



Université de Montpellier - Faculté des Sciences

Année universitaire 2023-2024

Licence 3 Mathématiques



---

# Théorie des probabilités

---

Nicolas Meyer

`nicolas.meyer@umontpellier.fr`



# Table des matières

<b>1</b>	<b>Bases de la théorie des probabilités</b>	<b>5</b>
1	Espace probabilisé . . . . .	5
1.1	Probabilité . . . . .	5
1.2	Exemples d'espaces probabilisés . . . . .	6
1.3	Quelques propriétés d'une probabilité . . . . .	9
2	Variable aléatoire . . . . .	11
2.1	Loi d'une variable aléatoire . . . . .	11
2.2	Lois usuelles . . . . .	14
2.3	Moments d'une variable aléatoire . . . . .	17
2.4	Fonction associée à une variable aléatoire . . . . .	23
<b>2</b>	<b>Indépendance</b>	<b>33</b>
1	Indépendance d'événements . . . . .	33
1.1	Conditionnement . . . . .	33
1.2	Quelques formules . . . . .	34
1.3	Indépendance d'événements . . . . .	35
2	Indépendance de variables aléatoires . . . . .	36
2.1	Indépendance de tribus . . . . .	36
2.2	Indépendance de variables aléatoires . . . . .	37
2.3	Critères d'indépendance . . . . .	40
3	Résultats asymptotiques . . . . .	41
3.1	Lemme de Borel-Cantelli . . . . .	41
3.2	Loi du 0-1 de Kolmogorov . . . . .	45
4	Appendice au Chapitre 2 : démonstrations techniques . . . . .	47
4.1	Preuve du Lemme 2.9 . . . . .	47
4.2	Critère d'indépendance via les fonctions caractéristiques . . . . .	47
<b>3</b>	<b>Loi des grands nombres</b>	<b>49</b>
1	Différentes notions de convergence . . . . .	49
1.1	Convergence p.s. et convergence en probabilité . . . . .	49
1.2	Convergence $L^p$ . . . . .	52

2	Loi forte des grands nombres . . . . .	53
2.1	Le résultat . . . . .	54
2.2	Quelques applications . . . . .	56
3	Appendice au Chapitre 3 : démonstrations techniques . . . . .	58
<b>4</b>	<b>Convergence en loi et théorème central limite</b>	<b>61</b>
1	Convergence en loi . . . . .	61
1.1	Définition et premiers exemples . . . . .	61
1.2	Deux cas particuliers . . . . .	63
1.3	Lien avec les autres modes de convergence . . . . .	65
2	Caractérisation de la convergence en loi . . . . .	67
2.1	Restriction des fonctions tests . . . . .	67
2.2	Caractérisation en termes de fonction de répartition et de fonction caractéristique . . . . .	68
3	Théorème central limite . . . . .	70
3.1	Énoncé . . . . .	70
3.2	Application à la statistique . . . . .	72
4	Appendice au Chapitre 4 : démonstrations techniques . . . . .	75
4.1	Preuve du théorème de Lévy (Théorème 4.15) . . . . .	75
<b>A</b>	<b>Rappels : théorie de la mesure et de l'intégration</b>	<b>77</b>
1	Résultats généraux de théorie de la mesure . . . . .	77
1.1	Tribu . . . . .	77
1.2	Mesure . . . . .	78
1.3	Lemme de classe monotone . . . . .	80
2	Transformée de Fourier . . . . .	81
2.1	Définitions et propriétés . . . . .	81
2.2	Le cas de la densité gaussienne . . . . .	83
2.3	Inversion de Fourier . . . . .	84
2.4	Transformée de Fourier d'une mesure . . . . .	85

# Chapitre 1

## Bases de la théorie des probabilités

L'objectif de ce chapitre est de construire un cadre rigoureux pour calculer des probabilités. Des cas particuliers d'expériences aléatoires ont été vus en L2, essentiellement pour des univers finis ou infini dénombrables. Cependant, certains phénomènes aléatoires naturels ne peuvent pas être modélisés par un espace dénombrable. C'est par exemple en s'intéressant au phénomène de turbulence que le scientifique Kolmogorov a proposé, au milieu du XXe siècle, d'étendre les outils probabilistes de base à un univers quelconque.

### 1 Espace probabilisé

#### 1.1 Probabilité

On considère une expérience aléatoire dont on note  $\Omega$  l'ensemble des résultats possibles. On appelle cet ensemble l'*univers*. On souhaite associer à une partie  $A \subset \Omega$  un réel entre 0 et 1 qui donne une idée de la plausibilité qu'un résultat issu de l'expérience tombe dans  $A$ .

**Définition 1.1** (Probabilité). *Soit  $(\Omega, \mathcal{F})$  un espace mesurable. Une mesure sur  $(\Omega, \mathcal{F})$  est une application  $\mu : \mathcal{F} \rightarrow [0, \infty]$  qui vérifie :*

1.  $\mu(\emptyset) = 0$  ;
2. pour toute suite  $(A_n)_{n \in \mathbb{N}}$  d'éléments 2 à 2 disjoints de  $\mathcal{F}$ ,

$$\mu\left(\bigcup_{n=0}^{+\infty} A_n\right) = \sum_{n=0}^{+\infty} \mu(A_n).$$

On dit alors que  $(\Omega, \mathcal{F}, \mu)$  est un espace mesuré.

Si de plus la mesure  $\mu$  vérifie  $\mu(\Omega) = 1$ , alors on dit alors que  $\mu$  est une mesure de probabilité ou loi de probabilité (ou plus simplement une probabilité). On parle alors d'espace probabilisé. Le plus souvent on notera  $\mathbb{P}$  la mesure de probabilité :  $(\Omega, \mathcal{F}, \mathbb{P})$  est un espace probabilisé.

Remarquons que puisque  $\mathbb{P}(\Omega) = 1$ , une mesure de probabilité est à valeurs dans  $[0, 1]$ . Un événement  $A \in \mathcal{A}$  est dit *presque sûr* (p.s. en abrégé) si  $\mathbb{P}(A) = 1$ .

Une probabilité n'étant rien d'autre qu'une mesure de masse totale 1, le cours de théorie de la mesure regorge d'exemples de probabilités. En voici quelques unes usuelles.

### Exemples 1.2.

1. Soit  $(\Omega, \mathcal{F})$  un espace mesurable et  $\omega$  un élément fixé dans  $\Omega$ . La mesure (ou masse) de Dirac en  $\omega$  est la mesure définie pour tout  $A \in \mathcal{F}$  par

$$\delta_\omega(A) = \mathbb{1}_A(\omega).$$

On vérifie facilement que c'est bien une probabilité.

2. Sur le segment  $[0, 1]$  muni de sa tribu des boréliens la mesure de Lebesgue est une probabilité.
3. Si  $(\Omega, \mathcal{F}, \mu)$  est un espace mesuré avec  $0 < \mu(\Omega) < \infty$ , alors on obtient une probabilité en considérant la mesure  $\mathbb{P} = \frac{\mu(\cdot)}{\mu(\Omega)}$ .

**Interprétation** Un espace probabilisé est donc un cas particulier d'espace mesuré pour lequel la masse totale de la mesure est égale à 1. En fait, le point de vue diffère de la théorie de l'intégration : dans le cadre de la théorie des probabilités, on cherche à fournir un modèle mathématique pour une "expérience aléatoire".

- L'ensemble  $\Omega$  est appelé l'*univers* : il représente l'ensemble de toutes les éventualités possibles, toutes les déterminations du hasard dans l'expérience considérée. Les éléments  $\omega$  de  $\Omega$ , parfois appelés *événements élémentaires*, correspondent donc aux issues possibles de l'expérience aléatoire.
- La tribu  $\mathcal{F}$  correspond à l'ensemble des *événements* : ce sont les parties de  $\Omega$  dont on peut évaluer la probabilité. Il faut voir un événement  $A$  de  $\mathcal{F}$  comme un sous-ensemble de  $\Omega$  contenant toutes les éventualités  $\omega$  pour lesquelles une certaine propriété est vérifiée.
- On associe à chaque événement  $A \in \mathcal{F}$  un réel  $\mathbb{P}(A) \in [0, 1]$  qui donne la plausibilité que le résultat de l'expérience soit dans  $A$ .

**Remarque 1.3.** Dans les premiers traités de théorie des probabilités, longtemps avant l'introduction de la théorie de la mesure, la probabilité  $\mathbb{P}(A)$  était définie de la manière suivante : on imagine qu'on répète l'expérience aléatoire un nombre  $N$  de fois, et on note  $N_A$  le nombre de répétitions pour lesquelles l'événement  $A$  est réalisé ; alors, la proportion  $N_A/N$  converge quand  $N \rightarrow \infty$  vers la probabilité  $\mathbb{P}(A)$ . Nous verrons plus loin le lien entre cette définition "historique" et l'approche moderne.

## 1.2 Exemples d'espaces probabilisés

### Quelques exemples classiques

1. On considère l'univers  $\Omega = \{1, \dots, 6\}$  muni de la tribu pleine  $\mathcal{F} = \mathcal{P}(\Omega)$ . La mesure

$$\mathbb{P} = \frac{1}{6} \sum_{k=1}^6 \delta_k$$

est alors une mesure de probabilité sur l'espace mesurable  $(\Omega, \mathcal{F})$  : c'est la probabilité (ou loi) uniforme sur  $\{1, \dots, 6\}$ . Elle modélise le lancer d'un dé équilibré.

Si  $A \in \mathcal{F}$ , alors  $\mathbb{P}(A)$  correspond à la probabilité que le dé donne un chiffre dans  $A$  :

$$\mathbb{P}(A) = \frac{1}{6} \sum_{k=1}^6 \delta_k(A) = \frac{\text{Card}(A)}{6}.$$

Par exemple,  $\mathbb{P}(\{1, 3, 5\}) = 1/2$ .

2. Si on répète cette expérience  $n$  fois, alors un espace probabilisé naturel est

$$\Omega = \{1, \dots, 6\}^n, \quad \mathcal{F} = \mathcal{P}(\Omega), \quad \mathbb{P} = \frac{1}{6^n} \sum_{\omega \in \Omega} \delta_\omega.$$

Plus simplement, on a  $\mathbb{P}(\{\omega_1, \dots, \omega_n\}) = 1/6^n$ , pour  $\omega_1, \dots, \omega_n \in \{1, \dots, 6\}$ .

3. Pour modéliser le résultat d'un pile ou face biaisé, on peut considérer  $\Omega = \{0, 1\}$  muni de la tribu pleine  $\mathcal{F} = \mathcal{P}(\{0, 1\})$  et de la probabilité

$$\mathbb{P} = (1 - p)\delta_0 + p\delta_1.$$

où  $p \in [0, 1]$  correspond le biais de la pièce. On a alors par exemple

$$\mathbb{P}(\{0\}) = (1 - p)\delta_0(\{0\}) + p\delta_1(\{0\}) = 1 - p.$$

**Un exemple non dénombrable** On considère une infinité de lancers de dé. Le choix de l'espace probabilisé est déjà moins évident. On prend pour univers  $\Omega$  l'ensemble des suites à valeurs dans  $\{1, 2, 3, 4, 5, 6\}$  :

$$\Omega = \{1, \dots, 6\}^{\mathbb{N}^*} = \{(\omega_n)_{n \geq 1} : \omega_n \in \{1, \dots, 6\}\},$$

de sorte qu'un élément  $(\omega_n)$  est une suite qui donne les résultats successifs des différents lancers. La tribu de toutes les parties de  $\Omega$  est ici trop grande, on se restreint à une sous-tribu  $\mathcal{F}$  que l'on définit de la manière suivante. Pour  $k \geq 1$  et  $x_1, \dots, x_k$  des entiers dans  $\{1, \dots, 6\}$ , on pose

$$A_{x_1, \dots, x_k} = \{\omega = (\omega_n)_{n \geq 1} \in \Omega : \omega_1 = x_1, \dots, \omega_k = x_k\}.$$

On définit alors la tribu  $\mathcal{F}$  comme la plus petite tribu rendant mesurables les ensembles  $A_{x_1, \dots, x_k}$  pour tout choix de  $x_1, \dots, x_k$  et pour tout  $k \geq 1$ . Enfin, on souhaite munir l'espace  $(\Omega, \mathcal{F})$  d'une probabilité  $\mathbb{P}$  vérifiant

$$\mathbb{P}(A_{x_1, \dots, x_k}) = \mathbb{P}(\{\omega = (\omega_n)_{n \geq 1} \in \Omega : \omega_1 = x_1, \dots, \omega_k = x_k\}) = \left(\frac{1}{6}\right)^k,$$

pour tout  $x_1, \dots, x_k$  et tout  $k \geq 1$ . L'unicité de cette probabilité  $\mathbb{P}$  résulte du lemme de classe monotone. Son existence est délicate à prouver, elle peut se construire à partir de la mesure de Lebesgue.

**Des espaces probabilisés généraux** Soit  $(\Omega, \mathcal{F}, \mu)$  un espace mesuré, avec  $\mu$  une mesure pas nécessairement de probabilité, ni même finie. Soit  $f : \Omega \rightarrow [0, \infty[$  une fonction mesurable positive telle que  $\int_{\Omega} f(x) d\mu(x) = 1$ . Alors, l'application  $\mathbb{P} : \mathcal{F} \rightarrow [0, 1]$ , définie par

$$\mathbb{P}(A) = \int_A f(x) d\mu(x) = \int_{\Omega} f(x) \mathbf{1}_A(x) d\mu(x),$$

est une mesure de probabilité sur  $(\Omega, \mathcal{F})$ . On obtient donc un espace probabilisé  $(\Omega, \mathcal{F}, \mathbb{P})$ . La fonction  $f$  est alors appelée *densité* de  $\mathbb{P}$  par rapport à  $\mu$ .

On utilisera souvent l'abus de notation suivant :

$$d\mathbb{P}(x) = f(x) d\mu(x), \quad \text{ou plus simplement} \quad d\mathbb{P} = f d\mu, \quad \text{ou même} \quad \mathbb{P} = f\mu.$$

On se rappellera que pour évaluer une telle probabilité sur un ensemble  $A \in \mathcal{F}$  il faut intégrer la fonction  $f$  par rapport à la mesure  $\mu$  sur  $A$ .

#### Exemples 1.4.

1. On retrouve l'exemple précédent avec  $(\Omega, \mathcal{F}, \mu) = (\mathbb{N}, \mathcal{P}(\mathbb{N}), \text{Compt})$ , où  $\text{Compt}$  désigne la mesure de comptage sur  $\mathbb{N}$ , et  $f(n) = p_n$ . En effet, dans ce cas, on a l'égalité, pour  $A \subset \mathbb{N}$ ,

$$\int_A f(x) d\mu(x) = \int_A f(n) d\text{Compt}(n) = \sum_{n \in A} p_n.$$

Par exemple, avec  $p_n = e^{-\theta} \frac{\theta^n}{n!}$ ,  $\theta > 0$ , on obtient la probabilité

$$\mathbb{P} = \sum_{n=0}^{\infty} e^{-\theta} \frac{\theta^n}{n!} \delta_n,$$

qu'on appelle mesure (ou loi) de Poisson.

2. Avec  $(\Omega, \mathcal{F}, \mu) = (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \lambda_d)$  où  $\lambda_d$  désigne la mesure de Lebesgue sur  $\mathbb{R}^d$ , on obtient

$$\mathbb{P}(A) = \int_{\mathbb{R}^d} f(x) \mathbf{1}_A(x) d\lambda_d(x), \quad \text{pour tout } A \in \mathcal{B}(\mathbb{R}^d).$$

Par exemple, dans le cas  $d = 1$  avec  $f(x) = e^{-x} \mathbf{1}_{\mathbb{R}_+}(x)$ , on obtient la mesure de probabilité  $\mathbb{P}$  définie par

$$\mathbb{P}(A) = \int_{\mathbb{R}} e^{-x} \mathbf{1}_{\mathbb{R}_+}(x) \mathbf{1}_A(x) d\lambda_1(x), \quad \text{pour tout } A \in \mathcal{B}(\mathbb{R}),$$

qu'on appelle mesure (ou loi) exponentielle.

Dans toute la suite, on utilisera exclusivement la mesure de Lebesgue sur  $\mathbb{R}^d$ , donc on abrègera souvent  $d\lambda_d(x)$  en  $dx$  s'il n'y a pas d'ambiguïté sur la dimension  $d$ .



**Un dernier exemple** On s'intéresse au déplacement d'une bouteille à la surface de l'océan pendant un intervalle de temps de longueur 1 (avec une unité de temps quelconque fixée). L'univers naturel est alors  $\Omega = \mathcal{C}([0, 1], \mathbb{R}^2)$ , ensemble des fonctions continues de  $[0, 1]$  dans  $\mathbb{R}^2$ . Un élément de  $\Omega$  est donc une fonction continue de  $[0, 1]$  dans  $\mathbb{R}^2$ . Là encore, il n'est pas évident de proposer une tribu et une probabilité qui modélisent cette expérience aléatoire. Si on s'intéresse aux positions de la bouteille à un instant  $t$  donné, on souhaite que les applications coordonnées  $f_t : \omega \mapsto \omega(t)$  pour  $t \in [0, 1]$  soient mesurables. On considère donc la plus petite tribu qui rende mesurables ces applications. Il reste alors à proposer une probabilité  $\mathbb{P}$  qui modélise correctement le caractère erratique de la bouteille. Un choix classique est de prendre pour  $\mathbb{P}$  la mesure de Wiener qui est la loi du mouvement brownien (vu bien plus tard, sûrement en M2).

### 1.3 Quelques propriétés d'une probabilité

Une probabilité étant une mesure, elle vérifie toutes les propriétés des mesures étudiées en cours d'intégration. On liste ci-dessous les principales.

**Proposition 1.5.** *Soit  $(\Omega, \mathcal{F}, \mathbb{P})$  un espace probabilisé.*

1. (Monotonie). *Si  $A, B \in \mathcal{F}$  sont tels que  $A \subset B$ , alors*

$$\mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(B \setminus A),$$

*où  $B \setminus A = \{\omega \in B : \omega \notin A\}$  désigne la soustraction ensembliste. On en déduit en particulier que  $\mathbb{P}(A) \leq \mathbb{P}(B)$  et, avec  $B = \Omega$ , que  $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$ . La propriété de monotonie est illustrée en Figure 1.1.*

2. (Additivité forte). *Si  $A, B \in \mathcal{F}$ , alors*

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

3. (Sous- $\sigma$ -additivité). *Si  $(A_n)_{n \in \mathbb{N}}$  est une suite d'événements, alors*

$$\mathbb{P}\left(\bigcup_{n=0}^{+\infty} A_n\right) \leq \sum_{n=0}^{+\infty} \mathbb{P}(A_n).$$

4. (Croissance monotone). *Si  $(A_n)_{n \in \mathbb{N}}$  est une suite d'événements croissante pour l'inclusion ( $A_n \subset A_{n+1}$ ), alors*

$$\mathbb{P}\left(\bigcup_{n=0}^{+\infty} A_n\right) = \lim_{n \rightarrow +\infty} \mathbb{P}(A_n).$$

*Une illustration de cette propriété est donnée Figure 1.2.*

5. (Décroissance monotone). *Si  $(A_n)_{n \in \mathbb{N}}$  est une suite d'événements décroissante pour l'inclusion ( $A_{n+1} \subset A_n$ ), alors*

$$\mathbb{P}\left(\bigcap_{n=0}^{+\infty} A_n\right) = \lim_{n \rightarrow +\infty} \mathbb{P}(A_n).$$

*Une illustration de cette propriété est donnée Figure 1.3.*

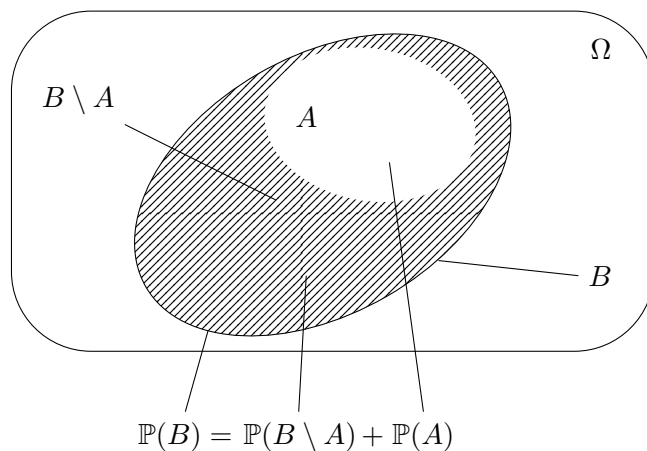
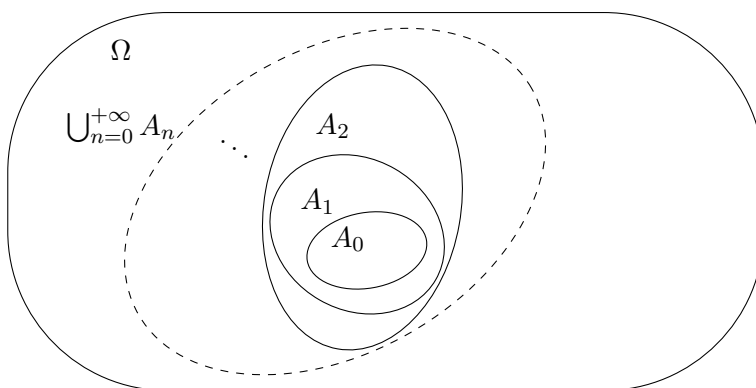
FIGURE 1.1 – Illustration de la monotonie :  $\mathbb{P}(A) \leq \mathbb{P}(B)$ .

FIGURE 1.2 – Une suite d'ensembles croissante pour l'inclusion.

**Exemple 1.6.** Reprenons le deuxième exemple de la section 1.2 sur l'infinité de lancers de dé. On considère les événements

$$B_n = \{\text{on obtient un 6 lors des } n \text{ premiers lancers}\} \quad \text{et} \quad B = \{\text{on obtient un 6}\}.$$

Les ensembles  $B_n$  forment une suite croissante dont l'union est  $B$ . La propriété de croissance monotone implique alors que

$$\mathbb{P}(B) = \mathbb{P}\left(\bigcup_{n \geq 1} B_n\right) = \lim_{n \rightarrow \infty} \mathbb{P}(B_n)$$

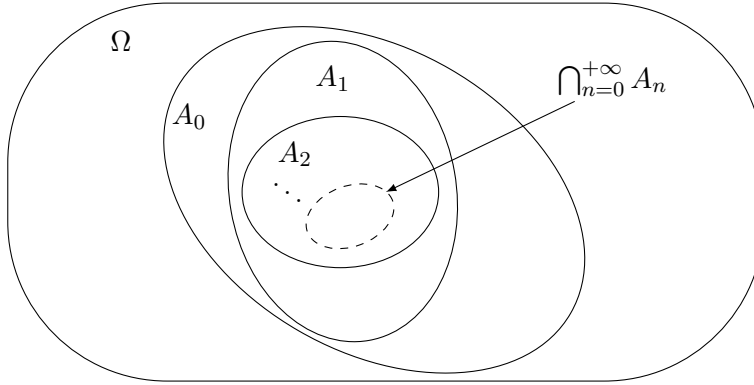


FIGURE 1.3 – Une suite d'ensembles décroissante pour l'inclusion.

On peut réécrire  $B_n$  en fonction des ensembles  $A_{x_1, \dots, x_k}$  définis dans la section 1.2 via

$$B_n^c = \bigcup_{x_1=1}^5 \cdots \bigcup_{x_n=1}^5 A_{x_1, \dots, x_n},$$

ce qui prouve en particulier que  $B_n$  appartient à la tribu  $\mathcal{F}$  (et donc  $B$  aussi). Par ailleurs, pour tout  $k = 1, \dots, n$ , l'union  $\bigcup_{x_k=1}^5 A_{x_1, \dots, x_k, \dots, x_n}$  est formée de cinq événements deux à deux disjoints. On obtient alors

$$\begin{aligned} \mathbb{P}(B_n) &= 1 - \mathbb{P}(B_n^c) = 1 - \mathbb{P}\left(\bigcup_{x_1=1}^5 \cdots \bigcup_{x_n=1}^5 A_{x_1, \dots, x_n}\right) \\ &= 1 - \sum_{x_1=1}^5 \cdots \sum_{x_n=1}^5 \mathbb{P}(A_{x_1, \dots, x_n}) = 1 - \sum_{x_1=1}^5 \cdots \sum_{x_n=1}^5 \left(\frac{1}{6}\right)^n, \end{aligned}$$

par définition de la probabilité  $\mathbb{P}$ . On en déduit alors que

$$\mathbb{P}(B_n) = 1 - \left(\frac{5}{6}\right)^n,$$

et donc  $\mathbb{P}(B) = \lim_{n \rightarrow \infty} \mathbb{P}(B_n) = 1$ . L'événement  $B$  est presque sûr. Remarquons cependant que  $B$  ne coïncide pas avec  $\Omega$  : il manque toute les suites ne comportant pas de 6.

## 2 Variable aléatoire

### 2.1 Loi d'une variable aléatoire

**Définition 1.7** (Variable aléatoire). Soit  $(\Omega, \mathcal{F}, \mathbb{P})$  un espace probabilisé et  $(E, \mathcal{E})$  un espace mesurable. Une variable aléatoire est une application mesurable  $X : \Omega \rightarrow E$ , c'est-à-dire

$$\forall B \in \mathcal{E}, \quad X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\} \in \mathcal{F}.$$

Si l'espace d'arrivée est  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$  on parle de variable aléatoire réelle, si c'est  $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$  on parle de vecteur aléatoire. Ce sera le cas de la majorité des exemples de ce cours.

On remarque en particulier que la probabilité  $\mathbb{P}$  n'intervient pas dans la définition d'une variable aléatoire. Elle interviendra plus tard quand on parlera de loi.

**Exemples 1.8.** Reprenons quelques exemples développés précédemment en section 1.2.

1. Lors du lancer de deux dés, on s'intéresse à la somme des résultats. Cela incite à regarder l'application

$$\begin{aligned} X : \Omega &\rightarrow \{2, 3, \dots, 12\} \\ (i, j) &\mapsto i + j. \end{aligned}$$

Cette application est mesurable puisque l'ensemble de départ est muni de la tribu pleine.

2. Lors du lancer d'une infinité de dés, on s'intéresse au nombre de lancers nécessaires pour obtenir le premier 6. On définit alors

$$\begin{aligned} Y : \Omega &\rightarrow \mathbb{N}^* \cup \{\infty\} \\ \omega = (\omega_n)_{n \geq 1} &\mapsto \inf\{n \geq 1 : \omega_n = 6\}, \end{aligned}$$

avec la convention  $\inf \emptyset = \infty$ . En munissant l'ensemble d'arrivée de la tribu pleine, on obtient, pour tout  $k \geq 1$ ,

$$Y^{-1}(\{k\}) = \{\omega \in \Omega : \omega_1 \neq 6, \dots, \omega_{k-1} \neq 6, \omega_k = 6\} = \bigcup_{x_1, \dots, x_{k-1} \in \{1, \dots, 5\}} A_{x_1, \dots, x_{k-1}, 6},$$

et

$$Y^{-1}(\{\infty\}) = \{\omega \in \Omega : \forall k \geq 1, \omega_k \in \{1, \dots, 5\}\} = \bigcap_{k=1}^{\infty} \bigcup_{x_1, \dots, x_k \in \{1, \dots, 5\}} A_{x_1, \dots, x_k},$$

qui sont dans la tribu  $\mathcal{F}$ . Ainsi,  $Y$  est une application mesurable donc une variable aléatoire.

3. On s'intéresse à la position de la bouteille à l'instant  $t = 1$ . On définit l'application  $Z$  par

$$\begin{aligned} Z : \Omega &\rightarrow \mathbb{R}^2 \\ \omega = (\omega(t))_{t \in [0,1]} &\mapsto \omega(1). \end{aligned}$$

Cette application correspond en fait à  $f_1$  définie dans l'exemple 3 de la section 1.2. Elle est donc mesurable, ce qui prouve que  $Z$  est une variable aléatoire.

Puisqu'une variable aléatoire n'est rien d'autre qu'une application mesurable, la notion de variable aléatoire est stable par toutes les opérations algébriques envisageables.

**Définition 1.9** (Loi d'une variable aléatoire). *La loi de la variable aléatoire  $X$  est la mesure image de  $\mathbb{P}$  par  $X$ . C'est donc une mesure de probabilité sur  $(E, \mathcal{E})$ , notée  $\mathbb{P}_X$ , définie par*

$$\forall A \in \mathcal{E}, \quad \mathbb{P}_X(A) = \mathbb{P}(X^{-1}(A)).$$

En pratique, on écrit plutôt

$$\mathbb{P}_X(A) = \mathbb{P}(X \in A) \quad (= \mathbb{P}(\{\omega \in \Omega : X(\omega) \in A\})),$$

ce qui permet d'oublier en quelque sorte le rôle de  $\Omega$ .

La loi  $\mathbb{P}_X$  permet de calculer la probabilité d'événements qui "dépendent" de la variable aléatoire  $X$ . À chaque "point aléatoire"  $\omega \in \Omega$  on associe un "point aléatoire"  $X(\omega) \in E$ , et  $\mathbb{P}_X(A)$  correspond à la probabilité que ce point aléatoire  $X(\omega)$  tombe dans  $A$ . Une illustration de ce procédé est donnée en Figure 1.4.

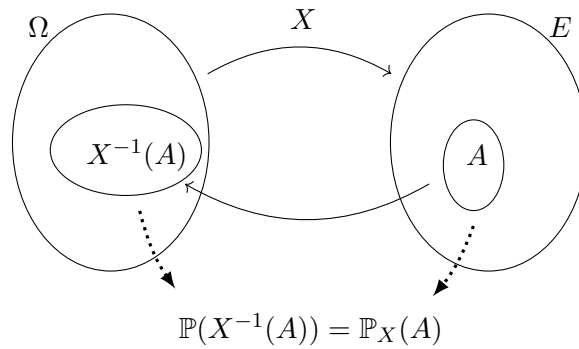


FIGURE 1.4 – Une variable aléatoire  $X$  est une fonction d'un espace probabilisé  $(\Omega, \mathcal{F}, \mathbb{P})$  vers un espace mesurable  $(E, \mathcal{E})$ . La loi d'une variable aléatoire est la mesure image de  $X$  par  $\mathbb{P}$  : c'est donc une mesure de probabilité sur  $(E, \mathcal{E})$ .

**Remarque 1.10.** En pratique, l'espace de probabilité  $(\Omega, \mathcal{F}, \mathbb{P})$  est un peu mystérieux. Si on se donne un espace d'arrivée  $(E, \mathcal{E})$  muni d'une loi  $\mu$ , on peut toujours écrire cette loi comme une mesure image par une application mesurable  $X$  : il suffit de prendre  $\Omega = E$ ,  $\mathcal{F} = \mathcal{E}$ ,  $\mathbb{P} = \mu$  et  $X = \text{id}_E$ . Ainsi, toute mesure de probabilité est la loi d'une variable aléatoire. Dans les exemples, on se concentrera quasi exclusivement sur la loi d'une variable aléatoire et non pas sur l'espace probabilisé de départ, celui-ci n'étant qu'une représentation théorique du hasard. En particulier, on confondra parfois de manière abusive une variable aléatoire et sa loi. Il faut garder en tête que la première est une application tandis que la seconde est une mesure de probabilité.

**Exemple 1.11.** Reprenons l'exemple de la succession de lancers de dés. La variable aléatoire  $Y = \inf\{n \geq 1 : \omega_n = 6\}$  est à valeurs dans  $\mathbb{N}^* \cup \{\infty\}$  donc sa loi est une mesure de probabilité sur  $\mathbb{N}^* \cup \{\infty\}$ . Elle est définie par

$$\mathbb{P}_Y(\{k\}) = \mathbb{P}(Y^{-1}(\{k\})) = \mathbb{P}(\omega = (\omega_n)_{n \geq 1} : \omega_1 \neq 6, \dots, \omega_{k-1} \neq 6, \omega_k = 6) = \left(\frac{5}{6}\right)^{k-1} \times \frac{1}{6}.$$

Par ailleurs, on a vu dans l'Exemple 1.6 que  $\mathbb{P}_Y(\{\infty\}) = 0$  ce qui permet de restreindre la loi de

$Y$  à  $\mathbb{N}^*$ . Ainsi, la loi de  $Y$  est la mesure

$$\mathbb{P}_Y = \sum_{k=1}^{\infty} \left(\frac{5}{6}\right)^{k-1} \times \frac{1}{6} \delta_k.$$

On verra plus loin que c'est la loi géométrique de paramètre  $1/6$ .

