Université des Sciences et Technologies de Lille
U.F.R. de Mathématiques Pures et Appliquées
Bât. M2, F-59655 Villeneuve d'Ascq Cedex

Intégration
Analyse de Fourier
Probabilités

Réédition 2 septembre 2010

Charles SUQUET

Licence de Mathématiques **2003–2004**

Table des matières

Index

Chapitre 1

Mesures

Ce chapitre introduit les mesures sur un ensemble abstrait et leurs propriétés générales. On discute aussi la construction de ces mesures sur $\mathbb{R}$ et $\mathbb{R}^d$. L'ensemble de définition d'une mesure est une *tribu* et ceci nous amène à étudier ce type de structure.

1.1 Prologue

Définition 1.1. *Soit $\mathcal{F}$ une* tribu *de parties de Ω. On appelle* mesure positive *sur $(\Omega, \mathcal{F})$ une application*

$$\mu : \mathcal{F} \longrightarrow [0, +\infty],$$

vérifiant

a) $\mu(\emptyset) = 0$ *;*

b) *μ est σ-additive : pour toute suite $(A_i)_{i\in\mathbb{N}^*}$ d'éléments de $\mathcal{F}$ deux à deux disjoints,*

$$\mu\Big(\bigcup_{i\in\mathbb{N}^*} A_i\Big) = \sum_{i=1}^{+\infty} \mu(A_i). \tag{1.1}$$

La définition de la structure de tribu est précisée ci-dessous. D'ores et déjà, en raison de (1.1), une tribu doit posséder l'ensemble vide et être stable par union dénombrable. Remarquons que l'union des A_i est invariante par permutation sur les indices i et aussi sous l'effet de regroupements en paquets des A_i. La cohérence de la définition de la σ-additivité utilise donc implicitement, mais fortement, les propriétés des séries à termes dans $\overline{\mathbb{R}}_+$. Si l'on voulait définir des mesures à signe ou à valeurs complexes, il faudrait ajouter dans la définition la condition de sommabilité de la famille des nombres $\mu(A_i)$. Dans ce cours, nous nous limiterons aux mesures à valeurs dans $\overline{\mathbb{R}}_+$ que nous désignerons désormais simplement par le mot mesure.

Dans la définition 1.1, la propriété importante est clairement b). On pourrait remplacer a) par $\mu(\emptyset) < +\infty$, car en prenant chaque A_i égal à l'ensemble vide dans (1.1), on en déduirait $\mu(\emptyset) = 0$.

Comme « fonction d'ensembles », une mesure n'est pas nécessairement définie sur la famille $\mathcal{P}(\Omega)$ de toutes les parties de Ω. Elle a un ensemble de définition qui est généralement plus petit, c'est la tribu $\mathcal{F}$. Pour $A \in \mathcal{F}$, $\mu(A)$ s'appelle mesure de l'ensemble A.

Si $\mu(\Omega) = 1$, la mesure μ est appelée *probabilité* sur $(\Omega, \mathcal{F})$ et la condition a) devient superflue (prendre $A_1 = \Omega$ et $A_i = \emptyset$ pour $i \geq 2$ dans (1.1)).

Définition 1.2. *Une famille $\mathcal{F}$ de parties de Ω est appelée* tribu *(ou σ-algèbre) sur Ω si elle*

a) possède l'ensemble vide : $\emptyset \in \mathcal{F}$;

b) est stable par complémentaire : $\forall A \in \mathcal{F}$, $A^c \in \mathcal{F}$;

c) est stable par union dénombrable : $(\forall i \in \mathbb{N}^*,\ A_i \in \mathcal{F}) \Rightarrow \bigcup_{i \in \mathbb{N}^*} A_i \in \mathcal{F}$.

La définition d'une mesure présuppose la donnée d'une tribu $\mathcal{F}$. Par contre, on peut très bien définir une tribu sans parler de mesure, ou définir plusieurs mesures sur la même tribu.

1.2 Tribus

1.2.1 Exemples immédiats

Les deux exemples les plus simples de tribu sur un ensemble Ω sont $\mathcal{F}_0 := \{\emptyset, \Omega\}$ et $\mathcal{F}_1 = \mathcal{P}(\Omega)$, famille de tous les sous-ensembles de Ω. Ce sont des cas extrémaux puisqu'il est clair que toute tribu $\mathcal{F}$ sur Ω vérifie $\{\emptyset, \Omega\} \subset \mathcal{F} \subset \mathcal{P}(\Omega)$.

Si A est une partie de Ω, la famille

$$\mathcal{F} := \{\emptyset, A, A^c, \Omega\}$$

est une tribu sur Ω. En effet, elle vérifie par construction les propriétés a) et b) de la définition 1.2. Regardons la stabilité par union dénombrable en prenant une suite quelconque $(A_i)_{i \geq 1}$ d'éléments de $\mathcal{F}$ et en notant E leur réunion. Notons $\mathcal{H}$ l'ensemble de toutes les valeurs prises par la suite $(A_i)_{i \geq 1}$. Il y a 16 possibilités pour $\mathcal{H}$ (autant que de parties de $\mathcal{F}$, soit 2^4). On a seulement 4 cas pour la valeur de E :

- $\mathcal{H} = \{\emptyset\}$, $E = \emptyset$;
- $\mathcal{H} = \{\emptyset, A\}$ ou $\{A\}$, alors $E = A$;
- $\mathcal{H} = \{\emptyset, A^c\}$ ou $\{A^c\}$, alors $E = A^c$;
- dans tous les autres cas, $\mathcal{H}$ a pour éléments A et A^c simultanément ou Ω, alors $E = \Omega$.

Ainsi on trouve toujours pour E un élément de $\mathcal{F}$, qui est donc bien stable par union dénombrable.

En raison des propriétés a) et b), $\mathcal{F}$ est incluse dans toute tribu $\mathcal{G}$ ayant A comme élément. On l'appelle tribu engendrée par A. Rien que sur cet exemple qui est le plus simple possible en dehors des tribus $\mathcal{F}_0$ et $\mathcal{F}_1$ ci-dessus, on perçoit qu'il peut être très laborieux de vérifier qu'un ensemble même fini est une tribu. Les masochistes pourront s'exercer en déterminant manuellement la plus petite tribu $\mathcal{F}'$ contenant deux parties A et B de Ω telles que $A \neq B$ et $A \neq B^c$ ($\mathcal{F}'$ a 16 éléments et une fois qu'on les a trouvés, il reste à vérifier la stabilité par union dénombrable). Pour les autres, il est conseillé de résoudre cet exercice à l'aide de la proposition 1.7 ci-dessous.

1.2.2 Opérations ensemblistes dans une tribu

Une conséquence de la définition d'une tribu $\mathcal{F}$ est *grosso modo* qu'en effectuant des opérations ensemblistes finies ou dénombrables sur les éléments de $\mathcal{F}$, on ne sort jamais de $\mathcal{F}$. Cet énoncé est bien sûr trop flou pour pouvoir être démontré et doit seulement être considéré comme un guide pratique. En pratique, pour démontrer l'appartenance à $\mathcal{F}$, on dispose outre la définition, des propriétés suivantes.

Proposition 1.3. *Soit $\mathcal{F}$ une tribu sur l'ensemble Ω.*

a) $\mathcal{F}$ est stable par union finie ;

b) $\mathcal{F}$ est stable par intersections dénombrables ou finies ;

c) si $A \in \mathcal{F}$ et $B \in \mathcal{F}$, $A \setminus B$ et $A\Delta B$ sont dans $\mathcal{F}$.

Preuve. Soit $A_1, \dots, A_n$ une suite finie d'éléments de $\mathcal{F}$. Pour vérifier l'appartenance à $\mathcal{F}$ de $A_1 \cup \dots \cup A_n$, il suffit de compléter cette suite en posant pour $i > n$, $A_i := \emptyset$ et d'appliquer la stabilité de $\mathcal{F}$ par union dénombrable.

Pour la stabilité par intersection dénombrable, il suffit de remarquer que pour toute suite $(A_i)_{i\in\mathbb{N}^*}$ d'éléments de $\mathcal{F}$,

$$\bigcap_{i\in\mathbb{N}^*} A_i = \left(\left(\bigcap_{i\in\mathbb{N}^*} A_i\right)^c\right)^c = \left(\bigcup_{i\in\mathbb{N}^*} A_i^c\right)^c,$$

et d'utiliser la stabilité de $\mathcal{F}$ par complémentaire et par union dénombrable. La stabilité par intersection finie s'en déduit en prenant tous les A_i égaux à Ω à partir d'un certain rang.

Pour le c), il suffit de rappeler que $A \setminus B = A \cap B^c$ et $A\Delta B = (A\cap B^c) \cup (A^c \cap B)$. □

Le résultat suivant est parfois utile pour vérifier pratiquement qu'une famille donnée est une tribu.

Proposition 1.4. *La famille non vide $\mathcal{F}$ de parties de Ω est une tribu sur Ω si et seulement si elle satisfait aux trois conditions :*

i) stabilité par complémentaire : $\forall A \in \mathcal{F}$, $A^c \in \mathcal{F}$;

ii) stabilité par intersections finies : $\forall A, B \in \mathcal{F}$, $A \cap B \in \mathcal{F}$;

iii) stabilité par réunion dénombrable disjointe *: pour toute suite $(E_i)_{i\in\mathbb{N}}$ d'éléments deux à deux disjoints de $\mathcal{F}$, $\bigcup_{i\in\mathbb{N}} E_i \in \mathcal{F}$.*

En raison de l'associativité de l'intersection : $(A\cap B)\cap C = A\cap B\cap C$, la stabilité pour l'intersection de *deux* éléments A et B quelconques de $\mathcal{F}$ est équivalente à la stabilité pour l'intersection d'un nombre fini $n \geq 2$ quelconque d'éléments de $\mathcal{F}$. En pratique pour vérifier la stabilité par intersection finie, il suffit donc de traiter le cas $n = 2$.

Preuve. La condition i) est déjà dans la définition d'une tribu ; ii) est nécessaire par la prop. 1.3 b) ; iii) est un cas particulier de la stabilité par union dénombrable. Ainsi la nécessité des conditions i), ii) et iii) est claire et il nous reste à montrer leur suffisance.

Puisque $\mathcal{F}$ est non vide, il y a au moins un élément A dans $\mathcal{F}$, alors par i), $A^c \in \mathcal{F}$ et par ii), $A \cap A^c \in \mathcal{F}$, d'où $\emptyset \in \mathcal{F}$. Il reste donc à montrer la stabilité de $\mathcal{F}$ par union dénombrable quelconque. Auparavant deux remarques nous seront utiles. Si A et B sont dans $\mathcal{F}$, $B \setminus A = B \cap A^c$ aussi. En combinant i) et ii), on obtient la stabilité de $\mathcal{F}$ par réunion finie.

Soit $(A_i)_{i\in\mathbb{N}}$ une suite quelconque dans $\mathcal{F}$. Posons pour tout $n \in \mathbb{N}$,

$$B_n := \bigcup_{i=0}^{n} A_i, \quad B := \bigcup_{i\in\mathbb{N}} A_i.$$

Il est clair que les suites $(A_i)_{i\in\mathbb{N}}$ et $(B_n)_{n\in\mathbb{N}}$ ont même réunion, le problème est donc de vérifier l'appartenance de B à $\mathcal{F}$.

Admettons, pour l'instant, que pour tout $n \geq 1$, B_n vérifie la décomposition en union disjointe :

$$B_n = B_0 \cup \left(\bigcup_{i=1}^{n} (B_i \setminus B_{i-1}) \right).$$

En écrivant la réunion infinie des B_n à l'aide de cette décomposition et en « effaçant » toutes les répétitions des $B_i \setminus B_{i-1}$, on en déduit immédiatement que B vérifie la décomposition en union dénombrable disjointe :

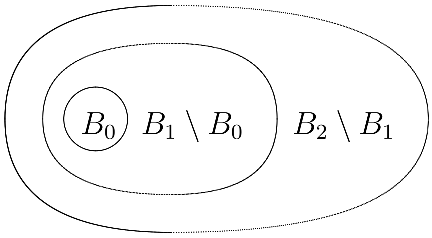
$$B = B_0 \cup \left(\bigcup_{i\in\mathbb{N}^*} (B_i \setminus B_{i-1}) \right).$$

Par stabilité de $\mathcal{F}$ par union finie, chaque B_n est dans $\mathcal{F}$. Alors chaque $B_i \setminus B_{i-1}$ est dans $\mathcal{F}$ (stabilité par différence). Enfin la réunion dénombrable disjointe des $B_i \setminus B_{i-1}$ est dans $\mathcal{F}$.

Ainsi pour compléter la preuve, il ne reste plus qu'à justifier la décomposition de B_n. Posons :

$$D_n = B_0 \cup \left(\bigcup_{i=1}^{n} (B_i \setminus B_{i-1}) \right).$$

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 puisque pour tout $i \leq n$, $B_i \setminus B_{i-1} \subset B_i \subset B_n$. 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 $0 \leq i_0 \leq n$. Si $i_0 = 0$, $\omega \in B_0$ et comme $B_0 \subset D_n$, $\omega \in D_n$. Si $i_0 \geq 1$, par la définition même de i_0, on a $\omega \in B_{i_0}$ et $\omega \notin B_{i_0-1}$, donc $\omega \in B_{i_0} \setminus B_{i_0-1}$ et comme $i_0 \leq n$, $B_{i_0} \setminus B_{i_0-1} \subset D_n$ donc $\omega \in D_n$. Le raisonnement précédent étant valable pour tout ω de B_n, on en déduit $B_n \subset D_n$.



□

1.2.3 Génération de tribus

Nous examinons maintenant plusieurs méthodes de construction de tribus qui auront toutes leur utilité dans la suite de ce cours.

Commençons par un résultat négatif qu'il convient de connaître : si $\mathcal{F}_1$ et $\mathcal{F}_2$ sont des tribus sur Ω, $\mathcal{F}_1 \cup \mathcal{F}_2$ n'est en général pas une tribu. Voici un contre-exemple simple. Prenons $\Omega := \{1, 2, 3\}$ et pour $i = 1$, 2, $A_i := \{i\}$, $\mathcal{F}_i := \{\emptyset, A_i, A_i^c, \Omega\}$. La réunion de ces deux tribus est

$$\mathcal{F}_1 \cup \mathcal{F}_2 = \Big\{\emptyset, \{1\}, \{2\}, \{2, 3\}, \{1, 3\}, \Omega\Big\}.$$

On constate que A_1 et A_2 sont éléments de $\mathcal{F}_1 \cup \mathcal{F}_2$, alors que $A_1 \cup A_2 = \{1, 2\}$ n'est pas élément de $\mathcal{F}_1 \cup \mathcal{F}_2$.

La situation est bien plus agréable pour les intersections de tribus.

Proposition 1.5. *Si I est une famille* quelconque *d'indices et $(\mathcal{F}_i)_{i \in I}$ une famille de tribus sur Ω, alors $\bigcap\limits_{i \in I} \mathcal{F}_i$ est une tribu sur Ω.*

Vouloir démontrer une proposition aussi évidente reviendrait simplement à expliquer sur cet exemple ce qu'est une intersection. Noter que l'on ne suppose pas I dénombrable. Cette proposition permet notamment de justifier la définition suivante (noter l'analogie formelle avec l'intersection d'espaces vectoriels et le sous-espace vectoriel engendré par une famille de vecteurs).

Définition 1.6. *Soit $\mathcal{S}$ une famille quelconque de parties de Ω ($\mathcal{S} \subset \mathcal{P}(\Omega)$). On appelle tribu engendrée par $\mathcal{S}$ et on note $\sigma(\mathcal{S})$, l'intersection de toutes les tribus sur Ω contenant $\mathcal{S}$. C'est la plus petite (au sens de l'inclusion) des tribus $\mathcal{F}$ telles que $\mathcal{S} \subset \mathcal{F}$.*

Au risque d'insister lourdement, notons comme conséquence pratique de cette définition *l'argument de minimalité* que nous utiliserons intensivement :

$$(\mathcal{S} \subset \mathcal{F} \text{ et } \mathcal{F} \text{ tribu}) \Rightarrow (\sigma(\mathcal{S}) \subset \mathcal{F}). \tag{1.2}$$

En général, on ne sait pas donner une description constructive d'une tribu infinie différente de $\mathcal{P}(\Omega)$. Le résultat suivant est une exception notable.

Proposition 1.7. *Si I est fini ou dénombrable et si $\Pi = \{A_i; i \in I\}$ est une* partition *de Ω, alors*

$$\sigma(\Pi) = \Big\{\bigcup_{i\in J} A_i;\ J \subset I\Big\}.$$

Rappelons que par convention, une union indexée par l'ensemble vide est elle même vide. Une partition de Ω est une famille de parties dont aucune n'est vide, deux à deux disjointes et dont la réunion est Ω. La proposition 1.7 pourra être démontrée en T.D.

Nous examinons maintenant le « transport » de tribus par une application. Commençons par quelques précisions sur les images ensemblistes. Dans ce qui suit, on fixe deux ensembles E et F et une application *quelconque* $f : E \to F$. À partir de f, on va construire des applications opérant à 3 niveaux de structures. Nous noterons provisoirement ces niveaux par un indice avec la signification suivante :

0 : on opère sur des éléments de E ou F (niveau E, F),

1 : on opère sur des parties de E ou F, (niveau $\mathcal{P}(E)$, $\mathcal{P}(F)$),

2 : on opère sur des *familles* de parties de E ou F (niveau $\mathcal{P}\big(\mathcal{P}(E)\big)$, $\mathcal{P}\big(\mathcal{P}(F)\big)$).

Au niveau 0, on pose simplement $f_0 := f$. Ensuite au niveau 1, on définit les images ensemblistes directe et réciproque par

$$\begin{array}{rcl}\forall A \in \mathcal{P}(E),\quad f_1(A) & := & \{f_0(x) \in F;\ x \in A\};\\ \forall B \in \mathcal{P}(F),\quad f_1^{-1}(B) & := & \{x \in E;\ f_0(x) \in B\}.\end{array}$$

En particulier $f_1(\emptyset) = \emptyset$ et $f_1^{-1}(\emptyset) = \emptyset$. Il importe de noter que f_1^{-1} n'est pas une application réciproque au sens classique : $f_1^{-1}(B)$ est défini via f_0 pour tout $B \in \mathcal{P}(F)$ et f_1 peut très bien ne pas être bijective. Remarquons ici qu'on a toujours $f_1\big(f_1^{-1}(B)\big) \subset B$ et que l'inclusion peut être stricte (prendre f_0 non surjective et B non inclus dans $f_1(E)$).

Passons au niveau 2, où nous définirons seulement f_2^{-1} en posant pour toute famille $\mathcal{B}$ de parties de F (donc $\mathcal{B} \subset \mathcal{P}(F)$ ou encore $\mathcal{B} \in \mathcal{P}\big(\mathcal{P}(F)\big)$) :

$$f_2^{-1}(\mathcal{B}) := \{A \in \mathcal{P}(E);\ \exists B \in \mathcal{B}, A = f_1^{-1}(B)\} = \{f_1^{-1}(B); B \in \mathcal{B}\}.$$

Signalons ici le piège : $f_2^{-1}(\mathcal{B})$ n'est pas forcément égal à $\{A \in \mathcal{P}(E);\ f_1(A) \in \mathcal{B}\}$. Voici un contre-exemple. On prend $f = f_0 : [-\pi, \pi] \to \mathbb{R}$, $x \mapsto \sin x$ et $\mathcal{B} = \{[0, b],\ b \in]1, +\infty[\}$. Alors $f_2^{-1}(\mathcal{B}) = \{[0, \pi]\}$, mais $\{A \in \mathcal{P}([-\pi, \pi]); f_1(A) \in \mathcal{B}\} = \emptyset$.

Dans l'usage courant, que nous adoptons désormais, on utilise la même lettre f pour désigner f_0, f_1 ou f_2, le contexte permettant de comprendre à quel niveau on se situe. C'est assez facile pour peu que l'on respecte les conventions typographiques désignant par des minuscules italiques comme x, y, ω, les éléments de E ou F, par des majuscules italiques A, B, E, les sous-ensembles de E ou F et par des majuscules cursives $\mathcal{A}$, $\mathcal{B}$, $\mathcal{F}$, $\mathcal{P}$, les familles d'ensembles (familles de parties de E ou F).

Proposition 1.8. *Soient E et F des ensembles et $f : E \to F$ une application. L'inverse ensembliste f^{-1} vérifie :*

a) pour tout $B \subset F$, $f^{-1}(B^c) = \big(f^{-1}(B)\big)^c$;

b) pour toute famille $(B_i)_{i\in I}$ de parties de F, $f^{-1}\Bigl(\bigcup\limits_{i\in I} B_i\Bigr) = \bigcup\limits_{i\in I} f^{-1}(B_i)$.

c) pour toute famille $(B_i)_{i\in I}$ de parties de F, $f^{-1}\Bigl(\bigcap\limits_{i\in I} B_i\Bigr) = \bigcap\limits_{i\in I} f^{-1}(B_i)$.

Preuve. Dans le a), les deux complémentaires utilisés n'ont pas la même signification : $B^c = F \setminus B$ tandis que $\bigl(f^{-1}(B)\bigr)^c = E \setminus f^{-1}(B)$. On vérifie a) ainsi :

$$\begin{aligned} f^{-1}(B^c) &= \{x \in E; f(x) \in B^c\} \\ &= \{x \in E; f(x) \notin B\} \\ &= \{x \in E; x \notin f^{-1}(B)\} \\ &= E \setminus f^{-1}(B) = \bigl(f^{-1}(B)\bigr)^c. \end{aligned}$$

Pour montrer b), on écrit

$$\begin{aligned} f^{-1}\Bigl(\bigcup_{i\in I} B_i\Bigr) &= \Bigl\{x \in E; f(x) \in \bigcup_{i\in I} B_i\Bigr\} \\ &= \{x \in E;\ \exists i \in I, f(x) \in B_i\} \\ &= \{x \in E;\ \exists i \in I, x \in f^{-1}(B_i)\} \\ &= \bigcup_{i\in I} f^{-1}(B_i). \end{aligned}$$

La vérification de c) est analogue en remplaçant « $\exists i$ » par « $\forall i$ ». □

Remarque 1.9. La situation est moins confortable pour l'image ensembliste directe. On pourra vérifier en exercice que $f\Bigl(\bigcup\limits_{i\in I} A_i\Bigr) = \bigcup\limits_{i\in I} f(A_i)$. Par contre l'image directe ne commute pas en général avec l'intersection. Comme contre-exemple, on peut prendre $E = \mathbb{R}$, $F = \mathbb{R}^+$, $f : x \mapsto x^2$, $A_1 = [-1, 0[$ et $A_2 =]0, 1]$. Ainsi $f(A_1 \cap A_2) = f(\emptyset) = \emptyset$, alors que $f(A_1) \cap f(A_2) =]0, 1]$. Elle ne commute pas davantage avec le passage au complémentaire, contre-exemple : $E = F = \mathbb{R}$, $f(x) = x^2$, $A =]0, +\infty[$, on trouve $f(A^c) = [0, +\infty[$ et $f(A)^c =]-\infty, 0]$.

Proposition 1.10. *Soient E et F des ensembles et $f : E \to F$ une application.*

i) Si $\mathcal{B}$ est une tribu sur F, $f^{-1}(\mathcal{B})$ est une tribu sur E.

ii) Si $\mathcal{A}$ est une tribu sur E, $\mathcal{T} := \{B \in \mathcal{P}(F);\ f^{-1}(B) \in \mathcal{A}\}$ est une tribu sur F. On l'appelle tribu image de $\mathcal{A}$ par f et on la note $f(\mathcal{A})$.

Preuve de i). L'ensemble vide est bien un élément de $f^{-1}(\mathcal{B})$, car $\mathcal{B}$ étant une tribu, $\emptyset \in \mathcal{B}$ et $f^{-1}(\emptyset) = \emptyset$. Pour vérifier la stabilité par complémentaire, soit A un élément quelconque de $f^{-1}(\mathcal{B})$. Cela signifie qu'il existe $B \in \mathcal{B}$ tel que $A = f^{-1}(B)$. Or $A^c = f^{-1}(B^c)$ et la tribu $\mathcal{B}$ est stable par complémentaire donc $B^c \in \mathcal{B}$ et ainsi $A^c \in f^{-1}(\mathcal{B})$. Enfin, pour vérifier la stabilité par union dénombrable, soit $(A_i)_{i\geq 1}$ une suite d'éléments de $f^{-1}(B)$. Il existe alors une suite $(B_i)_{i\geq 1}$ dans $\mathcal{B}$ telle que pour tout $i \geq 1$, $A_i = f^{-1}(B_i)$. On peut alors écrire

$$\bigcup_{i\geq 1} A_i = \bigcup_{i\geq 1} f^{-1}(B_i) = f^{-1}\Bigl(\bigcup_{i\geq 1} B_i\Bigr) \in f^{-1}(\mathcal{B}),$$

car la tribu $\mathcal{B}$ est stable par union dénombrable. □

Preuve de ii). Laissée en exercice. □

Proposition 1.11. *Soient E et F des ensembles, $f : E \to F$ une application et $\mathcal{S} \subset \mathcal{P}(F)$ une famille quelconque de parties de F. Alors*

$$f^{-1}\big(\sigma(\mathcal{S})\big) = \sigma\big(f^{-1}(\mathcal{S})\big). \tag{1.3}$$

Preuve. On prouve l'égalité (1.3) en montrant l'inclusion dans les deux sens.

a) Vérification de $\sigma\big(f^{-1}(\mathcal{S})\big) \subset f^{-1}\big(\sigma(\mathcal{S})\big)$. De l'inclusion $\mathcal{S} \subset \sigma(\mathcal{S})$, on tire $f^{-1}(\mathcal{S}) \subset f^{-1}\big(\sigma(\mathcal{S})\big)$. Par la proposition 1.10 i), $f^{-1}\big(\sigma(\mathcal{S})\big)$ est une tribu. Par minimalité (voir (1.2)) on en déduit

$$\sigma\big(f^{-1}(\mathcal{S})\big) \subset f^{-1}\big(\sigma(\mathcal{S})\big). \tag{1.4}$$

b) Vérification de $f^{-1}\big(\sigma(\mathcal{S})\big) \subset \sigma\big(f^{-1}(\mathcal{S})\big)$. Posons

$$\mathcal{A} := \sigma\big(f^{-1}(\mathcal{S})\big), \qquad \mathcal{F} := \{B \in \mathcal{P}(F);\ f^{-1}(B) \in \mathcal{A}\}.$$

Par la proposition 1.10 ii), $\mathcal{F}$ est une tribu sur F (tribu image de $\mathcal{A}$ par f). On voit que

$$\mathcal{S} \subset \mathcal{F},$$

car pour tout $B \in \mathcal{S}$, $f^{-1}(B) \in f^{-1}(\mathcal{S}) \subset \sigma\big(f^{-1}(\mathcal{S})\big) = \mathcal{A}$. En invoquant à nouveau l'argument de minimalité (1.2), on en déduit que $\sigma(\mathcal{S}) \subset \mathcal{F}$, d'où

$$f^{-1}\big(\sigma(\mathcal{S})\big) \subset f^{-1}(\mathcal{F}).$$

Remarquons maintenant que $f^{-1}(\mathcal{F}) \subset \mathcal{A}$ en raison de la définition de $\mathcal{F}$ puisque tout $B \in \mathcal{F}$ est tel que $f^{-1}(B) \in \mathcal{A}$. Finalement on a

$$f^{-1}\big(\sigma(\mathcal{S})\big) \subset f^{-1}(\mathcal{F}) \subset \mathcal{A} = \sigma\big(f^{-1}(\mathcal{S})\big). \tag{1.5}$$

L'égalité (1.3) découle de (1.4) et (1.5). □

1.2.4 Le théorème π-λ

Pour montrer une propriété relative à une tribu $\mathcal{F}$, il est souvent plus commode de montrer cette propriété pour une famille $\mathcal{L}$ de parties de Ω plus grande que $\mathcal{F}$ et d'obtenir le résultat souhaité par restriction à $\mathcal{F}$. Un outil important dans cette perspective est le théorème de Dynkin, connu aussi sous le nom de théorème π-λ.

Définition 1.12. *On dit que $\mathcal{P}$ est une π-classe (ou π-système) de parties de Ω si elle est stable pour l'intersection finie : $\forall A, B \in \mathcal{P}$, $A \cap B \in \mathcal{P}$.*

Définition 1.13. *Une famille $\mathcal{L}$ de parties de Ω est appelée λ-classe (ou λ-système) si elle vérifie :*

a) $\Omega \in \mathcal{L}$ *;*

b) stabilité par différence propre : si $A, B \in \mathcal{L}$ *et* $A \subset B$*, alors* $B \setminus A \in \mathcal{L}$ *;*

c) stabilité par union dénombrable croissante : si $(A_k)_{k\in\mathbb{N}}$ *est une suite dans* $\mathcal{L}$ *telle que* $\forall k \in \mathbb{N}$*,* $A_k \subset A_{k+1}$*, l'ensemble* $A := \bigcup_{k\in\mathbb{N}} A_k$ *est élément de* $\mathcal{L}$*.*

Toute tribu est à la fois une π-classe et une λ-classe (cf. prop. 1.3). La réciproque est vraie.

Proposition 1.14. *Si la famille* $\mathcal{F}$ *de parties de* Ω *est à la fois une* π*-classe et une* λ*-classe, c'est une tribu.*

Preuve. $\mathcal{F}$ est stable par complémentaire puisque $\Omega \in \mathcal{F}$ (déf. 1.13-a)) et $A^c = \Omega \setminus A$ (déf. 1.13-b)). En particulier, $\emptyset = \Omega^c$ est élément de $\mathcal{F}$. En combinant la stabilité de $\mathcal{F}$ par intersection finie (comme π-classe) et la stabilité par complémentaire on obtient la stabilité par union finie. Soit $(B_k)_{k\in\mathbb{N}}$ une suite *quelconque* dans $\mathcal{F}$. Pour tout $n \in \mathbb{N}$, l'union finie $A_n := \cup_{0\leq k\leq n} B_k$ est dans $\mathcal{F}$. De plus la suite (A_n) est *croissante* pour l'inclusion et a pour réunion $A = \cup_{k\in\mathbb{N}} B_k$. En raison de déf. 1.13-c), $\cup_{k\in\mathbb{N}} B_k \in \mathcal{F}$, ce qui établit la stabilité de $\mathcal{F}$ par union dénombrable. $\mathcal{F}$ est donc bien une tribu. □

Théorème 1.15 (Dynkin)**.** *Si une* λ*-classe* $\mathcal{L}$ *de parties de* Ω *contient une* π*-classe* $\mathcal{P}$*, elle contient aussi la tribu engendrée par* $\mathcal{P}$*.*

Preuve. Il est facile de voir que l'intersection d'une famillle quelconque non vide de λ-classes de parties de Ω est encore une λ-classe de parties de Ω. De plus $\mathcal{P}(\Omega)$ est une λ-classe. La famille de toutes les λ-classes contenant $\mathcal{P}$ est donc non vide et on peut définir $\Lambda(\mathcal{P})$, λ-classe engendrée par $\mathcal{P}$, comme l'intersection de toutes les λ-classes contenant $\mathcal{P}$. On va montrer que $\Lambda(\mathcal{P}) = \sigma(\mathcal{P})$ et comme $\Lambda(\mathcal{P}) \subset \mathcal{L}$, ceci prouvera le théorème.

Comme $\sigma(\mathcal{P})$ est une λ-classe et contient $\mathcal{P}$, il est clair que $\Lambda(\mathcal{P}) \subset \sigma(\mathcal{P})$. Il reste à prouver l'inclusion inverse. Pour cela, il suffit de montrer que la λ-classe $\Lambda(\mathcal{P})$ est aussi une π-classe car alors elle sera une tribu (prop. 1.14) contenant $\mathcal{P}$, donc contenant aussi la tribu $\sigma(\mathcal{P})$.

Il s'agit donc de prouver que si A et B sont deux éléments quelconques de $\Lambda(\mathcal{P})$, $A \cap B \in \Lambda(\mathcal{P})$. Définissons pour tout $C \in \mathcal{P}(\Omega)$ la famille de parties

$$\mathcal{G}_C := \{D \in \mathcal{P}(\Omega);\ C \cap D \in \Lambda(\mathcal{P})\}.$$

Vérifions d'abord que si $C \in \Lambda(\mathcal{P})$, $\mathcal{G}_C$ est une λ-classe. Ω est bien élément de $\mathcal{G}_C$ puisque $C \cap \Omega = C \in \Lambda(\mathcal{P})$. Si $D, E \in \mathcal{G}_C$ et $D \subset E$, alors $(E \setminus D) \cap C = (E \cap C) \setminus (D \cap C)$ est bien élément de $\Lambda(\mathcal{P})$ en appliquant la propriété b) de la définition 1.13 avec $D \cap C$ et $E \cap C$. Ainsi, $E \setminus D$ satisfait la condition d'appartenance à $\mathcal{G}_C$. Ceci établit la stabilité de $\mathcal{G}_C$ par différence propre. Soit (D_n) une suite dans $\mathcal{G}_C$, croissante pour l'inclusion et D sa réunion. Alors $(C \cap D_n)$ est une suite dans $\Lambda(\mathcal{P})$, croissante et de réunion $C \cap D$. Comme $\Lambda(\mathcal{P})$ est une λ-classe, $C \cap D \in \Lambda(\mathcal{P})$, donc D est élément de $\mathcal{G}_C$ et ceci établit la stabilité de $\mathcal{G}_C$ par union dénombrable croissante.

Si $C \in \mathcal{P}$, alors $\mathcal{G}_C$ contient $\mathcal{P}$ puisque pour tout $D \in \mathcal{P}$, $C \cap D \in \mathcal{P} \subset \Lambda(\mathcal{P})$. Or $\mathcal{G}_C$ est une λ-classe, donc elle contient aussi $\Lambda(\mathcal{P})$. Nous venons ainsi d'établir que

$$\forall C \in \mathcal{P},\ \forall B \in \Lambda(\mathcal{P}), \quad C \cap B \in \Lambda(\mathcal{P}).$$

Ce résultat peut se réécrire : $\forall B \in \Lambda(\mathcal{P})$, $\mathcal{G}_B \supset \mathcal{P}$. Or $\mathcal{G}_B$ est une λ-classe, donc par minimalité elle contient aussi $\Lambda(\mathcal{P})$. L'inclusion $\mathcal{G}_B \supset \Lambda(\mathcal{P})$ vraie pour tout $B \in \Lambda(\mathcal{P})$ se traduit par : $\forall A \in \Lambda(\mathcal{P})$, $\forall B \in \Lambda(\mathcal{P})$, $A \cap B \in \Lambda(\mathcal{P})$. Ceci établit la stabilité de $\Lambda(\mathcal{P})$ par intersections finies et achève la preuve du théorème de Dynkin. □

1.2.5 Tribus boréliennes

Dans un espace métrique E, on peut définir les notions d'ensemble ouvert, fermé, de voisinage, d'intérieur, de frontière, etc. Toutes ces notions qui permettent de faire de l'analyse dans ce cadre se déduisent de la notion d'ouvert. On peut aussi faire de l'analyse dans un cadre plus général, celui d'*espace topologique* $(E, \mathcal{O})$. Dans un tel espace, on se donne *a priori* la famille $\mathcal{O}$ des ouverts appelée *topologie*. Une topologie $\mathcal{O}$ sur E est une famille de parties de E vérifiant :

a) $\emptyset \in \mathcal{O}$ et $E \in \mathcal{O}$;

b) $\mathcal{O}$ est stable par intersection finie ;

c) $\mathcal{O}$ est stable par union quelconque.

Tout espace métrique est aussi un espace topologique en prenant pour $\mathcal{O}$ la famille des unions quelconques de boules ouvertes (y compris l'ensemble vide, considéré comme union indexée par $\emptyset$). Une topologie sur E n'est pas une tribu, il lui manque la stabilité par complémentaire. Notons que c) est une condition plus forte que la stabilité par union dénombrable.

Définition 1.16. *Soit* $(E, \mathcal{O})$ *un espace topologique. La tribu borélienne de* E *est la tribu engendrée par la famille* $\mathcal{O}$ *des* ouverts *de* E*. On la note* $\mathrm{Bor}(E)$*, ou* $\mathcal{B}(E)$ *ou* $\mathcal{B}_E$*.*

En raison de la stabilité d'une tribu par passage au complémentaire, il est facile de voir que $\mathrm{Bor}(E)$ est aussi la tribu engendrée par les fermés de E. Dans la suite, nous nous intéressons plus particulièrement au cas $E = \mathbb{R}^d$, $d \geq 1$. Sauf mention explicite du contraire, $\mathbb{R}^d$ sera toujours supposé muni de sa topologie d'espace vectoriel [1].

Nous aurons besoin du résultat suivant.

Lemme 1.17.

i) Tout ouvert de $\mathbb{R}$ *(resp.* $\mathbb{R}^d$*) est union au plus dénombrable d'intervalles ouverts (resp. de pavés ouverts) à extrémités rationnelles (resp. dans* $\mathbb{Q}^d$*).*

1. On démontre (cf. cours d'Analyse Réelle) que sur un espace vectoriel de *dimension finie*, il existe une seule topologie séparée rendant continues l'addition des vecteurs et la multiplication d'un vecteur par un scalaire. Toutes les normes sur cet espace sont équivalentes et engendrent cette topologie. On est donc libre de choisir sur $\mathbb{R}^d$ les boules associées à la norme de son choix, la famille des ouverts correspondante et la tribu borélienne associée seront toujours les mêmes.

ii) Tout ouvert de $\mathbb{R}$ est union au plus dénombrable d'intervalles ouverts deux à deux disjoints.

Preuve de i). Traitons d'abord le cas particulier où l'ouvert est un intervalle ouvert à extrémités réelles $a < b$. Il suffit de noter que

$$]a, b[= \bigcup_{n\in\mathbb{N}}]r_n, s_n[,$$

où $\big(r_n\big)_{n\in\mathbb{N}}$ et $\big(s_n\big)_{n\in\mathbb{N}}$ sont deux suites de rationnels telles que $r_n \downarrow a$, $s_n \uparrow b$ et pour tout n, $a < r_n < s_n < b$.

Soit V un ouvert quelconque de $\mathbb{R}$. Alors il est réunion d'intervalles ouverts et en exploitant le cas précédent, on peut l'écrire

$$V = \bigcup_{i\in I}]a_i, b_i[= \bigcup_{i\in I} \bigcup_{n\in\mathbb{N}}]r_{n,i}, s_{n,i}[= \bigcup_{(r,s)\in D}]r, s[,$$

où l'on a noté $D := \Big\{\big(r_{n,i}, s_{n,i}\big);\ i \in I, n \in \mathbb{N}\Big\}$. L'introduction de cette indexation par D revient à nettoyer l'indexation par I et $\mathbb{N}$ en effaçant les répétitions (il peut y en avoir une infinité non dénombrable !) de chaque intervalle $]r, s[$ apparaissant dans la réunion de tous les $]r_{n,i}, s_{n,i}[$. Comme D est inclus dans $\mathbb{Q}^2$, il est au plus dénombrable et on a bien réussi à écrire V comme union au plus dénombrable d'intervalles à extrémités rationnelles. Le résultat est donc établi en dimension 1. La généralisation en dimension $d > 1$ est une simple adaptation du cas unidimensionnel et est laissée en exercice. $\square$

Preuve de ii). Remarquons qu'un sous-ensemble E de $\mathbb{R}$ est un intervalle si et seulement si pour tous x, y dans E avec $x \leq y$, $[x, y] \subset E$. On part de la décomposition $V = \bigcup_{(r,s)\in D}]r, s[$, établie dans la preuve du i). On définit sur D la relation binaire :

$$(r, s) \sim (r', s') \quad \Leftrightarrow \quad \exists J \text{ intervalle inclus dans } V \text{ tel que }]r, s[\subset J \text{ et }]r's'[\subset J.$$

Il est clair qu'il s'agit d'une relation d'équivalence. Les classes $\{D_k; k \in K\}$ forment une partition de D, donc l'ensemble d'indices K est au plus dénombrable. Posant alors

$$\forall k \in K, \quad E_k := \bigcup_{(r,s)\in D_k}]r, s[,$$

on a clairement

$$V = \bigcup_{k\in K} E_k.$$

Par construction, chaque E_k est un ouvert. Vérifions que c'est un intervalle. Soient x et y quelconques dans E_k avec $x \leq y$. Alors il existe des couples de rationnels $(r, s) \in D_k$ et $(r', s') \in D_k$ tels que $x \in]r, s[$ et $y \in]r', s'[$. De plus comme ces deux couples sont dans la même classe D_k, il existe un intervalle J inclus dans V et contenant à la fois $]r, s[$ et $]r', s'[$. On en déduit que x et y sont dans cet intervalle J, donc aussi $[x, y] \subset J \subset V$. Pour conclure que E_k est un intervalle (nécessairement ouvert), il reste à vérifier que $[x, y]$ est inclus dans E_k. Si tel n'est pas le cas, il existe un $z \in]x, y[$ tel que $z \notin E_k$ et comme $z \in V$, nécessairement $z \in E_j$ pour une autre classe E_j disjointe de E_k. Dans ce cas il existe un couple de rationnels $(r'', s'') \in D_j$ tels que $z \in]r'', s''[$. Or $J_0 := J \cup]r'', s''[$ est un intervalle (comme réunion de deux intervalles ayant au moins un point commun z) et est inclus dans V. Comme J_0 contient à la fois $]r, s[$, $]r', s'[$ et $]r'', s''[$, les trois couples de rationnels correspondants sont dans la même classe, d'où la contradiction. Donc $z \in E_k$ et E_k est bien un intervalle. □

Proposition 1.18. *La tribu borélienne de $\mathbb{R}$ est engendrée par chacune des familles*
- *intervalles ouverts $]a, b[$, $a, b \in \mathbb{R}$;*
- *intervalles fermés $[a, b]$, $a, b \in \mathbb{R}$;*
- *intervalles semi-ouverts $]a, b]$, $a, b \in \mathbb{R}$;*
- *demi-droites ouvertes $]-\infty, a[$, $a \in \mathbb{R}$;*
- *demi-droites fermées $]-\infty, a]$, $a \in \mathbb{R}$.*

Preuve. On se contentera de traiter le cas de la famille

$$\mathcal{I} := \{]a, b[;\ a, b \in \mathbb{R}, a < b\},$$

les autres cas étant laissés en exercice. Comme $\mathcal{I}$ est incluse dans la famille des ouverts de $\mathbb{R}$, elle est aussi incluse dans la tribu $\mathrm{Bor}(\mathbb{R})$, donc $\sigma(\mathcal{I}) \subset \mathrm{Bor}(\mathbb{R})$. Dans l'autre sens, l'ensemble des ouverts de $\mathbb{R}$ est inclus dans la tribu $\sigma(\mathcal{I})$ grâce au lemme 1.17 i). Donc $\mathrm{Bor}(\mathbb{R}) \subset \sigma(\mathcal{I})$. □

Proposition 1.19. *La tribu borélienne de $\mathbb{R}^d$ est engendrée par chacune des familles suivantes :*
- *les pavés ouverts $\prod_{i=1}^d]a_i, b_i[$, $a_i, b_i \in \mathbb{R}$;*
- *les pavés fermés $\prod_{i=1}^d [a_i, b_i]$, $a_i, b_i \in \mathbb{R}$;*
- *les pavés de la forme $\prod_{i=1}^d]a_i, b_i]$, $a_i, b_i \in \mathbb{R}$;*
- *les hyperquadrants $\prod_{i=1}^d]-\infty, a_i]$, $a_i \in \mathbb{R}$.*

Preuve. Le cas de la famille $\mathcal{P}$ des pavés ouverts se traite exactement comme celui des intervalles ouverts à la proposition 1.18. Laissant au lecteur le cas des pavés fermés et des hyperquadrants, nous traiterons seulement le cas de la famille $\mathcal{C}$ des pavés $\prod_{i=1}^d]a_i, b_i]$, qui nous sera utile dans la suite. On vérifie d'abord que tout pavé ouvert est élément de $\sigma(\mathcal{C})$, en écrivant par exemple

$$Q := \prod_{i=1}^d]a_i, b_i[= \bigcup_{n\geq 1} \prod_{i=1}^d]a_i, b_i - 1/n]. \tag{1.6}$$

Noter que si pour un i, $b_i \leq a_i$, les deux intervalles $]a_i, b_i[$ et $]a_i, b_i - 1/n]$ se réduisent à l'ensemble vide et Q est lui même vide. On déduit de (1.6) l'inclusion

$$\mathrm{Bor}(\mathbb{R}^d) = \sigma(\mathcal{P}) \subset \sigma(\mathcal{C}). \tag{1.7}$$

Dans l'autre sens, l'écriture de Q comme intersection dénombrable de pavés ouverts

$$Q := \prod_{i=1}^d]a_i, b_i] = \bigcap_{n\geq 1} \prod_{i=1}^d]a_i, b_i + 1/n[, \tag{1.8}$$

nous donne l'inclusion $\mathcal{C} \subset \sigma(\mathcal{P})$, d'où

$$\sigma(\mathcal{C}) \subset \sigma(\mathcal{P}) = \mathrm{Bor}(\mathbb{R}^d). \tag{1.9}$$

Au vu de (1.7) et (1.9), on a bien $\sigma(\mathcal{C}) = \mathrm{Bor}(\mathbb{R}^d)$. $\square$

1.3 Mesures

1.3.1 Un peu de vocabulaire

On appelle *espace mesurable* un couple $(\Omega, \mathcal{F})$ où Ω est un ensemble et $\mathcal{F}$ une tribu sur Ω. Un tel espace mesurable peut être muni d'une (ou de plusieurs) mesure(s) (cf. définition 1.1). Si μ est une mesure sur $(\Omega, \mathcal{F})$, le triplet $(\Omega, \mathcal{F}, \mu)$ s'appelle *espace mesuré*.

On dit qu'une propriété (π) vérifiée par des éléments de Ω est $\mathcal{F}$ mesurable si $B := \{\omega \in \Omega; \omega \text{ vérifie } (\pi)\}$ est un élément de $\mathcal{F}$. Une propriété $\mathcal{F}$-mesurable (π) est dite vraie μ-presque partout (ou μ-presque sûrement si μ est une probabilité) si $\mu(B^c) = 0$. Dans le cas où μ est une probabilité, $\mu(B^c) = 0$ équivaut à $\mu(B) = 1$.

Une mesure μ est dite finie si $\mu(\Omega) < +\infty$. D'après la proposition 1.21 2) ci-dessous, pour tout $A \in \mathcal{F}$, $\mu(A) \leq \mu(\Omega)$. En conséquence, pour une mesure considérée comme fonction d'ensembles ayant pour domaine de définition $\mathcal{F}$, « finie » équivaut à « bornée ». Cette équivalence est bien sûr fausse pour une fonction quelconque. Par exemple $f :]0, 1] \to \mathbb{R}$, $x \mapsto 1/x$. On a $|f(x)| < +\infty$ pour tout $x \in]0, 1]$, donc f est bien finie en tout point de son domaine de définition. Par contre, $\sup_{x \in]0,1]} |f(x)| = +\infty$, donc f n'est pas bornée sur son domaine de définition.

Une mesure μ sur $(\Omega, \mathcal{F})$ est dite σ-finie si Ω est union dénombrable d'ensembles A_n membres de $\mathcal{F}$ et chacun de mesure finie.

1.3.2 Premiers exemples

Exemple 1.1 (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 tout espace mesurable $(\Omega, \mathcal{F})$ où $\mathcal{F}$ est une tribu sur Ω.

Vérification. Il est clair que $\delta_{x_0}(\emptyset) = 0$. Pour montrer la σ-additivité, soit $(A_n)_{n \in \mathbb{N}}$ une suite d'éléments de $\mathcal{P}(\Omega)$, deux à deux disjoints.

Cas 1 : $x_0 \notin \bigcup_{n \in \mathbb{N}} A_n$. Alors $\delta_{x_0}\Big(\bigcup_{n \in \mathbb{N}} A_n\Big) = 0$. 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 $\sum_{n \in \mathbb{N}} \delta_{x_0}(A_n) = 0$.

Cas 2 : $x_0 \in \underset{n\in\mathbb{N}}{\cup} A_n$. Alors $\delta_{x_0}\Big(\underset{n\in\mathbb{N}}{\cup} A_n\Big) = 1$. 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 $\sum_{n\in\mathbb{N}} \delta_{x_0}(A_n) = 1$.

Dans les deux cas, on a

$$\delta_{x_0}\Big(\underset{n\in\mathbb{N}}{\cup} A_n\Big) = \sum_{n\in\mathbb{N}} \delta_{x_0}(A_n),$$

δ_{x_0} est donc bien σ-additive. C'est une mesure sur $(\Omega, \mathcal{P}(\Omega))$. □

Exemple 1.2 (séries de mesures finies)**.** Si les $(\mu_k)_{k\in\mathbb{N}}$ sont des mesures finies 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 $(\Omega, \mathcal{F})$.

Vérification. Pour tout k, $0 \leq a_k < +\infty$, donc $a_k\mu_k(\emptyset) = 0$ et $\mu(\emptyset) = \sum_{k\in\mathbb{N}} a_k\mu_k(\emptyset) = 0$. 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 la propriété de sommation par paquets dans $\overline{\mathbb{R}}_+$, on peut écrire :

$$\begin{aligned}
\mu\Big(\underset{n\in\mathbb{N}}{\cup} A_n\Big) = \sum_{k\in\mathbb{N}} a_k\mu_k\Big(\underset{n\in\mathbb{N}}{\cup} A_n\Big) &= \sum_{k\in\mathbb{N}} a_k \left(\sum_{n\in\mathbb{N}} \mu_k(A_n)\right) \\
&= \sum_{k\in\mathbb{N}} \left(\sum_{n\in\mathbb{N}} a_k\mu_k(A_n)\right) \\
&= \sum_{n\in\mathbb{N}} \left(\sum_{k\in\mathbb{N}} a_k\mu_k(A_n)\right) \\
&= \sum_{n\in\mathbb{N}} \mu(A_n),
\end{aligned}$$

ce qui établit la σ-additivité de μ. □

Exemple 1.3 (mesures ponctuelles)**.** Soient $(x_k)_{k\in\mathbb{N}}$ une suite dans Ω et $(a_k)_{k\in\mathbb{N}}$ une suite de réels positifs. Alors $\mu := \sum_{k\in\mathbb{N}} a_k\delta_{x_k}$ 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. Il est clair que le résultat demeure si l'on remplace l'ensemble d'indexation $\mathbb{N}$ par n'importe quel ensemble dénombrable. Les mesures de ce type sont appelées *mesures ponctuelles*.

Si de plus $\sum_{k\in\mathbb{N}} a_k = 1$, μ est une probabilité, appelée « probabilité discrète ». C'est le cas notamment pour les *lois* des variables aléatoires discrètes, considérées comme mesures sur $(\mathbb{R}, \mathcal{P}(\mathbb{R}))$. Si $X : \Omega \to \mathbb{R}$ est une variable aléatoire discrète, l'ensemble $X(\Omega)$ de toutes les valeurs possibles est dénombrable et peut s'écrire $X(\Omega) = \{x_k,\ k \in \mathbb{N}\}$ et la loi de X est la mesure $P_X = \sum_{k\in\mathbb{N}} \mathbf{P}(X = x_k)\delta_{x_k}$ (nous reviendrons sur cette question au chapitre 2).

Remarque 1.20. Si $\Omega = \{\omega_i;\ i \in I\}$ est fini ou dénombrable, toute mesure sur $(\Omega, \mathcal{P}(\Omega))$ qui est *finie sur les singletons* est une mesure ponctuelle. En particulier une probabilité P sur $(\Omega, \mathcal{P}(\Omega))$ est caractérisée par les relations

$$\forall i \in I, \quad P(\{\omega_i\}) = p_i,$$

où les p_i sont des réels positifs tels que $\sum_{i \in I} p_i = 1$.

Vérification. La tribu $\mathcal{P}(\Omega)$ possède les singletons (quand Ω est fini ou dénombrable, c'est même la seule tribu ayant cette propriété). 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, $0 \leq a_i < +\infty$. Considérons alors la mesure ponctuelle

$$\nu := \sum_{i \in I} a_i \delta_{\omega_i}.$$

Soit $A \in \mathcal{P}(\Omega)$, alors A est fini ou dénombrable et on peut l'écrire comme union disjointe finie ou dénombrable de singletons :

$$A = \bigcup_{\omega_i \in A} \{\omega_i\}.$$

Par σ-additivité (ou par additivité quand A est fini, cf. prop. 1.21 ci-dessous), on a donc

$$\mu(A) = \sum_{\omega_i \in A} \mu(\{\omega_i\}) = \sum_{\omega_i \in A} a_i = \sum_{i \in I} a_i \delta_{\omega_i}(A) = \nu(A).$$

Ainsi pour tout $A \in \mathcal{P}(\Omega)$, $\mu(A) = \nu(A)$, donc $\mu = \nu$ et μ est une mesure ponctuelle. □

Exemple 1.4 (trace et conditionnement)**.** 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 $(\Omega, \mathcal{F})$, que l'on peut appeler mesure trace de μ sur B. Si $0 < \mu(B) < +\infty$, 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 $(\Omega, \mathcal{F})$, c'est même une probabilité. 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 $i \neq j$, $A_i \cap A_j = \emptyset$ 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. De plus on a

$$\Big(\bigcup_{i \in \mathbb{N}} A_i\Big) \cap B = \bigcup_{i \in \mathbb{N}} (A_i \cap B).$$

La σ-additivité de ν découle alors clairement de celle de μ. □

Exemple 1.5 (mesure de Lebesgue λ_d)**.** Nous admettons provisoirement 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[2],

$$\mu\left(\prod_{i=1}^d]a_i, b_i]\right) = \prod_{i=1}^d (b_i - a_i).$$

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.

1.3.3 Propriétés générales d'une mesure

Proposition 1.21. *Toute mesure μ sur $(\Omega, \mathcal{F})$ vérifie les propriétés suivantes :*

1. *Additivité finie : si les A_i $(1 \leq i \leq n)$ sont deux à deux disjoints :*

$$\mu\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n \mu(A_i).$$

2. *Croissance : $\forall A \in \mathcal{F}$, $\forall B \in \mathcal{F}$, $A \subset B \;\Rightarrow\; \mu(A) \leq \mu(B)$.*
3. *Continuité monotone séquentielle*
   (a) *Si $(B_n)_{n\geq 0}$ est une suite croissante dans $\mathcal{F}$, convergente[3] vers $B \in \mathcal{F}$, alors $\mu(B) = \lim_{n\to+\infty} \mu(B_n)$. Notation :*

$$B_n \uparrow B \Rightarrow \mu(B_n) \uparrow \mu(B) \quad (n \to +\infty).$$

   (b) *Si $(C_n)_{n\geq 0}$ est une suite décroissante dans $\mathcal{F}$, convergente[4] vers $C \in \mathcal{F}$* et si de plus *$\mu(C_0)$ est* fini, *alors $\mu(C) = \lim_{n\to+\infty} \mu(C_n)$. Notation :*

$$\mu(C_0) < +\infty \text{ et } C_n \downarrow C \Rightarrow \mu(C_n) \downarrow \mu(C) \quad (n \to +\infty).$$

4. (a) *Sous-additivité : $\forall A_1, \dots, A_n \in \mathcal{F}$, $\quad \mu\left(\bigcup_{i=1}^n A_i\right) \leq \sum_{i=1}^n \mu(A_i)$.*

   (b) *Sous σ-additivité : $\forall A_1, \dots, A_n, \dots \in \mathcal{F}$, $\quad \mu\left(\bigcup_{i\in\mathbb{N}^*} A_i\right) \leq \sum_{i=1}^{+\infty} \mu(A_i)$.*

Preuve de 1. Soient $A_1, \dots, A_n$, n éléments de $\mathcal{F}$ deux à deux disjoints. Pour $j > n$, posons $A_j = \emptyset$. On a ainsi une suite infinie $(A_j)_{j\geq 1}$ d'ensembles deux à deux disjoints, éléments de $\mathcal{F}$. En utilisant la σ-additivité, on obtient alors :

$$\mu\left(\bigcup_{j=1}^n A_j\right) = \mu\left(\bigcup_{j\in\mathbb{N}^*} A_j\right) = \sum_{j=1}^n \mu(A_j) + \sum_{j=n+1}^{+\infty} \mu(A_j).$$

Comme $\mu(\emptyset) = 0$, la somme pour $j \geq n + 1$ vaut 0. □

2. Ce qui suppose implicitement que $a_i < b_i$ pour chaque $i = 1, \dots, d$.
3. Ce qui signifie : $\forall n \geq 0$, $B_n \subset B_{n+1}$ et $B = \bigcup_{n\geq 0} B_n$.
4. Ce qui signifie : $\forall n \geq 0$, $C_{n+1} \subset C_n$ et $C = \bigcap_{n\geq 0} C_n$.

Remarque 1.22. Il résulte de l'additivité finie que si μ est une mesure *finie*,

$$\forall A, B \in \mathcal{F}, \text{ tels que } A \subset B, \quad \mu(B \setminus A) = \mu(B) - \mu(A). \tag{1.10}$$

Plus généralement (1.10) est vraie avec une mesure non finie à condition que $\mu(A)$ soit fini. Pour vérifier (1.10), on remarque simplement que B est la réunion disjointe de A et $B \setminus A$, d'où $\mu(B) = \mu(A) + \mu(B \setminus A)$. Quand $\mu(A) < +\infty$, on obtient une égalité équivalente en le retranchant des deux membres, d'où le résultat.

Preuve de 2. Si $A \subset B$, alors $B = A \cup (B \cap A^c)$ et cette réunion est disjointe. D'après *1* on a $\mu(B) = \mu(A) + \mu(B \cap A^c)$ et comme $\mu(B \cap A^c) \geq 0$, on en déduit $\mu(B) \geq \mu(A)$. □

Preuve de 3(a). On utilise les décompositions suivantes de B_n et de B en réunion au plus dénombrable disjointe (cf. la preuve de la proposition 1.4 pour la justification) :

$$B_n = B_0 \cup \left(\bigcup_{i=1}^{n} (B_i \setminus B_{i-1}) \right), \qquad B = B_0 \cup \left(\bigcup_{i \in \mathbb{N}^*} (B_i \setminus B_{i-1}) \right).$$

Par σ-additivité de μ, ces deux décompositions nous donnent :

$$\mu(B_n) = \mu(B_0) + \sum_{i=1}^{n} \mu(B_i \setminus B_{i-1}), \quad \mu(B) = \mu(B_0) + \sum_{i=1}^{+\infty} \mu(B_i \setminus B_{i-1}).$$

Comme cette série converge, sa somme est la limite de la suite de ses sommes partielles de rang n, ce qui s'écrit :

$$\mu(B) = \lim_{n \to +\infty} \left\{ \mu(B_0) + \sum_{i=1}^{n} \mu(B_i \setminus B_{i-1}) \right\} = \lim_{n \to +\infty} \mu(B_n).$$

□

Preuve de 3(b). Posons $B_n = C_0 \setminus C_n$. Comme $C_n \subset C_0$, on a la réunion disjointe $C_0 = C_n \cup B_n$. Par additivité finie, $\mu(C_0) = \mu(C_n) + \mu(B_n)$. De plus comme $B_n \subset C_0$, $\mu(B_n) < +\infty$. On peut donc écrire

$$\mu(C_n) = \mu(C_0) - \mu(B_n).$$

La suite (B_n) est croissante de réunion $C_0 \setminus C$, car $B_n = C_0 \cap C_n^c$ d'où

$$\bigcup_{n \geq 0} B_n = C_0 \cap \left(\bigcup_{n \geq 0} C_n^c \right) = C_0 \cap \left(\bigcap_{n \geq 0} C_n \right)^c = C_0 \cap C^c = C_0 \setminus C.$$

Par 3(a) $\mu(B_n) \uparrow \mu(C_0 \setminus C)$, donc $\mu(C_n)$ converge vers $\mu(C_0) - \mu(C_0 \setminus C) = \mu(C)$. Noter que $\mu(C_0)$ étant fini et C étant inclus dans C_0, $\mu(C)$ et $\mu(C_0 \setminus C)$ sont tous deux finis. □

Il importe de noter l'utilisation de l'hypothèse $\mu(C_0) < +\infty$. Voici un contre-exemple sans elle : on prend $\Omega = \mathbb{N}$, $\mathcal{F} = \mathcal{P}(\Omega)$ et $\mu = \sum_{n \in \mathbb{N}} \delta_n$ mesure de comptage. La suite des $C_n = \mathbb{N} \cap [n, +\infty[$ est décroissante d'intersection vide, mais comme pour tout n, $\mu(C_n) = +\infty$, $\mu(C_n)$ ne peut converger vers 0.

Preuve de 4(a). On remarque que pour tout $n \geq 1$, on a :

$$\bigcup_{i=0}^{n} A_i = \bigcup_{i=0}^{n} B_i,$$

où les B_i sont des ensembles deux à deux disjoints définis comme suit :

$$\begin{aligned} B_0 = A_0, \quad B_1 &= A_1 \cap B_0^c, \quad B_2 = A_2 \cap (B_0 \cup B_1)^c, \dots \\ \dots B_n &= A_n \cap (B_0 \cup B_1 \cup \dots B_{n-1})^c, \dots \end{aligned}$$

Par additivité :

$$\mu\Bigl(\bigcup_{i=0}^{n} A_i\Bigr) = \mu\Bigl(\bigcup_{i=0}^{n} B_i\Bigr) = \sum_{i=0}^{n} \mu(B_i).$$

Par construction pour tout i, $B_i \subset A_i$, d'où $\mu(B_i) \leq \mu(A_i)$ et :

$$\mu\Bigl(\bigcup_{i=0}^{n} A_i\Bigr) = \sum_{i=0}^{n} \mu(B_i) \leq \sum_{i=0}^{n} \mu(A_i).$$

$\square$

Preuve de 4(b). Posons pour tout $n \geq 1$,

$$D_n = \bigcup_{i=0}^{n} A_i, \qquad D = \bigcup_{n \geq 1} D_n = \bigcup_{i \in \mathbb{N}} A_i.$$

La suite $(D_n)_{n \geq 1}$ est *croissante* et a pour limite D. Donc d'après *3(a)*, $\mu(D_n) \uparrow \mu(D)$ $(n \to +\infty)$. D'après *4(a)* on a :

$$\forall n \geq 1, \quad \mu(D_n) \leq \sum_{i=0}^{n} \mu(A_i).$$

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 $+\infty$:

$$\mu(\bigcup_{i \in \mathbb{N}} A_i) = \mu(D) \leq \sum_{i=0}^{+\infty} \mu(A_i).$$

Ce qui prouve *4(b)*. Remarquons que les sommes partielles de la série convergent dans $\mathbb{R}^+ \cup \{+\infty\}$. $\square$

Proposition 1.23 (Formule de Poincaré)**.**
Soit μ une mesure finie *sur $(\Omega, \mathcal{F})$. Pour tout entier $n \geq 2$ et tous $A_1, \dots, A_n \in \mathcal{F}$:*

$$\mu\Bigl(\bigcup_{i=1}^{n} A_i\Bigr) = \sum_{i=1}^{n} \mu(A_i) + \sum_{k=2}^{n} (-1)^{k+1} \sum_{1 \leq i_1 < i_2 < \dots < i_k \leq n} \mu(A_{i_1} \cap \dots \cap A_{i_k}). \tag{1.11}$$

Preuve. On raisonne par récurrence. Pour $n = 2$, la formule se réduit à

$$\mu(A \cup B) = \mu(A) + \mu(B) - \mu(A \cap B). \tag{1.12}$$

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 l'additivité on en déduit :

$$\begin{aligned} \mu(A \cup B) &= \mu(A \cap B^c) + \mu(A \cap B) + \mu(A^c \cap B) \\ &= \big[\mu(A \cap B^c) + \mu(A \cap B)\big] + \big[\mu(A \cap B) + \mu(A^c \cap B)\big] - \mu(A \cap B) \\ &= \mu(A) + \mu(B) - \mu(A \cap B). \end{aligned}$$

Supposons la formule de Poincaré vraie au rang n (plus précisément on suppose que pour *toute* suite de n éléments $A_1, \dots, A_n$ de $\mathcal{F}$, l'égalité (1.11) est vérifiée). Pour en déduire qu'elle est alors vraie au rang $n + 1$, il nous faut calculer $\mu\Big(\overset{n+1}{\underset{i=1}{\cup}} A_i\Big)$. On commence par appliquer (1.12) avec $A = \overset{n}{\underset{i=1}{\cup}} A_i$ et $B = A_{n+1}$. On obtient ainsi :

$$\begin{aligned} \mu\Big(\overset{n+1}{\underset{i=1}{\cup}} A_i\Big) &= \mu\Big(\overset{n}{\underset{i=1}{\cup}} A_i\Big) + \mu(A_{n+1}) - \mu\Big(\Big[\overset{n}{\underset{i=1}{\cup}} A_i\Big] \cap A_{n+1}\Big) \\ &= \mu\Big(\overset{n}{\underset{i=1}{\cup}} A_i\Big) + \mu(A_{n+1}) - \mu\Big(\overset{n}{\underset{i=1}{\cup}} (A_i \cap A_{n+1})\Big). \end{aligned}$$

On applique maintenant l'hypothèse de récurrence (formule de Poincaré (1.11)) d'abord avec les n ensembles $A_1, \dots, A_n$ puis avec les n ensembles $A'_1, \dots, A'_n$, où l'on a posé $A'_i := A_i \cap A_{n+1}$. Il vient :

$$\begin{aligned} \mu\Big(\overset{n+1}{\underset{i=1}{\cup}} A_i\Big) = \quad & \sum_{i=1}^{n} \mu(A_i) + \sum_{k=2}^{n} (-1)^{k+1} \sum_{1 \le i_1 < i_2 < \ldots < i_k \le n} \mu(A_{i_1} \cap \cdots \cap A_{i_k}) \\ & + \mu(A_{n+1}) \\ & - \sum_{i=1}^{n} \mu(A'_i) - \sum_{j=2}^{n} (-1)^{j+1} \sum_{1 \le i_1 < i_2 < \ldots < i_j \le n} \mu(A'_{i_1} \cap \cdots \cap A'_{i_j}) \\ = \quad & \sum_{i=1}^{n+1} \mu(A_i) & (1.13) \\ & + \sum_{k=2}^{n} (-1)^{k+1} \sum_{1 \le i_1 < i_2 < \ldots < i_k \le n} \mu(A_{i_1} \cap \cdots \cap A_{i_k}) & (1.14) \\ & + (-1)^{2+1} \sum_{i=1}^{n} \mu(A_i \cap A_{n+1}) & (1.15) \\ & + \sum_{j=2}^{n} (-1)^{(j+1)+1} \sum_{1 \le i_1 < i_2 < \ldots < i_j \le n} \mu(A_{i_1} \cap \cdots \cap A_{i_j} \cap A_{n+1}). & (1.16) \end{aligned}$$

Comparons ce résultat avec ce que l'on espère trouver, c'est-à-dire avec

$$\sum_{i=1}^{n+1} \mu(A_i) + \underbrace{\sum_{k=2}^{n+1} (-1)^{k+1} \sum_{1 \leq i_1 < i_2 < \ldots < i_k \leq n+1} \mu(A_{i_1} \cap \cdots \cap A_{i_k})}_{=:T_{n+1}}.$$

Cela revient à vérifier que T_{n+1} est égal à la somme des lignes (1.14) à (1.16) 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 $i_k \leq n$ et $k \leq n$). On le retrouve exactement à la ligne (1.14). 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 (1.15). Il reste alors la somme des termes pour lesquels $3 \leq k \leq n+1$ et $i_k = n+1$ (donc $i_{k-1} \leq n$). Cette somme est exactement le contenu de la ligne (1.16), comme on peut le voir en faisant le changement d'indice $k = j+1$ dans (1.16). Ceci achève la récurrence. □

1.3.4 Fonction de répartition

Cette sous-section est une illustration de la proposition 1.21 et un prélude à la caractérisation des mesures finies sur la tribu borélienne de $\mathbb{R}$.

Définition 1.24. *Soit μ une mesure* finie *sur* $(\mathbb{R}, \mathrm{Bor}(\mathbb{R}))$*. Sa fonction de répartition (f.d.r.) est l'application*

$$F : \mathbb{R} \to \mathbb{R}_+, \quad x \mapsto F(x) := \mu\big(]-\infty, x]\big).$$

Théorème 1.25 (propriétés des f.d.r.)**.** *Soit F la fonction de répartition d'une mesure finie μ sur* $(\mathbb{R}, \mathrm{Bor}(\mathbb{R}))$*.*

i) F est croissante sur $\mathbb{R}$.

ii) F est continue à droite et pourvue d'une limite à gauche en tout point (fonction càdlàg*). De plus*

$$\lim_{x \to -\infty} F(x) = 0, \qquad \lim_{x \to +\infty} F(x) = \mu(\mathbb{R}).$$

iii) Pour tout $a \in \mathbb{R}$, $F(a) - F(a^-) = \mu(\{a\})$ et il n'y a qu'un ensemble au plus dénombrable de points a où $F(a) - F(a^-) > 0$.

iv) Si deux mesures finies μ et ν sur $(\mathbb{R}, \mathrm{Bor}(\mathbb{R}))$ *ont même f.d.r., elles sont égales.*

La propriété iv) est listée ici pour des raisons de commodité, mais nous n'avons pas encore les outils pour la prouver complètement. On la démontrera seulement pour des mesures ponctuelles au sens de l'exemple 1.3. Le cas général sera vu ultérieurement comme corollaire du théorème d'unicité (th. 1.34).

Preuve de i). Si $x < x'$, $]-\infty, x] \subset]-\infty, x']$ et par croissance de μ (prop. 1.21–2.), $\mu\big(]-\infty, x]\big) \leq \mu\big(]-\infty, x']\big)$, soit encore $F(x) \leq F(x')$. F est bien croissante. □

Preuve de ii). Comme fonction croissante, F a en tout point $a \in \mathbb{R}$, une limite à gauche $F(a^-) := \lim_{\varepsilon \to 0, \varepsilon > 0} F(a - \varepsilon)$ et une limite à droite $F(a^+) := \lim_{\varepsilon \to 0, \varepsilon > 0} F(a + \varepsilon)$. Pour établir la continuité à droite de F en a, il suffit donc de vérifier que $F(a^+) = F(a)$. L'*existence* de la limite à droite $F(a^+)$ étant acquise, il suffit finalement de vérifier que $F(a) = \lim_{n \to +\infty} F(a + 1/n)$. En posant $A_n :=]-\infty, a + 1/n]$ et $A =]-\infty, a]$, on a $F(a + 1/n) = \mu(A_n)$ et $F(a) = \mu(A)$. On voit immédiatement que $(A_n)_{n \geq 1}$ est une suite décroissante pour l'inclusion dans le tribu $\mathrm{Bor}(\mathbb{R})$ et que $\cap_{k \geq 1} A_k = A$. Par continuité décroissante séquentielle de μ (prop. 1.21–3.b)), on en déduit que $\mu(A) = \lim_{n \to +\infty} \mu(A_n)$.

La croissance de F sur $\mathbb{R}$ assure l'existence de limites en $-\infty$ et $+\infty$. Pour les identifier, il suffit donc de trouver $\lim_{n \to +\infty} F(-n)$ et $\lim_{n \to +\infty} F(n)$. En utilisant la continuité monotone séquentielle de μ (prop. 1.21–3.), on obtient quand n tend vers $+\infty$:

$$F(-n) = \mu\Big(]-\infty, -n]\Big) \ \downarrow\ \mu\Big(\bigcap_{k \geq 1}]-\infty, -k]\Big) = \mu(\emptyset) = 0;$$
$$F(n) = \mu\Big(]-\infty, n]\Big) \ \uparrow\ \mu\Big(\bigcup_{k \geq 1}]-\infty, k]\Big) = \mu(\mathbb{R}).$$

□

Preuve de iii). Posons $C_n :=]a - 1/n, a]$. Parce que μ est une mesure finie, il résulte de l'additivité finie de μ (Rem. 1.22) que

$$\mu(C_n) = \mu\Big(]-\infty, a] \setminus]-\infty, a - 1/n]\Big) = \mu\Big(]-\infty, a]\Big) - \mu\Big(]-\infty, a - 1/n]\Big) = F(a) - F(a - 1/n).$$

D'autre part $(C_n)_{n \geq 1}$ est une suite dans la tribu $\mathrm{Bor}(\mathbb{R})$, décroissante pour l'inclusion et son intersection est $\{a\}$. Donc par continuité décroissante séquentielle de μ, $\mu(C_n) = F(a) - F(a - 1/n)$ converge vers $\mu(\{a\})$, d'où $F(a) - F(a^-) = \mu(\{a\})$. Comme $F(a) = F(a^+)$, on voit ainsi que F est discontinue au point a si et seulement si $\mu(\{a\}) > 0$. Lorsque ceci arrive, $\mu(\{a\})$ est l'amplitude du « saut » de la fonction F en a et la discontinuité est de première espèce. On sait qu'une fonction monotone ne peut avoir de discontinuités que de première espèce (i.e. avec existence d'une limite à gauche et d'une limite à droite) et que l'ensemble de ces discontinuités est au plus dénombrable. □

Preuve de iv) dans le cas particulier des mesures finies ponctuelles. Soient μ et ν deux mesures finies ponctuelles au sens de l'exemple 1.3. On peut les représenter sous la forme

$$\mu = \sum_{k \in \mathbb{N}} a_k \delta_{x_k}, \qquad \nu = \sum_{k \in \mathbb{N}} b_k \delta_{y_k}.$$

Quitte à modifier les coefficients a_k, b_k, on peut toujours se ramener au cas où les x_k sont tous distincts et de même pour les y_k. On a alors pour tout $x \in \mathbb{R}$,

$$\mu(\{x\}) = \sum_{k \in \mathbb{N}} a_k \delta_{x_k}(\{x\}) = \begin{cases} a_j & \text{si } x = x_j \text{ pour un } j \in \mathbb{N}, \\ 0 & \text{si } x \notin \{x_k;\ k \in \mathbb{N}\}. \end{cases}$$

De même,

$$\nu(\{x\}) = \sum_{k\in\mathbb{N}} b_k \delta_{y_k}(\{x\}) = \begin{cases} b_j & \text{si } x = y_j \text{ pour un } j \in \mathbb{N}, \\ 0 & \text{si } x \notin \{y_k;\ k \in \mathbb{N}\}. \end{cases}$$

Notant F et G les fonctions de répartitions respectives de μ et ν, on déduit de iii) et de l'hypothèse $F = G$ que $\mu(\{x\}) = \nu(\{x\})$ pour tout $x \in \mathbb{R}$. Le calcul de $\mu(\{x\})$ et $\nu(\{x\})$ ci-dessus nous montre alors que pour tout $k \in \mathbb{N}$, $x_k = y_k$ et $a_k = b_k$. L'égalité de μ et ν en découle. □

1.3.5 Mesures extérieures

Définition 1.26. *Soit Ω un ensemble. Une* mesure extérieure *sur Ω est une application $\mu^* : \mathcal{P}(\Omega) \to \overline{\mathbb{R}}_+$, telle que*

i) $\mu^*(\emptyset) = 0$;

ii) μ^* est croissante : si $A \subset B \subset \Omega$, $\mu^*(A) \leq \mu^*(B)$.

iii) μ^* est sous-σ-additive : pour toute suite $(A_i)_{i\in\mathbb{N}}$ de parties de Ω,

$$\mu^*\Bigl(\bigcup_{i\in\mathbb{N}} A_i\Bigr) \leq \sum_{i\in\mathbb{N}} \mu^*(A_i). \tag{1.17}$$

Toute mesure sur $(\Omega, \mathcal{P}(\Omega))$ est une mesure extérieure (prop. 1.21–2. et 4.b)), mais bien sûr, l'intérêt de cette notion est qu'elle est plus générale que celle de mesure. Elle sert notamment à construire des mesures par restriction d'une mesure extérieure à une sous-tribu de $\mathcal{P}(\Omega)$. L'exemple fondamental suivant donne une construction très générale de mesures extérieures.

Exemple 1.6. Soit $\mathcal{E}$ une famille de parties de Ω, telle que $\emptyset \in \mathcal{E}$ et $\theta : \mathcal{E} \to \overline{\mathbb{R}}_+$ une fonction d'ensembles sur $\mathcal{E}$ telle que $\theta(\emptyset) = 0$. Pour tout $A \in \mathcal{P}(\Omega)$, on pose

$$\mu^*(A) := \inf \sum_{k\in\mathbb{N}} \theta(E_k), \tag{1.18}$$

où l'infimum est pris sur toutes les suites $(E_k)_{k\in\mathbb{N}}$ d'éléments de $\mathcal{E}$, dont la réunion $\cup_{k\in\mathbb{N}} E_k$ recouvre A. On fait la convention habituelle attribuant la valeur $+\infty$ à l'infimum d'un ensemble de réels lorsque ce dernier est vide (donc ici s'il n'y a aucun recouvrement de A par une suite dans $\mathcal{E}$). La fonction d'ensembles μ^* ainsi définie est une mesure extérieure. Remarquons la faiblesse des hypothèses faites sur $\mathcal{E}$ et θ. Pour obtenir des applications intéressantes, il nous faudra bien sûr enrichir la structure $(\mathcal{E}, \theta)$.

Vérification. D'abord il est clair que $\mu^*(A) \geq 0$ pour tout $A \in \mathcal{P}(\Omega)$ parce que la fonction θ est à valeurs dans $\overline{\mathbb{R}}_+$. Puisque $\emptyset$ est élément de $\mathcal{E}$, son recouvrement par la suite (E_k) dont tous les éléments sont $\emptyset$ donne $\mu^*(\emptyset) \leq 0$, donc $\mu^*(\emptyset) = 0$. Si $A \subset B \subset \Omega$, tout recouvrement de B est un recouvrement de A et on en déduit immédiatement que $\mu^*(A) \leq \mu^*(B)$. Pour montrer la sous-σ-additivité, soit $(A_i)_{i\in\mathbb{N}}$ une suite quelconque dans $\mathcal{P}(\Omega)$. Si l'un au moins des $\mu^*(A_i)$ vaut $+\infty$, (1.17) est automatiquement vérifiée.

Si tous les $\mu^*(A_i)$ sont finis, la définition de l'infimum (1.18) nous assure pour chaque i et chaque $\varepsilon > 0$, de l'existence d'un recouvrement $(E_{i,k},\ k \in \mathbb{N})$ de A_i tel que

$$\sum_{k\in\mathbb{N}} \theta(E_{i,k}) \leq \mu^*(A_i) + \frac{\varepsilon}{2^{i+1}}.$$

D'autre part, $A := \cup_{i\in\mathbb{N}} A_i$ a au moins un recouvrement par une union *dénombrable* d'éléments de $\mathcal{E}$, c'est $\cup_{(i,k)\in\mathbb{N}^2} E_{i,k}$, donc d'après la définition (1.18), on a

$$\mu^*(A) \leq \sum_{(i,k)\in\mathbb{N}^2} \theta(E_{i,k}) = \sum_{i\in\mathbb{N}} \sum_{k\in\mathbb{N}} \theta(E_{i,k}) \leq \sum_{i\in\mathbb{N}} \Bigl(\mu^*(A_i) + \frac{\varepsilon}{2^{i+1}}\Bigr) = \varepsilon + \sum_{i\in\mathbb{N}} \mu^*(A_i).$$

Ceci étant vrai pour tout $\varepsilon > 0$, on en déduit (1.17). $\square$

Définition 1.27. *Soit μ^* une mesure extérieure sur Ω. On dit que $A \in \mathcal{P}(\Omega)$ est μ^*-mesurable si*

$$\forall E \in \mathcal{P}(\Omega), \quad \mu^*(E) = \mu^*(A \cap E) + \mu^*(A^c \cap E). \tag{1.19}$$

En raison de la sous-additivité de μ^*, (1.19) équivaut à

$$\forall E \in \mathcal{P}(\Omega), \quad \mu^*(E) \geq \mu^*(A \cap E) + \mu^*(A^c \cap E). \tag{1.20}$$

On note $\mathcal{M}(\mu^*)$ l'ensemble des parties μ^*-mesurables de Ω.

Nous prouverons dans un instant que $\mathcal{M}(\mu^*)$ est une tribu. Il n'est guère aisé de donner une interprétation convaincante de (1.19). Pour en saisir la pertinence, le plus simple est peut-être de regarder comment elle fonctionne dans la preuve du lemme suivant.

Lemme 1.28. *La trace de μ^* sur tout $E \in \mathcal{P}(\Omega)$ est additive sur $\mathcal{M}(\mu^*)$: pour tout $n \geq 1$ et tous $A_1, \dots, A_n \in \mathcal{M}(\mu^*)$ deux à deux disjoints,*

$$\mu^*\Bigl(E \cap \Bigl(\bigcup_{j=1}^{n} A_j\Bigr)\Bigr) = \sum_{j=1}^{n} \mu^*(E \cap A_j). \tag{1.21}$$

En particulier pour $E = \Omega$, on obtient l'additivité de μ^* sur $\mathcal{M}(\mu^*)$.

Preuve. Remarquons qu'ici on ne peut pas se contenter de vérifier (1.21) pour $n = 2$ car on ne sait pas encore si $\mathcal{M}(\mu^*)$ est stable par union finie, ce qui empêche d'utiliser directement l'associativité de la réunion. L'égalité (1.21), vraie pour $n = 1$, se montre par récurrence. Supposons la vérifiée au rang n et soit $A_{n+1} \in \mathcal{M}(\mu^*)$, les A_j, $1 \leq j \leq n+1$ étant deux à deux disjoints. En utilisant (1.19) avec $E' := E \cap (\cup_{1\leq j\leq n+1} A_j)$ et $A := A_{n+1} \in \mathcal{M}(\mu^*)$, il vient

$$\mu^*(E') = \mu^*(E' \cap A_{n+1}) + \mu^*(E' \cap A_{n+1}^c). \tag{1.22}$$

Grâce à la disjonction des A_j, on a

$$E' \cap A_{n+1} = E \cap \Bigl(\bigcup_{j=1}^{n+1} A_j\Bigr) \cap A_{n+1} = \bigcup_{j=1}^{n+1} (E \cap A_j \cap A_{n+1}) = E \cap A_{n+1}$$

et comme les A_j pour $j \leq n$ sont tous inclus dans A_{n+1}^c, donc aussi leur réunion,

$$E' \cap A_{n+1}^c = \Big(E \cap \Big(\bigcup_{j=1}^{n} A_j\Big) \cap A_{n+1}^c\Big) \cup \big(E \cap A_{n+1} \cap A_{n+1}^c\big) = E \cap \Big(\bigcup_{j=1}^{n} A_j\Big).$$

En reportant ces évaluations dans (1.22), on obtient

$$\mu^*\Big(E \cap \Big(\bigcup_{j=1}^{n+1} A_j\Big)\Big) = \mu^*(E \cap A_{n+1}) + \mu^*\Big(E \cap \Big(\bigcup_{j=1}^{n} A_j\Big)\Big),$$

ce qui permet d'achever la récurrence. □

Théorème 1.29. $\mathcal{M}(\mu^*)$ *est une tribu et la restriction de* μ^* *à* $\mathcal{M}(\mu^*)$ *est une mesure.*

Preuve. Pour montrer que $\mathcal{M}(\mu^*)$ est une tribu, nous utilisons la proposition 1.4. L'ensemble vide est clairement un élément de $\mathcal{M}(\mu^*)$, donc $\mathcal{M}(\mu^*)$ n'est pas vide (sic !). La stabilité par complémentaire étant évidente, il reste à prouver la stabilité pour l'intersection finie et pour la réunion dénombrable disjointe.
a) Stabilité pour l'intersection finie. Soient A et B dans $\mathcal{M}(\mu^*)$. On a pour tout E les trois égalités suivantes :

$$\begin{aligned} \mu^*(E) &= \mu^*(A \cap E) + \mu^*(A^c \cap E) \\ \mu^*(A \cap E) &= \mu^*(B \cap A \cap E) + \mu^*(B^c \cap A \cap E) \\ \mu^*(A^c \cap E) &= \mu^*(B \cap A^c \cap E) + \mu^*(B^c \cap A^c \cap E) \end{aligned}$$

Puis, remarquant que $(B \cap A^c \cap E) \cup (B^c \cap A \cap E) \cup (B^c \cap A^c \cap E) = (A^c \cup B^c) \cap E$ on obtient par sous σ-additivité de μ^*,

$$\mu^*(A^c \cup B^c \cap E) \leq \mu^*(B \cap A^c \cap E) + \mu^*(B^c \cap A \cap E) + \mu^*(B^c \cap A^c \cap E),$$

d'où découle finalement $\mu^*(E) \geq \mu^*(A^c \cup B^c \cap E) + \mu^*(B \cap A \cap E)$, qui prouve que $A \cap B$ est dans $\mathcal{M}(\mu^*)$.
b) Stabilité pour la réunion dénombrable disjointe. Soit $(A_n)_{n\geq 1}$ une suite d'éléments deux à deux disjoints de $\mathcal{M}(\mu^*)$ et A sa réunion. La stabilité par complémentaire et par intersection finie de $\mathcal{M}(\mu^*)$ entraînent sa stabilité par union finie. Ainsi pour tout $n \geq 1$, $B_n := \bigcup_{j=1}^{n} A_j$ est dans $\mathcal{M}(\mu^*)$ et l'on a pour tout $E \subset \Omega$,

$$\mu^*(E) = \mu^*(E \cap B_n) + \mu^*(E \cap B_n^c).$$

Comme $B_n \subset A$, $A^c \subset B_n^c$, et par croissance de la mesure extérieure μ^*, on en déduit

$$\mu^*(E) \geq \mu^*(E \cap B_n) + \mu^*(E \cap A^c). \tag{1.23}$$

Par le lemme 1.28,

$$\mu^*(E \cap B_n) = \sum_{j=1}^{n} \mu^*(E \cap A_j),$$

ce qui reporté dans (1.23) nous donne

$$\mu^*(E) \geq \sum_{j=1}^{n} \mu^*(E \cap A_j) + \mu^*\Big(E \cap A^c\Big).$$

Ceci étant vrai pour tout n, on en déduit

$$\mu^*(E) \geq \sum_{j=1}^{+\infty} \mu^*(E \cap A_j) + \mu^*\Big(E \cap A^c\Big). \tag{1.24}$$

Or $E \cap A = \cup_{j \in \mathbb{N}^*}(E \cap A_j)$, d'où par sous-$\sigma$-additivité de μ^*,

$$\mu^*(E) \geq \mu^*\Big(E \cap A\Big) + \mu^*\Big(E \cap A^c\Big),$$

ce qui prouve bien que A est dans $\mathcal{M}(\mu^*)$. Ainsi $\mathcal{M}(\mu^*)$ est stable par union dénombrable disjointe et est bien une tribu.

c) μ^* restreinte à $\mathcal{M}(\mu^*)$ est une mesure. C'est une retombée immédiate de la preuve du b). En effet, on sait déjà que $\mu^*(\emptyset) = 0$. Soit $(A_n)_{n\geq 1}$ une suite d'éléments deux à deux disjoints de $\mathcal{M}(\mu^*)$ et A sa réunion. Le cas particulier $E = A$ dans (1.24) s'écrit

$$\mu^*(A) \geq \sum_{j=1}^{+\infty} \mu^*(A_j). \tag{1.25}$$

La sous-σ-additivité de μ^* fournit l'inégalité inverse, de sorte qu'on a l'égalité dans (1.25), ce qui exprime la σ-additivité de μ^* restreinte à $\mathcal{M}(\mu^*)$. □

1.3.6 Théorème d'extension

Grosso modo, le problème que nous allons étudier maintenant peut se décrire ainsi. On dispose d'une fonction d'ensemble μ sur une famille $\mathcal{C}$ de parties de Ω. À titre d'exemple, on peut prendre pour $\mathcal{C}$ la classe $\mathcal{I}$ des intervalles $]a, b]$ de $\Omega = \mathbb{R}$ et y définir μ par $\mu(]a, b]) := b - a$. On cherche des conditions suffisantes assez générales permettant de prolonger μ à toute la tribu $\sigma(\mathcal{C})$ de sorte que ce prolongement soit une mesure sur $\mathcal{C}$. Avec l'exemple mentionné, comme $\sigma(\mathcal{I}) = \mathrm{Bor}(\mathbb{R})$, on aura ainsi construit la mesure de Lebesgue sur $\mathbb{R}$, qui étend la fonction « longueur » de $\mathcal{I}$ à toute la tribu borélienne de $\mathbb{R}$. Commençons par préciser la structure de $\mathcal{C}$.

Définition 1.30. *On appelle* semi-algèbre *de parties de Ω, toute famille $\mathcal{C}$ de parties de Ω telle que :*

a) $\emptyset \in \mathcal{C}$,

b) $\mathcal{C}$ est stable par intersection finie,

c) pour tous $A \in \mathcal{C}$, $B \in \mathcal{C}$ tels que $A \subset B$, il existe une suite finie $C_1, \ldots, C_n$ d'éléments de $\mathcal{C}$, deux à deux disjoints tels que $B \setminus A = \bigcup_{k=1}^{n} C_k$.

Bien sûr le c) doit être compris comme englobant le cas particulier où $B \setminus A \in \mathcal{C}$, auquel cas on peut prendre $n = 1$, $C_1 = B \setminus A$ et la clause « deux à deux disjoints » devient sans objet. L'exemple fondamental suivant permet de motiver cette définition.

Exemple 1.7. La classe $\mathcal{I}$ des intervalles $]a, b]$ de $\mathbb{R}$ est une semi-algèbre de parties de $\mathbb{R}$. De même, la classe $\mathcal{C}$ des pavés de $\mathbb{R}^d$ de la forme $Q = \prod_{i=1}^{d}]a_i, b_i]$ est une semi-algèbre.

Vérification. L'intervalle $]a, b]$ est par définition $\{x \in \mathbb{R};\ a < x \leq b\}$. Ainsi pour $a = b$, $]a, a] = \emptyset \in \mathcal{I}$. Cette définition de l'intervalle nous montre immédiatement que $\cap_{1\leq k\leq n}]a_k, b_k] =]a, b]$ en posant $a := \max_{1\leq k\leq n} a_k$ et $b := \min_{1\leq k\leq n} b_k$ (et donc le résultat est $\emptyset$ quand $a \geq b$). Soit $]c, d] \subset]a, b]$ et supposons pour éviter le cas trivial que $c < d$. Alors nécessairement $a \leq c < d \leq b$ et $]a, b] \setminus]c, d] =]a, c] \cup]d, b]$. Ainsi $\mathcal{I}$ est bien une semi-algèbre. Notons au passage que notre vérification de la propriété c) n'est pas transposable aux cas des classes $\mathcal{I}_o$ d'intervalles ouverts $]a, b[$ ou $\mathcal{I}_f$ d'intervalles fermés $[a, b]$. Par contre la transposition à la classe $\mathcal{I}'$ des intervalles $[a, b[$ est immédiate.

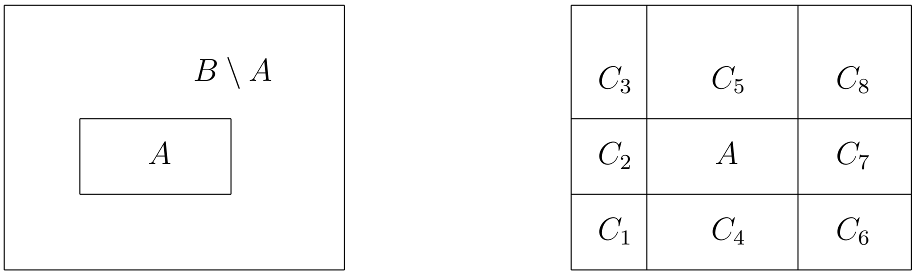
Passons à la classe $\mathcal{C}$ des pavés de $\mathbb{R}^d$ du type

$$Q = \prod_{i=1}^{d}]a_i, b_i] = \{x = (x_1, \dots, x_d) \in \mathbb{R}^d;\ \forall i = 1, \dots, d,\ a_i < x_i \leq b_i\}.$$

Sous cette dernière forme, on voit immédiatement que $\mathcal{C}$ contient $\emptyset$ (si $a_i \geq b_i$ pour au moins un i, $Q = \emptyset$) et est stable par intersection finie : en posant $a_i := \max_{1\leq k\leq n} a_{i,k}$ et $b_i := \min_{1\leq k\leq n} b_{i,k}$, on a

$$\bigcap_{k=1}^{n} \prod_{i=1}^{d}]a_{i,k}, b_{i,k}] = \prod_{i=1}^{d}]a_i, b_i].$$

Pour vérifier la condition c), soit $R := \prod_{i=1}^{d}]a'_i, b'_i]$ inclus dans $Q := \prod_{i=1}^{d}]a_i, b_i]$. Le cas $R = \emptyset$ est trivial. Si $R \neq \emptyset$, alors $Q \neq \emptyset$, d'où $a'_i < b'_i$ et $a_i < b_i$ pour tout $i = 1, \dots, d$. Le point $x = (b'_1, \dots, b'_d)$ appartient à R donc aussi à Q d'où $a_i < b'_i \leq b_i$ pour tout i. Si $m := \min_{1\leq i\leq d}(b'_i - a'_i)$, m est strictement positif et pour tout $0 < \varepsilon \leq m$, $y_\varepsilon := (a'_1 + \varepsilon, \dots, a'_d + \varepsilon)$ est dans R donc dans Q d'où $a_i < a'_i + \varepsilon \leq b'_i \leq b_i$ pour tout i. En faisant tendre ε vers 0, on en déduit que pour tout $i = 1, \dots, d$, $a_i \leq a'_i < b'_i \leq b_i$. On voit alors que le pavé Q peut être découpé en 3^d pavés $\prod_{i=1}^{d} I_{i,j(i)}$ où $j(i) \in \{1, 2, 3\}$ et $I_{i,1} :=]a_i, a'_i]$, $I_{i,2} :=]a'_i, b'_i]$, $I_{i,3} :=]b'_i, b_i]$. Ces 3^d pavés sont deux à deux disjoints (certains pouvant être vides). Le choix $j(i) = 2$ pour tout i donne le pavé R. On a donc trouvé $n = 3^d - 1$ pavés $C_1, \dots, C_n$ deux à deux disjoints tels que $Q \setminus R = \cup_{1\leq k\leq n} C_k$. La condition c) est vérifiée et $\mathcal{C}$ est une semi-algèbre. $\square$



La décomposition $B \setminus A = \overset{8}{\underset{i=1}{\cup}} C_i$, pour $d = 2$.

Théorème 1.31. *Soit $\mathcal{C}$ une semi-algèbre de parties de Ω et μ une application $\mathcal{C} \to \overline{\mathbb{R}}_+$ vérifiant :*

i) $\mu(\emptyset) = 0$ *;*

ii) *μ est additive sur $\mathcal{C}$: pour toute suite finie $C_0, \dots, C_n$ d'éléments deux à deux disjoints de $\mathcal{C}$ et telle que $\overset{n}{\underset{i=0}{\cup}} C_i$ appartienne à $\mathcal{C}$,*

$$\mu\Big(\overset{n}{\underset{i=0}{\cup}} C_i\Big) = \sum_{i=0}^{n} \mu(C_i);$$

iii) *μ est sous-σ-additive sur $\mathcal{C}$: pour toute suite $(A_i)_{i\in\mathbb{N}}$ d'éléments de $\mathcal{C}$ telle que $\underset{i\in\mathbb{N}}{\cup} A_i$ appartienne à $\mathcal{C}$,*

$$\mu\Big(\underset{i\in\mathbb{N}}{\cup} A_i\Big) \leq \sum_{i=0}^{+\infty} \mu(A_i).$$

Alors μ se prolonge en une mesure sur $\sigma(\mathcal{C})$.

Remarquons que la conclusion implique que μ devait déjà être σ-additive sur $\mathcal{C}$. Réciproquement, si μ est σ-additive sur $\mathcal{C}$ et $\mu(\emptyset) = 0$, alors elle vérifie ii) et iii). On n'aurait donc pas perdu de généralité en remplaçant dans l'énoncé ii) et iii) par l'hypothèse de σ-additivité de μ sur $\mathcal{C}$. Nous avons préféré la version énoncée ci-dessus pour des raisons pratiques.

Preuve. La preuve s'appuie sur le théorème 1.29 : on associe à μ la mesure extérieure μ^* définie par

$$\forall A \in \mathcal{P}(\Omega), \quad \mu^*(A) := \inf \sum_{k\in\mathbb{N}} \mu(D_k), \tag{1.26}$$

où l'infimum est pris sur toutes les suites $(D_k)_{k\in\mathbb{N}}$ d'éléments de $\mathcal{C}$, dont la réunion $\cup_{k\in\mathbb{N}} D_k$ recouvre A. $\mathcal{M}(\mu^*)$ étant la tribu associée à μ^* (cf. déf. 1.27 et th. 1.29), on sait que la restriction de μ^* à $\mathcal{M}(\mu^*)$ est une mesure. Le théorème 1.31 sera donc prouvé si nous montrons que

a) $\sigma(\mathcal{C}) \subset \mathcal{M}(\mu^*)$;

b) μ^* coïncide avec μ sur $\mathcal{C}$: $\forall B \in \mathcal{C}$, $\mu^*(B) = \mu(B)$.

En préliminaire, vérifions d'abord que μ est croissante sur $\mathcal{C}$. Si $A, B \in \mathcal{C}$ et $A \subset B$, alors il existe une suite finie $C_1, \dots, C_n$ d'éléments deux à deux disjoints de $\mathcal{C}$ telle que $B \setminus A = \cup_{1 \leq i \leq n} C_i$. Comme $B \setminus A$ est disjoint de A, chacun de ces C_i est disjoint de $A =: C_0$ et on peut décomposer B en réunion finie disjointe $B = \cup_{0 \leq i \leq n} C_i$ d'éléments de $\mathcal{C}$. Par additivité et positivité de μ sur $\mathcal{C}$, on a alors :

$$\mu(B) = \sum_{i=0}^{n} \mu(C_i) = \mu(A) + \sum_{i=1}^{n} \mu(C_i) \geq \mu(A).$$

Ainsi $\mu(A) \leq \mu(B)$ et μ est croissante sur $\mathcal{C}$.

a) $\mathcal{M}(\mu^*)$ contient $\sigma(\mathcal{C})$. Par minimalité de $\sigma(\mathcal{C})$, il suffit de montrer que la tribu $\mathcal{M}(\mu^*)$ contient $\mathcal{C}$. Soit B quelconque dans $\mathcal{C}$, nous allons établir l'appartenance de B à $\mathcal{M}(\mu^*)$ en vérifiant (1.20) :

$$\forall E \in \mathcal{P}(\Omega), \quad \mu^*(E) \geq \mu^*(B \cap E) + \mu^*(B^c \cap E).$$

Si $\mu^*(E) = +\infty$, cette inégalité est automatiquement vérifiée. Si $\mu^*(E) < +\infty$, pour tout $\varepsilon > 0$, il existe une suite $(D_k)_{k \in \mathbb{N}}$ dans $\mathcal{C}$ telle que $E \subset \cup_{k \in \mathbb{N}} D_k$ et

$$\sum_{k \in \mathbb{N}} \mu(D_k) \leq \mu^*(E) + \varepsilon. \tag{1.27}$$

Comme $\mathcal{C}$ est stable par intersection finie, chaque $B \cap D_k$ est élément de $\mathcal{C}$. Ainsi l'inclusion $B \cap E \subset \cup_{k \in \mathbb{N}} (B \cap D_k)$ nous fournit un recouvrement de $B \cap E$ par une suite de $\mathcal{C}$ et la définition de μ^* implique

$$\mu^*(B \cap E) \leq \sum_{k \in \mathbb{N}} \mu(B \cap D_k). \tag{1.28}$$

Le même argument ne s'applique pas aux $B^c \cap D_k$, car $\mathcal{C}$ n'est pas nécessairement stable par complémentaire. Cependant en remarquant que $B^c \cap D_k = D_k \setminus (B \cap D_k)$, on dispose pour chaque k d'une suite finie $(C_{j,k})_{1 \leq j \leq n_k}$ d'éléments deux à deux disjoints de $\mathcal{C}$ tels que

$$B^c \cap D_k = \bigcup_{j=1}^{n_k} C_{j,k}.$$

L'inclusion

$$B^c \cap E \subset \bigcup_{k \in \mathbb{N}} \bigcup_{j=1}^{n_k} C_{j,k}$$

nous fournit maintenant un recouvrement de $B^c \cap E$ par une union dénombrable d'éléments de $\mathcal{C}$, donc

$$\mu^*(B^c \cap E) \leq \sum_{k \in \mathbb{N}} \sum_{j=1}^{n_k} \mu(C_{j,k}). \tag{1.29}$$

En sommant (1.28) et (1.29), il vient

$$\mu^*(B \cap E) + \mu^*(B^c \cap E) \leq \sum_{k \in \mathbb{N}} \Big\{ \mu(B \cap D_k) + \sum_{j=1}^{n_k} \mu(C_{j,k}) \Big\}.$$

Grâce à l'additivité de μ sur $\mathcal{C}$ et à la décomposition

$$D_k = (B \cap D_k) \cup \bigcup_{j=1}^{n_k} C_{j,k}$$

en union finie disjointe d'éléments de $\mathcal{C}$, ceci peut se réécrire

$$\mu^*(B \cap E) + \mu^*(B^c \cap E) \leq \sum_{k \in \mathbb{N}} \mu(D_k).$$

En reportant dans (1.27), on obtient

$$\mu^*(E) \geq \mu^*(B \cap E) + \mu^*(B^c \cap E) - \varepsilon.$$

Cette inégalité étant vraie pour tout $\varepsilon > 0$, on en déduit $\mu^*(E) \geq \mu^*(B \cap E) + \mu^*(B^c \cap E)$. Comme E était quelconque dans $\mathcal{P}(\Omega)$, on a bien vérifié (1.20) pour B quelconque dans $\mathcal{C}$ et ainsi la tribu $\mathcal{M}(\mu^*)$ contient $\mathcal{C}$, donc aussi $\sigma(\mathcal{C})$.

b) μ^* coïncide avec μ sur $\mathcal{C}$. Soit $B \in \mathcal{C}$. En considérant le recouvrement particulier de B par la suite $(D_k)_{k \in \mathbb{N}}$ où $D_0 = B$ et $D_k = \emptyset$ pour tout $k \geq 1$, on obtient immédiatement

$$\mu^*(B) \leq \mu(B). \tag{1.30}$$

Soit maintenant $(D_k)_{k \in \mathbb{N}}$ un recouvrement dénombrable quelconque de B par des éléments de $\mathcal{C}$. On a alors $B = \cup_{k \in \mathbb{N}}(B \cap D_k)$, d'où par sous-$\sigma$-additivité de μ sur $\mathcal{C}$,

$$\mu(B) \leq \sum_{k \in \mathbb{N}} \mu(B \cap D_k).$$

Par croissance de μ sur $\mathcal{C}$, $\mu(B \cap D_k) \leq \mu(D_k)$, d'où

$$\mu(B) \leq \sum_{k \in \mathbb{N}} \mu(D_k).$$

Ceci étant vrai pour tout recouvrement dénombrable $(D_k)_{k \in \mathbb{N}}$ de B par des éléments de $\mathcal{C}$, on en déduit en prenant l'infimum sur tous ces recouvrements :

$$\mu(B) \leq \mu^*(B). \tag{1.31}$$

Cette inégalité jointe à (1.30) nous donne l'égalité $\mu(B) = \mu^*(B)$ pour tout $B \in \mathcal{C}$ et ceci termine la preuve du théorème. $\square$

Nous terminons cette sous-section par deux lemmes élémentaires qui ont leur utilité pour la vérification pratique de la sous-σ-additivité sur $\mathcal{C}$.

Lemme 1.32. *Soient $\mathcal{C}$ une semi-algèbre de parties de Ω, $A \in \mathcal{C}$ et $A_1, A_2, \dots, A_n$ une suite finie dans $\mathcal{C}$, de réunion incluse dans A. Alors $A \setminus (A_1 \cup \dots \cup A_n)$ est union finie disjointe d'éléments de $\mathcal{C}$.*

Preuve. On le vérifie par récurrence sur n. Pour $n = 1$, la propriété énoncée se réduit au c) de la définition 1.30 d'une semi-algèbre. Supposons la propriété vraie au rang n. Soit $A_1, A_2, \dots, A_n, A_{n+1}$ une suite finie dans $\mathcal{C}$, de réunion incluse dans A. Par hypothèse de récurrence, on a une décomposition

$$A \setminus (A_1 \cup \dots \cup A_n) = \bigcup_{j=1}^{m_n} C_j,$$

les C_j étant des éléments deux à deux disjoints de $\mathcal{C}$. On peut alors écrire

$$\begin{aligned} A \setminus (A_1 \cup \dots \cup A_n \cup A_{n+1}) &= A \cap (A_1^c \cap \dots \cap A_n^c) \cap A_{n+1}^c \\ &= \big(A \setminus (A_1 \cup \dots \cup A_n)\big) \cap A_{n+1}^c \\ &= \bigcup_{j=1}^{m_n} (C_j \cap A_{n+1}^c) \\ &= \bigcup_{j=1}^{m_n} \big(C_j \setminus (C_j \cap A_{n+1})\big). \end{aligned}$$

Comme $\mathcal{C}$ est stable par intersection finie, $C_j \cap A_{n+1}$ appartient à $\mathcal{C}$ et en appliquant la propriété c) de la définition d'une semi-algèbre, on dispose pour chaque j d'une suite finie $(D_{j,k})_{1 \leq k \leq n_j}$ d'éléments deux à deux disjoints de $\mathcal{C}$ telle que

$$C_j \setminus (C_j \cap A_{n+1}) = \bigcup_{k=1}^{n_j} D_{j,k}.$$

La suite finie dans $\mathcal{C}$ obtenue en rassemblant tous les $D_{j,k}$ pour $1 \leq j \leq m_n$, $1 \leq k \leq n_j$ a ses éléments deux à deux disjoints, par construction pour les paires ayant le même indice j et pour les autres paires, parce que les $C_j \setminus (C_j \cap A_{n+1})$ sont inclus dans les C_j eux-mêmes deux à deux disjoints. On a donc obtenu la décomposition en réunion finie disjointe d'éléments de $\mathcal{C}$:

$$A \setminus (A_1 \cup \dots \cup A_n \cup A_{n+1}) = \bigcup_{j=1}^{m_n} \bigcup_{k=1}^{n_j} D_{j,k},$$

ce qui achève la récurrence. □

Lemme 1.33. *Soient $\mathcal{C}$ une semi-algèbre de parties de Ω, μ une fonction d'ensembles définie sur $\mathcal{C}$, à valeurs dans $\overline{\mathbb{R}}_+$ et additive sur $\mathcal{C}$ et $A \in \mathcal{C}$.*

a) Si $A_1, \dots, A_n \in \mathcal{C}$ sont deux à deux disjoints et de réunion incluse dans A,

$$\sum_{j=1}^{n} \mu(A_j) \leq \mu(A).$$

b) Soient $B_1, \dots, B_n \in \mathcal{C}$. On suppose que $A \subset \bigcup_{j=1}^{n} B_j$. Alors

$$\mu(A) \leq \sum_{j=1}^{n} \mu(B_j).$$

Preuve. Le a) est une conséquence immédiate du lemme 1.32 qui nous fournit la décomposition :

$$A = \Big(\bigcup_{j=1}^{n} A_j\Big) \cup \Big(\bigcup_{k=1}^{m_n} C_k\Big),$$

les C_k et les A_j étant tous dans $\mathcal{C}$ et deux à deux disjoints. Par additivité de μ sur $\mathcal{C}$, il en résulte

$$\mu(A) = \sum_{j=1}^{n} \mu(A_j) + \sum_{k=1}^{m_n} \mu(C_k) \geq \sum_{j=1}^{n} \mu(A_j).$$

Pour prouver le b), remarquons d'abord que

$$A = A \cap \Big(\bigcup_{j=1}^{n} B_j\Big) = \bigcup_{j=1}^{n} (A \cap B_j).$$

Ensuite, en posant $C_1 := A \cap B_1$ et pour $j > 1$, $C_j := (A \cap B_j) \setminus \bigcup_{i=1}^{j-1} (A \cap B_i)$, on a (cf. la preuve de la proposition 1.4)

$$A = \bigcup_{j=1}^{n} C_j \quad \text{(union disjointe)}.$$

Comme $\mathcal{C}$ est stable par intersection finie, on peut appliquer le lemme 1.32 à chaque C_j pour le décomposer en union finie disjointe d'éléments $D_{j,k}$ $(1 \leq k \leq m_j)$ de $\mathcal{C}$. On a alors l'union disjointe

$$A = \bigcup_{j=1}^{n} \bigcup_{k=1}^{m_j} D_{j,k} \quad \text{d'où} \quad \mu(A) = \sum_{j=1}^{n} \sum_{k=1}^{m_j} \mu(D_{j,k}).$$

Pour j fixé, les $D_{j,k}$ sont deux à deux disjoints et de réunion égale à C_j, donc incluse dans B_j. Par le a), on a alors

$$\sum_{k=1}^{m_j} \mu(D_{j,k}) \leq \mu(B_j).$$

En reportant dans la décomposition de $\mu(A)$ ci-dessus, on aboutit à

$$\mu(A) = \sum_{j=1}^{n} \sum_{k=1}^{m_j} \mu(D_{j,k}) \leq \sum_{j=1}^{n} \mu(B_j),$$

ce qui est bien la conclusion annoncée. □

1.3.7 Théorème d'unicité

Théorème 1.34. *Soient μ et ν deux mesures définies sur le même espace mesurable $(\Omega, \mathcal{F})$ et qui coïncident sur une π-classe $\mathcal{P}$ ($\mathcal{P} \subset \mathcal{F}$).*

a) Si $\mu(\Omega) = \nu(\Omega) < +\infty$, alors μ et ν coïncident sur $\sigma(\mathcal{P})$ (qui est une sous-tribu de $\mathcal{F}$).

b) *Si $\mu(\Omega) = +\infty$ et s'il existe dans $\mathcal{P}$ une suite $(A_n)_{n\in\mathbb{N}^*}$ de réunion Ω et telle que pour tout $n \in \mathbb{N}^*$, $\mu(A_n) < +\infty$, alors μ et ν coïncident sur $\sigma(\mathcal{P})$.*

Preuve de a). Soit $\mathcal{L} := \{A \in \mathcal{F};\ \mu(A) = \nu(A)\}$. Par hypothèse, $\mathcal{L}$ contient $\mathcal{P}$. Si on montre que $\mathcal{L}$ est une λ-classe, le théorème de Dynkin (th. 1.15) nous donnera l'inclusion $\sigma(\mathcal{P}) \subset \mathcal{L}$, qui signifie que $\mu(A) = \nu(A)$ pour tout $A \in \sigma(\mathcal{P})$, ce qui est bien la conclusion annoncée. Vérifions donc que $\mathcal{L}$ est une λ-classe. $\Omega \in \mathcal{L}$ puisque par hypothèse $\mu(\Omega) = \nu(\Omega)$. Si A et B sont dans $\mathcal{F}$ et $A \subset B$, comme μ et ν sont finies, on a (cf. remarque 1.22),

$$\mu(B \setminus A) = \mu(B) - \mu(A) = \nu(B) - \nu(A) = \nu(B \setminus A),$$

donc $B \setminus A$ est dans $\mathcal{L}$, ce qui établit la stabilité par différence propre. Soit enfin $(A_n)_{n\in\mathbb{N}}$ une suite dans $\mathcal{L}$, croissante pour l'inclusion et A sa réunion. Pour tout $n \in \mathbb{N}$, on a $\mu(A_n) = \nu(A_n)$ et par continuité croissante séquentielle des mesures, $\mu(A_n) \uparrow \mu(A)$ et $\nu(A_n) \uparrow \nu(A)$ quand n tend vers l'infini. On en déduit que $\mu(A) = \nu(A)$, d'où la stabilité de $\mathcal{L}$ par union dénombrable croissante. Ainsi $\mathcal{L}$ est bien une λ-classe et ceci achève la preuve du cas a). $\square$

Preuve de b). Soit $A \in \sigma(\mathcal{P})$. On montre que $\mu(A) = \nu(A)$ en utilisant un argument d'approximation par des mesures finies et le résultat du a). Pour celà, on note d'abord que

$$\mu(A) = \mu\Bigl(A \cap \Bigl(\bigcup_{i=1}^{+\infty} A_i\Bigr)\Bigr) = \mu\Bigl(\bigcup_{i=1}^{+\infty} (A \cap A_i)\Bigr) = \lim_{n\to+\infty} \mu\Bigl(\bigcup_{i=1}^{n} (A \cap A_i)\Bigr),$$

et que la même expression est valable en remplaçant μ par ν. Remarquons que sous cette forme et même sous l'hypothèse plus restrictive $A \in \mathcal{P}$ au lieu de $A \in \sigma(\mathcal{P})$, il n'est pas clair directement que $\mu\Bigl(\bigcup_{i=1}^{n} (A \cap A_i)\Bigr) = \nu\Bigl(\bigcup_{i=1}^{n} (A \cap A_i)\Bigr)$. En effet $\mathcal{P}$ n'est pas nécessairement stable par réunion finie. Pour se débarasser de cette réunion, on fait appel à la formule de Poincaré (prop. 1.23) qui s'écrit ici :

$$\mu\Bigl(\bigcup_{i=1}^{n} (A \cap A_i)\Bigr) = \sum_{i=1}^{n} \mu(A \cap A_i) + \sum_{k=2}^{n} (-1)^{k+1} \sum_{1\le i_1<i_2<\ldots<i_k\le n} \mu(A \cap A_{i_1} \cap \cdots \cap A_{i_k}).$$

En appliquant la même formule avec ν à la place de μ, on voit que le problème sera résolu si l'on prouve que pour tous choix de $1 \le k \le n$ et $1 \le i_1 < i_2 < \ldots < i_k \le n$,

$$\forall A \in \sigma(\mathcal{P}), \quad \mu(A \cap A_{i_1} \cap \cdots \cap A_{i_k}) = \nu(A \cap A_{i_1} \cap \cdots \cap A_{i_k}).$$

Remarquons maintenant que si l'on pose $B = A_{i_1} \cap \cdots \cap A_{i_k}$, B est dans $\mathcal{P}$ et $\nu(B) = \mu(B) \le \max_{1\le j\le k} \mu(A_{i_j}) < +\infty$. L'ultime réduction du problème consiste ainsi à vérifier que :

$$\forall A \in \sigma(\mathcal{P}), \quad \forall B \in \mathcal{P} \text{ tel que } \mu(B) = \nu(B) < +\infty, \quad \mu(A \cap B) = \nu(A \cap B). \tag{1.32}$$

Pour cela, on considère les mesures μ' et ν' définies par

$$\forall A \in \mathcal{F}, \quad \mu'(A) := \mu(A \cap B), \quad \nu'(A) := \nu(A \cap B).$$

Ce sont bien des mesures sur $\mathcal{F}$, d'après l'exemple 1.4. Elles sont finies puisque $\mu'(\Omega) = \mu(B) < +\infty$ (idem pour ν'). Elles coïncident sur $\mathcal{P}$ puisque si $A \in \mathcal{P}$, $A \cap B \in \mathcal{P}$ et $\mu = \nu$ sur $\mathcal{P}$. D'après le a), μ' et ν' coïncident sur $\sigma(\mathcal{P})$. Ceci étant vrai pour tout $B \in \mathcal{P}$ tel que $\mu(B) = \nu(B) < +\infty$, (1.32) est établie, ce qui achève la preuve du théorème d'unicité. □

À titre de première application du théorème d'unicité, nous pouvons maintenant compléter la preuve du théorème 1.25 iv).

Corollaire 1.35. *Si deux mesures finies μ et ν sur $(\mathbb{R}, \mathrm{Bor}(\mathbb{R}))$ ont même fonction de répartition, elles sont égales.*

Preuve. La famille d'intervalles $\mathcal{P} := \{\,]-\infty, x];\ x \in \mathbb{R}\}$ est une π-classe qui engendre la tribu borélienne de $\mathbb{R}$. L'égalité des fonctions de répartition F de μ et G de ν n'est qu'une autre manière d'exprimer la coïncidence de μ et ν sur $\mathcal{P}$. De plus par continuité croissante séquentielle :

$$\mu(\mathbb{R}) = \lim_{n\to+\infty} \mu(]-\infty, n]) = \lim_{n\to+\infty} F(n) = \lim_{n\to+\infty} G(n) = \lim_{n\to+\infty} \nu(]-\infty, n]) = \nu(\mathbb{R}).$$

Donc par le théorème 1.34 a), μ et ν coïncident sur $\sigma(\mathcal{P}) = \mathrm{Bor}(\mathbb{R})$. □

1.3.8 Mesures de Stieltjes sur $\mathbb{R}$

Une application importante des théorèmes 1.31 et 1.34 est la construction de mesures sur $(\mathbb{R}, \mathrm{Bor}(\mathbb{R}))$ à partir d'une fonction croissante. En particulier on obtient ainsi la mesure de Lebesgue λ_1 sur $\mathbb{R}$.

Théorème 1.36. *Soit $F : \mathbb{R} \to \mathbb{R}$ une fonction croissante et continue à droite en tout point. Définissons sur $\mathcal{C} := \{\,]a, b];\ a, b \in \mathbb{R}\}$ la fonction d'ensembles μ par $\mu(\emptyset) = 0$ et $\mu(]a, b]) = F(b) - F(a)$ si $a < b$. Alors μ se prolonge de manière unique en une mesure sur la tribu borélienne de $\mathbb{R}$. Cette mesure est appelée* mesure de Stieltjes *associée à F. En particulier lorsque $F : x \mapsto x$ est l'identité sur $\mathbb{R}$, on obtient ainsi la mesure de Lebesgue λ (ou λ_1) qui prolonge à $\mathrm{Bor}(\mathbb{R})$ la fonction longueur des intervalles de $\mathcal{C}$ ($\lambda(]a, b]) = b - a$, si $a < b$).*

Preuve. On sait que $\sigma(\mathcal{C}) = \mathrm{Bor}(\mathbb{R})$. On montre l'existence par le théorème 1.31 et l'unicité par le théorème 1.34.
Additivité de μ sur $\mathcal{C}$. Il s'agit de vérifier que pour toute décomposition en union finie disjointe $]a, b] = \bigcup_{k=1}^{n} I_k$, avec $I_k :=]a_k, b_k]$,

$$F(b) - F(a) = \sum_{k=1}^{n} \Big(F(b_k) - F(a_k) \Big). \tag{1.33}$$

On ne perd pas de généralité en supposant qu'aucun I_k n'est vide, donc que $a_k < b_k$ pour tout k. L'égalité (1.33) devient évidente par téléscopage des termes lorsque les extrémités des I_k forment une subdivision de $]a, b]$ de la forme

$$a = c_0 < c_1 < c_2 < \cdots < c_{n-1} < c_n = b, \quad a_k = c_{k-1},\ b_k = c_k, \quad 1 \le k \le n. \tag{1.34}$$

Il suffit donc de montrer que l'on peut se ramener à cette situation en réindexant les intervalles I_k. Réindexons les en intervalles I'_j $(1 \leq j \leq n)$ avec $I'_j =]a'_j, b'_j]$ de façon que la suite $(b'_j)_{1\leq j\leq n}$ soit croissante. S'il existait un j tel que $b'_j > a'_{j+1}$, on aurait $a'_{j+1} < b'_j \leq b'_{j+1}$, donc $b'_j \in I'_{j+1}$, d'où $I'_j \cap I'_{j+1} \neq \emptyset$, ce qui contredirait l'hypothèse que les intervalles sont deux à deux disjoints. Ainsi pour tout $j = 1, \dots, n-1$, $b'_j \leq a'_{j+1}$. Supposons maintenant qu'il existe $1 \leq j_0 < n$ pour lequel cette inégalité soit stricte : $b'_{j_0} < a'_{j_0+1}$. Alors a'_{j_0+1} ne pourrait être dans $\cup_{j=1}^n I'_j$ puisque

$$a'_{j_0+1} > b_{j_0} \Rightarrow a'_{j_0+1} \notin \bigcup_{j=1}^{j_0} I'_j, \quad a'_{j_0+1} \notin]a'_{j_0+1}, b'_{j_0+1}], \quad a'_{j_0+1} < b'_{j_0+1} \Rightarrow a'_{j_0+1} \notin \bigcup_{j=j_0+2}^{n} I'_j.$$

Ceci est impossible car $a'_{j_0+1} \in]a,b]$ puisque d'une part $a'_{j_0+1} \geq b'_{j_0} \geq b'_1 > a$ et d'autre part $a'_{j_0+1} \leq b'_n \leq b$. On a donc pour tout $1 \leq j \leq n-1$, $b'_j = a'_{j+1}$. Tout ceci nous donne la configuration

$$a \leq a'_1 < b'_1 = a'_2 < b'_2 = a'_3 < \cdots < b'_{n-1} = a'_n < b'_n \leq b.$$

Si $a < a'_1$, alors $(a + a'_1)/2 \in]a,b]$ et $(a+a'_1)/2 \notin \cup_{1\leq j\leq n} I'_j$, donc nécessairement $a = a'_1$. De même $b'_n = b$. On a donc bien une configuration du type (1.34) avec $c_j = b'_j$, $c_{j-1} = a'_j$, $1 \leq j \leq n$. D'où

$$\sum_{k=1}^n \mu(I_k) = \sum_{j=1}^n \mu(I'_j) = \sum_{j=1}^n \Big(F(c_j) - F(c_{j-1})\Big) = F(c_n) - F(c_0) = F(b) - F(a) = \mu(]a,b]).$$

L'additivité de μ sur $\mathcal{C}$ est ainsi établie.

Sous-σ-additivité de μ sur $\mathcal{C}$. Il s'agit de montrer que

$$]a,b] = \bigcup_{k\in\mathbb{N}^*}]a_k, b_k] \quad \Rightarrow \quad F(b) - F(a) \leq \sum_{k=1}^{+\infty} \Big(F(b_k) - F(a_k)\Big). \tag{1.35}$$

Nous allons utiliser un argument de compacité, la continuité à droite de F permettant de remplacer le recouvrement de $]a,b]$ par les $]a_k, b_k]$ par un recouvrement d'ouverts avec une erreur aussi petite que l'on veut. Fixons pour commencer $\varepsilon > 0$. La croissance et la continuité à droite de F au point a nous assurent de l'existence d'un $\delta_0 \in]0, b-a[$ tel que

$$F(a) \leq F(a + \delta_0) \leq F(a) + \varepsilon. \tag{1.36}$$

On recouvre alors comme suit $[a + \delta_0, b]$:

$$[a+\delta_0, b] \subset]a,b] = \bigcup_{k\in\mathbb{N}^*}]a_k, b_k] \subset \bigcup_{k\in\mathbb{N}^*}]a_k, b_k + \delta_k[, \tag{1.37}$$

où les δ_k $(k \geq 1)$ sont choisis grâce à la continuité à droite de F en b_k, de telle sorte que

$$\forall k \geq 1, \quad F(b_k) \leq F(b_k + \delta_k) \leq F(b_k) + \varepsilon 2^{-k}. \tag{1.38}$$

Du recouvrement du compact $[a + \delta_0, b]$ par la suite d'ouverts $\Big(]a_k, b_k + \delta_k[\Big)_{k\geq 1}$, on peut extraire un recouvrement *fini*. Il existe donc un entier n (dépendant de F, de ε et des suites (a_k), (b_k)) tel que

$$]a + \delta_0, b] \subset [a + \delta_0, b] \subset \bigcup_{k=1}^n]a_k, b_k + \delta_k[\subset \bigcup_{k=1}^n]a_k, b_k + \delta_k].$$

Ayant déjà établi l'additivité de μ sur la semi-algèbre $\mathcal{C}$, nous pouvons appliquer le lemme 1.33 b) à cette suite d'inclusions pour obtenir $\mu(\,]a+\delta_0, b]) \leq \sum_{k=1}^n \mu(\,]a_k, b_k+\delta_k])$, ce qui s'écrit aussi :

$$F(b) - F(a+\delta_0) \leq \sum_{k=1}^{n} \Big(F(b_k+\delta_k) - F(a_k) \Big). \tag{1.39}$$

Compte-tenu de (1.36) et (1.38) et de la positivité des $F(b_k) - F(a_k)$ par croissance de F, on en déduit :

$$F(b) - F(a) - \varepsilon \leq \sum_{k=1}^{n} \Big(F(b_k) - F(a_k) \Big) + \sum_{k=1}^{n} \varepsilon 2^{-k} \leq \sum_{k=1}^{+\infty} \Big(F(b_k) - F(a_k) \Big) + \varepsilon.$$

Finalement, nous avons montré que :

$$\forall \varepsilon > 0, \quad F(b) - F(a) \leq \sum_{k=1}^{+\infty} \Big(F(b_k) - F(a_k) \Big) + 2\varepsilon.$$

En faisant tendre ε vers 0, on en déduit (1.35). Ceci termine la vérification des hypothèses du théorème 1.31. L'existence d'une mesure prolongeant μ de $\mathcal{C}$ à $\operatorname{Bor}(\mathbb{R})$ est maintenant acquise.

Unicité du prolongement de μ *à* $\operatorname{Bor}(\mathbb{R})$. Si μ et ν sont deux mesures sur $\operatorname{Bor}(\mathbb{R})$, qui coïncident sur $\mathcal{C}$, elles sont σ-finies puisque $\mathbb{R} = \cup_{n\in\mathbb{N}^*}]-n, n]$ et

$$\mu(]-n, n]) = \nu(]-n, n]) = F(n) - F(-n) < +\infty,$$

en rappelant que F est à valeurs dans $\mathbb{R}$, donc $F(n) - F(-n)$ est un *nombre* réel, donc fini. Par continuité séquentielle croissante des mesures, comme $]-n, n] \uparrow \mathbb{R}$, on en déduit que $\mu(\mathbb{R}) = \nu(\mathbb{R}) \leq +\infty$. D'autre part, $\mathcal{C}$ est une π-classe qui engendre $\operatorname{Bor}(\mathbb{R})$. Par le théorème 1.34, cas a) si $\mu(\mathbb{R}) < +\infty$ et b) sinon, les deux mesures μ et ν coïncident sur $\sigma(\mathcal{C}) = \operatorname{Bor}(\mathbb{R})$. □

Nous disposons donc maintenant d'une grande variété de mesures sur $(\mathbb{R}, \operatorname{Bor}(\mathbb{R}))$, puisque toute fonction croissante continue à droite F génère une mesure de Stieltjes $\mu = \mu_F$. Bien sûr cette représentation $\mu = \mu_F$ n'est pas unique puisque pour toute constante c, les fonctions F et $F + c$ génèrent la même mesure de Stieltjes. Malgré sa richesse, ce procédé de construction de mesures sur $(\mathbb{R}, \operatorname{Bor}(\mathbb{R}))$ ne les fournit pas toutes, comme on peut le voir par la caractérisation suivante des mesures de Stieltjes.

Proposition 1.37. *Une mesure* μ *sur* $(\mathbb{R}, \operatorname{Bor}(\mathbb{R}))$ *est de Stieltjes si et seulement si elle est finie sur les intervalles* bornés *de* $\mathbb{R}$.

