

# Outils mathématiques pour les sciences

Charles SUQUET

Chapitres 1 à 6



# Chapitre 1

## Séries numériques

### 1.1 Révision

En préliminaire, on rappelle ici le calcul de la somme des termes d'une suite géométrique.

**Définition 1.1** (Suite géométrique). *Une suite  $(u_n)_{n \in \mathbb{N}}$  de nombres réels ou complexes est une suite géométrique si elle vérifie pour tout  $n \in \mathbb{N}$  la relation :*

$$u_{n+1} = qu_n,$$

où  $q$  est une constante, réelle ou complexe, indépendante de  $n$  et appelée raison de la suite.

**Remarque 1.2.** La connaissance du premier terme et de la raison d'une suite géométrique permet de calculer tous les termes de la suite. On vérifie facilement que :

$$\forall n \in \mathbb{N}, \quad u_n = u_0 q^n.$$

On peut généraliser cette formule en commençant avec un terme d'indice quelconque :

$$u_n = u_1 q^{n-1} = u_2 q^{n-2} = u_3 q^{n-3} = \dots = u_k q^{n-k}.$$

Remarquez que dans ces formules, la somme de l'indice du premier terme utilisé et de l'exposant de  $q$  vaut toujours  $n$ .

Il est essentiel de savoir calculer la somme des termes d'une suite géométrique.

**Proposition 1.3** (somme des termes d'une suite géométrique). *Soit  $(u_n)_{n \in \mathbb{N}}$  une suite géométrique de raison  $q$ . Posons pour tout entier  $n$  :*

$$S_n := \sum_{k=0}^n u_k = u_0 + u_1 + u_2 + \dots + u_n.$$

Alors

- Si  $q \neq 1$ ,

$$S_n = \frac{u_0(1 - q^{n+1})}{1 - q}.$$

- Si  $q = 1$ ,  $S_n = (n + 1)u_0$ .

*Preuve.* En factorisant par le premier terme :

$$S_n = u_0 + u_0q^1 + u_0q^2 + u_0q^3 + \cdots + u_0q^n = u_0(1 + q + q^2 + q^3 + \cdots + q^n),$$

on se ramène immédiatement au calcul de la somme  $S'_n$  définie par :

$$S'_n = 1 + q + q^2 + q^3 + \cdots + q^n = \sum_{k=0}^n q^k.$$

Dans le cas particulier où  $q = 1$ ,  $S'_n$  est la somme de  $n + 1$  termes (on commence à  $k = 0$  avec  $q^0 = 1$  et on va jusqu'à  $n$ ) tous égaux à 1. Elle vaut donc  $(n + 1)$  et donc  $S_n = (n + 1)u_0$ .

Supposons désormais que  $q \neq 1$ . Pour calculer  $S'_n$ , il y a une petite astuce qui consiste à remarquer que  $(1 - q)S'_n = S'_n - qS'_n$  se simplifie considérablement. Pour le voir, il est commode de disposer le calcul comme ci-dessous.

$$\begin{array}{r} S'_n \quad \quad = 1 + q + q^2 + q^3 + \cdots \cdots \cdots + q^{n-1} + q^n \\ -qS'_n \quad = -q - q^2 - q^3 - q^4 - \cdots \cdots \cdots - q^n - q^{n+1} \\ \hline S'_n - qS'_n = 1 + 0 + 0 + 0 + \cdots \cdots \cdots \cdots \cdots + 0 - q^{n+1} \end{array}$$

On obtient ainsi

$$S'_n - qS'_n = (1 - q)S'_n = 1 - q^{n+1},$$

d'où le résultat en divisant par  $(1 - q) \neq 0$ . □

La formule pour le calcul de  $S'_n$  s'adapte facilement à des situations analogues. Par exemple calculons pour  $q \neq 1$ ,

$$T_n = q^3 + q^4 + q^5 + \cdots + q^n = \sum_{k=3}^n q^k.$$

En mettant le premier terme en facteur, on voit qu'on se ramène au calcul de  $S'_{n-3}$  :

$$\begin{aligned} T_n &= q^3 + q^4 + q^5 + \cdots + q^{n-1} + q^n \\ &= q^3(1 + q + q^2 + \cdots + q^{n-4} + q^{n-3}) \\ &= q^3 \frac{1 - q^{n-3+1}}{1 - q} \\ &= q^3 \frac{1 - q^{n-2}}{1 - q} = \frac{q^3 - q^{n+1}}{1 - q}. \end{aligned}$$

Notons que l'on serait arrivé plus vite au même résultat en remarquant que

$$T_n = S'_n - S'_2 = \frac{1 - q^{n+1}}{1 - q} - \frac{1 - q^{2+1}}{1 - q} = \frac{q^3 - q^{n+1}}{1 - q}.$$

On peut compléter la proposition 1.3 dans le cas  $q \neq 1$  avec la règle-formule suivante permettant de s'y retrouver dans tous les cas de sommes de termes consécutifs d'une suite géométrique :

$$\text{Somme de termes consécutifs} = (\text{1er terme}) \times \frac{1 - \text{raison}^{\text{nombre de termes}}}{1 - \text{raison}}.$$

## 1.2 Deux exemples introductifs

Peut-on donner un sens à la somme de *tous* les termes d'une suite infinie de nombres réels? Voici deux exemples « familiers » montrant que c'est possible.

**Exemple 1.4** (la tablette de chocolat). Peut-on partager un rectangle en une infinité de rectangles de façon à enlever à chaque étape la moitié de l'aire restante? La figure 1.1 montre une solution.

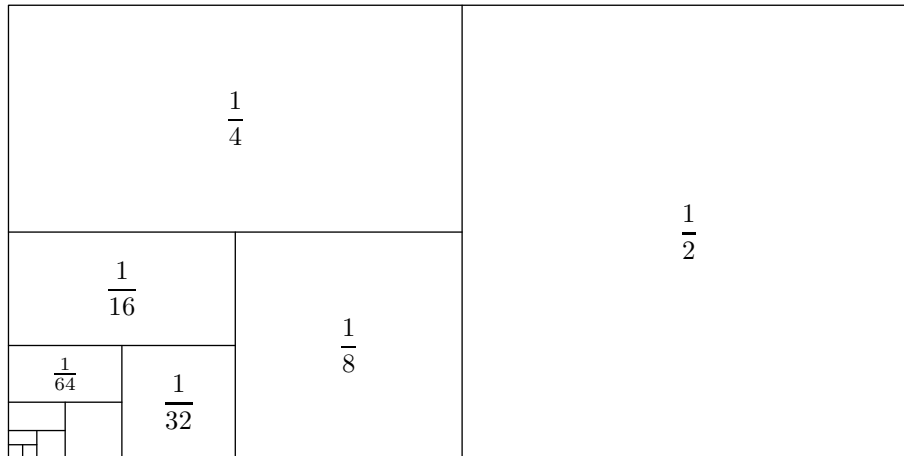


FIGURE 1.1 – Découpage infini d'une tablette de chocolat

En prenant comme unité d'aire l'aire du rectangle de départ (la tablette), on en déduit la relation :

$$1 = \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} + \frac{1}{32} + \frac{1}{64} + \frac{1}{128} + \dots = \lim_{n \rightarrow +\infty} \sum_{k=1}^n 2^{-k}.$$

**Exemple 1.5** (une division qui « ne tombe pas juste »). Quelle écriture décimale peut-on donner du nombre rationnel  $\frac{4}{33}$ ? On sait depuis l'école primaire qu'il suffit de poser la division.

$$\begin{array}{r} 4 \\ 40 \\ 70 \\ 40 \\ 70 \\ 40 \\ 70 \\ 40 \\ 70 \\ 4 \end{array} \quad \left| \begin{array}{r} 33 \\ \hline 0,121212121212 \end{array} \right.$$

On voit bien que les restes 4 et 7 se reproduisent périodiquement et que la division pourrait continuer à l'infini. Si on veut donner un sens mathématique précis à cette « division infinie », on peut écrire :

$$\frac{4}{33} = 0,12121212121212\dots = \lim_{n \rightarrow +\infty} S_n,$$

en définissant  $S_n$  par

$$S_n = 0, \underbrace{12\dots 12}_{2n \text{ chiffres}}.$$

Vérifions que  $S_n$  ainsi définie a bien pour limite  $\frac{4}{33}$  quand  $n$  tend vers l'infini. Pour cela on écrit  $S_n$  sous la forme :

$$S_n = \frac{12}{100} + \frac{12}{10\,000} + \cdots + \frac{12}{10^{2n}} = \sum_{k=1}^n \frac{12}{100^k} = \frac{12}{100} \sum_{k=1}^n \frac{1}{100^{k-1}} = \frac{12}{100} \sum_{j=0}^{n-1} \left(\frac{1}{100}\right)^j,$$

ce qui montre que  $S_n$  est la somme des  $n$  premiers termes de la suite géométrique de premier terme  $u_0 = \frac{12}{100}$  et de raison  $\frac{1}{100} = 10^{-2}$ . Par conséquent,

$$S_n = \frac{12}{100} \times \frac{1 - (10^{-2})^n}{1 - \frac{1}{100}} = \frac{12}{100} \times \frac{100}{99} (1 - 10^{-2n}) = \frac{4}{33} (1 - 10^{-2n}),$$

ce qui converge bien vers  $\frac{4}{33}$  quand  $n$  tend vers l'infini.

### 1.3 Généralités

**Définition 1.6.** Soit  $(u_k)_{k \in \mathbb{N}}$  une suite de nombres réels. Pour tout  $n \in \mathbb{N}$ , notons

$$S_n = \sum_{k=0}^n u_k = u_0 + u_1 + \cdots + u_n.$$

a) Si  $S_n$  tend vers une limite finie  $S$  quand  $n$  tend vers  $+\infty$ , on dit que la série de terme général  $u_k$  converge et on note :

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

b) Si  $S_n$  n'a pas de limite finie quand  $n$  tend vers l'infini (c.-à-d. tend vers  $-\infty$  ou vers  $+\infty$  ou n'a aucune limite), on dit que la série  $\sum_{k=0}^{+\infty} u_k$  diverge.

Nous venons de voir deux séries convergentes (exemples 1.4 et 1.5), justifiant les égalités :

$$1 = \sum_{k=1}^{+\infty} 2^{-k}, \quad \frac{4}{33} = \sum_{k=1}^{+\infty} \frac{12}{100^k}.$$

Ici la borne inférieure dans le symbole  $\sum$  est «  $k = 1$  » au lieu de «  $k = 0$  », mais on peut considérer que cela relève de la définition ci-dessus en posant  $u_0 = 0$  (ou adapter cette définition de manière évidente).

**Remarque 1.7.** On s'autorise dans les deux cas a) et b) l'écriture «  $\sum_{k=0}^{+\infty} u_k$  », mais on ne pourra considérer cet objet comme un nombre réel (et donc le faire intervenir dans des calculs) qu'après avoir établi la convergence.

**Remarque 1.8.** Les écritures «  $\sum_{k=0}^{+\infty} u_k$  » et «  $\sum_{k=0}^{+\infty} u_k$  » ont la même signification mathématique. La différence n'est que typographique. La première est utilisée pour les formules « en ligne » à l'intérieur d'une phrase pour ne pas agrandir l'interligne, la deuxième dans les formules « centrées » et mises en évidence. Pour l'écriture manuscrite, on n'utilise généralement que la deuxième forme.

**Définition 1.9** (somme partielle et reste). Avec les notations de la définition 1.6,  $S_n$  est appelée somme partielle de rang  $n$  de la série ( $S_0 = u_0$  étant la somme partielle de rang zéro). Si la série converge,  $R_n = S - S_n$  est appelé reste de rang  $n$  de la série.

**Remarque 1.10.** On peut adapter facilement la définition 1.6 pour définir les séries  $\sum_{k=n_0}^{+\infty} u_k$ , pour  $n_0$  entier quelconque. La convergence d'une série ne dépend pas de ses premiers termes. En effet pour tout  $n_0$  fixé, puisque  $S_{n_0}$  est une constante<sup>1</sup>,  $S_n$  converge vers une limite finie  $S$  si et seulement si  $S_n - S_{n_0}$  converge vers  $S - S_{n_0}$ . Autrement dit :

$$\forall n_0 \in \mathbb{N}, \quad \sum_{k=0}^{+\infty} u_k \quad \text{et} \quad \sum_{k=n_0}^{+\infty} u_k \quad \text{sont de même nature,}$$

c'est-à-dire ou toutes deux convergentes ou toutes deux divergentes.

**Remarque 1.11.** Ce qui précède montre que le reste de rang  $n$  d'une série, lorsqu'il existe, peut lui même s'écrire comme une série convergente :

$$\text{si } \sum_{k=0}^{+\infty} u_k \text{ converge, son reste de rang } n \text{ est } R_n = \sum_{k=n+1}^{+\infty} u_k.$$

Nous n'avons vu pour l'instant que deux exemples de séries. Ces deux exemples ont d'ailleurs un point commun, c'est d'être bâtis sur des suites géométriques. Nous pouvons systématiser ceci en examinant le cas particulier important des séries géométriques. Pour de telles séries, on dispose d'un critère de convergence très simple et en cas de convergence, on sait facilement calculer  $S_n$ ,  $S$  et  $R_n$ . Si  $(u_k)_{k \in \mathbb{N}}$  est une suite géométrique de raison  $q$ , l'étude de la série géométrique  $\sum_{k=0}^{+\infty} u_k = \sum_{k=0}^{+\infty} u_0 q^k$  se ramène à celle de la série géométrique standard  $\sum_{k=0}^{+\infty} q^k$  par mise en facteur du premier terme dans les sommes partielles.

**Proposition 1.12** (série géométrique standard).

1. La série géométrique standard  $\sum_{k=0}^{+\infty} q^k$  converge si et seulement si  $|q| < 1$ .

2. Si  $|q| < 1$ ,  $\sum_{k=0}^{+\infty} q^k = \frac{1}{1-q}$ .

3. Si  $|q| < 1$ ,  $R_n = \sum_{k=n+1}^{+\infty} q^k = \frac{q^{n+1}}{1-q}$ .

*Preuve.* On utilise la formule de calcul de  $S_n$  somme partielle de rang  $n$ , cf. Prop. 1.3. Dans le cas particulier  $q = 1$ ,  $S_n = (n+1)$  tend vers l'infini avec  $n$ , donc la série diverge. Si  $q \neq 1$ ,

$$S_n = \frac{1 - q^{n+1}}{1 - q}.$$

a) Si  $|q| > 1$ , la valeur absolue de  $q^{n+1}$  tend vers l'infini et il en est de même pour celle du numérateur  $1 - q^{n+1}$ . Par conséquent la valeur absolue de  $S_n$  tend vers l'infini, ce qui empêche que  $S_n$  ait une limite finie. Donc la série diverge<sup>2</sup>.

1. Quitte à prolonger la définition des  $u_k$  s'ils ne sont définis que pour  $k \geq n_0$  en posant  $u_k = 0$  pour  $k < n_0$ .

2. Si  $q > 1$ ,  $S_n$  tend vers  $+\infty$ , si  $q < -1$ ,  $S_{2n}$  tend vers  $+\infty$  et  $S_{2n+1}$  tend vers  $-\infty$ .

- b) Si  $q = -1$ ,  $S_{2n} = 1$  et  $S_{2n+1} = 0$  pour tout  $n$  (évident directement), la suite des sommes partielles oscille indéfiniment entre les valeurs 0 et 1 et n'a donc pas de limite.
- c) Si  $|q| < 1$ ,  $|q^{n+1}| = |q|^{n+1}$  tend vers 0 quand  $n$  tend vers l'infini. Il en va de même pour  $q^{n+1}$ , donc  $S_n$  a pour limite finie  $1/(1 - q)$ .

À ce stade, nous avons justifié les points 1 et 2 de la proposition. Pour le calcul du reste  $R_n$  lorsque  $|q| < 1$ , on écrit  $R_n = S - S_n$ , avec ici  $S = 1/(1 - q)$ , d'où :

$$R_n = S - S_n = \frac{1}{1 - q} - \frac{1 - q^{n+1}}{1 - q} = \frac{q^{n+1}}{1 - q},$$

le point 3 est ainsi vérifié. □

Lorsque l'on dispose de deux séries convergentes, on peut fabriquer une nouvelle série convergente par combinaison linéaire.

**Proposition 1.13.** *Si  $\sum_{k=0}^{+\infty} u_k$  et  $\sum_{k=0}^{+\infty} v_k$  convergent, la série  $\sum_{k=0}^{+\infty} (au_k + bv_k)$  converge pour toutes constantes réelles  $a, b$ .*

La preuve est laissée en exercice. Grâce à cette proposition, on peut voir immédiatement par exemple que la série de terme général  $u_k = 3(\sqrt{2})^{-k} - 5 \times (0,995)^k$  converge<sup>3</sup>.

Peut-on trouver d'autres exemples de séries convergentes que les séries géométriques et leurs combinaisons linéaires? Voici une méthode simple pour construire de nombreux exemples. On part d'une suite réelle  $(v_n)_{n \in \mathbb{N}}$  ayant une limite finie  $\ell$  quand  $n$  tend vers l'infini. On pose  $u_0 = v_0$  et pour tout  $k \geq 1$ ,  $u_k = v_k - v_{k-1}$ . Alors la série de terme général  $u_k$  converge et a pour somme  $S = \ell$ . En effet,

$$S_n = \sum_{k=0}^n u_k = v_0 + (v_1 - v_0) + (v_2 - v_1) + (v_3 - v_2) + \cdots + (v_{n-1} - v_{n-2}) + (v_n - v_{n-1}) = v_n$$

Par exemple, la suite de terme général

$$v_n = \frac{n}{2n + 1}$$

est convergente avec pour limite  $1/2$ . On calcule alors

$$u_k = v_k - v_{k-1} = \frac{k}{2k + 1} - \frac{k - 1}{2(k - 1) + 1} = \cdots = \frac{1}{4k^2 - 1}$$

et on en déduit que la série de terme général  $u_k$  converge et a pour somme :

$$\sum_{k=0}^{+\infty} \frac{1}{4k^2 - 1} = \frac{1}{2}.$$

Réciproquement, toute série convergente  $\sum_{k=0}^{+\infty} u_k$  peut être construite sur ce modèle. Il suffit de prendre  $v_n = S_n$  car  $u_k = S_k - S_{k-1}$  pour tout  $k \geq 1$ .

Finalement, puisqu'il y a une telle correspondance entre suites convergentes et séries convergentes, à quoi bon étudier les séries et pourquoi ne pas se contenter d'étudier les suites? La réponse à cette objection naturelle est que la correspondance vue ci-dessus a surtout un intérêt théorique. En effet, si on peut toujours calculer  $u_k$  à partir de  $v_k$ ,

---

3. Ne me croyez pas sur parole, vérifiez.



*l'opération inverse est rarement possible en pratique.* La solution qui consiste à prendre  $v_n = S_n$  n'a d'intérêt que si l'on a une formule de calcul explicite de  $S_n$ , c.-à-d. permettant de calculer directement  $S_n$  à partir de  $n$ , sans additionner un par un tous les  $u_k$ . C'est le cas pour les séries géométriques, mais c'est très loin d'être la situation générale. En pratique, *seuls les  $u_k$  sont connus.* Le premier problème est alors de savoir si la série converge. Si c'est le cas, on ne sait pas forcément calculer la valeur exacte de sa somme  $S$ . En fait la convergence permet d'approximer numériquement le nombre  $S$  par  $S_n$  (calculable par additions) et l'erreur d'approximation commise est  $R_n$ . Il est alors utile de pouvoir contrôler la vitesse de convergence de  $R_n$  vers 0, pour savoir combien de termes de la série il va falloir additionner pour obtenir une précision donnée.

**Proposition 1.14.** *Si une série converge, son terme général tend vers 0.*

Ceci nous donne le premier test à faire lorsqu'on est en présence d'une série : regarder si son terme général tend vers 0. Si cette condition n'est pas vérifiée, alors on est certain que la série diverge.

*Preuve de la proposition 1.14.* Soit  $S_n = \sum_{k=0}^n u_k$  la somme partielle de rang  $n$  d'une série convergente. Alors pour tout  $n \in \mathbb{N}^*$ ,  $u_n = S_n - S_{n-1}$ . Puisque la série converge, notons  $S$  sa somme qui est un nombre réel. Par définition de la convergence de la série,  $S_n$  tend vers  $S$  quand  $n$  tend vers l'infini. Il en va de même pour  $S_{n-1}$  :  $(S_{n-1})_{n \in \mathbb{N}^*}$  met ses pieds dans les traces de  $(S_n)_{n \in \mathbb{N}}$  avec un pas de retard, donc les deux suites parcourent le même chemin et ont même limite  $S$ . Par conséquent :

$$\lim_{n \rightarrow +\infty} u_n = \lim_{n \rightarrow +\infty} (S_n - S_{n-1}) = \lim_{n \rightarrow +\infty} S_n - \lim_{n \rightarrow +\infty} S_{n-1} = S - S = 0.$$

□

**Remarque 1.15.** Attention, la réciproque de la proposition est fautive ! Pour que la série  $\sum_{k=0}^{+\infty} u_k$  converge, il est *nécessaire* que  $u_k$  tende vers 0 (c'est ce que dit la proposition 1.14), mais ce n'est pas *suffisant*. En effet, il existe des séries divergentes dont le terme général tend vers 0. La plus célèbre de toutes est la *série harmonique*.

**Exemple 1.16** (divergence de la série harmonique). La série

$$\sum_{k=1}^{+\infty} \frac{1}{k}$$

est appelée série harmonique. Son terme général tend évidemment vers zéro. Pour voir qu'elle diverge, nous allons minorer  $S_{2n} - S_n$  comme suit. Pour tout  $n \geq 1$ ,

$$S_{2n} - S_n = \underbrace{\frac{1}{n+1} + \frac{1}{n+2} + \frac{1}{n+3} + \cdots + \frac{1}{n+n}}_{n \text{ termes supérieurs ou égaux au dernier } \frac{1}{2n}} \geq n \times \frac{1}{2n} = \frac{1}{2}.$$

Raisonnons maintenant par l'absurde en supposant que la série harmonique converge et notons  $S$  sa somme. Alors  $S_{2n}$  comme  $S_n$  doivent converger vers  $S$  et en passant à la limite dans l'inégalité ci-dessus, on obtiendrait  $S - S = 0 \geq \frac{1}{2}$ , ce qui est bien sûr faux. Par conséquent, la série harmonique ne peut converger.

## 1.4 Séries à termes positifs

Les séries à termes positifs jouent un rôle clé dans la théorie des séries car leur divergence n'arrive que si et seulement si la suite des sommes partielles tend vers  $+\infty$ . Les possibilités de divergence sont bien plus diverses pour une série dont les termes changent de signe une infinité de fois.

Pour comprendre le comportement d'une série à termes positifs  $u_k$ , il convient d'observer que dans ce cas  $(S_n)_{n \in \mathbb{N}}$  est une suite *croissante* de réels positifs. En effet, pour tout  $n \geq 1$ ,  $S_n - S_{n-1} = u_n \geq 0$ , donc  $S_n \geq S_{n-1}$ . Il est commode de représenter une telle suite croissante  $(S_n)_{n \in \mathbb{N}}$  (ou la série  $\sum_{k=0}^{+\infty} u_k$ ) par un escalier infini, où la  $n^{\text{e}}$  marche est le segment horizontal d'extrémités les points de coordonnées  $(n, S_n)$  et  $(n+1, S_n)$ . Ainsi la figure 1.2 représente (les premières marches de) l'escalier associé à la série harmonique (en posant  $u_0 = 0$ ,  $S_0 = u_0$ ).

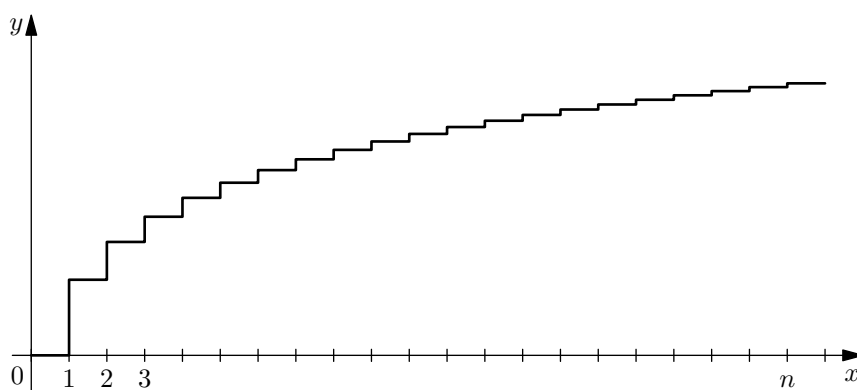


FIGURE 1.2 – Escalier infini de la série harmonique

Sur cette figure, on voit que la pente est de plus en plus douce, mais comme dans le cas d'une représentation graphique de fonction, cela ne suffit pas pour conclure à l'existence d'une asymptote horizontale. On observe le même phénomène avec la représentation graphique de la fonction logarithme et il est bien connu que  $\ln x$  tend vers l'infini quand  $x$  tend vers l'infini et donc qu'il ne peut y avoir d'asymptote horizontale<sup>4</sup>.

Pour un escalier infini représentant une série à termes positifs, il y a une alternative :

- ou bien l'escalier est *plafonné*, cf. figure 1.3, autrement dit il existe une altitude  $M$  qu'il ne peut *jamaïs* franchir (mathématiquement,  $M$  est un *majorant* de l'ensemble de toutes les valeurs de la suite  $(S_n)_{n \in \mathbb{N}} : \forall n \in \mathbb{N}, S_n \leq M$ ) ;
- ou bien il n'est pas plafonné et finit donc par dépasser toute altitude fixée à l'avance.

Dans le deuxième cas, comme la suite  $(S_n)_{n \in \mathbb{N}}$  est croissante, pour chaque altitude  $A$  fixée, dès que  $S_{n_0} > A$  pour un certain indice  $n_0$ ,  $S_n > A$  pour tous les indices suivants. Comme  $A$  est quelconque, cela signifie que  $S_n$  *tend vers l'infini*.

4. En fait, on peut montrer que pour la série harmonique,  $S_n / \ln n$  tend vers 1 en l'infini et donc  $S_n$  tend vers  $+\infty$  avec la même vitesse que  $\ln n$  (cf. T.D.).

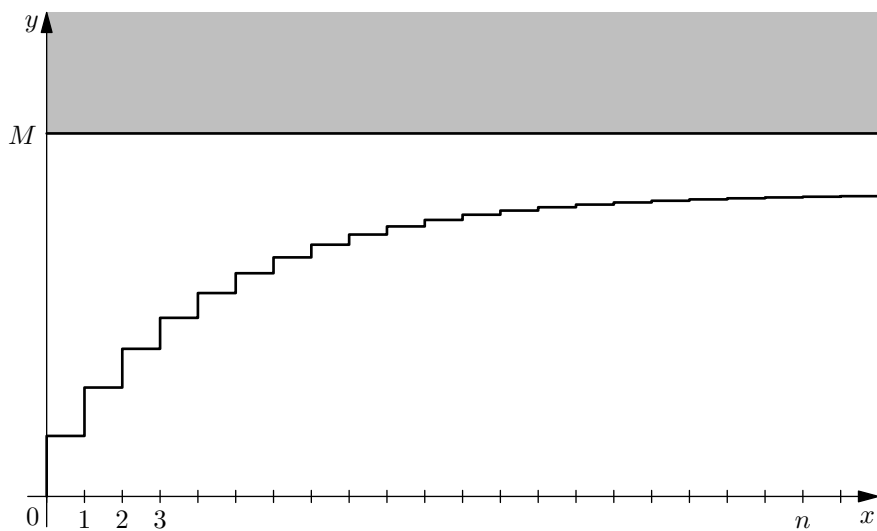


FIGURE 1.3 – Escalier infini plafonné

Revenons au premier cas. Notons  $S$  la *plus petite altitude possible d'un plafond* pour l'escalier<sup>5</sup>, cf. figure 1.4. Alors  $S_n$  a pour limite  $S$ . Pour le voir, on note d'abord que pour tout  $n \in \mathbb{N}$ ,  $S_n \leq S$  (puisque  $S$  est l'altitude d'un plafond pour l'escalier). Si on choisit

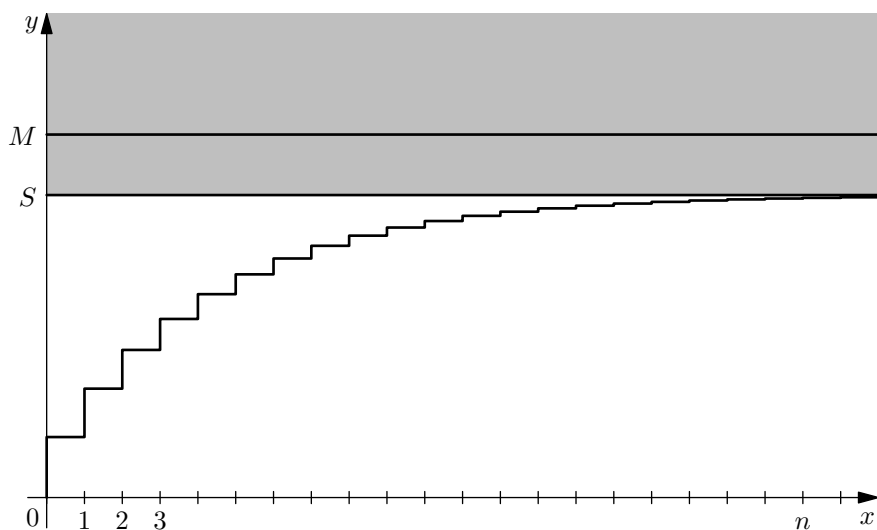


FIGURE 1.4 – Escalier infini plafonné et son asymptote

une altitude quelconque inférieure strictement à  $S$ , disons  $S - \varepsilon$  avec  $\varepsilon > 0$ , alors cette altitude ne peut être celle d'un plafond, puisque le plus bas des plafonds possibles a pour altitude  $S$ . Donc il existe au moins une marche qui franchit cette altitude  $S - \varepsilon$ , notons  $n_0$  son numéro ( $S_{n_0} > S - \varepsilon$ ). Mais comme l'escalier est croissant, toutes les marches suivantes sont au dessus de l'altitude  $S - \varepsilon$ , autrement dit :  $S_n > S - \varepsilon$  pour tout  $n \geq n_0$ . Comme le raisonnement a été fait avec  $\varepsilon > 0$  *quelconque*, nous avons montré que :

$$\forall \varepsilon > 0, \exists n_0 \in \mathbb{N}, \forall n \geq n_0, \quad S - \varepsilon < S_n \leq S.$$

Ceci prouve que  $S_n$  tend vers  $S$  quand  $n$  tend vers l'infini.

5. L'ensemble des majorants d'un ensemble non vide  $E$  de nombres réels a un plus petit élément, appelé borne supérieure de  $E$ . C'est une propriété fondamentale de l'ensemble  $\mathbb{R}$  des nombres réels. Ici on prend pour  $E$  l'ensemble de toutes les valeurs prises par  $S_n$  lorsque  $n$  décrit  $\mathbb{N}$  et  $S$  est la borne supérieure de  $E$ .

Résumons cette discussion comme suit.

**Proposition 1.17.** *Pour une série à termes positifs  $\sum_{k=0}^{+\infty} u_k$ , la suite des sommes partielles  $(S_n)_{n \in \mathbb{N}}$  est croissante.*

1. *Ou bien  $(S_n)_{n \in \mathbb{N}}$  est majorée. Elle converge alors vers un réel  $S$  fini. La série  $\sum_{k=0}^{+\infty} u_k$  converge et a pour somme  $S$ .*
2. *Ou bien  $(S_n)_{n \in \mathbb{N}}$  n'est pas majorée. Elle tend alors vers  $+\infty$  et la série  $\sum_{k=0}^{+\infty} u_k$  diverge.*

**Remarque 1.18.** On a clairement un résultat analogue pour une série dont tous les termes sont positifs à partir d'un certain rang  $k_0$ , la seule modification étant que la croissance de  $(S_n)_{n \in \mathbb{N}}$  a lieu à partir de  $k_0$ . On peut aussi étendre ce résultat aux séries à termes tous négatifs à partir d'un certain rang  $k_0$ , dans ce cas  $(S_n)_{n \in \mathbb{N}}$  décroît à partir de  $k_0$ , il faut remplacer « majorée » par « minorée » et  $+\infty$  par  $-\infty$  pour la limite de  $S_n$  dans le cas 2.

Voyons maintenant une application importante de la proposition 1.17.

**Théorème 1.19** (théorème de comparaison des séries à termes positifs). *On suppose qu'il existe un entier  $k_0$  tel que :*

$$\forall k \geq k_0, \quad 0 \leq u_k \leq v_k.$$

Alors

- i) *la convergence de  $\sum_{k=0}^{+\infty} v_k$  implique celle de  $\sum_{k=0}^{+\infty} u_k$ , de plus  $\sum_{k=0}^{+\infty} u_k \leq \sum_{k=0}^{+\infty} v_k$  ;*
- ii) *la divergence de  $\sum_{k=0}^{+\infty} u_k$  implique celle de  $\sum_{k=0}^{+\infty} v_k$ .*

Attention, l'hypothèse  $0 \leq u_k$  est indispensable. Par exemple en prenant pour tout  $k$ ,  $u_k = -1$  et  $v_k = 2^{-k}$ , on a bien  $u_k \leq v_k$ , la série  $\sum_{k=0}^{+\infty} v_k$  converge (série géométrique de raison  $1/2$ ), mais  $S_n = \sum_{k=0}^n u_k = -n - 1$  tend vers  $-\infty$  et donc  $\sum_{k=0}^{+\infty} u_k$  diverge.

*Preuve du théorème 1.19.* Comme la convergence ou la divergence d'une série ne dépendent pas des  $k_0$  premiers termes, on peut se contenter de faire la preuve avec  $k_0 = 0$ . Notons pour tout  $n \in \mathbb{N}$ ,

$$S_n = \sum_{k=0}^n u_k, \quad T_n = \sum_{k=0}^n v_k.$$

Des inégalités  $0 \leq u_k \leq v_k$  valables pour tout  $k$  on déduit immédiatement que

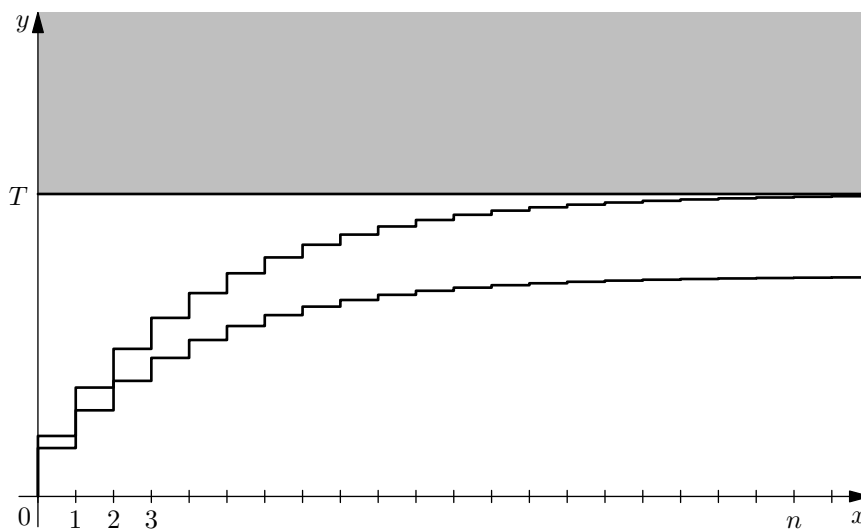
$$\forall n \in \mathbb{N}, \quad 0 \leq S_n \leq T_n.$$

- i) Si  $\sum_{k=0}^{+\infty} v_k$  converge, la suite croissante  $(T_n)_{n \in \mathbb{N}}$  est majorée par sa limite finie  $T$  qui est la somme de cette série, d'où :

$$\forall n \in \mathbb{N}, \quad 0 \leq S_n \leq T_n \leq T < +\infty.$$

La suite croissante  $(S_n)_{n \in \mathbb{N}}$  est donc majorée par le réel  $T$  et converge donc vers une limite finie  $S \leq T$ , cf. figure 1.5. Autrement dit,  $\sum_{k=0}^{+\infty} u_k$  converge et  $S = \sum_{k=0}^{+\infty} u_k \leq \sum_{k=0}^{+\infty} v_k = T < +\infty$ .

- ii) Si  $\sum_{k=0}^{+\infty} u_k$  diverge, la suite croissante  $(S_n)_{n \in \mathbb{N}}$  tend vers  $+\infty$ . Comme  $S_n \leq T_n$  pour tout  $n$ ,  $(T_n)_{n \in \mathbb{N}}$  tend aussi vers  $+\infty$  et  $\sum_{k=0}^{+\infty} v_k$  diverge.

FIGURE 1.5 – Le plafonnement de  $(T_n)$  implique celui de  $(S_n)$ 

□

Voyons maintenant trois exemples d'application de ce théorème pour prouver la convergence ou la divergence d'une série à termes positifs.

**Exemple 1.20.** La série à termes positifs  $\sum_{k=0}^{+\infty} \exp(-k^2)$  converge. En effet, pour tout  $k \in \mathbb{N}$ ,  $k \leq k^2$  (avec inégalité stricte dès que  $k > 1$ ), d'où  $-k^2 \leq -k$  et par croissance de la fonction exponentielle,  $\exp(-k^2) \leq \exp(-k) = (1/e)^k$ . Ce majorant est le terme général d'une série géométrique de raison  $1/e$  positive et strictement inférieure à 1, donc convergente. Par le point i) du théorème 1.19, on en déduit la convergence de  $\sum_{k=0}^{+\infty} \exp(-k^2)$ .

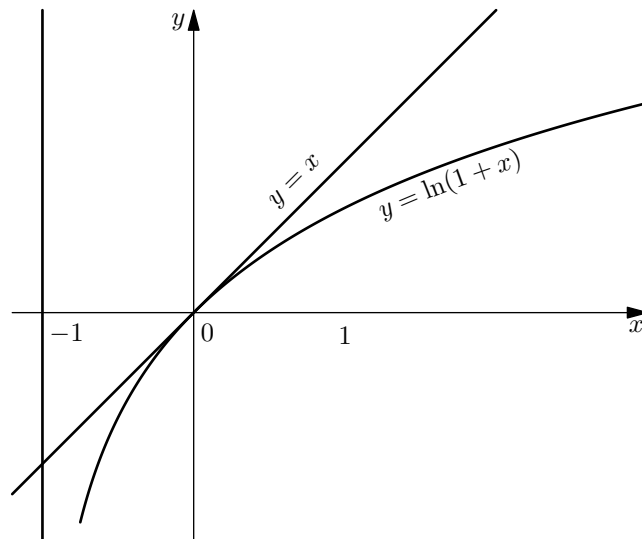
**Exemple 1.21.** La série à termes positifs  $\sum_{k=1}^{+\infty} k^{-a}$  diverge pour  $a < 1$ . Le cas où  $a \leq 0$  est évident puisqu'alors le terme général ne tend pas vers 0. Pour  $0 < a < 1$ , on note que  $k^a \leq k$  pour  $k \geq 1$ , d'où  $k^{-a} \geq k^{-1}$ . Or on sait déjà (cf. exemple 1.16) que la série harmonique, de terme général positif  $k^{-1}$  diverge. Donc par le point ii) du théorème 1.19,  $\sum_{k=1}^{+\infty} k^{-a}$  diverge.

**Exemple 1.22.** La série à termes positifs  $\sum_{k=1}^{+\infty} \frac{1}{\ln(1+k)}$  diverge. Pour le voir, on part de l'inégalité  $\ln(1+x) \leq x$ , valide pour tout réel  $x > -1$ , voir figure 1.6 et vérifier en exercice. En particulier, pour tout  $k \in \mathbb{N}^*$ ,  $\ln(1+k) \leq k$ , d'où  $\frac{1}{k} \leq \frac{1}{\ln(1+k)}$ . On conclut comme dans l'exemple précédent en utilisant la divergence de la série harmonique.

Le théorème de comparaison des séries à termes positifs permet de justifier la *règle des équivalents*, un outil fort utile pour étudier les séries à termes positifs. Il permet en effet de remplacer un terme général  $u_k$  d'expression compliquée par un terme général  $v_k$  plus simple qui lui est équivalent pour décider de la convergence d'une série à termes positifs. Commençons par un rappel sur l'équivalence des suites.

**Définition 1.23** (suites équivalentes). *Deux suites réelles  $(u_k)_{k \in \mathbb{N}}$  et  $(v_k)_{k \in \mathbb{N}}$  sont dites équivalentes en  $+\infty$  ou plus simplement équivalentes, notation  $u_k \sim v_k$ , s'il existe une suite réelle  $(w_k)_{k \in \mathbb{N}}$  et un entier  $k_0$  tels que*

$$\forall k \geq k_0, \quad u_k = v_k w_k \quad \text{et} \quad \lim_{k \rightarrow +\infty} w_k = 1.$$

FIGURE 1.6 – Inégalité de concavité  $\ln(1+x) \leq x$ 

En particulier si les  $v_k$  ne s'annulent jamais à partir d'un certain rang et si  $u_k/v_k$  tend vers 1, les deux suites sont équivalentes. Mais la définition ci-dessus est plus générale car elle autorise des suites  $(v_k)_{k \in \mathbb{N}}$  prenant la valeur zéro pour une infinité d'indices, pourvu que la suite équivalente  $(u_k)_{k \in \mathbb{N}}$  prenne aussi la valeur zéro pour ces mêmes indices.

**Théorème 1.24** (règle des équivalents). *Soient  $(u_k)_{k \in \mathbb{N}}$  et  $(v_k)_{k \in \mathbb{N}}$  deux suites réelles équivalentes. On suppose de plus que tous les  $v_k$  sont positifs à partir d'un certain rang. Alors les séries  $\sum_{k=0}^{+\infty} u_k$  et  $\sum_{k=0}^{+\infty} v_k$  sont de même nature (si l'une converge, l'autre aussi, si l'une diverge, l'autre aussi).*

*Preuve.* Par hypothèse, il existe un entier  $k_0$  tel que pour tout  $k \geq k_0$ ,  $v_k \geq 0$ . Il existe aussi une suite  $(w_k)_{k \in \mathbb{N}}$  tendant vers 1 et un entier  $k_1$  tels que pour tout  $k \geq k_1$   $u_k = v_k w_k$ . À cause de la convergence vers 1 de la suite  $(w_k)_{k \in \mathbb{N}}$ , il existe aussi un entier  $k_2$  tel que pour tout  $k \geq k_2$ ,  $1/2 \leq w_k \leq 3/2$ . En prenant  $k_3 = \max(k_0, k_1, k_2)$ , on obtient :

$$\forall k \geq k_3, \quad 0 \leq \frac{1}{2}v_k \leq u_k \leq \frac{3}{2}v_k.$$

Comme les séries de terme général  $\frac{1}{2}v_k$  ou  $\frac{3}{2}v_k$  sont clairement de même nature que celle de terme général  $v_k$ , on en déduit que  $\sum_{k=0}^{+\infty} u_k$  et  $\sum_{k=0}^{+\infty} v_k$  sont de même nature par application du théorème de comparaison.  $\square$

Attention, dans le théorème ci-dessus, l'hypothèse de *positivité* de  $v_k$  à partir d'un certain rang<sup>6</sup> est essentielle. Sans elle le résultat peut devenir faux. Par exemple si on prend pour  $k \geq 1$ ,  $u_k = (-1)^k k^{-1/2} + k^{-1}$  et  $v_k = (-1)^k k^{-1/2}$ , on a bien  $u_k \sim (-1)^k k^{-1/2} = v_k$ , mais on peut montrer que  $\sum_{k=0}^{+\infty} u_k$  diverge, tandis que  $\sum_{k=0}^{+\infty} v_k$  converge grâce au théorème des séries alternées, voir le th. 1.35 p. 17 et l'exemple 1.36.

Voici deux exemples d'application de la règle des équivalents.

**Exemple 1.25.** La série  $\sum_{k=0}^{+\infty} \frac{2^k - 4k^5}{3^k - k^2}$  converge. En effet,

$$u_k = \frac{2^k - 4k^5}{3^k - k^2} = \frac{2^k}{3^k} \times \frac{1 - 4k^5 2^{-k}}{1 - k^2 3^{-k}} \sim \left(\frac{2}{3}\right)^k = v_k.$$

6. L'équivalence de  $u_k$  et  $v_k$  implique alors que  $u_k$  est aussi positif à partir d'un certain rang, pas forcément le même que pour  $v_k$ .

Le terme  $v_k$  est positif pour tout  $k$  et  $\sum_{k=0}^{+\infty} v_k$  est la série géométrique standard de raison  $q = 2/3$ , convergente puisque  $|q| < 1$ . Donc par la règle des équivalents,  $\sum_{k=0}^{+\infty} u_k$  converge.

**Exemple 1.26.** La série  $\sum_{k=0}^{+\infty} \sqrt{\frac{k-1}{k^2+3}}$  diverge. En effet,

$$u_k = \sqrt{\frac{k-1}{k^2+3}} = \sqrt{\frac{k}{k^2} \times \frac{1-k^{-1}}{1+3k^{-2}}} = \frac{1}{k^{1/2}} \times \sqrt{\frac{1-k^{-1}}{1+3k^{-2}}} \sim k^{-1/2} = v_k.$$

Le terme  $v_k$  est défini et positif pour tout  $k \geq 1$  et  $\sum_{k=0}^{+\infty} v_k$  diverge, cf. exemple 1.21. Donc par la règle des équivalents,  $\sum_{k=0}^{+\infty} u_k$  diverge.

Une autre conséquence importante du théorème 1.19 est le théorème de comparaison série-intégrale.

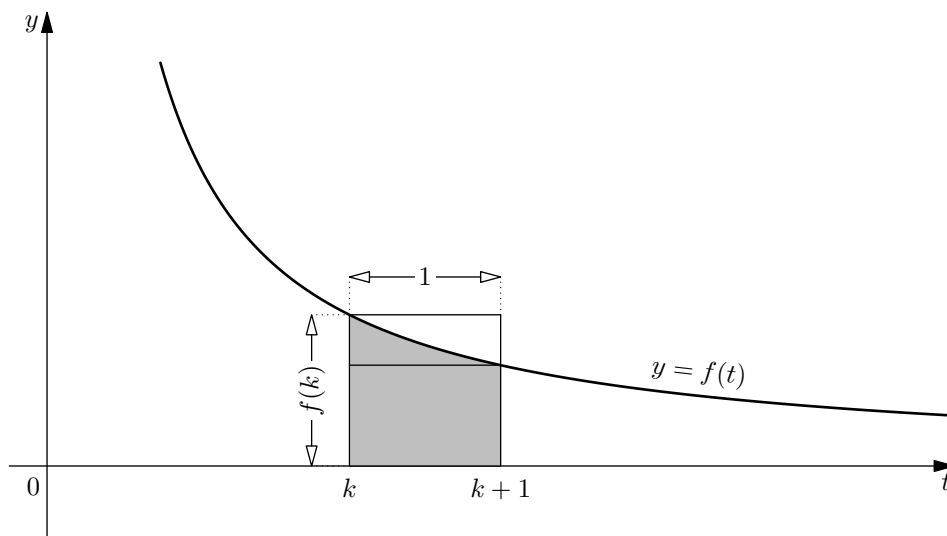


FIGURE 1.7 – Comparaison série-intégrale

**Théorème 1.27** (comparaison série-intégrale). *Soit  $f$  continue sur  $[k_0, +\infty[$ , décroissante et positive sur cet intervalle. La série  $\sum_{k=k_0}^{+\infty} f(k)$  converge si et seulement si  $\int_{k_0}^n f(t) dt$  a une limite finie quand  $n$  tend vers l'infini.*

*Preuve.* Par décroissance de  $f$ , pour tout  $k \geq k_0$  et tout  $t \in [k, k+1]$ ,  $f(k) \geq f(t) \geq f(k+1)$ . En intégrant ces inégalités sur  $[k, k+1]$ , on obtient :  $\int_k^{k+1} f(k) dt \geq \int_k^{k+1} f(t) dt \geq \int_k^{k+1} f(k+1) dt$ . La première et la troisième intégrale sont des intégrales d'une fonction constante sur un segment de longueur 1 (ce sont des aires de rectangles, cf. la figure 1.7). L'encadrement peut donc se réécrire sous la forme :

$$f(k) \geq \int_k^{k+1} f(t) dt \geq f(k+1).$$

En sommant ces inégalités depuis  $k = k_0$  jusqu'à  $k = n-1$ , les intégrales se raccordent en une intégrale sur  $[k_0, n]$  et l'on obtient :

$$\sum_{k=k_0}^{n-1} f(k) \geq \int_{k_0}^n f(t) dt \geq \sum_{k=k_0}^{n-1} f(k+1) = \sum_{j=k_0+1}^n f(j).$$

En notant  $S_n = \sum_{k=0}^n f(k)$  et  $I_n = \int_{k_0}^n f(t) dt$ , on peut réécrire cet encadrement sous la forme :

$$S_{n-1} - S_{k_0-1} \geq I_n \geq S_n - S_{k_0}$$

Comme  $f$  est positive sur  $[k_0, +\infty[$ , les suites  $(S_n)_{n \geq k_0}$  et  $(I_n)_{n \geq k_0}$  sont *croissantes*<sup>7</sup>. En notant que pour tout  $n > k_0$ ,  $S_n \geq S_{n-1}$ , on tire de l'encadrement ci-dessus les inégalités :

$$\forall n > k_0, \quad S_n \leq I_n + S_{k_0}, \quad S_n \geq S_{n-1} \geq I_n + S_{k_0-1}.$$

Il n'y a que deux possibilités pour la suite *croissante*  $(I_n)_{n \geq k_0}$  : ou bien elle converge vers une limite finie  $I$ , ou bien elle tend vers  $+\infty$ .

Dans le premier cas, la suite croissante  $(I_n)_{n \geq k_0}$  est majorée par sa limite  $I$ , donc pour tout  $n > k_0$ ,  $S_n \leq I_n + S_{k_0} \leq I + S_{k_0}$ . La suite *croissante*  $(S_n)_{n \geq k_0}$  est donc majorée par  $I + S_{k_0}$  constante indépendante de  $n$ , donc elle converge vers une limite finie  $S$  et  $S \leq I + S_{k_0}$ . Autrement dit, la série  $\sum_{k=0}^{+\infty} f(k)$  converge et  $\sum_{k=0}^{+\infty} f(k) \leq I + S_{k_0}$ .

Dans le deuxième cas,  $I_n + S_{k_0-1}$  tend vers  $+\infty$  et comme  $S_n \geq I_n + S_{k_0-1}$ ,  $S_n$  tend aussi vers  $+\infty$ , autrement dit, la série  $\sum_{k=0}^{+\infty} f(k)$  diverge.  $\square$

**Corollaire 1.28** (séries de Riemann).

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

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

*Preuve.* On applique le théorème 1.19 avec  $f(t) = t^{-a}$ ,  $a > 0$  et  $k_0 = 1$ . Un simple calcul de primitive nous donne pour  $a \neq 1$  :

$$\int_1^n \frac{dt}{t^a} = \left[ \frac{t^{-a+1}}{-a+1} \right]_1^n = \frac{n^{-a+1} - 1}{-a+1} \xrightarrow{n \rightarrow +\infty} \begin{cases} \frac{1}{a-1} & \text{si } a > 1, \\ +\infty & \text{si } a < 1. \end{cases}$$

Le cas particulier  $a = 1$  est celui de la série harmonique dont on connaît déjà la divergence. L'application du théorème permet de retrouver ce résultat puisque

$$\int_1^n \frac{dt}{t} = \left[ \ln t \right]_1^n = \ln n - \ln 1 = \ln n \xrightarrow{n \rightarrow +\infty} +\infty.$$

$\square$

**Corollaire 1.29** (séries de Bertrand).

1. La série  $\sum_{k=2}^{+\infty} \frac{1}{k^a (\ln k)^b}$  est  $\begin{cases} \text{convergente pour } a > 1, \text{ quel que soit } b \in \mathbb{R}, \\ \text{divergente pour } 0 < a < 1, \text{ quel que soit } b \in \mathbb{R}. \end{cases}$
2. La série  $\sum_{k=2}^{+\infty} \frac{1}{k (\ln k)^b}$  est  $\begin{cases} \text{convergente pour } b > 1, \\ \text{divergente pour } b \leq 1. \end{cases}$

7. Pour la croissance de  $(I_n)_{n \geq k_0}$ , on note que  $I_{n+1} - I_n = \int_n^{n+1} f(t) dt \geq 0$  puisque  $f(t) \geq 0$  pour tout  $t \in [n, n+1]$ .



*Preuve.* Pour  $a \neq 1$  (ou pour  $a = 1$  et  $b \leq 0$ ), la nature de la série de Bertrand  $\sum_{k=2}^{+\infty} k^{-a} (\ln k)^{-b}$  s'obtient directement par comparaison avec une série de Riemann (vérification laissée en exercice). Traitons le cas  $a = 1$  et  $b > 0$  en appliquant le théorème 1.19 avec  $k_0 = 2$  et

$$f(t) = \frac{1}{t(\ln t)^b}.$$

La fonction  $f$  est continue et décroissante sur  $[2, +\infty[$  et un calcul d'intégrale par changement de variable  $u = \ln t$  nous donne pour  $b \neq 1$  :

$$\int_2^n \frac{dt}{t(\ln t)^b} = \int_{\ln 2}^{\ln n} \frac{du}{u^b} = \frac{(\ln 2)^{-b+1} - (\ln n)^{-b+1}}{b-1} \xrightarrow{n \rightarrow +\infty} \begin{cases} \frac{(\ln 2)^{-b+1}}{b-1} & \text{si } b > 1, \\ +\infty & \text{si } b < 1. \end{cases}$$

Pour  $b = 1$ ,

$$\int_2^n \frac{dt}{t \ln t} = \int_{\ln 2}^{\ln n} \frac{du}{u} = \left[ \ln u \right]_{\ln 2}^{\ln n} = \ln(\ln n) - \ln(\ln 2) \xrightarrow{n \rightarrow +\infty} +\infty.$$

□

## 1.5 Séries à termes de signe variable

Nous passons maintenant à l'étude des séries qui ne sont pas à termes positifs. Pour expliquer le titre de cette section, il convient de remarquer que si le signe de  $u_k$  est *constant* à partir d'un certain rang  $k_0$ , l'étude de la convergence de  $\sum_{k=0}^{+\infty} u_k$  se ramène à celle d'une série à termes positifs. En effet,  $\sum_{k=0}^{+\infty} u_k$  et  $\sum_{k=k_0}^{+\infty} u_k$  sont de même nature et si  $u_k \leq 0$  pour  $k \geq k_0$ , il suffit de considérer  $\sum_{k=k_0}^{+\infty} (-u_k)$ . Ce que nous allons voir dans cette section n'est donc réellement nouveau que pour les séries dont le terme général change de signe une infinité de fois.

Une première idée pour l'étude des séries à terme de signe variable est d'exprimer ses sommes partielles comme différences de sommes partielles de deux séries à termes positifs. Pour cela, il suffit d'utiliser la décomposition  $u_k = u_k^+ - u_k^-$  où  $u_k^+$  est la partie positive du réel  $u_k$  et  $u_k^-$  sa partie négative, définies comme suit.

**Définition 1.30** (partie positive et partie négative d'un réel). *Soit  $x$  un nombre réel. Sa partie positive  $x^+$  et sa partie négative  $x^-$  sont les réels positifs :*

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

Il importe de noter que  $x^-$  est un réel positif. Voici deux exemples.

$$(2,35)^+ = 2,35, \quad (2,35)^- = 0, \quad (-2,35)^+ = 0, \quad (-2,35)^- = 2,35.$$

On remarque aussi que  $|x|$  est égal à  $x^+$  si  $x$  est positif et à  $x^-$  si  $x$  est négatif. Plus généralement, on a les relations suivantes :

$$\forall x \in \mathbb{R}, \quad x = x^+ - x^-, \quad |x| = x^+ + x^-.$$

Notons aussi que

$$\forall x \in \mathbb{R}, \quad 0 \leq x^+ \leq |x|, \quad 0 \leq x^- \leq |x|.$$

Revenons au cas d'une série  $\sum_{k=0}^{+\infty} u_k$  dont le terme général est de signe quelconque. Considérons les sommes partielles<sup>8</sup> :

$$S_n = \sum_{k=0}^n u_k, \quad S'_n = \sum_{k=0}^n u_k^+, \quad S''_n = \sum_{k=0}^n u_k^-.$$

On voit alors que pour tout  $n \in \mathbb{N}$ ,  $S_n = S'_n - S''_n$  et qu'une condition *suffisante* pour que  $S_n$  converge vers un réel est que chacune des deux suites  $(S'_n)_{n \in \mathbb{N}}$  et  $(S''_n)_{n \in \mathbb{N}}$  converge dans  $\mathbb{R}_+$ , autrement dit que chacune des séries à termes *positifs*  $\sum_{k=0}^{+\infty} u_k^+$  et  $\sum_{k=0}^{+\infty} u_k^-$  converge dans  $\mathbb{R}_+$ . Compte tenu des inégalités  $0 \leq u_k^+ \leq |u_k|$  et  $0 \leq u_k^- \leq |u_k|$ , le théorème de comparaison des séries à termes positifs nous permet de conclure qu'une condition suffisante pour la convergence de  $\sum_{k=0}^{+\infty} u_k^+$  et  $\sum_{k=0}^{+\infty} u_k^-$  et donc aussi de  $\sum_{k=0}^{+\infty} u_k$  est que  $\sum_{k=0}^{+\infty} |u_k|$  converge. Nous venons ainsi d'établir le résultat le plus important pour la convergence d'une série à termes de signe variable.

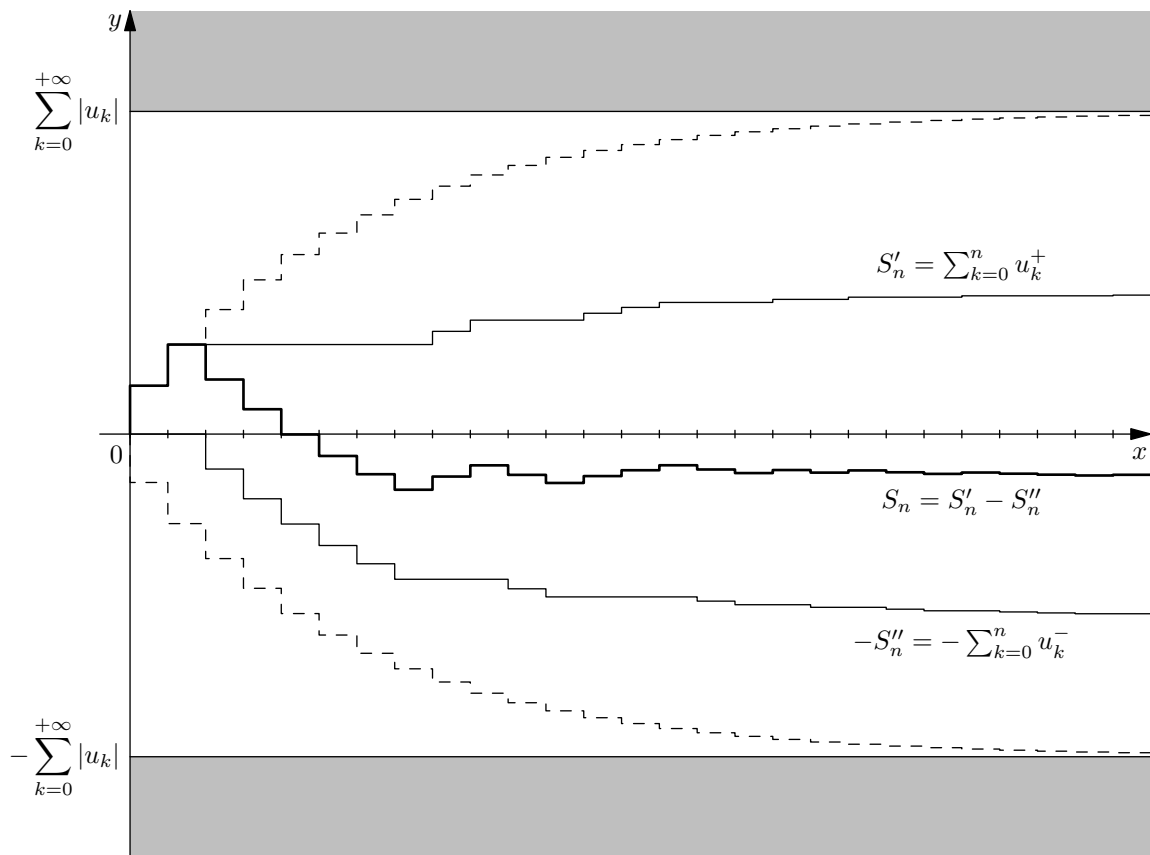


FIGURE 1.8 – La convergence de  $\sum_{k=0}^{+\infty} |u_k|$  implique celle de  $\sum_{k=0}^{+\infty} u_k$

**Théorème 1.31.** *Pour que la série  $\sum_{k=0}^{+\infty} u_k$  converge, il suffit que la série à termes positifs  $\sum_{k=0}^{+\infty} |u_k|$  converge.*

**Définition 1.32** (série absolument convergente). *La série  $\sum_{k=0}^{+\infty} u_k$  est dite absolument convergente si la série  $\sum_{k=0}^{+\infty} |u_k|$  converge.*

8. On peut naturellement se demander pourquoi on n'a pas noté  $S_n^+$  au lieu de  $S'_n$  et  $S_n^-$  au lieu de  $S''_n$ . C'est parce que  $S'_n$  et  $S''_n$  ne sont en général pas les parties positive ou négative de  $S_n$ . Pour s'en convaincre, on pourra comparer  $(a+b)^+$  et  $a^+ + b^+$  en prenant pour  $a$  et  $b$  deux réels de signes contraires.