**Variable aléatoire discrète** On parle de variable aléatoire discrète si  $X$  est à valeurs dans un ensemble  $E$  dénombrable. On choisit alors pour  $\mathcal{E}$  l'ensemble des parties de  $E$ . Pour  $A \in \mathcal{E}$ , on a alors

$$\mathbb{P}_X(A) = \mathbb{P}(X \in A) = \mathbb{P}\left(X \in \bigcup_{x \in A} \{x\}\right) = \sum_{x \in A} \mathbb{P}(X = x).$$

La loi de  $X$  est donc entièrement déterminée par les quantités  $p_x = \mathbb{P}(X = x)$ . La probabilité  $\mathbb{P}_X$  s'écrit alors

$$\mathbb{P}_X = \sum_{x \in E} p_x \delta_x.$$

**Variable aléatoire à densité** Une variable aléatoire  $X$  à valeurs dans  $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$  est dite à densité s'il existe une fonction mesurable  $f_X : \mathbb{R}^d \rightarrow [0, \infty[$  d'intégrale 1 par rapport à la mesure de Lebesgue  $\lambda_d$  telle que

$$d\mathbb{P}_X(x) = f_X(x) d\lambda_d(x) = f_X(x) dx.$$

On rappelle que cela signifie que pour tout  $A \in \mathcal{B}(\mathbb{R}^d)$ ,

$$\mathbb{P}_X(A) = \int_A f_X(x) dx.$$

La fonction  $f_X$  est appelée la *densité de (la loi de)  $X$* , l'unicité de cette fonction n'étant valable que  $\lambda_d$ -presque partout. Si  $d = 1$ , on a en particulier pour tous réels  $a \leq b$ ,

$$\mathbb{P}_X([a, b]) = \int_a^b f_X(x) dx.$$

Il faut cependant garder à l'esprit que certaines variables aléatoires ne sont ni discrète ni à densité.

## 2.2 Lois usuelles

On liste ci-dessous les lois les plus utilisées en probabilité. Des illustrations graphiques de ces lois sont proposées en fin de chapitre.

### 2.2.1 Lois discrètes

**Loi uniforme sur un ensemble fini** Soit  $E = \{x_1, \dots, x_n\}$  est un ensemble fini. On dit que  $X : \Omega \rightarrow E$  suit une *loi uniforme sur  $E$* , et on note  $X \sim \mathcal{U}(E)$ , si

$$\mathbb{P}(X = x) = \frac{1}{\text{Card}(E)} = \frac{1}{n}, \quad x \in E.$$

Autrement dit,

$$\mathbb{P}_X = \frac{1}{n} \sum_{k=1}^n \delta_{x_k}.$$

En particulier, pour tout  $A \subset E$ ,

$$\mathbb{P}(A) = \frac{\text{Card}(A)}{\text{Card}(E)}.$$

**Exercice 1.12.** Montrer que si  $E$  est infini dénombrable, alors il n'existe pas de loi uniforme sur  $E$ , c'est-à-dire telle que  $\mathbb{P}(X = x) = p$ , pour tout  $x \in E$ .

**Loi de Bernoulli** Soit  $p \in [0, 1]$ . On dit que  $X : \Omega \rightarrow \{0, 1\}$  suit une loi de Bernoulli de paramètre  $p$ , et on note  $X \sim \mathcal{B}(p)$ , si

$$\mathbb{P}(X = 1) = p = 1 - \mathbb{P}(X = 0).$$

Autrement dit,

$$\mathbb{P}_X = (1 - p)\delta_0 + p\delta_1.$$

Pour  $p = 1/2$  on retrouve la loi uniforme sur  $\{0, 1\}$ . La loi de Bernoulli modélise une expérience aléatoire binaire avec deux issues possibles, succès ( $= 1$ ) ou échec ( $= 0$ ). Une telle expérience est parfois appelée expérience de Bernoulli. Le paramètre  $p$  représente donc la probabilité de réussite de l'expérience.

**Loi binomiale** Soit  $p \in [0, 1]$  et  $n \in \mathbb{N}^*$ . On dit que  $X : \Omega \rightarrow \{0, 1, \dots, n\}$  suit une loi binomiale de paramètres  $(n, p)$ , et on note  $X \sim \mathcal{B}(n, p)$ , si

$$\mathbb{P}(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}, \quad k = 0, \dots, n.$$

Autrement dit,

$$\mathbb{P}_X = \sum_{k=0}^n \binom{n}{k} p^k (1 - p)^{n-k} \delta_k.$$

Cela correspond à la répétition de  $n$  expériences de Bernoulli dont la probabilité du succès est  $p$  (par exemple le lancer d'une pièce biaisée avec probabilité  $p$  d'obtenir pile).

**Loi géométrique** Soit  $p \in ]0, 1]$ . On dit que  $X : \Omega \rightarrow \mathbb{N}^*$  suit une loi géométrique de paramètre  $p$ , et on note  $X \sim \mathcal{G}(p)$ , si

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

Autrement dit,

$$\mathbb{P}_X = \sum_{k=1}^{\infty} (1 - p)^{k-1} p \delta_k.$$

Cette loi modélise le premier succès lors d'une succession d'expériences de Bernoulli de paramètre  $p$ . Dans la probabilité  $\mathbb{P}(X = k)$ , le facteur  $(1 - p)^{k-1}$  correspond aux  $k - 1$  échecs et le facteur  $p$  au succès lors de la  $k$ -ème expérience.

**Loi de Poisson** Soit  $\theta > 0$ . On dit que  $X : \Omega \rightarrow \mathbb{N}$  suit une loi de Poisson de paramètre  $\theta$ , et on note  $X \sim \mathcal{P}(\theta)$ , si

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

Autrement dit,

$$\mathbb{P}_X = \sum_{k=0}^{\infty} e^{-\theta} \frac{\theta^k}{k!} \delta_k.$$

On montrera dans un chapitre ultérieur qu'une loi de Poisson peut être vue comme une loi de Bernoulli de paramètres  $(n, \theta/n)$  pour  $n$  très grand. Cela signifie qu'une loi de Poisson correspond au nombre d'occurrences d'un phénomène rare (de probabilité inversement proportionnelle au nombre d'expériences réalisées). En effet, on peut montrer que pour tout entier  $k$ ,

$$\binom{n}{k} \left(\frac{\theta}{n}\right)^k \left(1 - \frac{\theta}{n}\right)^{n-k} \xrightarrow{n \rightarrow \infty} e^{-\theta} \frac{\theta^k}{k!}.$$

On appelle la loi de Poisson la *loi des événements rares*.

### 2.2.2 Loi à densité

**Loi uniforme sur un sous-ensemble mesurable de  $\mathbb{R}^d$**  Soit  $A \in \mathcal{B}(\mathbb{R}^d)$  un borélien de mesure de Lebesgue finie non nulle :  $0 < \lambda_d(A) < \infty$ . On dit que la variable aléatoire  $X : \Omega \rightarrow \mathbb{R}^d$  suit une loi uniforme sur  $A$ , et on note  $X \sim \mathcal{U}(A)$ , si

$$d\mathbb{P}_X = \frac{\mathbb{1}_A(x)}{\lambda_d(A)} dx.$$

Dans le cas  $d = 1$  et  $A = [a, b]$ , avec  $a < b$ , la loi uniforme correspond à  $(b - a)^{-1} \mathbb{1}_{[a, b]}(x) dx$ .



**Loi exponentielle** Soit  $\theta > 0$ . La variable aléatoire  $X : \Omega \rightarrow \mathbb{R}$  suit une loi exponentielle de paramètre  $\theta$ , et on note  $X \sim \mathcal{E}(\theta)$ , si

$$d\mathbb{P}_X = \theta e^{-\theta x} \mathbf{1}_{\mathbb{R}_+}(x) dx.$$

La loi exponentielle possède une propriété fondamentale qui est l'absence de mémoire ou de non-vieillessement : pour tout  $x, y \geq 0$ ,

$$\mathbb{P}(X > x + y \mid X > x) = \mathbb{P}(X > y).$$

C'est d'ailleurs une propriété caractéristique de la loi exponentielle (si on inclut la masse de Dirac en 0 comme une loi exponentielle dégénérée de paramètre  $+\infty$ ).

**Loi normale** Soit  $\mu \in \mathbb{R}$  et  $\sigma > 0$ . Une variable aléatoire  $X : \Omega \rightarrow \mathbb{R}$  suit une loi normale (ou gaussienne) de moyenne  $\mu$  et de variance  $\sigma^2$  si sa loi est donnée par

$$d\mathbb{P}_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx.$$

On peut vérifier que la densité associée est bien d'intégrale 1. En effet, le changement de variable  $y = (x - \mu)/\sigma$  nous ramène en effet à l'intégrale de Gauss :

$$\int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy = 1.$$

## 2.3 Moments d'une variable aléatoire

On considère dans toute cette section un espace probabilisé  $(\Omega, \mathcal{F}, \mathbb{P})$ .

### 2.3.1 Espérance

**Définition 1.13** (Espérance). Soit  $X : \Omega \rightarrow \mathbb{R}$  une variable aléatoire réelle. On appelle espérance de  $X$  l'intégrale

$$\mathbb{E}[X] = \int_{\Omega} X(\omega) d\mathbb{P}(\omega),$$

qui est bien définie si  $X$  est positive ou si  $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ . On définit de même l'espérance d'une variable aléatoire intégrable à valeurs dans  $\mathbb{C}$ .

Plus généralement, si  $X : \Omega \rightarrow \mathbb{C}^d$  est à valeurs vectorielles, alors  $X$  s'écrit  $(X_1, \dots, X_d)$  où les  $X_i$  sont les coordonnées de  $X$ , qui sont des variables aléatoires. On pose alors  $\mathbb{E}[X] = (\mathbb{E}[X_1], \dots, \mathbb{E}[X_d])$ , à condition que les espérances  $\mathbb{E}[X_i]$  soient bien définies.

En particulier, si  $X = \mathbf{1}_A$ , avec  $A \in \mathcal{F}$ , alors  $\mathbb{E}[\mathbf{1}_A] = \mathbb{P}(A)$ .

**Remarques 1.14.**

1. En général,  $\mathbb{E}[X]$  s'interprète comme la moyenne de  $X$ . Si  $\Omega$  est fini et si  $\mathbb{P}$  attribue la même valeur à chaque élément de  $\Omega$ , alors  $\mathbb{E}[X]$  est bien la moyenne au sens usuel.
2. Une variable aléatoire d'espérance nulle est dite *centrée*.
3. Un cas un peu pathologique est celui des variables aléatoires positives d'espérance infinie. L'espérance est alors bien définie puisque c'est l'intégrale d'une fonction positive. Cependant on évitera de considérer ces variables aléatoires.

L'espérance étant définie via une intégrale, elle jouit des mêmes propriétés. La proposition suivante en donne deux fondamentales.

**Proposition 1.15.** *L'espérance est une application linéaire : si  $X$  et  $Y$  sont des variables aléatoires intégrables ou positives, et si  $a, b \in \mathbb{R}$ , alors*

$$\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y],$$

sous réserve que ces quantités existent.

De plus, si  $X \geq 0$ , alors  $\mathbb{E}[X] \geq 0$ , avec égalité si et seulement si  $X = 0$   $\mathbb{P}$ -presque partout.

En pratique, la définition de l'espérance n'est pas utilisable pour les calculs. La proposition ci-dessous permet de passer de l'espace  $(\Omega, \mathcal{F}, \mathbb{P})$  à l'espace  $(E, \mathcal{E}, \mathbb{P}_X)$ .

**Proposition 1.16** (Formule de transfert). *Soit  $X : \Omega \rightarrow E$  une variable aléatoire à valeurs dans un espace mesurable  $(E, \mathcal{E})$ . Si  $h : E \rightarrow [0, \infty[$  est une application mesurable positive, alors  $h(X) = h \circ X$  est une variable aléatoire et*

$$\mathbb{E}[h(X)] = \int_{\Omega} h(X(\omega)) \, d\mathbb{P}(\omega) = \int_E h(x) \, d\mathbb{P}_X(x).$$

Si maintenant  $h : E \rightarrow \mathbb{R}$  est mesurable, alors  $h(X) \in L^1(\Omega, \mathcal{F}, \mathbb{P})$  si et seulement si  $h \in L^1(E, \mathcal{E}, \mathbb{P}_X)$ , et dans ce cas cette même formule reste valable.

*Démonstration.* C'est une propriété générale des mesures images qui a été abordée en cours d'intégration. La preuve se fait par étapes : d'abord pour une fonction indicatrice  $f = \mathbf{1}_A$  pour laquelle le résultat est évident, puis on étend le résultat à toute fonction étagée par linéarité, puis à toute fonction mesurable positive avec le théorème de convergence monotone, et enfin à toute fonction intégrable en considérant la partie positive et la partie négative.  $\square$

On remarque que le dernier membre de la formule de transfert ne fait plus intervenir l'espace  $\Omega$  : la variable aléatoire  $X$  n'apparaît plus qu'à travers sa loi  $\mathbb{P}_X$ . En particulier, si  $X$  est une variable aléatoire réelle on obtient avec  $f = id_{\mathbb{R}}$  :

$$\mathbb{E}[X] = \int_E x \, d\mathbb{P}_X(x),$$

sous réserve que cette quantité existe.

**Exemple 1.17.** Si  $X$  est une variable aléatoire discrète sur  $E = (x_k)_{k \geq 0}$ , avec  $x_k \in \mathbb{R}$ , alors  $\mathbb{P}_X = \sum_{k=0}^{\infty} p_k \delta_{x_k}$ , donc

$$\mathbb{E}[X] = \int_{\mathbb{R}} x \left( d \sum_{k=0}^{\infty} p_k \delta_{x_k} \right) = \sum_{k=0}^{\infty} p_k \underbrace{\int_{\mathbb{R}} x d\delta_{x_k}}_{=x_k} = \sum_{k=0}^{\infty} x_k p_k.$$

Par exemple, si  $X$  suit une loi de Bernoulli de paramètre  $p$ , alors

$$\mathbb{E}[X] = (1-p) \times 0 + p \times 1 = p.$$

**Proposition 1.18** (Caractérisation de la loi à l'aide de l'espérance). *Soit  $X : \Omega \rightarrow E$  une variable aléatoire. La loi de  $X$  est caractérisée par les quantités  $\mathbb{E}[h(X)]$ , où  $h : E \rightarrow \mathbb{R}$  décrit l'ensemble des fonctions mesurables bornées (ou mesurables positives).*

*Autrement dit, si  $X$  et  $X'$  sont deux variables aléatoires telles que pour toute fonction  $h : E \rightarrow \mathbb{R}$  mesurable bornée (ou mesurable positive)  $\mathbb{E}[h(X)] = \mathbb{E}[h(X')]$ , alors  $X$  et  $X'$  ont la même loi.*

*Démonstration.* Si  $X$  et  $X'$  ont la même loi, c'est-à-dire vérifient  $\mathbb{P}_X = \mathbb{P}_{X'}$ , alors pour toute fonction  $h : E \rightarrow \mathbb{R}$  mesurable bornée (ou mesurable positive), on a

$$\mathbb{E}[h(X)] = \int_E h(x) d\mathbb{P}_X(x) = \int_E h(x) d\mathbb{P}_{X'}(x) = \mathbb{E}[h(X')].$$

Réciproquement, si  $X$  et  $X'$  vérifie cette égalité, alors en prenant  $h = \mathbb{1}_A$  pour  $A \in \mathcal{E}$ , on obtient

$$\mathbb{P}_X(A) = \mathbb{E}[h(X)] = \mathbb{E}[h(X')] = \mathbb{P}_{X'}(A),$$

et ceci étant vrai pour tout  $A \in \mathcal{E}$  on en déduit que  $\mathbb{P}_X = \mathbb{P}_{X'}$ .  $\square$

**Conséquence.** Ce résultat est souvent utilisé de la manière suivante. On dispose d'une variable aléatoire  $Y$  dont on ne connaît pas la loi, mais pour laquelle on a certaines informations (par exemple qu'elle est fonction d'une autre variable aléatoire  $X$  dont on connaît la loi). Alors si on parvient à avoir une expression pour l'espérance  $\mathbb{E}[h(Y)]$  pour tout  $h$  mesurable bornée (ou mesurable positive), alors on peut identifier la loi de  $Y$  avec une loi connue.

**Exemple 1.19.** Soit  $X$  une variable aléatoire de loi uniforme sur  $[0, 1]$  et  $Y = -\ln(X)$ . Déterminons la loi de  $Y$ . Soit  $h : \mathbb{R} \rightarrow \mathbb{R}$  une fonction mesurable bornée. La formule de transfert donne

$$\mathbb{E}[h(Y)] = \mathbb{E}[h(-\ln(X))] = \int_{\mathbb{R}} h(-\ln(x)) \mathbb{1}_{[0,1]}(x) dx = \int_0^1 h(-\ln(x)) dx.$$

Le changement de variable  $t = -\ln(x)$  donne alors

$$\mathbb{E}[h(Y)] = - \int_{\infty}^0 h(t) e^{-t} dt = \int_0^{\infty} h(t) e^{-t} dt = \int_{\mathbb{R}} h(t) e^{-t} \mathbb{1}_{[0,\infty[} dt.$$

Cette dernière intégrale correspond à l'espérance de  $h(Y')$  pour  $Y'$  une variable aléatoire de loi exponentielle de paramètre 1. D'après la proposition, la variable aléatoire  $Y$  a la même loi que  $Y'$  et suit donc une loi exponentielle de paramètre 1.

Cette technique est fondamentale pour déterminer une loi. Elle est parfois appelée technique de la fonction muette puisque la fonction  $h$  qui apparaît dans l'espérance ne joue en fait aucun rôle.

### 2.3.2 Moments d'ordre $p$

Dans un contexte probabiliste, l'espace  $L^p(\Omega, \mathcal{F}, \mathbb{P})$ , pour  $p \in [1, \infty]$ , correspond à l'ensemble des variables aléatoires  $X : \Omega \rightarrow \mathbb{R}$  telle que

$$\|X\|_p = \left( \int_{\Omega} |X(\omega)|^p d\mathbb{P}(\omega) \right)^{1/p} < \infty.$$

L'application  $\|\cdot\|_p$  est une norme qui fait de  $L^p(\Omega, \mathcal{F}, \mathbb{P})$  un espace de Banach (c'est-à-dire un espace vectoriel normé complet). Dans le cas  $p = \infty$ , on peut montrer que

$$\|X\|_{\infty} = \sup_{\omega \in \Omega} |X(\omega)|,$$

et  $X \in L^{\infty}(\Omega, \mathcal{F}, \mathbb{P})$  si  $X$  est bornée  $\mathbb{P}$ -presque partout.

Précisons que les espaces  $L^p(\Omega, \mathcal{F}, \mathbb{P})$  contiennent les fonctions mesurables définies à un ensemble de mesure nulle près pour  $\mathbb{P}$ . Donc si  $\|X\|_p = 0$ , on en déduit alors seulement que  $X$  est nulle  $\mathbb{P}$ -presque partout.

**Définition 1.20** (Moment). *Soit  $p \in [1, \infty]$ . Soit  $X$  une variable aléatoire réelle dans  $L^p(\Omega, \mathcal{F}, \mathbb{P})$ . Le moment d'ordre  $p$  de  $X$  est*

$$\mathbb{E}[X^p] = \int_{\Omega} (X(\omega))^p d\mathbb{P}(\omega) = \int_{\mathbb{R}} x^p d\mathbb{P}_X(x).$$

La deuxième égalité résulte de la formule de transfert. On remarque que le moment d'ordre  $p$  ne dépend donc de  $X$  qu'à travers sa loi.

Le moment d'ordre 1 n'est rien d'autre que l'espérance de  $X$ , si celle-ci est bien définie. La mesure de probabilité  $\mathbb{P}$  étant finie, les espaces  $L^p(\Omega, \mathcal{F}, \mathbb{P})$  sont décroissants en  $p \in [1, \infty]$ . En particulier, si  $X$  admet un moment d'ordre  $p$  alors elle admet un moment d'ordre  $p'$  pour tout  $p' \leq p$ . Remarquons que la définition du moment d'ordre  $p$  peut s'étendre à  $p \in ]0, 1[$ , mais dans ce cas  $\|\cdot\|_p$  n'est pas une norme.

Les moments d'une variable aléatoire  $X$  permettent de contrôler le comportement de  $X$  à l'infini, comme l'énonce le résultat suivant.

**Proposition 1.21** (Inégalité de Markov). *Soit  $X$  une variable aléatoire réelle positive et  $p \in [1, \infty[$ . Alors pour tout  $x > 0$ ,*

$$\mathbb{P}(X \geq x) \leq \frac{\mathbb{E}[X^p]}{x^p}.$$

Remarquons que si l'espérance de  $X^p$  est infinie, alors l'inégalité est clairement vérifiée.

*Démonstration.* Il suffit de partir de l'inégalité  $x^p \mathbb{1}_{\{X \geq x\}} \leq X^p$  (si l'indicatrice est nulle c'est évident, sinon cela signifie que  $X \geq x$  donc que  $X^p \geq x^p$ ). La linéarité de l'espérance donne alors

$$x^p \mathbb{E}[\mathbb{1}_{\{X \geq x\}}] \leq \mathbb{E}[X^p],$$

d'où le résultat puisque  $\mathbb{E}[\mathbb{1}_{\{X \geq x\}}] = \mathbb{P}(X \geq x)$ .  $\square$

Rappelons que l'espace  $L^2(\Omega, \mathcal{F}, \mathbb{P})$  a de bonnes propriétés puisque c'est un espace de Hilbert pour le produit scalaire

$$\langle X, Y \rangle = \int_{\Omega} X(\omega) Y(\omega) d\mathbb{P}(\omega) = \mathbb{E}[XY],$$

pour  $X$  et  $Y$  dans  $L^2(\Omega, \mathcal{F}, \mathbb{P})$ . La norme associée à ce produit scalaire est

$$\|X\|_2 = \sqrt{\int_{\Omega} X(\omega)^2 d\mathbb{P}(\omega)} = \sqrt{\mathbb{E}[X^2]}.$$

**Définition 1.22** (Variance). *La variance d'une variable aléatoire  $X$ , notée  $\text{Var}(X)$ , est le moment d'ordre 2 de la variable aléatoire  $X - \mathbb{E}[X]$  :*

$$\text{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = \|X - \mathbb{E}[X]\|_2^2,$$

qui est bien définie dès lors que  $X \in L^2(\Omega, \mathcal{F}, \mathbb{P})$ .

On définit également l'écart-type  $\sigma(X)$  comme la racine carrée de la variance :

$$\sigma(X) = \sqrt{\text{Var}(X)} = \|X - \mathbb{E}[X]\|_2.$$

Une variable aléatoire de variance 1 est dite réduite.

La variance est bien définie puisque si  $X \in L^2(\Omega, \mathcal{F}, \mathbb{P})$ , alors  $X$  est intégrable car les espaces  $L^p$  sont décroissants pour l'inclusion. Par suite, l'espérance de  $X$  existe et la variable aléatoire  $X^2 - \mathbb{E}[X]X + \mathbb{E}[X^2]$  est intégrable. On en déduit également la formule de Koenig-Huygens :

$$\text{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2.$$

L'écart-type (resp. la variance) correspond donc à la distance (resp. la distance au carré) au sens  $L^2(\Omega, \mathcal{F}, \mathbb{P})$  de  $X$  à son espérance : c'est une mesure de dispersion. Notons que la variance d'une variable aléatoire  $X$  est nulle si et seulement si  $X$  est constante presque sûrement.

L'interprétation de la variance en tant que mesure de la dispersion est formalisée par le résultat suivant, qui est une conséquence de l'inégalité de Markov avec  $p = 2$  et la variable aléatoire positive  $|X - \mathbb{E}[X]|$ .

**Proposition 1.23** (Inégalité de Bienaymé-Tchebychev). *Soit  $X$  une variable aléatoire dans  $L^2(\Omega, \mathcal{F}, \mathbb{P})$ . Alors, pour tout  $x > 0$ ,*

$$\mathbb{P}(|X - \mathbb{E}[X]| \geq x) \leq \frac{\text{Var}(X)}{x^2}.$$

La variance permet donc d'estimer la probabilité que  $X$  s'écarte de sa moyenne.

**Définition 1.24** (Covariance). *Soit  $X$  et  $Y$  deux variables aléatoires réelles dans  $L^2(\Omega, \mathcal{F}, \mathbb{P})$ . La covariance de  $X$  et  $Y$  est la quantité*

$$\text{Cov}(X, Y) = \langle X - \mathbb{E}[X], Y - \mathbb{E}[Y] \rangle = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y].$$

La covariance vérifie donc les propriétés d'un produit scalaire : elle est bilinéaire et on a les formules  $\text{Cov}(X, Y) = \text{Cov}(Y, X)$  et  $\text{Cov}(X, X) = \|X - \mathbb{E}[X]\|_2^2 = \text{Var}(X)$ . Par ailleurs, la relation de Pythagore correspond dans ce cas à

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y),$$

et plus généralement :

$$\text{Var}(X_1 + \dots + X_n) = \sum_{i=1}^n \text{Var}(X_i) + 2 \sum_{1 \leq i < j \leq n} \text{Cov}(X_i, X_j) = \sum_{i,j=1}^n \text{Cov}(X_i, X_j).$$

On définit alors le *coefficient de corrélation* comme le produit scalaire normalisé :

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)}\sqrt{\text{Var}(Y)}} = \frac{\langle X - \mathbb{E}[X], Y - \mathbb{E}[Y] \rangle}{\|X - \mathbb{E}[X]\|_2 \|Y - \mathbb{E}[Y]\|_2}.$$

Il correspond au cosinus de l'angle entre les vecteurs (au sens de l'algèbre linéaire)  $X - \mathbb{E}[X]$  et  $Y - \mathbb{E}[Y]$ . L'inégalité de Cauchy-Schwarz se reformule dans ce cas particulier

$$\text{Cov}(X, Y) \leq \sqrt{\text{Var}(X)}\sqrt{\text{Var}(Y)},$$

ce qui assure que  $\rho(X, Y)$  est dans  $[-1, 1]$ .

### 2.3.3 Moments des lois usuelles

On donne ci-dessous les valeurs de l'espérance et de la variance pour les lois usuelles. Les résultats s'obtiennent via la formule de transfert et la formule de Koenig-Huygens.

**Loi uniforme sur  $\{1, \dots, n\}$**  Si  $X \sim \mathcal{U}(\{1, \dots, n\})$ , alors

$$\mathbb{E}[X] = \frac{n+1}{2} \quad \text{et} \quad \text{Var}(X) = \frac{n^2-1}{12}.$$

**Loi de Bernoulli de paramètre  $p$**  Si  $X \sim \mathcal{B}(p)$ , alors

$$\mathbb{E}[X] = p \quad \text{et} \quad \text{Var}(X) = p(1 - p).$$

**Loi binomiale de paramètres  $n, p$**  Si  $X \sim \mathcal{B}(n, p)$ , alors

$$\mathbb{E}[X] = np \quad \text{et} \quad \text{Var}(X) = np(1 - p).$$

**Loi géométrique de paramètres  $p$**  Si  $X \sim \mathcal{G}(p)$ , alors

$$\mathbb{E}[X] = \frac{1}{p} \quad \text{et} \quad \text{Var}(X) = \frac{1 - p}{p^2}.$$

**Loi de Poisson de paramètre  $\theta$**  Si  $X \sim \mathcal{P}(\theta)$ , alors

$$\mathbb{E}[X] = \theta \quad \text{et} \quad \text{Var}(X) = \theta.$$

**Loi uniforme sur  $[a, b]$**  Si  $X \sim \mathcal{U}([a, b])$ , alors

$$\mathbb{E}[X] = \frac{a + b}{2} \quad \text{et} \quad \text{Var}(X) = \frac{(b - a)^2}{12}.$$

**Loi exponentielle de paramètre  $\theta$**  Si  $X \sim \mathcal{E}(\theta)$ , alors

$$\mathbb{E}[X] = \frac{1}{\theta} \quad \text{et} \quad \text{Var}(X) = \frac{1}{\theta^2}.$$

**Loi normale de paramètre  $\mu$  et  $\sigma > 0$**  Si  $X \sim \mathcal{N}(\mu, \sigma^2)$ , alors

$$\mathbb{E}[X] = \mu \quad \text{et} \quad \text{Var}(X) = \sigma^2.$$

En particulier, la loi normale  $\mathcal{N}(0, 1)$  est centrée réduite.

**Exercice 1.25.** Ces résultats ont tous été démontrés dans le cours de L2. C'est toutefois un bon exercice de refaire ces calculs.

## 2.4 Fonction associée à une variable aléatoire

La loi d'une variable aléatoire  $X$  étant une mesure, ce n'est pas un objet facile à manipuler. Dans cette section on montre qu'on peut caractériser une loi via des fonctions à valeurs réelles.

### 2.4.1 Fonction de répartition

**Définition 1.26** (Fonction de répartition). *Soit  $X : \Omega \rightarrow \mathbb{R}$  une variable aléatoire réelle. On appelle fonction de répartition la fonction  $F_X : \mathbb{R} \rightarrow [0, 1]$  donnée par*

$$F_X(x) = \mathbb{P}(X \leq x), \quad x \in \mathbb{R}.$$

On notera plus simplement  $F = F_X$  s'il n'y pas d'ambiguïté sur la variable aléatoire considérée.

**Exemple 1.27.** Si  $X$  est une variable aléatoire discrète de loi

$$\mathbb{P}_X = \sum_{k=1}^{\infty} p_k \delta_{x_k},$$

avec  $p_1, p_2, \dots$  une suite de réels positifs qui somment à 1 et les  $x_k$  supposés ordonnés, alors

$$F_X(t) = \int_{\mathbb{R}} \mathbb{1}_{]-\infty, t]}(u) d\left(\sum_{k=1}^{\infty} p_k \delta_{x_k}(u)\right) = \sum_{k=1}^{\infty} p_k \int_{\mathbb{R}} \mathbb{1}_{]-\infty, t]}(u) d\delta_{x_k}(u) = \sum_{k=1}^{\infty} p_k \delta_{x_k}([-\infty, t]),$$

ce qui donne  $F_X(t) = \sum_{k=1}^{\infty} p_k \mathbb{1}_{[x_k, \infty[}(t)$ ,  $t \in \mathbb{R}$ .

**Exemple 1.28.** Si  $X$  est une variable aléatoire absolument continue par rapport à la mesure de Lebesgue sur  $\mathbb{R}$  de densité  $f$ , alors sa fonction de répartition est

$$F_X(x) = \int_{-\infty}^x f(t) dt.$$

Les résultats de dérivation sous le signe intégrale impliquent alors que  $F_X$  est dérivable  $\lambda_d$ -presque partout de dérivée  $f$  (c'est une conséquence de théorème de différentiation de Lebesgue qui n'est pas un résultat trivial!).

Par exemple, la loi exponentielle de paramètre 1 caractérisée par la densité  $f(x) = e^{-x} \mathbb{1}_{[0, \infty[}(x)$  admet pour fonction de répartition

$$F(x) = (1 - e^{-x}) \mathbb{1}_{[0, \infty[}(x).$$

Remarquons qu'il est parfois impossible de donner une expression explicite de la fonction de répartition d'une variable aléatoire. C'est le cas de la loi normale pour laquelle on peut seulement écrire

$$F(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt.$$

La fonction de répartition de la loi normale est parfois notée  $\Phi$ . Les valeurs numériques de  $\Phi(x)$  pour différentes valeurs de  $x$  sont données dans une table.



**Remarque 1.29.** Dans certains cas, il est plus utile de travailler avec la fonction de survie  $\bar{F}_X$  (parfois aussi notée  $S_X$ ), définie par

$$\bar{F}_X(x) = 1 - F_X(x) = \mathbb{P}(X > x), \quad x \in \mathbb{R},$$

et qui a des propriétés similaires à la fonction de répartition.

