vecteurs aléatoires.

Proposition 10.7 (convergence en loi de vecteurs par image continue)**.** *Si les Y_n et Y sont des vecteurs aléatoires de $\mathbb{R}^d$ tels que Y_n converge en loi vers Y, alors pour toute f continue $\mathbb{R}^d \to \mathbb{R}^j$, $f(Y_n)$ converge en loi vers $f(Y)$ dans $\mathbb{R}^j$.*

En appliquant la proposition 10.7 avec pour f les projections canoniques $\pi_k : \mathbb{R}^d \to \mathbb{R}$, $x = (x_1, \ldots, x_d) \mapsto x_k$, on en déduit le corollaire suivant.

Corollaire 10.8 (convergence en loi des composantes)**.** *Si le vecteur aléatoire Y_n converge en loi dans $\mathbb{R}^d$ vers un vecteur aléatoire Y, alors chacune de ses composantes converge en loi dans $\mathbb{R}$ vers la composante correspondante de Y.*

Il importe de comprendre que la réciproque *est fausse*. Pour s'en convaincre, examinons le cas $d = 2$ en supposant que les variables aléatoires réelles X_n et Y_n convergent en loi respectivement vers X et Y. Si ces convergences en loi impliquaient celle du couple (X_n, Y_n) vers le couple (X, Y), alors en prenant l'image par l'application continue $(s, t) \mapsto s + t$, on en déduirait la convergence en loi de $X_n + Y_n$ vers $X + Y$. Ceci serait en contradiction avec la remarque 10.4-3.

Le diagramme des convergences de la figure 10.1, qui reprend et complète celui de la figure 9.2 avec les mêmes conventions, indique que la convergence en loi est la plus faible des convergences de suites de variables aléatoires. Cette affirmation se justifie par le résultat suivant.

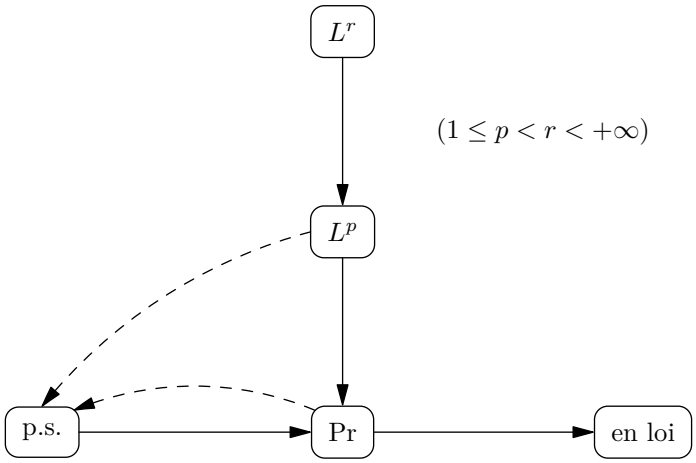


FIGURE 10.1 – Diagramme des convergences des suites de v.a.

Proposition 10.9. *La convergence en probabilité implique la convergence en loi : si les Y_n ($n \geqslant 1$) et Y sont des variables aléatoires réelles définies sur le même espace probabilisé $(\Omega, \mathcal{F}, P)$ telles que Y_n converge en probabilité vers Y, alors Y_n converge aussi en loi vers Y.*

Nous allons prouver la proposition en utilisant la définition 10.2, cette méthode ayant l'avantage de se généraliser immédiatement au cas des vecteurs aléatoires [4] de $\mathbb{R}^d$. Nous aurons besoin du lemme élémentaire d'analyse suivant.

Lemme 10.10 (convergence par sous-sous-suites)**.** *La suite de réels $(u_n)_{n\geqslant 1}$ converge vers le réel ℓ si de toute sous-suite de $(u_n)_{n\geqslant 1}$ on peut extraire une nouvelle sous-suite qui converge vers ℓ.*

Preuve. Par confort typographique, nous noterons une sous-suite de $(u_n)_{n\in\mathbb{N}*}$ comme une suite $(u_n)_{n\in A}$, où A est une partie infinie de $\mathbb{N}^*$. Une sous-suite de $(u_n)_{n\in A}$ s'écrira alors $(u_n)_{n\in B}$, pour une partie infinie B de A. La convergence de cette sous-suite sera notée :

$$u_n \xrightarrow[n\in B,\ n\to+\infty]{} \ell.$$

4. La convergence en probabilité de Y_n vers Y dans $\mathbb{R}^d$ se définit comme en dimension 1, mais en remplaçant $|Y_n - Y|$ par $\|Y_n - Y\|$ après le choix d'une norme dans $\mathbb{R}^d$. Peu importe laquelle, puisqu'en dimension finie elles sont toutes équivalentes.

L'hypothèse du lemme s'écrit donc

$$\forall A \text{ infini} \subset \mathbb{N}^*, \quad \exists B \text{ infini} \subset A, \quad u_n \xrightarrow[n\in B,\, n\to+\infty]{} \ell. \tag{10.9}$$

Supposons que $(u_n)_{n\in\mathbb{N}*}$ *ne converge pas* vers ℓ. Il existe alors un $\varepsilon > 0$ tel que

$$\forall j \in \mathbb{N}^*, \ \exists n \geqslant j, \quad |u_n - \ell| \geqslant \varepsilon. \tag{10.10}$$

Autrement dit, il existe *une infinité* d'entiers n tels que $|u_n - \ell| \geqslant \varepsilon$. Notons A l'ensemble de ces entiers. Par l'hypothèse (10.9), il existe une partie infinie B de cet ensemble A telle que la sous-suite $(u_n)_{n\in B}$ converge vers ℓ. On peut alors trouver $n \in B$ assez grand pour que $|u_n - \ell| < \varepsilon$. Mais comme cet n est aussi dans A, on aboutit à une contradiction. □

Preuve de la proposition 10.9. Par hypothèse, Y_n converge en probabilité vers Y. Soit $h : \mathbb{R} \to \mathbb{R}$ continue bornée quelconque, il s'agit de prouver que $\mathbf{E}h(Y_n)$ converge vers $\mathbf{E}h(Y)$. Nous allons utiliser pour cela le lemme 10.10 avec $u_n = \mathbf{E}h(Y_n)$ et $\ell = \mathbf{E}h(Y)$. Comme h est bornée, il existe un réel $b > 0$ tel que $h(x) \in [-b, b]$ pour tout $x \in \mathbb{R}$. On en déduit l'inégalité entre variables aléatoires $-b \leqslant h(Y_n) \leqslant b$. Soit A une partie infinie quelconque de $\mathbb{N}^*$. Puisque $(Y_n)_{n\in A}$ converge en probabilité vers Y, on peut en extraire une sous-suite $(Y_n)_{n\in B}$ qui converge p.s. vers Y. Par continuité de h, on en déduit que

$$h(Y_n) \xrightarrow[n\in B,\, n\to+\infty]{\text{p.s.}} h(Y).$$

Comme $|h(Y_n)| \leqslant b$, on en déduit par le théorème de convergence dominée, la v.a. constante b étant évidemment intégrable, que

$$\mathbf{E}h(Y_n) \xrightarrow[n\in B,\, n\to+\infty]{} \mathbf{E}h(Y).$$

Comme A était quelconque, on conclut par le lemme 10.10 que c'est *toute la suite* $(\mathbf{E}h(Y_n))_{n\in\mathbb{N}*}$ qui converge vers $\mathbf{E}h(Y)$. Ceci étant vrai pour toute fonction continue bornée h, la proposition 10.9 est démontrée. □

Remarques 10.11 (convergences en probabilité et en loi)**.**

1. La convergence en loi n'implique pas la convergence en probabilité. Voici un contre exemple qui est l'un des plus simples possibles. On prend $\Omega = \{-1, 1\}$, $\mathcal{F} = \mathcal{P}(\Omega)$ et P l'équiprobabilité sur Ω. On définit sur Ω la variable aléatoire Y_1 comme l'identité sur Ω. On construit la suite (Y_n) en posant pour $n \geqslant 2$, $Y_n = (-1)^{n-1}Y_1$. Alors la loi sous P de chaque Y_n est la loi de Rademacher : $P(Y_n = -1) = P(Y_n = 1) = 1/2$. Donc Y_n converge en loi trivialement vers Y_1 (et aussi vers $-Y_1$ qui a même loi). Mais il n'y a pas convergence en probabilité car pour tout $k \in \mathbb{N}^*$ et tout $\varepsilon \in]0, 2[$, $P(|Y_{2k} - Y_1| > \varepsilon) = P(2|Y_1| > \varepsilon) = 1$.
2. La réciproque de la proposition 10.9 est vraie dans le cas très particulier où la v.a. limite est *constante*. Supposons en effet que Y_n converge en loi vers la constante c. La fonction de répartition de la variable aléatoire c est $\mathbf{1}_{[c,+\infty[}$ qui est continue

partout sauf au point c. Désignant par F_n la fonction de répartition de Y_n, on note que pour tout $\varepsilon > 0$,

$$\begin{aligned} P(|Y_n - c| \geqslant \varepsilon) &= P(Y_n \leqslant c - \varepsilon) + P(Y_n \geqslant c + \varepsilon) \\ &\leqslant P(Y_n \leqslant c - \varepsilon) + P\Big(Y_n > c + \frac{\varepsilon}{2}\Big) \\ &= F_n(c - \varepsilon) + 1 - F_n\Big(c + \frac{\varepsilon}{2}\Big). \end{aligned}$$

La convergence de $P(|Y_n - c| \geqslant \varepsilon)$ vers 0 résulte de cette majoration puisque $F_n(x)$ converge vers $\mathbf{1}_{[c,+\infty[}(x)$ pour tout $x \neq c$.

Le lemme suivant est d'une grande utilité dans les problèmes de convergence en loi rencontrés en statistique mathématique. Pour en voir l'intérêt, rappelons que la convergence en loi de chacune des composantes d'un vecteur aléatoire n'implique pas le convergence en loi du vecteur.

Lemme 10.12 (Slutsky)**.** *Soient $(Y_n)_{n\geqslant 1}$ et $(Z_n)_{n\geqslant 1}$ deux suites de variables aléatoires définies sur le même espace probabilisé et telles que*

$$Y_n \xrightarrow[n\to+\infty]{\text{loi}} Y \quad \text{et} \quad Z_n \xrightarrow[n\to+\infty]{\Pr} c,$$

où c est une constante. Alors,

$$(Y_n, Z_n) \xrightarrow[n\to+\infty]{\text{loi}} (Y, c). \tag{10.11}$$

Nous admettons ce lemme (une preuve presque complète est donnée p. 392). En pratique, le lemme de Slutsky est utilisé le plus souvent *via* le corollaire suivant.

Corollaire 10.13 (Slutsky)**.** *Soient $(Y_n)_{n\geqslant 1}$ et $(Z_n)_{n\geqslant 1}$ deux suites de v.a. définies sur le même espace probabilisé et telles que Y_n converge en loi vers Y et Z_n converge en loi vers une constante c. Alors*

a) $Y_n + Z_n$ converge en loi vers $Y + c$;

b) $Y_n Z_n$ converge en loi vers cY.

Preuve du corollaire. Par le lemme de Slutsky, (Y_n, Z_n) converge en loi vers le vecteur aléatoire (Y, c). Par image continue, $g(Y_n, Z_n)$ converge en loi vers $g(Y, c)$ pour toute application continue $g : \mathbb{R}^2 \to \mathbb{R}$ (prop. 10.7). En choisissant $g : (s, t) \mapsto s + t$, puis $g : (s, t) \mapsto st$, on obtient successivement le a) et le b). □

10.2 Normalité asymptotique

Nous sommes maintenant en mesure d'aborder l'étude de la convergence en loi de sommes de variables aléatoires indépendantes convenablement normalisées vers une limite gaussienne. Nous examinerons diverses situations donnant lieu à chaque fois à un « théorème limite central » ou « TLC » dans le jargon probabiliste.

10.2.1 Sommes de variables aléatoires i.i.d.

Théorème 10.14 (théorème limite central, cas i.i.d.)**.** *Soit $(X_k)_{k\geqslant 1}$ une suite de variables aléatoires définies sur le même espace probabilisé $(\Omega, \mathcal{F}, P)$, indépendantes, de même loi et de carré intégrable et non* p.s. *constantes. Notons $\mu := \mathbf{E}X_1$, $\sigma^2 := \operatorname{Var} X_1$ avec $\sigma > 0$ et $S_n = \sum_{k=1}^n X_k$. Alors*

$$S_n^* := \frac{S_n - \mathbf{E}S_n}{\sqrt{\operatorname{Var} S_n}} = \frac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow[n\to+\infty]{\text{loi}} Z, \tag{10.12}$$

où Z est une variable de loi gaussienne $\mathfrak{N}(0,1)$.

Il est possible d'énoncer ce théorème de manière plus élémentaire, sans parler de convergence en loi, ni même de loi gaussienne. En exploitant la *continuité* sur $\mathbb{R}$ de la f.d.r. Φ de la loi $\mathfrak{N}(0,1)$ et la définition 10.1 de la convergence en loi, on voit en effet qu'une formulation équivalente de la conclusion (10.12) du théorème est que

$$\forall x \in \mathbb{R}, \quad P(S_n^* \leqslant x) \xrightarrow[n\to+\infty]{} \Phi(x) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left(\frac{-t^2}{2}\right) \mathrm{d}t. \tag{10.13}$$

Une conséquence pratique de (10.13) est que pour tous réels a, b, tels que $a < b$,

$$P\big(S_n^* \in I(a,b)\big) \xrightarrow[n\to+\infty]{} \Phi(b) - \Phi(a) = \frac{1}{\sqrt{2\pi}} \int_{a}^{b} \exp\left(\frac{-t^2}{2}\right) \mathrm{d}t, \tag{10.14}$$

où $I(a,b)$ est n'importe lequel des quatre intervalles d'extrémités a et b. Noter que sous cette forme on pourrait énoncer une version du théorème limite central compréhensible par un public ne connaissant que la notion d'intégrale de Riemann ordinaire [5].

Corollaire 10.15 (théorème de de Moivre-Laplace)**.** *Si S_n est une variable aléatoire de loi binomiale de paramètres n et $p \in]0,1[$, on a avec $q := 1-p$,*

$$S_n^* := \frac{S_n - np}{\sqrt{npq}} = \sqrt{\frac{n}{pq}}\left(\frac{S_n}{n} - p\right) \xrightarrow[n\to+\infty]{\text{loi}} Z,$$

où Z est une variable de loi gaussienne $\mathfrak{N}(0,1)$.

Preuve. C'est une conséquence immédiate du théorème 10.14 en remarquant que S_n a *même loi*[6] que $X_1 + \cdots + X_n$, où les X_k sont des variables aléatoires de Bernoulli indépendantes et de même paramètre p et en rappelant que l'espérance et la variance de la loi $\mathrm{Bin}(n,p)$ sont respectivement np et npq. □

5. On peut même laisser tomber la forme intégrale de la limite dans (10.14) en se contentant de dire que Φ est une fonction croissante continue que l'on a tabulée.

6. Il est clair d'après la définition de la convergence en loi que si Y_n converge en loi vers Y et si pour chaque n, Y'_n a même loi que Y_n, Y'_n converge aussi en loi vers Y.

La démonstration historique du théorème de de Moivre-Laplace repose sur un bon contrôle des coefficients binomiaux *via* la formule de Stirling. L'intérêt de cette approche « élémentaire » est de donner une idée de la vitesse de convergence qui est en $O(n^{-1/2})$, voir [14, chap. 7].

Nous admettrons le théorème 10.14 dont la démonstration n'est pas au programme (les lecteurs curieux ou « avancés » pourront consulter la section 10.4). Ce théorème a de multiples applications, notamment en statistique. À ce stade, on peut souligner deux idées.

D'abord, on peut noter que le comportement asymptotique en loi de S_n^* *ne dépend pas de la loi* de X_1. La seule condition pour que la loi de S_n^* soit approximativement gaussienne pour les grandes valeurs de n est que X_1 soit de carré intégrable. Ceci donne un caractère universel aux lois gaussiennes et explique la fréquence de l'utilisation de ces lois en modélisation [7]. On peut dire que le comportement asymptotique en loi de sommes S_n^* et donc aussi de S_n « oublie » tout de la loi des X_i, sauf le paramètre de localisation $\mu = \mathbf{E}X_1$ et le paramètre de dispersion $\sigma^2 = \operatorname{Var} X_1$. C'est l'une des raisons de l'importance donnée à ces deux paramètres en théorie des probabilités.

La deuxième idée importante est que le théorème limite central donne une idée de la vitesse de convergence dans la loi des grands nombres. *Grosso modo*, on peut dire que dans le bon cas où $\mathbf{E}X_1^2 < +\infty$, cette vitesse est en $O(n^{-1/2})$. Précisons le sens de cette affirmation. Par (10.14) appliqué avec $[a, b] = [-t, t]$, on obtient :

$$\lim_{n\to+\infty} P\big(S_n^* \in [-t, t]\big) = \Phi(t) - \Phi(-t) = 2\Phi(t) - 1, \tag{10.15}$$

en utilisant la relation $\Phi(-t) = 1 - \Phi(t)$ due à la parité de la densité de $\mathfrak{N}(0, 1)$. En remarquant maintenant que

$$S_n^* = \frac{S_n - n\mathbf{E}X_1}{\sigma\sqrt{n}} = \frac{\sqrt{n}}{\sigma}\Big(\frac{S_n}{n} - \mathbf{E}X_1\Big), \tag{10.16}$$

on peut réécrire (10.15) sous la forme

$$P\Big(\Big|\frac{S_n}{n} - \mathbf{E}X_1\Big| \leqslant \frac{\sigma t}{\sqrt{n}}\Big) = 2\Phi(t) - 1 + \varepsilon_n, \tag{10.17}$$

où ε_n est une suite de réels (pas forcément positifs), convergente vers 0. Pour tout $\delta > 0$, on peut choisir un $t = t(\delta)$ assez grand pour que $2\Phi(t)-1 > 1-\delta/2$ car $2\Phi(t)-1$ tend vers 1 quand t tend vers $+\infty$. Ensuite pour $n \geqslant n_0(\delta)$, on aura $|\varepsilon_n| < \delta/2$ et finalement

$$\forall\delta > 0,\ \exists t(\delta), n(\delta),\ \forall n \geqslant n(\delta), \quad P\Big(\Big|\frac{S_n}{n} - \mathbf{E}X_1\Big| \leqslant \frac{\sigma t(\delta)}{\sqrt{n}}\Big) > 1 - \delta. \tag{10.18}$$

C'est au sens de (10.18) que l'on peut dire que S_n/n converge vers $\mathbf{E}X_1$ avec une vitesse en $O(n^{-1/2})$. On peut résumer (10.18) par l'écriture $|n^{-1}S_n - \mathbf{E}X_1| = O_P(n^{-1/2})$, dont le deuxième membre se lit « grand O en probabilité de $n^{-1/2}$ ».

7. D'autant plus qu'il existe de nombreuses généralisations du théorème limite central, avec des v.a. indépendantes mais de lois différentes, avec des vecteurs aléatoires, avec des v.a. « faiblement dépendantes » ...

Dans l'utilisation pratique du théorème 10.14, on travaille souvent avec n « grand » fixé et on approxime la loi de S_n^* par la loi $\mathfrak{N}(0,1)$, ou ce qui revient au même, *on approxime la loi* de S_n par la *loi gaussienne* $\mathfrak{N}(n\mathbf{E}X_1, \sigma\sqrt{n})$ *de même espérance et même variance* que S_n. Plus précisément, en notant que $g_n : x \mapsto \frac{x-n\mathbf{E}X_1}{\sigma\sqrt{n}}$ est une bijection croissante de $\mathbb{R}$ sur $\mathbb{R}$ et en posant pour $a < b$ réels,

$$a_n = g_n(a) = \frac{a - n\mathbf{E}X_1}{\sigma\sqrt{n}}, \quad b_n = g_n(b) = \frac{b - n\mathbf{E}X_1}{\sigma\sqrt{n}},$$

on a

$$P(a \leqslant S_n \leqslant b) = P(a_n \leqslant S_n^* \leqslant b_n) = \Phi(b_n) - \Phi(a_n) + \varepsilon_n. \tag{10.19}$$

On néglige alors le terme d'erreur ε_n et on termine le calcul en utilisant la table des valeurs de Φ.

10.2.2 Vitesse de convergence dans le TLC

La question qui se pose dans le calcul précédent est « que signifie n *grand* ? », ou encore « comment peut-on contrôler l'erreur ε_n ? », autrement dit, quelle est la vitesse de convergence vers 0 de ε_n ? La réponse est que dans le « bon cas » où X_1 a un moment d'ordre 3, la vitesse de convergence dans le théorème limite central est en $O(n^{-1/2})$.

Théorème 10.16 (Berry-Esséen, 1941–42)**.** *Soit $(X_i)_{i\geqslant 1}$ une suite de variables aléatoires i.i.d. telle que $\mathbf{E}|X_i|^3 < +\infty$. On note $\sigma^2 := \operatorname{Var} X_1$, $\rho^3 := \mathbf{E}|X_1 - \mathbf{E}X_1|^3$, avec $\sigma > 0$ et $\rho > 0$. Il existe alors une constante universelle $C > 0$ telle que pour tout $n \geqslant 1$,*

$$\Delta_n := \sup_{x\in\mathbb{R}} \left|P(S_n^* \leqslant x) - \Phi(x)\right| \leqslant C\frac{\rho^3}{\sigma^3}\frac{1}{\sqrt{n}}.$$

Pour une preuve du théorème, voir [6]. L'obtention de la meilleure constante C a été l'objet d'une longue quête. La valeur initiale de Esséen était $C = 7{,}59$. Une valeur plus moderne et proche de l'optimale est $C = 0{,}797\,5$ (Van Beek (1972)).

Il est intéressant de regarder ce que donne le théorème de Berry-Esséen pour le cas de de Moivre-Laplace, donc avec des X_i suivant la loi de Bernoulli de paramètre p. En calculant ρ^3 on vérifie facilement que

$$\Delta_n \leqslant C\frac{p^2+q^2}{\sqrt{pq}}\frac{1}{\sqrt{n}}, \quad q := 1-p.$$

On dispose en fait de résultats plus précis d'Uspensky [16] concernant ce cas particulier, cf. [14, chap. 7].

Voici un exemple tout à fait élémentaire permettant de comprendre qu'il n'y a pas lieu d'espérer une vitesse de convergence meilleure que $O(n^{-1/2})$ pour Δ_n. Prenons X_1 de loi de Bernoulli de paramètre $1/2$. On a alors

$$S_{2n} \sim \operatorname{Bin}(2n, \frac{1}{2}), \quad \mathbf{E}S_{2n} = 2n\frac{1}{2} = n.$$

On cherche un équivalent de $P(S^*_{2n} \leqslant 0) - \Phi(0)$. Remarquons d'abord que

$$\{S^*_{2n} < 0\} = \{0 \leqslant S_{2n} < n\} \quad \text{et} \quad \{S^*_{2n} > 0\} = \{n < S_{2n} \leqslant 2n\}.$$

En raison de la symétrie des coefficients binomiaux ($C^k_{2n} = C^{2n-k}_{2n}$),

$$P(S^*_{2n} < 0) = \sum_{k=0}^{n-1} C^k_{2n} 2^{-2n} = \sum_{j=n+1}^{2n} C^j_{2n} 2^{-2n} = P(S^*_{2n} > 0).$$

On a ainsi $2P(S^*_{2n} < 0) + P(S^*_{2n} = 0) = 1$ d'où l'on tire $P(S^*_{2n} < 0) = \frac{1}{2} - \frac{1}{2}P(S^*_{2n} = 0)$ et $P(S^*_{2n} \leqslant 0) = \frac{1}{2} + \frac{1}{2}P(S^*_{2n} = 0)$. En rappelant que $\Phi(0) = \frac{1}{2}$, on aboutit à

$$P(S^*_{2n} \leqslant 0) - \Phi(0) = \frac{1}{2}P(S^*_{2n} = 0) = \frac{1}{2}P(S_{2n} = n) = C^n_{2n} 2^{-2n-1}.$$

Par la formule de Stirling ($n! \sim \sqrt{2\pi} n^{n+1/2} \mathrm{e}^{-n}$), on obtient l'équivalent

$$P(S^*_{2n} \leqslant 0) - \Phi(0) \sim \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{2n}}.$$

Comme $(2\pi)^{-1/2} > 0{,}398\,9$, on a pour $n \geqslant n_0$, $|P(S^*_{2n} \leqslant 0) - \Phi(0)| \geqslant 0{,}398(2n)^{-1/2}$, minorant à comparer avec le majorant uniforme $\Delta_{2n} \leqslant 0{,}798(2n)^{-1/2}$ fourni dans ce cas par le théorème de Berry-Esséen [8].

Revenons à la situation générale du théorème 10.16. Que se passe-t-il dans la zone intermédiaire où le théorème limite central est vérifié parce que X_1 a un moment d'ordre 2, mais X_1 n'a pas de moment d'ordre 3 ? On a toujours une vitesse de convergence, pourvu que l'intégrabilité de X_1 soit un peu plus forte que la seule existence d'un moment d'ordre 2. Le résultat précis est le suivant.

Théorème 10.17 (Katz 1963, Petrov 1965)**.** *Soit $(X_i)_{i\geqslant 1}$ une suite de variables aléatoires i.i.d. telle que $\mathbf{E}X_1 = 0$, $\mathbf{E}\big(X_1^2 g(X_1)\big) < +\infty$, où g est une fonction positive, paire, croissante sur $]0, +\infty[$ et telle que $x/g(x)$ soit croissante sur $]0, +\infty[$. Il existe alors une constante universelle $A > 0$ telle que pour tout $n \geqslant 1$,*

$$\Delta_n := \sup_{x\in\mathbb{R}} \big|P(S^*_n \leqslant x) - \Phi(x)\big| \leqslant A \frac{\mathbf{E}\big(X_1^2 g(X_1)\big)}{\sigma^2 g(\sigma\sqrt{n})}.$$

En particulier si $\mathbf{E}|X_1|^{2+\delta} < +\infty$ pour un $\delta \in]0, 1]$,

$$\Delta_n \leqslant A \frac{\mathbf{E}|X_1|^{2+\delta}}{\sigma^{2+\delta}} \frac{1}{n^{\delta/2}}.$$

8. Signalons cependant que dans le cas de la loi de Bernoulli de paramètre $p = 1/2$, il est possible d'obtenir une vitesse de convergence en $O(n^{-1})$ pour une variante de la distance Δ_n obtenue en appliquant la *correction de continuité* [14, chap. 7].

10.2.3 Intervalle de confiance pour une probabilité inconnue

Voyons maintenant une première application statistique du théorème limite central à l'estimation d'une probabilité inconnue p à partir de l'observation d'un *échantillon* de n variables de Bernoulli indépendantes de paramètre p. Par exemple on se trouve en présence d'une urne dont on sait seulement qu'elle contient des boules rouges et des vertes, la proportion de chaque couleur étant inconnue. On cherche à estimer la proportion p de boules rouges en effectuant n tirages avec remise et en notant à chaque tirage si la couleur de la boule est rouge ou non. On pose alors $X_i = 1$ si le i^{e} tirage donne une boule rouge et $X_i = 0$ sinon. $S_n = X_1 + \cdots + X_n$ est le nombre aléatoire de boules rouges sorties en n tirages et S_n/n la fréquence observée de sortie d'une boule rouge. Cette situation est comparable à un sondage avant le deuxième tour d'une élection opposant deux candidats A et B. On interroge n personnes et on pose $X_i = 1$ si la i^{e} personne interrogée annonce son intention de voter pour A. Ici S_n/n est la fréquence des intentions de votes pour A dans l'échantillon des sondés. La proportion p d'électeurs ayant l'intention de voter A dans la population totale est inconnue avant l'élection et c'est elle que l'on cherche à estimer [9].

Nous allons voir comment le théorème de de Moivre-Laplace permet de construire des *intervalles de confiance* pour p (ce que les médias appellent la « fourchette » des instituts de sondage). Considérons pour $t > 0$ l'évènement

$$A_{n,t} := \left\{ \omega \in \Omega \; ; \; -t \leqslant \sqrt{\frac{n}{pq}} \Big(\frac{S_n(\omega)}{n} - p \Big) \leqslant t \right\}.$$

Le théorème de de Moivre-Laplace nous dit que pour n assez grand, on peut utiliser l'approximation :

$$P\big(A_{n,t}\big) \simeq \Phi(t) - \Phi(-t) = 2\Phi(t) - 1.$$

Ceci peut se réécrire

$$P\left(\frac{S_n}{n} - t\sqrt{\frac{pq}{n}} \leqslant p \leqslant \frac{S_n}{n} + t\sqrt{\frac{pq}{n}} \right) = 2\Phi(t) - 1 + \varepsilon_n.$$

On ignore la valeur de p, donc a fortiori celle de $\sqrt{pq}$. Heureusement, il est possible de la majorer car $p(1-p)$ est maximal pour $p = 1/2$, cf. figure 10.2. D'où

$$\sqrt{pq} \leqslant \sqrt{\frac{1}{4}} = \frac{1}{2}, \tag{10.20}$$

de sorte qu'en notant

$$B_{n,t} := \left\{ \omega \in \Omega \; ; \; \frac{S_n(\omega)}{n} - \frac{t}{2\sqrt{n}} \leqslant p \leqslant \frac{S_n(\omega)}{n} + \frac{t}{2\sqrt{n}} \right\},$$

9. En toute rigueur, les 2 modèles ne sont pas équivalents car dans l'échantillon des sondés, on n'interroge jamais deux fois la même personne tandis que lors des tirages avec remise, la même boule peut sortir plus d'une fois. Néanmoins en raison du théorème de convergence de la loi hypergéométrique vers la loi binomiale, on considère généralement que le modèle d'urne ci-dessus est une bonne représentation du sondage.

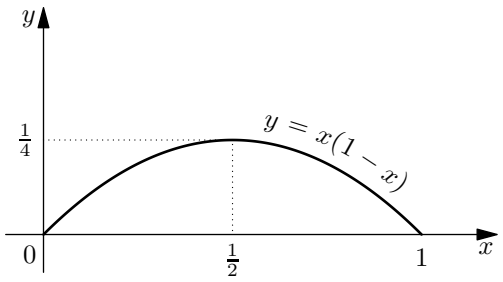


FIGURE 10.2 – Majoration de la variance d'une loi de Bernoulli

l'inclusion $A_{n,t} \subset B_{n,t}$ nous donne :

$$P\big(B_{n,t}\big) \geqslant 2\Phi(t) - 1 + \varepsilon_n. \tag{10.21}$$

En pratique, n est fixé et on a observé des valeurs numériques explicites $x_1, \ldots, x_n$ que l'on interprète comme les valeurs de $X_1(\omega), \ldots, X_n(\omega)$ pour *un même* ω tiré au sort (suivant P). On est donc en présence d'une valeur numérique explicite, $S_n(\omega)/n = (x_1 + \cdots + x_n)/n$, disons pour fixer les idées $S_n(\omega)/n = 0{,}53$. Proposer pour le paramètre inconnu p l'intervalle de confiance

$$I_{n,t} = \left[0{,}53 - \frac{t}{2\sqrt{n}}\ ;\ 0{,}53 + \frac{t}{2\sqrt{n}}\right],$$

c'est faire le *pari* que le ω observé est bien dans $B_{n,t}$. La probabilité de gagner ce pari est minorée par $2\Phi(t) - 1 + \varepsilon_n$. On dit que $I_{n,t}$ est un intervalle de confiance pour p avec un *niveau*[10] d'au moins $2\Phi(t) - 1 + \varepsilon_n$. En pratique, on laisse tomber le ε_n et on détermine t de façon approchée grâce à la tabulation de Φ. Par exemple pour un niveau de confiance de 95%, on est ramené à la résolution de l'équation $\Phi(t) = 1{,}95/2 = 0{,}975$ d'où $t \simeq 1{,}96$, ce qui nous donne l'intervalle

$$I_n = \left[\frac{S_n(\omega)}{n} - \frac{1{,}96}{2\sqrt{n}}\ ;\ \frac{S_n(\omega)}{n} + \frac{1{,}96}{2\sqrt{n}}\right], \quad \text{au niveau de confiance 95\%.}$$

En fait les statisticiens préfèrent une variante de cette méthode pour obtenir des intervalles de confiance plus étroits, notamment quand p n'est pas trop proche de $1/2$. L'idée est de remplacer la variance inconnue pq de X_1 par un *estimateur* au lieu de la majorer de façon certaine par (10.20). Ainsi en estimant pq par $\overline{X}(1 - \overline{X})$ où

10. Il y a ici un piège sémantique : supposons qu'on ait trouvé $I_{n,t} = [0{,}51; 0{,}55]$ avec un niveau de confiance de 95 %. Il est tout à fait incorrect d'écrire « $P(p \in [0{,}51; 0{,}55]) \geqslant 0{,}95$ ». En effet, p n'a rien d'aléatoire, c'est une constante. L'aléatoire concerne notre ignorance sur sa valeur. Même si on considère p comme une variable aléatoire constante, la probabilité de son appartenance à $[0{,}51; 0{,}55]$ vaut 0 ou 1 et, comme on ne peut pas exclure le premier cas, on ne peut pas minorer cette probabilité par 0,95.

$\overline{X} = S_n/n$, on obtient au niveau de confiance 95% l'intervalle

$$J_n = \left[\overline{X}(\omega) - 1{,}96\sqrt{\frac{\overline{X}(\omega)(1-\overline{X}(\omega))}{n}}\ ;\ \overline{X}(\omega) + 1{,}96\sqrt{\frac{\overline{X}(\omega)(1-\overline{X}(\omega))}{n}}\right].$$

Cette deuxième méthode sera justifiée ultérieurement par le théorème limite central avec autonormalisation, th. 10.21 p. 378.

10.2.4 Généralisation du TLC

Le théorème limite central a un domaine de validité bien plus vaste que celui des suites de variables i.i.d. de carré intégrable. Nous allons en voir trois généralisations. Mais auparavant signalons que dans le cas i.i.d., le théorème 10.14 est optimal : la convergence en loi de S_n^* *équivaut* à la condition $\mathbf{E}X_1^2 < +\infty$ (noter l'analogie avec la LFGN dans le cas i.i.d. où la convergence presque-sûre de S_n/n équivaut à la condition $\mathbf{E}|X_1| < +\infty$). On peut en effet démontrer le résultat suivant.

Théorème 10.18. *Soit $(X_i)_{i\geqslant 1}$ une suite de variables aléatoires i.i.d. définies sur le même espace probabilisé et telle que $S_n/\sqrt{n}$ converge en loi vers $\mathfrak{N}(0,1)$. Alors X_1 est de carré intégrable, $\mathbf{E}X_1^2 = 1$ et $\mathbf{E}X_1 = 0$.*

Ce théorème contient essentiellement la réciproque du théorème 10.14. En effet, en appliquant le théorème 10.18 aux variables aléatoires $X_i' = (X_i - \mu)/\sigma$ on en déduit que s'il existe $\mu \in \mathbb{R}$ et $\sigma > 0$ constantes [11] telles que $(S_n - n\mu)/(\sigma\sqrt{n})$ converge en loi vers [12] une v.a. Z de loi $\mathfrak{N}(0,1)$, alors $\mathbf{E}X_1^2 < +\infty$, $\mathbf{E}X_1 = \mu$ et $\operatorname{Var} X_1 = \sigma^2$.

Passons aux généralisations du TLC. La première abandonne l'hypothèse d'équidistribution [13], au prix d'une plus forte intégrabilité des X_i, c.-à-d. l'existence de moments d'ordre supérieur à 2.

Théorème 10.19 (Liapounov)**.** *Soit $(X_k)_{k\geqslant 1}$ une suite de variables aléatoires définies sur le même espace probabilisé, indépendantes (mais pas forcément de même loi), centrées, ayant toutes un moment d'ordre $2+\delta$ (pour une certaine constante $\delta > 0$). On note $s_n^2 := \operatorname{Var}(S_n)$ et on suppose $s_n > 0$. On suppose de plus vérifiée la* condition de Liapounov *:*

$$\frac{1}{s_n^{2+\delta}} \sum_{k=1}^{n} \mathbf{E}|X_k|^{2+\delta} \xrightarrow[n\to+\infty]{} 0. \tag{10.22}$$

Alors S_n/s_n converge en loi vers Z de loi $\mathfrak{N}(0,1)$.

Bien entendu, ce théorème n'est énoncé avec des variables centrées que pour alléger les notations. Si les X_k ne sont pas centrées, on se ramène à des variables centrées en appliquant le théorème aux $X_k' := X_k - \mathbf{E}X_k$, ce qui donne la convergence en loi de $(S_n - \mathbf{E}S_n)/s_n$ vers Z.

11. Bien sûr sans supposer *a priori* que ces constantes ont un lien avec d'éventuels moments de X_1.
12. En fait, on peut même se passer de l'hypothèse que la loi limite est gaussienne.
13. C'est-à-dire l'hypothèse que les X_i ont même loi.

En fait, on peut se passer de l'hypothèse d'existence de moments d'ordre strictement supérieur à 2, mais au prix d'une hypothèse qui peut sembler de prime abord un peu artificielle, la *condition de Lindeberg*. Les courageux et les curieux désireux de comprendre sa provenance sont invités à lire la section 10.4. Le résultat suivant peut être considéré comme la généralisation optimale du TLC dans le cas de variables aléatoires indépendantes non équidistribuées.

Théorème 10.20 (Lindeberg)**.** *Considérons le « tableau triangulaire » de variables aléatoires de ligne numéro n ($n \in \mathbb{N}^*$) :*

$$X_{n,1}, \dots, X_{n,i}, \dots, X_{n,k_n},$$

où ces k_n variables sont définies sur le même $(\Omega_n, \mathcal{F}_n, P_n)$, indépendantes, centrées, de carré intégrable. On note

$$S_n := \sum_{i=1}^{k_n} X_{n,i}, \qquad \sigma_{n,i}^2 := \operatorname{Var} X_{n,i}, \qquad s_n^2 := \operatorname{Var} S_n = \sum_{i=1}^{k_n} \sigma_{n,i}^2.$$

On suppose $s_n > 0$ pour tout n. Si de plus le tableau vérifie la condition de Lindeberg *:*

$$\forall \varepsilon > 0, \quad \frac{1}{s_n^2} \sum_{i=1}^{k_n} \mathbf{E}\big(X_{n,i}^2 \mathbf{1}_{\{|X_{n,i}| > \varepsilon s_n\}}\big) \xrightarrow[n \to +\infty]{} 0, \tag{10.23}$$

alors S_n/s_n converge en loi vers une v.a. Z de loi $\mathfrak{N}(0,1)$.

On peut montrer que la condition de Lindeberg implique que $\lim_{n\to+\infty} k_n = +\infty$, ce qui explique que l'on s'abstienne de mentionner que k_n tend vers l'infini dans les hypothèses. L'appelation « tableau triangulaire » est un abus de langage commode, la forme triangulaire n'apparaissant réellement que dans le cas particulier où $k_n = cn$. Remarquons qu'avec un tableau triangulaire, *toutes* les v.a. termes de la somme S_n peuvent changer quand on passe de S_n à S_{n+1}, c'est pour cela qu'il n'est pas nécessaire de supposer que des variables situées sur des lignes différentes du tableau sont définies sur le même espace probabilisé. L'espérance utilisée dans la formule (10.23) est définie relativement à la mesure de probabilité P_n, il serait plus correct de la noter $\mathbf{E}_n$ au lieu de $\mathbf{E}$. Pour toute v.a. Y positive définie sur $(\Omega_n, \mathcal{F}_n, P_n)$, $\mathbf{E}_n Y := \int_0^{+\infty} P_n(Y > t)\,\mathrm{d}t$, etc.

Nous allons vérifier à titre d'exercice que le théorème de Lindeberg contient à la fois le TLC cas i.i.d. (th. 10.14) et le théorème de Liapounov.

Commençons par le cas d'une suite $(X_i)_{i\geqslant 1}$, de variables aléatoires i.i.d. de carré intégrable. Considérons le tableau triangulaire dont la ligne n est formée des $k_n = n$ variables aléatoires $X_1, X_2, \dots, X_n$ (c.-à-d. $X_{n,i} = X_i$) définies sur $(\Omega_n, \mathcal{F}_n, P_n) = (\Omega, \mathcal{F}, P)$. Ces variables étant indépendantes et de même loi, $s_n^2 = n\sigma^2$ et les variables $X_{n,i}^2 \mathbf{1}_{\{|X_{n,i}| > \varepsilon s_n\}}$ ont même loi. Par conséquent, la quantité censée tendre vers 0 dans la condition de Lindeberg s'écrit pour tout $\varepsilon > 0$,

$$\frac{1}{s_n^2} \sum_{i=1}^{k_n} \mathbf{E}\big(X_{n,i}^2 \mathbf{1}_{\{|X_{n,i}| > \varepsilon s_n\}}\big) = \frac{1}{n\sigma^2} n \mathbf{E}\big(X_1^2 \mathbf{1}_{\{|X_1| > \varepsilon \sigma n^{1/2}\}}\big) = \frac{1}{\sigma^2} \mathbf{E}\big(X_1^2 \mathbf{1}_{\{|X_1| > \varepsilon \sigma n^{1/2}\}}\big).$$

La convergence vers 0 de cette dernière expression quand n tend vers l'infini, découle de l'hypothèse $\mathbf{E}(X_1^2) < +\infty$ *via* le théorème de convergence dominée (exercice). De façon plus élémentaire, on peut aussi remarquer que si u_n est une suite de réels positifs tendant vers l'infini et Y une v.a. positive,

$$\mathbf{E}\big(Y\mathbf{1}_{\{Y>u_n\}}\big) = \int_0^{+\infty} P(Y\mathbf{1}_{\{Y>u_n\}} > t)\,\mathrm{d}t \overset{\star}{=} u_n P(Y > u_n) + \int_{u_n}^{+\infty} P(Y > t)\,\mathrm{d}t.$$

Si Y est intégrable, cette dernière expression [14] tend vers 0 quand n tend vers l'infini en raison de la convergence de l'intégrale généralisée $\int_0^{+\infty} P(Y > t)\,\mathrm{d}t$ et par application à Y de l'inégalité de Markov raffinée (prop. 7.19). Ensuite on prend $Y = X_1^2$ et $u_n = n\varepsilon^2\sigma^2$ en notant que $\{X_1^2 > u_n\} = \{|X_1| > u_n^{1/2}\}$. Nous avons donc bien vérifié que le théorème de Lindeberg implique le théorème 10.14.

Considérons maintenant le cas d'une suite de variables aléatoires X_i indépendantes, d'espérance nulle et vérifiant $\mathbf{E}|X_i|^{2+\delta} < +\infty$ pour une constante $\delta > 0$ ne dépendant pas de $i \in \mathbb{N}^*$. Supposons de plus vérifiée la condition de Liapounov (10.22). Nous allons voir que l'on peut alors appliquer le théorème de Lindeberg au tableau triangulaire où $k_n = n$, la ligne n étant formée des variables aléatoires $X_1, X_2, \ldots, X_n$ (c.-à-d. $X_{n,i} = X_i$) définies sur $(\Omega_n, \mathcal{F}_n, P_n) = (\Omega, \mathcal{F}, P)$. En notant que sur l'évènement $\{|X_i| \geqslant \varepsilon s_n\}$ on a $|X_i|^\delta/(\varepsilon s_n)^\delta \geqslant 1$, la majoration

$$\mathbf{E}\big(X_i^2\mathbf{1}_{\{|X_i|\geqslant\varepsilon s_n\}}\big) \leqslant \mathbf{E}\Big(X_i^2\Big(\frac{|X_i|}{\varepsilon s_n}\Big)^\delta \mathbf{1}_{\{|X_i|\geqslant\varepsilon s_n\}}\Big) \leqslant \frac{1}{\varepsilon^\delta s_n^\delta}\mathbf{E}|X_k|^{2+\delta},$$

nous montre que la condition de Lindeberg est vérifiée *via* l'inégalité

$$\frac{1}{s_n^2}\sum_{i=1}^n \mathbf{E}\big(X_i^2\mathbf{1}_{\{|X_i|>\varepsilon s_n\}}\big) \leqslant \frac{1}{\varepsilon^\delta s_n^{2+\delta}}\sum_{k=1}^n \mathbf{E}|X_k|^{2+\delta}.$$

En conclusion, S_n/s_n converge en loi vers $\mathfrak{N}(0,1)$ et nous avons ainsi vérifié que le théorème de Lindeberg implique celui de Liapounov.

La version suivante du théorème limite central est directement motivée par les applications statistiques. Pour construire à l'aide du TLC un intervalle de confiance pour une espérance inconnue, on a besoin de bornes qui ne dépendent pas de la variance lorsque cette dernière est aussi inconnue, cf. l'exemple page 376. Une technique courante est alors de remplacer cette variance inconnue par un estimateur de celle-ci. La légitimation mathématique repose sur le théorème suivant.

Théorème 10.21 (TLC avec autonormalisation)**.** *Soient $X_1, \ldots, X_n, \ldots$ des variables aléatoires indépendantes et de même loi telle que $\mathbf{E}X_1^2 < +\infty$ et $\sigma^2 := \operatorname{Var} X_1 > 0$. On note $S_n := X_1 + \cdots + X_n$. On suppose de plus que $(V_n)_{n\geqslant 1}$ est une suite de variables aléatoires positives qui converge en probabilité vers σ^2. Alors*

$$T_n := \sqrt{\frac{n}{V_n}}\left(\frac{S_n}{n} - \mathbf{E}X_1\right) \xrightarrow[n\to+\infty]{\text{loi}} Z,$$

14. La justification soigneuse de l'égalité étoilée est laissée en exercice.

où Z suit la loi gaussienne standard $\mathfrak{N}(0,1)$.

L'énoncé est volontairement légèrement incorrect pour respecter l'usage. La preuve est reportée à la section 10.4 où l'on trouvera un énoncé rectifié. En général on applique ce théorème en prenant pour V_n la « variance empirique ». Cette variance empirique est pour chaque ω, la variance calculée sur la série statistique réellement observée $x_1 = X_1(\omega), \ldots, x_n = X_n(\omega)$. C'est donc la variable aléatoire :

$$V_n = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - (\overline{X})^2,$$

où $\overline{X} = S_n/n$. En appliquant deux fois la loi forte des grands nombres, on voit facilement que la variance empirique converge presque sûrement (donc aussi en probabilité) vers la variance théorique $\sigma^2 = \mathbf{E}X_1^2 - (\mathbf{E}X_1)^2$.

Si on revient au cas où les X_i sont des variables de Bernoulli, discuté page 376, on peut appliquer directement le théorème ci-dessus avec $V_n = \overline{X}(1-\overline{X})$ en notant que par la loi forte des grands nombres, V_n converge presque-sûrement vers $p(1-p) = \sigma^2$. On peut aussi remarquer que pour des variables de Bernoulli, la variance empirique n'est autre que $\overline{X}(1-\overline{X})$. La vérification de cette affirmation est facile une fois que l'on a noté que $X_i = X_i^2$ car X_i ne prend que les valeurs 0 et 1.

10.3 Théorème limite central vectoriel

Il est courant en statistique d'observer des données de nature vectorielle. On a souvent recours dans ce contexte à des approximations gaussiennes dont la justification mathématique repose sur un théorème limite central vectoriel. Avant de présenter la version du TLC pour des suites i.i.d. de vecteurs aléatoires, il convient d'apporter quelques précisions sur les vecteurs aléatoires gaussiens.

10.3.1 Vecteurs aléatoires gaussiens

Il est commode d'élargir la famille des lois gaussiennes sur $\mathbb{R}$ en considérant qu'une variable aléatoire constante $X = c$, ou $X = c$ p.s., suit la loi gaussienne de paramètres $m = c$ et $\sigma = 0$. Bien sûr, cette loi est la mesure de Dirac au point c et n'a donc pas de densité. Avec cette convention, nous pouvons donner la définition suivante.

Définition 10.22 (vecteur aléatoire gaussien)**.** *On dit que le vecteur aléatoire $X = (X_1, \ldots, X_d)$ de $\mathbb{R}^d$ est* gaussien *si pour toute forme linéaire u sur $\mathbb{R}^d$, la variable aléatoire réelle $u(X)$ suit une loi gaussienne sur $\mathbb{R}$, autrement dit si*

$$\forall (a_1, \ldots, a_d) \in \mathbb{R}^d, \quad a_1X_1 + \cdots + a_dX_d \text{ est une v.a. gaussienne.} \tag{10.24}$$

Remarque 10.23. En prenant pour formes linéaires u les formes coordonnées $x \mapsto x_i$, $i \in [\![1, d]\!]$, on voit immédiatement que si X est un vecteur gaussien, ses composantes $X_1, \ldots, X_d$ sont des variables aléatoires gaussiennes [15]. *La réciproque est*

15. Au sens élargi donné ci-dessus, certaines pouvant être constantes p.s.

fausse. Pour s'en convaincre, on pourra étudier en exercice l'exemple suivant. On définit $X = (X_1, X_2)$ en prenant X_1 gaussienne $\mathfrak{N}(0,1)$ et en posant $X_2 := X_1 \mathbf{1}_{\{|X_1| \leqslant 1\}} - X_1 \mathbf{1}_{\{|X_1|>1\}}$. On peut alors montrer que

1. X_2 suit la loi $\mathfrak{N}(0,1)$ (en distinguant les trois cas $x < -1$, $-1 \leqslant x \leqslant 1$ et $x > 1$, vérifier que $P(X_2 \leqslant x) = \Phi(x) = P(X_1 \leqslant x)$) ;
2. $X_1 + X_2$ ne suit pas une loi gaussienne, donc le vecteur (X_1, X_2) ne peut pas être gaussien.

Proposition 10.24 (paramètres d'un vecteur gaussien)**.** *La loi d'un vecteur aléatoire gaussien $X = (X_1, \dots, X_d)$ est caractérisée par son vecteur d'espérances*

$$m = \mathbf{E}X = \big(\mathbf{E}X_1, \dots, \mathbf{E}X_d\big) \tag{10.25}$$

et sa matrice de covariance

$$K = \big[\operatorname{Cov}(X_i, X_j)\big]_{1 \leqslant i,j \leqslant d}. \tag{10.26}$$

La loi de X sera notée $\mathfrak{N}(m, K)$.

Preuve. Soient $X = (X_1, \dots, X_d)$ et $Y = (Y_1, \dots, Y_d)$ deux vecteurs aléatoires *gaussiens* ayant même vecteur d'espérances et même matrice de covariance. D'après (8.13) et (8.15), pour toute forme linéaire u sur $\mathbb{R}^d$, $\mathbf{E}u(X) = u(\mathbf{E}X) = u(\mathbf{E}Y) = \mathbf{E}u(Y)$ et $\operatorname{Var} u(X) = \operatorname{Var} u(Y)$. Les deux *variables aléatoires réelles gaussiennes* $u(X)$ et $u(Y)$ ont donc mêmes paramètres, donc même loi $\mathfrak{N}\big(\mathbf{E}u(X), \operatorname{Var}^{1/2} u(X)\big)$. Ceci étant vrai pour toute forme linéaire u, on en déduit grâce à la proposition 8.32, que les *vecteurs* aléatoires X et Y ont même loi. La loi d'un vecteur gaussien X est donc bien caractérisée par les seuls paramètres $\mathbf{E}X$ et K. □

Proposition 10.25 (famille des vecteurs gaussiens)**.** *La famille des vecteurs aléatoires gaussiens de $\mathbb{R}^d$ est stable par addition de vecteur constant, multiplication par un scalaire constant et addition de vecteurs aléatoires gaussiens* indépendants*. L'image d'un vecteur aléatoire gaussien de $\mathbb{R}^d$ par une application linéaire $\mathbb{R}^d \to \mathbb{R}^j$ est un vecteur gaussien de $\mathbb{R}^j$.*

Preuve. On vérifie facilement que la famille des *variables* aléatoires gaussiennes est stable par transformations affines : si Z est gaussienne de loi $\mathfrak{N}(m, \sigma)$, $aZ + b$ est gaussienne [16] de loi $\mathfrak{N}(am + b, |a|\sigma)$. En particulier, les formes linéaires sur $\mathbb{R}$ étant du type $t \mapsto at$, Z peut aussi être vu comme un *vecteur* gaussien de dimension 1.

Soit $z = (z_1, \dots, z_d)$ un vecteur (non aléatoire) de $\mathbb{R}^d$ et X un vecteur aléatoire gaussien. Pour vérifier que le vecteur aléatoire $Y := X + z$ est gaussien, on prend une forme linéaire quelconque u et on note que $u(Y) = u(X) + u(z)$ est la somme de la variable aléatoire gaussienne $u(X)$ et de la constante réelle $u(z)$, c'est donc une v.a.

16. Démontrez le en exercice de révision en cherchant une relation entre la f.d.r. de Z et celle de $aZ + b$ à l'aide d'un changement de variable dans l'intégrale de la densité (cas $a \neq 0$ et Z non constante)...

gaussienne ayant même variance que $u(X)$ et d'espérance $\mathbf{E}u(X) + u(z)$. Ainsi $u(Y)$ est une v.a. gaussienne pour toute forme linéaire u, donc Y est un vecteur gaussien. De même si c est une constante réelle, cX est gaussien car $u(cX) = cu(X)$ est une variable aléatoire gaussienne de loi $\mathfrak{N}\big(c\mathbf{E}u(X), |c| \operatorname{Var}^{1/2} u(X)\big)$.

Soient X et Y deux vecteurs aléatoires gaussiens de $\mathbb{R}^d$ indépendants. Pour toute forme linéaire u sur $\mathbb{R}^d$, $u(X+Y) = u(X) + u(Y)$ est la somme de deux variables aléatoires gaussiennes $u(X)$ et $u(Y)$ qui ont hérité de l'indépendance de X et Y. Or on peut vérifier, cf. lemme 10.27 ci-dessous, que la somme de deux variables aléatoires gaussiennes *indépendantes* est encore une v.a. gaussienne. Ainsi $u(X+Y)$ est une v.a. gaussienne pour toute u, donc le vecteur $X+Y$ est gaussien.

Enfin si X est un vecteur gaussien de $\mathbb{R}^d$ et $F : \mathbb{R}^d \to \mathbb{R}^j$ est une application linéaire, $Y = F(X)$ est un vecteur gaussien de $\mathbb{R}^j$ car pour toute forme linéaire $v : \mathbb{R}^j \to \mathbb{R}$, $v(Y) = (v \circ F)(X) = u(X)$ avec $u := v \circ F$ forme linéaire sur $\mathbb{R}^d$. □

Corollaire 10.26. *Si les* variables *aléatoires réelles $X_1, \dots, X_d$ sont indépendantes et gaussiennes, alors le* vecteur *aléatoire $X = (X_1, \dots, X_d)$ est gaussien dans $\mathbb{R}^d$.*

Preuve. On raisonne par récurrence sur la dimension d de l'espace. D'abord pour $d = 2$, si X_1 et X_2 sont deux variables aléatoires gaussiennes indépendantes, donc aussi des vecteurs aléatoires gaussiens de dimension 1, elles engendrent par tranformation linéaire les vecteurs gaussiens $(X_1, 0)$ et $(0, X_2)$ de $\mathbb{R}^2$ qui héritent de leur indépendance. Par conséquent leur *somme* $(X_1, 0) + (0, X_2) = (X_1, X_2)$ est encore un vecteur gaussien de $\mathbb{R}^2$. Supposons maintenant le corollaire vrai pour $\mathbb{R}^j$ pour tout $j \leqslant d$ et montrons qu'il est encore vrai pour $j = d+1$. En effet par transformations linéaires, $(X_1, \dots, X_d, 0)$ et $(0, \dots, 0, X_{d+1})$ sont des vecteurs gaussiens de $\mathbb{R}^{d+1}$, fonctions mesurables des blocs disjoints $(X_1, \dots, X_d)$ et X_{d+1} dont ils héritent l'indépendance (proposition 8.37). Donc leur somme $(X_1, \dots, X_d, X_{d+1})$ est un vecteur gaussien de $\mathbb{R}^{d+1}$ par la proposition 10.25. □

Lemme 10.27. *La somme de deux variables aléatoires gaussiennes indépendantes est encore une variable aléatoire gaussienne.*

Preuve. Soient Y_1 et Y_2 deux v.a. gaussiennes indépendantes de paramètres respectifs (m_1, σ_1) et (m_2, σ_2). Si ces deux v.a. sont constantes, $\sigma_1 = \sigma_2 = 0$, leur somme est la constante $m_1 + m_2$ donc est encore gaussienne. Si une seule des deux est constante, disons Y_2, on sait que $Y_1 + Y_2 = Y_1 + m_2$ est gaussienne de paramètres $m_1 + m_2$ et σ_1. Il nous reste à traiter le cas où $\sigma_1 > 0$ et $\sigma_2 > 0$. En posant $X_1 := Y_1 - m_1$ et $X_2 := Y_2 - m_2$, on obtient deux nouvelles v.a. gaussiennes indépendantes de loi respective $\mathfrak{N}(0, \sigma_1)$ et $\mathfrak{N}(0, \sigma_2)$. Il suffit de vérifier que $X_1 + X_2$ est une v.a. gaussienne $\mathfrak{N}(0, \sigma)$ et on en déduira que $Y_1 + Y_2$ a pour loi $\mathfrak{N}(m_1 + m_2, \sigma)$.

Puisque σ_1 et σ_2 sont non nuls, X_1 et X_2 ont chacune une densité donnée par :

$$f_i(t) = \frac{1}{\sigma_i \sqrt{2\pi}} \exp\left(- \frac{t^2}{2\sigma_i^2} \right), \quad i = 1, 2.$$

On sait alors qu'en raison de l'indépendance de X_1 et X_2, la v.a. $X_1 + X_2$ admet elle aussi une densité g qui est la convoluée $f_1 * f_2$ définie par $g(t) = \int_{-\infty}^{+\infty} f_1(s) f_2(t-s)\,\mathrm{d}s$, soit ici :

$$\forall t \in \mathbb{R}, \qquad g(t) = \frac{1}{2\pi\sigma_1\sigma_2} \int_{-\infty}^{+\infty} \exp\left(-\frac{s^2}{2\sigma_1^2} - \frac{(t-s)^2}{2\sigma_2^2} \right) \mathrm{d}s. \tag{10.27}$$

Pour vérifier que g est bien une densité gaussienne, nous allons transformer la forme quadratique en s, t à l'intérieur de l'exponentielle pour la mettre sous la forme

$$\frac{s^2}{2\sigma_1^2} + \frac{(t-s)^2}{2\sigma_2^2} = c\big((s-at)^2 + bt^2\big).$$

Ensuite le changement de variable $x = s - at$ dans l'intégrale (on intègre relativement à s, donc at joue le rôle d'une constante) nous permettra de sortir tout ce qui dépend de t de l'intégrale pour aboutir à $g(t) = c' \exp(-b't^2)$, l'intégrale devenant alors une constante absorbée dans le c'.

Passons aux détails du calcul. Il est commode de poser $\sigma^2 := \sigma_1^2 + \sigma_2^2$ et $\tau_i := \sigma_i^2/\sigma^2$.

$$\begin{aligned}
\frac{s^2}{2\sigma_1^2} + \frac{(t-s)^2}{2\sigma_2^2} = \frac{\sigma_2^2 s^2 + \sigma_1^2 (t-s)^2}{2\sigma_1^2\sigma_2^2} &= \frac{\sigma^2 s^2 - 2\sigma_1^2 st + \sigma_1^2 t^2}{2\sigma_1^2\sigma_2^2} \\
&= \frac{\sigma^2}{2\sigma_1^2\sigma_2^2}\big(s^2 - 2\tau_1 st + \tau_1 t^2\big) \\
&= \frac{\sigma^2}{2\sigma_1^2\sigma_2^2}\big((s-\tau_1 t)^2 + (\tau_1 - \tau_1^2)t^2\big) \\
&= \frac{\sigma^2}{2\sigma_1^2\sigma_2^2}(s-\tau_1 t)^2 + \frac{t^2}{2\sigma^2},
\end{aligned}$$

en notant que $\tau_1 + \tau_2 = 1$, d'où $\tau_1 - \tau_1^2 = \tau_1\tau_2 = \sigma_1^2\sigma_2^2\sigma^{-4}$. En reportant ce calcul dans (10.27), on a donc

$$g(t) = \frac{1}{2\pi\sigma_1\sigma_2} \exp\left(-\frac{t^2}{2\sigma^2} \right) \int_{-\infty}^{+\infty} \exp\left(-\frac{\sigma^2}{2\sigma_1^2\sigma_2^2}(s-\tau_1 t)^2 \right) \mathrm{d}s.$$

Le changement de variable $x = \sigma(\sigma_1\sigma_2)^{-1}(s - \tau_1 t)$ dans cette intégrale généralisée nous ramène après simplifications à l'intégrale de la densité gaussienne standard :

$$g(t) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{t^2}{2\sigma^2} \right) \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2} \right) \mathrm{d}x = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{t^2}{2\sigma^2} \right).$$

On en déduit que $X_1 + X_2$ suit la loi gaussienne $\mathfrak{N}(0, \sigma)$ avec $\sigma = (\sigma_1^2 + \sigma_2^2)^{1/2}$. □

Remarque 10.28. Jusqu'à présent nous n'avions pas donné d'exemple de vecteur gaussien en dimension $d > 1$. Le corollaire 10.26 nous donne une première famille d'exemples, les vecteurs gaussiens construits en prenant leurs composantes gaussiennes

indépendantes. En combinant ceci avec la proposition 10.25, on obtient une classe d'exemples bien plus large, celle des vecteurs aléatoires Y qui peuvent s'écrire

$$Y = A(X) + b \tag{10.28}$$

où $X = (X_1, \dots, X_d)$ est gaussien à composantes indépendantes, A est une application linéaire quelconque de $\mathbb{R}^d$ dans $\mathbb{R}^d$ et $b = (b_1, \dots, b_d)$ un vecteur constant.

En fait on peut montrer que *tout* vecteur gaussien de $\mathbb{R}^d$ admet une représentation de la forme (10.28).

10.3.2 TLC vectoriel

Nous admettons le résultat suivant qui permet de ramener la convergence en loi en dimension d à de la convergence en loi en dimension 1. Dans la littérature anglo-américaine, ce résultat est connu sous le sobriquet de *Cramér-Wold device*, ce que l'on pourrait traduire par « truc de Cramér-Wold » ou de façon moins irrévérencieuse, par « astuce de Cramér-Wold ».

Lemme 10.29 (Cramér-Wold)**.** *La suite $(Y_n)_{n\geqslant 1}$ de vecteurs aléatoires dans $\mathbb{R}^d$ converge en loi dans $\mathbb{R}^d$ vers le vecteur aléatoire Y si et seulement si pour toute forme linéaire $u : \mathbb{R}^d \to \mathbb{R}$, la suite de variables aléatoires réelles $\big(u(Y_n)\big)_{n\geqslant 1}$ converge en loi dans $\mathbb{R}$ vers $u(Y)$.*

Théorème 10.30. *Soit (X_k) une suite de vecteurs aléatoires de $\mathbb{R}^d$, indépendants, de même loi et de carré intégrable, c.-à-d. $\mathbf{E}\|X_1\|^2 < +\infty$, et $S_n := \sum_{k=1}^n X_k$. Alors*

$$\frac{S_n - \mathbf{E}S_n}{\sqrt{n}} \xrightarrow[n\to+\infty]{\text{loi}} Z, \tag{10.29}$$

où Z est gaussien de loi $\mathfrak{N}(0, K)$, K étant la matrice de covariance de X_1.

Malgré la ressemblance formelle de cet énoncé avec celui du théorème 10.14, il n'est pas possible ici de modifier la normalisation $\sqrt{n}$ par $\sigma\sqrt{n}$ pour avoir toujours la même loi limite $\mathfrak{N}(0, I)$, où I est la matrice identité par rapport à la base canonique de $\mathbb{R}^d$.

Preuve. Par le lemme de Cramér-Wold, il suffit de vérifier que pour toute forme linéaire u sur $\mathbb{R}^d$, $u\big(n^{-1/2}(S_n - \mathbf{E}S_n)\big)$ converge en loi vers $u(Z)$ où Z désigne un vecteur aléatoire gaussien de loi $\mathfrak{N}(0, K)$. Notons d'emblée que $\mathbf{E}u(Z) = u(\mathbf{E}Z) = u(0) = 0$ et $\operatorname{Var} u(Z) = \operatorname{Var} u(X_1)$ puisque Z et X_1 ont même matrice de covariance, cf. proposition 8.30.

Par linéarité de u,

$$u\left(\frac{S_n - \mathbf{E}S_n}{\sqrt{n}}\right) = \frac{1}{\sqrt{n}} \sum_{k=1}^n \big(u(X_k) - \mathbf{E}u(X_k)\big).$$

Les $u(X_k)$ sont des variables aléatoires réelles indépendantes, de même loi, de carré intégrable ($\mathbf{E}|u(X_1)|^2 \leqslant \|u\|^2 \mathbf{E}\|X_1\|^2$), on peut donc appliquer le théorème 10.14 qui donne ici

$$u\left(\frac{S_n - \mathbf{E}S_n}{\sqrt{n}}\right) \xrightarrow[n\to+\infty]{\text{loi}} W,$$

où W désigne *n'importe quelle v.a. gaussienne de loi* $\mathfrak{N}\big(0, \operatorname{Var} u(X_1)\big)$. Cette loi est celle de $u(Z)$, donc $u\big(n^{-1/2}(S_n - \mathbf{E}S_n)\big)$ converge en loi vers $u(Z)$. □

Une application importante du théorème 10.30 est la convergence d'une loi multinomiale vers une loi gaussienne, ce qui en un certain sens, généralise le théorème de de Moivre-Laplace.

Rappelons que la loi multinomiale sert à modéliser le total des résultats observés pour chaque type dans une suite d'épreuves répétées indépendantes ayant chacune d types de résultats possibles. Par exemple si on lance 200 fois un dé, on obtient un vecteur de dimension 6 dont la i^{e} composante est le nombre total d'apparitions de la face numéro i au cours des 200 lancers. Ce vecteur suit la loi multinomiale de paramètres 200 et $(p_1, p_2, p_3, p_4, p_5, p_6)$, où les p_i valent tous 1/6 si le dé est équilibré. Plus formellement, le vecteur aléatoire N suit la loi multinomiale de paramètres n et $p = (p_1, \dots, p_d)$ où $n \in \mathbb{N}^*$ et les p_i sont strictement positifs et de somme 1, si pour tout d-uple $(j_1, j_2, \dots, j_d)$ d'entiers tels que $j_1 + j_2 + \cdots + j_d = n$,

$$P\big(N = (j_1, j_2, \dots, j_d)\big) = \frac{n!}{j_1! j_2! \dots j_d!} p_1^{j_1} p_2^{j_2} \dots p_d^{j_d}$$

et $P\big(N = (j_1, j_2, \dots, j_d)\big) = 0$ si $j_1 + j_2 + \cdots + j_d \neq n$.

