Dernière édition : 23 avril 2020 Notes prises par Téofil ADAMSKI

FONDEMENT DES PROBABILITÉS

(FPR)

Guillaume POLY

L3 maths recherche Université de Rennes 1

Chapitre 1

Espaces probabilisés

1.1 Introduction

Le concept d'espace probabilisé est lié aux expériences aléatoires : c'est une expérience dont on ne peut pas prédire l'issue de façon certaine. Par exemple, de telles expériences peuvent être un lancer de dé, un lancer de pièce jusqu'à ce qu'on obtienne « face », la durée de vie d'un appareil électronique ou encore la trajectoire d'un grain de pollen porté par le vent.

On appelle *univers des possibles* (souvent noté Ω) l'ensemble des issues de l'expérience aléatoire. On appelle *événement* un ensemble de résultats possibles de l'expérience aléatoires. Par exemple, pour le premier exemple, un événement est « le résultat est pair » représenté par l'ensemble $\{2, 4, 6\}$.

Le but principal de la théorie des probabilités est d'associer à chaque événement A un nombre $\mathbb{P}(A) \in [0, 1]$ qui traduise les chances que cet événement puisse se réaliser. Celui qui a formalisé cette théorie telle que présentée dans ce cours est Kolmogorov en 1933.

Attention. En général, pour un univers donné Ω, on ne peut pas prendre $\mathscr{P}(\Omega)$ pour ensemble des événements car il y a des ensembles $A \subset \Omega$ dont on ne pourra pas calculer la probabilité.

1.2 Espaces probabilisés

Définition 1.1 *(tribu).* Soit Ω un ensemble. On appelle *tribu* sur Ω toute classe de parties $\mathscr{F} \subset \mathscr{P}(\Omega)$ vérifiant

(i) $\Omega \in \mathscr{F}$;

(ii) pour tout $A \in \mathscr{F}$, on a $A^c \in \mathscr{F}$;

(iii) pour toute suite $(A_n)_{n \geqslant 0}$ de $\mathscr{F}$, on a $\bigcup_{n \geqslant 0} A_n \in \mathscr{F}$.

On dit alors que le couple $(\Omega, \mathscr{F})$ est un *espace mesurable.*

◇ Remarque. Les trois points (i), (ii) et (iii) impliquent qu'une tribu $\mathscr{F}$ est aussi stable par union et intersections finies ou infinies dénombrables.

▷ Exemples. – Les classes de parties $\{\emptyset, \Omega\}$ et $\mathscr{P}(\Omega)$ sont des tribus.
– Pour tout $A \subset \Omega$, la classe de parties $\{\emptyset, A, A^c, \Omega\}$ est un tribu.

Propriété 1.2. Si $(\mathscr{F}_i)_{i \in I}$ est une famille de tribus sur Ω, alors $\bigcup_{i \in I} \mathscr{F}_i$ est aussi une tribu sur Ω

Définition-proposition 1.3. Soit $A \subset \mathscr{P}(\Omega)$. On note $\sigma(A)$ l'intersection de toutes les tribus sur Ω contenant A, appelé *tribu engendrée* par A. Cette définition a un sens car $\mathscr{P}(\Omega)$ est une tribu contenant A.

▷ Exemples. – On note $\mathscr{B}(\mathbb{R})$ la tribu engendrée par les ouverts de $\mathbb{R}$, appelée *tribu borélienne* de $\mathbb{R}$. On a

$$\mathscr{B}(\mathbb{R}) = \sigma(\{]-\infty, t] \mid t \in \mathbb{R}\}).$$

– Pour tout $A \subset \Omega$, on a $\sigma(\{A\}) = \{\emptyset, A, A^c, \Omega\}$.

Définition 1.4 *(mesure de probabilités).* Soit $(\Omega, \mathscr{F})$ un espace mesurable. On appelle *probabilité* sur $(\Omega, \mathscr{F})$ toute application $\mathbb{P} \colon \mathscr{F} \to [0, 1]$ vérifiant

(i) $\mathbb{P}(\Omega) = 1$ et $\mathbb{P}(\emptyset) = 0$;

(ii) pour toute suite $(A_n)_{n \geqslant 0}$ de $\mathscr{F}$ d'élements deux à deux disjoints, on a

$$\mathbb{P}\Big(\bigcup_{n \geqslant 0} A_n\Big) = \sum_{n \geqslant 0} \mathbb{P}(A_n).$$

On dit alors que le triplet $(\Omega, \mathscr{F}, \mathbb{P})$ est un *espace probabilisé.*

◇ Remarque. La propriété (ii) est appelé la σ-additivité de $\mathbb{P}$.

▷ EXEMPLES. – Soit Ω un ensemble fini. On pose $\mathscr{F} \coloneqq \mathscr{P}(\Omega)$. Alors l'application

$$\mathbb{P}\colon A \in \mathscr{F} \longmapsto \frac{\operatorname{Card} A}{\operatorname{Card} \Omega} \in [0,1]$$

est une probabilité sur $(\Omega, \mathscr{F})$.

– Soit Ω un ensemble. On pose $\mathscr{F} \coloneqq \mathscr{P}(\Omega)$. Soit $a \in \Omega$. Alors l'application

$$\mathbb{P}\colon A \in \mathscr{F} \longmapsto \mathbb{1}_A(a)$$

est une probabilité sur $(\Omega, \mathscr{F})$.

– La mesure de LEBESGUE λ est une probabilité sur $([0,1], \mathscr{B}([0,1]))$.

◇ REMARQUE. On peut enrichir la tribu des boréliens par $\mathscr{B}([0,1])$ en y ajoutant les ensembles négligeables et on obtient la tribu de LEBESGUE $\mathscr{F}_{\mathrm{c}} = \mathscr{L}([0,1])$ et on peut y définir une mesure complétée $\overline{\lambda}$ sur cette tribu. On peut montrer que $\mathscr{F}_{\mathrm{c}} \neq \mathscr{P}([0,1])$.

On définit sur $\mathbb{R}$ une relation d'équivalence $\mathscr{R}$ telle que $x \mathrel{\mathscr{R}} y \Leftrightarrow x - y \in \mathbb{Q}$ pour tous $x, y \in \mathbb{R}$. On prend V l'ensemble des représentants dans $[0,1]$, appelé ensemble de VITALI. Alors V n'est pas mesurable pour la mesure de LEBESGUE. En effet, supposons que $V \in \mathscr{F}_{\mathrm{c}}$. On pose

$$A \coloneqq \bigcup_{r \in [-1,1] \cap \mathbb{Q}} (V + r) \subset [-1, 2].$$

Comme $V \in \mathscr{F}_{\mathrm{c}}$, on a $A \in \mathscr{F}_{\mathrm{c}}$. Les ensembles $V + r$ avec $r \in \left]-1, 1\right[\cap \mathbb{Q}$ sont deux à deux disjoints, donc la σ-additivité et l'invariance par translation donnent

$$\overline{\lambda}(A) = \sum_{r \in [-1,1] \cap \mathbb{Q}} \overline{\lambda}(V) \leqslant 3,$$

donc $\overline{\lambda}(V) = 0$, donc $\overline{\lambda}(A) = 0$ ce qui est absurde car $[0,1] \subset A$. Justifions ce dernier point. Soit $x \in [0,1]$. Montrons que $x \in A$. Il existe $y \in V$ tel que $x - y \in \mathbb{Q}$. Comme $V \subset [0,1]$, on a $x - y \in \mathbb{Q} \cap [-1,1]$, donc

$$x \in \bigcup_{r \in [-1,1] \cap \mathbb{Q}} (V + r) \subset [-1, 2] = A.$$

PARADOXE DE BANACH-TARSKI. Soient B_1 et B_2 deux boules de $\mathbb{R}^3$ de rayons différents. Alors on peut partitionner B_1 en un nombre fini de morceaux et les réarranger pour former B_2.

Ceci explique pourquoi, en général, on ne peut pas prendre $\mathscr{P}(\Omega)$ comme tribu.

1.3 Propriétés élémentaires des probabilités

PROPRIÉTÉ 1.5. Soit $(\Omega, \mathscr{F}, \mathbb{P})$ un espace probabilisé. Alors

1. pour tous $A, B \in \mathscr{F}$ disjoints, on a $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$;
2. pour tous $A, B \in \mathscr{F}$ tels que $A \subset B$, on a $\mathbb{P}(A) \leqslant \mathbb{P}(B)$;
3. pour tout $A \in \mathscr{F}$, on a $\mathbb{P}(A^{\mathrm{c}}) = 1 - \mathbb{P}(A)$;
4. $\mathbb{P}$ est σ-sous-additive, *i. e.* pour toute suite $(A_n)_{n \geqslant 0}$ de $\mathscr{F}$, on a
$$\mathbb{P}\Big(\bigcup_{n \geqslant 0} A_n\Big) \leqslant \sum_{n \geqslant 0} \mathbb{P}(A_n).$$
5. $\mathbb{P}$ est continue sur des suites croissantes, *i. e.* pour toute suite $(A_n)_{n \geqslant 0}$ croissante de $\mathscr{F}$, on a
$$\mathbb{P}\Big(\bigcup_{n \geqslant 0} A_n\Big) = \lim_{n \to +\infty} \mathbb{P}(A_n).$$
6. $\mathbb{P}$ est continue sur des suites décroissantes, *i. e.* pour toute suite $(A_n)_{n \geqslant 0}$ décroissante de $\mathscr{F}$, on a
$$\mathbb{P}\Big(\bigcap_{n \geqslant 0} A_n\Big) = \lim_{n \to +\infty} \mathbb{P}(A_n).$$

Preuve Voir le cours d'intégration de LEBESGUE. □

THÉORÈME 1.6. Soient $(\Omega, \mathscr{F})$ un espace mesurable et $\mathbb{P}\colon \mathscr{F} \to [0,1]$ finiement additive vérifiant $\mathbb{P}(\Omega) = 1$. Alors les propriétés suivantes sont équivalentes :

(i) $\mathbb{P}$ est un probabilité ;
(ii) $\mathbb{P}$ est σ-sous-additive ;

(iii) $\mathbb{P}$ est continue sur des suites croissantes ;
(iv) $\mathbb{P}$ est continue sur des suites décroissantes ;
(v) $\mathbb{P}$ est continue sur des suites décroissantes tendant vers $\emptyset$, *i. e.* pour toute suite $(A_n)_{n\geqslant 0}$ décroissante de $\mathscr{F}$ telle que $\bigcap_{n\in\mathbb{N}} A_n = \emptyset$, on a

$$\mathbb{P}\Big(\bigcap_{n\geqslant 0} A_n\Big) = \lim_{n\to+\infty} \mathbb{P}(A_n).$$

Preuve Par les propriétés précédentes, les implications (i) $\Rightarrow$ $(k) \in \{(\mathrm{ii}),(\mathrm{iii}),(\mathrm{iv}),(\mathrm{v})\}$ sont vraies. De plus, l'équivalence (iii) $\Leftrightarrow$ (iv) est évidente en passant au complémentaire. L'implication (iv) $\Rightarrow$ (v) est évidente.

• *(ii) $\Rightarrow$ (iii).* On suppose que $\mathbb{P}$ est σ-sous-additive. Soit $(A_n)_{n\in\mathbb{N}}$ une suite décroissante de $\mathscr{F}$. Pour $n\in\mathbb{N}^*$, on pose $B_n = A_n \setminus A_{n-1}$ et $B_0 = A_0$. Alors $\bigcup_{n\in\mathbb{N}} A_n = \bigcup_{n\in\mathbb{N}} B_n$, donc

$$\mathbb{P}\Big(\bigcup_{n\in\mathbb{N}} A_n\Big) = \mathbb{P}\Big(\bigcup_{n\in\mathbb{N}} B_n\Big) \leqslant \sum_{n\in\mathbb{N}} \mathbb{P}(B_n) = \lim_{p\to+\infty} \sum_{n=0}^{p} [\mathbb{P}(A_n) - \mathbb{P}(A_{n-1})] = \lim_{n\to+\infty} \mathbb{P}(A_n).$$

De plus, on a $A_n \in \bigcup_{n\in\mathbb{N}} A_n$ pour tout $n\in\mathbb{N}$, on a

$$\mathbb{P}(A_n) \leqslant \mathbb{P}\Big(\bigcup_{n\in\mathbb{N}} A_n\Big).$$

En laissant tendre n vers $+\infty$, on obtient que

$$\lim_{n\to+\infty} \mathbb{P}(A_n) \leqslant \mathbb{P}\Big(\bigcup_{n\in\mathbb{N}} A_n\Big).$$

• *(v) $\Rightarrow$ (i).* On suppose que $\mathbb{P}$ est continue sur des suites décroissantes tendant vers $\emptyset$. Montrons qu'elle est σ-additive. Soit $(A_n)_{n\in\mathbb{N}}$ une suite de $\mathscr{F}$ d'éléments deux à deux disjoints. Pour $n\in\mathbb{N}$, on pose $B_n = \bigcup_{k=n}^{+\infty} A_k$. Alors la suite $(B_n)_{n\in\mathbb{N}}$ est décroissante d'intersection vide, on a $\lim_{n\to+\infty} \mathbb{P}(B_n) = 0$. Or, pour tout $n\in\mathbb{N}$, on a

$$\begin{aligned}\mathbb{P}\Big(\bigcup_{k\in\mathbb{N}} A_k\Big) &= \mathbb{P}\Big(\bigsqcup_{k=0}^{n} A_k\Big) + \mathbb{P}(B_{n+1})\\ &= \sum_{k=0}^{n} \mathbb{P}(A_k) + \mathbb{P}(B_{n+1}) \longrightarrow \sum_{k=0}^{+\infty} \mathbb{P}(A_k).\end{aligned}$$

• *Conclusion.* On a ainsi montré les implications (i) $\Rightarrow$ (ii) $\Rightarrow$ (iii) $\Rightarrow$ (iv) $\Rightarrow$ (v) $\Rightarrow$ (i) ce qui montre l'équivalence entre toutes ces propositions. $\square$

1.4 Limites supérieure et inférieures d'ensembles

Définition 1.7. Soient $(\Omega, \mathscr{F})$ un espace mesurable et $(A_n)_{n\in\mathbb{N}}$ une suite de $\mathscr{F}$. On pose

$$\limsup_n A_n := \bigcap_{n\in\mathbb{N}} \bigcup_{k\geqslant n} A_k \quad \text{et} \quad \liminf_n A_n := \bigcup_{n\in\mathbb{N}} \bigcap_{k\geqslant n} A_k.$$

Proposition 1.8. Soit $\omega\in\Omega$. Alors

1. $\omega \in \liminf_n A_n$ si et seulement si ω appartient à tous les A_k à partir d'un certain rang ;
2. $\omega \in \limsup_n A_n$ si et seulement si ω appartient à une infinité de A_k.

Preuve Il suffit d'écrire ce que signifie $\omega \in \liminf_n A_n$ à l'aide de quantificateurs. $\square$

Proposition 1.9. Alors

1. $\liminf_n A_n \subset \limsup_n A_n$;
2. $\mathbb{1}_{\liminf_n A_n} = \liminf_n \mathbb{1}_{A_n}$ et $\mathbb{1}_{\limsup_n A_n} = \limsup_n \mathbb{1}_{A_n}$;
3. $(\liminf_n A_n)^{\mathrm{c}} = \limsup_n A_n^{\mathrm{c}}$;
4. $\mathbb{P}(\liminf_n A_n) \leqslant \liminf_n \mathbb{P}(A_n) \leqslant \limsup_n \mathbb{P}(A_n) \leqslant \mathbb{P}(\limsup_n A_n)$.

◇ Remarque. En général, les inégalités du point 4 sont strictes. Par exemples, soient $A, B \in \mathscr{F}$ tels que $A\cap B = \emptyset$ et $0 < \mathbb{P}(A) < \mathbb{P}(B)$. Pour $n\in\mathbb{N}$, on pose $A_n := A$ si n est pair et $A_n := B$ sinon. Alors le suite d'inégalités s'écrit

$$\mathbb{P}(A\cap B) = 0 < \mathbb{P}(A) < \mathbb{P}(B) < \mathbb{P}(A\cup B).$$

1.5 Classe monotones

Définition 1.10. Soit Ω un ensemble. On appelle *classe monotone* sur Ω toute classe de parties $\mathscr{M} \subset \mathscr{P}(\Omega)$ vérifiant

(i) $\Omega \in \mathscr{M}$;

(ii) pour tous $A, B \in \mathscr{M}$ tels que $A \subset B$, on a $B \setminus A \in \mathscr{M}$;

(iii) pour toute suite croissante $(A_n)_{n\geqslant 0}$ de $\mathscr{M}$, on a $\bigcup_{n\geqslant 0} A_n \in \mathscr{M}$.

◇ Remarques. – Une classe monotone n'est, en général, pas une tribu.

– Un intersection de classes monotones est encore une classe monotone. Si $A \subset \mathscr{P}(\Omega)$, on définit alors $m(\Omega)$ l'intersection de toutes les classes monotones contenant A : c'est la plus petite classe monotone contenant A et on l'appelle la *classe monotone engendrée* par A.

– Si on ajoute la stabilité par union ou intersection fini, on a une tribu.

Définition 1.11. On appelle π-*système* toute classe de parties stable par intersection finie.

Théorème 1.12 *(Dynkin)*. Soit $\mathscr{A} \subset \mathscr{P}(\Omega)$ un π-système. Alors $m(\mathscr{A}) = \sigma(\mathscr{A})$.

Preuve On fixe $A \in \mathscr{A}$. Montrons que

$$\mathscr{M}_1 \coloneqq \{B \in m(\mathscr{A}) \mid A \cap B \in m(\mathscr{A})\}$$

est une classe monotone. On a $\Omega \cap A = A \in m(\mathscr{A})$, donc $\Omega \in \mathscr{M}_1$. Soient $C, D \in \mathscr{M}_1$ tels que $C \in D$. Alors

$$(D \setminus C) \cap A = (D \cap A) \setminus (C \cap A) \in m(\mathscr{A}),$$

donc $D \setminus C \in \mathscr{M}_1$. Enfin, soit $(C_n)_{n\geqslant 0}$ une suite croissante de $\mathscr{M}_1$. Alors

$$\Big[\bigcup_{n\geqslant 0} C_n\Big] \cap A = \bigcup_{n\geqslant 0} [C_n \cap A] \in m(\mathscr{A}),$$

donc $\bigcup_{n\geqslant 0} C_n \in \mathscr{M}_1$. Cela montre que $\mathscr{M}_1$ est bien une classe monotone. Comme $\mathscr{A}$ est stable par intersection finie, on a $\mathscr{M}_1 \supset \mathscr{A}$, donc $\mathscr{M}_1 \supset m(\mathscr{A})$. Comme l'inclusion réciproque est vraie, on a $\mathscr{M}_1 = m(\mathscr{A})$. On a donc

$$\forall A \in \mathscr{A},\ \forall B \in m(\mathscr{A}), \quad A \cap B \in m(\mathscr{A}).$$

Fixons $B \in m(\mathscr{A})$. De même, la classe de parties

$$\mathscr{M}_2 \coloneqq \{A \in m(\mathscr{A}) \mid A \cap B \in m(\mathscr{A})\}$$

est une classe monotone vérifiant $\mathscr{M}_2 \supset \mathscr{A}$. On en déduit que $\mathscr{M}_2 = m(\mathscr{A})$. En conclusion, la classe $\mathscr{M}_2$ est stable par intersection finie, donc c'est une tribu contenant A. D'où $\sigma(\mathscr{A}) \subset m(\mathscr{A})$. Réciproquement, la tribu $\sigma(\mathscr{A})$ est une classe monotone contenant $\mathscr{A}$, donc $m(\mathscr{A}) \subset \sigma(A)$. D'où $m(\mathscr{A}) = \sigma(A)$. □

Théorème 1.13. Soit $(\Omega, \mathscr{F})$ un espace mesurable. On suppose qu'il existe un π-système $\mathscr{C}$ tel que $\mathscr{F} = \sigma(\mathscr{C})$. Soient $\mathbb{P}_1$ et $\mathbb{P}_2$ deux mesures de probabilités sur $(\Omega, \mathscr{F})$ coïncidant sur $\mathscr{C}$. Alors $\mathbb{P}_1 = \mathbb{P}_2$.

Preuve On note

$$\mathscr{M} \coloneqq \{A \in \mathscr{F} \mid \mathbb{P}_1(A) = \mathbb{P}_2(B)\}.$$

Il suffit de montrer que $\mathscr{M} = \mathscr{F}$. Pour cela, on va d'abord montrer que $\mathscr{M}$ est une classe monotone. Comme $\mathbb{P}_1(\Omega) = \mathbb{P}_2(\Omega) = 1$, on a $\Omega \in \mathscr{M}$. Soient $A, B \in \mathscr{M}$ tels que $A \subset B$. Alors

$$\mathbb{P}_1(B \setminus A) = \mathbb{P}_1(B) - \mathbb{P}_1(A) = \mathbb{P}_2(B) - \mathbb{P}_2(A) = \mathbb{P}_2(B \setminus A),$$

donc $B \setminus A \in \mathscr{M}$. Enfin, soit $(A_n)_{n\in\mathbb{N}}$ une suite croissante de $\mathscr{M}$. Par continuité des probabilités, on a

$$\mathbb{P}_1\Big(\bigcup_{n\in\mathbb{N}} A_n\Big) = \lim_{n\to+\infty} \mathbb{P}_1(A_n) = \lim_{n\to+\infty} \mathbb{P}_2(A_n) = \mathbb{P}_2\Big(\bigcup_{n\in\mathbb{N}} A_n\Big),$$

donc $\bigcup_{n\in\mathbb{N}} A_n \in \mathscr{M}$. On en déduit que $\mathscr{M}$ est une classe monotone et elle contient $\mathscr{C}$, donc $m(\mathscr{C}) \subset \mathscr{M}$. Comme $\mathscr{C}$ est un π-système, le théorème de Dynkin donne $m(\mathscr{C}) = \sigma(\mathscr{C}) = \mathscr{F}$, donc $\mathscr{F} \subset \mathscr{M}$. Comme l'inclusion réciproque est évidente, on a $\mathscr{M} = \mathscr{F}$. D'où $\mathbb{P}_1 = \mathbb{P}_2$. □

Chapitre 2

Variables aléatoires

2.1 Définition

Définition 2.1. Soient $(\Omega, \mathscr{F})$ et $(E, \mathscr{E})$ deux espaces mesurables. On appelle *variable aléatoire* de $(\Omega, \mathscr{F})$ dans $(E, \mathscr{E})$ toute fonction $X \colon \Omega \to E$ vérifiant

$$\forall A \in \mathscr{E}, \quad X^{-1}(A) \in \mathscr{F}.$$

◇ Remarque. On suppose que $(E, \mathscr{E}) = (\mathbb{R}, \mathscr{B}(\mathbb{R}))$. Si $d = 1$, on dit que X est une variable aléatoire réelle. Dans tous les cas, on dit que X est un vecteur aléatoire réel. On écrit $X = (X_1, \dots, X_d)$. Alors pour tout $i \in [\![1, n]\!]$, la fonction X_i est une variable aléatoire réelle appelée la i-ième marginale de X.

▷ Exemples. Soit $A \in \mathscr{F}$. Alors $\mathbb{1}_A \colon (\Omega, \mathscr{F}) \to (\mathbb{R}, \mathscr{B}(\mathbb{R}))$ est un variable aléatoire réelle. Plus généralement, soient $A_1, \dots, A_p \in \mathscr{F}$ et $a_1, \dots, a_p \in \mathbb{R}$. Alors $\sum_{i=1}^{n} a_i \mathbb{1}_{A_i}$ est une variable aléatoire réelle.

Rappel. Toute variable aléatoire réelle positive est limite simple et croissante de variables aléatoires étagées.

Proposition 2.2. On suppose qu'il existe $\mathscr{C} \subset \mathscr{P}(\mathscr{E})$ tel que $\mathscr{E} = \sigma(\mathscr{C})$. Soit $X \colon (\Omega, \mathscr{F}) \to (E, \mathscr{E})$. Alors X est un variable aléatoire si et seulement si

$$\forall C \in \mathscr{C}, \quad X^{-1}(C) \in \mathscr{F}.$$

Preuve Le sens direct est évident. Réciproquement, on note

$$\mathscr{G} := \{A \in \mathscr{E} \mid X^{-1}(A) \in \mathscr{F}\}.$$

Alors $\mathscr{G}$ est une tribu contenant $\mathscr{C}$, donc $\sigma(\mathscr{C}) \subset \mathscr{G}$. D'où $\mathscr{G} = \sigma(\mathscr{C}) = \mathscr{E}$. □

2.2 Loi d'une variable aléatoire

Définition 2.3. Soit $X \colon (\Omega, \mathscr{F}, \mathbb{P}) \to (E, \mathscr{E})$ une variable aléatoire. La *loi* de X est l'application

$$\mathbb{P}_X \colon \left| \begin{aligned} &\mathscr{E} \longrightarrow [0, 1], \\ &A \longmapsto \mathbb{P}(X \in A). \end{aligned} \right.$$

C'est un probabilité sur $(E, \mathscr{E})$.

Preuve On vérifie aisément qu'il s'agit d'une probabilité, notamment grâce à la σ-additivité de $\mathbb{P}$. □

Définition 2.4. Soient $X \colon (\Omega_1, \mathscr{F}_1, \mathbb{P}_1) \to (E, \mathscr{E})$ et $Y \colon (\Omega_2, \mathscr{F}_2, \mathbb{P}_2) \to (E, \mathscr{E})$ deux variables aléatoires. On dit que X et Y sont égales en loi si $\mathbb{P}_X = \mathbb{P}_Y$.

◇ Remarque. En général, en probabilité, on ne s'intéresse qu'aux égalités en loi, l'espace de probabilité de départ ne joue que peu de rôle.

Propriété 2.5. On suppose qu'il existe un π-système $\mathscr{C}$ tel que $\mathscr{E} = \sigma(\mathscr{C})$. Alors X et Y sont égales en loi si et seulement si $\mathbb{P}_X$ et $\mathbb{P}_Y$ coïncident sur $\mathscr{C}$.

Définition 2.6. Soit $X \colon (\Omega, \mathscr{F}, \mathbb{P}) \to (\mathbb{R}, \mathscr{B}(\mathbb{R}))$. La *fonction de répartition* de X est l'application

$$F_X \colon \left| \begin{aligned} &\mathbb{R} \longrightarrow [0, 1], \\ &t \longmapsto \mathbb{P}(X \leqslant t). \end{aligned} \right.$$

Propriété 2.7. Alors

1. F_X est croissante ;
2. F_X tend vers 0 en $-\infty$ et 1 en $+\infty$;

3. F_X est CADLAG (continue à droite et limitée à gauche), *i. e.* pour tout $t_0 \in \mathbb{R}$, on a
$$\lim_{t \to t_0^+} F_X(t) = F_X(t_0) \quad \text{et} \quad \lim_{t \to t_0^-} F_X(t) = \mathbb{P}(X < t_0).$$
4. X et Y sont égales en loi si et seulement si $F_X = F_Y$;
5. F_X admet un nombre au plus dénombrable de discontinuités, appelés *atomes* de la loi de X.

Preuve 1. La croissance de F_X vient de la croissance de $\mathbb{P}$.
2. Soit $(t_n)_{n\in\mathbb{N}}$ une suite croissante de réels telle que $t_n \to +\infty$. Comme $\mathbb{R} = \bigcup_{n\in\mathbb{N}}]-\infty, t_n]$, la continuité de la probabilité donne
$$1 = \mathbb{P}_X(\mathbb{R}) = \lim_{n\to+\infty} \mathbb{P}_X(]-\infty, t_n]) = \lim_{n\to+\infty} F_X(t_n).$$
De même pour la limite à droite.
3. Soit $t_0 \in \mathbb{R}$. Comme $]-\infty, t_0] = \bigcap_{n\in\mathbb{N}^*}]-\infty, t_0 + 1/n]$, on a $F_X(t_0) = \lim_{n\to+\infty} F_X(t_0 + 1/n)$. Par caractérisation séquentielle, on considère une suite réelle décroissante $(t_n)_{n\in\mathbb{N}}$ telle que $t_n \to t_0$. Alors
$$\lim_{n\to+\infty} F_X(t_n) = F_X(t_0).$$
On considère une suite réelle croissante $(t_n)_{n\in\mathbb{N}}$ telle que $t_n n \to t_0$. Comme $]-\infty, t_0[= \bigcup_{n\in\mathbb{N}}]-\infty, t_n]$, on a
$$\mathbb{P}_X(]-\infty, t_0[) = \lim_{n\to+\infty} F_X(t_n).$$
4. Comme les boréliens de $\mathbb{R}$ sont engendrés par les intervalles $]-\infty, t]$ avec $t \in \mathbb{R}$, les lois $\mathbb{P}_X$ et $\mathbb{P}_Y$ coïncident sur la classe $\{]-\infty, t] \mid t \in \mathbb{R}\}$ si et seulement si elles sont égales sur $\mathscr{B}(\mathbb{R})$. Donc les variables X et Y sont égales en loi si et seulement si $F_X = F_Y$.
5. C'est un résultat plus général sur les fonctions $\mathbb{R} \to \mathbb{R}$ admettant des limites finies à gauche et à droite en tout point. Voir le lemme suivant. □

Lemme 2.8. Si une fonction $f : \mathbb{R} \to \mathbb{R}$ admet une limite à gauche et à droite en tout point, alors elle admet un nombre au plus dénombrable de points de discontinuité.