**Propriétés 1.30.** Soit  $X$  une variable aléatoire réelle de fonction de répartition  $F_X$ .

1.  $F_X$  est une fonction croissante et continue à droite.
2.  $\lim_{x \rightarrow \infty} F_X(x) = 1$  et  $\lim_{x \rightarrow -\infty} F_X(x) = 0$ .
3. Notons  $F_X(x-)$  la limite à gauche de  $F_X$  en  $x$  définie par

$$F_X(x-) = \lim_{\substack{t \rightarrow x \\ t < x}} F_X(t).$$

Alors  $F_X(x-) = \mathbb{P}(X < x)$ . En particulier, on en déduit que

$$\mathbb{P}(X = x) = F_X(x) - F_X(x-).$$

*Démonstration.* 1. Si  $x \leq y$ , alors  $] -\infty, x] \subset ] -\infty, y]$ , et donc par croissance de la probabilité  $\mathbb{P}$ , on en déduit que  $F_X(x) \leq F_X(y)$ . Par ailleurs, soit  $(x_n)_{n \geq 1}$  une suite qui décroît vers un réel  $x$ . Les ensembles  $] -\infty, x_n]$  étant décroissants pour l'inclusion, la propriété de décroissance monotone donne

$$\lim_{n \rightarrow \infty} \mathbb{P}(X \leq x_n) = \mathbb{P}\left(\bigcap_{n \geq 1} \{X \leq x_n\}\right).$$

et cette dernière probabilité correspond bien à  $\mathbb{P}(X \leq x)$ . Ainsi,  $F_X(x_n) \rightarrow F_X(x)$  quand  $x_n \downarrow x$ , ce qui prouve la continuité à droite de  $F_X$ .

2. On montre la première égalité. Soit  $(x_n)_{n \geq 1}$  une suite croissante qui tend vers  $\infty$ . Les ensembles  $] -\infty, x_n]$  étant croissants, la propriété de croissance monotone implique que

$$\lim_{n \rightarrow \infty} \mathbb{P}(X \leq x_n) = \mathbb{P}\left(\bigcup_{n \geq 1} \{X \leq x_n\}\right) = \mathbb{P}(X \in \mathbb{R}) = 1,$$

Donc  $F_X(x) \rightarrow 1$  quand  $x \rightarrow \infty$ .

3. Notons déjà que  $F_X(x-)$  est bien définie car  $F_X$  est une fonction croissante donc admet des limites à gauche et à droite en tout point. Soit maintenant  $(x_n)_n$  une suite qui croît vers un réel  $x$  en restant strictement inférieure à  $x$ . Les ensembles  $] -\infty, x_n]$  sont alors croissants et vérifient  $\bigcup_{n \geq 1} \{X \leq x_n\} = \{X < x\}$ . Par propriété de limite monotone de  $\mathbb{P}$ , on obtient

$$\lim_{n \rightarrow \infty} \mathbb{P}(X \leq x_n) = \mathbb{P}\left(\bigcup_{n \geq 1} \{X \leq x_n\}\right) = \mathbb{P}(X < x).$$

□

Le dernier point montre que la fonction  $F_X$  est continue en  $x$  si et seulement si  $\mathbb{P}(X = x) = 0$ . Une fonction croissante n'admettant qu'un nombre dénombrable de points de discontinuité, on en déduit que la loi d'une variable aléatoire  $X$  ne peut charger au plus qu'un nombre dénombrable de points.

L'intérêt de la fonction de répartition est qu'elle caractérise la loi.

**Proposition 1.31.** *Soit  $X$  et  $Y$  deux variables aléatoires réelles de fonction de répartition respective  $F_X$  et  $F_Y$ . Si  $F_X = F_Y$ , alors  $X$  et  $Y$  ont même loi.*

*Démonstration.* Dire que  $F_X = F_Y$  revient à dire que les mesures  $\mathbb{P}_X$  et  $\mathbb{P}_Y$  coïncident sur les ensembles de la forme  $] -\infty, x]$  qui engendrent la tribu borélienne  $\mathcal{B}(\mathbb{R})$  et sont stables par intersection finie. Le lemme de classe monotone assure alors que  $\mathbb{P}_X = \mathbb{P}_Y$  partout.  $\square$

Notons qu'on peut comparer les lois  $\mathbb{P}_X$  et  $\mathbb{P}_Y$  même si les variables aléatoires  $X$  et  $Y$  ne sont pas définies sur le même espace probabilisé, seul l'ensemble image, ici  $\mathbb{R}$ , est important.

## 2.4.2 Fonction caractéristique

On considère dans cette section des variables aléatoires à valeurs dans  $\mathbb{R}^d$  que l'on munit du produit scalaire standard  $\langle x, y \rangle = x_1 y_1 + \dots + x_d y_d$ .

**Définition 1.32** (Fonction caractéristique). *Soit  $X$  une variable aléatoire à valeurs dans  $\mathbb{R}^d$ . La fonction caractéristique de  $X$  est la fonction  $\varphi_X : \mathbb{R}^d \rightarrow \mathbb{C}$  définie par*

$$\varphi_X(t) = \mathbb{E}[e^{i\langle t, X \rangle}] = \int_{\Omega} e^{i\langle t, X(\omega) \rangle} d\mathbb{P}(\omega) = \int_{\mathbb{R}^d} e^{i\langle t, x \rangle} d\mathbb{P}_X(x), \quad t \in \mathbb{R}^d.$$

L'inégalité  $|e^{i\langle t, X(\omega) \rangle}| \leq 1$  implique que  $\int_{\Omega} |e^{i\langle t, X(\omega) \rangle}| d\mathbb{P}(\omega) \leq \int_{\Omega} d\mathbb{P}(\omega) = \mathbb{P}(\Omega) = 1$ , ce qui assure que la fonction  $\omega \mapsto e^{i\langle t, X(\omega) \rangle}$  est intégrable donc que  $\varphi_X$  est bien définie.

La fonction caractéristique correspond donc, au signe près, à la transformée de Fourier de la variable aléatoire  $X$  vue comme une fonction mesurable de  $\omega$  dans  $\mathbb{R}^d$  (voir la section 2 de l'annexe dédiée à ce sujet). En écrivant  $X = (X_1, \dots, X_d)$ , on peut réécrire  $\varphi_X$  sous la forme

$$\varphi_X(t) = \mathbb{E}[e^{i(t_1 X_1 + \dots + t_d X_d)}] = \int_{\Omega} e^{i(t_1 X_1(\omega) + \dots + t_d X_d(\omega))} d\mathbb{P}(\omega) = \int_{\mathbb{R}^d} e^{i(t_1 x_1 + \dots + t_d x_d)} d\mathbb{P}_X(x),$$

pour tout  $t = (t_1, \dots, t_d) \in \mathbb{R}^d$ . Dans le cas  $d = 1$ , on obtient pour tout  $t \in \mathbb{R}$ ,

$$\varphi_X(t) = \mathbb{E}[e^{itX}] = \int_{\Omega} e^{itX(\omega)} d\mathbb{P}(\omega) = \int_{\mathbb{R}} e^{itx} d\mathbb{P}_X(x).$$

### Exemples 1.33.

1. Si  $X$  est une variable aléatoire discrète de loi

$$\mathbb{P}_X = \sum_{k=1}^{\infty} p_k \delta_{x_k},$$

alors

$$\varphi_X(t) = \sum_{k=1}^{\infty} p_k e^{itx_k}, \quad t \in \mathbb{R}.$$

On obtient ainsi les fonctions caractéristiques des lois discrètes usuelles. Par exemple, la fonction caractéristique d'une variable aléatoire  $X$  de loi de Bernoulli de paramètre  $p$  vaut

$$\varphi_X(t) = (1 - p) + p e^{it}, \quad t \in \mathbb{R}.$$

2. Si  $X$  suit une loi uniforme sur  $[a, b]$ , alors

$$\varphi_X(t) = \frac{1}{b-a} \int_{\mathbb{R}} \mathbb{1}_{[a,b]}(x) e^{ixt} dx = \frac{1}{b-a} \left[ \frac{e^{ixt}}{it} \right]_a^b = \frac{e^{ibt} - e^{iat}}{it(b-a)}.$$

3. Si  $X$  suit une loi normale  $\mathcal{N}(0, 1)$ , alors

$$\varphi_X(t) = \exp\left(-\frac{t^2}{2}\right), \quad t \in \mathbb{R}.$$

Donnons la preuve de ce résultat. On part de l'égalité

$$\varphi_X(t) = \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} e^{itx} dx.$$

La fonction  $(t, x) \mapsto e^{-\frac{x^2}{2}} e^{itx}$  est intégrable par rapport à  $x$ , dérivable par rapport à  $t$  de dérivée vérifiant  $|e^{-\frac{x^2}{2}} e^{itx} ix| \leq e^{-\frac{x^2}{2}} |x|$  qui est intégrable par rapport à  $x$ . Ainsi, le théorème de dérivation sous le signe intégrale donne

$$\varphi'_X(t) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} ix e^{-\frac{x^2}{2}} e^{itx} dx.$$

On effectue alors une intégration par parties :

$$\varphi'_X(t) = \frac{1}{\sqrt{2\pi}} \left[ -i e^{-\frac{x^2}{2}} e^{itx} dx \right]_{-\infty}^{\infty} - \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} (-i) e^{-\frac{x^2}{2}} it e^{itx} dx = -\frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-\frac{x^2}{2}} t e^{itx} dx.$$

Ainsi,  $\varphi_X$  est solution de l'équation différentielle  $y'(t) = -ty(t)$  avec condition initiale  $y(0) = 1$ . On en déduit alors que  $\varphi_X(t) = e^{-\frac{t^2}{2}}$ .

Par suite, si  $Y \sim \mathcal{N}(\mu, \sigma^2)$ , alors  $X = (Y - \mu)/\sigma \sim \mathcal{N}(0, 1)$  et donc

$$\varphi_Y(t) = \mathbb{E}[e^{itY}] = \mathbb{E}[e^{it(\mu + \sigma X)}] = e^{it\mu} \varphi_X(\sigma t) = \exp\left(it\mu - \frac{\sigma^2 t^2}{2}\right).$$

La transformée de Fourier permet de calculer les moments d'une variable aléatoire via le résultat suivant.

**Proposition 1.34.** *Soit  $X$  une variable aléatoire réelle. Si  $X$  admet un moment d'ordre  $p \in \mathbb{N}$ , alors  $\varphi_X$  est de classe  $\mathcal{C}^p(\mathbb{R}, \mathbb{C})$  et on a la relation*

$$\mathbb{E}[X^p] = (-i)^p \varphi_X^{(p)}(0).$$

En particulier,  $\varphi_X$  est toujours continue (le moment d'ordre 0 existe toujours).

*Démonstration.* La fonction  $(\omega, t) \mapsto e^{itX(\omega)}$  est intégrable par rapport à  $\omega$  de dérivable par rapport à  $t$  de dérivée  $t \mapsto iX(\omega) e^{itX(\omega)}$  qui vérifie

$$|iX(\omega) e^{itX(\omega)}| \leq |X(\omega)|, \quad \text{pour } \mathbb{P}\text{-presque tout } \omega \in \Omega.$$

Ainsi, si  $X$  admet un moment d'ordre  $p = 1$ , alors  $|X|$  est intégrable et le théorème de dérivation sous l'intégrale implique alors que

$$\varphi_X'(t) = \mathbb{E}[iX e^{itX}].$$

En évaluant cette égalité en  $t = 0$  on obtient le résultat pour  $p = 1$ . Une récurrence donne le résultat pour  $p$  quelconque.  $\square$

**Théorème 1.35.** *Soit  $X$  et  $Y$  deux variables aléatoires à valeurs dans  $\mathbb{R}^d$  telles que  $\varphi_X = \varphi_Y$ . Alors  $X$  et  $Y$  ont même loi.*

*Démonstration.* Admise. Elle s'appuie sur les propriétés des transformées de Fourier des mesures. On trouvera des éléments de preuve dans la section 2 de l'annexe.  $\square$

### 2.4.3 Fonction génératrice pour des variables aléatoires à valeurs dans $\mathbb{N}$

Dans le cas de variables aléatoires à valeurs dans  $\mathbb{N}$  on utilise la fonction génératrice plutôt que la fonction caractéristique.

**Définition 1.36** (Fonction génératrice). *Soit  $X$  une variable aléatoire à valeurs dans  $\mathbb{N}$ . La fonction génératrice de  $X$  est la série entière*

$$G_X(z) = \mathbb{E}[z^X] = \sum_{n=0}^{\infty} \mathbb{P}(X = n) z^n.$$

Comme la somme des coefficients  $\mathbb{P}(X = n)$  vaut 1, la fonction génératrice a un rayon de convergence supérieur ou égal à 1. Ainsi, la fonction génératrice est bien définie au moins sur  $[-1, 1]$  et est continue sur  $[-1, 1]$ . Par ailleurs,  $G_X$  est  $C^\infty$  sur  $] -1, 1[$  et on a la relation

$$\mathbb{P}(X = n) = \frac{G_X^{(n)}(0)}{n!}, \quad n \in \mathbb{N}.$$

En particulier, la fonction génératrice d'une variable aléatoire  $X$  caractérise sa loi : la donnée de  $G_X$  permet de retrouver les valeurs de  $\mathbb{P}(X = n)$  à partir desquelles on obtient toutes les probabilités  $\mathbb{P}(X \in A)$ ,  $A \subset \mathbb{N}$ , donc la loi de  $X$ .

On utilise la variable  $z$  car on peut en fait définir la fonction génératrice pour tout nombre complexe  $z$  tel que  $|z| \leq 1$  et faire alors le lien avec la fonction caractéristique. Cependant, on se contentera ici de l'intervalle  $] -1, 1[$ .

Comme pour la fonction caractéristique, l'intérêt de la fonction génératrice réside dans son lien avec les moments de  $X$ . En effet, la fonction génératrice étant une série entière, on peut la dériver terme à terme :

$$G'_X(z) = \mathbb{E}[Xz^{X-1}] = \sum_{n=1}^{\infty} n\mathbb{P}(X=n)z^{n-1}, \quad z \in ]-1, 1[.$$

Soit alors  $(z_k)$  une suite dans  $[0, 1[$  qui croît vers 1. Les fonctions  $f_k(\omega) = X(\omega)z_k^{X(\omega)-1}$  forment une suite croissante (de fonctions) donc le théorème de convergence monotone donne

$$\lim_{k \rightarrow \infty} \mathbb{E}[Xz_k^{X-1}] = \mathbb{E}[X],$$

ce qui signifie que

$$\lim_{z \uparrow 1} G'_X(z) = G'_X(1-) = \mathbb{E}[X].$$

où la limite vaut éventuellement  $\infty$ .

Plus généralement, on peut montrer par récurrence la formule

$$G_X^{(k)}(1-) = \mathbb{E}[X(X-1)\cdots(X-k+1)], \quad k \in \mathbb{N}.$$

### Exemples 1.37.

1. Si  $X$  suit une loi de Bernoulli de paramètre  $p$ , alors pour  $z \in \mathbb{R}$ ,

$$G_X(z) = \mathbb{E}[z^X] = (1-p)z^0 + pz^1 = (1-p) + pz.$$

2. Si  $X$  suit une loi binomiale de paramètres  $n$  et  $p$ , alors pour  $z \in \mathbb{R}$ ,

$$G_X(z) = \mathbb{E}[z^X] = \sum_{k=0}^n \binom{n}{k} p^k (1-p)^{n-k} z^k = \sum_{k=0}^n \binom{n}{k} (pz)^k (1-p)^{n-k} = [(1-p) + pz]^n.$$

3. Si  $X$  suit une loi géométrique de paramètre  $p$ , alors pour  $|z| < 1/(1-p)$ ,

$$G_X(z) = \mathbb{E}[z^X] = \sum_{k=1}^{\infty} p(1-p)^{k-1} z^k = pz \sum_{k=0}^{\infty} [(1-p)z]^k = \frac{pz}{1 - (1-p)z}.$$

4. Si  $X$  suit une loi de Poisson de paramètre  $\lambda$ , alors pour  $z \in \mathbb{R}$ ,

$$G_X(z) = \mathbb{E}[z^X] = \sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!} z^k = e^{-\lambda} \sum_{k=0}^{\infty} \frac{(\lambda z)^k}{k!} = e^{-\lambda} e^{z\lambda} = e^{\lambda(z-1)}.$$

**Exercice 1.38.** Retrouver la valeur de  $\mathbb{E}[X]$  dans les quatre cas précédents en utilisant la fonction génératrice.

## Illustrations graphiques des lois discrètes usuelles

### Loi uniforme sur un ensemble fini

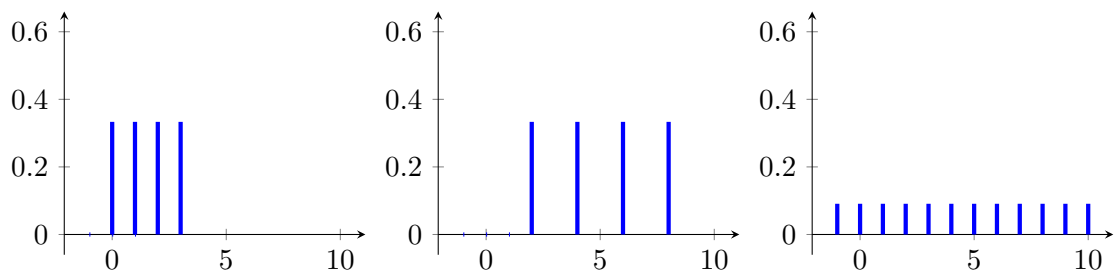


FIGURE 1.5 – Loi uniforme  $\mathcal{U}(\{0, 1, 2, 3\})$ ,  $\mathcal{U}(\{2, 4, 6, 8\})$  et  $\mathcal{U}(\{-1, 0, \dots, 10\})$ .

### Loi de Bernoulli

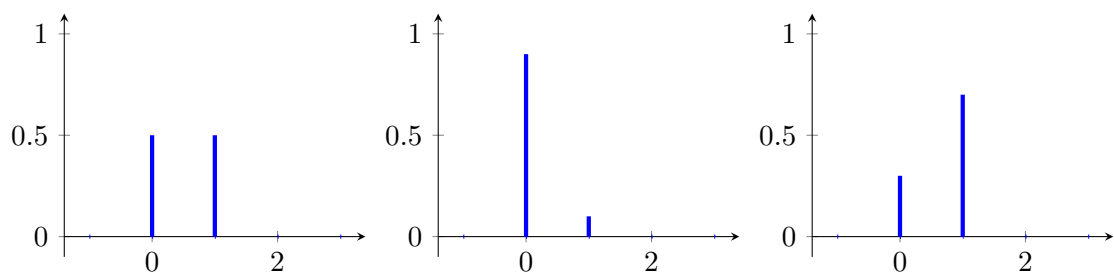


FIGURE 1.6 – Loi de Bernoulli de paramètre  $\mathcal{B}(0.5)$ ,  $\mathcal{B}(0.1)$  et  $\mathcal{B}(0.7)$ .

### Loi binomiale

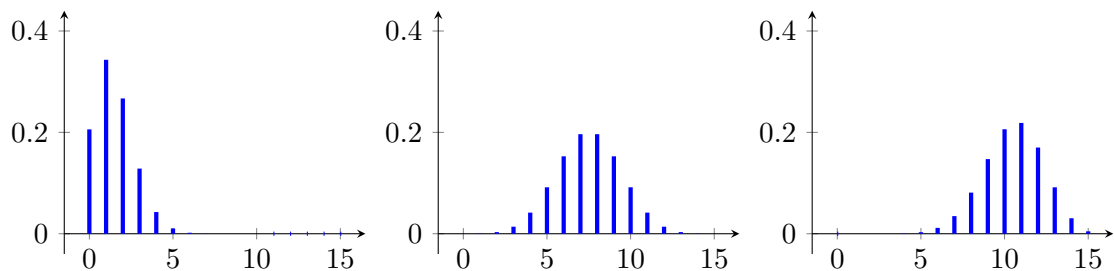
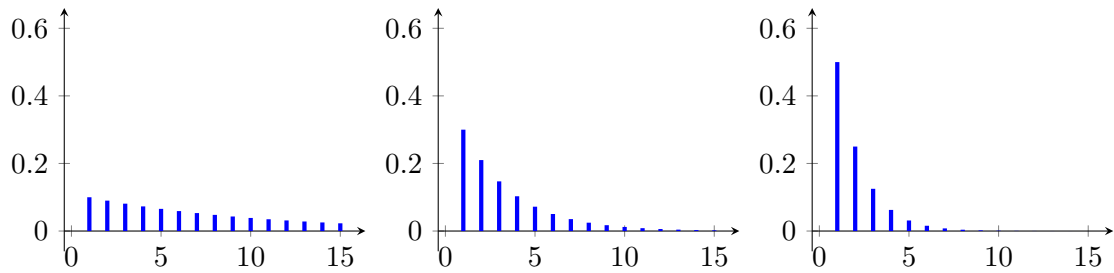
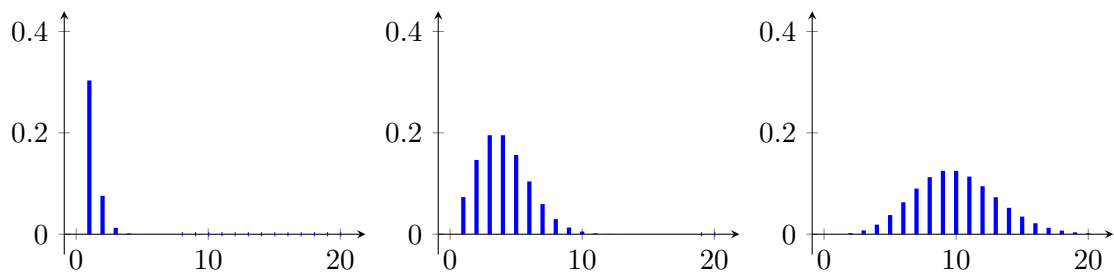


FIGURE 1.7 – Loi Binomiale de paramètre  $\mathcal{B}(15, 0.1)$ ,  $\mathcal{B}(15, 0.5)$  et  $\mathcal{B}(15, 0.7)$ .

**Loi géométrique**FIGURE 1.8 – Loi géométrique de paramètre  $\mathcal{G}(0.1)$ ,  $\mathcal{G}(0.3)$  et  $\mathcal{G}(0.5)$ .**Loi de Poisson**FIGURE 1.9 – Loi de Poisson de paramètre  $\mathcal{P}(0.5)$ ,  $\mathcal{P}(4)$  et  $\mathcal{P}(10)$ .





## Chapitre 2

# Indépendance

L'indépendance est la première notion fondamentale qui diffère grandement de la théorie de la mesure. Intuitivement, l'indépendance d'événements ou de variables aléatoires est assez aisée à appréhender. En pratique, c'est l'indépendance de sous-tribus qui sera intéressante à étudier. Le résultat clé de cette section concerne la caractérisation de l'indépendance via la mesure produit : dire que deux variables aléatoires sont indépendantes revient à dire que leur loi correspond au produit de leurs marginales.

Dans tout le chapitre on fixe un espace probabilisé  $(\Omega, \mathcal{F}, \mathbb{P})$ .

### 1 Indépendance d'événements

#### 1.1 Conditionnement

Si  $A \in \mathcal{F}$  est un événement tel que  $\mathbb{P}(A) > 0$ , on définit alors, pour tout  $B \in \mathcal{F}$ ,

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

que l'on appelle *probabilité conditionnelle de B sachant A*. La distribution de l'union sur l'intersection et les égalités  $\mathbb{P}(\emptyset \mid A) = \mathbb{P}(\emptyset \cap A)/\mathbb{P}(A) = 0$  et  $\mathbb{P}(\Omega \mid A) = \mathbb{P}(\Omega \cap A)/\mathbb{P}(A) = 1$  impliquent que l'application

$$\begin{array}{ccc} \mathbb{P}_A & : & \mathcal{F} \rightarrow [0, 1] \\ & & B \mapsto \mathbb{P}(B \mid A), \end{array}$$

est une mesure de probabilité sur  $(\Omega, \mathcal{F})$ . Intuitivement, l'espace probabilisé  $(\Omega, \mathcal{F}, \mathbb{P}_A)$  est l'espace correspondant à une expérience aléatoire pour laquelle on sait a priori que l'événement  $A$  est réalisé.

Si  $A$  et  $B$  sont deux événements de probabilité strictement positive, alors

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

égalité parfois appelée *formule de Bayes*. Cette formule permet d'inverser le conditionnement.

## 1.2 Quelques formules

L'idée du conditionnement est de pouvoir calculer la probabilité d'un événement  $B$  si on connaît la probabilité de  $B$  sachant  $A_i$  pour une famille assez riche de  $A_i$ .

Soit  $I \subset \mathbb{N}$  un ensemble d'indices fini ou dénombrable. La famille d'événements  $(A_i)_{i \in I}$  forme une partition si les  $A_i$  sont deux à deux disjoints et de réunion  $\Omega$ . Dans ce cas, on a la *formule des probabilités totales* : si les  $A_i$  sont de probabilité strictement positive, alors pour  $B \in \mathcal{F}$ ,

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

Un cas particulier intéressant de cette formule concerne le cas d'une partition formée par deux événements  $A$  et  $A^c$ . La formule devient alors

$$\mathbb{P}(B) = \mathbb{P}(B \mid A) \mathbb{P}(A) + \mathbb{P}(B \mid A^c) \mathbb{P}(A^c).$$

À partir de cette relation on peut réécrire la *formule de Bayes* :

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

pour tout  $B$  tel que  $\mathbb{P}(B) > 0$  et pour  $j \in I$  fixé. Ici aussi le cas particulier de la partition  $(A, A^c)$  est d'usage courant :

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(B \mid A) \mathbb{P}(A)}{\mathbb{P}(B \mid A) \mathbb{P}(A) + \mathbb{P}(B \mid A^c) \mathbb{P}(A^c)}.$$

**Exemple 2.1.** Une maladie touche une personne sur 100. Un test de dépistage est proposé. Si une personne est malade, le test est positif à 99%. Si une personne n'est pas malade, le test est positif dans 1% des cas. On souhaite savoir si ce test est efficace, c'est-à-dire si une personne testée positive est effectivement malade. Pour cela, on note  $M$  l'événement "la personne est malade" et  $P$  l'événement "le test est positif". Les données se traduisent alors par

$$\mathbb{P}(M) = \frac{1}{100} = 1 - \mathbb{P}(M^c), \quad \mathbb{P}(P \mid M) = \frac{99}{100}, \quad \mathbb{P}(P \mid M^c) = \frac{99}{100}.$$

La formule de Bayes donne alors

$$\mathbb{P}(M \mid P) = \frac{\mathbb{P}(P \mid M) \mathbb{P}(M)}{\mathbb{P}(P \mid M) \mathbb{P}(M) + \mathbb{P}(P \mid M^c) \mathbb{P}(M^c)} = \frac{\frac{99}{100} \times \frac{1}{100}}{\frac{99}{100} \times \frac{1}{100} + \frac{1}{100} \times \frac{99}{100}} = \frac{1}{2}.$$

Il y a une chance sur deux que la personne testée positive soit effectivement malade : autant effectuer le dépistage à pile ou face !

### 1.3 Indépendance d'événements

**Définition 2.2** (Indépendance de deux événements). *Deux événements  $A$  et  $B$  sont dits indépendants si  $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$ .*

Autrement dit, si  $\mathbb{P}(A) > 0$ , cela signifie que  $\mathbb{P}(B \mid A) = \mathbb{P}(B)$ , c'est-à-dire que la réalisation de  $A$  n'influence pas celle de  $B$ .

**Exemple 2.3.** On lance deux dés. L'univers est alors  $\Omega = \{1, \dots, 6\}^2$  que l'on munit de la tribu pleine et de la probabilité uniforme :  $\mathbb{P}(\{\omega\}) = 1/36$  pour tout  $\omega \in \Omega$ . Les événements  $A = \{6\} \times \{1, \dots, 6\}$  et  $B = \{1, \dots, 6\} \times \{6\}$  vérifient

$$\mathbb{P}(A) = \frac{6}{36} = \mathbb{P}(B) \quad \text{et} \quad \mathbb{P}(A \cap B) = \mathbb{P}(\{(6, 6)\}) = \frac{1}{36},$$

ce qui prouve l'indépendance de  $A$  et  $B$ .

**Définition 2.4** (Indépendance de plusieurs événements). *Plus généralement, des événements  $A_1, \dots, A_n$  sont dits indépendantes si pour tout  $J \subset \{1, \dots, n\}$ , on a*

$$\mathbb{P}\left(\bigcap_{j \in J} A_j\right) = \prod_{j \in J} \mathbb{P}(A_j) \quad (2.1)$$

Attention, il convient bien de vérifier la propriété pour toute partie  $J \subset \{1, \dots, n\}$ . En particulier, il ne suffit pas de vérifier que pour chaque paire  $(i, j) \subset \{1, \dots, n\}$ , les événements  $A_i$  et  $A_j$  sont indépendants.

**Exemple 2.5.** On lance deux dés et on considère les événements  $A$  : "le premier dé donne un résultat impair" ;  $B$  : "le deuxième dé donne un résultat pair" ;  $C$  : "la somme des deux dés est un nombre pair".

On modélise cette expérience par l'univers  $\Omega = \{1, \dots, 6\}^2$  muni de la tribu pleine et de la probabilité uniforme (chaque tirage a la même probabilité d'apparaître). On obtient rapidement les probabilités suivantes

$$\mathbb{P}(A) = \frac{1}{2}, \quad \mathbb{P}(B) = \frac{1}{2}, \quad \mathbb{P}(C) = \frac{1}{2}.$$

$$\mathbb{P}(A \cap B) = \frac{1}{4}, \quad \mathbb{P}(A \cap C) = \frac{1}{4}, \quad \mathbb{P}(B \cap C) = \frac{1}{2}.$$

Ainsi les événements  $A$ ,  $B$ , et  $C$  sont deux à deux indépendants, mais comme  $\mathbb{P}(A \cap B \cap C) = \mathbb{P}(\emptyset) = 0 \neq \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C)$ , ils ne sont pas indépendants.

**Remarque 2.6.** Vérifier l'indépendance de  $n$  événements revient donc à vérifier  $2^n - n - 1$  conditions (toutes les parties  $I$  de  $\{1, \dots, n\}$  de cardinal supérieur ou égal à 2). Cela peut donc s'avérer fastidieux à vérifier en pratique. Souvent, c'est le contexte qui assure l'indépendance d'événements, par exemple dans le cas de plusieurs lancers de dés, chaque lancer n'influençant pas le résultat des autres.

La définition précédente s'étend à une famille quelconque  $(A_i)_{i \in I}$ . On dit que ces événements sont indépendants si l'égalité (2.1) est vraie pour tout sous-ensemble  $J \subset I$  fini.

**Proposition 2.7.** *On peut vérifier l'indépendance en passant au complémentaire. Autrement dit, si  $A_1, A_2, \dots, A_n$  sont des événements indépendants, alors  $A_1^c, A_2, \dots, A_n$  sont également indépendants.*

*Démonstration.* Il suffit d'écrire, pour tout  $1 < i_2 < \dots < i_k \leq n$ ,

$$\begin{aligned} \mathbb{P}(A_1^c \cap A_{i_2} \cap \dots \cap A_{i_k}) &= \mathbb{P}(A_{i_2} \cap \dots \cap A_{i_k}) - \mathbb{P}(A_1 \cap A_{i_2} \cap \dots \cap A_{i_k}) \\ &= (1 - \mathbb{P}(A_1)) \prod_{j=2}^k \mathbb{P}(A_{i_j}) \\ &= \mathbb{P}(A_1^c) \mathbb{P}(A_{i_2}) \dots \mathbb{P}(A_{i_k}), \end{aligned}$$

ce qui permet de conclure.  $\square$

## 2 Indépendance de variables aléatoires

### 2.1 Indépendance de tribus

**Définition 2.8** (Indépendance de tribus). *Soit  $\mathcal{F}_1, \dots, \mathcal{F}_n$  des sous-tribus de  $\mathcal{F}$ . On dit que  $\mathcal{F}_1, \dots, \mathcal{F}_n$  sont indépendantes si*

$$\forall A_1 \in \mathcal{F}_1, \dots, \forall A_n \in \mathcal{F}_n, \quad \mathbb{P}(A_1 \cap \dots \cap A_n) = \prod_{i=1}^n \mathbb{P}(A_i). \quad (2.2)$$

Cela signifie que pour tout  $A_1 \in \mathcal{F}_1, \dots, A_n \in \mathcal{F}_n$ , les événements  $A_1, \dots, A_n$  sont indépendants (attention, un sens de l'équivalence nécessite un peu de travail).