Théorème 10.31. *Si $(N_n)_{n\geqslant 1}$ est une suite de vecteurs aléatoires de $\mathbb{R}^d$ de loi multinomiale de paramètres n et $p = (p_1, \dots, p_d)$,*

$$\frac{1}{\sqrt{n}}(N_n - np) \xrightarrow[n\to+\infty]{\text{loi}} Z \quad \textit{de loi } \mathfrak{N}(0, K), \tag{10.30}$$

où K est la matrice de covariance de terme général

$$K_{i,j} = p_i \delta_{i,j} - p_i p_j, \quad (i,j) \in [\![1, d]\!]^2.$$

Preuve. Le vecteur aléatoire N_n a *même loi* que $S_n := \sum_{k=1}^n X_k$, où les X_k sont des vecteurs aléatoires de $\mathbb{R}^d$, indépendants et de même loi donnée par

$$P\big(X_k = (0, \dots, 0, \underset{\substack{\uparrow \\ i}}{1}, 0, \dots, 0)\big) = p_i, \quad i \in [\![1, d]\!].$$

Autrement dit, on se ramène à un modèle d'épreuves répétées indépendantes et le vecteur aléatoire X_k est un codage binaire du résultat de la k^{e} épreuve. En notant $X_{k,i}$ la i^{e} composante de X_k, on voit que cette variable aléatoire suit la loi de Bernoulli

de paramètre p_i et que $\sum_{i=1}^d X_{k,i} = 1$. L'espérance de X_k est le *vecteur* p. Le vecteur X_k étant borné est clairement de carré intégrable. On est ainsi dans les conditions de validité du théorème 10.30 qui nous donne :

$$T_n := \frac{1}{\sqrt{n}} \sum_{k=1}^{n} (X_k - p) \xrightarrow[n\to+\infty]{\text{loi}} Z \quad \text{de loi } \mathfrak{N}(0, K), \tag{10.31}$$

où le terme général de la matrice de K par rapport à la base canonique de $\mathbb{R}^d$ est

$$K_{i,j} = \mathrm{Cov}(X_{1,i}, X_{1,j}) = \mathbf{E}(X_{1,i}X_{1,j}) - p_i p_j.$$

Si $i \neq j$, $X_{1,i}X_{1,j} = 0$ comme produit d'indicatrices d'évènements disjoints. Pour $i = j$, $X_{1,i}^2 = X_{1,i}$ car $X_{1,i}$ ne peut prendre que la valeur 0 ou 1. On a donc bien dans les deux cas $K_{i,j} = p_i\delta_{i,j} - p_i p_j$ et on conclut en remarquant que T_n et $n^{-1/2}(N_n - np)$ ont même loi. □

Remarque 10.32. La loi limite dans ce théorème est un exemple naturel de loi gaussienne sur $\mathbb{R}^d$ *sans densité*. On montre en effet que

$$P(Z_1 + Z_2 + \cdots + Z_d = 0) = 1 \quad \text{où} \quad Z = (Z_1, \ldots, Z_d) \sim \mathfrak{N}(0, K). \tag{10.32}$$

Pour cela, considérons la forme linéaire $u : (x_1, x_2, \ldots, x_d) \mapsto x_1 + x_2 + \cdots + x_d$ définie sur $\mathbb{R}^d$. La continuité de u et (10.31) entraînent la convergence en loi de $u(T_n)$ vers $u(Z)$. D'autre part, $u(T_n) = n^{-1/2} \sum_{k=1}^n \big(u(X_k) - u(p)\big)$ par linéarité de u. Comme $u(X_k) = 1 = u(p)$, $u(T_n)$ est la variable aléatoire nulle, sa loi est la masse de Dirac en 0 et il en est de même pour la loi limite quand n tend vers $+\infty$. Mais dire que $u(Z)$ a pour loi δ_0 revient exactement à (10.32). Ainsi toute la masse de la loi de Z est portée par l'hyperplan $x_1 + x_2 + \cdots + x_d = 0$ qui est de λ_d-mesure nulle et la loi de Z ne peut avoir de densité. Pour un autre argument, voir l'exercice 8.5.

10.4 Compléments sur la convergence en loi et le TLC

Cette section développe les aspects techniques de la convergence en loi dans le but de fournir autant que possible des preuves de résultats admis au cours de ce chapitre. En première lecture on pourra se contenter de jeter un coup d'oeil aux énoncés, quitte à y revenir plus tard en cas de besoin.

10.4.1 Outillage pour la convergence en loi

Avant de nous focaliser sur l'étude de la convergence en loi par les moments fonctionnels donc au sens de la définition 10.2, nous montrons que celle ci implique la convergence en loi au sens de la définition 10.1. La réciproque est vraie mais plus

difficile à démontrer et n'est de toutes façons quasiment pas utilisée dans cet ouvrage. On s'en dispensera donc [17].

Proposition 10.33. *Notons F_n et F les fonctions de répartition respectives des variables aléatoires réelles Y_n $(n \geqslant 1)$ et Y. On suppose que*

$$\text{pour toute } h \text{ continue bornée } \mathbb{R} \to \mathbb{R}, \quad \mathbf{E}h(Y_n) \xrightarrow[n\to+\infty]{} \mathbf{E}h(Y). \tag{10.33}$$

Alors

$$\text{pour tout } x \text{ point de continuité de } F, \quad F_n(x) \xrightarrow[n\to+\infty]{} F(x). \tag{10.34}$$

Preuve. Fixons $x \in \mathbb{R}$ et remarquons que $F_n(x) = P(Y_n \leqslant x) = \mathbf{E}\mathbf{1}_{]-\infty,x]}(Y_n)$ et $F(x) = \mathbf{E}\mathbf{1}_{]-\infty,x]}(Y)$. La fonction $f : t \mapsto \mathbf{1}_{]-\infty,x]}(t)$ est bornée sur $\mathbb{R}$ mais malheureusement, à cause de son unique discontinuité au point x, on ne peut pas appliquer directement (10.33) pour conclure.

Nous allons approximer f en l'encadrant par deux fonctions continues bornées, affines par morceaux, cf. figure 10.3.

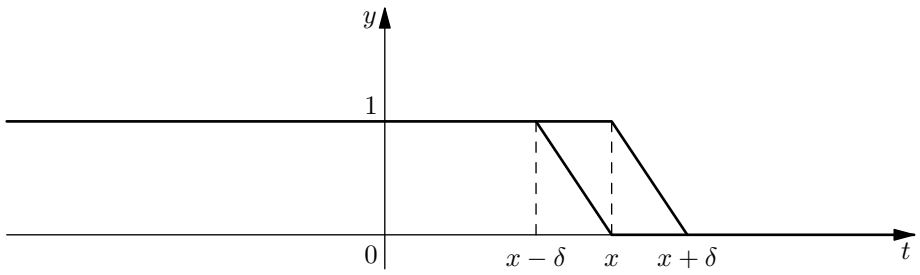


FIGURE 10.3 – Encadrement de $\mathbf{1}_{]-\infty,x]}$ par 2 fonctions continues : $g(\,.\,+\delta) \leqslant f \leqslant g$.

De façon plus formelle, si avec un paramètre $\delta > 0$, on définit la fonction g par

$$g(t) = \begin{cases} 1 & \text{si } t \in]-\infty, x] \\ 1 - (t-x)/\delta & \text{si } t \in]x, x+\delta] \\ 0 & \text{si } t \in]x+\delta, +\infty[, \end{cases}$$

alors $g(t+\delta) \leqslant f(t) \leqslant g(t)$ pour tout $t \in \mathbb{R}$, d'où

$$\forall t \in \mathbb{R}, \quad f(t+\delta) \leqslant g(t+\delta) \leqslant f(t) \leqslant g(t) \leqslant f(t-\delta). \tag{10.35}$$

Grâce à ces encadrements nous pouvons contrôler pour Z v.a. quelconque, la différence

17. Les curieux peuvent aller voir la preuve du théorème 9.32 dans [15].

$|P(Z \leqslant x) - P(Y \leqslant x)|$ à l'aide de moments fonctionnels des v.a. Y et Z. En effet

$$\begin{aligned} P(Z \leqslant x) - P(Y \leqslant x) &= \mathbf{E}f(Z) - \mathbf{E}f(Y) \\ &\leqslant \mathbf{E}g(Z) - \mathbf{E}f(Y) \\ &= \mathbf{E}g(Z) - \mathbf{E}g(Y) + \mathbf{E}g(Y) - \mathbf{E}f(Y) \\ &\leqslant \mathbf{E}g(Z) - \mathbf{E}g(Y) + \mathbf{E}f(Y-\delta) - \mathbf{E}f(Y) \\ &= \mathbf{E}g(Z) - \mathbf{E}g(Y) + F(x+\delta) - F(x), \end{aligned}$$

en rappelant que F est la f.d.r. de Y. De même

$$\begin{aligned} P(Y \leqslant x) - P(Z \leqslant x) &= \mathbf{E}f(Y) - \mathbf{E}f(Z) \\ &\leqslant \mathbf{E}f(Y) - \mathbf{E}g(Z+\delta) \\ &= \mathbf{E}f(Y) - \mathbf{E}g(Y+\delta) + \mathbf{E}g(Y+\delta) - \mathbf{E}g(Z+\delta) \\ &\leqslant \mathbf{E}f(Y) - \mathbf{E}f(Y+\delta) + \mathbf{E}g(Y+\delta) - \mathbf{E}g(Z+\delta) \\ &= F(x) - F(x-\delta) + \mathbf{E}g(Y+\delta) - \mathbf{E}g(Z+\delta). \end{aligned}$$

On en déduit la majoration

$$|P(Z \leqslant x) - P(Y \leqslant x)| \leqslant \max\big(|\mathbf{E}g(Z) - \mathbf{E}g(Y)| \,;\, |\mathbf{E}g(Y+\delta) - \mathbf{E}g(Z+\delta)|\big) + \Delta(F, x, \delta), \tag{10.36}$$

avec

$$\Delta(F, x, \delta) := \max\big(F(x) - F(x-\delta) \,;\, F(x+\delta) - F(x)\big). \tag{10.37}$$

L'inégalité (10.36) est valable pour tout x réel, tout $\delta > 0$, toute v.a. Z, avec la fonction g dépendant de x et δ. Si x est un point de continuité de F, $\Delta(F, x, \delta)$ tend vers 0 quand δ tend vers 0. Fixons $\varepsilon > 0$ quelconque. On peut alors trouver un δ assez petit pour que $\Delta(F, x, \delta) < \varepsilon$. Une fois choisi ce δ, les fonctions continues bornées g et $g(\,.\, + \delta)$ sont fixées. En prenant alors $Z = Y_n$ dans (10.36), l'hypothèse (10.33) appliquée avec $h = g$ et $h = g(\,.\, + \delta)$ nous donne l'existence d'un entier n_0 tel que

$$\forall n \geqslant n_0, \quad \max\big(|\mathbf{E}g(Y_n) - \mathbf{E}g(Y)| \,;\, |\mathbf{E}g(Y+\delta) - \mathbf{E}g(Y_n+\delta)|\big) < \varepsilon.$$

Ainsi pour tout $n \geqslant n_0$, $|P(Y_n \leqslant x) - P(Y \leqslant x)| < 2\varepsilon$. Comme $\varepsilon > 0$ était arbitraire, la convergence de $F_n(x)$ vers $F(x)$ est établie. □

La définition 10.2 de la convergence en loi par les moments fonctionnels $\mathbf{E}h(Y_n)$ pour *toute fonction h continue bornée* est bien commode pour démontrer des propriétés de cette convergence, cf. la preuve de la proposition 10.6. En revanche, quand on veut vérifier pratiquement qu'une suite donnée de v.a. ou de vecteurs aléatoires converge en loi, on a tout intérêt à pouvoir réduire la classe des fonctions h considérée. Nous débutons l'étude de ce type de réduction par l'importante notion d'équitension.

Définition 10.34 (équitension)**.** *Une famille $\{P_i,\ i \in I\}$ de mesures de probabilité sur $\mathbb{R}^d$ est dite* équitendue *si*

$$\forall \varepsilon > 0,\ \exists K \text{ compact de } \mathbb{R}^d \text{ tel que} \quad \forall i \in I,\ P_i(\mathbb{R}^d \backslash K) < \varepsilon. \tag{10.38}$$

Une suite $(X_n)_{n\geqslant 1}$ de vecteurs aléatoires de $\mathbb{R}^d$ est dite équitendue si la suite des lois $(P_{X_n})_{n\geqslant 1}$ est équitendue au sens précédent. Ceci se traduit par

$$\forall \varepsilon > 0,\ \exists K \text{ compact de } \mathbb{R}^d \text{ tel que } \quad \forall n \geqslant 1,\ P(X_n \notin K) < \varepsilon, \tag{10.39}$$

ou de manière équivalente par

$$\forall \varepsilon > 0,\ \exists K \text{ compact de } \mathbb{R}^d \text{ tel que } \quad \forall n \geqslant 1,\ P(X_n \in K) > 1 - \varepsilon. \tag{10.40}$$

Remarque 10.35. On sait que tout compact est borné et qu'en dimension finie, la fermeture de tout ensemble borné est compacte. On pourrait donc sans dommage remplacer dans la définition ci-dessus K compact par K borné et même plus simplement remplacer K par $[-a,a]^d$ pour $a > 0$. L'importance de prendre vraiment K compact au lieu de borné n'apparaît que lorsqu'on étudie les vecteurs aléatoires en dimension infinie. En pratique, une condition nécessaire et suffisante d'équitension de $(X_n)_{n\geqslant 1}$ dans $\mathbb{R}^d$ est donc

$$\forall \varepsilon > 0,\ \exists a > 0, \quad \forall n \geqslant 1,\ P(\|X_n\| > a) < \varepsilon. \tag{10.41}$$

Bien entendu, le choix de la norme n'a ici aucune importance, autre que de confort, puisque toutes les normes sur $\mathbb{R}^d$ sont équivalentes.

Remarque 10.36. Une famille finie $\{P_i,\ i \in I\}$ de mesures de probabilité sur $\mathbb{R}^d$ est toujours équitendue. Vérifions le d'abord dans le cas où la famille est réduite à une seule mesure [18] P_0. Posons $C_k := [-k,k]^d$. Les C_k sont des compacts de $\mathbb{R}^d$ et la suite $(C_k)_{k\geqslant 1}$ est croissante de réunion $\mathbb{R}^d$. Donc $P_0(C_k) \uparrow 1$ quand k tend vers l'infini, par continuité séquentielle croissante de P_0. Pour tout $\varepsilon > 0$, on a alors un $k_0 = k_0(\varepsilon)$ tel que $P_0(C_{k_0}) > 1 - \varepsilon$, ce qui équivaut à $P_0(\mathbb{R}^d \backslash C_{k_0}) < \varepsilon$. Maintenant si I est un ensemble fini quelconque d'indices, on peut construire comme ci-dessus pour chaque $i \in I$ un $k_i = k_i(\varepsilon)$ tel que $P_i(C_{k_i}) > 1 - \varepsilon$. En prenant le plus grand élément de cet ensemble fini d'entiers k_i, appelons le m, tous les C_{k_i} sont inclus dans le compact C_m et on a $P_i(C_m) > 1 - \varepsilon$ pour tout $i \in I$, la famille est donc équitendue. Évidemment ce raisonnement s'écroule si I est infini, car alors rien ne garantit que m, défini maintenant comme le sup des $k_i(\varepsilon)$, soit fini. Or si m vaut $+\infty$, $C_m = \mathbb{R}^d$ n'est plus compact.

Remarque 10.37. La réunion de deux familles équitendues $\{P_i,\ i \in I\}$ et $\{Q_j,\ j \in J\}$ de mesures de probabilité sur $\mathbb{R}^d$ est encore équitendue. En effet si K et K' sont des compacts tels que pour tout i, $P_i(K) > 1 - \varepsilon$ et pour tout j, $Q_j(K') > 1 - \varepsilon$, alors $K \cup K'$ est un compact vérifiant $P_i(K \cup K') > 1 - \varepsilon$ et $Q_j(K \cup K') > 1 - \varepsilon$ pour tout $i \in I$ et tout $j \in J$. En particulier, compte-tenu de la remarque 10.36, si $(X_n)_{n\geqslant 1}$ est une suite équitendue de vecteurs aléatoires de $\mathbb{R}^d$ et X un vecteur aléatoire *quelconque* de $\mathbb{R}^d$, la famille $\{X_n,\ n \geqslant 1\} \cup \{X\}$ est équitendue.

18. On dit alors que cette mesure est « tendue », le préfixe « équi » sous-entendant qu'il y a plusieurs mesures dans la famille. De même on parle de fonction continue et de famille équicontinue de fonctions.

Voici maintenant une première utilisation de l'équitension pour établir un résultat intermédiaire.

Lemme 10.38. *Soient Z_n ($n \geqslant 1$) et Z des variables aléatoires réelles vérifiant*

i) $(Z_n)_{n\geqslant 1}$ *est équitendue,*

ii) pour toute fonction $g : \mathbb{R} \to \mathbb{R}$ *continue* à support compact [19]*,* $\mathbf{E}g(Z_n)$ *converge vers* $\mathbf{E}g(Z)$ *quand* n *tend vers l'infini.*

Alors Z_n converge en loi vers Z.

Preuve. L'hypothèse ii) est *a priori* plus faible que la convergence en loi. En effet toute fonction continue à support compact est évidemment continue bornée mais la réciproque est fausse. Il nous faut montrer que pour toute fonction *continue bornée* $h : \mathbb{R} \to \mathbb{R}$, $\mathbf{E}h(Z_n)$ converge vers $\mathbf{E}h(Z)$ quand n tend vers l'infini. Soit donc h une telle fonction et notons $\|h\|_\infty := \sup_{x\in\mathbb{R}} |h(x)|$. Fixons un $\varepsilon > 0$ arbitraire. Par l'hypothèse i) et la remarque 10.37, la famille $\{P_{Z_n},\ n \geqslant 1\} \cup \{P_Z\}$ est équitendue. On dispose donc d'un réel $a > 0$ tel que

$$\forall n \geqslant 1, \quad P(|Z_n| > a) < \varepsilon \quad \text{et} \quad P(|Z| > a) < \varepsilon. \tag{10.42}$$

Notons f_a la fonction « trapèze » définie [20] par la figure 10.4 et posons $g := hf_a$.

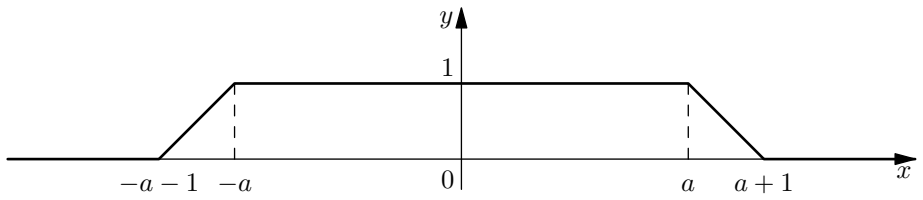


FIGURE 10.4 – Fonction « trapèze » f_a

Alors g est continue comme produit de deux fonctions continues et nulle en dehors du compact $[-a-1, a+1]$. L'erreur commise en approximant $\mathbf{E}h(Z_n)$ par $\mathbf{E}g(Z_n)$ est égale par linéarité de l'espérance à $\mathbf{E}(h-g)(Z_n)$. Pour majorer cette quantité, on remarque que $|h-g|(Z_n) \leqslant \|h\|_\infty \mathbf{1}_{\{|Z_n|>a\}}$, car $0 \leqslant |h-g| = |h|(1-f_a) \leqslant |h|$ et $h-g$ est nulle sur $[-a, a]$. Par conséquent, compte-tenu de (10.42),

$$|\mathbf{E}(h-g)(Z_n)| \leqslant \|h\|_\infty \mathbf{E}\mathbf{1}_{\{|Z_n|>a\}} = \|h\|_\infty P(|Z_n| > a) < \|h\|_\infty \varepsilon. \tag{10.43}$$

Il est clair que la même inégalité est vérifiée avec Z à la place de Z_n.

D'autre part grâce à l'hypothèse ii), il existe un n_0 dépendant seulement de g et de ε tel que

$$\forall n \geqslant n_0, \quad |\mathbf{E}g(Z_n) - \mathbf{E}g(Z)| < \varepsilon. \tag{10.44}$$

19. Donc continue sur $\mathbb{R}$ et nulle en dehors d'un certain compact de $\mathbb{R}$.

20. Pour ceux qui pensent qu'une figure ne définit pas une fonction et qui préfèrent les formules obscures, disons $f_a(x) := \mathbf{1}_{[0,a]}(|x|) + (a+1-|x|)\mathbf{1}_{]a,a+1]}(|x|)$.

Finalement en utilisant (10.43) appliquée à Z_n et à Z et (10.44), on voit que pour tout $n \geqslant n_0$,

$$\begin{aligned} |\mathbf{E}h(Z_n) - \mathbf{E}h(Z)| &\leqslant |\mathbf{E}(h-g)(Z_n)| + |\mathbf{E}(h-g)(Z)| + \big|\mathbf{E}g(Z_n) - \mathbf{E}g(Z)\big| \\ &\leqslant (2\|h\|_\infty + 1)\varepsilon. \end{aligned}$$

Comme ε était arbitraire, la convergence de $\mathbf{E}h(Z_n)$ vers $\mathbf{E}h(Z)$ est établie. □

Proposition 10.39. *Soient Z_n ($n \geqslant 1$) et Z des variables aléatoires réelles telles que pour toute fonction $g : \mathbb{R} \to \mathbb{R}$ continue* à support compact, *$\mathbf{E}g(Z_n)$ converge vers $\mathbf{E}g(Z)$ quand n tend vers l'infini. Alors Z_n converge en loi vers Z.*

Preuve. D'après le lemme 10.38, il suffit de prouver l'équitension de $(Z_n)_{n\geqslant 1}$. Fixons $\varepsilon > 0$ arbitraire. La variable aléatoire Z étant tendue (remarque 10.36), il existe un réel $a > 0$ tel que $P(|Z| \leqslant a) > 1 - \varepsilon$. La f.d.r. de $|Z|$ n'ayant qu'un ensemble au plus dénombrable de points de discontinuité, a au moins un point de continuité c dans l'ensemble non dénombrable $[a, +\infty[$. On a alors

$$P(|Z| \leqslant c) > 1 - \varepsilon \quad \text{et} \quad \lim_{x \to c} P(|Z| \leqslant x) = P(|Z| \leqslant c). \tag{10.45}$$

Pour $0 < \delta < c$, soit g_δ la fonction trapèze nulle en dehors de $[-c, c]$, valant 1 sur $[-c+\delta, c-\delta]$, cf. figure 10.5, donc vérifiant l'encadrement $\mathbf{1}_{[-c+\delta,c-\delta]} \leqslant g_\delta \leqslant \mathbf{1}_{[-c,c]}$.

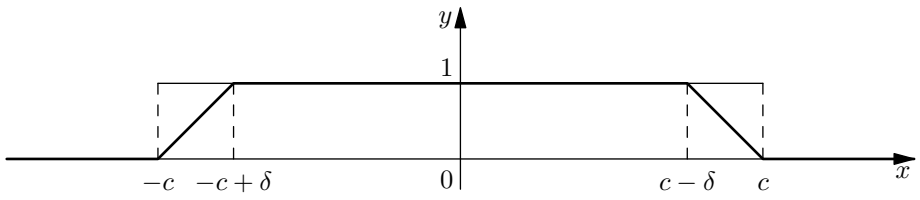


FIGURE 10.5 – Fonction « trapèze » g_δ vérifiant $\mathbf{1}_{[-c+\delta,c-\delta]} \leqslant g_\delta \leqslant \mathbf{1}_{[-c,c]}$

De cet encadrement on déduit que

$$P(|Z| \leqslant c - \delta) = \mathbf{E}\mathbf{1}_{[-c+\delta,c-\delta]}(Z) \leqslant \mathbf{E}g_\delta(Z) \leqslant \mathbf{E}\mathbf{1}_{[-c,c]}(Z) = P(|Z| \leqslant c).$$

D'après (10.45) on en déduit qu'en prenant δ suffisamment petit,

$$\mathbf{E}g_\delta(Z) \geqslant P(|Z| \leqslant c - \delta) > P(|Z| \leqslant c) - \varepsilon > 1 - 2\varepsilon. \tag{10.46}$$

Fixons désormais δ vérifiant (10.46). Comme g_δ est continue à support compact, l'hypothèse nous donne un n_0 tel que

$$\forall n \geqslant n_0, \quad \mathbf{E}g_\delta(Z_n) > \mathbf{E}g_\delta(Z) - \varepsilon. \tag{10.47}$$

Comme $g_\delta \leqslant \mathbf{1}_{[-c,c]}$, $P(|Z_n| \leqslant c) = \mathbf{E}\mathbf{1}_{[-c,c]}(Z_n) \geqslant \mathbf{E}g_\delta(Z_n)$, d'où en combinant cette minoration avec (10.46) et (10.47),

$$\forall n \geqslant n_0, \quad P(|Z_n| \leqslant c) > 1 - 3\varepsilon. \tag{10.48}$$

D'autre part la famille *finie* $\{P_{X_n},\ n < n_0\}$ est équitendue, cf. remarque 10.36, on peut donc trouver un réel $c' > 0$ tel que

$$\forall n < n_0, \quad P(|Z_n| \leqslant c') > 1 - 3\varepsilon. \tag{10.49}$$

Finalement en posant $c'' := \max(c, c')$, nous déduisons de (10.48) et (10.49) que

$$\forall n \in \mathbb{N}^*, \quad P(|Z_n| \leqslant c'') > 1 - 3\varepsilon.$$

Comme ε était arbitraire, ceci établit l'équitension de $(X_n)_{n\geqslant 1}$. □

Corollaire 10.40. *Soient Z_n ($n \geqslant 1$) et Z des variables aléatoires réelles telles que pour toute fonction $g : \mathbb{R} \to \mathbb{R}$ continue* tendant vers 0 à l'infini, *$\mathbf{E}g(Z_n)$ converge vers $\mathbf{E}g(Z)$ quand n tend vers l'infini. Alors Z_n converge en loi vers Z.*

Preuve. L'espace $\mathcal{C}_c(\mathbb{R})$ des fonctions continues à support compact est inclus dans l'espace $\mathcal{C}_0(\mathbb{R})$ des fonctions continues tendant vers 0 à l'infini. □

Corollaire 10.41. *Si $\mathbf{E}h(Z_n)$ converge quand n tend vers $+\infty$ vers $\mathbf{E}h(Z)$ pour toute $h \in \mathcal{H}$, où $\mathcal{H}$ est une famille dense dans $\mathcal{C}_0(\mathbb{R})$ pour la norme $\|\ \|_\infty$, alors Z_n converge en loi vers Z.*

Preuve. On utilise le corollaire 10.40 en montrant que pour toute $g \in \mathcal{C}_0(\mathbb{R})$, $\mathbf{E}g(Z_n)$ converge vers $\mathbf{E}g(Z)$. Soit donc g quelconque dans $\mathcal{C}_0(\mathbb{R})$. Par densité de $\mathcal{H}$ dans $\mathcal{C}_0(\mathbb{R})$, il existe pour tout $\varepsilon > 0$, une fonction $h \in \mathcal{H}$ telle que $\|h - g\|_\infty < \varepsilon$. Cette inégalité s'écrit encore $\sup_{x\in\mathbb{R}} |(g-h)(x)| < \varepsilon$, ce qui nous permet de voir que pour *toute v.a. réelle Y,*

$$|\mathbf{E}(g-h)(Y)| \leqslant \mathbf{E}|(g-h)(Y)| \leqslant \varepsilon. \tag{10.50}$$

En écrivant maintenant $\mathbf{E}g(Z_n) = \mathbf{E}h(Z_n) + \mathbf{E}(g-h)(Z_n)$ et de même avec Z au lieu de Z_n, en appliquant (10.50) avec $Y = Z_n$, puis avec $Y = Z$, on obtient

$$\forall n \in \mathbb{N}^*, \quad |\mathbf{E}g(Z_n) - \mathbf{E}g(Z)| \leqslant |\mathbf{E}h(Z_n) - \mathbf{E}h(Z)| + 2\varepsilon.$$

Par hypothèse $\mathbf{E}h(Z_n)$ converge vers $\mathbf{E}h(Z)$, donc pour $n \geqslant n_0$, $|\mathbf{E}h(Z_n) - \mathbf{E}h(Z)| < \varepsilon$. Finalement

$$\forall n \geqslant n_0, \quad |\mathbf{E}g(Z_n) - \mathbf{E}g(Z)| < 3\varepsilon.$$

Comme ε était arbitraire, on a bien vérifié la convergence de $\mathbf{E}g(Z_n)$ vers $\mathbf{E}g(Z)$. La fonction $g \in \mathcal{C}_0(\mathbb{R})$ était elle aussi arbitraire, on conclut donc par le corollaire 10.40 que Z_n converge en loi vers Z. □

Corollaire 10.42. *Si $\mathbf{E}f(Z_n)$ converge quand n tend vers $+\infty$ vers $\mathbf{E}f(Z)$ pour toute $f \in \mathcal{H}'$, où $\mathcal{H}'$ est une famille* totale *dans $\mathcal{C}_0(\mathbb{R})$ pour la norme $\|\ \|_\infty$, alors Z_n converge en loi vers Z.*

Preuve. Dire que $\mathcal{H}'$ est totale dans $\mathcal{C}_0(\mathbb{R})$ signifie que la famille $\mathcal{H}$ des combinaisons linéaires de fonctions de $\mathcal{H}'$ est dense dans $\mathcal{C}_0(\mathbb{R})$. Il suffit donc d'après le corollaire 10.41 de vérifier que $\mathbf{E}h(Z_n)$ converge vers $\mathbf{E}h(Z)$ pour toute h de la forme

$$h = \sum_{i\in I} a_i f_i, \quad I \text{ fini, les } a_i \in \mathbb{R}, \text{ les } f_i \in \mathcal{H}'.$$

Comme $\mathcal{H}'$ est incluse dans $\mathcal{C}_0(\mathbb{R})$, les f_i sont bornées donc les v.a. $f_i(Z_n)$ et $f_i(Z)$ sont intégrables. Par linéarité de l'espérance on a donc

$$\mathbf{E}h(Z_n) = \sum_{i\in I} a_i \mathbf{E}f_i(Z_n), \quad \mathbf{E}h(Z) = \sum_{i\in I} a_i \mathbf{E}f_i(Z).$$

Par hypothèse $\mathbf{E}f_i(Z_n)$ converge vers $\mathbf{E}f_i(Z)$ pour chaque $i \in I$. Comme I est fini, on en déduit que $\mathbf{E}h(Z_n)$ converge vers $\mathbf{E}h(Z)$. □

Corollaire 10.43. *Si $\mathbf{E}f(Z_n)$ converge quand n tend vers $+\infty$ vers $\mathbf{E}f(Z)$ pour toute $f \in \mathcal{C}_0^\infty(\mathbb{R})$, Z_n converge en loi vers Z.*

Preuve. L'espace $\mathcal{C}_0^\infty(\mathbb{R})$ des fonctions indéfiniment dérivables sur $\mathbb{R}$ et tendant vers 0 à l'infini est dense dans $\mathcal{C}_0(\mathbb{R})$, cf. cours d'analyse[21]. On peut donc appliquer le corollaire 10.41. □

Nous terminons cette sous-section outillage par une preuve partielle du lemme de Slutsky, basée sur l'emploi de moments fonctionnels pour une classe plus restreinte de fonctions que l'espace des fonctions continues bornées.

Preuve du lemme de Slutsky. En revenant à la définition de la convergence en loi des vecteurs aléatoires, il s'agit de montrer que si

$$Y_n \xrightarrow[n\to+\infty]{\text{loi}} Y \quad \text{et} \quad Z_n \xrightarrow[n\to+\infty]{\Pr} c, \ (c \text{ constante})$$

alors

$$\forall h \in \mathcal{C}_b(\mathbb{R}^2), \quad \mathbf{E}h(Y_n, Z_n) \xrightarrow[n\to+\infty]{} \mathbf{E}h(Y, c), \tag{10.51}$$

où $\mathcal{C}_b(\mathbb{R}^2)$ désigne l'espace des fonctions continues bornées sur $\mathbb{R}^2$. En fait il suffit de vérifier (10.51) pour des fonctions h de la forme particulière $h(y,z) = f(y)g(z)$ avec $f, g \in \mathcal{C}_b(\mathbb{R})$. Nous admettrons ce point qui pourrait se démontrer en utilisant l'idée de la preuve du corollaire 10.42 combinée avec un corollaire du théorème de Stone Weierstrass, après compactification de $\mathbb{R}^2$ par l'ajout d'un seul point à l'infini. On réduit ainsi la preuve du lemme de Slutsky à celle de la convergence suivante :

$$\forall f \in \mathcal{C}_b(\mathbb{R}), \ \forall g \in \mathcal{C}_b(\mathbb{R}) \quad \mathbf{E}\big(f(Y_n)g(Z_n)\big) \xrightarrow[n\to+\infty]{} \mathbf{E}\big(f(Y)g(c)\big). \tag{10.52}$$

21. On peut le démontrer soit en faisant de la régularisation par convolution, soit en utilisant le théorème de Stone-Weierstrass après compactification de $\mathbb{R}$ par l'ajout d'un seul point à l'infini.

Comme f et g sont bornées, les v.a. $f(Y_n)$, $f(Y)$, $g(Z_n)$ ainsi que leurs produits et leurs différences le sont aussi. Ceci légitime l'existence de toutes les espérances figurant dans les calculs ci-dessous. En utilisant la linéarité de l'espérance et la constance de $g(c)$, on peut écrire :

$$\begin{aligned}\mathbf{E}\big(f(Y_n)g(Z_n)\big) - \mathbf{E}\big(f(Y)g(c)\big) &= \mathbf{E}\Big(f(Y_n)\big(g(Z_n)-g(c)\big) + g(c)\big(f(Y_n)-f(Y)\big)\Big)\\ &= \mathbf{E}\Big(f(Y_n)\big(g(Z_n)-g(c)\big)\Big) + g(c)\big(\mathbf{E}f(Y_n)-\mathbf{E}f(Y)\big).\end{aligned}$$

On en déduit la majoration

$$\big|\mathbf{E}\big(f(Y_n)g(Z_n)\big) - \mathbf{E}\big(f(Y)g(c)\big)\big| \leqslant \|f\|_\infty \mathbf{E}|g(Z_n)-g(c)| + |g(c)|\big|\mathbf{E}f(Y_n)-\mathbf{E}f(Y)\big|.$$

Comme g est continue, la convergence en probabilité de Z_n vers c implique celle de $g(Z_n)$ vers $g(c)$, cf. l'exercice 9.4. Cette convergence en probabilité de $g(Z_n)$ vers $g(c)$ implique aussi sa convergence au sens L^1 puisque g est bornée (appliquer la proposition 9.19 avec comme constante majorante $\|g\|_\infty$). Par conséquent,

$$\|f\|_\infty \mathbf{E}|g(Z_n)-g(c)| \xrightarrow[n\to+\infty]{} 0. \tag{10.53}$$

D'autre part il résulte immédiatement de la définition de la convergence en loi de Y_n vers Y que

$$|g(c)|\big|\mathbf{E}f(Y_n)-\mathbf{E}f(Y)\big| \xrightarrow[n\to+\infty]{} 0. \tag{10.54}$$

De (10.53) et (10.54) on déduit la convergence (10.52), ce qui achève la preuve. □

10.4.2 Démonstration du TLC

Dans toutes les versions du TLC dans $\mathbb{R}$ exposées ci-dessus, on a une somme de variables aléatoires indépendantes que l'on centre et divise par l'écart type de la somme et c'est cette somme normalisée qui converge en loi vers une gaussienne $\mathfrak{N}(0,1)$. D'après le corollaire 10.43, il suffit de montrer la convergence des f-moments pour toute $f \in \mathcal{C}_0^\infty(\mathbb{R})$. Nous allons étudier l'erreur commise en remplaçant ces f-moments d'une somme par les f-moments de gaussiennes ayant même espérance et même variance que la somme [22]. Dans cette approche le lemme suivant est très utile.

Lemme 10.44. *Soit $f \in \mathcal{C}_0^3(\mathbb{R})$ et Y une v.a. réelle. Soient U, V des variables aléatoires réelles indépendantes de Y telles que $\mathbf{E}|U|^3 < +\infty$, $\mathbf{E}|V|^3 < +\infty$, $\mathbf{E}U = \mathbf{E}V$ et $\mathbf{E}U^2 = \mathbf{E}V^2$. Alors*

$$|\mathbf{E}f(Y+U) - \mathbf{E}f(Y+V)| \leqslant c(\mathbf{E}|U|^3 + \mathbf{E}|V|^3), \tag{10.55}$$

où c est une constante ne dépendant que de f.

22. La technique de preuve du TLC exposée dans cette section n'est pas la plus répandue dans la littérature où on utilise souvent les fonctions caractéristiques. Nous suivons d'assez près l'exposé donné dans le livre de D. Pollard [11].

Preuve. La formule de Taylor au point y pour un accroissement t s'écrit

$$f(y+t) = f(y) + tf'(y) + \frac{t^2}{2}f''(y) + r(y,t), \tag{10.56}$$

où le reste $r(y,t)$ est de la forme $r(y,t) = \frac{t^3}{6}f^{(3)}(y+\theta t)$ avec $\theta = \theta(f,y,t) \in [0,1]$. On a donc la majoration

$$|r(y,t)| \leqslant \frac{1}{6}\|f^{(3)}\|_\infty |t|^3. \tag{10.57}$$

En prenant $y = Y(\omega)$, $t = U(\omega)$ pour ω quelconque dans Ω, ces égalités et cette majoration passent aux variables aléatoires :

$$f(Y+U) = f(Y) + Uf'(Y) + \frac{U^2}{2}f''(Y) + r(Y,U), \tag{10.58}$$

$$|r(Y,U)| \leqslant \frac{1}{6}\|f^{(3)}\|_\infty |U|^3. \tag{10.59}$$

Les v.a. $f(Y+U)$, $f(Y)$, $f'(Y)$, $f''(Y)$ sont bornées et U est de cube intégrable, de plus U et Y sont indépendantes. Donc $\mathbf{E}|Uf'(Y)| = \mathbf{E}|U|\mathbf{E}|f'(Y)| < +\infty$, $\mathbf{E}|U^2 f''(Y)| = \mathbf{E}U^2\mathbf{E}|f''(Y)| < +\infty$. Dans (10.58) et (10.59), on peut alors prendre l'espérance des deux membres pour obtenir

$$\mathbf{E}f(Y+U) = \mathbf{E}f(Y) + \mathbf{E}U\mathbf{E}f'(Y) + \frac{1}{2}\mathbf{E}U^2\mathbf{E}f''(Y) + \mathbf{E}r(Y,U), \tag{10.60}$$

avec

$$\mathbf{E}|r(Y,U)| \leqslant \frac{1}{6}\|f^{(3)}\|_\infty \mathbf{E}|U|^3. \tag{10.61}$$

Évidemment (10.60) et (10.61) restent vérifiées avec V à la place de U. On en déduit que

$$\mathbf{E}f(Y+U) - \mathbf{E}f(Y+V) = (\mathbf{E}U - \mathbf{E}V)\mathbf{E}f'(Y) + \frac{1}{2}(\mathbf{E}U^2 - \mathbf{E}V^2)\mathbf{E}f''(Y) + \mathbf{E}r(Y,U) - \mathbf{E}r(Y,V),$$

d'où compte tenu des égalités de moments $\mathbf{E}U = \mathbf{E}V$ et $\mathbf{E}U^2 = \mathbf{E}V^2$,

$$|\mathbf{E}f(Y+U) - \mathbf{E}f(Y+V)| = |\mathbf{E}r(Y,U) - \mathbf{E}r(Y,V)| \leqslant \frac{1}{6}\|f^{(3)}\|_\infty (\mathbf{E}|U|^3 + \mathbf{E}|V|^3),$$

ce qui nous donne (10.55) avec $c = \frac{1}{6}\|f^{(3)}\|_\infty$. □

Le lemme 10.44 permet de majorer l'erreur commise en approximant le f-moment d'une somme centrée normalisée de v.a. indépendantes par le f-moment de la loi gaussienne $\mathfrak{N}(0,1)$.

Lemme 10.45. *Soient $X_1, \dots, X_n$ des v.a. indépendantes vérifiant $\mathbf{E}|X_i|^3 < +\infty$, $\mathbf{E}X_i = 0$, $\sum_{i=1}^n \sigma_i^2 = 1$ avec $\sigma_i^2 := \mathbf{E}X_i^2$. On note $S_n^* := \sum_{i=1}^n X_i$. Soient $Z_1, \dots, Z_n$ des v.a. telles que Z_i ait pour loi $\mathfrak{N}(0, \sigma_i)$ et que la suite $X_1, \dots, X_n, Z_1, \dots, Z_n$ soit indépendante. Soit Z de loi $\mathfrak{N}(0,1)$. Alors pour toute $f \in \mathcal{C}_0^3(\mathbb{R})$,*

$$|\mathbf{E}f(S_n^*) - \mathbf{E}f(Z)| \leqslant c \sum_{i=1}^n \big(\mathbf{E}|X_i|^3 + \mathbf{E}|Z_i|^3\big), \tag{10.62}$$

avec $c = \frac{1}{6}\|f^{(3)}\|_\infty$.

Preuve. Notons d'abord que Z a même loi que $W_n := Z_1 + \cdots + Z_n$. En effet par indépendance des Z_i, le vecteur $(Z_1, \dots, Z_n)$ est gaussien, cf. corollaire 10.26. Donc la combinaison linéaire $W_n = Z_1 + \cdots + Z_n$ est une v.a. gaussienne. Sa loi est déterminée par son espérance qui vaut 0 puisque les Z_i sont centrées et sa variance $\sigma^2 = \operatorname{Var} W_n = \sum_{i=1}^n \sigma_i^2 = 1$ à cause de l'indépendance des Z_i. Donc W_n suit comme Z, la loi $\mathfrak{N}(0,1)$. Par conséquent, pour toute $f \in \mathcal{C}_0^3(\mathbb{R})$, $\mathbf{E}f(Z) = \mathbf{E}f(W_n)$.

L'idée de la preuve est de faire du chaînage pour passer progressivement de $\mathbf{E}f(S_n^*)$ à $\mathbf{E}f(W_n)$ en remplaçant à chaque pas un X_i par un Z_i. Le lemme 10.45 permet de majorer l'erreur d'approximation commise à chaque pas. Définissons les sommes « trouées » T_j (T_j trouée au rang j) en posant :

$$T_j := X_1 + \cdots + X_{j-1} + Z_{j+1} + \cdots + Z_n, \quad j \in [\![1, n]\!].$$

Pour bien comprendre le mécanisme de chaînage que nous allons utiliser, voyons le à l'oeuvre dans le cas $n = 3$. On a alors $T_1 = Z_2 + Z_3$, $T_2 = X_1 + Z_3$ et $T_3 = X_1 + X_2$. On part de la somme téléscopique :

$$\begin{aligned} f(X_1 + X_2 + X_3) - f(Z_1 + Z_2 + Z_3) \quad = \quad & f(X_1 + X_2 + X_3) - f(X_1 + X_2 + Z_3) \\ & + f(X_1 + X_2 + Z_3) - f(X_1 + Z_2 + Z_3) \\ & + f(X_1 + Z_2 + Z_3) - f(Z_1 + Z_2 + Z_3), \end{aligned}$$

que l'on réécrit sous la forme

$$\begin{aligned} f(X_1 + X_2 + X_3) - f(Z_1 + Z_2 + Z_3) \quad = \quad & f(T_3 + X_3) - f(T_3 + Z_3) \\ & + f(T_2 + X_2) - f(T_2 + Z_2) \\ & + f(T_1 + X_1) - f(T_1 + Z_1), \end{aligned}$$

en utilisant les égalités $T_2 + X_2 = T_3 + Z_3$ et $T_1 + X_1 = T_2 + Z_2$.

Pour le cas général, il est commode de poser $T_0 := W_n$ et $X_0 := 0$. On note alors que

$$\forall j \in [\![1, n]\!], \quad T_j + Z_j = T_{j-1} + X_{j-1}. \tag{10.63}$$

On peut alors écrire la somme téléscopique

$$f(S_n^*) - f(W_n) = f(T_n + X_n) - f(T_0 + X_0) = \sum_{j=1}^n \big(f(T_j + X_j) - f(T_{j-1} + X_{j-1})\big)$$

En prenant l'espérance et en utilisant (10.63), il vient

$$\mathbf{E}f(S_n^*) - \mathbf{E}f(W_n) = \sum_{j=1}^{n} \big(\mathbf{E}f(T_j + X_j) - \mathbf{E}f(T_j + Z_j)\big).$$

Pour tout j, X_j et Z_j sont indépendantes de T_j, sont de cube intégrable et ont mêmes moments d'ordre 1 et 2. On peut donc appliquer le lemme 10.44 avec $Y = T_j$, $U = X_j$ et $V = Z_j$ à chacun des termes de la somme ci-dessus, ce qui donne

$$|\mathbf{E}f(S_n^*) - \mathbf{E}f(W_n)| \leqslant \sum_{j=1}^{n} \big|\mathbf{E}f(T_j + X_j) - \mathbf{E}f(T_j + Z_j)\big| \leqslant c \sum_{j=1}^{n} (\mathbf{E}|X_j|^3 + \mathbf{E}|Z_j|^3),$$

avec $c = \frac{1}{6}\|f^{(3)}\|_\infty$. □

Le lemme 10.45 permet de démontrer une version du théorème de Liapounov pour des v.a. de cube intégrable et d'expliciter dans le cas i.i.d. une vitesse de convergence en $O(n^{-1/2})$ pour les f-moments.

Théorème 10.46 (Liapounov)**.** *Soit $(X_k)_{k\geqslant 1}$ une suite de variables aléatoires définies sur le même espace probabilisé, indépendantes (mais pas forcément de même loi), centrées, ayant toutes un moment d'ordre 3 ($\mathbf{E}|X_k|^3 < +\infty$). On note $S_n := X_1 + \cdots + X_n$, $\sigma_k^2 := \operatorname{Var} X_k$, $s_n^2 := \operatorname{Var}(S_n) = \sigma_1^2 + \cdots + \sigma_n^2$ et on suppose $s_n > 0$. Si de plus la* condition de Liapounov *:*

$$\frac{1}{s_n^3} \sum_{k=1}^{n} \mathbf{E}|X_k|^3 \xrightarrow[n\to+\infty]{} 0 \tag{10.64}$$

est vérifiée, alors S_n/s_n converge en loi vers Z de loi $\mathfrak{N}(0,1)$.

Dans le cas où les X_k sont de plus i.i.d., on a pour toute $f \in \mathcal{C}_0^3(\mathbb{R})$,

$$|\mathbf{E}f(S_n/s_n) - \mathbf{E}f(Z)| \leqslant Cn^{-1/2}, \tag{10.65}$$

où la constante C vaut

$$C = \frac{1}{6}\left(\frac{\mathbf{E}|X_1|^3}{\sigma_1^3} + \frac{4}{\sqrt{2\pi}}\right)\|f^{(3)}\|_\infty. \tag{10.66}$$

Preuve. D'après le corollaire 10.43 et l'inclusion $\mathcal{C}_0^\infty(\mathbb{R}) \subset \mathcal{C}_0^3(\mathbb{R})$, il suffit de montrer la convergence des f-moments de S_n/s_n vers ceux de Z pour toute $f \in \mathcal{C}_0^3(\mathbb{R})$ pour obtenir la convergence en loi de S_n/s_n vers Z. Posons $X_{n,k} := s_n^{-1}X_k$, $\sigma_{n,k}^2 := \operatorname{Var} X_{n,k} = s_n^{-2}\mathbf{E}X_k^2$. En appliquant le lemme 10.45 aux $X_{n,k}$, notons qu'alors $S_n^* = S_n/s_n$, on obtient :

$$\forall f \in \mathcal{C}_0^3(\mathbb{R}), \quad |\mathbf{E}f(S_n/s_n) - \mathbf{E}f(Z)| \leqslant c \sum_{k=1}^{n} \left(\frac{\mathbf{E}|X_k|^3}{s_n^3} + \left(\frac{\sigma_k}{s_n}\right)^3 \mathbf{E}|Z|^3\right). \tag{10.67}$$

En effet la gaussienne centrée $Z_{n,k}$ de variance $\sigma_{n,k}^2$ a même loi que $\sigma_{n,k}Z$, donc même moment absolu d'ordre 3, d'où $\mathbf{E}|Z_{n,k}|^3 = \sigma_{n,k}^3\mathbf{E}|Z|^3 = \sigma_k^3 s_n^{-3}\mathbf{E}|Z|^3$. Compte-tenu de la condition de Liapounov (10.64) et de (10.67), il ne nous reste plus pour obtenir la convergence en loi qu'à vérifier la convergence :

$$\frac{1}{s_n^3}\sum_{k=1}^{n}\sigma_k^3 \xrightarrow[n\to+\infty]{} 0. \tag{10.68}$$

Pour pouvoir exploiter la condition de Liapounov, on souhaite disposer d'une majoration du type $\sigma_k^3 \leqslant a\mathbf{E}|X_k|^3$, avec a constante, qui ferait immédiatement découler (10.68) de (10.64). L'inégalité de Jensen, cf. exercice 7.14, nous dit que si φ est une fonction convexe sur $\mathbb{R}$ et Y une v.a. telle que Y et $\varphi(Y)$ soient intégrables, $\varphi(\mathbf{E}Y) \leqslant \mathbf{E}\varphi(Y)$. En appliquant ceci à la v.a. $Y = X^2$ et avec la fonction convexe $\varphi : y \mapsto |y|^{3/2}$, on obtient pour toute v.a. de cube intégrable :

$$\big(\mathbf{E}(X^2)\big)^{3/2} \leqslant \mathbf{E}\big((X^2)^{3/2}\big) = \mathbf{E}\big(|X|^3\big).$$

On en déduit pour chaque X_k la majoration annoncée avec $a = 1$ puisque $\sigma_k = (\mathbf{E}(X_k^2))^{1/2}$.

Voici une variante « manuelle » évitant le recours à l'inégalité de Jensen au prix d'une moins bonne constante a. On commence par vérifier que pour tout $t > 0$, on a

$$\mathbf{E}X^2 = \mathbf{E}\big(X^2\mathbf{1}_{\{|X|\leqslant t\}}\big) + \mathbf{E}\big(X^2\mathbf{1}_{\{|X|>t\}}\big) \leqslant t^2 + \frac{\mathbf{E}|X|^3}{t}. \tag{10.69}$$

En effet $\mathbf{E}\big(X^2\mathbf{1}_{\{|X|\leqslant t\}}\big)$ se majore naturellement par $t^2P(|X| \leqslant t) \leqslant t^2$. Pour le deuxième terme, on remarque que sur l'évènement $\{|X| > t\}$, on a $|X|^3 > t|X|^2$, d'où $|X|^2 < t^{-1}|X|^3$. Cette inégalité vraie seulement sur $\{|X| > t\}$ implique l'inégalité entre variables aléatoires, vraie sur tout Ω :

$$X^2\mathbf{1}_{\{|X|>t\}} \leqslant \frac{1}{t}|X|^3\mathbf{1}_{\{|X|>t\}} \leqslant \frac{1}{t}|X|^3,$$

puis en prenant les espérances, $\mathbf{E}\big(X^2\mathbf{1}_{\{|X|>t\}}\big) \leqslant t^{-1}\mathbf{E}|X|^3$. Puisque (10.69) est ainsi vérifiée pour tout $t > 0$, on a intérêt à choisir la valeur de t qui minimise le majorant obtenu. Un simple calcul de dérivée nous montre que la fonction $g : t \mapsto t^2 + \frac{\mathbf{E}|X|^3}{t}$ atteint son minimum sur $]0, +\infty[$ au point $t = 2^{-1/3}\big(\mathbf{E}|X|^3\big)^{1/3}$ et donc que la valeur minimale de g est $3 \times 2^{-2/3}\big(\mathbf{E}|X|^3\big)^{2/3}$. On en déduit l'inégalité $\sigma_k^3 \leqslant 3^{3/2} \times 2^{-1}\mathbf{E}|X|^3$, ce qui est moins bon que l'inégalité de Jensen puisque ici $a = 3^{3/2} \times 2^{-1} \simeq 2{,}599$.

Enfin dans le cas i.i.d., on a $s_n = n^{1/2}\sigma_1$. En revenant à (10.67) et en notant que $\mathbf{E}|Z|^3 = 4(2\pi)^{-1/2}$ (vérifiez), on obtient :

$$|\mathbf{E}f(S_n/s_n) - \mathbf{E}f(Z)| \leqslant \frac{c}{\sigma_1^3 n^{3/2}}\big(n\mathbf{E}|X_1|^3 + n\sigma_1^3\mathbf{E}|Z|^2\big) = \frac{c}{n^{1/2}}\left(\frac{\mathbf{E}|X_1|^3}{\sigma_1^3} + \frac{4}{\sqrt{2\pi}}\right),$$

d'où (10.65) et (10.66). □

Jusqu'ici nous n'avons prouvé le TLC qu'en supposant l'existence de moments d'ordre 3. Nous allons maintenant nous affranchir de cette hypothèse en revenant à la formule de Taylor (10.56) et en notant que le reste $r(y,t)$ s'écrit aussi :

$$r(y,t) = \big(f(y+t) - f(y) - tf'(y)\big) - \frac{t^2}{2}f''(y) = \frac{t^2}{2}\big(f''(y+\tau t) - f''(y)\big),$$

pour un certain $\tau = \tau(f,y,t) \in [0,1]$. Par conséquent,

$$\forall y,t \in \mathbb{R}, \quad |r(y,t)| \leqslant t^2 \|f''\|_\infty. \tag{10.70}$$

Cette nouvelle majoration de $|r(y,t)|$ est meilleure que (10.57) lorsque $|t|$ tend vers l'infini, en revanche elle est moins bonne lorsque t tend vers 0. Pour garder le meilleur des deux majorations, on fait un compromis en utilisant (10.57) pour $|t| < \varepsilon$ et (10.70) pour $|t| \geqslant \varepsilon$, ce qui donne pour tout $\varepsilon > 0$,

$$\forall y,t \in \mathbb{R}, \quad |r(y,t)| \leqslant C|t|^3 \mathbf{1}_{\{|t|<\varepsilon\}} + Ct^2 \mathbf{1}_{\{|t|\geqslant\varepsilon\}} \leqslant C\varepsilon t^2 \mathbf{1}_{\{|t|<\varepsilon\}} + Ct^2 \mathbf{1}_{\{|t|\geqslant\varepsilon\}}, \tag{10.71}$$

avec

$$C = \max\Big(\|f''\|_\infty, \frac{1}{6}\|f^{(3)}\|_\infty\Big). \tag{10.72}$$

Par conséquent, pour toute v.a. Y et toute v.a. U de carré intégrable,

$$\mathbf{E}|r(Y,U)| \leqslant C\varepsilon \mathbf{E}\big(U^2 \mathbf{1}_{\{|U|<\varepsilon\}}\big) + C\mathbf{E}\big(U^2 \mathbf{1}_{\{|U|\geqslant\varepsilon\}}\big) \leqslant C\varepsilon \mathbf{E}U^2 + C\mathbf{E}\big(U^2 \mathbf{1}_{\{|U|\geqslant\varepsilon\}}\big).$$

Reprenant la preuve des lemmes (10.44) et (10.45) avec cette nouvelle majoration de $\mathbf{E}|r(Y,U)|$, on obtient la clé de la preuve du théorème de Lindeberg.

Lemme 10.47. *Soient $X_1, \dots, X_n$ des v.a. indépendantes vérifiant $\mathbf{E}X_i^2 < +\infty$, $\mathbf{E}X_i = 0$, $\sum_{i=1}^n \sigma_i^2 = 1$ avec $\sigma_i^2 := \mathbf{E}X_i^2$. On note $S_n^* := \sum_{i=1}^n X_i$. Soient $Z_1, \dots, Z_n$ des v.a. telles que Z_i ait pour loi $\mathfrak{N}(0,\sigma_i)$ et que la suite $X_1, \dots, X_n, Z_1, \dots, Z_n$ soit indépendante. Soit Z de loi $\mathfrak{N}(0,1)$. Alors pour toute $f \in \mathcal{C}_0^3(\mathbb{R})$ et tout $\varepsilon > 0$,*

$$|\mathbf{E}f(S_n^*) - \mathbf{E}f(Z)| \leqslant C \sum_{i=1}^n \Big(\varepsilon\sigma_i^2 + \mathbf{E}\big(X_i^2 \mathbf{1}_{\{|X_i|\geqslant\varepsilon\}}\big) + \mathbf{E}|Z_i|^3\Big), \tag{10.73}$$

avec C donnée par (10.72).

Notez le traitement dissymétrique des X_i et des Z_i, dû au fait que les Z_i étant gaussiennes ont des moments d'ordre 3. Il est alors plus commode de leur appliquer (10.57) pour contrôler $|r(Y,Z_i)|$, tandis que pour les X_i, on tient à n'utiliser que des majorations valables même si ces variables sont seulement de carré intégrable.

Preuve du théorème de Lindeberg. Avec les notations et les hypothèses du th. 10.20, il suffit de montrer la convergence de $\mathbf{E}f(S_n/s_n)$ vers $\mathbf{E}f(Z)$ pour toute $f \in \mathcal{C}_0^3(\mathbb{R})$.

En appliquant le lemme (10.47) aux k_n variables $X_i = X_{n,i}/s_n$, on dispose de la majoration suivante valable pour tout $\varepsilon > 0$:

$$|\mathbf{E}f(S_n/s_n) - \mathbf{E}f(Z)| \leqslant C \sum_{i=1}^{k_n} \Big(\varepsilon \frac{\sigma_{n,i}^2}{s_n^2} + \frac{1}{s_n^2} \mathbf{E}\big(X_{n,i}^2 \mathbf{1}_{\{|X_{n,i}|>\varepsilon s_n\}}\big) + \frac{1}{s_n^3} \mathbf{E}|Z_{n,i}|^3 \Big)$$
$$= C\varepsilon + \frac{C}{s_n^2} \sum_{i=1}^{k_n} \mathbf{E}\big(X_{n,i}^2 \mathbf{1}_{\{|X_{n,i}|>\varepsilon s_n\}}\big) + \frac{1}{s_n^3} \sum_{i=1}^{k_n} \mathbf{E}|Z_{n,i}|^3. \quad (10.74)$$

Dans ce majorant, le deuxième terme tend vers 0 quand n tend vers l'infini par la condition de Lindeberg, cf. hypothèse (10.23). Il peut donc être rendu inférieur à ε pour tout $n \geqslant n_1$. Pour le troisième terme, on rappelle que la v.a. gaussienne $Z_{n,i}$ a même loi que $\sigma_{n,i}Z$, d'où

$$\frac{1}{s_n^3} \sum_{i=1}^{k_n} \mathbf{E}|Z_{n,i}|^3 = \frac{\mathbf{E}|Z|^3}{s_n^3} \sum_{i=1}^{k_n} \sigma_{n,i}^3 \leqslant \frac{\mathbf{E}|Z|^3}{s_n^3} \Big(\max_{1\leqslant i\leqslant k_n} \sigma_{n,i} \Big) \sum_{i=1}^{k_n} \sigma_{n,i}^2 = \frac{4}{\sqrt{2\pi}} \frac{1}{s_n} \max_{1\leqslant i\leqslant k_n} \sigma_{n,i}.$$

Il nous reste à montrer la convergence vers 0 de $s_n^{-1} \max_{1\leqslant i\leqslant k_n} \sigma_{n,i}$, ce qui équivaut à celle de $s_n^{-2} \max_{1\leqslant i\leqslant k_n} \sigma_{n,i}^2$. On se rapelle opportunément que $\sigma_{n,i}^2 = \mathbf{E}X_{n,i}^2$, ce qui va nous permettre de réexploiter la condition de Lindeberg. En effet

$$\mathbf{E}X_{n,i}^2 = \mathbf{E}\big(X_{n,i}^2 \mathbf{1}_{\{|X_{n,i}|<s_n\varepsilon\}}\big) + \mathbf{E}\big(X_{n,i}^2 \mathbf{1}_{\{|X_{n,i}|\geqslant s_n\varepsilon\}}\big) \leqslant \varepsilon^2 s_n^2 + \mathbf{E}\big(X_{n,i}^2 \mathbf{1}_{\{|X_{n,i}|\geqslant s_n\varepsilon\}}\big).$$

D'où

$$\begin{aligned} \frac{1}{s_n^2} \max_{1\leqslant i\leqslant k_n} \sigma_{n,i}^2 &\leqslant \varepsilon^2 + \frac{1}{s_n^2} \max_{1\leqslant i\leqslant k_n} \mathbf{E}\big(X_{n,i}^2 \mathbf{1}_{\{|X_{n,i}|\geqslant s_n\varepsilon\}}\big) \\ &\leqslant \varepsilon^2 + \frac{1}{s_n^2} \sum_{i=1}^{k_n} \mathbf{E}\big(X_{n,i}^2 \mathbf{1}_{\{|X_{n,i}|\geqslant s_n\varepsilon\}}\big). \end{aligned}$$

Le deuxième terme de ce majorant tend vers 0 quand n tend vers $+\infty$ par la condition de Lindeberg. Il est donc inférieur à ε^2 pour tout $n \geqslant n_2$. Par conséquent,

$$\forall n \geqslant n_2, \quad \frac{1}{s_n} \max_{1\leqslant i\leqslant k_n} \sigma_{n,i} \leqslant \varepsilon\sqrt{2}.$$

En prenant $n_0 := \max(n_1, n_2)$ et en revenant à (10.74), on obtient finalement

$$\forall n \geqslant n_0, \quad |\mathbf{E}f(S_n/s_n) - \mathbf{E}f(Z)| \leqslant \Big(C + 1 + \frac{4}{\sqrt{\pi}}\Big)\varepsilon.$$

Comme $\varepsilon > 0$ était arbitraire, on a bien établi la convergence de $\mathbf{E}f(S_n/s_n)$ vers $\mathbf{E}f(Z)$. Ceci étant vérifié pour toute $f \in \mathcal{C}_0^3(\mathbb{R})$, S_n/s_n converge en loi vers Z. □

Preuve du TLC avec autonormalisation. En toute rigueur, l'énoncé du théorème 10.21 est incorrect car il est possible que V_n prenne la valeur 0 avec probabilité non-nulle et

alors T_n n'est pas définie. Pour simplifier, nous supposerons dans un premier temps qu'il existe un rang n_0 non aléatoire, à partir duquel $\{V_n = 0\} = \varnothing$, en sorte que pour tout $n \geqslant n_0$, T_n est bien définie sur tout Ω et est une variable aléatoire. L'idée est de se ramener au TLC classique grâce au lemme de Slutsky. Pour cela on écrit :

$$T_n = \sqrt{\frac{n}{V_n}} \left(\frac{S_n}{n} - \mathbf{E}X_1 \right) = \sqrt{\frac{n}{\sigma^2}} \left(\frac{S_n}{n} - \mathbf{E}X_1 \right) \sqrt{\frac{\sigma^2}{V_n}} = S_n^* \sqrt{\frac{\sigma^2}{V_n}}.$$

Par le théorème 10.14, S_n^* converge en loi vers Y gaussienne $\mathfrak{N}(0,1)$.

On sait que si $(V_n)_{n\geqslant 1}$ converge en probabilité vers une constante c et si la fonction f définie sur un voisinage de c est continue au point c, alors $f(V_n)$ converge en probabilité vers $f(c)$, cf. l'exercice 9.4. En appliquant ceci à la fonction $f : x \mapsto \sigma x^{-1/2}$ qui est continue sur $\mathbb{R}^*$, donc en particulier au point $c = \sigma^2 > 0$, on obtient :

$$\sqrt{\frac{\sigma^2}{V_n}} \xrightarrow[n\to+\infty]{\Pr} 1. \tag{10.75}$$

Par application du lemme de Slutsky, plus précisément du corollaire 10.13 b), la convergence en loi de S_n^* vers Y et (10.75) impliquent la convergence en loi de T_n vers Y de loi $\mathfrak{N}(0,1)$.

Pour compléter la preuve, il nous reste à définir proprement T_n sans utiliser l'hypothèse $\{V_n = 0\} = \varnothing$ pour $n \geqslant n_0$ et à vérifier la convergence en loi de ce « nouveau » T_n vers Y. Posons donc pour tout $\omega \in \Omega$,

$$T_n(\omega) := \begin{cases} \sqrt{\frac{n}{V_n(\omega)}} \left(\frac{S_n(\omega)}{n} - \mathbf{E}X_1 \right) & \text{si } V_n(\omega) \neq 0, \\ 0 & \text{si } V_n(\omega) = 0. \end{cases}$$

Avec la convention habituelle sur les indicatrices dans les variables aléatoires, la valeur nulle d'un indicatrice s'impose toujours lorsqu'elle est facteur d'une quantité infinie ou non définie, en clair :

$$\sqrt{\frac{\sigma^2}{V_n}} \mathbf{1}_{\{V_n \neq 0\}} = \begin{cases} \sqrt{\frac{\sigma^2}{V_n}} & \text{sur } \{V_n \neq 0\}, \\ 0 & \text{sur } \{V_n = 0\}. \end{cases}$$

Avec cette convention, le T_n redéfini proprement ci-dessus peut s'écrire :

$$T_n = \frac{\sqrt{n}}{\sigma} \left(\frac{S_n}{n} - \mathbf{E}X_1 \right) \left(\sqrt{\frac{\sigma^2}{V_n}} \mathbf{1}_{\{V_n \neq 0\}} \right).$$

Il est alors immédiat d'adapter la démonstration donnée dans le cas où $\{V_n = 0\}$ est vide en notant que $\mathbf{1}_{\{V_n \neq 0\}}$ converge en probabilité vers 1. □

10.5 Exercices

Ex. 10.1 *Convergence en loi dans* $\mathbb{N}$

Soit $(X_n)_{n\geqslant 1}$ une suite de variables aléatoires à valeurs dans $\mathbb{N}$.

1) Montrez que si $(X_n)_{n\geqslant 1}$ converge en loi vers une variable aléatoire réelle X, alors $P(X \in \mathbb{N}) = 1$ et $P(X_n = k)$ converge vers $P(X = k)$ pour tout $k \in \mathbb{N}$. *Indication.* On pourra utiliser la définition fonctionnelle de la convergence en loi avec la fonction en dents de scie $g_j = \sum_{k\in\mathbb{N}} g_{j,k}$ où $g_{j,k}$ désigne la fonction triangle isocèle de base $[k - 2^{-j}, k + 2^{-j}]$ et de sommet $(k, 1)$.

2) On suppose désormais que pour tout $k \in \mathbb{N}$, $P(X_n = k)$ converge vers un réel a_k (nécessairement positif). Montrez que la série $\sum_{k\in\mathbb{N}} a_k$ converge et a une somme inférieure ou égale à 1.

3) Montrez que si $\sum_{k\in\mathbb{N}} a_k = 1$, alors X_n converge en loi vers une variable aléatoire X dont la loi P_X vérifie pour tout $k \in \mathbb{N}$, $P_X(\{k\}) = a_k$ et donc $P_X(\mathbb{N}) = 1$.

4) Montrez que si $\sum_{k\in\mathbb{N}} a_k < 1$, X_n ne peut converger en loi. Donnez un exemple de cette situation.

5) Reformulez les théorèmes 6.37 et 6.41 en termes de convergence en loi.

Solution p. 494

Ex. 10.2 *Questions élémentaires sur la convergence en loi*

1) Montrez que si X_n converge en loi vers une constante c, elle converge aussi en probabilité.

2) Soit $(U_n)_{n\geqslant 1}$ une suite de variables aléatoires où chaque U_n suit la loi uniforme sur l'ensemble fini $\{k/n,\ k \in [\![1, n]\!]\}$. Étudiez la convergence en loi de U_n, de deux façons : avec la définition par les f.d.r. et avec celle par les moments fonctionnels.

3) Soit $(U_n)_{n\geqslant 1}$ une suite de variables aléatoires où chaque U_n suit la loi uniforme sur l'intervalle $[1, 1 + 1/n]$. Étudiez la convergence en loi de U_n.

4) Soit $(Y_n)_{n\geqslant 1}$ une suite de variables aléatoires gaussiennes de loi $\mathfrak{N}(0, \sigma_n)$, avec σ_n tendant vers l'infini. Étudiez la convergence de la suite des fonctions de répartitions. Y a-t-il convergence en loi ?

Ex. 10.3 Soit $(c_n)_{n\geqslant 1}$ une suite de réels tendant vers $+\infty$, a une constante et $(Y_n)_{n\geqslant 1}$ une suite de variables aléatoires telle que $c_n(Y_n - a)$ converge en loi.

1) Montrez qu'une fonction de répartition quelconque a des points de continuité aussi grands que l'on veut en valeur absolue.

2) Montrez que Y_n converge en probabilité vers a.

3) Application : le théorème de de Moivre-Laplace implique la loi faible des grands nombres pour les fréquences.

Ex. 10.4 *Estimation d'un paramètre de loi uniforme*

Soit $(X_i)_{i\geqslant 1}$ une suite de variables aléatoires indépendantes de même loi uniforme sur $[0, \theta]$, $\theta > 0$. On pose $M_n = \max_{1\leqslant i\leqslant n} X_i$.

1) Montrez que M_n converge presque-sûrement vers θ.

2) Montrez que $n(\theta - M_n)$ converge en loi et déterminez la loi limite par sa f.d.r.

Ex. 10.5 *La cantine*

Une entreprise emploie 500 personnes qui déjeunent à la cantine à l'un ou l'autre des deux services avec une probabilité égale de manger au premier ou au second. Si le gérant veut avoir une probabilité supérieure à 95 % de disposer d'assez de couverts, combien devra-t-il en prévoir pour chacun des deux services ?

Indication. On numérote les 500 personnes par ordre alphabétique (ou n'importe quel ordre, *sauf* l'ordre d'arrivée au restaurant !). On note X_i la variable aléatoire valant 1 si la i^{e} personne de la liste mange au premier service, 0 si elle mange au deuxième. Exprimer l'évènement dont on cherche à minorer la probabilité à l'aide de $S_n = \sum_{i=1}^{500} X_i$.