Avec cette définition, le théorème ci-dessus peut aussi s'énoncer : « toute série absolument convergente est convergente ».

**Exemple 1.33.** La série  $\sum_{k=1}^{+\infty} \frac{\cos(k^2)}{k\sqrt{k}}$  converge.

Pour le vérifier, il suffit de montrer qu'elle est *absolument* convergente. Or  $|\cos(k^2)| \leq 1$ , donc par le théorème de comparaison des séries à termes positifs et par convergence de la série de Riemann  $\sum_{k=1}^{+\infty} k^{-3/2}$  (car  $3/2 > 1$ ), la série  $\sum_{k=1}^{+\infty} |\cos(k^2)k^{-3/2}|$  converge.

À ce stade, il est naturel de se demander s'il existe des séries convergentes mais pas absolument convergentes. Pour les séries  $\sum_{k=0}^{+\infty} u_k$  à terme général positif à partir d'un certain rang  $k_0$ ,  $u_k = |u_k|$  pour  $k \geq k_0$ . Si  $\sum_{k=0}^{+\infty} u_k$  converge, alors  $\sum_{k=k_0}^{+\infty} u_k$  converge aussi, donc aussi  $\sum_{k=k_0}^{+\infty} |u_k|$  qui est la même série. Par conséquent  $\sum_{k=0}^{+\infty} |u_k|$  converge. Ainsi pour une série à termes positifs à partir d'un certain rang, convergence et convergence absolue sont deux notions équivalentes. Le même raisonnement s'applique aussi aux séries à termes négatifs à partir d'un certain rang, en remplaçant  $u_k$  par  $-u_k$ .

Donc si nous cherchons des séries convergentes mais pas absolument convergentes, les seuls candidats admissibles sont les séries dont le terme général n'est pas « constant à partir d'un certain rang », c.-à-d. *change de signe une infinité de fois*. Par exemple, nous allons voir que la série  $\sum_{k=1}^{+\infty} (-1)^k/k$  converge mais pas absolument. C'est une application directe du théorème des *séries alternées* ci-dessous.

**Définition 1.34.** La suite  $(u_k)_{k \in \mathbb{N}}$  est dite *alternée* si pour tout  $k \in \mathbb{N}$ ,  $u_k$  et  $u_{k+1}$  sont de signes contraires.

**Théorème 1.35** (théorème des séries alternées). Soit  $(u_k)_{k \in \mathbb{N}}$  une suite alternée telle que  $(|u_k|)_{k \in \mathbb{N}}$  décroît et tend vers 0. Alors,

- i) la série  $\sum_{k=0}^{+\infty} u_k$  converge ;
- ii) sa somme  $S$  est encadrée pour tout  $n \in \mathbb{N}$  par les deux sommes partielles consécutives  $S_n$  et  $S_{n+1}$  ;
- iii) le reste  $R_n = \sum_{k=n+1}^{+\infty} u_k$  vérifie pour tout  $n \in \mathbb{N}$  :

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

*Preuve.*

*Preuve de i).* Posons  $u'_k = |u_k|$ . Alors  $u'_k \geq 0$  et  $u'_k$  tend vers 0 en décroissant. Quitte à multiplier par  $-1$  tous les termes de la série, on peut toujours supposer que  $u_0$  est positif ( $u_0 \geq 0$ ) et donc par alternance des signes que tous les termes d'indice impair sont négatifs ( $u_{2j+1} \leq 0$  pour tout  $j \in \mathbb{N}$ ) et tous les termes d'indice pair sont positifs ( $u_{2j} \geq 0$  pour tout  $j \in \mathbb{N}$ ). La somme partielle de rang  $n$  s'écrit alors :

$$S_n = \sum_{k=0}^n u_k = \sum_{k=0}^n (-1)^k u'_k.$$

Pour tout  $j \in \mathbb{N}$ ,

$$S_{2j+1} = (u'_0 - u'_1) + (u'_2 - u'_3) + \cdots + (u'_{2j} - u'_{2j+1}) = \sum_{l=0}^j (u'_{2l} - u'_{2l+1}).$$

Par décroissance de  $u'_k$ , chacune des différences entre parenthèses ci-dessus est positive.

En utilisant la positivité de  $(u'_{2j} - u'_{2j+1})$ , on voit immédiatement que pour tout  $j \geq 1$ ,

$$S_{2j+1} = S_{2j-1} + (u'_{2j} - u'_{2j+1}) \geq S_{2j-1}.$$

On en déduit que la suite  $(S_{2j+1})_{j \in \mathbb{N}}$  est *croissante*.

D'autre part,

$$S_{2j+1} = u'_0 - (u'_1 - u'_2) - (u'_3 - u'_4) - \cdots - (u'_{2j-1} - u'_{2j}) - u'_{2j+1} \leq u'_0,$$

puisque toutes les parenthèses *retranchées* à  $u'_0$  et le terme  $u'_{2j+1}$  sont positifs. Ceci étant valable pour tout  $j \in \mathbb{N}$ , on en déduit que la suite  $(S_{2j+1})_{j \in \mathbb{N}}$  est *majorée* par le réel  $u'_0$ .

En tant que suite croissante et majorée par  $u'_0$ , la suite  $(S_{2j+1})_{j \in \mathbb{N}}$  converge dans  $\mathbb{R}$  vers une limite  $S$  et

$$\forall j \in \mathbb{N}, \quad S_{2j+1} \leq S \leq u'_0.$$

De plus en notant que pour tout  $j \geq 1$ ,  $S_{2j} = S_{2j-1} + u_{2j}$  et en rappelant que  $u_{2j}$  tend vers 0 quand  $j$  tend vers l'infini, on en déduit que  $(S_{2j})_{j \in \mathbb{N}}$  converge aussi vers  $S$ .

Ceci termine la preuve de i) puisque la convergence vers  $S$  des deux sous-suites « complémentaires »  $(S_{2j+1})_{j \in \mathbb{N}}$  et  $(S_{2j})_{j \in \mathbb{N}}$  implique celle de la suite  $(S_n)_{n \in \mathbb{N}}$  vers la même limite  $S$ . Autrement dit, la série  $\sum_{k=0}^{+\infty} u_k$  converge et a pour somme  $S$ .

*Preuve de ii).* En écrivant

$$S_{2j} = u'_0 - (u'_1 - u'_2) - (u'_3 - u'_4) - \cdots - (u'_{2j-1} - u'_{2j})$$

et en rappelant la positivité de chacune des parenthèses retranchées à  $u'_0$ , on voit que la suite  $(S_{2j})_{j \in \mathbb{N}}$  est décroissante. Comme on sait déjà qu'elle converge, elle est minorée par sa limite  $S$ . Donc, pour tout  $j \in \mathbb{N}$ ,

$$S_{2j+1} \leq S \leq S_{2j},$$

ce qui établit l'encadrement de la somme de la série  $\sum_{k=0}^{+\infty} u_k$  par deux sommes partielles consécutives de rang quelconque<sup>9</sup>.

*Preuve de iii).* De l'encadrement de  $S$  par les sommes partielles consécutives  $S_{2j}$  et  $S_{2j+1}$ , on déduit la majoration :

$$|R_{2j}| = |S_{2j} - S| \leq |S_{2j} - S_{2j+1}| = |u_{2j+1}|.$$

Comme  $j$  est quelconque, ceci montre que pour  $n$  pair  $|R_n| \leq |u_{n+1}|$ .

Comme on dispose aussi de l'encadrement  $S_{2j+1} \leq S \leq S_{2j+2}$ , on en déduit la majoration :

$$|R_{2j+1}| = |S - S_{2j+1}| \leq |S_{2j+1} - S_{2j+2}| = |u_{2j+2}|,$$

valable pour tout entier  $j$ , ce qui montre que pour  $n$  impair  $|R_n| \leq |u_{n+1}|$  et achève la preuve du théorème.  $\square$

**Exemple 1.36.** Pour  $0 < a \leq 1$ , la série  $\sum_{k=1}^{+\infty} (-1)^k k^{-a}$  est convergente (th. 1.35) mais pas absolument (cor. 1.28). Elle converge absolument pour  $a \geq 1$  (cor. 1.28).

9. Notons au passage que si l'on avait pris  $u_0 < 0$ , les rôles de  $S_{2j}$  et  $S_{2j+1}$  seraient échangés et l'encadrement ci-dessus deviendrait  $S_{2j} \leq S \leq S_{2j+1}$ .

**Exemple 1.37.** La série  $\sum_{k=2}^{+\infty} \frac{(-1)^k}{\ln k}$  converge (th. 1.35). On peut voir qu'elle ne converge pas absolument en appliquant le théorème de comparaison, par exemple avec  $\ln k$  et  $\sqrt{k}$ .

**Remarque 1.38.** Même dans le cas où la série alternée est absolument convergente, la propriété d'encadrement de sa somme par deux sommes partielles consécutives (th. 1.35-ii)) est intéressante pour le calcul numérique de  $S$ .

Nous donnons maintenant un théorème qui contient le théorème des séries alternées comme cas particulier et que nous admettrons.

**Théorème 1.39** (théorème d'Abel). *Soit  $(u_k)_{k \in \mathbb{N}}$  une suite dont le terme général est de la forme :*

$$u_k = a_k b_k,$$

où

- i) la suite  $(a_k)_{k \in \mathbb{N}}$  est décroissante et tend vers 0 ;
- ii) les sommes de termes consécutifs de  $(b_k)_{k \in \mathbb{N}}$  sont bornées, autrement dit, il existe une constante  $M > 0$  telle que pour tous entiers  $m$  et  $n$  ( $m \leq n$ ),

$$\left| \sum_{k=m}^n b_k \right| \leq M.$$

Alors la série  $\sum_{k=0}^{+\infty} u_k$  converge et son reste  $R_n = \sum_{k=n+1}^{+\infty} u_k$  vérifie pour tout  $n \in \mathbb{N}$ ,

$$|R_n| \leq M a_{n+1}.$$

Le théorème d'Abel a une application importante aux séries trigonométriques.

**Corollaire 1.40.** *Si  $(a_k)_{k \in \mathbb{N}}$  est une suite décroissante et convergente vers 0, alors pour tout réel  $\theta$  n'appartenant pas à  $2\pi\mathbb{Z}$ , les séries*

$$\sum_{k=0}^{+\infty} a_k \cos(k\theta) \quad \text{et} \quad \sum_{k=0}^{+\infty} a_k \sin(k\theta)$$

convergent.

*Preuve.* Il suffit de vérifier l'hypothèse ii) du théorème d'Abel avec  $b_k = \cos(k\theta)$  et  $b_k = \sin(k\theta)$ . Pour  $0 \leq m \leq n$  entiers quelconques, posons :

$$T_{m,n} = \sum_{k=m}^n e^{ik\theta},$$

en rappelant que  $e^{ik\theta} = \cos(k\theta) + i \sin(k\theta)$ . Comme

$$\left| \sum_{k=m}^n \cos(k\theta) \right| = |\operatorname{Re}(T_{m,n})| \leq |T_{m,n}| \quad \text{et} \quad \left| \sum_{k=m}^n \sin(k\theta) \right| = |\operatorname{Im}(T_{m,n})| \leq |T_{m,n}|,$$

il suffit de montrer que  $T_{m,n}$  est bornée dans  $\mathbb{C}$  indépendamment de  $m$  et  $n$ . Or  $T_{m,n}$  est une somme de termes consécutifs d'une suite géométrique de raison  $e^{i\theta}$ , qui se calcule facilement (cf. section 1.1) :

$$T_{m,n} = e^{im\theta} (1 + e^{i\theta} + e^{2i\theta} + \dots + e^{i(n-m)\theta}) = e^{im\theta} \frac{1 - e^{i(n-m+1)\theta}}{1 - e^{i\theta}} \quad \text{si } e^{i\theta} \neq 1.$$

On en déduit que

$$\text{si } e^{i\theta} \neq 1, \quad |T_{m,n}| \leq \frac{2}{|1 - e^{i\theta}|},$$

ce qui est bien une constante indépendante de  $m$  et  $n$ . La condition  $e^{i\theta} \neq 1$ , signifie que l'affixe du nombre complexe  $e^{i\theta}$  peut être n'importe quel point du cercle trigonométrique, sauf le point d'affixe 1. Cela signifie que  $\theta$  ne doit pas être de la forme  $2k\pi$  pour  $k \in \mathbb{Z}$ , autrement dit,  $\theta \notin 2\pi\mathbb{Z}$ .  $\square$

## 1.6 Compléments sur les séries absolument convergentes

Lorsque l'on additionne un nombre fini de réels, les règles de l'algèbre nous permettent de permuter l'ordre des termes ou de les grouper comme on le souhaite, sans que cela ne change le résultat final. Peut-on généraliser ceci à la somme d'un ensemble infini de réels, c'est-à-dire à une série? La réponse est non en général, mais oui dans le cas des séries *absolument* convergentes. Nous donnons sans démonstration le théorème correspondant.

**Théorème 1.41.** *Soit  $\sum_{k=0}^{+\infty} u_k$  une série absolument convergente.*

1. *Elle reste absolument convergente et de même somme pour toute permutation de l'ordre des termes :*

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

*pour toute bijection  $f : \mathbb{N} \rightarrow \mathbb{N}$ .*

2. *Elle est sommable par paquets : pour toute décomposition  $\mathbb{N} = \bigcup_{j \in J} I_j$  en sous-ensembles non vides et deux à deux disjoints,*

$$\sum_{k=0}^{+\infty} u_k = \sum_{j \in J} \left( \sum_{k \in I_j} u_k \right).$$

**Remarque 1.42.** La série harmonique alternée  $\sum_{k=1}^{+\infty} (-1)^k/k$  est convergente mais pas absolument. Avec une permutation judicieusement choisie de l'ordre des termes, on peut s'amuser à faire converger ses sommes partielles vers n'importe quel nombre réel choisi à l'avance, ou à les faire osciller de façon qu'elles passent une infinité de fois sous  $a$  et une infinité de fois au-dessus de  $b$  pour tous réels  $a < b$  fixés à l'avance, on peut aussi les faire tendre vers  $+\infty$  ou vers  $-\infty$ , etc. La démonstration précise de cette affirmation est un peu longue à écrire, mais l'idée sous-jacente est très simple. On imagine deux puits de profondeur infinie, le puits gauche et le puits droit. Dans le puits gauche, on range tous les termes d'indice impair de la série, la profondeur croissant avec l'indice. On aura donc du haut vers le bas :  $-1, -1/3, -1/5, -1/7, \dots$ . Dans le puits droit on range de même les termes de rang pair :  $1/2, 1/4, 1/6, \dots$ . Essayons maintenant de construire une bijection  $f$  de  $\mathbb{N}$  dans  $\mathbb{N}$  de façon que les sommes partielles  $T_n = \sum_{k=1}^n u_{f(k)}$  oscillent indéfiniment entre disons  $-3$  et  $+7$ . Le point clé est le suivant : comme chacune des séries  $\sum_{j=0}^{+\infty} u_{2j-1}$  et  $\sum_{j=1}^{+\infty} u_{2j}$  divergent, la première vers  $-\infty$  et la deuxième vers  $+\infty$ , chaque fois que l'on retranche les termes jusqu'à un certain rang, la série restante diverge de la même façon. Autrement dit, quel que soit le nombre de termes déjà puisé dans chacun des deux puits, il en reste suffisamment pour bâtir un déplacement vers la gauche (si on puise dans le puits

gauche) ou vers la droite (si on puise dans le puits droit), d'une longueur supérieure à tout nombre que l'on a choisi à l'avance. L'algorithme est alors très simple. On commence par puiser à gauche suffisamment de termes pour que leur somme  $s_1$  passe au dessous de  $-3$ . Puis on puise à droite suffisamment de termes pour que leur somme  $s_2$  vérifie  $s_1 + s_2 > 10$  (donc  $s_2 > 10 - s_1$ ). Ensuite on puise à gauche suffisamment de termes pour construire une somme  $s_3$  qui nous ramène à gauche de  $-3$  :  $s_1 + s_2 + s_3 < -3$ , etc. On construit ainsi  $f$  proche en proche en écrivant les indices  $f(k)$  dans l'ordre de leur puisage :  $f(k)$  est l'indice du  $k^e$  terme de la série puisé par cet algorithme. Il est clair qu'avec cette construction,  $T_n = \sum_{k=1}^n u_{f(k)}$  ne peut converger.

**Remarque 1.43.** Il importe de comprendre que dans la propriété ii) de sommation par paquets au théorème 1.41, la taille des paquets  $\sum_{k \in I_j}$  peut être variable et éventuellement infinie. Si on ne considère que des paquets de taille finie fixe, alors la propriété est vraie pour toute série convergente, même non absolument.

**Théorème 1.44** (sommation par paquets de taille fixe). *Soit  $(u_n)_{n \in \mathbb{N}}$  une suite convergente vers 0. Fixons un entier  $p \geq 2$  et définissons les « paquets de  $p$  termes consécutifs » :*

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

Alors les séries  $\sum_{k=0}^{+\infty} u_k$  et  $\sum_{j=0}^{+\infty} v_j$  sont ou toutes deux convergentes ou toutes deux divergentes et ont même somme lorsqu'elles convergent.

*Preuve.* Posons

$$S_n = \sum_{k=0}^n u_k, \quad T_n = \sum_{j=0}^n v_j.$$

De la convergence vers 0 de la suite  $(u_n)_{n \in \mathbb{N}}$  on déduit facilement (vérification détaillée laissée en exercice) que

- i)  $\lim_{n \rightarrow +\infty} |S_n - T_{[n/p]}| = 0$ ,  $[n/p]$  désignant le quotient dans la division euclidienne de  $n$  par  $p$ ;
- ii)  $\lim_{n \rightarrow +\infty} |T_n - S_{np}| = 0$ .

Grâce à ces deux convergences, on peut dire que :

- si  $S_n$  converge vers un réel  $S$ , alors  $S_{np}$  converge aussi vers  $S$  et par ii),  $T_n$  converge vers  $S$ ;
- si  $T_n$  converge vers un réel  $T$ , alors  $T_{[n/p]}$  converge aussi vers  $T$  et par i),  $S_n$  converge vers  $T$ .

Donc  $S_n$  converge si et seulement si  $T_n$  converge et en cas de convergence les deux suites ont même limite.  $\square$

**Remarque 1.45.** Le groupement de termes par paquets de taille fixe peut parfois changer une série convergente mais pas absolument, en série absolument convergente. En voici deux exemples.

**Exemple 1.46.** Groupons deux par deux les termes de la série  $\sum_{k=0}^{+\infty} (-1)^k / (k+1)$  :

$$v_1 = u_0 + u_1 = 1 - \frac{1}{2} = \frac{1}{2}, \dots, v_j = u_{2j} + u_{2j+1} = \frac{1}{2j+1} - \frac{1}{2j+2} = \frac{1}{(2j+1)(2j+2)}$$

La série de terme général  $u_k = (-1)^k / (k+1)$  converge par le théorème sur les séries alternées. La série de terme général  $v_j$  obtenue en groupant deux par deux les termes de

$\sum_{k=0}^{+\infty} u_k$  est donc convergente et de même somme par le théorème 1.44. Comme elle est à termes positifs, elle est aussi absolument convergente. La positivité de  $v_j$  permet d'ailleurs de vérifier directement la convergence puisque  $v_j \sim 1/(4j^2)$ . L'égalité des sommes des deux séries s'écrit ici :

$$\sum_{k=0}^{+\infty} \frac{(-1)^{k+1}}{k+1} = \sum_{j=0}^{+\infty} \frac{1}{(2j+1)(2j+2)}.$$

**Exemple 1.47.** Considérons la série de terme général

$$u_k = \frac{(-1)^k}{k + (-1)^k}, \quad k \geq 2.$$

La suite  $(u_k)_{k \geq 2}$  est alternée, mais on ne peut pas appliquer le théorème des séries alternées car  $(|u_k|)_{k \geq 2}$  n'est pas décroissante. Groupons les termes deux par deux en posant pour tout  $j \geq 1$ ,

$$v_j = u_{2j} + u_{2j+1} = \frac{1}{2j+1} + \frac{-1}{2j} = \frac{-1}{2j(2j+1)}$$

La série  $\sum_{j=1}^{+\infty} v_j$  est absolument convergente puisque  $|v_j| \sim 1/(4j^2)$ , donc elle converge. Par le théorème 1.44,  $\sum_{k=2}^{+\infty} u_k$  converge aussi et a même somme :

$$\sum_{k=2}^{+\infty} \frac{(-1)^k}{k + (-1)^k} = - \sum_{j=1}^{+\infty} \frac{1}{2j(2j+1)}.$$

Lorsque l'on effectue le produit de deux sommes d'un nombre fini de termes, on peut l'écrire comme la somme de tous les produits obtenus en choisissant un facteur dans la première somme et un facteur dans la deuxième. Il y a une formule analogue pour les séries absolument convergentes, mais elle fait appel à la notion de série double qui n'est pas au programme de ce cours. Nous nous contenterons d'une formule de développement du produit de deux séries qui ressemble formellement au produit de deux polynômes.

**Théorème 1.48** (produit de deux séries absolument convergentes). *Soient  $\sum_{j=0}^{+\infty} u_j$  et  $\sum_{k=0}^{+\infty} v_k$  deux séries absolument convergentes. Alors la série de terme général*

$$w_\ell = \sum_{j=0}^{\ell} u_j v_{\ell-j} = \sum_{j+k=\ell} u_j v_k$$

*est absolument convergente et*

$$\left( \sum_{j=0}^{+\infty} u_j \right) \times \left( \sum_{k=0}^{+\infty} v_k \right) = \sum_{\ell=0}^{+\infty} w_\ell = \sum_{\ell=0}^{+\infty} \left( \sum_{j+k=\ell} u_j v_k \right).$$

**Exemple 1.49.** On sait que la série  $\sum_{j=0}^{+\infty} \frac{a^j}{j!}$  est absolument convergente pour tout réel  $a$ . En appliquant le théorème 1.48 et modulo un petit calcul laissé en exercice, on peut en déduire que pour tous réels  $a$  et  $b$  :

$$\left( \sum_{j=0}^{+\infty} \frac{a^j}{j!} \right) \times \left( \sum_{k=0}^{+\infty} \frac{b^k}{k!} \right) = \sum_{\ell=0}^{+\infty} \frac{(a+b)^\ell}{\ell!}.$$



# Chapitre 2

## Séries entières

On s'intéresse dans ce chapitre à des séries de fonctions de la forme

$$\sum_{k=0}^{+\infty} a_k z^k, \quad z \in \mathbb{C},$$

où  $(a_k)$  est une suite de nombres complexes. Une telle série est appelée *série entière*. Comme sous-produit de leur étude on obtient des résultats valables pour les séries de la forme

$$\sum_{k=0}^{+\infty} a_k x^k, \quad x \in \mathbb{R},$$

où  $(a_k)$  est une suite de nombres réels, appelées séries entières réelles.

Tout polynôme  $\sum_{k=0}^n a_k z^k$  est un cas particulier de série entière. Lorsqu'une série entière converge, sa somme est limite d'une suite de polynômes. Nous nous demanderons dans quelle mesure les séries entières héritent des propriétés des polynômes.

### 2.1 Préliminaires

Avant d'entreprendre cette étude des séries entières, il nous faut préciser la notion de convergence pour des suites de nombres complexes et donner quelques explications sur la convergence d'une suite de fonctions.

#### 2.1.1 Suites et séries dans $\mathbb{C}$

Tout nombre complexe  $z$  peut s'écrire de manière unique sous la forme

$$z = x + iy, \quad x \in \mathbb{R}, \quad y \in \mathbb{R}.$$

Les réels  $x$  et  $y$  sont appelés respectivement *partie réelle* de  $z$ , notation  $x = \operatorname{Re}(z)$  et *partie imaginaire* de  $z$ , notation  $y = \operatorname{Im}(z)$ . Le module de  $z$  noté  $|z|$  est le réel positif défini par

$$|z| = \sqrt{x^2 + y^2}.$$

Notons les inégalités entre valeur absolue des parties réelle ou imaginaire et module :

$$|x| = |\operatorname{Re}(z)| \leq |z|, \quad |y| = |\operatorname{Im}(z)| \leq |z|.$$

La définition de la convergence d'une suite  $(z_n)_{n \in \mathbb{N}}$  de nombres complexes est la même que celle d'une suite de réels, en remplaçant les distances exprimées comme valeurs absolues de différences par les distances exprimées comme modules de différences.

**Définition 2.1** (convergence d'une suite de nombres complexes). *La suite  $(z_n)_{n \in \mathbb{N}}$  de nombres complexes converge vers une limite  $L \in \mathbb{C}$  si :*

$$\forall \varepsilon > 0, \exists n_0 \in \mathbb{N}, \forall n \geq n_0, |z_n - L| < \varepsilon. \quad (2.1)$$

Autrement dit, quelque soit le niveau de précision  $\varepsilon$  choisi, à partir d'un certain rang  $n_0$  (dépendant de  $\varepsilon$ ), tous les termes de la suite vont se trouver dans le même disque de centre  $L$  et de rayon  $\varepsilon$ .

**Remarque 2.2.** Dans le cas particulier où le nombre complexe  $L$  vaut 0, (2.1) s'écrit :

$$\forall \varepsilon > 0, \exists n_0 \in \mathbb{N}, \forall n \geq n_0, |z_n - 0| = |z_n| < \varepsilon.$$

C'est exactement la définition de la convergence dans  $\mathbb{R}$  de la suite de réels positifs  $(|z_n|)_{n \in \mathbb{N}}$ . Il y a donc équivalence entre la convergence de  $z_n$  vers 0 (dans  $\mathbb{C}$ ) et celle de  $|z_n|$  vers 0 (dans  $\mathbb{R}_+$ ).

Voici deux façons de ramener une convergence dans  $\mathbb{C}$  à de la convergence dans  $\mathbb{R}$ .

**Proposition 2.3.** *Pour la convergence dans  $\mathbb{C}$  d'une suite  $(z_n)_{n \in \mathbb{N}}$ , on a les équivalences suivantes :*

- a)  $z_n$  converge vers  $L$  dans  $\mathbb{C}$  si et seulement si  $|z_n - L|$  converge vers 0 dans  $\mathbb{R}$ .
- b)  $z_n$  converge vers  $L$  dans  $\mathbb{C}$  si et seulement si

$$x_n = \operatorname{Re}(z_n) \xrightarrow{n \rightarrow +\infty} \operatorname{Re}(L), \quad \text{et} \quad y_n = \operatorname{Im}(z_n) \xrightarrow{n \rightarrow +\infty} \operatorname{Im}(L).$$

*Preuve.*

- a) On généralise immédiatement l'argument donné ci-dessus pour la convergence vers 0 : en posant  $u_n = |z_n - L|$  et en remplaçant  $|z_n - L|$  par  $u_n$  dans (2.1), on obtient la convergence de  $u_n$  vers 0.
- b) On remarque que

$$z_n - L = (x_n + iy_n) - (\operatorname{Re}(L) + i \operatorname{Im}(L)) = (x_n - \operatorname{Re}(L)) + i(y_n - \operatorname{Im}(L))$$

d'où

$$|z_n - L| = \sqrt{(x_n - \operatorname{Re}(L))^2 + (y_n - \operatorname{Im}(L))^2},$$

ce qui montre que la convergence  $x_n$  vers  $\operatorname{Re}(L)$  et celle de  $y_n$  vers  $\operatorname{Im}(L)$  impliquent celle de  $|z_n - L|$  vers 0 dans  $\mathbb{R}$  et donc celle de  $z_n$  vers  $L$  dans  $\mathbb{C}$  d'après le a).

Pour la réciproque, on note que

$$|x_n - \operatorname{Re}(L)| \leq |z_n - L| \quad \text{et} \quad |y_n - \operatorname{Im}(L)| \leq |z_n - L|.$$

Par conséquent la convergence de  $z_n$  vers  $L$  dans  $\mathbb{C}$  implique les convergences vers 0 dans  $\mathbb{R}$  de  $x_n - \operatorname{Re}(L)$  et de  $y_n - \operatorname{Im}(L)$ , donc la convergence de  $x_n$  vers  $\operatorname{Re}(L)$  et celle de  $y_n$  vers  $\operatorname{Im}(L)$ . □

Nous pouvons maintenant généraliser la notion de série au cas des termes complexes.

**Définition 2.4** (série à termes complexes). Soit  $(z_n)_{n \in \mathbb{N}}$  une suite dans  $\mathbb{C}$ . Notons  $(S_n)_{n \in \mathbb{N}}$  la suite associée de ses sommes partielles :

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

On dit que la série de terme général  $z_n$  converge et a pour somme le nombre complexe  $S$  si  $(S_n)_{n \in \mathbb{N}}$  converge vers  $S$  dans  $\mathbb{C}$ . Dans ce cas on note :

$$S = \sum_{k=0}^{+\infty} z_k.$$

Comme dans le cas des séries à termes réels, on utilisera aussi la notation formelle  $\sum_{k=0}^{+\infty} z_k$  pour désigner la série sans préjuger de sa convergence.

On peut ramener l'étude de la convergence d'une série à termes complexes à celle de deux séries à termes réels.

**Proposition 2.5.** La série de terme général  $w_k = u_k + iv_k$  ( $u_k, v_k \in \mathbb{R}$ ) converge dans  $\mathbb{C}$  si et seulement si  $\sum_{k=0}^{+\infty} u_k$  et  $\sum_{k=0}^{+\infty} v_k$  convergent dans  $\mathbb{R}$  et dans ce cas,

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

*Preuve.* Il suffit d'appliquer la proposition 2.3 à la suite complexe de terme général  $S_n = \sum_{k=0}^n w_k$  en notant que  $\operatorname{Re}(S_n) = \sum_{k=0}^n u_k$  et  $\operatorname{Im}(S_n) = \sum_{k=0}^n v_k$ .  $\square$

**Corollaire 2.6.** Soit  $(w_k)_{k \in \mathbb{N}}$  une suite de nombres complexes. Si la série des modules  $\sum_{k=0}^{+\infty} |w_k|$  converge dans  $\mathbb{R}_+$ , la série  $\sum_{k=0}^{+\infty} w_k$  converge dans  $\mathbb{C}$ . On dit qu'elle est absolument convergente.

*Preuve.* Avec les notations de la proposition 2.5,  $|u_k| \leq |w_k|$  et  $|v_k| \leq |w_k|$  pour tout entier  $k$ . On en déduit par le théorème de comparaison des séries à termes positifs (th. 1.19, p. 10) que la convergence de la série des modules  $|w_k|$  implique la convergence des séries à termes réels positifs  $\sum_{k=0}^{+\infty} |u_k|$  et  $\sum_{k=0}^{+\infty} |v_k|$  et donc la convergence (absolue) des séries à termes réels  $\sum_{k=0}^{+\infty} u_k$  et  $\sum_{k=0}^{+\infty} v_k$ . Par la proposition 2.5, ceci entraîne la convergence dans  $\mathbb{C}$  de  $\sum_{k=0}^{+\infty} w_k$ .  $\square$

**Remarque 2.7.** Dans ce chapitre, nous allons étudier des séries à termes complexes de la forme  $w_k = a_k z^k$ , pour  $z$  complexe. Sous cette forme, la réduction à l'étude de la série des parties réelles et de celle des parties imaginaires n'est généralement d'aucune aide. Pour vous en convaincre, calculez en fonction des réels  $x$  et  $y$  les parties réelle et imaginaire de  $(1 + 3i)(x + iy)^2$ . Si l'on veut une expression commode des parties réelle et imaginaire de  $z^k$ , il convient de passer par l'écriture exponentielle de  $z$  sous la forme  $z = re^{i\theta}$  ( $r \geq 0$ ) car alors  $z^k = r^k e^{ik\theta} = r^k (\cos(k\theta) + i \sin(k\theta))$ .

**Proposition 2.8.** Si  $z_n$  converge vers  $L$  dans  $\mathbb{C}$ , alors  $|z_n|$  converge vers  $|L|$  dans  $\mathbb{R}_+$  (réciproque fautive).

*Preuve.* En appliquant l'inégalité triangulaire pour les modules, on obtient en particulier  $||z_n| - |L|| < |z_n - L|$ . En supposant que  $z_n$  converge vers  $L$  et en implantant cette inégalité dans (2.1), on obtient :

$$\forall \varepsilon > 0, \exists n_0 \in \mathbb{N}, \forall n \geq n_0, \quad ||z_n| - |L|| \leq |z_n - L| < \varepsilon,$$

ce qui montre que  $|z_n|$  converge vers  $|L|$ . Pour voir que la réciproque est fautive, le contre exemple le plus simple est certainement  $z_k = (-1)^k L$ .  $\square$

**Remarque 2.9.** En particulier si une suite  $(z_n)_{n \in \mathbb{N}}$  est telle que  $|z_n|$  tende vers l'infini (ou plus généralement ne soit pas borné), alors  $z_n$  ne peut converger dans  $\mathbb{C}$ .

**Exemple 2.10** (série géométrique dans  $\mathbb{C}$ ). La série géométrique complexe standard  $\sum_{k=0}^{+\infty} z^k$ , où  $z \in \mathbb{C}$ , converge si et seulement si  $|z| < 1$ . Dans ce cas elle est aussi absolument convergente et sa somme vaut :

$$\sum_{k=0}^{+\infty} z^k = \frac{1}{1-z}, \quad \text{pour tout } z \text{ tel que } |z| < 1.$$

Pour le vérifier, rappelons que le calcul de la somme partielle de rang  $n$  d'une suite géométrique (cf. prop. 1.3) reste valide pour une suite géométrique à valeurs complexes. On a donc

$$S_n = \sum_{k=0}^n z^k = \frac{1-z^{n+1}}{1-z} \text{ si } z \neq 1, \quad S_n = n+1 \text{ si } z=1.$$

Pour la convergence de  $S_n$ , on distingue deux cas.

1.  $|z| < 1$ . Alors  $|z|^{n+1}$  tend vers 0, autrement dit  $|z^{n+1}|$  tend vers 0, donc  $z^{n+1}$  tend vers 0 par la remarque 2.2. Par conséquent

$$S_n = \frac{1-z^{n+1}}{1-z} \xrightarrow{n \rightarrow +\infty} \frac{1}{1-z}.$$

2.  $|z| \geq 1$ . Alors  $|z^n| = |z|^n \geq 1$  pour tout  $n$ . Par conséquent  $z^n = S_n - S_{n-1}$  ne peut tendre vers 0 et ceci empêche, comme dans le cas réel, la convergence de la série.

## 2.1.2 Fonctions d'une variable complexe

Comme pour la limite d'une suite, la définition de la limite d'une fonction de variable complexe à valeurs complexes s'obtient à partir du cas réel en remplaçant simplement valeur absolue par module.

**Définition 2.11.** Soit  $D$  une partie de  $\mathbb{C}$  et  $f : D \rightarrow \mathbb{C}$  une fonction définie sur  $D$  et à valeurs dans  $\mathbb{C}$ . On dit que  $f(z)$  tend vers  $L \in \mathbb{C}$  quand  $z$  tend vers  $z_0$  dans  $D$  si :

$$\forall \varepsilon > 0, \exists \delta > 0, \forall z \in D \text{ tel que } |z - z_0| < \delta, \quad |f(z) - L| < \varepsilon. \quad (2.2)$$

Les propriétés « algébriques » des limites des fonctions d'une variable réelle (somme, différence, produit, composition, ...) restent vraies pour les fonctions d'une variable complexe, à l'exception notable de celles qui font intervenir la relation d'ordre. Nous l'admettrons.

Disposant maintenant de la notion de limite, on peut généraliser les notions de continuité et de dérivabilité.

**Définition 2.12** (continuité). Soit  $D$  une partie de  $\mathbb{C}$  et  $f : D \rightarrow \mathbb{C}$  une fonction définie sur  $D$  et à valeurs dans  $\mathbb{C}$ .

- a) On dit que  $f$  est continue au point  $z_0 \in D$  si, quand  $z$  tend vers  $z_0$  dans  $D$ ,  $f(z)$  tend vers  $f(z_0)$ .
- b) Soit  $E \subset D$ . On dit que  $f$  est continue sur  $E$  si elle est continue en tout point  $z_0$  de  $E$ .

Pour définir la dérivabilité au point  $z_0$ , il nous faudra être un peu plus restrictif en ne considérant que le cas des points  $z_0$  « intérieurs » à  $D$ , c'est à dire tels que  $D$  contienne un disque  $\Delta(z_0, r) = \{z \in \mathbb{C} ; |z - z_0| < r\}$  de rayon  $r > 0$ . Ceci exclut<sup>1</sup> notamment les points de la « frontière » de  $D$ .

**Définition 2.13** (dérivabilité). *Soit  $D$  une partie de  $\mathbb{C}$  et  $f : D \rightarrow \mathbb{C}$  une fonction définie sur  $D$  et à valeurs dans  $\mathbb{C}$ .*

a) *On dit que  $f$  est dérivable au point  $z_0$  intérieur à  $D$  et on note  $f'(z_0)$  sa dérivée en ce point si*

$$\lim_{z \rightarrow z_0} \frac{f(z) - f(z_0)}{z - z_0} = f'(z_0).$$

b) *Soit  $E \subset D$  tel que tous les points de  $E$  soient intérieurs à  $D$ . On dit que  $f$  est dérivable sur  $E$  si elle est dérivable en tout point  $z_0$  de  $E$ .*

Les règles « algébriques » de calcul des dérivées (somme, combinaison linéaire, produit, quotient, composée) restent valables pour les fonctions  $D \rightarrow \mathbb{C}$ , par héritage des règles sur les calculs de limite. En particulier, un polynôme à variable complexe se dérive exactement comme un polynôme à variable réelle :

$$\forall z \in \mathbb{C}, \quad (c_d z^d + \dots + c_j z^j + \dots + c_1 z + c_0)' = d c_d z^{d-1} + \dots + j c_j z^{j-1} + \dots + c_1. \quad (2.3)$$

### 2.1.3 Suites et séries de fonctions d'une variable complexe

Considérons maintenant une suite  $(f_n)_{n \in \mathbb{N}}$  de fonctions  $D \rightarrow \mathbb{C}$  et supposons que  $f_n$  converge vers une fonction  $f$  sur  $D$ . La notion la plus simple de convergence à envisager ici est définie par «  $f_n(z)$  tend vers  $f(z)$  pour tout  $z \in D$  ». C'est ce que l'on appelle la *convergence simple* sur  $D$ .

**Définition 2.14** (convergence simple). *Soit  $D$  une partie de  $\mathbb{C}$ . La suite  $(f_n)_{n \in \mathbb{N}}$  de fonctions  $D \rightarrow \mathbb{C}$  converge simplement sur  $D$  vers la fonction  $f : D \rightarrow \mathbb{C}$  si*

$$\forall z \in D, \forall \varepsilon > 0, \exists n_0 = n_0(z, \varepsilon), \forall n \geq n_0, \quad |f_n(z) - f(z)| < \varepsilon. \quad (2.4)$$

Une question naturelle est alors : « dans quelles conditions est-ce que  $f$  peut hériter des éventuelles propriétés de régularité de  $f_n$  (continuité, dérivabilité) ? ». À titre d'exemple, supposons que les  $f_n$  soient toutes continues au point  $z_0$ , peut-on en déduire que  $f$  l'est aussi ? En notant que  $f = \lim_{n \rightarrow +\infty} f_n$ , on voit qu'il s'agit d'un problème d'*interversion de limites*. En effet, la continuité de  $f$  au point  $z_0$  s'écrit  $\lim_{z \rightarrow z_0} f(z) = f(z_0)$  et la question posée s'écrit donc :

$$\lim_{z \rightarrow z_0} \left( \lim_{n \rightarrow +\infty} f_n(z) \right) = \lim_{n \rightarrow +\infty} f_n(z_0) = \lim_{n \rightarrow +\infty} \left( \lim_{z \rightarrow z_0} f_n(z) \right) ?$$

Il nous faut donc regarder si  $f(z)$  est aussi proche que l'on veut de  $f(z_0)$  à condition de prendre  $z$  suffisamment proche de  $z_0$  et la difficulté est que l'on ne sait rien *a priori* de la continuité de  $f$ . Pour pouvoir exploiter la continuité des  $f_n$  qui approximent  $f$ , une idée naturelle est alors de découper en trois morceaux la distance entre  $f(z)$  et  $f(z_0)$  par inégalité triangulaire :

$$|f(z) - f(z_0)| \leq |f(z) - f_n(z)| + |f_n(z) - f_n(z_0)| + |f_n(z_0) - f(z_0)|.$$

1. Pour des raisons qui sortent du programme de ce cours.

On est alors naturellement tenté de dire que pour  $n$  grand les morceaux  $|f(z) - f_n(z)|$  et  $|f_n(z_0) - f(z_0)|$  sont petits par convergence simple de  $f_n$  vers  $f$  et que pour  $z$  suffisamment proche de  $z_0$ , le morceau  $|f_n(z) - f_n(z_0)|$  est lui aussi petit par continuité de  $f_n$ . On peut ainsi rendre  $|f(z) - f(z_0)|$  aussi petit que l'on veut (disons inférieur à  $3\varepsilon$  pour  $\varepsilon$  arbitraire) à condition de choisir d'abord  $n$  assez grand pour que  $|f(z) - f_n(z)| < \varepsilon$  et  $|f_n(z_0) - f(z_0)| < \varepsilon$  (tous deux pour le même  $n$ ); ensuite,  $n$  étant ainsi fixé, la continuité de  $f_n$  au point  $z_0$  nous permet d'affirmer l'existence d'un  $\delta > 0$  tel que pour tout  $z$  dans  $D$  tel que  $|z - z_0| < \delta$ ,  $|f_n(z) - f_n(z_0)| < \varepsilon$ . En recollant les trois morceaux, on obtient donc  $|f(z) - f(z_0)| < 3\varepsilon$  pour  $|z - z_0| < \delta$ , ce qui montre la continuité<sup>2</sup> de  $f$  au point  $z_0$ .

Arrivé à ce point, l'auteur déclare que tout lecteur ayant perçu l'erreur dans le raisonnement ci-dessus est un mathématicien ou a les capacités pour le devenir.

Revoyons les choses pas à pas pour détecter la faille dans le raisonnement. Commençons par écrire précisément ce que nous voulons montrer, à savoir la continuité de  $f$  au point  $z_0$  :

$$\forall \varepsilon > 0, \exists \delta > 0, \forall z \in D \text{ tel que } |z - z_0| < \delta, \quad |f(z) - f(z_0)| < 3\varepsilon. \quad (2.5)$$

D'abord la petitesse des morceaux  $|f_n(z_0) - f(z_0)|$  et  $|f(z) - f_n(z)|$  par convergence simple de  $f_n$  vers  $f$  signifie précisément que :

$$\forall \varepsilon > 0, \exists n_0 = n_0(z_0, \varepsilon), \quad \forall n \geq n_0, \quad |f_n(z_0) - f(z_0)| < \varepsilon, \quad (2.6)$$

$$\forall \varepsilon > 0, \exists n_1 = n_1(z, \varepsilon), \quad \forall n \geq n_1, \quad |f_n(z) - f(z)| < \varepsilon. \quad (2.7)$$

Jusqu'ici pas d'erreur, puisqu'on s'est contenté de recopier la définition de la convergence de  $f_n(z)$  vers  $f(z)$  et celle de  $f_n(z_0)$  vers  $f(z_0)$ . Il nous faut maintenant choisir *le même*  $n$  vérifiant les inégalités dans (2.6) et (2.7) pour pouvoir exploiter la continuité de  $f_n$  au point  $z_0$ . Il suffit de prendre n'importe quel  $n$  supérieur ou égal à la fois à  $n_0$  et à  $n_1$ . L'ensemble des  $n$  possibles est donc formé des entiers  $n \geq m_0 = \max(n_0, n_1)$ . Donc  $n$  doit être pris dans l'intervalle :

$$[m_0, +\infty[ \quad \text{avec } m_0 = m_0(z, z_0, \varepsilon) = \max(n_0(z_0, \varepsilon), n_1(z, \varepsilon)).$$

Ce choix étant fait, on travaille avec une seule fonction  $f_n$  et sa continuité nous dit que :

$$\forall \varepsilon > 0, \exists \delta = \delta(\varepsilon, f_n, z_0) > 0, \forall w \in D \text{ tel que } |w - z_0| < \delta, \quad |f_n(w) - f_n(z_0)| < \varepsilon. \quad (2.8)$$

Donc il semble bien qu'en appliquant ceci avec  $w = z$ , on obtienne (2.5). Malheureusement (2.5) doit être vérifiée pour *tous les*  $z \in D$  dont la distance à  $z_0$  est inférieure à  $\delta$ . Et en général, il y en a une infinité pour chaque  $\delta > 0$ , sauf si  $z_0$  est un point « isolé » de  $D$ . Or  $f_n$  tel que nous l'avons choisi ci-dessus *dépend* de  $z$  et c'est là la faille du raisonnement. Pour que le raisonnement devienne correct, il faudrait que l'on puisse garantir la possibilité de choisir le même  $f_n$  par la procédure ci-dessus au moins pour tous les  $z \in D$  tels que  $|z - z_0| < r$  pour un certain  $r$  (même très petit, mais strictement positif). Dans ce cas, (2.8) est valable avec le même  $n$  en imposant la condition supplémentaire  $0 < \delta \leq r$  dans le choix de  $\delta$ , ce qui est toujours possible sans altérer la définition de la continuité. Notons maintenant  $K = \{z \in D ; |z - z_0| < r\}$ . Comment peut-on choisir le même  $n$  vérifiant les inégalités dans (2.6) et (2.7) pour tous les  $z \in K$ ? Un tel  $n$  devrait être supérieur ou égal à tous les  $n_1(z, \varepsilon)$  pour  $z \in K$ . Et le problème est que s'il y a une infinité d'éléments

2. Le fait que l'on ait pour majorant  $3\varepsilon$  au lieu de  $\varepsilon$  est sans importance puisque  $\varepsilon > 0$  est arbitraire. Si vous n'en êtes pas convaincu, remplacez  $\varepsilon$  par  $\varepsilon' = \varepsilon/3$  dans les opérations de contrôle des trois morceaux et à l'arrivée vous obtiendrez après recollage  $\varepsilon$ .

dans  $K$  (aussi petit que l'on choisisse  $r$ ), il peut très bien y avoir une infinité de valeurs *distinctes* dans l'ensemble des entiers  $n_1(z, \varepsilon)$  et dans ce cas cet ensemble d'entiers n'est pas borné et on ne peut pas trouver d'entier  $n$  les majorant tous.

En conclusion de cette discussion, on peut dire que la convergence simple semble insuffisante<sup>3</sup> pour que  $f$  hérite de la continuité des  $f_n$  au point  $z_0$ , mais que si l'on sait contrôler l'approximation de  $f(z)$  par  $f_n(z)$  *uniformément* par rapport à  $z$  dans un certain « voisinage » de  $z_0$ , alors on peut déduire la continuité de  $f$  de celle des  $f_n$ . Ceci nous conduit à la définition suivante.

**Définition 2.15** (convergence uniforme). *Soit  $D$  une partie de  $\mathbb{C}$ ,  $f$  une fonction  $D \rightarrow \mathbb{C}$  et  $K \subset D$ . La suite  $(f_n)_{n \in \mathbb{N}}$  de fonctions  $D \rightarrow \mathbb{C}$  converge uniformément sur  $K$  vers la fonction  $f$  si*

$$\forall \varepsilon > 0, \exists n_0 = n_0(\varepsilon), \forall n \geq n_0, \forall z \in K, \quad |f_n(z) - f(z)| < \varepsilon. \quad (2.9)$$

Quitte à paraître un peu lourd, insistons encore une fois sur le fait que le  $n_0$  qui figure dans (2.9) est commun à tous les  $z$  de  $K$ , notez la différence avec la définition de la convergence simple par (2.4) où l'on écrit d'abord «  $\forall z$  » et ensuite «  $\exists n_0 \dots$  », ce qui laisse entendre que  $n_0$  peut dépendre de  $z$  tandis que l'écriture dans l'ordre inverse dans (2.9) impose que  $n_0$  soit le même pour tous les  $z \in K$ .

**Proposition 2.16.** *La convergence uniforme sur  $K$  implique la convergence simple sur  $K$ .*

Si vous avez lu tout ce qui précède, cela devrait vous paraître évident. Pour voir définitivement que la réciproque est fautive, voici un contre exemple.

**Exemple 2.17.** Soit  $D = \{z \in \mathbb{C}; |z| \leq 1\}$  et  $(f_n)_{n \in \mathbb{N}}$  la suite de fonctions  $D \rightarrow \mathbb{C}$  définie par :

$$\forall n \in \mathbb{N}, \forall z \in D, \quad f_n(z) = (1 - |z|)^n.$$

Cette suite converge simplement sur  $D$  vers la fonction  $f$  définie par

$$\forall z \in D, \quad f(z) = \begin{cases} 1 & \text{si } z = 0, \\ 0 & \text{si } 0 < |z| \leq 1. \end{cases}$$

On remarque que  $f$  n'est pas continue au point 0. Il en résulte que pour  $K = \{z \in D; |z| \leq r\}$ , avec  $0 < r < 1$ ,  $f_n$  ne peut converger uniformément vers  $f$  sur  $K$ . Les vérifications sont laissées en exercice.

La convergence uniforme est le mode de convergence utile pour transférer les propriétés de régularité des fonctions  $f_n$  à la limite  $f$ . Les deux théorèmes suivants illustrent ce fait. Ils trouveront une application spécifique au cas des séries entières.

**Théorème 2.18** (continuité par convergence uniforme). *Soit  $D$  une partie de  $\mathbb{C}$ ,  $(f_n)_{n \in \mathbb{N}}$  une suite de fonctions  $D \rightarrow \mathbb{C}$  et  $f$  une fonction  $D \rightarrow \mathbb{C}$ .*

- i) *Si les  $f_n$  sont continues au point  $z_0 \in D$  et s'il existe un disque ouvert  $\Delta$  centré en  $z_0$  tel que  $(f_n)_{n \in \mathbb{N}}$  converge uniformément vers  $f$  sur  $D \cap \Delta$ , alors  $f$  est continue au point  $z_0$ .*
- ii) *Si les  $f_n$  sont continues sur un disque ouvert  $\Delta$  inclus dans  $D$  et  $(f_n)_{n \in \mathbb{N}}$  converge uniformément sur  $\Delta$  vers  $f$  alors  $f$  est continue sur  $\Delta$ .*

---

3. Cette insuffisance ne peut se prouver à partir d'une tentative infructueuse de démonstration, mais un contre exemple simple permettra de l'établir (ex. 2.17).

*Preuve.* Le point i) a déjà été démontré pour l'essentiel dans la discussion préliminaire à la définition de la convergence uniforme. Bien sûr cette remarque ne dispense pas l'étudiant consciencieux de réécrire cette preuve dans le contexte de ce théorème.

Pour le point ii), il suffit de remarquer que si  $\Delta$  est un disque ouvert inclus dans  $D$ , tout point de  $\Delta$  est centre d'un disque ouvert  $\Delta'$  inclus dans  $\Delta$  et que la convergence uniforme sur  $\Delta$  implique la convergence uniforme sur tout sous-ensemble de  $\Delta$  donc ici sur  $\Delta'$ . Il ne reste plus alors qu'à appliquer le i) avec  $\Delta'$  à la place de  $\Delta$ .  $\square$

**Corollaire 2.19.** *Soit  $D = D(0, R) = \{z \in \mathbb{C} ; |z| < R\}$ . On suppose que les  $f_n$  sont continues sur  $D$  et que  $(f_n)_{n \in \mathbb{N}}$  converge uniformément vers  $f$  sur chaque disque fermé  $\overline{D}(0, r)$  pour  $0 < r < R$ . Alors  $f$  est continue sur  $D(0, R)$ .*

Pour comprendre ce corollaire, il importe de remarquer que la convergence uniforme sur chaque  $\overline{D}(0, r)$  pour  $0 < r < R$  n'implique pas la convergence uniforme sur la réunion de ces disques, c'est-à-dire sur  $D(0, R)$ . La situation décrite par ce corollaire est typiquement celle des séries entières.

*Preuve.* Soit  $z_0$  un point quelconque de  $D(0, R)$ . Alors  $|z_0| < R$ , donc on peut trouver un  $r$  tel que  $|z_0| < r < R$ , par exemple  $r = (|z_0| + R)/2$ . Alors  $z_0$  appartient à un disque ouvert  $\Delta = D(0, r) = \{z \in \mathbb{C} ; |z| < r\}$  inclus dans  $D$  et  $f_n$  converge uniformément vers  $f$  sur  $\Delta$  donc par le point ii) du théorème  $f$  est continue sur  $\Delta$  et en particulier en  $z_0$ . Comme ce raisonnement vaut pour tout  $z_0 \in D$ ,  $f$  est continue sur  $D$ .  $\square$

Nous donnons maintenant sans démonstration un théorème de dérivabilité par convergence uniforme.

**Théorème 2.20** (dérivabilité par convergence uniforme). *Pour  $w_0 \in \mathbb{C}$  et  $r > 0$ , notons  $D(w_0, r) = \{z \in \mathbb{C} ; |w_0 - z| < r\}$  le disque ouvert de centre  $w_0$  et de rayon  $r$ . On suppose que*

- a)  $(f_n)_{n \in \mathbb{N}}$  converge simplement vers  $f$  sur  $D(w_0, R)$ , pour un  $R > 0$ ;
- b) chaque  $f_n$  est dérivable sur  $D(w_0, R)$ ;
- c) pour tout  $0 < r < R$ ,  $(f'_n)_{n \in \mathbb{N}}$  converge uniformément sur  $D(w_0, r)$  vers une fonction  $g$ .

Alors  $f$  est dérivable sur  $D(w_0, R)$  et  $f' = \lim_{n \rightarrow +\infty} f'_n = g$ .

Tout ce que nous venons de voir sur les suites de fonctions d'une variable complexe s'applique en particulier aux séries de fonctions d'une variable complexe :

$$S : D \rightarrow \mathbb{C}, \quad z \longmapsto S(z) = \sum_{k=0}^{+\infty} u_k(z),$$

où les  $u_k$  sont définies sur  $D$  et à valeurs dans  $\mathbb{C}$  et la série ci-dessus converge pour tout  $z \in D$ . Le lien entre suites de fonctions et série se fait par la suite des sommes partielles  $(S_n)_{n \in \mathbb{N}}$  :

$$S_n : D \rightarrow \mathbb{C}, \quad z \longmapsto S_n(z) = \sum_{k=0}^n u_k(z).$$

On dira ainsi que la série  $S$  converge simplement sur  $D$  (resp uniformément sur  $K \subset D$ ) si la suite de fonctions  $S_n$  converge simplement vers  $S$  sur  $D$  (resp. uniformément vers  $S$  sur  $K$ ).

Pour montrer la convergence uniforme d'une série de fonctions, une condition suffisante bien pratique est la notion de convergence normale.



**Définition 2.21** (convergence normale). *On dit que la série de fonctions  $S(z) = \sum_{k=0}^{+\infty} u_k(z)$  converge normalement sur  $A \subset \mathbb{C}$ , s'il existe une série convergente à termes positifs  $\sum_{k=0}^{+\infty} b_k$  telle que*

$$\forall k \in \mathbb{N}, \forall z \in A, \quad |u_k(z)| \leq b_k.$$

**Proposition 2.22.** *La convergence normale d'une série de fonctions sur  $A$  implique sa convergence uniforme sur  $A$ .*

*Preuve.* Par hypothèse, pour tout  $z \in A$  et tout  $k \in \mathbb{N}$ ,  $|u_k(z)| \leq b_k$ , donc la série  $\sum_{k=0}^{+\infty} u_k(z)$  est absolument convergente et la fonction  $S(z) = \sum_{k=0}^{+\infty} u_k(z)$  est bien définie au moins sur  $A$ . On peut alors écrire :

$$\forall z \in A, \quad |S(z) - S_n(z)| = \left| \sum_{k=n+1}^{+\infty} u_k(z) \right| \leq \sum_{k=n+1}^{+\infty} |u_k(z)| \leq \sum_{k=n+1}^{+\infty} b_k.$$

Comme la série à termes positifs  $\sum_{k=0}^{+\infty} b_k$  converge, son reste de rang  $n$  tend vers 0. Alors pour tout  $\varepsilon > 0$ , il existe un  $n_0 = n_0(\varepsilon)$  tel que pour tout  $n \geq n_0$ ,  $\sum_{k=n+1}^{+\infty} b_k < \varepsilon$ . Remarquons que ce  $n_0$  dépend de  $\varepsilon$  et de la suite  $(b_k)_{k \in \mathbb{N}}$ , mais pas de  $z$ . On en déduit :

$$\forall \varepsilon > 0, \exists n_0 = n_0(\varepsilon), \forall z \in A, \forall n \geq n_0, \quad |S(z) - S_n(z)| < \varepsilon.$$

Ceci exprime exactement la convergence uniforme sur  $A$  de  $S_n$  vers  $S$ . □

## 2.2 Disque de convergence

Nous pouvons maintenant aborder l'étude des séries entières  $\sum_{k=0}^{+\infty} a_k z^k$ . La première question qui se pose est de déterminer l'ensemble des  $z \in \mathbb{C}$  pour lesquels cette série converge. Il est commode pour cela d'utiliser les notations suivantes.

Pour  $w_0 \in \mathbb{C}$  et  $r \in [0, +\infty[$ , les ensembles

$$\begin{aligned} D(w_0, r) &:= \{z \in \mathbb{C} ; |z - w_0| < r\}, \\ \overline{D}(w_0, r) &:= \{z \in \mathbb{C} ; |z - w_0| \leq r\}, \end{aligned}$$

désignent respectivement les disques ouvert et fermé de centre  $w_0$  et de rayon  $r$ , voir la figure 2.1 (ces deux disques devraient être superposés et on ne verrait alors que  $\overline{D}(w_0, r)$ ; ils ont été séparés pour des raisons de lisibilité).