Plus généralement, si  $(\mathcal{F}_i)_{i \in I}$  est une famille quelconques de tribus, on dit qu'elles sont indépendantes si les tribus  $(\mathcal{F}_j)_{j \in J}$  le sont pour tout sous-ensemble  $J \subset I$  fini.

En pratique, il n'est pas nécessaire de vérifier la formule (2.2) pour tous les événements. On peut se limiter à une sous-classe suffisamment riche d'événements comme le montre le lemme suivant.

**Lemme 2.9.** *Soit  $\mathcal{F}_1, \dots, \mathcal{F}_n$  des sous-tribus de  $\mathcal{F}$ . On considère  $\mathcal{C}_1 \subset \mathcal{F}_1, \dots, \mathcal{C}_n \subset \mathcal{F}_n$  tels que*

1. *les  $\mathcal{C}_1, \dots, \mathcal{C}_n$  sont stables par intersection finie,*
2. *pour tout  $i \in \{1, \dots, n\}$ ,  $\Omega \in \mathcal{C}_i$ ,*
3. *pour tout  $i \in \{1, \dots, n\}$ ,  $\sigma(\mathcal{C}_i) = \mathcal{F}_i$ ,*
4. *pour tout  $C_1 \in \mathcal{C}_1, \dots, C_n \in \mathcal{C}_n$ , on a*

$$\mathbb{P}\left(\bigcap_{i=1}^n C_i\right) = \prod_{i=1}^n \mathbb{P}(C_i).$$

Alors les tribus  $\mathcal{F}_1, \dots, \mathcal{F}_n$  sont indépendantes.

La démonstration de ce résultat est donnée en fin de chapitre.

**Corollaire 2.10** (Lemme de regroupement par paquets). *Soit  $\mathcal{F}_1, \dots, \mathcal{F}_n$  des sous-tribus indépendantes et  $1 \leq n_1 < \dots < n_p = n$ . Alors les tribus*

$$\begin{aligned}\mathcal{G}_1 &= \sigma(\mathcal{F}_1 \cup \dots \cup \mathcal{F}_{n_1}) = \sigma(\mathcal{F}_1, \dots, \mathcal{F}_{n_1}), \\ \mathcal{G}_2 &= \sigma(\mathcal{F}_{n_1+1} \cup \dots \cup \mathcal{F}_{n_2}) = \sigma(\mathcal{F}_{n_1+1}, \dots, \mathcal{F}_{n_2}), \\ &\vdots \\ \mathcal{G}_p &= \sigma(\mathcal{F}_{n_{p-1}+1} \cup \dots \cup \mathcal{F}_{n_p}) = \sigma(\mathcal{F}_{n_{p-1}+1}, \dots, \mathcal{F}_{n_p})\end{aligned}$$

sont indépendantes.

Par exemple, si  $\mathcal{F}_1, \mathcal{F}_2, \mathcal{F}_3, \mathcal{F}_4$  sont des sous-tribus indépendantes d'une tribu  $\mathcal{F}$ , alors  $\sigma(\mathcal{F}_1, \mathcal{F}_2)$  et  $\sigma(\mathcal{F}_3, \mathcal{F}_4)$  sont indépendantes.

*Démonstration.* Pour  $j = 1, \dots, p$ , on définit  $\mathcal{C}_j$  comme la classe des parties de la forme

$$B_{n_{j-1}+1} \cap \dots \cap B_{n_j},$$

où  $B_i \subset \mathcal{F}_i$  pour tout  $i \in \{n_{j-1} + 1, \dots, n_j\}$

Alors les  $\mathcal{C}_j$  sont stables par intersection finie et contiennent  $\Omega$ . Par ailleurs,  $\mathcal{C}_j$  contient tous les  $\mathcal{F}_{n_{j-1}+1}, \dots, \mathcal{F}_{n_j}$ , donc  $\sigma(\mathcal{C}_j) = \mathcal{G}_j$ . L'indépendance des sous-tribus  $\mathcal{F}_i$  assure alors que la dernière hypothèse du lemme 2.9 est vérifiée, ce qui conclut la preuve.  $\square$

## 2.2 Indépendance de variables aléatoires

Soit  $X$  une variable aléatoire à valeurs dans un espace mesurable  $(E, \mathcal{E})$ . On peut lui associer la tribu

$$\sigma(X) = \{X^{-1}(A), A \in \mathcal{E}\},$$

qui est une sous-tribu de  $\mathcal{F}$ , appelée *tribu engendrée par la variable aléatoire  $X$* . On interprète  $\sigma(X)$  comme l'information contenue dans la variable aléatoire  $X$ , ou encore les événements mesurables par rapport à  $X$ .

**Exemple 2.11.** On considère le lancer de deux pièces équilibrées et on désigne par  $X$  la variable aléatoire valant 1 si la première pièce donne Pile et 0 sinon. L'espace probabilisé est alors  $(\Omega, \mathcal{F}, \mathbb{P})$ , avec  $\Omega = \{P, F\}^2$ ,  $\mathcal{F} = \mathcal{P}(\Omega)$ , et  $\mathbb{P}$  la probabilité uniforme sur  $\Omega$ . L'ensemble d'arrivée est  $\{0, 1\}$  muni de la tribu pleine

$$\mathcal{F}' = \mathcal{P}(\{0, 1\}) = \{\{0\}, \{1\}, \{0, 1\}, \emptyset\},$$

ce qui donne

$$X^{-1}(\{1\}) = \{PP, PF\}, \quad X^{-1}(\{0\}) = \{FF, FP\}, \quad X^{-1}(\{0, 1\}) = \Omega, \quad X^{-1}(\emptyset) = \emptyset.$$

On en déduit alors que  $\sigma(X) = \{\emptyset, \Omega, \{PP, PF\}, \{FF, FP\}\}$ .

Plus généralement, si  $X_1, \dots, X_n$  est une suite de variables aléatoires avec  $X_j : (\Omega, \mathcal{F}) \rightarrow (E_j, \mathcal{E}_j)$ , la tribu engendrée par ces variables aléatoires est définie par

$$\sigma(X_1, \dots, X_n) := \sigma(X_j^{-1}(A), A \in \mathcal{E}_j, j = 1, \dots, n).$$

On dispose de tous les éléments pour définir l'indépendance de variables aléatoires.

**Définition 2.12** (Indépendance de variables aléatoires). *Soit  $X_1, \dots, X_n$  des variables aléatoires à valeurs dans  $(E_1, \mathcal{E}_1), \dots, (E_n, \mathcal{E}_n)$ . On dit que ces variables aléatoires sont indépendantes si les tribus  $\sigma(X_1), \dots, \sigma(X_n)$  le sont. Cela signifie que*

$$\forall i \in \{1, \dots, n\}, \quad \forall A_i \in \mathcal{E}_i, \quad \mathbb{P}\left(\bigcap_{i=1}^n \{X_i \in A_i\}\right) = \prod_{i=1}^n \mathbb{P}(X_i \in A_i).$$

Plus généralement, les  $(X_i)_{i \in I}$  forme une famille de variables aléatoires indépendantes si pour tout sous-ensemble fini  $J \subset I$ ,  $(X_j)_{j \in J}$  est une famille de variables aléatoires indépendantes.

La définition précédente dit que les variables aléatoires  $X_1, \dots, X_n$  sont indépendantes si et seulement si

$$\mathbb{P}((X_1, \dots, X_n) \in A_1 \times \dots \times A_n) = \mathbb{P}(X_1 \in A_1) \cdots \mathbb{P}(X_n \in A_n).$$

pour tout  $A_1 \in \mathcal{E}_1, \dots, A_n \in \mathcal{E}_n$ . Notons que le vecteur aléatoire  $(X_1, \dots, X_n)$  est définie sur l'espace produit  $E_1 \times \dots \times E_n$  muni de la tribu produit  $\mathcal{E}_1 \otimes \dots \otimes \mathcal{E}_n$ . Du point de vue des mesures, l'égalité précédente revient à dire que

$$\mathbb{P}_{(X_1, \dots, X_n)} = \mathbb{P}_{X_1} \otimes \dots \otimes \mathbb{P}_{X_n},$$

où le membre de droite correspond à une mesure produit. (Notons qu'en toute rigueur il faudrait invoquer le lemme de classe monotone pour justifier le passage de l'égalité avec les  $A_i$  à l'égalité des mesures.)

**Proposition 2.13.** *Soit  $X_1, \dots, X_n$  des variables aléatoires à valeurs dans  $(E_1, \mathcal{E}_1), \dots, (E_n, \mathcal{E}_n)$ . Les assertions suivantes sont équivalentes.*

1. *Les variables aléatoires  $X_1, \dots, X_n$  sont indépendantes.*
2. *Pour toutes fonctions mesurables  $h_1 : E_1 \rightarrow \mathbb{R}, \dots, h_n : E_n \rightarrow \mathbb{R}$ , soit positives, soit dans  $L^1(E_j, \mathcal{E}_j, \mathbb{P}_{X_j})$ , on a l'égalité*

$$\mathbb{E}[h_1(X_1) \cdots h_n(X_n)] = \mathbb{E}[h_1(X_1)] \cdots \mathbb{E}[h_n(X_n)].$$

Ce résultat s'étend sans problème si les fonctions  $h_1, \dots, h_n$  sont à valeurs dans  $\mathbb{C}$  (et on se trouve alors dans le cas  $L^1(E_j, \mathcal{E}_j, \mathbb{P}_{X_j})$ ).

*Démonstration.* Supposons que les variables aléatoires  $X_1, \dots, X_n$  sont indépendantes. Alors, la formule de transfert appliquée au vecteur aléatoire

$$X = (X_1, \dots, X_n) : \begin{array}{ccc} \Omega & \rightarrow & E_1 \times \dots \times E_n \\ \omega & \mapsto & (X_1(\omega), \dots, X_n(\omega)) \end{array}$$

implique que

$$\begin{aligned} \mathbb{E}[h_1(X_1) \cdots h_n(X_n)] &= \int_{\Omega} h_1(X_1(\omega)) \cdots h_n(X_n(\omega)) \, d\mathbb{P}(\omega) \\ &= \int_{E_1 \times \dots \times E_n} h_1(x_1) \cdots h_n(x_n) \, d\mathbb{P}_{(X_1, \dots, X_n)}(x_1, \dots, x_n) \\ &= \int_{E_1 \times \dots \times E_n} h_1(x_1) \cdots h_n(x_n) \, d(\mathbb{P}_{X_1} \otimes \dots \otimes \mathbb{P}_{X_n})(x_1, \dots, x_n). \end{aligned}$$

Le théorème de Fubini (Tonelli ou Lebesgue en fonction de l'hypothèse sur les  $h_i$ ) implique alors que

$$\mathbb{E}[h_1(X_1) \cdots h_n(X_n)] = \int_{E_1} h_1(x_1) \, d\mathbb{P}_{X_1}(x_1) \cdots \int_{E_n} h_n(x_n) \, d\mathbb{P}_{X_n}(x_n).$$

et cette dernière quantité n'est rien d'autre que  $\mathbb{E}[h_1(X_1)] \cdots \mathbb{E}[h_n(X_n)]$ .

Réciproquement, supposons l'égalité des espérances vérifiée et fixons  $A_1 \in \mathcal{E}_1, \dots, A_n \in \mathcal{E}_n$ . On pose  $h_j = \mathbb{1}_{A_j}$  et l'égalité devient

$$\mathbb{E}[\mathbb{1}_{A_1}(X_1) \cdots \mathbb{1}_{A_n}(X_n)] = \mathbb{E}[\mathbb{1}_{A_1}(X_1)] \cdots \mathbb{E}[\mathbb{1}_{A_n}(X_n)]$$

ce qui donne

$$\mathbb{P}(X_1 \in A_1, \dots, X_n \in A_n) = \mathbb{P}(X_1 \in A_1) \cdots \mathbb{P}(X_n \in A_n),$$

d'où l'indépendance des  $X_j$ . □

Par exemple si  $X_1, \dots, X_n$  sont des variables aléatoires réelles intégrables et indépendantes, alors  $\mathbb{E}[X_1 \cdots X_n] = \mathbb{E}[X_1] \cdots \mathbb{E}[X_n]$ .

Un autre exemple concerne les fonctions caractéristiques. Si  $X$  et  $Y$  sont deux variables aléatoires indépendantes à valeurs dans  $\mathbb{R}^d$ , alors la proposition précédente implique que

$$\mathbb{E}[e^{i\langle t, X \rangle} e^{i\langle t, Y \rangle}] = \mathbb{E}[e^{i\langle t, X \rangle}] \mathbb{E}[e^{i\langle t, Y \rangle}],$$

pour tout  $t \in \mathbb{R}^d$ , ce qui signifie que  $\varphi_{X+Y} = \varphi_X \varphi_Y$ .

**Conséquence.** Pour des variables aléatoires  $X$  et  $Y$  dans  $L^2(\Omega, \mathcal{F}, \mathbb{P})$ , on rappelle que

$$\text{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y].$$

Si  $X$  et  $Y$  sont indépendantes, la proposition précédente implique alors que  $\text{Cov}(X, Y) = 0$ . La réciproque de ce résultat est fausse (cf votre cours de L2). En particulier, sous l'hypothèse d'indépendance la variance est additive :  $\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y)$ . Ce résultat se généralise à  $n$  variables aléatoires  $X_1, \dots, X_n$  indépendantes :  $\text{Var}(X_1 + \dots + X_n) = \text{Var}(X_1) + \dots + \text{Var}(X_n)$ .

### 2.3 Critères d'indépendance

On donne ici des critères d'indépendance utiles en pratique. Il faut retenir que si des variables aléatoires  $X_1, \dots, X_n$  sont indépendantes alors des quantités liées à ces variables se factorisent.

**Cas discret** Soit  $X_1, \dots, X_n$  des variables aléatoires à valeurs dans des ensembles discrets  $E_1, \dots, E_n$ . Elles sont indépendantes si et seulement si pour tout  $(x_1, \dots, x_n) \in E_1 \times \dots \times E_n$ ,

$$\mathbb{P}(X_1 = x_1, \dots, X_n = x_n) = \prod_{i=1}^n \mathbb{P}(X_i = x_i) \cdots \mathbb{P}(X_n = x_n).$$

Pour le sens direct, il suffit de prendre  $A_i = \{x_i\}$ . Pour la réciproque, on écrit  $A_i = \bigcup_{x \in A_i} \{x\}$ , et, cette union étant dénombrable et formée d'événements disjoints, on en déduit que

$$\begin{aligned} \mathbb{P}\left(\bigcap_{i=1}^n \{X_i \in A_i\}\right) &= \mathbb{P}\left(X_1 \in \bigcup_{x_1 \in A_1} \{x_1\}, \dots, X_n \in \bigcup_{x_n \in A_n} \{x_n\}\right) \\ &= \sum_{x_1 \in A_1} \cdots \sum_{x_n \in A_n} \mathbb{P}(X_1 = x_1, \dots, X_n = x_n) \\ &= \sum_{x_1 \in A_1} \cdots \sum_{x_n \in A_n} \mathbb{P}(X_1 = x_1) \cdots \mathbb{P}(X_n = x_n) \\ &= \mathbb{P}(X_1 \in A_1) \cdots \mathbb{P}(X_n \in A_n). \end{aligned}$$

**Cas de variables aléatoires à valeurs dans  $\mathbb{R}$**  Soit  $X_1, \dots, X_n$  des variables aléatoires à valeurs réelles. Elles sont indépendantes si et seulement si  $x_1, \dots, x_n \in \mathbb{R}$ ,

$$\mathbb{P}(X_1 \leq x_1, \dots, X_n \leq x_n) = \mathbb{P}(X_1 \leq x_1) \cdots \mathbb{P}(X_n \leq x_n)$$

Autrement dit, la fonction de répartition du vecteur aléatoire  $(X_1, \dots, X_n)$  est égale au produit des fonctions de répartition des marginales  $X_k$ . Ce résultat est une conséquence du lemme 2.9 avec comme classe  $\mathcal{C}_i$  les ensembles de la forme  $\{X_i \leq x\}$  et  $\Omega$  tout entier.

**Cas de variables aléatoires à valeurs dans  $\mathbb{R}^d$**  Soit  $X_1, \dots, X_n$  des variables aléatoires à valeurs dans  $\mathbb{R}^d$ . Elles sont indépendantes si et seulement si pour tout  $t_1, \dots, t_n \in \mathbb{R}$ ,

$$\varphi_{(X_1, \dots, X_n)}(t_1, \dots, t_n) = \varphi_{X_1}(t_1) \cdots \varphi_{X_n}(t_n),$$

c'est-à-dire que

$$\mathbb{E}[e^{i\langle t, X \rangle}] = \mathbb{E}[e^{it_1 X_1}] \cdots \mathbb{E}[e^{it_n X_n}],$$

où  $t = (t_1, \dots, t_n)$  et  $X = (X_1, \dots, X_n)$ . Le sens direct est une conséquence de la proposition 2.13. La réciproque est plus délicate à montrer, elle résulte de l'injectivité de la transformée de Fourier. La démonstration est donnée en fin de chapitre.



**Cas de variables aléatoires à densité** Soit  $X = (X_1, \dots, X_d)$  un vecteur aléatoire à valeurs dans  $\mathbb{R}^d$ . On suppose que  $X$  admet une densité  $f_X : \mathbb{R}^d \rightarrow \mathbb{R}_+$  qui se factorise sous la forme

$$f_X(x) = \prod_{i=1}^d f_i(x_i), \quad x = (x_1, \dots, x_d),$$

où les  $f_i : \mathbb{R} \rightarrow \mathbb{R}_+$  sont des fonctions mesurables telles que  $\int_{\mathbb{R}} f_i(y) dy = 1$ . Alors les variables aléatoires  $X_1, \dots, X_d$  sont indépendantes et pour tout  $i = 1, \dots, d$ ,  $f_i$  est la densité de  $X_i$ .

*Démonstration.* Si  $h_1, \dots, h_d : \mathbb{R} \rightarrow \mathbb{R}$  sont des fonctions mesurables bornées, alors le théorème de Fubini-Lebesgue implique que

$$\mathbb{E}\left[\prod_{i=1}^d h_i(X_i)\right] = \int_{\mathbb{R}^d} \prod_{i=1}^d h_i(x_i) f_i(x_i) dx_i = \prod_{i=1}^d \left( \int_{\mathbb{R}} h_i(t) f_i(t) dt \right).$$

Maintenant, en prenant  $h_i = 1$  sauf une (disons  $h_j$ ), on obtient

$$\mathbb{E}[h_j(X_j)] = \int_{\mathbb{R}} h_j(t) f_j(t) dt,$$

ce qui prouve que  $X_j$  est une variable aléatoire de densité  $f_j$ . En reprenant des  $h_i$  quelconques, on a donc montré que

$$\mathbb{E}\left[\prod_{i=1}^d h_i(X_i)\right] = \prod_{i=1}^d \mathbb{E}[h_i(X_i)]$$

La proposition 2.13 donne alors l'indépendance des  $X_i$ . □

**Exemple 2.14.** Soit  $(X, Y)$  un vecteur aléatoire de densité  $f$  sur  $\mathbb{R}^2$  donnée par

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

Le résultat précédent implique que les marginales  $X$  et  $Y$  sont indépendantes de loi normale centrée réduite.

### 3 Résultats asymptotiques

#### 3.1 Lemme de Borel-Cantelli

Si  $(A_n)_{n \in \mathbb{N}}$  est une suite d'événements de  $\Omega$ , on pose

$$\limsup_{n \rightarrow \infty} A_n = \bigcap_{n=0}^{\infty} \bigcup_{k=n}^{\infty} A_k = \{\omega \in \Omega : \omega \text{ appartient à une infinité de } A_n\},$$

et

$$\liminf_{n \rightarrow \infty} A_n = \bigcup_{n=0}^{\infty} \bigcap_{k=n}^{\infty} A_k = \{\omega \in \Omega : \omega \text{ appartient tous les } A_n \text{ à partir d'un certain rang}\}.$$

Ce sont clairement des éléments de la tribu  $\mathcal{F}$  comme union et intersection d'éléments de  $\mathcal{F}$ . De plus, on a les relations

$$\limsup_{n \rightarrow \infty} A_n^c = (\liminf_{n \rightarrow \infty} A_n)^c \quad \text{et} \quad \liminf_{n \rightarrow \infty} A_n^c = (\limsup_{n \rightarrow \infty} A_n)^c.$$

**Proposition 2.15** (Lemme de Borel-Cantelli). *Soit  $(A_n)_{n \in \mathbb{N}}$  une suite d'événements.*

1. *Si  $\sum_{n=0}^{\infty} \mathbb{P}(A_n) < \infty$ , alors  $\mathbb{P}(\limsup_{n \rightarrow \infty} A_n) = 0$ , ou de manière équivalente,*

$$\{n \in \mathbb{N} : \omega \in A_n\} \text{ est fini p.s.}$$

2. *Si  $\sum_{n=0}^{\infty} \mathbb{P}(A_n) = \infty$  et que les  $A_n$  sont indépendants, alors  $\mathbb{P}(\limsup_{n \rightarrow \infty} A_n) = 1$ , ou de manière équivalente,*

$$\{n \in \mathbb{N} : \omega \in A_n\} \text{ est infini p.s.}$$

L'hypothèse d'indépendance est nécessaire comme le montre le contre-exemple  $A_n = A$  avec  $0 < \mathbb{P}(A) < 1$ .

*Démonstration.*

1. Les ensembles  $B_n = \bigcup_{k=n}^{\infty} A_k$  sont décroissants pour l'inclusion, donc

$$\mathbb{P}\left(\bigcap_{n=0}^{\infty} B_n\right) = \lim_{n \rightarrow \infty} \mathbb{P}(B_n).$$

Or, par sous-additivité, on a

$$\mathbb{P}(B_n) = \mathbb{P}\left(\bigcup_{k=n}^{\infty} A_k\right) \leq \sum_{k=n}^{\infty} \mathbb{P}(A_k),$$

et ce dernier terme tend vers 0 car c'est le reste d'une série qui est convergente par hypothèse. Ainsi,

$$\mathbb{P}(\limsup_{n \rightarrow \infty} A_n) = \mathbb{P}\left(\bigcap_{n=0}^{\infty} B_n\right) = \lim_{n \rightarrow \infty} \mathbb{P}(B_n) = 0.$$

2. Montrons que  $\mathbb{P}(\liminf_{n \rightarrow \infty} A_n^c) = 0$ .

Pour  $n \in \mathbb{N}$ , le théorème de convergence dominée assure que

$$\mathbb{P}\left(\bigcap_{k=n}^{\infty} A_k^c\right) = \mathbb{E}\left[\mathbf{1}_{\bigcap_{k=n}^{\infty} A_k^c}\right] = \mathbb{E}\left[\prod_{k=n}^{\infty} \mathbf{1}_{A_k^c}\right] = \lim_{N \rightarrow \infty} \mathbb{E}\left[\prod_{k=n}^N \mathbf{1}_{A_k^c}\right].$$

Par indépendance des  $A_k$  la proposition 2.13 donne l'égalité  $\mathbb{E}[\prod_{k=n}^N \mathbb{1}_{A_k^c}] = \prod_{k=n}^N \mathbb{E}[\mathbb{1}_{A_k^c}]$ . En utilisant l'inégalité de convexité  $1 - x \leq e^{-x}$  valable pour tout réel  $x$ , on en déduit alors que

$$\mathbb{P}\left(\bigcap_{k=n}^{\infty} A_k^c\right) = \lim_{N \rightarrow \infty} \prod_{k=n}^N \mathbb{E}[\mathbb{1}_{A_k^c}] = \lim_{N \rightarrow \infty} \prod_{k=n}^N (1 - \mathbb{P}(A_k)) \leq \exp\left(-\sum_{k=n}^{\infty} \mathbb{P}(A_k)\right).$$

L'hypothèse  $\sum_{k=0}^{\infty} \mathbb{P}(A_k) = \infty$  implique que cette dernière quantité est nulle pour tout  $n \in \mathbb{N}$ . On conclut alors par sous  $\sigma$ -additivité :

$$\mathbb{P}(\limsup_{n \rightarrow \infty} A_n^c) = \mathbb{P}\left(\bigcup_{n=0}^{\infty} \bigcap_{k=n}^{\infty} A_k^c\right) \leq \sum_{n=0}^{\infty} \mathbb{P}\left(\bigcap_{k=n}^{\infty} A_k^c\right) = 0.$$

Un passage au complémentaire conclut alors la preuve. □

Il faut retenir en particulier que sous l'hypothèse d'indépendance des  $A_n$ , seuls deux cas peuvent arriver, suivant la nature de la série  $\sum \mathbb{P}(A_n)$  :

$$\mathbb{P}(\limsup_{n \rightarrow \infty} A_n) = 0 \quad \text{ou} \quad \mathbb{P}(\limsup_{n \rightarrow \infty} A_n) = 1.$$

On parle parfois de *loi du 0-1 de Borel*.

**Exemple 2.16.** Soit  $X_k$  des variables aléatoires de loi de Bernoulli de paramètre respectif  $p_k$ . On considère les événements  $A_k = \{X_k = 1\}$  de sorte que  $p_k = \mathbb{P}(A_k)$ . Le lemme de Borel-Cantelli donne alors

1. Si  $\sum_{k=1}^{\infty} p_k < \infty$ , alors seul un nombre fini des  $X_k$  est non nul presque sûrement.
2. Si  $\sum_{k=1}^{\infty} p_k = \infty$  et si les variables aléatoires  $X_k$  sont indépendantes, alors une infinité des  $X_k$  vaut 1 presque sûrement.

**Mesure uniforme sur  $\mathbb{N}$**  Le lemme de Borel-Cantelli permet de montrer qu'il n'existe pas de mesure uniforme sur  $\mathbb{N}$ , au sens où  $\mathbb{P}(n\mathbb{N}) = 1/n$  pour tout  $n \geq 1$ . Supposons qu'une telle mesure existe. On considère les événements  $A_p = p\mathbb{N}$ , pour  $p$  un nombre premier. Ces événements sont indépendants puisque pour tous  $p_1, \dots, p_k$ , nombres premiers distincts, on a

$$\mathbb{P}(A_{p_1} \cap \dots \cap A_{p_k}) = \mathbb{P}(p_1 \dots p_k \mathbb{N}) = \frac{1}{p_1 \dots p_k} = \prod_{i=1}^k \mathbb{P}(A_{p_i}).$$

Comme  $\sum_{p \text{ premiers}} \frac{1}{p} = \infty$ , on en déduit alors par le deuxième point du lemme de Borel-Cantelli que  $\mathbb{P}$ -presque tout entier  $n$  est dans une infinité d'ensemble  $p\mathbb{N}$  avec  $p$  premier, et donc est multiple d'une infinité de nombres premiers distincts. C'est clairement absurde.

**Motifs dans une suite indépendante et identiquement distribuée** Soit  $X = (X_k)_{k \geq 1}$  une suite de variables aléatoires indépendantes de loi de Bernoulli de paramètre  $1/2$ . Soit  $\epsilon = (\epsilon_1, \dots, \epsilon_n) \in \{0, 1\}^n$  une suite déterministe de longueur  $n$ . Pour  $k \in \mathbb{N}$ , on pose

$$B_k(\epsilon) = \{\omega = (\omega_i)_{i \geq 1} \in \{0, 1\}^{\mathbb{N}^*} : (\omega_{k+1}, \dots, \omega_{k+n}) = (\epsilon_1, \dots, \epsilon_n)\}.$$

L'événement  $A_k(\epsilon) = \{X \in B_k(\epsilon)\}$  est alors réalisé si et seulement si  $X_{k+1} = \epsilon_1, \dots, X_{k+n} = \epsilon_n$ , c'est-à-dire si et seulement si la suite finie  $\epsilon$  apparaît dans la suite  $X$  entre le rang  $k+1$  et le rang  $n+k$ . On pose alors

$$A(\epsilon) = \limsup_{k \rightarrow \infty} A_k(\epsilon) = \limsup_{k \rightarrow \infty} \{X \in B_k(\epsilon)\} = \limsup_{k \rightarrow \infty} \{X_{k+1} = \epsilon_1, \dots, X_{k+n} = \epsilon_n\}$$

qui correspond à l'événement "la suite  $\epsilon$  apparaît une infinité de fois dans le développement dyadique de  $X$ ". Les événements  $A_k(\epsilon) = \{X \in B_k(\epsilon)\}$ ,  $k \in \mathbb{N}$ , ne sont pas indépendants mais les événements  $A_{kn}(\epsilon) = \{X \in B_{kn}(\epsilon)\}$ ,  $k \in \mathbb{N}$ , sont mesurables respectivement par rapport aux tribus  $\sigma(X_{kn+1}, \dots, X_{(k+1)n})$  qui sont indépendantes d'après le lemme de regroupement par paquets. Par ailleurs, l'indépendance des  $X_i$  donne

$$\mathbb{P}(A_{kn}(\epsilon)) = \mathbb{P}(X_{kn+1} = \epsilon_1, \dots, X_{(k+1)n} = \epsilon_n) = \mathbb{P}(X_{kn+1} = \epsilon_1) \cdots \mathbb{P}(X_{(k+1)n} = \epsilon_n) = \frac{1}{2^n}.$$

Donc la série  $\sum_k \mathbb{P}(A_{kn}(\epsilon))$  diverge (attention, on somme sur  $k$ ). On en déduit, par le lemme de Borel-Cantelli, que

$$\mathbb{P}(\limsup_{k \rightarrow \infty} A_{kn}(\epsilon)) = 1,$$

puis que

$$\mathbb{P}(A(\epsilon)) = \mathbb{P}(\limsup_{k \rightarrow \infty} A_k(\epsilon)) \geq \mathbb{P}(\limsup_{k \rightarrow \infty} A_{kn}(\epsilon)) = 1.$$

Cela signifie que presque sûrement la suite finie  $\epsilon$  apparaît une infinité de fois dans la suite  $X = (X_k)_{k \geq 1}$ .

Comme il y a un nombre dénombrable de suite finie de 0 et de 1, l'intersection des  $B_\epsilon$  sur toutes les suites finies  $\epsilon$  est encore de probabilité 1. Autrement dit, avec probabilité 1, toute suite finie apparaît une infinité de fois dans une suite  $X = (X_k)_{k \geq 1}$  constituées de variables aléatoires de Bernoulli indépendantes.

La suite  $X$  étant constituée de 0 et de 1, on peut la voir comme un texte infini écrit en binaire. On vient donc de montrer que si on tape aléatoirement en binaire sur une machine, alors tout les romans déjà écrits, et ceux qu'ils restent à écrire, vont apparaître une infinité de fois p.s. De manière équivalente, si on écrit un réel  $u \in [0, 1[$  en binaire, alors son développement dyadique propre contient toutes les suites finies une infinité de fois p.s.<sup>1</sup> Bref, si vous tirez un nombre au hasard dans  $[0, 1[$ , vous obtenez tous les romans du monde...

