# Fondements des probabilités

# Table des matières

# Chapitre 1

# Espace de probabilité et vocabulaire probabiliste

## 1.1 Espace de probabilité

L'espace fondamental est un triplet généralement noté $(\Omega, \mathfrak{F}, \mathbb{P})$. L'ensemble $\Omega$ est un ensemble non vide qui décrit toutes les réalisations possibles d'une expérience aléatoire. Un point $\omega \in \Omega$ est une réalisation de l'expérience.

**Exemples :** 1. Si on lance un dé,

$$\Omega = \{1, 2, 3, 4, 5, 6\} \qquad \text{(ensemble fini)}$$

2. On lance une pièce jusqu'à obtenir « face ». Dans ce cas,

$$\Omega = \{F, PF, PPF, PPPF, \ldots\} \qquad \text{(ensemble infini dénombrable)}$$

3. On considère la durée de vie d'une ampoule. Dans ce cas,

$$\Omega = \mathbb{R}_+ \qquad \text{(non dénombrable)}$$

Au lieu de s'intéresser à toute l'expérience, on ne s'intéresse souvent qu'à un certain type de réalisations. Un évènement $A$ est une combinaison de résultats possibles. On souhaite associer à un événement $A$ un nombre $\mathbb{P}(A)$ qui permet de quantifier les chances de réalisation de $A$. Pour cela, on introduit la notion de tribu.

Soit $\Omega$ un ensemble. $\mathfrak{P}(\Omega)$ est l'ensemble des parties de $\Omega$.

**Définition 1.1**

*Une tribu $\mathcal{F}$ sur $\Omega$ est un sous ensemble de $\mathfrak{P}(\Omega)$ tel que :*

- $\Omega \in \mathcal{F}$ ;
- *si $A \in \mathcal{F}$ alors $A^c \in \mathcal{F}$ (stabilité par passage au complémentaire) ;*
- *si $(A_i)_{i\geq 0}$ avec $A_i \in \mathcal{F}$ alors $\bigcup_{i\geq 0} A_i \in \mathcal{F}$ (stabilité par union dénombrable).*

**Exemples :** 1. Tribu triviale $\mathcal{F} = \{\varnothing, \Omega\}$.
2. Tribu grossière $\mathcal{F} = \mathfrak{P}(E)$.
3. $\mathcal{F} = \{\varnothing, A, A^c, \Omega\}$ avec $A \in \mathfrak{P}(\Omega)$.

**Proposition 1.2**

*Si $\mathcal{F}$ est une tribu sur $\Omega$, $\varnothing \in \mathcal{F}$, $\mathcal{F}$ est stable par intersection finie ou dénombrable, union finie, par différence,...*

*Une intersection quelconque de tribus est une tribu.*

Soit $A \subset \mathfrak{P}(\Omega)$, on définit $\sigma(A)$ comme la plus petite tribu contenant $A$ :

$$\sigma(A) = \bigcap_{\substack{\mathcal{F} \text{ tribu} \\ A \in \mathcal{F}}} \mathcal{F}$$

$\sigma(A)$ est appelée tribu engendrée par $A$.

**Exemple :** Si $\Omega$ est topologique, la tribu borélienne $\mathcal{B}(\Omega)$ est la tribu engendrée par les ouverts.

Si $\Omega = \mathbb{R}$, $\mathcal{B}(\mathbb{R})$ est engendrée par les intervalles de la forme $]s, t[$, avec $s, t \in \mathbb{R}$ et même $]-\infty, t[,\ t \in \mathbb{R}, \ldots$

Si $\Omega = \mathbb{R}^d$, $\mathcal{B}(\mathbb{R}^d)$ est engendrée par les pavés $\displaystyle\prod_{i=1}^{d} ]s_i, t_i[$.

**Définition 1.3**

*On considère un espace mesurable $(\Omega, \mathcal{F})$. Une (mesure de) probabilité sur $(\Omega, \mathcal{F})$ est une application $\mathbb{P} : \mathcal{F} \to [0, 1]$ telle que :*

– $\mathbb{P}(\varnothing) = 0$ ;

– *si $(A_i) \in \mathcal{F}$ avec $A_i \cap A_j = \varnothing\ \forall i \neq j$ alors*

$$\mathbb{P}\left(\bigcup_{i \geq 0} A_i\right) = \sum_{i \geq 0} \mathbb{P}(A_i) \qquad \sigma\text{-additiivté} ;$$

– $\mathbb{P}(\Omega) = 1$.

*Le triplet $(\Omega, \mathcal{F}, \mathbb{P})$ où $(\Omega, \mathcal{F})$ est un espace mesurable et $\mathbb{P}$ une probabilité est appelé espace de probabilité.*

**Exemples :** 1. Mesure de Dirac. Soit $a \in \Omega$.

$$\delta_a(A) = \begin{cases} 1 & \text{si } a \in A \\ 0 & \text{sinon} \end{cases}$$

2. $([0, 1], \mathcal{B}([0, 1]), \mathrm{d}\lambda)$ avec $\mathrm{d}\lambda$ mesure de Lebesgue.
3. $(\mathbb{N}, \mathfrak{P}(\mathbb{N}))$. Une probabilité sur cet espace est de la forme

$$\mathbb{P} = \sum_{n \geq 0} p_n \delta_n$$

où $p_n \in [0,1]$ tels que $\sum_{n\geq 0} p_n = 1$. Si $A \in \mathfrak{P}(\mathbb{N})$,

$$\mathbb{P}(A)) = \sum_{n\geq 0} p_n \mathbf{1}_A(n).$$

## 1.2 Propriétés élémentaires

Ici $\cup$ est interprété comme « ou », $\cap$ est interprété comme « et ».

**Proposition 1.4**

*Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité.*

*$(i)$ Soient $A, B \in \mathcal{F}$. Si $A \cap B = \varnothing$, alors*

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

*$(ii)$ Soit $A \in \mathcal{F}$.*

$$\mathbb{P}(A^c) = 1 - \mathbb{P}(A).$$

*$(iii)$ Soient $A, B \in \mathcal{F}$. Si $A \subset B$, alors*

$$\mathbb{P}(A) \subset \mathbb{P}(B)$$

*$(iv)$ Soient $A, B \in \mathcal{F}$.*

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

*$(v)$ Soit $(A_i) \in \mathcal{F}$.*

$$\mathbb{P}\left(\bigcup_{i\geq 0} A_i\right) \leq \sum_{i\geq 0} \mathbb{P}(A_i)$$

*$(vi)$ Soit $(A_i) \in \mathcal{F}$ une suite croissante d'événements : $A_i \subset A_{i+1}$. Alors*

$$\mathbb{P}\left(\bigcup_{i\geq 0} A_i\right) = \lim_{i\to+\infty} \mathbb{P}(A_i) = \sup_{i\geq 0} \mathbb{P}(A_i).$$

*$(vii)$ Soit $(A_i) \in \mathcal{F}$ une suite décroissante d'événements : $A_{i+1} \subset A_i$. Alors*

$$\mathbb{P}\left(\bigcap_{i\geq 0} A_i\right) = \lim_{i\to+\infty} \mathbb{P}(A_i) = \inf_{i\geq 0} \mathbb{P}(A_i).$$

$\triangleright$ $(ii)$ $\Omega = A \cup A^c$ et $\mathbb{P}(\Omega) = 1 = \mathbb{P}(A) + \mathbb{P}(A^c)$.

$(iv)$ Soient $A, B \in \mathcal{F}$, $A \cup B = (A \backslash B) \sqcup (B \backslash A) \sqcup (A \cap B)$ Alors

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

Par ailleurs, $A = (A \backslash B) \sqcup (A \cap B)$ donc

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

Ainsi,

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

□

**Proposition 1.5 (*Formule de Poincaré*)**

*Soient* $A_1, \dots, A_n \in \mathcal{F}$

$$\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} \mathbb{P}(A_i) + \sum_{k=2}^{n} (-1)^{k+1} \sum_{1 \le i_1 < i_2 < \cdots < i_k \le n} \mathbb{P}(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}).$$

▷ Par récurrence (exercice). □

Pour trois événements $A, B, C \in \mathcal{F}$, on obtient

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

## 1.3 $\liminf$ et $\limsup$ d'événements

Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité. Soit $(A_n)_{n \ge 0}$ une suite d'événements.

**Définition 1.6**

*On définit*

$$\liminf_{n \to +\infty} A_n = \bigcup_{n \ge 0} \bigcap_{k > n} A_k$$

*et*

$$\limsup_{n \to +\infty} A_n = \bigcap_{n \ge 0} \bigcup_{k > n} A_k.$$

**Rappel :** Si $(u_n)$ est une suite de réels,

$$\liminf_{n \to +\infty} u_n = \sup_n \inf_{k \ge n} u_k \le \limsup_{n \to +\infty} u_n = \inf_n \sup_{k \ge n} u_k.$$

On remarque que

$$\liminf_n A_n \subset \limsup_n A_n$$

et

$$(\liminf A_n)^c = \limsup (A_n)^c.$$

Si $\omega \in \liminf A_n$, il existe $n \ge 0$ tel que $\forall\ k > n,\ \omega \in A_k$ : « à partir d'un certain rang, $\omega$ est dans tous les $A_k$ »ou « $\omega$ est dans tous les $A_n$ sauf un nombre fini ».

Si $\omega \in \limsup A_n$, pour tout $n \ge 0$, il existe $k > n$ tel que $\omega \in A_k$ : « $\omega$ est dans une infinité de $A_n$ ».

**Proposition 1.7**

$$\mathbf{1}_{\liminf A_n} = \liminf \mathbf{1}_{A_n},$$

$$\mathbf{1}_{\limsup A_n} = \limsup \mathbf{1}_{A_n},$$
$$\mathbb{P}(\liminf A_n) \leq \liminf \mathbb{P}(A_n) \leq \limsup \mathbb{P}(A_n) \leq \mathbb{P}(\limsup A_n).$$

▷ cf. TD. □

## 1.4 Classe monotone

**Définition 1.8**

*Une famille $\mathcal{M}$ de parties de $\Omega$ est appelée classe monotone si :*
- $\Omega \in \mathcal{M}$ ;
- *$\mathcal{M}$ est stable par différence : si $A, B \in \mathcal{M}$ et $B \subset A$, alors $A \backslash B \in \mathcal{M}$ ;*
- *$\mathcal{M}$ est stable par union croissante : si $(A_i) \in \mathcal{M}$ et $A_i \subset A_{i+1}$ alors $\bigcup_{i \geq 0} A_i \in \mathcal{M}$.*

Soit $A \subset \mathfrak{P}(\Omega)$, on note $\mathcal{M}(A)$ la classe monotone engendrée par $A$ :

$$\mathcal{M}(A) = \bigcap_{\substack{A \in \mathcal{M} \\ \mathcal{M} \text{ classe monotone}}} \mathcal{M}.$$

**Remarques :** – Une tribu est une classe monotone.
– Une classe monotone stable par intersection finie est une tribu.

**Théorème 1.9 (*Théorème des classes montones*)**

*Soit $\mathcal{C} \subset \mathfrak{P}(\Omega)$ stable par intersection finie. Alors*

$$\mathcal{M}(\mathcal{C}) = \sigma(\mathcal{C}).$$

**Application :** Théorème de Dynkin : Soient $\mathcal{C} \subset \mathfrak{P}(\Omega)$ stable par intersection finie, $\mathbb{P}_1$ et $\mathbb{P}_2$ deux probabilité sur $\Omega$ qui coïncident sur $\mathcal{C}$. Alors, elles coïncident sur $\sigma(\mathcal{C})$.

▷ Considérons $\mathcal{M} = \{A \in \mathfrak{P}(\Omega),\ \mathbb{P}_1(A) = \mathbb{P}_2(A)\}$. On a :
– $\Omega \in \mathcal{M}$ car $\mathbb{P}_1(\Omega) = 1 = \mathbb{P}_2(\Omega)$ ;
– si $A, B \in \mathcal{M}$ et $B \subset A$, $A \backslash B \in \mathcal{M}$ car

$$\mathbb{P}_1(A \backslash B) \underset{\mathbb{P}_1 \text{ proba}}{=} \mathbb{P}_1(A) - \mathbb{P}_1(B) \underset{A,B \in \mathcal{M}}{=} \mathbb{P}_2(A) - \mathbb{P}_2(B) \underset{\mathbb{P}_2 \text{ proba}}{=} \mathbb{P}_2(A \backslash B)\ ;$$

– si $(A_i)$ est une suite croissante de $\mathcal{M}$,

$$\mathbb{P}_1\left(\bigcup_{i \geq 0} A_i\right) = \lim_{i \to +\infty} \mathbb{P}_1(A_i) = \lim_{i \to +\infty} \mathbb{P}_2(A_i) = \mathbb{P}_2\left(\bigcup_{i \geq 0} A_i\right)$$

donc $\bigcup_{i \geq 0} \in \mathcal{M}$.

Par conséquent, $\mathcal{M}$ est une classe monotone. Comme $\mathcal{C} \subset \mathcal{M}$, on conclut par le théorème des classes monotones. □

# Chapitre 2

# Variables aléatoires

**Exemple :** On lance dix fois un dé à six faces. À chaque lancer, le joueur gagne 2 euros si le résultat est 5 ou 6, 1 euros si c'est un 4, rien si c'est un 3 et perd 1 euros si c'est un 2, 3 euros si c'est un 5.

$$\Omega = \{1, 2, 3, 4, 5, 6\}^{10}$$
$$\mathcal{F} = \mathfrak{P}(\Omega)$$
$$\forall\, \omega \in \Omega \qquad \mathbb{P}(\{\omega\}) = \left(\frac{1}{6}\right)^{10}.$$

Pour le joueur, la description complète de l'expérience est inutile, il veut juste connaître son gain final. Le fain est une application de $\Omega$ à valeurs dans $\{-30, \ldots, 20\}$. On parle de variable aléatoire.

## 2.1 Définition et propriétés

Soient $(\Omega, \mathcal{F})$ et $(E, \mathcal{E})$ deux espaces mesurables et $X$ une application de $\Omega$ dans $E$. On note, pour $A \in \mathcal{E}$,

$$\{X \in A\} = X^{-1}(A) = \{\omega \in \Omega,\ X(\omega) \in A\}.$$

Si $A$ est un singleton, $A = \{a\}$, on fait l'abus de notation suivant :

$$\{X \in \{a\}\} = \{X = a\}.$$

**Définition 2.1**

*On appelle variable aléatoire (v.a.) toute application mesurable $X$ d'un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$ à valeurs dans un espace mesurable $(E, \mathcal{E})$ mesurable, c'est-à-dire $X^{-1}(\mathcal{E}) \subset \mathcal{F}$.*

Si $(E, \mathcal{E}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ on parle variable aléatoire réelle.

Si $(E, \mathcal{E}) = (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ on parle de vecteur aléatoire. On peut écrire $X = (X_1, \ldots, X_d)$ avec $X_i : (\Omega, \mathcal{F}, \mathbb{P}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ une variable aléatoire réelle. Dans ce cas, $X_i$ s'appelle la $i$-ème marginale du vecteur $X$.

Si $(E, \mathcal{E}) = (\mathbb{N}, \mathfrak{P}(\mathbb{N}))$ on parle de variable aléatoire entière.

**Exemple :** Une fonction étagée est une fonction mesurable qui prend un nombre fini de valeurs :

$$X = \sum_{i=1}^{n} x_i \mathbf{1}_{A_i} \qquad \text{avec } x_i \in E,\ A_i \in \mathcal{F}.$$

Rappel : toute variable aléatoire est limite simple de fonctions étagées et si la variable aléatoire est positive, la limite peut être choisie croissante.

**Proposition 2.2**

*Si $\mathcal{E} = \sigma(\mathcal{C})$, $X$ est une variable aléatoire si et seulement si $X^{-1}(\mathcal{C}) \subset \mathcal{F}$.*

**Conséquence :** $X$ est une variable aléatoire réelle si et seulement si $\{X \leq t\} \in \mathcal{F}\ \forall\ t \in \mathbb{R}$, car les intervalles $]-\infty, t],\ t \in \mathbb{R}$, engendrent la tribu borélienne.

▷ $\Rightarrow$ OK

$\Leftarrow$ Supposons que $X^{-1}(\mathcal{C}) \subset \mathcal{F}$. On définit

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

- $\mathcal{C} \subset \mathcal{A}$.

- $X^{-1}(E) = \{\omega \in \Omega,\ X(\omega) \in E\} = \Omega \in \mathcal{F}$ donc $E \in \mathcal{A}$.

- Soit $A \in \mathcal{A}$

$$X^{-1}(A^c) = \{\omega \in \Omega,\ X(\omega) \notin A\} = (\underbrace{X^{-1}(A)}_{\in \mathcal{F}})^c \in \mathcal{F} \text{ (tribu)}$$

donc $A^c \in \mathcal{A}$.