Preuve. Soit μ une mesure de Stieltjes, associée à la fonction croissante $F : \mathbb{R} \to \mathbb{R}$. Si I est un intervalle borné de $\mathbb{R}$, on peut l'inclure dans un $]a, b]$ avec a et b réels finis. Alors par croissance des mesures pour l'inclusion,

$$0 \leq \mu(I) \leq \mu(]a, b]) = F(b) - F(a) < +\infty,$$

puisque F est à valeurs dans $\mathbb{R}$. Ceci montre qu'une mesure de Stieltjes est finie sur tout intervalle borné [5].

Réciproquement, soit μ une mesure sur $(\mathbb{R}, \mathrm{Bor}(\mathbb{R}))$, telle que pour tout intervalle borné I de $\mathbb{R}$, $\mu(I) < +\infty$. Montrons que μ est de Stieltjes. Traitons d'abord le cas particulier où μ est une mesure finie ($\mu(\mathbb{R}) < +\infty$). On prend alors pour F la fonction de répartition de μ : $F(x) := \mu(]-\infty, x])$. On sait que F est croissante et continue à droite (th. 1.25). Notons μ_F la mesure de Stieltjes associée à F. Comme μ est finie, la remarque 1.22 nous permet d'écrire pour tout intervalle $]a, b]$,

$$\mu(]a,b]) = \mu(]-\infty,b] \setminus]-\infty,a]) = \mu(]-\infty,b]) - \mu(]-\infty,a]) = F(b) - F(a) = \mu_F(]a,b]).$$

Ainsi les mesures μ et μ_F coïncident sur la classe des intervalles $]a, b]$ donc sur $\mathrm{Bor}(\mathbb{R})$, par le théorème d'unicité. Donc $\mu = \mu_F$ est bien une mesure de Stieltjes.

Dans le cas général, associons à μ la fonction F définie par

$$F(x) := \begin{cases} \mu(]0,x]) & \text{si } x \geq 0, \\ -\mu(]x,0]) & \text{si } x < 0. \end{cases}$$

μ étant finie sur les intervalles bornés, F est à valeurs dans $\mathbb{R}$. F est croissante car

- si $0 \leq x \leq y$, $]0,x] \subset]0,y]$ donc $F(x) = \mu(]0,x]) \leq \mu(]0,y]) = F(y)$;
- si $x < y < 0$, $]x,0] \supset]y,0]$, donc $\mu(]x,0]) \geq \mu(]y,0])$ et $-\mu(]x,0]) \leq -\mu(]y,0])$, d'où $F(x) \leq F(y)$;
- si $x < 0 \leq y$, $F(x) \leq 0$ et $F(y) \geq 0$, donc $F(x) \leq F(y)$.

Comme fonction monotone, F admet une limite à droite en tout point a de $\mathbb{R}$. Pour montrer que F est continue à droite au point a, il suffit donc de vérifier que $F(a + 1/n)$ tend vers $F(a)$ quand n tend vers $+\infty$. Ceci résulte immédiatement de la continuité monotone séquentielle de la mesure μ en notant que

- si $a \geq 0$, la suite d'intervalles $(]0, a + 1/n])_{n\geq 1}$ décroît (pour l'inclusion) et a pour intersection $]0, a]$;
- si $a < 0$, la suite $(]a + 1/n, 0])_{n\geq 1}$ croît (pour l'inclusion) et a pour réunion $]a, 0]$.

Ainsi la fonction croissante continue à droite F engendre une mesure de Stieltjes μ_F et pour voir que $\mu = \mu_F$, il ne reste plus qu'à vérifier que si $-\infty < a < b < +\infty$, $\mu(]a, b]) = F(b) - F(a)$. Par additivité finie de μ et la remarque 1.22, on a :

- si $0 \leq a < b$, $\mu(]a,b]) = \mu(]0,b] \setminus]0,a]) = \mu(]0,b]) - \mu(]0,a]) = F(b) - F(a)$;
- si $a < 0 \leq b$, le découpage en union disjointe $]a,b] =]a,0] \cup]0,b]$ donne $\mu(]a,b]) = \mu(]a,0]) + \mu(]0,b]) = -F(a) + F(b)$;
- si $a < b < 0$, $\mu(]a,b]) = \mu(]a,0] \setminus]b,0]) = \mu(]a,0]) - \mu(]b,0]) = -F(a) - (-F(b)) = F(b) - F(a)$.

Le théorème d'unicité nous permet de conclure que $\mu = \mu_F$ est bien une mesure de Stieltjes. □

5. Attention, ceci ne signifie pas qu'elle soit *bornée* sur la classe de tous les intervalles bornés. La situation n'est pas la même qu'à la sous-section 1.3.1 (μ finie sur $\mathcal{F}$ équivaut à μ bornée sur $\mathcal{F}$), car contrairement à la tribu $\mathcal{F}$, la classe des intervalles bornés n'a pas de plus grand élément pour l'inclusion.

Exemple 1.8. Pour illustrer simplement la proposition 1.37, considérons les deux mesures ponctuelles

$$\mu := \sum_{k\in\mathbb{N}^*} \delta_k, \quad \nu := \sum_{k\in\mathbb{N}^*} \delta_{1/k}.$$

μ est de Stieltjes car un intervalle borné ne contient qu'un ensemble fini d'entiers. Par contre ν n'est pas de Stieltjes puisque $\nu(]0,1]) = +\infty$.

1.3.9 Mesure de Lebesgue sur $\mathbb{R}^d$

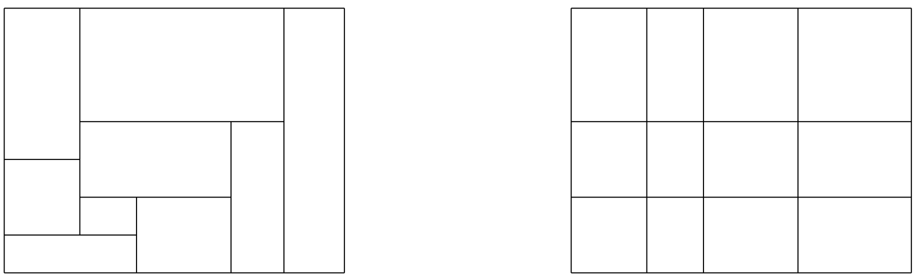
Théorème 1.38. *Il existe une unique mesure μ sur $(\mathbb{R}^d, \mathrm{Bor}(\mathbb{R}^d))$ telle que pour tout pavé $\prod_{i=1}^d]a_i, b_i]$ (avec $a_i < b_i$ pour tout $i = 1, \dots, d$),*

$$\mu\left(\prod_{i=1}^d]a_i, b_i]\right) = \prod_{i=1}^d (b_i - a_i). \tag{1.40}$$

On l'appelle mesure de Lebesgue sur $\mathbb{R}^d$ *et on la note λ_d.*

Preuve. On sait déjà que la classe $\mathcal{C}$ des pavés de la forme $\prod_{i=1}^d]a_i, b_i]$ est une semi-algèbre (exemple 1.7) qui engendre la tribu borélienne de $\mathbb{R}^d$ (prop. 1.19). Pour vérifier l'existence de la mesure de Lebesgue, on applique le théorème 1.31 à la fonction d'ensembles μ définie sur $\mathcal{C}$ par (1.40) lorsque le pavé $\prod_{i=1}^d]a_i, b_i]$ est non vide (c'est-à-dire lorsque $a_i < b_i$ pour chaque i) et par $\mu(\emptyset) = 0$ sinon. Il s'agit donc de vérifier que μ est additive sur $\mathcal{C}$ et sous-σ-additive sur $\mathcal{C}$.

Additivité de μ sur $\mathcal{C}$. Il nous faut vérifier que pour chaque décomposition d'un pavé $Q \in \mathcal{C}$ en réunion disjointe d'un nombre fini de pavés $C_j \in \mathcal{C}$, $1 \leq j \leq n$, $\mu(Q) = \sum_{j=1}^n \mu(C_j)$. Bien sûr, on suppose sans perdre de généralité qu'aucun des C_j n'est vide et dans cette égalité, $\mu(Q)$ et les $\mu(C_j)$ se calculent par la formule (1.40). Le cas facile est celui où les C_j forment ce que nous appelerons une « partition cartésienne » de Q, le cas d'une partition quelconque demandant un peu plus de travail. La figure ci-dessous illustre ces deux cas en dimension $d = 2$.



Partition quelconque Partition cartésienne

Plus précisément, nous dirons que les $\{C_j;\ 1 \leq j \leq n\}$ forment une partition cartésienne de $Q = \prod_{i=1}^d]a_i, b_i]$ si pour chaque i, il existe une subdivision

$$a_i = c_{i,0} < \cdots < c_{i,k} < \cdots < c_{i,n_i} = b_i, \tag{1.41}$$

telle que $\{C_j;\ 1 \leq j \leq n\}$ soit l'ensemble de tous les pavés

$$Q_{k_1,\dots,k_d} := \prod_{i=1}^{d}]c_{i,k_i-1}, c_{i,k_i}], \quad 1 \leq k_i \leq n_i,\ 1 \leq i \leq d. \tag{1.42}$$

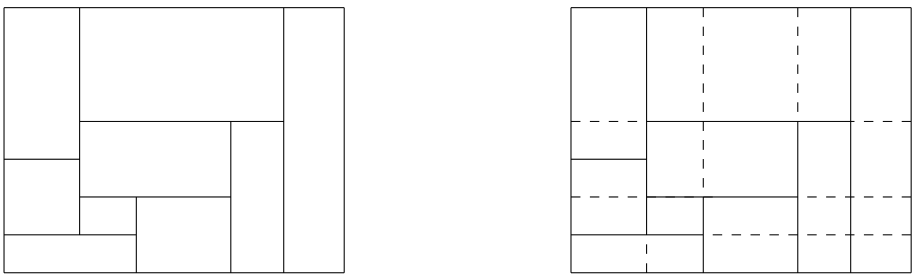
Notons $K := \{1,\dots,n_1\} \times \dots \times \{1,\dots,n_d\}$ l'ensemble de tous les d-uples $(k_1,\dots,k_d)$, et remarquons que $n = \operatorname{card} K = n_1 \dots n_d$. Dans ce cas particulier, la formule traduisant l'additivité de μ se réduit au développement d'un produit fini de sommes finies, en écrivant $b_i - a_i = \sum_{k=1}^{n_i}(c_{i,k} - c_{i,k-1})$:

$$\begin{aligned}
\mu(Q) = \prod_{i=1}^{d}\left(\sum_{k=1}^{n_i}(c_{i,k} - c_{i,k-1})\right) &= \sum_{(k_1,\dots,k_d)\in K} \prod_{i=1}^{d}(c_{i,k_i} - c_{i,k_i-1}) \\
&= \sum_{(k_1,\dots,k_d)\in K} \mu\Big(Q_{k_1,\dots,k_d}\Big) \\
&= \sum_{j=1}^{n} \mu(C_j).
\end{aligned}$$

Passons au cas d'une partition quelconque de Q par les $C_j := \prod_{i=1}^{d}]a_{i,j}, b_{i,j}]$. L'idée est de se ramener à une partition cartésienne, en définissant pour tout $i = 1,\dots,d$, une subdivision (1.41) où les $c_{i,k}$ sont tels que

$$\Big\{c_{i,k};\ 0 \leq k \leq n_i\Big\} = \bigcup_{j=1}^{n}\{a_{i,j}, b_{i,j}\}.$$

Autrement dit, on prend toutes les extrémités des segments projections des C_j sur la i-ème coordonnée et on les range par ordre croissant après avoir effacé les répétitions.



Partition quelconque — Partition cartésienne induite

On induit ainsi une partition cartésienne de Q par les pavés $Q_{k_1,\dots,k_d}$ définis par (1.42), avec $n_i \leq n + 2$. Avec K défini ci-dessus, on a donc

$$\mu(Q) = \sum_{(k_1,\dots,k_d)\in K} \mu\Big(Q_{k_1,\dots,k_d}\Big). \tag{1.43}$$

D'autre part, chacun des pavés C_j a lui aussi une partition cartésienne induite par les $Q_{k_1,\dots,k_d}$. En effet, pour tout i, $a_{i,j}$ et $b_{i,j}$ font partie par construction, de la subdivision

$\{c_{i,k};\ 0 \le k \le n_i\}$. Il existe donc deux indices $l(i,j)$ et $m(i,j)$ tels que $a_{i,j} = c_{i,l(i,j)}$ et $b_{i,j} = c_{i,m(i,j)}$. Les $\{c_{i,k};\ l(i,j) \le k \le m(i,j)\}$ forment une subdivision de $]a_{i,j}, b_{i,j}]$. En notant $K_j := \{l(1,j), \dots, m(1,j)\} \times \dots \times \{l(d,j), \dots, m(d,j)\}$, on a donc

$$\mu(C_j) = \sum_{(k_1,\dots,k_d)\in K_j} \mu\Big(Q_{k_1,\dots,k_d}\Big). \tag{1.44}$$

Pour conclure, il ne reste plus qu'à remarquer que les K_j ($(1 \le j \le n)$ forment une partition de l'ensemble de multi-indices K. Compte-tenu de (1.43) et (1.44), on en déduit

$$\sum_{j=1}^{n} \mu(C_j) = \sum_{j=1}^{n} \sum_{(k_1,\dots,k_d)\in K_j} \mu\Big(Q_{k_1,\dots,k_d}\Big) = \sum_{(k_1,\dots,k_d)\in K} \mu\Big(Q_{k_1,\dots,k_d}\Big) = \mu(Q).$$

Sous-σ-additivité de μ sur $\mathcal{C}$. La preuve est quasiment la même que celle déjà vue au théorème 1.36. On considère une décomposition en union dénombrable dans $\mathcal{C}$:

$$A = \bigcup_{j\in\mathbb{N}^*} A_j, \quad \text{où} \quad A := \prod_{i=1}^{d}]a_i, b_i], \quad A_j := \prod_{i=1}^{d}]a_{i,j}, b_{i,j}],\ j \in \mathbb{N}^*$$

et il s'agit de montrer que

$$\mu(A) \le \sum_{j=1}^{+\infty} \mu(A_j). \tag{1.45}$$

Fixons $\varepsilon > 0$. Par continuité en zéro de la fonction polynomiale $t \mapsto \prod_{i=1}^{d}(b_i - a_i - t)$, il existe un $\delta > 0$ tel que

$$\mu(B) > \mu(A) - \varepsilon, \text{ où } B := \prod_{i=1}^{d}]a_i + \delta, b_i]. \tag{1.46}$$

Par un argument similaire, on a aussi

$$\forall j \ge 1,\ \exists \delta_j > 0 \text{ tel qu'avec } B_j := \prod_{i=1}^{d}]a_{i,j}, b_{i,j} + \delta_j], \quad \mu(B_j) < \mu(A_j) + \varepsilon 2^{-j}. \tag{1.47}$$

On a alors en notant $\overline{B} := \prod_{i=1}^{d}[a_i + \delta, b_i]$ et $\mathring{B}_j := \prod_{i=1}^{d}]a_{i,j}, b_{i,j} + \delta_j[$, les inclusions

$$\overline{B} \subset A = \bigcup_{j\in\mathbb{N}^*} A_j \subset \bigcup_{j\in\mathbb{N}^*} \mathring{B}_j.$$

Du recouvrement du compact $\overline{B}$ par les ouverts $\mathring{B}_j$, on peut extraire un recouvrement fini. Il existe donc un entier n tel que

$$B \subset \overline{B} \subset \bigcup_{j=1}^{n} \mathring{B}_j \subset \bigcup_{j=1}^{n} B_j.$$

L'additivité de μ sur $\mathcal{C}$ étant déjà prouvée, le lemme 1.33 b) nous dit alors que

$$\mu(B) \le \sum_{j=1}^{n} \mu(B_j).$$

Compte-tenu de (1.46) et (1.47), on en déduit

$$\mu(A) - \varepsilon \leq \sum_{j=1}^{n} \mu(B_j) \leq \sum_{j=1}^{n} \mu(A_j) + \sum_{j=1}^{n} \varepsilon 2^{-j} \leq \sum_{j=1}^{+\infty} \mu(A_j) + \varepsilon.$$

Nous avons ainsi montré que

$$\forall \varepsilon > 0, \quad \mu(A) \leq \sum_{j=1}^{+\infty} \mu(A_j) + 2\varepsilon.$$

On en déduit (1.45), ce qui établit la sous-σ-additivité sur $\mathcal{C}$. Les hypothèses du théorème 1.31 sont maintenant vérifiées, ce qui nous assure de l'existence d'un prolongement de la fonction d'ensembles μ sur $\mathcal{C}$ en une mesure sur $\sigma(\mathcal{C}) = \mathrm{Bor}(\mathbb{R}^d)$.

Unicité de μ. Remarquons d'abord que toute mesure μ vérifiant (1.40) est σ-finie puisque :

$$\mathbb{R}^d = \bigcup_{n\in\mathbb{N}^*}]-n,n]^d \quad \text{et} \quad \forall n \in \mathbb{N}^*, \quad \mu\Big(]-n,n]^d\Big) = (2n)^d < +\infty.$$

Comme $\mathcal{C}$ est une π-classe qui engendre la tribu borélienne de $\mathbb{R}^d$, le théorème 1.34 b), appliqué avec $A_n =]-n,n]^d$ nous donne immédiatement l'unicité de μ. □

Proposition 1.39. *La mesure de Lebesgue λ_d sur $(\mathbb{R}^d, \mathrm{Bor}(\mathbb{R}^d))$ a les propriétés suivantes.*

i) *λ_d est σ-finie.*

ii) *λ_d est invariante par translations.*

iii) *λ_d est invariante par les « symétries » du type s_ε définies comme les applications linéaires transformant la base canonique $(e_1, \dots, e_d)$ en $(\varepsilon_1 e_1, \dots, \varepsilon_d e_d)$, où $\varepsilon = (\varepsilon_1, \dots, \varepsilon_d) \in \{-1, 1\}^d$.*

iv) *Si h est l'homothétie $x \mapsto cx$ dans $\mathbb{R}^d$, pour tout borélien B, $\lambda_d\big(h(B)\big) = |c|^d \lambda_d(B)$.*

v) *λ_d ne charge pas les points : $\forall x \in \mathbb{R}^d$, $\lambda_d(\{x\}) = 0$. Si $A \subset \mathbb{R}^d$ est fini ou dénombrable, $\lambda_d(A) = 0$.*

vi) $\lambda_d\left(\prod_{i=1}^{d}[a_i, b_i]\right) = \lambda_d\left(\prod_{i=1}^{d}]a_i, b_i[\right)$ *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.*

vii) *Si E est un sous-espace affine de $\mathbb{R}^d$ et $E \neq \mathbb{R}^d$, $\lambda_d(E) = 0$.*

La σ-finitude a déjà été vue ci-dessus (unicité de μ).

Preuve de ii)–iv). Notons encore $\mathcal{C}$ la classe des pavés $\prod_{i=1}^d]a_i, b_i]$. Soit g l'une des transformations translation, symétrie ou homothétie. En admettant provisoirement que $\lambda_d \circ g$ est une mesure sur $(\mathbb{R}^d, \mathrm{Bor}(\mathbb{R}^d))$, on voit immédiatement qu'elle coïncide sur $\mathcal{C}$ avec la mesure λ_d (cas ii) et iii)) ou $|c|^d \lambda_d$ (cas iv)). Le théorème d'unicité permet alors de conclure que ces mesures coïncident sur toute la tribu borélienne de $\mathbb{R}^d$ (détails laissés au lecteur). On en déduit que pour tout borélien B, $\lambda_d\big(g(B)\big) = \lambda_d(B)$ dans les cas ii) et iii) et $\lambda_d\big(g(B)\big) = |c|^d\lambda_d(B)$ dans le cas iv).

Revenons à la nature de la fonction d'ensembles $\lambda_d \circ g$. D'abord, est-elle bien définie sur $\mathrm{Bor}(\mathbb{R}^d)$? Il s'agit de vérifier que pour tout borélien B de $\mathbb{R}^d$, son image ensembliste directe $g(B)$ est encore un borélien de $\mathbb{R}^d$. L'application g étant bijective, soit $f := g^{-1}$ son inverse ponctuel. On a pour toute partie B de $\mathbb{R}^d$,

$$g(B) = \{g(y);\ y \in B\} = \{f^{-1}(y);\ y \in B\} = \{x \in \mathbb{R}^d;\ f(x) \in B\} = f^{-1}(B).$$

Par la proposition 1.11, $f^{-1}\big(\sigma(\mathcal{C})\big) = \sigma\big(f^{-1}(\mathcal{C})\big)$. On voit immédiatement que la classe des pavés $\mathcal{C}$ est globalement invariante par g : $f^{-1}(\mathcal{C}) = \mathcal{C}$. Par conséquent, $f^{-1}\big(\mathrm{Bor}(\mathbb{R}^d)\big) = \sigma(\mathcal{C}) = \mathrm{Bor}(\mathbb{R}^d)$. Ainsi pour tout borélien B, $f^{-1}(B)$ est lui-même un borélien et comme $f^{-1}(B) = g(B)$, ceci montre que la fonction d'ensembles $\lambda_d \circ g$ est bien définie sur $\mathrm{Bor}(\mathbb{R}^d)$.

Pour vérifier la σ-additivité de $\lambda_d \circ g$, soit $B = \cup_{i\in\mathbb{N}} B_i$ une réunion dénombrable disjointe de boréliens de $\mathbb{R}^d$. On a $g(\cup_{i\in\mathbb{N}} B_i) = \cup_{i\in\mathbb{N}} g(B_i)$ (c'est vrai pour une application quelconque) et les B_i étant deux à deux disjoints, les $g(B_i)$ le sont aussi, parce que g est *injective*. La σ-additivité de $\lambda_d \circ g$ se déduit alors de celle de λ_d en écrivant :

$$(\lambda_d \circ g)(B) = \lambda_d\big(g(B)\big) = \lambda_d\Big(\bigcup_{i\in\mathbb{N}} g(B_i)\Big) = \sum_{i\in\mathbb{N}} \lambda_d\big(g(B_i)\big) = \sum_{i\in\mathbb{N}} (\lambda_d \circ g)(B_i).$$

En outre, on a immédiatement $(\lambda_d \circ g)(\emptyset) = \lambda_d\big(g(\emptyset)\big) = \lambda_d(\emptyset) = 0$. Ainsi la fonction d'ensembles $\lambda_d \circ g$ est bien une mesure sur $\mathrm{Bor}(\mathbb{R}^d)$. □

Remarque 1.40. Dans ce qui précède, nous avons délibérément adopté le point de vue naïf selon lequel λ_d est invariante par une transformation f si $\lambda_d \circ f = \lambda_d$. Ceci a occasionné quelques contorsions pour vérifier que $\lambda_d \circ f$ est une mesure, la bijectivité de f jouant un rôle clé. Nous verrons au chapitre suivant que l'on peut définir la mesure image $\lambda_d \circ f^{-1}$ sous des conditions bien moins restrictives sur f (on demandera seulement que f soit mesurable). La « bonne définition » de l'invariance de λ_d par la transformation f s'écrira alors $\lambda_d \circ f^{-1} = \lambda_d$. Le lecteur pourra, à titre d'exercice, redémontrer ii) et iii) avec cette nouvelle définition.

Preuve de v) et vi). Ces deux propriétés se montrent par des arguments de continuité monotone séquentielle. Pour tout $x = (x_1, \dots, x_d) \in \mathbb{R}^d$, le singleton $\{x\}$ peut s'écrire comme l'intersection de la suite décroissante de pavés $C_n := \prod_{i=1}^d]x_i - 1/n, x_i]$. Par continuité décroissante de λ_d,

$$C_n \downarrow \{x\} \quad \Rightarrow \quad \lambda_d(\{x\}) = \lim_{n\to+\infty} \lambda_d(C_n) = \lim_{n\to+\infty} n^{-d} = 0.$$

Si A est fini ou dénombrable, par additivité finie ou σ-additivité de λ_d, on peut écrire

$$\lambda_d(A) = \lambda_d\Big(\bigcup_{x\in A} \{x\}\Big) = \sum_{x\in A} \lambda_d(\{x\}) = 0.$$

Noter que cet argument n'est plus valable si A n'est pas dénombrable : par exemple on a bien l'union disjointe $[0,1] = \bigcup_{x\in[0,1]}\{x\}$, mais $\lambda_1([0,1]) = 1 \neq 0$.

Pour prouver vi), on ne perd pas de généralité en supposant que $a_i \leq b_i$ pour tout i. Posons pour tout $n \geq 1$,

$$A_n := \prod_{i=1}^{d} \big]a_i - 1/n, b_i\big], \quad B_n := \prod_{i=1}^{d} \big]a_i, b_i - 1/n\big].$$

On voit facilement que le pavé fermé $\prod_{i=1}^{d}[a_i, b_i]$ est l'intersection de la suite décroissante $(A_n)_{n\geq 1}$ et que le pavé ouvert $\prod_{i=1}^{d}]a_i, b_i[$ est la réunion de la suite croissante $(B_n)_{n\geq 1}$. Par continuité monotone séquentielle de λ_d, on en déduit

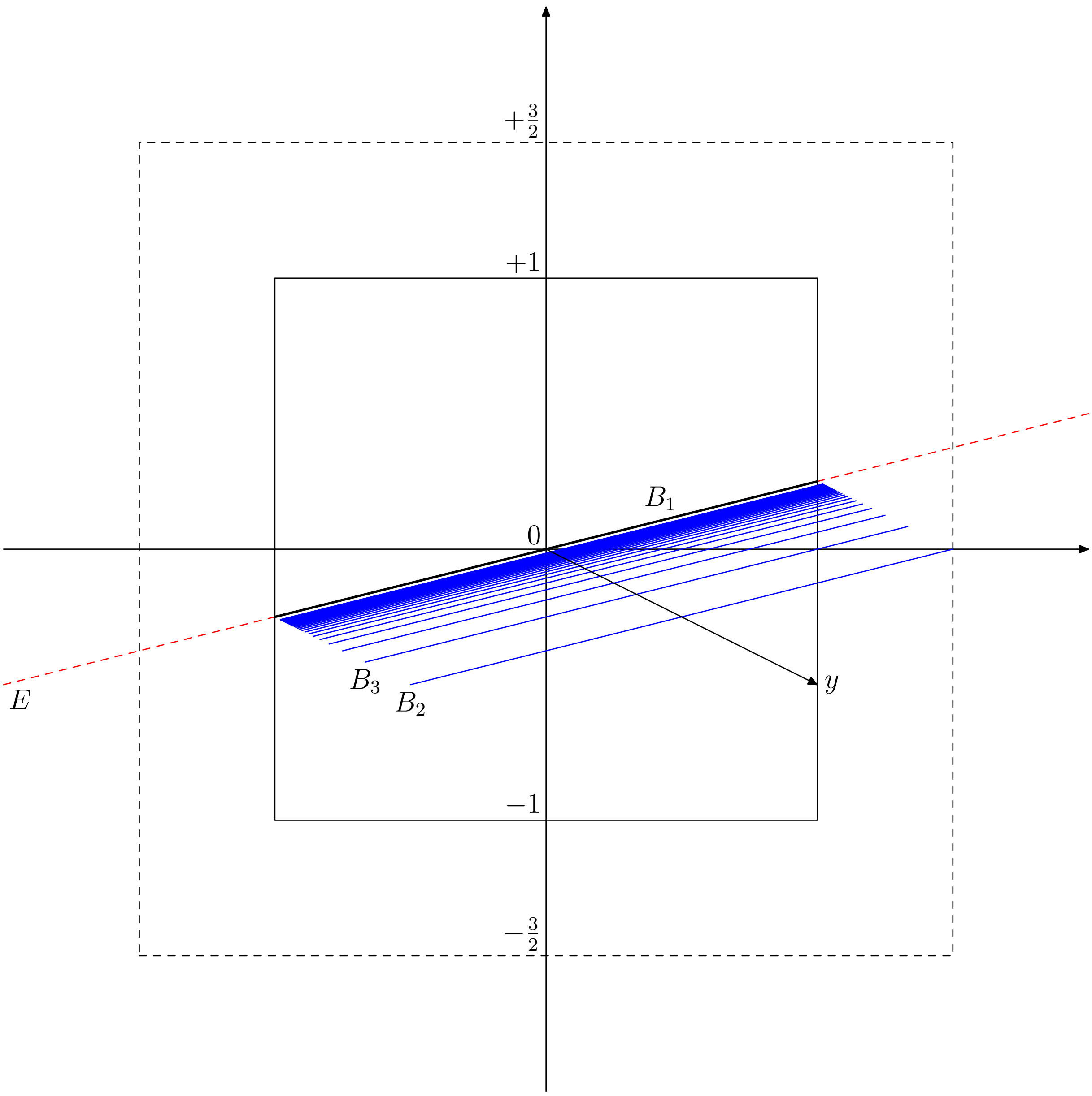
$$\lambda_d\left(\prod_{i=1}^{d}]a_i, b_i[\right) = \lim_{n\to+\infty} \lambda_d(A_n) = \lim_{n\to+\infty} \prod_{i=1}^{d}(b_i - a_i + 1/n) = \prod_{i=1}^{d}(b_i - a_i)$$

et

$$\lambda_d\left(\prod_{i=1}^{d}[a_i, b_i]\right) = \lim_{n\to+\infty} \lambda_d(B_n) = \lim_{n\to+\infty} \prod_{i=1}^{d}(b_i - a_i - 1/n)^+ = \prod_{i=1}^{d}(b_i - a_i),$$

d'où la conclusion. On a utilisé la notation $(b_i - a_i - 1/n)^+$ pour prendre en compte l'éventualité $a_i \geq b_i - 1/n$, auquel cas B_n est l'ensemble vide, de mesure nulle. $\square$

Preuve de vii). L'idée est de regarder la trace B_1 de E sur le cube unité et d'en accumuler dans un volume *fini*, une infinité de translatés parallèles, donc disjoints. Comme tous ces translatés ont même mesure, celle-ci ne peut être que nulle et on en déduit $\lambda_d(E) = 0$ en utilisant les homothétiques de B_1. La figure suivante, dessinée dans le cas $d = 2$, illustre cette idée.



Voici les détails de la preuve en dimension d quelconque. En raison de l'invariance de λ_d par translation, on peut se ramener au cas où E est un sous-espace vectoriel de $\mathbb{R}^d$ (i.e. $0 \in E$). Par hypothèse, il existe $y \in \mathbb{R}^d$ tel que $y \notin E$ et comme $0 \in E$, y n'est pas nul. Comme E est un espace vectoriel, aucun vecteur cy, $c \in \mathbb{R}^*$ n'est dans E. Quitte à remplacer y par $\frac{1}{\|y\|_\infty}y$, on peut supposer que $\|y\|_\infty = 1$ (on note $\|x\|_\infty := \max_{1\leq i\leq d} |x_i|$). Pour tout c réel, notons $E + cy$ le sous-espace affine, image de E par la translation de vecteur cy : $E+cy := \{x+cy; x \in E\}$. Pour $c \neq c'$, les sous-espaces affines parallèles $E+cy$ et $E + c'y$ sont disjoints ; sinon ils auraient un élément commun z qui pourrait s'écrire $z = x + cy = x' + c'y$ pour des vecteurs x et x' de E et on en déduirait $y = \frac{1}{c-c'}(x - x')$, donc $y \in E$, ce qui est exclu. Définissons maintenant :

$$B_1 := E \cap [-1, 1]^d, \quad B_k := B_1 + \frac{1}{k}y, \quad (k \geq 2).$$

Comme $\|y\|_\infty = 1$, on a $\bigcup_{k\geq 1} B_k \subset [-3/2, 3/2]^d$, donc

$$\lambda_d\Big(\bigcup_{k\geq 1} B_k\Big) \leq 3^d < +\infty. \tag{1.48}$$

Les B_k sont deux à deux disjoints car chaque B_k est inclus dans $E + \frac{1}{k}y$. En utilisant la σ-additivité de λ_d et son invariance par translation, on voit que pour tout entier $n \geq 1$,

$$\lambda_d\Big(\bigcup_{k\geq 1} B_k\Big) = \sum_{k=1}^{+\infty} \lambda_d(B_k) \geq \sum_{k=1}^{n} \lambda_d(B_k) = n\lambda_d(B_1).$$

Compte-tenu de (1.48), on a ainsi $n\lambda_d(B_1) \leq 3^d$ pour tout $n \geq 1$ et ceci impose $\lambda_d(B_1) = 0$. Soit maintenant h_n l'homothétie $x \mapsto nx$. Par la propriété iv), $\lambda_d\big(h_n(B_1)\big) = n^d\lambda_d(B_1) = 0$. D'autre part il est clair que $h_n(B_1) = E \cap [-n, n]^d$. On en déduit que $\lambda_d(E) = 0$ par continuité croissante séquentielle puisque $E \cap [-n, n]^d \uparrow E$. $\square$

Chapitre 2

Applications mesurables

2.1 Topologie et tribus boréliennes de $\overline{\mathbb{R}}$ et $\overline{\mathbb{R}}_+$

Dans la théorie de l'intégration de Lebesgue, il est très commode de travailler avec des fonctions à valeurs dans la droite achevée $\overline{\mathbb{R}}$, réunion de $\mathbb{R}$ et des points à l'infini $-\infty$ et $+\infty$. Nous ferons aussi un usage intensif de la demi droite achevée $\overline{\mathbb{R}}_+ := \mathbb{R}_+ \cup \{+\infty\}$.

Pour munir $\overline{\mathbb{R}}$ d'une topologie [1] compatible avec celle de $\mathbb{R}$, on utilise une application strictement croissante continue f de $]-1,+1[$ sur $\mathbb{R}$, par exemple $t \mapsto \tan(\pi t/2)$. L'application f est une bijection *bicontinue* de $]-1,+1[$ sur $\mathbb{R}$ (un homéomorphisme). On la prolonge en une bijection $\tilde{f}$ de $[-1,+1]$ sur $\overline{\mathbb{R}}$ en posant $\tilde{f}(x) := f(x)$ pour $x \in]-1,+1[$, $\tilde{f}(-1) := -\infty$ et $\tilde{f}(+1) := +\infty$. En transportant par $\tilde{f}$ la topologie de $[-1,+1]$ sur $\overline{\mathbb{R}}$, on fait de $\tilde{f}$ un homéomorphisme. Ainsi les intervalles $[-1,a[$, $(-1 < a < +1)$ forment un système fondamental de voisinages de -1 dans $[-1,+1]$. Leurs images par $\tilde{f}$ nous donnent comme système fondamental de voisinages de $-\infty$ dans $\overline{\mathbb{R}}$, les intervalles $[-\infty,b[$, $(b \in \mathbb{R})$. De même un système fondamental de voisinages de $+\infty$ est formé des $]b,+\infty]$. Comme $]-1,+1[$ est un ouvert de $[-1,+1]$, son image $\mathbb{R}$ par $\tilde{f}$ est un ouvert de $\overline{\mathbb{R}}$. On utilise aussi $\tilde{f}$ pour transporter la relation d'ordre dans $[-1,+1]$ à $\overline{\mathbb{R}}$ de façon à ce que $\tilde{f}$ soit strictement croissante.

Dans $\overline{\mathbb{R}}$, toute suite monotone a une limite. Toute série à termes positifs converge dans $\overline{\mathbb{R}}_+$. Tout sous-ensemble E de $\overline{\mathbb{R}}$ possède une borne inférieure $\inf E$ et une borne supérieure $\sup E$ dans $\overline{\mathbb{R}}$. Cependant on réservera l'appellation *borné* à un ensemble E pour lequel $-\infty < \inf E$ et $\sup E < +\infty$. De même on dira qu'une fonction $g : \Omega \to \overline{\mathbb{R}}$ est bornée si $E = g(\Omega)$ est une partie bornée de $\overline{\mathbb{R}}$ au sens précédent.

Reprenons la construction de la topologie de $\overline{\mathbb{R}}$ avec une approche métrique. On fixe $\tilde{f}$ ayant les propriétés ci-dessus, disons $f(t) = \tan(\pi t/2)$ pour tout $t \in \mathbb{R}$, $\tilde{f}(-1) = -\infty$ et $\tilde{f}(1) = +\infty$. On munit alors $\overline{\mathbb{R}}$ de la distance

$$d(x,y) := |\tilde{f}^{-1}(x) - \tilde{f}^{-1}(y)| = \frac{2}{\pi}|\arctan x - \arctan y| = \left| \int_x^y \frac{2\,\mathrm{d}u}{\pi(1+u^2)} \right|. \tag{2.1}$$

1. Pour les lecteurs ne connaissant que les espaces métriques, nous définirons aussi cette topologie à partir d'une distance.

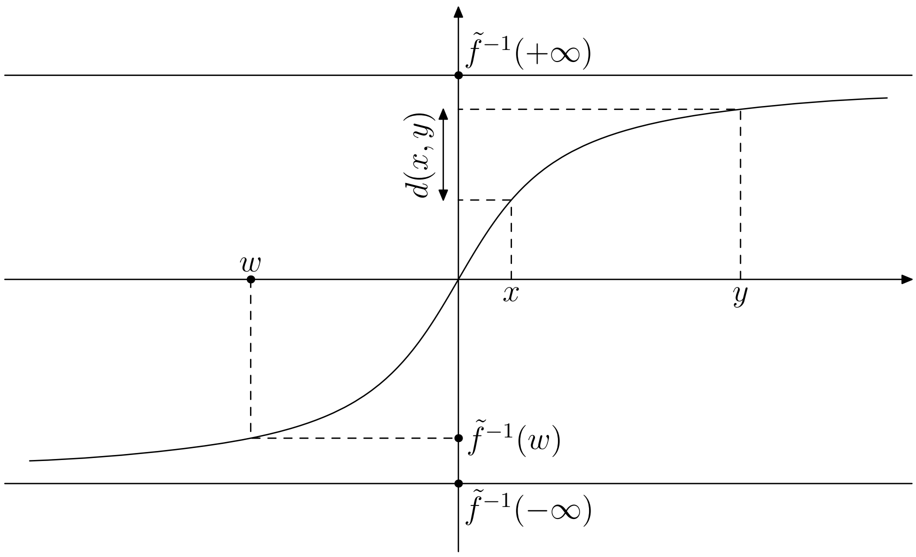


FIGURE 2.1 – La distance d sur $\overline{\mathbb{R}}$

Remarquons que pour cette distance, $d(-\infty,+\infty) = 2$. Notons au passage que dans l'espace $(\overline{\mathbb{R}}, d)$, $\mathbb{R}$ est la boule ouverte de centre 0 et de rayon 1, c'est donc un ouvert de $\overline{\mathbb{R}}$. Voyons de plus près la relation entre les boules ouvertes de $[-1,1]$ et celles de $(\overline{\mathbb{R}}, d)$. Notons $\Delta(t_0, r)$ la boule ouverte de centre t_0 et de rayon r dans l'espace métrique $([-1,1], \delta)$ où δ est la métrique usuelle $\delta(s,t) := |s-t|$. Il est clair que $\Delta(t_0, r) =]t_0 - r, t_0 + r[\cap[-1,1]$. Soit $B(c,r)$ la boule de centre $c \in \overline{\mathbb{R}}$ et de rayon r dans l'espace métrique $(\overline{\mathbb{R}}, d)$. Vu la définition de d, cette boule peut s'écrire en posant $t_0 := \tilde{f}^{-1}(c)$,

$$\begin{aligned} B(c,r) &= \Big\{x \in \overline{\mathbb{R}};\ |\tilde{f}^{-1}(x) - \tilde{f}^{-1}(c)| < r\Big\} \\ &= \Big\{x \in \overline{\mathbb{R}};\ t_0 - r < \tilde{f}^{-1}(x) < t_0 + r\Big\} \\ &= \Big\{x \in \overline{\mathbb{R}};\ \exists t \in [-1,1], x = \tilde{f}(t), t_0 - r < t < t_0 + r\Big\} \\ &= \tilde{f}\Big(]t_0 - r, t_0 + r[\cap[-1,1]\Big) \\ &= \tilde{f}\Big(\Delta(t_0, r)\Big). \end{aligned}$$

Ainsi les boules ouvertes de $(\overline{\mathbb{R}}, d)$ sont exactement les images par $\tilde{f}$ des boules ouvertes de $([-1,1], \delta)$. En voici la liste exhaustive, classée en trois types, les droites, les demi-droites et les segments ouverts.

- droites : $[-\infty,+\infty]$, $]-\infty,+\infty]$, $[-\infty,+\infty[$, $]-\infty,+\infty[$;
- demi-droites : $]-\infty,a[$, $[-\infty,a[$, $]a,+\infty[$, $]a,+\infty]$, a étant un réel quelconque ;
- segments ouverts : $]a,b[$ (a, b réels quelconques, avec $a < b$). L'intervalle $]a,b[$ est la boule ouverte de centre c tel que $2\arctan c = \arctan a + \arctan b$ et de rayon $r = \frac{1}{\pi}(\arctan b - \arctan a)$.

Par restriction à $\mathbb{R}$, on voit ainsi que dans l'espace métrique $(\mathbb{R}, d)$, la famille des boules ouvertes est constituée des segments ouverts et des intervalles de la forme $]-\infty, a[$, $]a,+\infty[$, $a \in \mathbb{R}$ et de $\mathbb{R} =]-\infty,+\infty[$. Les boules ouvertes de $(\mathbb{R}, d)$ qui ne sont pas des segments ouverts sont clairement réunions de segments ouverts. Les ouverts de $(\mathbb{R}, d)$ qui sont par définition, les réunions de boules ouvertes sont donc aussi les réunions de segments ouverts. Or la famille des segments ouverts est exactement la famille des boules ouvertes de l'espace métrique $(\mathbb{R}, \delta)$ où δ est la métrique usuelle $\delta(x, y) := |x - y|$. Ainsi les deux métriques d et δ génèrent la même topologie sur $\mathbb{R}$ (i.e. ont les mêmes ouverts). Elles sont donc équivalentes.

Remarque 2.1.

i) Tout ouvert de $\mathbb{R}$ est aussi un ouvert de $\overline{\mathbb{R}}$.

ii) Si W est un ouvert de $\overline{\mathbb{R}}$, $W \cap \mathbb{R}$ est ouvert de $\mathbb{R}$.

Justification. Si V est ouvert de $\mathbb{R}$, il est union de segments ouverts qui sont aussi des boules ouvertes de $\overline{\mathbb{R}}$, donc V est ouvert de $\overline{\mathbb{R}}$. Soit W un ouvert de $\overline{\mathbb{R}}$. Il s'écrit $\cup_{i \in I} B_i$ où les B_i sont des boules ouvertes de $\overline{\mathbb{R}}$ et $W \cap \mathbb{R} = \cup_{i \in I}(B_i \cap \mathbb{R})$. Si certaines des B_i contiennent $-\infty$ ou $+\infty$, l'intersection avec $\mathbb{R}$ les transforme en intervalles ouverts de $\mathbb{R}$ (exemple : $[-\infty, a[\cap \mathbb{R} =]-\infty, a[$) qui sont eux-mêmes unions de *segments* ouverts. Ainsi $W \cap \mathbb{R}$ est union de segments ouverts de $\mathbb{R}$, donc ouvert de $\mathbb{R}$. □

Ayant maintenant défini une topologie sur $\overline{\mathbb{R}}$, on peut le munir de la tribu borélienne correspondante. Il est alors naturel de se demander si les boréliens de $\overline{\mathbb{R}}$ diffèrent beaucoup de ceux de $\mathbb{R}$. Notons auparavant que $\mathrm{Bor}(\mathbb{R})$ n'est pas une tribu sur $\overline{\mathbb{R}}$ (puisque $\overline{\mathbb{R}} \notin \mathrm{Bor}(\mathbb{R})$) et que $\mathrm{Bor}(\overline{\mathbb{R}})$ n'est pas davantage une tribu sur $\mathbb{R}$ puisque $\overline{\mathbb{R}} \in \mathrm{Bor}(\overline{\mathbb{R}})$ et $\overline{\mathbb{R}} \not\subset \mathbb{R}$.

Proposition 2.2.

a) Si B est un borélien de $\overline{\mathbb{R}}$, $B \cap \mathbb{R}$ est un borélien de $\mathbb{R}$.

b) Tout borélien de $\mathbb{R}$ est aussi un borélien de $\overline{\mathbb{R}}$.

c) B est un borélien de $\overline{\mathbb{R}}$ si et seulement si $B = A \cup C$ où A est un borélien de $\mathbb{R}$ et C est l'un des ensembles $\emptyset$, $\{-\infty\}$, $\{+\infty\}$, $\{-\infty, +\infty\}$.

Preuve du a). Considérons la famille $\mathcal{F}$ de parties de $\overline{\mathbb{R}}$ définie par :

$$\mathcal{F} := \Big\{E \in \mathcal{P}(\overline{\mathbb{R}});\ E \cap \mathbb{R} \in \mathrm{Bor}(\mathbb{R})\Big\}.$$

Il est immédiat de vérifier que $\mathcal{F}$ est une tribu sur $\overline{\mathbb{R}}$. Grâce à la remarque 2.1 ii), $\mathcal{F}$ possède tous les ouverts de $\overline{\mathbb{R}}$. Par minimalité, elle contient donc la tribu engendrée par ces ouverts, c'est-à-dire $\mathrm{Bor}(\overline{\mathbb{R}})$. L'inclusion $\mathrm{Bor}(\overline{\mathbb{R}}) \subset \mathcal{F}$ n'est qu'une autre façon d'écrire a) qui est ainsi établie. □

Preuve du b). La famille $\mathcal{G}$ de parties de $\mathbb{R}$ définie par :

$$\mathcal{G} := \Big\{A \in \mathrm{Bor}(\mathbb{R});\ A \in \mathrm{Bor}(\overline{\mathbb{R}})\Big\} = \mathrm{Bor}(\mathbb{R}) \cap \mathrm{Bor}(\overline{\mathbb{R}})$$

est une tribu[2] sur $\mathbb{R}$. En effet il est clair que $\mathcal{G}$ possède l'ensemble vide et est stable par union dénombrable. Pour la stabilité par complémentaire, on observe que si $A \in \mathcal{G}$,

2. Attention, c'est l'intersection de deux tribus sur des ensembles Ω_1 et Ω_2 différents, on ne peut donc pas appliquer ici la proposition 1.5.

alors $\overline{\mathbb{R}} \setminus A \in \mathrm{Bor}(\overline{\mathbb{R}})$, puisque $A \in \mathrm{Bor}(\overline{\mathbb{R}})$; en écrivant $\mathbb{R} \setminus A = (\overline{\mathbb{R}} \setminus A) \cap \mathbb{R}$, on voit que $\mathbb{R} \setminus A$ est l'intersection de deux boréliens de $\overline{\mathbb{R}}$, donc appartient aussi à $\mathrm{Bor}(\overline{\mathbb{R}})$ et $\mathbb{R} \setminus A$ est bien élément de $\mathcal{G}$. Par construction, $\mathcal{G}$ est une sous tribu de $\mathrm{Bor}(\mathbb{R})$. D'autre part elle contient les ouverts de $\mathbb{R}$ (remarque 2.1 i)), donc par minimalité la tribu qu'ils engendrent, c'est à dire $\mathrm{Bor}(\mathbb{R})$. Finalement $\mathcal{G} = \mathrm{Bor}(\mathbb{R})$ et $\mathrm{Bor}(\mathbb{R}) \subset \mathrm{Bor}(\overline{\mathbb{R}})$, ce qui établit b). Attention $\mathrm{Bor}(\mathbb{R})$ n'est pas une *sous-tribu* de $\mathrm{Bor}(\overline{\mathbb{R}})$ car elles ne sont pas construites sur le même ensemble Ω. □

Preuve du c). Soit $B \in \mathrm{Bor}(\overline{\mathbb{R}})$, on peut l'écrire $B = (B \cap \mathbb{R}) \cup C$ où C est inclus dans $\{-\infty, +\infty\}$. Par le a), $A := B \cap \mathbb{R}$ est un borélien de $\mathbb{R}$.

Pour la réciproque, notons d'abord que dans $(\overline{\mathbb{R}}, d)$, les singletons sont des fermés (c'est vrai dans n'importe quel espace métrique) donc des boréliens. Par conséquent, chacun des quatre sous-ensembles de $\{-\infty, +\infty\}$ est un borélien de $\overline{\mathbb{R}}$. Cette remarque et le b) nous montrent que si A est un borélien de $\mathbb{R}$ et C un sous-ensemble de $\{-\infty, +\infty\}$, $A \cup C$ est un borélien de $\overline{\mathbb{R}}$. □

Passons maintenant à la topologie de $\overline{\mathbb{R}}_+$. On l'obtient naturellement comme la topologie trace de $\overline{\mathbb{R}}$, ou pour rester dans un cadre métrique, comme celle de l'espace métrique $(\overline{\mathbb{R}}_+, d)$, d étant la restriction à $\overline{\mathbb{R}}_+$ de la métrique définie en (2.1). La restriction à $[0, 1]$ de $\tilde{f}$ est alors un homéomorphisme de cet intervalle sur $\overline{\mathbb{R}}_+$. Les boules ouvertes de $(\overline{\mathbb{R}}_+, d)$ sont les images par $\tilde{f}$ de celles de $([0, 1], \delta)$. On peut alors reprendre tout le travail que nous venons de faire pour $\overline{\mathbb{R}}$. Le lecteur pourra établir en exercice les analogues de la remarque 2.1 et de la proposition 2.2 a) et b) obtenus en remplaçant $\mathbb{R}$ par $\mathbb{R}_+$ et $\overline{\mathbb{R}}$ par $\overline{\mathbb{R}}_+$ et en déduire le résultat suivant.

Proposition 2.3. *B est un borélien de $\overline{\mathbb{R}}_+$ si et seulement si B est un borélien de $\mathbb{R}_+$ ou s'écrit $B = A \cup \{+\infty\}$, A étant un borélien de $\mathbb{R}_+$.*

Proposition 2.4.

a) La tribu borélienne de $\overline{\mathbb{R}}$ est engendrée par chacune des familles

$$\mathcal{I}_1 := \{\, [-\infty, x[;\ x \in \mathbb{R}\}, \quad \mathcal{I}_2 := \{\, [-\infty, x];\ x \in \mathbb{R}\}.$$

b) La tribu borélienne de $\overline{\mathbb{R}}_+$ est engendrée par chacune des familles

$$\mathcal{J}_1 := \{\, [0, x[;\ x \in \mathbb{R}_+\}, \quad \mathcal{J}_2 := \{\, [0, x];\ x \in \mathbb{R}_+\}.$$

Preuve. Montrons l'égalité $\sigma(\mathcal{I}_1) = \mathrm{Bor}(\overline{\mathbb{R}})$ en établissant l'inclusion dans les deux sens. Soit V un ouvert de $\overline{\mathbb{R}}$. On peut l'écrire $V = (V \cap \mathbb{R}) \cup (V \cap \{-\infty, +\infty\})$. En notant que $\{-\infty\} = \bigcap_{n \in \mathbb{N}} [-\infty, -n[$ et $\{+\infty\} = \bigcap_{n \in \mathbb{N}}]n, +\infty]$, on voit que les ensembles $\{-\infty\}$, $\{+\infty\}$ et $]-\infty, x[$ appartiennent à la tribu $\sigma(\mathcal{I}_1)$. Par la remarque 2.1 ii), $V \cap \mathbb{R}$ appartient à $\mathrm{Bor}(\mathbb{R})$, laquelle est engendrée par les $]-\infty, x[$ donc contenue dans $\sigma(\mathcal{I}_1)$. On en déduit l'appartenance de V à $\sigma(\mathcal{I}_1)$, puis par minimalité, l'inclusion $\mathrm{Bor}(\overline{\mathbb{R}}) \subset \sigma(\mathcal{I}_1)$. L'inclusion dans l'autre sens résulte immédiatement de ce que les intervalles $[-\infty, x[$ sont dans $\overline{\mathbb{R}}$ des boules ouvertes de centre $-\infty$, donc des ouverts.

La justification des égalités $\sigma(\mathcal{I}_2) = \mathrm{Bor}(\overline{\mathbb{R}})$, $\sigma(\mathcal{J}_1) = \mathrm{Bor}(\overline{\mathbb{R}}_+)$ et $\sigma(\mathcal{J}_2) = \mathrm{Bor}(\overline{\mathbb{R}}_+)$ est analogue et laissée en exercice au lecteur. □

2.2 Arithmétique dans $\overline{\mathbb{R}}_+$

La nécessité d'intégrer des fonctions sur des ensembles de mesure infinie ou d'intégrer des fonctions pouvant naturellement prendre une valeur infinie en certains points (par exemple une limite supérieure d'une suite de fonctions ou la somme d'une série à termes positifs) nous amène à *prolonger* à $\overline{\mathbb{R}}_+$ l'addition et la multiplication usuelles en adoptant les conventions suivantes.

$$a + (+\infty) = (+\infty) + a = +\infty, \quad \forall a \in \overline{\mathbb{R}}_+. \tag{2.2}$$

$$a \times (+\infty) = (+\infty) \times a = \begin{cases} +\infty & \text{si} \quad 0 < a \leq +\infty, \\ 0 & \text{si} \quad a = 0. \end{cases} \tag{2.3}$$

L'addition et la multiplication ainsi prolongées restent commutatives et associatives. La multiplication reste distributive par rapport à l'addition. Par contre il convient d'être prudent chaque fois qu'apparaissent une soustraction ou une division. En particulier les règles de simplification habituelles ne s'étendent pas : $a+b = a+c$ n'implique $b = c$ que si a est fini et $ab = ac$ n'implique $b = c$ que si $0 < a < +\infty$.

Au premier abord, la convention $0 \times (+\infty) = 0$ dégage une odeur sulfureuse car elle paraît contradictoire avec la notion de *forme indéterminée du type* « $0 \times (+\infty)$ ». Cette contradiction n'est qu'apparente. La multiplication M, définie sur $\mathbb{R}_+^2$ par $M(a, b) = ab$, a été prolongée par (2.3) à $\overline{\mathbb{R}}_+^2$ en conservant de bonnes propriétés algébriques, mais on ne prétend pas que ce prolongement de M soit *continu* aux points $(0, +\infty)$ et $(+\infty, 0)$. La notion de *forme indéterminée du type* « $0 \times (+\infty)$ » nous dit précisément qu'il est *impossible de prolonger M en ces points en préservant sa continuité* : si (x_n) et (y_n) sont deux suites dans $\mathbb{R}_+$ convergentes l'une vers 0 et l'autre vers $+\infty$ (donc (x_n, y_n) converge vers $(0, +\infty)$ dans $\overline{\mathbb{R}}_+^2$), la suite $M(x_n, y_n) = x_n y_n$ peut selon les cas converger vers n'importe quel élément de $\overline{\mathbb{R}}_+$ ou même n'avoir aucune limite.

2.3 Mesurabilité

Définition 2.5. *Soient $(\Omega_1, \mathcal{F}_1)$ et $(\Omega_2, \mathcal{F}_2)$ deux espaces mesurables. L'application $f : \Omega_1 \to \Omega_2$ est dite $\mathcal{F}_1$-$\mathcal{F}_2$ mesurable si pour tout $B \in \mathcal{F}_2$, $f^{-1}(B) \in \mathcal{F}_1$, autrement dit si $f^{-1}(\mathcal{F}_2) \subset \mathcal{F}_1$.*

Remarque 2.6. Il est clair d'après la définition de la mesurabilité d'une application, que celle-ci est conservée chaque fois que l'on *diminue* (au sens de l'inclusion) la tribu de l'espace d'arrivée ou que l'on *agrandit* celle de l'espace de départ.

S'il n'y a pas d'ambiguïté sur les tribus concernées, on pourra se contenter de dire « mesurable » au lieu de « $\mathcal{F}_1$-$\mathcal{F}_2$ mesurable ». Les applications mesurables jouent un rôle central dans la théorie de l'intégration abstraite. Elles permettent de « transporter » la mesure d'un espace vers un autre. Quand l'espace d'arrivée est $\Omega_2 = \overline{\mathbb{R}}$ avec $\mathcal{F}_2 = \mathrm{Bor}(\overline{\mathbb{R}})$, ce sont elles qui ont vocation à être intégrées. Dans le langage de la théorie des probabilités, les applications mesurables correspondent à la notion de variable aléatoire.

Définition 2.7. *Soit* $(\Omega, \mathcal{F})$ *un espace probabilisable. On appelle* variable aléatoire réelle *sur* $(\Omega, \mathcal{F})$ *toute application* $X : \Omega \to \mathbb{R}$*,* $\mathcal{F}$-$\mathrm{Bor}(\mathbb{R})$ *mesurable. De même on appelle* variable aléatoire complexe *sur* $(\Omega, \mathcal{F})$ *toute application* $X : \Omega \to \mathbb{C}$*,* $\mathcal{F}$-$\mathrm{Bor}(\mathbb{C})$ *mesurable et* vecteur aléatoire*, toute application* $X : \Omega \to \mathbb{R}^d$*,* $\mathcal{F}$-$\mathrm{Bor}(\mathbb{R}^d)$ *mesurable* $(d > 1)$*.*

Rappelons que du point de vue mathématique, un espace probabilisable n'est rien d'autre qu'un espace mesurable, c'est-à-dire un ensemble Ω muni d'une tribu de parties de Ω. L'emploi de l'adjectif « probabilisable » au lieu de mesurable indique seulement l'intention que l'on a de munir cet espace d'une (ou de plusieurs !) mesure de probabilité $\mathbf{P}$.

Définition 2.8. *Soit* $(\Omega, \mathcal{F})$ *un espace probabilisable. On appelle* variable aléatoire discrète *(réelle, resp. complexe) sur* $(\Omega, \mathcal{F})$ *toute application* $X : \Omega \to \mathbb{K}$ *(*$\mathbb{K} = \mathbb{R}$*, resp.* $\mathbb{C}$*), telle que* $X(\Omega)$ *est au plus dénombrable et* X *est* $\mathcal{F}$-$\mathcal{P}(\mathbb{K})$ *mesurable. De même on appelle* vecteur aléatoire discret*, toute application* $X : \Omega \to \mathbb{R}^d$*, telle que* $X(\Omega)$ *est au plus dénombrable et* X *est* $\mathcal{F}$-$\mathcal{P}(\mathbb{R}^d)$ *mesurable* $(d > 1)$*.*

En vertu de la remarque 2.6, une variable aléatoire *discrète* réelle (resp. complexe) est aussi une variable aléatoire réelle (resp. complexe) et un vecteur aléatoire *discret* est aussi un vecteur aléatoire, au sens de la définition 2.7.

Un exemple simple, mais important, d'application mesurable de Ω dans $\mathbb{R}$ (et de variable aléatoire discrète) est l'indicatrice $\mathbf{1}_A$ d'un élément A de la tribu $\mathcal{F}$:

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

En effet pour tout $B \subset \mathbb{R}$, on a

$$(\mathbf{1}_A)^{-1}(B) = \begin{cases} \emptyset & \text{si} \quad 0 \notin B \text{ et } 1 \notin B, \\ A & \text{si} \quad 0 \notin B \text{ et } 1 \in B, \\ A^c & \text{si} \quad 0 \in B \text{ et } 1 \notin B, \\ \Omega & \text{si} \quad 0 \in B \text{ et } 1 \in B. \end{cases}$$

On vient de vérifier que $(\mathbf{1}_A)^{-1}\big(\mathcal{P}(\mathbb{R})\big) = \sigma(\{A\})$. L'application $\mathbf{1}_A$ est donc $\sigma(\{A\})$-$\mathcal{P}(\mathbb{R})$ mesurable. Elle est donc aussi $\mathcal{F}$-$\mathcal{B}$ mesurable pour toute tribu $\mathcal{F}$ possédant A et toute tribu $\mathcal{B}$ incluse dans $\mathcal{P}(\mathbb{R})$ d'après la remarque 2.6. De même, le lecteur vérifiera aisément qu'une fonction constante $\Omega \to \mathbb{R}$ est mesurable pour toute tribu sur Ω et toute tribu sur $\mathbb{R}$.

En dehors de ces exemples élémentaires, il est généralement assez difficile, voire parfois impossible, de tester directement l'appartenance de $f^{-1}(B)$ à $\mathcal{F}_1$ pour *tout* $B \in \mathcal{F}_2$, en raison notamment de l'absence de description exhaustive des éléments de $\mathcal{F}_2$ quand cette tribu est assez riche (cas de la tribu $\mathrm{Bor}(\mathbb{R})$). La proposition suivante nous montre qu'en fait, il suffit de restreindre ce test aux éléments B d'une famille génératrice de $\mathcal{F}_2$.

Proposition 2.9. *Soient* $(\Omega_1, \mathcal{F}_1)$ *et* $(\Omega_2, \mathcal{F}_2)$ *deux espaces mesurables et* $\mathcal{S}$ *une famille de parties de* Ω_2 *engendrant* $\mathcal{F}_2$ *(*$\sigma(\mathcal{S}) = \mathcal{F}_2$*). L'application* $f : \Omega_1 \to \Omega_2$ *est* $\mathcal{F}_1$-$\mathcal{F}_2$ *mesurable si pour tout* $B \in \mathcal{S}$*,* $f^{-1}(B) \in \mathcal{F}_1$*, autrement dit si* $f^{-1}(\mathcal{S}) \subset \mathcal{F}_1$*.*

Preuve. Comme $\mathcal{F}_1$ est une tribu sur Ω_1, l'hypothèse $f^{-1}(\mathcal{S}) \subset \mathcal{F}_1$ implique par minimalité l'inclusion de tribus $\sigma\big(f^{-1}(\mathcal{S})\big) \subset \mathcal{F}_1$. Or nous savons par la Proposition 1.11) que $\sigma\big(f^{-1}(\mathcal{S})\big) = f^{-1}\big(\sigma(\mathcal{S})\big)$. Comme $\sigma(\mathcal{S}) = \mathcal{F}_2$, nous venons ainsi de vérifier l'inclusion $f^{-1}(\mathcal{F}_2) \subset \mathcal{F}_1$, autrement dit la mesurabilité de f. □

Voici une liste (non exhaustive) d'applications courantes de la proposition 2.9.

Corollaire 2.10. *Soient Ω un ensemble muni d'une tribu $\mathcal{F}$ et une application $f : \Omega \to \mathbb{K}$ où $\mathbb{K}$ désigne l'un des ensembles $\mathbb{R}$, $\overline{\mathbb{R}}$, $\mathbb{R}_+$, $\overline{\mathbb{R}}_+$, ou $\mathbb{R}^d$. Notons $\mathcal{B} := \mathrm{Bor}(\mathbb{K})$.*

i) Pour $\mathbb{K} = \mathbb{R}$, f est $\mathcal{F}$-$\mathcal{B}$ mesurable si pour tous $a, b \in \mathbb{R}$, $f^{-1}(]a, b[) \in \mathcal{F}$.

ii) Pour $\mathbb{K} = \mathbb{R}$, f est $\mathcal{F}$-$\mathcal{B}$ mesurable si pour tout $x \in \mathbb{R}$, $f^{-1}(]-\infty, x]) \in \mathcal{F}$.

iii) Pour $\mathbb{K} = \overline{\mathbb{R}}$, f est $\mathcal{F}$-$\mathcal{B}$ mesurable si pour tout $x \in \mathbb{R}$, $f^{-1}([-\infty, x]) \in \mathcal{F}$.

iv) Pour $\mathbb{K} = \mathbb{R}_+$ ou $\overline{\mathbb{R}}_+$, f est $\mathcal{F}$-$\mathcal{B}$ mesurable si pour tout $x \in \mathbb{R}_+$, $f^{-1}([0, x]) \in \mathcal{F}$.

v) Pour $\mathbb{K} = \mathbb{R}^d$, f est $\mathcal{F}$-$\mathcal{B}$ mesurable si pour tout pavé $C := \prod_{i=1}^d]a_i, b_i]$, $f^{-1}(C) \in \mathcal{F}$.

Corollaire 2.11. *Soit $(\Omega, \mathcal{F})$ un espace probabilisable et X une application $\Omega \to \mathbb{K}$ (où $\mathbb{K} = \mathbb{R}$, $\mathbb{C}$ ou $\mathbb{R}^d$, $d > 1$) telle que $X(\Omega)$ soit une partie* au plus dénombrable *de $\mathbb{K}$. Alors X est une variable aléatoire* discrète *(resp. un vecteur aléatoire discret) si :*

$$\forall x \in X(\Omega), \quad X^{-1}(\{x\}) \in \mathcal{F}. \tag{2.4}$$

Comme la tribu $\mathcal{P}(\mathbb{K})$ possède les singletons, il est clair que cette condition suffisante est aussi nécessaire. La caractérisation de la $\mathcal{F}$-$\mathcal{P}(\mathbb{K})$ mesurabilité de X par (2.4) avait été prise comme définition d'une variable aléatoire discrète en DEUG (voir [ICP, Déf. 3.1]).

Preuve. Posons $\Omega_2 := X(\Omega)$, $\Omega_2^c := \mathbb{K} \setminus \Omega_2$ et notons $\mathcal{S}$ la famille des singletons de Ω_2. Comme Ω_2 est au plus dénombrable, $\sigma(\mathcal{S}) = \mathcal{P}(\Omega_2)$. Donc par la proposition 2.9, X *considérée comme application* $\Omega \to \Omega_2$ est $\mathcal{F}$-$\mathcal{P}(\Omega_2)$ mesurable, d'où

$$\forall A \in \mathcal{P}(\Omega_2), \quad X^{-1}(A) \in \mathcal{F}. \tag{2.5}$$

Notons qu'il n'y a pas ici d'ambiguïté sur l'écriture $X^{-1}(A)$ qui représente le même ensemble, que l'on prenne Ω_2 ou $\mathbb{K}$ pour ensemble d'arrivée de l'application X. Soit maintenant $B \in \mathcal{P}(\mathbb{K})$ quelconque. En écrivant

$$X^{-1}(B) = X^{-1}(B \cap \Omega_2) \cup X^{-1}(B \cap \Omega_2^c),$$

en remarquant que $X^{-1}(B \cap \Omega_2^c) = \emptyset$ et en appliquant (2.5) à $A := B \cap \Omega_2$, on obtient l'appartenance à $\mathcal{F}$ de $X^{-1}(B)$. L'application X est donc bien $\mathcal{F}$-$\mathcal{P}(\mathbb{K})$ mesurable. □

Revenant à la définition 2.5, on observe son analogie formelle avec la définition de la continuité d'une application entre deux espaces topologiques : en remplaçant les tribus $\mathcal{F}_1$ et $\mathcal{F}_2$ par des *topologies* $\mathcal{T}_1$ et $\mathcal{T}_2$, la définition deviendrait celle d'une application continue. Un cas important est celui où Ω_1 et Ω_2 sont munis de leurs tribus boréliennes $\sigma(\mathcal{T}_1)$ et $\sigma(\mathcal{T}_2)$.

Définition 2.12. *Soient* $(\Omega_1, \mathcal{T}_1)$ *et* $(\Omega_2, \mathcal{T}_2)$ *deux espaces topologiques, munis de leurs tribus boréliennes respectives* $\mathcal{B}_1 = \sigma(\mathcal{T}_1)$ *et* $\mathcal{B}_2 = \sigma(\mathcal{T}_2)$*. Une application* $f : \Omega_1 \to \Omega_2$ *est dite borélienne si elle est* $\mathcal{B}_1$*-*$\mathcal{B}_2$ *mesurable.*

Proposition 2.13. *Soient* $(\Omega_1, \mathcal{T}_1)$ *et* $(\Omega_2, \mathcal{T}_2)$ *deux espaces topologiques. Toute application continue* $f : \Omega_1 \to \Omega_2$ *est borélienne.*

Preuve. Soit $V \in \mathcal{T}_2$ un ouvert de Ω_2. Par continuité de f, $f^{-1}(V)$ est un ouvert de Ω_1, c'est-à-dire un élément de $\mathcal{T}_1$. C'est donc aussi un élément de $\mathcal{B}_1 = \sigma(\mathcal{T}_1)$. Ceci étant vrai pour tout V, on a ainsi établi l'inclusion $f^{-1}(\mathcal{T}_2) \subset \mathcal{B}_1$. Par la proposition 2.9, ceci implique la $\mathcal{B}_1$-$\mathcal{B}_2$ mesurabilité de f qui est donc bien borélienne. □

Nous examinons maintenant l'effet sur la mesurabilité des opérations usuelles sur les fonctions.

Proposition 2.14. *Soient* $(\Omega_i, \mathcal{F}_i)$*,* $i = 1, 2, 3$ *des espaces mesurables,* $f : \Omega_1 \to \Omega_2$ *une application* $\mathcal{F}_1$*-*$\mathcal{F}_2$ *mesurable et* $g : \Omega_2 \to \Omega_3$*,* $\mathcal{F}_2$*-*$\mathcal{F}_3$ *mesurable. Alors* $g \circ f$ *est* $\mathcal{F}_1$*-*$\mathcal{F}_3$ *mesurable.*

Preuve. Soit $C \subset \Omega_3$. En appliquant de façon répétée la définition de l'inverse ensembliste, on obtient :

$$\begin{aligned}
(g \circ f)^{-1}(C) &= \{\omega \in \Omega_1;\ (g \circ f)(\omega) \in C\} \\
&= \{\omega \in \Omega_1;\ g\big(f(\omega)\big) \in C\} \\
&= \{\omega \in \Omega_1;\ f(\omega) \in g^{-1}(C)\} \\
&= \{\omega \in \Omega_1;\ \omega \in f^{-1}\big(g^{-1}(C)\big)\} \\
&= f^{-1}\big(g^{-1}(C)\big).
\end{aligned}$$

En prenant C quelconque dans $\mathcal{F}_3$, on en déduit

$$(g \circ f)^{-1}(\mathcal{F}_3) = f^{-1}\big(g^{-1}(\mathcal{F}_3)\big).$$

Les mesurabilités respectives de g et f se traduisant par $g^{-1}(\mathcal{F}_3) \subset \mathcal{F}_2$ et $f^{-1}(\mathcal{F}_2) \subset \mathcal{F}_1$, on en déduit l'inclusion $(g \circ f)^{-1}(\mathcal{F}_3) \subset \mathcal{F}_1$ qui exprime la $\mathcal{F}_1$-$\mathcal{F}_3$ mesurabilité de $g \circ f$. □

Proposition 2.15. *Soient* $(\Omega, \mathcal{F})$ *un espace mesurable,* f *et* g *deux applications de* Ω *dans* $\mathbb{R}$*. L'application* $h = (f, g) : \Omega \to \mathbb{R}^2$ *est* $\mathcal{F}$-$\mathrm{Bor}(\mathbb{R}^2)$ *mesurable si et seulement si* f *et* g *sont* $\mathcal{F}$-$\mathrm{Bor}(\mathbb{R})$ *mesurables.*

Preuve. Vérifions d'abord que la mesurabilité de f et g est nécessaire à celle de h. Supposons donc h mesurable. Les projections $\pi_1 : (x, y) \mapsto x$ et $\pi_2 : (x, y) \mapsto y$ étant continues sont boréliennes et il suffit d'appliquer la proposition 2.14 à $f = \pi_1 \circ h$ et $g = \pi_2 \circ h$ pour obtenir la mesurabilité de f et g.

Réciproquement, supposons f et g $\mathcal{F}$-Bor($\mathbb{R}$) mesurables. Soit $\mathcal{C}$ la classe des pavés $]a,b]\times]c,d]$ dans $\mathbb{R}^2$. On a

$$\begin{aligned} h^{-1}(]a,b]\times]c,d]) &= \{\omega\in\Omega;\ \big(f(\omega),g(\omega)\big)\in]a,b]\times]c,d]\} \\ &= \{\omega\in\Omega;\ f(\omega)\in]a,b] \text{ et } g(\omega)\in]c,d]\} \\ &= \{\omega\in\Omega;\ \omega\in f^{-1}(]a,b]) \text{ et } \omega\in g^{-1}(]c,d])\} \\ &= f^{-1}(]a,b])\cap g^{-1}(]c,d]). \end{aligned}$$

Grâce à la mesurabilité de f et g on en déduit $h^{-1}(\mathcal{C})\subset\mathcal{F}$. Comme $\sigma(\mathcal{C})=\mathrm{Bor}(\mathbb{R}^2)$, la mesurabilité de h en découle via la proposition 2.9. □

Corollaire 2.16. *Soit $(\Omega,\mathcal{F})$ un espace mesurable, f et g deux applications de Ω dans $\mathbb{R}$.*

i) *La fonction complexe $f+ig:\Omega\to\mathbb{C}$ est $\mathcal{F}$-Bor($\mathbb{C}$) mesurable si et seulement si f et g sont $\mathcal{F}$-Bor($\mathbb{R}$) mesurables.*

ii) *Si f et g sont $\mathcal{F}$-Bor($\mathbb{R}$) mesurables, il en est de même pour les fonctions $f+g$, fg, cf ($c\in\mathbb{R}$ constante) et $|f|$. Ceci se généralise aux fonctions à valeurs complexes en remplaçant Bor($\mathbb{R}$) par Bor($\mathbb{C}$).*

iii) *Si f et g sont $\mathcal{F}$-Bor($\mathbb{R}$) mesurables, il en est de même pour les fonctions $\min(f,g)$ et $\max(f,g)$.*

iv) *f est $\mathcal{F}$-Bor($\mathbb{R}$) mesurable si et seulement si $f^+=\max(f,0)$ et $f^-=-\min(f,0)$ le sont.*

Preuve. Par souci de concision (on ne rit pas…), nous laissons au lecteur le soin de préciser les tribus concernées par les mesurabilités évoquées ci-dessous.

Pour i), supposons f et g mesurables, alors $(f,g):\Omega\to\mathbb{R}^2$ l'est aussi (cf. Prop. 2.15). En composant avec l'application continue $\mathbb{R}^2\to\mathbb{C}$, $(x,y)\mapsto x+iy$, on obtient la mesurabilité de $f+ig$ (cf. Prop. 2.13). Réciproquement, les projections π_1 et π_2 de $\mathbb{C}$ sur respectivement l'axe réel et l'axe des imaginaires purs étant continues, la mesurabilité de $f+ig$ entraîne celles de $f=\pi_1\circ(f+ig)$ et $g=\pi_2\circ(f+ig)$ (cf. Prop. 2.13 et Prop. 2.14).

Pour ii), on pose $h=(f,g)$, $s:\mathbb{R}^2\to\mathbb{R}$, $(x,y)\mapsto x+y$, $p:\mathbb{R}^2\to\mathbb{R}$, $(x,y)\mapsto xy$, $p_c:\mathbb{R}\to\mathbb{R}$, $x\mapsto cx$, $a:\mathbb{R}\to\mathbb{R}$, $x\mapsto|x|$. Les applications p, s, p_c et a sont continues donc boréliennes. On utilise alors les propositions 2.14 et 2.15 en notant que : $f+g=s\circ h$, $fg=p\circ h$, $cf=p_c\circ f$, $|f|=a\circ f$. Le cas complexe se ramène au cas réel après séparation des parties réelles et imaginaires des fonctions.

Pour iii), on note m et M les applications continues de $\mathbb{R}^2$ dans $\mathbb{R}$ définies par $m(x,y)=\min(x,y)$ et $M(x,y)=\max(x,y)$ et on obtient les mesurabilités souhaitées par composition puisque $\min(f,g)=m\circ h$ et $\max(f,g)=M\circ h$. Rappelons à cette occasion, que si $\min(x,y)$ est toujours égal à l'un des deux *nombres* x ou y, la *fonction* $\min(f,g)$ n'est en général égale à aucune des deux fonctions f et g.

La nécessité de la mesurabilité de f^+ et f^- dans iv) découle de iii) avec $g=0$ (et aussi de ii) avec $c=-1$). La suffisance provient de ii) en remarquant que $f=f^+-f^-$. □

La mesurabilité des applications est un outil commode pour prouver l'appartenance à $\mathcal{F}$ de certains sous ensembles de Ω. Voici un exemple d'usage courant. Soient f, g : $\Omega \to \mathbb{R}$, $\mathcal{F}$-Bor($\mathbb{R}$) mesurables. Définissons l'ensemble $\{f \leq g\}$ par

$$\{f \leq g\} := \Big\{\omega \in \Omega;\ f(\omega) \leq g(\omega)\Big\}.$$

Alors $\{f \leq g\}$ appartient à la tribu $\mathcal{F}$. Il y a au moins deux façons de le voir. On peut noter que $h := g - f$ est $\mathcal{F}$-Bor($\mathbb{R}$) mesurable (corollaire 2.16 ii) et que $\{f \leq g\} = h^{-1}([0, +\infty[)$ est élément de $\mathcal{F}$ comme image réciproque par h d'un fermé (donc borélien) de $\mathbb{R}$. On peut aussi utiliser directement la proposition 2.15 en exploitant la $\mathcal{F}$-Bor($\mathbb{R}^2$) mesurabilité de $H := (f, g)$ et en notant que $\{f \leq g\} = H^{-1}(A)$, où $A := \{(x, y) \in \mathbb{R}^2;\ x \leq y\}$ est fermé dans $\mathbb{R}^2$. Cette digression sournoise permet à l'auteur d'amener la question : a-t-on la même conclusion si f et g sont à valeurs dans $\overline{\mathbb{R}}_+$ et sont $\mathcal{F}$-Bor($\overline{\mathbb{R}}_+$) mesurables ? Il est clair que la première solution basée sur $g - f$ est irrecevable puisque $g - f$ n'est pas forcément défini sur tout Ω. La deuxième solution supposerait une exploration préalable de la topologie de $\overline{\mathbb{R}}_+^2$ et de sa tribu borélienne (a-t-elle une famille génératrice formée de produits cartésiens de boréliens de $\overline{\mathbb{R}}_+$?...) en vue de prouver l'analogue de la proposition 2.15. Nous proposons une troisième voie basée sur une méthode de discrétisation pour établir le résultat. Celui-ci sera réutilisé au chapitre 3.

Proposition 2.17. *Si les applications $f, g : \Omega \to \overline{\mathbb{R}}_+$ sont $\mathcal{F}$-*Bor($\overline{\mathbb{R}}_+$) *mesurables, l'ensemble $\{f \leq g\}$ appartient à $\mathcal{F}$.*

Preuve. En remarquant que si $g(\omega) = +\infty$, l'inégalité $f(\omega) \leq g(\omega)$ est automatiquement satisfaite, autrement dit que $\{g = +\infty\} \subset \{f \leq g\}$, on peut écrire

$$\{f \leq g\} = \{f \leq g \text{ et } g = +\infty\} \cup \{f \leq g < +\infty\} = \{g = +\infty\} \cup \{f \leq g < +\infty\}.$$

Comme g est $\mathcal{F}$-Bor($\overline{\mathbb{R}}_+$) mesurable, $g^{-1}(\{+\infty\}) \in \mathcal{F}$ et on est ramenés à établir l'appartenance à $\mathcal{F}$ de $\{f \leq g < +\infty\}$. Pour cela, commençons par prouver l'équivalence

$$f(\omega) \leq g(\omega) < +\infty \Leftrightarrow \forall n \in \mathbb{N}, \exists k \in \mathbb{N},\ \frac{k}{2^n} \leq g(\omega) < \frac{k+1}{2^n} \text{ et } f(\omega) < \frac{k+1}{2^n}. \qquad (2.6)$$

L'implication directe est évidente. Pour la réciproque, le cas $n = 0$ donne la finitude de $g(\omega)$. D'autre part $k = k(n, \omega)$ est unique pour ω et n fixés. Il est clair que $u_n := 2^{-n}(k(n, \omega) + 1)$ tend vers $g(\omega)$ quand n tend vers $+\infty$ puisque $0 \leq u_n - g(\omega) \leq 2^{-n}$. En passant à la limite dans l'inégalité $f(\omega) < u_n$, on en déduit $f(\omega) \leq g(\omega)$. La traduction du second membre de l'équivalence (2.6) en opérations ensemblistes nous donne maintenant la représentation

$$\{f \leq g < +\infty\} = \bigcap_{n\in\mathbb{N}} \bigcup_{k\in\mathbb{N}} \biggl(g^{-1}\Bigl(\Bigl[\frac{k}{2^n}, \frac{k+1}{2^n}\Bigr[\Bigr) \cap f^{-1}\Bigl(\Bigl[0, \frac{k+1}{2^n}\Bigr[\Bigr)\biggr).$$

L'appartenance à $\mathcal{F}$ de $\{f \leq g < +\infty\}$ résulte alors clairement de la mesurabilité de f et de g. $\square$

Dans l'étude des suites de fonctions mesurables à valeurs dans $\mathbb{R}$ ou $\mathbb{R}_+$, il est souvent commode d'agrandir l'espace d'arrivée en y rajoutant les (ou le) points à l'infini. La proposition suivante nous rassure quant à la préservation de la mesurabilité.

Proposition 2.18. *Soit $f : \Omega \to \mathbb{R}$, une application $\mathcal{F}$-$\mathrm{Bor}(\mathbb{R})$ mesurable. Elle est aussi $\mathcal{F}$-$\mathrm{Bor}(\overline{\mathbb{R}})$ mesurable lorsqu'on la considère comme application $\Omega \to \overline{\mathbb{R}}$. De même si $f : \Omega \to \mathbb{R}_+$ est $\mathcal{F}$-$\mathrm{Bor}(\mathbb{R}_+)$ mesurable, elle est aussi $\mathcal{F}$-$\mathrm{Bor}(\mathbb{K})$ mesurable comme application $\Omega \to \mathbb{K}$, pour chacun des ensembles d'arrivée $\mathbb{K} = \overline{\mathbb{R}}_+$, $\mathbb{R}$, $\overline{\mathbb{R}}$.*

Preuve. Pour traiter d'un coup les deux situations, notons $\mathbb{K}_0$ l'ensemble d'arrivée « initial » ($\mathbb{K}_0 = \mathbb{R}$ ou $\mathbb{R}_+$), $\mathbb{K}$ l'ensemble d'arrivée « élargi » et $\mathbb{K}_0^c := \mathbb{K} \setminus \mathbb{K}_0$.

Soit $B \in \mathrm{Bor}(\mathbb{K})$, on peut l'écrire $B = (B \cap \mathbb{K}_0) \cup (B \cap \mathbb{K}_0^c)$, d'où

$$f^{-1}(B) = f^{-1}(B \cap \mathbb{K}_0) \cup f^{-1}(B \cap \mathbb{K}_0^c) = f^{-1}(B \cap \mathbb{K}_0),$$

puisque f étant à valeurs dans $\mathbb{K}_0$, $f^{-1}(B \cap \mathbb{K}_0^c) = \emptyset$ (l'agrandissement de l'ensemble d'arrivée n'a pas changé $f(\Omega)$). Grâce aux propositions 2.2 et 2.3, on voit que $B \cap \mathbb{K}_0$ est aussi un borélien de $\mathbb{K}_0$. Par l'hypothèse de mesurabilité $\mathcal{F}$-$\mathrm{Bor}(\mathbb{K}_0)$ de f, on en déduit l'appartenance à $\mathcal{F}$ de $f^{-1}(B \cap \mathbb{K}_0)$. Ainsi $f^{-1}(B) \in \mathcal{F}$ pour tout $B \in \mathrm{Bor}(\mathbb{K})$. $\square$

Proposition 2.19. *Soient $(\Omega, \mathcal{F})$ un espace mesurable et $(f_n)_{n\geq 1}$ une suite d'applications $\mathcal{F}$-$\mathrm{Bor}(\overline{\mathbb{R}})$ mesurables $\Omega \to \overline{\mathbb{R}}$. Alors $h := \sup_{n\geq 1} f_n$ est $\mathcal{F}$-$\mathrm{Bor}(\overline{\mathbb{R}})$ mesurable. De même, si les f_n sont à valeurs dans $\overline{\mathbb{R}}_+$ et $\mathcal{F}$-$\mathrm{Bor}(\overline{\mathbb{R}}_+)$ mesurables, h est $\mathcal{F}$-$\mathrm{Bor}(\overline{\mathbb{R}}_+)$ mesurable.*

En raison de la proposition 2.18, la mesurabilité $\mathcal{F}$-$\mathrm{Bor}(\overline{\mathbb{K}})$ de h reste vraie si les f_n sont à valeurs dans $\mathbb{K} = \mathbb{R}$ ou $\mathbb{R}_+$ et $\mathcal{F}$-$\mathrm{Bor}(\mathbb{K})$ mesurables [3].

Preuve. La tribu borélienne de $\overline{\mathbb{R}}$ étant engendrée par les $[-\infty, a]$, il suffit de montrer que pour tout $a \in \mathbb{R}$, $h^{-1}([-\infty, a]) \in \mathcal{F}$ (cf. Cor. 2.10 iii). On remarque d'abord que si $(u_n)_{n\geq 1}$ est une suite dans $\overline{\mathbb{R}}$, on a l'équivalence [4]

$$\sup_{n\in\mathbb{N}^*} u_n \leq a \iff \forall n \in \mathbb{N}^*,\ u_n \leq a.$$

Cette équivalence nous permet d'écrire :

$$\begin{aligned} h^{-1}([-\infty, a]) &= \{\omega \in \Omega;\ \sup_{n\geq 1} f_n(\omega) \leq a\} \\ &= \{\omega \in \Omega;\ \forall n \in \mathbb{N}^*,\ f_n(\omega) \leq a\} \\ &= \{\omega \in \Omega;\ \forall n \in \mathbb{N}^*,\ f_n(\omega) \in [-\infty, a]\} \\ &= \bigcap_{n\in\mathbb{N}^*} f_n^{-1}([-\infty, a]), \end{aligned}$$

ce qui nous donne la conclusion souhaitée grâce à la mesurabilité de chaque f_n et à la stabilité de la tribu $\mathcal{F}$ par intersection *dénombrable.*

Le cas où les f_n sont à valeurs dans $\overline{\mathbb{R}}_+$ se traite de la même façon, en appliquant le point iv) au lieu du iii) dans le corollaire 2.10. $\square$

3. On est obligé de garder $\overline{\mathbb{K}}$ comme ensemble d'arrivée pour h car le supremum d'une suite de fonctions à valeurs finies peut très bien prendre la valeur $+\infty$.

4. Attention, cette équivalence ne subsiste pas si l'on remplace les inégalités larges par des inégalités strictes. Dans ce cas on aurait toujours « $\Rightarrow$ » mais plus « $\Leftarrow$ ».

Corollaire 2.20. *Soient* $(\Omega, \mathcal{F})$ *un espace mesurable et* $(f_n)_{n\geq 1}$ *une suite d'applications* $\mathcal{F}$-$\mathrm{Bor}(\overline{\mathbb{R}})$ *mesurables* $\Omega \to \overline{\mathbb{R}}$.

i) Les fonctions $\sup_{n\geq 1} f_n$ *et* $\inf_{n\geq 1} f_n$ *sont* $\mathcal{F}$-$\mathrm{Bor}(\overline{\mathbb{R}})$ *mesurables.*

ii) Les fonctions $\limsup_{n\to+\infty} f_n$ *et* $\liminf_{n\to+\infty} f_n$ *sont* $\mathcal{F}$-$\mathrm{Bor}(\overline{\mathbb{R}})$ *mesurables.*

iii) Si f *est limite simple sur* Ω *de* f_n *(c'est-à-dire si* $\forall \omega \in \Omega,\ f_n(\omega) \to f(\omega) \in \overline{\mathbb{R}}$*),* f *est* $\mathcal{F}$-$\mathrm{Bor}(\overline{\mathbb{R}})$ *mesurable.*

Cet énoncé reste valable en remplaçant partout $\overline{\mathbb{R}}$ *par* $\overline{\mathbb{R}}_+$.

On peut faire ici aussi la même remarque qu'après la proposition 2.19, pour le cas où les f_n sont à valeurs dans $\mathbb{R}$ ou $\mathbb{R}_+$.

Preuve. Pour i), il suffit de remarquer que $\inf_{n\geq 1} f_n = -\sup_{n\geq 1}(-f_n)$. Pour ii), on se ramène à i) en écrivant

$$\liminf_{n\to+\infty} f_n = \sup_{n\geq 1} \inf_{k\geq n} f_k, \qquad \limsup_{n\to+\infty} f_n = \inf_{n\geq 1} \sup_{k\geq n} f_k.$$

Enfin, si f est limite simple de f_n, $f = \limsup_{n\to+\infty} f_n$, donc iii) découle de ii).

Dans le cas où les f_n sont à valeurs dans $\overline{\mathbb{R}}_+$, la seule chose à changer est la justification de la mesurabilité de $g := \inf_{n\geq 1} f_n$. La tribu borélienne de $\overline{\mathbb{R}}_+$ est engendrée par la famille $\mathcal{J}_1' := \{[a, +\infty];\ a \in \mathbb{R}_+\}$ et l'équivalence

$$\inf_{n\in\mathbb{N}^*} u_n \geq a \Longleftrightarrow \forall n \in \mathbb{N}^*,\ u_n \geq a,$$

nous permet d'écrire $g^{-1}([a, +\infty]) = \bigcap_{n\in\mathbb{N}^*} f_n^{-1}([a, +\infty])$, justifiant ainsi l'appartenance à $\mathcal{F}$ de $g^{-1}([a, +\infty])$ et la mesurabilité de g. □

2.4 Fonctions étagées

Les fonctions étagées sont les fonctions les plus simples à partir desquelles on pourra construire l'intégrale abstraite par approximation.

Définition 2.21. *On dit qu'une application* $f : \Omega \to \mathbb{R}$ *est une fonction étagée si l'ensemble* $f(\Omega)$ *de ses valeurs est fini. En notant* $f(\Omega) = \{y_1, y_2, \dots y_n\}$ *(les* y_i *étant tous distincts) et* $A_i = f^{-1}(\{y_i\})$*, on a la décomposition canonique*

$$f = \sum_{i=1}^{n} y_i \mathbf{1}_{A_i} \tag{2.7}$$

et $\{A_i;\ 1 \leq i \leq n\}$ *est une partition finie de* Ω.

Si $0 \in f(\Omega)$, il faut l'écrire dans la décomposition (2.7) pour avoir vraiment la décomposition canonique. Dans le même ordre d'idées, si une fonction g s'écrit $g = \sum_{j=1}^{m} c_j \mathbf{1}_{B_j}$, les c_j étant distincts et tous non nuls, les B_j étant deux à deux disjoints mais

$B := \bigcup_{j=1}^{m} B_j \neq \Omega$, alors g est étagée et sa décomposition canonique est $g = \sum_{j=1}^{m+1} c_j \mathbf{1}_{B_j}$, où l'on a posé $c_{m+1} = 0$ et $B_{m+1} = B^c$. Enfin, si $g = \sum_{j=1}^{m} c_j \mathbf{1}_{B_j}$, les c_j étant distincts et tous non nuls, les B_j étant quelconques, g est encore étagée. Pour avoir sa décompostion canonique, il faut considérer la partition de Ω engendrée par la famille $B_1, \dots, B_m, B_{m+1}$ (cette partition est formée de celles des intersections $D = C_1 \cap C_2 \cap \dots \cap C_{m+1}$ qui sont non vides, C_j désignant l'un des deux ensembles B_j ou B_j^c) et calculer la valeur constante de g sur chacun des ensembles D et éventuellement réunir les D donnant la même valeur pour g.

Proposition 2.22. *Soient* $(\Omega, \mathcal{F})$ *un espace mesurable et* $f : \Omega \to \mathbb{R}$ *une fonction étagée, de décomposition canonique* $f = \sum_{i=1}^{n} y_i \mathbf{1}_{A_i}$. *Alors* f *est* $\mathcal{F}$-Bor$(\mathbb{R})$ *mesurable si et seulement si tous les* A_i *sont dans* $\mathcal{F}$.

Preuve. Si tous les A_i sont dans $\mathcal{F}$, f est mesurable comme combinaison linéaire d'indicatrices d'éléments de $\mathcal{F}$ (cf. p. 62 et Cor. 2.16 ii)).

Réciproquement, si la fonction étagée f est mesurable, comme la tribu borélienne de $\mathbb{R}$ contient les singletons, chaque $A_i = f^{-1}(\{y_i\})$ est dans $\mathcal{F}$. □

Si $f = \sum_{i=1}^{n} y_i \mathbf{1}_{A_i}$, les A_i étant tous dans $\mathcal{F}$ et les y_i des réels positifs, f est mesurable $\mathcal{F}$-Bor$(\mathbb{R}_+)$. Par la proposition 2.18, elle est aussi mesurable $\mathcal{F}$-Bor$(\mathbb{K})$ pour $\mathbb{K} = \overline{\mathbb{R}}_+$, $\mathbb{R}$, $\overline{\mathbb{R}}$. Nous appelerons dans la suite « fonction étagée mesurable positive » une telle fonction f sans préciser l'ensemble d'arrivée ni la tribu borélienne considérés.

Théorème 2.23. *Soit* $(\Omega, \mathcal{F})$ *un espace mesurable. Toute fonction* $f : \Omega \to \overline{\mathbb{R}}_+$ *et mesurable* $\mathcal{F}$-Bor$(\overline{\mathbb{R}}_+)$ *est limite simple sur* Ω *d'une suite* croissante $(f_n)_{n\geq 1}$ *de fonctions étagées mesurables positives.*

Preuve. L'idée est d'utiliser pour construire f_n, les valeurs approchées dyadiques par défaut de f au niveau de résolution n. On peut procéder comme suit en définissant pour $n \in \mathbb{N}$, les ensembles

$$\begin{aligned} A_{n,k} &:= f^{-1}\Big([k2^{-n}, (k+1)2^{-n}[\Big), \quad 0 \leq k \leq n2^n - 1; \\ A_{n,n2^n} &:= f^{-1}\Big([n, +\infty]\Big). \end{aligned}$$

On prend alors pour f_n la fonction étagée positive de décomposition canonique

$$f_n = \sum_{k=0}^{n2^n} \frac{k}{2^n} \mathbf{1}_{A_{n,k}}.$$

Comme f est mesurable, les $A_{n,k}$ sont dans $\mathcal{F}$, ce qui entraîne la mesurabilité de f_n (Prop. 2.22).

Il reste à vérifier que pour tout $\omega \in \Omega$, la suite de *réels* $\big(f_n(\omega)\big)_{n\geq 1}$ est croissante et converge dans $\overline{\mathbb{R}}$ vers $f(\omega)$.

Réglons d'abord le cas où $f(\omega) = +\infty$. Alors pour tout n, $f_n(\omega) = n$, ce qui nous donne bien une suite croissante convergente vers $+\infty = f(\omega)$.

Si $f(\omega) < +\infty$, $f_n(\omega) = n$ pour $n \leq [f(\omega)]$ (on note $[x]$ la partie entière de x, unique entier m tel que $m \leq x < m+1$) et pour $n > [f(\omega)]$,

$$f_n(\omega) = \frac{k(n,\omega)}{2^n} = \max\Big\{\frac{l}{2^n};\ \frac{l}{2^n} \leq f(\omega),\ l \in \mathbb{N}\Big\}. \tag{2.8}$$

La suite *finie* $\big(f_n(\omega)\big)_{n\leq[f(\omega)]}$ est clairement croissante. Regardons la suite $\big(f_n(\omega)\big)_{n>[f(\omega)]}$. D'après (2.8), on a

$$f_n(\omega) = \frac{k(n,\omega)}{2^n} = \frac{2k(n,\omega)}{2^{n+1}} \leq f(\omega) \quad \Rightarrow \quad \frac{2k(n,\omega)}{2^{n+1}} \leq \frac{k(n+1,\omega)}{2^{n+1}} = f_{n+1}(\omega),$$

d'où la croissance de la suite $\big(f_n(\omega)\big)_{n>[f(\omega)]}$. Pour établir définitivement la croissance de toute la suite $\big(f_n(\omega)\big)_{n\geq 1}$, il ne reste plus qu'à examiner le point de raccord des deux sous-suites, donc à comparer $f_n(\omega)$ et $f_{n+1}(\omega)$ pour $n = [f(\omega)]$. Il suffit de remarquer que $f_n(\omega) = n = (n2^{n+1})2^{-n-1} \leq f(\omega)$ et comme $f_{n+1}(\omega)$ est donné par (2.8), on a $(n2^{n+1})2^{-n-1} \leq k(n+1,\omega)2^{-n-1} = f_{n+1}(\omega)$.

La convergence est immédiate, puisque pour $n > [f(\omega)]$, on a d'après (2.8)

$$f_n(\omega) \leq f(\omega) < f_n(\omega) + \frac{1}{2^n},$$

d'où $0 \leq f(\omega) - f_n(\omega) < 2^{-n}$. □

Remarque : La convergence est *uniforme* sur Ω si f est bornée (*i.e.* $M := \sup_\Omega f < +\infty$) car pour $n > M$, on a pour tout $\omega \in \Omega$, $0 \leq f(\omega) - f_n(\omega) < 2^{-n}$.

Pour illustrer la convergence de f_n vers f, on a choisi ci-dessous $\Omega = [0,4]$, $f(x) = \big(2+\sin(x^2)\big)\sqrt{|x-1|}$ et représenté les fonctions étagées f_1, f_2 et f_3 (figures 2.2 à 2.4).

Corollaire 2.24. *Soit* $(\Omega, \mathcal{F})$ *un espace mesurable. Une fonction* $g : \Omega \to \overline{\mathbb{R}}$ *est* $\mathcal{F}$-$\mathrm{Bor}(\overline{\mathbb{R}})$ *mesurable si et seulement si elle est limite simple sur* Ω *d'une suite de fonctions étagées* $\mathcal{F}$-$\mathrm{Bor}(\mathbb{R})$ *mesurables.*

Preuve. On sait déjà qu'une limite simple de fonctions mesurables est mesurable (Cor. 2.20 iii). Dans l'autre sens, on écrit $g = g^+ - g^-$ et on utilise le théorème 2.23 en remarquant que pour tout ω au plus une seule des deux valeurs $g^+(\omega)$ et $g^-(\omega)$ est non nulle et que si $f(\omega) = 0$ dans le Th. 2.23, alors tous les $f_n(\omega)$ sont nuls. Ceci permet de découper Ω en trois parties disjointes $\{g = 0\}$, $\{g > 0\}$ et $\{g < 0\}$ sur chacune desquelles on applique le théorème 2.23. □

2.5 Mesures images, lois

Les applications mesurables permettent de « transporter » la mesure d'un espace à un autre. Dans le cas d'une variable (ou d'un vecteur) aléatoire, la « mesure image » ainsi obtenue s'appelle la loi de la variable (du vecteur) aléatoire.

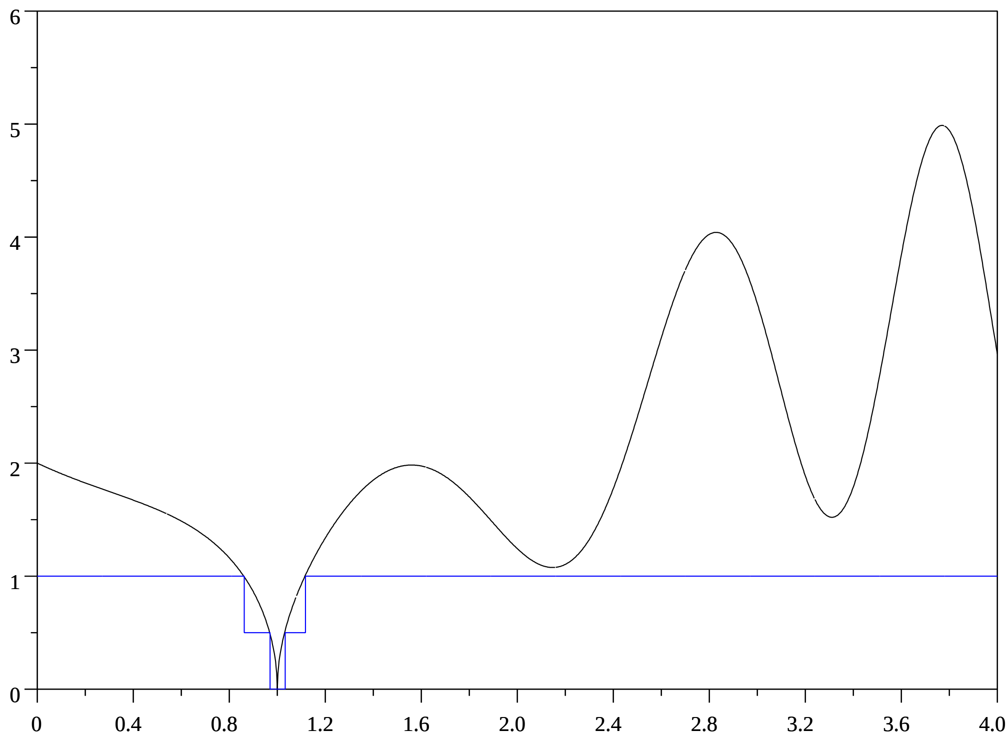


FIGURE 2.2 – f et f_1

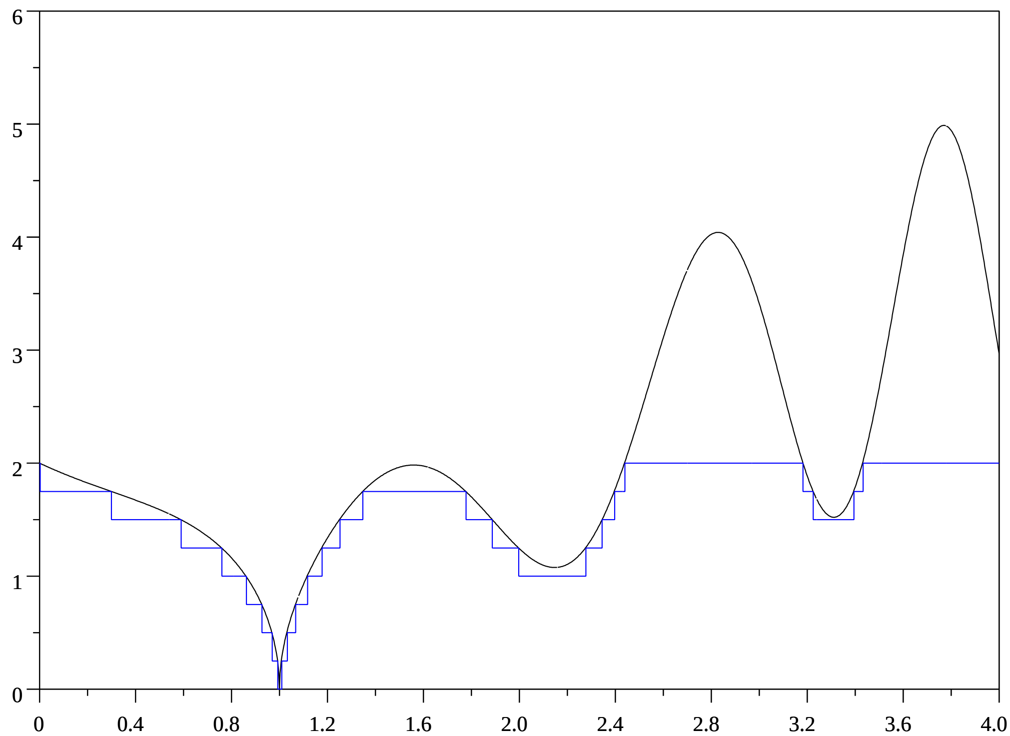


FIGURE 2.3 – f et f_2

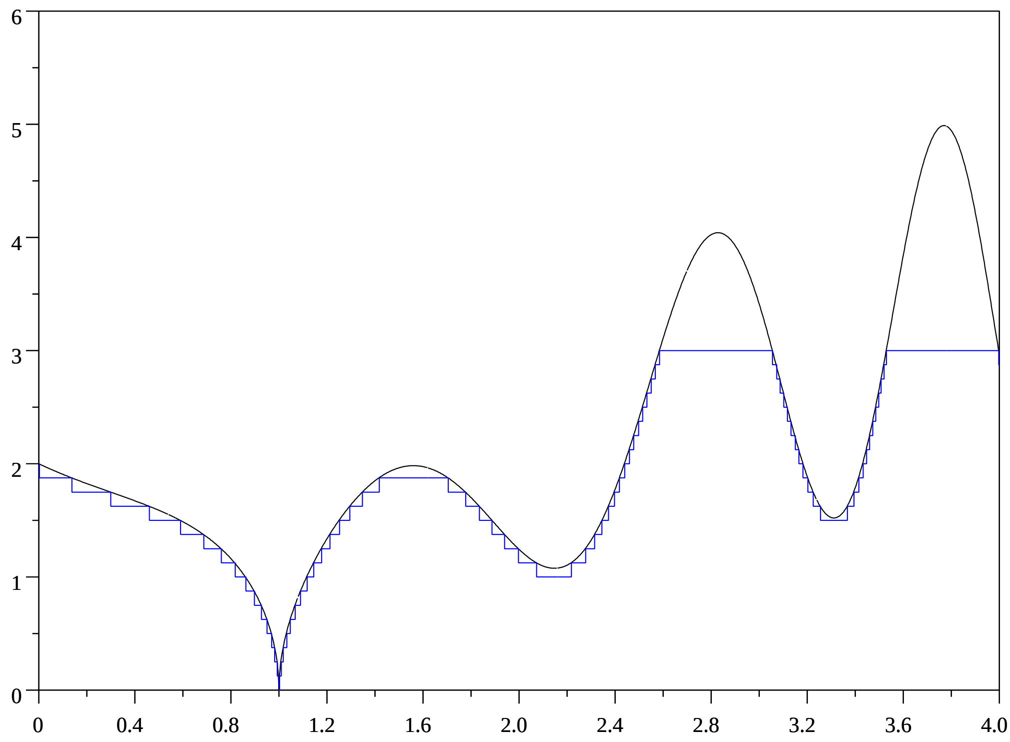


FIGURE 2.4 – f et f_3

2.5.1 Généralités

Proposition 2.25. *Soient* $(\Omega_i, \mathcal{F}_i)$, $i = 1, 2$ *des espaces mesurables,* μ *une mesure sur* $(\Omega_1, \mathcal{F}_1)$ *et* $f : \Omega_1 \to \Omega_2$ *une application* $\mathcal{F}_1$-$\mathcal{F}_2$ *mesurable. La fonction d'ensembles* $\mu \circ f^{-1}$ *définie sur* $\mathcal{F}_2$ *par*

$$\forall B \in \mathcal{F}_2, \quad (\mu \circ f^{-1})(B) := \mu\bigl(f^{-1}(B)\bigr)$$

est une mesure sur $(\Omega_2, \mathcal{F}_2)$. *On l'appelle* mesure image *de* μ *par* f.

Preuve. La fonction d'ensembles $\nu = \mu \circ f^{-1}$ est bien définie sur $\mathcal{F}_2$ puisque puisque pour $B \in \mathcal{F}_2$, $f^{-1}(B) \in \mathcal{F}_1$ par mesurabilité de f et μ est définie sur $\mathcal{F}_1$. Comme $f^{-1}(\emptyset) = \emptyset$, on a immédiatement $\nu(\emptyset) = \mu(\emptyset) = 0$. Pour la σ-additivité de ν, on note que l'image inverse ensembliste f^{-1} commute avec les intersections et les réunions. Soit $(B_k)_{n\in\mathbb{N}}$ une suite d'ensembles deux à deux disjoints, membres de $\mathcal{F}_2$. Pour $j \neq k$, $f^{-1}(B_j) \cap f^{-1}(B_k) = f^{-1}(B_j \cap B_k) = f^{-1}(\emptyset) = \emptyset$, donc les $f^{-1}(B_k)$ sont deux à deux disjoints et dans $\mathcal{F}_1$. La σ-additivité de ν résulte alors de celle de μ en écrivant

$$\nu\Bigl(\bigcup_{k\in\mathbb{N}} B_k\Bigr) = \mu\Bigl(f^{-1}\Bigl(\bigcup_{k\in\mathbb{N}} B_k\Bigr)\Bigr) = \mu\Bigl(\bigcup_{k\in\mathbb{N}} f^{-1}(B_k)\Bigr) = \sum_{k\in\mathbb{N}} \mu\bigl(f^{-1}(B_k)\bigr) = \sum_{k\in\mathbb{N}} \nu(B_k).$$

On a donc vérifié que ν est une mesure sur $(\Omega_2, \mathcal{F}_2)$. Remarquons au passage que $\nu(\Omega_2) = \mu(\Omega_1)$ car $f^{-1}(\Omega_2) = \Omega_1$. □

La notion de mesure image est particulièrement utile en théorie des probabilités puisqu'elle permet de transporter la probabilité d'un espace probabilisé abstrait sur un espace plus familier, comme $\mathbb{R}^d$ ou l'une de ses parties.

Définition 2.26. *Soit* $(\Omega, \mathcal{F}, \mathbf{P})$ *un espace probabilisé et* $X : \Omega \to \mathbb{R}^d$ *un vecteur aléatoire. La mesure image* $P_X := \mathbf{P} \circ X^{-1}$ *qui est une probabilité sur* $\mathrm{Bor}(\mathbb{R}^d)$, *est appelée* loi *du vecteur aléatoire* X *sous* $\mathbf{P}$ *(de la variable aléatoire* X *si* $d = 1$*).*

La loi de X sous $\mathbf{P}$ est donc la mesure de probabilité définie par

$$\forall B \in \mathrm{Bor}(\mathbb{R}^d), \quad P_X(B) = \mathbf{P}\bigl(X^{-1}(B)\bigr) = \mathbf{P}\bigl(\{\omega \in \Omega;\ X(\omega) \in B\}\bigr).$$

Il est courant de simplifier cette écriture en

$$P_X(B) = \mathbf{P}(X \in B).$$

L'usage est aussi d'abréger « loi de X sous $\mathbf{P}$ » en « loi de X » quand il n'y a pas d'ambiguïté, c'est-à-dire quand on n'envisage qu'une seule mesure de probabilité sur $(\Omega, \mathcal{F})$. Remarquons que la définition de X ne présuppose pas l'existence de $\mathbf{P}$. Remarquons aussi que deux variables aléatoires définies sur le même espace peuvent avoir même loi sans être égales, voire même avoir même loi sans être définies sur le même espace probabilisé...

Notons enfin que les situations où l'on munit $(\Omega, \mathcal{F})$ de plusieurs mesures de probabilité (et où il peut être utile de préciser sous quelle mesure on considère la loi de X) n'ont rien d'exceptionnel. On peut en mentionner deux d'importance fondamentale : le conditionnement et le modèle statistique (voir les sous-sections 2.5.4 et 2.5.5 ci-après).

Proposition 2.27 (Loi d'une variable aléatoire discrète)**.** *Soient* $(\Omega, \mathcal{F}, \mathbf{P})$ *un espace probabilisé et* $X : \Omega \to \mathbb{K}$ *une variable aléatoire discrète ou un vecteur aléatoire discret (au sens de la définition 2.8, avec* $\mathbb{K} = \mathbb{R}$*,* $\mathbb{C}$ *ou* $\mathbb{R}^d$*,* $d > 1$*). La loi de* X *sous* $\mathbf{P}$ *est la mesure* P_X *sur* $\mathcal{P}(\mathbb{K})$*, donc aussi par restriction sur* $\mathrm{Bor}(\mathbb{K})$*, donnée par :*

$$P_X = \sum_{x \in X(\Omega)} \mathbf{P}(X = x)\delta_x, \tag{2.9}$$

où δ_x *désigne la mesure de Dirac au point* x*. Rappelons qu'ici* $X(\Omega)$ *est au plus dénombrable, le second membre de (2.9) est donc une somme finie ou une série de mesures finies.*

Preuve. Nous savons déjà (cf. preuve du corollaire 2.11) que pour tout $B \in \mathcal{P}(\mathbb{K})$, $X^{-1}(B) = X^{-1}\big(B \cap X(\Omega)\big)$. Par conséquent, comme $P_X = \mathbf{P} \circ X^{-1}$,

$$\forall B \in \mathcal{P}(\mathbb{K}), \quad P_X(B) = P_X\big(B \cap X(\Omega)\big). \tag{2.10}$$

L'ensemble $B \cap X(\Omega)$ est au plus dénombrable, donc union finie ou dénombrable de ses singletons. Par additivité ou σ-additivité de la mesure P_X, on en déduit :

$$\begin{aligned} P_X\big(B \cap X(\Omega)\big) &= \sum_{x \in B \cap X(\Omega)} P_X(\{x\}) \\ &= \sum_{x \in B \cap X(\Omega)} \mathbf{P}(X = x) \\ &= \sum_{x \in X(\Omega)} \mathbf{P}(X = x)\mathbf{1}_B(x) \\ &= \sum_{x \in X(\Omega)} \mathbf{P}(X = x)\delta_x(B). \end{aligned}$$

Compte tenu de (2.10), nous venons ainsi de vérifier que :

$$\forall B \in \mathcal{P}(\mathbb{K}), \quad P_X(B) = \sum_{x \in X(\Omega)} \mathbf{P}(X = x)\delta_x(B),$$

ce qui est précisément la traduction de (2.9). □

Définition 2.28 (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,* $(x_i)_{i \in I}$ *est une famille de nombres réels et* $(p_i)_{i \in I}$ *est une famille sommable de réels positifs, de somme* $\sum_{i \in I} p_i = 1$*.*

Il est clair que par restriction, μ est aussi une probabilité sur $(\mathbb{R}, \mathrm{Bor}(\mathbb{R}))$.

Remarque 2.29. 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 2.28. C'est une conséquence de la proposition 2.27. 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 ($Y(\Omega)$ est 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 2.28.

2.5.2 Lois discrètes classiques

Nous revoyons ici les principales lois discrètes classiques, déjà étudiées en DEUG. Pour les détails, voir [ICP]. La notion d'indépendance des évènements ou des variables aléatoires utilisée dans cette présentation a été définie en DEUG. Elle sera généralisée et étudiée systématiquement au chapitre sur les mesures produits. Lorsque la loi de X est une mesure connue μ, ce qui est le cas pour tous les exemples qui suivent, il est d'usage de dire que X « suit » la loi μ ou encore que X « obéit » à la loi μ. Il est commode d'introduire ces lois discrètes à l'aide de v.a. discrètes, même si elles peuvent être suivies par des v.a. plus générales, cf. la remarque 2.29.

Lois de Bernoulli

Définition 2.30. *La variable aléatoire discrète X suit la loi de Bernoulli de paramètre p ($p \in [0, 1]$) si $X(\Omega) = \{0, 1\}$ et*

$$\mathbf{P}(X = 1) = p, \qquad \mathbf{P}(X = 0) = 1 - p = q.$$

On notera $X \sim \mathrm{Bern}(p)$.

Si A est un événement de probabilité p, son indicatrice $\mathbf{1}_A$ 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 $X = \mathbf{1}_A$ en définissant $A := \{\omega \in \Omega,\ X(\omega) = 1\}$.

Loi uniforme sur un ensemble fini de réels

Définition 2.31. *La variable aléatoire discrète X suit la loi uniforme sur l'ensemble* fini *de réels $A = \{x_1, \dots, x_n\}$ si $X(\Omega) = A$ et si P_X est l'équiprobabilité sur cet ensemble :*

$$\forall k = 1, \dots, n, \qquad \mathbf{P}(X = x_k) = \frac{1}{n}.$$

Par exemple le nombre de points obtenus lors du jet d'un dé équilibré suit la loi uniforme sur $\{1, 2, 3, 4, 5, 6\}$. Dire que P_X est la loi uniforme sur $A := \{x_1, \dots, x_n\}$ équivaut à

$$\forall B \in \mathcal{P}(\mathbb{R}), \quad \mathbf{P}(X \in B) = \frac{\operatorname{card}(A \cap B)}{\operatorname{card}(A)}. \tag{2.11}$$

La loi P_X apparaît ainsi comme le conditionnement [5] par A de la *mesure de comptage* ν sur $\mathbb{R}$ définie par : $\nu(B) = \operatorname{card} B$ si B est une partie finie de $\mathbb{R}$ et $\nu(B) = +\infty$ si B est une partie infinie de $\mathbb{R}$.

Lois binomiales

Définition 2.32. *La variable aléatoire discrète X suit la loi binomiale de paramètres n et p ($n \in \mathbb{N}^*$ et $p \in [0, 1]$) si l'ensemble des valeurs possibles est $X(\Omega) = \{0, 1, \dots, n\}$ et*

$$\forall k = 0, 1, \dots, n, \qquad \mathbf{P}(X = k) = C_n^k p^k (1 - p)^{n-k}.$$

Notation : $X \sim \operatorname{Bin}(n, p)$.

La loi binomiale $\operatorname{Bin}(n, p)$ 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.

Plus généralement, soit $A_1, \dots, A_n$ une famille de n événements mutuellement indépendants ayant tous même probabilité p et notons $X_i = \mathbf{1}_{A_i}$. Alors la variable aléatoire $S_n = \sum_{i=1}^{n} X_i$ suit la loi binomiale $\operatorname{Bin}(n, p)$.

Lois hypergéométriques

Alors que la loi binomiale intervient dans les tirages avec remise, la loi hypergéométrique correspond aux tirages sans remise.

Exemple 2.1. 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 ?

En considérant tous les échantillons possibles comme équiprobables, un peu de dénombrement mène à la formule suivante :

$$\mathbf{P}(X = k) = \frac{C_M^k \times C_{N-M}^{n-k}}{C_N^n} \quad \text{si} \quad \begin{cases} 0 \le k \le M, \\ 0 \le n - k \le N - M. \end{cases} \tag{2.12}$$

Définition 2.33. *La loi définie par (2.12) s'appelle loi hypergéométrique de paramètres N, M et n. Notation : $X \sim \operatorname{Hypg}(N, M, n)$. 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.*

5. Au sens de l'exemple 1.4 du chapitre précédent.

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 2.34. *On suppose que quand N tend vers $+\infty$, $M = M(N)$ tend vers $+\infty$ en vérifiant la condition :*

$$\lim_{N\to+\infty} \frac{M}{N} = p \quad \textit{avec} \quad 0 < p < 1. \tag{2.13}$$

Alors, n restant fixé, la loi hypergéométrique $\operatorname{Hypg}(N, M, n)$ *converge vers la loi binomiale* $\operatorname{Bin}(n, p)$*, ce qui signifie ici que si $(X_N)_{N\geq 1}$ est une suite de v.a. avec $X_N \sim$* $\operatorname{Hypg}(N, M, n)$ *et Y est une v.a. de loi* $\operatorname{Bin}(n, p)$*,alors :*

$$\forall k = 0, 1, \ldots, n, \qquad \lim_{N\to+\infty} \mathbf{P}(X_N = k) = \mathbf{P}(Y = k), \tag{2.14}$$

autrement dit :

$$\forall k = 0, 1, \ldots, n, \qquad \lim_{N\to+\infty} \frac{C_M^k \times C_{N-M}^{n-k}}{C_N^n} = C_n^k p^k (1-p)^{n-k}. \tag{2.15}$$

Pour la preuve, voir [ICP].

Lois multinomiales

Définition 2.35. *Le vecteur aléatoire discret $X : \Omega \to \mathbb{R}^d$ suit la loi multinomiale de paramètres n et $(p_1, \ldots, p_d)$ où $n \in \mathbb{N}^*$ et les p_i sont strictement positifs et de somme* 1 *si $X(\Omega)$ est l'ensemble des d-uples $(j_1, j_2, \ldots, j_d)$ d'entiers tels que $j_1 + j_2 + \cdots + j_d = n$ et si*

$$\forall (j_1, j_2, \ldots, j_d) \in X(\Omega), \quad \mathbf{P}\Big\{X = (j_1, j_2, \ldots, j_d)\Big\} = \frac{n!}{j_1! j_2! \ldots j_d!} p_1^{j_1} p_2^{j_2} \ldots p_d^{j_d}.$$

Notation : $X \sim$ $\operatorname{Mult}(n; p_1, \ldots, p_d)$

Rappelons que la loi multinomiale sert à modéliser le total des résultats de chaque type observés 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-ème 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 p_i est la probabilité d'apparition de la face n° i lors d'un lancer.

Lois géométriques

Définition 2.36. *Une variable aléatoire discrète X suit la loi géométrique de paramètre $p \in]0, 1[$, si $X(\Omega) = \mathbb{N}^*$ et :*

$$\forall k \in \mathbb{N}^*, \quad \mathbf{P}(X = k) = (1-p)^{k-1} p.$$

Notation : $X \sim$ $\operatorname{Geom}(p)$*.*

La situation typique où intervient la loi géométrique est celle du « temps d'attente du premier succès » dans une suite infinie d'épreuves répétées indépendantes avec même probabilité de succès $p \in]0,1[$. Si X désigne le numéro (aléatoire) de la première épreuve où l'on obtient un succès, on vérifie facilement que $\mathbf{P}(X = k) = (1-p)^{k-1}p$ pour tout $k \in \mathbb{N}^*$. En toute rigueur X est à valeurs dans $\overline{\mathbb{N}^*}$ en attribuant à X la valeur $+\infty$ lorsqu'aucune épreuve de la suite ne donne un succès. On voit facilement que $\mathbf{P}(X = +\infty) = 0$, ce qui permet de considérer X comme une variable à valeurs dans $\mathbb{N}^*$ (en modifiant Ω et $\mathcal{F}$, voir à ce sujet le corrigé du Problème de l'examen de septembre 2003).

Lorsque X suit une loi géométrique, les probabilités $P(X > n)$ ont une expression particulièrement simple en fonction de $q = 1 - p$.

$$\mathbf{P}(X > n) = q^n.$$

Cette formule permet de vérifier facilement la propriété d'« absence de mémoire en temps discret » :

Proposition 2.37. *Si X suit la loi géométrique de paramètre p,*

$$\forall n, k \in \mathbb{N}^*, \quad \mathbf{P}(X > n + k \mid X > n) = \mathbf{P}(X > k). \tag{2.16}$$

La preuve est laissée en exercice, de même que la réciproque : si une variable aléatoire X à valeurs dans $\mathbb{N}^*$ vérifie (2.16), elle suit une loi géométrique.

Lois de Poisson

Définition 2.38. *On dit que la variable aléatoire discrète X suit la loi de Poisson de paramètre $\alpha > 0$ si $X(\Omega) = \mathbb{N}$ et*

$$\forall k \in \mathbb{N}, \quad \mathbf{P}(X = k) = \frac{\mathrm{e}^{-\alpha}\alpha^k}{k!}.$$

Notation : $X \sim \operatorname{Pois}(\alpha)$.

Une des raisons de l'importance de cette loi est le théorème de convergence de la loi binomiale vers la loi de Poisson.

Théorème 2.39. *Si $(p_n)_{n\geq 1}$ est une suite de réels de $[0,1]$ vérifiant*

$$np_n \to \alpha \in]0, +\infty[, \quad \textit{quand} \quad n \to +\infty, \tag{2.17}$$

alors :

$$\forall k \in \mathbb{N}, \quad C_n^k p_n^k (1 - p_n)^{n-k} \longrightarrow \frac{\mathrm{e}^{-\alpha}\alpha^k}{k!}, \quad \textit{quand} \quad n \to +\infty.$$

2.5.3 Lois uniformes

Sur un ensemble infini A, il n'existe pas d'équiprobabilité P. Sinon on pourrait trouver une suite infinie $(x_k)_{k\in\mathbb{N}}$ d'éléments de A tous distincts tels que pour tout $k \in \mathbb{N}$, $P(\{x_k\}) = p > 0$ avec p indépendent de k, d'où pour tout $n \geq 1$, $P(\{x_1, x_2, \dots, x_n\}) = np \leq P(A)$ et en faisant tendre n vers $+\infty$, on trouverait $P(A) = +\infty$, ce qui est absurde. On peut néanmoins définir la loi uniforme sur une partie (borélienne) infinie de $\mathbb{R}$ ou $\mathbb{R}^d$ de façon analogue à (2.11), en remplaçant la mesure de comptage par la mesure de Lebesgue.

Définition 2.40. *Soit A un borélien de $\mathbb{R}^d$ tel que $0 < \lambda_d(A) < +\infty$, λ_d désignant la mesure de Lebesgue sur $\mathbb{R}^d$. Le vecteur aléatoire $X : \Omega \to \mathbb{R}^d$ suit la loi uniforme sur A si*

$$\forall B \in \mathrm{Bor}(\mathbb{R}^d), \quad \mathbf{P}(X \in B) = \frac{\lambda_d(A \cap B)}{\lambda_d(A)}. \tag{2.18}$$

Notation : $X \sim \mathrm{Unif}(A)$.

Le cas $d = 1$ et $A = [0, 1]$ revêt une importance particulière. En effet on peut démontrer que si Y est une variable aléatoire de fonction de répartition F et si $X \sim \mathrm{Unif}([0, 1])$, alors $F^{-1}(X)$ a même loi que Y. Ici F^{-1} désigne l'inverse généralisé de F défini par

$$\forall u \in]0, 1[, \quad F^{-1}(u) = \inf\{t \in \mathbb{R};\ F(t) \geq u\}.$$

Ceci permet de *simuler* n'importe quelle loi sur $\mathbb{R}$ dès que l'on dispose d'un générateur de nombres aléatoires suivant la loi uniforme sur $[0, 1]$. Cette propriété sera vue en T.D.

2.5.4 Lois conditionnelles

Rappelons que si $(\Omega, \mathcal{F}, \mathbf{P})$ est un espace probabilisé et $H \in \mathcal{F}$ un évènement tel que $\mathbf{P}(H) > 0$, on peut définir sur $\mathcal{F}$ une nouvelle *mesure* de probabilité $\mathbf{P}_H = \mathbf{P}(\,.\mid H)$ par

$$\forall B \in \mathcal{F}, \quad \mathbf{P}_H(B) := \mathbf{P}(B \mid H) = \frac{\mathbf{P}(B \cap H)}{\mathbf{P}(H)}.$$

Définition 2.41. *Soient $(\Omega, \mathcal{F}, \mathbf{P})$ un espace probabilisé, $H \in \mathcal{F}$ tel que $\mathbf{P}(H) > 0$, X un vecteur aléatoire à valeurs dans $\mathbb{R}^d$ et défini sur $(\Omega, \mathcal{F})$. On appelle* loi conditionnelle de X sachant H, *la loi de X sous $\mathbf{P}_H$. En la notant $P_{X|H}$, on a donc*

$$\forall B \in \mathrm{Bor}(\mathbb{R}^d), \quad P_{X|H}(B) = \mathbf{P}_H\Big(X^{-1}(B)\Big) = \mathbf{P}(X \in B \mid H).$$

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 mesure dont on munit $(\Omega, \mathcal{F})$ et sous laquelle on considère la loi de X.

Voici un exemple de calcul de loi conditionnelle particulièrement simple. Soit A un borélien de $\mathbb{R}^d$ tel que $0 < \lambda_d(A) < +\infty$ et X un vecteur aléatoire de loi uniforme sur A.

Soit C un borélien inclus dans A et tel que $\lambda_d(C) > 0$. Alors la loi de X sachant $X \in C$ est la loi uniforme sur C. En effet en rappelant (2.18) et en notant que $A \cap C = C$, on peut écrire pour tout $B \in \mathrm{Bor}(\mathbb{R}^d)$:

$$\begin{aligned} P_{X|X\in C}(B) = \mathbf{P}(X \in B \mid X \in C) &= \frac{\mathbf{P}(X \in B \text{ et } X \in C)}{\mathbf{P}(X \in C)} \\ &= \frac{\mathbf{P}(X \in B \cap C)}{\mathbf{P}(X \in C)} \\ &= \frac{\lambda_d(A \cap B \cap C)/\lambda_d(A)}{\lambda_d(A \cap C)/\lambda_d(A)} \\ &= \frac{\lambda_d(B \cap C)}{\lambda_d(C)}. \end{aligned}$$

2.5.5 Un premier exemple de modèle statistique

Terminons avec un exemple simple de modèle statistique où coexistent naturellement plusieurs lois pour *une même* variable aléatoire. On dispose d'une urne contenant des boules rouges et des boules vertes. La proportion θ de boules vertes est inconnue. On se propose de l'estimer en effectuant n tirages avec remise d'une boule et en notant sa couleur. C'est un problème de *sondage*.