1. Rappelons que si  $u \in [0, 1[$ , alors son développement dyadique est la suite  $(x_k)_{k \geq 1}$  non stationnaire à 1 telle que  $u = \sum_{k=1}^{\infty} 2^{-k} x_k$ . Cette suite vérifie  $x_k = \lfloor 2^k u \rfloor - 2 \lfloor 2^{k-1} u \rfloor \in \{0, 1\}$ . On peut montrer que si  $U$  suit une loi uniforme sur  $[0, 1[$ , alors son développement dyadique propre  $X = (X_k)_{k \geq 1}$  est une suite de variables aléatoires indépendantes de loi de Bernoulli de paramètre  $1/2$

### 3.2 Loi du 0-1 de Kolmogorov

Soit  $(\mathcal{F}_1, \mathcal{F}_2, \dots)$  une suite de sous-tribus de  $\mathcal{F}$ . On pose

$$\overline{\mathcal{F}}_n = \sigma(\mathcal{F}_n, \mathcal{F}_{n+1}, \dots)$$

la sous-tribu engendrée par les  $\mathcal{F}_k$  pour  $k \geq n$ , et

$$\mathcal{F}_\infty = \bigcap_{n \geq 1} \overline{\mathcal{F}}_n = \bigcap_{n \geq 1} \sigma(\mathcal{F}_n, \mathcal{F}_{n+1}, \dots).$$

C'est une sous-tribu de  $\mathcal{F}$  que l'on appelle *tribu terminale (ou asymptotique)* associée à la suite  $(\mathcal{F}_1, \mathcal{F}_2, \dots)$ . Un événement  $A \in \mathcal{F}_\infty$  est appelé *événement de queue*. On remarque que cette tribu ne dépend pas des premières tribus, d'où son nom. Par exemple, si  $\mathcal{F}_n = \sigma(X_n)$  pour une suite  $(X_n)_{n \geq 1}$  de variables aléatoires, alors la tribu asymptotique  $\mathcal{F}_\infty$  ne dépend pas des premiers termes de la suite  $(X_n)$ . Donnons un exemple.

**Exemple 2.17.** Montrons que l'événement  $\{(X_n)_{n \geq 1} \text{ converge}\}$  est dans la tribu asymptotique de  $\mathcal{F}_n = \sigma(X_n)$ . On remarque que la suite  $(X_n)_{n \geq 1}$  converge si et seulement si pour tout  $k \geq 1$  la suite  $(X_n)_{n \geq k}$  converge. Or pour  $k \geq 1$  fixé, la suite  $(X_n)_{n \geq k} = (X_k, X_{k+1}, \dots)$  est mesurable par rapport à la tribu  $\overline{\mathcal{F}}_k$ , et donc

$$\{(X_n)_{n \geq 1} \text{ converge}\} = \{(X_n)_{n \geq k} \text{ converge}\} \in \overline{\mathcal{F}}_k.$$

Ceci étant vrai pour tout  $k \geq 1$  on en déduit que cet événement appartient à  $\mathcal{F}_\infty$ .

Attention, si l'événement  $\{(X_n)_{n \geq 1} \text{ converge}\}$  est dans la tribu asymptotique, l'événement  $\{\lim_{n \rightarrow \infty} S_n \geq 0\}$ , avec  $S_n = X_1 + \dots + X_n$  ne l'est pas car la valeur de la limite de  $(S_n)_{n \geq 1}$  dépend des premiers termes de la suite.

**Proposition 2.18** (Loi du 0-1 de Kolmogorov). *Si les tribus  $\mathcal{F}_1, \mathcal{F}_2, \dots$  sont indépendantes, alors la tribu asymptotique est triviale : pour tout  $A \in \mathcal{F}_\infty$ ,  $\mathbb{P}(A) = 0$  ou  $\mathbb{P}(A) = 1$ .*

*Démonstration.* Par le lemme de regroupement par paquets, la tribu  $\overline{\mathcal{F}}_{n+1}$  est indépendante de  $\mathcal{F}_i$  pour tout  $i \leq n$  et tout  $n \geq 1$ . L'inclusion  $\mathcal{F}_\infty \subset \overline{\mathcal{F}}_{n+1}$  implique alors que  $\mathcal{F}_\infty$  est indépendante de  $\mathcal{F}_i$  pour tout  $i \geq 1$ . On en déduit que  $\mathcal{F}_\infty$  est indépendante de  $\sigma(\mathcal{F}_1, \mathcal{F}_2, \dots) = \overline{\mathcal{F}}_1$  en utilisant encore une fois le lemme de regroupement par paquets. Mais comme  $\mathcal{F}_\infty \subset \overline{\mathcal{F}}_1$ , cela signifie que  $\mathcal{F}_\infty$  est indépendante d'elle-même. Ainsi, pour tout  $A \in \mathcal{F}_\infty$ ,  $\mathbb{P}(A) = \mathbb{P}(A \cap A) = \mathbb{P}(A)^2$ . On obtient alors  $\mathbb{P}(A) \in \{0, 1\}$  ce qui conclut la preuve.  $\square$

On applique souvent la loi du 0-1 au contexte suivant. On a une suite de variables aléatoires indépendantes  $X_1, X_2, \dots$  et on pose  $\mathcal{F}_n = \sigma(X_n)$  qui sont alors des sous-tribus indépendantes de  $\mathcal{F}$ .

**Exemple 2.19.** Soit  $(X_n)_{n \geq 1}$  une suite de variables aléatoires réelles indépendantes. On pose  $S_n = X_1 + \dots + X_n$  pour tout  $n \geq 1$ . Montrons alors que les événements

$$\left\{ \limsup_{n \rightarrow \infty} S_n = \infty \right\} \quad \text{et} \quad \left\{ \limsup_{n \rightarrow \infty} S_n = -\infty \right\}$$

sont de probabilité 0 ou 1.

On montre ce résultat pour la limsup et en changeant les  $X_n$  en leur opposée on obtient alors le résultat pour la liminf. Pour  $k \geq 1$  fixé, dire que  $\limsup_{n \rightarrow \infty} X_1 + \cdots + X_n = \infty$  revient à dire que  $\limsup_{n \rightarrow \infty} X_k + \cdots + X_n = \infty$  pour tout  $k \geq 1$  (les premiers termes n'ont pas d'influence sur le caractère infini ou non de la limsup). Or cette dernière quantité ne dépend que des variables aléatoires  $X_k, X_{k+1}, \dots, X_n$  donc est mesurable par rapport à  $\bar{\mathcal{F}}_k$ , avec  $\mathcal{F}_n = \sigma(X_n)$ . Ceci étant valable pour tout  $k \geq 1$ , on en déduit que l'événement  $\limsup_{n \rightarrow \infty} S_n = \infty$  est dans la tribu asymptotique  $\mathcal{F}_\infty$ . Les tribus  $\mathcal{F}_n = \sigma(X_n)$  étant indépendantes par hypothèse sur les  $X_n$ , la loi du 0-1 de Kolmogorov donne le résultat.

## 4 Appendice au Chapitre 2 : démonstrations techniques

### 4.1 Preuve du Lemme 2.9

On fixe  $C_2 \in \mathcal{C}_2, \dots, C_n \in \mathcal{C}_n$  et on considère

$$\mathcal{M}_1 = \{A \in \mathcal{F}_1 : \mathbb{P}(A \cap C_2 \cap \dots \cap C_n) = \mathbb{P}(A)\mathbb{P}(C_2) \cdots \mathbb{P}(C_n)\}.$$

Alors  $\mathcal{M}_1$  contient  $\mathcal{C}_1$  par hypothèse et donc en particulier  $\Omega \in \mathcal{M}_1$ . On vérifie alors que  $\mathcal{M}_1$  est une classe monotone. Le lemme de classe monotone entraîne alors que  $\mathcal{M}_1$  contient  $\sigma(\mathcal{C}_1) = \mathcal{F}_1$  et on a montré que pour tout  $A \in \mathcal{F}_1$  et  $C_2 \in \mathcal{C}_2, \dots, C_n \in \mathcal{C}_n$ ,

$$\mathbb{P}(A \cap C_2 \cap \dots \cap C_n) = \mathbb{P}(A)\mathbb{P}(C_2) \cdots \mathbb{P}(C_n).$$

On pose alors, pour  $A_1 \in \mathcal{F}_1, C_3 \in \mathcal{C}_3, \dots, C_n \in \mathcal{C}_n$ ,

$$\mathcal{M}_2 = \{A \in \mathcal{F}_2 : \mathbb{P}(A_1 \cap A \cap C_3 \cap \dots \cap C_n) = \mathbb{P}(A_1)\mathbb{P}(A)\mathbb{P}(C_3) \cdots \mathbb{P}(C_n)\}.$$

Les mêmes arguments que précédemment entraînent que  $\mathcal{M}_2$  contient  $\sigma(\mathcal{C}_2) = \mathcal{F}_2$ . On continue ainsi de suite par récurrence.

### 4.2 Critère d'indépendance via les fonctions caractéristiques

Rappelons le résultat. Les variables aléatoires  $X_1, \dots, X_d$  à valeurs dans  $\mathbb{R}^d$  sont indépendantes si et seulement si pour tout  $t = (t_1, \dots, t_d) \in \mathbb{R}^d$ ,

$$\varphi_{(X_1, \dots, X_d)}(t) = \varphi_{X_1}(t_1) \cdots \varphi_{X_d}(t_d),$$

c'est-à-dire

$$\mathbb{E}[e^{i\langle t, X \rangle}] = \mathbb{E}[e^{it_1 X_1}] \cdots \mathbb{E}[e^{it_d X_d}],$$

où  $X = (X_1, \dots, X_d)$ . On peut également écrire cette égalité avec la transformée de Fourier d'une mesure :

$$\widehat{\mathbb{P}}_X(\xi) = \widehat{\mathbb{P}}_{X_1}(\xi_1) \cdots \widehat{\mathbb{P}}_{X_d}(\xi_d),$$

où  $\widehat{\mathbb{P}}_X$  représente la transformée de Fourier de la mesure  $\mathbb{P}_X$ .

*Démonstration.* Le sens direct est une conséquence de la proposition 2.13 avec les fonctions  $h_i(x) = e^{it_i x}$ , pour  $t_i$  fixé.

Réciproquement, supposons que  $\widehat{\mathbb{P}}_X = \widehat{\mathbb{P}}_{X_1} \cdots \widehat{\mathbb{P}}_{X_d}$ . D'après la Proposition A.18, on sait que  $\widehat{\mathbb{P}}_{X_1} \cdots \widehat{\mathbb{P}}_{X_d}$  est également la transformée de Fourier de la mesure  $\mathbb{P}_{X_1} \otimes \cdots \otimes \mathbb{P}_{X_d}$ . Par injectivité de la transformée de Fourier d'une mesure (cf Corollaire A.20), on en déduit que

$$\mathbb{P}_X = \mathbb{P}_{X_1} \otimes \cdots \otimes \mathbb{P}_{X_d},$$

ce qui prouve l'indépendance de  $X_i$ . □





# Chapitre 3

## Loi des grands nombres

Dans tout le chapitre, on considère une suite  $(X_n)_{n \geq 1}$  de variables aléatoires et  $X$  une variable aléatoire à valeurs dans  $\mathbb{R}$  ou  $\mathbb{C}$ . On va définir différentes notions de convergence en regardant ces variables aléatoires comme des fonctions de  $\Omega$  dans  $\mathbb{R}$  ou  $\mathbb{C}$ . En particulier, on suppose que toutes les variables aléatoires considérées sont définies sur un même espace de probabilité  $(\Omega, \mathcal{F}, \mathbb{P})$ .

### 1 Différentes notions de convergence

#### 1.1 Convergence p.s. et convergence en probabilité

**Définition 3.1** (Convergence presque sûre). *On dit que  $(X_n)$  converge presque sûrement (p.s. en abrégé) vers  $X$ , et on note  $X_n \xrightarrow[n \rightarrow \infty]{p.s.} X$ , si l'événement  $\{X_n \rightarrow X\}$  est presque sûr, c'est-à-dire*

$$\mathbb{P}\left(\left\{\omega \in \Omega : X_n(\omega) \xrightarrow[n \rightarrow \infty]{} X(\omega)\right\}\right) = 1.$$

Notons que cela revient à dire que l'événement  $\{\limsup_{n \rightarrow \infty} |X_n - X| = 0\}$  est presque sûr. Le lemme de Borel-Cantelli sera donc utile pour établir une convergence presque sûre (attention toutefois à la différence entre la limsup d'une suite de variables aléatoires et la limsup d'une suite d'événements).

La convergence presque sûre correspond à la convergence ponctuelle de la suite de fonctions mesurables  $(X_n)$  vers  $X$  (à un ensemble de mesure nulle près). Donc si  $(X_n)$  converge p.s. vers  $X$  et si  $h$  est une fonction continue, alors  $(h(X_n))$  converge p.s. vers  $h(X)$ . En particulier, si  $(X_n)$  et  $(Y_n)$  convergent p.s. vers  $X$  et  $Y$  respectivement, alors  $aX_n + bY_n$  converge p.s. vers  $aX + bY$  et  $X_n Y_n$  converge p.s. vers  $XY$ .

La convergence p.s. est assez intuitive mais pas forcément facile à manipuler en pratique. On propose donc un mode de convergence plus souple.

**Définition 3.2** (Convergence en probabilité). *On dit que  $(X_n)$  converge en probabilité vers  $X$ ,*

et on note  $X_n \xrightarrow[n \rightarrow \infty]{\mathbb{P}} X$ , si pour tout  $\epsilon > 0$ ,

$$\mathbb{P}(|X_n - X| \geq \epsilon) \xrightarrow[n \rightarrow \infty]{} 0.$$

Cela signifie qu'on s'autorise à ce que  $X_n$  s'écarte de  $X$  de plus de  $\epsilon$ , mais la probabilité que cet événement se produise tend vers 0.

**Exemple 3.3.** Soit  $X_n \sim \mathcal{E}(n)$ . Pour  $\epsilon > 0$  fixé, on a

$$\mathbb{P}(|X_n| \geq \epsilon) = \mathbb{P}(X_n \geq \epsilon) = e^{-n\epsilon} \xrightarrow[n \rightarrow \infty]{} 0.$$

Donc  $(X_n)$  converge en probabilité vers 0. On peut montrer que la convergence vers 0 a aussi lieu au sens presque sûr, mais cela demande plus de travail (cf TD).

**Proposition 3.4.** Si  $(X_n)$  converge presque sûrement vers  $X$ , alors  $(X_n)$  converge en probabilité vers  $X$ .

*Démonstration.* Supposons que  $(X_n)$  converge presque sûrement vers  $X$  et fixons  $\epsilon > 0$ . On écrit alors

$$\mathbb{P}(|X_n - X| \geq \epsilon) = \mathbb{E}[\mathbf{1}_{\{|X_n - X| \geq \epsilon\}}].$$

L'indicatrice tend vers zéro presque sûrement par hypothèse et est bornée par 1 qui est intégrable pour  $\mathbb{P}$ . Le théorème de convergence dominée implique alors que

$$\mathbb{E}[\mathbf{1}_{\{|X_n - X| \geq \epsilon\}}] \xrightarrow[n \rightarrow \infty]{} 0,$$

ce qui prouve la convergence en probabilité de  $(X_n)$  vers  $X$ . □

**Exemple 3.5.** Soit  $(X_n)$  la suite des variables aléatoires dont la loi est donnée par

$$\mathbb{P}(X_n = n) = \frac{1}{n} = 1 - \mathbb{P}(X_n = 0).$$

Alors pour tout  $\epsilon > 0$ ,  $\mathbb{P}(|X_n| \geq \epsilon) \leq \mathbb{P}(X_n = n) = \frac{1}{n} \rightarrow 0$ , donc  $(X_n)$  converge vers 0 en probabilité.

Or, la série de terme général  $\mathbb{P}(X_n = n)$  diverge donc par indépendance des  $X_n$  le lemme de Borel-Cantelli implique que  $\mathbb{P}(\limsup_{n \rightarrow \infty} \{X_n = n\}) = 1$ . Ainsi, presque sûrement  $X_n = n$  infiniment souvent, donc presque sûrement  $\limsup_{n \rightarrow \infty} X_n = \infty$ . Donc  $(X_n)$  ne converge pas presque sûrement.

**Remarque 3.6.** En pratique, la convergence en probabilité peut être établie via l'inégalité de Markov. Pour  $p \in \mathbb{N}^*$ , on écrit

$$\mathbb{P}(|X_n - X| \geq \epsilon) \leq \frac{\mathbb{E}[|X_n - X|^p]}{\epsilon^p},$$

et on essaye de montrer que cette espérance converge vers zéro. On verra que cette dernière quantité est liée à la convergence dans l'espace  $L^p(\Omega, \mathcal{F}, \mathbb{P})$ .

On étudie souvent la convergence d'une somme  $S_n = X_1 + \dots + X_n$  où les  $X_k$  sont des variables aléatoires indépendantes et identiquement distribuées (iid en abrégé), c'est-à-dire indépendantes et de même loi. On se demande si, après normalisation, la somme  $S_n$  converge. Le résultat suivant traite le cas de variables aléatoires de carré intégrable. Le théorème fondamental plus général est la loi forte des grands nombres qui sera énoncée plus loin.

**Proposition 3.7** (Loi faible des grands nombres). *Soit  $(X_n)$  une suite de variables aléatoires dans  $L^2(\Omega, \mathcal{F}, \mathbb{P})$ . On note  $\mu$  leur espérance. On pose  $S_n = X_1 + \dots + X_n$  et on définit la moyenne empirique  $\bar{X}_n$  par*

$$\bar{X}_n = \frac{S_n}{n} = \frac{X_1 + \dots + X_n}{n}.$$

Alors  $\bar{X}_n \xrightarrow[n \rightarrow \infty]{\mathbb{P}} \mu$ .

Ce résultat est assez intuitif : si on répète une expérience aléatoire de manière indépendante et dans les mêmes conditions (même loi pour les  $X_i$ ), alors la moyenne empirique se rapproche de l'espérance qui est la moyenne théorique.

*Démonstration.* Les variables aléatoires  $X_k$  étant dans  $L^2(\Omega, \mathcal{F}, \mathbb{P})$ , leur moyenne empirique  $\bar{X}_n$  l'est aussi et vérifie

$$\mathbb{E}[\bar{X}_n] = \frac{1}{n} \mathbb{E}[X_1 + \dots + X_n] = \mu \quad \text{et} \quad \text{Var}(\bar{X}_n) = \frac{1}{n^2} \text{Var}(X_1 + \dots + X_n) = \frac{\text{Var}(X_1)}{n},$$

où on a utilisé la linéarité de l'espérance dans la première égalité et la linéarité de la variance pour des variables aléatoires indépendantes dans la deuxième égalité.

Soit alors  $\epsilon > 0$ . L'inégalité de Bienaymé-Tchebychev donne

$$\mathbb{P}(|\bar{X}_n - \mu| \geq \epsilon) = \mathbb{P}(|\bar{X}_n - \mathbb{E}[\bar{X}_n]| \geq \epsilon) \leq \frac{\text{Var}(\bar{X}_n)}{\epsilon^2} = \frac{\text{Var}(X_1)}{n\epsilon^2} \xrightarrow[n \rightarrow \infty]{} 0,$$

ce qui prouve que  $\bar{X}_n \xrightarrow[n \rightarrow \infty]{\mathbb{P}} \mu$ . □

**Exemple 3.8.** Si les  $X_n$  sont indépendantes de loi de Bernoulli de paramètre  $p$ , alors la loi faible des grands nombres implique que

$$\frac{X_1 + \dots + X_n}{n} \xrightarrow[n \rightarrow \infty]{\mathbb{P}} p.$$

Par exemple, si vous lancez un grand nombre de fois une pièce équilibrée, la fréquence de Pile obtenu se rapproche de  $1/2$  quand le nombre de lancers augmente. C'est parfois ce cas particulier avec des variables aléatoires de loi de Bernoulli qui est appelé loi faible des grands nombres.

L'objectif de la fin du chapitre consiste à montrer que cette convergence a en fait lieu presque sûrement et sous l'hypothèse plus faible d'intégrabilité des  $X_k$ . La preuve nécessite un peu d'efforts. On va d'abord introduire une troisième notion de convergence liée aux espaces  $L^p(\Omega, \mathcal{F}, \mathbb{P})$ .

## 1.2 Convergence $L^p$

**Définition 3.9** (Convergence dans  $L^p$ ). Soit  $p \in [1, \infty]$ . On dit que  $(X_n)$  converge vers  $X$  dans  $L^p(\Omega, \mathcal{F}, \mathbb{P})$ , et on note  $X_n \xrightarrow[n \rightarrow \infty]{L^p} X$ , si

$$\mathbb{E}[|X_n - X|^p] \xrightarrow[n \rightarrow \infty]{} 0.$$

Il s'agit donc de la convergence usuelle dans l'espace de Banach  $L^p = L^p(\Omega, \mathcal{F}, \mathbb{P})$  muni de la norme  $\|\cdot\|_p$ .

On rappelle que dans le cas d'une mesure finie, les espaces  $L^p$  sont décroissants : si  $p \leq q$ , alors  $L^q \subset L^p$ . Dans le cas d'une mesure de probabilité, on a l'inégalité  $\|\cdot\|_p \leq \|\cdot\|_q$ , donc la convergence  $L^q$  implique la convergence  $L^p$ .

**Proposition 3.10.** Si  $(X_n)$  converge vers  $X$  dans  $L^p$ , alors la convergence a lieu en probabilité.

*Démonstration.* C'est une conséquence de l'inégalité de Markov : pour  $\epsilon > 0$ ,

$$\mathbb{P}(|X_n - X| \geq \epsilon) \leq \frac{\mathbb{E}[|X_n - X|^p]}{\epsilon^p},$$

et cette quantité tend vers 0 sous l'hypothèse de convergence dans  $L^p$ .  $\square$

**Exemple 3.11.** On reprend l'exemple 3.5 en le modifiant légèrement. On se fixe  $p \in [1, \infty[$  et on considère la suite de variables aléatoires  $(X_n)$  de loi

$$\mathbb{P}(X_n = n^{1/p}) = \frac{1}{n} = 1 - \mathbb{P}(X_n = 0).$$

Alors

$$\mathbb{E}[|X_n|^p] = \frac{1}{n} \times (n^{1/p})^p = 1,$$

donc la suite  $(X_n)$  ne converge pas dans  $L^p$ . On remarque cependant qu'elle converge vers 0 dans  $L^q$  pour tout  $q < p$ .

Si  $(X_n)$  converge dans  $L^p$ , alors il n'y a pas forcément convergence presque sûre (sauf si  $p = \infty$ ). On a cependant le résultat suivant, conséquence du théorème de Riesz-Fisher.

**Proposition 3.12.** Si  $(X_n)$  converge dans  $L^p$ , alors il existe une suite extraite  $(n_k)$  telle que  $X_{n_k} \xrightarrow[k \rightarrow \infty]{p.s.} X$ .

Réciproquement, si  $(X_n)$  converge presque sûrement vers  $X$ , alors la convergence n'a pas forcément lieu dans  $L^p$ . Toutefois, si  $|X_n| \leq Y$  avec  $Y \in L^p$ , alors le théorème de convergence dominée donne la convergence dans  $L^p$ .

**Remarque 3.13.** On a vu que les convergences p.s. et  $L^p$  impliquaient la convergence en probabilité et que les réciproques étaient fausses. Toutefois, si  $(X_n)$  converge en probabilité vers  $X$ , alors

1. il existe une suite extraite  $(n_k)$  telle que  $X_{n_k}$  converge presque sûrement ;
2. si de plus la suite est bornée dans  $L^q$ , c'est-à-dire s'il existe  $M > 0$  tel que pour tout  $n \in \mathbb{N}$ ,  $\mathbb{E}[|X_n|^q] \leq M$ , alors  $(X_n)$  converge dans  $L^p$  pour tout  $p \in [1, q[$ .

Ces résultats ne sont pas forcément évidents à démontrer mais font appel à des techniques standard d'intégration. Les preuves sont donnés en fin de chapitre. En résumé, on a le diagramme d'implication illustré en Figure 3.1.

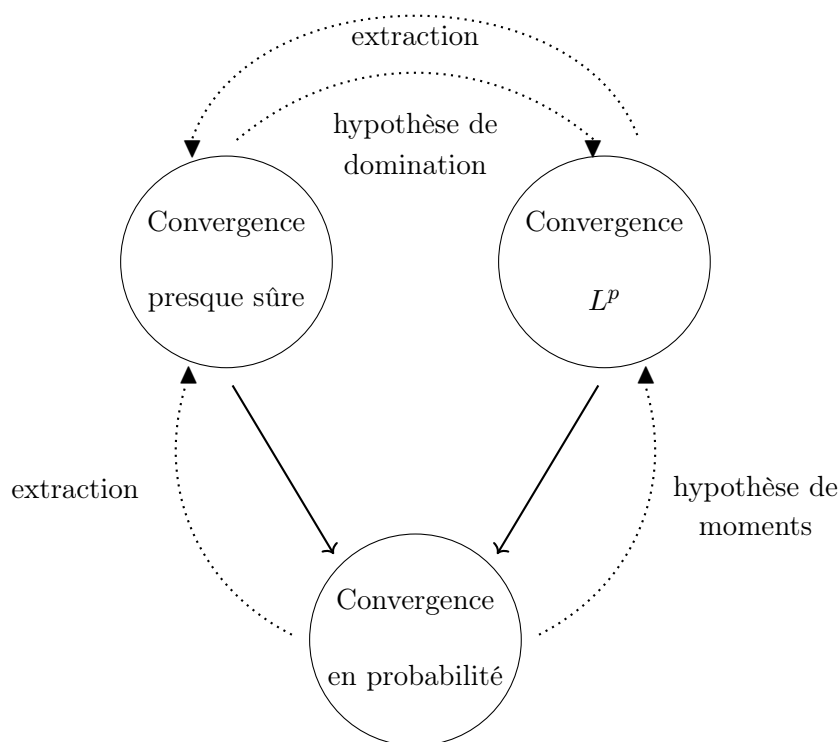


FIGURE 3.1 – Diagramme d'implication.

## 2 Loi forte des grands nombres

On a déjà établi plus haut la loi faible des grands nombres qui donne la convergence en probabilité de la moyenne empirique  $\bar{X}_n$  vers l'espérance  $\mathbb{E}[X_1]$  pour des variables aléatoires  $X_k$  indépendante, de même loi et dans  $L^2$ . La loi forte des grands nombres donne la convergence presque sûre sous l'hypothèse plus faible d'intégrabilité des  $X_k$ .

## 2.1 Le résultat

**Théorème 3.14** (Loi forte des grands nombres). *Soit  $X_1, X_2, \dots$  une suite de variables aléatoires réelles indépendantes et de même loi dans  $L^1(\Omega, \mathcal{F}, \mathbb{P})$ . Alors*

$$\bar{X}_n = \frac{X_1 + \dots + X_n}{n} \xrightarrow[n \rightarrow \infty]{p.s.} \mathbb{E}[X_1]. \quad (3.1)$$

Remarquons que ce théorème est également vrai pour des vecteurs aléatoires (c'est-à-dire des variables aléatoires à valeurs dans  $\mathbb{R}^d$ ) : il suffit d'appliquer le théorème précédent coordonnée par coordonnée.

L'hypothèse d'intégrabilité est nécessaire puisqu'elle assure que la limite  $\mathbb{E}[X_1]$  existe bien et est finie. Pour des variables aléatoires positives d'espérance infinie, le théorème reste vrai au sens où

$$\frac{X_1 + \dots + X_n}{n} \xrightarrow[n \rightarrow \infty]{p.s.} \infty,$$

Il suffit d'appliquer la loi forte des grands nombres aux variables aléatoires bornées  $\min(X_n, K)$ , pour  $K > 0$ , puis d'utiliser le théorème de convergence dominée quand  $K \rightarrow \infty$ .

Notons que la convergence (3.1) a également lieu dans  $L^1$ . En effet, on tronque les  $X_i$  en définissant

$$Y_i = X_i \mathbf{1}_{\{|X_i| \leq K\}} \quad \text{et} \quad Z_i = X_i \mathbf{1}_{\{|X_i| > K\}},$$

pour un réel  $K > 0$  fixé, de sorte que  $X_i = Y_i + Z_i$ . Les variables aléatoires  $Y_i$  sont alors indépendantes, de même loi, et bornées, tandis que les variables aléatoires  $Z_i$  sont dans  $L^1$ . Il s'ensuit :

$$\begin{aligned} \mathbb{E} \left[ \left| \frac{X_1 + \dots + X_n}{n} - \mathbb{E}[X_1] \right| \right] &\leq \mathbb{E} \left[ \left| \frac{Y_1 + \dots + Y_n}{n} - \mathbb{E}[Y_1] \right| \right] + \mathbb{E} \left[ \left| \frac{Z_1 + \dots + Z_n}{n} - \mathbb{E}[Z_1] \right| \right] \\ &\leq \mathbb{E} \left[ \left| \frac{Y_1 + \dots + Y_n}{n} - \mathbb{E}[Y_1] \right| \right] + 2\mathbb{E}[|Z_1|]. \end{aligned}$$

Pour le premier terme, on écrit

$$\begin{aligned} \mathbb{E} \left[ \left| \frac{Y_1 + \dots + Y_n}{n} - \mathbb{E}[Y_1] \right|^2 \right] &= \mathbb{E} \left[ \left| \frac{Y_1 - \mathbb{E}[Y_1] + \dots + Y_n - \mathbb{E}[Y_n]}{n} \right|^2 \right] \\ &= \text{Var} \left( \frac{Y_1 + \dots + Y_n}{n} \right) = \frac{\text{Var}(Y_1)}{n} \end{aligned}$$

et ce terme tend vers 0 quand  $n \rightarrow \infty$ . Ainsi,  $((Y_1 + \dots + Y_n)/n)$  converge vers  $\mathbb{E}[Y_1]$  dans  $L^2$  donc dans  $L^1$  et donc

$$\limsup_{n \rightarrow \infty} \mathbb{E} \left[ \left| \frac{X_1 + \dots + X_n}{n} - \mathbb{E}[X_1] \right| \right] \leq 2\mathbb{E}[|Z_1|],$$

ceci étant vrai pour tout  $K > 0$ . Enfin, la variable aléatoire  $Z_1 = X_1 \mathbf{1}_{\{|X_1| > K\}}$  converge vers 0 quand  $K \rightarrow \infty$  et est bornée par  $K$ , donc le théorème de convergence dominée assure que  $2\mathbb{E}[|Z_1|] \xrightarrow[K \rightarrow \infty]{} 0$ , ce qui conclut la preuve.

Notons cependant que du point de vue probabiliste, c'est la convergence p.s. qui a le plus de signification.

*Démonstration.* Il existe plusieurs preuves de la loi des grands nombres, toutes nécessitent un peu de travail.

On pose  $S_n = X_1 + \cdots + X_n$  et  $S_0 = 0$ , et on fixe  $a > \mathbb{E}[X_1]$ . La suite  $(M_n)$  définie par

$$M_n = \max_{0 \leq k \leq n} (S_k - ka)$$

est positive et croissante donc converge presque sûrement vers  $M = \sup_{n \geq 0} (S_n - na) \in [0, \infty]$ .

**Étape 1** Montrons que  $\mathbb{P}(M = \infty) \in \{0, 1\}$ .

Pour tout  $k \geq 0$ , on part de l'égalité

$$\{M = \infty\} = \{\sup_{n \geq 0} (S_n - na) = \infty\} = \{\sup_{n \geq k} (X_{k+1} + \cdots + X_n - na) = \infty\},$$

qui implique que cet événement est dans la tribu  $\overline{\mathcal{F}}_k$ , avec  $\mathcal{F}_k = \sigma(X_k)$ . Ceci étant vrai pour tout  $k \geq 0$ , on en déduit que  $\{M = \infty\}$  est dans la tribu asymptotique  $\mathcal{F}_\infty$ , et comme les tribus  $\mathcal{F}_k = \sigma(X_k)$  sont indépendantes, la loi du 0-1 de Kolmogorov implique que  $\mathbb{P}(M = \infty) \in \{0, 1\}$ .

**Étape 2** Montrons que  $\mathbb{P}(M = \infty) = 0$ .

On pose

$$S'_n = X_2 + X_3 + \cdots + X_{n+1}, \quad n \geq 1,$$

et  $S'_0 = 0$ , et de même

$$M'_n = \max_{0 \leq k \leq n} (S'_k - ka) \quad \text{et} \quad M' = \sup_{n \geq 0} (S'_n - na) \in [0, \infty].$$

La suite  $(S_n)_{n \geq 0}$  est alors de même loi que  $(S'_n)_{n \geq 0}$  et est indépendante de  $X_1$  par le lemme de regroupement par paquets. Par suite,  $M_n$  et  $M'_n$  ont également la même loi. Donc pour tout  $x \in \mathbb{R}$ ,

$$\mathbb{P}(M \leq x) = \lim_{n \rightarrow \infty} \mathbb{P}(M_n \leq x) = \lim_{n \rightarrow \infty} \mathbb{P}(M'_n \leq x) = \mathbb{P}(M' \leq x),$$

par la propriété de décroissance monotone appliquée aux événements  $\{M_n \leq x\}$ . Ainsi,  $M$  et  $M'$  ont la même fonction de répartition donc la même loi.

