Université Gustave Eiffel

Licence 3 — Mathématiques

# Intégration et probabilités

Notes de cours

Solesne Bourguin

Année universitaire 2026–2027

*Ce polycopié a été élaboré à partir de plusieurs sources. Une partie s'inspire notamment des notes de cours de Florence Merlevède, en particulier de ses notes manuscrites. Je la remercie chaleureusement pour ces documents, qui ont constitué une source précieuse pour la préparation de ce cours.*

# Table des matières


**I Intégration** **3**

**1 Introduction à l'intégrale de Lebesgue** **4**
1.1 Rappel sur l'intégrale de Riemann . . . . . 4
1.2 Limites de l'intégrale de Riemann . . . . . 4
1.3 Idée de l'intégrale de Lebesgue . . . . . 4
1.4 Un rappel de cardinalité . . . . . 5

**2 Espaces mesurés** **7**
2.1 Espaces mesurables et tribus . . . . . 7
2.1.1 Tribus . . . . . 7
2.1.2 Tribu borélienne sur $\mathbb{R}$ . . . . . 8
2.1.3 Tribu borélienne sur un espace métrique . . . . . 9
2.1.4 Autres exemples de tribus . . . . . 10
2.2 Fonctions mesurables . . . . . 11
2.2.1 Définition et exemples . . . . . 11
2.2.2 Propriétés de stabilité de l'ensemble des fonctions mesurables . . . . . 12
2.3 Mesures positives . . . . . 13
2.4 Construction de mesures et unicité . . . . . 14
2.4.1 Applications . . . . . 16
2.5 Mesure image . . . . . 17

**3 Intégrale par rapport à une mesure** **18**
3.1 Fonctions étagées . . . . . 18
3.2 Intégrale d'une fonction mesurable positive . . . . . 21
3.3 Mesures à densité . . . . . 23
3.4 Intégrale des fonctions de signe quelconque . . . . . 24
3.5 Intégrale par rapport à la mesure de Lebesgue . . . . . 25
3.6 Théorème de convergence dominée et applications . . . . . 26

**4 Espaces $L^p$ et $\ell^p$** **29**
4.1 Espaces $\mathcal{L}^p$ . . . . . 29
4.2 Espaces $L^p$ . . . . . 30
4.3 Liens entre convergence p.p. et convergence $L^p$ . . . . . 31
4.4 Quelques théorèmes de densité dans les espaces $L^p$ . . . . . 32

**5 Intégrales multiples** **33**
5.1 Intégrale sur un espace produit . . . . . 33
5.1.1 Tribu produit . . . . . 33
5.1.2 Sections d'un ensemble mesurable . . . . . 34
5.1.3 Mesure produit . . . . . 34
5.2 Théorèmes de Fubini . . . . . 35

5.3 Intégrales multiples sur $\mathbb{R}^d$ . . . . . 36

**II Probabilités** **38**

**6 Introduction aux probabilités** **39**
6.1 Expériences aléatoires et événements . . . . . 39
6.1.1 Espace des états . . . . . 40
6.1.2 Événements . . . . . 40
6.2 Probabilité . . . . . 43
6.3 Probabilité conditionnelle et indépendance . . . . . 45

**7 Variables aléatoires réelles** **49**
7.1 Variables aléatoires réelles . . . . . 49
7.2 Loi d'une variable aléatoire réelle . . . . . 50
7.2.1 Variables aléatoires discrètes . . . . . 50
7.2.2 Variables aléatoires continues . . . . . 51
7.2.3 Fonction de répartition . . . . . 52
7.2.4 Moments des variables aléatoires réelles . . . . . 54
7.3 Méthode de la fonction muette . . . . . 57

**8 Vecteurs aléatoires** **58**
8.1 Introduction . . . . . 58
8.2 Généralités . . . . . 58
8.3 Identification de la loi d'un vecteur aléatoire . . . . . 60
8.4 Intégration des vecteurs aléatoires et moments . . . . . 60
8.4.1 Espérance . . . . . 60
8.4.2 Matrice de covariance . . . . . 61
8.4.3 Coefficient de corrélation linéaire . . . . . 61
8.5 Vecteurs aléatoires indépendants . . . . . 61
8.6 Fonction caractéristique . . . . . 64
8.7 Somme de variables aléatoires indépendantes . . . . . 65
8.7.1 Produit de convolution . . . . . 65
8.8 Vecteurs gaussiens . . . . . 66
8.8.1 Loi normale dans $\mathbb{R}^n$ . . . . . 66
8.8.2 Fonction caractéristique d'un vecteur gaussien . . . . . 67
8.8.3 Indépendance dans le cas gaussien . . . . . 67

**9 Théorèmes limites** **68**
9.1 Convergences de suites de variables aléatoires . . . . . 68
9.1.1 Convergence presque sûre . . . . . 68
9.1.2 Convergence en probabilité . . . . . 69
9.1.3 Convergence dans $L^p$ . . . . . 70
9.2 Loi des grands nombres . . . . . 71
9.3 Convergence en loi . . . . . 72
9.4 Complément : preuve de la loi forte des grands nombres . . . . . 75

# Première partie

# Intégration

# Chapitre 1

# Introduction à l'intégrale de Lebesgue

## 1.1 Rappel sur l'intégrale de Riemann

**Rappel 1.1** (Intégrale de Riemann sur un segment)**.** Soit $f$ une fonction bornée sur $[\alpha, \beta]$. Pour une subdivision $\tau = \{\alpha = t_0 < t_1 < \cdots < t_p = \beta\}$ de $[\alpha, \beta]$, on pose

$$I_+(f, \tau) = \sum_{i=1}^{p} (t_i - t_{i-1}) \sup_{x \in [t_{i-1}, t_i]} f(x), \qquad I_-(f, \tau) = \sum_{i=1}^{p} (t_i - t_{i-1}) \inf_{x \in [t_{i-1}, t_i]} f(x).$$

Le pas de la subdivision est $|\tau| = \sup_{1 \le i \le p} (t_i - t_{i-1})$. On dit que $f$ est Riemann-intégrable sur $[\alpha, \beta]$ si, pour toute suite $(\tau_n)_{n \ge 1}$ de subdivisions telle que $|\tau_n| \to 0$, on a $I_+(f, \tau_n) - I_-(f, \tau_n) \to 0$. Dans ce cas, $I_+(f, \tau_n)$ et $I_-(f, \tau_n)$ convergent vers une même limite, indépendante de la suite de subdivisions. On la note $\int_\alpha^\beta f(x)\, dx$.

L'intégrale de Riemann est liée à la notion de longueur. Si $A = [a, b] \subset [\alpha, \beta]$, alors $\mathbf{1}_A$ est Riemann-intégrable et $\int_\alpha^\beta \mathbf{1}_A(x)\, dx = b - a$, où $\mathbf{1}_A(x) = 1$ si $x \in A$ et $\mathbf{1}_A(x) = 0$ sinon. En revanche, si $B = \mathbb{Q} \cap [0, 1]$, alors, pour toute subdivision $\tau$ de $[0, 1]$, on a $I_+(\mathbf{1}_B, \tau) = 1$ et $I_-(\mathbf{1}_B, \tau) = 0$, car tout intervalle non trivial contient à la fois des rationnels et des irrationnels. Ainsi $\mathbf{1}_B$ n'est pas Riemann-intégrable sur $[0, 1]$. Dans la théorie de Lebesgue, $\mathbf{1}_B$ sera intégrable et son intégrale sera $\lambda(B) = 0$, où $\lambda$ désigne la longueur de Lebesgue : l'ensemble $B$ est une réunion dénombrable de singletons, et chaque singleton aura une longueur nulle.

Pour étendre la notion d'intégrale, on procédera en quatre temps : définir une classe $\mathcal{A}$ de parties auxquelles on peut attribuer une longueur $\mu(A)$ ; définir les propriétés que doit vérifier $\mu$ ; poser $\int \mathbf{1}_A\, d\mu = \mu(A)$ ; puis étendre l'intégrale aux fonctions générales par approximation.

## 1.2 Limites de l'intégrale de Riemann

L'intégrale de Riemann présente deux limitations importantes. D'abord, une fonction Riemann-intégrable sur un segment est nécessairement bornée. Ensuite, la classe des fonctions Riemann-intégrables n'est pas stable par convergence simple. Elle n'est pas non plus stable par composition en toute généralité : si $f$ et $g$ sont Riemann-intégrables, $g \circ f$ ne l'est pas forcément, sauf hypothèses supplémentaires, par exemple si $g$ est continue.

## 1.3 Idée de l'intégrale de Lebesgue

Pour intégrer une fonction positive $f$, l'approche de Riemann découpe l'axe des abscisses et approxime l'aire sous le graphe par des rectangles verticaux. L'idée de Lebesgue est de raisonner plutôt par niveaux de valeurs : on découpe l'axe des ordonnées en petits intervalles. Si $(B_k)_k$ est une partition

de l'image de $f$ et si $c_k \in B_k$, on approche $f$ par la fonction simple $s = \sum_k c_k \mathbf{1}_{f^{-1}(B_k)}$, ce qui conduit à l'approximation

$$\int f \, d\mu \simeq \sum_k c_k \, \mu\left(f^{-1}(B_k)\right).$$

La difficulté nouvelle est donc de donner un sens à la taille des ensembles de la forme $f^{-1}(B_k)$. Même lorsque $B_k$ est un intervalle de valeurs, sa préimage n'est pas forcément un intervalle de l'axe des abscisses. Dans la figure ci-dessous, par exemple, l'intervalle de valeurs est $[y_1, y_2]$, mais sa préimage est l'union de plusieurs intervalles disjoints : $f^{-1}([y_1, y_2]) = I_1 \cup I_2 \cup I_3 \cup I_4$.

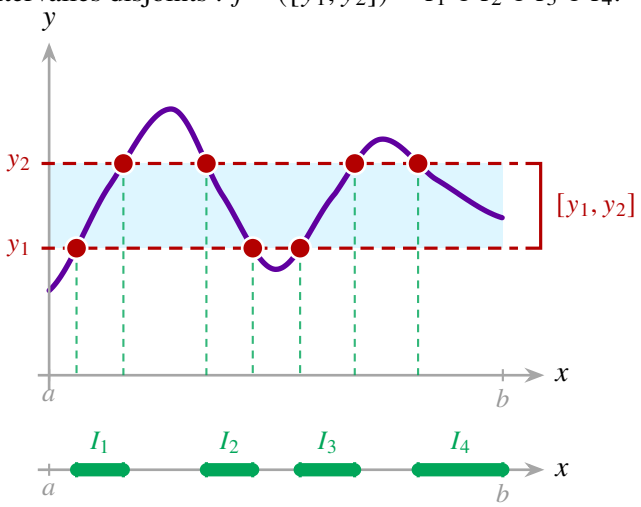



FIGURE 1.1 – Préimage d'un intervalle de valeurs : l'ensemble $f^{-1}([y_1, y_2])$ est représenté en vert sur l'axe inférieur et vaut ici $I_1 \cup I_2 \cup I_3 \cup I_4$ ; il n'est donc pas forcément un intervalle.

Cette figure reste très simple. En réalité, les ensembles de la forme $f^{-1}(B_k)$ ne sont en général pas des intervalles, ni même nécessairement des unions finies d'intervalles. Leur associer une longueur, ou plus généralement une mesure, est alors très délicat, et parfois impossible sans imposer une structure adaptée aux parties que l'on veut mesurer. C'est précisément ce qui conduit à introduire les tribus et les mesures.

## 1.4 Un rappel de cardinalité

**Exemple 1.2.** Les ensembles $\mathcal{P}(E)$ et $\{0,1\}^E$ sont équipotents : l'application $A \mapsto \mathbf{1}_A$ est une bijection de $\mathcal{P}(E)$ sur l'ensemble des applications de $E$ dans $\{0,1\}$.

**Exemple 1.3.** Les ensembles $\mathbb{N}$ et $2\mathbb{N}$ sont équipotents, via $n \mapsto 2n$. Les ensembles $\mathbb{N}$ et $\mathbb{N} \times \mathbb{N}$ le sont également : on peut énumérer les couples d'entiers en parcourant les diagonales $m + n = c$, $c \in \mathbb{N}$.

**Proposition 1.4** (Théorème de Cantor)**.** *Pour tout ensemble $E$, on a* $\mathrm{Card}(E) < \mathrm{Card}(\mathcal{P}(E))$.

*Démonstration.* Soit $f : E \to \mathcal{P}(E)$. Montrons que $f$ n'est jamais surjective. Posons $D_f = \{x \in E : x \notin f(x)\}$. Si $D_f$ avait un antécédent par $f$, il existerait $x_0 \in E$ tel que $f(x_0) = D_f$. Alors $x_0 \in D_f$ équivaudrait à $x_0 \notin f(x_0) = D_f$, contradiction. □

**Exemple 1.5.** Les ensembles $\mathbb{Z}$ et $\mathbb{N}$ sont équipotents. Par exemple, l'application $f : \mathbb{Z} \to \mathbb{N}$ définie par $f(n) = -2n$ si $n \leq 0$, et $f(n) = 2n - 1$ si $n > 0$, est une bijection.

**Proposition 1.6.** *Tout produit cartésien infini dénombrable d'ensembles non vides dont aucun n'est réduit à un singleton est non dénombrable.*

*Démonstration.* Soit $(E_i)_{i\geq 0}$ une suite d'ensembles contenant chacun au moins deux éléments. Pour tout $i$, il existe une injection $\psi_i\colon \{0,1\} \to E_i$. L'application

$$\varphi\colon \{0,1\}^{\mathbb{N}} \to \prod_{i\geq 0} E_i, \qquad \varphi((x_i)_{i\geq 0}) = (\psi_i(x_i))_{i\geq 0},$$

est injective. Ainsi $\mathrm{Card}(\prod_{i\geq 0} E_i) \geq \mathrm{Card}(\{0,1\}^{\mathbb{N}}) = \mathrm{Card}(\mathcal{P}(\mathbb{N})) > \mathrm{Card}(\mathbb{N})$, d'après le théorème de Cantor. □

**Théorème 1.7.** *Les ensembles $\mathbb{R}$ et $\mathcal{P}(\mathbb{N})$ sont équipotents.*

# Chapitre 2

# Espaces mesurés

## 2.1 Espaces mesurables et tribus

### 2.1.1 Tribus

Dans toute cette section, $\Omega$ désigne un ensemble quelconque et $\mathcal{P}(\Omega)$ l'ensemble de ses parties. Si $\Omega$ est fini de cardinal $n$, alors $\mathcal{P}(\Omega)$ est de cardinal $2^n$. Par exemple, pour $\Omega = \{a, b, c\}$, on a

$$\mathcal{P}(\Omega) = \{\emptyset, \{a\}, \{b\}, \{c\}, \{a, b\}, \{a, c\}, \{b, c\}, \Omega\}.$$

On dit que $\Omega$ est dénombrable s'il est en bijection avec $\mathbb{N}$ ; dans ce cas, on peut numéroter ses éléments.

**Exemple 2.1.** Les ensembles $\mathbb{N}$, $\mathbb{Z}$ et $\mathbb{Q}$ sont dénombrables, tandis que $\mathbb{R}$ et $[0, 1]$ ne le sont pas. En particulier, $\mathcal{P}(\mathbb{N})$ est en bijection avec $\{0, 1\}^{\mathbb{N}}$ et n'est pas dénombrable.

On commence par préciser la classe de parties sur laquelle une mesure sera définie.

**Définition 2.2.** Une classe $\mathcal{A} \subset \mathcal{P}(\Omega)$ est appelée *tribu*, ou *$\sigma$-algèbre*, si elle vérifie les trois propriétés suivantes :

(i) $\Omega \in \mathcal{A}$ ;

(ii) si $A \in \mathcal{A}$, alors $A^c \in \mathcal{A}$ ;

(iii) si $(A_n)_{n\in I}$ est une famille dénombrable d'éléments de $\mathcal{A}$, alors $\bigcup_{n\in I} A_n \in \mathcal{A}$.

On dit alors que $(\Omega, \mathcal{A})$ est un *espace mesurable*. Si $A \in \mathcal{A}$, on dit que $A$ est *$\mathcal{A}$-mesurable*.

*Remarque* 2.3. Une classe $C$ de parties de $\Omega$ est une algèbre, ou algèbre de Boole, si elle contient $\Omega$, est stable par passage au complémentaire et par réunion finie. Une tribu est donc une algèbre stable par réunion dénombrable, d'où le nom de $\sigma$-algèbre.

**Proposition 2.4.** *Si $\mathcal{A}$ est une tribu sur $\Omega$, alors $\emptyset \in \mathcal{A}$, $\mathcal{A}$ est stable par intersection dénombrable, par différence, par différence symétrique, ainsi que par limites supérieure et inférieure de suites d'ensembles.*

*Démonstration.* On a $\emptyset = \Omega^c \in \mathcal{A}$. Si $(A_n)_{n\geq 1} \subset \mathcal{A}$, alors $\bigcap_{n\geq 1} A_n = (\bigcup_{n\geq 1} A_n^c)^c \in \mathcal{A}$. Si $A, B \in \mathcal{A}$, alors $A \setminus B = A \cap B^c \in \mathcal{A}$ et $A\Delta B = (A \cap B^c) \cup (B \cap A^c) \in \mathcal{A}$. Enfin,

$$\limsup_{n\to\infty} A_n = \bigcap_{n\geq 1} \bigcup_{k\geq n} A_k, \qquad \liminf_{n\to\infty} A_n = \bigcup_{n\geq 1} \bigcap_{k\geq n} A_k,$$

qui appartiennent à $\mathcal{A}$. □

**Rappel 2.5.** Pour une suite $(A_n)_{n\geq 1}$ de parties de $\Omega$, l'ensemble $\limsup A_n$ est l'ensemble des $\omega \in \Omega$ qui appartiennent à une infinité de $A_n$ ; l'ensemble $\liminf A_n$ est l'ensemble des $\omega$ qui appartiennent à tous les $A_n$ sauf peut-être un nombre fini. De façon équivalente,

$$\limsup A_n = \left\{\omega : \sum_{n\geq 1} \mathbf{1}_{A_n}(\omega) = +\infty\right\}, \qquad \liminf A_n = \left\{\omega : \sum_{n\geq 1} \mathbf{1}_{A_n^c}(\omega) < +\infty\right\}.$$

**Exemple 2.6.** Si $A_n = [(-1)^n, 3 + 1/n]$, alors $\limsup A_n = [-1, 3]$ et $\liminf A_n = [1, 3]$.

**Exemple 2.7** (Exemples de tribus)**.**

(i) La tribu triviale est $\{\emptyset, \Omega\}$.

(ii) La tribu discrète est $\mathcal{P}(\Omega)$. Si $\Omega$ est dénombrable, on considère souvent $\mathcal{P}(\Omega)$ comme la tribu naturelle sur $\Omega$.

(iii) Si $(A_n)_{n\in\mathbb{N}}$ est une partition finie ou dénombrable de $\Omega$, alors $\mathcal{A} = \{\bigcup_{i\in I} A_i : I \subset \mathbb{N}\}$ est une tribu.

(iv) Si $A \subset \Omega$, alors $\{\emptyset, \Omega, A, A^c\}$ est la plus petite tribu contenant $A$.

**Proposition 2.8.** *La classe $\mathcal{A} = \{A \subset \Omega : A$ ou $A^c$ est dénombrable$\}$ est une tribu sur $\Omega$.*

*Démonstration.* On a $\Omega \in \mathcal{A}$ car $\Omega^c = \emptyset$ est dénombrable, et la stabilité par complémentaire est immédiate. Soit maintenant $(A_n)_{n\geq 1} \subset \mathcal{A}$. Si tous les $A_n$ sont dénombrables, alors $\bigcup_n A_n$ est dénombrable. Sinon, il existe $n_0$ tel que $A_{n_0}$ ne soit pas dénombrable ; puisque $A_{n_0} \in \mathcal{A}$, l'ensemble $A_{n_0}^c$ est dénombrable. Or $(\bigcup_n A_n)^c \subset A_{n_0}^c$, donc $(\bigcup_n A_n)^c$ est dénombrable. Dans les deux cas, $\bigcup_n A_n \in \mathcal{A}$. □

*Remarque* 2.9. La tribu code l'information disponible sur une expérience. Si $\Omega = \{a, b, c\}$, on peut par exemple prendre $\mathcal{A} = \{\Omega, \emptyset, \{a\}, \{b, c\}\} = \sigma(\{a\})$ ou $\mathcal{A} = \{\Omega, \emptyset, \{c\}, \{a, b\}\}$. Le choix de la tribu dépend des événements que l'on souhaite distinguer.

**Proposition 2.10** (Tribu engendrée)**.**

(a) *L'intersection d'une collection non vide quelconque de tribus sur $\Omega$ est une tribu.*

(b) *Si $\mathcal{G} \subset \mathcal{P}(\Omega)$, il existe une plus petite tribu contenant $\mathcal{G}$. On la note $\sigma(\mathcal{G})$ et on l'appelle la tribu engendrée par $\mathcal{G}$. Plus précisément,*

$$\sigma(\mathcal{G}) = \bigcap_{\mathcal{A}\ tribu :\ \mathcal{G}\subset\mathcal{A}} \mathcal{A}.$$

*Démonstration.* Une intersection non vide de tribus contient $\Omega$, est stable par complémentaire et par réunion dénombrable ; c'est donc une tribu. Pour le second point, l'intersection ci-dessus est prise sur une famille non vide, car $\mathcal{P}(\Omega)$ est une tribu contenant $\mathcal{G}$. Elle contient $\mathcal{G}$ et est incluse dans toute tribu contenant $\mathcal{G}$, ce qui prouve la minimalité. □

*Remarque* 2.11. On peut avoir $\mathcal{G} \neq \mathcal{G}'$ mais $\sigma(\mathcal{G}) = \sigma(\mathcal{G}')$, par exemple pour $\mathcal{G} = \{A\}$ et $\mathcal{G}' = \{A^c\}$. On a toujours $\mathcal{G} \subset \sigma(\mathcal{G})$ ; si $\mathcal{A}$ est une tribu, alors $\mathcal{A} = \sigma(\mathcal{A})$ ; et si $\mathcal{G} \subset C$, alors $\sigma(\mathcal{G}) \subset \sigma(C)$. Enfin, si $\mathcal{A}$ et $\mathcal{B}$ sont deux tribus, on note souvent $\mathcal{A} \vee \mathcal{B} = \sigma(\mathcal{A} \cup \mathcal{B})$.

Attention, l'union de deux tribus n'est pas forcément une tribu. Par exemple, si $\Omega = \{1, 2, 3, 4\}$,

$$\mathcal{A} = \{\emptyset, \{1, 2\}, \{3, 4\}, \Omega\}, \qquad \mathcal{B} = \{\emptyset, \{1, 3\}, \{2, 4\}, \Omega\},$$

alors $\mathcal{A} \cup \mathcal{B}$ n'est pas stable par réunion puisque $\{1, 2\} \cup \{1, 3\} = \{1, 2, 3\} \notin \mathcal{A} \cup \mathcal{B}$.

### 2.1.2 Tribu borélienne sur $\mathbb{R}$

Dans cette section, $\Omega = \mathbb{R}$. Il existe plusieurs définitions équivalentes de la tribu borélienne.

**Définition 2.12.** La tribu borélienne sur $\mathbb{R}$, notée $\mathcal{B}(\mathbb{R})$, est la tribu engendrée par les intervalles ouverts de $\mathbb{R}$ : $\mathcal{B}(\mathbb{R}) = \sigma(\{]a, b[ : a, b \in \mathbb{R}\})$.

**Proposition 2.13.** *Tout intervalle ouvert, fermé ou semi-ouvert de $\mathbb{R}$ appartient à $\mathcal{B}(\mathbb{R})$.*

*Démonstration.* Les intervalles ouverts appartiennent à $\mathcal{B}(\mathbb{R})$ par définition. De plus, $[a, b[ = \bigcap_{n\geq 1} ]a - 1/n, b[$ et $[a, b] = \bigcap_{n\geq 1} ]a - 1/n, b + 1/n[$. Ces ensembles sont donc boréliens. Les autres types d'intervalles se traitent de même ou par passage au complémentaire. □

**Proposition 2.14.** *On a $\mathcal{B}(\mathbb{R}) = \sigma(\mathcal{G}) = \sigma(\mathcal{G}')$, où $\mathcal{G} = \{]-\infty, x] : x \in \mathbb{R}\}$ et $\mathcal{G}' = \{]-\infty, x] : x \in \mathbb{Q}\}$.*

*Démonstration.* Comme $\mathcal{G}' \subset \mathcal{G}$ et $\mathcal{G} \subset \mathcal{B}(\mathbb{R})$, on a $\sigma(\mathcal{G}') \subset \sigma(\mathcal{G}) \subset \mathcal{B}(\mathbb{R})$. Montrons l'inclusion réciproque. Soit $]a, b[$ un intervalle ouvert. On choisit des suites rationnelles $(a_n)$ et $(b_n)$ telles que $a < a_n < b_n < b$ pour tout $n$, avec $a_n \downarrow a$ et $b_n \uparrow b$. Alors $]a, b[ = \bigcup_{n\geq 1} ]a_n, b_n]$. Or $]a_n, b_n] = ]-\infty, b_n] \cap (]-\infty, a_n])^c \in \sigma(\mathcal{G}')$. Ainsi tous les intervalles ouverts appartiennent à $\sigma(\mathcal{G}')$, donc $\mathcal{B}(\mathbb{R}) \subset \sigma(\mathcal{G}')$. □

*Remarque* 2.15. On définit souvent la tribu borélienne à partir des ouverts. Une partie $A \subset \mathbb{R}$ est ouverte si, pour tout $x \in A$, il existe $\varepsilon > 0$ tel que $]x - \varepsilon, x + \varepsilon[ \subset A$. Cette notion s'étend naturellement aux espaces métriques $(E, d)$.

**Proposition 2.16.** *Tout ouvert non vide de $\mathbb{R}$ est réunion dénombrable d'intervalles ouverts.*

*Démonstration.* Cette propriété est admise ici. Elle repose sur la densité de $\mathbb{Q}$ dans $\mathbb{R}$ : on peut écrire un ouvert comme réunion d'intervalles à extrémités rationnelles contenus dans cet ouvert. □

**Proposition 2.17.** *La tribu borélienne sur $\mathbb{R}$ est la tribu engendrée par les ouverts de $\mathbb{R}$.*

*Démonstration.* Notons $O$ l'ensemble des ouverts de $\mathbb{R}$. Puisque tout intervalle ouvert est un ouvert, $\mathcal{B}(\mathbb{R}) \subset \sigma(O)$. Réciproquement, tout ouvert $O$ de $\mathbb{R}$ est, d'après la proposition précédente, réunion dénombrable d'intervalles ouverts ; ces intervalles sont boréliens, donc $O \in \mathcal{B}(\mathbb{R})$. Ainsi $O \subset \mathcal{B}(\mathbb{R})$ et, comme $\mathcal{B}(\mathbb{R})$ est une tribu, $\sigma(O) \subset \mathcal{B}(\mathbb{R})$. Les deux inclusions donnent le résultat. □

### 2.1.3 Tribu borélienne sur un espace métrique

Soit $(E, d)$ un espace métrique. La notion d'ouvert définie sur $\mathbb{R}$ se formule de la même manière : une partie $O \subset E$ est ouverte si, pour tout $x \in O$, il existe $r > 0$ tel que la boule ouverte $B(x, r)$ soit incluse dans $O$.

**Définition 2.18.** La tribu borélienne sur $E$, notée $\mathcal{B}(E)$, est la tribu engendrée par les ouverts de $E$. De façon équivalente, c'est aussi la tribu engendrée par les fermés de $E$.

Dans le cas de $\mathbb{R}^d$, muni de sa topologie usuelle, la tribu borélienne $\mathcal{B}(\mathbb{R}^d)$ est engendrée par les pavés ouverts

$$\prod_{i=1}^{d} ]a_i, b_i[, \qquad a_i < b_i,$$

ou même par les pavés à extrémités rationnelles. On a par ailleurs

$$\mathcal{B}(\mathbb{R}^d) = \mathcal{B}(\mathbb{R})^{\otimes d}.$$

Pour préciser cette notation, rappelons la définition de la tribu produit.

**Définition 2.19.** Soient $(\Omega_1, \mathcal{A}_1)$ et $(\Omega_2, \mathcal{A}_2)$ deux espaces mesurables. On appelle tribu produit sur $\Omega_1 \times \Omega_2$, et l'on note $\mathcal{A}_1 \otimes \mathcal{A}_2$, la tribu

$$\mathcal{A}_1 \otimes \mathcal{A}_2 = \sigma(\mathcal{A}_1 \times \mathcal{A}_2),$$

où

$$\mathcal{A}_1 \times \mathcal{A}_2 = \left\{A_1 \times A_2 : A_1 \in \mathcal{A}_1, A_2 \in \mathcal{A}_2\right\}.$$

Autrement dit, $\mathcal{A}_1 \otimes \mathcal{A}_2$ est la plus petite tribu contenant tous les pavés mesurables.

Par définition, on a toujours $\mathcal{B}(E_1) \otimes \mathcal{B}(E_2) \subset \mathcal{B}(E_1 \times E_2)$, lorsque $E_1 \times E_2$ est muni de la distance produit. Pour des espaces métriques séparables, l'égalité est vraie :

$$\mathcal{B}(E_1) \otimes \mathcal{B}(E_2) = \mathcal{B}(E_1 \times E_2).$$

### 2.1.4 Autres exemples de tribus

**Proposition 2.20** (Tribu image réciproque)**.** *Soit $f\colon X \to (Y, \mathcal{B})$ une application à valeurs dans un espace mesurable. Alors*

$$f^{-1}(\mathcal{B}) = \left\{f^{-1}(B)\colon B \in \mathcal{B}\right\}$$

*est une tribu sur $X$.*

*Démonstration.* C'est une conséquence immédiate des formules de transport par image réciproque : $f^{-1}(Y) = X$, $(f^{-1}(B))^c = f^{-1}(B^c)$ et, pour toute suite $(B_n)_{n\geq 1}$,

$$\bigcup_{n\geq 1} f^{-1}(B_n) = f^{-1}\left(\bigcup_{n\geq 1} B_n\right).$$

□

**Définition 2.21.** La tribu $f^{-1}(\mathcal{B})$ est appelée tribu image réciproque de $\mathcal{B}$ par $f$. On la note aussi parfois $\sigma(f)$ lorsqu'il n'y a pas d'ambiguïté sur la tribu de l'espace d'arrivée.

**Exemple 2.22** (Tribu trace et tribu bande)**.** Soit $(X, \mathcal{A})$ un espace mesurable et soit $Y \subset X$. En appliquant la construction précédente à l'injection canonique $i\colon Y \to X$, on obtient la tribu trace de $\mathcal{A}$ sur $Y$ :

$$i^{-1}(\mathcal{A}) = \{A \cap Y\colon A \in \mathcal{A}\}.$$

Si $Y \in \mathcal{A}$, alors cette tribu est incluse dans $\mathcal{A}$.

Soit maintenant $\pi\colon X \times Y \to X$ la projection canonique. La tribu image réciproque $\pi^{-1}(\mathcal{A})$ est la tribu bande :

$$\pi^{-1}(\mathcal{A}) = \{A \times Y\colon A \in \mathcal{A}\}.$$

L'image directe d'une tribu n'est en général pas une tribu. Si $f\colon (X, \mathcal{A}) \to Y$, la classe $\{f(A)\colon A \in \mathcal{A}\}$ n'est donc pas, en général, une bonne façon de transporter la structure mesurable sur $Y$.

**Définition 2.23** (Tribu image)**.** Soit $f\colon (X, \mathcal{A}) \to Y$ une application. On appelle tribu image de $\mathcal{A}$ par $f$ la tribu sur $Y$ définie par

$$\mathcal{B} = \left\{B \subset Y\colon f^{-1}(B) \in \mathcal{A}\right\}.$$

Cette classe est bien une tribu, encore une fois grâce aux formules de transport par image réciproque.