- Soit $(A_i)_{i \geq 0}$, $A_i \in \mathcal{A}$.

$$X^{-1}\left(\bigcup_{i \geq 0} A_i\right) = \{\omega \in \Omega,\ X(\omega) \in \bigcup_{i \geq 0} A_i\} = \bigcup_{i \geq 0} X^{-1}(A_i) \in \mathcal{F}$$

donc $\bigcup_{i \geq 0} A_i \in \mathcal{A}$.

Ainsi, $\mathcal{A}$ est une tribu qui contient $\mathcal{C}$, donc $\mathcal{A} = \mathcal{E} = \sigma(\mathcal{C})$. □

**Rappel :** Comme la mesurabilité est stable par composition, la somme, la différence, le quotient de variables aléatoires réelles est une variable aléatoire réelle. Si l'on considère une suite de variables aléatoires, alors l'inf, le sup, la liminf, limsup sont des variables aléatoires.

## 2.2 Loi d'une variable aléatoire

**Définition 2.3**

*Soit $X : (\Omega, \mathcal{F}, \mathbb{P}) \to (E, \mathcal{E})$ une variable aléatoire. On appelle loi de $X$, notée $\mathbb{P}_X$, la mesure* *image de $\mathbb{P}$ par l'application $X$ : pour $A \in \mathcal{E}$,*

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

*Dans le cas d'un vecteur aléatoire $X = (X_1, \ldots, X_d)$, la loi $\mathbb{P}_{(X_1,\ldots,X_d)}$ est appelée loi jointe du vecteur $X$. Les lois des marginales $X_1, \ldots, X_d$ sont appelées lois marginales*

▷ Montrons que $\mathbb{P}_X$ est une mesure de probabilité.
- $\mathbb{P}_X(E) = \mathbb{P}(\Omega) = 1$.
- Soit $(A_i)$ une famille d'ensembles disjoints de $\mathcal{E}$.

$$\mathbb{P}_X\left(\bigcup_{i\geq 0} A_i\right) = \mathbb{P}\left(X \in \bigcup_{i\geq 0} A_i\right) = \mathbb{P}\left(\bigcup_{i\geq 0}\{X \in A_i\}\right)$$

De plus, si $i \neq j$ et $\omega \in \{X \in A_i\}$, $X(\omega) \in A_i$. Si $\omega \in \{X \in A_j\}$ alors $\omega \in A_i \cap A_j = \varnothing$ donc $\{X \in A_i\} \cap \{X \in A_j\} = \varnothing$. Ainsi,

$$\mathbb{P}_X\left(\bigcup_{i\geq 0} A_i\right) = \sum_{i\geq 0} \mathbb{P}(\{X \in A_i\}) = \sum_{i\geq 0} \mathbb{P}_X(A_i).$$

Ainsi, $\mathbb{P}_X$ est bien une mesure de probabilité. □

**Exemple :** On lance deux fois un dé et on regarde le minimum des chiffres obtenus.

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

$$\mathcal{F} = \mathfrak{P}(\Omega)$$

$$\mathbb{P}(\{\omega\}) = \frac{1}{36}$$

$$X : \begin{array}{rcl} (\Omega, \mathcal{F}, \mathbb{P}) & \to & (E, \mathfrak{P}(E)) \\ \omega = (\omega_1, \omega_2) & \mapsto & \min(\omega_1, \omega_2) \end{array}$$

Loi de $X$ :

| $i$ | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| $\mathbb{P}(X = i)$ | $\frac{11}{36}$ | $\frac{9}{36}$ | $\frac{7}{36}$ | $\frac{5}{36}$ | $\frac{3}{36}$ | $\frac{1}{36}$ |

$X^{-1}(\{6\}) = \{6, 6\}$
$X^{-1}(\{5\}) = \{(5,6), (5,5), (6,5)\}$

**Définition 2.4**

*On dit que deux variables aléatoires $X$ et $Y$ ont la même loi si $\mathbb{P}_X = \mathbb{P}_Y$. On note $X \overset{\mathcal{L}}{=} Y$.*

⚠ Le cas $X : (\Omega_1, \mathcal{F}_1, \mathbb{P}_2) \to (E, \mathcal{E})$ et $Y : (\Omega_2, \mathcal{F}_2, \mathbb{P}_2) \to (E, \mathcal{E})$ et $\mathbb{P}_X = \mathbb{P}_Y$ est possible !

**Propriété 2.5**

*Si $(X, Y) \in E^2$ est une couple de variables aléatoires de loi $\mathbb{P}_{(X,Y)}$ alors les lois marginales $\mathbb{P}_X$*

*et $\mathbb{P}_Y$ de $X$ et $Y$ sont données par*

$$\forall A \in \mathcal{E} \qquad \mathbb{P}_X(A) = \mathbb{P}_{(X,Y)}(A \times E) \qquad \mathbb{P}_Y(A) = \mathbb{P}_{(X,Y)}(E \times A)$$

*Plus généralement, si on a un $n$-uplet $(X_1, \ldots, X_n)$ de variables aléatoires*

$$\mathbb{P}_{X_i}(A) = \mathbb{P}_{(X_1,\ldots,X_n)}(E^{i-1} \times A \times E^{n-i})$$

▷ $\{X \in A\} = \{(X, Y) \in A \times E\}$ □

**Définition 2.6**

*Soit $X$ une variable aléatoire. On appelle atome de $X$ (ou de sa loi) tout point $x \in E$ tel que $\mathbb{P}(X = x) \neq 0$.*

**Définition 2.7**

*On appelle support d'une variable aléatoire $X : \Omega \to \mathbb{R}$ le plus petit ensemble fermé $S(X)$ de $\mathbb{R}$ tel que $\mathbb{P}_X(S(X)) = 1$.*

On considère maintenant des variables aléatoires réelles : $E = \mathbb{R}$.

## 2.3 Fonction de répartition

**Rappel :** Si $\mathbb{P}_1$ et $\mathbb{P}_2$ coïncident sur $\mathcal{C}$ et $\mathcal{C}$ est stable par intersection finie, alors $\mathbb{P}_1 = \mathbb{P}_2$ sur $\sigma(\mathcal{C})$.

**Conséquence :** Soient $X, Y$ sont deux variables aléatoires à valeurs dans $(E, \mathcal{E})$ avec $\mathcal{E} = \sigma(\mathcal{C})$ et $\mathcal{C}$ stable par intersection finie et $\mu$ est une probabilité sur $(E, \mathcal{E})$.

$(i)$ $X$ suit la loi $\mu$ si et seulement si $\forall A \in \mathcal{C}\ \mathbb{P}_X(A) = \mu(A)$

$(ii)$ $X \overset{\mathcal{L}}{=} Y$ si et seulement si $\forall\ A \in \mathcal{C},\ \mathbb{P}(X \in A) = \mathbb{P}(Y \in A)$.

**Définition 2.8**

*On appelle fonction de répartition d'une variable aléatoire réelle $X : (\Omega, \mathcal{F}, \mathbb{P}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ la fonction*

$$F_X : \begin{array}{ccc} \mathbb{R} & \to & [0,1] \\ x & \mapsto & \mathbb{P}(X \leq x) \end{array}$$

**Proposition 2.9**

*La fonction de répartition caractérise la loi d'une variable aléatoire, c'est-à-dire, si $X$ et $Y$ sont deux variables aléatoires telles que $F_X = F_Y$ alors $X \overset{\mathcal{L}}{=} Y$.*

▷ $\mathcal{C} = \{]-\infty, x],\ x \in \mathbb{R}\}$. $\mathcal{B}(\mathbb{R}) = \sigma(\mathcal{C})$ et $\mathcal{C}$ est stable par intersection finie. □

**Exemple :** On reprend l'exemple précédent du dé jeté deux fois. On considère le minimum des chiffres obtenus.

On a :

$$\forall\ x < 1,\ F_X(x) = \mathbb{P}(X \leq x) = 0$$

$$\forall\ x \in [1, 2[,\ F_X(x) = \mathbb{P}(X = 1)$$

$$\forall x \in [2, 3[,\ F_X(x) = \mathbb{P}(X = 1 \text{ ou } X = 2) = \mathbb{P}(X = 1) + \mathbb{P}(X = 2)$$

etc.

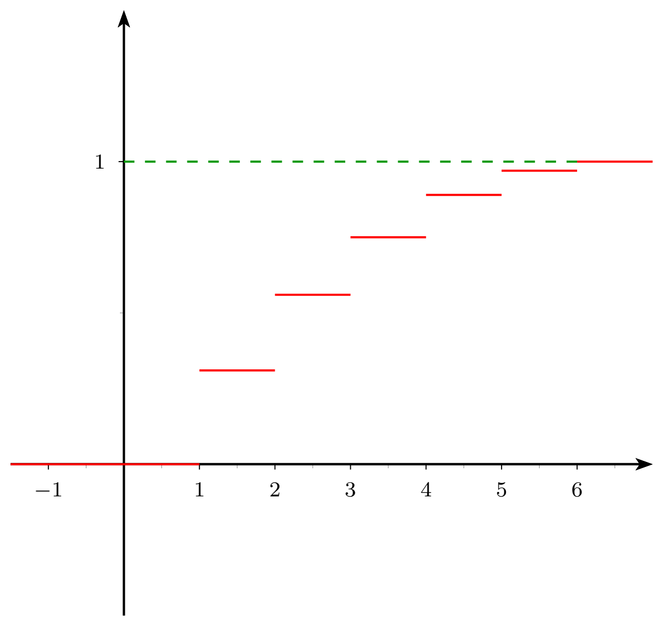


FIGURE 2.1 – Fonction de répartition de la variable aléatoire $X$

**Proposition 2.10**

*(i) $F_X$ est une fonction croissante.*

*(ii) $\lim\limits_{x\to-\infty} F_X(x) = 0$ et $\lim\limits_{x\to+\infty} F_X(x) = 1$.*

*(iii) $F_X$ est continue à droite et admet une limite à gauche en tout point de $\mathbb{R}$.*

$$F_X(x^-) = \lim_{\substack{y\to x\\ y<x}} F_X(y) = \mathbb{P}(X < x).$$

*(iv) $\mathbb{P}(X \in [a, b]) = F_X(b) - F_X(a^-)$.*

*(v) Si $x$ n'est pas un atome de $X$ alors $F_X$ est continue en $x$. L'ensemble des points de discontinuité de $F_X$ est l'ensemble des atomes de $X$.*

▷ $(i)$ Soient $x, y \in \mathbb{R},\ x \leq y$. Alors

$$F_X(x) = \mathbb{P}_X(]-\infty, x]) \leq \mathbb{P}_X(]-\infty, y]) = F_X(y)$$

car la mesure de probabilité est croissante.

$(ii)$ On a $\lim\limits_{x\to-\infty} F_X(x) = \lim\limits_{n\to+\infty} F_X(-n) = \lim\limits_{n\to+\infty} \mathbb{P}_X(]-\infty, -n])$. On pose $A_n = ]-\infty, -n]$, alors $(A_n)_{n\geq 0}$ est décroissante. Par décroissance séquentielle,

$$\lim_{x\to-\infty} F_X(x) = \lim_{n\to+\infty} \mathbb{P}_X(A_n) = \mathbb{P}_X\left(\bigcap_{n\geq 0} A_n\right) = 0.$$

En $+\infty$, $\lim_{x\to+\infty} F_X(x) = \lim_{n\to+\infty} F_X(n)$ avec $F_X(n) = \mathbb{P}_X(]-\infty, n])$. On pose $B_n = ]-\infty, n]$, alors $(B_n)_{n\geq 0}$ est une suite croissante donc

$$\lim_{x\to+\infty} F_X(x) = \lim_{n\to+\infty} \mathbb{P}_X(B_n) = \mathbb{P}_X\left(\bigcup_{n\geq 0} B_n\right) = \mathbb{P}_X(\mathbb{R}) = 1.$$

$(iii)$ Soit $x \in \mathbb{R}$.

$$\lim_{\substack{y\to x\\ y>x}} F_X(y) = \lim_{n\to+\infty} F_X\left(x+\frac{1}{n}\right).$$

Soit $A_n = \{X \leq x + \frac{1}{n}\}$. $(A_n)_{n\geq 0}$ est une suite décroissante et $\bigcap_{n\geq 0} A_n = \{X \leq x\}$ donc

$$\lim_{\substack{y\to x\\ y>x}} F_X(y) = \mathbb{P}(X \leq x) = F_X(x)$$

et $F_X$ est continue à droite.

Les limites existent car $F_X$ est croissante. Soit $x \in \mathbb{R}$.

$$\lim_{\substack{y\to x\\ y<x}} F_X(y) = \lim_{n\to+\infty} F_X\left(x-\frac{1}{n}\right).$$

Soit $B_n = \{X \leq x - \frac{1}{n}\}$. $(B_n)_{n\geq 0}$ est une suite croissante et $\bigcup_{n\geq 0} B_n = \{X < x\}$ donc

$$\lim_{\substack{y\to x\\ y<x}} F_X(y) = \mathbb{P}(X < x).$$

$(iv)$ On a

$$\mathbb{P}(X \in [a,b]) = \mathbb{P}(X \in ]-\infty, b] \backslash ]-\infty, a[) = \mathbb{P}(X \in ]-\infty, b]) - \mathbb{P}(]-\infty, a[) = F_X(b) - F_X(a^-).$$

$(v)$ $F_X$ est continue en $x$ si et seulement si $F_X(x^-) = F_X(x) = F_X(x^+)$ *ie.* $\mathbb{P}(X < x) = \mathbb{P}(X \leq x)$ avec

$$\mathbb{P}(X \leq x) = \mathbb{P}(X < x \text{ ou } X = x) = \mathbb{P}(X < x) + \mathbb{P}(X = x).$$

Donc $F_X$ est continue sur $x$ si et seulement si $x$ n'est pas un atome de $X$. □

**Proposition 2.11**

*La fonction de répartition admet au plus un nombre dénombrable de points de discontinuité. Par conséquent, une variable aléatoire réelle a au plus un nombre dénombrable d'atomes.*

▷ Soit $D_n$ l'ensemble des points de discontinuité avec un saut d'amplitude supérieure ou égale à $\frac{1}{n}$ :

$$D_n = \left\{x \in \mathbb{R},\ F_X(x) - F_X(x^-) \geq \frac{1}{n}\right\}.$$

Comme $F_X(x) \in [0,1]$ pour tout $x \in \mathbb{R}$, alors $\#D_n \leq n$. De plus, l'ensemble des points de discontinuité de $F_X$ est $\bigcup_{n\geq 1} D_n$ donc il est au plus dénombrable. □

**Théorème 2.12**

*Une fonction $F : \mathbb{R} \to [0,1]$ croissante, continue à droite et admettant une limite à gauche en* *tout point de $\mathbb{R}$ telle que*

$$\lim_{x\to-\infty} F_X(x) = 0 \qquad et \qquad \lim_{x\to+\infty} F_X(x) = 1$$

*est la fonction de réparation d'une variable aléatoire $X$.*

$\triangleright$ On se place dans le cas particulier $(\Omega, \mathcal{F}, \mathbb{P}) = (]0,1[, \mathcal{B}(]0,1[), \lambda)$. On note $X(\omega) = \inf\{t \in \mathbb{R},\ F(t) \geq \omega\}$ (remarque : si $F$ est bijective, alors $X(\omega) = F^{-1}(\omega)$). $X$ est une variable aléatoire sur $(]0,1[, \mathcal{B}(]0,1[), \lambda)$. Montrons que $\{t \in \mathbb{R},\ F(t) \geq \omega\} = [X(\omega), +\infty[$.

$\subset$ : Soit $t \in \mathbb{R}$ tel que $F(t) \geq \omega$. Comme $X(\omega)$ est l'inf, $t \in [X(\omega), +\infty[$.

$\supset$ : Soit $t \in \mathbb{R}$ tel que $t \geq X(\omega)$. $F$ étant croissante, $F(t) \geq F(X(\omega)) \geq \omega$ d'où l'égalité.

De plus,

$$\begin{aligned} F_X(x) &= \mathbb{P}(X \leq x) = \mathbb{P}(\{\omega \in \Omega,\ X(\omega) \leq x\}) = \mathbb{P}(\{\omega \in \Omega,\ x \in [X(\omega), +\infty[\}) \\ &= \mathbb{P}(\{\omega \in \Omega,\ x \in \{t \in \mathbb{R},\ F(t) \geq \omega\}\}) = \mathbb{P}(\{\omega \in \Omega,\ F(x) \geq \omega\}) \\ &= \lambda(]0, F(x)] = F(x) \end{aligned}$$

$\square$

**Définition 2.13**

*Soit $X = (X_1, \ldots, X_d)$ un vecteur aléatoire de $\mathbb{R}^d$. On appelle fonction de répartition de $X$ la fonction définie sur $\mathbb{R}^d$ par*

$$F_X(x) = F_X(x_1, \ldots, x_d) = \mathbb{P}(X_1 \leq x_1,\ X_2 \leq x_2, \ldots, X_d \leq x_d).$$

*La fonction de répartition de $X_i$ est*

$$F_{X_i}(x_i) = \lim_{\substack{x_j \to +\infty \\ j \neq i}} F_X(x_1, \ldots, x_d)$$

**Remarque :** La loi d'un vecteur aléatoire détermine la loi des marginales. Mais, de la loi des marginales on ne peut pas déduire la loi des vecteurs aléatoires.