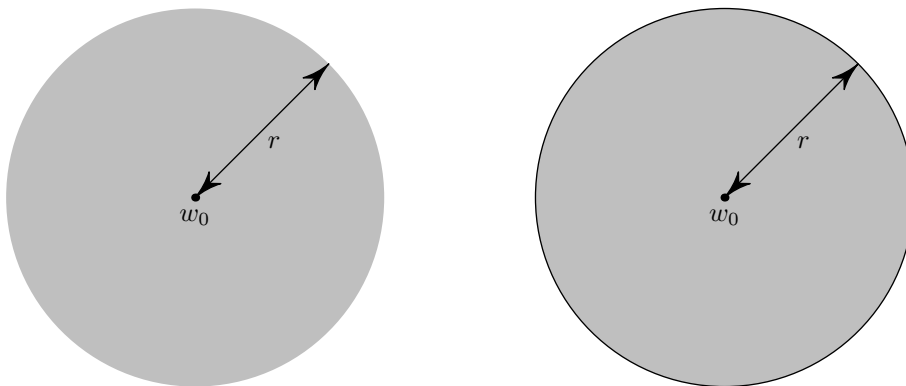


FIGURE 2.1 – Disque ouvert  $D(w_0, r)$  et disque fermé  $\overline{D}(w_0, r)$

Il y a toujours au moins une valeur de  $z$  pour laquelle la série  $\sum_{k=0}^{+\infty} a_k z^k$  converge, c'est  $z = 0$ . En effet dans ce cas, tous les termes de la série sont nuls à partir du rang  $k = 1$ . La série se réduit donc à son premier terme d'indice 0. Notons ici que l'on fait la même convention que pour les polynômes, à savoir que  $z^0$  vaut toujours 1, même si  $z = 0$ . Donc pour  $z = 0$ , la série converge trivialement et sa somme vaut  $a_0$ .

La clé de l'étude de la convergence des séries entières est leur comparaison avec des séries géométriques. Nous allons voir *grosso modo* qu'il existe pour chaque série entière une valeur critique  $R \in [0, +\infty]$  telle si  $|z| < R$ , la série converge à vitesse géométrique (ou exponentielle) et si  $|z| > R$  (avec  $R < +\infty$ ) la série diverge grossièrement : son terme général ne tend même pas vers 0.

**Lemme 2.23** (d'Abel). *Supposons qu'il existe  $r_0 > 0$  tel que la suite  $(|a_k| r_0^k)_{k \in \mathbb{N}}$  soit bornée. Alors la série entière  $\sum_{k=0}^{+\infty} a_k z^k$*

- converge absolument en tout point  $z$  du disque ouvert  $D(0, r_0)$  ;
- converge normalement sur chaque disque fermé  $\overline{D}(0, r)$  pour  $r < r_0$ .

*Preuve.* Par hypothèse, il existe une constante  $M \in ]0, +\infty[$  telle que

$$\forall k \in \mathbb{N}, \quad |a_k| r_0^k \leq M.$$

- Soit  $z$  quelconque dans  $D(0, r_0)$ . En écrivant

$$|a_k z^k| = |a_k| r_0^k \left| \frac{z}{r_0} \right|^k \leq M \left( \frac{|z|}{r_0} \right)^k$$

et en notant que par appartenance de  $z$  à  $D(0, r_0)$ ,  $|z| < r_0$  et donc  $|z|/r_0 < 1$ , on voit que la série  $\sum_{k=0}^{+\infty} a_k z^k$  converge absolument par comparaison avec la série géométrique de terme général  $M(|z|/r_0)^k$ .

- Soit  $r < r_0$ . Pour tout  $z \in \overline{D}(0, r)$ ,  $|z| \leq r$ , d'où

$$|a_k z^k| \leq M \left( \frac{r}{r_0} \right)^k$$

et puisque  $r/r_0 < 1$ , on a bien convergence normale de  $\sum_{k=0}^{+\infty} a_k z^k$  sur  $\overline{D}(0, r)$ . □

Considérons maintenant l'ensemble

$$I := \{r \in \mathbb{R}_+ ; (|a_k| r^k)_{k \in \mathbb{N}} \text{ bornée}\}. \quad (2.10)$$

On remarque d'abord que  $r = 0$  appartient toujours à  $I$  et en est le plus petit élément. De plus  $I$  est un *intervalle* de  $\mathbb{R}_+$  car si  $r > 0$  est dans  $I$ , tout  $r' \in [0, r[$  y est aussi. En effet pour tout  $k$  entier,  $r'^k \leq r^k$  donc  $|a_k| r'^k \leq |a_k| r^k$  et la bornitude de la suite  $(|a_k| r^k)_{k \in \mathbb{N}}$  implique celle de  $(|a_k| r'^k)_{k \in \mathbb{N}}$ . L'ensemble  $I$  est donc de la forme  $I = [0, R[$  ou  $I = [0, R]$  pour un  $R \in \mathbb{R}_+$  ou  $I = [0, +\infty[$  (cas où  $(|a_k| r^k)_{k \in \mathbb{N}}$  est bornée pour tout  $r > 0$ ).

**Définition 2.24** (rayon de convergence). *On appelle rayon de convergence de la série entière  $\sum_{k=0}^{+\infty} a_k z^k$  la borne supérieure (dans  $\overline{\mathbb{R}_+}$ ) de l'ensemble des réels  $r \geq 0$  tels que la suite  $(|a_k| r^k)_{k \in \mathbb{N}}$  soit bornée.*

Notons  $R$  le rayon de convergence de la série entière. C'est aussi la borne droite de l'intervalle  $I$  défini ci-dessus. Voici un exemple pour chacune des situations possibles. Les vérifications sont laissées en exercice.

**Exemple 2.25.**

$$\sum_{k=0}^{+\infty} 2^k z^k \quad R = \frac{1}{2} \quad I = \left[0, \frac{1}{2}\right], \quad (2.11)$$

$$\sum_{k=0}^{+\infty} 2^k k^5 z^k \quad R = \frac{1}{2} \quad I = \left[0, \frac{1}{2}\right], \quad (2.12)$$

$$\sum_{k=0}^{+\infty} \frac{z^k}{k!} \quad R = +\infty \quad I = [0, +\infty[. \quad (2.13)$$

L'appellation « rayon de convergence » se justifie par le théorème suivant.

**Théorème 2.26.** Soit  $R$  le rayon de convergence de la série entière  $\sum_{k=0}^{+\infty} a_k z^k$ .

Cas 1. Si  $R = 0$ , la série ne converge que pour  $z = 0$ .

Cas 2. Si  $0 < R < +\infty$ ,

- a) la série converge absolument pour tout  $z \in D(0, R)$  ;
- b) elle converge normalement sur chaque disque  $\overline{D}(0, r)$  pour  $r < R$  ;
- c) elle diverge pour tout  $z$  tel que  $|z| > R$ .

Cas 3. Si  $R = +\infty$ ,

- a) la série converge absolument pour tout  $z \in \mathbb{C}$  ;
- b) elle converge normalement sur toute partie bornée de  $\mathbb{C}$ .

*Preuve.* Rappelons que l'intervalle  $I$  défini par (2.10) est de la forme  $[0, R[$  pour  $0 \leq R \leq +\infty$  ou  $[0, R]$  pour  $0 \leq R < +\infty$ .

- Si  $R < +\infty$ , soit  $z$  tel que  $|z| > R$ . Comme  $R$  est la borne supérieure de l'intervalle  $I$ ,  $|z| \notin I$ , donc la suite  $(|a_k z^k|)_{k \in \mathbb{N}}$  n'est pas bornée. Donc  $|a_k z^k|$  ne peut tendre vers zéro<sup>4</sup> et de même pour  $a_k z^k$ , cf. remarque 2.2. La série  $\sum_{k=0}^{+\infty} a_k z^k$  diverge donc grossièrement. Ceci établit les cas 1 et 2c).
- Si  $0 < r < R \leq +\infty$ , il existe  $r'$  tel que  $r < r' < R$  et donc  $r' \in I$  et  $(|a_k| r'^k)_{k \in \mathbb{N}}$  est bornée. Par le lemme d'Abel (lemme 2.23), la série converge absolument en tout point  $z$  de  $D(0, r')$ , donc en particulier pour tout  $z$  de module  $r$ . Ceci étant valable pour tout  $r < R$ , la série converge absolument sur  $D(0, R)$ , ce qui établit le cas 2a) et aussi le cas 3a), en convenant que  $D(0, +\infty) = \mathbb{C}$ . De plus, par la deuxième partie du lemme d'Abel, la série  $\sum_{k=0}^{+\infty} a_k z^k$  converge normalement sur  $\overline{D}(0, r)$ . Ceci établit le cas 2b) et aussi le cas 3b) car toute partie bornée de  $\mathbb{C}$  peut être incluse dans un disque  $\overline{D}(0, r)$  pour  $r$  assez grand. □

**Remarque 2.27.**

- i) Le théorème 2.26 ne dit rien sur ce qui se passe pour  $|z| = R$ .
- ii) En général il n'y a pas convergence normale de la série entière sur  $D(0, R)$ .
- iii) Le théorème 2.26 reste valable pour les séries entières de variable réelle  $\sum_{k=0}^{+\infty} a_k x^k$  en remplaçant  $D(0, R)$  par  $] - R, R[$ .

4. Une suite convergente dans  $\mathbb{R}$  est bornée.

**Définition 2.28.** Si le rayon de convergence  $R$  de la série entière  $\sum_{k=0}^{+\infty} a_k z^k$  est non nul, on dit que  $D(0, R)$  est le disque de convergence de  $\sum_{k=0}^{+\infty} a_k z^k$ . De même  $] - R, R[$  est appelé intervalle de convergence de la série entière de variable réelle  $\sum_{k=0}^{+\infty} a_k x^k$ .

Il convient de ne pas confondre « disque de convergence » et ensemble de tous les points de convergence. En effet, si l'on est sûr de la convergence dans le disque ouvert  $D(0, R)$  et de la divergence pour  $|z| > R$  (si  $R < +\infty$ ), il reste une zone du plan pour laquelle on ne sait rien dire en général. C'est le cercle  $\mathcal{C}(0, R) = \{z \in \mathbb{C} ; |z| = R\}$  appelé *cercle frontière* de la série entière. Selon les exemples, tout peut arriver pour  $z \in \mathcal{C}(0, R)$ . Voici trois exemples où le rayon de convergence est 1 et le comportement de la série sur son cercle frontière varie considérablement d'un exemple à l'autre.

**Exemple 2.29** (divergence grossière en tout point du cercle frontière). La série entière  $\sum_{k=0}^{+\infty} z^k$  a pour rayon de convergence 1 et diverge en tout point de son cercle frontière  $\mathcal{C}(0, 1)$ . En effet si  $|z| = 1$ ,  $|z^k| = |z|^k = 1$ , donc le terme général de la série ne tend pas vers 0, cf. remarque 2.2.

**Exemple 2.30** (convergence normale sur le cercle frontière). La série  $\sum_{k=0}^{+\infty} (1 + k^2)^{-1} z^k$  a pour rayon de convergence 1 et converge normalement, donc aussi absolument, sur le disque fermé  $\overline{D}(0, 1)$ . En particulier elle converge normalement et absolument sur son cercle frontière  $\mathcal{C}(0, 1)$ . En effet, pour tout  $z \in \overline{D}(0, 1)$ ,  $|z| \leq 1$  et on a la majoration :

$$\left| \frac{z^k}{1 + k^2} \right| = \frac{|z|^k}{1 + k^2} \leq \frac{1}{1 + k^2}$$

et ce majorant indépendant de  $z$  est le terme général d'une série à terme positifs convergente par équivalence avec  $k^{-2}$ , terme général d'une série de Riemann convergente.

**Exemple 2.31** (divergence en un point du cercle frontière, convergence non absolue aux autres points du cercle). La série  $\sum_{k=1}^{+\infty} z^k/k$  a pour rayon de convergence 1. Elle diverge clairement au point  $z = 1$  (série harmonique) et converge non absolument au point  $z = -1$  (série harmonique alternée). En écrivant tout  $z \in \mathcal{C}(0, 1)$  sous la forme  $z = e^{i\theta}$ , on se ramène par la proposition 2.5 à l'étude des séries  $\sum_{k=1}^{+\infty} k^{-1} \cos(k\theta)$  et  $\sum_{k=1}^{+\infty} k^{-1} \sin(k\theta)$ . Par le théorème d'Abel (voir le corollaire 1.40), on sait que ces séries convergent pour  $e^{i\theta} \neq 1$ . Par ailleurs la non convergence absolue de  $\sum_{k=1}^{+\infty} k^{-1} e^{ik\theta}$  est évidente directement puisque  $|k^{-1} e^{ik\theta}| = k^{-1}$  qui est le terme général de la série harmonique. Finalement la série entière diverge au point  $z = 1$  de  $\mathcal{C}(0, 1)$  et converge (non absolument) en tous les autres points de son cercle frontière.

Le théorème 2.26 et sa preuve nous fournissent une méthode pratique pour trouver le rayon de convergence  $R$  d'une série entière. On a vu en effet que pour tout  $|z| < R$  la série converge absolument donc en posant  $r = |z|$ ,  $|a_k| r^k$  tend vers 0. *A contrario*, si  $|z| > R$  (lorsque  $R < +\infty$ ), la série diverge grossièrement, c.-à-d. son terme général ne tend pas vers 0, donc  $|a_k| r^k$  tend vers 0. Ceci nous conduit à la règle suivante qui devrait vous permettre de déterminer explicitement le rayon de convergence dans la plupart des cas.

**Proposition 2.32** (détermination pratique du rayon de convergence). *Pour trouver le rayon de convergence de la série entière  $\sum_{k=0}^{+\infty} a_k z^k$ , on cherche l'ensemble de toutes les valeurs  $r \geq 0$  telles que*

$$\lim_{k \rightarrow +\infty} |a_k| r^k = 0.$$

*C'est un intervalle de la forme  $[0, R[$  ou  $[0, R]$  et la borne droite de cet intervalle est le rayon de convergence.*

**Exemple 2.33.** La série entière  $\sum_{k=0}^{+\infty} (k^2 + 3k + 5)z^k$  a pour rayon de convergence 1. Pour le voir, on vérifie que :

$$\lim_{k \rightarrow +\infty} |k^2 + 3k + 5|r^k = \begin{cases} 0 & \text{si } r < 1, \\ +\infty & \text{si } r > 1. \end{cases}$$

En effet dans le cas  $r < 1$ , on a une forme indéterminée du type «  $\infty \times 0$  », mais la suite exponentielle  $(r^k)_{k \in \mathbb{N}}$  tend vers 0 suffisamment vite pour l'emporter sur la convergence vers l'infini de n'importe quel polynôme. Dans le cas  $r > 1$ , les deux facteurs tendent vers l'infini. À titre anecdotique, on peut remarquer que si  $r = 1$ , la limite est encore  $+\infty$ , mais c'est sans importance pour la détermination de  $R$ .

D'autres exemples sont proposés en T.D. En attendant, voici une généralisation de l'exemple 2.33 qui nous sera utile par la suite.

**Lemme 2.34.** Si  $P$  est un polynôme, la série entière  $\sum_{k=0}^{+\infty} P(k)a_k z^k$  a même rayon de convergence que  $\sum_{k=0}^{+\infty} a_k z^k$ .

*Preuve.* Le résultat étant évident si  $P$  est une constante, on peut supposer que le degré  $d$  de  $P$  est au moins 1. La fonction polynôme  $P$  est de la forme

$$P : t \longmapsto c_d t^d + c_{d-1} t^{d-1} + \dots + c_1 t + c_0.$$

Notons  $R$  le rayon de convergence de  $\sum_{k=0}^{+\infty} a_k z^k$ . D'après la proposition 2.32, il nous suffit de montrer que

1. si  $r < R$ ,  $|P(k)a_k r^k|$  tend vers 0 ;
2. si  $r > R$ ,  $|P(k)a_k r^k|$  ne tend pas vers 0.

Dans le cas où  $R = 0$  il suffit de vérifier le point 2 et si  $R = +\infty$ , il suffit de vérifier le point 1.

Pour vérifier le point 1 (lorsque  $R > 0$ ), soit  $r < R$ . Alors il existe  $r_0$  tel que  $r < r_0 < R$  et par définition du rayon de convergence, la suite  $(|a_k| r_0^k)_{k \in \mathbb{N}}$  est bornée. Il existe donc une constante  $M$  telle que pour tout  $k \in \mathbb{N}$ ,  $|a_k| r_0^k \leq M$ . On en déduit que

$$\forall k \in \mathbb{N}, \quad |P(k)||a_k| r^k = |P(k)||a_k| r_0^k \left(\frac{r}{r_0}\right)^k \leq M |P(k)| \left(\frac{r}{r_0}\right)^k,$$

Or  $|P(k)|$  est équivalent à  $|c_d| k^d$  quand  $k$  tend vers l'infini et comme  $0 \leq q = r/r_0 < 1$ ,  $k^d q^k$  tend vers 0 quand  $k$  tend vers l'infini. On en déduit que  $|P(k)||a_k| r^k$  tend aussi vers 0, ce qui établit le point 1.

Supposons maintenant que  $R < r < +\infty$ . Comme  $r > R$ , la suite  $(|a_k| r^k)_{k \in \mathbb{N}}$  n'est pas bornée. Elle a donc une sous-suite tendant vers l'infini, et comme  $|P(k)|$  tend vers l'infini avec  $k$ , la suite  $(|P(k)a_k r^k|)_{k \in \mathbb{N}}$  a une sous-suite tendant vers l'infini. Le point 2 est donc établi.  $\square$

## 2.3 Opérations sur les séries entières

**Proposition 2.35** (somme de deux séries entières). Soient  $f(z) = \sum_{k=0}^{+\infty} a_k z^k$  et  $g(z) = \sum_{k=0}^{+\infty} b_k z^k$  deux séries entières de rayons de convergence respectifs  $R_1$  et  $R_2$ . Alors la série entière  $\sum_{k=0}^{+\infty} (a_k + b_k) z^k$  a pour rayon de convergence  $R \geq \min(R_1, R_2)$  et

$$\forall z \in D(0, \min(R_1, R_2)), \quad f(z) + g(z) = \sum_{k=0}^{+\infty} (a_k + b_k) z^k. \quad (2.14)$$

*Preuve.* Soit  $z$  quelconque dans  $D(0, \min(R_1, R_2))$ . Puisque  $|z| < R_1$  et  $|z| < R_2$ , les séries  $\sum_{k=0}^{+\infty} a_k z^k$  et  $\sum_{k=0}^{+\infty} b_k z^k$  convergent absolument. En revenant aux sommes partielles et en rappelant que  $|a_k + b_k| \leq |a_k| + |b_k|$ , on en déduit que la série somme  $\sum_{k=0}^{+\infty} (a_k + b_k) z^k$  converge aussi absolument et vérifie (2.14). En particulier avec  $z = r$ , on voit que pour tout réel positif  $r < \min(R_1, R_2)$ ,  $|a_k + b_k| r^k$  tend vers 0, donc d'après la proposition 2.32, le rayon de convergence  $R$  de  $\sum_{k=0}^{+\infty} (a_k + b_k) z^k$  vérifie  $R \geq r$ . Ceci étant vrai pour tout  $r < \min(R_1, R_2)$ , on en déduit que  $R \geq \min(R_1, R_2)$ .  $\square$

**Remarque 2.36.** Il se peut que l'inégalité  $R \geq \min(R_1, R_2)$  soit stricte. Voici un exemple simple. On prend  $a_k = 2^{-k}$ ,  $b_k = 3^{-k} - 2^{-k}$ . Alors les séries  $\sum_{k=0}^{+\infty} a_k z^k$  et  $\sum_{k=0}^{+\infty} b_k z^k$  ont pour rayon de convergence  $R_1 = R_2 = 2$  mais  $\sum_{k=0}^{+\infty} (a_k + b_k) z^k = \sum_{k=0}^{+\infty} 3^{-k} z^k$  a pour rayon de convergence  $R = 3$ .

**Proposition 2.37** (produit d'une série entière par une constante). *Soit  $f(z) = \sum_{k=0}^{+\infty} a_k z^k$  une série entière de rayon de convergence  $R$  et  $c$  une constante complexe non nulle. Alors la série  $\sum_{k=0}^{+\infty} c a_k z^k$  a pour rayon de convergence  $R$  et pour tout  $z \in D(0, R)$ ,  $cf(z) = \sum_{k=0}^{+\infty} c a_k z^k$ .*

La vérification est immédiate<sup>5</sup>. En combinant les propositions 2.35 et 2.37, on voit que pour toutes constantes complexes  $c$  et  $c'$  et toutes séries entières  $\sum_{k=0}^{+\infty} a_k z^k$  et  $\sum_{k=0}^{+\infty} b_k z^k$  de rayons de convergence respectifs  $R_1$  et  $R_2$  non nuls, la série combinaison linéaire  $\sum_{k=0}^{+\infty} (c a_k + c' b_k) z^k$  a un rayon de convergence  $R \geq \min(R_1, R_2)$  et

$$\forall z \in D(0, \min(R_1, R_2)), \quad \sum_{k=0}^{+\infty} (c a_k + c' b_k) z^k = c \sum_{k=0}^{+\infty} a_k z^k + c' \sum_{k=0}^{+\infty} b_k z^k. \quad (2.15)$$

**Proposition 2.38** (produit de deux séries entières). *Soient  $f(z) = \sum_{k=0}^{+\infty} a_k z^k$  et  $g(z) = \sum_{k=0}^{+\infty} b_k z^k$  deux séries entières de rayons de convergence respectifs  $R_1$  et  $R_2$ . alors la série entière produit  $\sum_{k=0}^{+\infty} (\sum_{j=0}^k a_j b_{k-j}) z^k$  a pour rayon de convergence  $R \geq \min(R_1, R_2)$  et*

$$\forall z \in D(0, \min(R_1, R_2)), \quad \sum_{k=0}^{+\infty} \left( \sum_{j=0}^k a_j b_{k-j} \right) z^k = f(z)g(z). \quad (2.16)$$

*Preuve.* Pour  $z$  quelconque dans  $D(0, \min(R_1, R_2))$ , la version complexe du théorème sur le produit de deux séries absolument convergentes (th. 1.48) nous donne la convergence absolue de la série produit  $\sum_{k=0}^{+\infty} (\sum_{j=0}^k a_j b_{k-j}) z^k$  et l'égalité 2.16. Comme dans la preuve de la proposition 2.35, on en déduit que le rayon de convergence la série produit est supérieur ou égal à  $\min(R_1, R_2)$ .  $\square$

## 2.4 Régularité de la somme d'une série entière

Pour une série entière  $\sum_{k=0}^{+\infty} a_k z^k$  de rayon de convergence  $R > 0$ , notons

$$S_n(z) = \sum_{k=0}^n a_k z^k \quad \text{et} \quad S(z) = \sum_{k=0}^{+\infty} a_k z^k.$$

Les fonctions polynômes  $S_n$  sont définies sur  $\mathbb{C}$  où elles sont continues et indéfiniment dérivables. La fonction  $S$  est définie au moins sur  $D(0, R)$  et peut-être en certains points

5. Demandez-vous quand même pourquoi on suppose  $c \neq 0$  et ce qu'il advient si  $c = 0$ .

du cercle frontière  $\mathcal{C}(0, R)$ . Puisque la série converge normalement sur le disque fermé  $\overline{D}(0, r)$  pour tout  $r < R$ , elle converge aussi uniformément sur ce disque. Ainsi,  $S$  est limite uniforme sur  $\overline{D}(0, r)$  de la suite de fonctions continues  $(S_n)_{n \in \mathbb{N}}$ , donc  $S$  est *continue* sur  $\overline{D}(0, r)$ . Ceci étant vrai pour tout  $r < R$ ,  $S$  est continue en tout point  $z$  de  $D(0, R)$ , autrement dit, continue sur  $D(0, R)$ .

En fait, on peut obtenir un bien meilleur résultat. Nous allons voir que  $S$  est indéfiniment dérivable sur  $D(0, R)$  et que ses dérivées successives s'obtiennent par dérivation « terme à terme » de la série  $\sum_{k=0}^{+\infty} a_k z^k$ .

**Théorème 2.39.** *Soit  $S(z) = \sum_{k=0}^{+\infty} a_k z^k$  une série entière de rayon de convergence  $R > 0$ . Alors la fonction  $S$  est dérivable sur  $D(0, R)$  et sa dérivée  $S'$  est égale à la série des dérivées « terme à terme » :*

$$\forall z \in D(0, R), \quad S'(z) = \left( \sum_{k=0}^{+\infty} a_k z^k \right)' = \sum_{k=0}^{+\infty} (a_k z^k)' = \sum_{k=1}^{+\infty} k a_k z^k,$$

la série entière  $S'$  ayant même rayon de convergence  $R$  que  $S$ .

*Preuve.* Notons d'abord que la dérivation terme à terme que nous voulons justifier pour la série est valable pour chacune de ses sommes partielles :

$$\forall z \in \mathbb{C}, \quad S'_n(z) = \left( \sum_{k=0}^n a_k z^k \right)' = \sum_{k=1}^n k a_k z^{k-1},$$

par la formule de dérivation d'un polynôme.

Fixons maintenant un point quelconque  $z_0 \in D(0, R)$ . Comme  $|z_0| < R$ , on peut trouver  $r$  tel que  $|z_0| < r < R$ , par exemple  $r = (R + |z_0|)/2$ . Par le lemme 2.34 appliqué avec  $P(k) = k$ , le rayon de convergence de la série entière  $\sum_{k=1}^{+\infty} k a_k z^{k-1}$  est  $R$ . Cette série converge donc normalement sur  $\overline{D}(0, r)$ . Donc  $S'_n$  converge uniformément vers la somme de cette série sur  $\overline{D}(0, r)$ . Par le théorème 2.20,  $S$  est dérivable au point  $z_0$  et

$$S'(z_0) = \lim_{n \rightarrow +\infty} S'_n(z_0) = \sum_{k=1}^{+\infty} k a_k z_0^{k-1}.$$

Ceci étant valable pour tout  $z_0 \in D(0, R)$ , le théorème est démontré.  $\square$

**Corollaire 2.40.** *Soit  $S(z) = \sum_{k=0}^{+\infty} a_k z^k$  une série entière de rayon de convergence  $R > 0$ . La fonction  $S$  est indéfiniment dérivable sur  $D(0, R)$  et ses dérivées successives sont données par :*

$$\forall \ell \in \mathbb{N}^*, \quad \forall z \in D(0, R), \quad S^{(\ell)}(z) = \sum_{k=\ell}^{+\infty} k(k-1) \dots (k-\ell+1) a_k z^{k-\ell}. \quad (2.17)$$

Chaque série entière  $S^{(\ell)}(z)$  a même rayon de convergence que  $S$ .

Ce corollaire se vérifie en appliquant le th. 2.39 avec  $S'$  à la place de  $S$ , ce qui donne le résultat pour  $S'' = S^{(2)}$ . Notons que si on remplace  $S$  par  $S'$ , le rayon de convergence est inchangé, de même quand on passe de  $S'$  à  $S''$ . En itérant ce processus, on obtient le résultat souhaité.

6. Après avoir remarqué que la série entière  $\sum_{k=1}^{+\infty} k a_k z^k = z \sum_{k=1}^{+\infty} k a_k z^{k-1}$  a même rayon de convergence que  $\sum_{k=1}^{+\infty} k a_k z^{k-1}$ .

**Corollaire 2.41** (primitivation terme à terme). Soit  $S(z) = \sum_{k=0}^{+\infty} a_k z^k$  une série entière de rayon de convergence  $R > 0$ . Alors la série entière

$$T(z) = \sum_{k=0}^{+\infty} \frac{a_k}{k+1} z^{k+1}$$

a même rayon de convergence  $R$  que  $S$ , est sa dérivée sur  $D(0, R)$  est  $S$ .

La preuve est laissée en exercice.

Voici une application de la dérivation des séries entières à un calcul de somme de série qui nous sera utile en probabilités (espérance d'une loi géométrique).

**Exemple 2.42** (dérivée de la série géométrique standard). La série géométrique standard  $S(z) = \sum_{k=0}^{+\infty} z^k$  a pour rayon de convergence 1, cf. l'exemple 2.10 p. 26. On dispose pour sa somme de la formule explicite :

$$\forall z \in D(0, 1), \quad S(z) = \sum_{k=0}^{+\infty} z^k = \frac{1}{1-z}.$$

On peut calculer de deux façons  $S'(z)$ .

1. En appliquant le théorème de dérivation terme à terme (th. 2.39) on obtient :

$$\forall z \in D(0, 1), \quad S'(z) = \sum_{k=1}^{+\infty} k z^{k-1}. \quad (2.18)$$

2. En dérivant directement la fonction  $F : z \mapsto 1/(1-z)$  comme dérivée de l'inverse d'une fonction polynôme :

$$\forall z \in D(0, 1), \quad S'(z) = F'(z) = \frac{1}{(1-z)^2}. \quad (2.19)$$

En comparant (2.18) et (2.19), on obtient :

$$\forall z \in D(0, 1), \quad \sum_{k=1}^{+\infty} k z^{k-1} = \frac{1}{(1-z)^2}.$$

En particulier pour  $z = q \in ]0, 1[$  et en posant  $p = 1 - q$ , on obtient

$$\forall q \in ]0, 1[, \quad \sum_{k=1}^{+\infty} k q^{k-1} p = \frac{p}{(1-q)^2} = \frac{1}{p}. \quad (2.20)$$

C'est cette formule qui donne l'espérance du temps d'attente du premier succès dans une suite d'épreuves répétées indépendantes (par exemple lancer d'un dé) avec pour chacune probabilité de succès  $p$ .

## 2.5 Développements en série entière

Quand une série entière a un rayon de convergence non nul, sa somme définit une fonction sur le disque de convergence  $D(0, R)$  et cette fonction est indéfiniment dérivable sur  $D(0, R)$  d'après le corollaire 2.40. On examine maintenant le problème inverse du développement en série entière d'une fonction  $f$  donnée.



**Définition 2.43.** Soit  $f$  une fonction  $f$  définie sur un disque ouvert contenant 0. On dit qu'elle est développable en série entière au voisinage de 0 s'il existe une série entière  $\sum_{k=0}^{+\infty} a_k z^k$  de rayon de convergence  $R > 0$  et un réel  $r \in ]0, R[$  tel que

$$\forall z \in D(0, r), \quad f(z) = \sum_{k=0}^{+\infty} a_k z^k. \quad (2.21)$$

**Remarque 2.44.** Cette définition se spécialise naturellement au cas où  $f$  est de variable réelle  $x$  en remplaçant le systématiquement le mot « disque » par « intervalle ». Nous ferons l'étude pour  $f$  de variable complexe  $z$ , même si en pratique les résultats les plus courants sont utilisés en variable réelle.

**Remarque 2.45.** La série  $\sum_{k=0}^{+\infty} a_k z^k$  dans (2.21) est appelée développement en série entière de  $f$ . Un tel développement lorsqu'il existe est unique. En effet d'après le corollaire 2.40,  $f$  est nécessairement indéfiniment dérivable sur  $D(0, r)$  et en appliquant la formule (2.17) du corollaire, on obtient :

$$\begin{aligned} \forall n \in \mathbb{N}, \forall z \in D(0, r), \quad f^{(n)}(z) &= \sum_{k=n}^{+\infty} k(k-1) \dots (k-n+1) a_k z^{k-n} \\ &= n! a_n + z \sum_{k=n+1}^{+\infty} k(k-1) \dots (k-n+1) a_k z^{k-n-1}. \end{aligned}$$

En particulier pour  $z = 0$ , ceci s'écrit

$$\forall n \in \mathbb{N}, \quad f^{(n)}(0) = n! a_n. \quad (2.22)$$

Donc si  $f$  est développable en série entière au voisinage de 0, les coefficients  $a_k$  de son développement vérifient :

$$\forall k \in \mathbb{N}, \quad a_k = \frac{f^{(k)}(0)}{k!}. \quad (2.23)$$

Ceci montre bien que les coefficients d'un éventuel développement de  $f$  en série entière sont parfaitement déterminés dès que l'on connaît  $f$ , ce qui prouve l'unicité du développement en série entière de  $f$ , sous réserve d'existence.

**Définition 2.46** (série de Taylor). Soit  $f$  une fonction à valeurs complexes définie sur un disque ouvert contenant 0 et indéfiniment dérivable. On appelle série de Taylor de  $f$  en 0, la série entière

$$\sum_{k=0}^{+\infty} \frac{f^{(k)}(0)}{k!} z^k.$$

Nous pouvons maintenant reformuler le problème du développement en série entière comme suit.

**Proposition 2.47.** Une fonction  $f$  définie sur un disque ouvert contenant 0 est développable en série entière au voisinage de 0 si

- $f$  est indéfiniment dérivable sur un disque ouvert  $D(0, R_1)$  ;
- sa série de Taylor en 0 a un rayon de convergence  $R_2 > 0$  ;
- il existe un disque ouvert  $D(0, r)$  avec  $0 < r \leq \min(R_1, R_2)$  sur lequel  $f$  soit égale à la somme de sa série de Taylor en 0, c'est-à-dire :

$$\forall z \in D(0, r), \quad f(z) = \sum_{k=0}^{+\infty} \frac{f^{(k)}(0)}{k!} z^k.$$

Pour pouvoir approximer  $f(z)$  par la somme partielle d'ordre  $n$  de sa série de Taylor, nous aurons besoin de la version suivante de la formule de Taylor avec reste intégral pour des fonctions d'une variable réelle à valeurs complexes.

**Théorème 2.48.** *Soit  $r > 0$  et  $g$  une fonction  $] -r, r[ \rightarrow \mathbb{C}$ ,  $n + 1$  fois dérivable sur  $] -r, r[$  et telle que  $g^{(n+1)}$  soit continue sur  $] -r, r[$ . Alors*

$$\forall s \in ] -r, r[, \quad g(s) = g(0) + \sum_{k=1}^n \frac{s^k}{k!} g^{(k)}(0) + s^{n+1} \int_0^1 \frac{(1-t)^{n+1}}{n!} g^{(n+1)}(ts) dt. \quad (2.24)$$

Soit  $f$  définie sur  $D(0, R_1)$  et indéfiniment dérivable et fixons un nombre complexe  $z \neq 0$  quelconque dans le disque  $D(0, R_1)$ . Nous allons exploiter la formule (2.24) en paramétrant le segment de droite d'extrémités 0 et  $z$  dans  $\mathbb{C}$ , tout simplement en écrivant qu'un point du plan complexe appartient à ce segment si et seulement s'il peut s'écrire  $sz$  avec un  $s \in [0, 1]$ . On pose ensuite  $g(s) = f(sz)$ . Alors puisque  $|z| < R_1$ ,  $g$  est définie sur l'intervalle  $] -r, r[$  avec  $r = R_1/|z|$  et hérite de l'indéfinie dérivabilité de  $f$  par dérivation de fonctions composées. On vérifie facilement que pour tout  $s \in ] -r, r[$  :

$$g'(s) = z f'(sz), g''(s) = z^2 f''(sz), \dots, g^{(k)}(s) = z^k f^{(k)}(sz), \dots$$

En particulier, pour tout entier  $k$ ,  $g^{(k)}(0) = z^k f^{(k)}(0)$ . En appliquant maintenant la formule (2.24) avec  $s = 1$ , on obtient :

$$f(z) = f(0) + \sum_{k=1}^n \frac{z^k}{k!} f^{(k)}(0) + z^{n+1} \int_0^1 \frac{(1-t)^{n+1}}{n!} f^{(n+1)}(tz) dt. \quad (2.25)$$

Comme  $z \neq 0$  était quelconque dans  $D(0, R_1)$  et comme la formule ci-dessus se réduit à  $f(0) = f(0)$  quand  $z = 0$ , (2.25) est valide pour tout  $z \in D(0, R_1)$ .

En notant  $S_n^{\text{Tay}}(f, z)$  la somme partielle de rang  $n$  de la série de Taylor de  $f$ , on peut réécrire (2.25) sous la forme

$$\forall z \in D(0, R_1), \quad f(z) = S_n^{\text{Tay}}(f, z) + z^{n+1} \int_0^1 \frac{(1-t)^{n+1}}{n!} f^{(n+1)}(tz) dt. \quad (2.26)$$

On peut synthétiser cette discussion par la condition nécessaire et suffisante suivante.

**Théorème 2.49.** *Soit  $f$  une fonction à valeurs complexes définie et indéfiniment dérivable sur un disque ouvert de  $\mathbb{C}$  contenant 0. Alors  $f$  est développable en série entière sur un disque  $D(0, r)$  si et seulement si :*

$$\forall z \in D(0, r), \quad \lim_{n \rightarrow +\infty} \frac{z^{n+1}}{n!} \int_0^1 (1-t)^{n+1} f^{(n+1)}(tz) dt = 0. \quad (2.27)$$

*Preuve.* Supposons d'abord que  $f$  est développable en série entière sur le disque  $D(0, r)$ . Alors pour tout  $z \in D(0, r)$ ,  $f(z) = \sum_{k=0}^{+\infty} a_k z^k$ , le rayon de convergence de cette série est au moins  $r$ ,  $f$  est indéfiniment dérivable sur  $D(0, r)$  et d'après (2.23), les  $a_k$  vérifient  $a_k = f^{(k)}(0)/k!$ , d'où  $\sum_{k=0}^n a_k = S_n^{\text{Tay}}(f, z)$ . Comme  $f$  vérifie (2.26) sur  $D(0, r)$ , la convergence de la série  $\sum_{k=0}^{+\infty} a_k z^k$  sur  $D(0, r)$  implique la convergence de son reste vers 0, ce qui donne (2.27).

Réciproquement, supposons  $f$  définie et indéfiniment dérivable sur  $D(0, r)$ . Elle vérifie donc (2.26) sur  $D(0, r)$ . Supposons de plus que (2.27) est vérifiée pour tout  $z \in D(0, r)$ . Alors pour tout  $z \in D(0, r)$ ,  $S_n^{\text{Tay}}(f, z)$  tend vers  $f(z)$  quand  $n$  tend vers l'infini. Cela signifie que la série de Taylor de  $f$  converge vers  $f$  sur  $D(0, r)$ . Donc  $f$  est bien développable en série entière. Le rayon de convergence de sa série de Taylor est alors supérieur ou égal à  $r$ .  $\square$

**Corollaire 2.50.** *Si pour un  $r > 0$ ,  $f$  est indéfiniment dérivable sur  $D(0, r)$  et vérifie pour une constante  $C > 0$  :*

$$\forall n \in \mathbb{N}, \forall z \in D(0, r), \quad |f^{(n)}(z)| \leq C,$$

*alors  $f$  est la somme de sa série de Taylor sur  $D(0, r)$  et le rayon de convergence de cette série est supérieur ou égal à  $r$ .*

Dans la condition suffisante ci-dessus, la constante  $C$  ne doit évidemment pas dépendre de  $n$  ni de  $z$ , mais peut dépendre de  $r$ .

*Preuve.* On applique le théorème 2.49 en vérifiant la condition (2.27) par la majoration suivante.

$$\begin{aligned} \left| \frac{z^{n+1}}{n!} \int_0^1 (1-t)^{n+1} f^{(n+1)}(tz) dt \right| &\leq \frac{r^{n+1}}{n!} \int_0^1 (1-t)^{n+1} |f^{(n+1)}(tz)| dt \\ &\leq C \frac{r^{n+1}}{n!} \int_0^1 (1-t)^{n+1} dt = C \frac{r^{n+1}}{n!(n+2)} \leq C \frac{r^{n+1}}{(n+1)!}. \end{aligned}$$

On se ramène ainsi à vérifier que  $r^{n+1}/(n+1)!$  tend vers 0 quand  $n$  tend vers l'infini. Comme  $r$  est ici une constante, on peut trouver un entier  $n_0 > r$ . Alors

$$\forall n \geq n_0, \quad \frac{r^{n+1}}{(n+1)!} = \frac{r^{n_0+1}}{(n_0+1)!} \times \frac{r^{n-n_0}}{(n_0+2) \times \cdots \times (n+1)}.$$

En introduisant les constantes  $C_0 = r^{n_0+1}/(n_0+1)!$  et  $q = r_0/(n_0+2) < 1$ , on en déduit que

$$\forall n \geq n_0, \quad \frac{r^{n+1}}{(n+1)!} \leq C_0 q^{n-n_0}.$$

Ce dernier majorant étant le terme général d'une suite géométrique de raison positive  $q < 1$ , tend vers 0.  $\square$

Par restriction aux fonctions d'une variable réelle, le corollaire 2.50 permet d'obtenir le développement en série entière de la plupart des fonctions usuelles.

**Exemple 2.51** (développement en série entière des fonctions sinus et cosinus). La fonction sinus a pour dérivée la fonction cosinus, laquelle a pour dérivée la fonction sinus. La fonction sinus est donc indéfiniment dérivable sur  $\mathbb{R}$  et la suite de ses dérivées successives se reproduit périodiquement tous les 4 termes consécutifs. Autrement dit en classant l'entier  $k$  selon le reste de sa division par 4 :

$$\forall x \in \mathbb{R}, \forall k \in \mathbb{N}, \quad \sin^{(k)} x = \begin{cases} \sin x & \text{si } k = 4j, \\ \cos x & \text{si } k = 4j + 1, \\ -\sin x & \text{si } k = 4j + 2, \\ -\cos x & \text{si } k = 4j + 3. \end{cases}$$

Comme les fonctions sinus et cosinus sont bornées en valeur absolue par 1 sur  $\mathbb{R}$ , on peut appliquer la version variable réelle du corollaire 2.50 qui nous dit ici que la fonction sinus est la somme de sa série de Taylor sur tout  $\mathbb{R}$ . En remarquant que

$$\forall k \in \mathbb{N}, \quad \sin^{(k)}(0) = \begin{cases} 0 & \text{si } k = 4j \text{ ou } 4j + 2, \\ 1 & \text{si } k = 4j + 1, \\ -1 & \text{si } k = 4j + 3, \end{cases}$$

on obtient le développement en série entière :

$$\forall x \in \mathbb{R}, \quad \sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \frac{x^9}{9!} - \frac{x^{11}}{11!} + \dots = \sum_{l=0}^{+\infty} \frac{(-1)^l}{(2l+1)!} x^{2l+1}. \quad (2.28)$$

Par la même méthode (faites-le), on obtient aussi le développement en série entière de la fonction cosinus :

$$\forall x \in \mathbb{R}, \quad \cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \frac{x^8}{8!} - \frac{x^{10}}{10!} + \dots = \sum_{l=0}^{+\infty} \frac{(-1)^l}{(2l)!} x^{2l}. \quad (2.29)$$

**Exemple 2.52** (développement en série entière de la fonction exponentielle). La fonction exponentielle est indéfiniment dérivable sur  $\mathbb{R}$  et toutes ses dérivées successives sont égales :

$$\forall x \in \mathbb{R}, \forall k \in \mathbb{N}, \quad \exp^{(k)}(x) = \exp(x).$$

En particulier, pour tout  $k \in \mathbb{N}$ ,  $\exp^{(k)}(x) = \exp(0) = 1$ . Soit  $r > 0$  quelconque. Comme la fonction  $\exp$  est croissante et positive, on a la majoration :

$$\forall x \in ]-r, r[, \forall n \in \mathbb{N}, \quad |\exp^{(n)}(x)| = \exp(x) \leq \exp(r).$$

Comme ce majorant est indépendant de  $n$  et de  $x$ , on est dans les conditions d'application du corollaire 2.50 et on obtient le développement en série de Taylor :

$$\exp(x) = \sum_{k=0}^{+\infty} \frac{x^k}{k!}, \quad (2.30)$$

valable pour tout  $x \in ]-r, r[$ . Mais comme  $r$  était quelconque, ce développement est valable en fait pour tout  $x$  réel.

**Remarque 2.53.** Nous avons déjà rencontré la série entière au second membre de (2.30) et vu qu'elle a un rayon de convergence infini et vérifié, sans savoir qu'elle représentait la fonction exponentielle, que sa somme  $f(x)$  vérifie pour tous réels  $a$  et  $b$  la relation  $f(a+b) = f(a)f(b)$ . On peut aussi remarquer que si on dérive terme à terme la série entière  $f(x)$ , on obtient bien  $f'(x) = f(x)$ . Ces deux propriétés subsistent pour la série entière  $f(z)$  avec  $z \in \mathbb{C}$ . Cela légitime l'extension de  $\mathbb{R}$  à  $\mathbb{C}$  de la fonction exponentielle via la définition suivante.

**Définition 2.54** (fonction exponentielle complexe). On définit la fonction exponentielle complexe  $\exp$  par :

$$\forall z \in \mathbb{C}, \quad \exp(z) := \sum_{k=0}^{+\infty} \frac{z^k}{k!}. \quad (2.31)$$

Comme pour la fonction exponentielle de variable réelle, on pourra utiliser la notation abrégée  $\exp(z) = e^z$ .

**Remarque 2.55.** En classe de Terminale, on introduit sans pouvoir la démontrer la formule

$$\forall y \in \mathbb{R}, \quad e^{iy} = \cos y + i \sin y. \quad (2.32)$$

Avec la définition que nous venons de donner de l'exponentielle complexe, nous pouvons maintenant établir cette formule. En effet en prenant  $z = iy$  dans (2.31), on obtient :

$$\forall y \in \mathbb{R}, \quad e^{iy} = \sum_{k=0}^{+\infty} \frac{i^k y^k}{k!} = 1 + iy - \frac{y^2}{2!} - i \frac{y^3}{3!} + \frac{y^4}{4!} + i \frac{y^5}{5!} - \frac{y^6}{6!} + \dots$$

En séparant partie réelle et partie imaginaire de cette série convergente, cf. proposition 2.5, il vient :

$$\begin{aligned}\forall y \in \mathbb{R}, \quad e^{iy} &= \left( 1 - \frac{y^2}{2!} + \frac{y^4}{4!} - \frac{y^6}{6!} + \frac{y^8}{8!} + \dots \right) + i \left( y - \frac{y^3}{3!} + \frac{y^5}{5!} - \frac{y^7}{7!} + \frac{y^9}{9!} + \dots \right) \\ &= \sum_{l=0}^{+\infty} \frac{(-1)^l}{(2l)!} y^{2l} + i \sum_{l=0}^{+\infty} \frac{(-1)^l}{(2l+1)!} y^{2l+1} \\ &= \cos y + i \sin y,\end{aligned}$$

compte-tenu des développements (2.28) et (2.29) déjà obtenus.

*Cette section sera complétée ultérieurement.*



# Chapitre 3

## Séries de Fourier

### 3.1 Séries trigonométriques

**Définition 3.1.** On appelle *série trigonométrique* une série de fonctions d'une variable réelle dont le terme général  $u_k : \mathbb{R} \rightarrow \mathbb{R}$ ,  $x \mapsto u_k(x)$  est de la forme

$$u_k(x) = a_k \cos(k\omega x) + b_k \sin(k\omega x), \quad k \in \mathbb{N}, \quad a_k, b_k \in \mathbb{R},$$

où  $\omega \in \mathbb{R}_+^*$  est une constante.

Pour  $k = 0$ ,  $u_0$  se réduit à la fonction constante  $u_0 = a_0$ , la valeur de  $b_0$  n'intervient pas dans la somme de la série. Par convention nous prendrons  $b_0 = 0$ .

Lorsque la série trigonométrique

$$S(x) = \sum_{k=0}^{+\infty} (a_k \cos(k\omega x) + b_k \sin(k\omega x))$$

converge pour tout  $x \in \mathbb{R}$ , la fonction  $S : x \mapsto S(x)$  est périodique de période

$$T = \frac{2\pi}{\omega}.$$

Pour démontrer cette affirmation, il suffit de vérifier que chaque  $u_k$  est périodique de période  $T$ .

$$\begin{aligned} u_k(x + T) &= a_k \cos(k\omega(x + T)) + b_k \sin(k\omega(x + T)) \\ &= a_k \cos(k\omega x + k\omega T) + b_k \sin(k\omega x + k\omega T) \\ &= a_k \cos(k\omega x + 2k\pi) + b_k \sin(k\omega x + 2k\pi) \\ &= a_k \cos(k\omega x) + b_k \sin(k\omega x) = u_k(x). \end{aligned}$$

Ceci étant vrai pour tout  $x \in \mathbb{R}$  et tout  $k \in \mathbb{N}$ , on en déduit que

$$S(x + T) = \sum_{k=0}^{+\infty} u_k(x + T) = \sum_{k=0}^{+\infty} u_k(x) = S(x),$$

ce qui montre que  $S$  est  $T$ -périodique.

**Proposition 3.2.** Lorsqu'elle converge, la série trigonométrique peut s'écrire aussi sous la forme de série d'exponentielles complexes (forme complexe de la série trigonométrique) :

$$S(x) = \sum_{k=0}^{+\infty} (a_k \cos(k\omega x) + b_k \sin(k\omega x)) = \sum_{k \in \mathbb{Z}} c_k e^{ik\omega x} := \lim_{n \rightarrow +\infty} \sum_{k=-n}^{k=n} c_k e^{ik\omega x},$$

les  $c_k$  étant reliés aux  $a_k, b_k$  par les formules suivantes valides pour tout  $k \in \mathbb{N}^*$  :

$$c_k = \frac{a_k - ib_k}{2}, \quad c_{-k} = \frac{a_k + ib_k}{2}, \quad (3.1)$$

$$a_k = c_k + c_{-k}, \quad b_k = i(c_k - c_{-k}), \quad (3.2)$$

avec  $a_0 = c_0$  et  $b_0 = 0$ .

*Preuve.* Pour tout  $k \in \mathbb{N}^*$ ,

$$\begin{aligned} a_k \cos(k\omega x) + b_k \sin(k\omega x) &= a_k \frac{e^{ik\omega x} + e^{-ik\omega x}}{2} + b_k \frac{e^{ik\omega x} - e^{-ik\omega x}}{2i} \\ &= \left( \frac{a_k}{2} + \frac{b_k}{2i} \right) e^{ik\omega x} + \left( \frac{a_k}{2} - \frac{b_k}{2i} \right) e^{-ik\omega x} \\ &= \left( \frac{a_k}{2} - \frac{ib_k}{2} \right) e^{ik\omega x} + \left( \frac{a_k}{2} + \frac{ib_k}{2} \right) e^{-ik\omega x} \\ &= c_k e^{ik\omega x} + c_{-k} e^{-ik\omega x}. \end{aligned}$$

Dans le cas particulier  $k = 0$ ,

$$a_0 \cos(0 \times \omega x) + b_0 \sin(0 \times \omega x) = a_0 = a_0 e^{i \times 0 \times \omega x},$$

d'où  $c_0 = a_0$ . □

Le cas facile pour les séries trigonométriques est celui où les séries numériques de terme général respectif  $a_k$  et  $b_k$  convergent absolument.

**Théorème 3.3** (convergence normale). *Si les séries numériques  $\sum_{k=0}^{+\infty} |a_k|$  et  $\sum_{k=0}^{+\infty} |b_k|$  convergent, la série de fonctions  $\sum_{k=0}^{+\infty} (a_k \cos(k\omega x) + b_k \sin(k\omega x))$  converge normalement sur  $\mathbb{R}$ .*

*Preuve.* En notant que pour tout  $x \in \mathbb{R}$ ,  $|\cos(k\omega x)|$  et  $|\sin(k\omega x)|$  sont majorés par 1, on obtient :

$$\begin{aligned} |a_k \cos(k\omega x) + b_k \sin(k\omega x)| &\leq |a_k \cos(k\omega x)| + |b_k \sin(k\omega x)| \\ &= |a_k| |\cos(k\omega x)| + |b_k| |\sin(k\omega x)| \\ &\leq |a_k| + |b_k|. \end{aligned}$$

Ce dernier majorant est indépendant de  $x$ . De plus,  $\sum_{k=0}^{+\infty} (|a_k| + |b_k|)$  converge comme somme de deux séries numériques convergentes. Donc la série trigonométrique considérée converge normalement sur  $\mathbb{R}$ . □

#### Corollaire 3.4.

1. Si les séries  $\sum_{k=0}^{+\infty} |a_k|$  et  $\sum_{k=0}^{+\infty} |b_k|$  convergent,
  - a) la fonction  $S : \mathbb{R} \rightarrow \mathbb{R}$ ,  $x \mapsto S(x) = \sum_{k=0}^{+\infty} (a_k \cos(k\omega x) + b_k \sin(k\omega x))$  est continue sur  $\mathbb{R}$  ;
  - b) pour tout  $x \in \mathbb{R}$ ,

$$\begin{aligned} \int_0^x S(t) dt &= \sum_{k=0}^{+\infty} \int_0^x (a_k \cos(k\omega t) + b_k \sin(k\omega t)) dt \\ &= \sum_{k=1}^{+\infty} \frac{b_k}{k\omega} + a_0 x + \sum_{k=1}^{+\infty} \left( \frac{a_k}{k\omega} \sin(k\omega x) - \frac{b_k}{k\omega} \cos(k\omega x) \right). \end{aligned} \quad (3.3)$$



2. Si les séries  $\sum_{k=0}^{+\infty} k|a_k|$  et  $\sum_{k=0}^{+\infty} k|b_k|$  convergent, alors  $S$  est dérivable sur  $\mathbb{R}$  et pour tout  $x \in \mathbb{R}$ ,

$$S'(x) = \sum_{k=0}^{+\infty} u'_k(x) = \sum_{k=0}^{+\infty} (-ka_k\omega \sin(k\omega x) + kb_k\omega \cos(k\omega x)). \quad (3.4)$$

*Esquisse de preuve.* On sait que la convergence normale implique la convergence uniforme et que la continuité (ici celle des sommes partielles de la série trigonométrique) est préservée par convergence uniforme, cf. th. 2.18, l'adaptation au cas considéré étant laissée au lecteur. La possibilité d'intégrer terme à terme la série trigonométrique se justifie comme suit. Fixons  $x$  quelconque et notons  $S_n(t)$  la somme partielle de rang  $n$  :

$$S_n(t) = \sum_{k=0}^n (a_k \cos(k\omega t) + b_k \sin(k\omega t)).$$

Par additivité de l'intégrale,

$$\forall n \in \mathbb{N}, \quad \int_0^x S_n(t) dt = \sum_{k=0}^n \int_0^x (a_k \cos(k\omega t) + b_k \sin(k\omega t)) dt. \quad (3.5)$$

Si on montre que  $\int_0^x S_n(t) dt$  converge vers  $\int_0^x S(t) dt$  quand  $n$  tend vers l'infini, on obtiendra (3.3) en faisant tendre  $n$  vers l'infini dans (3.5). On sait que  $S_n$  converge normalement, donc aussi uniformément, vers  $S$  sur  $\mathbb{R}$ . Traduisons cette convergence uniforme :

$$\forall \varepsilon > 0, \exists n_0, \forall n \geq n_0, \forall t \in \mathbb{R}, \quad |S_n(t) - S(t)| < \varepsilon.$$

On en déduit que si  $x > 0$ ,

$$\left| \int_0^x S_n(t) dt - \int_0^x S(t) dt \right| \leq \int_0^x |S_n(t) - S(t)| dt \leq \int_0^x \varepsilon dt = \varepsilon x.$$

Le cas  $x < 0$  est analogue mais en remplaçant  $\int_0^x$  par  $-\int_x^0$  et majorant obtenu est  $\varepsilon|x|$ . Le cas  $x = 0$  étant trivial, on obtient dans tous les cas :  $\left| \int_0^x S_n(t) dt - \int_0^x S(t) dt \right| \leq |x|\varepsilon$ . Et comme  $x$  était fixé, on a bien montré que  $\int_0^x S_n(t) dt$  converge vers  $\int_0^x S(t) dt$  quand  $n$  tend vers l'infini.

Pour le point 2, on commence par noter que si  $\sum_{k=0}^{+\infty} k|a_k|$  et  $\sum_{k=0}^{+\infty} k|b_k|$  convergent, les séries  $\sum_{k=0}^{+\infty} |a_k|$  et  $\sum_{k=0}^{+\infty} |b_k|$  convergent aussi par le théorème de comparaison des séries à termes positifs. Ainsi la série des dérivées terme à terme converge uniformément sur  $\mathbb{R}$  et la série initiale  $S$  converge sur  $\mathbb{R}$ . Par le théorème de dérivabilité par convergence uniforme (une version adaptée du th. 2.20), appliqué à la suite  $S_n$ , la fonction  $S$  est dérivable sur  $\mathbb{R}$  et sa dérivée peut se calculer par dérivation terme à terme de la série définissant  $S$ , autrement dit, vérifie (3.4).  $\square$

En dehors du cas confortable où les séries  $\sum_{k=0}^{+\infty} |a_k|$  et  $\sum_{k=0}^{+\infty} |b_k|$  convergent, il y a d'autres possibilités pour obtenir la convergence d'une série trigonométrique. Nous nous contenterons du résultat suivant qui est une conséquence directe du théorème d'Abel (th. 1.39).

**Théorème 3.5.** *Si les suites réelles  $(a_k)_{k \in \mathbb{N}}$  et  $(b_k)_{k \in \mathbb{N}}$  sont décroissantes à partir d'un certain rang et tendent vers 0, la série trigonométrique  $\sum_{k=0}^{+\infty} (a_k \cos(k\omega x) + b_k \sin(k\omega x))$  converge pour tout  $x \in \mathbb{R}$ , sauf éventuellement les  $x$  de la forme  $jT$  pour  $j \in \mathbb{Z}$ .*

Lorsque la série trigonométrique associée aux suites réelles  $(a_k)_{k \in \mathbb{N}}$  et  $(b_k)_{k \in \mathbb{N}}$  converge sur  $\mathbb{R}$  (sauf peut-être en des points isolés) vers une fonction  $S$ , il est naturel de se demander si la connaissance de la fonction  $S$  permet de retrouver les coefficients  $a_k$  et  $b_k$ . Pour répondre à cette question, la proposition suivante est un préliminaire commode.

**Proposition 3.6.** *Soit  $f : \mathbb{R} \rightarrow \mathbb{R}$ , une fonction  $T$ -périodique et Riemann intégrable sur tout intervalle fermé borné de  $\mathbb{R}$ . Alors pour tous  $a, b \in \mathbb{R}$ ,*

$$\int_a^{a+T} f(t) dt = \int_b^{b+T} f(t) dt.$$

*Autrement dit, l'intégrale  $\int_a^{a+T} f(t) dt$  ne dépend pas de  $a$ . Dans la suite on notera cette intégrale  $\int_T f(t) dt$ .*

*Preuve.* Par la relation de Chasles pour l'intégrale,  $\int_a^{a+T} = \int_a^b + \int_b^{b+T} + \int_{b+T}^{a+T}$ . De plus,  $f$  étant  $T$ -périodique,  $f(s+T) = f(s)$ , donc par le changement de variable  $t = s+T$ ,

$$\int_{b+T}^{a+T} f(t) dt = \int_b^a f(s+T) ds = \int_b^a f(s) ds = - \int_a^b f(t) dt.$$

en revenant à la relation de Chasles, on en déduit :

$$\int_a^{a+T} f(t) dt = \int_a^b f(t) dt + \int_b^{b+T} f(t) dt - \int_a^b f(t) dt = \int_b^{b+T} f(t) dt.$$

□

La proposition 3.6 permet de choisir judicieusement l'intervalle d'intégration pour exploiter une éventuelle parité ou imparité de  $f$  pour calculer l'intégrale  $\int_T f(t) dt$ .