**Lemme 2.24** (Lemme de transport)**.** *Soit $f\colon X \to Y$ et soit $C \subset \mathcal{P}(Y)$. Alors les deux tribus sur $X$ suivantes coïncident :*

$$\sigma\big(f^{-1}(C)\big) = f^{-1}\big(\sigma(C)\big).$$

*Démonstration.* Comme $C \subset \sigma(C)$, on a $f^{-1}(C) \subset f^{-1}(\sigma(C))$. Or $f^{-1}(\sigma(C))$ est une tribu sur $X$, donc

$$\sigma\big(f^{-1}(C)\big) \subset f^{-1}\big(\sigma(C)\big).$$

Pour l'inclusion réciproque, considérons la tribu image de $\sigma(f^{-1}(C))$ par $f$, c'est-à-dire

$$\mathcal{B} = \left\{B \subset Y\colon f^{-1}(B) \in \sigma(f^{-1}(C))\right\}.$$

Alors $\mathcal{B}$ est une tribu sur $Y$ contenant $C$. Ainsi $\sigma(C) \subset \mathcal{B}$, ce qui donne $f^{-1}(\sigma(C)) \subset \sigma(f^{-1}(C))$. □

## 2.2 Fonctions mesurables

### 2.2.1 Définition et exemples

**Définition 2.25.** Soient $(E, \mathcal{E})$ et $(F, \mathcal{F})$ deux espaces mesurables. Une application $f : E \to F$ est dite $\mathcal{E}$-$\mathcal{F}$ mesurable, ou simplement mesurable lorsque les tribus sont sous-entendues, si

$$f^{-1}(\mathcal{F}) \subset \mathcal{E}.$$

La tribu $\sigma(f) = f^{-1}(\mathcal{F})$ est donc la plus petite tribu sur $E$ rendant $f$ mesurable.

**Exemple 2.26** (Indicatrices)**.** Soit $A \subset E$. La fonction $\mathbf{1}_A$ est mesurable de $(E, \mathcal{E})$ dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ si et seulement si $A \in \mathcal{E}$. En effet, pour tout borélien $B \subset \mathbb{R}$,

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

Ainsi $\sigma(\mathbf{1}_A) = \{\emptyset, E, A, A^c\}$.

*Remarque* 2.27. Lorsque $E$ et $F$ sont des espaces métriques munis de leurs tribus boréliennes, on parle simplement de fonction borélienne.

**Exemple 2.28** (Exemples élémentaires)**.** Si $\mathcal{E} = \mathcal{P}(E)$, toute application de $E$ dans $(F, \mathcal{F})$ est mesurable. Si $\mathcal{E} = \{\emptyset, E\}$ et si les singletons de $F$ sont mesurables, alors les seules applications mesurables de $E$ dans $F$ sont les applications constantes.

**Quelques propriétés.**

**Proposition 2.29.** *Soit $f : (E, \mathcal{E}) \to (F, \mathcal{F})$. Si $\mathcal{F} = \sigma(C)$, où $C \subset \mathcal{P}(F)$, alors $f$ est mesurable si et seulement si*

$$f^{-1}(C) \in \mathcal{E} \qquad \textit{pour tout } C \in C.$$

*Autrement dit, il suffit de tester la mesurabilité sur une classe engendrant la tribu d'arrivée.*

*Démonstration.* On a $f$ mesurable si et seulement si $f^{-1}(\sigma(C)) \subset \mathcal{E}$. D'après le lemme de transport, cela équivaut à $\sigma(f^{-1}(C)) \subset \mathcal{E}$, donc, puisque $\mathcal{E}$ est une tribu, à $f^{-1}(C) \subset \mathcal{E}$. □

**Exemple 2.30** (Applications)**.** Soit $f : (X, \mathcal{A}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Alors $f$ est mesurable si et seulement si, pour tout $a \in \mathbb{R}$, l'ensemble

$$\{f \geq a\} = \{x \in X : f(x) \geq a\}$$

appartient à $\mathcal{A}$. On peut remplacer $\{f \geq a\}$ par $\{f > a\}$, $\{f \leq a\}$ ou $\{f < a\}$.

Plus généralement, si $f : (X, \mathcal{A}) \to (E, \mathcal{B}(E))$, où $E$ est un espace métrique, alors $f$ est mesurable si et seulement si $f^{-1}(O) \in \mathcal{A}$ pour tout ouvert $O$ de $E$. En particulier, toute application continue entre espaces métriques est borélienne.

**Proposition 2.31** (Stabilité par composition)**.** *Soient $f : (E, \mathcal{E}) \to (F, \mathcal{F})$ et $g : (F, \mathcal{F}) \to (G, \mathcal{G})$ deux applications mesurables. Alors $g \circ f$ est mesurable de $(E, \mathcal{E})$ dans $(G, \mathcal{G})$.*

*Démonstration.* Pour tout $C \in \mathcal{G}$,

$$(g \circ f)^{-1}(C) = f^{-1}\big(g^{-1}(C)\big).$$

Comme $g$ est mesurable, $g^{-1}(C) \in \mathcal{F}$ ; comme $f$ est mesurable, $f^{-1}(g^{-1}(C)) \in \mathcal{E}$. □

**Exemple 2.32** (Applications)**.** Si $f\colon (X,\mathcal{A}) \to (\mathbb{R},\mathcal{B}(\mathbb{R}))$ est mesurable, alors, pour tout $a \in \mathbb{R}$, les fonctions $\max(f,a)$, $\min(f,a)$, $|f|$, $f^+ = \max(f,0)$ et $f^- = -\min(f,0)$ sont mesurables. De plus, si $f$ ne s'annule pas, alors $1/f$ est mesurable, car $x \mapsto 1/x$ est continue sur $\mathbb{R}^*$.

**Proposition 2.33** (Fonctions vectorielles)**.** *Soit $f = (f_1,\dots,f_d)\colon (X,\mathcal{A}) \to (\mathbb{R}^d,\mathcal{B}(\mathbb{R}^d))$. Alors $f$ est mesurable si et seulement si chaque coordonnée $f_i\colon (X,\mathcal{A}) \to (\mathbb{R},\mathcal{B}(\mathbb{R}))$ est mesurable.*

*Démonstration.* Il suffit de traiter le cas $d = 2$, le cas général étant identique. Si $f$ est mesurable, alors $f_i = \pi_i \circ f$, où $\pi_i$ est la projection canonique de $\mathbb{R}^2$ sur sa $i$-ème coordonnée ; comme $\pi_i$ est continue, elle est borélienne, et donc $f_i$ est mesurable.

Réciproquement, la tribu $\mathcal{B}(\mathbb{R}^2)$ est engendrée par les rectangles ouverts $U \times V$, avec $U, V$ ouverts de $\mathbb{R}$. Si $f_1$ et $f_2$ sont mesurables, alors

$$f^{-1}(U \times V) = f_1^{-1}(U) \cap f_2^{-1}(V) \in \mathcal{A}.$$

Le critère de mesurabilité par classe génératrice donne la mesurabilité de $f$. □

**Exemple 2.34.** En identifiant $\mathbb{C}$ et $\mathbb{R}^2$, une fonction $f\colon (X,\mathcal{A}) \to (\mathbb{C},\mathcal{B}(\mathbb{C}))$ est mesurable si et seulement si $\operatorname{Re} f$ et $\operatorname{Im} f$ le sont.

### 2.2.2 Propriétés de stabilité de l'ensemble des fonctions mesurables

**Proposition 2.35.** *Soient $f$ et $g$ deux fonctions mesurables de $(E,\mathcal{E})$ dans $(\mathbb{R},\mathcal{B}(\mathbb{R}))$. Alors :*

(i) *pour tout $\lambda \in \mathbb{R}$, la fonction $\lambda f$ est mesurable ;*

(ii) *$f + g$ est mesurable ;*

(iii) *$fg$ est mesurable ;*

(iv) *$\max(f,g)$ et $\min(f,g)$ sont mesurables.*

*Démonstration.* L'application $\psi\colon E \to \mathbb{R}^2$, $x \mapsto (f(x),g(x))$, est mesurable par la proposition précédente. Les applications continues de $\mathbb{R}^2$ dans $\mathbb{R}$ données par $(u,v) \mapsto u+v$, $(u,v) \mapsto uv$, $(u,v) \mapsto \max(u,v)$ et $(u,v) \mapsto \min(u,v)$ sont boréliennes. Le résultat suit par composition. Le premier point se déduit du second en prenant $g = \lambda$ constant. □

**Passage à la limite.**

**Proposition 2.36.** *Soit $(f_n)_{n\geq 1}$ une suite de fonctions mesurables de $(E,\mathcal{E})$ dans $(\mathbb{R},\mathcal{B}(\mathbb{R}))$. Alors :*

(i) *$\sup_n f_n$ et $\inf_n f_n$ sont mesurables ;*

(ii) *$\limsup_n f_n$ et $\liminf_n f_n$ sont mesurables ;*

(iii) *l'ensemble des points $x \in E$ où la suite $(f_n(x))_{n\geq 1}$ converge est mesurable ;*

(iv) *si $f_n$ converge simplement vers $f$, alors $f$ est mesurable.*

*Démonstration.* Pour tout $a \in \mathbb{R}$,

$$\left\{\sup_n f_n \leq a\right\} = \bigcap_{n\geq 1}\{f_n \leq a\} \in \mathcal{E}, \qquad \left\{\inf_n f_n < a\right\} = \bigcup_{n\geq 1}\{f_n < a\} \in \mathcal{E}.$$

Cela prouve (i). Le point (ii) suit de (i) et des formules

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

Pour (iii), notons $A$ l'ensemble des $x \in E$ tels que $(f_n(x))_{n\geq 1}$ converge. Comme $\mathbb{R}$ est complet, la convergence équivaut au critère de Cauchy. En prenant des $\varepsilon$ rationnels positifs, on obtient

$$A = \bigcap_{\varepsilon\in\mathbb{Q}_+^*} \bigcup_{N\geq 1} \bigcap_{n\geq N} \bigcap_{p\geq 1} \left\{x \in E\colon \left|f_{n+p}(x) - f_n(x)\right| \leq \varepsilon\right\}.$$

Chaque ensemble qui apparaît à droite est mesurable, donc $A \in \mathcal{E}$. Enfin, si $f_n \to f$ simplement, alors $\limsup_n f_n = \liminf_n f_n = f$, et (iv) découle de (ii). □

**Exemple 2.37.** Si $f\colon \mathbb{R} \to \mathbb{R}$ est dérivable, alors sa dérivée est borélienne. En effet,

$$f'(x) = \lim_{n\to\infty} n\left(f\left(x + \frac{1}{n}\right) - f(x)\right),$$

limite simple de fonctions boréliennes.

## 2.3 Mesures positives

Dans tout ce cours, les mesures seront positives ; on omettra donc souvent cet adjectif dans la suite.

**Définition 2.38.** Soit $(\Omega, \mathcal{A})$ un espace mesurable. Une mesure positive sur $(\Omega, \mathcal{A})$ est une application $\mu\colon \mathcal{A} \to \overline{\mathbb{R}}_+ = [0, +\infty]$ vérifiant :

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

(ii) pour toute suite $(A_n)_{n\geq 1}$ d'éléments de $\mathcal{A}$ deux à deux disjoints,

$$\mu\left(\bigsqcup_{n\geq 1} A_n\right) = \sum_{n\geq 1} \mu(A_n).$$

Le triplet $(\Omega, \mathcal{A}, \mu)$ est appelé espace mesuré. On dit aussi que $\mu$ est une fonctionnelle d'ensemble $\sigma$-additive.

**Définition 2.39.** Une mesure $\mu$ sur $(\Omega, \mathcal{A})$ est dite finie si $\mu(A) < +\infty$ pour tout $A \in \mathcal{A}$. Elle est dite $\sigma$-finie s'il existe une suite croissante $(A_n)_{n\geq 1}$ d'ensembles mesurables telle que

$$\bigcup_{n\geq 1} A_n = \Omega \qquad \text{et} \qquad \mu(A_n) < +\infty \quad \text{pour tout } n.$$

**Exemple 2.40.**

(1) La mesure de Dirac au point $a \in \Omega$, notée $\delta_a$, est définie par $\delta_a(A) = 1$ si $a \in A$ et $\delta_a(A) = 0$ sinon.

(2) Sur $\{0, 1\}$, la mesure $(1 - p)\delta_0 + p\delta_1$ est la mesure de Bernoulli de paramètre $p \in [0, 1]$.

(3) La mesure de comptage sur $\mathcal{P}(\Omega)$ est définie par

$$\mu(A) = \begin{cases} \mathrm{Card}(A), & A \text{ est fini,} \\ +\infty, & \text{sinon.} \end{cases}$$

**Proposition 2.41** (Premières propriétés d'une mesure)**.** *Soit $\mu$ une mesure sur un espace mesurable $(\Omega, \mathcal{A})$. Alors :*

(i) *si $A_0, \ldots, A_n \in \mathcal{A}$ sont deux à deux disjoints, alors*

$$\mu\left(\bigsqcup_{i=0}^{n} A_i\right) = \sum_{i=0}^{n} \mu(A_i);$$

(ii) *si* $A, B \in \mathcal{A}$ *et* $A \subset B$, *alors* $\mu(A) \leq \mu(B)$ ;

(iii) *pour tous* $A, B \in \mathcal{A}$,

$$\mu(A \cup B) + \mu(A \cap B) = \mu(A) + \mu(B);$$

(iv) *pour toute famille dénombrable* $(A_i)_{i \in I}$ *d'éléments de* $\mathcal{A}$,

$$\mu\left(\bigcup_{i \in I} A_i\right) \leq \sum_{i \in I} \mu(A_i);$$

(v) *si* $(A_n)_{n \geq 1}$ *est croissante, c'est-à-dire* $A_n \subset A_{n+1}$, *alors*

$$\mu\left(\bigcup_{n \geq 1} A_n\right) = \lim_{n \to \infty} \mu(A_n);$$

(vi) *si* $(A_n)_{n \geq 1}$ *est décroissante, c'est-à-dire* $A_{n+1} \subset A_n$, *et s'il existe* $n_0$ *tel que* $\mu(A_{n_0}) < +\infty$, *alors*

$$\mu\left(\bigcap_{n \geq 1} A_n\right) = \lim_{n \to \infty} \mu(A_n).$$

*Démonstration.* Le point (i) découle de la $\sigma$-additivité en complétant la famille finie par des ensembles vides. Le point (ii) vient de l'écriture disjointe $B = A \sqcup (B \setminus A)$.

Pour (iii), on décompose $A \cup B = (A \setminus B) \sqcup (A \cap B) \sqcup (B \setminus A)$, puis on compare avec les décompositions disjointes de $A$ et de $B$. Pour (iv), on pose, après avoir numéroté la famille, $B_1 = A_1$ et $B_n = A_n \setminus \bigcup_{k<n} A_k$ pour $n \geq 2$. Les $B_n$ sont disjoints, $\bigcup_n B_n = \bigcup_n A_n$ et $B_n \subset A_n$, ce qui donne l'inégalité.

Pour (v), on pose $B_1 = A_1$ et $B_n = A_n \setminus A_{n-1}$ pour $n \geq 2$. Alors $A_n = \bigsqcup_{k=1}^{n} B_k$ et $\bigcup_n A_n = \bigsqcup_{k \geq 1} B_k$, d'où le résultat. Enfin, pour (vi), quitte à remplacer la suite par $(A_{n+n_0})_{n \geq 0}$, on suppose $\mu(A_1) < +\infty$. Les ensembles $B_n = A_1 \setminus A_n$ sont croissants et $\bigcup_n B_n = A_1 \setminus \bigcap_n A_n$. Par (v),

$$\mu(A_1) - \mu\left(\bigcap_{n \geq 1} A_n\right) = \lim_{n \to \infty} \big(\mu(A_1) - \mu(A_n)\big),$$

ce qui donne l'égalité annoncée. □

*Remarque* 2.42. L'hypothèse de finitude dans la continuité à droite est indispensable. Pour la mesure de comptage sur $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$, les ensembles $A_i = \{k \in \mathbb{N} : k \geq i\}$ forment une suite décroissante telle que $\bigcap_i A_i = \emptyset$, mais $\mu(A_i) = +\infty$ pour tout $i$.

## 2.4 Construction de mesures et unicité

La construction d'une mesure suit souvent le schéma suivant : on commence par prescrire la mesure sur une famille concrète d'ensembles, par exemple les intervalles de $\mathbb{R}$ ou les pavés de $\mathbb{R}^d$, puis on cherche à prolonger cette fonction d'ensemble à la tribu engendrée. Le théorème de prolongement qui répond à cette question est un résultat profond, que nous n'énonçons pas ici. En revanche, nous allons établir des critères d'unicité : si deux mesures coïncident sur une classe d'ensembles suffisamment grande, alors elles coïncident sur toute la tribu engendrée.

**Définition 2.43.** Une classe $\Lambda$ de parties d'un ensemble $E$ est appelée un $\lambda$-système si :

(i) $\emptyset \in \Lambda$ ;

(ii) si $A, B \in \Lambda$ et $A \subset B$, alors $B \setminus A \in \Lambda$ ;

(iii) si $(A_n)_{n\geq 1}$ est une suite croissante d'éléments de $\Lambda$, alors $\bigcup_{n\geq 1} A_n \in \Lambda$.

*Remarque* 2.44. Toute tribu est un $\lambda$-système. De plus, l'intersection d'une collection non vide de $\lambda$-systèmes est encore un $\lambda$-système.

**Proposition 2.45.** *Soit $C$ une famille de parties de $E$.*

(a) *Il existe un plus petit $\lambda$-système contenant $C$, noté $\lambda(C)$.*

(b) *Si $E \in \Lambda$ et si $\Lambda$ est un $\lambda$-système stable par intersections finies, alors $\Lambda$ est une tribu.*

*Démonstration.* Pour (a), on prend l'intersection de tous les $\lambda$-systèmes contenant $C$. Cette intersection est non vide, car $\mathcal{P}(E)$ est un $\lambda$-système qui contient $C$.

Pour (b), comme $E \in \Lambda$, la stabilité par différence donne la stabilité par complémentaire. Il reste à montrer la stabilité par réunion dénombrable. Pour une suite $(A_n)_{n\geq 1}$ d'éléments de $\Lambda$, on écrit

$$\bigcup_{n\geq 1} A_n = \bigcup_{n\geq 1} \left( \bigcup_{k=1}^{n} A_k \right).$$

Les réunions finies appartiennent à $\Lambda$ car $A \cup B = (A^c \cap B^c)^c$ et $\Lambda$ est stable par complémentaire et par intersections finies. La suite des réunions finies est croissante, donc la réunion dénombrable appartient à $\Lambda$. □

**Théorème 2.46** (Théorème de la classe monotone)**.** *Soit $C$ une famille de parties de $E$, stable par intersections finies et contenant $E$ ; une telle famille est appelée un $\pi$-système. Alors*

$$\lambda(C) = \sigma(C).$$

*En particulier, le plus petit $\lambda$-système contenant un $\pi$-système est une tribu.*

*Démonstration.* Toute tribu est un $\lambda$-système, donc $\lambda(C) \subset \sigma(C)$. Pour obtenir l'inclusion réciproque, il suffit de montrer que $\lambda(C)$ est stable par intersections finies.

Posons

$$\mathcal{A}_1 = \{A \in \lambda(C) \colon A \cap C \in \lambda(C) \text{ pour tout } C \in C\}.$$

Comme $C$ est stable par intersections finies, on a $C \subset \mathcal{A}_1$. Les propriétés de $\lambda$-système montrent que $\mathcal{A}_1$ est un $\lambda$-système ; ainsi $\lambda(C) \subset \mathcal{A}_1$. On en déduit que, pour tout $A \in \lambda(C)$ et tout $C \in C$, $A \cap C \in \lambda(C)$.

Posons maintenant

$$\mathcal{A}_2 = \{A \in \lambda(C) \colon A \cap B \in \lambda(C) \text{ pour tout } B \in \lambda(C)\}.$$

Ce que l'on vient de montrer donne $C \subset \mathcal{A}_2$. De nouveau, $\mathcal{A}_2$ est un $\lambda$-système, donc $\lambda(C) \subset \mathcal{A}_2$. Cela signifie que $\lambda(C)$ est stable par intersections finies. Par la proposition précédente, $\lambda(C)$ est une tribu, et donc $\sigma(C) \subset \lambda(C)$. □

**Corollaire 2.47** (Unicité pour des mesures finies)**.** *Soient $\mu$ et $\nu$ deux mesures finies sur un espace mesurable $(\Omega, \mathcal{A})$. On suppose qu'elles coïncident sur un $\pi$-système $C \subset \mathcal{A}$ tel que $\sigma(C) = \mathcal{A}$. Alors $\mu$ et $\nu$ coïncident sur $\mathcal{A}$.*

*Démonstration.* Posons

$$\Lambda = \{A \in \mathcal{A} \colon \mu(A) = \nu(A)\}.$$

On vérifie que $\Lambda$ est un $\lambda$-système. En effet, $\emptyset \in \Lambda$. Si $A, B \in \Lambda$ et $A \subset B$, alors, puisque les mesures sont finies,

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

Enfin, si $(A_n)$ est une suite croissante d'éléments de $\Lambda$, la continuité à gauche donne

$$\mu\left(\bigcup_n A_n\right) = \lim_n \mu(A_n) = \lim_n \nu(A_n) = \nu\left(\bigcup_n A_n\right).$$

Ainsi $\Lambda$ est un $\lambda$-système contenant $C$. Par le théorème de la classe monotone, $\mathcal{A} = \sigma(C) = \lambda(C) \subset \Lambda$. □

**Corollaire 2.48** (Unicité pour des mesures $\sigma$-finies)**.** *Soient $\mu$ et $\nu$ deux mesures $\sigma$-finies sur $(\Omega, \mathcal{A})$. On suppose qu'il existe une suite croissante $(E_n)_{n\geq 1}$ d'éléments de $\mathcal{A}$ telle que $\bigcup_n E_n = \Omega$, $\mu(E_n) = \nu(E_n) < +\infty$ pour tout n, et que $\mu$ et $\nu$ coïncident sur un $\pi$-système $C$ qui engendre $\mathcal{A}$ et contient chaque $E_n$. Alors $\mu$ et $\nu$ coïncident sur $\mathcal{A}$.*

*Démonstration.* Pour chaque $n$, on applique le corollaire précédent aux mesures traces

$$\mu_n(A) = \mu(A \cap E_n), \qquad \nu_n(A) = \nu(A \cap E_n), \qquad A \in \mathcal{A}.$$

Ces mesures sont finies, grâce à l'hypothèse $\mu(E_n) = \nu(E_n) < +\infty$. Elles coïncident sur $C$ : si $C \in C$, alors $C \cap E_n \in C$, car $C$ est stable par intersections finies et contient $E_n$, et donc

$$\mu_n(C) = \mu(C \cap E_n) = \nu(C \cap E_n) = \nu_n(C).$$

Comme $\sigma(C) = \mathcal{A}$, on obtient $\mu_n = \nu_n$ sur $\mathcal{A}$.

Soit maintenant $A \in \mathcal{A}$. La suite $(A \cap E_n)_{n\geq 1}$ est croissante et $\bigcup_n (A \cap E_n) = A$. Par continuité à gauche,

$$\mu(A) = \lim_{n\to\infty} \mu(A \cap E_n) = \lim_{n\to\infty} \nu(A \cap E_n) = \nu(A).$$

Ainsi $\mu$ et $\nu$ coïncident sur $\mathcal{A}$. □

### 2.4.1 Applications

**Proposition 2.49** (Unicité de la mesure de Lebesgue)**.** *Supposons qu'il existe deux mesures $\mu$ et $\nu$ sur $\mathcal{B}(\mathbb{R}^d)$ telles que, pour tout pavé ouvert*

$$R = \prod_{k=1}^{d} I_k, \qquad I_k = ]a_k, b_k[,$$

*avec $a_k, b_k \in \overline{\mathbb{R}}$, on ait*

$$\mu(R) = \nu(R) = \prod_{k=1}^{d} (b_k - a_k),$$

*avec la convention usuelle $0 \cdot +\infty = 0$. Alors $\mu = \nu$ sur $\mathcal{B}(\mathbb{R}^d)$. Cette propriété assure l'unicité de la mesure de Lebesgue sur $\mathbb{R}^d$ ; son existence est admise ici.*

*Démonstration.* On traite le cas $d = 1$, le cas général étant identique en remplaçant les intervalles par des pavés. Posons

$$C = \{]a, b[ : a, b \in \overline{\mathbb{R}},\ a < b\}.$$

La classe $C$ est stable par intersections finies et engendre $\mathcal{B}(\mathbb{R})$. De plus, $\mathbb{R} = \bigcup_{n\geq 1} ]-n, n[$, avec $]-n, n[ \in C$, et les deux mesures donnent la même valeur finie à chacun de ces intervalles. Le corollaire d'unicité pour les mesures $\sigma$-finies s'applique donc, et $\mu = \nu$ sur $\mathcal{B}(\mathbb{R})$. □

*Remarque* 2.50. La mesure de Lebesgue $\lambda$ vérifie $\lambda(\{a\}) = 0$ pour tout $a \in \mathbb{R}$. En effet,

$$\{a\} = \bigcap_{n \geq 1} ]a - 1/n, a + 1/n[$$

et la continuité à droite donne

$$\lambda(\{a\}) = \lim_{n \to \infty} \lambda(]a - 1/n, a + 1/n[) = 0.$$

*Remarque* 2.51 (Invariance par translation). La mesure de Lebesgue est invariante par translation. En effet, si $x \in \mathbb{R}^d$ et si l'on pose $\mu_x(A) = \lambda(A - x)$ pour $A \in \mathcal{B}(\mathbb{R}^d)$, alors $\mu_x$ est encore une mesure sur $\mathcal{B}(\mathbb{R}^d)$ et elle prend la même valeur que $\lambda$ sur les pavés ouverts. Par unicité, $\mu_x = \lambda$.

## 2.5 Mesure image

Pour clore ce chapitre sur les espaces mesurés, nous introduisons la notion de mesure image. C'est l'une des notions les plus utiles en probabilités : la loi d'une variable aléatoire est exactement une mesure image.

**Proposition 2.52** (Définition de la mesure image)**.** *Soient $(E, \mathcal{E}, \mu)$ un espace mesuré, $(S, \mathcal{S})$ un espace mesurable et $f : E \to S$ une application mesurable. On définit, pour $A \in \mathcal{S}$,*

$$\mu_f(A) = \mu\left(f^{-1}(A)\right).$$

*Alors $\mu_f$ est une mesure sur $(S, \mathcal{S})$. On l'appelle la mesure image de $\mu$ par $f$ ; on la note aussi $f_*\mu$.*

*Démonstration.* On a $\mu_f(\emptyset) = \mu(f^{-1}(\emptyset)) = \mu(\emptyset) = 0$. Si $(A_n)_{n \geq 1}$ est une suite d'éléments deux à deux disjoints de $\mathcal{S}$, alors les ensembles $f^{-1}(A_n)$ sont deux à deux disjoints et

$$f^{-1}\left(\bigsqcup_{n \geq 1} A_n\right) = \bigsqcup_{n \geq 1} f^{-1}(A_n).$$

La $\sigma$-additivité de $\mu$ donne donc

$$\mu_f\left(\bigsqcup_{n \geq 1} A_n\right) = \sum_{n \geq 1} \mu_f(A_n).$$

Ainsi $\mu_f$ est bien une mesure. □

**Exemple 2.53** (Loi d'une variable aléatoire)**.** Soit $X$ une variable aléatoire réelle sur un espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$, c'est-à-dire une application mesurable de $(\Omega, \mathcal{F})$ dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. La loi de $X$, notée $\mathbb{P}_X$, est la mesure image de $\mathbb{P}$ par $X$ : pour tout borélien $B \in \mathcal{B}(\mathbb{R})$,

$$\mathbb{P}_X(B) = \mathbb{P}(X \in B) = \mathbb{P}\left(X^{-1}(B)\right).$$

# Chapitre 3

# Intégrale par rapport à une mesure

Soit $(X, \mathcal{A}, \mu)$ un espace mesuré. Nous allons définir l'intégrale d'une fonction $f$ par rapport à la mesure $\mu$ pour une classe aussi large que possible de fonctions. On veut en particulier que

$$\int_X \mathbf{1}_A \, d\mu = \mu(A), \qquad A \in \mathcal{A},$$

et que l'application $f \mapsto \int_X f \, d\mu$ soit linéaire lorsque cela a un sens.

Le principe de construction est le suivant. On définit d'abord l'intégrale pour les combinaisons linéaires finies de fonctions indicatrices, c'est-à-dire pour les fonctions étagées. On en déduit ensuite l'intégrale des fonctions mesurables positives par passage à la limite croissante. Enfin, pour une fonction mesurable de signe quelconque, on écrit $f = f^+ - f^-$, où $f^+ = \max(f, 0)$ et $f^- = \max(-f, 0)$.

## 3.1 Fonctions étagées

**Définition 3.1.** Soit $K = \mathbb{R}$ ou $\mathbb{C}$. Une fonction $f \colon (X, \mathcal{A}) \to (K, \mathcal{B}(K))$ est dite étagée si elle est mesurable et ne prend qu'un nombre fini de valeurs. Dans ce cas, il existe une partition finie mesurable $(A_i)_{i \in I}$ de $X$ et des scalaires $(\alpha_i)_{i \in I}$ de $K$ tels que

$$f = \sum_{i \in I} \alpha_i \mathbf{1}_{A_i}.$$

Si les $\alpha_i$ sont deux à deux distincts, cette décomposition est unique : on a alors $A_i = \{f = \alpha_i\}$. De manière canonique, on peut donc écrire

$$f = \sum_{\alpha \in f(X)} \alpha \, \mathbf{1}_{\{f = \alpha\}}.$$

Les fonctions étagées jouent, dans l'intégration de Lebesgue, le rôle des fonctions en escaliers dans l'intégration de Riemann.

**Exemple 3.2.** Si $A \in \mathcal{A}$, alors $\mathbf{1}_A$ est une fonction étagée, car $\mathbf{1}_A = 1 \cdot \mathbf{1}_A + 0 \cdot \mathbf{1}_{A^c}$.

**Proposition 3.3.** *L'ensemble des fonctions étagées de $(X, \mathcal{A})$ dans $K$, noté $\mathcal{E}_K(X, \mathcal{A})$, est une $K$-algèbre. Autrement dit, il est stable par combinaisons linéaires et par produit. Lorsque $K = \mathbb{R}$, il est aussi stable par maximum et minimum finis.*

*Démonstration.* La fonction nulle est étagée. Soient $f = \sum_{i \in I} \alpha_i \mathbf{1}_{A_i}$ et $g = \sum_{j \in J} \beta_j \mathbf{1}_{B_j}$, où $(A_i)_{i \in I}$ et $(B_j)_{j \in J}$ sont deux partitions finies mesurables de $X$. Alors $(A_i \cap B_j)_{(i,j) \in I \times J}$ est encore une partition finie mesurable de $X$, et

$$\lambda f + g = \sum_{i,j} (\lambda \alpha_i + \beta_j) \mathbf{1}_{A_i \cap B_j}, \qquad fg = \sum_{i,j} \alpha_i \beta_j \mathbf{1}_{A_i \cap B_j}.$$

Lorsque $K = \mathbb{R}$, on a de même

$$\max(f,g) = \sum_{i,j} \max(\alpha_i, \beta_j) \mathbf{1}_{A_i \cap B_j}, \qquad \min(f,g) = \sum_{i,j} \min(\alpha_i, \beta_j) \mathbf{1}_{A_i \cap B_j}.$$

□

**Lemme 3.4** (Lemme fondamental d'approximation)**.** *Soit $f\colon (X, \mathcal{A}) \to (\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}}))$ une fonction mesurable. Alors il existe une suite $(f_n)_{n\geq 1}$ de fonctions étagées qui converge simplement vers $f$. De plus :*

(i) *si $f \geq 0$, on peut choisir $f_n \geq 0$ et $f_n \leq f_{n+1}$ pour tout $n$ ;*

(ii) *si $f$ est bornée, on peut choisir une suite qui converge uniformément vers $f$.*

*Démonstration.* Supposons d'abord $f \geq 0$. Pour $n \geq 1$ et $0 \leq k < n2^n$, posons

$$E_{n,k} = \left\{ x \in X\colon \frac{k}{2^n} \leq f(x) < \frac{k+1}{2^n} \right\}, \qquad E_{n,\infty} = \{x \in X\colon f(x) \geq n\}.$$

Ces ensembles sont mesurables, et l'on définit

$$f_n(x) = \sum_{k=0}^{n2^n - 1} \frac{k}{2^n} \mathbf{1}_{E_{n,k}}(x) + n \mathbf{1}_{E_{n,\infty}}(x).$$

La fonction $f_n$ est étagée, positive, et $f_n \leq f$. Le raffinement des intervalles dyadiques montre que $f_n \leq f_{n+1}$. Si $f(x) = +\infty$, alors $f_n(x) = n$ pour tout $n$ assez grand, donc $f_n(x) \to +\infty$. Si $f(x) < +\infty$, alors, pour $n$ assez grand, $f(x) < n$ et

$$0 \leq f(x) - f_n(x) \leq 2^{-n},$$

donc $f_n(x) \to f(x)$.