On peut utiliser la modélisation suivante. Prenons $\Omega_n = \{\mathrm{r}, \mathrm{v}\}^n$. Un évènement élémentaire ω est donc ici une suite finie de n caractères r ou v codant les résultats des n tirages. Comme Ω_n est fini, on le munit de la tribu $\mathcal{F}_n = \mathcal{P}(\Omega_n)$. Avec cette tribu, n'importe quelle application de Ω_n dans $\mathbb{R}$ est mesurable, donc est une variable aléatoire. Définissons maintenant les n variables aléatoires X_i $(i = 1, \dots, n)$ par

$$X_i(\omega) = \begin{cases} 0 & \text{si la } i\text{-ème composante de } \omega \text{ est r,} \\ 1 & \text{si la } i\text{-ème composante de } \omega \text{ est v.} \end{cases}$$

Définissons aussi

$$S_n := \sum_{i=1}^{n} X_i, \qquad M_n := \frac{1}{n} S_n.$$

L'introduction de ces deux variables est bien naturelle. En effet $S_n(\omega)$ est le nombre de caractères v dans l'évènement élémentaire ω, autrement dit le nombre de boules vertes observées dans la suite de n tirages codée par ω. De même $M_n(\omega)$ est la proportion de boules vertes dans cette même suite de tirages.

Pour l'instant nous avons des variables aléatoires discrètes définies sur $(\Omega_n, \mathcal{F}_n)$, mais pas encore de mesure de probabilité sur cet espace. On ne peut donc pas encore parler de loi pour ces variables. Comme nous avons choisi un Ω_n ne dépendant pas de θ, il est clair que la probabilité dont on va munir $(\Omega_n, \mathcal{F}_n)$ doit dépendre de θ. Malheureusement nous ignorons la valeur de θ, il faut donc se résigner à payer cette ignorance en munissant $(\Omega_n, \mathcal{F}_n)$ non pas d'une mesure de probabilité $\mathbf{P}_\theta$, mais de *toute une famille* $(\mathbf{P}_\theta)_{\theta \in \Theta}$. La famille d'espaces probabilisés $\big(\Omega_n, \mathcal{F}_n, (\mathbf{P}_\theta)_{\theta\in\Theta}\big)$ s'appelle un *modèle statistique*. L'ensemble Θ est l'espace des paramètres du modèle. Dans la situation qui nous intéresse, θ est une proportion, donc un nombre rationnel. On pourra donc prendre $\Theta = \mathbb{Q} \cap [0,1]$ si on ignore le nombre total N de boules dans l'urne ou se restreindre à $\Theta = \{0, 1/N, 2/N, \dots, 1\}$ si on connaît la valeur de N.

Comment définir maintenant $\mathbf{P}_\theta$? Notons d'abord que Ω_n étant fini et muni de $\mathcal{F}_n = \mathcal{P}(\Omega_n)$, $\mathbf{P}_\theta$ sera caractérisée par les $\mathbf{P}_\theta(\{\omega\})$, pour ω décrivant Ω. Arrivés à ce point, nous allons voir qu'il n'y plus qu'un seul choix admissible pour $\mathbf{P}_\theta$. Rappelons en effet que les tirages sont avec remise, donc la composition de l'urne est la même avant chaque nouveau tirage. La probabilité de sortir une boule verte lors du i-ème tirage doit donc être θ. Autrement dit la loi de X_i sous $\mathbf{P}_\theta$ doit être la loi de Bernoulli de paramètre θ. D'autre part les résultats des tirages passés n'influencent pas le tirage à venir puisque l'urne est toujours dans la même composition avant chaque tirage. On modélise ceci par l'indépendance des tirages, autrement dit sur l'espace $(\Omega_n, \mathcal{F}_n, \mathbf{P}_\theta)$, les X_i doivent être des variables de Bernoulli de même loi et indépendantes. Ceci nous conduit à poser :

$$\forall \omega \in \Omega_n, \quad \mathbf{P}_\theta(\{\omega\}) := \theta^{S_n(\omega)}(1-\theta)^{n-S_n(\omega)}.$$

Il est alors facile de voir que sous $\mathbf{P}_\theta$, S_n suit la loi binomiale $\operatorname{Bin}(n,\theta)$. Quant à M_n, elle suit une loi qui a les mêmes masses que $\operatorname{Bin}(n,\theta)$, mais localisées sur les rationnels $0, 1/n, 2/n, \dots 1$, au lieu des entiers $0, 1, 2, \dots, n$. En notant cette loi P_{θ,M_n}, on a

$$P_{\theta,M_n} = \sum_{k=0}^{n} C_n^k \theta^k (1-\theta)^{n-k} \delta_{k/n}.$$

Quand n est grand, cette loi est bien concentrée dans un petit voisinage de θ, voir les diagrammes en bâtons[6] de la figure 2.5 pour une illustration. Ceci permet de proposer une estimation du paramètre inconnu θ par un *intervalle de confiance* construit à partir de la valeur observée $M_n(\omega)$ (voir [ICP] chapitre 6 pour les détails).

On peut aussi proposer une *estimation ponctuelle* de θ en utilisant la convergence presque sûre de M_n vers θ (loi forte des grands nombres, voir [ICP] chapitre 6). On estime alors θ par la valeur observée $M_n(\omega)$. Cette approche suppose que l'on remplace l'espace mesurable $(\Omega_n, \mathcal{F}_n)$ par un espace $(\Omega, \mathcal{F})$ ne dépendant pas de n et assez « riche » pour supporter une suite infinie $(X_i)_{i\geq 1}$ de variables aléatoires de Bernoulli qui soient sous chaque $\mathbf{P}_\theta$, indépendantes et de même loi $\operatorname{Bern}(\theta)$. Ce problème est celui de la modélisation du jeu de pile ou face infini[7]. La difficulté est exactement la même que celle que nous avons rencontrée pour construire la mesure de Lebesgue sur $\mathbb{R}$.

6. Pour représenter graphiquement la loi discrète $\mu = \sum p_k \delta_{x_k}$, on trace à partir de chaque point $(x_k, 0)$ un segment vertical de hauteur proportionnelle à p_k. La figure 2.5 affiche théoriquement 101 bâtons pour chacune des deux lois de M_{100}, sous $\mathbf{P}_{0,15}$ et sous $\mathbf{P}_{0,7}$. En pratique seuls sont visibles ceux dont la hauteur est supérieure à l'épaisseur d'un trait d'imprimante.

7. Voir les Annales d'IFP 2002-2003, D.M. n° 2.

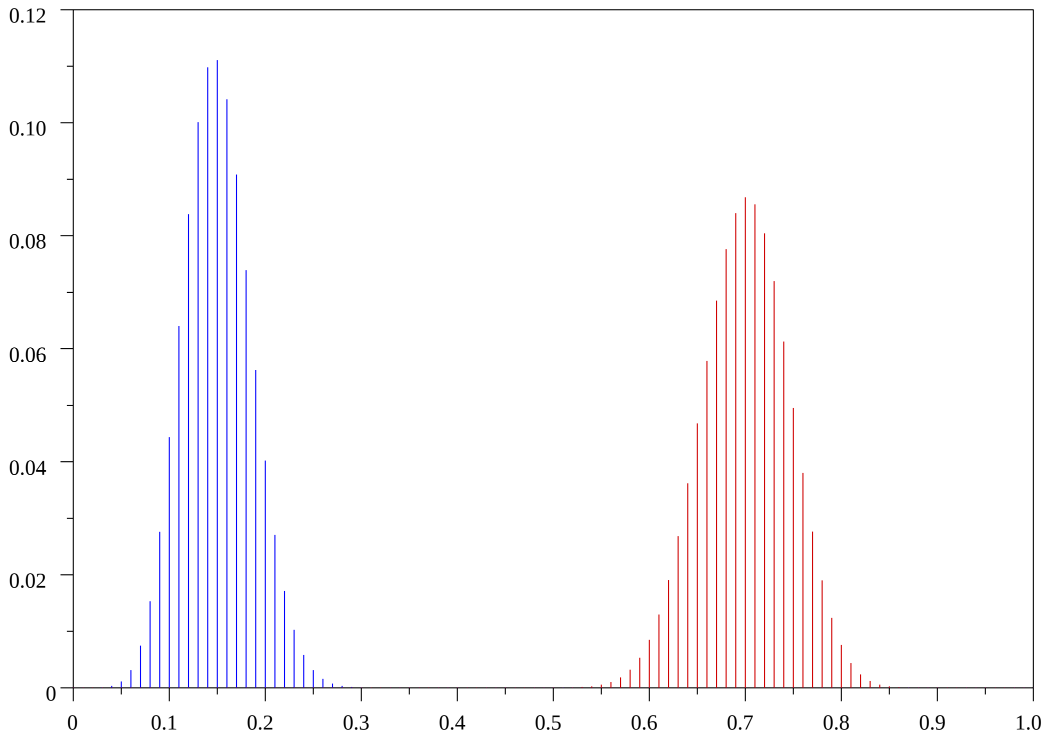


FIGURE 2.5 – Loi de M_{100} sous $\mathbf{P}_{0,15}$ et sous $\mathbf{P}_{0,7}$

Chapitre 3

Intégration des fonctions mesurables positives

Dans tout ce chapitre, $(\Omega, \mathcal{F}, \mu)$ désigne un espace mesuré. Notre but est de construire et d'étudier l'intégrale sur Ω par rapport à μ, d'abord pour les fonctions étagées positives, puis par extension pour toutes les applications $f : \Omega \to \overline{\mathbb{R}}_+$, mesurables $\mathcal{F}$-Bor$(\overline{\mathbb{R}}_+)$.

3.1 Intégration des fonctions étagées positives

On note $\mathcal{E}_+$ l'ensemble des fonctions étagées $\Omega \to \mathbb{R}_+$, mesurables $\mathcal{F}$-Bor$(\mathbb{R}_+)$.

Définition 3.1. *Soit $u \in \mathcal{E}_+$, de décomposition canonique $u = \sum_{i=1}^{n} y_i \mathbf{1}_{A_i}$. On pose*

$$\int_\Omega u \,\mathrm{d}\mu := \sum_{i=1}^{n} y_i \mu(A_i). \tag{3.1}$$

Cette quantité s'appelle intégrale sur Ω de u par rapport à la mesure μ.

L'intégrale $\int_\Omega u \,\mathrm{d}\mu$ est un élément de $\overline{\mathbb{R}}_+$. Bien que les y_i soient tous réels (d'après la définition des fonctions étagées), elle peut très bien valoir $+\infty$, si pour un $y_i > 0$, le $\mu(A_i)$ correspondant vaut $+\infty$. Rappelons la convention « $0 \times (+\infty) := 0$ » qui est bien utile ici lorsque $\mu\Big(u^{-1}(\{0\})\Big) = +\infty$.

Exemple 3.1. Si u est une fonction constante, alors sa décomposition canonique s'écrit $u = c\mathbf{1}_\Omega$, avec $c \in \mathbb{R}_+$. La formule (3.1) nous donne alors

$$\int_\Omega c \,\mathrm{d}\mu = c\mu(\Omega).$$

Exemple 3.2. Si $u = c\mathbf{1}_A$, avec $c \in]0, +\infty[$, $A \in \mathcal{F}$ et $A \neq \Omega$, sa décomposition canonique est $u = c\mathbf{1}_A + 0 \times \mathbf{1}_{A^c}$ d'où

$$\int_\Omega c\mathbf{1}_A \,\mathrm{d}\mu = c\mu(A) + 0 \times \mu(A^c) = c\mu(A).$$

Ce résultat reste vrai pour $c = 0$ par l'exemple 3.1 (la décomposition canonique est alors $u = 0 \times \mathbf{1}_\Omega$). Le cas particulier $c = 1$ est d'une grande utilité car il permet d'exprimer la mesure d'un ensemble sous la forme d'une intégrale :

$$\forall A \in \mathcal{F}, \quad \int_\Omega \mathbf{1}_A \,\mathrm{d}\mu = \mu(A). \tag{3.2}$$

Le lemme suivant donne plus de souplesse dans le calcul de l'intégrale d'une fonction étagée en évitant le recours systématique à la décomposition canonique. Il sera utile pour prouver l'additivité de l'intégrale sur $\mathcal{E}_+$.

Lemme 3.2. *Si $u \in \mathcal{E}_+$ s'écrit $u = \sum_{k=1}^{r} t_k \mathbf{1}_{C_k}$, les C_k appartenant tous à $\mathcal{F}$ et étant deux à deux disjoints, les réels t_k n'étant pas forcément tous distincts,*

$$\int_\Omega u \,\mathrm{d}\mu = \sum_{k=1}^{r} t_k \mu(C_k).$$

Preuve. En notant $u(\Omega) = \{y_1, \dots, y_n\}$, les y_i étant tous distincts et $A_i := u^{-1}(\{y_i\})$, la décomposition canonique est

$$u = \sum_{i=1}^{n} y_i \mathbf{1}_{A_i}$$

Comme les C_k sont deux à deux disjoints, u est constante de valeur t_k sur chaque C_k *non vide*. Chaque C_k non vide est donc inclus dans un A_i, car sinon C_k aurait au moins un élément ω dans un A_i et un élément ω' dans un A_j avec $i \neq j$, d'où $u(\omega) = y_i \neq y_j = u(\omega')$, ce qui contredirait la constance de u sur C_k. Notons

$$K(i) := \Big\{ k \in \{1, \dots, r\};\ \emptyset \neq C_k \subset A_i \Big\}, \quad i = 1, \dots, n.$$

Si $k \in K(i)$, C_k n'étant pas vide a au moins un élément ω qui nécessairement est aussi dans A_i, d'où $t_k = u(\omega) = y_i$. Ainsi

$$\forall i = 1, \dots, n,\ \forall k \in K(i), \quad t_k = y_i. \tag{3.3}$$

Par construction de $K(i)$, on a clairement

$$\forall i = 1, \dots, n, \quad \bigcup_{k \in K(i)} C_k \subset A_i.$$

A-t-on l'égalité ? Supposons que l'ensemble $A_i \setminus \bigcup_{k \in K(i)} C_k$ ait au moins un élément ω. D'une part $\omega \in A_i$ donc $u(\omega) = y_i$. D'autre part ω n'appartient à aucun des C_k indexés par $K(i)$, ni à aucun des C_k indexés par un autre $K(j)$ pour $j \neq i$ puisqu'alors il serait dans A_j qui est disjoint de A_i. Ainsi ω n'appartient à aucun des C_k, puisqu'on vient de les éliminer tous… sauf ceux qui sont vides ! Alors ou bien le complémentaire dans Ω de $\bigcup_{1 \leq k \leq r} C_k$ est vide et on a une contradiction, ou bien ω appartient à ce complémentaire et dans ce cas, $u(\omega) = 0$ puisque tous les $\mathbf{1}_{C_k}(\omega)$ sont nuls. Là aussi on a une contradiction, sauf si $y_i = 0$. Nous venons donc de vérifier que :

$$\text{si } y_i \neq 0, \quad A_i \setminus \bigcup_{k \in K(i)} C_k = \emptyset, \quad \text{d'où } A_i = \bigcup_{k \in K(i)} C_k.$$

Les C_k indexés par $K(i)$ formant ainsi une partition finie de A_i (pour $y_i \neq 0$), on en déduit par additivité de la mesure μ et (3.3) :

$$\text{si } y_i \neq 0, \quad y_i \mu(A_i) = y_i \sum_{k \in K(i)} \mu(C_k) = \sum_{k \in K(i)} y_i \mu(C_k) = \sum_{k \in K(i)} t_k \mu(C_k).$$

Finalement,

$$\int_\Omega u \, \mathrm{d}\mu = \sum_{i=1}^n y_i \mu(A_i) = \sum_{\substack{i=1 \\ y_i \neq 0}}^n y_i \mu(A_i) = \sum_{\substack{i=1 \\ y_i \neq 0}}^n \sum_{k \in K(i)} t_k \mu(C_k) = \sum_{\substack{k=1 \\ t_k \neq 0 \\ C_k \neq \emptyset}}^r t_k \mu(C_k) = \sum_{k=1}^r t_k \mu(C_k).$$

□

Proposition 3.3. *L'intégrale sur $\mathcal{E}_+$ est homogène, additive et croissante : pour tous $u, v \in \mathcal{E}_+$, $c \in \mathbb{R}_+$,*

i) $\displaystyle\int_\Omega cu \, \mathrm{d}\mu = c \int_\Omega u \, \mathrm{d}\mu$ *;*

ii) $\displaystyle\int_\Omega (u + v) \, \mathrm{d}\mu = \int_\Omega u \, \mathrm{d}\mu + \int_\Omega v \, \mathrm{d}\mu$ *;*

iii) si $u \leq v$, $\displaystyle\int_\Omega u \, \mathrm{d}\mu \leq \int_\Omega v \, \mathrm{d}\mu$.

Preuve. D'abord, il est clair que pour $u, v \in \mathcal{E}_+$ et $c \in \mathbb{R}_+$, les fonctions cu et $u + v$ appartiennent à $\mathcal{E}_+$. Notons $u = \sum_{i=1}^m y_i \mathbf{1}_{A_i}$ et $v = \sum_{j=1}^n z_j \mathbf{1}_{B_j}$ les décompositions canoniques respectives de u et v. Ainsi $\{A_i;\ 1 \leq i \leq m\}$ et $\{B_j;\ 1 \leq j \leq n\}$ sont des partitions de Ω.

L'homogénéité i) est évidente car la décomposition canonique de cu est $\sum_{i=1}^m c y_i \mathbf{1}_{A_i}$ si $c > 0$ et $0 \times \mathbf{1}_\Omega$ si $c = 0$. Dans ce cas particulier, il convient de remarquer que si $\int_\Omega u \, \mathrm{d}\mu = +\infty$, on a encore $0 \times \int_\Omega u \, \mathrm{d}\mu = 0 = \int_\Omega (0 \times u) \, \mathrm{d}\mu$ grâce à la convention « $0 \times (+\infty) = 0$ ».

Pour vérifier l'additivité ii), posons

$$C_{i,j} := A_i \cap B_j, \quad 1 \leq i \leq m,\ 1 \leq j \leq n.$$

Ces ensembles sont deux à deux disjoints, certains pouvant être vides. On voit que

$$\forall i = 1, \dots, m, \quad A_i = \bigcup_{j=1}^n C_{i,j} \quad \text{et} \quad \forall j = 1, \dots, n, \quad B_j = \bigcup_{i=1}^m C_{i,j}. \tag{3.4}$$

Vérifions le pour la première égalité :

$$\bigcup_{j=1}^n C_{i,j} = \bigcup_{j=1}^n A_i \cap B_j = A_i \cap \left(\bigcup_{j=1}^n B_j \right) = A_i \cap \Omega = A_i.$$

Sur chaque $C_{i,j}$ non vide, la fonction $u+v$ est constante et vaut y_i+z_j. Les $C_{i,j}$ sont deux à deux disjoints et la réunion de *tous* les $C_{i,j}$ est Ω. Enfin si $C_{i,j}=\emptyset$, $(y_i+z_j)\mathbf{1}_{C_{i,j}}$ est la fonction identiquement nulle. Ces remarques justifient l'écriture

$$u+v=\sum_{i=1}^{m}\sum_{j=1}^{n}(y_i+z_j)\mathbf{1}_{C_{i,j}}.$$

Cette écriture n'est pas forcément la décomposition canonique de $u+v$ car certains des $C_{i,j}$ peuvent être vides et les (y_i+z_j) ne sont pas nécessairement tous distincts. C'est ici qu'intervient le lemme 3.2 qui légitime la première des égalités suivantes :

$$\begin{aligned}
\int_\Omega (u+v)\,\mathrm{d}\mu &= \sum_{i=1}^{m}\sum_{j=1}^{n}(y_i+z_j)\mu(C_{i,j}) \\
&= \sum_{i=1}^{m}\sum_{j=1}^{n} y_i\mu(C_{i,j}) + \sum_{i=1}^{m}\sum_{j=1}^{n} z_j\mu(C_{i,j}) \\
&= \sum_{i=1}^{m} y_i \sum_{j=1}^{n}\mu(C_{i,j}) + \sum_{j=1}^{n} z_j\sum_{i=1}^{m}\mu(C_{i,j}) \\
&= \sum_{i=1}^{m} y_i\mu(A_i) + \sum_{j=1}^{n} z_j\mu(B_j) \qquad (3.5)\\
&= \int_\Omega u\,\mathrm{d}\mu + \int_\Omega v\,\mathrm{d}\mu,
\end{aligned}$$

l'égalité (3.5) découlant de (3.4) et de l'additivité de la mesure μ.

La croissance iii) résulte de l'additivité. En effet rappelons que $u\leq v$ signifie que pour tout $\omega\in\Omega$, $u(\omega)\leq v(\omega)$ et comme u et v sont étagées, donc à valeurs *réelles*, $v(\omega)-u(\omega)$ est bien défini et $v(\omega)-u(\omega)\geq 0$. Alors $v-u\in\mathcal{E}_+$, et $v=u+(v-u)$. Par ii), on a

$$\int_\Omega v\,\mathrm{d}\mu=\int_\Omega u\,\mathrm{d}\mu+\int_\Omega (v-u)\,\mathrm{d}\mu\geq\int_\Omega u\,\mathrm{d}\mu,$$

puisque $\int_\Omega (v-u)\,\mathrm{d}\mu\in\overline{\mathbb{R}}_+$. □

Lemme 3.4. *Soit $u\in\mathcal{E}_+$. La fonction d'ensembles*

$$\nu:\mathcal{F}\to\overline{\mathbb{R}}_+,\quad B\mapsto\nu(B)=\int_\Omega u\mathbf{1}_B\,\mathrm{d}\mu=:\int_B u\,\mathrm{d}\mu$$

est une mesure sur $(\Omega,\mathcal{F})$.

Preuve. Si $B\in\mathcal{F}$ et $u\in\mathcal{E}_+$, la fonction $u\mathbf{1}_B$ appartient elle aussi à $\mathcal{E}_+$ et en notant $u=\sum_{i=1}^n y_i\mathbf{1}_{A_i}$ la décomposition canonique de u, on a

$$u\mathbf{1}_B=\sum_{i=1}^{n}y_i(\mathbf{1}_{A_i}\cdot\mathbf{1}_B)=\sum_{i=1}^{n}y_i\mathbf{1}_{A_i\cap B}.$$

Les $A_i\cap B$ étant deux à deux disjoints, le lemme 3.2 nous donne

$$\forall B\in\mathcal{F},\quad \int_\Omega u\mathbf{1}_B\,\mathrm{d}\mu=\sum_{i=1}^{n}y_i\mu(A_i\cap B).$$

Cette égalité s'écrit aussi $\nu = \sum_{i=1}^{n} y_i\nu_i$, où les fonctions d'ensembles ν_i sont définies par

$$\nu_i : \mathcal{F} \to \overline{\mathbb{R}}_+, \quad B \mapsto \nu_i(B) := \mu(A_i \cap B).$$

On voit ainsi que ν_i est la mesure trace de μ sur A_i (cf. Chapitre 1, exemple 1.4). Pour chaque i, $y_i\nu_i$ est donc aussi une mesure (si $y_i = 0$, c'est la mesure nulle, même si ν_i n'est pas finie, en raison de la convention « $0 \times (+\infty) := 0$ »). Une somme finie de mesures est clairement une mesure (la somme hérite de la σ-additivité de chacun des termes par la formule de sommation par paquets dans $\overline{\mathbb{R}}_+$). La fonction d'ensembles ν est donc bien une mesure sur $\mathcal{F}$. □

3.2 Intégrale dans $\mathcal{M}_+$

Nous notons $\mathcal{M}_+$ l'ensemble des applications $f : \Omega \to \overline{\mathbb{R}}_+$, mesurables $\mathcal{F}$-Bor$(\overline{\mathbb{R}}_+)$. Grâce au corollaire 2.18, $\mathcal{E}_+$ est inclus dans $\mathcal{M}_+$.

Définition 3.5. *Pour $f \in \mathcal{M}_+$, on appelle intégrale sur Ω de f par rapport à μ l'élément de $\overline{\mathbb{R}}_+$ noté $\int_\Omega f \,\mathrm{d}\mu$ et défini par*

$$\int_\Omega f \,\mathrm{d}\mu := \sup\left\{\int_\Omega u \,\mathrm{d}\mu;\ u \in \mathcal{E}_+,\ u \leq f\right\}. \tag{3.6}$$

Pour $A \in \mathcal{F}$, $f\mathbf{1}_A$ appartient encore à $\mathcal{M}_+$ et on définit aussi

$$\int_A f \,\mathrm{d}\mu := \int_\Omega f\mathbf{1}_A \,\mathrm{d}\mu. \tag{3.7}$$

Cette définition pose deux problèmes de cohérence, résolus par les deux remarques suivantes.

Remarque 3.6. Si $f \in \mathcal{E}_+$, les deux définitions de son intégrale par (3.1) et (3.6) coïncident. Pour le voir, notons provisoirement $\int_\Omega^{\text{étag}} f \,\mathrm{d}\mu$ l'intégrale de f au sens de (3.1) et $\int_\Omega^{\text{mes+}} f \,\mathrm{d}\mu$ celle au sens de (3.6). Pour toute $u \in \mathcal{E}_+$ telle que $u \leq f$, on a par croissance de l'intégrale des fonctions étagées (cf. proposition 3.3 iii) l'inégalité $\int_\Omega^{\text{étag}} u \,\mathrm{d}\mu \leq \int_\Omega^{\text{étag}} f \,\mathrm{d}\mu$. Par conséquent dans (3.6), le supremum est *atteint* pour $u = f$, ce qui implique $\int_\Omega^{\text{mes+}} f \,\mathrm{d}\mu = \int_\Omega^{\text{étag}} f \,\mathrm{d}\mu$.

Remarque 3.7. Pour s'assurer de la cohérence de la définition de $\int_A f \,\mathrm{d}\mu$ par (3.7), le lecteur résoudra avec profit l'exercice suivant.
Soit $A \in \mathcal{F}$, on note $\mathcal{F}_A := \{A \cap B;\ B \in \mathcal{F}\}$.

a) *Vérifier que $\mathcal{F}_A$ est une tribu sur A (tribu trace). Par construction $\mathcal{F}_A \subset \mathcal{F}$, donc la restriction μ_A de μ à $\mathcal{F}_A$ est bien définie (noter que $\mathcal{F}_A$ n'est pas une* sous-tribu *de $\mathcal{F}$). Vérifier que μ_A est une mesure sur $(A, \mathcal{F}_A)$.*
b) *Montrer que pour toute $f \in \mathcal{M}_+(\mathcal{F})$, sa restriction f_A à A appartient à $\mathcal{M}_+(\mathcal{F}_A)$ et que*

$$\int_A f_A \,\mathrm{d}\mu_A = \int_\Omega f\mathbf{1}_A \,\mathrm{d}\mu,$$

la première intégrale étant comprise au sens de (3.6), appliquée à l'espace mesuré $(A, \mathcal{F}_A, \mu_A)$.

Dans le langage de la théorie des probabilités, les notations traditionnelles sont un peu différentes de celles utilisées ci-dessus.

Définition 3.8. *Soit* $(\Omega, \mathcal{F})$ *un espace probabilisable. On appelle* variable aléatoire positive *sur* $(\Omega, \mathcal{F})$ *toute application* $X : \Omega \to \overline{\mathbb{R}}_+$*, mesurable* $\mathcal{F}$-Bor$(\overline{\mathbb{R}}_+)$*. Si* $\mathbf{P}$ *est une mesure de probabilité sur* $(\Omega, \mathcal{F})$*, on appelle* espérance *de* X *sous* $\mathbf{P}$*, l'intégrale*

$$\mathbf{E}X := \int_\Omega X \, \mathrm{d}\mathbf{P},$$

au sens de (3.6).

Notons au passage qu'une variable aléatoire positive n'est pas nécessairement une variable aléatoire réelle (elle l'est si $+\infty \notin X(\Omega)$, vérification laissée en exercice).

Proposition 3.9 (Croissance de l'intégrale)**.** *Pour toutes* $f, g \in \mathcal{M}_+$*,*

$$f \leq g \quad \Rightarrow \quad \int_\Omega f \, \mathrm{d}\mu \leq \int_\Omega g \, \mathrm{d}\mu. \tag{3.8}$$

Preuve. C'est une conséquence immédiate de la définition 3.5 et de l'inclusion

$$\{u \in \mathcal{E}_+;\ u \leq f\} \subset \{u \in \mathcal{E}_+;\ u \leq g\}.$$

□

Nous abordons maintenant le premier des grands théorèmes d'interversion limite intégrale « $\int \lim = \lim \int$ » qui font toute la puissance de la théorie de l'intégration au sens de Lebesgue.

Théorème 3.10 (de convergence monotone ou de Beppo Levi)**.** *Soit* $(f_n)_{n\in\mathbb{N}}$ *une suite* croissante *dans* $\mathcal{M}_+$*. Alors* $f := \sup_{n\in\mathbb{N}} f_n = \lim_{n\to+\infty} f_n$ *est aussi dans* $\mathcal{M}_+$ *et*

$$\int_\Omega f \, \mathrm{d}\mu = \lim_{n\to+\infty} \int_\Omega f_n \, \mathrm{d}\mu = \sup_{n\in\mathbb{N}} \int_\Omega f_n \, \mathrm{d}\mu. \tag{3.9}$$

Preuve. L'appartenance de f à $\mathcal{M}_+$ a déjà été vue (proposition 2.19). Par croissance de l'intégrale (prop. 3.9), la suite $\left(\int_\Omega f_n \, \mathrm{d}\mu\right)_{n\in\mathbb{N}}$ est croissante dans $\overline{\mathbb{R}}_+$, donc convergente dans $\overline{\mathbb{R}}_+$ vers une limite que nous noterons L :

$$L := \lim_{n\to+\infty} \int_\Omega f_n \, \mathrm{d}\mu = \sup_{n\in\mathbb{N}} \int_\Omega f_n \, \mathrm{d}\mu. \tag{3.10}$$

L'inégalité $f_n \leq f$, vraie pour tout $n \in \mathbb{N}$, implique par croissance de l'intégrale, $\int_\Omega f_n \, \mathrm{d}\mu \leq \int_\Omega f \, \mathrm{d}\mu$, puis en prenant le supremum sur n (ou par passage à la limite) :

$$L \leq \int_\Omega f \, \mathrm{d}\mu. \tag{3.11}$$

Pour obtenir l'inégalité inverse et achever ainsi la preuve de (3.9), il suffit de montrer que :

$$\forall u \in \mathcal{E}_+, \qquad u \leq f \quad \Rightarrow \quad \int_\Omega u \, \mathrm{d}\mu \leq L, \tag{3.12}$$

en effet en prenant dans (3.12) le supremum sur les u majorées par f il vient :

$$\int_\Omega f \, \mathrm{d}\mu = \sup\left\{ \int_\Omega u \, \mathrm{d}\mu;\ u \leq f,\ u \in \mathcal{E}_+ \right\} \leq L. \tag{3.13}$$

Preuve de (3.12). Soit $u \in \mathcal{E}_+$, $u \leq f$. Fixons $p \in]0,1[$ et définissons :

$$A_n := \Big\{ \omega \in \Omega;\ pu(\omega) \leq f_n(\omega) \Big\}.$$

Grâce à la proposition 2.17, on sait que A_n appartient à la tribu $\mathcal{F}$. La suite $(A_n)_{n\in\mathbb{N}}$ est *croissante pour l'inclusion* car si $\omega \in A_n$, alors $pu(\omega) \leq f_n(\omega) \leq f_{n+1}(\omega)$ par croissance de $(f_n)_{n\in\mathbb{N}}$, donc $\omega \in A_{n+1}$. La réunion de la suite $(A_n)_{n\in\mathbb{N}}$ est égale à Ω. En effet pour tout $\omega \in \Omega$,

– ou bien $0 < f(\omega) \leq +\infty$, alors comme $u(\omega) \leq f(\omega)$ et $p < 1$, on a $pu(\omega) < f(\omega)$; or $f_n(\omega)$ converge vers $f(\omega)$, il existe donc un $n_0 \in \mathbb{N}$ tel que $f_{n_0}(\omega) > pu(\omega)$, d'où $\omega \in A_{n_0}$ et $\omega \in \bigcup_{n\in\mathbb{N}} A_n$;
– ou bien $f(\omega) = 0$ alors d'une part l'inégalité $u \leq f$ implique $u(\omega) = 0$ d'où $pu(\omega) = 0$ et d'autre part les inégalités $0 \leq f_n \leq f$ impliquent pour tout n, $f_n(\omega) = 0$; dans ce cas, ω appartient à tous les A_n donc *a fortiori* à leur réunion.

Nous venons ainsi de vérifier que :

$$A_n \uparrow \Omega, \quad \text{dans } \mathcal{F}. \tag{3.14}$$

Nous disposons des inégalités suivantes :

$$\forall \omega \in \Omega, \quad pu(\omega)\mathbf{1}_{A_n}(\omega) \leq f_n(\omega)\mathbf{1}_{A_n}(\omega) \leq f_n(\omega), \tag{3.15}$$

en effet si $\omega \in A_n$, $\mathbf{1}_{A_n}(\omega) = 1$ et la première inégalité résulte de la définition de A_n, si $\omega \in A_n^c$, $\mathbf{1}_{A_n}(\omega) = 0$ et (3.15) se réduit à $0 \leq 0 \leq f_n(\omega)$. Réécrivant (3.15) sous la forme $pu\mathbf{1}_{A_n} \leq f_n\mathbf{1}_{A_n} \leq f_n$ d'une inégalité entre *fonctions* appartenant à $\mathcal{M}_+$, on en déduit par croissance de l'intégrale :

$$\int_{A_n} pu \, \mathrm{d}\mu = \int_\Omega pu\mathbf{1}_{A_n} \, \mathrm{d}\mu \leq \int_\Omega f_n\mathbf{1}_{A_n} \, \mathrm{d}\mu \leq \int_\Omega f_n \, \mathrm{d}\mu \leq L.$$

Compte-tenu de l'homogénéité de l'intégrale sur $\mathcal{E}_+$ (noter que $u\mathbf{1}_{A_n}$ est étagée), nous retenons de ces inégalités vraies pour tout n, que

$$\forall n \in \mathbb{N}, \qquad p\int_{A_n} u \, \mathrm{d}\mu \leq L. \tag{3.16}$$

Cette inégalité peut encore s'écrire $p\nu(A_n) \leq L$, en introduisant la fonction d'ensembles $\nu : B \mapsto \nu(B) := \int_B u \, \mathrm{d}\mu$. Par le lemme 3.4, ν est une *mesure* sur $(\Omega, \mathcal{F})$. Sa continuité croissante séquentielle et (3.14) nous donnent alors la convergence $\nu(A_n) \uparrow \nu(\Omega) = \int_\Omega u \, \mathrm{d}\mu$. Faisant tendre n vers l'infini dans (3.16) on obtient ainsi :

$$p \int_\Omega u \, \mathrm{d}\mu \leq L. \tag{3.17}$$

Jusqu'ici, nous avons travaillé avec p *fixé*, mais quelconque dans $]0,1[$. L'ensemble A_n dépendait de p, mais nous venons de l'éliminer. L'inégalité (3.17) est donc vraie *pour tout* $p \in]0,1[$ et on peut y faire tendre p vers 1 (par valeurs inférieures) pour obtenir finalement

$$\int_\Omega u \, \mathrm{d}\mu \leq L.$$

Comme les seules hypothèses faites sur u pour aboutir à ce résultat étaient $u \in \mathcal{E}_+$ et $u \leq f$, ceci établit (3.12) et achève la preuve du théorème. □

Corollaire 3.11 (homogénéité et additivité de l'intégrale dans $\mathcal{M}_+$)**.** *Pour toutes fonctions $f, g \in \mathcal{M}_+$ et toute constante $c \in \overline{\mathbb{R}}_+$,*

a) $\displaystyle\int_\Omega cf \, \mathrm{d}\mu = c \int_\Omega f \, \mathrm{d}\mu$ *;*

b) $\displaystyle\int_\Omega (f+g) \, \mathrm{d}\mu = \int_\Omega f \, \mathrm{d}\mu + \int_\Omega g \, \mathrm{d}\mu.$

Preuve. Le cas $c < +\infty$ dans le a) est une conséquence immédiate de la définition 3.5 et de la proposition 3.3 i) et ne requiert pas le théorème de Beppo Levi. Pour traiter le cas $c = +\infty$, posons $f_n := nf$. La suite $(f_n)_{n \in \mathbb{N}}$ est croissante dans $\mathcal{M}_+$. Sa limite est cf. En effet, si $f(\omega) = 0$, $cf(\omega) = (+\infty) \times 0 = 0$ et $f_n(\omega) = nf(\omega) = 0$ converge bien vers 0. Si $f(\omega) \in]0, +\infty]$, $cf(\omega) = (+\infty) \times f(\omega) = +\infty$ et $f_n(\omega) = nf(\omega)$ tend vers $+\infty$ puisque $f(\omega) > 0$. Par le théorème de Beppo Levi, $\int_\Omega cf \, \mathrm{d}\mu = \int_\Omega \lim_{n\to+\infty} f_n \, \mathrm{d}\mu = \lim_{n\to+\infty} \int_\Omega f_n \, \mathrm{d}\mu$. Par le cas $c < +\infty$ (avec $c = n$), $\int_\Omega f_n \, \mathrm{d}\mu = n \int_\Omega f \, \mathrm{d}\mu$. Si $\int_\Omega f \, \mathrm{d}\mu > 0$, la limite de $n \int_\Omega f \, \mathrm{d}\mu$ est $+\infty$. Si $\int_\Omega f \, \mathrm{d}\mu = 0$, cette limite vaut 0. Dans les deux cas, cette limite s'écrit $(+\infty) \times \int_\Omega f \, \mathrm{d}\mu$ grâce à nos conventions sur la multiplication dans $\overline{\mathbb{R}}_+$. Finalement on a bien $\int_\Omega (+\infty) \times f \, \mathrm{d}\mu = (+\infty) \times \int_\Omega f \, \mathrm{d}\mu$.

Pour prouver le b), le théorème 2.23 nous fournit deux suites croissantes (u_n) et (v_n) dans $\mathcal{E}_+$ convergentes vers f et g respectivement. Clairement la suite $(u_n + v_n)$ est alors croissante dans $\mathcal{E}_+$ et converge vers $f + g$. Par additivité de l'intégrale des fonctions étagées,

$$\forall n \in \mathbb{N}, \qquad \int_\Omega (u_n + v_n) \, \mathrm{d}\mu = \int_\Omega u_n \, \mathrm{d}\mu + \int_\Omega v_n \, \mathrm{d}\mu. \tag{3.18}$$

Une triple application du théorème de Beppo Levi nous donne quand n tend vers l'infini :

$$\int_\Omega (u_n + v_n) \, \mathrm{d}\mu \uparrow \int_\Omega (f+g) \, \mathrm{d}\mu, \quad \int_\Omega u_n \, \mathrm{d}\mu \uparrow \int_\Omega f \, \mathrm{d}\mu, \quad \int_\Omega v_n \, \mathrm{d}\mu \uparrow \int_\Omega g \, \mathrm{d}\mu,$$

ce qui permet de conclure par passage à la limite dans (3.18). □

Corollaire 3.12 (Interversion série-intégrale dans $\mathcal{M}_+$). *Soit $(f_k)_{k\in\mathbb{N}}$ une suite dans $\mathcal{M}_+$. La fonction $\sum_{k=0}^{+\infty} f_k$ est aussi dans $\mathcal{M}_+$ et*

$$\int_\Omega \left(\sum_{k=0}^{+\infty} f_k\right) \mathrm{d}\mu = \sum_{k=0}^{+\infty} \int_\Omega f_k \, \mathrm{d}\mu \qquad \textit{(égalité dans } \overline{\mathbb{R}}_+\textit{)}. \tag{3.19}$$

Preuve. Posons

$$s_n := \sum_{k=0}^{n} f_k, \quad s := \sum_{k=0}^{+\infty} f_k.$$

L'application s_n est mesurable positive comme somme d'un nombre fini d'applications mesurables positives. La suite (s_n) converge en croissant (dans $\overline{\mathbb{R}}_+$) vers s. L'additivité de l'intégrale vue au corollaire 3.11 b) pour la somme de deux éléments de $\mathcal{M}_+$ s'étend immédiatement à la somme d'un nombre *fini* quelconque de termes dans $\mathcal{M}_+$, d'où

$$\forall n \geq 1, \qquad \int_\Omega s_n \, \mathrm{d}\mu = \sum_{k=0}^{n} \int_\Omega f_k \, \mathrm{d}\mu. \tag{3.20}$$

Par le théorème de Beppo Levi, $s \in \mathcal{M}_+$ et $\int_\Omega s_n \, \mathrm{d}\mu$ converge en croissant dans $\overline{\mathbb{R}}_+$ vers $\int_\Omega s \, \mathrm{d}\mu$. Ainsi le premier membre de (3.20) converge vers celui de (3.19). D'autre part, $\int_\Omega f_k \, \mathrm{d}\mu$ est dans $\overline{\mathbb{R}}_+$ donc positif. Par conséquent la suite $\left(\sum_{k=0}^{n} \int_\Omega f_k \, \mathrm{d}\mu\right)_{n\in\mathbb{N}}$ est croissante donc convergente dans $\overline{\mathbb{R}}_+$ et sa limite est $\sum_{k=0}^{+\infty} \int_\Omega f_k \, \mathrm{d}\mu$. Le second membre de (3.20) converge ainsi vers celui de (3.19). L'égalité (3.20) étant vraie pour tout $n \geq 1$ nous donne donc l'égalité (3.19) par passage à la limite. □

Corollaire 3.13 (Lemme de Fatou). *Si $(f_n)_{n\in\mathbb{N}}$ est une suite dans $\mathcal{M}_+$,*

$$\int_\Omega \liminf_{n\to+\infty} f_n \, \mathrm{d}\mu \leq \liminf_{n\to+\infty} \int_\Omega f_n \, \mathrm{d}\mu. \tag{3.21}$$

Preuve. Posons $g := \liminf_{n\to+\infty} f_n$. Par définition de la limite inférieure,

$$g = \lim_{n\to+\infty} \inf_{k\geq n} f_k = \sup_{n\in\mathbb{N}} \inf_{k\geq n} f_k.$$

La fonction $g_n := \inf_{k\geq n} f_k$ appartient à $\mathcal{M}_+$ (cf. corollaire 2.20 i) et la suite $(g_n)_{n\in\mathbb{N}}$ converge en croissant vers g. Par le théorème de Beppo Levi, on a donc

$$\int_\Omega g_n \, \mathrm{d}\mu \uparrow \int_\Omega g \, \mathrm{d}\mu = \int_\Omega \liminf_{n\to+\infty} f_n \, \mathrm{d}\mu. \tag{3.22}$$

D'autre part, on a clairement pour tout $n \in \mathbb{N}$, $g_n \leq f_n$, d'où par croissance de l'intégrale,

$$\forall n \in \mathbb{N}, \qquad \int_\Omega g_n \, \mathrm{d}\mu \leq \int_\Omega f_n \, \mathrm{d}\mu. \tag{3.23}$$

Nous ne savons pas si le second membre de (3.23) a une limite quand n tend vers l'infini, mais par contre sa limite inférieure existe toujours. On peut ainsi passer à la limite inférieure dans (3.23), ce qui nous donne par conservation de l'inégalité *large* :

$$\liminf_{n\to+\infty} \int_\Omega g_n \, \mathrm{d}\mu \leq \liminf_{n\to+\infty} \int_\Omega f_n \, \mathrm{d}\mu. \tag{3.24}$$

Par (3.22), on sait que la limite inférieure du premier membre de (3.24) est en fait une limite et vaut :

$$\liminf_{n\to+\infty} \int_\Omega g_n \, \mathrm{d}\mu = \lim_{n\to+\infty} \int_\Omega g_n \, \mathrm{d}\mu = \int_\Omega \liminf_{n\to+\infty} f_n \, \mathrm{d}\mu,$$

d'où la conclusion. □

Remarque 3.14. Voici un exemple simple qui peut servir à mémoriser le sens de l'inégalité dans le lemme de Fatou et à montrer que celle-ci peut être stricte. Prenons $\Omega = \mathbb{R}$, $\mathcal{F} = \operatorname{Bor}(\mathbb{R})$ et $\mu = \lambda$, mesure de Lebesgue. Prenons pour (f_n) la suite de fonctions étagées

$$f_n = \begin{cases} \mathbf{1}_{[0,1]} & \text{si } n \text{ est pair,} \\ \mathbf{1}_{]1,2]} & \text{si } n \text{ est impair.} \end{cases}$$

Il est clair que $\liminf f_n$ est la fonction identiquement nulle sur $\mathbb{R}$, d'où $\int_\mathbb{R} \liminf f_n \, \mathrm{d}\lambda = 0$. D'autre part l'intégrale de fonction étagée $\int_\mathbb{R} f_n \, \mathrm{d}\lambda$ vaut $\lambda([0,1])$ pour n pair et $\lambda(]1,2])$ pour n impair, donc pour tout n, $\int_\mathbb{R} f_n \, \mathrm{d}\lambda = 1$ et la suite constante $(\int_\mathbb{R} f_n \, \mathrm{d}\mu)$ a pour limite inférieure 1. Finalement sur cet exemple nous avons

$$0 = \int_\mathbb{R} \liminf f_n \, \mathrm{d}\lambda < \liminf \int_\mathbb{R} f_n \, \mathrm{d}\lambda = 1.$$

Remarque 3.15. Une lecture rapide de la démonstration du lemme de Fatou pourrait laisser l'impression qu'elle s'adapte avec les limites supérieures au lieu des limites inférieures. Il n'en est rien. En écrivant $h_n := \sup_{k\geq n} f_k$, on voit immédiatement que $f_n \leq h_n$, d'où $\int_\Omega f_n \, \mathrm{d}\mu \leq \int_\Omega h_n \, \mathrm{d}\mu$ et par conservation de l'inégalité large $\limsup \int_\Omega f_n \, \mathrm{d}\mu \leq \limsup \int_\Omega h_n \, \mathrm{d}\mu$. La deuxième partie de la preuve s'adapte ainsi facilement. Par contre la première partie ne fonctionne plus. En effet h_n converge en *décroissant* vers $h = \limsup f_n$ et nous n'avons pas de théorème de Beppo Levi pour les suites décroissantes. Il n'est donc plus possible d'identifier $\limsup \int_\Omega h_n \, \mathrm{d}\mu$ comme $\int_\Omega h \, \mathrm{d}\mu$.

Remarque 3.16. Le fait que la *preuve* du lemme de Fatou ne s'adapte pas aux limites supérieures n'interdit pas l'existence d'une inégalité analogue pour les limites supérieures. La confrontation des deux exemples suivants montre qu'il faut renoncer définitivement à l'idée d'un lemme de Fatou pour les limites supérieures. Le premier est celui de la remarque 3.14, où $\limsup f_n = \mathbf{1}_{[0,2]}$ et

$$1 = \limsup \int_\mathbb{R} f_n \, \mathrm{d}\lambda < \int_\mathbb{R} \limsup f_n \, \mathrm{d}\lambda = 2.$$

Le second exemple est sur le même espace mesuré, la suite (g_n) de fonctions étagées définies par $g_n := \frac{1}{n}\mathbf{1}_{[0,+\infty[}$. Clairement $\limsup g_n = 0$ d'où $\int_\mathbb{R} \limsup g_n \, \mathrm{d}\lambda = 0$. Par contre, $\int_{\mathbb{R}} g_n \,\mathrm{d}\lambda = +\infty$ pour tout n, d'où $\limsup \int_{\mathbb{R}} g_n \,\mathrm{d}\lambda = +\infty$. Contrairement au premier exemple, nous avons ici

$$0 = \int_{\mathbb{R}} \limsup g_n \,\mathrm{d}\lambda < \limsup \int_{\mathbb{R}} g_n \,\mathrm{d}\lambda = +\infty.$$

3.3 Applications aux mesures

Nous étudions dans cette section quelques retombées du théorème de Beppo Levi pour la construction de nouvelles mesures et l'intégration par rapport à une mesure image.

3.3.1 Mesure à densité par rapport à une autre mesure

Théorème 3.17. *Soit* $(\Omega, \mathcal{F}, \mu)$ *un espace mesuré et* $f : \Omega \to \overline{\mathbb{R}}_+$*, mesurable* $\mathcal{F}$-$\mathrm{Bor}(\overline{\mathbb{R}}_+)$*. Définisssons la fonction d'ensembles*

$$\nu : \mathcal{F} \to \overline{\mathbb{R}}_+, \quad A \mapsto \nu(A) := \int_A f \,\mathrm{d}\mu.$$

a) ν *est une mesure sur* $(\Omega, \mathcal{F})$*. On dit qu'elle est de densité* f *par rapport à* μ*.*

b) Pour toute $g : \Omega \to \overline{\mathbb{R}}_+$*, mesurable* $\mathcal{F}$-$\mathrm{Bor}(\overline{\mathbb{R}}_+)$*,*

$$\int_\Omega g \,\mathrm{d}\nu = \int_\Omega g f \,\mathrm{d}\mu. \tag{3.25}$$

Pour exprimer que ν est à densité f par rapport à μ, on pourra écrire *symboliquement* :

$$\mathrm{d}\nu = f \,\mathrm{d}\mu \quad \text{ou} \quad \frac{\mathrm{d}\nu}{\mathrm{d}\mu} = f.$$

L'écriture $\frac{\mathrm{d}\nu}{\mathrm{d}\mu} = f$ est un peu abusive car elle sous-entend que lorsqu'il existe une densité f, celle ci est unique. Il est facile de voir pourtant que si h est mesurable $\mathcal{F}$-$\mathrm{Bor}(\overline{\mathbb{R}}_+)$ et égale μ-presque partout à f, h est aussi une densité de ν par rapport à μ. Nous reviendrons sur cette question après la preuve du théorème.

Preuve du a). Calculons $\nu(\emptyset)$ en appliquant la définition de ν :

$$\nu(\emptyset) = \int_\emptyset f \,\mathrm{d}\mu = \int_\Omega f \mathbf{1}_\emptyset \,\mathrm{d}\mu = \int_\Omega 0 \,\mathrm{d}\mu = 0.$$

Pour prouver la σ-additivité de ν, soit $(A_i)_{i\in\mathbb{N}}$ une suite dans $\mathcal{F}$, à termes deux à deux disjoints et A sa réunion. Nous commençons par vérifier que

$$\mathbf{1}_A = \sum_{i\in\mathbb{N}} \mathbf{1}_{A_i}. \tag{3.26}$$

Pour cela, comparons les images de $\omega \in \Omega$ par chacun des deux membres de l'égalité fonctionnelle (3.26).

– Ou bien ω n'appartient pas à A et alors il ne peut appartenir à aucun des A_i, donc $\mathbf{1}_A(\omega) = 0$ et $\mathbf{1}_{A_i}(\omega) = 0$ pour tout i. Dans ce cas $\mathbf{1}_A(\omega) = 0 = \sum_{i\in\mathbb{N}} \mathbf{1}_{A_i}(\omega)$.
– Ou bien $\omega \in A$ et dans ce cas il existe un *unique* $i_0 = i_0(\omega)$ tel que $\omega \in A_{i_0}$: l'existence vient de la définition de A comme réunion des A_i et l'unicité du fait que les A_i sont deux à deux disjoints. Dans ce cas on a $\mathbf{1}_A(\omega) = 1$, $\mathbf{1}_{A_{i_0}}(\omega) = 1$ et pour tout $i \neq i_0$, $\mathbf{1}_{A_i}(\omega) = 0$, d'où $\mathbf{1}_A(\omega) = 1 = \sum_{i\in\mathbb{N}} \mathbf{1}_{A_i}(\omega)$.

En appliquant la définition de ν et (3.26), on justifie la première ligne du calcul suivant :

$$\begin{aligned} \nu\Big(\bigcup_{i\in\mathbb{N}} A_i\Big) = \int_A f \,\mathrm{d}\mu = \int_\Omega \mathbf{1}_A f \,\mathrm{d}\mu \quad &= \quad \int_\Omega \Big(\sum_{i\in\mathbb{N}} \mathbf{1}_{A_i}\Big) f \,\mathrm{d}\mu \\ &= \quad \int_\Omega \Big(\sum_{i\in\mathbb{N}} (\mathbf{1}_{A_i} f)\Big) \,\mathrm{d}\mu \qquad (3.27) \\ &= \quad \sum_{i\in\mathbb{N}} \int_\Omega \mathbf{1}_{A_i} f \,\mathrm{d}\mu \qquad (3.28) \\ &= \quad \sum_{i\in\mathbb{N}} \nu(A_i). \end{aligned}$$

L'égalité (3.27) résulte de la distributivité de la multiplication sur l'addition dans $\overline{\mathbb{R}}_+$ (noter que dans la série $\sum_{i\in\mathbb{N}} \mathbf{1}_{A_i}(\omega)$, il y a au plus un terme non nul). Pour chaque i, la fonction $\mathbf{1}_{A_i} f$ appartient à $\mathcal{M}_+$, ce qui permet d'appliquer le corollaire 3.12 pour légitimer (3.28). Ainsi la σ-additivité de ν est établie et ν est une mesure sur $(\Omega, \mathcal{F})$. □

Preuve du b). Décrivons d'abord la démarche utilisée pour prouver l'égalité (3.25), il s'agit d'une méthode standard d'usage courant dans ce cours. On commence par vérifier (3.25) pour les fonctions indicatrices $g = \mathbf{1}_A$, $A \in \mathcal{F}$. Par additivité et homogénéité, on étend cette égalité aux fonctions étagées. Enfin on obtient l'égalité pour les fonctions g mesurables positives quelconques par un passage à la limite en les approximant par une suite de fonctions étagées. Voyons cela en détail.

L'égalité (3.25) est vraie pour toute g de la forme $g = \mathbf{1}_A$, $A \in \mathcal{F}$. En effet,

$$\int_\Omega \mathbf{1}_A \,\mathrm{d}\nu = \nu(A) = \int_A f \,\mathrm{d}\mu = \int_\Omega \mathbf{1}_A f \,\mathrm{d}\mu.$$

La première de ces égalités est simplement le calcul de l'intégrale de la fonction étagée $\mathbf{1}_A$ par rapport à la *mesure* ν, cf. (3.2). Les deux autres égalités sont la définition de ν et celle de l'intégrale sur A.

Soit maintenant $g \in \mathcal{E}_+$, de décomposition canonique $g = \sum_{i=1}^n y_i \mathbf{1}_{A_i}$. Son intégrale sur Ω par rapport à la mesure ν est par définition

$$\int_\Omega g \,\mathrm{d}\nu = \sum_{i=1}^n y_i \nu(A_i).$$

En utilisant successivement la définition de $\nu(A_i)$, celle de l'intégrale sur A_i, l'additivité et l'homogénéité de l'intégrale par rapport à μ dans $\mathcal{M}_+$, on en déduit :

$$\int_\Omega g \,\mathrm{d}\nu = \sum_{i=1}^n y_i \int_{A_i} f \,\mathrm{d}\mu = \sum_{i=1}^n y_i \int_\Omega \mathbf{1}_{A_i} f \,\mathrm{d}\mu = \int_\Omega \sum_{i=1}^n y_i \mathbf{1}_{A_i} f \,\mathrm{d}\mu = \int_\Omega g f \,\mathrm{d}\mu.$$

Ainsi (3.25) est vérifiée pour toute $g \in \mathcal{E}_+$.

Soit $g \in \mathcal{M}_+$ quelconque. Par le théorème 2.23, il existe une suite croissante (g_n) dans $\mathcal{E}_+$, convergeant vers g. La fonction f étant mesurable positive, le produit $g_n f$ est aussi mesurable positif (en général, ce n'est plus une fonction étagée, seulement un élément de $\mathcal{M}_+$). Par positivité de f et croissance de (g_n), la suite $(g_n f)$ est aussi croissante dans $\mathcal{M}_+$. Vérifions la convergence de $(g_n f)$ vers gf, c'est-à-dire la convergence dans $\overline{\mathbb{R}}_+$ de $\big(g_n(\omega)f(\omega)\big)$ vers $g(\omega)f(\omega)$, pour tout $\omega \in \Omega$. Cette convergence résulte des règles classiques sur la limite d'un produit [1] lorsque le couple $\big(g(\omega), f(\omega)\big)$ n'est ni $(0, +\infty)$, ni $(+\infty, 0)$. Voyons donc ces deux cas particuliers. Si $g(\omega) = 0$ et $f(\omega) = +\infty$ alors, parce que la suite $\big(g_n(\omega)\big)$ est *croissante* [2] dans $\overline{\mathbb{R}}_+$, tous les $g_n(\omega)$ sont nuls et $g_n(\omega)f(\omega) = 0 \times (+\infty) = 0$ converge trivialement vers $g(\omega)f(\omega) = 0 \times (+\infty) = 0$. Si $g(\omega) = +\infty$ et $f(\omega) = 0$, alors pour tout n, $g_n(\omega)f(\omega) = g_n(\omega) \times 0 = 0$ et cette suite nulle converge trivialement vers $0 = g(\omega)f(\omega) = (+\infty) \times 0$.

Les hypothèses du théorème de Beppo Levi sont donc satisfaites aussi bien par la suite (g_n) que par $(g_n f)$. L'application du théorème de Beppo Levi relativement à ν pour (g_n) et à μ pour $(g_n f)$ nous donne les convergences :

$$\int_\Omega g_n \,\mathrm{d}\nu \uparrow \int_\Omega g \,\mathrm{d}\nu, \qquad \int_\Omega g_n f \,\mathrm{d}\mu \uparrow \int_\Omega g f \,\mathrm{d}\mu. \tag{3.29}$$

Comme g_n appartient à $\mathcal{E}_+$, elle vérifie (3.25) :

$$\forall n \in \mathbb{N}, \quad \int_\Omega g_n \,\mathrm{d}\nu = \int_\Omega g_n f \,\mathrm{d}\mu. \tag{3.30}$$

Les convergences (3.29) permettent de passer à la limite dans l'égalité (3.30) pour conclure que g vérifie (3.25). □

Proposition 3.18. *Soient f et g deux applications $\Omega \to \overline{\mathbb{R}}_+$, mesurables $\mathcal{F}$-*$\mathrm{Bor}(\overline{\mathbb{R}}_+)$ *telles que*

$$\forall A \in \mathcal{F}, \quad \int_A f \,\mathrm{d}\mu = \int_A g \,\mathrm{d}\mu. \tag{3.31}$$

On suppose de plus que $\int_\Omega f \,\mathrm{d}\mu < +\infty$. Alors $f = g$ μ-presque partout. La même conclusion reste vraie sous l'hypothèse plus générale de l'existence dans $\mathcal{F}$ d'une suite croissante $(\Omega_n)_{n\geq 1}$ de réunion Ω, telle que $\int_{\Omega_n} f \,\mathrm{d}\mu$ soit finie pour chaque n.

La démonstration est proposée en exercice de T.D. Cette proposition signifie que si ν est la mesure de densité f par rapport à μ, alors toute autre densité g de ν est égale à f μ-presque partout dans le cas où ν est finie ou plus généralement σ-finie.

Exemple 3.3. Soit ν la loi uniforme sur le borélien B de $\mathbb{R}^d$ $(0 < \lambda_d(B) < +\infty)$. Alors ν admet pour densité par rapport à λ_d la fonction $f = c\mathbf{1}_B$, avec $c := 1/\lambda_d(B)$.

1. On pourrait dire aussi que la multiplication $m : (x, y) \mapsto xy$ est continue en tout point de $\overline{\mathbb{R}}_+^2$, sauf aux points $(0, +\infty)$ et $(+\infty, 0)$.

2. Voici un contre-exemple simple montrant l'utilité de la croissance de g_n ici : si $f(\omega) = +\infty$ et $g_n(\omega) = 1/n$, alors $g_n(\omega)f(\omega) = +\infty$ (pour tout n) ne converge pas vers $g(\omega)f(\omega) = 0 \times (+\infty) = 0$.

Réciproquement si une loi de probabilité ν sur $(\mathbb{R}^d, \mathrm{Bor}(\mathbb{R}^d))$ admet par rapport à λ_d une densité de la forme $f = c\mathbf{1}_B$, avec $c > 0$ et $B \in \mathrm{Bor}(\mathbb{R}^d)$, alors ν est la loi uniforme sur B, $0 < \lambda_d(B) < +\infty$ et $c = 1/\lambda_d(B)$.

Vérification. Par définition, si ν est la loi uniforme sur B, $\nu(A) = c\lambda_d(A \cap B)$ pour tout $A \in \mathrm{Bor}(\mathbb{R}^d)$, avec $c := 1/\lambda_d(B)$. En remarquant que $\mathbf{1}_{A\cap B} = \mathbf{1}_A \mathbf{1}_B$, on a donc

$$\nu(A) = c\lambda_d(A \cap B) = c \int_{\mathbb{R}^d} \mathbf{1}_{A\cap B} \,\mathrm{d}\lambda_d = c \int_{\mathbb{R}^d} \mathbf{1}_A \mathbf{1}_B \,\mathrm{d}\lambda_d = \int_A c\mathbf{1}_B \,\mathrm{d}\lambda_d.$$

Ces égalités étant vérifiées pour tout $A \in \mathrm{Bor}(\mathbb{R}^d)$, la mesure ν a pour densité $c\mathbf{1}_B$ par rapport à λ_d.

Réciproquement si une loi de probabilité ν sur $(\mathbb{R}^d, \mathrm{Bor}(\mathbb{R}^d))$ admet par rapport à λ_d une densité de la forme $f = c\mathbf{1}_B$, avec $c > 0$ et $B \in \mathrm{Bor}(\mathbb{R}^d)$, alors pour tout $A \in \mathrm{Bor}(\mathbb{R}^d)$,

$$\nu(A) = \int_A c\mathbf{1}_B \,\mathrm{d}\lambda_d = c \int_{\mathbb{R}^d} \mathbf{1}_A \mathbf{1}_B \,\mathrm{d}\lambda_d = c \int_{\mathbb{R}^d} \mathbf{1}_{A\cap B} \,\mathrm{d}\lambda_d = c\lambda_d(A \cap B).$$

En particulier pour $A = \mathbb{R}^d$, on obtient $1 = \nu(\mathbb{R}^d) = c\lambda_d(B)$. Cette égalité impose la finitude et la stricte positivité de $\lambda_d(B)$ et nous donne $c = 1/\lambda_d(B)$. en revenant au cas général, on a donc $\nu(A) = \lambda_d(A \cap B)/\lambda_d(B)$, ce qui montre que ν est la loi uniforme sur B. □

Pour voir d'autres exemples intéressants de lois à densité par rapport à λ_d, il nous faudra attendre de savoir intégrer pratiquement par rapport à λ_d, ce qui sera vu aux chapitres 4 pour la dimension 1 et 5 pour $d > 1$.

3.3.2 Transfert

Nous étudions maintenant l'intégration par rapport à une mesure image. Le résultat fondamental est le *théorème de transfert*. Travaillant avec deux espaces mesurables $(\Omega_1, \mathcal{F}_1)$ et $(\Omega_2, \mathcal{F}_2)$, nous noterons pour $i = 1, 2$, $\mathcal{E}_+(\Omega_i)$ et $\mathcal{M}_+(\Omega_i)$ les ensembles de fonctions $\Omega_i \to \overline{\mathbb{R}}_+$ respectivement étagées mesurables ou mesurables $\mathcal{F}_i$-$\mathrm{Bor}(\overline{\mathbb{R}}_+)$.

Théorème 3.19 (de transfert)**.** *Soient $(\Omega_1, \mathcal{F}_1)$ et $(\Omega_2, \mathcal{F}_2)$ deux espaces mesurables et $\varphi : \Omega_1 \to \Omega_2$, une application $\mathcal{F}_1$-$\mathcal{F}_2$ mesurable. Soit μ une mesure sur $(\Omega_1, \mathcal{F}_1)$ et $\nu := \mu \circ \varphi^{-1}$ sa mesure image sur $(\Omega_2, \mathcal{F}_2)$. Alors pour toute $h \in \mathcal{M}_+(\Omega_2)$,*

$$\int_{\Omega_1} (h \circ \varphi) \,\mathrm{d}\mu = \int_{\Omega_2} h \,\mathrm{d}\nu. \tag{3.32}$$

Ce théorème est essentiellement un théorème de changement de variable dans l'intégrale sur un espace abstrait. En écrivant le premier membre de (3.32) sous la forme $\int_{\Omega_1} h\big(\varphi(\omega_1)\big) \,\mathrm{d}\mu(\omega_1)$, on voit qu'il suffit de poser $\omega_2 = \varphi(\omega_1)$ et de remplacer μ par sa mesure image $\nu = \mu \circ \varphi^{-1}$ pour effectuer le changement de variable et obtenir $\int_{\Omega_2} h(\omega_2) \,\mathrm{d}\nu(\omega_2)$. Autrement dit, *le changement de variable dans une intégrale abstraite se réduit à la détermination de la mesure image.* Remarquer que l'on ne fait aucune hypothèse sur φ, à part sa mesurabilité. En particulier, φ n'a nul besoin d'être bijective.

Preuve. D'abord, puisque $h \in \mathcal{M}_+(\Omega_2)$ et φ est $\mathcal{F}_1$-$\mathcal{F}_2$ mesurable, $h \circ \varphi$ appartient à $\mathcal{M}_+(\Omega_1)$, donc le premier membre de (3.32) est bien défini. Nous utilisons à nouveau la méthode standard vue ci-dessus en vérifiant (3.32) successivement pour les indicatrices, les fonctions étagées et enfin les applications mesurables positives.

Cas des indicatrices : $h = \mathbf{1}_B$, $B \in \mathcal{F}_2$. Par mesurabilité de φ, $A := \varphi^{-1}(B)$ appartient à $\mathcal{F}_1$ et on a :

$$\int_{\Omega_2} \mathbf{1}_B \,\mathrm{d}\nu = \nu(B) = \mu\Big(\varphi^{-1}(B)\Big) = \mu(A) = \int_{\Omega_1} \mathbf{1}_A \,\mathrm{d}\mu.$$

Il ne reste plus qu'à vérifier que $\mathbf{1}_A = \mathbf{1}_B \circ \varphi$, ce qui résulte des égalités suivantes :

$$\forall \omega \in \Omega_1, \quad \mathbf{1}_A(\omega) = \begin{cases} 1 & \text{si } \omega \in \varphi^{-1}(B) \\ 0 & \text{si } \omega \notin \varphi^{-1}(B) \end{cases} = \begin{cases} 1 & \text{si } \varphi(\omega) \in B \\ 0 & \text{si } \varphi(\omega) \notin B \end{cases} = \mathbf{1}_B\Big(\varphi(\omega)\Big).$$

Cas des fonctions étagées : $h \in \mathcal{E}_+(\Omega_2)$, de décomposition canonique $h = \sum_{i=1}^n y_i \mathbf{1}_{B_i}$. Alors $h \circ \varphi = \sum_{i=1}^n y_i (\mathbf{1}_{B_i} \circ \varphi)$. En utilisant l'additivité et l'homogénéité de l'intégrale par rapport à μ et à ν et l'égalité (3.32) pour les indicatrices, on obtient :

$$\int_{\Omega_1} (h \circ \varphi) \,\mathrm{d}\mu = \sum_{i=1}^n y_i \int_{\Omega_1} (\mathbf{1}_{B_i} \circ \varphi) \,\mathrm{d}\mu = \sum_{i=1}^n y_i \int_{\Omega_2} \mathbf{1}_{B_i} \,\mathrm{d}\nu = \int_{\Omega_2} \sum_{i=1}^n y_i \mathbf{1}_{B_i} \,\mathrm{d}\nu = \int_{\Omega_2} h \,\mathrm{d}\nu,$$

ce qui établit (3.32) pour $h \in \mathcal{E}_+(\Omega_2)$.

Cas des fonctions mesurables positives : $h \in \mathcal{M}_+(\Omega_2)$. Alors il existe une suite $(h_n)_{n \in \mathbb{N}}$ dans $\mathcal{E}_+(\Omega_2)$, croissante et convergente vers h. On voit alors immédiatement que la suite $(h_n \circ \varphi)_{n \in \mathbb{N}}$ est dans $\mathcal{E}_+(\Omega_1)$, croissante et convergente vers $h \circ \varphi$. L'égalité (3.32) établie ci-dessus pour les éléments de $\mathcal{E}_+(\Omega_2)$ nous donne :

$$\forall n \in \mathbb{N}, \quad \int_{\Omega_1} (h_n \circ \varphi) \,\mathrm{d}\mu = \int_{\Omega_2} h_n \,\mathrm{d}\nu. \tag{3.33}$$

Le théorème de Beppo Levi appliqué aux suites $(h_n \circ \varphi)$ et (h_n) permet de passer à la limite dans (3.33) pour obtenir $\int_{\Omega_1} (h \circ \varphi) \,\mathrm{d}\mu = \int_{\Omega_2} h \,\mathrm{d}\nu$ et achever ainsi la preuve du théorème. □

Voici une importante application du théorème de transfert aux calculs d'espérances. Nous l'énonçons dans le cas des vecteurs aléatoires $\Omega \to \mathbb{R}^d$, le lecteur pourra l'adapter à celui des variables aléatoires réelles en faisant $d = 1$, ou à celui des variables aléatoires complexes.

Corollaire 3.20 (Calcul d'espérance par transfert)**.** *Soit* $(\Omega, \mathcal{F}, \mathbf{P})$ *un espace probabilisé et* $X : \Omega \to \mathbb{R}^d$ *un vecteur aléatoire sur* $(\Omega, \mathcal{F})$*. Notons* $\mathbf{E}$ *l'espérance relativement à* $\mathbf{P}$ *et* P_X *la loi de* X *sous* $\mathbf{P}$*. Alors pour toute fonction mesurable positive* $h : \mathbb{R}^d \to \overline{\mathbb{R}}_+$,

$$\mathbf{E}h(X) = \int_\Omega h(X) \,\mathrm{d}\mathbf{P} = \int_\Omega (h \circ X) \,\mathrm{d}\mathbf{P} = \int_{\mathbb{R}^d} h \,\mathrm{d}P_X. \tag{3.34}$$

Vérification. Il suffit d'appliquer le théorème de transfert avec $\Omega_1 = \Omega$, $\mathcal{F}_1 = \mathcal{F}$, $\varphi = X$, $\Omega_2 = \mathbb{R}^d$, $\mathcal{F}_2 = \mathrm{Bor}(\mathbb{R}^d)$ et $\nu = P_X = \mathbf{P} \circ \varphi^{-1}$. □

Ce corollaire 3.20 est d'importance capitale. À lui seul il serait déjà une motivation suffisante pour notre entreprise de construction de l'intégrale abstraite. Le corollaire 3.20 montre que $\mathbf{E}h(X)$ ne dépend que de h et de la loi de X. On pourrait ainsi calculer cette espérance en ignorant tout de l'espace $(\Omega, \mathcal{F}, \mathbf{P})$, pourvu que l'on connaisse la loi de X. De plus (3.34) ramène l'intégration sur l'espace abstrait Ω à une intégration sur l'espace beaucoup plus familier $\mathbb{R}^d$. Pour que ce procédé soit effectif, il reste à savoir identifier P_X et à savoir intégrer sur $\mathbb{R}^d$. Un cas important en pratique est celui où la loi de X a une densité par rapport à la mesure de Lebesgue.

Corollaire 3.21 (Espérances pour des lois à densité)**.** *Si la loi du vecteur aléatoire X a une densité f par rapport à la mesure de Lebesgue λ_d de $\mathbb{R}^d$, alors pour toute fonction mesurable positive $h : \mathbb{R}^d \to \overline{\mathbb{R}}_+$,*

$$\mathbf{E}h(X) = \int_{\mathbb{R}^d} hf \, \mathrm{d}\lambda_d. \tag{3.35}$$

Vérification. L'égalité (3.35) résulte immédiatement de la combinaison du corollaire 3.20 et du théorème 3.17 b). □

Terminons par un corollaire qui aura son utilité dans les calculs effectifs de lois.

Corollaire 3.22 (Identification de lois)**.** *Les vecteurs aléatoires X et Y à valeurs dans $\mathbb{R}^d$ ont même loi si et seulement si pour toute $h : \mathbb{R}^d \to \overline{\mathbb{R}}_+$ mesurable positive,*

$$\mathbf{E}h(X) = \mathbf{E}h(Y).$$

Preuve. Remarquons d'abord qu'il n'est pas nécessaire que X et Y soient définis sur le même espace probabilisé $(\Omega, \mathcal{F}, \mathbf{P})$. Si X et Y ont même loi, cela signifie l'égalité des mesures P_X et P_Y définies sur $(\mathbb{R}^d, \mathrm{Bor}(\mathbb{R}^d))$. Alors pour toute $h \in \mathcal{M}_+(\mathbb{R}^d)$, on a l'égalité d'intégrales $\int_{\mathbb{R}^d} h \, \mathrm{d}P_X = \int_{\mathbb{R}^d} h \, \mathrm{d}P_Y$ qui grâce au corollaire 3.20, s'écrit aussi $\mathbf{E}h(X) = \mathbf{E}h(Y)$.

Pour la réciproque, il suffit de noter que si ν_1 et ν_2 sont deux mesures sur $(\mathbb{R}^d, \mathrm{Bor}(\mathbb{R}^d))$ telles que pour toute $h \in \mathcal{M}_+(\mathbb{R}^d)$, $\int_{\mathbb{R}^d} h \, \mathrm{d}\nu_1 = \int_{\mathbb{R}^d} h \, \mathrm{d}\nu_2$, alors $\nu_1 = \nu_2$, ce qui est évident en prenant des h de la forme $\mathbf{1}_B$, où B est un borélien quelconque de $\mathbb{R}^d$. □

3.3.3 Intégration par rapport à une série de mesures

Pour clore ce chapitre, nous examinons le calcul d'intégrales par rapport à une mesure de Dirac ou à une série de mesures, avec une application aux variables aléatoires discrètes.

Proposition 3.23 (Intégrale par rapport à une mesure de Dirac)**.** *Soit δ_{ω_0} la mesure de Dirac au point $\omega_0 \in \Omega$.*

$$\forall f \in \mathcal{M}_+, \qquad \int_\Omega f \, \mathrm{d}\delta_{\omega_0} = f(\omega_0). \tag{3.36}$$

Preuve. On utilise une nouvelle fois la méthode standard en vérifiant (3.36) successivement pour les indicatrices, les fonctions étagées, puis les fonctions mesurables positives.

Vérification pour $f = \mathbf{1}_A$, $A \in \mathcal{F}$. D'après (3.2) et la définition d'une mesure de Dirac (cf. Exemple 1.1, p. 25),

$$\int_\Omega \mathbf{1}_A \,\mathrm{d}\delta_{\omega_0} = \delta_{\omega_0}(A) = \begin{cases} 1 & \text{si } \omega_0 \in A \\ 0 & \text{si } \omega_0 \notin A \end{cases} = \mathbf{1}_A(\omega_0) = f(\omega_0).$$

Vérification pour $f \in \mathcal{E}_+$. Soit $\sum_{i=1}^n y_i \mathbf{1}_{A_i}$ la décomposition canonique de f. Par définition de l'intégrale sur $\mathcal{E}_+$,

$$\int_\Omega f \,\mathrm{d}\delta_{\omega_0} = \sum_{i=1}^n y_i \delta_{\omega_0}(A_i) = \sum_{i=1}^n y_i \mathbf{1}_{A_i}(\omega_0) = f(\omega_0).$$

Vérification pour $f \in \mathcal{M}_+$. Alors f est limite d'une suite croissante (f_n) d'éléments de $\mathcal{E}_+$. Comme (3.36) est vraie dans $\mathcal{E}_+$,

$$\forall n \in \mathbb{N}, \qquad \int_\Omega f_n \,\mathrm{d}\delta_{\omega_0} = f_n(\omega_0).$$

Faisons tendre n vers $+\infty$ dans cette égalité. Le premier membre converge vers $\int_\Omega f \,\mathrm{d}\delta_{\omega_0}$ par le théorème de Beppo Levi, le second membre converge vers $f(\omega_0)$ par construction de f_n. Ainsi f vérifie elle aussi (3.36). □

Proposition 3.24 (Intégrale par rapport à une série de mesures)**.** *Soient $(\mu_k)_{k\in\mathbb{N}}$ une suite de mesures sur le même espace mesurable $(\Omega, \mathcal{F})$ et $(a_k)_{k\in\mathbb{N}}$ une suite dans $\mathbb{R}_+$. La fonction d'ensembles*

$$\mu := \sum_{k\in\mathbb{N}} a_k \mu_k$$

est une mesure sur $(\Omega, \mathcal{F})$ et

$$\forall f \in \mathcal{M}_+, \qquad \int_\Omega f \,\mathrm{d}\mu = \sum_{k\in\mathbb{N}} a_k \int_\Omega f \,\mathrm{d}\mu_k. \tag{3.37}$$

Preuve. Le fait que μ soit une mesure a déjà été établi au chapitre 1 (exemple 1.2), en supposant la finitude des mesures μ_k. Le seul rôle de cette hypothèse était d'éviter un conflit du type $0 \times (+\infty)$ lorsque $a_k = 0$ et $\mu_k(A) = +\infty$ qui nous aurait empêché de définir la fonction d'ensembles $a_k\mu_k$. Ce conflit a été résolu par les conventions arithmétiques faites au chapitre 2 qui entraînent que $a_k\mu_k$ est la mesure nulle, que μ_k soit finie ou non. La restriction faite à l'exemple 1.2 n'a donc plus lieu d'être et la démonstration que μ est une mesure peut se recopier telle quelle.

Pour vérifier (3.37), on fait appel une dernière fois[3] à la méthode standard, réduite ici à deux étapes.

3. Pour ce chapitre...

Vérification pour $f \in \mathcal{E}_+$. Soit $\sum_{i=1}^n y_i \mathbf{1}_{A_i}$ la décomposition canonique de f. Par définition de l'intégrale sur $\mathcal{E}_+$ et la propriété de sommation par paquets dans $\overline{\mathbb{R}}_+$,

$$\int_\Omega f \,\mathrm{d}\mu = \sum_{i=1}^n y_i \mu(A_i) = \sum_{i=1}^n y_i \sum_{k\in\mathbb{N}} a_k \mu_k(A_i) = \sum_{k\in\mathbb{N}} a_k \sum_{i=1}^n y_i \mu_k(A_i) = \sum_{k\in\mathbb{N}} a_k \int_\Omega f \,\mathrm{d}\mu_k.$$

Vérification pour $f \in \mathcal{M}_+$. Alors f est limite d'une suite croissante (f_n) d'éléments de $\mathcal{E}_+$. Comme (3.37) est vraie dans $\mathcal{E}_+$,

$$\forall n \in \mathbb{N}, \quad \int_\Omega f_n \,\mathrm{d}\mu = \sum_{k\in\mathbb{N}} a_k \int_\Omega f_n \,\mathrm{d}\mu_k. \tag{3.38}$$

En appliquant le théorème de Beppo Levi (une infinité de fois !), on obtient les convergences :

$$\int_\Omega f_n \,\mathrm{d}\mu \uparrow \int_\Omega f \,\mathrm{d}\mu, \quad \text{et} \quad \forall k \in \mathbb{N}, \quad \int_\Omega f_n \,\mathrm{d}\mu_k \uparrow \int_\Omega f \,\mathrm{d}\mu_k. \tag{3.39}$$

Pour obtenir (3.37) par passage à la limite dans (3.38), il n'y a aucun problème avec le premier membre, grâce à la première des convergences (3.39). Par contre il reste à justifier le passage à la limite dans la série. En posant $u_{n,k} := a_k \int_\Omega f_n \,\mathrm{d}\mu_k$ et $u_k := a_k \int_\Omega f \,\mathrm{d}\mu_k$, on a pour chaque k par (3.39), $u_{n,k} \uparrow u_k$ quand n tend vers l'infini. Pour conclure, il suffit alors d'appliquer le lemme suivant dont la preuve n'utilise pas la théorie de l'intégration et a été vue en introduction à ce cours. □

Lemme 3.25 (Théorème de Beppo Levi pour les séries)**.** *Soit* $(u_{n,k};\ (n,k) \in \mathbb{N}^2)$ *une suite double dans* $\overline{\mathbb{R}}_+$ *telle que pour tout* $k \in \mathbb{N}$*, la suite simple* $(u_{n,k})_{n\in\mathbb{N}}$ *converge* en croissant *dans* $\overline{\mathbb{R}}_+$ *vers* u_k *(*$u_{n,k} \uparrow u_k \leq +\infty$*). Alors*

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

Corollaire 3.26 (Espérance d'une variable aléatoire positive discrète)**.** *Soit* $(\Omega, \mathcal{F}, \mathbf{P})$ *un espace probabilisé et* X *une variable aléatoire positive discrète sur* $(\Omega, \mathcal{F})$*. Alors l'espérance de* X *sous* $\mathbf{P}$ *peut se calculer par*

$$\mathbf{E}X = \sum_{x\in X(\Omega)} x \mathbf{P}(X = x). \tag{3.40}$$

Preuve. Une variable aléatoire positive discrète X sur $(\Omega, \mathcal{F})$ est une application $\Omega \to \overline{\mathbb{R}}_+$, telle que $X(\Omega)$ soit au plus *dénombrable* et X soit mesurable [4] $\mathcal{F}$-$\mathcal{P}(\overline{\mathbb{R}}_+)$. Elle est donc *a fortiori* mesurable $\mathcal{F}$-Bor$(\overline{\mathbb{R}}_+)$. La loi de X sous $\mathbf{P}$ est la mesure image $P_X = \mathbf{P} \circ X^{-1}$ et peut s'écrire (cf. Prop. 2.27) comme une série (ou une somme finie si $X(\Omega)$ est fini) de mesures :

$$P_X = \sum_{x\in X(\Omega)} \mathbf{P}(X = x)\delta_x.$$

4. Cette mesurabilité équivaut ici à $\forall x \in X(\Omega)$, $X^{-1}(\{x\}) \in \mathcal{F}$, comme on peut le voir par une adaptation immédiate de la preuve du corollaire 2.11.

En appliquant le théorème de transfert avec $(\Omega_1, \mathcal{F}_1, \mu) = (\Omega, \mathcal{F}, \mathbf{P})$, $\varphi = X$, $(\Omega_2, \mathcal{F}_2, \nu) = (\overline{\mathbb{R}}_+, \mathrm{Bor}(\overline{\mathbb{R}}_+), P_X)$ et h égale à *l'identité* sur $\overline{\mathbb{R}}_+$, on obtient :

$$\mathbf{E}X = \int_\Omega X \,\mathrm{d}P = \int_{\overline{\mathbb{R}}_+} h \,\mathrm{d}P_X.$$

Numérotons les éléments de $X(\Omega)$ par des indices entiers $X(\Omega) = \{x_0, x_1, \dots, x_k, \dots\}$. Par la proposition 3.24 avec $a_k := \mathbf{P}(X = x_k)$ et $\mu_k := \delta_{x_k}$ et la proposition 3.23,

$$\int_{\overline{\mathbb{R}}_+} h \,\mathrm{d}P_X = \sum_{x_k \in X(\Omega)} a_k \int_{\overline{\mathbb{R}}_+} h \,\mathrm{d}\delta_{x_k} = \sum_{x_k \in X(\Omega)} \mathbf{P}(X = x_k) h(x_k) = \sum_{x_k \in X(\Omega)} \mathbf{P}(X = x_k) x_k.$$

Il ne reste plus qu'à dégraisser les notations en oubliant la numérotation entière de $X(\Omega)$ pour obtenir (3.40). □

Chapitre 4

Intégration, espace $\mathbf{L}^1$

Dans tout ce chapitre, $(\Omega, \mathcal{F}, \mu)$ est un espace mesuré. Nous allons définir et étudier l'intégrale sur cet espace, de fonctions mesurables de signe *non constant*. Ayant déjà vu l'intégrale de fonctions mesurables positives, on pourrait définir de manière analogue l'intégrale des fonctions mesurables *négatives* (i.e. $\forall \omega \in \Omega$, $f(\omega) \in [-\infty, 0]$) par un simple changement de signe : $\int_\Omega f \, \mathrm{d}\mu := -\int_\Omega (-f) \, \mathrm{d}\mu$. L'étape suivante serait de décomposer une fonction f de signe non constant en $f = f^+ - f^-$, différence de deux fonctions positives et de poser $\int_\Omega f \, \mathrm{d}\mu := \int_\Omega f^+ \, \mathrm{d}\mu - \int_\Omega f^- \, \mathrm{d}\mu$. On voit tout de suite poindre le danger d'une écriture « $+\infty - \infty$ », dépourvue de sens. Ceci nous conduit à renoncer à définir une intégrale pour *toutes* les fonctions mesurables et à introduire l'ensemble $\mathcal{L}^1(\mu)$ des fonctions μ-intégrables pour lesquelles l'intégrale $\int_\Omega f \, \mathrm{d}\mu$ sera toujours un *réel* ou un complexe, mais jamais infini.

4.1 Fonctions μ-intégrables

Définition 4.1 (Intégrabilités). *Soit f une application définie sur Ω et à valeurs dans $\mathbb{K} = \mathbb{R}$, $\overline{\mathbb{R}}$ ou $\mathbb{C}$. On dit qu'elle est μ-intégrable si*

a) elle est mesurable $\mathcal{F}$-$\mathrm{Bor}(\mathbb{K})$ *;*

b) l'intégrale $\int_\Omega |f| \, \mathrm{d}\mu$ est finie, *$|f|$ désignant la valeur absolue (cas $\mathbb{R}$ et $\overline{\mathbb{R}}$) ou le module (cas $\mathbb{C}$).*

Soit $A \in \mathcal{F}$, on dit que f est μ-intégrable sur A si la fonction $f\mathbf{1}_A$ est μ-intégrable au sens ci-dessus.

Remarques 4.2.

1. La mesurabilité ne concerne que les tribus ; par contre l'intégrabilité est toujours relative à une mesure particulière.
2. Une fonction mesurable positive a toujours une intégrale (et ce par rapport à n'importe quelle mesure). Elle n'est pas forcément intégra**b**le par rapport à une mesure donnée.

3. Il importe de ne pas omettre la condition a), car $|f|$ peut très bien être mesurable et avoir une intégrale finie sans que f soit mesurable. Pour s'en convaincre, prenons pour μ une mesure de probabilité et pour f la fonction $\mathbf{1}_B - \mathbf{1}_{B^c}$ où $B \notin \mathcal{F}$. Alors f n'est pas mesurable puisque $f^{-1}(\{1\}) = B \notin \mathcal{F}$. Par contre $|f|$ est la fonction constante sur Ω valant 1, elle est mesurable et $\int_\Omega |f| \,\mathrm{d}\mu = 1 < +\infty$.
4. On peut vérifier que la μ-intégrabilité de f sur $A \in \mathcal{F}$ équivaut à la μ_A-intégrabilité de la restriction de f à A, où μ_A est la mesure restriction de μ à la tribu trace de $\mathcal{F}$ sur A (cf. remarque 3.7).

Si f est à valeurs dans $\overline{\mathbb{R}}$ ou $\mathbb{R}$, rappelons que sa partie positive f^+ et sa partie négative f^- sont les fonctions positives

$$f^+ = \max(f; 0), \qquad f^- = -\min(f; 0) = \max(-f; 0).$$

Si $f(\omega) > 0$, $f^+(\omega) = f(\omega) = |f(\omega)|$ et $f^-(\omega) = 0$. Si $f(\omega) < 0$, $f^+(\omega) = 0$ et $f^-(\omega) = -f(\omega) = |f(\omega)|$. Enfin si $f(\omega) = 0$, $f^+(\omega) = f^-(\omega) = 0$. On voit ainsi que

$$|f| = f^+ + f^-, \qquad f = f^+ - f^-, \tag{4.1}$$

en notant au passage que cette dernière égalité ne génère *jamais* l'expression non définie « $+\infty - \infty$ », car pour chaque ω, au moins l'un des deux termes $f^+(\omega)$ et $f^-(\omega)$ vaut zéro. Il résulte immédiatement de (4.1) que f est μ-intégrable si et seulement si f^+ et f^- sont mesurables et ont des intégrales finies.

Concernant les fonctions f à valeurs complexes, l'équivalence entre la mesurabilité de f et celles de $\operatorname{Re} f$ et $\operatorname{Im} f$ ainsi que les inégalités

$$\max\big(|\operatorname{Re} f|; |\operatorname{Im} f|\big) \leq |f| \leq |\operatorname{Re} f| + |\operatorname{Im} f|, \tag{4.2}$$

montrent que f est μ-intégrable si et seulement si les fonctions réelles $\operatorname{Re} f$ et $\operatorname{Im} f$ le sont.

Définition 4.3 (Intégrales). *Soit f une application définie sur Ω, à valeurs dans $\mathbb{K} = \mathbb{R}$, $\overline{\mathbb{R}}$ ou $\mathbb{C}$ et μ-intégrable. Son intégrale par rapport à μ sur Ω est alors définie par :*

- *si* $\mathbb{K} = \mathbb{R}$ *ou* $\overline{\mathbb{R}}$, $\displaystyle\int_\Omega f \,\mathrm{d}\mu := \int_\Omega f^+ \,\mathrm{d}\mu - \int_\Omega f^- \,\mathrm{d}\mu$;
- *si* $\mathbb{K} = \mathbb{C}$, $\displaystyle\int_\Omega f \,\mathrm{d}\mu := \int_\Omega \operatorname{Re} f \,\mathrm{d}\mu + i \int_\Omega \operatorname{Im} f \,\mathrm{d}\mu$.

Si $A \in \mathcal{F}$, l'intégrale de f sur A est définie par

$$\int_A f \,\mathrm{d}\mu := \int_\Omega f \mathbf{1}_A \,\mathrm{d}\mu.$$

Cette intégrale est aussi notée $\int_A f(\omega) \,\mathrm{d}\mu(\omega)$ ou $\int_A f(\omega)\mu(\mathrm{d}\omega)$, y compris lorsque $A = \Omega$.

Lorsque l'on considère l'intégrale comme une application définie sur un ensemble de fonctions et à valeurs dans $\mathbb{K}$, il est parfois commode d'utiliser la notation $\mu(f) := \int_\Omega f \,\mathrm{d}\mu$. De ce point de vue, $\int_A f \,\mathrm{d}\mu = \mu(f\mathbf{1}_A)$ et $\mu(A) = \mu(\mathbf{1}_A)$.

Soit f à valeurs dans $\mathbb{R}_+$ ou $\overline{\mathbb{R}}_+$ et μ-intégrable. Il est clair que la définition 4.3 de $\int_\Omega f \,\mathrm{d}\mu$ est compatible avec celle de l'intégrale d'une fonction mesurable positive au sens du chapitre 3.

Si f est à valeurs complexes et intégrable, il est clair par la définition 4.3 que

$$\mathrm{Re}\Big(\int_\Omega f \,\mathrm{d}\mu\Big) = \int_\Omega \mathrm{Re}\, f \,\mathrm{d}\mu, \quad \mathrm{Im}\Big(\int_\Omega f \,\mathrm{d}\mu\Big) = \int_\Omega \mathrm{Im}\, f \,\mathrm{d}\mu$$

Dans le langage des probabilités, la définition 4.3 est celle de l'espérance mathématique d'une variable aléatoire réelle ou complexe. Rappelons qu'une telle variable aléatoire est simplement une application mesurable $\mathcal{F}$-$\mathrm{Bor}(\mathbb{K})$, avec $\mathbb{K} = \mathbb{R}$ ou $\mathbb{C}$.

Définition 4.4 (Espérance)**.** *Soit $(\Omega, \mathcal{F}, \mathbf{P})$ un espace probabilisé et X une variable aléatoire définie sur Ω et à valeurs dans $\mathbb{R}$ ou $\mathbb{C}$. Si X est $\mathbf{P}$-intégrable, on définit son espérance mathématique (sous $\mathbf{P}$) par*

$$\mathbf{E}X := \int_\Omega X \,\mathrm{d}\mathbf{P} = \int_\Omega X(\omega) \,\mathrm{d}\mathbf{P}(\omega).$$

Avant d'établir les propriétés générales de l'intégrale, examinons deux exemples importants dont la combinaison va nous donner le calcul d'intégrales par rapport à une mesure discrète quelconque.

Proposition 4.5. *Soit ω_0 fixé dans Ω et δ_{ω_0} la mesure de Dirac en ce point. Une fonction mesurable $f : \Omega \to \mathbb{K}$ est toujours δ_{ω_0}-intégrable si $\mathbb{K} = \mathbb{R}$, ou $\mathbb{C}$. Si $\mathbb{K} = \overline{\mathbb{R}}$, la δ_{ω_0}-intégrabilité de f équivaut à la finitude de $f(\omega_0)$. Quand f est δ_{ω_0}-intégrable, on a*

$$\int_\Omega f \,\mathrm{d}\delta_{\omega_0} = f(\omega_0).$$

Preuve. Par mesurabilité de f, $|f| \in \mathcal{M}_+$ et par la proposition 3.23,

$$\int_\Omega |f| \,\mathrm{d}\delta_{\omega_0} = |f|(\omega_0) = |f(\omega_0)|.$$

La δ_{ω_0}-intégrabilité de f se réduit donc dans ce cas à la finitude de $f(\omega_0)$, laquelle est automatique si $\mathbb{K} = \mathbb{R}$, ou $\mathbb{C}$. Quand f est intégrable et $\mathbb{K} = \mathbb{R}$, la définition 4.3 et la proposition 3.23 appliquée aux fonctions mesurables positives f^+ et f^- nous donnent :

$$\int_\Omega f \,\mathrm{d}\delta_{\omega_0} = \int_\Omega f^+ \,\mathrm{d}\delta_{\omega_0} - \int_\Omega f^- \,\mathrm{d}\delta_{\omega_0} = f^+(\omega_0) - f^-(\omega_0) = f(\omega_0).$$

Le cas complexe se ramène au cas réel par séparation des parties réelles et imaginaires. $\square$

Proposition 4.6. *Soit J une partie de $\mathbb{N}$, $(\mu_j)_{j\in J}$ une suite (finie ou infinie) de mesures sur $(\Omega, \mathcal{F})$ et $(a_j)_{j\in J}$ une suite de réels* strictement *positifs. Notons μ la mesure $\sum_{j\in J} a_j\mu_j$ sur $(\Omega, \mathcal{F})$. Une application $f : \Omega \to \mathbb{K}$ avec $\mathbb{K} = \mathbb{R}$, $\overline{\mathbb{R}}$ ou $\mathbb{C}$, est μ-intégrable si et seulement si elle est* mesurable $\mathcal{F}$-Bor$(\mathbb{K})$ *et*

$$\sum_{j\in J} a_j \int_\Omega |f| \,\mathrm{d}\mu_j < +\infty. \tag{4.3}$$

Quand f est μ intégrable,

$$\int_\Omega f \,\mathrm{d}\mu = \sum_{j\in J} a_j \int_\Omega f \,\mathrm{d}\mu_j. \tag{4.4}$$

Preuve. La caractérisation de la μ-intégrabilité n'est que la reproduction de la définition 4.1, dans laquelle on a implanté la formule de calcul de $\int_\Omega |f| \,\mathrm{d}\mu$ fournie par la proposition 3.24. Notons que dans la proposition 3.24, on n'interdit pas aux coefficients a_k d'être nuls, donc la formule de calcul des intégrales de fonctions mesurables *positives* vaut aussi pour les mesures $\sum_{j\in J} a_j\mu_j$ avec $J \subset \mathbb{N}$.

Supposons maintenant f μ-intégrable. La condition (4.3) et la stricte positivité de chaque a_j, $j \in J$ impliquent la finitude de $\int_\Omega |f| \,\mathrm{d}\mu_j$ et donc la μ_j-intégrabilité [1] de f.

Lorsque $\mathbb{K} = \mathbb{R}$ ou $\overline{\mathbb{R}}$, en appliquant la proposition 3.24 aux fonctions mesurables positives f^+ et f^-, on obtient

$$\int_\Omega f^+ \,\mathrm{d}\mu = \sum_{j\in J} a_j \int_\Omega f^+ \,\mathrm{d}\mu_j, \quad \int_\Omega f^- \,\mathrm{d}\mu = \sum_{j\in J} a_j \int_\Omega f^- \,\mathrm{d}\mu_j. \tag{4.5}$$

Comme $\int_\Omega f^+ \,\mathrm{d}\mu$ et $\int_\Omega f^- \,\mathrm{d}\mu$ sont majorées par $\int_\Omega |f| \,\mathrm{d}\mu$, la μ-intégrabilité de f entraîne la convergence vers un réel positif *fini* des séries [2] figurant dans (4.5). On peut donc former leur différence qui est ainsi une série convergente dans $\mathbb{R}$ et l'application de la définition 4.3 nous donne

$$\int_\Omega f \,\mathrm{d}\mu = \int_\Omega f^+ \,\mathrm{d}\mu - \int_\Omega f^- \,\mathrm{d}\mu = \sum_{j\in J} a_j \left\{ \int_\Omega f^+ \,\mathrm{d}\mu_j - \int_\Omega f^- \,\mathrm{d}\mu_j \right\} = \sum_{j\in J} a_j \int_\Omega f \,\mathrm{d}\mu_j.$$

□

Corollaire 4.7 (Représentation d'une série comme intégrale)**.** *Soit $\sum_{k=0}^{+\infty} u_k$ une série à termes réels ou complexes,* absolument *convergente. Elle peut s'écrire comme intégrale par rapport à la mesure de comptage $\nu := \sum_{k\in\mathbb{N}} \delta_k$. En effet, en notant $f : \mathbb{N} \to \mathbb{C}$, $k \mapsto f(k) := u_k$, on a :*

$$\sum_{k=0}^{+\infty} u_k = \int_\mathbb{N} f \,\mathrm{d}\nu.$$

1. C'est précisément ici que l'hypothèse $a_j > 0$ est importante. En effet si $a_j = 0$ et $\int_\Omega |f| \,\mathrm{d}\mu_j = +\infty$, on a $a_j \int_\Omega |f| \,\mathrm{d}\mu_j = 0$, ce qui n'empêche aucunement la réalisation de (4.3). Par contre $\int_\Omega f \,\mathrm{d}\mu_j$ *n'est pas défini* et (4.4) devient illégitime (ne pas confondre « non défini » et « infini »).

2. Si J est fini, ce sont des somme finies de réels et il n'y a aucune justification à donner.

Preuve. Munissons $\mathbb{N}$ de la tribu $\mathcal{P}(\mathbb{N})$, pour laquelle toute application $f : \mathbb{N} \to \mathbb{C}$ est mesurable. L'intégrabilité de f se réduit alors à la condition $\int_{\mathbb{N}} |f| \, \mathrm{d}\nu < +\infty$. En appliquant la proposition 3.24 à la mesure ν et à la fonction mesurable positive $|f|$, cette condition s'écrit

$$\int_{\mathbb{N}} |f| \, \mathrm{d}\nu = \sum_{k \in \mathbb{N}} \int_{\mathbb{N}} |f| \, \mathrm{d}\delta_k = \sum_{k \in \mathbb{N}} |f|(k) = \sum_{k \in \mathbb{N}} |f(k)| = \sum_{k \in \mathbb{N}} |u_k| < +\infty.$$

Ainsi la ν-intégrabilité de f résulte de la convergence absolue de la série $\sum_{k=0}^{+\infty} u_k$. Par application de (4.4), on a

$$\int_{\mathbb{N}} f \, \mathrm{d}\nu = \sum_{k \in \mathbb{N}} \int_{\mathbb{N}} f \, \mathrm{d}\delta_k = \sum_{k \in \mathbb{N}} f(k) = \sum_{k \in \mathbb{N}} u_k.$$

□

Nous passons maintenant à l'étude des propriétés générales de l'intégrale.

Définition 4.8 (Espace $\mathcal{L}^1(\mu)$)**.** *Pour $\mathbb{K} = \mathbb{R}$ ou $\mathbb{C}$, on note $\mathcal{L}^1_{\mathbb{K}}(\mu)$ l'ensemble des fonctions μ-intégrables $f : \Omega \to \mathbb{K}$.*

Proposition 4.9 (Linéarité de l'intégrale)**.** *Pour $\mathbb{K} = \mathbb{R}$ ou $\mathbb{C}$, l'ensemble $\mathcal{L}^1_{\mathbb{K}}(\mu)$ est un espace vectoriel sur $\mathbb{K}$. L'intégration $f \mapsto \int_\Omega f \, \mathrm{d}\mu$ considérée comme application de $\mathcal{L}^1_{\mathbb{K}}(\mu)$ dans $\mathbb{K}$ est une forme linéaire :*

$$\forall a, b \in \mathbb{K}, \ \forall f, g \in \mathcal{L}^1_{\mathbb{K}}(\mu), \quad \int_\Omega (af + bg) \, \mathrm{d}\mu = a \int_\Omega f \, \mathrm{d}\mu + b \int_\Omega g \, \mathrm{d}\mu. \tag{4.6}$$

On a exclu délibérément le cas $\mathbb{K} = \overline{\mathbb{R}}$ dans la définition de $\mathcal{L}^1_{\mathbb{K}}(\mu)$, car si f et g à valeurs dans $\mathbb{K} = \overline{\mathbb{R}}$ sont μ-intégrables, on peut très bien avoir pour certains ω, $f(\omega) = +\infty$ et $g(\omega) = -\infty$ de sorte que $(f+g)(\omega)$ n'est pas défini et $\mathcal{L}^1_{\overline{\mathbb{R}}}(\mu)$ ne peut pas être un espace vectoriel.

Preuve de la proposition 4.9. Soient f et g dans $\mathcal{L}^1_{\mathbb{K}}(\mu)$ et $h := f + g$. Alors h est mesurable et compte-tenu de la croissance et de l'additivité de l'intégrale dans $\mathcal{M}_+$, l'inégalité $|h| \leq |f| + |g|$ nous donne :

$$\int_\Omega |h| \, \mathrm{d}\mu \leq \int_\Omega (|f| + |g|) \, \mathrm{d}\mu = \int_\Omega |f| \, \mathrm{d}\mu + \int_\Omega |g| \, \mathrm{d}\mu < +\infty.$$

Ainsi h est μ-intégrable et $\mathcal{L}^1_{\mathbb{K}}(\mu)$ est stable par addition. Pour vérifier l'additivité de l'intégrale dans $\mathcal{L}^1_{\mathbb{K}}(\mu)$, regardons d'abord le cas $\mathbb{K} = \mathbb{R}$. On peut alors écrire

$$h^+ - h^- = f^+ - f^- + g^+ - g^-,$$

d'où l'égalité dans $\mathcal{M}_+$:

$$h^+ + f^- + g^- = h^- + f^+ + g^+.$$

Par additivité de l'intégrale dans $\mathcal{M}_+$, on obtient

$$\int_\Omega h^+ \,\mathrm{d}\mu + \int_\Omega f^- \,\mathrm{d}\mu + \int_\Omega g^- \,\mathrm{d}\mu = \int_\Omega h^- \,\mathrm{d}\mu + \int_\Omega f^+ \,\mathrm{d}\mu + \int_\Omega g^+ \,\mathrm{d}\mu$$

et la finitude de ces 6 intégrales nous permet d'écrire :

$$\int_\Omega h^+ \,\mathrm{d}\mu - \int_\Omega h^- \,\mathrm{d}\mu = \int_\Omega f^+ \,\mathrm{d}\mu - \int_\Omega f^- \,\mathrm{d}\mu + \int_\Omega g^+ \,\mathrm{d}\mu - \int_\Omega g^- \,\mathrm{d}\mu,$$

ce qui donne bien $\int_\Omega (f+g)\,\mathrm{d}\mu = \int_\Omega f\,\mathrm{d}\mu + \int_\Omega g\,\mathrm{d}\mu$.

Lorsque $\mathbb{K} = \mathbb{C}$, la définition de l'intégrale : $\int_\Omega h\,\mathrm{d}\mu = \int_\Omega \operatorname{Re} h\,\mathrm{d}\mu + i\int_\Omega \operatorname{Im} h\,\mathrm{d}\mu$ et les égalités $\operatorname{Re}(f+g) = \operatorname{Re} f + \operatorname{Re} g$, $\operatorname{Im}(f+g) = \operatorname{Im} f + \operatorname{Im} g$ nous ramènent au cas précédent.

Soient $f \in \mathcal{L}^1_{\mathbb{K}}(\mu)$ et $a \in \mathbb{K}$. L'intégrabilité de af résulte de la relation $|af| = |a||f|$ et de l'homogénéité de l'intégrale dans $\mathcal{M}_+$:

$$\forall c \in \mathbb{R}_+,\ \forall \varphi \in \mathcal{M}_+, \qquad \int_\Omega c\varphi\,\mathrm{d}\mu = c\int_\Omega \varphi\,\mathrm{d}\mu. \tag{4.7}$$

Donc $af \in \mathcal{L}^1_{\mathbb{K}}(\mu)$ et ceci achève de prouver que $\mathcal{L}^1_{\mathbb{K}}(\mu)$ est bien un espace vectoriel.

Pour vérifier que $\int_\Omega af\,\mathrm{d}\mu = a\int_\Omega f\,\mathrm{d}\mu$, commençons par le cas $\mathbb{K} = \mathbb{R}$. Si $a > 0$, $(af)^+ = af^+$ et $(af)^- = af^-$ d'où grâce à (4.7) :

$$\int_\Omega af\,\mathrm{d}\mu = \int_\Omega af^+\,\mathrm{d}\mu - \int_\Omega af^-\,\mathrm{d}\mu = a\int_\Omega f^+\,\mathrm{d}\mu - a\int_\Omega f^-\,\mathrm{d}\mu = a\int_\Omega f\,\mathrm{d}\mu.$$

Si $a < 0$, $(af)^+ = |a|f^-$ et $(af)^- = |a|f^+$, d'où

$$\begin{aligned}\int_\Omega af\,\mathrm{d}\mu &= \int_\Omega |a|f^-\,\mathrm{d}\mu - \int_\Omega |a|f^+\,\mathrm{d}\mu \\ &= |a|\int_\Omega f^-\,\mathrm{d}\mu - |a|\int_\Omega f^+\,\mathrm{d}\mu = -a\int_\Omega f^-\,\mathrm{d}\mu + a\int_\Omega f^+\,\mathrm{d}\mu = a\int_\Omega f\,\mathrm{d}\mu.\end{aligned}$$

Le cas $a = 0$ est immédiat.

Enfin, prenons $\mathbb{K} = \mathbb{C}$ et notons $\alpha = \operatorname{Re} a$, $\beta = \operatorname{Im} a$. Alors

$$af = (\alpha + i\beta)(\operatorname{Re} f + i\operatorname{Im} f) = (\alpha\operatorname{Re} f - \beta\operatorname{Im} f) + i(\alpha\operatorname{Im} f + \beta\operatorname{Re} f),$$

de sorte qu'en utilisant la définition de l'intégrale dans le cas $\mathbb{K} = \mathbb{C}$ et (4.6) déjà établie dans le cas réel, on obtient

$$\begin{aligned}\int_\Omega af\,\mathrm{d}\mu &= \int_\Omega (\alpha\operatorname{Re} f - \beta\operatorname{Im} f)\,\mathrm{d}\mu + i\int_\Omega (\alpha\operatorname{Im} f + \beta\operatorname{Re} f)\,\mathrm{d}\mu \\ &= (\alpha + i\beta)\int_\Omega \operatorname{Re} f\,\mathrm{d}\mu + (\alpha i - \beta)\int_\Omega \operatorname{Im} f\,\mathrm{d}\mu \\ &= (\alpha + i\beta)\Bigl(\int_\Omega \operatorname{Re} f\,\mathrm{d}\mu + i\int_\Omega \operatorname{Im} f\,\mathrm{d}\mu\Bigr) = a\int_\Omega f\,\mathrm{d}\mu.\end{aligned}$$

Ceci achève la vérification de (4.6) dans le cas complexe. □

Avant de poursuivre, il est utile de reformuler (4.6) dans le cadre probabiliste.

Corollaire 4.10 (Linéarité de l'espérance)**.** *L'espérance des variables aléatoires à valeurs réelles ou complexes, définies sur* $(\Omega, \mathcal{F}, \mathbf{P})$ *et* $\mathbf{P}$*-intégrables est* linéaire.

Il est intéressant de noter qu'avec les définitions de l'espérance données en DEUG, il était *impossible* de prouver cette linéarité (à moins de se restreindre aux variables aléatoires discrètes). En effet dans le cas d'une variable aléatoire à densité g, on définissait l'espérance comme une intégrale $\int_{-\infty}^{+\infty} xg(x)\,\mathrm{d}x$. L'ennui c'est que si X et Y ont des densités, rien ne dit que $X+Y$ soit aussi à densité.

Proposition 4.11. *Avec* $\mathbb{K} = \mathbb{R}$ *ou* $\mathbb{C}$*, on a*

a) Si $f \in \mathcal{L}^1_{\mathbb{K}}(\mu)$*, alors* $|f| \in \mathcal{L}^1_{\mathbb{K}}(\mu)$ *et*

$$\left|\int_\Omega f\,\mathrm{d}\mu\right| \leq \int_\Omega |f|\,\mathrm{d}\mu. \tag{4.8}$$

b) Si $f, g : \Omega \to \mathbb{K}$ *sont deux fonctions mesurables telles que* $|f| \leq |g|$ *et si* g *est* μ*-intégrable, alors* f *l'est aussi.*

c) Si $f, g \in \mathcal{L}^1_{\mathbb{R}}(\mu)$ *sont telles que* $f \leq g$*, alors* $\int_\Omega f\,\mathrm{d}\mu \leq \int_\Omega g\,\mathrm{d}\mu$*.*

Preuve. Pour prouver a), on remarque d'abord que l'intégrabilité de $|f|$ découle immédiatement de la définition de celle de f : si f est mesurable, $|f|$ l'est aussi et les deux fonctions ont même valeur absolue (module). Pour établir (4.8) dans le cas $\mathbb{K} = \mathbb{R}$, il suffit de noter que f^+ et f^- étant positives, leurs intégrales le sont aussi, d'où :

$$\begin{aligned}\left|\int_\Omega f\,\mathrm{d}\mu\right| = \left|\int_\Omega f^+\,\mathrm{d}\mu - \int_\Omega f^-\,\mathrm{d}\mu\right| &\leq \left|\int_\Omega f^+\,\mathrm{d}\mu\right| + \left|\int_\Omega f^-\,\mathrm{d}\mu\right| \\ &= \int_\Omega f^+\,\mathrm{d}\mu + \int_\Omega f^-\,\mathrm{d}\mu = \int_\Omega (f^+ + f^-)\,\mathrm{d}\mu = \int_\Omega |f|\,\mathrm{d}\mu.\end{aligned}$$

Dans le cas $\mathbb{K} = \mathbb{C}$, puisque f est dans $\mathcal{L}^1_{\mathbb{C}}(\mu)$, $w := \int_\Omega f\,\mathrm{d}\mu$ est un nombre complexe. Il existe alors $c \in \mathbb{C}$ tel que $|c| = 1$ et cw soit un réel positif ou nul (si $w \neq 0$, prendre $c = \overline{w}/|w|$, si $w = 0$, $c = 1$ convient). On a ainsi $|w| = |cw| = cw = \int_\Omega cf\,\mathrm{d}\mu \in \mathbb{R}$ d'après (4.6), d'où

$$\left|\int_\Omega f\,\mathrm{d}\mu\right| = \int_\Omega cf\,\mathrm{d}\mu = \mathrm{Re}\Big(\int_\Omega cf\,\mathrm{d}\mu\Big) = \int_\Omega \mathrm{Re}(cf)\,\mathrm{d}\mu \leq \int_\Omega |cf|\,\mathrm{d}\mu = \int_\Omega |f|\,\mathrm{d}\mu.$$

Le b) est quasi immédiat en se souvenant que la mesurabilité de f et g implique celle de $|f|$ et $|g|$ et que ces deux dernières fonctions mesurables positives ont toujours une intégrale élément de $\overline{\mathbb{R}}_+$. L'intégrabilité de f résulte alors de celle de g et de la croissance de l'intégrale dans $\mathcal{M}_+$:

$$\int_\Omega |f|\,\mathrm{d}\mu \leq \int_\Omega |g|\,\mathrm{d}\mu < +\infty.$$

Pour vérifier c), on observe que pour des fonctions à valeurs *réelles* (donc finies en chaque point ω) $f \leq g$ équivaut à $f^+ + g^- \leq g^+ + f^-$. La croissance de l'intégrale dans $\mathcal{M}_+$ nous donne alors

$$\int_\Omega (f^+ + g^-)\,\mathrm{d}\mu \leq \int_\Omega (g^+ + f^-)\,\mathrm{d}\mu$$

Par additivité de l'intégrale et finitude de $\int_\Omega g^- \,\mathrm{d}\mu$ et $\int_\Omega f^- \,\mathrm{d}\mu$, on en déduit

$$\int_\Omega f^+ \,\mathrm{d}\mu - \int_\Omega f^- \,\mathrm{d}\mu \leq \int_\Omega g^+ \,\mathrm{d}\mu - \int_\Omega g^- \,\mathrm{d}\mu,$$

d'où $\int_\Omega f \,\mathrm{d}\mu \leq \int_\Omega g \,\mathrm{d}\mu$. □

Proposition 4.12 (Intégration par rapport à une mesure à densité)**.** *Soient* $(\Omega, \mathcal{F}, \mu)$ *un espace mesuré et* $f : \Omega \to \overline{\mathbb{R}}_+$*, mesurable* $\mathcal{F}$-Bor$(\overline{\mathbb{R}}_+)$*. On note* ν *la mesure de densité* f *par rapport à* μ *(cf. théorème 3.17). Soit* $g : \Omega \to \mathbb{K} = \mathbb{R}$ *ou* $\overline{\mathbb{R}}$*, mesurable* $\mathcal{F}$-Bor$(\mathbb{K})$*. L'application* g *est* ν*-intégrable si et seulement si le produit* gf *est* μ*-intégrable et dans ce cas,*

$$\int_\Omega g \,\mathrm{d}\nu = \int_\Omega gf \,\mathrm{d}\mu. \tag{4.9}$$

On a le même énoncé pour $\mathbb{K} = \mathbb{C}$*, à condition de prendre* f *à valeurs dans* $\mathbb{R}_+$ *au lieu de* $\overline{\mathbb{R}}_+$*.*

Preuve. La condition d'intégrabilité résulte immédiatement du théorème 3.17 appliqué à la fonction mesurable positive $|g|$ en notant que $|g|f = |gf|$. Remarquons au passage que les mesurabilités de gf et de f n'impliquent pas celle de g. L'hypothèse de mesurabilité de g n'est donc pas superflue.

La formule (4.9) se vérifie d'abord pour $\mathbb{K} = \mathbb{R}, \overline{\mathbb{R}}$ en appliquant la formule analogue pour les fonctions mesurables positives à g^+ et g^- et en remarquant que grâce à la positivité de f, $(gf)^+ = g^+ f$ et $(gf)^- = g^- f$. Le cas complexe se ramène au cas réel en séparant partie réelle et imaginaire de g et en notant que f étant à valeurs *réelles*[3], $\mathrm{Re}(gf) = \mathrm{Re}(g)f$ et $\mathrm{Im}(gf) = \mathrm{Im}(g)f$. □

Théorème 4.13 (de transfert)**.** *Soient* $(\Omega_1, \mathcal{F}_1)$ *et* $(\Omega_2, \mathcal{F}_2)$ *deux espaces mesurables et* $\varphi : \Omega_1 \to \Omega_2$ *une application mesurable* $\mathcal{F}_1$-$\mathcal{F}_2$*. Soit* μ *une mesure sur* $(\Omega_1, \mathcal{F}_1)$ *et* $\nu := \mu \circ \varphi^{-1}$ *sa mesure image sur* $(\Omega_2, \mathcal{F}_2)$*. Soit* h *une application* $h : \Omega_2 \to \mathbb{K}$, $\mathbb{K} = \mathbb{R}$, $\overline{\mathbb{R}}$ *ou* $\mathbb{C}$*, mesurable* $\mathcal{F}$-Bor$(\mathbb{K})$*. Alors* h *est* ν*-intégrable si et seulement si* $h \circ \varphi$ *est* μ*-intégrable et dans ce cas,*

$$\int_{\Omega_1} (h \circ \varphi) \,\mathrm{d}\mu = \int_{\Omega_2} h \,\mathrm{d}\nu. \tag{4.10}$$

Preuve. L'équivalence entre la μ-intégrabilité de $h \circ \varphi$ et la ν-intégrabilité de h est une conséquence directe du théorème de transfert dans $\mathcal{M}_+$ et de la relation $|h \circ \varphi| = |h| \circ \varphi$. Dans le cas $\mathbb{K} = \mathbb{R}$ ou $\overline{\mathbb{R}}$, (4.10) se vérifie en notant que $(h \circ \varphi)^+ = h^+ \circ \varphi$, et $(h \circ \varphi)^- = h^- \circ \varphi$ et en appliquant le théorème de transfert dans $\mathcal{M}^+$ à h^+ et h^-. Le cas complexe se ramène au cas réel par séparation des parties réelles et imaginaires en notant que $\mathrm{Re}(h \circ \varphi) = (\mathrm{Re}\, h) \circ \varphi$ et $\mathrm{Im}(h \circ \varphi) = (\mathrm{Im}\, h) \circ \varphi$. □

Corollaire 4.14 (Calculs d'espérances et de moments)**.** *Soit* X *une variable aléatoire réelle définie sur l'espace probabilisé* $(\Omega, \mathcal{F}, \mathbf{P})$ *et* P_X *sa loi. Alors* X *est* $\mathbf{P}$*-intégrable si et seulement si* $\int_\mathbb{R} |x| \,\mathrm{d}P_X(x) < +\infty$ *et dans ce cas son espérance est donnée par*

$$\mathbf{E}X := \int_\Omega X \,\mathrm{d}\mathbf{P} = \int_\mathbb{R} x \,\mathrm{d}P_X(x). \tag{4.11}$$

3. Si on autorisait $f(\omega)$ à valoir $+\infty$, on aurait un problème de définition avec le produit $g(\omega)f(\omega)$, puisque nous n'avons pas donné de sens à $z \times (+\infty)$ pour $z \in \mathbb{C}$ quelconque.

Plus généralement, si h est une application borélienne de $\mathbb{R}$ dans $\mathbb{R}$, la variable aléatoire $Y = h(X)$ est $\mathbf{P}$-intégrable si et seulement si $\int_{\mathbb{R}} |h(x)| \, \mathrm{d}P_X(x) < +\infty$ et dans ce cas

$$\mathbf{E}h(X) := \int_\Omega h(X) \, \mathrm{d}\mathbf{P} = \int_{\mathbb{R}} h(x) \, \mathrm{d}P_X(x) = \int_{\mathbb{R}} y \, \mathrm{d}P_Y(y). \tag{4.12}$$

Lorsqu'elle existe, la quantité $\mathbf{E}h(X)$ est appelée moment fonctionnel de X (pour la fonction h). Lorsque $h(x) = x^n$, on parle plus simplement de moment d'ordre n.

Preuve. Pour la c.n.s. d'intégrabilité de $Y = h(X)$ et le calcul de son espérance par (4.12), il suffit d'appliquer le théorème de transfert avec $\Omega_1 = \Omega$, $\mathcal{F}_1 = \mathcal{F}$, $\Omega_2 = \mathbb{R}$, $\mathcal{F}_2 = \mathrm{Bor}(\mathbb{R})$ et $\varphi = X$. Le cas où h est l'identité sur $\mathbb{R}$ ($h(x) = x$) donne la condition d'existence et le calcul de $\mathbf{E}X$ par (4.11)... et la dernière égalité dans (4.12) en prenant cette fois $\varphi = Y$. □

Remarque 4.15. Sur le modèle du corollaire 4.14, on peut décliner plusieurs variantes dont la preuve s'obtient par une adaptation immédiate de celle du corollaire. En voici deux.

a) La variable aléatoire complexe Z sur $(\Omega, \mathcal{F}, \mathbf{P})$ est $\mathbf{P}$-intégrable si et seulement si $\int_{\mathbb{C}} |z| \, \mathrm{d}P_Z(z) < +\infty$ et dans ce cas, $\mathbf{E}Z = \int_{\mathbb{C}} z \, \mathrm{d}P_Z(z)$.

b) Soit $X : \Omega \to \mathbb{R}^d$ un vecteur aléatoire et $h : \mathbb{R}^d \to \mathbb{K} = \mathbb{R}$ ou $\mathbb{C}$, borélienne. La variable aléatoire réelle ou complexe $Y = h(X)$ est $\mathbf{P}$-intégrable si et seulement si $\int_{\mathbb{R}^d} |h(x)| \, \mathrm{d}P_X(x) < +\infty$ et dans ce cas

$$\mathbf{E}h(X) := \int_\Omega h(X) \, \mathrm{d}\mathbf{P} = \int_{\mathbb{R}^d} h(x) \, \mathrm{d}P_X(x) = \int_{\mathbb{K}} y \, \mathrm{d}P_Y(y).$$

Corollaire 4.16 (Espérance d'une variable aléatoire discrète)**.** *Soit $(\Omega, \mathcal{F}, \mathbf{P})$ un espace probabilisé et X une variable aléatoire discrète sur $(\Omega, \mathcal{F})$, à valeurs dans $\mathbb{K} = \mathbb{R}$ ou $\mathbb{C}$. Alors X est $\mathbf{P}$-intégrable si et seulement si*

$$\sum_{x \in X(\Omega)} |x| \mathbf{P}(X = x) < +\infty \tag{4.13}$$

et dans ce cas l'espérance de X sous $\mathbf{P}$ peut se calculer par

$$\mathbf{E}X = \sum_{x \in X(\Omega)} x \mathbf{P}(X = x). \tag{4.14}$$

Si h est une application quelconque $\mathbb{K} \to \mathbb{C}$, $Y := h(X)$ est encore une variable aléatoire discrète. Elle est $\mathbf{P}$-intégrable si et seulement si

$$\sum_{x \in X(\Omega)} |h(x)| \mathbf{P}(X = x) < +\infty \tag{4.15}$$

et dans ce cas l'espérance de Y sous $\mathbf{P}$ peut se calculer par

$$\mathbf{E}h(X) = \sum_{x \in X(\Omega)} h(x) \mathbf{P}(X = x). \tag{4.16}$$

Preuve. L'ensemble $X(\Omega) = \{x_0, x_1, \dots, x_n, \dots\}$ est ici une partie au plus dénombrable de $\mathbb{R}$ ou $\mathbb{C}$. Selon la proposition 2.27, la loi de X est donnée par

$$P_X = \sum_{x_j \in X(\Omega)} \mathbf{P}(X = x_j)\delta_{x_j}.$$

Il suffit alors d'appliquer le corollaire 4.14, en utilisant les propositions 4.6 et 4.5 pour le calcul des intégrales relatives à P_X. Deux précisions s'imposent ici. Nous n'avons pas besoin de l'hypothèse h borélienne, car X est mesurable $\mathcal{F}$-$\mathcal{P}(\mathbb{K})$, donc pour n'importe quelle tribu $\mathcal{G}$ sur $\mathbb{C}$, h étant automatiquement mesurable $\mathcal{P}(\mathbb{K})$-$\mathcal{G}$, $h \circ X$ est mesurable $\mathcal{F}$-$\mathcal{G}$. Ceci est vrai en particulier pour $\mathcal{G} = \operatorname{Bor}(\mathbb{C})$. En fait on adapte légèrement la preuve du corollaire 4.14 en prenant $\mathcal{F}_2 = \mathcal{P}(\mathbb{K})$ au lieu de $\mathcal{F}_2 = \operatorname{Bor}(\mathbb{K})$ pour la mesurabilité $\mathcal{F}$-$\mathcal{F}_2$ de X dans l'application du théorème de transfert. D'autre part dans l'application de la proposition 4.6, il n'y a pas lieu d'écarter les indices j tels que $a_j := \mathbf{P}(X = x_j) = 0$. En effet, ici $\mu_j = \delta_{x_j}$ et h étant à valeurs dans $\mathbb{R}$ ou $\mathbb{C}$, est automatiquement δ_{x_j}-intégrable (voir note 1 p. 108). □

Remarque 4.17. En DEUG (voir [ICP, 5.1]), on avait utilisé (4.13) et (4.14) pour *définir* l'existence et la valeur de $\mathbf{E}X$. Le remplacement de cette définition initiale de $\mathbf{E}X$ par l'intégrale abstraite $\int_\Omega X \, \mathrm{d}\mathbf{P}$ permet *d'unifier la théorie de l'espérance* et de bénéficier des bonnes propriétés de l'intégrale abstraite. Par exemple si X est une variable aléatoire discrète et Y une variable aléatoire à densité, ayant chacune une espérance, la variable aléatoire $Z = X + Y$ qui n'est en général ni discrète ni à densité, a aussi une espérance et $\mathbf{E}Z = \mathbf{E}X + \mathbf{E}Y$.

Exemple 4.1 (Fonction caractéristique d'un vecteur aléatoire)**.** Soit $X : \Omega \to \mathbb{R}^d$ un vecteur aléatoire de composantes $X_1, \dots, X_d$, sur l'espace probabilisé $(\Omega, \mathcal{F}, \mathbf{P})$. On appelle *fonction caractéristique* de X l'application $\varphi_X : \mathbb{R}^d \to \mathbb{C}$, définie par

$$\forall t = (t_1, \dots, t_d) \in \mathbb{R}^d, \quad \varphi_X(t) := \mathbf{E} \exp(i\langle t, X\rangle) = \mathbf{E} \exp\Big(i(t_1 X_1 + \dots + t_d X_d)\Big). \tag{4.17}$$

Cette fonction caractéristique joue un grand rôle en théorie des Probabilités, notamment dans l'étude de la convergence en loi. Justifions l'existence de $\varphi_X(t)$. Pour t fixé, l'application $h := \exp(i\langle t, .\rangle) : \mathbb{R}^d \to \mathbb{C}$ est continue donc borélienne et bornée donc $h(X)$ est une variable aléatoire complexe bornée, donc $\mathbf{P}$-intégrable sur Ω. En notant P_X la loi du vecteur aléatoire X, on a en appliquant la remarque 4.15 b)

$$\varphi_X(t) = \int_{\mathbb{R}^d} \exp(i\langle t, x\rangle) \, \mathrm{d}P_X(x). \tag{4.18}$$

En particulier si X a une densité f par rapport à la mesure de Lebesgue,

$$\varphi_X(t) = \int_{\mathbb{R}^d} \exp(i\langle t, x\rangle) f(x) \, \mathrm{d}\lambda_d(x). \tag{4.19}$$

La poursuite du calcul explicite de φ_X nécessite alors de savoir calculer pratiquement une intégrale par rapport à λ_d, ce que nous étudierons au chapitre 5.