**Exemple :** Soit $(X, Y)$ un vecteur de loi

$$\mathbb{P}_{(X,Y)} = \frac{1}{6}\delta_{(0,0)} + \frac{1}{3}\delta_{(0,1)} + \frac{1}{12}\delta_{(1,0)} + \frac{5}{12}\delta_{(1,1)}$$

La loi de $X$ est

$$\mathbb{P}_X = \frac{1}{2}\delta_0 + \frac{1}{2}\delta_1$$

et celle de $Y$ est

$$\mathbb{P}_Y = \frac{1}{4}\delta_0 + \frac{3}{4}\delta_1.$$

Soit $(U, V)$ des vecteurs aléatoires de loi

$$\mathbb{P}_{(U,V)} = \frac{1}{4}\delta_{(0,0)} + \frac{1}{4}\delta_{(0,1)} + \frac{1}{2}\delta_{(1,1)}.$$

Alors $\mathbb{P}_U = \mathbb{P}_X$ et $\mathbb{P}_Y = \mathbb{P}_V$ mais $\mathbb{P}_{(X,Y)} \neq \mathbb{P}_{(U,V)}$.

## 2.4 Exemples de variables aléatoires

### 2.4.1 Variables discrètes

**Définition 2.14**

*Une variable aléatoire $X : (\Omega, \mathcal{F}, \mathbb{P}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ est dite discrète si elle prend un nombre au plus dénombrable de valeurs.*

Soit $A$ l'ensemble des atomes d'une variable aléatoire discrète $X$. $A$ est au plus dénombrable et $\mathbb{P}_X(A) = 1$.

Si $X$ est une variable aléatoire discrète alors sa loi est la somme pondérée des mesures de Dirac en ses atomes :

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

▷ $A = \bigcup_{x \in A} \{x\}$. Soit $B \in \mathcal{B}(\mathbb{R})$. Alors

$$\begin{aligned}
\mathbb{P}_X(B) &= \mathbb{P}(X \in B) = \mathbb{P}\left(X \in B \cap \textstyle\bigcup_{x \in A}\{x\}\right) \\
&= \mathbb{P}\left(X \in \bigcup_{x \in A} B \cap \{x\}\right) = \sum_{x \in A} \underbrace{\mathbb{P}(B \cap \{x\})}_{\mathbb{P}(X = x \text{ et } x \in B)} = \sum_{x \in A} \mathbb{P}(X = x) \\
&= \sum_{x \in A} \mathbb{P}(X = x)\mathbf{1}_{x \in B} = \sum_{x \in A} \mathbb{P}(X = x)\delta_x(B)
\end{aligned}$$

□

**Proposition 2.15**

*Soit $X$ une variable aléatoire discrète et $A$ l'ensemble de ses atomes $A = \{x_i,\ i \in I\}$ ordonné $(x_i < x_{i+1})$. Alors, la fonction de réparation $F_X$ de $X$ est croissante et constante sur chaque intervalle $[x_i, x_{i+1}[$. Le saut est égal à $p_i = \mathbb{P}(X = x_i)$.*

▷ Soient $s, t \in [x_i, x_{i+1}[$, $s < t$.

$$F_X(t) - F_X(s) = \mathbb{P}(X \in ]s, t]) = \sum_{\{k,\ s < x_k \leq t\}} p_i = 0.$$

□

**Exemples :** $(i)$ $X$ variable aléatoire constante : $\forall \omega \in \Omega,\ X(\omega) = c$. Alors $\mathbb{P}_X = \delta_c$ car $\mathbb{P}_X(A) = \mathbb{P}(X \in A) = \mathbb{P}(c \in A) = \mathbf{1}_{c \in A}$.

$(ii)$ Si $X$ est à valeurs dans un ensemble fini $\{x_1, \ldots, x_n\}$ alors $\mathbb{P}_X = \sum_{i=1}^{n} p_i \delta_{x_i}$ avec $p_i = \mathbb{P}(X = x_i) \in [0, 1]$.

$(iii)$ Si tous les atomes ont la même probabilité de se réaliser on parle de loi équirépartie.

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

exemple : dé non truqué : $\mathbb{P}_X = \sum_{i=1}^{6} \frac{1}{6}\delta_i$.

Loi équirépartie sur $\mathbb{N}$ ? $\sum_{n\in\mathbb{N}} p = 1$ : impossible.

$(iv)$ $A \in \mathcal{F}$, $X(\omega) = \mathbf{1}_A(\omega)$. $X$ est est à valeurs dans $\{0, 1\}$

$$\mathbb{P}(X = 0) = \mathbb{P}_X(\mathbf{1}_A = 0) = \mathbb{P}(A^c) = 1 - \mathbb{P}(A)$$

et $\mathbb{P}(X = 1) = \mathbb{P}(A)$. Alors, $\mathbb{P}_X = \mathbb{P}(A)\delta_1 + (1 - \mathbb{P}(A))\delta_0$ est appelée loi de Bernoulli.

$(v)$ Loi binomiale : on compte le nombre de succès lorsqu'on réalise $n$ expériences de manière indépendante. Par exemple : lancer d'une pièce $n$ fois, et on compte le nombre de face. $X$ est à valeurs dans $\{0, \ldots, n\}$. On note $p$ la probabilité que la pièce tombe sur face. Alors

$$\mathbb{P}(X = k) = \mathbb{P}(k \text{ fois face et } (n-k) \text{ fois pile}) = \binom{n}{k} p^k(1-p)^{n-k}$$

où $\binom{n}{k} = \frac{n!}{k!(n-k)!}$.

**Remarque :** $1 = \mathbb{P}_X(\{0, \ldots, n\}) = \sum_{k=0}^{n} \mathbb{P}(X = k) = \sum_{k=0}^{n} \binom{n}{k} p^k(1-p)^{n-k} = (p+1-p)^n = 1$ ok.

Formule du binôme :

$$\sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k} = (a+b)^n.$$

Triangle de Pascal :

$$\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}.$$

$(v)$ Loi géométrique : loi du premier succès. Par exemple, $X$ est le nombre de lancers nécessaires pour avoir un "pile". $X$ est à valeurs dans $\mathbb{N}^*$. En notant $p \in ]0, 1[$ la probabilité d'un succès, on a :

$$\forall k \in \mathbb{N}^*, \qquad \mathbb{P}(X = k) = \mathbb{P}((k-1) \text{ faces puis pile}) = (1-p)^{k-1}p.$$

Vérification : $\mathbb{P}(X = k) \in [0, 1]$ et $\sum_{k=1}^{+\infty} \mathbb{P}(X = k) = p \sum_{k=1}^{+\infty} (1-p)^{k-1} = \frac{p}{1-(1-p)} = 1$.

Ici,

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

$(vi)$ Loi de Poisson sur $\mathbb{N}$.

$$\forall k \in \mathbb{N}, \qquad \mathbb{P}(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}$$

où $\lambda$ est un paramètre strictement positif.

**Remarques :** 1. Soit $X$ une variable aléatoire discrète à valeurs dans $\mathcal{X} = \{x_i,\ i \in I\}$ avec $\mathbb{P}(X = x_i) = p_i$. Soit $g : \mathbb{R} \to \mathbb{R}$ une application mesurable. Alors, $Y = g(X)$ est une variable aléatoire discrète à valeurs dans $\{y_j,\ j \in J\} = \{g(x_i),\ i \in I\} = g(\mathcal{X})$. De plus,

$$\mathbb{P}(Y = y_j) = \sum_{\substack{i \in I \\ g(x_i)=y_j}} p_i.$$

$\triangleright$ $Y$ est bien une variable aléatoire à valeurs dans $g(\mathcal{X})$. De plus,

$$\mathbb{P}(Y = y_j) = \mathbb{P}(g(X) = y_j) = \mathbb{P}(g \circ X = y_j) = \mathbb{P}(X \in g^{-1}(\{y_j\})) = \sum_{\substack{i \in I \\ g(x_i)=y_j}} i$$

$\square$

2. Soit $(X, Y)$ un vecteur aléatoire discret à valeurs dans $\{(x_i, y_j), (i, j) \in I \times J\}$. Alors, les lois des marginales sont

$$\mathbb{P}_X(x_i) = \sum_{j \in J} \mathbb{P}((X, Y) = (x_i, y_j)) \qquad \text{et} \qquad \mathbb{P}_Y(y_j) = \sum_{i \in I} \mathbb{P}((X, Y)) = (x_i, y_j).$$

Exemple : on lance deux dés. $X$ est le nombre de chiffres pairs. $Y$ est le maximum des deux chiffres. Quelle est la loi de $(X, Y)$ ?
$(X, Y)$ est à valeurs dans $\{0, 1, 2\} \times \{1, \ldots, 6\}$.

| $X \backslash Y$ | 1 | 2 | 3 | 4 | 5 | 6 | Loi de $X$ |
|---|---|---|---|---|---|---|---|
| 0 | $\frac{1}{36}$ | 0 | $\frac{3}{36}$ | 0 | $\frac{5}{36}$ | 0 | $\frac{1}{4}$ |
| 1 | 0 | $\frac{2}{36}$ | $\frac{2}{36}$ | $\frac{4}{36}$ | $\frac{4}{36}$ | $\frac{6}{36}$ | $\frac{1}{2}$ |
| 2 | 0 | $\frac{1}{36}$ | 0 | $\frac{3}{36}$ | 0 | $\frac{5}{36}$ | $\frac{1}{4}$ |
| Loi de $Y$ | $\frac{1}{36}$ | $\frac{1}{12}$ | $\frac{5}{36}$ | $\frac{7}{36}$ | $\frac{9}{36}$ | $\frac{11}{36}$ | (1) |

### 2.4.2 Théorème de Radon-Nikodym

**Définition 2.16**

*Soient $\mu$ et $\nu$ deux mesures sur $(\Omega, \mathcal{F})$. On dit que $\mu$ est absolument continue par rapport à $\nu$ ($\mu << \nu$) si*

$$\forall A \in \mathcal{F}, \qquad \nu(A) = 0 \quad \Longrightarrow \quad \mu(A) = 0.$$

**Théorème 2.17**

*Soient $\mu$ et $\nu$ deux mesures $\sigma$-finies. Si $\mu << \nu$ alors il existe $f : (\Omega, \mathcal{F}) \to \mathbb{R}$ mesurable telle que*

$$\forall A \in \mathcal{F}, \qquad \mu(A) = \int_A f \, d\nu$$

*$f$ est appelée densité de $\mu$ par rapport à $\nu$, notée $f = \dfrac{d\mu}{d\nu}$. De plus, si $\mu$ est une mesure finie, alors $f \in L^1(\nu)$.*

**Remarque :** (*i*) Dans ce cas, $\int g \mathrm{d}\mu = \int gf \mathrm{d}\nu$.

(*ii*) Si les mesures $\mu$ et $\nu$ sont positives alors $f$ est positive.

**Conséquence :** Si $\mu$ est une mesure de probabilité et $\nu$ une mesure positive, si $\mu << \nu$ alors $f$ est positive et

$$\int_\Omega f \mathrm{d}\nu = 1.$$

**Exemples :** – Prenons $\nu$ la mesure de comptage sur $\mathbb{N}$ :

$$\nu = \sum_{k \geq 0} \delta_k$$

et $\mu$ la loi de Bernoulli :

$$\mu = p\delta_1 + (1-p)\delta_0$$

Soit $A \in \mathbb{P}(\mathbb{N})$ tel que $\nu(A) = 0$. Alors $A = \mu(A)$ et donc $\mu(A)$. Ainsi, $\mu << \nu$.

$$f(k) = \begin{cases} 1-p & \text{si } k = 0 \\ p & \text{si } k = 1 \\ 0 & \text{si } k \geq 2. \end{cases}$$

– Si $\nu = \lambda$ la mesure de Lebesgue. Aucune loi discrète est absolument continue par rapport à $\lambda$ car $\lambda(\{x\}) = 0$ et $\mu(\{x\}) \neq 0$ pour $x$ bien choisi.

– Soit $X$ une variable aléatoire positive telle que sa loi vérifie :

$$\forall t, s \geq 0, \qquad \mathbb{P}(X > t + s) = \mathbb{P}(X > t)\mathbb{P}(X > s).$$

Alors,

$$\mathbb{P}(X > t) = e^{-\alpha t} \qquad \alpha > 0$$

$$\mathbb{P}(X \leq t) = \mathbb{P}(X \in [0, t]) = 1 - e^{-\alpha t} = \int_0^t \alpha e^{-\alpha s} \mathrm{d}s.$$

Par conséquent, la loi de $X$ est absolument continue par rapport à $\lambda$ de densité

$$f(t) = \alpha e^{-\alpha t} \mathbf{1}_{t \geq 0}$$

(Loi exponentielle)

### 2.4.3 Variables aléatoires à densité

**Définition 2.18**

*$X$ est une variable aléatoire à densité si $\mathbb{P}_X << \lambda$. On note $f$ sa densité.*

$$\mathbb{P}(X \in A) = \int_A f(x) dx$$

*$f$ est positive et*

$$\int_{\mathbb{R}} f(x) dx = 1.$$

*On note alors la loi de $X$ : $f(x)dx$.*

**Exemples :** – Loi exponentielle de densité

$$f(x) = \alpha e^{-\alpha x}\mathbf{1}_{x\geq 0} \qquad \alpha > 0.$$

– Loi uniforme sur $[a,b]$ de densité

$$f(x) = \frac{1}{b-a}\mathbf{1}_{[a,b]}(x)$$

– Loi normale (ou gaussienne) de densité

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

Rappel :

$$\left(\int_{\mathbb{R}} e^{-\frac{x^2}{2}}\mathrm{d}x\right)^2 = \int_{\mathbb{R}} e^{-\frac{x^2}{2}}\mathrm{d}x \int_{\mathbb{R}} e^{-\frac{y^2}{2}}\mathrm{d}y \underset{\text{Fubini-Tonneli}}{=} \int_{\mathbb{R}\times\mathbb{R}} e^{-\frac{x^2+y^2}{2}}\mathrm{d}x\mathrm{d}y$$

$$= \int_0^{+\infty}\int_0^{2\pi} e^{-\frac{\rho^2}{2}}\rho\mathrm{d}\rho\mathrm{d}\theta = 2\pi\left[-e^{-\frac{\rho^2}{2}}\right]_0^{+\infty} = 2\pi$$

Généralisation : loi gaussienne de moyenne $m$ et de variance $\sigma^2$ :

$$f(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{(x-m)^2}{2\sigma^2}}.$$

**Proposition 2.19**

*Soit $X$ une variable à densité $f$. Alors sa fonction de répartition $F_X$ vérifie :*

*(i) $\forall x \in \mathbb{R}, \qquad F_X(x) = \int_{-\infty}^{x} f(t)dt.$*

*(ii) $F_X$ est continue sur $\mathbb{R}$.*

*(iii) Si $f$ est continue en $x_0$, alors $F_X$ est dérivable en $x_0$ et $F_X'(x_0) = f(x_0)$.*

▷ *(ii)* $F_X$ est discontinue en les atomes de $X$. Soit $x$ un atome.

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

ce qui est absurde donc si $X$ est à densité alors $X$ n'a pas d'atomes et donc $F_X$ est continue sur $\mathbb{R}$.

*(iii)* $f$ est continue en $x_0$ donc

$$\forall\varepsilon > 0,\ \exists\delta > 0,\ \forall t \in ]x_0-\delta, x_0+\delta[, \quad |f(t) - f(x_0)| < \varepsilon.$$

Soit $\varepsilon > 0$ et $\delta$ comme ci-dessus. Soit $h \in \mathbb{R}$ tel que $0 < |h| < \delta$.

$$F_X(x_0+h) - F_X(x_0) = \int_{x_0}^{x_0+h} f(t)\mathrm{d}t$$

donc

$$|F_X(x_0+h) - F_X(x_0) - hf(x_0)| = \left|\int_{x_0}^{x_0+h}(f(t) - f(x_0))\mathrm{d}t\right| \leq \left|\int_{x_0}^{x_0+h}|f(t) - f(x_0)|\mathrm{d}t\right| \leq \varepsilon|h|.$$

Ainsi, $F_X$ est dérivable en $x_0$ et $F_X'(x_0) = f(x_0)$. □

⚠ $X$ variable aléatoire de fonction de répartition continue $\not\Rightarrow$ $X$ à densité.
Exemple de fonction croissante $[0,1] \to [0,1]$ continue : l'escalier de Cantor.
$f_0(x) = x$ pour $x \in [0,1]$.
$f_1(x)$ fonction affine par morceaux telle que $f_1\left([\frac{1}{3}, \frac{2}{3}]\right) = \left\{\frac{1}{2}\right\}$, $f_1(0) = 0$ et $f_1(1) = 1$.
Passage de $f_n$ à $f_{n+1}$) : sur chaque intervalle $[u,v]$ où $f_n$ n'est pas constante, on remplace $f_n$ par la fonction affine par morceaux qui vaut $\dfrac{f_n(u)+f_n(v)}{2}$ sur $\left[\dfrac{2u}{3}+\dfrac{v}{3}, \dfrac{2v}{3}+\dfrac{u}{3}\right]$.
On remarque que

$$f_{n+1}(x) - f_n(x)| \leq 2^{-n} \qquad x \in [0,1]$$

donc $\sum_{n\geq 0}(f_{n+1} - f_n)$ converge uniformément. Ainsi, la suite $(f_n)$ converge uniformément donc $f = \lim f_n$ est croissante continue telle que $f(0) = 0$, $f(1) = 1$ et $f$ est constante sur $K^c$ où $K$ est l'ensemble de Cantor (compact de $[0,1]$, d'intérieur vide, de mesure de Lebesgue nulle). $f$ est dérivable presque partout de dérivée nulle. Si $X$ était à densité, sa densité serait nulle ce qui est absurde.