Si $f$ est positive et bornée, l'inégalité précédente est uniforme pour $n$ assez grand, d'où la convergence uniforme. Enfin, pour une fonction réelle de signe quelconque, on écrit $f = f^+ - f^-$ et l'on applique ce qui précède à $f^+$ et $f^-$. Dans le cas borné, les deux approximations peuvent être choisies uniformes. Le cas complexe se traite en approchant séparément les parties réelle et imaginaire. □

**Définition 3.5.** On note $\mathcal{E}_+(\mathcal{A})$ l'ensemble des fonctions étagées positives de $(X, \mathcal{A})$ dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Si $f \in \mathcal{E}_+(\mathcal{A})$, on appelle intégrale de $f$ relativement à $\mu$ la quantité

$$\mu(f) = \int_X f \, d\mu = \sum_{\alpha \in f(X)} \alpha \, \mu(\{f = \alpha\}),$$

avec la convention $0 \cdot (+\infty) = 0$.

*Remarque* 3.6. On note aussi cette intégrale sous les formes $\int_X f(x)\, d\mu(x)$ ou $\int_X f(x)\, \mu(dx)$. Si $f = \sum_{i\in I} \alpha_i \mathbf{1}_{A_i}$, où $(A_i)_{i\in I}$ est une partition finie mesurable de $X$ et $\alpha_i \geq 0$, alors

$$\int_X f \, d\mu = \sum_{i \in I} \alpha_i \, \mu(A_i).$$

Ainsi la définition ne dépend pas de la représentation choisie.

**Proposition 3.7** (Premières propriétés)**.** *Soient $f, g \in \mathcal{E}_+(\mathcal{A})$. Alors :*

(i) $\mu(f+g) = \mu(f) + \mu(g)$ ;

(ii) *si $f \leq g$, alors $\mu(f) \leq \mu(g)$ ;*

(iii) *pour tout* $a \geq 0$, $\mu(af) = a\mu(f)$.

*Démonstration.* Écrivons $f = \sum_{i\in I} \alpha_i \mathbf{1}_{A_i}$ et $g = \sum_{j\in J} \beta_j \mathbf{1}_{B_j}$ sur deux partitions finies mesurables de $X$. En raffinant par les intersections $A_i \cap B_j$, on obtient

$$f + g = \sum_{i,j} (\alpha_i + \beta_j) \mathbf{1}_{A_i \cap B_j}.$$

L'additivité finie de $\mu$ donne alors

$$\mu(f + g) = \sum_i \alpha_i \mu(A_i) + \sum_j \beta_j \mu(B_j) = \mu(f) + \mu(g).$$

Si $f \leq g$, alors $g = f + (g - f)$, où $g - f$ est étagée positive ; le point précédent donne $\mu(g) = \mu(f) + \mu(g - f) \geq \mu(f)$. L'homogénéité positive est immédiate. □

*Remarque* 3.8. Si $A \in \mathcal{A}$ et $f \in \mathcal{E}_+(\mathcal{A})$, alors $\mathbf{1}_A f \in \mathcal{E}_+(\mathcal{A})$. On pose

$$\int_A f \, d\mu = \int_X \mathbf{1}_A f \, d\mu.$$

*Remarque* 3.9 (Approximation par restriction à une suite croissante). Soit $(E_n)_{n\geq 1}$ une suite croissante d'éléments de $\mathcal{A}$ telle que $X = \bigcup_{n\geq 1} E_n$. Alors, pour toute fonction $f \in \mathcal{E}_+(\mathcal{A})$,

$$\int_X f \, d\mu = \lim_{n\to\infty} \uparrow \int_{E_n} f \, d\mu.$$

En effet, si $f = \sum_{i\in I} \alpha_i \mathbf{1}_{A_i}$, alors $\mathbf{1}_{E_n} f$ est encore étagée positive et

$$\int_{E_n} f \, d\mu = \sum_{i\in I} \alpha_i \, \mu(E_n \cap A_i).$$

Par continuité à gauche de la mesure, $\mu(E_n \cap A_i) \uparrow \mu(A_i)$ pour tout $i \in I$, d'où le résultat.

**Proposition 3.10** (Passage à la limite croissante pour les fonctions étagées)**.** *Soit* $(f_n)_{n\geq 1}$ *une suite croissante de fonctions de* $\mathcal{E}_+(\mathcal{A})$ *telle que* $f_n \uparrow f$*, où* $f \in \mathcal{E}_+(\mathcal{A})$*. Alors*

$$\int_X f_n \, d\mu \uparrow \int_X f \, d\mu.$$

*Démonstration.* Le cas $f = 0$ est immédiat. Sinon, par monotonie, $\int f_n \, d\mu \leq \int f \, d\mu$, donc la suite des intégrales a une limite $\ell \leq \int f \, d\mu$. Fixons $c \in ]0, 1[$ et posons

$$E_n = \{x \in X : \, f_n(x) \geq c f(x)\}.$$

Comme $f_n \leq f_{n+1}$, les ensembles $E_n$ sont croissants, et $\bigcup_n E_n = X$ : en effet, si $f(x) > 0$, alors $f_n(x)/f(x) \to 1$, tandis que si $f(x) = 0$, l'inégalité est automatique. Ainsi, par la remarque précédente,

$$\lim_n \int_X \mathbf{1}_{E_n} f \, d\mu = \int_X f \, d\mu.$$

Or $f_n \geq c f \mathbf{1}_{E_n}$, donc

$$\ell \geq c \lim_n \int_X \mathbf{1}_{E_n} f \, d\mu = c \int_X f \, d\mu.$$

En faisant tendre $c$ vers 1, on obtient $\ell = \int_X f \, d\mu$. □

## 3.2 Intégrale d'une fonction mesurable positive

**Définition 3.11.** On note $\mathcal{M}_+(\mathcal{A})$ l'ensemble des fonctions mesurables positives de $(X, \mathcal{A})$ dans $(\overline{\mathbb{R}}_+, \mathcal{B}(\overline{\mathbb{R}}_+))$, où $\overline{\mathbb{R}}_+ = [0, +\infty]$. Pour $f \in \mathcal{M}_+(\mathcal{A})$, on définit son intégrale relativement à $\mu$ par

$$\mu(f) = \int_X f \, d\mu = \sup\left\{ \int_X g \, d\mu \colon g \in \mathcal{E}_+(\mathcal{A}),\ 0 \le g \le f \right\}.$$

Si $\mu(f) < +\infty$, on dit que $f$ est intégrable.

**Lemme 3.12.** *Soit $f \in \mathcal{M}_+(\mathcal{A})$. Si $(f_n)_{n\ge1}$ est une suite croissante de fonctions de $\mathcal{E}_+(\mathcal{A})$ telle que $f_n \uparrow f$, alors*

$$\int_X f \, d\mu = \lim_{n\to\infty} \uparrow \int_X f_n \, d\mu.$$

*Démonstration.* Comme $f_n \le f$, on a $\int f_n \, d\mu \le \int f \, d\mu$, donc la limite $\ell = \lim_n \int f_n \, d\mu$ existe dans $\overline{\mathbb{R}}_+$ et vérifie $\ell \le \int f \, d\mu$. Réciproquement, soit $h \in \mathcal{E}_+(\mathcal{A})$ avec $h \le f$. Posons $h_n = \min(f_n, h)$. Alors $h_n \in \mathcal{E}_+(\mathcal{A})$, $h_n \uparrow h$, et le passage à la limite croissante pour les fonctions étagées donne

$$\int_X h \, d\mu = \lim_n \int_X h_n \, d\mu \le \ell.$$

En prenant le supremum sur tous les $h \le f$ étagés positifs, on obtient $\int f \, d\mu \le \ell$. □

**Proposition 3.13** (Premières propriétés)**.** *Soient $f, g \in \mathcal{M}_+(\mathcal{A})$. Alors :*

(i) *si $f \le g$, alors $\mu(f) \le \mu(g)$ ;*

(ii) *$\mu(f + g) = \mu(f) + \mu(g)$ ;*

(iii) *pour tout $a \ge 0$, $\mu(af) = a\mu(f)$.*

*Démonstration.* La croissance est immédiate d'après la définition. Pour l'additivité, prenons des suites croissantes de fonctions étagées positives $f_n \uparrow f$ et $g_n \uparrow g$. Alors $f_n + g_n \uparrow f + g$, et le lemme précédent donne

$$\mu(f + g) = \lim_n \mu(f_n + g_n) = \lim_n \big(\mu(f_n) + \mu(g_n)\big) = \mu(f) + \mu(g).$$

L'homogénéité positive se démontre de la même manière à partir de l'homogénéité pour les fonctions étagées. □

**Théorème 3.14** (Beppo-Levi, ou théorème de convergence monotone)**.** *Soit $(f_n)_{n\ge1}$ une suite croissante de fonctions de $\mathcal{M}_+(\mathcal{A})$. Alors $f = \lim_n f_n$ appartient à $\mathcal{M}_+(\mathcal{A})$ et*

$$\int_X f \, d\mu = \lim_{n\to\infty} \uparrow \int_X f_n \, d\mu.$$

*Démonstration.* La mesurabilité de $f$ a déjà été établie. Comme $f_n \le f$, on a $\int f_n \, d\mu \le \int f \, d\mu$, donc, en posant $\ell = \lim_n \int f_n \, d\mu$, on a $\ell \le \int f \, d\mu$.

Pour l'inégalité réciproque, choisissons, pour chaque $n$, une suite $(g_{n,p})_{p\ge1}$ de fonctions étagées positives telle que $g_{n,p} \uparrow f_n$ lorsque $p \to \infty$. Posons

$$h_p = \max_{1\le n\le p} g_{n,p}.$$

Alors $h_p \in \mathcal{E}_+(\mathcal{A})$ et $(h_p)$ est croissante. Si $h = \lim_p h_p$, on a $h \le f$. D'autre part, pour $p \ge n$, $h_p \ge g_{n,p}$, donc $h \ge f_n$ pour tout $n$, et finalement $h = f$. Le lemme précédent donne donc

$$\int_X f \, d\mu = \lim_p \int_X h_p \, d\mu.$$

Or $h_p \le f_p$, donc

$$\int_X f\,d\mu \le \lim_p \int_X f_p\,d\mu = \ell.$$

□

**Proposition 3.15** (Autres propriétés)**.** *Soit $f \in \mathcal{M}_+(\mathcal{A})$. Alors :*

(i) $\int_X f\,d\mu = 0$ *si et seulement si* $f = 0$ $\mu$*-p.p. ;*

(ii) *si* $g \in \mathcal{M}_+(\mathcal{A})$ *et* $f = g$ $\mu$*-p.p., alors* $\int_X f\,d\mu = \int_X g\,d\mu$ *;*

(iii) *pour tout* $A > 0$*,*

$$\mu(\{f \ge A\}) \le \frac{1}{A}\int_X f\,d\mu$$

(inégalité de Markov)*;*

(iv) *si* $\int_X f\,d\mu < +\infty$*, alors* $f < +\infty$ $\mu$*-p.p. ;*

(v) *si* $\mu(A) = 0$*, alors* $\int_A f\,d\mu = 0$*.*

*Démonstration.* Pour (i), le résultat est immédiat pour les fonctions étagées positives. Dans le cas général, prenons $f_n \in \mathcal{E}_+(\mathcal{A})$ avec $f_n \uparrow f$. Alors

$$\{f \ne 0\} = \{f > 0\} = \bigcup_{n\ge 1}\{f_n > 0\}.$$

Par continuité à gauche et par Beppo-Levi, $\mu(\{f \ne 0\}) = 0$ équivaut à $\mu(f_n) = 0$ pour tout $n$, ce qui équivaut à $\mu(f) = 0$.

Pour (ii), posons

$$h = \begin{cases} \max(f,g) - \min(f,g), & \text{sur } \{\min(f,g) < +\infty\}, \\ 0, & \text{sur } \{f = g = +\infty\}. \end{cases}$$

Alors $h \in \mathcal{M}_+(\mathcal{A})$ et $h = 0$ $\mu$-p.p. ; par (i), $\mu(h) = 0$. Comme $\max(f,g) = \min(f,g) + h$, les deux fonctions $\max(f,g)$ et $\min(f,g)$ ont la même intégrale. On en déduit, par monotonie, que $f$ et $g$ ont la même intégrale.

Pour (iii), on utilise l'inégalité ponctuelle $f \ge A\mathbf{1}_{\{f\ge A\}}$, qui donne $\int f\,d\mu \ge A\mu(\{f \ge A\})$. Pour (iv), on écrit $\{f = +\infty\} = \bigcap_{n\ge 1}\{f \ge n\}$ et l'on applique la continuité à droite à la suite décroissante $\{f \ge n\}$, en utilisant (iii). Enfin, (v) résulte de (i) appliqué à $f\mathbf{1}_A$. □

*Remarque* 3.16. L'implication réciproque de (iv) est fausse : une fonction peut être finie $\mu$-p.p. sans être intégrable. Par exemple, pour la mesure de comptage sur $\mathbb{N}$, la fonction $f(0) = 0$ et $f(n) = 1/n$ pour $n \ge 1$ est finie partout, mais $\int_{\mathbb{N}} f\,d\mu = \sum_{n\ge 1} 1/n = +\infty$.

**Exemple 3.17** (Mesure de Dirac)**.** Pour la mesure de Dirac $\delta_a$, on a

$$\int_X f\,d\delta_a = f(a), \qquad f \in \mathcal{M}_+(\mathcal{A}).$$

En effet, c'est immédiat pour les indicatrices, puis pour les fonctions étagées, et enfin pour les fonctions mesurables positives par passage à la limite croissante.

**Exemple 3.18** (Mesure de comptage sur $\mathbb{N}$)**.** Si $\mu = \sum_{n\in\mathbb{N}} \delta_n$ est la mesure de comptage sur $\mathbb{N}$, alors, pour toute fonction $f\colon \mathbb{N} \to \overline{\mathbb{R}}_+$,

$$\int_{\mathbb{N}} f\,d\mu = \sum_{n\in\mathbb{N}} f(n).$$

**Proposition 3.19.** *Soit $(f_n)_{n\geq 0}$ une suite de fonctions de $\mathcal{M}_+(\mathcal{A})$ et $f = \sum_{n\geq 0} f_n$. Alors*

$$\int_X f\, d\mu = \sum_{n\geq 0} \int_X f_n\, d\mu.$$

*Démonstration.* On applique le théorème de Beppo-Levi aux sommes partielles $S_N = \sum_{n=0}^{N} f_n$. Comme $S_N \uparrow f$, on obtient

$$\int_X f\, d\mu = \lim_{N\to\infty} \int_X S_N\, d\mu = \lim_{N\to\infty} \sum_{n=0}^{N} \int_X f_n\, d\mu.$$

□

**Exemple 3.20.** Pour la mesure de comptage sur $\mathbb{N}$, si $f_n \colon \mathbb{N} \to \overline{\mathbb{R}}_+$ et $f = \sum_n f_n$, alors

$$\sum_{k\in\mathbb{N}} f(k) = \sum_{k\in\mathbb{N}} \sum_{n\in\mathbb{N}} f_n(k) = \sum_{n\in\mathbb{N}} \sum_{k\in\mathbb{N}} f_n(k).$$

**Lemme 3.21** (Lemme de Fatou)**.** *Soit $(f_n)_{n\geq 1}$ une suite de fonctions de $\mathcal{M}_+(\mathcal{A})$. Alors*

$$\int_X \liminf_{n\to\infty} f_n\, d\mu \leq \liminf_{n\to\infty} \int_X f_n\, d\mu.$$

*Démonstration.* Posons

$$g_p = \inf_{n\geq p} f_n.$$

Alors $(g_p)_{p\geq 1}$ est une suite croissante de fonctions de $\mathcal{M}_+(\mathcal{A})$ et $g_p \uparrow \liminf_n f_n$. Par Beppo-Levi,

$$\int_X \liminf_n f_n\, d\mu = \lim_p \int_X g_p\, d\mu.$$

Comme $g_p \leq f_n$ pour tout $n \geq p$, on a

$$\int_X g_p\, d\mu \leq \inf_{n\geq p} \int_X f_n\, d\mu.$$

On conclut en faisant tendre $p$ vers l'infini. □

## 3.3 Mesures à densité

**Rappel 3.22.** Dès que l'intégrale est définie, on pose

$$\int_A f\, d\mu = \int_X f \mathbf{1}_A\, d\mu.$$

**Théorème 3.23** (Mesure à densité)**.** *Soit $f \colon (X, \mathcal{A}) \to [0, +\infty]$ une fonction mesurable positive. Pour $A \in \mathcal{A}$, posons*

$$\nu(A) = \int_A f\, d\mu.$$

*Alors $\nu$ est une mesure sur $(X, \mathcal{A})$. De plus, pour toute fonction mesurable positive $g \colon (X, \mathcal{A}) \to [0, +\infty]$, on a*

$$\int_X g\, d\nu = \int_X fg\, d\mu.$$

*Démonstration.* La fonction $\nu$ est bien à valeurs dans $\overline{\mathbb{R}}_+$. On a $\nu(\emptyset) = 0$. Si $(A_n)_{n\geq 1}$ est une suite d'éléments deux à deux disjoints de $\mathcal{A}$, alors

$$\mathbf{1}_{\bigcup_{n\geq 1} A_n} = \lim_{N\to\infty} \uparrow \sum_{n=1}^{N} \mathbf{1}_{A_n}.$$

Par Beppo-Levi,

$$\nu\left(\bigcup_{n\geq 1} A_n\right) = \int_X f\mathbf{1}_{\bigcup_n A_n}\, d\mu = \sum_{n\geq 1} \int_X f\mathbf{1}_{A_n}\, d\mu = \sum_{n\geq 1} \nu(A_n),$$

ce qui montre que $\nu$ est une mesure.

L'identité $\int g\, d\nu = \int fg\, d\mu$ est vraie pour $g = \mathbf{1}_A$ par définition de $\nu$. Par linéarité positive, elle est vraie pour toute fonction étagée positive $g$. Enfin, si $g \in \mathcal{M}_+(\mathcal{A})$, on prend une suite $g_n \in \mathcal{E}_+(\mathcal{A})$ telle que $g_n \uparrow g$, et l'on applique Beppo-Levi des deux côtés. □

*Remarque* 3.24. La mesure $\nu$ est souvent notée $f\mu$ ou $f\, d\mu$. On dit que $f$ est une densité de $\nu$ par rapport à $\mu$, ou encore la dérivée de Radon–Nikodym de $\nu$ par rapport à $\mu$ lorsque cette notation est justifiée ; on écrit alors souvent $f = d\nu/d\mu$.

## 3.4 Intégrale des fonctions de signe quelconque

Soit $f\colon (X, \mathcal{A}) \to \overline{\mathbb{R}}$ une fonction mesurable. On écrit

$$f = f^+ - f^-, \qquad f^+ = \max(f, 0), \qquad f^- = \max(-f, 0).$$

Les fonctions $f^+$ et $f^-$ sont mesurables positives, et l'on a $|f| = f^+ + f^-$. On aimerait poser $\int f\, d\mu = \int f^+\, d\mu - \int f^-\, d\mu$, mais cette expression n'a un sens que si l'on n'est pas dans la situation indéterminée $+\infty - (+\infty)$.

**Définition 3.25.** On note $\mathcal{L}^1(\mu)$, ou $\mathcal{L}^1(X, \mathcal{A}, \mu)$, l'ensemble des fonctions mesurables $f\colon (X, \mathcal{A}) \to \overline{\mathbb{R}}$ telles que

$$\int_X |f|\, d\mu < +\infty.$$

Si $f \in \mathcal{L}^1(\mu)$, on définit son intégrale par

$$\int_X f\, d\mu = \int_X f^+\, d\mu - \int_X f^-\, d\mu.$$

*Remarque* 3.26. Si $f \in \mathcal{L}^1(\mu)$, alors $f$ est finie $\mu$-presque partout. En effet, $\{f = \pm\infty\} = \{|f| = +\infty\}$ est de mesure nulle dès que $\int |f|\, d\mu < +\infty$.

**Proposition 3.27** (Premières propriétés)**.** *Soient $f, g \in \mathcal{L}^1(\mu)$. Alors :*

(i) *pour tout $a \in \mathbb{R}$, $af + g \in \mathcal{L}^1(\mu)$ et*

$$\int_X (af + g)\, d\mu = a \int_X f\, d\mu + \int_X g\, d\mu;$$

(ii) *si $f \leq g$, alors $\int_X f\, d\mu \leq \int_X g\, d\mu$ ;*

(iii) *si $(f_n)_{n\geq 1}$ est une suite croissante de fonctions intégrables telle que $f_n \uparrow f$ et $f \in \mathcal{L}^1(\mu)$, alors*

$$\int_X f_n\, d\mu \uparrow \int_X f\, d\mu;$$

(iv) $\left|\int_X f\,d\mu\right| \le \int_X |f|\,d\mu$ ;

(v) *si* $\mu(A) = 0$, *alors* $\int_A |f|\,d\mu = 0$, *et donc* $f\mathbf{1}_A = 0$ $\mu$*-p.p.*

*Démonstration.* La linéarité se déduit des identités $f + g = (f + g)^+ - (f + g)^-$ et des propriétés de l'intégrale positive ; une façon rapide de procéder est de regrouper les parties positives et négatives pour ne faire intervenir que des fonctions positives. La monotonie vient de $g - f \ge 0$, donc $\int (g - f)\,d\mu \ge 0$. Le passage à la limite croissante s'obtient en appliquant Beppo-Levi à $g_n = f_n - f_1 \ge 0$. Enfin, $-|f| \le f \le |f|$ donne l'inégalité triangulaire pour l'intégrale. Le dernier point est une conséquence des propriétés déjà établies pour les fonctions positives. □

*Remarque* 3.28 (Modification sur un ensemble négligeable). Si $f$ et $g$ sont mesurables à valeurs dans $\overline{\mathbb{R}}$ et si $f = g$ $\mu$-p.p., alors $f \in \mathcal{L}^1(\mu)$ si et seulement si $g \in \mathcal{L}^1(\mu)$ ; dans ce cas,

$$\int_X f\,d\mu = \int_X g\,d\mu.$$

Ainsi l'intégrale ne voit pas les modifications sur un ensemble de mesure nulle.

**Définition 3.29** (Fonctions à valeurs complexes)**.** Soit $f\colon (X, \mathcal{A}) \to \mathbb{C}$ mesurable. On dit que $f \in \mathcal{L}^1(\mu)$ si $\operatorname{Re} f$ et $\operatorname{Im} f$ appartiennent à $\mathcal{L}^1(\mu)$, et l'on pose alors

$$\int_X f\,d\mu = \int_X \operatorname{Re} f\,d\mu + i\int_X \operatorname{Im} f\,d\mu.$$

## 3.5 Intégrale par rapport à la mesure de Lebesgue

On note $\lambda$ la mesure de Lebesgue sur $\mathbb{R}$, ou sur un intervalle muni de la tribu borélienne. Le résultat suivant fait le lien avec l'intégrale de Riemann.

**Théorème 3.30** (Compatibilité avec l'intégrale de Riemann)**.** *Soit* $[a, b]$ *un intervalle borné de* $\mathbb{R}$*. Si* $f\colon [a, b] \to \mathbb{R}$ *est bornée, mesurable et Riemann-intégrable, alors* $f\mathbf{1}_{[a,b]} \in \mathcal{L}^1(\lambda)$ *et*

$$\int_{[a,b]} f\,d\lambda = \int_a^b f(x)\,dx,$$

*où le membre de droite désigne l'intégrale de Riemann.*

*Démonstration.* Soit $\tau = \big(a = t_0 < t_1 < \cdots < t_n = b\big)$ une subdivision. Posons

$$m_i = \inf_{x\in[t_{i-1},t_i]} f(x), \qquad M_i = \sup_{x\in[t_{i-1},t_i]} f(x),$$

et définissons deux fonctions étagées $\ell_\tau$ et $u_\tau$ telles que $\ell_\tau \le f \le u_\tau$ et

$$\int \ell_\tau\,d\lambda = \sum_{i=1}^n m_i(t_i - t_{i-1}), \qquad \int u_\tau\,d\lambda = \sum_{i=1}^n M_i(t_i - t_{i-1}).$$

Ces deux quantités sont les sommes inférieure et supérieure de Riemann. En faisant tendre le pas de la subdivision vers 0 et en utilisant l'intégrabilité de Riemann, on obtient l'égalité des deux intégrales. □

*Remarque* 3.31 (Intégrales impropres). Soit $]a, b[$ un intervalle, avec éventuellement $a = -\infty$ ou $b = +\infty$, et soit $f\colon ]a, b[\to \mathbb{R}$ mesurable. Si $f$ est bornée et Riemann-intégrable sur tout segment $[c, d] \subset ]a, b[$, et si son intégrale de Riemann impropre converge absolument, alors $f \in \mathcal{L}^1(\lambda)$ sur $]a, b[$, et l'intégrale de Lebesgue coïncide avec cette intégrale impropre.

## 3.6 Théorème de convergence dominée et applications

**Théorème 3.32** (Convergence dominée de Lebesgue)**.** *Soit $(f_n)_{n\geq 1}$ une suite de fonctions mesurables définies sur $(X, \mathcal{A}, \mu)$ et à valeurs réelles ou complexes. On suppose que :*

(i) $f_n \to f$ *$\mu$-p.p. ;*

(ii) *il existe une fonction $g \in \mathcal{L}^1(\mu)$ telle que $|f_n| \leq g$ $\mu$-p.p. pour tout $n$.*

*Alors $f \in \mathcal{L}^1(\mu)$ et*

$$\lim_{n\to\infty} \int_X f_n \, d\mu = \int_X f \, d\mu.$$

*De plus,*

$$\lim_{n\to\infty} \int_X |f_n - f| \, d\mu = 0.$$

*Démonstration.* On peut supposer les fonctions réelles. Sur l'ensemble de convergence, on a $|f| = \lim_n |f_n| \leq g$, donc $f \in \mathcal{L}^1(\mu)$. Posons $A$ l'ensemble où $f_n \to f$ et où $g < +\infty$ ; alors $\mu(A^c) = 0$. Les fonctions

$$h_n = (2g - \mathbf{1}_A |f_n - f|)_{n\geq 1}$$

sont mesurables positives. Par le lemme de Fatou,

$$\int_X 2g \, d\mu \leq \liminf_{n\to\infty} \int_X (2g - \mathbf{1}_A |f_n - f|) \, d\mu.$$

On en déduit

$$0 \leq \limsup_{n\to\infty} \int_X |f_n - f| \, d\mu \leq 0.$$

Ainsi $\int |f_n - f| \, d\mu \to 0$, ce qui implique la convergence des intégrales par l'inégalité

$$\left| \int_X f_n \, d\mu - \int_X f \, d\mu \right| \leq \int_X |f_n - f| \, d\mu.$$

□

**Exemple 3.33.** Pour

$$I_n = \int_0^1 n^2 x^2 e^{-n^2 x^2} \, d\lambda(x),$$

on a $n^2 x^2 e^{-n^2 x^2} \to 0$ pour tout $x \in [0, 1]$ et $\sup_{n\geq 1} n^2 x^2 e^{-n^2 x^2} \leq C$ pour une constante $C$ indépendante de $x$. Par convergence dominée, $I_n \to 0$.

**Exemple 3.34.** L'hypothèse de domination est essentielle. Pour

$$J_n = \int_0^1 n^2 x e^{-n^2 x^2} \, d\lambda(x),$$

l'intégrande converge ponctuellement vers 0, mais

$$J_n = \left[ -\frac{1}{2} e^{-n^2 x^2} \right]_0^1 = \frac{1}{2}(1 - e^{-n^2}) \longrightarrow \frac{1}{2}.$$

Il n'y a donc pas de domination intégrable uniforme.

**Exemple 3.35.** Si $f \in \mathcal{L}^1(\mu)$, alors

$$\int_{\{|f|>t\}} f \, d\mu \longrightarrow 0 \qquad \text{lorsque } t \to +\infty.$$

En effet, $f\mathbf{1}_{\{|f|>t\}} \to 0$ p.p. et $\left|f\mathbf{1}_{\{|f|>t\}}\right| \leq |f|$.

**Proposition 3.36** (Séries de fonctions)**.** *Soit $(f_n)_{n\geq 1}$ une suite de fonctions mesurables. Si*

$$\sum_{n\geq 1} \int_X |f_n| \, d\mu < +\infty,$$

*alors la série $\sum_{n\geq 1} f_n$ converge absolument $\mu$-p.p., sa somme est intégrable, et*

$$\int_X \sum_{n\geq 1} f_n \, d\mu = \sum_{n\geq 1} \int_X f_n \, d\mu.$$

*Démonstration.* Par Beppo-Levi appliqué aux fonctions positives $|f_n|$,

$$\int_X \sum_{n\geq 1} |f_n| \, d\mu = \sum_{n\geq 1} \int_X |f_n| \, d\mu < +\infty.$$

Donc $\sum_n |f_n| < +\infty$ $\mu$-p.p. On applique ensuite le théorème de convergence dominée aux sommes partielles, dominées par $\sum_n |f_n|$. □

**Lemme 3.37** (Borel–Cantelli)**.** *Soit $(A_n)_{n\geq 1}$ une suite d'éléments de $\mathcal{A}$. Si $\sum_{n\geq 1} \mu(A_n) < +\infty$, alors*

$$\mu\left(\limsup_{n\to\infty} A_n\right) = 0.$$

*Démonstration.* On applique la proposition précédente à la série de fonctions indicatrices :

$$\int_X \sum_{n\geq 1} \mathbf{1}_{A_n} \, d\mu = \sum_{n\geq 1} \mu(A_n) < +\infty.$$

Ainsi $\sum_n \mathbf{1}_{A_n} < +\infty$ p.p., ce qui signifie que presque tout point n'appartient qu'à un nombre fini de $A_n$. □

**Théorème 3.38** (Continuité sous le signe intégrale)**.** *Soit $I$ un intervalle de $\mathbb{R}$ et soit $f : I \times X \to \mathbb{R}$, $(t,x) \mapsto f(t,x)$. Pour $t \in I$, posons*

$$F(t) = \int_X f(t,x) \, d\mu(x).$$

*On suppose que :*

(i) *pour tout $t \in I$, la fonction $x \mapsto f(t,x)$ est mesurable ;*

(ii) *pour $\mu$-presque tout $x$, la fonction $t \mapsto f(t,x)$ est continue en un point $t_0 \in I$ ;*

(iii) *il existe $g \in \mathcal{L}^1(\mu)$ telle que, pour tout $t \in I$, $|f(t,x)| \leq g(x)$ pour $\mu$-presque tout $x$.*

*Alors $F$ est bien définie sur $I$ et continue en $t_0$.*

*Démonstration.* Si $t_n \to t_0$, alors $f(t_n,x) \to f(t_0,x)$ pour $\mu$-presque tout $x$, et $|f(t_n,x)| \leq g(x)$. Le théorème de convergence dominée donne $F(t_n) \to F(t_0)$. □

**Théorème 3.39** (Dérivation sous le signe intégrale)**.** *Soit $I$ un intervalle ouvert non vide de $\mathbb{R}$ et soit $f : I \times X \to \mathbb{R}$. Soit $t_0 \in I$. On suppose que :*

(i) *pour tout $t \in I$, $x \mapsto f(t,x)$ est $\mu$-intégrable ;*

(ii) *pour $\mu$-presque tout $x$, la dérivée partielle $\partial_t f(t_0,x)$ existe ;*

(iii) *il existe $g \in \mathcal{L}^1(\mu)$ telle que, pour tout $t \in I$,*

$$|f(t,x) - f(t_0,x)| \leq g(x)\,|t - t_0|$$

*pour $\mu$-presque tout $x$.*

*Alors* $F(t) = \int_X f(t,x)\,d\mu(x)$ *est dérivable en* $t_0$ *et*

$$F'(t_0) = \int_X \partial_t f(t_0,x)\,d\mu(x).$$

*Démonstration.* Soit $(t_n)$ une suite de points de $I$, distincts de $t_0$, telle que $t_n \to t_0$. Posons

$$\varphi_n(x) = \frac{f(t_n,x) - f(t_0,x)}{t_n - t_0}.$$

Alors $\varphi_n(x) \to \partial_t f(t_0,x)$ pour $\mu$-presque tout $x$, et $|\varphi_n(x)| \le g(x)$. Par convergence dominée,

$$\frac{F(t_n) - F(t_0)}{t_n - t_0} = \int_X \varphi_n\,d\mu \longrightarrow \int_X \partial_t f(t_0,x)\,d\mu(x).$$

□

**Corollaire 3.40** (Dérivation globale sur un intervalle ouvert)**.** *Soit* $I$ *un intervalle ouvert de* $\mathbb{R}$*. On suppose que :*

(i) *pour tout* $t \in I$, $f(t,\cdot)$ *est* $\mu$*-intégrable ;*

(ii) *pour* $\mu$*-presque tout* $x$, $t \mapsto f(t,x)$ *est dérivable sur* $I$ *;*

(iii) *il existe* $g \in \mathcal{L}^1(\mu)$ *telle que, pour tout* $t \in I$, $|\partial_t f(t,x)| \le g(x)$ *pour* $\mu$*-presque tout* $x$*.*

*Alors* $F(t) = \int_X f(t,x)\,d\mu(x)$ *est dérivable sur* $I$ *et*

$$F'(t) = \int_X \partial_t f(t,x)\,d\mu(x), \qquad t \in I.$$

*Démonstration.* Fixons $t_0 \in I$. Par le théorème des accroissements finis, pour $t$ assez proche de $t_0$ et pour presque tout $x$,

$$|f(t,x) - f(t_0,x)| \le \sup_{s \in [t \wedge t_0, t \vee t_0]} |\partial_t f(s,x)|\ |t - t_0| \le g(x)\,|t - t_0|\,.$$

On applique le théorème précédent en $t_0$. □

*Remarque* 3.41. Les résultats de continuité et de dérivation sous le signe intégrale restent valables pour des fonctions à valeurs complexes, en les appliquant séparément aux parties réelle et imaginaire.