Dans le cas où X est une variable aléatoire discrète, nous disposons déjà de l'outillage nécessaire pour achever le calcul de φ_X. Par exemple si X suit la loi de Poisson de paramètre α, en appliquant le corollaire 4.16, on obtient

$$\forall t \in \mathbb{R}, \quad \varphi_X(t) = \mathbf{E} \exp(itX) = \sum_{k\in\mathbb{N}} \frac{\mathrm{e}^{-\alpha}\alpha^k}{k!}\mathrm{e}^{itk} = \mathrm{e}^{-\alpha}\sum_{k\in\mathbb{N}} \frac{(\alpha\mathrm{e}^{it})^k}{k!} = \exp\Big(\alpha(\mathrm{e}^{it}-1)\Big).$$

4.2 Négligeabilité

Deux fonctions intégrables qui diffèrent seulement sur un ensemble de mesure nulle ont même intégrale. Ceci motive la recherche d'une plus grande souplesse dans la théorie de l'intégration, notamment en remplaçant dans les hypothèses des théorèmes d'interversion limite intégrale, les conditions requises *en tout point* de Ω par leurs analogues *presque partout*. Une autre conséquence importante est que l'application $f \mapsto \int_\Omega |f| \,\mathrm{d}\mu$ n'est qu'une semi-norme sur $\mathcal{L}^1_{\mathbb{K}}(\mu)$. Ceci nous amènera à quotienter $\mathcal{L}^1_{\mathbb{K}}(\mu)$ par l'espace des fonctions mesurables nulles presque partout de façon à en faire un espace vectoriel normé (qui sera de plus complet).

4.2.1 Ensembles négligeables

Une propriété (π) relative à certains éléments de Ω peut toujours être vue comme une application π de Ω dans l'ensemble booléen $\{\mathrm{Faux}, \mathrm{Vrai}\}$ que l'on peut munir de la tribu $\mathcal{P}$ de toutes ses parties. On dit alors que la propriété (π) est $\mathcal{F}$-mesurable si l'application π est $\mathcal{F}$-$\mathcal{P}$ mesurable. Ceci est équivalent à $A := \pi^{-1}(\mathrm{Vrai}) \in \mathcal{F}$.

Définition 4.18. *Soit* $(\Omega, \mathcal{F}, \mu)$ *un espace mesuré.*

a) Un élément A de la tribu $\mathcal{F}$ est une partie μ-négligeable *(ou simplement négligeable s'il n'y a pas d'ambiguité sur la mesure μ) si* $\mu(A) = 0$.

b) Une propriété $\mathcal{F}$-mesurable (π) est dite vraie μ-presque partout si $\mu\Big(\pi^{-1}(\mathrm{Faux})\Big) = 0$. *Si $\mu = \mathbf{P}$ est une mesure de probabilité, on parle de propriété vraie $\mathbf{P}$-presque-sûrement. Dans ce cas* $\mathbf{P}\Big(\pi^{-1}(\mathrm{Faux})\Big) = 0$ *est équivalent à* $\mathbf{P}\Big(\pi^{-1}(\mathrm{Vrai})\Big) = 1$.

c) Une fonction $f : \Omega \to \mathbb{K}$, $\mathbb{K} = \mathbb{R}$ ou $\mathbb{C}$ est négligeable si elle est $\mathcal{F}$-$\mathrm{Bor}(\mathbb{K})$ *mesurable et nulle μ-presque partout.*

Remarquons que dans le cas d'une mesure μ telle que $\mu(\Omega) = +\infty$, on peut avoir $\mu\Big(\pi^{-1}(\mathrm{Vrai})\Big) = \mu(\Omega)$ sans que (π) soit vraie μ-presque partout.

Dans la suite, on abrègera « μ-presque partout » en μ-p.p. ou en p.p. s'il n'y a pas d'ambiguité sur la mesure μ concernée.

Si A est une partie mesurable (i.e. $A \in \mathcal{F}$) d'un ensemble négligeable, A est aussi négligeable. Toute réunion dénombrable d'ensembles négligeables est négligeable.

4.2.2 Finitude presque partout

Un exemple important de propriété vraie presque partout est la finitude des fonctions intégrables.

Proposition 4.19. *Soit* $f : \Omega \to \overline{\mathbb{R}}$ *une application* μ*-intégrable. Alors* f *est* finie μ *presque partout.*

La preuve de ce résultat très utile repose sur le lemme suivant.

Lemme 4.20. *Soit* $g : \Omega \to \overline{\mathbb{R}}_+$ *mesurable positive. Pour tout réel* $t > 0$,

$$\mu(\{g \geq t\}) \leq \frac{1}{t} \int_\Omega g \, \mathrm{d}\mu. \tag{4.20}$$

Dans la théorie des Probabilités, ce lemme est connu sous le nom d'inégalité de Markov. Nous le traduisons dans ce langage pour la commodité de référence ultérieure :

Lemme 4.21 (Inégalité de Markov)**.** *Soit* Y *une variable aléatoire positive (à valeurs dans* $\overline{\mathbb{R}}_+$*) sur l'espace probabilisé* $(\Omega, \mathfrak{F}, \mathbf{P})$*. Alors*

$$\forall t > 0, \quad \mathbf{P}(Y \geq t) \leq \frac{\mathbf{E}Y}{t}.$$

Remarque 4.22. L'inégalité de Markov n'a d'intérêt que si $\mathbf{E}Y < +\infty$ et dans ce cas, seulement pour $t > \mathbf{E}Y$. Pour n'importe quelle variable aléatoire positive Y telle que $\mathbf{P}(Y = +\infty) = 0$, $\mathbf{P}(Y > t)$ tend vers 0 quand t tend vers $+\infty$ par continuité séquentielle décroissante de $\mathbf{P}$. L'inégalité de Markov nous dit que si de plus $\mathbf{E}Y$ est fini, cette convergence a lieu (au moins) avec la vitesse $O(t^{-1})$. En fait on peut raffiner et montrer que cette vitesse est au moins $o(t^{-1})$, cf. T.D.

Preuve du lemme 4.20. Comme g est mesurable, $A := \{g \geq t\} \in \mathfrak{F}$ et comme g est positive et minorée par la constante t sur A :

$$\int_\Omega g \, \mathrm{d}\mu \geq \int_A g \, \mathrm{d}\mu \geq \int_A t \, \mathrm{d}\mu = t\mu(\{g \geq t\}).$$

On en déduit (4.20) en divisant par t. □

Preuve de la proposition 4.19. Définissons les ensembles A_n et A dans $\mathfrak{F}$ par

$$\forall n \in \mathbb{N}^*, \quad A_n := \{|f| \geq n\}, \qquad A := \{|f| = +\infty\}.$$

On a clairement $A \subset A_n$. En appliquant le lemme 4.20 on obtient

$$0 \leq \mu(A) \leq \mu(A_n) \leq \frac{1}{n} \int_\Omega |f| \, \mathrm{d}\mu$$

et comme $\int_\Omega |f| \, \mathrm{d}\mu < +\infty$ (par intégrabilité de f), ce majorant tend vers 0 quand n tend vers $+\infty$. □

Corollaire 4.23. *Soit (f_n) une suite de fonctions mesurables positives (à valeurs dans $\mathbb{R}_+$ ou $\overline{\mathbb{R}}_+$) telle que*

$$\sum_{n=0}^{+\infty} \int_\Omega f_n \, \mathrm{d}\mu < +\infty.$$

Alors $f := \sum_{n=0}^{+\infty} f_n$ est finie μ-presque partout sur Ω.

Preuve. Il suffit d'appliquer la proposition 4.19, en notant que grâce au théorème d'interversion intégrale série dans $\mathcal{M}_+$ (corollaire du théorème de Beppo Levi),

$$\int_\Omega f \, \mathrm{d}\mu = \sum_{n=0}^{+\infty} \int_\Omega f_n \, \mathrm{d}\mu < +\infty.$$

□

En prenant pour f_n des fonctions indicatrices dans le corollaire 4.23, on obtient un résultat connu dans la théorie des Probabilités sous le nom de *premier lemme de Borel Cantelli.*

Corollaire 4.24 (Borel-Cantelli)**.** *Soit $(A_n)_{n\geq 1}$ une suite d'évènements sur l'espace probabilisé $(\Omega, \mathcal{F}, \mathbf{P})$ et A l'ensemble des ω qui appartiennent à une infinité d'évènements de la suite. On suppose que*

$$\sum_{n=1}^{+\infty} \mathbf{P}(A_n) < +\infty.$$

Alors $\mathbf{P}(A) = 0$.

Preuve. On pose $f_n = \mathbf{1}_{A_n}$ et on remarque que

$$A = \Big\{ \omega \in \Omega;\ \sum_{n=1}^{+\infty} \mathbf{1}_{A_n}(\omega) = +\infty \Big\}.$$

Alors

$$\sum_{n=1}^{+\infty} \int_\Omega f_n \, \mathrm{d}\mathbf{P} = \sum_{n=1}^{+\infty} \mathbf{P}(A_n) < +\infty.$$

Par le corollaire 4.23, $f := \sum_{n=1}^{+\infty} f_n$ est finie $\mathbf{P}$ presque partout, donc $A = \{f = +\infty\}$ est de mesure nulle. □

On peut écrire l'évènement A avec des opérations ensemblistes dénombrables à partir des A_n. En effet ω appartient à A si et seulement si il existe une infinité d'entiers k tels que $\omega \in A_k$. Ceci s'écrit

$$\forall n \in \mathbb{N}, \exists k \geq n;\ \omega \in A_k.$$

Avec la traduction automatique des quantificateurs, ceci s'écrit encore $\omega \in \bigcap_{n\in\mathbb{N}} \bigcup_{k\geq n} A_k$. Ainsi

$$A = \bigcap_{n\in\mathbb{N}} \bigcup_{k\geq n} A_k. \tag{4.21}$$

Cette écriture fait penser à la définition de la limite supérieure d'une suite de réels : $\limsup u_n = \inf_{n\in\mathbb{N}} \sup_{k\geq n} u_k$, à condition de remplacer intersection par inf et union par sup. Or pour la relation d'ordre partiel définie par l'inclusion, l'intersection donne justement la borne inférieure d'une famille d'ensembles et la réunion la borne supérieure. Ceci nous conduit à poser

$$\limsup A_n := \bigcap_{n\in\mathbb{N}} \bigcup_{k\geq n} A_k.$$

De manière analogue, on définit

$$\liminf A_n := \bigcup_{n\in\mathbb{N}} \bigcap_{k\geq n} A_k.$$

En traduisant en sens inverse les opérations ensemblistes avec des quantificateurs, on voit que ω appartient à $\liminf A_n$ si et seulement si $\exists n \in \mathbb{N}, \forall k \geq n,\ \omega \in A_k$. Autrement dit, $\liminf A_n$ est l'ensemble des ω qui appartiennent à *tous* les A_k à partir d'un certain rang $n = n(\omega)$.

En passant au complémentaire dans (4.21), on obtient

$$A^c = \bigcup_{n\in\mathbb{N}} \bigcap_{k\geq n} A_k^c = \liminf A_n^c.$$

En revenant à la conclusion du lemme de Borel Cantelli, $\mathbf{P}(A) = 0$ équivaut à $\mathbf{P}(A^c) = 1$ et cette conclusion peut alors s'interpréter ainsi : « presque sûrement, plus aucun des évènements A_n ne se réalise au-delà d'un certain rang (aléatoire) ». Nous verrons ultérieurement des exemples d'application du lemme de Borel Cantelli.

4.2.3 Égalité presque partout

Définition 4.25. *Soit* $(\Omega, \mathcal{F}, \mu)$ *une espace mesuré et* f, g *deux fonctions* $\Omega \to \mathbb{K}$, $\mathbb{K} = \mathbb{R}$, $\overline{\mathbb{R}}$ *ou* $\mathbb{C}$, $\mathcal{F}$-Bor$(\mathbb{K})$ *mesurables. On dit qu'elles sont égales* μ*-presque partout lorsque* $\{f \neq g\}$ *est négligeable. Notations :*

$$f = g \quad \mu\text{−p.p.} \qquad \textit{ou} \qquad f \overset{\text{p.p.}}{=} g,$$

s'il n'y a pas d'ambiguïté sur la mesure μ *concernée. En particulier,* f *est dite négligeable si elle est égale* μ*-p.p. à zéro.*

Proposition 4.26. *Pour* $\mathbb{K} = \mathbb{R}$ *ou* $\mathbb{C}$*, la relation d'égalité* μ*-p.p. est une relation d'équivalence sur l'ensemble des fonctions mesurables. Elle est compatible avec l'addition et la multiplication par un scalaire.*

Preuve. La relation est évidemment réflexive ($f \overset{\text{p.p.}}{=} f$), symétrique ($f \overset{\text{p.p.}}{=} g \Rightarrow g \overset{\text{p.p.}}{=} f$). Montrons qu'elle est transitive ($f \overset{\text{p.p.}}{=} g$ et $g \overset{\text{p.p.}}{=} h$ impliquent $f \overset{\text{p.p.}}{=} h$). Les ensembles $A := \{f \neq g\}$ et $B := \{g \neq h\}$ sont négligeables, donc aussi leur réunion. Il suffit alors de remarquer que pour tout $\omega \in (A \cup B)^c$, $f(\omega) = g(\omega) = h(\omega)$.

Si $f_1 \overset{\text{p.p.}}{=} g_1$ et $f_2 \overset{\text{p.p.}}{=} g_2$, $A := \{f_1 \neq g_1\}$ et $B := \{f_2 \neq g_2\}$ sont négligeables, donc aussi leur réunion. Pour tout $\omega \in (A \cup B)^c$, $f_1(\omega) = g_1(\omega)$ et $f_2(\omega) = g_2(\omega)$ donc $(f_1+f_2)(\omega) = (g_1+g_2)(\omega)$ d'où $f_1+f_2 \overset{\text{p.p.}}{=} g_1+g_2$. De même sur A^c, on a $cf_1(\omega) = cg_1(\omega)$ pour toute constante c, donc $cf_1 \overset{\text{p.p.}}{=} cg_1$. □

Proposition 4.27. *Soient (f_n) et (g_n) sont deux suites d'applications mesurables à valeurs dans $\mathbb{R}$ ou $\overline{\mathbb{R}}$.*

a) Si $f_1 \overset{\text{p.p.}}{=} g_1$ et $f_2 \overset{\text{p.p.}}{=} g_2$, alors $\min(f_1, f_2) \overset{\text{p.p.}}{=} \min(g_1, g_2)$ et $\max(f_1, f_2) \overset{\text{p.p.}}{=} \max(g_1, g_2)$.

b) Si pour tout n, $f_n \overset{\text{p.p.}}{=} g_n$, alors

$$\inf_{n\in\mathbb{N}} f_n \overset{\text{p.p.}}{=} \inf_{n\in\mathbb{N}} g_n \qquad \sup_{n\in\mathbb{N}} f_n \overset{\text{p.p.}}{=} \sup_{n\in\mathbb{N}} g_n \tag{4.22}$$

$$\liminf_{n\to+\infty} f_n \overset{\text{p.p.}}{=} \liminf_{n\to+\infty} g_n \qquad \limsup_{n\to+\infty} f_n \overset{\text{p.p.}}{=} \limsup_{n\to+\infty} g_n \tag{4.23}$$

c) Si de plus $\lim\limits_{n\to+\infty} f_n \overset{\text{p.p.}}{=} f$, avec f mesurable, $\lim\limits_{n\to+\infty} g_n \overset{\text{p.p.}}{=} f$.

Preuve. Le a) est un cas particulier de (4.22). Notons pour tout $n \in \mathbb{N}$, $A_n := \{\omega \in \Omega;\ f_n(\omega) \neq g_n(\omega)\}$. Alors chaque A_n est un ensemble négligeable. La réunion dénombrable $A = \bigcup\limits_{n\in\mathbb{N}} A_n$ l'est donc aussi et

$$\forall \omega \in N^c,\ \forall n \in \mathbb{N}, \quad f_n(\omega) = g_n(\omega).$$

Le b) et le c) en résultent immédiatement. □

Proposition 4.28. *Une fonction mesurable $f : \Omega \to \mathbb{K}$ ($\mathbb{K} = \mathbb{R}$, $\overline{\mathbb{R}}$ ou $\mathbb{C}$) est négligeable si et seulement si $\int_\Omega |f| \,\mathrm{d}\mu = 0$.*

Preuve. Notons $A := \{\omega \in \Omega;\ |f(\omega)| \neq 0\}$.

Si $\int_\Omega |f| \,\mathrm{d}\mu = 0$, notons pour $n \in \mathbb{N}^*$, $A_n := \{\omega \in \Omega;\ |f(\omega)| \geq 1/n\}$. Par le lemme 4.20

$$\forall n \in \mathbb{N}^*, \quad \mu(A_n) \leq n \int_\Omega |f| \,\mathrm{d}\mu = 0.$$

Comme A est limite croissante (pour l'inclusion) de la suite (A_n), on a par continuité croissante séquentielle de μ, $\mu(A) = \lim_{n\to+\infty} \mu(A_n) = 0$. Donc f est μ-négligeable.

Réciproquement, si f est négligeable, supposons d'abord $|f|$ bornée sur tout Ω par une constante $M < +\infty$. Alors il suffit d'intégrer l'inégalité $|f| \leq M\mathbf{1}_A$ pour obtenir $\int_\Omega |f| \,\mathrm{d}\mu \leq M\mu(A) = 0$. Si $|f|$ n'est pas bornée, on prend $g_n := \min(n; |f|)$ et par le cas $|f|$ bornée, $\int_\Omega g_n \,\mathrm{d}\mu = 0$. On conclut avec le théorème de Beppo Levi puisque $g_n \uparrow |f|$. □

Corollaire 4.29.

i) Soient f et g deux fonctions mesurables $\Omega \to \overline{\mathbb{R}}_+$, égales μ-p.p. Alors $\int_\Omega f \,\mathrm{d}\mu = \int_\Omega g \,\mathrm{d}\mu$.

ii) Si $f \in \mathcal{L}^1_{\mathbb{K}}(\mu)$ ($\mathbb{K} = \mathbb{R}$ ou $\mathbb{C}$) et si g est une fonction mesurable égale à f p.p. (éventuellement g à valeurs dans $\overline{\mathbb{R}}$ quand $\mathbb{K} = \mathbb{R}$), g est intégrable et $\int_\Omega f \,\mathrm{d}\mu = \int_\Omega g \,\mathrm{d}\mu$.

Preuve. Dans les égalités

$$f = f\mathbf{1}_{\{f=g\}} + f\mathbf{1}_{\{f\neq g\}} \qquad g = g\mathbf{1}_{\{f=g\}} + g\mathbf{1}_{\{f\neq g\}},$$

les termes $f\mathbf{1}_{\{f\neq g\}}$ et $g\mathbf{1}_{\{f\neq g\}}$ sont des fonctions négligeables mesurables positives. Par la proposition 4.28, leurs intégrales sont nulles. Par additivité de l'intégrale dans $\mathcal{M}_+$ nous avons ainsi :

$$\int_\Omega f\,\mathrm{d}\mu = \int_{\{f=g\}} f\,\mathrm{d}\mu + \int_{\{f\neq g\}} f\,\mathrm{d}\mu = \int_{\{f=g\}} g\,\mathrm{d}\mu = \int_{\{f=g\}} g\,\mathrm{d}\mu + \int_{\{f\neq g\}} g\,\mathrm{d}\mu = \int_\Omega g\,\mathrm{d}\mu.$$

Pour vérifier ii), considérons d'abord le cas $\mathbb{K} = \mathbb{R}$. Par la proposition 4.27 a), $f^+ \overset{\text{p.p.}}{=} g^+$ et $f^- \overset{\text{p.p.}}{=} g^-$. On a donc d'après le i) $\int_\Omega f^+\,\mathrm{d}\mu = \int_\Omega g^+\,\mathrm{d}\mu$ et $\int_\Omega f^-\,\mathrm{d}\mu = \int_\Omega g^-\,\mathrm{d}\mu$. L'intégrabilité de f entraîne la finitude de ces 4 intégrales donc l'intégrabilité de g. On a alors l'égalité $\int_\Omega f\,\mathrm{d}\mu = \int_\Omega (g^+ - g^-)\,\mathrm{d}\mu = \int_\Omega g\,\mathrm{d}\mu$. Le cas complexe se ramène au cas $\mathbb{K} = \mathbb{R}$ par séparation des parties réelles et imaginaires. □

4.3 L'espace de Lebesgue $\mathrm{L}^1(\mu)$

Pour $\mathbb{K} = \mathbb{R}$ ou $\mathbb{C}$, notons $\mathcal{N}_\mathbb{K}(\mu)$ l'ensemble des applications mesurables négligeables $\Omega \to \mathbb{K}$. Grâce à la proposition 4.26, on voit immédiatement que $\mathcal{N}_\mathbb{K}(\mu)$ est un espace vectoriel sur $\mathbb{K}$. C'est un sous-espace vectoriel de $\mathcal{L}^1_\mathbb{K}(\mu)$ par la proposition 4.28. La relation d'équivalence qu'est l'égalité μ-p.p. sur $\mathcal{L}^1_\mathbb{K}(\mu)$ peut ainsi s'écrire

$$\forall f, g \in \mathcal{L}^1_\mathbb{K}(\mu), \qquad f \overset{\text{p.p.}}{=} g \ \text{ si et seulement si } \ f - g \in \mathcal{N}_\mathbb{K}(\mu).$$

Définition 4.30. *Soit $(\Omega, \mathcal{F}, \mu)$ un espace mesuré et $\mathbb{K} = \mathbb{R}$ ou $\mathbb{C}$. On note $\mathrm{L}^1_\mathbb{K}(\mu)$ l'espace vectoriel quotient $\mathcal{L}^1_\mathbb{K}(\mu)/\mathcal{N}_\mathbb{K}(\mu)$ de l'espace des fonctions μ-intégrables $\Omega \to \mathbb{K}$ par son sous-espace des applications μ-négligeables.*

La notation $\mathrm{L}^1_\mathbb{K}(\mu)$ peut être remplacée, lorsque le contexte l'exige ou le permet, par des notations encore plus explicites comme $\mathrm{L}^1_\mathbb{K}(\Omega, \mu)$, voire $\mathrm{L}^1_\mathbb{K}(\Omega, \mathcal{F}, \mu)$ ou plus brèves comme $\mathrm{L}^1(\mu)$ ou L^1, etc.

Remarque 4.31. Un élément de $\mathrm{L}^1_\mathbb{K}(\mu)$ est une classe d'équivalence de fonctions μ-intégrables pour l'égalité μ-presque partout. Si f et g sont deux fonctions membres de la même classe de $\mathrm{L}^1_\mathbb{K}(\mu)$, alors $f \overset{\text{p.p.}}{=} g$ et clairement pour tout $A \in \mathcal{F}$, $f\mathbf{1}_A \overset{\text{p.p.}}{=} g\mathbf{1}_A$. On a donc par le corollaire 4.29, $\int_A f\,\mathrm{d}\mu = \int_A g\,\mathrm{d}\mu$. Il en résulte que f et g sont indistinguables pour tout ce qui concerne le calcul d'intégrales (par rapport à μ).

Remarque 4.32. La construction de L^1 exposée ci-dessus exclut les fonctions à valeurs dans $\overline{\mathbb{R}}$. Rappelons que l'on ne peut pas munir l'ensemble des fonctions $\Omega \to \overline{\mathbb{R}}$, μ-intégrables d'une structure de $\mathbb{R}$ espace vectoriel car on ne peut garantir que la somme de deux telles fonctions soit définie sur tout Ω.

Néanmoins, si $g : \Omega \to \overline{\mathbb{R}}$ est μ-intégrable, la fonction $f := g\mathbf{1}_{\{|g|<+\infty\}}$ est égale μ-p.p. à g et est dans $\mathcal{L}^1_\mathbb{R}(\mu)$: en effet par la proposition 4.19, $\{f \neq g\}$ est négligeable et par le corollaire 4.29 i), appliqué à $|f|$ et $|g|$, f est μ intégrable. On s'autorisera donc parfois à parler par abus de langage de la classe de g dans $\mathrm{L}^1_\mathbb{R}(\mu)$.

Définition 4.33. *Soit φ un élément de $\mathrm{L}^1_{\mathbb{K}}(\mu)$. On définit son intégrale sur $A \in \mathcal{F}$ par*

$$\int_A \varphi \, \mathrm{d}\mu := \int_A f \, \mathrm{d}\mu,$$

où $f \in \mathcal{L}^1_{\mathbb{K}}(\mu)$ est un représentant quelconque de la classe φ.

D'après la remarque 4.31, cette définition est cohérente : la valeur de $\int_A \varphi \, \mathrm{d}\mu$ ne dépend pas du choix du représentant f de φ. Dans le même esprit, on peut grâce à la proposition 4.27, définir φ^+, φ^-, $|\varphi|$.

Proposition 4.34. $\mathrm{L}^1_{\mathbb{K}}(\mu)$ *est un espace vectoriel* normé *par*

$$\|\varphi\|_1 := \int_\Omega |\varphi| \, \mathrm{d}\mu. \tag{4.24}$$

L'espace $\mathrm{L}^1_{\mathbb{K}}(\mu)$ sera toujours supposé muni de cette norme. L'application $\mathrm{L}^1_{\mathbb{K}}(\mu) \to \mathbb{K}$, $\varphi \mapsto \int_\Omega \varphi \, \mathrm{d}\mu$ est une forme linéaire continue *sur $\mathrm{L}^1_{\mathbb{K}}(\mu)$. Sa norme dans l'espace des formes linéaires continues sur $\mathrm{L}^1_{\mathbb{K}}(\mu)$ (le dual topologique de $\mathrm{L}^1_{\mathbb{K}}(\mu)$) est inférieure ou égale à 1, ce qui se traduit par :*

$$\forall \varphi \in \mathrm{L}^1_{\mathbb{K}}(\mu), \quad \left| \int_\Omega \varphi \, \mathrm{d}\mu \right| \leq \|\varphi\|_1. \tag{4.25}$$

Preuve. Vérifions d'abord que (4.24) définit bien une norme sur $\mathrm{L}^1_{\mathbb{K}}(\mu)$. Soient φ, ψ, quelconques dans $\mathrm{L}^1_{\mathbb{K}}(\mu)$ et $c \in \mathbb{K}$. Notons f et g des représentants respectifs de φ et ψ dans $\mathcal{L}^1_{\mathbb{K}}(\mu)$, alors $f + g$ est un représentant de $\varphi + \psi$ et cf un représentant de $c\varphi$. D'où l'homogénéité :

$$\|c\varphi\|_1 = \int_\Omega |cf| \, \mathrm{d}\mu = |c| \int_\Omega |f| \, \mathrm{d}\mu = |c| \cdot \|\varphi\|_1$$

et la sous-additivité

$$\|\varphi + \psi\|_1 = \int_\Omega |f + g| \, \mathrm{d}\mu \leq \int_\Omega (|f| + |g|) \, \mathrm{d}\mu = \int_\Omega |f| \, \mathrm{d}\mu + \int_\Omega |g| \, \mathrm{d}\mu = \|\varphi\|_1 + \|\psi\|_1.$$

D'autre part le réel $\|\varphi\|_1$ est clairement positif ou nul. Il reste à vérifier qu'il ne s'annule que si φ est égal au zéro de l'espace $\mathrm{L}^1_{\mathbb{K}}(\mu)$, classe d'équivalence de la fonction nulle, autrement dit $\mathcal{N}_{\mathbb{K}}(\mu)$. C'est bien le cas d'après la proposition 4.28.

L'inégalité (4.25) découle de la proposition 4.11 a) via le choix d'un représentant f. Rappelons que si ℓ est une application linéaire d'un espace vectoriel normé E dans un e.v.n. F, elle est continue si et seulement s'il existe une constante C telle que

$$\forall v \in E, \quad \|\ell(v)\|_F \leq C \|v\|_E.$$

La « norme opérateur » de ℓ dans l'espace vectoriel des applications linéaires continues de E dans F est alors définie comme la meilleure valeur possible C_ℓ pour la constante C dans l'inégalité ci-dessus, autrement dit

$$\|\ell\| := C_\ell = \sup_{\substack{v \in E \\ v \neq 0}} \frac{\|\ell(v)\|_F}{\|v\|_E} = \sup_{\|v\|_E = 1} \|\ell(v)\|_F = \sup_{\|v\|_E \leq 1} \|\ell(v)\|_F.$$

Revenons à (4.25), en prenant $E = \mathrm{L}^1_{\mathbb{K}}(\mu)$ muni de la norme $\| \ \|_1$ et $F = \mathbb{K}$ muni de la norme $| \ |$, on voit que (4.25) exprime bien la continuité de la forme linéaire $\varphi \mapsto \int_\Omega \varphi \,\mathrm{d}\mu$ et que la norme opérateur de cette forme linéaire est au plus 1. En fait elle vaut exactement 1 dès qu'il existe $A \in \mathcal{F}$ tel que $0 < \mu(A) < +\infty$. En effet $\mathbf{1}_A$ est alors μ-intégrable et sa classe φ dans $\mathrm{L}^1_{\mathbb{K}}(\mu)$ vérifie

$$\left|\int_\Omega \varphi \,\mathrm{d}\mu\right| = \left|\int_\Omega \mathbf{1}_A \,\mathrm{d}\mu\right| = \mu(A) = \int_\Omega |\mathbf{1}_A| \,\mathrm{d}\mu = \int_\Omega |\varphi| \,\mathrm{d}\mu = \|\varphi\|_1,$$

ce qui montre qu'on ne peut pas choisir de constante C plus petite que 1 dans l'inégalité $|\int_\Omega \psi \,\mathrm{d}\mu| \leq C\|\psi\|_1$ $(\forall \psi \in \mathrm{L}^1_{\mathbb{K}}(\mu))$. □

Nous verrons ultérieurement, comme conséquence du théorème de convergence dominée que l'espace vectoriel normé $\mathrm{L}^1_{\mathbb{K}}(\mu)$ est *complet*. C'est donc un espace de Banach.

4.4 Le théorème de convergence dominée

Nous abordons maintenant le théorème principal d'interversion limite intégrale connu sous le nom de théorème de convergence dominée ou théorème de Lebesgue. L'appellation « convergence dominée » est en soi une bonne description de ce théorème qui dit *grosso modo* que l'interversion $\lim \int f_n \,\mathrm{d}\mu = \int \lim f_n \,\mathrm{d}\mu$ est valide dès que la suite (f_n) vérifie une hypothèse de convergence (μ-p.p. sur Ω) et une hypothèse de *domination* par une fonction μ-intégrable au sens suivant.

Définition 4.35. *La suite (f_n) d'applications mesurables $\Omega \to \mathbb{K}$ ($\mathbb{K} = \mathbb{R}$, $\overline{\mathbb{R}}$ ou $\mathbb{C}$) est* dominée *sur Ω par la fonction mesurable $g : \Omega \to \overline{\mathbb{R}}_+$ si*

$$\forall n \in \mathbb{N},\ \forall \omega \in \Omega, \quad |f_n(\omega)| \leq g(\omega).$$

Elle est dite dominée μ-p.p. si l'inégalité ci-dessus a lieu μ-presque partout sur Ω.

Il est clair que dans la définition de la domination μ-p.p., on peut intervertir « $\forall n$ » et « μ-p.p. » puisqu'une réunion dénombrable d'ensembles négligeables est encore négligeable. D'un point de vue théorique, la meilleure fonction dominant une suite (f_n) est simplement la fonction $\sup_{n\in\mathbb{N}} |f_n|$. En pratique, il est rare que l'on sache expliciter cette fonction et étudier directement son intégrabilité, c'est ce qui motive l'introduction d'une fonction g donnant plus de souplesse.

Dans sa version finale, le théorème de Lebesgue est en fait un théorème de convergence dans L^1, d'où l'intérêt de la clause « μ-p.p. » dans ses hypothèses. L'approche en pente douce (au détriment de la concision !) présentée dans les trois lemmes suivants diffère l'utilisation de cette clause. Le lecteur pressé peut sauter directement à l'énoncé du théorème 4.39 et se contenter de la démonstration plus rapide, basée sur le lemme de Fatou, qui sera exposée après la proposition 4.40.

Lemme 4.36 (de convergence décroissante)**.** *Soit (f_n) une suite de fonctions mesurables $\Omega \to [0, +\infty[$ vérifiant les trois hypothèses* *a) La suite* (f_n) *converge simplement sur* Ω *vers* 0 *:* $\forall\omega\in\Omega$, $\lim_{n\to+\infty} f_n(\omega)=0$.

b) La suite $(f_n)_{n\geq n_0}$ *est décroissante :* $\forall\omega\in\Omega$, $\forall n\geq n_0$, $f_n(\omega)\geq f_{n+1}(\omega)$.

c) f_{n_0} *est* μ*-intégrable :* $\int_\Omega f_{n_0}\,\mathrm{d}\mu<+\infty$.

Alors

$$\lim_{n\to+\infty}\int_\Omega f_n\,\mathrm{d}\mu=0.$$

Preuve. Posons pour $n\geq n_0$, $h_n:=f_{n_0}-f_n$. Cette fonction est bien définie sur tout Ω (parce que les f_n sont à valeurs dans $[0,+\infty[$ et non dans $\overline{\mathbb{R}}_+$). Elle est mesurable comme différence de deux fonctions mesurables à valeurs dans $\mathbb{R}$. Elle est positive d'après b). La suite $(h_n)_{n\geq n_0}$ converge en croissant vers f_{n_0} d'après a) et b).

Par b) et c), pour $n\geq n_0$, les f_n sont intégrables puisque qu'elles sont positives et

$$\forall n\geq n_0,\qquad \int_\Omega f_n\,\mathrm{d}\mu\leq\int_\Omega f_{n_0}\,\mathrm{d}\mu<+\infty.$$

Les h_n sont alors elles aussi intégrables comme différences d'éléments de $\mathcal{L}^1_{\mathbb{R}}(\mu)$. Comme $f_n=f_{n_0}-h_n$, on a par linéarité de l'intégrale dans $\mathcal{L}^1_{\mathbb{R}}(\mu)$,

$$\forall n\geq n_0,\qquad \int_\Omega f_n\,\mathrm{d}\mu=\int_\Omega f_{n_0}\,\mathrm{d}\mu-\int_\Omega h_n\,\mathrm{d}\mu. \tag{4.26}$$

Quand n tend vers l'infini, le membre de droite de (4.26) converge vers 0 par le théorème de Beppo Levi appliqué à la suite $(h_n)_{n\geq n_0}$ et parce que $\int_\Omega f_{n_0}\,\mathrm{d}\mu$ a une valeur finie. □

Remarquons que l'hypothèse c) est indispensable. Sur l'espace mesuré $(\mathbb{R},\mathrm{Bor}(\mathbb{R}),\lambda)$ la suite des fonctions $f_n=\mathbf{1}_{[n,+\infty[}$ vérifie les hypothèses a) et b), mais pas la conclusion. Par contre on peut se passer de l'hypothèse de décroissance b), comme le montre le lemme suivant.

Lemme 4.37. *Soit* (f_n) *une suite de fonctions mesurables* $\Omega\to[0,+\infty[$, *convergeant simplement vers zéro sur* Ω *et dominée sur* Ω *par une fonction* μ*-intégrable* $g:\Omega\to\overline{\mathbb{R}}_+$. *Alors* $\int_\Omega f_n\,\mathrm{d}\mu$ *converge vers* 0.

Preuve. Pour tout $\omega\in\Omega$, on a $0\leq\sup_{n\in\mathbb{N}}f_n\leq g(\omega)\leq+\infty$. Comme g est μ-intégrable, l'ensemble $A:=\{g<+\infty\}$ est de complémentaire négligeable (proposition 4.19). Posons alors

$$g_n:=\sup_{k\geq n}(f_k\mathbf{1}_A).$$

Comme

$$0\leq\int_\Omega f_n\,\mathrm{d}\mu=\int_A f_n\,\mathrm{d}\mu\leq\int_A g_n\,\mathrm{d}\mu=\int_\Omega g_n\,\mathrm{d}\mu,$$

il suffit de prouver la convergence vers 0 de $\int_\Omega g_n\,\mathrm{d}\mu$. Pour ce faire, on vérifie que la suite des fonctions mesurables g_n à valeurs dans $[0,+\infty[$ satisfait aux conditions du lemme 4.36. La décroissance de g_n est claire. L'intégrabilité de g_0 résulte immédiatement de la domination de la suite (f_k) et donc *a fortiori* de la suite $(f_k\mathbf{1}_A)$ par la fonction μ-intégrable g puisque

$$\int_\Omega g_0\,\mathrm{d}\mu=\int_\Omega\sup_{k\geq0}(f_k\mathbf{1}_A)\,\mathrm{d}\mu\leq\int_\Omega g\,\mathrm{d}\mu<+\infty.$$

D'autre part, la convergence simple sur Ω vers zéro de $(f_k \mathbf{1}_A)$ découle immédiatement de celle de (f_k) et s'écrit :

$$\forall \omega \in \Omega, \quad \forall \varepsilon > 0,\ \exists k_0(\omega),\ \forall k \geq k_0(\omega), \quad f_k(\omega)\mathbf{1}_A(\omega) < \varepsilon.$$

On en déduit que

$$\forall \omega \in \Omega, \quad \forall \varepsilon > 0,\ \exists k_0(\omega),\ g_{k_0(\omega)}(\omega) = \sup_{k \geq k_0(\omega)} f_k(\omega)\mathbf{1}_A(\omega) \leq \varepsilon,$$

puis en raison de la décroissance de (g_n), $g_n(\omega) \leq \varepsilon$ pour tout $n \geq k_0(\omega)$, ce qui prouve la convergence simple sur Ω vers zéro de (g_n). Ainsi (g_n) vérifiant toutes les hypothèses du lemme 4.36, $\int_\Omega g_n \,\mathrm{d}\mu$ converge vers zéro et il en va de même pour $\int_\Omega f_n \,\mathrm{d}\mu$. □

Lemme 4.38. *Soit (f_n) une suite de fonctions mesurables $\Omega \to \mathbb{K}$ ($\mathbb{K} = \mathbb{R}$ ou $\mathbb{C}$). On suppose que la suite (f_n) converge simplement sur Ω vers une fonction f à valeurs dans $\mathbb{K}$ et est dominée sur Ω par une fonction μ-intégrable g. Alors f et les f_n sont μ-intégrables et*

$$\lim_{n \to +\infty} \int_\Omega |f_n - f| \,\mathrm{d}\mu = 0.$$

Preuve. La domination sur Ω de la suite (f_n) par g s'écrit

$$\forall \omega \in \Omega, \quad |f_n(\omega)| \leq g(\omega),$$

ce qui nous donne déjà la μ-intégrabilité des f_n (cf. proposition 4.11 b). En passant à la limite dans cette inégalité, on en déduit $|f(\omega)| \leq g(\omega)$, d'où la μ-intégrabilité de f.

La suite des fonctions mesurables $h_n := |f_n - f|$ est alors dominée par la fonction μ-intégrable $2g$. Ainsi la suite (h_n) satisfait alors aux conditions du lemme 4.37 donc $\int_\Omega h_n \,\mathrm{d}\mu$ converge vers zéro. □

Théorème 4.39 (de convergence dominée)**.** *Avec $\mathbb{K} = \mathbb{R}$ ou $\mathbb{C}$, soit (f_n) une suite de fonctions mesurables $\Omega \to \mathbb{K}$ vérifiant les deux hypothèses suivantes.*

*i) La suite (f_n) converge μ-presque partout sur Ω vers une fonction $f : \Omega \to \mathbb{K}$, mesurable $\mathcal{F}$-*Bor$(\mathbb{K})$*.*

*ii) La suite (f_n) est dominée μ-p.p. par une fonction μ-*intégrable $g : \Omega \to \overline{\mathbb{R}}_+$.

Sous ces conditions,

a) f et les f_n sont μ-intégrables ;

b) (f_n) converge vers f au sens L^1 *:* $\displaystyle\lim_{n \to +\infty} \int_\Omega |f_n - f| \,\mathrm{d}\mu = 0$ *;*

c) on a l'interversion limite-intégrale : $\displaystyle\lim_{n \to +\infty} \int_\Omega f_n \,\mathrm{d}\mu = \int_\Omega f \,\mathrm{d}\mu.$

Preuve. Soit A l'ensemble des $\omega \in \Omega$ où $f_n(\omega)$ converge vers $f(\omega)$. D'après l'hypothèse i), A est membre de la tribu $\mathcal{F}$ et $\mu(A^c) = 0$. Pour tout entier n, notons $B_n := \{\omega \in \Omega;\ |f_n(\omega)| \leq g(\omega)\}$. Par l'hypothèse ii), chaque B_n est membre de $\mathcal{F}$ et $\mu(B_n^c) = 0$. Notons B

l'intersection de tous les B_n. Comme une réunion dénombrable d'ensembles négligeables est encore négligeable, $\mu(B^c) = 0$. Enfin, notant $E = A \cap B$, on a $\mu(E^c) = 0$ et

$$\forall \omega \in E, \qquad \lim_{n \to +\infty} f_n(\omega) = f(\omega), \tag{4.27}$$

$$\forall \omega \in E,\ \forall n \in \mathbb{N}, \qquad |f_n(\omega)| \leq g(\omega). \tag{4.28}$$

Il est clair que si l'on remplace f_n par $h_n := f_n \mathbf{1}_E$ et f par $h := f\mathbf{1}_E$, les conditions (4.27) et (4.28) deviennent valides *sur tout* Ω. La suite (h_n) vérifie donc les hypothèses du lemme 4.38 et par conséquent h et les h_n sont μ-intégrables et $\int_\Omega |h_n - h| \,\mathrm{d}\mu$ converge vers zéro.

D'autre part, les f_n et f étant mesurables et à valeurs dans $\mathbb{R}$ ou $\mathbb{C}$, la fonction $f_n - f$ est bien *définie* en tout point de Ω et mesurable. Comme $h_n = f_n$ μ-p.p. et $h = f$ μ-p.p., f et les f_n sont μ-intégrables (cf. corollaire 4.29 ii). De plus $|h_n - h| = |f_n - f|$ μ-presque partout et $\int_\Omega |f_n - f| \,\mathrm{d}\mu = \int_\Omega |h_n - h| \,\mathrm{d}\mu$ (cf. corollaire 4.29 i). Par conséquent, $\int_\Omega |f_n - f| \,\mathrm{d}\mu$ converge vers zéro. L'interversion limite intégrale du c) est alors une conséquence immédiate de l'inégalité

$$0 \leq \left| \int_\Omega f_n \,\mathrm{d}\mu - \int_\Omega f \,\mathrm{d}\mu \right| = \left| \int_\Omega (f_n - f) \,\mathrm{d}\mu \right| \leq \int_\Omega |f_n - f| \,\mathrm{d}\mu.$$

□

En pratique, l'hypothèse de convergence i) est un peu plus simple à vérifier qu'il n'y paraît. En effet on a simplement besoin de montrer que f_n converge μ-presque partout, sans se préoccuper des valeurs à attribuer à f sur l'ensemble de divergence, ni de la mesurabilité de f. Cette simplification est une conséquence de la proposition suivante.

Proposition 4.40. *Soit (f_n) une suite de fonctions mesurables $\Omega \to \mathbb{K}$ qui converge μ-p.p., ce qui signifie que $A := \{\omega \in \Omega;\ f_n(\omega) \text{ converge dans } \mathbb{K}\}$ est un élément de la tribu $\mathcal{F}$ et $\mu(A^c) = 0$. Alors*

a) Il existe une fonction f définie sur tout *Ω et à valeurs dans $\mathbb{K}$,* mesurable $\mathcal{F}$-$\mathrm{Bor}(\mathbb{K})$*, telle que (f_n) converge μ-p.p. vers f. On peut prendre*

$$f := \lim_{n \to +\infty} \big(f_n \mathbf{1}_A\big). \tag{4.29}$$

b) (f_n) converge μ-p.p. vers h mesurable si et seulement si $h = f$ μ-p.p.

Preuve. Vérifions que la suite de fonctions $(f_n \mathbf{1}_A)$ converge en tout point de Ω. En effet si $\omega \in A$, $f_n(\omega)$ converge vers un élément de $\mathbb{K}$ par définition de A. Si $\omega \in A^c$, $(f_n \mathbf{1}_A)(\omega) = f_n(\omega)\mathbf{1}_A(\omega) = 0$ pour tout n donc $(f_n \mathbf{1}_A)(\omega)$ converge vers 0 quand n tend vers l'infini. La formule (4.29) définit donc bien une fonction f sur *tout* Ω. Pour tout n, $f_n \mathbf{1}_A$ est mesurable comme produit de deux fonctions mesurables et donc f l'est aussi comme limite d'une suite de fonctions mesurables convergeant simplement sur Ω.

D'autre part, (f_n) converge μ-p.p. vers f puisque si $\omega \in A$, $f_n(\omega) \to f(\omega)$ et si $\omega \notin A$, $(f_n(\omega))$ ne converge pas dans $\mathbb{K}$, donc ne peut converger vers $0 = f(\omega)$. Ceci montre que

A^c est exactement l'ensemble de tous les points ω tels que $f_n(\omega)$ ne converge pas vers $f(\omega)$. Comme $\mu(A^c) = 0$, on a bien vérifié que (f_n) converge μ-p.p. vers f.

Soit maintenant h mesurable, égale à f μ-presque partout. Notons B l'ensemble des $\omega \in \Omega$ tels que $(f_n(\omega))$ ne converge pas vers $h(\omega)$. Alors B est dans $\mathcal{F}$ (parce que $B = \{\limsup |f_n - h| > 0\}$). Pour $\omega \in A \cap \{h = f\}$, $(f_n(\omega))$ converge vers $h(\omega)$. On a donc l'inclusion

$$B \subset \Big(A \cap \{h = f\}\Big)^c = A^c \cup \{h \neq f\}$$

qui montre que $\mu(B) = 0$. Ainsi f_n converge vers h μ-presque partout.

Réciproquement, supposons que (f_n) converge vers h μ-presque partout, peut-on en déduire l'égalité μ-p.p. de f et h ? Notons

$$A_1 := \{\omega \in \Omega; f_n(\omega) \to f(\omega)\}, \quad A_2 := \{\omega \in \Omega; f_n(\omega) \to h(\omega)\}.$$

Alors $A_1 \cap A_2 \subset \{f = h\}$ par unicité de la limite de $(f_n(\omega))$ dans $\mathbb{K}$. D'où $\{f \neq h\} \subset A_1^c \cup A_2^c$ d'où $\mu(\{f \neq h\}) = 0$. $\square$

Voici la deuxième preuve du théorème de convergence dominée, annoncée page 122.

Preuve du théorème de Lebesgue par le lemme de Fatou. Les notations et les hypothèses étant celles du théorème 4.39, posons $\Omega' := \{\omega \in \Omega;\ \forall n \geq 1,\ |f_n(\omega)| \leq g(\omega)\}$ et

$$g_n := \Big(2g - |f - f_n|\Big)\mathbf{1}_{\Omega'}.$$

Nous allons prouver le b) de la conclusion du théorème 4.39, le a) et le c) se vérifiant comme ci-dessus. Notons que l'hypothèse de domination μ-p.p. se traduit par $\mu(\Omega \setminus \Omega') = 0$ et que vu la définition de Ω', elle entraîne aussi la positivité des g_n *sur tout* Ω. Les g_n étant mesurables positives, on peut appliquer le lemme de Fatou à la suite (g_n) :

$$\int_\Omega \liminf_{n\to+\infty} g_n \,\mathrm{d}\mu \leq \liminf_{n\to+\infty} \int_\Omega g_n \,\mathrm{d}\mu. \tag{4.30}$$

L'hypothèse de convergence μ-p.p. de f_n vers f nous donne

$$\liminf_{n\to+\infty} g_n = 2g, \quad \mu\text{-p.p.} \tag{4.31}$$

Notons que les deux membres de cette égalité sont des fonctions définies sur tout Ω et mesurables positives. De (4.31) on déduit

$$\int_\Omega \liminf_{n\to+\infty} g_n \,\mathrm{d}\mu = \int_\Omega 2g \,\mathrm{d}\mu. \tag{4.32}$$

Regardons maintenant le deuxième membre de (4.30). Par l'hypothèse de domination, $|f - f_n|$ est μ-intégrable, on peut donc écrire $\int_\Omega g_n \,\mathrm{d}\mu$ comme différence de deux intégrales [4] d'où :

$$\liminf_{n\to+\infty} \int_\Omega g_n \,\mathrm{d}\mu = \liminf_{n\to+\infty} \left\{\int_\Omega 2g \,\mathrm{d}\mu - \int_\Omega |f - f_n| \,\mathrm{d}\mu\right\} \tag{4.33}$$

$$= \int_\Omega 2g \,\mathrm{d}\mu - \limsup_{n\to+\infty} \int_\Omega |f - f_n| \,\mathrm{d}\mu. \tag{4.34}$$

4. En toute rigueur, il faudrait d'abord écrire $\int_\Omega g_n \,\mathrm{d}\mu = \int_{\Omega'} g_n \,\mathrm{d}\mu = \dots$, détails laissés au lecteur consciencieux.

Pour passer de (4.33) à (4.34), on utilise le fait que si (u_n) est une suite de réels et c une constante (finie), $\liminf_{n\to+\infty}(c-u_n) = c - \limsup_{n\to+\infty} u_n$, dont la vérification est laissée en exercice [5]. En combinant (4.30), (4.32) et (4.34), on obtient

$$\int_\Omega 2g \,\mathrm{d}\mu \leq \int_\Omega 2g \,\mathrm{d}\mu - \limsup_{n\to+\infty} \int_\Omega |f - f_n| \,\mathrm{d}\mu.$$

Comme $\int_\Omega 2g \,\mathrm{d}\mu$ est *finie*, cette inégalité équivaut à

$$\limsup_{n\to+\infty} \int_\Omega |f - f_n| \,\mathrm{d}\mu \leq 0.$$

On en déduit immédiatement la nullité de $\limsup \int_\Omega |f - f_n| \,\mathrm{d}\mu$ puis l'existence et la nullité de $\lim \int_\Omega |f - f_n| \,\mathrm{d}\mu$. Ceci établit le b) de la conclusion du théorème 4.39. □

Des trois conclusions du théorème de convergence dominée, l'interversion limite intégrale c) est incontestablement la plus populaire, car c'est elle que l'on cherche à obtenir dans la plupart des applications. Néanmoins, il importe de comprendre que la convergence L^1 de f_n vers f est un résultat *nettement plus fort* que la convergence de $\int_\Omega f_n \,\mathrm{d}\mu$ vers $\int_\Omega f \,\mathrm{d}\mu$. Le résultat suivant permet de s'en convaincre.

Proposition 4.41. *Soit* $(\Omega, \mathcal{F}, \mu)$ *un espace mesuré,* (f_n) *une suite dans* $\mathrm{L}^1_{\mathbb{K}}(\mu)$ *et* $f \in \mathrm{L}^1_{\mathbb{K}}(\mu)$. *Alors* (f_n) *converge au sens* L^1 *vers* f *si et seulement si* $\int_A f_n \,\mathrm{d}\mu$ *converge vers* $\int_A f \,\mathrm{d}\mu$, uniformément *en* A *sur* $\mathcal{F}$, *c'est-à-dire*

$$\sup_{A\in\mathcal{F}} \left| \int_A f_n \,\mathrm{d}\mu - \int_A f \,\mathrm{d}\mu \right| \xrightarrow[n\to+\infty]{} 0.$$

La proposition 4.41 est une conséquence immédiate du lemme suivant

Lemme 4.42. *La formule*

$$N(\varphi) := \sup_{A\in\mathcal{F}} \left| \int_A \varphi \,\mathrm{d}\mu \right|, \quad \varphi \in \mathrm{L}^1_{\mathbb{K}}(\mu)$$

définit une norme sur $\mathrm{L}^1_{\mathbb{K}}(\mu)$, *équivalente à la norme* $\| \; \|_1$.

Preuve. Vérifions d'abord que N est une norme. L'homogénéité est évidente. La sous-additivité de N résulte de l'additivité de l'intégrale, de la sous-additivité de la valeur absolue (ou du module) et de la sous-additivité du sup. Soit maintenant $\varphi \in \mathrm{L}^1_{\mathbb{K}}(\mu)$ telle que $N(\varphi) = 0$ et f un représentant de la classe de φ (f est donc une vraie fonction, $f \in \mathcal{L}^1_{\mathbb{K}}(\mu)$). Il s'agit de montrer que f est négligeable. Regardons d'abord le cas $\mathbb{K} = \mathbb{R}$. En prenant successivement $A := \{f > 0\}$ et $A' := \{f < 0\}$, on voit que $\int_A f \,\mathrm{d}\mu = 0$ et $\int_{A'} (-f) \,\mathrm{d}\mu = 0$. Or $\int_A f \,\mathrm{d}\mu = \int_\Omega f^+ \,\mathrm{d}\mu$ et $\int_{A'} (-f) \,\mathrm{d}\mu = \int_\Omega f^- \,\mathrm{d}\mu$. La proposition 4.28 appliquée aux fonctions mesurables positives f^+ et f^- montre alors qu'elles sont négligeables, donc que leur différence f l'est aussi. Le cas $\mathbb{K} = \mathbb{C}$ se ramène au cas réel en

5. Noter au passage qu'ici la positivité de $\liminf(c - u_n)$ et celle de $\limsup u_n$ impliquent la *finitude* de $\limsup u_n$.

observant que les inégalités $|\operatorname{Re} z| \leq |z|$ et $|\operatorname{Im} z| \leq |z|$ impliquent $N(\operatorname{Re} f) \leq N(f)$ et $N(\operatorname{Im} f) \leq N(f)$. Ainsi N est bien une norme.

Pour montrer l'équivalence avec la norme $\| \|_1$, on remarque d'abord que

$$\forall \varphi \in \mathrm{L}^1_{\mathbb{K}}(\mu),\ \forall A \in \mathcal{F}, \qquad \left| \int_A \varphi \,\mathrm{d}\mu \right| \leq \int_A |\varphi| \,\mathrm{d}\mu \leq \int_\Omega |\varphi| \,\mathrm{d}\mu = \|\varphi\|_1,$$

d'où en passant au sup sur $A \in \mathcal{F}$,

$$\forall \varphi \in \mathrm{L}^1_{\mathbb{K}}(\mu), \quad N(\varphi) \leq \|\varphi\|_1.$$

Dans l'autre sens, pour $\mathbb{K} = \mathbb{R}$, en prenant f représentant de la classe de φ,

$$\|f\|_1 \leq \|f^+\|_1 + \|f^-\|_1 = \int_\Omega f^+ \,\mathrm{d}\mu + \int_\Omega f^- \,\mathrm{d}\mu = \int_{\{f>0\}} f \,\mathrm{d}\mu + \int_{\{f<0\}} (-f) \,\mathrm{d}\mu \leq 2N(f),$$

d'où en revenant aux classes :

$$\forall \varphi \in \mathrm{L}^1_{\mathbb{R}}(\mu), \quad \|\varphi\|_1 \leq 2N(\varphi).$$

Dans le cas complexe, on a en utilisant le cas réel

$$\|f\|_1 \leq \|\operatorname{Re} f\|_1 + \|\operatorname{Im} f\|_1 \leq 2N(\operatorname{Re} f) + 2N(\operatorname{Im} f) \leq 4N(f).$$

En revenant aux classes on obtient :

$$\forall \varphi \in \mathrm{L}^1_{\mathbb{C}}(\mu), \quad \|\varphi\|_1 \leq 4N(\varphi),$$

ce qui achève la preuve de l'équivalence des normes. □

4.5 Comparaison avec l'intégrale de Riemann

Une des premières applications du théorème de convergence dominée est le recyclage de la plupart des intégrales de Riemann courantes en intégrales par rapport à la mesure de Lebesgue. Commençons par un bref rappel sur l'intégrabilité au sens de Riemann.

4.5.1 Intégrabilité au sens de Riemann

Soit $[a, b]$ un intervalle borné de $\mathbb{R}$. On appelle subdivision de $[a, b]$ toute suite finie du type $\Delta = \{x_0 = a < x_1 < \cdots < x_n = b\}$. Pour une fonction *bornée* $f : [a, b] \to \mathbb{R}$, $(-\infty < a < b < +\infty)$, on définit ses sommes de Darboux inférieure $S_\Delta(f)$ et supérieure $S^\Delta(f)$ par

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

Pour une illustration, voir les figures 4.1 et 4.2.

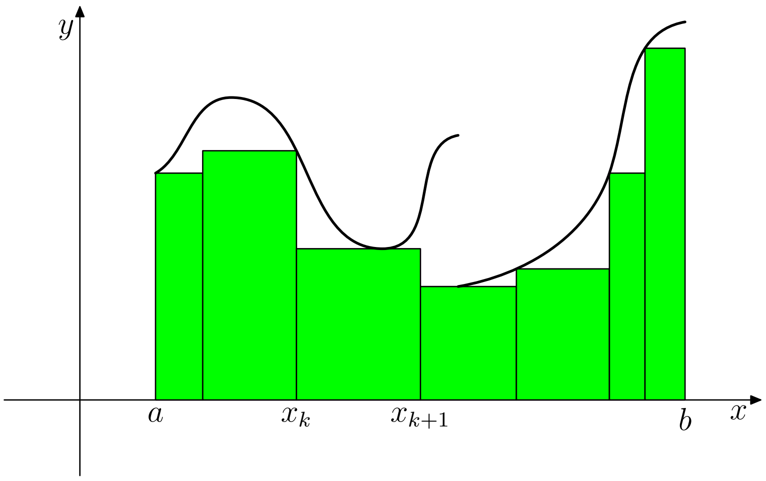


FIGURE 4.1 – $S_\Delta(f)$

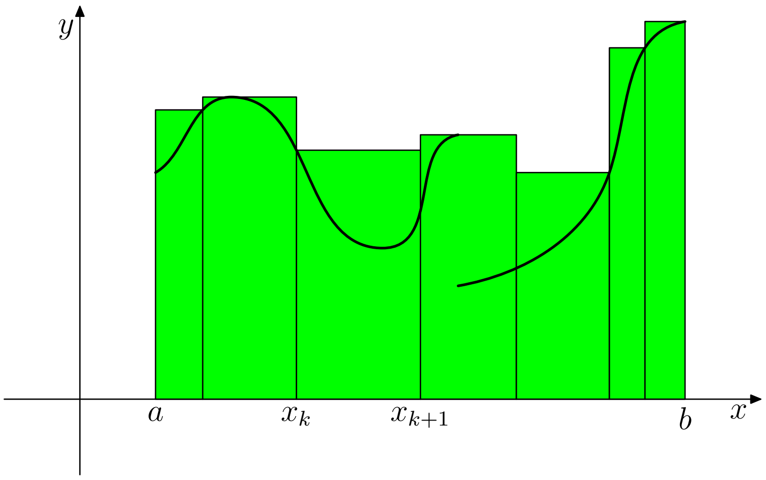


FIGURE 4.2 – $S^\Delta(f)$

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

$$\Delta \subset \Delta' \quad \Rightarrow \quad S_\Delta(f) \leq S_{\Delta'}(f) \text{ et } S^\Delta(f) \geq S^{\Delta'}(f).$$

Les figures 4.3 et 4.4 illustrent l'effet de l'adjonction à la subdivision Δ des figures 4.1 et 4.2 de deux nouveaux points.

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 $[a, b]$.

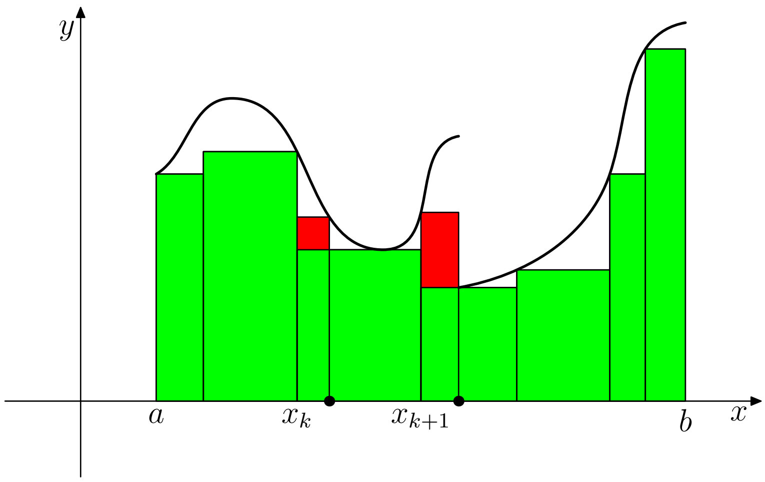


FIGURE 4.3 – Si $\Delta \subset \Delta'$, $S_\Delta(f) \leq S_{\Delta'}(f)$

Pour Δ_1 et Δ_2 subdivisions de $[a, b]$ on a clairement

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

d'où $S_{\Delta_1}(f) \leq S^{\Delta_2}(f)$. En prenant successivement le sup sur tous les Δ_1, puis l'inf sur tous les Δ_2, on en déduit

$$I_*(f) \leq I^*(f),$$

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

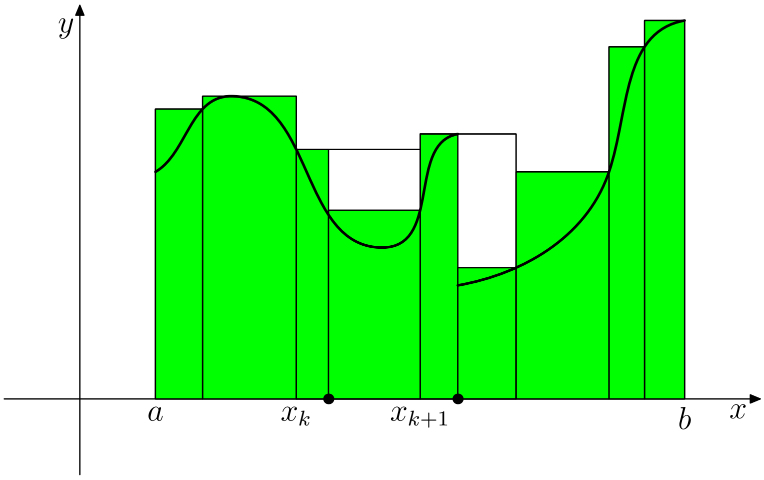


FIGURE 4.4 – Si $\Delta \subset \Delta'$, $S^\Delta(f) \geq S^{\Delta'}(f)$

Définition 4.43. *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*

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

Nous prendrons bien garde dans cette section de distinguer les notations $\int_a^b f(x)\,\mathrm{d}x$ intégrale de Riemann (éventuellement généralisée) et $\int_{[a,b]} f\,\mathrm{d}\lambda$, intégrale de Lebesgue (i.e. au sens de la définition 4.3) par rapport à la mesure de Lebesgue λ. Par la suite, une fois établie leur coïncidence sous des conditions assez générales, on pourra confondre les deux notations, comme il est d'usage courant dans la littérature mathématique.

Nous terminons ces « rappels » sur la Riemann intégrabilité par le cas important des fonctions continues.

Proposition 4.44. *Si $f : [a,b] \to \mathbb{R}$ est continue, elle est Riemann intégrable. De plus si F est une primitive de f sur $[a,b]$,*

$$\int_a^b f(x)\,\mathrm{d}x = F(b) - F(a). \tag{4.35}$$

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 $\varepsilon > 0$, on peut trouver une subdivision $\Delta = \{x_0 = a < x_1 < \cdots < x_n = b\}$ (dépendant de δ) telle que pour $k = 1, \ldots, n$, $x_k - x_{k-1} < \delta$. Comme les bornes inférieure m_k et supérieure M_k de f sur le compact $[x_{k-1}, x_k]$ sont atteintes, on a pour tout k $0 \leq M_k - m_k < \varepsilon$. On a alors

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

En raison de l'encadrement $S_\Delta(f) \leq I_*(f) \leq I^*(f) \leq S^\Delta(f)$, nous avons ainsi établi que

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

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, 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

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

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

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

et comme nous savons déjà que f est Riemann intégrable on en déduit $F(b) - F(a) = I_*(f) = I^*(f)$, ce qui établit (4.35). □

La preuve de (4.35) 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 $[a,b]$.

4.5.2 Comparaison

Proposition 4.45. *Soit $[a,b]$ un intervalle borné de $\mathbb{R}$. Si f borélienne bornée $[a,b] \to \mathbb{R}$ est Riemann intégrable, elle est aussi intégrable sur $[a,b]$ par rapport à la mesure de Lebesgue et*

$$\int_a^b f(x)\,\mathrm{d}x = \int_{[a,b]} f\,\mathrm{d}\lambda. \tag{4.36}$$

Preuve. La λ intégrabilité de f résulte de sa bornitude et du fait que la mesure de Lebesgue sur $[a,b]$ est *finie*, cf. Prop. 4.11 b). Ainsi les deux intégrales dans (4.36) sont bien définies. Pour montrer leur égalité, commençons par associer à chaque subdivision Δ les fonctions *en escaliers*[6]

$$f_\Delta(x) := \sum_{k=1}^{n} \mathbf{1}_{[x_{k-1}, x_k[}(x) \inf_{[x_{k-1}, x_k]} f, \qquad f^\Delta(x) := \sum_{k=1}^{n} \mathbf{1}_{[x_{k-1}, x_k[}(x) \sup_{[x_{k-1}, x_k]} f.$$

6. Une fonction en escaliers est une combinaison linéaire d'indicatrices d'intervalles disjoints. C'est donc aussi une fonction étagée, la réciproque étant bien entendu fausse.

On choisit alors une suite de subdivisions (Δ_n), croissante pour l'inclusion et telle que $S_{\Delta_n}(f) \uparrow I_*(f)$ et $S^{\Delta_n}(f) \downarrow I^*(f)$. Pour vérifier l'existence d'une telle suite (Δ_n), remarquons que la définition de $I_*(f)$ comme borne supérieure et celle de $I^*(f)$ comme borne inférieure nous fournissent deux suites (D_n) et (D'_n) pas forcément monotones de subdivisions telles que $\lim_{n\to+\infty} S_{D_n}(f) = I_*(f)$ et $\lim_{n\to+\infty} S^{D'_n}(f) = I^*(f)$. Il suffit alors de prendre $\Delta_n := \cup_{k\leq n}(D_k \cup D'_k)$. La croissance pour l'inclusion de la suite (Δ_n) implique la croissance de la suite (f_{Δ_n}) et la décroissance de (f^{Δ_n}). On a alors

$$\lim \uparrow f_{\Delta_n} =: g \leq f \leq h := \lim \downarrow f^{\Delta_n}.$$

Les fonctions g et h sont boréliennes comme limites de fonctions boréliennes. Remarquons maintenant que les fonctions en escaliers f_{Δ_n} et f^{Δ_n} étant étagées, on a

$$\int_{[a,b]} f_{\Delta_n} \,\mathrm{d}\lambda = \sum_{k=1}^{n} \lambda([x_{k-1}, x_k[) \inf_{[x_{k-1},x_k]} f = \sum_{k=1}^{n} (x_k - x_{k-1}) \inf_{[x_{k-1},x_k]} f = S_{\Delta_n}(f)$$

et de manière analogue

$$\int_{[a,b]} f^{\Delta_n} \,\mathrm{d}\lambda = S^{\Delta_n}(f).$$

Par construction de (Δ_n), $S_{\Delta_n}(f)$ et $S^{\Delta_n}(f)$ convergent respectivement vers $I_*(f)$ et $I^*(f)$, donc vers la même limite $\int_a^b f(x)\,\mathrm{d}x$ puisque f est supposée Riemann intégrable.

D'autre part, les suites (f_{Δ_n}) et (f^{Δ_n}) sont *dominées* par la fonction constante $M := \sup_{[a,b]} |f|$ qui est λ-intégrable puisque $\lambda([a,b]) = b - a < +\infty$. Par le théorème de convergence dominée, on a donc

$$\lim_{n\to+\infty} \int_{[a,b]} f_{\Delta_n} \,\mathrm{d}\lambda = \int_{[a,b]} g \,\mathrm{d}\lambda \quad \text{et} \quad \lim_{n\to+\infty} \int_{[a,b]} f^{\Delta_n} \,\mathrm{d}\lambda = \int_{[a,b]} h \,\mathrm{d}\lambda.$$

On voit ainsi que

$$\int_{[a,b]} g \,\mathrm{d}\lambda = I_*(f) = \int_a^b f(x)\,\mathrm{d}x = I^*(f) = \int_{[a,b]} h \,\mathrm{d}\lambda.$$

Enfin f étant λ-intégrable sur $[a,b]$, en intégrant les inégalités $g \leq f \leq h$, on obtient

$$\int_{[a,b]} g \,\mathrm{d}\lambda \leq \int_{[a,b]} f \,\mathrm{d}\lambda \leq \int_{[a,b]} h \,\mathrm{d}\lambda.$$

On en déduit que $\int_{[a,b]} f \,\mathrm{d}\lambda = \int_a^b f(x)\,\mathrm{d}x$. □

Remarque 4.46. On ne peut pas s'affranchir de l'hypothèse de Riemann intégrabilité de f dans la proposition 4.45. En effet la fonction indicatrice des rationnels de $[a,b]$ est bornée et borélienne donc λ-intégrable sur $[a,b]$. Par contre elle n'est pas Riemann intégrable car $I_*(f) = 0$ et $I^*(f) = b - a$ (exercice).

Proposition 4.47. *Soit f une fonction $]a,b[\to \mathbb{R}$ $(-\infty \leq a < b \leq +\infty)$. On suppose que* *i)* f *est borélienne ;*

ii) sur tout $[\alpha,\beta]\subset]a,b[$, f *est bornée et Riemann intégrable ;*

iii) l'intégrale de Riemann généralisée $\int_a^b f(x)\,\mathrm{d}x$ *est* absolument *convergente.*

Alors f *est Lebesgue intégrable sur* $]a,b[$ *(donc aussi sur* $[a,b]$*) et*

$$\int_{[a,b]} f\,\mathrm{d}\lambda = \int_{]a,b[} f\,\mathrm{d}\lambda = \int_a^b f(x)\,\mathrm{d}x.$$

Preuve. L'hypothèse iii) et les inégalités $0\le f^+\le |f|$ et $0\le f^-\le |f|$ impliquent

$$0\le \int_a^b f^+(x)\,\mathrm{d}x < +\infty,\quad 0\le \int_a^b f^-(x)\,\mathrm{d}x < +\infty. \tag{4.37}$$

Soient (a_n) une suite décroissante de réels de limite a et (b_n) une suite croissante de limite b, telles que pour tout $n\in\mathbb{N}$, $a<a_n<b_n<b$. Par i), ii) et la proposition 4.45, on a pour tout n, $\int_{[a_n,b_n]} f^+\,\mathrm{d}\lambda = \int_{a_n}^{b_n} f^+(x)\,\mathrm{d}x$ et $\int_{[a_n,b_n]} f^-\,\mathrm{d}\lambda = \int_{a_n}^{b_n} f^-(x)\,\mathrm{d}x$. De plus les suites de fonctions boréliennes positives $f^+\mathbf{1}_{[a_n,b_n]}$ et $f^-\mathbf{1}_{[a_n,b_n]}$ convergent sur tout $]a,b[$ en croissant vers f^+ et f^- respectivement. Ainsi en combinant le théorème de Beppo Levi et la convergence des intégrales de Riemann généralisées sur $]a,b[$ de f^+ et f^-, on obtient

$$\int_{]a,b[} f^+\,\mathrm{d}\lambda = \int_a^b f^+(x)\,\mathrm{d}x,\quad \int_{]a,b[} f^-\,\mathrm{d}\lambda = \int_a^b f^-(x)\,\mathrm{d}x. \tag{4.38}$$

De (4.37) et (4.38) on déduit

$$\int_{]a,b[} |f|\,\mathrm{d}\lambda = \int_a^b f^+(x)\,\mathrm{d}x + \int_a^b f^-(x)\,\mathrm{d}x < +\infty,$$

ce qui établit la λ-intégrabilité de f. La définition de l'intégrale de f par rapport à λ et (4.38) nous donnent enfin

$$\int_{]a,b[} f\,\mathrm{d}\lambda = \int_{]a,b[} f^+\,\mathrm{d}\lambda - \int_{]a,b[} f^-\,\mathrm{d}\lambda = \int_a^b f^+(x)\,\mathrm{d}x - \int_a^b f^-(x)\,\mathrm{d}x = \int_a^b f(x)\,\mathrm{d}x.$$

□

L'hypothèse de convergence *absolue* de l'intégrale généralisée $\int_a^b f(x)\,\mathrm{d}x$ est indispensable, comme le montre le résultat suivant.

Proposition 4.48. *Soit* f *borélienne sur* $]a,b[$ *et Riemann intégrable sur tout* $[\alpha,\beta]\subset\,]a,b[$*. On suppose de plus que l'intégrale de Riemann généralisée* $\int_a^b f(x)\,\mathrm{d}x$ n'est pas *absolument convergente. Alors*

$$\int_{]a,b[} |f|\,\mathrm{d}\lambda = +\infty$$

et donc f n'est pas *Lebesgue intégrable sur* $]a,b[$.

Preuve. La croissance de l'intégrale sur $\mathcal{M}_+$ et la proposition 4.45 nous donnent la minoration

$$\int_{]a,b[} |f| \,\mathrm{d}\lambda \geq \int_{[\alpha,\beta]} |f| \,\mathrm{d}\lambda = \int_\alpha^\beta |f(x)| \,\mathrm{d}x.$$

Ce minorant tend vers $+\infty$ quand α tend vers a et β vers b en raison de la non convergence absolue de l'intégrale généralisée $\int_a^b f(x)\,\mathrm{d}x$. □

Pour conclure cette sous-section, il n'est pas inutile d'énoncer et de mémoriser le cas particulier suivant de la proposition 4.47 qui devrait couvrir la plupart des intégrales rencontrées en pratique.

Proposition 4.49. *Soient $a, b \in \overline{\mathbb{R}}$, $a < b$. Si $f :]a, b[\to \mathbb{R}$ est* continue *sur $]a, b[$ et d'intégrale de Riemann généralisée $\int_a^b f(x)\,\mathrm{d}x$* absolument *convergente, alors f est Lebesgue-intégrable sur $]a, b[$ et*

$$\int_{]a,b[} f \,\mathrm{d}\lambda = \int_a^b f(x) \,\mathrm{d}x.$$

Preuve. Si $[\alpha, \beta] \subset]a, b[$, $a < \alpha \leq \beta < b$, donc α et β sont finis et l'intervalle $[\alpha, \beta]$ est un *compact* sur lequel f est continue donc bornée et Riemann intégrable. D'autre part f étant continue $]a, b[\to \mathbb{R}$ est clairement borélienne. Ainsi toutes les hypothèses de la proposition 4.47 sont vérifiées, d'où la conclusion. □

4.5.3 Changement de variable dans une intégrale $\int f \,\mathrm{d}\lambda$

La proposition 4.45 permet l'extension suivante aux intégrales par rapport à λ de la formule de changement de variable dans une intégrale de Riemann.

Proposition 4.50. *Soit I un intervalle ouvert (pas nécessairement borné) de $\mathbb{R}$, $\varphi : I \to \mathbb{R}$, strictement monotone et de classe C^1. On pose $J := \varphi(I)$. Alors pour toute application $f : J \to \mathbb{R}$ mesurable positive ou λ-intégrable sur J, on a*

$$\forall A \in \mathrm{Bor}(I), \quad \int_A f\big(\varphi(x)\big)|\varphi'(x)| \,\mathrm{d}\lambda(x) = \int_{\varphi(A)} f(y) \,\mathrm{d}\lambda(y). \tag{4.39}$$

Si de plus φ' ne s'annule en aucun point de I, on a aussi

$$\forall A \in \mathrm{Bor}(I), \quad \int_A f\big(\varphi(x)\big) \,\mathrm{d}\lambda(x) = \int_{\varphi(A)} f(y)|(\varphi^{-1})'(y)| \,\mathrm{d}\lambda(y). \tag{4.40}$$

Remarquons que les hypothèses de régularité portent sur le changement de variable φ et non sur f. On ne suppose pas f Riemann intégrable. La portée de cette formule de changement de variable dépasse donc largement celle d'une simple traduction du langage de l'intégrale de Riemann dans celui de l'intégrale de Lebesgue.

Preuve. La première intégrale dans (4.39) peut s'écrire $\int_A f\circ\varphi \,\mathrm{d}\mu$ où μ est la mesure de densité $|\varphi'|$ par rapport à λ. On prouve (4.39) en appliquant le théorème de transfert entre les espaces mesurés :

$$(I, \mathrm{Bor}(I), \mu) \xrightarrow{\quad\varphi\quad} (J, \mathrm{Bor}(J), \nu := \mu\circ\varphi^{-1}).$$

Commençons par identifier la mesure image ν. Comme φ est continue strictement monotone, J est un intervalle ouvert. Sa tribu borélienne est engendrée par la classe des intervalles bornés $[c,d]\subset J$. De plus φ est une bijection de I sur J et l'inverse ensembliste se confond avec la bijection inverse.
Cas 1 : φ décroissante. Alors $|\varphi'| = -\varphi'$ et

$$\nu([c,d]) = \mu\circ\varphi^{-1}([c,d]) = \mu([\varphi^{-1}(d), \varphi^{-1}(c)]) = \int_{[\varphi^{-1}(d),\varphi^{-1}(c)]} (-\varphi')\,\mathrm{d}\lambda.$$

La fonction φ' est continue donc bornée et Riemann intégrable sur l'intervalle fermé borné $[\varphi^{-1}(d), \varphi^{-1}(c)]$. Par la proposition 4.45, on a

$$\int_{[\varphi^{-1}(d),\varphi^{-1}(c)]} (-\varphi')\,\mathrm{d}\lambda = -\int_{\varphi^{-1}(d)}^{\varphi^{-1}(c)} \varphi'(y)\,\mathrm{d}y = -\Big[\varphi(y)\Big]_{\varphi^{-1}(d)}^{\varphi^{-1}(c)} = d - c = \lambda([c,d]).$$

Ainsi les deux mesures ν et λ σ-finies sur J (qui n'est pas forcément borné) coïncident sur la π-classe des intervalles $[c,d]$ qui engendre $\mathrm{Bor}(J)$. Ceci montre que ν est la mesure de Lebesgue sur J.
Cas 2 : φ croissante. La vérification de l'égalité $\nu = \lambda$ étant analogue (et légèrement plus simple !) sera laissée au lecteur.

À ce stade, le théorème de transfert nous donne (4.39) pour $A = I$. Pour conclure dans le cas général, il suffit d'appliquer le théorème de transfert avec $(f\circ\varphi)\mathbf{1}_A$ en remarquant que :

$$\forall x\in I,\quad f\big(\varphi(x)\big)\mathbf{1}_A(x) = f\big(\varphi(x)\big)\mathbf{1}_{\varphi(A)}\big(\varphi(x)\big),$$

la transformation d'indicatrice se justifiant par l'équivalence de $x\in A$ et $\varphi(x)\in\varphi(A)$. Ainsi (4.39) est établie.

Pour montrer la formule jumelle (4.40), on utilise la même méthode en appliquant le théorème de transfert avec $\mu = \lambda$ et $\nu = \lambda\circ\varphi^{-1}$. Nous nous contentons de décrire l'identification de la mesure ν dans le cas φ décroissante, laissant au lecteur le soin de compléter la preuve. L'hypothèse supplémentaire que φ' ne s'annule en aucun point de I nous garantit que φ^{-1} a une dérivée *continue* en tout point de J :

$$\forall y\in J,\quad (\varphi^{-1})'(y) = \frac{1}{\varphi'\big(\varphi^{-1}(y)\big)}.$$

On peut donc écrire la variation de φ^{-1} entre deux points de J comme l'intégrale de Riemann de $(\varphi^{-1})'$ entre ces deux points. Pour $[c,d]\subset J$, $-\infty < c \le d < +\infty$, on a ainsi

$$\begin{aligned}\nu([c,d]) = \lambda\Big([\varphi^{-1}(d), \varphi^{-1}(c)]\Big) = \varphi^{-1}(c) - \varphi^{-1}(d) &= \int_c^d -(\varphi^{-1})'(y)\,\mathrm{d}y \\ &= \int_c^d |(\varphi^{-1})'(y)|\,\mathrm{d}y \\ &= \int_{[c,d]} |(\varphi^{-1})'(y)|\,\mathrm{d}\lambda(y).\end{aligned}$$

Ceci montre que ν est la mesure de densité $|(\varphi^{-1})'|$ par rapport à λ. Notons que nous avons utilisé deux fois la continuité de $(\varphi^{-1})'$ sur $[c,d]$. D'abord pour écrire sa primitive φ^{-1} à l'aide d'une intégrale de Riemann, ensuite pour convertir cette intégrale de Riemann en intégrale de Lebesgue. □

4.5.4 Fonction de répartition et densité

Soit X une variable aléatoire réelle sur $(\Omega, \mathcal{F}, \mathbf{P})$, P_X sa loi et F sa fonction de répartition : $F(x) = \mathbf{P}(X \leq x) = P_X(]-\infty, x])$. Supposons de plus que P_X ait une densité f par rapport à la mesure de Lebesgue. On a alors :

$$\forall x \in \mathbb{R}, \quad F(x) = P_X(]-\infty, x]) = \int_{]-\infty,x]} f\,\mathrm{d}\lambda. \tag{4.41}$$

Il est naturel de se demander quelle relation y a-t-il entre F et f ? La réponse est que F est dérivable λ-presque partout et que sa dérivée (définie sauf sur un ensemble de mesure nulle) est égale λ-p.p. à f. La preuve de ce résultat sortirait du programme de la Licence. Le lecteur curieux pourra se reporter à l'ouvrage de Kolmogorov, Fomine *Éléments de la Théorie des Fonctions et de l'Analyse Fonctionnelle*, chap. VI, §3. Nous nous contenterons du résultat moins général suivant, qui devrait cependant couvrir la plupart des cas rencontrés en pratique.

Proposition 4.51. *Soit X une variable aléatoire réelle et F sa fonction de répartition. On suppose que F est continue sur $\mathbb{R}$ et de classe C^1 par morceaux. Alors la loi de X a une densité f par rapport à la mesure de Lebesgue et $f = F'$ λ-presque partout.*

Preuve. Dire que F est continue et C^1 par morceaux sur $\mathbb{R}$ signifie qu'il existe une suite finie $-\infty < a_1 < a_2 < \cdots < a_n < +\infty$ telle que F est dérivable partout sur $\mathbb{R}$, sauf peut-être aux points a_i, que sa dérivée F' est continue sur $\mathbb{R} \setminus \{a_1, \dots, a_n\}$ et que F est aussi continue en chaque point a_i (les « raccords » sont continus). En raison de la croissance de F, F' est positive partout où elle est définie. Notons $a_0 := -\infty$, $a_{n+1} := +\infty$ et $I_i :=]a_i, a_{i+1}[$, $0 \leq i \leq n$. Comme F a pour limites 0 en $-\infty$ et 1 en $+\infty$, il est commode de la prolonger par continuité à $\overline{\mathbb{R}}$ en posant $F(a_0) := 0$ et $F(a_{n+1}) := 1$.

Si $[c,d]$ est inclus dans I_i, F' est continue sur $[c,d]$ et $\int_c^d F'(t)\,\mathrm{d}t = F(d) - F(c)$. Par continuité de F aux extrémités de I_i, on en déduit que l'intégrale généralisée $\int_{a_i}^{a_{i+1}} F'(t)\,\mathrm{d}t$ converge absolument ($F' \geq 0$ sur I_i) et que pour tout $x \in]a_i, a_{i+1}]$, $F(x) = F(a_i) + \int_{a_i}^x F'(t)\,\mathrm{d}t$. Par la proposition 4.49, on a $F(x) = F(a_i) + \int_{]a_i,x[} F'(t)\,\mathrm{d}\lambda(t)$.

Posons $f := F'$ sur $\mathbb{R} \setminus \{a_1, \dots, a_n\}$ et $f(a_i) := 0$ pour $1 \leq i \leq n$. Alors f est mesurable positive (exercice) et pour tout $x \in [a_i, a_{i+1}]$, $F(x) = F(a_i) + \int_{]a_i,x[} f \,\mathrm{d}\lambda$. On en déduit en particulier que

$$\int_{\mathbb{R}} |f| \,\mathrm{d}\lambda = \int_{\mathbb{R}\setminus\{a_1,\dots,a_n\}} f \,\mathrm{d}\lambda = \sum_{i=0}^{n} \int_{]a_i,a_{i+1}[} f \,\mathrm{d}\lambda = \sum_{i=0}^{n} \Big(F(a_{i+1}) - F(a_i) \Big) = 1 < +\infty.$$

Donc la fonction mesurable positive f est λ intégrable sur $\mathbb{R}$. Notons μ la mesure de densité f par rapport à λ. En recollant les morceaux, on voit aussi que

$$\forall x \in \mathbb{R}, \quad F(x) = \int_{]-\infty,x]} f \,\mathrm{d}\lambda = \mu(]-\infty, x]), \tag{4.42}$$

c'est clair si x n'est pas l'un des a_i (noter que l'on peut modifier l'ensemble d'intégration en rajoutant un nombre fini de points, donc un ensemble de λ-mesure nulle, sans changer l'intégrale). Si x est l'un des a_i $(1 \leq i \leq n)$, il suffit de remarquer que les deux membres de (4.42) représentent chacun une fonction de répartition, continue à droite sur $\mathbb{R}$ et de faire tendre x vers a_i par valeurs supérieures. L'égalité (4.42) montre donc que P_X, loi de X et μ ont même fonction de répartition donc que $P_X = \mu$. Ainsi P_X a la densité f par rapport à λ. □

Exemple 4.2. Soient $X_1, \dots, X_n$, n variables aléatoires définies sur le même $(\Omega, \mathcal{F}, \mathbf{P})$, indépendantes et de même loi, de fonction de répartition F, continue sur $\mathbb{R}$ et C^1 par morceaux. On pose

$$M_n := \max_{1\leq i\leq n} X_i, \quad H(x) := \mathbf{P}(M_n \leq x).$$

Il est facile d'exprimer la fonction de répartition H de M_n à l'aide F. En effet, pour tout $x \in \mathbb{R}$,

$$\begin{aligned} H(x) = \mathbf{P}\Big(\max_{1\leq i\leq n} X_i \leq x\Big) &= \mathbf{P}\Big(\forall i = 1, \dots, n,\ X_i \leq x\Big) \\ &= \mathbf{P}\Big(\bigcap_{i=1}^{n} \{X_i \leq x\}\Big) \\ &= \prod_{i=1}^{n} \mathbf{P}(X_i \leq x) \qquad (4.43) \\ &= \Big(F(x)\Big)^n, \qquad (4.44) \end{aligned}$$

où (4.43) se justifie (en anticipant légèrement sur le chapitre 5) par le fait que l'indépendance de la suite (X_i) implique l'indépendance mutuelle des évènements $\{X_i \leq x\}$ et où (4.44) vient de ce que les X_i ont même loi de f.d.r. F, donc $\mathbf{P}(X_i \leq x) = F(x)$.

Par la proposition 4.51, la loi de X_i a une densité f égale λ-presque partout à F'. Les théorèmes classiques sur la continuité et la dérivabilité des fonctions composées montrent que $H = F^n$ est elle aussi continue sur $\mathbb{R}$ et C^1 par morceaux. Il en résulte que la loi de M_n a une densité h par rapport à λ et que

$$h = H' = nF^{n-1}F', \quad \lambda - \text{p.p.}$$

Par exemple si les X_i suivent la loi exponentielle de paramètre $a > 0$,

$$F(x) = \left(1 - \mathrm{e}^{-ax}\right)\mathbf{1}_{[0,+\infty[}(x),$$

F est donc continue sur $\mathbb{R}$ et dérivable partout sauf au point $x = 0$ et M_n a pour densité :

$$h(x) = na\left(1 - \mathrm{e}^{-ax}\right)^{n-1}\mathrm{e}^{-ax}\mathbf{1}_{[0,+\infty[}(x).$$

Remarque 4.52. Attention à ne pas tranformer abusivement la proposition 4.51 en une règle pratique du style « si la fonction de répartition F est dérivable presque partout, alors il y a une densité, égale à F' p.p. ». Cette pseudo règle est doublement fausse. D'abord l'hypothèse de continuité de F *en tout point* de $\mathbb{R}$ est indispensable. Si X suit une loi discrète avec $X(\Omega)$ fini, sa f.d.r. F est dérivable et de dérivée nulle en tout point d'un ensemble de complémentaire fini. Elle est donc bien C^1 par morceaux, mais comme F a des discontinuités (en nombre fini), P_X ne peut avoir de densité par rapport à λ, car s'il y avait une densité f, on aurait :

$$\forall x \in \mathbb{R}, \quad \mathbf{P}(X = x) = \int_{\{x\}} f(t)\,\mathrm{d}\lambda(t) = 0, \quad \text{car } \lambda(\{x\}) = 0.$$

D'autre part, il existe des fonctions de répartition F, continues sur tout $\mathbb{R}$, dérivables presque partout sur $\mathbb{R}$ et de dérivée nulle presque partout sur $\mathbb{R}$. Pour un exemple, voir le problème sur « l'escalier de Cantor » dans les Annales d'IFP 2001–02. Si la loi correspondante avait une densité égale[7] presque partout à F', on aurait $P_X(\mathbb{R}) = \int_{\mathbb{R}} F'\,\mathrm{d}\lambda = 0$, ce qui est impossible puisque $P_X(\mathbb{R}) = 1$.

4.6 Applications du théorème de convergence dominée

Le théorème de convergence dominée est un outil efficace pour étudier les propriétés de régularité de fonctions définies par une intégrale de la forme

$$F(x) := \int_{\Omega} f(x, \omega)\,\mathrm{d}\mu(\omega),$$

où f est définie sur un produit cartésien $E\times\Omega$ et à valeurs dans $\mathbb{K} = \mathbb{R}$ ou $\mathbb{C}$. Pour pouvoir ramener les questions de régularité de F à des problèmes de convergences de suites, on se limitera au cas où E est un espace *métrique*. En pratique E sera le plus souvent une partie de $\mathbb{R}$, $\mathbb{R}^d$ ou $\mathbb{C}$. Nous utilisons les notations classiques pour les applications partielles :

$$\begin{aligned} f(x, \,.\,) &: \Omega \to \mathbb{K}, & \omega &\mapsto f(x, \,.\,)(\omega) := f(x, \omega), \\ f(\,.\,, \omega) &: E \to \mathbb{K}, & x &\mapsto f(\,.\,, \omega)(x) := f(x, \omega). \end{aligned}$$

Les théorèmes que nous avons en vue établissent une régularité de F à partir d'un triplet d'hypothèses :

7. En fait cette loi n'a *aucune* densité par rapport à λ, il s'agit d'une loi dite *singulière*.

(Mi) Mesurabilité ou intégrabilité des applications partielles $f(x,\,.\,) : \Omega \to \mathbb{K}$.

(Ré) Régularité (pour μ-presque tout ω) des applications partielles $f(\,.\,,\omega) : E \to \mathbb{K}$.

(Do) Domination (localement en x) de f (ou de $\frac{\partial^k}{\partial x^k} f$) par une fonction μ-intégrable $g : \Omega \to \overline{\mathbb{R}}_+$.

Théorème 4.53 (de continuité sous le signe $\int$)**.** *Soit* (E,d) *un espace métrique,* $\mathbb{K} = \mathbb{R}$ *ou* $\mathbb{C}$ *et* f *une application*

$$f : E \times \Omega \to \mathbb{K}, \quad (x,\omega) \mapsto f(x,\omega).$$

Soit $x_0 \in E$*. On suppose qu'il existe un voisinage* V_0 *de* x_0 *tel que*

a) Pour tout $x \in V_0$*, l'application* $f(x,\,.\,) : \Omega \to \mathbb{K}$ *est mesurable.*

b) Pour μ*-presque tout* $\omega \in \Omega$*, l'application* $f(\,.\,,\omega) : E \to \mathbb{K}$*, est continue au point* x_0*.*

c) Il existe une fonction μ*-intégrable* g *(dépendant en général de* V_0*) telle que pour tout* $x \in V_0$ *et pour* μ*-presque tout* $\omega \in \Omega$*,* $|f(x,\omega)| \leq g(\omega)$*.*

Alors la fonction $F(x) := \int_\Omega f(x,\omega)\,\mathrm{d}\mu(\omega)$ *est bien définie sur* V_0 *et est* continue *au point* x_0*.*

Preuve. La μ-intégrabilité de $f(x,\,.\,)$ pour tout $x \in V_0$ résulte de sa mesurabilité a) et de sa domination c) par g. La fonction F est donc bien définie au moins sur V_0. Pour la continuité, il s'agit de montrer que $\lim_{x\to x_0} F(x) = F(x_0)$. Comme E est un espace métrique, il suffit pour cela de prouver que pour *toute suite* (y_n) *convergeant vers* x_0 dans E, $F(y_n)$ converge vers $F(x_0)$ dans $\mathbb{K}$. Soit donc (y_n) une telle suite. Comme V_0 est un voisinage de x_0, tous les y_n sont dans V_0 à partir d'un certain rang et on ne perd pas de généralité en supposant qu'ils y sont tous. Posons alors $h_n := f(y_n,\,.\,)$. Grâce à l'hypothèse b), la suite de fonctions mesurables (h_n) converge μ-p.p. vers $h := f(x_0,\,.\,)$. Elle est dominée μ-p.p. par la fonction intégrable g grâce à c). Le théorème de convergence dominée appliqué à (h_n) donne la conclusion. □

Remarque 4.54. Ce théorème est de nature essentiellement *locale*, on montre la régularité en un point x_0 fixé, en utilisant un voisinage de ce point. Il n'est pas difficile d'en faire un théorème *global* en remplaçant V_0 par E, en supposant dans b) que $f(\,.\,,\omega)$ est continue en tout point x de E et dans c) que l'on a la même fonction dominante g pour tous les x de E. C'est d'ailleurs sous cette forme que le théorème est souvent énoncé. Si nous lui avons préféré une version plus laborieuse, c'est parce qu'elle colle mieux à l'utilisation pratique de ce théorème où il arrive souvent qu'on ne puisse choisir la même fonction dominante g pour tous les points x de E. Voir l'exemple de la transformation de Laplace ci-après.

Exemple 4.3 (Continuité des fonctions caractéristiques)**.** Soit $X : \Omega \to \mathbb{R}^d$ un vecteur aléatoire sur l'espace probabilisé $(\Omega, \mathcal{F}, \mathbf{P})$ et $\varphi_X : \mathbb{R}^d \to \mathbb{C}$, sa *fonction caractéristique* (cf. exemple 4.1, p. 114). Rappelons que

$$\forall t = (t_1,\dots,t_d) \in \mathbb{R}^d, \quad \varphi_X(t) := \mathbf{E}\exp(i\langle t, X\rangle) = \int_\Omega \exp(i\langle t, X\rangle)\,\mathrm{d}\mathbf{P}.$$

Alors φ_X est continue sur $\mathbb{R}^d$. C'est une application immédiate du théorème 4.53 (version globale) en prenant $E = \mathbb{R}^d$, $\mathbb{K} = \mathbb{C}$, $f : \mathbb{R}^d \times \Omega \to \mathbb{C}$, $(t, \omega) \mapsto \exp(i\langle t, X(\omega)\rangle)$, $V_0 = E = \mathbb{R}^d$, $\mu = \mathbf{P}$ et pour g la variable aléatoire constante 1 qui est bien $\mathbf{P}$-intégrable sur Ω.

Exemple 4.4 (Continuité d'une transformée de Laplace)**.** Soit $f : \mathbb{R}_+ \to \mathbb{C}$ une application borélienne. On pose

$$a := \inf\Big\{x \in \mathbb{R};\ \int_{\mathbb{R}_+} \mathrm{e}^{-xu}|f(u)|\,\mathrm{d}\lambda(u) < +\infty\Big\}. \tag{4.45}$$

On suppose $a < +\infty$. Alors

$$F : x \mapsto \int_{\mathbb{R}_+} \mathrm{e}^{-xu} f(u)\,\mathrm{d}\lambda(u)$$

est définie et continue sur $]a, +\infty[$.

Preuve. On commence par noter que l'application $H : \mathbb{R} \to \overline{\mathbb{R}}_+$, $x \mapsto \int_{\mathbb{R}_+} \mathrm{e}^{-xu}|f(u)|\,\mathrm{d}\lambda(u)$ est décroissante. En effet par décroissance pour $u \geq 0$ fixé de $x \mapsto \mathrm{e}^{-xu}$, on a

$$\text{si } x_1 \leq x_2, \quad \forall u \in \mathbb{R}_+, \quad \mathrm{e}^{-x_2 u}|f(u)| \leq \mathrm{e}^{-x_1 u}|f(u)|. \tag{4.46}$$

Par croissance de l'intégrale sur $\mathcal{M}_+$, l'intégration de cette inégalité sur $\mathbb{R}_+$ (relativement à $\mathrm{d}\lambda(u)$) nous donne $H(x_2) \leq H(x_1)$.

Il en résulte que $A := \{x \in \mathbb{R};\ H(x) < +\infty\}$ est un *intervalle* non borné à droite. En effet si $x \in A$, tout $y \geq x$ vérifie $H(y) \leq H(x) < +\infty$, donc y appartient lui aussi à A. Par conséquent $A =]a, +\infty[$ ou $[a, +\infty[$ avec a défini par (4.45).

En tout cas nous disposons au moins de l'inclusion $]a, +\infty[\subset A$. On en déduit que pour $x \in]a, +\infty[$, la fonction borélienne $u \mapsto \mathrm{e}^{-xu} f(u)$ est λ-intégrable sur $\mathbb{R}_+$ et qu'ainsi $F(x)$ est bien défini[8] (à valeur dans $\mathbb{C}$).

Fixons maintenant $x_0 \in]a, +\infty[$. On se donne une marge de sécurité à gauche de x_0 en prenant $a < m_0 < x_0$, par exemple $m_0 := \frac{a+x_0}{2}$ si $a > -\infty$ ou $m_0 := x_0 - 1$ si $a = -\infty$. Alors $m_0 \in A$ et donc $H(m_0) < +\infty$. Par l'inégalité 4.46, on a

$$\forall x \in [m_0, +\infty[, \quad \forall u \in \mathbb{R}_+, \quad \mathrm{e}^{-xu}|f(u)| \leq \mathrm{e}^{-m_0 u}|f(u)|.$$

Comme $H(m_0) < +\infty$, la fonction $g : u \mapsto \mathrm{e}^{-m_0 u}|f(u)|$ est λ-intégrable sur $\mathbb{R}_+$. L'intervalle $V_0 := [m_0, +\infty[$ est un *voisinage* de x_0 et la condition de domination c) du théorème 4.53 est bien vérifiée. Peut-être est-il temps de préciser ici que nous prenons $\Omega = \mathbb{R}_+$, $\omega = u$, $\mu = \lambda$, $E =]a, +\infty[$, $\mathbb{K} = \mathbb{C}$ et que l'application f du théorème 4.53 n'a rien à voir avec le f de cet exemple. Il faudrait la remplacer par $\varphi : (x, u) \mapsto \mathrm{e}^{-xu} f(u)$ et noter que $F(x) = \int_\Omega \varphi(x, \omega)\,\mathrm{d}\lambda(\omega)$. La vérification des conditions a) et b) du théorème 4.53 étant immédiate, nous pouvons conclure que F est continue au point x_0.

Comme la seule hypothèse faite sur x_0 pour établir la continuité de F en ce point était $x_0 \in]a, +\infty[$, nous pouvons dans un deuxième temps en conclure que F est continue *en tout point* de $]a, +\infty[$. $\square$

8. Que F soit ou non définie au point $x = a$ (quand $a > -\infty$) dépend du choix de f. À titre d'exercice, examiner les cas $f(u) = u \sin u$ et $f(u) = (1 + u)^{-2}$.

Théorème 4.55 (de dérivabilité sous le signe $\int$)**.** *Soient V un ouvert de $\mathbb{R}$ et f une application $V \times \Omega \to \mathbb{K}$. On suppose que*

a) Pour tout $x \in V$, l'application $f(x, \,.\,)$ est dans $\mathcal{L}^1_{\mathbb{K}}(\Omega)$.

b) Pour μ-presque tout $\omega \in \Omega$, l'application $f(\,.\,,\omega) : V \to \mathbb{K}$, est dérivable sur V.

c) Il existe une fonction μ-intégrable g telle que pour tout $x \in V$ et pour μ-presque tout $\omega \in \Omega$, $|\frac{\partial}{\partial x} f(x,\omega)| \le g(\omega)$.

Alors la fonction $F(x) := \int_\Omega f(x,\omega)\,\mathrm{d}\mu(\omega)$ est dérivable sur V et

$$\forall x \in V, \quad F'(x) = \int_\Omega \frac{\partial f}{\partial x}(x,\omega)\,\mathrm{d}\mu(\omega).$$

Nous avons cette fois énoncé le théorème sous une forme *globale*, mais la remarque 4.54 reste valable : en pratique, on sera souvent amené à l'utiliser comme un théorème local avec f définie sur $E \times \Omega$ en prenant pour V un « petit » intervalle ouvert contenant x_0 lorsqu'on ne parvient pas à trouver une fonction dominante g qui fasse l'affaire pour tous les x du « grand » ouvert E.

Preuve. L'existence de $F(x)$ pour tout $x \in V$ découle de a). Par le même argument que dans la preuve du théorème 4.53, il suffit de montrer que pour une suite (y_n) convergente vers x_0 dans V (avec $y_n \neq x_0$ pour tout n),

$$\lim_{n\to+\infty} \frac{F(y_n) - F(x_0)}{y_n - x_0} = \int_\Omega \frac{\partial f}{\partial x}(x_0,\omega)\,\mathrm{d}\mu(\omega).$$

Posons

$$h_n(\omega) := \frac{1}{y_n - x_0}\Big(f(y_n,\omega) - f(x_0,\omega)\Big).$$

Par l'hypothèse b), il existe $D_1 \in \mathcal{F}$, tel que $\mu(D_1^c) = 0$ et pour tout $\omega \in D_1$, $f(\,.\,,\omega)$ est dérivable sur tout V. Par le théorème des accroissements finis (appliqué séparément à $\operatorname{Re} f(\,.\,,\omega)$ et $\operatorname{Im} f(\,.\,,\omega)$ quand f est à valeurs complexes), on a la majoration

$$\forall \omega \in D_1, \quad |f(y_n,\omega) - f(x_0,\omega)| \le \sup_{x\in V}\left|\frac{\partial f}{\partial x}(x,\omega)\right| |y_n - x_0|.$$

Par l'hypothèse c), il existe $D_2 \in \mathcal{F}$, tel que $\mu(D_2^c) = 0$ et pour tout $\omega \in D_2$ et tout $x \in V$, $|\frac{\partial}{\partial x} f(x,\omega)| \le g(\omega)$. L'ensemble $D_1 \cap D_2$ est de complémentaire μ-négligeable et

$$\forall \omega \in D_1 \cap D_2, \quad \lim_{n\to+\infty} h_n(\omega) = \frac{\partial f}{\partial x}(x_0,\omega) \quad \text{et} \quad \forall n \in \mathbb{N},\ |h_n(\omega)| \le g(\omega).$$

Le théorème de convergence dominée appliqué à la suite (h_n) nous dit maintenant que sa limite μ-p.p. $\frac{\partial f}{\partial x}(x_0, \,.\,)$ est μ-intégrable et que

$$\lim_{n\to+\infty} \frac{F(y_n) - F(x_0)}{y_n - x_0} = \lim_{n\to+\infty} \int_\Omega h_n(\omega)\,\mathrm{d}\mu(\omega) = \int_\Omega \frac{\partial f}{\partial x}(x_0,\omega)\,\mathrm{d}\mu(\omega).$$

$\square$

Pour des exemples d'application du théorème 4.55, on pourra consulter les Annales d'IFP 2001-02.

Théorème 4.56 (d'holomorphie sous le signe $\int$)**.** *Soit V un ouvert de $\mathbb{C}$ et $f : V \times \Omega \to \mathbb{C}$. On suppose que*

a) Pour tout $z \in V$, l'application $f(z, \,.\,) : \Omega \to \mathbb{C}$ est mesurable.

b) Pour μ-presque tout $\omega \in \Omega$, l'application $f(\,.\,, \omega) : V \to \mathbb{C}$, est holomorphe sur V.

c) Pour tout $z_0 \in V$, il existe un disque fermé *$\Delta \subset V$ de centre z_0 et une fonction μ-intégrable g (dépendant de Δ) telle que pour tout $z \in \Delta$ et pour μ-presque tout $\omega \in \Omega$, $|f(z, \omega)| \leq g(\omega)$.*

Alors la fonction $F(z) := \int_\Omega f(z, \omega) \,\mathrm{d}\mu(\omega)$ est holomorphe sur V et

$$\forall z \in V, \quad F'(z) = \int_\Omega \frac{\partial f}{\partial z}(z, \omega) \,\mathrm{d}\mu(\omega).$$

Preuve. L'existence de F en tout point de V résulte de a) et de c). Il convient de remarquer la différence d'hypothèse avec le théorème 4.55 : on n'essaie pas de contrôler *a priori* la dérivée $\frac{\partial f}{\partial z}$, on se contente de la domination de f. On va montrer que F est dérivable (en la variable complexe z) au point z_0. Pour cela, notons R le rayon du disque fermé Δ de l'hypothèse c) et $\Delta(z_0, r)$ le disque fermé concentrique de rayon $r = R/2$. La preuve sera ramenée à une simple reproduction de celle du théorème 4.55 une fois que l'on aura justifié les deux affirmations suivantes énoncées pour alléger avec une fonction holomorphe $h : V \to \mathbb{C}$.

$$\forall v \in \Delta(z_0, r), \quad |h'(v)| \leq \frac{1}{r} \sup_{z \in \Delta(z_0, R)} |h(z)|. \tag{4.47}$$

$$\forall z_1 \in \Delta(z_0, r), \quad |h(z_1) - h(z_0)| \leq |z_1 - z_0| \sup_{v \in \Delta(z_0, r)} |h'(v)|. \tag{4.48}$$

Pour vérifier (4.47), on utilise la formule de Cauchy en notant que le disque fermé de centre v et de rayon r est contenu dans $\Delta(z_0, R)$ et que comme ce dernier disque fermé est inclus dans l'ouvert V, il existe une marge de sécurité $\varepsilon > 0$ tel que le disque ouvert de centre z_0 et de rayon $R + \varepsilon$ soit lui même inclus dans V. En notant γ le cercle (orienté positivement) de centre v et de rayon r, on a en paramétrant γ par $z = v + r\mathrm{e}^{it}$, $t \in [0, 2\pi]$,

$$h'(v) = \frac{1}{2\pi i} \int_\gamma \frac{h(z)}{(z - v)^2} \,\mathrm{d}z = \frac{1}{2\pi r} \int_0^{2\pi} h(v + r\mathrm{e}^{it})\mathrm{e}^{-it} \,\mathrm{d}t,$$

ce qui donne immédiatement (4.47).

Pour établir (4.48), paramétrons le segment radial $[z_0, z_1]$ par $z = z_0 + t(z_1 - z_0)$, $t \in [0, 1]$. On a alors

$$h(z_1) - h(z_0) = \int_{[z_0, z_1]} h'(z) \,\mathrm{d}z = \int_0^1 h'\Big(z_0 + t(z_1 - z_0)\Big)(z_1 - z_0) \,\mathrm{d}t,$$

d'où l'on déduit facilement (4.48). □

Chapitre 5

Intégration sur des espaces produits

5.1 Produit de deux mesures

Étant donnés deux espaces mesurés $(\Omega_1, \mathcal{F}_1, \mu_1)$ et $(\Omega_2, \mathcal{F}_1, \mu_2)$, le but de cette section est de construire une mesure μ sur $\Omega_1 \times \Omega_2$ telle que $\mu(A_1 \times A_2) = \mu_1(A_1)\mu_2(A_2)$ pour tous $A_1 \in \mathcal{F}_1$ et $A_2 \in \mathcal{F}_2$. La première chose à faire est de munir $\Omega_1 \times \Omega_2$ d'une tribu adéquate.

5.1.1 Produit de deux tribus

Définition 5.1 (Tribu produit). *Soient $(\Omega_1, \mathcal{F}_1)$ et $(\Omega_2, \mathcal{F}_2)$ deux espaces mesurables. On appelle tribu produit de $\mathcal{F}_1$ par $\mathcal{F}_2$ et on note $\mathcal{F}_1 \otimes \mathcal{F}_2$, la tribu sur $\Omega_1 \times \Omega_2$ engendrée par la famille $\mathcal{R}$ des « rectangles mesurables » $A_1 \times A_2$, où $A_1 \in \mathcal{F}_1$ et $A_2 \in \mathcal{F}_2$:*

$$\mathcal{F}_1 \otimes \mathcal{F}_2 := \sigma\Big(\big\{A_1 \times A_2;\ A_1 \in \mathcal{F}_1,\ A_2 \in \mathcal{F}_2\big\}\Big). \tag{5.1}$$

Remarques 5.2. Il n'est sans doute pas superflu de noter d'emblée que :

a) La famille des rectangles mesurables[1] n'est pas elle même une tribu en général. Elle n'est stable ni par réunion finie ni par complémentaire.

b) Le produit de deux tribus n'est pas commutatif. La tribu $\mathcal{F}_2 \otimes \mathcal{F}_1$ n'est pas en général égale à $\mathcal{F}_1 \otimes \mathcal{F}_2$, tout simplement parce que le produit cartésien n'est pas commutatif : $\Omega_1 \times \Omega_2 \neq \Omega_2 \times \Omega_1$. Par conséquent le produit de deux mesures $\mu_1 \otimes \mu_2$, que nous allons définir sur $\mathcal{F}_1 \otimes \mathcal{F}_2$ ne sera pas non plus commutatif.

Proposition 5.3. *Les deux projections canoniques*

$$\pi_i : \Omega_1 \times \Omega_2, \quad (\omega_1, \omega_2) \mapsto \omega_i \quad (i = 1, 2)$$

sont mesurables $(\mathcal{F}_1 \otimes \mathcal{F}_2)$ - $\mathcal{F}_i$. La tribu produit $\mathcal{F}_1 \otimes \mathcal{F}_2$ est la plus petite tribu rendant mesurables les projections canoniques π_i.

1. On l'aura compris, le mot rectangle n'est pas à prendre ici au pied de la lettre.

Preuve. En effet pour tout $A_1 \in \mathcal{F}_1$, $\pi_1^{-1}(A_1) = A_1 \times \Omega_2 \in \mathcal{F}_1 \otimes \mathcal{F}_2$, ce qui établit la mesurabilité de π_1. La vérification pour π_2 est analogue. Soit maintenant $\mathcal{F}$ une tribu sur $\Omega_1 \times \Omega_2$ telle que les projections canoniques π_i soient $\mathcal{F}$ - $\mathcal{F}_i$ mesurables $(i = 1, 2)$. Nécessairement $\mathcal{F}$ contient les rectangles $\pi_1^{-1}(A_1) = A_1 \times \Omega_2$ pour tout $A_1 \in \mathcal{F}_1$ et $\pi_2^{-1}(A_2) = \Omega_1 \times A_2$ pour tout $A_2 \in \mathcal{F}_2$. La tribu $\mathcal{F}$ étant stable par intersection doit alors contenir tous les

$$(A_1 \times \Omega_2) \cap (\Omega_1 \times A_2) = A_1 \times A_2.$$

Elle contient donc tous les rectangles mesurables, donc aussi la tribu qu'ils engendrent, c'est à dire $\mathcal{F}_1 \otimes \mathcal{F}_2$. On voit ainsi que la tribu produit $\mathcal{F}_1 \otimes \mathcal{F}_2$ est la plus petite tribu rendant mesurables les projections canoniques π_i. $\square$

On introduit maintenant la notion de *section* d'une partie d'un produit cartésien. La définition et le lemme qui suit sont purement ensemblistes et ne font pas intervenir les propriétés des tribus.

Définition 5.4. *Soit $E \subset \Omega_1 \times \Omega_2$ et $\omega_1 \in \Omega_1$ fixé. On appelle section de E en ω_1, le sous-ensemble E_{ω_1} de Ω_2 défini par*

$$E_{\omega_1} := \{\omega_2 \in \Omega_2;\ (\omega_1, \omega_2) \in E\}.$$

De même pour tout $\omega_2 \in \Omega_2$, on définit la section de E en ω_2 par

$$E_{\omega_2} := \{\omega_1 \in \Omega_1;\ (\omega_1, \omega_2) \in E\}.$$

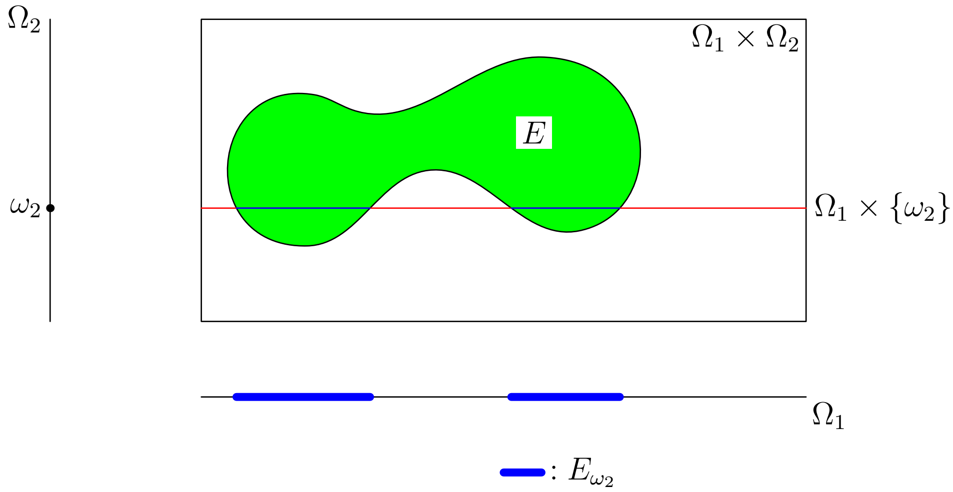


FIGURE 5.1 – Section E_{ω_2} de E en ω_2

Lemme 5.5. *La section commute avec le complémentaire, la différence propre, l'union et l'intersection. Plus précisément, si $(E_i)_{i \in I}$ est une famille quelconque (pas forcément dénombrable) de parties de $\Omega_1 \times \Omega_2$, on a pour tout $\omega_1 \in \Omega_1$,* *a)* $\big((\Omega_1 \times \Omega_2) \setminus E\big)_{\omega_1} = \Omega_2 \setminus E_{\omega_1}$ *et si* $E \subset F \subset \Omega_1 \times \Omega_2$, $(F \setminus E)_{\omega_1} = F_{\omega_1} \setminus E_{\omega_1}$ *;*

b) $\Big(\bigcup_{i\in I} E_i\Big)_{\omega_1} = \bigcup_{i\in I}(E_i)_{\omega_1}$ *;*

c) $\Big(\bigcap_{i\in I} E_i\Big)_{\omega_1} = \bigcap_{i\in I}(E_i)_{\omega_1}$

et les propriétés analogues pour les sections en $\omega_2 \in \Omega_2$*.*

Preuve. La vérification de a) s'écrit

$$\begin{aligned}
\Omega_2 \setminus E_{\omega_1} &= \Omega_2 \setminus \{\omega_2 \in \Omega_2;\ (\omega_1,\omega_2) \in E\} \\
&= \{\omega_2 \in \Omega_2;\ (\omega_1,\omega_2) \notin E\} \\
&= \{\omega_2 \in \Omega_2;\ (\omega_1,\omega_2) \in \Omega_1 \times \Omega_2 \setminus E\} \\
&= (\Omega_1 \times \Omega_2 \setminus E)_{\omega_1}.
\end{aligned}$$

Celle de b) est tout aussi facile :

$$\begin{aligned}
\Big(\bigcup_{i\in I} E_i\Big)_{\omega_1} &= \{\omega_2 \in \Omega_2;\ (\omega_1,\omega_2) \in \bigcup_{i\in I} E_i\} \\
&= \{\omega_2 \in \Omega_2;\ \exists i \in I,\ (\omega_1,\omega_2) \in E_i\} \\
&= \bigcup_{i\in I}\{\omega_2 \in \Omega_2; (\omega_1,\omega_2) \in E_i\} \\
&= \bigcup_{i\in I}(E_i)_{\omega_1}.
\end{aligned}$$

Enfin, c) s'obtient soit directement, soit par combinaison de a) et b). □

Proposition 5.6. *Toutes les sections d'un ensemble* $E \in \mathcal{F}_1 \otimes \mathcal{F}_2$ *sont mesurables, au sens suivant :*

$$\forall \omega_1 \in \Omega_1,\ E_{\omega_1} \in \mathcal{F}_2, \qquad \forall \omega_2 \in \Omega_2,\ E_{\omega_2} \in \mathcal{F}_1.$$

Preuve. Il suffit bien sûr, de faire la preuve pour les sections en ω_1. Fixons donc $\omega_1 \in \Omega_1$ et définissons

$$\mathcal{G}_1 := \{E \subset \Omega_1 \times \Omega_2;\ E_{\omega_1} \in \mathcal{F}_2\}.$$

Si on montre que $\mathcal{G}_1$ est une tribu sur $\Omega_1 \times \Omega_2$ et qu'elle contient la famille des rectangles mesurables $\mathcal{R} = \{A_1 \times A_2\,;\ A_1 \in \mathcal{F}_1,\ A_2 \in \mathcal{F}_2\}$, alors elle contiendra la tribu $\mathcal{F}_1 \otimes \mathcal{F}_2 = \sigma(\mathcal{R})$ et on aura établi que pour tout $E \in \mathcal{F}_1 \otimes \mathcal{F}_2$, $E \in \mathcal{G}_1$, donc $E_{\omega_1} \in \mathcal{F}_2$. Ceci étant vrai pour tout ω_1, la proposition sera prouvée.

Pour voir que $\mathcal{G}_1$ contient tous les rectangles mesurables $A_1 \times A_2$, il suffit d'écrire

$$\begin{aligned}
(A_1 \times A_2)_{\omega_1} &= \{\omega_2 \in \Omega_2;\ (\omega_1,\omega_2) \in A_1 \times A_2\} \\
&= \{\omega_2 \in \Omega_2;\ \omega_1 \in A_1 \text{ et } \omega_2 \in A_2\} \\
&= \begin{cases} \emptyset & \text{si } \omega_1 \notin A_1, \\ A_2 & \text{si } \omega_1 \in A_1. \end{cases}
\end{aligned} \tag{5.2}$$

Dans les deux cas, on trouve un élément de $\mathcal{F}_2$.

Vérifions que $\mathcal{G}_1$ est une tribu. La section en ω_1 de l'ensemble vide (considéré comme sous-ensemble de $\Omega_1 \times \Omega_2$) est bien évidemment l'ensemble vide (considéré comme sous-ensemble de Ω_2) donc $\emptyset \in \mathcal{G}_1$. La stabilité de $\mathcal{G}_1$ par complémentaire résulte de celle de $\mathcal{F}_2$, via le lemme 5.5 a). La stabilité de $\mathcal{G}_1$ par réunion dénombrable résulte de même de celle de $\mathcal{F}_2$, via le lemme 5.5 b). □

Voyons maintenant la question de la mesurabilité des applications partielles.

Lemme 5.7. *Soit* $(\Omega, \mathcal{F})$ *un espace mesurable quelconque et* $f : \Omega_1 \times \Omega_2 \to \Omega$ *une application* $\mathcal{F}_1 \otimes \mathcal{F}_2$ - $\mathcal{F}$ *mesurable. Alors pour tout* $\omega_1 \in \Omega_1$*, l'application partielle*

$$f(\omega_1, \,.\,) : \Omega_2 \to \Omega, \quad \omega_2 \mapsto f(\omega_1, \omega_2),$$

est $\mathcal{F}_2$ - $\mathcal{F}$ *mesurable. De même* $f(\,.\,, \omega_2)$ *est* $\mathcal{F}_1$ - $\mathcal{F}$ *mesurable pour tout* $\omega_2 \in \Omega_2$*.*

Preuve. Par symétrie, il suffit de traiter le cas de $f(\omega_1, \,.\,)$. Fixons donc ω_1 quelconque dans Ω_1 et prenons un élément quelconque B de la tribu $\mathcal{F}$.

$$\begin{aligned} f(\omega_1, \,.\,)^{-1}(B) &= \{\omega_2 \in \Omega_2;\ f(\omega_1, \,.\,)(\omega_2) \in B\} \\ &= \{\omega_2 \in \Omega_2;\ f(\omega_1, \omega_2) \in B\} \\ &= \{\omega_2 \in \Omega_2;\ (\omega_1, \omega_2) \in f^{-1}(B)\} = \bigl(f^{-1}(B)\bigr)_{\omega_1}. \end{aligned}$$

Par la $\mathcal{F}_1 \otimes \mathcal{F}_2$ - $\mathcal{F}$ mesurabilité de f, $f^{-1}(B)$ est dans la tribu $\mathcal{F}_1 \otimes \mathcal{F}_2$. La proposition 5.6 nous donne alors l'appartenance de sa section en ω_1 à la tribu $\mathcal{F}_2$. La $\mathcal{F}_2$ - $\mathcal{F}$ mesurabilité de $f(\omega_1, \,.\,)$ est ainsi établie. □

5.1.2 Construction de la mesure produit

Théorème 5.8 (mesure produit)**.** *Pour* $i = 1, 2$*, on suppose que* μ_i *est une mesure* σ*-finie sur* $(\Omega_i, \mathcal{F}_i)$*. Alors il existe une unique mesure* μ *sur* $(\Omega_1 \times \Omega_2, \mathcal{F}_1 \otimes \mathcal{F}_2)$ *telle que*

$$\forall A_1 \in \mathcal{F}_1,\ \forall A_2 \in \mathcal{F}_2, \quad \mu(A_1 \times A_2) = \mu_1(A_1)\mu_2(A_2). \tag{5.3}$$

On la note $\mu = \mu_1 \otimes \mu_2$ *et on l'appelle* mesure produit *de* μ_1 *par* μ_2*. Elle vérifie*

$$\forall E \in \mathcal{F}_1 \otimes \mathcal{F}_2, \quad \mu_1 \otimes \mu_2(E) = \int_{\Omega_1} \mu_2(E_{\omega_1})\,\mathrm{d}\mu_1(\omega_1) = \int_{\Omega_2} \mu_1(E_{\omega_2})\,\mathrm{d}\mu_2(\omega_2). \tag{5.4}$$

Preuve de l'unicité de μ*.* Réglons d'abord cette question qui est la plus facile, en supposant que μ et ν sont deux mesures sur $\mathcal{F}_1 \otimes \mathcal{F}_2$ vérifiant (5.3). Elles coïncident donc sur la classe $\mathcal{R}$ des rectangles mesurables :

$$\forall A_1 \in \mathcal{F}_1,\ \forall A_2 \in \mathcal{F}_2, \quad \mu(A_1 \times A_2) = \nu(A_1 \times A_2) = \mu_1(A_1)\mu_2(A_2). \tag{5.5}$$

On remarque alors que $\mathcal{R}$ est un π-système (*i.e.* stable par intersections finies). Ceci résulte de l'égalité

$$(A_1 \times A_2) \cap (B_1 \times B_2) = (A_1 \cap B_1) \times (A_2 \cap B_2),$$

vraie pour tous $A_1, B_1 \subset \Omega_1$ et $A_2, B_2 \subset \Omega_2$. Comme μ_i $(i = 1, 2)$ est σ-finie, il existe une suite $(\Omega_{i,n})_{n\geq 1}$ dans $\mathcal{F}_i$, croissante pour l'inclusion et telle que

$$\forall n \geq 1,\ \mu_i(\Omega_{i,n}) < +\infty \quad \text{et} \quad \Omega_{i,n} \uparrow \Omega_i, \quad (n \to +\infty), \quad i = 1, 2.$$

Il est facile d'en déduire que $\Omega_{1,n} \times \Omega_{2,n} \uparrow \Omega_1 \times \Omega_2$ et que

$$\forall n \geq 1,\ \mu(\Omega_{1,n} \times \Omega_{2,n}) = \nu(\Omega_{1,n} \times \Omega_{2,n}) = \mu_1(\Omega_{1,n})\mu_2(\Omega_{2,n}) < +\infty.$$

On peut alors conclure par le théorème d'unicité des mesures (th. 1.34) que μ et ν coïncident sur $\sigma(\mathcal{R}) = \mathcal{F}_1 \otimes \mathcal{F}_2$. $\square$

Preuve de l'existence de μ.

On va définir μ par la formule

$$\forall E \in \mathcal{F}_1 \otimes \mathcal{F}_2, \quad \mu(E) := \int_{\Omega_1} \mu_2(E_{\omega_1})\, \mathrm{d}\mu_1(\omega_1). \tag{5.6}$$

Pour motiver cette formule, regardons ce qu'elle donne dans le cas particulier où $E = A_1 \times A_2$ est un rectangle mesurable. Alors grâce à (5.2), on peut l'écrire

$$\begin{aligned} \mu(A_1 \times A_2) &= \int_{A_1} \mu_2\big((A_1 \times A_2)_{\omega_1}\big)\, \mathrm{d}\mu_1(\omega_1) + \int_{\Omega_1 \setminus A_1} \mu_2\big((A_1 \times A_2)_{\omega_1}\big)\, \mathrm{d}\mu_1(\omega_1) \\ &= \int_{A_1} \mu_2(A_2)\, \mathrm{d}\mu_1(\omega_1) + \int_{\Omega_1 \setminus A_1} \mu_2(\emptyset)\, \mathrm{d}\mu_1(\omega_1) \\ &= \mu_2(A_2) \int_{A_1} \mathrm{d}\mu_1(\omega_1) + 0 = \mu_1(A_1)\mu_2(A_2). \end{aligned}$$

Pour montrer *l'existence* de μ ayant la propriété (5.3), il suffit donc de prouver que (5.6) définit bien une fonction d'ensembles μ sur $\mathcal{F}_1 \otimes \mathcal{F}_2$ et que μ est une mesure sur cette tribu. Pour réaliser ce programme, supposons dans un premier temps que μ_2 est une mesure *finie*.

Définition de μ.

On sait par la proposition 5.6 que $\mu_2(E_{\omega_1})$ a bien un sens pour tout $E \in \mathcal{F}_1 \otimes \mathcal{F}_2$. Pour que l'intégrale définissant $\mu(E)$ dans (5.6) soit elle aussi définie, il reste à vérifier que l'application

$$h_E : \Omega_1 \to \mathbb{R}_+, \quad \omega_1 \mapsto \mu_2(E_{\omega_1})$$

est bien mesurable $\mathcal{F}_1$ - $\mathrm{Bor}(\mathbb{R}_+)$.

Mesurabilité de h_E. On utilise la même approche indirecte que dans la preuve de la proposition 5.6, en introduisant la famille d'ensembles

$$\mathcal{E} := \{E \in \mathcal{F}_1 \otimes \mathcal{F}_2;\ h_E \text{ est } \mathcal{F}_1 \text{ - } \mathrm{Bor}(\mathbb{R}_+) \text{ mesurable}\}.$$

On va montrer que $\mathcal{E}$ est une tribu qui contient la classe $\mathcal{R}$ des rectangles mesurables $A_1 \times A_2$, donc aussi $\mathcal{F}_1 \otimes \mathcal{F}_2 = \sigma(\mathcal{R})$, d'où $\mathcal{E} = \mathcal{F}_1 \otimes \mathcal{F}_2$. Ceci entraînera la mesurabilité de h_E pour tout $E \in \mathcal{F}_1 \otimes \mathcal{F}_2$.