Enfin, un calcul direct donne

$$M_{n+1} = \max(0, \max_{1 \leq k \leq n+1} (S_k - ka)) = \max(0, M'_n + X_1 - a) = M'_n - \min(a - X_1, M'_n).$$

Comme les variables aléatoires  $M_n$  et  $M'_n$  sont intégrables et de même loi, on obtient

$$\mathbb{E}[\min(a - X_1, M'_n)] = \mathbb{E}[M'_n] - \mathbb{E}[M_{n+1}] = \mathbb{E}[M_n] - \mathbb{E}[M_{n+1}] \leq 0,$$

par croissance de la suite  $(M_n)_{n \geq 0}$ . Le terme  $\min(a - X_1, M'_n)$  converge vers  $\min(a - X_1, M')$  et est majoré par  $|a - X_1|$  (on rappelle que  $M'_n \geq 0$ ), donc le théorème de convergence dominée donne l'inégalité  $\mathbb{E}[\min(a - X_1, M')] \leq 0$ .

Supposons maintenant que  $\mathbb{P}(M = \infty) = 1$ . Alors  $\mathbb{P}(M' = \infty) = 1$  car  $M$  et  $M'$  ont la même loi, et donc  $\min(a - X_1, M'_n) = a - X_1$  presque sûrement. L'inégalité précédente devient alors  $\mathbb{E}[a - X_1] \leq 0$ , ce qui est absurde car on a choisi  $a > \mathbb{E}[X_1]$ . Ainsi  $\mathbb{P}(M = \infty) = 0$ .

**Étape 3** Conclusion.

Par définition de  $M$ , on a l'inégalité  $S_n \leq na + M$  pour tout  $n \geq 0$ , donc comme  $M$  est fini presque sûrement on en déduit que

$$\mathbb{P}\left(\limsup_{n \rightarrow \infty} \frac{S_n}{n} \leq a\right) = 1.$$

On choisit alors  $a$  rationnel qui décroît vers  $\mathbb{E}[X_1]$ . Par décroissance monotone, on obtient

$$\begin{aligned} \mathbb{P}\left(\limsup_{n \rightarrow \infty} \frac{S_n}{n} \leq \mathbb{E}[X_1]\right) &= \mathbb{P}\left(\bigcap_{\substack{a \downarrow \mathbb{E}[X_1] \\ a \in \mathbb{Q}}} \left\{\limsup_{n \rightarrow \infty} \frac{S_n}{n} \leq a\right\}\right) \\ &= \lim_{\substack{a \downarrow \mathbb{E}[X_1] \\ a \in \mathbb{Q}}} \mathbb{P}\left(\limsup_{n \rightarrow \infty} \frac{S_n}{n} \leq a\right) = \lim_{\substack{a \downarrow \mathbb{E}[X_1] \\ a \in \mathbb{Q}}} 1 = 1. \end{aligned}$$

En remplaçant les  $X_n$  par  $-X_n$ , on obtient l'inégalité inverse

$$\mathbb{P}\left(\liminf_{n \rightarrow \infty} \frac{S_n}{n} \geq \mathbb{E}[X_1]\right) = 1.$$

Ainsi,

$$\mathbb{P}\left(\lim_{n \rightarrow \infty} \frac{S_n}{n} = \mathbb{E}[X_1]\right) = \mathbb{P}\left(\limsup_{n \rightarrow \infty} \frac{S_n}{n} \leq \mathbb{E}[X_1], \liminf_{n \rightarrow \infty} \frac{S_n}{n} \geq \mathbb{E}[X_1]\right) = 1,$$

comme intersection de deux événements de probabilité 1. □

## 2.2 Quelques applications

**Marche aléatoire non centrée** Soit  $X_1, X_2, \dots$  une suite de variables aléatoires iid intégrables. Posons  $S_n = X_1 + \dots + X_n$ . On appelle parfois la suite  $(S_n)$  une *marche aléatoire* et les  $X_k$  sont appelés les pas de cette marche aléatoire<sup>1</sup>. Supposons que  $\mathbb{E}[X_1] \neq 0$ . La loi forte des grands nombres donne alors que

$$\mathbb{P}(S_n/n \rightarrow \mathbb{E}[X_1]) = 1,$$

ce qui implique que

$$1 = \mathbb{P}(S_n/n \rightarrow \mathbb{E}[X_1]) \leq \mathbb{P}(|S_n| \rightarrow \infty).$$

Ainsi,  $|S_n| \rightarrow \infty$  presque sûrement.

---

1. L'idée de marche aléatoire (mais pas le terme) a été introduite par le biostatisticien Karl Pearson en 1905 lors de l'étude des migrations : chaque  $X_k$  représente le déplacement d'un individu dans une direction donnée lors d'une unité de temps et  $S_n = X_1 + \dots + X_n$  correspond alors à l'endroit où se trouve l'individu au bout de  $n$  unités de temps.



**Approximation d'intégrales par la méthode de Monte-Carlo** Soit  $f : \mathbb{R}^d \rightarrow \mathbb{R}$  une fonction intégrable. Supposons que l'on veuille calculer une intégrale de la forme

$$I(f) = \int_{[0,1]^d} f(t) \, dt.$$

On considère des variables aléatoires  $U_1, U_2, \dots$  iid de loi uniforme sur le cube  $[0, 1]^d$ . Alors les variables aléatoires  $f(X_1), f(X_2), \dots$  sont iid, intégrables et d'espérance  $I(f)$ . La loi forte des grands nombres implique alors que

$$\frac{f(U_1) + \dots + f(U_n)}{n} \xrightarrow[n \rightarrow \infty]{p.s.} I(f).$$

Ceci donne donc une méthode pour approcher une intégrale sur  $[0, 1]^d$ , à condition de savoir simuler facilement la suite  $U_1, U_2, \dots$ . L'avantage de cette méthode est qu'elle ne suppose aucune régularité sur la fonction  $f$ , contrairement aux méthodes déterministes standards (rectangles, trapèzes, etc.) et qu'elle est valable en dimension quelconque. Cependant la vitesse de convergence vers l'intégrale est faible (de l'ordre de  $\sqrt{n}$  comme on le verra dans le chapitre suivant), donc il faut simuler beaucoup de variables aléatoires  $U_k$  pour obtenir une approximation correcte.

### 3 Appendice au Chapitre 3 : démonstrations techniques

On prouve dans cette section les deux réciproques partielles permettant de remonter de la convergence en probabilité à la convergence p.s. ou  $L^p$  (voir Remarque 3.13).

1. Si  $(X_n)_{n \in \mathbb{N}}$  est une suite de variables aléatoires réelles qui converge en probabilité vers une variable aléatoire réelle  $X$ , alors il existe une suite extraite  $(n_k)_{k \geq 1}$  telle que  $X_{n_k}$  converge presque sûrement vers  $X$ .

*Démonstration.* Soit  $\epsilon \in ]0, 1[$ . On décompose  $\mathbb{E}[|X_n - X| \wedge 1]$  de la manière suivante :

$$\mathbb{E}[|X_n - X| \wedge 1] = \mathbb{E}[ (|X_n - X| \wedge 1) \mathbf{1}_{|X_n - X| \leq \epsilon} ] + \mathbb{E}[ (|X_n - X| \wedge 1) \mathbf{1}_{|X_n - X| > \epsilon} ] .$$

On obtient alors la majoration suivante :

$$\mathbb{E}[|X_n - X| \wedge 1] \leq \mathbb{E}[\epsilon \mathbf{1}_{|X_n - X| \leq \epsilon}] + \mathbb{E}[\mathbf{1}_{|X_n - X| > \epsilon}] = \epsilon + \mathbb{P}(|X_n - X| > \epsilon) .$$

La convergence  $X_n \xrightarrow[n \rightarrow \infty]{\mathbb{P}} X$  donne alors

$$\limsup_{n \rightarrow \infty} \mathbb{E}[|X_n - X| \wedge 1] \leq \epsilon + \limsup_{n \rightarrow \infty} \mathbb{P}(|X_n - X| > \epsilon) = \epsilon .$$

Ceci étant vrai pour tout  $\epsilon \in ]0, 1[$ , on en déduit alors que

$$\lim_{n \rightarrow \infty} \mathbb{E}[|X_n - X| \wedge 1] = 0 .$$

Il existe donc  $n_1$  tel que  $\mathbb{E}[|X_{n_1} - X| \wedge 1] \leq 1/2$ , puis  $n_2 \geq n_1$  tel que  $\mathbb{E}[|X_{n_2} - X| \wedge 1] \leq 1/2^2$ , etc.  $n_k \geq n_{k-1}$  tel que

$$\mathbb{E}[|X_{n_k} - X| \wedge 1] \leq \frac{1}{2^k} .$$

Ainsi la série  $\sum_{k \geq 1} \mathbb{E}[|X_{n_k} - X| \wedge 1] = \mathbb{E}[\sum_{k \geq 1} |X_{n_k} - X| \wedge 1]$  converge, où l'égalité est une conséquence du théorème de Fubini. On en déduit que la série de variables aléatoires  $\sum_{k \geq 1} |X_{n_k} - X| \wedge 1$  converge presque sûrement. En décomposant

$$\sum_{k \geq 1} |X_{n_k} - X| \wedge 1 = \sum_{k \geq 1 : |X_{n_k} - X| < 1} |X_{n_k} - X| + \sum_{k \geq 1 : |X_{n_k} - X| \geq 1} 1 ,$$

on en déduit qu'il ne peut y avoir qu'un nombre fini de  $k$  tel que  $|X_{n_k} - X| \geq 1$ , et donc que  $\sum_{k \geq 1} |X_{n_k} - X|$  converge p.s. Ceci donne bien la convergence p.s. de  $(X_{n_k})_{k \geq 1}$  vers  $X$ .  $\square$

2. Soit  $(X_n)_{n \in \mathbb{N}}$  est une suite de variables aléatoires réelles qui converge en probabilité vers une variable aléatoire réelle  $X$  et  $q \in [1, \infty]$ . S'il existe  $M > 0$  tel que pour tout  $n \in \mathbb{N}$ ,  $\mathbb{E}[|X_n|^q] \leq M$ , alors  $(X_n)$  converge dans  $L^p$  pour tout  $p \in [1, q[$ .

*Démonstration.* D'après le résultat précédent, il existe une sous-suite  $(n_k)_{k \geq 1}$  telle que  $X_{n_k} \xrightarrow[k \rightarrow \infty]{p.s.} X$ . Le lemme de Fatou donne alors

$$\mathbb{E}[|X|^q] \leq \liminf_{n \rightarrow \infty} \mathbb{E}[|X_{n_k}|^q] \leq M,$$

donc  $X \in L^q$ . Les espaces  $L^p$  étant décroissants, on en déduit que  $X \in L^p$  pour tout  $p \in [1, q[$ .

Soit  $\epsilon > 0$ . La même décomposition que dans la preuve précédente donne

$$\mathbb{E}[|X_n - X|^p] = \mathbb{E}[|X_n - X|^p \mathbf{1}_{|X_n - X| \leq \epsilon}] + \mathbb{E}[|X_n - X|^p \mathbf{1}_{|X_n - X| > \epsilon}].$$

Le premier terme est majorée par  $\epsilon^p$  tandis que l'inégalité de Hölder avec les exposants  $q/p$  et  $q/(q-p)$  donne la majoration

$$\mathbb{E}[|X_n - X|^p \mathbf{1}_{|X_n - X| > \epsilon}] \leq \mathbb{E}[|X_n - X|^{p/q}]^p \mathbb{E}[(\mathbf{1}_{|X_n - X| > \epsilon})^{q/(q-p)}]^{(q-p)/q}.$$

En utilisant le fait que  $\mathbb{E}[|X_n - X|^q]$  est borné, disons par une constante  $C > 0$ , on obtient

$$\mathbb{E}[|X_n - X|^p] \leq \epsilon^p + C^{p/q} \mathbb{P}(|X_n - X| > \epsilon)^{(q-p)/q},$$

et l'hypothèse de convergence en probabilité entraîne alors que

$$\limsup_{n \rightarrow \infty} \mathbb{E}[|X_n - X|^p] \leq \epsilon^p.$$

Ceci étant vrai pour tout  $\epsilon > 0$ , on obtient bien  $\lim_{n \rightarrow \infty} \mathbb{E}[|X_n - X|^p] = 0$ , ce qui prouve la convergence  $X_n \xrightarrow[n \rightarrow \infty]{L^p} X$ .  $\square$



## Chapitre 4

# Convergence en loi et théorème central limite

La loi forte des grands nombres introduite au chapitre précédent donnait un comportement asymptotique au premier ordre d'une somme de variables aléatoires indépendantes d'espérance finie  $\mu$  :

$$\frac{X_1 + \cdots + X_n}{n} = \mu + o(1).$$

La limite est une constante déterministe, ce qui peut surprendre de prime abord : l'agrégation des résultats aléatoires donne asymptotiquement un résultat non aléatoire.

On peut naturellement se demander quel est l'ordre suivant du développement asymptotique précédent. C'est le théorème central-limite, énoncé dans ce chapitre, qui donne la réponse : pour  $n$  grand,

$$\frac{X_1 + \cdots + X_n}{n} \stackrel{(\text{loi})}{\simeq} \mu + \frac{\sigma N}{\sqrt{n}},$$

où  $N$  désigne une variable aléatoire de loi normale centrée réduite et  $\sigma$  l'écart-type des  $X_k$ . Le caractère aléatoire réapparaît dans cette expression. Il reste à préciser la nature de l'approximation.

### 1 Convergence en loi

On note  $\mathcal{C}_b(\mathbb{R}^d)$  l'espace des fonctions continues bornées sur  $\mathbb{R}^d$  et à valeurs réelles que l'on munit de la norme uniforme :

$$\|h\|_\infty = \sup_{x \in \mathbb{R}^d} |h(x)|.$$

#### 1.1 Définition et premiers exemples

**Définition 4.1** (Convergence en loi). *On dit qu'une suite  $(X_n)_{n \geq 1}$  de variables aléatoires à valeurs dans  $\mathbb{R}^d$  converge en loi (ou en distribution) vers une variable aléatoire  $X$  à valeurs dans*

$\mathbb{R}^d$  si pour tout  $h \in \mathcal{C}_b(\mathbb{R}^d)$ ,

$$\mathbb{E}[h(X_n)] \xrightarrow{n \rightarrow \infty} \mathbb{E}[h(X)].$$

On note alors  $X_n \xrightarrow{\mathcal{L}} X$ . On trouvera aussi les notations  $X_n \xrightarrow{\text{loi}} X$  ou  $X_n \xrightarrow{(d)} X$  (pour convergence en distribution).

#### Remarques 4.2.

1. On peut considérer la convergence en loi de variables aléatoires définies sur des espaces de probabilité différents, ce qui n'était pas le cas des autres modes de convergence. Ici, on supposera implicitement pour simplifier que les variables aléatoires considérées sont toutes définies sur un même espace de probabilité.
2. On insiste sur un point essentiel. La convergence en loi ne fait intervenir les variables aléatoires qu'à travers leur loi. En particulier, si  $(X_n)$  converge en loi vers  $X$ , alors  $(X_n)$  converge également en loi vers n'importe quelle variable aléatoire  $X'$  de même loi que  $X$ . Ainsi, si  $(X_n)$  converge en loi vers  $X$  de loi  $\mu$ , alors on fera parfois un abus de langage du type " $(X_n)$  converge en loi vers  $\mu$ ".
3. On remarque donc que la convergence en loi est une convergence de mesures. En fait elle correspond à la convergence étroite de mesures. On dit qu'une suite de mesures  $(\mu_n)_{n \geq 1}$  sur  $\mathbb{R}^d$  converge étroitement vers une mesure  $\mu$  sur  $\mathbb{R}^d$  si pour toute fonction  $h : \mathbb{R}^d \rightarrow \mathbb{R}$  continue bornée

$$\int_{\mathbb{R}^d} h(x) d\mu_n(x) \xrightarrow{n \rightarrow \infty} \int_{\mathbb{R}^d} h(x) d\mu(x).$$

**Exemple 4.3.** Soit  $(X_n)_{n \geq 1}$  une suite de variables aléatoires de loi uniforme sur  $\{0, \frac{1}{n}, \frac{2}{n}, \dots, \frac{n-1}{n}, \frac{n}{n}\}$ , c'est-à-dire

$$\forall n \geq 1, \quad \mathbb{P}(X_n = k/n) = \frac{1}{n+1}, \quad k = 0, \dots, n.$$

Étudions la convergence en loi de la suite  $(X_n)_{n \geq 1}$ . Si  $h \in \mathcal{C}_b(\mathbb{R}^d)$ , alors le théorème de transfert donne

$$\mathbb{E}[h(X_n)] = \frac{1}{n+1} \sum_{k=0}^n h(k/n).$$

On reconnaît une somme de Riemann, d'où

$$\mathbb{E}[h(X_n)] = \frac{1}{n+1} \sum_{k=0}^n h(k/n) \xrightarrow{n \rightarrow \infty} \int_0^1 h(x) dx.$$

Donc si  $X$  est une variable aléatoire de loi uniforme sur  $[0, 1]$ , alors  $\mathbb{E}[h(X_n)] \rightarrow \mathbb{E}[h(X)]$ , ce qui prouve la convergence en loi  $X_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} X$ . Cela signifie qu'une loi uniforme (discrète) sur  $\{0, \frac{1}{n}, \frac{2}{n}, \dots, \frac{n-1}{n}, \frac{n}{n}\}$  converge en loi vers une loi uniforme (à densité) sur  $[0, 1]$ .

Cet exemple précédent montre en particulier qu'une suite de loi discrète peut converger vers une loi à densité. L'inverse peut aussi arriver.

**Exemple 4.4.** Soit  $(X_n)_{n \geq 1}$  une suite de variables aléatoires de loi normale  $\mathcal{N}(0, \sigma_n^2)$  avec  $\sigma_n > 0$  tel que  $\sigma_n \rightarrow 0$ . Étudions la convergence en loi de la suite  $(X_n)_{n \geq 1}$ . Pour  $h \in \mathcal{C}_b(\mathbb{R}^d)$ , le théorème de transfert et le changement de variable  $y = x/\sigma_n$  donnent

$$\mathbb{E}[h(X_n)] = \int_{\mathbb{R}} h(x) \frac{1}{\sigma_n \sqrt{2\pi}} e^{-\frac{x^2}{2\sigma_n^2}} dx = \int_{\mathbb{R}} h(\sigma_n y) \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy.$$

Posons  $h_n(y) = h(\sigma_n y) \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}$ . La fonction  $h_n$  est intégrable sur  $\mathbb{R}$  et vérifie

$$h_n(y) \xrightarrow{n \rightarrow \infty} h(0) \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} \quad \text{et} \quad h_n(y) \leq \|h\|_{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}.$$

Le théorème de convergence dominée assure alors que

$$\int_{\mathbb{R}} h_n(y) dy \xrightarrow{n \rightarrow \infty} \int_{\mathbb{R}} h(0) \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy = h(0).$$

En notant  $X$  une variable aléatoire de loi de Dirac en 0, on vient de montrer que

$$\mathbb{E}[h(X_n)] = \int_{\mathbb{R}} h_n(y) dy \xrightarrow{n \rightarrow \infty} h(0) = \mathbb{E}[h(X)].$$

Ainsi  $X_n \xrightarrow{\mathcal{L}} \delta_0$ .

## 1.2 Deux cas particuliers

### 1.2.1 Loi sur $\mathbb{N}$

**Proposition 4.5.** Une suite de variables aléatoires  $(X_n)_{n \geq 1}$  à valeurs dans  $\mathbb{N}$  converge vers une variable aléatoire  $X$  à valeurs dans  $\mathbb{N}$  si et seulement si

$$\forall k \in \mathbb{N}, \quad \mathbb{P}(X_n = k) \xrightarrow{n \rightarrow \infty} \mathbb{P}(X = k).$$

*Démonstration.* Supposons que  $(X_n)_{n \geq 1}$  converge en loi vers  $X$ . Pour  $k \in \mathbb{N}$  fixé, on considère la fonction  $h(x) = (1 - |x - k|)_+$  qui vaut 1 en  $k$  et est nulle en tout autre entier. La convergence  $\mathbb{E}[h(X_n)] \rightarrow \mathbb{E}[h(X)]$  se réécrit alors

$$\mathbb{P}(X_n = k) \xrightarrow{n \rightarrow \infty} \mathbb{P}(X = k).$$

Réciproquement, supposons que pour tout entier  $k$ ,  $\mathbb{P}(X_n = k) \rightarrow \mathbb{P}(X = k)$ . Soit  $h \in \mathcal{C}_b(\mathbb{R})$  et  $\epsilon > 0$ . Remarquons tout d'abord que comme  $\mathbb{P}(X > k) \rightarrow 0$  quand  $k \rightarrow \infty$ , il existe  $k_0$  tel que  $\mathbb{P}(X > k_0) \leq \epsilon$ . Or par hypothèse sur les  $X_n$ ,

$$\mathbb{P}(X_n > k_0) = 1 - \sum_{k=0}^{k_0} \mathbb{P}(X_n = k) \xrightarrow{n \rightarrow \infty} 1 - \sum_{k=0}^{k_0} \mathbb{P}(X = k) = \mathbb{P}(X > k_0),$$

ce qui implique qu'il existe  $n_0$  tel que  $\mathbb{P}(X > k_0) \leq 2\epsilon$ . On part de la décomposition

$$|\mathbb{E}[h(X_n)] - \mathbb{E}[h(X)]| = \left| \sum_{k=0}^{k_0} h(k)(\mathbb{P}(X_n = k) - \mathbb{P}(X = k)) \right| + \left| \sum_{k=k_0+1}^{\infty} h(k)\mathbb{P}(X_n = k) - \mathbb{P}(X = k) \right|.$$

Concernant le premier terme, pour tout  $k = 0, \dots, n$ , la convergence  $\mathbb{P}(X_n = k) \rightarrow \mathbb{P}(X = k)$  assure que pour  $n$  assez grand, on a la majoration  $|\mathbb{P}(X_n = k) - \mathbb{P}(X = k)| \leq \epsilon$  et donc

$$\left| \sum_{k=0}^{k_0} h(k)(\mathbb{P}(X_n = k) - \mathbb{P}(X = k)) \right| \leq \sum_{k=0}^{k_0} \|h\|_{\infty} |\mathbb{P}(X_n = k) - \mathbb{P}(X = k)| \leq \|h\|_{\infty} (k_0 + 1)\epsilon.$$

Par ailleurs, le deuxième terme est majoré par

$$\|h\|_{\infty} \left( \sum_{k=k_0+1}^{\infty} \mathbb{P}(X_n = k) + \sum_{k=k_0+1}^{\infty} \mathbb{P}(X = k) \right) = \|h\|_{\infty} (\mathbb{P}(X_n > k_0) + \mathbb{P}(X > k_0)) \leq \|h\|_{\infty} 3\epsilon.$$

Bref, on vient de montrer que pour  $n$  assez grand,  $|\mathbb{E}[h(X_n)] - \mathbb{E}[h(X)]| \leq \|h\|_{\infty} (k_0 + 4)\epsilon$ , ce qui implique que  $\mathbb{E}[h(X_n)] \rightarrow \mathbb{E}[h(X)]$ .  $\square$

**Exemple 4.6** (Approximation binomiale-Poisson). Soit  $(X_n)_{n \geq 1}$  une suite de variables aléatoires de loi binomiale de paramètre  $p_n$  tel que  $np_n \rightarrow \lambda > 0$ . Alors  $(X_n)_{n \geq 1}$  converge en loi vers une loi de Poisson de paramètre  $\lambda$ .

En effet, en notant  $X$  une variable aléatoire de loi de Poisson de paramètre  $\lambda$ , il suffit de montrer que pour tout  $k \in \mathbb{N}$ ,

$$\mathbb{P}(X_n = k) \xrightarrow{n \rightarrow \infty} \mathbb{P}(X = k),$$

c'est-à-dire que

$$\binom{n}{k} p_n^k (1 - p_n)^{n-k} \mathbb{1}_{\{k \leq n\}} \xrightarrow{n \rightarrow \infty} \frac{e^{-\lambda} \lambda^k}{k!}, \quad (4.1)$$

Soit donc  $k \in \mathbb{N}$  fixé. Pour  $n$  assez grand, l'indicatrice vaut 1, donc on regarde le comportement asymptotique de

$$\frac{n!}{(n-k)!k!} p_n^k (1 - p_n)^{n-k} = \frac{np_n \times (n-1)p_n \times \dots \times (n-k+1)p_n}{k!} (1 - p_n)^{n-k}.$$

Par hypothèse sur les  $p_n$ , on a la convergence  $(n-j+1)p_j \xrightarrow{n \rightarrow \infty} \lambda$  pour tout  $j$  fixé, et

$$(1 - p_n)^{n-k} = \exp[(n-k) \ln(1 - p_n)] = \exp[(n-k)(-p_n + o(p_n))] \rightarrow e^{-\lambda}.$$

Ceci prouve la convergence (4.1), et donc que  $\mathbb{P}(X_n = k) \xrightarrow{n \rightarrow \infty} \mathbb{P}(X = k)$  pour tout  $k \in \mathbb{N}$ . On a donc établi la convergence en loi de la suite  $(X_n)$  vers une loi de Poisson de paramètre  $\lambda$ .



### 1.2.2 Loi à densité sur $\mathbb{R}^d$

Le résultat suivant, appelé lemme de Scheffé, est une version continue de la proposition précédente.

**Proposition 4.7** (Lemme de Scheffé). *Soit  $(X_n)_{n \geq 1}$  une suite de variables aléatoires de densité respective  $(f_n)_{n \geq 1}$  telle que  $f_n$  converge presque partout vers une densité  $f$ . Alors  $X_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} X$ , où  $X$  est une variable aléatoire de densité  $f$ .*

*Démonstration.* On souhaiterait appliquer le théorème de convergence dominée mais on n'a aucune majoration pour  $f_n$ . On va donc s'appuyer sur le lemme de Fatou, qu'on applique à la suite de fonctions positives  $h_n = \min(f_n, f) = (f_n + f - |f_n - f|)/2$ , on obtient

$$\begin{aligned} \int_{\mathbb{R}^d} f(x) dx &= \int_{\mathbb{R}^d} \liminf_{n \rightarrow \infty} h_n(x) dx \\ &\leq \liminf_{n \rightarrow \infty} \int_{\mathbb{R}^d} h_n(x) dx = \liminf_{n \rightarrow \infty} \int_{\mathbb{R}^d} \frac{f_n(x) + f(x) - |f_n(x) - f(x)|}{2} dx. \end{aligned}$$

Comme  $f_n$  et  $f$  sont des densités, on en déduit que  $\limsup_{n \rightarrow \infty} \int_{\mathbb{R}^d} |f_n(x) - f(x)| dx = 0$  (ce qui signifie que  $(f_n)_{n \geq 1}$  converge vers  $f$  dans  $L^1(\mathbb{R}^d)$ ).

Par suite, si  $h \in \mathcal{C}_b(\mathbb{R}^d)$ , on a

$$\left| \int_{\mathbb{R}^d} h(x) f_n(x) dx - \int_{\mathbb{R}^d} h(x) f(x) dx \right| \leq \|h\|_{\infty} \int_{\mathbb{R}^d} |f_n(x) - f(x)| dx \xrightarrow[n \rightarrow \infty]{} 0,$$

ce qui donne le résultat. □

**Exemple 4.8.** Soit  $(X_n)$  une suite de variables aléatoires de loi exponentielle de paramètre  $\theta_n$  et  $X$  une variable aléatoire de loi exponentielle de paramètre  $\theta$ . Si  $\theta_n \rightarrow \theta$  quand  $n \rightarrow \infty$ , alors les densités de ces variables aléatoires vérifient

$$\theta_n e^{-\theta_n x} \xrightarrow[n \rightarrow \infty]{} \theta e^{-\theta x},$$

et la proposition précédente assure que  $X_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} X$ .

### 1.3 Lien avec les autres modes de convergence

Rappelons que la convergence en loi fait uniquement intervenir les lois des variables aléatoires et pas les variables aléatoires elles-mêmes. À l'inverse, les autres modes de convergence définies dans le chapitre précédent s'appuient sur la définition d'une variable aléatoire en tant que fonction mesurable. La comparaison de la convergence en loi avec d'autres modes de convergence nécessitent donc de considérer les variables aléatoires définies sur un même espace de probabilité (ce que l'on a fait jusqu'à présent).

Le résultat suivant donne le lien entre convergence en loi et convergence en probabilité.

**Proposition 4.9.** Soit  $(X_n)_{n \geq 1}$  une suite de variables aléatoires réelles convergeant en probabilité vers  $X$ . Alors  $(X_n)_{n \geq 1}$  converge en loi vers  $X$ .

*Démonstration.* Supposons que  $(X_n)_{n \geq 1}$  converge en probabilité vers  $X$ . Soit  $h \in \mathcal{C}_b(\mathbb{R})$  et  $\epsilon > 0$ . On a

$$\begin{aligned} |\mathbb{E}[h(X_n)] - \mathbb{E}[h(X)]| &\leq \mathbb{E}[|h(X_n) - h(X)| \mathbb{1}_{|h(X_n) - h(X)| < \epsilon}] + \mathbb{E}[|h(X_n) - h(X)| \mathbb{1}_{|h(X_n) - h(X)| \geq \epsilon}] \\ &\leq \epsilon + 2\|h\|_\infty \mathbb{P}(|h(X_n) - h(X)| \geq \epsilon). \end{aligned}$$

Or, si  $(X_n)$  converge en probabilité vers  $X$ , alors  $(h(X_n))$  converge en probabilité vers  $h(X)$ , donc il existe  $n_0$  tel que pour tout  $n \geq n_0$ ,  $\mathbb{P}(|h(X_n) - h(X)| \geq \epsilon) \leq \epsilon$ . Ainsi,

$$|\mathbb{E}[h(X_n)] - \mathbb{E}[h(X)]| \leq \epsilon(1 + 2\|h\|_\infty).$$

ce qui permet de conclure que  $X_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} X$ . □

**Contre-exemple 4.10.** Soit  $X$  une variable aléatoire de loi de Bernoulli de paramètre  $1/2$  et  $(X_n)_{n \geq 1}$  une suite constante de variables aléatoires telle que  $X_n = X$  pour tout  $n$ . Alors  $(X_n)$  converge en loi vers  $X$ . Par ailleurs, on peut remarquer que la variable aléatoire  $Y = 1 - X$  suit aussi une loi de Bernoulli de paramètre  $1/2$ , donc  $(X_n)_{n \geq 1}$  converge en loi vers  $Y = 1 - X$ . Cela peut surprendre mais c'est cohérent si on se souvient que la convergence en loi ne fait intervenir que les lois des variables aléatoires. Cependant, comme  $|X_n - Y| = |2X - 1| = 1$  pour tout  $n$ , il s'ensuit que pour  $\epsilon > 0$  assez petit,  $\mathbb{P}(|X_n - Y| \geq \epsilon) = 1$ , donc  $(X_n)_{n \geq 1}$  ne converge pas en probabilité vers  $Y$ .

Il existe cependant un cas où la réciproque de la proposition précédente est vraie.

**Proposition 4.11.** Soit  $(X_n)$  une suite de variables aléatoires convergeant en loi vers une loi de Dirac  $\delta_a$ . Alors la convergence a lieu en probabilité :  $X_n \xrightarrow[n \rightarrow \infty]{\mathbb{P}} a$ .