**Remarque :** Si $X$ est à densité, alors $F_X$ est absolument continue.

**Proposition 2.20**

*Si $X$ est une variable aléatoire ayant une fonction de répartition $F_X$ de la forme*

$$F_X(x) = \int_{-\infty}^{x} f(t)\mathrm{d}t$$

*où $f$ est mesurable positive, alors $X$ est une variable aléatoire de densité $f$.*

▷ $f$ est mesurable positive donc

$$\int_{\mathbb{R}} f(t)\mathrm{d}t = \lim_{x\to+\infty} F_X(x) = 1$$

donc $f$ est une densité.

Soit $\mu$ la mesure définie par

$$\forall A \in \mathcal{B}(\mathbb{R}), \qquad \mu(A) = \int_A f(x)\mathrm{d}x.$$

On a

$$\forall x \in \mathbb{R}, \qquad \mu(]-\infty,x]) = \mathbb{P}_X(]-\infty,x]).$$

Comme l'ensemble $\{]-\infty,x],\ x \in \mathbb{R}\}$ est stable par intersection, $\mu = \mathbb{P}_X$ sur $\sigma(\{]-\infty,x],\ x \in \mathbb{R}\}) = \mathcal{B}(\mathbb{R})$. □

**Cas d'un vecteur aléatoire à densité**

**Définition 2.21**

*Soit $(X_1,\dots,X_d)$ un vecteur aléatoire sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. On dit que $(X_1,\dots,X_d)$ est à densité* *s'il existe une fonction $f : \mathbb{R}^d \to \mathbb{R}$ mesurable positive telle que*

$$\forall A \in \mathcal{B}(\mathbb{R}^d), \qquad \mathbb{P}((X_1, \ldots, X_d) \in A) = \int_A f(t)dt = \int_A f(t_1, \ldots, t_d)dt_1 \ldots dt_d$$

*et*

$$\int_{\mathbb{R}^d} f(t)dt = 1.$$

**Proposition 2.22**

*Si $(X, Y)$ est un vecteur aléatoire de densité $f_{(X,Y)}$ alors les marginales sont à densité :*

$$f_X(x) = \int_{\mathbb{R}} f_{(X,Y)}(x, y)dy \qquad f_Y(y) = \int_{\mathbb{R}} f_{(X,Y)}(x, y)dx.$$

$\triangleright$ $\mathbb{P}_X(A) = \mathbb{P}(X \in A, Y \in \mathbb{R}) = \displaystyle\int_{A\times\mathbb{R}} f_{(X,Y)}(x, y)\mathrm{d}x\mathrm{d}y \underset{\text{Fubini-Tonelli}}{=} \int_A \left(\int_{\mathbb{R}} f_{(X,Y)}(x, y)\mathrm{d}y\right) \mathrm{d}x.$ $\square$

**Proposition 2.23**

*Soit $X$ une variable aléatoire à densité $f$. On note*

$$S_f = \overline{\left\{x \in \mathbb{R},\ \forall V(x) \text{ voisinage de } x,\ \int_{V(x)} f(t)dt > 0\right\}}$$

*On a*

$$S(X) = S_f.$$

$\triangleright$ $- S(X) \subset S_f$ : $S_f^c$ est un ouvert. Pour chaque $x \in S_f^c$ il existe une voisinage $V(x)$ de $x$ tel que $\displaystyle\int_{V(x)} f(t)\mathrm{d}t = 0$. $S_f^c$ est un ouvert de $\mathbb{R}$ donc il peut s'écrire sous la forme

$$S_f^c = \bigcup_{i\in I} ]a_i, b_i[, \qquad a_i, b_i \in \overline{\mathbb{R}}$$

avec $I$ un ensemble dénombrable. Pour tout $i \in I$, $]a_i, b_i[\subset \bigcup_{x\in]a_i,b_i[\cap\mathbb{Q}} V(x)$. Alors,

$$0 \leq \mathbb{P}(X \in ]a_i, b_i[) =\leq \mathbb{P}(X \in [a_i, b_i]) = 0 \leq \mathbb{P}\left(X \in \bigcup_{x\in]a_i,b_i[\cap\mathbb{Q}} V(x)\right) \leq \sum_{x\in]a_i,b_i[\cap\mathbb{Q}} \mathbb{P}(X \in V(x)) = 0.$$

Par conséquent,

$$\mathbb{P}(X \in [a_i, b_i]) = 0$$

donc $[a_i, b_i]$ est négligeable pour la loi de $X$. $I$ étant au plus dénombrable, $\mathbb{P}(X \in S_f^c) = 0$ donc $\mathbb{P}(X \in S_f) = 1$. On en déduit que $S(X) \subset S_f$.

$- S_f \subset S(X)$ : $\mathbb{P}(X \in S(X)) = 1$ donc $\mathbb{P}(X \in S(X)^c) = 0$ et $S(X)^c$ est ouvert. SI $x \in S(X)^c$, avec $V(x) = S(X)^c$ $x \in S_f^c$. Donc $S_f \subset S(X)$. $\square$

**Proposition 2.24**

*Soit $X$ une variable aléatoire de densité $f_X$. Soit $g : \mathbb{R}^d \to \mathbb{R}^d$ un $\mathcal{C}^1$-difféomorphisme ($g \in \mathcal{C}^1$, bijective, $g^{-1} \in \mathcal{C}^1$). Alors, $Y = g(X)$ est une variable aléatoire à densité donnée par*

$$\forall y \in \mathbb{R}^d,\ f_Y(y) = f_X(g^{-1}(y))|Jac(g^{-1}(y))|$$

*où* $|Jac(g(x))| = \left|\det\left(\frac{\partial g_i}{\partial x_j}\right)\right|$.

▷ On applique le théorème de changement de variable :

$$\mathbb{P}(Y \in B) = \mathbb{P}(g(X) \in B) = \int_{g^{-1}(B)} f_X(x)\mathrm{d}x = \int_B f_X(g^{-1}(y))|\mathrm{Jac}(g^{-1}(y))|\mathrm{d}y.$$

□

**Exemples :** – Soit $(X, Y)$ un vecteur aléatoire de densité $f_{(X,Y)}(x, y) = c\mathbf{1}_{[0,1]\times[-1,2]}(x, y) = c\mathbf{1}_{[0,1]}(x)\mathbf{1}_{[-1,2]}(y)$. $f_{(X,Y)}$ est une densité si :

- elle est mesurable : ok ;
- elle est positive : $\Rightarrow c \geq 0$ ;
- $\int_{\mathbb{R}^2} f_{(X,Y)}(x, y)\mathrm{d}x\mathrm{d}y = 1$. Or

$$\int_{\mathbb{R}^2} f_{(X,Y)}(x, y)\mathrm{d}x\mathrm{d}y = 3c$$

donc $f_{(X,Y)}$ est une densité si et seulement si $c = \frac{1}{3}$.

$X$ est variable aléatoire de densité

$$f_X(x) = \int_{\mathbb{R}} f_{(X,Y)}(x, y)\mathrm{d}y = \frac{1}{3}\mathbf{1}_{[0,1]}(x)\int_{\mathbb{R}} \mathbf{1}_{[-1,2]}(y)\mathrm{d}y = \mathbf{1}_{[0,1]}$$

donc $X \sim \mathcal{U}([0, 1])$.

De même $Y$ est une variable aléatoire de densité

$$f_Y(y) = \frac{1}{3}\mathbf{1}_{[-1,2]}(y)$$

donc $Y \sim \mathcal{U}([-1, 2])$.

– Soit $(X, Y)$ un couple de densité

$$\forall (x, y) \in \mathbb{R}^2, \qquad f_{(X,Y)}(x, y) = \frac{1}{3\pi}e^{-\frac{x^2+2xy+5y^2}{6}}.$$

Loi de $X$ :

$$f_X(x) = \frac{1}{3\pi}\int_{\mathbb{R}} e^{-\frac{x^2+2xy+5y^2}{6}}\mathrm{d}y = \frac{1}{3}e^{-\frac{x^2}{6}}\int_{-\infty}^{+\infty} e^{-\frac{2xy+5y^2}{6}}\mathrm{d}y.$$

Or $2xy + 5y^2 = 5(y + \frac{x}{5})^2 - \frac{x^2}{5}$ donc

$$f_X(x) = \frac{1}{3\pi}e^{-\frac{x^2}{6}}e^{\frac{x^2}{30}}\int_{-\infty}^{+\infty} e^{-\frac{5}{6}(y+\frac{x}{5})^2}\mathrm{d}y = \frac{1}{3\pi}e^{-\frac{2x^2}{15}}\sqrt{\frac{3}{5}}\underbrace{\int_{-\infty}^{+\infty} e^{-\frac{z^2}{2}}\mathrm{d}z}_{\sqrt{2\pi}}$$

en effectuant le changement de variables $z = \sqrt{\frac{5}{3}}(y + \frac{x}{5})$. Alors,

$$f_X(x) = \sqrt{\frac{2}{15\pi}}e^{-\frac{2x^2}{15}}$$

donc $X \sim \mathcal{N}(0, \frac{15}{4}) = \mathcal{N}(m, \sigma^2)$ : loi normale.
Exercice : montrer que la loi de $Y$ est

$$f_Y(y) = \sqrt{\frac{2}{3\pi}} e^{\frac{-2}{3}y^2}.$$

# Chapitre 3

# Espérance d'une variable aléatoire

**Théorème de transfert**

Soient $(E, \mathcal{E})$, et $(F, \mathcal{F})$ deux espaces mesurables, $\varphi : (E, \mathcal{E}) \to (F, \mathcal{F})$ une fonction mesurable et $\mu$ une mesure sur $(E, \mathcal{E})$. La mesure image $\nu = \mu_\varphi$ définie par

$$\forall B \in \mathcal{F}, \qquad \nu(B) = \mu(\varphi^{-1}(B))$$

est une mesure sur $(F, \mathcal{F})$.

**Théorème 3.1**

*Soit $h : (F, \mathcal{F}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ une fonction mesurable. Alors $h$ est $\nu$-intégrale si et seulement si $h \circ \varphi$ est $\mu$-intégrable et alors*

$$\int_F h d\nu = \int_E h \circ \varphi d\mu.$$

## 3.1 Définition et propriétés de l'espérance d'une variable aléatoire

Soient $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité et $X : (\Omega, \mathcal{F}) \to \mathbb{R}^d$ une variable aléatoire soit positive (si $d = 1$), soit intégrable *ie.*

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

**Définition 3.2**

*On définit l'espérance d'une variable aléatoire réelle positive ou intégrable comme*

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

*Si $X$ est à valeurs dans $\mathbb{R}^d$, $X = (X_1, \ldots, X_d)$,*

$$\mathbb{E}[X] = (\mathbb{E}[X_1], \ldots, \mathbb{E}[X_d]).$$

$\mathbb{E}[X]$ est la valeur moyenne de la variable aléatoire $X$.

**Remarques :** – $\mathbb{E}[X] \in \mathbb{R}$ et n'est pas aléatoire !

– Comme les probabilités sont de masse finie, les constantes sont intégrables. Si $X = c$, $\mathbb{E}[X] = c$. En particulier, si $X$ est une variable aléatoire bornée, $X$ est intégrable.

– Si $X = \mathbf{1}_A$ alors $\mathbb{E}[X] = \mathbb{P}(A)$.

– Si $X$ est une fonction étagée positive $X = \sum_{i \in I} a_i \mathbf{1}_{A_i}$ où $I$ est fini, alors

$$\mathbb{E}[X] = \sum_{i \in I} a_i \mathbb{P}(A_i).$$

– Si $X$ est une fonction étagée quelconque, $X = X^+ - X^-$ où $X^+ = \max(X, 0)$, $X^- = \max(-X, 0)$ sont étagées positives et

$$\mathbb{E}[X] = \mathbb{E}[X^+] - \mathbb{E}[X^-].$$

– Si $X$ est une variable aléatoire positive alors $X$ est limite croissante de variables aléatoires $X_n$ étagées positives. Alors,

$$\mathbb{E}[X] = \lim_{n \to +\infty} \mathbb{E}[X_n] = \sup\{\mathbb{E}[Y],\ Y \text{ variable aléatoire étagée telle que } Y \leq X\}.$$

– Si $X$ est une variable aléatoire intégrable, *ie.* $\mathbb{E}[|X|] < \infty$, $X = X^+ - X^-$ où $X^+$ et $X^-$ sont des variables aléatoires positives. Alors,

$$\mathbb{E}[X] = \mathbb{E}[X^+] - \mathbb{E}[X^-].$$

– Si $X$ est complexe avec $\mathbb{E}[|X|] < \infty$ alors

$$\mathbb{E}[X] = \mathbb{E}[\Re(X)] + i\mathbb{E}[\Im(X)].$$

L'espérance étant définie comme une intégrale, on a les mêmes propriétés que pour l'intégrale.

**Théorème 3.3 (*convergence monotone*)**

*Si $(X_n)$ est une suite croissante de variables aléatoires positives alors*

$$\sup \mathbb{E}[X_n] = \lim_{n \to +\infty} \mathbb{E}[X_n] = \mathbb{E}[\lim_{n \to +\infty} X_n] = \mathbb{E}[\sup X_n].$$

**Lemme 3.4 (*Fatou*)**

*Si $(X_n)$ est une suite de variables aléatoires positives alors*

$$\mathbb{E}[\liminf_{n \to +\infty}] \leq \liminf_{n \to +\infty} \mathbb{E}[X_n].$$

**Théorème 3.5 (*convergence dominée*)**

*Si $(X_n)$ est une suite de variables aléatoires qui converge presque sûrement ie.*

$$\exists N \in \mathcal{F},\ \mathbb{P}(N) = 0,\ \forall \omega \notin N,\ X_n(\omega) \xrightarrow[n \to +\infty]{} X(\omega)$$

*et s'il existe une variable aléatoire $Y$ positive telle que $|X_n| \leq Y$ presque sûrement et $\mathbb{E}[Y] < \infty$*

*alors*

$$\lim_{n\to+\infty} \mathbb{E}[X_n] = \mathbb{E}[X].$$

*On a en fait*

$$\lim_{n\to+\infty} \mathbb{E}[|X_n - X|] = 0.$$

**Définition 3.6**

*Soit $X$ une variable aléatoire intégrable. $X$ est dite centrée si $\mathbb{E}[X] = 0$.*

**Proposition 3.7**

*Soient $X, Y$ deux variables aléatoires intégrables.*
$(i)$ *Si $X \leq Y$ alors $\mathbb{E}[X] \leq \mathbb{E}[Y]$.*
$(ii)$ *Si $A, B \in \mathcal{F}$, $A \subset B$ et $X \geq 0$ alors $\mathbb{E}[X\mathbf{1}_A] \leq \mathbb{E}[X\mathbf{1}_B]$.*
$(iii)$ *Si $a, b \in \mathbb{R}$, $\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]$.*
$(iv)$ $|\mathbb{E}[X]| \leq \mathbb{E}[|X|]$
$(v)$ *Si $X$ est une variable aléatoire positive telle que $\mathbb{E}[X] = 0$ alors $X = 0$ presque sûrement.*

**Proposition 3.8**

*Soit $X : (\Omega, \mathcal{F}, \mathbb{P}) \to \mathbb{R}$ une variable aléatoire de loi $\mathbb{P}_X$. $X$ est $\mathbb{P}$-intégrable si et seulement si*

$$\int_{\mathbb{R}} |x|\mathbb{P}_X(dx) < +\infty$$

*et alors,*

$$\mathbb{E}[X] = \int_{\mathbb{R}} x\mathbb{P}_X(dx).$$

▷ On applique le théorème de transfert avec $\varphi = X$ et $h(x) = x$ :

$$\int_{\Omega} h \circ \varphi \mathrm{d}\mathbb{P} = \int_{\mathbb{R}} h\mathrm{d}\mathbb{P}_X.$$

□

**Inégalité de Markov :** Si $X$ est une variable aléatoire positive alors

$$\forall t > 0, \qquad \mathbb{P}(X \geq t) \leq \frac{\mathbb{E}[X]}{t}.$$

▷ Par définition, $\mathbb{P}(X \geq t) = \mathbb{E}[\mathbf{1}_{\{X\geq t\}}]$. Or $X \geq X\mathbf{1}_{\{X\geq t\}} \geq t\mathbf{1}_{\{X\geq t\}}$. Donc

$$\mathbb{E}[X] \geq \mathbb{E}[t\mathbf{1}_{\{X\geq t\}}] = t\mathbb{P}(X \geq t).$$

□

**Proposition 3.9**

$(i)$ *Si $X$ est une variable aléatoire positive alors*

$$\mathbb{E}[X] = \int_0^{+\infty} \mathbb{P}(X \geq t)dt = \int_0^{+\infty} (1 - F_X(t))dt.$$

*(ii) De plus,* $\mathbb{E}[X] < \infty$ *si et seulement si*

$$\sum_{n\geq 0} \mathbb{P}(X \geq n\varepsilon) < +\infty$$

*ou*

$$\sum_{n\geq 0} 2^n \mathbb{P}(X \geq \varepsilon 2^n) < \infty$$

*pou* $\varepsilon > 0$.