Vérifions d'abord que tous les rectangles mesurables sont dans $\mathcal{E}$. Grâce à (5.2), on vérifie immédiatement que

$$\forall\ A_1 \times A_2 \in \mathcal{R},\ \forall \omega_1 \in \Omega_1, \quad h_{A_1\times A_2}(\omega_1) = \mu_2(A_2)\mathbf{1}_{A_1}(\omega_1).$$

Ainsi $h_{A_1\times A_2}$ est le produit de la constante positive $\mu_2(A_2)$ par la fonction $\mathcal{F}_1$ - $\mathrm{Bor}(\mathbb{R}_+)$ mesurable $\mathbf{1}_{A_1}$ (sa mesurabilité équivaut à l'appartenance de A_1 à $\mathcal{F}_1$). $h_{A_1\times A_2}$ est donc elle aussi $\mathcal{F}_1$ - $\mathrm{Bor}(\mathbb{R}_+)$ mesurable, d'où l'appartenance de $A_1 \times A_2$ à $\mathcal{E}$.

Essayons de montrer que $\mathcal{E}$ est une tribu [2]. L'ensemble vide appartient à $\mathcal{E}$. En effet, $h_\emptyset$ est la fonction nulle donc mesurable comme fonction constante [3]. Pour vérifier la stabilité par complémentaire de $\mathcal{E}$, notons $E^c := \Omega_1 \times \Omega_2 \setminus E$. On peut écrire grâce au lemme 5.5 a) et à la *finitude* de μ_2,

$$h_{E^c}(\omega_1) = \mu_2\Big((E^c)_{\omega_1}\Big) = \mu_2\Big(\Omega_2 \setminus E_{\omega_1}\Big) = \mu_2(\Omega_2) - \mu_2(E_{\omega_1}) = \mu_2(\Omega_2) - h_E(\omega_1) \geq 0.$$

Ainsi pour $E \in \mathcal{E}$, h_{E^c} est la différence de la fonction constante (finie) $\mu_2(\Omega_2)$ et de la fonction mesurable positive (à valeurs finies) h_E. Elle est donc $\mathcal{F}_1$ - $\mathrm{Bor}(\mathbb{R})$ mesurable et comme elle est positive, elle est aussi $\mathcal{F}_1$ - $\mathrm{Bor}(\mathbb{R}_+)$ mesurable, ce qui établit l'appartenance de E^c à $\mathcal{E}$. Pour la stabilité de $\mathcal{E}$ par union dénombrable, soit $(E_k)_{k\geq 1}$ une suite dans $\mathcal{E}$ et $E := \bigcup\limits_{k\in\mathbb{N}^*} E_k$. En notant $F_n := \bigcup\limits_{1\leq k\leq n} E_k$, on a $F_n \uparrow E$ et $h_{F_n} \uparrow h_E$ par continuité séquentielle croissante de μ_2. Il suffit donc d'établir la mesurabilité de h_{F_n} pour tout $n \geq 1$, autrement dit la stabilité de $\mathcal{E}$ par union finie. On s'est ainsi ramené à vérifier que si A et B sont dans $\mathcal{E}$, $A \cup B$ y est aussi. C'est facile dans le cas particulier où A et B sont disjoints puisque le lemme 5.5 b) et l'additivité de μ_2 nous permettent d'écrire $h_{A\cup B} = h_A + h_B$, de sorte que $h_{A\cup B}$ est mesurable comme somme de deux fonctions mesurables et $A \cup B \in \mathcal{E}$. Malheureusement dans le cas où A et B ne sont pas disjoints, on ne voit pas bien comment montrer directement la mesurabilité de $h_{A\cup B}$. Notons au passage que le cas A et B disjoints nous donne la stabilité de $\mathcal{E}$ pour la différence propre : si $C, D \in \mathcal{E}$ et $C \subset D$, en posant $A = C$ et $B = D \setminus C$, l'égalité $h_{A\cup B} = h_A + h_B$ nous donne $h_{D\setminus C} = h_D - h_C$ et assure la mesurabilité de $h_{D\setminus C}$, donc l'appartenance de $D \setminus C$ à $\mathcal{E}$. On voit aussi que l'appartenance de $A \cup B$ (dans le cas général) à $\mathcal{E}$ équivaut à celle de $A \cap B$, mais cela ne nous avance guère.

Dans cette tentative infructueuse pour montrer que $\mathcal{E}$ est une tribu, on a quand même établi entre autres que $\mathcal{E}$ contient la π-classe $\mathcal{R}$ des rectangles mesurables, qu'elle est stable par réunion croissante et par différence propre, donc que $\mathcal{E}$ est une λ-classe. Par le théorème de Dynkin (cf. th. 1.15), $\mathcal{E}$ contient $\sigma(\mathcal{R}) = \mathcal{F}_1 \otimes \mathcal{F}_2$. Comme par construction, $\mathcal{E} \subset \mathcal{F}_1 \otimes \mathcal{F}_2$, on a l'égalité $\mathcal{E} = \mathcal{F}_1 \otimes \mathcal{F}_2$, ce qui achève la preuve de la mesurabilité de h_E pour tout $E \in \mathcal{F}_1 \otimes \mathcal{F}_2$.

On a donc montré que (5.6) définissait bien une fonction d'ensembles sur $\mathcal{F}_1 \otimes \mathcal{F}_2$ lorsque μ_2 est une mesure finie. Lorsque μ_2 n'est pas finie, sa σ-finitude nous assure de

2. L'approche exposée ci-dessous n'est délibérément pas la plus courte, elle a pour but de localiser la difficulté et de motiver l'invocation du théorème de Dynkin pour la solution de ce problème.

3. Une fonction constante $\Omega_1 \to \mathbb{R}_+$ est déjà mesurable lorsque l'on munit Ω_1 de la tribu triviale $\{\emptyset, \Omega_1\}$, donc *a fortiori* pour toute tribu sur Ω_1.

l'existence dans $\mathcal{F}_2$ d'une suite $\Omega_{2,n} \uparrow \Omega_2$ telle que pour tout $n \geq 1$, $\mu_2(\Omega_{2,n}) < +\infty$. Le cas μ_2 finie montre alors que $\mathcal{E}$ contient tous les $E \in \mathcal{F}_1 \otimes \mathcal{F}_2$ qui sont inclus dans au moins un $\Omega_1 \times \Omega_{2,n}$. Soit alors F quelconque dans $\mathcal{F}_1 \otimes \mathcal{F}_2$. On a $F_n := F \cap (\Omega_1 \times \Omega_{2,n}) \uparrow F$ et chaque F_n est dans $\mathcal{E}$. Comme la propriété de stabilité par union croissante de $\mathcal{E}$ n'utilisait pas la finitude de μ_2, on en déduit que $F = \cup_{n \geq 1} F_n$ est lui aussi dans $\mathcal{E}$. On a ainsi établi l'inclusion $\mathcal{F}_1 \otimes \mathcal{F}_2 \subset \mathcal{E}$ et donc l'égalité $\mathcal{F}_1 \otimes \mathcal{F}_2 = \mathcal{E}$ dans le cas où μ_2 est σ-finie.

La fonction d'ensembles μ définie par (5.6) est une mesure. Il est clair que

$$\mu(\emptyset) = \int_{\Omega_1} h_\emptyset \, \mathrm{d}\mu_1 = \int_{\Omega_1} 0 \, \mathrm{d}\mu_1 = 0.$$

Soient $(E_n)_{n \geq 1}$ une suite d'éléments deux à deux disjoints de $\mathcal{F}_1 \otimes \mathcal{F}_2$ et $E := \cup_{n \geq 1} E_n$. Alors pour tout $\omega_1 \in \Omega_1$ les sections E_{n,ω_1} sont des éléments deux à deux disjoints de $\mathcal{F}_2$ (proposition 5.6 et lemme 5.5 c)). Par le lemme 5.5 b), E_{ω_1} s'écrit comme la réunion disjointe $E_{\omega_1} = \cup_{n \geq 1} E_{n,\omega_1}$ et par σ additivité de μ_2,

$$h_E = \sum_{n=1}^{+\infty} h_{E_n}.$$

En utilisant le corollaire du théorème de Beppo Levi pour les séries de fonctions mesurables positives, on en déduit

$$\mu(E) = \int_{\Omega_1} h_E \, \mathrm{d}\mu_1 = \int_{\Omega_1} \sum_{n=1}^{+\infty} h_{E_n} \, \mathrm{d}\mu_1 = \sum_{n=1}^{+\infty} \int_{\Omega_1} h_{E_n} \, \mathrm{d}\mu_1 = \sum_{n=1}^{+\infty} \mu(E_n),$$

ce qui établit la σ-additivité de μ et achève la preuve de l'existence de μ.

Dans la preuve de *l'existence* de μ, on a utilisé la σ-finitude de μ_2, pas celle de μ_1. Cependant nous avons eu besoin de la σ-finitude des deux mesures pour prouver l'unicité de μ. Un autre avantage de supposer μ_1 σ-finie est que l'on peut échanger les rôles de μ_1 et μ_2 dans la construction de μ et obtenir ainsi la formule (5.4). □

5.1.3 Application aux calculs de volumes et d'espérances

Voici une première application importante de la mesure produit.

Proposition 5.9. *Notons λ_d la mesure de Lebesgue sur $\mathbb{R}^d$, $(\Omega, \mathcal{F}, \nu)$ un espace mesuré, la mesure ν étant σ-finie et $f : \Omega \to \mathbb{R}$ une application $\mathcal{F}$ - $\mathrm{Bor}(\mathbb{R})$ mesurable positive. On définit son hypographe par*

$$G := \{(\omega, y) \in \Omega \times \mathbb{R};\ 0 \leq y \leq f(\omega)\}.$$

Alors G appartient à la tribu $\mathcal{F} \otimes \mathrm{Bor}(\mathbb{R})$ et

$$\nu \otimes \lambda_1(G) = \int_\Omega f \, \mathrm{d}\nu. \tag{5.7}$$

En particulier quand $\Omega = \mathbb{R}$ et $\nu = \lambda_1$, (5.7) donne (grâce à l'égalité $\lambda_1 \otimes \lambda_1 = \lambda_2$) l'interprétation de $\int_{\mathbb{R}} f \,\mathrm{d}\lambda_1$ comme l'aire de la région délimitée par le graphe de f et l'axe des abscisses. De même, quand $\Omega = \mathbb{R}^d$ et $\nu = \lambda_d$, identifions $\mathbb{R}^d \times \mathbb{R}$ avec $\mathbb{R}^{d+1}$, notons x un élément générique de $\mathbb{R}^d$ et y la dernière coordonnée dans $\mathbb{R}^{d+1}$. La formule (5.7) permet alors, grâce à l'égalité $\lambda_{d+1} = \lambda_d \otimes \lambda_1$, d'interpréter $\int_{\mathbb{R}^d} f(x) \,\mathrm{d}\lambda_d(x)$ comme le volume de la région délimitée par l'hyperplan de $\mathbb{R}^{d+1}$ d'équation $y = 0$ et l'hypersurface d'équation $y = f(x)$.

L'égalité $\lambda_{d+1} = \lambda_d \otimes \lambda_1$ est admise provisoirement. Elle sera établie à la sous-section 5.3.1, remarque 5.17.

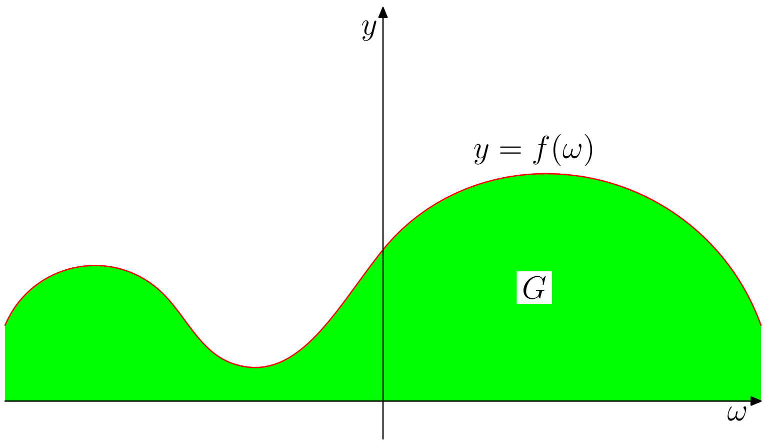


FIGURE 5.2 – Hypographe de f

Preuve. Vérifions d'abord l'appartenance de G à la tribu produit. Pour cela, on introduit l'application

$$\varphi : \Omega \times \mathbb{R} \to \mathbb{R}^2, \quad (\omega, y) \mapsto \big(f(\omega), y\big)$$

et on note que $G = \varphi^{-1}(H)$ où $H := \{(x, y) \in \mathbb{R}^2 ; \, 0 \leq y \leq x\}$ est fermé (donc borélien) de $\mathbb{R}^2$. Il suffit alors de montrer que φ est mesurable $\mathcal{F} \otimes \mathrm{Bor}(\mathbb{R})$ - $\mathrm{Bor}(\mathbb{R}^2)$ pour établir l'appartenance de G à $\mathcal{F} \otimes \mathrm{Bor}(\mathbb{R})$. Pour la mesurabilité de φ, on vérifie facilement que pour tout pavé de $\mathbb{R}^2$ de la forme $]a, b] \times]c, d]$,

$$\varphi^{-1}(]a, b] \times]c, d]) = f^{-1}(]a, b]) \times]c, d].$$

Comme f est $\mathcal{F}$ - $\mathrm{Bor}(\mathbb{R})$ mesurable, $f^{-1}(]a, b]) \in \mathcal{F}$ et on voit ainsi que $\varphi^{-1}(]a, b] \times]c, d])$ est un rectangle mesurable de la tribu $\mathcal{F} \otimes \mathrm{Bor}(\mathbb{R})$. La classe des $]a, b] \times]c, d]$ engendrant $\mathrm{Bor}(\mathbb{R}^2)$, on a prouvé la mesurabilité requise pour φ.

On applique alors la formule (5.6) de définition de la mesure produit en notant que la section G_ω est le segment $[0, f(\omega)]$ donc $\lambda_1(G_\omega) = f(\omega)$ (par hypothèse, $f(\omega) \geq 0$) :

$$\nu \otimes \lambda_1(G) = \int_\Omega \lambda_1(G_\omega) \,\mathrm{d}\nu(\omega) = \int_\Omega f(\omega) \,\mathrm{d}\nu(\omega).$$

□

En adaptant l'argument utilisé dans la preuve ci-dessus, il est facile de voir que le graphe de f, *i.e.* l'ensemble des couples (ω, y) tels que $y = f(\omega)$ est de $\nu \otimes \lambda_1$ mesure nulle (exercice).

Voici maintenant une application probabiliste analogue à la proposition 5.9.

Proposition 5.10. *Si X est une variable aléatoire* positive *sur l'espace probabilisé* $(\Omega, \mathcal{F}, \mathbf{P})$,

$$\mathbf{E}X = \int_{\mathbb{R}_+} \mathbf{P}(X \geq x)\, \mathrm{d}\lambda_1(x) = \int_{\mathbb{R}_+} \mathbf{P}(X > x)\, \mathrm{d}\lambda_1(x). \tag{5.8}$$

Cette formule est utile pour calculer directement l'espérance d'une variable aléatoire positive à partir de sa fonction de répartition $F(x) = \mathbf{P}(X \leq x)$ puisque $\mathbf{P}(X > x) = 1 - F(x)$. La deuxième égalité dans (5.8) n'est pas surprenante puisque les fonctions intégrées ne diffèrent qu'aux points de discontinuité de la fonction monotone F dont l'ensemble est au plus dénombrable, donc de λ_1 mesure nulle.

Preuve. Pour la première égalité dans (5.8), posons

$$G := \{(\omega, x) \in \Omega \times \mathbb{R};\ 0 \leq x \leq X(\omega)\}.$$

Par la proposition 5.9, on a $\mathbf{E}X = \int_\Omega X\, \mathrm{d}\mathbf{P} = \mathbf{P} \otimes \lambda_1(G)$ et en utilisant [4] (5.4),

$$\begin{aligned} \mathbf{P} \otimes \lambda_1(G) = \int_{\mathbb{R}_+} \mathbf{P}(G_x)\, \mathrm{d}\lambda_1(x) &= \int_{\mathbb{R}_+} \mathbf{P}(\{\omega \in \Omega;\ 0 \leq x \leq X(\omega)\})\, \mathrm{d}\lambda_1(x) \\ &= \int_{\mathbb{R}_+} \mathbf{P}(X \geq x)\, \mathrm{d}\lambda_1(x). \end{aligned}$$

À titre d'exercice, on obtiendra la deuxième égalité de (5.8) de la même façon en remplaçant G par $\{(\omega, x) \in \Omega \times \mathbb{R}_+;\ 0 \leq x < f(\omega)\}$. □

5.2 Intégrales doubles

On étudie maintenant l'intégration par rapport à une mesure produit. Il s'agit de voir sous quelles conditions une intégrale $\int_{\Omega_1 \times \Omega_2} f\, \mathrm{d}(\mu_1 \otimes \mu_2)$ est égale à une intégrale double (ou itérée) de la forme $\int_{\Omega_1} \left\{ \int_{\Omega_2} f(\omega_1, \,.\,)\, \mathrm{d}\mu_2 \right\} \mathrm{d}\mu_1(\omega_1)$.

5.2.1 Cas des fonctions mesurables positives

L'intégrale par rapport à une mesure produit d'une fonction mesurable positive sur $\Omega_1 \times \Omega_2$ est toujours égale (dans $\overline{\mathbb{R}}_+$) à chacune des deux intégrales itérées $\int_{\Omega_1} \int_{\Omega_2}$ et $\int_{\Omega_2} \int_{\Omega_1}$. Ce sympathique résultat généralise aux intégrales la propriété d'échange des sommations pour les séries doubles à termes positifs. L'énoncé précis est le suivant.

4. Remarquer que pour $x < 0$, la section G_x se réduit à l'ensemble vide d'où $\mathbf{P}(G_x) = 0$, d'où l'intégrale sur $\mathbb{R}_+$ au lieu de $\mathbb{R}$.

Théorème 5.11 (Fubini - Tonelli)**.** *Pour $i = 1, 2$, on suppose que μ_i est une mesure σ-finie sur $(\Omega_i, \mathcal{F}_i)$. Soit $f : \Omega_1 \times \Omega_2 \to \overline{\mathbb{R}}_+$ une application $\mathcal{F}_1 \otimes \mathcal{F}_2$ - $\mathrm{Bor}(\overline{\mathbb{R}}_+)$ mesurable positive. Les applications F_i*

$$F_1 : \omega_1 \mapsto \int_{\Omega_2} f(\omega_1, \omega_2)\,\mathrm{d}\mu_2(\omega_2) \quad \textit{et} \quad F_2 : \omega_2 \mapsto \int_{\Omega_1} f(\omega_1, \omega_2)\,\mathrm{d}\mu_1(\omega_1)$$

sont $\mathcal{F}_i$ - $\mathrm{Bor}(\overline{\mathbb{R}}_+)$ mesurables et vérifient

$$\begin{aligned} \int_{\Omega_1 \times \Omega_2} f\,\mathrm{d}(\mu_1 \otimes \mu_2) &= \int_{\Omega_1} \left\{ \int_{\Omega_2} f(\omega_1, \omega_2)\,\mathrm{d}\mu_2(\omega_2) \right\} \mathrm{d}\mu_1(\omega_1) && (5.9) \\ &= \int_{\Omega_2} \left\{ \int_{\Omega_1} f(\omega_1, \omega_2)\,\mathrm{d}\mu_1(\omega_1) \right\} \mathrm{d}\mu_2(\omega_2). && (5.10) \end{aligned}$$

Preuve. Avant d'examiner la mesurabilité des F_i, notons que grâce au lemme 5.7 appliqué avec $\Omega = \overline{\mathbb{R}}_+$, les applications partielles $f(\omega_1, \,.\,)$ et $f(\,.\,, \omega_2)$ sont mesurables positives et les intégrales définissant les F_i ont un sens comme éléments de $\overline{\mathbb{R}}_+$.

Considérons maintenant le cas particulier où f est l'indicatrice d'un ensemble E appartenant à la tribu $\mathcal{F}_1 \otimes \mathcal{F}_2$. On a alors d'après (5.4),

$$\int_{\Omega_1 \times \Omega_2} f\,\mathrm{d}(\mu_1 \otimes \mu_2) = \mu_1 \otimes \mu_2(E) = \int_{\Omega_1} \mu_2(E_{\omega_1})\,\mathrm{d}\mu_1(\omega_1) = \int_{\Omega_2} \mu_1(E_{\omega_2})\,\mathrm{d}\mu_2(\omega_2). \quad (5.11)$$

D'autre part, on vérifie facilement que

$$\mathbf{1}_E(\omega_1, \omega_2) = \mathbf{1}_{E_{\omega_1}}(\omega_2) = \mathbf{1}_{E_{\omega_2}}(\omega_1),$$

d'où

$$F_1(\omega_1) = \int_{\Omega_2} \mathbf{1}_E(\omega_1, \omega_2)\,\mathrm{d}\mu_2(\omega_2) = \int_{\Omega_2} \mathbf{1}_{E_{\omega_1}}(\omega_2)\,\mathrm{d}\mu_2(\omega_2) = \mu_2(E_{\omega_1}).$$

et symétriquement $F_2(\omega_2) = \mu_1(E_{\omega_2})$. La mesurabilité de $F_1 : \omega_1 \mapsto \mu_2(E_{\omega_1})$ a été établie dans la preuve du théorème 5.8. Celle de F_2 s'en déduit par symétrie. En reportant les expressions trouvées pour F_1 et F_2 dans (5.11), on a bien la double égalité (5.9)–(5.10) et le théorème est prouvé dans le cas particulier où f est l'indicatrice d'un ensemble $E \in \mathcal{F}_1 \otimes \mathcal{F}_2$.

La mesurabilité des F_i et la vérification de la double égalité (5.9)–(5.10) dans le cas général s'en déduisent par la méthode d'extension standard en considérant successivement les fonctions étagées puis les limites croissantes de fonctions étagées... la rédaction détaillée est laissée en exercice. □

5.2.2 Cas général

Théorème 5.12 (Fubini)**.** *Pour $i = 1, 2$, on suppose que μ_i est une mesure σ-finie sur $(\Omega_i, \mathcal{F}_i)$. Soit $f : \Omega_1 \times \Omega_2 \to \mathbb{K}$, $\mathbb{K} = \mathbb{R}$ ou $\mathbb{C}$ une application $\mathcal{F}_1 \otimes \mathcal{F}_2$ - $\mathrm{Bor}(\mathbb{K})$ mesurable.*

a) Pour que f soit $\mu_1 \otimes \mu_2$ intégrable, il faut et il suffit que l'une des deux intégrales

$$I = \int_{\Omega_1} \left\{ \int_{\Omega_2} |f(\omega_1, \omega_2)|\,\mathrm{d}\mu_2(\omega_2) \right\} \mathrm{d}\mu_1(\omega_1), \quad J = \int_{\Omega_2} \left\{ \int_{\Omega_1} |f(\omega_1, \omega_2)|\,\mathrm{d}\mu_1(\omega_1) \right\} \mathrm{d}\mu_2(\omega_2)$$

soit finie.

b) Si f est $\mu_1 \otimes \mu_2$ intégrable, la fonction $f(\omega_1, .)$ est μ_2 intégrable sur Ω_2 pour μ_1 presque tout $\omega_1 \in \Omega_1$. La fonction $\omega_1 \mapsto \int_{\Omega_2} f(\omega_1, \omega_2) \,\mathrm{d}\mu_2(\omega_2)$ est définie μ_1 presque partout sur Ω_1 et μ_1 intégrable sur Ω_1. L'énoncé analogue obtenu en permutant les rôles de μ_1 et μ_2 est aussi valable.

c) Si f est $\mu_1 \otimes \mu_2$ intégrable, on a

$$\int_{\Omega_1 \times \Omega_2} f \,\mathrm{d}(\mu_1 \otimes \mu_2) = \int_{\Omega_1} \left\{ \int_{\Omega_2} f(\omega_1, \omega_2) \,\mathrm{d}\mu_2(\omega_2) \right\} \mathrm{d}\mu_1(\omega_1) \tag{5.12}$$

$$= \int_{\Omega_2} \left\{ \int_{\Omega_1} f(\omega_1, \omega_2) \,\mathrm{d}\mu_1(\omega_1) \right\} \mathrm{d}\mu_2(\omega_2). \tag{5.13}$$

Preuve. Pour tout $\omega_1 \in \Omega_1$, $f(\omega_1, .)$ est $\mathcal{F}_2$ - Bor($\mathbb{K}$) mesurable par le lemme 5.7. On peut alors définir la fonction

$$G_1 : \Omega_1 \to \overline{\mathbb{R}}_+, \quad G_1(\omega_1) := \int_{\Omega_2} |f(\omega_1, .)| \,\mathrm{d}\mu_2.$$

Par le théorème de Fubini-Tonelli, G_1 est une fonction mesurable positive sur Ω_1. Il en est de même pour G_2 obtenue en échangeant les rôles de Ω_1 et Ω_2. On a alors $I = \int_{\Omega_1} G_1 \,\mathrm{d}\mu_1$ et $J = \int_{\Omega_2} G_2 \,\mathrm{d}\mu_2$.

La condition nécessaire et suffisante d'intégrabilité de f par rapport à $\mu_1 \otimes \mu_2$ s'écrit comme d'habitude $\int_{\Omega_1 \times \Omega_2} |f| \,\mathrm{d}(\mu_1 \otimes \mu_2) < +\infty$. Le théorème de Fubini-Tonelli appliqué à la fonction mesurable positive $|f|$ nous dit que cette intégrale est *toujours* égale à I et à J, qu'elle soit finie ou non. La c.n.s. d'intégrabilité a) en résulte.

Pour vérifier le point b), on suppose désormais que $I = \int_{\Omega_1} G_1 \,\mathrm{d}\mu_1 < +\infty$. Alors G_1 est finie μ_1 presque partout sur Ω_1 :

$$\Omega_1' := \{\omega_1 \in \Omega_1;\ G_1(\omega_1) < +\infty\} \in \mathcal{F}_1 \quad \text{et} \quad \mu_1(\Omega_1 \setminus \Omega_1') = 0.$$

Par définition de G_1 et Ω_1', on voit alors que pour tout $\omega_1 \in \Omega_1'$, l'application partielle mesurable $f(\omega_1, .)$ est μ_2 intégrable sur Ω_2. On peut donc définir

$$F_1 : \Omega_1 \to \mathbb{K}, \quad F_1(\omega_1) := \begin{cases} \int_{\Omega_2} f(\omega_1, .) \,\mathrm{d}\mu_2 & \text{si } \omega_1 \in \Omega_1', \\ 0 & \text{si } \omega_1 \notin \Omega_1'. \end{cases}$$

Si f est réelle, en séparant f^+ et f^- et en leur appliquant le théorème 5.11, on voit que $\int_{\Omega_2} f^+(\omega_1, .) \,\mathrm{d}\mu_2$ et $\int_{\Omega_2} f^-(\omega_1, .) \,\mathrm{d}\mu_2$ sont à valeurs positives *finies* pour chaque $\omega_1 \in \Omega_1'$ et mesurables comme fonctions $\Omega_1 \to \overline{\mathbb{R}}_+$. En écrivant

$$F_1(\omega_1) = \mathbf{1}_{\Omega_1'}(\omega_1) \int_{\Omega_2} f^+(\omega_1, .) \,\mathrm{d}\mu_2 - \mathbf{1}_{\Omega_1'}(\omega_1) \int_{\Omega_2} f^-(\omega_1, .) \,\mathrm{d}\mu_2, \tag{5.14}$$

on voit que F_1 est $\mathcal{F}_1$ - Bor($\mathbb{R}$) mesurable. En effet, chacun des deux termes de cette différence est mesurable $\mathcal{F}_1$ - Bor($\overline{\mathbb{R}}_+$) comme produit de deux fonctions ayant cette même mesurabilité. Pour $\int_{\Omega_2} f^{\pm}(\omega_1, .) \,\mathrm{d}\mu_2$, cette mesurabilité est garantie par le théorème de Fubini-Tonelli. Grâce à l'indicatrice et à la convention arithmétique « $0 \times (+\infty) = 0$ », la fonction $\omega_1 \mapsto \mathbf{1}_{\Omega_1'}(\omega_1) \int_{\Omega_2} f^+(\omega_1, \,.\,) \,\mathrm{d}\mu_2$ ne prend que des valeurs *finies*. On en déduit facilement qu'elle est *aussi* mesurable $\mathcal{F}_1$ - Bor($\mathbb{R}$) (utiliser le corollaire 2.10). On a clairement le même résultat avec f^- à la place de f^+. Par conséquent F_1 est mesurable $\mathcal{F}_1$ - Bor($\mathbb{R}$) comme différence de deux fonctions ayant cette même mesurabilité [5].

Lorsque f est à valeurs complexes, on obtient la mesurabilité de F_1 en séparant partie réelle et partie imaginaire de f. Enfin, la μ_1 intégrabilité de F_1 résulte de l'inégalité $|F_1(\omega_1)| \leq \int_{\Omega_2} |f(\omega_1, \,.\,)| \,\mathrm{d}\mu_2$ et de l'hypothèse $I = \int_{\Omega_1} G_1 \,\mathrm{d}\mu_1 < +\infty$.

Pour vérifier c), si $\mathbb{K} = \mathbb{R}$ et f est $\mu_1 \otimes \mu_2$ intégrable, on obtient les égalités (5.12) et (5.13) en écrivant $f = f^+ - f^-$ et en appliquant les égalités analogues (5.9) et (5.10) du théorème de Fubini-Tonelli aux fonctions mesurables positives f^+ et f^-. Il n'y a aucun problème pour les différences lorsqu'on recolle les morceaux puisque toutes les intégrales concernées sont finies. Le cas complexe se déduit du cas réel par séparation des parties réelle et imaginaire. $\square$

5.2.3 Fonctions $f_1 \otimes f_2$

Un cas particulier important est celui où f peut s'écrire comme le produit tensoriel $f_1 \otimes f_2$ de deux fonctions $f_i : \Omega_i \to \mathbb{K}$, $i = 1, 2$ d'une variable :

$$f_1 \otimes f_2 : \Omega_1 \times \Omega_2 \to \mathbb{K}, \quad (\omega_1, \omega_2) \mapsto f_1(\omega_1) f_2(\omega_2).$$

Étudions la mesurabilité [6] de $f_1 \otimes f_2$. Supposons d'abord que chaque f_i soit $\mathcal{F}_i$ - Bor($\mathbb{K}$) mesurable ($\mathbb{K} = \overline{\mathbb{R}}_+$, $\mathbb{R}$ ou $\mathbb{C}$). Écrivons alors $f_1 \otimes f_2 = p \circ \varphi$ où $p : (x, y) \mapsto xy$ est la multiplication $\mathbb{K}^2 \to \mathbb{K}$ et $\varphi : \Omega_1 \times \Omega_2 \to \mathbb{K}^2$, $(\omega_1, \omega_2) \mapsto \big(f_1(\omega_1), f_2(\omega_2)\big)$. Pour $\mathbb{K} = \mathbb{R}$ ou $\mathbb{C}$, p est continue, donc borélienne et il suffit de montrer la mesurabilité de φ. Si $\mathbb{K} = \overline{\mathbb{R}}_+$, p est discontinue aux points $(0, +\infty)$ et $(+\infty, 0)$ et on établit sa mesurabilité directement (exercice). Pour montrer la mesurabilité de φ, on utilise le fait que la tribu borélienne de $\mathbb{K}^2$ est engendrée par une classe $\mathcal{C}$ de boréliens de la forme $B_1 \times B_2$, où les B_i sont eux mêmes membres d'une classe particulière de boréliens de $\mathbb{K}$. Pour $\mathbb{K} = \mathbb{R}$ ou $\overline{\mathbb{R}}_+$, on peut prendre par exemple, les B_i de la forme $[a, b]$ avec $a, b \in \mathbb{K}$. Pour $\mathbb{K} = \mathbb{C}$, on peut en passant par l'identification de $\mathbb{C}$ et $\mathbb{R}^2$, prendre les $B = \{z \in \mathbb{C};\ \mathrm{Re}(z) \in [a, b],\ \mathrm{Im}(z) \in [c, d]\}$, avec a, b, c, d réels. Il suffit alors de noter que pour $B_1 \times B_2 \in \mathcal{C}$,

$$\begin{aligned}
\varphi^{-1}(B_1 \times B_2) &= \Big\{(\omega_1, \omega_2) \in \Omega_1 \times \Omega_2;\ \big(f_1(\omega_1), f_2(\omega_2)\big) \in B_1 \times B_2\Big\} \\
&= \Big\{(\omega_1, \omega_2) \in \Omega_1 \times \Omega_2;\ f_1(\omega_1) \in B_1 \text{ et } f_2(\omega_2) \in B_2\Big\} \\
&= \Big\{(\omega_1, \omega_2) \in \Omega_1 \times \Omega_2;\ \omega_1 \in f_1^{-1}(B_1) \text{ et } \omega_2 \in f_2^{-1}(B_2)\Big\} \\
&= f_1^{-1}(B_1) \times f_2^{-1}(B_2)
\end{aligned}$$

Chaque f_i étant $\mathcal{F}_i$ - Bor($\mathbb{K}$) mesurable, $f_i^{-1}(B_i) \in \mathcal{F}_i$ d'où $f_1^{-1}(B_1) \times f_2^{-1}(B_2) \in \mathcal{F}_1 \otimes \mathcal{F}_2$. Ceci étant vrai pour tout $B_1 \times B_2 \in \mathcal{C}$, on obtient $\varphi^{-1}(\mathcal{C}) \subset \mathcal{F}_1 \otimes \mathcal{F}_2$. On en déduit

5. Dans (5.14) il serait catastrophique de *mettre en facteur* $\mathbf{1}_{\Omega_1'}$. Voyez vous pourquoi ?
6. Passage à sauter en première lecture. Allez directement au corollaire 5.13.

$\sigma\Big(\varphi^{-1}(\mathcal{C})\Big) \subset \mathcal{F}_1 \otimes \mathcal{F}_2$. D'autre part $\sigma\Big(\varphi^{-1}(\mathcal{C})\Big) = \varphi^{-1}\Big(\sigma(\mathcal{C})\Big) = \varphi^{-1}\Big(\mathrm{Bor}(\mathbb{K}^2)\Big)$. On a ainsi établi l'inclusion $\varphi^{-1}\Big(\mathrm{Bor}(\mathbb{K}^2)\Big) \subset \mathcal{F}_1 \otimes \mathcal{F}_2$ qui signifie la mesurabilité de φ.

Réciproquement, la mesurabilité de $f = f_1 \otimes f_2$ implique-t-elle celle des f_i ? La réponse est oui, sauf dans des cas pathologiques. En effet d'après le lemme 5.7, on sait que $f(\omega_1, \,.\,)$ et $f(\,.\,, \omega_2)$ sont mesurables. Traitons d'abord le cas $\mathbb{K} = \mathbb{R}$ ou $\mathbb{C}$. Si donc on peut choisir ω_i tel que $f_i(\omega_i) \neq 0$ $(i = 1, 2)$, l'égalité évidente $f(\omega_1, \,.\,) = f_1(\omega_1)f_2$ nous donne $f_2 = c_1 f(\omega_1, \,.\,)$, la constante $c = 1/f_1(\omega_1)$ étant un nombre réel ou complexe[7]. De même $f_1 = c_2 f(\,.\,, \omega_2)$. Si $\mathbb{K} = \overline{\mathbb{R}}_+$, s'il existe ω_i tel que $0 < f_i(\omega_i) < +\infty$, on est ramenés à la situation précédente. Par contre si disons f_1 ne prend que les valeurs 0 ou $+\infty$, on peut avoir f mesurable sans que f_2 le soit. Pour le voir, prendre $f_1 = (+\infty)\mathbf{1}_A$, où $A \in \mathcal{F}_1$ et $f_2 = 1 + \mathbf{1}_B$ où B est une partie de Ω_2 n'appartenant pas à $\mathcal{F}_2$. Alors $f = (+\infty)\mathbf{1}_{A\times\Omega_2}$ est mesurable puisque $A \times \Omega_2 \in \mathcal{F}_1 \otimes \mathcal{F}_2$. Par contre f_2 n'est pas mesurable car $B \notin \mathcal{F}_2$.

Corollaire 5.13 (Variables séparées)**.** *Dans le cas où f est de la forme $f_1 \otimes f_2$, si chaque f_i est μ_i intégrable $(i = 1, 2)$, alors f est $\mu_1 \otimes \mu_2$ intégrable et*

$$\int_{\Omega_1\times\Omega_2} f_1 \otimes f_2 \,\mathrm{d}(\mu_1 \otimes \mu_2) = \left\{\int_{\Omega_1} f_1 \,\mathrm{d}\mu_1\right\}\left\{\int_{\Omega_2} f_2 \,\mathrm{d}\mu_2\right\}. \tag{5.15}$$

Preuve. Vérifions d'abord la $\mu_1 \otimes \mu_2$ intégrabilité de $f_1 \otimes f_2$. Les f_i étant intégrables sont en particulier mesurables, donc $f_1 \otimes f_2$ est aussi mesurable, comme nous venons de le voir. Pour établir la finitude de $\int_{\Omega_1\times\Omega_2} |f_1 \otimes f_2| \,\mathrm{d}(\mu_1 \otimes \mu_2)$, on utilise le théorème de Fubini-Tonelli :

$$\begin{aligned}
\int_{\Omega_1\times\Omega_2} |f_1 \otimes f_2| \,\mathrm{d}(\mu_1 \otimes \mu_2) &= \int_{\Omega_1} \left\{\int_{\Omega_2} |f_1(\omega_1)||f_2(\omega_2)| \,\mathrm{d}\mu_2(\omega_2)\right\} \mathrm{d}\mu_1(\omega_1) \\
&= \int_{\Omega_1} |f_1(\omega_1)| \left\{\int_{\Omega_2} |f_2| \,\mathrm{d}\mu_2\right\} \mathrm{d}\mu_1(\omega_1) \qquad (5.16) \\
&= \left\{\int_{\Omega_2} |f_2| \,\mathrm{d}\mu_2\right\} \int_{\Omega_1} |f_1| \,\mathrm{d}\mu_1 < +\infty. \qquad (5.17)
\end{aligned}$$

Justifications. Pour (5.16), on utilise le fait que $|f_1(\omega_1)|$ est une constante (dépendant de ω_1) pour l'intégration sur Ω_2. Remarquons d'ailleurs que cette constante est finie pour μ_1 presque tout ω_1 puisque f_1 est μ_1 intégrable. Pour (5.17), on a simplement sorti la constante finie $\int_{\Omega_2} |f_2| \,\mathrm{d}\mu_2$ de l'intégrale sur Ω_1. Enfin la finitude du produit des deux intégrales résulte des intégrabilités respectives de f_1 et f_2.

Une fois acquise l'intégrabilité de $f_1 \otimes f_2$, le théorème de Fubini légitime le même calcul que ci-dessus sans les valeurs absolues et donne (5.15). □

5.3 Intégration sur un produit fini d'espaces

5.3.1 Produits finis de tribus et de mesures

Nous abordons brièvement l'extension de l'étude précédente à un produit de n espaces mesurés $(\Omega_i, \mathcal{F}_i, \mu_i)$, où les mesures μ_i sont σ-finies. Notons $\mathcal{R}_n$ la classe des pavés

7. Notons que la valeur $f_1(\omega_1) = +\infty$ est exclue puisque $\mathbb{K} = \mathbb{R}$ ou $\mathbb{C}$.

mesurables

$$\mathcal{R}_n := \{A_1 \times A_2 \times \cdots \times A_n;\ A_i \in \mathcal{F}_i,\ 1 \leq i \leq n\}.$$

On définit la tribu produit $\mathcal{E}_n$ sur $\Omega_1 \times \cdots \times \Omega_n$ par

$$\mathcal{E}_n = \bigotimes_{i=1}^{n} \mathcal{F}_i := \sigma(\mathcal{R}_n). \tag{5.18}$$

La construction de la mesure produit $\mu_1 \otimes \cdots \otimes \mu_n$ se fait par récurrence en se ramenant au cas de deux espaces facteurs. On utilise pour cela le léger abus qui consiste à identifier les espaces [8]

$$(\Omega_1 \times \cdots \times \Omega_{n-1}) \times \Omega_n \quad \text{et} \quad \Omega_1 \times \cdots \times \Omega_n. \tag{5.19}$$

On pourrait se passer de cette identification en montrant que l'application

$$\big((\omega_1, \dots, \omega_{n-1}), \omega_n\big) \mapsto (\omega_1, \dots, \omega_n)$$

définit une bijection bimesurable entre ces deux espaces munis respectivement des tribus produits $\mathcal{E}_{n-1} \otimes \mathcal{F}_n$ et $\mathcal{E}_n$ ...

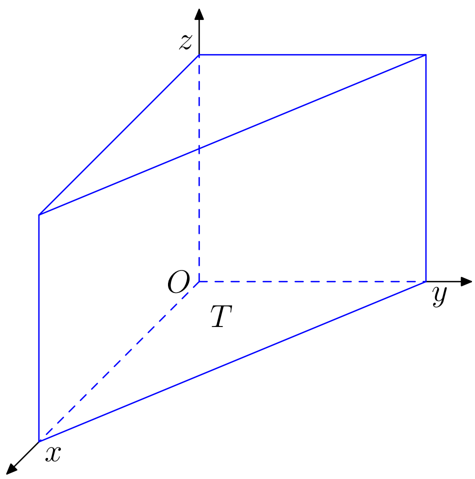


FIGURE 5.3 – Prisme $C = T \times [0, 1]$

En réalité, *le vrai problème n'est pas tant l'identification de ces deux espaces que celle de leurs tribus.* La tribu $\mathcal{E}_{n-1} \otimes \mathcal{F}_n$ est engendrée par la classe $\mathcal{R}'$ des rectangles mesurables de la forme $E \times A_n$ avec $E \in \mathcal{E}_{n-1}$ et $A_n \in \mathcal{F}_n$. Clairement $\mathcal{R}'$ contient $\mathcal{R}_n$ (via l'identification ci-dessus) mais est plus riche que $\mathcal{R}_n$. Par exemple avec $n = 3$ et $(\Omega_i, \mathcal{F}_i) = \big(\mathbb{R}, \mathrm{Bor}(\mathbb{R})\big)$, $(i = 1, 2, 3)$, le prisme de la figure 5.3

$$C := \{(x, y, z) \in \mathbb{R}^3;\ 0 \leq x + y \leq 1,\ x, y \geq 0,\ 0 \leq z \leq 1\}$$

8. Formellement, $\mathbb{R}^2 \times \mathbb{R}$ n'est pas exactement le même ensemble que $\mathbb{R}^3$. Un élément de $\mathbb{R}^2 \times \mathbb{R}$ est un *couple* $\big((x, y), z\big)$ dont la première composante est elle-même un couple de réels et la deuxième un réel, tandis qu'un élément de $\mathbb{R}^3$ est un *triplet* (x, y, z) de réels.

est dans $\mathcal{R}'$ comme produit cartésien du triangle $T = \{(x, y) \in \mathbb{R}_+^2;\ 0 \leq x + y \leq 1\}$ de $\mathbb{R}^2$ par le segment $[0, 1]$. Il n'est pas dans $\mathcal{R}_3$ car T n'est pas un produit cartésien. La proposition suivante permet de résoudre ce problème.

Proposition 5.14. *Avec l'identification des espaces (5.19), on a l'égalité de tribus*

$$\mathcal{E}_n = \mathcal{E}_{n-1} \otimes \mathcal{F}_n, \quad n \geq 2.$$

Plus généralement, notons pour $0 \leq i < j \leq n$,

$$\Omega_{i,j} := \Omega_{i+1} \times \cdots \times \Omega_j, \quad \mathcal{E}_{i,j} := \mathcal{F}_{i+1} \otimes \cdots \otimes \mathcal{F}_j = \sigma(\mathcal{R}_{i,j}),$$

où $\mathcal{R}_{i,j}$ est la classe des pavés mesurables $A_{i+1} \times \cdots \times A_j$ avec $A_k \in \mathcal{F}_k$, $i < k \leq j$. Alors en identifiant les espaces $\Omega_{0,n}$ et $\Omega_{0,j} \times \Omega_{j,n}$, on a

$$\mathcal{E}_n = \mathcal{E}_{0,n} = \mathcal{E}_{0,j} \otimes \mathcal{E}_{j,n}.$$

Preuve. L'égalité $\mathcal{E}_n = \mathcal{E}_{n-1} \otimes \mathcal{F}_n$ n'est que le cas $j = n - 1$ de $\mathcal{E}_n = \mathcal{E}_{0,j} \otimes \mathcal{E}_{j,n}$. Pour établir cette dernière égalité, on montre l'inclusion dans les deux sens.

L'inclusion $\mathcal{E}_n \subset \mathcal{E}_{0,j} \otimes \mathcal{E}_{j,n}$ est immédiate puisque si $R = A_1 \times \cdots \times A_n$ est dans $\mathcal{R}_n$, on peut l'identifier avec $(A_1 \times \cdots \times A_j) \times (A_{j+1} \times \cdots \times A_n)$. Il est donc membre de la classe $\mathcal{R}'$ des rectangles mesurables qui engendre $\mathcal{E}_{0,j} \otimes \mathcal{E}_{j,n}$. Ainsi $\mathcal{R}_n \subset \mathcal{R}'$ d'où $\mathcal{E}_n = \sigma(\mathcal{R}_n) \subset \sigma(\mathcal{R}') = \mathcal{E}_{0,j} \otimes \mathcal{E}_{j,n}$.

Pour obtenir l'inclusion dans l'autre sens, fixons d'abord un pavé quelconque $P = A_{j+1} \times \cdots \times A_n$ dans $\mathcal{R}_{j,n}$ et notons

$$\mathcal{G}_P := \{E \in \mathcal{E}_{0,j};\ E \times P \in \mathcal{E}_n\}.$$

Si $E \in \mathcal{R}_{0,j}$, $E \times P \in \mathcal{R}_n$ (via l'identification $\Omega_{0,n} = \Omega_{0,j} \times \Omega_{j,n}$). Donc $\mathcal{G}_P$ contient $\mathcal{R}_{0,j}$. On vérifie que $\mathcal{G}_P$ est une sous-tribu de $\mathcal{E}_{0,j}$ grâce au lemme élémentaire suivant qui est laissé en exercice.

Lemme 5.15. *Soient Ω et Ω' deux ensembles.*

i) Pour tout $F \subset \Omega'$, $\emptyset \times F = \emptyset$.

ii) Pour tout $E \subset \Omega$ et tout $F \subset \Omega'$, $\big(\Omega \setminus E\big) \times F = \big(\Omega \times F\big) \setminus (E \times F)$.

iii) Pour tout ensemble d'indices I, toute famille $(E_i)_{i \in I}$ de parties de Ω et tout $F \subset \Omega'$,

$$\Big(\bigcup_{i \in I} E_i\Big) \times F = \bigcup_{i \in I} (E_i \times F).$$

On a donc $\mathcal{E}_{0,j} \supset \mathcal{G}_P \supset \sigma(\mathcal{R}_{0,j}) = \mathcal{E}_{0,j}$. Ainsi $\mathcal{G}_P = \mathcal{E}_{0,j}$ et donc pour tout $E \in \mathcal{E}_{0,j}$, $E \times P$ est dans la tribu $\mathcal{E}_n$. Comme P a été pris quelconque dans $\mathcal{R}_{j,n}$, on a montré que

$$\forall E \in \mathcal{E}_{0,j},\ \forall P \in \mathcal{R}_{j,n}, \quad E \times P \in \mathcal{E}_n.$$

Fixons maintenant E quelconque dans $\mathcal{E}_{0,j}$ et définissons

$$\mathcal{H}_E := \{F \in \mathcal{E}_{j,n};\ E \times F \in \mathcal{E}_n\}.$$

D'après ce qui précède, $\mathcal{H}_E$ contient $\mathcal{R}_{j,n}$. On vérifie que c'est une sous-tribu de $\mathcal{E}_{j,n}$ en utilisant le lemme 5.15 (en permutant Ω et Ω'). Donc $\mathcal{E}_{j,n} \supset \mathcal{H}_E \supset \sigma(\mathcal{R}_{j,n}) = \mathcal{E}_{j,n}$ d'où $\mathcal{H}_E = \mathcal{E}_{j,n}$. Ainsi $E \times F \in \mathcal{E}_n$ pour tout $F \in \mathcal{E}_{j,n}$. Comme E est quelconque dans $\mathcal{E}_{0,j}$, on a finalement montré que

$$\forall E \in \mathcal{E}_{0,j},\ \forall F \in \mathcal{E}_{j,n}, \quad E \times F \in \mathcal{E}_n,$$

ce qui s'écrit aussi $\mathcal{R}' \subset \mathcal{E}_n$. On en déduit $\sigma(\mathcal{R}') \subset \mathcal{E}_n$, ce qui achève la preuve puisque $\sigma(\mathcal{R}') = \mathcal{E}_{0,j} \otimes \mathcal{E}_{j,n}$. □

Un exemple important de tribu produit à n facteurs est la tribu borélienne de $\mathbb{R}^n$. Cependant l'égalité

$$\operatorname{Bor}(\mathbb{R}^n) = \underbrace{\operatorname{Bor}(\mathbb{R}) \otimes \cdots \otimes \operatorname{Bor}(\mathbb{R})}_{n \text{ facteurs}} =: \Big(\operatorname{Bor}(\mathbb{R})\Big)^{\otimes n}, \tag{5.20}$$

n'est pas une simple conséquence de la proposition 5.14. En effet la tribu borélienne de $\mathbb{R}^n$ n'a pas été définie comme tribu *produit* des tribus boréliennes de $\mathbb{R}$, autrement dit la tribu engendrée par les $B_1 \times \cdots \times B_n$ où $B_i \in \operatorname{Bor}(\mathbb{R})$, mais comme la tribu engendrée par l'ensemble des ouverts de $\mathbb{R}^n$. La justification de (5.20) demande le petit effort supplémentaire suivant.

Proposition 5.16. *Soient p, $q \in \mathbb{N}^*$. En identifiant $\mathbb{R}^p \times \mathbb{R}^q$ et $\mathbb{R}^{p+q}$, on a*

$$\operatorname{Bor}(\mathbb{R}^p) \otimes \operatorname{Bor}(\mathbb{R}^q) = \operatorname{Bor}(\mathbb{R}^{p+q}).$$

Preuve. Notons $\mathcal{C}$ la classe des parallélépipèdes $Q = \prod_{i=1}^{p+q}]a_i, b_i]$. L'identification $\mathbb{R}^{p+q} = \mathbb{R}^p \times \mathbb{R}^q$ nous permet d'écrire

$$Q = \prod_{i=1}^{p+q}]a_i, b_i] = \left(\prod_{i=1}^{p}]a_i, b_i]\right) \times \left(\prod_{i=p+1}^{p+q}]a_i, b_i]\right),$$

faisant ainsi apparaître Q comme un élément de la classe $\mathcal{R}$ des rectangles mesurables de la tribu produit $\operatorname{Bor}(\mathbb{R}^p) \otimes \operatorname{Bor}(\mathbb{R}^q)$. On a donc $\mathcal{C} \subset \mathcal{R}$, d'où $\sigma(\mathcal{C}) \subset \sigma(\mathcal{R}) = \operatorname{Bor}(\mathbb{R}^p) \otimes \operatorname{Bor}(\mathbb{R}^q)$. D'autre part on sait que la tribu borélienne de $\mathbb{R}^{p+q}$ est engendrée par $\mathcal{C}$. On vient ainsi d'établir l'inclusion $\operatorname{Bor}(\mathbb{R}^{p+q}) \subset \operatorname{Bor}(\mathbb{R}^p) \otimes \operatorname{Bor}(\mathbb{R}^q)$.

Pour obtenir l'inclusion dans l'autre sens, il suffit de montrer que $\mathcal{R} \subset \operatorname{Bor}(\mathbb{R}^{p+q})$, autrement dit que pour tous $A \in \operatorname{Bor}(\mathbb{R}^p)$ et $B \in \operatorname{Bor}(\mathbb{R}^q)$, $A \times B$ est dans $\operatorname{Bor}(\mathbb{R}^{p+q})$. Considérons les projections canoniques

$$\begin{aligned} \pi_1 : \mathbb{R}^{p+q} &\to \mathbb{R}^p, \qquad (x_1, \dots, x_{p+q}) \longmapsto (x_1, \dots, x_p) \\ \pi_2 : \mathbb{R}^{p+q} &\to \mathbb{R}^q, \qquad (x_1, \dots, x_{p+q}) \longmapsto (x_{p+1}, \dots, x_{p+q}). \end{aligned}$$

Ce sont des applications continues, donc boréliennes. On en déduit l'appartenance à $\operatorname{Bor}(\mathbb{R}^{p+q})$ des ensembles $\pi_1^{-1}(A) = A \times \mathbb{R}^q$ et $\pi_2^{-1}(B) = \mathbb{R}^p \times B$. On conclut en notant que $A \times B = (A \times \mathbb{R}^q) \cap (\mathbb{R}^p \times B)$. □

Revenons à la construction de la mesure produit sur la tribu $\mathcal{E}_n$. On peut maintenant définir

$$\nu_n := \bigotimes_{i=1}^{n} \mu_i$$

par récurrence en posant

$$\nu_2 := \mu_1 \otimes \mu_2, \quad \nu_n := \nu_{n-1} \otimes \mu_n, \quad (n > 2).$$

Bien sûr, ν_n est l'unique mesure dont la restriction à $\mathcal{R}_n$ vérifie

$$\forall\, A_1 \times \cdots \times A_n \in \mathcal{R}_n, \quad \nu_n(A_1 \times \cdots \times A_n) = \mu_1(A_1)\mu_2(A_2)\ldots\mu_n(A_n). \tag{5.21}$$

Remarque 5.17. En notant λ_d la mesure de Lebesgue sur $\mathbb{R}^d$, on a

$$\lambda_d = \underbrace{\lambda_1 \otimes \cdots \otimes \lambda_1}_{d \text{ facteurs}} =: \lambda_1^{\otimes d}.$$

En effet par les propositions 5.14 et 5.16, on a $\mathrm{Bor}(\mathbb{R}^d) = \mathrm{Bor}(\mathbb{R})^{\otimes d}$ et grâce à (5.21) les mesures λ_d et $\lambda_1^{\otimes d}$ coïncident sur la π-classe des parallélépipèdes $]a_1, b_1] \times \cdots \times]a_d, b_d]$. Par le théorème d'unicité des mesures elles coïncident sur toute la tribu engendrée, donc sur la tribu borélienne de $\mathbb{R}^d$.

5.3.2 Intégrales multiples

Les théorèmes de Tonelli et de Fubini se généralisent comme suit.

Théorème 5.18. *On suppose que pour $i = 1, \ldots, n$, μ_i est une mesure σ-finie sur l'espace mesurable $(\Omega_i, \mathcal{F}_i)$ et que*

$$f : \Omega_{0,n} := \Omega_1 \times \cdots \times \Omega_n \longrightarrow \overline{\mathbb{R}}_+$$

est mesurable $\bigotimes_{i=1}^{n} \mathcal{F}_i$ - $\mathrm{Bor}(\overline{\mathbb{R}}_+)$. Alors pour toute permutation τ des indices $1, 2, \ldots, n$, on a

$$\begin{aligned} \int_{\Omega_{0,n}} f \,\mathrm{d}(\mu_1 \otimes \cdots \otimes \mu_n) &= \int_{\Omega_1}\int_{\Omega_2}\cdots\int_{\Omega_n} f(\omega_1, \ldots, \omega_n)\,\mathrm{d}\mu_n(\omega_n)\ldots\mathrm{d}\mu_2(\omega_2)\,\mathrm{d}\mu_1(\omega_1) \\ &= \int_{\Omega_{\tau(1)}}\cdots\int_{\Omega_{\tau(n)}} f(\omega_1, \ldots, \omega_n)\,\mathrm{d}\mu_{\tau(n)}(\omega_{\tau(n)})\ldots\mathrm{d}\mu_{\tau(1)}(\omega_{\tau(1)}). \end{aligned}$$

Toutes les intégrales itérées d'ordre k ($1 \leq k \leq n-1$) contenues dans ces formules définissent des fonctions mesurables positives relativement aux tribus produits concernées.

Théorème 5.19. *Avec les notations du théorème précédent, si $f : \Omega_{0,n} \to \mathbb{K} = \mathbb{R}$ ou $\mathbb{C}$ est $\mathcal{E}_n$ - $\mathrm{Bor}(\mathbb{K})$ mesurable, elle est $\mu_1 \otimes \cdots \otimes \mu_n$ intégrable si et seulement si l'une des $n!$ intégrales itérées d'ordre n*

$$\int_{\Omega_{\tau(1)}}\cdots\int_{\Omega_{\tau(n)}} |f(\omega_1, \ldots, \omega_n)|\,\mathrm{d}\mu_{\tau(n)}(\omega_{\tau(n)})\ldots\mathrm{d}\mu_{\tau(1)}(\omega_{\tau(1)})$$

est finie. Dans ce cas les $n!$ égalités d'intégrales du théorème précédent restent vraies et toutes les intégrales itérées d'ordre $k < n$ apparaissant dans ces formules définissent presque partout des fonctions intégrables (relativement aux mesures concernées).

Corollaire 5.20. *Si pour* $i = 1, \dots n$, $f_i : \Omega_i \to \mathbb{K}$ *est* μ_i *intégrable, la fonction*

$$f := f_1 \otimes \cdots \otimes f_n : \Omega_{0,n} \longrightarrow \mathbb{K} = \mathbb{R}, \quad (\omega_1, \dots, \omega_n) \longmapsto f_1(\omega_1) \dots f_n(\omega_n)$$

est $\mu_1 \otimes \cdots \otimes \mu_n$ *intégrable et*

$$\int_{\Omega_{0,n}} f \,\mathrm{d}(\mu_1 \otimes \cdots \otimes \mu_n) = \prod_{i=1}^{n} \int_{\Omega_i} f_i \,\mathrm{d}\mu_i.$$

5.3.3 Application aux lois marginales d'un vecteur aléatoire

Pour terminer cette section, nous examinons une application aux vecteurs aléatoires. Rappelons que si $(\Omega, \mathcal{F}, \mathbf{P})$ est un espace probabilisé, un vecteur aléatoire sur cet espace est une application $X : \Omega \to \mathbb{R}^d$, mesurable $\mathcal{F}$ - $\mathrm{Bor}(\mathbb{R}^d)$. Sa loi P_X est la mesure image $\mathbf{P} \circ X^{-1}$ qui est une probabilité sur $\mathrm{Bor}(\mathbb{R}^d)$. Pour $i = 1, \dots, d$, soit π_i la i-ème projection canonique de $\mathbb{R}^d$ sur $\mathbb{R}$. Notons $X_i = \pi_i \circ X$. Alors X_i est mesurable $\mathcal{F}$ - $\mathrm{Bor}(\mathbb{R})$, c'est donc une variable aléatoire réelle. Pour tout $\omega \in \Omega$, on a

$$X(\omega) = \bigl(X_1(\omega), X_2(\omega), \dots, X_d(\omega)\bigr).$$

La loi P_{X_i} de la variable aléatoire X_i s'appelle i-ème *loi marginale* du vecteur aléatoire X.
Rappelons encore que P_X a une densité f par rapport à λ_d, mesure de Lebesgue sur $\mathbb{R}^d$, si

$$\forall B \in \mathrm{Bor}(\mathbb{R}^d), \quad \mathbf{P}(X \in B) = P_X(B) = \int_B f(x_1, \dots, x_d) \,\mathrm{d}\lambda_d(x_1, \dots, x_d). \tag{5.22}$$

Dans ce cas, f est une fonction mesurable positive et $\int_{\mathbb{R}^d} f \,\mathrm{d}\lambda_d = 1$.

Proposition 5.21. *Soit* $X = (X_1, \dots, X_d)$ *un vecteur aléatoire à valeurs dans* $\mathbb{R}^d$, *dont la loi a une densité* f *par rapport à* λ_d. *Alors sa* i*-ème loi marginale* P_{X_i} *admet une densité* f_{X_i} *par rapport à* λ_1, *donnée par*

$$f_{X_i}(t) = \int_{\mathbb{R}^{d-1}} f(x_1, \dots, x_{i-1}, t, x_{i+1}, \dots, x_d) \,\mathrm{d}\lambda_{d-1}(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_d).$$

Preuve. Pour tout borélien A de $\mathbb{R}$,

$$P_{X_i}(A) = \mathbf{P}(X_i \in A) = \mathbf{P}(\pi_i \circ X \in A) = \mathbf{P}(X \in \mathbb{R}^{i-1} \times A \times \mathbb{R}^{d-i}).$$

En utilisant (5.22), la remarque 5.17 et le théorème 5.18, on en déduit

$$\begin{aligned}
P_{X_i}(A) &= \int_{\mathbb{R}^{i-1} \times A \times \mathbb{R}^{d-i}} f(x_1, \dots, x_d) \,\mathrm{d}\lambda_1^{\otimes d}(x_1, \dots, x_d) \\
&= \int_A \int_{\mathbb{R}^{i-1} \times \mathbb{R}^{d-i}} f(x_1, \dots, x_d) \,\mathrm{d}\lambda_1^{\otimes (d-1)}(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_d) \,\mathrm{d}\lambda_1(x_i) \\
&= \int_A f_{X_i}(x_i) \,\mathrm{d}\lambda_1(x_i)
\end{aligned}$$

On en déduit par (5.22) appliqué à P_{X_i}, que cette loi a pour densité f_{X_i}. □

5.4 Changement de variable dans $\mathbb{R}^d$

5.4.1 Introduction

Cette section ne relève qu'indirectement de l'intégration sur un espace produit. Les résultats qu'elle contient constituent, avec le théorème de Fubini, l'un des deux outils pour le calcul pratique d'intégrales $\int_{\mathbb{R}^d} f \,\mathrm{d}\lambda$, où λ est la mesure de Lebesgue sur $\mathbb{R}^d$. Il s'agit de donner une forme explicite au théorème de transfert pour effectuer pratiquement un changement de variable dans ce type d'intégrales. Rappelons le cadre général du théorème de transfert. On considère le diagramme suivant

$$(\Omega_1, \mathcal{F}_1, \mu) \xrightarrow{\varphi} (\Omega_2, \mathcal{F}_2, \nu := \mu \circ \varphi^{-1}) \xrightarrow{h} (\mathbb{K}, \mathrm{Bor}(\mathbb{K})),$$

où φ est mesurable $\mathcal{F}_1$ - $\mathcal{F}_2$ et h est mesurable $\mathcal{F}_2$ - $\mathrm{Bor}(\mathbb{K})$. La formule de transfert s'écrit alors

$$\int_{\Omega_1} (h \circ \varphi) \,\mathrm{d}\mu = \int_{\Omega_2} h \,\mathrm{d}\nu. \tag{5.23}$$

Cette formule est valable pour toute h mesurable positive (cas $\mathbb{K} = \overline{\mathbb{R}}_+$) et si $\mathbb{K} = \mathbb{R}$ ou $\mathbb{C}$, pour toute $h \in \mathcal{L}^1_{\mathbb{K}}(\nu)$ ou ce qui est équivalent, telle que $h \circ \varphi \in \mathcal{L}^1_{\mathbb{K}}(\mu)$.

Dans tout ce qui suit, on prendra $\Omega_1 = \mathbb{R}^d$ ou V ouvert de $\mathbb{R}^d$ et $\Omega_2 = \varphi(\Omega_1)$. La mesure μ sera la mesure de Lebesgue λ sur $\mathbb{R}^d$ ou sa restriction à V. Nous supposerons de plus φ *bijective*. Notons provisoirement ψ son inverse ponctuel : pour $y \in \Omega_2$, $\psi(y)$ est l'unique $x \in \Omega_1$ tel que $\varphi(x) = y$. Alors l'inverse ensembliste $\varphi^{-1}(B)$ coïncide avec l'image directe de B par l'application ψ :

$$\begin{aligned}
\varphi^{-1}(B) &= \{x \in \Omega_1;\ \varphi(x) \in B\} \\
&= \{x \in \Omega_1;\ \exists y \in B, y = \varphi(x)\} \\
&= \{x \in \Omega_1;\ \exists y \in B, x = \psi(y)\} \\
&= \{\psi(y);\ y \in B\} \\
&= \psi(B).
\end{aligned}$$

Cette remarque étant faite, nous revenons à la notation traditionnelle φ^{-1} pour l'inverse ponctuel ψ de φ. Posons $f := h \circ \varphi$ d'où $h = f \circ \varphi^{-1}$. La formule (5.23) peut alors s'écrire sous l'une des deux formes équivalentes :

$$\int_{\Omega_1} f(x) \,\mathrm{d}\lambda(x) = \int_{\varphi(\Omega_1)} f\big(\varphi^{-1}(y)\big) \,\mathrm{d}(\lambda \circ \varphi^{-1})(y) \tag{5.24}$$

$$\int_{\Omega_1} h\big(\varphi(x)\big) \,\mathrm{d}\lambda(x) = \int_{\varphi(\Omega_1)} h(y) \,\mathrm{d}(\lambda \circ \varphi^{-1})(y). \tag{5.25}$$

Dans (5.24), φ^{-1} intervient comme inverse ponctuel et comme inverse ensembliste, on pourrait écrire le second membre $\int_{\varphi(\Omega_1)} f\big(\psi(y)\big) \,\mathrm{d}(\lambda \circ \varphi^{-1})(y)$. Pour donner à ces formules un caractère effectif, il reste à identifier la mesure $\lambda \circ \varphi^{-1}$. Les outils à notre disposition dans le cadre du programme de la Licence nous permettent de traiter complètement le cas où φ est une bijection linéaire et nous conduiront à admettre le résultat dans le cas plus général où φ est un C^1 difféomorphisme[9].

5.4.2 Changement de variable linéaire

Soit donc φ une bijection linéaire $\mathbb{R}^d \to \mathbb{R}^d$. Elle est évidemment borélienne puisque toutes les applications linéaires entre espaces de dimension finie sont continues. Son inverse ponctuel φ^{-1} est aussi linéaire. Pour tout vecteur $u \in \mathbb{R}^d$, nous notons T_u la translation de vecteur u. C'est une bijection d'inverse $T_u^{-1} = T_{-u}$. Nous posons $\nu := \lambda \circ \varphi^{-1}$. Nous commençons par vérifier que ν est invariante par translations. Pour cela en appliquant la remarque ci-dessus sur la coïncidence entre inverse ensembliste et image directe par inverse ponctuel pour une bijection, on peut écrire pour tout borélien B de $\mathbb{R}^d$

$$(\nu \circ T_u^{-1})(B) = \nu\Big(T_{-u}(B)\Big) = \lambda\Big((\varphi^{-1} \circ T_{-u})(B)\Big).$$

Grâce à la linéarité de φ^{-1}, on a pour tout $y \in \mathbb{R}^d$,

$$(\varphi^{-1} \circ T_{-u})(y) = \varphi^{-1}(y - u) = \varphi^{-1}(y) - \varphi^{-1}(u).$$

Posant $v := \varphi^{-1}(u)$, on voit ainsi que $\varphi^{-1} \circ T_{-u} = T_{-v} \circ \varphi^{-1} = T_v^{-1} \circ \varphi^{-1}$. On en déduit

$$\lambda\Big((\varphi^{-1} \circ T_{-u})(B)\Big) = \lambda\Big(T_v^{-1}(\varphi^{-1}(B))\Big) = (\lambda \circ T_v^{-1})(\varphi^{-1}(B)).$$

Comme λ est invariante par translations, $\lambda \circ T_v^{-1} = \lambda$ d'où

$$(\lambda \circ T_v^{-1})(\varphi^{-1}(B)) = \lambda(\varphi^{-1}(B)) = \nu(B).$$

On a donc vérifié que $(\nu \circ T_u^{-1})(B) = \nu(B)$ pour tout borélien B et toute translation T_u. La mesure ν est invariante par translations.

Notons $C := [0,1]^d$ le cube unité. On a $0 < \nu(C) < +\infty$. En effet, $\varphi^{-1}(C)$ est borné comme image d'un borné par l'application linéaire continue φ^{-1}. Sa mesure de Lebesgue est donc finie d'où $\nu(C) < +\infty$. D'autre part $\varphi^{-1}(C)$ contient $\varphi^{-1}(]0,1[^d)$ qui est *ouvert* comme image réciproque de l'ouvert $]0,1[^d$ par l'application continue φ. Cet ouvert contient au moins un cube de la forme $]x_0 - \varepsilon, x_0 + \varepsilon[^d$ de mesure de Lebesgue $(2\varepsilon)^d > 0$ donc $\nu(C) > 0$.

On vérifie facilement (exercice déjà fait pour $d = 1$) qu'une mesure μ invariante par translation et telle que $0 < \mu(C) < +\infty$ est proportionnelle à la mesure de Lebesgue : $\mu = K\lambda$ pour une constante $0 < K < +\infty$. Nous venons donc de prouver le résultat suivant.

Lemme 5.22. *Si $\varphi : \mathbb{R}^d \to \mathbb{R}^d$ est une bijection linéaire, il existe $K(\varphi) \in]0, +\infty[$ constante telle que*

$$\lambda \circ \varphi^{-1} = K(\varphi)\lambda. \tag{5.26}$$

9. Le lecteur curieux et courageux pourra consulter W. Rudin *Analyse réelle et complexe* ou P. Billingsley *Probability and measure*...

Pour déterminer explicitement la constante $K(\varphi)$, il est utile d'établir les propriétés suivantes.

Lemme 5.23. *La constante* $K(\varphi)$ *définie par* (5.26) *vérifie*

a) $K(\varphi) = \lambda(\varphi^{-1}(C))$ *où* $C = [0,1]^d$.

b) Si φ *et* ψ *sont deux bijections linéaires* $\mathbb{R}^d \to \mathbb{R}^d$, $K(\varphi \circ \psi) = K(\varphi)K(\psi)$.

c) $K(\varphi) = 1/\lambda(\varphi(C))$.

Preuve. Pour a), on note que $\lambda(C) = 1$ et on calcule $\nu(C)$ par (5.26) :

$$\lambda(\varphi^{-1}(C)) = K(\varphi)\lambda(C) = K(\varphi).$$

Le b) s'obtient en appliquant successivement a) avec $\varphi \circ \psi$, (5.26) avec ψ et a) avec φ :

$$K(\varphi\circ\psi) = \big(\lambda\circ(\psi^{-1}\circ\varphi^{-1})\big)(C) = (\lambda\circ\psi^{-1})\big(\varphi^{-1}(C)\big) = K(\psi)\lambda\big(\varphi^{-1}(C)\big) = K(\psi)K(\varphi).$$

En choisissant $\psi = \varphi^{-1}$ dans b), on obtient c). □

Arrivés à ce stade, nous avons réduit le problème d'identification de la mesure $\lambda\circ\varphi^{-1}$ au calcul du volume de $\varphi(C)$ (ou de $\varphi^{-1}(C)$). Essayons de nous en faire une idée plus précise en examinant le cas de la dimension $d = 2$. Notons (e_1, e_2) la base canonique de $\mathbb{R}^2$. Le carré unité C est

$$C = \{x = x_1e_1 + x_2e_2 \in \mathbb{R}^2;\ 0 \le x_1 \le 1,\ 0 \le x_2 \le 1\}.$$

Son image par l'application linéaire φ est donc

$$\varphi(C) = \{y = x_1\varphi(e_1) + x_2\varphi(e_2) \in \mathbb{R}^2;\ 0 \le x_1 \le 1,\ 0 \le x_2 \le 1\}.$$

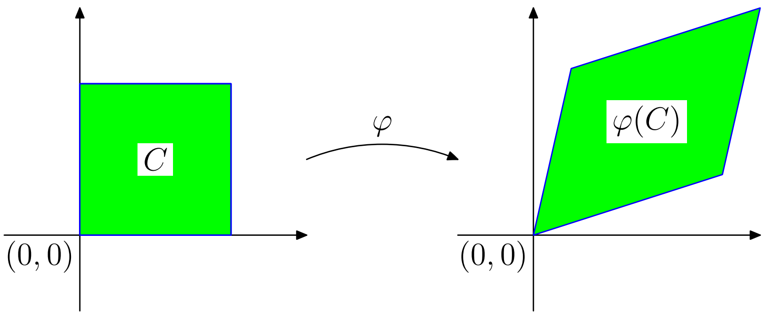


FIGURE 5.4 – Image du carré unité

Il s'agit du *parallélogramme* de sommets $\big(0, \varphi(e_1), \varphi(e_1)+\varphi(e_2), \varphi(e_2)\big)$. Notons $\varphi(e_1) = ae_1 + be_2$ et $\varphi(e_2) = ce_1 + de_2$. On vérifie facilement (exercice) que l'aire de ce parallélogramme est

$$\lambda\big(\varphi(C)\big) = |ad - bc|.$$

On remarque que c'est la valeur absolue du déterminant de la matrice $\begin{pmatrix} a & c \\ b & d \end{pmatrix}$ de φ dans la base (e_1, e_2). Cette propriété est générale.

Lemme 5.24. *Si φ est une bijection linéaire $\mathbb{R}^d \to \mathbb{R}^d$, la constante $K(\varphi)$ définie par (5.26) vaut*

$$K(\varphi) = \frac{1}{|\det \varphi|} = |\det(\varphi^{-1})|. \tag{5.27}$$

Preuve. Comme $\det(\varphi \circ \psi) = \det(\varphi)\det(\psi)$ et $K(\varphi \circ \psi) = K(\varphi)K(\psi)$, le lemme sera prouvé par décomposition de φ en un produit de bijections linéaires simples pour lesquelles on vérifie (5.27). Ces applications sont des trois types suivants décrits en donnant l'image de la base canonique $(e_1, \dots, e_d)$ de $\mathbb{R}^d$:

(I) $\big(\varphi(e_1), \dots, \varphi(e_d)\big)$ est une permutation de $(e_1, \dots, e_d)$;

(II) $\varphi(e_1) = ae_1$ $(a \neq 0)$ et $\varphi(e_i) = e_i$, $(\forall i \geq 2)$;

(III) $\varphi(e_1) = e_1 + e_2$ et $\varphi(e_i) = e_i$, $(\forall i \geq 2)$.

La décomposition de toute bijection linéaire $\mathbb{R}^d \to \mathbb{R}^d$ en un produit d'applications de ce type est un résultat purement algébrique que nous admettrons.

Si φ est du type (I), son déterminant vaut $+1$ ou -1 et comme la mesure de Lebesgue est invariante par permutation des coordonnées, $\lambda\big(\varphi(C)\big) = \lambda(C) = 1$. Donc $K(\varphi) = 1$ et (5.27) est vérifiée.

Si φ est du type (II), son déterminant vaut a. Si $a > 0$, $\varphi(C) = [0, a] \times [0, 1]^{d-1}$ et $\lambda\big(\varphi(C)\big) = (a - 0) \times 1^{d-1} = a = |a|$. Si $a < 0$, $\varphi(C) = [a, 0] \times [0, 1]^{d-1}$ et $\lambda\big(\varphi(C)\big) = (0 - a) \times 1^{d-1} = -a = |a|$. Dans les deux cas $\lambda\big(\varphi(C)\big) = |a| = |\det \varphi|$, et (5.27) est vérifiée via le lemme 5.23 c).

Si φ est du type (III), son déterminant vaut $+1$ (évident en développant suivant la première colonne). D'autre part, $\varphi(C) = A \times [0, 1]^{d-2}$, où A est le parallélogramme de $\mathbb{R}^2$ de sommets $(0, e_1 + e_2, e_1 + 2e_2, e_2)$. Comme $\lambda_d = \lambda_2 \otimes \lambda_{d-2}$, on en déduit que $\lambda_d\big(\varphi(C)\big) = \lambda_2(A)\lambda_{d-2}([0, 1]^{d-2}) = \lambda_2(A)$. L'aire de A étant clairement la même que celle du carré unité de $\mathbb{R}^2$, on a $\lambda_d\big(\varphi(C)\big) = 1 = \det \varphi$ et (5.27) est vérifiée via le lemme 5.23 c). □

En revenant aux formules (5.24) et (5.25), nous pouvons maintenant énoncer :

Proposition 5.25. *Si φ est une bijection linéaire $\mathbb{R}^d \to \mathbb{R}^d$,*

$$\int_{\mathbb{R}^d} f(x)\,\mathrm{d}\lambda(x) = \int_{\mathbb{R}^d} f\big(\varphi^{-1}(y)\big)|\det(\varphi^{-1})|\,\mathrm{d}\lambda(y) \tag{5.28}$$

$$\int_{\mathbb{R}^d} h\big(\varphi(x)\big)\,\mathrm{d}\lambda(x) = \int_{\mathbb{R}^d} h(y)|\det(\varphi^{-1})|\,\mathrm{d}\lambda(y). \tag{5.29}$$

Ces égalités sont valables pour toutes fonctions mesurables positives f, h. L'égalité (5.28) est vraie pour toute f à valeurs réelles ou complexes, λ-intégrable sur $\mathbb{R}^d$ (ce qui équivaut à la λ-intégrabilité de $f \circ \varphi^{-1}$). De même (5.29) vaut pour toute $h \in \mathcal{L}^1_{\mathbb{K}}(\mathbb{R}^d, \lambda)$ ($\Leftrightarrow h \circ \varphi \in \mathcal{L}^1_{\mathbb{K}}(\mathbb{R}^d, \lambda)$).

Remarque 5.26. La portée des formules (5.28) et (5.29) est plus grande qu'il n'y paraît. Elles permettent de faire un changement de variable linéaire bijectif dans une intégrale sur *n'importe quel borélien* B de $\mathbb{R}^d$. Il suffit pour cela de remplacer f par $f\mathbf{1}_B$ et de remarquer que $\mathbf{1}_B\big(\varphi^{-1}(y)\big) = \mathbf{1}_{\varphi(B)}(y)$. On obtient ainsi

$$\int_B f(x)\,\mathrm{d}\lambda(x) = \int_{\varphi(B)} f\Big(\varphi^{-1}(y)\Big)|\det(\varphi^{-1})|\,\mathrm{d}\lambda(y). \tag{5.30}$$

Corollaire 5.27. *La mesure de Lebesgue de $\mathbb{R}^d$ est invariante par les isométries euclidiennes de $\mathbb{R}^d$.*

Preuve. Soit φ une isométrie euclidienne de $\mathbb{R}^d$. Quitte à la composer avec une translation (qui laisse λ_d invariante), on se ramène au cas où $\varphi(0) = 0$. Alors φ est une application linéaire $\mathbb{R}^d \to \mathbb{R}^d$ qui conserve le produit scalaire. Il est bien connu que pour une telle application $|\det(\varphi)| = 1$, d'où la conclusion grâce à (5.26) et (5.27). □

5.4.3 Le théorème de changement de variable C^1

Que peut-on dire lorsqu'on a un changement de variable non linéaire ? Rappelons que dans le cas $d = 1$, nous avons déjà établi au chapitre 4 que si I est un intervalle ouvert (pas nécessairement borné) de $\mathbb{R}$, si $\varphi : I \to \mathbb{R}$ est strictement monotone, de classe C^1 et telle que φ' ne s'annule en aucun point point de I, on a [10]

$$\int_I f(x)\,\mathrm{d}\lambda(x) = \int_{\varphi(I)} f\Big(\varphi^{-1}(y)\Big)|(\varphi^{-1})'(y)|\,\mathrm{d}\lambda(y).$$

Cette formule se généralise à la dimension d en remplaçant intervalle ouvert par ouvert et $(\varphi^{-1})'(y)$ par le *déterminant jacobien* de φ^{-1}, *i.e.* le déterminant de l'application linéaire tangente à φ^{-1} au point y. Nous admettrons le résultat dont l'énoncé précis est le suivant.

Théorème 5.28. *Soit V un ouvert de $\mathbb{R}^d$ et φ un C^1-difféomorphisme de $V \to \varphi(V) \subset \mathbb{R}^d$. Alors les formules de changement de variable équivalentes*

$$\begin{aligned}\int_V f(x)\,\mathrm{d}\lambda(x) &= \int_{\varphi(V)} f\Big(\varphi^{-1}(y)\Big)|\operatorname{Jac}(\varphi^{-1})(y)|\,\mathrm{d}\lambda(y),\\ \int_V h\Big(\varphi(x)\Big)\,\mathrm{d}\lambda(x) &= \int_{\varphi(V)} h(y)|\operatorname{Jac}(\varphi^{-1})(y)|\,\mathrm{d}\lambda(y),\end{aligned}$$

sont vérifiées pour toutes f, h mesurables positives ou telles que $f \in \mathcal{L}^1_{\mathbb{K}}(\lambda)$, $h\circ\varphi \in \mathcal{L}^1_{\mathbb{K}}(\lambda)$.

En notant $\psi_1, \dots, \psi_d$ les applications coordonnées de φ^{-1}, on a

$$y = (y_1, \dots, y_d) = \varphi(x) \quad \Leftrightarrow \quad x = (x_1, \dots, x_d) = \Big(\psi_1(y), \dots, \psi_d(y)\Big)$$

et

$$\operatorname{Jac}(\varphi^{-1})(y) := \det\left[\frac{\partial\psi_i(y)}{\partial y_j}\right]_{1\le i,j\le d}.$$

10. Appliquer (4.40) avec f à la place de $f\circ\varphi$.

Voici une notation mnémotechnique basée sur l'analogie formelle avec le changement de variable en dimension 1, que l'on peut utiliser *au brouillon* :

$$\text{« } \mathrm{d}x_1 \ldots \mathrm{d}x_d = \frac{\mathrm{D}(x_1, \ldots, x_d)}{\mathrm{D}(y_1, \ldots, y_d)} \mathrm{d}y_1 \ldots \mathrm{d}y_d \text{ »} \quad \text{pour} \quad \mathrm{d}\lambda(x) = |\operatorname{Jac}(\varphi^{-1})(y)| \, \mathrm{d}\lambda(y).$$

Avant d'aborder la mise en pratique du théorème 5.28, il n'est sans doute pas superflu de faire le point sur la notion de C^1-difféomorphisme.

5.4.4 Rappels sur les C^1-difféomorphismes

Définition 5.29. *Soient E et F deux ouverts de $\mathbb{R}^d$. L'application $\varphi : E \to F$ est appelée* C^1-difféomorphisme *de E* sur *F si elle vérifie les deux conditions :*

i) φ est une bijection de E sur F.

ii) φ et φ^{-1} sont C^1, c'est-à-dire possèdent des dérivées partielles continues, en tout point de E pour φ et en tout point de F pour φ^{-1}.

Théorème 5.30 (d'inversion locale)**.** *Soit $\varphi : E \to F$ une application C^1 sur E. Soit $a \in E$ tel que* $\operatorname{Jac}(\varphi)(a) \neq 0$*. Notons $b := \varphi(a) \in F$. Alors il existe un ouvert A de E contenant a et un ouvert B de F contenant b tels que la restriction φ_A de φ à A soit une bijection de A sur B, que φ_A soit C^1 sur A et que φ_A^{-1} soit C^1 sur B (autrement dit que φ_A soit un C^1-difféomorphisme de A sur B).*

Examinons l'exemple suivant. On prend $E = F = \mathbb{R}^2$ et on définit φ par $\varphi(x, y) = (x^2 - y^2, 2xy)$. On voit immédiatement que

$$\operatorname{Jac}(\varphi)(x_0, y_0) = \begin{vmatrix} 2x_0 & -2y_0 \\ 2y_0 & 2x_0 \end{vmatrix} = 4(x_0^2 + y_0^2).$$

Si $(x_0, y_0) \neq (0, 0)$, on peut donc inverser localement φ au voisinage de (x_0, y_0). Par ailleurs, on vérifie directement que

$$\varphi(x, y) = \varphi(x', y') \quad \Leftrightarrow \quad (x, y) = (x', y') \text{ ou } (x, y) = (-x', -y').$$

Donc $\varphi : \mathbb{R}^2 \to \mathbb{R}^2$ n'est pas injective et il est impossible de l'inverser globalement. En particulier il est impossible de l'inverser sur un domaine contenant les deux points distincts (x_0, y_0) et $(-x_0, -y_0)$. En fait on peut inverser φ sur tout demi-plan dont la frontière passe par le point $(0, 0)$. Ce problème est celui de la détermination de la racine carrée complexe puisque si $z = x + iy$, $\varphi(x, y) = \big(\operatorname{Re}(z^2), \operatorname{Im}(z^2)\big)$.

Théorème 5.31 (d'inversion globale)**.** *Si $\varphi : E \to F$ est C^1 sur E et si* pour tout *$a \in E$,* $\operatorname{Jac}(\varphi)(a) \neq 0$*, alors l'image par φ de tout ouvert de E est un ouvert de $\mathbb{R}^d$ (on dit que φ est une application ouverte).*

Si de plus *φ est* injective*, alors φ est un C^1-difféomorphisme de E sur $\varphi(E)$.*

À titre d'exercice, on pourra confronter ce théorème à l'exemple précédent dans les deux cas $E := \mathbb{R}^2 \setminus \{(0,0)\}$ et $E := \{(x,y) \in \mathbb{R}^2;\ y > 0\}$.
Conséquence pratique. Pour vérifier qu'un changement de variable φ est un C^1-difféomorphisme, il *suffit* de vérifier que φ possède des dérivées partielles continues en tout point de l'ouvert de départ, que $\operatorname{Jac}(\varphi)(a) \neq 0$ en tout point a de cet ouvert et que φ est une bijection (en précisant soigneusement les ensembles de départ et d'arrivée !).

5.4.5 Méthode pratique

Pour alléger les écritures, nous nous plaçons dans le cas $d = 2$, mais les conseils donnés ci-dessous sont valables avec d quelconque, modulo le changement de notations. Soit

$$\varphi : (x,y) \longmapsto \varphi(x,y) = (s,t) \tag{5.31}$$

le changement de variable à effectuer dans l'intégrale

$$I := \int_V f(x,y)\, \mathrm{d}\lambda_2(x,y),$$

où V est un ouvert de $\mathbb{R}^2$.

1. On regarde si on peut inverser la relation $(s,t) = \varphi(x,y)$. Cela amène en général à introduire des conditions supplémentaires dans (5.31), en précisant les ensembles de départ E et d'arrivée F, de façon à considérer φ comme une bijection de E sur F. Si $V \subset E$, on passe au point suivant.
   Sinon, on découpe V en une réunion finie (éventuellement infinie dénombrable) d'ouverts V_i *disjoints* (plus éventuellement des ensembles de mesure nulle) tels que la restriction de φ à chaque V_i soit injective. On a alors
   $$I = \sum_i \int_{V_i} f(x,y)\, \mathrm{d}\lambda_2(x,y).$$
2. On détermine ensuite le nouvel ensemble d'intégration $\varphi(V)$ (resp. $\varphi(V_i)$) en écrivant :
   $$(s,t) \in \varphi(V) \quad \Leftrightarrow \quad (x,y) = \varphi^{-1}(s,t) \in V,$$
   ce qui en pratique revient à remplacer dans les inéquations définissant V (resp. V_i) les anciennes variables x et y par leurs expressions en fonction des nouvelles variables s et t .
3. À partir de la relation $(x,y) = \varphi^{-1}(s,t)$, on calcule les dérivées partielles $\frac{\partial x}{\partial s}$, $\frac{\partial x}{\partial t}$, $\frac{\partial y}{\partial s}$ et $\frac{\partial y}{\partial t}$ de φ^{-1} et on vérifie qu'elles sont *continues* sur $\varphi(V)$ (resp. sur $\varphi(V_i)$), donc que φ^{-1} est C^1. On calcule ensuite le jacobien de φ^{-1} et on vérifie qu'il ne s'annule en aucun point de $\varphi(V)$ (resp. $\varphi(V_i)$) [11].

11. Une variante consiste à calculer les dérivées partielles $\frac{\partial s}{\partial x}, \frac{\partial t}{\partial x} \dots$ de φ, vérifier leur continuité sur V, calculer $\operatorname{Jac}(\varphi)(x,y)$ et vérifier qu'il ne s'annule en aucun point de V. On établit ainsi que φ est un C^1 difféomorphisme de V sur $\varphi(V)$ et il ne reste plus qu'à calculer $\operatorname{Jac}(\varphi^{-1})(s,t)$ par la relation $\operatorname{Jac}(\varphi^{-1})(s,t) = 1/\operatorname{Jac}(\varphi)(\varphi^{-1}(s,t))$. Le choix entre ces deux méthodes sera guidé par la commodité du calcul des dérivées partielles. En tout état de cause, la nécessité d'inverser proprement φ reste incontournable.

Une fois tout ce travail effectué, on peut appliquer le théorème 5.28 pour conclure :

$$\int_V f(x,y)\,\mathrm{d}\lambda_2(x,y) = \int_{\varphi(V)} f\Big(\varphi^{-1}(s,t)\Big)|\operatorname{Jac}(\varphi^{-1})(s,t)|\,\mathrm{d}\lambda_2(s,t),$$

(resp. $\int_{V_i} = \int_{\varphi(V_i)}\dots$).

5.4.6 Coordonnées polaires, coordonnées sphériques

Parmi les applications les plus connues du théorème 5.28 figurent le changement de variable par passage des coordonnées cartésiennes aux coordonnées polaires ($d=2$) ou aux coordonnées sphériques ($d=3$). Voyons cela plus en détail.

Coordonnées polaires

Le passage des coordonnées cartésiennes aux coordonnées polaires est donné par les formules

$$x = r\cos\theta, \quad y = r\sin\theta. \tag{5.32}$$

Il est commode ici de travailler avec l'application $\psi = \varphi^{-1} : (r,\theta) \mapsto (x,y)$. Cette application réalise une bijection de $]0,+\infty[\times[\alpha,\alpha+2\pi[$ sur $\mathbb{R}^2\setminus\{(0,0)\}$, où α est un réel quelconque fixé (en pratique on prend le plus souvent $\alpha=-\pi$ ou $\alpha=0$). Notons ici l'exclusion de tous les couples $(0,\theta)$ de l'ensemble de départ et de leur image commune $(0,0)$ dans l'ensemble d'arrivée : leur conservation eût empêché la bijectivité de ψ (du moins avec un ensemble de départ produit cartésien). Nous réduisons l'ensemble de définition de ψ à $W_\alpha :=]0,+\infty[\times]\alpha,\alpha+2\pi[$ de façon à avoir un *ouvert*. Ceci nous amène à prendre comme ensemble d'arrivée $V_\alpha := \mathbb{R}^2\setminus D_\alpha$, où D_α est la demi-droite fermée d'origine $(0,0)$ et de vecteur directeur $(\cos\alpha,\sin\alpha)$. Comme $\lambda_2(D_\alpha)=0$, on pourra toujours écrire $\int_{\mathbb{R}^2} f\,\mathrm{d}\lambda_2 = \int_{V_\alpha} f\,\mathrm{d}\lambda_2$ pour f λ_2 intégrable. Finalement, ψ est définie par :

$$\psi : W_\alpha =]0,+\infty[\times]\alpha,\alpha+2\pi[\longrightarrow V_\alpha = \mathbb{R}^2\setminus D_\alpha, \quad (r,\theta)\longmapsto(x,y)=(r\cos\theta,r\sin\theta).$$

Le calcul des dérivées partielles de ψ est immédiat :

$$\frac{\partial x}{\partial r} = \cos\theta, \quad \frac{\partial y}{\partial r} = \sin\theta, \quad \frac{\partial x}{\partial\theta} = -r\sin\theta, \quad \frac{\partial y}{\partial\theta} = r\cos\theta.$$

Ces dérivées partielles sont continues en tout point de W_α, donc ψ est C^1 sur cet ouvert. D'autre part le jacobien s'écrit

$$\operatorname{Jac}(\psi)(r,\theta) = \begin{vmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial\theta} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial\theta} \end{vmatrix} = \begin{vmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{vmatrix} = r(\cos^2\theta+\sin^2\theta) = r.$$

On voit ainsi que $\operatorname{Jac}(\psi)(r,\theta)$ ne s'annule en aucun point de W_α. Comme ψ est une bijection de W_α sur V_α, le théorème d'inversion globale nous permet d'affirmer que ψ est un C^1-difféomorphisme de W_α sur V_α. Son inverse φ est donc un C^1-difféomorphisme de V_α

sur W_α. Nous pouvons maintenant appliquer le théorème de changement de variable 5.28 pour obtenir en rappelant que $\int_{\mathbb{R}^2} f \, \mathrm{d}\lambda_2 = \int_{V_\alpha} f \, \mathrm{d}\lambda_2$,

$$\int_{\mathbb{R}^2} f(x,y) \, \mathrm{d}\lambda_2(x,y) = \int_{]0,+\infty[\times]\alpha,\alpha+2\pi[} f(r\cos\theta, r\sin\theta) r \, \mathrm{d}\lambda_2(r,\theta), \tag{5.33}$$

formule valide pour toute fonction f mesurable positive ou λ_2 intégrable sur $\mathbb{R}^2$.

Exemple 5.1 (Aire d'un disque)**.** Le passage en coordonnées polaires permet de calculer facilement la mesure de Lebesgue d'un disque. En raison de l'invariance par translation, on ne perd pas de généralité en le supposant centré à l'origine. Soit donc $\Delta = \Delta(0,R) := \{(x,y) \in \mathbb{R}^2;\ x^2+y^2 \leq R^2\}$. En appliquant la formule (5.33) à la fonction $f = \mathbf{1}_\Delta$, on obtient :

$$\begin{aligned}
\lambda_2(\Delta) = \int_{\mathbb{R}^2} \mathbf{1}_\Delta \, \mathrm{d}\lambda_2 &= \int_{\mathbb{R}^2} \mathbf{1}_{\{x^2+y^2\leq R^2\}} \, \mathrm{d}\lambda_2(x,y) \\
&= \int_{]0,+\infty[\times]0,2\pi[} \mathbf{1}_{\{r\leq R\}} r \, \mathrm{d}\lambda_2(r,\theta) \\
&= \left\{ \int_0^R r \, \mathrm{d}r \right\} \left\{ \int_{]0,2\pi[} \mathrm{d}\lambda_1(\theta) \right\} \\
&= (2\pi)\frac{R^2}{2}.
\end{aligned}$$

Sans surprise, on trouve ainsi $\lambda_2\big(\Delta(0,R)\big) = \pi R^2$.

Exemple 5.2 (Densité gaussienne)**.** Pour tous $\sigma > 0$ et $m \in \mathbb{R}$, les fonctions $x \mapsto (2\pi\sigma)^{-1/2} \exp\big(-(x-m)^2/(2\sigma^2)\big)$ sont des densités de probabilité par rapport à λ_1. La vérification de cette affirmation se ramène après changement de variable au calcul de l'intégrale de Riemann généralisée $I = \int_{-\infty}^{+\infty} \exp(-x^2) \, \mathrm{d}x$. En utilisant le corollaire du théorème de Fubini Tonelli pour les fonctions à variables séparées, on vérifie facilement que :

$$I^2 = \int_{\mathbb{R}^2} \exp\big(-(x^2+y^2)\big) \, \mathrm{d}\lambda_2(x,y).$$

On calcule ensuite cette intégrale double par passage en coordonnées polaires...et on trouve $I^2 = \pi$, d'où $I = \sqrt{\pi}$. La rédaction détaillée est laissée en exercice (pour un corrigé voir les Annales d'IFP, session de septembre 2002).

Coordonnées sphériques

Le passage des coordonnées cartésiennes de $\mathbb{R}^3$ aux coordonnées sphériques peut être défini par les formules :

$$\left\{ \begin{array}{lcl} x &=& r\cos s \cos t, \\ y &=& r \sin s \cos t, \\ z &=& r \sin t. \end{array} \right. \tag{5.34}$$

La variable r est la distance à l'origine, s peut s'interpréter comme la longitude et t comme la latitude.

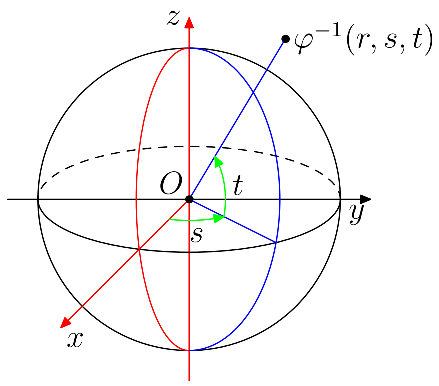


FIGURE 5.5 – Coordonnées sphériques

Là aussi il est plus commode de travailler avec φ^{-1}. Pour avoir un C^1 difféomorphisme d'un ouvert de $\mathbb{R}^3$ sur son image, on prive $\mathbb{R}^3$ du demi plan méridien $\{x \geq 0,\ y = 0\}$. Les formules (5.34) définissent une bijection $\varphi^{-1} : W \to V$, où :

$$W :=]0,+\infty[\times]0,2\pi[\times\Big]\frac{-\pi}{2},\frac{\pi}{2}\Big[, \qquad V := \mathbb{R}^3 \setminus \{(x,y,z) \in \mathbb{R}^3;\ x \geq 0,\ y = 0\}.$$

Les dérivées partielles de φ^{-1} apparaissent dans le calcul suivant de son jacobien et sont visiblement des fonctions continues de (r,s,t).

$$J := \operatorname{Jac}(\varphi^{-1})(r,s,t) = \begin{vmatrix} \frac{\partial x}{\partial r} & \frac{\partial y}{\partial r} & \frac{\partial z}{\partial r} \\ \frac{\partial x}{\partial s} & \frac{\partial y}{\partial s} & \frac{\partial z}{\partial s} \\ \frac{\partial x}{\partial t} & \frac{\partial y}{\partial t} & \frac{\partial z}{\partial t} \end{vmatrix} = \begin{vmatrix} \cos s\cos t & \sin s\cos t & \sin t \\ -r\sin s\cos t & r\cos s\cos t & 0 \\ -r\cos s\sin t & -r\sin s\sin t & r\cos t \end{vmatrix}$$

En développant suivant la troisième colonne, on trouve

$$\begin{aligned} J &= \sin t \begin{vmatrix} -r\sin s\cos t & r\cos s\cos t \\ -r\cos s\sin t & -r\sin s\sin t \end{vmatrix} + r\cos t \begin{vmatrix} \cos s\cos t & \sin s\cos t \\ -r\sin s\cos t & r\cos s\cos t \end{vmatrix} \\ &= r^2\sin^2 t\cos t \begin{vmatrix} -\sin s & \cos s \\ -\cos s & -\sin s \end{vmatrix} + r^2\cos^3 t \begin{vmatrix} \cos s & \sin s \\ -\sin s & \cos s \end{vmatrix} \\ &= r^2(\sin^2 t + \cos^2 t)\cos t = r^2\cos t. \end{aligned}$$

On remarque que J ne s'annule en aucun point de W. Par le théorème d'inversion globale, φ^{-1} (et donc aussi φ) est un C^1 difféomorphisme. Ainsi en notant que $\lambda_3(\mathbb{R}^3 \setminus V) = 0$ (le demi-plan méridien exclu est inclus dans un plan qui est un sous-espace affine de $\mathbb{R}^3$ de mesure λ_3 nulle), on peut écrire la formule de changement de variable

$$\int_{\mathbb{R}^3} f\,\mathrm{d}\lambda_3 = \int_{]0,+\infty[\times]0,2\pi[\times\left]\frac{-\pi}{2},\frac{\pi}{2}\right[} f(r\cos s\cos t, r\sin s\cos t, r\sin t)r^2\cos t\,\mathrm{d}\lambda_3(r,s,t).$$

Cette formule est valide pour toute fonction f mesurable positive ou λ_3 intégrable.

Exemple 5.3 (Volume d'une boule euclidienne)**.** En prenant dans la formule ci-dessus $f = \mathbf{1}_B$, où $B = B(0,R) := \{(x,y,z) \in \mathbb{R}^3;\ x^2+y^2+z^2 \leq R^2\}$ est la boule euclidienne de centre 0 et de rayon R, on obtient :

$$\begin{aligned}
\lambda_3(B) &= \int_{]0,R[\times]0,2\pi[\times]-\pi/2,\pi/2[} r^2 \cos t \, \mathrm{d}\lambda_3(r,s,t) \\
&= \left\{\int_{]0,R[} r^2 \, \mathrm{d}\lambda_1(r)\right\}\left\{\int_{]0,2\pi[} \mathrm{d}\lambda_1(s)\right\}\left\{\int_{]-\pi/2,\pi/2[} \cos t \, \mathrm{d}\lambda_1(t)\right\} \\
&= 2\pi \left\{\int_0^R r^2 \, \mathrm{d}r\right\}\left\{\int_{-\pi/2}^{\pi/2} \cos t \, \mathrm{d}t\right\} \\
&= 2\pi \times \frac{R^3}{3} \times 2.
\end{aligned}$$

On vérifie ainsi que le volume d'une boule euclidienne de rayon R est

$$\lambda_3\big(B(0,R)\big) = \frac{4}{3}\pi R^3.$$

5.4.7 Calculs de lois par changement de variable dans $\mathbb{R}^d$

Soit $X = (X_1, \dots, X_d)$ un vecteur aléatoire dans $\mathbb{R}^d$ ayant une densité f par rapport à la mesure de Lebesgue λ. On cherche la loi P_Y de $Y = \varphi(X)$ où l'on suppose que φ est un C^1-difféomorphisme d'un ouvert V sur $\varphi(V) \subset \mathbb{R}^d$ et que $P_X(\mathbb{R}^d \setminus V) = 0$.

Pour ce faire, on calcule de deux façons $\mathbf{E}h(Y)$ où $h : \mathbb{R}^d \to \mathbb{R}_+$ est une fonction mesurable positive quelconque.

Première façon : par transfert $\Omega \xrightarrow{Y} \mathbb{R}^d$.

$$\mathbf{E}h(Y) = \int_\Omega h(Y) \, \mathrm{d}\mathbf{P} = \int_{\mathbb{R}^d} h(y) \, \mathrm{d}P_Y(y). \tag{5.35}$$

Deuxième façon : par transfert $\Omega \xrightarrow{X} \mathbb{R}^d$.

$$\mathbf{E}h(Y) = \int_\Omega h(\varphi(X)) \, \mathrm{d}\mathbf{P} = \int_{\mathbb{R}^d} h(\varphi(x) \, \mathrm{d}P_X(x) = \int_{\mathbb{R}^d} h(\varphi(x) f(x) \, \mathrm{d}\lambda(x). \tag{5.36}$$

Comme nous avons supposé que $P_X(\mathbb{R}^d \setminus V) = 0$, la densité f de X est nulle λ-p.p. sur le complémentaire de l'ouvert V d'où

$$\mathbf{E}h(Y) = \int_V h(\varphi(x) f(x) \, \mathrm{d}\lambda(x). \tag{5.37}$$

Les hypothèses du théorème 5.28 étant satisfaites, la formule de changement de variable nous donne

$$\mathbf{E}h(Y) = \int_{\varphi(V)} h(y) f\Big(\varphi^{-1}(y)\Big) |\operatorname{Jac}(\varphi^{-1})(y)| \, \mathrm{d}\lambda(y). \tag{5.38}$$

En notant μ la mesure sur $(\mathbb{R}^d, \mathrm{Bor}(\mathbb{R}^d))$ de densité par rapport à λ

$$g(y) := f\Big(\varphi^{-1}(y)\Big) |\operatorname{Jac}(\varphi^{-1})(y)| \mathbf{1}_{\varphi(V)}(y),$$

(5.38) s'écrit $\mathbf{E}h(Y) = \int_{\mathbb{R}^d} h \,\mathrm{d}\mu$. En comparant avec (5.35), nous voyons que pour toute fonction h mesurable positive, $\int_{\mathbb{R}^d} h \,\mathrm{d}P_Y = \int_{\mathbb{R}^d} h \,\mathrm{d}\mu$. Il en résulte immédiatement que $\mu = P_Y$ (prendre $h = \mathbf{1}_B$, B borélien quelconque). En conclusion la loi de Y est la mesure à densité g par rapport à λ.

Exemple 5.4. Soit (X_1, X_2) un vecteur aléatoire de densité

$$f(s,t) = \frac{1}{2\pi} \exp\Bigl(-\frac{s^2+t^2}{2}\Bigr).$$

On définit la variable aléatoire réelle

$$Y_1 := \begin{cases} X_1/X_2 & \text{si } X_2 \neq 0, \\ 0 & \text{si } X_2 = 0. \end{cases}$$

Quelle est la loi de Y_1 ?

Solution. Il est facile de vérifier que la fonction continue (donc mesurable) positive f est bien une densité, soit en remarquant que $f(s,t) = (g \otimes g)(s,t)$, où g est la densité gaussienne standard [12], soit en calculant $\int_{\mathbb{R}^2} f \,\mathrm{d}\lambda_2$ par passage en coordonnées polaires. Y_1 est bien une variable aléatoire comme fonction mesurable d'un vecteur aléatoire.

Pour pouvoir exploiter le théorème de changement de variable, on complète Y_1 en un vecteur aléatoire (Y_1, Y_2) de même dimension que (X_1, X_2) par adjonction de la variable aléatoire $Y_2 := X_2$. On va chercher la loi de (Y_1, Y_2) par la méthode du changement de variable. La loi de Y_1, première loi marginale, s'obtiendra alors par intégration partielle de la densité du vecteur. On a ainsi $(Y_1, Y_2) = \varphi(X_1, X_2)$, où φ est définie sur $\mathbb{R}^2$ par

$$(u,v) = \varphi(s,t) = \begin{cases} (s/t, t) & \text{si } t \neq 0, \\ (0,t) & \text{si } t = 0. \end{cases}$$

On vérifie facilement que φ restreinte à l'ouvert $W := \{(s,t) \in \mathbb{R}^2;\ t \neq 0\}$ est une bijection de cet ouvert sur lui même. Son inverse s'écrit

$$(s,t) = \varphi^{-1}(u,v) = (uv, v), \quad (u,v) \in \varphi(W) = W.$$

Le déterminant jacobien de φ^{-1} est donc

$$\operatorname{Jac}(\varphi^{-1})(u,v) = \begin{vmatrix} v & 0 \\ u & 1 \end{vmatrix} = v.$$

Les 4 dérivées partielles de φ^{-1} sont des fonctions continues de (u,v) et le jacobien ne s'annule en aucun point de W, on a donc bien un C^1-difféomorphisme de W sur W. Remarquons aussi que $\mathbf{P}\Bigl((X_1, X_2) \notin W\Bigr) = \mathbf{P}(X_2 = 0) = 0$ parce que X_2 est une variable aléatoire à densité. On aura donc pour toute fonction g mesurable positive,

$$\int_{\mathbb{R}^2} g \,\mathrm{d}P_{(X_1,X_2)} = \int_W g \,\mathrm{d}P_{(X_1,X_2)}. \tag{5.39}$$

12. En anticipant légèrement sur la section suivante, on voit ainsi que X_1 et X_2 sont deux variables aléatoires gaussiennes indépendantes.

Soit h borélienne $\mathbb{R}^2 \to \mathbb{R}_+$, calculons de deux façons $\mathbf{E}h(Y_1, Y_2)$. D'abord par transfert $\Omega \xrightarrow{(Y_1,Y_2)} \mathbb{R}^2$,

$$\mathbf{E}h(Y_1, Y_2) = \int_{\mathbb{R}^2} h(u, v) \,\mathrm{d}P_{(Y_1,Y_2)}(u, v). \tag{5.40}$$

D'autre part comme $\mathbf{E}h(Y_1, Y_2) = \mathbf{E}h\bigl(\varphi(X_1, X_2)\bigr)$, le transfert $\Omega \xrightarrow{(X_1,X_2)} \mathbb{R}^2$, (5.39) et l'expression de la densité f de (X_1, X_2) nous donnent

$$\begin{aligned}
\mathbf{E}h(Y_1, Y_2) &= \int_{\mathbb{R}^2} h\bigl(\varphi(s, t)\bigr) \,\mathrm{d}P_{(X_1,X_2)}(s, t) \\
&= \int_W h\bigl(\varphi(s, t)\bigr) \,\mathrm{d}P_{(X_1,X_2)}(s, t) \\
&= \int_W h\bigl(\varphi(s, t)\bigr) \frac{1}{2\pi} \exp\Bigl(-\frac{s^2 + t^2}{2}\Bigr) \,\mathrm{d}\lambda_2(s, t). \qquad (5.41)
\end{aligned}$$

Le changement de variable φ est un C^1 difféomorphisme de l'ouvert W sur lui même. Appliqué à l'intégrale (5.41), il conduit à

$$\mathbf{E}h(Y_1, Y_2) = \int_W h(u, v) \frac{1}{2\pi} \exp\Bigl(-\frac{1}{2}(u^2 + 1)v^2\Bigr)|v| \,\mathrm{d}\lambda_2(u, v). \tag{5.42}$$

Les égalités (5.40) et (5.42) étant valables pour toute h mesurable positive, leur comparaison montre que la loi $P_{(Y_1,Y_2)}$ du vecteur aléatoire (Y_1, Y_2) est la mesure à densité par rapport à λ_2 :

$$(u, v) \longmapsto \frac{1}{2\pi} \exp\Bigl(-\frac{1}{2}(u^2 + 1)v^2\Bigr)|v|\mathbf{1}_W(u, v).$$

La densité f_{Y_1} de la loi marginale P_{Y_1} s'en déduit par application de la proposition 5.21 :

$$f_{Y_1}(u) = \int_{\mathbb{R}} \frac{1}{2\pi} \exp\Bigl(-\frac{1}{2}(u^2 + 1)v^2\Bigr)|v|\mathbf{1}_W(u, v) \,\mathrm{d}\lambda_1(v) \tag{5.43}$$

En se souvenant que $W = \{(u, v) \in \mathbb{R}^2;\ v \neq 0\}$, on voit que $\mathbf{1}_W(u, v) = \mathbf{1}_{\mathbb{R}^*}(v)$ (quel que soit $u \in \mathbb{R}$). On calcule alors l'intégrale (5.43) ainsi (les justifications suivent) :

$$\begin{aligned}
f_{Y_1}(u) &= \int_{]-\infty,0[} \frac{1}{2\pi} \exp\Bigl(-\frac{1}{2}(u^2 + 1)v^2\Bigr)(-v) \,\mathrm{d}\lambda_1(v) \\
&\quad + \int_{]0,+\infty[} \frac{1}{2\pi} \exp\Bigl(-\frac{1}{2}(u^2 + 1)v^2\Bigr)v \,\mathrm{d}\lambda_1(v) \qquad (5.44) \\
&= -\int_{-\infty}^{0} \frac{1}{2\pi} \exp\Bigl(-\frac{1}{2}(u^2 + 1)v^2\Bigr)v \,\mathrm{d}v + \int_0^{+\infty} \frac{1}{2\pi} \exp\Bigl(-\frac{1}{2}(u^2 + 1)v^2\Bigr)v \,\mathrm{d}v \\
&= \frac{1}{\pi} \int_0^{+\infty} \exp\Bigl(-\frac{1}{2}(u^2 + 1)v^2\Bigr)v \,\mathrm{d}v \qquad (5.45) \\
&= \frac{1}{\pi} \int_0^{+\infty} \exp(-z) \frac{1}{1 + u^2} \,\mathrm{d}z \qquad (5.46) \\
&= \frac{1}{\pi(1 + u^2)} \qquad (5.47)
\end{aligned}$$

Justifications :

1. Découpage (5.44) : $\{0\}$ est de λ_1 mesure nulle.
2. De (5.44) à (5.45) : la fonction à intégrer est continue et a une intégrale de Riemann généralisée absolument convergente sur chacun des intervalles $]-\infty, 0[$ et $]0, +\infty[$. Ceci justifie la conversion des intégrales de Lebesgue en intégrales de Riemann. La réduction à une seule intégrale de Riemann s'obtient par le changement de variable $v \mapsto -v$.
3. De (5.45) à (5.46) : changement de variable $z = \frac{1+u^2}{2} v^2$. Bien noter qu'ici $\frac{1+u^2}{2}$ est une constante.

Conclusion : l'égalité (5.47) montre que $Y_1 = X_1/X_2$ suit la loi de Cauchy. □

5.5 Indépendance

L'indépendance d'événements et de variables aléatoires discrètes a été étudiée [13] en DEUG. La notion de tribu produit et de mesure produit permet de généraliser et de systématiser cette étude. Dans toute cette section, on travaille avec un même espace probabilisé $(\Omega, \mathcal{F}, \mathbf{P})$ sur lequel seront définis les variables et vecteurs aléatoires considérés. Nous réservons l'appelation « évènement » aux ensembles membres de la tribu $\mathcal{F}$.

5.5.1 Indépendance d'évènements

Définition 5.32. *Deux évènements A et B (i.e. $A \in \mathcal{F}$ et $B \in \mathcal{F}$) sont indépendants si*

$$\mathbf{P}(A \cap B) = \mathbf{P}(A)\mathbf{P}(B).$$

Remarques 5.33.

a) Si A est un événement tel que $P(A) = 0$ ou $P(A) = 1$, alors il est indépendant de tout événement, *y compris de lui même* (c'est le cas en particulier pour Ω et $\emptyset$).

b) Deux événements *incompatibles* A et B avec $P(A) > 0$ et $P(B) > 0$ ne sont *jamais indépendants*. En effet $A \cap B = \emptyset$ implique $P(A \cap B) = 0$ or $P(A)P(B) \neq 0$.

c) L'indépendance de deux événements A et B n'est pas une propriété intrinsèque aux événements, elle est toujours relative à l'espace probabilisé $(\Omega, \mathcal{F}, \mathbf{P})$.

d) Si $\mathbf{P}(B) \neq 0$, l'indépendance de A et B équivaut à $\mathbf{P}(A \mid B) = \mathbf{P}(A)$. Ceci exprime bien l'idée intuitive d'indépendance : la connaissance de la réalisation de B ne modifie pas notre degré d'incertitude sur celle de A.

Définition 5.34. *Soit I un ensemble quelconque d'indices.*

a) Les évènements d'une famille $(A_i)_{i \in I}$ sont mutuellement indépendants si

$$\forall J \textit{ fini } \subset I, \qquad \mathbf{P}\Big(\bigcap_{j \in J} A_j\Big) = \prod_{j \in J} \mathbf{P}(A_j).$$

13. Il est conseillé de (re)lire les sections 2.2, 4.3 et 5.4 de [ICP].

b) *Les classes d'évènements* $(\mathcal{C}_i)_{i\in I}$ *(i.e.* $\forall i \in I$, $\mathcal{C}_i \subset \mathcal{F}$*) sont mutuellement indépendantes si pour tout choix d'un* A_i *dans chaque* $\mathcal{C}_i$*, les* $(A_i)_{i\in I}$ *sont mutuellement indépendants au sens du a).*

L'indépendance mutuelle est plus forte que l'indépendance deux à deux. Les évènements A, B, C sont indépendants deux à deux s'ils vérifient

$$\mathbf{P}(A\cap B) = \mathbf{P}(A)\mathbf{P}(B), \quad \mathbf{P}(B\cap C) = \mathbf{P}(B)\mathbf{P}(C), \quad \mathbf{P}(C\cap A) = \mathbf{P}(C)\mathbf{P}(A).$$

Pour être mutuellement indépendants, ils devraient vérifier *en plus* la relation

$$\mathbf{P}(A\cap B\cap C) = \mathbf{P}(A)\mathbf{P}(B)\mathbf{P}(C),$$

laquelle ne peut se déduire des trois précédentes. Dans toute la suite, sauf mention explicite du contraire, nous allègerons « mutuellement indépendant(e)s » en « indépendant(e)s ».

Proposition 5.35. *Si* $\mathcal{C}_1,\dots,\mathcal{C}_n$ *sont des* π*-classes indépendantes d'évènements, alors les tribus engendrées* $\sigma(\mathcal{C}_1),\dots,\sigma(\mathcal{C}_n)$ *sont indépendantes.*

Preuve. Nous nous contenterons de faire la preuve pour $n = 2$. Fixons $A_1 \in \mathcal{C}_1$ et définissons

$$\mathcal{E}_{A_1} := \{A_2 \in \sigma(\mathcal{C}_2);\ \mathbf{P}(A_1\cap A_2) = \mathbf{P}(A_1)\mathbf{P}(A_2)\}.$$

Montrons que $\mathcal{E}_{A_1}$ est une λ-classe. Comme Ω est indépendant de tout événement, $\Omega \in \mathcal{E}_{A_1}$. Pour vérifier la stabilité par union dénombrable croissante, soit $(A_{2,n})_{n\geq 1}$ une suite croissante pour l'inclusion dans $\mathcal{E}_{A_1}$. Sa limite A_2, c'est-à-dire l'union des $A_{2,n}$, est dans $\sigma(\mathcal{C}_2)$ et il s'agit de prouver l'indépendance de A_1 et A_2. La convergence ensembliste $A_{2,n} \uparrow A_2$ implique $A_1 \cap A_{2,n} \uparrow A_1 \cap A_2$ et toutes deux donnent par continuité croissante de $\mathbf{P}$

$$\mathbf{P}(A_{2,n}) \uparrow \mathbf{P}(A_2) \quad \text{et} \quad \mathbf{P}(A_1\cap A_{2,n}) \uparrow \mathbf{P}(A_1\cap A_2).$$

On en déduit l'indépendance de A_1 et A_2 en passant à la limite dans les égalités

$$\mathbf{P}(A_1\cap A_{2,n}) = \mathbf{P}(A_1)\mathbf{P}(A_{2,n})$$

qui traduisent l'appartenance des $A_{2,n}$ à $\mathcal{E}_{A_1}$. Pour vérifier la stabilité par différence propre, soient A_2 et B_2 dans $\mathcal{E}_{A_1}$ tels que $A_2 \subset B_2$. On peut alors écrire

$$\begin{aligned}
\mathbf{P}\Big(A_1\cap (B_2\setminus A_2)\Big) &= \mathbf{P}\Big((A_1\cap B_2)\setminus (A_1\cap A_2)\Big)\\
&= \mathbf{P}(A_1\cap B_2) - \mathbf{P}(A_1\cap A_2)\\
&= \mathbf{P}(A_1)\mathbf{P}(B_2) - \mathbf{P}(A_1)\mathbf{P}(A_2)\\
&= \mathbf{P}(A_1)\Big(\mathbf{P}(B_2) - \mathbf{P}(A_2)\Big)\\
&= \mathbf{P}(A_1)\mathbf{P}(B_2\setminus A_2).
\end{aligned}$$

Donc $B_2 \setminus A_2$ est dans $\mathcal{E}_{A_1}$. Ainsi $\mathcal{E}_{A_1}$ est une λ-classe qui contient $\mathcal{C}_2$. Par le théorème de Dynkin, elle contient aussi la tribu $\sigma(\mathcal{C}_2)$, donc $\mathcal{E}_{A_1} = \sigma(\mathcal{C}_2)$. Ceci étant vrai pour tout $A_1 \in \mathcal{C}_1$, nous avons établi que

$$\forall A_1 \in \mathcal{C}_1, \forall A_2 \in \sigma(\mathcal{C}_2), \quad \mathbf{P}(A_1 \cap A_2) = \mathbf{P}(A_1)\mathbf{P}(A_2).$$

Fixons $A_2 \in \sigma(\mathcal{C}_2)$ et définissons

$$\mathcal{G}_{A_2} := \{A_1 \in \sigma(\mathcal{C}_1);\ \mathbf{P}(A_1 \cap A_2) = \mathbf{P}(A_1)\mathbf{P}(A_2)\}.$$

En procédant comme ci-dessus, on montre que $\mathcal{G}_{A_2}$ est un λ-système qui contient $\mathcal{C}_1$, on montre que $\mathcal{G}_{A_2} = \sigma(\mathcal{C}_1)$. Ceci étant vrai pour tout $A_2 \in \sigma(\mathcal{C}_2)$, on en déduit que $\mathbf{P}(A_1 \cap A_2) = \mathbf{P}(A_1)\mathbf{P}(A_2)$ pour tous $A_1 \in \sigma(\mathcal{C}_1)$ et $A_2 \in \sigma(\mathcal{C}_2)$. L'indépendance des tribus $\sigma(\mathcal{C}_1)$ et $\sigma(\mathcal{C}_2)$ est établie. □

Pour une application élémentaire de la proposition 5.35, examinons le cas où les $\mathcal{C}_i$ sont réduites à un seul évènement A_i (ce sont alors trivialement des π classes). Les tribus engendrées sont alors les $\sigma(\mathcal{C}_i) = \{A_i, A_i^c, \Omega, \emptyset\}$. On en déduit le corollaire suivant (que l'on pourra aussi démontrer directement à titre d'exercice ou trouver dans [ICP, Prop. 2.12]).

Corollaire 5.36. *Si $A_1, \dots, A_n$ est une suite finie d'évènements indépendants, alors toute suite $B_1, \dots, B_n$ telle que pour chaque i, $B_i = A_i$ ou $B_i = A_i^c$ est encore une suite d'évènements indépendants. La même propriété reste valable pour les suites infinies d'évènements indépendants.*

Voici maintenant une application de l'indépendance d'une suite d'évènements connue sous le nom de *deuxième lemme de Borel Cantelli.* Rappelons que si $(A_n)_{n\in\mathbb{N}}$ est une suite quelconque d'évènements,

$$\limsup_{n\to\infty} A_n = \bigcap_{n\in\mathbb{N}} \bigcup_{k\geq n} A_k = \{\omega \in \Omega;\ \omega \text{ réalise un infinité de } A_k\}$$

Lemme 5.37 (Borel Cantelli II)**.** *Soit $(A_n)_{n\in\mathbb{N}}$ une suite d'évènements* indépendants *telle que*

$$\sum_{n=0}^{+\infty} \mathbf{P}(A_n) = +\infty. \tag{5.48}$$

Alors

$$\mathbf{P}\Big(\limsup_{n\to\infty} A_n\Big) = 1,$$

autrement dit, presque sûrement une infinité d'évènements A_k se réalisent.

Preuve. Posons

$$B_{n,m} := \bigcup_{n\leq k\leq m} A_k, \qquad B_n := \bigcup_{k\geq n} A_k.$$

Par le corollaire 5.36, les A_k^c sont indépendants, d'où

$$\mathbf{P}(B_{n,m}) = 1 - \mathbf{P}(B_{n,m}^c) = 1 - \mathbf{P}\Big(\bigcap_{k=n}^{m} A_k^c\Big) = 1 - \prod_{k=n}^{m}\big(1 - \mathbf{P}(A_k)\big).$$

On utilise alors l'inégalité de convexité [14] $\mathrm{e}^{-x} \geq 1 - x$ avec $x = \mathbf{P}(A_k)$ pour obtenir la minoration

$$1 \geq \mathbf{P}(B_{n,m}) \geq 1 - \prod_{k=n}^{m} \exp\big(-\mathbf{P}(A_k)\big) = 1 - \exp\Big(-\sum_{k=n}^{m} \mathbf{P}(A_k)\Big). \tag{5.49}$$

En laissant n fixe et faisant tendre m vers l'infini dans (5.49), on en déduit grâce à l'hypothèse (5.48) que $\mathbf{P}(B_{n,m})$ tend vers 1. D'autre part B_n est limite croissante pour l'inclusion des $B_{n,m}$, donc par continuité croissante séquentielle de $\mathbf{P}$,

$$\mathbf{P}(B_n) = \lim_{m\to+\infty} \mathbf{P}(B_{n,m}) = 1.$$

cette égalité étant vraie pour tout $n \in \mathbb{N}$, on en déduit que

$$\mathbf{P}\Big(\bigcap_{n\in\mathbb{N}} B_n\Big) = 1,$$

puisque

$$\mathbf{P}\Big(\bigcup_{n\in\mathbb{N}} B_n^c\Big) \leq \sum_{k\in\mathbb{N}} \mathbf{P}(B_n^c) = 0.$$

Comme l'intersection de tous les B_n est l'évènement $\limsup A_n$, le lemme est démontré. □

5.5.2 Indépendance de variables aléatoires

Nous définissons maintenant l'indépendance d'une famille quelconque $(X_i)_{i\in I}$ de variables ou vecteurs aléatoires. Cette indépendance est celle de la famille de sous-tribus de $\mathcal{F}$ engendrées par les X_i. Rapplons que si X est une application $(\Omega, \mathcal{F}) \to (E, \mathcal{B})$, la tribu engendrée par X est

$$\sigma(X) := X^{-1}(\mathcal{B}) = \{X^{-1}(B);\ B \in \mathcal{B}\}.$$

Si de plus, X est $\mathcal{F}$ - $\mathcal{B}$ mesurable, alors $\sigma(X)$ est une sous-tribu de $\mathcal{F}$. On peut dire alors que X est une variable aléatoire à valeurs dans l'espace mesurable $(E, \mathcal{B})$. En pratique, E peut être $\overline{\mathbb{R}}_+$, $\mathbb{R}$, $\mathbb{C}$, $\mathbb{N}$, $\mathbb{R}^d$, ...et la tribu $\mathcal{B}$ est la tribu borélienne correspondante ou la tribu $\mathcal{P}(E)$ lorsque E est dénombrable.

14. La représentation graphique de la fonction convexe $x \mapsto \mathrm{e}^{-x}$ est toujours au-dessus de sa tangente à l'origine d'où $\mathrm{e}^{-x} \geq 1 - x$ pour tout $x \in \mathbb{R}$.

Définition 5.38. *Soit I une famille quelconque d'indices et pour tout $i \in I$, X_i une variable aléatoire à valeurs dans l'espace mesurable $(E_i, \mathcal{B}_i)$. On dit que $(X_i)_{i\in I}$ est une famille de variables aléatoires indépendantes si la famille de tribus $(\sigma(X_i))_{i\in I}$ est indépendante, autrement dit si*

$$\forall J \text{ fini } \subset I, \quad \forall j \in J, \forall B_j \in \mathcal{B}_j, \quad \mathbf{P}(\forall j \in J,\ X_j \in B_j) = \prod_{j\in J} \mathbf{P}(X_j \in B_j).$$

Notons que cette définition est assez souple pour englober l'indépendance d'une collection complètement hétéroclite de « variables aléatoires », certaines pouvant être des variables aléatoires réelles, d'autres des vecteurs aléatoires (de dimensions diverses), d'autres des variables aléatoires discrètes, ...

Proposition 5.39. *Un vecteur aléatoire $(X_1, \dots, X_d)$ à valeurs dans $\mathbb{R}^d$ est à composantes indépendantes si et seulement si sa loi est le produit de ses lois marginales, i.e.*

$$P_{(X_1,\dots,X_d)} = P_{X_1} \otimes \cdots \otimes P_{X_d}.$$

Preuve. Supposons d'abord que les variables aléatoires réelles $X_1, \dots, X_d$ soient indépendantes et soit B un pavé mesurable pour la tribu produit $\mathrm{Bor}(\mathbb{R}^1)^{\otimes d}$, B s'écrit donc $B = B_1 \times \cdots \times B_d$, avec pour chaque $i = 1, \dots d$, $B_i \in \mathrm{Bor}(\mathbb{R}^1)$. On a alors

$$\begin{aligned}
P_{(X_1,\dots,X_d)}(B) &= \mathbf{P}\Big((X_1, \dots, X_d) \in B_1 \times \cdots \times B_d\Big) && (5.50)\\
&= \mathbf{P}\Big(\forall i \in \{1, \dots, d\}, X_i \in B_i\Big) && (5.51)\\
&= \mathbf{P}\left(\bigcap_{i=1}^{d} X_i^{-1}(B_i)\right) && (5.52)\\
&= \mathbf{P}\Big(X_1^{-1}(B_1)\Big) \dots \mathbf{P}\Big(X_d^{-1}(B_d)\Big) && (5.53)\\
&= P_{X_1}(B_1) \dots P_{X_d}(B_d) && (5.54)\\
&= \Big(P_{X_1} \otimes \cdots \otimes P_{X_d}\Big)(B). && (5.55)
\end{aligned}$$

Justifications : (5.50) vient de la définition de la loi du vecteur aléatoire ; (5.51) et (5.52) utilisent la définition d'un produit cartésien et celle d'une intersection. Le passage de (5.52) à (5.53) repose sur *l'hypothèse d'indépendance* des X_i. La définition de la loi de X_i nous fait passer de (5.53) à (5.54). De là on arrive à (5.55) par la définition de la mesure produit sur la classe des pavés mesurables.