**Théorème 3.7.** *Pour toute série trigonométrique  $S(x) = \sum_{k=0}^{+\infty} (a_k \cos(k\omega x) + b_k \sin(k\omega x))$  telle que  $\sum_{k=0}^{+\infty} |a_k|$  et  $\sum_{k=0}^{+\infty} |b_k|$  convergent,*

$$a_0 = \frac{1}{T} \int_T S(t) dt \tag{3.6}$$

*et pour tout  $k \geq 1$ ,*

$$a_k = \frac{2}{T} \int_T S(t) \cos(k\omega t) dt, \quad b_k = \frac{2}{T} \int_T S(t) \sin(k\omega t) dt. \tag{3.7}$$

*En utilisant la représentation complexe de la série trigonométrique  $S(x) = \sum_{k \in \mathbb{Z}} c_k e^{ik\omega x}$ , on a*

$$\forall k \in \mathbb{Z}, \quad c_k = \frac{1}{T} \int_T S(t) e^{-ik\omega t} dt. \tag{3.8}$$

*Preuve.* Il est plus commode de démontrer (3.8), les formules pour les  $a_k$  et  $b_k$  s'en déduisent facilement grâce à aux égalités (3.2). En justifiant l'interversion série intégrale par un argument de convergence uniforme analogue à celui employé dans la preuve du corollaire 3.4-1 b), détails laissés en exercice, on a

$$\frac{1}{T} \int_T S(t) e^{-ik\omega t} dt = \sum_{j \in \mathbb{Z}} c_j \frac{1}{T} \int_T e^{i(j-k)\omega t} dt. \tag{3.9}$$

On est ainsi amenés à calculer les intégrales  $I_m = \int_T e^{im\omega t} dt$  pour  $m \in \mathbb{Z}$ . Le cas particulier  $m = 0$  donne immédiatement  $I_0 = T$ . Pour  $m \neq 0$ , en rappelant que  $T = 2\pi/\omega$ , on a :

$$\forall m \in \mathbb{Z}^*, \quad I_m = \int_0^T e^{im\omega t} dt = \left[ \frac{e^{im\omega t}}{im\omega} \right]_0^T = \frac{1}{im\omega} (e^{im\omega T} - 1) = \frac{1}{im\omega} (e^{2im\pi} - 1) = 0.$$

En posant  $j - k = m$ , on en déduit que

$$\frac{1}{T} \int_T e^{i(j-k)\omega t} dt = \begin{cases} 0 & \text{si } j \neq k, \\ 1 & \text{si } j = k. \end{cases} \quad (3.10)$$

La série au second membre de (3.9) a donc tous ses termes d'indice  $j \neq k$  nuls et le terme d'indice  $j = k$  est égal à  $c_k$ . La somme de la série est donc  $c_k$ , ce qui établit la formule (3.8).  $\square$

## 3.2 Développement en série de Fourier

Après avoir étudié les propriétés des séries trigonométriques, on s'intéresse maintenant au problème inverse (analogue au problème du développement en série entière d'une fonction donnée vu au chapitre précédent). Étant donnée une fonction périodique  $f$ , peut-on l'écrire comme la somme d'une série trigonométrique convergente? On a vu au théorème 3.7, que si une série trigonométrique converge normalement, ses coefficients sont reliés à sa somme par les relations (3.6)–(3.8). Ces formules nous fournissent donc, en remplaçant  $S$  par  $f$ , des candidats naturels pour les coefficients d'un développement de  $f$  en série trigonométrique. Ceci motive la définition suivante.

**Définition 3.8** (série de Fourier d'une fonction). *Soit  $f : \mathbb{R} \rightarrow \mathbb{R}$ , une fonction  $T$ -périodique et intégrable sur  $[0, T]$ . On appelle série de Fourier de  $f$ , la série trigonométrique*

$$S(f, x) := \sum_{k=0}^{+\infty} (a_k(f) \cos(k\omega x) + b_k(f) \sin(k\omega x)),$$

où les  $a_k(f)$  et  $b_k(f)$  sont donnés par

$$a_0(f) = \frac{1}{T} \int_T f(t) dt, \quad b_0(f) = 0$$

et pour tout  $k \in \mathbb{N}^*$ ,

$$a_k(f) = \frac{2}{T} \int_T f(t) \cos(k\omega t) dt, \quad b_k(f) = \frac{2}{T} \int_T f(t) \sin(k\omega t) dt.$$

En notation complexe, ceci s'écrit :

$$S(f, x) = \sum_{k \in \mathbb{Z}} c_k(f) e^{ik\omega x}, \quad c_k(f) = \frac{1}{T} \int_T f(t) e^{-ik\omega t} dt.$$

La question de savoir pour quelles fonctions périodiques  $f$  la série de Fourier converge (et en quel sens) vers  $f$  a été le point de départ de nombreuses recherches en analyse mathématique. Au niveau de ce cours, nous nous contenterons du théorème de Dirichlet qui concerne les fonction de classe  $C^1$  par morceaux, ce qui est suffisant pour les applications courantes.

**Définition 3.9.** On dit que la fonction  $T$ -périodique est  $C^1$  par morceaux si on peut découper l'intervalle  $]0, T[$  (ou n'importe quel intervalle  $]a, a + T[$ ) en un nombre fini d'intervalles  $]a_{i-1}, a_i[$  avec  $0 = a_0 < a_1 < \dots < a_d = T$  tels que pour tout  $1 \leq i \leq d$ ,  $f$  soit dérivable et  $f'$  continue sur  $]a_{i-1}, a_i[$  et qu'en chaque  $a_i$ ,  $f$  et  $f'$  aient des limites finies à droite et à gauche.

**Théorème 3.10** (Dirichlet). Soit  $f$  une fonction  $T$  périodique sur  $\mathbb{R}$  et de classe  $C^1$  par morceaux. La série de Fourier de  $f$  converge en tout point  $x$  de  $\mathbb{R}$  et

$$\sum_{k=0}^{+\infty} (a_k(f) \cos(k\omega x) + b_k(f) \sin(k\omega x)) = \frac{f(x-) + f(x+)}{2} \quad (3.11)$$

$$= \frac{1}{2} \lim_{\substack{\varepsilon \rightarrow 0, \\ \varepsilon > 0}} f(x - \varepsilon) + \frac{1}{2} \lim_{\substack{\varepsilon \rightarrow 0, \\ \varepsilon > 0}} f(x + \varepsilon).$$

En particulier, en tout point de continuité  $x$  de  $f$ , la somme de la série de Fourier de  $f$  au point  $x$  est égale à  $f(x)$ .

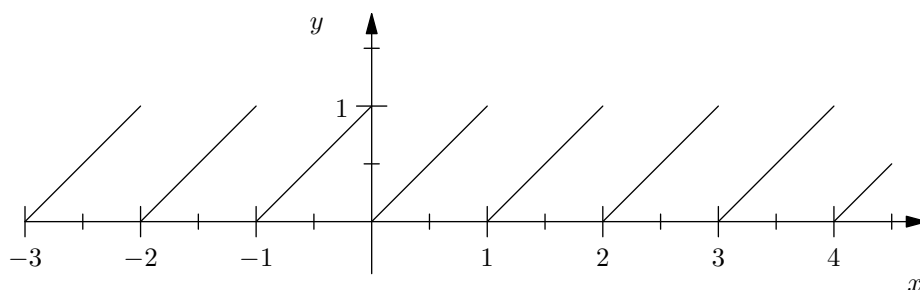


FIGURE 3.1 – Fonction 1 périodique vérifiant  $f(x) = x$  pour  $x \in [0, 1[$

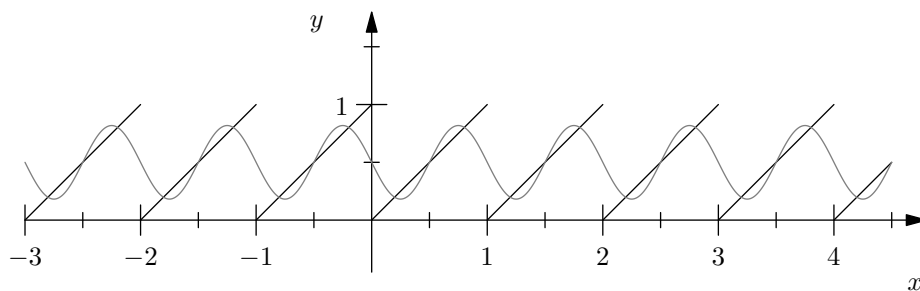
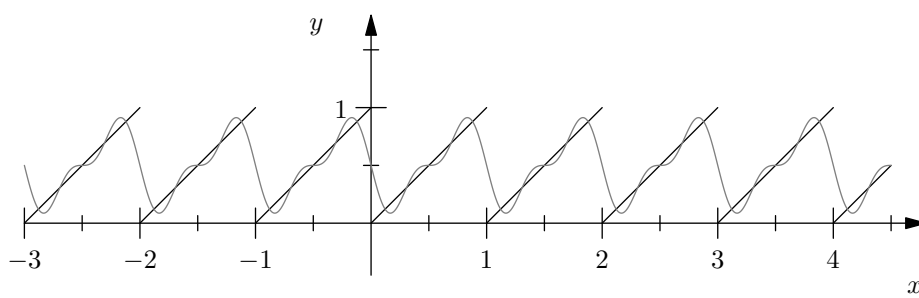
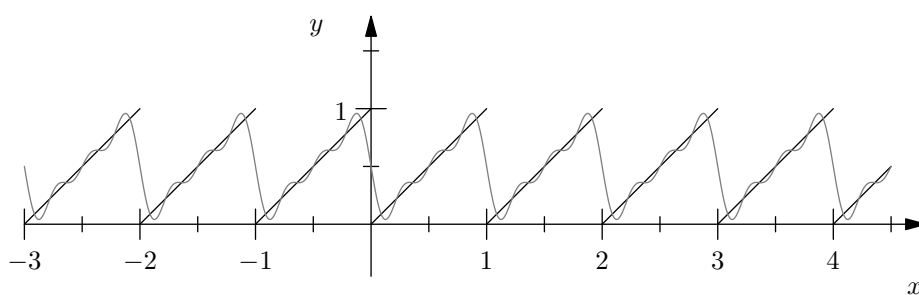
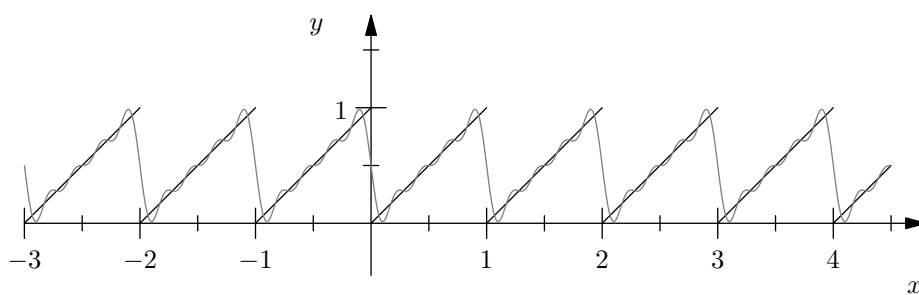


FIGURE 3.2 –  $f$  et  $S_1(f)$

FIGURE 3.3 –  $f$  et  $S_2(f)$ FIGURE 3.4 –  $f$  et  $S_3(f)$ FIGURE 3.5 –  $f$  et  $S_4(f)$

**Théorème 3.11** (identité de Parseval). *Si  $f$  est  $T$ -périodique et si l'intégrale, éventuellement généralisée,  $\int_T f(t)^2 dt$  converge,*

$$\frac{2}{T} \int_T f(t)^2 dt = 2a_0(f)^2 + \sum_{k=1}^{+\infty} (a_k(f)^2 + b_k(f)^2). \quad (3.12)$$

*Cette formule est vraie en particulier si  $f$  est bornée et continue par morceaux sur  $[0, T]$ .*

D'un point de vue physique, si on voit  $f(t)$  comme un signal périodique et  $\frac{2}{T} \int_T f(t)^2 dt$  comme l'énergie de ce signal par période, l'identité de Parseval exprime cette énergie comme la somme des énergies des ondes sinusoïdales en lesquelles on peut décomposer ce signal, cette décomposition étant la série de Fourier de  $f$ .

D'un point de vue mathématique, l'identité de Parseval est un théorème de Pythagore en dimension infinie. Pour cela on introduit sur l'espace des fonctions de carré intégrable le produit scalaire  $\langle f, g \rangle = \frac{1}{T} \int_T f(t)g(t) dt$  et on montre que les fonctions  $1/\sqrt{2}$ ,  $\cos(k\omega x)$ ,  $\sin(k\omega x)$ ,  $k \geq 1$ , constituent une base orthonormée de cet espace. L'identité de Parseval est alors la formule qui exprime la carré scalaire du « vecteur »  $f$  comme la somme des carré des coordonnées de ce vecteur dans cette base. Bien sûr, la justification rigoureuse de ces affirmations est hors programme. Nous nous contenterons de donner la première étape de la démonstration qui consiste à établir l'identité de Parseval pour des « polynômes trigonométriques ».

*Amorce de preuve.* Nous allons montrer l'identité de Parseval pour une fonction qui est égale à la  $n^e$  somme partielle de sa série de Fourier, autrement dit pour une fonction

$$f_n(t) = \sum_{j=-n}^n c_j e^{ij\omega t}.$$

Il est facile de voir que dans ce cas,  $c_k(f_n) = c_k$  pour  $-n \leq k \leq n$  et  $c_k(f_n) = 0$  pour  $|k| > n$ .

$$\begin{aligned} \int_T f_n(t)^2 dt &= \int_T \left| \sum_{j=-n}^n c_j e^{ij\omega t} \right|^2 dt = \int_T \left( \sum_{j=-n}^n c_j e^{ij\omega t} \right) \left( \sum_{k=-n}^n \bar{c}_k e^{-ik\omega t} \right) dt \\ &= \int_T \sum_{j,k=-n}^n c_j \bar{c}_k e^{i(j-k)\omega t} dt \\ &= \sum_{j,k=-n}^n c_j \bar{c}_k \int_T e^{i(j-k)\omega t} dt. \end{aligned}$$

D'après (3.10), cette somme de  $(2n+1)^2$  termes se réduit à la somme des termes pour lesquels  $j = k$  et on obtient :

$$\int_T f_n(t)^2 dt = T \sum_{k=-n}^n |c_k|^2 = T|c_0|^2 + T \sum_{k=1}^n |c_k|^2 + T \sum_{k=1}^n |c_{-k}|^2.$$

En se rappelant (3.1), vraie pour une série trigonométrique convergente, donc a fortiori dans le cas particulier du polynôme trigonométrique  $f_n$  vu comme une série trigonométrique dont tous les coefficients sont nuls à partir du rang  $n+1$ , on a  $|c_0|^2 = a_0^2$  et

$$|c_k|^2 = \left| \frac{a_k - ib_k}{2} \right|^2 = \frac{a_k^2 + b_k^2}{4}, \quad |c_{-k}|^2 = \left| \frac{a_k + ib_k}{2} \right|^2 = \frac{a_k^2 + b_k^2}{4},$$

d'où

$$\int_T f_n(t)^2 dt = T a_0^2 + \frac{T}{2} \sum_{k=1}^n (a_k^2 + b_k^2).$$

On en déduit immédiatement l'inégalité de Parseval pour  $f_n$ .

Soit maintenant  $f$  de carré intégrable sur  $[0, T]$  et  $S_n(f)$  la  $n^{\text{e}}$  somme partielle de sa série de Fourier. En notant de pour tout  $k \leq n$ ,  $a_k(S_n(f)) = a_k(f)$  et  $b_k(S_n(f)) = b_k(f)$  et en appliquant ce que nous venons de démontrer à  $f_n = S_n(f)$ , on obtient :

$$\forall n \geq 1, \quad \frac{2}{T} \int_T S_n(f)(t)^2 dt = 2a_0(f)^2 + \sum_{k=1}^n (a_k(f)^2 + b_k(f)^2). \quad (3.13)$$

À partir de là, il suffit de savoir démontrer que

$$\lim_{n \rightarrow +\infty} \int_T S_n(f)(t)^2 dt = \int_T f(t)^2 dt, \quad (3.14)$$

pour pouvoir établir l'identité de Parseval pour  $f$  en passant à la limite dans (3.13). La justification de (3.14) est la partie hors programme que nous admettrons.  $\square$





# Chapitre 4

## Évènements et probabilités

### 4.1 Modéliser l'aléatoire

#### 4.1.1 Notion d'expérience aléatoire

La théorie des probabilités fournit des modèles mathématiques permettant l'étude d'expériences dont le résultat ne peut être prévu avec une totale certitude. Le tableau 4.1 en donne quelques exemples.

Expérience	Résultat observable
Lancer d'un dé	Un entier $k \in \llbracket 1, 6 \rrbracket$
Prélèvement de $n$ objets en sortie d'une chaîne de production	Nombre d'objets défectueux dans l'échantillon
Questionnaire à 100 questions binaires	Suite $\omega$ de 100 réponses $\omega \in \{\text{oui}, \text{non}\}^{100}$
Lancer d'une pièce jusqu'à la première obtention de pile	Un entier $k \in \mathbb{N}$ : le temps d'attente du premier succès
Mise en service d'une ampoule	Durée de vie $T \in \mathbb{R}_+$
Lancer d'une fléchette sur une cible	Point d'impact
Mouvement d'un grain de pollen dans un liquide	Une fonction continue : la trajectoire
Mélange de deux gaz	Répartition spatiale de deux types de molécules

TABLE 4.1 – Quelques expériences aléatoires typiques

Bien que le résultat précis de chacune de ces expériences soit imprévisible, l'observation et l'intuition nous amènent à penser que ces phénomènes obéissent à certaines lois. Par exemple si on jette 6 000 fois un dé, on s'attend à ce que le nombre d'apparitions de la face « 3 » soit *voisin* de 1 000. Si on met en service 100 ampoules, leurs durées de vie observées seront *concentrées* autour d'une certaine valeur moyenne.

La théorie des probabilités permet de donner un sens précis à ces considérations un peu vagues. La *statistique* permet de confronter les modèles probabilistes avec la réalité observée afin de les valider ou de les invalider. Par exemple si quelqu'un a 60 bonnes réponses sur 100 à un questionnaire, est-il légitime de considérer qu'il a « mieux fait » que le hasard ? Sur les  $n$  objets prélevés en sortie de chaîne,  $k$  sont défectueux. Peut-on en déduire quelque chose sur la qualité de la production globale ?

### 4.1.2 Évènements

La théorie moderne des probabilités utilise le langage des ensembles pour modéliser une expérience aléatoire. Nous noterons  $\Omega$  un ensemble dont les éléments représentent tous les résultats possibles ou *évènements élémentaires* d'une expérience aléatoire donnée. Les *évènements* (ou évènements composés) seront représentés par des parties (sous-ensembles) de  $\Omega$ .

Il n'est pas toujours facile de trouver un ensemble  $\Omega$  permettant de modéliser l'expérience aléatoire. Voici une règle pratique pour y arriver : les évènements élémentaires sont ceux qui contiennent *l'information maximale* qu'il est possible d'obtenir de l'expérience. Par exemple si on jette un dé, l'évènement  $A$  : « obtention d'un chiffre pair » n'est pas élémentaire. Il est composé des trois évènements élémentaires 2, 4, 6 :  $A = \{2, 4, 6\}$ . Ici  $\Omega = \{1, 2, 3, 4, 5, 6\}$ . De même si on lance trois fois une pièce de monnaie, les évènements élémentaires sont des triplets comme (p,f,p) indiquant le résultat précis de chacun des trois lancers. Ici  $\Omega = \{f, p\}^3$ . L'évènement  $B$  « obtention de pile au deuxième des trois lancers » est composé :  $B = \{(f, p, f); (f, p, p); (p, p, f); (p, p, p)\}$ .

Avec ce mode de représentation, les opérations logiques sur les évènements : « et », « ou », « négation » se traduisent par des opérations ensemblistes : intersection, réunion, passage au complémentaire. Le tableau 4.2 page 56 présente la correspondance entre les deux langages.

Notations	Vocabulaire ensembliste	Vocabulaire probabiliste
$\emptyset$	ensemble vide	évènement impossible
$\Omega$	ensemble plein	évènement certain
$\omega$	élément de $\Omega$	évènement élémentaire
$A$	sous-ensemble de $\Omega$	évènement
$\omega \in A$	$\omega$ appartient à $A$	Le résultat $\omega$ est une des réalisations possibles de $A$
$A \subset B$	$A$ inclus dans $B$	$A$ implique $B$
$A \cup B$	réunion de $A$ et $B$	$A$ ou $B$
$A \cap B$	intersection de $A$ et $B$	$A$ et $B$
$A^c$	complémentaire de $A$ dans $\Omega$	évènement contraire de $A$
$A \cap B = \emptyset$	$A$ et $B$ sont disjoints	$A$ et $B$ sont incompatibles

TABLE 4.2 – Langage ensembliste - langage probabiliste

Les opérations logiques sur les évènements peuvent bien sûr faire intervenir plus de deux évènements. Ainsi, si  $A_1, \dots, A_n$  sont des évènements,

$$\bigcup_{i=1}^n A_i = A_1 \cup A_2 \cup \dots \cup A_n$$

est l'ensemble des  $\omega$  qui sont dans *l'un au moins* des  $A_i$ . C'est donc l'évènement « réalisation de l'un au moins des  $A_i$  ( $1 \leq i \leq n$ ) ». De même :

$$\bigcap_{i=1}^n A_i = A_1 \cap A_2 \cap \dots \cap A_n$$

est l'ensemble des  $\omega$  qui sont dans *tous* les  $A_i$ . C'est donc l'évènement « réalisation de chacun des  $A_i$  ( $1 \leq i \leq n$ ) ». Ceci s'étend aux réunions et intersections d'une suite infinie d'évènements :

$$\begin{aligned} \bigcup_{i \in \mathbb{N}^*} A_i &= \{\text{réalisation de l'un au moins des } A_i, i \in \mathbb{N}^*\}, \\ \bigcap_{i \in \mathbb{N}^*} A_i &= \{\text{réalisation de tous les } A_i, i \in \mathbb{N}^*\}. \end{aligned}$$

Ces opérations logiques sur des suites d'évènements sont très utiles pour analyser des évènements complexes à l'aide d'évènements plus simples et, comme nous le verrons plus tard, calculer ainsi des probabilités.

### 4.1.3 Un exemple informel

Voici une façon simple de générer un nombre entier au hasard sans le limiter *a priori* en taille : on lance un dé jusqu'à la première obtention du six et on note le nombre de lancers ainsi réalisés. Pour tout entier  $n \geq 1$ , notons  $E_n$  l'évènement :

$$E_n = \{\text{la première apparition du six a lieu lors du } n^{\text{e}} \text{ lancer}\}.$$

Nous n'avons pas précisé l'ensemble  $\Omega$  des évènements élémentaires associés à cette expérience, mais il est clair que si,  $n$  étant fixé, on veut attribuer une probabilité  $P_n(E_n)$  à l'évènement  $E_n$ , il faut se placer dans un  $\Omega$  permettant de modéliser au moins les  $n$  premiers lancers. A minima, on pourrait prendre :

$$\Omega_n = \llbracket 1, 6 \rrbracket^n = \{(u_1, \dots, u_n) ; u_k \in \llbracket 1, 6 \rrbracket, \forall k \leq n\}.$$

Dans cette modélisation,  $u_k$  représente le numéro entre 1 et 6 sorti au  $k^{\text{e}}$  lancer et l'évènement  $E_n$  s'écrit :

$$E_n = \{(u_1, \dots, u_{n-1}, 6) ; u_k \in \llbracket 1, 5 \rrbracket, \forall k \leq n-1\}.$$

Faisons l'hypothèse d'équiprobabilité de tous les évènements élémentaires  $\omega = (u_1, \dots, u_n)$ , ce qui revient ici à supposer que le dé est « équilibré », c'est-à-dire que lors d'un lancer, chaque face a même probabilité d'apparition  $1/6$ . Alors la probabilité  $P_n(E_n)$  se calcule en faisant du dénombrement :

$$P_n(E_n) = \frac{\text{card } E_n}{\text{card } \Omega_n} = \frac{5^{n-1} \times 1}{6^n} = \left(\frac{5}{6}\right)^{n-1} \frac{1}{6}.$$

Bien entendu la modélisation ci-dessus n'est pas satisfaisante car on ne sait pas *a priori* de combien de lancers on aura besoin pour une première apparition du six et on voudrait donc pouvoir calculer la probabilité de  $E_n$  pour toute valeur de  $n$ . Ceci nous amène naturellement à remplacer  $\Omega_n$  par l'ensemble  $\Omega$  des suites *infinies* d'entiers compris entre 1 et 6 :

$$\Omega = \{(u_k)_{k \in \mathbb{N}^*} ; u_k \in \llbracket 1, 6 \rrbracket, \forall k \in \mathbb{N}^*\}.$$

Nous n'explicitons pas ici la famille  $\mathcal{F}$  d'évènements observables<sup>1</sup>, mais il est clair qu'elle doit contenir *au moins* tous les évènements dont la réalisation ne dépend que des  $n$  premiers lancers (comme  $E_n$ ), et ce pour toute valeur de l'entier  $n \geq 1$ . Une autre condition

1. En fait il n'est pas judicieux ici de prendre comme évènements observables *tous* les sous-ensembles de  $\Omega$ , mais l'expliquer proprement nous emmènerait bien au delà du niveau de ce cours.

naturelle à imposer au modèle est que la probabilité de réalisation de  $E_n$  ne dépende pas du nombre de lancers, pourvu qu'il soit au moins égal à  $n$ . Autrement dit, on posera  $P(E_n) = P_n(E_n) = (5/6)^{n-1}(1/6)$  pour tout  $n \geq 1$ .

Considérons maintenant l'évènement

$$E = \{ \text{le six finit par sortir} \}.$$

Comme  $E$  se décompose en réunion des  $E_n$  :

$$E = \bigcup_{n \in \mathbb{N}^*} E_n,$$

on aimerait pouvoir calculer sa probabilité à partir des probabilités des  $E_n$ . Pour cela, nous avons besoin de deux choses :

- que  $E$  soit lui-même un *évènement observable*, autrement dit qu'on puisse lui attribuer une probabilité ;
- que l'on dispose d'une formule exprimant  $P(E)$  en fonction des  $P(E_n)$ .

Pour la première condition, il suffit de convenir qu'une famille d'évènements observables doit être *stable* par réunion de *suites* d'évènements observables (donc ici comme les  $E_n$  sont tous dans  $\mathcal{F}$  et  $E$  est la réunion de la suite des  $E_n$ ,  $E$  appartient lui aussi à  $\mathcal{F}$ ).

Pour la deuxième condition, en procédant par analogie avec le cas d'une réunion finie d'évènements *deux à deux incompatibles*, et les  $E_n$  le sont (pourquoi?), on a bien envie d'écrire :

$$P(E) = P\left(\bigcup_{n \in \mathbb{N}^*} E_n\right) = \sum_{n=1}^{+\infty} P(E_n).$$

Admettons que ceci soit légitime. Alors la probabilité ainsi attribuée à  $E$  se calcule comme la somme de la série géométrique de raison  $5/6$  (donc convergente) et de premier terme  $1/6$  :

$$P(E) = \sum_{n=1}^{+\infty} P(E_n) = \frac{1}{6} \sum_{n=1}^{+\infty} \left(\frac{5}{6}\right)^{n-1} = \frac{1}{6} \sum_{k=0}^{+\infty} \left(\frac{5}{6}\right)^k = \frac{1}{6} \times \frac{1}{1 - 5/6} = 1.$$

Ainsi  $P(E) = 1$ , ce qui est conforme à l'intuition : si le dé est équilibré, on est sûr que le six finira bien par sortir.

Regardons maintenant l'évènement complémentaire  $E^c$  :

$$E^c = \{ \text{le six ne sort jamais} \}.$$

Naturellement,  $P(E^c) = 1 - P(E) = 1 - 1 = 0$ . Mais attention à ne pas en déduire que  $E^c$  est l'ensemble vide. Bien au contraire, il contient une infinité d'évènements élémentaires :

$$E^c = \{(u_k)_{k \in \mathbb{N}^*} ; u_k \in \llbracket 1, 5 \rrbracket, \forall k \in \mathbb{N}^*\}.$$

Avant de passer à la théorie, listons quelques idées à retenir de cet exemple étudié de manière informelle.

1. Quand l'ensemble des évènements élémentaires est infini, il n'est pas toujours facile d'explicitier le choix de la famille  $\mathcal{F}$  d'évènements observables. On devra parfois se contenter d'admettre qu'elle contient au moins les évènements auxquels on s'intéresse vraiment.
2. Cette famille  $\mathcal{F}$  devrait être stable par réunion des suites d'évènements observables et par passage au complémentaire.

3. Les séries jouent un rôle essentiel lorsque  $\Omega$  est infini.
4. La propriété d'additivité des probabilités (la probabilité d'une réunion finie d'évènements deux à deux incompatibles est égale à la somme de leurs probabilités) devrait pouvoir s'étendre aux *suites* infinies d'évènements deux à deux incompatibles en remplaçant « somme » par « série convergente ».

## 4.2 Le modèle probabiliste

La *probabilité*  $P$ , telle que nous allons la définir ci-dessous, est une fonction qui, à un évènement, associe un nombre compris entre 0 et 1 et censé mesurer les chances de réalisation de cet évènement. Pour des raisons sortant du cadre de ce cours, il n'est pas toujours possible d'attribuer ainsi de manière cohérente une probabilité à *chaque* partie de  $\Omega$ . En d'autres termes,  $P$  ne peut pas être considérée comme une *application* de  $\mathcal{P}(\Omega) \rightarrow [0, 1]$ , où  $\mathcal{P}(\Omega)$  est l'ensemble de toutes les parties de  $\Omega$ , mais comme une *fonction* ayant pour domaine de définition une famille  $\mathcal{F}$  de sous-ensembles de  $\Omega$ , généralement plus petite que  $\mathcal{P}(\Omega)$ . Cette famille  $\mathcal{F}$  est appelée famille des évènements observables ou tribu. Pour mériter ce nom, elle doit vérifier certaines propriétés données dans la définition suivante.

**Définition 4.1** (tribu). *Une famille  $\mathcal{F}$  de parties de  $\Omega$  est appelée tribu sur  $\Omega$  si elle*

- a) possède l'ensemble vide :  $\emptyset \in \mathcal{F}$  ;
- b) est stable par passage au complémentaire :  $\forall A \in \mathcal{F}, A^c \in \mathcal{F}$  ;
- c) est stable par union dénombrable :  $(\forall i \in \mathbb{N}^*, A_i \in \mathcal{F}) \Rightarrow \bigcup_{i \in \mathbb{N}^*} A_i \in \mathcal{F}$ .

On vérifie à partir de cette définition qu'une tribu est stable par unions finies (prendre tous les  $A_i$  vides à partir d'un certain rang), intersections finies et par intersections dénombrables (combinaison b) et c)).

**Remarque 4.2.** La définition générale d'une tribu  $\mathcal{F}$  ne suppose pas que tous les singletons  $\{\omega\}$  soient des éléments de  $\mathcal{F}$ . Donc un « évènement élémentaire » n'est pas toujours un évènement observable. Néanmoins dans la plupart des exemples que nous étudierons, la tribu possèdera les singletons.

Voici trois exemples simples de tribu, nous en verrons d'autres par la suite.

- La tribu triviale sur  $\Omega$  est  $\mathcal{F} = \{\Omega, \emptyset\}$ .
- L'ensemble  $\mathcal{P}(\Omega)$  de toutes les parties de  $\Omega$  est une tribu.
- Si  $A$  est une partie de  $\Omega$ , alors  $\mathcal{F} := \{\Omega, \emptyset, A, A^c\}$  est une tribu. C'est la *plus petite* tribu possédant  $A$  comme élément, au sens où toute tribu  $\mathcal{G}$  telle que  $A \in \mathcal{G}$  contient  $\mathcal{F}$ . On dit que  $\mathcal{F}$  est la tribu *engendrée* par  $A$ .

**Définition 4.3.** *Soit  $\Omega$  un ensemble et  $\mathcal{F}$  une tribu sur  $\Omega$ . On appelle probabilité sur  $(\Omega, \mathcal{F})$  toute application  $P$  de  $\mathcal{F}$  dans  $[0, 1]$  vérifiant :*

- (i)  $P(\Omega) = 1$ .
- (ii) Pour toute suite  $(A_j)_{j \geq 1}$  d'évènements de  $\mathcal{F}$  deux à deux disjoints (incompatibles) :

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

Le triplet  $(\Omega, \mathcal{F}, P)$  s'appelle *espace probabilisé*.

Définir une probabilité sur  $(\Omega, \mathcal{F})$  c'est en quelque sorte attribuer une « masse » à chaque évènement observable, avec par convention une masse totale égale à 1 pour l'évènement certain  $\Omega$ .

Nous donnons maintenant sept propriétés importantes d'une probabilité qui se déduisent de la définition 4.3. Il faut les connaître pour pouvoir faire du calcul des probabilités. La connaissance de la démonstration est facultative.

**Proposition 4.4** (propriétés générales d'une probabilité).

Toute probabilité  $P$  sur  $(\Omega, \mathcal{F})$  vérifie les propriétés suivantes :

1.  $P(\emptyset) = 0$ .
2. Additivité.
  - a) Si  $A \cap B = \emptyset$ ,  $P(A \cup B) = P(A) + P(B)$ .
  - b) Si les  $A_i$  ( $1 \leq i \leq n$ ) sont deux à deux disjoints :  $P(\bigcup_{i=1}^n A_i) = \sum_{i=1}^n P(A_i)$ .
3.  $\forall A \in \mathcal{F}$ ,  $P(A^c) = 1 - P(A)$ .
4.  $\forall A \in \mathcal{F}$ ,  $\forall B \in \mathcal{F}$ ,  $A \subset B \Rightarrow P(A) \leq P(B)$ .
5.  $\forall A \in \mathcal{F}$ ,  $\forall B \in \mathcal{F}$ ,  $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ .
6. Continuité monotone séquentielle.
  - a) Si  $(B_n)_{n \geq 0}$  est une suite croissante d'évènements de  $\mathcal{F}$  convergente<sup>2</sup> vers  $B \in \mathcal{F}$ , alors  $P(B) = \lim_{n \rightarrow +\infty} P(B_n)$ , en notation abrégée :
 
$$B_n \uparrow B \Rightarrow P(B_n) \uparrow P(B) \quad (n \rightarrow +\infty).$$
  - b) Si  $(C_n)_{n \geq 0}$  est une suite décroissante d'évènements de  $\mathcal{F}$  convergente<sup>3</sup> vers  $C \in \mathcal{F}$ , alors  $P(C) = \lim_{n \rightarrow +\infty} P(C_n)$  en notation abrégée :
 
$$C_n \downarrow C \Rightarrow P(C_n) \downarrow P(C) \quad (n \rightarrow +\infty).$$
7. Sous-additivité et sous- $\sigma$ -additivité.
  - a)  $\forall A \in \mathcal{F}$ ,  $\forall B \in \mathcal{F}$ ,  $P(A \cup B) \leq P(A) + P(B)$ .
  - b)  $\forall A_1, \dots, A_n \in \mathcal{F}$ ,  $P(\bigcup_{i=1}^n A_i) \leq \sum_{i=1}^n P(A_i)$ .
  - c)  $\forall A_1, \dots, A_n, \dots \in \mathcal{F}$ ,  $P(\bigcup_{i \in \mathbb{N}^*} A_i) \leq \sum_{i=1}^{+\infty} P(A_i)$ .

*Preuve.* Soit  $P$  une fonction d'ensembles  $\mathcal{F} \rightarrow [0, 1]$  satisfaisant aux conditions (i) et (ii) de la définition 4.3, il s'agit de démontrer que  $P$  vérifie les propriétés 1 à 7.

*Preuve de 1.* Comme  $P(A_j) \geq 0$  pour tout  $A_j \in \mathcal{F}$ , on a toujours

$$\sum_{j \in \mathbb{N}^*} P(A_j) \geq P(A_1) + P(A_2),$$

le premier membre pouvant être égal à  $+\infty$ . En choisissant  $A_j = \emptyset$  pour tout  $j \in \mathbb{N}^*$  et en utilisant la  $\sigma$ -additivité (ii), on en déduit :

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

Par conséquent,  $P(\emptyset) \geq 2P(\emptyset)$  et comme  $P(\emptyset) \geq 0$ , ceci entraîne  $P(\emptyset) = 0$ . □

2. Ce qui signifie :  $\forall n \geq 0$ ,  $B_n \subset B_{n+1}$  et  $B = \bigcup_{n \geq 0} B_n$ .

3. Ce qui signifie :  $\forall n \geq 0$ ,  $C_{n+1} \subset C_n$  et  $C = \bigcap_{n \geq 0} C_n$ .

*Preuve de 2.* Soient  $A_1, \dots, A_n$ ,  $n$  évènements 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'évènements deux à deux disjoints. En utilisant la  $\sigma$ -additivité, on obtient alors :

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

D'après 1, la somme pour  $j \geq n + 1$  vaut 0, ceci prouve 2 b). Bien sûr, 2 a) n'est que le cas particulier  $n = 2$ .  $\square$

*Preuve de 3.* Prendre  $B = A^c$  dans 2 a) et utiliser (i).  $\square$

*Preuve de 4.* Si  $A \subset B$ , alors  $B = A \cup (B \cap A^c)$  et cette réunion est disjointe. D'après 2 a) on a  $P(B) = P(A) + P(B \cap A^c)$  et comme  $P(B \cap A^c) \geq 0$ , on en déduit  $P(B) \geq P(A)$ .  $\square$

*Preuve de 5.* On a les décompositions suivantes en unions disjointes :

$$\begin{aligned} A \cup B &= (A \cap B^c) \cup (A \cap B) \cup (A^c \cap B), \\ A &= (A \cap B^c) \cup (A \cap B), \\ B &= (A \cap B) \cup (A^c \cap B). \end{aligned}$$

En utilisant l'additivité on en déduit :

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

$\square$

*Preuve de 6.* Il suffit de prouver 6 a), la propriété 6 b) s'en déduit en appliquant 6 a) à la suite d'évènements  $B_n = C_n^c$ . Admettons, pour l'instant, que pour tout  $n \geq 1$ ,  $B_n$  vérifie la décomposition suivante en union disjointe (cf. figure 4.1)

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

En écrivant la réunion infinie des  $B_n$  à l'aide de cette décomposition et en « effaçant »

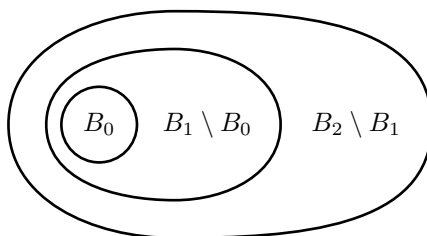


FIGURE 4.1 – Décomposition de  $B_0 \cup B_1 \cup B_2$  en union disjointe

toutes les répétitions des  $B_i \setminus B_{i-1}$ , on en déduit immédiatement que  $B$  vérifie la décomposition en union disjointe :

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

Passant aux probabilités, ces deux décompositions nous donnent :

$$P(B_n) = P(B_0) + \sum_{i=1}^n P(B_i \setminus B_{i-1}),$$

$$P(B) = P(B_0) + \sum_{i=1}^{+\infty} P(B_i \setminus B_{i-1}).$$

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

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

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

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

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 car  $B_i \setminus B_{i-1} \subset B_i \subset B_n$ , pour  $i \leq n$ . Pour prouver l'inclusion inverse, on note  $\omega$  un élément *quelconque* de  $B_n$  et on montre que  $\omega$  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  $\omega$  de  $B_n$ , on en déduit  $B_n \subset D_n$ .  $\square$

*Preuve de 7 a).* D'après 5 :

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \leq P(A) + P(B),$$

car  $P(A \cap B) \geq 0$ .  $\square$

*Preuve de 7 b).* On remarque que pour tout  $n \geq 1$ ,  $\bigcup_{i=1}^n A_i = \bigcup_{i=1}^n B_i$ , où les  $B_i$  sont des évènements *deux à deux disjoints* définis comme suit :

$$B_1 = A_1, B_2 = A_2 \cap B_1^c, B_3 = A_3 \cap (B_1 \cup B_2)^c, \dots, B_n = A_n \cap (B_1 \cup \dots \cup B_{n-1})^c.$$

Par additivité :

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

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

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

$\square$

*Preuve de 7 c).* Posons pour tout  $n \geq 1$ ,

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



La suite  $(D_n)_{n \geq 1}$  est *croissante* et a pour limite  $D$ . Donc d'après 6 a),  $P(D_n) \uparrow P(D)$  ( $n \rightarrow +\infty$ ). D'après 7 b) on a :

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

Les deux membres de cette inégalité étant les termes généraux de deux suites croissantes de réels positifs, on obtient en passant à la limite quand  $n$  tend vers l'infini :

$$P\left(\bigcup_{i \in \mathbb{N}^*} A_i\right) = P(D) \leq \sum_{i=1}^{+\infty} P(A_i),$$

ce qui prouve 7 c). Remarquons que les sommes partielles de la série convergent dans  $\overline{\mathbb{R}}_+$ . Bien sûr, l'inégalité obtenue n'a d'intérêt que lorsque la série de terme général  $P(A_i)$  converge et a une somme strictement inférieure à 1.  $\square$

La vérification de la proposition 4.4 est maintenant complète.  $\square$

Le calcul de probabilités de réunions ou d'intersections est une question cruciale. La propriété 5 montre qu'en général on ne peut pas calculer  $P(A \cup B)$  à partir de la *seule connaissance* de  $P(A)$  et  $P(B)$  et qu'on se heurte à la même difficulté pour  $P(A \cap B)$ . Le calcul des probabilités d'intersections sera discuté plus tard, à propos du conditionnement. Pour les probabilités de réunions, on peut se demander comment se généralise la propriété 5 lorsqu'on réunit plus de deux événements. Il est facile de vérifier (faites-le!) que :

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C).$$

Le cas général est donné par la formule de Poincaré qui exprime  $P(A_1 \cup \dots \cup A_n)$  à l'aide des probabilités de *toutes les intersections* des  $A_i$  : 2 à 2, 3 à 3, etc.

**Proposition 4.5** (formule de Poincaré).

Pour tout entier  $n \geq 2$  et tous événements  $A_1, \dots, A_n$  :

$$\begin{aligned} P\left(\bigcup_{i=1}^n A_i\right) &= \sum_{i=1}^n P(A_i) + \sum_{k=2}^n (-1)^{k+1} \sum_{1 \leq i_1 < i_2 < \dots < i_k \leq n} P(A_{i_1} \cap \dots \cap A_{i_k}) \\ &= \sum_{i=1}^n P(A_i) + \sum_{k=2}^n (-1)^{k+1} \sum_{\substack{I \subset \llbracket 1, n \rrbracket \\ \text{card } I = k}} P\left(\bigcap_{j \in I} A_j\right). \end{aligned} \quad (4.1)$$

La formule peut se démontrer par récurrence. Nous l'admettrons.

## 4.3 Exemples

Nous examinons maintenant quelques exemples d'espaces probabilisés et de calcul de probabilités d'événements.

**Exemple 4.6.** On effectue une partie de pile ou face en trois coups. Quelle est la probabilité d'obtenir pile aux premier et troisième lancers ?

On peut modéliser cette expérience en prenant  $\Omega = \{\mathbf{f}, \mathbf{p}\}^3$  ensemble de tous les triplets donc chaque composante est l'un des lettres  $\mathbf{f}$  (pour « face ») ou  $\mathbf{p}$  (pour « pile »). Pour

famille d'évènements observables nous prenons  $\mathcal{F} = \mathcal{P}(\Omega)$  l'ensemble de toutes les parties<sup>4</sup> de  $\Omega$ . La pièce étant supposée symétrique, nous n'avons *a priori* pas de raison de supposer que l'un des 8 triplets de résultats possibles soit favorisé ou défavorisé par rapport aux autres. Nous choisissons donc  $P$  de sorte que tous les évènements élémentaires aient même probabilité (hypothèse d'équiprobabilité), soit :

$$\forall \omega \in \Omega, \quad P(\{\omega\}) = \frac{1}{\text{Card } \Omega} = \frac{1}{2^3}.$$

L'évènement  $B$  dont on veut calculer la probabilité s'écrit :

$$B = \{(p, f, p); (p, p, p)\}.$$

D'où :

$$P(B) = \frac{1}{8} + \frac{1}{8} = \frac{1}{4}.$$

**Exemple 4.7.** On fait remplir un questionnaire comportant 20 questions binaires. Quelle est la probabilité qu'un candidat répondant au hasard obtienne au moins 16 bonnes réponses ?

On choisit ici :

$$\Omega = \{\text{oui, non}\}^{20}, \quad \mathcal{F} = \mathcal{P}(\Omega).$$

Un évènement élémentaire est donc une grille complétée à 20 réponses ou liste de 20 réponses oui ou non. Si le candidat répond complètement au hasard, on peut considérer que chacune des  $2^{20}$  grilles de réponses possibles a la même probabilité d'apparaître (hypothèse d'équiprobabilité sur  $\Omega$ ). Pour tout  $B \subset \Omega$ , on a alors :

$$P(B) = \frac{\text{Card } B}{\text{Card } \Omega}.$$

En particulier pour  $B = \{\text{obtention d'au moins 16 bonnes réponses}\}$ ,

$$P(B) = \frac{C_{20}^{16} + C_{20}^{17} + C_{20}^{18} + C_{20}^{19} + C_{20}^{20}}{2^{20}} = \frac{6196}{2^{20}} \simeq 0,006.$$

La notation  $C_n^k$  (coefficient binomial) désigne ici le nombre de parties à  $k$  éléments d'un ensemble à  $n$  éléments ( $0 \leq k \leq n$ ).

**Exemple 4.8** (contrôle de production). On prélève au hasard un échantillon de  $k$  pièces dans une production totale de  $N$  pièces comprenant en tout  $n$  pièces défectueuses. Le prélèvement est *sans remise* (donc  $k \leq N$ ). Cherchons la probabilité de :

$$A_j = \{\text{il y a exactement } j \text{ pièces défectueuses dans l'échantillon}\}.$$

On prend pour  $\Omega$  l'ensemble de toutes les parties à  $k$  éléments d'un ensemble à  $N$  éléments (ensemble de tous les échantillons possibles de taille  $k$ ),  $\mathcal{F} = \mathcal{P}(\Omega)$  et  $P$  l'équiprobabilité sur  $\Omega$ . Il suffit alors de dénombrer tous les échantillons ayant exactement  $j$  pièces défectueuses. Un tel échantillon se construit en prenant  $j$  pièces dans le sous-ensemble des défectueuses ( $C_n^j$  choix possibles) et en complétant par  $k - j$  pièces prises dans le sous-ensemble des non-défectueuses ( $C_{N-n}^{k-j}$  choix possibles). On en déduit :

$$P(A_j) = \frac{C_n^j C_{N-n}^{k-j}}{C_N^k} \quad \text{si} \quad \begin{cases} 0 \leq j \leq n, \\ 0 \leq j \leq k, \\ k - j \leq N - n. \end{cases}$$

Si l'une de ces trois conditions n'est pas vérifiée,  $P(A_j) = 0$ .

4. Lorsque  $\Omega$  est fini, il est toujours possible de faire ce choix.

Dans tous les exemples précédents, nous avons fait systématiquement l'hypothèse d'équiprobabilité. Ce n'est pas une obligation et il y a bien d'autres façons d'obtenir une probabilité  $P$  sur un  $\Omega$  fini muni de la famille d'évènements  $\mathcal{F} = \mathcal{P}(\Omega)$ . Voici comment les trouver toutes. L'énoncé suivant est à connaître, sa preuve peut être sautée en première lecture.

**Proposition 4.9.** *Soit  $\Omega = \{\omega_1, \dots, \omega_n\}$  un ensemble fini à  $n$  éléments et  $\mathcal{F} = \mathcal{P}(\Omega)$  la famille des parties de  $\Omega$ .*

1. *À toute probabilité  $P$  sur  $(\Omega, \mathcal{F})$ , on peut associer une suite finie de réels positifs  $p_1, \dots, p_n$  telle que*

$$\sum_{i=1}^n p_i = 1 \quad (4.2)$$

et

$$\forall i \in \llbracket 1, n \rrbracket, \quad P(\{\omega_i\}) = p_i. \quad (4.3)$$

2. *Réciproquement, à toute suite finie de réels positifs  $p_1, \dots, p_n$  de somme 1, on peut associer une unique probabilité  $P$  sur  $(\Omega, \mathcal{F})$  vérifiant (4.3). Cette probabilité est définie par :*

$$\forall A \in \mathcal{F}, \quad P(A) = \sum_{\omega_i \in A} p_i. \quad (4.4)$$

*Preuve.* Si  $P$  est une probabilité sur  $(\Omega, \mathcal{F})$ , en définissant les  $p_i$  par (4.3) et en notant que  $\Omega$  peut s'écrire comme l'union finie

$$\Omega = \bigcup_{i=1}^n \{\omega_i\}$$

où les  $\{\omega_i\}$  sont évidemment deux à deux disjoints (la numérotation  $i \mapsto \omega_i$  des éléments de  $\Omega$  est bijective, sinon  $\Omega$  aurait moins de  $n$  éléments), on obtient par additivité de  $P$  (prop. 4.4 2b)) :

$$1 = P(\Omega) = P\left(\bigcup_{i=1}^n \{\omega_i\}\right) = \sum_{i=1}^n P(\{\omega_i\}) = \sum_{i=1}^n p_i,$$

ce qui établit (4.2).

Pour montrer la réciproque, on se donne une suite finie quelconque  $p_1, \dots, p_n$  de réels positifs vérifiant (4.2) et on définit la fonction d'ensembles  $P : \mathcal{F} \rightarrow [0, 1]$  par la formule (4.4) avec la convention qu'une somme indexée par l'ensemble vide vaut 0, donc ici  $P(\emptyset) = 0$ . En prenant pour  $A$  successivement chacun des  $\{\omega_i\}$ , on voit que la condition (4.3) est automatiquement vérifiée par  $P$ . Il nous reste à montrer que  $P$  est bien une probabilité sur  $(\Omega, \mathcal{F})$  au sens de la définition 4.3 et que c'est *la seule* qui vérifie les égalités (4.3).

- i) Vérifions d'abord que  $P(\Omega) = 1$ .  $P$  étant définie comme fonction d'ensembles par la formule (4.4), on a

$$P(\Omega) = \sum_{\omega_i \in \Omega} p_i = \sum_{i=1}^n p_i = 1,$$

puisque les  $p_i$  vérifient (4.2) par hypothèse.

ii) Vérifions la  $\sigma$ -additivité de  $P$ . Soit  $(A_j)_{j \geq 1}$  une suite quelconque de parties de  $\Omega$  deux à deux disjointes. Comme  $\Omega$  est fini, seul un nombre fini  $m \leq n$  de  $A_j$  sont non vides<sup>5</sup>, notons les  $A_{j_1}, \dots, A_{j_m}$ . La réunion de tous les  $A_j$ , que nous noterons  $A$  s'écrit alors :

$$A = \bigcup_{j \in \mathbb{N}^*} A_j = \bigcup_{k=1}^m A_{j_k}.$$

En revenant à la définition de  $P$ , nous avons

$$P\left(\bigcup_{j \in \mathbb{N}^*} A_j\right) = P\left(\bigcup_{k=1}^m A_{j_k}\right) = \sum_{\omega_i \in A} p_i.$$

Dans cette dernière somme, l'ensemble d'indexation  $A$  se découpe en les  $m$  ensemble disjointes  $A_{j_k}$ , ce qui permet de découper la somme indexée par  $A$  des  $p_i$  en les  $m$  paquets correspondants<sup>6</sup> :

$$\sum_{\omega_i \in A} p_i = \sum_{k=1}^m \sum_{\omega_i \in A_{j_k}} p_i = \sum_{k=1}^m P(A_{j_k}) = \sum_{j=1}^{+\infty} P(A_j),$$

en notant que tous les termes rajoutés pour passer de la somme  $\sum_{k=1}^m$  à la série  $\sum_{j=1}^{+\infty}$  ci-dessus sont de la forme  $P(\emptyset)$ , donc nuls. Comme la suite  $(A_j)_{j \geq 1}$  de parties deux à deux disjointes de  $\Omega$  était quelconque, ceci prouve la  $\sigma$ -additivité de  $P$ .

L'application  $P : \mathcal{F} \rightarrow [0, 1]$  est donc bien une probabilité.

Il nous reste à voir que  $P$  est la seule probabilité sur  $(\Omega, \mathcal{F})$  vérifiant (4.3). Soit  $Q$  une probabilité sur  $(\Omega, \mathcal{F})$  vérifiant (4.3). Comme tout  $A \in \mathcal{F}$  peut s'écrire comme l'union finie  $A = \bigcup_{\omega_i \in A} \{\omega_i\}$ , par additivité de  $Q$  on obtient :

$$\forall A \in \mathcal{F}, \quad Q(A) = Q\left(\bigcup_{\omega_i \in A} \{\omega_i\}\right) = \sum_{\omega_i \in A} Q(\{\omega_i\}) = \sum_{\omega_i \in A} p_i = P(A),$$

donc  $Q = P$ , ce qui établit l'unicité et termine la preuve.  $\square$

La proposition 4.9 peut sembler essentiellement théorique et il est naturel de se demander si étant donnée une suite finie  $p_1, \dots, p_n$  de réels positifs et de somme 1, il existe toujours une expérience modélisable par l'espace probabilisé  $(\Omega, \mathcal{P}(\Omega), P)$  associé à cette suite par la proposition 4.9. En voici un exemple.

**Exemple 4.10** (roue de la chance). On se donne une suite finie  $p_1, \dots, p_n$  de réels positifs de somme 1. Sur un disque on trace les  $n$  secteurs angulaires consécutifs d'angle au centre  $2\pi p_k$  radians,  $1 \leq k \leq n$ . On choisit un point de repère fixe (par exemple la pointe de la flèche sur la figure 4.2). On fait tourner le disque avec un impulsion initiale et on regarde quel secteur indique la flèche quand le disque s'arrête. Cela revient à tirer au hasard par ce procédé un numéro de secteur. Et la probabilité que le numéro  $k$  sorte ainsi est  $p_k$ .

**Remarque 4.11.** Lorsque  $\Omega$  est fini, la façon la plus simple de construire une probabilité sur  $(\Omega, \mathcal{P}(\Omega))$  est de choisir  $P(\{\omega\}) = 1/\text{card } \Omega$ . On parle alors d'*équiprobabilité* ou de *probabilité uniforme* sur  $(\Omega, \mathcal{P}(\Omega))$ . C'est la modélisation qui s'impose naturellement

5. S'ils sont tous vides,  $m = 0$  et la formule de  $\sigma$ -additivité s'écrit  $P(\emptyset) = \sum_{j=1}^{+\infty} P(\emptyset)$  soit  $0 = \sum_{j=1}^{+\infty} 0$ , ce qui est trivial.

6. Dans une somme d'un nombre fini de termes, on peut grouper les termes comm l'on veut sans changer la valeur de somme globale.

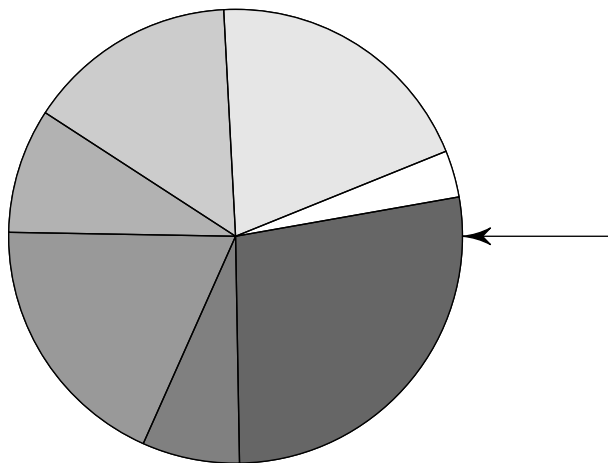


FIGURE 4.2 – Une roulette quelconque

lorsqu'on n'a pas de raison de penser *a priori* qu'un résultat élémentaire de l'expérience soit favorisé ou défavorisé par rapport aux autres. La situation est radicalement différente lorsque  $\Omega$  est infini *dénombrable*<sup>7</sup>. Sur un tel ensemble, *il ne peut pas y avoir d'équiprobabilité*. Imaginons que l'on veuille tirer une boule *au hasard* dans une urne contenant une infinité de boules numérotées de manière bijective par les entiers naturels. Soit  $\{\omega_i\}$  l'évènement *tirage de la boule numérotée  $i$*  ( $i \in \mathbb{N}$ ) et  $p_i$  sa probabilité. Par  $\sigma$ -additivité, les  $p_i$  vérifient nécessairement :

$$\sum_{i \in \mathbb{N}} p_i = 1.$$

Mais si les  $p_i$  sont égaux, tous les termes de la série ci-dessus valent  $p_0$ . Sa somme est alors  $+\infty$  si  $p_0 > 0$  ou 0 si  $p_0 = 0$ , il y a donc une contradiction.

Voici maintenant une caractérisation de toutes les probabilités sur les espaces *au plus dénombrables*, c'est-à-dire finis ou dénombrables. Elle généralise la proposition 4.9. La preuve est analogue, en utilisant pour la  $\sigma$ -additivité de  $P$  la propriété de sommation par paquets des séries à termes *positifs*.

**Proposition 4.12.** *Soit  $\Omega = \{\omega_i ; i \in I\}$  un ensemble au plus dénombrable. La donnée d'une probabilité sur  $(\Omega, \mathcal{P}(\Omega))$  équivaut à la donnée d'une famille  $(p_i)_{i \in I}$  dans  $\mathbb{R}_+$  telle que :*

$$\sum_{i \in I} p_i = 1$$

et des égalités

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

La probabilité  $P$  est alors définie par

$$\forall A \in \mathcal{P}(\Omega), \quad P(A) = \sum_{\omega_i \in A} p_i.$$

7. Un ensemble  $I$  est dit dénombrable s'il est en bijection avec  $\mathbb{N}$ , c'est-à-dire si on peut numéroter ses éléments par les nombres entiers de façon qu'à chaque entier corresponde un élément de  $I$ , que deux éléments différents aient des numéros différents et que tout élément de  $I$  ait un numéro entier. Par exemple, l'ensemble  $\mathbb{Z}$  des entiers relatifs, l'ensemble  $2\mathbb{N}$  des entiers pairs, l'ensemble  $\mathbb{N}^2$  des couples d'entiers sont dénombrables.

Une conséquence de la proposition 4.12 est qu'à partir de toute série convergente à termes  $u_k \geq 0$  avec au moins un  $u_k > 0$ , on peut construire une probabilité sur  $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$  ou sur n'importe quel ensemble dénombrable. En effet, soit  $S = \sum_{k=0}^{+\infty} u_k$ . Comme la série converge dans  $\mathbb{R}_+$  et a au moins un terme strictement positif,  $0 < S < +\infty$ . Alors en posant

$$\forall k \in \mathbb{N}, \quad p_k = \frac{u_k}{S},$$

on voit que les  $p_k$  sont positifs et que  $\sum_{k=0}^{+\infty} p_k = 1$ . On peut donc définir une probabilité  $P$  sur  $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$  en posant :

$$\forall A \subset \mathbb{N}, \quad P(A) = \frac{1}{S} \sum_{k \in A} u_k.$$

**Exemple 4.13** (une probabilité définie sur  $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$ ).

Soit  $a$  un réel strictement positif fixé. On sait que la fonction exponentielle est développable en série entière avec rayon de convergence infini, d'où en particulier :

$$e^a = \sum_{k=0}^{+\infty} \frac{a^k}{k!}.$$

Les termes de cette série convergente étant tous positifs, on pose :

$$\forall k \in \mathbb{N}, \quad p_k = \frac{e^{-a} a^k}{k!}.$$

Alors

$$\sum_{k=0}^{+\infty} \frac{e^{-a} a^k}{k!} = e^{-a} \sum_{k=0}^{+\infty} \frac{a^k}{k!} = e^{-a} e^a = 1.$$

Pour tout  $A \subset \mathbb{N}$ , on définit :

$$P(A) = \sum_{k \in A} p_k = \sum_{k \in A} \frac{e^{-a} a^k}{k!}.$$

D'après la proposition 4.12,  $P$  est une probabilité sur  $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$ . On l'appelle *loi de Poisson*<sup>8</sup> de paramètre  $a$ . Calculons par exemple  $P(2\mathbb{N})$  où  $2\mathbb{N}$  désigne l'ensemble des entiers pairs.

$$P(2\mathbb{N}) = \sum_{k \in 2\mathbb{N}} p_k = \sum_{l=0}^{+\infty} \frac{e^{-a} a^{2l}}{(2l)!} = e^{-a} \frac{e^a + e^{-a}}{2} = \frac{1}{2}(1 + e^{-2a}).$$

Une conséquence de ce résultat est : si l'on tire un nombre entier au hasard suivant une loi de Poisson, la probabilité qu'il soit pair est strictement supérieure à  $\frac{1}{2}$ .