# Chapitre 4

# Espaces $L^p$ et $\ell^p$

Dans ce chapitre, on fixe un espace mesuré $(X, \mathcal{A}, \mu)$ et un exposant $1 \leq p \leq +\infty$. Les fonctions sont à valeurs réelles ou complexes, sauf mention contraire. Le point important est que l'intégrale permet de mesurer la "taille moyenne" d'une fonction, mais seulement à modification sur un ensemble négligeable près. C'est pourquoi les véritables espaces normés seront des espaces de classes de fonctions égales presque partout.

## 4.1 Espaces $\mathcal{L}^p$

**Définition 4.1.** Soit $1 \leq p < +\infty$. On note $\mathcal{L}^p = \mathcal{L}^p(X, \mathcal{A}, \mu)$ l'ensemble des fonctions mesurables $f\colon X \to \mathbb{K}$, où $\mathbb{K} = \mathbb{R}$ ou $\mathbb{C}$, telles que $|f|^p$ soit $\mu$-intégrable. On pose

$$\|f\|_p = \left(\int_X |f|^p \, d\mu\right)^{1/p}.$$

**Proposition 4.2.** *Pour $1 \leq p < +\infty$, $\mathcal{L}^p$ est un espace vectoriel.*

*Démonstration.* La fonction nulle appartient à $\mathcal{L}^p$. Si $f, g \in \mathcal{L}^p$ et $\lambda \in \mathbb{K}$, alors

$$|\lambda f + g|^p \leq (|\lambda|\,|f| + |g|)^p \leq 2^p \left(|\lambda|^p\,|f|^p + |g|^p\right),$$

ce qui est intégrable. Donc $\lambda f + g \in \mathcal{L}^p$. □

On a immédiatement $\|\lambda f\|_p = |\lambda|\,\|f\|_p$. En revanche, $\|f\|_p = 0$ entraîne seulement $f = 0$ $\mu$-presque partout. Ainsi $\|\cdot\|_p$ est d'abord une semi-norme sur $\mathcal{L}^p$. Pour obtenir l'inégalité triangulaire, on utilise l'inégalité de Hölder.

**Théorème 4.3** (Inégalité de Hölder)**.** *Soient $p, q \in ]1, +\infty[$ tels que $1/p + 1/q = 1$. Si $f \in \mathcal{L}^p$ et $g \in \mathcal{L}^q$, alors $fg \in \mathcal{L}^1$ et*

$$\int_X |fg| \, d\mu \leq \|f\|_p \, \|g\|_q.$$

*En particulier, lorsque $p = q = 2$, on retrouve l'inégalité de Cauchy–Schwarz.*

*Démonstration.* On part de l'inégalité de Young : pour $u, v \geq 0$ et $\alpha \in ]0, 1[$,

$$u^\alpha v^{1-\alpha} \leq \alpha u + (1-\alpha)v.$$

Elle découle de la concavité de log, ou encore de la convexité de $x \mapsto -\log x$. Si $\int |f|^p \, d\mu = 0$ ou $\int |g|^q \, d\mu = 0$, le résultat est immédiat. Sinon, on applique Young à

$$u = \frac{|f(x)|^p}{\int_X |f|^p \, d\mu}, \qquad v = \frac{|g(x)|^q}{\int_X |g|^q \, d\mu}, \qquad \alpha = \frac{1}{p}.$$

Après intégration, on obtient

$$\frac{1}{\|f\|_p \|g\|_q} \int_X |fg| \, d\mu \leq 1,$$

ce qui donne l'inégalité annoncée. □

*Remarque* 4.4. Dans le cas non trivial, l'égalité dans Hölder a lieu lorsque $|f|^p$ et $|g|^q$ sont proportionnelles presque partout, plus précisément lorsqu'il existe $\alpha, \beta \geq 0$, non tous deux nuls, tels que $\alpha |f|^p = \beta |g|^q$ $\mu$-p.p.

**Théorème 4.5** (Inégalité de Minkowski)**.** *Pour* $1 \leq p < +\infty$ *et* $f, g \in \mathcal{L}^p$*, on a*

$$\|f + g\|_p \leq \|f\|_p + \|g\|_p .$$

*Démonstration.* Le cas $p = 1$ est l'inégalité triangulaire intégrée. Soit donc $p > 1$ et soit $q$ l'exposant conjugué de $p$. On écrit

$$|f + g|^p \leq |f| \, |f + g|^{p-1} + |g| \, |f + g|^{p-1} .$$

En appliquant Hölder aux deux termes et en remarquant que $(p - 1)q = p$, on obtient

$$\|f + g\|_p^p \leq \left( \|f\|_p + \|g\|_p \right) \|f + g\|_p^{p-1} .$$

Si $\|f + g\|_p = 0$, il n'y a rien à montrer ; sinon on divise par $\|f + g\|_p^{p-1}$. □

**Corollaire 4.6.** *Pour tout* $p \in [1, +\infty[$*,* $\mathcal{L}^p$ *est un espace vectoriel semi-normé pour la semi-norme* $\|\cdot\|_p$*.*

**Définition 4.7** (Espace $\mathcal{L}^\infty$)**.** On note $\mathcal{L}^\infty = \mathcal{L}^\infty(X, \mathcal{A}, \mu)$ l'ensemble des fonctions mesurables $f$ essentiellement bornées, c'est-à-dire telles qu'il existe $a \in \mathbb{R}_+$ vérifiant $|f| \leq a$ $\mu$-p.p. On pose

$$\|f\|_\infty = \inf\{a \in \mathbb{R}_+ : \ |f| \leq a \quad \mu\text{-p.p.}\}.$$

Cette quantité est appelée borne supérieure essentielle de $|f|$.

*Remarque* 4.8. Comme pour $p < +\infty$, $\|\cdot\|_\infty$ est seulement une semi-norme sur $\mathcal{L}^\infty$, et l'on a

$$\|f + g\|_\infty \leq \|f\|_\infty + \|g\|_\infty .$$

donc $f \mapsto \|f\|_p$ est une semi-norme pour tout $1 \leq p \leq +\infty$.

## 4.2 Espaces $L^p$

Le défaut de $\mathcal{L}^p$ est que $\|f\|_p = 0$ implique seulement $f = 0$ presque partout. On passe donc aux classes de fonctions modulo égalité presque partout.

**Définition 4.9.** On note $\mathcal{N}$ l'ensemble des fonctions mesurables $f$ telles que $f = 0$ $\mu$-p.p. Sur les fonctions mesurables, on définit la relation

$$f \mathcal{R} g \quad \Longleftrightarrow \quad f - g \in \mathcal{N},$$

où encore $f = g$ $\mu$-p.p. C'est une relation d'équivalence. On note $\widetilde{f}$ la classe d'équivalence de $f$.

Si $f \mathcal{R} f'$ et $g \mathcal{R} g'$, alors $f + g \mathcal{R} f' + g'$ et $\lambda f \mathcal{R} \lambda f'$. De plus $\|f\|_p = \|f'\|_p$ dès que $f = f'$ $\mu$-p.p. On peut donc poser

$$\left\| \widetilde{f} \right\|_p = \|f\|_p .$$

**Définition 4.10.** On note $L^p = L^p(X, \mathcal{A}, \mu)$ l'ensemble des classes d'équivalence de fonctions de $\mathcal{L}^p$ pour la relation d'égalité presque partout. On munit $L^p$ de la norme $\|\cdot\|_p$. En pratique, on confond souvent une fonction et sa classe d'équivalence.

**Théorème 4.11** (Complétude de $L^p$)**.** *Pour tout $p \in [1, +\infty]$, l'espace $(L^p, \|\cdot\|_p)$ est complet. C'est donc un espace de Banach.*

*Démonstration.* On traite d'abord le cas $1 \leq p < +\infty$. Il suffit de rappeler qu'un espace vectoriel normé est complet si toute série normalement convergente y est convergente. On utilisera le fait suivant, conséquence de Beppo–Levi et de Minkowski : si $(f_n)_{n\geq 1}$ est une suite de fonctions mesurables positives et

$$\sum_{n\geq 1} \|f_n\|_p < +\infty,$$

alors

$$\left\| \sum_{n\geq 1} f_n \right\|_p \leq \sum_{n\geq 1} \|f_n\|_p .$$

Soit donc $(f_n)$ une suite de $L^p$ telle que $\sum_n \|f_n\|_p < +\infty$. Alors $\sum_n |f_n(x)| < +\infty$ pour presque tout $x$. Sur cet ensemble, la série $\sum_n f_n(x)$ converge dans $\mathbb{K}$ ; on note $S$ sa somme, et on pose $S = 0$ ailleurs. On a $S \in L^p$, et

$$\left\| S - \sum_{k=0}^{n} f_k \right\|_p \leq \sum_{k\geq n+1} \|f_k\|_p \longrightarrow 0.$$

Cela montre la complétude de $L^p$.

Pour $p = +\infty$, soit $(f_n)$ une série telle que $\sum_n \|f_n\|_\infty < +\infty$. Pour chaque $n$, il existe un ensemble négligeable $E_n$ tel que $|f_n(x)| \leq \|f_n\|_\infty$ pour $x \notin E_n$. Sur le complémentaire de $E = \bigcup_n E_n$, la série $\sum_n f_n(x)$ converge absolument et uniformément au sens de la norme essentielle. Si $S$ désigne sa somme, alors

$$\left\| S - \sum_{k=0}^{n} f_k \right\|_\infty \leq \sum_{k\geq n+1} \|f_k\|_\infty \longrightarrow 0.$$

Donc $L^\infty$ est complet. □

## 4.3 Liens entre convergence p.p. et convergence $L^p$

On sait déjà que, si $1 \leq p < +\infty$, si $f_n \to f$ $\mu$-p.p. et s'il existe $g \in L^p$ telle que $|f_n| \leq g$ $\mu$-p.p. pour tout $n$, alors $f_n \to f$ dans $L^p$. En effet, on applique le théorème de convergence dominée à $h_n = |f_n - f|^p$.

La réciproque est fausse en général, mais on dispose du résultat suivant.

**Théorème 4.12.** *Soit $1 \leq p < +\infty$. Si $(f_n)$ est une suite de $L^p$ qui converge vers $f$ pour $\|\cdot\|_p$, alors on peut extraire une sous-suite $(f_{n_k})_k$ qui converge $\mu$-presque partout vers $f$.*

*Démonstration.* Comme $\|f_n - f\|_p \to 0$, on choisit par récurrence une sous-suite $(f_{n_k})$ telle que

$$\left\|f_{n_k} - f\right\|_p^p \leq 2^{-k}.$$

Pour $q \geq 1$, posons

$$A_{k,q} = \left\{ \left|f_{n_k} - f\right|^p > \frac{1}{q} \right\}.$$

Par l'inégalité de Markov,

$$\mu(A_{k,q}) \leq q \left\|f_{n_k} - f\right\|_p^p \leq q 2^{-k}.$$

Le lemme de Borel–Cantelli donne $\mu(\limsup_k A_{k,q}) = 0$ pour chaque $q$. Ainsi, hors d'un ensemble négligeable, pour tout $q$ fixé, il n'y a qu'un nombre fini de $k$ tels que $x \in A_{k,q}$. Cela signifie que $f_{n_k}(x) \to f(x)$. □

**Corollaire 4.13.** *Soit $1 \le p < +\infty$. Si $f_n \to f$ dans $L^p$ et si $f_n \to g$ $\mu$-p.p., alors $f = g$ $\mu$-p.p.*

*Démonstration.* D'après le théorème précédent, on extrait une sous-suite de $(f_n)$ qui converge $\mu$-p.p. vers $f$. Cette même sous-suite converge aussi $\mu$-p.p. vers $g$, donc $f = g$ $\mu$-p.p. □

**Exemple 4.14** (Convergence p.p. sans convergence $L^1$)**.** Sur $([0,1], \mathcal{B}([0,1]), \lambda)$, posons $f_n = n\mathbf{1}_{[0,1/n]}$. Alors $f_n \to 0$ $\lambda$-p.p., mais

$$\int_0^1 f_n \, d\lambda = 1$$

pour tout $n$. Il n'y a donc pas convergence vers 0 dans $L^1$.

**Exemple 4.15** (Convergence $L^1$ sans convergence p.p.)**.** Sur $[0,1]$, posons $u_n = \sum_{i=1}^n 1/i$. On choisit $A_n \subset [0,1]$ de longueur $u_{n+1} - u_n = 1/(n+1)$, en faisant glisser ces intervalles le long de $[0,1]$ de façon périodique, et l'on pose $f_n = \mathbf{1}_{A_n}$. Alors $\|f_n\|_1 = \lambda(A_n) = 1/(n+1) \to 0$, donc $f_n \to 0$ dans $L^1$. En revanche, les ensembles $A_n$ passent infiniment souvent devant tout point ; la suite ne converge donc pas presque partout vers 0.

## 4.4 Quelques théorèmes de densité dans les espaces $L^p$

**Rappel 4.16.** Une fonction étagée $\varphi$ est intégrable si et seulement si

$$\mu(\{\varphi \neq 0\}) < +\infty,$$

avec la convention $0 \cdot \mu(A) = 0$ pour tout $A \in \mathcal{A}$. Dans ce cas, $\varphi \in L^p(\mu)$ pour tout $p \in [1, +\infty[$.

**Proposition 4.17.** *Pour tout $p \in [1, +\infty[$, l'ensemble des fonctions étagées intégrables est dense dans $L^p(\mu)$.*

*Démonstration.* Il suffit de traiter le cas d'une fonction $f \in L^p(\mu)$ positive. Par le lemme fondamental d'approximation, il existe une suite croissante $(\varphi_n)$ de fonctions étagées positives telle que $0 \le \varphi_n \le f$ et $\varphi_n \uparrow f$. Alors

$$|f - \varphi_n|^p \le f^p, \qquad |f - \varphi_n|^p \longrightarrow 0.$$

Le théorème de convergence dominée donne $\|f - \varphi_n\|_p \to 0$. Le cas général s'obtient en appliquant l'argument aux parties positive et négative, ou aux parties réelle et imaginaire. □

**Lemme 4.18.** *Soient $C \subset \mathcal{D} \subset L^p(\mu)$. Si $C$ est dense dans $\mathcal{D}$ et si $\mathcal{D}$ est dense dans $L^p(\mu)$, alors $C$ est dense dans $L^p(\mu)$.*

**Théorème 4.19** (Densité dans $L^p(\mathbb{R})$)**.** *On se place sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$ et $1 \le p < +\infty$.*

(a) *L'ensemble des fonctions en escalier à support compact est dense dans $L^p(\lambda)$.*

(b) *L'ensemble $C_c(\mathbb{R})$ des fonctions continues à support compact est dense dans $L^p(\lambda)$.*

*Démonstration.* On approche d'abord une fonction étagée intégrable par une fonction en escalier à support compact. Il suffit pour cela d'approcher les indicatrices $\mathbf{1}_A$ par des combinaisons finies d'indicatrices d'intervalles, puis de couper hors d'un compact de mesure suffisamment petite. Ensuite, pour passer des fonctions en escalier aux fonctions continues à support compact, on remplace les indicatrices d'intervalles $\mathbf{1}_{[a,b]}$ par des fonctions continues qui valent 1 sur $[a+\varepsilon, b-\varepsilon]$, 0 hors de $[a,b]$, et sont affines sur les deux petits intervalles de raccord. L'erreur $L^p$ tend vers 0 avec $\varepsilon$. □

# Chapitre 5

# Intégrales multiples

Nous voulons maintenant donner un sens rigoureux à l'intégration sur un produit cartésien. Dans le cas de $\mathbb{R}^2$, cela revient à donner un sens à la mesure d'une surface et à justifier les calculs par intégrales itérées. Le cadre abstrait est celui des espaces mesurés produits.

## 5.1 Intégrale sur un espace produit

### 5.1.1 Tribu produit

Soient $(\Omega_i, \mathcal{A}_i)$, $1 \leq i \leq n$, des espaces mesurables, et posons $\Omega = \Omega_1 \times \cdots \times \Omega_n$. Un point de $\Omega$ s'écrit $\omega = (\omega_1, \ldots, \omega_n)$, avec $\omega_i \in \Omega_i$.

**Définition 5.1.** La tribu produit sur $\Omega$ est la tribu engendrée par les pavés mesurables :

$$\mathcal{A}_1 \otimes \cdots \otimes \mathcal{A}_n = \sigma\left(\{A_1 \times \cdots \times A_n \colon A_i \in \mathcal{A}_i,\ 1 \leq i \leq n\}\right).$$

Pour $n = 2$, on écrit simplement $\mathcal{A} \otimes \mathcal{B}$.

La tribu produit est la plus petite tribu qui rende les projections canoniques mesurables. Plus précisément, si $\pi_i \colon \Omega_1 \times \cdots \times \Omega_n \to \Omega_i$ désigne la projection $\pi_i(\omega) = \omega_i$, alors chaque $\pi_i$ est $(\mathcal{A}_1 \otimes \cdots \otimes \mathcal{A}_n, \mathcal{A}_i)$-mesurable. Réciproquement, toute tribu sur $\Omega$ qui rend toutes les projections mesurables contient la tribu produit. En effet,

$$A_1 \times \cdots \times A_n = \bigcap_{i=1}^{n} \pi_i^{-1}(A_i).$$

**Proposition 5.2.** *Soit $f \colon (Z, C) \to (X \times Y, \mathcal{A} \otimes \mathcal{B})$, que l'on écrit $f(z) = (f_X(z), f_Y(z))$. Alors $f$ est mesurable si et seulement si $f_X$ est $(C, \mathcal{A})$-mesurable et $f_Y$ est $(C, \mathcal{B})$-mesurable.*

*Démonstration.* Si $f$ est mesurable, alors $f_X = \pi_X \circ f$ et $f_Y = \pi_Y \circ f$ sont mesurables par composition. Réciproquement, si $f_X$ et $f_Y$ sont mesurables, alors, pour tout pavé $A \times B$ avec $A \in \mathcal{A}$ et $B \in \mathcal{B}$,

$$f^{-1}(A \times B) = f_X^{-1}(A) \cap f_Y^{-1}(B) \in C.$$

Ainsi $f$ est mesurable puisque les pavés engendrent $\mathcal{A} \otimes \mathcal{B}$. □

*Remarque* 5.3. Si $X$ et $Y$ sont des espaces topologiques, alors

$$\mathcal{B}(X) \otimes \mathcal{B}(Y) \subset \mathcal{B}(X \times Y),$$

car les projections sont continues pour la topologie produit. Si $X$ et $Y$ sont à bases dénombrables d'ouverts, par exemple s'ils sont métriques séparables, alors

$$\mathcal{B}(X) \otimes \mathcal{B}(Y) = \mathcal{B}(X \times Y).$$

En particulier,

$$\mathcal{B}(\mathbb{R}^d) = \mathcal{B}(\mathbb{R})^{\otimes d}.$$

### 5.1.2 Sections d'un ensemble mesurable

Soient $(X, \mathcal{A})$ et $(Y, \mathcal{B})$ deux espaces mesurables. Pour $C \subset X \times Y$, on définit les sections

$$C_x = \{y \in Y : (x, y) \in C\}, \qquad C^y = \{x \in X : (x, y) \in C\}.$$

**Proposition 5.4.** *Si $C \in \mathcal{A} \otimes \mathcal{B}$, alors, pour tout $x \in X$ et tout $y \in Y$, on a $C_x \in \mathcal{B}$ et $C^y \in \mathcal{A}$.*

*Démonstration.* Fixons $x \in X$ et considérons

$$\mathcal{D}_x = \{C \in \mathcal{A} \otimes \mathcal{B} : C_x \in \mathcal{B}\}.$$

C'est une tribu. Elle contient les pavés mesurables, car $(A \times B)_x = B$ si $x \in A$ et $(A \times B)_x = \emptyset$ sinon. Donc $\mathcal{D}_x = \mathcal{A} \otimes \mathcal{B}$. Le raisonnement pour les sections $C^y$ est le même. □

**Corollaire 5.5.** *Soit $f : (X \times Y, \mathcal{A} \otimes \mathcal{B}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ une fonction mesurable. Alors, pour tout $x \in X$, la fonction $f_x : y \mapsto f(x, y)$ est $(\mathcal{B}, \mathcal{B}(\mathbb{R}))$-mesurable, et, pour tout $y \in Y$, la fonction $f^y : x \mapsto f(x, y)$ est $(\mathcal{A}, \mathcal{B}(\mathbb{R}))$-mesurable.*

*Démonstration.* Pour une indicatrice $f = \mathbf{1}_C$, cela découle de la proposition précédente. Par linéarité, le résultat vaut pour les fonctions étagées, puis pour les fonctions mesurables positives par approximation croissante. Le cas général s'obtient en séparant parties positive et négative. □

### 5.1.3 Mesure produit

Pour construire une mesure sur $X \times Y$, on veut imposer la règle naturelle

$$m(A \times B) = \mu(A)\nu(B).$$

L'existence et surtout l'unicité de cette mesure nécessitent une hypothèse de $\sigma$-finitude.

**Théorème 5.6** (Mesure produit)**.** *Soient $(X, \mathcal{A}, \mu)$ et $(Y, \mathcal{B}, \nu)$ deux espaces mesurés $\sigma$-finis. Il existe une unique mesure $m$ sur $(X \times Y, \mathcal{A} \otimes \mathcal{B})$ telle que*

$$m(A \times B) = \mu(A)\nu(B), \qquad A \in \mathcal{A}, \quad B \in \mathcal{B}.$$

*Cette mesure est $\sigma$-finie et se note $\mu \otimes \nu$. De plus, pour tout $C \in \mathcal{A} \otimes \mathcal{B}$,*

$$(\mu \otimes \nu)(C) = \int_X \nu(C_x)\, d\mu(x) = \int_Y \mu(C^y)\, d\nu(y).$$

*Démonstration.* L'unicité provient du théorème d'unicité des mesures $\sigma$-finies. En effet, si $(A_n)$ et $(B_n)$ sont des suites croissantes d'ensembles mesurables de mesure finie recouvrant respectivement $X$ et $Y$, alors les pavés $A_n \times B_n$ recouvrent $X \times Y$ et ont une mesure finie. La classe des pavés mesurables est stable par intersections finies et engendre $\mathcal{A} \otimes \mathcal{B}$.

Pour la formule avec les sections, on commence par montrer que, pour $C \in \mathcal{A} \otimes \mathcal{B}$, les fonctions $x \mapsto \nu(C_x)$ et $y \mapsto \mu(C^y)$ sont mesurables positives. On le vérifie d'abord pour les pavés, puis on utilise un argument de classe monotone. Les deux applications

$$C \longmapsto \int_X \nu(C_x)\, d\mu(x), \qquad C \longmapsto \int_Y \mu(C^y)\, d\nu(y)$$

définissent alors des mesures sur $\mathcal{A} \otimes \mathcal{B}$ qui coïncident avec $\mu(A)\nu(B)$ sur les pavés. Par unicité, elles sont égales à $\mu \otimes \nu$. □

*Remarque* 5.7. L'hypothèse de $\sigma$-finitude ne peut pas être simplement omise. Par exemple, prenons $X = Y = \mathbb{R}$, $\mu = \lambda$ la mesure de Lebesgue et $\nu$ la mesure de comptage sur $\mathbb{R}$, qui n'est pas $\sigma$-finie. Pour la diagonale $\Delta = \{(x, x) : x \in \mathbb{R}\}$, les sections vérifient $\Delta_x = \{x\}$ et $\Delta^y = \{y\}$. On obtient alors

$$\int_{\mathbb{R}} \nu(\Delta_x)\, d\lambda(x) = +\infty, \qquad \int_{\mathbb{R}} \lambda(\Delta^y)\, d\nu(y) = 0.$$

Les deux formules ne donnent donc pas le même résultat.

*Remarque* 5.8. Le théorème s'étend par récurrence au produit d'un nombre fini de mesures $\sigma$-finies. En particulier, la mesure de Lebesgue sur $\mathbb{R}^d$ est

$$\lambda_d = \underbrace{\lambda \otimes \cdots \otimes \lambda}_{d \text{ fois}}.$$

## 5.2 Théorèmes de Fubini

**Théorème 5.9** (Fubini–Tonelli)**.** *Soient $(X, \mathcal{A}, \mu)$ et $(Y, \mathcal{B}, \nu)$ deux espaces mesurés $\sigma$-finis, et soit $f : X \times Y \to \overline{\mathbb{R}}_+$ une fonction $\mathcal{A} \otimes \mathcal{B}$-mesurable positive. Alors les fonctions*

$$x \longmapsto \int_Y f(x, y)\, d\nu(y), \qquad y \longmapsto \int_X f(x, y)\, d\mu(x)$$

*sont mesurables positives, et*

$$\int_{X\times Y} f\, d(\mu \otimes \nu) = \int_X \left( \int_Y f(x, y)\, d\nu(y) \right) d\mu(x) = \int_Y \left( \int_X f(x, y)\, d\mu(x) \right) d\nu(y).$$

*Démonstration.* Pour $f = \mathbf{1}_C$, le résultat est exactement la formule de la mesure produit. Par linéarité, il vaut pour les fonctions étagées positives. Si $f$ est mesurable positive, on choisit une suite croissante $(f_n)$ de fonctions étagées positives convergeant vers $f$. On applique le résultat précédent à chaque $f_n$ et l'on passe à la limite par le théorème de Beppo–Levi. □

**Théorème 5.10** (Fubini–Lebesgue)**.** *Soient $(X, \mathcal{A}, \mu)$ et $(Y, \mathcal{B}, \nu)$ deux espaces mesurés $\sigma$-finis, et soit $f : X \times Y \to \mathbb{K}$, avec $\mathbb{K} = \mathbb{R}$ ou $\mathbb{C}$, une fonction $\mathcal{A} \otimes \mathcal{B}$-mesurable telle que $f \in L^1(\mu \otimes \nu)$. Alors :*

(i) *pour $\mu$-presque tout $x$, la fonction $y \mapsto f(x, y)$ appartient à $L^1(\nu)$ ; pour $\nu$-presque tout $y$, la fonction $x \mapsto f(x, y)$ appartient à $L^1(\mu)$ ;*

(ii) *les fonctions*

$$x \longmapsto \int_Y f(x, y)\, d\nu(y), \qquad y \longmapsto \int_X f(x, y)\, d\mu(x)$$

*sont intégrables, définies presque partout ;*

(iii) *on a*

$$\int_{X\times Y} f\, d(\mu \otimes \nu) = \int_X \left( \int_Y f(x, y)\, d\nu(y) \right) d\mu(x) = \int_Y \left( \int_X f(x, y)\, d\mu(x) \right) d\nu(y).$$

*Démonstration.* Dans le cas réel, on applique Fubini–Tonelli à $f^+$ et $f^-$. Comme

$$\int_{X\times Y} |f|\, d(\mu \otimes \nu) < +\infty,$$

les ensembles où l'une des intégrales de $f^+$ ou $f^-$ vaut $+\infty$ sont de mesure nulle. Les identités s'obtiennent par différence. Le cas complexe se traite en séparant partie réelle et partie imaginaire. □

**Exemple 5.11** (Séries doubles)**.** Prenons $X = Y = \mathbb{N}$ munis de la tribu des parties et de la mesure de comptage. Pour une fonction $f : \mathbb{N}^2 \to \mathbb{K}$, le théorème de Fubini donne, dès que

$$\sum_{n,m\geq 0} |f(n,m)| < +\infty,$$

l'égalité

$$\sum_{n,m\geq 0} f(n,m) = \sum_{n\geq 0}\sum_{m\geq 0} f(n,m) = \sum_{m\geq 0}\sum_{n\geq 0} f(n,m).$$

Pour une fonction positive, cette identité reste vraie avec des valeurs dans $[0, +\infty]$.

**Exemple 5.12** (Deux intégrales itérées différentes)**.** Sur $]0,1]^2$, posons

$$f(x,y) = \frac{x^2 - y^2}{(x^2+y^2)^2}.$$

Alors

$$\int_0^1 f(x,y)\,dy = \left[\frac{y}{x^2+y^2}\right]_0^1 = \frac{1}{1+x^2},$$

et donc

$$\int_0^1 \left(\int_0^1 f(x,y)\,dy\right) dx = \frac{\pi}{4}.$$

En intervertissant les rôles de $x$ et $y$, on obtient au contraire

$$\int_0^1 \left(\int_0^1 f(x,y)\,dx\right) dy = -\frac{\pi}{4}.$$

Ainsi $f$ n'est pas intégrable sur $]0,1]^2$ au sens de Lebesgue, et les hypothèses du théorème de Fubini–Lebesgue sont indispensables.

## 5.3 Intégrales multiples sur $\mathbb{R}^d$

On rappelle d'abord la formule de changement de variables. Soient $U$ et $V$ deux ouverts de $\mathbb{R}^d$ et $\varphi : U \to V$ un $C^1$-difféomorphisme. On note

$$D\varphi(x) = \left(\frac{\partial \varphi_i}{\partial x_j}(x)\right)_{1\leq i,j\leq d}, \qquad J_\varphi(x) = \det D\varphi(x).$$

**Théorème 5.13** (Changement de variables)**.** *Soient $U$ et $V$ deux ouverts de $\mathbb{R}^d$, et soit $\varphi : U \to V$ un $C^1$-difféomorphisme. Pour toute fonction borélienne $f : V \to \overline{\mathbb{R}}_+$, on a*

$$\int_V f(y)\,d\lambda_d(y) = \int_U f(\varphi(x))\,\left|J_\varphi(x)\right|\,d\lambda_d(x).$$

*La même formule vaut pour toute fonction $f$ intégrable sur $V$.*

**Exemple 5.14** (Coordonnées polaires)**.** L'application

$$\varphi(r,\theta) = (r\cos\theta, r\sin\theta)$$

est un $C^1$-difféomorphisme de $\mathbb{R}_+^*\times]0,2\pi[$ sur $\mathbb{R}^2 \setminus ([0,+\infty[\times\{0\})$. Son jacobien vaut $r$. Ainsi, pour une fonction borélienne positive $f$ sur $\mathbb{R}^2$,

$$\int_{\mathbb{R}^2} f(x,y)\,d\lambda_2(x,y) = \int_0^{2\pi}\int_0^{+\infty} f(r\cos\theta, r\sin\theta) r\,dr\,d\theta,$$

avec la convention usuelle que l'on peut négliger les ensembles de mesure nulle.