$\triangleright$ $(i)$ Par le théorème de Fubini-Tonelli,

$$\int_0^{+\infty} \mathbb{P}(X > t)\mathrm{d}t = \int_0^{+\infty} \mathbb{E}[\mathbf{1}_{\{X>t\}}]\mathrm{d}t = \mathbb{E}\left[\int_0^{+\infty} \mathbf{1}_{\{X>t\}}\mathrm{d}t\right] = \mathbb{E}\left[\int_0^X \mathrm{d}t\right] = \mathbb{E}[X].$$

$(ii)$ Pour $t \in [n, n+1]$, $\{X \geq n+1\} \subset \{X \geq t\} \subset \{X \geq n\}$ donc

$$\forall t \in [n, n+1], \qquad \mathbb{P}(X > n+1) \leq \mathbb{P}(X > t) \leq \mathbb{P}(X > n)$$

d'où

$$\mathbb{P}(X > n+1) \leq \int_n^{n+1} \mathbb{P}(X > t)\mathrm{d}t \leq \mathbb{P}(X > n)$$

et

$$\sum_{n\geq 0} \mathbb{P}(X > n+1) \leq \underbrace{\int_0^{+\infty} \mathbb{P}(X > t)\mathrm{d}t}_{\mathbb{E}[X]} \leq \sum_{n\geq 0} \mathbb{P}(X > n).$$

Ainsi, la propriété est vraie pour $\varepsilon = 1$ et on remplace $X$ par $\frac{1}{\varepsilon}X$ pour obtenir le résultat pour tout $\varepsilon$.

On obtient la seconde équivalence en utilisant

$$[1, +\infty[= \bigcup_{n\geq 0} [2^n, 2^{n+1}[$$

$$0, +\infty[= [0, 1[\cup \bigcup_{n\geq 0} [2^n, 2^{n+1}[$$

et en remarquant que

$$0 \leq \int_0^1 \mathbb{P}(X > t)\mathrm{d}t \leq 1.$$

□

**Proposition 3.10 (*Inégalité de Jensen*)**

*Soient* $\varphi : \mathbb{R} \to \mathbb{R}$ *une fonction convexe (donc mesurable) et* $X : \Omega \to \mathbb{R}$ *une variable aléatoire intégrable telle que* $\varphi \circ X$ *est intégrable. Alors*

$$\varphi(\mathbb{E}[X]) \leq \mathbb{E}[\varphi(X)].$$

$\triangleright$ Pour tout $x$, $\varphi(x) = \sup\{g(x),\ g \text{ affine } \leq \varphi\}$. Si $g$ est affine, comme $\mathbb{E}$ est une application linéaire,

$$g(\mathbb{E}[X]) = \mathbb{E}[g(X)] \leq \mathbb{E}[\sup_{g \text{ affine}\leq\varphi} g(X)] \leq \mathbb{E}[\varphi(X)]$$

En passant au sup, on obtient $\varphi(\mathbb{E}[X]) \leq \mathbb{E}[\varphi(X)]$. □

**Exemple :** Soit $c \in \mathbb{R}$. On pose $\varphi : x \mapsto x \vee c = \sup(x, c)$ est convexe donc

$$\mathbb{E}[X] \vee c \leq \mathbb{E}[x \vee c]$$

Avec $\varphi : x \mapsto x^2$,

$$\mathbb{E}[X]^2 \leq \mathbb{E}[X^2]$$

Avec $\varphi : x \mapsto e^x$,

$$e^{\alpha \mathbb{E}[X]} \leq \mathbb{E}[e^{\alpha X}].$$

## 3.2 Exemples de calcul d'espérance

### 3.2.1 Variables discrètes

Soit $X$ une variable aléatoire à valeurs dans $\mathcal{A} = \{x_i,\ i \in I\}$ où $I$ est au plus dénombrable. Sa loi est

$$\mathbb{P}_X = \sum_{i \in I} \underbrace{\mathbb{P}(X = x_i)}_{p_i} \delta_{x_i}.$$

Alors $X$ est intégrable si $\mathbb{E}[|X|] = \sum_{i \in I} |x_i| p_i < \infty$ (c'est le cas par exemple si $\#I < \infty$) et dans ce cas,

$$\mathbb{E}[X] = \sum_{i \in I} p_i$$

Si $h : \mathbb{R} \to \mathbb{R}$ est mesurable, $h \circ X$ est une variable aléatoire discrète intégrable si

$$\mathbb{E}[|h(X)|] = \sum_{i \in I} |h(x_i)| p_i < \infty$$

et dans ce cas

$$\mathbb{E}[h(X)] = \sum_{i \in I} h(x_i) p_i.$$

**Exemples :** – Si $X = c$ est constante, $\mathcal{E}[X] = c$.

– Si $X$ est à valeurs dans $\{0, 1\}$ et suit la loi $\mathcal{B}(p)$ :

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

alors

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

⚠Si $X$ est à valeurs dans $\{a, b\}$, $\mathbb{P}(X = a) = 1 - p$, $\mathbb{P}(X = b) = p$ et

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

– Si $X$ est équirépartie sur $\mathcal{A} = \{x_1, \dots, x_n\}$ deux à deux distincts.

$$\mathbb{E}[X] = \sum_{i=1}^{n} x_i \frac{1}{n}$$

est la moyenne arithmétique des $x_i$.

– Si $X \sim \mathcal{B}(n,p)$ $X \in [\![0,n]\!]$ (où $n$ est le nombre d'expériences et $p$ la probabilité de succès)

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

et

$$\begin{aligned}\mathbb{E}[X] &= \sum_{k=0}^{n} k \binom{n}{k} p^k (1-p)^{n-k} = \sum_{k=1}^{n} n \binom{n-1}{k-1} p^k (1-p)^{n-k} \\ &= np \sum_{k=1}^{n} \binom{n-1}{k-1} p^{k-1} (1-p)^{(n-1)-(k-1)} = np \sum_{i=0}^{n-1} \binom{n-1}{i} p^i (1-p)^{(n-1)-i} = np.\end{aligned}$$

– $X \sim \mathrm{Geom}(p)$ avec $X \in \mathbb{N}^*$.

$$\mathbb{P}(X=k) = p(1-p)^{k-1} \qquad k \geq 1$$

$$\mathbb{E}[X] = \sum_{k=1}^{+\infty} kp(1-p)^{k-1} = p \sum_{k=1}^{+\infty} k(1-p)^{k-1} = p \times \frac{1}{(1-1+p)^2} = \frac{1}{p}$$

– $X \sim \mathcal{P}(\lambda)$ $X \in \mathbb{N}$ et

$$\mathbb{P}(X=k) = \frac{\lambda^k}{k!} e^{-\lambda}$$

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

### 3.2.2 Variables aléatoires à densité

Soit $X : (\Omega, \mathcal{F}, \mathbb{P}) \to \mathbb{R}$ une variable aléatoire à densité. $X$ est intégrable si et seulement si $\int_{\mathbb{R}} |x| f(x) \mathrm{d}x < \infty$ et dans ce cas

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

**Remarque :** Si $X$ est bornée alors elle est intégrable car $\int f = 1$.

Si $h : \mathbb{R} \to \mathbb{R}$ est mesurable, si $\mathbb{E}[|h(X)|] = \int_{\mathbb{R}} |h(x)| f(x) \mathrm{d}x < \infty$ alors

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

**Exemples :** – Si $X \sim \mathcal{U}([a,b])$ alors $f : x \mapsto \frac{1}{b-1} \mathbf{1}_{[a,b]}(x)$, $X$ est bornée et

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

– Si $X \sim \mathrm{Exp}(\alpha)$ $\alpha \geq 0$ alors

$$f : x \mapsto \alpha e^{-\alpha x} \mathbf{1}_{x \geq 0}$$

$X$ est intégrable

$$\mathbb{E}[X] = \int_0^{+\infty} x\alpha e^{-\alpha x}\mathrm{d}x = \frac{1}{\alpha}$$

– $X$ suit la loi de Cauchy de paramètre $a > 0$ :

$$f : \begin{array}{ccc} \mathbb{R} & \to & \mathbb{R} \\ x & \mapsto & \frac{a}{\pi}\frac{1}{a^2+x^2} \end{array}$$

comme $x \mapsto \frac{|x|}{a^2+x^2}$ n'est pas intégrable, $X$ n'est pas intégrable donc $\mathbb{E}[X]$ n'existe pas.

– $X \sim \mathcal{N}(0,1)$

$$f : \begin{array}{ccc} \mathbb{R} & \to & \mathbb{R} \\ x & \mapsto & \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}} \end{array}$$

$x \mapsto |x|e^{-\frac{x^2}{2}}$ est intégrable sur $\mathbb{R}$ donc

$$\mathbb{E}[X] = \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} x e^{-\frac{x^2}{2}}\mathrm{d}x = 0$$

**Exercice :** $\forall p \geq,\ X^p$ est intégrable. Montrer que

$$\begin{cases} \mathbb{E}[X^{2k+1}] = 0 \\ \mathbb{E}[X^{2k}] = \frac{(2k)!}{2^k k!} \end{cases}$$

**Proposition 3.11**

*Soit $m \in \mathbb{R}$ et $\sigma^2 > 0$.*

$$X \sim \mathcal{N}(m,\sigma^2) \qquad \Longleftrightarrow \qquad Z = \frac{X-m}{\sigma} \sim \mathcal{N}(0,1)$$

**Conséquence :** $X = m + \sigma Z$ donc $\mathbb{E}[X] = m + \sigma\mathbb{E}[Z] = m$.

$\triangleright$ Si $X \sim \mathcal{N}(m,\sigma^2)$,

$$\mathbb{P}(Z \leq t) = \mathbb{P}\left(\frac{X-m}{\sigma} \leq t\right) = \mathbb{P}(X \leq m+\sigma t) = \int_{-\infty}^{m+\sigma t} \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x-m)^2}{2\sigma^2}}\mathrm{d}x = \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}\mathrm{d}x$$

donc $Z \sim \mathcal{N}(0,1)$. La réciproque se démontre en remontant le calcul □

### 3.2.3 Identification de la loi d'une variable aléatoire

**Théorème 3.12**

*Soit $X : (\Omega, \mathcal{F}, \mathbb{P}) \to (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ une variable aléatoire. Soit $\mu$ une mesure de probabilité sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. Les assertions suivantes sont équivalentes :*

*(i) La loi de $X$ est égale à $\mu$.*
*(ii)* $\forall A \in \mathcal{B}(\mathbb{R})$ $\mathbb{P}(X \in A) = \mu(A)$
*(iii)* $\forall (t_1, \ldots, t_d) \in \mathbb{R}^d,\ \mathbb{P}(X1 \leq t_1, \ldots, X_d \leq t_d) = \mu\left(\prod_{i=1}^{d} ]-\infty, t_i]\right)$
*(iv)* $\forall h$ *mesurable positive* $\mathbb{E}[h(X)] = \int_{\mathbb{R}^d} h d\mu$
*(v)* $\forall h \in \mathcal{C}_c^\infty(\mathbb{R}^d)$ $\mathbb{E}[h(X)] = \int_{\mathbb{R}^d} h d\mu.$

**Exemple :** $X \sim \mathcal{U}(]0,1[)$ et $Y = \ln(X)$. Soit $h$ une fonction mesurable positive

$$\mathbb{E}[h(Y)] = \mathbb{E}[[h(-\ln(X)) = \int_0^1 h(-\ln(x))\mathrm{d}x = \int_0^{+\infty} h(y)e^{-y}\mathrm{d}y = \int_0^{+\infty} h\mathbb{P}_Y(\mathrm{d}y)$$

donc

$$Y \sim \mathrm{Exp}(1)$$

$\triangleright$ $(v) \Rightarrow (ii)$ Soit $K$ un compact de $\mathbb{R}^d$ et $G$ un ouvert tel que $K \subset G$ et $\overline{G}$ compact.

$$\tilde{h} : x \mapsto \frac{d(x, G^c)}{d(x, K) + d(x, G^c)}$$

est continue à valeurs dans $[0,1]$ et à support compact. $\forall x \in K,\ h(x) = 1$ et $\mathbf{1}_K(x) \leq h(x) \leq \mathbf{1}_G(x)$.

On construit une suite décroissante $(G_n)$ d'ouverts tels que $\bigcap G_n = K$. On le fait en dimension 1 :

$$K_n = \frac{d(x, G_n^c)}{d(x, K) + d(x, G_n^c)} \xrightarrow[n \to +\infty]{} \mathbf{1}_K(x)$$

De plus $(h_n)_n$ est décroissante positive donc par le théorème de convergence dominée,

$$\mathbb{E}[h_n(X)] \xrightarrow[n \to +\infty]{} \mathbb{E}[\mathbf{1}_K(x)] = \mathbb{P}(X \in K)$$

Par hypothèse $\int h_n \mathrm{d}\mu \xrightarrow[n \to +\infty]{\mu} (K)$ onc pour tout compact $\mathbb{P}(X \in K) = \mu(K)$ puis $\mathbb{P}(X \in A) = \mu(A)$ pour tout $A \in \mathcal{B}(\mathbb{R})$. $\square$

## 3.3 Moments d'une variable aléatoire

**Définition 3.13**

*Soit* $X : (\Omega, \mathcal{F}, \mathbb{P}) \to E = \mathbb{R}, \mathbb{C}, \mathbb{R}^d$, *ou* $\mathbb{C}^d$. *On dit que* $X$ *a un moment d'ordre* $p > 0$ *fini si* $\mathbb{E}[|X|^p] < \infty$ *et alors le moment d'ordre* $p$ *de* $X$ *est* $\mathbb{E}[X^p]$.

**Définition 3.14**

*On définit l'espace*

$$\mathcal{L}^p = \{X : (\Omega, \mathcal{F}, \mathbb{P}) \to \mathbb{E} \text{ variable aléatoire telle que } \mathbb{E}[|X|^p|] < \infty\}$$

*pour* $p \in ]0, +\infty[$ *et*

$$\mathcal{L}^\infty = \{X : (\Omega, \mathcal{F}, \mathbb{P}) \to \mathbb{R} \text{ variable aléatoire telle qu'il existe } C > 0,\ \mathbb{P}(X > C) = 0.$$

On réutilise les espaces $L^p$ vus en intégration.

**Théorème 3.15 (*Inégalité de Hölder*)**

*Soient* $p, q \geq 1$ *tels que* $\frac{1}{p} + \frac{1}{q} = 1$. *Soient* $X \in L^p(\Omega, \mathcal{F}, \mathbb{P})$ *et* $Y \in L^q(\Omega, \mathcal{F}, \mathbb{P})$. *Alors*

$$XY \in L^1(\Omega, \mathcal{F}, \mathbb{P}) \qquad \text{et} \qquad \mathbb{E}[|XY|] \leq \mathbb{E}[|X|^p]^{\frac{1}{p}} \mathbb{E}[|X|^q]^{\frac{1}{q}}$$

**Corollaire 3.16 (*Inégalité de Cauchy-Schwarz*)**

$$X, Y \in L^2(\Omega, \mathcal{F}, \mathbb{P}) \quad \Longrightarrow \quad XY \in L^1(\Omega, \mathcal{F}, \mathbb{P}) \quad \text{et} \quad \mathbb{E}[|XY|] \leq \mathbb{E}[|X|^2]^{\frac{1}{2}} \mathbb{E}[|Y|^2]^{\frac{1}{2}}.$$

▷ On a

$$\forall a \in \mathbb{R}, \qquad \mathbb{E}[(aX + Y)^2] = a^2 \mathbb{E}[X^2] + 2a\mathbb{E}[XY] + \mathbb{E}[Y^2] = \mathbb{E}[(aX + Y)^2] \geq 0$$

donc

$$\delta = \mathbb{E}[XY]^2 - \mathbb{E}[X^2]\mathbb{E}[Y^2] \leq 0$$

d'où le résultat. □

Cas d'égalité : si on a égalité dans l'inégalité de Cauchy-Schwarz, $\delta' = 0$ et donc il existe $a \in \mathbb{R}$ tel que

$$\mathbb{E}[(aX + Y)^2] = 0$$

donc $aX + Y = 0$ ps. Ainsi, il y a égalité dans l'inégalité de Cauchy-Schwarz si et seulement s'il existe une relation linéaire entre $X$ et $Y$.