Preuve Pour $x \in \mathbb{R}$, on note
$$\ell^+(x) \coloneqq \lim_{y\to x^+} f(y) \quad \text{et} \quad \ell^-(x) \coloneqq \lim_{y\to x^-} f(y).$$
Soit $\eta > 0$. On considère les ensembles
$$D \coloneqq \{x \in \mathbb{R} \mid \ell^+(x) \neq f(x) \text{ ou } \ell^-(x) \neq f(x)\} \quad \text{et} \quad D_\eta^\pm \coloneqq \{x \in \mathbb{R} \mid |f(x) - \ell^\pm(x)| \geqslant \eta\}.$$
Montrons que D_η^+ est dénombrable. Soit $x \in D_\eta^+$. Comme f admet une limite à droite, il existe $\alpha > 0$ tel que
$$\forall y \in]x, x+\alpha[, \quad |\ell^+(x) - f(y)| \leqslant \frac{\eta}{3}.$$
L'inégalité triangulaire donne alors
$$\forall y, y' \in]x, x+\alpha[, \quad |f(y) - f(y')| \leqslant \frac{2\eta}{3}.$$
En laissant tendre y' vers y^+, on obtient que
$$\forall y \in]x, x+\alpha[, \quad |f(y) - \ell^+(y)| \leqslant \frac{2\eta}{3} < \eta.$$
D'où $]x, x+\alpha[\in (D_\eta^+)^c$. De même, il existe $\beta > 0$ tel que $]x-\beta, x[\in (D_\eta^+)^c$. Donc $D_\eta^+ \cap]x-\beta, x+\alpha[= \{x\}$, *i. e.* le point x est isolé dans D_η^+. On en déduit que
$$D^+ \coloneqq \{x \in \mathbb{R} \mid f(x) \neq \ell^+(x)\} = \bigcup_{p\in\mathbb{N}^*} D_{1/p}^+$$
est dénombrable. De même, on vérifie que $D^- \coloneqq \{x \in \mathbb{R} \mid f(x) \neq \ell^-(x)\}$ est dénombrable. On en déduit que l'ensemble des points de discontinuité $D = D^+ \cup D^-$ est dénombrable.

• *Autre preuve du point 5.* Si $x \in \mathbb{R}$ un point de discontinuité de F_X, alors $\mathbb{P}(X = x) > 0$. Pour $n \in \mathbb{N}^*$, on pose
$$A_n \coloneqq \{x \in \mathbb{R} \mid \mathbb{P}(X = x) \geqslant 1/n\}$$
de telle sorte que $\operatorname{Card} A_n \leqslant n$. En effet, si tel n'était pas le cas, alors on prend des éléments deux à deux distincts $x_1, \dots, x_{n+1} \in A_n$ et on a
$$\mathbb{P}(X \in \{x_1, \dots, x_{n+1}\}) = \sum_{k=1}^{n+1} \mathbb{P}(X = x_k) \geqslant \frac{n+1}{n} > 1$$
ce qui est impossible. Donc l'ensemble des points de discontinuité $\bigcup_{n\in\mathbb{N}^*} A_n$ est dénombrable. □

Théorème 2.9. Soit $F\colon \mathbb{R} \to [0,1]$ croissante, continue à droite et admettent comme limites 0 et 1 en $-\infty$ et $+\infty$. Alors il existe un espace probabilisé $(\Omega, \mathscr{F}, \mathbb{P})$ et une variable aléatoire $X\colon \Omega \to \mathbb{R}$ tels que $F = F_X$.

Preuve On considère l'espace probabilisé $(\Omega, \mathscr{F}, \mathbb{P}) = (]0,1[, \mathscr{B}(]0,1[), \lambda)$. Pour $u \in]0,1[$, on pose

$$X(u) \coloneqq \inf \{t \in \mathbb{R} \mid F(t) \geqslant u\}.$$

Cette quantité est bien définie puisque, comme $\lim_{t\to+\infty} F(t) = 1$, l'ensemble $\{t \in \mathbb{R} \mid F(t) \geqslant u\}$ est non vide et, comme $\lim_{t\to-\infty} F(t) = 1$, cet ensemble est minorée.

Soit $s \in]0,1[$. Montrons que $\{u \in]0,1[\mid X(u) \leqslant s\} = \{u \in]0,1[\mid u \leqslant F(s)\}$ par double inclusion. Soit $u \in]0,1[$ tel que $X(u) \leqslant s$. La croissance de F donne $F(X(u)) \leqslant F(s)$. Or il existe une suite réelles $(t_n)_{n\in\mathbb{N}}$ telles que $t_n \to X(u)$, $t_n \geqslant X(u)$ et $F(t_n) \geqslant u$ pour tout $n \in \mathbb{N}$. Par la continuité à droite de F, on a $F(t_n) \to F(X(u))$, donc $F(X(u)) \geqslant u$, donc $F(s) \geqslant u$. Réciproquement, soit $u \in]0,1[$ tel que $F(s) \geqslant u$. Alors $s \in \{t \in \mathbb{R} \mid F(t) \geqslant u\}$, donc $X(u) \leqslant s$.

On en déduit alors que X est une variable aléatoire puisque $X^{-1}(]0,s]) =]0, F(s)[$ et les boréliens de $]0,1[$ sont engendrés par les intervalles $]0,s]$ avec $s \in]0,1[$. De plus, on a $\lambda(\{X(u) \geqslant s\}) = \lambda(\{u \leqslant F(s)\}) = F(s)$. ☐

Loi marginale

Définition 2.10. Soit $(X_1, \dots, X_n)$ un vecteur aléatoire de $(\Omega, \mathscr{F}, \mathbb{P})$ dans $(E_1 \times \dots \times E_n, \mathscr{E})$. Pour $i \in [\![1,n]\!]$, on appelle i-ième loi marginale de $(X_1, \dots, X_n)$ la probabilité

$$\mathbb{P}_{X_i} : \left| \begin{array}{l} \mathscr{P}(E_i) \longrightarrow [0,1], \\ \quad A \longmapsto \mathbb{P}((X_1, \dots, X_n) \in E_1 \times \dots \times E_{i-1} \times A \times E_{i+1} \times \dots, E_n). \end{array} \right.$$

◇ Remarque. Deux vecteurs peuvent avoir les mêmes lois marginales sans pour autant avoir la même loi.

Rappel. Soit $(E, \mathscr{E})$ un espace mesuré. On note P l'ensemble des mesures de probabilité sur $(E, \mathscr{E})$. Ce dernier est convexe.

Variables aléatoires discrètes

Définition-proposition 2.11. On dit qu'une variable aléatoire $X\colon (\Omega, \mathscr{F}, \mathbb{P}) \to (E, \mathscr{E})$ est *discrète* s'il existe une partie dénombrable D de E telle que $\mathbb{P}(X \in D) = 1$. Dans ce cas, la loi de X est donnée par

$$\mathbb{P}_X = \sum_{x\in D} \mathbb{P}(X = x)\delta_x.$$

◇ Remarque. La loi d'une variable aléatoire discrète X est entièrement caractérisée par la partie dénombrable $D \subset E$ telle que $\mathbb{P}(X \in D) = 1$ et par la famille $(\mathbb{P}(X = x))_{x\in D}$.

Lois discrètes principales. Soit $X\colon (\Omega, \mathscr{F}, \mathbb{P}) \to (\mathbb{R}, \mathscr{B}(\mathbb{R}))$ une variable aléatoire discrète.

- On dit que X suit une loi constante égale à $c \in \mathbb{R}$ si $\mathbb{P}_X = \delta_c$.
- On dit que X suit une loi uniforme sur un ensemble fini $\{x_1, \dots, x_n\}$ si

$$\mathbb{P}_X = \frac{1}{n} \sum_{i=1}^{n} \delta_{x_i}.$$

- On dit que X suit une loi de Bernoulli de paramètre $p \in [0,1]$ et on note $X \sim \mathrm{Ber}(p)$ si

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

Pour $A \in \mathscr{F}$, on a $\mathbb{1}_A \sim \mathrm{Ber}(\mathbb{P}(A))$.

- On dit que X suit une loi binomiale de paramètre $(n,p) \in \mathbb{N} \times [0,1]$ et on note $X \sim \mathrm{Bin}(n,p)$ si

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

- On dit que X suit une loi de Poisson de paramètre $\lambda > 0$ et on note $X \sim \mathscr{P}(\lambda)$ si

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

Variables aléatoire à densité (ou continue)

THÉORÈME 2.12 *(RADON-NYKODYM).* Soit $(\Omega, \mathscr{F})$ un espace mesurable. On suppose qu'il existe deux mesures σ-finis μ_1 et μ_2 sur $(\Omega, \mathscr{F})$ telle que μ_1 soit absolument continue par rapport à μ_2, noté $\mu_1 \ll \mu_2$, *i. e.*

$$\forall A \in \mathscr{F}, \quad \mu_2(A) = 0 \quad \Longrightarrow \quad \mu_1(A) = 0.$$

Alors il existe une unique fonction f mesurable et μ_2-intégrable telle que

$$\mu_1(A) = \int_A f \, \mathrm{d}\mu_2, \quad \forall A \in \mathscr{F}.$$

On notera cette fonction $\mathrm{d}\mu_1 / \mathrm{d}\mu_2 \coloneqq f$, appelée *densité relative* de μ_2 par rapport à μ_1.

DÉFINITION 2.13. On dit qu'une variable aléatoire $X \colon (\Omega, \mathscr{F}, \mathbb{P}) \to (\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n))$ admet une *densité* par rapport à la mesure de LEBESGUE s'il existe une fonction f borélienne, positive presque-partout et de masse 1 telle que

$$\mathbb{P}(X \in A) = \int_A f(x_1, \dots, x_n) \, \mathrm{d}x_1 \cdots \mathrm{d}x_n, \quad \forall A \in \mathscr{B}(\mathbb{R}^n).$$

$\diamond$ REMARQUE. Une variable aléatoire X est à densité si et seulement si $\mathbb{P}_X \ll \lambda_n$.

LOIS À DENSITÉ PRINCIPALES. Soit $X \colon (\Omega, \mathscr{F}, \mathbb{P}) \to (\mathbb{R}, \mathscr{B}(\mathbb{R}))$ une variable aléatoire à densité. On dit que X suit une loi

– uniforme sur $[a, b]$ avec $a < b$ et on note $X \sim \mathscr{U}([a, b])$ si X admet pour densité $(b-a)^{-1} \mathbb{1}_{[a,b]}$.
– gaussienne de paramètre (μ, σ^2) avec $\mu, \sigma \in \mathbb{R}$ et on note $X \sim \mathscr{N}(\mu, \sigma^2)$ si X admet pour densité

$$x \in \mathbb{R} \longmapsto \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2} \right).$$

– de CAUCHY de paramètre $a > 0$ et on note $X \sim \mathscr{C}(a)$ si X admet pour densité

$$x \in \mathbb{R} \longmapsto \frac{a}{\pi} \frac{1}{a^2 + x^2}.$$

PROPOSITION 2.14. Soit X un variable aléatoire réelle admettant une densité f. Alors

1. F_X est continue ;
2. si f est continue en $t_0 \in \mathbb{R}$, alors F_X est dérivable en t_0 et $F_X'(t_0) = f(t_0)$.

Preuve 1. La fonction F_X est continue à droite. Il suffit de montrer qu'elle est continue à gauche. Soit $x \in \mathbb{R}$. Comme $\lambda(\{x\}) = 0$, on a

$$\mathbb{P}(X = x) = \int_{\{x\}} f(t) \, \mathrm{d}t = 0.$$

Or $F_X(t) \to \mathbb{P}(X < x) = \mathbb{P}(X \leqslant x)$ quand $t \to x^-$, donc F_X est continue à gauche. en X.
2. On suppose que f est continue en $t_0 \in \mathbb{R}$. Soit $\varepsilon > 0$. Alors il existe $\alpha > 0$ tel que

$$\forall h \in \left]-\alpha, \alpha\right[, \quad |f(t_0 + h) - f(t_0)| \leqslant \varepsilon.$$

Alors pour tout $h \in \left]-\alpha, \alpha\right[$, on a

$$\left| \frac{F_X(t_0 + h) - F_X(t_0)}{h} - f(t_0) \right| = \left| \frac{1}{h} \int_{t_0}^{t_0+h} [f(t) - f(t_0)] \, \mathrm{d}t \right| \leqslant \varepsilon$$

ce qui conclut.

$\square$

THÉORÈME 2.15 *(RADEMACHER).* Soit X une variable aléatoire réelles. Si F_X est lipschitzienne, alors X admet une densité f. De plus, on a $F_X' = f$ presque partout.

PROPOSITION 2.16. Soit X une variable aléatoire réelles. Si F_X est continue et de classe $\mathscr{C}^1$ par morceaux, alors X admet une densité f et $F_X' = f$ presque partout.

Chapitre 3
Espérances

3.1 Définition et premières propriétés

Définition 3.1. Soit $X\colon (\Omega, \mathscr{F}, \mathbb{P}) \to (\mathbb{R}, \mathscr{B}(\mathbb{R}))$ une variable aléatoire. On dit que X est *intégrable* si

$$\int_\Omega |X(\omega)| \,\mathrm{d}\mathbb{P}(\omega) < +\infty.$$

Dans ce cas, l'espérance de X est la quantité

$$\mathbb{E}(X) \coloneqq \int_\Omega X(\omega) \,\mathrm{d}\mathbb{P}(\omega).$$

◇ Remarque. Comme $\mathbb{P}$ est une mesure positive finie, les variables aléatoires bornées sont intégrables.

Propriété 3.2. Soient X et Y deux variables aléatoires intégrables et $a, b \in \mathbb{R}$. Alors

1. $aX + bY$ est intégrable et $\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y)$;
2. si $X \leqslant Y$ presque sûrement, alors $\mathbb{E}(X) \leqslant \mathbb{E}(Y)$;
3. $|\mathbb{E}(X)| \leqslant \mathbb{E}(|X|)$; (inégalité triangulaire)
4. si $X \geqslant 0$ presque sûrement et $\mathbb{E}(X) = 0$, alors $X = 0$ presque sûrement ;
5. pour tout $t > 0$, on a

$$\mathbb{P}(|X| > t) \leqslant \frac{\mathbb{E}(|X|)}{t}. \qquad \text{(inégalité de Markov)}$$

◇ Remarque. En général, l'espace $(\Omega, \mathscr{F}, \mathbb{P})$ est inconnu et l'intégrale

$$\int_\Omega |X(\omega)| \,\mathrm{d}\mathbb{P}(\omega)$$

est difficile à évaluer. Heureusement, si la loi de X est connue, alors on peut la calculer grâce au théorème suivant, dit de transfert, que l'on rappel.

Théorème 3.3 *(de transfert)*. Soit $X\colon (\Omega, \mathscr{F}, \mathbb{P}) \to (\mathbb{R}, \mathscr{B}(\mathbb{R}))$ une variable aléatoire.

1. Alors

$$\int_\Omega |X(\omega)| \,\mathrm{d}\mathbb{P}(\omega) = \int_\mathbb{R} |x| \,\mathrm{d}\mathbb{P}_X(x).$$

2. Soit $h\colon \mathbb{R} \to \mathbb{R}$ borélienne. Alors $h(X)$ est $\mathbb{P}$-intégrable si et seulement si h est $\mathbb{P}_X$-intégrable. Dans ce cas, on a

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

Preuve Montrons le second point. On suppose que $h \geqslant 0$. Alors on approche h par une suite croissante de fonctions étagées $(h_n)_{n\in\mathbb{N}}$. On suppose que $h = \mathbb{1}_A$ avec $A \in \mathscr{F}$. Alors

$$\mathbb{E}(h(X)) = \mathbb{P}(X \in A) \quad \text{et} \quad \int_\mathbb{R} h(x) \,\mathrm{d}\mathbb{P}_X(x) = \mathbb{P}_X(A).$$

Par linéarité, c'est vraie pour des fonctions étagées, *i. e.*

$$\forall n \in \mathbb{N}, \quad \mathbb{E}(h_n(X)) = \int_\Omega h_n(X(\omega)) \,\mathrm{d}\mathbb{P}(\omega) = \int_\mathbb{R} h_n(x) \,\mathrm{d}\mathbb{P}_X(x).$$

Le théorème de Beppo Levi donne alors

$$\mathbb{E}(h_n(X)) \longrightarrow \mathbb{E}(h(X)) \quad \text{et} \quad \int_\mathbb{R} h_n(x) \,\mathrm{d}\mathbb{P}_X(x) \longrightarrow \int_\mathbb{R} h(x) \,\mathrm{d}\mathbb{P}_X(x)$$

ce qui conclut si h est borélienne positive.

On ne suppose plus que h est positive. Alors le cas précédent assure

$$\mathbb{E}(|h(X)|) = \int_\mathbb{R} |h(x)| \,\mathrm{d}\mathbb{P}_X(x)$$

ce qui montre l'équivalence. Comme

$$\mathbb{E}(h^{\pm}(X)) = \int_{\mathbb{R}} h^{\pm}(x)\,\mathrm{d}\mathbb{P}_X(x)$$

et $h = h^+ - h^-$, on en déduit le résultat. □

Proposition 3.4. Soit X une variable aléatoire positive. Alors

$$\mathbb{E}(X) = \int_0^{+\infty} \mathbb{P}(X > t)\,\mathrm{d}t.$$

Preuve Pour tout $t \geqslant 0$, on a $\mathbb{P}(X > t) = \mathbb{E}(\mathbb{1}_{\{X>t\}})$. Alors le théorème de Fubini-Tonelli donne

$$\begin{aligned}\int_0^{+\infty} \mathbb{P}(X > t)\,\mathrm{d}t &= \int_0^{+\infty} \int_\Omega \mathbb{1}_{\{X>t\}}(\omega)\,\mathrm{d}\mathbb{P}(\omega)\,\mathrm{d}t \\ &= \int_\Omega \int_0^{+\infty} \mathbb{1}_{\{X>t\}}(\omega)\,\mathrm{d}t\,\mathrm{d}\mathbb{P}(\omega) \\ &= \int_\Omega \int_0^{X(\omega)} \mathrm{d}t\,\mathrm{d}\mathbb{P}(\omega) \\ &= \int_\Omega X(\omega)\,\mathrm{d}\mathbb{P}(\omega) = \mathbb{E}(X).\end{aligned}$$

□

◇ Remarque. En particulier, une variable aléatoire positive X est intégrable si et seulement si $t \longmapsto \mathbb{P}(X > t)$ est intégrable sur $\mathbb{R}_+$. Avec l'inégalité de Markov, on obtient que $\mathbb{P}(X > t) = O(1/t)$. On perd donc un peu d'informations.

Corollaire 3.5. Soit X une variable aléatoire positive. Alors X est intégrable si et seulement si

$$\sum_{n\in\mathbb{N}} \mathbb{P}(X > n) < +\infty.$$

Preuve On fait une comparaison série-intégrale. On a

$$\mathbb{E}(X) = \int_0^{+\infty} \mathbb{P}(X > t)\,\mathrm{d}t = \sum_{k=0}^{+\infty} \int_k^{k+1} \mathbb{P}(X > t)\,\mathrm{d}t.$$

Or la fonction $t \longmapsto \mathbb{P}(X > t)$ est décroissante, donc

$$\sum_{k=0}^{+\infty} \mathbb{P}(X > k+1) \leqslant \mathbb{E}(X) \leqslant \sum_{k=0}^{+\infty} \mathbb{P}(X > t)$$

ce qui permet de conclure. □

Proposition 3.6 *(inégalité de Jensen).* Soient X une variable aléatoire positive et $\varphi\colon \mathbb{R}_+ \to \mathbb{R}$ une fonction convexe. Alors $\varphi(\mathbb{E}(X)) \leqslant \mathbb{E}(\varphi(X))$.

Preuve Comme φ est convexe, pour tout $x \in \mathbb{R}$, on a

$$\varphi(x) = \sup\{ax + b \mid \forall t \in \mathbb{R}, at + b < \varphi(t)\}.$$

Soient $a, b \in \mathbb{R}$ tels que $at + b \leqslant \varphi(t)$ pour tout $t \in \mathbb{R}$. Alors $aX + b \leqslant \varphi(X)$. Par croissance et linéarité de l'espérance on a $a\mathbb{E}(X) + b \leqslant \mathbb{E}(\varphi(x))$. En passant à la borne supérieure sur (a, b), on a $\varphi(\mathbb{E}(X)) \leqslant \mathbb{E}(\varphi(X))$. □

Espérances de lois discrètes classiques. Soit X une variable aléatoire discrète. Alors

– si $X = c$ presque sûrement, alors $\mathbb{P}_X = \delta_c$ et $\mathbb{E}(X) = c$;
– si $X = \mathbb{1}_A$ avec $A \in \mathscr{F}$, alors $X \sim \mathrm{Ber}(\mathbb{P}(A))$ et $\mathbb{E}(X) = \mathbb{P}(A)$;
– si $\mathbb{P}_X = n^{-1}\sum_{k=0}^n \delta_{x_n}$, alors

$$\mathbb{E}(X) = \int_{\mathbb{R}} x\,\mathrm{d}\Big(\frac{1}{n}\sum_{k=0}^n \delta_{x_n}\Big) = \frac{1}{n}\sum_{k=0}^n x_n\ ;$$

– si $X \sim \mathrm{Bin}(n, p)$ avec $n \in \mathbb{N}^*$ et $p \in [0, 1]$, alors $\mathbb{P}_X = \sum_{k=0}^n \binom{n}{k} p^k (1-p)^{n-k} \delta_k$ et

$$\mathbb{E}(X) = \sum_{k=0}^n k\binom{n}{k} p^k (1-p)^{n-k}$$

$$
\begin{aligned}
&= \sum_{k=1}^{n} k\binom{n}{k} p^k(1-p)^{n-k}\\
&= np\sum_{k=0}^{n-1}\binom{n-1}{k} p^k(1-p)^{n-1-k} = np\ ;
\end{aligned}
$$

– si $X \sim \mathscr{P}(\lambda)$ avec $\lambda \in \mathbb{R}$, alors $\mathbb{P}_X = \sum_{k=0}^{+\infty} e^{-\lambda}(k!)^{-1}\lambda^k\delta_k$ et

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

Question. Comment écrire l'intégrale $\int_{\mathbb{R}} h(x)\,\mathrm{d}\mathbb{P}_X(x)$ selon les hypothèses sur $\mathbb{P}_X$?

- *Cas discret.* On suppose que

$$
\mathbb{P}_X = \sum_{i=1}^{+\infty} p_i\delta_{x_i} \quad \text{avec} \quad \sum_{i=1}^{+\infty} p_i = 1.
$$

Pour tout $i \in \mathbb{N}^*$, on a $\mathbb{P}_X(\{x_i\}) = p_i$, donc $\mathbb{P}(X = x_i) = p_i$. On en déduit que, si $h = \mathbb{1}_A$ avec $A \in \mathscr{F}$, on a

$$
\mathbb{E}(h(X)) = \sum_{i=1}^{+\infty} h(x_i)\mathbb{P}(X = x_i).
$$

Par linéarité, cette relation est vraie si h est une fonction étagée puis, par approximation, si $h(X)$ est intégrable.

- *Cas continue.* On suppose que X admet une densité f. Par linéarité puis par approximation, si $h(X)$ est intégrable, on a

$$
\mathbb{E}(h(X)) = \int_{\mathbb{R}} h(x)f(x)\,\mathrm{d}x.
$$

Espérances de lois continues classiques. Soit X une variable aléatoire. Alors

– si $X \sim \mathscr{U}_{[a,b]}$, alors $\mathbb{P}(X \in [a,b]) = 1$, donc X est une variable aléatoire bornée, donc elle est intégrable et

$$
\mathbb{E}(X) = \int_{\mathbb{R}} x\,\mathrm{d}\mathbb{P}_X(x) = \int_{\mathbb{R}} x\frac{\mathbb{1}_{[a,b]}(x)}{b-a}\,\mathrm{d}x = \frac{1}{b-a}\int_a^b x\,\mathrm{d}x = \frac{a+b}{2}\ ;
$$

– si $X \sim \mathscr{E}(\lambda)$ avec $\lambda > 0$, alors

$$
\mathbb{E}(|X|) = \int_{\mathbb{R}} |x|\,\lambda e^{-\lambda x}\mathbb{1}_{[0,+\infty[}(x)\,\mathrm{d}x = \int_0^{+\infty} \lambda x e^{-\lambda x}\,\mathrm{d}x = \frac{1}{\lambda},
$$

donc X est intégrable et, comme $\mathbb{P}(X \geqslant 0) = 1$, on a $X = |X|$, donc $\mathbb{E}(X) = 1/\lambda$;

– si $X \sim \mathscr{N}(\mu,\sigma^2)$, alors

$$
\mathbb{E}(|X|) = \int_{\mathbb{R}} |x|\,\frac{1}{\sqrt{2\pi\sigma^2}}\exp\Big(-\frac{(x-\mu)^2}{2\sigma^2}\Big)\,\mathrm{d}x < +\infty,
$$

donc X est intégrable et

$$
\begin{aligned}
\mathbb{E}(X) &= \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi\sigma^2}}\exp\Big(-\frac{(x-\mu)^2}{2\sigma^2}\Big)\,\mathrm{d}x\\
&= \int_{\mathbb{R}} (x-\mu)\frac{1}{\sqrt{2\pi\sigma^2}}\exp\Big(-\frac{(x-\mu)^2}{2\sigma^2}\Big)\,\mathrm{d}x + \mu\int_{\mathbb{R}} \frac{1}{\sqrt{2\pi\sigma^2}}\exp\Big(-\frac{(x-\mu)^2}{2\sigma^2}\Big) = \mu.
\end{aligned}
$$

3.2 Moments d'une variable aléatoire

Définition 3.7. Soit $X\colon (\Omega,\mathscr{F},\mathbb{P}) \to (\mathbb{R},\mathscr{B}(\mathbb{R}))$ une variable aléatoire. On dit que X admet un *moment* d'ordre $p \geqslant 1$ si $\mathbb{E}(|X|^p) < +\infty$.

Proposition 3.8 *(inégalité de Hölder).* Soient $p, q \geqslant 1$ tels que $1/p + 1/q = 1$. Soient X et Y deux variables aléatoires admettent respectivement des moments d'ordre p et q. Alors XY est intégrable et

$$
\mathbb{E}(|XY|) \leqslant \mathbb{E}(|X|^p)^{1/p}\mathbb{E}(|Y|^q)^{1/q}.
$$

De plus, si $p \leqslant q$, alors $\mathbb{E}(|X|^p)^{1/p} \leqslant \mathbb{E}(|X|^q)^{1/q}$. Autrement dit, si X admet un moment d'ordre q, alors elle admet un moment d'ordre $p \in [1,q]$.

Preuve Se référer au cours d'intégration de Lebesgue. □

Proposition 3.9 *(inégalité de Minkowski).* Soient $p \geqslant 1$ et X et Y deux variables aléatoires admettant un moment d'ordre p. Alors $X+Y$ admet un moment d'ordre p et

$$\mathbb{E}(|X+Y|^p)^{1/p} \leqslant \mathbb{E}(|X|^p)^{1/p} + \mathbb{E}(|Y|^p)^{1/p}.$$

Notation. On note $\|X\|_p := \mathbb{E}(|X|^p)^{1/p}$. L'application $\| \ \|_p$ n'est pas définie. Pour obtenir une norme, il va falloir quotienter par la relation d'équivalence « égale presque partout ». On obtient alors l'espace vectoriel normé $\mathrm{L}^p(\Omega, \mathscr{F}, \mathbb{P})$ qui est un espace de Banach d'après le théorème de Riesz-Fischer.

Définition 3.10. La variance d'un variable aléatoire $X \in \mathrm{L}^2(\Omega, \mathscr{F}, \mathbb{P})$ est la quantité $\mathrm{Var}(X) := \mathbb{E}((X - \mathbb{E}(X))^2)$.