**Exemple 5.15** (Intégrale gaussienne)**.** Soit

$$g(x,y) = \frac{1}{2\pi} e^{-(x^2+y^2)/2}.$$

Par Fubini et le changement de variables polaires,

$$\int_{\mathbb{R}^2} g(x,y)\, d\lambda_2(x,y) = \frac{1}{2\pi} \int_0^{2\pi} \int_0^{+\infty} e^{-r^2/2} r\, dr\, d\theta = 1.$$

On retrouve donc

$$\int_{\mathbb{R}} e^{-x^2/2}\, dx = \sqrt{2\pi}.$$

**Exemple 5.16** (Volume de la boule unité)**.** On note

$$\omega_d = \lambda_d\left(\left\{x \in \mathbb{R}^d :\ \|x\| \le 1\right\}\right)$$

le volume de la boule euclidienne unité de $\mathbb{R}^d$. Pour $r \ge 0$,

$$\lambda_d\left(\left\{x \in \mathbb{R}^d :\ \|x\| \le r\right\}\right) = r^d \omega_d.$$

Considérons $f(x,t) = e^{-t}\mathbf{1}_{\{\|x\|^2 \le t\}}$ sur $\mathbb{R}^d \times \mathbb{R}_+$. Par Fubini,

$$\int_{\mathbb{R}^d} \int_{\|x\|^2}^{+\infty} e^{-t}\, dt\, d\lambda_d(x) = \int_0^{+\infty} e^{-t} \lambda_d\left(\left\{x :\ \|x\| \le \sqrt{t}\right\}\right) dt.$$

Le membre de gauche vaut

$$\int_{\mathbb{R}^d} e^{-\|x\|^2}\, d\lambda_d(x) = \pi^{d/2},$$

et le membre de droite vaut

$$\omega_d \int_0^{+\infty} t^{d/2} e^{-t}\, dt = \omega_d\, \Gamma\left(\frac{d}{2} + 1\right).$$

Ainsi

$$\omega_d = \frac{\pi^{d/2}}{\Gamma\left(\frac{d}{2} + 1\right)}.$$

En particulier, pour $d = 3$, on retrouve le volume $4\pi/3$ de la boule unité.

# Deuxième partie

# Probabilités

# Chapitre 6

# Introduction aux probabilités

En 1933, N. Kolmogorov, en s'appuyant sur la théorie de la mesure, a donné un cadre axiomatique précis au calcul des probabilités. L'idée générale est la suivante : on décrit d'abord l'ensemble $\Omega$ des issues possibles d'une expérience, puis on choisit une tribu $\mathcal{A}$ d'événements observables, et enfin on attribue à ces événements une mesure positive de masse totale 1, appelée probabilité.

Ce point de vue permet de traiter dans un même langage les expériences finies, comme un lancer de dé, les expériences dénombrables, comme une suite infinie de lancers, et les modèles continus, comme une durée de vie ou une trajectoire aléatoire. Il explique aussi pourquoi la théorie des probabilités repose naturellement sur les notions vues dans la première partie : tribus, mesures, fonctions mesurables et intégrales.

## 6.1 Expériences aléatoires et événements

Une expérience est aléatoire dès lors que son résultat ne peut pas être prédit exactement. Par exemple, on ne connaît pas à l'avance la valeur lue sur la face d'un dé que l'on jette, la température demain matin à huit heures, ou encore le prochain résultat d'un tirage du loto. Dans chacun de ces exemples, le résultat précis est incertain ; en revanche, on connaît, au moins idéalement, l'ensemble des issues possibles.

Il est important de souligner que le mot "aléatoire" ne signifie pas nécessairement que le phénomène n'obéit à aucune loi physique. Pour un lancer de pièce, si l'on connaissait avec une précision parfaite la vitesse initiale, la hauteur du lancer, l'angle de rotation, les frottements et les courants d'air, on pourrait en principe écrire des équations de mécanique. Mais le mécanisme est trop complexe et trop sensible aux conditions initiales pour que ce modèle déterministe soit exploitable. On préfère donc décrire l'expérience à l'aide d'un modèle probabiliste.

On note $\Omega$ l'ensemble des valeurs possibles de l'expérience. La théorie des probabilités modélise ensuite la fréquence de réalisation des événements. Si une expérience est répétée $n$ fois dans des conditions identiques et de façon indépendante, et si un événement $A$ s'est produit $k_n(A)$ fois, alors $k_n(A)/n$ est la fréquence empirique de $A$. L'idée fondamentale est que, dans les situations régulières,

$$\frac{k_n(A)}{n} \longrightarrow \mathbb{P}(A), \qquad n \to \infty.$$

Cette convergence sera formalisée plus tard par la loi des grands nombres. Elle justifie aussi l'additivité : si $A \cap B = \emptyset$, alors

$$\frac{k_n(A \cup B)}{n} = \frac{k_n(A)}{n} + \frac{k_n(B)}{n},$$

et donc, au passage à la limite, on doit avoir

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) \qquad \text{si } A \cap B = \emptyset.$$

### 6.1.1 Espace des états

**Définition 6.1.** Une *expérience aléatoire* est une expérience dont le résultat ne peut pas être déterminé avec certitude. On note souvent $E$ une telle expérience. Son *univers*, ou *espace des états*, est l'ensemble de tous les résultats possibles ; on le note $\Omega$. Un *événement élémentaire* est un résultat possible de l'expérience, c'est-à-dire un élément $\omega \in \Omega$.

L'univers peut être fini, dénombrable ou non dénombrable. Le choix de $\Omega$ n'est pas toujours unique : il dépend de l'information que l'on souhaite conserver sur l'expérience.

**Exemple 6.2** (Choix d'univers)**.** Voici quelques exemples d'expériences aléatoires et de choix naturels d'univers.

(1) Pour un lancer d'une pièce de monnaie, on peut prendre $\Omega = \{P, F\}$, où $P$ signifie pile et $F$ face.

(2) Pour deux lancers consécutifs d'une même pièce,

$$\Omega = \{PP, PF, FP, FF\}.$$

(3) Pour un lancer d'un dé régulier,

$$\Omega = \{1, 2, 3, 4, 5, 6\}.$$

(4) Pour deux lancers consécutifs d'un même dé, si l'on garde le détail des deux résultats, on prend $\Omega_1 = \{1, \dots, 6\}^2$. Si l'on ne s'intéresse qu'à la somme obtenue, on peut aussi prendre $\Omega_2 = \{2, 3, \dots, 12\}$.

(5) Pour un nombre infini de lancers d'un même dé,

$$\Omega = \{(u_n)_{n\geq 1} : u_n \in \{1, \dots, 6\} \text{ pour tout } n \geq 1\} = \{1, \dots, 6\}^{\mathbb{N}^*}.$$

(6) Pour la durée de vie d'un individu, on peut choisir par exemple $\Omega = [0, 120]$, mesuré en années.

(7) Pour la promenade d'un ivrogne dans une rue, qui à chaque instant fait un pas en avant ou en arrière, on peut prendre

$$\Omega = \{-1, 1\}^{\mathbb{N}^*}.$$

(8) Pour la trajectoire continue d'une poussière à la surface de l'eau pendant un intervalle de temps $[0, T]$, un modèle possible est

$$\Omega = C([0, T], \mathbb{R}^2).$$

Dans la suite, on va vouloir calculer la probabilité de certaines parties de l'espace fondamental $\Omega$. Ces parties seront les événements.

### 6.1.2 Événements

**Définition 6.3.** Un *événement* est une propriété dont on peut dire si elle est réalisée ou non une fois l'issue de l'expérience connue. Mathématiquement, un événement est donc un sous-ensemble de l'univers $\Omega$.

**Exemple 6.4** (Événements dans les exemples précédents)**.**

(1) Pour un lancer de pièce, l'événement "obtenir pile" est le sous-ensemble $\{P\}$ de $\{P, F\}$.

(2) Pour deux lancers de pièce, l'événement "le premier lancer donne pile" est

$$A = \{PP, PF\}.$$

(3) Pour deux lancers de dé, l'événement "la somme vaut 4" est

$$A_1 = \{(1,3),(2,2),(3,1)\} \subset \{1,\dots,6\}^2,$$

si l'on garde le détail des deux dés, et simplement $\{4\}$ si l'univers choisi est $\{2,3,\dots,12\}$.

(4) Pour des lancers indéfinis d'un dé, l'événement "le premier 1 est obtenu au $N$-ième lancer" est

$$\{(u_n)_{n\geq 1} : u_1 \geq 2, \dots, u_{N-1} \geq 2, u_N = 1\}.$$

(5) Pour une durée de vie, l'événement "l'individu atteint au moins 50 ans" est $\{x \in \mathbb{R}_+ : 50 \leq x \leq 120\}$.

(6) Pour la promenade d'un ivrogne, l'événement "l'ivrogne avance au $N$-ième pas" est $\{(u_n)_{n\geq 1} : u_N = 1\}$.

Puisque les événements sont des sous-ensembles de $\Omega$, on peut effectuer sur eux les opérations usuelles de la théorie des ensembles. Ces opérations ont une interprétation probabiliste naturelle.

| Notation | Terminologie ensembliste | Terminologie probabiliste |
|---|---|---|
| $\Omega$ | ensemble entier | espace des états, événement certain |
| $\omega$ | élément de $\Omega$ | événement élémentaire |
| $A$ | sous-ensemble de $\Omega$ | événement |
| $\omega \in A$ | $\omega$ appartient à $A$ | $A$ est réalisé si $\omega$ est l'issue de l'expérience |
| $A \subset B$ | $A$ est inclus dans $B$ | si $A$ est réalisé, alors $B$ aussi |
| $A \cup B$ | réunion de $A$ et $B$ | événement "$A$ ou $B$", le ou n'étant pas exclusif |
| $A \cap B$ | intersection de $A$ et $B$ | événement "$A$ et $B$" |
| $A^c$ | complémentaire de $A$ | événement contraire de $A$ |
| $\emptyset$ | ensemble vide | événement impossible |
| $A \cap B = \emptyset$ | $A$ et $B$ sont disjoints | $A$ et $B$ sont incompatibles |

**Exemple 6.5.** Pour deux lancers successifs d'une même pièce de monnaie, posons $A = \{PP\}$, $B = \{PF\}$ et $C = \{FP, FF\}$. Alors

$$A \cup B = \{PP, PF\} = C^c$$

est l'événement "le premier lancer donne pile". En revanche $A \cap B = \emptyset$ est l'événement impossible : les événements $A$ et $B$ sont incompatibles.

*Propriété* 6.6 (Règles de calcul ensembliste). Pour tous événements $A, B, C$, on a les identités suivantes :

$$A \cup B = B \cup A, \qquad (A \cup B) \cup C = A \cup (B \cup C),$$

$$(A \cup B) \cap C = (A \cap C) \cup (B \cap C),$$

et les lois de Morgan

$$(A \cup B)^c = A^c \cap B^c, \qquad (A \cap B)^c = A^c \cup B^c.$$

Les mêmes règles valent pour les intersections et réunions finies ou dénombrables.

*Remarque* 6.7. Si $A_1, \dots, A_n$ sont des événements, alors $\bigcup_{i=1}^n A_i$ est l'événement "au moins un des $A_i$ se réalise", tandis que $\bigcap_{i=1}^n A_i$ est l'événement "tous les $A_i$ se réalisent". De même, $\bigcup_{i\geq 1} A_i$ signifie qu'au moins un des événements de la suite se réalise, et $\bigcap_{i\geq 1} A_i$ signifie que tous se réalisent.

**Exemple 6.8** (Alice et Bob)**.** Alice et Bob lancent le même dé à tour de rôle, Alice commençant. Le gagnant est le premier à obtenir un six. On considère les événements

$$A = \{\text{victoire d'Alice}\}, \qquad B = \{\text{victoire de Bob}\}, \qquad D = \{\text{il n'y a pas de vainqueur}\}.$$

Pour analyser ces événements, introduisons

$$F_n = \{\text{fin de la partie au } n\text{-ième lancer}\}, \qquad S_j = \{\text{le } j\text{-ième lancer donne un six}\}.$$

Il est clair que l'absence de vainqueur signifie qu'aucun lancer ne donne six :

$$D = \bigcap_{j\geq 1} S_j^c.$$

Alice ne peut gagner qu'aux lancers de rang impair, et Bob qu'aux lancers de rang pair. Ainsi

$$A = \bigcup_{k\geq 0} F_{2k+1}, \qquad B = \bigcup_{k\geq 1} F_{2k}.$$

Enfin, la partie finit au $n$-ième lancer si les $n-1$ premiers lancers ne donnent pas six et si le $n$-ième donne six. Donc

$$F_1 = S_1, \qquad F_n = \left(\bigcap_{j=1}^{n-1} S_j^c\right) \cap S_n \quad (n \geq 2).$$

En particulier, avec la convention $\bigcap_{j=1}^{0} S_j^c = \Omega$,

$$A = \bigcup_{k\geq 0}\left[\left(\bigcap_{j=1}^{2k} S_j^c\right) \cap S_{2k+1}\right], \qquad B = \bigcup_{k\geq 1}\left[\left(\bigcap_{j=1}^{2k-1} S_j^c\right) \cap S_{2k}\right].$$

Cet exemple illustre l'intérêt de décomposer un événement compliqué à l'aide d'événements plus simples.

Un événement est donc une partie de $\Omega$, autrement dit un élément de $\mathcal{P}(\Omega)$. Si $\Omega = \{a, b, c\}$, alors $\mathcal{P}(\Omega)$ a 8 éléments :

$$\emptyset, \{a\}, \{b\}, \{c\}, \{a, b\}, \{a, c\}, \{b, c\}, \{a, b, c\} = \Omega.$$

Plus généralement, si $\Omega$ a $n$ éléments, alors

$$\mathrm{Card}(\mathcal{P}(\Omega)) = 2^n.$$

Il pourrait sembler naturel d'associer une probabilité à toute partie de $\Omega$, c'est-à-dire de prendre $\mathcal{P}(\Omega)$ comme ensemble des événements. C'est ce que l'on fait lorsque $\Omega$ est fini ou dénombrable. En revanche, pour certains univers non dénombrables, par exemple $\Omega = \mathbb{R}$, la classe $\mathcal{P}(\Omega)$ est trop grande pour que l'on puisse attribuer une probabilité à toutes ses parties de manière cohérente. On se restreint alors à une sous-famille $\mathcal{A} \subset \mathcal{P}(\Omega)$ dont les éléments seront les événements auxquels on peut attribuer une probabilité. Cette famille doit être stable par les opérations logiques usuelles : complémentaire, réunion dénombrable, intersection dénombrable. C'est la notion de tribu.

**Définition 6.9** (Tribu)**.** Soit $\Omega$ un univers. Une classe $\mathcal{A} \subset \mathcal{P}(\Omega)$ est une *tribu*, ou *$\sigma$-algèbre*, si :

(i) $\Omega \in \mathcal{A}$ ;

(ii) si $A \in \mathcal{A}$, alors $A^c \in \mathcal{A}$ ;

(iii) si $(A_n)_{n\geq 1}$ est une suite d'éléments de $\mathcal{A}$, alors $\bigcup_{n\geq 1} A_n \in \mathcal{A}$.

Le couple $(\Omega, \mathcal{A})$ est appelé *espace probabilisable*.

*Remarque* 6.10. Il faut bien distinguer les deux notations $A \subset \Omega$ et $A \in \mathcal{A}$. La première signifie que $A$ est une partie de l'univers ; la seconde signifie que cette partie est un événement mesurable, c'est-à-dire qu'elle appartient à la tribu. Écrire $A \subset \mathcal{A}$ n'a en général aucun sens dans ce contexte.

*Remarque* 6.11. La stabilité par intersections dénombrables découle de la définition. En effet, si $A_n \in \mathcal{A}$, alors

$$\bigcap_{n\geq 1} A_n = \left(\bigcup_{n\geq 1} A_n^c\right)^c \in \mathcal{A}.$$

En particulier, une tribu est stable par différence, car $A \setminus B = A \cap B^c$, et par différence symétrique.

**Exemple 6.12.** La classe $\{\emptyset, \Omega\}$ est la tribu triviale. La classe $\mathcal{P}(\Omega)$ est la tribu totale, parfois appelée tribu discrète. Sur $\mathbb{R}$, on utilise généralement la tribu borélienne

$$\mathcal{B}(\mathbb{R}) = \sigma\left(\{]a, b[ : a < b\}\right).$$

Si $A \subset \Omega$, la plus petite tribu contenant $A$ est $\{\emptyset, \Omega, A, A^c\}$. Si $\Omega = \{a, b, c, d\}$, on peut par exemple considérer les tribus $\{\emptyset, \Omega, \{a\}, \{b, c, d\}\}$ ou $\{\emptyset, \Omega, \{a, b\}, \{c, d\}\}$. Le choix de la tribu reflète l'information que l'on décide d'observer.

## 6.2 Probabilité

Une fois fixés un univers $\Omega$ et une tribu $\mathcal{A}$, on peut définir proprement ce qu'est une probabilité sur $(\Omega, \mathcal{A})$. Intuitivement, une probabilité mesure la chance de réalisation d'un événement. Formellement, c'est une mesure positive de masse totale 1.

**Définition 6.13.** Soit $(\Omega, \mathcal{A})$ un espace probabilisable. Une *probabilité* sur $(\Omega, \mathcal{A})$ est une application $\mathbb{P} : \mathcal{A} \to [0, 1]$ telle que :

(i) $\mathbb{P}(\Omega) = 1$ ;

(ii) si $(A_n)_{n\geq 1}$ est une suite d'événements deux à deux incompatibles, alors

$$\mathbb{P}\left(\bigsqcup_{n\geq 1} A_n\right) = \sum_{n\geq 1} \mathbb{P}(A_n).$$

Le triplet $(\Omega, \mathcal{A}, \mathbb{P})$ est appelé *espace probabilisé*.

*Remarque* 6.14. La propriété précédente est la $\sigma$-additivité. Elle implique en particulier $\mathbb{P}(\emptyset) = 0$. En effet, en prenant $A_n = \emptyset$ pour tout $n \geq 1$, on obtient

$$\mathbb{P}(\emptyset) = \sum_{n\geq 1} \mathbb{P}(\emptyset),$$

ce qui force $\mathbb{P}(\emptyset) = 0$, puisque $\mathbb{P}(\emptyset)$ est finie.

**Exemple 6.15** (Équiprobabilité sur un univers fini)**.** Soit $\Omega$ fini. Si les événements élémentaires sont équiprobables, on définit

$$\mathbb{P}(\{\omega\}) = \frac{1}{\mathrm{Card}(\Omega)} \qquad (\omega \in \Omega).$$

Alors, pour tout $A \subset \Omega$,

$$\mathbb{P}(A) = \frac{\mathrm{Card}(A)}{\mathrm{Card}(\Omega)}.$$

C'est le modèle usuel pour un dé équilibré, une pièce équilibrée ou un tirage uniforme parmi un nombre fini de possibilités.

**Proposition 6.16** (Propriétés élémentaires d'une probabilité)**.** *Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. Alors :*

(i) *si $A_1, \dots, A_n$ sont deux à deux disjoints, alors*

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

(ii) $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$ *;*

(iii) *si $A \subset B$, alors $\mathbb{P}(A) \leq \mathbb{P}(B)$ ;*

(iv) $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$ *;*

(v) *plus généralement, pour $A_1, \dots, A_n$,*

$$\mathbb{P}(A_1 \cup \cdots \cup A_n) = \sum_{k=1}^{n} (-1)^{k-1} \sum_{1 \leq i_1 < \cdots < i_k \leq n} \mathbb{P}(A_{i_1} \cap \cdots \cap A_{i_k});$$

(vi) *pour toute suite $(A_n)_{n \geq 1}$,*

$$\mathbb{P}\left(\bigcup_{n \geq 1} A_n\right) \leq \sum_{n \geq 1} \mathbb{P}(A_n);$$

(vii) *si $A_n \uparrow A$, alors $\mathbb{P}(A_n) \uparrow \mathbb{P}(A)$ ;*

(viii) *si $A_n \downarrow A$, alors $\mathbb{P}(A_n) \downarrow \mathbb{P}(A)$.*

*Démonstration.* Ces propriétés sont les propriétés générales des mesures positives, appliquées au cas où $\mathbb{P}(\Omega) = 1$. Par exemple, si $A \subset B$, alors $B = A \sqcup (B \setminus A)$, d'où $\mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(B \setminus A) \geq \mathbb{P}(A)$. La formule d'inclusion-exclusion pour deux événements vient de la décomposition disjointe

$$A \cup B = (A \setminus B) \sqcup (A \cap B) \sqcup (B \setminus A).$$

Les propriétés de continuité croissante et décroissante sont celles déjà obtenues pour les mesures. □

*Remarque* 6.17. Toute réunion dénombrable d'événements de probabilité nulle est encore de probabilité nulle. De même, toute intersection dénombrable d'événements de probabilité 1 est encore de probabilité 1. Cette observation justifie le vocabulaire "presque sûrement". On dit qu'une propriété est vraie $\mathbb{P}$-presque sûrement si elle est vraie sur un événement de probabilité 1.

**Lemme 6.18** (Borel–Cantelli, première partie)**.** *Soit $(A_n)_{n \geq 1}$ une suite d'événements. Si $\sum_{n \geq 1} \mathbb{P}(A_n) < +\infty$, alors*

$$\mathbb{P}\left(\limsup_{n \to \infty} A_n\right) = 0.$$

*Autrement dit, presque sûrement, seul un nombre fini des événements $A_n$ se réalisent.*

*Démonstration.* Pour tout $N \geq 1$, on a

$$\limsup_{n \to \infty} A_n = \bigcap_{N \geq 1} \bigcup_{n \geq N} A_n \subset \bigcup_{n \geq N} A_n.$$

Par sous-additivité,

$$\mathbb{P}\left(\bigcup_{n \geq N} A_n\right) \leq \sum_{n \geq N} \mathbb{P}(A_n).$$

Le membre de droite tend vers 0, puisque la série converge. En faisant tendre $N$ vers l'infini, on obtient le résultat. □

## 6.3 Probabilité conditionnelle et indépendance

La connaissance d'une information sur une expérience peut modifier la probabilité que l'on attribue à un événement. Par exemple, dans un jeu de cartes, la probabilité que la deuxième carte soit un as n'est pas la même selon que l'on sait que la première carte est un as ou non. Cette idée conduit à la notion de probabilité conditionnelle.

**Définition 6.19.** Soient $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé et $A, B \in \mathcal{A}$ avec $\mathbb{P}(B) > 0$. La *probabilité conditionnelle* de $A$ sachant $B$ est

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

Pour $B$ fixé, l'application $A \mapsto \mathbb{P}(A \mid B)$ est une probabilité sur $(\Omega, \mathcal{A})$.

*Démonstration.* On a $\mathbb{P}(\Omega \mid B) = 1$. Si $(A_n)$ est une suite d'événements deux à deux disjoints, alors les $A_n \cap B$ sont encore deux à deux disjoints, et

$$\mathbb{P}\left(\bigcup_n A_n \mid B\right) = \frac{\mathbb{P}\left(\bigcup_n (A_n \cap B)\right)}{\mathbb{P}(B)} = \sum_n \frac{\mathbb{P}(A_n \cap B)}{\mathbb{P}(B)}.$$

□

**Proposition 6.20** (Conditionnements successifs)**.** *Si $A_1, \ldots, A_n$ sont des événements tels que $\mathbb{P}(A_1 \cap \cdots \cap A_{n-1}) > 0$, alors*

$$\mathbb{P}(A_1 \cap \cdots \cap A_n) = \mathbb{P}(A_1)\mathbb{P}(A_2 \mid A_1) \cdots \mathbb{P}(A_n \mid A_1 \cap \cdots \cap A_{n-1}).$$

**Définition 6.21** (Partition)**.** Une famille dénombrable $(A_i)_{i \in I}$ d'événements forme une partition de $\Omega$ si $A_i \neq \emptyset$ pour tout $i$, si les $A_i$ sont deux à deux disjoints, et si $\bigcup_{i \in I} A_i = \Omega$.

**Proposition 6.22** (Formule des probabilités totales)**.** *Si $(A_i)_{i \in I}$ est une partition dénombrable de $\Omega$, alors, pour tout événement $B$,*

$$\mathbb{P}(B) = \sum_{i \in I} \mathbb{P}(B \cap A_i).$$

*Si de plus $\mathbb{P}(A_i) > 0$ pour tout $i$, alors*

$$\mathbb{P}(B) = \sum_{i \in I} \mathbb{P}(B \mid A_i)\mathbb{P}(A_i).$$

**Exemple 6.23** (Deux cartes dans un jeu de 52)**.** On tire deux cartes dans un jeu de 52 cartes. On veut connaître la probabilité que la deuxième carte soit un as. Soit $A_1$ l'événement "la première carte est un as" et $A_2$ l'événement "la deuxième carte est un as". La famille $A_1, A_1^c$ est une partition, donc

$$\mathbb{P}(A_2) = \mathbb{P}(A_2 \mid A_1)\mathbb{P}(A_1) + \mathbb{P}(A_2 \mid A_1^c)\mathbb{P}(A_1^c).$$

Or

$$\mathbb{P}(A_2 \mid A_1) = \frac{3}{51}, \qquad \mathbb{P}(A_1) = \frac{4}{52},$$

et

$$\mathbb{P}(A_2 \mid A_1^c) = \frac{4}{51}, \qquad \mathbb{P}(A_1^c) = \frac{48}{52}.$$

Ainsi

$$\mathbb{P}(A_2) = \frac{3}{51}\frac{4}{52} + \frac{4}{51}\frac{48}{52} = \frac{4}{52} = \frac{1}{13}.$$

**Proposition 6.24** (Formule de Bayes)**.** *Soit $(A_i)_{i\in I}$ une partition dénombrable de $\Omega$ telle que $\mathbb{P}(A_i) > 0$ pour tout $i$. Si $B \in \mathcal{A}$ et $\mathbb{P}(B) > 0$, alors*

$$\mathbb{P}(A_n \mid B) = \frac{\mathbb{P}(B \mid A_n)\mathbb{P}(A_n)}{\sum_{i\in I} \mathbb{P}(B \mid A_i)\mathbb{P}(A_i)}.$$

**Exemple 6.25** (Test médical)**.** On considère une maladie $M$ et un test $T$, où $T$ désigne l'événement "le test est positif". Supposons que

$$\mathbb{P}(M) = 0.01, \qquad \mathbb{P}(T \mid M) = 0.95, \qquad \mathbb{P}(T \mid M^c) = 0.01.$$

La probabilité d'être malade sachant que le test est positif est

$$\mathbb{P}(M \mid T) = \frac{\mathbb{P}(T \mid M)\mathbb{P}(M)}{\mathbb{P}(T \mid M)\mathbb{P}(M) + \mathbb{P}(T \mid M^c)\mathbb{P}(M^c)}.$$

Numériquement,

$$\mathbb{P}(M \mid T) = \frac{0.95 \cdot 0.01}{0.95 \cdot 0.01 + 0.01 \cdot 0.99} \simeq 0.49.$$

Même avec un test très sensible, la faible fréquence de la maladie dans la population a donc un effet important sur la probabilité a posteriori.

L'idée intuitive de l'indépendance est la suivante : deux événements $A$ et $B$ sont indépendants si le fait de savoir que $B$ se produit ne modifie pas la probabilité de $A$. Lorsque $\mathbb{P}(B) > 0$, cela s'écrit $\mathbb{P}(A \mid B) = \mathbb{P}(A)$, ce qui équivaut à $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$. Cette dernière formule a l'avantage de rester valable même lorsque $\mathbb{P}(B) = 0$.

**Définition 6.26.** Deux événements $A$ et $B$ sont *indépendants* si

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B).$$

**Exemple 6.27.** On lance un dé deux fois de suite. Soit $A$ l'événement "le premier lancer donne un nombre pair" et $B$ l'événement "le second lancer donne un nombre pair". L'univers naturel est $\{1, \dots, 6\}^2$, muni de l'équiprobabilité. On a

$$\mathbb{P}(A) = \mathbb{P}(B) = \frac{1}{2}, \qquad \mathbb{P}(A \cap B) = \frac{9}{36} = \frac{1}{4}.$$

Donc $A$ et $B$ sont indépendants.

**Proposition 6.28** (Passage au complémentaire)**.** *Si $A$ et $B$ sont indépendants, alors les paires $A, B^c$, $A^c, B$ et $A^c, B^c$ sont aussi indépendantes.*

*Démonstration.* Par exemple,

$$\mathbb{P}(A \cap B^c) = \mathbb{P}(A) - \mathbb{P}(A \cap B) = \mathbb{P}(A) - \mathbb{P}(A)\mathbb{P}(B) = \mathbb{P}(A)\mathbb{P}(B^c).$$

Les autres cas se démontrent de la même façon. □

*Remarque* 6.29. Il ne faut pas confondre indépendance et incompatibilité. Si $A \cap B = \emptyset$ et si $\mathbb{P}(A) > 0$, $\mathbb{P}(B) > 0$, alors $\mathbb{P}(A \cap B) = 0$ tandis que $\mathbb{P}(A)\mathbb{P}(B) > 0$. Les événements ne sont donc pas indépendants.

**Définition 6.30.** Soient $A_1, \dots, A_n$ des événements.

(i) Ils sont *deux à deux indépendants* si $A_i$ et $A_j$ sont indépendants pour tous $i \neq j$.

(ii) Ils sont *mutuellement indépendants* si, pour tout $J \subset \{1, \dots, n\}$,

$$\mathbb{P}\left(\bigcap_{j\in J} A_j\right) = \prod_{j\in J} \mathbb{P}(A_j).$$

(iii) Une famille finie ou dénombrable $(A_i)_{i\in I}$ est mutuellement indépendante si toute sous-famille finie vérifie la propriété précédente.

**Exemple 6.31** (Indépendance deux à deux et indépendance mutuelle)**.** On lance deux fois une pièce équilibrée et l'on considère

$$A_1 = \{PP, PF\}, \qquad A_2 = \{PP, FP\}, \qquad A_3 = \{PP, FF\}.$$

Chaque événement a probabilité $1/2$, et chaque intersection deux à deux a probabilité $1/4$. Les événements sont donc deux à deux indépendants. En revanche,

$$\mathbb{P}(A_1 \cap A_2 \cap A_3) = \frac{1}{4} \neq \frac{1}{8} = \mathbb{P}(A_1)\mathbb{P}(A_2)\mathbb{P}(A_3).$$

Il n'y a donc pas indépendance mutuelle.

**Proposition 6.32.** *Soit $(A_i)_{i\in I}$ une famille finie ou dénombrable d'événements indépendants. Pour chaque $i$, soit $B_i$ égal à $A_i$, $A_i^c$, $\emptyset$ ou $\Omega$. Alors la famille $(B_i)_{i\in I}$ est encore indépendante.*

**Lemme 6.33** (Borel–Cantelli, deuxième partie)**.** *Soit $(A_n)_{n\geq 1}$ une suite d'événements indépendants. Si $\sum_{n\geq 1} \mathbb{P}(A_n) = +\infty$, alors*

$$\mathbb{P}\left(\limsup_{n\to\infty} A_n\right) = 1.$$

*Ainsi, sous indépendance, si la somme des probabilités diverge, alors presque sûrement une infinité des événements $A_n$ se réalisent.*

*Démonstration.* On montre que le complémentaire de la limite supérieure a probabilité nulle. Pour $N \leq M$, par indépendance,

$$\mathbb{P}\left(\bigcap_{n=N}^{M} A_n^c\right) = \prod_{n=N}^{M} (1 - \mathbb{P}(A_n)) \leq \exp\left(-\sum_{n=N}^{M} \mathbb{P}(A_n)\right).$$

Comme la série $\sum_n \mathbb{P}(A_n)$ diverge, le membre de droite tend vers 0 lorsque $M \to \infty$. Donc $\mathbb{P}(\bigcap_{n\geq N} A_n^c) = 0$ pour tout $N$, et finalement

$$\mathbb{P}\left(\bigcup_{N\geq 1} \bigcap_{n\geq N} A_n^c\right) = 0.$$

Or cet événement est le complémentaire de $\limsup_n A_n$. □

**Définition 6.34.** Des expériences aléatoires $E_1, \ldots, E_n$ sont indépendantes si, pour tout choix d'événements $A_i$ observables à l'issue de $E_i$, les événements $A_1, \ldots, A_n$ sont mutuellement indépendants. Une suite d'expériences est indépendante si toute sous-famille finie l'est.

**Exemple 6.35** (Lancers infinis d'une pièce)**.** On jette indéfiniment, de façon indépendante et dans les mêmes conditions, une pièce de monnaie. On note $p \in ]0, 1[$ la probabilité d'obtenir face à un lancer.

Pour $n \geq 1$, soit $A_n$ l'événement "on obtient face au lancer numéro $n$". Les événements $A_n$ sont indépendants et $\mathbb{P}(A_n) = p$. Comme $\sum_n \mathbb{P}(A_n) = +\infty$, la deuxième partie du lemme de Borel–Cantelli donne

$$\mathbb{P}(A_n \text{ se produit une infinité de fois}) = 1.$$

Autrement dit, presque sûrement, face apparaît une infinité de fois.

Soit maintenant $B_n$ l'événement "on obtient face à tous les lancers numéro $n, n+1, \ldots, 2n-1$". Alors $B_n$ impose $n$ succès consécutifs, donc

$$\mathbb{P}(B_n) = p^n.$$

La série $\sum_n p^n$ converge. La première partie du lemme de Borel–Cantelli donne donc

$$\mathbb{P}(B_n \text{ se produit une infinité de fois}) = 0.$$

Ainsi, presque sûrement, il n'y a qu'un nombre fini de blocs de la forme $n, n+1, \ldots, 2n-1$ entièrement constitués de faces.