**Théorème 3.17 (*Inégalité de Minkowski*)**

*Soient* $X, Y$ *deux variables aléatoires dans* $L^p$, $1 \leq p \leq +\infty$ *alors*

$$\|X + Y\|_p \leq \|X\|_p + \|Y\|_p$$

**Théorème 3.18**

*Si* $1 \leq p \leq q \leq +\infty$, *alors*

$$L^\infty \subset L^q \subset L^p \subset L^1$$

*et* $p \mapsto \|.\|_p$ *est une fonction croissante.*

**Théorème 3.19 (*Riesz-Fisher*)**

*Soit* $p \geq 1$. $(L^p(\Omega, \mathcal{F}, \mathbb{P}), \|.\|_p)$ *est un espace vectoriel normé complet.*
$L^2(\Omega, \mathcal{F}, \mathbb{P})$ *muni du produit scalaire* $\langle X, Y \rangle = \mathbb{E}[XY]$ *est un espace de Hilbert.*

## 3.4 Variance et covariance

**Définition 3.20**

*Soit $X \in L^2(\Omega, \mathcal{F}, \mathbb{P})$. On définit la variance de $X$ comme*

$$\mathrm{Var}(X) = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right]$$

*et l'écart-type de $X$ comme*

$$\sigma_X = \sqrt{\mathrm{Var}(X)}.$$

**Exemple :** Soient $X$ et $Y$ de lois respectives

$$\mathbb{P}_X = \frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_1, \qquad \mathbb{P}_Y = \frac{1}{2}\delta_{-100} + \frac{1}{2}\delta_{100}.$$

Alors

$$\mathbb{E}[X] = 0 = \mathbb{E}[Y]$$

mais

$$\sigma(X) = 1 \qquad \text{et} \qquad \sigma(Y) = 100.$$

**Remarque :** L'espérance donne la valeur moyenne d'une variable aléatoire et l'écart-type donc la distance moyenne entre la variable aléatoire et son espérance. Plus l'écart-type est grand, plus la variable aléatoire est diffuse. Plus l'écart-type est proche de zéro, plus la variable aléatoire est concentrée autour de sa moyenne. C'est une mesure de dispersion.

**Proposition 3.21**

*Soit $X$ une variable aléatoire dans $L^2$.*

$(i)$ $\mathrm{Var}(X) \geq 0$.

$(ii)$ $\mathrm{Var}(X) = 0$ *si et seulement si* $X = \mathbb{E}[X]$ *presque sûrement.*

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

$(iv)$ $\forall a, b \in \mathbb{R}, \qquad \mathrm{Var}(aX + b) = a^2\,\mathrm{Var}(X)$.

$\triangleright$ $(iii)$ En développant le carré, on obtient :

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

$(iv)$ On a :

$$\mathrm{Var}(aX) = a^2\mathbb{E}[X^2] - a^2\mathbb{E}[X]^2 = a^2\,\mathrm{Var}(X)$$

et

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

$\square$

**Exemples :** – Si $X$ est une variable aléatoire discrète à valeurs dans $\mathcal{A} = \{x_i,\ i \in I\}$ avec $I$ au plus dénombrable. Alors, en notant $p_i = \mathbb{P}(X = x_i)$.

$$\mathrm{Var}(X) = \sum_{i\in I}(x_i - \mathbb{E}[X])^2 p_i = \sum_{i\in I} x_i^2 p_i - \left(\sum_{i\in I} x_i p_i\right)^2$$

- Si $\mathbb{P}_X = \delta_a$ alors $\mathrm{Var}(X) = 0$.
- Si $X \sim \mathcal{B}(p)$ avec $p \in ]0,1[$,

$$\mathbb{E}[X] = p \qquad \text{et} \qquad \mathbb{E}[X^2] = 0^2 \times (1-p) + 1^2 \times p = p$$

donc

$$\mathrm{Var}(X) = p - p^2 = p(1-p).$$

- Si $X \sim \mathrm{Geom}(p)$ alors $\forall k \geq 1, \qquad \mathbb{P}(X = k) = (1-p)^{k-1}p$ donc

$$\mathbb{E}[X] = \frac{1}{p} \qquad \text{et} \qquad \mathbb{E}[X^2] = \sum_{k=1}^{+\infty} k^2 (1-p)^{k-1} p$$

Or $k^2 = k(k-1) + k$ donc

$$\mathbb{E}[X^2] = p(1-p)\sum_{k=1}^{+\infty} k(k-1)(1-p)^{k-2} + \sum_{k=1}^{+\infty} k(1-p)^{k-1}p = p(1-p)\frac{2}{(1-(1-p))^3} + \frac{1}{p} = \frac{2}{p^2} - \frac{1}{p}.$$

Alors,

$$\mathrm{Var}(X) = \frac{2}{p^2} - \frac{1}{p} - \left(\frac{1}{p}\right)^2 = \frac{1}{p^2} - \frac{1}{p} = \frac{p-1}{p^2}.$$

- Montrer que si $X \sim \mathcal{B}(n,p)$ alors

$$\mathrm{Var}(X) = np(1-p).$$

Montrer que si $X \sim \mathcal{P}(\lambda)$ alors

$$\mathrm{Var}(X) = \lambda.$$

- Si $X$ est une variable aléatoire à densité $f_X$ dans $L^2$ alors

$$\mathrm{Var}(X) = \int_{\mathbb{R}} (x - \mathbb{E}[X])^2 f_X(\mathrm{d}x) = \int_{\mathbb{R}} x^2 f_X(\mathrm{d}x) - \left(\int_{\mathbb{R}} x f_X(\mathrm{d}x)\right)^2.$$

- Si $X \sim \mathcal{U}([a,b])$,

$$\mathbb{E}[X] = \frac{a+b}{2} \qquad \text{et} \qquad \mathbb{E}[X^2] = \frac{1}{b-a}\int_a^b x^2 \mathrm{d}x = \frac{b^3 - a^3}{3(b-a)} = \frac{b^2 + ab + a^2}{3}$$

donc

$$\mathrm{Var}(X) = \frac{4(b^2 + ab + a^2) - 3(a+b)^2}{12} = \frac{(b-a)^2}{12}.$$

- Si $X \sim \mathrm{Exp}(\alpha)$ alors $f_X(x) = \alpha e^{-\alpha x}\mathbf{1}_{x>0}$ donc

$$\mathbb{E}[X] = \frac{1}{\alpha}$$

et

$$\mathbb{E}[X^2] = \int_0^{+\infty} \alpha x^2 e^{-\alpha x} \mathrm{d}x = \left[\left(-x^2 - \frac{2x}{\alpha} - \frac{2}{\alpha^2}\right) e^{-\alpha x}\right]_0^{+\infty} = \frac{2}{\alpha^2}$$

donc

$$\mathrm{Var}(X) = \frac{2}{\alpha^2} - \frac{1}{\alpha^2} = \frac{1}{\alpha^2}.$$

– Si $X \sim \mathcal{N}(0,1)$ alors

$$\mathbb{E}[X] = 0$$

donc

$$\mathrm{Var}(X) = \mathbb{E}[X^2] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x^2 e^{-\frac{x^2}{2}} \mathrm{d}x$$

Par intégration par parties ($u(x) = x$ et $v(x) = xe^{-\frac{x^2}{2}}$), on obtient

$$\mathrm{Var}(X) = \frac{1}{\sqrt{2\pi}} \left[-xe^{-\frac{x^2}{2}}\right]_{-\infty}^{+\infty} + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{x^2}{2}} \mathrm{d}x = 1.$$

Si $X \sim \mathcal{N}(m, \sigma^2)$ on peut écrire $X = m + \sigma Z$ avec $Z \sim \mathcal{N}(0,1)$ ($Z = \frac{X-m}{\sigma}$). Alors,

$$\mathrm{Var}(X) = \mathrm{Var}(m + \sigma Z) = \mathrm{Var}(\sigma Z) = \sigma^2 \,\mathrm{Var}(Z) = \sigma^2.$$

**Théorème 3.22 (*Inégalité de Bienaymé-Tchebychev*)**

*Soit $X$ une variable aléatoire dans $L^2(\Omega, \mathcal{F}, \mathbb{P})$. Alors,*

$$\forall a > 0, \qquad \mathbb{P}(|X - \mathbb{E}[X]| > a) \leq \frac{\mathrm{Var}(X)}{a^2}.$$

▷ $\mathbb{P}(|x - \mathbb{E}[X]| > a) = \mathbb{P}((X - \mathbb{E}[X])^2 > a^2) \underset{\text{Markov}}{\leq} \dfrac{\mathbb{E}[(X - \mathbb{E}[X])^2]}{a^2} = \dfrac{\mathrm{Var}(X)}{a^2}.$ □

**Application :** Combien de fois faut-il jouer à Pile ou Face avec une pièce bien équilibrée pour que la proportion de piles soit dans l'intervalle $\left[\frac{1}{2} - 0,01, \frac{1}{2} + 0,01\right]$ avec probabilité supérieure à 0,96 ? Notons $X$ le nombre de piles sur $n$ lancers. Alors, $X \sim \mathcal{B}(n, \frac{1}{2})$. On cherche $n$ tel que $\mathbb{P}\left(\frac{X}{n} \in \left[\frac{1}{2} - 0,01, \frac{1}{2} + 0,01\right.\right.$ $0,96$. Mais

$$\mathbb{P}\left(\frac{X}{n} \in \left[\frac{1}{2} - 0,01, \frac{1}{2} + 0,01\right]\right) = \mathbb{P}\left(\left|X - \frac{n}{2}\right| \leq 0,01n\right) = 1 - \mathbb{P}\left(\left|X - \frac{n}{2}\right| > 0,01n\right) \geq 1 - \frac{\mathrm{Var}(X)}{(0,01n)^2}.$$

Or $\mathrm{Var}(X) = \frac{n}{4}$ donc on a le résultat si $n$ vérifie

$$1 - \frac{1}{4.10^{-4} n} \geq 0,96 \quad \Longleftrightarrow \quad n \geq \frac{1}{0,0004 \times 0,04} = 62500.$$

Si $X$ est une variable aléatoire dans $\mathbb{R}^d$, $d > 1$, que faire ? Si $d = 2$, on définit la quantité suivante.

**Définition 3.23**

*Soient $X, Y$ deux variables aléatoires réelles dans $L^2$. La covariance de $(X, Y)$ est définie par*

$$\mathrm{cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])].$$

**Proposition 3.24**

*(i)* $(X, Y) \mapsto \operatorname{cov}(X, Y)$ *est une application bilinéaire.*
*(ii)* $\operatorname{cov}(X, X) = \operatorname{Var}(X)$.
*(iii)* $\forall a \in \mathbb{R}, \qquad \operatorname{cov}(X, a) = 0$.
*(iv)* $\operatorname{cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$ *donc si $X$ ou $Y$ est centrée,* $\operatorname{cov}(X, Y) = \mathbb{E}[XY]$.
*(v)* $\operatorname{Var}(X + Y) = \operatorname{Var}(X) + \operatorname{Var}(Y) + 2\operatorname{cov}(X, Y)$.

▷ *(v)* $\operatorname{Var}(X+Y) = \mathbb{E}\left[(X+Y-\mathbb{E}[X+Y])^2\right] = \mathbb{E}[((X-\mathbb{E}[X]) + (Y-\mathbb{E}[Y]))^2]$ donc

$$\operatorname{Var}(X+Y) = \operatorname{Var}(X) + \operatorname{Var}(Y) + 2\operatorname{cov}(X, Y).$$

□

**Définition 3.25**

*Le coefficient de corrélation de $X, Y \in L^2$ (non constantes) est*

$$\rho(X, Y) = \frac{\operatorname{cov}(X, Y)}{\sigma(X)\sigma(Y)}.$$

**Remarque :** $\rho(X, Y) \in [-1, 1]$ d'après l'inégalité de Cauchy-Schwarz. Si $|\rho(X, Y)| = 1$ alors il y a égalité dans l'inégalité de Cauchy-Schwarz et donc il existe une relation linéaire entre $X - \mathbb{E}[X]$ et $Y - \mathbb{E}[Y]$ et donc une relation affine entre $X$ et $Y$ :

$$Y = aX + b \qquad \text{avec} \begin{cases} a = \dfrac{\operatorname{cov}(X, Y)}{\operatorname{Var}(X)} \\ b = \mathbb{E}[Y] - a\mathbb{E}[X] \end{cases}$$

**Définition 3.26**

*Soit $(X_1, \dots, X_d)$ un vecteur aléatoire dans $L^2$. On appelle matrice de covariance la matrice*

$$\Gamma = (\operatorname{cov}(X_i, X_j))_{1 \le i,j \le d}$$

*$\Gamma$ est une matrice symétrique et pour tout $1 \le i \le d$, $\Gamma_{i,i} = \operatorname{Var}(X_i)$. De plus, si $X = (X_1, \dots, X_d)^T$, on peut écrire*

$$\Gamma = \mathbb{E}[XX^T] - \mathbb{E}[X]\mathbb{E}[X]^T = \mathbb{E}[(X - \mathbb{E}[X])(X - \mathbb{E}[X])^T].$$

*Ainsi, $\Gamma$ est une matrice positive :*

$$\forall a \in \mathbb{R}^d, \qquad a^T\Gamma a = \mathbb{E}\left[a^T(X - \mathbb{E}[X])(a^T(X - \mathbb{E}[X]))^T\right] = \mathbb{E}\left[\left|a^T(X - \mathbb{E}[X])\right|^2\right] \ge 0.$$

# Chapitre 4

# Fonction caractéristique

## 4.1 Définition

**Définition 4.1**

*Soit $X : (\Omega, \mathcal{F}, \mathbb{P}) \to \mathbb{R}^d$ un vecteur aléatoire. On définit la fonction caractéristique de $X$ par*

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

**Remarque :** $\varphi_X$ est bien définie sur $\mathbb{R}^d$ car $|e^{i\langle t,X\rangle}| \leq 1$ donc intégrable.

**Théorème 4.2**

*La fonction caractéristique caractérise la loi d'une variable aléatoire : si $X$ et $Y$ sont deux variables aléatoires telles que $\varphi_X = \varphi_Y$ alors $\mathbb{P}_X = \mathbb{P}_Y$.*

$\triangleright$ On démontre le théorème en dimension 1. Soit $g$ une fonction continue à support compact. Exprimons $\mathbb{E}[g(X)]$ à l'aide de $\varphi_X$. On remarque que, par le théorème de convergence dominée,

$$g(x) = \lim_{\lambda\to 0^+} \frac{1}{2\pi}\int_{\mathbb{R}} g(x-t)\frac{2}{1+t^2}\mathrm{d}t.$$

Alors, par le théorème de convergence dominée,

$$\begin{aligned}
\mathbb{E}[g(X)] &= \mathbb{E}\left[\lim_{\lambda\to 0}\frac{1}{2\pi}\int_{\mathbb{R}} g(X-\lambda t)\frac{2}{1+t^2}\mathrm{d}t\right] = \lim_{\lambda\to 0}\mathbb{E}\left[\frac{1}{2\pi}\int_{\mathbb{R}} g(X-\lambda t)\frac{2}{1+t^2}\mathrm{d}t\right] \\
&= \lim_{\lambda\to 0}\int_{\mathbb{R}}\frac{1}{2\pi}\int_{\mathbb{R}} g(x-\lambda t)\frac{2}{1+t^2}\mathrm{d}t\mathbb{P}_X(\mathrm{d}x) \underset{s=x-\lambda t}{=} \lim_{\lambda\to 0}\int_{\mathbb{R}}\frac{1}{2\pi}\int_{\mathbb{R}} f(s)\frac{2}{1+\left(\frac{x-s}{\lambda}\right)^2}\frac{\mathrm{d}s}{\lambda}\mathbb{P}_X(\mathrm{d}x) \\
&\underset{\text{Fubini}}{=} \lim_{\lambda\to 0}\int_{\mathbb{R}} g(s)\left(\frac{1}{2\pi}\int_{\mathbb{R}}\frac{2\lambda}{\lambda^2+(x-s)^2}\mathbb{P}_X(\mathrm{d}x)\right)\mathrm{d}s
\end{aligned}$$

Or

$$\begin{aligned}
\frac{2\lambda}{\lambda^2+(x-s)^2} &= \frac{1}{\lambda+i(x-s)}+\frac{1}{\lambda-i(x-s)} = \int_0^{+\infty} e^{-(\lambda+i(x-s)u)}\mathrm{d}u + \int_0^{+\infty} e^{-(\lambda-i(x-s))u}\mathrm{d}u \\
&= \int_{-\infty}^{+\infty} e^{\lambda|u|-i(x-s)u}\mathrm{d}u
\end{aligned}$$

donc

$$\begin{aligned}
E(g(X)) &= \lim_{\lambda\to 0}\int_{\mathbb{R}} g(s)\frac{1}{2\pi}\left(\int_{\mathbb{R}}\left(\int_{\mathbb{R}} e^{\lambda|u|-i(x-s)u}\mathrm{d}u\right)\mathbb{P}_X(\mathrm{d}x)\right)\mathrm{d}s \\
&\underset{\text{Fubini}}{=} \lim_{\lambda\to 0}\int_{\mathbb{R}}\frac{g(s)}{2\pi}\left(\int_{\mathbb{R}} e^{-\lambda|u|}e^{isu}\left(\underbrace{\int_{\mathbb{R}} e^{-ixu}\mathbb{P}_X(\mathrm{d}x)}_{\varphi_X(-u)}\right)\mathrm{d}u\right)\mathrm{d}s = \mathbb{E}[g(Y)]
\end{aligned}$$

si $\varphi_X=\varphi_Y$ et ce pour tout fonction $g$ continue à support compact. On a finalement

$$\mathbb{E}[g(X)] = \int_{\mathbb{R}}\frac{g(s)}{2\pi}\int_{\mathbb{R}} e^{isu}\varphi_X(-u)\mathrm{d}u\mathrm{d}s.$$

On a pu appliquer les théorèmes de Fubini et de convergence dominée car

$$\left|\frac{g(s)}{2\pi}e^{-\lambda|u|}e^{-i(x-s)u}\right| \le \frac{1}{2\pi}|g(s)|e^{-|u|} \in L^1(\mathrm{d}s\otimes\mathrm{d}u\otimes\mathbb{P}_X(\mathrm{d}x))$$

□

Pour trouver la loi d'une variable aléatoire on peut utiliser :
- si $X$ est discrète : $\mathbb{P}(X=x_k)$,
- la fonction de répartition,
- $\mathbb{E}[h(X)]$ pour $h$ mesurable bornée,
- $\varphi_X$.