Solution p. 496

Ex. 10.6 *Cumul d'erreurs*

Un programme de calcul utilise J chiffres significatifs après la virgule et arrondit tous les résultats d'opérations à ce format (donc à $\frac{1}{2} \times 10^{-J}$ près). On suppose qu'il effectue 10^6 opérations élémentaires successives, que les erreurs commises pour chacune sont indépendantes, de loi uniforme sur l'intervalle $\left[-\frac{1}{2} \times 10^{-J}; \frac{1}{2} \times 10^{-J}\right]$ et que l'erreur sur le résultat final est la somme des erreurs commises sur chaque opération. Évaluer la probabilité pour que l'erreur finale soit inférieure ou égale en valeur absolue à $\frac{1}{2} \times 10^{-J+3}$.

Ex. 10.7 *Surcharge au décollage*

Un avion a une charge maximum au décollage (hors kérosène) de 25 tonnes. Il embarque 318 personnes et leurs bagages (équipage compris). La masse X_i du i^{e} individu embarqué est une v.a. d'espérance 65 kg et d'écart-type 10 kg. La masse de ses bagages est une v.a. Y_i d'espérance 12 kg et d'écart-type 2 kg.

1) Évaluez la probabilité d'une surcharge au décollage en précisant les hypothèses d'indépendance que vous serez amenés à utiliser.

2) On rappelle que la somme de deux v.a. indépendantes Z_1 et Z_2 de lois gaussiennes respectives $\mathfrak{N}(m_1, \sigma_1)$ et $\mathfrak{N}(m_2, \sigma_2)$ est encore une v.a. gaussienne. Déterminez ses paramètres.

3) La charge maximale de 25 tonnes a été déterminée en tenant compte d'une masse volumique moyenne du kérosène de 0,8 kg/l. En réalité la masse volumique du kérosène peut varier entre 0,755 kg/l et 0,845 kg/l. Du fait des avitaillements [23] successifs d'un aéroport à l'autre, il est impossible de prévoir assez à l'avance la masse volumique du kérosène embarqué pour un vol donné. On est donc amené à modéliser cette masse volumique par une variable aléatoire gaussienne d'espérance 0,8 kg/l et d'écart-type 0,011 kg/l. Commentez ce choix.

4) Les données techniques concernant l'avion sont les suivantes. Masse maximale au décollage 269 tonnes, masse à vide 120 tonnes, volume des réservoirs de kérosène 152 500 litres. En supposant que la densité du kérosène et la masse des personnes embarquées et de leurs bagages sont indépendantes, recalculez la probabilité de surcharge au décollage.

23. Avitaillement : ravitaillement d'un avion en carburant.

Ex. 10.8 *Différence de sommes*

On suppose que les deux suites de variables aléatoires $(X_i)_{i\geqslant 1}$ et $(Y_i)_{i\geqslant 1}$ sont définies sur le même espace probabilisé $(\Omega, \mathcal{F}, P)$ et vérifient les hypothèses suivantes :

a) chacune de ces suites est i.i.d. ;

b) X_i et Y_i sont de carré intégrable, $\mathbf{E}X_i =: m_1$, $\mathbf{E}Y_i =: m_2$, $\operatorname{Var} X_i =: \sigma_1^2 > 0$, $\operatorname{Var} Y_i =: \sigma_2^2 > 0$;

c) les suites $(X_i)_{i\geqslant 1}$ et $(Y_i)_{i\geqslant 1}$ sont indépendantes l'une de l'autre, ce qui implique notamment que pour toute fonction mesurable $g : \mathbb{R}^2 \to \mathbb{R}$, $(g(X_i, Y_i))_{i\geqslant 1}$ est une suite de v.a. indépendantes et aussi que pour tout $i \geqslant 1$, X_i et Y_i sont indépendantes.

On pose pour tout $n \geqslant 1$,

$$S_n := \sum_{i=1}^{n} X_i, \quad T_n := \sum_{i=1}^{n} Y_i.$$

1) On suppose que $m_2 > m_1$. Expliquez pourquoi

$$W_n := \frac{1}{\sqrt{n}}\big(T_n - S_n\big) \xrightarrow[n\to+\infty]{\text{p.s.}} +\infty.$$

Indication. Pensez à la loi forte des grands nombres.

2) On suppose maintenant que $m_1 = m_2$. On pose $Z_1 := Y_1 - X_1$. Quelle est la variance de Z_1 ? Montrez que W_n converge en loi vers une v.a. gaussienne dont on précisera les paramètres.

Ex. 10.9 *Records et loi des grands nombres*

On note X une variable aléatoire positive dont la loi vérifie $P(X > t) = t^{-r}$ pour tout réel $t \geqslant 1$. Dans tout l'exercice, on suppose que la constante r est dans $]0, 1[$. On note $(X_k)_{k\geqslant 1}$ une suite de variables aléatoires positives définies sur le même espace probabilisé, indépendantes, de même loi que X et on pose :

$$M_n := \max_{1\leqslant k\leqslant n} X_k, \qquad S_n := \sum_{k=1}^{n} X_k.$$

1) Que vaut $P(X > 1/2)$? Que vaut $\mathbf{E}X$?

2) Calculez la fonction de répartition de M_n.

3) Montrez que $n^{-1/r}M_n$ converge en loi, quand n tend vers l'infini, vers une variable aléatoire positive Z dont vous donnerez la fonction de répartition.

4) Justifiez la convergence de série à termes positifs :

$$\forall \delta > 0, \quad \sum_{n=1}^{+\infty} \exp(-n^{\delta}) < +\infty.$$

5) Soit a une constante telle que $0 < a < 1/r$. Montrez que

$$\sum_{n=1}^{+\infty} P(M_n \leqslant n^a) < +\infty.$$

6) En déduire que presque-sûrement $M_n > n^a$ pour tout $n \geqslant N$, où N est un rang aléatoire.

7) En prenant $a > 1$, en déduire que

$$\frac{S_n}{n} \xrightarrow[n\to+\infty]{\text{p.s.}} +\infty.$$

Ceci contredit-il la loi forte des grands nombres ?

8) Montrez qu'en fait

$$\frac{S_n}{n^c} \xrightarrow[n\to+\infty]{\text{p.s.}} +\infty,$$

pour tout exposant c tel que $0 < c < 1/r$.

Solution p. 497

Ex. 10.10 *Convergences de moyennes géométriques*

Soit $(X_k)_{k\geqslant 1}$ une suite de variables aléatoires réelles définies sur le même espace probabilisé, à valeurs dans $]0, +\infty[$, indépendantes, de même loi telle que $\mathbf{E}|\ln X_1| < +\infty$. On note $b := \mathbf{E}\ln X_1$.

1) On note $T_n = X_1 \dots X_n$ le produit des n premières variables aléatoires de la suite $(X_k)_{k\geqslant 1}$. La moyenne géométrique de ces variables est alors $T_n^{1/n}$. Montrez que cette moyenne géométrique converge presque sûrement, quand n tend vers l'infini, vers une constante que vous exprimerez en fonction de b.

2) Calculez cette limite dans le cas des v.a. X_k de l'exercice 10.9.

3) On revient au cas général et on pose $Y_k := \mathrm{e}^{-b}X_k$ et $R_n := Y_1 \dots Y_n$. Vérifiez que $R_n^{1/n}$ converge presque-sûrement vers 1 quand n tend vers $+\infty$.

4) On suppose de plus que $\ln X_1$ est de carré intégrable et que $\operatorname{Var}\ln X_1$ est strictement positive. Montrez que $R_n^{1/\sqrt{n}}$ converge *en loi* vers une variable aléatoire W qui s'exprime comme fonction d'une v.a. de loi classique. La loi de W est appelée *log-normale*.

Solution p. 500

Ex. 10.11 On considère la variable aléatoire X (non p.s. constante) et une suite $(X_k)_{k\geqslant 1}$ de variables aléatoires indépendantes et de même loi que X. On suppose de plus que X est de carré intégrable et que

$$\frac{1}{\sqrt{2}}X_1 + \frac{1}{\sqrt{2}}X_2 \text{ a même loi que } X. \tag{10.76}$$

Le but de cet exercice est de montrer que ces propriétés caractérisent une loi classique.

1) Montrez que nécessairement, $\mathbf{E}X = 0$.

2) Expliquez pourquoi

$$T_n := 2^{-n/2}\sum_{k=1}^{2^n} X_k$$

a même loi que X.

3) On pose $\sigma^2 = \mathbf{E}X^2$, $\sigma > 0$. Montrez que $\sigma^{-1}T_n$ converge en loi et précisez la loi limite. Même question pour T_n.

4) Quelle est la loi de X ?

Ex. 10.12 *La méthode de Monte-Carlo pour le calcul d'intégrales*

Soit $h : [0,1] \to \mathbb{R}$ une fonction continue. On se propose de donner une valeur approchée de

$$I := \int_0^1 h(x)\,\mathrm{d}x,$$

par une méthode probabiliste appelée « méthode de Monte-Carlo ». Pour cela on utilise la simulation informatique d'une suite $(U_i)_{i\geqslant 1}$ de variables aléatoires indépendantes et de même loi uniforme sur $[0,1]$. On pose

$$S_n := \sum_{i=1}^{n} h(U_i).$$

1) Expliquer pourquoi $X_1 := h(U_1)$ est intégrable et exprimer son espérance à l'aide de l'intégrale I.

2) À l'aide d'un théorème du cours, montrer que

$$\frac{S_n}{n} \xrightarrow[n\to+\infty]{\text{p.s.}} I.$$

Ce résultat légitime pour n « grand » l'approximation de I par la valeur $S_n(\omega)/n$ calculée à partir de l'échantillon généré par l'ordinateur. On s'intéresse maintenant au contrôle de l'erreur d'approximation.

3) Expliquer pourquoi X_1 est de carré intégrable et exprimer sa variance σ^2 à l'aide d'intégrales de la fonction h. Est-il raisonnable dans ce contexte de supposer σ connue ?

4) En utilisant le théorème limite central, proposer un intervalle centré sur $\frac{S_n}{n}$ et contenant I avec une probabilité de 95% (cet intervalle étant de longueur minimale et en négligeant l'erreur due à l'approximation gaussienne).

5) Cet intervalle dépendant de la quantité inconnue σ, n'est pas utilisable en pratique. Expliquer comment résoudre ce problème en majorant σ si on connaît un majorant M de $\sup_{x\in[0,1]} |h(x)|$.

6) Proposer de même un intervalle de confiance avec « variance estimée » en le justifiant par un théorème du cours. Pour cela on complète en pratique le programme de calcul de S_n/n en lui faisant calculer en plus la quantité

$$W_n := \frac{1}{n}\sum_{i=1}^{n} h(U_i)^2.$$

Solution p. 502

Ex. 10.13 *Comportement asymptotique de variables de loi $\chi^2(n)$*

On rappelle que si X suit la loi gaussienne $\mathfrak{N}(m,\sigma)$, $\mathbf{E}X = m$ et $\operatorname{Var} X = \sigma^2$. Dans tout l'exercice, on considère une suite $(X_i)_{i\geqslant 1}$ de variables aléatoires indépendantes et de même loi $\mathfrak{N}(0,1)$. On pose pour tout $n \geqslant 1$,

$$Z_n := \sum_{i=1}^{n} X_i^2.$$

On sait qu'alors Z_n suit la loi $\chi^2(n)$.

1) Dans cette question, X est une v.a. de loi $\mathfrak{N}(0,1)$.

1.a) Expliquez pourquoi X a des moments de tout ordre.

1.b) Que valent les $\mathbf{E}(X^{2n+1})$ pour $n \in \mathbb{N}$?

1.c) On pose $c_n := \mathbf{E}(X^{2n})$ pour $n \in \mathbb{N}$. Montrez en intégrant par parties que

$$\forall n \in \mathbb{N}, \quad c_n = \frac{1}{2n+1} c_{n+1}.$$

1.d) En déduire une formule explicite pour c_n et calculez $\mathbf{E}(X^4)$.

2) En justifiant votre réponse, calculez $\mathbf{E}Z_n$ et $\operatorname{Var} Z_n$.

3) Quel est le comportement de $(n^{-1}Z_n)_{n\geqslant 1}$ quand n tend vers l'infini ?

4) Justifiez la convergence

$$\frac{Z_n - n}{\sqrt{2n}} \xrightarrow[n\to+\infty]{\text{loi}} Z, \tag{10.77}$$

où Z suit la loi $\mathfrak{N}(0,1)$.

5) On note Y une variable gaussienne de loi $\mathfrak{N}(0, 2^{-1/2})$. Prouvez la convergence

$$\sqrt{Z_n} - \sqrt{n} \xrightarrow[n\to+\infty]{\text{loi}} Y. \tag{10.78}$$

Indication. On peut remarquer que Z_n étant une v.a. positive, $Z_n = (\sqrt{Z_n})^2$, d'où

$$\sqrt{Z_n} - \sqrt{n} = \frac{Z_n - n}{\sqrt{Z_n} + \sqrt{n}}.$$

On peut alors exploiter les deux questions précédentes.

Ex. 10.14 *Vecteurs gaussiens*

Soient X_1, X_2, X_3, trois variables aléatoires réelles définies sur le même espace probabilisé, indépendantes et de même loi gaussienne $\mathfrak{N}(0,1)$.

1) Pour chacun des vecteurs aléatoires suivants, indiquez s'il est gaussien en justifiant votre réponse.

$$U = (X_1, X_2, X_3), \quad V = (X_1, X_2, X_2, X_2 + X_3), \quad W = (X_1, X_2, X_3, 7 - X_1).$$

2) Donnez la matrice de covariance de chacun de ces vecteurs, en évitant de détailler tous les calculs, mais en donnant les justifications pertinentes.

Ex. 10.15 *Vecteurs gaussiens*

Soient X_1 et X_2 deux variables aléatoires définies sur le même espace probabilisé $(\Omega, \mathcal{F}, P)$, indépendantes et de même loi gaussienne $\mathfrak{N}(0,1)$. Soient a et b des réels vérifiant $a^2+b^2 = 1$. On définit les variables aléatoires Y_1 et Y_2 par l'égalité matricielle :

$$\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \begin{pmatrix} a & -b \\ b & a \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}.$$

1) Expliquez *sans aucun calcul* pourquoi $V = (Y_1, Y_2)$ est un vecteur gaussien de $\mathbb{R}^2$.

2) Calculez le vecteur espérance et la matrice de covariance de (Y_1, Y_2). Comparez les lois des vecteurs (X_1, X_2) et (Y_1, Y_2).

3) Que peut-on dire du vecteur aléatoire image de (X_1, X_2) par une rotation de centre $(0,0)$ et d'angle 38° ?

Ex. 10.16 *Matrices de covariance et couples gaussiens*

En dimension 1, il est facile de vérifier que tout nombre réel positif peut être la variance d'une certaine variable aléatoire (sauriez vous expliquer pourquoi en trois lignes maximum ?). On se propose d'examiner une question analogue en dimension 2 et d'en déduire une classification des lois de vecteurs gaussiens de $\mathbb{R}^2$.

1) Montrez que si M est la matrice de covariance d'un vecteur aléatoire (X, Y) de $\mathbb{R}^2$, elle est nécessairement de la forme :

$$M = \begin{pmatrix} a & c \\ c & b \end{pmatrix}, \qquad a, b \in \mathbb{R}^+,\ c \in \mathbb{R}, \quad ab - c^2 \geqslant 0. \tag{10.79}$$

2) Réciproquement, soit M une matrice 2×2 de la forme (10.79), montrons qu'elle est la matrice de covariance d'un certain couple aléatoire (X, Y) de $\mathbb{R}^2$. Le cas où $ab = 0$ est trivial, puisqu'alors M est la matrice nulle donc la matrice de covariance de n'importe quel vecteur aléatoire *constant* (justifiez ces affirmations). On suppose désormais que ab est non nul. On pose alors

$$\sigma_1 = \sqrt{a},\ \sigma_2 = \sqrt{b}, \quad \rho = \frac{c}{\sqrt{ab}} \quad (\text{donc } \rho \in [-1,1]),$$

on se donne un couple (U, V) de variables aléatoires réelles *indépendantes*, de carré intégrable et de variance 1 et on définit l'application linéaire g par :

$$g : \mathbb{R}^2 \to \mathbb{R}^2, \quad (u,v) \longmapsto (x,y), \text{ avec } \begin{cases} x = \sigma_1 u, \\ y = \sigma_2 \rho u + \sigma_2 \sqrt{1-\rho^2}\, v. \end{cases} \tag{10.80}$$

Vérifiez que le vecteur aléatoire $(X, Y) = g(U, V)$ a pour matrice de covariance M.

3) Que peut-on dire de la loi de (X, Y) lorsque U et V ci-dessus sont gaussiennes $\mathfrak{N}(0,1)$?

4) Soit (X', Y'), un vecteur aléatoire gaussien d'espérance $(0,0)$ et de matrice de covariance K. Utilisez ce qui précède pour montrer que si le déterminant de K

est non nul, alors la loi de (X', Y') a une densité qui se déduit de celle de (U, V) par un changement de variable linéaire. Donner l'expression de cette densité à l'aide des variances σ_1^2 et σ_2^2 de X' et Y' et de leur coefficient de corrélation.

5) Montrez que si $\det K = 0$, soit le vecteur aléatoire (X', Y') est p.s. constant égal à $(0, 0)$, soit sa loi est portée par une droite D, c'est-à-dire $P((X', Y') \in D) = 1$, dont vous donnerez l'équation.

Solution p. 506

Ex. 10.17 *Une application du théorème de Berry-Esséen*

1) On considère une suite $((\Omega_n, \mathcal{F}_n, P_n))_{n\geqslant 1}$ d'espaces probabilisés et sur chaque Ω_n une suite finie $(X_{n,i})_{1\leqslant i\leqslant n}$ de variables aléatoires P_n-i.i.d. ayant un moment absolu d'ordre 3 fini ($\mathbf{E}_n|X_{n,1}|^3 < +\infty$). On pose $S_n := \sum_{i=1}^n X_{n,i}$ et on définit $S_n^* := (S_n - \mathbf{E}S_n)(\operatorname{Var} S_n)^{-1/2}$. En utilisant l'inégalité de Berry-Esséen, donnez une condition suffisante pour que S_n^* converge en loi vers vers $\mathfrak{N}(0, 1)$.

2) On suppose désormais que S_n est une variable aléatoire de loi binomiale de paramètres n et p_n. En utilisant ce qui précède, montrez que S_n^* converge en loi vers $\mathfrak{N}(0, 1)$ si p_n converge vers $p \in]0, 1[$ ou si p_n tend vers zéro mais pas trop vite (précisez comment). Que se passe-t-il si $p_n \sim cn^{-1}$?

Ex. 10.18 *Inversions d'une permutation aléatoire*

On note Ω_n l'ensemble des permutations de $[\![1, n]\!]$, c'est à dire des bijections $\omega : [\![1, n]\!] \to [\![1, n]\!]$. Soient $k = \omega(i) \in [\![1, n]\!]$. On dit que k présente une inversion avec $\ell = \omega(j)$ si $j > i$ et $\ell < k$. Ainsi le nombre d'inversions dans ω de $k = \omega(i)$ est le nombre d'éléments inférieurs à $\omega(i)$ et situés à sa droite dans la liste $\omega(1), \ldots, \omega(i), \ldots, \omega(n)$. Le nombre d'inversions de k est ainsi toujours compris entre 0 et $k - 1$, le nombre d'inversions de 1 est toujours nul.

Le but de cet exercice est d'établir une loi faible des grands nombres et un théorème limite central pour le nombre total d'inversions d'une permutation aléatoire. Notons pour cela $X_k(\omega)$ le nombre d'inversions de k dans ω, pour $k \in [\![1, n]\!]$ et $S_n(\omega) = \sum_{k=1}^n X_k(\omega)$.

1) Lorsque n n'est pas trop grand, on peut représenter ω en donnant la liste explicite des $\omega(i)$. Par commodité, nous écrirons chaque $k \in [\![1, n]\!]$ sur un jeton carré et chaque disposition sur une même ligne des n jetons représentera une permutation ω. Par exemple

$$\boxed{4}\boxed{3}\boxed{5}\boxed{1}\boxed{2}$$

représente la permutation ω définie par : $1 \mapsto 4$, $2 \mapsto 3$, $3 \mapsto 5$, $4 \mapsto 1$, $5 \mapsto 2$. L'élément $4 = \omega(1)$ a 3 inversions, puisqu'à sa droite on trouve 3, 1 et 2 qui lui sont inférieurs, donc $X_4(\omega) = 3$. De proche en proche, on voit ainsi que le vecteur d'inversions associé à ω est $(X_1(\omega), \ldots, X_5(\omega)) = (0, 0, 2, 3, 2)$. Trouvez la (ou les) permutation(s) ayant pour vecteur d'inversion $(0, 1, 0, 3, 2)$. Même question pour le vecteur d'inversions $(0, 0, 1, 2, 0)$.

2) On note Δ_n le produit cartésien

$$\Delta_n = \{0\} \times [\![0, 1]\!] \times [\![0, 2]\!] \times \cdots \times [\![0, n-1]\!].$$

Tout vecteur d'inversions d'une permutation $\omega \in \Omega_n$ est dans Δ_n (pourquoi ?). Montrez que pour tout $x \in \Delta_n$, il existe *un unique* $\omega \in \Omega_n$ tel que $(X_1(\omega), \dots, X_n(\omega)) = x$. Autrement dit, l'application $\omega \mapsto (X_1(\omega), \dots, X_n(\omega))$ est une bijection de Ω_n sur Δ_n.

3) On munit Ω_n de la tribu $\mathcal{F}_n = \mathcal{P}(\Omega_n)$ de tous ses sous-ensembles et de la probabilité P_n, loi uniforme sur Ω_n. On a donc $P_n(\{\omega\}) = \frac{1}{n!}$ pour tout $\omega \in \Omega_n$. Que vaut $P_n((X_1, \dots, X_n) = x)$ pour $x \in \Delta_n$? Quelle est la loi du vecteur aléatoire $(X_1, \dots, X_n)$?

4) Soit un espace probabilisé $(\Omega', \mathcal{F}', P')$ sur lequel on puisse définir des variables aléatoires discrètes $U_1, \dots, U_n$ indépendantes et telles que pour tout $k \in [\![1, n]\!]$, U_k suive la loi uniforme sur $[\![0, k-1]\!]$ (on ne vous demande pas de justifier l'existence de cet espace). Vérifiez que $(U_1, \dots, U_n)$ suit la loi uniforme sur Δ_n. En déduire que les X_k sont indépendantes et de même loi que les U_k.

5) Justifiez brièvement les formules sommatoires

$$\sum_{j=1}^{n} j = \frac{n(n+1)}{2} \tag{10.81}$$

et

$$\sum_{j=1}^{n} j^2 = \frac{n(n+1)(2n+1)}{6}. \tag{10.82}$$

6) On rappelle que $S_n = X_1 + \cdots + X_n$. Vérifiez que pour $n \geqslant 2$,

$$\mathbf{E}S_n = \frac{n(n-1)}{4} \sim \frac{n^2}{4}, \quad \operatorname{Var} S_n = \frac{n(n-1)(2n+5)}{72} \sim \frac{n^3}{36}.$$

7) Montrez que S_n vérifie la loi faible des grands nombres suivante :

$$\forall \varepsilon > 0, \quad P_n\left(\left|\frac{S_n}{n^2} - \frac{1}{4}\right| > \varepsilon)\right) \xrightarrow[n\to+\infty]{} 0.$$

8) On rappelle le théorème limite central de Lindeberg que l'on énonce ici sous une forme adaptée à notre problème.

Théorème (Lindeberg)**.** *Considérons le « tableau triangulaire » de variables aléatoires de ligne numéro n* ($n \in \mathbb{N}^*$) *:*

$$X_{n,1}, \dots, X_{n,i}, \dots, X_{n,n},$$

où ces n variables sont définies sur le même $(\Omega_n, \mathcal{F}_n, P_n)$*, indépendantes, de carré intégrable. On note*

$$S_n := \sum_{i=1}^{n} X_{n,i}, \qquad \sigma_{n,i}^2 := \operatorname{Var} X_{n,i}, \qquad s_n^2 := \operatorname{Var} S_n = \sum_{i=1}^{n} \sigma_{n,i}^2.$$

On suppose $s_n > 0$ *pour tout n. Si de plus le tableau vérifie la* condition de Lindeberg *:*

$$\forall \varepsilon > 0, \quad R_{n,\varepsilon} := \frac{1}{s_n^2} \sum_{i=1}^{n} \mathbf{E}\big((X_{n,i} - \mathbf{E}X_{n,i})^2 \mathbf{1}_{\{|X_{n,i} - \mathbf{E}X_{n,i}| > \varepsilon s_n\}}\big) \xrightarrow[n\to+\infty]{} 0, \tag{10.83}$$

alors $s_n^{-1}(S_n - \mathbf{E}S_n)$ *converge en loi vers une v.a.* Z *de loi* $\mathfrak{N}(0,1)$.

Dans notre problème, les espaces probabilisés $(\Omega_n, \mathcal{F}_n, P_n)_{n\geqslant 1}$ sont ceux définis à la question 3. La v.a. notée X_k, quand on travaillait à n fixé, est notée maintenant $X_{n,k}$ (elle dépend de n puisqu'elle s'applique à des permutations ω sur $[\![1,n]\!]$). En utilisant le résultat de la question 6, montrez qu'il existe un rang $n_0 = n_0(\varepsilon)$ à partir duquel toutes les indicatrices dans $R_{n,\varepsilon}$ sont nulles (sur Ω_n tout entier) et qu'ainsi la condition de Lindeberg (10.83) est trivialement vérifiée. Utilisez ceci pour prouver que :

$$\frac{6}{n^{3/2}}\left(S_n - \frac{n^2}{4}\right) \xrightarrow[n\to+\infty]{\text{loi}} Z,$$

où Z suit la loi $\mathfrak{N}(0,1)$.

9) Justifiez l'affirmation « pour n grand, environ la moitié des permutations sur $[\![1,n]\!]$ ont leur nombre total d'inversions compris entre les bornes $\dfrac{n^2}{4} \pm 0{,}113\, n^{3/2}$ ».

Solution p. 510

Ex. 10.19 *Application du th. de Lindeberg à un test d'égalité de deux espérances*

On dispose d'un échantillon $X_1, \ldots, X_l$ de la loi inconnue P_X (les X_i sont i.i.d. de même loi qu'une v.a. aléatoire générique X) et d'un échantillon $Y_1, \ldots, Y_m$ de la loi P_Y (les Y_i sont i.i.d. de même loi qu'une v.a. aléatoire générique Y). On suppose de plus que ces deux échantillons sont indépendants. On voudrait tester l'hypothèse

$$(\mathcal{H}_0): \quad \mathbf{E}X = \mathbf{E}Y \qquad \text{contre} \qquad (\mathcal{H}_1): \quad \mathbf{E}X \neq \mathbf{E}Y.$$

Cette situation se présente notamment en médecine quand on veut tester l'efficacité d'un nouveau médicament et où l'on observe les durées de guérisons X_i d'un premier échantillon de patients auxquels on a administré le nouveau médicament, tandis que les Y_i sont les durées de guérison des patients d'un autre échantillon ayant reçu un médicament ancien (ou un placebo !). A priori on ne connaît pas les variances σ_X^2 et σ_Y^2 des lois P_X et P_Y. On construit une statistique de test

$$T := \frac{\overline{Y} - \overline{X}}{\sqrt{\frac{1}{l}S_X^2 + \frac{1}{m}S_Y^2}}, \tag{10.84}$$

où $\overline{X}$, $\overline{Y}$ sont les moyennes et S_X^2, S_Y^2 les variances empiriques :

$$\overline{X} := \frac{1}{l}\sum_{i=1}^{l} X_i, \quad S_X^2 := \frac{1}{l}\sum_{i=1}^{l}(X_i - \overline{X})^2, \quad \text{etc.}$$

Pour l et m « grands », on considère que sous $(\mathcal{H}_0)$, T doit être approximativement gaussienne $\mathfrak{N}(0,1)$. Ceci amène à définir un test de niveau ε en décidant de rejeter $(\mathcal{H}_0)$ si l'on observe $|T| > t_\varepsilon$, où t_ε est défini par $P(|Z| > t_\varepsilon) = \varepsilon$, avec $Z \sim \mathfrak{N}(0,1)$.

La légitimation théorique de ce test est contenue dans le théorème suivant que l'on vous propose de démontrer.

Théorème 10.48. *On suppose que X et Y sont de carré intégrable (avec $\sigma_X, \sigma_Y > 0$) et que $l = l(n)$ et $m = m(n)$ tendent vers l'infini avec n. Alors, sous $(\mathcal{H}_0)$, la statistique de test $T = T_n$ définie par (10.84) converge en loi vers $\mathfrak{N}(0,1)$ lorsque n tend vers l'infini. En revanche, sous $(\mathcal{H}_1)$, $|T_n|$ tend p.s. vers l'infini.*

1) Sous $(\mathcal{H}_0)$ ou $(\mathcal{H}_1)$, donner les limites presque sûres du numérateur et du dénominateur de T_n. En déduire la limite p.s. de $|T_n|$ sous $(\mathcal{H}_1)$.

2) On se place désormais sous $(\mathcal{H}_0)$. Comme on a une forme indéterminée du type « 0/0 » pour la convergence p.s. de T_n, on étudie plutôt la convergence en loi. Considérons pour cela le tableau triangulaire de n^{e} ligne (rappelons que $l = l(n)$ et $m = m(n)$) :

$$\frac{-X_1'}{l}, \ldots, \frac{-X_l'}{l}, \frac{Y_1'}{m}, \ldots, \frac{Y_m'}{m}, \tag{10.85}$$

où l'on a posé $X_i' := X_i - \mathbf{E}X_i$ et $Y_i' = Y_i - \mathbf{E}Y_i$. La somme de la ligne est $S_n' = \overline{Y'} - \overline{X'} = \overline{Y} - \overline{X} + \mathbf{E}X - \mathbf{E}Y$, mais comme on est sous $(\mathcal{H}_0)$, $\mathbf{E}X - \mathbf{E}Y = 0$, d'où $S_n' = S_n$. Grâce à l'indépendance, sa variance est

$$s_n^2 = \operatorname{Var} S_n = \frac{1}{l^2} \sum_{i=1}^{l} \operatorname{Var} X_i + \frac{1}{m^2} \sum_{j=1}^{m} \operatorname{Var} Y_j = \frac{1}{l}\sigma_X^2 + \frac{1}{m}\sigma_Y^2.$$

Justifiez les égalités $S_X^2 = S_{X'}^2$, $S_Y^2 = S_{Y'}^2$, d'où $T_n' = T_n$, en notant T_n' la statistique obtenue en centrant toutes les observations.

3) Vérifiez la convergence

$$\frac{\frac{1}{l(n)}\sigma_X^2 + \frac{1}{m(n)}\sigma_Y^2}{\frac{1}{l(n)}S_X^2 + \frac{1}{m(n)}S_Y^2} \xrightarrow[n\to+\infty]{\text{p.s.}} 1.$$

En écrivant T_n sous la forme

$$T_n = \left(\frac{\frac{1}{l}\sigma_X^2 + \frac{1}{m}\sigma_Y^2}{\frac{1}{l}S_X^2 + \frac{1}{m}S_Y^2} \right)^{1/2} \times \frac{S_n'}{s_n}$$

on réduit la preuve du théorème à la vérification de la condition de Lindeberg pour le tableau triangulaire (10.85), pourquoi ?

4) Il s'agit donc de montrer, pour tout $\varepsilon > 0$, la convergence vers 0 de

$$R_n(\varepsilon) = \frac{1}{s_n^2} \left(\sum_{i=1}^{l} \mathbf{E}\left(\left(\frac{-X_i'}{l} \right)^2 \mathbf{1}_{\{|X_i'/l| > \varepsilon s_n\}} \right) + \sum_{j=1}^{m} \mathbf{E}\left(\left(\frac{Y_j'}{m} \right)^2 \mathbf{1}_{\{|Y_j'/m| > \varepsilon s_n\}} \right) \right).$$

Vérifiez que

$$R_n(\varepsilon) = \frac{1}{l s_n^2} \mathbf{E}\left(X_1'^2 \mathbf{1}_{\{X_1'^2 > \varepsilon^2 l^2 s_n^2\}} \right) + \frac{1}{m s_n^2} \mathbf{E}\left(Y_1'^2 \mathbf{1}_{\{Y_1'^2 > \varepsilon^2 m^2 s_n^2\}} \right).$$

puis, que

$$R_n(\varepsilon) \leqslant \frac{1}{\sigma_X^2}\mathbf{E}\left(X_1'^2\mathbf{1}_{\{X_1'^2>\varepsilon^2 l^2 s_n^2\}}\right) + \frac{1}{\sigma_Y^2}\mathbf{E}\left(Y_1'^2\mathbf{1}_{\{Y_1'^2>\varepsilon^2 m^2 s_n^2\}}\right).$$

5) Expliquez pourquoi $l(n)^2 s_n^2$ tend vers l'infini quand n tend vers l'infini et en déduire que

$$\mathbf{E}\left(X_1'^2\mathbf{1}_{\{X_1'^2>\varepsilon^2 l^2 s_n^2\}}\right) \xrightarrow[n\to+\infty]{} 0.$$

Conclure.

Chapitre 11

Indications et solutions

Ce chapitre regroupe des solutions rédigées de certains exercices et des indications pour une partie des exercices. Leur numérotation est celle des exercices correspondants.

11.1 Exercices du chapitre 1

Solution 1.6

1) Vérifions d'abord la Proposition 1.42 b) sur l'exemple proposé par l'énoncé ($n = 9$ et $r = 4$) en utilisant le codage des répartitions possibles par des chaînes de caractères. On représente d'abord chaque jeton par le caractère 'O', *avec un espace* entre deux caractères consécutifs :

O O O O O O O O O.

Pour représenter les 4 boîtes, il suffit d'insérer 3 caractères '|', chacun dans un espace libre. Par exemple, la chaîne

O O|O O O|O|O O O

code la répartition avec 2 jetons dans la première boîte, 3 dans la deuxième, 1 dans la troisième et 3 dans la quatrième, autrement dit la solution $(x_1, x_2, x_3, x_4) = (2, 3, 1, 3)$.

Avec ce codage, il est clair qu'il y a autant de répartitions des 9 jetons dans les 4 boîtes que de façons de placer les 3 caractères '|' parmi les 8 espaces disponibles. Il y a donc en tout $C_8^3 = 56$ répartitions possibles des 9 jetons dans les 4 boîtes. C'est bien la formule annoncée : C_{9-1}^{4-1}. La généralisation est immédiate : n jetons délimitent $n-1$ espaces de séparation entre jetons consécutifs et il faut exactement $r-1$ caractères '|' pour délimiter r boîtes, il y a donc bien C_{n-1}^{r-1} répartitions possibles. Notons qu'entre deux espaces de séparation, il y a toujours au moins un caractère 'O', donc le codage utilisé interdit la représentation d'une boîte vide.

2) On prouve la Proposition 1.42 a), en utilisant le codage suivant. Le caractère '|' représente à la fois le début et la fin d'une boîte (sauf s'il est le premier ou le dernier

de la chaîne). Il nous en faut donc 5 pour représenter 4 boîtes. On représente encore chaque jeton par un 'O', mais cette fois on ne laisse plus d'espace entre les caractères de la chaîne. L'exemple de la question précédente sera alors codé par la chaîne de 14 caractères sans espaces :

$$|\mathsf{OO}|\mathsf{OOO}|\mathsf{O}|\mathsf{OOO}|.$$

Ceci permet de représenter une boîte vide par deux caractères '|' consécutifs. Par exemple,

$$|\mathsf{OO}||\mathsf{OOOO}|\mathsf{OOO}|, \qquad |||\mathsf{OOOO}|\mathsf{OOOOO}|,$$

représentent respectivement les solutions $(2, 0, 4, 3)$ et $(0, 0, 4, 5)$.

Raisonnons à nouveau sur l'exemple $n = 9$ et $r = 4$. Le codage comporte en tout 9 caractères 'O' et 5 caractères '|'. Il faut obligatoirement un '|' en début de chaîne et un en fin. Il en reste donc 3 à répartir parmi $14 - 2 = 12$ positions disponibles. Une fois placés, il reste 9 places inoccupées qui recevront chacune un 'O'. Le codage est donc parfaitement déterminé dès que l'on connaît la place des trois '|' libres. Il y donc exactement $C_{12}^3 = 220$ répartitions possibles et ceci correspond bien à la formule annoncée.

Généralisation. Notons d'abord que contrairement au b), il n'est pas nécessaire de supposer $n \geqslant r$ puisqu'il peut y avoir des boîtes vides. On remarque qu'il faut $r + 1$ caractères '|' pour délimiter r boîtes : pour $1 \leqslant k \leqslant r$, le k^{e} '|' marque le début de la k^{e} boîte et le $(k + 1)^{\text{e}}$ en marque la fin (ainsi le $(r + 1)^{\text{e}}$ marque la fin de la r^{e} boîte). On a donc en tout une chaîne de $n + r + 1$ caractères commençant et finissant obligatoirement par un '|'. Le placement des $(r + 1) - 2$ caractères '|' libres parmi les $(n + r + 1) - 2$ positions disponibles peut se faire de $C_{(n+r+1)-2}^{(r+1)-2} = C_{n+r-1}^{r-1}$ façons différentes et il ne reste alors plus aucun choix pour les n places restantes qui recevront obligatoirement chacune un 'O'. En utilisant la relation $C_m^k = C_m^{m-k}$, avec $m = n + r - 1$ et $k = r - 1$, on voit que le nombre de répartitions trouvé s'écrit aussi C_{n+r-1}^n. On peut d'ailleurs trouver directement ce résultat en plaçant les n caractères 'O' parmi les $(n + r + 1) - 2$ positions disponibles une fois placés les deux '|' des extrémités. C'est alors pour les $r - 1$ '|' restants qu'il n'y a plus de choix : ils vont occuper les $(n + r + 1) - 2 - n = r - 1$ places restantes.

3) *Passage du a) au b) par changement de variable.*
Considérons pour $m, r > 1$ et $n \geqslant r$, les ensembles

$$\begin{aligned} E_{m,r} &:= \{x = (x_1, \ldots, x_r) \in \mathbb{N}^r \ ; \ x_1 + x_2 + \cdots + x_r = m\}; \\ F_{n,r} &:= \{y = (y_1, \ldots, y_r) \in \mathbb{N}_*^r \ ; \ y_1 + y_2 + \cdots + y_r = n\}. \end{aligned}$$

Définissons l'application $f : \mathbb{N}^r \to \mathbb{N}_*^r$, $x \mapsto y$ par

$$y = (y_1, \ldots, y_r) = f(x) = (1 + x_1, 1 + x_2, \ldots, 1 + x_r).$$

C'est une *injection* car si $x \neq x'$, il existe au moins un indice $1 \leqslant i \leqslant r$ pour lequel $x_i \neq x'_i$. En notant $y = f(x)$ et $y' = f(x')$, on a alors pour cet indice i, $y_i = 1 + x_i \neq 1 + x'_i = y'_i$ d'où $y \neq y'$. Cherchons l'ensemble image de $E_{m,r}$ par f.

Soit $x \in E_{m,r}$. Son image $y = f(x)$ a une somme des composantes qui vaut

$$y_1+y_2+\cdots+y_r = (1+x_1)+(1+x_2)+\cdots+(1+x_r) = \underbrace{1+\cdots+1}_{r \text{ termes}}+x_1+\cdots+x_r = r+m.$$

Donc $y = f(x)$ est un élément de $F_{m+r,r}$.

Réciproquement, soit $z = (z_1, \ldots, z_r)$ un élément quelconque de $F_{m+r,r}$. Cela signifie que chaque z_i $(1 \leqslant i \leqslant r)$ est un entier strictement positif et que $z_1 + \cdots + z_r = m + r$. Cette égalité peut encore s'écrire : $(z_1 - 1) + \cdots + (z_r - 1) = m$. En posant $x_i = z_i - 1$, on constate que chaque x_i est un entier positif ou nul et que $x_1 + \cdots + x_r = m$. Ainsi $x := (x_1, \ldots, x_r)$ est dans $E_{m,r}$. Il est clair que $f(x) = z$. Donc z a un antécédent x par f dans $E_{m,r}$.

Ainsi tout x dans $E_{m,r}$ a son image par f dans $F_{m+r,r}$ et tout z de $F_{m+r,r}$ est image d'un élément de $E_{m,r}$. L'ensemble image de $E_{m,r}$ par f est donc bien $F_{m+r,r}$. Il en résulte que la restriction de f à $E_{m,r}$ est une *surjection* de cet ensemble sur $F_{m+r,r}$. Comme elle est aussi injective, c'est une *bijection* entre les deux ensembles.

Grâce à cette bijection, $\operatorname{card} E_{m,r} = \operatorname{card} F_{m+r,r}$. On peut donc retrouver le b) à partir du a). En effet d'après a), le nombre de solutions à composantes dans $\mathbb{N}$ de l'équation $x_1 + \cdots + x_r = m$ est C^{r-1}_{m+r-1}. Ce nombre est $\operatorname{card} E_{m,r}$ donc aussi $\operatorname{card} F_{m+r,r}$, c'est-à-dire le nombre de solutions à composantes dans $\mathbb{N}_*$ de l'équation $x_1 + \cdots + x_r = n$ avec $n = m + r$ (donc $n \geqslant r$).

Solution 1.7 *En attendant le métro*

1) Une chaîne comportant m lettres L et n lettres O est construite au hasard. On se propose de calculer la probabilité qu'elle comporte exactement r séries de O. On utilise les résultats de l'exercice 1 comme suit. Pour $i \in [\![1, r]\!]$, notons x_i le nombre de lettres O de la i^{e} série de O (tous les x_i sont strictement positifs). Soit y_1 le nombre (éventuellement nul) de L avant la première série de O, y_{r+1} le nombre (éventuellement nul) de L après la dernière série de O. Pour $2 \leqslant i \leqslant r$, y_i désigne le nombre strictement positif de lettres L dans la série qui précède la i^{e} série de O.

$$\underbrace{\text{L}\ldots\text{L}}_{y_1}\ \underbrace{\text{O}\ldots\text{O}}_{x_1}\ \underbrace{\text{L}\ldots\text{L}}_{y_2}\ \underbrace{\text{O}\ldots\text{O}}_{x_2}\ \cdots\cdots\ \underbrace{\text{L}\ldots\text{L}}_{y_r}\ \underbrace{\text{O}\ldots\text{O}}_{x_r}\ \underbrace{\text{L}\ldots\text{L}}_{y_{r+1}}$$

Les r-uplets $(x_1, \ldots, x_r)$ possibles sont les solutions à composantes strictement positives de l'équation $x_1 + \cdots + x_r = n$. On se ramène à une condition du même type pour les y_i en faisant le changement de variable :

$$z_1 = y_1 + 1, \quad z_{r+1} = y_r + 1, \qquad z_i = y_i \ (2 \leqslant i \leqslant r).$$

Si les voyageurs se répartissent au hasard, on peut faire l'hypothèse d'équiprobabilité sur l'ensemble Ω des répartitions élémentaires possibles. Une répartition élémentaire peut se représenter par une chaîne de O et de L comme ci-dessus. Le cardinal de Ω est $C^n_{m+n} = C^m_{m+n}$. En effet, il y a C^n_{m+n} façons de placer les n lettres O (donc les voyageurs) parmi les $m + n$ positions possibles (les sièges).

Notons G_r l'ensemble des répartitions ayant exactement r séries de O. En raison de l'hypothèse d'équiprobabilité, la probabilité p_r d'avoir r séries de O est donc :

$$p_r = \frac{\operatorname{card} G_r}{\operatorname{card} \Omega} = \frac{\operatorname{card} G_r}{C_{m+n}^n}.$$

Le problème se réduit donc au calcul de $\operatorname{card} G_r$. L'étude ci-dessus montre qu'un élément de G_r se caractérise par un couple (x, z) où $x = (x_1, \dots, x_r) \in \mathbb{N}_*^r$ et $z = (z_1, \dots, z_{r+1}) \in \mathbb{N}_*^{r+1}$ vérifient $x_1 + \cdots + x_r = n$ et $z_1 + \cdots + z_{r+1} = m + 2$. Il y a donc exactement autant d'éléments de G_r que de couples de ce type. Autrement dit, avec les notations de l'exercice 1 3),

$$\operatorname{card} G_r = \operatorname{card}\big(F_{n,r} \times F_{m+2,r+1}\big) = \operatorname{card}\big(F_{n,r}\big) \cdot \operatorname{card}\big(F_{m+2,r+1}\big) = C_{n-1}^{r-1} C_{(m+2)-1}^{(r+1)-1}.$$

Finalement, on a bien

$$p_r = \frac{C_{m+1}^r C_{n-1}^{r-1}}{C_{m+n}^n}.$$

2) *Retour au sociologue.* Supposons que les voyageurs se répartissent au hasard. On peut alors appliquer la formule ci-dessus avec $n = 7$, $m = 9$ et $r = 7$. La probabilité d'observer 7 séries de O est donc

$$p_7 = \frac{C_{10}^7 C_6^6}{C_{16}^7} = \frac{C_{10}^7}{C_{16}^7} = \frac{10!}{3!\,7!}\,\frac{7!\,9!}{16!} = \frac{10!\,9!}{3!\,16!} \simeq 0{,}010\,5.$$

Cette valeur étant petite, le sociologue peut à bon droit mettre en doute l'hypothèse d'une répartition due au seul hasard.

Indication 1.11 Dans tout intervalle ouvert, il y a au moins un rationnel. Utiliser ceci pour *injecter* la famille d'intervalles dans $\mathbb{Q}$.

Indications 1.12 Raisonner par l'absurde en supposant qu'il n'existe aucune droite ne contenant pas au moins deux points à coordonnées rationnelles et exploiter la dénombrabilité de $\mathbb{Q}^2 \times \mathbb{Q}^2$. Alternativement, on peut aussi remarquer que si une droite passe par deux points à coordonnées rationnelles et d'abscisses différentes, sa pente est rationnelle.

11.2 Exercices du chapitre 2

Indication 2.2 Dénombrer l'ensemble $C_k := C \cap [\![10^k, 10^{k+1}[\![$ et en déduire un majorant de $\sum_{n \in C_k} u_n$.

Solution 2.3 *Développements d'un nombre décimal*

Par définition, l'écriture $x = 0{,}379\,999\,999\,999\dots$ signifie que

$$x = \frac{37}{100} + \sum_{k=3}^{+\infty} \frac{9}{10^k}.$$

La série $\sum_{k\geqslant 3}$ ci-dessus est une série géométrique de premier terme $9/1000$ et de raison $1/10$. Sa somme est donc

$$\sum_{k=3}^{+\infty} \frac{9}{10^k} = \frac{9}{1000} \sum_{j=0}^{+\infty} \Big(\frac{1}{10}\Big)^j = \frac{9}{1000}\, \frac{1}{1-\frac{1}{10}} = \frac{9}{1000}\, \frac{10}{9} = \frac{1}{100}.$$

Par conséquent

$$x = \frac{37}{100} + \frac{1}{100} = 0{,}38.$$

L'écriture $0{,}379\,999\,999\,999\ldots$ porte le nom de « développement décimal illimité impropre » du nombre 0,38. On appelle « développement décimal illimité propre » d'un réel x, un développement qui ne comporte pas la répétition indéfinie du chiffre 9 à partir d'un certain rang.

Pour généraliser le calcul fait ci-dessus, il est utile d'établir une fois pour toutes la formule suivante :

$$\forall m \in \mathbb{N}, \quad R_m := \sum_{k=m+1}^{+\infty} \frac{9}{10^k} = \frac{1}{10^m}. \tag{11.1}$$

En effet, la série $\sum_{k=m+1}^{+\infty}$ ci-dessus est géométrique de premier terme $9 \times 10^{-m-1}$ et de raison $1/10$. Sa somme se calcule en mettant en facteur le premier terme de façon à se ramener à la série géométrique standard de raison $1/10$, ce qui donne :

$$R_m = \frac{9}{10^{m+1}} \sum_{j=0}^{+\infty} \Big(\frac{1}{10}\Big)^j = \frac{9}{10^{m+1}}\, \frac{1}{1-\frac{1}{10}} = \frac{9}{10^{m+1}}\, \frac{10}{9} = \frac{1}{10^m}.$$

Cherchons une caractérisation des nombres réels *positifs* ayant un développement décimal illimité impropre. Soit x un tel réel. Dans ce cas le développement impropre de x peut s'écrire $x = c_{-n}\ldots c_0, c_1\ldots c_m 999999\ldots$, pour des entiers $n \geqslant 0$, $m \geqslant 0$ (si $m = 0$, la partie $c_1\ldots c_m$ disparaît) avec $c_m \neq 9$ si $m \geqslant 1$. Bien entendu dans cette écriture, les c_i sont des « chiffres décimaux » pris dans $[\![0, 9]\!]$. Alors

$$x = \frac{a}{10^m} + \sum_{k=m+1}^{+\infty} \frac{9}{10^k}, \quad \text{où } a \text{ est l'entier } \sum_{i=-n}^{m} c_i 10^{-m-i}.$$

Grâce à (11.1), on en déduit que

$$x = \frac{a}{10^m} + \frac{1}{10^m} = \frac{a+1}{10^m}.$$

Comme $a+1$ est entier, x est donc un nombre décimal *non nul* (car $a \geqslant 0$). Réciproquement, tout nombre décimal x strictement positif admet un développement décimal illimité impropre. En effet si $x = \frac{b}{10^m}$ avec $b \in \mathbb{N}^*$ et $m \in \mathbb{N}$, on peut toujours écrire

$$x = \frac{b-1}{10^m} + \frac{1}{10^m} = \sum_{i=-n}^{m} c_i 10^{-i} + \sum_{k=m+1}^{+\infty} \frac{9}{10^k},$$

où $\sum_{i=-n}^{m} c_i 10^{m-i}$ est l'écriture en base 10 de l'entier positif ou nul $b-1$. Ceci nous fournit le développement décimal illimité $x = c_{-n} \dots c_0, c_1 \dots c_m 9999 \dots$.

La (ou les) représentation(s) décimale(s) illimitée(s) d'un nombre réel strictement négatif x s'obtient par symétrie en écrivant la représentation correspondante de $-x$ précédée d'un signe moins. Compte-tenu de ce que nous venons de voir pour les réels positifs, on en déduit que $x < 0$ admet un développement décimal illimité impropre si et seulement s'il est décimal.

Reste à examiner le cas particulier de 0. Tout développement décimal illimité de 0 est de la forme

$$0 = \sum_{i=-n}^{+\infty} c_i 10^{-i} \quad \text{ou} \quad 0 = -\sum_{i=-n}^{+\infty} c_i 10^{-i},$$

où les c_i sont dans les deux cas des chiffres décimaux pris dans $[\![0, 9]\!]$. Dans le premier cas, la série est à termes positifs et sa somme ne peut être nulle que si tous ses termes sont nuls, donc tous les c_i valent 0. C'est la même chose dans le deuxième cas après multiplication des deux membres de l'égalité par -1.

En conclusion, l'ensemble des réels qui n'ont pas de développement décimal illimité impropre est $\mathbb{R}\backslash\mathbb{D}^*$. Dans cet ensemble, on trouve des rationnels non décimaux (exemple 1/3), des nombres algébriques non rationnels comme $\sqrt{2}$ et des nombres transcendants comme π.

Solution 2.5 *Une somme doublement géométrique*

1) $\mathcal{U}$ est une famille de réels *positifs*, sa somme S a donc toujours un sens comme élément de $\overline{\mathbb{R}}_+$, on peut d'ailleurs la définir directement comme le supremum des sommes S_K où K décrit l'ensemble de toutes les parties finies de $\mathbb{N}^2$. Dans ce cas la famille est sommable si S est finie. Nous pouvons appliquer à cette famille le théorème de sommation par paquets dans $\overline{\mathbb{R}}_+$, en découpant $\mathbb{N}^2$ en la réunion des « paquets » $\{i\} \times \mathbb{N}$, où i décrit $\mathbb{N}$. Nous obtenons ainsi

$$S := \sum_{(i,j)\in\mathbb{N}^2} u_{i,j} = \sum_{i\in\mathbb{N}} \left(\sum_{j\in\mathbb{N}} \frac{1}{18} \left(\frac{5}{6}\right)^i \left(\frac{2}{3}\right)^j \right) = \frac{1}{18} \sum_{i\in\mathbb{N}} \left(\frac{5}{6}\right)^i \left(\sum_{j\in\mathbb{N}} \left(\frac{2}{3}\right)^j \right).$$

Rappelons ici que $\sum_{k=0}^{+\infty} q^k$, série géométrique standard de raison $q \in \mathbb{R}$, converge dans $\mathbb{R}$ si et seulement si $|q| < 1$ et que sa somme est alors

$$\sum_{k=0}^{+\infty} q^k = \frac{1}{1-q} \qquad (|q| < 1).$$

En appliquant cette règle successivement à la série indexée par j avec $q = 2/3$ puis à celle indexée par i avec $q = 5/6$, nous obtenons [1] :

$$S = \frac{1}{18} \sum_{i\in\mathbb{N}} \left(\frac{5}{6}\right)^i \frac{1}{1-2/3} = \frac{3}{18} \sum_{i\in\mathbb{N}} \left(\frac{5}{6}\right)^i = \frac{1}{6} \frac{1}{1-5/6} = 1 < +\infty.$$

1. Lorsque q est positif, on peut écrire indifféremment $\sum_{k=0}^{+\infty} q^k$ ou $\sum_{k\in\mathbb{N}} q^k$, avant même de savoir s'il y a convergence absolue dans $\mathbb{R}$, en raison de la convergence commutative dans $\overline{\mathbb{R}}_+$ des séries à termes positifs.

La famille $\mathcal{U}$ est donc sommable et sa somme S vaut 1.

2) On note L le sous ensemble de $\mathbb{N}^2$ défini par $L := \big\{(i,j) \in \mathbb{N}^2 \ ; \ 0 \leqslant i \leqslant j\big\}$ et on s'intéresse à la famille $\mathcal{U}_L := \{u_{i,j} \ ; \ (i,j) \in L\}$.

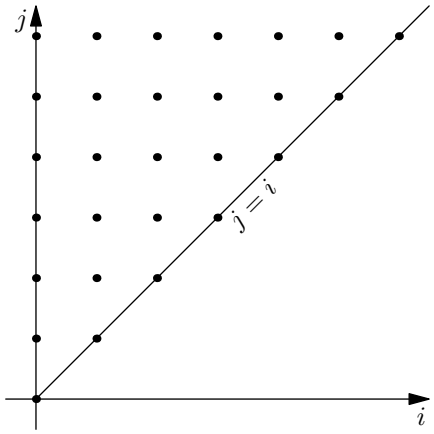


FIGURE 11.1 – Ensemble d'indices L

a) Une représentation graphique de L est proposée figure 11.1.

b) La famille $\mathcal{U}_L$ est sommable comme sous-famille de la famille sommable $\mathcal{U}$.

c) Pour calculer la somme S_L de cette famille, on peut utiliser le théorème de sommation par paquets en découpant L en paquets « horizontaux » $[\![0, j]\!] \times \{j\}$, avec j décrivant $\mathbb{N}$, cf. figure 11.2.

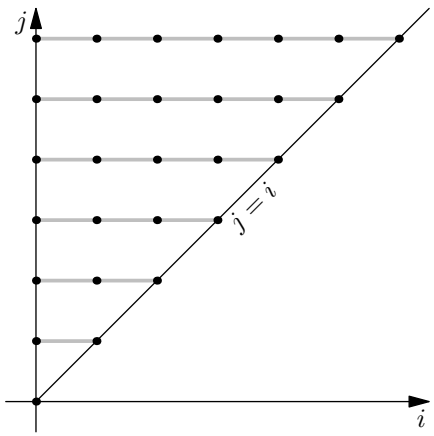


FIGURE 11.2 – Découpage de L en paquets « horizontaux »

$$S_L := \sum_{(i,j)\in L} u_{i,j} = \sum_{j\in\mathbb{N}} \left(\sum_{i=0}^{j} \frac{1}{18} \Big(\frac{5}{6}\Big)^i \Big(\frac{2}{3}\Big)^j \right) = \frac{1}{18} \sum_{j\in\mathbb{N}} \Big(\frac{2}{3}\Big)^j \left(\sum_{i=0}^{j} \Big(\frac{5}{6}\Big)^i \right) \qquad (11.2)$$

Rappelons ici que la somme des termes d'une suite géométrique finie de raison q est donnée par la formule

$$\sum_{k=0}^{n} q^k = \frac{1-q^{n+1}}{1-q} \qquad \forall q \neq 1.$$

En appliquant ceci avec $q = 5/6$ et en reportant dans (11.2), il vient

$$S_L = \frac{1}{18} \sum_{j\in\mathbb{N}} \Big(\frac{2}{3}\Big)^j \frac{1-(5/6)^{j+1}}{1-5/6} = \frac{1}{3} \sum_{j\in\mathbb{N}} \Big(\frac{2}{3}\Big)^j \left(1 - \Big(\frac{5}{6}\Big)^{j+1}\right).$$

On voit ainsi que S_L peut s'écrire comme la différence de deux séries géométriques de raison $q = \frac{2}{3} < 1$ et $q' = \frac{2}{3} \times \frac{5}{6} = \frac{5}{9} < 1$. La convergence de ces deux séries légitime cette écriture comme différence.

$$S_L = \frac{1}{3} \sum_{j=0}^{+\infty} \Big(\frac{2}{3}\Big)^j - \frac{5}{18} \sum_{j=0}^{+\infty} \Big(\frac{2}{3}\Big)^j \Big(\frac{5}{6}\Big)^j = \frac{1}{3} \frac{1}{1-2/3} - \frac{5}{18} \frac{1}{1-10/18} = 1 - \frac{5}{8} = \frac{3}{8}.$$

Une autre façon de calculer S_L est d'appliquer le théorème de sommation par paquets en découpant L en paquets « verticaux » $\{i\} \times [\![j, +\infty[\![$, i décrivant $\mathbb{N}$, cf. figure 11.3. On obtient ainsi

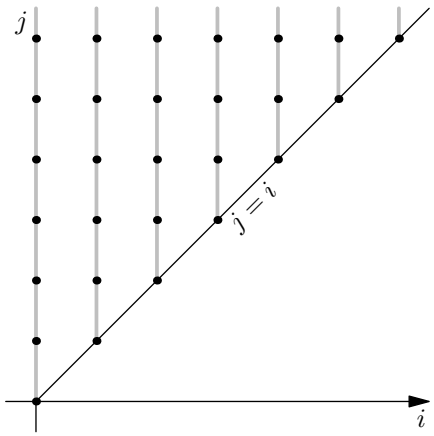


FIGURE 11.3 – Découpage de L en paquets « verticaux »

$$S_L = \sum_{i=0}^{+\infty} \left(\sum_{j=i}^{+\infty} \frac{1}{18} \Big(\frac{5}{6}\Big)^i \Big(\frac{2}{3}\Big)^j \right) = \frac{1}{18} \sum_{i=0}^{+\infty} \Big(\frac{5}{6}\Big)^i \left(\sum_{j=i}^{+\infty} \Big(\frac{2}{3}\Big)^j \right). \qquad (11.3)$$

C'est l'occasion de rappeler le calcul de la somme d'une série géométrique amputée de ses premiers termes. Il suffit de mettre en facteur le premier terme pour faire apparaître la série géométrique standard :

$$\sum_{k=m}^{+\infty} q^k = q^m \sum_{k=m}^{+\infty} q^{k-m} = q^m \sum_{\ell=0}^{+\infty} q^\ell = \frac{q^m}{1-q}, \quad (|q|<1).$$

En reportant ceci dans (11.3) avec $m = i$ et $q = 2/3$, il vient

$$S_L = \frac{1}{18} \sum_{i=0}^{+\infty} \Big(\frac{5}{6}\Big)^i \Big(\frac{2}{3}\Big)^i \frac{1}{1-2/3} = \frac{1}{6} \sum_{i=0}^{+\infty} \Big(\frac{5}{9}\Big)^i = \frac{1}{6} \frac{1}{1-5/9} = \frac{1}{6} \times \frac{9}{4} = \frac{3}{8}.$$

11.3 Exercices du chapitre 3

Indication 3.2 Calculez $s + t - |s - t|$ en fonction de $\min(s, t)$ pour s et t réels.

Indication 3.3 Raisonner par l'absurde en utilisant la continuité de f pour obtenir un sous-intervalle de $[a, b]$ sur lequel f a un minorant constant strictement positif.

Solution 3.4 En posant $g(t) = f(t)/M$, on est ramené à montrer la convergence vers 1 de $I_n := \big(\int_a^b g(t)^n \,\mathrm{d}t\big)^{1/n}$ pour toute fonction continue positive g sur $[a, b]$ telle que $\sup_{[a,b]} g = 1$. On peut déjà remarquer que puisque g positive est bornée par 1, $I_n \leqslant 1$ pour tout $n \geqslant 1$.

La fonction g étant continue sur le compact $[a, b]$ atteint son maximum (qui est donc aussi son supremum 1) en au moins un point t_0 de $[a, b]$. Fixons ε arbitraire dans $]0, 1[$. Par continuité de g au point t_0, il existe $\delta > 0$ tel que pour tout $t \in [t_0 - \delta, t_0 + \delta] \cap [a, b]$, $g(t) \geqslant g(t_0) - \varepsilon = 1 - \varepsilon$. Quitte à diminuer δ, on peut toujours s'arranger pour l'un au moins des intervalles $[t_0 - \delta, t_0]$ et $[t_0, t_0 + \delta]$ soit inclus dans $[a, b]$. En minorant alors I_n par l'intégrale de g sur un tel intervalle de longueur δ, puis en minorant g par $1 - \varepsilon$ sur cet intervalle, on obtient :

$$\forall n \in \mathbb{N}^*, \quad J_n := (\delta(1-\varepsilon)^{-n})^{1/n} \leqslant I_n \leqslant 1.$$

Comme $J_n = (1 - \varepsilon)\delta^{1/n}$ converge vers $1 - \varepsilon$, il existe un entier n_ε tel que pour tout $n \geqslant n_\varepsilon$, $J_n \geqslant 1 - 2\varepsilon$ et par conséquent :

$$\forall n \geqslant n_\varepsilon, \quad 1 - 2\varepsilon \leqslant I_n \leqslant 1.$$

Comme $\varepsilon \in]0, 1[$ était arbitriare, ceci établit la convergence I_n vers 1.

Solution 3.5 *Sommes de Riemann*

1) Pour alléger les écritures, posons pour $k \in [\![0, n]\!]$, $s_{n,k} = a + \frac{k}{n}(b - a)$. Il s'agit de vérifier que

$$T_n := \int_a^b f(t) \,\mathrm{d}t - \frac{b-a}{n} \sum_{k=1}^{n} f(t_{n,k})$$

tend vers 0 quand n tend vers l'infini. En remarquant que $s_{n,k} - s_{n,k-1} = (b-a)/n$, on peut réécrire T_n sous la forme :

$$T_n = \sum_{k=1}^{n} \int_{s_{n,k-1}}^{s_{n,k}} f(t)\,\mathrm{d}t - \sum_{k=1}^{n} \int_{s_{n,k-1}}^{s_{n,k}} f(t_{n,k})\,\mathrm{d}t = \sum_{k=1}^{n} \int_{s_{n,k-1}}^{s_{n,k}} \big(f(t) - f(t_{n,k})\big)\,\mathrm{d}t,$$

d'où la majoration :

$$|T_n| \leqslant \sum_{k=1}^{n} \int_{s_{n,k-1}}^{s_{n,k}} |f(t) - f(t_{n,k})|\ \mathrm{d}t.$$

Par continuité uniforme de f sur le compact $[a,b]$,

$$\forall \varepsilon > 0, \exists \delta_\varepsilon > 0, \quad |t-s| < \delta_\varepsilon \Rightarrow |f(t) - f(s)| < \frac{\varepsilon}{b-a}.$$

Pour chaque $\varepsilon > 0$, il existe un entier n_ε tel que pour tout $n \geqslant n_\varepsilon$, $(b-a)/n < \delta_\varepsilon$ et donc pour tout $t \in [s_{n,k-1}, s_{n,k}]$, $|f(t) - f(t_{n,k})| < \varepsilon/(b-a)$. Par conséquent,

$$\forall n \geqslant n_\varepsilon, \quad |T_n| \leqslant \sum_{k=1}^{n} \int_{s_{n,k-1}}^{s_{n,k}} \frac{\varepsilon}{b-a}\,\mathrm{d}t = \varepsilon.$$

Ceci étant vrai pour tout $\varepsilon > 0$, la convergence de T_n vers 0 est prouvée.

2) Si f est lipschitzienne sur $[a,b]$, il existe une constante $C > 0$ telle que pour tous $s, t \in [a,b]$, $|f(t) - f(s)| \leqslant C|t-s|$. En appliquant cette inégalité avec $t \in [s_{n,k-1}, s_{n,k}]$ et $s = t_{n,k}$, on obtient :

$$|f(t) - f(t_{n,k})| \leqslant C\,|t - t_{n,k}| \leqslant \frac{C(b-a)}{n}.$$

On en déduit que

$$|T_n| \leqslant \sum_{k=1}^{n} \int_{s_{n,k-1}}^{s_{n,k}} \frac{C(b-a)}{n}\,\mathrm{d}t = \frac{C(b-a)^2}{n},$$

donc T_n est un $O(1/n)$.