Proposition 3.11. Soient $X, Y \in \mathrm{L}^2(\Omega, \mathscr{F}, \mathbb{P})$ et $\lambda, c \in \mathbb{R}$. Alors
1. $\mathrm{Var}(X+c) = \mathrm{Var}(X)$;
2. $\mathrm{Var}(\lambda X) = \lambda^2 \mathrm{Var}(X)$;
3. si $\mathrm{Var}(X) = 0$, alors $X = \mathbb{E}(X)$ presque sûrement ;
4. $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\,\mathrm{Cov}(X,Y)$ avec $\mathrm{Cov}(X,Y) := \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$.

Définition 3.12. On dit que deux variables aléatoires X et $Y \in \mathrm{L}^2(\Omega, \mathscr{F}, \mathbb{P})$ sont *décorrélées* si $\mathrm{Cov}(X,Y) = 0$.

Variances de lois classiques. Soit X une variable aléatoire. Alors
– si $X \sim \mathrm{Ber}(p)$ avec $p \in \,]0,1[$, alors $X^2 = X$ et $\mathrm{Var}(X) = p(1-p)$;
– si $X \sim \mathrm{Bin}(n,p)$ avec $n \in \mathbb{N}$ et $p \in \,]0,1[$, alors

$$\begin{aligned}\mathbb{E}(X^2) &= \sum_{k=0}^{n} k^2 \binom{n}{k} p^k (1-p)^{n-k}\\ &= \sum_{k=0}^{n} k(k-1)\binom{n}{k} p^k(1-p)^k + \sum_{k=0}^{n} k \binom{n}{k} p^k (1-p)^{n-k}\\ &= \sum_{k=2}^{n} n(n-1)p^2 \binom{n-2}{k-2} p^{k-2}(1-p)^{n-2-(k-2)} + np\\ &= p^2 n(n-1) + np,\end{aligned}$$

donc

$$\mathrm{Var}(X) = p^2 n(n-1) + np - n^2p^2 = np(1-p) \ ;$$

– si $X \sim \mathscr{P}(\lambda)$ avec $\lambda > 0$, alors

$$\mathbb{E}(X^2) = \sum_{k=0}^{+\infty} e^{-\lambda} k^2 \frac{\lambda^k}{k!} = \sum_{k=0}^{+\infty} e^{-\lambda} k(k-1)\frac{\lambda^k}{k!} + \lambda = \lambda^2 + \lambda,$$

donc $\mathrm{Var}(X) = \lambda$;
– si $X \sim \mathscr{G}(p)$ avec $p \in \,]0,1[$, alors une manipulation avec des séries entières donne $\mathrm{Var}(X) = (1-p)/p^2$;
– si $X \sim \mathscr{U}_{[a,b]}$, alors $\mathrm{Var}(X) = (b-a)^2/12$;
– si $X \sim \mathscr{E}(\lambda)$ avec $\lambda > 0$, alors le théorème de convergence dominée donne

$$\mathbb{E}(X^2) = \int_{\mathbb{R}} \lambda x^2 e^{-\lambda x}\,\mathrm{d}x = \lambda \int_0^{+\infty} \frac{\mathrm{d}^2}{\mathrm{d}\lambda^2}[e^{-\lambda x}]\,\mathrm{d}x = \lambda \frac{\mathrm{d}^2}{\mathrm{d}\lambda^2}\left[\int_0^{+\infty} e^{-\lambda x}\right] = \lambda \frac{\mathrm{d}^2}{\mathrm{d}\lambda^2}[1/\lambda] = \frac{2}{\lambda^2},$$

donc $\mathrm{Var}(X) = 1/\lambda^2$;
– si $X \sim \mathscr{N}(\mu, \sigma^2)$, alors le changement de variables $y = (x-\mu)/\sigma$ donne

$$\begin{aligned}\mathbb{E}(X^2) &= \int_{\mathbb{R}} x^2 \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big(-\frac{(x-\mu)^2}{2\sigma^2}\Big)\,\mathrm{d}x\\ &= \int_{\mathbb{R}} (\mu + \sigma y)^2 e^{-y^2/2} \frac{\mathrm{d}y}{\sqrt{2\pi}}\\ &= \mu^2 \int_{\mathbb{R}} e^{-y^2/2}\frac{\mathrm{d}y}{\sqrt{2\pi}} + 2\mu\sigma \underbrace{\int y e^{-y^2/2} \frac{\mathrm{d}y}{\sqrt{2\pi}}}_{=0 \text{ par imparité}} + \sigma^2 \int_{\mathbb{R}} y^2 e^{-y^2/2} \frac{\mathrm{d}y}{\sqrt{2\pi}}\\ &= \mu^2 + \sigma^2,\end{aligned}$$

donc $\mathrm{Var}(X) = \sigma^2$.

THÉORÈME 3.13 *(de caractérisation par les espérances).* Soient $X \coloneqq (X_1, \dots, X_d)$ et $Y \coloneqq (Y_1, \dots, Y_d)$ deux vecteurs aléatoires à valeurs dans $\mathbb{R}^d$. Alors les propositions suivantes sont équivalentes :

(i) X et Y ont le même loi ;

(ii) pour toute $\varphi \in \mathscr{C}_{\mathrm{c}}^0(\mathbb{R}^d, \mathbb{R})$, on a $\mathbb{E}(\varphi(X_1, \dots, X_d)) = \mathbb{E}(\varphi(Y_1, \dots, Y_d))$.

Preuve Le sens direct est évident.

• *En dimension 1.* On sait que $\mathbb{P}_X = \mathbb{P}_Y$ si et seulement si

$$\forall a, b \in \mathbb{R}, \quad a < b \quad \Longrightarrow \quad \mathbb{P}(X \in [a, b]) = \mathbb{P}(Y \in [a, b]).$$

Cela vient du fait que les boréliens de $\mathbb{R}$ sont engendrés par les intervalles fermés. Soient $a, b \in \mathbb{R}$ tels que $a < b$. Soit $\varepsilon > 0$. On considère $\varphi_\varepsilon \in \mathscr{C}_{\mathrm{c}}^0(\mathbb{R})$ telle que

– $\varphi = 1$ sur $[a, b]$;
– $\varphi = 0$ sur $]-\infty, a - \varepsilon] \cup [b + \varepsilon, +\infty[$;
– affine sur $[a - \varepsilon, a]$ et sur $[b, b + \varepsilon]$.

On suppose alors (ii). Par hypothèse, on a $\mathbb{E}(\varphi_\varepsilon(X)) = \mathbb{E}(\varphi_\varepsilon(Y))$. Or

$$\mathbb{E}(\varphi_\varepsilon(X)) = \mathbb{E}(\mathbb{1}_{[a,b]}(X)) + \mathbb{E}(\mathbb{1}_{[a-\varepsilon,a[}(X)\varphi_\varepsilon(X)) + \mathbb{E}(\mathbb{1}_{]b,b+\varepsilon]}(X)\varphi_\varepsilon(X)).$$

Sur $[a - \varepsilon, a[$ et $]b, b - \varepsilon]$, on a $\varphi_\varepsilon \leqslant 1$, donc

$$\mathbb{E}(\mathbb{1}_{[a-\varepsilon,a[}(X)\varphi_\varepsilon(X)) \leqslant \mathbb{P}(X \in [a - \varepsilon, a[) \quad \text{et} \quad \mathbb{E}(\mathbb{1}_{]b,b+\varepsilon]}(X)\varphi_\varepsilon(X)) \leqslant \mathbb{P}(X \in]b, b + \varepsilon]).$$

En posant $\varepsilon = 1/n$ avec $n \in \mathbb{N}^*$, comme la suite $([a - 1/n, a[)_{n \in \mathbb{N}^*}$ est décroissante dont l'intersection est vide, la continuité donne

$$\mathbb{P}(X \in [a - 1/n, a[) \longrightarrow 0.$$

De même, on a $\mathbb{P}(X \in]b, b + \varepsilon]) \longrightarrow 0$. On en déduit que $\mathbb{E}(\varphi_{1/n}(X)) \longrightarrow \mathbb{P}(X \in [a, b])$. On procède de même pour la variable Y et, par passage à la limite, on trouve que $\mathbb{P}(X \in [a, b]) = \mathbb{P}(Y \in [a, b])$.

• *Cas général.* En dimension quelconque, la même preuve avec des pavés marche. Une autre preuve est la suivante. Soit K un compact de $\mathbb{R}^d$. Pour $n \in \mathbb{N}^*$, on pose $O_n \coloneqq \{x \in \mathbb{R}^d \mid \mathrm{d}(x, K) < 1/n\}$ et

$$h_n : \left| \begin{array}{l} \mathbb{R}^d \longrightarrow \mathbb{R}, \\ x \longmapsto \dfrac{\mathrm{d}(x, O_n^{\mathrm{c}})}{\mathrm{d}(x, K) + \mathrm{d}(x, O_n^{\mathrm{c}})}. \end{array} \right.$$

Remarquons que le dénominateur ne s'annule jamais. On montre que la suite $(h_n)_{n \in \mathbb{N}}$ est une suite de fonctions continues à support compact qui converge simplement vers l'indicatrice $\mathbb{1}_K$. Comme $0 \leqslant h_n \leqslant 1$ pour $n \in \mathbb{N}^*$, si on suppose (i), le théorème de convergence dominée donne

$$\mathbb{P}(X \in K) = \mathbb{E}(\mathbb{1}_K(X)) = \lim_{n \to +\infty} \mathbb{E}(h_n(X)) = \lim_{n \to +\infty} \mathbb{E}(h_n(Y)) = \mathbb{E}(\mathbb{1}_K(Y)) = \mathbb{P}(Y \in K).$$

ceci est vrai pour tout compact et donc pour tout pavé de $\mathbb{R}^d$. Ces derniers engendrant les boréliens de $\mathbb{R}^d$, on obtient que $\mathbb{P}_X = \mathbb{P}_Y$. □

◇ REMARQUES. Comme les fonctions $\mathscr{C}^\infty$ à support compact sont denses dans les fonctions $\mathscr{C}^0$ à support compacts pour $\| \, \|_\infty$, la condition (ii) peut être remplacée par

(ii′) pour toute $\varphi \in \mathscr{C}_{\mathrm{c}}^\infty(\mathbb{R}^d, \mathbb{R})$, on a $\mathbb{E}(\varphi(X_1, \dots, X_d)) = \mathbb{E}(\varphi(Y_1, \dots, Y_d))$.

Chapitre 4
Fonctions caractéristiques

4.1 Définition et théorème de caractérisation

Définition 4.1. Soit $X \coloneqq (X_1, \dots, X_d)$ une vecteur aléatoire à valeurs dans $(\mathbb{R}^d, \mathscr{B}(\mathbb{R}^d))$. La *fonction caractéristique* de X est la fonction

$$\varphi_X : \left| \begin{array}{rcl} \mathbb{R}^d & \longrightarrow & \mathbb{R}, \\ (t_1, \dots, t_d) & \longmapsto & \mathbb{E}(\exp[i(t_1 X_1 + \dots + t_n X_n)]). \end{array} \right.$$

Théorème 4.2 *(de caractérisation).* Soient $X \coloneqq (X_1, \dots, X_d)$ et $Y \coloneqq (Y_1, \dots, Y_d)$ deux vecteurs aléatoires à valeurs dans $(\mathbb{R}^d, \mathscr{B}(\mathbb{R}^d))$ tels que $\varphi_X = \varphi_Y$. Alors X et Y ont la même loi.

Lemme 4.3. Soit $X \sim \mathscr{N}(0,1)$. Alors $\varphi_X(t) = e^{-t^2/2}$ pour tout $t \in \mathbb{R}$.

Preuve Pour tout $t \in \mathbb{R}$, on a

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

Après application du théorème de convergence dominée avec la domination

$$\forall t, x \in \mathbb{R}, \quad \left| \frac{\mathrm{d}}{\mathrm{d}t} [e^{itx} e^{-x^2/2}] \right| = |ixe^{itx} e^{-x^2/2}| \leqslant |x|\, e^{-x^2/2} \in \mathrm{L}^1(\mathbb{R}, \mathrm{d}x),$$

la fonction φ_X est dérivable et, pour tout $t \in \mathbb{R}$, on a

$$\begin{aligned} \varphi_X'(t) &= \int_{\mathbb{R}} ixe^{itx} e^{-x^2/2} \frac{\mathrm{d}x}{\sqrt{2\pi}} \\ &= i \int_{\mathbb{R}} xe^{-x^2/2} \times e^{itx} \frac{\mathrm{d}x}{\sqrt{2\pi}} \\ &= i \left(\left[-e^{-x^2/2} \frac{e^{itx}}{\sqrt{2\pi}} \right]_{\mathbb{R}} + \int_{\mathbb{R}} e^{-x^2/2} ite^{itx} \frac{\mathrm{d}x}{\sqrt{2\pi}} \right) \\ &= t\varphi_X(t). \end{aligned}$$

En résolvant cette équation différentielle, il existe $\lambda \in \mathbb{R}$ tel que $\varphi_X(t) = \lambda e^{-t^2/2}$ pour tout $t \in \mathbb{R}$. Pour $t = 0$, on a $\varphi_X(0) = \mathbb{E}(1) = 1$, donc $\lambda = 1$. D'où $\varphi_X(t) = e^{-t^2/2}$ pour tout $t \in \mathbb{R}$. $\square$

Preuve du théorème Soient $h \in \mathscr{C}_c^\infty(\mathbb{R})$ et $\varepsilon > 0$. On pose

$$h_\varepsilon : \left| \begin{array}{l} \mathbb{R} \longrightarrow \mathbb{R}, \\ x \longmapsto \displaystyle\int_{\mathbb{R}} h(x-y) e^{-y^2/2\varepsilon^2} \frac{\mathrm{d}y}{\sqrt{2\pi\varepsilon^2}}. \end{array} \right.$$

Pour tout $x \in \mathbb{R}$, on a

$$\begin{aligned} |h(x) - h_\varepsilon(x)| &= \left| \int_{\mathbb{R}} [h(x-y) - h(x)] \frac{e^{-y^2/2\varepsilon^2}}{\sqrt{2\pi\varepsilon^2}} \,\mathrm{d}y \right| \\ &\leqslant \int_{\mathbb{R}} |h(x-y) - h(x)| \frac{e^{-y^2/2\varepsilon^2}}{\sqrt{2\pi\varepsilon^2}} \,\mathrm{d}y \\ &\leqslant \|h'\|_\infty \int_{\mathbb{R}} |y| \frac{e^{-y^2/2\varepsilon^2}}{\sqrt{2\pi\varepsilon^2}} \,\mathrm{d}y && \text{(inégalité des accroissememnts finis)} \\ &= \|h'\|_\infty\, \varepsilon \int_{\mathbb{R}} |u|\, e^{-u^2/2} \frac{\mathrm{d}u}{\sqrt{2\pi}} = O(\varepsilon). && \text{(changement de variables } u = y/\varepsilon) \end{aligned}$$

Donc $\|h - h_\varepsilon\|_\infty \to 0$ quand $\varepsilon \to 0$. De plus, on a

$$\mathbb{E}(h_\varepsilon(X)) = \int_{\mathbb{R}} h_\varepsilon(x) \,\mathrm{d}\mathbb{P}_X(x)$$

$$= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} h(x-y) e^{-y^2/2\varepsilon^2} \frac{\mathrm{d}y}{\sqrt{2\pi\varepsilon^2}} \right) \mathrm{d}\mathbb{P}_X(x)$$
$$= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} h(u) e^{-(x-u)^2/2\varepsilon^2} \frac{\mathrm{d}u}{\sqrt{2\pi\varepsilon^2}} \right) \mathrm{d}\mathbb{P}_X(x). \qquad \text{(changement de variables } u = x-y\text{)}$$

L'idée est de remplacer le terme $e^{-(x-u)^2/2\varepsilon^2}$ par l'intégrale

$$\int_{\mathbb{R}} e^{i(x-u)z/\varepsilon} e^{-z^2/2} \frac{\mathrm{d}z}{\sqrt{2\pi}}.$$

On obtient alors

$$\mathbb{E}(h_\varepsilon(X)) = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} h(u) \left(\int_{\mathbb{R}} e^{i(x-u)z/\varepsilon} e^{-z^2/2} \frac{\mathrm{d}z}{2\pi\varepsilon} \right) \mathrm{d}u \right) \mathrm{d}\mathbb{P}_X(x)$$
$$= \int_{\mathbb{R}^3} F(x,u,z)\, \mathrm{d}\mathbb{P}_X(x)\, \mathrm{d}z\, \mathrm{d}u \quad \text{avec} \quad F(x,u,z) \coloneqq h(x) e^{i(x-u)z/\varepsilon} \frac{e^{-z^2/2}}{2\pi\varepsilon}.$$

Or pour tout $(x,u,z) \in \mathbb{R}^3$, on a

$$|F(x,u,z)| \leqslant |h(u)| \frac{e^{-z^2/2}}{2\pi\varepsilon} \in \mathrm{L}^1(\mathbb{R}^3, \mathrm{d}\mathbb{P}_X(x) \otimes \mathrm{d}u \otimes \mathrm{d}z).$$

puisque h est à support compact. Le théorème de FUBINI donne alors

$$\mathbb{E}(h_\varepsilon(X)) = \int_{\mathbb{R}} \int_{\mathbb{R}} h(u) \frac{e^{-iu/\varepsilon}}{\sqrt{2\pi\varepsilon^2}} \frac{e^{-z^2/2}}{\sqrt{2\pi}} \left(\int_{\mathbb{R}} e^{xz/\varepsilon}\, \mathrm{d}\mathbb{P}_X(x) \right) \mathrm{d}u\, \mathrm{d}z$$
$$= \int_{\mathbb{R}} \int_{\mathbb{R}} \frac{h(u) e^{-iu/\varepsilon} e^{-z^2/2}}{\sqrt{2\pi\varepsilon^2}\sqrt{2\pi}} \varphi_X(z/\varepsilon)\, \mathrm{d}u\, \mathrm{d}z.$$

Si $\varphi_X = \varphi_Y$, alors $\mathbb{E}(h_\varepsilon(X)) = \mathbb{E}(h_\varepsilon(Y))$. En laissant tendre ε vers 0^+, on montre que $\mathbb{E}(h(X)) = \mathbb{E}(h(Y))$. Le théorème de caractérisation par les espérances assure alors que $\mathbb{P}_X = \mathbb{P}_Y$. □

PROPRIÉTÉ 4.4. Soit $X \coloneqq (X_1, \dots, X_d)$ un vecteur aléatoire. Alors
1. $\varphi_X(0) = 1$;
2. pour tout $t \in \mathbb{R}^d$, on a $\varphi_X(-t) = \overline{\varphi_X(t)}$;
3. pour tout $t \in \mathbb{R}^d$, on a $|\varphi(t)| \leqslant 1$;
4. φ_X est uniformément continue sur $\mathbb{R}^d$.

PROPOSITION 4.5. 1. Soit $X \in \mathrm{L}^p(\Omega, \mathscr{F}, \mathbb{P})$. Alors φ_X est de classe $\mathscr{C}^p$ et vérifie

$$\varphi_X^{(k)}(0) = i^k \mathbb{E}(X^k), \quad k \in [\![0,p]\!].$$

2. Soit X une variable aléatoire telle que φ_X soit de classe $\mathscr{C}^p$. Alors $X \in \mathrm{L}^q(\Omega, \mathscr{F}, \mathbb{P})$ avec $q \coloneqq 2\lfloor p/2 \rfloor$.

Preuve 1. Par récurrence, montrons que, pour $k \in [\![0,p]\!]$, la fonction φ_X est de classe $\mathscr{C}^k$ et

$$\varphi_X^{(k)}(t) = i^k \mathbb{E}(X^k e^{itX}), \quad t \in \mathbb{R}.$$

Si $k = 0$, c'est évident. Soit $k \leqslant p-1$. On suppose que la propriété est vraie au rang k. Pour tous $\omega, t \in \mathbb{R}$, on a

$$\left| \frac{\mathrm{d}}{\mathrm{d}t} [X^k(\omega) e^{itX(\omega)}] \right| = |iX^{k+1}(\Omega) e^{itX(\omega)}| \leqslant |X^{k+1}(\omega)| \in \mathrm{L}^1(\mathbb{R}, \mathrm{d}\mathbb{P}(\omega))$$

car $k \leqslant p-1$. Le théorème de convergence dominée assure alors que la fonction $\varphi_X^{(k)}$ est dérivable et

$$\varphi_X^{(k+1)}(t) = i^{k+1} \mathbb{E}(X^{k+1} e^{itX}), \quad t \in \mathbb{R}.$$

De plus, la fonction $(\omega, t) \longmapsto X^{k+1}(\omega) e^{itX(\omega)}$ est continue par rapport à t et est dominée par $|X^{k+1}|$. Par le même théorème, la fonction $\varphi_X^{(k+1)}$ est donc continue ce qui terme la récurrence. Ceci montre que φ_X est de classe $\mathscr{C}^p$.

2. On procède par récurrence sur p. Traitons alors uniquement le cas $p = 2$. On suppose que la fonction φ_X est de classe $\mathscr{C}^2$. Par un développement de TAYLOR, on sait que

$$\forall t \in \mathbb{R}, \quad \frac{\varphi_X(t+h) + \varphi_X(t-h) - 2\varphi_X(t)}{h^2} \xrightarrow[h \to 0]{} \varphi_X''(t).$$

En particulier, pour $t = 0$, on a

$$\mathrm{Re}\left(\frac{2\varphi_X(0) - \varphi_X(h) - \varphi_X(-h)}{h^2} \right) \xrightarrow[h \to 0]{} -\mathrm{Re}\, \varphi_X''(0).$$

Or pour tout $h \in \mathbb{R}^*$, on a

$$\frac{2\varphi_X(0) - \varphi_X(h) - \varphi_X(-h)}{h^2} = \frac{2 - \mathbb{E}(i^{ihX} + e^{-ihX})}{h^2} = 2\mathbb{E}\Big(\underbrace{\frac{1 - \cos hX}{h^2}}_{\geqslant 0}\Big).$$

Le lemme de Fatou donne

$$+\infty > \liminf_{h \to 0} \mathbb{E}\Big(\frac{1 - \cos hX}{h^2}\Big) \geqslant \mathbb{E}\Big(\liminf_{h \to 0} \frac{1 - \cos hX}{h^2}\Big) = \mathbb{E}(X^2/2).$$

On en déduit que $\mathbb{E}(X^2) < +\infty$ ce qu'on souhaitait démontrer.

□

Fonctions caractéristiques des lois classiques. Soient X une variable aléatoire et $t \in \mathbb{R}$. Alors

– si $X \sim \operatorname{Ber}(p)$ avec $p \in]0,1[$, alors

$$\varphi_X(t) = e^{it}p + (1-p) \ ;$$

– si $X \sim \operatorname{Bin}(n,p)$ avec $n \in \mathbb{N}^*$ et $p \in]0,1[$, alors

$$\varphi_X(t) = \sum_{k=0}^{n} e^{itk} \binom{n}{k} p^k (1-p)^{n-k} = (e^{it}p + 1 - p)^n \ ;$$

– si $X \sim \mathscr{P}(\lambda)$ avec $\lambda > 0$, alors

$$\varphi_X(t) = \sum_{k=0}^{+\infty} e^{itk} \frac{\lambda^k}{k!} e^{-\lambda} = \exp[\lambda(e^{it} - 1)] \ ;$$

– si $X \sim \mathscr{G}(\lambda)$ avec $p \in]0,1[$, alors

$$\varphi_X(t) = \sum_{k=1}^{+\infty} e^{itk} (1-p)^{k-1} p = \frac{pe^{it}}{1 - e^{it}(1-p)} = \frac{p}{e^{-it} - 1 + p} \ ;$$

– si $X \sim \mathscr{U}_{[a,b]}$ avec $a < b$, alors

$$t \neq 0 \quad \Longrightarrow \quad \varphi_X(t) = \frac{1}{b-a} \int_a^b e^{itx} \, \mathrm{d}x = \frac{e^{itb} - e^{ita}}{it(b-a)} = e^{it(a+b)/2} \operatorname{sinc}\Big(\frac{b-a}{2} t\Big) \ ;$$

– si $X \sim \mathscr{E}(\lambda)$ avec $\lambda > 0$, alors

$$\varphi_X(t) = \int_0^{+\infty} e^{itx} \lambda e^{-\lambda x} \, \mathrm{d}x = \lambda \int_0^{+\infty} e^{(it-\lambda)x} \, \mathrm{d}x = \lambda \left[\frac{e^{(it-\lambda)x}}{it - \lambda} \right]_0^{+\infty} = \frac{1}{1 - it\lambda^{-1}}.$$

– Par le lemme, on sait que, si $X \sim \mathscr{N}(0,1)$, alors $\varphi_X(t) = e^{-t^2/2}$. On suppose que $X \sim \mathscr{N}(\mu, \sigma^2)$ avec $\mu, \sigma \in \mathbb{R}$. En effectuant le changement de variables $y = (x - \mu)/\sigma$, on a

$$\begin{aligned} \varphi_X(t) &= \int_{\mathbb{R}} e^{itx} \exp\Big(-\frac{(x-\mu)^2}{\sigma^2}\Big) \frac{\mathrm{d}x}{\sqrt{2\pi\sigma^2}} \\ &= \int_{\mathbb{R}} e^{it(\sigma^2 y + \mu)} e^{-y^2/2} \frac{\mathrm{d}x}{\sqrt{2\pi\sigma^2}} \\ &= e^{it\mu} \int_{\mathbb{R}} e^{it\sigma y} e^{-y^2/2} \frac{\mathrm{d}x}{\sqrt{2\pi\sigma^2}} \\ &= e^{it\mu} e^{-t^2\sigma^2/2}. \end{aligned}$$

4.2 Formule d'inversion de Fourier

4.2.1 Énoncé

Définition 4.6. Soit $f \in \mathrm{L}^1(\mathbb{R}, \mathrm{d}x)$. La *transformée de Fourier* de f est la fonction

$$\mathscr{F}[f] \colon \left| \begin{array}{l} \mathbb{R} \longrightarrow \mathbb{R}, \\ t \longmapsto \displaystyle\int_{\mathbb{R}} e^{-itx} f(x) \, \mathrm{d}x. \end{array} \right.$$

THÉORÈME 4.7 *(d'inversion de FOURIER).* Soit $f \in \mathrm{L}^1(\mathbb{R})$ telle que $\mathscr{F}[f] \in \mathrm{L}^1(\mathbb{R})$. Alors

$$f(x) = \int_{\mathbb{R}} e^{itx} \mathscr{F}[f](t) \frac{\mathrm{d}t}{2\pi} = (2\pi)^{-1} \mathscr{F}[\mathscr{F}[f]](-x), \quad x \in \mathbb{R}.$$

Preuve Soient $f, g \in \mathrm{L}^1(\mathbb{R})$. Alors $f * g \in \mathrm{L}^1(\mathbb{R})$ et $\mathscr{F}[f * g] = \mathscr{F}[f] \times \mathscr{F}[g]$. Ce dernier résultat découle du théorème de FUBINI. Soit $\varepsilon > 0$. On pose

$$\gamma_\varepsilon : \left| \begin{array}{l} \mathbb{R} \longrightarrow \mathbb{R}, \\ x \longmapsto (2\pi\varepsilon^2)^{-1/2} e^{-x^2/2\varepsilon}. \end{array} \right.$$

Soit $x \in \mathbb{R}$. Alors

$$\begin{aligned} \mathscr{F}[f * \gamma_\varepsilon](x) &= \mathscr{F}[f](t) \mathscr{F}[\gamma_\varepsilon](t) \\ &= \mathscr{F}[f](t) \int_{\mathbb{R}} e^{-ixt} \gamma_\varepsilon(x) \,\mathrm{d}x \\ &= \mathscr{F}[f](t) e^{-t^2\varepsilon^2/2}, \end{aligned}$$

donc

$$\begin{aligned} \frac{1}{2\pi} \mathscr{F}[\mathscr{F}[f * \gamma_\varepsilon]](-x) &= \frac{1}{2\pi} \int_{\mathbb{R}} e^{ixt} \mathscr{F}[f * \gamma_\varepsilon](t) \,\mathrm{d}t \\ &= \frac{1}{2\pi} \int_{\mathbb{R}} e^{ixt} \mathscr{F}[f](t) e^{-t^2\varepsilon^2/2} \,\mathrm{d}t \\ &= \frac{1}{2\pi} \int_{\mathbb{R}} \left(\int_{\mathbb{R}} f(y) e^{-iyt} \,\mathrm{d}y \right) e^{-t^2\varepsilon^2/2} e^{ixt} \,\mathrm{d}t. \end{aligned}$$