# Chapitre 7

# Variables aléatoires réelles

Après avoir réalisé une expérience aléatoire, on ne s'intéresse souvent pas au résultat complet $\omega \in \Omega$, mais seulement à une quantité numérique qui dépend de ce résultat : le nombre de succès, le plus grand résultat obtenu, une durée de vie, un temps d'attente, etc. Cette idée conduit à la notion de variable aléatoire. Formellement, une variable aléatoire est une application mesurable définie sur l'espace probabilisé.

## 7.1 Variables aléatoires réelles

**Définition 7.1.** Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. Une variable aléatoire réelle est une application mesurable

$$X : (\Omega, \mathcal{A}) \longrightarrow (E, \mathcal{B}),$$

où $E \subset \mathbb{R}$ est l'ensemble des valeurs possibles de $X$, muni d'une tribu $\mathcal{B}$, le plus souvent la tribu borélienne induite par $\mathcal{B}(\mathbb{R})$. La condition de mesurabilité signifie que, pour tout $B \in \mathcal{B}$,

$$X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\} \in \mathcal{A}.$$

On note souvent cet événement $\{X \in B\}$. Par exemple, $\{X \in [a, b]\} = X^{-1}([a, b])$.

La mesurabilité est ce qui permet d'attribuer une probabilité aux événements formulés en termes de $X$. Sans elle, l'expression $\mathbb{P}(X \in B)$ n'aurait pas toujours de sens.

**Exemple 7.2.** On jette deux dés distincts. L'univers naturel est

$$\Omega = \{(i, j) : 1 \leq i, j \leq 6\},$$

muni de l'équiprobabilité.

(1) Si l'on s'intéresse à la somme des points, on définit

$$X(i, j) = i + j.$$

Alors $X$ prend ses valeurs dans $\{2, 3, \ldots, 12\}$.

(2) Si l'on s'intéresse au plus grand des deux chiffres obtenus, on définit

$$Y(i, j) = \max(i, j).$$

Alors $Y$ prend ses valeurs dans $\{1, 2, \ldots, 6\}$.

**Exemple 7.3.** On observe deux micro-processeurs et l'on s'intéresse à la durée de vie du premier qui tombe en panne. Si $\omega_1$ et $\omega_2$ désignent les durées de vie des deux composants, on peut prendre $\Omega = \mathbb{R}_+^* \times \mathbb{R}_+^*$ et définir

$$T(\omega_1, \omega_2) = \min(\omega_1, \omega_2).$$

La variable $T$ est à valeurs dans $\mathbb{R}_+^*$ et représente le premier temps de panne.

## 7.2 Loi d'une variable aléatoire réelle

**Définition 7.4.** Soit $X : (\Omega, \mathcal{A}, \mathbb{P}) \to (E, \mathcal{B})$ une variable aléatoire. La loi de $X$, notée $\mathbb{P}_X$ ou $P_X$, est la probabilité sur $(E, \mathcal{B})$ définie par

$$\mathbb{P}_X(B) = \mathbb{P}(X^{-1}(B)) = \mathbb{P}(X \in B), \qquad B \in \mathcal{B}.$$

On dit que $\mathbb{P}_X$ est la loi image de $\mathbb{P}$ par $X$.

Cette définition est exactement la notion de mesure image introduite dans la partie d'intégration :

$$\mathbb{P}_X = X_* \mathbb{P}.$$

Elle permet de transporter la probabilité de l'espace abstrait $\Omega$ vers l'espace concret des valeurs prises par $X$.

**Exemple 7.5.** Si $\Omega$ est une population de cinq individus, dont trois ont les yeux noirs et deux les yeux bleus, et si $X$ désigne la couleur des yeux, à valeurs dans $E = \{\text{noir}, \text{bleu}\}$, alors

$$\mathbb{P}_X(\{\text{noir}\}) = \mathbb{P}(X = \text{noir}) = \frac{3}{5}, \qquad \mathbb{P}_X(\{\text{bleu}\}) = \frac{2}{5}.$$

*Remarque* 7.6. Toute probabilité $Q$ sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ est la loi d'une variable aléatoire réelle. Il suffit de prendre $(\Omega, \mathcal{A}, \mathbb{P}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}), Q)$ et $X(\omega) = \omega$. Alors $\mathbb{P}_X = Q$.

*Remarque* 7.7. La loi de $X$ ne détermine pas la variable aléatoire $X$ elle-même. Elle ne retient que la répartition probabiliste des valeurs prises par $X$, et non la manière dont ces valeurs sont attachées aux résultats $\omega \in \Omega$.

### 7.2.1 Variables aléatoires discrètes

**Définition 7.8.** Une variable aléatoire $X$, à valeurs dans un ensemble $E \subset \mathbb{R}$, est dite discrète si elle ne prend qu'un nombre fini ou dénombrable de valeurs. On écrit alors

$$X(\Omega) = \{x_k : k \in K\}, \qquad K \subset \mathbb{N}.$$

Pour chaque valeur $x_k$, l'ensemble

$$A_k = \{\omega \in \Omega : X(\omega) = x_k\} = X^{-1}(\{x_k\})$$

est un événement. La loi de $X$ est entièrement déterminée par les nombres

$$p_k = \mathbb{P}(X = x_k), \qquad k \in K,$$

qui vérifient $p_k \geq 0$ et $\sum_{k \in K} p_k = 1$. On a alors

$$\mathbb{P}_X = \sum_{k \in K} p_k \delta_{x_k}.$$

**Exemple 7.9** (Somme de deux dés)**.** Pour la somme $X$ de deux dés distincts, on obtient

| $k$ | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $\mathbb{P}(X = k)$ | $\frac{1}{36}$ | $\frac{2}{36}$ | $\frac{3}{36}$ | $\frac{4}{36}$ | $\frac{5}{36}$ | $\frac{6}{36}$ | $\frac{5}{36}$ | $\frac{4}{36}$ | $\frac{3}{36}$ | $\frac{2}{36}$ | $\frac{1}{36}$ |

La loi est triangulaire : les sommes centrales sont plus probables que les sommes extrêmes.

**Exemple 7.10** (Maximum de deux dés)**.** Pour le maximum $Y$ de deux dés distincts, on obtient

| $k$ | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| $\mathbb{P}(Y = k)$ | $\frac{1}{36}$ | $\frac{3}{36}$ | $\frac{5}{36}$ | $\frac{7}{36}$ | $\frac{9}{36}$ | $\frac{11}{36}$ |

En effet, $\mathbb{P}(Y \leq k) = k^2/36$, donc $\mathbb{P}(Y = k) = k^2/36 - (k-1)^2/36 = (2k-1)/36$.

*Remarque* 7.11 (Même loi, variables différentes). On lance deux dés, l'un bleu et l'autre rouge. Soit $X$ le résultat du dé bleu et $Y$ celui du dé rouge. Alors $X$ et $Y$ ont la même loi uniforme sur $\{1, \ldots, 6\}$, mais ce ne sont pas les mêmes variables. En effet, l'événement $\{X = Y\}$ est l'événement "obtenir un double", et

$$\mathbb{P}(X = Y) = \frac{6}{36} = \frac{1}{6}, \qquad \mathbb{P}(X \neq Y) = \frac{5}{6}.$$

Ainsi $X$ et $Y$ ont même loi, mais ne sont pas égales presque sûrement.

**Exemple 7.12** (Lois discrètes usuelles)**.**

(1) *Loi uniforme finie.* Si $E = \{x_1, \ldots, x_n\}$ et $\mathbb{P}(X = x_i) = 1/n$, on écrit

$$X \sim \mathcal{U}(\{x_1, \ldots, x_n\}).$$

(2) *Loi de Bernoulli.* Si $A$ est un événement de probabilité $p$, alors $X = \mathbf{1}_A$ suit une loi de Bernoulli de paramètre $p$, notée $\mathcal{B}(p)$, et

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

(3) *Loi binomiale.* On répète $n$ fois, de façon indépendante, une expérience qui donne un succès avec probabilité $p$ et un échec avec probabilité $1 - p$. Si $N$ est le nombre de succès, alors

$$\mathbb{P}(N = k) = \binom{n}{k} p^k (1-p)^{n-k}, \qquad k = 0, \ldots, n.$$

On écrit $N \sim \mathcal{B}(n, p)$.

(4) *Loi géométrique.* On répète des expériences indépendantes, avec succès de probabilité $p > 0$, jusqu'au premier succès. Si $N$ est le rang du premier succès, alors

$$\mathbb{P}(N = k) = (1-p)^{k-1} p, \qquad k \geq 1.$$

On écrit $N \sim \mathcal{G}(p)$.

(5) *Loi de Poisson.* Pour $\lambda > 0$, on dit que $X \sim \mathcal{P}(\lambda)$ si $X$ est à valeurs dans $\mathbb{N}$ et

$$\mathbb{P}(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}, \qquad k \in \mathbb{N}.$$

Cette loi apparaît comme limite de lois binomiales : si $X_n \sim \mathcal{B}(n, p_n)$, avec $p_n \to 0$ et $np_n \to \lambda$, alors, pour tout $k \in \mathbb{N}$,

$$\mathbb{P}(X_n = k) \longrightarrow e^{-\lambda} \frac{\lambda^k}{k!}.$$

### 7.2.2 Variables aléatoires continues

**Définition 7.13.** Soit $X$ une variable aléatoire réelle. On dit que $X$ admet une densité par rapport à la mesure de Lebesgue, ou que sa loi $\mathbb{P}_X$ est absolument continue par rapport à $\lambda$, s'il existe une fonction mesurable positive $f_X : \mathbb{R} \to \mathbb{R}_+$, d'intégrale égale à 1, telle que, pour tout $A \in \mathcal{B}(\mathbb{R})$,

$$\mathbb{P}_X(A) = \mathbb{P}(X \in A) = \int_A f_X(x)\, d\lambda(x).$$

On écrit alors $\mathbb{P}_X \ll \lambda$.

*Remarque* 7.14. Si $X$ admet une densité, alors, pour tout $a \in \mathbb{R}$,

$$\mathbb{P}(X = a) = \int_{\{a\}} f_X(x)\, d\lambda(x) = 0.$$

Par conséquent,

$$\mathbb{P}(a < X < b) = \mathbb{P}(a \leq X \leq b) = \mathbb{P}(a < X \leq b) = \mathbb{P}(a \leq X < b).$$

Les extrémités d'un intervalle n'ont aucune influence pour une loi à densité.

*Remarque* 7.15. Une densité n'est pas définie de manière unique point par point. Si $f$ et $g$ sont deux densités de la même loi, alors $f = g$ $\lambda$-presque partout. On note la classe d'équivalence des densités

$$\frac{d\mathbb{P}_X}{d\lambda}.$$

**Exemple 7.16** (Lois à densité usuelles)**.**

(1) *Loi uniforme sur* $[a, b]$. Sa densité est

$$f_X(x) = \frac{1}{b-a}\mathbf{1}_{[a,b]}(x).$$

(2) *Loi uniforme sur un borélien.* Si $B \in \mathcal{B}(\mathbb{R})$ vérifie $0 < \lambda(B) < +\infty$, la loi uniforme sur $B$ a pour densité

$$f_X(x) = \frac{1}{\lambda(B)}\mathbf{1}_B(x).$$

(3) *Loi exponentielle.* Pour $\lambda > 0$, $X \sim \mathcal{E}(\lambda)$ si

$$f_X(x) = \lambda e^{-\lambda x}\mathbf{1}_{\{x \geq 0\}}.$$

Elle sert notamment à modéliser des durées de vie ou des temps d'attente.

(4) *Loi normale.* Pour $m \in \mathbb{R}$ et $\sigma^2 > 0$, $X \sim \mathcal{N}(m, \sigma^2)$ si

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right).$$

(5) *Loi du chi-deux.* Pour $n \geq 1$, $X \sim \chi^2(n)$ si

$$f_X(x) = \frac{(1/2)^{n/2}}{\Gamma(n/2)} x^{n/2-1} e^{-x/2}\mathbf{1}_{\{x>0\}},$$

où, pour $a > 0$,

$$\Gamma(a) = \int_0^{+\infty} x^{a-1}e^{-x}\, dx.$$

### 7.2.3 Fonction de répartition

**Définition 7.17.** La fonction de répartition d'une variable aléatoire réelle $X$ est la fonction $F_X : \mathbb{R} \to [0, 1]$ définie par

$$F_X(x) = \mathbb{P}(X \leq x) = \mathbb{P}_X(]-\infty, x]).$$

*Propriété* 7.18 (Propriétés d'une fonction de répartition). La fonction $F_X$ vérifie les propriétés suivantes.

(a) Elle caractérise la loi de $X$.

(b) Elle est croissante, continue à droite, et admet une limite à gauche en tout point.

(c) $\lim_{x\to-\infty} F_X(x) = 0$ et $\lim_{x\to+\infty} F_X(x) = 1$.

(d) Pour tout $x \in \mathbb{R}$,

$$\mathbb{P}(X = x) = F_X(x) - F_X(x^-), \qquad F_X(x^-) = \lim_{t\uparrow x} F_X(t).$$

(e) L'ensemble des points de discontinuité de $F_X$ est au plus dénombrable.

*Démonstration.* Le point (a) vient du théorème d'unicité des mesures : la classe $\{]-\infty, x] : x \in \mathbb{R}\} \cup \{\mathbb{R}\}$ est stable par intersections finies et engendre $\mathcal{B}(\mathbb{R})$. Les points (b) et (c) découlent de la croissance de la famille d'événements $\{X \le x\}$ et des propriétés de continuité des mesures. Pour (d), on écrit

$$\{X = x\} = \bigcap_{n\ge 1} \left\{ x - \frac{1}{n} < X \le x \right\}$$

et l'on utilise la continuité à droite des mesures sur la suite décroissante correspondante. Enfin, si $D$ est l'ensemble des discontinuités et si

$$D_n = \left\{ x \in \mathbb{R} : F_X(x) - F_X(x^-) > \frac{1}{n} \right\},$$

alors $D \subset \bigcup_n D_n$. Comme $F_X$ est croissante et à valeurs dans $[0, 1]$, l'ensemble $D_n$ contient au plus $n$ points. Ainsi $D$ est dénombrable. □

Si $X$ est discrète et prend ses valeurs dans $x_1 < x_2 < \cdots$, alors $F_X$ est une fonction en escaliers. Plus précisément, pour $x_k \le x < x_{k+1}$,

$$F_X(x) = \sum_{i\le k} \mathbb{P}(X = x_i).$$

**Exemple 7.19.** Pour le maximum $Y$ de deux dés, on a

| $k$ | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| $\mathbb{P}(Y = k)$ | $\frac{1}{36}$ | $\frac{3}{36}$ | $\frac{5}{36}$ | $\frac{7}{36}$ | $\frac{9}{36}$ | $\frac{11}{36}$ |
| $F_Y(k)$ | $\frac{1}{36}$ | $\frac{4}{36}$ | $\frac{9}{36}$ | $\frac{16}{36}$ | $\frac{25}{36}$ | 1 |

Si $X$ admet une densité $f_X$, alors

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\, dt.$$

La fonction $F_X$ est continue, et en tout point où $f_X$ est continue, $F_X$ est dérivable et $F_X'(x) = f_X(x)$. Par exemple, si $X \sim \mathcal{E}(\lambda)$, alors

$$F_X(x) = (1 - e^{-\lambda x})\mathbf{1}_{\{x\ge 0\}}.$$

*Remarque* 7.20. On dit qu'une probabilité est diffuse si elle ne charge aucun point. Une loi à densité est diffuse, mais la réciproque est fausse. Il existe des lois diffuses singulières, c'est-à-dire portées par un ensemble de mesure de Lebesgue nulle. De plus, une fonction de répartition peut être presque partout dérivable sans que la loi soit absolument continue par rapport à la mesure de Lebesgue : une loi de Bernoulli donne déjà une fonction de répartition dérivable presque partout avec dérivée nulle presque partout.

La fonction de répartition est l'outil privilégié pour calculer la loi d'une variable de la forme $Y = h(X)$.

**Exemple 7.21.** Soit $X \sim \mathcal{N}(0,1)$ et soit $Y = X^2$. Pour $y < 0$, $F_Y(y) = 0$. Pour $y \geq 0$,

$$F_Y(y) = \mathbb{P}(-\sqrt{y} \leq X \leq \sqrt{y}) = F_X(\sqrt{y}) - F_X(-\sqrt{y}) = 2F_X(\sqrt{y}) - 1.$$

En dérivant, on obtient, pour $y > 0$,

$$f_Y(y) = \frac{1}{\sqrt{y}} \frac{1}{\sqrt{2\pi}} e^{-y/2}.$$

Ainsi $Y \sim \chi^2(1)$.

**Proposition 7.22** (Changement de variable)**.** *Soit $X$ une variable aléatoire à valeurs dans un intervalle $I$, admettant une densité $f_X$. Soit $Y = h(X)$, où $h : I \to J$ est dérivable et bijective. Alors $Y$ admet une densité donnée, aux points $y \in J$ tels que $h'(h^{-1}(y)) \neq 0$, par*

$$f_Y(y) = \frac{1}{\left|h'(h^{-1}(y))\right|} f_X(h^{-1}(y)).$$

*Démonstration.* Si $h$ est croissante, alors

$$F_Y(y) = \mathbb{P}(h(X) \leq y) = \mathbb{P}(X \leq h^{-1}(y)) = F_X(h^{-1}(y)).$$

En dérivant, on obtient

$$f_Y(y) = f_X(h^{-1}(y))(h^{-1})'(y) = \frac{1}{h'(h^{-1}(y))} f_X(h^{-1}(y)).$$

Si $h$ est décroissante, alors

$$F_Y(y) = \mathbb{P}(X \geq h^{-1}(y)) = 1 - F_X(h^{-1}(y)),$$

et la dérivation donne la même formule avec une valeur absolue. □

**Exemple 7.23.** Si $X \sim \mathcal{N}(0,1)$, $m \in \mathbb{R}$, $\sigma > 0$, et $Y = \sigma X + m$, alors

$$F_Y(y) = F_X\left(\frac{y-m}{\sigma}\right), \qquad f_Y(y) = \frac{1}{\sigma} f_X\left(\frac{y-m}{\sigma}\right).$$

Ainsi $Y \sim \mathcal{N}(m, \sigma^2)$.

### 7.2.4 Moments des variables aléatoires réelles

L'espérance mathématique se comprend comme le barycentre de la loi de la variable aléatoire. Si $X$ est étagée, par exemple $X = \sum_{i=1}^k x_i \mathbf{1}_{A_i}$, où les $A_i$ sont disjoints et $x_i \in \mathbb{R}$, on pose

$$\mathbb{E}(X) = \sum_{i=1}^k x_i \mathbb{P}(A_i).$$

**Définition 7.24.** Si $X$ est une variable aléatoire positive, on définit

$$\mathbb{E}(X) = \int_\Omega X(\omega)\, d\mathbb{P}(\omega) = \sup\{\mathbb{E}(Y) : Y \text{ étagée positive, } 0 \leq Y \leq X\}.$$

Si $X$ est de signe quelconque, on dit qu'elle est intégrable si $\mathbb{E}(X^+) < +\infty$ et $\mathbb{E}(X^-) < +\infty$, et l'on pose

$$\mathbb{E}(X) = \mathbb{E}(X^+) - \mathbb{E}(X^-).$$

On a $X$ intégrable si et seulement si $|X|$ est intégrable.

Les résultats fondamentaux de l'intégration se traduisent immédiatement en probabilités.

**Théorème 7.25** (Convergence monotone)**.** *Si $(X_n)_{n\geq 1}$ est une suite de variables aléatoires positives telle que $X_n \uparrow X$ presque sûrement, alors*

$$\mathbb{E}(X_n) \uparrow \mathbb{E}(X).$$

**Lemme 7.26** (Fatou)**.** *Si $(X_n)_{n\geq 1}$ est une suite de variables aléatoires positives, alors*

$$\mathbb{E}\left(\liminf_{n\to\infty} X_n\right) \leq \liminf_{n\to\infty} \mathbb{E}(X_n).$$

**Théorème 7.27** (Convergence dominée)**.** *Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires. On suppose que $X_n \to X$ presque sûrement et qu'il existe une variable aléatoire positive $Y$ telle que $\mathbb{E}(Y) < +\infty$ et $|X_n| \leq Y$ presque sûrement pour tout $n$. Alors $X_n$ et $X$ sont intégrables, et*

$$\mathbb{E}(|X_n - X|) \longrightarrow 0, \qquad \mathbb{E}(X_n) \longrightarrow \mathbb{E}(X).$$

**Théorème 7.28** (Théorème de transfert)**.** *Soit $X$ une variable aléatoire réelle et soit $g : (\mathbb{R}, \mathcal{B}(\mathbb{R})) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ une fonction mesurable. Alors $g(X) \in L^1(\mathbb{P})$ si et seulement si $g \in L^1(\mathbb{P}_X)$, et, dans ce cas,*

$$\mathbb{E}(g(X)) = \int_\Omega g(X(\omega))\, d\mathbb{P}(\omega) = \int_{\mathbb{R}} g(x)\, d\mathbb{P}_X(x).$$

*Démonstration.* On commence par $g = \mathbf{1}_B$, avec $B \in \mathcal{B}(\mathbb{R})$. Alors

$$\mathbb{E}(\mathbf{1}_B(X)) = \mathbb{P}(X \in B) = \mathbb{P}_X(B) = \int_{\mathbb{R}} \mathbf{1}_B(x)\, d\mathbb{P}_X(x).$$

Par linéarité, le résultat vaut pour les fonctions étagées positives. On l'étend ensuite aux fonctions mesurables positives par le lemme d'approximation et le théorème de convergence monotone, puis aux fonctions intégrables par séparation des parties positive et négative. □

**Exemple 7.29** (Calculs d'espérance)**.**

(1) Si $X$ est discrète à valeurs dans $\{a_i : i \geq 0\}$ et si $p_i = \mathbb{P}(X = a_i)$, alors, lorsque la somme a un sens,
$$\mathbb{E}(X) = \sum_{i\geq 0} a_i p_i.$$

(2) Si $X$ admet une densité $f$, alors, lorsque $\int_{\mathbb{R}} |x|\, f(x)\, dx < +\infty$,
$$\mathbb{E}(X) = \int_{\mathbb{R}} x f(x)\, dx.$$

(3) Si $X \sim \mathcal{U}([a, b])$, alors $\mathbb{E}(X) = (a + b)/2$.

(4) Si $X \sim \mathcal{E}(\lambda)$, alors $\mathbb{E}(X) = 1/\lambda$.

*Remarque* 7.30. Deux variables de même loi ont la même espérance lorsqu'elle existe, mais la réciproque est fausse. Par exemple, une variable uniforme sur $\{-1, 1\}$ et une variable uniforme sur $\{-2, 0, 2\}$ ont toutes deux espérance 0, mais des lois différentes.

*Remarque* 7.31. Certaines lois n'ont pas d'espérance. Par exemple, si $\mathbb{P}(X = k) = c/k^2$ sur $\mathbb{N}^*$, avec $c = 6/\pi^2$, alors

$$\mathbb{E}(X) = \sum_{k\geq 1} k \frac{c}{k^2} = c \sum_{k\geq 1} \frac{1}{k} = +\infty.$$

La loi de Cauchy, de densité $x \mapsto 1/(\pi(1 + x^2))$, est un autre exemple classique de loi sans espérance.

**Proposition 7.32** (Propriétés des moments)**.** *Soit $p \geq 1$. On note $L^p = L^p(\Omega, \mathcal{A}, \mathbb{P})$ l'espace des variables aléatoires de puissance p-ième intégrable, modulo égalité presque sûre. Alors :*

(i) *si $X \in L^p$ et $1 \leq q \leq p$, alors $X \in L^q$ ;*

(ii) *$X \in L^p$ si et seulement si, pour tout $a \in \mathbb{R}$, $X - a \in L^p$ ;*

(iii) *$X \in L^2$ si et seulement si $X \in L^1$ et $X - \mathbb{E}(X) \in L^2$.*

*Si $X \in L^2$, on définit la variance par*

$$\mathrm{Var}(X) = \mathbb{E}\left((X - \mathbb{E}(X))^2\right).$$

*On a*

$$\mathrm{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2, \qquad \mathrm{Var}(aX + b) = a^2 \,\mathrm{Var}(X).$$

*De plus, $\mathbb{E}(X)$ est la meilleure approximation de X par une constante au sens quadratique :*

$$\min_{a \in \mathbb{R}} \mathbb{E}((X - a)^2) = \mathbb{E}((X - \mathbb{E}(X))^2).$$

**Proposition 7.33** (Inégalités et propriétés supplémentaires des moments)**.** *Soient les variables aléatoires considérées définies sur le même espace probabilisé. On a les propriétés suivantes.*

(i) ***Hölder.*** *Si $p, q \in [1, +\infty]$ vérifient $1/p + 1/q = 1$, si $X \in L^p$ et $Y \in L^q$, alors $XY \in L^1$ et*

$$\|XY\|_1 \leq \|X\|_p \, \|Y\|_q .$$

*Lorsque $p = q = 2$, on retrouve l'inégalité de Cauchy–Schwarz.*

(ii) ***Minkowski.*** *Si $p \in [1, +\infty]$, $X \in L^p$ et $Y \in L^p$, alors $X + Y \in L^p$ et*

$$\|X + Y\|_p \leq \|X\|_p + \|Y\|_p .$$

(iii) *Pour tout $p \in [1, +\infty]$, $L^p, \|\cdot\|_p$ est un espace de Banach.*

(iv) ***Markov.*** *Soit X une variable aléatoire réelle et soit h une fonction mesurable positive croissante. Alors, pour tout $a \in \mathbb{R}$ tel que $h(a) > 0$,*

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}(h(X))}{h(a)}.$$

(v) ***Chebyshev.*** *Si $X \in L^2$, alors, pour tout $x > 0$,*

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq x) \leq \frac{\mathrm{Var}(X)}{x^2}.$$

(vi) ***Jensen.*** *Si $X \in L^1$, si $f$ est convexe et si $f(X) \in L^1$, alors*

$$\mathbb{E}(f(X)) \geq f(\mathbb{E}(X)).$$

(vii) *Si X est positive, alors*

$$\mathbb{E}(X) = \int_0^{+\infty} \mathbb{P}(X > t)\, dt = \int_0^{+\infty} \mathbb{P}(X \geq t)\, dt,$$

*avec égalité dans $\overline{\mathbb{R}}_+$. Si X est positive et intégrable, alors*

$$x\,\mathbb{P}(X \geq x) \longrightarrow 0 \qquad \text{lorsque } x \to +\infty.$$

*Démonstration.* Les inégalités de Hölder et de Minkowski sont celles du cours d'intégration, appliquées à l'espace mesuré $(\Omega, \mathcal{A}, \mathbb{P})$. L'inégalité de Markov vient de l'inégalité ponctuelle $h(X) \geq h(a)\mathbf{1}_{\{X \geq a\}}$. L'inégalité de Chebyshev s'obtient en appliquant Markov à la variable positive $(X - \mathbb{E}(X))^2$ avec $h(t) = t^2$. Pour Jensen, par convexité, en $a = \mathbb{E}(X)$, il existe $\lambda_a \in \mathbb{R}$ tel que $f(x) \geq f(a) + \lambda_a(x - a)$. En prenant $x = X(\omega)$ et en intégrant, on obtient le résultat.

La formule des couches s'obtient en écrivant $X = \int_0^{+\infty} \mathbf{1}_{\{X > t\}}\, dt$ et en appliquant Fubini–Tonelli. Enfin, si $X \in L^1$, alors $x\mathbf{1}_{\{X \geq x\}} \leq X\mathbf{1}_{\{X \geq x\}}$, et la convergence dominée appliquée à $X\mathbf{1}_{\{X \geq x\}}$ donne la limite annoncée. □

*Remarque* 7.34. L'inégalité de Jensen permet de retrouver l'inclusion $L^p \subset L^q$ sur un espace probabilisé lorsque $1 \leq q \leq p$. En prenant $Z = |Y|^q$ et $f(x) = x^{p/q}$, on obtient

$$\big(\mathbb{E}(|Y|^q)\big)^{p/q} \leq \mathbb{E}(|Y|^p).$$

## 7.3 Méthode de la fonction muette

La méthode de la fonction muette consiste à identifier une loi en testant l'intégrale contre une classe suffisamment riche de fonctions. On note $\mathcal{H}$ l'une des classes suivantes : fonctions mesurables positives, fonctions continues bornées, ou fonctions continues à support compact.

**Proposition 7.35** (Identification d'une loi par fonctions tests)**.** *Soit $X$ une variable aléatoire réelle et soit $\mu$ une probabilité sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Si, pour toute fonction $h \in \mathcal{H}$,*

$$\mathbb{E}(h(X)) = \int_{\mathbb{R}} h(x)\, d\mu(x),$$

*alors $\mu$ est la loi de $X$. Réciproquement, si $\mu = \mathbb{P}_X$, cette égalité est le théorème de transfert.*

*Démonstration.* Pour les fonctions indicatrices $h = \mathbf{1}_A$, avec $A \in \mathcal{B}(\mathbb{R})$, l'égalité donne $\mathbb{P}_X(A) = \mu(A)$. Si l'on ne teste que contre des fonctions continues, on approche les indicatrices d'intervalles fermés $[a, b]$ par une suite décroissante de fonctions continues bornées $(h_n)$ telle que $h_n \downarrow \mathbf{1}_{[a,b]}$. Par convergence monotone appliquée à $h_1 - h_n$, on obtient $\mathbb{P}(X \in [a, b]) = \mu([a, b])$. La classe des intervalles fermés engendre la tribu borélienne et le théorème d'unicité des mesures finies conclut. □

**Exemple 7.36** (Loi de $X^2$)**.** Soit $X$ une variable aléatoire de densité $f$ sur $\mathbb{R}$, et soit $Y = X^2$. Pour toute fonction mesurable positive $h$,

$$\mathbb{E}(h(Y)) = \int_{\mathbb{R}} h(x^2) f(x)\, dx.$$

On sépare les parties positive et négative, puis on effectue le changement de variable $y = x^2$ :

$$\int_{\mathbb{R}} h(x^2) f(x)\, dx = \int_0^{+\infty} h(y) \frac{f(\sqrt{y}) + f(-\sqrt{y})}{2\sqrt{y}}\, dy.$$

Ainsi $Y$ admet la densité

$$g(y) = \frac{f(\sqrt{y}) + f(-\sqrt{y})}{2\sqrt{y}} \mathbf{1}_{\{y > 0\}}.$$

# Chapitre 8

# Vecteurs aléatoires

## 8.1 Introduction

Lorsque plusieurs variables aléatoires interviennent dans une même expérience, la probabilité d'un événement dépend souvent de leurs valeurs prises ensemble. On étudie donc la loi du vecteur complet, et non seulement les lois de chacune de ses composantes.

**Exemple 8.1** (Tirage simultané dans une urne)**.** Une urne contient 7 boules : 2 bleues, 3 blanches et 2 rouges. On prélève 3 boules d'un coup. On note $X$ le nombre de boules bleues et $Y$ le nombre de boules blanches. Le nombre de rouges est $3 - X - Y$. Pour $i \in \{0, 1, 2\}$ et $j \in \{0, 1, 2, 3\}$,

$$\mathbb{P}(X = i, Y = j) = \frac{\binom{2}{i}\binom{3}{j}\binom{2}{3-i-j}}{\binom{7}{3}},$$

avec la convention $\binom{n}{k} = 0$ si $k < 0$ ou $k > n$. On obtient

| $i \backslash j$ | 0 | 1 | 2 | 3 |
|---|---|---|---|---|
| 0 | 0 | 3/35 | 6/35 | 1/35 |
| 1 | 2/35 | 12/35 | 6/35 | 0 |
| 2 | 2/35 | 3/35 | 0 | 0 |

D'où $\mathbb{P}(X > Y) = 1/5$, $\mathbb{P}(X = Y) = 12/35$ et $\mathbb{P}(2 \text{ rouges}) = \mathbb{P}(X + Y = 1) = 1/7$.