**Exemple 4.14** (loi uniforme sur un segment). Prenons  $\Omega = \mathbb{R}$ ,  $\mathcal{F} = \mathcal{B}$  la plus petite famille d'évènements observables (tribu) contenant tous les intervalles de  $\mathbb{R}$ . Pour un intervalle  $I$  de  $\mathbb{R}$ , notons  $\ell(I)$  sa longueur (éventuellement infinie). Soit  $[a, b]$  un segment fixé de  $\mathbb{R}$ . On définit une probabilité  $P$  sur  $(\mathbb{R}, \mathcal{B})$  en posant :

$$\forall A \in \mathcal{B}, \quad P(A) = \frac{\ell(A \cap [a, b])}{\ell([a, b])} = \frac{\ell(A \cap [a, b])}{b - a}. \quad (4.5)$$

8. Cet exemple peut paraître à première vue complètement artificiel. Nous verrons qu'il n'en est rien et que la loi de Poisson est sans doute la plus importante des probabilités que l'on peut définir sur  $\mathbb{N}$  du point de vue des applications.

Nous admettrons ce résultat. La définition générale de la longueur d'un ensemble  $A \cap [a, b]$  qui n'est pas forcément un intervalle, sort du cadre de ce cours. En pratique, nous n'appliquerons (4.5) qu'avec un  $A$  qui est soit un intervalle, soit une réunion finie d'intervalles disjoints.

La probabilité  $P$  définie par (4.5) est appelée *loi uniforme sur  $[a, b]$* . Remarquons que pour cette probabilité, tout singleton est de probabilité nulle ( $\forall x \in \mathbb{R}, P(\{x\}) = 0$ ), ce qui résulte de la propriété analogue de  $\ell$ . On voit sur cet exemple que la probabilité d'une union infinie *non dénombrable* d'évènements deux à deux disjoints n'est pas forcément égale à la somme de la famille correspondante de probabilités d'évènements. En effet,

$$1 = P([a, b]) = P\left(\bigcup_{x \in [a, b]} \{x\}\right) \neq \sum_{x \in [a, b]} P(\{x\}) = 0.$$

## 4.4 Probabilités conditionnelles

### 4.4.1 Introduction

Comment doit-on modifier la probabilité que l'on attribue à un évènement lorsque l'on dispose d'une *information supplémentaire*? Le concept de probabilité conditionnelle permet de répondre à cette question.

Par exemple, soit  $E$  l'ensemble des licenciés d'une fédération d'escrime,  $F$  le sous-ensemble des escrimeuses,  $H$  celui des escrimeurs,  $G$  l'ensemble des membres de la fédération qui sont gauchers et  $D$  celui des droitiers. On choisit un individu au hasard dans  $E$ , chaque licencié a donc une probabilité  $1/\text{card } E$  d'être choisi, ce qui revient à prendre comme espace probabilisé  $\Omega = E$ ,  $\mathcal{F} = \mathcal{P}(E)$  et  $P$  la probabilité uniforme (ou équiprobabilité) sur  $E$ . Si cet individu est un homme, quelle est la probabilité qu'il soit gaucher? Sans condition de sexe, la probabilité que l'individu choisi au hasard soit gaucher est naturellement  $\text{card } G/\text{card } E$ . Comment tenir compte de l'information supplémentaire apportée ici par la réalisation de l'évènement  $H$ ? On considère que tout se passe comme si d'emblée, on avait limité le tirage au sort aux éléments du sous-ensemble  $H$ . Ceci conduit à attribuer *dans ce nouveau contexte*, la probabilité  $\text{card}(G \cap H)/\text{card } H$  au choix d'un individu gaucher. Cette probabilité que l'individu choisi soit gaucher *sachant* que c'est un homme peut s'écrire :

$$\frac{\text{card}(G \cap H)}{\text{card } H} = \frac{\frac{\text{card}(G \cap H)}{\text{card } E}}{\frac{\text{card } H}{\text{card } E}} = \frac{P(G \cap H)}{P(H)}.$$

Par analogie, nous donnons dans le cas général la définition formelle suivante.

**Définition 4.15** (probabilité conditionnelle). *Soit  $H$  un évènement de probabilité non nulle. Pour tout évènement  $A$ , on définit :*

$$P(A | H) = \frac{P(A \cap H)}{P(H)},$$

*appelée probabilité conditionnelle de l'évènement  $A$  sous l'hypothèse  $H$ .*

Remarquons que pour l'instant, il ne s'agit que d'un jeu d'écriture. On a simplement défini un réel  $P(A | H)$  pour que :

$$P(A \cap H) = P(A | H)P(H).$$

Ce qui fait l'intérêt du concept de probabilité conditionnelle, c'est qu'il est souvent bien plus facile d'attribuer *directement* une valeur à  $P(A | H)$  en tenant compte des conditions expérimentales (liées à l'*information*  $H$ ) et d'en déduire ensuite la valeur de  $P(A \cap H)$ . Le raisonnement implicite alors utilisé est : tout espace probabilisé modélisant correctement la réalité expérimentale devrait fournir cette valeur pour  $P(A | H)$ .

**Exemple 4.16.** Une urne contient  $r$  boules rouges et  $v$  boules vertes. On en tire deux l'une après l'autre, sans remise. Quelle est la probabilité d'obtenir deux rouges ?

Notons  $H$  et  $A$  les évènements :

$$H = \{\text{rouge au 1}^{\text{er}} \text{ tirage}\}, \quad A = \{\text{rouge au 2}^{\text{e}} \text{ tirage}\}.$$

Un espace probabilisé  $(\Omega, \mathcal{F}, P)$  modélisant correctement cette expérience devrait vérifier :

$$\begin{aligned} P(H) &= \frac{r}{r+v}, \\ P(A | H) &= \frac{r-1}{r+v-1}. \end{aligned}$$

En effet, si  $H$  est réalisé, le deuxième tirage a lieu dans une urne contenant  $r+v-1$  boules dont  $r-1$  rouges. On en déduit :

$$P(\text{deux rouges}) = P(A \cap H) = P(A | H)P(H) = \frac{r-1}{r+v-1} \times \frac{r}{r+v}.$$

On aurait pu arriver au même résultat en prenant pour  $\Omega$  l'ensemble des arrangements de deux boules parmi  $r+v$ , muni de l'équiprobabilité et en faisant du dénombrement :

$$\text{card } \Omega = (r+v)(r+v-1), \quad \text{card}(A \cap H) = r(r-1).$$

d'où :

$$P(A \cap H) = \frac{r(r-1)}{(r+v)(r+v-1)}.$$

Notons d'ailleurs que  $\text{card } H = r(r+v-1)$  d'où

$$P(H) = \frac{\text{card } H}{\text{card } \Omega} = \frac{r(r+v-1)}{(r+v)(r+v-1)} = \frac{r}{r+v}.$$

En appliquant la définition formelle de  $P(A | H)$  on retrouve :

$$P(A | H) = \frac{P(A \cap H)}{P(H)} = \frac{r(r-1)}{(r+v)(r+v-1)} \times \frac{r+v}{r} = \frac{r-1}{r+v-1},$$

ce qui est bien la valeur que nous avons attribuée *a priori* en analysant les conditions expérimentales.

**Remarque 4.17.** Il importe de bien comprendre que l'écriture «  $A | H$  » ne désigne pas un nouvel évènement<sup>9</sup> différent de  $A$ . Quand on écrit  $P(A | H)$ , ce que l'on a modifié, ce n'est pas l'évènement  $A$ , mais la valeur numérique qui lui était attribuée par la fonction d'ensembles  $P$ . Il serait donc en fait plus correct d'écrire  $P_H(A)$  que  $P(A | H)$ . On conservera néanmoins cette dernière notation essentiellement pour des raisons typographiques :  $P(A | H_1 \cap H_2 \cap H_3)$  est plus lisible que  $P_{H_1 \cap H_2 \cap H_3}(A)$ .

9. En fait cette écriture prise isolément (sans le  $P$ ) *n'a pas de sens* et ne devrait *jamais* être utilisée. Le symbole  $|$  ne représente pas une opération sur les évènements qui l'entourent.



### 4.4.2 Propriétés

**Proposition 4.18.** Soit  $(\Omega, \mathcal{F}, P)$  un espace probabilisé et  $H$  un évènement fixé tel que  $P(H) \neq 0$ . Alors la fonction d'ensembles  $P_H = P(\cdot | H)$  définie par :

$$P(\cdot | H) : \mathcal{F} \rightarrow [0, 1] \quad B \mapsto P(B | H)$$

est une nouvelle probabilité sur  $(\Omega, \mathcal{F})$ .

*Preuve.* Il s'agit essentiellement de vérifier que  $P_H$  est  $\sigma$ -additive. Ceci pourra être vu en exercice.  $\square$

Une conséquence immédiate est que la fonction d'ensembles  $P(\cdot | H)$  vérifie toutes les propriétés de la proposition 4.4.

**Corollaire 4.19.** La fonction d'ensembles  $P(\cdot | H)$  vérifie :

1.  $P(\emptyset | H) = 0$ ,  $P(\Omega | H) = 1$  et si  $A \supset H$ ,  $P(A | H) = 1$ .
2. Si les  $A_i$  sont deux à deux disjoints :

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

3. Pour tout  $B \in \mathcal{F}$ ,  $P(B^c | H) = 1 - P(B | H)$ .
4. Pour tous  $A \in \mathcal{F}$  et  $B \in \mathcal{F}$ , si  $A \subset B$ ,  $P(A | H) \leq P(B | H)$ .
5. Pour tous  $A \in \mathcal{F}$  et  $B \in \mathcal{F}$ ,

$$P(A \cup B | H) = P(A | H) + P(B | H) - P(A \cap B | H).$$

6. Pour toute suite  $(A_i)_{i \geq 1}$  d'évènements :

$$P\left(\bigcup_{i \in \mathbb{N}^*} A_i | H\right) \leq \sum_{i=1}^{+\infty} P(A_i | H).$$

7. Si  $B_n \uparrow B$ ,  $P(B_n | H) \uparrow P(B | H)$ , ( $n \rightarrow +\infty$ ).
8. Si  $C_n \downarrow C$ ,  $P(C_n | H) \downarrow P(C | H)$ , ( $n \rightarrow +\infty$ ).

Nous n'avons vu jusqu'ici aucune formule permettant de calculer la probabilité d'une intersection d'évènements à l'aide des probabilités de ces évènements. Une telle formule n'existe pas dans le cas général. Les probabilités conditionnelles fournissent une méthode générale tout à fait naturelle pour calculer une probabilité d'intersection.

**Proposition 4.20** (règle des conditionnements successifs).

Si  $A_1, \dots, A_n$  sont  $n$  évènements tels que  $P(A_1 \cap \dots \cap A_{n-1}) \neq 0$ , on a :

$$P(A_1 \cap \dots \cap A_n) = P(A_1)P(A_2 | A_1)P(A_3 | A_1 \cap A_2) \times \dots \\ \dots \times P(A_n | A_1 \cap \dots \cap A_{n-1}).$$

*Preuve.* Pour  $1 \leq i \leq n-1$ , on a  $\bigcap_{j=1}^{n-1} A_j \subset \bigcap_{j=1}^i A_j$  d'où :

$$0 < P\left(\bigcap_{j=1}^{n-1} A_j\right) \leq P\left(\bigcap_{j=1}^i A_j\right).$$

Donc  $P(\bigcap_{j=1}^i A_j)$  n'est nul pour aucun  $i \leq n-1$  et on peut conditionner par l'évènement  $\bigcap_{j=1}^i A_j$ . Ceci légitime le calcul suivant :

$$\begin{aligned} & P(A_1)P(A_2 | A_1)P(A_3 | A_1 \cap A_2) \times \cdots \times P(A_n | A_1 \cap \dots \cap A_{n-1}) \\ = & P(A_1) \times \frac{P(A_1 \cap A_2)}{P(A_1)} \times \frac{P(A_1 \cap A_2 \cap A_3)}{P(A_1 \cap A_2)} \times \cdots \times \frac{P(A_1 \cap \dots \cap A_n)}{P(A_1 \cap \dots \cap A_{n-1})} \\ = & P(A_1 \cap \dots \cap A_n), \end{aligned}$$

après simplifications en chaîne de toutes ces fractions. □

**Exemple 4.21.** Une urne contient 30 boules : 20 bleues et 10 jaunes. On effectue trois tirages consécutifs et sans remise d'une boule. Quelle est la probabilité d'obtenir :

- trois boules jaunes ?
- une bleue, une jaune, une bleue dans cet ordre ?

Notons pour  $i = 1, 2, 3$ ,  $B_i$  et  $J_i$  les évènements « sortie d'une boule bleue au  $i^e$  tirage » et « sortie d'une boule jaune au  $i^e$  tirage ».

Nous cherchons d'abord  $P(J_1 \cap J_2 \cap J_3)$ . Par conditionnements successifs,

$$P(J_1 \cap J_2 \cap J_3) = P(J_1)P(J_2 | J_1)P(J_3 | J_1 \cap J_2).$$

Si  $J_1$  est réalisé, il y a dans l'urne juste avant le deuxième tirage 20 boules bleues et 9 boules jaunes, d'où  $P(J_2 | J_1) = 9/29$ . Si  $J_1 \cap J_2$  est réalisé, autrement dit si une boule jaune est sortie lors de chacun des deux premiers tirages, il reste 20 boules bleues et 8 jaunes juste avant le 3<sup>e</sup> tirage, d'où  $P(J_3 | J_1 \cap J_2) = 8/28$ . Ainsi

$$P(J_1 \cap J_2 \cap J_3) = \frac{10}{30} \times \frac{9}{29} \times \frac{8}{28} = \frac{1}{3} \times \frac{9}{29} \times \frac{2}{7} = \frac{6}{203} \simeq 0,029\ 557.$$

On procède de même pour calculer  $P(B_1 \cap J_2 \cap B_3)$  :

$$\begin{aligned} P(B_1 \cap J_2 \cap B_3) &= P(B_1)P(J_2 | B_1)P(B_3 | B_1 \cap J_2) \\ &= \frac{20}{30} \times \frac{10}{29} \times \frac{19}{28} \\ &= \frac{1}{3} \times \frac{5}{29} \times \frac{19}{7} = \frac{95}{609} \simeq 0,155\ 993. \end{aligned}$$

Les probabilités conditionnelles permettent aussi de calculer la probabilité d'un évènement en conditionnant par *tous les cas possibles*. Du point de vue ensembliste, ces *cas possibles* correspondent à une *partition* de  $\Omega$ .

**Définition 4.22** (partition). *On dit qu'une famille  $(H_i)_{i \in I}$  de parties de  $\Omega$  est une partition de  $\Omega$  si elle vérifie les trois conditions :*

- $\forall i \in I, H_i \neq \emptyset$ .
- $\Omega = \bigcup_{i \in I} H_i$ .
- Les  $H_i$  sont deux à deux disjoints ( $i \neq j \Rightarrow H_i \cap H_j = \emptyset$ ).

**Proposition 4.23** (conditionnement par les cas possibles<sup>10</sup>).

(i) *Si  $H$  est tel que  $P(H) \neq 0$  et  $P(H^c) \neq 0$ , on a*

$$\forall A \in \mathcal{F}, \quad P(A) = P(A | H)P(H) + P(A | H^c)P(H^c).$$

10. Ou formule des probabilités totales.

(ii) Si  $H_1, \dots, H_n$  est une partition finie de  $\Omega$  en évènements de probabilité non nulle,

$$\forall A \in \mathcal{F}, \quad P(A) = \sum_{i=1}^n P(A | H_i)P(H_i).$$

(iii) Si  $(H_i)_{i \in \mathbb{N}}$  est une partition de  $\Omega$  telle qu'aucun  $P(H_i)$  ne soit nul,

$$\forall A \in \mathcal{F}, \quad P(A) = \sum_{i=0}^{+\infty} P(A | H_i)P(H_i).$$

*Preuve.* Il suffit de vérifier (iii), les deux premières propriétés se démontrant de façon analogue. Comme  $(H_i)_{i \in \mathbb{N}}$  est une partition de  $\Omega$ ,

$$A = A \cap \Omega = A \cap \left( \bigcup_{i \in \mathbb{N}} H_i \right) = \bigcup_{i \in \mathbb{N}} (A \cap H_i)$$

et cette réunion est disjointe car les  $H_i$  étant deux à deux disjoints, il en est de même pour les  $A \cap H_i$ . Donc par  $\sigma$ -additivité :

$$P(A) = \sum_{i=0}^{+\infty} P(A \cap H_i) = \sum_{i=0}^{+\infty} P(A | H_i)P(H_i).$$

□

**Exemple 4.24.** On considère deux urnes  $U_1$  et  $U_2$ . L'urne  $U_1$  contient  $r_1$  boules rouges et  $v_1$  boules vertes. L'urne  $U_2$  contient  $r_2$  boules rouges et  $v_2$  boules vertes. On lance un dé. S'il indique le chiffre 1, on choisit l'urne  $U_1$ , sinon on choisit  $U_2$ . Dans chaque cas on effectue deux tirages avec remise dans l'urne choisie. Quelle est la probabilité d'obtenir une rouge au premier tirage ? deux rouges en tout ?

Adoptons les notations d'évènements suivantes :

$$R = \{\text{rouge au 1}^{\text{er}} \text{ tirage}\}, \quad R' = \{\text{rouge au 2}^{\text{e}} \text{ tirage}\},$$

$$H_1 = \{\text{choix de l'urne } U_1\}, \quad H_2 = H_1^c = \{\text{choix de l'urne } U_2\}.$$

Il est facile de calculer directement  $P(R | H_i)$  et  $P(R \cap R' | H_i)$  pour  $i = 1, 2$ . En effet, une fois l'urne  $U_i$  choisie, on a un problème classique de tirages avec remise dans la même urne que l'on peut traiter (par exemple) par le dénombrement. On a ainsi :

$$P(R | H_i) = \frac{r_i}{r_i + v_i}, \quad P(R \cap R' | H_i) = \left( \frac{r_i}{r_i + v_i} \right)^2.$$

La formule de conditionnement par la partition  $\{H_1, H_2\}$  donne :

$$\begin{aligned} P(R) &= P(R | H_1)P(H_1) + P(R | H_2)P(H_2) \\ &= \frac{1}{6} \frac{r_1}{r_1 + v_1} + \frac{5}{6} \frac{r_2}{r_2 + v_2} \end{aligned}$$

et

$$\begin{aligned} P(R \cap R') &= P(R \cap R' | H_1)P(H_1) + P(R \cap R' | H_2)P(H_2) \\ &= \frac{1}{6} \left( \frac{r_1}{r_1 + v_1} \right)^2 + \frac{5}{6} \left( \frac{r_2}{r_2 + v_2} \right)^2. \end{aligned}$$

Lorsqu'on a une partition de  $\Omega$  en  $n$  hypothèses ou cas possibles  $H_i$  et que l'on sait calculer les  $P(H_i)$  et les  $P(A | H_i)$ , on peut se poser le problème inverse : calculer  $P(H_j | A)$  à l'aide des quantités précédentes. La solution est donnée par la formule suivante quelquefois appelée (abusivement) formule des probabilités des causes.

**Proposition 4.25** (formule de Bayes).

Soit  $A$  un évènement de probabilité non nulle. Si les évènements  $H_i$  ( $1 \leq i \leq n$ ) forment une partition de  $\Omega$  et si aucun  $P(H_i)$  n'est nul, on a pour tout  $j \in \llbracket 1, n \rrbracket$  :

$$P(H_j | A) = \frac{P(A | H_j)P(H_j)}{\sum_{i=1}^n P(A | H_i)P(H_i)}.$$

*Preuve.* On note que par définition des probabilités conditionnelles :

$$P(H_j | A) = \frac{P(A \cap H_j)}{P(A)} = \frac{P(A | H_j)P(H_j)}{P(A)}.$$

Il ne reste plus qu'à développer  $P(A)$  en conditionnant par la partition  $(H_i, 1 \leq i \leq n)$  comme à la proposition 4.23.  $\square$

La même formule se généralise au cas d'une partition dénombrable. Ces formules sont plus faciles à retrouver qu'à mémoriser.

**Exemple 4.26.** Un questionnaire à choix multiples propose  $m$  réponses pour chaque question. Soit  $p \in ]0, 1[$  la probabilité qu'un étudiant connaisse la réponse à une question donnée. S'il ignore la réponse, il choisit au hasard l'une des réponses proposées. Quelle est pour le correcteur, la probabilité qu'un étudiant connaisse vraiment la bonne réponse lorsqu'il l'a donnée ?

Notons :

$$\begin{aligned} B &= \{\text{l'étudiant donne la bonne réponse}\}, \\ C &= \{\text{l'étudiant connaît la bonne réponse}\}. \end{aligned}$$

On cherche  $P(C | B)$ . Avec ces notations, les données de l'énoncé se traduisent par :

$$P(C) = p, \quad P(C^c) = 1 - p, \quad P(B | C) = 1, \quad P(B | C^c) = \frac{1}{m}.$$

On en déduit :

$$\begin{aligned} P(C | B) &= \frac{P(B \cap C)}{P(B)} \\ &= \frac{P(B | C)P(C)}{P(B | C)P(C) + P(B | C^c)P(C^c)} \\ &= \frac{1 \times p}{1 \times p + \frac{1}{m}(1 - p)} = \frac{mp}{1 + (m - 1)p}. \end{aligned}$$

Notons  $P(C | B) = f(m, p)$ . Pour  $p$  fixé,  $f(\cdot, p)$  croît sur  $\mathbb{N}^*$  depuis  $f(1, p) = p$  jusqu'à  $\lim_{m \rightarrow +\infty} f(m, p) = 1$ . D'autre part pour  $m$  fixé,  $f(m, \cdot)$  croît sur  $]0, 1[$  avec  $\lim_{p \rightarrow 0} f(m, p) = 0$  et  $\lim_{p \rightarrow 1} f(m, p) = 1$ . On a pour  $0 < p < 1$  :

$$\frac{P(C | B)}{p} = \frac{m}{1 + (m - 1)p} \geq 1, \quad \text{d'où } P(C | B) > p.$$

Si  $p = 0$  (resp.  $p = 1$ ), on ne peut plus conditionner par  $C$  (resp. par  $C^c$ ) et dans le calcul ci-dessus, seule l'égalité  $P(C | B) = P(B \cap C)/P(B)$  reste valable. Pour  $p = 0$ ,  $P(B \cap C) \leq P(C) = 0$ , d'où  $P(C | B) = 0$ . Pour  $p = 1$ ,  $P(B \cap C^c) \leq P(C^c) = 0$  d'où  $P(B \cap C^c) = 0$  et  $P(B \cap C) = P(B) - P(B \cap C^c) = P(B)$  d'où  $P(C | B) = 1$ . Tous ces résultats sont conformes à l'intuition.

**Exemple 4.27.** Un test sanguin a une probabilité de 0,95 de détecter un certain virus lorsque celui-ci est effectivement présent. Il donne néanmoins un *faux résultat positif* pour 1% des personnes non infectées. Si 0,5% de la population est porteuse du virus, quelle est la probabilité qu'une personne ait le virus sachant qu'elle a un test positif ?

Notons :

$$\begin{aligned} V &= \{\text{la personne testée a le virus}\}, \\ T^+ &= \{\text{la personne testée a un test positif}\}. \end{aligned}$$

On cherche  $P(V | T^+)$ . Or on sait que  $P(V) = 0,005$ ,  $P(T^+ | V) = 0,95$  et  $P(T^+ | V^c) = 0,01$ . On en déduit :

$$\begin{aligned} P(V | T^+) &= \frac{P(T^+ \cap V)}{P(T^+)} = \frac{P(T^+ | V)P(V)}{P(T^+ | V)P(V) + P(T^+ | V^c)P(V^c)} \\ &= \frac{0,95 \times 0,005}{0,95 \times 0,005 + 0,01 \times 0,995} \simeq 0,323. \end{aligned}$$

On voit ainsi que contrairement à ce que l'on aurait pu croire, le test n'est pas fiable : si la personne présente un test positif, il y a approximativement deux chances sur trois qu'elle ne soit pas porteuse du virus !

## 4.5 Indépendance

### 4.5.1 Indépendance de deux événements

Soient  $A$  et  $B$  deux événements de probabilité non nulle. Il arrive que la connaissance de la réalisation de  $A$  ne modifie pas notre information sur celle de  $B$ , autrement dit que  $P(B | A) = P(B)$ . C'est le cas par exemple lorsque l'on fait deux tirages avec remise et que la réalisation de  $A$  ne dépend que du résultat du premier tirage, celle de  $B$  que du deuxième. Symétriquement on aura dans cet exemple  $P(A | B) = P(A)$ . Cette remarque se généralise :

**Proposition 4.28.** *Si  $A$  et  $B$  sont des événements de probabilité non nulle, les trois égalités suivantes sont équivalentes :*

- (i)  $P(B | A) = P(B)$ ,
- (ii)  $P(A | B) = P(A)$ ,
- (iii)  $P(A \cap B) = P(A)P(B)$ .

*Preuve.* Comme  $P(A) \neq 0$  et  $P(B) \neq 0$ , on a la chaîne d'équivalences :

$$\frac{P(A \cap B)}{P(A)} = P(B) \Leftrightarrow P(A \cap B) = P(A)P(B) \Leftrightarrow \frac{P(A \cap B)}{P(B)} = P(A).$$

□

D'autre part la relation (iii) est toujours vérifiée dans le cas dégénéré où  $P(A) = 0$  ou  $P(B) = 0$ . En effet, on a alors à la fois  $P(A)P(B) = 0$  et  $0 \leq P(A \cap B) \leq \min(P(A), P(B)) = 0$  d'où  $P(A \cap B) = 0$ . Ainsi la relation (iii) est un peu plus générale que (i) et (ii). Elle a aussi sur les deux autres l'avantage de la symétrie d'écriture. C'est elle que l'on retient pour définir l'indépendance.

**Définition 4.29.** Soit  $(\Omega, \mathcal{F}, P)$  un espace probabilisé. Deux évènements  $A$  et  $B$  de cet espace sont dits indépendants lorsque :

$$P(A \cap B) = P(A)P(B).$$

**Exemple 4.30.** On jette deux fois le même dé. Les évènements

$$\begin{aligned} A &= \{\text{obtention d'un chiffre pair au premier lancer}\}, \\ B &= \{\text{obtention du 1 au deuxième lancer}\}, \end{aligned}$$

sont indépendants.

En effet, en prenant  $\Omega = \llbracket 1, 6 \rrbracket^2$ ,  $\mathcal{F} = \mathcal{P}(\Omega)$  et  $P$  l'équiprobabilité, on vérifie que :

$$\begin{aligned} P(A) &= \frac{3 \times 6}{36} = \frac{1}{2}, & P(B) &= \frac{6 \times 1}{36} = \frac{1}{6}, \\ P(A \cap B) &= \frac{3 \times 1}{36} = \frac{1}{12}, & P(A)P(B) &= \frac{1}{2} \times \frac{1}{6} = \frac{1}{12}. \end{aligned}$$

**Remarques 4.31.**

- Si  $A$  est un évènement tel que  $P(A) = 0$  ou  $P(A) = 1$ , alors il est indépendant de tout évènement, *y compris de lui-même* (c'est le cas en particulier pour  $\Omega$  et  $\emptyset$ ).
- Deux évènements *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  $P(A)P(B) \neq 0$ .
- L'indépendance de deux évènements  $A$  et  $B$  n'est pas une propriété intrinsèque aux évènements, elle est toujours relative au modèle  $(\Omega, \mathcal{F}, P)$  que l'on a choisi. Voici un exemple pour l'illustrer.

**Exemple 4.32.** Une urne contient 12 boules numérotées de 1 à 12. On en tire une au hasard et on considère les évènements :

$$A = \{\text{tirage d'un nombre pair}\}, \quad B = \{\text{tirage d'un multiple de 3}\}.$$

L'espace probabilisé qui s'impose naturellement ici est  $\Omega = \llbracket 1, 12 \rrbracket$  muni de l'équiprobabilité  $P$ . Les évènements  $A$  et  $B$  s'écrivent :

$$A = \{2, 4, 6, 8, 10, 12\}, \quad B = \{3, 6, 9, 12\}, \quad A \cap B = \{6, 12\}.$$

Dans ce modèle,

$$\begin{aligned} P(A) &= \frac{6}{12} = \frac{1}{2}, & P(B) &= \frac{4}{12} = \frac{1}{3}, \\ P(A \cap B) &= \frac{2}{12} = \frac{1}{6}, & P(A)P(B) &= \frac{1}{2} \times \frac{1}{3} = \frac{1}{6}, \end{aligned}$$

donc  $A$  et  $B$  sont indépendants.

On rajoute maintenant dans l'urne une boule numérotée treize et on recommence l'expérience. Les évènements  $A$  et  $B$  restent les mêmes, mais *le modèle a changé*. On a maintenant l'équiprobabilité  $P'$  sur  $\Omega' = \llbracket 1, 13 \rrbracket$  et

$$P'(A) = \frac{6}{13}, \quad P'(B) = \frac{4}{13}, \quad P'(A \cap B) = \frac{2}{13},$$

mais

$$P'(A)P'(B) = \frac{6}{13} \times \frac{4}{13} = \frac{24}{169} \neq \frac{2}{13},$$

donc  $A$  et  $B$  ne sont plus indépendants. Un peu de réflexion permet de relier ces résultats calculatoires avec la notion intuitive d'indépendance présentée en introduction. Dans le premier cas, la proportion des multiples de trois parmi les pairs est la même que parmi les impairs. Le fait de savoir que la boule tirée est paire ne modifie en rien notre information sur  $B$ . Par contre dans le deuxième cas, l'ajout de la treizième boule modifie la proportion des multiples de trois : elle est plus élevée chez les pairs que chez les impairs. Donc le fait de savoir que la boule tirée est paire augmente un peu la probabilité que nous pouvons attribuer à  $B$ .

**Proposition 4.33.** *Si  $A$  et  $B$  sont indépendants, il en est de même pour les paires d'évènements  $A$  et  $B^c$ ,  $A^c$  et  $B$ ,  $A^c$  et  $B^c$ .*

*Preuve.* Par hypothèse,  $P(A \cap B) = P(A)P(B)$ . En considérant la réunion disjointe  $A = (A \cap B) \cup (A \cap B^c)$ , nous avons :  $P(A \cap B^c) = P(A) - P(A \cap B)$ , d'où :

$$P(A \cap B^c) = P(A) - P(A)P(B) = P(A)(1 - P(B)) = P(A)P(B^c).$$

Donc  $A$  et  $B^c$  sont indépendants. L'échange des rôles de  $A$  et  $B$  dans ce raisonnement donne l'indépendance de  $A^c$  et  $B$ . En réutilisant le premier résultat avec  $A^c$  à la place de  $A$ , on obtient alors celle de  $A^c$  et  $B^c$ .  $\square$

## 4.5.2 Indépendance mutuelle

On se propose de généraliser la notion d'indépendance à plus de deux évènements. Examinons d'abord la situation suivante.

**Exemple 4.34.** Une urne contient quatre jetons : un bleu, un blanc, un rouge et un bleu-blanc-rouge. On en tire un au hasard. Considérons les trois évènements

$$\begin{aligned} A &= \{\text{le jeton tiré contient du bleu}\}, \\ B &= \{\text{le jeton tiré contient du blanc}\}, \\ C &= \{\text{le jeton tiré contient du rouge}\}. \end{aligned}$$

Il est clair que  $P(A) = P(B) = P(C) = 2/4 = 1/2$ . D'autre part :

$$P(A \cap B) = P(\text{tricolore}) = \frac{1}{4} = P(A)P(B)$$

et de même  $P(B \cap C) = 1/4 = P(B)P(C)$ ,  $P(C \cap A) = 1/4 = P(C)P(A)$ . Ainsi les évènements  $A$ ,  $B$ ,  $C$  sont *deux à deux indépendants*.

D'autre part  $P(A | B \cap C) = 1$  car  $B \cap C = \{\text{tricolore}\}$ . Donc la connaissance de la réalisation *simultanée* de  $B$  et  $C$  modifie notre information sur  $A$ . La notion d'indépendance deux à deux n'est donc pas suffisante pour traduire l'idée intuitive d'indépendance de plusieurs évènements. Ceci motive la définition suivante.

**Définition 4.35.** *Trois évènements  $A$ ,  $B$ ,  $C$  sont dits mutuellement indépendants lorsqu'ils vérifient les quatre conditions :*

$$\begin{aligned} P(A \cap B) &= P(A)P(B), \\ P(B \cap C) &= P(B)P(C), \\ P(C \cap A) &= P(C)P(A), \\ P(A \cap B \cap C) &= P(A)P(B)P(C). \end{aligned}$$

Avec cette définition de l'indépendance des évènements  $A$ ,  $B$  et  $C$  on a bien<sup>11</sup>  $P(A | B) = P(A)$ ,  $P(A | B \cap C) = P(A)$ , ainsi que toutes les égalités qui s'en déduisent par permutation sur les lettres  $A$ ,  $B$ ,  $C$ . On peut généraliser cette définition comme suit.

**Définition 4.36.** *Les  $n$  évènements  $A_1, \dots, A_n$  sont dits mutuellement indépendants si pour toute sous-famille  $A_{i_1}, \dots, A_{i_k}$  avec  $1 \leq i_1 < \dots < i_k \leq n$ ,*

$$P(A_{i_1} \cap \dots \cap A_{i_k}) = P(A_{i_1}) \times \dots \times P(A_{i_k}). \quad (4.6)$$

L'indépendance mutuelle implique évidemment l'indépendance deux à deux et la réciproque est fautive comme le montre l'exemple 4.34. Dans toute la suite, lorsque nous parlerons d'une famille de plusieurs évènements indépendants sans autre précision, nous sous-entendrons systématiquement *mutuellement* indépendants.

**Proposition 4.37.** *Si  $\{A_1, \dots, A_n\}$  est une famille de  $n$  évènements indépendants, toute famille obtenue en remplaçant certains des  $A_i$  par leur complémentaire est encore indépendante.*

*Preuve.* Supposons la proposition démontrée dans le cas où l'on a remplacé *un seul*  $A_i$  par son complémentaire. Le cas général s'en déduit en utilisant cette propriété autant de fois qu'il y a de  $A_i$  changés en leur complémentaire. Dans le cas d'un seul  $A_i$  remplacé, on ne perd pas de généralité en supposant qu'il s'agit de  $A_1$  (il suffit de changer l'indexation des évènements, ce qui n'affecte pas leur indépendance mutuelle). Il nous reste alors à vérifier (4.6) dans le cas où  $i_1 = 1$  avec  $A_{i_1}^c$  à la place de  $A_{i_1}$  (dans le cas  $i_1 > 1$ , l'égalité ne fait intervenir que des éléments de la famille initiale et il n'y a donc rien à vérifier). Posons  $B = A_{i_2} \cap \dots \cap A_{i_k}$ . L'hypothèse (4.6) appliquée à la famille  $A_{i_2}, \dots, A_{i_k}$  nous donne  $P(B) = P(A_{i_2}) \times \dots \times P(A_{i_k})$ . La même hypothèse appliquée à  $A_{i_1}, \dots, A_{i_k}$  nous donne alors l'indépendance de  $A_{i_1}$  et de  $B$ . Par la proposition 4.33, on en déduit :

$$P(A_{i_1}^c \cap B) = P(A_{i_1}^c) \times P(B) = P(A_{i_1}^c) \times P(A_{i_2}) \times \dots \times P(A_{i_k}),$$

ce qui achève la preuve. □

**Définition 4.38** (indépendance d'une suite d'évènements). *Une suite infinie d'évènements est dite indépendante si toute sous-suite finie est formée d'évènements mutuellement indépendants.*

**Remarque 4.39.** Compte-tenu de la proposition 4.37, on voit immédiatement que si  $(A_i)_{i \in \mathbb{N}}$  est une suite indépendante d'évènements, toute suite formée en remplaçant certains des  $A_i$  (éventuellement tous) par leur complémentaire est encore indépendante.

### 4.5.3 Épreuves répétées

Considérons une suite d'épreuves réalisées dans les mêmes conditions expérimentales, par exemple tirages avec remise dans la même urne, lancers successifs d'un dé, etc. Il est alors raisonnable de supposer que les résultats de tout sous-ensemble fini d'épreuves n'ont aucune influence sur ceux des autres épreuves.

**Définition 4.40.** *On dit que les épreuves sont indépendantes si toute suite  $(A_i)_{i \geq 1}$  telle que la réalisation de chaque  $A_i$  est déterminée uniquement par le résultat de la  $i^e$  épreuve est une suite indépendante d'évènements.*

11. Lorsque les probabilités conditionnelles existent.



**Exemple 4.41.** On réalise une suite d'épreuves indépendantes. Chaque épreuve résulte en un succès avec probabilité  $p \in ]0, 1[$  ou en un échec avec probabilité  $q = 1 - p$ . Quelle est la probabilité des évènements suivants ?

- a)  $A = \{\text{Au moins un succès au cours des } n \text{ premières épreuves}\}.$
- b)  $B = \{\text{Exactement } k \text{ succès au cours des } n \text{ premières épreuves}\}.$
- c)  $C = \{\text{Toutes les épreuves donnent un succès}\}.$

Notons pour tout  $i \geq 1$  :  $R_i = \{\text{succès à la } i^{\text{e}} \text{ épreuve}\}$ ,  $R_i^c$  est alors l'échec à la  $i^{\text{e}}$  épreuve.

- a)  $A = \bigcup_{i=1}^n R_i$ , d'où  $A^c = \bigcap_{i=1}^n R_i^c$ . Les  $R_i^c$  étant indépendants, on en déduit :

$$P(A^c) = \prod_{i=1}^n P(R_i^c) = (1 - p)^n = q^n,$$

d'où  $P(A) = 1 - q^n$ .

- b) Traitons d'abord le cas  $0 < k < n$ . L'évènement  $B$  est la réunion disjointe de tous les évènements du type :

$$B_I = \left( \bigcap_{i \in I} R_i \right) \cap \left( \bigcap_{j \in J} R_j^c \right),$$

où  $I$  est une partie de cardinal  $k$  de  $\llbracket 1, n \rrbracket$  et  $J$  son complémentaire dans  $\llbracket 1, n \rrbracket$ . L'ensemble d'indices  $I$  représente un choix possible des  $k$  épreuves donnant un succès, les autres épreuves indexées par  $J$  donnant alors un échec. En considérant tous les choix possibles de l'ensemble  $I$  (il y en a  $C_n^k$ ), on obtient une partition de  $B$  par les  $B_I$ . Par indépendance des épreuves, pour tout  $I$  on a :

$$P(B_I) = \prod_{i \in I} P(R_i) \times \prod_{j \in J} P(R_j^c) = p^k q^{n-k}.$$

On voit ainsi que  $P(B_I)$  ne dépend pas de  $I$ . On en déduit :

$$P(B) = \sum_{\substack{I \subset \llbracket 1, n \rrbracket \\ \text{card } I = k}} P(B_I) = C_n^k p^k q^{n-k}.$$

La vérification de la validité de la formule  $P(B) = C_n^k p^k q^{n-k}$  dans les cas  $k = 0$  et  $k = n$  est laissée au lecteur.

- c) Pour  $n \geq 1$ , soit  $C_n = \{\text{succès aux } n \text{ premières épreuves}\}$ . Clairement  $C$  est inclus dans  $C_n$  donc  $0 \leq P(C) \leq P(C_n)$ . En utilisant l'indépendance des  $R_i$  on obtient :

$$P(C_n) = P\left(\bigcap_{i=1}^n R_i\right) = \prod_{i=1}^n P(R_i) = p^n.$$

donc pour tout  $n \geq 1$ ,  $0 \leq P(C) \leq p^n$ . En faisant tendre  $n$  vers  $+\infty$ , on en déduit  $P(C) = 0$ .



# Chapitre 5

## Variables aléatoires

### 5.1 Introduction

DANS de nombreux jeux, on fait intervenir le hasard en observant la somme des points marqués par deux dés. Considérons le jet d'un dé bleu et d'un dé rouge et notons  $S$  la somme des points obtenus. On modélise cette expérience en prenant l'équiprobabilité sur  $\Omega = \llbracket 1, 6 \rrbracket^2$ . Un événement élémentaire  $\omega$  est ici un couple  $(b, r)$  où  $b$  désigne le résultat du dé bleu et  $r$  celui du rouge et  $S(\omega) = b + r$ . Il est commode de décrire la situation par un tableau à 36 cases en écrivant la valeur de  $S(\omega)$  dans la case représentant  $\omega = (b, r)$  à l'intersection de la ligne  $b$  et de la colonne  $r$ .

	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

FIGURE 5.1 – Somme des points de deux dés

On a ainsi défini une application  $S$  de  $\Omega$  dans l'ensemble des sommes de points possibles :  $\llbracket 2, 12 \rrbracket$ . On dit que  $S$  est une *variable aléatoire* sur  $\Omega$ . En fait, l'observation qui nous intéresse dans cette expérience, ce n'est pas  $\omega$ , mais seulement  $S(\omega)$ . Ce que l'on aimerait connaître, c'est la probabilité que la somme des points prenne une valeur donnée, soit  $P(S = k)$  pour  $k \in \llbracket 2, 12 \rrbracket$ . Ici la notation «  $P(S = k)$  » est un abus d'écriture commode pour désigner  $P(\{\omega \in \Omega ; S(\omega) = k\})$ . En utilisant l'équiprobabilité sur  $\Omega$  et la figure 5.1, nous obtenons le tableau 5.1.

$k$	2	3	4	5	6	7	8	9	10	11	12
$P(S = k)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

TABLE 5.1 – Probabilités des valeurs de la somme des points

Cela revient à considérer un nouvel ensemble d'événements élémentaires :

$$\Omega' = S(\Omega) = \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$$

et à le munir de la probabilité  $P_S$  définie par le tableau 5.1, plus précisément :

$$\forall B \in \mathcal{P}(\Omega'), \quad P_S(B) := \sum_{k \in B} P(S = k) = P(S \in B). \quad (5.1)$$

Cette nouvelle probabilité  $P_S$  s'appelle *loi de la variable aléatoire  $S$* . En d'autres termes, nous avons réalisé *via  $S$*  un *transfert* de l'espace probabilisé  $(\Omega, \mathcal{P}(\Omega), P)$  sur l'espace probabilisé  $(\Omega', \mathcal{P}(\Omega'), P_S)$ . Remarquons maintenant que l'on peut facilement agrandir l'ensemble d'arrivée de  $S$  en remplaçant  $\Omega'$  par  $\Omega'' = \mathbb{R}$ . On peut munir  $\mathbb{R}$  au choix de la tribu  $\mathcal{P}(\mathbb{R})$  ou de la tribu

$$\mathcal{B} \text{ plus petite tribu contenant les intervalles.} \quad (5.2)$$

Il y a de bonnes raisons à cet agrandissement de l'ensemble d'arrivée, ne serait-ce que pour pouvoir faire des opérations sur les variables aléatoires. Imaginons par exemple que l'on jette  $n$  fois la paire de dés et que l'on s'intéresse à la moyenne arithmétique  $M_n$  des sommes obtenues. Il ne serait guère commode de travailler avec un espace d'arrivée  $\Omega'_n := M_n(\Omega^n)$ , surtout si l'on s'intéresse au comportement de  $M_n$  pour  $n$  tendant vers l'infini. Nous prendrons donc désormais comme ensemble d'arrivée  $\mathbb{R}$  pour les variables aléatoires que nous étudierons.

Si  $X(\omega)$  est une grandeur physique (masse, température, pression, longueur, etc.) mesurée à partir du résultat  $\omega$  d'une expérience, il n'est en général pas possible de la déterminer avec une précision absolue. Tout ce que l'on peut dire est que  $X$  appartient à un certain intervalle, dont la longueur dépend de la précision de l'instrument de mesure utilisé. Les quantités pertinentes pour identifier la *loi* de  $X$  sont alors les  $P(X \in I)$  pour  $I$  intervalle de  $\mathbb{R}$ , plutôt que les  $P(X = x)$  qui pourraient être nulles pour tout  $x$  réel. Nous définirons la loi de  $X$  en posant  $P_X(B) = P(X \in B)$  pour  $B \in \mathcal{B}$ , *sous réserve que cela ait un sens*. Voyons cela de plus près.

Soit  $X$  une application  $\Omega \rightarrow \mathbb{R}$ . Pour tout  $B \subset \mathbb{R}$ , notons

$$\{X \in B\} := \{\omega \in \Omega; X(\omega) \in B\} =: X^{-1}(B).$$

Cette écriture «  $X^{-1}$  » ne suppose aucunement la bijectivité de  $X$ . Il s'agit seulement d'une notation commode pour « l'ensemble des antécédents des éléments de  $B$  par l'application  $X$  ». On aimerait pouvoir transporter par  $X$  la probabilité  $P$ , mesure définie sur  $(\Omega, \mathcal{F})$  en une probabilité  $P_X$ , définie sur  $(\mathbb{R}, \mathcal{B})$  en posant :

$$\forall B \in \mathcal{B}, \quad P_X(B) := P(X^{-1}(B)) = P(X \in B).$$

Pour que cette écriture ait un sens, encore faut-il que  $X^{-1}(B)$  soit un élément de la tribu  $\mathcal{F}$  sur laquelle est définie  $P$ . Nous supposons donc que  $X$  vérifie la condition :

$$\forall B \in \mathcal{B}, \quad X^{-1}(B) \in \mathcal{F}. \quad (5.3)$$

Nous réserverons le nom de *variable aléatoire* aux applications  $X : \Omega \rightarrow \mathbb{R}$  vérifiant (5.3). En pratique, cette condition (5.3) sera toujours vérifiée.

## 5.2 Généralités

### 5.2.1 Variables aléatoires réelles

**Définition 5.1** (variable aléatoire réelle). *Soit  $(\Omega, \mathcal{F}, P)$  un espace probabilisé. On appelle variable aléatoire réelle sur  $(\Omega, \mathcal{F})$ , ou plus simplement variable aléatoire, toute application  $X$  :*

$$X : \Omega \longrightarrow \mathbb{R} \quad \omega \mapsto X(\omega),$$

vérifiant :

$$\text{pour tout intervalle } I \text{ de } \mathbb{R}, \quad X^{-1}(I) \in \mathcal{F}.$$

Nous admettrons que cette condition implique (5.3) et que cela permet de définir une probabilité  $P_X$  sur  $(\mathbb{R}, \mathcal{B})$  par

$$\forall B \in \mathcal{B}, \quad P_X(B) = P(X^{-1}(B)) = P(X \in B).$$

**Remarque 5.2.** Il importe de noter que la mesure de probabilité  $P$  sur  $(\Omega, \mathcal{F})$  ne joue *aucun rôle* dans la définition de la notion de variable aléatoire. C'est pour cela que nous parlons de « variable aléatoire sur  $(\Omega, \mathcal{F})$  » plutôt que sur  $(\Omega, \mathcal{F}, P)$ .

**Définition 5.3** (variable aléatoire discrète). *On appelle variable aléatoire discrète sur  $(\Omega, \mathcal{F})$ , toute application  $X : \Omega \rightarrow \mathbb{R}$  vérifiant les deux conditions suivantes.*

(i) *L'ensemble des images  $X(\Omega) = \{X(\omega), \omega \in \Omega\}$  est une partie au plus dénombrable de  $\mathbb{R}$ . On peut donc numéroter ses éléments par des indices entiers<sup>1</sup>*

$$X(\Omega) = \{x_0, x_1, \dots, x_k, \dots\}.$$

(ii)  *$X$  est une variable aléatoire réelle, ce qui équivaut ici à*

$$\forall x \in X(\Omega), \quad X^{-1}(\{x\}) \in \mathcal{F}. \quad (5.4)$$

Notons que si  $X(\Omega)$  est au plus dénombrable, pour tout intervalle  $I$ ,  $I \cap X(\Omega)$  est aussi au plus dénombrable et c'est ce qui permet de écrire  $X^{-1}(I)$  comme union au plus dénombrable d'évènements  $X^{-1}(\{x\})$  pour  $x \in X(\Omega)$ .

### 5.2.2 Loi d'une variable aléatoire

Nous admettrons la proposition suivante.

**Proposition 5.4.** *Soit  $X$  une variable aléatoire sur  $(\Omega, \mathcal{F})$  et  $P$  une probabilité sur  $(\Omega, \mathcal{F})$ . La fonction d'ensembles  $P_X = P \circ X^{-1}$  définie sur  $\mathcal{B}$  par*

$$\forall B \in \mathcal{B}, \quad P_X(B) := P(X^{-1}(B)) = P(X \in B) \quad (5.5)$$

*est une probabilité sur  $(\mathbb{R}, \mathcal{B})$ .*

Autrement dit, une variable aléatoire  $X$  permet de *transporter* la probabilité  $P$  définie sur  $(\Omega, \mathcal{F})$  en une probabilité  $P_X$  définie sur  $(\mathbb{R}, \mathcal{B})$ .

**Définition 5.5** (loi d'une variable aléatoire). *Soient  $(\Omega, \mathcal{F}, P)$  un espace probabilisé et  $X$  une variable aléatoire sur  $(\Omega, \mathcal{F})$ . On appelle loi de  $X$  sous  $P$ , ou plus simplement loi de  $X$ , la probabilité  $P_X$  sur  $(\mathbb{R}, \mathcal{B})$  définie par (5.5).*

Si  $\mu$  est une mesure de probabilité sur  $(\mathbb{R}, \mathcal{B})$ , on dit que  $X$  « suit » la loi  $\mu$  si  $P_X = P \circ X^{-1} = \mu$ , autrement dit, si la loi de  $X$  sous  $P$  est la mesure  $\mu$ .

---

1. Pour tous les exemples classiques que nous rencontrerons, il est possible de les numéroter de manière croissante :  $x_0 < x_1 < x_2 \dots$ . Mais ce n'est pas toujours le cas, car  $X(\Omega)$  peut être par exemple, l'ensemble des décimaux (ou des rationnels) de  $[0, 1]$ .

**Remarque 5.6.** Dans les problèmes usuels de probabilités, on travaille souvent avec un seul  $(\Omega, \mathcal{F}, P)$  et on se contente alors de l'appellation *loi de  $X$* . Il n'en va pas de même en statistique où l'on met généralement en concurrence *plusieurs* modèles  $(\Omega, \mathcal{F}, P_\theta)$ , où  $\theta$  est un paramètre inconnu et où on se propose de choisir un de ces modèles au vu des valeurs  $X(\omega)$  observées. C'est là que l'appellation *loi de  $X$  sous  $P_\theta$*  s'impose. Pour donner un exemple simple, considérons le problème du sondage d'un échantillon de 500 personnes avant le second tour d'une élection présidentielle opposant le candidat  $A$  au candidat  $B$ . Ici  $\theta$  est la proportion *inconnue* d'électeurs votant  $A$  dans la population totale. Si  $X$  est le nombre de personnes interrogées favorables à  $A$ , la loi de  $X$  sous  $P_\theta$  est la loi binomiale<sup>2</sup>  $\text{Bin}(500, \theta)$ .

Une autre situation où il est naturel de considérer plusieurs lois pour une même variable aléatoire est celle du *conditionnement*. Rappelons que si  $(\Omega, \mathcal{F}, P)$  est un espace probabilisé et  $H \in \mathcal{F}$  un évènement tel que  $P(H) > 0$ , on peut définir sur  $\mathcal{F}$  une nouvelle probabilité  $P_H = P(\cdot | H)$  par

$$\forall A \in \mathcal{F}, \quad P_H(A) := P(A | H) = \frac{P(A \cap H)}{P(H)}.$$

**Définition 5.7** (loi conditionnelle). *Soient  $(\Omega, \mathcal{F}, P)$  un espace probabilisé,  $H \in \mathcal{F}$  tel que  $P(H) > 0$ ,  $X$  une variable aléatoire sur  $(\Omega, \mathcal{F})$ . On appelle loi conditionnelle de  $X$  sachant  $H$ , la loi de  $X$  sous  $P_H$ . En la notant  $P_{X|H}$ , on a donc*

$$\forall B \in \mathcal{B}, \quad P_{X|H}(B) = P_H(X^{-1}(B)) = P(X \in B | H).$$

Il importe de ne pas se laisser induire en erreur par la notation  $P_{X|H}$ , elle ne concerne pas une nouvelle variable aléatoire «  $X | H$  » mais bien toujours *la même* variable aléatoire  $X$ . Ce qui a changé, c'est la probabilité dont on munit  $(\Omega, \mathcal{F})$  et sous laquelle on considère la loi de  $X$ .

**Remarque 5.8.** Deux variables aléatoires peuvent *avoir même loi sans être égales*. Par exemple considérons le jet de deux dés, l'un bleu et l'autre rouge. Notons  $X$  le nombre de points indiqué par le dé bleu et  $Y$  celui du rouge. Les variables aléatoires  $X$  et  $Y$  sont définies sur le même espace probabilisé  $\Omega = \{1, 2, 3, 4, 5, 6\}^2$  muni de l'équiprobabilité. On a  $X(\Omega) = Y(\Omega) = \llbracket 1, 6 \rrbracket$  et :

$$\forall k \in \llbracket 1, 6 \rrbracket, \quad P(X = k) = \frac{1}{6}, \quad P(Y = k) = \frac{1}{6}.$$

Donc  $X$  et  $Y$  ont même loi :  $P_X = P_Y = \sum_{k=1}^6 \frac{1}{6} \delta_k$ . Pour autant, on n'a pas l'égalité des variables aléatoires  $X$  et  $Y$  qui signifierait  $X(\omega) = Y(\omega)$  pour tout  $\omega \in \Omega$  (égalité de deux applications). Autrement dit, en lançant deux dés on obtiendrait à *coup sûr* un double. En revanche nous pouvons considérer l'évènement  $\{X = Y\}$  dont la réalisation n'est pas certaine et calculer sa probabilité :

$$P(X = Y) = P\left(\bigcup_{k=1}^6 \{(X, Y) = (k, k)\}\right) = \frac{6}{36} = \frac{1}{6}.$$

On en déduit :  $P(X \neq Y) = 5/6$ .

2. En fait c'est une loi hypergéométrique (tirages sans remise), mais en raison du théorème 5.26, on peut la remplacer en pratique par une binomiale.

### 5.2.3 Fonction de répartition

**Définition 5.9** (f.d.r. d'une variable aléatoire). Soient  $(\Omega, \mathcal{F}, P)$  un espace probabilisé et  $X$  une variable aléatoire sur  $(\Omega, \mathcal{F})$ . On appelle fonction de répartition (f.d.r.) de  $X$ , la fonction  $F_X$  définie sur  $\mathbb{R}$  par :

$$\forall x \in \mathbb{R}, \quad F_X(x) = P_X([\!-\infty, x]) = P(X \leq x).$$

La fonction  $F_X$  ne dépend que de la loi<sup>3</sup> de  $X$ . Deux variables aléatoires de même loi ont même fonction de répartition. La proposition suivante donne les propriétés générales des fonctions de répartition des variables aléatoires.

**Proposition 5.10.** La fonction de répartition  $F_X$  d'une variable aléatoire  $X$  est croissante sur  $\mathbb{R}$ , avec limite 0 en  $-\infty$  et 1 en  $+\infty$ . Elle est continue à droite et limitée à gauche en tout point de  $\mathbb{R}$  et vérifie :

$$\forall x \in \mathbb{R}, \quad P(X = x) = F(x) - F(x-). \quad (5.6)$$

La fonction de répartition d'une variable aléatoire caractérise sa loi, autrement dit :  $F_X = F_Y$  si et seulement si les variables aléatoires  $X$  et  $Y$  ont même loi.

Rappelons ici que la notation  $F(x-)$  désigne la limite à gauche de  $F$  au point  $x$  :

$$F(x-) = \lim_{\substack{t \rightarrow x \\ t < x}} F(t).$$

Nous admettrons cette proposition qui peut se démontrer en utilisant les propriétés générales d'une probabilité (prop. 4.4), sauf la partie «  $F_X = F_Y \Rightarrow X$  et  $Y$  ont même loi » dont la difficulté dépasse le niveau de ce cours.

En combinant la définition de  $F_X$  et la propriété (5.6), on obtient les formules suivantes de calcul à l'aide de  $F_X$  des  $P(X \in I)$  pour  $I$  intervalle de  $\mathbb{R}$  :

$$P(a < X \leq b) = F_X(b) - F_X(a), \quad (5.7)$$

$$P(a \leq X \leq b) = F_X(b) - F_X(a-), \quad (5.8)$$

$$P(a \leq X < b) = F_X(b-) - F_X(a-), \quad (5.9)$$

$$P(a < X < b) = F_X(b-) - F_X(a), \quad (5.10)$$

$$P(X \leq a) = F_X(a), \quad (5.11)$$

$$P(X < a) = F_X(a-), \quad (5.12)$$

$$P(X > b) = 1 - F_X(b), \quad (5.13)$$

$$P(X \geq b) = 1 - F_X(b-). \quad (5.14)$$

Dans le cas particulier des variables aléatoires discrètes, on peut donner une formule explicite de calcul de la fonction de répartition.

**Proposition 5.11** (f.d.r. d'une variable aléatoire discrète). Soient  $(\Omega, \mathcal{F}, P)$  un espace probabilisé et  $X$  une variable aléatoire discrète sur  $(\Omega, \mathcal{F})$ . Fixons une numérotation de l'ensemble au plus dénombrable  $X(\Omega)$  par les entiers :  $X(\Omega) = \{x_0, x_1, \dots, x_k, \dots\}$  et notons  $p_k := P(X = x_k)$ . La fonction de répartition  $F_X$  vérifie alors :

$$\forall x \in \mathbb{R}, \quad F_X(x) = \sum_{x_k \in X(\Omega)} p_k \mathbf{1}_{[x_k, +\infty[}(x), \quad (5.15)$$

3. Il serait plus correct, mais plus long, de parler de f.d.r. de la loi de  $X$  ou même de f.d.r. de la loi de  $X$  sous  $P$ .

ce qui s'écrit aussi

$$\forall x \in \mathbb{R}, \quad F_X(x) = \sum_{\substack{x_k \in X(\Omega) \\ x_k \leq x}} P(X = x_k). \quad (5.16)$$

De plus, si on peut numérotter les éléments de  $X(\Omega)$  de manière croissante ( $x_k < x_{k+1}$  pour tout  $k$ ), la fonction  $F_X$  est constante sur chaque intervalle  $[x_n, x_{n+1}[$  et vaut sur cet intervalle  $F_X(x) = \sum_{k \leq n} p_k$ .

*Preuve.* Soit  $x$  quelconque dans  $\mathbb{R}$ . En notant que l'évènement  $\{X \in ]-\infty, x]\}$  peut s'écrire ici comme l'union au plus dénombrable des  $\{X = x_k\}$  pour  $x_k \leq x$ , on obtient

$$F_X(x) = P_X(] - \infty, x]) = \sum_{x_k \in ] - \infty, x]} p_k = \sum_{x_k \in X(\Omega)} p_k \mathbf{1}_{[x_k, +\infty[}(x), \quad (5.17)$$

en notant que

$$\mathbf{1}_{[x_k, +\infty[}(x) = \begin{cases} 1 & \text{si } x_k \leq x \\ 0 & \text{si } x_k > x \end{cases} = \begin{cases} 1 & \text{si } x_k \in ] - \infty, x] \\ 0 & \text{si } x_k \notin ] - \infty, x] \end{cases}.$$

Si la suite  $(x_k)$  est strictement croissante, le réel  $x$  appartient à un seul intervalle  $[x_n, x_{n+1}[$ . Pour  $k \leq n$ , on a alors  $x_k \leq x_n \leq x$  et  $\mathbf{1}_{[x_k, +\infty[}(x) = 1$ , tandis que si  $k > n$ ,  $x_k \geq x_{n+1} > x$ , donc  $\mathbf{1}_{[x_k, +\infty[}(x) = 0$ . On a donc  $F_X(x) = \sum_{k \leq n} p_k$  et ceci étant valable pour tout  $x \in [x_n, x_{n+1}[$ , la fonction  $F_X$  est constante sur cet intervalle.  $\square$

À titre d'exemple, la figure 5.2 donne la représentation graphique de  $F_S$  où  $S$  est la variable aléatoire somme des points de deux dés.