Or l'application $(t, y) \longmapsto f(y) e^{-iyt} e^{-t^2\varepsilon^2/2} e^{ixt}$ est intégrale, donc e théorème de FUBINI donne alors

$$\begin{aligned} \frac{1}{2\pi} \mathscr{F}[\mathscr{F}[f * \gamma_\varepsilon]](-x) &= \frac{1}{2\pi} \int_{\mathbb{R}} f(y) \left(\int_{\mathbb{R}} e^{-t^2\varepsilon^2/2} e^{i(x-y)t} \,\mathrm{d}t \right) \mathrm{d}y \\ &= \frac{1}{2\pi} \sqrt{\frac{2\pi}{\varepsilon^2}} \int_{\mathbb{R}} f(y) \left(\int_{\mathbb{R}} \frac{e^{-t^2\varepsilon^2/2}}{\sqrt{2\pi/\varepsilon^2}} e^{i(x-y)t} \,\mathrm{d}t \right) \mathrm{d}y \\ &= \frac{1}{2\pi} \sqrt{\frac{2\pi}{\varepsilon^2}} \int_{\mathbb{R}} f(y) e^{-(x-y)^2/2\varepsilon^2} \,\mathrm{d}y \\ &= \frac{1}{\sqrt{2\pi\varepsilon^2}} \int_{\mathbb{R}} f(y) e^{-(x-y)^2/2\varepsilon^2} \,\mathrm{d}y = f * \gamma_\varepsilon(x). \end{aligned}$$

D'une part, on a $\|f * \gamma_\varepsilon - f\|_1 \to 0$ quand $\varepsilon \to 0$. Quitte à extraire une sous-suite de $(\gamma_\varepsilon)_{\varepsilon > 0}$, on peut supposer que $f * \gamma_\varepsilon(x) \to f(x)$ pour presque tout $x \in \mathbb{R}$. D'autre part, le théorème de convergence dominée assure

$$\frac{1}{2\pi} \mathscr{F}[\mathscr{F}[f * \gamma_\varepsilon]](x) \longrightarrow \mathscr{F}[\mathscr{F}[f]](x)$$

pour presque tout $x \in \mathbb{R}$. Par unicité de la limite, on a $\mathscr{F}[F[f]](x) = f(x)$ pour presque tout $x \in \mathbb{R}$ ☐

4.2.2 Calcul de la fonction caractéristique d'une loi de CAUCHY

Soit X une variable aléatoire suivant une loi de CAUCHY de paramètre 1. Soit $\xi \in \mathbb{R}$. On veut calculer sa fonction caractéristique

$$\mathbb{E}(e^{i\xi X}) = \int_{\mathbb{R}} \frac{e^{i\xi x}}{\pi(x^2 + 1)} \,\mathrm{d}x.$$

On considère la fonction $g \colon x \in \mathbb{R} \longmapsto e^{-|x|}$. Alors $g \in \mathrm{L}^1(\mathbb{R}, \mathrm{d}x)$. Calculons la transformée de FOURIER de g. Pour tout $t \in \mathbb{R}$, on a

$$\begin{aligned} \mathscr{F}[g](t) &= \int_{\mathbb{R}} e^{-ixt} g(x) \,\mathrm{d}x \\ &= \int_{\mathbb{R}_+} e^{-ixt} e^{-x} \,\mathrm{d}x + \int_{\mathbb{R}_-} e^{-ixt} e^{x} \,\mathrm{d}x \\ &= \left[-\frac{e^{-(1+it)x}}{1 + it} \right]_0^{+\infty} + \left[\frac{e^{(1-it)x}}{1 - it} \right]_{-\infty}^0 \\ &= \frac{1}{1 + it} + \frac{1}{1 - it} = \frac{2}{1 + t^2}. \end{aligned}$$

Alors $\mathscr{F}[g] \in \mathrm{L}^1(\mathbb{R})$. La formule d'inversion donne alors, pour tout $x \in \mathbb{R}$,

$$\begin{aligned} g(x) &= \frac{1}{2\pi} \int_{\mathbb{R}} e^{itx} \mathscr{F}[g](t) \, \mathrm{d}t \\ &= \frac{1}{2\pi} \int_{\mathbb{R}} e^{itx} \frac{2}{1+t^2} \, \mathrm{d}t \\ &= \frac{1}{\pi} \int_{\mathbb{R}} \frac{e^{itx}}{1+t^2} \, \mathrm{d}t. \end{aligned}$$

On en déduit alors que

$$\mathbb{E}(e^{i\xi X}) = e^{-|\xi|}.$$

Maintenant, on suppose que le paramètre vaut $a \in \mathbb{R}$. En effectuant le changement de variables $x = ay$, on obtient que

$$\begin{aligned} \mathbb{E}(e^{i\xi X}) &= \int_{\mathbb{R}} \frac{e^{i\xi x} a}{\pi(x^2+a^2)} \, \mathrm{d}x \\ &= \int_{\mathbb{R}} e^{ia\xi y} \frac{a^2}{a^2 \pi (y^2+1)} \, \mathrm{d}y = e^{-a|\xi|}. \end{aligned}$$

◇ Remarque. La fonction caractéristique de X n'est pas de classe $\mathscr{C}^1$. Ce n'est pas surprenant puisque X n'appartient pas à $\mathrm{L}^1(\mathbb{R})$.

Chapitre 5
Indépendances

5.1 Indépendance d'événements et famille d'événements

Définition 5.1. Soit $(\Omega, \mathscr{F}, \mathbb{P})$ un espace probabilisé. Deux événements A et B de $\mathscr{F}$ sont *indépendants* si

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

On note alors $A \perp\!\!\!\perp B$. Plus généralement, une famille $(A_i)_{i\in I}$ d'événements de $\mathscr{F}$ est *mutuellement indépendante* si, pour tout partie finie $J \subset I$, on a

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

◇ Remarque. La notion d'indépendance mutuelle est plus forte que l'indépendance deux à deux. En effet, on considère l'espace probabilisé $(\Omega, \mathscr{P}(\Omega), \mathbb{P})$ avec

$$\Omega := \{\mathrm{PP}, \mathrm{PF}, \mathrm{FP}, \mathrm{FF}\} \quad \text{et} \quad \mathbb{P} := \frac{\delta_{\mathrm{PP}} + \delta_{\mathrm{PF}} + \delta_{\mathrm{FP}} + \delta_{\mathrm{FF}}}{4}.$$

On pose $A := \{\mathrm{PF}, \mathrm{PP}\}$, $B := \{\mathrm{FP}, \mathrm{PP}\}$ et $C := \{\mathrm{FF}, \mathrm{PP}\}$. Alors (A, B, C) est une famille d'événements deux à deux indépendants, mais elle n'est pas mutuellement indépendantes.

Définition 5.2. Une famille $(\mathscr{A}_i)_{i\in I}$ de parties de $\mathscr{F}$ est mutuellement indépendantes si, pour toute partie finie $J \subset I$ et tout $A_j \in \mathscr{A}_j$ pour chaque $j \in J$, on a

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

Si les classes de parties $\mathscr{A}_i$ sont des tribus, on parle de tribus mutuellement indépendantes.

Proposition 5.3. Soit $(\mathscr{C}_i)_{i\in I}$ une famille mutuellement indépendante de π-systèmes de $\mathscr{F}$. Alors $(\sigma(\mathscr{C}_i))_{i\in I}$ est une famille mutuellement indépendante de sous-tribus de $\mathscr{F}$.

Preuve • *Cas particulier.* Traitons le cas où $I = \{1, 2\}$. Soient $\mathscr{C}_1$ et $\mathscr{C}_2$ deux π-systèmes indépendants. On fixe $B \in \mathscr{C}_2$. Alors la classe de parties

$$\mathscr{M} := \{A \in \sigma(\mathscr{C}_1) \mid \mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)\}$$

est une classe monotone contenant $\mathscr{C}_1$, donc $\mathscr{M} \supset m(\mathscr{C}_1) = \sigma(\mathscr{C}_1)$. Or $\mathscr{M} \subset \sigma(\mathscr{C}_1)$, donc $\mathscr{M} = \sigma(\mathscr{C}_1)$. On fixe maintenant $A \in \sigma(\mathscr{C}_1)$. De même, la classe de parties

$$\mathscr{M}' := \{B \in \sigma(\mathscr{C}_2) \mid \mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)\}$$

est une classe monotone égale à $\sigma(\mathscr{C}_2)$. Ceci montre que

$$\forall A \in \sigma(\mathscr{C}_1),\ \forall B \in \sigma(\mathscr{C}_2), \quad \mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B).$$

et termine la preuve.

• *Cas général.* On veut montrer que $(\sigma(\mathscr{C}_i))_{i\in I}$ est mutuellement indépendante. Soit $J := \{j_1, \dots, j_p\}$ une partie finie de I. Montrons que $(\sigma(\mathscr{C}_{j_k}))_{1\leqslant k\leqslant p}$ est mutuellement indépendante. On procède comme précédemment en fixant $A_2 \in \mathscr{C}_{j_2}$, …, $A_p \in \mathscr{C}_{j_p}$. □

Théorème 5.4 *(principe de coalition).* Soit $(\mathscr{F}_i)_{i\in I}$ une famille de sous-tribus de $\mathscr{F}$ qui sont mutuellement indépendantes. On suppose que l'on peut écrire $I = \bigsqcup_{a\in A} I_a$. Pour $a \in A$, on note

$$\mathscr{G}_a := \sigma\Big(\bigcup_{i\in I_a} \mathscr{F}_i\Big).$$

Alors $(\mathscr{G}_a)_{a\in A}$ est une famille mutuellement indépendante de tribus.

Preuve Pour $a \in A$, on note $\mathscr{A}_a \coloneqq \bigcup_{i\in I_a} \mathscr{F}_i$. Alors $(\mathscr{A}_a)_{a\in A}$ est mutuellement indépendante. On voudrait en déduire que $(\mathscr{G}_a)_{a\in A}$ est mutuellement indépendantes. Soit $a \in A$. On pose

$$\mathscr{C}_a \coloneqq \Big\{B \in \mathscr{F} \;\Big|\; \exists J \subset I_a \text{ fini}, \exists (A_j)_{j\in J} \subset \prod_{j\in J} \mathscr{F}_j,\ B = \bigcap_{j\in J} A_j\Big\}.$$

Les classes $\mathscr{C}_a$ sont des π-systèmes mutuellement indépendants. De plus, on a

$$\bigcup_{i\in I_a} \mathscr{F}_i \subset \mathscr{C}_a \subset \mathscr{G}_a, \quad \text{donc} \quad \mathscr{G}_a = \sigma\Big(\bigcup_{i\in I_a} \mathscr{F}_i\Big) \subset \sigma(\mathscr{C}_a) \subset \sigma(\mathscr{G}_a) = \mathscr{G}_a.$$

On en déduit que $\mathscr{G}_a = \sigma(\mathscr{C}_a)$. D'après la proposition précédente, comme les classes $\mathscr{C}_a$ sont des π-systèmes indépendants, la famille $(\mathscr{G}_a)_{a\in A}$ est une famille de tribus indépendante. ☐

5.2 Indépendance de variables aléatoires

Définition 5.5. La tribu engendrée par une variable aléatoire $X\colon (\Omega, \mathscr{F}, \mathbb{P}) \to (E, \mathscr{E})$ est

$$\sigma(X) \coloneqq \{X^{-1}(A) \mid A \in \mathscr{E}\}.$$

Définition 5.6. Une famille de variables aléatoires $(X_i)_{i\in I}$ est mutuellement indépendantes si $(\sigma(X_i))_{i\in I}$ est une famille mutuellement indépendantes de tribus.

Théorème 5.7 *(critère d'indépendance).* Soit $(X_1, \dots, X_n)$ un vecteur aléatoire à valeurs dans $(\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n))$. Alors les propriétés suivantes sont équivalentes :

(i) $X_1, \dots X_n$ sont mutuellement indépendantes ;
(ii) $\mathbb{P}_{(X_1,\dots,X_n)} = \mathbb{P}_{X_1} \otimes \cdots \otimes \mathbb{P}_{X_n}$;
(iii) pour tous fonctions boréliennes positives $h_1, \dots, h_n$, on a

$$\mathbb{E}\Big(\prod_{k=1}^n h_k(X_k)\Big) = \prod_{k=1}^n \mathbb{E}(h_k(X_k)) \ ; \tag{$*$}$$

(iv) pour tous fonctions boréliennes bornées $h_1, \dots, h_n$, on a $(*)$;
(v) pour tout $(t_1, \dots, t_n) \in \mathbb{R}^n$, on a

$$\mathbb{E}\Big(\exp\Big[i\sum_{k=1}^n t_k X_k\Big]\Big) = \prod_{k=1}^n \mathbb{E}(e^{it_k X_k}) \ ;$$

(vi) si les variables X_i sont à valeurs dans un ensemble dénombrable $E = \{x_p \mid p \in \mathbb{N}^*\}$, alors

$$\mathbb{P}(X_1 = x_{i_1}, \dots, X_n = x_{i_n}) = \mathbb{P}(X_1 = x_{i_1}) \cdots \mathbb{P}(X_n = x_{i_n})$$

pour tous $i_1, \dots, i_n \in \mathbb{N}^*$;
(vii) si le vecteur $(X_1, \dots, X_n)$ admet une densité f, alors

$$f(x_1, \dots, x_n) = f_{X_1}(x_1) \cdots f_{X_n}(x_n)$$

pour tout $(x_1, \dots, x_n) \in \mathbb{R}^n$.

Preuve • *(i) ⇔ (ii).* Il y a équivalence entre
– $X_1, \dots X_n$ sont mutuellement indépendantes ;
– pour tous $A_1, \dots, A_n \in \mathscr{B}(\mathbb{R})$, on a

$$\begin{aligned}\mathbb{P}_{(x_1,\dots,x_n)}(A_1 \times \cdots \times A_n) &= \mathbb{P}(X_1 \in A_1, \dots, X_n \in A_n) \\ &= \mathbb{P}(X_1 \in A_1) \cdots \mathbb{P}(X_n \in A_n) \\ &= \mathbb{P}_{X_1} \otimes \cdots \otimes \mathbb{P}_{X_n}(A_1 \times \cdots \times A_n) \ ;\end{aligned}$$

– $\mathbb{P}_{(X_1,\dots,X_n)}$ et $\mathbb{P}_{X_1} \otimes \cdots \otimes \mathbb{P}_{X_n}$ coïncident sur les produits cartésiens de boréliens ;
– $\mathbb{P}_{(X_1,\dots,X_n)} = \mathbb{P}_{X_1} \otimes \cdots \otimes \mathbb{P}_{X_n}$.

• *(ii) ⇒ (iii).* On suppose (ii). Soient $h_1, \dots, h_n$ des fonctions boréliennes positives. Le théorème de transfert et de Fubini-Tonelli donnent

$$\begin{aligned}\mathbb{E}(h_1(X_1) \cdots h_n(X_n)) &= \int_{\mathbb{R}^n} h_1(x_1) \cdots h_n(x_n) \,\mathrm{d}\mathbb{P}_{(X_1,\dots,X_n)}(x_1, \dots, x_n) \\ &= \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} h_1(x_1) \cdots h_n(x_n) \,\mathrm{d}\mathbb{P}_{X_1}(x_1) \cdots \mathrm{d}\mathbb{P}_{X_n}(x_n)\end{aligned}$$

$$
\begin{aligned}
&= \int_{\mathbb{R}} h_1(x_1)\,\mathrm{d}\mathbb{P}_{X_1}(x_1) \times \cdots \times \int_{\mathbb{R}} h_n(x_n)\,\mathrm{d}\mathbb{P}_{X_n}(x_n) \\
&= \mathbb{E}(h_1(X_1)) \cdots \mathbb{E}(h_n(X_n)).
\end{aligned}
$$

• *(iii) ⇒ (iv).* On suppose (iii). Soient h_1, …, h_n des fonctions boréliennes bornées. Pour chaque $i \in [\![1, n]\!]$, on écrit $h_i = h_i^+ - h_i^-$. On obtient alors (iv) puisque on peut développer

$$\mathbb{E}\Big(\prod_{k=1}^n h_k(X_k)\Big) = \mathbb{E}\Big(\prod_{k=1}^n [h_k^+ - h_k^-]\Big).$$

• *(iv) ⇒ (v).* Comme les fonctions $h_k \colon x \longmapsto e^{t_k x}$ sont boréliennes et bornées, cette implication est triviale.

• *(v) ⇒ (ii).* On suppose (v). Pour tout $(t_1, \dots, t_n) \in \mathbb{R}^n$, on a

$$\mathbb{E}\Big(\exp\Big[i\sum_{k=1}^n t_k X_k\Big]\Big) = \int_{\mathbb{R}^n} \exp\Big[i\sum_{k=1}^n t_k x_k\Big]\,\mathrm{d}\mathbb{P}_{(X_1,\dots,X_n)}(x_1, \dots, x_n)$$

et

$$
\begin{aligned}
\mathbb{E}\Big(\exp\Big[i\sum_{k=1}^n t_k X_k\Big]\Big) &= \prod_{k=1}^n \int_{\mathbb{R}} e^{it_k x_k}\,\mathrm{d}\mathbb{P}_{X_k}(x_k) \\
&= \int_{\mathbb{R}^n} \exp\Big[i\sum_{k=1}^n t_k x_k\Big]\,\mathrm{d}\mathbb{P}_{X_1} \otimes \cdots \otimes \mathbb{P}_{X_n}(x_1, \dots, x_n).
\end{aligned}
$$

Alors $\mathbb{P}_{(X_1,\dots,X_n)}(x_1, \dots, x_n)$ et $\mathbb{P}_{X_1} \otimes \cdots \otimes \mathbb{P}_{X_n}$ donnent les mêmes fonctions caractéristiques, donc elles sont égales et on a montré (ii).

• *(i) ⇔ (vi).* On suppose que X_1, … X_n sont mutuellement indépendantes. Soient $i_1, \dots, i_n \in \mathbb{N}^*$. Comme les fonctions $\mathbb{1}_{\{X_j = x_{i_j}\}}$ sont bornées, le point (iv) donne

$$
\begin{aligned}
\mathbb{P}(X_1 = x_{i_1}, \dots, X_n = x_{i_n}) &= \mathbb{E}(\mathbb{1}_{\{X_1 = x_{i_1}\}} \cdots \mathbb{1}_{\{X_n = x_{i_n}\}}) \\
&= \mathbb{E}(\mathbb{1}_{\{X_1 = x_{i_1}\}}) \cdots \mathbb{E}(\mathbb{1}_{\{X_n = x_{i_n}\}}) \\
&= \mathbb{P}(X_1 = x_{i_1}) \cdots \mathbb{P}(X_n = x_{i_n})
\end{aligned}
$$

Réciproquement, on suppose (vi). Soit $(t_1, \dots, t_d) \in \mathbb{R}^d$. Alors le théorème de transfert assure

$$
\begin{aligned}
\mathbb{E}(e^{i(t_1 X_{i_1} + \cdots + t_n X_{i_n})}) &= \sum_{i_1=1}^{+\infty} \cdots \sum_{i_n=1}^{+\infty} e^{i(t_1 X_{i_1} + \cdots + t_n X_{i_n})} \mathbb{P}(X_1 = x_{i_1}, \dots, X_n = x_{i_n}) \\
&= \sum_{i_1=1}^{+\infty} \cdots \sum_{i_n=1}^{+\infty} e^{i(t_1 X_{i_1} + \cdots + t_n X_{i_n})} \mathbb{P}(X_1 = x_{i_1}) \cdots \mathbb{P}(X_n = x_{i_n}) \\
&= \sum_{i_1=1}^{+\infty} \cdots \sum_{i_n=1}^{+\infty} \prod_{k=1}^n e^{it_k X_{i_k}} \mathbb{P}(X_k = x_{i_k}) \\
&= \prod_{k=1}^n \mathbb{E}(e^{itkX_k}).
\end{aligned}
$$

On en déduit que X_1, …, x_n sont mutuellement indépendantes.

• *(i) ⇔ (vii).* On a

$$\mathbb{P}_{(X_1,\dots,X_n)} = f(x_1, \dots, x_n)\,\mathrm{d}x_1 \cdots \mathrm{d}x_n \quad \text{et} \quad \mathbb{P}_{X_1} \otimes \cdots \otimes \mathbb{P}_{X_n} = f_{X_1}(x_1) \cdots f_{X_n}(x_n)\,\mathrm{d}x_1 \cdots \mathrm{d}x_n.$$

Avec le point (ii), on en déduit que X_1, …, X_n sont mutuellement indépendantes si et seulement si

$$f(x_1, \dots, x_n)\,\mathrm{d}x_1 \cdots \mathrm{d}x_n = f_{X_1}(x_1) \cdots f_{X_n}(x_n)\,\mathrm{d}x_1 \cdots \mathrm{d}x_n,$$

i. e. si et seulement si (vii). □

◇ Remarques. – Si les variables X_1, …, X_n sont indépendantes, alors la fonction de densité du vecteur $(X_1, \dots, X_n)$ est à « variables séparées ». Par exemple, la fonction $(2\pi)^{-1} e^{-(x^2+y^2)/2} = (2\pi)^{-1} e^{-x^2/2} e^{-y^2/2}$ est à variables séparées, mais la fonction $ce^{-(ax^2+bxy+y^2)}$ ne l'est pas si $b \neq 0$.

– Lorsqu'on veut montrer que X et Y sont indépendantes, il suffit de vérifier que $\mathbb{E}(e^{i(tX+sY)}) = \mathbb{E}(e^{itX})\mathbb{E}(e^{isY})$ pour tous $t, s \in \mathbb{R}$. Par contre, si on a uniquement cette égalité pour $t = s$, les variables X et Y ne sont pas forcément indépendantes : il suffit de prendre $X = Y \sim \mathscr{C}(1)$ et la variable X n'est pas indépendante avec elle-même car sinon elle serait constante. Montrons ce dernier point. On suppose que $X \perp\!\!\!\perp X$. Pour tout $t \in \mathbb{R}$, on a

$$F_X(t) = \mathbb{P}(X \leqslant t) = \mathbb{P}(\{X \leqslant t\} \cap \{X \leqslant t\}) = \mathbb{P}(X \leqslant t)\mathbb{P}(X \leqslant t) = F_X(t)^2,$$

donc $F_X(t) \in \{0,1\}$. On pose

$$t_0 \coloneqq \inf\{t \in \mathbb{R} \mid F_X(t) = 1\}.$$

Par continuité à droite de F_X, on a $F_X(t_0) = 1$. De plus, pour tout $\varepsilon > 0$, on a $\mathbb{P}(X \leqslant t_0 - \varepsilon) = 0$. Donc

$$\mathbb{P}(X < t_0) = \mathbb{P}\Big(\bigcup_{\varepsilon \in \mathbb{Q}_+^*} \{X \leqslant t_0 - \varepsilon\}\Big) = 0.$$

D'où $\mathbb{P}(X = t_0) = 1$, donc $X = t_0$ presque sûrement.

5.3 Sommes de variables indépendantes

Définition 5.8 *(convolution de lois).* Soient $(X_1, \dots, X_d)$ et $(Y_1, \dots, Y_d)$ deux vecteurs aléatoires indépendants à valeurs dans $\mathbb{R}^d$. On définit la loi de la somme de ceux deux vecteurs comme

$$\mathbb{P}_{(X_1+Y_1,\dots,X_d+Y_d)} \coloneqq \mathbb{P}_{(X_1,\dots,X_d)} * \mathbb{P}_{(Y_1,\dots,Y_d)}.$$

◇ Remarque. Si les deux vecteurs admettent respectivement des densités f et g, alors le vecteur somme a admet pour densité $f * g$.

Preuve Soit $A \in \mathscr{B}(\mathbb{R}^d)$. Alors successivement, le théorème de transfert, le théorème de Fubini et le changement de variables $(u_1, \dots, u_d) = (x_1 + y_1, \dots, x_d + y_d)$ assurent

$$\begin{aligned}
&\mathbb{P}((X_1 + Y_1, \dots, X_d + Y_d) \in A) \\
&= \mathbb{E}(\mathbb{1}_A(X_1 + Y_1, \dots, X_d + Y_d)) \\
&= \int_{\mathbb{R}^{2d}} \mathbb{1}_A(x_1 + y_1, \dots, x_d + y_d) \, \mathrm{d}\mathbb{P}_{(X_1,\dots,X_d,Y_1,\dots,Y_d)}(x_1, \dots, x_d, y_1, \dots, y_d) \\
&= \int_{\mathbb{R}^{2d}} \mathbb{1}_A(x_1 + y_1, \dots, x_d + y_d) \, \mathrm{d}\mathbb{P}_{(X_1,\dots,X_d)}(x_1, \dots, x_d) \otimes \mathbb{P}_{(Y_1,\dots,Y_d)}(y_1, \dots, y_d) \\
&= \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} \mathbb{1}_A(x_1 + y_1, \dots, x_d + y_d) f(x_1, \dots, x_d) g(y_1, \dots, y_d) \, \mathrm{d}x_1 \cdots \mathrm{d}x_d \, \mathrm{d}y_1 \cdots \mathrm{d}y_d \\
&= \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} g(y_1, \dots, y_d) \left(\int_{\mathbb{R}} \cdots \int_{\mathbb{R}} \mathbb{1}_A(x_1 + y_1, \dots, x_d + y_d) f(x_1, \dots, x_d) \, \mathrm{d}x_1 \cdots \mathrm{d}x_d \right) \mathrm{d}y_1 \cdots \mathrm{d}y_d \\
&= \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} g(y_1, \dots, y_d) \left(\int_{\mathbb{R}} \cdots \int_{\mathbb{R}} \mathbb{1}_A(u_1, \dots, u_d) f(u_1 - y_1, \dots, u_d - y_d) \, \mathrm{d}u_1 \cdots \mathrm{d}u_d \right) \mathrm{d}y_1 \cdots \mathrm{d}y_d \\
&= \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} \mathbb{1}_A(u_1, \dots, u_d) \left(\int_{\mathbb{R}} \cdots \int_{\mathbb{R}} g(y_1, \dots, y_d) f(u_1 - y_1, \dots, u_d - y_d) \, \mathrm{d}y_1 \cdots \mathrm{d}y_d \right) \mathrm{d}u_1 \cdots \mathrm{d}u_d \\
&= \int_{\mathbb{R}^d} \mathbb{1}_A(u)(f * g)(u) \, \mathrm{d}u.
\end{aligned}$$

□

Convolution des lois classiques.

– Soient $n, m \in \mathbb{N}^*$ et $p \in]0,1]$. Soient X et Y deux variables aléatoires indépendantes telles que $X \sim \mathrm{Bin}(n,p)$ et $Y \sim \mathrm{Bin}(m,p)$. Alors $X + Y \sim \mathrm{Bin}(n+m,p)$. En effet, pour tout $t \in \mathbb{R}$, on a

$$\begin{aligned}
\mathbb{E}(e^{it(X+Y)}) &= \mathbb{E}(e^{itX})\mathbb{E}(e^{itY}) \\
&= [pe^{it} + (1-p)]^n [pe^{it} + (1-p)]^m \\
&= [pe^{it} + (1-p)]^{n+m}
\end{aligned}$$

où on reconnaît la fonction caractéristique d'une loi binomiale de paramètre $(n+m,p)$.

On peut montrer la sorte de réciproque suivante : si X et Y sont deux variables indépendantes telles que $X + Y \sim \mathrm{Bin}(n,p)$, alors il existe $c > 0$ et $n_1, n_2 \in \mathbb{N}^*$ avec $n_1 + n_2 = n$ tels que $X - c \sim \mathrm{Bin}(n_1,p)$ et $Y + c \sim \mathrm{Bin}(n_2,p)$.

– Soient $\lambda_1, \lambda_2 > 0$. Soient X et Y deux variables aléatoires indépendantes telles que $X \sim \mathscr{P}(\lambda_1)$ et $Y \sim \mathscr{P}(\lambda_2)$. Alors $X + Y \sim \mathscr{P}(\lambda_1 + \lambda_2)$. En effet, on procède de même que précédemment avec les fonctions caractéristiques. On a la réciproque suivante.

Théorème 5.9 *(Raikov).* Soient X et Y deux variables aléatoires indépendantes telles que $X + Y \sim \mathscr{P}(\lambda)$. Alors il existe $c > 0$ et $\lambda_1, \lambda_2 > 0$ avec $\lambda_1 + \lambda_2 = n$ tels que $X - c \sim \mathscr{P}(\lambda_1)$ et $Y + c \sim \mathscr{P}(\lambda_2)$.