3) On remarque que

$$S_n := \sum_{k=1}^{n} \frac{1}{n+k} = \frac{1}{n} \sum_{k=1}^{n} \frac{1}{1+\frac{k}{n}} = \frac{1}{n} \sum_{k=1}^{n} f\Big(\frac{k}{n}\Big),$$

avec $f(x) = (1+x)^{-1}$. En appliquant le résultat de la question 1) avec $a = 0$, $b = 1$, $s_{n,k} = t_{n,k} = k/n$ pour tout $k \in [\![1, n]\!]$, on obtient

$$S_n \xrightarrow[n\to+\infty]{} \int_0^1 \frac{\mathrm{d}x}{1+x} = \ln 2.$$

Solution 3.7 *Riemann intégrabilité par prolongement*
Soit y_0 la valeur de $\tilde{f}$ au point b et notons

$$M := \max(|y_0|, \sup_{[a,b[} |f|).$$

Fixons $\varepsilon > 0$ arbitraire. On peut toujours trouver $c = c(\varepsilon) \in]a, b[$ tel que $b - c < \varepsilon/M$. Puisque f est Riemann intégrable sur $[a, c]$, il existe une subdivision Δ de $[a, c]$ telle qu'en notant $I_\varepsilon := \int_a^c f(x)\,\mathrm{d}x$:

$$I_\varepsilon - \varepsilon < S_\Delta(f, [a, c]) \leqslant S^\Delta(f, [a, c]) < I_\varepsilon + \varepsilon. \tag{11.4}$$

Adjoignons à Δ le point b pour en faire une subdivision Δ' de $[a, b]$. Par construction, on a clairement l'inégalité

$$|S_\Delta(f, [a, c]) - S_{\Delta'}(\tilde{f}, [a, b])| \leqslant (b - c)M < \varepsilon$$

ainsi que son analogue pour les sommes de Darboux supérieures. En combinant ceci avec l'encadrement (11.4), on en déduit :

$$I_\varepsilon - 2\varepsilon < S_{\Delta'}(\tilde{f}, [a, b]) \leqslant I_*(\tilde{f}, [a, b]) \leqslant I^*(\tilde{f}, [a, b]) \leqslant S^{\Delta'}(\tilde{f}, [a, b]) < I_\varepsilon + 2\varepsilon.$$

Ceci montre en particulier que

$$0 \leqslant I^*(f, [a, b]) - I_*(f, [a, b]) < 4\varepsilon$$

et comme $\varepsilon > 0$ était arbitraire, $I^*(\tilde{f}, [a, b]) = I_*(\tilde{f}, [a, b])$. Ainsi $\tilde{f}$ est bien Reimann intégrable sur $[a, b]$. De plus, grâce au théorème 3.32,

$$\int_a^b \tilde{f}(x)\,\mathrm{d}x = \lim_{\substack{c \to b \\ c < b}} \int_a^c \tilde{f}(x)\,\mathrm{d}x = \lim_{\substack{c \to b \\ c < b}} \int_a^c f(x)\,\mathrm{d}x.$$

11.4 Exercices du chapitre 4

Solution 4.5 *Propriétés de la frontière*
a) Vérification de : $\partial A = \overline{A} \backslash A^\circ$. Par définition,

$$\partial A = (A^\circ \cup A^{\mathrm{ext}})^c = (A^{\mathrm{ext}})^c \cap (A^\circ)^c = (A^{\mathrm{ext}})^c \backslash A^\circ.$$

Il suffit donc de vérifier que $(A^{\mathrm{ext}})^c = \overline{A}$. Or $x \in A^{\mathrm{ext}}$ si et seulement si x est intérieur à A^c, c.-à-d. s'il existe un $r > 0$ tel que la boule ouverte $B(x, r)$ soit incluse dans A^c. Donc $x \in (A^{\mathrm{ext}})^c$ si et seulement si *pour tout* $r > 0$, $B(x, r)$ *n'est pas* incluse dans A^c, autrement dit a une intersection non vide avec A. Or cette propriété caractérise l'appartenance de x à $\overline{A}$. Il y a donc équivalence entre l'appartenance à $(A^{\mathrm{ext}})^c$ et à $\overline{A}$, ce qui établit l'égalité des ces deux ensembles.

b) Vérification de : $\partial(A\cup B)\subset(\partial A\cup\partial B)$. Soit x un point quelconque de $\partial(A\cup B)$. On sait qu'alors tout voisinage de x contient au moins un point de $A\cup B$ et au moins un point de $(A\cup B)^c=A^c\cap B^c$. Par conséquent pour tout $n\in\mathbb{N}^*$, il existe $y_n\in A\cup B$ et $z_n\in A^c\cap B^c$ tels que $\|x-y_n\|<1/n$ et $\|x-z_n\|<1/n$. Définissons alors le sous-ensemble $\mathbb{N}_A$ de $\mathbb{N}^*$ par $\mathbb{N}_A:=\{n\in\mathbb{N}^*\ ;\ y_n\in A \text{ et } z_n\in A^c\}$ et définissons de même $\mathbb{N}_B$. D'après ce qui précède, $\mathbb{N}^*=\mathbb{N}_A\cup\mathbb{N}_B$, donc l'un au moins des ensembles $\mathbb{N}_A$ et $\mathbb{N}_B$ est infini. Si $\mathbb{N}_A$ est infini, x est adhérent à la fois à A et à A^c, donc $x\in\partial A$. Si $\mathbb{N}_B$ est infini $x\in\partial B$. Dans les deux cas, $x\in(\partial A\cup\partial B)$. Comme x était quelconque dans $\partial(A\cup B)$, l'inclusion recherchée est démontrée.

c) Vérification de : $\partial(A\cap B)\subset(\partial A\cup\partial B)$. Soit E une partie quelconque de $\mathbb{R}^d$. Comme E et E^c jouent un rôle symétrique dans la définition de ∂E, il est clair que $\partial(E^c)=\partial E$. En combinant cette remarque avec la propriété b), on obtient :

$$\partial(A\cap B)=\partial((A\cap B)^c)=\partial(A^c\cup B^c)\subset\partial(A^c)\cup\partial(B^c)=\partial A\cup\partial B,$$

ce qui établit la propriété c).

d) Vérification de : $\partial(A\backslash B)\subset(\partial A\cup\partial B)$. En utilisant c) et $\partial(E^c)=\partial E$, il vient :

$$\partial(A\backslash B)=\partial(A\cap B^c)\subset\partial A\cup\partial(B^c)=\partial A\cup\partial B.$$

Indication 4.7 *Admissibilité*

1) f est *uniformément* continue sur $[a,b]$.
2) La frontière de A peut se découper en 4 morceaux $\mathcal{R}$-négligeables.

Indication 4.9 *Fonction Bêta*

Justifiez l'écriture de $\Gamma(a)\Gamma(b)$ comme intégrale double généralisée :

$$\Gamma(a)\Gamma(b)=\int_{\mathbb{R}_+^2}u^{a-1}v^{b-1}\mathrm{e}^{-(u+v)}\,\mathrm{d}u\,\mathrm{d}v$$

et utilisez le changement de variables :

$$s=u+v,\quad t=\frac{u}{u+v}.$$

Solution 4.10 *Compatibilité des définitions de l'intégrale généralisée*

Soit f telle que $I=\int_{-\infty}^{+\infty}f(x)\,\mathrm{d}x$ converge *absolument* au sens des définitions 4.8 et 4.25. Il y a donc un ensemble fini de « trous » $T=\{t_1,\dots,t_k\}$ (éventuellement vide) tel qu'avec $t_0:=-\infty$, $t_{k+1}:=+\infty$, f soit Riemann intégrable sur tout $[a,b]\subset\mathbb{R}\backslash T$. De plus, il y a une suite de réels $c_0,c_1,\dots,c_k$, telle que pour tout $i\in[\![0,k]\!]$, $t_i<c_i<t_{i+1}$ et que chacune des limites suivantes existe et soit finie :

$$\forall i\in[\![0,k]\!],\quad \lim_{\substack{x_i\to t_{i+1},\\ x_i<t_{i+1}}}\int_{c_i}^{x_i}|f(t)|\,\mathrm{d}t,\quad \lim_{\substack{x'_i\to t_i,\\ x'_i>t_i}}\int_{x'_i}^{c_i}|f(t)|\,\mathrm{d}t.$$

Alors en remplaçant $|f|$ par f ci-dessus on obtient aussi des limites finies et $\int_{-\infty}^{+\infty} f(x)\,\mathrm{d}x$ est la somme de toutes ces limites.

Examinons d'abord le cas où f est bornée[2] sur $\mathbb{R}$. On prolonge f en chaque trou t_i par $f(t_i) := 0$. Définissons pour tout $n \in \mathbb{N}^*$ les compacts :

$$K_n := \bigcup_{i=0}^{k+1} [a_{n,i}, b_{n,i}], \quad \tilde{K}_n := K_n \cup T,$$

où pour tout i, $[a_{n,i}, b_{n,i}] \subset]t_i, t_{i+1}[$ et $a_{n,i} \downarrow t_i$, $b_{n,i} \uparrow t_{i+1}$ quand n tend vers l'infini. Clairement f est Riemann intégrable sur chacun de ces compacts admissibles. Les suites $(K_n)_{n\geqslant 1}$ et $(\tilde{K}_n)_{n\geqslant 1}$ épuisent respectivement $\mathbb{R}\backslash T$ et $\mathbb{R}$. De plus il découle immédiatement de la définition des intégrales généralisées $\int_{-\infty}^{+\infty} f(x)\,\mathrm{d}x$ et $\int_{-\infty}^{+\infty} |f(x)|\,\mathrm{d}x$ rappelée ci-dessus que

$$\lim_{n\to+\infty} \int_{K_n} |f(x)|\,\mathrm{d}x = \lim_{n\to+\infty} \int_{\tilde{K}_n} |f(x)|\,\mathrm{d}x = \int_{-\infty}^{+\infty} |f(x)|\,\mathrm{d}x$$

et que ces convergences restent valides en remplaçant partout $|f|$ par f. Ceci montre que les intégrales généralisées $\int_{\mathbb{R}\backslash T} f(x)\,\mathrm{d}x$ et $\int_{\mathbb{R}} f(x)\,\mathrm{d}x$ existent au sens de la définition 4.88 appliquée avec $d = 1$ et que ces deux intégrales sont égales à l'intégrale généralisée $\int_{-\infty}^{+\infty} f(x)\,\mathrm{d}x$ au sens de la définition 4.8.

Passons au cas où f n'est pas bornée sur $\mathbb{R}$. Il nous faut utiliser la troncature f_N de f définie par (4.62) p. 148. Remarquons aussi que si f est Riemann intégrable sur un intervalle fermé borné $[a, b]$, f_N l'est aussi car $f_N = \max(\min(f, N), -N)$, et le max ou le min de deux fonctions Riemann intégrables sur $[a, b]$ héritent de la même intégrabilité (cf. exercice 3.2). D'après le cas borné, $\int_{\mathbb{R}\backslash T} |f_N(x)|\,\mathrm{d}x$ et $\int_{\mathbb{R}} |f_N(x)|\,\mathrm{d}x$ comprises au sens de la définition 4.88 coïncident avec l'intégrale généralisée $\int_{-\infty}^{+\infty} |f_N(x)|\,\mathrm{d}x$ au sens de la définition 4.8. Il nous reste donc seulement à vérifier que

$$\lim_{N\to+\infty} \int_{-\infty}^{+\infty} |f_N(x)|\,\mathrm{d}x = \int_{-\infty}^{+\infty} |f(x)|\,\mathrm{d}x. \tag{11.5}$$

Fixons $\varepsilon > 0$ arbitraire. Grâce à la convergence de $\int_{-\infty}^{+\infty} |f(x)|\,\mathrm{d}x$, on peut trouver un $\delta < \min_{1\leqslant i<k}(t_{i+1} - t_i)$ et des réels $\alpha < t_1 - \delta$, $\beta > t_k + \delta$ tels que

$$\int_{-\infty}^{\alpha} |f(x)|\,\mathrm{d}x + \sum_{i=1}^{k} \int_{t_i-\delta}^{t_i+\delta} |f(x)|\,\mathrm{d}x + \int_{\beta}^{+\infty} |f(x)|\,\mathrm{d}x < \varepsilon.$$

Comme $|f_N| \leqslant |f|$, cette inégalité reste valable avec f_N à la place de f, avec les mêmes α, β, δ qui ne dépendent pas de N.

2. Rappelons au passage que si $-\infty < t_i < t_{i+1} < +\infty$ et f est bornée sur $]t_i, t_{i+1}[$, alors les intégrales généralisées $\int_{t_i}^{c_i}$ et $\int_{c_i}^{t_{i+1}}$ sont de fausses intégrales généralisées puisque tout prolongement fini de f aux points t_i et t_{i+1} est Riemann intégrable sur $[t_i, t_{i+1}]$, cf. l'exercice 3.7.

Par Riemann intégrabilité locale, $|f|$ est bornée par, disons $M = M(\varepsilon)$, sur l'ensemble

$$K = [\alpha, t_1 - \delta] \cup \left(\bigcup_{i=1}^{k-1} [t_i + \delta, t_{i+1} - \delta] \right) \cup [t_k + \delta, \beta].$$

Alors pour tout $N \geqslant M$, f_N coïncide avec f sur K. En recollant les morceaux, on obtient :

$$0 \leqslant \int_{-\infty}^{+\infty} |f(x)| \, \mathrm{d}x - \int_{-\infty}^{+\infty} |f_N(x)| \, \mathrm{d}x < \varepsilon.$$

On en déduit la convergence (11.5) par arbitrarité de ε. Une adaptation immédiate de ce qui précède nous donne la même convergence avec f au lieu de $|f|$.

Solution 4.11 *Espérance et fonction de survie*

1) En notant T le « triangle » $\{(x,t) \in [a,b]^2 \, ; \, t \leqslant x\}$ et $\tilde{f}$ la fonction $\mathbb{R}_+^2 \to \mathbb{R}_+$, $(x,t) \mapsto f(x)$, on voit que sur $[a,b]^2$, $h = \tilde{f}\mathbf{1}_T$. Comme T est un compact admissible, $\mathbf{1}_T$ est Riemann intégrable sur $[a,b]^2$ et il nous suffit donc de vérifier que $\tilde{f}$ l'est aussi. Or il est immédiat de vérifier que pour toute subdivision $\Delta = \Delta_1 \times \Delta_2$ de $[a,b]^2$,

$$S_\Delta(\tilde{f}) = (b-a) S_{\Delta_1}(f), \quad S^\Delta(\tilde{f}) = (b-a) S^{\Delta_1}(f).$$

Par conséquent, la Riemann intégrabilité de f sur $[a,b]$ entraîne celle de $\tilde{f}$ sur $[a,b]^2$. Ceci établit l'existence de

$$I_{a,b} := \int_{[a,b]^2} h(x,t) \, \mathrm{d}x \, \mathrm{d}t.$$

Pour les calculs qui suivent, il est utile de remarquer que

$$\forall (x,t) \in [a,b]^2, \quad h(x,t) = f(x)\mathbf{1}_T(x,t) = f(x)\mathbf{1}_{[a,x]}(t) = f(x)\mathbf{1}_{[t,b]}(x).$$

Pour calculer $I_{a,b}$, appliquons une première fois le théorème « de Fubini » (th. 4.80) en intégrant d'abord par rapport à t. On remarque au préalable que pour tout $x \in [a,b]$, la fonction $h(x, \, .\,) : t \mapsto f(x)\mathbf{1}_{[a,x]}(t)$ est Riemann intégrable sur $[a,b]$ puisque constante sur $[a,x]$ et sur $]x,b]$. Donc $\{x \in [a,b] \, ; \, h(x, \, .\,) \notin \mathcal{R}([a,b])\} = \varnothing$ et les deux hypothèses du th. 4.80 sont satisfaites. On obtient ainsi :

$$\begin{aligned} I_{a,b} = \int_a^b \left\{ \int_a^b f(x)\mathbf{1}_{[a,x]}(t) \, \mathrm{d}t \right\} \mathrm{d}x &= \int_a^b f(x) \left\{ \int_a^b \mathbf{1}_{[a,x]}(t) \, \mathrm{d}t \right\} \mathrm{d}x \\ &= \int_a^b f(x) \left\{ \int_a^x \mathrm{d}t \right\} \mathrm{d}x \\ &= \int_a^b f(x)(x-a) \, \mathrm{d}x. \end{aligned}$$

Les fonctions f et $x \mapsto x$ étant intégrables sur $[a,b]$, leur produit l'est aussi. Par linéarité on en déduit que

$$I_{a,b} = \int_a^b x f(x) \, \mathrm{d}x - a \int_a^b f(x) \, \mathrm{d}x = \int_a^b x f(x) \, \mathrm{d}x - a\big(G(a) - G(b)\big).$$

Appliquons une deuxième fois le théorème 4.80 en intégrant d'abord par rapport à x. Pour tout $t \in [a,b]$, la fonction $h(\,.\,,t) : x \mapsto f(x)\mathbf{1}_{[t,b]}(x)$ est égale soit à la fonction nulle, soit à f. Dans les deux cas, elle est Riemann intégrable sur $[a,b]$ donc $\{t \in [a,b]\ ;\ h(\,.\,,t) \notin \mathcal{R}([a,b])\} = \varnothing$ et les deux hypothèses du théorème 4.80 sont satisfaites. On obtient ainsi :

$$\begin{aligned} I_{a,b} = \int_a^b \left\{ \int_a^b f(x)\mathbf{1}_{[t,b]}(x)\,\mathrm{d}x \right\} \mathrm{d}t &= \int_a^b \left\{ \int_t^b f(x)\,\mathrm{d}x \right\} \mathrm{d}t \\ &= \int_a^b \big(G(t) - G(b)\big)\,\mathrm{d}t \\ &= \int_a^b G(t)\,\mathrm{d}t - (b-a)G(b). \end{aligned}$$

En égalant les deux valeurs trouvées pour $I_{a,b}$, on obtient :

$$\int_a^b xf(x)\,\mathrm{d}x = \int_a^b G(t)\,\mathrm{d}t + aG(a) - bG(b). \tag{11.6}$$

Dans le cas où f est continue, G est dérivable et $G' = -f$, par une adaptation immédiate du théorème 3.32. La formule (11.6) peut alors s'obtenir par une simple intégration par parties.

2) La formule (11.6) s'étend au cas où $\int_a^b f(x)\,\mathrm{d}x$ est généralisée en a ou en b, mais Riemann intégrable sur tout $[a',b'] \subset]a,b[$. En effet, prenons deux suites $(a_n)_{n\geqslant 1}$ et $(b_n)_{n\geqslant 1}$ dans $]a,b[$ telles que $a < a_1 < b_1 < b$ et $(a_n)_{n\geqslant 1}$ converge en décroissant vers a, $(b_n)_{n\geqslant 1}$ converge en croissant vers b. La formule (11.6) est vérifiée sur chaque $[a_n,b_n] \subset]a,b[$:

$$\forall n \geqslant 1, \quad \int_{a_n}^{b_n} xf(x)\,\mathrm{d}x = \int_{a_n}^{b_n} G(t)\,\mathrm{d}t + a_nG(a_n) - b_nG(b_n).$$

Par continuité de G, quand n tend vers l'infini, $\int_{a_n}^{b_n} G(t)\,\mathrm{d}t$ converge vers l'intégrale de Riemann ordinaire $\int_a^b G(t)\,\mathrm{d}t$ et $a_nG(a_n) - b_nG(b_n)$ tend vers $aG(a) - bG(b)$. On en déduit que l'intégrale éventuellement généralisée $\int_a^b xf(x)\,\mathrm{d}x$ converge et vaut $\int_a^b G(t)\,\mathrm{d}t + aG(a) - bG(b)$.

3) Pour montrer que $nG(n)$ tend vers zéro sous l'une des deux hypothèses proposées par l'énoncé, il suffit d'utiliser les inégalités :

$$nG(n) = \int_n^{+\infty} nf(x)\,\mathrm{d}x \leqslant \int_n^{+\infty} xf(x)\,\mathrm{d}x,$$

$$\frac{n}{2}G(n) \leqslant \int_{n/2}^{n} G(t)\,\mathrm{d}t \leqslant \int_{n/2}^{+\infty} G(t)\,\mathrm{d}t,$$

en rappelant la décroissance de G.

4) Nous allons exploiter ce qui précède pour prouver l'égalité dans $\overline{\mathbb{R}}_+$:

$$\int_0^{+\infty} xf(x)\,\mathrm{d}x = \int_0^{+\infty} G(t)\,\mathrm{d}t. \tag{11.7}$$

Par hypothèse, l'intégrale généralisée de f entre 0 et $+\infty$ est convergente au sens de la définition 4.8. Donc f est définie sur $\mathbb{R}^+$, sauf peut-être aux points $x_0 = 0 < x_1 < \cdots < x_k < x_{k+1} := +\infty$ (en prenant $k = 0$ si f est définie sur $]0, +\infty[$ ou sur $[0, +\infty[$). De plus f est Riemann intégrable sur chaque $[a, b] \subset \mathbb{R}_+ \backslash \{x_1, \dots x_k\}$. Ceci vaut aussi pour la fonction $x \mapsto xf(x)$. D'après la question 2, on a

$$\forall j \in [\![1, k]\!], \quad \int_{x_{j-1}}^{x_j} xf(x)\,\mathrm{d}x = \int_{x_{j-1}}^{x_j} G(t)\,\mathrm{d}t + x_{j-1}G(x_{j-1}) - x_jG(x_j)$$

et pour tout entier $n > x_k$, $\int_{x_k}^n xf(x)\,\mathrm{d}x = \int_{x_k}^n G(t)\,\mathrm{d}t + x_kG(x_k) - nG(n)$. En recollant les morceaux, il vient :

$$\forall n > k, \quad \int_0^n xf(x)\,\mathrm{d}x = \int_0^n G(t)\,\mathrm{d}t - nG(n). \tag{11.8}$$

Pour l'instant, nous n'avons utilisé que la convergence de l'intégrale généralisée $\int_0^{+\infty} f(x)\,\mathrm{d}x$ et toutes les intégrales généralisées $\int_{x_{j-1}}^{x_j} xf(x)\,\mathrm{d}x$ sont convergentes d'après la question 2. On peut d'ailleurs le voir aussi directement puisque x reste borné sur chaque $[x_{j-1}, x_j]$ pour $j \leqslant k$. Supposons maintenant que l'une des deux intégrales généralisées $\int_0^{+\infty} xf(x)\,\mathrm{d}x$ ou $\int_0^{+\infty} G(t)\,\mathrm{d}t$ converge (dans $\mathbb{R}_+$). Alors $nG(n)$ tend vers 0 et en faisant tendre n vers l'infini dans (11.8), on voit que l'autre intégrale généralisée converge aussi dans $\mathbb{R}_+$ et que l'égalité (11.7) a lieu dans $\mathbb{R}_+$.

Supposons maintenant que $\int_0^{+\infty} xf(x)\,\mathrm{d}x = +\infty$. D'après (11.8) et par positivité de G, $\int_0^n G(t)\,\mathrm{d}t \geqslant \int_0^n xf(x)\,\mathrm{d}x$, d'où en faisant tendre n vers l'infini, $\int_0^{+\infty} G(t)\,\mathrm{d}t = +\infty$.

Alternativement, supposons que $\int_0^{+\infty} G(t)\,\mathrm{d}t = +\infty$ et montrons que nécessairement $\int_0^{+\infty} xf(x)\,\mathrm{d}x = +\infty$. Considérons la suite $(u_n)_{n>k}$ de terme général $u_n = \int_0^n G(t)\,\mathrm{d}t - nG(n)$. D'après (11.8) cette suite de réels positifs est croissante, donc convergente vers une limite L dans $\overline{\mathbb{R}}_+$. Si $L = +\infty$, l'affaire est réglée puisqu'alors $\int_0^n xf(x)\,\mathrm{d}x = u_n$ tend vers l'infini et donc $\int_0^{+\infty} xf(x)\,\mathrm{d}x = +\infty$. Si L est fini, alors on peut écrire $u_n = L + \varepsilon_n$ avec ε_n tendant vers 0, d'où $nG(n) = \int_0^n G(t)\,\mathrm{d}t - L - \varepsilon_n$, ce qui montre que $nG(n)$ tend vers l'infini avec n. Mais comme $nG(n) \leqslant \int_n^{+\infty} xf(x)\,\mathrm{d}x$, ceci entraîne que $\int_0^{+\infty} xf(x)\,\mathrm{d}x = +\infty$.

11.5 Exercices du chapitre 5

Indication 5.1 Passez à l'évènement complémentaire. Et si cela ne vous inspire pas davantage, allez voir le lemme 9.2, p. 335.

Indication 5.4 On peut montrer qu'il existe un unique $n \in \mathbb{N}$ tel que $P\big(]n, n+1]\big) = 1$, puis une suite décroissante de sous-intervalles emboîtés de $]n, n+1]$ de probabilité un, dont la longueur tend vers 0.

Solution 5.5 *La piscine (permutations aléatoires)*

On se propose de calculer les probabilités des évènements :

$$E_{N,k} = \{\text{exactement } k \text{ nageurs retrouvent leurs affaires}\}.$$

On choisit comme espace probabilisé Ω_N ensemble des toutes les permutations (bijections) sur $[\![1, N]\!]$ muni de l'équiprobabilité P_N. On notera :

$$B_i = \{\text{le } i^{\text{e}} \text{ nageur retrouve ses affaires}\}.$$

1) Pour $j \leqslant N$ et $1 \leqslant i_1 < i_2 < \cdots < i_j \leqslant N$, calculons $P_N(B_{i_1} \cap \cdots \cap B_{i_j})$. Puisque P_N est *l'équiprobabilité* sur Ω_N,

$$P_N(B_{i_1} \cap \cdots \cap B_{i_j}) = \frac{\operatorname{card}(B_{i_1} \cap \cdots \cap B_{i_j})}{\operatorname{card} \Omega_N} = \frac{\operatorname{card}(B_{i_1} \cap \cdots \cap B_{i_j})}{N!},$$

en rappelant que le nombre de permutations d'un ensemble à N éléments est $N!$. Le problème se réduit donc au calcul de $\operatorname{card}(B_{i_1} \cap \cdots \cap B_{i_j})$.

Essayons déjà de nous faire une idée sur un exemple numérique simple, disons $N = 6$, $j = 3$, $i_1 = 2$, $i_2 = 4$, $i_3 = 5$. Alors $B_2 \cap B_4 \cap B_5$ est l'ensemble des permutations de $[\![1, 6]\!]$ laissant invariants 2, 4 et 5. En voici la liste complète :

1	2	3	4	5	6
1	2	6	4	5	3
3	2	6	4	5	1
3	2	1	4	5	6
6	2	3	4	5	1
6	2	1	4	5	3

Notons que parmi ces permutations, certaines ont plus de 3 points fixes. Pour dénombrer $B_2 \cap B_4 \cap B_5$, il suffit de remarquer qu'une permutation quelconque laissant 2, 4 et 5 invariants se construit sur le schéma suivant :

	2		4	5		

les case vides devant être complétées en « piochant » dans $\{1, 3, 6\}$. Il y a clairement autant de façons de compléter ces cases qu'il y a de permutations de l'ensemble à 3 éléments $\{1, 3, 6\}$, c'est-à-dire $3! = 6$.

Généralisons. Une permutation ω évènement élémentaire de $B_{i_1} \cap \cdots \cap B_{i_j}$ se construit en plaçant d'abord les j points fixes $i_1, \ldots, i_j$ ($\omega(i_k) = i_k$ pour $k \in [\![1, j]\!]$). Il reste alors $N - j$ cases à compléter en piochant dans l'ensemble $[\![1, N]\!] \backslash \{i_1, \ldots, i_j\}$. Il y a donc autant d'éléments dans $B_{i_1} \cap \cdots \cap B_{i_j}$ que de permutations de $[\![1, N]\!] \backslash \{i_1, \ldots, i_j\}$. Par conséquent,

$$\operatorname{card}(B_{i_1} \cap \cdots \cap B_{i_j}) = (N - j)! \quad \text{et} \quad P_N(B_{i_1} \cap \cdots \cap B_{i_j}) = \frac{(N - j)!}{N!}.$$

Notons que dans le cas particulier $j = N$, on a nécessairement $i_k = k$ pour $1 \leqslant k \leqslant N$ et on voit directement que $B_1 \cap \cdots \cap B_N$ ne contient qu'un seul élément la permutation identité. Donc $P_N(B_1 \cap \cdots \cap B_N) = 1/N!$, ce qui coïncide avec le cas $j = N$ dans la formule ci-dessus puisque $0! = 1$.

Pour $j = 1$, la formule ci-dessus nous donne $P_N(B_i) = \frac{1}{N}$. Pour $j = 2$, on voit que si $i_1 \neq i_2$,

$$P_N(B_{i_1} \cap B_{i_2}) = \frac{(N-2)!}{N!} = \frac{1}{N(N-1)} \neq P_N(B_{i_1})P_N(B_{i_2}) = \frac{1}{N^2}.$$

Par conséquent les B_i *ne sont pas* 2 à 2 indépendants.

On constate, comme on pouvait s'en douter, que $P_N(B_{i_1} \cap \cdots \cap B_{i_j})$ ne dépend que de N et de j et pas des valeurs particulières des indices $i_1, \ldots, i_j$. Il est commode de réécrire ce résultat en notant J une partie de cardinal $j \geqslant 1$ de $[\![1, N]\!]$ et en posant :

$$D_J := \bigcap_{i \in J} B_i,$$

sous la forme

$$P_N(D_J) = \frac{(N-j)!}{N!}, \quad j = \operatorname{card} D_J.$$

Comme il y a en tout C_N^j parties J de $[\![1, N]\!]$ ayant pour cardinal j, on en déduit que pour $1 \leqslant j \leqslant N$:

$$\sum_{i_1 < i_2 < \cdots < i_j} P_N(B_{i_1} \cap \cdots \cap B_{i_j}) = \sum_{\substack{J \subset [\![1,N]\!] \\ \operatorname{card} J = j}} P_N(D_J) = C_N^j \frac{(N-j)!}{N!} = \frac{N!(N-j)!}{j!(N-j)!N!} = \frac{1}{j!}.$$

2) L'évènement $E_{N,0}$ « aucun nageur ne retrouve ses affaires », s'exprime facilement en fonction des B_i :

$$E_{N,0} = \bigcap_{i=1}^{N} B_i^c.$$

Comme les B_i ne sont même pas 2 à 2 indépendants, il ne sont *a fortiori* pas mutuellement indépendants et il en va de même pour les complémentaires. On ne peut donc pas calculer $P_N(E_{N,0})$ comme le produit des $P_N(B_i^c)$.

Les calculs faits à la question précédente nous incitent alors à passer à l'évènement complémentaire $\bigcup_{1 \leqslant i \leqslant N} B_i$ et à calculer la probabilité de cette réunion par la formule de Poincaré [3].

Avec les notations D_J introduites ci-dessus, cette formule s'écrit :

$$P\left(\bigcup_{i=1}^{N} B_i\right) = \sum_{j=1}^{N} (-1)^{j+1} \sum_{\substack{J \subset [\![1,N]\!] \\ \operatorname{card} J = j}} P_N(D_J),$$

3. Cette union n'est pas disjointe puisque $P_N(B_{i_1} \cap B_{i_2}) = \frac{1}{N(N-1)} > 0$ interdit que $B_{i_1} \cap B_{i_2}$ ne soit vide. On ne peut donc pas utiliser ici l'additivité de P_N.

d'où

$$P_N(E_{N,0}) = 1 - P\left(\bigcup_{i=1}^{N} B_i\right) = 1 - \sum_{j=1}^{N} (-1)^{j+1} \frac{1}{j!} = \sum_{j=0}^{N} \frac{(-1)^j}{j!}.$$

3) Fixons k nageurs retrouvant leurs affaires. On peut exprimer le *nombre* de permutations des $N-k$ autres nageurs telles qu'aucun d'eux ne retrouve ses affaires à l'aide de $P_{N-k}(E_{N-k,0})$. En effet, puisque P_{N-k} est l'équiprobabilité sur Ω_{N-k},

$$P_{N-k}(E_{N-k,0}) = \frac{\operatorname{card} E_{N-k,0}}{\operatorname{card} \Omega_{N-k}} = \frac{\operatorname{card} E_{N-k,0}}{(N-k)!},$$

d'où

$$\operatorname{card} E_{N-k,0} = (N-k)!\, P_{N-k}(E_{N-k,0}).$$

On remarque ensuite qu'il y a une relation simple entre $\operatorname{card} E_{N-k,0}$ et $\operatorname{card} E_{N,k}$. En effet une permutation de $[\![1, N]\!]$ ayant exactement k points fixes se caractérise par l'ensemble J de cardinal k formé des points fixes et une permutation ω' sans aucun point fixe de l'ensemble $[\![1, N]\!] \backslash J$. Il est clair que le nombre des permutations du type ω' ne dépend de J que par k et N, donc est égal au nombre de permutations sans point fixe de $[\![1, N-k]\!] = [\![1, N]\!] \backslash J_0$ pour $J_0 = [\![N-k+1, N]\!]$. Ce nombre est exactement $\operatorname{card} E_{N-k,0}$. Comme le nombre de choix possibles pour J est C_N^k, on en déduit que

$$\operatorname{card} E_{N,k} = C_N^k \operatorname{card} E_{N-k,0}.$$

Ceci nous permet de calculer $P_N(E_{N,k})$:

$$\begin{aligned} P_N(E_{N,k}) = \frac{\operatorname{card} E_{N,k}}{N!} &= \frac{C_N^k \operatorname{card} E_{N-k,0}}{N!} \\ &= \frac{C_N^k (N-k)!\, P_{N-k}(E_{N-k,0})}{N!} \\ &= \frac{1}{k!} P_{N-k}(E_{N-k,0}), \end{aligned}$$

d'où finalement, en appliquant la formule obtenue à la question précédente pour $P_N(E_{N,0})$ avec $N-k$ au lieu de N,

$$P_N(E_{N,k}) = \frac{1}{k!} \sum_{j=0}^{N-k} \frac{(-1)^j}{j!}.$$

4) Pour k fixé, on peut écrire

$$p_k = \lim_{N \to \infty} P_N(E_{N,k}) = \frac{1}{k!} \sum_{j=0}^{+\infty} \frac{(-1)^j}{j!},$$

sous réserve que cette série converge dans $\mathbb{R}$. Il y a au moins deux façons de justifier cette convergence.

1. On peut remarquer qu'il s'agit d'une *série alternée* dont la valeur absolue du terme général tend vers 0 en décroissant. On sait qu'une telle série converge toujours dans $\mathbb{R}$. En prime on sait aussi que sa somme, ici p_k, est encadrée par deux sommes partielles consécutives.
2. On peut aussi remarquer que la fonction exponentielle a le développement en série entière suivant avec rayon de convergence infini :

$$\mathrm{e}^x = \sum_{j=0}^{+\infty} \frac{x^j}{j!},$$

cette série convergeant absolument pour tout x réel. En particulier, pour $x = -1$,

$$\sum_{j=0}^{+\infty} \frac{(-1)^j}{j!} = \mathrm{e}^{-1}.$$

Nous obtenons ainsi :

$$p_k = \lim_{N\to\infty} P_N(E_{N,k}) = \frac{\mathrm{e}^{-1}}{k!}.$$

Pour voir que la suite de réels *positifs* $(p_k)_{k\in\mathbb{N}}$ définit une probabilité P sur $\mathbb{N}$, il suffit de vérifier que la série de terme général p_k a pour somme 1. On pourra alors poser pour tout $k \in \mathbb{N}$, $P(\{k\}) := p_k$ et pour tout $A \subset \mathbb{N}$, donc au plus dénombrable, $P(A) := \sum_{k\in A} p_k$, ce qui définit bien une probabilité sur $\mathbb{N}$. Cette vérification est immédiate grâce au développement en série entière de la fonction exponentielle :

$$\sum_{k=0}^{+\infty} p_k = \mathrm{e}^{-1} \sum_{k=0}^{+\infty} \frac{1}{k!} = \mathrm{e}^{-1}\mathrm{e}^{1} = 1.$$

On reconnaît en P la loi de Poisson de paramètre 1. Rappelons que la loi de Poisson de paramètre $a > 0$ est l'unique probabilité Q_a sur $\mathbb{N}$ vérifiant :

$$\forall k \in \mathbb{N}, \quad Q_a(\{k\}) = \frac{\mathrm{e}^{-a}a^k}{k!}.$$

Avec ces notations, $P = Q_1$.

5) On s'intéresse maintenant à la vitesse de convergence de $P_N(E_{N,k})$. Pour cela, on peut exploiter le fait que, pour chaque k fixé, $P_N(E_{N,k})$ est la somme partielle de rang $N-k$ d'une série alternée dont la valeur absolue du terme général tend vers 0 en décroissant. On sait qu'alors la somme de cette série, soit ici p_k, est encadrée par deux sommes partielles consécutives, quel que soit leur rang [4]. En particulier, la distance entre p_k et la somme partielle de rang $N-k$ est inférieure strictement à la distance entre cette somme partielle et celle de rang $N-k+1$, ce qui s'écrit :

$$|P_N(E_{N,k}) - p_k| < \left| \frac{1}{k!} \sum_{j=0}^{N-k+1} \frac{(-1)^j}{j!} - \frac{1}{k!} \sum_{j=0}^{N-k} \frac{(-1)^j}{j!} \right| = \frac{1}{k!(N-k+1)!}.$$

4. Pour une série alternée dont aucun terme n'est nul, cet encadrement est strict.

Cette majoration est valide pour tout $k \in [\![0, N]\!]$. Par sommation de l'inégalité ci-dessus pour $k \in [\![0, n]\!]$, avec $n \leqslant N$, on obtient :

$$\sum_{j=0}^{n} |P_N(E_{N,j}) - p_j| \leqslant \sum_{j=0}^{n} \frac{1}{j!(N-j+1)!} \leqslant \sum_{j=0}^{n} \frac{1}{j!(N-n+1)!} \leqslant \frac{1}{(N-n+1)!} \sum_{j=0}^{+\infty} \frac{1}{j!}.$$

On reconnaît dans la somme de cette dernière série le nombre e. Nous avons ainsi établi que :

$$\forall n \in [\![0, N]\!], \qquad \sum_{j=0}^{n} |P_N(E_{N,j}) - p_j| \leqslant \frac{\mathrm{e}}{(N+1-n)!}.$$

6) *Application.* Le majorant d'erreur obtenu ci-dessus pour l'approximation de $P_N(E_{N,k})$ par p_k sera numériqueemnt très petit pourvu que k ne soit pas trop proche de N. Or la masse de la probabilité P dont le support est $\mathbb{N}$ est en fait très concentrée sur les entiers proches de 0. Par exemple, à 6×10^{-4} près, elle est portée par $[\![0, 5]\!]$, comme le montre le petit calcul suivant :

$$P([\![0,5]\!]) = \sum_{j=0}^{5} p_j = \sum_{j=0}^{5} \frac{\mathrm{e}^{-1}}{j!} = \mathrm{e}^{-1}\Big(1 + \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \frac{1}{4!} + \frac{1}{5!}\Big) \simeq 0{,}999\,406.$$

Si on regarde alors l'approximation de $P_{N,j}$ par p_j pour $j \leqslant 5$ et $N \geqslant 12$, on aura

$$|P_N(E_{N,j}) - p_j| < \frac{1}{j!(N-j+1)!} \leqslant \frac{1}{8!} \simeq 0{,}000\,024\,8 < 3 \times 10^{-5}.$$

Cette majoration assez grossière est donc déjà suffisante pour avoir une approximation des $P_{N,j}$ pour $N \geqslant 12$ et $0 \leqslant j \leqslant 5$ par p_j qui ne dépend pas de N, avec une erreur inférieure à 10^{-4}. Voici les valeurs numériques obtenues.

j	0	1	2	3	4	5
p_j	0,367 879	0,367 879	0,183 940	0,061 313	0,015 328	0,003 066

La probabilité qu'au moins 6 nageurs retrouvent leurs affaires est :

$$1 - \sum_{j=0}^{5} P_N(E_{N,j}) \simeq 1 - \sum_{j=0}^{5} p_j \simeq 1 - 0{,}999\,406 = 0{,}000\,594.$$

Pour $N \geqslant 12$, l'erreur ε_N commise dans l'approximation ci-dessus de $\sum_{j=0}^{5} P_N(E_{N,j})$ par $\sum_{j=0}^{5} p_j$ est majorée par :

$$\varepsilon_N \leqslant \sum_{j=0}^{5} |P_N(E_{N,j}) - p_j| \leqslant \frac{\mathrm{e}}{(N+1-5)!} \leqslant \frac{\mathrm{e}}{8!} \leqslant 0{,}000\,067\,5 < 10^{-4}.$$

Solution 5.6 *Pile ou face triphasé*

Trois personnes nommées A, B, C lancent à tour de rôle la même pièce de monnaie (ABCABCABC...). La probabilité d'obtenir pile lors d'un lancer est p $(0 < p < 1)$. Le gagnant est le premier qui obtient pile (la partie *s'arrête* alors). On note A_n (resp. B_n, C_n) l'évènement *A gagne la partie lors du n^e lancer* (resp. B, C)

Remarque préliminaire. La règle du jeu adoptée ne permet pas, en toute rigueur, de considérer les lancers de la pièce comme une suite d'épreuves répétées indépendantes. Par exemple, la probabilité d'obtenir pile au deuxième lancer n'est pas p. En fait, p est la probabilité d'obtenir pile au deuxième lancer *sachant* que ce lancer a bien lieu, c'est-à-dire sachant que l'on a obtenu face au premier lancer. La résolution correcte des questions 1) et 2) requiert donc l'utilisation du conditionnement en chaîne au lieu de l'indépendance.

1) *Calcul de $P(A_1)$, $P(B_2)$ et $P(C_3)$.*

Le premier lancer ayant de toutes façons lieu, on a clairement $P(A_1) = p$. Pour calculer $P(B_2)$, on remarque que B_2 ne peut avoir lieu que si A a perdu le premier lancer. Donc B_2 implique A_1^c, autrement dit $B_2 \subset A_1^c$. Par conséquent $B_2 = B_2 \cap A_1^c$ et :

$$P(B_2) = P(B_2 \cap A_1^c) = P(B_2 \mid A_1^c)P(A_1^c) = pq,$$

en notant $q = 1 - p$ la probabilité d'obtenir face à un lancer sachant qu'il a lieu.

De même, la réalisation de C_3 nécessite celle de A_1^c et de B_2^c d'où $C_3 = C_3 \cap B_2^c \cap A_1^c$ et :

$$P(C_3) = P(A_1^c \cap B_2^c \cap C_3) = P(A_1^c)P(B_2^c \mid A_1^c)P(C_3 \mid A_1^c \cap B_2^c) = pq^2.$$

En effet $P(C_3 \mid A_1^c \cap B_2^c)$ est la probabilité d'obtenir pile au troisième lancer *sachant* que ce lancer a lieu. Donc $P(C_3 \mid A_1^c \cap B_2^c) = p$.

Les évènements A_1 et B_2 étant incompatibles $(A_1 \cap B_2 = \varnothing)$ ne peuvent être indépendants :

$$P(A_1 \cap B_2) = P(\varnothing) = 0 \neq P(A_1)P(B_2) = p^2q.$$

2) *Calcul de $P(A_n)$, $P(B_n)$ et $P(C_n)$.*

Il est clair que chaque joueur ne peut gagner qu'à un lancer où c'est son tour de jouer. Ainsi :

- A ne peut gagner qu'à un lancer dont le numéro est de la forme $n = 3k + 1$, $k \in \mathbb{N}$. Pour les autres valeurs de n, $P(A_n) = 0$.
- B ne peut gagner qu'à un lancer dont le numéro est de la forme $n = 3k + 2$, $k \in \mathbb{N}$. Pour les autres valeurs de n, $P(B_n) = 0$.
- C ne peut gagner qu'à un lancer dont le numéro est de la forme $n = 3k$, $k \in \mathbb{N}^*$. Pour les autres valeurs de n, $P(C_n) = 0$.

Afin de généraliser les calculs de la question 1), il est commode d'introduire les notations suivantes. Pour $i \in \mathbb{N}^*$, désignons par F_i l'évènement *obtention de face au i^e lancer* et par G_i l'évènement *obtention de pile au i^e lancer*. Pour tout $n \geqslant 2$, les

probabilités d'obtenir pile au n^{e} lancer ou face au même lancer *sachant que ce lancer a bien lieu* sont respectivement :

$$P\Big(G_n \mid \bigcap_{i=1}^{n-1} F_i\Big) = p, \quad \text{et} \quad P\Big(F_n \mid \bigcap_{i=1}^{n-1} F_i\Big) = q.$$

D'autre part la règle du conditionnement en chaîne nous fournit :

$$P\Big(\bigcap_{i=1}^{n} F_i\Big) = P(F_1)P(F_2 \mid F_1)P(F_3 \mid F_1 \cap F_2)\cdots P\Big(F_n \mid \bigcap_{i=1}^{n-1} F_i\Big) = q^n.$$

En remarquant que

$$A_{3k+1} = G_{3k+1} \cap \Big(\bigcap_{i=1}^{3k} F_i\Big),$$

on a alors :

$$P(A_{3k+1}) = P\Big(G_{3k+1} \mid \bigcap_{i=1}^{3k} F_i\Big)P\Big(\bigcap_{i=1}^{3k} F_i\Big) = pq^{3k}.$$

La même méthode nous donne :

$$P(B_{3k+2}) = pq^{3k+1},\ k \in \mathbb{N} \quad \text{et} \quad P(C_{3k}) = pq^{3k-1},\ k \in \mathbb{N}^*.$$

3) *Probabilité de gagner de chacun des trois joueurs.*
Notons V (resp. V', V'') l'évènement $\{A$ gagne la partie$\}$ (resp. B, C). On a :

$$V = \bigcap_{k\in\mathbb{N}} A_{3k+1}, \quad V' = \bigcap_{k\in\mathbb{N}} B_{3k+2}, \quad V'' = \bigcap_{k\in\mathbb{N}*} C_{3k-1},$$

ces réunions portant sur des évènements deux à deux disjoints. On en déduit :

$$\begin{aligned}
P(V) &= \sum_{k=0}^{+\infty} pq^{3k} = \frac{p}{1-q^3} = \frac{p}{(1-q)(1+q+q^2)} = \frac{1}{1+q+q^2}\\
P(V') &= \sum_{k=0}^{+\infty} pq^{3k+1} = \frac{pq}{1-q^3} = \frac{pq}{(1-q)(1+q+q^2)} = \frac{q}{1+q+q^2}\\
P(V'') &= \sum_{k=1}^{+\infty} pq^{3k-1} = \frac{pq^2}{1-q^3} = \frac{pq^2}{(1-q)(1+q+q^2)} = \frac{q^2}{1+q+q^2}.
\end{aligned}$$

Les séries géométriques qui interviennent dans ces calculs ont toutes pour raison q^3 qui vérifie $0 < q^3 < 1$. Elles sont donc convergentes. On remarque, comme on pouvait s'y attendre, que A est avantagé par rapport à B qui est lui même avantagé par rapport à C : $P(V'') = qP(V') < P(V') = qP(V) < P(V)$ puisque $q < 1$. Par exemple si la pièce est équilibrée, $p = q = 1/2$, on obtient : $P(A) = 4/7$, $P(B) = 2/7$, $P(C) = 1/7$.

4) *Probabilité qu'il y ait un vainqueur*

L'évènement $W = \{$il y a un vainqueur$\}$ s'écrit comme la réunion disjointe : $W = V \cup V' \cup V''$. On a ainsi :

$$P(W) = P(V) + P(V') + P(V'') = \frac{1}{1+q+q^2} + \frac{q}{1+q+q^2} + \frac{q^2}{1+q+q^2} = 1.$$

Il est donc certain que la partie durera un temps fini. Mais bien sûr comme ce temps est lui même aléatoire, on ne peut pas le majorer avec certitude par une durée déterministe.

Remarque sur la modélisation. Lors de la première utilisation de cet exercice dans un sujet d'examen, une solution plus simple mais manquant de rigueur a été présentée dans plusieurs copies et considérée avec bienveillance par les correcteurs. Elle consiste à admettre l'indépendance des lancers et à en déduire que $P(A_{3k+1}) = P(X = 3k+1)$, $P(B_{3k+2}) = P(X = 3k+2)$ et $P(C_{3k}) = P(X = 3k)$ où X suit une loi géométrique de paramètre p. comme nous l'avons vu ci-dessus, l'indépendance des lancers n'est pas compatible avec la règle du jeu adoptée. On peut néanmoins accepter ce raisonnement si l'on admet que tout se passe comme s'il y avait un arbitre qui lance indéfiniment la pièce, les trois joueurs A, B, C pariant chacun son tour que le lancer va donner pile [5].

Solution 5.9 *Probabilité de gagner un jeu sur son service au tennis*

On rappelle les notations d'évènements suivantes.

$$\begin{aligned} A_n &:= \{A \text{ gagne le } n^{\text{e}} \text{ échange}\}, \quad B_n := \{B \text{ gagne le } n^{\text{e}} \text{ échange}\}, \\ E_{i,j} &:= \{\text{Au bout de } i+j \text{ échanges, } A \text{ a } i \text{ points et } B \text{ en a } j\}. \end{aligned}$$

1) A peut gagner avec 4 points si le score est 4 à 0 ou 4 à 1 ou 4 à 2, autrement dit si l'un des évènements $E_{4,0}$, $E_{4,1}$ ou $E_{4,2}$ se réalise. On a donc

$$G' := \{A \text{ gagne le jeu avec 4 points}\} = E_{4,0} \cup E_{4,1} \cup E_{4,2}.$$

L'autre possibilité est que A gagne avec plus de 4 points et cela se produit si et seulement si le jeu se termine sur un score 5 à 3 ou 6 à 4 ou 7 à 5 ou ..., bref sur un score de la forme $k+2$ à k pour k décrivant $[\![3, +\infty[\![$. Ceci correspond à la réalisation de l'un des évènements de la suite infinie $(E_{k+2,k})_{k\geqslant 3}$. Ainsi

$$G'' := \{A \text{ gagne le jeu avec plus de 4 points}\} = \bigcup_{k\geqslant 3} E_{k+2,k}.$$

Il est clair que $G = G' \cup G''$, d'où la décomposition

$$G = E_{4,0} \cup E_{4,1} \cup E_{4,2} \cup \Big(\bigcup_{k\geqslant 3} E_{k+2,k}\Big). \tag{11.9}$$

5. Par ailleurs, le nombre X de lancers nécessaire pour avoir un vainqueur suit effectivement une loi géométrique d'après les résultat de la question 2).

2) *Calcul de $P(E_{4,0})$, $P(E_{4,1})$ et $P(E_{4,2})$.*
Comme $E_{4,0} = A_1 \cap A_2 \cap A_3 \cap A_4$, l'indépendance des échanges nous donne :

$$P(E_{4,0}) = P(A_1)P(A_2)P(A_3)P(A_4) = p^4.$$

Le gain du jeu par A sur le score de 4 à 1 (évènement $E_{4,1}$) se réalise si et seulement si A gagne le cinquième échange (A_5) et B marque un seul point lors des quatre premiers échanges ($E_{3,1}$). Ainsi $E_{4,1} = A_5 \cap E_{3,1}$. En considérant les 5 premiers échanges comme indépendants [6] on a donc $P(E_{4,1}) = P(E_{3,1})P(A_5) = P(E_{3,1})p$. On remarque que $P(E_{3,1})$ est la probabilité que A ait exactement trois succès dans une suite de 4 épreuves répétées indépendantes, c'est donc $C_4^3 p^3(1-p) = 4p^3(1-p)$. On a donc

$$P(E_{4,1}) = P(E_{3,1})p = 4p^4(1-p).$$

Par un raisonnement analogue, on voit que $E_{4,2} = A_6 \cap E_{3,2}$ et

$$P(E_{4,2}) = P(E_{3,2})P(A_6) = C_5^3 p^3(1-p)^2 p = 10p^4(1-p)^2.$$

3) Pour tout $k \geqslant 3$, on a :

$$E_{k+2,k} = E_{3,3} \cap \Big(\bigcap_{4 \leqslant j \leqslant k} C_j \Big) \cap A_{2k+1} \cap A_{2k+2}, \tag{11.10}$$

où l'on a posé $C_j := (A_{2j-1} \cap B_{2j}) \cup (B_{2j-1} \cap A_{2j})$. Pour justifier cette décomposition, on remarque que $E_{k+2,k}$ est aussi l'évènement « A *gagne* le jeu avec $k+2$ points contre k à B ». Le jeu dure donc dans ce cas $(2k+2)$ échanges, donc au moins 8 échanges puisque $k \geqslant 3$. Celà implique l'égalité des 2 joueurs au bout des 6 premiers échanges. En effet, compte tenu des règles sur le gain du jeu, les seuls scores possibles au bout de 6 échanges sont 4 à 2 (et alors A remporte le jeu) ou 3 à 3 (les échanges doivent continuer) ou 2 à 4 (et B remporte le jeu). Les scores du type 6 à 0 ou 5 à 1 sont impossibles car alors un des deux joueurs serait arrivé le premier à 4 points avec plus de deux points d'avance et le jeu se serait arrêté avant le sixième échange.

Regardons d'abord le cas particulier de $E_{5,3}$. Cet évènement se réalise si et seulement si les deux joueurs sont à égalité au bout des 6 premiers échanges *et* A remporte le septième et le huitième. Donc

$$P(E_{5,3}) = E_{3,3} \cap A_7 \cap A_8.$$

Ceci est un cas particulier de la formule (11.10) en convenant qu'une intersection indexée par « $4 \leqslant j \leqslant 3$ » est égale à Ω (ce qui est logique puisque qu'aucun j ne vérifie cette condition, on ne met donc aucune restriction d'appartenance à un C_j pour qu'un évènement élémentaire ω soit dans cette « intersection »).

Pour $k \geqslant 4$, la réalisation de $E_{k+2,k}$ signifie la victoire de A au bout de $2k+2$ échanges avec cette fois $2k+2 \geqslant 10$. Ceci se produit si et seulement si *chacune* des trois conditions suivantes se réalise (on aura donc l'intersection des trois évènements correspondants) :

6. En toute rigueur, ce n'est pas vrai, mais du point de vue du calcul tout se passe comme s'ils l'étaient, voir la remarque 11.3 page 440.

– A et B sont à égalité au bout des 6 premiers échanges
– A et B se retrouvent à égalité à l'issue de chaque paire d'échanges depuis la paire 7^{e}, 8^{e} jusqu'à la paire $(2k-1)^{\text{e}}$, $(2k)^{\text{e}}$.
– A gagne les deux derniers échanges : le $(2k+1)^{\text{e}}$ et le $(2k+2)^{\text{e}}$.

L'évènement « A et B marquent chacun un point lors de la paire d'échanges $(2j-1)^{\text{e}}$ et $(2j)^{\text{e}}$ » s'écrit $C_j = (A_{2j-1} \cap B_{2j}) \cup (B_{2j-1} \cap A_{2j})$ puisque ou bien A marque le premier et B égalise, ou c'est l'inverse. La justification de (11.10) est donc complète.

4) *Calcul des $P(E_{k+2,k})$.*

L'évènement $E_{3,3}$ se réalise si et seulement si lors des 6 premiers échanges, A marque exactement 3 points (donc B aussi). En considérant ces 6 premiers échanges comme une suite d'épreuves répétées indépendantes, on a immédiatement

$$P(E_{3,3}) = C_6^3 p^3 (1-p)^3 = 20p^3(1-p)^3.$$

Les évènements $A_{2j-1} \cap B_{2j}$ et $B_{2j-1} \cap A_{2j}$ sont incompatibles (par exemple parce que le premier implique que A remporte l'échange numéro $(2j-1)$ tandis que le deuxième implique que B remporte ce même échange). On a donc

$$P(C_j) = P(A_{2j-1} \cap B_{2j}) + P(B_{2j-1} \cap A_{2j}) = p(1-p) + (1-p)p = 2p(1-p),$$

en utilisant l'indépendance des échanges pour la deuxième égalité. On remarque que la valeur trouvée ne dépend pas de j. On la notera r dans la suite pour alléger les écritures. Finalement, par indépendance des échanges on obtient :

$$\begin{aligned} P(E_{k+2,k}) &= P(E_{3,3}) \times \Big(\prod_{j=4}^{k} P(C_j)\Big) \times P(A_{2k+1})P(A_{2k+2}) \\ &= 20p^3(1-p)^3 r^{k-3} p^2 \\ &= 20p^5(1-p)^3 r^{k-3}. \end{aligned}$$

Cette formule reste valable dans le cas particulier $k = 3$ où l'on a directement

$$P(E_{5,3}) = P(E_{3,3} \cap A_7 \cap A_8) = P(E_{3,3})P(A_7)P(A_8) = 20p^3(1-p)^3 p^2.$$

5) Les $E_{k+2,k}$ sont deux à deux disjoints. En effet, soient $k \neq l$, la réalisation de $E_{k+2,k}$ implique que le jeu se termine au $(2k+2)^{\text{e}}$ échange tandis que celle de $E_{l+2,l}$ implique qu'il se termine au $(2l+2)^{\text{e}}$ échange. Comme $k \neq l$, $2k+2 \neq 2l+2$, donc les évènements $E_{k+2,k}$ et $E_{l+2,l}$ sont incompatibles. On peut ainsi utiliser la σ-additivité :

$$\begin{aligned} P\Big(\bigcup_{k \geqslant 3} E_{k+2,k}\Big) &= \sum_{k=3}^{+\infty} P(E_{k+2,k}) \\ &= 20p^5(1-p)^3 \sum_{k=3}^{+\infty} r^{k-3} \\ &= 20p^5(1-p)^3 \sum_{j=0}^{+\infty} r^{j} \\ &= 20p^5(1-p)^3 \frac{1}{1-r}. \end{aligned}$$

La série géométrique de raison r qui intervient dans ce calcul est bien convergente puisque $r = 2p(1-p)$ est dans $[0,1[$. En effet si $p < 1/2$ alors $2p < 1$ et $2p(1-p) < 1-p \leqslant 1$, si $p > 1/2$ alors $1-p < 1/2$ et on a la même majoration stricte pour r en échangeant les rôles de p et $1-p$, enfin si $p = 1/2$, $r = 2 \times 1/2 \times 1/2 = 1/2 < 1$. En fait on peut vérifier que le maximum de $2p(1-p)$ lorsque p parcourt $[0,1]$ est atteint en $p = 1/2$.

6) Les évènements figurant dans la décomposition (11.9) correspondent chacun au gain du jeu sur un score différent et sont donc deux à deux incompatibles. On en déduit

$$\begin{aligned} P(G) &= P(E_{4,0}) + P(E_{4,1}) + P(E_{4,2}) + P\Big(\bigcup_{k\geqslant 3} E_{k+2,k}\Big) \\ &= p^4\Big(1 + 4(1-p) + 10(1-p)^2 + \frac{20p(1-p)^3}{1-2p(1-p)}\Big) =: f(p). \quad (11.11) \end{aligned}$$

On remarque que pour $p = 0$, $P(G) = 0$ et pour $p = 1$, $P(G) = 1$, ce qui est conforme à l'intuition.

7) L'évènement $N = \{$le jeu continue indéfiniment sans vainqueur$\}$ s'écrit :

$$N = E_{3,3} \cap \Big(\bigcap_{j\geqslant 4} C_j\Big).$$

Il est donc inclus pour tout $n \geqslant 4$ dans l'évènement

$$N_n := E_{3,3} \cap \Big(\bigcap_{j=4}^{n} C_j\Big),$$

d'où

$$0 \leqslant P(N) \leqslant P(N_n) = 20p^3(1-p)^3 r^{n-3}.$$

Ceci étant vrai pour *tout* $n \geqslant 4$, ces inégalités *larges* se conservent par passage à la limite quand n tend vers l'infini. Comme $0 \leqslant r < 1$, la limite de $P(N_n)$ est nulle, on en déduit $P(N) = 0$. La probabilité que le jeu continue indéfiniment est nulle (bien que l'évènement N ne soit pas l'ensemble vide).

8) Notons H l'évènement « B gagne le jeu ». Les trois évènements G, H et N forment une partition de Ω, donc $P(G) + P(H) + P(N) = 1$. Comme nous venons de voir que $P(N) = 0$, cette égalité se réduit à $P(G) + P(H) = 1$. Dans le cas particulier où $p = 1/2$, A ne tire aucun avantage de son service, et on doit avoir par symétrie $P(G) = P(H)$, d'où $P(G) = 1/2$. Vérifions le sur la formule générale (11.11) pour $P(G)$. Elle s'écrit ici

$$P(G) = \frac{1}{16}\Big(1 + 2 + \frac{10}{4} + \frac{20/16}{1-1/2}\Big) = \frac{1}{16}\Big(3 + \frac{5}{2} + \frac{5}{2}\Big) = \frac{1}{2}.$$

Remarque 11.1. On aurait pu dans le cas général calculer $P(H)$ par la même méthode qui nous a conduit à (11.11) pour $P(G)$. Il suffit de remarquer que la probabilité que B gagne l'échange est $q = 1-p$. Il est clair alors que $P(H) = f(q)$. Comme

$P(N) = 0$, nécessairement $f(p) + f(q) = 1$. Graphiquement ceci se traduit par le fait que la représentation graphique de f admet le point $(1/2, 1/2)$ comme centre de symétrie. La figure 11.4 (réalisée à l'aide du logiciel Scilab) semble le confirmer.

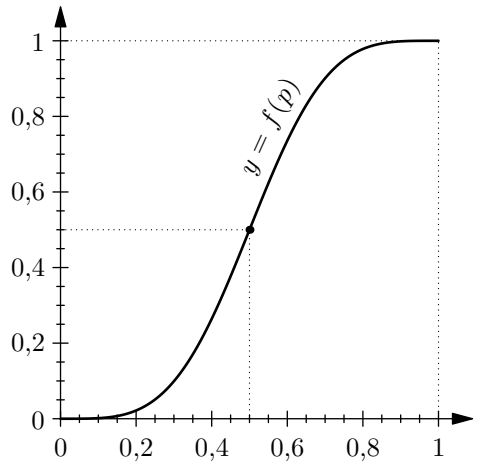


FIGURE 11.4 – Représentation graphique de $P(G) = f(p)$

Il faut bien avouer que cette symétrie ne saute pas aux yeux à la lecture de (11.11). Pour le vérifier par le calcul, commençons par réécrire (11.11) sous la forme :

$$f(p) = p^4\Big(1 + 4q + 10q^2 + \frac{20pq^3}{1-2pq}\Big).$$

On a alors

$$f(p) + f(q) = p^4 + q^4 + 4pq(p^3 + q^3) + 10p^2q^2(p^2 + q^2) + \frac{20p^3q^3(p^2 + q^2)}{1 - 2pq}.$$

De là on arrive facilement à $f(p) + f(q) = 1$ grâce aux formules suivantes obtenues en développant $(p + q)^n$ par la formule du binôme en notant que $p + q = 1$:

$$p^2 + q^2 = 1 - 2pq, \quad p^3 + q^3 = 1 - 3pq, \quad p^4 + q^4 = 1 - 4pq + 2p^2q^2.$$

Remarque 11.2. Un coup d'oeil sur la figure 11.4 et sur le tableau de valeurs numériques ci-dessous permet de voir l'effet amplificateur des déséquilibres dû aux règles du tennis. Dès que p s'éloigne de $1/2$, $f(p)$ se rapproche très vite de 0 ou de 1.

p	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9
$f(p)$	0,001	0,022	0,099	0,264	0,500	0,736	0,901	0,978	0,999

Remarque 11.3. L'hypothèse d'indépendance des échanges n'est en réalité valable que pour les 4 premiers échanges (puisqu'on est sûr qu'ils auront lieu). Pour les

autres échanges, en raison de l'arrêt au bout d'un nombre aléatoire d'échanges, on ne peut pas avoir indépendance. Pour s'en convaincre, il suffit de remarquer que les évènements $E_{4,0}$ et A_5 ont des probabilités qui ne valent ni 0 ni 1 et sont *incompatibles* : la réalisation de $E_{4,0}$ impliquant le gain du jeu au bout de 4 échanges, A ne peut gagner le cinquième ! Ils ne sont donc pas indépendants. Pourtant la réalisation de $E_{4,0}$ ne dépend que du résultat des 4 premiers échanges et celle de A_5 que du résultat du cinquième échange. Si ces échanges étaient indépendants, ces deux évènements devraient l'être.

Il y a deux façons de remédier à ce problème. La première est de dire que toutes les probabilités calculées dans ce problème sont les mêmes que si les échanges continuaient indéfiniment après le gain du jeu par l'un des deux joueurs. On aurait ainsi une suite infinie d'épreuves répétées indépendantes. L'autre façon est de faire du conditionnement en chaîne en considérant que ce n'est pas $P(A_n)$ qui est constante égale à p comme le fait abusivement l'énoncé mais les probabilités *conditionnelles* $P(A_n \mid H_{n-1})$ où H_{n-1} représente n'importe quelle suite de résultats détaillés des $n-1$ premiers échanges compatible avec la tenue du n^{e} échange. Par exemple pour $n = 5$, on peut mettre à la place de H_4 les évènements $A_1 \cap A_2 \cap B_3 \cap A_4$ ou $B_1 \cap B_2 \cap B_3 \cap A_4$.

Solution 5.10 *Inégalité et ball trap accéléré*

Soit $(A_i)_{i\geqslant 1}$ une suite d'évènements *indépendants*. On note $p_i = P(A_i)$ et on suppose que

$$a := \sum_{i=1}^{+\infty} p_i < +\infty.$$

Pour tout $n \geqslant 1$, on définit l'application $S_n : \Omega \to \mathbb{R}$ par :

$$\forall \omega \in \Omega, \quad S_n(\omega) = \sum_{i=1}^{n} \mathbf{1}_{A_i}(\omega).$$

Le but de cet exercice est de prouver l'inégalité :

$$\forall n, k \in \mathbb{N}^*, \quad P\Big(S_n \geqslant k\Big) \leqslant \frac{a^k}{k!}.$$

1) Remarquons d'abord que si $k > n$, $P(S_n \geqslant k) = P(\varnothing) = 0$. En effet comme $\mathbf{1}_{A_i}(\omega)$ est majoré par 1, on a pour tout $\omega \in \Omega$, $S_n(\omega) \leqslant n$ et par conséquent l'évènement $\{S_n \geqslant k\}$ se réduit à l'ensemble vide. La majoration $P(S_n \geqslant k) \leqslant \frac{a^k}{k!}$ est alors trivialement vérifiée. Dans tout ce qui suit, on suppose $1 \leqslant k \leqslant n$.

2) On note $B_{n,k} := \{\omega \in \Omega \ ; \ \sum_{i=1}^{n} \mathbf{1}_{A_i}(\omega) \geqslant k\}$. C'est l'évènement « réalisation d'*au moins* k des A_i ». Si ω appartient à $B_{n,k}$ (ou réalise l'évènement $B_{n,k}$), il existe donc un ensemble d'indices $F = F_\omega$ de cardinal k tel que ω réalise chacun des A_i pour $i \in F$. Ceci s'écrit :

$$\forall \omega \in B_{n,k}, \ \exists F = F_\omega \subset [\![1, n]\!], \ \operatorname{card} F = k, \quad \omega \in \bigcap_{i \in F} A_i,$$

soit encore :

$$\forall \omega \in B_{n,k}, \quad \omega \in \bigcup_{\substack{F \subset [\![1,n]\!] \\ \text{card } F = k}} \bigcap_{i \in F} A_i,$$

ce qui justifie l'inclusion :

$$B_{n,k} \subset \bigcup_{\substack{F \subset [\![1,n]\!] \\ \text{card } F = k}} \bigcap_{i \in F} A_i.$$

3) En utilisant successivement la sous-additivité de P et l'indépendance des A_i, on en déduit immédiatement que

$$P(B_{n,k}) \leqslant \sum_{\substack{F \subset [\![1,n]\!] \\ \text{card } F = k}} P\Big(\bigcap_{i \in F} A_i \Big) = \sum_{\substack{F \subset [\![1,n]\!] \\ \text{card } F = k}} \prod_{i \in F} p_i.$$

4) On note $a_n := \sum_{i=1}^{n} p_i$. On peut alors voir a_n^k comme le produit de k sommes identiques de n termes. La formule de développement d'un produit de sommes nous donne :

$$a_n^k = \sum_{(i_1,\dots,i_k) \in [\![1,n]\!]^k} p_{i_1} \dots p_{i_k}.$$

Sous cette forme, on a une somme de n^k termes *tous positifs ou nuls*. En laissant tomber tous ceux d'entre eux qui correspondent à un vecteur d'indices $(i_1, \dots, i_k)$ où les i_j ne sont pas tous distincts, on obtient la minoration suivante de a_n^k :

$$a_n^k \geqslant \sum_{\substack{(i_1,\dots,i_k) \in [\![1,n]\!]^k \\ \text{les } i_j \text{ tous distincts}}} p_{i_1} \dots p_{i_k} = k! \sum_{1 \leqslant i_1 < i_2 < \dots < i_k \leqslant n} p_{i_1} \dots p_{i_k} = k! \sum_{\substack{F \subset [\![1,n]\!] \\ \text{card } F = k}} \prod_{i \in F} p_i.$$

5) Comme a_n est la n^{e} somme partielle d'une série à termes positifs ou nuls, elle vérifie $a_n \leqslant \sum_{i=1}^{+\infty} p_i = a$. En combinant ceci avec les résultats des questions 3) et 4), on aboutit à

$$\forall n \geqslant 1, \ \forall k \in [\![1,n]\!], \quad P\Big(S_n \geqslant k\Big) = P(B_{n,k}) \leqslant \frac{a_n^k}{k!} \leqslant \frac{a^k}{k!}.$$

Ce dernier majorant est aussi valable pour $k > n$ d'après la question 1).

6) *Application à un problème de tir.* Dans un stand de tir, une cible mobile traverse le champ visuel d'un tireur (une traversée par épreuve). À chaque épreuve, le tireur tire un coup sur la cible. D'une épreuve à la suivante, la vitesse de la cible augmente de 20 %. On suppose que pour un tireur donné, la probabilité de toucher la cible est *inversement proportionnelle* à la vitesse de la cible. Les tirs sont supposés indépendants, le tireur dispose d'autant de cartouches qu'il le souhaite et le défi qu'il doit relever est celui de toucher au moins 20 fois la cible. Notons p la probabilité que le tireur touche la cible lors du premier passage.

En utilisant le résultat démontré ci-dessus, nous nous proposons de majorer indépendamment de p la probabilité de réussir le défi. On note n le nombre de tirs et on suppose dans un premier temps que ce nombre est choisi par le tireur *avant* le début

de l'épreuve. Pour $i \geqslant 1$, v_i et p_i désignent respectivement la vitesse de la cible et la probabilité qu'elle soit touchée au i^{e} passage. Soit S_n le nombre de tirs réussis. Nous devons majorer $P(S_n \geqslant 20)$.