*Démonstration.* On suppose que  $X_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} c$ , où  $c \in \mathbb{R}$  est une constante. Soit  $\epsilon > 0$ . On pose  $h(x) = \min(|x - c|/\epsilon, 1)$  de sorte que  $h$  est continue, bornée, nulle en  $c$  et vérifie  $\mathbb{1}_{\{|x - c| \geq \epsilon\}} \leq h(x)$ . Ainsi,

$$\mathbb{P}(|X_n - c| \geq \epsilon) = \mathbb{E}[\mathbb{1}_{\{|x - c| \geq \epsilon\}}] \leq \mathbb{E}[h(X_n)] \xrightarrow[n \rightarrow \infty]{} \mathbb{E}[h(c)] = 0.$$

□

Ce résultat peut sembler surprenant à première vue : on arrive à passer de la convergence en loi, qui ne dépend pas de l'espace de probabilité, à la convergence en probabilité, qui en dépend. L'astuce se cache dans la probabilité  $\mathbb{P}(|X_n - X| \leq \epsilon)$  qui ne dépend en fait que de l'aléa contenu dans  $X_n$ , puisque  $X$  est constant.

## 2 Caractérisation de la convergence en loi

### 2.1 Restriction des fonctions tests

On note  $\mathcal{C}_c(\mathbb{R}^d)$  l'ensemble des fonctions continues à support compact définies sur  $\mathbb{R}^d$  et à valeurs réelles.

**Proposition 4.12** (Restriction des fonctions tests). *Soit  $(X_n)_{n \geq 1}$  et  $X$  des variables aléatoires à valeurs dans  $\mathbb{R}^d$ . Les assertions suivantes sont équivalentes.*

1. La suite  $(X_n)_{n \geq 1}$  converge en loi vers  $X$ .
2. Pour tout  $h \in \mathcal{C}_c(\mathbb{R}^d)$ ,

$$\mathbb{E}[h(X_n)] \xrightarrow{n \rightarrow \infty} \mathbb{E}[h(X)].$$

Par ailleurs, soit  $H$  un ensemble de fonctions mesurables bornées définies sur  $\mathbb{R}^d$  et dont l'adhérence pour la norme uniforme  $\|\cdot\|_\infty$  contient  $\mathcal{C}_c(\mathbb{R}^d)$ <sup>1</sup>. Si pour tout  $h \in H$ ,

$$\mathbb{E}[h(X_n)] \xrightarrow{n \rightarrow \infty} \mathbb{E}[h(X)],$$

alors  $X_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} X$ .

Autrement dit, il suffit de vérifier la convergence sur un ensemble de fonctions "assez grand".

*Démonstration.* On a clairement  $1. \implies 2.$

Montrons que  $2. \implies 1.$  Soit  $h \in \mathcal{C}_b(\mathbb{R}^d)$ . Pour  $k \geq 1$ , on définit la fonction  $g_k(x) = \min(1, k + 1 - \|x\|_\infty)_+$  de sorte que  $g_k$  est continue, bornée entre 0 et 1, est nulle en-dehors de la boule  $B(0, k + 1)$  (pour la norme infinie), et vaut 1 sur la boule  $B(0, k)$ , et converge vers la fonction constante égale à 1 quand  $k \rightarrow \infty$ . On en déduit donc que  $g_k h$  est dans  $\mathcal{C}_c(\mathbb{R}^d)$  et converge vers  $h$  pour la norme infinie. Soit alors  $h \in \mathcal{C}_b(\mathbb{R}^d)$ . Alors  $h g_k \in \mathcal{C}_c(\mathbb{R}^d)$  et donc par hypothèse

$$\mathbb{E}[h g_k(X_n)] \xrightarrow{n \rightarrow \infty} \mathbb{E}[h g_k(X)].$$

On part alors de la décomposition

$$\begin{aligned} & |\mathbb{E}[h(X_n)] - \mathbb{E}[h(X)]| \\ & \leq |\mathbb{E}[h(X_n)] - \mathbb{E}[h g_k(X_n)]| + |\mathbb{E}[h g_k(X_n)] - \mathbb{E}[h g_k(X)]| + |\mathbb{E}[h g_k(X)] - \mathbb{E}[h(X)]| \\ & \leq \|h\|_\infty \mathbb{E}[(1 - g_k)(X_n)] + \mathbb{E}[h g_k(X_n)] - \mathbb{E}[h g_k(X)] + \|h\|_\infty \mathbb{E}[(1 - g_k)(X)], \end{aligned}$$

et on en déduit que

$$\begin{aligned} \limsup_{n \rightarrow \infty} |\mathbb{E}[h(X_n)] - \mathbb{E}[h(X)]| & \leq \|h\|_\infty \limsup_{n \rightarrow \infty} \left( (1 - \mathbb{E}[h_k(X_n)]) + (1 - \mathbb{E}[h_k(X)]) \right) \\ & = 2\|h\|_\infty (1 - \mathbb{E}[h_k(X)]). \end{aligned}$$

---

1. Rappelons que l'adhérence d'un tel ensemble  $H$  correspond à toutes les fonctions qui sont limites pour la norme uniforme de suites de fonctions de  $H$ .

Comme  $h_k \xrightarrow[k \rightarrow \infty]{} 1$ , avec  $h_k$  bornée par 1, le théorème de convergence dominée assure que le membre de droite tend vers 0 quand  $k \rightarrow \infty$ . Ainsi

$$\limsup_{n \rightarrow \infty} |\mathbb{E}[h(X_n)] - \mathbb{E}[h(X)]| = 0,$$

ce qui donne la convergence en loi de  $(X_n)$  vers  $X$ . Ceci prouve l'équivalence.

Supposons maintenant que  $\mathbb{E}[h(X_n)] \rightarrow \mathbb{E}[h(X)]$  pour tout  $h \in H$ , ensemble de fonctions mesurables bornées dont l'adhérence contient  $\mathcal{C}_c(\mathbb{R}^d)$ . On va montrer qu'alors le point 2. de l'équivalence précédente est vérifiée. Soit donc  $h \in \mathcal{C}_c(\mathbb{R}^d) \subset \bar{H}$ . Par définition de l'adhérence de  $H$ , il existe une suite de fonctions  $(h_k)$  dans  $H$  telle que  $h_k \rightarrow h$  pour la norme infinie. On part toujours de la décomposition

$$\begin{aligned} |\mathbb{E}[h(X_n)] - \mathbb{E}[h(X)]| &\leq |\mathbb{E}[h(X_n)] - \mathbb{E}[h_k(X_n)]| + |\mathbb{E}[h_k(X_n)] - \mathbb{E}[h_k(X)]| + |\mathbb{E}[h_k(X)] - \mathbb{E}[h(X)]| \\ &\leq \|h - h_k\|_\infty + |\mathbb{E}[h_k(X_n)] - \mathbb{E}[h_k(X)]| + \|h - h_k\|_\infty. \end{aligned}$$

Comme  $h_k \in H$ , l'hypothèse 3. assure que le terme central converge vers 0 quand  $n \rightarrow \infty$  et donc

$$\limsup_{n \rightarrow \infty} |\mathbb{E}[h(X_n)] - \mathbb{E}[h(X)]| \leq 2\|h - h_k\|_\infty.$$

Ceci étant valable pour tout  $k \geq 1$ , on obtient  $\mathbb{E}[h(X_n)] \xrightarrow[n \rightarrow \infty]{} \mathbb{E}[h(X)]$ , ce qui donne le résultat voulu.  $\square$

**Remarque 4.13.** La convergence du point 2. est appelée convergence vague. On vient de montrer qu'elle équivaut à la convergence en loi. Ceci est vrai car les mesures considérées (ici les lois de  $(X_n)_{n \geq 1}$  et  $X$  sont toutes de même masse totale finie. La preuve s'appuie en effet sur le fait que la fonction constante égale à 1 est intégrable. Dans d'autres espaces mesurés, on sait juste que la convergence en loi implique la convergence vague.

## 2.2 Caractérisation en termes de fonction de répartition et de fonction caractéristique

**Proposition 4.14** (Convergence en loi et fonction de répartition). *La suite de variables aléatoires réelles  $(X_n)_{n \geq 1}$  converge en loi vers la variable aléatoire réelle  $X$  si et seulement si  $F_{X_n}(t) \rightarrow F_X(t)$  pour tout réel  $t$  point de continuité de  $F_X$ .*

Rappelons que la fonction de répartition  $F_X$  est croissante et continue à droite, et que pour tout  $t \in \mathbb{R}$ ,

$$\mathbb{P}(X = t) = F_X(t) - \lim_{\substack{\epsilon \rightarrow 0 \\ \epsilon > 0}} F_X(t - \epsilon).$$

Ainsi,  $F_X$  est continue en  $t$  si et seulement si  $\mathbb{P}(X = t) = 0$ .

*Démonstration.* Supposons  $X_n \xrightarrow{\mathcal{L}} X$  et fixons  $t \in \mathbb{R}$  un point de continuité de  $F_X$ . Comme  $F_X(t) = \mathbb{E}[\mathbb{1}_{]-\infty, t]}(X)]$ , l'idée est d'encadrer  $x \mapsto \mathbb{1}_{]-\infty, t]}(x)$  par des fonctions continues bornées. On considère la fonction  $f_{t, \epsilon}$  définie par

$$h_{t, \epsilon}(x) = \frac{(t - x)_+}{\epsilon} \wedge 1, \quad x \in \mathbb{R}.$$

On remarque que cette fonction est continue et bornée et qu'elle vérifie

$$\mathbb{1}_{]-\infty, t-\epsilon]}(x) \leq h_{t, \epsilon}(x) \leq \mathbb{1}_{]-\infty, t]}(x), \quad x \in \mathbb{R}.$$

On en déduit alors que

$$\mathbb{E}[\mathbb{1}_{]-\infty, t-\epsilon]}(X)] \leq \mathbb{E}[h_{t, \epsilon}(X)] = \liminf_{n \rightarrow \infty} \mathbb{E}[h_{t, \epsilon}(X_n)] \leq \liminf_{n \rightarrow \infty} \mathbb{E}[\mathbb{1}_{]-\infty, t]}(X_n)],$$

ce qui se réécrit

$$F_X(t - \epsilon) \leq \liminf_{n \rightarrow \infty} F_{X_n}(t).$$

De manière analogue, on a les inégalités

$$\mathbb{1}_{]-\infty, t]}(x) \leq h_{t+\epsilon, \epsilon}(x) \leq \mathbb{1}_{]-\infty, t+\epsilon]}(x), \quad x \in \mathbb{R}.$$

On en déduit alors que

$$\limsup_{n \rightarrow \infty} \mathbb{E}[\mathbb{1}_{]-\infty, t]}(X_n)] \leq \limsup_{n \rightarrow \infty} \mathbb{E}[h_{t+\epsilon, \epsilon}(X_n)] = \mathbb{E}[h_{t+\epsilon, \epsilon}(X)] \leq \mathbb{E}[\mathbb{1}_{]-\infty, t+\epsilon]}(X)],$$

ce qui se réécrit

$$\limsup_{n \rightarrow \infty} F_{X_n}(t) \leq F_X(t + \epsilon).$$

Bref, en recollant les morceaux on trouve que

$$F_X(t - \epsilon) \leq \liminf_{n \rightarrow \infty} F_{X_n}(t) \leq \limsup_{n \rightarrow \infty} F_{X_n}(t) \leq F_X(t + \epsilon).$$

En faisant tendre  $\epsilon$  vers 0 et en utilisant la continuité de  $F_X$  en  $t$  on en déduit alors que

$$F_X(t) \leq \liminf_{n \rightarrow \infty} F_{X_n}(t) \leq \limsup_{n \rightarrow \infty} F_{X_n}(t) \leq F_X(t),$$

en finalement que  $\lim_{n \rightarrow \infty} F_{X_n}(t) = F_X(t)$ .

Réciproquement, supposons que  $F_{X_n}(x) \rightarrow F_X(x)$  en tout point de continuité  $x$  de  $F_X$ . On considère l'ensemble  $D$  des points de discontinuité de  $F_X$ . Comme  $F_X$  est croissante,  $D$  est au plus dénombrable. De plus, si  $a, b \notin D$  avec  $a < b$ , alors

$$\mathbb{P}(X_n \in ]a, b]) = F_{X_n}(b) - F_{X_n}(a) \xrightarrow{n \rightarrow \infty} F_X(b) - F_X(a) = \mathbb{P}(X \in ]a, b]).$$

Soit alors  $H$  l'espace vectoriel engendré par les fonctions  $\mathbb{1}_{]a, b]}$ , pour tout  $a, b \in \mathbb{R} \setminus D$ . La convergence ci-dessus s'étend par linéarité à  $\mathbb{E}[f(X_n)] \rightarrow \mathbb{E}[f(X)]$  pour tout  $f \in H$ . Comme  $\mathcal{C}_c(\mathbb{R}) \subset \bar{H}$  pour la norme uniforme, on conclut par la proposition 4.12.  $\square$

**Théorème 4.15** (Théorème de Lévy). *Une suite  $(X_n)$  de variables aléatoires à valeurs dans  $\mathbb{R}^d$  converge en loi vers une variable aléatoire  $X$  à valeurs dans  $\mathbb{R}^d$  si et seulement si pour tout  $t \in \mathbb{R}^d$ , on a*

$$\varphi_{X_n}(t) \xrightarrow{n \rightarrow \infty} \varphi_X(t).$$

La preuve de ce résultat étant compliquée, on la repousse en fin de chapitre.

### 3 Théorème central limite

#### 3.1 Énoncé

Si  $X_1, \dots, X_n$  sont des variables aléatoires iid intégrables, alors la loi des grands nombres assure que la moyenne  $(X_1 + \dots + X_n)/n$  converge presque sûrement vers  $\mathbb{E}[X_1]$ . L'idée est de d'étudier la vitesse de convergence vers zéro de la différence

$$\frac{X_1 + \dots + X_n}{n} - \mathbb{E}[X_1].$$

Sous l'hypothèse supplémentaire que les variables aléatoires  $X_i$  sont dans  $L^2$ , on peut intuitiver la réponse en calculant, comme dans la preuve de la loi faible des grands nombres,

$$\mathbb{E}[(X_1 + \dots + X_n - n\mathbb{E}[X_1])^2] = \text{Var}(X_1 + \dots + X_n) = n \text{Var}(X_1).$$

ce qui indique que la valeur moyenne de  $(X_1 + \dots + X_n - n\mathbb{E}[X_1])^2$  croît linéairement avec  $n$ . Il semble donc que la vitesse de convergence soit de l'ordre que  $\sqrt{n}$ . Le théorème central-limite formalise ceci.

**Théorème 4.16** (Théorème central limite). *Soit  $(X_n)_{n \geq 1}$  une suite de variables aléatoires réelles indépendantes et de même loi dans  $L^2(\Omega, \mathcal{F}, \mathbb{P})$ . On pose  $S_n = X_1 + \dots + X_n$  et on suppose  $\sigma^2 = \text{Var}(X_1) > 0$ . Alors*

$$\frac{S_n - n\mathbb{E}[X_1]}{\sigma\sqrt{n}} \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, 1).$$

On en déduit le caractère fondamental de la loi normale en probabilités : elle intervient comme limite en loi d'une somme de variables aléatoires correctement normalisées. Il faut se rendre compte des hypothèses très faibles de ce théorèmes. Il suffit en effet de supposer que les variables aléatoires iid considérées admettent une variance pour que la convergence en loi ait lieu. En particulier, leur loi peut être quelconque (par exemple discrète), la somme normalisée converge vers une loi normale.

#### Remarques 4.17.

1. Pour se souvenir de la normalisation, il suffit de déterminer l'espérance et la variance de  $S_n$  :

$$\mathbb{E}[S_n] = \mathbb{E}[X_1 + \dots + X_n] = n\mathbb{E}[X_1],$$

et

$$\text{Var}(X_1) = \text{Var}(X_1 + \cdots + X_n) = n \text{Var}(X_1).$$

Ainsi, la variable aléatoire  $\frac{S_n - n\mathbb{E}[X_1]}{\sigma\sqrt{n}}$  est centrée réduite, comme la loi normale qui intervient dans la limite.

2. On peut en particulier réécrire la convergence du TCL en plaçant la variance dans la limite :

$$\frac{S_n - n\mathbb{E}[X_1]}{\sqrt{n}} \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, \sigma^2).$$

On remarque encore une fois que les membres de gauche et de droite ont la même espérance et la même variance.

3. En divisant le numérateur et le dénominateur du membre de gauche par  $n$ , on en déduit une version pour la moyenne  $\bar{X}_n = S_n/n$  :

$$\sqrt{n} \frac{\bar{X}_n - \mathbb{E}[X_1]}{\sigma} \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, 1).$$

4. Via la caractérisation de la convergence en loi par les fonctions de répartition, et comme la loi normale admet une fonction de répartition continue en tout point, le théorème équivaut à dire que pour tout  $x \in \mathbb{R}$ ,

$$\mathbb{P}\left(\frac{S_n - n\mathbb{E}[X_1]}{\sigma\sqrt{n}} \leq x\right) \xrightarrow[n \rightarrow \infty]{} \int_{-\infty}^x \frac{e^{-\frac{t^2}{2}}}{\sqrt{2\pi}} dt.$$

ou encore que pour tout  $a, b \in \mathbb{R}$ , avec  $a < b$ ,

$$\mathbb{P}\left(a \leq \frac{S_n - n\mathbb{E}[X_1]}{\sigma\sqrt{n}} \leq b\right) \xrightarrow[n \rightarrow \infty]{} \int_a^b \frac{e^{-\frac{t^2}{2}}}{\sqrt{2\pi}} dt.$$

Démontrons à présent le théorème central limite. La preuve s'appuie sur le théorème de Lévy.

*Démonstration.* Quitte à remplacer  $X_n$  par  $X_n - \mathbb{E}[X_n]$  on peut supposer que les variables aléatoires sont centrées. On souhaite donc établir la convergence en loi de  $\frac{S_n}{\sigma\sqrt{n}}$  vers  $\mathcal{N}(0, 1)$ . D'après le théorème de Lévy, il suffit de montrer que

$$\varphi_{\frac{S_n}{\sigma\sqrt{n}}}(t) \xrightarrow[t \rightarrow \infty]{} e^{-\frac{t^2}{2}}.$$

Or, les  $X_i$  étant iid, on a

$$\varphi_{\frac{S_n}{\sigma\sqrt{n}}}(t) = \mathbb{E}\left[e^{it\frac{X_1 + \cdots + X_n}{\sigma\sqrt{n}}}\right] = \prod_{j=1}^n \mathbb{E}\left[e^{it\frac{X_j}{\sigma\sqrt{n}}}\right] = \left(\varphi_{X_1}\left(\frac{t}{\sigma\sqrt{n}}\right)\right)^n.$$

Par suite, comme  $X_1 \in L^2(\Omega, \mathcal{F}, \mathbb{P})$ , sa fonction caractéristique  $\varphi_{X_1}$  est de classe  $\mathcal{C}^2(\mathbb{R}, \mathbb{C})$  et vérifie  $\varphi'_{X_1}(0) = i\mathbb{E}[X_1] = 0$  et  $\varphi''_{X_1}(0) = -\mathbb{E}[X_1^2] = -\sigma^2$ . Un développement limité en 0 donne alors

$$\varphi_{X_1}(x) = 1 - \frac{\sigma^2}{2}x^2 + o(x^2), \quad x \rightarrow 0.$$

On obtient alors

$$\varphi_{\frac{S_n}{\sigma\sqrt{n}}}(t) = \left(1 - \frac{t^2}{2n} + o\left(\frac{1}{n}\right)\right)^n \xrightarrow{n \rightarrow \infty} e^{-\frac{t^2}{2}},$$

et le théorème de Lévy permet de conclure.  $\square$

**Exemple 4.18.** Soit  $X_1, \dots, X_n$  une suite de variables aléatoires indépendantes de loi de Bernoulli de paramètre  $1/2$ . Posons  $S_n = X_1 + \dots + X_n$ . On rappelle que  $\mathbb{E}[X_1] = 1/2$  et que  $\text{Var}(X_1) = 1/4$ . Le théorème central limite donne alors la convergence en loi

$$2\sqrt{n}\left(\frac{S_n}{n} - \frac{1}{2}\right) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, 1).$$

### 3.2 Application à la statistique

L'idée générale de la statistique inférentielle est la suivante. On dispose de données  $x_1, \dots, x_n$  que l'on considère comme la réalisation d'une suite de variables aléatoires iid  $X_1, \dots, X_n$ . Cela signifie qu'il existe  $\omega \in \Omega$  tel que  $X_1(\omega) = x_1, \dots, X_n(\omega) = x_n$ . L'objectif est alors de déterminer au mieux la loi des  $X_k$ . On va donc travailler directement avec la suite de variables aléatoires  $X_1, \dots, X_n$ .

Si on veut connaître l'espérance  $\mu := \mathbb{E}[X_1]$  (qui est une quantité inconnue puisqu'on ne connaît pas la loi de  $X_1$ ), on l'estime via  $\bar{X}_n = \frac{X_1 + \dots + X_n}{n}$  et, sous l'hypothèse d'intégrabilité des  $X_k$ , la loi des grands nombres assure que

$$\bar{X}_n \xrightarrow[n \rightarrow \infty]{p.s.} \mathbb{E}[X_1].$$

On dit que  $\bar{X}_n$  est un estimateur (fortement) convergeant de  $\mu$ .

Si on suppose maintenant que  $X_1$  est de carré intégrable, le théorème central limite donne alors l'erreur asymptotique que l'on commet en assimilant  $\bar{X}_n$  à  $\mu$  :

$$\mathbb{P}\left(|\bar{X}_n - \mu| \geq \frac{\sigma t}{\sqrt{n}}\right) \xrightarrow[n \rightarrow \infty]{} \mathbb{P}(|N| \geq t) = 2(1 - \Phi(t)) = 2 \int_t^\infty \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx,$$

où  $N \sim \mathcal{N}(0, 1)$  et  $\Phi$  désigne la fonction de répartition de  $N$ . Si on prend  $t = 1.96$ , une lecture de la table de la loi normale donne alors  $2(1 - \Phi(t)) = 0.05$ . Ainsi, quand  $n$  est assez grand, l'erreur commise en assimilant  $\mu$  à  $\bar{X}_n$  sera d'au plus  $1.96\sigma/\sqrt{n}$  avec une probabilité inférieure à 0.05%. Autrement dit, la paramètre  $\mu$  appartient à l'intervalle

$$\left[\bar{X}_n - \frac{1.96\sigma}{\sqrt{n}}, \bar{X}_n + \frac{1.96\sigma}{\sqrt{n}}\right]$$



avec une probabilité de 95%. Un tel intervalle est appelé *intervalle de confiance* (asymptotique) de niveau 95%.

Insistons sur un point. Le paramètre inconnu  $\mu$  correspond à l'espérance des  $X_k$  : il est donc déterministe. Si on dit qu'il appartient à l'intervalle précédent avec une probabilité de 95% c'est parce que l'intervalle est aléatoire : il dépend en effet de  $\bar{X}_n$ . On notera toutefois que cet intervalle dépend de  $\sigma$ , l'écart-type des  $X_k$ , qui est sûrement aussi inconnu.

**Application aux sondages** Lors d'une élection, deux candidats  $A$  et  $B$  s'opposent. On note  $p$  la proportion de voix (inconnue avant l'élection) en faveur du candidat  $A$ . Un sondage consiste alors à interroger  $n$  individus de la population pour connaître leur intention de vote (en pratique on prend souvent  $n$  au moins égal à 1000). On note  $X_k$  la variable aléatoire valant 1 si le  $k$ -ème individu interrogé déclare voter pour  $A$  et 0 s'il va voter pour  $B$ . Les variables aléatoires  $X_1, \dots, X_n$  suivent alors une loi de Bernoulli de paramètre  $p$ . On suppose que les  $X_k$  sont des variables aléatoires indépendantes (ça se discute...).

À l'issue du sondage on obtient un échantillon  $x_1, \dots, x_n$  de *données* à valeurs dans  $\{0, 1\}$ . On travaille cependant au niveau des variables aléatoires. Le sondage est une expérience aléatoire (les sondés sont choisis au hasard) et un nouveau sondage ne donnerait pas forcément le même échantillon  $x_1, \dots, x_n$ , mais les variables aléatoires  $X_1, \dots, X_n$  auraient toujours la même loi.

Il semble naturel d'estimer la proportion inconnue  $p$  par

$$\bar{X}_n = \frac{X_1 + \dots + X_n}{n},$$

qui correspond à la proportion empirique de voix en faveur du candidat  $A$ . La loi des grands nombres donne la convergence

$$\bar{X}_n \xrightarrow[n \rightarrow \infty]{p.s.} \mathbb{E}[X_1] = p,$$

où on a reconnu l'espérance d'une variable aléatoire de Bernoulli de paramètre  $p$ . Quand  $n$  est grand, la proportion empirique de voix se rapproche de la vraie proportion  $p$ . Mais avec quelle erreur ? C'est le théorème central limite qui donne une estimation de l'erreur. La variance de  $X_1$  vaut  $p(1-p)$  donc

$$\sqrt{n} \frac{\bar{X}_n - p}{\sqrt{p(1-p)}} \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, 1),$$

où, de manière équivalente,

$$\mathbb{P}\left(\left|\sqrt{n} \frac{\bar{X}_n - p}{\sqrt{p(1-p)}}\right| \geq t\right) \xrightarrow[n \rightarrow \infty]{} 2(1 - \Phi(t)) = 2 \int_t^\infty \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx.$$

En choisissant  $t = 1.96$ , le membre de droite vaut alors 0.05. Ainsi, avec une probabilité de 95%,  $p$  appartient à l'intervalle

$$\left[\bar{X}_n - \frac{1.96\sqrt{p(1-p)}}{\sqrt{n}}, \bar{X}_n + \frac{1.96\sqrt{p(1-p)}}{\sqrt{n}}\right].$$

Mais ceci n'est pas du tout satisfaisant : le paramètre  $p$  apparaît dans l'intervalle ! On peut toutefois majorer l'écart-type  $\sqrt{p(1-p)}$  par  $1/2$ , de sorte que

$$\left[ \bar{X}_n - \frac{0.98}{\sqrt{n}}, \bar{X}_n + \frac{0.98}{\sqrt{n}} \right]$$

est un intervalle de confiance (asymptotique) de niveau au moins 95% pour  $p$ . Pour  $n = 1000$  personnes, l'intervalle de confiance est de rayon  $0.98/\sqrt{1000} \approx 0.03$ , soit 3%. Autrement dit, avec une probabilité de l'ordre de 95%, le paramètre  $p$  égale le résultat du sondage  $\bar{X}_n$  à plus ou moins 3% près. Ces 3% sont parfois appelés *marge d'erreur*.

Si on revient à l'intervalle de confiance, on peut remarquer qu'il est centré en l'estimateur  $\bar{X}_n$ . C'est la moindre des choses : il n'y a pas de raison de favoriser une erreur par excès ou par défaut dans l'approximation de  $p$  par  $\bar{X}_n$ . Par ailleurs, la largeur de l'intervalle diminue quand  $n$  augmente : interroger davantage de personnes permet de réduire l'erreur d'estimation.

On peut se demander s'il est possible d'obtenir un intervalle de confiance non-asymptotique, c'est-à-dire de trouver  $t > 0$  tel que

$$\mathbb{P}(|\bar{X}_n - p| \geq t) \leq 0.05.$$

Comme  $\mathbb{E}[\bar{X}_n] = p$ , l'inégalité de Bienaymé-Tchebychev donne

$$\mathbb{P}(|\bar{X}_n - p| \geq t) = \mathbb{P}(|\bar{X}_n - \mathbb{E}[\bar{X}_n]| \geq t) \leq \frac{\text{Var}(\bar{X}_n)}{t^2} = \frac{p(1-p)}{nt^2}.$$

Le numérateur de la borne étant inconnue (elle dépend de  $p$ ), on le majore par  $1/4$ , ce qui donne

$$\mathbb{P}(|\bar{X}_n - p| \geq t) \leq \frac{1}{4nt^2}.$$

En choisit alors  $t$  tel que  $1/(4nt^2) = 0.05$ , ce qui donne  $t = 1/\sqrt{0.2n} \approx 2.24/\sqrt{n}$ . Ainsi, on est sûr à 95% que  $p$  est dans l'intervalle de confiance

$$\left[ \bar{X}_n - \frac{2.24}{\sqrt{n}}, \bar{X}_n + \frac{2.24}{\sqrt{n}} \right].$$

Cet intervalle est plus grand que celui trouvé via le théorème central limite (2.24 vs 0.98), donc de moins bonne qualité. Cependant, il est valable pour tout  $n$  alors que le précédent est un intervalle de confiance asymptotique.

## 4 Appendice au Chapitre 4 : démonstrations techniques

### 4.1 Preuve du théorème de Lévy (Théorème 4.15)

La preuve s'appuie sur la transformée de Fourier d'une fonction  $f \in L^1(\mathbb{R}^d)$  définie par

$$\forall \xi \in \mathbb{R}^d, \quad \hat{f}(\xi) = \int_{\mathbb{R}^d} f(x) e^{-ix\xi} dx.$$

dont on rappelle ci-dessous quelques propriétés.

1. *Injectivité de la transformée de Fourier.* Si  $f \in L^1(\mathbb{R}^d)$  est telle que  $\hat{f} \in L^1(\mathbb{R}^d)$ , alors  $\hat{\hat{f}}(x) = (2\pi)^d f(-x)$  pour presque tout  $x \in \mathbb{R}^d$ .
2. *Transformée de Fourier et espérance.* Pour tout  $f \in L^1(\mathbb{R}^d)$  telle que  $\hat{f} \in L^1(\mathbb{R}^d)$ , et pour toute variable aléatoire  $X$  à valeurs dans  $\mathbb{R}^d$ , on a la relation

$$\mathbb{E}[\hat{f}(X)] = \mathbb{E}\left[\int_{\mathbb{R}^d} e^{itX} dt\right] = \int_{\mathbb{R}^d} \mathbb{E}[e^{itX}] dt \quad (4.2)$$

3. *Transformée de Fourier d'une gaussienne.* Pour  $\sigma > 0$ , on désigne par  $g_\sigma$  la densité gaussienne centrée de variance  $\sigma^2 > 0$  :

$$g_\sigma(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}}, \quad x \in \mathbb{R}.$$

Alors :

- (a) La transformée de Fourier de  $g_\sigma$  est donnée par

$$\hat{g}_\sigma(\xi) = \frac{\sqrt{2\pi}}{\sigma} g_{1/\sigma}(\xi), \quad \xi \in \mathbb{R}.$$

Ce résultat peut se montrer à l'aide d'une dérivation sous la signe intégrale puis d'une intégration par parties.

- (b) Pour tout  $f, g \in L^1(\mathbb{R})$ ,  $\widehat{f * g} = \hat{f} \hat{g}$ .
- (c) Pour tout  $f \in \mathcal{C}_c(\mathbb{R})$ ,  $g_\sigma * f \rightarrow f$  pour la norme uniforme.

On pose  $H = \{g_\sigma * f : f \in \mathcal{C}_c(\mathbb{R}), \sigma > 0\}$ . Le point 3c implique que l'adhérence de  $H$  pour la norme uniforme contient  $\mathcal{C}_c(\mathbb{R})$ . D'après la proposition 4.12, il suffit donc de montrer que pour tout  $\sigma > 0$  et tout  $f \in \mathcal{C}_c(\mathbb{R})$ ,

$$\mathbb{E}[g_\sigma * f(X_n)] \xrightarrow{n \rightarrow \infty} \mathbb{E}[g_\sigma * f(X)].$$

Soit donc  $\sigma > 0$  et  $f \in \mathcal{C}_c(\mathbb{R})$ . En combinant 3a et 3b, on obtient

$$\widehat{g_\sigma * f} = \frac{\sqrt{2\pi}}{\sigma} g_{1/\sigma} \hat{f},$$

et cette fonction est intégrable puisque  $\hat{f}$  est bornée car  $f$  est à support compact. On utilise alors la propriété 1 qui donne pour presque tout  $x \in \mathbb{R}$ ,

$$g_\sigma * f(x) = \frac{\sqrt{2\pi}}{\sigma} g_{1/\sigma} \hat{f}(-x).$$

On applique alors la relation (4.2) qui donne

$$\mathbb{E}[g_\sigma * f(X_n)] = \int_{\mathbb{R}} \frac{\sqrt{2\pi}}{\sigma} g_{1/\sigma} \hat{f}(-t) \mathbb{E}[e^{itX_n}] dt.$$

La fonction à intégrer est bornée par  $\frac{\sqrt{2\pi}}{\sigma} g_{1/\sigma} \hat{f}(-t)$  qui est intégrable (voir plus haut). Le théorème de convergence dominée donne alors

$$\int_{\mathbb{R}} \frac{\sqrt{2\pi}}{\sigma} g_{1/\sigma} \hat{f}(-t) \mathbb{E}[e^{itX_n}] dt \xrightarrow{n \rightarrow \infty} \int_{\mathbb{R}} \frac{\sqrt{2\pi}}{\sigma} g_{1/\sigma} \hat{f}(-t) \mathbb{E}[e^{itX}] dt.$$

En utilisant encore une fois la relation (4.2), on obtient alors que

$$\mathbb{E}[g_\sigma * f(X_n)] \xrightarrow{n \rightarrow \infty} \mathbb{E}[g_\sigma * f(X)].$$

ce qui conclut la preuve.