## 8.2 Généralités

**Définition 8.2.** Un vecteur aléatoire à valeurs dans $\mathbb{R}^d$ est une application mesurable

$$X = (X_1, \ldots, X_d) : (\Omega, \mathcal{A}) \longrightarrow (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d)).$$

**Proposition 8.3** (Loi d'un vecteur aléatoire)**.** *Si $X$ est un vecteur aléatoire et si $\mathbb{P}$ est une probabilité sur $(\Omega, \mathcal{A})$, alors*

$$\mathbb{P}_X(B) = \mathbb{P}(X \in B), \qquad B \in \mathcal{B}(\mathbb{R}^d),$$

*définit une probabilité sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, appelée loi de $X$.*

**Proposition 8.4** (Lois marginales)**.** *Si $X = (X_1, \ldots, X_d)$ est un vecteur aléatoire, chaque composante $X_i$ est une variable aléatoire réelle. Sa loi, appelée* loi marginale*, est donnée par*

$$\mathbb{P}_{X_i}(B_i) = \mathbb{P}(X_i \in B_i) = \mathbb{P}(X \in \mathbb{R}^{i-1} \times B_i \times \mathbb{R}^{d-i}).$$

*Démonstration.* La composante $X_i$ s'écrit $X_i = \pi_i \circ X$, où $\pi_i$ est la projection canonique. Comme $\pi_i$ est continue, elle est borélienne. L'identité sur les lois vient de $\{X_i \in B_i\} = \{X \in \mathbb{R}^{i-1} \times B_i \times \mathbb{R}^{d-i}\}$. □

**Exemple 8.5** (Des marginales ne déterminent pas la loi jointe)**.** Soit $X = (X_1, X_2)$ uniforme sur $[0,1]^2$. Ses deux marginales sont uniformes sur $[0,1]$. Si $Y_1$ est uniforme sur $[0,1]$, $Y_2 = Y_1$ et $Y = (Y_1, Y_2)$, alors $Y$ a les mêmes marginales que $X$. Pourtant, pour la diagonale $\Delta = \{(s,t) : s = t\}$, on a $\mathbb{P}(Y \in \Delta) = 1$ alors que $\mathbb{P}(X \in \Delta) = 0$.

**Définition 8.6** (Fonction de répartition)**.** La fonction de répartition d'un vecteur $X = (X_1, \dots, X_d)$ est

$$F_X(x_1, \dots, x_d) = \mathbb{P}(X_1 \le x_1, \dots, X_d \le x_d).$$

Elle caractérise la loi, car les pavés $\prod_i ]-\infty, x_i]$ engendrent $\mathcal{B}(\mathbb{R}^d)$.

**Définition 8.7.** Une probabilité sur $\mathbb{R}^d$ est discrète si elle est portée par un ensemble au plus dénombrable. Elle est absolument continue par rapport à $\lambda_d$ s'il existe $f \in L^1_+(\lambda_d)$ telle que

$$\mu(B) = \int_B f(x_1, \dots, x_d)\, d\lambda_d(x_1, \dots, x_d).$$

On écrit alors $\mu = f\lambda_d$.

**Théorème 8.8.** *Si une probabilité sur $\mathbb{R}^d$ admet une densité $f$, alors la dérivée mixte $\partial^d F / \partial x_1 \cdots \partial x_d$ existe $\lambda_d$-p.p. et coïncide avec $f$ p.p. La réciproque est fausse sans hypothèses de régularité supplémentaires.*

**Théorème 8.9** (Densité d'une marginale)**.** *Si $X = (X_1, \dots, X_d)$ admet une densité $f_X$, alors, pour $p \le d$, le vecteur $(X_1, \dots, X_p)$ admet pour densité*

$$f_{(X_1,\dots,X_p)}(x_1, \dots, x_p) = \int_{\mathbb{R}^{d-p}} f_X(x_1, \dots, x_d)\, dx_{p+1} \cdots dx_d.$$

*Démonstration.* Pour tout borélien $B \in \mathcal{B}(\mathbb{R}^p)$, Fubini donne

$$\begin{aligned}\mathbb{P}((X_1, \dots, X_p) \in B) &= \int_{B \times \mathbb{R}^{d-p}} f_X(x_1, \dots, x_d)\, dx_1 \cdots dx_d \\ &= \int_B \left( \int_{\mathbb{R}^{d-p}} f_X(x_1, \dots, x_d)\, dx_{p+1} \cdots dx_d \right) dx_1 \cdots dx_p.\end{aligned}$$

□

**Exemple 8.10** (Loi multinomiale)**.** On considère $N$ objets de $d$ types, avec $n_i$ objets du type $i$ et $n_1 + \cdots + n_d = N$. On tire avec remise $k$ objets. Si $X_i$ est le nombre d'objets de type $i$ tirés, alors

$$\mathbb{P}(X = (x_1, \dots, x_d)) = \begin{cases} \dfrac{k!}{x_1! \cdots x_d!} \left(\dfrac{n_1}{N}\right)^{x_1} \cdots \left(\dfrac{n_d}{N}\right)^{x_d}, & \text{si } x_1 + \cdots + x_d = k, \\ 0, & \text{sinon.} \end{cases}$$

Pour $d = 2$, on retrouve une loi binomiale marginale.

**Exemple 8.11** (Loi uniforme sur un domaine)**.** Si $D \in \mathcal{B}(\mathbb{R}^d)$ vérifie $0 < \lambda_d(D) < +\infty$, la loi uniforme sur $D$ admet la densité $f_X = \lambda_d(D)^{-1} \mathbf{1}_D$. Pour le disque unité de $\mathbb{R}^2$,

$$f_{X_1}(x_1) = \frac{1}{\pi} \int_{\mathbb{R}} \mathbf{1}_{\{x_1^2 + x_2^2 \le 1\}}\, dx_2 = \frac{2}{\pi} \sqrt{1 - x_1^2}\, \mathbf{1}_{]-1,1[}(x_1).$$

## 8.3 Identification de la loi d'un vecteur aléatoire

Soit $\mathcal{H}_d$ la classe des fonctions mesurables positives, ou continues bornées, ou continues à support compact sur $\mathbb{R}^d$.

**Proposition 8.12.** *Soit $X$ un vecteur aléatoire à valeurs dans $\mathbb{R}^d$. Une probabilité $\mu$ sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ est la loi de $X$ si et seulement si, pour toute fonction $h \in \mathcal{H}_d$,*

$$\mathbb{E}(h(X)) = \int_{\mathbb{R}^d} h(x)\, d\mu(x).$$

**Proposition 8.13** (Densité d'un vecteur image)**.** *Soit $X$ de densité $f_X$. On suppose que $\mathbb{P}(X \in D) = 1$, où $D$ est un ouvert de $\mathbb{R}^d$, et que $g : D \to D' = g(D)$ est un $C^1$-difféomorphisme. Alors $Y = g(X)$ admet la densité*

$$f_Y(y) = f_X(g^{-1}(y)) \left|J_{g^{-1}}(y)\right| \mathbf{1}_{D'}(y).$$

*Démonstration.* Pour toute fonction mesurable positive $h$, on écrit

$$\mathbb{E}(h(Y)) = \int_D h(g(x)) f_X(x)\, dx.$$

Le changement de variable $y = g(x)$ donne l'expression voulue, puis on identifie la loi par la méthode de la fonction muette. □

**Exemple 8.14** (Produit et quotient)**.** Soit $(X, Y)$ de densité

$$f_{X,Y}(x, y) = \frac{1}{2\sqrt{x}} \frac{1}{y^2} \mathbf{1}_{]0,1[}(x) \mathbf{1}_{]1,+\infty[}(y).$$

Pour $U = XY$ et $V = X/Y$, on a $x = \sqrt{uv}$, $y = \sqrt{u/v}$ et $\left|J_{g^{-1}}(u, v)\right| = 1/(2v)$. L'image du domaine est

$$D' = \{(u, v) \in \mathbb{R}^2 : 0 < v < u < 1/v,\ 0 < v < 1\},$$

et donc

$$f_{U,V}(u, v) = \frac{1}{4} \frac{1}{u^{5/4}} \frac{1}{v^{1/4}} \mathbf{1}_{D'}(u, v).$$

**Exemple 8.15** (Rapport de deux gaussiennes)**.** Si $(X_1, X_2)$ a pour densité $(2\pi)^{-1} e^{-(x_1^2+x_2^2)/2}$, alors $Y_1 = X_1/X_2$ a pour densité

$$f_{Y_1}(y) = \int_{\mathbb{R}} |z| \frac{1}{2\pi} e^{-z^2(1+y^2)/2}\, dz = \frac{1}{\pi(1 + y^2)}.$$

C'est la loi de Cauchy standard.

## 8.4 Intégration des vecteurs aléatoires et moments

### 8.4.1 Espérance

**Définition 8.16.** Un vecteur aléatoire $X = (X_1, \ldots, X_d)$ est intégrable si $\|X\|$ est intégrable, pour une norme quelconque sur $\mathbb{R}^d$. Dans ce cas,

$$\mathbb{E}(X) = (\mathbb{E}(X_1), \ldots, \mathbb{E}(X_d)).$$

*Remarque* 8.17. Toutes les normes étant équivalentes sur $\mathbb{R}^d$, le choix de la norme n'a pas d'importance. De plus, $X$ est intégrable si et seulement si chaque composante $X_i$ appartient à $L^1$.

### 8.4.2 Matrice de covariance

**Définition 8.18.** Si $X$ et $Y$ sont deux variables réelles de carré intégrable, leur covariance est

$$\mathrm{Cov}(X,Y) = \mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))).$$

Si $Z = (X,Y)^t$, sa matrice de covariance est

$$C_Z = \begin{pmatrix} \mathrm{Var}(X) & \mathrm{Cov}(X,Y) \\ \mathrm{Cov}(X,Y) & \mathrm{Var}(Y) \end{pmatrix}.$$

Pour $X = (X_1, \dots, X_d)$, on pose $C_X = (\mathrm{Cov}(X_i, X_j))_{i,j}$.

**Proposition 8.19.** *On a* $\mathrm{Cov}(X,Y) = \mathrm{Cov}(Y,X)$, $\mathrm{Cov}(X,X) = \mathrm{Var}(X)$,

$$\mathrm{Cov}(X,Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y),$$

*ainsi que* $\mathrm{Cov}(X + a, Y + b) = \mathrm{Cov}(X,Y)$ *et* $\mathrm{Cov}(aX, bY) = ab\,\mathrm{Cov}(X,Y)$.

*Remarque* 8.20. La matrice $C_X$ est semi-définie positive, car

$$\alpha^t C_X \alpha = \mathbb{E}\left(\sum_{i=1}^{d} \alpha_i (X_i - \mathbb{E}(X_i))\right)^2 \geq 0.$$

### 8.4.3 Coefficient de corrélation linéaire

**Définition 8.21.** Si $X$ et $Y$ sont de carré intégrable et de variance non nulle, leur coefficient de corrélation linéaire est

$$\rho(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sigma_X \sigma_Y}.$$

**Proposition 8.22.** *On a* $|\rho(X,Y)| \leq 1$. *De plus,* $|\rho(X,Y)| = 1$ *si et seulement s'il existe* $a, b, c$, *avec* $ab \neq 0$, *tels que* $aX + bY + c = 0$ *p.s.*

*Démonstration.* Le premier point est Cauchy–Schwarz. Pour le second, on centre et réduit $X$ et $Y$. Si $\rho = \rho(X,Y)$, alors $\mathbb{E}(Y - \rho X)^2 = 1 - \rho^2$, d'où le critère. □

**Proposition 8.23.** *Si* $X_1, \dots, X_n$ *sont de carré intégrable, alors*

$$\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{Var}(X_i) + 2 \sum_{1 \leq i < j \leq n} \mathrm{Cov}(X_i, X_j).$$

*Remarque* 8.24 (Interprétation hilbertienne). Dans $L^2$, muni de $\langle X, Y \rangle = \mathbb{E}(XY)$, on a

$$\mathrm{Cov}(X,Y) = \langle X - \mathbb{E}(X), Y - \mathbb{E}(Y) \rangle.$$

Donc la covariance nulle signifie orthogonalité des parties centrées ; $|\rho| = 1$ signifie colinéarité. Le coefficient $\rho$ ne détecte pas les relations non linéaires.

## 8.5 Vecteurs aléatoires indépendants

**Définition 8.25.** Des vecteurs aléatoires $X_1, \dots, X_n$, à valeurs dans $\mathbb{R}^{d_1}, \dots, \mathbb{R}^{d_n}$, sont indépendants si, pour tous boréliens $A_i \in \mathcal{B}(\mathbb{R}^{d_i})$,

$$\mathbb{P}(X_1 \in A_1, \dots, X_n \in A_n) = \prod_{i=1}^{n} \mathbb{P}(X_i \in A_i).$$

Équivalemment,

$$\mathbb{P}_{(X_1, \dots, X_n)} = \mathbb{P}_{X_1} \otimes \cdots \otimes \mathbb{P}_{X_n}.$$

*Remarque* 8.26. L'indépendance des variables aléatoires équivaut à l'indépendance des tribus engendrées $\sigma(X_i)$. Plus généralement, des sous-tribus $\mathcal{F}_i$ sont indépendantes si l'intersection d'un événement choisi dans chacune d'elles a pour probabilité le produit des probabilités.

**Proposition 8.27** (Hérédité par blocs)**.** *Soient $X_1, \dots, X_n$ indépendantes. Découpons $\{1, \dots, n\}$ en blocs disjoints $E_1, \dots, E_m$. Si $Y_j = (X_k)_{k \in E_j}$ et $Z_j = h_j(Y_j)$, avec $h_j$ borélienne, alors $Z_1, \dots, Z_m$ sont indépendants.*

*Démonstration.* La tribu $\sigma(Z_j)$ est incluse dans $\sigma(Y_j)$, et les tribus associées aux blocs disjoints sont indépendantes. □

**Définition 8.28** (Suite infinie indépendante)**.** Une suite $(X_i)_{i \in \mathbb{N}}$ est indépendante si toute sous-famille finie est indépendante : pour tout ensemble fini non vide $K \subset \mathbb{N}$,

$$\mathbb{P}\left(\bigcap_{i \in K} \{X_i \in B_i\}\right) = \prod_{i \in K} \mathbb{P}(X_i \in B_i).$$

**Théorème 8.29** (Critères d'indépendance)**.** *Soient $X_1, \dots, X_n$ des variables aléatoires réelles.*

(i) *Elles sont indépendantes si et seulement si*

$$\mathbb{P}_{(X_1, \dots, X_n)} = \mathbb{P}_{X_1} \otimes \cdots \otimes \mathbb{P}_{X_n}.$$

(ii) *Si $X = (X_1, \dots, X_n)$ est discret, alors les composantes sont indépendantes si et seulement si, pour tout $(x_1, \dots, x_n) \in X(\Omega)$,*

$$\mathbb{P}(X_1 = x_1, \dots, X_n = x_n) = \prod_{i=1}^{n} \mathbb{P}(X_i = x_i).$$

(iii) *Si $X$ admet une densité $f_X$, alors les composantes sont indépendantes si et seulement s'il existe des fonctions mesurables positives $f_1, \dots, f_n$ telles que*

$$f_X(x_1, \dots, x_n) = f_1(x_1) \cdots f_n(x_n) \quad \lambda_n\text{-p.p.}$$

*Dans ce cas, chaque marginale admet une densité proportionnelle au facteur correspondant.*

*Démonstration.* Le premier point résulte du théorème d'unicité des mesures finies, appliqué à la classe des pavés $]-\infty, x_1] \times \cdots \times ]-\infty, x_n]$, qui engendre $\mathcal{B}(\mathbb{R}^n)$ et qui est stable par intersections finies.

Supposons maintenant $X$ discret. Si les composantes sont indépendantes, il suffit de prendre les singletons $\{x_i\}$ dans la définition de l'indépendance pour obtenir la formule annoncée. Réciproquement, soit $B = B_1 \times \cdots \times B_n$ un pavé borélien. Posons $E = X_1(\Omega) \times \cdots \times X_n(\Omega)$ et $C = E \cap B = C_1 \times \cdots \times C_n$, où $C_i = X_i(\Omega) \cap B_i$. Comme $E$ contient $X(\Omega)$, et comme les points de $E \setminus X(\Omega)$ ont probabilité nulle pour $X$, on obtient

$$\begin{aligned}
\mathbb{P}_X(B) &= \sum_{(x_1, \dots, x_n) \in C} \mathbb{P}(X_1 = x_1, \dots, X_n = x_n) \\
&= \sum_{(x_1, \dots, x_n) \in C} \prod_{i=1}^{n} \mathbb{P}(X_i = x_i) = \prod_{i=1}^{n} \sum_{x_i \in C_i} \mathbb{P}(X_i = x_i) \\
&= \prod_{i=1}^{n} \mathbb{P}(X_i \in B_i).
\end{aligned}$$

Par unicité, cela donne l'indépendance.

Supposons enfin que la densité de $X$ soit de la forme $f_1 \otimes \cdots \otimes f_n$, avec des fonctions mesurables positives. La densité de la marginale $X_i$ est alors

$$f_{X_i}(t) = \int_{\mathbb{R}^{n-1}} f_1(x_1) \cdots f_{i-1}(x_{i-1}) f_i(t) f_{i+1}(x_{i+1}) \cdots f_n(x_n)\, dx_1 \cdots dx_{i-1} dx_{i+1} \cdots dx_n,$$

qui est de la forme $c_i f_i(t)$, avec $c_i > 0$. Pour un pavé $B = B_1 \times \cdots \times B_n$, Fubini–Tonelli donne

$$\mathbb{P}_X(B) = \int_B f_1 \otimes \cdots \otimes f_n \, d\lambda_n = \prod_{i=1}^n \int_{B_i} f_i \, d\lambda.$$

En prenant $B_i = \mathbb{R}$, on voit que les constantes de normalisation se compensent et l'on retrouve le produit des marginales. Réciproquement, si les $X_i$ sont indépendantes, la loi jointe est le produit des lois marginales ; si celles-ci ont pour densités $f_{X_i}$, alors la densité jointe est $f_{X_1} \otimes \cdots \otimes f_{X_n}$. □

**Exemple 8.30.** Soit $X = (X_1, X_2)$ un vecteur aléatoire de densité

$$f(s,t) = \frac{1}{(2\pi^3)^{1/2}} \frac{1}{(1+s^2)(te^t)^{1/2}} \mathbf{1}_{]0,+\infty[}(t).$$

Cette densité se factorise en un produit d'une fonction de $s$ et d'une fonction de $t$. Les variables $X_1$ et $X_2$ sont donc indépendantes.

**Proposition 8.31** (Critère fonctionnel d'indépendance)**.** *Les vecteurs aléatoires $X_1, \ldots, X_n$, à valeurs dans $\mathbb{R}^{d_1}, \ldots, \mathbb{R}^{d_n}$, sont indépendants si et seulement si, pour toutes fonctions mesurables positives $h_i : \mathbb{R}^{d_i} \to \mathbb{R}_+$, on a*

$$\mathbb{E}\left(\prod_{i=1}^n h_i(X_i)\right) = \prod_{i=1}^n \mathbb{E}(h_i(X_i)).$$

*Démonstration.* Si les vecteurs sont indépendants, alors la loi jointe est le produit des lois marginales, et Fubini–Tonelli donne

$$\begin{aligned}\mathbb{E}\left(\prod_{i=1}^n h_i(X_i)\right) &= \int \prod_{i=1}^n h_i(x_i)\, d(\mathbb{P}_{X_1} \otimes \cdots \otimes \mathbb{P}_{X_n})(x_1, \ldots, x_n) \\ &= \prod_{i=1}^n \int h_i(x_i)\, d\mathbb{P}_{X_i}(x_i) = \prod_{i=1}^n \mathbb{E}(h_i(X_i)).\end{aligned}$$

Réciproquement, il suffit de prendre $h_i = \mathbf{1}_{A_i}$ pour retrouver la définition de l'indépendance. □

**Corollaire 8.32.** *Si $X_1, \ldots, X_n$ sont indépendantes et intégrables, alors*

$$\mathbb{E}\left(\prod_{i=1}^n X_i\right) = \prod_{i=1}^n \mathbb{E}(X_i),$$

*à condition que le produit soit intégrable. En particulier, pour deux variables indépendantes intégrables, $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$.*

*Remarque* 8.33. La réciproque est fausse : l'égalité $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$ n'implique pas l'indépendance. Par exemple, si $X \sim \mathcal{U}([-1,1])$ et $Y = X^2$, alors $\mathbb{E}(X) = 0$ et donc $\mathbb{E}(XY) = \mathbb{E}(X^3) = 0 = \mathbb{E}(X)\mathbb{E}(Y)$, mais $Y$ est entièrement déterminée par $X$.

## 8.6 Fonction caractéristique

**Définition 8.34.** Soit $\mu$ une mesure finie sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. On appelle transformée de Fourier de $\mu$ la fonction $\widehat{\mu} : \mathbb{R}^d \to \mathbb{C}$ définie par

$$\widehat{\mu}(t) = \int_{\mathbb{R}^d} e^{i\langle t,x\rangle}\, d\mu(x), \qquad t \in \mathbb{R}^d.$$

Si $\mu = \mathbb{P}_X$, on écrit

$$\varphi_X(t) = \mathbb{E}\left(e^{i\langle t,X\rangle}\right)$$

et l'on appelle $\varphi_X$ la fonction caractéristique de $X$.

La transformée de Fourier d'une mesure finie est bornée et continue. En effet,

$$|\widehat{\mu}(t)| \leq \mu(\mathbb{R}^d) = \widehat{\mu}(0),$$

et la continuité découle du théorème de continuité sous le signe intégrale.

**Théorème 8.35** (Unicité)**.** *Si deux mesures finies sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ ont la même transformée de Fourier, alors elles sont égales. En particulier, la fonction caractéristique caractérise la loi d'un vecteur aléatoire.*

**Proposition 8.36** (Moments et dérivées de la fonction caractéristique)**.** *Soit $X = (X_1, \ldots, X_d)$ un vecteur aléatoire de $\mathbb{R}^d$.*

(i) *Si $X$ est intégrable, alors $\varphi_X$ admet des dérivées partielles d'ordre un continues et*

$$\mathbb{E}(X_j) = -i\frac{\partial \varphi_X}{\partial t_j}(0), \qquad 1 \leq j \leq d.$$

(ii) *Si $X$ est de carré intégrable, alors $\varphi_X$ admet des dérivées partielles d'ordre deux continues et*

$$\mathbb{E}(X_j X_k) = -\frac{\partial^2 \varphi_X}{\partial t_j \partial t_k}(0), \qquad 1 \leq j, k \leq d.$$

*Démonstration.* C'est une application directe du théorème de dérivation sous le signe intégrale. Par exemple,

$$\frac{\partial}{\partial t_j} e^{i\langle t,X\rangle} = iX_j e^{i\langle t,X\rangle},$$

et la domination est fournie par $\left|X_j\right|$ lorsque $X_j$ est intégrable. □

Si $X$ est de carré intégrable, le développement de Taylor de $\varphi_X$ au voisinage de zéro s'écrit, pour $u \in \mathbb{R}^d$,

$$\varphi_X(u) = 1 + i\langle u, \mathbb{E}(X)\rangle - \frac{1}{2} u\, \mathbb{E}(XX^t)\, u^t + o(\|u\|^2).$$

**Exemple 8.37.** Si $X \sim \mathcal{P}(\lambda)$, alors

$$\varphi_X(t) = \exp(\lambda(e^{it} - 1)).$$

Si $X \sim \mathcal{N}(0,1)$, alors $\varphi_X(t) = e^{-t^2/2}$. En effet,

$$\varphi_X'(t) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} ixe^{-x^2/2} e^{itx}\, dx = -t\varphi_X(t),$$

après intégration par parties. Comme $\varphi_X(0) = 1$, on obtient bien $\varphi_X(t) = e^{-t^2/2}$. Plus généralement, si $X \sim \mathcal{N}(m, \sigma^2)$, alors

$$\varphi_X(t) = e^{itm} e^{-\sigma^2 t^2/2}.$$

**Théorème 8.38** (Critère d'indépendance par fonctions caractéristiques)**.** *Les variables aléatoires réelles $X_1, \dots, X_n$ sont indépendantes si et seulement si*

$$\varphi_{(X_1,\dots,X_n)}(t_1,\dots,t_n) = \prod_{k=1}^{n} \varphi_{X_k}(t_k), \qquad (t_1,\dots,t_n) \in \mathbb{R}^n.$$

*Autrement dit,*

$$\mathbb{E}\left(e^{i\sum_{k=1}^n t_k X_k}\right) = \prod_{k=1}^{n} \mathbb{E}(e^{it_k X_k}).$$

*Démonstration.* Si les variables sont indépendantes, on applique le critère fonctionnel avec $h_k(x) = e^{it_k x}$, puis on sépare parties réelle et imaginaire. Réciproquement, l'égalité des fonctions caractéristiques montre que la loi jointe et le produit des lois marginales ont même transformée de Fourier ; par unicité, elles sont égales. □

## 8.7 Somme de variables aléatoires indépendantes

La loi de $X + Y$ n'est pas déterminée par les seules lois marginales de $X$ et $Y$ lorsque les variables ne sont pas indépendantes. Par exemple, si $X \sim \mathcal{B}(1/2)$ et $Y = 1 - X$, alors $Y \sim \mathcal{B}(1/2)$, mais $X + Y = 1$ p.s. En revanche, si $Y' = X$, alors $X + Y' = 2X$, qui suit $\frac{1}{2}\delta_0 + \frac{1}{2}\delta_2$.

**Proposition 8.39.** *Si $X_1, \dots, X_n$ sont indépendants à valeurs dans $\mathbb{R}^d$, alors, pour tout $t \in \mathbb{R}^d$,*

$$\varphi_{\sum_{k=1}^n X_k}(t) = \prod_{k=1}^{n} \varphi_{X_k}(t).$$

*Démonstration.* Par indépendance,

$$\varphi_{\sum_k X_k}(t) = \mathbb{E}\left(\prod_{k=1}^{n} e^{i\langle t, X_k\rangle}\right) = \prod_{k=1}^{n} \mathbb{E}\left(e^{i\langle t, X_k\rangle}\right).$$

□

**Exemple 8.40.** Si $X \sim \mathcal{P}(\lambda)$, $Y \sim \mathcal{P}(\mu)$, et si $X$ et $Y$ sont indépendantes, alors

$$\varphi_{X+Y}(t) = \exp(\lambda(e^{it} - 1)) \exp(\mu(e^{it} - 1)) = \exp((\lambda + \mu)(e^{it} - 1)),$$

donc $X + Y \sim \mathcal{P}(\lambda + \mu)$.

### 8.7.1 Produit de convolution

Si $X$ et $Y$ sont indépendantes, la loi de $X + Y$ est déterminée par les lois de $X$ et de $Y$. On pose

$$\mathbb{P}_{X+Y} = \mathbb{P}_X * \mathbb{P}_Y.$$

Plus généralement, si $P$ et $Q$ sont deux probabilités sur $\mathbb{R}$, leur convolution est

$$P * Q = (P \otimes Q) \circ s^{-1}, \qquad s(x, y) = x + y.$$

Ainsi, pour tout borélien $B$,

$$P * Q(B) = \iint \mathbf{1}_B(x + y)\, dP(x) dQ(y).$$

Si $h$ est mesurable positive, ou intégrable pour $P * Q$, alors

$$\int_{\mathbb{R}} h(z)\, d(P * Q)(z) = \iint_{\mathbb{R}^2} h(x+y)\, dP(x) dQ(y).$$

Si $P$ a une densité $f$, alors $P * Q$ a pour densité

$$z \longmapsto \int_{\mathbb{R}} f(z-y)\, dQ(y).$$

Si $Q$ a aussi une densité $g$, alors

$$f * g(z) = \int_{\mathbb{R}} f(z-y) g(y)\, dy$$

est une densité de $P * Q$.

**Exemple 8.41** (Estimateur à noyau)**.** Soit $\mu_n = n^{-1} \sum_{i=1}^n \delta_{X_i}$ la mesure empirique et soit $\nu_n$ la mesure de densité $h_n^{-1} K(\cdot/h_n)$, où $K$ est positive et d'intégrale un. Alors $\nu_n * \mu_n$ a pour densité

$$z \longmapsto \frac{1}{nh_n} \sum_{i=1}^n K\left(\frac{z - X_i}{h_n}\right).$$

C'est l'estimateur à noyau de la densité commune lorsque les $X_i$ ont une densité commune.

*Remarque* 8.42. La convolution est associative et commutative, et $\delta_0$ est son élément neutre. Sauf dans le cas des mesures de Dirac, il n'y a pas d'inverse. Parmi les familles stables par convolution, on a

$$\mathcal{N}(m_1, \sigma_1^2) * \mathcal{N}(m_2, \sigma_2^2) = \mathcal{N}(m_1 + m_2, \sigma_1^2 + \sigma_2^2),$$

$$\mathcal{P}(\lambda) * \mathcal{P}(\mu) = \mathcal{P}(\lambda + \mu), \qquad \mathcal{B}(n_1, p) * \mathcal{B}(n_2, p) = \mathcal{B}(n_1 + n_2, p).$$

**Exemple 8.43** (Somme de deux lois exponentielles)**.** Si $X \sim \mathcal{E}(\lambda)$, $Y \sim \mathcal{E}(\mu)$, et si $X$ et $Y$ sont indépendantes, alors $Z = X + Y$ a pour densité

$$h(z) = \int_0^z \lambda\mu e^{-\lambda(z-y)} e^{-\mu y}\, dy\, \mathbf{1}_{z \geq 0}.$$

Ainsi, si $\lambda = \mu$,

$$h(z) = \lambda^2 z e^{-\lambda z} \mathbf{1}_{z \geq 0},$$

et si $\lambda \neq \mu$,

$$h(z) = \frac{\lambda\mu}{\lambda - \mu} \left(e^{-\mu z} - e^{-\lambda z}\right) \mathbf{1}_{z \geq 0}.$$

## 8.8 Vecteurs gaussiens

### 8.8.1 Loi normale dans $\mathbb{R}^n$

Une variable aléatoire réelle $X_0$ est normale centrée réduite si $X_0 \sim \mathcal{N}(0,1)$, c'est-à-dire si elle admet la densité $(2\pi)^{-1/2} e^{-x^2/2}$. Une variable réelle $X$ est normale si elle peut s'écrire

$$X = aX_0 + b$$

avec $a, b \in \mathbb{R}$ et $X_0 \sim \mathcal{N}(0,1)$. Alors $\mathbb{E}(X) = b$ et $\mathrm{Var}(X) = a^2$. Si $a \neq 0$, sa densité est

$$\frac{1}{|a|\sqrt{2\pi}} \exp\left(-\frac{1}{2} \frac{(x-b)^2}{a^2}\right).$$

**Définition 8.44.** Un vecteur aléatoire $X$ à valeurs dans $\mathbb{R}^n$ est dit gaussien si toute combinaison linéaire de ses composantes est une variable aléatoire réelle gaussienne :

$$X \text{ gaussien} \iff \langle t, X\rangle \text{ est gaussienne pour tout } t \in \mathbb{R}^n.$$

Si $A : \mathbb{R}^n \to \mathbb{R}^m$ est linéaire et $B \in \mathbb{R}^m$, alors $AX + B$ est encore un vecteur gaussien. En effet, pour tout $y \in \mathbb{R}^m$,

$$\langle y, AX + B\rangle = \langle A^t y, X\rangle + \langle y, B\rangle,$$

qui est une variable gaussienne réelle.

### 8.8.2 Fonction caractéristique d'un vecteur gaussien

**Théorème 8.45.** *Si $X$ est un vecteur gaussien de $\mathbb{R}^n$, de moyenne $\mu$ et de matrice de covariance $C_X$, alors*

$$\varphi_X(t) = \exp\left(i\langle t, \mu\rangle - \frac{1}{2}\langle t, C_X t\rangle\right), \qquad t \in \mathbb{R}^n.$$

*Démonstration.* On peut supposer $X$ centré, car le recentrage ne fait apparaître que le facteur $e^{i\langle t,\mu\rangle}$. Comme $\langle t, X\rangle$ est une variable gaussienne réelle centrée,

$$\mathbb{E}\left(e^{iu\langle t,X\rangle}\right) = \exp\left(-\frac{1}{2}u^2 \operatorname{Var}(\langle t, X\rangle)\right).$$

Or

$$\operatorname{Var}(\langle t, X\rangle) = \sum_{i,j=1}^{n} t_i t_j \mathbb{E}(X_i X_j) = \langle t, C_X t\rangle.$$

On prend $u = 1$. □

La loi d'un vecteur gaussien est donc entièrement déterminée par son espérance et sa matrice de covariance. Lorsque $C_X$ est inversible, la densité s'écrit

$$f_X(x) = \frac{1}{(2\pi)^{n/2}\sqrt{\det C_X}} \exp\left(-\frac{1}{2}(x-\mu)^t C_X^{-1}(x-\mu)\right).$$

### 8.8.3 Indépendance dans le cas gaussien

**Théorème 8.46.** *Si $X = (X_1, \ldots, X_n)$ est un vecteur gaussien, alors les variables $X_i$ sont indépendantes si et seulement si la matrice de covariance $C_X$ est diagonale.*

*Démonstration.* On peut supposer $X$ centré. Si les $X_i$ sont indépendantes, alors les covariances hors diagonale sont nulles. Réciproquement, si $C_X$ est diagonale, la fonction caractéristique de $X$ se factorise :

$$\varphi_X(t) = \exp\left(-\frac{1}{2}\sum_{i=1}^{n} t_i^2 \mathbb{E}(X_i^2)\right) = \prod_{i=1}^{n} \varphi_{X_i}(t_i).$$

Le critère d'indépendance par fonctions caractéristiques donne le résultat. □

*Remarque* 8.47. Le fait que les composantes soient gaussiennes ne suffit pas à rendre le vecteur gaussien. Soient $X \sim \mathcal{N}(0,1)$ et $\varepsilon$ indépendante de $X$, avec $\mathbb{P}(\varepsilon = 1) = \mathbb{P}(\varepsilon = -1) = 1/2$. Posons $Y = \varepsilon X$. Alors $Y \sim \mathcal{N}(0,1)$ et $\operatorname{Cov}(X, Y) = 0$, mais $\mathbb{P}(X + Y = 0) = 1/2$. Le couple $(X, Y)$ n'est donc pas gaussien.