– Soient $\mu_1, \sigma_1, \mu_2, \sigma_2 \in \mathbb{R}$. Soient X et Y deux variables aléatoires indépendantes telles que $X \sim \mathscr{N}(\mu_1, \sigma_1^2)$ et $Y \sim \mathscr{N}(\mu_2, \sigma_2^2)$ telles que $X \perp\!\!\!\perp Y$. Alors $X + Y \sim \mathscr{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$. De même, on a la réciproque suivant

Théorème 5.10 *(Cramèr).* Soient X et Y deux variables aléatoires indépendantes telles que $X+Y \sim \mathscr{N}(\mu, \sigma^2)$. Alors il existe $\mu_1, \mu_2, \sigma_1, \sigma_2 \in \mathbb{R}$ avec $\mu_1 + \mu_2 = \mu$ et $\sigma_1^2 + \sigma_2^2 = \sigma^2$ tels que $X \sim \mathscr{N}(\mu_1, \sigma_1^2)$ et $X \sim \mathscr{N}(\mu_2, \sigma_2^2)$.

– Soient $a, b \in \mathbb{R}$. Soient X et Y deux variables aléatoires indépendantes telles que $X \sim \mathscr{C}(a)$ et $Y \sim \mathscr{C}(b)$. Alors $X + Y \sim \mathscr{C}(a+b)$.

Proposition 5.11. Soient X_1, …, X_n des variables aléatoires deux à deux indépendantes admettant un moment d'ordre 2. Alors

$$\operatorname{Var}(X_1 + \cdots + X_n) = \operatorname{Var}(X_1) \cdots \operatorname{Var}(X_n).$$

Preuve Comme

$$\operatorname{Var}(X_1 + \cdots + X_n) = \operatorname{Var}(X_1) \cdots \operatorname{Var}(X_n) + \sum_{i \neq j} \operatorname{Cov}(X_i, X_j),$$

il suffit de montrer que $\operatorname{Cov}(X_i, X_j) = 0$ pour tous $i \neq j$. Pour tous $i \neq j$, comme les variables X_i et X_j sont indépendantes, on a $\mathbb{E}(X_i X_j) = \mathbb{E}(X_i)\mathbb{E}(X_j)$, donc $\operatorname{Cov}(X_i, X_j) = 0$. □

Théorème 5.12. 1. Il existe un espace probabilisé $(\Omega, \mathscr{F}, \mathbb{P})$ et une suite $(X_i)_{i \in \mathbb{N}}$ de variables aléatoires mutuellement indépendantes suivant une loi de Bernoulli de paramètre $1/2$.
2. Étant donnée une suite quelconque $(\mu_i)_{i \geqslant 1}$ de mesures de probabilités sur $\mathbb{R}$, il existe un espace probabilisé $(\Omega, \mathscr{F}, \mathbb{P})$ et une suite $(X_i)_{i \geqslant 1}$ de variables aléatoires mutuellement indépendantes tels que $\mathbb{P}_{X_i} = \mu_i$ pour tout $i \geqslant 1$.

Preuve 1. Soient $(\Omega, \mathscr{F}, \mathbb{P})$ un espace probabilisé et X une variable aléatoire définie sur $(\Omega, \mathscr{F}, \mathbb{P})$ de loi uniforme sur $[0, 1]$. Par exemple, on prend l'espace $([0, 1], \mathscr{B}([0, 1], \lambda)$ et $X : x \in [0, 1] \longmapsto x \in \mathbb{R}$ de sorte que $\mathbb{P}_X = \lambda$. Par l'écriture diadique, on peut écrire X sous la forme

$$X = \sum_{k=1}^{+\infty} \frac{\varepsilon_k}{2^k} \quad \text{avec} \quad \varepsilon_k \in \{0, 1\}.$$

Soient $n \in \mathbb{N}^*$ et $k \in [\![0, 2^n - 1]\!]$. Alors

$$\mathbb{P}(X \in [k/2^n, (k+1)/2^n[) = 1/2^n.$$

Or

$$\{X \in [k/2^n, (k+1)/2^n[\} = \{2^n X \in [k, k+1[\} = \{\lfloor 2^n X \rfloor = k\}.$$

On écrit $k = a_1 + \cdots + a_n 2^{n-1}$ avec $a_i \in \{0, 1\}$. Alors

$$2^n X = \sum_{k=1}^{n} \varepsilon_k 2^{n-k} + \underbrace{\sum_{k=n+1}^{+\infty} \varepsilon_k 2^{n-k}}_{<1},$$

donc $\mathbb{P}(\varepsilon_1 = a_n, \dots, \varepsilon_n = a_1) = 2^{-n}$. C'est vrai pour $k \in [\![0, 2^n - 1]\!]$. Finalement, pour tout $(a_1, \dots, a_n) \in \{0, 1\}^n$, on a $\mathbb{P}(\varepsilon_1 = a_n, \dots, \varepsilon_n = a_1) = 1/2^n$ ce qui implique que, pour tout $i \geqslant 1$, on a

$$\mathbb{P}(\varepsilon_i = 1) = \sum_{(a_1, \dots, a_{i-1}, a_{i+1}, \dots, a_n) \in \{0,1\}^{n-1}} \mathbb{P}(\varepsilon_1 = a_1, \dots, \varepsilon_i = 1, \dots, \varepsilon_n = a_n) = \frac{1}{2^n} 2^{n-1} = \frac{1}{2},$$

donc $\varepsilon_i \sim \operatorname{Ber}(1/2)$. On en déduit que les variables ε_i sont mutuellement indépendantes. □

◇ Remarque. – Comme $\mathbb{N}$ et $\mathbb{N}^2$ sont en bijections, il existe une suite $(\varepsilon_{p,q})_{p,q \geqslant 1}$ de variables aléatoires indépendantes de loi de Bernoulli de paramètre $1/2$. Pour $p \geqslant 1$, on pose

$$u_p \coloneqq \sum_{q=1}^{+\infty} \frac{\varepsilon_{p,q}}{2^k}.$$

Alors les variables u_p sont mutuellement indépendantes. En effet, l'écriture

$$\mathbb{N}^* \times \mathbb{N}^* = \bigcup_{p \geqslant 1} A_p \quad \text{avec} \quad A_p \coloneqq \{(p, q) \mid q \geqslant 1\}$$

est une partition de $\mathbb{N}^* \times \mathbb{N}^*$, donc la fonction u_p est mesurable par rapport à la tribu $\sigma(\varepsilon_{i,j} \mid (i, j) \in A_p)$, donc c'est une variable aléatoire. Montrons que $u_p \sim \mathscr{U}_{[0,1]}$. Pour cela, considérons une variable U qui s'écrit

$$U = \sum_{k=1}^{+\infty} \frac{\varepsilon_k}{2^k}$$

où les variables ε_k sont indépendantes et suivent une loi de Bernoulli de paramètre 1/2. Montrons que $U \sim \mathscr{U}_{[0,1]}$. Soit $t \in \mathbb{R}$. Pour $n \geqslant 1$, on a

$$\begin{aligned}
\mathbb{E}\Big(\exp\Big[it\sum_{k=1}^{n}\frac{\varepsilon_k}{2^k}\Big]\Big) &= \sum_{k=1}^{n}\mathbb{E}(e^{it\varepsilon_k/2^k}) \\
&= \prod_{k=1}^{n}\left(\frac{1}{2}e^{it/2^k}+\frac{1}{2}\right) \\
&= \prod_{k=1}^{n} e^{it/2^{k+1}}\frac{e^{it/2^{k+1}}+e^{-it/2^{k+1}}}{2} \\
&= \prod_{k=1}^{n} e^{it/2^{k+1}}\cos\Big(\frac{t}{2^{k+1}}\Big) \\
&= \exp\Big(it\sum_{k=1}^{n}2^{-(k+1)}\Big)\prod_{k=1}^{n}\cos\Big(\frac{t}{2^{k+1}}\Big) \\
&= \exp\Big(it\sum_{k=1}^{n}2^{-(k+1)}\Big)\frac{1}{2^n\sin(t/2^{n+1})}\prod_{k=1}^{n}\cos\Big(\frac{t}{2^{k+1}}\Big)\,2^n\sin\Big(\frac{t}{2^{k+1}}\Big) \\
&= \exp\Big(it\sum_{k=1}^{n}2^{-(k+1)}\Big)\frac{\sin(t/2)}{2^n\sin(t/2^{n+1})} \sim e^{it/2}\frac{\sin(t/2)}{t/2}.
\end{aligned}$$

En laissant tendre n vers $+\infty$, le théorème de convergence dominée donne

$$\mathbb{E}(e^{itU}) = \mathbb{E}\Big(\exp\Big[it\sum_{k=1}^{+\infty}\frac{\varepsilon_k}{2^k}\Big]\Big) = e^{it/2}\frac{\sin(t/2)}{t/2}.$$

Or

$$\int_0^1 e^{itx}\,\mathrm{d}x = \frac{e^{it}-1}{it} = e^{it/2}\frac{\sin(t/2)}{t/2}.$$

Par la caractérisation par les fonctions caractéristiques, on obtient que $U \sim \mathscr{U}_{[0,1]}$.

– À partir d'un loi uniforme, on peut créer une infinité de variables aléatoires indépendantes de loi uniforme. Soient μ une mesure de probabilité sur $\mathbb{R}$ et F_μ sa fonction de répartition. Soit U une variable aléatoire suivant une loi uniforme. Si F_μ est inversible, alors $F_\mu^{-1}(U) \sim \mu$.

5.4 Lemmes de Borel-Cantelli

Lemme 5.13. Soient $(\Omega, \mathscr{F}, \mathbb{P})$ un espace probabilisé et $(A_k)_{k\geqslant 1}$ une suite d'événements de $\mathscr{F}$. Alors

$$\sum_{k=1}^{+\infty}\mathbb{P}(A_k) < +\infty \quad\Longrightarrow\quad \mathbb{P}(\limsup_{k\to+\infty} A_k) = 0.$$

Preuve On suppose que $\sum_{k=1}^{+\infty}\mathbb{P}(A_k) < +\infty$. On rappel que

$$\limsup_{k\to+\infty} A_k = \bigcap_{n\geqslant 1}\bigcup_{k\geqslant n} A_k.$$

Par la σ-sous-additivité, pour tout $n \geqslant N$, on a

$$\mathbb{P}\Big(\bigcup_{k\geqslant n} A_k\Big) \leqslant \sum_{k\geqslant n}\mathbb{P}(A_k) \longrightarrow 0.$$

D'autre part, la continuité donne

$$\mathbb{P}\Big(\bigcap_{n\geqslant 1}\bigcup_{k\geqslant n} A_k\Big) = \lim_{n\to+\infty}\mathbb{P}\Big(\bigcup_{k\geqslant n} A_k\Big).$$

D'où

$$\mathbb{P}(\limsup_{k\to+\infty} A_k) = \mathbb{P}\Big(\bigcap_{n\geqslant 1}\bigcup_{k\geqslant n} A_k\Big) \leqslant \lim_{n\to+\infty}\sum_{k\geqslant n}\mathbb{P}(A_k) \longrightarrow 0.$$ □

LEMME 5.14. On suppose que la suite $(A_k)_{k\geqslant 1}$ est mutuellement indépendante. Alors

$$\sum_{k=1}^{+\infty} \mathbb{P}(A_k) = +\infty \quad \Longrightarrow \quad \mathbb{P}(\limsup_{k\to+\infty} A_k) = 1.$$

Preuve On suppose que $\sum_{k=1}^{+\infty} \mathbb{P}(A_k) = +\infty$. Alors

$$\mathbb{P}([\limsup_{k\to+\infty} A_k]^{\mathrm{c}}) = 1 - \mathbb{P}(\limsup_{k\to+\infty} A_k) = \mathbb{P}(\liminf_{k\to+\infty} A_k^{\mathrm{c}}).$$

Or

$$\liminf_{k\to+\infty} A_k^{\mathrm{c}} = \bigcup_{n\geqslant 1} \bigcap_{k\geqslant n} A_k^{\mathrm{c}}.$$

Il suffit de montrer que $\mathbb{P}(\bigcap_{k\geqslant n} A_k^{\mathrm{c}}) = 0$ pour tout $n \geqslant 1$. La continuité donne

$$\mathbb{P}\Big(\bigcap_{k\geqslant n} A_k^{\mathrm{c}}\Big) = \mathbb{P}\Big(\bigcap_{p\geqslant n} \bigcap_{k=n}^{p} A_k^{\mathrm{c}}\Big) = \lim_{p\to+\infty} \mathbb{P}\Big(\bigcup_{k=n}^{p} A_k^{\mathrm{c}}\Big) = \lim_{p\to+\infty} \prod_{k=n}^{p} [1 - \mathbb{P}(A_k)].$$

On adopte la convention $\log 0 = -\infty$. Alors pour tout $p \geqslant 1$, on a

$$\log\Big(\prod_{k=n}^{p} [1 - \mathbb{P}(A_k)]\Big) = \sum_{k=n}^{p} \ln(1 - \mathbb{P}(A_k)) \leqslant -\sum_{k=n}^{p} \mathbb{P}(A_k) \longrightarrow -\infty.$$

On en déduit que

$$\prod_{k=n}^{p} [1 - \mathbb{P}(A_k)] \longrightarrow 0.$$

On en déduit que $\mathbb{P}([\limsup_{k\to+\infty} A_k]^{\mathrm{c}}) = 0$, donc $\mathbb{P}(\limsup_{k\to+\infty} A_k) = 1$. ☐

LEMME 5.15 *(super BOREL-CANTELLI)*. On suppose que les variables A_k sont deux à deux indépendantes. Alors

$$\sum_{k=1}^{+\infty} \mathbb{P}(A_k) = +\infty \quad \Longrightarrow \quad \mathbb{P}(\limsup_{k\to+\infty} A_k) = 1.$$

Preuve On suppose que $\sum_{k=1}^{+\infty} \mathbb{P}(A_k) = +\infty$. On remarque que, pour tous $n, m \in \mathbb{N}$ tels que $n \leqslant m$, on a

$$\mathbb{1}_{A_n\cup\cdots\cup A_k} \sum_{k=1}^{m} \mathbb{1}_{A_k} = \sum_{k=1}^{m} \mathbb{1}_{A_k}.$$

En intégrant et en utilisant l'inégalite de CAUCHY-SCHWARZ, on obtient que

$$\begin{aligned}
\mathbb{P}\Big(\bigcup_{k=n}^{m} A_k\Big) = \sum_{k=n}^{m} \mathbb{P}(A_k) = \mathbb{E}\Big(\sum_{k=n}^{m} \mathbb{1}_{A_k}\Big) &\leqslant \sqrt{\mathbb{P}\Big(\bigcup_{k=n}^{m} A_k\Big)}\sqrt{\mathbb{E}\Big(\Big(\sum_{k=n}^{m} \mathbb{1}_{A_k}\Big)^2\Big)} \\
&\leqslant \sqrt{\mathbb{P}\Big(\bigcup_{k=n}^{m} A_k\Big)}\sqrt{\sum_{k=n}^{m} \mathbb{P}(A_k) + \sum_{k\neq \ell} \mathbb{P}(A_k \cap A_\ell)} \\
&\leqslant \sqrt{\mathbb{P}\Big(\bigcup_{k=n}^{m} A_k\Big)}\sqrt{\sum_{k=n}^{m} \mathbb{P}(A_k) + \sum_{k\neq \ell} \mathbb{P}(A_k)\mathbb{P}(A_\ell)} \\
&\leqslant \sqrt{\mathbb{P}\Big(\bigcup_{k=n}^{m} A_k\Big)}\sqrt{\sum_{k=n}^{m} \mathbb{P}(A_k) + \Big(\sum_{k=n}^{m} \mathbb{P}(A_k)\Big)^2}.
\end{aligned}$$

On en déduit que

$$\mathbb{P}\Big(\bigcup_{k=n}^{m} A_k\Big) \geqslant \frac{(\sum_{k=n}^{m} \mathbb{P}(A_k))^2}{(\sum_{k=n}^{m} \mathbb{P}(A_k))^2 + \sum_{k=n}^{m} \mathbb{P}(A_k)}.$$

Comme $x^2/(x^2 + x) \longrightarrow 1$ quand $x \to +\infty$, en laissant tendre m vers $+\infty$, on obtient que $\mathbb{P}(\bigcup_{k=n}^{+\infty} A_k) = 1$. Par continuité, on a

$$\mathbb{P}(\limsup_{k\to+\infty} A_k) = \lim_{n\to+\infty} \mathbb{P}\Big(\bigcup_{k=n}^{+\infty} A_k\Big) = 1.$$ ☐

5.5 Tribu asymptotique et loi du 0-1 de Kolmogorov

Définition 5.16. Soit $(X_n)_{n\geqslant 1}$ une suite de variables aléatoires définies sur un espace probabilisé. Pour $n \geqslant 1$, la *tribu du futur* et le *tribu asymptotique* sont respectivement les tribus

$$\mathscr{F}^n := \sigma(X_k \mid k \geqslant n) \quad \text{et} \quad \mathscr{F}^\infty := \bigcap_{n\geqslant 1} \mathscr{F}^n.$$

◇ Remarque. Les événements de $\mathscr{F}^\infty$ mesurent le comportement asymptotiques de la suite $(X_n)_{n\geqslant 1}$.

– On note $A := \{(X_n)_{n\geqslant 1}$ converge dans $\mathbb{R}\}$. Comme $\mathbb{R}$ est complet, on a

$$A = \bigcap_{\varepsilon\in\mathbb{Q}_+} \bigcup_{p\geqslant 1} \bigcap_{n,m\geqslant p} \{|X_n - X_m| < \varepsilon\} \in \mathscr{F}.$$

De plus, comme $A = \{\limsup X_n = \liminf X_n\}$ et les fonctions $\limsup X_n$ et $\liminf X_n$ sont $\mathscr{F}^p$-mesurable pour tout $p \geqslant 1$, on en déduit que $A \in \mathscr{F}^p$ pour tout $p \geqslant 1$ ce qui assure que $A \in \mathscr{F}^\infty$.

– On note $B := \{(X_n)_{n\geqslant 1}$ est bornée dans $\mathbb{R}\}$. On a

$$B = \bigcup_{N\geqslant 1} \bigcap_{n\geqslant 1} \{|X_n| \leqslant N\} \in \mathscr{F}.$$

De plus, pour tout $p \geqslant 1$, on a $B = \{(X_{n+p})_{n\geqslant 1}$ est bornée dans $\mathbb{R}\} \in \mathscr{F}^p$ ce qui montre que $B \in \mathscr{F}^\infty$.

– Mais l'événement $\{(X_n)_{n\geqslant 1}$ est croissante$\}$ n'appartient pas à $\mathscr{F}^\infty$.

Théorème 5.17 *(loi du 0-1 de Kolmogorov).* Soit $(X_n)_{n\geqslant 1}$ une suite mutuellement indépendante de variables aléatoires. Alors la tribu $\mathscr{F}^\infty$ est triviale, *i. e.* pour tout événement $A \in \mathscr{F}^\infty$, on a $\mathscr{P}(A) \in \{0, 1\}$.

Preuve Montrons que la tribu $\mathscr{F}^\infty$ est indépendante d'elle-même. Soit $n \geqslant 1$. Le lemme de coalition donne

$$\sigma(X_1, \dots, X_n) \perp\!\!\!\perp \mathscr{F}^{n+1}.$$

Or $\mathscr{F}^\infty \subset \mathscr{F}^{n+1}$, donc $\sigma(X_1, \dots, X_n) \perp\!\!\!\perp \mathscr{F}^\infty$. Donc on obtient que

$$\mathscr{C} := \bigcup_{n\geqslant 1} \sigma(X_1, \dots, X_n) \perp\!\!\!\perp \mathscr{F}^\infty.$$

On montre que la classe de parties $\mathscr{C}$ est un π-système, donc $\sigma(\mathscr{C}) \perp\!\!\!\perp \mathscr{F}^\infty$. Enfin, on a $\sigma(\mathscr{C}) = \sigma(X_n \mid n \geqslant 1)$. Or $\mathscr{F}^\infty \subset \sigma(X_n \mid n \geqslant 1)$, donc $\mathscr{F}^\infty \perp\!\!\!\perp \mathscr{F}^\infty$. On conclut que la tribu $\mathscr{F}^\infty$ est triviale. □

◇ Remarque. Si X est une variable aléatoire mesurable par rapport à $\mathscr{F}^\infty$, alors elle est presque sûrement constante (cf. remarque p. 21).

Chapitre 6

Convergences de variables aléatoires

6.1 Convergence presque sûre

Définition 6.1. On dit qu'une suite de variables aléatoires réelles $(X_n)_{n\geqslant 0}$ converge presque sûrement vers une variable aléatoire X_∞ si

$$\mathbb{P}(\{\omega \in \Omega \mid X_n(\omega) \longrightarrow X_\infty(\omega)\}) = 1.$$

On note alors $X_n \xrightarrow{\text{ps}} X_\infty$.

◇ Remarques. – En général, on n'a pas la convergence partout, *i. e.* pour tout $\omega \in \Omega$. La convergence presque sûre revient à dire que l'ensemble des « exceptions » est négligeable.
– L'ensemble $\{\omega \in \Omega \mid X_n(\omega) \longrightarrow X_\infty(\omega)\}$ est bien un élément de la tribu $\mathscr{F}$.

Définition 6.2. Soit $d \geqslant 1$. On dit qu'une suite de vecteurs aléatoires $(X_n)_{n\geqslant 0} \coloneqq (X_{n,1}, \dots, X_{n,d})_{n\geqslant 0}$ converge presque sûrement vers un vecteur aléatoire X_∞ si

$$\mathbb{P}(\{\omega \in \Omega \mid X_n(\omega) \longrightarrow X_\infty(\omega)\}) = 1.$$

◇ Remarque. La converge presque sûrement d'un vecteur aléatoire est équivalent à la converge de ce vecteur coordonnées par coordonnées.

Proposition 6.3. Soient $(X_n)_{n\geqslant 0}$ une suite de variables aléatoires et X et Y deux variables aléatoires. Alors
1. si $X_n \xrightarrow{\text{ps}} X$ et $X_n \xrightarrow{\text{ps}} Y$, alors $\mathbb{P}(X = Y) = 1$;
2. on a $X_n \xrightarrow{\text{ps}} X$ si et seulement si

$$\forall \varepsilon > 0, \quad \mathbb{P}(\limsup_{n\to+\infty} \{|X_n - X| > \varepsilon\}) = 0 \ ;$$

3. on a

$$\left(\forall \varepsilon > 0, \quad \sum_{n=0}^{+\infty} \mathbb{P}(|X_n - X| > \varepsilon) < +\infty\right) \quad \Longrightarrow \quad X_n \xrightarrow{\text{ps}} X \ ;$$

4. si la suite $(X_n)_{n\geqslant 0}$ est mutuellement indépendante, alors l'implication du point 3 est une équivalence ;
5. pour toute $f \in \mathscr{C}^0(\mathbb{R}^d, \mathbb{R}^m)$, si $X_n \xrightarrow{\text{ps}} X$, alors $f(X_n) \xrightarrow{\text{ps}} f(X)$.

Preuve 1. On pose

$$A \coloneqq \{\omega \in \Omega \mid X_n(\omega) \longrightarrow X(\omega)\} \quad \text{et} \quad A \coloneqq \{\omega \in \Omega \mid Y_n(\omega) \longrightarrow X(\omega)\}.$$

Par hypothèse, on a $\mathbb{P}(A) = \mathbb{P}(B) = 1$, donc $\mathbb{P}(A \cap B) = 1$. Or $A \cap B \subset \{X = Y\}$, donc $\mathbb{P}(X = Y) = 1$.
2. $\Rightarrow$ On suppose que $X_n \xrightarrow{\text{ps}} X$. Soit $\varepsilon > 0$. Alors l'événement $A \coloneqq \bigcup_{n\geqslant 0} \bigcap_{p\geqslant n} \{|X_p - X| < \varepsilon\}$ est de probabilité 1. Par ailleurs, on a

$$\limsup_{n\to+\infty} \{|X_n - X| > \varepsilon\} = \{\omega \in \Omega \mid |X_n(\omega) - X(\omega)| > \varepsilon \text{ une infinité de fois}\} = A^{\mathrm{c}},$$

donc $\mathbb{P}(\limsup_{n\to+\infty} \{|X_n - X| > \varepsilon\}) = 0$.
$\Leftarrow$ On suppose que $\mathbb{P}(\limsup_{n\to+\infty} \{|X_n - X| > \varepsilon\}) = 0$ pour tout $\varepsilon > 0$. La σ-sous-additivité donne alors

$$\mathbb{P}\Big(\bigcup_{\varepsilon\in\mathbb{Q}_+} \limsup_{n\to+\infty} \{|X_n - X| > \varepsilon\}\Big) = 0.$$

Par ailleurs, on a

$$\Big(\bigcup_{\varepsilon\in\mathbb{Q}_+} \limsup_{n\to+\infty} \{|X_n - X| > \varepsilon\}\Big)^{\mathrm{c}} = \bigcap_{\varepsilon\in\mathbb{Q}_+} (\limsup_{n\to+\infty} \{|X_n - X| > \varepsilon\})^{\mathrm{c}} = \bigcap_{\varepsilon\in\mathbb{Q}_+} \bigcup_{n\geqslant 1} \bigcap_{p\geqslant n} \{\{X_p - X\} \leqslant \varepsilon\}$$

$$= \{\omega \in \Omega \mid X_n(\omega) \longrightarrow X(\omega)\}$$

ce qui montre l'implication.

3. On suppose que la série $\sum \mathbb{P}(\{X_n - X\} > \varepsilon)$ converge pour tout $\varepsilon > 0$. Par le premier lemme de Borel-Cantelli, on en déduit que $\mathbb{P}(\limsup_{n\to+\infty}\{|X_n - X| > \varepsilon\}) = 0$ pour tout $\varepsilon > 0$. D'où $X_n \xrightarrow{\text{ps}} X$.

4. Il suffit de traiter le sens réciproque. On suppose que $X_n \xrightarrow{\text{ps}} X$. Comme la suite est mutuellement indépendante, la loi du 0-1 de Kolmogorov assure que la variable X est mesurable par rapport à la tribu asymptotique qui est triviale. Par conséquent, la variable X est constante presque sûrement. Ainsi, les événements $\{|X_n - X| > \varepsilon\}$ sont mutuellement indépendants, donc le deuxième de Borel-Cantelli affirme

$$\mathbb{P}(\limsup_{n\to+\infty}\{|X_n - X| > \varepsilon\}) = 0 \iff \sum_{n=0}^{+\infty}\mathbb{P}(|X_n - X| > \varepsilon) < +\infty.$$

Or $X_n \xrightarrow{\text{ps}} X$, donc le point 2 donne $\mathbb{P}(\limsup_{n\to+\infty}|X_n - X| > \varepsilon) = 0$, donc la série converge. □

◇ Remarque. La notion de convergence presque sûre ne correspond pas à une topologie sur l'ensemble des variables aléatoires réelles sur $(\Omega, \mathscr{F}, \mathbb{P})$. En effet, soit $(X_n)_{n\geqslant 1}$ une suite de variables aléatoires réelles telle que chaque variable X_n suive une loi de Bernoulli de paramètre $1/n$. Alors pour $\varepsilon \in [0,1]$, on a $\mathbb{P}(|X_n| > \varepsilon) = 1/n$. Or la série $\sum 1/n$ ne converge pas, donc la suite $(X_n)_{n\geqslant 1}$ ne converge pas presque sûrement vers 0. Par l'absurde, supposons qu'une telle topologie τ existe. Alors $X_n \not\longrightarrow X$ pour cette topologie, *i. e.* il existe $O \in \tau$ tel que

$$\forall N \geqslant 1,\ \exists n \geqslant N, \quad X_n \in O^{\mathrm{c}}.$$

On peut donc extraire une sous-suite $(X_{n_k})_{k\geqslant 1}$ dont les termes appartiennent à O^{c}. Alors $\mathbb{E}(|X_{n_k}|) = 1/n_k \longrightarrow 0$, donc la suite $(X_{n_k})_{k\geqslant 1}$ tend vers 0 dans $\mathrm{L}^1(\mathbb{P})$, donc on peut en extraire une sous-suite qui converge presque sûrement vers 0 ce qui est impossible.

6.2 Convergence dans L^p

Définition 6.4. Soit $p \geqslant 1$. On dit qu'une suite de variables aléatoires $(X_n)_{n\geqslant 0}$ de $\mathrm{L}^p(\mathbb{P})$ converge dans L^p vers une variable aléatoire $X \in \mathrm{L}^p(\mathbb{P})$ si

$$\mathbb{E}(|X_n - X|^p) \longrightarrow 0.$$

On note alors $X_p \xrightarrow{\mathrm{L}^p} X$

◇ Remarque. On étend la définition à des vecteurs aléatoires en prenant n'importe quelle norme sur $\mathbb{R}^d$.