Les hypothèses de l'énoncé se traduisent par les égalités :

$$p_1 = p, \quad \text{et} \quad \forall i \geqslant 1, \quad p_i = \frac{c}{v_i}, \quad v_{i+1} = 1{,}2\, v_i,$$

où c est une constante. On en déduit immédiatement :

$$\forall i \geqslant 1, \quad \frac{p_{i+1}}{p_i} = \frac{\frac{c}{v_{i+1}}}{\frac{c}{v_i}} = \frac{v_i}{v_{i+1}} = \frac{1}{1{,}2} = \frac{5}{6}.$$

On voit ainsi que $(p_i)_{i\geqslant 1}$ est la suite géométrique de premier terme $p_1 = p$ et de raison $5/6$, d'où

$$\forall i \geqslant 1, \quad p_i = p\Big(\frac{5}{6}\Big)^{i-1}.$$

La série géométrique associée de raison $5/6$ est donc convergente ($|5/6| < 1$) et sa somme a se calcule comme suit :

$$a = \sum_{i=1}^{+\infty} p_i = p \sum_{j=0}^{+\infty} \Big(\frac{5}{6}\Big)^j = \frac{p}{1 - 5/6} = 6p.$$

En majorant la probabilité inconnue p par 1, on voit alors, grâce au résultat de la question 5), que

$$\forall n \geqslant 1, \quad P(S_n \geqslant 20) \leqslant \frac{(6p)^{20}}{20!} \leqslant \frac{6^{20}}{20!} \leqslant 1{,}51 \times 10^{-3}.$$

Remarque 11.4. On peut donc dire que quel que soit le nombre de tirs choisi *a priori* par le tireur, sa probabilité de gagner le défi est majorée par 0,001 51. On notera cependant que l'énoncé était un peu flou et ne précisait pas que le nombre de tirs devait être choisi *avant* le début de l'épreuve. Une modélisation plus réaliste serait de considérer que le nombre de tirs est une variable aléatoire N : le tireur s'arrête quand il a gagné, ou quand il renonce. La valeur de N n'est pas connue à l'avance. On peut quand même répondre à la question de l'énoncé dans ce cas, en s'appuyant sur les hypothèses suivantes.

- Pour tout $i \leqslant n$, $P(A_i \mid N = n) = p_i$ ne dépend pas de n.
- L'indépendance des tirs a lieu conditionnellement à N. Cela revient à dire que pour tout $n \geqslant 1$ tel que $P(N = n) \neq 0$, les n premiers tirs sont $P(\,.\,|N = n)$-indépendants.

Notons I l'ensemble des entiers n tels que $P(N = n) > 0$ et S_N le nombre de succès en N tirs. Avec les hypothèses formulées ci-dessus, la majoration établie aux questions 1)–5) reste valable sous la forme : $P(S_n \geqslant k \mid N = n) \leqslant \frac{a^k}{k!}$. En appliquant

la formule de conditionnement par tous les cas possibles, on obtient alors :

$$\begin{aligned} P(S_N \geqslant 20) &= \sum_{n\in I} P(S_N \geqslant 20 \mid N=n)P(N=n) \\ &= \sum_{n\in I} P(S_n \geqslant 20 \mid N=n)P(N=n) \\ &\leqslant \sum_{n\in I} \frac{6^{20}}{20!} P(N=n) \\ &= \frac{6^{20}}{20!} \sum_{n\in I} P(N=n) = \frac{6^{20}}{20!}. \end{aligned}$$

On retrouve ainsi le même majorant qu'avec un nombre de tirs non aléatoire.

11.6 Exercices du chapitre 6

Solution 11.6 *Code de la Route I*

1) Le nombre S de bonnes réponses du candidat est ici le nombre de succès dans une suite de 40 épreuves répétées indépendantes, avec pour chacune probabilité de succès 1/4. La variable aléatoire S suit donc la loi binomiale de paramètres 40 et 1/4.

2) Pour calculer $P(S \geqslant 36)$, on utilise la décomposition

$$P(S \geqslant 36) = \sum_{k=36}^{40} P(S=k).$$

$$\begin{aligned} P(S \geqslant 36) &= C_{40}^{36}\Big(\frac{1}{4}\Big)^{36}\Big(\frac{3}{4}\Big)^{4} + C_{40}^{37}\Big(\frac{1}{4}\Big)^{37}\Big(\frac{3}{4}\Big)^{3} + C_{40}^{38}\Big(\frac{1}{4}\Big)^{38}\Big(\frac{3}{4}\Big)^{2} \\ &\quad + C_{40}^{39}\Big(\frac{1}{4}\Big)^{39}\Big(\frac{3}{4}\Big)^{1} + C_{40}^{40}\Big(\frac{1}{4}\Big)^{40}\Big(\frac{3}{4}\Big)^{0} \\ &= \frac{1}{4^{40}}\big(C_{40}^{36}\times 3^4 + C_{40}^{37}\times 3^3 + C_{40}^{38}\times 3^2 + C_{40}^{39}\times 3 + C_{40}^{40}\times 1\big) \\ &= \frac{1}{4^{40}}\big(91390\times 81 + 9880\times 27 + 780\times 9 + 40\times 1 + 1\times 1\big) \\ &\simeq 6{,}35\times 10^{-18}. \end{aligned}$$

Cette probabilité est infime. Elle est du même ordre de grandeur que celle de trouver 10 milliards de fois *consécutives* (sans tricher !) les 6 bons numéros au Loto en jouant une grille à 6 numéros à chaque tirage. Pour cela, l'heureux candidat et ses descendants devraient jouer pendant 100 millions d'années en gagnant à chaque fois ! On ne prend donc aucun risque en pratique en considérant qu'un candidat ayant obtenu au moins 36 bonnes réponses n'a certainement pas répondu au hasard à *toutes* les questions.

Solution 6.5 *Code de la Route II*

Dans un modèle plus réaliste, le candidat répond à coup sûr lorsqu'il connaît la réponse à la question et s'il l'ignore, choisit au hasard entre les 4 réponses proposées. On suppose que toutes les questions sont indépendantes et que pour chacune de ces questions, la probabilité que le candidat connaisse la vraie réponse est p. Ce paramètre p mesure donc le vrai niveau du candidat.

1) Soit A_i l'évènement « le candidat *connaît* la réponse à la i^{e} question ». Notons $\mathbf{1}_{A_i}$ la variable aléatoire indicatrice de cet évènement :

$$\mathbf{1}_{A_i} = \begin{cases} 1 & \text{si le candidat connaît la réponse à la } i^{\text{e}} \text{ question ;} \\ 0 & \text{s'il l'ignore.} \end{cases}$$

Le nombre S de questions à réponse connue du candidat est donc

$$S = \sum_{i=1}^{40} \mathbf{1}_{A_i}.$$

L'indépendance des questions entraîne celle des évènements A_i. Comme ils ont chacun la même probabilité p, S suit la loi binomiale :

$$S \sim \text{Bin}(40; p).$$

2) On note B_i l'évènement « le candidat donne la bonne réponse à la i^{e} question ». On a clairement

$$P(B_i \mid A_i) = 1 \quad \text{et} \quad P(B_i \mid A_i^c) = \frac{1}{4}.$$

On calcule alors $P(B_i)$ en conditionnant par les deux cas possibles :

$$P(B_i) = P(B_i \mid A_i)P(A_i) + P(B_i \mid A_i^c)P(A_i^c) = 1 \times p + \frac{1}{4}(1-p) = \frac{1+3p}{4}.$$

3) Introduisons à nouveau les variables aléatoires indicatrices :

$$\mathbf{1}_{B_i} = \begin{cases} 1 & \text{si le candidat donne la bonne réponse à la } i^{\text{e}} \text{ question ;} \\ 0 & \text{si le candidat donne une réponse fausse à la } i^{\text{e}} \text{ question.} \end{cases}$$

Le nombre total U de bonnes réponses données par le candidat peut s'écrire :

$$U = \sum_{i=1}^{40} \mathbf{1}_{B_i}.$$

Les B_i sont indépendants et de même probabilité $(1 + 3p)/4$, donc U suit la loi binomiale :

$$U \sim \text{Bin}\Big(40; \frac{1+3p}{4}\Big).$$

De même le nombre total V de réponses fausses données par le candidat est

$$V = \sum_{i=1}^{40} \mathbf{1}_{B_i^c}.$$

Les B_i^c étant indépendants et de même probabilité $1 - (1 + 3p)/4 = 3(1 - p)/4$,

$$V \sim \mathrm{Bin}\Big(40; \frac{3 - 3p}{4}\Big).$$

4) Notons C_i l'évènement « le candidat répond au hasard à la i^{e} question *et* trouve la bonne réponse ». Clairement pour tout $i \in [\![1, 40]\!]$, $B_i = A_i \cup C_i$ avec $A_i \cap C_i = \varnothing$. On en déduit $P(B_i) = P(A_i) + P(C_i)$, d'où

$$P(C_i) = P(B_i) - P(A_i) = \frac{1 + 3p}{4} - p = \frac{1 - p}{4}.$$

Le nombre T de bonnes réponses « chanceuses » est

$$T = \sum_{i=1}^{40} \mathbf{1}_{C_i}.$$

Les C_i étant indépendants et de même probabilité $(1 - p)/4$,

$$T \sim \mathrm{Bin}\Big(40; \frac{1 - p}{4}\Big).$$

Solution 6.6 *Épreuves répétées à trois issues*

1) L'événement E consistant en l'obtention dans cet ordre de deux A suivis d'un B et de deux C s'écrit :

$$E = A_1 \cap A_2 \cap B_3 \cap C_4 \cap C_5.$$

En raison de l'indépendance des épreuves, les cinq événements intervenant dans cette intersection sont *mutuellement indépendants*. D'où :

$$P(E) = P(A_1)P(A_2)P(B_3)P(C_4)P(C_5) = a^2bc^2.$$

L'événement F obtention (sans condition d'ordre) de deux A, un B et deux C est la réunion disjointe de tous les sous-événements du type E construits en imposant parmi les 5 épreuves les rangs des deux A et celui du B parmi les trois rangs restants (les deux C occupant alors les deux derniers rangs non utilisés). Chacun de ces sous-événements a même probabilité que E et il y a en tout $C_5^2 \times C_3^1$ dispositions possibles. Ainsi

$$P(F) = C_5^2 C_3^1 a^2bc^2 = \frac{5!}{2!\,3!}\frac{3!}{1!\,2!}a^2bc^2 = \frac{5!}{2!\,1!\,2!}a^2bc^2 = 30a^2bc^2.$$

2) Généralisons. Pour $i+j+k=n$, notons $E_{i,j,k}$ l'évènement obtention dans cet ordre de i résultats A, j résultats B et k résultats C :

$$E_{i,j,k} = \left(\bigcap_{\ell=1}^{i} A_\ell\right) \cap \left(\bigcap_{\ell=i+1}^{i+j} B_\ell\right) \cap \left(\bigcap_{\ell=i+j+1}^{n} C_\ell\right).$$

Par indépendance des épreuves,

$$P(E_{i,j,k}) = a^i b^j c^k.$$

Dans ces écritures, nous avons fait la convention habituelle selon laquelle une intersection indexée par une condition impossible à réaliser (par exemple $i+1 \leqslant \ell \leqslant i+j$ lorsque $j=0$) vaut Ω. Le calcul ci-dessus inclut donc les cas particuliers d'indices i, j ou k nuls. Si l'on supprime la condition d'ordre, on obtient l'évènement $F_{i,j,k}$, réunion disjointe d'événements de même probabilité que $E_{i,j,k}$. Leur dénombrement s'effectue comme ci-dessus, il y en a $C_n^i C_{n-i}^j$: en effet, il y a C_n^i façons de placer les i A, puis C_{n-i}^j façons de placer les j B parmi les $n-i$ positions restant libres, après quoi il ne reste plus que $k=n-i-j$ positions pour les k C, donc une seule façon de les placer. La probabilité d'obtenir au cours des n épreuves (et sans condition d'ordre) i A, j B et k C ($i+j+k=n$) vaut donc :

$$P(F_{i,j,k}) = C_n^i C_{n-j}^j a^i b^j c^k = \frac{n!}{i!\,(n-i)!}\frac{(n-i)!}{j!\,(n-i-j)!}a^i b^j c^k = \frac{n!}{i!\,j!\,k!}a^i b^j c^k.$$

Solution 6.7 *Code de la Route III*

1) Notons $E = \{S=i, U=m\}$ et $F = \{S=i, T=m-i, V=40-m\}$. Nous allons vérifier que $E=F$.

Soit $\omega \in E$. Cela signifie que $S(\omega)=i$ et $U(\omega)=m$. Comme $T=U-S$ et $V=40-U$, on a $T(\omega)=m-i$ et $V(\omega)=40-m$. Donc ω vérifie les trois conditions d'appartenance à F. Donc tout $\omega \in E$ est aussi élément de F (autrement dit, $E \subset F$).

Réciproquement, soit $\omega \in F$. Alors $S(\omega)=i$, $T(\omega)=m-i$ et $V(\omega)=40-m$. Comme $U=T+S$, $U(\omega)=T(\omega)+S(\omega)=(m-i)+i=m$. Ainsi $S(\omega)=i$ et $U(\omega)=m$, donc ω est dans E. On a vérifié que tout ω de F est aussi dans E, autrement dit que $F \subset E$.

Les inclusions $E \subset F$ *et* $F \subset E$ montrent que $E=F$, d'où l'égalité des probabilités :

$$P(S=i, U=m) = P(S=i, T=m-i, V=40-m), \quad i \leqslant m \leqslant 40.$$

2) On peut alors utiliser cette égalité pour le calcul de $P(S=i \mid U=m)$:

$$P(S=i \mid U=m) = \frac{P(S=i, U=m)}{P(U=m)} = \frac{P(S=i, T=m-i, V=40-m)}{P(U=m)}.$$

La loi de U étant connue, il suffit de calculer $P(S=i, T=m-i, V=40-m)$. Pour ce faire, on utilise le résultat de l'exercice 6.6 en considérant que l'on a une série de 40 épreuves répétées indépendantes avec pour chacune, trois résultats possibles :

A réponse connue du candidat avec probabilité $a = p$;
B bonne réponse chanceuse avec probabilité $b = (1-p)/4$;
C réponse fausse avec probabilité $c = (3-3p)/4$.
Dans ces conditions, la probabilité d'avoir i réponses connues, $m-i$ bonnes réponses chanceuses et $40-m$ réponses fausses est

$$P(S=i, T=m-i, V=40-m) = \frac{40!}{i!(m-i)!(40-m)!} p^i \Big(\frac{1-p}{4}\Big)^{m-i} \Big(\frac{3-3p}{4}\Big)^{40-m}.$$

Comme U suit la loi $\mathrm{Bin}\Big(40; \frac{1+3p}{4}\Big)$,

$$P(U=m) = \frac{40!}{m!(40-m)!} \Big(\frac{1+3p}{4}\Big)^{m} \Big(\frac{3-3p}{4}\Big)^{40-m}.$$

En reportant dans le calcul de $P(S=i \mid U=m)$, il vient après simplifications

$$\begin{aligned} P(S=i \mid U=m) &= \frac{m!}{i!(m-i)!} p^i \Big(\frac{1-p}{4}\Big)^{m-i} \Big(\frac{4}{1+3p}\Big)^{m} \\ &= C_m^i \Big(\frac{4p}{1+3p}\Big)^{i} \Big(\frac{1-p}{1+3p}\Big)^{m-i}. \end{aligned}$$

Rappelons que ce résultat est valide pour $0 \leqslant i \leqslant m \leqslant 40$. On remarque que

$$0 \leqslant \frac{4p}{1+3p} \leqslant 1 \quad \text{et} \quad \frac{4p}{1+3p} + \frac{1-p}{1+3p} = 1.$$

On peut donc énoncer le résultat sous la forme suivante : *conditionnellement à l'évènement* $\{U=m\}$, S suit la loi binomiale de paramètres m et $4p/(1+3p)$.

3) En remarquant que si $i > m$, $P(S=i \mid U=m) = 0$ (ceci revient à considérer que le candidat n'est pas distrait et ne peut donner moins de bonnes réponses qu'il n'en connaît !), on en déduit

$$P(S \geqslant 36 \mid U=38) = P(S=36 \mid U=38) + P(S=37 \mid U=38) + P(S=38 \mid U=38).$$

Posons pour alléger les écritures $r = 4p/(1+3p)$.

$$\begin{aligned} P(S \geqslant 36 \mid U=38) &= C_{38}^{36} r^{36}(1-r)^2 + C_{38}^{37} r^{37}(1-r) + C_{38}^{38} r^{38} \\ &= r^{36}\big(703(1-r)^2 + 38r(1-r) + r^2\big) \\ &= r^{36}\big(666r^2 - 1368r + 703\big). \end{aligned}$$

En remplaçant r par sa valeur, on aboutit après calcul à

$$P(S \geqslant 36 \mid U=38) = \frac{(4p)^{36}}{(1+3p)^{38}} (567p^2 - 1254p + 703)$$

Solution 6.8 *Loi triangulaire*

1) La fonction f est bien une densité de probabilité sur $\mathbb{R}$. En effet, elle est positive, continue à support compact, donc son intégrale généralisée sur $\mathbb{R}$ se réduit à une intégrale de Riemann ordinaire sur $[0,2]$ donc est finie. Cette intégrale vaut clairement 1 puisque c'est l'aire d'un triangle de base 2 et de hauteur 1.

2) Soit Z une variable aléatoire positive de densité f et F sa fonction de répartition. Les valeurs de F demandées sont regroupées dans le tableau suivant :

x	0	$\frac{1}{2}$	1	2
$F(x)$	0	$\frac{1}{8}$	$\frac{1}{2}$	1

On les obtient par un simple calcul d'aire puisque $F(x)$ est l'aire du domaine délimité par l'axe des abscisses, la droite verticale d'abscisse x et le graphe de F. Ainsi $F(1/2)$ est l'aire d'un triangle de base $1/2$ et de hauteur $1/2$, $F(1)$ est l'aire d'un triangle de base 1 et de hauteur 1, $F(2)$ a déjà été calculé à la question précédente.

3) Calculons $\mathbf{E}Z$ en utilisant la densité f. Cette fonction est affine par morceaux :

$$f(t) = \begin{cases} 0 & \text{si } t < 0, \\ t & \text{si } 0 \leqslant t < 1, \\ 2-t & \text{si } 1 \leqslant t < 2, \\ 0 & \text{si } t \geqslant 2. \end{cases}$$

Comme Z est presque-sûrement bornée ($P(0 \leqslant Z \leqslant 2) = 1$), elle est évidemment intégrable, donc $\mathbf{E}Z$ existe et peut se calculer par :

$$\mathbf{E}Z = \int_{-\infty}^{+\infty} t f(t)\,\mathrm{d}t = \int_0^2 t f(t)\,\mathrm{d}t = \int_0^1 t^2\,\mathrm{d}t + \int_1^2 t(2-t)\,\mathrm{d}t.$$

On obtient ainsi [7]

$$\mathbf{E}Z = \left[\frac{t^3}{3}\right]_0^1 + \left[t^2 - \frac{t^3}{3}\right]_1^2 = \frac{1}{3} + 4 - \frac{8}{3} - 1 + \frac{1}{3} = 3 - \frac{6}{3} = 1.$$

Ce résultat était intuitivement évident avant tout calcul, en interprétant $\mathbf{E}Z$ comme le *barycentre* de la répartition de masse de densité f sur le segment $[0,2]$ et en remarquant que cette répartition est symétrique par rapport à 1 puisque le graphe de f admet la droite d'équation $t = 1$ comme axe de symétrie.

Solution 6.9 *Une variable aléatoire tronquée*

Sur l'espace probabilisé $(\Omega, \mathcal{F}, P)$, X est une variable aléatoire de loi exponentielle. Soient a et b deux réels fixés $0 < a < b$. On définit sur le même espace $(\Omega, \mathcal{F}, P)$, la

7. Il serait plus astucieux de poser $t = 1+u$ dans la dernière intégrale ce qui donnerait $\int_0^1 (1-u^2)\,\mathrm{d}u$, mais comme l'auteur du sujet recherchait un effet de contraste entre la méthode « bourrin » donnée ci-dessus et l'interprétation barycentrique, je ne vais pas le contrarier (Ch. S.).

variable aléatoire Y par

$$\forall \omega \in \Omega, \quad Y(\omega) := \begin{cases} 0 & \text{si } X(\omega) \leqslant a, \\ X(\omega) & \text{si } a < X(\omega) \leqslant b, \\ b & \text{si } X(\omega) > b. \end{cases}$$

1) L'allure de la représentation graphique de la fonction de répartition de Y est donnée par la figure 11.5.

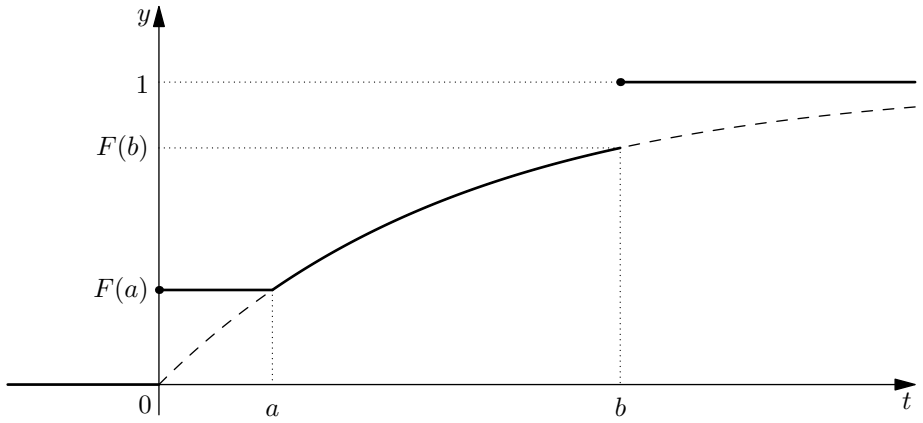


FIGURE 11.5 – Les f.d.r. de X (tirets) et de Y (trait épais)

Voici les justifications. Il s'agit de calculer $H(t) := P(Y \leqslant t)$ pour $t \in \mathbb{R}$. On notera F la fonction de répartition de X.

Première façon, partant de l'intuition. On pressent que les valeurs de t qui vont jouer un rôle important sont 0, a et b. On va donc essayer de calculer H sur chacun des 4 intervalles découpés sur la droite réelle par ces trois nombres.

1. Pour $t \in]-\infty, 0[$, l'évènement $\{Y \leqslant t\}$ est inclus dans $\{Y < 0\}$ lequel se réduit à l'ensemble vide puisque par construction, Y ne peut prendre que des valeurs positives ou nulles. Donc $P(Y \leqslant t) \leqslant P(\varnothing) = 0$ et pour $t \in]-\infty, 0]$, $H(t) = 0$.
2. Calculons $H(0)$. D'après ce qui précède, $P(Y \leqslant 0) = P(Y < 0) + P(Y = 0) = P(Y = 0)$. La définition de Y nous montre que $Y(\omega) = 0$ si et seulement si $X(\omega) \leqslant a$. On a donc l'égalité d'évènements $\{Y = 0\} = \{X \leqslant a\}$, d'où $H(0) = P(Y = 0) = P(X \leqslant a) = F(a)$.
3. Pour $t \in]0, a[$, l'évènement $\{Y \leqslant t\}$ se réduit à l'évènement $\{Y = 0\}$ puisque Y ne peut prendre que soit la valeur 0 soit une valeur supérieure ou égale à a. Donc $H(t) = P(Y = 0) = H(0) = F(a)$. Ainsi la fonction H est constante, égale à $F(a)$, sur l'intervalle $[0, a[$.
4. Pour $t \in [a, b[$, l'évènement $\{Y \leqslant t\}$ est la réunion des deux évènements disjoints $\{Y = 0\}$ et $\{a \leqslant Y \leqslant t\}$. D'après la définition de Y, le deuxième s'écrit aussi $\{a \leqslant$

$X \leqslant t\}$. On a donc $H(t) = P(Y = 0)+P(a \leqslant X \leqslant t) = F(a)+F(t)-F(a-) = F(t)$ car F étant continue sur $\mathbb{R}$ donc au point a, $F(a-) = F(a)$. Ainsi H coïncide avec F sur $[a, b[$.

5. Calculons $H(b)$. La définition de Y nous montre que pour tout $\omega \in \Omega$, $Y(\omega) \leqslant b$. Donc $\{Y \leqslant b\} = \Omega$ et $H(b) = P(Y \leqslant b) = 1$. On a ainsi une discontinuité de H au point b car la limite à gauche de H en b est $F(b)$ par le cas précédent et la continuité de F au point b. Or $F(b) < 1$ car, X suivant une loi exponentielle, $P(X > b) > 0$.
6. Pour $t > b$, on a $1 = H(b) \leqslant H(t) \leqslant 1$, donc $H(t) = 1$.

Deuxième façon, plus automatique. On décompose l'évènement $\{Y \leqslant t\}$ sur la partition de Ω constituée des trois évènements $\{X \leqslant a\}$, $\{a < X \leqslant b\}$ et $\{X > b\}$. On obtient ainsi l'union disjointe $\{Y \leqslant t\} = E_1 \cup E_2 \cup E_3$, avec $E_1 = \{Y \leqslant t\} \cap \{X \leqslant a\}$, $E_2 = \{Y \leqslant t\} \cap \{a < X \leqslant b\}$ et $E_3 = \{Y \leqslant t\} \cap \{X > b\}$. Pour simplifier les écritures, on se limite au cas où $t \geqslant 0$, puisque, Y étant une v.a. positive, on a clairement $H(t) = 0$ pour tout $t < 0$. On peut alors vérifier les égalités suivantes (faites-le) :

$$E_1 = \{X \leqslant a\}, \qquad E_2 = \begin{cases} \varnothing & \text{si } t \leqslant a \\ \{a < X \leqslant t\} & \text{si } a < t \leqslant b \\ \{a < X \leqslant b\} & \text{si } t > b \end{cases} \qquad E_3 = \begin{cases} \varnothing & \text{si } t < b \\ \{X > b\} & \text{si } t \geqslant b \end{cases}$$

En recollant soigneusement les morceaux, on retrouve les résultats de la 1re méthode.

2) La fonction de répartition H de la variable aléatoire Y n'est pas continue sur $\mathbb{R}$ (discontinuités en 0 et en b), la loi de Y ne peut donc avoir de densité. Rappelons que la continuité de la fonction de répartition est *nécessaire* (mais *pas suffisante*) pour que la loi ait une densité.

Solution 6.11 *Un peu de trigonométrie aléatoire*

On définit une variable aléatoire X grâce à la construction représentée à la figure 11.6.

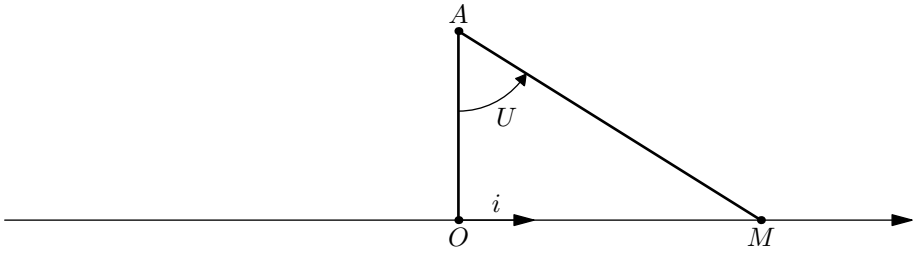


FIGURE 11.6 – Construction de M

L'angle $\left(\overrightarrow{AO}, \overrightarrow{AM}\right)$ a pour mesure en radians U, *variable aléatoire* de loi uniforme sur $]-\pi/2, \pi/2[$. La distance AO vaut 1 et X est l'abscisse du point M sur la droite de repère $(O, \vec{i}\,)$: $\overrightarrow{OM} = X\vec{i}$. Les angles sont orientés dans le sens trigonométrique,

la figure est faite pour U positif. Pour $U = 0$, M coïncide avec le point O et pour $-\pi/2 < U < 0$, M est « à gauche » de O.

1) En utilisant le triangle rectangle AOM et en tenant compte de l'orientation, on obtient :

$$X = \tan U.$$

En effet si U est positif, l'angle aigu non orienté OAM a pour mesure U et $OM = X$. Or $\tan(OAM) = \frac{OM}{OA}$. Si U est négatif, OAM a pour mesure $-U = |U|$, $X = -OM$ donc la relation $\tan(OAM) = \frac{OM}{OA}$ donne $\tan(-U) = -X$ d'où $\tan U = X$ car la fonction tangente est impaire.

Une autre façon de vérifier la formule $X = \tan U$ est de tracer le cercle trigonométrique de centre A et de rayon 1, en prenant l'origine des arcs au point O et d'utiliser l'interprétation géométrique de la fonction tangente. Pour retrouver la figure « habituelle » (?), il suffit de tourner la feuille de $\pi/2$ dans le sens trigonométrique...

2) En notant F la fonction de répartition de la variable aléatoire réelle X, on a pour tout x réel,

$$F(x) = P(X \leqslant x) = P(\tan U \leqslant x) = P(U \leqslant \arctan x).$$

La dernière égalité se justifie comme suit. Rappelons d'abord que la restriction de la fonction tangente à l'intervalle $]-\pi/2, \pi/2[$ est continue strictement croissante avec pour limite à droite $-\infty$ en $-\pi/2$ et pour limite à gauche $+\infty$ en $\pi/2$. Cette restriction réalise donc une bijection [8] de $]-\pi/2, \pi/2[$ sur $\mathbb{R}$. La fonction arctan est la réciproque de cette bijection. En particulier elle est strictement croissante sur $\mathbb{R}$ et pour tout $u \in]-\pi/2, \pi/2[$, $\arctan(\tan u) = u$, pour tout $x \in \mathbb{R}$, $\tan(\arctan x) = x$. Ceci nous permet de vérifier l'égalité d'évènements $\{\tan U \leqslant x\} = \{U \leqslant \arctan x\}$ et donc de leur probabilité. En effet on a les deux implications :

$\tan U \leqslant x \Rightarrow U \leqslant \arctan x$, par croissance d'arctangente et $\arctan(\tan U) = U$,

$U \leqslant \arctan x \Rightarrow \tan U \leqslant x$, par croissance de tangente et $\tan(\arctan x) = x$.

Ceci justifie l'équivalence $(\tan U \leqslant x) \Leftrightarrow (U \leqslant \arctan x)$ et donc l'égalité des évènements correspondants.

D'autre part, U suivant la loi uniforme sur $]-\pi/2, \pi/2[$, on a en notant λ la mesure de Lebesgue sur $\mathbb{R}$ (donc la longueur pour un segment) :

$$\begin{aligned} \forall y \in \left]\frac{-\pi}{2}, \frac{\pi}{2}\right[, \quad P(U \leqslant y) = P(U \in]-\infty, y]) &= \frac{\lambda(]-\infty, y] \cap]-\pi/2, \pi/2[)}{\lambda(]-\pi/2, \pi/2[)} \\ &= \frac{\lambda(]-\pi/2, y[)}{\pi} \\ &= \frac{y - (-\pi/2)}{\pi} = \frac{1}{2} + \frac{y}{\pi}. \end{aligned}$$

8. L'injectivité résulte de la croissance stricte et la surjectivité vient de la continuité et des limites aux bornes $\pm\pi/2$ *via* le théorème des valeurs intermédiaires.

En appliquant ceci à $y = \arctan x$, on aboutit à

$$\forall x \in \mathbb{R}, \quad F(x) = P(U \leqslant \arctan x) = \frac{1}{2} + \frac{1}{\pi} \arctan x.$$

On vérifie facilement qu'il s'agit bien d'une fonction de répartition : elle est croissante et continue sur $\mathbb{R}$ (propriétés héritées d'arctangente par composition) ; sa limite en $-\infty$ est 0, car la limite en $-\infty$ de l'arctangente est $-\pi/2$, de même sa limite en $+\infty$ est 1 car celle de l'arctangente est $\pi/2$.

3) La fonction F est de classe C^1 sur $\mathbb{R}$ comme l'arctangente. La loi de X admet donc une densité f qui s'obtient par dérivation de F :

$$\forall t \in \mathbb{R}, \quad f(t) = F'(t) = \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{1}{2} + \frac{1}{\pi}\arctan t\right) = \frac{1}{\pi}\frac{\mathrm{d}}{\mathrm{d}t}(\arctan t) = \frac{1}{\pi(1+t^2)}.$$

On reconnaît la densité de la loi de Cauchy.

4) Pour calculer $P(|X| \leqslant 1)$, le plus simple est de remarquer que dans le triangle rectangle AOM, $OM \leqslant OA = 1$ si et seulement si l'angle aigu (non orienté) de sommet A a une mesure d'au plus $\pi/4$, c'est-à-dire si et seulement si $|U| \leqslant \pi/4$. D'où

$$P(|X| \leqslant 1) = P(U \in [-\pi/4, \pi/4]) = \frac{\lambda([-\pi/4, \pi/4])}{\lambda(]-\pi/2, \pi/2[)} = \frac{1}{2}.$$

Cette méthode avait l'avantage de ne pas utiliser les questions précédentes.

On pouvait retrouver ce résultat à partir de la fonction de répartition F en notant que :

$$\begin{aligned} P(|X| \leqslant 1) = P(X \in [-1, 1]) = F(1) - F(-1-) &= F(1) - F(-1) \\ &= \frac{\arctan(1) - \arctan(-1)}{\pi} \\ &= \frac{\pi/4 - (-\pi/4)}{\pi} = \frac{1}{2}. \end{aligned}$$

On a utilisé la continuité de F pour remplacer la limite à gauche $F(-1-)$ par $F(-1)$.

Solution 6.12

1) Il est clair que f est positive. Pour que ce soit une densité, il suffit donc de vérifier que son intégrale sur $\mathbb{R}$ vaut 1 :

$$\int_{-\infty}^{+\infty} f(t)\,\mathrm{d}t = \int_0^1 \frac{1}{(1+t)\ln 2}\,\mathrm{d}t = \left[\frac{\ln(1+t)}{\ln 2}\right]_0^1 = \frac{\ln 2 - \ln 1}{\ln 2} = 1.$$

La fonction de répartition s'obtient en intégrant la densité :

$$F(x) = \int_{-\infty}^{x} f(t)\,\mathrm{d}t = \int_{-\infty}^{x} \frac{1}{(1+t)\ln 2}\mathbf{1}_{]0,1[}(t)\,\mathrm{d}t.$$

Pour $x < 0$, cette intégrale vaut 0, pour $x \geqslant 1$, elle vaut $\int_0^1 f(t)\,\mathrm{d}t$, c'est à dire 1, pour $0 \leqslant x \leqslant 1$, elle vaut :

$$\int_0^x f(t)\,\mathrm{d}t = \left[\frac{\ln(1+t)}{\ln 2}\right]_0^x = \frac{\ln(1+x)}{\ln 2}.$$

En rassemblant ces trois cas en une seule formule :

$$\forall x \in \mathbb{R}, \quad F(x) = \frac{\ln(1+x)}{\ln 2}\mathbf{1}_{[0,1[}(x) + \mathbf{1}_{[1,+\infty[}(x).$$

2) La fonction $x \mapsto 1/x$ étant décroissante sur $]0,+\infty[$, on a :

$$\forall k \in \mathbb{N}^*,\ \forall t \in]0,1[, \quad k \leqslant \frac{1}{X} \leqslant k+t \Longleftrightarrow \frac{1}{k+t} \leqslant X \leqslant \frac{1}{k}.$$

Ainsi les deux évènements définis par les deux membres de cette équivalence sont confondus, d'où :

$$P\left(k \leqslant \frac{1}{X} \leqslant k+t\right) = P\left(\frac{1}{k+t} \leqslant X \leqslant \frac{1}{k}\right) = P\left(X \leqslant \frac{1}{k}\right) - P\left(X < \frac{1}{k+t}\right).$$

La variable X étant à densité, on a pour tout $x \in \mathbb{R}$, $P(X = x) = 0$ d'où $P(X < x) = P(X \leqslant x)$. En appliquant ceci à $x = 1/(k+t)$, on obtient :

$$\begin{aligned} P\left(X \leqslant \frac{1}{k}\right) - P\left(X < \frac{1}{k+t}\right) &= F\left(\frac{1}{k}\right) - F\left(\frac{1}{k+t}\right) \\ &= \frac{1}{\ln 2}\left(\ln\left(\frac{k+1}{k}\right) - \ln\left(\frac{k+1+t}{k+t}\right)\right). \end{aligned}$$

Finalement, on a pour tout $k \in \mathbb{N}^*$ et tout $t \in]0,1[$:

$$P\left(k \leqslant \frac{1}{X} \leqslant k+t\right) = \frac{1}{\ln 2}\big(\ln(k+1) - \ln k + \ln(k+t) - \ln(k+1+t)\big) \qquad (11.12)$$

3) Soit $Y = \frac{1}{X} - \left[\frac{1}{X}\right]$. Cherchons sa fonction de répartition G définie par $G(t) = P(Y \leqslant t)$. D'abord, il est clair qu'avec probabilité 1, $0 \leqslant Y \leqslant 1$. Donc $G(t) = 0$ pour tout $t < 0$ et $G(t) = 1$ pour tout $t \geqslant 1$. Ainsi $F(t) = G(t)$ dans ces deux cas. Calculons $G(t)$ pour $0 < t < 1$. On note pour cela que :

$$\begin{aligned} \{Y \leqslant t\} &= \left\{\frac{1}{X} - \left[\frac{1}{X}\right] \leqslant t\right\} \\ &= \bigcup_{k\in\mathbb{N}^*}\left(\left\{\left[\frac{1}{X}\right] = k\right\} \cap \left\{\frac{1}{X} - k \leqslant t\right\}\right) \\ &= \bigcup_{k\in\mathbb{N}^*}\left(\left\{k \leqslant \frac{1}{X} < k+1\right\} \cap \left\{\frac{1}{X} \leqslant k+t\right\}\right) \\ &= \bigcup_{k\in\mathbb{N}^*}\left\{k \leqslant \frac{1}{X} \leqslant k+t\right\}. \end{aligned}$$

On a ainsi décomposé l'évènement $\{Y \leqslant t\}$ en une réunion indexée par une famille dénombrable ($\mathbb{N}^*$) d'évènements disjoints. Par conséquent,

$$P(Y \leqslant t) = \sum_{k=1}^{+\infty} P\left(k \leqslant \frac{1}{X} \leqslant k+t\right).$$

Pour calculer cette série, on utilise (11.12) en revenant aux sommes partielles et en remarquant que chaque terme se simplifie avec son successeur :

$$\begin{aligned}\sum_{k=1}^{m} P\left(k \leqslant \frac{1}{X} \leqslant k+t\right) &= \frac{1}{\ln 2}\left(\sum_{k=1}^{m}\big(\ln(k+1)-\ln k\big) + \sum_{k=1}^{m}\big(\ln(k+t)-\ln(k+1+t)\big)\right)\\ &= \frac{1}{\ln 2}\big(\ln(m+1)+\ln(1+t)-\ln(m+1+t)\big).\end{aligned}$$

Pour trouver la limite quand m tend vers $+\infty$ de cette expression, on écrit :

$$\ln(m+1+t) = \ln(m+1) + \ln\left(1+\frac{t}{m+1}\right),$$

d'où l'on déduit :

$$\lim_{m\to+\infty} \frac{1}{\ln 2}\big(\ln(m+1)+\ln(1+t)-\ln(m+1+t)\big) = \frac{\ln(1+t)}{\ln 2}.$$

Ainsi les fonctions F et G sont égales et Y a même loi que X.

Solution 6.17 *Une loi discrète pathologique*

Le but de cet exercice est de construire et d'étudier une v.a. discrète ayant pour ensemble $X(\Omega)$ de valeurs possibles l'ensemble $\mathbb{D}$ des nombres *décimaux* de $[0,1[$. Un nombre décimal peut se représenter par une suite *finie* de *chiffres décimaux*. On peut aussi l'écrire sous la forme $k10^{-n}$ avec $k \in \mathbb{N}$ et $n \in \mathbb{N}^*$. Cette représentation n'est pas unique, elle est appelée forme réduite si k n'est pas divisible par 10. On note $\mathbb{D}_n$ l'ensemble des décimaux de $[0,1[$ ayant le niveau de résolution n, c'est-à-dire de forme réduite $k10^{-n}$ (on pose $\mathbb{D}_0 = \{0\}$).

Nous construisons X par la procédure suivante. On dispose de deux urnes. La première contient des boules rouges en proportion $0 < p < 1$ et des vertes en proportion q. La deuxième urne contient 10 boules numérotées de 0 à 9. On effectue des tirages avec remise dans la première urne jusqu'à la première apparition d'une rouge. On note N le nombre (aléatoire) de tirages nécessaires. Une fois connue la valeur de N, on effectue N tirages avec remise d'une boule dans la deuxième urne. En notant Y_j le *chiffre* sorti lors du j^{e} tirage dans la deuxième urne ($i \leqslant N$), on forme le nombre décimal :

$$X(\omega) = 0, Y_1(\omega)Y_2(\omega)\dots Y_{N(\omega)}(\omega) = \sum_{j=1}^{N(\omega)} \frac{Y_j(\omega)}{10^j}.$$

1) N est le *temps d'attente* du premier succès dans une suite d'épreuves répétées indépendantes [9] avec à chaque épreuve même probabilité de succès p. On sait que N suit la loi géométrique de paramètre p qui est caractérisée par :

$$N(\Omega) = \mathbb{N}^*, \qquad \forall k \in \mathbb{N}^*,\ P(N = k) = q^{k-1}p.$$

2) Soit n fixé. Lorsque l'on effectue n tirages *avec remise* dans la deuxième urne, on considère que toutes les n-suites possibles de résultats (il y en 10^n) ont même probabilité de sortie. On modélise ainsi les tirages avec remise par $\Omega_n = [\![0, 9]\!]^n$, muni de la tribu $\mathcal{F}_n = \mathcal{P}(\Omega_n)$ et de la probabilité uniforme définie par $P_n(\{\omega\}) = 1/\operatorname{card}\Omega$. Avec ce modèle, la probabilité d'obtenir une suite $\omega = (c_1, \dots, c_n)$ de n chiffres décimaux choisie à l'avance est donc 10^{-n}.

Tout espace probabilisé $(\Omega, \mathcal{F}, P)$ modélisant correctement l'expérience étudiée doit donner pour $P(Y_1 = c_1, Y_2 = c_2, \dots, Y_n = c_n \mid N = n)$ la même valeur que lorsqu'on fait dès le départ n tirages avec remise, autrement dit

$$P(Y_1 = c_1, Y_2 = c_2, \dots, Y_n = c_n \mid N = n) = P_n(\{\omega\}) = 10^{-n}, \quad \omega = (c_1, \dots, c_n) \in \Omega_n.$$

Ce raisonnement de type « cahier des charges » est standard en théorie des probabilités. Il est assez rare que l'on calcule $P(A \mid B)$ par sa définition. On lui *attribue* directement une valeur en analysant l'expérience (lorsque la connaissance de la réalisation de B simplifie cette expérience) et on se sert de cette valeur pour en déduire celle que devrait avoir $P(A \cap B)$ dans tout modèle « raisonnable ».

3) Il y a une infinité de façons d'écrire 0,375 sous la forme d'un développement décimal *fini*. La forme générale d'un tel développement est

$$0,375\ \underbrace{0\dots\dots0}_{(n-3)\text{ chiffres}}\ , \qquad n \geqslant 3.$$

Il y a aussi un autre développement décimal de $0,375$, mais celui-la est *illimité*, c'est $0{,}374\,999\,999\,999\,999\,9\ldots9\ldots$. Il est exclu que l'on puisse obtenir un tel développement avec l'expérience envisagée car comme $p > 0$, $P(N < +\infty) = 1$, voir p. 224.

Avant de se lancer dans le calcul, remarquons que $\{X = 0{,}375\}$ implique $\{N \geqslant 3\}$, ce qui s'écrit aussi $\{X = 0{,}375\} \subset \{N \geqslant 3\}$. On en déduit

$$\{X = 0{,}375\} = \{X = 0{,}375\} \cap \{N \geqslant 3\} = \{X = 0{,}375\} \cap \Big(\bigcup_{k \geqslant 3} \{N = k\}\Big).$$

Ceci nous donne une décomposition en réunion dénombrable d'évènements deux à deux incompatibles :

$$\{X = 0{,}375\} = \bigcup_{j \geqslant 3} \Big(\{X = 0{,}375\} \cap \{N = j\}\Big).$$

9. Il est légitime de considérer les épreuves comme indépendantes parce que les tirages se font *avec remise*.

En utilisant la σ-additivité de P, on en déduit

$$\begin{aligned} P(X = 0{,}375) &= \sum_{j=3}^{+\infty} P\Big(\{X = 0{,}375\} \cap \{N = j\}\Big) \\ &= \sum_{j=3}^{+\infty} P\Big(\{Y_1 = 3, Y_2 = 7, Y_3 = 5, Y_4 = 0, \ldots, Y_j = 0\} \cap \{N = j\}\Big). \end{aligned}$$

La formule $P(A \cap B) = P(A \mid B)P(B)$ (quand $P(B) > 0$) et les résultats des questions 1) et 2) nous donnent

$$\begin{aligned} &P\Big(\{Y_1 = 3, Y_2 = 7, Y_3 = 5, Y_4 = 0, \ldots, Y_j = 0\} \cap \{N = j\}\Big) = \\ &\quad P\Big(Y_1 = 3, Y_2 = 7, Y_3 = 5, Y_4 = 0, \ldots, Y_j = 0 \mid N = j\Big) P(N = j) = 10^{-j} q^{j-1} p. \end{aligned}$$

Finalement le calcul de $P(X = 0{,}375)$ se ramène à un calcul de série géométrique :

$$P(X = 0{,}375) = \sum_{j=3}^{+\infty} 10^{-j} q^{j-1} p = \frac{p}{q} \sum_{j=3}^{+\infty} \Big(\frac{q}{10}\Big)^j = \frac{p}{q} \Big(\frac{q}{10}\Big)^3 \frac{1}{1 - \frac{q}{10}} = \frac{pq^2}{100(10 - q)}.$$

4) Pour généraliser ce résultat, soit $d \in \mathbb{D}_n$ et $d = k10^{-n}$ sa forme réduite. Soit

$$k = u_1 10^{n-1} + u_2 10^{n-2} + \cdots + u_{n-1} 10 + u_n, \quad u_i \in [\![0, 9]\!],\ i \in [\![1, n]\!],$$

l'écriture en base 10 de l'entier k. Une adaptation immédiate de la décomposition utilisée pour le cas $d = 0{,}375$ nous conduit à

$$\begin{aligned} P(X = d) &= \sum_{j=n}^{+\infty} P\big(\forall i \in [\![1, n]\!],\ Y_i = u_i \text{ et } \forall i \in]\!]n, j]\!],\ Y_i = 0 \mid N = j\big) P(N = j) \\ &= \sum_{j=n}^{+\infty} 10^{-j} q^{j-1} p \\ &= \frac{p}{q} \Big(\frac{q}{10}\Big)^n \frac{1}{1 - \frac{q}{10}}. \end{aligned}$$

Après simplification, on obtient

$$P(X = d) = \frac{pq^{n-1}}{(10 - q)10^{n-1}} = \frac{p}{9 + p} \Big(\frac{q}{10}\Big)^{n-1}, \qquad d \in \mathbb{D}_n,\ n \geqslant 1. \tag{11.13}$$

Le cas particulier $d = 0$ se traite de manière analogue :

$$P(X = 0) = \sum_{j=1}^{+\infty} P\big(Y_i = 0\ (1 \leqslant i \leqslant j) \mid N = j\big) P(N = j) = \sum_{j=1}^{+\infty} \frac{q^{j-1} p}{10^j} = \frac{p}{9 + p}.$$

5) Pour vérifier que $\mathbb{D}$ est dénombrable, on pourrait se contenter de remarquer que $\mathbb{D}$ est union dénombrable des ensembles finis deux à deux disjoints $\mathbb{D}_n$. Pour construire une numérotation explicite des éléments de $\mathbb{D}$, une idée simple est de considérer successivement chaque $\mathbb{D}_n$ et de numéroter de gauche à droite les éléments de l'ensemble fini $\mathbb{D}_n$ avec les plus petits entiers non déjà utilisés à l'étape $n-1$. Notons $\varphi : \mathbb{D} \to \mathbb{N}$ cette numérotation. Voici le début de sa construction :

$\mathbb{D}_0$

d	0
$\varphi(d)$	0

$\mathbb{D}_1$

$10d$	1	2	3	4	5	6	7	8	9
$\varphi(d)$	1	2	3	4	5	6	7	8	9

$\mathbb{D}_2$

$100d$	1	2	…	8	9	11	12	…	19	21	…	88	89	91	…	99
$\varphi(d)$	10	11	…	17	18	19	20	…	27	28	…	89	90	91	…	99

Afin d'obtenir une formule pour le calcul de $\varphi(d)$ on dénombre $\mathbb{D}_n$. Il suffit pour cela de remarquer que les éléments de $\mathbb{D}_n$ sont les $k10^{-n}$ lorsque $0 \leqslant k < 10^n$, sauf ceux pour lesquels k est multiple de 10 :

$$\mathbb{D}_n = \Big\{ \frac{k}{10^n} \ ; \ 0 \leqslant k < 10^n \Big\} \backslash \Big\{ \frac{10l}{10^n} \ ; \ 0 \leqslant l < 10^{n-1} \Big\}. \tag{11.14}$$

On en déduit immédiatement que

$$\operatorname{card} \mathbb{D}_n = 10^n - 10^{n-1}.$$

De même on vérifie, soit par un raisonnement direct, soit par cette formule que

$$\operatorname{card}\Big(\bigcup_{j=0}^{n-1} \mathbb{D}_j \Big) = 1 + \sum_{j=1}^{n-1} (10^j - 10^{j-1}) = 10^{n-1}, \quad n \geqslant 1.$$

On en déduit (en prenant garde au fait que la numérotation commence à 0) que le plus grand élément de D_{n-1} et le plus petit élément de $\mathbb{D}_n$ ont pour numéros respectifs,

$$\varphi\Big(\frac{10^{n-1}-1}{10^{n-1}} \Big) = 10^{n-1} - 1, \quad \varphi\Big(\frac{1}{10^n} \Big) = 10^{n-1}.$$

Finalement, soit $d \in \mathbb{D}_n$, on peut l'écrire $d = k10^{-n}$ avec $k = 10l + r$, $1 \leqslant r \leqslant 9$. Dans cette division euclidienne de k par 10, la valeur du quotient l permet de compter tous les multiples de 10 inférieurs à k : il y en a $l+1$ (zéro compris). Le numéro de k dans la liste des entiers de 0 à k étant $k+1$, on obtient

$$\varphi(d) = (10^{n-1} - 1) + (k+1) - (l+1) = 10^{n-1} + k - l.$$

Notons $[x]$ la partie entière du réel x, unique entier m tel que $m \leqslant x < m+1$. Comme $l = [k/10]$ et $k = 10^n d$, on aboutit à

$$\varphi(d) = 10^n d + 10^{n-1} - \big[10^{n-1} d\big] - 1, \quad d \in \mathbb{D}_n, \ n \geqslant 1.$$

6) La bijection $\varphi : \mathbb{D} \to \mathbb{N}$ construite ci-dessus est croissante séparément sur chaque $\mathbb{D}_n$, mais pas globalement sur $\mathbb{D}$. Il en résulte que $\varphi^{-1} : \mathbb{N} \to \mathbb{D}$ n'est pas croissante.

Montrons qu'il n'existe aucune numérotation *croissante* des éléments de $X(\Omega) = \mathbb{D}$ (c.-à-d. vérifiant $X(\Omega) = \mathbb{D} = \{x_k \ ; \ k \in \mathbb{N}\}$ avec $\forall k \in \mathbb{N}$, $x_k < x_{k+1}$). Supposons que $\psi : \mathbb{N} \to \mathbb{D}$ soit une telle numérotation croissante ($\psi(k) = x_k$). D'abord, nécessairement, $\psi(0) = x_0 = 0$ car sinon $x_0 > 0$ et comme ψ est croissante, $\psi(k) > x_0 > 0$ pour tout $k \geqslant 1$ et zéro ne serait jamais numéroté par ψ. Ensuite par croissance de ψ, $x_1 > x_0 = 0$ et $\forall k \geqslant 2$, $x_k > x_1$. Comme ψ est surjective, il n'y a donc *aucun* élément de $\mathbb{D}$ dans $]0, x_1[$. Ceci contredit la densité de $\mathbb{D}$ dans $[0, 1]$.

7) *Calcul de* $\sum_{d\in\mathbb{D}_n} d$. Il suffit d'utiliser (11.14) et la formule donnant la somme des termes d'une suite arithmétique (nombre de termes multiplié par la moyenne du premier et du dernier) :

$$\begin{aligned}\sum_{d\in\mathbb{D}_n} d &= \sum_{k=1}^{10^n-1} \frac{k}{10^n} - \sum_{l=1}^{10^{n-1}-1} \frac{l}{10^{n-1}} \\ &= \frac{1}{10^n}\,\frac{(10^n-1)10^n}{2} - \frac{1}{10^{n-1}}\,\frac{(10^{n-1}-1)10^{n-1}}{2}.\end{aligned}$$

Après simplification, on obtient

$$\sum_{d\in\mathbb{D}_n} d = \frac{9}{2} \times 10^{n-1}, \quad n \geqslant 1. \tag{11.15}$$

8) Soit F la fonction de répartition de X : $F(x) = P(X \leqslant x)$. On sait que F est continue à droite et a une limite à gauche en tout point. De plus pour tout $a \in \mathbb{R}$,

$$F(a) - F(a-) = P(X = a) = P_X(\{a\}), \tag{11.16}$$

en notant $F(a-)$ la limite à gauche de F au point a, cf. la proposition 6.19. D'après (11.16), en tout point $d \in \mathbb{D}$, on a $F(d) > F(d-) \geqslant F(x)$ pour $x < d$. Comme tout intervalle ouvert de $[0, 1[$ contient un décimal, F n'est constante sur aucun intervalle ouvert de $[0, 1[$. Elle n'est donc pas en escaliers sur $[0, 1[$. Ceci contraste avec les f.d.r. des lois discrètes usuelles (pour lesquelles on a une numérotation croissante de $X(\Omega)$ et où F est constante sur $[x_k, x_{k+1}[$, donc en escaliers).

9) Toujours d'après (11.16), F est continue au point a si $P(X = a) = 0$, autrement dit si $a \notin \mathbb{D}$, et discontinue si $P(X = a) > 0$, donc si $a \in \mathbb{D}$. On a donc construit à partir d'une expérience très simple une fonction de répartition discontinue en tout décimal de $[0, 1[$ et continue en tout réel non décimal de $[0, 1[$ (et aussi en tout $x \notin [0, 1[$).

10) *Calcul de* $P(X \leqslant 0{,}375 \mid N = j)$.
Pour calculer cette probabilité conditionnelle, on se place dans la situation où l'évènement $\{N = j\}$ est *réalisé*. On effectue alors exactement j tirages avec remise dans la deuxième urne. Chaque suite de chiffres possible a alors même probabilité d'apparition 10^{-j}. Il revient au même de dire que le nombre décimal X ainsi généré suit la

loi uniforme sur $\Delta_j = \{k10^{-j} \ ; \ 0 \leqslant k < 10^j\}$, ou encore que la variable aléatoire à valeurs *entières* $10^j X$ suit la loi uniforme sur $[\![0, 10^j[\![$. On en déduit que

$$\begin{aligned} P(X \leqslant 0{,}375 \mid N = j) &= P(10^j X \leqslant 0{,}375 \times 10^j \mid N = j) \\ &= P(10^j X \leqslant \lfloor 0{,}375 \times 10^j \rfloor \mid N = j) \\ &= \frac{1 + \lfloor 0{,}375 \times 10^j \rfloor}{10^j}. \end{aligned}$$

La partie entière $\lfloor 0{,}375 \times 10^j \rfloor$ utilisée dans cette formule n'a d'importance que pour $j < 3$. En effet $0{,}375 \times 10^j$ est entier pour $j \geqslant 3$. Pour $j < 3$, on peut d'ailleurs vérifier directement que

$$\begin{aligned} P(X \leqslant 0{,}375 \mid N = 1) &= P(Y_1 \leqslant 3 \mid N = 1) = \frac{4}{10}, \\ P(X \leqslant 0{,}375 \mid N = 2) &= P\big((Y_1 \leqslant 2 \text{ et } Y_2 \text{ quelconque}) \text{ ou } (Y_1 = 3 \text{ et } Y_2 \leqslant 7) \mid N = 2\big) \\ &= \frac{38}{100}. \end{aligned}$$

11) Le raisonnement fait à la question précédente se généralise immédiatement à d quelconque dans $\mathbb{D}$.

$$\forall d \in \mathbb{D}, \qquad P(X \leqslant d \mid N = j) = \frac{\lfloor 10^j d \rfloor + 1}{10^j}. \tag{11.17}$$

12) *Calcul de $F(k/10)$ pour $0 \leqslant k \leqslant 9$.*

En appliquant la formule des probabilités totales,

$$\begin{aligned} F\Big(\frac{k}{10}\Big) = P\Big(X \leqslant \frac{k}{10}\Big) &= \sum_{j=1}^{+\infty} P\Big(X \leqslant \frac{k}{10} \mid N = j\Big) P(N = j) \\ &= \sum_{j=1}^{+\infty} \frac{\lfloor k10^{j-1} \rfloor + 1}{10^j} \, q^{j-1} p \\ &= \sum_{j=1}^{+\infty} \frac{k}{10} \, q^{j-1} p + \sum_{j=1}^{+\infty} \frac{q^{j-1} p}{10^j} \\ &= \frac{k}{10} + \frac{p}{10} \sum_{j=1}^{+\infty} \Big(\frac{q}{10}\Big)^{j-1} \\ &= \frac{k}{10} + \frac{p}{9+p}. \end{aligned}$$

Si $d = l/100$ avec $0 \leqslant l \leqslant 99$, on a $\lceil 10^j d \rceil = 10^j d$ dès que $j \geqslant 2$. D'où le calcul

$$\begin{aligned} F(d) = P\Big(X \leqslant d\Big) &= \sum_{j=1}^{+\infty} P(X \leqslant d \mid N = j) P(N = j) \\ &= \frac{\lceil 10d \rceil}{10} p + \sum_{j=2}^{+\infty} d q^{j-1} p + \sum_{j=1}^{+\infty} \frac{q^{j-1} p}{10^j} \\ &= \frac{\lceil 10d \rceil}{10} p + dq + \frac{p}{9+p}. \end{aligned}$$

En particulier quand l est multiple de 10, $\lceil 10d \rceil / 10 = d$ et on retrouve la formule établie pour d de la forme $k/10$, $F(d) = d + p/(9+p) = d + F(0)$.

13) *Formule générale pour $F(d)$, $d \in \mathbb{D}$.*

Prenons $d = k10^{-n}$ (pas forcément sous forme réduite), $0 \leqslant k < 10^n$. Alors pour $j \geqslant n$ on a $\lceil 10^j d \rceil 10^{-j} = d$. En appliquant la même méthode que ci-dessus, on arrive à

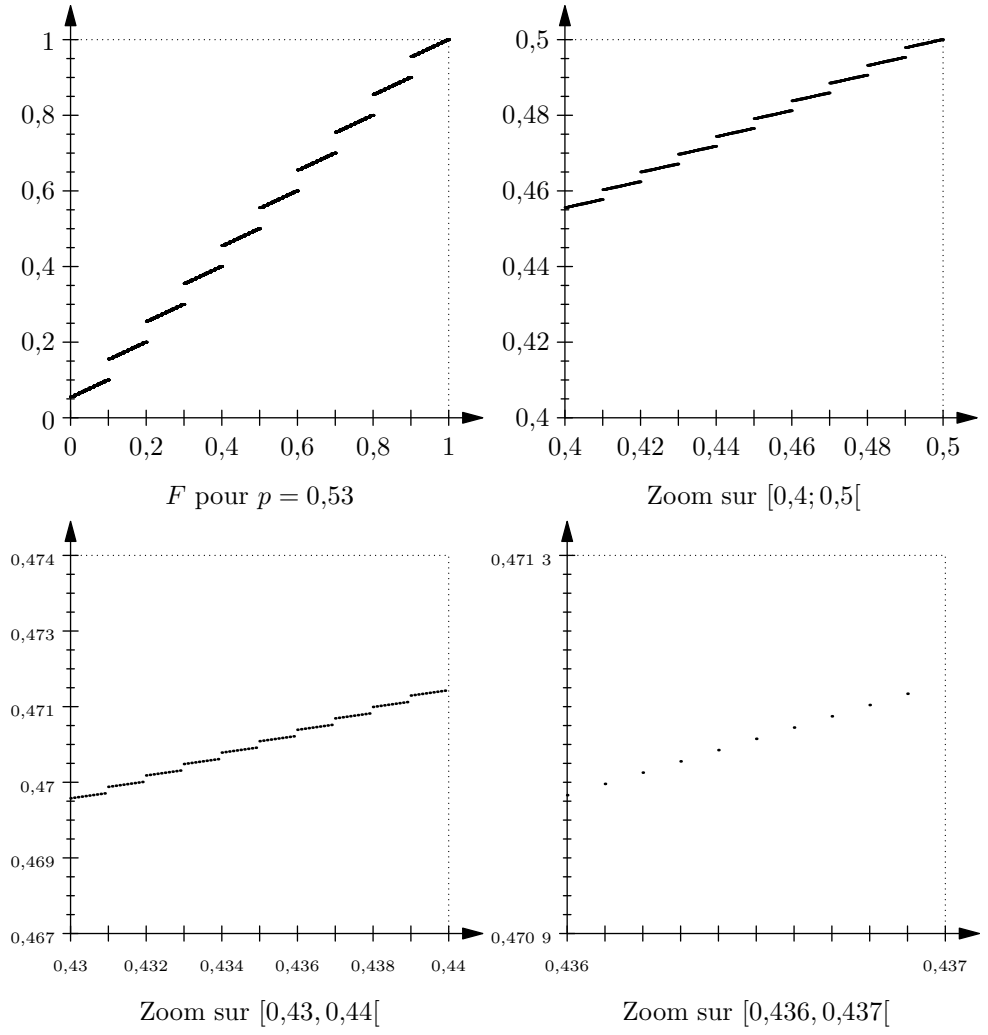
$$\begin{aligned} P(X \leqslant d) &= \sum_{j=1}^{n-1} \frac{\lceil 10^j d \rceil}{10^j} q^{j-1} p + \sum_{j=n}^{+\infty} d q^{j-1} p + \frac{p}{9+p} \\ &= \sum_{j=1}^{n-1} \frac{\lceil 10^j d \rceil}{10^j} q^{j-1} p + d q^{n-1} + \frac{p}{9+p}. \end{aligned}$$

On peut voir que cette formule ne dépend pas de la représentation choisie pour d, en remarquant que si $k = 10^m l$, pour $j \geqslant n - m$, $\lceil 10^j d \rceil 10^{-j} = d$ et pour avoir la même valeur de $F(d)$ avec les deux représentations $d = k10^{-n}$ et $d = l10^{-n+m}$, il reste seulement à vérifier que

$$\sum_{j=n-m}^{n-1} q^{j-1} p + q^{n-1} = q^{n-m-1},$$

exercice élémentaire sur les suites géométriques, laissé au lecteur. On présente maintenant quelques représentations graphiques de la fonction de répartition F pour $p = 0{,}53$ (figure 11.7, p. 462). À proprement parler, il est impossible de dessiner physiquement le graphe de F puisqu'il présente une infinité de discontinuités dans tout intervalle. On s'est contenté de faire calculer [10] les valeurs de $F(d)$ pour $d = k10^{-4}$ et d'afficher les points $(d, F(d))$. Le niveau de résolution de l'écran et celui de l'imprimante donnent une illusion de continuité quand l'amplitude des sauts est inférieure à un pixel. Les zooms successifs montrent qu'il n'en est rien et font apparaître la structure fractale de F.

10. Les calculs et les graphiques ont été réalisés avec le logiciel libre Asymptote.



F pour $p = 0{,}53$ — Zoom sur $[0{,}4; 0{,}5[$

Zoom sur $[0{,}43, 0{,}44[$ — Zoom sur $[0{,}436, 0{,}437[$

FIGURE 11.7 – Représentation graphique de 10 000 points $(d, F(d))$

11.7 Exercices du chapitre 7

Solution 7.2 *Contrôleur contre fraudeur*

1) Pour $1 \leqslant j < k$, notons $E_{j,k}$ l'évènement : « le premier contrôle du fraudeur a lieu lors de son j^{e} trajet et le deuxième contrôle lors du k^{e} trajet ». On a clairement

la décomposition en union disjointe :

$$\forall k \geqslant 2, \quad \{T = k\} = \bigcup_{j=1}^{k-1} E_{j,k}. \tag{11.18}$$

Par additivité de la probabilité, le calcul de $P(T = k)$ se ramène donc à celui des $P(E_{j,k})$. Notons F_i l'évènement « le fraudeur n'est pas contrôlé lors de son i^{e} trajet ». Son complémentaire F_i^c est donc l'évènement « le fraudeur est contrôlé lors de son i^{e} trajet ». L'évènement $E_{j,k}$ peut s'écrire sous la forme

$$E_{j,k} = F_j^c \cap F_k^c \cap \Big(\bigcap_{\substack{1 \leqslant i < k \\ i \neq j}} F_i \Big),$$

avec le cas particulier $E_{1,2} = F_1^c \cap F_2^c$. Nous faisons maintenant l'hypothèse d'indépendance [11] de la suite d'évènements $(F_i)_{i \in \mathbb{N}*}$. Alors les k évènements F_i $(i \neq j, k)$ et F_j^c, F_k^c sont mutuellement indépendants (cf. la proposition 5.53) et

$$P(E_{j,k}) = P(F_j^c) P(F_k^c) \prod_{\substack{1 \leqslant i < k \\ i \neq j}} P(F_i) = p^2 q^{k-2}.$$

On note au passage que $P(E_{j,k})$ ne dépend pas de j. En utilisant l'additivité de P et (11.18), on aboutit à :

$$\forall k \geqslant 2, \quad P(T = k) = \sum_{j=1}^{k-1} P(E_{j,k}) = (k-1) p^2 q^{k-2}.$$

La variable aléatoire T suit ainsi la loi binomiale négative de paramètres 2 et p. C'est la loi du temps d'attente du deuxième succès (ici pour le contrôleur !) dans une suite d'épreuves répétées indépendantes.

2) La probabilité $P(T > n)$ s'exprime à l'aide des $P(T = k)$ par σ-additivité :

$$P(T > n) = \sum_{k=n+1}^{+\infty} P(T = k) = p^2 \sum_{k=n+1}^{+\infty} (k-1) q^{k-2}.$$

On voit ainsi que

$$P(T > n) = p^2 f_n'(q), \tag{11.19}$$

où la fonction f_n est définie comme la somme de la série entière

$$f_n(x) := \sum_{k=n+1}^{+\infty} x^{k-1}.$$

11. Cette hypothèse revient pratiquement à dire que les contrôles sont faits au hasard, sans tenir compte des résultats des contrôles précédents. Elle ne serait pas pertinente si par exemple, le fraudeur était pris en filature par un contrôleur particulièrement hargneux et décidé à le coincer.

Cette série est une série géométrique de raison x, donc de rayon de convergence $R = 1$. On sait alors que sa somme est dérivable sur l'intervalle $]-R, R[$ et que sa dérivée peut se calculer par dérivation terme à terme :

$$\forall x \in]-1, 1[, \quad f_n'(x) := \sum_{k=n+1}^{+\infty} (k-1)x^{k-2}.$$

Comme $0 < p < 1$ (sinon le problème est sans intérêt), $q = 1 - p$ est aussi dans l'intervalle $]0, 1[$ et on peut appliquer l'égalité ci-dessus avec $x = q$. Par ailleurs, on peut calculer explicitement $f_n(x)$ comme somme d'une série géométrique :

$$f_n(x) = x^n + x^{n+1} + x^{n+2} + \cdots = x^n \sum_{j=0}^{+\infty} x^j = \frac{x^n}{1-x},$$

formule valable pour tout $x \in]-1, 1[$. On en déduit une formule explicite pour la dérivée :

$$f_n'(x) = \frac{(1-x)nx^{n-1} + x^n}{(1-x)^2} = \frac{(n-(n-1)x)x^{n-1}}{(1-x)^2}.$$

En faisant $x = q$ (donc $1 - x = p$) et en reportant ce résultat dans (11.19), il vient :

$$P(T > n) = (n-(n-1)q)q^{n-1} = (np+q)q^{n-1}.$$

3) Le bilan financier de la stratégie du fraudeur est la variable aléatoire $S := T - 60$. En effet lorsqu'il arrête de frauder après son deuxième contrôle, il a dépensé en tout 60 € d'amende et gagné T € en ne payant pas ses tickets de métro. La probabilité que sa stratégie soit bénéficiaire est donc $P(S > 0) = P(T > 60)$.