Nous venons ainsi de vérifier que les deux mesures $P_{(X_1,\dots,X_d)}$ et $P_{X_1} \otimes \cdots \otimes P_{X_d}$ coïncident sur la classe des pavés mesurables de la tribu produit $\mathrm{Bor}(\mathbb{R}^1)^{\otimes d}$. Par unicité de la mesure produit, ces deux mesures coïncident sur toute la tribu $\mathrm{Bor}(\mathbb{R}^1)^{\otimes d}$. Rappelons que cette tribu n'est autre que $\mathrm{Bor}(\mathbb{R}^d)$, d'après (5.20) et la proposition 5.16.

Réciproquement, supposons l'égalité des deux mesures $P_{(X_1,\dots,X_d)}$ et $P_{X_1} \otimes \cdots \otimes P_{X_d}$ sur la tribu borélienne de $\mathbb{R}^d$. Elles coïncident en particulier sur la classe des pavés mesurables de $\mathrm{Bor}(\mathbb{R}^1)^{\otimes d}$. On a donc pour tout $B = B_1 \times \cdots \times B_d$ avec $B_i \in \mathrm{Bor}(\mathbb{R})$,

$$P_{(X_1,\dots,X_d)}(B) = \Big(P_{X_1} \otimes \cdots \otimes P_{X_d}\Big)(B). \tag{5.56}$$

La justification détaillée ci-dessus des égalités (5.50) à (5.55) montre que seul le passage de (5.52) à (5.53) utilisait l'indépendance des X_i. Nous pouvons donc recycler toutes les autres égalités de la manière suivante. On a égalité des seconds membres de (5.52), (5.51) et (5.50). Les égalités (5.50) et (5.56) nous conduisent au second membre de (5.55) d'où l'on peut remonter jusqu'à celui de (5.53). Ainsi

$$\forall i \in \{1, \dots, d\},\ \forall B_i \in \operatorname{Bor}(\mathbb{R}), \quad \mathbf{P}\left(\bigcap_{i=1}^{d} X_i^{-1}(B_i)\right) = \mathbf{P}\Big(X_1^{-1}(B_1)\Big) \dots \mathbf{P}\Big(X_d^{-1}(B_d)\Big).$$

Soit J une partie quelconque de $\{1, \dots, d\}$ (automatiquement finie) et choisissons $B_i = \mathbb{R}$ lorsque $i \notin J$ dans l'égalité ci-dessus. Pour ces indices, $X_i^{-1}(B_i) = X_i^{-1}(\mathbb{R}) = \Omega$ et $\mathbf{P}\Big(X_i^{-1}(B_i)\Big) = 1$. On peut donc effacer sans inconvénient les évènements correspondants et leurs probabilités dans l'égalité ci-dessus. On obtient alors, avec des boréliens B_j, $(j \in J)$ quelconques :

$$\mathbf{P}\left(\bigcap_{j\in J} X_j^{-1}(B_j)\right) = \prod_{j\in J} \mathbf{P}\Big(X_j^{-1}(B_j)\Big),$$

ce qui prouve l'indépendance des variables aléatoires réelles $X_1, \dots, X_d$. □

Corollaire 5.40. *Soit $X_1, \dots, X_n$ une suite finie de variables aléatoires réelles indépendantes. Alors les vecteurs aléatoires obtenus en découpant dans cette suite des blocs consécutifs disjoints sont indépendants.*

Corollaire 5.41. *Soit $X_1, \dots, X_n$ une suite finie de variables aléatoires réelles indépendantes. Pour tout découpage $n_0 = 0,\ 1 \leq n_1 < n_2 < \dots < n_k = n$ en blocs $Y_j = (X_{1+n_{j-1}}, \dots, X_{n_j})$ $(1 \leq j \leq k)$ et toutes fonctions mesurables $h_j = \mathbb{R}^{n_j - n_{j-1}} \to (E_j, \mathcal{B}_j)$, les variables aléatoires $h_1(Y_1), \dots, h_k(Y_k)$ sont indépendantes.*

Preuve des corollaires 5.40 et 5.41. Le corollaire 5.40 est à l'évidence un cas particulier du corollaire 5.41 en prenant pour h_j l'identité sur $\mathbb{R}^{n_j - n_{j-1}}$. Notons à ce propos que la collection de « variables aléatoires » $Z_j := h_j(Y_j)$ peut être hétéroclite : vecteurs aléatoires dans $\mathbb{R}^{d_j}$, variables aléatoires réelles, complexes ou discrètes. L'indépendance des Y_j, $(1 \leq j \leq k)$ résulte facilement de la proposition 5.39 grâce à l'associativité du produit des tribus et des mesures. Pour le corollaire 5.41, il s'agit de vérifier l'indépendance des tribus $\sigma(Z_j)$ $(1 \leq j \leq k)$. Or ces tribus s'écrivent $\sigma(Z_j) = Y_j^{-1}\Big(h_j^{-1}(\mathcal{B}_j)\Big)$. La mesurabilité de h_j se traduit par l'inclusion de tribus $h_j^{-1}(\mathcal{B}_j) \subset \operatorname{Bor}(\mathbb{R}^{n_j - n_{j-1}})$. On a ainsi

$$\sigma(Z_j) = Y_j^{-1}\Big(h_j^{-1}(\mathcal{B}_j)\Big) \subset Y_j^{-1}\Big(\operatorname{Bor}(\mathbb{R}^{n_j - n_{j-1}})\Big) = \sigma(Y_j).$$

L'indépendance des tribus $\sigma(Y_j)$ résulte du corollaire 5.40 et passe évidemment à leurs sous-tribus $\sigma(Z_j)$. □

Proposition 5.42.

a) Les variables aléatoires réelles $X_1, \dots, X_d$ sont indépendantes si et seulement si

$$\forall (x_1, \dots, x_d) \in \mathbb{R}^d, \quad \mathbf{P}(X_1 \leq x_1, \dots, X_d \leq x_d) = \mathbf{P}(X_1 \leq x_1) \dots \mathbf{P}(X_d \leq x_d). \tag{5.57}$$

b) *Soit* $(X_1, \dots, X_d)$ *un vecteur aléatoire dans* $\mathbb{R}^d$*, de densité* f*. Ses composantes* X_i *sont indépendantes si et seulement s'il existe* $f_1, \dots, f_d$*,* $\mathbb{R} \to \overline{\mathbb{R}}_+$*, mesurables telles que*

$$f = f_1 \otimes \cdots \otimes f_d.$$

Dans ce cas les densités marginales f_{X_i} *sont de la forme* $c_i f_i$ *où les constantes* $c_i > 0$ *sont liées par* $c_1 \dots c_d = 1$.

c) *Soient* $(X_1, \dots, X_d)$ *un vecteur aléatoire discret à valeurs dans* $\mathbb{N}^d$*. Ses composantes* X_i *sont indépendantes si et seulement si*

$$\forall (k_1, \dots, k_d) \in \mathbb{N}^d, \quad \mathbf{P}(X_1 = k_1, \dots, X_d = k_d) = \mathbf{P}(X_1 = k_1) \dots \mathbf{P}(X_d = k_d). \tag{5.58}$$

Preuve du a). Il suffit de remarquer que (5.57) exprime l'égalité des deux mesures finies $P_{(X_1,\dots,X_d)}$ et $P_{X_1} \otimes \cdots \otimes P_{X_d}$ sur la π-classe $\mathcal{C}$ des pavés mesurables de la forme

$$B =]-\infty, x_1] \times \cdots \times]-\infty, x_d], \quad (x_1, \dots, x_d) \in \mathbb{R}^d.$$

Comme $\sigma(\mathcal{C}) = \mathrm{Bor}(\mathbb{R}^d)$, le théorème d'unicité des mesures implique l'égalité de ces mesures sur toute la tribu borélienne de $\mathbb{R}^d$. L'indépendance des X_i en découle par la proposition 5.39. Réciproquement cette indépendance implique immédiatement (5.57).

Notons que le membre de gauche dans (5.57) définit la fonction de répartition $F_{X_1,\dots,X_d}$ du vecteur $(X_1, \dots, X_d)$ et que le membre de droite est le produit (tensoriel) des fonctions de répartition des X_i. On a ainsi une caractérisation de l'indépendance :

$$X_1, \dots, X_d \text{ indépendantes} \quad \Leftrightarrow \quad F_{X_1,\dots,X_d} = F_{X_1} \otimes \cdots \otimes F_{X_d}. \tag{5.59}$$

exprimée en termes de fonctions de répartition. □

Preuve du b).

Supposons d'abord que la densité f du vecteur aléatoire $(X_1, \dots, X_d)$ soit de la forme $f = f_1 \otimes \cdots \otimes f_d$ où les f_i sont des fonctions d'une variables réelle, mesurables positives. Par la proposition 5.21, la loi de chaque composante X_i a aussi une densité f_{X_i}, qui s'obtient en intégrant f par rapport à tous les x_j pour $j \neq i$:

$$\begin{aligned}
f_{X_i}(t) &= \int_{\mathbb{R}^{d-1}} f_1(x_1) \dots f_{i-1}(x_{i-1}) f_i(t) f_{i+1}(x_{i+1}) \dots f_d(x_d) \mathrm{d}\lambda_{d-1}(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_d) \\
&= f_i(t) \int_{\mathbb{R}^{d-1}} f_1(x_1) \dots f_{i-1}(x_{i-1}) f_{i+1}(x_{i+1}) \dots f_d(x_d) \mathrm{d}\lambda_{d-1}(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_d) \\
&= c_i f_i(t).
\end{aligned}$$

Comme f_{X_i} est une densité de probabilité, il est clair que la constante positive c_i ne peut pas être nulle. Calculons $P_{(X_1,\dots,X_d)}(B)$ pour $B = B_1 \times \cdots \times B_d$, pavé mesurable quelconque de $\mathrm{Bor}(\mathbb{R})^{\otimes d}$. En utilisant le corollaire 5.20, on obtient :

$$\begin{aligned} P_{(X_1,\dots,X_d)}(B) = \int_B f \,\mathrm{d}\lambda_d &= \int_{B_1\times\dots\times B_d} (f_1 \otimes \dots \otimes f_d)\,\mathrm{d}(\lambda_1^{\otimes d}) \\ &= \prod_{i=1}^d \int_{B_i} f_i \,\mathrm{d}\lambda_1 \\ &= \prod_{i=1}^d \frac{1}{c_i} \int_{B_i} f_{X_i} \,\mathrm{d}\lambda_1 \\ &= \prod_{i=1}^d \frac{1}{c_i} P_{X_i}(B_i). \end{aligned}$$

Ainsi

$$P_{(X_1,\dots,X_d)}(B_1 \times \dots \times B_d) = a P_{X_1}(B_1) \dots P_{X_d}(B_d), \tag{5.60}$$

où la constante a vaut $(c_1 \dots c_d)^{-1}$. En particulier en choisissant dans (5.60) tous les B_i égaux à $\mathbb{R}$, on obtient $1 = a \times 1$ d'où $a = 1$ et $c_1 \dots c_d = 1$. En reportant cette valeur de a dans (5.60), on voit alors que (5.60) exprime l'indépendance des X_i.

Réciproquement, supposons les X_i indépendantes. Par la proposition 5.21, chaque X_i a une densité f_{X_i}. Évaluons $P_{(X_1,\dots,X_d)}(B)$ pour $B = B_1 \times \dots \times B_d$ pavé mesurable quelconque de $\mathrm{Bor}(\mathbb{R})^{\otimes d}$. En utilisant l'indépendance des X_i, la définition des f_{X_i} et le corollaire 5.20, on obtient :

$$P_{(X_1,\dots,X_d)}(B) = \prod_{i=1}^d P_{X_i}(B_i) = \prod_{i=1}^d \int_{B_i} f_{X_i} \,\mathrm{d}\lambda_1.$$

En appliquant le corollaire 5.20, on en déduit

$$P_{(X_1,\dots,X_d)}(B_1 \times \dots \times B_d) = \int_{B_1\times\dots\times B_d} (f_{X_1} \otimes \dots \otimes f_{X_d})\,\mathrm{d}\lambda_d. \tag{5.61}$$

En particulier en prenant dans (5.61) tous les B_i égaux à $\mathbb{R}$, on voit que

$$\int_{\mathbb{R}^d} (f_{X_1} \otimes \dots \otimes f_{X_d})\,\mathrm{d}\lambda_d = 1,$$

ce qui montre que la fonction mesurable positive $g := f_{X_1} \otimes \dots \otimes f_{X_d}$ est une densité de probabilité. En notant provisoirement μ la mesure de densité g par rapport à λ_d, l'égalité (5.61) s'interprète alors comme l'égalité de μ et de $P_{(X_1,\dots,X_d)}$ sur la classe $\mathcal{R}_d$ des pavés mesurables de $\mathrm{Bor}(\mathbb{R})^{\otimes d}$, donc sur toute la tribu $\mathrm{Bor}(\mathbb{R})^{\otimes d}$ par unicité de la mesure produit. Ainsi f et g sont deux densités par rapport à λ_d de la même mesure. Elles sont donc égales λ_d presque partout [15]. □

15. Exercice : si f et g sont deux densités de la même mesure μ par rapport à une mesure ν, alors $f = g$ ν-presque partout. Indication : soit $A := \{f < g\}$, montrer que $\nu(A) = 0$ en raisonnant par l'absurde en remarquant que A est l'union croissante des $A_n := \{f + 1/n < g\}$.

Preuve du c). La tribu concernée par la mesurabilité du vecteur aléatoire $(X_1, \dots, X_d)$ est ici $\mathcal{P}(\mathbb{N}^d)$. On voit que cette tribu coïncide avec $\mathcal{P}(\mathbb{N})^{\otimes d}$ car tout élément de $\mathcal{P}(\mathbb{N}^d)$ est une partie de $\mathbb{N}^d$, donc est au plus dénombrable et réunion au plus dénombrable de singletons $\{\mathbf{k}\} = \{(k_1, \dots, k_d)\}$. La condition (5.58) traduit l'égalité des mesures $P_{(X_1,\dots,X_d)}$ et $P_{X_1} \otimes \dots \otimes P_{X_d}$ sur la classe des singletons de $\mathbb{N}^d$. L'indépendance des X_i se traduit elle, par l'égalité de ces mesures sur la classe des pavés $A = A_1 \times \dots \times A_d$ où chaque A_i peut être une partie quelconque de $\mathbb{N}$. Elle implique donc (5.58) en prenant pour chaque A_i un singleton $\{k_i\}$ de $\mathbb{N}$. Pour la réciproque, on remarque que A et les A_i sont au plus dénombrables et on utilise les propriétés des séries multiples à termes positifs :

$$\begin{aligned}
P_{(X_1,\dots,X_d)}(A) &= \sum_{(k_1,\dots,k_d)\in A} P_{(X_1,\dots,X_d)}\Big(\{(k_1,\dots,k_d)\}\Big) \\
&= \sum_{(k_1,\dots,k_d)\in A} \mathbf{P}(X_1 = k_1, \dots, X_d = k_d) \\
&= \sum_{(k_1,\dots,k_d)\in A} \mathbf{P}(X_1 = k_1) \dots \mathbf{P}(X_d = k_d) \\
&= \left(\sum_{k_1\in A_1} \mathbf{P}(X_1 = k_1)\right) \times \dots \times \left(\sum_{k_d\in A_d} \mathbf{P}(X_d = k_d)\right) \\
&= \mathbf{P}(X_1 \in A_1) \dots \mathbf{P}(X_d \in A_d) \\
&= (P_{X_1} \otimes \dots \otimes P_{X_d})(A).
\end{aligned}$$

□

Proposition 5.43. *Soit $(X_i)_{i\in I}$ une famille quelconque de variables aléatoires réelles indépendantes. Alors pour toute partie finie J de I et toute famille $\{h_j,\ j \in J\}$ de fonctions boréliennes $\mathbb{R} \to \mathbb{C}$ telles que les $h_j(X_j)$ soient $\mathbf{P}$-intégrables, la variable aléatoire produit $\prod_{j\in J} h_j(X_j)$ est $\mathbf{P}$-intégrable et*

$$\mathbf{E}\left(\prod_{j\in J} h_j(X_j)\right) = \prod_{j\in J} \mathbf{E} h_j(X_j).$$

Preuve. En réindexant l'ensemble fini $J = \{i_1, \dots, i_n\}$, on se ramène au cas $J = \{1, \dots, n\}$. La fonction $\omega \mapsto \prod_{j\in J} h_j\big(X_j(\omega)\big)$ est mesurable comme produit des fonctions mesurables $h_j \circ X_j$. Par le théorème de transfert et la proposition 5.39, on a

$$\begin{aligned}
\mathbf{E}\left|\prod_{j=1}^{n} h_j(X_j)\right| &= \int_{\mathbb{R}^n} |h_1(x_1) \dots h_n(x_n)| \,\mathrm{d}P_{(X_1,\dots,X_n)}(x_1, \dots, x_n) \\
&= \int_{\mathbb{R}^n} |h_1| \otimes \dots \otimes |h_n| \,\mathrm{d}(P_{X_1} \otimes \dots \otimes P_{X_n}).
\end{aligned}$$

Par le corollaire 5.20, cette dernière intégrale est finie dès que chacune des intégrales $\int_{\mathbb{R}} |h_j| \,\mathrm{d}P_{X_j}$ est finie, autrement dit (par transfert) dès que $\mathbf{E}|h_j(X_j)| < +\infty$. Ayant ainsi réglé la question de l'intégrabilité de la variable aléatoire $\prod_{j\in J} h_j(X_j)$, on peut écrire en utilisant successivement le transfert $\Omega \to \mathbb{R}^n$, la proposition 5.39, le corollaire 5.20 et les transferts $\mathbb{K} \xrightarrow{X_j} \Omega$:

$$\begin{aligned}
\mathbf{E}\left(\prod_{j=1}^n h_j(X_j)\right) &= \int_{\mathbb{R}^n} h_1(x_1)\dots h_n(x_n)\,\mathrm{d}P_{(X_1,\dots,X_n)}(x_1,\dots,x_n)\\
&= \int_{\mathbb{R}^n} (h_1\otimes\dots\otimes h_n)\,\mathrm{d}(P_{X_1}\otimes\dots\otimes P_{X_n})\\
&= \prod_{j=1}^n \int_{\mathbb{R}} h_j\,\mathrm{d}P_{X_j}\\
&= \prod_{j=1}^n \mathbf{E}h_j(X_j).
\end{aligned}$$

□

Corollaire 5.44. *Si les variables aléatoires réelles (ou complexes) $X_1,\dots,X_n$ sont indépendantes et intégrables, leur produit est aussi intégrable et*

$$\mathbf{E}(X_1\dots X_n) = (\mathbf{E}X_1)\dots(\mathbf{E}X_n).$$

Il est facile de voir que la réciproque du corollaire 5.44 est fausse en construisant un couple (X_1,X_2) de variables aléatoires réelles non indépendantes telles que $\mathbf{E}(X_1X_2) = \mathbf{E}X_1\mathbf{E}X_2$. Prenons par exemple X_1 de loi uniforme sur $[-1,+1]$ et $X_2 := X_1^2$. On a alors

$$\mathbf{E}X_1X_2 = \mathbf{E}X_1^3 = \int_{[-1,+1]} x^3\,\mathrm{d}\lambda(x) = 0.$$

D'autre part $\mathbf{E}X_1 = 0$ donc $\mathbf{E}X_1\mathbf{E}X_2 = 0$. Il est clair intuitivement que X_1 et X_2 ne sont pas indépendantes puisque X_2 est une fonction déterministe de X_1. Pour vérifier cette non-indépendance par le calcul, on peut remarquer que d'une part

$$\begin{aligned}
\mathbf{P}\Big(X_1\in[0,1/2] \text{ et } X_2\in[0,1/4]\Big) &= \mathbf{P}\Big(X_1\in[0,1/2] \text{ et } X_1\in[-1/2,1/2]\Big)\\
&= \mathbf{P}\Big(X_1\in[0,1/2]\Big) = \frac{1}{4}
\end{aligned}$$

et d'autre part

$$\mathbf{P}\Big(X_1\in[0,1/2]\Big)\mathbf{P}\Big(X_2\in[0,1/4]\Big) = \frac{1}{4}\mathbf{P}\Big(X_1\in[-1/2,1/2]\Big) = \frac{1}{8}.$$

5.5.3 Covariance

Nous venons de voir que si X et Y sont deux variables aléatoires indépendantes et intégrables, $\mathbf{E}XY = \mathbf{E}X\mathbf{E}Y$ et que la réciproque était fausse. Ceci est l'une des motivations pour l'étude des propriétés de la quantité $\mathbf{E}(XY) - \mathbf{E}X\mathbf{E}Y$.

Définition 5.45 (Covariance)**.** *Soient X et Y deux variables aléatoires de carré intégrables. La quantité*

$$\mathrm{Cov}(X,Y) := \mathbf{E}\Big[(X-\mathbf{E}X)(Y-\mathbf{E}Y)\Big]$$

est bien définie et est appelée covariance de X et Y.

Pour légitimer cette définition, il nous faut vérifier que l'hypothèse d'intégrabilité de X^2 et Y^2 entraîne l'existence de $\mathbf{E}X$ et $\mathbf{E}Y$ et l'intégrabilité de $(X-\mathbf{E}X)(Y-\mathbf{E}Y)$. Pour le premier point, il suffit d'écrire

$$\mathbf{E}|X| = \int_{\{|X|\leq 1\}} |X|\,\mathrm{d}\mathbf{P} + \int_{\{|X|>1\}} |X|\,\mathrm{d}\mathbf{P} \leq 1 + \int_{\{|X|>1\}} X^2\,\mathrm{d}\mathbf{P} \leq 1 + \mathbf{E}X^2 < +\infty.$$

Pour le deuxième point, on utilise d'abord l'inégalité $|xy| \leq x^2+y^2$ vraie pour tous réels x, y. Posons $x = X(\omega) - \mathbf{E}X$ et $y = Y(\omega) - \mathbf{E}Y$ et intégrons l'inégalité obtenue sur Ω :

$$\mathbf{E}\Big|(X-\mathbf{E}X)(Y-\mathbf{E}Y)\Big| \leq \mathbf{E}\Big[(X-\mathbf{E}X)^2\Big] + \mathbf{E}\Big[(Y-\mathbf{E}Y)^2\Big]. \tag{5.62}$$

On utilise ensuite l'inégalité $(a+b)^2 \leq 2a^2+2b^2$ en posant $a = X(\omega)$ et $b = \mathbf{E}X$. On en déduit après intégration sur Ω :

$$\mathbf{E}\Big[(X-\mathbf{E}X)^2\Big] \leq 2\mathbf{E}(X^2) + 2(\mathbf{E}X)^2 < +\infty.$$

En reportant cette majoration et son analogue pour Y dans (5.62), on obtient l'intégrabilité de $(X-\mathbf{E}X)(Y-\mathbf{E}Y)$.

Proposition 5.46 (Formule de Koenig)**.** *Si X et Y sont deux variables aléatoires de carré intégrables,*

$$\mathrm{Cov}(X,Y) = \mathbf{E}(XY) - \mathbf{E}X\mathbf{E}Y. \tag{5.63}$$

Preuve. Comme nous venons de le voir, l'intégrabilité de X^2 et de Y^2 implique celles de X, de Y et de XY. Le second membre de (5.63) est donc bien défini. Introduisons pour alléger, les *constantes* $c = \mathbf{E}X$ et $c' = \mathbf{E}Y$. En utilisant la linéarité de l'espérance et le fait que l'espérance d'une constante est égale à cette constante, on obtient :

$$\begin{aligned}
\mathrm{Cov}(X,Y) = \mathbf{E}\Big[(X-c)(Y-c')] &= \mathbf{E}(XY - cY - c'X + cc') \\
&= \mathbf{E}(XY) - c\mathbf{E}Y - c'\mathbf{E}X + cc' \\
&= \mathbf{E}(XY) - cc',
\end{aligned}$$

ce qui établit (5.63). □

Définition 5.47. *On dit que deux variables aléatoires réelles de carré intégrable X et Y sont non-corrélées si*

$$\mathbf{E}(XY) = \mathbf{E}X\mathbf{E}Y.$$

Cette condition équivaut (pour des variables de carré intégrable) à $\mathrm{Cov}(X,Y) = 0$.

Nous savons déjà que deux variables aléatoires indépendantes et de carré intégrable sont non-corrélées et que la réciproque est fausse. Notons aussi que pour des variables aléatoires X et Y indépendantes et intégrables, on a $\mathbf{E}(XY) = \mathbf{E}X\mathbf{E}Y$, même si ces variables ne sont pas de *carré* intégrable.

Définition 5.48 (Variance)**.** *Soit X une variable aléatoire de carré intégrable. On appelle variance de X la quantité*

$$\operatorname{Var} X := \mathbf{E}\Big[(X - \mathbf{E}X)^2\Big] = \operatorname{Cov}(X, X).$$

L'existence a déjà été justifiée lors de la définition de la covariance. La formule de Koenig appliquée à ce cas particulier s'écrit

$$\operatorname{Var} X = \mathbf{E}(X^2) - (\mathbf{E}X)^2. \tag{5.64}$$

Proposition 5.49 (Variance d'une somme)**.**

a) Si les variables aléatoires réelles $X_1, \dots, X_n$ sont de carré intégrable :

$$\operatorname{Var}\Big(\sum_{i=1}^n X_i\Big) = \sum_{i,j=1}^n \operatorname{Cov}(X_i, X_j) \tag{5.65}$$

$$= \sum_{i=1}^n \operatorname{Var} X_i + \sum_{\substack{i,j=1\\ i\neq j}}^n \operatorname{Cov}(X_i, X_j). \tag{5.66}$$

b) Si de plus elles sont deux à deux non corrélées ($\operatorname{Cov}(X_i, X_j) = 0$ pour $i \neq j$),

$$\operatorname{Var}\Big(\sum_{i=1}^n X_i\Big) = \sum_{i=1}^n \operatorname{Var} X_i.$$

Preuve. Il suffit de remarquer que l'application $(X, Y) \mapsto \operatorname{Cov}(X, Y)$ est une forme bilinéaire sur l'espace vectoriel des variables aléatoires de carré intégrable. □

5.6 Convolution

Dans cette section, nous désignons par $(E, \mathcal{B})$ l'espace mesurable $E = \mathbb{R}^d$ ou $\mathbb{Z}^d$ ($d \geq 1$) muni de la tribu $\mathcal{B} = \operatorname{Bor}(\mathbb{R}^d)$ ou $\mathcal{B} = \mathcal{P}(\mathbb{Z}^d)$ respectivement. Dans les deux cas la tribu $\mathcal{B}$ est invariante par toute translation T_u de vecteur u dans E.

Définition 5.50. *Avec les notations ci-dessus, soient μ_1, μ_2 deux mesures finies sur $(E, \mathcal{B})$. Le produit de convolution de μ_1 et μ_2 est la mesure*

$$\mu_1 * \mu_2 := (\mu_1 \otimes \mu_2) \circ s^{-1},$$

où $s : E^2 \to E$, $\quad (x, y) \mapsto x + y$ est l'addition sur $E \times E$.

Notons que $\mu_1 * \mu_2$ est une mesure finie sur $(E, \mathcal{B})$. En effet $s^{-1}(E) = E^2$ et

$$\mu_1 * \mu_2(E) = \mu_1 \otimes \mu_2(E^2) = \mu_1(E)\mu_2(E) < +\infty.$$

En particulier quand μ_1 et μ_2 sont des probabilités sur $(E, \mathcal{B})$, $\mu_1 * \mu_2$ est aussi une probabilité sur $(E, \mathcal{B})$.

Proposition 5.51. *Pour tout $B \in \mathcal{B}$,*

$$\mu_1 * \mu_2(B) = \int_E \mu_2(B - x)\,\mathrm{d}\mu_1(x) = \int_E \mu_1(B - y)\,\mathrm{d}\mu_2(y), \tag{5.67}$$

où $B - x := \{z - x;\ z \in B\} = T_{-x}(B)$.

Preuve. Notons $A := s^{-1}(B) = \{(x, y) \in E^2;\ x + y \in B\}$. Cherchons la section de A en x :

$$\begin{aligned} A_x = \{y \in E;\ (x, y) \in A\} &= \{y \in E;\ x + y \in B\} \\ &= \{y \in E;\ \exists z \in B, y = z - x\} \\ &= \{z - x;\ z \in B\} \\ &= B - x. \end{aligned}$$

Il suffit alors d'appliquer la formule (5.4) de calcul de la mesure produit :

$$\mu_1 * \mu_2(B) = \mu_1 \otimes \mu_2\big(s^{-1}(B)\big) = \mu_1 \otimes \mu_2(A) = \int_E \mu_2(A_x)\,\mathrm{d}\mu_1(x) = \int_E \mu_2(B - x)\,\mathrm{d}\mu_1(x).$$

La deuxième égalité dans (5.67) s'obtient en échangeant les rôles de x et y. □

Remarque 5.52. Formellement, rien n'interdit de généraliser la définition 5.50 à des mesures σ-finies quelconques. Le calcul ci-dessus reste valable. Mais si on prend $E = \mathbb{R}$ et $\mu_1 = \mu_2 = \lambda$ mesure de Lebesgue, on voit que $\mu_1 * \mu_2(B) = +\infty$ pour tout borélien B tel que $\lambda(B) > 0$ puisque $\lambda(B - x) = \lambda(B)$. Ce type de pathologie nous incite à nous restreindre au cas des mesures finies.

Proposition 5.53. *Si X et Y sont deux variables aléatoires* indépendantes *à valeurs dans $(E, \mathcal{B})$, la loi de $X + Y$ est le produit de convolution des lois de X et Y :*

$$X \text{ et } Y \text{ indépendantes} \quad \Rightarrow \quad P_{X+Y} = P_X * P_Y. \tag{5.68}$$

Preuve. Puisque $X + Y = s(X, Y)$, $P_{X+Y} = P_{(X,Y)} \circ s^{-1}$. Cette égalité est vérifiée pour la loi de n'importe quelle somme, que les termes soient indépendants ou non. La propriété supplémentaire apportée par l'indépendance est que $P_{(X,Y)} = P_X \otimes P_Y$. La conclusion en découle puisque par la définition 5.50, $(P_X \otimes P_Y) \circ s^{-1} = P_X * P_Y$. □

Voyons maintenant comment intégrer par rapport à $\mu_1 * \mu_2$.

Proposition 5.54. *La formule*

$$\int_E h \,\mathrm{d}(\mu_1 * \mu_2) = \int_{E^2} h(x+y) \,\mathrm{d}(\mu_1 \otimes \mu_2)(x,y)$$

est vérifiée par

- *toute fonction h mesurable positive $E \to \overline{\mathbb{R}}_+$;*
- *toute $h \in \mathcal{L}^1_{\mathbb{K}}(\mu_1 * \mu_2)$, $\mathbb{K} = \mathbb{R}$ ou $\mathbb{C}$, ce qui équivaut à $h \circ s \in \mathcal{L}^1_{\mathbb{K}}(\mu_1 \otimes \mu_2)$.*

Preuve. Puisque $\mu_1 * \mu_2$ est la mesure image de $\mu_1 \otimes \mu_2$ par s, c'est simplement le théorème de transfert appliqué à $E \times E \xrightarrow{s} E$. □

Même si le produit de convolution de la mesure de Lebesgue avec elle même est sans intérêt, on peut regarder ce que donne la convolution de deux mesures finies à densité par rapport à λ.

Proposition 5.55. *Si μ_1 et μ_2 sont les mesures finies de densités respectives f et g par rapport à la mesure de Lebesgue λ sur $\mathbb{R}^d$, $\mu_1 * \mu_2$ est la mesure à densité h définie λ-p.p. par*

$$h(y) = \int_{\mathbb{R}^d} f(x)g(y-x) \,\mathrm{d}\lambda(x) = \int_{\mathbb{R}^d} f(y-x)g(x) \,\mathrm{d}\lambda(x). \tag{5.69}$$

Preuve. Soit B un borélien quelconque de $\mathbb{R}^d$. L'application de (5.67) nous donne

$$\mu_1 * \mu_2(B) = \int_{\mathbb{R}^d} \mu_2(B-x) f(x) \,\mathrm{d}\lambda(x) = \int_{\mathbb{R}^d} \left\{ \int_{\mathbb{R}^d} \mathbf{1}_{B-x}(u) g(u) \,\mathrm{d}\lambda(u) \right\} f(x) \,\mathrm{d}\lambda(x).$$

Notons que $\mathbf{1}_{B-x}(u) = \mathbf{1}_B(x+u)$. Le changement de variable $y = x + u$ (à x fixé) dans l'intégrale interne s'écrit compte-tenu de l'invariance de λ et de $\mathbb{R}^d$ par translation :

$$\int_{\mathbb{R}^d} \mathbf{1}_{B-x}(u) g(u) \,\mathrm{d}\lambda(u) = \int_{\mathbb{R}^d} \mathbf{1}_B(y) g(y-x) \,\mathrm{d}\lambda(y).$$

En reportant cette expression dans l'intégrale itérée ci-dessus et en permutant l'ordre des intégrations grâce au théorème de Fubini Tonelli, il vient

$$\mu_1 * \mu_2(B) = \int_{\mathbb{R}^d} \mathbf{1}_B(y) \left\{ \int_{\mathbb{R}^d} f(x) g(y-x) \,\mathrm{d}\lambda(x) \right\} \mathrm{d}\lambda(y) = \int_B h(y) \,\mathrm{d}\lambda(y).$$

Nous venons ainsi d'établir que pour tout $B \in \operatorname{Bor}(\mathbb{R}^d)$, $\mu_1 * \mu_2(B) = \int_B h \,\mathrm{d}\lambda$, autrement dit que $\mu_1 * \mu_2$ est la mesure de densité h par rapport à λ. La deuxième expression de h contenue dans la formule (5.69) s'obtient de manière symétrique en permutant les rôles de μ_1 et μ_2 dans le calcul ci-dessus. Notons enfin qu'en prenant $B = \mathbb{R}^d$ dans le calcul ci-dessus, on voit que h est λ-intégrable sur $\mathbb{R}^d$, en particulier, h est finie λ-presque partout. Par ailleurs, toute fonction égale λ-p.p. à h est aussi une densité de $\mu_1 * \mu_2$. □

La proposition 5.55 incite à définir le produit de convolution de deux densités f et g (par rapport à λ de mesures finies) par $f * g = h$ où h est donnée par (5.69). Le pas suivant consiste à remarquer que si f et g sont dans $\mathrm{L}^1_{\mathbb{K}}(\lambda)$, $|f|$ et $|g|$ sont des densités par rapport à λ de mesures finies. Ceci nous conduit à la définition du produit de convolution de deux éléments de $\mathrm{L}^1_{\mathbb{K}}(\lambda)$.

Définition 5.56. *Si f et g sont dans $\mathrm{L}^1_{\mathbb{K}}(\lambda)$, leur produit de convolution est la classe de fonctions dont un représentant est défini λ-p.p. par*

$$(f * g)(y) := \int_{\mathbb{R}^d} f(x)g(y-x)\,\mathrm{d}\lambda(x) = \int_{\mathbb{R}^d} f(y-x)g(x)\,\mathrm{d}\lambda(x).$$

Proposition 5.57. *Si f et g sont dans $\mathrm{L}^1_{\mathbb{K}}(\lambda)$, alors $f * g$ est aussi dans $\mathrm{L}^1_{\mathbb{K}}(\lambda)$ et*

$$\|f * g\|_1 \leq \|f\|_1 \|g\|_1.$$

Preuve. Il suffit d'appliquer le théorème de Fubini Tonelli et l'invariance par translation de la mesure de Lebesgue :

$$\begin{aligned}
\int_{\mathbb{R}^d} \left| \int_{\mathbb{R}^d} f(x)g(y-x)\,\mathrm{d}\lambda(x) \right| \mathrm{d}\lambda(y) &\leq \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} |f(x)||g(y-x)|\,\mathrm{d}\lambda(x)\,\mathrm{d}\lambda(y) \\
&= \int_{\mathbb{R}^d} |f(x)| \left\{ \int_{\mathbb{R}^d} |g(y-x)|\,\mathrm{d}\lambda(y) \right\} \mathrm{d}\lambda(x) \\
&= \int_{\mathbb{R}^d} |f(x)| \left\{ \int_{\mathbb{R}^d} |g(y)|\,\mathrm{d}\lambda(y) \right\} \mathrm{d}\lambda(x) \\
&= \left\{ \int_{\mathbb{R}^d} |g(y)|\,\mathrm{d}\lambda(y) \right\} \int_{\mathbb{R}^d} |f(x)|\,\mathrm{d}\lambda(x) \\
&= \|f\|_1 \|g\|_1.
\end{aligned}$$

□

Chapitre 6

Espaces L^p

Dans tout ce chapitre, $(\Omega, \mathcal{F}, \mu)$ est un espace mesuré.

6.1 Construction des espaces L^p

Définition 6.1. *Soit $p \in [1, +\infty[$ et $\mathbb{K} = \mathbb{R}$ ou $\mathbb{C}$. On note $\mathcal{L}^p_{\mathbb{K}}(\mu)$ l'ensemble des fonctions $f : \Omega \to \mathbb{K}$ telles que*

i) f est mesurable $\mathcal{F}$ - $\mathrm{Bor}(\mathbb{K})$;

ii) $\int_\Omega |f|^p \,\mathrm{d}\mu < +\infty$.

Remarque 6.2. Pour $p \in [1, +\infty[$, la fonction $\varphi : \mathbb{R}_+ \to \mathbb{R}_+$, $x \mapsto x^p$ est *convexe* [1], ce qui implique notamment que pour tous $a, b \geq 0$, $\varphi\big(\frac{a+b}{2}\big) \leq \frac{1}{2}\varphi(a) + \frac{1}{2}\varphi(b)$. Appliquant cette inégalité avec $a = |f(\omega)|$ et $b = |g(\omega)|$, on voit que pour toutes $f, g : \Omega \to \mathbb{K}$,

$$\Big(\frac{|f| + |g|}{2}\Big)^p \leq \frac{1}{2}|f|^p + \frac{1}{2}|g|^p.$$

On en déduit immédiatement que si f et g sont dans $\mathcal{L}^p_{\mathbb{K}}(\mu)$, leur somme y est aussi. La stabilité de $\mathcal{L}^p_{\mathbb{K}}(\mu)$ sous l'effet de la multiplication par un scalaire étant évidente, il en résulte que $\mathcal{L}^p_{\mathbb{K}}(\mu)$ est un *espace vectoriel* de fonctions.

Exemple 6.1. Si $\Omega = \mathbb{N}$, $\mathcal{F} = \mathcal{P}(\mathbb{N})$ et μ est la mesure de comptage sur $\mathbb{N}$, on retrouve l'espace bien connu des suites de puissance p-ième absolument sommable :

$$\mathcal{L}^p_{\mathbb{K}}(\mu) = \ell^p_{\mathbb{K}}(\mathbb{N}) := \Big\{(a_n)_{n\in\mathbb{N}};\ \forall n \in \mathbb{N},\ a_n \in \mathbb{K} \text{ et } \sum_{n=0}^{+\infty} |a_n|^p < +\infty\Big\}.$$

1. La « corde » entre deux points de la courbe est « au dessus » de l'arc de courbe correspondant, ou encore l'image du barycentre est majorée par le barycentre des images, voir la figure 6.1.

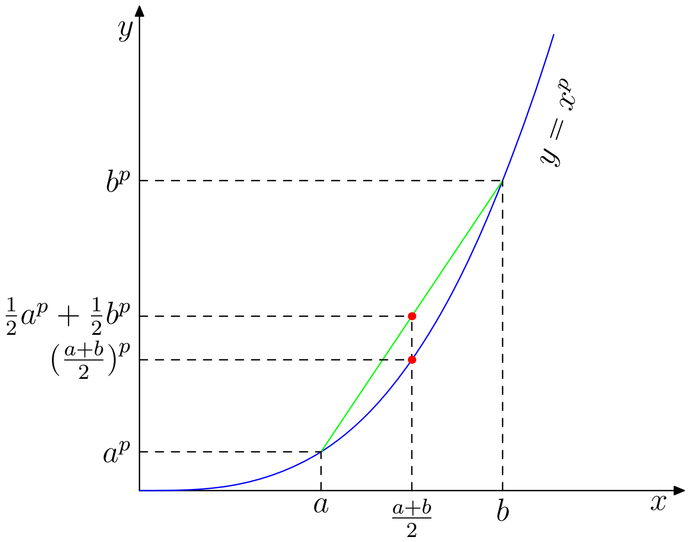


FIGURE 6.1 – Convexité de $x \mapsto x^p$

L'idée qui est derrière l'introduction des espaces $\mathcal{L}^p$ est de fournir une échelle permettant de quantifier le degré de μ-intégrabilité d'une fonction. Intuitivement, le paramètre p sert à amplifier les grandes valeurs de $|f|$ et l'appartenance de f à $\mathcal{L}^p$ signifie que ces grandes valeurs ne pèsent pas trop lourd (relativement à la mesure μ). Il est souhaitable de compléter cette échelle d'espaces ($\mathcal{L}^p$, $1 \leq p < +\infty$) en lui adjoignant le cas limite $p = +\infty$. Ce cas nécessite un traitement particulier.

Définition 6.3. *Si $f : \Omega \to \mathbb{K}$ est mesurable, on définit sa borne supérieure essentielle*

$$\operatorname{ess\,sup} f := \inf\{c \in \overline{\mathbb{R}}_+;\ |f| \leq c \ \ \mu\text{-p.p.}\}.$$

$\mathcal{L}^\infty_{\mathbb{K}}(\mu)$ est l'ensemble des fonctions $f : \Omega \to \mathbb{K}$ mesurables telles que $\operatorname{ess\,sup} f < +\infty$.

Il ne saute pas aux yeux que le supremum essentiel soit la quantité pertinente pour définir le cas limite $p = +\infty$ dans l'échelle des $\mathcal{L}^p$. Ce choix s'explique par le fait que si $f \in \mathcal{L}^r_{\mathbb{K}}(\mu)$ pour au moins un $r \in [1, +\infty[$,

$$\lim_{p\to+\infty} \left\{ \int_\Omega |f|^p \,\mathrm{d}\mu \right\}^{1/p} = \operatorname{ess\,sup} f.$$

Ce résultat pourra être vu en exercice.

La manipulation pratique du supremum essentiel est grandement facilitée par le résultat suivant.

Proposition 6.4.

$$\operatorname{ess\,sup} f = \min\{c \in \overline{\mathbb{R}}_+;\ |f| \leq c \ \ \mu\text{-p.p.}\}.$$

Autrement dit, $M = \operatorname{ess\,sup} f$ si et seulement si $|f| \leq M$ μ-p.p. et pour tout $\varepsilon > 0$, $\mu(|f| > M - \varepsilon) > 0$.

Preuve. Notons $B := \{c \in \overline{\mathbb{R}}_+;\ |f| \leq c\ \mu\text{-p.p.}\}$ et $M := \inf B = \operatorname{ess\,sup} f$. Nous allons vérifier que $M \in B$, donc que $M = \min B$.

Soit $b \in B$, et $c \geq b$, alors μ-p.p., $|f| \leq b \leq c$ donc tout $c \geq b$ est aussi dans B et $B \supset [b, +\infty[$. Par conséquent, $B = \bigcup_{b \in B} [b, +\infty[$ est un intervalle, donc $B =]M, +\infty[$ ou $B = [M, +\infty[$. En tout cas, pour tout $n \geq 1$, $M + 1/n \in B$. Soit $E_n := \{\omega \in \Omega;\ |f(\omega)| \leq M + 1/n\}$. Comme $M + 1/n \in B$, $\mu(E_n^c) = 0$. On a clairement $\{|f| \leq M\} = \bigcap_{n \geq 1} E_n$ d'où par σ-additivité de μ,

$$\mu\Big(\{|f| > M\}\Big) = \mu\Big(\bigcup_{n \geq 1} E_n^c\Big) \leq \sum_{n \geq 1} \mu(E_n^c) = 0.$$

Ainsi $|f| \leq M$ μ-p.p., donc $M \in B$. $\square$

Théorème 6.5 (Inégalité de Hölder)**.** *Soient $p \in]1, +\infty[$ et q tel que $\frac{1}{p} + \frac{1}{q} = 1$. Pour toutes f, g mesurables positives $\Omega \to \overline{\mathbb{R}}_+$,*

$$\int_\Omega fg \,\mathrm{d}\mu \leq \left\{\int_\Omega f^p \,\mathrm{d}\mu\right\}^{1/p} \left\{\int_\Omega g^q \,\mathrm{d}\mu\right\}^{1/q}. \tag{6.1}$$

La démonstration du théorème 6.5 repose sur le lemme de concavité suivant dont la justification sera donnée après la preuve du théorème.

Lemme 6.6. *Si $p \in]1, +\infty[$ et q est tel que $\frac{1}{p} + \frac{1}{q} = 1$, alors*

$$\forall x, y \in \overline{\mathbb{R}}_+, \quad xy \leq \frac{x^p}{p} + \frac{y^q}{q}. \tag{6.2}$$

Preuve du théorème 6.5. Posons

$$A := \left\{\int_\Omega f^p \,\mathrm{d}\mu\right\}^{1/p}, \quad B := \left\{\int_\Omega g^q \,\mathrm{d}\mu\right\}^{1/q}.$$

Traitons d'abord les cas particuliers. Si $A = 0$, alors $f = 0$ μ-p.p. et $fg = 0$ μ-p.p., donc (6.1) est vraie (idem si $B = 0$). Si A ou B vaut $+\infty$ avec $A, B > 0$, alors $AB = +\infty$ et (6.1) est vraie.

On suppose désormais que $0 < A < +\infty$ et $0 < B < +\infty$. Pour tout $\omega \in \Omega$, on peut appliquer le lemme 6.6 avec $x = f(\omega)/A$ et $y = g(\omega)/B$:

$$\forall \omega \in \Omega, \quad \frac{f(\omega)g(\omega)}{AB} \leq \frac{1}{pA^p} f(\omega)^p + \frac{1}{qB^q} g(\omega)^q. \tag{6.3}$$

En intégrant (6.3) sur Ω, il vient :

$$\frac{1}{AB}\int_\Omega fg \,\mathrm{d}\mu \leq \frac{1}{pA^p}\int_\Omega f^p \,\mathrm{d}\mu + \frac{1}{qB^q}\int_\Omega g^q \,\mathrm{d}\mu = \frac{1}{p} + \frac{1}{q} = 1.$$

On en déduit que $\int_\Omega fg \,\mathrm{d}\mu \leq AB$, ce qui établit (6.1). $\square$

Preuve du lemme 6.6. Si $x = 0$ ou $y = 0$, (6.2) est vraie automatiquement puisque son premier membre est nul. Si $x = +\infty$ ou $y = +\infty$, (6.2) est encore vraie puisque son second membre vaut $+\infty$. Supposons désormais que $0 < x, y < +\infty$. Alors en prenant le logarithme de l'égalité $xy = (x^p)^{1/p}(y^q)^{1/q}$, et en utilisant la *concavité* de la fonction logarithme, il vient

$$\ln(xy) = \frac{1}{p}\ln(x^p) + \frac{1}{q}\ln(y^q) \leq \ln\Big(\frac{1}{p}x^p + \frac{1}{q}y^q\Big).$$

Par croissance de la fonction exponentielle, on en déduit l'inégalité (6.2). □

Remarque 6.7 (cas $p = 1$, $q = +\infty$)**.** Si f et g sont mesurables positives,

$$\int_\Omega fg\,\mathrm{d}\mu \leq (\text{ess sup}\, g)\int_\Omega f\,\mathrm{d}\mu. \tag{6.4}$$

En effet, en posant $M := \text{ess sup}\, g$, la proposition 6.4 nous assure que $g \leq M$ μ-p.p., donc que $fg \leq M$ μ-p.p. et en intégrant cette inégalité sur Ω, on obtient (6.4).

Corollaire 6.8. *Soient $f, g : \Omega \to \mathbb{K} = \mathbb{R}$ ou $\mathbb{C}$, mesurables.*

a) Si $1 < p < +\infty$,

$$\int_\Omega |fg|\,\mathrm{d}\mu \leq \Big\{\int_\Omega |f|^p\,\mathrm{d}\mu\Big\}^{1/p}\Big\{\int_\Omega |g|^q\,\mathrm{d}\mu\Big\}^{1/q}. \tag{6.5}$$

Il en résulte que si $f \in \mathcal{L}^p(\mu)$ et $g \in \mathcal{L}^q(\mu)$, alors $fg \in \mathcal{L}^1(\mu)$.

b) Si $p = 1$, $q = +\infty$,

$$\int_\Omega |fg|\,\mathrm{d}\mu \leq (\text{ess sup}\, g)\int_\Omega |f|\,\mathrm{d}\mu. \tag{6.6}$$

Preuve. Il suffit de noter que $|fg| = |f||g|$ et d'appliquer aux fonctions mesurables positives $|f|$ et $|g|$ l'inégalité de Hölder dans le cas $1 < p < +\infty$ et la remarque 6.7 dans le cas $p = 1$, $q = +\infty$. □

Théorème 6.9 (Inégalité de Minkowski)**.** *Pour $1 < p < +\infty$ et toutes f, g mesurables positives $\Omega \to \overline{\mathbb{R}}_+$,*

$$\Big\{\int_\Omega (f+g)^p\,\mathrm{d}\mu\Big\}^{1/p} \leq \Big\{\int_\Omega f^p\,\mathrm{d}\mu\Big\}^{1/p} + \Big\{\int_\Omega g^p\,\mathrm{d}\mu\Big\}^{1/p}. \tag{6.7}$$

Preuve. Soit q l'exposant conjugué de p, *i.e.* l'unique réel tel que $\frac{1}{p} + \frac{1}{q} = 1$. L'inégalité de Hölder nous fournit les majorations :

$$\begin{aligned}
\int_\Omega f(f+g)^{p-1}\,\mathrm{d}\mu &\leq \Big\{\int_\Omega f^p\,\mathrm{d}\mu\Big\}^{1/p}\Big\{\int_\Omega (f+g)^{(p-1)q}\,\mathrm{d}\mu\Big\}^{1/q},\\
\int_\Omega g(f+g)^{p-1}\,\mathrm{d}\mu &\leq \Big\{\int_\Omega g^p\,\mathrm{d}\mu\Big\}^{1/p}\Big\{\int_\Omega (f+g)^{(p-1)q}\,\mathrm{d}\mu\Big\}^{1/q}.
\end{aligned}$$

Par addition membre à membre et en notant que $(p-1)q = p$, on obtient

$$\int_\Omega (f+g)^p \,\mathrm{d}\mu \leq \left[\left\{ \int_\Omega f^p \,\mathrm{d}\mu \right\}^{1/p} + \left\{ \int_\Omega g^p \,\mathrm{d}\mu \right\}^{1/p} \right] \left\{ \int_\Omega (f+g)^p \,\mathrm{d}\mu \right\}^{1/q}. \tag{6.8}$$

Si $0 < \int_\Omega (f+g)^p \,\mathrm{d}\mu < +\infty$, on en déduit (6.7) en divisant par cette intégrale ($1 - \frac{1}{q} = \frac{1}{p}$). Si $\int_\Omega (f+g)^p \,\mathrm{d}\mu = 0$, (6.7) est automatiquement vérifiée. Si $\int_\Omega (f+g)^p \,\mathrm{d}\mu = +\infty$, on utilise l'inégalité de convexité ($p > 1$) :

$$\left(\frac{f+g}{2} \right)^p \leq \frac{1}{2} f^p + \frac{1}{2} g^p,$$

pour voir que nécessairement au moins l'une des intégrales $\int_\Omega f^p \,\mathrm{d}\mu$ ou $\int_\Omega g^p \,\mathrm{d}\mu$ vaut $+\infty$ et qu'ainsi (6.7) est automatiquement vérifiée. □

Proposition 6.10 (semi-normes)**.**

a) Pour $1 \leq p < +\infty$*, on définit une semi-norme sur* $\mathcal{L}^p_{\mathbb{K}}(\mu)$ *en posant*

$$\|f\|_p := \left\{ \int_\Omega |f|^p \,\mathrm{d}\mu \right\}^{1/p}.$$

b) Pour $p = +\infty$*, on définit une semi-norme sur* $\mathcal{L}^\infty_{\mathbb{K}}(\mu)$ *en posant*

$$\|f\|_\infty := \text{ess sup}\, f.$$

Ce ne sont pas des normes car :

$$\|f\|_p = 0 \quad \Leftrightarrow \quad f = 0 \quad \mu\text{-p.p. (i.e. } f \in \mathcal{N}_{\mathbb{K}}(\mu)).$$

Preuve. Pour le a) comme pour le b), l'homogénéité ($\|cf\| = |c|\|f\|$) est immédiate. Vérifions la sous-additivité ($\|f+g\| \leq \|f\| + \|g\|$) dans le cas a). Par croissance sur $\overline{\mathbb{R}}_+$ de la fonction puissance $x \mapsto x^p$, on a $|f+g|^p \leq (|f|+|g|)^p$ d'où $\int_\Omega |f+g|^p \,\mathrm{d}\mu \leq \int_\Omega (|f|+|g|)^p \,\mathrm{d}\mu$ et il ne reste plus qu'à appliquer l'inégalité de Minkowski aux fonctions mesurables positives $|f|$ et $|g|$. Pour la sous-additivité dans le cas b), remarquons que grâce à la proposition 6.4, on a μ-presque partout $|f| \leq \|f\|_\infty$ et $|g| \leq \|g\|_\infty$. Ainsi $|f+g| \leq \|f\|_\infty + \|g\|_\infty$ μ-presque partout et vu la définition du supremum essentiel de $f+g$, on en déduit que $\|f+g\|_\infty \leq \|f\|_\infty + \|g\|_\infty$. □

Rappelons que $\mathcal{N}_{\mathbb{K}}(\mu)$ est l'espace vectoriel des applications $\Omega \to \mathbb{K}$ qui sont μ-négligeables. Il est clair que c'est un s.e.v. de chaque $\mathcal{L}^p_{\mathbb{K}}(\mu)$ pour $1 \leq p \leq +\infty$.

Définition 6.11. *Pour* $1 \leq p \leq +\infty$*, on définit l'espace quotient*

$$\mathrm{L}^p_{\mathbb{K}}(\mu) := \mathcal{L}^p_{\mathbb{K}}(\mu) / \mathcal{N}_{\mathbb{K}}(\mu)$$

des classes de fonctions de puissance p-ième μ*-intégrable (ou* μ*-essentiellement bornées si* $p = +\infty$*) modulo l'égalité* μ-p.p.*. La semi-norme* $\|\ \|_p$ *sur* $\mathcal{L}^p$ *devient une vraie norme sur* L^p *par ce passage au quotient.*

Sauf mention explicite du contraire, l'espace vectoriel $\mathrm{L}^p_{\mathbb{K}}(\mu)$ sera toujours muni de la norme $\| \ \|_p$ définie ci-dessus. En particulier si $1 \leq p < +\infty$, la convergence dans $\mathrm{L}^p(\mu)$ de la suite φ_n vers φ a la traduction suivante :

$$\varphi_n \xrightarrow[n\to+\infty]{\mathrm{L}^p(\mu)} \varphi \quad \Leftrightarrow \quad \lim_{n\to+\infty} \int_\Omega |\varphi_n - \varphi|^p \,\mathrm{d}\mu = 0.$$

Remarque 6.12. Dans le cas particulier où μ est une mesure telle que pour tout $\omega \in \Omega$, $\mu(\{\omega\}) > 0$, la nullité μ-presque partout équivaut à la nullité partout et $\mathcal{N}_{\mathbb{K}}(\mu)$ se réduit à $\{0\}$ (singleton fonction nulle). Alors $\|f\|_p$ et $\|f\|_\infty$ sont déjà des normes sur $\mathcal{L}^p_{\mathbb{K}}(\mu)$ $(1 \leq p \leq +\infty)$ et les espaces $\mathcal{L}^p_{\mathbb{K}}(\mu)$ et $\mathrm{L}^p_{\mathbb{K}}(\mu)$ sont les mêmes[2]. Cette situation se produit notamment pour les les espaces $\ell^p(\mathbb{N})$ et $\ell^\infty(\mathbb{N})$.

Théorème 6.13 (convergence dominée dans L^p, $1 \leq p < +\infty$)**.** *Soit* $(f_n)_{n\in\mathbb{N}}$ *une suite de fonctions mesurables* $\Omega \to \mathbb{K}$. *On suppose que*

a) $f_n \xrightarrow[n\to+\infty]{\mu\text{-p.p.}} f$.

b) Il existe $g \in \mathcal{L}^p(\mu)$ *telle que pour tout* n, $|f_n| \leq g$ μ-p.p..

Alors la classe de f *appartient à* $\mathrm{L}^p_{\mathbb{K}}(\mu)$ *et* f_n *tend vers* f *au sens* L^p, *i.e.* $\|f_n - f\|_p$ *tend vers* 0.

Attention, ce théorème serait grossièrement faux avec $p = +\infty$. Voici un contre-exemple : $\Omega = [0,1]$, $\mathcal{F} = \mathrm{Bor}([0,1])$, $\mu = \lambda$, $f_n = \mathbf{1}_{[1/n,2/n]}$ $(n \geq 2)$ et $f = 0$.

Preuve. Soit $\Omega' := \{\omega \in \Omega;\ f_n(\omega) \to f(\omega)\}$. Par une argumentation déjà vue lorsque nous avons établi le théorème de convergence dominée classique, on sait que $\Omega' \in \mathcal{F}$ et que $\mu(\Omega'^c) = 0$. On définit alors la suite des fonctions mesurables positives

$$h_n := \begin{cases} |f_n - f|^p & \text{sur } \Omega' \\ 0 & \text{sur } \Omega'^c \end{cases}$$

et on obtient la conclusion en appliquant à (h_n) le théorème de convergence dominée classique avec pour fonction dominante $2^p g^p$. □

Théorème 6.14. *Pour* $1 \leq p \leq +\infty$, $\mathrm{L}^p_{\mathbb{K}}(\mu)$ *est un espace vectoriel normé* complet, *c'est donc un* espace de Banach.

Preuve dans le cas $1 \leq p < +\infty$. Soit $(\varphi_n)_{n\in\mathbb{N}}$ une suite de Cauchy dans $\mathrm{L}^p_{\mathbb{K}}(\mu)$. Pour chaque classe d'équivalence φ_n, fixons l'un de ses représentants dans $\mathcal{L}^p_{\mathbb{K}}(\mu)$ et notons le f_n. La propriété de Cauchy se traduit alors par :

$$\forall \varepsilon > 0,\ \exists N(\varepsilon) \in \mathbb{N}^*,\ \forall j,k \geq N(\varepsilon), \quad \|f_j - f_k\|_p < \varepsilon. \tag{6.9}$$

Nous allons extraire une sous-suite[3] $(f_{n_i})_{i\in\mathbb{N}}$ telle que

$$\forall n \geq n_i, \quad \|f_n - f_{n_i}\|_p < 2^{-i}. \tag{6.10}$$

2. Notons aussi que dans ce cas, le supremum essentiel et le supremum coïncident.
3. La suite $(n_i)_{i\in\mathbb{N}}$ d'indices extraits est donc strictement croissante : $\forall i \in \mathbb{N}$, $n_i < n_{i+1}$.

En particulier on a

$$\forall i \in \mathbb{N}, \quad \|f_{n_{i+1}} - f_{n_i}\|_p < 2^{-i}. \tag{6.11}$$

L'extraction de la sous-suite $(f_{n_i})_{i\in\mathbb{N}}$ repose sur (6.9) en prenant (pour $\varepsilon = 2^{-i}$)

$$n_0 := N(1) \quad \text{et pour } i \geq 1, \quad n_i := 1 + \max\Bigl(n_{i-1}, N(2^{-i})\Bigr).$$

Posons désormais pour alléger $g_i := f_{n_i}$. La décomposition en somme téléscopique

$$g_j = g_0 + \sum_{i=1}^{j}(g_i - g_{i-1})$$

nous fournit l'inégalité

$$|g_j| \leq |g_0| + \sum_{i=1}^{j} |g_i - g_{i-1}| =: h_j. \tag{6.12}$$

La suite de fonctions mesurables positives (h_j) ainsi définie est croissante, notons h sa limite. Par croissance et continuité de la fonction $x \mapsto x^p$ sur $\overline{\mathbb{R}}_+$, on a $h_j^p \uparrow h^p$ et par le théorème de Beppo Levi et la continuité sur $\overline{\mathbb{R}}_+$ de la fonction $x \mapsto x^{1/p}$, on en déduit :

$$\|h_j\|_p \uparrow \left\{\int_\Omega |h|^p \,\mathrm{d}\mu\right\}^{1/p} \qquad (j \to +\infty). \tag{6.13}$$

À ce stade, il n'est pas exclu que cette limite de $\|h_j\|_p$ soit $+\infty$.

Par ailleurs la définition de h_j et l'inégalité de Minkowski nous procurent la majoration :

$$\begin{aligned}
\|h_j\|_p &\leq \|g_0\|_p + \sum_{i=1}^{j} \|g_i - g_{i-1}\|_p \\
&\leq \|g_0\|_p + \sum_{i=1}^{j} 2^{-i} \\
&\leq \|g_0\|_p + \sum_{i=1}^{+\infty} 2^{-i} = \|g_0\|_p + 1.
\end{aligned}$$

Ce majorant de $\|h_j\|_p$ étant valable pour tout j, on en déduit en faisant tendre j vers $+\infty$ et en tenant compte de (6.13) que

$$\left\{\int_\Omega |h|^p \,\mathrm{d}\mu\right\}^{1/p} \leq \|g_0\|_p + 1 < +\infty.$$

Il en résulte que h^p est μ-intégrable, donc que h est finie μ-presque partout sur Ω. Comme $h_j \leq h$, on en déduit que la série de sommes partielles $g_j = g_0 + \sum_{i=1}^{j}(g_i - g_{i-1})$ est *absolument* convergente μ-p.p. sur Ω. Notons g sa somme [4]. La suite (g_j) converge

4. A priori définie seulement μ-p.p., mais il suffit de prendre g nulle sur l'ensemble de mesure nulle où la série n'est pas absolument convergente pour que g soit mesurable et limite μ-p.p. de g_j.

μ-p.p. vers g et est dominée μ-p.p. par $h \in \mathrm{L}^p$ puisque $|g_j| \leq h_j \leq h$. Le théorème de convergence dominée dans L^p nous donne alors :

$$\|g_j - g\|_p \xrightarrow[j\to+\infty]{} 0.$$

On en déduit en procédant comme suit, que

$$\|f_n - g\|_p \xrightarrow[j\to+\infty]{} 0.$$

Soit $\varepsilon > 0$ fixé. Il existe i_0 tel que $2^{-i_0} < \varepsilon$. Comme g_j tend vers g dans L^p, il existe un $i_1 \geq i_0$ tel que $\|g_{i_1} - g\|_p < \varepsilon$. Alors par (6.10), nous avons

$$\forall n \geq n_{i_1}, \quad \|f_n - f_{n_{i_1}}\|_p = \|f_n - g_{i_1}\|_p < 2^{-i_1} \leq 2^{-i_0} < \varepsilon.$$

Ainsi pour tout $\varepsilon > 0$, on peut trouver $n_{i_1}(\varepsilon)$ tel que pour tout $n \geq n_{i_1}$, $\|f_n - g\|_p < 2\varepsilon$, autrement dit f_n converge vers g au sens L^p. En notant $\varphi \in \mathrm{L}^p_{\mathbb{K}}(\mu)$ la classe d'équivalence de $g \in \mathcal{L}^p_{\mathbb{K}}(\mu)$, on a donc la convergence vers zéro de $\|\varphi_n - \varphi\|_p$, autrement dit la convergence de φ_n vers φ dans l'espace $\mathrm{L}^p_{\mathbb{K}}(\mu)$.

Partis d'une suite de Cauchy quelconque dans $\mathrm{L}^p(\mu)$, nous avons montré qu'elle converge dans cet espace. La complétude de $\mathrm{L}^p(\mu)$ est établie. □

Corollaire 6.15. *Si pour un $p \in [1, +\infty[$, la suite (f_n) converge vers f au sens $\mathrm{L}^p(\mu)$, on peut en extraire une sous-suite (f_{n_i}) qui converge μ-p.p. vers f.*

Écrit avec l'abus de langage traditionnel, cet énoncé signifie bien sûr qu'après avoir choisi dans chaque classe f_n un représentant (une vraie fonction notée encore f_n et appartenant à $\mathcal{L}^p(\mu)$), on peut extraire de cette suite de vraies fonctions une sous-suite qui converge μ-p.p. vers un représentant de f.

Preuve. La suite (f_n) converge dans l'espace métrique L^p donc est de Cauchy. La sous-suite (de vraies fonctions) (f_{n_i}) construite dans la preuve du théorème 6.14 converge μ-p.p. et au sens L^p vers une certaine fonction g qui est aussi dans $\mathcal{L}^p$. Pour vérifier que $f = g$ (μ-p.p.), il suffit d'écrire

$$\|f - g\|_p \leq \|f - f_{n_i}\|_p + \|f_{n_i} - g\|_p.$$

Quand i tend vers l'infini, ce majorant tend vers 0, d'où $\|f - g\|_p = 0$. □

Preuve du théorème 6.14 dans le cas $p = +\infty$. Soit (φ_n) une suite de Cauchy dans $\mathrm{L}^\infty_{\mathbb{K}}(\mu)$. Choisissons dans chaque classe φ_n un représentant noté f_n (une vraie fonction appartenant à $\mathcal{L}^\infty_{\mathbb{K}}(\mu)$). On a donc

$$\forall \varepsilon > 0,\ \exists N(\varepsilon),\ \forall k, l \geq N(\varepsilon), \quad \operatorname{ess\,sup}(f_k - f_l) < \varepsilon.$$

Grâce à la proposition 6.4, nous en déduisons que

$$\forall \varepsilon > 0,\ \exists N(\varepsilon),\ \forall k, l \geq N(\varepsilon), \quad |f_k - f_l| < \varepsilon \quad \mu\text{-p.p.} \tag{6.14}$$

Pour éviter la confrontation avec une réunion non dénombrable d'ensembles de mesure nulle, discrétisons le ε en prenant disons $\varepsilon_i = 2^{-i}$, $i \in \mathbb{N}$. Posons $N_i := N(\varepsilon_i)$ et

$$A_{i,k,l} := \Big\{\omega \in \Omega;\ |f_k(\omega) - f_l(\omega)| < \varepsilon\Big\}, \quad A_i := \bigcap_{k,l \geq N_i} A_{i,k,l}, \quad \Omega' := \bigcap_{i \in \mathbb{N}} A_i.$$

Par (6.14), on a $\mu(A_{i,k,l}^c) = 0$ pour tout $i \in \mathbb{N}$ et tous $k, l \geq N_i$. On en déduit successivement par sous-σ-additivité de μ que $\mu(A_i^c) = 0$ et $\mu(\Omega'^c) = 0$. Nous avons donc maintenant :

$$\forall \omega \in \Omega',\ \forall i \in \mathbb{N},\ \forall k, l \geq N_i, \quad |f_k(\omega) - f_l(\omega)| < \varepsilon_i. \tag{6.15}$$

Autrement dit, pour tout $\omega \in \Omega'$, la suite $\big(f_n(\omega)\big)$ est de Cauchy. Comme $\mathbb{K}$ est complet, cette suite converge dans $\mathbb{K}$. Posant alors

$$f(\omega) := \begin{cases} \lim\limits_{n \to +\infty} f_n(\omega) & \text{si } \omega \in \Omega', \\ 0 & \text{si } \omega \notin \Omega', \end{cases}$$

on définit une fonction mesurable $f : \Omega \to \mathbb{K}$ et f_n converge μ presque partout vers f. Cette fonction f appartient à $\mathcal{L}^\infty_{\mathbb{K}}(\mu)$. En effet elle est nulle en dehors de Ω' et pour $\omega \in \Omega'$, on a pour tout $n \geq N_0$, $|f_n(\omega) - f_{N_0}(\omega)| < \varepsilon_0$, d'où $|f_n(\omega)| < |f_{N_0}(\omega)| + \varepsilon_0$ et en faisant tendre n vers l'infini, $|f(\omega)| \leq |f_{N_0}(\omega)| + \varepsilon_0$. Comme $f_{N_0} \in \mathcal{L}^\infty_{\mathbb{K}}(\mu)$, en posant $\Omega'' := \Omega' \cap \{\omega \in \Omega;\ |f_{N_0}(\omega)| \leq \operatorname{ess\,sup} f_{N_0}\}$, on a $\mu(\Omega''^c) = 0$ et pour tout $\omega \in \Omega''$, $|f(\omega)| \leq \operatorname{ess\,sup} f_{N_0} + \varepsilon_0$ et ce majorant fini ne dépend pas de ω.

Cette fonction f est notre candidate pour être la limite au sens L^∞ de la suite de Cauchy (f_n). Pour vérifier cette convergence L^∞, on déduit de (6.15) en faisant tendre l vers $+\infty$ que

$$\forall \omega \in \Omega',\ \forall i \in \mathbb{N},\ \forall k \geq N_i, \quad |f_k(\omega) - f(\omega)| \leq \varepsilon_i. \tag{6.16}$$

Comme $\mu(\Omega'^c) = 0$, l'inégalité $|f_k(\omega) - f(\omega)| \leq \varepsilon_i$ vraie sur Ω' implique $\operatorname{ess\,sup}(f_k - f) \leq \varepsilon_i$. Ainsi la convergence L^∞ de f_n vers f résulte de (6.16). En notant $\varphi \in \mathrm{L}^\infty_{\mathbb{K}}(\mu)$ la classe de $f \in \mathcal{L}^\infty_{\mathbb{K}}(\mu)$, on a ainsi établi la convergence vers φ dans l'espace $\mathrm{L}^\infty_{\mathbb{K}}(\mu)$ de la suite de Cauchy (φ_n). Comme la suite de Cauchy (φ_n) était quelconque, l'espace $\mathrm{L}^\infty_{\mathbb{K}}(\mu)$ est complet. □

Remarque 6.16. Le corollaire 6.15 a été énoncé avec $1 \leq p < +\infty$. Pour $p = +\infty$ on a un meilleur résultat. Si f_n converge vers f au sens L^∞, c'est *toute la suite* (f_n) (et non pas seulement une sous-suite) qui converge μ-p.p. vers f. On peut voir ce résultat comme un sous-produit de la preuve du cas $p = +\infty$ ci-dessus, ou le démontrer directement (en exercice) en s'inspirant de la technique de discrétisation du ε utilisée ci-dessus.

6.2 Comparaison des L^p

Il n'y a pas de relations générales d'inclusion entre les $\mathrm{L}^p(\mu)$, indépendantes de l'espace mesuré $(\Omega, \mathcal{F}, \mu)$. Pour s'en convaincre, prenons pour espace mesuré $(\mathbb{R}, \mathrm{Bor}(\mathbb{R}), \lambda)$ et comparons les espaces $\mathrm{L}^1(\lambda)$ et $\mathrm{L}^2(\lambda)$. Soient les fonctions

$$f(x) = \frac{1}{\sqrt{x}} \mathbf{1}_{]0,1]}(x) \quad \text{et} \quad g(x) = \frac{1}{x} \mathbf{1}_{[1,+\infty[}(x).$$

On voit immédiatement que f appartient à $\mathrm{L}^1(\lambda)$, mais pas à $\mathrm{L}^2(\lambda)$ tandis que g appartient à $\mathrm{L}^2(\lambda)$ mais pas à $\mathrm{L}^1(\lambda)$. Ainsi, aucun des deux espaces $\mathrm{L}^1(\lambda)$ et $\mathrm{L}^2(\lambda)$ n'est inclus dans l'autre.

Dans le cas d'une mesure *finie* μ, le paysage est beaucoup plus agréable. Les espaces $\mathrm{L}^p(\mu)$ sont alors emboîtés dans l'ordre inverse de leurs exposants et, ce qui est bien plus important, on a des inclusions *topologiques*.

Définition 6.17. *Soient* $(E, \| \ \|_E)$ *et* $(F, \| \ \|_F)$ *deux espaces vectoriels normés. On dit que* E *est inclus* topologiquement *dans* F *(notation* $E \hookrightarrow F$*) si*

a) $E \subset F$ *;*

b) l'injection canonique $J : E \to F$, $x \mapsto x$ *est continue.*

La condition b) équivaut à l'existence d'une constante $C < +\infty$ *telle que*

$$\forall x \in E, \quad \|x\|_F \leq C\|x\|_E.$$

Une conséquence pratique de l'inclusion topologique de E dans F est que si une suite x_n converge vers x dans E (i.e. $\|x_n - x\|_E \to 0$), alors elle converge aussi dans F vers la même limite (i.e. $\|x_n - x\|_F \to 0$).

Théorème 6.18. *Si* μ *est une mesure finie, on a pour* $1 \leq p \leq r \leq +\infty$ *les inclusions topologiques*

$$\mathrm{L}^\infty(\mu) \hookrightarrow \mathrm{L}^r(\mu) \hookrightarrow \mathrm{L}^p(\mu) \hookrightarrow \mathrm{L}^1(\mu).$$

Preuve. Il suffit de montrer les deux premières inclusions topologiques, la troisième n'étant qu'un cas particulier de la seconde.

Pour l'inclusion $\mathrm{L}^\infty(\mu) \hookrightarrow \mathrm{L}^r(\mu)$, soit $f \in \mathcal{L}^\infty(\mu)$ et $M := \operatorname{ess\,sup} f$. Alors $|f| \leq M$ μ-p.p. d'où

$$\forall r \in [1, +\infty[, \quad \left\{\int_\Omega |f|^r \,\mathrm{d}\mu\right\}^{1/r} \leq \left\{\int_\Omega M^r \,\mathrm{d}\mu\right\}^{1/r} = M\mu(\Omega).$$

En passant aux classes d'équivalence modulo l'égalité μ-p.p., on en déduit que

$$\forall f \in \mathrm{L}^\infty(\mu), \quad \|f\|_r \leq \mu(\Omega)\|f\|_\infty.$$

Comme $\mu(\Omega) < +\infty$, ceci établit l'inclusion topologique de $\mathrm{L}^\infty(\mu)$ dans $\mathrm{L}^r(\mu)$.

Pour la deuxième inclusion, fixons $1 \leq p \leq r < +\infty$ et soit f mesurable $\Omega \to \mathbb{K}$. Notons a et b deux exposants conjugués ($\frac{1}{a} + \frac{1}{b} = 1$) à préciser ultérieurement. L'inégalité de Hölder appliquée au produit FG avec $F = |f|^p$ et $G = 1$ nous donne :

$$\int_\Omega |f|^p \,\mathrm{d}\mu \leq \left\{\int_\Omega |f|^{pa} \,\mathrm{d}\mu\right\}^{1/a} \left\{\int_\Omega 1^b \,\mathrm{d}\mu\right\}^{1/b} = \mu(\Omega)^{1/b} \left\{\int_\Omega |f|^{pa} \,\mathrm{d}\mu\right\}^{1/a}.$$

Cette inégalité est vraie pour tout $a \geq 1$. Choisissant a de sorte que $pa = r$, on en déduit par croissance sur $\overline{\mathbb{R}}_+$ de $x \mapsto x^{1/p}$,

$$\left\{\int_\Omega |f|^p \,\mathrm{d}\mu\right\}^{1/p} \leq \mu(\Omega)^{1/(pb)} \left\{\int_\Omega |f|^r \,\mathrm{d}\mu\right\}^{1/r}. \tag{6.17}$$

On peut se débarasser de l'exposant b en notant que $\frac{1}{pa}+\frac{1}{pb}=\frac{1}{p}$ d'où $\frac{1}{pb}=\frac{1}{p}-\frac{1}{pa}=\frac{1}{p}-\frac{1}{r}$. Prenant maintenant f quelconque dans $\mathcal{L}^r(\mu)$ et passant aux classes d'équivalence pour l'égalité μ-p.p., on déduit de (6.17) que :

$$\forall f \in \mathrm{L}^r(\mu), \quad \|f\|_p \leq \mu(\Omega)^{1/p-1/r}\|f\|_r,$$

ce qui établit l'inclusion topologique de $\mathrm{L}^r(\mu)$ dans $\mathrm{L}^p(\mu)$. $\square$

6.3 Théorèmes de densité

Théorème 6.19. *Notons $\mathcal{E}$ l'espace des fonctions mesurables étagées $h : \Omega \to \mathbb{K}$ telles que*

$$\mu(\{\omega \in \Omega;\ h(\omega) \neq 0\}) < +\infty.$$

Pour $1 \leq p < +\infty$, $\mathcal{E}$ est dense dans $\mathrm{L}^p(\mu)$.

L'écriture de cet énoncé sacrifie une fois de plus à l'abus de langage traditionnel. Il faut bien sûr comprendre la conclusion comme « l'ensemble des classes des éléments de $\mathcal{E}$ est dense dans $\mathrm{L}^p(\mu)$ ».

Preuve. D'abord il est évident que $\mathcal{E}$ est un s.e.v. de $\mathcal{L}^p_{\mathbb{K}}(\mu)$. Soit $\varphi \geq 0$ dans $\mathrm{L}^p(\mu)$ et f un représentant de φ. On a donc $f \geq 0$, μ-p.p.. Quitte à modifier f sur un ensemble de mesure nulle, donc sans changer sa classe φ, on peut supposer $f \geq 0$ partout. On sait qu'il existe alors une suite croissante h_n de fonctions étagées mesurables positives qui converge partout vers f sur Ω. Les inégalités $0 \leq h_n \leq f$ impliquent l'appartenance des h_n à $\mathcal{L}^p_{\mathbb{K}}(\mu)$. Si $m_n > 0$ est la plus petite valeur non nulle prise par la fonction *étagée* h_n, l'inégalité de Markov nous permet d'écrire :

$$\begin{aligned}\mu(\{\omega \in \Omega;\ h_n(\omega) \neq 0\}) = \mu(\{\omega \in \Omega;\ h_n(\omega) \geq m_n\}) &= \mu(\{\omega \in \Omega;\ h_n(\omega)^p \geq m_n^p\}) \\ &\leq \frac{1}{m_n^p}\int_\Omega h_n^p \,\mathrm{d}\mu \\ &\leq \frac{1}{m_n^p}\int_\Omega f^p \,\mathrm{d}\mu < +\infty.\end{aligned}$$

On voit ainsi que les h_n sont dans $\mathcal{E}$. D'autre part le théorème de convergence dominée dans L^p nous donne la convergence de h_n vers f au sens L^p :

$$\lim_{n\to+\infty} \|f - h_n\|_p = 0.$$

Le cas f à valeurs réelles se déduit du cas $f \geq 0$ par décomposition $f = f^+ - f^-$. Le cas complexe se ramène au cas réel par séparation des parties réelle et imaginaire. $\square$

Corollaire 6.20. *Pour $1 \leq p < +\infty$ et $1 \leq r < +\infty$, l'espace $\mathrm{L}^p(\mu) \cap \mathrm{L}^r(\mu)$ est dense dans chacun des espaces $\mathrm{L}^p(\mu)$ et $\mathrm{L}^r(\mu)$ pour la topologie correspondante.*

Preuve. Comme $\mathcal{E}$ est un sous-espace de tous les L^p pour $1 \leq p < +\infty$, c'est aussi un sous-espace de $\mathrm{L}^p(\mu) \cap \mathrm{L}^r(\mu)$. Il suffit alors d'appliquer le théorème 6.19 dans L^p puis dans L^r. $\square$

Nous discutons maintenant la densité d'espaces de fonctions continues dans les L^p. Bien sûr, cela suppose Ω muni d'une topologie. Nous nous limiterons au cas des espaces métriques.

Théorème 6.21. *Soit E un espace métrique et μ une mesure* finie *sur $(E, \mathrm{Bor}(E))$. On note $\mathcal{C}_b(E)$ l'espace des fonctions continues bornées $E \to \mathbb{K}$. Pour $1 \leq p < +\infty$, $\mathcal{C}_b(E)$ est dense dans $\mathrm{L}^p_{\mathbb{K}}(\mu)$.*

Preuve. La mesure μ étant finie, toute fonction mesurable bornée est dans $\mathcal{L}^p_{\mathbb{K}}(\mu)$. On a donc l'inclusion ensembliste $\mathcal{C}_b(E) \subset \mathcal{L}^p(\mu)$. Comme dans la preuve du théorème 6.19, il suffit de montrer que toute fonction positive f de $\mathcal{L}^p_{\mathbb{K}}(\mu)$ peut être approchée à ε près en distance L^p par une fonction continue bornée. En raison du théorème 6.19, il suffit de le prouver lorsque f est étagée positive. Par combinaison linéaire finie et inégalité triangulaire, la réduction ultime du problème est l'approximation en distance L^p de $f = \mathbf{1}_A$ pour A borélien par une fonction continue. Rappelons un résultat déjà démontré en exercice (cf. Annales I.F.P 2001-2002, D.M. n$^{\mathrm{o}}$ 2).

Lemme 6.22. *Si E est un espace métrique, toute mesure* finie *μ sur $(E, \mathrm{Bor}(E))$ est* régulière*, ce qui signifie que pour tout borélien A,*

$$\begin{aligned} \mu(A) &= \sup\{\mu(F);\ F \text{ fermé} \subset A\}; \\ \mu(A) &= \inf\{\mu(V);\ V \text{ ouvert} \supset A\}. \end{aligned}$$

Par le lemme 6.22, pour ε positif fixé, il existe un fermé F et un ouvert V tels que $F \subset A \subset V$ et $\mu(A) - \varepsilon^p/2 \leq \mu(F) \leq \mu(V) \leq \mu(A) + \varepsilon^p/2$. Comme μ est finie, on en déduit $\mu(V \setminus F) = \mu(V) - \mu(F) \leq \varepsilon^p$. Admettons provisoirement l'existence d'une fonction continue g telle que $0 \leq g \leq 1$, nulle en dehors de V et de valeur constante 1 sur F. Ce point fait l'objet du lemme 6.23 ci-dessous (appliqué avec le fermé $H = E \setminus V$). Ainsi, $f = g = 1$ sur F, $0 \leq g \leq f = 1$ sur $A \setminus F$, $0 = f \leq g \leq 1$ sur $V \setminus A$ et $f = g = 0$ sur V^c (dans le cas particulier où $E = V$, cette dernière condition est sans objet et il suffit de prendre pour g la fonction constante 1 sur E). On en déduit

$$\int_\Omega |f - g|^p \,\mathrm{d}\mu = \int_{V \setminus F} |f - g|^p \,\mathrm{d}\mu \leq \int_{V \setminus F} 1^p \,\mathrm{d}\mu = \mu(V \setminus F) \leq \varepsilon^p,$$

d'où $\|f - g\|_p \leq \varepsilon$. □

Lemme 6.23. *Soient F et H deux fermés disjoints non vides d'un espace métrique E. Il existe une fonction continue $g : E \to \mathbb{R}_+$ telle que $0 \leq g \leq 1$, $g = 1$ sur F et $g = 0$ sur H.*

Preuve. Rappelons la définition de la distance d'un point x à un fermé F non vide dans un espace métrique (E, d) :

$$d(x, F) := \inf_{y \in F} d(x, y).$$

Il est bien connu que $d(\,.\,, F)$ est continue et que $d(x, F) = 0$ si et seulement si [5] $x \in F$.

5. Quand F est *fermé*. Dans le cas F quelconque, $d(x, F) = 0$ équivaut à $x \in \overline{F}$.

Définissons alors g par

$$g(x) := \frac{d(x,H)}{d(x,H)+d(x,F)}, \quad x \in E.$$

Le dénominateur ne peut être nul que si $d(x,F) = d(x,H) = 0$. Or F et H étant fermés, cette égalité implique $x \in F \cap H$, ce qui est impossible puisque F et H sont disjoints. La fonction g apparaît ainsi comme le quotient de deux fonctions continues avec dénominateur ne s'annulant en aucun point de E. Elle est donc continue sur E. On vérifie immédiatement que $0 \leq g \leq 1$, et que si $x \in F$, $g(x) = 1$, si $x \in H$, $g(x) = 0$. $\square$

Notre prochain théorème de densité concerne l'approximation en distance L^p par des fonctions continues à support compact. Rappelons que si f est une fonction de l'espace topologique E dans $\mathbb{R}$ ou $\mathbb{C}$, le support de f est la fermeture de l'ensemble $\{x \in E;\ f(x) \neq 0\}$. Ainsi une fonction à support compact K est nulle en tout point de $E \setminus K$... s'il en existe! En particulier si E est lui même compact, toute fonction sur E est à support compact.

Théorème 6.24. *Supposons que l'espace métrique E est* localement compact *et réunion d'une suite croissante de compacts. Notons $\mathcal{C}_c(E)$ l'espace des fonctions continues $E \to \mathbb{K}$, à support compact. Soit μ une mesure finie sur $(E, \mathrm{Bor}(E))$. Pour $1 \leq p < +\infty$, $\mathcal{C}_c(E)$ est dense dans $\mathrm{L}^p_{\mathbb{K}}(\mu)$.*

Preuve. Par hypothèse, il existe une suite K_n de compacts croissante pour l'inclusion et de réunion E. On effectue la même réduction du problème que dans la preuve du théorème 6.21 en se ramenant au cas $f = \mathbf{1}_A$, pour A borélien quelconque. On remarque alors que si $A_n := A \cap K_n$, $A_n \uparrow A$ et donc par continuité séquentielle croissante de μ, $\mu(A_n) \uparrow \mu(A)$. Comme

$$\|\mathbf{1}_A - \mathbf{1}_{A_n}\|_p = \mu(A \setminus A_n)^{1/p} = \Big(\mu(A) - \mu(A_n)\Big)^{1/p} \xrightarrow[n\to+\infty]{} 0,$$

il suffit de montrer l'approximation à ε près en distance L^p de $\mathbf{1}_{A_n}$ par une fonction continue à support compact. En procédant comme dans la preuve du théorème 6.21, on trouve F fermé et V ouverts tels que $F \subset A_n \subset V$ et $\mu(V \setminus F) \leq \varepsilon^p$. Notons que F est maintenant compact comme fermé dans le compact K_n. Par contre on ne peut rien dire de l'éventuelle compacité de $\overline{V}$. Et c'est bien ennuyeux car si on définit g comme dans la preuve du théorème 6.21, on ne pourra garantir qu'elle soit à support compact. On se tire de ce mauvais pas grâce au lemme suivant dont la preuve est légèrement différée.

Lemme 6.25. *Soit E un espace métrique localement compact, K un compact de E et V un ouvert contenant K. Alors il existe W ouvert de fermeture compacte tel que $K \subset W \subset \overline{W} \subset V$.*

Appliquons ce lemme avec le compact $K = F$ et l'ouvert V. On obtient un ouvert W de fermeture compacte tel que $F \subset W \subset \overline{W} \subset V$. Nous ne prétendons pas que A_n est inclus dans W. Écartant provisoirement le cas particulier $W^c = \emptyset$, on définit alors

$$g(x) := \frac{d(x,W^c)}{d(x,W^c)+d(x,F)}, \quad x \in E.$$

C'est bien une fonction continue (car les fermés F et W^c sont disjoints), nulle en dehors de W, donc à support inclus dans le compact $\overline{W}$. On a $g = 1 = f$ sur F et $g = f = 0$ sur V^c. Sur $V \setminus F$, f et g sont encadrées par 0 et 1 donc $|f - g| \leq 1$. On en déduit que $\|f - g\|_p \leq \mu(V \setminus F)^{1/p} \leq \varepsilon$.

Dans le cas particulier où $W^c = \emptyset$, on a $\overline{W} = E$, donc E est compact et la fonction constante $g = 1$ sur E est à support compact. On a aussi $\|f - g\|_p \leq \mu(V \setminus F)^{1/p} \leq \varepsilon$. □

Preuve du lemme 6.25. Traitons d'abord le cas où $V \neq E$ et donc $V^c \neq \emptyset$. Posons

$$\delta := \inf_{x \in K} d(x, V^c).$$

Par compacité de K et continuité de $d(\,.\,, V^c)$, cet infimum est un minimum : il existe $x_0 \in K$ tel que $\delta = d(x_0, V^c)$. Les fermés K et V^c étant disjoints, $\delta = d(x_0, V^c) > 0$.

Par locale compacité de E, tout $x \in E$ possède un voisinage ouvert U_x de fermeture compacte. Définissons alors les ouverts

$$W_x := B(x, \delta/2) \cap U_x, \quad x \in K.$$

Ainsi W_x est un voisinage ouvert de x, de fermeture compacte puisque $\overline{W_x}$ est fermé dans le compact $\overline{U_x}$. Le compact K est recouvert par la famille d'ouverts $\{W_x,\, x \in K\}$, dont on peut extraire un recouvrement fini :

$$K \subset W_{x_1} \cup \cdots \cup W_{x_m} =: W.$$

W est un ouvert comme union finie d'ouverts. $\overline{W}$ est compact car fermé et inclus dans $\overline{W_{x_1}} \cup \cdots \cup \overline{W_{x_m}}$ qui est compact comme union finie de compacts. Il reste à vérifier l'inclusion de $\overline{W}$ dans V. Prenons y quelconque dans $\overline{W}$. L'inclusion

$$\overline{W} \subset \bigcup_{i=1}^{m} \overline{B(x_i, \delta/2)}$$

nous fournit un x_i tel que $d(x_i, y) \leq \delta/2$. Pour tout z de V^c, l'inégalité triangulaire nous donne alors la minoration

$$d(y, z) \geq d(x_i, z) - d(x_i, y) \geq \delta - \delta/2 = \delta/2.$$

En prenant la borne inférieure pour z décrivant V^c, on en déduit $d(y, V^c) \geq \delta/2 > 0$. Ainsi y n'appartient pas à V^c donc il est dans V. Le raisonnement étant valable pour y quelconque dans $\overline{W}$, l'inclusion $\overline{W} \subset V$ est établie.

Pour compléter la preuve du lemme, il reste à examiner le cas où $V^c = \emptyset$ et donc $V = E$. Il suffit alors de modifier la définition de W_x en $W_x := U_x$, l'inclusion $\overline{W} \subset V$ étant maintenant triviale. □

Théorème 6.26. *Soit λ la mesure de Lebesgue sur $\mathbb{R}^d$. Pour $1 \leq p < +\infty$, l'espace $\mathcal{C}_c(\mathbb{R}^d)$ des fonctions continues $\mathbb{R}^d \to \mathbb{K}$, à support compact est dense dans $\mathrm{L}^p_{\mathbb{K}}(\lambda)$.*

Preuve. Grâce au théorème 6.19, il suffit de montrer que si A est un borélien tel que $\lambda(A) < +\infty$, on peut approcher $\mathbf{1}_A$ en distance L^p à ε près par une fonction g continue à support compact. On peut encore réduire la problème en se ramenant au cas où A est un borélien *borné* [6]. En effet, posons $A_n = A \cap [-n, n]^d$. Par continuité séquentielle croissante de λ, $\lambda(A_n) \uparrow \lambda(A)$. Comme $\lambda(A)$ est *fini*, on a pour n assez grand, $0 \leq \lambda(A) - \lambda(A_n) \leq \varepsilon^p$ et $\lambda(A \setminus A_n) \leq \varepsilon^p$ d'où $\|\mathbf{1}_A - \mathbf{1}_{A_n}\|_p \leq \varepsilon$.

Dans la suite, nous supposons donc A borné, disons

$$A \subset C := [-a, a]^d.$$

Prenons pour espace métrique $E =]-(a+1), a+1[$, muni de la distance euclidienne (ou de n'importe quelle distance associée à une norme sur $\mathbb{R}^d$) et pour μ la (restriction à la tribu borélienne de E de la) mesure à densité 1_C par rapport à λ. Par régularité de la mesure finie μ, on peut trouver F fermé et V ouvert de E tels que $F \subset A \subset V$ et $\mu(V \setminus F) \leq \varepsilon^p$. Notons au passage que comme E est ouvert de $\mathbb{R}^d$, V est aussi ouvert dans $\mathbb{R}^d$. De plus comme $\mu(E \setminus C) = 0$, on peut toujours remplacer V par $V \cap]-a-\delta, a+\delta[^d$ pour $0 < \delta < 1$. Dans la suite nous supposerons donc $V \subset]-a-\delta, a+\delta[^d$ pour un δ qui sera précisé ultérieurement. Ce remplacement n'a pas affecté l'inclusion $A \subset V$ puisque A est inclus dans C et que les points éventuellement éliminés sont tous hors de C. Nous avons donc maintenant :

$$F \subset A \subset V \subset]-a-\delta, a+\delta[^d.$$

En reprenant la construction de g dans la preuve du théorème 6.24, on obtient une fonction continue $g : E \to [0, 1]$ à support compact inclus dans V et telle que $\int_E |\mathbf{1}_A - g|^p \, \mathrm{d}\mu \leq \varepsilon^p$. Remarquons que g est nulle sur $\Delta := \{x \in \mathbb{R}^d ;\ a + \delta < |x_i| \leq a + 1,\ 1 \leq i \leq d\}$. Ainsi il est clair qu'en posant $g(x) := 0$ pour $x \in \mathbb{R}^d \setminus E$, on a un prolongement continu de g à tout $\mathbb{R}^d$, sans modifier le support.

En notant $C_\delta :=]-a-\delta, a+\delta[^d$, nous avons alors

$$\begin{aligned}
\int_{\mathbb{R}^d} |\mathbf{1}_A - g|^p \, \mathrm{d}\lambda &= \int_C |\mathbf{1}_A - g|^p \, \mathrm{d}\lambda + \int_{C_\delta \setminus C} |\mathbf{1}_A - g|^p \, \mathrm{d}\lambda \\
&= \int_E |\mathbf{1}_A - g|^p \, \mathrm{d}\mu + \int_{C_\delta \setminus C} |\mathbf{1}_A - g|^p \, \mathrm{d}\lambda \\
&\leq \varepsilon^p + \lambda(C_\delta \setminus C).
\end{aligned}$$

On voit immédiatement que $\lambda(C_\delta \setminus C) = (2a + 2\delta)^d - (2a)^d$ peut être rendu inférieur à ε^p en choisissant δ suffisamment petit. Ce choix étant fait, nous avons donc établi l'existence de g continue sur $\mathbb{R}^d$ et à support compact telle que $\|\mathbf{1}_A - g\|_p \leq 2^{1/p}\varepsilon$, ce qui achève la démonstration. $\square$

6. Tout borélien borné a une mesure de Lebesgue finie, mais la réciproque est fausse : $\mathbb{Q}$ est de mesure nulle mais pas borné.

6.4 Dualité

Le dual *topologique* d'un espace vectoriel normé est l'espace des formes linéaires *continues* sur cet espace. En dimension infinie, une forme linéaire n'est pas automatiquement continue et il convient donc de distinguer dual topologique et dual algébrique.

Proposition 6.27. *Soit $p \in [1, +\infty[$ et q son exposant conjugué $\frac{1}{p} + \frac{1}{q} = 1$. Pour g fixé dans $\mathrm{L}^q_{\mathbb{K}}(\mu)$, l'application*

$$\Phi : \mathrm{L}^p_{\mathbb{K}}(\mu) \to \mathbb{K}, \quad f \mapsto \Phi(f) := \int_\Omega fg \,\mathrm{d}\mu$$

est une forme linéaire continue sur $\mathrm{L}^p_{\mathbb{K}}(\mu)$. Si $1 < p < +\infty$, la norme de Φ est égale à $\|g\|_q$. Si $p = 1$ (et donc $q = +\infty$) et si la mesure μ est σ-finie, la norme de Φ est $\|g\|_\infty$.

Preuve. D'abord Φ est bien définie puisque d'après l'inégalité de Hölder

$$\int_\Omega |fg| \,\mathrm{d}\mu \leq \|f\|_p \|g\|_q < +\infty.$$

On déduit immédiatement de cette inégalité que

$$\forall f \in \mathrm{L}^p(\mu), \quad |\Phi(f)| = \left| \int_\Omega fg \,\mathrm{d}\mu \right| \leq \int_\Omega |fg| \,\mathrm{d}\mu \leq \|f\|_p \|g\|_q.$$

Ceci montre que Φ est continue et que sa norme (dans l'espace des applications linéaires continues de $\mathrm{L}^p_{\mathbb{K}}(\mu)$ dans $\mathbb{K}$) vérifie

$$\|\Phi\| \leq \|g\|_q. \tag{6.18}$$

Cas $1 < p < +\infty$. Si $g = 0$ (le zéro de L^q, i.e. tout représentant de g est une fonction nulle μ p.p.), Φ est la forme linéaire nulle donc sa norme est $0 = \|g\|_q$. En dehors de ce cas particulier, $\|g\|_q > 0$. Choisissons un représentant de g (noté encore g) et définissons la fonction f par

$$f = h|g|^{q-1}, \quad \text{où} \quad h(\omega) = \begin{cases} \dfrac{\overline{g(\omega)}}{|g(\omega)|} & \text{si } g(\omega) \neq 0, \\ 0 & \text{si } g(\omega) = 0. \end{cases} \tag{6.19}$$

L'exposant q étant le conjugué de p, vérifie $q = p(q-1)$, d'où $|f|^p = |g|^{p(q-1)} = |g|^q$. L'appartenance de f à $\mathrm{L}^p(\mu)$ découle ainsi de celle de g à $\mathrm{L}^q(\mu)$. En outre $\|f\|_p = \|g\|_q^{q/p} > 0$. Calculons maintenant $\Phi(f)$:

$$\Phi(f) = \int_\Omega fg \,\mathrm{d}\mu = \int_{\{g \neq 0\}} \frac{\bar{g}}{|g|} |g|^{q-1} g \,\mathrm{d}\mu = \int_{\{g \neq 0\}} g\bar{g}|g|^{q-2} \,\mathrm{d}\mu = \int_\Omega |g|^q \,\mathrm{d}\mu = \|g\|_q^q.$$

On en déduit la minoration

$$\|\Phi\| \geq \frac{|\Phi(f)|}{\|f\|_p} = \|g\|_q^{q-q/p} = \|g\|_q$$

qui avec (6.18) entraîne $\|\Phi\| = \|g\|_q$.

Cas $p = +\infty$ et μ σ-finie. Le cas particulier $\|g\|_\infty$ étant immédiat, on suppose $\|g\|_\infty > 0$ et on fixe un représentant de g. Définissons pour $0 < \varepsilon < \|g\|_\infty$

$$A_\varepsilon := \{\omega \in \Omega;\ \|g\|_\infty - \varepsilon \leq |g(\omega)| \leq \|g\|_\infty\}.$$

D'après la définition du supremum essentiel de g, $\mu(A_\varepsilon)$ est strictement positif. La mesure μ étant σ-finie, il existe $B_\varepsilon \in \mathfrak{F}$ tel que $0 < \mu(B_\varepsilon) < +\infty$ et $B_\varepsilon \subset A_\varepsilon$. Soit alors

$$f_\varepsilon := \frac{\mathbf{1}_{B_\varepsilon}}{\mu(B_\varepsilon)} h,$$

où h est définie comme dans (6.19). Clairement, $f_\varepsilon \in \mathrm{L}^1(\mu)$ et $\|f_\varepsilon\|_1 = 1$. De plus comme sur B_ε, $g(\omega) \neq 0$,

$$\Phi(f_\varepsilon) = \int_\Omega f_\varepsilon g \,\mathrm{d}\mu = \int_{B_\varepsilon} \frac{\bar{g}}{|g|\mu(B_\varepsilon)} g \,\mathrm{d}\mu = \frac{1}{\mu(B_\varepsilon)} \int_{B_\varepsilon} |g| \,\mathrm{d}\mu \geq \|g\|_\infty - \varepsilon > 0.$$

On peut ainsi minorer la norme de Φ par

$$\|\Phi\| \geq \frac{|\Phi(f_\varepsilon)|}{\|f_\varepsilon\|_1} = \Phi(f_\varepsilon) \geq \|g\|_\infty - \varepsilon > 0.$$

Cette minoration étant vraie pour $0 < \varepsilon < \|g\|_\infty$, on en déduit en faisant tendre ε vers 0 que $\|\Phi\| \geq \|g\|_\infty$, ce qui avec (6.18) permet de conclure $\|\Phi\| = \|g\|_\infty$. □

Corollaire 6.28. *Si $1 < p < +\infty$ et $g \in \mathrm{L}^q(\mu)$,*

$$\|g\|_q = \sup_{\|f\|_p \leq 1} \left| \int_\Omega fg \,\mathrm{d}\mu \right|.$$

Ceci s'étend au cas $q = +\infty$, $p = 1$ si μ est σ-finie.

La proposition 6.27 admet une réciproque que nous admettrons.

Théorème 6.29 (Riesz)**.** *Si μ est σ-finie et $1 \leq p < +\infty$, toute forme linéaire continue Ψ sur $\mathrm{L}^p_{\mathbb{K}}(\mu)$ peut s'écrire*

$$\Psi : \mathrm{L}^p_{\mathbb{K}}(\mu) \to \mathbb{K}, \quad f \mapsto \int_\Omega fg \,\mathrm{d}\mu,$$

pour un unique élément g de L^q où $\frac{1}{p} + \frac{1}{q} = 1$. On peut ainsi identifier le dual topologique de $\mathrm{L}^p_{\mathbb{K}}(\mu)$ à $\mathrm{L}^q_{\mathbb{K}}(\mu)$:

$$\left(\mathrm{L}^p_{\mathbb{K}}(\mu)\right)' = \mathrm{L}^q_{\mathbb{K}}(\mu), \quad 1 \leq p < +\infty,\ \frac{1}{p} + \frac{1}{q} = 1.$$

Corollaire 6.30. *Si μ est σ-finie et $1 < p < +\infty$, $\mathrm{L}^p(\mu)$ est réflexif (on peut l'identifier avec son bidual topologique).*

Remarque 6.31. En général le dual de L^∞ n'est pas L^1. Il contient L^1.

Chapitre 7

Espaces de Hilbert et séries de Fourier

7.1 Espaces de Hilbert

7.1.1 Quelques définitions et rappels

Définition 7.1. *Un espace vectoriel normé* $(H, \| \ \|)$ *sur* $\mathbb{C}$ *(ou* $\mathbb{R}$*) est de Hilbert si sa norme provient d'un produit scalaire et s'il est* complet.

Nous nous plaçons dans toute la suite dans le cas d'un espace vectoriel complexe, tous les résultats étant immédiatement adaptables au cas réel. Pour $z = x + \mathrm{i}y \in \mathbb{C}$ $(x, y \in \mathbb{R})$, nous notons $\bar{z}$ son conjugué : $\bar{z} = x - \mathrm{i}y$.

Rappelons qu'un produit scalaire $\langle . , . \rangle$ est une *forme sesquilinéaire hermitienne définie positive*, autrement dit vérifiant pour tous f, g, $h \in H$ et tout $a \in \mathbb{C}$

i) $\langle g, f \rangle = \overline{\langle f, g \rangle}$;

ii) $\langle f + g, h \rangle = \langle f, h \rangle + \langle g, h \rangle$;

iii) $\langle af, g \rangle = a\langle f, g \rangle, \quad \langle f, ag \rangle = \bar{a}\langle f, g \rangle$;

iv) $\langle f, f \rangle \in \mathbb{R}_+$;

v) $\langle f, f \rangle = 0 \Rightarrow f = 0$.

Grâce à iv), on peut poser

$$\|f\| := (\langle f, f \rangle)^{1/2}.$$

Par iii) on a l'homogénéité $\|af\| = (a\bar{a})^{1/2}\|f\| = |a|\|f\|$. Les propriétés i) à iv) impliquent classiquement l'inégalité de Cauchy Schwarz :

$$\forall f, g \in H, \quad |\langle f, g \rangle| \leq \|f\|\|g\|. \tag{7.1}$$

De cette inégalité on déduit la sous-additivité de $\| \ \|$:

$$\|f + g\| \leq \|f\| + \|g\|.$$

En effet,

$$\begin{aligned}\|f+g\|^2 = \langle f+g, f+g\rangle &= \langle f,f\rangle + \langle f,g\rangle + \langle g,f\rangle + \langle g,g\rangle \\ &\leq \|f\|^2 + 2\|f\|\|g\| + \|g\|^2 \\ &= \big(\|f\| + \|g\|\big)^2.\end{aligned}$$

Ainsi l'homogénéité, la sous-additivité et v) montrent que $\| \ \|$ est bien une norme. La distance associée est $d(f,g) := \|f-g\|$ et c'est relativement à cette distance que H est complet.

Exemple 7.1. Soit $(\Omega, \mathcal{F}, \mu)$ est un espace mesuré, $\mathbb{K} = \mathbb{R}$ ou $\mathbb{C}$, l'espace $\mathrm{L}^2_{\mathbb{K}}(\mu)$, muni du produit scalaire

$$\langle f, g\rangle := \int_\Omega f\bar{g}\,\mathrm{d}\mu$$

est un espace de Hilbert.

Définition 7.2. *Une partie G de H est dite* dense *dans H si*

$$\forall h \in H, \forall \varepsilon > 0,\ \exists g \in G; \quad \|g-h\| < \varepsilon,$$

ou de manière équivalente si tout h de H est limite d'une suite d'éléments g_n de G : $\|g_n - h\| \to 0$.

Définition 7.3. *Une partie F de H est dite* totale *si l'ensemble des combinaisons linéaires finies des éléments de F est dense dans H.*

Définition 7.4. *Une famille $\{e_i,\ i \in I\}$ d'éléments de H est dite* orthonormée *si*

$$\forall i \in I,\ \forall j \in I, \quad \langle e_i, e_j\rangle = \delta_{i,j} := \begin{cases} 1 & \textit{si } i = j, \\ 0 & \textit{si } i \neq j. \end{cases}$$

Définition 7.5. *Une base hilbertienne de H ou base orthonormée est une famille orthonormée totale dans H.*

Pour l'instant cette définition reste formelle et nous ne savons pas s'il est possible (et en quel sens) de décomposer un élément quelconque de H sur une base hilbertienne de H (s'il en existe). Le cas particulier où H est de dimension finie est bien connu et l'on sait qu'il possède toujours une base orthonormée de cardinal égal à la dimension de l'espace.

7.1.2 Bases hilbertiennes et séparabilité

Rappelons qu'un espace métrique est séparable s'il possède un sous-ensemble dénombrable dense. La plupart des espaces de Hilbert utilisés en pratique sont séparables. Tout espace de Hilbert possédant une *suite* totale $(f_k)_{k\in\mathbb{N}}$ est séparable. Il suffit pour le voir de prendre comme partie dénombrable dense l'ensemble des combinaisons linéaires finies des f_k à coefficients rationnels (ou dans $\mathbb{Q} + \mathrm{i}\mathbb{Q}$ dans le cas complexe).

Proposition 7.6. *Tout espace de Hilbert séparable de dimension infinie possède une base hilbertienne dénombrable.*

Preuve. Soit $\{g_n,\ n \in \mathbb{N}\}$ une suite dense dans H. On peut en extraire par récurrence une sous-suite $(g_{n_i})_{i\in\mathbb{N}}$ telle que

a) Pour tout $i \in \mathbb{N}$, $\{g_{n_j};\ 0 \leq j \leq i\}$ est une famille libre.

b) Pour tout $i \in \mathbb{N}$, $\mathrm{Vect}\{g_k;\ 0 \leq k \leq n_i\} = \mathrm{Vect}\{g_{n_j};\ 0 \leq j \leq i\} =: E_i$.

c) $E := \bigcup_{i\in\mathbb{N}} E_i \supset \{g_n,\ n \in \mathbb{N}\}$.

En orthogonalisant $(g_{n_i})_{i\in\mathbb{N}}$ par le procédé de Gram Schmidt on construit une suite orthonormée $(f_{n_i})_{i\in\mathbb{N}}$ telle que pour tout $i \in \mathbb{N}$, $(f_{n_j})_{0\leq j\leq i}$ soit une base orthonormée de l'espace de dimension finie E_i. Pour montrer que la suite $(f_{n_i})_{i\in\mathbb{N}}$ est une base hilbertienne de H, il reste à vérifier qu'elle est totale dans H. Soit f quelconque dans H et $\varepsilon > 0$. Par c) et la densité de $\{g_n,\ n \in \mathbb{N}\}$ dans H, il existe $g \in E$ tel que $\|f - g\| < \varepsilon$. Comme E est la réunion des E_i, il existe un indice $i \in \mathbb{N}$ tel que $g \in E_i$. Or les $(f_{n_j})_{0\leq j\leq i}$ forment une base de l'espace de dimension finie E_i donc g se décompose dans cette base sous la forme $g = \sum_{j=0}^{i} a_j f_{n_j}$. Bien sûr, g, i et les a_j dépendent de f et de ε, mais le raisonnement est valable pour tout f et tout ε. La suite $(f_{n_i})_{i\in\mathbb{N}}$ est donc totale dans H. □

Proposition 7.7. *Si H est un espace de Hilbert séparable, toute base orthonormée de H est au plus dénombrable.*

Preuve. Soit $\{e_i,\ i \in I\}$ une base orthonormée dans H. Pour prouver que I est au plus dénombrable, il suffit de montrer l'existence d'une injection de I *dans* $\mathbb{N}$. Remarquons d'abord que si i et j sont deux éléments distincts de I,

$$\|e_i - e_j\| = \sqrt{2}.$$

Par séparabilité de H, il existe un ensemble dénombrable $\{g_n,\ n \in \mathbb{N}\}$ dense dans H. Pour chaque $i \in I$, il existe alors un entier n_i tel que

$$\|e_i - g_{n_i}\| < \frac{\sqrt{2}}{3}. \tag{7.2}$$

En choisissant pour chaque $i \in I$, un seul entier n_i parmi tous ceux vérifiant (7.2), on définit une application $\varphi : I \to \mathbb{N}$, $i \mapsto n_i$.

Par inégalité triangulaire on a alors pour $i \neq j$

$$\sqrt{2} = \|e_i - e_j\| \leq \|e_i - g_{n_i}\| + \|g_{n_i} - g_{n_j}\| + \|g_{n_j} - e_j\| \leq \|g_{n_i} - g_{n_j}\| + \frac{2\sqrt{2}}{3},$$

d'où

$$\|g_{n_i} - g_{n_j}\| \geq \frac{\sqrt{2}}{3}.$$

On en déduit que $g_{n_i} \neq g_{n_j}$ et donc que $n_i \neq n_j$. Ainsi pour tous i, j distincts dans I, $\varphi(i) \neq \varphi(j)$ ce qui exprime l'injectivité de φ. □

Remarque 7.8. La même démonstration prouve que si $\{e_i,\ i \in I\}$ est seulement une famille orthonormée dans H séparable, alors I est au plus dénombrable.

Proposition 7.9. *Soit H un espace de Hilbert ayant une base hilbertienne $\{e_i,\ i \in I\}$. Si $f \in H$ est orthogonale à tous les e_i, $i \in I$, alors $f = 0$. Ainsi une base hilbertienne est une famille orthonormée maximale pour l'inclusion.*

Preuve. Comme $\{e_i,\ i \in I\}$ est totale dans H, pour tout ε, on peut trouver une combinaison linéaire finie $\sum_{j\in J} a_j e_j$ telle que

$$\Bigl\| f - \sum_{j\in J} a_j e_j \Bigr\| < \varepsilon. \tag{7.3}$$

D'autre part, l'orthogonalité de f à tous les e_j et l'orthonormalité de la famille finie $\{e_j,\ j \in J\}$ nous permettent d'écrire

$$\Bigl\| f - \sum_{j\in J} a_j e_j \Bigr\|^2 = \|f\|^2 + 2\operatorname{Re}\Bigl\langle f, \sum_{j\in J} a_j e_j \Bigr\rangle + \sum_{j\in J} |a_j|^2 = \|f\|^2 + \sum_{j\in J} |a_j|^2.$$

Par conséquent $\|f\|^2 \leq \bigl\| f - \sum_{j\in J} a_j e_j \bigr\|^2$. En reportant cette inégalité dans (7.3), on en déduit $\|f\| < \varepsilon$. Ceci étant vrai pour tout ε, $\|f\| = 0$ et $f = 0$. $\square$

Corollaire 7.10. *Soit H un espace de Hilbert séparable et $\{e_i,\ i \in I\}$ une base hilbertienne. Alors ou bien H est de dimension finie et $\operatorname{card} I = \dim H$, ou bien H est de dimension infinie et I est exactement dénombrable (i.e. en bijection avec $\mathbb{N}$).*

Remarque 7.11. Dans ce qui suit, nous continuerons dans le cas où H est de dimension infinie et séparable, à indexer la base hilbertienne par I plutôt que par $\mathbb{N}$. Ce choix est motivé par l'existence d'exemples importants où la base est naturellement indexée par un I dénombrable autre que $\mathbb{N}$. C'est le cas notamment pour la base des exponentielles complexes de $\mathrm{L}^2(\mathbb{T})$ avec $I = \mathbb{Z}$ (cf. théorème 7.25) et pour les bases d'ondelettes de $\mathrm{L}^2(\mathbb{R})$, où l'on peut prendre pour I l'ensemble des nombres dyadiques.

7.1.3 Meilleure approximation

Théorème 7.12. *Soit $\{e_i,\ i \in I\}$ une famille orthonormée de H. Alors pour tout $f \in H$, toute partie finie $J \neq \emptyset$ de I et tous scalaires a_j, $j \in J$,*

$$\Bigl\| f - \sum_{j\in J} c_j(f) e_j \Bigr\| \leq \Bigl\| f - \sum_{j\in J} a_j e_j \Bigr\|, \tag{7.4}$$

avec

$$c_j(f) := \langle f, e_j \rangle. \tag{7.5}$$

L'égalité a lieu dans (7.4) si et seulement si $a_j = c_j(f)$ pour tout $j \in J$.

Preuve. En appliquant la formule $\|x+y\|^2 = \|x\|^2 + \|y\|^2 + 2\operatorname{Re}\langle x, y\rangle$,

$$\begin{aligned}
\Big\| f - \sum_{j\in J} a_j e_j \Big\|^2 &= \Big\| \Big\{ f - \sum_{j\in J} c_j(f) e_j \Big\} + \Big\{ \sum_{j\in J} \big(c_j(f) - a_j\big) e_j \Big\} \Big\|^2 \\
&= \Big\| f - \sum_{j\in J} c_j(f) e_j \Big\|^2 + \Big\| \sum_{j\in J} \big(c_j(f) - a_j\big) e_j \Big\|^2 \\
&\quad + 2\operatorname{Re} \Big\langle f - \sum_{j\in J} c_j(f) e_j, \sum_{j\in J} \big(c_j(f) - a_j\big) e_j \Big\rangle.
\end{aligned}$$

Le terme $2\operatorname{Re}\langle \dots, \dots \rangle$ dans cette expression est nul car en développant le produit scalaire par rapport à son deuxième argument, on obtient

$$\sum_{k\in J} \overline{\big(c_k(f) - a_k\big)} \Big\langle f - \sum_{j\in J} c_j(f) e_j, e_k \Big\rangle = \sum_{k\in J} \overline{\big(c_k(f) - a_k\big)} \big(\langle f, e_k\rangle - c_k(f)\big) = 0.$$

D'autre part, grâce à l'orthonormalité des e_j on a

$$\Big\| \sum_{j\in J} \big(c_j(f) - a_j\big) e_j \Big\|^2 = \sum_{j\in J} |c_j(f) - a_j|^2.$$

Finalement

$$\begin{aligned}
\Big\| f - \sum_{j\in J} a_j e_j \Big\|^2 &= \Big\| f - \sum_{j\in J} c_j(f) e_j \Big\| + \sum_{j\in J} |c_j(f) - a_j|^2 \\
&\geq \Big\| f - \sum_{j\in J} c_j(f) e_j \Big\|^2,
\end{aligned}$$

l'égalité n'étant possible que si $\sum_{j\in J} |c_j(f) - a_j|^2 = 0$, autrement dit $c_j(f) = a_j$ pour tout $j \in J$. □

Interprétation géométrique : le vecteur $\pi_J(f) := \sum_{j\in J} c_j(f) e_j$ est la projection orthogonale de f sur le s.e.v. de dimension finie $E_J := \operatorname{Vect}\{e_j,\ j \in J\}$. Il est orthogonal à $f - \pi_J(f)$. Définissons la distance de f à E_J par

$$d(f, E_J) := \inf_{g\in E_J} \|f - g\|.$$

Le théorème 7.12 nous dit alors que cet infimum est un *minimum* et que ce minimum est atteint en l'unique point $g = \pi_J(f)$. Ainsi la projection orthogonale de f sur E_J est *la* meilleure approximation de f par un élément de E_J. Il est clair que ce résultat se généralise immédiatement avec n'importe quel sous-espace vectoriel de dimension finie de H puisqu'un tel espace possède toujours une base orthonormée et que seuls les e_j indexés par J ont été utilisés dans la preuve du théorème 7.12.

Corollaire 7.13. *Soit $\{e_i,\ i \in I\}$ une famille orthonormée de H. Alors pour tout $f \in H$, toutes parties finies $K \subset J$ non vides de I*

$$\Big\| f - \sum_{j\in J} c_j(f) e_j \Big\| \leq \Big\| f - \sum_{j\in K} c_j(f) e_j \Big\|. \tag{7.6}$$

Preuve. Il suffit d'appliquer le théorème 7.12 avec la suite finie $(a_j)_{j\in J}$ définie par $a_j := c_j(f)$ si $j \in K$, $a_j := 0$ si $j \in J \setminus K$. □

Théorème 7.14 (Inégalité de Bessel)**.** *Si $\{e_i,\ i \in I\}$ est une famille orthonormée au plus dénombrable de H, alors*

$$\forall f \in H, \quad \sum_{i\in I} |\langle f, e_i\rangle|^2 \leq \|f\|^2. \tag{7.7}$$

Preuve. Notons encore $c_i(f) := \langle f, e_i\rangle$, $i \in I$. Pour toute partie finie J de I,

$$\begin{aligned}\Big\langle f - \sum_{j\in J} c_j(f)e_j, \sum_{j\in J} c_j(f)e_j\Big\rangle &= \sum_{k\in J} \overline{c_k(f)}\Big\langle f - \sum_{j\in J} c_j(f)e_j, e_k\Big\rangle \\ &= \sum_{k\in J} \overline{c_k(f)}\Big(\langle f, e_k\rangle - c_k(f)\Big) = 0.\end{aligned}$$

Cette orthogonalité nous permet d'écrire

$$\|f\|^2 = \Big\|f - \sum_{j\in J} c_j(f)e_j\Big\|^2 + \Big\|\sum_{j\in J} c_j(f)e_j\Big\|^2 = \Big\|f - \sum_{j\in J} c_j(f)e_j\Big\|^2 + \sum_{j\in J} |c_j(f)|^2,$$

d'où

$$\sum_{j\in J} |c_j(f)|^2 \leq \|f\|^2.$$

Cette inégalité étant vraie pour toute partie finie J de I, on en déduit (7.7) en prenant le sup sur J. □

Corollaire 7.15. *Soit $(e_k)_{k\in\mathbb{N}}$ une suite orthonormée dans H et pour tout $f \in H$, $c_k(f) := \langle f, e_k\rangle$. La série*

$$\sum_{k=0}^{+\infty} c_k(f)e_k \quad \textit{converge}\ \text{commutativement}\ \textit{dans}\ H. \tag{7.8}$$

La convergence commutative dans H se traduit par les conditions suivantes.

1. Il existe $g \in H$, telle que $\lim_{n\to+\infty} \Big\|g - \sum_{k=0}^{n} c_k(f)e_k\Big\| = 0$.
2. Pour toute bijection $\tau : \mathbb{N} \to \mathbb{N}$, $\lim_{n\to+\infty} \Big\|g - \sum_{k=0}^{n} c_{\tau(k)}(f)e_{\tau(k)}\Big\| = 0$.

Preuve. Par l'inégalité de Bessel,

$$\sum_{k=0}^{+\infty} |c_k(f)|^2 \leq \|f\|^2 < +\infty. \tag{7.9}$$

Cette série de réels positifs est donc convergente dans $\mathbb{R}_+$. La suite de ses sommes partielles vérifie le critère de Cauchy dans $\mathbb{R}_+$. En raison de l'orthonormalité de la suite (e_k), on a pour tous $n < m$ dans $\mathbb{N}$,

$$\Big\|\sum_{k=n}^{m} c_k(f)e_k\Big\|^2 = \sum_{k=n}^{m} |c_k(f)|^2.$$

On en déduit que la série *vectorielle* de terme général $c_k(f)e_k$ vérifie le critère de Cauchy dans H. Comme H est complet, elle converge donc (pour la distance de H) vers un élément g de H.

Soit maintenant une bijection $\tau : \mathbb{N} \to \mathbb{N}$. Notons :

$$I_n := \{0, 1, 2, \dots, n\} \triangle \{\tau(0), \tau(1), \tau(2), \dots, \tau(n)\}, \quad m(n) := \min I_n,$$

où $\triangle$ est la différence symétrique ensembliste [1]. Par orthonormalité de la suite (e_k) et l'inégalité de Bessel, on a pour tout $n \in \mathbb{N}$,

$$\left\| \sum_{k=0}^{n} c_k(f)e_k - \sum_{k=0}^{n} c_{\tau(k)}(f)e_{\tau(k)} \right\|^2 = \sum_{i\in I_n} |c_i(f)|^2 \leq \sum_{i=m(n)}^{+\infty} |c_i(f)|^2.$$

Il ne reste alors plus qu'à voir que $m(n)$ tend vers l'infini avec n pour conclure par (7.9) que ce majorant tend vers 0. Fixons un entier quelconque K. La condition $m(n) > K$ est réalisée dès que tous les entiers $j \leq K$ sont dans $\{\tau(0), \tau(1), \tau(2), \dots, \tau(n)\}$. Pour cela il suffit que $n \geq \max\{\tau^{-1}(j);\ 0 \leq j \leq K\} =: N(K)$. Ainsi $m(n)$ tend bien vers l'infini. □

7.1.4 Développement dans une base hilbertienne

Théorème 7.16. *Soit H un espace de Hilbert séparable.*

a) Si $\{e_i,\ i \in I\}$ est une base hilbertienne de H, on a le développement

$$\forall f \in H, \quad f = \sum_{i\in I} \langle f, e_i\rangle e_i, \tag{7.10}$$

où la série converge (commutativement) pour la topologie de H.

b) Si $\{e_i,\ i \in I\}$ est une famille orthonormée de H et si (7.10) est vérifiée, alors $\{e_i,\ i \in I\}$ est une base hilbertienne de H.

*c) Si $\{e_i,\ i \in I\}$ est une base hilbertienne de H, on a l'*identité de Plancherel :

$$\forall f \in H, \quad \|f\|^2 = \sum_{i\in I} |\langle f, e_i\rangle|^2. \tag{7.11}$$

et celle de Parseval :

$$\forall f, g \in H, \quad \langle f, g\rangle = \sum_{i\in I} \langle f, e_i\rangle \overline{\langle g, e_i\rangle}. \tag{7.12}$$

d) Si $\{e_i,\ i \in I\}$ est une famille de vecteurs unitaires de H (i.e. $\forall i \in I$, $\|e_i\| = 1$), vérifiant (7.11), alors c'est une base hilbertienne de H.

1. $A \triangle B := (A \cap B^c) \cup (A^c \cap B)$ est l'ensemble des éléments qui appartiennent à *un seul* des deux ensembles A et B.

Preuve. Le cas I fini étant bien connu, on peut se restreindre pour la preuve au cas I infini. D'après le corollaire 7.15 et les propriétés des séries à termes positifs, il est clair que toutes les convergences intervenant dans l'énoncé du théorème sont commutatives (sauf peut-être (7.12) que nous traiterons à part). On est donc libre de choisir une numérotation particulière de l'ensemble dénombrable I (cf. le corollaire 7.10 et la remarque 7.8) pour faire la preuve du théorème. Pour alléger les notations, on identifiera carrément I et $\mathbb{N}$.

Preuve du a). Puisque par hypothèse $\{e_n,\ n\in\mathbb{N}\}$ est totale dans H, on peut construire un tableau triangulaire de scalaires $\{a_{n,j},\ n\in\mathbb{N},\ 0\leq j\leq k_n\}$ tel que

$$f_n := \sum_{j=0}^{k_n} a_{n,j}e_j \xrightarrow[n\to+\infty]{H} f.$$

Posons avec $c_i(f) := \langle f, e_i\rangle$,

$$r_n := \left\| f - \sum_{j=0}^{n} c_j(f)e_j \right\|, \quad n\in\mathbb{N}.$$

Par la propriété de meilleure approximation (théorème 7.12), on a alors

$$0 \leq r_{k_n} = \left\| f - \sum_{j=0}^{k_n} c_j(f)e_j \right\| \leq \|f - f_n\| \xrightarrow[n\to+\infty]{} 0.$$

La sous-suite de réels positifs (r_{k_n}) converge ainsi vers 0. D'autre part en raison du corollaire 7.13, (r_n) est décroissante. C'est donc *la suite* (r_n) *toute entière* qui converge vers 0. Ceci prouve la convergence dans H de la série $\sum_{j=0}^{+\infty} c_j(f)e_j$ vers f. Cette convergence est commutative en raison du corollaire 7.15.

Preuve du b). La convergence (7.10) signifie que

$$\forall f\in H, \quad \sum_{j=0}^{n} c_j(f)e_j \xrightarrow[n\to+\infty]{H} f,$$

ce qui implique que $\{e_k,\ k\in\mathbb{N}\}$ est totale dans H. C'est donc bien une base hilbertienne.

Preuve du c). Si $\{e_k,\ k\in\mathbb{N}\}$ est une base hilbertienne, on a

$$\forall f\in H, \quad \sum_{k=0}^{n} c_k(f)e_k \xrightarrow[n\to+\infty]{H} f.$$

Par continuité de la norme sur H, on en déduit la convergence dans $\mathbb{R}_+$:

$$\left\| \sum_{k=0}^{n} c_k(f)e_k \right\|^2 \xrightarrow[n\to+\infty]{} \|f\|^2. \tag{7.13}$$

Par orthonormalité des e_k,

$$\left\| \sum_{k=0}^{n} c_k(f)e_k \right\|^2 = \sum_{k=0}^{n} |c_k(f)|^2$$

et en reportant dans (7.13) on en déduit l'identité de Plancherel.

Montrons l'identité de Parseval. D'abord l'inégalité $|zz'| \leq |z|^2 + |z'|^2$ appliquée au terme général de la série (7.12) montre (via l'identité de Plancherel) que celle-ci est absolument convergente donc commutativement convergente. Par le a), on a les convergences :

$$\sum_{k=0}^{n} c_k(f)e_k \xrightarrow[n\to+\infty]{H} f, \quad \sum_{k=0}^{n} c_k(g)e_k \xrightarrow[n\to+\infty]{H} g.$$

Le produit scalaire étant une forme bilinéaire *continue* sur H (en raison de l'inégalité de Cauchy Schwarz), on en déduit :

$$\left\langle \sum_{k=0}^{n} c_k(f)e_k, \sum_{k=0}^{n} c_k(g)e_k \right\rangle \xrightarrow[n\to+\infty]{} \langle f, g\rangle. \tag{7.14}$$

Par orthonormalité des e_k et bilinéarité du produit scalaire, on a

$$\left\langle \sum_{j=0}^{n} c_j(f)e_k, \sum_{k=0}^{n} c_k(g)e_k \right\rangle = \sum_{j,k=0}^{n} c_j(f)\overline{c_k(g)}\langle e_j, e_k\rangle = \sum_{k=0}^{n} c_k(f)\overline{c_k(g)}.$$

En reportant le résultat de ce calcul dans (7.14) on en déduit l'identité de Parseval.