Propriété 6.5. Soient $(X_n)_{n\geqslant 0}$ une suite de $\mathrm{L}^p(\mathbb{P})$ et $X, Y \in \mathrm{L}^p(\mathbb{P})$. Alors

1. si $X_n \xrightarrow{\mathrm{L}^p} X$ et $X_n \xrightarrow{\mathrm{L}^p} Y$, alors $X = Y$ presque sûrement ;
2. si $X_n \xrightarrow{\mathrm{L}^p} X$ et $q \in [1,p]$, alors $X_n \xrightarrow{\mathrm{L}^q} X$;
3. si $X_n \xrightarrow{\text{ps}} X$ et il existe $Y \in \mathrm{L}^p(\mathbb{P})$ tel que $|X_n| \leqslant Y$ pour tout $n \geqslant 0$, alors $X \in \mathrm{L}^p(\mathbb{P})$ et $X_n \xrightarrow{\mathrm{L}^p} X$;
4. si $X_n \xrightarrow{\mathrm{L}^p} X$, alors il existe une suite strictement croissante d'entiers $(n_k)_{k\geqslant 0}$ telle que $X_{n_k} \xrightarrow{\text{ps}} X$.

Preuve 1. Pour tout $n \geqslant 0$, l'inégalité de Minkowski donne

$$\mathbb{E}(|X - Y|^p)^{1/p} \leqslant \mathbb{E}(|X - X_n|^p)^{1/p} + \mathbb{E}(|X_n - Y|^p)^{1/p}.$$

En laissant tendre n vers $+\infty$, on obtient $\mathbb{E}(|X - Y|^p)^{1/p} = 0$, donc $X = Y$ presque sûrement.

2. Soit $q \in [1,p]$. L'inégalité de Hölder donne $\mathbb{E}(|X_n - Y|^q) \leqslant \mathbb{E}(|X_n - X|^p)^{q/p} \longrightarrow 0$.

3. Pour $n \geqslant 0$, on pose $A_n = \{|X_n| \leqslant Y\}$ de sorte que $\mathbb{P}(A_n) = 1$. La continuité de $\mathbb{P}$ donne $\mathbb{P}(\bigcap_{n\geqslant 0} A_n) = 1$, donc $(\forall n \geqslant 0, |X_n| \leqslant Y)$ presque sûrement. En laissant tend n vers $+\infty$, on obtient que $|X| \leqslant Y$ presque sûrement. On en déduit que $X \in \mathrm{L}^p(\mathbb{P})$. De plus, pour tout $n \geqslant 0$, on a

$$|X_n - X|^p \leqslant (|X_n| + |X|)^p \leqslant 2^p Y^p \in \mathrm{L}^1(\mathbb{P}).$$

Le théorème de convergence dominée assure alors $\mathbb{E}(|X_n - X|)^p \longrightarrow 0$, donc $X_n \xrightarrow{\mathrm{L}^p} X$.

4. Il s'agit d'un résultat du cours d'intégration. □

▷ Exemples. On considère une suite indépendante de variables aléatoires $(X_n)_{n\geqslant 1}$ et une suite $(p_n)_{n\geqslant 1}$ de $[0,1]$ telles que $\mathbb{P}(X_n = x_n) = p_n$ et $\mathbb{P}(X_n = 0) = 1 - p_n$ pour tout $n \geqslant 1$. Que peut-on dire de la convergence presque sûre ou L^p dans les cas suivants ?

– On suppose que $x_n = n$ et $p_n = 1/n$ pour tout $n \geqslant 1$. Pour tout $n \geqslant 1$, on a

$$\mathbb{E}(X_n) = x_n\mathbb{P}(X_n = x_n) + 0\mathbb{P}(X_n = 0) = 1,$$

donc la suite ne tend pas vers 0 dans $\mathrm{L}^1(\mathbb{P})$ et donc dans aucun $\mathrm{L}^p(\mathbb{P})$ avec $p > 1$. Par ailleurs, pour tout $\varepsilon > 0$ et tout $n > \varepsilon$, on a $\mathbb{P}(|X_n| > \varepsilon) = 1/n$. Or la série $\sum 1/n$ diverge, donc la suite $(X_n)_{n\geqslant 1}$ ne converge pas presque sûrement vers 0

– On suppose que $x_n = \sqrt{n}$ et $p_n = 1/n$ pour tout $n \geqslant 1$. Pour tous $\varepsilon > 0$ et $n \in \mathbb{N}^*$ tels que $\varepsilon < \sqrt{n}$, on a

$$\mathbb{P}(|X_n| > \varepsilon) = \mathbb{P}(X_n = \sqrt{n}) = 1/n,$$

donc la série $\sum \mathbb{P}(|X_n| > \varepsilon)$ ne converge pas, donc la suite $(X_n)_{n\geqslant 1}$ ne converge pas presque sûrement. Par ailleurs, on a

$$\mathbb{E}(|X_n|^p) = \sqrt{n}^p \frac{1}{n} = \frac{n^{p/2}}{n},$$

donc la suite $(X_n)_{n\geqslant 1}$ converge dans L^p vers 0 pour tout $p \in [1, 2[$.

– On suppose que $x_n = n^2$ et $p_n = 1/n^2$ pour tout $n \geqslant 1$. De même, la suite converge presque sûrement vers 0, mais elle ne converge dans aucun L^p.

– On suppose que $x_n = 1$ et $p_n = 1/2^n$ pour tout $n \geqslant 1$. La suite converge presque sûrement vers 0 et elle converge vers 0 dans L^p pour tout $p \geqslant 1$.

6.3 Convergence en probabilité

Définition 6.6. On dit qu'une suite de variables aléatoires $(X_n)_{n\geqslant 0}$ converge en probabilité vers une variable aléatoire X si

$$\forall \varepsilon > 0, \quad \mathbb{P}(|X_n - X| > \varepsilon) \xrightarrow[n\to+\infty]{} 0.$$

On note alors $X_n \xrightarrow{\text{proba}} X$.

Proposition 6.7 *(unicité de la limite).* Soient $(X_n)_{n\geqslant 0}$ une suite de variables aléatoires et X et Y deux variables aléatoires telles que $X_n \xrightarrow{\text{proba}} X$ et $X_n \xrightarrow{\text{proba}} Y$. Alors $X = Y$ presque sûrement.

Preuve On a

$$\{X \neq Y\} = \bigcup_{\varepsilon\in\mathbb{Q}_+} \{|X - Y| > \varepsilon\}.$$

Il suffit alors que montrer que $\mathbb{P}(|X - Y| > \varepsilon) = 0$ pour tout $\varepsilon \in \mathbb{Q}_+$. Soit $\varepsilon \in \mathbb{Q}_+$. Or pour tout $n \geqslant 0$, on a

$$|X - Y| \leqslant |X - X_n| + |X_n - Y|,$$

donc

$$\{|X - Y| > \varepsilon\} \subset \{|X - X_n| > \varepsilon/2\} \cup \{|Y - X_n| > \varepsilon/2\},$$

donc

$$\mathbb{P}(|X - Y| > \varepsilon) \leqslant \mathbb{P}(|X - X_n| > \varepsilon/2) + \mathbb{P}(|Y - X_n| > \varepsilon/2) \longrightarrow 0$$

ce qui montre que $\mathbb{P}(|X - Y| > \varepsilon) = 0$. On conclut ensuite par la σ-sous-additivité. □

Proposition 6.8. Soient $f \in \mathscr{C}^0(\mathbb{R}^d, \mathbb{R}^m)$ et $(X_n)_{n\geqslant 0}$ une suite de vecteurs aléatoires convergeant en probabilité vers un vecteur aléatoire X. Alors $f(X_n) \xrightarrow{\text{proba}} f(X)$.

Preuve Soit $M > 0$. Alors le théorème de Heine assure que la fonction f est uniformément continue sur le compact $\overline{\mathrm{B}}(0, M)$. Soit $\varepsilon > 0$. Il existe $\eta_{\varepsilon,M} > 0$ tel que

$$\forall x, y \in \overline{\mathrm{B}}(0, M), \quad \|x - y\| \leqslant \eta_{\varepsilon,M} \quad \Longrightarrow \quad \|f(x) - f(y)\| \leqslant \varepsilon.$$

On veut montrer que

$$\mathbb{P}(\|f(X_n) - f(X)\| > \varepsilon) \longrightarrow 0.$$

Pour tout $n \geqslant 0$, on a

$$\{\|f(X_n) - f(X)\| > \varepsilon\} \cap \{\|X_n\| \leqslant M\} \cap \{\|X\| \leqslant M\} \subset \{\|X_n - X\| > \eta_{\varepsilon,M}\},$$

donc

$$\begin{aligned}\{\|f(X_n) - f(X)\| > \varepsilon\} &\subset \{\|X_n - X\| > \eta_{\varepsilon,M}\} \cup \{\|X_n\| > M\} \cup \{\|X\| > M\} \\ &\subset \{\|X_n - X\| > \eta_{\varepsilon,M}\} \cup \{\|X_n - X\| > 1\} \cup \{\|X\| > M - 1\} \cup \{\|X\| > M\},\end{aligned}$$

donc

$$\mathbb{P}(\|f(X_n) - f(X)\| > \varepsilon) \leqslant \mathbb{P}(\|X_n - X\| > \eta_{\varepsilon,M}) + \mathbb{P}(\|X_n - X\| > 1) + \mathbb{P}(\|X\| > M - 1) + \mathbb{P}(\|X\| > M).$$

En passant à la limite supérieure, on obtient que

$$\limsup_{n\to+\infty} \mathbb{P}(\|f(X_n) - f(X)\| > \varepsilon) \leqslant \mathbb{P}(\|X\| > M-1) + \mathbb{P}(\|X\| > M).$$

De plus, la propriété de la limite monotone donne

$$\mathbb{P}(\|X\| > M) \xrightarrow[M\to+\infty]{} 0 \quad \text{et} \quad \mathbb{P}(\|X\| > M-1) \xrightarrow[M\to+\infty]{} 0.$$

En laissant tendre M vers $+\infty$, on trouve que

$$\limsup_{n\to+\infty} \mathbb{P}(\|f(X_n) - f(X)\| > \varepsilon) = 0$$

ce qui montre la proposition. □

PROPOSITION 6.9. Les convergences presque sûre et L^p impliquent la convergence en probabilité.

Preuve Soient $(X_n)_{n\geqslant 0}$ une suite de variables aléatoires et X une variable aléatoire. On suppose que $X_n \xrightarrow{\text{ps}} X$. Soit $\varepsilon > 0$. Pour $n \geqslant 0$, on a $\mathbb{P}(|X_n - X| > \varepsilon) = \mathbb{E}(\mathbb{1}_{\{|X_n - X| > \varepsilon\}})$. Comme $X_n \xrightarrow{\text{ps}} X$, on a $\mathbb{1}_{\{|X_n - X| > \varepsilon\}} \xrightarrow{\text{ps}} 0$. Le théorème de convergence dominée donne alors $\mathbb{P}(|X_n - X| > \varepsilon) \longrightarrow 0$. D'où $X_n \xrightarrow{\text{proba}} X$.

Soit $p \geqslant 1$. On suppose que les variables X_n et X appartiennent à L^p et vérifient $X_n \xrightarrow{\mathrm{L}^p} X$. Soit $\varepsilon > 0$. Pour tout $n \geqslant 0$, l'inégalité de MARKOV assure

$$\mathbb{P}(|X_n - X| > \varepsilon) = \mathbb{P}(|X_n - X|^p > \varepsilon^p) \leqslant \frac{\mathbb{E}(|X_n - X|^p)}{\varepsilon^p} \longrightarrow 0.$$

D'où $X_n \xrightarrow{\text{proba}} X$. □

CONTRE-EXEMPLE. Soit $(X_n)_{n\geqslant 0}$ une suite de variables aléatoires telles que

$$\mathbb{P}(X_n = n) = 1/n \quad \text{et} \quad \mathbb{P}(X_n = 0) = 1 - 1/n$$

pour tout $n \geqslant 0$. Alors pour tout $\varepsilon > 0$, on a $\mathbb{P}(|X_n| > \varepsilon) = \mathbb{P}(X_n = n) \longrightarrow 0$, donc la suite converge vers 0 en probabilité. Cependant, on a $\mathbb{E}(|X_n|) = 1 \not\longrightarrow 0$, donc la suite ne converge pas vers 0 dans L^1. De plus, la série $\sum \mathbb{P}(|X_n| > 0) = \sum 1/n$, diverge, donc elle ne converge pas vers 0 presque sûrement.

PROPOSITION 6.10. Soient $(X_n)_{n\geqslant 0}$ une suite de variables aléatoires convergeant en probabilité vers une variable aléatoire X. Alors il existe une extraction $(n_k)_{k\geqslant 1}$ telle que $X_{n_k} \xrightarrow{\text{ps}} X$.

Preuve Soit $k \geqslant 1$. Comme $\mathbb{P}(|X_n - X| > 1/k) \longrightarrow 0$, il existe $n_k \in \mathbb{N}$ tel que

$$\forall n \geqslant n_k, \quad \mathbb{P}(|X_n - X| > 1/k) \leqslant 1/k^2.$$

On construit alors une suite strictement croissante d'entiers $(n_k)_{k\geqslant 1}$ telle que

$$\forall k \geqslant 1, \quad \mathbb{P}(|X_{n_k} - X| > 1/k) \leqslant 1/k^2.$$

Alors la série $\sum \mathbb{P}(|X_{n_k} - X| > 1/k)$. Le lemme de BOREL-CANTELLI assure

$$\mathbb{P}(\limsup_{k\to+\infty} \{|X_{n_k} - X| > 1/k\} = 0,$$

donc

$$\mathbb{P}(\liminf_{k\to+\infty} \{|X_{n_k} - X| \leqslant 1/k\} = 1.$$

Cela montre que, presque sûrement, à partir d'un certain rang, on a $|X_{n_k} - X| \leqslant 1/k$, donc $X_{n_k} \xrightarrow{\text{ps}} X$. □

THÉORÈME 6.11. Soient $(X_n)_{n\geqslant 0}$ une suite de variables aléatoires convergeant en probabilité vers une variable aléatoire X. Alors les deux propositions suivantes sont équivalentes :

(i) $X_n \xrightarrow{\text{proba}} X$;

(ii) pour toute extraction $(n_k)_{k\geqslant 0}$, il existe une extraction $(k_p)_{p\geqslant 0}$ telle que $X_{n_{k_p}} \xrightarrow{\text{ps}} X$.

Preuve $\Rightarrow$ On suppose (i). Soit $(n_k)_{k\geqslant 0}$ une extraction. Comme $X_n \xrightarrow{\text{proba}} X$, on a $X_{n_k} \xrightarrow{\text{proba}} X$. De plus, la suite $(\mathbb{P}(|X_{n_k} - X| > \varepsilon))_{k\geqslant 0}$ est une sous-suite de la suite $(\mathbb{P}(|X_n - X| > \varepsilon))_{,\geqslant 0}$. Par la proposition précédente, on obtient le résultat voulue.

$\Leftarrow$ Par l'absurde, on suppose que la suite $(X_n)_{n\geqslant 0}$ ne converge pas en probabilité. Soit $\varepsilon > 0$. Alors

$$\mathbb{P}(|X_n - X| > \varepsilon) \not\longrightarrow 0,$$

donc il existe $\alpha > 0$ et une extraction $(n_k)_{k\geqslant 0}$ tels que

$$\forall k \geqslant 0, \quad \mathbb{P}(|X_{n_k} - X| > \varepsilon) \geqslant \alpha. \tag{$*$}$$

Par l'hypothèse, il existe une extraction $(k_p)_{p\geqslant 0}$ telle que $X_{n_{k_p}} \xrightarrow{\text{ps}} X$, donc $X_{n_{k_p}} \xrightarrow{\text{proba}} X$. Pour $n \geqslant 0$ assez grand, on a alors $\mathbb{P}(|X_{n_{k_p}} - X| < \varepsilon) < \alpha$ ce qui contredit la relation $(*)$. □

Proposition 6.12. Soient $(X_n)_{n\geqslant 0}$ une suite de variables aléatoires de L^p convergeant en probabilité vers une variable aléatoire X de L^p. On suppose qu'il existe une variable aléatoire Y de L^p telle que $|X_n| \leqslant Y$ presque sûrement pour tout $n \geqslant 0$. Alors $X \in \mathrm{L}^p$ et $X_n \xrightarrow{\mathrm{L}^p} X$.

Preuve Montrons que $|X| \leqslant Y$ presque sûrement. Comme $X_n \xrightarrow{\text{proba}} X$, il existe une extraction $(n_k)_{k\geqslant 0}$ telle que $X_{n_k} \xrightarrow{\text{ps}} X$. Par l'hypothèse de domination, on a $|X_{n_k}| \leqslant Y$ presque sûrement pour tout $k \geqslant 0$. En passant à la limite, on a $|X| \leqslant Y$ presque sûrement.

Par l'absurde, supposons que $\mathbb{E}(|X_n - X|^p) \not\longrightarrow 0$. Alors il existe $\alpha > 0$ et une extraction $(m_k)_{k\geqslant 0}$ tels que

$$\forall k \geqslant 0, \quad \mathbb{E}(|X_{m_k} - X|^p) \geqslant \alpha.$$

Or $X_{m_k} \xrightarrow{\text{proba}} X$, donc il existe une extraction $(k_p)_{p\geqslant 0}$ telle que $X_{m_{k_p}} \xrightarrow{\text{ps}} X$. La théorème de convergence dominée donne alors $\mathbb{E}(|X_{m_{k_p}} - X|^p) \longrightarrow 0$ ce qui est impossible. D'où $X_n \xrightarrow{\mathrm{L}^p} X$. □

6.4 Convergence en loi

6.4.1 Convergence étroite

Définition 6.13. On dit qu'une suite de mesures de probabilité $(\mu_n)_{n\geqslant 0}$ sur $\mathbb{R}^d$ converge étroitement vers une mesure de probabilité μ sur $\mathbb{R}^d$ si

$$\forall f \in \mathscr{C}_{\mathrm{b}}(\mathbb{R}), \quad \int_{\mathbb{R}^d} f(x)\,\mathrm{d}\mu_n(x) \longrightarrow \int_{\mathbb{R}^d} f(x)\,\mathrm{d}\mu(x).$$

Proposition 6.14. Soit $H \subset \mathscr{C}_{\mathrm{b}}(\mathbb{R})$ tel que $\mathscr{C}_{\mathrm{c}}(\mathbb{R}) \subset \overline{H}^{\infty}$. Alors une suite de mesures de probabilité $(\mu_n)_{n\geqslant 0}$ sur $\mathbb{R}^d$ converge étroitement vers une mesure de probabilité μ sur $\mathbb{R}^d$ si et seulement si

$$\forall f \in H, \quad \int_{\mathbb{R}^d} f(x)\,\mathrm{d}\mu_n(x) \longrightarrow \int_{\mathbb{R}^d} f(x)\,\mathrm{d}\mu(x). \tag{$*$}$$

Preuve Le sens direct est évident puisque $H \subset \mathscr{C}_{\mathrm{b}}(\mathbb{R})$. Montrons le sens réciproque. On suppose la relation $(*)$.

- *Cas particulier.* On suppose que $H = \mathscr{C}_{\mathrm{c}}(\mathbb{R})$. Pour $p > 0$, on considère une fonction $\theta_p \colon \mathbb{R}_+ \to [0,1]$ telle que $\theta_p = 1$ sur $[0,p]$, affine sur $[p,2p]$ et $\theta_p = 0$ sur $[2p,+\infty[$. Soient $f \in \mathscr{C}_{\mathrm{b}}(\mathbb{R})$, $p > 0$ et $n \geqslant 0$. Alors

$$\int_{\mathbb{R}^d} f(x)\,\mathrm{d}\mu_n(x) = \int_{\mathbb{R}^d} f(x)\theta_p(\|x\|)\,\mathrm{d}\mu_n(x) + \int_{\mathbb{R}^d} f(x)(1-\theta_p(\|x\|))\,\mathrm{d}\mu_n(x).$$

On a

$$\int_{\mathbb{R}^d} f(x)(1-\theta_p(\|x\|))\,\mathrm{d}\mu_n(x) \leqslant \|f\|_\infty \int_{\mathbb{R}^d} (1-\theta_p(\|x\|))\,\mathrm{d}\mu_n(x).$$

On obtient alors que

$$\left|\int_{\mathbb{R}^d} f(x)\,\mathrm{d}\mu_n(x) - \int_{\mathbb{R}^d} f(x)\,\mathrm{d}\mu(x)\right| \leqslant \left|\int_{\mathbb{R}^d} f(x)\theta_p(\|x\|)\,\mathrm{d}\mu_n(x) - \int_{\mathbb{R}^d} f(x)\theta_p(\|x\|)\,\mathrm{d}\mu(x)\right| + \\ \|f\|_\infty\left(2 - \int_{\mathbb{R}^d} \theta_p(\|x\|)\,\mathrm{d}\mu_n(x) - \int_{\mathbb{R}^d} \theta_p(\|x\|)\,\mathrm{d}\mu(x)\right).$$

Comme $f\theta_p \in \mathscr{C}_{\mathrm{c}}(\mathbb{R})$, le premier terme converge vers 0. Comme $\theta_p \in \mathscr{C}_{\mathrm{c}}(\mathbb{R})$, on a

$$\int_{\mathbb{R}^d} \theta_p(\|x\|)\,\mathrm{d}\mu_n(x) \longrightarrow \int_{\mathbb{R}^d} \theta_p(\|x\|)\,\mathrm{d}\mu(x).$$

On en déduit que

$$\limsup_{n\to+\infty} \left|\int_{\mathbb{R}^d} f(x)\,\mathrm{d}\mu_n(x) - \int_{\mathbb{R}^d} f(x)\,\mathrm{d}\mu(x)\right| \leqslant 2\|f\|_\infty\left(1 - \int_{\mathbb{R}^d} \theta_p(\|x\|)\,\mathrm{d}\mu(x)\right).$$

Or pour tout $x \in \mathbb{R}^d$, on a $\theta_p(\|x\|) \longrightarrow 1$ quand $p \to +\infty$ et $|\theta_p(\|x\|)| \leqslant 1$ pour tout $p > 0$. Le théorème de convergence dominée assure donc le résultat.

- *Cas général.* Soient $f \in \mathscr{C}_{\mathrm{c}}(\mathbb{R})$ et $\varepsilon > 0$. Comme $\mathscr{C}_{\mathrm{c}}(\mathbb{R}) \subset \overline{H}$, il existe $h_\varepsilon \in H$ tel que $\|f - h_\varepsilon\|_\infty \leqslant \varepsilon$. Alors

$$\left|\int_{\mathbb{R}^d} f(x)\,\mathrm{d}\mu_n(x) - \int_{\mathbb{R}^d} f(x)\,\mathrm{d}\mu(x)\right| \leqslant \left|\int_{\mathbb{R}^d} h_\varepsilon(x)\,\mathrm{d}\mu_n(x) - \int_{\mathbb{R}^d} h_\varepsilon(x)\,\mathrm{d}\mu(x)\right| + 2\|f - h_\varepsilon\|_\infty.$$

Le premier terme converge quand $\varepsilon \to 0$. On en déduit que la conclusion. □

◇ Remarque. La continuité de f est essentielle. De plus, s'il y a convergence étroite, on n'a pas nécessairement $\mu_n(A) \longrightarrow \mu(A)$ pour tout $A \in (\mathbb{R}^d)$: il suffit de considérer la suite $(\delta_{1/n})_{n\geqslant 1}$ et $A := \{0\}$.

Théorème 6.15. Soient $(\mu_n)_{n\geqslant 0}$ une suite de mesures de probabilité sur $\mathbb{R}^d$ et μ une mesure de probabilité sur $\mathbb{R}^d$. Alors les propositions suivantes sont équivalentes :

(i) la suite $(\mu_n)_{n\geqslant 0}$ converge étroitement vers μ ;
(ii) pour tout fermé F de $\mathbb{R}^d$, on a $\limsup_{n\to+\infty}\mu_n(F)\leqslant\mu(F)$;
(iii) pour tout ouvert O de $\mathbb{R}^d$, on a $\liminf_{n\to+\infty}\mu_n(O)\geqslant\mu(O)$;
(iv) pour tout $B\in\mathscr{B}(\mathbb{R}^d)$ tel que $\mu(\partial B)=0$, on a $\mu_n(B)\longrightarrow\mu(B)$;
(v) pour toute fonction mesurable bornée $f\colon\mathbb{R}^d\to\mathbb{R}$ dont l'ensemble des points de discontinuité est μ-négligeable, on a

$$\int_{\mathbb{R}^d} f(x)\,\mathrm{d}\mu_n(x)\longrightarrow\int_{\mathbb{R}^d} f(x)\,\mathrm{d}\mu(x).$$

Preuve • *(i) ⇒ (ii).* On suppose (i). Soit F un fermé de $\mathbb{R}^d$. Pour $k\geqslant 1$, on considère la fonction

$$f_k\colon x\in\mathbb{R}^d\longmapsto(1+\mathrm{d}(x,F))^{-k}.$$

Cette suite $(f_k)_{k\geqslant 1}$ est une suite de bornée continue et bornée, elle converge simplement vers $\mathbb{1}_F$ en décroissant. Ainsi pour $k\geqslant 1$, on a

$$\limsup_{n\to+\infty}\mu_n(F)=\limsup_{n\to+\infty}\int_{\mathbb{R}^d}\mathbb{1}_F(x)\,\mathrm{d}\mu_n(x)\leqslant\limsup_{n\to+\infty}\int_{\mathbb{R}^d}f_k(x)\,\mathrm{d}\mu_n(x)=\int_{\mathbb{R}^d}f_k(x)\,\mathrm{d}\mu(x).$$

Par ailleurs, le théorème de convergence dominée assure que $\int_{\mathbb{R}^d}f_k(x)\,\mathrm{d}\mu(x)\longrightarrow\mu(F)$ ce qui montre (ii). L'équivalence (ii) ⇔ (iii) se montre en passant au complémentaire.

• *(iii) ⇒ (iv).* On suppose (iii). Soit $B\in\mathscr{B}(\mathbb{R}^d)$ tel que $\mu(\partial B)=\mu(\overline{B}\setminus\mathring{B})=0$. Comme $\mathring{B}\subset B\subset\overline{B}$, les points (ii) et (iii) assure

$$\mu(\mathring{B})\leqslant\liminf_{n\to+\infty}\mu_n(\mathring{B})\leqslant\liminf_{n\to+\infty}\mu_n(B)\leqslant\limsup_{n\to+\infty}\mu_n(B)\leqslant\limsup_{n\to+\infty}\mu_n(\overline{B})\leqslant\mu(\overline{B}).$$

Comme $\mu(\overline{B})=\mu(\mathring{B})$, on en déduit que $\mu(B)=\liminf_{n\to+\infty}\mu_n(B)=\limsup_{n\to+\infty}\mu_n(B)$. D'où (iv).

• *(v) ⇒ (i).* On suppose (v). Soit $f\in\mathscr{C}_{\mathrm{b}}(\mathbb{R})$. Alors l'ensemble des points de discontinuité de f est vide. On obtient alors la convergence étroite et (i).

• *(iv) ⇒ (v).* On suppose (iv). Soit $f\colon\mathbb{R}^d\to\mathbb{R}$ une fonction mesurable bornée. Alors l'ensemble

$$\{t\in\mathbb{R}\mid\mu(\{x\in\mathbb{R}^d\mid f(x)=t\})>0\}$$

est au plus dénombre car c'est l'ensemble des points de discontinuité de la fonction de répartition F_f. Son complémentaire D est donc dense dans $\mathbb{R}^d$. Comme f est bornée, il existe $M>0$ tel que $\|f\|_\infty\leqslant M$. Soit $\varepsilon>0$. Alors il existe une subdivision $-M<t_1<\cdots<t_k<M$ de D telle que $\max_{2\leqslant i\leqslant k}|t_i-t_{i-1}|\leqslant\varepsilon$. Pour $x\in\mathbb{R}^d$, on pose

$$g(x)\coloneqq\sum_{i=1}^{k-1}\mathbb{1}_{[t_i,t_{i+1}[}(f(x))t_i.$$

Alors $\|f-g\|_\infty\leqslant\varepsilon$. Pour tout $n\geqslant 0$, on a donc

$$\left|\int_{\mathbb{R}^d}f(x)\,\mathrm{d}\mu_n(x)-\int_{\mathbb{R}^d}f(x)\,\mathrm{d}\mu(x)\right|\leqslant\left|\int_{\mathbb{R}^d}g(x)\,\mathrm{d}\mu_n(x)-\int_{\mathbb{R}^d}g(x)\,\mathrm{d}\mu(x)\right|+2\varepsilon.$$

Or

$$\int_{\mathbb{R}^d}g(x)\,\mathrm{d}\mu(x)=\sum_{i=1}^{k-1}t_i\mu(\{x\in\mathbb{R}^d\mid f(x)\in[t_i,t_{i+1}[\})$$

et de même pour μ_n. Soit $i\in[\![1,k-1]\!]$. On pose $B\coloneqq\{x\in\mathbb{R}^d\mid f(x)\in[t_i,t_{i+1}[\}$. On note D_f l'ensemble des points de discontinuité de f et on suppose que $\mu(D_f)=0$. Alors

$$\mathring{B}\supset\{x\in D_f^{\mathrm{c}}\mid f(x)\in\,]t_i,t_{i+1}[\}\quad\text{et}\quad\overline{B}\subset\{x\in D_f^{\mathrm{c}}\mid f(x)\in[t_i,t_{i+1}]\}\cup D_f.$$

Par conséquent, on a

$$\partial B\subset\bigcup_{i=1}^{k}\left\{x\in\mathbb{R}^d\;\middle|\;f(x)=t_i\right\}\cup D_f.$$

Comme les réels t_i appartiennent à D^{c} qui est au plus dénombrable, l'ensemble ∂B est négligeable. On en déduit que $\mu_n(B_i)\longrightarrow\mu(B)$ ce qui induit

$$\left|\int_{\mathbb{R}^d}g(x)\,\mathrm{d}\mu_n(x)-\int_{\mathbb{R}^d}g(x)\,\mathrm{d}\mu(x)\right|\longrightarrow 0$$

et ce qui conclut la preuve. ☐

6.4.2 Convergence en loi

DÉFINITION 6.16. On dit qu'une suite de variables aléatoires $(X_n)_{n\geqslant 0}$ converge en loi vers une variable aléatoire X si la suite $(\mathbb{P}_{X_n})_{n\geqslant 0}$ converge étroitement vers $\mathbb{P}_X$. On note alors $X_n \xrightarrow{\text{loi}} X$.