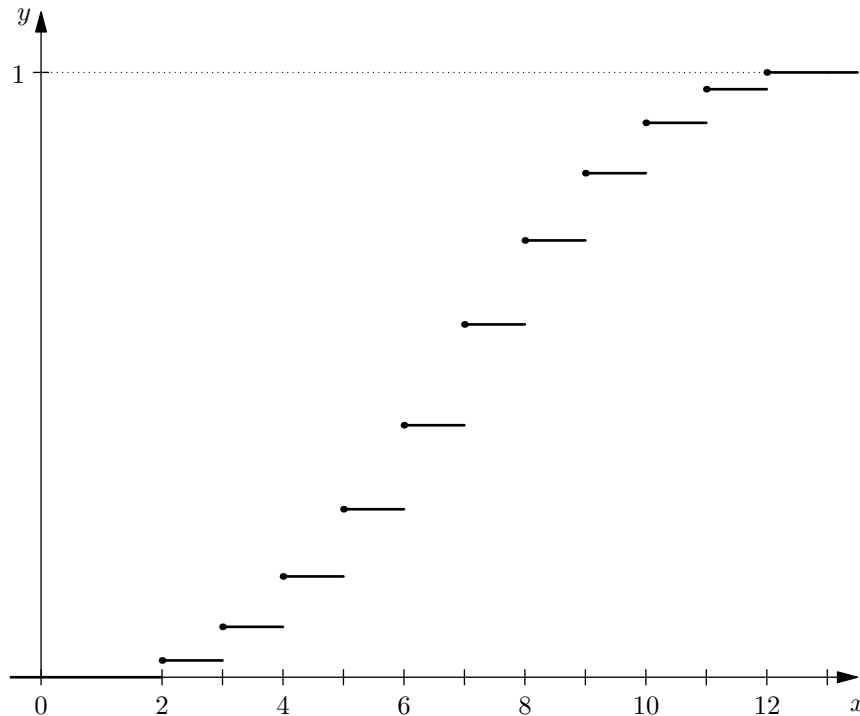


FIGURE 5.2 – f.d.r. de  $S$  somme des points de deux dés

Si l'on veut esquisser une classification sommaire des lois des variables aléatoires, on peut commencer par les partager entre les lois à f.d.r. continue sur  $\mathbb{R}$  et les lois à f.d.r. non



continue<sup>4</sup> sur  $\mathbb{R}$ . On parle plus simplement de *lois continues* ou encore *lois diffuses* dans le premier cas et de lois non continues ou non diffuses dans le deuxième. Dans la famille des lois non continues, nous connaissons déjà la sous-famille des lois discrètes. Dans la famille des lois continues, une importante sous-famille est celle des *lois à densité* que nous allons examiner maintenant.

### 5.2.4 Lois à densité

La loi d'une variable aléatoire  $X$  est à densité  $f$  si pour tout intervalle de  $\mathbb{R}$ , la probabilité d'appartenance de  $X$  à cet intervalle peut s'écrire comme l'intégrale de  $f$  sur cet intervalle.

**Définition 5.12** (densité de probabilité). *On appelle densité de probabilité sur  $\mathbb{R}$  toute fonction  $f$  vérifiant*

- a)  $f$  est définie et positive sur  $\mathbb{R} \setminus K$ , où  $K$  est une partie finie (éventuellement vide) de  $\mathbb{R}$  ;
- b)  $f$  est Riemann intégrable sur tout intervalle  $[a, b] \subset \mathbb{R} \setminus K$  ;
- c) l'intégrale généralisée de  $f$  sur  $] -\infty, +\infty[$  converge et

$$\int_{-\infty}^{+\infty} f(t) dt = 1.$$

Si  $f$  est une fonction positive définie seulement sur un intervalle  $]a, b[$  de  $\mathbb{R}$  et telle que  $\int_a^b f(t) dt = 1$ , on peut en faire une densité en la prolongeant à tout  $\mathbb{R}$  en posant  $f(t) := 0$  pour  $t \notin ]a, b[$ . Voici quatre exemples simples de densités :

$$f_1(t) := \frac{1}{b-a} \mathbf{1}_{[a,b]}(t); \quad f_2(t) := \frac{1}{2\sqrt{t}} \mathbf{1}_{]0,1]}(t);$$

$$f_3(t) := e^{-t} \mathbf{1}_{[0,+\infty[}(t); \quad f_4(t) := \frac{1}{\pi(1+t^2)}.$$

**Remarque 5.13** (usage des indicatrices dans les formules explicites). La définition de  $f_2$  repose sur un *abus d'écriture* d'usage courant. En effet il y a en toute rigueur un problème pour calculer  $f_2(t)$  lorsque  $t \leq 0$ , puisqu'alors il nous faut former le produit de l'expression  $\frac{1}{2\sqrt{t}}$  non définie (du moins en tant que nombre réel) par 0. La convention adoptée est que si la formule de calcul d'une fonction contient le produit d'une indicatrice par une expression non définie lorsque cette indicatrice est nulle, le produit vaut 0 dans ce cas. Ceci permet de considérer que la « définition » de  $f_2$  comme ci-dessus est un raccourci d'écriture commode pour :

$$f_2(t) := \begin{cases} \frac{1}{2\sqrt{t}} & \text{si } t \in ]0, 1], \\ 0 & \text{si } t \notin ]0, 1]. \end{cases}$$

**Définition 5.14.** *Soient  $(\Omega, \mathcal{F}, P)$  un espace probabilisé et  $X$  une variable aléatoire réelle sur  $(\Omega, \mathcal{F})$ . La loi de  $X$  sous  $P$  a pour densité  $f$  si :*

$$\forall a \in \mathbb{R}, \forall b \geq a, \quad P(X \in ]a, b]) = \int_a^b f(t) dt. \quad (5.18)$$

*On dit aussi par abus de langage que  $X$  a pour densité  $f$  (lorsqu'il n'y a pas ambiguïté sur  $P$ ).*

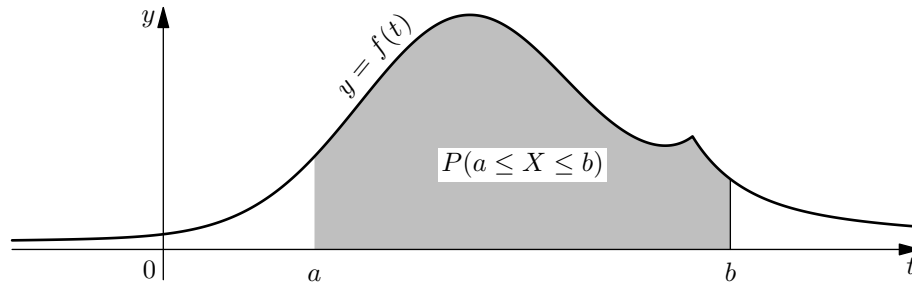


FIGURE 5.3 –  $P(a < X \leq b) = \int_a^b f(t) dt$  pour  $X$  de densité  $f$

**Remarque 5.15.** Il est clair d'après cette définition que si  $Y$  est une autre variable aléatoire ayant même loi que  $X$  (donc mêmes probabilités d'appartenance aux intervalles), elle a aussi la densité  $f$ . D'autre part, il n'y a pas unicité de la densité d'une variable aléatoire. Par exemple  $g_1 = \mathbf{1}_{[0,1]}$  et  $g_2 = \mathbf{1}_{]0,1]}$  sont deux densités de probabilité qui donnent les mêmes intégrales :  $\int_a^b g_1(t) dt = \int_a^b g_2(t) dt$  pour toute paire de réels  $a$  et  $b$ . Ces deux fonctions peuvent chacune être prise comme densité de la loi uniforme sur  $[0, 1]$ .

L'exemple ci-dessus montre qu'il ne suffit pas de vérifier que deux variables aléatoires ont des densités qui diffèrent en un point pour en déduire qu'elles n'ont pas même loi. Nous admettrons le lemme suivant qui donne une condition suffisante pratique pour que deux variables à densité n'aient pas même loi.

**Lemme 5.16.** Soient  $X$  et  $Y$  deux variables aléatoires admettant respectivement pour densité les fonctions  $f$  et  $g$ . On suppose qu'il existe un réel  $t_0$  tel que  $f(t_0) \neq g(t_0)$  et que de plus,  $f$  et  $g$  sont toutes deux continues au point  $t_0$ . Alors  $X$  et  $Y$  n'ont pas même loi.

Examinons maintenant les relations entre la densité d'une variable aléatoire, lorsqu'elle existe, et sa fonction de répartition, qui elle, existe toujours, .

**Proposition 5.17.** Si la variable aléatoire  $X$  a pour densité  $f$ , sa fonction de répartition  $F$  vérifie :

- a)  $\forall x \in \mathbb{R}, F(x) = \int_{-\infty}^x f(t) dt$  ;
- b)  $F$  est continue sur  $\mathbb{R}$  ;
- c) si  $f$  est continue au point  $x_0$ , alors  $F$  est dérivable en ce point et  $F'(x_0) = f(x_0)$ .

**Corollaire 5.18.** Si la variable aléatoire  $X$  a pour densité  $f$ , les égalités suivantes sont vérifiées pour tous  $a, b \in \mathbb{R}$  tels que  $a < b$  :

$$P(a < X < b) = P(a < X \leq b) = P(a \leq X < b) = P(a \leq X \leq b) = \int_a^b f(t) dt,$$

$$P(X < a) = P(X \leq a) = \int_{-\infty}^a f(t) dt,$$

$$P(X > b) = P(X \geq b) = \int_b^{+\infty} f(t) dt.$$

*Preuve de la proposition 5.17 et du corollaire 5.18.*

---

4. On peut démontrer que l'ensemble des points de discontinuité d'une f.d.r. quelconque est au plus dénombrable.

*Preuve de a).* Puisque  $X$  a pour densité  $f$ , on a pour tous réels  $a < b$ ,

$$P(X \in ]a, b]) = F(b) - F(a) = \int_a^b f(t) dt. \quad (5.19)$$

Il suffit d'appliquer (5.19) avec  $b = x$  fixé et  $a = -n$  pour chaque  $n \in \mathbb{N}$  tel que  $-n < x$ . La suite d'évènements

$$A_n := \{X \in ]-n, x]\}, \quad n > -x,$$

est croissante pour l'inclusion et a pour réunion  $A = \{X \in ]-\infty, x]\}$ . Par continuité monotone séquentielle (cf. proposition 4.4), on a  $P(A_n) \uparrow P(A)$ , d'où

$$F(x) = P(A) = \lim_{n \rightarrow +\infty} P(X \in ]-n, x]) = \lim_{n \rightarrow +\infty} \int_{-n}^x f(t) dt = \int_{-\infty}^x f(t) dt,$$

en notant que l'intégrale généralisée de la densité  $f$  converge en  $-\infty$ .  $\square$

*Preuve de b).* Fixons  $x_0 \in \mathbb{R}$  quelconque. On sait déjà que  $F$  est continue à droite en tout point comme toute fonction de répartition. Il suffit donc de montrer la continuité à gauche en  $x_0$ . D'après le point b) de la définition 5.12, il existe  $a < x_0$  tel que  $f$  soit définie et Riemann intégrable sur tout intervalle  $[a, a'] \subset [a, x_0[$ . On a alors

$$\lim_{x \uparrow x_0} \int_a^x f(t) dt = \int_a^{x_0} f(t) dt,$$

où la deuxième intégrale est soit une intégrale de Riemann ordinaire soit une intégrale généralisée convergente. Cette relation peut aussi s'écrire à l'aide de  $F$  :

$$\lim_{x \uparrow x_0} (F(x) - F(a)) = F(x_0) - F(a).$$

On en déduit par addition de  $F(a)$  que  $F(x)$  tend vers  $F(x_0)$  quand  $x$  tend vers  $x_0$  par valeurs inférieures.  $\square$

*Preuve de c).* Puisque  $f$  est continue en  $x_0$ , elle est définie sur tout un voisinage de  $x_0$  et donc sur tout un intervalle  $]a, b[$  contenant  $x_0$ . La continuité de  $f$  en  $x_0$  peut alors s'écrire :

$$\forall \varepsilon > 0, \exists ]x_0 - \delta, x_0 + \delta[ \subset ]a, b[; \quad \forall t \in ]x_0 - \delta, x_0 + \delta[, \quad |f(t) - f(x_0)| < \varepsilon. \quad (5.20)$$

Pour tout  $h$  tel que  $0 < |h| < \delta$ , on a alors  $F(x_0 + h) - F(x_0) = \int_{x_0}^{x_0+h} f(t) dt$  d'où

$$|F(x_0 + h) - F(x_0) - hf(x_0)| = \left| \int_{x_0}^{x_0+h} (f(t) - f(x_0)) dt \right| \leq h\varepsilon.$$

En divisant par  $h$ , puis en faisant tendre  $h$  vers 0, on voit que  $F$  a bien une dérivée en  $x_0$  et que celle-ci vaut  $f(x_0)$ .  $\square$

La proposition 5.17 est maintenant complètement démontrée. Pour vérifier le corollaire 5.18, il suffit de combiner les relations générales (5.7)–(5.14) avec (5.18) et les points a) et b) de la proposition 5.17.  $\square$

**Remarques 5.19.**

1. D'après b) toute variable aléatoire à densité a une fonction de répartition continue. La réciproque *est fausse* : il existe des lois à fonction de répartition continue sans densité.
2. Par ailleurs si  $X$  a une densité, sa fonction de répartition n'est pas forcément dérivable en tout point. Par exemple la densité  $f_2$  ci-dessus a pour fonction de répartition associée  $F_2(x) = \sqrt{x}\mathbf{1}_{]0,1]}(x) + \mathbf{1}_{]1,+\infty[}(x)$  (cette écriture condensée signifie que  $F_2(x)$  est nul sur  $\mathbb{R}^-$ , vaut  $\sqrt{x}$  entre 0 et 1 et reste constant égal à 1 sur  $]1, +\infty[$ ).  $F_2$  est dérivable en tout point sauf en 0 et en 1.

La proposition suivante donne une règle pratique permettant de trouver la densité (lorsqu'elle existe!) à partir de la fonction de répartition dans les cas les plus courants.

**Proposition 5.20.** *On suppose que la fonction de répartition  $F$  de  $X$  est  $C^1$  par morceaux au sens suivant :  $F$  est continue sur  $\mathbb{R}$  et dérivable sur  $\mathbb{R}$  privé (éventuellement) d'un ensemble fini de points  $a_1 < \dots < a_n$ . Sur chacun des intervalles ouverts  $] -\infty, a_1[$ ,  $]a_i, a_{i+1}[$  ( $1 \leq i < n$ ),  $]a_n, +\infty[$ , la dérivée  $f$  de  $F$  est continue. Alors  $X$  a pour densité  $f$ .*

*Preuve.* Il est commode de poser  $a_0 := -\infty$  et  $a_{n+1} = +\infty$ . Sur chacun des intervalles ouverts  $I$  découpés par les  $a_i$ ,  $F$  est dérivable et sa dérivée  $f$  est continue. On sait alors que  $f$  a une infinité de primitives sur  $I$  et que si l'on fixe un  $\alpha$  dans  $I$ , toute primitive  $H$  de  $f$  sur  $I$  est de la forme  $H(x) = \int_{\alpha}^x f(t) dt + C$ , avec  $C$  constante. Comme  $F$  est l'une des primitives de  $f$  sur  $I$ , en prenant  $H = F$  et en faisant  $x = \alpha$ , on voit que la constante  $C$  vaut  $F(\alpha)$ . On a donc pour  $\alpha$  et  $x$  quelconques dans  $I$ ,  $F(x) - F(\alpha) = \int_{\alpha}^x f(t) dt$ . Fixons  $\alpha$  et prenons  $x \geq \alpha$ . Faisons tendre  $x$  vers la borne supérieure  $a_i$  de  $I$ . Comme  $F$  est continue (ou dans le cas  $a_i = +\infty$ ,  $F$  a une limite 1), l'intégrale généralisée  $\int_{\alpha}^{a_i} f(t) dt$  converge et vaut  $F(a_i) - F(\alpha)$  (ou  $1 - F(\alpha)$  quand  $a_i = +\infty$ ). De même en faisant tendre  $\alpha$  vers  $a_{i-1}$  on voit que l'intégrale généralisée  $\int_{a_{i-1}}^{a_i} f(t) dt$  converge et vaut  $F(a_i) - F(a_{i-1})$  (ou  $F(a_i)$  quand  $a_{i-1} = -\infty$ ). Finalement soient  $a$  et  $b > a$  quelconques dans  $\mathbb{R}$ . Si  $a$  et  $b$  sont dans le même intervalle  $I$  on a directement  $F(b) - F(a) = \int_a^b f(t) dt$ . Sinon on note  $(a_i)_{i_0 \leq i \leq i_1}$  l'ensemble de tous les  $a_i$  qui sont dans  $[a, b]$  et on écrit

$$F(b) - F(a) = F(a_{i_0}) - F(a) + \sum_{i=i_0}^{i_1-1} (F(a_{i+1}) - F(a_i)) + F(b) - F(a_{i_1}) = \int_a^b f(t) dt,$$

en utilisant la relation de Chasles pour les intégrales généralisées. On a donc toujours  $P(X \in ]a, b]) = F(b) - F(a) = \int_a^b f(t) dt$ , ce qui montre que  $X$  a pour densité  $f$ .  $\square$

### 5.3 Lois discrètes classiques

Dans toute la suite du chapitre, on fixe un espace probabilisé  $(\Omega, \mathcal{F}, P)$ , on désigne par  $X$  une variable aléatoire sur  $(\Omega, \mathcal{F})$  et par  $P_X$  sa loi sous  $P$ . Cette clause sera implicite chaque fois que nous écrirons dans les définitions « La variable aléatoire  $X$  suit la loi ... si ... ». Pour  $X$  de loi discrète, nous utiliserons la notation :

$$X_P(\Omega) := \{x \in \mathbb{R}; P(X = x) > 0\}. \quad (5.21)$$

Bien sûr,  $X_P(\Omega)$  est toujours inclus dans l'ensemble des valeurs « possibles »  $X(\Omega)$  et dans la plupart des situations pratiques d'utilisation des variables aléatoires, ces deux ensembles sont égaux<sup>5</sup>.

5. Ce distinguo entre  $X(\Omega)$  et  $X_P(\Omega)$ , à ignorer en première lecture, vient du fait que la loi d'une variable aléatoire réelle  $X$ , sous une certaine probabilité  $P$ , peut être discrète sans que  $X$  elle-même soit discrète au sens de la définition 5.3.

### 5.3.1 Lois de Bernoulli

**Définition 5.21.** La variable aléatoire  $X$  suit la loi de Bernoulli de paramètre  $p$  ( $p \in [0, 1]$ ) si  $X_P(\Omega) = \{0, 1\}$  avec :

$$P(X = 1) = p, \quad P(X = 0) = 1 - p = q.$$

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

Si  $A$  est un évènement de probabilité  $p$ , son indicatrice définie par :

$$\mathbf{1}_A(\omega) = \begin{cases} 1 & \text{si } \omega \in A \\ 0 & \text{si } \omega \notin A \end{cases} = \begin{cases} 1 & \text{si } A \text{ est réalisé} \\ 0 & \text{si } A \text{ n'est pas réalisé} \end{cases}$$

est une variable aléatoire suivant la loi de Bernoulli de paramètre  $p$ . Réciproquement, si  $X$  est une v.a. de Bernoulli, on peut toujours écrire que  $X = \mathbf{1}_A$  presque-sûrement, c'est-à-dire  $P(X = \mathbf{1}_A) = 1$ , en définissant  $A = \{\omega \in \Omega, X(\omega) = 1\}$ .

### 5.3.2 Loi uniforme sur un ensemble fini de réels

**Définition 5.22.** La variable aléatoire  $X$  suit la loi uniforme sur  $\{x_1, \dots, x_n\} \subset \mathbb{R}$  si  $P_X$  est l'équiprobabilité sur cet ensemble. Notation :  $X \sim \text{Unif}\{x_1, \dots, x_n\}$ .

Autrement dit,  $X_P(\Omega) = \{x_1, \dots, x_n\}$  et

$$\forall k \in \llbracket 1, n \rrbracket, \quad P(X = x_k) = \frac{1}{n}.$$

D'où

$$\forall B \in \mathcal{B}, \quad P_X(B) = \sum_{x_k \in B} \frac{1}{n} = \frac{\text{card}(B \cap \{x_1, \dots, x_n\})}{n}.$$

Par exemple, le nombre de points indiqué par un dé équilibré suit la loi uniforme sur  $\llbracket 1, 6 \rrbracket$ .

### 5.3.3 Lois binomiales

**Définition 5.23.** La variable aléatoire  $X$  suit la loi binomiale de paramètres  $n$  et  $p$ ,  $n \in \mathbb{N}^*$  et  $p \in [0, 1]$ , notation  $X \sim \text{Bin}(n, p)$ , si  $X_P(\Omega) = \llbracket 0, n \rrbracket$  et

$$\forall k \in \llbracket 0, n \rrbracket, \quad P(X = k) = C_n^k p^k (1 - p)^{n-k}.$$

La formule ci-dessus définit bien une loi de probabilité puisque les  $C_n^k p^k (1 - p)^{n-k}$  sont positifs et :

$$\sum_{k=0}^n C_n^k p^k (1 - p)^{n-k} = (p + (1 - p))^n = 1^n = 1,$$

en appliquant la formule du binôme de Newton (d'où le nom de la loi). La loi binomiale  $\text{Bin}(n, p)$  est la loi du nombre de succès obtenus en une suite de  $n$  épreuves répétées indépendantes avec pour chaque épreuve une probabilité de succès  $p$ . Ceci a été démontré dans l'exemple 4.41.

De même, soit  $A_1, \dots, A_n$  une famille d'évènements mutuellement indépendants tous de même probabilité  $p$  et notons  $X_i$  la variable de Bernoulli indicatrice de  $A_i$  :

$$X_i(\omega) = \begin{cases} 1 & \text{si } \omega \in A_i, \\ 0 & \text{si } \omega \in A_i^c. \end{cases}$$

Alors la variable aléatoire  $S_n = \sum_{i=1}^n X_i$  suit la loi binomiale  $\text{Bin}(n, p)$ .

### 5.3.4 Lois hypergéométriques

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

**Exemple 5.24.** Dans une production totale de  $N$  objets dont  $M$  sont défectueux, on prélève au hasard un échantillon de  $n$  objets (tirage sans remise). Soit  $X$  le nombre aléatoire d'objets défectueux dans l'échantillon. Quelle est sa loi ?

On peut prendre comme espace  $\Omega$  l'ensemble de tous les échantillons possibles (toutes les parties à  $n$  éléments d'un ensemble de cardinal  $N$ ) muni de l'équiprobabilité. Chaque échantillon a ainsi une probabilité  $1/C_N^n$  d'être choisi. Les échantillons (événements élémentaires) réalisant l'évènement  $\{X = k\}$  sont ceux qui contiennent  $k$  objets défectueux et  $n - k$  objets non défectueux. Ceci n'est réalisable que si  $0 \leq k \leq M$  et  $0 \leq n - k \leq N - M$ . Dénombrons ces échantillons. On les forme en choisissant  $k$  objets défectueux dans une sous-population de taille  $M$  et en complétant par  $n - k$  objets non défectueux choisis dans une sous-population de taille  $N - M$ . Il y en a donc  $C_M^k \times C_{N-M}^{n-k}$ . Finalement :

$$P(X = k) = \frac{C_M^k \times C_{N-M}^{n-k}}{C_N^n} \quad \text{si} \quad \begin{cases} 0 \leq k \leq M, \\ 0 \leq n - k \leq N - M. \end{cases} \quad (5.22)$$

**Définition 5.25.** La loi définie par (5.22) s'appelle loi hypergéométrique de paramètres  $N$ ,  $M$  et  $n$ . Notation :  $X \sim \text{Hypg}(N, M, n)$ . Le paramètre  $N$  est l'effectif de la population totale,  $M$  celui de la sous-population à laquelle on s'intéresse et  $n$  la taille de l'échantillon observé.

Pour une taille d'échantillon  $n$  fixée, plus  $N$  et  $M$  sont grands, moins les tirages sans remise diffèrent des tirages avec remise. Plus précisément, la loi hypergéométrique converge vers la loi binomiale au sens suivant.

**Théorème 5.26** (convergence de l'hypergéométrique vers la binomiale). *On suppose que quand  $N$  tend vers  $+\infty$ ,  $M = M(N)$  tend vers  $+\infty$  en vérifiant la condition :*

$$\lim_{N \rightarrow +\infty} \frac{M}{N} = p \quad \text{avec} \quad 0 < p < 1. \quad (5.23)$$

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

$$\forall k \in \llbracket 0, n \rrbracket, \quad \lim_{N \rightarrow +\infty} P(X_N = k) = P(Y = k), \quad (5.24)$$

autrement dit :

$$\forall k \in \llbracket 0, n \rrbracket, \quad \lim_{N \rightarrow +\infty} \frac{C_M^k \times C_{N-M}^{n-k}}{C_N^n} = C_n^k p^k (1-p)^{n-k}. \quad (5.25)$$

*Preuve.* Notons d'abord que comme  $p$  est strictement positif, l'hypothèse (5.23) implique que  $M$  tend vers  $+\infty$  avec  $N$  ; il en va de même pour  $N - M$  puisque  $p < 1$ . Ensuite,

$$\begin{aligned} \frac{C_M^k \times C_{N-M}^{n-k}}{C_N^n} &= \frac{M!}{k!(M-k)!} \times \frac{(N-M)!}{(n-k)!((N-M)-(n-k))!} \times \frac{n!(N-n)!}{N!} \\ &= C_n^k \frac{M!}{(M-k)!} \times \frac{(N-M)!}{((N-M)-(n-k))!} \times \frac{(N-n)!}{N!}. \end{aligned} \quad (5.26)$$

Comme  $k$  est fixé et  $M$  tend vers  $+\infty$ , la première fraction dans (5.26) est le produit de  $k$  facteurs  $M, (M-1), \dots, (M-k+1)$  tous équivalents<sup>6</sup> à  $M$  d'où :

$$\frac{M!}{(M-k)!} \sim M^k, \quad N \rightarrow +\infty. \quad (5.27)$$

Par le même argument avec  $n-k$  et  $N-M$  au lieu de  $k$  et  $M$  :

$$\frac{(N-M)!}{((N-M)-(n-k))!} \sim (N-M)^{n-k}, \quad N \rightarrow +\infty. \quad (5.28)$$

Enfin :

$$\frac{(N-n)!}{N!} \sim \frac{1}{N^n}, \quad N \rightarrow +\infty. \quad (5.29)$$

En reportant ces équivalents dans (5.26), on voit que lorsque  $N$  tend vers  $+\infty$  :

$$P(X_N = k) \sim C_n^k \frac{M^k (N-M)^{n-k}}{N^n} = C_n^k \left(\frac{M}{N}\right)^k \left(\frac{N-M}{N}\right)^{n-k}, \quad (5.30)$$

d'où :  $\lim_{N \rightarrow +\infty} P(X_N = k) = C_n^k p^k (1-p)^{n-k}$ . □

### 5.3.5 Lois géométriques

**Exemple 5.27** (un problème de temps d'attente).

Considérons une suite infinie d'épreuves répétées indépendantes avec même probabilité de succès  $p \in ]0, 1[$ . Soit  $X$  le numéro (aléatoire) de la première épreuve où l'on obtient un succès. Si l'on n'obtient jamais de succès, on conviendra que  $X = +\infty$ . Calculer  $P(X = k)$  pour tout  $k \in \mathbb{N}^*$ . En déduire les valeurs de  $P(X \in \mathbb{N}^*)$  et  $P(X = +\infty)$ .

En notant  $R_i = \{\text{succès à la } i\text{-ème épreuve}\}$ , on a :

$$\begin{aligned} \{X = k\} &= \{\text{échec aux } (k-1) \text{ premières et succès à la } k\text{-ième}\} \\ &= \left( \bigcap_{i=1}^{k-1} R_i^c \right) \cap R_k. \end{aligned}$$

D'où par indépendance des épreuves :

$$P(X = k) = \left( \prod_{i=1}^{k-1} P(R_i^c) \right) \times P(R_k) = (1-p)^{k-1} p.$$

Posons  $q = 1-p$  et notons que  $q \in ]0, 1[$ . La décomposition de l'évènement  $\{X \in \mathbb{N}^*\}$  en la réunion disjointe des  $\{X = k\}$  ( $k \in \mathbb{N}^*$ ) nous donne par  $\sigma$ -additivité :

$$\begin{aligned} P(X \in \mathbb{N}^*) &= \sum_{k \in \mathbb{N}^*} P(X = k) = \sum_{k \in \mathbb{N}^*} q^{k-1} p \\ &= p \sum_{l \in \mathbb{N}} q^l \quad (l = k-1) \\ &= p \frac{1}{1-q} = 1. \end{aligned}$$

6. Rappelons que deux suites  $(u_N)$  et  $(v_N)$  sont dites équivalentes lorsque  $u_N = v_N(1 + \varepsilon_N)$  avec  $\varepsilon_N$  tendant vers 0 quand  $N$  tend vers  $+\infty$  (notation :  $u_N \sim v_N$ ).

Ainsi avec probabilité 1, le premier succès a lieu au bout d'un nombre *fini* d'épreuves<sup>7</sup>. Remarquons qu'on aurait pu arriver au même résultat en montrant que  $P(X = +\infty) = 0$  par la méthode utilisée à l'exemple 4.41 c) en échangeant les rôles de succès et échec. En toute rigueur,  $X$  n'est pas une variable aléatoire discrète au sens de la définition 5.3 puisque  $X(\Omega)$  est une partie dénombrable de  $\overline{\mathbb{R}}$  au lieu de  $\mathbb{R}$ . Néanmoins  $X' := X\mathbf{1}_{\{X < +\infty\}}$  est une variable aléatoire discrète et ce qui précède montre que  $X'$  a même loi<sup>8</sup> que  $X$ . Cette loi est celle du temps d'attente du premier succès dans une suite d'épreuves répétées indépendantes, on l'appelle *loi géométrique de paramètre  $p$* .

**Définition 5.28.** Une variable aléatoire  $X$  suit la loi géométrique de paramètre  $p \in ]0, 1[$ , si  $X_P(\Omega) = \mathbb{N}^*$  et :

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

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

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

*Première méthode.* On calcule le reste d'une série géométrique :

$$\begin{aligned} P(X > n) &= \sum_{k=n+1}^{+\infty} q^{k-1}p = \sum_{l=n}^{+\infty} q^l p \\ &= pq^n \sum_{l=n}^{+\infty} q^{l-n} = pq^n \sum_{j=0}^{+\infty} q^j \\ &= \frac{pq^n}{1 - q} = q^n. \end{aligned}$$

*Deuxième méthode.* On se place dans la situation de l'exemple 5.27. L'évènement  $\{X > n\}$  se réalise si et seulement si les  $n$  premières épreuves donnent un échec, d'où  $\{X > n\} = \bigcap_{i=1}^n R_i^c$ . En utilisant l'indépendance des  $R_i$  on en déduit :

$$P(X > n) = \prod_{i=1}^n P(R_i^c) = q^n.$$

### 5.3.6 Lois de Poisson

**Définition 5.29.** On dit que la variable aléatoire discrète  $X$  suit la loi de Poisson de paramètre  $\alpha > 0$  si  $X_P(\Omega) = \mathbb{N}$  et

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

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

On sait que la fonction exponentielle a un développement en série entière avec rayon de convergence infini. En particulier :

$$\forall \alpha > 0, \quad e^\alpha = \sum_{k=0}^{+\infty} \frac{\alpha^k}{k!}.$$

7. Mais pas *borné* par un nombre fixé choisi avant le début des épreuves...

8. Pour être tout à fait rigoureux, il faudrait avoir défini les v.a. à valeurs dans  $\overline{\mathbb{R}}$  et leurs lois, ce qui nous aurait fait sortir du cadre de ce cours.



On a donc bien :

$$\sum_{k=0}^{+\infty} P(X = k) = e^{-\alpha} \sum_{k=0}^{+\infty} \frac{\alpha^k}{k!} = e^{-\alpha} e^{\alpha} = 1.$$

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

**Théorème 5.30.** *Si  $(p_n)_{n \geq 1}$  est une suite de réels de  $[0, 1]$  vérifiant*

$$np_n \xrightarrow[n \rightarrow +\infty]{} \alpha \in ]0, +\infty[, \quad (5.31)$$

alors pour tout  $k \in \mathbb{N}$ ,

$$C_n^k p_n^k (1 - p_n)^{n-k} \xrightarrow[n \rightarrow +\infty]{} \frac{e^{-\alpha} \alpha^k}{k!}.$$

*Preuve.* L'hypothèse (5.31) peut s'écrire sous la forme plus maniable :  $np_n = \alpha u_n$  avec  $u_n$  tendant vers 1 quand  $n$  tend vers  $+\infty$ . Ainsi  $p_n = \alpha u_n / n$  et

$$C_n^k p_n^k (1 - p_n)^{n-k} = \frac{n!}{(n-k)! k!} \left(\frac{\alpha}{n}\right)^k u_n^k \left(1 - \frac{\alpha u_n}{n}\right)^{n-k}. \quad (5.32)$$

Pour obtenir la limite de cette expression lorsque  $n$  tend vers  $+\infty$ ,  $k$  restant fixé, on remarque successivement que :

$$\lim_{n \rightarrow +\infty} \frac{n!}{(n-k)!} \frac{1}{n^k} = 1, \quad (5.33)$$

$$\lim_{n \rightarrow +\infty} u_n^k = 1, \quad (5.34)$$

$$\lim_{n \rightarrow +\infty} \left(1 - \frac{\alpha u_n}{n}\right)^{n-k} = e^{-\alpha}. \quad (5.35)$$

Pour justifier (5.35), on écrit :

$$\left(1 - \frac{\alpha u_n}{n}\right)^{n-k} = \exp\left[(n-k) \ln\left(1 - \frac{\alpha u_n}{n}\right)\right], \quad (5.36)$$

puis comme  $\alpha u_n / n$  tend vers 0 :

$$(n-k) \ln\left(1 - \frac{\alpha u_n}{n}\right) \sim n \left(-\frac{\alpha u_n}{n}\right) \sim -\alpha, \quad (n \rightarrow +\infty).$$

Par continuité de la fonction exponentielle, la limite du second membre de (5.36) est donc bien  $e^{-\alpha}$ , ce qui prouve (5.35). On obtient alors la conclusion du théorème en passant à la limite dans (5.32).  $\square$

Le théorème 5.30 sert de justification théorique à la règle pratique suivante : lorsque  $n$  est « grand » et  $np$  « petit », on peut remplacer la loi  $\text{Bin}(n, p)$  par la loi  $\text{Pois}(\alpha)$  où  $\alpha = np$ . En général on considère que  $n$  de l'ordre de quelques centaines et  $np$  de l'ordre de quelques unités donnent une bonne approximation. Sous cette forme, cette règle relève plus de la cuisine que des mathématiques. Il est possible par des techniques élémentaires de contrôler l'erreur commise en utilisant cette approximation. Nous nous contenterons ici d'un exemple classique et d'une comparaison graphique pour illustrer la qualité de cette approximation.

**Exemple 5.31.** Le président d'un bureau de vote est né un 1<sup>er</sup> avril. Il décide de noter le nombre  $X$  de personnes ayant leur anniversaire le même jour que lui parmi les 500 premiers électeurs qui se présentent.

La situation peut être assimilée à une suite d'épreuves répétées indépendantes et  $X$  est une variable aléatoire binomiale de paramètres  $n = 500$  et  $p = 1/365$  (en négligeant la question des années bissextiles sinon on prendrait  $p = 4/(3 \times 365 + 366)$ , ce qui ne changerait pas grand chose numériquement). Ainsi :

$$P(X = k) = C_{500}^k \left(\frac{1}{365}\right)^k \left(\frac{364}{365}\right)^{500-k}.$$

La règle énoncée ci-dessus nous conduit à approximer la loi de  $X$  par une loi de Poisson de paramètre :

$$\alpha = np = 500 \times \frac{1}{365}.$$

Voici une comparaison numérique pour les petites valeurs de  $k$  :

$k$	0	1	2	3	4	5
$P(X = k)$	0,253 7	0,348 4	0,238 8	0,108 9	0,037 2	0,010 1
$\frac{e^{-\alpha} \alpha^k}{k!}$	0,254 1	0,348 1	0,238 5	0,108 9	0,037 3	0,010 2

Remarquons que la probabilité d'observer plus de 5 anniversaires un 1<sup>er</sup> avril, calculée par la loi exacte de  $X$  ou par son approximation poissonnienne est inférieure à 0,003.

#### Comparaison graphique :

Les *diagrammes en bâtons* ci-dessous représentent la loi binomiale  $\text{Bin}(n, p)$  et la loi de Poisson approximante  $\text{Pois}(\alpha)$  avec  $\alpha = np$ . Les segments verticaux (les bâtons) du diagramme représentant la loi d'une variable discrète  $X$  (à valeurs dans  $\mathbb{N}$ ) ont une hauteur égale à  $P(X = k)$  avec une extrémité inférieure au point d'abscisse  $k$  de l'axe horizontal. Pour la lisibilité, on a légèrement décalé vers la gauche les bâtons de la loi de Poisson et vers la droite ceux de la loi binomiale. Bien que le diagramme en bâtons de la loi binomiale  $\text{Bin}(n, p)$  soit constitué théoriquement de  $n + 1$  bâtons (et que celui de la loi de Poisson en ait une infinité), seul un petit nombre de bâtons est visible sur les graphiques, les autres correspondant à des probabilités trop petites<sup>9</sup>. On constate que pour  $n = 200$  (figure 5.7), la différence entre les deux diagrammes n'est pratiquement plus discernable *visuellement*.

9. En fait, on s'est contenté d'afficher les probabilités correspondant à  $k$  inférieur ou égal à la partie entière supérieure de  $2\alpha + 4$ . On peut vérifier que la somme des probabilités ainsi négligées est inférieure à 1%, pour chacune des deux lois.

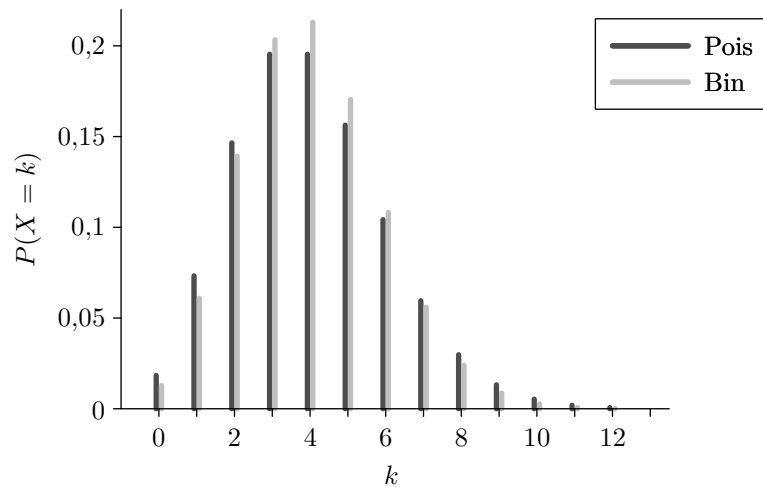


FIGURE 5.4 – Lois Bin(25; 0,16) et Pois(4)

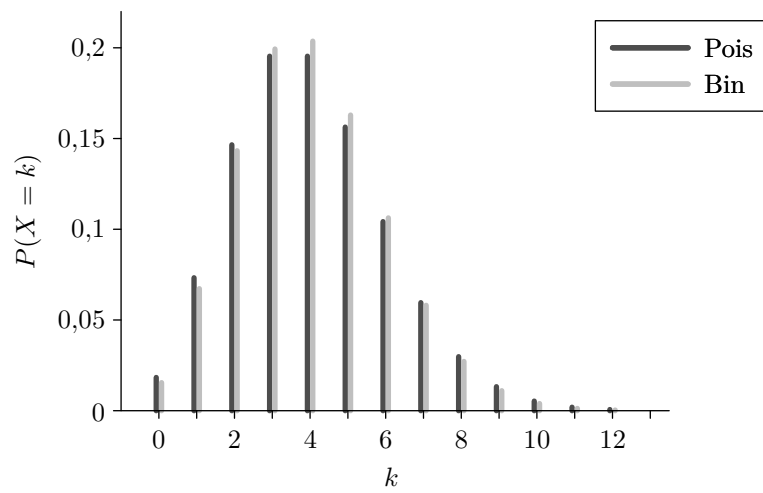


FIGURE 5.5 – Lois Bin(50; 0,08) et Pois(4)

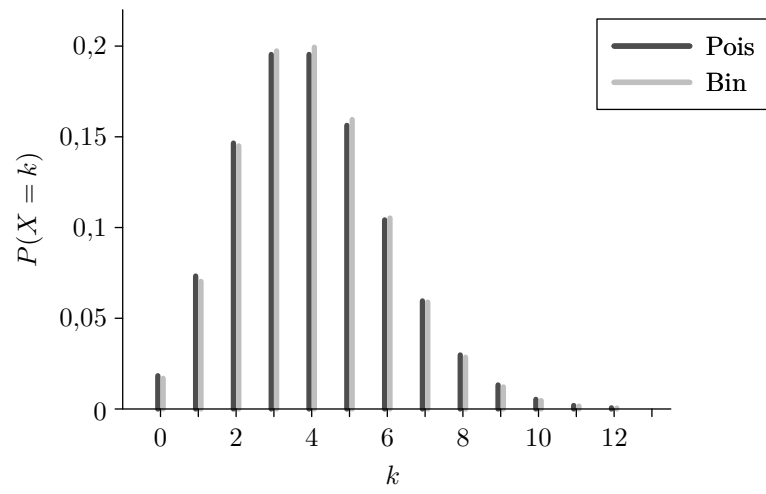


FIGURE 5.6 – Lois Bin(100; 0,04) et Pois(4)

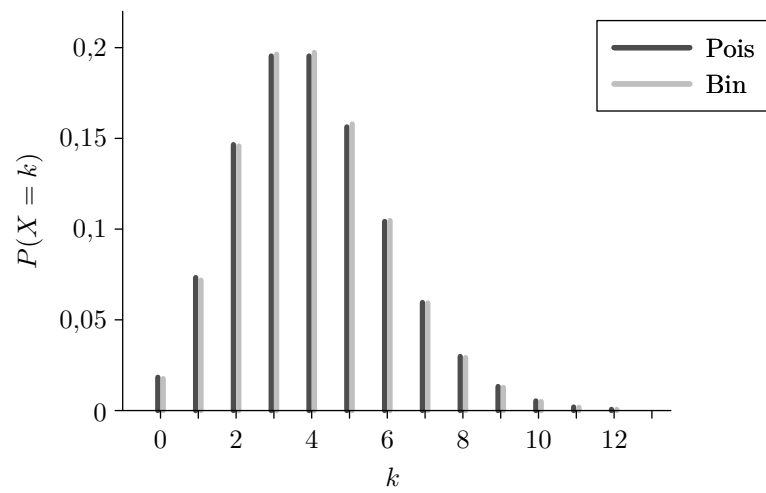


FIGURE 5.7 – Lois Bin(200; 0,02) et Pois(4)

Le théorème suivant, que nous admettrons, permet de saisir l'importance de la loi de Poisson, en justifiant au passage le bien fondé de l'hypothèse (5.31) du théorème de convergence de la loi binomiale vers la loi de Poisson.

**Théorème 5.32.** *Soit un phénomène donnant lieu à des observations aléatoires vérifiant les hypothèses suivantes.*

- (a) *Les observations dans des intervalles de temps disjoints sont indépendantes*
- (b) *Pour tout réel  $t$  tel que  $0 \leq t < t+T \leq 1$  la loi du nombre (aléatoire) d'observations dans l'intervalle  $[t, t+T[$  ne dépend que de la durée  $T$  de cet intervalle.*
- (c) *En notant  $p_n$  la probabilité d'avoir exactement une observation dans un intervalle de temps de durée  $1/n$  et  $r_n$  celle d'en avoir au moins une,  $\varepsilon_n = (r_n - p_n)/p_n$  tend vers 0 quand  $n$  tend vers l'infini.*

*Alors le nombre aléatoire d'observations dans l'intervalle  $[0, 1[$  suit la loi de Poisson de paramètre  $\alpha$  défini par  $\alpha = -\ln(1 - r_1)$ .*

**Remarque 5.33.** L'examen attentif de la démonstration du théorème ci-dessus montre que la structure d'ordre de l'intervalle  $[0, 1[$  n'y joue aucun rôle. L'important est la possibilité de réaliser une partition de  $[0, 1[$  en intervalles de même longueur tendant vers 0. Par conséquent en remplaçant la longueur par l'aire ou le volume, il est possible d'obtenir une version spatiale en dimension 2 ou 3 du théorème 5.32. Ceci permet de comprendre pourquoi la loi de Poisson fournit une bonne modélisation par exemple du nombre d'erreurs typographiques dans une page imprimée, du nombre d'impacts de météorites sur un territoire donné, du nombre d'accidents sur une portion d'autoroute pendant une période donnée, du nombre de raisins dans une portion de cake, du nombre d'étoiles dans une région de l'univers, ...

## 5.4 Lois à densité classiques

### 5.4.1 Lois uniformes

**Définition 5.34.** *La variable aléatoire réelle  $X$  suit la loi uniforme sur l'intervalle  $[a, b]$  ( $-\infty < a < b < +\infty$ ) si*

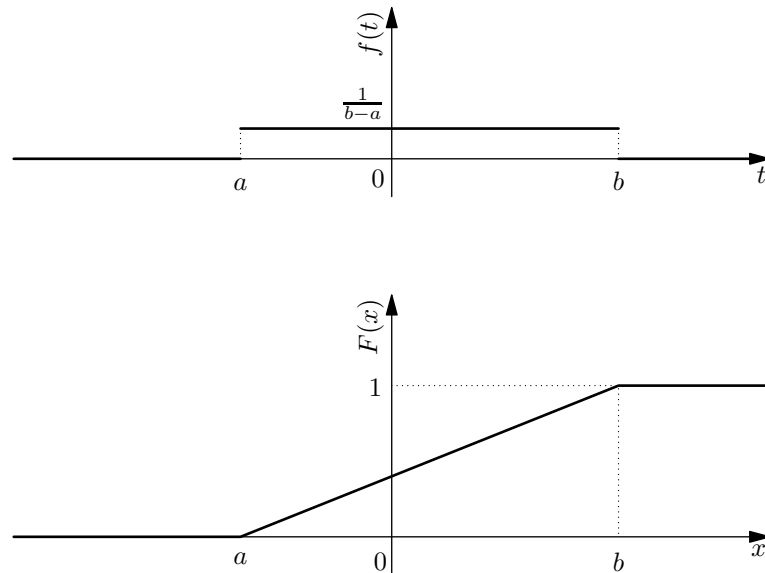
$$\text{pour tout } I \text{ intervalle de } \mathbb{R}, \quad P(X \in I) = P_X(I) = \frac{\ell([a, b] \cap I)}{\ell([a, b])}, \quad (5.37)$$

où  $\ell$  désigne la longueur (en particulier  $\ell([a, b]) = b - a$ ). Notation :  $X \sim \text{Unif}[a, b]$ .

Calculons la fonction de répartition  $F$  en prenant  $I = ]-\infty, x]$  pour  $x$  quelconque dans (5.37).

$$F(x) = P_X(]-\infty, x]) = \frac{\ell([a, b] \cap ]-\infty, x])}{\ell([a, b])} = \begin{cases} 0 & \text{si } -\infty < x < a; \\ \frac{x-a}{b-a} & \text{si } a \leq x < b; \\ 1 & \text{si } b \leq x < +\infty. \end{cases}$$

La fonction de répartition  $F$  est affine par morceaux, donc aussi  $C^1$  par morceaux au sens de la proposition 5.20, avec dérivabilité sur  $\mathbb{R} \setminus \{a, b\}$  (figure 5.8). La loi a donc une densité  $f$  qui s'obtient par dérivation de  $F$ , ce qui nous donne  $f(t) = 0$  si  $t < a$ ,  $f(t) = \frac{1}{b-a}$  si  $a < t < b$  et  $f(t) = 0$  si  $t > b$ . On complète la définition de  $f$  en la prolongeant en  $a$  et

FIGURE 5.8 – Densité  $f$  et f.d.r.  $F$  de la loi Unif $[a, b]$ 

$b$ , par exemple en posant  $f(a) = f(b) = \frac{1}{b-a}$ . La loi uniforme sur  $[a, b]$  admet donc pour densité

$$f = \frac{1}{b-a} \mathbf{1}_{[a,b]}.$$

Dans les calculs faisant intervenir la loi uniforme sur  $[a, b]$ , il est vivement conseillé d'utiliser chaque fois que c'est possible la formule (5.37) de préférence aux calculs d'intégrales de  $f$ .

**Remarque 5.35.** Comme  $\ell(\{a\}) = \ell(\{b\}) = 0$ , la loi uniforme sur  $[a, b]$  est aussi la loi uniforme sur  $]a, b]$ ,  $[a, b[$  ou  $]a, b[$ .

Une des raisons de l'importance de la loi uniforme sur  $[0, 1]$  est le théorème suivant.

**Théorème 5.36.** *Si  $X$  est une variable aléatoire réelle de fonction de répartition continue strictement croissante  $F$  et si  $U$  est une variable aléatoire de loi uniforme sur  $[0, 1]$ , alors la variable aléatoire  $Y := F^{-1}(U)$  a même loi que  $X$ .*

Rappelons qu'avoir même loi que  $X$  ne signifie aucunement être égale à  $X$ . Ce théorème permet de réduire la simulation informatique de la loi de  $X$  à celle de  $U$ . Nous verrons ultérieurement que ce résultat s'étend à toutes les fonctions de répartition, sans hypothèse de continuité ni de croissance stricte, *via* une redéfinition de  $F^{-1}$ .

*Preuve.* Comme  $F$  est continue strictement croissante, c'est une *bijection* de  $\mathbb{R}$  sur son image  $]0, 1[$  (en raison de la stricte monotonie de  $F$ , les bornes 0 et 1 ne sont pas atteintes). Par conséquent  $F^{-1} : ]0, 1[ \rightarrow \mathbb{R}$  est bien définie et vérifie :

$$\forall u \in ]0, 1[, \forall x \in \mathbb{R}, \quad F^{-1}(u) \leq x \text{ si et seulement si } u \leq F(x).$$

Comme  $P(0 < U < 1) = 1$ , on en déduit que les événements  $\{F^{-1}(U) \leq x\}$  et  $\{U \leq F(x)\}$  ont même probabilité. Pour obtenir la fonction de répartition de  $Y$ , on remarque alors que pour tout  $x \in \mathbb{R}$ ,

$$P(Y \leq x) = P(F^{-1}(U) \leq x) = P(U \leq F(x)) = \frac{\ell([0, F(x)])}{\ell([0, 1])} = F(x).$$

Ainsi  $Y$  a pour fonction de répartition  $F$ , donc a même loi que  $X$ . □

### 5.4.2 Lois exponentielles

**Définition 5.37.** Soit  $a$  un réel strictement positif. La variable aléatoire réelle  $X$  suit la loi exponentielle de paramètre  $a$  si elle admet pour densité

$$f(t) = ae^{-at} \mathbf{1}_{[0, +\infty[}(t).$$

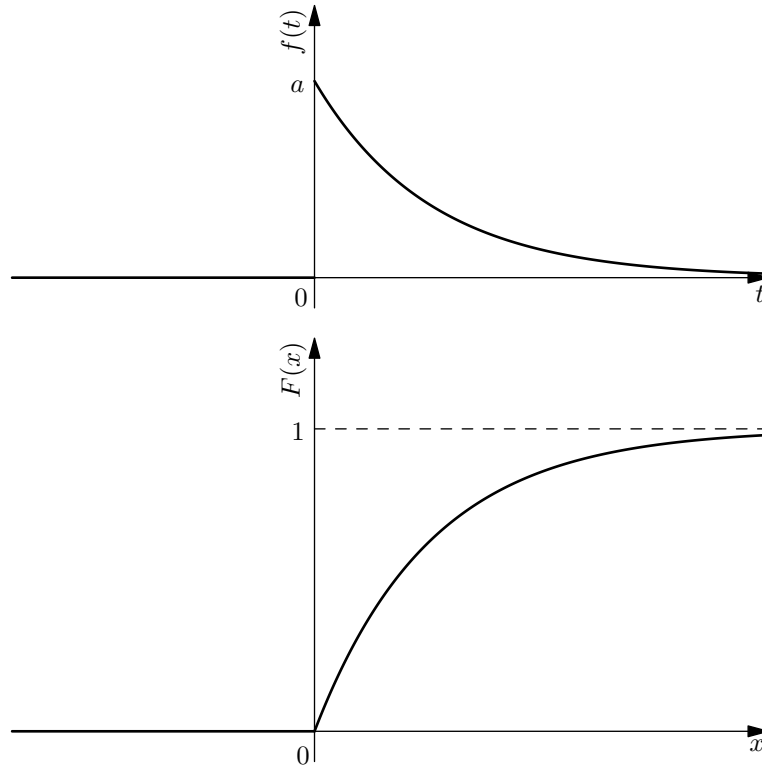


FIGURE 5.9 – Densité et f.d.r. de la loi  $\text{Exp}(a)$

En pratique, plutôt que de travailler avec la fonction de répartition d'une loi exponentielle, il est plus commode d'utiliser la *fonction de survie*  $G$  :

$$G(x) = P(X > x) = 1 - F(x) = \begin{cases} 1 & \text{si } x \leq 0, \\ e^{-ax} & \text{si } x > 0. \end{cases}$$

Les lois exponentielles sont souvent choisies pour modéliser des temps d'attente : temps d'attente à partir de maintenant du prochain tremblement de terre, du prochain faux numéro sur une ligne téléphonique, de la prochaine désintégration d'un atome de radium, etc.

La raison de ce choix est la propriété *d'absence de mémoire* en temps continu qui caractérise la famille des lois exponentielles.

**Théorème 5.38** (absence de mémoire).

i) Si la variable aléatoire  $X$  suit une loi exponentielle, elle vérifie la propriété *d'absence de mémoire* :

$$\forall s \in \mathbb{R}^+, \forall t \in \mathbb{R}^+, \quad P(X > t + s \mid X > t) = P(X > s). \quad (5.38)$$

ii) Réciproquement si une variable aléatoire  $X$  vérifie (5.38), elle suit une loi exponentielle.

Comme la fonction de survie caractérise la loi, (5.38) signifie que la loi de  $(X - t)$  conditionnelle à  $\{X > t\}$  est la même que la loi de  $X$ .

En préliminaire à la preuve du théorème<sup>10</sup>, remarquons que la probabilité conditionnelle dans (5.38) s'exprime commodément à l'aide de la fonction de survie  $G$ , définie par  $G(x) := P(X > x)$ . En effet,  $s$  étant positif, l'inégalité  $t + s \geq t$  entraîne l'inclusion  $\{X > t + s\} \subset \{X > t\}$  et l'égalité d'événements :  $\{X > t + s\} \cap \{X > t\} = \{X > t + s\}$ . On en déduit

$$P(X > t + s \mid X > t) = \frac{P(X > t + s)}{P(X > t)} = \frac{G(t + s)}{G(t)}. \quad (5.39)$$

*Preuve de i).* Si  $X$  suit la loi exponentielle de paramètre  $a$ ,  $G(x) = e^{-ax}$  pour tout  $x$  positif et (5.39) se traduit alors par :

$$P(X > t + s \mid X > t) = \frac{e^{-a(t+s)}}{e^{-at}} = e^{-as} = P(X > s).$$

Ainsi  $X$  de loi  $\text{Exp}(a)$  vérifie la propriété d'absence de mémoire (5.38).  $\square$

*Preuve de ii).* Soit  $X$  une variable aléatoire dont la loi vérifie (5.38) et  $G$  sa fonction de survie. Comme  $G = 1 - F$  (où  $F$  désigne la fonction de répartition de  $X$ ),  $G$  est décroissante et continue à droite et tend vers 0 en  $+\infty$ . De plus l'écriture de (5.38) suppose implicitement que  $G(t) > 0$  pour tout  $t \geq 0$  car sinon  $P(\cdot \mid X > t)$  ne serait pas définie. Grâce à (5.39), on voit que la propriété d'absence de mémoire (5.38) équivaut à

$$\forall s \in \mathbb{R}^+, \forall t \in \mathbb{R}^+, \quad \frac{G(t + s)}{G(t)} = G(s).$$

La fonction de survie  $G$  doit donc être une solution décroissante, continue à droite, tendant vers 0 en  $+\infty$  et telle que  $0 < G(t) \leq 1$  de l'équation fonctionnelle<sup>11</sup> :

$$\forall s \in \mathbb{R}^+, \forall t \in \mathbb{R}^+, \quad G(t + s) = G(t)G(s). \quad (5.40)$$

En faisant  $s = t = 0$  dans (5.40), on obtient  $G(0) = G(0)^2$  et comme  $G(0) > 0$ , on a

$$G(0) = 1. \quad (5.41)$$

En faisant  $s = t$  dans (5.40), on obtient  $G(2t) = G(t)^2$ , puis de proche en proche

$$\forall n \in \mathbb{N}^*, \forall t \geq 0, \quad G(nt) = G(t)^n. \quad (5.42)$$

En particulier pour  $t = 1/d$ ,  $d \in \mathbb{N}^*$  :

$$\forall n \in \mathbb{N}^*, \forall d \in \mathbb{N}^*, \quad G\left(\frac{n}{d}\right) = G\left(\frac{1}{d}\right)^n. \quad (5.43)$$

Lorsque  $n = d$ , (5.43) donne  $G(1) = G(1/d)^d$  d'où

$$\forall d \in \mathbb{N}^*, \quad G\left(\frac{1}{d}\right) = G(1)^{1/d}. \quad (5.44)$$

10. Seule la preuve du point i) est exigible des étudiants de ce cours.

11. Une équation fonctionnelle est une équation dont l'inconnue est ... une fonction! Les équations différentielles sont des exemples bien connus d'équations fonctionnelles.



Nous connaissons maintenant  $G$  sur l'ensemble des rationnels positifs puisque (5.41), (5.42), (5.43) et (5.44) nous donnent

$$\forall r \in \mathbb{Q}^+, \quad G(r) = G(1)^r. \quad (5.45)$$

Soit  $x \in \mathbb{R}^+ \setminus \mathbb{Q}^+$ ,  $x$  est limite d'une suite décroissante  $(r_n)$  de rationnels. Comme  $G$  est continue à droite,  $G(r_n)$  converge vers  $G(x)$ . D'autre part l'application  $y \mapsto G(1)^y$  est continue sur  $\mathbb{R}$ . Ainsi en appliquant (5.45) à  $r_n$  et en faisant tendre  $n$  vers l'infini on obtient

$$\forall x \in \mathbb{R}^+, \quad G(x) = G(1)^x. \quad (5.46)$$

*A priori* la constante  $G(1)$  est dans  $]0, 1]$ . On peut écarter la valeur  $G(1) = 1$  car sinon d'après (5.46), la limite en  $+\infty$  de  $G$  serait 1 alors qu'elle vaut 0.

Finalement, puisque  $0 < G(1) < 1$ , on peut poser  $G(1) = e^{-a}$  pour un réel  $a > 0$  (cela revient à prendre  $a = -\ln G(1)$ ). On peut alors réécrire (5.46) sous la forme

$$\forall x \in \mathbb{R}^+, \quad G(x) = e^{-ax}.$$

La fonction de survie  $G$  est donc la même que celle de la loi exponentielle de paramètre  $a$ , donc  $X$  suit cette loi (puisque la fonction de survie caractérise la loi au même titre que la fonction de répartition).  $\square$

### 5.4.3 Lois gaussiennes

Ces lois jouent un rôle capital dans l'étude des lois limites de sommes de variables aléatoires indépendantes. Par exemple (théorème de de Moivre Laplace) si  $S_n$  suit la loi  $\text{Bin}(n, p)$ , alors pour tout  $x \in \mathbb{R}$ ,  $P(S_n - np \leq x\sqrt{np(1-p)})$  converge quand  $n$  tend vers l'infini vers  $\Phi(x)$ , où  $\Phi$  est la f.d.r. de la loi gaussienne  $\mathfrak{N}(0, 1)$ .