## 4.2 Propriétés

**Proposition 4.3**

*Soit $X$ un vecteur aléatoire de fonction caractéristique $\varphi_X$.*
*(i) $\forall t\in\mathbb{R}^d$, $\quad |\varphi_X(t)|\le 1$.*
*(ii) $\forall t\in\mathbb{R}^d$, $\quad \overline{\varphi_X(t)}=\varphi_X(-t)$.*
*(iii) $\varphi_X(0)=1$.*
*(iv) $t\mapsto\varphi_X(t)$ est uniformément continue sur $\mathbb{R}^d$.*

▷ Soient $t,h\in\mathbb{R}^d$,

$$\varphi_X(t+h)-\varphi_X(t)=\mathbb{E}\left[e^{i\langle t+h,X\rangle}-e^{i\langle t,X\rangle}\right]=\mathbb{E}\left[e^{i\langle t+\frac{h}{2},X\rangle}\left(\underbrace{e^{i\langle\frac{h}{2},X\rangle}-e^{-i\langle\frac{h}{2},X\rangle}}_{2i\sin\left(\langle\frac{h}{2},X\rangle\right)}\right)\right]$$

donc

$$|\varphi_X(t+h) - \varphi_X(t)| \leq \mathbb{E}\left[2\left|\sin\left\langle\frac{h}{2}, X\right\rangle\right|\right] \leq 2\mathbb{E}\left[\min\left(\left|\left\langle\frac{h}{2}, X\right\rangle\right|, 1\right)\right] \xrightarrow[\|h\|\to 0]{} 0$$

uniformément en $t$, par le théorème de convergence dominée. Donc $\varphi_X$ est uniformément continue. □

**Exemples :** – Si $X$ est discrète à valeurs dans $\mathcal{A} = \{x_k,\ k \in I\}$,

$$\varphi_X(t) = \sum_{k\in I} e^{itx_k}\mathbb{P}(X = x_k).$$

– Si $X \sim \mathcal{B}(p)$,

$$\varphi_X(t) = \mathbb{E}[e^{itX}] = pe^{it} + (1-p).$$

– Si $X \sim \mathcal{B}(n,p)$, on a pour $k \in [\![0, n]\!]$

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

donc

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

– Si $X \sim \mathrm{Geom}(p)$, pour $k \in [\![1, n]\!]$,

$$\mathbb{P}(X = k) = p(1-p)^{k-1}$$

donc

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

– Si $X$ est une variable à densité,

$$\varphi_X(t) = \mathbb{E}[e^{itX}] = \int_{\mathbb{R}} e^{itx} f_X(x)\mathrm{d}x.$$

– Si $X \sim \mathcal{U}([a, b])$,

$$\varphi_X(t) = \frac{1}{b-a}\int_a^b e^{itx}\mathrm{d}x = \frac{e^{itb} - e^{ita}}{(b-a)it}.$$

– Si $X \sim \mathrm{Exp}(\alpha)$, $\alpha > 0$,

$$f_X(x) = \alpha e^{-\alpha x}\mathbf{1}_{x>0}$$

donc

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

– Si $X \sim \mathcal{N}(0, 1)$,

$$\varphi_X(t) = \int_{-\infty}^{+\infty} e^{itx}\frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}\mathrm{d}x = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} e^{-\frac{(x-it)^2}{2}}e^{-\frac{t^2}{2}}\mathrm{d}x = e^{-\frac{t^2}{2}}$$

– Si $X \sim \mathcal{N}(m, \sigma^2)$, alors $X = m + \sigma Z$ où $Z \sim \mathcal{N}(0,1)$ donc

$$\varphi_X(t) = \mathbb{E}[e^{it(m+\sigma Z)}] = e^{itm}\varphi_Z(t\sigma) = e^{itm}e^{-\frac{t^2\sigma^2}{2}}.$$

– Si $X$ suit la loi de Cauchy de paramètre $a$ : $\mathbb{P}_X = \frac{a}{\pi(a^2+x^2)}\mathrm{d}x$

$$\varphi_X(t) = e^{-a|t|}.$$

**Proposition 4.4**

*Soit $X$ une variable aléatoire réelle de fonction caractéristique $\varphi_X$. Si $X$ admet un moment d'ordre $p \geq 1$ fini alors $\varphi_X$ est de classe $\mathcal{C}^p$ et*

$$\forall k \leq p, \qquad \varphi^{(k)}(0) = i^k \mathbb{E}[X^k].$$

$\triangleright$ $h : t \mapsto e^{itX}$ est de classe $\mathcal{C}^\infty$ et $h^{(k)}(t) = (iX)^k e^{itX}$ donc $|h^{(k)}(t)| \leq |X^k|$ et le théorème de dérivation s'applique. $\square$

**Application :** – Si $X \sim \mathrm{Geom}(p)$ alors

$$\varphi_X'(t) = \frac{ipe^{it}}{(1 - e^{it}(1-p))^2}$$

et $\varphi_X'(0) = \frac{i}{p}$ d'où $\mathbb{E}[X] = \frac{1}{p}$.

**Théorème 4.5 (*Inversion de Fourier*)**

*Soit $X$ une variable aléatoire de fonction caractéristique $\varphi_X$. Si $\varphi_X$ est intégrable par rapport à la mesure de Lebesgue sur $\mathbb{R}^d$ alors la loi $\mathbb{P}_X$ de $X$ admet une densité (par rapport à la mesure de Lebesgue).*

$\triangleright$ En dimension 1, soit $g \in \mathcal{C}_c^0(\mathbb{R})$. On a vu que

$$\begin{aligned}\mathbb{E}[g(X)] &= \frac{1}{2\pi}\lim_{\lambda\to 0}\int_{\mathbb{R}}\int_{\mathbb{R}} g(s)\varphi_X(t)e^{-ist}e^{-\lambda|t|}\mathrm{d}t\mathrm{d}s \\ &= \int_{\mathbb{R}} g(s)\left(\frac{1}{2\pi}\int_{\mathbb{R}}\varphi_X(t)e^{-ist}\mathrm{d}t\right)\mathrm{d}s\end{aligned}$$

par les théorèmes de Fubini et de convergence dominée qui s'appliquent car :

$$\left|g(s)\varphi_X(t)e^{-ist}e^{-\lambda|t|}\right| \leq |g(s)||\varphi_X(t)| \in L^1(\mathbb{R}^2, \mathrm{d}s \otimes \mathrm{d}t).$$

$\square$

## 4.3 Autres fonctionnelles caractérisant la loi

### 4.3.1 Fonction génératrice pour des variables aléatoires entières

Soit $X$ une variable aléatoires à valeurs dans $\mathbb{N}$. On note, pour $n \in \mathbb{N}$, $p_n = \mathbb{P}(X = n)$.

**Définition 4.6**

*La fonction génératrice de $X$ est définie par*

$$M_X(t) = \mathbb{E}[t^X] = \sum_{n=0}^{+\infty} t^n p_n.$$

**Propriétés 4.7**

– *$M_X$ est bien définie sur $[0,1]$.*
– *$M_X$ est continue sur $[0,1]$.*
– *$M_X$ étant une série entière, $M_X$ est $\mathcal{C}^\infty$ sur $[0,1[$.*
– *On a :*

$$\forall n \in \mathbb{N}, \qquad M_X((n))(0) = n! p_n$$

**Conséquence :** On retrouve facilement la loi de $X$ à partir de $M_X$.

**Exemples :** – Si $X \sim \mathcal{B}(p)$, $\mathbb{P}(X=0) = 1-p$, $\mathbb{P}(X=1) = p$.

$$M_X(t) = 1 - p + p^t.$$

– Si $X \sim \mathrm{Geom}(p)$, $\forall n \geq 1,\ \mathbb{P}(X=n) = (1-p)^{n-1}p$.

$$M_X(t) = \sum_{n=1}^{+\infty} t^n (1-p)^{n-1} p = tp \sum_{n=1}^{+\infty} [t(1-p)]^{n-1} = \frac{tp}{1-t(1-p)} \quad \text{si } t < \frac{1}{1-p}.$$

### 4.3.2 Transformée de Laplace

**Définition 4.8**

*Soit $X$ un vecteur aléatoire de $\mathbb{R}^d$. On définit la transformée de Laplace de $X$ comme :*

$$L_X(t) = \mathbb{E}[e^{tX}] \qquad \text{si } d = 1$$

$$L_X(t) = \mathbb{E}[e^{\langle t, X\rangle}] \qquad \text{avec } t \in \mathbb{R}^d.$$

De manière générale, on peut seulement dire que $L_X$ est définie en 0.

**Propriétés 4.9**

– *Si $X$ est positive, s'il existe $t \neq 0$ tel que $L_X(t) < +\infty$ alors $X$ admet des moments finis de tout ordres.*
– *S'il existe $t > 0$ tel que $L_{|X|}(t) < +\infty$ alors $X$ admet des moments finis de tout ordres.*
– *S'il existe un voisinage $V$ de 0 tel que $\forall t \in V,\ L_X(t) < +\infty$ alors $X$ admet des moments finis de tout ordres.*
*Dans tous les cas précédents,*

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

– *Si $L_X$ est définie sur un voisinage de 0, alors $L_X$ caractérise la loi.*

**Remarque :** Si $X$ est positive, $L_X$ est définie sur $\mathbb{R}_-$.

# Chapitre 5

# Notion d'indépendance en probabilités

On se place sur $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilités.

## 5.1 Définitions

**Définition 5.1 (*Indépendance entre des événements*)**

*Deux événements $A, B \in \mathcal{F}$ sont dits indépendants si*

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

*Une famille d'événements $(A_i)_{i \in I}$ est dite (mutuellement) indépendante si*

$$\forall p \geq 2,\ \forall i_1, \dots, i_p \in I, \qquad \mathbb{P}(A_{i_1} \cap \dots \cap A_{i_p}) = \mathbb{P}(A_{i_1})\mathbb{P}(A_{i_2}) \dots \mathbb{P}(A_{i_p}).$$

⚠Les notions d'événements indépendants et d'événements disjoints sont deux notions bien distinctes. Si $A$ et $B$ sont disjoints $A \cap B = \varnothing$ donc $A$ et $B$ ne sont pas indépendants sauf si l'un est de probabilité nulle.

⚠Mutuellement indépendants $\Longrightarrow$ indépendants deux à deux mais la réciproque est fausse.
exemple : un joueur joue à Pile ou face. Il gagne 1 euro si Pile et perd 1 euro si Face.

$$A = \{\text{le joueur gagne 1 euro au } 1^{\text{er}} \text{ lancer}\}$$

$$B = \{\text{le joueur gagne 1 euro au } 2^{\text{e}} \text{ lancer}\}$$

$$C = \{\text{le joueur gagne 2 euros ou perd 2 euros lors des deux premiers lancers}\}$$

On a :

$$\mathbb{P}(A) = \frac{1}{2}, \qquad \mathbb{P}(B) = \frac{1}{2}$$

et $A$ et $B$ sont indépendants. Comme $A \cap B$ et $A^c \cap B^c$ sont disjoints et $A, B$ sont indépendants,

$$\mathbb{P}(C) = \mathbb{P}((A \cap B) \cup (A^c \cap B^c)) = \mathbb{P}(A \cap B) + \mathbb{P}(A^c \cap B^c) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}.$$

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

donc $A$ et $C$ sont indépendants. De même, $B$ et $C$ sont indépendants. Mais

$$\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A \cap B) = \frac{1}{4} \neq \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C)$$

donc $A, B, C$ ne sont pas mutuellement indépendants.

**Remarque :** Si $A$ et $B$ son indépendants, alors $A^c$ et $B$, $A^c$ et $B^c$, $A$ et $B^c$ sont respectivement indépendants.

**Définition 5.2 (*Indépendance pour des tribus*)**

*Soit $\mathcal{F}_1$ et $\mathcal{F}_2$ deux familles (en particulier tribus) d'ensembles mesurables. $\mathcal{F}_1$ et $\mathcal{F}_2$ sont indépendantes si pour tout $A_1 \in \mathcal{F}_1$ et tout $A_2 \in \mathcal{F}_2$, $A_1$ et $A_2$ sont indépendantes, ie.*

$$\forall A_1 \in \mathcal{F}_1,\ \forall A_2 \in \mathcal{F}_2, \qquad \mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_1)\mathbb{P}(A_2).$$

*Une famille $(\mathcal{F}_i)_{i\in I}$ de sous-ensembles de $\mathcal{F}$ est dite indépendante si*

$$(A_i)_{i\in I} \text{ est mutuellement indépendante avec } \forall i \in I,\ A_i \in \mathcal{F}_i.$$

**Proposition 5.3**

*Si $\mathcal{A}_1$ et $\mathcal{A}_2$ sont deux sous-ensembles de $\mathcal{F}$ stables par intersection finie et indépendants, alors $\sigma(\mathcal{A}_1)$ et $\sigma(\mathcal{A}_2)$ sont indépendantes.*

$\triangleright$ Soit $A_1 \in \mathcal{A}_1$. Notons

$$\mathcal{M}_1 = \{A_2 \in \sigma(\mathcal{A}_2),\ A_1 \text{ et } A_2,\ \text{sont indépendants}\}$$

On a $\Omega \in \mathcal{M}_1$, $\mathcal{M}_1$ est stable par différence, $\mathcal{M}_1$ est stable par union croissante, $\mathcal{A}_2 \subset \mathcal{M}_1$. Donc par le théorème des classes monotones, comme $\mathcal{A}_2$ est stable par intersection finie, $\mathcal{M}_1 = \sigma(\mathcal{A}_2)$. Soit $A_2 \in \sigma(A_2)$. On pose

$$\mathcal{M}_2 = \{A_1 \in \sigma(\mathcal{A}_1),\ A_1 \text{ et } A_2 \text{ sont indépendants}\}.$$

On a montré que $\mathcal{A}_1 \subset \mathcal{M}_2$ donc, comme précédemment, par le théorème des classes monotones, $\mathcal{M}_2 = \sigma(\mathcal{A}_1)$, $\forall A_2 \in \sigma(\mathcal{A}_2)$. Donc $\sigma(\mathcal{A}_1)$ et $\sigma(\mathcal{A}_2)$ sont indépendantes. $\square$

**Proposition 5.4**

*Si $(\mathcal{A}_i)_{i\in I}$ est une famille de sous-ensembles de $\mathcal{F}$ indépendants et stables par intersections finie alors $(\sigma(\mathcal{A}_i))_{i\in I}$ est une famille de tribus indépendantes.*

**Proposition 5.5**

*Si $(\mathcal{F}_i)_{i\in I}$ sont des tribus indépendantes, si $I = \bigcup_{k\in K} I_k$ est une partition de $I$, alors, en notant, pour $k \in K$, $\mathcal{G}_k = \sigma\left(\bigcup_{i\in I_k} \mathcal{F}_i\right)$, $(\mathcal{G}_k)_{k\in K}$ est indépendante.*

**Rappel :** Soit $X$ une variable aléatoire réelle. La tribu engendrée par $X$ est

$$\sigma(X) = \{X^{-1}(A),\ A \in \mathcal{B}(\mathbb{R})\}.$$

$\sigma(X)$ est la plus petite tribu qui rend mesurable l'application $X : \Omega \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$. $\sigma(X)$ est la tribu qui contient toutes les informations liées à la variable $X$.