PROPOSITION 6.17. Soient $(X_n)_{n\geqslant 0}$ une suite de variables aléatoires et X une variable aléatoire. Alors les propositions suivantes sont équivalentes :

(i) on a $X_n \xrightarrow{\text{loi}} X$;
(ii) pour toute $f \in \mathscr{C}_{\mathrm{b}}(\mathbb{R})$, on a $\mathbb{E}(f(X_n)) \longrightarrow \mathbb{E}(f(X))$;
(iii) pour toute partie $H \subset \mathscr{C}_{\mathrm{b}}(\mathbb{R})$ telle que $\mathscr{C}_{\mathrm{c}}(\mathbb{R}) \subset \overline{H}$ et toute $f \in H$, on a $\mathbb{E}(f(X_n)) \longrightarrow \mathbb{E}(f(X))$;
(iv) pour tout fermé F de $\mathbb{R}^d$, on a $\limsup_{n\to+\infty} \mathbb{P}(X_n \in F) \leqslant \mathbb{P}(X \in F)$;
(v) pour tout ouvert O de $\mathbb{R}^d$, on a $\limsup_{n\to+\infty} \mathbb{P}(X_n \in O) \geqslant \mathbb{P}(X \in O)$;
(vi) pour tout $B \in \mathscr{B}(\mathbb{R}^d)$ tel que $\mathbb{P}(X \in \partial B) = 0$, on a $\mathbb{P}(X_n \in B) \longrightarrow \mathbb{P}(X \in B)$;
(vii) pour tout point de continuité $t \in \mathbb{R}$ de F_X, on a $F_{X_n}(t) \longrightarrow F_X(t)$.

Preuve L'équivalence (i) $\Leftrightarrow$ (ii) découle du théorème de transfert. En vertu du théorème précédent, il suffit de montrer la dernière équivalence.

On suppose (i). Soit $t \in \mathbb{R}$ un point de continuité de F_x. On pose $B :=]-\infty, t] \in \mathscr{B}(\mathbb{R})$. La fonction $f := \mathbb{1}_B$ est mesurable et bornée. On en déduit que $F_{X_n}(t) = \mathbb{P}(X_n \leqslant t) = \mathbb{E}(f(X_n)) \longrightarrow \mathbb{E}(f(X)) = \mathbb{P}(X \leqslant t) = F_X(t)$.

Réciproquement, on suppose (vii). Soit $f \in \mathscr{C}_{\mathrm{c}}^\infty(\mathbb{R})$. Pour tout $x \in \mathbb{R}$, on a

$$f(x) = \int_{\mathbb{R}} f'(t)\mathbb{1}_{]-\infty,x[}(t)\,\mathrm{d}t.$$

Comme f est à support compact, le théorème de FUBINI donne

$$\mathbb{E}(f(X)) = \int_{\mathbb{R}} f'(t)\mathbb{E}(\mathbb{1}_{]-\infty,X[}(t))\,\mathrm{d}t = \int_{\mathbb{R}} f'(t)(1 - F_X(t))\,\mathrm{d}t.$$

Comme l'ensemble des points de discontinuité de F_X est au plus dénombrable, on a $F_{X_n} \to F_X$ presque partout. De plus, pour tout $t \in \mathbb{R}$, on a $|f'(t)(1 - F_{X_n}(t))| \leqslant |f'(t)| \in \mathrm{L}^1(\mathbb{R}, \mathrm{d}t)$. Le théorème de convergence dominée assure que

$$\mathbb{E}(f(X_n)) = \int_{\mathbb{R}} f'(t)(1 - F_{X_n}(t))\,\mathrm{d}t \longrightarrow \int_{\mathbb{R}} f'(t)(1 - F_X(t))\,\mathrm{d}t = \mathbb{E}(f(X)).$$

On conclut alors en utilisant la densité de $\mathscr{C}_{\mathrm{c}}^\infty(\mathbb{R})$ dans $\mathscr{C}_{\mathrm{c}}(\mathbb{R})$. D'où (ii). ☐

◇ REMARQUE. La convergence en loi peut être définies sans que les variables soient définies sur le même espace probabilisé. Ce mode de convergence ne regarde que les lois et fait fi de l'espace probabilisé. La proposition précédente reste vraie pour des vecteurs aléatoires.

PROPOSITION 6.18. Soient $(X_n)_{n\geqslant 0}$ une suite de variables aléatoires et X et Y deux variables aléatoires telles que $X_n \xrightarrow{\text{loi}} X$ et $X_n \xrightarrow{\text{loi}} Y$. Alors X et Y ont la même loi.

Preuve Soit $f \in \mathscr{C}_{\mathrm{b}}(\mathbb{R}^d)$. Alors $\mathbb{E}(f(X_n)) \to \mathbb{E}(f(X))$ et $\mathbb{E}(f(X_n)) \to \mathbb{E}(f(Y))$, donc $\mathbb{E}(f(X)) = \mathbb{E}(f(Y))$. Comme $\mathscr{C}_{\mathrm{c}}(\mathbb{R}^d) \subset \mathscr{C}_{\mathrm{b}}(\mathbb{R}^d)$, ceci montre que $\mathbb{P}_X = \mathbb{P}_Y$. ☐

▷ EXEMPLES. – Soit $(X_n)_{n\geqslant 1}$ une suite des variables aléatoires telles que $X_n \sim \mathrm{Ber}(1/n)$ pour tout $n \geqslant 1$. Pour toute $\varphi \in \mathscr{C}_{\mathrm{b}}(\mathbb{R})$ et pour tout $n \geqslant 1$, on a $\varphi(X_n) = \varphi(1)\frac{1}{n} + \varphi(0)(1 - \frac{1}{n}) \longrightarrow \varphi(0) = \mathbb{E}(\varphi(0))$. D'où $X_n \xrightarrow{\text{loi}} 0$.
– Soient $(X_n)_{n\geqslant 1}$ une suite de variables aléatoires telle que $X_n \sim \mathscr{U}([\![1, n]\!])$ pour tout $n \geqslant 1$ et U une variable aléatoire telle que $U \sim \mathscr{U}([0, 1])$. Montrons que $X_n/n \xrightarrow{\text{loi}} U$. Pour toute $\varphi \in \mathscr{C}_{\mathrm{b}}(\mathbb{R})$ et $n \geqslant 1$, le théorème de transfert pour les lois discrètes donne

$$\mathbb{E}(\varphi(X_n/n)) = \frac{1}{n}\sum_{k=1}^{n} \varphi(k/n) \longrightarrow \int_0^1 \varphi(x)\,\mathrm{d}x = \mathbb{E}(\varphi(U)).$$

PROPOSITION 6.19. Soient $(X_n)_{n\geqslant 0}$ une suite de variables aléatoires convergeant en loi vers une variable aléatoire X et $g \in \mathscr{C}(\mathbb{R}^d, \mathbb{R}^q)$. Alors $g(X_n) \xrightarrow{\text{loi}} g(X)$.

Preuve Il suffit de remarquer que la fonction $f \circ g \colon \mathbb{R}^d \to \mathbb{R}$ est continue et bornée pour toute fonction continue bornée $f \colon \mathbb{R}^q \to \mathbb{R}$. ☐

COROLLAIRE 6.20. Soit $(X_{n,1}, \dots, X_{n,d})_{n\geqslant 0}$ une suite de vecteurs aléatoires de $\mathbb{R}^d$ convergeant vers une vecteur aléatoire $(X_1, \dots, X_d)$. Alors pour tout $i \in [\![1, d]\!]$, on a $X_{n,i} \xrightarrow{\text{loi}} X_i$.

Preuve Il suffit de considérer les projections $\pi_i \colon \mathbb{R}^d \to \mathbb{R}$ sur la i-ième composante qui sont continues. □

◇ Remarque. La réciproque est fausse. Voici deux contre- exemples.

– Soit X une variable aléatoire telle que $\mathbb{P}(X=1) = \mathbb{P}(X=-1) = 1/2$. Pour $n \geqslant 0$, on pose $X_n \coloneqq X$ si n est pair et $X_n \coloneqq -X$ sinon et on pose $Y_n \coloneqq X$. Comme X et $-X$ ont la même loi, les variables X_n et X ainsi que les variables Y_n et X ont la même loi, donc $X_n \xrightarrow{\text{loi}} X$ et $Y_n \xrightarrow{\text{loi}} X$. Cependant, la suite $(X_n, Y_n)_{n\geqslant 0}$ ne converge pas en loi vers (X,Y) car sinon la suite $(X_n + Y_n)_{n\geqslant 0}$ convergerait en loi vers $2X$ ce qui est impossible car $X_n + Y_n = 0$ pour tout $n \in \mathbb{N}$ impair.

– Soient X et Y deux variables aléatoires indépendantes et de même loi $\mathscr{N}(0,1)$. Pour tous $t \in [0,1]$ et $\xi \in \mathbb{R}$, on a

$$\mathbb{E}(e^{i\xi(\sqrt{t}X + \sqrt{1-t}Y)}) = \mathbb{E}(e^{i\xi\sqrt{t}X})\mathbb{E}(e^{i\xi\sqrt{1-t}Y}) = e^{-\xi^2 t/2} e^{-\xi^2(1-t)/2} = e^{-\xi^2/2},$$

donc $\sqrt{t}X + \sqrt{1-t}Y \sim \mathscr{N}(0,1)$, donc les lois marginales du vecteur $V_t \coloneqq (X, \sqrt{t}X + \sqrt{1-t}Y)$ sont des lois $\mathscr{N}(0,1)$. De plus, on a $\mathbb{E}(X[\sqrt{t}X + \sqrt{1-t}Y]) = \sqrt{2}$, donc la loi du vecteur dépend de t. Soient $t, s \in [0,1]$ tels que $t \neq s$. Pour $n \geqslant 0$, on pose $U_n \coloneqq V_t$ si n est pair et $U_n \coloneqq V_s$ sinon. Alors les lois marginales convergent en loi, mais la suite $(V_n)_{n\geqslant 0}$ ne converge pas en loi.

Théorème 6.21 *(critère de Lévy).* La convergence en loi équivaut à la convergence simple des fonctions caractéristiques.

Preuve Montrons le théorème en dimension une. Soient $(X_n)_{n\geqslant 0}$ une suite de variables aléatoires et X une variable aléatoire.

$\Rightarrow$ On suppose que $X_n \xrightarrow{\text{loi}} X$. Soit $t \in \mathbb{R}$. Comme a fonction $x \in \mathbb{R} \longmapsto e^{itx}$ est continue est bornée, on obtient que $\mathbb{E}(e^{itX_n}) \longrightarrow \mathbb{E}(e^{itX})$. La suite $(\varphi_{X_n})_{n\geqslant 0}$ converge simplement donc vers φ_X.

$\Leftarrow$ On suppose que la suite $(\varphi_{X_n})_{n\geqslant 0}$ converge simplement vers φ_X. Soit $f \in \mathscr{C}_c^\infty(\mathbb{R})$. Alors sa transformée de Fourier $\mathscr{F}f$ appartient à $\mathrm{L}^1(\mathbb{R})$. La formule d'inversion donne alors

$$\forall x \in \mathbb{R}, \quad f(x) = \frac{1}{2\pi}\int_{\mathbb{R}} e^{-itx}\mathscr{F}f(t)\,\mathrm{d}t.$$

En utilisant le théorème de Fubini, on obtient alors que

$$\mathbb{E}(f(X_n)) = \frac{1}{2\pi}\int_{\mathbb{R}} \varphi_{X_n}(t)\mathscr{F}f(t)\,\mathrm{d}t.$$

Le théorème de convergence dominée permet alors de conclure que $\mathbb{E}(f(X_n)) \longrightarrow \mathbb{E}(f(X))$. On utilise enfin la densité ce qui termine la preuve. □

Théorème 6.22 *(Lévy, version forte).* Soit $(\varphi_n)_{n\geqslant 0}$ une suite de fonctions caractéristiques de variables aléatoires réelles X_n. Pour $t \in \mathbb{R}$, on pose $\varphi(t) \coloneqq \lim_{n\to+\infty} \varphi_n(t)$. On suppose que φ est continue en 0. Alors il existe une variable aléatoire X telle que $\varphi_X = \varphi$ et $X_n \xrightarrow{\text{loi}} X$.

Proposition 6.23. Si $X_n \xrightarrow{\text{proba}} X$, alors $X_n \xrightarrow{\text{loi}} X$.

◇ Remarque. La réciproque est fausse : il suffit de reprendre l'exemple précédent.

Preuve On suppose que $X_n \xrightarrow{\text{proba}} X$. Avec le critère de Lévy, il suffit de montrer la convergence simple des fonctions caractéristiques. Soit $t \in \mathbb{R}^d$. Pour tout $n \geqslant 0$, on a

$$|\varphi_{X_n}(t) - \varphi_X(t)| \leqslant \mathbb{E}(\left|e^{itX_n} - e^{itX}\right|) = \mathbb{E}(|e^{it(X_n - X)} - 1|) \leqslant \mathbb{E}(\min(2, |t|\,|X_n - X|)).$$

En effet, pour tout $s \in \mathbb{R}$, on a

$$\left|e^{is} - 1\right| \leqslant 2 \quad \text{et} \quad \left|e^{is} - 1\right| = \left|i\int_0^s e^{ix}\,\mathrm{d}x\right| \leqslant |s|\,.$$

Soit $\varepsilon > 0$. Pour tout $n \geqslant 0$, l'égalité $1 = \mathbb{1}_{|X_n - X| \leqslant \varepsilon} + \mathbb{1}_{|X_n - X| > \varepsilon}$ implique

$$\begin{aligned} |\varphi_{X_n}(t) - \varphi_X(t)| &\leqslant \varepsilon\,|t|\,\mathbb{P}(|X_n - X| \leqslant \varepsilon) + 2\mathbb{P}(|X_n - X| > \varepsilon) \\ &\leqslant \varepsilon\,|t| + 2\mathbb{P}(|X_n - X| > \varepsilon). \end{aligned}$$

On a alors $\lim_{n\to+\infty} |\varphi_{X_n}(t) - \varphi_X(t)| \leqslant \varepsilon\,|t|$. En faisant tendre ε vers 0^+, on a $|\varphi_{X_n}(t) - \varphi_X(t)| \longrightarrow 0$. □

Proposition 6.24. Soient $c \in \mathbb{R}^d$ et $(X_n)_{n\geqslant 0}$ une suite de variables aléatoires définies sur le même espace probabilisé tels que $X_n \xrightarrow{\text{loi}} c$. Alors $X_n \xrightarrow{\text{proba}} c$.

Preuve On se place en dimension une. Soit $\varepsilon > 0$. Pour tout $n \geqslant 0$, on a

$$\mathbb{P}(|X_n - c| > \varepsilon) = \mathbb{P}(X_n < c - \varepsilon) + \mathbb{P}(X_n > c + \varepsilon)$$

$$\leqslant F_{X_n}(c-\varepsilon) - F_{X_n}(c+\varepsilon).$$

Comme $\varepsilon > 0$, la fonction F_c est continue en $c \pm \varepsilon$. Comme $X_n \xrightarrow{\text{loi}} c$, on a

$$F_{X_n}(c-\varepsilon) \longrightarrow F_c(c-\varepsilon) = 0 \quad \text{et} \quad F_{X_n}(c+\varepsilon) \longrightarrow F_c(c+\varepsilon) = 1.$$

On en déduit que $\mathbb{P}(|X_n - c| > \varepsilon) \longrightarrow 0$ pour tout $\varepsilon > 0$. D'où $X_n \xrightarrow{\text{proba}} X$. □

Lemme 6.25 *(Slutsky).* Soient $(X_n)_{n\geqslant 0}$ une suite de variables aléatoires convergeant en loi vers une variable aléatoire X et $(Y_n)_{n\geqslant 0}$ une suite de variables aléatoires convergeant en probabilité vers une constante $c \in \mathbb{R}^d$. Alors $(X_n, Y_n) \xrightarrow{\text{loi}} (X, c)$.

Preuve Pour tout $(t,s) \in \mathbb{R}^2$, on a $\varphi_{(X,c)}(t,s) = e^{isc}\varphi_X(t)$ et

$$\begin{aligned}\left|\varphi_{(X_n,Y_n)}(t,s) - \varphi_{(X,c)}(t,s)\right| &= \left|\mathbb{E}(e^{itX_n}e^{isY_n}) - e^{isc}\mathbb{E}(e^{itX})\right| \\ &= \left|\mathbb{E}(e^{itX_n}[e^{isY_n} - e^{isc}]) + e^{isc}\mathbb{E}(e^{itX_n} - e^{itX})\right| \\ &= \mathbb{E}(\left|e^{isY_n} - e^{isc}\right|) + \left|\mathbb{E}(e^{itX_n} - e^{itX})\right|.\end{aligned}$$

Comme $Y_n \xrightarrow{\text{proba}} c$ et $X_n \xrightarrow{\text{loi}} X$, on en déduit que $\left|\varphi_{(X_n,Y_n)}(t,s) - \varphi_{(X,c)}(t,s)\right| \longrightarrow 0$ en utilisant la fonction continue $x \longmapsto |e^{isx} - e^{isc}|$. Le critère de Lévy donne la conclusion. □

Corollaire 6.26. Si $X_n \xrightarrow{\text{loi}} X$ et $X_n - Y_n \xrightarrow{\text{proba}} 0$, alors $Y_n \xrightarrow{\text{loi}} X$.

Proposition 6.27 *(généralisation du lemme de Fatou).* Si $X_n \xrightarrow{\text{loi}} X$, alors

$$\mathbb{E}(|x|) \leqslant \liminf_{n\to+\infty} \mathbb{E}(|X_n|).$$

Preuve On suppose que $X_n \to X$. Soit $k > 0$. Comme la fonction $x \longmapsto \min(|x|, k)$ est continue et bornée, on a

$$\mathbb{E}(\min(|X|, k)) = \lim_{n\to+\infty} \mathbb{E}(\min(|X_n|, k)) = \liminf_{n\to+\infty} \mathbb{E}(\min(|X_n|, k)) \leqslant \liminf_{n\to+\infty} \mathbb{E}(|X_n|).$$

On conclut par le théorème de convergence dominée en laissant tendre k vers $+\infty$. □

▷ Exemple. Soient $\lambda > 0$ et $(X_n)_{n\geqslant 1}$ une suite de variables aléatoires telle que $X_n \sim \mathscr{G}(\lambda/n)$ pour tout $n \geqslant 1$. Montrons que la suite $(X_n/n)_{n\geqslant 1}$ converge en loi vers la loi exponentielle $\mathscr{E}(\lambda)$. Pour tous $n \geqslant 1$ et $t \in \mathbb{R}$, on a

$$\varphi_{X_n}(t) = \frac{\frac{\lambda}{n}e^{it}}{1 - (1 - \frac{\lambda}{n})e^{it}}, \quad \text{donc} \quad \varphi_{X_n/n}(t) = \varphi_{X_n}(t/n) = \frac{\lambda}{n(e^{-it/n} - 1) + \lambda}.$$

Or $n(e^{-it/n} - 1) = n[\cos(t/n) - 1] - in\sin(t/n) \longrightarrow -it$, donc $\varphi_{X_n/n}(t) \longrightarrow \lambda/(\lambda - it)$. On reconnaît la fonction caractéristique d'une loi exponentielle de paramètre λ ce qui permet de conclure.

Récapitulatif. Voici un schéma récapitulatif des modes de convergence.

$$\begin{tikzcd}
\text{convergence dans } \mathrm{L}^\infty \arrow[r, Rightarrow] & \text{convergence dans } \mathrm{L}^p \arrow[r, Rightarrow] & \text{convergence dans } \mathrm{L}^1 \arrow[dr, Rightarrow] & & \\
& & & \text{convergence en probabilité} \arrow[r, Rightarrow] & \text{convergence en loi} \\
& & \text{convergence presque sûre} \arrow[ur, Rightarrow] & &
\end{tikzcd}$$

6.4.3 Un exemple fondamental : le théorème central limite

Théorème 6.28 *(central limite).* Soit $(X_n)_{n\geqslant 1}$ une suite indépendantes de variables aléatoires de même loi telle que $\mathbb{E}(X_1) = 0$ et $\mathbb{E}(X_1^2) = 1$. Soit N un variable aléatoire telle que $N \sim \mathscr{N}(0,1)$. Alors

$$\frac{X_1 + \cdots + X_n}{\sqrt{n}} \xrightarrow{\text{loi}} N.$$

Pour les besoins de la preuve, on utilisera la branche principale du logarithme complexe, noté log. On rappel que, pour tout $z \in \mathbb{C}$ tel que $|z| < 1$, on a

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

Preuve Remarquons que les variables X_k admettent un moment d'ordre deux, donc leurs fonctions caractéristiques sont de classe $\mathscr{C}^2$. Soient $k \geqslant 1$ et $t \in \mathbb{R}$. Le développement limité de la suite $(\mathbb{E}(e^{itX_k/\sqrt{n}}))_{n\geqslant 1}$ s'écrit, pour tout $n \geqslant 1$,

$$\begin{aligned}\mathbb{E}(e^{itX_k/\sqrt{n}}) &= 1 + \frac{i\mathbb{E}(X_k)}{\sqrt{n}}t - \frac{\mathbb{E}(X_k^2)}{2n}t^2 + \frac{\varepsilon_{t,n}}{n} \\ &= 1 - \frac{t^2}{2n} + \frac{\varepsilon_{t,n}}{n}\end{aligned}$$

où $(\varepsilon_{t,n})_{n\geqslant 1}$ est une suite tendant vers 0. Pour $n \geqslant 1$ assez grand, le module du terme $-\frac{t^2}{2n} + \frac{\varepsilon_{t,n}}{n}$ devient strictement inférieur à 1 ce qui permet d'appliquer le logarithme. En effectuant un développement limité de la fonction $z \longmapsto \log(1+z)$, on peut écrire

$$\log[\mathbb{E}(e^{itX_k/\sqrt{n}})] = -\frac{t^2}{2n} + \frac{\tilde{\varepsilon}_{t,n}}{n}$$

où $(\tilde{\varepsilon}_{t,n})_{n\geqslant 1}$ est une suite tendant vers 0. Finalement, comme les variables ont la même loi, on a

$$\sum_{k=1}^{n} \log[\mathbb{E}(e^{itX_k/\sqrt{n}})] = n\log[\mathbb{E}(e^{itX_1/\sqrt{n}})] = -\frac{t^2}{2} + \tilde{\varepsilon}_{t,n} \xrightarrow[n\to+\infty]{} -\frac{t^2}{2}.$$

En composant par la fonction exponentielle, on a

$$\exp\Big(\sum_{k=1}^{n} \log[\mathbb{E}(e^{itX_k/\sqrt{n}})]\Big) \xrightarrow[n\to+\infty]{} e^{-t^2/2}.$$

On reconnaît la fonction caractéristique d'une loi normale centrée réduite. Or pour $n \geqslant 1$, l'indépendance assure

$$\begin{aligned}\exp\Big(\sum_{k=1}^{n} \log[\mathbb{E}(e^{itX_k/\sqrt{n}})]\Big) &= \prod_{k=1}^{n} \mathbb{E}(e^{itX_k/\sqrt{n}}) \\ &= \exp\Big(it\frac{X_1+\cdots+X_n}{\sqrt{n}}\Big)\end{aligned}$$

ce qui permet de conclure par le critère de Lévy. □

Corollaire 6.29. Soit $(X_n)_{n\geqslant 1}$ une suite indépendantes de variables aléatoires de même loi telle que $\mathbb{E}(X_1) = \mu$ et $\mathrm{Var}(X_1) = \sigma^2$. Soit N un variable aléatoire telle que $N \sim \mathscr{N}(0, \sigma^2)$. Alors

$$\sqrt{n}\left(\frac{X_1+\cdots+X_n}{n} - \mu\right) \xrightarrow{\text{loi}} N.$$

Preuve Pour tout $n \geqslant 1$, on remarque que

$$\sqrt{n}\left(\frac{X_1+\cdots+X_n}{n} - \mu\right) = \frac{\sigma}{\sqrt{n}} \sum_{k=1}^{n} \frac{X_k-\mu}{\sigma}.$$

Les variables $(X_k-\mu)/\sigma$ sont indépendantes, de même loi, centrées et réduites. Le théorème central limite donne alors, pour tout $s \in \mathbb{R}$, la limite

$$\mathbb{E}\Big(\exp\Big[i\frac{s}{\sqrt{n}}\sum_{k=1}^{n}\frac{X_k-\mu}{\sigma}\Big]\Big) \xrightarrow[n\to+\infty]{} e^{-s^2/2}.$$

Pour tout $t \in \mathbb{R}$, la formule précédente pour $s = t\sigma$ assure que

$$\mathbb{E}\Big(it\sqrt{n}\left[\frac{X_1+\cdots+X_n}{n} - \mu\right]\Big) = \mathbb{E}\Big(\exp\Big[it\frac{\sigma}{\sqrt{n}}\sum_{k=1}^{n}\frac{X_k-\mu}{\sigma}\Big]\Big) \xrightarrow[n\to+\infty]{} e^{-\sigma^2t^2/2} = \mathbb{E}(e^{itN}).$$

Le critère de Lévy donne alors la conclusion. □

6.5 Loi des grands nombres

6.5.1 Loi faible des grands nombres

THÉORÈME 6.30. Soit $(X_n)_{n\geqslant 1}$ une suite indépendante de variables aléatoires de $\mathrm{L}^2(\mathbb{P})$ de même loi. Alors la suite $(\frac{1}{n}\sum_{k=1}^n X_k)_{n\geqslant 1}$ converge en probabilité et dans L^2 vers $\mathbb{E}(X_1)$.

Preuve Il suffit de montrer la convergence dans $\mathrm{L}^2(\mathbb{P})$. Pour tout $n \geqslant 0$, on a

$$\begin{aligned}\mathbb{E}\bigg(\bigg|\frac{1}{n}\sum_{k=1}^n X_k - \mathbb{E}(X_1)\bigg|^2\bigg) &= \mathbb{E}\bigg(\bigg|\frac{1}{n}\sum_{k=1}^n [X_k - \mathbb{E}(X_1)]\bigg|^2\bigg) \\ &= \frac{1}{n^2}\sum_{k=1}^n\sum_{\ell=1}^n \mathbb{E}((X_k-\mathbb{E}(X_1))(X_\ell - \mathbb{E}(X_\ell))) \\ &= \frac{1}{n^2}\sum_{k=1}^n \operatorname{Var}(X_k) = \frac{\operatorname{Var}(X_1)}{n} \longrightarrow 0.\end{aligned}$$

□

6.5.2 Loi des grands nombres pour des variable aléatoires décorrélées et bornées dans L^2

THÉORÈME 6.31. Soient $(X_i)_{i\geqslant 1}$ une suite de variables aléatoires et $\mu \in \mathbb{R}$ tels que
- $M \coloneqq \sup_{i\geqslant 1}\mathbb{E}(X_i^2) < +\infty$;
- pour tout $i \geqslant 1$, on ait $\mathbb{E}(X_i) = \mu$;
- pour tout $i, j \geqslant 1$ tels que $i \neq j$, on ait $\mathbb{E}((X_i-\mu)(X_j-\mu)) = 0$.

Alors la suite $(\frac{1}{n}\sum_{k=1}^n X_k)_{n\geqslant 1}$ converge presque sûrement et dans L^2 vers μ.