**Définition 5.39.** *On dit que la variable aléatoire  $X$  suit la loi gaussienne ou normale  $\mathfrak{N}(m, \sigma)$  si elle a pour densité la fonction :*

$$f_{m,\sigma} : \mathbb{R} \longrightarrow \mathbb{R}^+ \quad t \longmapsto \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(t-m)^2}{2\sigma^2}\right).$$

La loi  $\mathfrak{N}(0, 1)$  est appelée loi normale standard.

Tous les calculs de probabilités concernant une variable aléatoire de loi  $\mathfrak{N}(m, \sigma)$  peuvent se ramener à des calculs sur une variable de loi normale standard.

**Proposition 5.40.** *Si la variable aléatoire  $X$  suit la loi  $\mathfrak{N}(m, \sigma)$ ,  $Y := (X - m)/\sigma$  suit la loi  $\mathfrak{N}(0, 1)$ . Autrement dit, toute v.a. gaussienne  $X$  de loi  $\mathfrak{N}(m, \sigma)$  peut s'écrire  $X = \sigma Y + m$  avec  $Y$  de loi  $\mathfrak{N}(0, 1)$ .*

*Preuve.* On calcule  $P(a < Y \leq b)$  pour  $a$  et  $b$  réels quelconques ( $a < b$ ).

$$\begin{aligned} P\left(a < \frac{X - m}{\sigma} \leq b\right) &= P(\sigma a + m < X \leq \sigma b + m) \\ &= \int_{\sigma a + m}^{\sigma b + m} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x - m)^2}{2\sigma^2}\right) dx. \end{aligned}$$

Il suffit alors de faire le changement de variable  $y = (x - m)/\sigma$  pour obtenir

$$\forall a \in \mathbb{R}, \forall b > a, \quad P(a < Y \leq b) = \int_a^b \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right) dy.$$

Donc  $Y$  a bien pour densité  $f_{0,1}$ .  $\square$

**Remarque 5.41.** En adaptant le calcul fait dans la preuve de la proposition 5.40, on voit facilement que la famille des lois gaussiennes est *stable par transformations affines* : si  $X$  a pour loi  $\mathfrak{N}(m, \sigma)$ , alors pour tout  $(\alpha, \beta) \in \mathbb{R}^* \times \mathbb{R}$ , la v.a.  $\alpha X + \beta$  est encore gaussienne, de loi  $\mathfrak{N}(\alpha m + \beta, |\alpha|\sigma)$ .

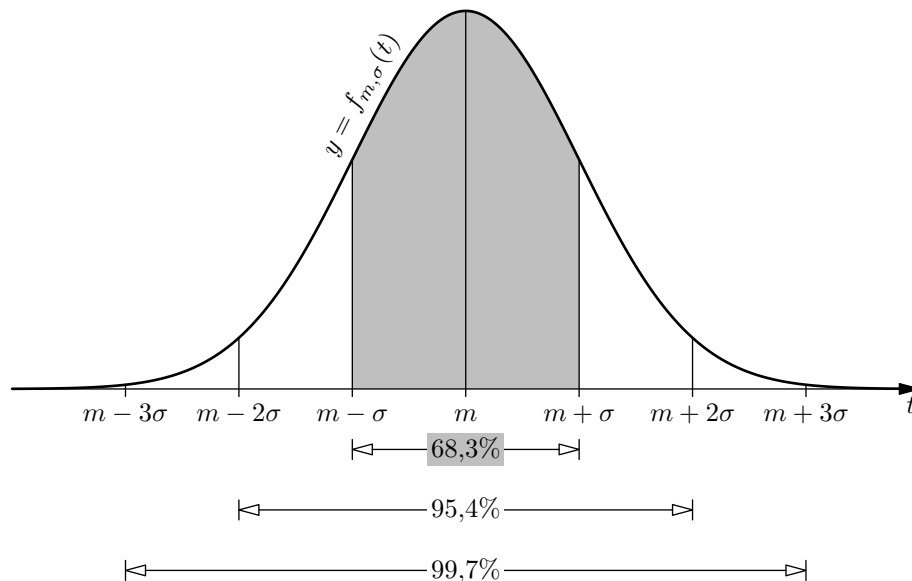


FIGURE 5.10 – Concentration de la loi  $\mathfrak{N}(m, \sigma)$  autour de  $m$

La figure 5.10 illustre la signification du paramètre de position  $m$  et du paramètre de dispersion  $\sigma$  pour la loi gaussienne  $\mathfrak{N}(m, \sigma)$ .

Cette concentration de pratiquement toute la probabilité dans l'intervalle  $[m - 3\sigma, m + 3\sigma]$  permet l'utilisation des lois gaussiennes pour modéliser des grandeurs aléatoires qui *a priori* prennent leurs valeurs seulement dans un petit intervalle de  $\mathbb{R}^+$  : taille, poids, ..., même si théoriquement une variable gaussienne peut prendre toute valeur entre  $-\infty$  et  $+\infty$ .

Il n'existe pas d'expression d'une primitive de la densité gaussienne  $f_{m, \sigma}$  à l'aide des fonctions usuelles. Les valeurs de la fonction de répartition  $\Phi$  de  $\mathfrak{N}(0, 1)$  sont tabulées, cf. page 106. D'après la proposition 5.40, ceci suffit pour calculer numériquement n'importe quelle f.d.r. de loi gaussienne.

#### 5.4.4 Lois de Cauchy

**Définition 5.42.** La variable aléatoire  $X$  suit la loi de Cauchy (ou loi de Cauchy de paramètres 0 et 1) si elle admet pour densité :

$$f(t) = \frac{1}{\pi(1+t^2)}.$$

*Notation* :  $X \sim \text{Cau}(0, 1)$ .

Cette loi est *symétrique*, ce qui signifie que  $X$  et  $-X$  ont même loi, ceci résultant ici de la parité de  $f$ . La fonction de répartition  $F$  est donnée par :

$$F(x) = \int_{-\infty}^x \frac{dt}{\pi(1+t^2)} = \frac{1}{\pi} \left( \frac{\pi}{2} + \arctan x \right),$$

où  $\arctan x$  est l'unique réel  $y \in ]-\pi/2, \pi/2[$  tel que  $\tan y = x$ .

Si  $Y = a + bX$ , avec  $X$  de loi  $\text{Cau}(0, 1)$ ,  $a \in \mathbb{R}$  et  $b \in \mathbb{R}_+^*$ , on dit encore que  $Y$  suit une loi de Cauchy, de paramètres  $(a, b)$ , notation  $Y \sim \text{Cau}(a, b)$ . La densité est alors

$$f_{a,b}(t) = \frac{1}{\pi b} \frac{1}{1 + \left(\frac{t-a}{b}\right)^2}.$$

### 5.4.5 Tables de la loi normale

On note  $\Phi$  la fonction de répartition de la loi normale standard  $\mathfrak{N}(0, 1)$  :

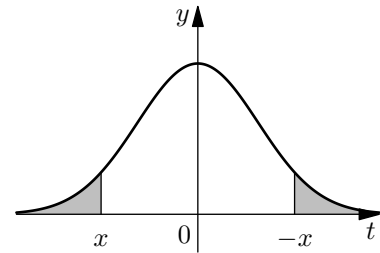
$$\forall x \in \mathbb{R}, \quad \Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right) dt.$$

La table de la page 106 donne des valeurs approchées à  $10^{-4}$  près de  $\Phi(x)$  pour  $x = \frac{k}{100}$ ,  $k \in \llbracket 0, 299 \rrbracket$ . Pour lire la valeur de  $\Phi(x)$ , on repère dans la première colonne la ligne qui comporte le chiffre des unités et celui des dixièmes de  $x$ . À l'intersection de cette ligne et de la colonne dont l'entête est le chiffre des centièmes de  $x$ , on lit la valeur de  $\Phi(x)$ . Par exemple pour  $\Phi(1,37)$ , on lit à l'intersection de la ligne commençant par 1,3 et de la colonne commençant par 0,07 la valeur  $\Phi(1,37) \simeq 0,9147$ .

La table donne les valeurs de  $\Phi(x)$  pour  $x$  positif. Lorsque  $x$  est négatif, on utilise la relation

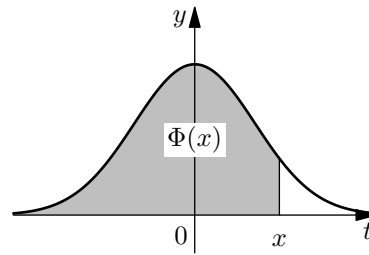
$$\Phi(x) = 1 - \Phi(-x)$$

qui résulte de la parité de la densité gaussienne  $\mathfrak{N}(0, 1)$ , voir ci-contre. Exemple : pour  $x = -1,8$ , on trouve :  $\Phi(x) = 1 - \Phi(1,8) \simeq 1 - 0,9641 = 0,0359$ .



Pour les « très grandes valeurs de  $x$  », disons  $|x| > 4$ , on peut utiliser l'encadrement suivant de la « queue » de la loi normale.

$$\forall x > 0, \quad \left(\frac{1}{x} - \frac{1}{x^3}\right) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) < 1 - \Phi(x) < \frac{1}{x} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right).$$

Table des valeurs de  $\Phi$ , f.d.r. de la loi normale standard  $\mathcal{N}(0, 1)$ 

$x$	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2,6	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974
2,8	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,9980	0,9981
2,9	0,9981	0,9982	0,9982	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986

Table pour les grandes valeurs de  $x$ 

$x$	3,0	3,1	3,2	3,3	3,4	3,5	3,6	3,8	4,0
$\Phi(x)$	0,99865	0,99904	0,99931	0,99952	0,99966	0,99976	0,999841	0,999928	0,999968

# Chapitre 6

## Espérance

### 6.1 Introduction

L'ESPÉRANCE  $\mathbf{E}X$  d'une variable aléatoire  $X$  est, lorsqu'elle existe, la *moyenne des valeurs de cette variable, pondérées par leurs probabilités de réalisation*.

Cette définition informelle est trop naïve pour permettre un traitement unifié de toutes les lois<sup>1</sup>. Rappelons qu'il existe des lois qui ne sont ni discrètes ni à densité et que la description la plus générale des lois de variables aléatoires réelles est donnée par leur fonction de répartition. Il est donc naturel de chercher à définir  $\mathbf{E}X$  à partir de la fonction de répartition  $F : t \mapsto P(X \leq t)$ . Nous allons motiver cette définition en nous restreignant au cas des variables aléatoires positives et en partant du cas simple où  $X$  est discrète avec  $X(\Omega) = \{x_1, \dots, x_n\}$  partie finie de  $\mathbb{R}_+$ . Dans ce cas, la définition informelle de  $\mathbf{E}X$  se traduit par la formule

$$\mathbf{E}X = \sum_{k=1}^n x_k P(X = x_k). \quad (6.1)$$

Les figures 6.1 et 6.2 nous montrent comment exprimer cette moyenne pondérée à l'aide de

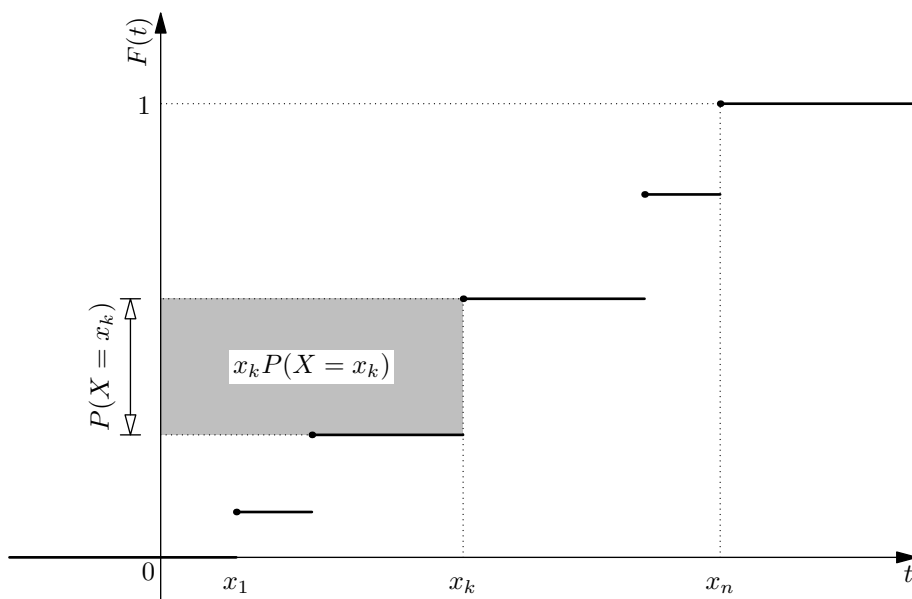


FIGURE 6.1 – Interprétation graphique des  $x_k P(X = x_k)$ , pour  $x_k \geq 0$

1. Cette définition de  $\mathbf{E}X$  nous fait pressentir que  $\mathbf{E}X$  ne doit dépendre que de la loi de  $X$ .

$F$ . Rappelons que dans ce cas,  $F$  présente en chaque  $x_k$  un saut d'amplitude  $P(X = x_k)$ .

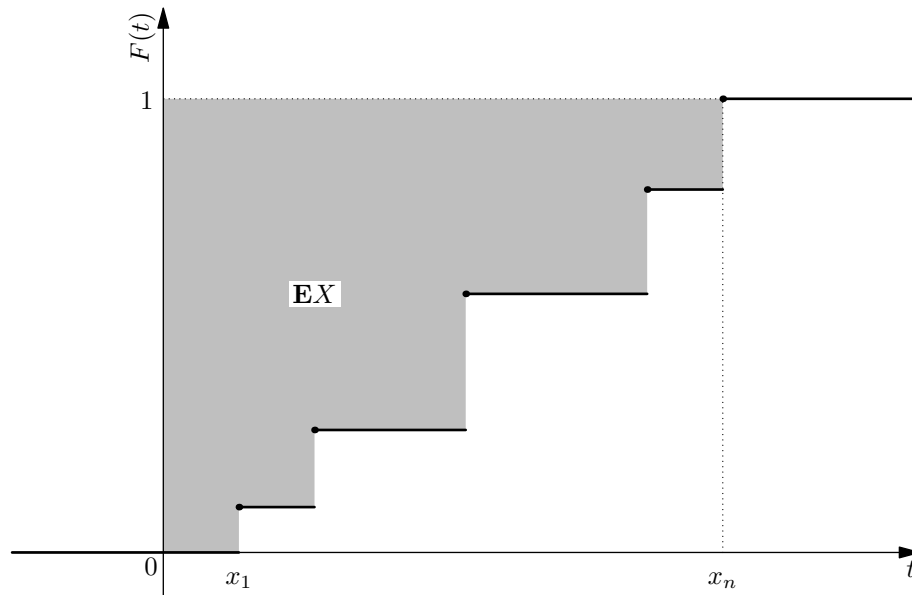


FIGURE 6.2 – Interprétation graphique de  $\mathbf{E}X = \sum_{k=1}^n x_k P(X = x_k)$ , les  $x_k \geq 0$ .

L'interprétation graphique en terme d'aires donnée par la figure 6.2 nous permet d'écrire  $\mathbf{E}X$  comme l'intégrale de Riemann ordinaire :  $\mathbf{E}X = \int_0^{x_n} (1 - F(t)) dt$  et aussi comme la fausse intégrale généralisée  $\int_0^{+\infty} (1 - F(t)) dt$ . Si on passe maintenant au cas d'une variable aléatoire positive quelconque, il paraît alors naturel de considérer que  $\mathbf{E}X$  est l'aire (éventuellement infinie) délimitée par le segment vertical  $t = 0$ ,  $y \in [0, 1]$ , la demi droite « asymptote »  $y = 1$ ,  $t \geq 0$  et le graphe de  $F$ , ce qui nous conduit à la formule

$$\mathbf{E}X := \int_0^{+\infty} (1 - F(t)) dt = \int_0^{+\infty} P(X > t) dt, \quad \text{pour toute v.a. positive } X.$$

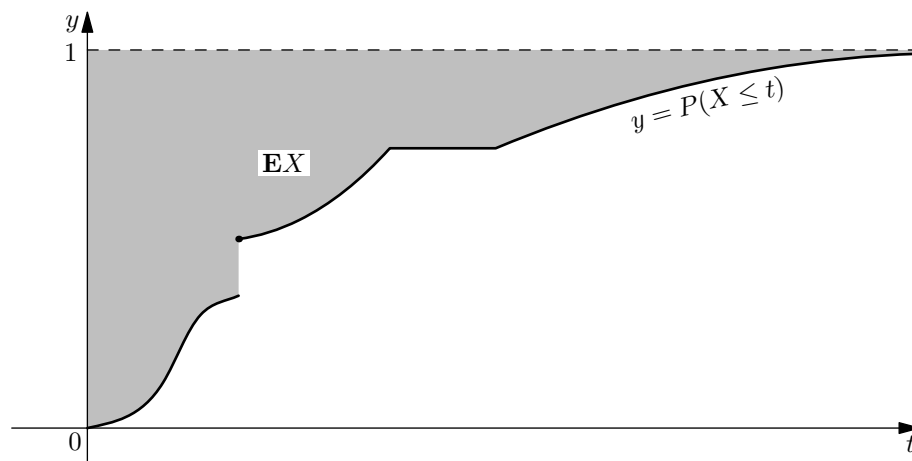


FIGURE 6.3 – Interprétation graphique de  $\mathbf{E}X$  via la f.d.r. de  $X$  v.a. positive.

## 6.2 Espérance d'une variable aléatoire positive

Dans toute la suite de ce chapitre, on fixe un espace probabilisé  $(\Omega, \mathcal{F}, P)$ . Toutes les variables aléatoires considérées seront, sauf mention explicite du contraire, définies sur cet

espace et leur loi sera la loi sous  $P$ .

**Définition 6.1** (espérance d'une v.a. positive). Soit  $X$  une variable aléatoire positive sur  $(\Omega, \mathcal{F})$ . On appelle espérance de  $X$  (ou espérance de  $X$  sous  $P$ ) la quantité

$$\mathbf{E}X := \int_0^{+\infty} P(X > t) dt = \lim_{b \rightarrow +\infty} \int_0^b P(X > t) dt, \quad (6.2)$$

qui est un élément de  $\overline{\mathbb{R}}_+$ .

Pour justifier l'existence de  $\mathbf{E}X$ , on commence par noter que l'application  $G : \mathbb{R}_+ \rightarrow [0, 1]$ ,  $t \mapsto G(t) := P(X > t)$  est décroissante sur  $\mathbb{R}_+$ , donc Riemann intégrable sur  $[0, b]$  pour tout  $b \in \mathbb{R}_+^*$ . L'intégrale  $\int_0^b G(t) dt = \int_0^b P(X > t) dt$  existe donc bien et est un réel positif pour tout  $b$ . Comme c'est une fonction croissante de sa borne supérieure  $b$ , elle converge dans  $\overline{\mathbb{R}}_+$  quand  $b$  tend vers  $+\infty$ .

Dans cette section, nous utiliserons l'interprétation graphique de  $\mathbf{E}X$  via la fonction de survie  $t \mapsto P(X > t)$ , cf. figure 6.4, plutôt que via la f.d.r.  $F : t \mapsto P(X \leq t)$ . On passe évidemment d'une représentation à l'autre en effectuant une symétrie orthogonale par rapport à la droite  $y = 1/2$ , puisque  $G = 1 - F$ . Cette symétrie conserve les aires.

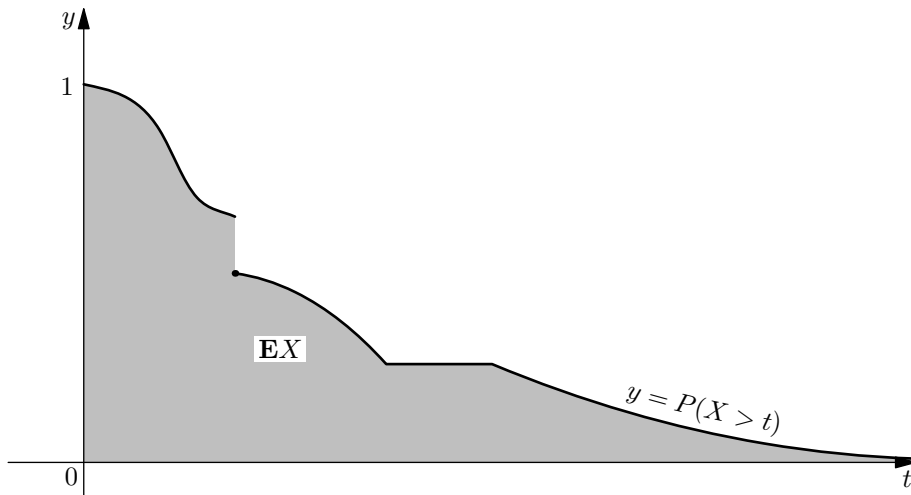


FIGURE 6.4 – Interprétation graphique de  $\mathbf{E}X$  via la fonction de survie de  $X$  v.a. positive.

**Remarque 6.2.**  $\mathbf{E}X$  ne dépend que de la loi de  $X$ , il serait donc plus correct de parler de l'espérance de la loi de  $X$  sous  $P$  au lieu de l'espérance de  $X$ . L'usage donne néanmoins la préférence à cette dernière appellation quand il n'y a pas d'ambiguïté sur  $P$ .

**Remarque 6.3** (espérance d'une v.a. presque sûrement positive). Dans les exercices, la variable aléatoire  $X$  n'est pas toujours donnée explicitement, il arrive assez souvent que l'on ne connaisse que sa loi  $P_X$ . Si  $P_X(\mathbb{R}_+) = 1$ , on s'autorisera une généralisation de la définition 6.1 en considérant que la formule (6.2) reste valable. Il s'agit bien d'une généralisation car on peut avoir  $P(X \geq 0) = 1$  sans que  $X(\omega)$  soit positif ou nul pour tout  $\omega$ , par exemple si  $\Omega = \mathbb{R}$ ,  $\mathcal{F} = \text{Bor}(\mathbb{R})$ ,  $P$  est la loi uniforme sur  $[0, 1]$  et  $X : \omega \mapsto \omega$  est l'identité sur  $\mathbb{R}$ . On a alors  $X(\omega) < 0$  pour une infinité non dénombrable de  $\omega$  et  $P(X \geq 0) = 1$ . Cette généralisation est cohérente avec la remarque 6.2 car si  $X$  est une variable aléatoire réelle définie sur un espace probabilisé  $(\Omega, \mathcal{F}, P)$  et telle que  $P(X \geq 0) = 1$ , on peut toujours trouver un espace probabilisé  $(\Omega', \mathcal{F}', P')$  et une variable aléatoire positive  $X'$  définie sur cet espace tels que  $X$  et  $X'$  aient même loi.

**Définition 6.4** (intégrabilité d'une v.a. positive). *On dit que la variable aléatoire positive  $X$  est intégrable si*

$$\int_0^{+\infty} P(X > t) dt < +\infty. \quad (6.3)$$

**Exemple 6.5.** Si la variable aléatoire positive  $X$  est bornée, c.-à-d. s'il existe une constante  $c$  telle que pour tout  $\omega \in \Omega$ ,  $0 \leq X(\omega) \leq c$ , alors elle est intégrable. En effet pour  $t \geq c$ ,  $P(X > t) = 0$ , ce qui réduit l'intégrale généralisée définissant  $\mathbf{E}X$  à une intégrale de Riemann ordinaire  $\int_0^c P(X > t) dt$  donc finie (et majorée par  $c$ ).

Plus généralement, si la loi de  $X$ , v.a. positive, vérifie  $P(X > t) \leq Ct^{-\alpha}$  pour un certain  $\alpha > 1$  et tout  $t \geq t_0 > 0$ , ou si  $P(X > t) \leq t^{-1}(\ln t)^{-\beta}$  pour un  $\beta > 1$  et tout  $t \geq t_0 > 0$ , alors  $X$  est intégrable. Réciproquement, l'intégrabilité de  $X$  nous donne un renseignement sur la vitesse de convergence<sup>2</sup> vers 0 de  $P(X > t)$  quand  $t$  tend vers  $+\infty$ . C'est l'inégalité de Markov que nous verrons ci-dessous (proposition 6.21).

Voyons maintenant quelques exemples simples de calcul d'espérance de variables aléatoires positives.

**Exemple 6.6** (espérance d'une constante positive). Si la variable aléatoire  $X$  est une constante positive  $c$ , c.-à-d.  $X(\omega) = c$  pour tout  $\omega \in \Omega$ , alors  $\mathbf{E}X = c$ . En effet :

$$P(X > t) = \begin{cases} 1 & \text{si } t < c \\ 0 & \text{si } t \geq c \end{cases} = \mathbf{1}_{]-\infty, c[}(t),$$

d'où

$$\mathbf{E}X = \int_0^{+\infty} \mathbf{1}_{]-\infty, c[}(t) dt = \int_0^c \mathbf{1}_{[0, c[}(t) dt = \int_0^c 1 dt = c.$$

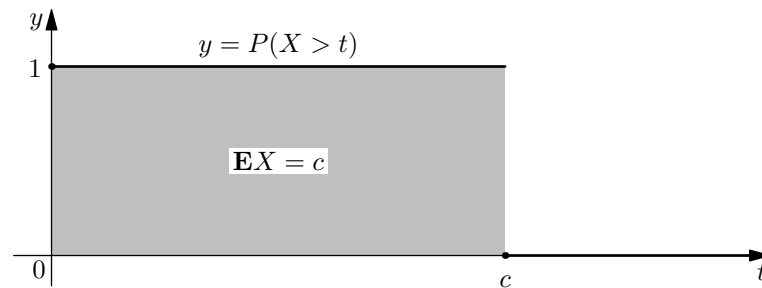


FIGURE 6.5 – Espérance de la v.a. constante  $X = c$

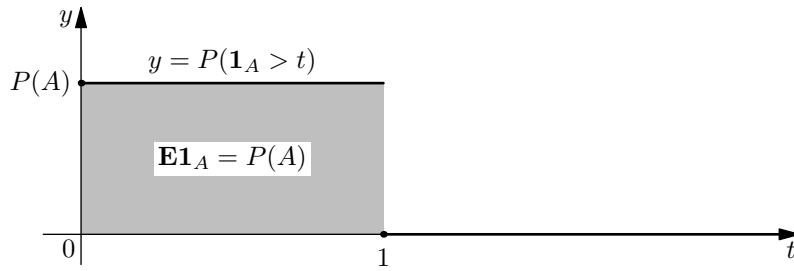
L'exemple suivant est d'une grande importance car il permet d'écrire toute probabilité d'évènement comme une espérance. Nous le formulons sous forme de proposition.

**Proposition 6.7** (espérance d'une indicatrice d'évènement). *Pour tout évènement  $A \in \mathcal{F}$ ,*

$$\mathbf{E}(\mathbf{1}_A) = P(A). \quad (6.4)$$

2. Pour n'importe quelle variable aléatoire  $X$ ,  $P(X > t)$  tend vers 0 quand  $t$  tend vers  $+\infty$ , car  $P(X > t) = 1 - F(t)$ , où  $F$  est la f.d.r. de  $X$  qui tend toujours vers 1 en  $+\infty$ .



FIGURE 6.6 – Espérance de la v.a. indicatrice  $X = \mathbf{1}_A$ 

*Preuve.* La variable aléatoire positive  $\mathbf{1}_A$  prend la valeur 1 sur l'évènement  $A$  et 0 sur  $A^c$ , elle suit la loi de Bernoulli de paramètre  $p = P(A)$ . L'évènement  $\{\mathbf{1}_A > t\}$  est donc égal à  $A$  si  $0 \leq t < 1$  et à l'ensemble vide si  $t \geq 1$ . On en déduit que

$$P(\mathbf{1}_A > t) = \begin{cases} P(A) & \text{si } 0 \leq t < 1, \\ 0 & \text{si } t \geq 1. \end{cases}$$

Par conséquent,

$$\mathbf{E}(\mathbf{1}_A) = \int_0^1 P(A) dt = P(A).$$

□

**Définition 6.8** (variable aléatoire simple). *On dit que la variable aléatoire réelle  $X$  définie sur  $(\Omega, \mathcal{F})$  est simple ou étagée si  $X(\Omega)$  est fini. En notant  $X(\Omega) = \{x_1, \dots, x_n\}$ , les  $x_i$  étant tous distincts,  $X$  admet la décomposition*

$$X = \sum_{k=1}^n x_k \mathbf{1}_{A_k}, \quad \text{où } A_k := \{X = x_k\}, \quad k \in \llbracket 1, n \rrbracket, \quad (6.5)$$

les évènements  $A_k$  formant une partition de  $\Omega$ .

**Proposition 6.9** (espérance d'une v.a. positive simple). *Si  $X$  est une variable aléatoire positive simple avec  $X(\Omega) = \{x_1, \dots, x_n\}$ ,*

$$\mathbf{E}X = \sum_{k=1}^n x_k P(X = x_k). \quad (6.6)$$

On retrouve ainsi la formule (6.1) de l'introduction.

*Preuve.* Notons en préliminaire qu'il nous faut résister ici à la tentation de dire « c'est immédiat en utilisant la décomposition (6.5), la proposition 6.7 et la linéarité de l'espérance », car nous n'avons pas encore prouvé que l'espérance est linéaire. En fait la proposition 6.9 est l'un des ingrédients de la preuve de la linéarité de l'espérance. Il nous faut donc vérifier (6.6) par un calcul direct basé sur la définition 6.1.

Quitte à réindexer, on peut toujours supposer que les  $x_k$  sont rangés par ordre croissant. Notons  $p_i := P(X = x_i)$  et  $s_k := \sum_{1 \leq i \leq k} p_i$ . La fonction de répartition  $F$  peut alors s'écrire (cf. la proposition 5.11) :

$$F(t) = \sum_{k=1}^{n-1} s_k \mathbf{1}_{[x_k, x_{k+1}[}(t) + s_n \mathbf{1}_{[x_n, +\infty[}(t). \quad (6.7)$$

Notons que pour  $t \geq x_n$ ,  $F(t) = s_n = 1$ , donc  $P(X > t) = 1 - F(t) = 0$ . Ainsi  $\mathbf{E}X = \int_0^{+\infty} P(X > t) dt = \int_0^{x_n} P(X > t) dt$ . On peut alors calculer  $\mathbf{E}X$  en utilisant la décomposition (6.7) et les propriétés de l'intégrale de Riemann sur l'intervalle fermé borné  $[0, x_n]$  :

$$\mathbf{E}X = \int_0^{x_n} (1 - F(t)) dt = x_n - \int_0^{x_n} F(t) dt = x_n - \sum_{k=1}^{n-1} \int_{x_k}^{x_{k+1}} s_k dt,$$

d'où

$$\begin{aligned} \mathbf{E}X &= x_n - \sum_{k=1}^{n-1} (x_{k+1} - x_k) s_k = x_n - \sum_{j=2}^n x_j s_{j-1} + \sum_{j=1}^{n-1} x_j s_j \\ &= x_n - x_n s_{n-1} + \sum_{j=2}^{n-1} x_j (s_j - s_{j-1}) + x_1 s_1 \\ &= x_n p_n + \sum_{j=2}^{n-1} x_j p_j + x_1 p_1 = \sum_{j=1}^n p_j x_j. \end{aligned}$$

Ceci établit (6.6). □

**Exemple 6.10** (espérance d'une loi uniforme discrète). Si  $X$  suit la loi uniforme sur l'ensemble fini  $\{x_1, \dots, x_n\}$ ,  $\mathbf{E}X$  est égale à la moyenne arithmétique des  $x_i$ . En effet :  $X(\Omega) = \{x_1, \dots, x_n\}$  et  $P(X = x_i) = 1/n$  pour tout  $i$  compris entre 1 et  $n$ . D'où :

$$\mathbf{E}X = \sum_{i=1}^n x_i P(X = x_i) = \frac{1}{n} \sum_{i=1}^n x_i.$$

Le calcul de l'espérance des lois binomiale et hypergéométrique est moins immédiat et sera vu ultérieurement.

**Proposition 6.11** (espérance d'une v.a. discrète positive). *Pour toute variable aléatoire discrète positive  $X$ ,*

$$\mathbf{E}X = \sum_{x \in X(\Omega)} x P(X = x), \quad (\text{égalité dans } \overline{\mathbb{R}}_+). \quad (6.8)$$

*Preuve dans le cas courant.* Le cas où  $X(\Omega)$  est fini est déjà traité par la proposition 6.9. Si  $X(\Omega)$  est infini dénombrable, nous supposons en plus qu'il existe une numérotation croissante de  $X(\Omega)$  par les entiers de  $\mathbb{N}$ . C'est le cas le plus courant<sup>3</sup>. Alors  $X(\Omega) = \{x_0, x_1, \dots, x_n, \dots\}$  et  $(x_n)_{n \in \mathbb{N}}$  est une suite croissante dans  $\mathbb{R}_+$ . Elle converge donc vers une limite  $c \leq +\infty$ .

1. Cas  $c = +\infty$ . On a alors par définition de l'intégrale généralisée :

$$\mathbf{E}X = \int_0^{+\infty} P(X > t) dt = \lim_{b \rightarrow +\infty} \int_0^b P(X > t) dt = \lim_{n \rightarrow +\infty} \int_0^{x_n} P(X > t) dt.$$

Sur l'intervalle  $[0, x_n]$ , la fonction de survie  $t \mapsto P(X > t)$  est en escaliers avec un nombre fini de marches. On en déduit facilement (revoir la construction de la figure 6.2) que :

$$\int_0^{x_n} P(X > t) dt = \sum_{k=0}^n x_k P(X = x_k).$$

3. Le cas où il n'existe pas de numérotation croissante de  $X(\Omega)$  est hors de portée de ce cours.

Par conséquent,

$$\lim_{n \rightarrow +\infty} \sum_{k=0}^n x_k P(X = x_k) = \lim_{n \rightarrow +\infty} \int_0^{x_n} P(X > t) dt,$$

d'où

$$\sum_{k=0}^{+\infty} x_k P(X = x_k) = \int_0^{+\infty} P(X > t) dt = \mathbf{E}X \quad (\text{égalité dans } \overline{\mathbb{R}}_+).$$

2. Cas  $c < +\infty$ . Si on montre que  $\int_0^{+\infty} P(X > t) dt = \int_0^c P(X > t) dt$ , on pourra conclure avec le même argument que dans le cas précédent. Or la suite  $(x_n)_{n \in \mathbb{N}}$  étant croissante, sa limite  $c$  est sa borne supérieure, ce qui entraîne qu'il n'existe aucun  $x_n$  à la droite de  $c$ . Comme  $X$  est discrète, on en déduit que :

$$P(X \leq c) = \sum_{x_k \leq c} P(X = x_k) = \sum_{x_k \in X(\Omega)} P(X = x_k) = 1.$$

Donc  $P(X > c) = 0$  et pour tout  $t \geq c$ ,  $P(X > t) = 0$ , d'où  $\int_0^{+\infty} P(X > t) dt = \int_0^c P(X > t) dt$ .

□

**Exemple 6.12** (espérance d'une loi géométrique). Si  $X$  suit la loi géométrique de paramètre  $p > 0$ ,  $\mathbf{E}X = \frac{1}{p}$ . On a ici  $X_P(\Omega) = \mathbb{N}^*$ ,  $P(X = k) = q^{k-1}p$  où  $q = 1 - p \in ]0, 1[$ . La série à termes *positifs* :

$$\sum_{k=1}^{+\infty} k P(X = k) = \sum_{k=1}^{+\infty} k q^{k-1} p$$

est convergente dans  $\mathbb{R}_+$ , nous l'avons déjà étudiée et calculé sa somme comme application de la dérivabilité d'une série entière, la série géométrique standard, cf. (2.20), p. 38.

**Exemple 6.13** (espérance d'une loi de Poisson). Si  $X$  suit la loi de Poisson de paramètre  $\alpha$ ,  $\mathbf{E}X = \alpha$ . En effet on a ici :  $X(\Omega) = \mathbb{N}$ ,  $P(x = k) = e^{-\alpha} \alpha^k / k!$ . La série de terme général *positif*  $u_k = k P(X = k)$  est convergente car le développement en série entière de  $e^z = \sum_{k=0}^{+\infty} z^k / k!$  a un rayon de convergence infini et la série entière  $\sum_{k=0}^{+\infty} k z^k / k!$  a même rayon de convergence, cf. lemme 2.34. Donc  $\mathbf{E}X$  existe dans  $\mathbb{R}_+$ . La somme de cette série se calcule ainsi :

$$\mathbf{E}X = \sum_{k=0}^{+\infty} k P(X = k) = \sum_{k=1}^{+\infty} \frac{e^{-\alpha} \alpha^k}{(k-1)!} = \alpha e^{-\alpha} \sum_{l=0}^{+\infty} \frac{\alpha^l}{l!} = \alpha.$$

**Proposition 6.14** (espérance d'une v.a. positive à densité). *Si la variable aléatoire positive  $X$  a pour densité  $f$ ,*

$$\mathbf{E}X = \int_0^{+\infty} x f(x) dx. \quad (6.9)$$

*Dans cette formule,  $\mathbf{E}X$  peut prendre la valeur  $+\infty$  si l'intégrale généralisée diverge.*

*Preuve.* Si  $X$  admet pour densité  $f$ ,  $P(X > t) = \int_t^{+\infty} f(x) dx$  pour tout  $t$ . En reportant cette égalité dans la définition de  $\mathbf{E}X$ , on obtient :

$$\int_0^{+\infty} P(X > t) dt = \int_0^{+\infty} \left\{ \int_t^{+\infty} f(x) dx \right\} dt = \int_0^{+\infty} \left\{ \int_0^{+\infty} f(x) \mathbf{1}_{[t, +\infty[}(x) dx \right\} dt.$$

Notons que pour  $t \geq 0$ ,  $\mathbf{1}_{[t, +\infty[}(x) = \mathbf{1}_{[0, x]}(t)$ . L'intégrande  $(x, t) \mapsto \mathbf{1}_{[0, x]}(t)f(x)$  étant *positive*, le théorème de Fubini-Tonelli légitime l'interversion des intégrations, ce qui donne :

$$\mathbf{E}X = \int_0^{+\infty} \left\{ \int_0^{+\infty} f(x) \mathbf{1}_{[0, x]}(t) dt \right\} dx = \int_0^{+\infty} f(x) \left\{ \int_0^{+\infty} \mathbf{1}_{[0, x]}(t) dt \right\} dx.$$

Comme pour  $x \geq 0$ ,  $\int_0^{+\infty} \mathbf{1}_{[0, x]}(t) dt = \int_0^x dt = x$ , on en déduit (6.9).  $\square$

**Remarque 6.15.** Notons que dans la démonstration ci-dessus, nous n'avons utilisé à aucun moment la positivité de la variable aléatoire  $X$ . On peut donc appliquer ce calcul à toute variable aléatoire réelle  $X$  ayant une densité  $f$  pour obtenir :

$$\int_0^{+\infty} P(X > t) dt = \int_0^{+\infty} xf(x) dx \quad (\text{égalité dans } \overline{\mathbb{R}}_+). \quad (6.10)$$

Attention à ne pas écrire  $\mathbf{E}X$  au premier membre de (6.10), cette quantité n'étant pour l'instant définie que pour  $X$  *positive*. La vraie formule pour  $\mathbf{E}X$  lorsque la v.a. réelle  $X$  est à densité est donnée à la proposition 6.28.

**Exemple 6.16** (espérance d'une loi exponentielle). Si  $X$  suit la loi exponentielle de paramètre  $a > 0$ ,  $\mathbf{E}X = \frac{1}{a}$ . En effet la densité est  $f(t) = ae^{-at}\mathbf{1}_{\mathbb{R}_+}(t)$ , d'où

$$\mathbf{E}X = \int_0^{+\infty} axe^{-ax} dx$$

et on obtient le résultat après intégration par parties (faites le!). On peut d'ailleurs retrouver ce résultat en notant que la fonction de survie  $G(x) = e^{-ax}$  pour  $x > 0$ , d'où

$$\mathbf{E}X = \int_0^{+\infty} e^{-ax} dx = \int_0^{+\infty} e^{-y} \frac{1}{a} dy = \lim_{b \rightarrow +\infty} \left[ \frac{-e^{-y}}{a} \right]_0^b = \frac{1}{a}.$$

**Proposition 6.17.** Si  $X$  est une variable aléatoire positive et  $c$  une constante réelle strictement positive, on a

$$\mathbf{E}(cX) = c\mathbf{E}X.$$

Cette égalité reste vraie pour  $c = 0$  si  $X$  est de plus intégrable.

*Preuve.* Puisque  $X$  est une variable aléatoire positive et  $c$  une constante positive,  $cX : \omega \mapsto (cX)(\omega) := cX(\omega)$  est une variable aléatoire positive. En lui appliquant la définition 6.1, on obtient :

$$\mathbf{E}(cX) = \int_0^{+\infty} P(cX > t) dt = \int_0^{+\infty} P\left(X > \frac{t}{c}\right) dt.$$

Dans cette intégrale généralisée d'une fonction *positive* localement intégrable sur  $[0, +\infty[$ , on peut effectuer le changement de variable  $s = t/c$ , qui nous donne :

$$\mathbf{E}(cX) = \int_0^{+\infty} P\left(X > \frac{t}{c}\right) dt = c \int_0^{+\infty} P(X > s) ds = c\mathbf{E}X.$$

Dans le cas particulier  $c = 0$ , cette méthode n'est plus valable (on ne peut déjà plus écrire «  $P(cX > t) = P(X > t/c)$  ») mais la formule est vraie trivialement, à condition que  $\mathbf{E}X$  soit *finie*, puisqu'alors  $\mathbf{E}(0 \times X) = \mathbf{E}(0) = 0$  et  $0 \times \mathbf{E}X = 0$ .  $\square$

**Proposition 6.18** (croissance de l'espérance). *Si  $X$  et  $Y$  sont deux variables aléatoires positives définies sur le même  $(\Omega, \mathcal{F}, P)$  et si  $X \leq Y$  c.-à-d.  $X(\omega) \leq Y(\omega)$  pour tout  $\omega \in \Omega$ , alors  $\mathbf{E}X \leq \mathbf{E}Y$ .*

*Preuve.* Si  $X(\omega) > t$ , alors comme  $Y(\omega) \geq X(\omega)$ , on a aussi  $Y(\omega) > t$ . Ceci justifie l'inclusion d'évènements  $\{X > t\} \subset \{Y > t\}$ , puis l'inégalité  $P(X > t) \leq P(Y > t)$ . Cette dernière inégalité étant vérifiée pour tout  $t$ , on peut l'intégrer entre 0 et  $+\infty$ , pour obtenir<sup>4</sup> :

$$\mathbf{E}X = \int_0^{+\infty} P(X > t) dt \leq \int_0^{+\infty} P(Y > t) dt = \mathbf{E}Y.$$

□

La proposition suivante nous donne un peu de confort pour l'expression des espérances de variables aléatoires positives et nous sera utile pour l'inégalité de Markov.

**Proposition 6.19.** *Pour toute variable aléatoire positive  $X$ ,*

$$\int_0^{+\infty} P(X > t) dt = \int_0^{+\infty} P(X \geq t) dt \quad (\text{égalité dans } \overline{\mathbb{R}}_+). \quad (6.11)$$

Avant d'en donner la preuve, notons que (6.11) n'a rien d'évident car  $P(X > t)$  et  $P(X \geq t)$  peuvent différer pour certaines valeurs de  $t$  (au plus pour une infinité dénombrable de valeurs de  $t$ ).

*Preuve.* Notons respectivement  $I$  et  $J$  le premier et le deuxième membre de (6.11). On prouve leur égalité en montrant l'inégalité dans les deux sens. L'inégalité  $I \leq J$  s'obtient par intégration de l'inégalité  $P(X > t) \leq P(X \geq t)$  vraie pour tout  $t$ .

Pour montrer que  $J \leq I$ , fixons  $\varepsilon > 0$  quelconque. L'intégrande dans  $J$  est une fonction *positive* localement Riemann intégrable sur  $[0, +\infty[$ . On peut donc effectuer dans  $J$  le changement de variable « translation »  $t = s + \varepsilon$ , qui nous donne :

$$J = \int_{-\varepsilon}^{+\infty} P(X \geq s + \varepsilon) ds = \int_{-\varepsilon}^0 P(X \geq s + \varepsilon) ds + \int_0^{+\infty} P(X \geq s + \varepsilon) ds.$$

En majorant  $P(X \geq s + \varepsilon)$  par 1 sur  $[-\varepsilon, 0]$  et par  $P(X > s)$  sur  $[0, +\infty[$ , on en déduit que

$$J \leq \varepsilon + \int_0^{+\infty} P(X > s) ds = \varepsilon + I.$$

L'inégalité  $J \leq I + \varepsilon$  étant ainsi vérifiée pour tout  $\varepsilon > 0$ , on en déduit en faisant tendre  $\varepsilon$  vers 0 que  $J \leq I$ . □

**Remarque 6.20.** Dans la démonstration ci-dessus, la positivité de  $X$  ne joue aucun rôle. Donc (6.11) reste valable pour n'importe quelle variable aléatoire réelle  $X$ . Une adaptation facile de la preuve ci-dessus montre que l'on a aussi

$$\int_{-\infty}^0 P(X < t) dt = \int_{-\infty}^0 P(X \leq t) dt. \quad (6.12)$$

---

4. La croissance de l'intégrale de Riemann (si  $f \leq g$  sur  $[a, b]$ ,  $\int_a^b f(t) dt \leq \int_a^b g(t) dt$ ) passe aux intégrales généralisées de fonctions *positives*. En effet, si  $f$  et  $g$  sont positives et localement Riemann intégrables sur  $[0, +\infty[$  et telles que  $f \leq g$  sur  $[0, +\infty[$ , alors  $\int_0^x f(t) dt \leq \int_0^x g(t) dt$  pour tout  $x \geq 0$  et cette inégalité entre deux fonctions *croissantes* de  $x$  se conserve dans  $\overline{\mathbb{R}}_+$  par passage à la limite quand  $x$  tend vers  $+\infty$ .

**Proposition 6.21** (inégalité de Markov). *Si  $X$  est une variable aléatoire positive,*

$$\forall x > 0, \quad P(X \geq x) \leq \frac{\mathbf{E}X}{x}. \quad (6.13)$$

**Remarques 6.22.** Cette inégalité n'a d'intérêt que lorsque le second membre est inférieur à 1, c'est-à-dire lorsque  $\mathbf{E}X < +\infty$  et  $x > \mathbf{E}X$ .

D'autre part il peut sembler un peu incongru de vouloir contrôler  $P(X \geq x)$  à l'aide de  $\mathbf{E}X$ , puisque le calcul de cette espérance par la définition 6.1 présuppose la connaissance des  $P(X > t)$  pour  $t \geq 0$  (dont on déduit facilement les  $P(X \geq t)$ ). Il se trouve qu'il arrive souvent en pratique que l'on sache calculer  $\mathbf{E}X$  sans connaître, ou sans avoir besoin de calculer, la loi de  $X$ . C'est le cas par exemple quand  $X$  est une somme finie de variables aléatoires d'espérances connues. On peut aussi savoir majorer  $\mathbf{E}X$  sans connaître la loi de  $X$ . Dans ces situations, l'inégalité de Markov est très utile. Pour ne citer qu'un exemple, l'inégalité de Markov est l'un des outils pour établir des « lois des grands nombres ».

Voici maintenant 3 preuves de l'inégalité de Markov, libre au lecteur de choisir celle qu'il préfère.

*Preuve n° 1.* C'est la preuve « muette » donnée par la figure 6.7.

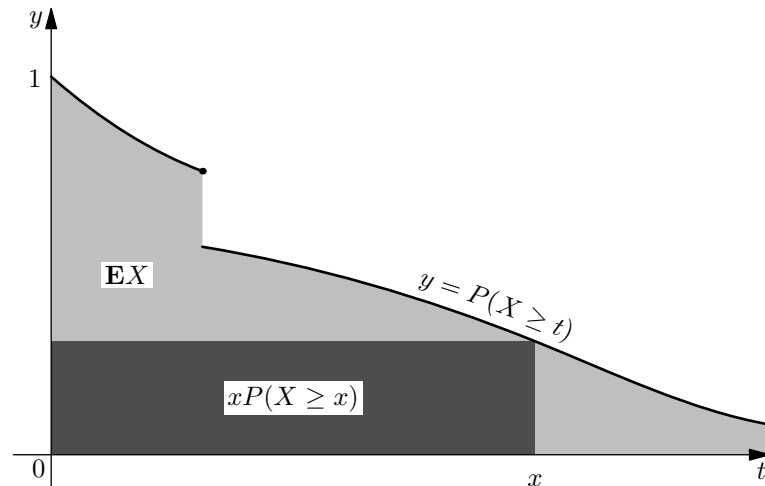


FIGURE 6.7 – Inégalité de Markov :  $xP(X \geq x) \leq \int_0^{+\infty} P(X \geq t) dt = \mathbf{E}X$ .

□

*Preuve n° 2.* Cette preuve ne fait que traduire explicitement la preuve graphique n° 1. Fixons  $x > 0$ , la quantité  $P(X \geq x)$  devenant ainsi une constante. À partir de cette constante, définissons la fonction  $h : [0, +\infty[ \rightarrow \mathbb{R}_+, t \mapsto P(X \geq x) \mathbf{1}_{[0, x]}(t)$ . Par décroissance de la fonction  $t \mapsto P(X \geq t)$ , on a  $h(t) \leq P(X \geq t)$  pour tout  $t \in [0, x]$ . D'autre part cette inégalité est aussi vérifiée pour tout  $t > x$  car alors  $h(t) = 0$ . En intégrant sur  $[0, +\infty[$  l'inégalité  $h(t) \leq P(X \geq t)$ , on obtient compte-tenu de (6.11) :

$$\int_0^{+\infty} h(t) dt \leq \int_0^{+\infty} P(X \geq t) dt = \mathbf{E}X.$$

D'autre part, puisque  $h$  est nulle sur  $]x, +\infty[$ ,

$$\int_0^{+\infty} h(t) dt = \int_0^x P(X \geq x) dt = xP(X \geq x).$$

Par conséquent

$$xP(X \geq x) \leq \mathbf{E}X,$$

ce qui nous donne (6.13) puisque  $x > 0$ .  $\square$

*Preuve n° 3.* Cette preuve plus abstraite exploite les propriétés déjà connues de l'espérance des v.a. positives. On fixe  $x$  qui joue donc le rôle d'une constante dans toute la preuve. On part de l'inégalité entre v.a. positives :  $x\mathbf{1}_{\{X \geq x\}} \leq X$  (vérifiez) dont on déduit par croissance de  $\mathbf{E}$  (proposition 6.18) :

$$\mathbf{E}(x\mathbf{1}_{\{X \geq x\}}) \leq \mathbf{E}X,$$

puis grâce aux propositions 6.17 et 6.7,  $xP(X \geq x) \leq \mathbf{E}X$ . On conclut en divisant par  $x > 0$ .  $\square$

**Corollaire 6.23.** *Si  $X$  est une variable aléatoire positive, on a l'équivalence*

$$\mathbf{E}X = 0 \Leftrightarrow P(X = 0) = 1,$$

*autrement dit,  $\mathbf{E}X$  est nulle si et seulement si  $X$  est presque sûrement nulle.*

*Preuve.* Supposons d'abord que  $\mathbf{E}X = 0$ . La v.a.  $X$  étant positive, l'égalité  $P(X = 0) = 1$  équivaut à  $P(X > 0) = 0$ . Introduisons la suite des événements  $A_n := \{X \geq 1/n\}$ ,  $n \in \mathbb{N}^*$ . Cette suite est croissante de réunion  $A := \{X > 0\}$ . Par continuité séquentielle croissante de  $P$ ,  $P(A) = \lim_{n \rightarrow +\infty} P(A_n)$ . Or l'inégalité de Markov appliquée avec  $x = 1/n$  nous montre que pour tout  $n \in \mathbb{N}^*$ ,  $0 \leq P(A_n) \leq n\mathbf{E}X = 0$ . Ainsi  $P(A_n) = 0$  pour tout  $n$  et  $P(A) = 0$  comme limite de la suite nulle.

Réciproquement, si  $P(X = 0) = 1$ , alors pour tout  $t \geq 0$ ,  $P(X > t) = 0$ , d'où  $\mathbf{E}X = \int_0^{+\infty} 0 dt = 0$ .  $\square$

## 6.3 Espérance d'une variable aléatoire réelle

L'extension de la notion d'espérance au cas des variables aléatoires réelles repose sur la décomposition  $X = X^+ - X^-$  et la formule «  $\mathbf{E}X := \mathbf{E}(X^+) - \mathbf{E}(X^-)$  », sous réserve que cette soustraction ait lieu dans  $\mathbb{R}$ . Avant d'expliquer cela, donnons quelques précisions sur les variables aléatoires positives  $X^+$  et  $X^-$ , définies par

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

Notons que  $X^+$  et  $X^-$  sont des variables aléatoires positives. En se rappelant la définition 1.30, on a donc :

$$X^+ : \Omega \rightarrow \mathbb{R}_+, \quad \omega \mapsto X^+(\omega) = (X(\omega))^+, \quad X^- : \Omega \rightarrow \mathbb{R}_+, \quad \omega \mapsto X^-(\omega) = (X(\omega))^-.$$

Au risque d'insister lourdement, notons encore que pour tout  $\omega \in \Omega$ ,

$$X^+(\omega) = \begin{cases} |X(\omega)| & \text{si } X(\omega) \geq 0, \\ 0 & \text{si } X(\omega) \leq 0, \end{cases} \quad X^-(\omega) = \begin{cases} 0 & \text{si } X(\omega) \geq 0, \\ |X(\omega)| & \text{si } X(\omega) \leq 0. \end{cases}$$

On en déduit les égalités

$$X = X^+ - X^-, \quad |X| = X^+ + X^-.$$

Pour définir  $\mathbf{E}X$  comme la différence  $\mathbf{E}(X^+) - \mathbf{E}(X^-)$ , il est clair qu'il nous faut interdire que ces deux espérances de variables aléatoires positives vaillent *simultanément*  $+\infty$ . On pourrait autoriser l'une des deux à valoir  $+\infty$  à condition que l'autre soit finie,  $\mathbf{E}X$  prenant alors la valeur  $+\infty$  si  $\mathbf{E}(X^+) = +\infty$  et  $\mathbf{E}(X^-) < +\infty$ , ou dans le cas inverse,  $\mathbf{E}X = -\infty$ . Nous ne retiendrons pas cette option, car le problème de «  $+\infty - \infty$  » réapparaîtrait de toutes façons pour la somme de deux variables aléatoires, ce qui empêcherait l'additivité de l'espérance pour les variables aléatoires réelles. Ainsi nous ne parlerons de l'espérance (sous-entendu dans  $\mathbb{R}$ ) de la variable aléatoire réelle  $X$  que si  $\mathbf{E}(X^+) < +\infty$  et  $\mathbf{E}(X^-) < +\infty$ , ou ce qui est équivalent si  $\mathbf{E}|X| < +\infty$ . Les propriétés de l'espérance établies dans cette section ne concernent que les variables aléatoires remplissant cette condition<sup>5</sup>. On dit que ces variables aléatoires réelles sont *intégrables*.

**Définition 6.24** (v.a. réelle intégrable). *La variable aléatoire réelle  $X$  est dite intégrable si  $\mathbf{E}|X| < +\infty$ , autrement dit si :*

$$\int_0^{+\infty} P(|X| > t) dt < +\infty.$$

**Définition 6.25** (espérance d'une v.a. réelle). *Si  $X$  est une variable aléatoire réelle intégrable, on appelle espérance de  $X$ , le réel  $\mathbf{E}X$  défini par*

$$\mathbf{E}X := \mathbf{E}(X^+) - \mathbf{E}(X^-). \quad (6.14)$$

Le second membre de (6.14) a bien un sens et représente un nombre réel, puisqu'en raison des inégalités  $0 \leq X^+ \leq |X|$  et  $0 \leq X^- \leq |X|$ , l'intégrabilité de  $X$  implique la finitude de  $\mathbf{E}(X^+)$  et de  $\mathbf{E}(X^-)$ .

**Remarque 6.26.** Une fois le réel  $\mathbf{E}X$  défini comme ci-dessus, on peut parler de sa partie positive  $(\mathbf{E}X)^+$  et de sa partie négative  $(\mathbf{E}X)^-$ . Il faut prendre garde à ne pas les confondre avec  $\mathbf{E}(X^+)$  et  $\mathbf{E}(X^-)$ . Pour voir un exemple simple où ces quantités diffèrent, prenons  $X$  variable aléatoire de Rademacher :  $X(\Omega) = \{-1, 1\}$  et  $P(X = -1) = P(X = 1) = 1/2$ . Comme  $X$  est bornée, elle est intégrable, donc  $\mathbf{E}X$  existe. Comme la loi de  $X$  est symétrique,  $\mathbf{E}X = 0$ , donc  $(\mathbf{E}X)^+ = 0^+ = 0$  et  $(\mathbf{E}X)^- = 0^- = 0$ . D'autre part, on vérifie que  $X^+$  et  $X^-$  sont des variables de Bernoulli de paramètre  $1/2$ , donc  $\mathbf{E}(X^+) = \mathbf{E}(X^-) = 1/2$ .

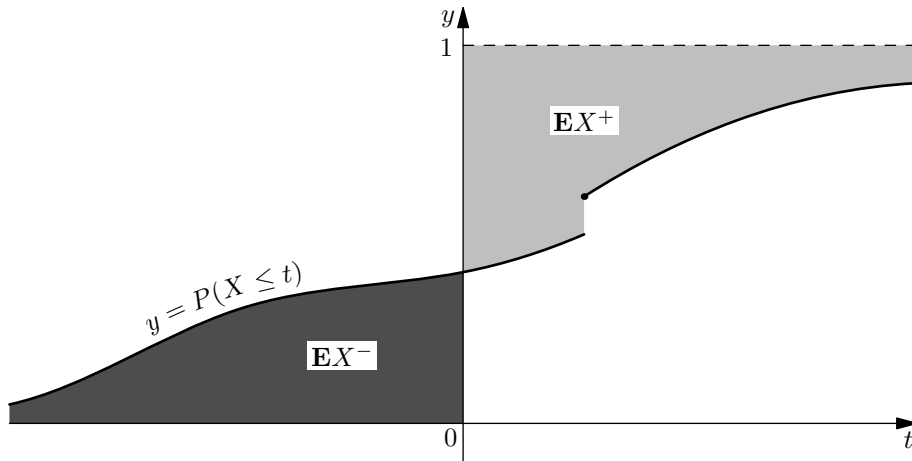
La formule (6.14) ne semble pas très commode pour le calcul explicite de  $\mathbf{E}X$ . On aimerait pouvoir exprimer  $\mathbf{E}X$ , directement à l'aide de la loi de  $X$ , par exemple de sa fonction de répartition, sans passer par la loi de  $X^+$  et celle de  $X^-$ . C'est l'objet de la proposition suivante.

**Proposition 6.27** (calcul de  $\mathbf{E}X$  à l'aide de la f.d.r.). *Soit  $X$  une variable aléatoire réelle intégrable de fonction de répartition  $F$ . Son espérance vérifie*

$$\mathbf{E}X = \int_0^{+\infty} P(X > t) dt - \int_{-\infty}^0 P(X < t) dt = \int_0^{+\infty} (1 - F(t)) dt - \int_{-\infty}^0 F(t) dt. \quad (6.15)$$

5. À titre exceptionnel, dans certaines situations, il peut être commode de dire qu'une variable aléatoire réelle a une espérance *infinie* si *une seule* des deux quantités  $\mathbf{E}(X^+)$  et  $\mathbf{E}(X^-)$  est infinie, par exemple dans la situation de la remarque 6.3, pour rester cohérent avec la définition de l'espérance d'une v.a. *positive* comme élément de  $\overline{\mathbb{R}}_+$ . Il convient alors de manipuler ces « espérances infinies » avec les plus grandes précautions. En tout état de cause, on ne parle *jamais* d'espérance de  $X$ , ni finie, ni infinie, lorsque  $\mathbf{E}X^+$  et  $\mathbf{E}X^-$  valent tous deux  $+\infty$ .