Preuve du d). Supposons que les vecteurs e_k soient tous de norme 1 et que

$$\forall f \in H, \quad \|f\|^2 = \sum_{k\in\mathbb{N}} |\langle f, e_k\rangle|^2. \tag{7.15}$$

En choisissant $f = e_j$, cette identité nous donne

$$\forall j \in \mathbb{N}, \quad \|e_j\|^2 = \|e_j\|^2 + \sum_{\substack{k\in\mathbb{N}\\ k\neq j}} |\langle e_j, e_k\rangle|^2.$$

On en déduit que pour tout $k \neq j$, $\langle e_j, e_k\rangle = 0$. Ainsi $\{e_k,\ k \in \mathbb{N}\}$ est orthonormée. Pour vérifier qu'elle est totale, on note (cf. la preuve de l'inégalité de Bessel) que

$$\left\| f - \sum_{k=0}^{n} c_k(f)e_k \right\|^2 = \|f\|^2 - \sum_{k=0}^{n} |c_k(f)|^2$$

et que cette quantité tend vers 0 par (7.15). □

Proposition 7.17 (unicité du développement)**.** *Soit $\{e_i,\ i \in I\}$ une base hilbertienne de l'espace de Hilbert séparable H et $(a_i)_{i\in I}$ une famille de scalaires telle que la série $\sum_{i\in I} a_i e_i$ converge au sens de la topologie de H vers un $f \in H$ pour au moins une numérotation de l'ensemble dénombrable I. Alors $a_i = \langle f, e_i\rangle$ pour tout $i \in I$.*

Preuve. Il est clair qu'il suffit de faire la preuve lorsque $I = \mathbb{N}$. Par hypothèse nous avons donc

$$f_n := \sum_{k=0}^{n} \xrightarrow[n\to+\infty]{H} f.$$

D'autre part pour tous j et n dans $\mathbb{N}$, on a par orthonormalité des e_k :

$$\langle f_n, e_j \rangle = \begin{cases} 0 & \text{si } n < j, \\ a_j & \text{si } n \geq j. \end{cases}$$

La suite de complexes $(\langle f_n, e_j \rangle)_{n\in\mathbb{N}}$ est donc constante égale à a_j à partir du rang j. Elle converge ainsi vers a_j.

D'autre part, $\langle\, . \,, e_j \rangle$ est une forme linéaire *continue* sur H (par Cauchy Schwarz) et f_n tend vers f pour la topologie de H. Donc

$$\langle f_n, e_j \rangle \xrightarrow[n\to+\infty]{} \langle f, e_j \rangle.$$

Par unicité de la limite dans $\mathbb{C}$, on en déduit $a_j = \langle f, e_j \rangle$. □

Théorème 7.18 (Riesz-Fischer)**.** *Tout espace de Hilbert de dimension infinie et séparable est isomorphe (isométriquement) à $\ell^2(\mathbb{N})$.*

Preuve. Rappelons que l'espace $\ell^2(\mathbb{N})$ est l'espace des suites $u = (u_k)_{k\in\mathbb{N}}$ de nombres complexes telles que

$$\|u\|_2^2 := \sum_{k\in\mathbb{N}} |u_k|^2 < +\infty$$

et que c'est un espace de Hilbert pour le produit scalaire

$$\langle u, v \rangle_{\ell^2} := \sum_{k\in\mathbb{N}} u_k \overline{v_k}.$$

Soit H un espace de Hilbert de dimension infinie et séparable. Il possède alors une base hilbertienne exactement dénombrable (corollaire 7.10) que l'on peut toujours, quitte à la réindexer, écrire sous la forme $\{e_k,\ k \in \mathbb{N}\}$. On pose $c_k(f) := \langle f, e_k \rangle$ et on définit :

$$\Psi : H \longrightarrow \ell^2(\mathbb{N}), \quad f \longmapsto \big(c_k(f)\big)_{k\in\mathbb{N}}.$$

L'identité de Plancherel montre que $\Psi(f)$ est bien un élément de $\ell^2(\mathbb{N})$ et que Ψ est une isométrie :

$$\forall f \in H, \quad \|\Psi(f)\|_2 = \|f\|. \tag{7.16}$$

L'injectivité de Ψ en découle immédiatement. Par ailleurs, Ψ étant clairement linéaire, (7.16) implique aussi la continuité de Ψ. Il reste à vérifier la surjectivité de Ψ. Soit donc $u = (u_k)_{k\in\mathbb{N}}$ un élément quelconque de $\ell^2(\mathbb{N})$. Montrons l'existence d'un $f \in H$ tel que $\Psi(f) = u$. Définissons la suite $(f_n)_{n\in\mathbb{N}}$ d'éléments de H par

$$\forall n \in \mathbb{N}, \quad f_n := \sum_{k=0}^{n} u_k e_k.$$

Par orthonormalité des e_k, on a

$$\forall n \in \mathbb{N}, \forall j \in \mathbb{N}^*, \quad \|f_{n+j} - f_n\|^2 = \sum_{k=n+1}^{n+j} |u_k|^2 \leq \sum_{k=n+1}^{+\infty} |u_k|^2 \xrightarrow[n\to+\infty]{} 0.$$

On en déduit que $(f_n)_{n\in\mathbb{N}}$ est de Cauchy dans H. Comme H est complet, $(f_n)_{n\in\mathbb{N}}$ a une limite f dans H, ce qui s'écrit encore

$$f = \sum_{k=0}^{+\infty} u_k e_k \quad \text{(série convergente pour la topologie de } H\text{)}.$$

Par la proposition 7.17, on a alors $u_k = c_k(f)$ pour tout $k \in \mathbb{N}$, donc $\Psi(f) = u$. □

7.2 Séries de Fourier

7.2.1 Espaces de fonctions 2π-périodiques

On note $\mathcal{C}(\mathbb{T})$ l'espace des fonctions $\mathbb{R} \to \mathbb{C}$, continues et 2π-périodiques. On le munit de la norme définie par :

$$\forall f \in \mathcal{C}(\mathbb{T}), \quad \|f\|_\infty := \sup_{t\in[0,2\pi]} |f(t)| = \sup_{t\in\mathbb{R}} |f(t)|. \tag{7.17}$$

En raison de la 2π-périodicité, il est clair que l'on peut remplacer l'intervalle $[0, 2\pi]$ par n'importe quel intervalle de longueur 2π dans le supremum ci-dessus.

Proposition 7.19. *Si $f \in \mathcal{C}(\mathbb{T})$, f est uniformément continue sur $\mathbb{R}$.*

Vérification. La fonction f est continue sur $\mathbb{R}$, donc *uniformément* continue sur tout *compact* de $\mathbb{R}$, en particulier sur $[-2\pi, 2\pi]$:

$$\forall \varepsilon > 0,\ \exists \delta \in]0, \pi[,\ \forall s, t \in [-2\pi, 2\pi], \quad |t - s| < \delta \Rightarrow |f(t) - f(s)| < \varepsilon. \tag{7.18}$$

Par rapport à la définition classique de la continuité uniforme, la contrainte supplémentaire $\delta < \pi$ n'est évidemment pas restrictive et la raison de ce choix apparaîtra dans un instant. Fixons $\varepsilon > 0$ et un δ correspondant fourni par (7.18). Pour tous réels x et y tels que $|x - y| < \delta$, il existe un unique $k \in \mathbb{Z}$ tel que $s := x + 2k\pi$ appartienne à $[-\pi, \pi]$. Alors $t := y + 2k\pi$ est à une distance de s inférieure à $\delta < \pi$, donc t est dans $[-2\pi, 2\pi]$. Grâce à (7.18), à la 2π-périodicité de f et à l'égalité $|y - x| = |t - s|$, on a donc

$$\forall x, y \in \mathbb{R}, \quad |y - x| < \delta \Rightarrow |f(y) - f(x)| = |f(t) - f(s)| < \varepsilon.$$

Ceci étant vrai pour tout $\varepsilon > 0$, la continuité uniforme de f sur $\mathbb{R}$ est établie. □

Dans toute la suite du chapitre, on désigne par m la mesure $\frac{1}{2\pi}\lambda$, où λ est la mesure de Lebesgue sur $\mathbb{R}$. Pour $1 \leq p < +\infty$, $\mathcal{L}^p(\mathbb{T})$ est l'espace des fonctions mesurables $f : \mathbb{R} \to \mathbb{C}$, telles que :

a) $f(t + 2\pi) = f(t)$, pour m-presque tout $t \in \mathbb{R}$;

b) $\int_{[0,2\pi]} |f|^p \,\mathrm{dm} < +\infty$.

La condition a) implique (vérification facile laissée au lecteur) que pour m-presque tout $t \in \mathbb{R}$, quel que soit $k \in \mathbb{Z}$, $f(t+2k\pi) = f(t)$. On peut aussi vérifier [2] que grâce au a) et à l'invariance de m par translation, on a pour tout réel c, $\int_{[0,2\pi]} |f|^p \,\mathrm{dm} = \int_{[c,c+2\pi]} |f|^p \,\mathrm{dm}$. On munit l'espace $\mathcal{L}^p(\mathbb{T})$ de la semi-norme définie par :

$$\forall f \in \mathcal{L}^p(\mathbb{T}), \quad \|f\|_p := \left\{\int_{[0,2\pi]} |f|^p \,\mathrm{dm}\right\}^{1/p} = \left\{\int_{[c,c+2\pi]} |f|^p \,\mathrm{dm}\right\}^{1/p}, c \in \mathbb{R}. \tag{7.19}$$

Il s'agit bien d'une semi-norme (homogénéité et sous-additivité ont été vues au chapitre sur les espaces L^p) car la nullité de $\|f\|_p$ équivaut à la nullité de l'intégrale $\int_{[0,2\pi]} |f|^p \,\mathrm{dm}$, donc à la nullité m-presque partout sur $[0, 2\pi]$ de f et grâce à la condition a), à la nullité m-presque partout sur $\mathbb{R}$. Pour en faire une vraie norme, on passe au quotient de $\mathcal{L}^p(\mathbb{T})$ par son sous-espace des fonctions nulles m-presque partout sur $\mathbb{R}$ et on définit ainsi l'espace $\mathrm{L}^p(\mathbb{T})$ des classes d'équivalences des éléments de $\mathcal{L}^p(\mathbb{T})$ modulo l'égalité m-p.p.

Proposition 7.20. *Pour $1 \leq p \leq r < +\infty$, on a les inclusions topologiques :*

$$\mathcal{C}(\mathbb{T}) \hookrightarrow \mathrm{L}^r(\mathbb{T}) \hookrightarrow \mathrm{L}^p(\mathbb{T}) \hookrightarrow \mathrm{L}^1(\mathbb{T}). \tag{7.20}$$

Preuve. Cet énoncé comporte un léger abus de langage puisque la première inclusion n'en est pas une au sens ensembliste : les éléments de $\mathcal{C}(\mathbb{T})$ sont des vraies fonctions tandis que ceux de $\mathrm{L}^r(\mathbb{T})$ sont des classes d'équivalence. Néanmoins, on peut définir une injection canonique $i : \mathcal{C}(\mathbb{T}) \to \mathrm{L}^r(\mathbb{T})$ en associant à toute $f \in \mathcal{C}(\mathbb{T})$ sa classe $\tilde{f}$ modulo l'égalité m presque partout. Si f est continue, elle est bornée sur $[0, 2\pi]$, donc $\int_{[0,2\pi]} |f|^r \,\mathrm{dm} < +\infty$ car m est finie (sur $[0, 2\pi]$). La condition a) étant évidememnt vérifiée, f appartient à $\mathcal{L}^r(\mathbb{T})$. Donc $\tilde{f}$ est bien un élément de L^r et i est une application $\mathcal{C}(\mathbb{T}) \to \mathrm{L}^r(\mathbb{T})$. Vérifions qu'elle est injective. Si f et g sont deux éléments distincts de $\mathcal{C}(\mathbb{T})$, l'ensemble $A := \{t \in [0, 2\pi];\ f(t) \neq g(t)\}$ est un *ouvert* non vide de $[0, 2\pi]$, en raison de la continuité de $f - g$. Or un ouvert non vide de $[0, 2\pi]$ ne peut être de mesure de Lebesgue nulle (il contient au moins un intervalle ouvert de longueur non nulle), donc $\mathrm{m}(A) > 0$ et ceci empêche f d'être égale m-p.p. à g, d'où $\tilde{f} \neq \tilde{g}$ et i est bien injective. La linéarité de i étant évidente, il ne reste plus qu'à vérifier sa continuité en notant que :

$$\forall f \in \mathcal{C}(\mathbb{T}), \quad \|i(f)\|_r = \left\{\int_{[0,2\pi]} |f|^r \,\mathrm{dm}\right\}^{1/r} \leq \left\{\int_{[0,2\pi]} \|f\|_\infty^r \,\mathrm{dm}\right\}^{1/r} = \|f\|_\infty.$$

La première inclusion topologique dans (7.20) est ainsi établie. Les autres ne sont qu'une réécriture du théorème 6.18, puisque m est une mesure finie. □

7.2.2 Coefficients de Fourier

Définition 7.21. *Notons $\mathsf{E} := \{e_n;\ n \in \mathbb{Z}\}$, la famille des fonctions*

$$e_n : \mathbb{R} \to \mathbb{C}, \quad e_n(t) := \mathrm{e}^{int}, \quad t \in \mathbb{R}. \tag{7.21}$$

2. Voir par exemple le corrigé de l'examen de juin dans les Annales d'I.F.P. 2001–2002.

Si $f \in \mathrm{L}^1(\mathbb{T})$, on appelle n-ième coefficient de Fourier de f, le nombre complexe

$$c_n(f) := \int_{[0,2\pi]} f\overline{e_n}\,\mathrm{dm} = \int_{[0,2\pi]} f(t)\mathrm{e}^{-int}\,\mathrm{dm}(t) = \frac{1}{2\pi}\int_{[0,2\pi]} f(t)\mathrm{e}^{-int}\,\mathrm{d}\lambda(t).$$

La série formelle $\sum_{n\in\mathbb{Z}} c_n(f)e_n$ *est appelée* série de Fourier *de f.*

Remarques 7.22. 1. Les $c_n(f)$ étant définis par des intégrales relativement à la mesure m, ne dépendent pas du choix du représentant de f modulo l'égalité m-p.p., il est donc légitime de parler de série de Fourier d'un élément de $\mathrm{L}^1(\mathbb{T})$.

2. Nous ne prétendons pas que la série de Fourier de f converge vers f. La question que nous nous proposons d'étudier est précisément : « sous quelles hypothèses relatives à f et pour quel type de convergence peut-on dire que f est la somme de sa série de Fourier ? »
3. Si $f \in \mathrm{L}^2(\mathbb{T})$, les $c_n(f)$ s'interprètent comme les produits scalaires :

$$c_n(f) = \big\langle f, e_n\big\rangle. \tag{7.22}$$

L'interprétation des $c_n(f)$ comme produits scalaires est limitée aux $f \in \mathrm{L}^2$. Voici une relation plus générale, basée sur les produits de convolution, qui nous sera utile dans la suite.

Proposition 7.23. *Pour toute $f \in \mathrm{L}^1(\mathbb{T})$,*

$$\forall n \in \mathbb{Z}, \quad f * e_n = c_n(f)e_n, \tag{7.23}$$

le produit de convolution étant défini ici par :

$$\forall x \in \mathbb{R}, \quad (f * e_n)(x) := \int_{[0,2\pi]} f(t)e_n(x-t)\,\mathrm{dm}(t).$$

Vérification. L'égalité (7.23) s'obtient simplement en écrivant pour tout $x \in \mathbb{R}$,

$$(f * e_n)(x) = \int_{[0,2\pi]} f(t)\mathrm{e}^{inx}\mathrm{e}^{-int}\,\mathrm{dm}(t) = \left\{\int_{[0,2\pi]} f(t)\mathrm{e}^{-int}\,\mathrm{dm}(t)\right\}\mathrm{e}^{inx} = c_n(f)e_n(x).$$

$\square$

7.2.3 Bases trigonométriques de $\mathrm{L}^2(\mathbb{T})$

Proposition 7.24. $\mathsf{E} = \{e_n;\ n \in \mathbb{Z}\}$ *est une famille orthonormée de $\mathrm{L}^2(\mathbb{T})$.*

Vérification. Par conversion en intégrale de Riemann de l'intégrale de Lebesgue de la fonction continue $e_k\overline{e_l}$ sur l'intervalle $[0, 2\pi]$, on a en effet pour tous k, $l \in \mathbb{Z}$,

$$\langle e_k, e_l\rangle = \int_{[0,2\pi]} e_k\overline{e_l}\,\mathrm{dm} = \frac{1}{2\pi}\int_0^{2\pi} \mathrm{e}^{ikt}\mathrm{e}^{-ilt}\,\mathrm{d}t = \frac{1}{2\pi}\int_0^{2\pi} \mathrm{e}^{i(k-l)t}\,\mathrm{d}t.$$

Si $k \neq l$, une primitive de $\mathrm{e}^{i(k-l)t}$ est $\frac{1}{i(k-l)}\mathrm{e}^{i(k-l)t}$, dont la variation entre 0 et 2π est nulle par périodicité. Si $k = l$, la fonction $\mathrm{e}^{i(k-l)t}$ est la constante 1 et $\frac{1}{2\pi}\int_0^{2\pi} \mathrm{d}t = 1$. Finalement,

$$\langle e_k, e_l \rangle = \begin{cases} 1 & \text{si } k = l, \\ 0 & \text{si } k \neq l, \end{cases}$$

la famille E est donc orthonormée. $\square$

Théorème 7.25. $\mathsf{E} = \{e_n;\ n \in \mathbb{Z}\}$ *est une base hilbertienne de* $\mathrm{L}^2(\mathbb{T})$*. On a donc le développement :*

$$\forall f \in \mathrm{L}^2(\mathbb{T}), \quad f = \sum_{k\in\mathbb{Z}} c_k(f)e_k, \tag{7.24}$$

avec convergence commutative de cette série pour la topologie de $\mathrm{L}^2(\mathbb{T})$*. En particulier on a*

$$S_n(f) := \sum_{k=-n}^{k=+n} c_k(f)e_k \xrightarrow[n\to+\infty]{\mathrm{L}^2(\mathbb{T})} f, \tag{7.25}$$

ce qui se traduit par :

$$\int_{[0,2\pi]} \left|S_n(f) - f\right|^2 \mathrm{dm} \xrightarrow[n\to+\infty]{} 0. \tag{7.26}$$

Remarque 7.26. Une autre façon d'énoncer (7.24) est de dire que la série de Fourier de toute fonction de $f \in \mathrm{L}^2(\mathbb{T})$ converge vers f au sens L^2. Insistons à nouveau sur le fait que (7.24) ne signifie aucunement que $f(t) = \sum_{k\in\mathbb{Z}} c_k(f)e_k(t)$ pour tout $t \in \mathbb{R}$. Du point de vue de la convergence ponctuelle, tout ce que l'on peut tirer de (7.25) en l'état actuel de nos connaissances, est l'existence d'une sous-suite $\left(S_{n_j}(f)(t)\right)$ qui converge λ presque partout vers $f(t)$ (en notant encore f un représentant quelconque de la classe f).

Preuve du théorème 7.25. Nous savons déjà que E est orthonormée, il reste à montrer qu'elle est totale dans $\mathrm{L}^2(\mathbb{T})$ (cf. les définitions 7.3 et 7.5).

Admettons provisoirement que E est totale dans l'espace $\left(\mathcal{C}(\mathbb{T}), \|\ \|_\infty\right)$. Cette propriété sera établie ci-dessous comme une conséquence immédiate du théorème de Fejér dont la démonstration est indépendante du théorème 7.25.

Soit $f \in \mathrm{L}^2(\mathbb{T})$ et notons encore f un de ses représentants (une vraie fonction). Il s'agit de montrer que pour tout $\varepsilon > 0$, on peut trouver une combinaison linéaire finie h_ε des e_k telle que $\|f - h_\varepsilon\|_2 < \varepsilon$.

Par densité de $\mathcal{C}(\mathbb{T})$ dans $\mathrm{L}^2(\mathbb{T})$ (pour la distance euclidienne associée à $\|\ \|_2$), on peut trouver une fonction $g_\varepsilon \in \mathcal{C}(\mathbb{T})$ telle que $\|f - g_\varepsilon\|_2 < \varepsilon/2$. Comme E est totale dans $\mathcal{C}(\mathbb{T})$, on peut approcher à $\varepsilon/2$ près (au sens cette fois de la distance associée à $\|\ \|_\infty$) par une combinaison linéaire finie h_ε des e_k : $\|g_\varepsilon - h_\varepsilon\|_\infty < \varepsilon/2$.

En utilisant l'inclusion topologique de $\mathcal{C}(\mathbb{T})$ dans $\mathrm{L}^2(\mathbb{T})$ (cf. 7.20), on peut alors contrôler la norme L^2 de la fonction continue $g_\varepsilon - h_\varepsilon$:

$$\|g_\varepsilon - h_\varepsilon\|_2 \leq \|g_\varepsilon - h_\varepsilon\|_\infty < \varepsilon/2.$$

Finalement en appliquant l'inégalité triangulaire dans L^2 on obtient :

$$\|f - h_\varepsilon\|_2 \leq \|f - g_\varepsilon\|_2 + \|g_\varepsilon - h_\varepsilon\|_2 < \varepsilon.$$

Comme f et ε étaient quelconques, E est bien totale dans $\mathrm{L}^2(\mathbb{T})$. □

Corollaire 7.27 (Bessel, Plancherel, Parseval)**.** *Pour toutes f, $g \in \mathrm{L}^2(\mathbb{T})$, on a*

a) Inégalité de Bessel : pour tout $J \subset \mathbb{Z}$,

$$\sum_{k\in J}\left|\int_{[0,2\pi]} f(t)\mathrm{e}^{-ikt}\,\mathrm{dm}(t)\right|^2 \leq \int_{[0,2\pi]} |f|^2\,\mathrm{dm}.$$

b) Identité de Plancherel :

$$\sum_{k\in\mathbb{Z}}\left|\int_{[0,2\pi]} f(t)\mathrm{e}^{-ikt}\,\mathrm{dm}(t)\right|^2 = \int_{[0,2\pi]} |f|^2\,\mathrm{dm}.$$

b) Identité de Parseval :

$$\sum_{k\in\mathbb{Z}}\int_{[0,2\pi]} f(t)\mathrm{e}^{-ikt}\,\mathrm{dm}(t)\int_{[0,2\pi]} \overline{g(t)}\mathrm{e}^{ikt}\,\mathrm{dm}(t) = \int_{[0,2\pi]} f\overline{g}\,\mathrm{dm}.$$

Vérification. Ce corollaire est une simple traduction des théorèmes 7.14 et 7.16 c) avec pour espace de Hilbert $H = \mathrm{L}^2(\mathbb{T})$ et comme base hilbertienne $\{e_k;\ k \in I\}$, la famille des e_n définies par (7.21), $I = \mathbb{Z}$. □

Corollaire 7.28. *La famille $\mathsf{F} := \{1\} \cup \{\sqrt{2}\cos(k\,.\,),\ \sqrt{2}\sin(k\,.\,),\ k \in \mathbb{N}^*\}$ est une base hilbertienne de $\mathrm{L}^2(\mathbb{T})$. On a pour toute $f \in \mathrm{L}^2(\mathbb{T})$, la décomposition en série commutativement convergente au sens L^2 :*

$$f = \frac{1}{2}a_0 + \sum_{k=1}^{+\infty}\Big(a_k\cos(k\,.\,) + b_k\sin(k\,.\,)\Big), \tag{7.27}$$

où les coefficients $a_k = a_k(f)$ et $b_k = b_k(f)$ sont définis par

$$a_k := \frac{1}{\pi}\int_{[0,2\pi]} f(t)\cos kt\,\mathrm{d}\lambda(t), \quad k \in \mathbb{N}, \tag{7.28}$$

$$b_k := \frac{1}{\pi}\int_{[0,2\pi]} f(t)\sin kt\,\mathrm{d}\lambda(t), \quad k \in \mathbb{N}^*. \tag{7.29}$$

Preuve. L'orthonormalité de la famille F résulte d'un calcul classique laissé au lecteur. Pour prouver que F est totale dans $\mathrm{L}^2(\mathbb{T})$, on utilise pour f quelconque dans $\mathrm{L}^2(\mathbb{T})$ la convergence (7.26) de $\|S_n(f) - f\|_2$ vers 0. Il suffit alors de vérifier que pour chaque

n, $S_n(f)$ est une combinaison linéaire finie de fonctions de la famille F. Ceci résulte du calcul élémentaire suivant :

$$\begin{aligned} S_n(f)(x) &= \sum_{k=-n}^{k=n} c_k(f)e_k(x) \\ &= c_0(f) + \sum_{k=1}^{n}\Big\{c_k(f)e_k(x) + c_{-k}(f)e_{-k}(x)\Big\} \\ &= c_0(f) + \sum_{k=1}^{n}\Big\{c_k(f)(\cos kx + i\sin kx) + c_{-k}(f)(\cos kx - i\sin kx)\Big\} \\ &= c_0(f) + \sum_{k=1}^{n}\Big\{\Big(c_k(f) + c_{-k}(f)\Big)\cos kx + i\Big(c_k(f) - c_{-k}(f)\Big)\sin kx\Big\}. \end{aligned}$$

Ainsi F est une base hilbertienne de $\mathrm{L}^2(\mathbb{T})$. Le développement de f dans cette base s'écrit :

$$f = \alpha_0 + \sum_{k=1}^{+\infty}\Big(\alpha_k\sqrt{2}\cos(k\,.\,) + \beta_k\sqrt{2}\sin(k\,.\,)\Big), \tag{7.30}$$

série commutativement convergente dans $\mathrm{L}^2(\mathbb{T})$, avec les coefficients définis par :

$$\begin{aligned} \alpha_0 &:= \langle f, 1\rangle = \frac{1}{2\pi}\int_{[0,2\pi]} f(t)\,\mathrm{d}\lambda(t) = \frac{a_0}{2}, \\ \alpha_k &:= \langle f, \sqrt{2}\cos(k\,.\,)\rangle = \frac{\sqrt{2}}{2\pi}\int_{[0,2\pi]} f(t)\cos kt\,\mathrm{d}\lambda(t) = \frac{a_k}{\sqrt{2}}, \quad k\in\mathbb{N}^*, \\ \beta_k &:= \langle f, \sqrt{2}\sin(k\,.\,)\rangle = \frac{\sqrt{2}}{2\pi}\int_{[0,2\pi]} f(t)\sin kt\,\mathrm{d}\lambda(t) = \frac{b_k}{\sqrt{2}}, \quad k\in\mathbb{N}^*. \end{aligned}$$

En reportant ces valeurs dans (7.30), on obtient le développement (7.27). □

Remarque 7.29. Les fonctions qui interviennent dans le développement (7.27) sont orthogonales, mais pas orthonormées. Il faut y prendre garde si l'on veut écrire l'analogue du corollaire 7.27 avec les coefficients a_k, b_k. Pour le faire proprement, il faut repasser par les α_k et les β_k. Par exemple on obtient ainsi pour l'identité de Parseval :

$$\forall f,g\in\mathrm{L}^2(\mathbb{T}),\quad \int_{[0,2\pi]} f\bar{g}\,\mathrm{dm} = \frac{1}{4}a_0(f)\overline{a_0(g)} + \frac{1}{2}\sum_{k=1}^{+\infty}\Big(a_k(f)\overline{a_k(g)} + b_k(f)\overline{b_k(g)}\Big).$$

Remarque 7.30. Si f est à valeurs réelles, les sommes partielles $S_n(f)$ et leur analogues pour la base F sont des fonctions à valeurs réelles. Il suffit de le vérifier pour $S_n(f)$ puisque :

$$\alpha_0 + \sum_{k=1}^{n}\Big(\alpha_k\sqrt{2}\cos(k\,.\,) + \beta_k\sqrt{2}\sin(k\,.\,)\Big) = c_0(f) + \sum_{k=1}^{n}\Big\{c_k(f)e_k + c_{-k}(f)e_{-k}\Big\} = S_n(f).$$

Or pour toute $g\in\mathrm{L}^1(\mathbb{T})$, même à valeurs complexes, on a $\overline{c_k(g)} = c_{-k}(\bar{g})$ pour tout $k\in\mathbb{Z}$. Comme f est réelle, $f=\bar{f}$, d'où pour tout $k\in\mathbb{Z}$, les égalités

$$c_k(f)e_k + c_{-k}(f)e_{-k} = c_k(f)e_k + c_{-k}(\bar{f})e_{-k} = c_k(f)e_k + \overline{c_k(f)}\overline{e_k} = 2\,\mathrm{Re}\Big(c_k(f)e_k\Big),$$

dont on déduit que $S_n(f)$ est à valeurs réelles.

Exemple 7.2. Soit f la fonction 2π périodique dont la restriction à $[-\pi, \pi]$ coïncide avec la fonction $\mathbf{1}_{[-\pi/2,\pi/2]}$. Un calcul immédiat nous donne

$$c_0(f) = \frac{1}{2}, \quad c_k(f) = \frac{\sin(k\pi/2)}{k\pi}, \quad k \in \mathbb{Z}^*.$$

Les sommes partielles $S_n(f)$ sont réelles et peuvent s'écrire pour tout $n \geq 1$,

$$S_n(f) = \frac{1}{2} + \frac{2}{\pi}\sum_{k=1}^{n} \frac{\sin(k\pi/2)}{k} \cos kx.$$

Les figures 7.1 et 7.2 présentent une représentation graphique de $S_{13}(f)$ et $S_{199}(f)$. On voit bien sur la figure 7.2 la convergence L^2 de $S_n(f)$ vers f. On voit aussi le mauvais comportement de $S_n(f)$ du point de vue de la convergence ponctuelle aux points de discontinuité $-\pi/2$ et $\pi/2$. Il est d'ailleurs facile de voir directement que $S_n(f)(\pi/2)$ ne converge pas vers $f(\pi/2) = 1$ puisque

$$S_n(f)(\pi/2) = \frac{1}{2} + \frac{2}{\pi}\sum_{k=1}^{n} \frac{\sin(k\pi/2)\cos(k\pi/2)}{k} = \frac{1}{2}.$$

Pour un oeil averti, la figure 7.2 illustre aussi le « phénomène de Gibbs ». On désigne ainsi le résultat suivant.

Soit f une fonction 2π périodique et C^1 par morceaux (les raccords n'étant pas nécessairement continus). Si x_0 est un point de discontinuité de f, $S_n(f)(x_0)$ converge vers $\frac{1}{2}(f(x_0^+) + f(x_0^-))$. Néanmoins, pour chaque n le maximum sur $]x_0 - c/n, x_0[$ (resp. sur $]x_0, x_0 + c/n[$) de $|S_n(f)(x) - f(x_0^-)|$ (resp. de $|S_n(f)(x) - f(x_0^+)|$) ne tend pas vers 0 quand n tend vers $+\infty$. Il est asymptotiquement de l'ordre d'environ 9% du saut $|f(x_0^+) - f(x_0^-)|$.

Après cette digression sur la convergence ponctuelle, revenons au point de vue L^2 et regardons ce que donne sur notre exemple l'identité de Plancherel. Compte-tenu de la valeurs des $c_k(f)$, celle ci s'écrit ici

$$\|f\|_2^2 = \sum_{k\in\mathbb{Z}} |c_k(f)|^2 = \frac{1}{4} + \frac{2}{\pi^2}\sum_{k=1}^{+\infty} \frac{\sin^2(k\pi/2)}{k^2} = \frac{1}{4} + \frac{2}{\pi^2}\sum_{j=0}^{+\infty} \frac{1}{(2j+1)^2}.$$

D'autre part le calcul direct de $\|f\|_2^2$ s'écrit

$$\|f\|_2^2 = \int_{[-\pi,\pi]} |f|^2 \,\mathrm{dm} = \frac{1}{2}.$$

Par comparaison des deux résultats, on en déduit

$$\sum_{j=1}^{+\infty} \frac{1}{(2j+1)^2} = \frac{\pi^2}{8}.$$

Cette formule permet de retrouver la valeur de $\zeta(2) := \sum_{n\geq 1} n^{-2}$. En effet

$$\zeta(2) = \sum_{j=1}^{+\infty} \frac{1}{(2j)^2} + \sum_{j=0}^{+\infty} \frac{1}{(2j+1)^2} = \frac{1}{4}\zeta(2) + \frac{\pi^2}{8},$$

d'où $(1 - 1/4)\zeta(2) = \pi^2/8$, puis $\zeta(2) = \pi^2/6$.

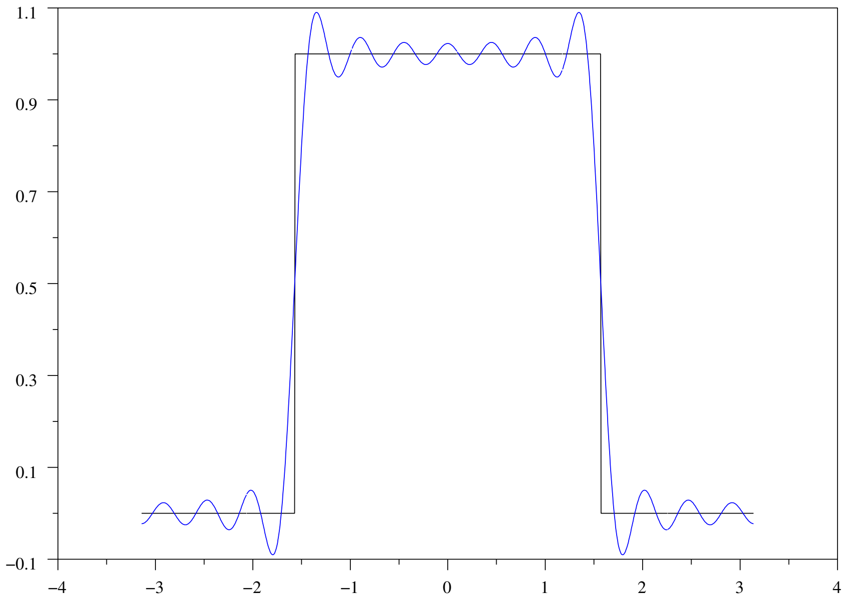


FIGURE 7.1 – $S_{13}(f)$ pour $f = \mathbf{1}_{[-\pi/2,\pi/2]}$

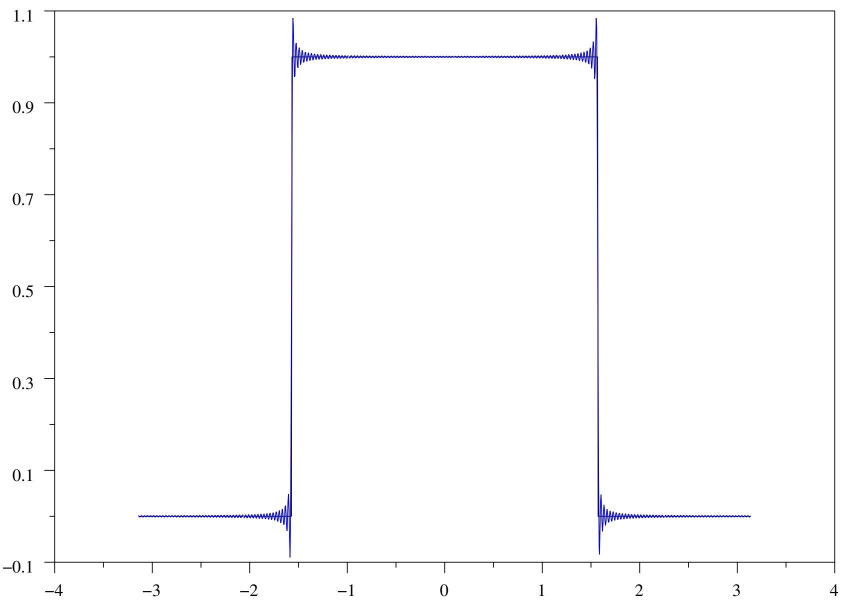


FIGURE 7.2 – $S_{199}(f)$ pour $f = \mathbf{1}_{[-\pi/2,\pi/2]}$

7.2.4 Noyaux de Dirichlet et de Fejér

Nous venons de voir que la série de Fourier d'une fonction $f \in \mathrm{L}^2(\mathbb{T})$ se comporte bien dans cet espace, c'est-à-dire converge pour la topologie de cet espace vers f : $\|S_n(f) - f\|_2 \to 0$. La même question concernant l'espace $\mathcal{C}(\mathbb{T})$ est naturelle et même antérieure. Malheureusement la réponse est négative. Il existe des fonctions continues f 2π-périodiques dont la série de Fourier diverge en certains points, donc ne peut converger uniformément vers f (la convergence au sens de l'espace $\mathcal{C}(\mathbb{T})$ est la convergence uniforme sur $[0, 2\pi]$, donc aussi sur $\mathbb{R}$ par périodicité). Nous verrons ci-dessous qu'un remède à ce mauvais comportement de $S_n(f)$ dans l'espace $\mathcal{C}(\mathbb{T})$ est de remplacer la convergence de la suite $(S_n(f))$ par celle de ses moyennes arithmétiques (convergence au sens de Césaro). Avant d'en arriver là, il nous faut en passer par les quelques préliminaires techniques contenus dans cette sous-section.

Définition 7.31. *On appelle* noyau de Dirichlet d'ordre n *la fonction :*

$$D_n := \sum_{k=-n}^{k=n} e_k, \quad n \in \mathbb{N}. \tag{7.31}$$

L'intérêt de ce noyau D_n est qu'il permet d'écrire la somme partielle $S_n(f)$ de la série de Fourier d'une $f \in \mathrm{L}^1(\mathbb{T})$ comme le produit de convolution

$$S_n(f) = f * D_n = \frac{1}{2\pi} \int_{[-\pi,\pi]} f(t) D_n(\,.\, - t)\, \mathrm{d}\lambda(t), \tag{7.32}$$

transformant ainsi l'étude de la convergence de la série de Fourier en un problème de convergence d'intégrale. La justification de (7.32) est le calcul suivant qui utilise la proposition 7.23 et la distributivité du produit de convolution d'une fonction de L^1 par rapport à l'addition de fonctions de L^∞ (qui résulte elle même de la linéarité de l'intégrale) :

$$S_n(f) = \sum_{k=-n}^{k=n} c_k(f) e_k = \sum_{k=-n}^{k=n} f * e_k = f * \left(\sum_{k=-n}^{k=n} e_k \right) = f * D_n.$$

Proposition 7.32. *Soit D_n le noyau de Dirichlet d'ordre n.*

a) Son intégrale sur une période ne dépend pas de n :

$$\int_{[-\pi,\pi]} D_n \,\mathrm{dm} = \frac{1}{2\pi} \int_{-\pi}^{\pi} D_n(t)\,\mathrm{d}t = 1.$$

b) La fonction continue 2π-périodique D_n a pour expression explicite :

$$D_n(t) = \begin{cases} \dfrac{\sin\bigl((n+1/2)t\bigr)}{\sin(t/2)} & \text{si } t \notin 2\pi\mathbb{Z}, \\ 2n+1 & \text{si } t \in 2\pi\mathbb{Z}. \end{cases}$$

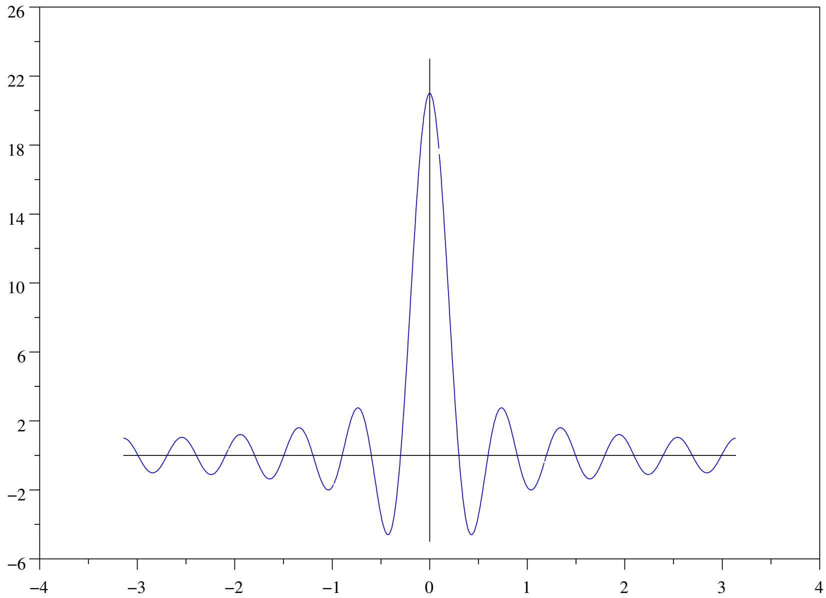


FIGURE 7.3 – Noyau de Dirichlet D_{10} sur $[-\pi, \pi]$

Vérification. Pour le a), il suffit de revenir à la définition de D_n et de remarquer que $\frac{1}{2\pi}\int_{-\pi}^{\pi} e_0(t)\,\mathrm{d}t = 1$ et si $k \neq 0$, $\int_{-\pi}^{\pi} e_k(t)\,\mathrm{d}t = 0$.

Pour le b), si $t \in 2\pi\mathbb{Z}$, on a pour tout $k \in \mathbb{Z}$, $e_k(t) = e_k(0) = 1$, d'où $D_n(t) = \sum_{k=-n}^{k=n} e_k(t) = 2n+1$. Si $t \notin 2\pi\mathbb{Z}$, on a $\mathrm{e}^{it} \neq 1$ et on peut calculer $D_n(t)$ comme la somme d'une suite géométrique de raison e^{it} :

$$\begin{aligned} D_n(t) = \mathrm{e}^{-int}\sum_{k=0}^{2n}\mathrm{e}^{ikt} = \mathrm{e}^{-int}\frac{\mathrm{e}^{(2n+1)it}-1}{\mathrm{e}^{it}-1} &= \frac{\mathrm{e}^{(n+1)it}-\mathrm{e}^{-nit}}{\mathrm{e}^{it/2}(\mathrm{e}^{it/2}-\mathrm{e}^{-it/2})} \\ &= \frac{\mathrm{e}^{(n+1/2)it}-\mathrm{e}^{-(n+1/2)it}}{\mathrm{e}^{it/2}-\mathrm{e}^{-it/2}} \\ &= \frac{\sin\big((n+1/2)t\big)}{\sin(t/2)}. \end{aligned}$$

Notons au passage que la limite de cette expression lorsque t tend vers 0 (ou vers $2j\pi$) est exactement $2n+1$, ce qui concorde avec la continuité de D_n, évidente directement puisque D_n est une somme finie de fonctions e_k. □

Définition 7.33 (Sommes de Fejér)**.** *Soit $f \in \mathrm{L}^1(\mathbb{T})$, on appelle* somme de Fejér *de rang n de f, la moyenne arithmétique des n premières sommes partielles de sa série de Fourier :*

$$\sigma_n(f) := \frac{1}{n}\sum_{j=0}^{n-1} S_j(f), \quad n \in \mathbb{N}^*. \tag{7.33}$$

À nouveau, on peut exprimer $\sigma_n(f)$ comme un produit de convolution

$$\sigma_n(f) = f * K_n = \frac{1}{2\pi}\int_{[-\pi,\pi]} f(t)K_n(\,.-t)\,\mathrm{d}\lambda(t) = \frac{1}{2\pi}\int_{[-\pi,\pi]} f(\,.-t)K_n(t)\,\mathrm{d}\lambda(t), \tag{7.34}$$

grâce au calcul élémentaire suivant qui exploite (7.32) :

$$\sigma_n(f) = \frac{1}{n}\sum_{j=0}^{n-1} f * D_j = f * \left(\frac{1}{n}\sum_{j=0}^{n-1} D_j\right).$$

Définition 7.34. *On appelle* noyau de Fejér *d'ordre n ($n \geq 1$) la fonction*

$$K_n := \frac{1}{n}\sum_{j=0}^{n-1} D_j.$$

Proposition 7.35 (Propriétés du noyau de Fejér)**.**

i) $K_n = \displaystyle\sum_{k=-n}^{k=n}\Big(1-\frac{|k|}{n}\Big)e_k$.

ii) $\forall f \in \mathrm{L}^1(\mathbb{T}), \quad \sigma_n(f) = \displaystyle\sum_{k=-n}^{k=n}\Big(1-\frac{|k|}{n}\Big)c_k(f)e_k$.

iii) $\forall n \in \mathbb{N}^*, \quad \displaystyle\int_{[-\pi,\pi]} K_n \,\mathrm{dm} = \frac{1}{2\pi}\int_{-\pi}^{\pi} K_n(t)\,\mathrm{d}t = 1.$

iv) Pour $t \in 2\pi\mathbb{Z}$, $K_n(t) = K_n(0) = n$ *et*

$$\forall t \notin 2\pi\mathbb{Z}, \quad K_n(t) = \frac{1}{n}\left(\frac{\sin(nt/2)}{\sin(t/2)}\right)^2.$$

En conséquence, K_n *est positif et compte-tenu de iii), c'est une densité de probabilité sur* $[-\pi,\pi]$ *par rapport à* m.

v) $\|K_n\|_1 = 1.$

vi) $\forall \delta \in]0,\pi], \quad I_n(\delta) := \displaystyle\int_{\{\delta \le |t| \le \pi\}} K_n \,\mathrm{dm} \xrightarrow[n\to+\infty]{} 0.$

Preuve. L'égalité i) s'obtient par la sommation triangulaire suivante :

$$\begin{aligned} nK_n = \sum_{j=0}^{n-1} D_j = \sum_{j=0}^{n-1}\sum_{|k|\le j} e_k &= \sum_{\substack{(j,k)\in\mathbb{N}\times\mathbb{Z}\\ |k|\le j\le n-1}} e_k \\ &= \sum_{|k|\le n-1} e_k \operatorname{card}\{j\in\mathbb{N};\ |k|\le j\le n-1\} \\ &= \sum_{|k|\le n-1}(n-|k|)e_k = \sum_{|k|\le n}(n-|k|)e_k. \end{aligned}$$

L'égalité ii) est une conséquence immédiate de i), (7.34) et (7.23).

Pour vérifier iii), on utilise simplement la définition de K_n, la proposition 7.32 a) et la linéarité de l'intégrale :

$$\int_{[-\pi,\pi]} K_n \,\mathrm{dm} = \int_{[-\pi,\pi]}\left(\frac{1}{n}\sum_{j=0}^{n-1} D_j\right)\mathrm{dm} = \frac{1}{n}\sum_{j=0}^{n-1}\int_{[-\pi,\pi]} D_j \,\mathrm{dm} = \frac{1}{n}\sum_{j=0}^{n-1} 1 = 1.$$

Vérifions maintenant la formule explicite iv) pour K_n. Si $t \in 2\pi\mathbb{Z}$, $K_n(t) = K_n(0)$ par 2π périodicité de K et le calcul de $K_n(0)$ est celui de la somme d'une progression arithmétique[3]

$$K_n(0) = \frac{1}{n}\sum_{j=0}^{n-1} D_j(0) = \frac{1}{n}\sum_{j=0}^{n-1}(2j+1) = \frac{1}{n}\times n\times\frac{1+(2n-1)}{2} = n.$$

Si $t \notin 2\pi\mathbb{Z}$, on utilise à nouveau un calcul de somme d'une progression géométrique de raison $\mathrm{e}^{it} \neq 1$. En effet, grâce à la définition de K_n et à la proposition 7.32 b), on peut

3. Il est bien connu que cette somme est égale au produit du nombre de termes multiplié par la moyenne du premier et du dernier termes.

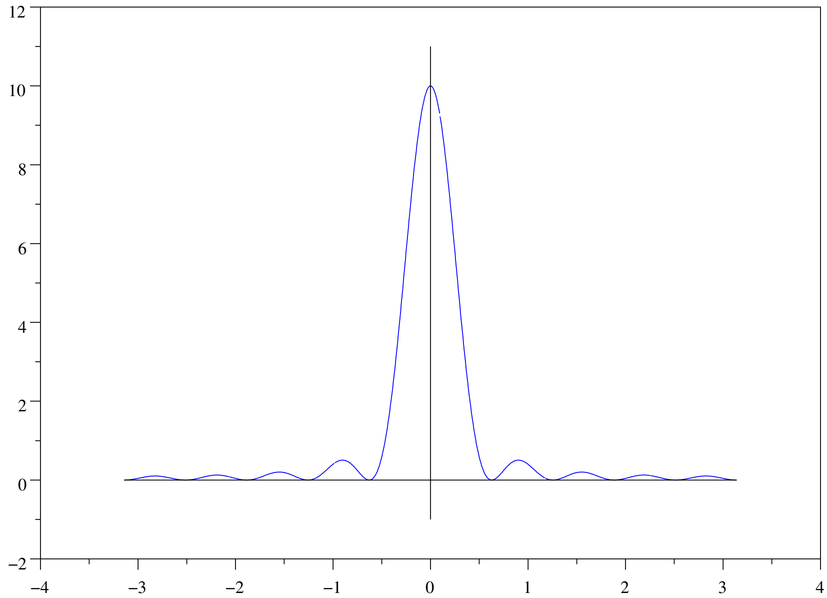


FIGURE 7.4 – Noyau de Fejér K_{10} sur $[-\pi, \pi]$

écrire :

$$\begin{aligned}
nK_n(t) = \sum_{j=0}^{n-1} D_j(t) = \sum_{j=0}^{n-1} \frac{\sin\bigl((j+1/2)t\bigr)}{\sin(t/2)} &= \frac{1}{\sin(t/2)} \sum_{j=0}^{n-1} \operatorname{Im}\bigl(\mathrm{e}^{ijt}\mathrm{e}^{it/2}\bigr) \\
&= \frac{1}{\sin(t/2)} \operatorname{Im}\left\{\frac{\mathrm{e}^{int}-1}{\mathrm{e}^{it}-1}\mathrm{e}^{it/2}\right\} \\
&= \frac{1}{\sin(t/2)} \operatorname{Im}\left\{\mathrm{e}^{int/2}\frac{\mathrm{e}^{int/2}-\mathrm{e}^{-int/2}}{\mathrm{e}^{it/2}-\mathrm{e}^{-it/2}}\right\} \\
&= \frac{1}{\sin(t/2)} \operatorname{Im}\left\{\mathrm{e}^{int/2}\frac{\sin(nt/2)}{\sin(t/2)}\right\} \\
&= \frac{\sin^2(nt/2)}{\sin^2(t/2)}.
\end{aligned}$$

L'égalité v) est une conséquence immédiate de iii) et de la positivité de K_n via iv).
Enfin pour établir vi), la parité et la continuité de K_n nous permettent d'écrire $I_n(\delta)$ comme l'intégrale de Riemann

$$I_n(\delta) = \frac{1}{\pi}\int_\delta^\pi K_n(t)\,\mathrm{d}t.$$

Lorsque t décrit $[\delta,\pi]$, $t/2$ décrit $[\delta/2,\pi/2]$ et par croissance de la fonction $\sin^2$ sur cet intervalle, on a la minoration :

$$\forall t\in[\delta,\pi],\quad \sin^2(t/2)\geq\sin^2(\delta/2).$$

Grâce à iv), on en déduit la majoration :

$$0\leq I_n(\delta) = \frac{1}{\pi}\int_\delta^\pi \frac{\sin^2(nt/2)}{n\sin^2(t/2)}\,\mathrm{d}t \leq \frac{1}{\pi}\int_\delta^\pi \frac{1}{n\sin^2(\delta/2)}\,\mathrm{d}t \leq \frac{1}{n\sin^2(\delta/2)}.$$

Ce majorant tend vers 0 quand n tend vers $+\infty$ ($\delta > 0$ restant fixé). Ceci achève la preuve de la proposition 7.35. □

Remarque 7.36. On peut donner une interprétation de la propriété vi) du noyau de Fejér en disant que la mesure de probabilité μ_n sur $[-\pi,\pi]$, de densité K_n par rapport à m, converge quand n tend vers l'infini vers la masse de Dirac au point 0. Nous reviendrons ultérieurement sur la question de la convergence des mesures.

7.2.5 Le théorème de Fejér

Théorème 7.37 (Fejér)**.** *Si f est continue 2π-périodique, ses sommes de Fejér $\sigma_n(f)$ convergent uniformément vers f sur $[-\pi,\pi]$ (donc aussi sur $\mathbb{R}$) :*

$$\forall f\in\mathcal{C}(\mathbb{T}),\quad \|\sigma_n(f)-f\|_\infty \xrightarrow[n\to+\infty]{} 0.$$

Preuve. Grâce à (7.34) et à la proposition 7.35 iii), on peut écrire la différence $\sigma_n(f)-f$ sous forme intégrale :

$$\sigma_n(f)(x)-f(x) = (f*K_n)(x)-f(x)\int_{[-\pi,\pi]} K_n\,\mathrm{dm} = \int_{[-\pi,\pi]} \big(f(x-t)-f(x)\big)K_n(t)\,\mathrm{dm}(t).$$

On en déduit la majoration :

$$\forall x\in[-\pi,\pi],\quad \Big|\sigma_n(f)(x)-f(x)\Big| \leq \int_{[-\pi,\pi]} |f(x-t)-f(x)|K_n(t)\,\mathrm{dm}(t). \tag{7.35}$$

Par la proposition 7.19, la fonction continue 2π-périodique f est *uniformément* continue sur tout $\mathbb{R}$:

$$\forall\varepsilon>0,\ \exists\delta>0,\ |u-v|\leq\delta \Rightarrow |f(u)-f(v)|\leq\varepsilon. \tag{7.36}$$

Fixons $\varepsilon>0$. On peut toujours imposer au δ fourni par (7.36) d'être inférieur à π. Découpons l'intégrale dans (7.35) selon $|t|<\delta$ et $\delta\leq|t|\leq\pi$. On majore $\int_{|t|<\delta}$ en appliquant (7.36) avec $u=x-t$ et $v=x$. Pour l'intégrale $\int_{\delta\leq|t|\leq\pi}$, il n'y a pas d'espoir que $|f(x-t)-f(x)|$ soit petit et on le majore brutalement [4] par $2\|f\|_\infty$.

$$\begin{aligned}
\Big|\sigma_n(f)(x)-f(x)\Big| &\leq \int_{[-\delta,\delta]} \varepsilon K_n(t)\,\mathrm{dm}(t) + \int_{\delta\leq|t|\leq\pi} 2\|f\|_\infty K_n(t)\,\mathrm{dm}(t)\\
&\leq \varepsilon\int_{[-\pi,\pi]} K_n\,\mathrm{dm} + 2\|f\|_\infty I_n(\delta) \qquad (7.37)\\
&= \varepsilon + 2\|f\|_\infty I_n(\delta). \qquad (7.38)
\end{aligned}$$

Le passage de (7.37) à (7.38) repose sur la propriété iii) du noyau de Fejér. Grâce à la propriété vi), il existe un entier n_0, tel que pour tout $n\geq n_0$, $2\|f\|_\infty I_n(\delta)\leq\varepsilon$. Ce n_0 dépend de f et δ (donc de f et ε), mais pas de x. On aboutit ainsi à

$$\forall\varepsilon>0,\ \exists n_0=n_0(f,\varepsilon),\ \forall n\geq n_0,\ \forall x\in[-\pi,\pi],\qquad \Big|\sigma_n(f)(x)-f(x)\Big|\leq 2\varepsilon,$$

ce qui signifie que $\sigma_n(f)$ converge vers f, *uniformément* sur $[-\pi,\pi]$. □

Parmi les multiples retombées du théorème de Fejér, figure le résultat de densité que nous avons utilisé par anticipation dans la preuve du théorème 7.25.

Corollaire 7.38. *La famille* $\mathsf{E}=\{e_k;\ k\in\mathbb{Z}\}$ *est totale dans l'espace* $\big(\mathcal{C}(\mathbb{T}),\|\ \|_\infty\big)$.

Preuve. Par le théorème de Fejér, pour f quelconque dans $\mathcal{C}(\mathbb{T})$ et tout $\varepsilon>0$, on peut trouver un entier n dépendant de f et ε tel que $\|\sigma_n(f)-f\|_\infty<\varepsilon$. Il suffit alors de remarquer que $\sigma_n(f)$ est une combinaison linéaire finie d'éléments de E. □

4. Mais uniformément en x.

Chapitre 8

Lois des grands nombres

Soit $(X_k)_{k\geq 1}$ une suite de variables aléatoires réelles définies sur le même espace probabilisé $(\Omega, \mathcal{F}, \mathbf{P})$. On étudie dans ce chapitre le comportement asymptotique de la suite $(M_n)_{n\geq 1}$ des moyennes arithmétiques :

$$M_n := \frac{1}{n} S_n = \frac{1}{n} \sum_{k=1}^{n} X_k.$$

(Re)lecture conseillée : [ICP], chapitre 6.

8.1 Convergences de suites de variables aléatoires

Commençons par faire le point sur les divers modes de convergences pour une suite $(Y_n)_{n\geq 1}$ de variables aléatoires réelles. Nous connaissons déjà la convergence presque sûre qui est l'autre nom de la convergence $\mathbf{P}$-presque partout sur Ω et les convergences L^p $(1 \leq p < +\infty)$:

$$Y_n \xrightarrow[n\to+\infty]{\mathrm{L}^p} Y \quad \Leftrightarrow \quad \|Y_n - Y\|_p \xrightarrow[n\to+\infty]{} 0 \quad \Leftrightarrow \quad \mathbf{E}|Y_n - Y|^p \xrightarrow[n\to+\infty]{} 0.$$

Il est temps d'y ajouter la convergence *en probabilité*.

Définition 8.1. *La suite Y_n converge en probabilité vers Y (notation $Y_n \xrightarrow[n\to+\infty]{\mathrm{Pr}} Y$) si*

$$\forall \varepsilon > 0, \quad \mathbf{P}\Big(|Y_n - Y| \geq \varepsilon\Big) \xrightarrow[n\to+\infty]{} 0.$$

Grâce à l'inégalité de Markov, on voit immédiatement que cette nouvelle convergence est impliquée par n'importe quelle convergence L^p $(1 \leq p < +\infty)$:

$$\forall \varepsilon > 0, \quad \mathbf{P}\Big(|Y_n - Y| \geq \varepsilon\Big) \leq \frac{1}{\varepsilon^p} \mathbf{E}|Y_n - Y|^p.$$

Comme la convergence L^∞ implique toutes les convergences L^p (cf. théorème 6.18), elle implique aussi la convergence en probabilité.

Proposition 8.2. *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

$$\Omega'_\varepsilon := \{\omega \in \Omega;\ \exists k_0 = k_0(\omega),\ \forall n \geq k_0,\ |Y_n(\omega) - Y(\omega)| < \varepsilon\}.$$

C'est bien un évènement (i.e. $\Omega'_\varepsilon \in \mathcal{F}$) puisqu'il s'écrit

$$\Omega'_\varepsilon = \bigcup_{k\in\mathbb{N}^*} \bigcap_{n\geq k} \{|Y_n - Y| < \varepsilon\}.$$

De plus Ω'_ε contient Ω', donc $\mathbf{P}(\Omega'_\varepsilon) = 1$. Pour tout $k \geq 1$, notons

$$A_k := \{\omega \in \Omega;\ \forall n \geq k,\ |Y_n(\omega) - Y(\omega)| < \varepsilon\} = \bigcap_{n\geq k} \{|Y_n - Y| < \varepsilon\}.$$

La suite $(A_k)_{k\geq 1}$ est clairement croissante pour l'inclusion et sa réunion est Ω'_ε. Par continuité séquentielle croissante de $\mathbf{P}$, on a donc $\mathbf{P}(A_k) \uparrow \mathbf{P}(\Omega'_\varepsilon) = 1$ $(k \to +\infty)$. Par conséquent,

$$\forall \delta > 0,\ \exists k_1, \quad P(A_{k_1}) > 1 - \delta.$$

Pour tout $n \geq k_1$, l'évènement $\{|Y_n - Y| < \varepsilon\}$ contient A_{k_1}, d'où

$$\forall n \geq k_1, \quad \mathbf{P}(|Y_n - Y| < \varepsilon) > 1 - \delta$$

et en passant à l'évènement complémentaire

$$\forall n \geq k_1, \quad \mathbf{P}(|Y_n - Y| \geq \varepsilon) < \delta.$$

Ceci établit la convergence vers 0 de $\mathbf{P}(|Y_n - Y| \geq \varepsilon)$. Comme ε était quelconque, on a bien convergence en probabilité de Y_n vers Y. □

Remarque 8.3. La convergence en probabilité n'implique pas la convergence presque-sûre. Voici un contre-exemple. On prend comme espace probabilisé $(]0,1], \mathrm{Bor}(]0,1]), \lambda)$. On définit les Y_n comme suit :

$$\begin{aligned}
&Y_1 = \mathbf{1}_{]0,1]},\\
&Y_2 = \mathbf{1}_{]0,1/2]}, Y_3 = \mathbf{1}_{]1/2,1]},\\
&Y_4 = \mathbf{1}_{]0,1/4]}, Y_5 = \mathbf{1}_{]1/4,1/2]}, Y_6 = \mathbf{1}_{]1/2,3/4]}, Y_7 = \mathbf{1}_{]3/4,1]},\\
&Y_8 = \mathbf{1}_{]0,1/8]}, Y_9 = \mathbf{1}_{]1/8,1/4]}, \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots, Y_{15} = \mathbf{1}_{]7/8,1]},\\
&Y_{16} = \mathbf{1}_{]0,1/16]}, \ldots\ldots
\end{aligned}$$

Le lecteur qui ne se satisferait pas de cette définition informelle peut toujours s'exercer à trouver une formule explicite pour Y_n. Sans entrer dans ces détails techniques, on peut facilement se convaincre de deux choses :

1. Pour tout $\omega \in]0,1]$, la suite de « bits » $(Y_n(\omega))_{n\geq 1}$ est formée d'une infinité de 0 et d'une infinité de 1. Elle ne peut donc converger (sa limite inférieure vaut 0 et sa limite supérieure 1). Ainsi non seulement on n'a pas de convergence presque sûre de Y_n, mais en plus $Y_n(\omega)$ ne converge pour *aucun* $\omega \in \Omega$.
2. Pour $0 < \varepsilon < 1$, $\mathbf{P}(|Y_n - 0| > \varepsilon) = \mathbf{P}(Y_n = 1) = \lambda(I_n)$, en notant I_n l'intervalle dyadique dont Y_n est l'indicatrice. La longueur de cet intervalle tend vers zéro quand n tend vers l'infini (à la même vitesse que l'inverse du logarithme en base deux de n). Donc Y_n converge vers 0 en probabilité.

Il est toutefois 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. C'est l'objet du résultat suivant.

Proposition 8.4. *Soient Y et $(Y_n)_{n\geq 1}$ des variables aléatoires réelles définies sur le même espace probabilisé $(\Omega, \mathcal{F}, \mathbf{P})$ et vérifiant*

$$\forall \varepsilon > 0, \quad \sum_{n=1}^{+\infty} \mathbf{P}(|Y_n - Y| \geq \varepsilon) < +\infty. \tag{8.1}$$

Alors Y_n converge presque sûrement vers Y.

La condition (8.1) implique évidemment la convergence en probabilité puisque le terme général de la série doit tendre vers 0. On dit d'une suite $(Y_n)_{n\geq 1}$ vérifiant (8.1) qu'elle converge *presque complètement* vers Y.

Preuve. Posons $B_{n,\varepsilon} := \{|Y_n - Y| > \varepsilon\}$ et $A_\varepsilon := \limsup B_{n,\varepsilon}$. Par le premier lemme de Borel-Cantelli, on déduit de (8.1) que $\mathbf{P}(A_\varepsilon) = 0$, d'où $\mathbf{P}(A_\varepsilon^c) = 1$. Notons que

$$A_\varepsilon^c = \Big\{\omega \in \Omega;\ \exists n_0 = n_0(\omega, \varepsilon),\ \forall n \geq n_0,\ |Y_n(\omega) - Y(\omega)| < \varepsilon\Big\}.$$

Discrétisons le ε en posant par exemple $\varepsilon_i = 2^{-i}$ et posons

$$\Omega' := \bigcap_{i\geq 1} A_{\varepsilon_i}^c.$$

Alors Ω' est un évènement de probabilité 1 comme intersection dénombrable d'évènements de probabilité 1. De plus,

$$\forall \omega \in \Omega', \quad \forall i \geq 1, \exists n_0 = n_0(\omega, \varepsilon_i),\ \forall n \geq n_0,\ |Y_n(\omega) - Y(\omega)| < \varepsilon_i\Big\}.$$

On en déduit immédiatement que si $\omega \in \Omega'$, $Y_n(\omega)$ converge vers $Y(\omega)$. Comme $\mathbf{P}(\Omega') = 1$, ceci entraîne la convergence presque sûre de Y_n vers Y. □

Corollaire 8.5. *Si la suite $(Y_n)_{n\geq 1}$ converge en probabilité vers Y, on peut en extraire une sous-suite $(Y_{n_i})_{i\geq 1}$ qui converge presque sûrement vers Y.*

Preuve. Notons encore $\varepsilon_i = 2^{-i}$. La convergence en probabilité de Y_n vers Y implique pour tout $i \geq 1$, la convergence de $\mathbf{P}(|Y_n - Y| > \varepsilon_i)$ vers 0 quand n tend vers l'infini. On en déduit l'existence d'une suite strictement croissante d'indices n_i telle que

$$\forall i \geq 1, \quad \mathbf{P}\Big(|Y_{n_i} - Y| \geq \varepsilon_i\Big) \leq \frac{1}{i^2}.$$

Vérifions maintenant que

$$\forall \varepsilon > 0, \quad \sum_{i=1}^{+\infty} \mathbf{P}(|Y_{n_i} - Y| \geq \varepsilon) < +\infty.$$

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 $i \geq i_0$, $\varepsilon_i < \varepsilon$. Pour $i \geq i_0$, on a donc $\{|Y_{n_i} - Y| \geq \varepsilon\} \subset \{|Y_{n_i} - Y| \geq \varepsilon_i\}$ 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. $\square$

Le diagramme de la figure 8.1 résume les relations entre les différents 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 signifient l'existence d'une sous-suite convergente selon le mode de la case d'arrivée. La convergence en loi sera étudiée ultérieurement.

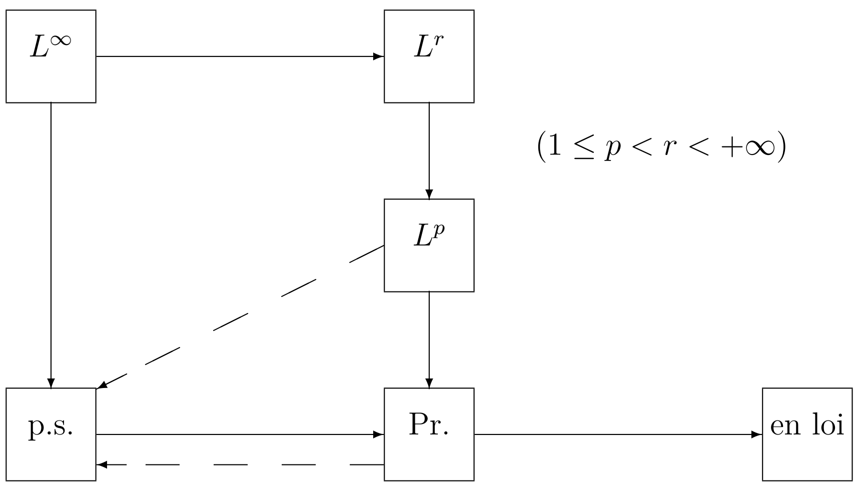


FIGURE 8.1 – Diagramme des convergences des suites de v.a.

8.2 Loi faible des grands nombres

Rappelons que si X est de carré intégrable ($\mathbf{E}X^2 < +\infty$), sa variance est définie par

$$\operatorname{Var} X := \mathbf{E}(X - \mathbf{E}X)^2.$$

Si les X_k sont deux à deux *non corrélées* (ce qui a lieu en particulier lorsqu'elles sont indépendantes), on a l'égalité

$$\operatorname{Var} S_n = \sum_{k=1}^{n} \operatorname{Var} X_k. \tag{8.2}$$

Proposition 8.6 (Inégalité de Bienaymé-Tchebycheff)**.** *Si les X_k sont de carré intégrable et deux à deux non corrélées,*

$$\forall t > 0, \quad \mathbf{P}\Big(\Big| \sum_{k=1}^{n} (X_k - \mathbf{E}X_k) \Big| \geq t \Big) \leq \frac{1}{t^2} \sum_{k=1}^{n} \operatorname{Var} X_k. \tag{8.3}$$

Preuve. Il suffit d'écrire :

$$\begin{aligned} \mathbf{P}\Big(|S_n - \mathbf{E}S_n| \geq t \Big) = \mathbf{P}\Big(|S_n - \mathbf{E}S_n|^2 \geq t^2 \Big) &\leq \frac{1}{t^2} \mathbf{E}(S_n - \mathbf{E}S_n)^2 \\ &= \frac{1}{t^2} \operatorname{Var} S_n, \end{aligned} \tag{8.4}$$

où l'inégalité dans (8.4) est l'inégalité de Markov appliquée à la variable aléatoire positive $|S_n - \mathbf{E}S_n|^2$. On conclut avec (8.2). □

Théorème 8.7 (Loi faible des G.N.)**.** *Si les X_k sont de même loi, de carré intégrable et deux à deux non corrélées, on a la convergence en probabilité :*

$$\frac{1}{n} \sum_{k=1}^{n} X_k \xrightarrow[n \to +\infty]{\mathrm{Pr}} \mathbf{E}X_1. \tag{8.5}$$

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, (8.2) 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 :

$$\forall t > 0, \quad \mathbf{P}\Big(|S_n - \mathbf{E}S_n| \geq t \Big) = \mathbf{P}\Big(|S_n - n\mathbf{E}X_1| \geq t \Big) \leq \frac{n \operatorname{Var} X_1}{t^2}.$$

Posant $t = n\varepsilon$, on en déduit :

$$\forall \varepsilon > 0,\ \forall n \in \mathbb{N}^*, \quad \mathbf{P}\Big(|S_n - n\mathbf{E}X_1| \geq n\varepsilon \Big) = \mathbf{P}\Big(\Big| \frac{S_n}{n} - \mathbf{E}X_1 \Big| \geq \varepsilon \Big) \leq \frac{n \operatorname{Var} X_1}{n^2 \varepsilon^2} = \frac{\operatorname{Var} X_1}{n\varepsilon^2}.$$

Pour tout $\varepsilon > 0$ fixé, on a ainsi

$$\mathbf{P}\Big(\Big| \frac{S_n}{n} - \mathbf{E}X_1 \Big| \geq \varepsilon \Big) \leq \frac{\operatorname{Var} X_1}{n\varepsilon^2} \xrightarrow[n \to +\infty]{} 0,$$

ce qui traduit exactement la convergence en probabilité de la suite de variables aléatoires S_n/n vers la variable aléatoire constante $\mathbf{E}X_1$. □

8.3 Lois fortes des grands nombres

Théorème 8.8 (Inégalité de Kolmogorov)**.** *Soient $X_1, \dots, X_n$ des variables aléatoires réelles indépendantes, de carré intégrable et d'espérance nulle et $S_k := \sum_{i=1}^k X_i$. Alors*

$$\forall t > 0, \quad \mathbf{P}\Bigl(\max_{1 \leq k \leq n} |S_k| \geq t\Bigr) \leq \frac{1}{t^2} \operatorname{Var} S_n. \tag{8.6}$$

Remarque 8.9. Si les X_k ne sont pas centrées ($\mathbf{E}X_k \neq 0$), en posant $X'_k := X_k - \mathbf{E}X_k$, en appliquant (8.6) aux X'_k et en notant que $\operatorname{Var} X_k = \operatorname{Var} X'_k$, on obtient la version plus générale de l'inégalité de Kolmogorov :

$$\forall t > 0, \quad \mathbf{P}\Bigl(\max_{1 \leq k \leq n} |S_k - \mathbf{E}S_k| \geq t\Bigr) \leq \frac{1}{t^2} \operatorname{Var} S_n. \tag{8.7}$$

Remarque 8.10. L'inégalité de Kolmogorov ressemble formellement à celle de Bienaymé-Tchebycheff. Elle est cependant beaucoup plus puissante, puisqu'elle permet de contrôler en probabilité les déviations de *toute la suite* finie $(S_k - \mathbf{E}S_k)_{1 \leq k \leq n}$ au lieu de seulement son dernier terme $S_n - \mathbf{E}S_n$ pour Bienaymé-Tchebycheff. L'hypothèse est aussi plus restrictive (indépendance mutuelle au lieu de non-corrélation deux à deux).

Preuve du théorème 8.8. Définissons les évènements

$$A_1 := \{|S_1| \geq t\} = \{|X_1| \geq t\} \quad \text{et} \quad A_k := \{|S_k| \geq t\} \cap \Bigl(\bigcap_{j=1}^{k-1} \{|S_j| < t\}\Bigr) \quad (2 \leq k \leq n).$$

Autrement dit $\omega \in A_k$ (l'évènement élémentaire ω réalise l'évènement A_k) si et seulement si la suite finie $\bigl(|S_j(\omega)|\bigr)_{1 \leq j \leq n}$ atteint ou dépasse t la première fois pour $j = k$. Avec cette interprétation, il est clair que les A_k sont deux à deux disjoints et que :

$$A := \bigcup_{k=1}^n A_k = \Bigl\{\max_{1 \leq k \leq n} |S_k| \geq t\Bigr\}. \tag{8.8}$$

On en déduit par croissance et additivité de l'intégrale (ou de la mesure de densité S_n^2 par rapport à $\mathbf{P}$) la minoration :

$$\mathbf{E}S_n^2 = \int_\Omega S_n^2 \,\mathrm{d}\mathbf{P} \geq \int_A S_n^2 \,\mathrm{d}\mathbf{P} = \sum_{k=1}^n \int_{A_k} S_n^2 \,\mathrm{d}\mathbf{P}. \tag{8.9}$$

En intégrant sur A_k l'inégalité élémentaire

$$S_n^2 = \bigl[S_k + (S_n - S_k)\bigr]^2 \geq S_k^2 + 2S_k(S_n - S_k),$$

on obtient

$$\int_{A_k} S_n^2 \,\mathrm{d}\mathbf{P} \geq \int_{A_k} S_k^2 \,\mathrm{d}\mathbf{P} + 2\int_{A_k} S_k(S_n - S_k) \,\mathrm{d}\mathbf{P}. \tag{8.10}$$

On remarque alors que $\int_{A_k} S_k(S_n - S_k)\,\mathrm{d}\mathbf{P}$ peut s'écrire $\mathbf{E}(YZ)$ avec $Y := \mathbf{1}_{A_k} S_k$ et $Z := S_n - S_k$. Ces deux variables aléatoires sont de carré intégrable comme combinaisons linéaires de variables aléatoires de carré intégrable. La première Y est une fonction mesurable du seul vecteur aléatoire $(X_1, \dots, X_k)$ tandis que Z est fonction mesurable du seul vecteur aléatoire $(X_{k+1}, \dots, X_n)$. Par conséquent Y et Z sont indépendantes (corollaire 5.41) et $\mathbf{E}(YZ) = \mathbf{E}Y\mathbf{E}Z$ (proposition 5.43). Or

$$\mathbf{E}Z = \mathbf{E}(S_n - S_k) = \mathbf{E} \sum_{j=k+1}^{n} X_j = \sum_{j=k+1}^{n} \mathbf{E}X_j = 0,$$

d'où $\int_{A_k} S_k(S_n - S_k)\,\mathrm{d}\mathbf{P} = 0$ (résultat évident directement dans le cas particulier où $k = n$). En reportant ceci dans (8.10), on en déduit compte tenu de (8.9) :

$$\operatorname{Var} S_n = \mathbf{E}S_n^2 \geq \sum_{k=1}^{n} \int_{A_k} S_k^2\,\mathrm{d}\mathbf{P}. \tag{8.11}$$

Vu la définition de A_k, on a sur cet évènement $|S_k| \geq t$, d'où $\int_{A_k} S_k^2\,\mathrm{d}\mathbf{P} \geq t^2\mathbf{P}(A_k)$. En injectant cette minoration dans (8.11), en utilisant (8.8) et en rappelant que les A_k sont deux à deux disjoints, on aboutit à :

$$\operatorname{Var} S_n \geq \sum_{k=1}^{n} t^2\mathbf{P}(A_k) = t^2\mathbf{P}\Bigl(\bigcup_{k=1}^{n} A_k\Bigr) = t^2\mathbf{P}(A).$$

L'inégalité $\mathbf{P}(A) \leq t^{-2} \operatorname{Var} S_n$ ainsi établie est exactement (8.6). □

Théorème 8.11 (Loi forte des grands nombres de Kolmogorov)**.** *On suppose que $(X_k)_{k\geq 1}$ est une suite de variables aléatoires réelles indépendantes, de carré intégrable, d'espérance nulle et que*

$$\sum_{j=1}^{+\infty} \frac{\operatorname{Var} X_j}{j^2} < +\infty. \tag{8.12}$$

Alors

$$\frac{1}{n}\sum_{k=1}^{n} X_k \xrightarrow[n\to+\infty]{\text{p.s.}} 0. \tag{8.13}$$

Remarque 8.12. Si les X_k ne sont pas centrées, en appliquant (8.13) aux $X'_k := X_k - \mathbf{E}X_k$, on obtient

$$\frac{1}{n}S_n - \frac{1}{n}\sum_{k=1}^{n} \mathbf{E}X_k \xrightarrow[n\to+\infty]{\text{p.s.}} 0. \tag{8.14}$$

Ce résultat est intéressant notamment dans le cas où la suite déterministe $(\mathbf{E}X_n)_{n\geq 1}$ a une limite finie ℓ quand n tend vers l'infini. En effet le théorème de Césaro nous donne alors la convergence vers ℓ de $n^{-1}\sum_{k=1}^{n} \mathbf{E}X_k$, d'où la convergence presque sûre de $n^{-1}S_n$ vers ℓ. En particulier si les X_k ont même espérance, $n^{-1}S_n$ converge p.s. vers $\mathbf{E}X_1$.

Démonstration du théorème 8.11. Une des difficultés lorsque l'on veut établir une convergence presque sûre est la gestion du « $\forall\varepsilon > 0$ » qui induit une intersection non dénombrable. Nous allons procéder en trois étapes, la première étant consacrée à la résolution de ce problème.

Première étape : réduction à un $\varepsilon > 0$ fixé. Supposons déjà prouvé que pour tout $\varepsilon > 0$,

$$\mathbf{P}\Big(\limsup_{n\to+\infty}\Big\{\Big|\frac{S_n}{n}\Big| \geq \varepsilon\Big\}\Big) = 0, \tag{8.15}$$

autrement dit qu'il y a une probabilité nulle que la suite (S_n/n) ait une infinité de termes hors de l'intervalle $]-\varepsilon,+\varepsilon[$. L'évènement contraire Ω_ε de cette limite supérieure s'énonce « à partir d'un certain rang (aléatoire) fini, plus aucun terme de la suite (S_n/n) n'est hors de $]-\varepsilon,+\varepsilon[$ », ce qui s'écrit encore :

$$\Omega_\varepsilon := \Big(\limsup_{n\to+\infty}\Big\{\Big|\frac{S_n}{n}\Big| \geq \varepsilon\Big\}\Big)^c = \Big\{\omega\in\Omega;\ \exists n_0 = n_0(\omega,\varepsilon),\ \forall n \geq n_0,\ \Big|\frac{S_n(\omega)}{n}\Big| < \varepsilon\Big\}.$$

D'après (8.15), on a $\mathbf{P}(\Omega_\varepsilon) = 1$. Discrétisons alors le ε en le remplaçant par une suite $(\varepsilon_i) \downarrow 0$, disons $\varepsilon_i := 2^{-i}$. Posons

$$\Omega' := \bigcap_{i\in\mathbb{N}} \Omega_{\varepsilon_i}.$$

Cet évènement est de probabilité 1 puisque :

$$\mathbf{P}(\Omega'^c) = \mathbf{P}\Big(\bigcup_{i\in\mathbb{N}} \Omega_{\varepsilon_i}^c\Big) \leq \sum_{i\in\mathbb{N}} \mathbf{P}(\Omega_{\varepsilon_i}^c) = 0.$$

Or par définition de Ω', tout $\omega \in \Omega'$ appartient à *chacun* des Ω_{ε_i}, autrement dit,

$$\forall \omega\in\Omega',\ \forall i\in\mathbb{N},\ \exists n_0 = n_0(\omega,\varepsilon_i),\ \forall n\geq n_0,\quad \Big|\frac{S_n(\omega)}{n}\Big| < \varepsilon_i.$$

On en déduit que pour tout $\omega\in\Omega'$, $n^{-1}S_n(\omega)$ tend vers zéro quand n tend vers $+\infty$. Donc Ω' est inclus dans l'évènement $\{n^{-1}S_n \to 0\}$ et comme $\mathbf{P}(\Omega') = 1$, $\mathbf{P}(n^{-1}S_n \to 0) = 1$, ce qui signifie la convergence presque sûre de $n^{-1}S_n$ vers zéro [1]. Ainsi le théorème sera démontré si on prouve que pour un $\varepsilon > 0$ quelconque fixé, on a l'égalité (8.15). □

Deuxième étape : réduction de (8.15) à une condition de Borel-Cantelli. On travaille désormais avec un $\varepsilon > 0$ fixé. On va se ramener au premier lemme de Borel-Cantelli en utilisant une partition de $\mathbb{N}^*$ en « octaves » $[2^k, 2^{k+1}[$. On remarque en effet que l'inégalité $n^{-1}|S_n(\omega)| \geq \varepsilon$ se réalise pour une infinité d'indices n si et seulement si ω appartient pour une infinité d'indices k, à l'évènement B_k défini par

$$B_k := \Big\{\omega\in\Omega;\ \exists n = n(\omega) \in [2^k, 2^{k+1}[,\ |S_n(\omega)| \geq n\varepsilon\Big\},\quad k\in\mathbb{N}.$$

Cette équivalence se traduit par l'égalité d'évènements :

$$\limsup_{n\to+\infty}\Big\{\Big|\frac{S_n}{n}\Big| \geq \varepsilon\Big\} = \limsup_{k\to+\infty} B_k.$$

1. En écrivant $\mathbf{P}(n^{-1}S_n \to 0)$, on a admis implicitement que l'ensemble $\Omega'' := \{\omega\in\Omega;\ n^{-1}S_n(\omega) \to 0\}$ est bien un *évènement*, c'est à dire appartient à la tribu $\mathcal{F}$. Une façon rapide de le justifier est d'utiliser la mesurabilité de la variable aléatoire $Y := \limsup_{n\to+\infty} |n^{-1}S_n|$ et l'équivalence entre $n^{-1}S_n \to 0$ et $Y = 0$, d'où $\Omega'' = Y^{-1}(\{0\})$.

Cette égalité et le premier lemme de Borel-Cantelli montrent que pour établir (8.15), il suffit de prouver la convergence de série :

$$\sum_{k=0}^{+\infty} \mathbf{P}(B_k) < +\infty. \tag{8.16}$$

$\square$

Troisième étape : preuve de (8.16) par l'inégalité de Kolmogorov. De l'écriture de l'évènement B_k sous la forme

$$B_k = \bigcup_{2^k \leq n < 2^{k+1}} \Big\{ |S_n| \geq n\varepsilon \Big\},$$

on déduit immédiatement les inclusions :

$$B_k \subset \bigcup_{2^k \leq n < 2^{k+1}} \Big\{ |S_n| \geq 2^k\varepsilon \Big\} = \Big\{ \max_{2^k \leq n < 2^{k+1}} |S_n| \geq 2^k\varepsilon \Big\} \subset \Big\{ \max_{1 \leq j \leq 2^{k+1}} |S_j| \geq 2^k\varepsilon \Big\}.$$

Ces inclusions nous permettent de majorer $\mathbf{P}(B_k)$ grâce à l'inégalité de Kolmogorov (c'est seulement maintenant que nous avons besoin de l'indépendance des X_j et de leur appartenance à $\mathrm{L}^2(\mathbf{P})$) :

$$\mathbf{P}(B_k) \leq \mathbf{P}\Big(\max_{1 \leq j \leq 2^{k+1}} |S_j| \geq 2^k\varepsilon \Big) \leq \frac{1}{\varepsilon^2 2^{2k}} \operatorname{Var}\Big(S_{2^{k+1}} \Big) = \frac{1}{\varepsilon^2 2^{2k}} \sum_{j=1}^{2^{k+1}} \operatorname{Var} X_j.$$

Cette inégalité nous conduit à la sommation triangulaire suivante pour contrôler la somme de la série des $\mathbf{P}(B_k)$.

$$\begin{aligned}
\sum_{k=0}^{+\infty} \mathbf{P}(B_k) &\leq \frac{1}{\varepsilon^2} \sum_{k=0}^{+\infty} \frac{1}{2^{2k}} \sum_{j=1}^{2^{k+1}} \operatorname{Var} X_j \\
&= \frac{1}{\varepsilon^2} \sum_{k=0}^{+\infty} \sum_{j=1}^{+\infty} \frac{1}{2^{2k}} \operatorname{Var} X_j \, \mathbf{1}_{\{1 \leq j \leq 2^{k+1}\}} \\
&= \frac{1}{\varepsilon^2} \sum_{j=1}^{+\infty} \operatorname{Var} X_j \sum_{k=0}^{+\infty} \frac{1}{2^{2k}} \mathbf{1}_{\{2^{k+1} \geq j\}} \\
&= \frac{1}{\varepsilon^2} \sum_{j=1}^{+\infty} \operatorname{Var} X_j \sum_{k=m_j}^{+\infty} \frac{1}{2^{2k}},
\end{aligned}$$

où l'on a posé $m_j := \min\Big\{ k \in \mathbb{N};\ 2^{k+1} \geq j \Big\}$. Un simple calcul de série géométrique nous donne :

$$\sum_{k=m_j}^{+\infty} \frac{1}{2^{2k}} = \sum_{k=m_j}^{+\infty} \frac{1}{4^k} = \frac{1}{4^{m_j}} \sum_{l=0}^{+\infty} \frac{1}{4^l} = \frac{1}{4^{m_j}} \frac{1}{1 - \frac{1}{4}} = \frac{4}{3 \times 4^{m_j}}.$$

De la définition de m_j, on tire les inégalités $2^{m_j} \geq j/2$, puis $4^{m_j} \geq j^2/4$, d'où

$$\frac{4}{3 \times 4^{m_j}} \leq \frac{16}{3j^2}.$$

Finalement,

$$\sum_{k=0}^{+\infty} \mathbf{P}(B_k) \leq \frac{16}{3\varepsilon^2} \sum_{j=1}^{+\infty} \frac{\operatorname{Var} X_j}{j^2},$$

ce qui établit (8.16) grâce à l'hypothèse (8.12). □

Par la deuxième étape, on en déduit (8.15) et comme ε était quelconque, ceci entraîne la convergence presque sûre vers 0 de S_n/n, d'après la première étape. Le théorème est maintenant complètement démontré. □

Théorème 8.13 (Loi forte des grands nombres de Khintchine)**.** *On suppose les X_k indépendantes,* de même loi *et* $\mathbf{E}|X_1| < +\infty$*. Alors*

$$\frac{1}{n}\sum_{k=1}^{n} X_k \xrightarrow[n\to+\infty]{\text{p.s.}} \mathbf{E}X_1. \tag{8.17}$$

Preuve. Quitte à remplacer X_k par $X'_k := X_k - \mathbf{E}X_k = X_k - \mathbf{E}X_1$, il suffit de faire la preuve dans le cas où $\mathbf{E}X_1 = 0$. On utilise une méthode de *troncature* permettant de remplacer les X_k par des variables aléatoires *bornées* Y_k (les bornes dépendant de k), donc ayant une variance et relevant ainsi de la loi forte des grands nombres de Kolmogorov. On pose pour cela

$$Y_k(\omega) := \begin{cases} X_k(\omega) & \text{si } |X_k(\omega)| \leq k, \\ 0 & \text{si } |X_k(\omega)| > k. \end{cases}$$

Autrement dit,

$$Y_k = X_k \mathbf{1}_{\{|X_k| \leq k\}}.$$

Bien que les X_k aient même loi, il n'en va pas de même des Y_k, car la troncature dépend de k. Remarquons cependant que

$$m_k := \mathbf{E}\Big(X_k \mathbf{1}_{\{|X_k|\leq k\}}\Big) = \mathbf{E}\Big(X_1 \mathbf{1}_{\{|X_1|\leq k\}}\Big) \xrightarrow[k\to+\infty]{} \mathbf{E}X_1, \tag{8.18}$$

par convergence dominée, grâce à l'hypothèse $\mathbf{E}|X_1| < +\infty$.

Première étape : réduction à la l.f.g.n. pour (Y_k)*.* Nous commençons par montrer que si $\mathbf{E}X_1 = 0$, (8.17) est impliquée par :

$$\frac{1}{n}\sum_{k=1}^{n} Y_k \xrightarrow[n\to+\infty]{\text{p.s.}} 0. \tag{8.19}$$

L'idée est d'utiliser le premier lemme de Borel-Cantelli pour prouver que les suites $(X_k)_{k\geq 1}$ et $(Y_k)_{k\geq 1}$ ont presque sûrement tous leurs termes égaux à partir d'un certain rang (aléatoire). Les moyennes arithmétiques auront alors p.s. le même comportement asymptotique. Ceci nous amène à étudier la convergence de la série $\sum_{k=1}^{+\infty} \mathbf{P}(X_k \neq Y_k)$. On note d'abord que

$$\sum_{k=1}^{+\infty} \mathbf{P}(X_k \neq Y_k) = \sum_{k=1}^{+\infty} \mathbf{P}(|X_k| > k) = \sum_{k=1}^{+\infty} \mathbf{P}(|X_1| > k),$$

la deuxième égalité reposant sur l'équidistribution [2] des X_k. Pour comparer cette dernière série à une intégrale, posons $g(t) := \mathbf{P}(|X_1| > t)$, ce qui définit une fonction décroissante $g : \mathbb{R}_+ \to [0,1]$. Cette fonction est clairement λ-intégrable sur tout intervalle borné de $\mathbb{R}_+$ et on a :

$$\forall k \in \mathbb{N}^*, \quad \int_{[k-1,k[} g(t)\,\mathrm{d}\lambda(t) \geq \int_{[k-1,k[} g(k)\,\mathrm{d}\lambda = g(k)\lambda([k-1,k[) = g(k).$$

On en déduit la majoration :

$$\sum_{k=1}^{+\infty} \mathbf{P}(X_k \neq Y_k) = \sum_{k=1}^{+\infty} g(k) \leq \sum_{k=1}^{+\infty} \int_{[k-1,k[} g(t)\,\mathrm{d}\lambda(t) = \int_{[0,+\infty[} \mathbf{P}(|X_1| > t)\,\mathrm{d}\lambda(t).$$

Or d'après la proposition 5.10 appliquée à la variable aléatoire positive $|X_1|$,

$$\int_{[0,+\infty[} \mathbf{P}(|X_1| > t)\,\mathrm{d}\lambda(t) = \mathbf{E}|X_1|.$$

Ainsi la convergence

$$\sum_{k=1}^{+\infty} \mathbf{P}(X_k \neq Y_k) < +\infty \tag{8.20}$$

résulte de l'hypothèse $\mathbf{E}|X_1| < +\infty$.

On sait maintenant que l'évènement

$$A := \Big\{\omega \in \Omega;\ \exists k_0 = k_0(\omega),\ \forall k \geq k_0,\ X_k(\omega) = Y_k(\omega)\Big\}$$

a pour probabilité 1, puisque (8.20) et le premier lemme de Borel-Cantelli nous donnent :

$$\mathbf{P}(A^c) = \mathbf{P}\Big(\{\omega \in \Omega; X_k(\omega) \neq Y_k(\omega) \text{ pour une infinité d'indices } k\}\Big) = 0.$$

Si $\omega \in A$, pour tout $n \geq k_0(\omega)$,

$$\frac{1}{n}\sum_{k=1}^{n} X_k(\omega) - \frac{1}{n}\sum_{k=1}^{n} Y_k(\omega) = \frac{1}{n}\underbrace{\left[\sum_{k=1}^{k_0(\omega)} X_k(\omega) - \sum_{k=1}^{k_0(\omega)} Y_k(\omega)\right]}_{\text{ne dépend pas de } n},$$

d'où la convergence vers 0 de cette différence des moyennes arithmétiques. Ainsi on a l'inclusion d'évènements :

$$A \subset \left\{\frac{1}{n}\sum_{k=1}^{n} X_k - \frac{1}{n}\sum_{k=1}^{n} Y_k \xrightarrow[n\to+\infty]{} 0\right\}$$

et comme $P(A) = 1$, on en déduit la convergence presque sûre :

$$\frac{1}{n}\sum_{k=1}^{n} X_k - \frac{1}{n}\sum_{k=1}^{n} Y_k \xrightarrow[n\to+\infty]{\text{p.s.}} 0.$$

Il est maintenant clair que (8.17) est impliquée par (8.19). □

2. L'équidistribution d'une suite de variables aléatoires X_k est la propriété : « les X_k ont même loi ».

Deuxième étape : réduction de la preuve de (8.19). Centrons les variables Y_k en les remplaçant par $Y'_k := Y_k - \mathbf{E}Y_k = Y_k - m_k$. Il suffit pour établir (8.19) de prouver que

$$\frac{1}{n}\sum_{k=1}^{n} Y'_k \xrightarrow[n\to+\infty]{\text{p.s.}} 0. \tag{8.21}$$

En effet $n^{-1}\sum_{k=1}^{n} Y_k = n^{-1}\sum_{k=1}^{n} Y'_k + n^{-1}\sum_{k=1}^{n} m_k$ et par (8.18) et le théorème de Césaro,

$$\frac{1}{n}\sum_{k=1}^{n} m_k \xrightarrow[n\to+\infty]{} \mathbf{E}X_1 = 0.$$

$\square$

Troisième étape : preuve de (8.21). Nous allons établir (8.21) en appliquant aux Y'_k la loi forte des grands nombres de Kolmogorov. Pour cela, il nous faut vérifier que

$$\sum_{k=1}^{+\infty} \frac{\operatorname{Var} Y'_k}{k^2} < +\infty. \tag{8.22}$$

Comme $\operatorname{Var} Y'_k = \operatorname{Var} Y_k$ et $\operatorname{Var} Y_k = \mathbf{E}Y_k^2 - m_k^2$, on voit que

$$\operatorname{Var} Y'_k \leq \mathbf{E}Y_k^2 = \int_{\{|X_k|\leq k\}} X_k^2 \,\mathrm{d}\mathbf{P}.$$

Comme X_k et X_1 ont même loi, on a par transferts :

$$\int_{\{|X_k|\leq k\}} X_k^2 \,\mathrm{d}\mathbf{P} = \int_{[-k,k]} x^2 \,\mathrm{d}P_{X_k}(x) = \int_{[-k,k]} x^2 \,\mathrm{d}P_{X_1}(x) = \int_{\{|X_1|\leq k\}} X_1^2 \,\mathrm{d}\mathbf{P}.$$

Nous pouvons maintenant écrire les majorations suivantes.

$$\begin{aligned}
\sum_{k=1}^{+\infty} \frac{\operatorname{Var} Y'_k}{k^2} &\leq \sum_{k=1}^{+\infty} \frac{1}{k^2} \int_{\{|X_1|\leq k\}} X_1^2 \,\mathrm{d}\mathbf{P} \\
&= \sum_{k=1}^{+\infty} \frac{1}{k^2} \sum_{i=1}^{k} \int_{\{i-1<|X_1|\leq i\}} X_1^2 \,\mathrm{d}\mathbf{P} \quad (\text{sur } \{i-1<|X_1|\leq i\},\ X_1^2 \leq i|X_1|) \\
&\leq \sum_{k=1}^{+\infty} \frac{1}{k^2} \sum_{i=1}^{k} \int_{\{i-1<|X_1|\leq i\}} i|X_1| \,\mathrm{d}\mathbf{P} \\
&= \sum_{k=1}^{+\infty} \sum_{i=1}^{+\infty} \frac{i}{k^2} \mathbf{1}_{\{1\leq i\leq k\}} \int_{\{i-1<|X_1|\leq i\}} |X_1| \,\mathrm{d}\mathbf{P} \\
&= \sum_{i=1}^{+\infty} i \left(\sum_{k=i}^{+\infty} \frac{1}{k^2} \right) \int_{\{i-1<|X_1|\leq i\}} |X_1| \,\mathrm{d}\mathbf{P}.
\end{aligned}$$

Pour achever la majoration, une classique comparaison série-intégrale nous donne

$$\forall i \geq 1, \quad \sum_{k=i}^{+\infty} \frac{1}{k^2} = \frac{1}{i^2} + \sum_{k=i+1}^{+\infty} \frac{1}{k^2} \leq \frac{1}{i^2} + \int_i^{+\infty} \frac{\mathrm{d}x}{x^2} = \frac{1}{i^2} + \frac{1}{i} \leq \frac{2}{i}.$$

Finalement,

$$\sum_{k=1}^{+\infty} \frac{\operatorname{Var} Y_k'}{k^2} \leq 2 \sum_{i=1}^{+\infty} \int_{\{i-1<|X_1|\leq i\}} |X_1| \,\mathrm{d}\mathbf{P} = 2\mathbf{E}|X_1| < +\infty,$$

ce qui établit (8.22) et donc aussi (8.21) par la loi forte des grands nombres de Kolmogorov. □

La loi forte des grands nombres de Khintchine est maintenant complètement démontrée. □

L'hypothèse d'intégrabilité de X_1 dans la loi forte des grands nombres de Khintchine est optimale. Plus précisément, dans le cas d'une suite (X_k) i.i.d., on a équivalence entre la convergence presque sûre de S_n/n et l'intégrabilité de X_1. Cette équivalence résulte du théorème 8.13 et du théorème suivant.

Théorème 8.14. *Soit $(X_k)_{k\geq 1}$ 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 $\mathbf{E}X_1$.*

Preuve. Nous exposons la preuve sous l'hypothèse supplémentaire que la limite p.s. de S_n/n est une *constante* c. La loi du zéro-un de Kolmogorov (qui figure au programme de Maîtrise) montre que si S_n/n converge p.s., sa limite est nécessairement une constante. Alors

$$\frac{X_n}{n} = \frac{S_n - S_{n-1}}{n} = \frac{S_n}{n} - \frac{n-1}{n} \times \frac{S_{n-1}}{n-1} \xrightarrow[n\to+\infty]{\text{p.s.}} c - c = 0.$$

En fixant $\varepsilon > 0$, on en déduit

$$\mathbf{P}\Big\{\omega \in \Omega; \exists n_0 = n_0(\omega), \forall n \geq n_0, \frac{|X_n(\omega)|}{n} < \varepsilon\Big\} = 1,$$

soit en passant au complémentaire

$$\mathbf{P}\Big\{\frac{|X_n|}{n} \geq \varepsilon \text{ une infinité de fois}\Big\} = 0.$$

Par le second lemme de Borel-Cantelli (lemme 5.37), on a alors

$$\sum_{n\geq 1} \mathbf{P}\Big\{\frac{|X_n|}{n} \geq \varepsilon\Big\} < +\infty.$$

Les X_i ayant même loi, ceci s'écrit

$$\sum_{n=1}^{+\infty} \mathbf{P}(|X_1| \geq n\varepsilon) < +\infty. \tag{8.23}$$

Pour finir la preuve, on observe que

$$\begin{aligned}
\mathbf{E}|X_1| &\leq \mathbf{E}\left(\sum_{n=0}^{+\infty}(n+1)\varepsilon\mathbf{1}_{\{n\varepsilon\leq|X_1|<(n+1)\varepsilon\}}\right)\\
&= \sum_{n=0}^{+\infty}(n+1)\varepsilon\mathbf{P}(n\varepsilon\leq|X_1|<(n+1)\varepsilon)\\
&= \varepsilon\sum_{n=0}^{+\infty}\mathbf{P}(n\varepsilon\leq|X_1|)<+\infty,
\end{aligned}$$

la dernière ligne s'obtenant par sommation triangulaire à partir des décompositions en unions disjointes

$$\{|X_1|\geq n\varepsilon\}=\bigcup_{k\geq n}\{k\varepsilon\leq|X_1|<(k+1)\varepsilon\}.$$

□

8.4 Quelques applications des lois des grands nombres

Nous présentons dans cette section quelques applications de la loi forte des grands nombres qui concernent essentiellement la statistique. Nous commençons par le cas facile des variables aléatoires indépendantes, de même loi et bornées. Elles sont donc automatiquement intégrables et vérifient la loi forte des grands nombres de Khintchine.

8.4.1 Applications de la LFGN pour des v.a. bornées

L'application la plus simple et aussi une des plus importantes du théorème 8.13 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 $(X_k)_{k\geq 1}$ est une suite de variables aléatoires de Bernoulli indépendantes et de même paramètre p, le théorème 8.13 nous donne la convergence presque sûre de $n^{-1}S_n$ vers $\mathbf{E}X_1=p$. La figure 8.2 montre une simulation réalisée en Scilab avec $n=1000$ et $p=0.6$. Soit maintenant $A\in\mathfrak{F}$ un évènement et $(A_k)_{k\geq 1}$ une suite d'évènements indépendants de même probabilité que A. En prenant $X_k=\mathbf{1}_{A_k}$ et en notant que $\mathbf{E}\mathbf{1}_{A_k}=\mathbf{P}(A_k)=\mathbf{P}(A)$, on obtient

$$\frac{1}{n}\sum_{k=1}^{n}\mathbf{1}_{A_k}\xrightarrow[n\to+\infty]{\text{p.s.}}\mathbf{P}(A).$$

Par exemple si A est l'évènement « obtention d'un double lors du lancer d'une paire de dés équilibrés », ceci nous dit que la *fréquence* d'obtention d'un double en n lancers converge presque sûrement vers $1/6$ lorsque n tend vers l'infini. À titre d'exemple historique, on peut mentionner le problème de l'aiguille de Buffon (cf. D.M. n° 4, Annales IFP 2002–03).

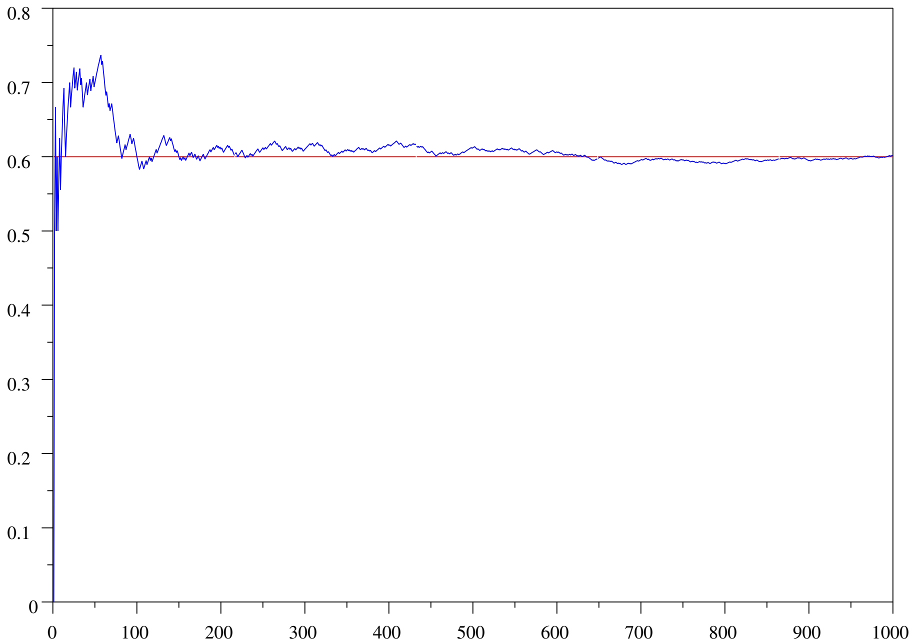


FIGURE 8.2 – Ligne polygonale de sommets $(k, S_k/k)$, $1 \leq k \leq 1000$, $X_k \sim \text{Bern}(0.6)$.

Le théorème 8.13 a une traduction statistique fondamentale : il permet de justifier la convergence de la fonction de répartition empirique. Considérons une suite (Y_k) de variables aléatoires indépendantes et de même loi de fonction de répartition F. On définit la *fonction de répartition empirique* F_n construite sur l'*échantillon* $Y_1, \dots, Y_n$ par

$$F_n(x) := \frac{1}{n} \sum_{k=1}^{n} \mathbf{1}_{\{Y_k \leq x\}}, \quad x \in \mathbb{R}. \tag{8.24}$$

Le théorème 8.13 appliqué aux variables aléatoires bornées $X_k = \mathbf{1}_{\{Y_k \leq x\}}$ nous donne immédiatement pour tout $x \in \mathbb{R}$ la convergence presque sûre de $F_n(x)$ vers $F(x)$, en remarquant que $\mathbf{E}X_1 = P(Y_1 \leq x) = F(x)$. Ainsi une *loi inconnue* peut être reconstituée approximativement à partir de l'observation d'un échantillon de grande taille. En fait, on peut obtenir mieux que la convergence simple presque sûre de F_n vers F.

Théorème 8.15 (Glivenko-Cantelli)**.** *Soit (Y_k) une suite de variables aléatoires indépendantes, de même loi et (F_n) la suite de fonctions de répartition empiriques associées. Alors*

$$\|F_n - F\|_\infty := \sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \xrightarrow[n \to +\infty]{\text{p.s.}} 0. \tag{8.25}$$

Ce théorème sera démontré en cours de Maîtrise.

La LFGN pour des variables aléatoires bornées donne aussi immédiatement la convergence presque sûre des fonctions caractéristiques empiriques.

Proposition 8.16. *Soit (Y_k) une suite de vecteurs aléatoires dans $\mathbb{R}^d$, indépendants et de même loi de fonction caractéristique φ définie par $\varphi(u) := \mathbf{E}\exp(i\langle u, Y_1\rangle)$, $u \in \mathbb{R}^d$. Alors la fonction caractéristique empirique*

$$\varphi_n(u) := \frac{1}{n} \sum_{k=1}^{n} \exp(i\langle u, Y_k\rangle)$$

converge ponctuellement presque sûrement sur $\mathbb{R}^d$ vers φ.

Vérification. Il suffit d'appliquer le théorème 8.13 aux variables aléatoires réelles $X_k' = \cos(\langle u, Y_k\rangle)$ et $X_k'' = \sin(\langle u, Y_k\rangle)$. □

8.4.2 La méthode de Monte Carlo

La loi des grands nombres fournit une méthode de calcul approché d'intégrales, intéressante lorsque la fonction à intégrer est très irrégulière ou lorsque la dimension de l'espace est élevée. Supposons que l'on veuille effectuer un calcul approché de

$$I := \int_{[0,1]^d} f(x) \,\mathrm{d}x,$$

où f est Lebesgue intégrable sur $[0,1]^d$. Soit $(U_i)_{i \geq 1}$ une suite de variables aléatoires indépendantes de même loi uniforme sur $[0,1]$. On déduit facilement de la LFGN de Kolmogorov-Khintchine que :

$$\frac{1}{n} \sum_{k=1}^{n} f\Big(U_{(k-1)d+1}, U_{(k-1)d+2}, \dots, U_{kd}\Big) \xrightarrow[n \to +\infty]{\text{p.s.}} \mathbf{E}f(U_1, \dots, U_d) = \int_{[0,1]^d} f(x) \,\mathrm{d}x.$$

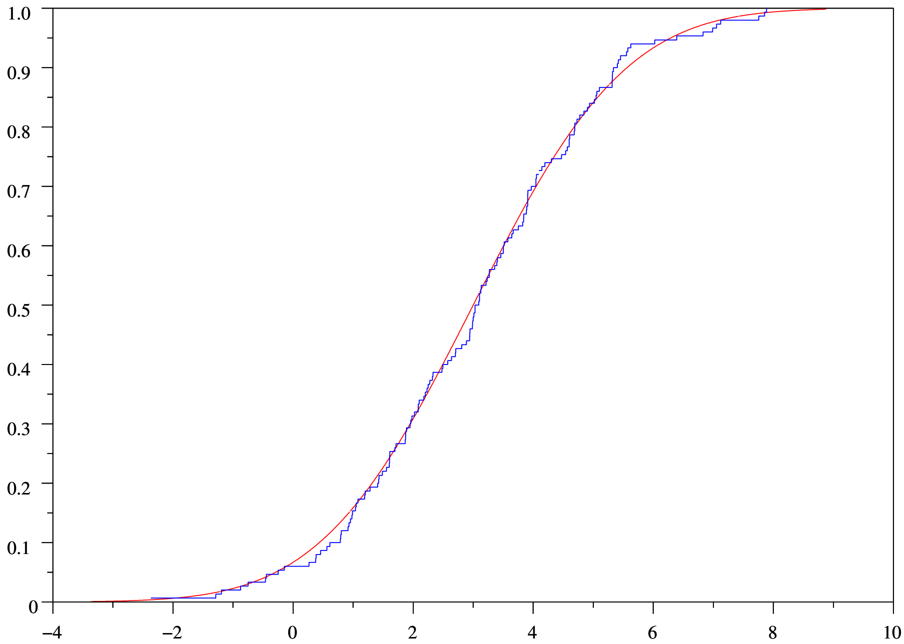


FIGURE 8.3 – F_{150} et F pour un échantillon de la loi $\mathfrak{N}(3, 2)$.

Le théorème limite central permet ensuite d'obtenir un *intervalle de confiance* pour I si l'on a des hypothèses supplémentaires permettant de contrôler la variance de $f(U_1, \dots, U_d)$, par exemple f bornée...

8.4.3 Estimation de paramètres

La LFGN permet de définir des estimateurs convergents de paramètres d'une loi inconnue μ (ou partiellement inconnue, par exemple on sait qu'il s'agit d'une loi de Poisson de paramètre α dont on ignore la valeur). Pour cela on utilise une suite d'observations indépendantes $X_1(\omega), \dots, X_n(\omega)$ où les X_i sont i.i.d. de même loi μ. On souhaite estimer un paramètre θ de la forme $\theta = \int_{\mathbb{R}} H \, \mathrm{d}\mu$. L'idée est de remplacer la mesure déterministe mais inconnue μ par la mesure *aléatoire* μ_n calculable à partir des observations :

$$\mu_n = \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i}.$$

Cette mesure est appelée *mesure empirique*. La fonction de répartition empirique déjà vue en (8.24) est simplement sa fonction de répartition : $F_n(x) = \mu_n(]-\infty, x])$. On propose d'estimer θ par

$$\widehat{\theta}_n := \int_{\mathbb{R}} H \, \mathrm{d}\mu_n = \frac{1}{n} \sum_{i=1}^{n} H(X_i).$$

La définition de θ suppose implicitement que H est μ intégrable. Cette intégrabilité s'écrit encore $\mathbf{E}|H(X_1)| < +\infty$. Ainsi par la loi forte des grands nombres,

$$\widehat{\theta}_n \xrightarrow[n\to+\infty]{\text{p.s.}} \mathbf{E}H(X_1) = \int_{\mathbb{R}} H \, \mathrm{d}\mu = \theta.$$

On dit que $\widehat{\theta}_n$ est une estimateur *fortement consistant* de θ. Il est aussi *sans biais* puisque $\mathbf{E}\widehat{\theta}_n = \theta$.

Cette méthode permet notamment d'estimer les moments de μ : en prenant $H(x) = x^r$, $\theta = \mathbf{E}X_1^r = \int_{\mathbb{R}} x^r \mu(\,\mathrm{d}x)$. Le cas $r = 1$ revêt une importance particulière. L'estimateur $\widehat{\theta}_n$ est alors simplement la *moyenne empirique*

$$\bar{X}_n := \frac{1}{n} \sum_{i=1}^{n} X_i.$$

On peut ainsi estimer notamment

- le paramètre p d'une loi de Bernoulli car $\mathbf{E}X_1 = p$;
- le paramètre α d'une loi de Poisson car $\mathbf{E}X_1 = \alpha$;
- le paramètre m d'une loi $\mathfrak{N}(m, \sigma^2)$ car $\mathbf{E}X_1 = m$;
- le paramètre θ d'une loi uniforme[3] sur $[0, \theta]$ car $\theta = 2\mathbf{E}X_1$.

Dans le même ordre d'idées, on peut estimer le paramètre a d'une loi exponentielle de densité $f(t) = a \exp(-at) \mathbf{1}_{\mathbb{R}^+}(t)$ par $\widetilde{\theta}_n = 1/\bar{X}_n$. En effet, $\mathbf{E}X_1 = 1/a$. On garde un estimateur fortement consistant, mais il n'est plus sans biais car $\mathbf{E}(1/\bar{X}_n) \neq 1/\mathbf{E}X_1$.

3. En fait dans ce cas, un meilleur estimateur est $\tilde{\theta}_n = \max_{1 \leq i \leq n} X_i$, affaire à suivre...

On peut de même estimer la variance σ^2 d'une loi μ d'espérance *connue* m. Il suffit de prendre $H(x) = (x-m)^2$ et on obtient l'estimateur fortement consistant et sans biais

$$\widehat{\theta}_n = \frac{1}{n}\sum_{i=1}^{n}(X_i - m)^2.$$

Quand m est inconnu, on l'estime par $\bar{X}_n$ et la variance σ^2 est estimée par la *variance empirique*

$$V_n = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2 = \frac{1}{n}\sum_{i=1}^{n}X_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n}X_i\right)^2,$$

la dernière égalité résultant simplement de la formule de Koenig pour la variance de la loi de probabilité $\mu_n(\omega)$ qui est exactement $V_n(\omega)$. On a toujours un estimateur fortement consistant par la LFGN, par contre il n'est plus qu'*asymptotiquement* sans biais puisque :

$$\mathbf{E}V_n = \mathbf{E}X_1^2 - \mathbf{E}\left(\frac{1}{n}\sum_{i=1}^{n}X_i\right)^2 = \cdots = \frac{n-1}{n}\sigma^2.$$

Ceci explique pourquoi pour les petites valeurs de n on préfère l'estimateur sans biais $\frac{n}{n-1}V_n$ noté souvent σ_{n-1}^2 par un de ces abus d'écriture qui font le charme si particulier de la littérature statistique...

Chapitre 9

Transformation de Fourier et convergence en loi

La transformation de Fourier déjà entrevue à propos des fonctions caractéristiques est un outil important, aussi bien en analyse qu'en théorie des probabilités. Après avoir introduit les propriétés fondamentales de la transformée de Fourier des mesures finies et des fonctions de $\mathrm{L}^1(\lambda_d)$, nous verrons comment la transformée de Fourier permet l'étude de la *convergence des lois de probabilité*. Cette étude débouche naturellement sur l'exemple le plus célèbre de convergence en loi. Il s'agit du *théorème limite central* qui affirme que la suite convenablement normalisée des sommes partielles de variables aléatoires i.i.d. et de carré intégrable suit asymptotiquement une loi gaussienne.

Dans tout le chapitre, $\mathbb{R}^d$ est muni du produit scalaire et de la norme euclidienne standards notés :

$$\langle t, x\rangle = \sum_{j=1}^{d} t_j x_j, \quad \|x\| = \left(x_1^2 + \cdots + x_d^2\right)^{1/2},$$

où $t = (t_1, \ldots, t_d)$ et $x = (x_1, \ldots, x_d)$ désignent deux vecteurs quelconques de $\mathbb{R}^d$.

9.1 Transformée de Fourier d'une mesure

9.1.1 Premières propriétés

Définition 9.1. *Soit μ une mesure finie sur $\left(\mathbb{R}^d, \operatorname{Bor}(\mathbb{R}^d)\right)$. On appelle* transformée de Fourier *de μ la* fonction $\widehat{\mu} : \mathbb{R}^d \to \mathbb{C}$, *définie par :*

$$\forall t \in \mathbb{R}^d, \quad \widehat{\mu}(t) := \int_{\mathbb{R}^d} \exp\bigl(i\langle t, x\rangle\bigr)\,\mathrm{d}\mu(x). \tag{9.1}$$

Lorsque μ est la loi d'un vecteur aléatoire, $\widehat{\mu}$ est la fonction caractéristique *de X.*

$$\textit{Si } \mu = P_X, \quad \widehat{\mu}(t) = \varphi_X(t) = \mathbf{E}\exp\bigl(i\langle t, X\rangle\bigr). \tag{9.2}$$

On notera que l'hypothèse μ finie (donc $\mu(\mathbb{R}^d) < +\infty$) rend automatiquement μ intégrable sur $\mathbb{R}^d$ la fonction $x \mapsto \exp\big(i\langle t, x\rangle\big)$, de module constant 1. Il est clair que cette fonction n'est λ_d intégrable sur $\mathbb{R}^d$ pour aucun t et donc (9.1) ne permet pas de définir $\widehat{\lambda_d}$.

Les deux tableaux suivants donnent les fonctions caractéristiques de quelques lois usuelles, discrètes ou à densité. Les calculs de vérification sont laissés en exercice.

Exemple 9.1 (Fonctions caractéristiques de lois discrètes usuelles, $d = 1$)**.**

Loi de X	Paramètres	$\mathbf{P}(X = k)$	$\varphi_X(t)$
δ_a	$a \in \mathbb{R}$	$\mathbf{P}(X = a) = 1$	e^{ita}
$\mathrm{Bern}(p)$	$p \in [0, 1]$	$\mathbf{P}(X = 0) = 1 - p,\ \mathbf{P}(X = 1) = p$	$p\mathrm{e}^{it} + 1 - p$
$\mathrm{Bin}(n, p)$	$p \in [0, 1],\ n \in \mathbb{N}^*$	$C_n^k p^k (1 - p)^{n-k}, \quad 0 \leq k \leq n$	$\big(p\mathrm{e}^{it} + (1 - p)\big)^n$
$\mathrm{Geom}(p)$	$p \in]0, 1[$	$(1 - p)^{k-1} p, \quad k \in \mathbb{N}^*$	$\dfrac{p\mathrm{e}^{it}}{1 - (1 - p)\mathrm{e}^{it}}$
$\mathrm{Pois}(\alpha)$	$\alpha \in]0, +\infty[$	$\dfrac{\mathrm{e}^{-\alpha}\alpha^k}{k!}, \quad k \in \mathbb{N}$	$\exp\big(\alpha(\mathrm{e}^{it} - 1)\big)$

Exemple 9.2 (Fonctions caractéristiques de lois usuelles à densité, $d = 1$)**.**

Loi	Paramètres	Densité $f_X(x)$	$\varphi_X(t)$
$\mathrm{Unif}[a, b]$	$a, b \in \mathbb{R}$	$\dfrac{1}{b - a}\mathbf{1}_{[a,b]}(x)$	$\dfrac{\mathrm{e}^{ibt} - \mathrm{e}^{iat}}{(b - a)it}$
$\mathrm{Exp}(a)$	$a \in]0, +\infty[$	$a\mathrm{e}^{-ax}\mathbf{1}_{[0,+\infty[}(x)$	$\dfrac{1}{1 - \frac{it}{a}}$
$\mathrm{Cau}(a, b)$	$a \in \mathbb{R},\ b \in]0, +\infty[$	$\dfrac{1}{\pi b}\dfrac{1}{1 + \left(\frac{x-a}{b}\right)^2}$	$\mathrm{e}^{iat - b\lvert t\rvert}$
Triangulaire		$\big(1 - \lvert x\rvert\big)\mathbf{1}_{[-1,1]}(x)$	$\dfrac{2(1 - \cos t)}{t^2}$
$\mathfrak{N}(m, \sigma)$	$m \in \mathbb{R},\ \sigma \in]0, +\infty[$	$\dfrac{1}{\sigma\sqrt{2\pi}}\exp\left(\dfrac{-(x - m)^2}{2\sigma^2}\right)$	$\exp\left(imt - \dfrac{\sigma^2 t^2}{2}\right)$

Le calcul de la fonction caractéristique de la loi gaussienne $\mathfrak{N}(m, \sigma)$ n'est pas immédiat. Il est détaillé ci-dessous (proposition 9.6).

Proposition 9.2. *La transformée de Fourier $\widehat{\mu}$ d'une mesure finie μ est une fonction continue et bornée sur $\mathbb{R}^d$. On a*

$$\forall t \in \mathbb{R}^d, \quad |\widehat{\mu}(t)| \leq \mu(\mathbb{R}^d) = \widehat{\mu}(0). \tag{9.3}$$

Vérification. Pour la bornitude, il suffit d'écrire

$$|\widehat{\mu}(t)| = \left|\int_{\mathbb{R}^d} \exp\bigl(i\langle t, x\rangle\bigr)\, \mathrm{d}\mu(x)\right| \le \int_{\mathbb{R}^d} \left|\exp\bigl(i\langle t, x\rangle\bigr)\right| \mathrm{d}\mu(x) = \int_{\mathbb{R}^d} \mathrm{d}\mu(x) = \mu(\mathbb{R}^d) < +\infty.$$

La continuité est une application immédiate du théorème de continuité sous le signe somme avec pour fonction dominante la constante 1. □

Proposition 9.3. *Soit X une variable aléatoire réelle ayant un moment d'ordre $r \in \mathbb{N}^*$ (i.e. $\mathbf{E}|X|^r < +\infty$) et ϕ sa fonction caractéristique. Alors ϕ est r fois dérivable sur $\mathbb{R}$ et*

$$\forall t \in \mathbb{R}, \quad \phi^{(k)}(t) = i^k \mathbf{E}\left(X^k \mathrm{e}^{itX}\right), \quad 1 \le k \le r.$$

En particulier

$$\phi^{(k)}(0) = i^k \mathbf{E} X^k.$$

Vérification. C'est une application du théorème de dérivation sous le signe somme en notant que si $\mathbf{E}|X|^r$ est fini, $\mathbf{E}|X|^k$ l'est aussi pour tout $k \le r$. Cette remarque jointe à la majoration évidente

$$\left|\frac{\mathrm{d}^k}{\mathrm{d}t^k} \exp\bigl(itX(\omega)\bigr)\right| = \left|i^k X(\omega)^k \mathrm{e}^{itX(\omega)}\right| \le |X(\omega)|^k,$$

nous permet lors de la k-ième dérivation, d'utiliser $|X|^k$ comme fonction dominante indépendante de t et $\mathbf{P}$-intégrable sur Ω. □

Proposition 9.4. *Si la mesure μ est une mesure produit $\mu = \mu_1 \otimes \cdots \otimes \mu_d$ de mesures finies sur $\mathbb{R}$,*

$$\forall t = (t_1, \dots, t_d) \in \mathbb{R}^d, \quad \widehat{\mu}(t) = \prod_{j=1}^{d} \widehat{\mu}_j(t_j), \tag{9.4}$$

autrement dit,

$$\widehat{\mu} = \widehat{\mu}_1 \otimes \cdots \otimes \widehat{\mu}_d.$$

Preuve. Comme $\mu(\mathbb{R}^d) = \mu_1(\mathbb{R}) \times \cdots \times \mu_d(\mathbb{R})$, μ est aussi une mesure finie et $\widehat{\mu}$ est donc bien définie. Comme $\int_{\mathbb{R}^d} |\exp(i\langle t, x\rangle)|\, \mathrm{d}\mu(x) \le \mu(\mathbb{R}^d)$, la fonction $x \mapsto \exp(i\langle t, x\rangle)$ est $\mu_1 \otimes \cdots \otimes \mu_d$ intégrable et on peut justifier par le corollaire 5.20 le calcul suivant :

$$\begin{aligned}
\widehat{\mu}(t) &= \int_{\mathbb{R}^d} \exp\bigl(i(t_1 x_1 + \cdots + t_d x_d)\bigr)\, \mathrm{d}(\mu_1 \otimes \cdots \otimes \mu_d)(x_1, \dots, x_d) \\
&= \int_{\mathbb{R}^d} \prod_{j=1}^{d} \exp(it_j x_j)\, \mathrm{d}(\mu_1 \otimes \cdots \otimes \mu_d)(x_1, \dots, x_d) \\
&= \prod_{j=1}^{d} \int_{\mathbb{R}} \exp(it_j x_j)\, \mathrm{d}\mu_j(x_j) \\
&= \prod_{j=1}^{d} \widehat{\mu}_j(t_j)
\end{aligned}$$

qui établit (9.4). □

Dans le langage des fonctions caractéristiques, le résultat que nous venons de vérifier se traduit comme suit.

Corollaire 9.5. *Soit $X = (X_1, \dots, X_d)$ un vecteur aléatoire à composantes* indépendantes*, notons φ sa fonction caractéristique et pour $j = 1, \dots, d$, φ_j celle de la variable aléatoire réelle X_j. Alors*

$$\forall t = (t_1, \dots, t_d) \in \mathbb{R}^d, \quad \varphi(t) = \varphi_1(t_1) \dots \varphi_d(t_d),$$

autrement dit

$$\varphi = \varphi_1 \otimes \dots \otimes \varphi_d.$$

Vérification. Si X est à composantes indépendantes, $P_X = P_{X_1} \otimes \dots \otimes P_{X_d}$. □

9.1.2 Fonction caractéristique de la loi normale standard

Pour des raisons qui apparaîtront progressivement, il importe de connaître la fonction caractéristique de la loi gaussienne standard $\mathfrak{N}(0, 1)$.

Proposition 9.6. *La loi normale $\mathfrak{N}(0, 1)$ de densité $f : x \mapsto (2\pi)^{-1/2} \exp(-x^2/2)$ a pour fonction caractéristique $\varphi = \sqrt{2\pi} f$:*

$$\forall t \in \mathbb{R}, \quad \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \mathrm{e}^{itx} \mathrm{e}^{-x^2/2} \, \mathrm{d}\lambda(x) = \mathrm{e}^{-t^2/2}. \tag{9.5}$$

Plus généralement, si X suit la loi gaussienne $\mathfrak{N}(m, \sigma)$, de densité

$$f_{m,\sigma} : x \longmapsto \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-(x-m)^2}{2\sigma^2}\right),$$

sa fonction caractéristique $\varphi_{m,\sigma}$ est donnée par

$$\forall t \in \mathbb{R}, \quad \varphi_{m,\sigma}(t) = \exp\left(imt - \frac{\sigma^2 t^2}{2}\right). \tag{9.6}$$

Des différentes méthodes possibles pour vérifier (9.5), la plus élégante est sans doute celle qui utilise l'analyse complexe [1]. Celle un peu plus laborieuse que nous exposons ci-dessous, a cependant le mérite de n'utiliser que des techniques du cours d'intégration. L'idée est de développer en série entière e^{itx} et d'intégrer terme à terme. Nous aurons besoin pour cela de connaître les moments de la loi $\mathfrak{N}(0, 1)$.

1. L'intégrale de la fonction *entière* $\exp(-z^2/2)$ le long du contour rectangulaire de sommets R, $R+it$, $-R+it$, $-R$, R est nulle. Elle le reste quand R tend vers l'infini et les intégrales sur les segments verticaux tendent vers 0. En paramétrant les segments horizontaux du contour, on en déduit (9.5). La rédaction complète à partir de cette esquisse est un exercice recommandé aux étudiants suivant le module Variable Complexe.