Preuve • *Convergence dans L^2.* Pour tout $n \geqslant 1$, on a

$$\begin{aligned}\mathbb{E}\bigg(\bigg|\frac{1}{n}\sum_{k=1}^n X_k - \mathbb{E}(X_1)\bigg|^2\bigg) &= \operatorname{Var}\Big(\frac{1}{n}\sum_{k=1}^n X_k\Big) \\ &= \frac{1}{n^2}\sum_{k=1}^n \operatorname{Var}(X_k) \\ &\leqslant \frac{1}{n^2}\sum_{k=1}^n \mathbb{E}(X_k^2) \leqslant \frac{M}{n} \longrightarrow 0.\end{aligned}$$

• *Converge presque sûre.* Soit $\varepsilon > 0$. Pour tout $n \geqslant 1$, l'inégalité de BIENAYMÉ-TCHEBYCHEV donne

$$\mathbb{P}\bigg(\bigg|\frac{1}{n^2}\sum_{k=1}^{n^2} X_k - \mu\bigg| \geqslant \varepsilon\bigg) \leqslant \frac{1}{\varepsilon^2}\operatorname{Var}\Big(\frac{1}{n^2}\sum_{k=1}^{n^2} X_k\Big) \leqslant \frac{M}{\varepsilon^2 n^2}$$

où la série $\sum 1/n^2$ converge. On en déduit

$$\frac{1}{n^2}\sum_{k=1}^{n^2} X_k \xrightarrow{\text{ps}} \mu. \tag{$*$}$$

On remarque que, pour tout $n \geqslant 1$, il existe un unique entier $m_n \geqslant 1$ tel que $m_n^2 \leqslant n \leqslant (m_n+1)^2$ (il suffit de considérer l'entier $m_n \coloneqq \lfloor\sqrt{n}\rfloor$). Pour tout $n \geqslant 1$, on peut écrire

$$\bigg|\frac{1}{n}\sum_{k=1}^n X_k - \mu\bigg| \leqslant \underbrace{\bigg|\frac{1}{n}\sum_{k=1}^{m_n}(X_k-\mu)\bigg|}_{A_n} + \underbrace{\bigg|\frac{1}{n}\sum_{k=m_n^2+1}^{n}(X_k-\mu)\bigg|}_{B_n}.$$

Avec le résultat $(*)$, on a

$$A_n = \bigg|\frac{m_n^2}{n}\frac{1}{m_n^2}\sum_{k=1}^{m_n}(X_k-\mu)\bigg| \longrightarrow 0.$$

Montrons que $B_n \xrightarrow{\text{ps}} 0$. Soit $\varepsilon > 0$. Pour tout $n \geqslant 1$, l'inégalité de BIENAYMÉ-TCHEBYCHEV donne

$$\begin{aligned}\mathbb{P}(|B_n| > \varepsilon) &\leqslant \frac{1}{\varepsilon^2 n^2}\operatorname{Var}\Big(\sum_{k=m_n^2+1}^{n}(X_k-\mu)\Big) \\ &= \frac{1}{\varepsilon^2 n^2}\sum_{k=m_n^2+1}^{n}\operatorname{Var}(X_k) \\ &\leqslant \frac{M}{\varepsilon^2}\frac{(m_n+1)-m_n^2}{n^2} \leqslant \frac{M}{\varepsilon^2}\frac{2\sqrt{n}+1}{n^2} = O\Big(\frac{1}{n^{3/2}}\Big).\end{aligned}$$

Cela montre $B_n \xrightarrow{\text{ps}} 0$. D'où le résultat. □

6.5.3 Loi des grands nombres pour des variable aléatoires indépendantes et de même loi dans L^1

Théorème 6.32. Soient $(X_i)_{i\geqslant 1}$ une suite de variables aléatoires tels que
- les variables X_i soient de même loi ;
- on ait $X_1 \in \mathrm{L}^1(\mathbb{P})$;
- les variables X_i soient mutuellement indépendantes

Alors la suite $(\frac{1}{n}\sum_{k=1}^n X_k)_{n\geqslant 1}$ converge presque sûrement et dans L^1 vers $\mathbb{E}(X_1)$.

Preuve Commençons par montrer la convergence presque sûre.

• *Première étape.* Quitte à considérer les suites $(X_n^+)_{n\geqslant 1}$ et $(X_n^-)_{n\geqslant 1}$ et utiliser la linéarité de l'espérance, on peut supposer que les variables X_n sont positives. Pour $k \geqslant 1$, on pose $Y_k \coloneqq X_k \mathbb{1}_{X_k\leqslant k}$. Comme les variables positives X_i ont la même loi et $X_1 \in \mathrm{L}^1(\mathbb{P})$, on a

$$\sum_{k=1}^{+\infty} \mathbb{P}(X_k > k) = \sum_{k=1}^{+\infty} \mathbb{P}(X_1 > k) < +\infty.$$

Le lemme de Borel-Cantelli donne alors

$$\mathbb{P}(\limsup_{k\to+\infty}\{X_k > k\}) = 0.$$

On en déduit que les suites $(Y_k)_{k\geqslant 1}$ et $(X_k)_{k\geqslant 1}$ ne différent que d'un nombre fini de terme, donc

$$\frac{1}{n}\sum_{k=1}^{n} X_k \xrightarrow{\text{ps}} \mathbb{E}(X_1) \iff \frac{1}{n}\sum_{k=1}^{n} Y_k \xrightarrow{\text{ps}} \mathbb{E}(Y_1).$$

• *Deuxième étape.* On remarque que les variables Y_k sont bornées, donc elles appartient à $\mathrm{L}^2(\mathbb{P})$. De plus, elles sont indépendantes. Soit $\alpha > 1$. Pour $n \geqslant 1$, on pose $k_n \coloneqq \lfloor \alpha^n \rfloor + 1$. Alors l'inégalité de Bienaymé-Tchebychev puis le théorème de Fubini donnent

$$\begin{aligned}\sum_{n=1}^{+\infty} \mathbb{P}\Big(\Big|\frac{1}{k_n}\sum_{i=1}^{k_n} Y_i - \frac{1}{k_n}\sum_{i=1}^{k_n}\mathbb{E}(Y_i)\Big| > \varepsilon\Big) &\leqslant \frac{1}{\varepsilon^2}\sum_{n=1}^{+\infty}\frac{1}{k_n^2}\sum_{i=1}^{k_n}\mathrm{Var}(Y_i)\\ &= \frac{1}{\varepsilon^2}\sum_{i=1}^{+\infty}\mathrm{Var}(Y_i)\sum_{k_n\geqslant i}\frac{1}{k_n^2}\\ &\leqslant \frac{1}{\varepsilon^2}\sum_{i=1}^{+\infty}\mathrm{Var}(Y_i)\sum_{k_n\geqslant i}\frac{1}{\alpha^{2n}}.\end{aligned}$$

Or pour tous $i \geqslant 1$ et $n \geqslant 1$, on a $k_n \geqslant i \Leftrightarrow \alpha^n \geqslant i-1$ et il existe $n_0 \geqslant 1$ tel que $\alpha^{2n_0} \geqslant i-1$, donc il existe des constantes $A_\alpha, B_\alpha \geqslant 0$ telles que

$$\sum_{k_n\geqslant i}\frac{1}{\alpha^{2n}} = \frac{1}{\alpha^{2n_0}}\frac{1}{1-1/\alpha} \leqslant A_\alpha\frac{1}{(i-1)^2} \leqslant B_\alpha\frac{1}{i^2}.$$

De plus, comme $\sum_{i\geqslant k+1} 1/i^2 = O(1/(k+1))$, il existe $C > 0$ tel que

$$\forall k \geqslant 1, \quad \sum_{i\geqslant k+1}\frac{1}{i^2} \leqslant \frac{C}{k+1}.$$

On pose $C_\alpha \coloneqq CB_\alpha$. Alors

$$\begin{aligned}\sum_{n=1}^{+\infty} \mathbb{P}\Big(\Big|\frac{1}{k_n}\sum_{i=1}^{k_n} Y_i - \frac{1}{k_n}\sum_{i=1}^{k_n}\mathbb{E}(Y_i)\Big| > \varepsilon\Big) &\leqslant \frac{1}{\varepsilon^2}\sum_{i=1}^{+\infty}\frac{\mathrm{Var}(Y_i)}{i^2}\\ &\leqslant \frac{B_\alpha}{\varepsilon^2}\sum_{i=1}^{+\infty}\frac{\mathbb{E}(Y_i^2)}{i^2}\\ &\leqslant \frac{B_\alpha}{\varepsilon^2}\sum_{i=1}^{+\infty}\frac{\mathbb{E}(X_i^2\mathbb{1}_{X_i\leqslant i})}{i^2}\\ &\leqslant \frac{B_\alpha}{\varepsilon^2}\sum_{i=1}^{+\infty}\frac{1}{i^2}\sum_{k=0}^{i-1}\mathbb{E}(X_1^2\mathbb{1}_{k\leqslant X_1<k+1})\\ &\leqslant \frac{B_\alpha}{\varepsilon^2}\mathbb{E}\Big(\sum_{k=0}^{+\infty}X_1^2\mathbb{1}_{k\leqslant X_1<k+1}\sum_{i=k+1}^{+\infty}\frac{1}{i^2}\Big)\end{aligned}$$

$$\begin{aligned}&\leqslant \frac{C_\alpha}{\varepsilon^2}\mathbb{E}\Big(\sum_{k=0}^{+\infty}\frac{X_1^2}{k+1}\mathbb{1}_{k\leqslant X_1<k+1}\Big)\\&\leqslant \frac{C_\alpha}{\varepsilon^2}\mathbb{E}\Big(\sum_{k=0}^{+\infty}X_1\mathbb{1}_{k\leqslant X_1<k+1}\Big)=\frac{C_\alpha}{\varepsilon^2}\mathbb{E}(X_1).\end{aligned}$$

D'où

$$\frac{1}{k_n}\sum_{i=1}^{k_n}Y_i-\frac{1}{k_n}\sum_{i=1}^{k_n}\mathbb{E}(Y_i)\xrightarrow{\text{ps}}0.$$

- *Troisième étape.* Il suffit de montrer que

$$\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}(Y_i)\longrightarrow\mathbb{E}(X_1).\tag{$*$}$$

D'une part, pour tout $n\geqslant 1$, on a $Y_i\leqslant X_i$ pour tout $i\in[\![1,n]\!]$, donc $\frac{1}{n}\sum_{i=1}^n\mathbb{E}(Y_i)\leqslant\mathbb{E}(X_1)$. Soit $N_0\geqslant 1$. D'autre part, on a

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}(Y_i)&\geqslant\frac{1}{n}\sum_{i=N_0+1}^{n}\mathbb{E}(Y_i)\\&\geqslant\frac{1}{n}\sum_{i=N_0+1}^{n}\mathbb{E}(X_i\mathbb{1}_{X_i\leqslant N_0})\\&=\frac{n-N_0}{n}\mathbb{E}(X_1\mathbb{1}_{X_1\leqslant N_0})\xrightarrow[n\to+\infty]{}\mathbb{E}(X_1\mathbb{1}_{X_1\leqslant N_0}).\end{aligned}$$

En passant aux limites inférieure et supérieure, on obtient

$$\mathbb{E}(X_1\mathbb{1}_{X_1\leqslant N_0})\leqslant\liminf_{n\to+\infty}\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}(Y_i)\leqslant\mathbb{E}(X_1)\quad\text{et}\quad\mathbb{E}(X_1\mathbb{1}_{X_1\leqslant N_0})\leqslant\limsup_{n\to+\infty}\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}(Y_i)\leqslant\mathbb{E}(X_1).$$

En laissant tendre N_0 vers $+\infty$ et en utilisant le théorème de convergence dominée, on récupère la limite $(*)$. On a donc montré

$$\frac{1}{k_n}\sum_{i=1}^{k_n}Y_i\xrightarrow{\text{ps}}\mathbb{E}(X_1).$$

- *Quatrième étape.* Soit $n\geqslant 1$. Il existe $m\geqslant 1$ tel que $k_m\leqslant n\leqslant k_{m+1}$. La positivité de variables Y_i donne

$$\frac{1}{n}\sum_{i=1}^{k_m}Y_i\leqslant\frac{1}{n}\sum_{i=1}^{n}Y_i\leqslant\frac{1}{n}\sum_{i=1}^{k_{m+1}}Y_i,$$

donc

$$\frac{k_m}{n}\frac{1}{k_m}\sum_{i=1}^{k_m}Y_i\leqslant\frac{1}{n}\sum_{i=1}^{n}Y_i\leqslant\frac{k_{m+1}}{n}\frac{1}{k_{m+1}}\sum_{i=1}^{k_{m+1}}Y_i,$$

donc

$$\frac{k_m}{k_{m+1}}\frac{1}{k_m}\sum_{i=1}^{k_m}Y_i\leqslant\frac{1}{n}\sum_{i=1}^{n}Y_i\leqslant\frac{k_{m+1}}{k_m}\frac{1}{k_{m+1}}\sum_{i=1}^{k_{m+1}}Y_i,$$

Or $k_m/k_{m+1}\longrightarrow 1/\alpha$ et, avec le troisième étape, on en déduit

$$\frac{\mathbb{E}(X_1)}{\alpha}\leqslant\frac{1}{n}\sum_{i=1}^{n}Y_i\leqslant\alpha\mathbb{E}(X_1).$$

- *Dernière étape.* Par conséquent, pour tout $\alpha\in\,]1,+\infty[\,\cap\,\mathbb{Q}$, on a

$$\frac{\mathbb{E}(X_1)}{\alpha}\leqslant\liminf_{n\to+\infty}\frac{1}{n}\sum_{i=1}^{n}Y_i\leqslant\limsup_{n\to+\infty}\frac{1}{n}\sum_{i=1}^{n}Y_i\leqslant\alpha\mathbb{E}(X_1).$$

Comme l'ensemble $]1,+\infty[\,\cap\,\mathbb{Q}$ est dénombrable, on obtient alors le théorème en laissant tendre α vers 1.

Montrons la convergence dans L^1. Soit $M>0$. Pour $k\geqslant 1$, on pose $Y_k^M\coloneqq X_k\mathbb{1}_{|X_k|\leqslant M}$ et $Z_k^M\coloneqq X_k\mathbb{1}_{|X_k|>M}$. Alors pour tout $n\geqslant 1$, on a

$$\frac{1}{n}\sum_{k=1}^{n}X_k-\mathbb{E}(X_1)=\Big(\frac{1}{n}\sum_{k=1}^{n}Y_k^M-\mathbb{E}(Y_1^M)\Big)+\Big(\frac{1}{n}\sum_{k=1}^{n}Z_k^M-\mathbb{E}(Z_1^M)\Big),$$

donc l'inégalité triangulaire assure

$$\mathbb{E}\left(\left|\frac{1}{n}\sum_{k=1}^{n} X_k - \mathbb{E}(X_1)\right|\right) \leqslant \underbrace{\mathbb{E}\left(\left|\frac{1}{n}\sum_{k=1}^{n} Y_k^M - \mathbb{E}(Y_1^M)\right|\right)}_{A_n^M} + \underbrace{2\mathbb{E}(|Z_1^M|)}_{B_M}.$$

D'après la convergence presque sûre que donne ce théorème, on a

$$\frac{1}{n}\sum_{k=1}^{n} Y_k^M - \mathbb{E}(Y_1^M) \xrightarrow{\text{ps}} 0 \quad \text{quand } n \to +\infty$$

où cette suite est bornée, donc le théorème de converge dominée assure $A_n^M \longrightarrow 0$ quand $n \to +\infty$. Donc

$$\limsup_{n\to+\infty} \mathbb{E}\left(\left|\frac{1}{n}\sum_{k=1}^{n} X_k - \mathbb{E}(X_1)\right|\right) \leqslant 2\mathbb{E}(|X_1|\,\mathbb{1}_{|X_1|>M}) \xrightarrow[M\to+\infty]{} 0$$

toujours par le théorème de convergence dominée. Ceci montre la convergence dans L^1. $\square$

Chapitre 7

Vecteurs gaussiens

7.1 Définition et premières propriétés

Définition 7.1. Un vecteur *gaussien* est un vecteur aléatoire $X \coloneqq (X_1, \dots, X_d)$ tel que, pour tout $\lambda \in \mathbb{R}^d$, il existe $\mu_\lambda \in \mathbb{R}$ et $\sigma_\lambda \geqslant 0$ tels que $\langle \lambda, X\rangle \sim \mathscr{N}(\mu_\lambda, \sigma_\lambda^2)$.

◇ Remarque. En particulier, les composantes d'un vecteur gaussien suivent des lois normales.

Définition 7.2. Soit $Y \coloneqq (Y_1, \dots, Y_d)$ un vecteur aléatoire dont les coordonnées sont dans $\mathrm{L}^2(\mathbb{P})$. La *matrice de covariance* de Y est la matrice $\Sigma_Y \coloneqq (\mathrm{Cov}(Y_i, Y_j))_{1 \leqslant i,j \leqslant n}$.

Proposition 7.3. Une matrice $\Sigma \in \mathscr{M}_d(\mathbb{R})$ est la matrice de covariance d'un vecteur aléatoire si et seulement si elle est symétrique positive.

Preuve On suppose que la matrice Σ est une matrice de covariance d'un vecteur aléatoire $(Y_1, \dots, Y_d)$. Alors elle est clairement symétrique. De plus, pour tout $(t_1, \dots, t_d) \in \mathbb{R}^d$, on a

$$\sum_{i=1}^d \sum_{j=1}^d t_i t_j \,\mathrm{Cov}(Y_i, Y_j) = \mathrm{Var}\Big(\sum_{i=1}^d t_i Y_i\Big) \geqslant 0.$$

La matrice Σ est donc symétrique positive.

Réciproquement, on suppose que la matrice Σ est symétrique positive. Soient N_1, …, N_d des variables aléatoires indépendantes et de même loi $\mathscr{N}(0,1)$. On pose $X \coloneqq (N_1, \dots, N_d)\sqrt{\Sigma}$. Alors $\mathbb{E}(X) = (0, \dots, 0)$ et, pour tous $i, j \in [\![1, n]\!]$, on a

$$\begin{aligned}
\mathbb{E}(X_i X_j) &= \mathbb{E}([N\sqrt{\Sigma}]_i [N\sqrt{\Sigma}]_j) \\
&= \mathbb{E}\Big(\Big[\sum_{k=0}^d N_k \sqrt{\Sigma}_{k,i}\Big]\Big[\sum_{\ell=0}^d N_\ell \sqrt{\Sigma}_{\ell,j}\Big]\Big) \\
&= \sum_{k=1}^d \sum_{\ell=1}^d \mathbb{E}(N_k N_\ell) \sqrt{\Sigma}_{k,i} \sqrt{\Sigma}_{\ell,j} \\
&= \sum_{k=1}^d \sqrt{\Sigma}_{k,i} \sqrt{\Sigma}_{k,j} = \sum_{k=1}^d \sqrt{\Sigma}_{i,k} \sqrt{\Sigma}_{k,j} = \Sigma_{i,j},
\end{aligned}$$

donc $\Sigma_X = \Sigma$. □

Proposition 7.4. Soit $X \coloneqq (X_1, \dots, X_d)$ un vecteur gaussien de moyenne $\mu \coloneqq (\mathbb{E}(X_1), \dots, \mathbb{E}(X_d))$. Alors
- pour tout $\lambda \in \mathbb{R}^d$, on a $\langle \lambda, X\rangle \sim \mathscr{N}(\langle \lambda, \mu\rangle, \lambda \Sigma_X {}^{\mathrm{t}}\lambda)$;
- pour tout $\xi \in \mathbb{R}^d$, on a $\mathbb{E}(e^{i\langle \xi, X\rangle}) = \exp(i\langle \xi, \mu\rangle - \frac{1}{2}\xi \Sigma_X {}^{\mathrm{t}}\xi)$.

Preuve 1. Soit $\lambda \coloneqq (\lambda_1, \dots, \lambda_d) \in \mathbb{R}^d$. Par définition, la variable aléatoire $\langle \lambda, X\rangle$ suit une loi $\mathscr{N}(\mu_\lambda, \sigma_\lambda^2)$ pour des réels $\mu_\lambda \in \mathbb{R}$ et $\sigma_\lambda \geqslant 0$. On obtient alors

$$\mu_\lambda = \sum_{i=1}^d \lambda_i \mathbb{E}(X_i) = \langle \lambda, \mu\rangle \quad \text{et} \quad \sigma_\lambda^2 = \mathrm{Cov}\Big(\sum_{i=1}^d \lambda_i X_i, \sum_{j=1}^d \lambda_j X_j\Big) = \sum_{i=1}^d \sum_{j=1}^d \lambda_i \lambda_j \,\mathrm{Cov}(X_i, X_i) = \lambda \Sigma_X {}^{\mathrm{t}}\lambda.$$

2. On rappel que, pour une variable gaussienne $Z \sim \mathscr{N}(\mu, \sigma^2)$, on a $\mathbb{E}(e^{itZ}) = \exp(it\mu - \frac{1}{2}t^2\sigma^2)$ pour tout $t \in \mathbb{R}$. Il suffit alors d'appliquer cette remarque avec $t = 1$ et $Z = \langle \xi, X\rangle$. □

◇ Remarque. Le matrice Σ_X et le vecteur μ caractérisent alors entièrement la loi de X. On note $X \sim \mathscr{N}(\mu, \Sigma_X)$.

Proposition 7.5. Soit $X \coloneqq (X_1, \dots, X_d)$ un vecteur gaussien. Alors les variables X_i sont mutuellement indépendantes si et seulement si, pour tous $i, j \in [\![1, n]\!]$ tels que $i \neq j$, on a $\mathrm{Cov}(X_i, X_j) = 0$.

Preuve Le sens direct est évident. Réciproquement, on suppose que les variables X_i sont décorrélées. Alors la matrice Σ_X est diagonale. On en déduit que, pour tout $(\xi_1, \dots, \xi_d) \in \mathbb{R}^d$, on a

$$\begin{aligned}\mathbb{E}\Big(\exp\Big(i\sum_{k=1}^{d}\xi_k X_k\Big)\Big) &= \exp\Big(i\sum_{k=1}^{d}\xi_k\mu_k - \frac{1}{2}\sum_{k=1}^{d}(\Sigma_X)_{k,k}\xi_k^2\Big)\\ &= \prod_{k=1}^{d}\exp(i\xi_k\mu_k - \tfrac{1}{2}(\Sigma_X)_{k,k}\xi_k^2)\\ &= \prod_{k=1}^{d}\mathbb{E}(\exp(i\xi_k X_k)).\end{aligned}$$

Cela montre l'indépendance. □

NOTATION. Pour une matrice aléatoire $M \coloneqq (M_{i,j})_{1\leqslant i,j\leqslant n}$, on note $\mathbb{E}(M) \coloneqq (\mathbb{E}(M_{i,j}))_{1\leqslant i,j\leqslant n}$.

PROPOSITION 7.6. Soit $X \sim \mathscr{N}(\mu, \Sigma)$ et $N \sim \mathscr{N}(0, I_d)$ deux vecteurs gaussiens. Alors X et $\mu + N\sqrt{\Sigma}$ ont la même loi.

Preuve Comme un vecteur gaussien est caractérisé par son espérance et sa matrice de covariance, il suffit de montrer qu'elles sont égales. On a bien $\mathbb{E}(\mu + N\sqrt{\Sigma}) = \mu = \mathbb{E}(X)$ et

$$\mathbb{E}({}^{\mathrm{t}}(\mu + N\sqrt{\Sigma} - \mathbb{E}(\mu))(\mu + N\sqrt{\Sigma} - \mathbb{E}(\mu))) = \mathbb{E}(\sqrt{\Sigma}\,{}^{\mathrm{t}}NN\sqrt{\Sigma}) = \sqrt{\Sigma}\mathbb{E}({}^{\mathrm{t}}NN)\sqrt{\Sigma} = \Sigma.$$ □

7.2 Vecteurs gaussiens et densité

THÉORÈME 7.7. Soit $X \coloneqq (X_1, \dots, X_d)$ une vecteur gaussien de loi $\mathscr{N}(\mu, \Sigma)$. Alors il admet une densité si et seulement si $\det\Sigma > 0$. Dans ce cas, sa densité est donnée par

$$f_X(x) = \frac{1}{\sqrt{(2\pi)^d\det\Sigma}}\exp\Big(-(x-\mu)\frac{1}{2}\Sigma^{-1}\,{}^{\mathrm{t}}(x-\mu)\Big).$$

Preuve On suppose $\det\Sigma = 0$. Alors il existe $z \in \mathbb{R}^d \setminus \{0\}$ tel que $z\Sigma = 0$. On obtient $\operatorname{Var}(\langle z, X\rangle) = z\Sigma\,{}^{\mathrm{t}}z = 0$, donc la variable $\langle z, X\rangle$ est presque sûrement constante. Alors le vecteur X est à valeurs dans un hyperplan, donc il n'admet pas de densité.

Réciproquement, on suppose $\det\Sigma > 0$. Soit $N \sim \mathscr{N}(0, I_d)$ un vecteur gaussien. Soit $A \in \mathscr{B}(\mathbb{R}^d)$. Avec la proposition précédente, on obtient

$$\mathbb{P}(X \in A) = \mathbb{E}(\mathbb{1}_A(\mu + N\sqrt{\Sigma})) = \int_{\mathbb{R}^d}\mathbb{1}_A(\mu + x\sqrt{\Sigma})\exp(-\tfrac{1}{2}\|x\|^2)\frac{\mathrm{d}x}{\sqrt{(2\pi)^d}}.$$

En effectuant le changement de variables $y = \mu + x\sqrt{\Sigma}$, on a

$$\begin{aligned}\mathbb{P}(X \in A) &= \int_{\mathbb{R}^d}\mathbb{1}_A(y)\exp(-\tfrac{1}{2}\|(y-\mu)\sqrt{\Sigma}^{-1}\|^2)\frac{\mathrm{d}y}{\sqrt{(2\pi)^d}\det\sqrt{\Sigma}}\\ &= \int_{\mathbb{R}^d}\mathbb{1}_A(y)\exp(-\tfrac{1}{2}(y-\mu)\Sigma^{-1}\,{}^{\mathrm{t}}(y-\mu))\frac{\mathrm{d}y}{\sqrt{(2\pi)^d\det\Sigma}}.\end{aligned}$$ □

7.3 Application linéaire de vecteurs gaussiens

PROPOSITION 7.8. Soient X un vecteur gaussien de dimension d de loi $\mathscr{N}(\mu, \Sigma)$ et $A \in \mathscr{M}_{d,m}(\mathbb{R})$. Alors

$$XA \sim \mathscr{N}(\mu A, {}^{\mathrm{t}}A\Sigma A).$$

Preuve On a $\mathbb{E}(XA) = \mathbb{E}(X)A = \mu A$ et

$$\mathbb{E}({}^{\mathrm{t}}(XA - \mu A)(XA - \mu A)) = \mathbb{E}({}^{\mathrm{t}}A\,{}^{\mathrm{t}}(X-\mu)(X-\mu)A) = {}^{\mathrm{t}}A\mathbb{E}({}^{\mathrm{t}}(X-\mu)(X-\mu))A = {}^{\mathrm{t}}A\Sigma\,{}^{\mathrm{t}}A.$$ □

◇ REMARQUE. Si (X, Y) est un vecteur gaussien, alors X et Y suivent des lois gaussiennes. La réciproque est fausse. Par exemple, soit (U, V) un vecteur gaussien de loi $\mathscr{N}(0, I_2)$. On pose $X \coloneqq U$ et $Y \coloneqq \operatorname{sgn}(V)U$. Alors la variable X suit une loi gaussienne $\mathscr{N}(0, 1)$. De plus, pour toute fonction $\varphi \in \mathscr{C}_{\mathrm{c}}(\mathbb{R}^2, \mathbb{R})$, on a

$$\begin{aligned}\mathbb{E}(\varphi(Y)) &= \mathbb{E}(\varphi(U)\mathbb{1}_{\{V>0\}}) + \mathbb{E}(\varphi(-U)\mathbb{1}_{\{V<0\}})\\ &= \mathbb{E}(\varphi(U))\mathbb{P}(V > 0) + \mathbb{E}(\varphi(-U))\mathbb{P}(V < 0)\end{aligned}$$

$$= \frac{1}{2}\mathbb{E}(\varphi(U)) + \frac{1}{2}\mathbb{E}(\varphi(-U)) = \mathbb{E}(\varphi(U)).$$

On en déduit que la variable Y suit une loi gaussienne $\mathscr{N}(0,1)$. Cependant, la variable $X+Y$ possède un atome en 0 et n'est pas nulle, donc elle n'est pas gaussienne, donc le vecteur (X,Y) n'est pas gaussien.