## Annexe A

# Rappels : théorie de la mesure et de l'intégration

## 1 Résultats généraux de théorie de la mesure

### 1.1 Tribu

**Définition A.1** (Tribu). Soit  $\Omega$  un ensemble et  $\mathcal{F}$  un sous-ensemble de  $\mathcal{P}(\Omega)$ . On dit que  $\mathcal{F}$  est une tribu (ou  $\sigma$ -algèbre) sur  $\Omega$  si les trois conditions suivantes sont vérifiées.

1.  $\Omega \in \mathcal{F}$ .
2.  $\mathcal{F}$  est stable par complémentaire : si  $A \in \mathcal{F}$ , alors  $A^c \in \mathcal{F}$ .
3.  $\mathcal{F}$  est stable par union dénombrable : si  $(A_n)_{n \in \mathbb{N}}$  est une suite de  $\mathcal{F}$ , alors la réunion  $\bigcup_{n=0}^{+\infty} A_n$  appartient à  $\mathcal{F}$ .

Dans un contexte probabiliste, l'espace mesurable  $(\Omega, \mathcal{F})$  est parfois appelé *espace probabilisable* puisqu'on y définit une mesure de probabilité.

**Exemples A.2.** Voici trois exemples classiques de tribus.

1. La tribu grossière ou triviale :  $\mathcal{F} = \{\emptyset, \Omega\}$ .
2. La tribu pleine :  $\mathcal{F} = \mathcal{P}(\Omega)$ , où  $\mathcal{P}(\Omega)$  désigne l'ensemble des parties de  $\Omega$ .
3. La tribu engendrée par une partie  $A$  de  $\Omega$  :  $\mathcal{F} = \{A, A^c, \emptyset, \Omega\}$ .

On déduit immédiatement de la définition qu'une tribu contient l'ensemble vide, est stable par union finie, et par intersection dénombrable ou finie. Bref, une tribu est stable par toutes les opérations ensemblistes souhaitées, du moment que cela concerne un nombre *dénombrable* d'opérations.

**Proposition A.3.** Une intersection quelconque de tribus est une tribu.

*Démonstration.* Soit  $(\mathcal{F}_i)_{i \in I}$  une famille quelconque de tribus.

1. Pour tout  $i$ ,  $\Omega \in \mathcal{F}_i$ , donc  $\Omega \in \bigcap_{i \in I} \mathcal{F}_i$ ;
2. Si  $A \in \bigcap_{i \in I} \mathcal{F}_i$ , alors pour tout  $i \in I$ ,  $A \in \mathcal{F}_i$  et donc  $A^c \in \mathcal{F}_i$  car les  $\mathcal{F}_i$  sont des tribus. Ainsi,  $A^c \in \bigcap_{i \in I} \mathcal{F}_i$ ;
3. Si pour tout  $n \in \mathbb{N}$ ,  $A_n \in \bigcap_{i \in I} \mathcal{F}_i$ , alors pour tout  $i \in I$  et tout  $n \in \mathbb{N}$ ,  $A_n \in \mathcal{F}_i$  donc  $\bigcup_{n \in \mathbb{N}} A_n \in \mathcal{F}_i$  car les  $\mathcal{F}_i$  sont des tribus. Par suite,  $\bigcup_{n \in \mathbb{N}} A_n \in \bigcap_{i \in I} \mathcal{F}_i$ .

□

Notons qu'en général l'union de deux tribus n'est pas une tribu. Il suffit de considérer l'ensemble  $\Omega = \{a, b, c\}$  muni des tribus  $\mathcal{F}_1 = \{\emptyset, \Omega, \{a\}, \{b, c\}\}$  et  $\mathcal{F}_2 = \{\emptyset, \Omega, \{b\}, \{a, c\}\}$ . L'union  $\mathcal{F}_1 \cup \mathcal{F}_2$  n'est alors pas une tribu car elle contient  $\{a\}$  et  $\{b\}$  mais pas  $\{a, b\}$ .

Il est souvent difficile d'expliciter tous les éléments d'une tribu. On s'appuie alors sur des éléments générateurs.

**Définition A.4** (Tribu engendrée par une partie). *Soit  $\mathcal{E}$  un sous-ensemble de  $\mathcal{P}(\Omega)$ . La tribu  $\sigma(\mathcal{E})$  engendrée par  $\mathcal{E}$  est l'intersection de toutes les tribus contenant  $\mathcal{E}$  :*

$$\sigma(\mathcal{E}) = \bigcap_{\tau \text{ tribu : } \mathcal{E} \subset \tau} \tau$$

*C'est la plus petite tribu contenant  $\mathcal{E}$ . La partie  $\mathcal{E}$  est alors appelée système générateur de la tribu  $\sigma(\mathcal{E})$ .*

On remarque que c'est bien une tribu, comme intersection de tribus.

### Exemples A.5.

1. Si  $A \subset \Omega$ , alors la tribu engendrée par  $A$  est  $\sigma(\{A\}) = \sigma(A) = \{\emptyset, \Omega, A, A^c\}$ .
2. Si  $\Omega$  est un espace métrique (même topologique), on appelle tribu borélienne de  $\Omega$  la tribu engendrée par les ouverts de  $\Omega$ . Un borélien est un ensemble appartenant à la tribu mesurable. Sur  $\mathbb{R}$ , on note  $\mathcal{B}(\mathbb{R})$  la tribu des boréliens. On rappelle que c'est aussi la tribu engendrée par les intervalles de la forme  $] - \infty, x]$ ,  $x \in \mathbb{R}$ <sup>1</sup>.

## 1.2 Mesure

**Définition A.6** (Mesure). *Soit  $(\Omega, \mathcal{F})$  un espace mesurable. Une mesure sur  $(\Omega, \mathcal{F})$  est une application  $\mu : \mathcal{F} \rightarrow [0, \infty]$  telle que*

1.  $\mu(\emptyset) = 0$ ,
2. pour toute  $(A_n)_{n \in \mathbb{N}}$  d'ensembles mesurables deux à deux disjoints, on a

$$\mu\left(\bigcup_{n=0}^{\infty} A_n\right) = \sum_{n=0}^{\infty} \mu(A_n).$$

---

1. C'est une conséquence du fait qu'on peut toujours écrire un ouvert de  $\mathbb{R}$  comme union dénombrable d'intervalles (même disjoints), et que tout intervalle peut s'écrire comme union et intersection dénombrable d'ensembles de la forme  $] - \infty, x]$ .

Une mesure  $\mu$  est dite

- $\sigma$ -finie s'il existe une suite croissante  $(A_n)_{n \in \mathbb{N}}$  d'ensembles mesurables telle que  $\Omega = \bigcup_{n \in \mathbb{N}} A_n$  et  $\mu(A_n) < \infty$ .
- finie si  $\mu(\Omega) < \infty$ ,
- de probabilité si  $\mu(\Omega) = 1$ ,

Il est clair qu'une mesure finie est  $\sigma$ -finie (prendre  $A_n = \Omega$ ) et qu'une mesure de probabilité est finie.

**Théorème A.7.** (*Existence et unicité de la mesure de Lebesgue*) Il existe une unique mesure  $\lambda_d$  sur  $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$  telle que

$$\lambda_d\left(\prod_{j=1}^d ]a_j, b_j[ \right) = \prod_{j=1}^d (b_j - a_j).$$

Cela signifie que la mesure de Lebesgue sur un pavé correspond à la notion habituelle de volume. En particulier en dimension 1 on obtient la longueur d'un intervalle :  $\lambda_1(]a, b[) = b - a$ . S'il n'y a pas d'ambiguïté, on abrègera souvent  $\lambda_1$  en  $\lambda$ .

**Exemple A.8.** Construction de mesure. Soit  $(\Omega, \mathcal{F}, \mu)$  un espace mesuré et  $f : \Omega \rightarrow [0, \infty]$  une application mesurable. Alors l'application  $\nu : \mathcal{B}(\mathbb{R}) \rightarrow [0, \infty]$ , définie par

$$\nu(A) = \int_{\mathbb{R}} f(x) \mathbf{1}_A(x) d\mu(x) = \int_A f(x) d\mu(x),$$

est une mesure sur  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ .

Dans ce cas, on dit que  $f$  est la densité de  $\nu$  par rapport à  $\mu$  et cette densité est unique  $\mu$  presque partout. On note alors

$$f = \frac{d\nu}{d\mu}.$$

On utilisera parfois la notation abusive  $d\nu = f d\mu$ .

Par exemple, si  $f(x) = e^{-x} \mathbf{1}_{\mathbb{R}_+}(x)$  et  $\mu$  est la mesure de Lebesgue, alors  $\nu$  est la loi exponentielle.

On peut se demander à quelle condition deux mesures sont associées par une telle fonction. Le théorème de Radon-Nikodym donne la réponse dans le cas de mesures  $\sigma$ -finies. Avant de l'énoncer, on définit la notion d'absolue continuité.

**Définition A.9** (Mesure absolument continue). Soit  $(\Omega, \mathcal{F})$  un espace mesurable, et  $\mu$  et  $\nu$  deux mesures sur  $(\Omega, \mathcal{F})$ . On dit que  $\mu$  est absolument continue par rapport à  $\mu$ , et on note  $\nu \ll \mu$ , si

$$\forall A \in \mathcal{F}, \quad \mu(A) = 0 \implies \nu(A) = 0.$$

**Théorème A.10** (Radon-Nikodym). Soit  $(\Omega, \mathcal{F})$  un espace mesurable, et  $\mu$  et  $\nu$  deux mesures  $\sigma$ -finies sur  $(\Omega, \mathcal{F})$ . Les propositions suivantes sont équivalentes.

1.  $\nu$  est absolument continue par rapport à  $\mu$  :  $\nu \ll \mu$ .
2. Il existe une application mesurable  $f : \Omega \rightarrow [0, \infty]$  telle que  $d\nu = f d\mu$ .

L'application  $f$  est alors unique  $\mu$ -presque partout et est appelée la densité de  $\nu$  par rapport à  $\mu$ .

### 1.3 Lemme de classe monotone

Le lemme de classe monotone est un outil central en théorie de la mesure.

**Définition A.11** (Classe monotone). *Soit  $\mathcal{M}$  une classe de parties d'un ensemble  $\Omega$ . On dit que  $\mathcal{M}$  est une classe monotone (ou  $\lambda$ -système) si*

1.  $\Omega \in \mathcal{M}$ ,
2. si  $A, B \in \mathcal{M}$  sont tels que  $A \subset B$ , alors  $B \setminus A \in \mathcal{M}$ ,
3. si  $(A_n)_{n \in \mathbb{N}}$  est une suite croissante de  $\mathcal{M}$  (au sens où  $A_n \subset A_{n+1}$ ), alors  $\cup_{n \in \mathbb{N}} A_n \in \mathcal{M}$ .

On peut remarquer qu'une classe monotone contient forcément l'ensemble vide, et qu'une tribu est un cas particulier de classe monotone. Par ailleurs, comme pour les tribus, l'intersection de classe monotone est une classe monotone. Ainsi, si  $\mathcal{C}$  est une classe de parties d'un ensemble  $\Omega$ , alors la classe monotone engendrée par  $\mathcal{C}$  est la plus petite (au sens de l'inclusion) classe monotone contenant  $\mathcal{C}$  et est définie par

$$\rho(\mathcal{C}) = \bigcap_{\substack{\mathcal{M} \text{ classe monotone} \\ \mathcal{C} \subset \mathcal{M}}} \mathcal{M}.$$

**Lemme A.12** (Lemme de classe monotone). *Soit  $\mathcal{F}$  une classe de parties d'un ensemble  $\Omega$  stable par intersection finie, c'est-à-dire telle que*

$$A, B \in \mathcal{F} \implies A \cap B \in \mathcal{F}.$$

*Alors  $\rho(\mathcal{F}) = \sigma(\mathcal{F})$ .*

On en déduit un résultat fondamental d'unicité de mesures de probabilité.

**Corollaire A.13.** *Soit  $(\Omega, \mathcal{F})$  un espace probabilisable,  $\mathcal{C}$  une classe de parties de  $\Omega$  stable par intersection finie, et  $\mu$  et  $\nu$  deux mesures de probabilités sur  $(\Omega, \mathcal{F})$ . Si  $\mu$  et  $\nu$  coïncident sur  $\mathcal{C}$ , alors elles coïncident sur  $\sigma(\mathcal{C})$ .*

*Autrement dit,*

$$\{\forall A \in \mathcal{C}, \mu(A) = \nu(A)\} \implies \{\forall A \in \sigma(\mathcal{C}), \mu(A) = \nu(A)\}.$$

*Démonstration.* La démonstration de ce corollaire est immédiate une fois que le lemme de classe monotone est établi. En effet, on pose

$$\mathcal{M} = \{A \in \mathcal{F} : \mu(A) = \nu(A)\}$$

et on vérifie que  $\mathcal{M}$  est une classe monotone. Comme elle contient  $\mathcal{C}$ , elle contient la plus petite classe monotone contenant  $\mathcal{C}$ , à savoir  $\sigma(\mathcal{C})$ .  $\square$



## 2 Transformée de Fourier

La transformée de Fourier est une transformation utile qui permet de 1) régulariser les fonctions, 2) simplifier les calculs. L'objectif de cette courte section est de donner les résultats sur la transformation de Fourier utiles en probabilités.

### 2.1 Définitions et propriétés

**Définition A.14** (Transformée de Fourier). Soit  $f \in L^1(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \lambda_d)$ . La transformée de Fourier de  $f$  est la fonction  $\hat{f}$ , définie pour tout  $\xi = (\xi_1, \dots, \xi_d) \in \mathbb{R}^d$ , par

$$\hat{f}(\xi) = \int_{\mathbb{R}^d} e^{-i\langle \xi, t \rangle} f(t) dt = \int_{\mathbb{R}^d} e^{-i(\xi_1 t_1 + \dots + \xi_d t_d)} f(t) dt,$$

où on a abrégé  $d\lambda_d(t)$  en  $dt$ .

Dans le cas  $d = 1$ , on obtient

$$\hat{f}(\xi) = \int_{\mathbb{R}} e^{-i\xi t} f(t) dt, \quad \xi \in \mathbb{R}.$$

Remarquons que  $\hat{f}$  est bien définie puisque  $|e^{-i\langle \xi, t \rangle} f(t)| = |f(t)|$  et que  $f$  est intégrable par hypothèse. On en déduit en particulier que  $\hat{f}$  est bornée avec  $\|\hat{f}\|_\infty \leq \|f\|_1$ . Cette propriété, associée à la linéarité de la transformée de Fourier, assure que l'application

$$\begin{array}{ccc} L^1(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \lambda_d) & \rightarrow & L^\infty(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \lambda_d) \\ f & \mapsto & \hat{f} \end{array}$$

est une application linéaire continue.

Si  $X$  est un vecteur aléatoire dans  $\mathbb{R}^d$  de densité  $f$  par rapport à la mesure de Lebesgue  $\lambda_d$ , alors sa fonction caractéristique est donnée, pour  $x \in \mathbb{R}^d$ , par

$$\varphi_X(x) = \mathbb{E}[e^{i\langle x, X \rangle}] = \int_{\mathbb{R}^d} e^{i\langle x, t \rangle} f(t) dt = \hat{f}(-x).$$

Donnons à présent quelques propriétés plus générales sur  $\hat{f}$ .

**Proposition A.15.** Soit  $f \in L^1(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \lambda_d)$ . Alors

1.  $\hat{f}$  est continue et  $\lim_{\|\xi\| \rightarrow \infty} \hat{f}(\xi) = 0$ .
2. Si  $x \mapsto \|x\|^k f(x)$  est une fonction intégrable pour un entier  $k \geq 1$ , alors  $\hat{f}$  est de classe  $\mathcal{C}^k(\mathbb{R}, \mathbb{C})$  et pour tout  $\alpha = (\alpha_1, \dots, \alpha_d) \in \mathbb{N}^d$  tel que  $|\alpha| = \sum_{j=1}^d \alpha_j \leq k$ , on a

$$\frac{\partial^{|\alpha|} \hat{f}}{\partial \xi_1^{\alpha_1} \dots \partial \xi_d^{\alpha_d}}(\xi) = \int_{\mathbb{R}^d} e^{-i\langle \xi, t \rangle} (-ix)^\alpha f(x) dx,$$

où  $x^\alpha = (x_1^{\alpha_1}, \dots, x_d^{\alpha_d})$ .

3. (Formule de réciprocité). Soit  $f, g \in L^1(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \lambda_d)$ . Alors

$$\int_{\mathbb{R}^d} f(x) \hat{g}(x) dx = \int_{\mathbb{R}^d} \hat{f}(\xi) g(\xi) d\xi.$$

4. (Convolution). Si  $f$  et  $g$  sont deux fonctions dans  $L^1(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \lambda_d)$ , alors  $\widehat{f * g} = \hat{f} \hat{g}$ .

Le résultat  $\lim_{\|\xi\| \rightarrow \infty} \hat{f}(\xi) = 0$  est connu sous le nom *lemme de Riemann-Lebesgue*. On peut utiliser n'importe quelle norme puisqu'elles sont toutes équivalentes dans  $\mathbb{R}^d$ .

*Démonstration.*

1. C'est une application directe du théorème de continuité sous le signe intégral. Pour le deuxième point, on montre d'abord la convergence vers 0 pour  $f$  dans  $\mathcal{C}_c^\infty(\mathbb{R}^d)$ , puis on étend au cas général en utilisant la densité de  $\mathcal{C}_c^\infty(\mathbb{R}^d)$  dans  $L^1(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \lambda_d)$ .
2. C'est une application du théorème de dérivation sous le signe intégral.
3. On sait que  $\hat{f}$  et  $\hat{g}$  sont bornées, donc les produits  $f\hat{g}$  et  $\hat{f}g$  sont bien dans  $L^1(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \lambda_d)$ . L'égalité est alors une conséquence du théorème de Fubini-Lebesgue :

$$\int_{\mathbb{R}^d} \int_{\mathbb{R}^d} e^{-i\langle x, \xi \rangle} f(x) g(\xi) d\xi dx = \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} e^{-i\langle x, \xi \rangle} f(x) g(\xi) dx d\xi.$$

En divisant des deux côtés par  $(2\pi)^{d/2}$ , on obtient bien

$$\int_{\mathbb{R}^d} f(x) \hat{g}(x) dx = \int_{\mathbb{R}^d} \hat{f}(\xi) g(\xi) d\xi.$$

4. Rappelons que  $f * g$  est définie par

$$(f * g)(x) = \int_{\mathbb{R}^d} f(x - y) g(y) dy.$$

Le théorème de Fubini-Lebesgue donne alors :

$$\int_{\mathbb{R}^d} \left( e^{-i\langle x, \xi \rangle} \int_{\mathbb{R}^d} f(x - y) g(y) dy \right) dx = \int_{\mathbb{R}^d} \left( g(y) \int_{\mathbb{R}^d} e^{-i\langle x, \xi \rangle} f(x - y) dx \right) dy$$

On applique le changement de variables  $t = x - y$  dans l'intégrale en  $x$  (dont le Jacobien est égal à l'identité de  $\mathbb{R}^d$ ) et on obtient

$$\int_{\mathbb{R}^d} e^{-i\langle x, \xi \rangle} f(x - y) dx = e^{-i\langle y, \xi \rangle} \int_{\mathbb{R}^d} e^{-i\langle t, \xi \rangle} f(t) dt = e^{-i\langle y, \xi \rangle} \hat{f}(\xi).$$

Ainsi,

$$\widehat{f * g}(\xi) = \int_{\mathbb{R}^d} g(y) e^{-i\langle y, \xi \rangle} \hat{f}(\xi) dy = \hat{f}(\xi) \hat{g}(\xi).$$

□

## 2.2 Le cas de la densité gaussienne

**Le cas  $d = 1$**  On traite dans cette section la transformée de Fourier de la fonction

$$g_\sigma(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}}, \quad x \in \mathbb{R},$$

avec  $\sigma > 0$ , qui correspond à la densité d'une loi normale  $\mathcal{N}(0, \sigma^2)$ . On utilisant le lien entre transformée de Fourier et fonction caractéristique, on remarque que le calcul a déjà été effectué dans l'exemple 1.33 pour  $X \sim \mathcal{N}(0, 1)$  : dans ce cas on a pour tout  $\xi \in \mathbb{R}$ ,

$$\varphi_X(\xi) = \int_{\mathbb{R}} e^{it\xi} g_\sigma(t) dt = e^{-\frac{\xi^2}{2}}.$$

Ainsi, si  $Y \sim \mathcal{N}(0, \sigma^2)$ , alors  $X = Y/\sigma \sim \mathcal{N}(0, 1)$  et donc

$$\varphi_Y(\xi) = \mathbb{E}[e^{i\xi Y}] = \mathbb{E}[e^{i\xi\sigma X}] = \varphi_X(\sigma\xi) = e^{-\frac{\sigma^2\xi^2}{2}} = \frac{\sqrt{2\pi}}{\sigma} g_{1/\sigma}(\xi).$$

Bref, en revenant aux transformées de Fourier, on vient de montrer que pour tout  $\xi \in \mathbb{R}$ ,

$$\hat{g}_\sigma(\xi) = \frac{\sqrt{2\pi}}{\sigma} g_{1/\sigma}(\xi).$$

**Le cas  $d$  quelconque** Posons maintenant, pour  $x = (x_1, \dots, x_d) \in \mathbb{R}^d$ ,

$$G_\sigma(x) = \frac{1}{(2\pi)^{d/2}\sigma^d} e^{-\frac{\|x\|^2}{2\sigma^2}} = \prod_{j=1}^d g_\sigma(x_j).$$

qui correspond à la densité d'un vecteur aléatoire  $X = (X_1, \dots, X_d)$  composé de variables aléatoires indépendantes de loi normales  $\mathcal{N}(0, \sigma^2)$ .

Comme  $G_\sigma$  s'écrit comme un produit de fonctions d'une seule variable, la transformée de Fourier de  $G_\sigma$  se détermine alors facilement : pour  $\xi = (\xi_1, \dots, \xi_d) \in \mathbb{R}^d$ ,

$$\hat{G}_\sigma(\xi) = \int_{\mathbb{R}^d} e^{i\langle x, \xi \rangle} \prod_{j=1}^d g_\sigma(x_j) dx = \prod_{j=1}^d \left( \int_{\mathbb{R}} e^{ix_j \xi_j} g_\sigma(x_j) dx_j \right) = \prod_{j=1}^d \hat{g}_\sigma(\xi_j),$$

par le théorème de Fubini-Lebesgue. On est ainsi ramené à la transformée de Fourier d'une gaussienne dans le cas  $d = 1$  et le paragraphe précédent donne

$$\hat{G}_\sigma(\xi) = \prod_{j=1}^d \frac{\sqrt{2\pi}}{\sigma} g_{1/\sigma}(\xi_j) = \frac{(2\pi)^{d/2}}{\sigma^d} G_{1/\sigma}(\xi), \quad \xi \in \mathbb{R}^d.$$

Remarquons qu'on obtient en particulier la formule suivante :

$$G_\sigma = \frac{1}{(2\pi)^{d/2}\sigma^d} \hat{G}_{1/\sigma} \tag{A.1}$$

### 2.3 Inversion de Fourier

Si la transformée de Fourier est intéressante, ne serait-ce que parce qu'elle régularise les fonctions, il apparaît utile de pouvoir revenir du monde de Fourier au monde des fonctions classiques. C'est l'objet du théorème d'inversion suivant.

**Théorème A.16.** *Soit  $f \in L^1(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \lambda_d)$  telle que  $\hat{f} \in L^1(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \lambda_d)$ . Alors la fonction  $g$  définie par*

$$g(x) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} e^{i\langle x, \xi \rangle} \hat{f}(\xi) d\xi$$

*vérifie  $f = g$   $\lambda_d$ -presque partout.*

*Démonstration.* Pour  $x \in \mathbb{R}^d$ , on définit la fonction  $e_x : \mathbb{R}^d \rightarrow \mathbb{R}^d$  par  $e_x(\xi) = e^{i\langle x, \xi \rangle}$ ,  $\xi \in \mathbb{R}^d$ . On pose également  $f_\sigma = G_\sigma * f$ . Rappelons qu'alors  $\hat{f}_\sigma = \hat{G}_\sigma \hat{f}$  d'après le quatrième point de la Proposition A.15. On en déduit alors pour tout  $x \in \mathbb{R}^d$ ,

$$\begin{aligned} f_\sigma(x) &= \int_{\mathbb{R}^d} G_\sigma(x-y) f(y) dy \\ &= \int_{\mathbb{R}^d} \frac{1}{(2\pi)^{d/2} \sigma^d} \hat{G}_{1/\sigma}(x-y) f(y) dy \\ &= \frac{1}{(2\pi)^{d/2} \sigma^d} \int_{\mathbb{R}^d} \left( \int_{\mathbb{R}^d} e^{-i\langle x-y, u \rangle} G_{1/\sigma}(u) du \right) f(y) dy, \end{aligned}$$

où on a utilisé la formule (A.1). Le changement de variable  $v = -u$  et la parité de  $G_{1/\sigma}$  donne alors

$$\int_{\mathbb{R}^d} e^{-i\langle x-y, u \rangle} G_{1/\sigma}(u) du = \int_{\mathbb{R}^d} e^{-i\langle y-x, v \rangle} G_{1/\sigma}(v) dv = \int_{\mathbb{R}^d} e^{-i\langle y, v \rangle} e_x(v) G_{1/\sigma}(v) dv = \widehat{e_x G_{1/\sigma}}(y).$$

La formule de réciprocity permet alors de conclure que

$$f_\sigma(x) = \frac{1}{(2\pi)^{d/2} \sigma^d} \int_{\mathbb{R}^d} e^{i\langle x, \xi \rangle} G_{1/\sigma}(\xi) \hat{f}(\xi) d\xi. \quad (\text{A.2})$$

Pour le terme de droite, on a la convergence suivante

$$\frac{1}{\sigma^d} G_{1/\sigma}(\xi) = e^{-\sigma^2 \frac{\|\xi\|^2}{2}} \xrightarrow{\sigma \rightarrow 0} \frac{1}{(2\pi)^{d/2}}.$$

Puisque pour tout  $x, \xi \in \mathbb{R}^d$ ,

$$\left| \frac{1}{\sigma^d} e^{i\langle x, \xi \rangle} G_{1/\sigma}(\xi) \hat{f}(\xi) \right| \leq e^{-\sigma^2 \frac{\|\xi\|^2}{2}} |\hat{f}(\xi)| \leq |\hat{f}(\xi)|,$$

et  $\hat{f} \in L^1$  par hypothèse, le théorème de convergence dominée assure alors que

$$\frac{1}{\sigma^d} \int_{\mathbb{R}^d} e^{i\langle x, \xi \rangle} G_{1/\sigma}(\xi) \hat{f}(\xi) d\xi \xrightarrow{\sigma \rightarrow 0} \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} e^{i\langle x, \xi \rangle} \hat{f}(\xi) d\xi.$$

Enfin, les résultats classiques du produit de convolution donne la convergence dans  $L^1$  :

$$f_\sigma = G_\sigma * f \xrightarrow[\sigma \rightarrow 0]{L^1} f.$$

En passant à la limite quand  $\sigma \rightarrow 0$  dans (A.2), on obtient alors bien que  $f(x)$  correspond pour  $\lambda_d$ -presque tout  $x \in \mathbb{R}^d$  à la fonction

$$x \mapsto \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} e^{i\langle x, \xi \rangle} \hat{f}(\xi) d\xi.$$

□

## 2.4 Transformée de Fourier d'une mesure

**Définition A.17.** Soit  $\mu$  une mesure sur  $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ . La transformée de Fourier de  $\mu$  est la fonction  $\hat{\mu}$  définie pour tout  $\xi \in \mathbb{R}^d$ , par

$$\hat{\mu}(\xi) = \int_{\mathbb{R}^d} e^{-i\langle \xi, x \rangle} d\mu(x).$$

On remarque immédiatement que si  $\mu$  est une mesure de densité  $f \in L^1(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \lambda_d)$  par rapport à la mesure de Lebesgue  $\lambda_d$ , c'est-à-dire  $d\mu = f d\lambda_d$ , alors

$$\hat{\mu}(\xi) = \int_{\mathbb{R}^d} e^{-i\langle \xi, t \rangle} f(t) dt = \hat{f}(\xi).$$

La définition de la transformée de Fourier d'une mesure est donc l'extension naturelle de donnée pour les fonctions de  $L^1(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \lambda_d)$ .

**Proposition A.18.** Si  $\mu = \mu_1 \otimes \cdots \otimes \mu_d$ , alors

$$\hat{\mu} = \hat{\mu}_1 \cdots \hat{\mu}_d,$$

dans le sens où pour tout  $\xi = (\xi_1, \dots, \xi_d) \in \mathbb{R}^d$ ,  $\hat{\mu}(\xi) = \hat{\mu}_1(\xi_1) \cdots \hat{\mu}_d(\xi_d)$ .

*Démonstration.* C'est une conséquence de la définition d'une mesure produit : si  $\xi = (\xi_1, \dots, \xi_d) \in \mathbb{R}^d$ , alors

$$\begin{aligned} \hat{\mu}(\xi) &= \int_{\mathbb{R}^d} e^{-i\langle \xi, t \rangle} d\mu(t_1, \dots, t_d) \\ &= \int_{\mathbb{R}^d} e^{-i\langle \xi, t \rangle} d(\mu_1 \otimes \cdots \otimes \mu_d)(t_1, \dots, t_d) \\ &= \int_{\mathbb{R}} e^{-i\xi_1 t_1} d\mu(t_1) \cdots \int_{\mathbb{R}} e^{-i\xi_d t_d} d\mu(t_d), \end{aligned}$$

ce qui implique que  $\hat{\mu}(\xi) = \hat{\mu}_1(\xi_1) \cdots \hat{\mu}_d(\xi_d)$ . □

**Théorème A.19.** *Soit  $\mu$  est une mesure telle que  $\hat{\mu} \in L^1$ . Alors  $\mu$  admet une densité par rapport à  $\lambda_d$ , qui est égale presque partout à la fonction*

$$x \mapsto \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} e^{i\langle x, \xi \rangle} \hat{\mu}(\xi) \, d\xi.$$

*Démonstration.* On ne détaillera pas entièrement la preuve, mais l'idée est similaire à celle du Théorème A.16 :

- formule donnant la transformée de Fourier de la convolution de deux mesures et formule de réciprocité,
- étude de la transformée de Fourier de  $\mu_\sigma = \mu * G_\sigma$ , où  $G_\sigma$  est la densité d'une gaussienne dans  $\mathbb{R}^d$ ,
- conclusion en faisant tendre  $\sigma$  vers 0.

□

**Corollaire A.20.** *La transformée de Fourier d'une mesure est une application injective : si deux mesures  $\mu$  et  $\nu$  vérifient  $\hat{\mu} = \hat{\nu}$ , alors  $\mu = \nu$ .*

*Démonstration.* Supposons que  $\widehat{\mu - \nu} = 0$ . Comme la fonction nulle est dans  $L^1$ , on déduit que  $\mu - \nu$  est à densité et que cette densité est nulle. Donc  $\mu = \nu$ . □