Le calcul numérique donne

$$P(T > 60) \simeq \begin{cases} 0{,}013\,8 & \text{pour } p = 1/10; \\ 0{,}191\,6 & \text{pour } p = 1/20. \end{cases}$$

4) D'un point de vue financier, le fraudeur a intérêt à avoir une idée de la valeur de p. Entre les deux exemples ci-dessus, la division par 2 de la probabilité de contrôle multiplie par presque 14 sa probabilité d'être bénéficaire. Pour diminuer son ignorance de p, le fraudeur peut l'estimer sur la base d'un grand nombre d'observations.

Considérons d'abord le cas du fraudeur « névrotique » : son objectif est de faire un bénéfice, peu lui importe lequel. Dans ce cas, la valeur critique de p est celle pour laquelle $P(T > 60) = 1/2$, soit approximativement $p \simeq 0{,}028$. En dessous de cette valeur de p, la stratégie du fraudeur lui donne plus d'une chance sur deux d'être bénéficiaire.

Voyons maintenant le cas plus réaliste du fraudeur « économique » : ce qui lui importe n'est pas tant la probabilité de faire un bénéfice, même minime, que le gain moyen qu'il peut espérer de sa stratégie. Pour lui, la quantité pertinente est non plus $P(T > 60)$, mais $\mathbf{E}(T - 60) = \mathbf{E}T - 60$. Un calcul simple (laissé en exercice) montre

que $\mathbf{E}T = 2/p$. La valeur critique de p en dessous de laquelle sa stratégie est gagnante « en moyenne » est alors donnée par l'équation $2/p = 60$, c'est $p = 1/30 \simeq 0{,}033\,3$. Notons que pour cette valeur de p, $P(T > 60) \simeq 0{,}401\,5$, ce qui montre bien que la prise en compte du montant des gains peut conduire à considérer comme gagnante une stratégie qui a moins d'une chance sur deux de générer un bénéfice.

Nota bene. L'auteur de l'énoncé *Contrôleur contre fraudeur* et de son corrigé récuse par avance toute accusation d'incitation à la fraude et décline toute responsabilité en cas de fraude de l'un de ses lecteurs. Il fait observer que des calculs du type ci-dessus [12] sont utiles à la compagnie de métro qui en a besoin pour déterminer le nombre de contrôleurs qu'elle doit employer (avec le coût salarial que cela implique)...

Solution 7.5 X est une variable aléatoire à valeurs dans $\mathbb{N}^*$, intégrable et en notant $p_k = P(X = k)$, $(p_k)_{k\geqslant 1}$ est *décroissante* sur $\mathbb{N}^*$.

1) Pour montrer l'inégalité :

$$\forall k \in \mathbb{N}^*, \quad P(X = k) < \frac{2\mathbf{E}X}{k^2}, \tag{11.20}$$

on remarque que :

$$\mathbf{E}X = \sum_{j=1}^{+\infty} jP(X = j) \geqslant \sum_{j=1}^{k} jP(X = j) \geqslant P(X = k) \sum_{j=1}^{k} j = P(X = k)\frac{k(k+1)}{2},$$

où la deuxième inégalité utilise la décroissance de $(p_j)_{j\geqslant 1}$. On note ensuite que $\mathbf{E}X$ est strictement positif car X est à valeur dans $\mathbb{N}^*$ donc il existe au moins un $p_j > 0$ et $\mathbf{E}X \geqslant jp_j > 0$. Si $P(X = k) > 0$, alors on déduit l'inégalité stricte dans (11.20) de la minoration $\mathbf{E}X \geqslant P(X = k)\frac{k(k+1)}{2}$ et de $k(k+1) > k^2$. Si $P(X = k) = 0$, l'inégalité stricte dans (11.20) est trivialement vérifiée puisque $\mathbf{E}X > 0$.

2) Il n'existe pas de constante $c > 0$ et d'entier k_0 tels que :

$$\forall k \geqslant k_0, \quad P(X = k) \geqslant \frac{c\mathbf{E}X}{k^2},$$

car sinon

$$\sum_{j=1}^{+\infty} jP(X = j) \geqslant c \sum_{k=k_0}^{+\infty} k\frac{c\mathbf{E}X}{k^2} = c\mathbf{E}X \sum_{k=k_0}^{+\infty} \frac{1}{k} = +\infty,$$

d'où $\mathbf{E}X = +\infty$, ce qui est contradictoire avec l'intégrabilité de X.

3) Voici un exemple de loi de variable aléatoire X vérifiant toutes les hypothèses de la question 1) sauf la décroissance de $(p_k)_{k\geqslant 1}$ et qui ne vérifie pas (11.20). On pose

$$p_k = P(X = k) = \begin{cases} a3^{-j} & \text{si } k = 2^j,\ j \in \mathbb{N} \\ 0 & \text{sinon,} \end{cases}$$

12. Notamment le cas du fraudeur « économique », en raison de la loi des grands nombres.

où la constante a est choisie de façon à avoir vraiment une loi de probabilité sur $\mathbb{N}^*$, c'est à dire telle que $a\sum_{j=0}^{+\infty} 3^{-j} = 1$, d'où $a = 2/3$. L'intégrabilité de X résulte de la convergence de la série géométrique $\sum_{j=0}^{+\infty} 2^j 3^{-j}$ de raison $2/3$. La sous-suite de $(k^2 P(X = k))$ indexée par les k de la forme $k = 2^j$ s'écrit $(a4^j 3^{-j})_{j\in\mathbb{N}}$ et tend vers l'infini. Ceci contredit (11.20).

Solution 7.7 *Espérance des statistiques d'ordre*

1) *Question préliminaire.* Pour tout $n \in \mathbb{N}^*$ et tout $j \in [\![0, n]\!]$, on pose

$$I_{j,n} := \int_0^1 t^j (1-t)^{n-j}\,\mathrm{d}t.$$

Le calcul de $I_{n,n}$ est immédiat :

$$I_{n,n} = \int_0^1 t^n\,\mathrm{d}t = \left[\frac{t^{n+1}}{n+1}\right]_0^1 = \frac{1}{n+1}.$$

En intégrant par parties $I_{j,n}$ pour $j \in [\![0, n[\![$, il vient

$$I_{j,n} = \left[\frac{t^{j+1}}{j+1}(1-t)^{n-j}\right]_0^1 + \frac{n-j}{j+1}\int_0^1 t^{j+1}(1-t)^{n-j-1}\,\mathrm{d}t,$$

d'où

$$I_{j,n} = \frac{n-j}{j+1} I_{j+1,n}.$$

En itérant cette égalité pour $j, j+1, j+2, \ldots, n-1$, on en déduit :

$$I_{j,n} = \frac{n-j}{j+1} \times \frac{n-j-1}{j+2} \times \cdots \times \frac{1}{n} \times I_{n,n} = \frac{(n-j)!}{\frac{n!}{j!}} \times \frac{1}{n+1},$$

d'où

$$I_{j,n} = \frac{j!(n-j)!}{(n+1)!}. \tag{11.21}$$

Dans toute la suite, on note $X_1, \ldots, X_n$ une suite de variables aléatoires *positives*, définies sur le même espace probabilisé $(\Omega, \mathcal{F}, P)$, *indépendantes et de même loi*, donc de même fonction de répartition notée F. Le but de ce problème est de calculer l'espérance des variables aléatoires $X_{n:1}, \ldots X_{n:n}$ obtenues par réarrangement croissant de $X_1, \ldots, X_n$ et appelées *statistiques d'ordre*. Plus précisément, pour tout $\omega \in \Omega$, on a *l'égalité des ensembles* $\{X_{n:k}(\omega)\ ;\ k \in [\![1, n]\!]\}$ et $\{X_k(\omega)\ ;\ k \in [\![1, n]\!]\}$ et de plus

$$\forall k \in [\![1, n[\![, \quad X_{n:k}(\omega) \leqslant X_{n:k+1}(\omega).$$

En particulier, $X_{n:1} = \min(X_1, \ldots, X_n)$ et $X_{n:n} = \max(X_1, \ldots, X_n)$.

2) Soit I une partie non vide de $[\![1, n]\!]$ et

$$p_I(x) := P(\forall i \in I,\ X_i \leqslant x \text{ et } \forall i \in [\![1, n]\!]\backslash I,\ X_i > x).$$

Cette probabilité peut s'écrire

$$p_I(x) = P\Big(\bigcap_{i\in I} A_i \cap \bigcap_{i\in[\![1,n]\!]\setminus I} A_i^c\Big),$$

en introduisant pour x fixé et $i \in [\![1,n]\!]$ les évènements

$$A_i := \{X_i \leqslant x\}, \quad A_i^c = \{X_i > x\}.$$

Ces évènements héritent de l'indépendance des X_i, par conséquent :

$$p_I(x) = \prod_{i\in I} P(A_i) \times \prod_{i\in[\![1,n]\!]\setminus I} P(A_i^c) = F(x)^{\operatorname{card} I}\big(1-F(x)\big)^{n-\operatorname{card} I}. \tag{11.22}$$

3) *Calcul de F_k, f.d.r. de $X_{n:k}$.*
Avec les notations de la question précédente, posons pour $I \subset [\![1,n]\!]$

$$B_I := \Big(\bigcap_{i\in I} A_i\Big) \cap \Big(\bigcap_{i\in[\![1,n]\!]\setminus I} A_i^c\Big).$$

On remarque que « $X_{n:k} \leqslant x$ » est équivalent à « au moins k parmi les X_i sont inférieurs ou égaux à x ». Ceci peut encore s'écrire

$$\{X_{n:k} \leqslant x\} = \bigcup_{\substack{I\subset[\![1,n]\!]\\ \operatorname{card} I\geqslant k}} B_I = \bigcup_{j=k}^{n} \bigcup_{\substack{I\subset[\![1,n]\!]\\ \operatorname{card} I=j}} B_I.$$

Les B_I sont deux à deux disjoints car si $I \neq I'$, il existe un indice i_0 appartenant à l'un de ces deux ensembles mais pas à l'autre et $B_I \cap B_{I'} \subset A_{i_0} \cap A_{i_0}^c = \varnothing$. On en déduit que

$$P(X_{n:k} \leqslant x) = \sum_{j=k}^{n} \sum_{\substack{I\subset[\![1,n]\!]\\ \operatorname{card} I=j}} P(B_I).$$

D'après (11.21), $P(B_I) = p_I(x)$ ne dépend de I que par $\operatorname{card} I$ et comme il y a exactement C_n^j parties de $[\![1,n]\!]$ ayant pour cardinal j, on obtient

$$\forall x \geqslant 0, \quad F_k(x) = P\big(X_{n:k} \leqslant x\big) = \sum_{j=k}^{n} C_n^j F(x)^j \big(1-F(x)\big)^{n-j}. \tag{11.23}$$

4) Dans la somme $\sum_{j=k}^{n}$ ci-dessus, on reconnaît un morceau du développement de $\big(F(x)+(1-F(x))\big)^n$ par la formule du binôme de Newton. Plus précisément on a

$$\begin{aligned} 1 = \big(F(x)+(1-F(x))\big)^n &= \sum_{j=0}^{n} C_n^j F(x)^j \big(1-F(x)\big)^{n-j} \\ &= \sum_{j=0}^{k-1} C_n^j F(x)^j \big(1-F(x)\big)^{n-j} + \sum_{j=k}^{n} C_n^j F(x)^j \big(1-F(x)\big)^{n-j}. \end{aligned}$$

Compte-tenu de (11.23), on a donc

$$\forall x \geqslant 0, \quad 1 - F_k(x) = \sum_{j=0}^{k-1} C_n^j F(x)^j \big(1 - F(x)\big)^{n-j}. \tag{11.24}$$

En intégrant sur $[0, +\infty[$ cette égalité entre fonctions positives localement Riemann intégrables, on obtient :

$$\mathbf{E}X_{n:k} = \int_0^{+\infty} (1 - F_k(x))\,\mathrm{d}x = \sum_{j=0}^{k-1} C_n^j \int_0^{+\infty} F(x)^j \big(1 - F(x)\big)^{n-j}\,\mathrm{d}x. \tag{11.25}$$

5) On suppose dans cette question que les X_i suivent la loi uniforme sur $[0,1]$. Dans ce cas $F(x) = x$ pour $x \in [0,1]$ et $F(x) = 1$ pour $x > 1$. En particulier $1 - F$ est nulle sur $[1, +\infty[$. On peut alors calculer explicitement les intégrales de la formule (11.25) qui se réduisent toutes [13] à des intégrales sur $[0,1]$:

$$\mathbf{E}X_{n:k} = \sum_{j=0}^{k-1} C_n^j \int_0^1 x^j (1-x)^{n-j}\,\mathrm{d}x$$

On reconnaît les intégrales $I_{j,n}$ calculées dans la question préliminaire, d'où

$$\mathbf{E}X_{n:k} = \sum_{j=0}^{k-1} C_n^j \frac{j!(n-j)!}{(n+1)!} = \sum_{j=0}^{k-1} \frac{n!}{j!(n-j)!} \frac{j!(n-j)!}{(n+1)!} = \sum_{j=0}^{k-1} \frac{1}{n+1}.$$

Cette dernière somme est constituée de k termes tous égaux à $1/(n+1)$. Finalement on obtient

$$\mathbf{E}X_{n:k} = \frac{k}{n+1}. \tag{11.26}$$

Par linéarité de l'espérance, on en déduit immédiatement les espérances des espacements $Y_{n,k} := X_{n:k+1} - X_{n:k}$:

$$\mathbf{E}Y_{n,k} = \mathbf{E}X_{n:k+1} - \mathbf{E}X_{n:k} = \frac{k+1}{n+1} - \frac{k}{n+1} = \frac{1}{n+1}. \tag{11.27}$$

Ce résultat généralise celui de l'exercice 7.6 question 4 qui traitait le cas $n = 2$.

6) On suppose dans cette question que les X_i suivent la loi exponentielle de paramètre 1. Donc pour tout $x \geqslant 0$, $P(X_i > x) = \mathrm{e}^{-x}$. On peut calculer les intégrales $\int_0^{+\infty} F(x)^j \big(1-F(x)\big)^{n-j}\,\mathrm{d}x$ en faisant le changement de variable $u = 1-\mathrm{e}^{-x}$ (croissant et C^1) dans cette intégrale généralisée de fonction positive, ce qui donne

$$\int_0^{+\infty} F(x)^j \big(1-F(x)\big)^{n-j}\,\mathrm{d}x = \int_0^1 u^j (1-u)^{n-j} \frac{\mathrm{d}u}{1-u} = \int_0^1 u^j (1-u)^{n-1-j}\,\mathrm{d}u = I_{j,n-1}.$$

13. Notons que $n - j \geqslant 1$ pour tout $j = 0, \ldots k-1 < n$. Pour $j = n$, $(1-F)^{n-j} = (1-F)^0$ est la fonction constante 1 qui ne s'annule pas sur $[1, +\infty[$.

En reportant ce résultat dans (11.25), il vient

$$\mathbf{E}X_{n:k} = \sum_{j=0}^{k-1} \frac{n!}{j!(n-j)!} \frac{j!(n-1-j)!}{n!} = \sum_{j=0}^{k-1} \frac{1}{n-j}.$$

En particulier,

$$\mathbf{E} \min_{i\leqslant n} X_i = \frac{1}{n}, \quad \mathbf{E} \max_{i\leqslant n} X_i = 1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n}.$$

7) Supposons que la variable aléatoire positive X_1 est *intégrable*. Les X_i ayant même loi sont donc toutes intégrables et de même espérance finie. Pour montrer que chacune des v.a. positives $X_{n:k}$ est intégrable, on remarque que $0 \leqslant X_{n:k} \leqslant X_{n:n} = \max_{i\leqslant n} X_i$. Il suffit donc de vérifier l'intégrabilité de $\max_{i\leqslant n} X_i$. Cette intégrabilité résulte de l'inégalité suivante et des propriétés de croissance et d'additivité de l'espérance des v.a. positives :

$$0 \leqslant \max_{i\leqslant n} X_i \leqslant \sum_{i=1}^{n} X_i,$$

d'où

$$0 \leqslant \mathbf{E} \max_{i\leqslant n} X_i \leqslant \sum_{i=1}^{n} \mathbf{E}X_i = n\mathbf{E}X_1 < +\infty.$$

8) On suppose dans cette question que F est C^1 sur $[0, +\infty[$, continue en zéro et on note f sa dérivée. La fonction de répartition F_k est alors continue sur $\mathbb{R}$, C^1 par morceaux comme composée par un polynôme de la fonction F elle même continue sur $\mathbb{R}$ et C^1 par morceaux. On sait qu'alors la loi correspondante admet une densité f_k qui s'obtient par dérivation de F_k (sauf peut-être au point de raccord 0). La fonction F_k est nulle en tout point de $]-\infty, 0[$, donc $f_k = F_k'$ aussi. Pour $x > 0$, en dérivant dans (11.23), on obtient

$$\begin{aligned} f_k(x) &= \sum_{j=k}^{n} \Big(C_n^j F(x)^j \big(1-F(x)\big)^{n-j} \Big)' \\ &= \sum_{j=k}^{n} \Big(jC_n^j F(x)^{j-1} \big(1-F(x)\big)^{n-j} - (n-j)C_n^j F(x)^j \big(1-F(x)\big)^{n-1-j} \Big) f(x) \end{aligned}$$

Les formules suivantes vont nous permettre de simplifier considérablement cette expression de f_k.

$$\begin{aligned} jC_n^j &= \frac{jn!}{j!(n-j)!} = \frac{n!}{(j-1)!(n-j)!} = nC_{n-1}^{j-1}, \\ (n-j)C_n^j &= \frac{(n-j)n!}{j!(n-j)!} = \frac{n!}{j!(n-j-1)!} = nC_{n-1}^{j}. \end{aligned}$$

En reportant ceci dans l'expression de $f_k(x)$ ci-dessus et en posant pour alléger les écritures

$$a_{n,j}(x) := C_{n-1}^{j} F(x)^j \big(1 - F(x)\big)^{n-1-j},$$

on peut écrire :

$$f_k(x) = nf(x) \sum_{j=k}^{n} \big(a_{n,j-1}(x) - a_{n,j}(x)\big) = nf(x) a_{n,k-1}(x),$$

après simplifications en chaîne et en notant que $a_{n,n} = 0$. Finalement,

$$\forall x > 0, \quad f_k(x) = \frac{n!}{(k-1)!(n-k)!} F(x)^{k-1} \big(1 - F(x)\big)^{n-k} f(x).$$

Solution 7.8 *Loi logistique*

1) Soit Y une variable aléatoire *positive* de loi exponentielle de paramètre 1, sa fonction de survie est donc donnée par

$$P(Y > x) = \begin{cases} 1 & \text{si } x < 0, \\ \mathrm{e}^{-x} & \text{si } x \geqslant 0. \end{cases}$$

Puisque Y est une v.a. positive,

$$\mathbf{E}Y = \int_0^{+\infty} P(Y > x)\,\mathrm{d}x = \int_0^{+\infty} \mathrm{e}^{-x}\,\mathrm{d}x = \lim_{b \to +\infty} \big[-\mathrm{e}^{-x}\big]_0^b = 1 < +\infty,$$

donc Y est intégrable et $\mathbf{E}Y = 1$.

Pour t réel quelconque, on a

$$P(Y' \leqslant t) = P(-Y \leqslant t) = P(Y \geqslant -t) = P(Y > -t) + P(Y = -t).$$

Comme Y est à densité, sa f.d.r. est continue donc $P(Y = -t) = 0$. On en déduit en distinguant les deux cas $-t < 0$ et $-t \geqslant 0$ que

$$F_{Y'}(t) = \begin{cases} \mathrm{e}^{t} & \text{si } t \leqslant 0, \\ 1 & \text{si } t > 0. \end{cases}$$

2) On note $Z := \ln U$, où U est une variable aléatoire réelle à valeurs dans $]0,1[$, de loi uniforme sur $]0,1[$. On cherche la loi et l'espérance de Z. Pour cela on va calculer la fonction de répartition de Z. Remarquons que U étant à valeurs dans $]0,1[$, $U(\omega) > 0$ pour tout $\omega \in \Omega$, donc $Z(\omega) = \ln(U(\omega))$ est bien défini pour tout $\omega \in \Omega$. Rappelons d'abord que la fonction de répartition de U est donnée par

$$P(U \leqslant x) = \begin{cases} 0 & \text{si } x \leqslant 0 \\ x & \text{si } 0 < x \leqslant 1, \\ 1 & \text{si } x > 1. \end{cases} \tag{11.28}$$

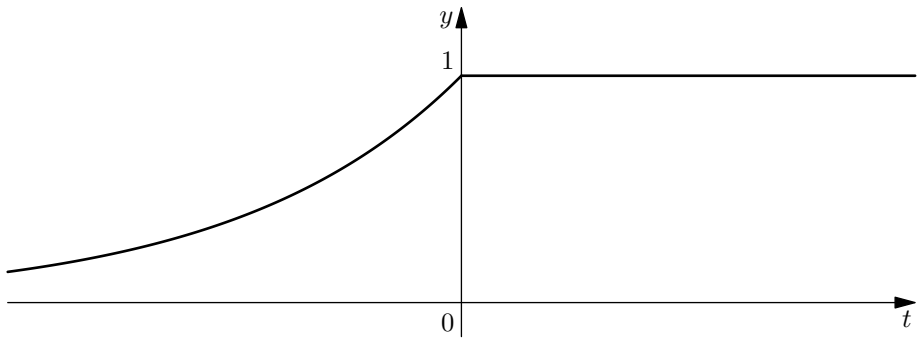


FIGURE 11.8 – Fonction de répartition de $Y' = -Y$

Pour tout $t \in \mathbb{R}$ et tout $\omega \in \Omega$, on a l'équivalence

$$\ln\big(U(\omega)\big) \leqslant t \Leftrightarrow U(\omega) \leqslant \mathrm{e}^t. \tag{11.29}$$

En effet l'implication de gauche à droite vient de la croissance sur $\mathbb{R}$ de la fonction exponentielle. L'implication réciproque vient de la croissance sur $]0, +\infty[$ de la fonction logarithme, en notant que $U(\omega)$ et e^t sont tous deux *strictement positifs*. On déduit de (11.29) l'égalité d'évènements $\{\ln U \leqslant t\} = \{U \leqslant \mathrm{e}^t\}$. En combinant cette égalité avec (11.28) appliquée avec $x = \mathrm{e}^t$, on justifie le calcul suivant :

$$P(Z \leqslant t) = P(\ln U \leqslant t) = P(U \leqslant \mathrm{e}^t) = \begin{cases} \mathrm{e}^t & \text{si } \mathrm{e}^t \leqslant 1 \\ 1 & \text{si } \mathrm{e}^t > 1 \end{cases} = \begin{cases} \mathrm{e}^t & \text{si } t \leqslant 0 \\ 1 & \text{si } t > 0 \end{cases} = P(Y' \leqslant t).$$

Ceci étant valable pour tout réel t, Z a même f.d.r. donc *a même loi* que Y'. Or $Y' = -Y$ est intégrable parce que Y l'est. Par linéarité de l'espérance, $\mathbf{E}Y' = -\mathbf{E}Y = -1$. Par conséquent, Z est intégrable et a pour espérance $\mathbf{E}Z = \mathbf{E}Y' = -1$.

3) On définit la variable aléatoire réelle X par

$$X := \ln\Big(\frac{1-U}{U}\Big).$$

Notons que U étant à valeurs dans $]0,1[$, $\frac{1-U(\omega)}{U(\omega)}$ est défini et strictement positif pour tout $\omega \in \Omega$, donc son logarithme $X(\omega)$ est bien défini pour tout $\omega \in \Omega$. Calculons la fonction de survie $G : x \mapsto G(x) := P(X > x)$. Sous réserve de justifications, on a :

$$\begin{aligned} \forall x \in \mathbb{R}, \quad P(X > x) = P\Big(\ln\Big(\frac{1-U}{U}\Big) > x\Big) &= P\Big(\frac{1-U}{U} > \mathrm{e}^x\Big) && (11.30) \\ &= P\big(1 - U > U\mathrm{e}^x\big) && (11.31) \\ &= P\Big(U < \frac{1}{1+\mathrm{e}^x}\Big) \\ &= \frac{1}{1+\mathrm{e}^x}. && (11.32) \end{aligned}$$

Justifications.

(11.30) : avec la même justification que pour (11.29), on a l'égalité d'évènements :

$$\left\{ \ln\left(\frac{1-U}{U}\right) > x \right\} = \left\{ \frac{1-U}{U} > \mathrm{e}^x \right\}.$$

(11.31) : par *stricte positivité* de U sur Ω, on a équivalence entre $\frac{1-U}{U} > \mathrm{e}^x$ et $1-U > U\mathrm{e}^x$.

(11.32) : on applique (11.28) en notant que $(1+\mathrm{e}^x)^{-1}$ est dans $]0,1[$ pour tout $x \in \mathbb{R}$.

La fonction de survie de X est donc $G : x \mapsto (1+\mathrm{e}^x)^{-1}$, d'où la fonction de répartition F donnée par :

$$\forall x \in \mathbb{R}, \quad F(x) = 1 - \frac{1}{1+\mathrm{e}^x}.$$

Cette fonction étant visiblement C^1 sur $\mathbb{R}$, la loi de X admet pour densité la dérivée f de F :

$$\forall x \in \mathbb{R}, \quad f(x) = \frac{\mathrm{e}^x}{(1+\mathrm{e}^x)^2}.$$

On remarque que $f(x) \sim \mathrm{e}^x$ pour $x \to -\infty$ et $f(x) \sim \mathrm{e}^{-x}$ pour $x \to +\infty$, ce qui permet de vérifier la convergence de l'intégrale gnéralisée $\int_{-\infty}^{+\infty} f(x)\,\mathrm{d}x$. On voit aussi que f est paire car

$$f(-x) = \frac{\mathrm{e}^{-x}}{(1+\mathrm{e}^{-x})^2} = \frac{\mathrm{e}^{2x}\mathrm{e}^{-x}}{\mathrm{e}^{2x}(1+\mathrm{e}^{-x})^2} = \frac{\mathrm{e}^x}{(\mathrm{e}^x+1)^2} = f(x).$$

On en déduit que la loi de X est *symétrique*, c.-à-d. que X et $-X$ ont même loi.

La fonction de répartition F et la densité f sont représentées sur les figures 11.9 et 11.10 respectivement.

4) Pour vérifier l'intégrabilité de X et calculer son espérance, on commence par noter que $X = \ln(1-U) - \ln U$. Nous connaissons déjà la loi et l'espérance de $\ln U$. La loi de $1-U$ est la même que celle de U, ce que l'on peut vérifier en calculant sa fonction de répartition comme suit. On remarque d'abord que

$$P(1-U \leqslant t) = P(U \geqslant 1-t) = 1 - P(U < 1-t) = 1 - P(U \leqslant 1-t),$$

la dernière égalité étant justifiée par la continuité de la f.d.r. de U. On en déduit

$$P(1-U \leqslant t) = \begin{cases} 1-0 & \text{si } 1-t \leqslant 0 \\ 1-(1-t) & \text{si } 0 < 1-t \leqslant 1 \\ 1-1 & \text{si } 1-t > 1 \end{cases} = \begin{cases} 1 & \text{si } t \geqslant 1 \\ t & \text{si } 0 \leqslant t < 1 \\ 0 & \text{si } t < 0 \end{cases}.$$

On constate ainsi que $P(1-U \leqslant t) = P(U \leqslant t)$ pour tout réel t (noter que la différence entre inégalités stricte et large en 0 et 1 avec la formule pour $P(U \leqslant t)$ est sans importance en raison de la continuité de la f.d.r. de U en ces points).

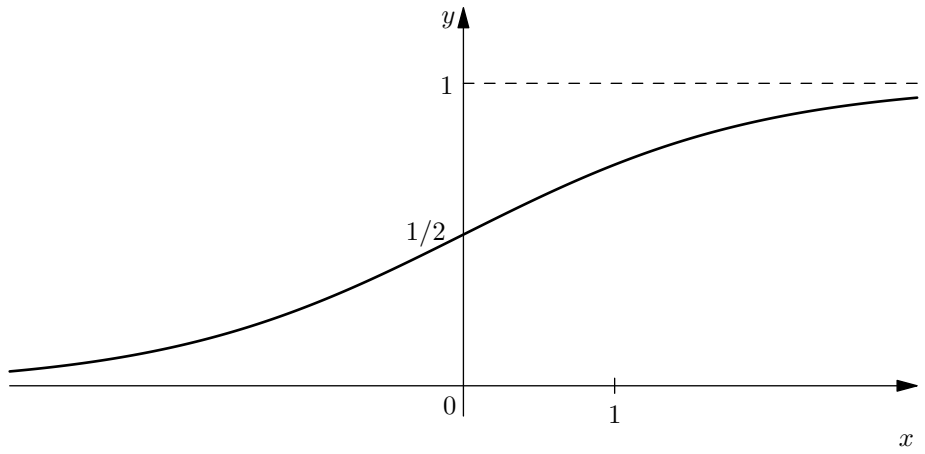


FIGURE 11.9 – Fonction de répartition F de la v.a. $X = \ln\left(\frac{1-U}{U}\right)$

L'égalité des lois de U et $1-U$ entraîne l'égalité des lois des v.a. réelles $\ln U$ et $\ln(1-U)$. On sait que $\ln U$ est intégrable. On en déduit que $\ln(1-U)$ est aussi intégrable et de même espérance. Finalement, $X = \ln U - \ln(1-U)$ est intégrable comme différence de deux v.a. intégrables et

$$\mathbf{E}X = \mathbf{E}(\ln U) - \mathbf{E}(\ln(1-U)) = 0.$$

Ce résultat est conforme à la symétrie de la loi de X. Cette observation est générale. Si une variable aléatoire X a une loi symétrique (c.-à-d. X et $-X$ ont même loi) et si X est intégrable, $\mathbf{E}X = 0$ (pourquoi ?). La condition d'intégrabilité est indispensable ici car il existe des lois symétriques n'ayant pas d'espérance, par exemple la loi de Cauchy standard de densité $t \mapsto \frac{1}{\pi(1+t^2)}$.

Solution 7.10 *Une somme doublement aléatoire*

Sur le même espace probabilisé $(\Omega, \mathcal{F}, P)$, on suppose définies
- une variable aléatoire N à valeurs dans $\mathbb{N}$,
- une suite de variables aléatoires positives $(X_i)_{i\geqslant 1}$, ayant toutes même loi.

On pose alors $S_0 := 0$ et pour $n \geqslant 1$, $S_n = \sum_{i=1}^{n} X_i$. On définit la variable aléatoire

$$T := \sum_{i=1}^{N} X_i, \tag{11.33}$$

autrement dit, pour tout $\omega \in \Omega$, $T(\omega) = S_{N(\omega)}(\omega)$. On suppose de plus que N est *indépendante de la suite* $(X_i)_{i\geqslant 1}$, ce qui implique notamment que pour tout $j \in \mathbb{N}$ et tous boréliens B et B' de $\mathbb{R}$, les évènements $\{S_j \in B\}$ et $\{N \in B'\}$ sont indépendants.

1) L'énoncé demande d'examiner le raisonnement erroné suivant. *Par additivité*

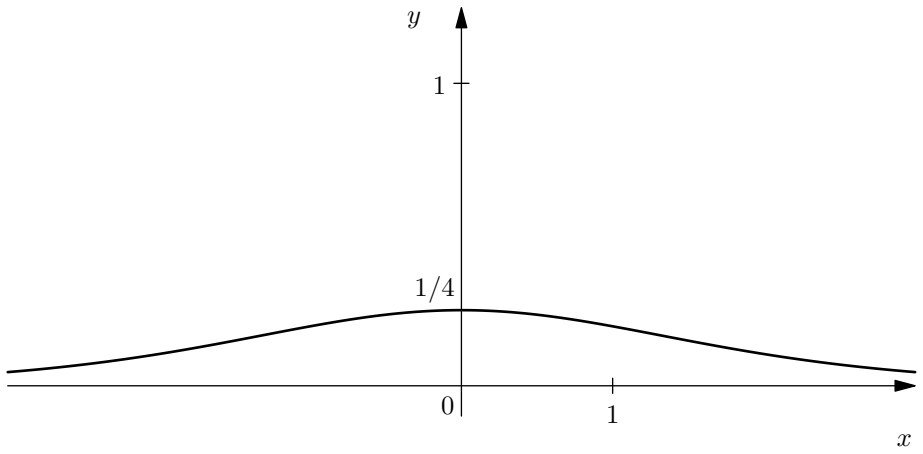


FIGURE 11.10 – Densité f de la v.a. $X = \ln\left(\frac{1-U}{U}\right)$

de l'espérance des variables aléatoires positives,

$$\mathbf{E}T = \mathbf{E}\left(\sum_{i=1}^{N} X_i\right) = \sum_{i=1}^{N} \mathbf{E}X_i = N\mathbf{E}X_1.$$

Comme la variable aléatoire T est positive, $\mathbf{E}T$ existe bien comme élément de $\overline{\mathbb{R}}_+$. Cette espérance ne peut pas être égale à $N\mathbf{E}X_1$ car cette dernière expression est une variable aléatoire, en général non constante, alors que $\mathbf{E}T$ est une constante dépendant seulement de la loi de T. L'erreur vient d'une utilisation abusive de l'additivité de l'espérance. Il est exact que si n est un entier *fixé* et les $X_1, \ldots, X_n$ sont des v.a. positives $\mathbf{E}(\sum_{i=1}^n X_i) = \sum_{i=1}^n \mathbf{E}X_i$. Mais T *n'est pas* la somme d'un nombre fixé de variables aléatoires puisque le nombre de termes dans $T(\omega)$ est $N(\omega)$ qui peut *varier* avec ω.

2) Soit $A \in \mathcal{F}$ un évènement et Y une variable aléatoire sur $(\Omega, \mathcal{F}, P)$ tels que pour tout $t \geqslant 0$, les évènements $\{Y > t\}$ et A soient indépendants. Montrons que

$$\forall t \geqslant 0, \quad P(Y\mathbf{1}_A > t) = P(Y > t)P(A). \tag{11.34}$$

0n part de la décomposition

$$P\big(Y\mathbf{1}_A > t\big) = P\big(\{Y\mathbf{1}_A > t\} \cap A\big) + P\big(\{Y\mathbf{1}_A > t\} \cap A^c\big). \tag{11.35}$$

On remarque alors que pour tout $t \geqslant 0$, $\{Y\mathbf{1}_A > t\} \cap A^c = \varnothing$. En effet si $\omega \in A^c$, $Y(\omega)\mathbf{1}_A(\omega)$ est nul et ne peut donc être *strictement* supérieur à t qui est lui même positif ou nul. Par conséquent

$$\forall t \geqslant 0, \quad P\big(\{Y\mathbf{1}_A > t\} \cap A^c\big) = 0. \tag{11.36}$$

D'autre part, pour tout $t \geqslant 0$, $\{Y\mathbf{1}_A > t\} \cap A = \{Y > t\} \cap A$. En effet si $\omega \in A$, $\mathbf{1}_A(\omega) = 1$ et donc *sur* A les variables aléatoires $Y\mathbf{1}_A$ et Y sont égales. En utilisant l'indépendance des évènements A et $\{Y > t\}$, on en déduit que :

$$\forall t \geqslant 0, \quad P\big(\{Y\mathbf{1}_A > t\} \cap A\big) = P\big(\{Y > t\} \cap A\big) = P(Y > t)P(A). \tag{11.37}$$

L'égalité (11.34) découle maintenant de (11.35), (11.36) et (11.37).

3) Considérons maintenant la variable aléatoire positive $S_j\mathbf{1}_{\{N=j\}}$. Par définition, son espérance est

$$\mathbf{E}\big(S_j\mathbf{1}_{\{N=j\}}\big) = \int_0^{+\infty} P\big(S_j\mathbf{1}_{\{N=j\}} > t\big)\,\mathrm{d}t.$$

En appliquant (11.34) avec $Y = S_j$ et $A = \{N = j\}$, on en déduit

$$\mathbf{E}\big(S_j\mathbf{1}_{\{N=j\}}\big) = \int_0^{+\infty} P(S_j > t)P(N = j)\,\mathrm{d}t.$$

Dans cette intégrale, $P(N = j)$ est constante relativement à la variable d'intégration t et peut donc « sortir » de l'intégrale :

$$\mathbf{E}\big(S_j\mathbf{1}_{\{N=j\}}\big) = P(N = j)\int_0^{+\infty} P(S_j > t)\,\mathrm{d}t = P(N = j)\mathbf{E}S_j.$$

La variable aléatoire S_j est somme d'un nombre fixé j de v.a. positives X_i, donc par additivité de l'espérance $\mathbf{E}S_j = \mathbf{E}X_1 + \cdots + \mathbf{E}X_j$ et comme les X_i ont même loi, elles ont même espérance, d'où $\mathbf{E}S_j = j\mathbf{E}X_1$. On a supposé implicitement dans ce calcul que $j \geqslant 2$, mais les cas $j = 0$ ou 1 sont évidents directement. Finalement nous avons établi que

$$\forall j \in \mathbb{N}, \quad \mathbf{E}\big(S_j\mathbf{1}_{\{N=j\}}\big) = P(N = j)j\mathbf{E}X_1. \tag{11.38}$$

4) Pour tout entier $n \geqslant 1$, on pose $T_n := T\mathbf{1}_{\{N\leqslant n\}}$. De la décomposition en union d'évènements deux à deux disjoints

$$\{N \leqslant n\} = \bigcup_{j=0}^{n} \{N = j\},$$

on déduit

$$\mathbf{1}_{\{N\leqslant n\}} = \sum_{j=0}^{n} \mathbf{1}_{\{N=j\}},$$

puis

$$T_n = \sum_{j=0}^{n} T\mathbf{1}_{\{N=j\}}.$$

Vérifions maintenant que $T\mathbf{1}_{\{N=j\}} = S_j\mathbf{1}_{\{N=j\}}$. En effet si $\omega \notin \{N = j\}$, alors $\mathbf{1}_{\{N=j\}}(\omega) = 0$, d'où $T(\omega)\mathbf{1}_{\{N=j\}}(\omega) = 0 = S_j(\omega)\mathbf{1}_{\{N=j\}}(\omega)$. Si $\omega \in \{N = j\}$, alors

$\mathbf{1}_{\{N=j\}}(\omega) = 1$ et $N(\omega) = j$, $T(\omega) = S_j(\omega)$, d'où $T(\omega)\mathbf{1}_{\{N=j\}}(\omega) = S_j(\omega)\mathbf{1}_{\{N=j\}}(\omega)$. On en déduit l'égalité

$$T_n = \sum_{j=0}^{n} S_j \mathbf{1}_{\{N=j\}}. \tag{11.39}$$

En prenant l'espérance de cette somme d'un nombre fixé n de v.a. positives et compte-tenu de (11.38), on en déduit :

$$\forall n \in \mathbb{N}^*, \quad \mathbf{E}T_n = \mathbf{E}X_1 \sum_{j=0}^{n} jP(N = j). \tag{11.40}$$

5) Vérifions que la suite de variables aléatoires positives $(T_n)_{n\geqslant 1}$ converge en croissant vers T. La croissance résulte de l'inclusion $\{N \leqslant n\} \subset \{N \leqslant n+1\}$ qui implique l'inégalité $\mathbf{1}_{\{N\leqslant n\}} \leqslant \mathbf{1}_{\{N\leqslant n+1\}}$. Pour établir la convergence de T_n vers T on montre que $T_n(\omega) = T(\omega)$ pour n assez grand, plus précisément pour tout $n \geqslant N(\omega)$ (noter que N est à valeurs dans $\mathbb{N}$, donc finies). Rappelons que $T(\omega) = S_{N(\omega)}(\omega)$. D'autre part, pour $n \geqslant N(\omega)$, $T_n(\omega) = S_{N(\omega)}(\omega)$ car $T_n(\omega) = \sum_{j=0}^{n} S_j(\omega)\mathbf{1}_{\{N=j\}}(\omega)$ par (11.39). Or $N(\omega) \in [\![0, n]\!]$ et à cause des indicatrices, tous les termes d'indice $j \neq N(\omega)$ sont nuls.

Par le théorème de Beppo Levi, la convergence croissante de la suite de variables aléatoires positives T_n vers T implique la convergence croissante de $\mathbf{E}T_n$ vers $\mathbf{E}T$. En passant à la limite quand n tend vers $+\infty$ dans (11.40), on en déduit :

$$\mathbf{E}T = \mathbf{E}X_1 \sum_{j=0}^{+\infty} jP(N = j) = (\mathbf{E}X_1)(\mathbf{E}N). \tag{11.41}$$

6) On peut retrouver (11.41) à partir de (11.38) en commençant par établir directement la formule

$$T = \sum_{j\in\mathbb{N}} S_j \mathbf{1}_{\{N=j\}} \tag{11.42}$$

Pour cela on compare les valeurs prises par les deux membres en un ω quelconque. D'abord, $T(\omega) = S_{N(\omega)}(\omega)$. D'autre part

$$\sum_{j\in\mathbb{N}} S_j(\omega)\mathbf{1}_{\{N=j\}}(\omega) = \sum_{j\neq N(\omega)} S_j(\omega)\mathbf{1}_{\{N=j\}}(\omega) + S_{N(\omega)}(\omega) = 0 + S_{N(\omega)}(\omega),$$

car pour tout $j \neq N(\omega)$, ω n'appartient pas à $\{N = j\}$ donc $\mathbf{1}_{\{N=j\}}(\omega) = 0$. Les deux membres de (11.42) prennent donc la même valeur $S_{N(\omega)}(\omega)$ en tout $\omega \in \Omega$, ce qui établit (11.42). Ainsi T apparaît comme la somme d'une série de variables aléatoires positives et cette série converge sur tout Ω (puisque $T(\omega)$ est fini pour tout ω). Par le corollaire « interversion série-espérance » du théorème de Beppo Levi, on en déduit :

$$\mathbf{E}T = \sum_{j\in\mathbb{N}} \mathbf{E}\big(S_j \mathbf{1}_{\{N=j\}}\big).$$

En utilisant (11.38), on en déduit que

$$\mathbf{E}T = \sum_{j\in\mathbb{N}} P(N=j) j \mathbf{E}X_1 = \mathbf{E}X_1 \sum_{j\in\mathbb{N}} jP(N=j) = (\mathbf{E}X_1)(\mathbf{E}N).$$

Solution 7.12 *Espérance d'une loi discrète pathologique*

Comme X est une variable aléatoire discrète positive et $X(\Omega) = \mathbb{D}$, l'espérance $\mathbf{E}X$ existe si la série $\sum_{d\in\mathbb{D}} d\,P(X=d)$ converge et $\mathbf{E}X$ vaut alors la somme de cette série.

La règle de sommation par paquets des séries à termes *positifs* (convergentes ou non dans $\mathbb{R}_+$) nous permet d'écrire

$$\sum_{d\in\mathbb{D}} d\,P(X=d) = \sum_{n=0}^{+\infty} \sum_{d\in\mathbb{D}_n} d\,P(X=d).$$

Pour tous les d dans D_n, les $P(X=d)$ ont la même valeur donnée par (11.13) ; compte-tenu de (11.15), ceci nous permet de voir que la convergence de la série équivaut à celle de la série géométrique de raison q $(0<q<1)$ et de calculer sa somme :

$$\begin{aligned}
\mathbf{E}X &= \sum_{n=1}^{+\infty} \frac{pq^{n-1}}{(10-q)10^{n-1}} \times \frac{9}{2} \times 10^{n-1} \\
&= \sum_{n=1}^{+\infty} \frac{9p}{2(10-q)} q^{n-1} \\
&= \frac{9p}{2(10-q)} \frac{1}{1-q}.
\end{aligned}$$

En se souvenant que $q = 1-p$, on obtient :

$$\mathbf{E}X = \frac{9}{2(10-q)} = \frac{9}{2(9+p)}. \tag{11.43}$$

La fonction $p \mapsto (9+p)^{-1}$ étant strictement décroissante sur $[0,1]$, on a l'encadrement

$$\frac{9}{20} < \mathbf{E}X < \frac{1}{2}.$$

Pour interpréter cet encadrement, considérons d'abord le cas limite $p=1$. Dans ce cas il n'y a que des boules rouges dans la première urne et on effectue un seul tirage dans la deuxième. La loi de X est alors la loi uniforme sur $\Delta_1 := \{k/10\ ;\ 0 \leqslant k \leqslant 9\} = \mathbb{D}_0 \cup \mathbb{D}_1$ son espérance est

$$\mathbf{E}X = \sum_{k=0}^{9} \frac{k}{10} \times \frac{1}{10} = \frac{1}{10^2} \sum_{k=0}^{9} k = \frac{1}{10^2} \times \frac{10(0+9)}{2} = \frac{9}{20} = 0{,}45 \qquad (p=1).$$

Cette espérance est le barycentre du système de masses ponctuelles de cette loi uniforme sur Δ_1. Si on avait considéré la loi uniforme sur $\Delta_1 \cup \{1\}$, on aurait trouvé comme barycentre 1/2. Ce décalage vers la gauche du barycentre se retrouve quand on fait n tirages dans la deuxième urne. La loi du nombre décimal aléatoire Z_n de $\Delta_n = \{k10^{-n}\ ;\ 0 \leqslant k < 10^n\} = \bigcup_{j=0}^{n} \mathbb{D}_j$ ainsi fabriqué a pour espérance

$$\mathbf{E}Z_n = \sum_{k=0}^{10^n-1} \frac{k}{10^n} \times \frac{1}{10^n} = \frac{1}{10^{2n}} \frac{10^n(10^n-1)}{2} = \frac{1}{2}\Big(1 - \frac{1}{10^n}\Big).$$

On observe que ce décalage du barycentre vers la gauche s'atténue quand n augmente. Notons Q_n la loi de Z_n. Comme on l'a vu à la question 2) de l'exercice 6.17 (corrigé p. 456), on a

$$Q_n(\{d\}) = P(Z_n = d) = P(X = d \mid N = n), \qquad d \in \Delta_n.$$

On peut donc dire que Q_n est aussi la *loi conditionnelle* de X sachant $\{N = n\}$, c'est-à-dire la mesure image [14] par X de la probabilité $P(\ .\ \mid N = n)$. La formule des probabilités totales nous dit que pour tout $A \in \mathrm{Bor}(\mathbb{R})$,

$$P_X(A) = P(X \in A) = \sum_{n\in\mathbb{N}*} P(X \in A \mid N = n)P(N = n) = \sum_{n\in\mathbb{N}*} Q_n(A)P(N = n).$$

On voit ainsi que la loi P_X de X est un *mélange* des lois Q_n avec pour coefficients les $P(N = n)$, ce qui se traduit par l'égalité de mesures :

$$P_X = \sum_{n\in\mathbb{N}*} P(N = n)Q_n.$$

À partir de cette formule, il est facile de vérifier que

$$\mathbf{E}X = \sum_{n=1}^{+\infty} P(N = n)\mathbf{E}Z_n. \tag{11.44}$$

Si l'objectif du problème avait seulement été le calcul de $\mathbf{E}X$, cette méthode aurait même été plus simple que celle exposée ci-dessus.

Avec (11.44), on comprend que le barycentre de la loi de X est d'autant plus proche de 1/2 que les valeurs de $\mathbf{E}Z_n$ pour n grand ont un poids plus élevé, autrement dit que $P(N = n)$ tend plus lentement vers 0 quand n tend vers $+\infty$. Et bien sûr, cette convergence est d'autant plus lente que q est plus grand donc p plus petit.

Considérons enfin le cas limite $p = 0$. Dans ce cas on n'obtient jamais de boule rouge et on ferait une infinité de tirages dans la première urne et donc on devrait faire une infinité de tirages avec remise dans la deuxième. Passant sur l'impossibilité

14. On considère X comme une variable aléatoire sur l'espace probabilisé $(\Omega, \mathcal{F}, P(\ .\ \mid N = n))$ au lieu de $(\Omega, \mathcal{F}, P)$.

physique de cette expérience [15], on peut quand même *définir* mathématiquement la variable aléatoire X dans ce cas par

$$X = \sum_{j=1}^{+\infty} \frac{Y_j(\omega)}{10^j}.$$

Cette série converge pour tout ω. La loi de X est alors la loi uniforme sur $[0,1[$ (lire à ce sujet [14, Sect. 6.5]) et son espérance ou barycentre est $1/2$.

Solution 7.13 *Transformée de Laplace*

La transformée de Laplace d'une variable aléatoire positive X est la fonction L_X définie sur $\mathbb{R}_+$ par :

$$\forall t \in \mathbb{R}_+, \quad L_X(t) = \mathbf{E}\exp(-tX).$$

Clairement L_X ne dépend que de la loi de X. On parlera donc aussi bien de la transformée de Laplace d'une variable que de sa loi.

Remarque simplificatrice. On convient une fois pour toutes que la *positivité* de la variable aléatoire X, signifie que $X(\omega) \geqslant 0$ *pour tout* $\omega \in \Omega$. Comme X n'intervient que *via* des espérances de la forme $\mathbf{E}h(X)$, les résultats obtenus dans cet exercice s'étendent facilement au cas d'une variable aléatoire X p.s. positive.

1) Pour tout $t \in \mathbb{R}$, $\exp(-tX)$ est une variable aléatoire positive, donc son espérance existe dans $\overline{\mathbb{R}}_+$. De plus comme X est une v.a. positive, $0 \leqslant \exp(-tX) \leqslant 1$, pour tout $t \in \mathbb{R}_+$. Dans ce cas, la variable $\exp(-tX)$ est bornée, donc intégrable et $L_X(t)$ est un nombre réel positif. La fonction L_X définit donc bien une application de $\mathbb{R}_+$ dans $\mathbb{R}_+$.

Remarque. Pour certaines variables aléatoires positives, l'ensemble de définition de la fonction L_X peut être plus grand que $[0,+\infty[$. Par exemple, si X suit une loi exponentielle de paramètre a, L_X est définie sur $]-a,+\infty[$.

Si X suit la loi de Bernoulli de paramètre p, on a pour toute fonction h définie au moins sur $\{0,1\}$,

$$\mathbf{E}h(X) = h(0)P(X=0) + h(1)P(X=1) = (1-p)h(0) + ph(1).$$

En appliquant ceci avec $h(x) = \exp(-tx)$ (avec t fixé), on obtient

$$L_X(t) = \mathbf{E}\exp(-tX) = 1 - p + p\exp(-t).$$

Cette formule est valide pour tout t réel (notez que si X suit une loi de Bernoulli, $\exp(-tX)$ est une variable aléatoire positive bornée par $\exp(|t|)$). Ceci nous fournit un exemple de transformée de Laplace définie sur $\mathbb{R}$ tout entier. De plus il est clair que L_X

15. On peut imaginer qu'il y ait deux expérimentateurs, le premier tirant les boules vertes et le second procédant à un nouveau tirage avec remise d'une boule numérotée chaque fois qu'une boule verte sort de la première urne, ainsi on n'attend pas un temps infini la première sortie d'une rouge pour commencer à construire X et on peut considérer que l'on réalise X par passage à la limite.

est décroissante et convexe et a pour limite 1 quand t tend vers 0 et $1-p = P(X=0)$ quand t tend vers $+\infty$.

2) Revenant au cas général d'une v.a. positive X quelconque, la décroissance de L_X se vérifie comme suit. On note d'abord que pour tout réel $x \geqslant 0$, la fonction $t \mapsto \exp(-xt)$ est décroissante sur $\mathbb{R}_+$. Par conséquent si $0 \leqslant s < t$, l'inégalité $\exp(-sX(\omega)) \geqslant \exp(-tX(\omega))$ est vérifiée pour tout $\omega \in \Omega$. On peut donc la réécrire comme l'inégalité entre variables aléatoires $\exp(-sX) \geqslant \exp(-tX)$. Par croissance de l'espérance, on en déduit $\mathbf{E}\exp(-sX) \geqslant \mathbf{E}\exp(-tX)$, autrement dit $L_X(s) \geqslant L_X(t)$. Ceci étant valide pour tous réels positifs s et t tels que $s < t$, la décroissance de la fonction L_X sur $\mathbb{R}_+$ est établie.

3) Vérifions la convexité de L_X sur $\mathbb{R}_+$. On rappelle que f est convexe sur $\mathbb{R}_+$ si pour tout $u \in [0,1]$, et tous $s,t \in \mathbb{R}_+$, $f(us+(1-u)t) \leqslant uf(s)+(1-u)f(t)$. Il est bien connu que la fonction exponentielle est convexe sur $\mathbb{R}$. On en déduit immédiatement que la fonction $t \mapsto \exp(-xt)$ est convexe sur $\mathbb{R}$ pour tout réel x (le fait que la corde soit toujours au dessus de la courbe est une propriété géométrique qui n'est pas affectée par le changement de paramétrage $t \mapsto -xt$ en abscisse). On en déduit que pour tout $u \in [0,1]$, et tous $s,t \in \mathbb{R}_+$,

$$\forall \omega \in \Omega, \quad \exp\big(-(us+(1-u)t)X(\omega)\big) \leqslant u\exp\big(-sX(\omega)\big) + (1-u)\exp\big(-tX(\omega)\big).$$

On peut réécrire ceci comme inégalité entre variables aléatoires

$$\exp\big(-\big(us+(1-u)t\big)X\big) \leqslant u\exp(-sX) + (1-u)\exp(-tX),$$

prendre l'espérance des deux membres et en déduire par croissance et linéarité de l'espérance :

$$\begin{aligned}\mathbf{E}\exp\big(-\big(us+(1-u)t\big)X\big) &\leqslant \mathbf{E}\big(u\exp(-sX) + (1-u)\exp(-tX)\big)\\ &= u\mathbf{E}\exp(-sX) + (1-u)\mathbf{E}\exp(-tX).\end{aligned}$$

Comme $s,t \in \mathbb{R}_+$ et $u \in [0,1]$ étaient quelconques, la convexité de L_X est vérifiée.

4) La convexité de L_X sur $\mathbb{R}_+$ implique sa continuité sur $]0,+\infty[$. Pour montrer la continuité de L_X au point 0, il suffit, compte-tenu de la monotonie de L_X, de montrer que $L_X(1/n)$ tend vers $L_X(0) = 1$ quand n tend vers l'infini. Justifions cette convergence à l'aide du théorème de B. Levi.

L'image d'une suite décroissante de réels par une fonction décroissante est une suite croissante de réels. En appliquant ceci avec la suite $(1/n)_{n\geqslant 1}$ et la fonction $t \mapsto \exp(-xt)$ (avec $x \geqslant 0$ quelconque mais fixé), on voit que la suite de réels $(\exp(-x/n))_{n\geqslant 1}$ est croissante pour tout $x \in \mathbb{R}_+$. En prenant $x = X(\omega)$ avec ω quelconque, on en déduit que la suite de variables aléatoires $(Y_n)_{n\geqslant 1} = (\exp(-X/n))_{n\geqslant 1}$ est *croissante*. Elle converge vers la v.a. constante $\exp(0) = 1$ par continuité de l'exponentielle. Comme ces variables aléatoires sont aussi positives, le théorème de B. Levi nous dit que $\mathbf{E}Y_n$ converge en croissant vers $\mathbf{E}(\lim_{n\to+\infty} Y_n) = \mathbf{E}1 = 1$. Ainsi $L_X(1/n)$ converge vers $1 = L_X(0)$ quand n tend vers l'infini. Et par monotonie de L_X, on en déduit que $L_X(t)$ converge vers $L_X(0)$ quand t tend vers 0. Ainsi L_X est bien continue au point 0.

Remarque. Comme $0 \leqslant Y_n \leqslant 1$, on pouvait aussi appliquer le théorème de convergence dominée (th. 9.20) au lieu de celui de B. Levi.

5) Par une argumentation similaire à celle de la question précédente, on obtient la décroissance de la suite de variables aléatoires $Y_n = \exp(-nX)$. La limite de $Y_n(\omega)$ lorsque n tend vers l'infini s'obtient en distinguant deux cas.
Cas 1. Si $X(\omega) > 0$, $-nX(\omega)$ tend vers $-\infty$ et son exponentielle $Y_n(\omega)$ tend vers 0.
Cas 2. Si $X(\omega) = 0$, $Y_n(\omega) = \exp(0) = 1$ pour tout n et la limite est trivialement 1.

On voit ainsi que la suite de variables aléatoires $(Y_n)_{n\geqslant 1}$ converge sur tout Ω vers $\mathbf{1}_{\{X=0\}}$. On ne peut pas utiliser directement le théorème de B. Levi pour intervertir limite et espérance, puisque ce théorème concerne les suites *croissantes* de variables aléatoires positives. Néanmoins puisque $0 \leqslant Y_n \leqslant 1$, en posant $Z_n = 1 - Y_n$, on retrouve une suite croissante de v.a. positives et en lui appliquant le théorème de B. Levi,

$$1 - \mathbf{E}Y_n = \mathbf{E}Z_n \uparrow \mathbf{E}(1 - \mathbf{1}_{\{X=0\}}) = 1 - \mathbf{E}\mathbf{1}_{\{X=0\}},$$

on obtient la convergence de $\mathbf{E}Y_n$ vers $\mathbf{E}\mathbf{1}_{\{X=0\}} = P(X = 0)$. On a donc montré la convergence de $L_X(n)$ vers $P(X = 0)$ quand n tend vers l'infini. En raison de la monotonie de L_X on en déduit que $L_X(t)$ tend vers $P(X = 0)$ quand t tend vers $+\infty$.

Là encore, on pouvait appliquer alternativement le théorème de convergence dominée pour obtenir la convergence de $\mathbf{E}Y_n$.

Commentaire sur l'interversion limite espérance pour une suite décroissante. Dans cette question, comme dans la précédente, l'utilisation de la monotonie de la suite (Y_n) était un chemin moins rapide que le théorème de convergence dominée. Ce détour est l'occasion d'attirer votre attention sur le caractère non symétrique du théorème de B. Levi dont l'énoncé devient faux si on se contente de remplacer la suite croissante de v.a. positives par une suite décroissante de v.a. positives. Voici un contre-exemple. On prend X suivant la loi de Cauchy et $Y_n = |X|\mathbf{1}_{\{|X|>n\}}$. Alors (Y_n) est décroissante et a pour limite 0, mais $\mathbf{E}Y_n = +\infty$ pour tout n et ne peut donc converger vers $\mathbf{E}0 = 0$.

6) Si X et Y sont des variables aléatoires indépendantes, leurs images respectives par des applications boréliennes g, h sont des variables aléatoires $g(X)$ et $h(Y)$ indépendantes. En choisissant pour g et h l'application continue donc borélienne $x \mapsto \exp(-tx)$, avec t fixé, on obtient l'indépendance des variables aléatoires positives $\exp(-tX)$ et $\exp(-tY)$. Ceci légitime le calcul suivant :

$$\begin{aligned} L_{X+Y}(t) = \mathbf{E}\exp\big(-t(X+Y)\big) &= \mathbf{E}\big(\exp(-tX)\exp(-tY)\big) \\ &= \mathbf{E}\exp(-tX)\mathbf{E}\exp(-tY) = L_X(t)L_Y(t). \end{aligned}$$

Comme ce calcul est valide pour tout $t \in \mathbb{R}_+$, il établit l'égalité fonctionnelle $L_{X+Y} = L_X L_Y$ vraie pour toute paire de v.a. positives et indépendantes X et Y.

Solution 7.15 *Moments fonctionnels et décroissance lente*

On utilise la formule (7.65) de calcul de moments fonctionnels pour apporter une réponse partielle au problème suivant. Quelle que soit la vitesse de convergence vers 0 de la fonction de survie G de la v.a. X, est-il possible de trouver un moment fonctionnel $\mathbf{E}h(X)$ fini avec h tendant vers $+\infty$ en $+\infty$? Les deux questions suivantes

sont consacrées à l'étude d'un exemple où G tend « lentement » vers 0, afin de se faire une intuition.

1) Soit a un réel. Considérons l'intégrale de Riemann ordinaire

$$I(a,x) := \int_2^x \frac{\mathrm{d}t}{t(\ln t)^a}$$

Notons $f : t \mapsto t^{-1}(\ln t)^{-a}$ l'intégrande. En notant que $f(t)$ peut s'écrire $(\ln t)^{-a}(\ln t)'$, on voit qu'une primitive de la fonction continue f est F donnée par

$$F(t) = \begin{cases} \dfrac{1}{(1-a)(\ln t)^{a-1}} & \text{si } a \neq 1, \\ \ln(\ln t) & \text{si } a = 1. \end{cases}$$

On en déduit que quand x tend vers $+\infty$, $F(x)$ tend vers $+\infty$ si $a \leqslant 1$ et vers 0 si $a > 1$. Comme $I(a,x) = F(x) - F(2)$, on voit ainsi que l'intégrale généralisée $I(a)$ converge si et seulement si $a > 1$.

2) Soit X une variable aléatoire positive de fonction de survie G définie par

$$\forall t \in \mathbb{R}, \qquad G(t) = \begin{cases} 1 & \text{si } t < 0, \\ \dfrac{1}{\ln(\mathrm{e}+t)} & \text{si } t \geqslant 0. \end{cases}$$

a) En appliquant la formule (7.65) avec $g(x) = x^\varepsilon$ et $\varepsilon > 0$ quelconque, on obtient :

$$\mathbf{E}X^\varepsilon = \int_0^{+\infty} \varepsilon t^{\varepsilon-1} \frac{1}{\ln(\mathrm{e}+t)}\,\mathrm{d}t.$$

Cette intégrale généralisée d'une fonction *positive* diverge, car lorsque t tend vers $+\infty$, l'intégrande est équivalente [16] à $\varepsilon t^{\varepsilon-1}(\ln t)^{-1}$. Or pour $\varepsilon > 1$, cet équivalent tend vers $+\infty$ en $+\infty$ et pour $0 < \varepsilon \leqslant 1$, il tend plus lentement vers 0 que $1/t$ dont l'intégrale généralisée diverge. Comme l'intégrande est positive, l'intégrale généralisée divergente vaut $+\infty$ dans les deux cas. On a ainsi vérifié que $\mathbf{E}X^\varepsilon = +\infty$ pour tout $\varepsilon > 0$.

b) La fonction $g(x) = (\ln(1+x))^b$ tend vers $+\infty$ pour $b > 0$, mais beaucoup plus lentement que x^ε. En appliquant à nouveau la formule (7.65), il vient :

$$\mathbf{E}(\ln(1+X))^b = \int_0^{+\infty} b\big(\ln(1+t)\big)^{b-1} \frac{\mathrm{d}t}{(1+t)\ln(\mathrm{e}+t)}.$$

L'intégrande est continue et positive, équivalente à $bt^{-1}(\ln t)^{b-2}$ en $+\infty$. Donc en appliquant le résultat de la question 1 avec $a = 2-b$, on voit que l'intégrale généralisée ci-dessus converge pour $2-b > 1$, autrement dit pour $b < 1$. Ainsi pour $0 < b < 1$, $g(x) = (\ln(1+x))^b$ tend vers $+\infty$ avec x et $\mathbf{E}(\ln(1+X))^b < +\infty$.

16. On peut trancher la question de la convergence en prenant un équivalent car l'intégrande est positive. Notons aussi que $\ln(\mathrm{e}+t) = \ln((1+\mathrm{e}/t)t) = \ln(1+\mathrm{e}/t) + \ln t = (1+o(1))\ln t$, d'où $\ln(\mathrm{e}+t) \sim \ln t$.

3) On suppose plus généralement que X est une variable aléatoire positive dont la fonction de survie G est C^1 sur $]0,+\infty[$ et strictement décroissante sur $[0,+\infty[$. Ceci implique en particulier que $G(t)$ est strictement positif pour tout $t \in [0,+\infty[$. En effet si $G(t_0) = 0$ pour un $t_0 \in [0,+\infty[$, alors par décroissance et positivité de G, pour tout $t \geqslant t_0$, $0 \leqslant G(t) \leqslant G(t_0) = 0$, donc G serait constante nulle sur tout l'intervalle $[t_0,+\infty[$, ce qui contredirait sa décroissance *stricte*. Comme la fonction $y \mapsto y^{-1/2}$ est C^1 sur $]0,+\infty[$, on en déduit par composition que $G^{-1/2}$ est C^1 sur $]0,+\infty[$. De plus, $G^{-1/2}$ est strictement croissante comme composée de fonctions strictement décroissantes. Enfin, comme la fonction de survie G a pour limite 0 en $+\infty$, $G^{-1/2}$ tend vers $+\infty$ en $+\infty$. Toutes les conditions de validité de la formule (7.65) étant satisfaites par $g = G^{-1/2}$, on a

$$\begin{aligned}
\mathbf{E}G(X)^{-1/2} &= G(0)^{-1/2} + \int_0^{+\infty} \Big(G(s)^{-1/2}\Big)' P(X > s)\,\mathrm{d}s \\
&= G(0)^{-1/2} + \int_0^{+\infty} \frac{-1}{2} G(s)^{-3/2} G'(s) G(s)\,\mathrm{d}s \\
&= G(0)^{-1/2} + \int_0^{+\infty} \frac{-1}{2} G(s)^{-1/2} G'(s)\,\mathrm{d}s \\
&= G(0)^{-1/2} + \lim_{x\to+\infty} \Big[-G(s)^{1/2}\Big]_0^x \\
&= G(0)^{-1/2} + G(0)^{1/2}.
\end{aligned}$$

On voit ainsi que la variable aléatoire positive $G(X)^{-1/2}$ a une espérance finie. Ceci montre l'existence d'une fonction positive h, strictement croissante sur $[0,+\infty[$, tendant vers $+\infty$ en $+\infty$ et telle que $\mathbf{E}h(X) < +\infty$.

La variable aléatoire X considérée dans cette question n'est pas forcément à densité, car on n'a pas eu besoin de supposer que G est continue *à gauche* en 0. En fait, compte-tenu des hypothèses faites sur G, X est à densité si et seulement si $G(0) = 1$, puisque la limite à gauche en 0 de la fonction de survie d'une v.a. positive est toujours 1.

La formule pour $\mathbf{E}G(X)^{-1/2}$ peut se réécrire :

$$\mathbf{E}G(X)^{-1/2} = \frac{1}{\sqrt{P(X>0)}} + \sqrt{P(X>0)}.$$

Dans le contexte de cette question, cette quantité vaut 2 si X est à densité et la réciproque est vraie (pourquoi ?).

Remarque. On a pris $h = G^{-1/2}$ pour une raison d'esthétique, mais il devrait être clair si vous avez compris ce qui précède, que $h = G^{-a}$ avec $0 < a < 1$ aurait aussi convenu.

11.8 Exercices du chapitre 8

Solution 8.2 *Usagers et usagères à la poste*

Notons $(\Omega, \mathcal{F}, P)$ un espace probabilisé modélisant cette situation. Comme on ne connaît pas le nombre d'arrivées, cet espace doit contenir une suite infinie d'épreuves répétées (les arrivées individuelles), la k^{e} épreuve ayant deux issues possibles H_k ou F_k avec $P(H_k) = p$, $P(F_k) = q$. Nous ferons l'hypothèse d'indépendance de ces épreuves : en l'absence d'autres informations, la connaissance des sexes d'un ensemble donné d'arrivants ne donne aucun renseignement sur ceux des autres arrivants ; ce ne serait pas le cas si l'on savait, par exemple, que le bureau est visité par des groupes organisés non mixtes.

1) *Calcul de* $P(X = i, Y = j \mid S = n)$ *pour* $i + j = n$. On dispose ici d'un renseignement supplémentaire : le nombre total d'arrivées est n. Cette information ne modifie en rien notre ignorance de la répartition par sexes des arrivants. Il est donc légitime de considérer que pour la *nouvelle probabilité* $P_n := P(\ . \mid S = n)$, les n premières arrivées restent indépendantes avec $P_n(H_k) = p$, $P_n(F_k) = q$ pour $1 \leqslant k \leqslant n$ (bien sûr pour $k > n$, $P_n(H_k) = P_n(F_k) = 0$).

Les i arrivées masculines et les j arrivées féminines peuvent se répartir de C_n^i façons possibles parmi les n arrivées (il suffit de connaître les numéros d'arrivées des i hommes). En raison de la P_n-indépendance des arrivées, chacune de ces répartitions a une P_n-probabilité égale à $p^i q^j$ d'où :

$$P(X = i, Y = j \mid S = n) = P_n(X = i, Y = j) = C_n^i p^i q^j \qquad (i + j = n).$$

Remarques. Comme $n = i + j$, on a

$$\{X = i, Y = j, S = n\} = \{X = i, S = n\} = \{Y = j, S = n\},$$

d'où :

$$\begin{aligned} P(X = i \mid S = n) &= C_n^i p^i q^j = C_n^i p^i q^{n-i}, & 0 \leqslant i \leqslant n, \\ P(Y = j \mid S = n) &= C_n^i p^i q^j = C_n^j p^i q^j = C_n^j q^j p^{n-j}, & 0 \leqslant j \leqslant n. \end{aligned}$$

Ainsi conditionnellement à $\{S = n\}$, X et Y suivent respectivement les lois binomiales $\mathrm{Bin}(n, p)$ et $\mathrm{Bin}(n, q)$. Notons que X et Y ne sont pas P_n-indépendantes : si l'on sait que $\{S = n\}$ est réalisé, $\{X = i\}$ impose $\{Y = n - i\}$.