FIGURE 6.8 – Espérance d'une v.a. réelle  $\mathbf{E}X := \mathbf{E}(X^+) - \mathbf{E}(X^-)$ 

*Preuve.* Compte-tenu des remarques faites au début de cette section, on vérifie facilement pour tout  $t \geq 0$ , les équivalences logiques suivantes :

$$X^+(\omega) > t \Leftrightarrow X(\omega) > t, \quad (6.16)$$

$$X^-(\omega) > t \Leftrightarrow X(\omega) < -t. \quad (6.17)$$

Laissant (6.16) au lecteur, détaillons la deuxième. Comme  $t \geq 0$ , l'inégalité  $X^-(\omega) > t$  implique que  $X^-(\omega)$  est strictement positif, donc que  $X(\omega)$  est strictement négatif et dans ce cas,  $X^-(\omega) = -X(\omega)$ , d'où  $-X(\omega) > t$  et donc  $X(\omega) < -t$ . Ceci justifie l'implication  $\Rightarrow$  dans (6.17). Réciproquement, si  $X(\omega) < -t$ ,  $X(\omega)$  est strictement négatif, donc  $X(\omega) = -X^-(\omega)$ , d'où  $-X^-(\omega) < -t$  et donc  $X^-(\omega) > t$ .

Les équivalences (6.16) et (6.17) donnent les égalités d'évènements  $\{X^+ > t\} = \{X > t\}$  et  $\{X^- > t\} = \{X < -t\}$ , valables pour tout <sup>6</sup>  $t \geq 0$ , d'où :

$$\forall t \geq 0, \quad P(X^+ > t) = P(X > t), \quad P(X^- > t) = P(X < -t). \quad (6.18)$$

En reportant ces égalités dans la définition de  $\mathbf{E}X$ , on justifie le passage à (6.19) dans le calcul suivant :

$$\begin{aligned} \mathbf{E}X = \mathbf{E}(X^+) - \mathbf{E}(X^-) &= \int_0^{+\infty} P(X^+ > t) dt - \int_0^{+\infty} P(X^- > t) dt \\ &= \int_0^{+\infty} P(X > t) dt - \int_0^{+\infty} P(X < -t) dt \end{aligned} \quad (6.19)$$

$$= \int_0^{+\infty} P(X > t) dt - \int_{-\infty}^0 P(X < s) ds \quad (6.20)$$

$$= \int_0^{+\infty} (1 - F(t)) dt - \int_{-\infty}^0 F(t) dt. \quad (6.21)$$

Le passage de (6.19) à (6.20) résulte bien sûr du changement de variable  $s = -t$  et justifie la première égalité dans (6.15). Le passage de (6.20) à (6.21) résulte de la remarque 6.20 et du remplacement de la « variable muette  $s$  » par  $t$ .  $\square$

6. Elles ne seraient pas valables pour tout  $t$  réel. La positivité de  $t$  est utilisée dans la vérification de (6.16) et (6.17).

**Proposition 6.28** (espérance d'une v.a. réelle à densité). *Soit  $X$  une variable aléatoire réelle de densité  $f$ . Alors  $X$  est intégrable si et seulement si*

$$\int_{-\infty}^{+\infty} |x|f(x) dx < +\infty \quad (6.22)$$

et dans ce cas,

$$\mathbf{E}X = \int_{-\infty}^{+\infty} xf(x) dx. \quad (6.23)$$

*Preuve.* Nous savons déjà par (6.18) et la remarque 6.15 que

$$\int_0^{+\infty} P(X^+ > t) dt = \int_0^{+\infty} P(X > t) dt = \int_0^{+\infty} xf(x) dx = \int_0^{+\infty} |x|f(x) dx, \quad (6.24)$$

sans aucune condition d'intégrabilité sur  $X$ .

Cherchons une formule analogue pour  $\int_0^{+\infty} P(X^- > t) dt$ . En utilisant (6.18), le passage de (6.19) à (6.20) et la remarque 6.20, on peut démarrer avec :

$$\int_0^{+\infty} P(X^- > t) dt = \int_{-\infty}^0 P(X < t) dt = \int_{-\infty}^0 P(X \leq t) dt.$$

On procède alors comme dans la preuve de la proposition 6.14 en injectant dans cette formule l'égalité  $P(X \leq t) = \int_{-\infty}^t f(x) dx$  et en justifiant l'interversion des intégrations par le théorème de Fubini-Tonelli.

$$\begin{aligned} \int_{-\infty}^0 P(X \leq t) dt &= \int_{-\infty}^0 \left\{ \int_{-\infty}^t f(x) dx \right\} dt \\ &= \int_{-\infty}^0 \left\{ \int_{-\infty}^0 \mathbf{1}_{]-\infty, t]}(x) f(x) dx \right\} dt \\ &= \int_{-\infty}^0 \left\{ \int_{-\infty}^0 \mathbf{1}_{[x, 0]}(t) f(x) dt \right\} dx \\ &= \int_{-\infty}^0 \left\{ \int_{-\infty}^0 \mathbf{1}_{[x, 0]}(t) dt \right\} f(x) dx. \end{aligned}$$

Comme pour  $x \leq 0$ ,  $\int_{-\infty}^0 \mathbf{1}_{[x, 0]}(t) dt = \int_x^0 dt = -x$ , on obtient finalement :

$$\int_0^{+\infty} P(X^- > t) dt = \int_{-\infty}^0 (-x)f(x) dx = \int_{-\infty}^0 |x|f(x) dx. \quad (6.25)$$

En appliquant l'additivité de l'espérance des v.a. positives à l'égalité  $|X| = X^+ + X^-$  et en rassemblant (6.24) et (6.25), on voit que

$$\mathbf{E}|X| = \mathbf{E}(X^+) + \mathbf{E}(X^-) = \int_0^{+\infty} |x|f(x) dx + \int_{-\infty}^0 |x|f(x) dx = \int_{-\infty}^{+\infty} |x|f(x) dx,$$

ce qui établit la condition nécessaire et suffisante d'intégrabilité (6.22). Si cette condition est réalisée,  $\mathbf{E}(X^+)$  et  $\mathbf{E}(X^-)$  sont finis et

$$\mathbf{E}X = \mathbf{E}(X^+) - \mathbf{E}(X^-) = \int_0^{+\infty} xf(x) dx - \int_{-\infty}^0 (-x)f(x) dx = \int_{-\infty}^{+\infty} xf(x) dx,$$

ce qui établit (6.23). □

**Exemple 6.29** (espérance de la loi  $\mathfrak{N}(0, 1)$ ). Si  $X$  suit la loi  $\mathfrak{N}(0, 1)$ ,  $X$  est intégrable et son espérance est nulle. En effet, la densité est  $f(x) = (2\pi)^{-1/2} \exp(-x^2/2)$ , donc l'intégrabilité se teste en vérifiant la convergence de l'intégrale généralisée :

$$I = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} |x| \exp\left(-\frac{x^2}{2}\right) dx.$$

Par parité, la convergence de cette intégrale se réduit à celle de  $\int_0^{+\infty} (2\pi)^{-1/2} |x| \exp(-x^2/2) dx$ . Il suffit maintenant de majorer pour  $x \geq x_0$  la fonction à intégrer par une fonction plus simple dont l'intégrale généralisée converge<sup>7</sup>. Par exemple, on peut remarquer que pour  $x \geq 4$ ,  $x^2 \geq 4x$ , d'où

$$\forall x \geq 4, \quad |x| \exp\left(-\frac{x^2}{2}\right) = \exp\left(-\frac{x^2}{2} + \ln x\right) \leq \exp(-2x + \ln x) \leq \exp(-x),$$

en utilisant l'inégalité  $\ln x \leq x - 1 \leq x$  valable pour tout  $x > 0$  donc pour  $x \geq 4$ . Et comme  $\lim_{b \rightarrow +\infty} \int_4^b \exp(-x) dx = e^{-4} < +\infty$ , la convergence  $I$  en découle.

Maintenant que l'on sait que  $I$  converge, l'existence de  $\mathbf{E}X$  est justifiée et

$$\mathbf{E}X = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} x \exp\left(-\frac{x^2}{2}\right) dx = 0,$$

parce que l'on intègre une fonction impaire sur un intervalle symétrique.

**Proposition 6.30** (espérance d'une v.a. discrète). *La variable aléatoire discrète  $X$  est intégrable si et seulement si*

$$\sum_{x \in X(\Omega)} |x| P(X = x) < +\infty \quad (6.26)$$

et dans ce cas,

$$\mathbf{E}X = \sum_{x \in X(\Omega)} x P(X = x), \quad (6.27)$$

cette somme désignant une série absolument convergente dans le cas où  $X(\Omega)$  est infini<sup>8</sup>.

*Preuve.*  $X$  étant discrète,  $X(\Omega)$  est au plus dénombrable. Notons  $B := X(\Omega) \cap ]0, +\infty[$  et  $B' := X(\Omega) \cap ]-\infty, 0[$ . Alors  $X^+(\Omega)$  est égal à  $B$  si tous les  $x \in X(\Omega)$  sont strictement positifs ou à  $B \cup \{0\}$  si l'un au moins d'entre eux est négatif ou nul<sup>9</sup>. Quoiqu'il en soit, on a toujours en utilisant la proposition 6.11 :

$$\mathbf{E}(X^+) = \sum_{x \in X^+(\Omega)} x P(X^+ = x) = \sum_{x \in X^+(\Omega)} x P(X = x) = \sum_{x \in B} x P(X = x), \quad (6.28)$$

puisque pour  $x = 0$  le terme éventuel  $x P(X = x)$  est nul. De même,  $X^-(\Omega)$  est égal à  $\{-x ; x \in B'\} =: -B'$  si tous les  $x \in X(\Omega)$  sont strictement négatifs ou à  $(-B') \cup \{0\}$  si l'un au moins d'entre eux est positif ou nul. En appliquant la proposition 6.11 à la v.a. positive discrète  $X^-$ ,

$$\mathbf{E}(X^-) = \sum_{y \in X^-(\Omega)} y P(X^- = y) = \sum_{x \in B'} (-x) P(X = x) = \sum_{x \in B'} |x| P(X = x). \quad (6.29)$$

7. Si vous ne connaissez pas cette règle de comparaison, vous pouvez l'admettre par analogie avec la règle pour les séries à termes positifs.

8.  $X(\Omega)$  est alors forcément dénombrable puisque  $X$  est discrète.

9. Il en résulte que  $X^+(\Omega)$  est au plus dénombrable et donc que la v.a. positive  $X^+$  est discrète.

En rassemblant (6.28) et (6.29), on obtient (noter que  $B \cap B' = \emptyset$ ) :

$$\begin{aligned} \mathbf{E}|X| &= \sum_{x \in B} xP(X = x) + \sum_{x \in B'} |x|P(X = x) \\ &= \sum_{x \in B \cup B'} |x|P(X = x) = \sum_{x \in X(\Omega)} |x|P(X = x), \end{aligned}$$

ce qui nous donne la CNS d'intégrabilité (6.26). Si cette condition est réalisée,  $\mathbf{E}(X^+)$  et  $\mathbf{E}(X^-)$  sont finis et

$$\begin{aligned} \mathbf{E}X &= \mathbf{E}(X^+) - \mathbf{E}(X^-) = \sum_{x \in B} xP(X = x) - \sum_{x \in B'} (-x)P(X = x) \\ &= \sum_{x \in B \cup B'} xP(X = x) \\ &= \sum_{x \in X(\Omega)} xP(X = x), \end{aligned}$$

ce qui établit (6.27). □

**Proposition 6.31** (Linéarité de l'espérance).

a) L'espérance des v.a. réelles intégrables est additive : si  $X$  et  $Y$  v.a. réelles définies sur le même  $(\Omega, \mathcal{F}, P)$  sont intégrables, alors  $X + Y$  l'est aussi et

$$\mathbf{E}(X + Y) = \mathbf{E}X + \mathbf{E}Y. \quad (6.30)$$

b) Si  $X$  est intégrable,  $cX$  l'est aussi pour toute constante réelle  $c$  et

$$\mathbf{E}(cX) = c\mathbf{E}X. \quad (6.31)$$

Nous admettons cette proposition. Voici trois applications à des calculs d'espérance de lois classiques.

**Exemple 6.32** (espérance d'une loi binomiale). Si  $X$  suit la loi binomiale  $\text{Bin}(n, p)$ ,  $\mathbf{E}X = np$ . On sait en effet que  $X$  a même loi que la somme

$$S = \sum_{i=1}^n \mathbf{1}_{A_i},$$

où les  $A_i$  sont des évènements indépendants de même probabilité  $p$ . Comme l'espérance d'une v. a. ne dépend que de sa loi,  $\mathbf{E}X = \mathbf{E}S$  et par linéarité de l'espérance :

$$\mathbf{E}X = \mathbf{E} \left( \sum_{i=1}^n \mathbf{1}_{A_i} \right) = \sum_{i=1}^n \mathbf{E}\mathbf{1}_{A_i} = np,$$

en utilisant la proposition 6.7.

**Exemple 6.33** (espérance d'une loi hypergéométrique). Si  $X$  suit la loi hypergéométrique  $\text{Hypg}(N, M, n)$ ,  $\mathbf{E}X = n \frac{M}{N}$ .

L'espérance d'une v. a. ne dépendant que de la loi, on ne perd pas de généralité en supposant que  $X$  est le nombre d'objets défectueux observé dans un échantillon de taille

$n$  prélevé sans remise dans une population totale de  $N$  objets dont  $M$  sont défectueux. En numérotant de 1 à  $M$  tous les objets défectueux, on a alors

$$X = X_1 + \cdots + X_M,$$

où

$$X_i = \begin{cases} 1 & \text{si le } i\text{-ème objet défectueux est prélevé,} \\ 0 & \text{sinon.} \end{cases}$$

Chaque  $X_i$  est une variable de Bernoulli de paramètre  $p_i = P(X_i = 1)$ . Le calcul de  $p_i$  se ramène au dénombrement de tous les échantillons possibles contenant le  $i^{\text{e}}$  objet défectueux. Un tel échantillon est constitué en prenant le  $i^{\text{e}}$  défectueux et en complétant par  $n - 1$  objets (défectueux ou non) choisis dans le reste de la population. D'où :

$$p_i = P(X_i = 1) = \frac{1 \times C_{N-1}^{n-1}}{C_N^n} = \frac{n}{N}.$$

Ainsi les  $X_i$  ont même loi et même espérance  $\mathbf{E}X_i = n/N$ . Par linéarité on en déduit :

$$\mathbf{E}X = \sum_{i=1}^M \mathbf{E}X_i = M \frac{n}{N}.$$

Remarquons que le résultat obtenu est le même que pour un prélèvement de  $n$  objets *avec remise* : dans ce cas, le nombre  $Y$  d'objets défectueux dans l'échantillon suit la loi binomiale  $\text{Bin}(n, M/N)$  et  $\mathbf{E}Y = nM/N$ .

**Exemple 6.34** (espérance de la loi gaussienne  $\mathfrak{N}(m, \sigma)$ ). Si  $X$  suit la loi gaussienne  $\mathfrak{N}(m, \sigma)$ ,  $X$  est intégrable et  $\mathbf{E}X = m$ . En effet, la v.a.  $Y = (X - m)/\sigma$  suit la loi  $\mathfrak{N}(0, 1)$  et  $X = \sigma Y + m$ . On sait déjà que  $Y$  est intégrable et que  $\mathbf{E}Y = 0$ , donc par linéarité, la constante  $m$  étant évidemment intégrable,  $\mathbf{E}X = \sigma \mathbf{E}Y + \mathbf{E}m = 0 + m = m$ .

**Proposition 6.35** (espérance et ordre).

- a) L'espérance des v.a. réelles intégrables est croissante : si  $X$  et  $Y$  v.a. réelles définies sur le même  $(\Omega, \mathcal{F}, P)$  sont intégrables et vérifient  $X \leq Y$ , c'est-à-dire pour tout  $\omega \in \Omega$ ,  $X(\omega) \leq Y(\omega)$ , alors  $\mathbf{E}X \leq \mathbf{E}Y$ .
- b) Si  $X$  est intégrable,  $|X|$  l'est aussi et

$$|\mathbf{E}X| \leq \mathbf{E}|X|. \quad (6.32)$$

*Preuve.* Pour le a), il suffit de noter que si  $X \leq Y$ ,  $Y - X$  est une v.a. positive et donc  $\mathbf{E}(Y - X) \geq 0$ . Comme  $X$  et  $Y$  sont intégrables, par linéarité de l'espérance,  $\mathbf{E}(Y - X) = \mathbf{E}Y - \mathbf{E}X$ . Ainsi la positivité de  $Y - X$  et la linéarité de  $\mathbf{E}$  impliquent  $\mathbf{E}Y - \mathbf{E}X \geq 0$  c'est-à-dire  $\mathbf{E}Y \geq \mathbf{E}X$ .

Une fois la croissance de  $\mathbf{E}$  ainsi établie, on en déduit le b) en partant de l'encadrement  $-|X| \leq X \leq |X|$  qui implique  $\mathbf{E}(-|X|) \leq \mathbf{E}X \leq \mathbf{E}|X|$ . Par linéarité  $\mathbf{E}(-|X|) = -\mathbf{E}|X|$ , ce qui permet de réécrire l'encadrement ci-dessus sous la forme :

$$-\mathbf{E}|X| \leq \mathbf{E}X \leq \mathbf{E}|X|,$$

qui équivaut à (6.32). □

## 6.4 Moments

On étudie dans cette section les  $\mathbf{E}h(X)$ , où  $h$  est une fonction réelle. Le cas le plus utile est celui où  $h$  est une fonction puissance,  $h(x) = x^r$ , on parle alors de moment d'ordre  $r$  de  $X$ .

**Définition 6.36.** *Soit  $r$  un réel positif. On appelle moment absolu d'ordre  $r$  de la variable aléatoire réelle  $X$  la quantité  $\mathbf{E}(|X|^r)$  (élément de  $\overline{\mathbb{R}}_+$ ). Si  $r$  est entier et  $X^r$  intégrable (donc si le moment absolu d'ordre  $r$  de  $X$  est fini), on appelle moment d'ordre  $r$  de  $X$  le réel  $\mathbf{E}(X^r)$ . On notera  $\mathbf{E}|X|^r$  pour  $\mathbf{E}(|X|^r)$  et  $\mathbf{E}X^r$  pour  $\mathbf{E}(X^r)$  en prenant garde de ne pas confondre ces quantités avec  $(\mathbf{E}|X|)^r$  et  $(\mathbf{E}X)^r$  respectivement.*

Remarquons qu'on ne définit pas le moment d'ordre  $r$  non entier pour  $X$  v.a. réelle, même si  $\mathbf{E}|X|^r < +\infty$ . En effet dans ce cas,  $X^r$  n'est pas définie sur l'évènement  $\{X < 0\}$  et si  $P(X < 0) \neq 0$ ,  $X^r$  ne peut être égale presque sûrement à une v.a. définie sur tout  $\Omega$ . Bien entendu, si  $X$  est une v.a. positive,  $\mathbf{E}X^r$  existe toujours dans  $\overline{\mathbb{R}}_+$ .

**Proposition 6.37.** *Si la variable aléatoire  $X$  a un moment absolu d'ordre  $r$  fini, elle a aussi un moment absolu d'ordre  $p$  fini pour tout  $p \in [0, r]$ .*

*Preuve.* Si  $0 \leq p \leq r$ , on a les inégalités suivantes entre variables aléatoires positives :

$$|X|^p \leq \mathbf{1}_{\{|X| \leq 1\}} + |X|^r \mathbf{1}_{\{|X| > 1\}} \leq 1 + |X|^r,$$

d'où

$$\mathbf{E}(|X|^p) \leq \mathbf{E}(1) + \mathbf{E}(|X|^r) = 1 + \mathbf{E}(|X|^r) < +\infty.$$

par croissance de l'espérance des v.a. positives.  $\square$

L'existence d'un moment absolu d'ordre  $r$  fini donne un renseignement sur la vitesse de convergence vers 0 de  $P(|X| \geq t)$  quand  $t$  tend vers  $+\infty$ . On a alors  $P(|X| \geq t) = O(t^{-r})$  par le corollaire suivant de l'inégalité de Markov.

**Proposition 6.38** (inégalité de Markov avec moment). *Pour toute variable aléatoire réelle  $X$ , pour tout réel  $r > 0$ ,*

$$\forall t > 0, \quad P(|X| \geq t) \leq \frac{\mathbf{E}|X|^r}{t^r}. \quad (6.33)$$

Bien entendu, cette inégalité n'a d'intérêt que si  $\mathbf{E}|X|^r < +\infty$  et  $t^{-r}\mathbf{E}|X|^r < 1$ .

*Preuve.* Il suffit de noter que  $P(|X| \geq t) \leq P(|X|^r \geq t^r) \leq t^{-r}\mathbf{E}|X|^r$ , en utilisant la croissance de l'application  $\mathbb{R}_+ \rightarrow \mathbb{R}_+$ ,  $x \mapsto x^r$  et l'inégalité de Markov pour la v.a. positive  $|X|^r$ .  $\square$

**Proposition 6.39** (moments d'une v.a. discrète). *Si  $X$  est une variable aléatoire discrète, pour tout réel  $r \geq 0$ ,*

$$\mathbf{E}|X|^r = \sum_{x \in X(\Omega)} |x|^r P(X = x). \quad (6.34)$$

De plus si  $r$  est entier et  $\mathbf{E}|X|^r < +\infty$ ,

$$\mathbf{E}X^r = \sum_{x \in X(\Omega)} x^r P(X = x). \quad (6.35)$$

Cette proposition n'est qu'une application de la formule de calcul de  $\mathbf{E}h(X)$  pour  $X$  discrète qui est établie dans toute sa généralité à la proposition 6.41 ci-dessous.

**Proposition 6.40** (moments d'une v.a. à densité). *Si  $X$  est une variable aléatoire réelle à densité  $f$ , pour tout réel  $r \geq 0$ ,*

$$\mathbf{E}|X|^r = \int_{-\infty}^{+\infty} |x|^r f(x) dx. \quad (6.36)$$

*Si de plus  $r$  est entier et  $\mathbf{E}|X|^r < +\infty$ ,*

$$\mathbf{E}X^r = \int_{-\infty}^{+\infty} x^r f(x) dx. \quad (6.37)$$

Là aussi, il s'agit d'un cas particulier d'une formule générale pour  $\mathbf{E}h(X)$  lorsque  $X$  est à densité, donnée ci-dessous. Nous invitons le lecteur à démontrer directement en exercice la proposition 6.40. La preuve est grandement facilitée par la monotonie de  $h : x \mapsto |x|^r$  sur chacun des ensembles  $\mathbb{R}_-$  et  $\mathbb{R}_+$ .

**Proposition 6.41** (calcul de  $\mathbf{E}h(X)$ ,  $X$  discrète). *Si  $X$  est une variable aléatoire discrète et  $h : \mathbb{R} \rightarrow \mathbb{R}$  une application borélienne, c.-à-d. mesurable  $\text{Bor}(\mathbb{R})$ - $\text{Bor}(\mathbb{R})$ ,*

$$\mathbf{E}|h(X)| = \sum_{x \in X(\Omega)} |h(x)| P(X = x). \quad (6.38)$$

*De plus, si  $\mathbf{E}|h(X)| < +\infty$ ,*

$$\mathbf{E}h(X) = \sum_{x \in X(\Omega)} h(x) P(X = x). \quad (6.39)$$

*Preuve.* Posons  $Y = |h(X)|$ . L'ensemble  $X(\Omega) = \{x_0, x_1, \dots\}$  est au plus dénombrable. L'application  $|h|$  n'étant pas supposée injective, il peut y avoir des répétitions dans la suite des  $|h(x_k)|$ . L'ensemble  $Y(\Omega)$  qui peut s'écrire en effaçant toutes les répétitions de cette suite est lui-même au plus dénombrable. La variable aléatoire  $Y = |h(X)|$  étant discrète positive, la formule établie au corollaire 6.11 nous donne :

$$\mathbf{E}|h(X)| = \mathbf{E}Y = \sum_{y \in Y(\Omega)} y P(Y = y). \quad (6.40)$$

Pour chaque réel  $y$  de  $Y(\Omega)$ , notons  $B_y$  l'ensemble de ses antécédents par  $|h|$  :

$$B_y = \{x \in X(\Omega) ; |h(x)| = y\}, \quad y \in Y(\Omega).$$

Ce sous-ensemble de  $X(\Omega)$  contient au moins un élément et est au plus dénombrable. On peut alors décomposer l'évènement  $\{Y = y\}$  en réunion disjointe d'une famille au plus dénombrable :

$$\{Y = y\} = \bigcup_{x \in B_y} \{X = x\}.$$

Le terme général de la somme (6.40) peut donc s'écrire :

$$y P(Y = y) = y \sum_{x \in B_y} P(X = x) = \sum_{x \in B_y} y P(X = x) = \sum_{x \in B_y} |h(x)| P(X = x).$$

Comme les  $B_y$  forment une partition de  $X(\Omega)$ , la propriété de sommation par paquets des familles à termes *positifs* nous donne :

$$\sum_{x \in X(\Omega)} |h(x)|P(X = x) = \sum_{y \in Y(\Omega)} \sum_{x \in B_y} |h(x)|P(X = x) = \sum_{y \in Y(\Omega)} yP(Y = y),$$

ce qui, compte-tenu de (6.40), établit la formule (6.38) pour  $\mathbf{E}|h(X)|$ .

Supposons maintenant que  $\mathbf{E}|h(X)| < +\infty$ . Grâce à (6.38), la famille de réels  $\{h(x)P(X = x); x \in X(\Omega)\}$  est *sommable*. On peut alors reprendre le calcul fait ci-dessus en remplaçant  $|h|$  par  $h$  et utiliser la propriété de sommation par paquets des familles sommables pour aboutir à la formule (6.39) de calcul de  $\mathbf{E}h(X)$ .  $\square$

**Proposition 6.42** (calcul de  $\mathbf{E}h(X)$ ,  $X$  à densité). *Si  $X$  est une variable aléatoire de densité  $f$  et  $h : \mathbb{R} \rightarrow \mathbb{R}$  une application réglée sur tout intervalle fermé borné de  $\mathbb{R}$ ,*

$$\mathbf{E}|h(X)| = \int_{-\infty}^{+\infty} |h(x)|f(x) dx. \quad (6.41)$$

De plus, si  $\mathbf{E}|h(X)| < +\infty$ ,

$$\mathbf{E}h(X) = \int_{-\infty}^{+\infty} h(x)f(x) dx. \quad (6.42)$$

Nous admettrons cette proposition. Une application  $h$  est réglée sur  $[a, b]$  si elle est limite uniforme sur  $[a, b]$  d'une suite de fonctions en escaliers. On démontre en analyse que  $h$  est réglée sur  $[a, b]$  si et seulement si elle admet en tout point de  $]a, b[$  une limite à gauche et une limite à droite (finies) ainsi qu'une limite finie à droite en  $a$  et à gauche en  $b$ . La classe des fonctions réglées, devrait donc suffire à nos besoins. Elle contient en particulier les fonctions continues et les fonctions monotones par morceaux.

Nous présentons maintenant une formule injustement méconnue, permettant de calculer des moments fonctionnels à partir de la fonction de survie. Son intérêt est de permettre un tel calcul pour des variables aléatoires qui ne sont ni discrètes ni à densité.

**Proposition 6.43.** *Soient  $X$  une variable aléatoire positive et  $g$  une application continue strictement croissante  $\mathbb{R}_+ \rightarrow \mathbb{R}$ , de classe  $C^1$  sur  $\mathbb{R}_+^*$ . Alors*

$$\mathbf{E}g(X) = g(0) + \int_0^{+\infty} P(X > s)g'(s) ds. \quad (6.43)$$

Le  $h$ -moment  $\mathbf{E}h(X)$  pour  $h : x \mapsto (x - \mathbf{E}X)^2$  occupe une place particulière dans la théorie des probabilités.

**Définition 6.44** (variance et écart type). *Si  $X$  est de carré intégrable ( $\mathbf{E}X^2 < +\infty$ ), on appelle variance de  $X$  le réel positif noté  $\text{Var } X$  défini par*

$$\text{Var } X := \mathbf{E}(X - \mathbf{E}X)^2. \quad (6.44)$$

On appelle alors écart type de  $X$  le réel  $\sigma(X) := (\text{Var } X)^{1/2}$ .

Remarquons que si  $\mathbf{E}X^2$  est fini,  $\mathbf{E}|X|$  l'est aussi (proposition 6.37), donc  $\mathbf{E}X$  est bien défini. De plus  $(X - \mathbf{E}X)^2 = X^2 - 2(\mathbf{E}X)X + (\mathbf{E}X)^2$  apparaît alors comme une combinaison linéaire de trois variables<sup>10</sup> intégrables, donc est aussi intégrable. Ainsi la

10. À savoir  $X^2$ ,  $X$  et la v.a. constante  $(\mathbf{E}X)^2$ .



v.a. positive  $(X - \mathbf{E}X)^2$  est intégrable et  $\mathbf{E}(X - \mathbf{E}X)^2$  est bien un réel positif, ce qui justifie la définition 6.44. Notons aussi que si  $X$  représente une grandeur physique,  $X$ ,  $\mathbf{E}X$  et  $\sigma(X)$  ont la même unité, mais pas  $\text{Var } X$ .

Lorsqu'elle existe, la variance de  $X$  est une façon de mesurer la *dispersion* de la loi de  $X$  autour de l'espérance. Les raisons de l'importance de la variance apparaîtront ultérieurement dans ce cours (inégalité de Tchebycheff, théorème limite central). L'application des propositions 6.41 et 6.42 nous donne (sous réserve d'intégrabilité de  $X^2$ ) les formules respectives :

$$\text{Var } X = \sum_{x \in X(\Omega)} (x - \mathbf{E}X)^2 P(X = x) \quad (\text{cas } X \text{ discrète}), \quad (6.45)$$

$$\text{Var } X = \int_{-\infty}^{+\infty} (x - \mathbf{E}X)^2 f(x) dx \quad (\text{cas } X \text{ à densité } f). \quad (6.46)$$

Dans la pratique, ces formules sont rarement utilisées, on leur préfère la formule suivante qui simplifie les calculs.

**Proposition 6.45** (formule de Koenig pour la variance). *Si la variable aléatoire  $X$  est de carré intégrable,*

$$\text{Var } X = \mathbf{E}X^2 - (\mathbf{E}X)^2. \quad (6.47)$$

*Preuve.* Rappelons que nous notons  $\mathbf{E}X^2$  pour  $\mathbf{E}(X^2)$  et que le second membre de la formule ci-dessus n'est donc généralement pas nul. On pose  $c = \mathbf{E}X$ .

$$\begin{aligned} \text{Var } X &= \mathbf{E}(X - c)^2 = \mathbf{E}(X^2 - 2cX + c^2) \\ &= \mathbf{E}X^2 - 2c\mathbf{E}X + \mathbf{E}c^2 \\ &= \mathbf{E}X^2 - 2c^2 + c^2 = \mathbf{E}X^2 - c^2, \end{aligned}$$

en utilisant la linéarité de l'espérance et l'espérance d'une constante. □

**Proposition 6.46** (translation et changement d'échelle). *Si  $\mathbf{E}X^2 < +\infty$ ,*

$$\forall a \in \mathbb{R}, \forall b \in \mathbb{R}, \quad \text{Var}(aX + b) = a^2 \text{Var } X, \quad \sigma(aX + b) = |a|\sigma(X). \quad (6.48)$$

*Preuve.* En utilisant la définition 6.44, la linéarité de l'espérance et le fait que l'espérance de la constante  $b$  est  $b$  :

$$\begin{aligned} \text{Var}(aX + b) &= \mathbf{E}(aX + b - \mathbf{E}(aX + b))^2 = \mathbf{E}(aX + b - a\mathbf{E}X - b)^2 \\ &= \mathbf{E}(a(X - \mathbf{E}X))^2 = \mathbf{E}(a^2(X - \mathbf{E}X)^2) \\ &= a^2\mathbf{E}((X - \mathbf{E}X)^2) = a^2 \text{Var } X, \end{aligned}$$

ce qui nous donne les formules (6.48). □

Il est clair, d'après la définition de la variance, que la variance d'une constante est nulle. La réciproque est *presque* vraie :

**Proposition 6.47** (nullité de la variance et constance p.s.).

$$\text{Var } X = 0 \Leftrightarrow X = \mathbf{E}X \text{ p.s.} \Leftrightarrow X \text{ est presque sûrement constante.} \quad (6.49)$$

*Preuve.* Les implications de droite à gauche dans (6.49) sont déjà acquises. On sait en effet que l'espérance d'une constante est cette constante et que si la v.a.  $Y := (X - \mathbf{E}X)^2$  vaut 0 avec probabilité 1, son espérance est nulle<sup>11</sup>.

Pour la première implication de gauche à droite, il suffit d'appliquer le corollaire 6.23 à la v.a. positive  $Y$ . La deuxième implication est triviale.  $\square$

Voyons maintenant quelques calculs de variance de lois classiques.

**Exemple 6.48** (Variance d'une loi de Bernoulli). La variance d'une loi de Bernoulli de paramètre  $p$  est  $p(1-p) = pq$ . En effet si  $X$  suit la loi  $\text{Bern}(p)$ , on a  $\mathbf{E}X = p$  et  $\mathbf{E}X^2 = 0^2 \times q + 1^2 \times p = p$  (avec  $q = 1-p$ ). D'où :

$$\text{Var } X = \mathbf{E}X^2 - (\mathbf{E}X)^2 = p - p^2 = p(1-p).$$

**Exemple 6.49** (variance de la loi uniforme sur  $\{1, \dots, n\}$ ). Si  $X$  suit la loi uniforme sur  $\{1, \dots, n\}$ ,  $\text{Var } X = \frac{n^2 - 1}{12}$ .

Dans ce cas,  $X(\Omega) = \{1, \dots, n\}$  et pour tout entier  $k$  compris entre 1 et  $n$ ,  $P(X = k) = 1/n$ . On calcule alors  $\mathbf{E}X$  et  $\mathbf{E}X^2$  :

$$\mathbf{E}X = \sum_{k=1}^n kP(X = k) = \frac{1}{n} \sum_{k=1}^n k = \frac{1}{n} \frac{n(n+1)}{2} = \frac{n+1}{2}.$$

$$\begin{aligned} \mathbf{E}X^2 &= \sum_{k=1}^n k^2 P(X = k) = \frac{1}{n} \sum_{k=1}^n k^2 \\ &= \frac{1}{n} \frac{n(n+1)(2n+1)}{6} = \frac{(n+1)(2n+1)}{6}. \end{aligned}$$

En appliquant la formule de Koenig, on obtient :

$$\text{Var } X = \frac{(n+1)(2n+1)}{6} - \frac{(n+1)^2}{4} = \frac{(n+1)(n-1)}{12}.$$

**Exemple 6.50** (variance d'une loi binomiale). Si la variable aléatoire  $X$  suit la loi  $\text{Bin}(n, p)$ ,  $\text{Var } X = np(1-p) = npq$ . On peut démontrer ce résultat par la formule de Koenig (exercice) ou le déduire d'une formule générale sur la variance d'une somme de v.a. indépendantes.

**Exemple 6.51** (variance d'une loi de Poisson). Si  $X$  suit la loi de Poisson  $\text{Pois}(\alpha)$ ,  $\text{Var } X = \alpha$ . Il est facile de vérifier (exercice) qu'une loi de Poisson possède des moments de tout ordre donc a fortiori une variance. On sait déjà que  $\mathbf{E}X = \alpha$ , il nous reste à calculer  $\mathbf{E}X^2$  :

$$\begin{aligned} \mathbf{E}X^2 &= \sum_{k=0}^{+\infty} k^2 \frac{e^{-\alpha} \alpha^k}{k!} = \sum_{k=1}^{+\infty} k \frac{e^{-\alpha} \alpha^k}{(k-1)!} \\ &= e^{-\alpha} \sum_{k=1}^{+\infty} ((k-1) + 1) \frac{\alpha^k}{(k-1)!} \\ &= \alpha^2 e^{-\alpha} \sum_{k=2}^{+\infty} \frac{\alpha^{k-2}}{(k-2)!} + \alpha e^{-\alpha} \sum_{k=1}^{+\infty} \frac{\alpha^{k-1}}{(k-1)!} \\ &= \alpha^2 + \alpha. \end{aligned}$$

D'où  $\text{Var } X = \mathbf{E}X^2 - (\mathbf{E}X)^2 = (\alpha^2 + \alpha) - \alpha^2 = \alpha$ .

11. En effet, pour tout  $t \geq 0$ ,  $P(Y > t) = 0$ , donc  $Y$  étant positive,  $\mathbf{E}Y = \int_0^{+\infty} P(Y > t) dt = 0$ .

**Exemple 6.52** (variance d'une loi géométrique). Si  $X$  suit la loi géométrique de paramètre  $p$ ,  $\text{Var } X = \frac{1-p}{p^2}$ . La vérification est laissée en exercice.

**Exemple 6.53** (variance d'une loi gaussienne). Si  $X$  suit la loi  $\mathfrak{N}(m, \sigma)$ ,  $\text{Var } X = \sigma^2$ . On sait en effet que l'on peut écrire  $X = \sigma Y + m$  avec  $Y$  et loi  $\mathfrak{N}(0, 1)$ . L'intégrabilité de  $Y^2$  se vérifie par une méthode analogue à celle utilisée dans l'exemple 6.34 (faites-le!). Ceci étant acquis, d'après (6.48),  $\text{Var } x = \sigma^2 \text{Var } Y$  et comme  $\mathbf{E}Y = 0$ , cf. exemple 6.34,  $\text{Var } Y = \mathbf{E}Y^2$ . il nous reste donc à vérifier que

$$\mathbf{E}Y^2 = \int_{-\infty}^{+\infty} \frac{y^2}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right) dy = 1.$$

Une façon de le voir est d'intégrer par parties la densité  $f_{0,1}$ . Pour cela considérons pour  $a > 0$  l'intégrale

$$I(a) := \int_{-a}^a \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right) dy$$

et posons  $u = \exp(-y^2/2)$ ,  $dv = (2\pi)^{-1/2} dy$ ,  $du = -y \exp(-y^2/2) dy$  et  $v = (2\pi)^{-1/2} y$ . Il vient

$$\begin{aligned} I(a) &= \left[ \frac{y}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right) \right]_{-a}^a - \int_{-a}^a \frac{-y^2}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right) dy \\ &= \frac{2a}{\sqrt{2\pi}} \exp\left(-\frac{a^2}{2}\right) + \int_{-a}^a \frac{y^2}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right) dy. \end{aligned}$$

En faisant tendre  $a$  vers l'infini<sup>12</sup> dans cette égalité, on obtient

$$1 = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right) dy = \int_{-\infty}^{+\infty} \frac{y^2}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right) dy.$$

## 6.5 Théorèmes limite

Nous allons voir maintenant comment l'outillage contenu dans ce chapitre permet d'aborder deux théorèmes importants donnant le comportement asymptotique des sommes de variables aléatoires indépendantes : une loi des grands nombres et le théorème limite central.

### 6.5.1 L'inégalité de Tchebycheff

Une des raisons de l'importance de la variance est le théorème suivant.

**Théorème 6.54** (Inégalité de Tchebycheff). *Si  $\text{Var } X$  existe,*

$$\forall t > 0, \quad P(|X - \mathbf{E}X| \geq t) \leq \frac{\text{Var } X}{t^2}. \quad (6.50)$$

*Preuve.* Il suffit d'appliquer le corollaire de l'inégalité de Markov avec moment, prop. 6.38 avec  $r = 2$ , à la v.a.  $Y = X - \mathbf{E}X$ .  $\square$

<sup>12</sup>. Pour aller plus vite, on a intégré sur un intervalle symétrique  $[-a, +a]$  parce que l'on sait déjà que les intégrales généralisées concernées sont convergentes. Si l'on voulait se servir de ce calcul pour montrer leur convergence, il faudrait bien sûr intégrer sur un intervalle  $[-a, b]$  et faire tendre  $a$  et  $b$  séparément vers  $+\infty$ .

**Remarque 6.55.** Si on pose  $t = u\sigma(X)$  dans (6.50), on obtient une nouvelle forme de l'inégalité de Tchebycheff :

$$\forall u > 0, \quad P(|X - \mathbf{E}X| \geq u\sigma(X)) \leq \frac{1}{u^2}. \quad (6.51)$$

Sous cette forme, le majorant obtenu est indépendant de la loi de  $X$ . Ceci permet de comprendre pourquoi  $\sigma(X)$  s'appelle *écart type* ou *unité d'écart*. Pour toute loi de probabilité ayant un moment d'ordre 2, la probabilité d'observer une déviation par rapport à l'espérance d'au moins  $u$  unités d'écart est majorée par  $u^{-2}$ .

**Exemple 6.56.** On jette 3600 fois un dé. Minorer la probabilité que le nombre d'apparitions du 1 soit compris strictement entre 480 et 720.

Notons  $S$  le nombre d'apparitions du 1. Cette variable aléatoire suit la loi binomiale  $\text{Bin}(3600, 1/6)$ . La valeur exacte de la probabilité qui nous intéresse est :

$$P(480 < S < 720) = \sum_{k=481}^{719} C_{3600}^k \left(\frac{1}{6}\right)^k \left(\frac{5}{6}\right)^{3600-k}.$$

Le calcul de la valeur exacte de cette somme nécessiterait l'écriture d'un programme et une certaine puissance de calcul informatique. L'inégalité de Tchebycheff est une alternative pratique à un calcul aussi déraisonnable. En effet on a :

$$\mathbf{E}S = 3600 \times \frac{1}{6} = 600, \quad \text{Var } S = 3600 \times \frac{1}{6} \times \frac{5}{6} = 500$$

et on remarque que  $480 - 600 = -120$ ,  $720 - 600 = 120$  d'où :

$$480 < S < 720 \Leftrightarrow -120 < S - 600 < +120 \Leftrightarrow |S - 600| < 120.$$

L'inégalité de Tchebycheff s'écrit ici :

$$\forall t > 0, \quad P(|S - 600| \geq t) \leq \frac{500}{t^2}.$$

En particulier pour  $t = 120$  :

$$P(|S - 600| \geq 120) \leq \frac{500}{120^2} = \frac{5}{144}.$$

D'où en passant à l'événement complémentaire :

$$P(480 < S < 720) = P(|S - 600| < 120) \geq 1 - \frac{5}{144} \geq 0.965.$$

On peut aussi utiliser l'inégalité de Tchebycheff avec un intervalle non centré sur l'espérance. Il suffit alors d'utiliser le plus grand intervalle centré qu'il contient. Ainsi pour minorer  $P(550 < S < 700)$ , il suffit de remarquer que

$$550 < S < 700 \Leftrightarrow 550 < S < 650,$$

d'où :

$$\{550 < S < 700\} \supset \{550 < S < 650\}$$

et d'appliquer l'inégalité de Tchebycheff avec  $t = 50$ . Bien sûr le résultat obtenu sera moins bon.

On peut aussi utiliser l'inégalité de Tchebycheff avec un intervalle unilatéral (i.e. du type  $[a, +\infty[$  ou  $] - \infty, a]$ ).

**Exemple 6.57.** Avec les notations de l'exemple précédent, proposer une valeur de  $u$  telle que  $P(S \geq u) \leq 0.05$ .

Comme la v.a.  $S$  est bornée par 3600, la réponse n'a d'intérêt que pour  $u < 3600$ . Une première idée est d'essayer l'inégalité de Markov qui s'écrit ici :

$$P(S \geq u) \leq \frac{\mathbf{E}S}{u} = \frac{600}{u}.$$

Par cette inégalité, il *suffirait* de prendre  $u$  tel que  $600/u \leq 0.05$ . Malheureusement, la plus petite solution de cette inéquation est  $u_0 = 600/0.05 = 12000$  et le résultat est sans intérêt. Pour utiliser l'inégalité de Tchebycheff, on commence par remarquer que pour tout  $t$  positif, l'inégalité  $S - \mathbf{E}S \geq t$  implique  $|S - \mathbf{E}S| \geq t$ , ce qui se traduit par l'inclusion d'événements :

$$\{S - \mathbf{E}S \geq t\} \subset \{|S - \mathbf{E}S| \geq t\}.$$

On en déduit :

$$P(S \geq t + \mathbf{E}S) = P(S - \mathbf{E}S \geq t) \leq P(|S - \mathbf{E}S| \geq t) \leq \frac{500}{t^2}.$$

Il *suffit* alors de choisir la plus petite valeur de  $t$  telle que  $500/t^2 \leq 0.05$  soit  $t_1 = 100$ . La valeur correspondante pour  $u$  étant  $u_1 = t_1 + \mathbf{E}S = 700$ .

## 6.5.2 Covariance

Le calcul de la variance de la somme de deux variables aléatoires de carré intégrable  $X$  et  $Y$  nous amène à introduire la covariance de  $(X, Y)$ .

**Proposition 6.58** (inégalité de Cauchy-Schwarz). *Si les variables aléatoires réelles  $X$  et  $Y$  ont des moments d'ordre 2, alors la variable aléatoire  $XY$  est intégrable et*

$$|\mathbf{E}(XY)| \leq (\mathbf{E}X^2)^{1/2}(\mathbf{E}Y^2)^{1/2}. \quad (6.52)$$

*Preuve.* L'intégrabilité de la v.a.  $XY$  résulte de l'inégalité élémentaire  $|XY| \leq X^2 + Y^2$ . On remarque alors que la fonction trinômiale suivante de la variable réelle  $t$ ,

$$g : t \longmapsto t^2\mathbf{E}Y^2 + 2t\mathbf{E}(XY) + \mathbf{E}X^2 = \mathbf{E}(X + tY)^2,$$

est définie sur  $\mathbb{R}$  et positive sur  $\mathbb{R}$ . Ceci n'est possible que si son discriminant est négatif, ce qui s'écrit

$$\Delta' = (\mathbf{E}(XY))^2 - (\mathbf{E}X^2)(\mathbf{E}Y^2) \leq 0,$$

d'où l'inégalité (6.52). □

**Définition 6.59** (covariance). *Si les variables aléatoires réelles  $X$  et  $Y$  ont des moments d'ordre 2, on appelle covariance du couple aléatoire  $(X, Y)$  la quantité :*

$$\text{Cov}(X, Y) = \mathbf{E}[(X - \mathbf{E}X)(Y - \mathbf{E}Y)].$$

Remarquons que  $\text{Cov}(X, X) = \text{Var } X$ .

**Proposition 6.60** (propriétés de la covariance). *Les propriétés suivantes sont vérifiées pour tout couple  $(X, Y)$  de v.a. réelles ayant des moments d'ordre 2.*

- (i)  $\text{Cov}(X, Y) = \text{Cov}(Y, X)$ .
- (ii) Pour tous réels  $a, b, c, d$  :  $\text{Cov}(aX + b, cY + d) = ac \text{Cov}(X, Y)$ .
- (iii)  $|\text{Cov}(X, Y)| \leq \sigma(X)\sigma(Y)$ .

La vérification est laissée au lecteur.

**Définition 6.61** (coefficient de corrélation). *Si  $X$  et  $Y$  sont des variables aléatoires réelles non constantes ayant des moments d'ordre 2, on appelle coefficient de corrélation entre  $X$  et  $Y$  la quantité :*

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma(X)\sigma(Y)}.$$

D'après (iii) on a toujours  $-1 \leq \rho(X, Y) \leq 1$ . D'autre part il résulte facilement du cas d'égalité dans l'inégalité de Cauchy-Schwarz que  $|\rho|$  est maximal lorsque  $Y$  est une fonction affine de  $X$  :  $Y = aX + b$ . Quand  $\rho = 0$ , ce qui arrive en particulier lorsque  $X$  et  $Y$  sont indépendantes, on dit que  $X$  et  $Y$  sont *non corrélées*.

**Proposition 6.62** (formule de Koenig pour la covariance). *Si la covariance de  $X$  et  $Y$  existe, elle peut se calculer par :*

$$\text{Cov}(X, Y) = \mathbf{E}(XY) - \mathbf{E}X\mathbf{E}Y.$$

*Preuve.* La vérification est analogue à celle de la formule de Koenig pour la variance (qui n'est que le cas particulier  $Y = X$ ) et est laissée en exercice.  $\square$

**Proposition 6.63** (variance d'une somme). *Si  $X_1, \dots, X_n$  sont des variables aléatoires réelles ayant des moments d'ordre 2,*

$$\text{Var}\left(\sum_{i=1}^n X_i\right) = \sum_{i,j=1}^n \text{Cov}(X_i, X_j) \quad (6.53)$$

$$= \sum_{i=1}^n \text{Var} X_i + \sum_{\substack{i,j=1 \\ i \neq j}}^n \text{Cov}(X_i, X_j). \quad (6.54)$$

Dans le cas  $n = 2$  (6.54) s'écrit :

$$\text{Var}(X + Y) = \text{Var} X + \text{Var} Y + 2 \text{Cov}(X, Y). \quad (6.55)$$

*Preuve.* Pour  $n \geq 2$  quelconque, l'identité algébrique :

$$\left(\sum_{i=1}^n Y_i\right)^2 = \sum_{i,j=1}^n Y_i Y_j,$$

utilisée avec  $Y_i = X_i - \mathbf{E}X_i$  et la linéarité de l'espérance nous donnent :

$$\begin{aligned} \text{Var}\left(\sum_{i=1}^n X_i\right) &= \mathbf{E}\left\{\sum_{i=1}^n X_i - \mathbf{E}\left(\sum_{i=1}^n X_i\right)\right\}^2 \\ &= \mathbf{E}\left\{\sum_{i=1}^n X_i - \sum_{i=1}^n \mathbf{E}X_i\right\}^2 \\ &= \mathbf{E}\left\{\sum_{i=1}^n Y_i\right\}^2 \\ &= \sum_{i,j=1}^n \mathbf{E}(Y_i Y_j) = \sum_{i,j=1}^n \text{Cov}(X_i, X_j). \end{aligned}$$

□

### 6.5.3 Variables aléatoires indépendantes

**Définition 6.64.** On dit que les variables aléatoires  $X_1, \dots, X_n$  définies sur le même espace probabilisé  $(\omega, \mathcal{F}, P)$  sont indépendantes si pour tout choix d'intervalles  $I_1, \dots, I_n$  de  $\mathbb{R}$ , les  $n$  événements  $\{X_1 \in I_1\}, \dots, \{X_n \in I_n\}$  sont indépendants.

**Théorème 6.65.** Si  $X$  et  $Y$  sont indépendantes et intégrables,  $XY$  est intégrable et

$$\mathbf{E}(XY) = \mathbf{E}X\mathbf{E}Y.$$

**Corollaire 6.66.** Si  $X$  et  $Y$  sont indépendantes et de carré intégrable,  $\text{Cov}(X, Y) = 0$ .

### 6.5.4 Loi faible des grands nombres

Il s'agit d'étudier la convergence des moyennes arithmétiques  $S_n/n$  construites à partir d'une suite de variables aléatoires  $(X_i)_{i \geq 1}$  en posant  $S_n := X_1 + \dots + X_n$ . Si  $S_n/n$  converge en probabilité, on parle d'une loi faible des grands nombres.

**Définition 6.67** (Convergence en probabilité). La suite  $Y_n$  converge en probabilité vers  $Y$  (notation  $Y_n \xrightarrow[n \rightarrow +\infty]{\text{Pr}} Y$ ) si

$$\forall \varepsilon > 0, \quad P(|Y_n - Y| \geq \varepsilon) \xrightarrow[n \rightarrow +\infty]{} 0.$$

Rappelons que si  $X$  est de carré intégrable ( $\mathbf{E}X^2 < +\infty$ ), sa variance est définie par

$$\text{Var } X := \mathbf{E}(X - \mathbf{E}X)^2.$$

Les  $X_k$  sont dites deux à deux *non-corrélées* si  $\text{Cov}(X_i, X_j) = 0$  pour tous  $i, j$  distincts. Ceci se produit en particulier lorsque les  $X_k$  sont deux à deux indépendantes, cf. cor. 6.66. Pour des  $X_k$  deux à deux non-corrélées, on déduit de (6.54) l'égalité

$$\text{Var } S_n = \sum_{k=1}^n \text{Var } X_k. \quad (6.56)$$

**Proposition 6.68** (inégalité de Bienaymé-Tchebycheff). Si les  $X_k$  sont de carré intégrable et deux à deux non-corrélées,

$$\forall t > 0, \quad P\left(\left|\sum_{k=1}^n (X_k - \mathbf{E}X_k)\right| \geq t\right) \leq \frac{1}{t^2} \sum_{k=1}^n \text{Var } X_k. \quad (6.57)$$

*Preuve.* Il suffit d'écrire :

$$\begin{aligned} P(|S_n - \mathbf{E}S_n| \geq t) &= P(|S_n - \mathbf{E}S_n|^2 \geq t^2) \leq \frac{1}{t^2} \mathbf{E}(S_n - \mathbf{E}S_n)^2 \\ &= \frac{1}{t^2} \text{Var } S_n, \end{aligned} \quad (6.58)$$

où l'inégalité dans (6.58) est l'inégalité de Markov appliquée à la variable aléatoire positive  $|S_n - \mathbf{E}S_n|^2$ . On conclut avec (6.56). □

**Théorème 6.69** (loi faible des grands nombres). *Si les  $X_k$  sont de même loi, de carré intégrable et deux à deux non-corrélées, on a la convergence en probabilité :*

$$\frac{1}{n} \sum_{k=1}^n X_k \xrightarrow[n \rightarrow +\infty]{\text{Pr}} \mathbf{E}X_1. \quad (6.59)$$

*Preuve.* Comme les  $X_k$  ont même loi, on a pour tout  $k$  les égalités  $\mathbf{E}X_k = \mathbf{E}X_1$  et  $\text{Var } X_k = \text{Var } X_1$ . Comme elles sont aussi deux à deux non-corrélées, (6.56) nous donne  $\text{Var } S_n = n \text{Var } X_1$ . Par linéarité de l'espérance on a aussi  $\mathbf{E}S_n = n\mathbf{E}X_1$ . L'inégalité de Bienaymé-Tchebycheff nous dit alors que :

$$\forall t > 0, \quad P(|S_n - \mathbf{E}S_n| \geq t) = P(|S_n - n\mathbf{E}X_1| \geq t) \leq \frac{n \text{Var } X_1}{t^2}.$$

Posant  $t = n\varepsilon$ , on en déduit :

$$\forall \varepsilon > 0, \forall n \in \mathbb{N}^*, \quad P(|S_n - n\mathbf{E}X_1| \geq n\varepsilon) = P\left(\left|\frac{S_n}{n} - \mathbf{E}X_1\right| \geq \varepsilon\right) \leq \frac{n \text{Var } X_1}{n^2\varepsilon^2} = \frac{\text{Var } X_1}{n\varepsilon^2}.$$

Pour tout  $\varepsilon > 0$  fixé, on a ainsi

$$P\left(\left|\frac{S_n}{n} - \mathbf{E}X_1\right| \geq \varepsilon\right) \leq \frac{\text{Var } X_1}{n\varepsilon^2} \xrightarrow[n \rightarrow +\infty]{} 0,$$

ce qui établit la convergence en probabilité de la suite de variables aléatoires  $S_n/n$  vers la variable aléatoire constante  $\mathbf{E}X_1$ .  $\square$

### 6.5.5 Théorème limite central

**Théorème 6.70** (théorème limite central). *Soit  $(X_k)_{k \geq 1}$  une suite de variables aléatoires définies sur le même espace probabilisé  $(\Omega, \mathcal{F}, P)$ , indépendantes, de même loi et de carré intégrable et non p.s. constantes. Notons  $\mu := \mathbf{E}X_1$ ,  $\sigma^2 := \text{Var } X_1$  avec  $\sigma > 0$  et  $S_n = \sum_{k=1}^n X_k$ . Posons*

$$S_n^* := \frac{S_n - \mathbf{E}S_n}{\sqrt{\text{Var } S_n}} = \frac{S_n - n\mu}{\sigma\sqrt{n}}. \quad (6.60)$$

Alors

$$\forall x \in \mathbb{R}, \quad P(S_n^* \leq x) \xrightarrow[n \rightarrow +\infty]{} \Phi(x) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x \exp\left(-\frac{t^2}{2}\right) dt. \quad (6.61)$$

Une conséquence pratique de (6.61) est que pour tous réels  $a, b$ , tels que  $a < b$ ,

$$P(S_n^* \in I(a, b)) \xrightarrow[n \rightarrow +\infty]{} \Phi(b) - \Phi(a) = \frac{1}{\sqrt{2\pi}} \int_a^b \exp\left(-\frac{t^2}{2}\right) dt, \quad (6.62)$$

où  $I(a, b)$  est n'importe lequel des quatre intervalles d'extrémités  $a$  et  $b$ .

**Corollaire 6.71** (théorème de de Moivre-Laplace). *Si  $S_n$  est une variable aléatoire de loi binomiale de paramètres  $n$  et  $p \in ]0, 1[$ , notons avec  $q := 1 - p$ ,*

$$S_n^* := \frac{S_n - np}{\sqrt{npq}} = \sqrt{\frac{n}{pq}} \left( \frac{S_n}{n} - p \right).$$

Alors les convergences (6.61) et (6.62) ont lieu.



*Preuve.* C'est une conséquence immédiate du théorème 6.70 en remarquant que  $S_n$  a même loi que  $X_1 + \dots + X_n$ , 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  $\text{Bin}(n, p)$  sont respectivement  $np$  et  $npq$ .  $\square$

**Exemple 6.72.** On lance 3600 fois un dé. Évaluer la probabilité que le nombre d'apparitions du 1 soit compris strictement entre 540 et 660.

Soit  $S$  le nombre d'apparitions du 1. Cette variable aléatoire suit la loi binomiale  $\text{Bin}(3600, 1/6)$ . On a  $\mathbf{E}S = 600$  et  $\text{Var } S = 500$ . En notant  $S^* = (S - \mathbf{E}S)/\sigma(S)$  la variable centrée réduite associée à  $S$  :

$$P(540 < S < 660) = P\left(\frac{540 - 600}{\sqrt{500}} < S^* < \frac{660 - 600}{\sqrt{500}}\right).$$

Comme  $n$  est grand on peut utiliser l'approximation liée au théorème de De Moivre-Laplace :

$$P\left(\frac{-60}{10\sqrt{5}} < S^* < \frac{60}{10\sqrt{5}}\right) \simeq \Phi\left(\frac{6}{\sqrt{5}}\right) - \Phi\left(\frac{-6}{\sqrt{5}}\right).$$

En utilisant la parité de la densité  $f_{0,1}$ , on a pour tout  $a > 0$  :  $\Phi(a) - \Phi(-a) = 2\Phi(a) - 1$ . En utilisant la table des valeurs de  $\Phi$  on obtient donc :

$$P(540 < S < 660) \simeq 2\Phi(2,68) - 1 \simeq 0,9926 \simeq 0,99.$$

(en raison de l'approximation de  $6/\sqrt{5}$  par 2,68, il n'y a pas lieu de conserver les deux derniers chiffres).

Il est intéressant de comparer ce résultat avec ce qu'aurait donné l'inégalité de Tchebycheff. En reprenant les calculs de l'exemple 6.56, on obtient :

$$P(540 < S < 660) \geq 1 - \frac{500}{60^2} = \frac{31}{36} > 0,86.$$

Pour être honnête, il convient néanmoins de noter que les deux résultats sont de nature différente. Le deuxième est une minoration dont on est certain, pas une approximation. Pour pouvoir affirmer que le théorème de De Moivre-Laplace nous donne ici un meilleur résultat que l'inégalité de Tchebycheff, il faut donc vérifier que l'erreur d'approximation est inférieure à  $0,99 - 0,86 = 0,13$ .