# Chapitre 9

# Théorèmes limites

## 9.1 Convergences de suites de variables aléatoires

### 9.1.1 Convergence presque sûre

**Définition 9.1.** Soient $(Y_n)_{n\geq 1}$ et $Y$ des variables aléatoires définies sur le même espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$. On dit que $Y_n$ converge presque sûrement vers $Y$ si

$$\mathbb{P}\left(\left\{\omega \in \Omega : \lim_{n\to\infty} Y_n(\omega) = Y(\omega)\right\}\right) = 1.$$

On écrit alors $Y_n \to Y$ p.s.

Pour justifier que l'ensemble de convergence est mesurable, on utilise une suite $(\varepsilon_i)_{i\geq 1}$ de réels strictement positifs tendant vers zéro. Alors

$$\{Y_n \to Y\} = \bigcap_{i\geq 1} \bigcup_{j\geq 1} \bigcap_{k\geq j} \{|Y_k - Y| \leq \varepsilon_i\}.$$

Cette écriture ne fait intervenir que des opérations dénombrables sur des événements mesurables.

**Lemme 9.2.** *Si $(B_i)_{i\geq 1}$ est une suite d'événements, alors*

$$\mathbb{P}\left(\bigcap_{i\geq 1} B_i\right) = 1 \quad \Longleftrightarrow \quad \mathbb{P}(B_i) = 1 \text{ pour tout } i \geq 1.$$

*Démonstration.* L'implication directe est immédiate. Réciproquement,

$$\mathbb{P}\left(\left(\bigcap_{i\geq 1} B_i\right)^c\right) = \mathbb{P}\left(\bigcup_{i\geq 1} B_i^c\right) \leq \sum_{i\geq 1} \mathbb{P}(B_i^c) = 0.$$

□

**Proposition 9.3** (Critère de convergence presque sûre)**.** *On a $Y_n \to Y$ p.s. si et seulement si, pour tout $\varepsilon > 0$,*

$$\mathbb{P}\left(\bigcup_{j\geq 1} \bigcap_{k\geq j} \{|Y_k - Y| \leq \varepsilon\}\right) = 1.$$

*De manière équivalente,*

$$\mathbb{P}\left(\bigcap_{j\geq 1} \bigcup_{k\geq j} \{|Y_k - Y| > \varepsilon\}\right) = 0.$$

*Démonstration.* La première équivalence se déduit de l'écriture précédente et de la croissance en $\varepsilon$. La seconde est obtenue par passage au complémentaire. L'ensemble

$$\bigcap_{j\geq 1}\bigcup_{k\geq j}\{|Y_k - Y| > \varepsilon\}$$

est le lim sup des événements $A_k = \{|Y_k - Y| > \varepsilon\}$, c'est-à-dire l'ensemble des réalisations pour lesquelles une infinité de ces événements se produisent. □

**Proposition 9.4** (Condition suffisante de convergence presque sûre)**.** *Si, pour tout $\varepsilon > 0$,*

$$\sum_{n\geq 1}\mathbb{P}(|Y_n - Y| > \varepsilon) < +\infty,$$

*alors $Y_n \to Y$ presque sûrement. On dit alors que $Y_n$ converge presque complètement vers $Y$.*

*Démonstration.* Pour chaque $\varepsilon > 0$, on applique le lemme de Borel–Cantelli aux événements $A_n = \{|Y_n - Y| > \varepsilon\}$. Leur lim sup est de probabilité nulle, ce qui donne le critère précédent. □

*Remarque* 9.5. La convergence presque sûre entraîne que, pour tout $\varepsilon > 0$,

$$\mathbb{P}\left(\sup_{k\geq j}|Y_k - Y| > \varepsilon\right) \longrightarrow 0.$$

En particulier,

$$\mathbb{P}(|Y_n - Y| > \varepsilon) \longrightarrow 0.$$

### 9.1.2 Convergence en probabilité

**Définition 9.6.** On dit que $Y_n$ converge en probabilité vers $Y$ si, pour tout $\varepsilon > 0$,

$$\mathbb{P}(|Y_n - Y| > \varepsilon) \longrightarrow 0.$$

On écrit $Y_n \to Y$ en probabilité, ou $Y_n \xrightarrow{\mathbb{P}} Y$.

**Proposition 9.7.** *La convergence presque sûre implique la convergence en probabilité.*

*Démonstration.* C'est la conséquence de la remarque précédente. □

**Exemple 9.8.** La réciproque est fausse. Soit $(X_n)$ une suite de variables indépendantes telles que, pour $x \geq 0$,

$$\mathbb{P}(X_n > x) = \frac{1}{1+x}.$$

Posons $Y_n = X_n/n$. Alors, pour tout $\varepsilon > 0$,

$$\mathbb{P}(|Y_n| > \varepsilon) = \frac{1}{1+n\varepsilon} \longrightarrow 0,$$

donc $Y_n \to 0$ en probabilité. En revanche,

$$\sum_{n\geq 1}\mathbb{P}(|Y_n| > \varepsilon) = +\infty,$$

et, par indépendance et Borel–Cantelli, l'événement $|Y_n| > \varepsilon$ se produit infiniment souvent presque sûrement. Il n'y a donc pas convergence presque sûre vers zéro.

**Proposition 9.9.** *Si $Y_n \to Y$ en probabilité, alors il existe une sous-suite $(Y_{n_i})_{i\geq 1}$ qui converge presque sûrement vers $Y$.*

*Démonstration.* Comme $Y_n \to Y$ en probabilité, pour tout $i \geq 1$ on peut choisir $n_i$ tel que

$$\mathbb{P}(|Y_{n_i} - Y| > 2^{-i}) \leq i^{-2}.$$

Pour tout $\varepsilon > 0$, il existe $i_0$ tel que $2^{-i} < \varepsilon$ pour $i \geq i_0$, et donc

$$\sum_{i \geq 1} \mathbb{P}(|Y_{n_i} - Y| > \varepsilon) < +\infty.$$

La condition suffisante de convergence presque sûre donne $Y_{n_i} \to Y$ p.s. □

### 9.1.3 Convergence dans $L^p$

**Définition 9.10.** Soit $p \geq 1$. On dit que $Y_n$ converge en moyenne d'ordre $p$, ou dans $L^p$, vers $Y$ si

$$\mathbb{E}(|Y_n - Y|^p) \longrightarrow 0.$$

On écrit $Y_n \to Y$ dans $L^p$.

*Remarque* 9.11. Si $Y_n \to Y$ dans $L^p$, alors $Y \in L^p$. Par Markov, la convergence dans $L^p$ implique la convergence en probabilité. De plus, si $1 \leq p < r < +\infty$, alors la convergence dans $L^r$ implique la convergence dans $L^p$, dès que l'on travaille sur un espace probabilisé.

On peut résumer les implications de base par le diagramme suivant, pour $1 \leq p < r < +\infty$ :

$$\text{convergence p.s.} \implies \text{convergence en probabilité,}$$
$$\text{convergence dans } L^r \implies \text{convergence dans } L^p \implies \text{convergence en probabilité.}$$

Il n'y a pas d'autre implication générale.

**Exemple 9.12** (Convergence dans $L^p$ sans convergence presque sûre)**.** Soient $(Y_n)_{n \geq 1}$ indépendantes avec

$$Y_n \sim \left(1 - \frac{1}{n}\right)\delta_0 + \frac{1}{n}\delta_1.$$

Alors $\mathbb{E}(|Y_n|^p) = 1/n \to 0$, donc $Y_n \to 0$ dans $L^p$ pour tout $p \geq 1$. Mais $\sum_n \mathbb{P}(Y_n = 1) = +\infty$, et Borel–Cantelli donne $Y_n = 1$ infiniment souvent presque sûrement. Donc $Y_n$ ne converge pas presque sûrement vers zéro.

**Exemple 9.13** (Convergence presque sûre sans convergence dans $L^p$)**.** La convergence presque sûre n'entraîne pas, en général, la convergence dans $L^p$. Soit $U$ une variable de loi uniforme sur $[0, 1]$, et posons

$$Y_n = n^2 \mathbf{1}_{[0,1/n]}(U).$$

Pour presque tout $\omega$, on a $U(\omega) > 0$, donc, pour $n$ assez grand, $U(\omega) > 1/n$ et ainsi $Y_n(\omega) = 0$. Par conséquent

$$Y_n \longrightarrow 0 \qquad \text{p.s.}$$

En revanche, pour tout $p \geq 1$,

$$\mathbb{E}(|Y_n|^p) = n^{2p}\,\mathbb{P}(U \in [0, 1/n]) = n^{2p-1} \longrightarrow +\infty.$$

Donc $Y_n$ ne converge pas vers 0 dans $L^p$.

## 9.2 Loi des grands nombres

Soit $(X_k)_{k\geq 1}$ une suite de variables aléatoires réelles définies sur le même espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$. On étudie le comportement asymptotique des moyennes arithmétiques

$$M_n = \frac{S_n}{n} = \frac{1}{n}\sum_{k=1}^{n} X_k.$$

L'idée générale est que, sous des hypothèses d'indépendance et d'intégrabilité, la moyenne empirique se stabilise autour de l'espérance commune.

**Théorème 9.14** (Loi faible des grands nombres)**.** *Soit $(X_k)_{k\geq 1}$ une suite de variables aléatoires de même loi, de carré intégrable et deux à deux non corrélées, c'est-à-dire*

$$\mathrm{Cov}(X_i, X_j) = 0 \qquad \text{si } i \neq j.$$

*Alors*

$$\frac{1}{n}\sum_{k=1}^{n} X_k \xrightarrow[n\to\infty]{\mathbb{P}} \mathbb{E}(X_1).$$

*Démonstration.* On a $\mathbb{E}(X_k) = \mathbb{E}(X_1)$ pour tout $k$. Pour $\varepsilon > 0$, l'inégalité de Chebychev donne

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{k=1}^{n} X_k - \mathbb{E}(X_1)\right| \geq \varepsilon\right) \leq \frac{1}{\varepsilon^2}\,\mathrm{Var}\left(\frac{1}{n}\sum_{k=1}^{n} X_k\right).$$

Par absence de corrélations,

$$\mathrm{Var}\left(\frac{1}{n}\sum_{k=1}^{n} X_k\right) = \frac{1}{n^2}\sum_{k=1}^{n}\mathrm{Var}(X_k) = \frac{1}{n}\mathrm{Var}(X_1).$$

Le membre de droite tend vers 0, ce qui donne la convergence en probabilité. □

La loi faible donne une convergence en probabilité. Sous des hypothèses naturelles légèrement différentes, on peut obtenir une convergence presque sûre.

**Théorème 9.15** (Loi forte des grands nombres de Khintchine)**.** *Soit $(X_k)_{k\geq 1}$ une suite de variables aléatoires réelles indépendantes, de même loi et intégrables. Alors*

$$\frac{1}{n}\sum_{k=1}^{n} X_k \longrightarrow \mathbb{E}(X_1) \qquad \text{p.s.}$$

*L'hypothèse $\mathbb{E}|X_1| < +\infty$ est optimale.*

*Remarque* 9.16. Si les $X_k$ sont indépendantes, de même loi, positives, mais non intégrables, alors la même conclusion reste vraie avec limite $+\infty$. En effet, si $\mathbb{E}(X_1) = +\infty$, on tronque $X_i$ par $Y_i = \min(X_i, a)$. Pour $a$ assez grand, $\mathbb{E}(Y_1)$ est aussi grand que l'on veut. Les variables $Y_i$ sont iid et bornées, donc la loi forte appliquée aux $Y_i$ donne

$$\frac{1}{n}\sum_{i=1}^{n} X_i \geq \frac{1}{n}\sum_{i=1}^{n} Y_i \longrightarrow \mathbb{E}(Y_1) \qquad \text{p.s.}$$

Puis on fait tendre $a$, et donc $\mathbb{E}(Y_1)$, vers l'infini.

*Preuve de la loi forte dans le cas $L^4$.* On donne ici une preuve simple lorsque les variables sont dans $L^4$ et centrées. Posons $S_n = \sum_{k=1}^n X_k$. Comme les variables sont iid et centrées,

$$\mathbb{E}(S_n^4) = n\mathbb{E}(X_1^4) + 3n(n-1)\big(\mathbb{E}(X_1^2)\big)^2.$$

Il existe donc une constante $C$ telle que $\mathbb{E}(S_n^4) \le Cn^2$. Par l'inégalité de Markov, pour tout $\varepsilon > 0$,

$$\mathbb{P}(|S_n| > \varepsilon n) \le \frac{\mathbb{E}(S_n^4)}{\varepsilon^4 n^4} \le \frac{C}{n^2}.$$

La série des probabilités est sommable ; Borel–Cantelli donne donc $S_n/n \to 0$ presque sûrement. □

**Exemple 9.17.** Soient $(Y_n)_{n\ge1}$ et $(Z_n)_{n\ge1}$ deux suites indépendantes de variables indépendantes de loi uniforme sur $[0,1]$. On veut étudier

$$\frac{1}{n}\sum_{i=1}^n \mathbf{1}_{\{Y_i+Z_i\le1\}}.$$

Les variables $\mathbf{1}_{\{Y_i+Z_i\le1\}}$ sont iid et intégrables. Leur espérance vaut

$$\mathbb{E}\left(\mathbf{1}_{\{Y_1+Z_1\le1\}}\right) = \mathbb{P}(Y_1+Z_1\le1) = \int_0^1\int_0^{1-y} dz\,dy = \frac{1}{2}.$$

La loi forte des grands nombres donne donc

$$\frac{1}{n}\sum_{i=1}^n \mathbf{1}_{\{Y_i+Z_i\le1\}} \longrightarrow \frac{1}{2} \qquad \text{p.s.}$$

## 9.3 Convergence en loi

**Définition 9.18.** Soient $(X_n)_{n\ge1}$ et $X$ des vecteurs aléatoires à valeurs dans $\mathbb{R}^d$. On dit que $X_n$ converge en loi vers $X$, et l'on écrit $X_n \xrightarrow{\mathcal{L}} X$, si, pour toute fonction continue bornée $f : \mathbb{R}^d \to \mathbb{R}$,

$$\mathbb{E}(f(X_n)) \longrightarrow \mathbb{E}(f(X)).$$

*Remarque* 9.19. La convergence en loi signifie seulement que la loi de $X_n$ converge vers la loi de $X$. La variable limite n'est donc déterminée qu'à travers sa loi. En particulier, elle n'est pas unique comme variable aléatoire définie sur un espace probabilisé donné.

*Remarque* 9.20. Même si $X_n \xrightarrow{\mathcal{L}} X$ et $Y_n \xrightarrow{\mathcal{L}} Y$, il n'est pas vrai en général que $X_n + Y_n \xrightarrow{\mathcal{L}} X + Y$. Par exemple, si $X \sim \mathcal{N}(0,1)$, si $X_n = X$ et $Y_n = -X$, alors $X_n$ et $Y_n$ ont toutes deux pour limite en loi une loi normale centrée réduite, mais $X_n + Y_n = 0$ pour tout $n$.

**Théorème 9.21** (Théorème de l'image continue)**.** *Soient $X_n$ et $X$ des vecteurs aléatoires de $\mathbb{R}^d$, et soit $g : \mathbb{R}^d \to \mathbb{R}^k$ continue. Si $X_n \xrightarrow{\mathcal{L}} X$, alors*

$$g(X_n) \xrightarrow{\mathcal{L}} g(X).$$

*Démonstration.* Si $h : \mathbb{R}^k \to \mathbb{R}$ est continue bornée, alors $h \circ g$ est continue bornée sur $\mathbb{R}^d$. Donc

$$\mathbb{E}(h(g(X_n))) = \mathbb{E}((h\circ g)(X_n)) \longrightarrow \mathbb{E}((h\circ g)(X)) = \mathbb{E}(h(g(X))).$$

□

**Corollaire 9.22.** *Si* $(X_n, Y_n) \xrightarrow{\mathcal{L}} (X, Y)$*, alors*

$$X_n \xrightarrow{\mathcal{L}} X, \qquad Y_n \xrightarrow{\mathcal{L}} Y,$$

*et, pour tous* $a, b \in \mathbb{R}$*,*

$$aX_n + bY_n \xrightarrow{\mathcal{L}} aX + bY, \qquad X_nY_n \xrightarrow{\mathcal{L}} XY.$$

**Proposition 9.23** (Critère par les fonctions de répartition en dimension un)**.** *Soient* $X_n$ *et* $X$ *des variables aléatoires réelles, de fonctions de répartition* $F_n$ *et* $F$*. Alors*

$$X_n \xrightarrow{\mathcal{L}} X$$

*si et seulement si, pour tout point de continuité* $x$ *de* $F$*,*

$$F_n(x) \longrightarrow F(x).$$

**Théorème 9.24** (Convergence en probabilité et convergence en loi)**.** *Si* $X_n$ *et* $X$ *sont définies sur le même espace probabilisé et si* $X_n \xrightarrow{\mathbb{P}} X$*, alors*

$$X_n \xrightarrow{\mathcal{L}} X.$$

*Démonstration.* On utilise la caractérisation par les fonctions uniformément continues bornées. Soit $f$ une telle fonction. Pour $\varepsilon > 0$, il existe $\eta > 0$ tel que $|x - y| < \eta$ implique $|f(x) - f(y)| < \varepsilon$. Alors

$$\begin{aligned} |\mathbb{E}(f(X_n)) - \mathbb{E}(f(X))| &\leq \int_{\{|X_n - X| < \eta\}} |f(X_n) - f(X)| \, d\mathbb{P} \\ &\quad + \int_{\{|X_n - X| \geq \eta\}} |f(X_n) - f(X)| \, d\mathbb{P} \\ &\leq \varepsilon + 2 \|f\|_\infty \mathbb{P}(|X_n - X| \geq \eta). \end{aligned}$$

Le dernier terme tend vers $\varepsilon$, puis on fait tendre $\varepsilon$ vers zéro. □

*Remarque* 9.25. La réciproque est fausse en général. Pour des variables de Bernoulli $\mathbb{P}(X_n = 1) = p_n$, $X_n$ converge en loi si et seulement si $p_n$ converge. Mais $X_n$ converge en probabilité vers 0 si et seulement si $p_n \to 0$, et vers 1 si et seulement si $p_n \to 1$. Pour une limite non constante, la convergence en loi ne force donc pas une convergence en probabilité.

**Proposition 9.26.** *Si* $X_n \xrightarrow{\mathcal{L}} a$*, où* $a \in \mathbb{R}^d$ *est constant, alors*

$$X_n \xrightarrow{\mathbb{P}} a.$$

*Démonstration.* Pour simplifier, prenons $d = 1$. Soit $\varepsilon > 0$. On choisit une fonction continue bornée $f$ telle que $f(a) = 0$ et $f(x) = 1$ lorsque $|x - a| \geq \varepsilon$. Alors

$$\mathbb{P}(|X_n - a| \geq \varepsilon) \leq \mathbb{E}(f(X_n)) \longrightarrow f(a) = 0.$$

□

**Théorème 9.27** (Théorème de Lévy)**.** *Soient* $(X_n)_{n \geq 1}$ *et* $X$ *des vecteurs aléatoires de* $\mathbb{R}^d$*. Alors*

$$X_n \xrightarrow{\mathcal{L}} X \quad \Longleftrightarrow \quad \varphi_{X_n}(t) \longrightarrow \varphi_X(t) \quad \textit{pour tout } t \in \mathbb{R}^d.$$

**Théorème 9.28** (Théorème de Slutsky)**.** *Soient $X_n$ et $Y_n$ des variables aléatoires réelles et $a \in \mathbb{R}$. Si $X_n \xrightarrow{\mathcal{L}} X$ et $Y_n \xrightarrow{\mathbb{P}} a$, alors*

$$(X_n, Y_n) \xrightarrow{\mathcal{L}} (X, a).$$

*En particulier,*

$$X_n + Y_n \xrightarrow{\mathcal{L}} X + a, \qquad X_n Y_n \xrightarrow{\mathcal{L}} aX.$$

*Démonstration.* On utilise les fonctions caractéristiques. Pour $(u, v) \in \mathbb{R}^2$, posons

$$\rho_n = \mathbb{E}\left(e^{i(uX_n + vY_n)}\right) - \mathbb{E}\left(e^{i(uX + va)}\right).$$

Alors

$$|\rho_n| \leq \mathbb{E}\left(\left|e^{ivY_n} - e^{iva}\right|\right) + \left|\mathbb{E}(e^{iva}(e^{iuX_n} - e^{iuX}))\right|.$$

Le second terme tend vers zéro par convergence en loi de $X_n$. Pour le premier, la fonction $y \mapsto \left|e^{ivy} - e^{iva}\right|$ est continue bornée et $Y_n \to a$ en probabilité, donc aussi en loi. Ce terme tend donc vers 0. Le théorème de Lévy donne la convergence du couple, puis le théorème de l'image continue donne les deux conséquences. □

**Théorème 9.29** (Théorème de la limite centrale)**.** *Soit $(X_i)_{i \geq 1}$ une suite de variables aléatoires réelles iid, de carré intégrable. Posons*

$$S_n = \sum_{i=1}^{n} X_i, \qquad m = \mathbb{E}(X_1), \qquad \sigma^2 = \mathrm{Var}(X_1).$$

*Alors*

$$\frac{S_n - nm}{\sqrt{n}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2).$$

*Si $\sigma^2 > 0$, cela équivaut à*

$$\frac{S_n - nm}{\sigma\sqrt{n}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1).$$

*Démonstration.* On suppose d'abord $m = 0$. Pour $t \in \mathbb{R}$, par indépendance,

$$\varphi_{S_n/\sqrt{n}}(t) = \mathbb{E}\left(e^{itS_n/\sqrt{n}}\right) = \left(\varphi_X\left(\frac{t}{\sqrt{n}}\right)\right)^n.$$

Comme $X_1 \in L^2$, la fonction caractéristique est deux fois dérivable en zéro et

$$\varphi_X(u) = 1 + u\varphi_X'(0) + \frac{u^2}{2}\varphi_X''(0) + u^2\varepsilon(u), \qquad \varepsilon(u) \to 0.$$

Or $\varphi_X'(0) = i\mathbb{E}(X_1) = 0$ et $\varphi_X''(0) = -\mathbb{E}(X_1^2) = -\sigma^2$. Donc

$$\varphi_X\left(\frac{t}{\sqrt{n}}\right) = 1 - \frac{t^2\sigma^2}{2n} + \frac{t^2}{n}\varepsilon\left(\frac{t}{\sqrt{n}}\right).$$

Il reste à utiliser le fait élémentaire suivant : si $a_n \to a$, alors $(1 + a_n/n)^n \to e^a$. Ici on obtient

$$\left(1 - \frac{t^2\sigma^2}{2n} + \frac{t^2}{n}\varepsilon\left(\frac{t}{\sqrt{n}}\right)\right)^n \longrightarrow e^{-t^2\sigma^2/2},$$

qui est la fonction caractéristique de $\mathcal{N}(0, \sigma^2)$. Le cas général se ramène au cas centré en remplaçant $X_i$ par $X_i - m$. □

**Exemple 9.30** (Une conséquence asymptotique pour la loi de Poisson)**.** Soit $(X_n)_{n\geq 1}$ une suite de variables indépendantes de loi de Poisson de paramètre 1, et posons $S_n = \sum_{k=1}^n X_k$. Alors $\mathbb{E}(X_1) = 1$ et $\mathrm{Var}(X_1) = 1$, donc

$$\frac{S_n - n}{\sqrt{n}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1).$$

En particulier, comme $S_n \sim \mathcal{P}(n)$, on peut déduire

$$e^{-n}\sum_{k=0}^{n}\frac{n^k}{k!} \longrightarrow \frac{1}{2}.$$

## 9.4 Complément : preuve de la loi forte des grands nombres

Nous donnons maintenant la preuve de la loi forte des grands nombres dans le cas intégrable. L'outil principal est l'inégalité maximale de Kolmogorov.

**Théorème 9.31** (Inégalité maximale de Kolmogorov)**.** *Soient $X_1, \ldots, X_n$ des variables aléatoires réelles indépendantes, de carré intégrable et d'espérance nulle. Posons $S_k = \sum_{i=1}^k X_i$. Alors, pour tout $t > 0$,*

$$\mathbb{P}\left(\max_{1\leq k\leq n} |S_k| \geq t\right) \leq \frac{1}{t^2}\mathbb{E}(S_n^2).$$

*Démonstration.* Pour $1 \leq k \leq n$, posons

$$A_1 = \{|S_1| \geq t\},$$

et, pour $k \geq 2$,

$$A_k = \{|S_k| \geq t\} \cap \bigcap_{j=1}^{k-1}\{|S_j| < t\}.$$

Les $A_k$ sont disjoints et leur union est

$$A = \left\{\max_{1\leq k\leq n} |S_k| \geq t\right\}.$$

Sur $A_k$, on a $S_k^2 \geq t^2$, donc

$$t^2\mathbb{P}(A) \leq \sum_{k=1}^{n}\mathbb{E}(S_k^2\mathbf{1}_{A_k}).$$

Or

$$S_n^2 = (S_k + S_n - S_k)^2 \geq S_k^2 + 2S_k(S_n - S_k).$$

Comme $A_k$ dépend seulement de $X_1, \ldots, X_k$, la variable $S_k\mathbf{1}_{A_k}$ est indépendante de $S_n - S_k$, et $\mathbb{E}(S_n - S_k) = 0$. Ainsi

$$\mathbb{E}(S_n^2\mathbf{1}_{A_k}) \geq \mathbb{E}(S_k^2\mathbf{1}_{A_k}).$$

En sommant,

$$t^2\mathbb{P}(A) \leq \sum_{k=1}^{n}\mathbb{E}(S_n^2\mathbf{1}_{A_k}) = \mathbb{E}(S_n^2\mathbf{1}_A) \leq \mathbb{E}(S_n^2).$$

□

**Théorème 9.32.** *Soit $(X_k)_{k\geq 1}$ une suite de variables aléatoires réelles indépendantes, de carré intégrable, centrées, telle que*

$$\sum_{k\geq 1} \frac{\mathbb{E}(X_k^2)}{k^2} < +\infty.$$

*Alors*

$$\frac{1}{n}\sum_{k=1}^{n} X_k \longrightarrow 0 \qquad p.s.$$

*Démonstration.* Posons $S_n = \sum_{k=1}^n X_k$. On veut montrer que, pour tout $\varepsilon > 0$,

$$\mathbb{P}\left(\limsup_{n\to\infty}\left\{\left|\frac{S_n}{n}\right| \geq \varepsilon\right\}\right) = 0.$$

On regroupe les entiers par blocs dyadiques. Pour $k \geq 0$, posons

$$B_k = \left\{\exists n \in [2^k, 2^{k+1}[\ |S_n| \geq n\varepsilon\right\}.$$

Alors

$$B_k \subset \left\{\max_{1\leq j\leq 2^{k+1}} |S_j| \geq 2^k\varepsilon\right\}.$$

Par l'inégalité maximale de Kolmogorov,

$$\mathbb{P}(B_k) \leq \frac{1}{2^{2k}\varepsilon^2}\mathbb{E}(S_{2^{k+1}}^2) = \frac{1}{2^{2k}\varepsilon^2}\sum_{j=1}^{2^{k+1}}\mathbb{E}(X_j^2).$$

En sommant sur $k$, puis en échangeant les sommes,

$$\sum_{k\geq 0}\mathbb{P}(B_k) \leq C_\varepsilon \sum_{j\geq 1}\frac{\mathbb{E}(X_j^2)}{j^2} < +\infty.$$

Borel–Cantelli donne que seuls un nombre fini de $B_k$ se réalisent presque sûrement, ce qui implique $S_n/n \to 0$ p.s. □

**Corollaire 9.33.** *Si les variables ne sont pas centrées et si*

$$\sum_{k\geq 1}\frac{\mathrm{Var}(X_k)}{k^2} < +\infty,$$

*alors*

$$\frac{1}{n}\sum_{k=1}^{n} X_k - \frac{1}{n}\sum_{k=1}^{n}\mathbb{E}(X_k) \longrightarrow 0 \qquad p.s.$$

*En particulier, si $\mathbb{E}(X_k) \to \ell \in \mathbb{R}$, alors, par Cesàro,*

$$\frac{1}{n}\sum_{k=1}^{n} X_k \longrightarrow \ell \qquad p.s.$$

*Preuve de la loi forte dans le cas intégrable.* On suppose, sans perte de généralité, que $\mathbb{E}(X_1) = 0$. On tronque les variables en posant

$$Y_k = X_k \mathbf{1}_{\{|X_k|\leq k\}}.$$

Comme les $X_k$ ont même loi,

$$\sum_{k\geq 1}\mathbb{P}(X_k \neq Y_k) = \sum_{k\geq 1}\mathbb{P}(|X_1| > k) \leq \mathbb{E}|X_1| < +\infty.$$

Par Borel–Cantelli, presque sûrement, $X_k = Y_k$ pour tout $k$ assez grand. Donc

$$\frac{1}{n}\sum_{k=1}^{n} X_k - \frac{1}{n}\sum_{k=1}^{n} Y_k \longrightarrow 0 \qquad \text{p.s.}$$

Il suffit donc d'étudier les $Y_k$. Comme $\mathbb{E}(Y_k) \to 0$, il suffit de montrer

$$\frac{1}{n}\sum_{k=1}^{n} (Y_k - \mathbb{E}(Y_k)) \longrightarrow 0 \qquad \text{p.s.}$$

On applique le théorème précédent à $Y_k - \mathbb{E}(Y_k)$. Il reste à vérifier

$$\sum_{k\geq 1} \frac{\mathrm{Var}(Y_k)}{k^2} < +\infty.$$

Comme $\mathrm{Var}(Y_k) \leq \mathbb{E}(Y_k^2)$,

$$\sum_{k\geq 1} \frac{\mathrm{Var}(Y_k)}{k^2} \leq \sum_{k\geq 1} \frac{\mathbb{E}(X_1^2 \mathbf{1}_{\{|X_1|\leq k\}})}{k^2}.$$

En échangeant somme et espérance et en utilisant la comparaison $\sum_{k\geq m} k^{-2} \leq C/m$, on obtient

$$\sum_{k\geq 1} \frac{\mathbb{E}(X_1^2 \mathbf{1}_{\{|X_1|\leq k\}})}{k^2} \leq C\mathbb{E}|X_1| < +\infty.$$

La loi forte est démontrée. □

**Exemple 9.34.** Soit $(X_n)_{n\geq 2}$ une suite de variables indépendantes telles que $X_n \in \{-n, 0, n\}$ et

$$\mathbb{P}(X_n = n) = \mathbb{P}(X_n = -n) = \frac{1}{2n(\log n)^2}, \qquad \mathbb{P}(X_n = 0) = 1 - \frac{1}{n(\log n)^2}.$$

Alors $\mathbb{E}(X_n) = 0$ et

$$\mathrm{Var}(X_n) = \mathbb{E}(X_n^2) = n^2 \frac{1}{n(\log n)^2} = \frac{n}{(\log n)^2}.$$

Ainsi

$$\sum_{n\geq 2} \frac{\mathrm{Var}(X_n)}{n^2} = \sum_{n\geq 2} \frac{1}{n(\log n)^2} < +\infty.$$

Le corollaire précédent donne donc

$$\frac{1}{n}\sum_{k=2}^{n} X_k \longrightarrow 0 \qquad \text{p.s.}$$