2) *Loi du vecteur* (X, Y).

Il s'agit de calculer $P(X = i, Y = j)$ pour chaque couple $(i, j) \in \mathbb{N}^2$. Comme le système d'évènements $(\{S = m\}, m \in \mathbb{N})$ forme une partition de Ω,

$$P(X = i, Y = j) = \sum_{m=0}^{+\infty} P(X = i, Y = j \mid S = m) P(S = m).$$

Il est clair que si $m \neq i + j$, $P(X = i, Y = j \mid S = m) = 0$. On en déduit :

$$P(X = i, Y = j) = P(X = i, Y = j \mid S = i + j) P(S = i + j) = C_{i+j}^i p^i q^j \mathrm{e}^{-\alpha} \frac{\alpha^{i+j}}{(i + j)!}.$$

Après simplifications et en notant que $\mathrm{e}^{-\alpha} = \mathrm{e}^{-\alpha(p+q)}$, il vient :

$$\forall (i,j) \in \mathbb{N}^2, \qquad P(X=i, Y=j) = \mathrm{e}^{-\alpha p}\frac{(\alpha p)^i}{i!}\,\mathrm{e}^{-\alpha q}\frac{(\alpha q)^j}{j!}.$$

3) *Lois de X et Y.*
Connaissant la loi du couple, on en déduit les lois marginales :

$$P(X=i) = \sum_{j=0}^{+\infty} P(X=i, Y=j) = \mathrm{e}^{-\alpha p}\frac{(\alpha p)^i}{i!}\sum_{j=0}^{+\infty}\mathrm{e}^{-\alpha q}\frac{(\alpha q)^j}{j!} = \mathrm{e}^{-\alpha p}\frac{(\alpha p)^i}{i!}.$$

Donc X suit la loi Pois(αp). On vérifie de même que la loi de Y est Pois(αq). Le résultat de la deuxième question montre alors que pour tout couple (i,j) de $\mathbb{N}^2$, $P(X=i, Y=j) = P(X=i)P(X=j)$. Les deux variables sont donc indépendantes.

Solution 8.4

1) Pour tous x, y réels, $|\sin \pi x \sin \pi y| \leqslant 1$, donc $1+\sin \pi x \sin \pi y \geqslant 0$ et h est positive sur $\mathbb{R}^2$. D'autre part, on sait que pour des fonctions de la forme $k(x,y) = f(x)g(y)$ avec f et g continues à support compact, $\int_{\mathbb{R}^2} f(x)g(y)\,\mathrm{d}x\mathrm{d}y = \big(\int_{\mathbb{R}} f(x)\,\mathrm{d}x\big)\big(\int_{\mathbb{R}} g(y)\,\mathrm{d}y\big)$. La fonction h proposée par l'énoncé peut s'écrire comme la somme de deux fonctions de ce type puisque $\mathbf{1}_{[-1,+1]^2}(x,y) = \mathbf{1}_{[-1,+1]}(x)\mathbf{1}_{[-1,+1]}(y)$. Il est alors facile de vérifier que l'intégrale de h sur $\mathbb{R}^2$ vaut 1 :

$$\begin{aligned}\int_{\mathbb{R}^2} h(x,y)\,\mathrm{d}x\mathrm{d}y &= \frac{1}{4}\int_{[-1,+1]^2}\mathrm{d}x\mathrm{d}y + \frac{1}{4}\int_{[-1,+1]^2}\sin \pi x \sin \pi y\,\mathrm{d}x\mathrm{d}y\\ &= 1 + \frac{1}{4}\left(\int_{-1}^{+1}\sin \pi x\,\mathrm{d}x\right)\left(\int_{-1}^{+1}\sin \pi y\,\mathrm{d}y\right) = 1.\end{aligned}$$

Par conséquent, h est bien une densité de probabilité sur $\mathbb{R}^2$.

2) Les densités de X et Y se calculent par intégration partielle de la densité du couple. D'autre part, comme la densité du couple est symétrique (domaine de non nullité symétrique et fonction symétrique), il est clair que X et Y ont même loi. Il suffit donc de calculer la densité f_X de la loi de X :

$$\begin{aligned}f_X(x) = \int_{-\infty}^{+\infty} h(x,y)\,\mathrm{d}y &= \frac{1}{4}\int_{-\infty}^{+\infty}(1+\sin \pi x \sin \pi y)\mathbf{1}_{[-1,+1]}(x)\mathbf{1}_{[-1,+1]}(y)\,\mathrm{d}y\\ &= \frac{1}{4}\mathbf{1}_{[-1,+1]}(x)\int_{-1}^{+1}\mathrm{d}y + \frac{1}{4}\mathbf{1}_{[-1,+1]}(x)\sin \pi x\int_{-1}^{+1}\sin \pi y\,\mathrm{d}y\\ &= \frac{1}{2}\mathbf{1}_{[-1,+1]}(x),\end{aligned}$$

ce qui est bien la densité uniforme sur $[-1,+1]$. Donc X et Y suivent la loi uniforme sur $[-1,+1]$. En particulier, par raison de symétrie, X et Y sont centrées : $\mathbf{E}X =$

$\mathbf{E}Y = 0$. Par conséquent, $\mathrm{Cov}(X,Y) = \mathbf{E}(XY) - (\mathbf{E}X)(\mathbf{E}Y) = \mathbf{E}(XY)$ et

$$\begin{aligned}\mathrm{Cov}(X,Y) &= \int_{\mathbb{R}^2} xyh(x,y)\,\mathrm{d}x\mathrm{d}y \\ &= \frac{1}{4}\int_{[-1,+1]^2} xy\,\mathrm{d}x\mathrm{d}y + \frac{1}{4}\int_{[-1,+1]^2} xy\sin\pi x\sin\pi y\,\mathrm{d}x\mathrm{d}y \\ &= \frac{1}{4}\left(\int_{-1}^{+1} x\,\mathrm{d}x\right)\left(\int_{-1}^{+1} y\,\mathrm{d}y\right) + \frac{1}{4}\left(\int_{-1}^{+1} x\sin\pi x\,\mathrm{d}x\right)\left(\int_{-1}^{+1} y\sin\pi y\,\mathrm{d}y\right) \\ &= 0 + \frac{1}{4}\left(\int_{-1}^{+1} x\sin\pi x\,\mathrm{d}x\right)^2.\end{aligned}$$

Cette dernière intégrale se calcule en intégrant par parties :

$$\int_{-1}^{+1} x\sin\pi x\,\mathrm{d}x = \Big[-\frac{x}{\pi}\cos\pi x\Big]_{-1}^{+1} + \int_{-1}^{+1}\frac{1}{\pi}\cos\pi x\,\mathrm{d}x = \frac{2}{\pi}.$$

On obtient finalement $\mathrm{Cov}(X,Y) = \dfrac{1}{\pi^2} \neq 0$, et donc X et Y ne sont pas indépendantes.

Solution 8.6 *Tir à l'arc*

On modélise le point d'impact de la flèche par le vecteur aléatoire (X,Y) de densité :

$$f : (x,y) \longmapsto \frac{1}{2\pi\sigma^2}\exp\left(\frac{-(x^2+y^2)}{2\sigma^2}\right).$$

1) Grâce aux propriétés de l'exponentielle, cette densité s'écrit aussi :

$$f(x,y) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left(\frac{-x^2}{2\sigma^2}\right) \times \frac{1}{\sigma\sqrt{2\pi}}\exp\left(\frac{-y^2}{2\sigma^2}\right) = g(x)g(y),$$

où g est la densité de la loi $\mathfrak{N}(0,\sigma)$. On sait alors, cf. proposition 8.43, que les composantes du vecteur aléatoire sont indépendantes et que la loi de X a pour densité g et celle de Y aussi [17].

2) Pour trouver la densité du vecteur aléatoire (R,U), nous allons appliquer la proposition 8.18 sur la densité d'un vecteur aléatoire image. Pour cela il nous faut calculer le jacobien de φ^{-1}. Le calcul des dérivées partielles de φ^{-1} est immédiat :

$$\frac{\partial x}{\partial r} = \cos u, \quad \frac{\partial y}{\partial r} = \sin u, \quad \frac{\partial x}{\partial u} = -r\sin u, \quad \frac{\partial y}{\partial u} = r\cos u.$$

Le jacobien s'écrit

$$\mathrm{Jac}(\varphi^{-1})(r,u) = \begin{vmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial u} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial u} \end{vmatrix} = \begin{vmatrix} \cos u & -r\sin u \\ \sin u & r\cos u \end{vmatrix} = r(\cos^2 u + \sin^2 u) = r.$$

17. Il est clair que les constantes de proportionnalité c telle que $f_X = cg$ et c' telle que $f_Y = c'g$ sont égales à 1 parce que g est déjà une densité de probabilité.

La densité du couple (R,U) est donc

$$f_{R,U} : (r,u) \longmapsto \frac{r}{2\pi\sigma^2} \exp\left(\frac{-r^2}{2\sigma^2}\right) \mathbf{1}_{]0,+\infty[\times]-\pi,\pi[}(r,u).$$

On remarque que cette densité est de la forme $g_1 \otimes g_2$, c'est-à-dire $f_{R,U}(r,u) = g_1(r)g_2(u))$ avec

$$g_1(r) = \frac{r}{\sigma^2} \exp\left(\frac{-r^2}{2\sigma^2}\right) \mathbf{1}_{]0,+\infty[}(r), \quad g_2(u) = \frac{1}{2\pi}\mathbf{1}_{]-\pi,\pi[}(u).$$

Il en résulte que les variables aléatoires R et U sont indépendantes et que leur loi a une densité proportionnelle à g_1 pour R et à g_2 pour U. En fait comme g_2 est déjà une densité de probabilité (c'est celle de la loi uniforme sur $]-\pi,\pi[$), les constantes de proportionnalité valent 1. En conclusion, R suit la loi de densité g_1 et U suit la loi uniforme sur $]-\pi,\pi[$.

Indications 8.8 *Covariances d'une loi multinomiale*

La v.a. $X_1 + \cdots + X_k$ est constante. Les v.a. X_i et $X_i + X_j$ suivent des lois binomiales.

Solution 8.11 *La méthode de Box Muller*

Le vecteur aléatoire (U_1,U_2) suit la loi uniforme sur $]0,1[^2$, de densité $f_{U_1,U_2} = \mathbf{1}_{]0,1[^2}$. Le changement de variable

$$g : (u_1,u_2) \mapsto (x,y) \quad \text{avec} \quad \begin{cases} x = (-2\ln u_1)^{1/2}\cos(2\pi u_2) \\ y = (-2\ln u_1)^{1/2}\sin(2\pi u_2) \end{cases}$$

réalise un C^1-difféomorphisme de l'ouvert $D =]0,1[^2$ sur l'ouvert $D' = \mathbb{R}^2\backslash\{(0,0)\}$. Par la proposition 8.18 sur la densité d'un vecteur aléatoire image, le vecteur aléatoire $(X,Y) = g(U_1,U_2)$ a une densité $f_{X,Y}$ donnée par

$$f_{X,Y}(x,y) = f_{U_1,U_2}\big(g^{-1}(x,y)\big)\big|\mathrm{Jac}(g^{-1})(x,y)\big|\mathbf{1}_{D'}(x,y).$$

Pour tout $(x,y) \in \mathbb{R}^2\backslash\{(0,0)\}$, $g^{-1}(x,y)$ appartient à $]0,1[^2$ et comme f_{U_1,U_2} vaut 1 sur $]0,1[^2$, le premier facteur dans la formule ci-dessus s'écrit $f_{U_1,U_2}\big(g^{-1}(x,y)\big) = 1$. Pour calculer $\mathrm{Jac}(g^{-1})(x,y)$, on calcule $\mathrm{Jac}(g)(u_1,u_2)$ et on utilise la formule

$$\mathrm{Jac}(g^{-1})(x,y) = \frac{1}{\mathrm{Jac}(g)\big(g^{-1}(x,y)\big)}.$$

En notant $g = (g_1,g_2)$, les dérivées partielles de g sont

$$\frac{\partial g_1}{\partial u_1} = -u_1^{-1}(-2\ln u_1)^{-1/2}\cos(2\pi u_2), \qquad \frac{\partial g_1}{\partial u_2} = -2\pi(-2\ln u_1)^{1/2}\sin(2\pi u_2),$$

$$\frac{\partial g_2}{\partial u_1} = -u_1^{-1}(-2\ln u_1)^{-1/2}\sin(2\pi u_2), \qquad \frac{\partial g_2}{\partial u_2} = 2\pi(-2\ln u_1)^{1/2}\cos(2\pi u_2).$$

On en déduit que

$$\mathrm{Jac}(g)(u_1,u_2) = \begin{vmatrix} \frac{\partial g_1}{\partial u_1} & \frac{\partial g_1}{\partial u_2} \\ \frac{\partial g_2}{\partial u_1} & \frac{\partial g_2}{\partial u_2} \end{vmatrix} = -\frac{2\pi}{u_1}\cos^2(2\pi u_2) - \frac{2\pi}{u_1}\sin^2(2\pi u_2) = -\frac{2\pi}{u_1}.$$

Pour exprimer u_1 en fonction de (x,y), on remarque que $x^2+y^2 = -2\ln u_1$, d'où $u_1 = \exp\big(-(x^2+y^2)/2\big)$ et

$$\big|\mathrm{Jac}(g^{-1})(x,y)\big| = \frac{1}{2\pi}\exp\Big(-\frac{x^2+y^2}{2}\Big).$$

Finalement la densité du couple (X,Y) est donnée par

$$f_{X,Y}(x,y) = \frac{1}{2\pi}\exp\Big(-\frac{x^2+y^2}{2}\Big)\mathbf{1}_{D'}(x,y).$$

Comme $\mathbf{1}_{D'}$ et $\mathbf{1}_{\mathbb{R}^2}$ ne diffèrent que sur $\{(0,0)\}$ qui est un ensemble de mesure nulle, (X,Y) admet aussi pour densité [18] $\tilde{f}_{X,Y}(x,y) = (2\pi)^{-1}\exp\big(-(x^2+y^2)/2\big)$ qui est la forme classique de la densité d'un couple de v.a. i.i.d. $\mathfrak{N}(0,1)$. Donc X et Y sont gaussiennes $\mathfrak{N}(0,1)$ indépendantes.

Solution 8.12 *Loi uniforme sur l'hypographe d'une densité*

Pour identifier la loi de Z lorsque $M=(Z,Y)$ suit la loi uniforme sur G, nous allons calculer $P(Z\in[a,b])$ pour tout intervalle $[a,b]\subset\mathbb{R}$. Notons $G_{a,b}$ l'intersection de l'hypographe G avec la « bande verticale » $[a,b]\times\mathbb{R}$, cf. figure 11.11. On peut aussi voir $G_{a,b}$ comme l'hypographe de la restriction de f à $[a,b]$.

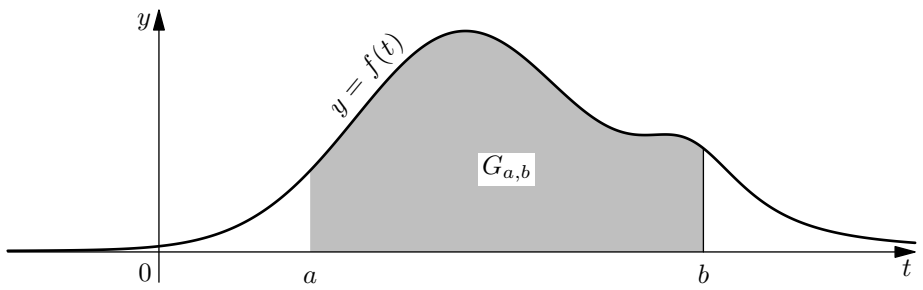


FIGURE 11.11 – Ensemble $G_{a,b}$

On s'appuie sur notre connaissance de la loi de M en notant que

$$P(Z\in[a,b]) = P\big((Z,Y)\in[a,b]\times\mathbb{R}\big) = \frac{\lambda_2\big(G\cap([a,b]\times\mathbb{R})\big)}{\lambda_2(G)} = \lambda_2(G_{a,b}), \quad (11.45)$$

18. On peut remplacer $f_{X,Y}$ par $\tilde{f}_{X,Y}$ dans les intégrales sans changer la valeur de ces intégrales.

car $\lambda_2(G) = \int_{-\infty}^{+\infty} f(t)\,\mathrm{d}t = 1$. L'aire de l'hypographe d'une fonction positive sur $[a,b]$ étant égale à son intégrale de Riemann (éventuellement généralisée) sur $[a,b]$, on a

$$P(Z \in [a,b]) = \lambda_2(G_{a,b}) = \int_a^b f(t)\,\mathrm{d}t.$$

De cette formule valable pour tous réels a et b tels que $a < b$, on déduit que Z a pour densité f.

Solution 8.13 *Simulation d'une loi uniforme sur un hypographe*

Pour identifier la loi de $M := \big(X, cg(X)U\big)$, nous allons calculer $\mathbf{E}h(M)$ pour $h : \mathbb{R}^2 \to \mathbb{R}_+$ continue à support compact quelconque. M étant une fonction du couple (X,U), il est commode d'exprimer $\mathbf{E}h(M)$ à l'aide de la densité de ce couple. Par indépendance de X et U, cette densité s'écrit $(x,u) \mapsto g(x)\mathbf{1}_{[0,1]}(u)$. Ainsi

$$\begin{aligned}\mathbf{E}h(M) &= \int_{\mathbb{R}^2} h\big(x, cg(x)u\big)g(x)\mathbf{1}_{[0,1]}(u)\,\mathrm{d}x\,\mathrm{d}u \\ &= \int_{-\infty}^{+\infty} \frac{1}{c}\left\{\int_0^1 h\big(x, cg(x)u\big)cg(x)\,\mathrm{d}u\right\}\mathrm{d}x. \end{aligned} \tag{11.46}$$

Notons $I(x)$ l'intégrale entre accolades. Si $g(x) > 0$, le changement de variable $y = cg(x)u$ à x fixé nous donne

$$I(x) = \int_0^{cg(x)} h(x,y)\,\mathrm{d}y = \int_{-\infty}^{+\infty} h(x,y)\mathbf{1}_{[0,cg(x)]}(y)\,\mathrm{d}y = \int_{-\infty}^{+\infty} h(x,y)\mathbf{1}_H(x,y)\,\mathrm{d}y. \tag{11.47}$$

Remarquons que si $g(x) = 0$, $[0, cg(x)] = \{0\}$ et $\int_{-\infty}^{+\infty} h(x,y)\mathbf{1}_{[0,cg(x)]}(y)\,\mathrm{d}y = 0$. On peut donc retenir la formule (11.47) comme l'expression générale de $I(x)$, y compris lorsque $g(x) = 0$. Par report de cette expression dans (11.46), on obtient

$$\mathbf{E}h(M) = \int_{\mathbb{R}^2} h(x,y)\frac{1}{c}\mathbf{1}_H(x,y)\,\mathrm{d}x\,\mathrm{d}y = \int_{\mathbb{R}^2} h(x,y)\frac{1}{\lambda_2(H)}\mathbf{1}_H(x,y)\,\mathrm{d}x\,\mathrm{d}y.$$

Cette formule étant vraie pour toute $h : \mathbb{R}^2 \to \mathbb{R}_+$ continue bornée, on en déduit que le vecteur aléatoire M suit la loi de densité $\lambda_2(H)^{-1}\mathbf{1}_H$, c.-à-d. la loi uniforme sur H.

11.9 Exercices du chapitre 9

Solution 9.10 *Tir à l'arc (suite)*

Comme R est une variable aléatoire *positive*, on peut écrire directement sans se soucier de l'intégrabilité et en notant que sa densité g_1 est nulle sur $]-\infty, 0]$:

$$\mathbf{E}R = \int_0^{+\infty} r g_1(r)\,\mathrm{d}r.$$

Grâce à la parité de la fonction à intégrer, on a alors

$$\mathbf{E}R = \frac{1}{2}\int_{-\infty}^{+\infty} \frac{r^2}{\sigma^2}\exp\left(\frac{-r^2}{2\sigma^2}\right)\mathrm{d}r = \frac{\sqrt{2\pi}}{2\sigma}\int_{-\infty}^{+\infty}\frac{1}{\sigma\sqrt{2\pi}}\,r^2\exp\left(\frac{-r^2}{2\sigma^2}\right)\mathrm{d}r = \frac{\sqrt{2\pi}}{2\sigma}\mathbf{E}Z^2,$$

où Z est une variable aléatoire de loi gaussienne [19] $\mathfrak{N}(0,\sigma)$. On voit immédiatement que $\mathbf{E}Z = 0$ et donc $\mathbf{E}Z^2 = \operatorname{Var} Z = \sigma^2$. On obtient ainsi

$$\mathbf{E}R = \sigma\sqrt{\frac{\pi}{2}}, \tag{11.48}$$

ce qui justifie *a posteriori* l'intégrabilité de la variable aléatoire *positive* R puisque $\mathbf{E}R = \mathbf{E}|R| < +\infty$.

La formule (11.48) confirme l'idée que le paramètre σ mesure l'adresse de l'archer. Plus il est petit, plus « en moyenne » le point d'impact sera proche du centre de la cible. L'interprétation de $\mathbf{E}R$ comme valeur moyenne des distances au centre observées sur un grand nombre de tirs se justifie théoriquement par la loi forte des grands nombres. En effet R est intégrable et la loi forte des grands nombres nous dit que si $(R_k)_{k\geqslant 1}$ est une suite de variables aléatoires indépendantes de même loi que R,

$$T_n := \frac{1}{n}\sum_{k=1}^{n} R_k \xrightarrow[n\to+\infty]{\text{p.s.}} \mathbf{E}R.$$

Compte-tenu de (11.48), on en déduit que

$$Z_n := \sqrt{\frac{2}{\pi}}T_n \xrightarrow[n\to+\infty]{\text{p.s.}} \sigma.$$

En pratique, on observe n tirs du même archer et on relève les valeurs numériques des distances au centre $r_1,\ldots,r_n$ des points d'impact de la flèche. On interprète ces valeurs numériques comme les valeurs $r_1 = R_1(\omega_0),\ldots,r_n = R_n(\omega_0)$, pour un même ω_0. On *parie* alors que cet ω_0 (qui ne nous est pas directement accessible) se trouve dans l'ensemble $\Omega_0 := \{\omega\in\Omega\ ;\ Z_n(\omega)\xrightarrow[n\to+\infty]{}\sigma\}$. La probabilité de gagner ce pari est $P(\Omega_0) = 1$. Ceci légitime l'estimation du paramètre inconnu σ par la valeur approchée $Z_n(\omega_0)$ fournie par les observations.

Solution 9.11 *Escargot aléatoire*

Soit $(X_i)_{i\in\mathbf{N}*}$ une suite de variables aléatoires, définies sur le même espace probabilisé $(\Omega,\mathcal{F},P)$, indépendantes et de même loi, de carré intégrable ($\mathbf{E}(X_1^2) < +\infty$). On définit par récurrence la suite de variables aléatoires positives $(R_n)_{n\geqslant 1}$ par

$$R_1 = |X_1|,\quad R_n = \sqrt{R_{n-1}^2 + X_n^2},\quad n\geqslant 2.$$

19. Le lecteur qui n'aimerait pas cette petite astuce de l'introduction de Z peut toujours effectuer le changement de variable $t = r/\sigma$ et intégrer par parties de façon à faire apparaître un exposant impair de t dans l'intégrale...

1) Pour étudier le comportement asymptotique de $(n^{-1/2}R_n)$, on commence par remarquer que :

$$R_n^2 = R_{n-1}^2 + X_n^2 \quad \text{d'où} \quad R_n^2 = \sum_{i=1}^{n} X_i^2.$$

Ainsi $\left(n^{-1}R_n^2\right)_{n\geqslant 1}$ apparaît comme la suite des moyennes arithmétiques des termes de la suite $(X_i^2)_{i\in\mathbf{N}^*}$. Comme ces variables s'écrivent $X_i^2 = h(X_i)$ où h est la fonction mesurable $h : x \mapsto x^2$, la suite $(X_i^2)_{i\in\mathbf{N}^*}$ hérite de l'indépendance des X_i. Comme h ne dépend pas de i et les X_i ont même loi, il en va de même pour les X_i^2. Enfin, $\mathbf{E}(X_i^2)$ est finie par hypothèse. Les X_i^2 vérifient donc les hypothèse de la loi forte des grands nombres, d'où

$$\frac{R_n^2}{n} = \frac{1}{n}\sum_{i=1}^{n} X_i^2 \xrightarrow[n\to+\infty]{\text{p.s.}} \mathbf{E}(X_1^2). \tag{11.49}$$

Par continuité de la fonction $g : \mathbb{R}_+ \to \mathbb{R}_+$, $x \mapsto \sqrt{x}$, on en déduit que

$$\frac{R_n}{n^{1/2}} \xrightarrow[n\to+\infty]{\text{p.s.}} \sqrt{\mathbf{E}(X_1^2)} =: \tau. \tag{11.50}$$

2) Les X_i étant maintenant de même loi uniforme sur $[0,1]$, on construit dans un repère orthonormé du plan la suite $(M_n)_{n\geqslant 1}$ de points aléatoires par la récurrence suivante. On pose d'abord $M_1 = (X_1, 0)$, puis connaissant M_{n-1} on obtient M_n comme l'unique point tel que $M_{n-1}M_n = X_n$ et que l'angle $(\overrightarrow{M_{n-1}M_n}, \overrightarrow{M_{n-1}O})$ ait pour mesure $+\pi/2$. On trace ainsi la ligne polygonale $\mathcal{E}_n$ de sommets $M_1, M_2, \ldots, M_n$ (l'escargot aléatoire). On fixe un $\varepsilon > 0$ et on se propose de montrer que presque sûrement, $\mathcal{E}_n$ est inclus dans le disque de centre O et de rayon $(1+\varepsilon)\sqrt{n/3}$ pour tout n assez grand.

Commençons par calculer τ lorsque X_1 suit la loi uniforme sur $[0,1]$.

$$\mathbf{E}X_1^2 = \int_0^1 x^2\,\mathrm{d}x = \frac{1}{3}, \quad \text{d'où} \quad \tau = \frac{1}{\sqrt{3}}.$$

Remarquons ensuite que la suite (R_n) étant croissante, $\mathcal{E}_n$ est inclus dans le disque fermé D_n de centre O et de rayon R_n. En effet puisque $OM_i = R_i$, tous les M_i pour $i \leqslant n$ sont dans D_n. De plus si $i < n$, M_i et M_{i+1} sont tous deux dans D_n qui par *convexité* contient tout le segment d'extrémités M_i et M_{i+1}. Ainsi D_n contient toute la ligne polygonale $\mathcal{E}_n$.

Notons $\Omega' := \{\omega \in \Omega\,;\ \lim_{n\to+\infty} n^{-1/2}R_n(\omega) = \tau\}$. On a alors

$$\forall\omega\in\Omega',\ \forall\varepsilon>0,\ \exists n_0 = n_0(\omega,\varepsilon),\ \forall n \geqslant n_0, \quad n^{-1/2}R_n(\omega) \leqslant \tau(1+\varepsilon).$$

Pour ε_0 fixé, on a clairement l'inclusion d'évènements

$$\Omega' \subset \left\{\omega\in\Omega,\ \exists n_0 = n_0(\omega,\varepsilon_0),\ \forall n\geqslant n_0,\ n^{-1/2}R_n(\omega) \leqslant \tau(1+\varepsilon)\right\} =: A_n.$$

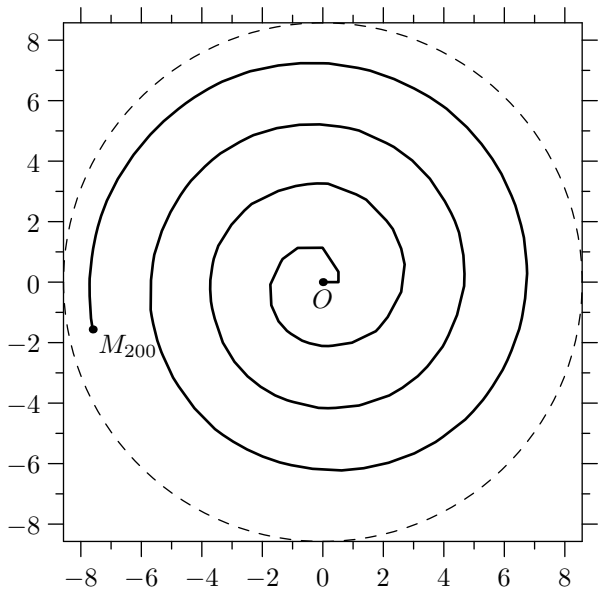


FIGURE 11.12 – Escargot aléatoire $\mathcal{E}_{200}$ (simulation)

Pour tout $\omega \in A_n$, on a

$$\forall n \geqslant n_0(\omega, \varepsilon_0) \quad R_n(\omega) \leqslant n^{1/2}\tau(1+\varepsilon), \quad \text{d'où} \quad \mathcal{E}_n(\omega) \subset D_n \subset D\Big(O, (1+\varepsilon)\sqrt{\frac{n}{3}}\Big).$$

Par la loi forte des grands nombres (question précédente), $P(\Omega') = 1$, d'où $P(A_n) = 1$. On a ainsi établi que presque sûrement, $\mathcal{E}_n$ est inclus dans le disque de centre O et de rayon $(1+\varepsilon)\sqrt{n/3}$ pour tout n assez grand.

La figure 11.12 illustre ceci. On a choisi pour cette simulation $n = 200$, $\varepsilon = 0{,}05$. Les deux cercles tracés ont respectivement pour rayon R_n et $(1+\varepsilon)\sqrt{n/3}$.

Indication 9.16 *Une convergence L^1*

Combiner loi forte des grands nombres et théorème de convergence dominée.

Solution 9.17 *Loi de Cauchy et L.F.G.N*

Le but de cet exercice est d'étudier la convergence des moyennes arithmétiques d'une suite de v.a. indépendantes et de même loi de Cauchy. Toutes les variables aléatoires intervenant dans l'énoncé sont supposées définies sur le même espace probabilisé $(\Omega, \mathcal{F}, P)$.

1) La variable aléatoire réelle X suivant la loi de Cauchy de densité :

$$f : \mathbb{R} \to \mathbb{R}_+, \quad t \mapsto \frac{1}{\pi(1+t^2)},$$

on a pour tout entier $n \geqslant 1$:

$$P(X > n) = \int_n^{+\infty} f(t)\,\mathrm{d}t = \int_n^{+\infty} \frac{1}{\pi(1+t^2)}\,\mathrm{d}t.$$

En notant que pour tout $t \in [n, +\infty[$, $t \geqslant 1$, d'où $1+t^2 \leqslant 2t^2$, on obtient la minoration :

$$\int_n^{+\infty} \frac{1}{\pi(1+t^2)}\,\mathrm{d}t \geqslant \int_n^{+\infty} \frac{1}{\pi(2t^2)}\,\mathrm{d}t = \frac{1}{2\pi}\left[\frac{-1}{t}\right]_n^{+\infty} = \frac{1}{2\pi n}.$$

Nous avons ainsi vérifié que

$$\forall n \geqslant 1, \quad P(X > n) \geqslant \frac{1}{2\pi n}.$$

2) Soit $(X_k)_{k\geqslant 1}$ une suite de variables aléatoires indépendantes et de même loi que X. On pose

$$A := \{\omega \in \Omega\ ;\ X_k(\omega) > k \text{ pour une infinité d'indices } k\}.$$

Notons pour tout entier $k \geqslant 1$, $A_k := \{X_k > n\}$. Les évènements A_k héritent de l'indépendance des variables aléatoires X_k. En utilisant la minoration établie à la question précédente, on obtient :

$$\sum_{k=1}^{+\infty} P(A_k) \geqslant \sum_{k=1}^{+\infty} \frac{1}{2\pi k} = +\infty,$$

en raison de la divergence bien connue de la série harmonique. Par le deuxième lemme de Borel-Cantelli, nous en déduisons que

$$P(\text{réalisation d'une infinité de } A_k) = 1,$$

autrement dit que $P(A) = 1$.

3) Nous allons déduire de ce qui précède que la suite $(X_k)_{k\geqslant 1}$ ne vérifie pas la loi forte des grands nombres. En posant $S_n := X_1 + \cdots + X_n$, on a

$$\forall \omega \in \Omega, \quad \frac{X_n(\omega)}{n} = \frac{S_n(\omega) - S_{n-1}(\omega)}{n} = \frac{S_n(\omega)}{n} - \frac{n-1}{n} \times \frac{S_{n-1}(\omega)}{n-1}. \tag{11.51}$$

Supposons maintenant que S_n/n converge presque-sûrement vers une certaine variable aléatoire Y. Notons $C := \{\omega \in \Omega\ ;\ \lim_{n\to+\infty} S_n(\omega)/n = Y(\omega)\}$. Si ω appartient à C, alors d'après (11.51) et la définition de C, $X_n(\omega)/n$ doit converger vers 0. Ceci montre que cet ω n'appartient pas à A puisque la convergence vers 0 de la suite $(X_n(\omega)/n)_{n\geqslant 1}$ interdit qu'une infinité de ses termes soient supérieurs à 1. Nous venons ainsi de vérifier que $A \cap C = \varnothing$. Par conséquent, C est inclus dans le complémentaire de A qui est de probabilité nulle, d'où $P(C) = 0$. Ceci contredit l'hypothèse de convergence presque-sûre de S_n/n vers Y. Comme nous n'avons fait aucune hypothèse sur Y, cela signifie

que S_n/n ne peut converger p.s. vers *aucune* v.a. Y. En particulier, la loi forte des grands nombres n'est pas vérifiée dans ce contexte car dans la LFGN, la limite p.s. est une v.a. constante [20].

4) Le résultat obtenu ci-dessus ne contredit pas la loi forte des grands nombres, puisque la condition nécessaire et suffisante pour la convergence p.s. de S_n/n est *l'intégrabilité de* X_1 (dans le cas d'une suite $(X_k)_{k\geqslant 1}$ i.i.d.). Or la loi de Cauchy n'a pas d'espérance : $\mathbf{E}|X_1| = +\infty$.

11.10 Exercices du chapitre 10

Solution 10.1 *Convergence en loi dans* $\mathbb{N}$

1) La fonction en dents de scie g_j est continue bornée sur $\mathbb{R}$, donc la convergence en loi de X_n vers X implique celle de $\mathbf{E}g_j(X_n)$ vers $\mathbf{E}g_j(X)$ quand n tend vers l'infini. D'autre part comme X_n est à valeurs dans $\mathbb{N}$,

$$\mathbf{E}g_j(X_n) = \sum_{k\in\mathbb{N}} g_j(k)P(X_n = k) = \sum_{k\in\mathbb{N}} P(X_n = k) = 1.$$

On en déduit que pour tout $j \geqslant 1$, $\mathbf{E}g_j(X) = 1$. Or quand j tend vers l'infini, $g_j(X(\omega))$ converge pour tout ω vers $\mathbf{1}_{\mathbb{N}}(X(\omega))$ et cette convergence est dominée par la variable aléatoire constante 1 qui est intégrable. Par le théorème de convergence dominée, $\mathbf{E}g_j(X)$ converge vers $\mathbf{E}\mathbf{1}_{\mathbb{N}}(X) = P(X \in \mathbb{N})$ quand j tend vers l'infini. Par conséquent, $P(X \in \mathbb{N}) = 1$ et la loi de X est discrète.

Fixons k quelconque dans $\mathbb{N}$. La convergence en loi de X_n vers X implique celle de $\mathbf{E}g_{j,k}(X_n)$ vers $\mathbf{E}g_{j,k}(X)$ quand n tend vers l'infini. Ceci est vrai pour tout $j \geqslant 1$, mais dans ce qui suit, on peut se contenter par exemple de $j = 1$. En notant que k est le seul entier dans le support de $g_{1,k}$, que $P(X_n \in \mathbb{N}) = P(X \in \mathbb{N}) = 1$ et que $g_{1,k}(k) = 1$, on voit que $\mathbf{E}g_{1,k}(X_n) = P(X_n = k)$ et $\mathbf{E}g_{1,k}(X) = P(X = k)$. La convergence d'espérances ci-dessus se réduit donc à la convergence de $P(X_n = k)$ vers $P(X = k)$, ce qu'il fallait démontrer.

2) On suppose désormais que pour tout $k \in \mathbb{N}$, $P(X_n = k)$ converge vers un réel a_k. Tous les a_k sont des réels positifs comme limites de probabilités. Pour tout entier m, on a

$$0 \leqslant \sum_{k=0}^{m} P(X_n = k) = P(X_n \leqslant m) \leqslant 1.$$

La somme ci-dessus a un nombre fini de termes, chacun convergeant vers le a_k de même indice. Donc c'est toute la somme qui converge vers la somme des a_k et en particulier :

$$0 \leqslant \sum_{k=0}^{m} a_k \leqslant 1.$$

20. En fait, nous avons démontré un petit peu plus que ce que demandait l'énoncé, à savoir que $(S_n/n)_{n\geqslant 1}$ *ne peut converger sur aucun évènement* C *tel que* $P(C) > 0$. Ceci est en accord avec la loi du zéro-un (non vue dans cet ouvrage) qui implique que l'évènement « asymptotique » $\{S_n/n$ converge$\}$ ne peut avoir pour probabilité que 0 ou 1, pourvu que les X_k soient indépendantes.

Ceci étant vrai pour tout entier m, les sommes partielles de la série de terme général positif a_k sont bornées par la constante 1, donc la série converge dans $\mathbb{R}_+$ et sa somme est majorée par 1 :

$$0 \leqslant \sum_{k=0}^{+\infty} a_k \leqslant 1.$$

3) Si $\sum_{k\in\mathbb{N}} a_k = 1$, on peut alors définir la probabilité Q sur $\mathbb{R}$ par $Q = \sum_{k\in\mathbb{N}} a_k \delta_k$ et prendre pour X l'identité sur l'espace probabilisé $(\mathbb{R}, \mathrm{Bor}(\mathbb{R}), Q)$. Alors la loi de X est simplement Q et vérifie bien $Q(\{k\}) = a_k$ pour tout entier k. On remarque alors que pour tout réel t,

$$P(X_n \leqslant t) = \sum_{0\leqslant k\leqslant t} P(X_n = k),$$

avec la convention qu'une somme indexée par l'ensemble vide (ce qui arrive ici pour $t < 0$) vaut zéro. Dans cette somme il n'y a qu'un nombre fini de termes, donc on peut passer à la limite quand n tend vers l'infini, ce qui donne

$$F_n(t) = P(X_n \leqslant t) \xrightarrow[n\to+\infty]{} \sum_{0\leqslant k\leqslant t} a_k = \sum_{0\leqslant k\leqslant t} Q(\{k\}) = Q(]-\infty, t]) = F(t),$$

en notant F la fonction de répartition de X. On a donc convergence ponctuelle sur $\mathbb{R}$ de la suite des fonctions de répartition F_n vers la f.d.r. de X, donc X_n converge en loi vers X.

Nota bene. Rappelons que pour la convergence en loi de X_n vers X, il n'est pas nécessaire que les X_n et X soient définis sur *le même* espace probabilisé. Ici nous avons supposé implicitement [21] que les X_n étaient définies sur le même $(\Omega, \mathcal{F}, P)$, mais il n'est pas clair que l'on puisse définir X ayant la loi Q (sous P) sur ce même espace [22]. C'est pourquoi nous l'avons défini sur $(\mathbb{R}, \mathrm{Bor}(\mathbb{R}), Q)$.

4) Supposons que $c = \sum_{k\in\mathbb{N}} a_k$ soit strictement inférieur à 1 et que X_n converge en loi vers une variable aléatoire Y de fonction de répartition G. Posons $\varepsilon = (1-c)/3$. Comme G a pour limite 1 en $+\infty$ et a au plus un ensemble dénombrable de points de discontinuité, on peut trouver un t point de continuité de G tel que $G(t) > 1-\varepsilon$. Alors par convergence en loi de X_n vers Y, $F_n(t)$ converge vers $G(t)$, donc en particulier $F_n(t) > G(t) - \varepsilon > 1 - 2\varepsilon$ pour tout n à partir d'un certain rang n_0. Mais d'autre part, $F_n(t)$ converge vers $\sum_{0\leqslant k\leqslant t} a_k$ qui est majoré par $\sum_{k=0}^{+\infty} a_k = c = 1 - 3\varepsilon$. Donc $F_n(t) < 1 - 2\varepsilon$ à partir d'un certain rang n_1. Pour $n = \max(n_0, n_1)$, on aboutit à une contradiction, ce qui montre que X_n ne peut converger en loi.

La situation où $c < 1$ peut effectivement se produire. Par exemple si X_n suit la loi uniforme sur $[\![1, n]\!]$, tous les a_k sont nuls et $c = 0$. Ou encore si la loi de X_n est donnée par $P(X_n = 0) = p \in]0, 1[$ et $P(X_n = n) = 1 - p$, on trouve $c = p$.

21. Pour le confort de notation qui permet d'écrire $P(X_n = k)$ au lieu de $P_n(X_n = k)$.

22. En revanche il n'est pas difficile de construire X sur $(\Omega, \mathcal{F}, P')$ avec P' convenablement choisie, pour que la loi de X sous P' soit Q, lorsque l'une au moins des X_n est telle qu'*aucun* des évènements $A_k = \{X_n = k\}$, $k \in \mathbb{N}$ ne soit vide. Alors la famille des A_k est une partition de Ω, on peut sélectionner un ω_k dans chaque A_k, poser $P' = \sum_{k\in\mathbb{N}} a_k \delta_{\omega_k}$ et définir $X(\omega) := k$ pour $\omega \in A_k$.

On peut synthétiser le résultat de cet exercice en disant qu'une suite de variables aléatoires X_n à valeurs dans $\mathbb{N}$ converge en loi *si et seulement si* $P(X_n = k)$ converge vers un réel a_k pour tout entier k *et* $\sum_{k=0}^{+\infty} a_k = 1$. Lorsque cette convergence en loi a lieu, la loi limite est $Q = \sum_{k\in\mathbb{N}} a_k \delta_k$.

5) Au vu de ce qui précède, la « convergence de la loi binomiale vers une loi de Poisson » et la « convergence de la loi hypergéométrique vers une loi binomiale », énoncées dans les théorèmes 6.41 et 6.37 au sens naïf de la convergence des $P(X_n = k)$, sont effectivement des convergences en loi au sens de la définition 10.1 ou 10.2.

Solution 10.5 *La cantine*

Voici une modélisation du problème. On numérote les 500 personnes par ordre alphabétique. On note X_i la variable aléatoire valant 1 si la i^{e} personne de la liste mange au premier service, 0 si elle mange au deuxième. Les X_i sont des variables de Bernoulli de même loi : $P(X_i = 0) = P(X_i = 1) = 1/2$. On suppose les X_i indépendantes, hypothèse raisonnable dans le cas d'une numérotation par ordre alphabétique (mais qui serait peut-être plus discutable si on avait numéroté les employés par bureau, les employés d'un même bureau ayant tendance à manger au même service...). Bien que l'énoncé ne le précise pas il s'agit évidemment de trouver le nombre *minimum* de couverts à disposer à chaque service, sinon avec 500 couverts à chaque service, le gérant ne prendrait aucun risque ! Notons k ce nombre minimum. Le nombre (aléatoire) de personnes mangeant au premier service est : $S_n = \sum_{i=1}^{500} X_i$ (avec $n = 500$). Le nombre de personnes mangeant au deuxième service est par conséquent $500 - S_n$ (on suppose que tout le monde mange). Le problème revient donc à trouver k minimal tel que :

$$P(S_n \leqslant k \text{ et } 500 - S_n \leqslant k) \geqslant 0{,}95,$$

ce qui s'écrit encore :

$$P(500 - k \leqslant S_n \leqslant k) \geqslant 0{,}95. \tag{11.52}$$

La loi de S_n est la binomiale $\mathrm{Bin}(500, 1/2)$, mais comme n est grand, il est légitime d'utiliser l'approximation de cette loi par une gaussienne grâce au théorème limite central. Pour ce faire, on normalise S_n en la centrant puis en divisant par l'écart type. L'espérance et la variance de S_n sont $\mathbf{E}S_n = n\frac{1}{2} = 250$ et $\mathrm{Var}\, S_n = n(\frac{1}{2})(1-\frac{1}{2}) = 125$. On peut écrire (11.52) sous la forme équivalente :

$$P\left(\frac{500 - k - 250}{\sqrt{125}} \leqslant \frac{S_n - 250}{\sqrt{125}} \leqslant \frac{k - 250}{\sqrt{125}}\right) \geqslant 0{,}95.$$

D'après le théorème limite central, la variable $(S_n - 250)/\sqrt{125}$ suit approximativement la loi normale $\mathfrak{N}(0, 1)$. En négligeant l'erreur due à cette approximation, le problème revient à trouver k minimal tel que :

$$P\left(\frac{250 - k}{\sqrt{125}} \leqslant Z \leqslant \frac{k - 250}{\sqrt{125}}\right) \geqslant 0{,}95,$$

où Z est une variable gaussienne de loi $\mathfrak{N}(0, 1)$. Notons Φ la fonction de répartition de la loi normale $\mathfrak{N}(0, 1)$. Pour tout $a > 0$, il est facile de vérifier que $P(-a \leqslant Z \leqslant$

$a) = 2\Phi(a) - 1$. En appliquant ce résultat avec $a = (k - 250)/\sqrt{125}$, on est ramené à la résolution de l'inéquation : $2\Phi(a) - 1 \geqslant 0{,}95$ équivalente à $\Phi(a) \geqslant 0{,}975$. La table nous donne $\Phi(1,96) = 0{,}975\,0$. On prendra donc pour k le plus petit entier tel que $(k - 250)/\sqrt{125} \geqslant 1,96$ d'où : $k \geqslant 250 + 1,96\sqrt{125} \approx 271{,}91$. Ainsi il suffit de prévoir 272 couverts par service pour avoir une probabilité supérieure à 0,95 que chacun puisse manger au service de son choix. On constate qu'en acceptant un risque faible de surcharge de la cantine, il y a moyen de réaliser une économie considérable en place et en mobilier.

Solution 10.9 *Records et loi des grands nombres*

On note X une variable aléatoire positive dont la loi vérifie $P(X > t) = t^{-r}$ pour tout réel $t \geqslant 1$. Dans tout l'exercice, on suppose que la constante r est dans $]0, 1[$. On note $(X_k)_{k\geqslant 1}$ une suite de variables aléatoires positives définies sur le même espace probabilisé, indépendantes, de même loi que X et on pose :

$$M_n := \max_{1\leqslant k\leqslant n} X_k, \qquad S_n := \sum_{k=1}^{n} X_k.$$

1) La fonction de survie G de la v.a. X vérifie $G(t) = P(X > t) = t^{-r}$ pour $t \geqslant 1$. Cette formule n'est évidemment plus valable pour $t < 1$ puisqu'elle fournirait une probabilité supérieure à 1. Il est néanmoins facile de voir que pour tout $t < 1$, on a par décroissance de G, $1 \geqslant P(X > t) \geqslant G(1) = 1$, d'où $P(X > t) = 1$. Ainsi $P(X > 1/2) = 1$.

Comme X est une variable aléatoire positive, son espérance (dans $\overline{\mathbb{R}}_+$) est donnée par

$$\mathbf{E}X = \int_0^{+\infty} P(X > t)\,\mathrm{d}t = \int_0^1 \mathrm{d}t + \int_1^{+\infty} t^{-r}\,\mathrm{d}t = +\infty,$$

car $r < 1$.

2) Pour calculer la fonction de répartition de M_n, on remarque d'abord que chacun des X_i est supérieur à 1 avec probabilité 1, donc leur maximum aussi. Il suffit donc de calculer pour tout réel $x \geqslant 1$, $F_n(x) := P(M_n \leqslant x)$, étant bien entendu que $F_n(x) = 0$ pour tout $x < 1$. On note alors que $M_n \leqslant x$ *si et seulement si* $X_k \leqslant x$ pour tout $k \in [\![1, n]\!]$. Cette équivalence logique se traduit par l'égalité d'évènements :

$$\big\{M_n \leqslant x\big\} = \bigcap_{k=1}^{n} \big\{X_k \leqslant x\big\}.$$

En invoquant successivement l'indépendance des X_k, puis leur équidistribution, on en déduit que pour tout $x \geqslant 1$,

$$F_n(x) = P(M_n \leqslant x) = \prod_{k=1}^{n} P(X_k \leqslant x) = \big(1 - G(x)\big)^n = \big(1 - x^{-r}\big)^n. \tag{11.53}$$

Remarquons à titre de vérification que cette fonction de x est continue sur $[1, +\infty[$, croissante comme composée de deux fonctions décroissantes et d'une fonction croissante. Sa limite quand x tend vers $+\infty$ est 1.

Finalement, la fonction de répartition F_n de la v.a. M_n est donnée par :

$$F_n(x) = \begin{cases} 0 & \text{si } x < 1, \\ \big(1 - x^{-r}\big)^n & \text{si } x \geqslant 1. \end{cases}$$

3) Pour étudier la convergence en loi de $n^{-1/r} M_n$, la définition de cette convergence par les fonctions de répartition nous conduit à chercher la limite pour x réel fixé de

$$H_n(x) := P\big(n^{-1/r} M_n \leqslant x\big).$$

Notons qu'en raison de la présence du facteur de normalisation $n^{-1/r}$, la v.a. $n^{-1/r} M_n$ peut prendre des valeurs inférieures à 1 avec probabilité non nulle, mais qu'elle reste une v.a. *positive*. On écartera donc d'emblée le cas trivial $x < 0$ pour lequel $H_n(x) = 0$. De même pour $x = 0$, $H_n(0) = 0$ car $H_n(0) \leqslant P(X_1 \leqslant 0) = 0$.

Pour $x > 0$, on se ramène à F_n en écrivant que

$$H_n(x) = P\big(n^{-1/r} M_n \leqslant x\big) = P\big(M_n \leqslant n^{1/r} x\big) = F_n\big(n^{1/r} x\big).$$

Notons $n_0 = n_0(x)$ la partie entière supérieure de x^{-r}. Alors pour $n \geqslant n_0$, $x^{-r} \leqslant n$, d'où $n^{1/r} x \geqslant 1$ et on peut appliquer (11.53) :

$$\forall n \geqslant n_0, \qquad H_n(x) = \big(1 - (n^{1/r} x)^{-r}\big)^n = \Big(1 - \frac{1}{n x^r}\Big)^n.$$

Lorsque n tend vers l'infini, x restant fixé, cette expression présente une forme indéterminée du type « 1^∞ ». Il est facile de lever l'indétermination en notant que

$$\lim_{n \to +\infty} n \ln\Big(1 - \frac{1}{n x^r}\Big) = -\frac{1}{x^r},$$

d'où en repassant à l'exponentielle (fonction continue),

$$\lim_{n \to +\infty} H_n(x) = \lim_{n \to +\infty} \exp\Big(n \ln\Big(1 - \frac{1}{n x^r}\Big)\Big) = \exp(-x^{-r}).$$

Récapitulons :

$$\forall x \in \mathbb{R}, \qquad \lim_{n \to +\infty} H_n(x) = H(x) := \begin{cases} 0 & \text{si } x \leqslant 0, \\ \exp(-x^{-r}) & \text{si } x > 0. \end{cases}$$

La fonction H est croissante sur $\mathbb{R}$ comme limite ponctuelle d'une suite d'applications croissantes sur $\mathbb{R}$ (vous pouvez aussi vérifier cette croissance sur la formule ci-dessus ... sans calculer de dérivée !). Elle est continue sur $\mathbb{R}$: la continuité est claire sur chacun des intervalles $]-\infty, 0[$ et $]0, +\infty[$ et au point de raccord 0, on remarque que quand x tend vers 0 par valeurs supérieures, $-x^{-r}$ tend vers $-\infty$, donc son exponentielle tend vers 0. La limite de H en $-\infty$ est trivialement 0. En $+\infty$, H tend vers 1 car $-x^{-r}$

tend vers 0. Ainsi H est une fonction de répartition. On a donc vérifié que $n^{-1/r}M_n$ converge en loi vers une v.a. Z de fonction de répartition H.
Nota bene. Pour avoir la convergence en loi, il est indispensable de vérifier que la fonction limite H obtenue ci-dessus est bien une fonction de répartition.

4) On peut montrer que

$$\forall \delta > 0, \quad \sum_{n=1}^{+\infty} \exp(-n^{\delta}) < +\infty, \tag{11.54}$$

en vérifiant que le terme général *positif* de cette série tend plus vite vers 0 que disons, n^{-2}. Pour cela, on regarde le quotient

$$\frac{\exp(-n^{\delta})}{n^{-2}} = n^2 \exp(-n^{\delta}) = \exp(-n^{\delta} + 2\ln n) = \exp\Big(-n^{\delta}\Big(1 - \frac{2\ln n}{n^{\delta}}\Big)\Big),$$

qui sous cette forme, tend clairement vers 0 quand n tend vers l'infini.

5) Soit a une constante telle que $0 < a < 1/r$. Montrons que

$$\sum_{n=1}^{+\infty} P(M_n \leqslant n^a) < +\infty. \tag{11.55}$$

Puisque $a > 0$, $n^a \geqslant 1$ pour tout $n \geqslant 1$ et on peut appliquer (11.53), d'où

$$\sum_{n=1}^{+\infty} P(M_n \leqslant n^a) = \sum_{n=1}^{+\infty} \big(1 - n^{-ar}\big)^n.$$

Comme cette série est à termes *positifs*, il suffit de montrer que l'on peut majorer son terme général par celui d'une série convergente. On remarque alors que

$$\big(1 - n^{-ar}\big)^n = \exp\big(n \ln(1 - n^{-ar})\big).$$

Par concavité de la fonction logarithme [23], $\ln(1 + u) \leqslant u$ pour tout $u > -1$. En appliquant ceci avec $u = -n^{-ar} > -1$, on obtient la majoration

$$\exp\big(n \ln(1 - n^{-ar})\big) \leqslant \exp\big(-n^{1-ar}\big). \tag{11.56}$$

Par (11.54) avec $\delta := 1 - ar$, ceci est bien le terme général d'une série convergente ($\delta > 0$ car $a < 1/r$).
Nota bene. Attention, de l'équivalence $n\ln(1-n^{-ar}) \sim -n^{1-ar}$, on ne peut pas déduire l'équivalence des exponentielles « $\exp(n\ln(1 - n^{-ar})) \sim \exp(-n^{1-ar})$ ». En effet, la première équivalence s'écrit $n\ln(1 - n^{-ar}) = -n^{1-ar}(1 + \varepsilon(n))$, avec $\lim_{n\to+\infty} \varepsilon(n) = 0$. En prenant l'exponentielle, $\exp(n\ln(1 - n^{-ar})) = \exp(-n^{1-ar})\exp(-n^{1-ar}\varepsilon(n))$ et on ne peut conclure à l'équivalence des exponentielles que si $\exp(-n^{1-ar}\varepsilon(n))$

23. L'inégalité qui suit exprime simplement que la courbe représentative de $u \mapsto \ln(1 + u)$ est « en dessous » de sa tangente au point $u = 0$.

tend vers 1, ce qui équivaut à $n^{1-ar}\varepsilon(n)$ tend vers 0. Cette dernière condition n'est en général pas réalisée. On peut néanmoins conclure avec cette approche en remarquant que $\varepsilon(n) \geqslant 0$ pour n assez grand (parce que $\varepsilon(n) \sim \frac{1}{2}n^{-ar}$, comme on peut le voir en développant $\ln(1-u)$ à l'ordre 2). On en déduit que pour n assez grand, $\exp(-n^{1-ar}\varepsilon(n)) \leqslant 1$, ce qui nous permet de retomber sur la majoration (11.56).

6) Grâce au premier lemme de Borel Cantelli, la convergence de série (11.55) nous dit que l'évènement

$$A := \big\{\omega \in \Omega\,;\ M_n(\omega) \leqslant n^a \text{ pour une infinité d'indices } n\big\}$$

est de probabilité nulle. Son complémentaire $\Omega' = A^c$ est donc de probabilité 1. L'appartenance d'un ω à Ω' signifie que l'inégalité $M_n(\omega) \leqslant n^a$ ne peut avoir lieu que pour un nombre fini (éventuellement nul) d'indices n, donc qu'il existe un rang $N = N(\omega)$ tel que pour tout $n \geqslant N$, $M_n(\omega) > n^a$. Comme $P(\Omega') = 1$, ceci implique que presque-sûrement $M_n > n^a$ pour tout $n \geqslant N$, avec N aléatoire (c'est à dire dépendant de ω).

7) En prenant $a > 1$, on en déduit que

$$\frac{S_n}{n} \xrightarrow[n\to+\infty]{\text{p.s.}} +\infty.$$

En effet comme les X_k sont positives, on a pour tout $\omega \in \Omega$, $M_n(\omega) \leqslant S_n(\omega)$. Si de plus $\omega \in \Omega'$, on a un entier $N(\omega)$ tel que :

$$\forall n \geqslant N(\omega), \quad \frac{S_n(\omega)}{n} \geqslant \frac{M_n(\omega)}{n} > \frac{n^a}{n} = n^{a-1}.$$

Comme $a > 1$, ce minorant tend vers $+\infty$ avec n et comme $P(\Omega') = 1$, on en déduit que S_n/n tend presque sûrement vers $+\infty$. Ceci ne contredit pas la loi forte des grands nombres qui ne s'applique pas ici puisque pour $r \leqslant 1$, X_1 n'est pas intégrable.

8) En fait

$$\frac{S_n}{n^c} \xrightarrow[n\to+\infty]{\text{p.s.}} +\infty,$$

pour tout exposant c tel que $0 < c < 1/r$. Pour le voir, on intercale un a entre c et $1/r$ et on adapte immédiatement le raisonnement ci-dessus en remplaçant le dénominateur n par n^c. Le minorant presque sûr à partir du rang aléatoire N devient n^{a-c} qui tend encore vers $+\infty$.

Solution 10.10 *Convergences de moyennes géométriques*

Soit $(X_k)_{k\geqslant 1}$ une suite de variables aléatoires réelles définies sur le même espace probabilisé, à valeurs dans $]0,+\infty[$, indépendantes, de même loi telle que $\mathbf{E}|\ln X_1| < +\infty$. On note $b := \mathbf{E}\ln X_1$.

1) On note $T_n = X_1 \cdots X_n$ le produit des n premières variables aléatoires de la suite $(X_k)_{k\geqslant 1}$, leur moyenne géométrique est $T_n^{1/n}$.

Posons pour tout $k \geqslant 1$, $Z_k := \ln X_k$. Comme X_k est à valeurs dans $]0,+\infty[$ et ln est définie et continue (donc borélienne) sur cet ensemble, on définit bien ainsi une

variable aléatoire. La suite $(Z_k)_{k\geqslant 1}$ hérite de l'indépendance et de l'équidistribution de $(X_k)_{k\geqslant 1}$, par image par la même fonction continue ln. De plus l'hypothèse ci-dessus s'écrit $\mathbf{E}|Z_1| < +\infty$. La loi forte des grands nombres nous dit alors que la *moyenne arithmétique* de $Z_1, \ldots, Z_n$ converge presque-sûrement vers $b = \mathbf{E}Z_1$. En prenant l'image par la fonction *continue* exponentielle, on en déduit que

$$T_n^{1/n} = \exp\Big(\frac{1}{n}(Z_1 + \cdots + Z_n)\Big) \xrightarrow[n\to+\infty]{\text{p.s.}} \mathrm{e}^b = \exp(\mathbf{E}\ln X_1). \tag{11.57}$$

2) Calculons cette limite dans le cas des v.a. X_k de l'exercice précédent. La fonction de répartition de X_1 est

$$F(t) = \begin{cases} 0 & \text{si } t < 1, \\ 1 - t^{-r} & \text{si } t \geqslant 1. \end{cases}$$

Cette fonction est C^1 par morceaux avec raccord continu au point 1. La loi de X_1 a donc une densité f, obtenue par dérivation de F :

$$f(t) = \begin{cases} 0 & \text{si } t \leqslant 1, \\ rt^{-r-1} & \text{si } t > 1. \end{cases}$$

La variable aléatoire X_1 étant p.s. supérieure à 1, $\ln X_1$ est p.s. positive, il n'y a donc pas lieu de vérifier l'intégrabilité avant de calculer $\mathbf{E}\ln X_1$.

$$\mathbf{E}\ln X_1 = \int_0^{+\infty} f(t)\ln t\,\mathrm{d}t = \int_1^{+\infty} rt^{-r-1}\ln t\,\mathrm{d}t = \lim_{a\to+\infty}\int_1^a rt^{-r-1}\ln t\,\mathrm{d}t.$$

On calcule $\int_1^a rt^{-r-1}\ln t\,\mathrm{d}t$ par parties en posant $u(t) = \ln t$ et $v'(t) = rt^{-r-1}$, ce qui donne :

$$\int_1^a rt^{-r-1}\ln t\,\mathrm{d}t = \big[-t^{-r}\ln t\big]_1^a + \int_1^a t^{-r-1}\,\mathrm{d}t = -a^{-r}\ln a - \frac{1}{r}(a^{-r} - 1).$$

Quand a tend vers $+\infty$, cette expression tend vers $1/r$, d'où

$$\mathbf{E}\ln X_1 = 1/r \quad \text{et} \quad T_n^{1/n} \xrightarrow[n\to+\infty]{\text{p.s.}} \exp(1/r).$$

On pouvait aussi calculer $\mathbf{E}\ln X_1$ sans passer par la densité f. En effet, comme $X_1 \geqslant 1$ p.s., $\ln X_1$ est positive p.s., d'où

$$\mathbf{E}\ln X_1 = \int_0^{+\infty} P(\ln X_1 > x)\,\mathrm{d}x.$$

Or la fonction exponentielle réalise une bijection croissante de $[0, +\infty[$ sur $[1, +\infty[$, d'où pour tout $x \in [0, +\infty[$, $\ln X_1 > x \Leftrightarrow X_1 > \mathrm{e}^x$ et par suite $P(\ln X_1 > x) =$

$P(X_1 > \mathrm{e}^x)$. Comme $t = \mathrm{e}^x \geqslant 1$, $P(X_1 > \mathrm{e}^x) = t^{-r} = \mathrm{e}^{-rx}$. En reportant ceci dans l'intégrale ci-dessus, il vient : $\mathbf{E} \ln X_1 = \int_0^{+\infty} \mathrm{e}^{-rx}\,\mathrm{d}x = \frac{1}{r}$.

3) On revient au cas général et on pose $Y_k := \mathrm{e}^{-b} X_k$ et $R_n := Y_1 \cdots Y_n$. Au vu de (11.57), on a

$$R_n^{1/n} = \exp\Big(\frac{1}{n}(Z_1 + \cdots + Z_n) - b\Big).$$

Par la loi forte des grands nombres appliquée à $(Z_k)_{k\geqslant 1}$, $n^{-1}(Z_1+\cdots+Z_n)-b$ converge presque sûrement vers 0. Par continuité, son exponentielle converge presque-sûrement vers $\mathrm{e}^0 = 1$.

4) Commençons par remarquer que $R_n^{1/\sqrt{n}} = (R_n^{1/n})^{\sqrt{n}}$ en sorte que si l'on essaie d'étudier la convergence presque-sûre de $R_n^{1/\sqrt{n}}$, on se heurte à une forme indéterminée du type « 1^∞ ». En fait nous allons montrer que si $\mathbf{E}Z_1^2 < +\infty$, il y a convergence *en loi* vers une v.a. non constante W, ce qui montrera au passage qu'il ne peut y avoir convergence p.s. vers une constante (pourquoi ?).

Pour alléger les écritures, posons $\sigma^2 = \operatorname{Var} Z_1 = \operatorname{Var} \ln X_1$ (on rappelle que cette variance est strictement positive [24]) et

$$S_n := Z_1 + \cdots + Z_n, \qquad S_n^* := \frac{S_n - \mathbf{E}S_n}{\sqrt{\operatorname{Var} S_n}} = \frac{S_n - nb}{\sigma\sqrt{n}}.$$

Par le théorème limite central (cas i.i.d.), S_n^* converge en loi vers une v.a. V de loi gaussienne $\mathfrak{N}(0,1)$.

On remarque alors que

$$\ln\big(R_n^{1/\sqrt{n}}\big) = \frac{1}{\sqrt{n}} \ln R_n = \frac{1}{\sqrt{n}}(S_n - nb) = \sigma S_n^*,$$

ce qui s'écrit aussi

$$R_n^{1/\sqrt{n}} = \exp\big(\sigma S_n^*\big).$$

Par conservation de la convergence en loi par image par la fonction *continue* $t \mapsto \exp(\sigma t)$, on en déduit que

$$R_n^{1/\sqrt{n}} \xrightarrow[n\to+\infty]{\text{loi}} \exp(\sigma V) =: W.$$

La loi de σV est gaussienne $\mathfrak{N}(0,\sigma)$. La loi de W est appelée *log-normale* car $\ln W$ est gaussienne.

Solution 10.12 *La méthode de Monte-Carlo pour le calcul d'intégrales*

Soit $h : [0,1] \to \mathbb{R}$ une fonction continue. On se propose de donner une valeur approchée de

$$I := \int_0^1 h(x)\,\mathrm{d}x,$$

24. Le cas où $\sigma = 0$ donne X_k constante p.s. et rend le problème sans intérêt.

basée sur le calcul de

$$S_n := \sum_{i=1}^{n} h(U_i),$$

où $(U_i)_{i\geqslant 1}$ est une suite de de variables aléatoires indépendantes et de même loi uniforme sur $[0,1]$ obtenues par simulation informatique.

1) Rappelons la proposition 7.42, réécrite ici avec un léger changement de notations, relative à l'existence et au calcul de $\mathbf{E}h(Y)$ lorsque Y est une v.a. à densité.

Proposition. *Soit Y une variable aléatoire de densité f et $h : \mathbb{R} \to \mathbb{R}$ une application réglée sur tout intervalle fermé borné de $\mathbb{R}$. Alors*

$$\mathbf{E}|h(Y)| = \int_{-\infty}^{+\infty} |h(x)| f(x)\,\mathrm{d}x. \tag{11.58}$$

Si $\mathbf{E}|h(Y)| < +\infty$, on a de plus

$$\mathbf{E}h(Y) = \int_{-\infty}^{+\infty} h(x) f(x)\,\mathrm{d}x. \tag{11.59}$$

En appliquant cette proposition avec $Y = U_1$, de densité $f = \mathbf{1}_{[0,1]}$ et h continue (donc réglée [25]), on obtient

$$\mathbf{E}|X_1| = \mathbf{E}|h(U_1)| = \int_{-\infty}^{+\infty} |h(x)| \mathbf{1}_{[0,1]}(x)\,\mathrm{d}x = \int_0^1 |h(x)|\,\mathrm{d}x.$$

Cette intégrale est une intégrale de Riemann ordinaire de la fonction continue $|h|$ sur l'intervalle fermé borné $[0,1]$, donc est finie. Ceci justifie l'intégrabilité de X_1 et nous permet d'écrire en traduisant (11.59),

$$\mathbf{E}X_1 = \mathbf{E}h(U_1) = \int_0^1 h(x)\,\mathrm{d}x = I.$$

2) La suite de variables aléatoires $(X_i)_{i\geqslant 1}$ est i.i.d. comme image de la suite i.i.d. $(U_i)_{i\geqslant 1}$ par la fonction continue (donc borélienne) h. On vient de voir que X_1 est intégrable. On peut donc appliquer à la suite $(X_i)_{i\geqslant 1}$ la loi forte des grands nombres, ce qui nous donne

$$\frac{S_n}{n} \xrightarrow[n\to+\infty]{\text{p.s.}} \mathbf{E}X_1 = I.$$

Ce résultat légitime pour n « grand » l'approximation de I par la valeur $S_n(\omega)/n$ calculée à partir de l'échantillon généré par l'ordinateur. On s'intéresse maintenant au contrôle de l'erreur d'approximation.