Lemme 9.7 (Moments de $\mathfrak{N}(0,1)$)**.** *Si la variable aléatoire réelle X suit la loi $\mathfrak{N}(0,1)$, elle a des moments de tout ordre. Ses moments d'ordre impair sont nuls et*

$$\forall k \in \mathbb{N}^*, \quad \mathbf{E}X^{2k} = \prod_{j=1}^{k}(2j-1) = \frac{(2k)!}{2^k k!}. \tag{9.7}$$

Preuve. Pour vérifier l'existence des moments, on écrit grâce au transfert de Ω vers $\mathbb{R}$ par X,

$$\mathbf{E}|X|^n = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} |x|^n \mathrm{e}^{-x^2/2} \,\mathrm{d}\lambda(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} |x|^n \mathrm{e}^{-x^2/2} \,\mathrm{d}x < +\infty, \tag{9.8}$$

la conversion de l'intégrale de Lebesgue en intégrale de Riemann étant justifiée par le fait que la fonction à intégrer est continue sur $\mathbb{R}$ et d'intégrale généralisée absolument convergente. Pour cette convergence, il suffit d'écrire $|x|^n\mathrm{e}^{-x^2/2} = |x|^n\mathrm{e}^{-x^2/4}\mathrm{e}^{-x^2/4}$, de noter que $|x|^n\mathrm{e}^{-x^2/4}$ tend vers 0 quand x tend vers $\pm\infty$ et que $\int_{-\infty}^{+\infty} \mathrm{e}^{-x^2/4}\,\mathrm{d}x$ converge en raison de l'inégalité $\mathrm{e}^{-x^2/4} \leq \mathrm{e}^{-|x|/4}$ pour $|x| \geq 1$, la convergence de $\int_{-\infty}^{+\infty} \mathrm{e}^{-|x|/4}\,\mathrm{d}x$ étant évidente par recours aux primitives. L'existence des moments de tout ordre est établie par (9.8) et le même argument pour la conversion de l'intégrale de Lebesgue en intégrale de Riemann nous permet maintenant d'écrire

$$\forall n \in \mathbb{N}, \quad \mathbf{E}X^n = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x^n \mathrm{e}^{-x^2/2} \,\mathrm{d}x.$$

Quand n est impair, la fonction à intégrer est impaire et cette intégrale de Riemann généralisée absolument convergente[2] est nulle. Reste à calculer les moments d'ordre pair : $c_k := \mathbf{E}X^{2k}$, $k \in \mathbb{N}$. On sait déjà que $c_0 = 1$. L'idée est de chercher une relation de récurrence en intégrant par parties.

$$\frac{1}{\sqrt{2\pi}} \int_a^b x^{2k-2}\mathrm{e}^{-x^2/2}\,\mathrm{d}x = \frac{1}{\sqrt{2\pi}} \left[\frac{x^{2k-1}}{2k-1}\mathrm{e}^{-x^2/2}\right]_a^b - \frac{1}{\sqrt{2\pi}} \int_a^b \frac{x^{2k-1}}{2k-1}(-x)\mathrm{e}^{-x^2/2}\,\mathrm{d}x.$$

En faisant tendre a vers $-\infty$ et b vers $+\infty$, on en déduit

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x^{2k-2}\mathrm{e}^{-x^2/2}\,\mathrm{d}x = \frac{1}{\sqrt{2\pi}(2k-1)} \int_{-\infty}^{+\infty} x^{2k}\mathrm{e}^{-x^2/2}\,\mathrm{d}x,$$

d'où la relation de récurrence

$$\forall k \in \mathbb{N}^*, \quad c_k = (2k-1)c_{k-1}.$$

On en déduit

$$c_k c_{k-1} \dots c_2 c_1 = (2k-1)(2k-3)\dots 3 c_{k-1} c_{k-2} \dots c_1 c_0,$$

2. La convergence est indispensable ici car sinon on pourrait écrire « $\int_{-\infty}^{+\infty} x\,\mathrm{d}x = 0$ », ce qui est une ânerie...

d'où après simplification (aucun des c_j n'est nul), la première égalité de (9.7). Pour la deuxième égalité, il suffit de remarquer que le produit des facteurs *pairs* dans le développement de $(2k)!$ s'écrit

$$2k \times 2(k-1) \times \cdots \times (2 \times 2) \times (2 \times 1) = 2^k k!,$$

ce qui permet d'exprimer le produit des facteurs *impairs* par :

$$\frac{(2k)!}{2^k k!} = (2k-1)(2k-3) \dots 1 = \prod_{j=1}^{k} (2j-1).$$

□

Preuve de la proposition 9.6. Soit X une variable aléatoire de loi $\mathfrak{N}(0,1)$. Sa fonction caractéristique φ peut d'écrire grâce au transfert $(\Omega, \mathbf{P}) \xrightarrow{X} (\mathbb{R}, P_X)$ sous la forme :

$$\varphi(t) = \mathbf{E}e^{itX} = \int_{\mathbb{R}} e^{itx} \, dP_X(x) = \int_{\mathbb{R}} e^{itx} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \, d\lambda(x).$$

Dans tout ce qui suit, on travaille avec t quelconque, mais fixé. Posons pour tout $x \in \mathbb{R}$,

$$S_n(x) := \sum_{l=0}^{n} \frac{(itx)^l}{l!} e^{-x^2/2}.$$

Comme la fonction exponentielle complexe est développable en série entière avec rayon de convergence infini, on a clairement :

$$\forall x \in \mathbb{R}, \quad S_n(x) \xrightarrow[n \to +\infty]{} e^{itx} e^{-x^2/2}. \tag{9.9}$$

On voit que cette convergence est *dominée* en écrivant :

$$\forall x \in \mathbb{R}, \quad \left| S_n(x) \right| \le \sum_{l=0}^{n} \frac{|tx|^l}{l!} e^{-x^2/2} \le e^{|tx|} e^{-x^2/2} = \exp\left(-\frac{x^2}{2} + |tx| \right) =: g(x).$$

Cette fonction g ne dépend pas de n. Comme elle est continue sur $\mathbb{R}$, on obtient sa λ-intégrabilité en vérifiant que l'intégrale de Riemann généralisée $\int_{-\infty}^{+\infty} g(x) \, dx$ est absolument convergente. Ceci résulte de la majoration $0 \le g(x) \le \exp(-|tx|)$ valable au voisinage de $-\infty$ et de $+\infty$ (plus précisément dès que $|x| \ge 4|t|$). Compte-tenu de (9.9), on peut donc appliquer le théorème de convergence dominée pour obtenir :

$$\lim_{n \to +\infty} \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} S_n \, d\lambda = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{itx} e^{-x^2/2} \, d\lambda(x). \tag{9.10}$$

Par ailleurs, par linéarité de l'intégrale (et grâce à l'existence des moments de tout ordre pour la loi $\mathfrak{N}(0,1)$), on a

$$\forall n \in \mathbb{N}, \quad \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} S_n \, d\lambda = \sum_{l=0}^{n} \frac{1}{\sqrt{2\pi}} \frac{(it)^l}{l!} \int_{\mathbb{R}} x^l e^{-x^2/2} \, d\lambda(x).$$

En reportant ceci dans (9.10), on aboutit à

$$\varphi(t) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \mathrm{e}^{itx} \mathrm{e}^{-x^2/2} \,\mathrm{d}\lambda(x) = \sum_{l=0}^{+\infty} \frac{(it)^l}{l!} \int_{\mathbb{R}} x^l \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-x^2/2} \,\mathrm{d}\lambda(x).$$

En utilisant maintenant le calcul des moments vu au lemme 9.7, il vient :

$$\varphi(t) = \sum_{k=0}^{+\infty} \frac{(it)^{2k}}{(2k)!} \mathbf{E} X^{2k} = \sum_{k=0}^{+\infty} (-1)^k \frac{t^{2k}}{(2k)!} \frac{(2k)!}{2^k k!} = \sum_{k=0}^{+\infty} \frac{(-t^2/2)^k}{k!} = \mathrm{e}^{-t^2/2}.$$

Comme t était quelconque, cette formule est vraie pour tout $t \in \mathbb{R}$ et (9.5) est démontrée.

On en déduit le calcul de $\varphi_{m,\sigma}$ en remarquant que si X suit la loi $\mathfrak{N}(m, \sigma)$, X a *même loi* que $\sigma Y + m$ où Y suit la loi $\mathfrak{N}(0, 1)$. Par conséquent, X a même fonction caractéristique que $\sigma Y + m$ et

$$\varphi_{m,\sigma}(t) = \mathbf{E} \exp\big(it(\sigma Y + m)\big) = \mathbf{E}\big(\mathrm{e}^{imt} \mathrm{e}^{it\sigma Y}\big) = \mathrm{e}^{imt} \mathbf{E}\big(\mathrm{e}^{it\sigma Y}\big) = \mathrm{e}^{imt} \varphi(\sigma t),$$

d'où l'égalité (9.6). □

9.1.3 Transformée d'un produit de convolution

Une propriété essentielle de la transformation de Fourier est qu'elle permet de ramener le produit de convolution à un produit ordinaire de deux fonctions.

Théorème 9.8. *Si μ et ν sont deux mesures finies sur* $\big(\mathbb{R}^d, \mathrm{Bor}(\mathbb{R}^d)\big)$,

$$\forall t \in \mathbb{R}^d, \quad \widehat{\mu * \nu}(t) = \widehat{\mu}(t) \widehat{\nu}(t). \tag{9.11}$$

De même si X et Y sont deux vecteurs aléatoires dans $\mathbb{R}^d$, indépendants, *la fonction caractéristique φ_{X+Y} de leur somme est donnée par*

$$\forall t \in \mathbb{R}^d, \quad \varphi_{X+Y}(t) = \varphi_X(t) \varphi_Y(t). \tag{9.12}$$

Preuve. Rappelons que $\mu * \nu$ est la mesure sur $\big(\mathbb{R}^d, \mathrm{Bor}(\mathbb{R}^d)\big)$, définie comme la mesure image de $\mu \otimes \nu$ (mesure sur $\big(\mathbb{R}^d \times \mathbb{R}^d, \mathrm{Bor}(\mathbb{R}^d) \otimes \mathrm{Bor}(\mathbb{R}^d)\big)$) par

$$s : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d, \quad (y, z) \mapsto y + z.$$

En particulier, une fonction borélienne $f : \mathbb{R}^d \to \mathbb{C}$ est $\mu * \nu$ intégrable si et seulement si $\tilde{f} : (y, z) \mapsto f(y + z)$ est $\mu \otimes \nu$ intégrable. Dans ce cas on a :

$$\int_{\mathbb{R}^d} f(x) \,\mathrm{d}(\mu * \nu)(x) = \int_{\mathbb{R}^d \times \mathbb{R}^d} f(y + z) \,\mathrm{d}(\mu \otimes \nu)(y, z).$$

Prenons comme fonction $f : x \mapsto \exp(i\langle t, x\rangle)$ pour $t \in \mathbb{R}^d$ fixé. La fonction $\tilde{f}$ est borélienne bornée sur $\mathbb{R}^d \times \mathbb{R}^d$, donc intégrable par rapport à la mesure finie $\mu \otimes \nu$. Par application du théorème de Fubini (plus précisément du corollaire 5.13), on obtient :

$$
\begin{aligned}
\widehat{\mu * \nu}(t) = \int_{\mathbb{R}^d} \exp(i\langle t, x\rangle)\,\mathrm{d}(\mu * \nu)(x) &= \int_{\mathbb{R}^d \times \mathbb{R}^d} \exp(i\langle t, y + z\rangle)\,\mathrm{d}(\mu \otimes \nu)(y, z) \\
&= \int_{\mathbb{R}^d \times \mathbb{R}^d} \exp(i\langle t, y\rangle)\exp(i\langle t, z\rangle)\,\mathrm{d}(\mu \otimes \nu)(y, z) \\
&= \int_{\mathbb{R}^d} \exp(i\langle t, y\rangle)\,\mathrm{d}\mu(y) \int_{\mathbb{R}^d} \exp(i\langle t, z\rangle)\,\mathrm{d}\nu(z) \\
&= \widehat{\mu}(t)\widehat{\nu}(t).
\end{aligned}
$$

Comme t était quelconque, cette égalité est vraie pour tout t et (9.11) est établie. Pour en déduire (9.12), il suffit de rappeler que si X et Y sont *indépendants*, la loi de leur somme est le produit de convolution : $P_{X+Y} = P_X * P_Y$. □

9.1.4 Caractérisation des mesures finies

Théorème 9.9. *Si deux mesures finies sur $\big(\mathbb{R}^d, \mathrm{Bor}(\mathbb{R}^d)\big)$ ont même transformée de Fourier, elles sont égales.*

Preuve. Il s'agit de montrer que l'égalité de *fonctions* $\widehat{\mu} = \widehat{\nu}$ implique l'égalité de mesures $\mu = \nu$.

Première étape. On réduit la preuve de l'égalité $\mu = \nu$ à celle de :

$$
\forall f : \mathbb{R}^d \to \mathbb{R}, \text{ continue à support compact,} \quad \int_{\mathbb{R}^d} f\,\mathrm{d}\mu = \int_{\mathbb{R}^d} f\,\mathrm{d}\nu. \tag{9.13}
$$

Supposons donc (9.13) vraie et soit

$$
C := \prod_{j=1}^{d}]a_j, b_j[,
$$

un pavé ouvert non vide de $\mathbb{R}^d$ (on a donc $a_j < b_j$ pour tout j). Définissons pour tout $n \geq n_0$, tel que $2/n_0 < \min_{1\leq j\leq d}(b_j - a_j)$, la fonction continue à support compact f_n par

$$
f_n := f_{n,1} \otimes \cdots \otimes f_{n,d},
$$

où $f_{n,j}$ 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 9.1.

On voit facilement que :

$$
\forall x \in \mathbb{R}^d, \quad f_n(x) \xrightarrow[n\to+\infty]{} \mathbf{1}_C(x).
$$

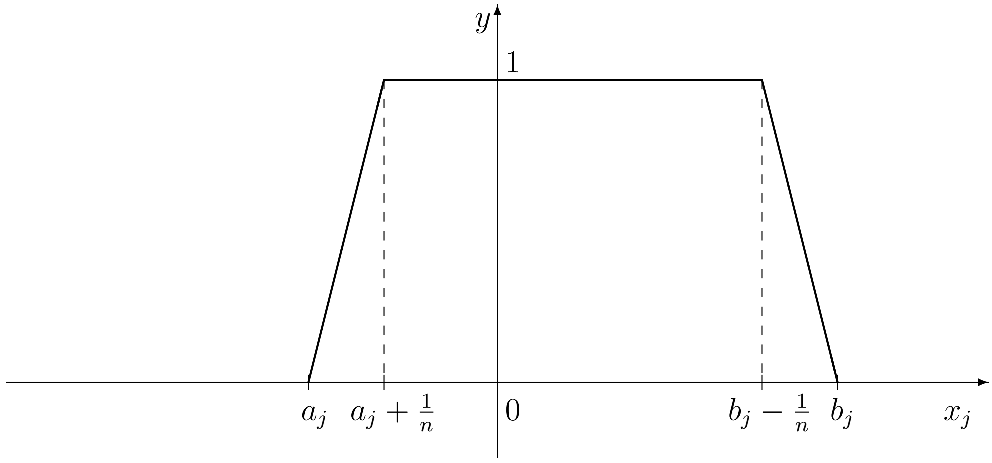


FIGURE 9.1 – Fonction $f_{n,j}$

Cette convergence ponctuelle[3] sur $\mathbb{R}^d$ est *dominée* par la fonction $\mathbf{1}_C$ qui est bornée, donc μ-intégrable puisque μ est une mesure finie. Par le théorème de convergence dominée, on en déduit :

$$\int_{\mathbb{R}^d} f_n \, \mathrm{d}\mu \xrightarrow[n\to+\infty]{} \int_{\mathbb{R}^d} \mathbf{1}_C \, \mathrm{d}\mu = \mu(C). \tag{9.14}$$

On a évidemment la même convergence d'intégrale avec ν à la place de μ. D'autre part, f_n étant continue à support compact, notre hypothèse provisoire (9.13) nous permet d'écrire :

$$\forall n \geq n_0, \quad \int_{\mathbb{R}^d} f_n \, \mathrm{d}\mu = \int_{\mathbb{R}^d} f_n \, \mathrm{d}\nu.$$

Compte-tenu de (9.14), on peut faire tendre n vers l'infini dans cette égalité pour obtenir $\mu(C) = \nu(C)$. Comme C était quelconque, on a établi la coïncidence des mesures finies μ et ν sur le π-système des pavés ouverts qui engendre la tribu borélienne de $\mathbb{R}^d$, donc μ et ν sont égales. □

Deuxième étape. On réduit la preuve de (9.13) à celle de :

$$\forall f : \mathbb{R}^d \to \mathbb{R} \text{ continue bornée}, \quad \forall n \in \mathbb{N}^*, \quad \int_{\mathbb{R}^d} f \, \mathrm{d}(\mu * \gamma_n) = \int_{\mathbb{R}^d} f \, \mathrm{d}(\nu * \gamma_n), \tag{9.15}$$

où $\gamma_n := \mathfrak{N}(0, n^{-1/2})^{\otimes d}$ est une mesure gaussienne. Pour cela il suffit de montrer que

$$\lim_{n\to+\infty} \int_{\mathbb{R}^d} f \, \mathrm{d}(\mu * \gamma_n) = \int_{\mathbb{R}^d} f \, \mathrm{d}\mu. \tag{9.16}$$

3. Remarquons qu'il est important ici d'avoir la convergence pour *tout* $x \in \mathbb{R}^d$ et pas seulement μ-presque partout. En effet la construction de f_n ne dépend aucunement de la mesure μ et le résultat doit être valable pour n'importe quelle mesure finie μ. S'il y avait ne serait-ce qu'un point x_0 tel que $f_n(x_0)$ ne converge pas vers $\mathbf{1}_C(x_0)$, il suffirait de prendre une mesure μ ayant une masse ponctuelle en x_0 pour que la convergence μ-p.p. de f_n vers $\mathbf{1}_C$ soit en défaut.

Remarquons d'abord que γ_n est la mesure image $\gamma_1 \circ h_n^{-1}$, de γ_1 par l'homothétie h_n : $x \mapsto n^{-1/2}x$. On en déduit que toute la masse de γ_n se concentre au voisinage du point 0 quand n tend vers l'infini. Plus précisément :

$$\forall \delta > 0, \quad \lim_{n\to+\infty} \gamma_n\Big(\mathbb{R}^d \setminus]-\delta, \delta[^d\Big) = 0. \tag{9.17}$$

En effet

$$\gamma_n\Big(]-\delta, \delta[^d\Big) = \gamma_1\Big(h_n^{-1}(]-\delta, \delta[^d)\Big) = \gamma_1\Big(]-\delta n^{1/2}, \delta n^{1/2}[^d\Big) \xrightarrow[n\to+\infty]{} \gamma_1(\mathbb{R}^d) = 1,$$

par continuité séquentielle croissante de la mesure de probabilité γ_1.

En exploitant (9.17) et la continuité de f, nous allons montrer que :

$$\forall x \in \mathbb{R}^d, \quad \int_{\mathbb{R}^d} f(x+z)\,\mathrm{d}\gamma_n(z) \xrightarrow[n\to+\infty]{} f(x). \tag{9.18}$$

En effet, fixons x quelconque dans $\mathbb{R}^d$, il existe pour tout $\varepsilon > 0$, un $\delta > 0$, dépendant de f, de x et de ε tel que

$$\forall z \in]-\delta, \delta[^d, \quad |f(x+z) - f(x)| < \varepsilon. \tag{9.19}$$

Comme γ_n est une probabilité sur $\mathbb{R}^d$, $f(x) = \int_{\mathbb{R}^d} f(x)\,\mathrm{d}\gamma_n(z)$, d'où

$$\begin{aligned}
\left|\int_{\mathbb{R}^d} f(x+z)\,\mathrm{d}\gamma_n(z) - f(x)\right| &= \left|\int_{\mathbb{R}^d}\Big(f(x+z) - f(x)\Big)\,\mathrm{d}\gamma_n(z)\right| \\
&\leq \int_{\mathbb{R}^d}\Big|f(x+z) - f(x)\Big|\,\mathrm{d}\gamma_n(z) \\
&\leq \int_{]-\delta,\delta[^d} \varepsilon\,\mathrm{d}\gamma_n + 2\|f\|_\infty \gamma_n\Big(\mathbb{R}^d \setminus]-\delta, \delta[^d\Big),
\end{aligned}$$

où l'on a posé $\|f\|_\infty := \sup_{\mathbb{R}^d} |f|$. Grâce à (9.17), il existe un entier n_0 tel que

$$\forall n \geq n_0, \quad 2\|f\|_\infty \gamma_n\Big(\mathbb{R}^d \setminus]-\delta, \delta[^d\Big) < \varepsilon.$$

Par conséquent,

$$\forall n \geq n_0, \quad \left|\int_{\mathbb{R}^d} f(x+z)\,\mathrm{d}\gamma_n(z) - f(x)\right| < \varepsilon\gamma_n\Big(]-\delta, \delta[^d\Big) + \varepsilon < 2\varepsilon.$$

La convergence (9.18) est ainsi établie [4].

Comme γ_n est une probabilité, on a la majoration uniforme :

$$\forall x \in \mathbb{R}^d, \quad \left|\int_{\mathbb{R}^d} f(x+z)\,\mathrm{d}\gamma_n(z)\right| \leq \|f\|_\infty. \tag{9.20}$$

La mesure μ étant finie, la constante $\|f\|_\infty$ est μ-intégrable. Grâce à (9.18) et (9.20), on peut appliquer le théorème de convergence dominée relativement à la mesure μ pour obtenir :

$$\int_{\mathbb{R}^d} f\,\mathrm{d}(\mu * \gamma_n) = \int_{\mathbb{R}^d}\left\{\int_{\mathbb{R}^d} f(x+z)\,\mathrm{d}\gamma_n(z)\right\}\mathrm{d}\mu(x) \xrightarrow[n\to+\infty]{} \int_{\mathbb{R}^d} f(x)\,\mathrm{d}\mu(x).$$

Ainsi nous avons établi (9.16) pour toute fonction f continue bornée, donc *a fortiori* pour toute fonction continue à support compact. □

4. Le lecteur attentif aura noté la ressemblance non fortuite avec la preuve du théorème de Fejér.

Troisième étape : preuve de (9.15). Il est temps de rappeler que l'hypothèse du théorème en cours de démonstration est l'égalité $\widehat{\mu}(t) = \widehat{\nu}(t)$ pour tout $t \in \mathbb{R}^d$. Pour achever la preuve du théorème, il nous reste à déduire de cette hypothèse l'égalité $\int f \,\mathrm{d}(\mu * \gamma_n) = \int f \,\mathrm{d}(\nu * \gamma_n)$. L'idée est d'exprimer $\int f \,\mathrm{d}(\mu * \gamma_n)$ comme une intégrale où μ n'intervient que par sa transformée de Fourier $\widehat{\mu}$. Il suffira alors de lui substituer $\widehat{\nu}$ pour conclure. Le point crucial est que la transformée de Fourier de la mesure gaussienne γ_n est, à un changement d'échelle près, la densité g_n de γ_n. Plus précisément, $\gamma_n = \mathfrak{N}(0, n^{-1/2})^{\otimes d}$ a pour densité

$$g_n(x) = \Big(\frac{n}{2\pi}\Big)^{d/2} \exp\Big(-\frac{n\|x\|^2}{2}\Big),$$

où $\|x\|$ désigne la norme euclidienne de $\mathbb{R}^d$. En combinant les propositions 9.4 et 9.6 on en déduit immédiatement que

$$\widehat{\gamma_n}(t) = \prod_{j=1}^{d} \exp\Big(-\frac{t_j^2}{2n}\Big) = \exp\Big(-\frac{\|t\|^2}{2n}\Big).$$

En posant $c_n := \big(\frac{n}{2\pi}\big)^{d/2}$, on voit ainsi que

$$\forall z \in \mathbb{R}^d, \quad g_n(z) = c_n \widehat{\gamma_n}(nz). \tag{9.21}$$

On en déduit que

$$\begin{aligned} \int_{\mathbb{R}^d} f \,\mathrm{d}(\mu * \gamma_n) &= \int_{\mathbb{R}^d} \Big\{ \int_{\mathbb{R}^d} f(x+z) g_n(z) \,\mathrm{d}\lambda_d(z) \Big\} \,\mathrm{d}\mu(x) \\ &= c_n \int_{\mathbb{R}^d} \Big\{ \int_{\mathbb{R}^d} f(x+z) \widehat{\gamma_n}(nz) \,\mathrm{d}\lambda_d(z) \Big\} \,\mathrm{d}\mu(x). \end{aligned} \tag{9.22}$$

Le changement de variable $y = x + z$ dans l'intégrale relativement à z dans (9.22) nous donne :

$$\int_{\mathbb{R}^d} f \,\mathrm{d}(\mu * \gamma_n) = c_n \int_{\mathbb{R}^d} \Big\{ \int_{\mathbb{R}^d} f(y) \widehat{\gamma_n}(ny - nx) \,\mathrm{d}\lambda_d(y) \Big\} \,\mathrm{d}\mu(x)$$

En remarquant que $\widehat{\gamma_n}(ny - nx) = \int_{\mathbb{R}^d} \mathrm{e}^{inyu} \mathrm{e}^{-inxu} \,\mathrm{d}\gamma_n(u)$, on obtient (sans problème pour la justification de l'interversion des intégrations via le théorème de Fubini puisque f est continue à support compact donc $\int_{\mathbb{R}^d} |f| \,\mathrm{d}\lambda_d < +\infty$) :

$$\begin{aligned} \int_{\mathbb{R}^d} f \,\mathrm{d}(\mu * \gamma_n) &= c_n \int_{\mathbb{R}^d} f(y) \int_{\mathbb{R}^d} \mathrm{e}^{inyu} \int_{\mathbb{R}^d} \mathrm{e}^{-inxu} \,\mathrm{d}\mu(x) \,\mathrm{d}\gamma_n(u) \,\mathrm{d}\lambda_d(y) \\ &= c_n \int_{\mathbb{R}^d} f(y) \int_{\mathbb{R}^d} \mathrm{e}^{inyu} \widehat{\mu}(-nu) \,\mathrm{d}\gamma_n(u) \,\mathrm{d}\lambda_d(y) && (9.23) \\ &= c_n \int_{\mathbb{R}^d} f(y) \int_{\mathbb{R}^d} \mathrm{e}^{inyu} \widehat{\nu}(-nu) \,\mathrm{d}\gamma_n(u) \,\mathrm{d}\lambda_d(y) && (9.24) \\ &= \int_{\mathbb{R}^d} f \,\mathrm{d}(\nu * \gamma_n). && (9.25) \end{aligned}$$

Le passage de (9.23) à (9.24) utilise l'hypothèse $\widehat{\mu} = \widehat{\nu}$, celui de (9.24) à (9.25) exploite le fait que le calcul menant à la formule (9.23) pour $\int_{\mathbb{R}^d} f \,\mathrm{d}(\mu * \gamma_n)$ est valable pour n'importe quelle mesure finie μ, donc en particulier pour ν. □

Le théorème 9.9 est maintenant complètement démontré. □

Voici une première application du théorème 9.9 à la caractérisation de l'indépendance des variables aléatoires par les fonctions caractéristiques.

Proposition 9.10. *Soit $X = (X_1, \dots, X_d)$ un vecteur aléatoire, notons φ sa fonction caractéristique et φ_j celle de X_j $(j = 1, \dots, d)$. Si*

$$\forall t = (t_1, \dots, t_d) \in \mathbb{R}^d, \quad \varphi(t) = \varphi_1(t_1) \dots \varphi_d(t_d), \tag{9.26}$$

alors les composantes de X sont mutuellement indépendantes.

Preuve. Notons ψ la fonction caractéristique de la loi $\nu := P_{X_1} \otimes \cdots \otimes P_{X_d}$. Par la proposition 9.4, $\psi(t) = \varphi_1(t_1) \dots \varphi_d(t_d)$, donc par l'hypothèse (9.26), $\psi = \varphi$, ce qui s'écrit aussi $\widehat{\nu} = \widehat{P_X}$. Le théorème 9.9 nous dit alors que $P_X = \nu$. Nous avons donc vérifié que $P_X = P_{X_1} \otimes \cdots \otimes P_{X_d}$, ce qui équivaut à l'indépendance des X_j. □

Comme la proposition 9.10 est la réciproque de la proposition 9.4, l'égalité (9.26) est une condition *nécessaire et suffisante* pour l'indépendance mutuelle de $X_1, \dots, X_d$.

Pour finir cette section, on donne une application des théorèmes 9.8 et 9.9 au calcul de lois de sommes de v.a. indépendantes.

Exemple 9.3 (Trois familles de lois stables par convolution)**.** La famille des lois gaussiennes sur $\mathbb{R}$, celle des lois de Poisson et celle des lois de Cauchy sont chacune stable par convolution. Plus précisément, on a les relations :

$$\operatorname{Pois}(\alpha) * \operatorname{Pois}(\beta) = \operatorname{Pois}(\alpha + \beta) \tag{9.27}$$

$$\mathfrak{N}(m_1, \sigma_1) * \mathfrak{N}(m_2, \sigma_2) = \mathfrak{N}\Big(m_1 + m_2, (\sigma_1^2 + \sigma_2^2)^{1/2}\Big) \tag{9.28}$$

$$\operatorname{Cau}(a_1, b_1) * \operatorname{Cau}(a_2, b_2) = \operatorname{Cau}(a_1 + a_2, b_1 + b_2). \tag{9.29}$$

Justifions (9.28), la vérification de (9.27) et (9.29) étant analogue sera laissée au lecteur. Pour alléger les écritures, notons $\mu_j := \mathfrak{N}(m_j, \sigma_j)$. En appliquant le théorème 9.8 et la proposition 9.6, on voit que :

$$\begin{aligned}
\forall t \in \mathbb{R}, \quad \widehat{\mu_1 * \mu_2}(t) = \widehat{\mu_1}(t)\widehat{\mu_2}(t) &= \exp\left(im_1 t - \frac{\sigma_1^2 t^2}{2}\right) \exp\left(im_2 t - \frac{\sigma_2^2 t^2}{2}\right) \\
&= \exp\left(i(m_1 + m_2)t - \frac{(\sigma_1^2 + \sigma_2^2)t^2}{2}\right) \\
&= \widehat{\mu}(t),
\end{aligned}$$

où $\mu := \mathfrak{N}\Big(m_1 + m_2, (\sigma_1^2 + \sigma_2^2)^{1/2}\Big)$. Ainsi $\mu_1 * \mu_2$ et μ ont même transformée de Fourier, donc d'après le théorème 9.9, $\mu_1 * \mu_2 = \mu$, ce qui n'est que l'écriture allégée de (9.28). Un énoncé équivalent à (9.28) est : *si X_1 et X_2 sont deux variables aléatoires gaussiennes* indépendantes *de paramètres respectifs (m_1, σ_1) et (m_2, σ_2), alors $S := X_1 + X_2$ est aussi gaussienne de paramètres (m, σ) avec $m = m_1 + m_2$ et $\sigma^2 = \sigma_1^2 + \sigma_2^2$.* Remarquons que dès que l'on sait que S est gaussienne, les relations entre les paramètres peuvent se retrouver en se rappelant que $m = \mathbf{E}S$ et $\sigma^2 = \operatorname{Var} S$. Il suffit alors d'appliquer la linéarité de l'espérance et la relation $\operatorname{Var} S = \operatorname{Var} X_1 + \operatorname{Var} X_2$ pour X_1 et X_2 *indépendantes*.

9.2 Transformée de Fourier d'une fonction

Définition 9.11. *Soit $f \in \mathrm{L}^1(\lambda_d)$, on définit sa transformée de Fourier $\widehat{f}$ par*

$$\forall t \in \mathbb{R}^d, \quad \widehat{f}(t) := \int_{\mathbb{R}^d} f(x) \exp(i\langle t, x\rangle) \, \mathrm{d}\lambda_d(x). \tag{9.30}$$

Remarque 9.12. Si μ est une mesure finie de densité f par rapport à λ_d, alors $\widehat{\mu} = \widehat{f}$.

Remarque 9.13. Il y a dans la littérature diverses autres définitions de la transformée de Fourier d'une fonction. Leur forme générale est $\widehat{f}(t) := a \int_{\mathbb{R}^d} f(x) \exp(ib\langle t, x\rangle) \, \mathrm{d}\lambda_d(x)$, le choix des constantes réelles a et b étant motivé par l'utilisation principale envisagée pour la transformée de Fourier. Par exemple si on veut la considérer comme isométrie euclidienne de $\mathrm{L}^2(\mathbb{R})$, il est commode de prendre $a = (2\pi)^{-1/2}$ et $b = 1$ ou -1. Nous avons adopté le choix qui permet d'identifier transformée de Fourier d'une densité de probabilité et fonction caractéristique de la loi correspondante.

9.2.1 Propriétés de la transformation de Fourier dans L^1

Nous exposons maintenant les propriétés de la transformée de Fourier en nous restreignant pour simplifier un peu, au cas $d = 1$.

Proposition 9.14. *Soient $f \in \mathrm{L}^1(\mathbb{R})$, $a \in \mathbb{R}$ et $c > 0$ quelconques.*

i) Si $g(x) = f(x)\mathrm{e}^{iax}$, alors $\widehat{g}(t) = \widehat{f}(t + a)$.

ii) Si $g(x) = f(x - a)$, alors $\widehat{g}(t) = \widehat{f}(t)\mathrm{e}^{ita}$.

iii) Si $g(x) = f(x/c)$, alors $\widehat{g}(t) = c\widehat{f}(ct)$.

Vérification. Si $g(x) = f(x)\mathrm{e}^{iax}$,

$$\widehat{g}(t) = \int_{\mathbb{R}} f(x)\mathrm{e}^{iax}\mathrm{e}^{itx} \, \mathrm{d}\lambda_1(x) = \int_{\mathbb{R}} f(x)\mathrm{e}^{i(a+t)x} \, \mathrm{d}\lambda_1(t) = \widehat{f}(t + a).$$

Si $g(x) = f(x - a)$, le changement de variable $y = x - a$ et l'invariance de λ_1 par translation nous donnent :

$$\widehat{g}(t) = \int_{\mathbb{R}} f(x - a)\mathrm{e}^{itx} \, \mathrm{d}\lambda_1(x) = \int_{\mathbb{R}} f(y)\mathrm{e}^{it(y+a)} \, \mathrm{d}\lambda_1(y) = \mathrm{e}^{ita} \int_{\mathbb{R}} f(y)\mathrm{e}^{ity} \, \mathrm{d}\lambda_1(y) = \widehat{f}(t)\mathrm{e}^{ita}.$$

Enfin si $g(x) = f(x/c)$, le changement de variable $y = x/c$ (la mesure image de λ_1 est alors $c\lambda_1$) donne :

$$\widehat{g}(t) = \int_{\mathbb{R}} f\Big(\frac{x}{c}\Big)\mathrm{e}^{itx} \, \mathrm{d}\lambda_1(x) = \int_{\mathbb{R}} f(y)\mathrm{e}^{itcy} c \, \mathrm{d}\lambda_1(y) = c\widehat{f}(ct).$$

□

Théorème 9.15. *Pour toutes f et $g \in \mathrm{L}^1(\mathbb{R})$,*

$$\widehat{f * g} = \widehat{f} \cdot \widehat{g}. \tag{9.31}$$

Preuve. Rappelons (définition 5.56) que si f et g sont dans $\mathrm{L}^1(\mathbb{R})$, leur produit de convolution est la classe de fonctions dont un représentant est défini λ-p.p. par :

$$(f*g)(y) := \int_{\mathbb{R}} f(x)g(y-x)\,\mathrm{d}\lambda_1(x) = \int_{\mathbb{R}} f(y-x)g(x)\,\mathrm{d}\lambda_1(x).$$

On sait (proposition 5.57) que si $f,g \in \mathrm{L}^1(\mathbb{R})$, alors $f*g$ est aussi un élément de L^1. Donc sa transformée de Fourier est bien définie et s'écrit :

$$\widehat{f*g}(t) = \int_{\mathbb{R}} \mathrm{e}^{ity}(f*g)(y)\,\mathrm{d}\lambda_1(y) = \int_{\mathbb{R}} \mathrm{e}^{ity}\left\{\int_{\mathbb{R}} f(x)g(y-x)\,\mathrm{d}\lambda_1(x)\right\}\mathrm{d}\lambda_1(y).$$

Comme $|f|*|g|$ est aussi dans L^1, on peut appliquer le théorème de Fubini pour permuter l'ordre des intégrations :

$$\widehat{f*g}(t) = \int_{\mathbb{R}}\left\{\int_{\mathbb{R}} \mathrm{e}^{ity} f(x)g(y-x)\,\mathrm{d}\lambda_1(y)\right\}\mathrm{d}\lambda_1(x).$$

Pour x fixé, on effectue dans l'intégrale intérieure le changement de variable $u = y - x$ qui nous donne (λ_1 étant invariante par translation) :

$$\begin{aligned}
\widehat{f*g}(t) &= \int_{\mathbb{R}}\left\{\int_{\mathbb{R}} \mathrm{e}^{it(x+u)} f(x)g(u)\,\mathrm{d}\lambda_1(u)\right\}\mathrm{d}\lambda_1(x)\\
&= \int_{\mathbb{R}} \mathrm{e}^{itx} f(x)\left\{\int_{\mathbb{R}} \mathrm{e}^{itu} g(u)\,\mathrm{d}\lambda_1(u)\right\}\mathrm{d}\lambda_1(x)\\
&= \left\{\int_{\mathbb{R}} \mathrm{e}^{itu} g(u)\,\mathrm{d}\lambda_1(u)\right\}\int_{\mathbb{R}} \mathrm{e}^{itx} f(x)\,\mathrm{d}\lambda_1(x)\\
&= \widehat{f}(t)\widehat{g}(t).
\end{aligned}$$

□

Proposition 9.16 (Dérivée d'une transformée de Fourier)**.** *Si f et la fonction $g : x \mapsto xf(x)$ sont toutes deux dans $\mathrm{L}^1(\mathbb{R})$, alors $\widehat{f}$ est dérivable et*

$$\frac{\mathrm{d}}{\mathrm{d}t}\widehat{f}(t) = i\,\widehat{g}(t) \tag{9.32}$$

Vérification. C'est une application directe du théorème de dérivation sous le signe somme. □

Théorème 9.17 (Transformée de Fourier d'une dérivée)**.** *Si f et f' sont dans L^1, alors*

$$\widehat{\left(\frac{\mathrm{d}f}{\mathrm{d}x}\right)}(t) = \widehat{f'}(t) = -it\widehat{f}(t). \tag{9.33}$$

Preuve. Examinons d'abord le cas particulier où f est C^1 à support compact $[a, b]$. Alors f et f' sont nulles aux points a et b et en dehors de $[a, b]$. On obtient facilement (9.33) en passant aux intégrales de Riemann et en intégrant par parties :

$$\begin{aligned}\widehat{f'}(t) = \int_{[a,b]} \mathrm{e}^{itx} f'(x)\,\mathrm{d}\lambda_1(x) &= \int_a^b \mathrm{e}^{itx} f'(x)\,\mathrm{d}x = \Big[\mathrm{e}^{itx} f(x)\Big]_a^b - it \int_a^b \mathrm{e}^{itx} f(x)\,\mathrm{d}x \\ &= -it \int_{-\infty}^{+\infty} \mathrm{e}^{itx} f(x)\,\mathrm{d}x \\ &= -it \int_{\mathbb{R}} \mathrm{e}^{itx} f(x)\,\mathrm{d}\lambda_1(x).\end{aligned}$$

Nous admettrons que l'on peut passer au cas général par un argument de densité. □

Remarque 9.18. Le théorème 9.17 énonce la deuxième propriété essentielle de la transformée de Fourier (la première étant l'effet sur la convolution). La transformée de Fourier transforme une dérivation en une multiplication par la fonction monôme $t \mapsto -it$. Plus généralement si D est un opérateur différentiel de la forme :

$$D := \sum_{k=1}^{n} c_k \frac{\mathrm{d}^k}{\mathrm{d}x^k},$$

on a pour toute f telle que $f, f', \dots, f^{(n)}$ soient dans L^1,

$$\widehat{Df}(t) = Q_n(t)\widehat{f}(t),$$

où Q est un polynôme. Cette propriété est utile pour la résolution d'équations différentielles de la forme $Df = g$ qui deviennent par transformée de Fourier $Q_n\widehat{f} = \widehat{g}$, d'où $\widehat{f} = Q_n^{-1}\widehat{g}$. On peut en déduire une solution f si l'on sait calculer la transformée de Fourier inverse de $Q_n^{-1}\widehat{g}$.

Théorème 9.19 (Riemann-Lebesgue)**.** *Si $f \in \mathrm{L}^1(\mathbb{R})$, $\widehat{f} \in \mathcal{C}_0(\mathbb{R})$, autrement dit $\widehat{f}$ est continue sur $\mathbb{R}$ et tend vers 0 en $-\infty$ et en $+\infty$. On a*

$$\|\widehat{f}\|_\infty \le \|f\|_1.$$

Preuve. La continuité de $\widehat{f}$ résulte d'une application immédiate du théorème de continuité sous le signe somme.

Pour la limite en $\pm\infty$, on peut écrire en remarquant que $\mathrm{e}^{i\pi} = -1$,

$$\forall t \in \mathbb{R}^*, \quad \widehat{f}(t) = -\int_{\mathbb{R}} f(x) \exp\Big(it(x + \pi/t)\Big)\,\mathrm{d}\lambda_1(x) = -\int_{\mathbb{R}} f(y - \pi/t)\mathrm{e}^{ity}\,\mathrm{d}\lambda_1(y).$$

On en déduit

$$2\widehat{f}(t) = \int_{\mathbb{R}} \Big(f(y) - f(y - \pi/t)\Big)\mathrm{e}^{ity}\,\mathrm{d}\lambda_1(y),$$

d'où

$$\forall t \in \mathbb{R}^*, \quad 2|\widehat{f}(t)| \le \|f - f(\,.\, - \pi/t)\|_1.$$

On conclut en faisant tendre t vers $+\infty$ (ou $-\infty$) et avec le lemme suivant. □

Lemme 9.20. *Pour* $1 \leq p < +\infty$*, les translations* opèrent continûment *sur* $\mathrm{L}^p(\mathbb{R})$*, ce qui signifie que pour tout* $f \in \mathrm{L}^p(\mathbb{R})$*,*

$$\lim_{h \to 0} \|f - f(\,.+h)\|_p = 0. \tag{9.34}$$

Preuve. Vérifions d'abord (9.34) dans le cas particulier où f est continue à support compact inclus dans $[a, b]$. Dans ce cas on a pour tout $h \in [-1, 1]$,

$$\|f - f(\,.+h)\|_p^p = \int_{[a-1,b+1]} |f(x) - f(x+h)|^p \,\mathrm{d}\lambda_1(x)$$

et on obtient facilement (9.34) par convergence dominée, en prenant pour fonction dominante la constante $2^p \sup_{\mathbb{R}} |f|^p$.

Dans le cas général, soit f un élément quelconque de $\mathrm{L}^p(\mathbb{R})$ et $\varepsilon > 0$. Par le théorème 6.26, l'espace $\mathcal{C}_c(\mathbb{R})$ des fonctions continues à support compact est dense dans $\mathrm{L}^p(\mathbb{R})$, on peut donc trouver une fonction $g \in \mathcal{C}_c(\mathbb{R})$ telle que $\|f - g_\varepsilon\|_p < \varepsilon$. En raison de l'invariance de la mesure de Lebesgue par translation, il est facile de voir que $\|f(\,.+h) - g_\varepsilon(\,.+h)\|_p = \|f - g_\varepsilon\|_p$. Par inégalité triangulaire dans $\mathrm{L}^p(\mathbb{R})$, on a :

$$\begin{aligned}
\|f - f(\,.+h)\|_p &\leq \|f - g_\varepsilon\|_p + \|g_\varepsilon - g_\varepsilon(\,.+h)\|_p + \|f(\,.+h) - g_\varepsilon(\,.+h)\|_p \\
&\leq \|g_\varepsilon - g_\varepsilon(\,.+h)\|_p + 2\varepsilon.
\end{aligned}$$

Par le cas particulier étudié ci-dessus, il existe $\delta > 0$ tel que pour tout h tel que $|h| < \delta$, $\|g_\varepsilon - g_\varepsilon(\,.+h)\|_p < \varepsilon$, d'où $\|f - f(\,.+h)\|_p < 3\varepsilon$. Ce raisonnement étant valable pour tout $\varepsilon > 0$, on a établi (9.34). □

Corollaire 9.21. *Si* $f \in \mathrm{L}^1(\mathbb{R})$*,* $\widehat{f}$ *est* uniformément *continue sur* $\mathbb{R}$.

Justification. C'est un exercice classique d'analyse : si f est continue sur $\mathbb{R}$ et tend vers 0 en $-\infty$ et $+\infty$, elle est uniformément continue sur $\mathbb{R}$. □

9.2.2 Formule d'inversion

Si l'on veut exploiter pleinement les bonnes propriétés de la transformée de Fourier pour résoudre des équations de convolution ou des équations différentielles, il importe de savoir l'inverser.

Théorème 9.22. *Si* f *et* $\widehat{f}$ *sont toutes deux dans* $\mathrm{L}^1(\mathbb{R})$*, on a la formule d'inversion,*

$$f(x) = \frac{1}{2\pi} \int_{\mathbb{R}} \widehat{f}(t) \mathrm{e}^{-ixt} \,\mathrm{d}\lambda_1(t) \quad \text{pour } \lambda_1\text{-presque tout } x \in \mathbb{R}. \tag{9.35}$$

Nous admettrons ce théorème.

9.3 Convergence en loi

9.3.1 Discussion introductive

Nous abordons maintenant l'étude de la convergence des lois de probabilité. Si μ_n et μ sont des probabilités sur $\bigl(\mathbb{R}^d, \mathrm{Bor}(\mathbb{R}^d)\bigr)$, la première notion de convergence de μ_n vers μ qui vienne à l'esprit est la convergence *ponctuelle* de la suite de *fonctions d'ensembles* μ_n définies sur $\mathrm{Bor}(\mathbb{R}^d)$ vers μ. Cette convergence est définie par :

$$\forall B \in \mathrm{Bor}(\mathbb{R}^d), \quad \mu_n(B) \xrightarrow[n\to+\infty]{} \mu(B). \tag{9.36}$$

Il se trouve que cette notion est trop restrictive quand on veut l'appliquer à l'étude de la convergence des lois des variables aléatoires. L'exemple simple suivant (avec $d = 1$) devrait nous en convaincre.

Exemple 9.4. Considérons la suite de variables aléatoires réelles $(X_n)_{n\geq 1}$, où pour chaque $n \geq 1$, X_n suit la loi uniforme sur l'intervalle $[-1/n, 1/n]$. Il est facile de vérifier que X_n converge vers la v.a. constante 0 pour chacun des modes de convergence déjà vus (p.s., en probabilité et dans tous les espaces $\mathrm{L}^p(\Omega)$, y compris $\mathrm{L}^\infty(\Omega)$). On s'attend donc légitimement à ce que la loi μ_n de X_n converge vers la loi de la v.a. constante 0, c'est à dire vers $\mu = \delta_0$, masse de Dirac en 0. Or cette convergence ne peut avoir lieu au sens de (9.36). En effet pour $B = [0, 1]$, on a clairement $\mu_n(B) = 1/2$ pour tout $n \geq 1$, la suite $\mu_n(B)$ converge donc vers $1/2$, tandis que $\mu(B) = \delta_0([0, 1]) = 1$. On a le même problème avec $B' =]0, 1]$, $\mu_n(B') = 1/2$ et $\mu(B') = 0$. Soit maintenant un intervalle I d'extrémités a et b avec $a \neq 0$ et $b \neq 0$. Vérifions que $\mu_n(I)$ converge vers $\delta_0(I)$. Puisque μ_n est une loi uniforme en dimension 1, $\mu_n(I)$ se calcule comme un rapport de longueurs :

$$\mu_n(I) = \frac{\lambda\bigl([-1/n, 1/n] \cap I\bigr)}{\lambda\bigl([-1/n, 1/n]\bigr)} = \frac{n}{2}\lambda\bigl([-1/n, 1/n] \cap I\bigr).$$

Comme ni a ni b n'est nul, nous avons seulement deux cas à envisager, $0 \in]a, b[$ et $0 \in \mathbb{R} \setminus [a, b]$. Dans le premier cas, pour tout n supérieur à un certain n_0, on a l'inclusion $[-1/n, 1/n] \subset]a, b[\subset I$, d'où $\mu_n(I) = \frac{n}{2}\lambda\bigl([-1/n, 1/n]\bigr) = 1$. Dans ce cas, $\mu_n(I)$ converge vers 1, donc vers $\delta_0(I)$, puisque si $0 \in I$, $\delta_0(I) = 1$. Dans le deuxième cas, on a pour tout n supérieur à un certain n_1, $[-1/n, 1/n] \cap [a, b] = \emptyset$, d'où *a fortiori* $[-1/n, 1/n] \cap I = \emptyset$ et $\mu_n(I) = 0$. Alors $\mu_n(I)$ converge vers $0 = \delta_0(I)$ $(0 \notin I)$.

Il semble donc que la pathologie présentée par l'exemple 9.4 soit liée à la présence sur la frontière du borélien B de masse ponctuelle pour la mesure limite μ. L'exemple suivant permet d'affiner cette impression et de voir que le problème n'est pas tant le fait que μ ait une masse ponctuelle sur la frontière, mais plus généralement que la frontière ait une μ-mesure non nulle.

Exemple 9.5. Considérons la suite de vecteurs aléatoires (X_n, Y_n) de $\mathbb{R}^2$ de loi μ_n uniforme sur $[0, 1] \times [0, 1/n]$. Intuitivement, la loi de μ_n devrait converger vers la loi μ

du vecteur aléatoire (X, Y), où X suit la loi uniforme sur $[0, 1]$ et Y est la v.a. constante zéro. Autrement dit $\mu = \mathrm{Unif}[0, 1] \otimes \delta_0$. À nouveau on constate que cette convergence ne peut avoir lieu au sens de (9.36), puisque pour $B = [0, 1/2] \times [-1, 0]$, $\mu_n(B) = 0$ pour tout n, tandis que $\mu(B) = \lambda_1([0, 1/2]) = 1/2$. Ici il n'y a pas de masse ponctuelle pour μ, mais la mesure de la frontière de B est non nulle.

Au vu des exemples 9.4 et 9.5 et avec un peu de bonne volonté, on se résout à modifier comme suit la définition de la convergence de μ_n vers μ proposée par (9.36) :

$$\forall B \in \mathrm{Bor}(\mathbb{R}^d) \text{ tel que } \mu(\partial B) = 0, \quad \mu_n(B) \xrightarrow[n \to +\infty]{} \mu(B). \tag{9.37}$$

Nous avons noté $\partial B = \overline{B} \setminus \overset{\circ}{B}$ la frontière de B. Pour une autre illustration de la pertinence de la condition $\mu(\partial B) = 0$, voir la remarque 9.39 sur le théorème de De Moivre-Laplace.

En fait, pour des raisons de maniabilité et pour permettre un lien ultérieur avec la notion de convergence faible étudiée en analyse fonctionnelle, nous choisirons comme définition de la convergence de μ_n vers μ la condition suivante dont l'équivalence avec (9.37) sera établie ci-dessous (théorème 9.32) :

$$\forall f : \mathbb{R}^d \to \mathbb{R}, \text{ continue bornée}, \quad \int_{\mathbb{R}^d} f \, \mathrm{d}\mu_n \xrightarrow[n \to +\infty]{} \int_{\mathbb{R}^d} f \, \mathrm{d}\mu. \tag{9.38}$$

À première vue, (9.38) semble très différente de (9.36) et (9.37). En fait la différence n'est pas si grande si l'on se rappelle que $\mu_n(B)$ peut aussi s'écrire sous la forme $\int_{\mathbb{R}^d} f \, \mathrm{d}\mu_n$ avec $f = \mathbf{1}_B$. Le fait de « tester » la convergence de μ_n vers μ sur les intégrales des fonctions continues bornées plutôt que sur celles des indicatrices des boréliens permet de « lisser » un peu les choses et d'éviter les pathologies présentées ci-dessus. On pourra vérifier à titre d'exercice que la condition (9.38) est réalisée dans les exemples 9.4 et 9.5.

9.3.2 Définition et propriétés

Définition 9.23 (Convergence étroite)**.** *Soient μ_n $(n \geq 1)$ et μ des mesures de probabilité sur $\big(\mathbb{R}^d, \mathrm{Bor}(\mathbb{R}^d)\big)$. On dit que μ_n converge* étroitement *vers μ quand n tend vers l'infini si pour toute fonction continue bornée $f : \mathbb{R}^d \to \mathbb{R}$, $\int_{\mathbb{R}^d} f \, \mathrm{d}\mu_n$ converge vers $\int_{\mathbb{R}^d} f \, \mathrm{d}\mu$ quand n tend vers l'infini.*

Définition 9.24 (Convergence en loi)**.** *Soient X_n $(n \geq 1)$ et X des vecteurs aléatoires de $\mathbb{R}^d$. On dit que X_n converge* en loi *vers X si la loi de X_n converge étroitement vers celle de X quand n tend vers l'infini. Autrement dit,*

$$X_n \xrightarrow[n \to +\infty]{\text{loi}} X \quad \text{si} \quad \forall f : \mathbb{R}^d \to \mathbb{R} \text{ continue bornée}, \quad \mathbf{E} f(X_n) \xrightarrow[n \to +\infty]{} \mathbf{E} f(X). \tag{9.39}$$

Remarque 9.25. Il n'est pas nécessaire dans la définition 9.24 de supposer les X_n définis sur le même espace probabilisé $(\Omega, \mathcal{F}, \mathbf{P})$, puisque seule intervient leur loi. On peut très bien envisager qu'ils soient définis sur des espaces probabilisés différents :

$X_n : (\Omega_n, \mathcal{F}_n, \mathbf{P}_n) \to \big(\mathbb{R}^d, \mathrm{Bor}(\mathbb{R}^d)\big)$. L'écriture de la convergence étroite de P_{X_n} vers P_X sous la forme (9.39) utilise implicitement le théorème de transfert puisque

$$\mathbf{E}f(X_n) = \int_{\Omega_n} f \circ X_n \,\mathrm{d}\mathbf{P}_n = \int_{\mathbb{R}^d} f \,\mathrm{d}P_{X_n}.$$

Remarque 9.26. Bien que traditionnelle et admise par l'ensemble des probabilistes et statisticiens, la définition 9.24 contient un *grossier abus de langage* qui est la source de bien des erreurs pour les débutants. La convergence en loi de la suite (X_n) vers le vecteur aléatoire X n'est pas une vraie convergence de suite de vecteurs aléatoires, c'est seulement la *convergence de la loi* de X_n vers la *loi* de X au sens de la définition 9.23 (qui elle, est une vraie convergence de suite de *mesures*). On peut donc dire que X_n converge aussi en loi vers *n'importe quel* vecteur aléatoire Y de *même loi* que X. Par conséquent, *il n'y a pas unicité de la limite* pour la convergence en loi. Il convient donc d'être très prudent avec la manipulation de cette convergence. Par exemple si X_n converge en loi vers X et Y_n converge en loi vers Y, on ne peut pas dire que $X_n + Y_n$ converge en loi vers $X + Y$. Pour le voir (avec $d = 1$), prenons X_n de loi symétrique, disons $\mathbf{P}(X_n = -1) = \mathbf{P}(X_n = +1) = 1/2$. Posons pour tout n, $Y_n := -X_n$. Il est clair que Y_n et X_n ont même loi et comme cette loi ne dépend pas de n, il y a bien convergence en loi de la suite (X_n) vers une variable aléatoire X de même loi que X_n. Comme Y_n a même loi que X_n pour tout n, il est correct de dire que Y_n converge elle aussi en loi vers X. On n'a pourtant pas convergence de $X_n + Y_n$ vers $X + X$, puisque pour tout n, $X_n + Y_n = 0$ donc $X_n + Y_n$ a pour loi δ_0 ne dépendant pas de n. Ainsi $X_n + Y_n$ converge en loi vers 0. La loi de $X + X$ n'est évidemment pas δ_0, c'est $\frac{1}{2}\delta_{-2} + \frac{1}{2}\delta_2$. Bien sûr, on peut *aussi* dire que Y_n converge en loi vers $Y := -X$ et avec ce choix particulier de Y (mais encore une fois une infinité d'autres choix sont possibles), $X_n + Y_n$ converge en loi vers $X + Y$.

Remarque 9.27. Maintenant que l'on a bien compris que dans la convergence en loi de X_n vers un vecteur aléatoire X de loi μ, la limite X n'est pas unique, on s'autorisera un abus de langage supplémentaire nous permettant d'écrire quand on ne souhaite pas choisir X parmi tous les vecteurs aléatoires ayant la loi μ : $X_n \xrightarrow{\text{loi}} \mu$. Nous userons de cette facilité dans l'énoncé du théorème limite central en écrivant

$$S_n^* \xrightarrow[n\to+\infty]{\text{loi}} \mathfrak{N}(0,1).$$

Voici un premier exemple de convergence étroite, obtenu par utilisation directe de la définition 9.23.

Exemple 9.6 (Convergence de lois uniformes discrètes)**.** Pour $n \geq 2$, soit μ_n la loi uniforme discrète sur l'ensemble fini $E_n := \{1/n, 2/n, \dots, n/n\}$. Autrement dit :

$$\mu_n = \frac{1}{n}\sum_{k=1}^{n} \delta_{k/n}.$$

Quand n tend vers l'infini, μ_n converge étroitement vers μ loi uniforme sur $[0,1]$, i.e. μ est la mesure de densité $\mathbf{1}_{[0,1]}$ par rapport à λ_1. En effet, soit $f : \mathbb{R} \to \mathbb{R}$ une fonction continue bornée. On a

$$\int_{\mathbb{R}} f \,\mathrm{d}\mu_n = \frac{1}{n} \sum_{k=1}^{n} \int_{\mathbb{R}} f \,\mathrm{d}\delta_{k/n} = \frac{1}{n} \sum_{k=1}^{n} f\Bigl(\frac{k}{n}\Bigr).$$

On reconnaît là une *somme de Riemann* associée à f et au partage de $[0,1]$ constitué par E_n. Il est bien connu que cette somme converge (puisque f est continue) vers l'intégrale de Riemann de f entre 0 et 1 :

$$\int_{\mathbb{R}} f \,\mathrm{d}\mu_n \xrightarrow[n\to+\infty]{} \int_0^1 f(x)\,\mathrm{d}x = \int_{[0,1]} f \,\mathrm{d}\lambda_1 = \int_{\mathbb{R}} f \mathbf{1}_{[0,1]} \,\mathrm{d}\lambda_1 = \int_{\mathbb{R}} f \,\mathrm{d}\mu.$$

Comme la fonction continue bornée f était quelconque, ceci établit la convergence étroite de μ_n vers μ. Sur le même sujet, voir aussi [ICP] section 8.1.

Le théorème suivant est un outil très puissant pour obtenir des convergences en loi.

Théorème 9.28 (de l'image continue)**.** *Soient X_n $(n \geq 1)$ et X des vecteurs aléatoires de $\mathbb{R}^d$ et $g : \mathbb{R}^d \to \mathbb{R}^k$ une application continue. Si X_n converge en loi vers X, alors la suite de vecteurs aléatoires $g(X_n)$ dans $\mathbb{R}^k$ converge en loi vers le vecteur aléatoire $g(X)$.*

Preuve. Par hypothèse, $\mathbf{E}f(X_n)$ tend vers $\mathbf{E}f(X)$ pour toute $f : \mathbb{R}^d \to \mathbb{R}$ continue bornée. Posons $Y_n := g(X_n)$ et $Y := g(X)$. Il s'agit de montrer que $\mathbf{E}h(Y_n)$ tend vers $\mathbf{E}h(Y)$ pour toute $h : \mathbb{R}^k \to \mathbb{R}$, continue bornée. Or $\mathbf{E}h(Y_n) = \mathbf{E}\bigl((h \circ g)(X_n)\bigr)$ et $\mathbf{E}h(Y) = \mathbf{E}\bigl((h \circ g)(X)\bigr)$, il suffit alors pour conclure de remarquer que $f := h \circ g$ est une fonction continue bornée de $\mathbb{R}^d \to \mathbb{R}$ et d'appliquer l'hypothèse. Il importe de remarquer que nous n'avons pas supposé g bornée. □

Voici une application simple mais bien utile du théorème de l'image continue.

Corollaire 9.29. *Si le vecteur aléatoire (X_n, Y_n) de $\mathbb{R}^2$ converge en loi vers le vecteur aléatoire (X, Y), alors*

a) $X_n \xrightarrow[n\to+\infty]{\text{loi}} X$ *et* $Y_n \xrightarrow[n\to+\infty]{\text{loi}} Y$.

b) $\forall (a,b) \in \mathbb{R}^2$, $aX_n + bY_n \xrightarrow[n\to+\infty]{\text{loi}} aX + bY$,

c) $X_n Y_n \xrightarrow[n\to+\infty]{\text{loi}} XY$.

Vérification. Il suffit de remarquer que d'après le théorème 9.28, si $g : \mathbb{R}^2 \to \mathbb{R}$ est continue, $g(X_n, Y_n)$ converge en loi vers $g(X, Y)$. Voici le choix explicite de g dans chaque cas :

a) $g(x,y) = x$, puis $g(x,y) = y$.

b) $g(x,y) = ax + by$.

c) $g(x,y) = xy$.

□

Remarque 9.30. Il convient de noter que la réciproque du a) est fausse, car sinon on pourrait grâce au b) en déduire que les convergences en loi respectives de X_n vers X et de Y_n vers Y impliquent celle de $X_n + Y_n$ vers $X + Y$. On sait que cette implication est fausse cf. remarque 9.25.

La remarque suivante sur le même sujet est à sauter en première lecture.

Remarque 9.31. Par ailleurs, la convergence en loi de X_n vers X et de Y_n vers Y n'interdit pas que pour chaque n, X_n et Y_n soient définis sur des espaces probabilisés $(\Omega_n, \mathcal{F}_n, \mathbf{P}_n)$ et $(\Omega'_n, \mathcal{F}'_n, \mathbf{P}'_n)$ différents[5]. La définition de $X_n + Y_n$ comme application mesurable n'est alors même pas claire. Une possibilité serait de la définir comme application de $(\Omega_n \times \Omega'_n, \mathcal{F}_n \otimes \mathcal{F}'_n)$ dans $(\mathbb{R}, \mathrm{Bor}(\mathbb{R}))$ en posant $(X_n + Y_n)(\omega, \omega') := X_n(\omega) + Y_n(\omega')$. Ceci laisse entière la question du choix de la probabilité sur $(\Omega_n \times \Omega'_n, \mathcal{F}_n \otimes \mathcal{F}'_n)$. Par exemple, munir cet espace mesurable de $\mathbf{P}_n \otimes \mathbf{P}'_n$ reviendrait à supposer que X_n et Y_n sont indépendantes, ce qui n'est pas dans l'hypothèse...Ceci explique pourquoi dans l'étude de la convergence en loi de sommes $S_n = X_{n,1} + \dots + X_{n,k_n}$ de variables aléatoires, on suppose en général que pour chaque n, les termes $X_{n,j}$ de S_n sont des variables aléatoires définies sur le même espace espace probabilisé $(\Omega_n, \mathcal{F}_n, \mathbf{P}_n)$.

Le théorème suivant rassemble les diverses caractérisations de la convergence étroite, à l'exception de celle relative à la convergence des transformées de Fourier qui fait l'objet de la sous-section 9.3.3. Dans la littérature anglaise, il est connu sous le sobriquet de *portmanteau theorem*, qu'il serait erroné de traduire par « théorème du portemanteau », puisque le faux-ami *portmanteau* désigne en fait une grande valise à deux compartiments[6]. Le premier compartiment *i)–iii)* contient les caractérisations fonctionnelles (i.e. en termes de convergences d'intégrales de fonctions) le deuxième *iv)–vi)* contient les caractérisations ensemblistes (convergences de mesures de boréliens).

Théorème 9.32. *Soient μ_n ($n \geq 1$) et μ des mesures de probabilité sur $\big(\mathbb{R}^d, \mathrm{Bor}(\mathbb{R}^d)\big)$. Les fonctions f ci-dessous sont toutes définies sur $\mathbb{R}^d$ et à valeurs dans $\mathbb{R}$. Les assertions suivantes sont équivalentes.*

i) Pour toute f continue bornée, $\int_{\mathbb{R}^d} f \,\mathrm{d}\mu_n \xrightarrow[n\to+\infty]{} \int_{\mathbb{R}^d} f \,\mathrm{d}\mu$.

ii) Pour toute f uniformément *continue sur $\mathbb{R}^d$ et bornée, $\int_{\mathbb{R}^d} f \,\mathrm{d}\mu_n \xrightarrow[n\to+\infty]{} \int_{\mathbb{R}^d} f \,\mathrm{d}\mu$.*

iii) Pour toute f continue à support compact, $\int_{\mathbb{R}^d} f \,\mathrm{d}\mu_n \xrightarrow[n\to+\infty]{} \int_{\mathbb{R}^d} f \,\mathrm{d}\mu$.

5. Cette difficulté ne se présente pas quand on parle du *vecteur aléatoire* (X_n, Y_n) qui est par définition, une application mesurable d'un espace $(\Omega_n, \mathcal{F}_n, \mathbf{P}_n)$ dans $(\mathbb{R}^2, \mathrm{Bor}(\mathbb{R}^2))$. Dans ce cas les composantes X_n et Y_n sont des applications mesurables du même espace probabilisé $(\Omega_n, \mathcal{F}_n, \mathbf{P}_n)$ dans $(\mathbb{R}, \mathrm{Bor}(\mathbb{R}))$.

6. D'après le Collins Cobuild : **1** *A portmanteau is a large travelling case which opens out into two equal compartments ; an old fashioned use.* **2** *Portmanteau is used to describe* **2.1** *a word that combines parts of the forms and meanings of two other words. For example, 'brunch' is formed from 'breakfast' and 'lunch'...* **2.2** *someone or something that combines many different features or uses.*

iv) Pour tout fermé C de $\mathbb{R}^d$, $\limsup_{n\to+\infty} \mu_n(C) \leq \mu(C)$.

v) Pour tout ouvert G de $\mathbb{R}^d$, $\liminf_{n\to+\infty} \mu_n(G) \geq \mu(G)$.

vi) Pour tout borélien B tel que $\mu(\partial B) = 0$, $\lim_{n\to+\infty} \mu_n(B) = \mu(B)$, où l'on a noté $\partial B := \overline{B} \setminus \overset{\circ}{B}$ la frontière de B.

Si $d = 1$, en notant F_n et F les fonctions de répartition respectives de μ_n et μ, les conditions i)–vi) sont équivalentes à :

vii) Pour tout point de continuité x de F, $\lim_{n\to+\infty} F_n(x) = F(x)$.

Preuve. La preuve de l'équivalence des conditions $i)$ à $vi)$ s'articule selon le schéma :

$$i) \Rightarrow ii) \Rightarrow iii) \Rightarrow i); \quad i) \Rightarrow iv); \quad iv) \Leftrightarrow v); \quad iv) \text{ et } v) \Rightarrow vi); \quad vi) \Rightarrow i).$$

Les implications $i) \Rightarrow ii) \Rightarrow iii)$ sont évidentes par emboîtement des ensembles de fonctions f concernées.

Preuve de $iii) \Rightarrow i)$. La propriété cruciale permettant de remonter de $iii)$ à $i)$ est le fait que pour une probabilité μ sur $\mathbb{R}^d$, toute la masse est portée à ε près par un *compact*. On dit que μ est *tendue*. Plus précisément, nous allons utiliser la version fonctionnelle suivante de cette propriété :

$$\forall \varepsilon > 0, \exists h_\varepsilon : \mathbb{R}^d \to [0,1], \text{ continue à support compact}, \quad 0 \leq 1 - \int_{\mathbb{R}^d} h_\varepsilon \,\mathrm{d}\mu < \varepsilon. \tag{9.40}$$

En effet, la suite croissante (pour l'inclusion) des compacts $B(k) := [-k,k]^d$, $k \in \mathbb{N}^*$ a pour réunion $\mathbb{R}^d$, donc par continuité séquentielle croissante de μ, $\mu\big(B(k)\big) \uparrow \mu(\mathbb{R}^d) = 1$. On peut donc trouver un entier k_ε tel que $\mu\big(B(k_\varepsilon)\big) > 1 - \varepsilon$. On définit alors

$$h_\varepsilon = h_{\varepsilon,1} \otimes \cdots \otimes h_{\varepsilon,d},$$

où $h_{\varepsilon,j} : \mathbb{R} \to [0,1]$ est la fonction continue valant 0 hors de $[-k_\varepsilon - 1, k_\varepsilon + 1]$, valant 1 sur $[-k_\varepsilon, k_\varepsilon]$ et est affine sur chacun des intervalles $[-k_\varepsilon - 1, -k_\varepsilon]$ et $[k_\varepsilon, k_\varepsilon + 1]$. Cette fonction h_ε est bien continue à support compact (le support est ici l'hypercube fermé $[-k_\varepsilon - 1, k_\varepsilon + 1]^d$). De plus h_ε vérifie l'encadrement $1 \geq h_\varepsilon \geq \mathbf{1}_{B(k_\varepsilon)}$, qui donne après intégration sur $\mathbb{R}^d$ relativement à μ :

$$1 \geq \int_{\mathbb{R}^d} h_\varepsilon \,\mathrm{d}\mu \geq \mu\big(B(k_\varepsilon)\big) > 1 - \varepsilon.$$

Ainsi (9.40) est vérifiée. Notons que la fonction h_ε dépend de ε et de μ.

Soit f continue bornée $\mathbb{R}^d \to \mathbb{R}$ et fixons $\varepsilon > 0$. Par inégalité triangulaire, on a la majoration :

$$\left| \int_{\mathbb{R}^d} f \,\mathrm{d}\mu_n - \int_{\mathbb{R}^d} f \,\mathrm{d}\mu \right| \leq T_1 + T_2 + T_3, \tag{9.41}$$

où

$$T_1 := \left|\int_{\mathbb{R}^d} f(1-h_\varepsilon)\,\mathrm{d}\mu_n\right| \quad T_2 := \left|\int_{\mathbb{R}^d} fh_\varepsilon\,\mathrm{d}\mu_n - \int_{\mathbb{R}^d} fh_\varepsilon\,\mathrm{d}\mu\right| \quad T_3 := \left|\int_{\mathbb{R}^d} f(1-h_\varepsilon)\,\mathrm{d}\mu\right|.$$

D'après (9.40) on a immédiatement (noter la positivité de $1-h_\varepsilon$) :

$$T_3 \leq \int_{\mathbb{R}^d} |f|(1-h_\varepsilon)\,\mathrm{d}\mu \leq \|f\|_\infty \int_{\mathbb{R}^d}(1-h_\varepsilon)\,\mathrm{d}\mu = \|f\|_\infty \left\{1 - \int_{\mathbb{R}^d} h_\varepsilon\,\mathrm{d}\mu\right\} < \|f\|_\infty\,\varepsilon. \quad (9.42)$$

De même avec μ_n au lieu de μ,

$$\forall n \in \mathbb{N}^*, \quad T_1 \leq \|f\|_\infty \left\{1 - \int_{\mathbb{R}^d} h_\varepsilon\,\mathrm{d}\mu_n\right\}. \quad (9.43)$$

La fonction h_ε étant continue à support compact, l'hypothèse iii) nous donne la convergence de $\int_{\mathbb{R}^d} h_\varepsilon\,\mathrm{d}\mu_n$ vers $\int_{\mathbb{R}^d} h_\varepsilon\,\mathrm{d}\mu$, qui avec (9.42) et (9.43) nous assure de l'existence d'un entier $n_1 = n_1(\varepsilon)$ tel que

$$\forall n \geq n_1, \quad T_1 \leq 2\|f\|_\infty\,\varepsilon. \quad (9.44)$$

Enfin, la fonction fh_ε est continue (comme produit de deux fonctions continues) et à support compact (puisque nulle en dehors du support compact de h). L'hypothèse *iii*) appliquée à cette fonction nous dit que T_2 tend vers 0 quand n tend vers l'infini, il existe donc un $n_2 = n_2(\varepsilon)$ tel que :

$$\forall n \geq n_2, \quad T_2 \leq \varepsilon. \quad (9.45)$$

En rassemblant (9.41), (9.42), (9.44) et (9.45), on voit que

$$\forall n \geq n_0 := \max(n_1, n_2), \qquad \left|\int_{\mathbb{R}^d} f\,\mathrm{d}\mu_n - \int_{\mathbb{R}^d} f\,\mathrm{d}\mu\right| < \big(3\|f\|_\infty + 1\big)\varepsilon.$$

Comme ε est arbitraire, on a ainsi prouvé la convergence de $\int_{\mathbb{R}^d} f\,\mathrm{d}\mu_n$ vers $\int_{\mathbb{R}^d} f\,\mathrm{d}\mu$. Ce résultat étant valable pour toute f continue bornée, l'implication *iii*) $\Rightarrow$ *i*) est établie. □

Preuve de i) $\Rightarrow$ *iv*). Fixons un fermé quelconque C de $\mathbb{R}^d$ et définissons la fonction continue $\operatorname{dist}(\,.\,, C)$ par :

$$\forall x \in \mathbb{R}^d, \quad \operatorname{dist}(x, C) := \inf\big\{\|x-y\|;\ y \in C\big\}.$$

Rappelons que puisque C est *fermé*,

$$\operatorname{dist}(x, C) = 0 \Leftrightarrow x \in C. \quad (9.46)$$

Définissons pour tout $k \in \mathbb{N}^*$, la fonction $h_k : \mathbb{R}_+ \to [0,1]$ par

$$h_k(t) := \begin{cases} 1-kt & \text{si } 0 \leq t \leq 1/k, \\ 0 & \text{si } t > 1/k. \end{cases}$$

Cette fonction est continue bornée et de plus

$$\lim_{k\to+\infty} h_k(t) = \begin{cases} 1 & \text{si } t = 0, \\ 0 & \text{si } t > 0. \end{cases} \tag{9.47}$$

En posant

$$\forall x \in \mathbb{R}^d, \quad f_k(x) := h_k\big(\operatorname{dist}(x, C)\big),$$

on construit une fonction continue bornée f_k telle que $\mathbf{1}_C \leq f_k \leq 1$ et compte-tenu de (9.46) et (9.47), f_k converge ponctuellement sur $\mathbb{R}^d$ vers $\mathbf{1}_C$. Cette convergence étant dominée par la constante 1 qui est μ intégrable (μ est une mesure finie),

$$\lim_{k\to+\infty} \int_{\mathbb{R}^d} f_k \,\mathrm{d}\mu = \int_{\mathbb{R}^d} \mathbf{1}_C \,\mathrm{d}\mu = \mu(C). \tag{9.48}$$

D'autre part pour chaque $k \in \mathbb{N}^*$, l'inégalité $\mathbf{1}_C \leq f_k$ nous donne :

$$\forall n \in \mathbb{N}^*, \quad \mu_n(C) \leq \int_{\mathbb{R}^d} f_k \,\mathrm{d}\mu_n. \tag{9.49}$$

Si l'on fait tendre n vers l'infini (k restant fixé), l'intégrale au second membre de (9.49) converge vers $\int_{\mathbb{R}^d} f_k \,\mathrm{d}\mu$ d'après l'hypothèse i). Par contre rien ne nous dit que la suite $\mu_n(C)$ converge. Pour remédier à cet inconvénient, on prend les limites supérieures [7]

$$\limsup_{n\to+\infty} \mu_n(C) \leq \limsup_{n\to+\infty} \int_{\mathbb{R}^d} f_k \,\mathrm{d}\mu_n. \tag{9.50}$$

Grâce à l'hypothèse i) le second membre de (9.50) s'écrit aussi $\lim_{n\to+\infty} \int_{\mathbb{R}^d} f_k \,\mathrm{d}\mu_n = \int_{\mathbb{R}^d} f_k \,\mathrm{d}\mu$, d'où

$$\forall k \in \mathbb{N}^*, \quad \limsup_{n\to+\infty} \mu_n(C) \leq \int_{\mathbb{R}^d} f_k \,\mathrm{d}\mu.$$

Dans cette inégalité le premier membre est une constante, de sorte qu'en faisant tendre k vers l'infini, on obtient grâce à (9.48),

$$\limsup_{n\to+\infty} \mu_n(C) \leq \mu(C).$$

Comme le fermé C était quelconque, l'implication $i) \Rightarrow iv)$ est établie. □

Preuve de iv) $\Leftrightarrow$ *v).* La preuve de cette équivalence est immédiate en remarquant que si G est un ouvert, son complémentaire $C := \mathbb{R}^d \setminus G$ est un fermé (et vice versa), que $\mu_n(G) = 1 - \mu_n(C)$ et en utilisant les égalités [8]

$$\limsup_{n\to+\infty}(1 - u_n) = 1 - \liminf_{n\to+\infty} u_n, \quad \liminf_{n\to+\infty}(1 - v_n) = 1 - \limsup_{n\to+\infty} v_n, \tag{9.51}$$

avec $u_n = \mu_n(G)$ et $v_n = \mu_n(C)$. □

---

7. On pourrait aussi prendre les limites inférieures et dérouler sans inconvénient la fin de la preuve. On aboutirait simplement à l'inégalité $\liminf_n \mu_n(C) \leq \mu(C)$ qui est plus faible que iv).

8. Si g est continue et décroissante $\overline{\mathbb{R}} \to \overline{\mathbb{R}}$ et $E \subset \overline{\mathbb{R}}$, on a $\sup g(E) = g(\inf E)$ et $\inf g(E) = g(\sup E)$ (exercice). En appliquant ceci avec $g(x) = 1 - x$ et en revenant aux définitions de lim inf et lim sup, on obtient (9.51).

Preuve de $\big(iv)$ *et* $v)\big) \Rightarrow vi)$. Soit B un borélien de $\mathbb{R}^d$ tel que $\mu(\partial B) = 0$. Comme $\overline{B}$ est l'union disjointe de son intérieur $\overset{\circ}{B}$ et de sa frontière ∂B, on a $\mu(\overline{B}) = \mu(\overset{\circ}{B}) + \mu(\partial B) = \mu(\overset{\circ}{B})$. En raison de l'inclusion $\overset{\circ}{B} \subset B \subset \overline{B}$, on en déduit :

$$\mu(\overset{\circ}{B}) = \mu(B) = \mu(\overline{B}). \tag{9.52}$$

L'inclusion $B \subset \overline{B}$ et l'hypothèse $iv)$ nous donnent

$$\limsup_{n\to+\infty} \mu_n(B) \le \limsup_{n\to+\infty} \mu_n(\overline{B}) \le \mu(\overline{B}), \tag{9.53}$$

tandis que $v)$ et l'inclusion $\overset{\circ}{B} \subset B$ nous donnent

$$\mu(\overset{\circ}{B}) \le \liminf_{n\to+\infty} \mu_n(\overset{\circ}{B}) \le \liminf_{n\to+\infty} \mu_n(B). \tag{9.54}$$

En combinant (9.52), (9.53) et (9.54), on en déduit

$$\mu(B) \le \liminf_{n\to+\infty} \mu_n(B) \le \limsup_{n\to+\infty} \mu_n(B) \le \mu(B),$$

d'où

$$\liminf_{n\to+\infty} \mu_n(B) = \limsup_{n\to+\infty} \mu_n(B) = \mu(B),$$

ce qui signifie que $\mu_n(B)$ converge vers $\mu(B)$. □

Preuve de $vi) \Rightarrow i)$. Soit $f : \mathbb{R}^d \to \mathbb{R}$ une fonction continue bornée. Nous allons construire pour chaque $\varepsilon > 0$, une fonction étagée $g = \sum_{k=1}^{l} c_k \mathbf{1}_{B_k}$ telle que *pour toute probabilité* ν *sur* $\big(\mathbb{R}^d, \mathrm{Bor}(\mathbb{R}^d)\big)$, $I(\nu) := \left|\int_{\mathbb{R}^d} (g - f)\,\mathrm{d}\nu\right| < 2\varepsilon$. Cette majoration uniforme relativement à la mesure permet de ramener le contrôle de la différence $|\int_{\mathbb{R}^d} f\,\mathrm{d}\mu_n - \int_{\mathbb{R}^d} f\,\mathrm{d}\mu|$ à celui de $|\int_{\mathbb{R}^d} g\,\mathrm{d}\mu_n - \int_{\mathbb{R}^d} g\,\mathrm{d}\mu|$. Cette différence d'intégrales s'écrit comme une combinaison linéaire finie des $\mu_n(B_k) - \mu(B_k)$ et on la majore par ε pour n assez grand grâce à l'hypothèse $vi)$, pour peu que l'on ait réussi à choisir les B_k tels que $\mu(\partial B_k) = 0$.

Pour réaliser ce programme, commençons par fixer $\varepsilon > 0$. Comme f est bornée, $f(\mathbb{R}^d)$ est une partie bornée de $\mathbb{R}$, on peut donc définir $a = a(\varepsilon)$ et $b = b(\varepsilon)$ par :

$$a := \max\Big\{k\varepsilon;\ k \in \mathbb{Z},\ (k+1)\varepsilon \le \inf f\Big\}, \quad b := \min\Big\{k\varepsilon;\ k \in \mathbb{Z},\ (k-1)\varepsilon \ge \sup f\Big\}.$$

Ainsi $[a, b]$ déborde de chaque côté de l'intervalle $[\inf f, \sup f]$ d'au moins ε. L'application continue f étant borélienne, la mesure image $\mu \circ f^{-1}$ est bien définie. C'est une mesure finie (et même une probabilité) sur $\big(\mathbb{R}, \mathrm{Bor}(\mathbb{R})\big)$. Notons F sa fonction de répartition. On sait que F est discontinue au point $y \in \mathbb{R}$ si et seulement si $(\mu \circ f^{-1})(\{y\}) \neq 0$ et que l'ensemble D de ces points de discontinuité de F est au plus dénombrable. Ainsi pour tout intervalle non vide J, $D \cap J$ est au plus dénombrable alors que J est infini non dénombrable. Par conséquent J contient au moins un point de D^c, autrement dit un point c tel que $(\mu\circ f^{-1})(\{c\})=0$. En appliquant cet argument à chacun des intervalles $[k\varepsilon,(k+1)\varepsilon[$ pour $a\le k\varepsilon<b$, on obtient une suite finie $(c_k)_{0\le k\le l}$ vérifiant :

$$\forall k=0,\dots,l,\quad (\mu\circ f^{-1})(\{c_k\})=0. \tag{9.55}$$

La construction de cette suite nous montre de plus que $c_0<\inf f$, $c_l>\sup f$ et que puisque $c_k\in[k\varepsilon,(k+1)\varepsilon[$,

$$\forall k=1,\dots,l,\quad 0<c_k-c_{k-1}<2\varepsilon. \tag{9.56}$$

Posons maintenant

$$g:=\sum_{k=1}^{l}c_k\mathbf{1}_{B_k},\quad \text{où}\quad B_k:=f^{-1}\Big(]c_{k-1},c_k]\Big).$$

Par construction les B_k sont deux à deux disjoints et comme $c_0<\inf f\le\sup f<c_l$, $]c_0,c_l]$ contient $f(\mathbb{R}^d)$ et par conséquent l'union des B_k est $\mathbb{R}^d$. On en déduit que pour toute mesure de probabilité ν sur $\mathbb{R}^d$,

$$I(\nu):=\left|\int_{\mathbb{R}^d}(g-f)\,\mathrm{d}\nu\right|=\sum_{k=1}^{l}\int_{B_k}(c_k-f)\,\mathrm{d}\nu\le 2\varepsilon\sum_{k=1}^{l}\nu(B_k)=2\varepsilon\nu(\mathbb{R}^d)=2\varepsilon, \tag{9.57}$$

en notant que si $x\in B_k$, $0\le c_k-f(x)\le c_k-c_{k-1}<2\varepsilon$, d'après (9.56).

L'inégalité triangulaire et (9.57) appliquée à $\nu=\mu_n$ puis à $\nu=\mu$, donnent

$$\begin{aligned}\left|\int_{\mathbb{R}^d}f\,\mathrm{d}\mu_n-\int_{\mathbb{R}^d}f\,\mathrm{d}\mu\right| &\le I(\mu_n)+\left|\int_{\mathbb{R}^d}g\,\mathrm{d}\mu_n-\int_{\mathbb{R}^d}g\,\mathrm{d}\mu\right|+I(\mu)\\ &\le 4\varepsilon+\sum_{k=1}^{l}|c_k||\mu_n(B_k)-\mu(B_k)|.\end{aligned} \tag{9.58}$$

Vérifions que $\mu(\partial B_k)=0$ pour chaque k. Par continuité de f,

– $G_k:=f^{-1}\Big(]c_{k-1},c_k[\Big)$ est un *ouvert*, inclus dans B_k, donc $G_k\subset\overset{\circ}{B_k}$;
– $C_k=f^{-1}\Big([c_{k-1},c_k]\Big)$ est un *fermé*, contenant B_k, donc $\overline{B_k}\subset C_k$.

Alors

$$\partial B_k=\overline{B_k}\setminus\overset{\circ}{B_k}\subset C_k\setminus G_k=\{x\in\mathbb{R}^d;\ f(x)=c_{k-1}\text{ ou }f(x)=c_k\},$$

d'où grâce à (9.55),

$$\mu(\partial B_k)=(\mu\circ f^{-1})(\{c_{k-1}\})+(\mu\circ f^{-1})(\{c_k\})=0.$$

On voit maintenant en appliquant l'hypothèse *vi*) que le sigma dans (9.58) est la somme d'un nombre fini (l dépendant de ε) de termes tendant chacun vers 0 quand n tend vers $+\infty$. On peut ainsi trouver un $n_0=n_0(\varepsilon)$ tel que ce sigma soit majoré par ε à partir du rang n_0. Ainsi

$$\exists n_0=n_0(\varepsilon),\ \forall n\ge n_0,\quad \left|\int_{\mathbb{R}^d}f\,\mathrm{d}\mu_n-\int_{\mathbb{R}^d}f\,\mathrm{d}\mu\right|<5\varepsilon.$$

Ceci étant vrai pour tout $\varepsilon>0$, $\int_{\mathbb{R}^d}f\,\mathrm{d}\mu_n$ converge vers $\int_{\mathbb{R}^d}f\,\mathrm{d}\mu$. Comme la fonction continue bornée f était quelconque, *i*) est vérifiée. □

Pour achever la preuve du théorème, il ne nous reste plus qu'à établir la caractérisation de la convergence étroite par les fonctions de répartition dans le cas $d = 1$. Exploitant l'équivalence déjà établie des conditions i) à vi), il nous suffit pour cela de vérifier que :

$$vi) \Rightarrow vii) \Rightarrow iii).$$

Preuve de $vi) \Rightarrow vii)$. En notant F_n et F les fonctions de répartition respectives de μ_n et μ, on a :

$$\forall x \in \mathbb{R}, \quad F_n(x) = \mu_n(]-\infty, x]), \quad F(x) = \mu(]-\infty, x]).$$

Supposons maintenant que x est point de continuité de F. On sait qu'alors $\mu(\{x\}) = 0$. Or $\{x\}$ est exactement la frontière du borélien $B :=]-\infty, x]$, donc par vi), $\mu_n(B)$ converge vers $\mu(B)$, autrement dit $F_n(x)$ converge vers $F(x)$. $\square$

Preuve de $vii) \Rightarrow iii)$. Soit $f : \mathbb{R} \to \mathbb{R}$, continue à support compact. Alors f est nulle en dehors d'un intervalle compact $[a, b]$ de $\mathbb{R}$ et sur cet intervalle, elle est *uniformément* continue :

$$\forall \varepsilon > 0,\ \exists \delta > 0,\ \forall x, y \in [a, b], \quad |x - y| \leq \delta \Rightarrow |f(x) - f(y)| < \varepsilon. \tag{9.59}$$

Fixons $\varepsilon > 0$ et un δ correspondant donné par (9.59). Découpons $[a, b]$ en intervalles consécutifs de longueur $\delta/2$ en partant de a (le dernier aura une longueur au plus $\delta/2$). L'ensemble des points de discontinuité de F étant au plus dénombrable, on peut trouver dans chacun de ces intervalles un point de continuité de F. On obtient ainsi une suite finie $x_1, \dots, x_l$ de points de continuité de F. Par construction, la distance entre x_{k-1} et x_k est au plus δ, d'où grâce à (9.59),

$$\forall k = 1, \dots, l,\ \forall x \in [x_{k-1}, x_k], \quad |f(x) - f(x_k)| < \varepsilon. \tag{9.60}$$

Considérons la fonction étagée

$$g := \sum_{k=1}^{l} f(x_k) \mathbf{1}_{]x_{k-1}, x_k]},$$

qui comme f, est nulle hors de $[a, b]$. Pour toute mesure de probabilité ν sur $\mathbb{R}$, on a

$$I(\nu) := \left| \int_{\mathbb{R}} (g - f) \,\mathrm{d}\nu \right| \tag{9.61}$$

$$\leq \int_{[a,x_0]} |f| \,\mathrm{d}\nu + \sum_{k=1}^{l} \int_{]x_{k-1},x_k]} \big| f(x) - f(x_k) \big| \,\mathrm{d}\nu(x) + \int_{]x_l,b]} |f| \,\mathrm{d}\nu. \tag{9.62}$$

Comme $f(a) = 0$ et $|x_0 - a| \leq \delta$, $|f(x)| = |f(x) - f(a)|$ est majoré par ε sur $[a, x_0]$, d'où

$$\int_{[a,x_0]} |f| \,\mathrm{d}\nu \leq \varepsilon \nu([a, x_0]).$$

Par le même argument avec b à la place de a,

$$\int_{]x_l,b]} |f|\,\mathrm{d}\nu \leq \varepsilon\nu(]x_l,b]).$$

De plus par (9.60),

$$\forall k = 1,\dots,l,\ \int_{]x_{k-1},x_k]} \big|f(x) - f(x_k)\big|\,\mathrm{d}\nu(x) \leq \varepsilon\nu(]x_{k-1},x_k]).$$

En reportant ces majorations dans (9.62), il vient :

$$I(\nu) \leq \varepsilon\nu([a,x_0]) + \varepsilon\sum_{k=1}^{l}\nu(]x_{k-1},x_k]) + \varepsilon\nu(]x_l,b]) = \varepsilon\nu([a,b]) \leq \varepsilon. \tag{9.63}$$

Par inégalité triangulaire et (9.63) appliquée avec $\nu = \mu_n$, puis $\nu = \mu$, on obtient

$$\begin{aligned}\left|\int_{\mathbb{R}} f\,\mathrm{d}\mu_n - \int_{\mathbb{R}} f\,\mathrm{d}\mu\right| &\leq I(\mu_n) + \left|\int_{\mathbb{R}} g\,\mathrm{d}\mu_n - \int_{\mathbb{R}} g\,\mathrm{d}\mu\right| + I(\mu)\\ &\leq 2\varepsilon + \sum_{k=1}^{l}|f(x_k)|\Big|\mu_n(]x_{k-1},x_k]) - \mu(]x_{k-1},x_k])\Big|. \end{aligned} \tag{9.64}$$

Remarquons maintenant que

$$\begin{aligned}\mu_n(]x_{k-1},x_k]) - \mu(]x_{k-1},x_k]) &= F_n(x_k) - F_n(x_{k-1}) - F(x_k) + F(x_{k-1})\\ &= \big(F_n(x_k) - F(x_k)\big) + \big(F(x_{k-1}) - F_n(x_{k-1})\big)\end{aligned}$$

tend vers 0 quand n tend vers $+\infty$, en raison de l'hypothèse *vii*) puisque les x_k sont des points de continuité de F. En notant que pour l'instant, ε est fixé et les x_k et l dépendent de ε et de F, mais pas de n, on en déduit l'existence d'un $n_0 = n_0(\varepsilon, F, f)$ tel que pour tout $n \geq n_0$, le sigma dans (9.64) est strictement inférieur à ε. Ainsi

$$\forall n \geq n_0 = n_0(\varepsilon, F, f),\quad \left|\int_{\mathbb{R}} f\,\mathrm{d}\mu_n - \int_{\mathbb{R}} f\,\mathrm{d}\mu\right| < 3\varepsilon.$$

Comme ε était arbitraire, ceci établit la convergence de $\int_{\mathbb{R}} f\,\mathrm{d}\mu_n$ vers $\int_{\mathbb{R}} f\,\mathrm{d}\mu$. La fonction continue à support compact f étant quelconque, on obtient finalement *iii*). □

Le théorème 9.32 est maintenant complètement démontré. □

Voici un exemple où la caractérisation de la convergence étroite par la convergence des f.d.r. au sens de *vii*) est un outil bien adapté.

Exemple 9.7 (Une loi limite d'extrêmes)**.** Soit $(X_k)_{k\geq 1}$ une suite de variables aléatoires indépendantes et de même loi avec fonction de répartition commune F. Définissons la suite de variables aléatoires (M_n) par :

$$M_n := \max_{1\leq k\leq n} X_k \qquad n \in \mathbb{N}^*. \tag{9.65}$$

Connaissant F, il est facile d'obtenir la fonction de répartition G_n de M_n :

$$G_n(x) = \mathbf{P}(M_n \leq x) = \mathbf{P}\big(\forall k \in \{1, \dots, n\},\ X_k \leq x\big) = \mathbf{P}\Big(\bigcap_{k=1}^{n} \{X_k \leq x\}\Big).$$

En utilisant l'indépendance des X_k, puis le fait qu'elles ont même loi, on en déduit :

$$G_n(x) = \prod_{k=1}^{n} \mathbf{P}(X_k \leq x) = \big(F(x)\big)^n. \tag{9.66}$$

Supposons désormais que les X_k ont pour loi commune la loi exponentielle de paramètre α. On a alors :

$$F(x) = 1 - \mathrm{e}^{-\alpha x} \text{ si } x \geq 0, \quad F(x) = 0 \text{ si } x < 0.$$

$$G_n(x) = \big(1 - \mathrm{e}^{-\alpha x}\big)^n \text{ si } x \geq 0, \quad G_n(x) = 0 \text{ si } x < 0.$$

Donc pour x réel fixé, on a $\lim_{n\to+\infty} G_n(x) = 0$. La signification intuitive de ce résultat est que M_n aura tendance à être grand pour les grandes valeurs de n. Afin de préciser cette idée, on essaie de normaliser M_n pour contrôler sa croissance aléatoire vers $+\infty$. Pour cela, on examine le comportement asymptotique de $\mathbf{P}(M_n - \alpha^{-1} \ln n \leq x)$:

$$\mathbf{P}\left(M_n - \frac{\ln n}{\alpha} \leq x\right) = G_n\left(x + \frac{\ln n}{\alpha}\right) = \big(1 - \mathrm{e}^{-\alpha x - \ln n}\big)^n = \left(1 - \frac{\mathrm{e}^{-\alpha x}}{n}\right)^n. \tag{9.67}$$

On en déduit que :

$$\lim_{n\to+\infty} \mathbf{P}\left(M_n - \frac{\ln n}{\alpha} \leq x\right) = \exp\big(-\mathrm{e}^{-\alpha x}\big) \tag{9.68}$$

Le calcul (9.67) est valable pour $\ln n \geq -\alpha x$, donc pour tout $n \in \mathbb{N}^*$ et tout $x \geq 0$. Pour $x < 0$ fixé, on aura $\ln n \geq -\alpha x$ pour $n \geq n_0(x)$ donc (9.68) est valable pour tout x réel. On peut donc dire qu'*asymptotiquement*, M_n est de l'ordre de grandeur de $\alpha^{-1} \ln n$ et que la dispersion aléatoire de M_n autour de cette valeur est donnée par la loi de fonction de répartition :

$$H(x) = \exp\big(-\mathrm{e}^{-\alpha x}\big) \qquad x \in \mathbb{R}. \tag{9.69}$$

Il s'agit d'une loi de Gumbel. D'après le théorème 9.32 *vii*) et (9.68), on peut reformuler la conclusion en disant que la suite de variables aléatoires $M_n - \alpha^{-1} \ln n$ converge en loi vers une v.a. suivant la loi de Gumbel définie par sa f.d.r. (9.69).

Théorème 9.33. *La convergence en probabilité implique la convergence en loi : si X_n ($n \geq 1$) et X sont des variables aléatoires réelles définies sur le même espace probabilisé $(\Omega, \mathcal{F}, \mathbf{P})$ telles que X_n converge en probabilité vers X, alors X_n converge aussi en loi vers X.*

Preuve. La convergence en probabilité de X_n vers X n'a de sens que si les X_n et X sont définis sur le même espace probabilisé $(\Omega, \mathcal{F}, \mathbf{P})$. Grâce au théorème 9.32, il suffit de montrer que $\int_{\mathbb{R}} f \, dP_{X_n}$ converge vers $\int_{\mathbb{R}} f \, dP_X$ pour toute fonction f uniformément continue bornée $\mathbb{R} \to \mathbb{R}$. La continuité uniforme se traduit par

$$\forall \varepsilon > 0, \ \exists \delta > 0, \quad |x - y| < \delta \Rightarrow |f(x) - f(y)| < \varepsilon.$$

On en déduit en prenant $x = X_n(\omega)$ et $y = X(\omega)$ que pour tout $\omega \in \{|X_n - X| < \delta\}$, $\left|f\bigl(X_n(\omega)\bigr) - f\bigl(X_n(\omega)\bigr)\right| < \varepsilon$. Par transfert, on a

$$\left|\int_{\mathbb{R}} f \, dP_{X_n} - \int_{\mathbb{R}} f \, dP_X\right| = \left|\int_{\Omega} \bigl(f(X_n) - f(X)\bigr) \, d\mathbf{P}\right| \leq \int_{\Omega} \bigl|f(X_n) - f(X)\bigr| \, d\mathbf{P}. \tag{9.70}$$

Découpant cette dernière intégrale suivant la partition de Ω en $\{|X_n - X| < \delta\}$ et son complémentaire, on obtient

$$\begin{aligned} \int_{\Omega} \bigl|f(X_n) - f(X)\bigr| \, d\mathbf{P} &\leq \int_{\{|X_n - X| < \delta\}} \varepsilon \, d\mathbf{P} + \int_{\{|X_n - X| \geq \delta\}} 2\|f\|_\infty \, d\mathbf{P} \\ &\leq \varepsilon + 2\|f\|_\infty \mathbf{P}\bigl(|X_n - X| \geq \delta\bigr). \end{aligned} \tag{9.71}$$

Cette majoration est valable pour tout n. En exploitant l'hypothèse de convergence en probabilité de X_n vers X, on peut trouver n_0 (dépendant de δ, donc de ε et de f) tel que pour tout $n \geq n_0$, le second terme dans (9.71) soit majoré par ε. En reportant dans (9.70), on en déduit que $\left|\int_{\mathbb{R}} f \, dP_{X_n} - \int_{\mathbb{R}} f \, dP_X\right| < 2\varepsilon$ pour tout $n \geq n_0(\varepsilon, f)$. Ceci établit la convergence de $\int_{\mathbb{R}} f \, dP_{X_n}$ vers $\int_{\mathbb{R}} f \, dP_X$. Ce résultat étant valable pour toute f uniformément continue bornée, P_{X_n} converge étroitement vers P_X par le théorème 9.32 *ii*), autrement dit, X_n converge en loi vers X. □

9.3.3 Convergence en loi et fonctions caractéristiques

Théorème 9.34. *Soient X_n ($n \in \mathbb{N}^*$) et X des vecteurs aléatoires de $\mathbb{R}^d$ de fonctions caractéristiques respectives φ_{X_n} et φ_X. Alors X_n converge en loi vers X si et seulement si*

$$\forall t \in \mathbb{R}^d, \quad \varphi_{X_n}(t) \xrightarrow[n \to +\infty]{} \varphi_X(t). \tag{9.72}$$

Preuve. On vérifie facilement la nécessité de (9.72) pour la convergence en loi de X_n vers X. En effet pour t quelconque fixé, les fonctions $x \mapsto \cos(\langle t, x \rangle)$ et $x \mapsto \sin(\langle t, x \rangle)$ sont continues et bornées sur $\mathbb{R}^d$. La convergence en loi de X_n vers X implique alors celle de $\mathbf{E} \cos(\langle t, X_n \rangle)$ vers $\mathbf{E} \cos(\langle t, X \rangle)$ et de $\mathbf{E} \sin(\langle t, X_n \rangle)$ vers $\mathbf{E} \sin(\langle t, X \rangle)$. Par linéarité de l'espérance, on en déduit la convergence $\mathbf{E} \exp(i\langle t, X_n \rangle) = \varphi_{X_n}(t)$ vers $\mathbf{E} \exp(i\langle t, X \rangle) = \varphi_X(t)$. Comme t était quelconque, (9.72) est vérifiée.

La réciproque demande un peu plus de travail. Notons pour alléger μ_n la loi de X_n et μ celle de X et supposons (9.72) vérifiée, *i.e.* que $\widehat{\mu}_n$ converge ponctuellement vers $\widehat{\mu}$ sur $\mathbb{R}^d$. Nous allons montrer la convergence étroite de μ_n vers μ en vérifiant que pour toute fonction f continue à support compact sur $\mathbb{R}^d$, $\int_{\mathbb{R}^d} f \, d\mu_n$ converge vers $\int_{\mathbb{R}^d} f \, d\mu$ (cf. théorème 9.32 *iii*)).

Suivant la même méthode que dans la preuve du théorème 9.9, introduisons dans le problème la loi gaussienne $\gamma_k := \mathfrak{N}(0, k^{-1/2})^{\otimes d}$. D'après la formule (9.23) établie dans la preuve du théorème 9.9 on a (en y remplaçant γ_n par γ_k) pour toute mesure de probabilité ν sur $(\mathbb{R}^d, \mathrm{Bor}(\mathbb{R}^d))$:

$$\forall f \in \mathcal{C}_c(\mathbb{R}^d), \quad \int_{\mathbb{R}^d} f \, \mathrm{d}(\nu * \gamma_k) = c_k \int_{\mathbb{R}^d} f(y) \int_{\mathbb{R}^d} \mathrm{e}^{ikyu} \widehat{\nu}(-ku) \, \mathrm{d}\gamma_k(u) \, \mathrm{d}\lambda_d(y). \tag{9.73}$$

Par inégalité triangulaire,

$$\left| \int_{\mathbb{R}^d} f \, \mathrm{d}\mu_n - \int_{\mathbb{R}^d} f \, \mathrm{d}\mu \right| \leq I_k(\mu_n) + \left| \int_{\mathbb{R}^d} f \, \mathrm{d}(\mu_n * \gamma_k) - \int_{\mathbb{R}^d} f \, \mathrm{d}(\mu * \gamma_k) \right| + I_k(\mu), \tag{9.74}$$

où l'on a posé pour toute mesure de probabilité ν sur $\mathbb{R}^d$:

$$I_k(\nu) := \left| \int_{\mathbb{R}^d} f \, \mathrm{d}\nu - \int_{\mathbb{R}^d} f \, \mathrm{d}(\nu * \gamma_k) \right|.$$

Nous aurons besoin d'une majoration de $I_k(\nu)$, uniforme par rapport à la mesure ν. Voici comment l'obtenir. La fonction f étant continue à support compact, elle est uniformément continue sur $\mathbb{R}^d$:

$$\forall \varepsilon > 0, \ \exists \delta > 0, \ \forall x \in \mathbb{R}^d, \ \forall y \in]-\delta, \delta[^d, \quad |f(x+y) - f(x)| < \varepsilon. \tag{9.75}$$

La mesure γ_k étant une probabilité sur $\mathbb{R}^d$, la constante $\int_{\mathbb{R}^d} f \, \mathrm{d}\nu$ est γ_k intégrable sur $\mathbb{R}^d$ et on a

$$\int_{\mathbb{R}^d} f \, \mathrm{d}\nu = \int_{\mathbb{R}^d} \left\{ \int_{\mathbb{R}^d} f(x) \, \mathrm{d}\nu(x) \right\} \mathrm{d}\gamma_k(y) = \int_{\mathbb{R}^d \times \mathbb{R}^d} f(x) \, \mathrm{d}(\nu \otimes \gamma_k)(x, y).$$

En écrivant $\int_{\mathbb{R}^d} f \, \mathrm{d}(\nu * \gamma_k)$ à l'aide de $\nu \otimes \gamma_k$, on en déduit :

$$\begin{aligned} I_k(\nu) &\leq \int_{\mathbb{R}^d \times \mathbb{R}^d} |f(x+y) - f(x)| \, \mathrm{d}(\nu \otimes \gamma_k)(x, y) \\ &= \int_{\mathbb{R}^d} \left\{ \int_{\mathbb{R}^d} |f(x+y) - f(x)| \, \mathrm{d}\gamma_k(y) \right\} \mathrm{d}\nu(x). \end{aligned} \tag{9.76}$$

Dans l'intégrale relative à γ_k, découpons $\mathbb{R}^d$ en $]-\delta, \delta[^d$ et son complémentaire, majorons $|f(x+y) - f(x)|$ par ε sur $]-\delta, \delta[^d$ (grâce à (9.75)) et par $2\|f\|_\infty$ sur son complémentaire. Le majorant ainsi obtenu pour l'intégrale est une constante indépendante de x. En reportant cette constante dans (9.76) et en intégrant sur $\mathbb{R}^d$ par rapport à la mesure de probabilité ν, on obtient finalement :

$$I_k(\nu) \leq \varepsilon + 2\|f\|_\infty \gamma_k\Big(\mathbb{R}^d \setminus]-\delta, \delta[^d\Big), \tag{9.77}$$

où δ est choisi en fonction de ε, conformément à (9.75). C'est bien la majoration uniforme en ν annoncée.

Fixons maintenant $\varepsilon > 0$ et le δ correspondant. Nous avons déjà vu, cf. (9.17), que $\gamma_k\Big(\mathbb{R}^d \setminus]-\delta, \delta[^d\Big)$ tend vers 0 quand k tend vers l'infini. On peut donc trouver un

$k_0 = k_0(\varepsilon, f)$ tel que $I_{k_0}(\nu) < 2\varepsilon$ et ceci *quelle que soit la mesure de probabilité* ν sur $\mathbb{R}^d$. Ceci vaut en particulier pour $\nu = \mu_n$ (pour tout n) et $\nu = \mu$. Par conséquent en choisissant $k = k_0$ dans (9.74), on a :

$$\forall n \in \mathbb{N}^*, \quad \left| \int_{\mathbb{R}^d} f \, \mathrm{d}\mu_n - \int_{\mathbb{R}^d} f \, \mathrm{d}\mu \right| \leq \left| \int_{\mathbb{R}^d} f \, \mathrm{d}(\mu_n * \gamma_{k_0}) - \int_{\mathbb{R}^d} f \, \mathrm{d}(\mu * \gamma_{k_0}) \right| + 4\varepsilon. \tag{9.78}$$

Admettons provisoirement que pour toute $f \in \mathcal{C}_c(\mathbb{R}^d)$,

$$\lim_{n \to +\infty} \int_{\mathbb{R}^d} f \, \mathrm{d}(\mu_n * \gamma_{k_0}) = \int_{\mathbb{R}^d} f \, \mathrm{d}(\mu * \gamma_{k_0}). \tag{9.79}$$

Alors on peut trouver un $n_0 = n_0(\varepsilon, f)$ tel que :

$$\forall n \geq n_0(\varepsilon), \quad \left| \int_{\mathbb{R}^d} f \, \mathrm{d}(\mu_n * \gamma_{k_0}) - \int_{\mathbb{R}^d} f \, \mathrm{d}(\mu * \gamma_{k_0}) \right| < \varepsilon.$$

En reportant dans (9.78), on obtient

$$\forall n \geq n_0(\varepsilon), \quad \left| \int_{\mathbb{R}^d} f \, \mathrm{d}\mu_n - \int_{\mathbb{R}^d} f \, \mathrm{d}\mu \right| < 5\varepsilon.$$

Comme ε était quelconque, ceci établit la convergence de $\int_{\mathbb{R}^d} f \, \mathrm{d}\mu_n$ vers $\int_{\mathbb{R}^d} f \, \mathrm{d}\mu$. Ce résultat étant valable pour toute $f \in \mathcal{C}_c(\mathbb{R}^d)$, la condition *iii*) du théorème 9.32 est vérifiée, donc μ_n converge étroitement vers μ.

Pour compléter la preuve, il ne reste plus qu'à justifier la convergence (9.79). En posant

$$h_n(y) := c_{k_0} f(y) \int_{\mathbb{R}^d} \mathrm{e}^{ik_0 yu} \widehat{\mu}_n(-k_0 u) \, \mathrm{d}\gamma_{k_0}(u), \quad h(y) := c_{k_0} f(y) \int_{\mathbb{R}^d} \mathrm{e}^{ik_0 yu} \widehat{\mu}(-k_0 u) \, \mathrm{d}\gamma_{k_0}(u),$$

la formule (9.73) nous permet de réécrire la convergence (9.79) sous la forme

$$\int_{\mathbb{R}^d} f \, \mathrm{d}(\mu_n * \gamma_{k_0}) = \int_{\mathbb{R}^d} h_n(y) \, \mathrm{d}\lambda_d(y) \xrightarrow[n \to +\infty]{} \int_{\mathbb{R}^d} h(y) \, \mathrm{d}\lambda_d(y) = \int_{\mathbb{R}^d} f \, \mathrm{d}(\mu * \gamma_{k_0}). \tag{9.80}$$

Par hypothèse, $\widehat{\mu}_n$ converge ponctuellement sur $\mathbb{R}^d$ vers $\widehat{\mu}$. On en déduit que

$$\forall y \in \mathbb{R}^d, \ \forall u \in \mathbb{R}^d, \quad \mathrm{e}^{ik_0 yu} \widehat{\mu}_n(-k_0 u) \xrightarrow[n \to +\infty]{} \mathrm{e}^{ik_0 yu} \widehat{\mu}(-k_0 u).$$

Pour chaque $y \in \mathbb{R}^d$, cette convergence est dominée relativement à n par la fonction constante 1 (considérée ici comme fonction de u) qui est γ_{k_0} intégrable sur $\mathbb{R}^d$ puisque γ_{k_0} est une probabilité. Par une première application du théorème de convergence dominée, relativement à la mesure γ_{k_0}, on en déduit :

$$\forall y \in \mathbb{R}^d, \quad h_n(y) \xrightarrow[n \to +\infty]{} h(y).$$

À son tour, cette convergence est dominée, en effet :

$$\forall n \in \mathbb{N}^*, \ \forall y \in \mathbb{R}^d, \quad |h_n(y)| \leq c_{k_0} |f(y)| \int_{\mathbb{R}^d} \left| \mathrm{e}^{ik_0 yu} \widehat{\mu}_n(-k_0 u) \right| \mathrm{d}\gamma_{k_0}(u) \leq c_{k_0} |f(y)|$$

et la fonction $c_{k_0}|f|$ indépendante de n est λ_d intégrable sur $\mathbb{R}^d$ parce que f est continue à support *compact*. Une deuxième application du théorème de convergence dominée, cette fois relativement à λ_d, nous donne finalement la convergence de $\int_{\mathbb{R}^d} h_n \, \mathrm{d}\lambda_d$ vers $\int_{\mathbb{R}^d} h \, \mathrm{d}\lambda_d$, ce qui établit (9.80) et achève la preuve du théorème. □

9.4 Théorème limite central

Théorème 9.35. *Soit* $(X_k)_{k\geq 1}$ *une suite de variables aléatoires définies sur le même espace probabilisé* $(\Omega, \mathcal{F}, \mathbf{P})$*, indépendantes, de même loi et de carré intégrable (et non* p.s. *constantes). Notons* $m := \mathbf{E}X_1$*,* $\sigma^2 := \operatorname{Var} X_1$ *avec* $\sigma > 0$*. Alors*

$$S_n^* := \frac{S_n - \mathbf{E}S_n}{\sqrt{\operatorname{Var} S_n}} = \frac{S_n - nm}{\sigma\sqrt{n}} \xrightarrow[n\to+\infty]{\text{loi}} \mathfrak{N}(0,1).$$

Preuve. On se ramène à des variables Y_k centrées et de variance 1 en écrivant

$$S_n^* = \frac{1}{\sqrt{n}} \sum_{k=1}^{n} Y_k \quad \text{où} \quad Y_k := \frac{X_k - m}{\sigma}.$$

Les Y_k sont indépendantes et de même loi. Notons ϕ la fonction caractéristique de Y_1 (et donc aussi de chaque Y_k) et φ_n celle de S_n^*. La fonction caractéristique de chacune des $n^{-1/2}Y_k$ est $t \mapsto \phi(n^{-1/2}t)$. Par indépendance et équidistribution de ces variables, on en déduit :

$$\forall t \in \mathbb{R}, \quad \varphi_n(t) = \Big\{\phi\Big(n^{-1/2}t\Big)\Big\}^n. \tag{9.81}$$

Comme Y_1 est de carré intégrable, ϕ est deux fois dérivable (cf. proposition 9.3), d'où

$$\forall u \in \mathbb{R}, \quad \phi(u) = \phi(0) + \phi'(0)u + \phi''(0)\frac{u^2}{2} + u^2\varepsilon(u)$$

avec $\varepsilon(u)$ tendant vers 0 quand u tend vers 0. La propositon 9.3 nous dit de plus que $\phi'(0) = i\mathbf{E}Y_1 = 0$ et $\phi''(0) = -\mathbf{E}Y_1^2 = -1$. Prenons $u = n^{-1/2}t$, il vient

$$\forall n \geq 1, \quad \phi\Big(n^{-1/2}t\Big) = 1 - \frac{t^2}{2n} + \frac{t^2}{n}\varepsilon\Big(n^{-1/2}t\Big).$$

Pour t fixé, on a par le lemme 9.36 b) ci-dessous :

$$\lim_{n\to+\infty} \left(1 - \frac{t^2}{2n} + \frac{t^2}{n}\varepsilon\Big(n^{-1/2}t\Big)\right)^n = \exp\left(\frac{-t^2}{2}\right). \tag{9.82}$$

Ce résultat étant valable pour tout t réel, on vient de montrer (compte-tenu de (9.81)) que φ_n converge ponctuellement sur $\mathbb{R}$ vers la fonction caractéristique de la loi $\mathfrak{N}(0,1)$. On conclut avec le théorème 9.34 que S_n^* converge en loi vers $\mathfrak{N}(0,1)$. □

Le lemme suivant permet d'éviter le recours au logarithme complexe pour justifier (9.82).

Lemme 9.36.

a) Pour tout $n \geq 1$*, si* $z_1, \dots, z_n$ *et* $z'_1, \dots, z'_n$ *sont des nombres complexes de module au plus un, on a*

$$|z_1 \dots z_n - z'_1 \dots z'_n| \leq \sum_{k=1}^{n} |z_k - z'_k|. \tag{9.83}$$

b) *Pour tout* $t \in \mathbb{R}$, *si* $\varepsilon(u)$ *est une fonction à valeurs complexes tendant vers* 0 *quand* u *tend vers* 0,

$$\left(1 - \frac{t^2}{2n} + \frac{t^2}{n}\varepsilon\left(n^{-1/2}t\right)\right)^n \xrightarrow[n\to+\infty]{} \exp(-t^2/2). \tag{9.84}$$

Preuve. Le a) se vérifie par récurrence. Le cas $n = 1$ est trivial. Supposons (9.83) vraie au rang n. Posons $p_n := z_1 \dots z_n$ et $p'_n := z'_1 \dots z'_n$. On peut alors écrire (justifications *a posteriori*)

$$\begin{aligned} |p_n z_{n+1} - p'_n z'_{n+1}| &= |(p_n - p'_n)z_{n+1} + p'_n(z_{n+1} - z'_{n+1})| \\ &\leq |p_n - p'_n||z_{n+1}| + |p'_n||z_{n+1} - z'_{n+1}| \quad (9.85) \\ &\leq |p_n - p'_n| + |z_{n+1} - z'_{n+1}| \quad (9.86) \\ &\leq \sum_{k=1}^{n} |z_k - z'_k| + |z_{n+1} - z'_{n+1}| = \sum_{k=1}^{n+1} |z_k - z'_k|. \quad (9.87) \end{aligned}$$

On a pu passer de (9.85) à (9.86) parce que $|z_{n+1}| \leq 1$ et $|p'_n| \leq 1$ (produit de n nombres complexes de module au plus 1). L'utilisation de l'hypothèse de récurrence pour majorer $|p_n - p'_n|$ a permis le passage de (9.86) à (9.87). Le a) est donc établi.

Pour prouver b), en prenant dans a) $z_1 = \dots = z_n = v$ et $z'_1 = \dots = z'_n = w$, on voit que pour tous nombres complexes v et w de module au plus 1,

$$\left|v^n - w^n\right| \leq n|v - w|. \tag{9.88}$$

Il est d'ailleurs facile d'obtenir directement cette inégalité en utilisant la factorisation de $v^n - w^n$ par $(v - w)$. Prenons maintenant

$$v := 1 - \frac{t^2}{2n} + \frac{t^2}{n}\varepsilon\left(n^{-1/2}t\right), \quad w := \exp\left(\frac{-t^2}{2n}\right).$$

Pour $n \geq n_0 = n_0(t)$, on a $|v| \leq 1$ et pour tout n, $|w| < 1$. En remarquant que

$$\exp\left(\frac{-t^2}{2n}\right) = 1 - \frac{t^2}{2n} + \frac{\delta_n(t)}{n}, \quad \lim_{n\to+\infty} \delta_n(t) = 0,$$

on déduit alors de (9.88) que

$$\begin{aligned} \forall n \geq n_0(t), \quad \left|\left(1 - \frac{t^2}{2n} + \frac{t^2}{n}\varepsilon\left(n^{-1/2}t\right)\right)^n - \exp\left(\frac{-t^2}{2}\right)\right| &\leq n\left|\frac{t^2}{n}\varepsilon\left(n^{-1/2}t\right) - \frac{\delta_n(t)}{n}\right| \\ &\leq t^2\left|\varepsilon\left(n^{-1/2}t\right)\right| + |\delta_n(t)| \end{aligned}$$

et ce majorant tend vers 0 quand n tend vers l'infini. □

Théorème 9.37 (De Moivre-Laplace)**.** *Si* S_n *est une variable aléatoire de loi binomiale de paramètres* n *et* $p \in]0, 1[$, *on a avec* $q := 1 - p$,

$$S_n^* := \frac{S_n - np}{\sqrt{npq}} = \sqrt{\frac{n}{pq}}\left(\frac{S_n}{n} - p\right) \xrightarrow[n\to+\infty]{\text{loi}} \mathfrak{N}(0, 1).$$

Comme la fonction de répartition Φ *de* $\mathfrak{N}(0,1)$ *est continue sur* $\mathbb{R}$*, ceci équivaut à*

$$\forall x \in \mathbb{R}, \quad \mathbf{P}(S_n^* \leq x) \xrightarrow[n\to+\infty]{} \Phi(x) = \int_{-\infty}^{x} \exp\Big(-\frac{t^2}{2}\Big)\frac{\mathrm{d}t}{\sqrt{2\pi}}.$$

Preuve. C'est un simple corollaire du théorème 9.35 en remarquant que S_n a *même loi* 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. □

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 $O(n^{-1/2})$, cf. [ICP] chap. 7.

Une application directe de ce théorème est la construction d'*intervalles de confiance* pour l'estimation d'une probabilité inconnue p à partir de l'observation d'un *échantillon* de n variables de Bernoulli indépendantes de paramètre p. Considérons pour $t > 0$ l'événement

$$A_{n,t} := \Big\{\omega \in \Omega; -t \leq \sqrt{\frac{n}{pq}}\Big(\frac{S_n(\omega)}{n} - p\Big) \leq t\Big\}.$$

Le théorème de De Moivre-Laplace nous dit que pour n assez grand, on peut utiliser l'approximation :

$$\mathbf{P}\big(A_{n,t}\big) \simeq \Phi(t) - \Phi(-t) = 2\Phi(t) - 1.$$

Ceci peut se réécrire

$$\mathbf{P}\Big(\frac{S_n}{n} - t\sqrt{\frac{pq}{n}} \leq p \leq \frac{S_n}{n} + t\sqrt{\frac{pq}{n}}\Big) = 2\Phi(t) - 1 + \varepsilon_n.$$

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 $p = 1/2$. D'où

$$\sqrt{pq} \leq \sqrt{\frac{1}{4}} = \frac{1}{2}, \tag{9.89}$$

de sorte qu'en notant

$$B_{n,t} := \Big\{\omega \in \Omega; \frac{S_n(\omega)}{n} - \frac{t}{2\sqrt{n}} \leq p \leq \frac{S_n(\omega)}{n} + \frac{t}{2\sqrt{n}}\Big\},$$

l'inclusion $A_{n,t} \subset B_{n,t}$ nous donne :

$$\mathbf{P}\big(B_{n,t}\big) \geq 2\Phi(t) - 1 + \varepsilon_n. \tag{9.90}$$

En pratique, n est fixé et on a observé des valeurs numériques explicites $x_1, \dots, x_n$ que l'on interprète comme les valeurs de $X_1(\omega), \dots, X_n(\omega)$ pour *un même* ω tiré au sort (suivant $\mathbf{P}$). 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*[9] 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 (9.89). Ainsi en estimant pq par $M_n(1 - M_n)$ où $M_n := S_n/n$, on obtient au niveau de confiance 95% l'intervalle

$$J_n = \left[M_n(\omega) - 1,96\sqrt{\frac{M_n(\omega)(1 - M_n(\omega))}{n}}; M_n(\omega) + 1,96\sqrt{\frac{M_n(\omega)(1 - M_n(\omega))}{n}}\right].$$

La figure 9.2 représente (verticalement) les intervalles I_n et J_n calculés sur le même échantillon simulé de taille 200 de la loi Bern(0.75). Sur 100 simulations de ce type, on observe 4 intervalles I ne contenant pas la vraie valeur de p (les numéros 50, 60, 69 et 95) et 6 intervalles J ne contenant pas p (numéros 2, 29, 50, 60, 69 et 95).

Corollaire 9.38 (du th. 9.35, approximation gaussienne d'une loi de Poisson)**.** *Si Y_n suit une loi de Poisson de paramètre $\alpha_n = na$ (a constante),*

$$Y_n^* := \frac{Y_n - \alpha_n}{\sqrt{\alpha_n}} \xrightarrow[n\to+\infty]{\text{loi}} \mathfrak{N}(0,1).$$

Preuve. Il suffit de remarquer que Y_n a *même loi* que $X_1 + \cdots + X_n$ où les X_k sont indépendantes et de même loi Pois(a). Par le théorème 9.35, le S_n^* bâti sur la suite des X_k converge en loi vers $\mathfrak{N}(0,1)$. Comme $\mathbf{E}Y_n = \alpha_n$ et $\operatorname{Var} Y_n = \alpha_n$, on voit que Y_n^* a même loi que S_n^*. Il en résulte que Y_n^* converge en loi vers la même limite que S_n^*. □

9. 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 « $\mathbf{P}(p \in [0,51; 0,55]) \geq 0,95$ ». 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 $0,95$.

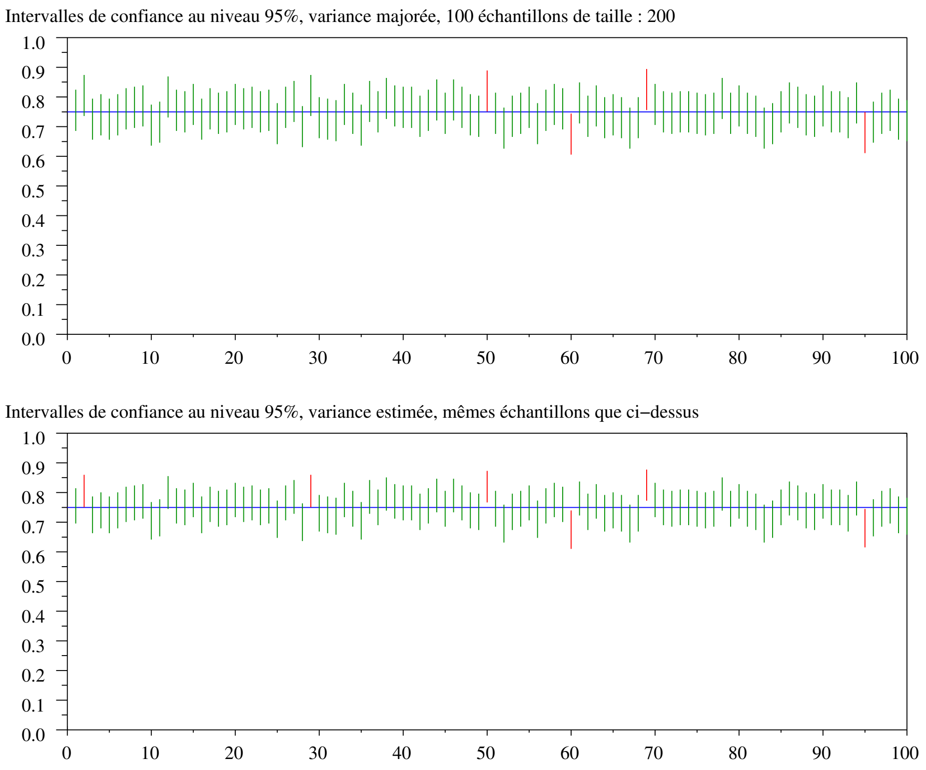


FIGURE 9.2 – Intervalles I_{200} et J_{200} pour 100 échantillons simulés avec $p = 0.75$.

Ce corollaire explique que l'approximation poissonienne d'une binomiale $\mathrm{Bin}(n, p)$ finisse par devenir gaussienne quand $\alpha_n = np$ augmente. En pratique, on préfère l'approximation gaussienne dès que $np \geq 20$.

Remarque 9.39. Revenons au théorème de De Moivre-Laplace. Notons μ_n la loi de S_n^*, $\mu := \mathfrak{N}(0, 1)$,

$$E_n := \left\{ \frac{k - nm}{\sigma\sqrt{n}}; \quad 0 \leq k \leq n \right\}$$

l'ensemble fini des valeurs possibles de S_n^* et B le borélien dénombrable $B := \cup_{n\geq 1} E_n$. Comme $\mu_n(E_n) = 1$ et $E_n \subset B$, on a $\mu_n(B) = 1$ pour tout n. Ainsi $\mu_n(B)$ converge vers 1. Or $\mu(B) = 0$ puisque B est dénombrable et μ est à densité par rapport à λ_1. Donc $\mu_n(B)$ ne converge pas vers $\mu(B)$. Il n'y a pas contradiction avec la convergence étroite de μ_n vers μ car B est d'intérieur vide (il ne contient aucun intervalle) et dense dans $\mathbb{R}$ (exercice), donc de fermeture $\mathbb{R}$. La frontière de B est donc. . .$\mathbb{R}$ lui même ! Donc $\mu(\partial B) = 1 \neq 0$. On voit ainsi que la restriction $\mu(\partial B) = 0$ dans (9.37) n'est pas seulement destinée à éliminer les cas pathologiques. Elle intervient aussi dans le plus classique et le plus utilisé de tous les exemples de convergence en loi.

On a le même phénomène avec l'exemple 9.6, où $B = \mathbb{Q} \cap]0, 1]$ et plus généralement, chaque fois que l'on a convergence étroite d'une suite de lois discrètes vers une loi continue.