3) En appliquant à nouveau la proposition 7.42 avec la fonction continue h^2 à la place de h, on obtient

$$\mathbf{E}|X_1^2| = \mathbf{E}h^2(U_1) = \int_0^1 h(x)^2\,\mathrm{d}x,$$

25. Ici, h n'est définie que sur $[0,1]$, donc en toute rigueur, il faudrait la prolonger à tout $\mathbb{R}$, soit en une fonction continue en posant $h(x) := h(0)$ pour $x < 0$ et $h(x) := h(1)$ pour $x > 1$, soit en une fonction continue par morceaux en posant $h(x) := 0$ pour $x \notin [0,1]$.

ce qui est encore une intégrale de Riemann ordinaire de la fonction continue h^2 sur l'intervalle fermé borné $[0,1]$. Cette quantité est donc finie et X_1 est bien de carré intégrable. De plus $\mathbf{E}X_1^2 = \int_0^1 h(x)^2\,\mathrm{d}x$.

En appliquant la formule de Koenig, on obtient

$$\sigma^2 := \operatorname{Var} X_1 = \mathbf{E}(X_1^2) - (\mathbf{E}X_1)^2 = \int_0^1 h(x)^2\,\mathrm{d}x - I^2.$$

Dans ce contexte, il n'est pas raisonnable de supposer σ connue, puisqu'on ne sait pas calculer $I = \int_0^1 h(x)\,\mathrm{d}x$. D'ailleurs, sauf cas particulier, le calcul exact de $\int_0^1 h(x)^2\,\mathrm{d}x$ risque lui aussi de ne pas être à notre portée.

Remarque. Pour les étudiants qui ont cru à tort ici qu'on leur demandait de vérifier la positivité de $\int_0^1 h(x)^2\,\mathrm{d}x - I^2$, voici quand même une réponse à cette question non posée. Il suffit d'appliquer l'inégalité de Cauchy-Schwarz pour les intégrales de Riemann au produit de fonctions $h = h \times 1$. Ceci nous donne

$$\left\{\int_0^1 h(x)\,\mathrm{d}x\right\}^2 = \left\{\int_0^1 h(x)\times 1\,\mathrm{d}x\right\}^2 \leqslant \left\{\int_0^1 h(x)^2\,\mathrm{d}x\right\}\left\{\int_0^1 1^2\,\mathrm{d}x\right\} = \int_0^1 h(x)^2\,\mathrm{d}x,$$

d'où $\int_0^1 h(x)^2\,\mathrm{d}x - I^2 \geqslant 0$.

4) Nous allons proposer un intervalle de confiance pour I en utilisant le théorème limite central. Ceci nous amène à rajouter l'hypothèse $\sigma^2 > 0$. En fait le cas d'égalité dans l'inégalité de Cauchy-Schwarz pour les intégrales de fonctions *continues* montre que si $\sigma^2 = 0$, alors h est constante sur $[0,1]$. Le calcul de son intégrale est alors trivial et il ne nécessite pas le recours à la méthode de Monte-Carlo ! Une autre façon de légitimer l'hypothèse $\sigma^2 > 0$ est d'utiliser le fait que les variables aléatoires de variance nulle sont constantes presque sûrement. Dans ce cas on aurait $h(U_i) = c$ pour tout i d'où $S_n/n = c$, presque-sûrement pour tout n. Alors $S_n/n = I = c = \mathbf{E}h(U_1)$ p.s. pour tout n. Dans ce cas, il est inutile de calculer un grand nombre de X_i pour évaluer I, un seul suffit ! Et il n'y a pas davantage lieu de proposer un intervalle de confiance pour I puisqu'on a immédiatement sa valeur exacte.

Puisque les X_i sont i.i.d., de carré intégrable et $\sigma^2 > 0$, le théorème limite central nous donne la convergence en loi vers une v.a. Z gaussienne $\mathfrak{N}(0,1)$ de la suite des sommes centrées normalisées S_n^* qui s'écrivent ici :

$$S_n^* := \frac{S_n - \mathbf{E}S_n}{\sqrt{\operatorname{Var} S_n}} = \frac{S_n - nI}{\sigma\sqrt{n}} = \frac{\sqrt{n}}{\sigma}\Big(\frac{S_n}{n} - I\Big).$$

En notant Φ la f.d.r. de $\mathfrak{N}(0,1)$, la convergence en loi de S_n^* vers Z implique, en tenant compte de la continuité de Φ, que pour tous réels $a < b$,

$$P(S_n^* \in [a,b]) \xrightarrow[n\to+\infty]{} \Phi(b) - \Phi(a).$$

On utilise cette convergence pour légitimer lorsque n est grand, l'approximation

$$P(S_n^* \in [a,b]) \simeq \Phi(b) - \Phi(a).$$

Ensuite, on note les équivalences logiques

$$S_n^* \in [a,b] \;\Leftrightarrow\; \frac{a\sigma}{\sqrt{n}} \leqslant \frac{S_n}{n} - I \leqslant \frac{b\sigma}{\sqrt{n}} \;\Leftrightarrow\; \frac{S_n}{n} - \frac{b\sigma}{\sqrt{n}} \leqslant I \leqslant \frac{S_n}{n} + \frac{a\sigma}{\sqrt{n}}.$$

En prenant $[a,b] = [-t,t]$, on obtient un intervalle d'encadrement de I centré sur S_n/n, la probabilité que cet encadrement soit vrai étant :

$$P\Big(\frac{S_n}{n} - \frac{t\sigma}{\sqrt{n}} \leqslant I \leqslant \frac{S_n}{n} + \frac{t\sigma}{\sqrt{n}}\Big) \simeq \Phi(t) - \Phi(-t) = 2\Phi(t) - 1,$$

la dernière égalité utilisant la parité de la *densité* Φ' de $\mathfrak{N}(0,1)$. En négligeant l'erreur due à l'approximation gaussienne, on peut dire que la probabilité de cet encadrement est au moins 95 % pour tout t tel que $2\Phi(t) - 1 \geqslant 0{,}95$ ou encore $\Phi(t) \geqslant 0{,}975$. On a évidemment intérêt à choisir l'intervalle de longueur minimale, ce qui revient à prendre pour t la plus petite solution de cette inéquation, soit en raison de la croissance et de la continuité de Φ la solution de l'équation $\Phi(t) = 0{,}975$. La table des valeurs de Φ nous donne comme valeur numérique approchée 1,96. Nous pouvons donc proposer comme intervalle d'encadrement pour I

$$\left[\frac{S_n}{n} - \frac{1{,}96\sigma}{\sqrt{n}}, \frac{S_n}{n} + \frac{1{,}96\sigma}{\sqrt{n}}\right] \quad \text{avec une probabilité d'au moins 95 \%,}$$

en négligeant l'erreur due à l'approximation gaussienne.

5) Cet intervalle dépendant de la quantité inconnue σ, n'est pas utilisable en pratique. Une façon de résoudre ce problème est de majorer σ lorsque l'on connaît un majorant M de $\sup_{x\in[0,1]} |h(x)|$. En effet on a alors $h^2 \leqslant M^2$, d'où

$$\sigma^2 = \int_0^1 h(x)^2 \,\mathrm{d}x - I^2 \leqslant \int_0^1 M^2 \,\mathrm{d}x = M^2.$$

On a donc $\sigma \leqslant M$. Remplacer σ par M élargit l'intervalle proposé ci-dessus et donc augmente (au sens large) la probabilité que cet intervalle aléatoire contienne I. Un intervalle de confiance pour I au niveau 95 % est donc

$$\left[\frac{S_n}{n} - \frac{1{,}96M}{\sqrt{n}}, \frac{S_n}{n} + \frac{1{,}96M}{\sqrt{n}}\right].$$

6) Une autre méthode pour construire un intervalle de confiance pour I, consiste à estimer la variance inconnue σ^2 par la variance empirique

$$V_n := \frac{1}{n}\sum_{i=1}^n h(U_i)^2 - \Big(\frac{S_n}{n}\Big)^2 = W_n - \Big(\frac{S_n}{n}\Big)^2,$$

qui est calculable à partir de la simulation des U_i. La justification théorique de cette méthode repose sur le théorème limite central avec autonormalisation. Ce théorème

nous dit que si on remplace σ par $V_n^{1/2}$ dans l'expression de S_n^*, alors la nouvelle suite des variables aléatoires S_n^{**} ainsi obtenues converge encore en loi vers Z de loi $\mathfrak{N}(0,1)$. On a donc

$$S_n^{**} := \frac{S_n - nI}{n^{1/2}V_n^{1/2}} = \frac{n^{1/2}}{V_n^{1/2}}\Big(\frac{S_n}{n} - I\Big) \xrightarrow[n\to+\infty]{\text{loi}} Z.$$

On en déduit comme à la question 4 que

$$P(S_n^{**} \in [a,b]) \xrightarrow[n\to+\infty]{} \Phi(b) - \Phi(a).$$

Une adaptation immédiate de ce qui a été fait à la question 4 conduit à proposer

$$\left[\frac{S_n}{n} - 1{,}96\sqrt{\frac{V_n}{n}}, \frac{S_n}{n} + 1{,}96\sqrt{\frac{V_n}{n}}\right]$$

comme intervalle de confiance pour I au niveau 95 %.

Solution 10.16 *Matrices de covariance et couples gaussiens*

En dimension 1, il est facile de vérifier que tout nombre réel positif t est la variance d'une certaine variable aléatoire. Par exemple, prenons X variable de Bernoulli de paramètre $1/2$, sa variance est $1/4$. Pour toute constante c, $\operatorname{Var}(cX) = c^2 \operatorname{Var} X = c^2/4$. En prenant $c = 2\sqrt{t}$ et $Y = cX$, on obtient $\operatorname{Var} Y = t$. On a choisi ici pour X une v.a. de Bernoulli par souci de simplicité, mais n'importe quelle variable aléatoire de variance non nulle aurait fait l'affaire. La question analogue en dimension 2 est beaucoup moins triviale et s'énonce : « étant donné une matrice 2×2, à quelles conditions est-elle matrice de covariance d'un certain vecteur aléatoire ? ».

1) Pour chercher des conditions *nécessaires*, supposons d'abord que M est effectivement la matrice de covariance d'un certain vecteur aléatoire (X,Y) de $\mathbb{R}^2$. Alors

$$M = \begin{pmatrix} \operatorname{Var} X & \operatorname{Cov}(X,Y) \\ \operatorname{Cov}(Y,X) & \operatorname{Var} Y \end{pmatrix}.$$

Comme la variance est positive, la diagonale principale de M est formée par les réels positifs $a = \operatorname{Var} X$ et $b = \operatorname{Var} Y$. La covariance étant symétrique, on a $\operatorname{Cov}(X,Y) = \operatorname{Cov}(Y,X) = c$. D'autre part avec ces notations, l'inégalité de Cauchy-Schwarz pour les covariances :

$$|\operatorname{Cov}(X,Y)|^2 \leqslant \operatorname{Var} X \operatorname{Var} Y$$

s'écrit $c^2 \leqslant ab$. Finalement M est de la forme :

$$M = \begin{pmatrix} a & c \\ c & b \end{pmatrix}, \qquad a,b \in \mathbb{R}^+,\ c \in \mathbb{R}, \quad ab - c^2 \geqslant 0. \tag{11.60}$$

Réciproquement, soit M une matrice 2×2 de la forme (11.60), nous allons montrer qu'elle est la matrice de covariance d'un certain couple aléatoire (X,Y) de $\mathbb{R}^2$.

2) Examinons d'abord le cas particulier où $ab = 0$. Puisque $0 \leqslant c^2 \leqslant ab$, on en déduit immédiatement que $c = 0$. Il y a trois possibilités.

1. $a = 0$ et $b = 0$, alors M est la matrice nulle, c'est la matrice de covariance d'un vecteur aléatoire constant (C_1, C_2) puisque la variance d'une constante est nulle.
2. $a = 0$ et $b > 0$, alors $M = \begin{pmatrix} 0 & 0 \\ 0 & b \end{pmatrix}$ est la matrice de covariance d'un vecteur aléatoire (C_1, Y), C_1 désignant n'importe quelle constante réelle et Y n'importe quelle variable aléatoire de variance b (d'après l'introduction il en existe toujours une).
3. $a > 0$ et $b = 0$, alors $M = \begin{pmatrix} a & 0 \\ 0 & 0 \end{pmatrix}$ est la matrice de covariance d'un vecteur aléatoire (X, C_2), où C_2 désigne n'importe quelle constante réelle et X n'importe quelle variable aléatoire de variance a.

Dans chacun de ces trois sous-cas, on a utilisé implicitement le fait que si l'une des composantes d'un couple aléatoire est constante, la covariance lorsqu'elle existe [26] est nulle. C'est une conséquence immédiate de l'inégalité de Cauchy-Schwarz ci-dessus et de la nullité de la variance d'une constante.

3) On suppose désormais que ab est non nul. On pose alors

$$\sigma_1 = \sqrt{a},\ \sigma_2 = \sqrt{b}, \quad \rho = \frac{c}{\sqrt{ab}} \quad (\text{donc } \rho \in [-1, 1]),$$

on se donne un couple (U, V) de variables aléatoires réelles *indépendantes*, de carré intégrable et de variance 1 et on définit l'application linéaire g par :

$$g : \mathbb{R}^2 \to \mathbb{R}^2, \quad (u, v) \longmapsto (x, y), \text{ avec } \begin{cases} x = \sigma_1 u, \\ y = \sigma_2 \rho u + \sigma_2 \sqrt{1 - \rho^2}\, v. \end{cases} \tag{11.61}$$

Les composantes X et Y du vecteur aléatoire $(X, Y) = g(U, V)$ sont de carré intégrable comme combinaisons linéaires des variables aléatoires U et V de carré intégrable. Par conséquent (X, Y) possède une matrice de covariance. Pour la calculer, on se ramène à U et V en rappelant que la variance d'une somme de deux v.a. indépendantes est la somme des variances et que la covariance de deux v.a. indépendantes est nulle.

$$\operatorname{Var} X = \operatorname{Var}(\sigma_1 U) = \sigma_1^2 \operatorname{Var} U = \sigma_1^2 = a. \tag{11.62}$$

$$\begin{aligned} \operatorname{Var} Y &= \operatorname{Var}\left(\sigma_2 \rho U + \sigma_2 \sqrt{1 - \rho^2}\, V\right) \\ &= \sigma_2^2 \rho^2 \operatorname{Var} U + 2\sigma_2^2 \rho \sqrt{1 - \rho^2} \operatorname{Cov}(U, V) + \sigma_2^2 (1 - \rho^2) \operatorname{Var} V \\ &= b\rho^2 + 0 + b(1 - \rho^2) = b. \end{aligned} \tag{11.63}$$

$$\begin{aligned} \operatorname{Cov}(X, Y) &= \operatorname{Cov}\left(\sigma_1 U, \sigma_2 \rho U + \sigma_2 \sqrt{1 - \rho^2}\, V\right) \\ &= \sigma_1 \sigma_2 \rho \operatorname{Var} U + \sigma_1 \sigma_2 \sqrt{1 - \rho^2} \operatorname{Cov}(U, V) \\ &= \sqrt{a}\sqrt{b}\frac{c}{\sqrt{ab}} + 0 = c. \end{aligned} \tag{11.64}$$

26. C'est-à-dire si l'autre composante est de carré intégrable.

Au vu de (11.62), (11.63) et (11.64), la matrice de covariance K du vecteur aléatoire (X, Y) est :

$$K = \begin{pmatrix} \operatorname{Var} X & \operatorname{Cov}(X,Y) \\ \operatorname{Cov}(Y,X) & \operatorname{Var} Y \end{pmatrix} = \begin{pmatrix} a & c \\ c & b \end{pmatrix} = M.$$

Nous avons ainsi construit un vecteur aléatoire (X, Y) ayant M pour matrice de covariance.

Tout ce qui précède montre qu'une matrice 2×2 à coefficients réels est la matrice de covariance d'un vecteur aléatoire de $\mathbb{R}^2$ *si et seulement si* elle est de la forme (11.60).

4) Lorsque U et V ci-dessus sont gaussiennes $\mathfrak{N}(0,1)$, le vecteur aléatoire (U, V) est *gaussien* en raison de *l'indépendance* de U et V. Cette indépendance nous permet aussi d'affirmer que le couple (U, V) admet une densité $f_{(U,V)}$ donnée par

$$f_{(U,V)}(u,v) = f_U(u) f_V(v) = \frac{1}{\sqrt{2\pi}} \exp\Big(\frac{-u^2}{2}\Big) \frac{1}{\sqrt{2\pi}} \exp\Big(\frac{-v^2}{2}\Big) = \frac{1}{2\pi} \exp\Big(-\frac{u^2+v^2}{2}\Big).$$

Comme (X, Y) est image du vecteur gaussien (U, V) par l'application *linéaire* g, c'est encore un vecteur gaussien.

En prime nous avons obtenu une méthode simple de simulation d'un vecteur aléatoire gaussien de $\mathbb{R}^2$ de matrice de covariance donnée.

5) Soit (X', Y'), un vecteur aléatoire gaussien d'espérance $(0, 0)$ et de matrice de covariance K. Puisque K est une matrice de covariance, elle est de la forme (11.60), avec $a = \sigma_1^2$, $b = \sigma_2^2$, $c = \operatorname{Cov}(X', Y')$. Le déterminant de K n'étant pas nul, le cas $ab = 0$ est exclu (car $\det K = ab - c^2 \geqslant 0$ donc si $ab = 0$, c est nul et $\det K = 0$). Les variances de X' et Y' sont donc strictement positives et le coefficient de corrélation ρ est bien défini :

$$\rho = \frac{\operatorname{Cov}(X', Y')}{\sqrt{\operatorname{Var} X' \operatorname{Var} Y'}} = \frac{c}{\sqrt{ab}}.$$

Le vecteur aléatoire gaussien $(X, Y) = g(U, V)$ construit aux deux questions précédentes à partir de U et V i.i.d. $\mathfrak{N}(0, 1)$ et de la matrice de covariance $M = K$ a pour espérance $(0, 0)$ et pour matrice de covariance K. Il a donc *même loi* que (X', Y'). On sait que si g est linéaire bijective, $g(U, V)$ a une densité qui se déduit de celle de (U, V) par la formule de changement de variable linéaire pour les densités. Pour vérifier la bijectivité de g, on note que

$$\det g = \sigma_1 \sigma_2 \sqrt{1-\rho^2} = \sqrt{ab(1-\rho^2)}$$

et

$$1 - \rho^2 = \frac{ab - c^2}{ab}$$

d'où

$$\det g = \sqrt{\det K} \neq 0.$$

D'après la fomule de changement de variable linéaire pour les densités, (X, Y) admet pour densité :

$$f_{X,Y}(x,y) = \frac{1}{|\det g|} f_{U,V}(g^{-1}(x,y)).$$

L'application linéaire inverse de g se calcule en résolvant le système (11.61) en considérant u et v comme des inconnues :

$$\begin{aligned} u &= \frac{x}{\sigma_1} \\ v &= \frac{-\sigma_2\rho x + \sigma_1 y}{\sigma_1\sigma_2\sqrt{1-\rho^2}} \end{aligned}$$

Pour alléger les écritures, calculons séparément :

$$\begin{aligned} u^2 + v^2 &= \frac{x^2}{\sigma_1^2} + \frac{(-\sigma_2\rho x + \sigma_1 y)^2}{\sigma_1^2\sigma_2^2(1-\rho^2)} \\ &= \frac{\sigma_2^2(1-\rho^2)x^2 + \sigma_2^2\rho^2 x^2 + \sigma_1^2 y^2 - 2\rho\sigma_1\sigma_2 xy}{\sigma_1^2\sigma_2^2(1-\rho^2)} \\ &= \frac{\sigma_2^2 x^2 + \sigma_1^2 y^2 - 2\rho\sigma_1\sigma_2 xy}{\sigma_1^2\sigma_2^2(1-\rho^2)} \end{aligned}$$

Finalement, la densité de (X, Y) est donnée par

$$f_{X,Y}(x,y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left(-\frac{\sigma_2^2 x^2 + \sigma_1^2 y^2 - 2\rho\sigma_1\sigma_2 xy}{2\sigma_1^2\sigma_2^2(1-\rho^2)}\right).$$

Comme (X', Y') a même loi que (X, Y), il admet lui aussi pour densité $f_{X,Y}$.

6) Vérifions finalement que si $\det K = 0$, soit le vecteur aléatoire (X', Y') est p.s. constant égal à $(0, 0)$, soit sa loi est portée par une droite D, c.-à-d. $P((X', Y') \in D) = 1$. En notant encore $a = \sigma_1^2$, $b = \sigma_2^2$, $c = \mathrm{Cov}(X', Y')$, l'hypothèse s'écrit $ab = c^2$.

Nous avons déjà examiné le cas $ab = 0$ à la question 2), revoyons le en prenant en compte le caractère gaussien de (X', Y').

1. Si $a = b = 0$, alors le vecteur gaussien constant $(0, 0)$ a pour espérance $(0, 0)$ et pour matrice de covariance K. Comme le loi gaussienne de (X', Y') est entièrement déterminée par son vecteur espérance et sa matrice de covariance, (X', Y') a *même loi* que le vecteur constant $(0, 0)$ et en particulier toute la masse de sa loi est portée par le point $(0, 0)$, ce qui s'écrit aussi $P((X', Y') = (0, 0)) = 1$.
2. Si $a = 0$ et $b > 0$, considérons le vecteur $(0, Y)$ où Y est gaussienne $\mathfrak{N}(0, \sigma_2)$. Ses composantes sont gaussiennes et indépendantes puisqu'une v.a. constante est toujours indépendante de n'importe quelle v.a. y compris d'elle même [27]. Donc $(0, Y)$ est gaussien. Comme il a même vecteur espérance et même matrice de covariance que (X', Y'), ces deux vecteurs gaussiens ont même loi. En particulier $P(X' = 0) = P((X', Y') \in \{0\} \times \mathbb{R}) = P((0, Y) \in \{0\} \times \mathbb{R}) = 1$, donc (X', Y') appartient avec probabilité 1 à la droite $D = \{0\} \times \mathbb{R}$ d'équation $x = 0$.

27. Si vous l'ignoriez, vérifiez-le en exercice.

3. Si $a > 0$ et $b = 0$, on vérifie de même que (X', Y') appartient avec probabilité 1 à la droite d'équation $y = 0$.

Il nous reste à étudier le cas $ab = c^2 > 0$. On peut alors définir le coefficient de corrélation $\rho = c(ab)^{-1/2}$ et noter que nécessairement $\rho = \pm 1$. L'application linéaire g définie par (11.61) se réécrit plus simplement :

$$g : \mathbb{R}^2 \to \mathbb{R}^2, \quad (u, v) \longmapsto (x, y), \text{ avec } \begin{cases} x = \sigma_1 u, \\ y = \sigma_2 \rho u. \end{cases}$$

Elle n'est plus bijective ($\det g = 0$) et elle envoie $\mathbb{R}^2$ sur la droite D d'équation

$$y = \frac{\sigma_2 \rho}{\sigma_1} x.$$

En notant encore (U, V) un vecteur gaussien à composantes indépendantes et de loi $\mathfrak{N}(0, 1)$, nous savons que le vecteur aléatoire $(X, Y) = g(U, V)$ est gaussien d'espérance $\mathbf{E}g(U, V) = g(\mathbf{E}U, \mathbf{E}V) = g(0, 0) = (0, 0)$ et de matrice de covariance K, donc de même loi que (X', Y'). En particulier,

$$P((X', Y') \in D) = P((X, Y) \in D) = P(g(U, V) \in D) = 1.$$

Solution 10.18 *Inversions d'une permutation aléatoire*

On note Ω_n l'ensemble des permutations de $[\![1, n]\!]$, c'est à dire des bijections $\omega : [\![1, n]\!] \to [\![1, n]\!]$. Soient $k = \omega(i) \in [\![1, n]\!]$. On dit que k présente une inversion avec $\ell = \omega(j)$ si $j > i$ et $\ell < k$. Ainsi le nombre d'inversions dans ω de $k = \omega(i)$ est le nombre d'éléments inférieurs à $\omega(i)$ et situés à sa droite dans la liste $\omega(1), \ldots, \omega(i), \ldots, \omega(n)$. Le nombre d'inversions de k est ainsi toujours compris entre 0 et $k - 1$, le nombre d'inversions de 1 est toujours nul.

Le but de cet exercice est d'établir une loi faible des grands nombres et un théorème limite central pour le nombre total d'inversions d'une permutation aléatoire. Notons pour cela $X_k(\omega)$ le nombre d'inversions de k dans ω, pour $k \in [\![1, n]\!]$ et $S_n(\omega) = \sum_{k=1}^{n} X_k(\omega)$.

1) Lorsque n n'est pas trop grand, on peut représenter ω en donnant la liste explicite des $\omega(i)$. Par commodité, nous écrirons chaque $k \in [\![1, n]\!]$ sur un jeton carré et chaque disposition sur une même ligne des n jetons représentera une permutation ω. Par exemple

4	3	5	1	2

représente la permutation ω définie par : $1 \mapsto 4$, $2 \mapsto 3$, $3 \mapsto 5$, $4 \mapsto 1$, $5 \mapsto 2$. L'élément $4 = \omega(1)$ a 3 inversions, puisqu'à sa droite on trouve 3, 1 et 2 qui lui sont inférieurs, donc $X_4(\omega) = 3$. De proche en proche, on voit ainsi que le vecteur d'inversions associé à ω est $(X_1(\omega), \ldots, X_5(\omega)) = (0, 0, 2, 3, 2)$.

Il y a une seule permutation ayant pour vecteur d'inversions $(0, 1, 0, 3, 2)$. Sa construction par placements successifs des 5 jetons est donnée par le tableau :

Jeton n° k	$X_k(\omega)$	Placement
1	0	$\boxed{1}$
2	1	$\boxed{2}\,\boxed{1}$
3	0	$\boxed{2}\,\boxed{1}\,\boxed{3}$
4	3	$\boxed{4}\,\boxed{2}\,\boxed{1}\,\boxed{3}$
5	2	$\boxed{4}\,\boxed{2}\,\boxed{5}\,\boxed{1}\,\boxed{3}$

De même, il y a une seule permutation ayant pour vecteur d'inversions $(0, 0, 1, 2, 0)$. Sa construction est donnée par le tableau :

Jeton n° k	$X_k(\omega)$	Placement
1	0	$\boxed{1}$
2	0	$\boxed{1}\,\boxed{2}$
3	1	$\boxed{1}\,\boxed{3}\,\boxed{2}$
4	2	$\boxed{1}\,\boxed{4}\,\boxed{3}\,\boxed{2}$
5	0	$\boxed{1}\,\boxed{4}\,\boxed{3}\,\boxed{2}\,\boxed{5}$

Sur ces deux exemples, on remarque que lors du placement du jeton n° k, le nombre d'inversions $X_k(\omega)$ ne laisse qu'une possibilité de placement du jeton n° k, par rapport aux jetons numérotés de 1 à $k-1$ déjà placés.

2) On note Δ_n le produit cartésien

$$\Delta_n = \{0\} \times [\![0, 1]\!] \times [\![0, 2]\!] \times \cdots \times [\![0, n-1]\!].$$

Tout vecteur d'inversions d'une permutation $\omega \in \Omega_n$ est dans Δ_n puisque, pour chaque $k \in [\![1, n]\!]$, le nombre d'inversions de k dans ω est le nombre d'entiers situés à sa droite dans la liste $\omega(1), \ldots, \omega(n)$ et inférieurs strictement à k. Ce nombre d'inversions de k est donc au plus $k-1$. Autrement dit, pour $k \in [\![1, n]\!]$, $X_k(\omega) \in [\![0, k-1]\!]$, d'où

$$(X_1(\omega), \ldots, X_n(\omega)) \in \prod_{k=1}^{n} [\![0, k-1]\!] = \Delta_n$$

et ceci est valide pour tout $\omega \in \Omega_n$.

Nous allons montrer que pour tout $x \in \Delta_n$, il existe *un unique* $\omega \in \Omega_n$ tel que $(X_1(\omega), \ldots, X_n(\omega)) = x$. Autrement dit que l'application $\omega \mapsto (X_1(\omega), \ldots, X_n(\omega))$ est une bijection de Ω_n sur Δ_n.

Les deux exemples étudiés à la question précédente indiquent la voie à suivre : à partir d'un vecteur x quelconque de Δ_n, on construit pas à pas une permutation ω ayant x pour vecteur d'inversions et on voit que cette construction nous donne à la fois l'existence et l'unicité.

- On commence par poser le jeton $\boxed{1}$.
- Si $x_2 = 0$, $\boxed{2}$ n'a pas d'inversion et on ne peut le placer qu'à la droite de $\boxed{1}$. Sinon, on le place à sa gauche.

– Si $x_3 = 0$, $\boxed{3}$ n'a pas d'inversion et on ne peut le placer qu'à la droite des deux premiers jetons déjà placés ; si $x_3 = 1$, $\boxed{3}$ a une seule inversion et on ne peut le placer qu'entre les deux jetons déjà placés ; si $x_3 = 2$, $\boxed{3}$ a deux inversions et on ne peut le placer qu'à la gauche des deux jetons déjà placés.
– ...
– Les jetons numérotés de 1 à $k-1$ étant déjà placés, la valeur j de x_k ($0 \leqslant j \leqslant k-1$) nous indique la seule possibilité de placement pour $\boxed{k}$: comme il doit présenter j inversions, on ne peut le mettre qu'en j^{e} position à partir de la droite, c'est-à-dire à la doite des $k-1$ jetons déjà placés si $j = 0$, intercalé entre les deux plus à droite si $j = 1$, etc.
– ...

On peut ainsi construire ω de proche en proche jusqu'à épuisement des n jetons. À chaque étape k, parmi les k positions relatives du jeton $\boxed{k}$ par rapport aux $k-1$ jetons déjà placés, une seule est compatible avec la valeur de x_k. Ceci nous donne bien l'existence et l'unicité de ω ayant pour vecteur d'inversions x.

3) On munit Ω_n de la tribu $\mathcal{F}_n = \mathcal{P}(\Omega_n)$ de tous ses sous-ensembles et de la probabilité P_n, loi uniforme sur Ω_n. On a donc $P_n(\{\omega\}) = \frac{1}{n!}$ pour tout $\omega \in \Omega_n$. On en déduit immédiatement que

$$\forall x \in \Delta_n, \quad P_n((X_1, \dots, X_n) = x) = \frac{1}{n!},$$

puisque pour chaque $x \in \Delta_n$, il existe un unique $\omega \in \Omega_n$ tel que $(X_1(\omega), \dots, X_n(\omega)) = x$, donc pour cet ω, $\{(X_1, \dots, X_n) = x\} = \{\omega\}$. Or $P_n(\{\omega\}) = \frac{1}{n!}$ pour tout $\omega \in \Omega_n$.

Ainsi le vecteur aléatoire discret $(X_1, \dots, X_n)$ peut prendre toute valeur vectorielle x dans Δ_n et nulle part ailleurs (d'après la question précédente). Il prend chacune de ces valeurs x avec *la même* probabilité $\frac{1}{n!}$. Sa loi est donc la *loi uniforme* (discrète) sur l'ensemble fini Δ_n.

Une autre façon de voir que $(X_1, \dots, X_n)$ suit la loi uniforme sur Δ_n est de remarquer que :

$$\operatorname{card} \Delta_n = \prod_{k=1}^{n} \operatorname{card} [\![0, k-1]\!] = n!$$

d'où

$$\forall x \in \Delta_n, \quad P_n((X_1, \dots, X_n) = x) = \frac{1}{\operatorname{card} \Delta_n}.$$

4) Soit un espace probabilisé $(\Omega', \mathcal{F}', P')$ sur lequel on puisse définir des variables aléatoires discrètes $U_1, \dots, U_n$ indépendantes et telles que pour tout $k \in [\![1, n]\!]$, U_k suive la loi uniforme sur $[\![0, k-1]\!]$. L'énoncé ne demandait pas de vérifier l'existence d'un tel espace [28]. Pour identifier la loi du vecteur aléatoire discret $U = (U_1, \dots, U_n)$,

28. On peut vérifier à titre d'exercice qu'une solution possible est de prendre $\Omega' = \Delta_n$, $\mathcal{F}'$ l'ensemble des parties de Δ_n, P' la probabilité uniforme sur Δ_n (donnée par $P'(\{x\}) = \frac{1}{n!}$) avec pour $U = (U_1, \dots, U_n)$ l'identité sur Δ_n.

on calcule pour tout $x = (x_1, \dots, x_n) \in \Delta_n$, $P'\big((U_1, \dots, U_n) = (x_1, \dots, x_n)\big)$:

$$\begin{aligned} P'\big((U_1, \dots, U_n) = (x_1, \dots, x_n)\big) &= P'\left(U_1 = x_1, \dots, U_n = x_n\right) \\ &= \prod_{k=1}^{n} P'(U_k = x_k) \quad \text{(par indépendance des } U_k\text{)} \\ &= \prod_{k=1}^{n} \frac{1}{k} \quad \text{(car } U_k \text{ suit la loi uniforme sur } [\![0, k-1]\!]\text{)} \\ &= \frac{1}{n!} = \frac{1}{\operatorname{card} \Delta_n}. \end{aligned}$$

Ceci étant vrai pour tout $x = (x_1, \dots, x_n) \in \Delta_n$, on en conclut que le vecteur aléatoire U suit la loi uniforme sur Δ_n.

Ainsi les vecteurs aléatoires $(X_1, \dots, X_n)$ et $(U_1, \dots, U_n)$ ont même loi. Ceci implique que chaque X_k a même loi que U_k, donc suit la loi uniforme sur $[\![0, k-1]\!]$. De plus comme les composantes de $(U_1, \dots, U_n)$ sont indépendantes, il en va de même pour celles de $(X_1, \dots, X_n)$. Cela devrait vous paraître évident si vous êtes familier avec les lois produits : un vecteur aléatoire est à composantes indépendantes si et seulement si sa loi est le produit de ses lois marginales, donc le fait d'être à composantes indépendantes est une propriété qui ne dépend que de la loi du vecteur. Sinon, il suffit d'écrire :

$$\begin{aligned} P\big((X_1, \dots, X_n) = (x_1, \dots, x_n)\big) &= P'\big((U_1, \dots, U_n) = (x_1, \dots, x_n)\big) \\ &= \prod_{k=1}^{n} P'(U_k = x_k) \quad \text{(par indépendance des } U_k\text{)} \\ &= \prod_{k=1}^{n} P(X_k = x_k) \quad \text{(car } U_k \text{ et } X_k \text{ ont même loi).} \end{aligned}$$

5) Justification des formules sommatoires

$$T_{n,1} := \sum_{j=1}^{n} j = \frac{n(n+1)}{2} \tag{11.65}$$

et

$$T_{n,2} := \sum_{j=1}^{n} j^2 = \frac{n(n+1)(2n+1)}{6}. \tag{11.66}$$

Ces deux formules peuvent se vérifier par récurrence. Il y a une méthode plus simple pour la première, illustrée par le tableau suivant :

1	2	3	$\dots$	k	$\dots$	$n-2$	$n-1$	n	$T_{n,1}$
n	$n-1$	$n-2$	$\dots$	$n-k+1$	$\dots$	3	2	1	$T_{n,1}$
$n+1$	$n+1$	$n+1$	$\dots$	$n+1$	$\dots$	$n+1$	$n+1$	$n+1$	$2T_{n,1}$

qui nous montre en sommant horizontalement et verticalement que $2T_{n,1} = n(n+1)$, ce qui donne (11.65).

Pour tout entier r, notons $T_{n,r} = \sum_{j=1}^{n} j^r$. Pour calculer $T_{n,2}$, on cherche une relation de récurrence entre $T_{n+1,3}$ et $T_{n,3}$ faisant intervenir $T_{n,2}$ et $T_{n,1}$. Pour cela, on note que

$$T_{n+1,3} = \sum_{j=1}^{n+1} j^3 = 1 + \sum_{k=1}^{n} (k+1)^3 = 1 + \sum_{k=1}^{n} (k^3 + 3k^2 + 3k + 1)$$
$$= T_{n,3} + 3T_{n,2} + 3T_{n,1} + (n+1).$$

En notant que l'on a directement $T_{n+1,3} = T_{n,3} + (n+1)^3$, on en déduit :

$$(n+1)^3 - (n+1) = 3T_{n,2} + 3T_{n,1}$$

d'où

$$3T_{n,2} = (n+1)((n+1)^2 - 1) - \frac{3}{2}n(n+1) = (n+1)(n^2 + 2n - \frac{3}{2}n) = n(n+1)(n+\frac{1}{2}),$$

ce qui nous donne bien (11.66).

On aurait pu utiliser la même méthode pour calculer $T_{n,1}$ en comparant $T_{n+1,2}$ et $T_{n,2}$.

6) Ces formules sommatoires vont nous servir à calculer l'espérance et la variance de $S_n = X_1 + \cdots + X_n$.

Puisque X_k suit la loi uniforme sur $[\![0, k-1]\!]$, son espérance est :

$$\mathbf{E}X_k = \sum_{j=0}^{k-1} jP(X_k = j) = \frac{1}{k}\sum_{j=1}^{k-1} j = \frac{1}{k}T_{k-1,1} = \frac{1}{k}\frac{(k-1)k}{2} = \frac{k-1}{2}.$$

Par additivité de l'espérance, on en déduit :

$$\mathbf{E}S_n = \sum_{k=1}^{n} \mathbf{E}X_k = \sum_{k=1}^{n} \frac{k-1}{2} = \frac{1}{2}T_{n-1,1} = \frac{n(n-1)}{4}.$$

Par indépendance des X_k, $\operatorname{Var} S_n = \operatorname{Var} X_1 + \cdots + \operatorname{Var} X_n$. Pour calculer la variance de X_k, on commence par calculer $\mathbf{E}X_k^2$:

$$\mathbf{E}X_k^2 = \sum_{j=0}^{k-1} j^2 P(X_k = j) = \frac{1}{k}\sum_{j=1}^{k-1} j^2 = \frac{1}{k}T_{k-1,2} = \frac{(k-1)(2k-1)}{6}.$$

Ensuite,

$$\operatorname{Var} X_k = \mathbf{E}X_k^2 - (\mathbf{E}X_k)^2 = \frac{(k-1)(2k-1)}{6} - \frac{(k-1)^2}{4}$$
$$= \frac{k-1}{2}\Big(\frac{2k-1}{3} - \frac{k-1}{2}\Big)$$
$$= \frac{(k-1)(k+1)}{12} = \frac{k^2-1}{12}.$$

Finalement,

$$\begin{aligned}\operatorname{Var} S_n = \sum_{k=1}^{n} \frac{k^2-1}{12} = \frac{1}{12}(T_{n,2}-n) &= \frac{1}{12}\left(\frac{n(n+1)(2n+1)}{6} - n\right)\\ &= \frac{n}{72}\big((n+1)(2n+1)-6\big)\\ &= \frac{n}{72}(2n^2+3n-5)\\ &= \frac{n(n-1)(2n+5)}{72}.\end{aligned}$$

7) Puisque $\mathbf{E}S_n$ est équivalent à $n^2/4$ quand n tend vers l'infini, il est naturel de se demander si S_n normalisé par n^2 satisfait une loi faible des grands nombres avec limite $1/4$. Plus précisément on regarde si la convergence « en probabilité [29] » suivante est vérifiée :

$$\forall \varepsilon > 0, \quad P_n\left(\left|\frac{S_n}{n^2} - \frac{1}{4}\right| > \varepsilon\right) \xrightarrow[n\to+\infty]{} 0. \tag{11.67}$$

L'idée est de se ramener à une inégalité de Tchebycheff en procédant comme suit :

$$\begin{aligned}P_n\left(\left|\frac{S_n}{n^2} - \frac{1}{4}\right| > \varepsilon\right) &= P_n\left(\left|S_n - \frac{n^2}{4}\right| > \varepsilon n^2\right)\\ &\leqslant P_n\left(|S_n - \mathbf{E}S_n| + \left|\mathbf{E}S_n - \frac{n^2}{4}\right| > \varepsilon n^2\right)\\ &\leqslant P_n\left(|S_n - \mathbf{E}S_n| > \frac{\varepsilon n^2}{2}\right) + P_n\left(\left|\mathbf{E}S_n - \frac{n^2}{4}\right| > \frac{\varepsilon n^2}{2}\right)\end{aligned}$$

Dans le deuxième terme de ce majorant, la « variable aléatoire » $|\mathbf{E}S_n - n^2/4|$ est constante et vaut $n/4$. La probabilité correspondante vaut donc 1 si $n/4 > \varepsilon n^2/2$ et 0 sinon. Il est donc clair qu'elle vaudra toujours zéro à partir d'un certain rang n_1. On peut prendre $n_1 \geqslant (2\varepsilon)^{-1}$.

Pour majorer le premier terme, on utilise l'inégalité de Tchebycheff :

$$P_n\left(|S_n - \mathbf{E}S_n| > \frac{\varepsilon n^2}{2}\right) \leqslant \frac{4\operatorname{Var} S_n}{\varepsilon^2 n^4} = \frac{n(n-1)(2n+5)}{18\varepsilon^2 n^4} \sim \frac{1}{9\varepsilon^2 n},$$

ce qui montre que $P_n\left(|S_n - \mathbf{E}S_n| > \frac{\varepsilon n^2}{2}\right)$ tend vers 0 quand n tend vers l'infini et complète la preuve de (11.67).

8) Dans notre problème, les espaces probabilisés $(\Omega_n, \mathcal{F}_n, P_n)_{n\geqslant 1}$ du théorème de Lindeberg rappelé dans l'énoncé sont ceux définis à la question 3. La v.a. notée X_k, quand on travaillait à n fixé, est notée maintenant $X_{n,k}$ (elle dépend de n puisqu'elle s'applique à des permutations ω sur $[\![1, n]\!]$).

29. Il s'agit ici d'une notion généralisée de convergence en probabilité puisque chaque S_n est définie sur un espace probabilisé dépendant de n et muni d'une probabilité P_n.

L'écart type s_n est ici équivalent à $n^{3/2}/6$ quand n tend vers l'infini d'après le calcul de $\operatorname{Var} S_n$ effectué ci-dessus. Or $X_{n,i}$ est à valeurs dans $[\![0, i-1]\!]$, ce qui implique que

$$|X_{n,i} - \mathbf{E}X_{n,i}| \leqslant i - 1 \leqslant n$$

Comme s_n tend vers l'infini plus vite que ce majorant n, il en va de même pour εs_n. Par conséquent il existe un rang $n_0 = n_0(\varepsilon)$ à partir duquel l'inégalité $|X_{n,i} - \mathbf{E}X_{n,i}| > \varepsilon s_n$ n'est vraie pour aucun $n \geqslant n_0$, aucun $i \leqslant n$ et aucun $\omega \in \Omega_n$. Il en résulte que pour $n \geqslant n_0(\varepsilon)$, toutes les indicatrices dans $R_{n,\varepsilon}$ sont nulles (sur Ω_n tout entier) et donc $R_{n,\varepsilon} = 0$ pour tout $n \geqslant n_0(\varepsilon)$. Comme ceci est valide pour tout $\varepsilon > 0$, la condition de Lindeberg (10.23) est trivialement vérifiée. Par le théorème de Lindeberg on a donc la convergence en loi :

$$S_n^\star := \frac{1}{s_n}(S_n - \mathbf{E}S_n) \xrightarrow[n\to+\infty]{\text{loi}} Z,$$

où Z est gaussienne $\mathfrak{N}(0,1)$.

Pour en déduire la convergence en loi vers Z de $6n^{-3/2}(S_n - n^2/4)$, on rappelle que $\mathbf{E}S_n - n^2/4 = -n/4$, d'où

$$\begin{aligned}\frac{6}{n^{3/2}}\left(S_n - \frac{n^2}{4}\right) = \frac{6}{n^{3/2}}\left(S_n - \mathbf{E}S_n - \frac{n}{4}\right) &= \frac{6}{n^{3/2}}\left(S_n - \mathbf{E}S_n\right) - \frac{3}{2n^{1/2}} \\ &= \frac{6s_n}{n^{3/2}}S_n^\star - \frac{3}{2n^{1/2}}.\end{aligned}$$

On obtient ainsi une expression de la forme :

$$\frac{6}{n^{3/2}}\left(S_n - \frac{n^2}{4}\right) = a_n S_n^\star + b_n, \quad \lim_{n\to+\infty} a_n = 1, \quad \lim_{n\to+\infty} b_n = 0.$$

En considérant a_n et b_n comme des variables aléatoires constantes, les limites ci-dessus nous donnent trivialement la convergence en probabilité de a_n vers 1 et la convergence en probabilité de b_n vers 0. En appliquant deux fois le lemme de Slutsky, on en déduit successivement que $a_n S_n^\star$ converge en loi vers Z, puis que $a_n S_n^\star + b_n$ converge en loi vers Z.

9) Nous venons de prouver la convergence en loi

$$\frac{6}{n^{3/2}}\left(S_n - \frac{n^2}{4}\right) \xrightarrow[n\to+\infty]{\text{loi}} Z,$$

qui implique, puisque la fonction de répartition Φ de la gaussienne standard Z est continue sur $\mathbb{R}$ que

$$\forall t > 0, \quad P_n\left(\frac{6}{n^{3/2}}\left|S_n - \frac{n^2}{4}\right| \leqslant t\right) \xrightarrow[n\to+\infty]{\text{loi}} P(|Z| \leqslant t) = \Phi(t) - \Phi(-t) = 2\Phi(t) - 1.$$

Ceci justifie pour n grand, l'approximation :

$$P_n\left(\frac{n^2}{4} - \frac{tn^{3/2}}{6} \leqslant S_n \leqslant \frac{n^2}{4} + \frac{tn^{3/2}}{6}\right) \simeq 2\Phi(t) - 1. \tag{11.68}$$

Cherchons t tel que $2\Phi(t) - 1 = 0{,}5$, ce qui équivaut à $\Phi(t) = 0{,}75$. La table donne

$$\Phi(0{,}67) = 0{,}748\,6, \quad \Phi(0{,}68) = 0{,}751\,7.$$

On en déduit une valeur approchée de t par interpolation affine :

$$t \simeq 0{,}67 + \frac{0{,}75 - 0{,}748\,6}{0{,}751\,7 - 0{,}748\,6}(0{,}68 - 0{,}67) = 0{,}67 + \frac{14}{31} \times 0{,}01 \simeq 0{,}674\,5.$$

En reportant cette valeur dans (11.68), on en déduit que pour n grand,

$$P_n\left(\frac{n^2}{4} - 0{,}113\,n^{3/2} \leqslant S_n \leqslant \frac{n^2}{4} + 0{,}113\,n^{3/2}\right) \simeq \frac{1}{2}$$

Et comme P_n est la probabilité *uniforme* sur l'espace Ω_n de toutes les permutations de $[\![1, n]\!]$, cela signifie bien que « pour n grand, environ la moitié des permutations sur $[\![1, n]\!]$ ont leur nombre total d'inversions compris entre les bornes $\dfrac{n^2}{4} \pm 0{,}113\,n^{3/2}$ ».

La source de cet exercice se trouve dans le livre de W. Feller [5], tome 1, chap. X, section 6, où elle est traitée en une demi-page.

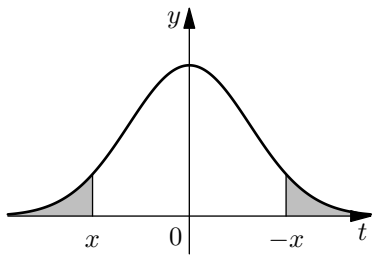
Tables de la loi normale

La table de la page 520 donne des valeurs approchées à 10^{-4} près de $\Phi(x)$ pour $x = \frac{k}{100}$, $k \in [\![0, 299]\!]$. Pour lire la valeur de $\Phi(x)$, on repère dans la première colonne la ligne qui comporte le chiffre des unités et celui des dixièmes de x. À l'intersection de cette ligne et de la colonne dont l'entête est le chiffre des centièmes de x, on lit la valeur de $\Phi(x)$. Par exemple pour $\Phi(1{,}37)$, on lit à l'intersection de la ligne commençant par 1,3 et de la colonne commençant par 0,07 la valeur $\Phi(1{,}37) \simeq 0{,}914\,7$.

La table donne les valeurs de $\Phi(x)$ pour x positif. Lorsque x est négatif, on utilise la relation

$$\Phi(x) = 1 - \Phi(-x)$$

qui résulte de la parité de la densité gaussienne $\mathfrak{N}(0, 1)$, voir ci-contre. Exemple : pour $x = -1,8$, on trouve : $\Phi(x) = 1 - \Phi(1{,}8) \simeq 1 - 0,9641 = 0{,}035\,9$.

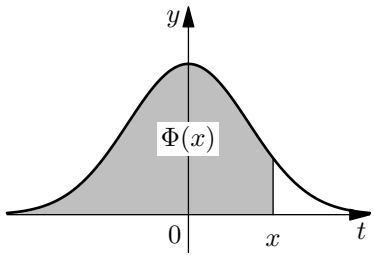


Pour les « très grandes valeurs de x », disons $|x| > 4$, on peut utiliser l'encadrement suivant de la « queue » de la loi normale.

$$\forall x > 0, \quad \left(\frac{1}{x} - \frac{1}{x^3}\right) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) < 1 - \Phi(x) < \frac{1}{x} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right).$$

Pour une preuve, voir l'exercice 4.3.

Table des valeurs de Φ, f.d.r. de la loi normale standard $\mathfrak{N}(0,1)$



x	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,500 0	0,504 0	0,508 0	0,512 0	0,516 0	0,519 9	0,523 9	0,527 9	0,531 9	0,535 9
0,1	0,539 8	0,543 8	0,547 8	0,551 7	0,555 7	0,559 6	0,563 6	0,567 5	0,571 4	0,575 3
0,2	0,579 3	0,583 2	0,587 1	0,591 0	0,594 8	0,598 7	0,602 6	0,606 4	0,610 3	0,614 1
0,3	0,617 9	0,621 7	0,625 5	0,629 3	0,633 1	0,636 8	0,640 6	0,644 3	0,648 0	0,651 7
0,4	0,655 4	0,659 1	0,662 8	0,666 4	0,670 0	0,673 6	0,677 2	0,680 8	0,684 4	0,687 9
0,5	0,691 5	0,695 0	0,698 5	0,701 9	0,705 4	0,708 8	0,712 3	0,715 7	0,719 0	0,722 4
0,6	0,725 7	0,729 1	0,732 4	0,735 7	0,738 9	0,742 2	0,745 4	0,748 6	0,751 7	0,754 9
0,7	0,758 0	0,761 1	0,764 2	0,767 3	0,770 4	0,773 4	0,776 4	0,779 4	0,782 3	0,785 2
0,8	0,788 1	0,791 0	0,793 9	0,796 7	0,799 5	0,802 3	0,805 1	0,807 8	0,810 6	0,813 3
0,9	0,815 9	0,818 6	0,821 2	0,823 8	0,826 4	0,828 9	0,831 5	0,834 0	0,836 5	0,838 9
1,0	0,841 3	0,843 8	0,846 1	0,848 5	0,850 8	0,853 1	0,855 4	0,857 7	0,859 9	0,862 1
1,1	0,864 3	0,866 5	0,868 6	0,870 8	0,872 9	0,874 9	0,877 0	0,879 0	0,881 0	0,883 0
1,2	0,884 9	0,886 9	0,888 8	0,890 7	0,892 5	0,894 4	0,896 2	0,898 0	0,899 7	0,901 5
1,3	0,903 2	0,904 9	0,906 6	0,908 2	0,909 9	0,911 5	0,913 1	0,914 7	0,916 2	0,917 7
1,4	0,919 2	0,920 7	0,922 2	0,923 6	0,925 1	0,926 5	0,927 9	0,929 2	0,930 6	0,931 9
1,5	0,933 2	0,934 5	0,935 7	0,937 0	0,938 2	0,939 4	0,940 6	0,941 8	0,942 9	0,944 1
1,6	0,945 2	0,946 3	0,947 4	0,948 4	0,949 5	0,950 5	0,951 5	0,952 5	0,953 5	0,954 5
1,7	0,955 4	0,956 4	0,957 3	0,958 2	0,959 1	0,959 9	0,960 8	0,961 6	0,962 5	0,963 3
1,8	0,964 1	0,964 9	0,965 6	0,966 4	0,967 1	0,967 8	0,968 6	0,969 3	0,969 9	0,970 6
1,9	0,971 3	0,971 9	0,972 6	0,973 2	0,973 8	0,974 4	0,975 0	0,975 6	0,976 1	0,976 7
2,0	0,977 2	0,977 8	0,978 3	0,978 8	0,979 3	0,979 8	0,980 3	0,980 8	0,981 2	0,981 7
2,1	0,982 1	0,982 6	0,983 0	0,983 4	0,983 8	0,984 2	0,984 6	0,985 0	0,985 4	0,985 7
2,2	0,986 1	0,986 4	0,986 8	0,987 1	0,987 5	0,987 8	0,988 1	0,988 4	0,988 7	0,989 0
2,3	0,989 3	0,989 6	0,989 8	0,990 1	0,990 4	0,990 6	0,990 9	0,991 1	0,991 3	0,991 6
2,4	0,991 8	0,992 0	0,992 2	0,992 5	0,992 7	0,992 9	0,993 1	0,993 2	0,993 4	0,993 6
2,5	0,993 8	0,994 0	0,994 1	0,994 3	0,994 5	0,994 6	0,994 8	0,994 9	0,995 1	0,995 2
2,6	0,995 3	0,995 5	0,995 6	0,995 7	0,995 9	0,996 0	0,996 1	0,996 2	0,996 3	0,996 4
2,7	0,996 5	0,996 6	0,996 7	0,996 8	0,996 9	0,997 0	0,997 1	0,997 2	0,997 3	0,997 4
2,8	0,997 4	0,997 5	0,997 6	0,997 7	0,997 7	0,997 8	0,997 9	0,997 9	0,998 0	0,998 1
2,9	0,998 1	0,998 2	0,998 2	0,998 3	0,998 4	0,998 4	0,998 5	0,998 5	0,998 6	0,998 6

Table pour les grandes valeurs de x

x	3,0	3,1	3,2	3,3	3,4	3,5	3,6	3,8	4,0
$\Phi(x)$	0,998 65	0,999 04	0,999 31	0,999 52	0,999 66	0,999 76	0,999 841	0,999 928	0,999 968

Bibliographie

[1] Ph. Barbe et M. Ledoux. *Probabilité*. EDP Sciences, 2007. ISBN : 978-2-86883-931-2.

[2] J. Dauxois et C. Hassenforder. *Toutes les probabilités et les statistiques, Cours et exercices corrigés*. Ellipses, 2004. ISBN : 2-7298-2163-5.

[3] P. Deheuvels. *L'intégrale*. Presses Universitaires de France, 1980. ISBN : 2-13-036648-1.

[4] *Enseignement spécifique et de spécialité de mathématiques de la série scientifique - classe terminale*. Bulletin officiel spécial n° 8 du 13 octobre 2011. URL : `http://www.education.gouv.fr/pid25535/bulletin_officiel.html?cid_bo=57529`.

[5] W. Feller. *An introduction to probability theory and its application*. Anglais. 2e éd. T. 1. Wiley, 1957.

[6] W. Feller. *An introduction to probability theory and its application*. Anglais. 2e éd. T. 2. Wiley, 1966.

[7] D. Foata, A. Fuchs et J. Franchi. *Calcul des probabilités*. 3e éd. Dunod, 2012. ISBN : 978-2-10-057424-7.

[8] O. Garet et A. Kurtzmann. *De l'intégration aux probabilités*. Ellipses, 2011. ISBN : 9782729870409.

[9] S. Lang. *Analysis I*. Anglais. Addison-Wesley, 1969.

[10] M. Ledoux et M. Talagrand. *Probability in Banach Spaces*. Anglais. Berlin, Heidelberg : Springer-Verlag, 1991.

[11] D. Pollard. *Convergence of stochastic processes*. Anglais. Springer, 1984.

[12] S. M. Ross. *Initiation aux probabilités*. Traduction de la septième édition américaine. Lausanne : Presses polytechniques et universitaires romandes, 2007. ISBN : 978-2-88074-738-1.

[13] L. Schwartz. *Topologie générale et analyse fonctionnelle*. Hermann, 1970.

[14] Ch. Suquet. *Introduction au calcul des probabilités*. UFR de Mathématiques, Université Lille 1, fév. 2012. URL : `http://math.univ-lille1.fr/~suquet/Polys/ICP.pdf`.

[15] Ch. SUQUET. *Intégration, Analyse de Fourier, Probabilités*. UFR de Mathématiques, Université Lille 1, fév. 2012. URL : `http://math.univ-lille1.fr/~suquet/Polys/IFP.pdf`.

[16] J.V. USPENSKY. *Introduction to mathematical probability*. Anglais. McGraw-Hill, 1937.

Index

absence de mémoire, 235
additivité
    d'une mesure, 156
    de l'espérance, 275
    de l'intégrale de Riemann, 83
admissible
    ensemble, 138
    fonction, 136
aiguille de Buffon, 352
aire, 73, 160
    algébrique, 90
    d'hypographe, 74, 90
    et intégrale de Riemann, 74
algorithme du rejet, 193, 332
arrangement, 20
autonormalisation, 378

Bayes (formule de), 186
Beppo Levi (théorème de), 265
bijection, 16
    réciproque, 16
borélien, 157
borélienne (application), 205
Borel Cantelli, 336, 337
bornitude stochastique, 356
Buffon (aiguille de), 352

cardinal, 17, 24
    d'un produit cartésien, 18
coefficient de corrélation, 306
coefficients
    binomiaux, 21
    multinomiaux, 23
combinaison, 21
complémentaire, 14
condition
    de Liapounov, 376, 396
    de Lindeberg, 377
conditionnement
    par les cas possibles, 185
conditionnements successifs, 184
continuité monotone séquentielle, 167
convergence
    étroite, 364
    dominée, 345, 347
    en loi, 363
        composantes, 366
        par image continue, 366
    en loi (pièges), 364
    en loi dans $\mathbb{R}^d$, 364
        formes linéaires, 383
    en moyenne d'ordre p, 342–344
    en probabilité, 339, 345, 356
    presque complète, 338
    presque sûre, 334, 336, 338, 341
convergence en loi
    binomiale vers gaussienne, 370
    binomiale vers Poisson, 225, 401
    hypergéométrique vers binomiale, 222, 401
    multinomiale vers gaussienne, 384
convolution
    de deux densités, 323
corrélation, 306

covariance, 306, 309
  changement d'échelle, 306
  de v.a. indépendantes, 327
  formule de Koenig, 306
  formules de calcul, 307
  matrice de, 309
  opérateur de, 311
  translation, 306
Cramér-Wold, 383
critère de Cauchy, 110
  famille sommable, 55
  intégrales généralisées, 111
  séries, 39
croissance
  d'une mesure, 155
  de l'espérance, 260, 276
  de l'intégrale de Riemmann, 85, 133

dénombrabilité, 24
  d'un produit cartésien, 27
  d'une union, 31
  de $\mathbb{N}^2$, 25
  de $\mathbb{N}^d$, $\mathbb{Z}^d$, 28
  de $\mathbb{Q}$, 28
  de $\mathbb{Z}$, 25
  et famille sommable, 52
  par image surjective, 32
dénombrable, 24
densité
  marginale, 300
  somme de v.a.r. indépendantes, 323
  vecteur aléatoire, 299
  vecteur aléatoire image, 304
densité de probabilité
  sur $\mathbb{R}$, 214
  sur $\mathbb{R}^d$, 299
difféomorphisme C^1, 146
dyadique (nombre), 30

écart type, 282
ensemble
  $\mathcal{R}$-négligeable, 136
  admissible, 138
  au plus dénombrable, 24
  dénombrable, 24
  fini, 17
épreuves indépendantes, 192
équidistribution, 376
équiprobabilité, 174
équitension, 387
espérance d'un vecteur aléatoire, 308
  et applications linéaires, 308
  linéarité, 308
espace
  de Banach, 39
  mesurable, 156
espérance
  croissance, 260
  d'une constante positive, 256
  d'une indicatrice, 256
  d'une v.a. discrète, 274
  d'une v.a. discrète positive, 269
  d'une v.a. positive, 254
  d'une v.a. positive à densité, 259
  d'une v.a. positive simple, 258
  d'une v.a. réelle, 271
  d'une v.a. réelle à densité, 272
  interversion série-espérance, 268
  linéarité, 275
  produit de v.a. indépendantes, 325
extérieur, 138

factorielle, 373
famille sommable
  absolument, 57
famille sommable, 49
  combinaison linéaire, 49
  critère de Cauchy, 55
  normalement, 57
  permutation d'indices, 50
  sommation par paquets, 58, 59
fermeture, 138
fonction
  absolument continue, 219
  admissible sur un pavé, 136

Bêta, 153
convexe, 342
en escalier, 76
Gamma (Γ), 152
localement Riemann intégrable, 104
réglée, 82
Riemann intégrable, 72, 131, 133
fonction de répartition
d'une probabilité sur $\mathbb{R}$, 176
d'une v.a. discrète, 212
d'une variable aléatoire, 211
fonction de survie, 235
formule
de Bayes, 186
de Koenig
covariance, 306
variance, 282
de Stirling, 373
des accroissements finis, 79
du binôme, 24
du multinôme, 22
fourchette, 374
frontière, 138

hypographe, 74

i.i.d., 349
inclusion, 14
indépendance
conservation par complémentaire, 192
covariance, 327
d'une suite d'évènements, 192
d'une suite de v.a., 316
de n variables aléatoires, 312
de n vecteurs aléatoires, 312
de deux évènements, 189
de sous-tribus, 315
des composantes (cas à densité), 317
des composantes (cas discret), 316
deux à deux, 191
espérance de produits, 324
hérédité, 313
mutuelle, 191
somme de v.a., 321, 323
indicatrice, 20
dans formules explicites, 214
inégalité de Bienaymé-Tchebycheff, 348
inégalité de Cauchy-Schwarz
espérance, 305
intégrale de Riemann, 88
intégrale généralisée, 127
inégalité de Jensen, 293, 397
inégalité de Markov, 261
avec moment, 277
injection, 15
intégrabilité
d'une v.a. discrète, 274
d'une v.a. réelle, 270
d'une v.a. réelle à densité, 272
intégrable
variable aléatoire positive, 256
variable aléatoire réelle, 270
intégrale de Riemann
additivité, 83
changement de variable, 91, 93
croissance, 85, 134
de a à b ($a > b$), 72
et primitive, 78
inférieure, 69
linéarité, 84, 133
positivité, 85, 134
relation de Chasles, 89
supérieure, 69
sur $[a, b]$ ($a < b$), 72
sur A borné dans $\mathbb{R}^d$, 133
sur un pavé de $\mathbb{R}^d$, 131
intégrale généralisée, 104
absolument convergente, 112, 113, 118
additivité, 127
changement de variable, 122, 123
comparaison, 117
convergente, 104

critère d'Abel, 113
critère de Cauchy, 111
croissance, 127
divergente, 104
équivalents, 118, 120
inégalité de Cauchy-Schwarz, 127
linéarité, 127
positivité, 127
produit, 127
relation de Chasles, 127
sur $\mathbb{R}^d$, 148
sur une partie de $\mathbb{R}^d$, 147
intégrale indéfinie, 90
intérieur, 138
intersection, 14
intervalle de confiance, 374
intervalle troué, 103
interversion
limite espérance, 265, 345, 347
limite intégrale, 95, 96
produit espérance, 325
série espérance, 268
inverse ensembliste, 206

jacobien, 146, 305

lemme
Borel Cantelli, 336, 337
de Slutsky, 369
loi
à densité, 215
binomiale, 192, 221
conditionnelle, 209
continue, 213
d'un vecteur aléatoire, 296
d'une variable aléatoire discrète, 209
d'une variable aléatoire réelle, 208
de Bernoulli, 221
de Cauchy, 239
de Gumbel, 366
de Poisson, 225, 322, 329
diffuse, 213
discrète sur $\mathbb{R}$, 209
exponentielle, 234
formes linéaires, 311
géométrique, 224
gaussienne $\mathfrak{N}(m, \sigma)$, 237
hypergéométrique, 222
marginale (vecteur aléatoire), 296
multinomiale, 298, 384
normale, 237
tables, 520
somme de v.a. discrètes, 321
indépendantes, 321
uniforme sur un borélien de $\mathbb{R}^d$, 176
uniforme sur un ensemble fini, 221
uniforme sur un segment, 175
loi faible des grands nombres, 349
loi forte des grands nombres, 349

masse de Dirac, 158
matrice de covariance, 309
mesurable (application), 205
mesure, 157
aire, 160
de comptage, 159
de Dirac, 158
de Lebesgue sur $\mathbb{R}^d$, 160
longueur, 160
ponctuelle, 159
propriétés, 171
série de mesures, 158
volume, 160
moment, 276
h-moment d'une v.a., 276
absolu, 276
d'une v.a. à densité, 278
d'une v.a. discrète, 277
fonctionnel, 276
Monte-Carlo, 405

n-uplet, 15
non dénombrabilité
de $\mathcal{P}(\mathbb{N})$, $[0, 1]$, $\mathbb{R}$, $\mathbb{C}$, 28
de $\{0, 1\}^{\mathbb{N}}$, 28

opérateur de covariance, 311

partie négative (d'un réel), 44
partie positive (d'un réel), 44
partition, 185
pas d'une subdivision, 78, 131
pavé, 128
permutation, 20
permutation aléatoire, 196
    inversions, 408
primitive, 79
probabilité
    conditionnelle, 182
    d'un intervalle, 178
    produit, 314
probabilités
    des causes, 186
    totales (formule des), 185
produit $\otimes$ de fonctions, 317
produit $\otimes$ de probabilités, 314
produit cartésien, 14

quantificateurs, 14, 334

record, 365, 403
réunion, 14
Riemann intégrabilité, 72, 131, 133
    d'un produit, 87
    de $|f|$, 85, 134
    de f^+ et f^-, 86
    fonction admissible, 98
    locale, 104
    par convergence uniforme, 81
Riemann intégrabilité de f
    admissible, 136
    bornée continue par morceaux, 79
    continue, 78
    continue sur un pavé, 131
    en escalier, 76
    monotone, 77
    réglée, 82

semi norme, 87
série, 37
    absolument convergente, 39
    alternée, 41
    commutativement convergente, 43
    convergence normale, 39
    de Bertrand, 40
    de Riemann, 40
    double, 60
    géométrique, 38
    harmonique, 38
    produit, 63
    reste (série convergente), 38
    somme partielle, 37
    vectorielle, 38
série double, 60
    interversion des sommations, 61
σ-additivité, 156
σ-algèbre, 156
sommation par paquets, 58
    dans $\overline{\mathbb{R}}_+$, 59
sommes de Darboux, 69, 129
sondage, 374
sous-additivité, 168
sous-σ-additivité, 168
stochastiquement bornée, 356
subdivision
    cellule, 129
    d'un pavé de $\mathbb{R}^d$, 129
    pas, 131
subdivision de $[a, b]$, 69
surjection, 15
    et dénombrabilité, 32

tableau triangulaire, 377
temps d'attente, 224
théorème
    absence de mémoire, 235
    caractérisation probabilités sur $\mathbb{R}$, 179
    comparaison série intégrale, 40
    comparaison séries à termes positifs, 39
    de B. Levi, 265
    de Berry-Esséen, 372

de convergence dominée, 345, 347
de de Moivre-Laplace, 370
de Fubini, 141
de Katz Petrov, 373
de Liapounov, 376, 396
de Lindeberg, 377
des séries alternées, 41
interversion des séries doubles, 61
limite central, 370
règle des équivalents (séries), 40
série produit, 63
sommation par paquets, 58
sommation par paquets dans $\overline{\mathbb{R}}_+$, 59
théorème limite central
avec autonormalisation, 378
cas i.i.d., 370
dans $\mathbb{R}^d$, 383
de de Moivre-Laplace, 370
de Liapounov, 376, 396
de Lindeberg, 377
et lois multinomiales, 384
vectoriel, 383
vitesse de convergence, 372, 373
tribu, 156
borélienne, 157
engendrée, 157
engendrée par une application, 315
produit, 314
sous-tribu, 315

variable aléatoire
étagée, 258
de Rademacher, 271
discrète, 207
positive, 254
positive intégrable, 256
réelle, 207
réelle intégrable, 270
simple, 258
variance, 282
changement d'échelle, 283
d'une somme, 307
empirique, 379
formule de Koenig, 282
translation, 283
vecteur aléatoire, 296
à densité, 299
discret, 298
espérance, 308
gaussien, 379
paramètres, 380
intégrable, 308
lois marginales, 296
vitesse de convergence
loi forte des grands nombres, 371
théorème limite central, 372, 373, 396
volume
pavé, 129