**Définition 5.6**

*Deux variables aléatoires $X$ et $Y$ sont dites indépendantes si $\sigma(X)$ et $\sigma(Y)$ sont indépendantes, ie.*

$$\forall A, B \in \mathcal{B}(\mathbb{R}), \qquad \mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B)$$

*Une famille de variables aléatoires $(X_i)_{i\in I}$ est (mutuellement) indépendante si les tribus associées sont indépendantes, ie.*

$$\forall p \geq 2,\ \forall i_1, \dots, i_p \in I, \forall A_{i_1}, \dots, A_{i_p} \in \mathcal{B}(\mathbb{R}), \qquad \mathbb{P}(X_{i_1} \in A_{i_1}, \dots, X_{i_p} \in A_{i_p}) = \prod_{k=1}^{p} \mathbb{P}(X_{i_k} \in A_{i_k}).$$

**Remarque :** $(X_1, \dots, X_d)$ est mutuellement indépendante si et seulement si $\forall A_1, \dots, A_d \in \mathcal{B}(\mathbb{R})$,

$$\mathbb{P}(X_1 \in A_1, \dots, X_d \in A_d) = \prod_{k=1}^{d} \mathbb{P}(X_k \in A_k).$$

## 5.2 Propriétés

**Proposition 5.7**

*Soit $X = (X_1, \dots, X_d)$ un vecteur aléatoire. $(X_1, \dots, X_d)$ est indépendante si et seulement si*

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

$\triangleright$ $\Rightarrow$ : Supposons l'indépendance. Soit $A = A_1 \times \cdots \times A_d \in \mathcal{B}(\mathbb{R})^d$.

$$\mathbb{P}_X(A) = \mathbb{P}(X \in A) = \mathbb{P}(X_1 \in A_1, \dots, X_d \in A_d) = \prod_{k=1}^{d} \mathbb{P}(X_k \in A_k) = \mathbb{P}_{X_1} \otimes \cdots \otimes \mathbb{P}_{X_d}(A).$$

Comme les pavés engendrent $\mathcal{B}(\mathbb{R}^d)$ et forment une famille stable par intersection finie, on a

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

$\Leftarrow$ : Idem. $\square$

**Remarque :** Si $X = (X_1, \dots, X_d)$ est mutuellement indépendant alors la connaissance des lois marginales suffit pour connaître la loi du vecteur.

**Corollaire 5.8**

*Les assertions suivantes sont équivalentes :*

- *$X_1, \dots, X_d$ $d$ variables aléatoires sont indépendantes.*
- *$\mathbb{P}_{(X_1,\dots,X_d)} = \mathbb{P}_{X_1} \otimes \cdots \otimes \mathbb{P}_{X_d}$.*
- *$\forall A_1, \dots, A_d \in \mathcal{B}(\mathbb{R}),\ \mathbb{P}(X_1 \in A_1, \dots, X_d \in A_d) = \prod_{i=1}^{d} \mathbb{P}(X_i \in A_i)$.*
- *$F_X(x) = \prod_{i=1}^{d} F_{X_i}(x_i)$ où $\forall x = (x_1, \dots, x_d) \in \mathbb{R}^d,\ F_X(x) = \mathbb{P}(X_1 \leq x_1, \dots, X_d \leq x_d)$.*
- *$\mathbb{E}[h_1(X_1) \dots h_d(X_d)] = \prod_{i=1}^{d} \mathbb{E}[h_i(X_i)]$ pour toute fonction $h_i$ mesurable bornée ou positive ou telle que $h_i(X_i) \in L^1$.*
- *$\forall t = (t_1, \dots, t_d) \in \mathbb{R}^d,\ \varphi_X(t) = \prod_{i=1}^d \varphi_{X_i}(t_i)$ où $\varphi_X(t) = \mathbb{E}[e^{i\langle t, X\rangle}]$.*
- *(cas discret) $\mathbb{P}(X_1 = x_1, \dots, X_d = x_d) = \prod_{i=1}^{d} \mathbb{P}(X_i = x_i)\ \forall x_1, \dots, x_d$.*
- *(cas discret à valeurs dans $\mathbb{N}^d$) $M_X(t) = \prod_{i=1}^{d} M_{X_i}(t_i)\ \forall t \in [0,1]^d$ où $M_X(t) = \mathbb{E}[t_1^{X_1} \dots t_d^{X_d}]$.*
- *$L_X(t) = \mathbb{E}[e^{\langle t, X\rangle}]$. Si $L_X$ est définie sur un voisinage $V$ de 0, $L_X(t) = \prod_{i=1}^{d} L_X(t_i),\ \forall t \in V$.*
- *(cas à densité) $f_{(X_1,\dots,X_d)}(x_1, \dots, x_d) = f_{X_1}(x_1) \dots f_{X_d}(x_d)$.*

$\triangleright$ – Fonction de répartition : Supposons que $\forall x \in \mathbb{R}^d,\ F_X(x) = \prod_{i=1}^{d} F_{X_i}(x_i)$. On note

$$\mathcal{D} = \{B = ]-\infty, x_1] \times \cdots \times ]-\infty, x_d],\ (x_1, \dots, x_d) \in \mathbb{R}^d\}$$

$\mathcal{D}$ est stable par intersection finie et $\mathcal{D}$ engendre $\mathcal{B}(\mathbb{R}^d)$. On conclut par le théorème des classes monotones que $X_1, \dots, X_d$ sont indépendantes.

Supposons que $X_1, \dots, X_d$ sont indépendantes, il suffit de prendre $A_1 = ]-\infty, x_1], \dots, A_d = ]-\infty, x_d]$.

– Espérance : $\Leftarrow$ Il suffit de prendre $h_i = \mathbf{1}_{A_i}$.
$\Rightarrow$ Soient $h_i$ comme voulue. Par le théorème de transfert,

$$\mathbb{E}[h_1(X_1) \dots h_d(X_d)] = \int_{\mathbb{R}^d} h_1(x_1) \dots h_d(x_d) \mathrm{d}\mathbb{P}_{(X_1,\dots,X_D)}(x_1, \dots, x_d).$$

Par indépendance,

$$\begin{aligned}
\mathbb{E}[h_1(X_1) \dots h_d(X_d)] &= \int_{\mathbb{R}^d} h_(x_1) \dots h_d(x_d) \mathbb{P}_{X_1}(\mathrm{d}x_1) \dots \mathbb{P}_{X_d}(\mathrm{d}x_d) \\
&= \int_{\mathbb{R}} h_1(x_1) \mathbb{P}_{X_1}(\mathrm{d}x_1) \times \cdots \times \int_{\mathbb{R}} h_d(x_d) \mathbb{P}_{X_d}(\mathrm{d}x_d) \text{ par le théorème de Fubini} \\
&= \mathbb{E}[h_1(X_1)] \dots \mathbb{E}[h_d(X_d)] \text{ par le théorème de transfert}
\end{aligned}$$

Le reste en découle. $\square$

**Exemples :** – Dans le cas discret. Soient $X, Y$ deux variables aléatoires à valeurs finies discrètes, $X$ et $Y$ sont indépendantes si et seulement si $\forall x, y$ $\mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x)\mathbb{P}(Y = y)$.
On jette deux dés.

$$X = \text{le nombre de chiffres pairs}$$

$$Y = \text{le maximum des deux chiffres}$$

$X$ et $Y$ ne sont pas indépendantes car $\mathbb{P}(X = 0, Y = 6) = 0$ mais $\mathbb{P}(X = 0) = \frac{1}{4} > 0$ et $\mathbb{P}(Y = 6) = \frac{11}{36} > 0$.

– Soit $(X, Y)$ un vecteur de densité $f_{(X,Y)}(x, y) = \frac{1}{3}\mathbf{1}_{[0,1]\times[-1,2]}(x, y) = \frac{1}{3}\mathbf{1}_{[0,1]}(x)\mathbf{1}_{[-1,2]}(y)$ on peut séparer les variables donc il y a indépendance.

$$f_{(X,Y)}(x, y) = \underbrace{\mathbf{1}_{[0,1]}(x)}_{=f_X(x)} \times \underbrace{\frac{1}{3}\mathbf{1}_{[-1,2]}(y)}_{=f_Y(y)}.$$

– Soit $(X, Y)$ un vecteur de densité $f_{(X,Y)}(x, y) = \frac{1}{3\pi}e^{-\frac{x^2+2xy+5y^2}{6}}$ : on ne peut pas séparer les variables, $X$ et $Y$ ne sont pas indépendantes.

$$\begin{aligned} f_X(x) &= \frac{1}{3}e^{-\frac{x^2}{6}} \int_{\mathbb{R}} e^{-\frac{2xy+5y^2}{6}} \mathrm{d}y = \frac{1}{3\pi}e^{-\frac{x^2}{6}} \int_{\mathbb{R}} e^{-\frac{5(ye+\frac{x}{5})^2}{6}} e^{\frac{x^2}{30}} \mathrm{d}y \\ &= \frac{1}{3\pi}e^{-\frac{2x^2}{15}} \underbrace{\int_{\mathbb{R}} e^{-\frac{5(y+\frac{x}{5})^2}{6}} \frac{\mathrm{d}y}{\sqrt{2\pi}\sigma}}_{\sigma^2=\frac{3}{5},\ \text{intégrale de la densité de } \mathcal{N}(-\frac{x}{5},\frac{3}{5})} \times\sqrt{2\pi}\sigma = \frac{1}{\sqrt{2\pi}}\frac{2}{\sqrt{15}}e^{-\frac{2x^2}{15}} \end{aligned}$$

donc $X \sim \mathcal{N}(0, \frac{15}{4})$.

$$\begin{aligned} f_Y(y) &= \frac{1}{3\pi}e^{-\frac{5y^2}{6}} \int_{\mathbb{R}} e^{-\frac{x^2+2xy}{6}} \mathrm{d}x = \frac{1}{3\pi}e^{-\frac{5y^2}{6}} \int_{\mathbb{R}} e^{-\frac{(x+y)^2}{6}} e^{\frac{y^2}{6}} \mathrm{d}x \\ &= \frac{1}{3\pi}e^{-\frac{2y^2}{3}} \left(\int_{\mathbb{R}} e^{-\frac{(x+y)^2}{6}} \frac{\mathrm{d}x}{\sqrt{2\pi}\sqrt{3}}\right) \times \sqrt{6\pi} = \sqrt{\frac{2}{3\pi}}e^{-\frac{2y^2}{3}} \end{aligned}$$

donc $Y \sim \mathcal{N}(0, \frac{3}{4})$.

$$f_X(x)f_Y(y) = \frac{2}{3\pi\sqrt{5}}e^{-\frac{2x^2}{15}}e^{-\frac{2y^2}{3}} \neq f_{(X,Y)}(x, y)$$

donc $X$ et $Y$ ne sont pas indépendants.

– $(X, Y)$ vecteur de densité

$$f_{(X,Y)}(x, y) = \begin{cases} \sqrt{\frac{x}{y}} & \text{si } 0 < y \leq x < 1 \\ 0 & \text{sinon} \end{cases}$$

Au vu du domaine de $f_{(X,Y)}$ on remarque que $Y \leq X$ : les variables sont dépendantes. Vérifier que $f_{(X,Y)}(x, y) \neq f_X(x)f_Y(y)$.

**Proposition 5.9**

*Soit $(X_i)_{i\in I}$ une famille de variables aléatoires. Les assertions suivantes sont équivalentes :*

– *$(X_i)_{i\in I}$ est indépendante.*
– *$\forall J \subset I$ partie finie de $I$, $(X_i)_{i\in I}$ est indépendante.*
– *$\forall J \subset I$, partie finie de $I$, $\mathbb{E}\left[\prod_{j\in J} h_j(X_j)\right] = \prod_{j\in J} \mathbb{E}[h_j(X_j)]$ pour toute $h_j$ borélienne telle que $h_j(X_j) \in L^1$.*

**Remarque :** Si $X_1, \ldots, X_d$ sont indépendantes alors

$$\mathbb{E}[X_1, \ldots, X_d] = \prod_{i=1}^{d} \mathbb{E}[X_i].$$

La réciproque est fausse : $X_1 \sim \mathcal{U}[-1,1]$, $X_2 = X_1^2$ $\mathbb{E}[X_1X_2] = \mathbb{E}[X_1^3] = 0 = \underbrace{\mathbb{E}[X_1]}_{=0}\mathbb{E}[X_2]$ mais $X_1$ et $X_2$ ne sont pas indépendantes :

$$\mathbb{P}\left(|X_1| \leq \frac{1}{2}, X_2 > \frac{1}{2}\right) = 0 \neq \underbrace{\mathbb{P}\left(|X_1| \leq \frac{1}{2}\right)\mathbb{P}\left(X_2 > \frac{1}{2}\right)}_{>0}.$$

## 5.3 Corrélation et indépendance

**Rappel :** Soient $X$ et $Y$ deux variables aléatoires dans $L^2$.

$$\mathrm{cov}(X,Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$$

**Définition 5.10**

*$X$ et $Y$ deux variables aléatoires dans $L^2$ sont dites non corrélées si* $\mathrm{cov}(X,Y) = 0$.

**Remarque :** $X, Y$ non corrélées si et seulement si $\langle X - \mathbb{E}[X], Y - \mathbb{E}[Y]\rangle = 0$ donc $X - \mathbb{E}[X]$ et $Y - \mathbb{E}[Y]$ sont orthogonaux pour le produit scalaire de $L^2$.

**Remarque :** Si $X$ et $Y$ sont indépendantes alors $X$ et $Y$ sont non corrélées. La réciproque est fausse (en général).

**Propriété 5.11**

*Si $X$ et $Y$ non corrélées alors* $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$.

▷ $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\,\mathrm{cov}(X,Y)$. □

**Propriété 5.12**

*Si $X_1, \ldots, X_d$ sont deux à deux non corrélées, alors*

$$\mathrm{Var}\left(\sum_{i=1}^{d} X_i\right) = \sum_{i=1}^{d} \mathrm{Var}(X_i)$$

$\triangleright$ $\text{Var}\left(\sum_{i=1}^{d}\right) = \text{cov}\left(\sum_{i=1}^{d} X_i, \sum_{j=1}^{d} X_j\right) = \sum_{i,j} \text{cov}(X_i, X_j)$ car cov est une fonction bilinéaire. Donc

$$\text{Var}\left(\sum_{i=1}^{d} X_i\right) = \sum_{i=1}^{d} \text{Var}(X_i) + \sum_{i \neq j} \text{cov}(X_i, X_j) = \sum_{i=1}^{d} \text{Var}(X_i).$$

$\square$

## 5.4 Événements asymptotiques

Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité.

### 5.4.1 Tribu du futur et tribu asymptotique

**Définition 5.13**

*Soit $(\mathcal{F}_n)_{n\in\mathbb{N}}$ une suite de sous-tribus de $\mathcal{F}$. La tribu du futur est définie par*

$$\mathcal{F}^n = \sigma\left(\bigcup_{k\geq n} \mathcal{F}_k\right)$$

*La tribu asymptotique est définie par*

$$\mathcal{F}^\infty = \bigcap_{n\geq 1} \mathcal{F}^n.$$

**Exemple :** On considère une suite $(X_n)_{n\geq 1}$ de variables aléatoires (on joue à pile ou face, $X_n$ est le résultat du $n$-ième lancer).

$$\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$$

tribu engendrée par les $n$ premières variables. $\mathcal{F}_n$ représente l'ensemble des événements avant le $n$-ième lancer. $\mathcal{F}_n$ est "l'histoire avant l'instant $n$". Ici $\mathcal{F}_n \subset \mathcal{F}_{n+1}$.

$$\mathcal{F}^n = \sigma\left(\bigcup_{k\geq n} \mathcal{F}_k\right) = \sigma(\{X_k,\ k \geq n\})$$

$\mathcal{F}^n$ représente les résultats après l'instant $n$, $\mathcal{F}^n$ est "le futur après l'instant $n$".

**Exemple :** $(X_n)_{n\geq 1}$ une suite de variables aléatoires. On définit $\mathcal{F}_n = \sigma(X_n)$. On suppose $\mathcal{F}_n \neq \mathcal{F}_k$ si $n \neq k$.

$$A = \{(X_n) \text{ passe une infinité de fois par } 0\} = \{\forall n \geq 1,\ \exists k \geq n,\ X_k = 0\} = \bigcap_{n\geq 1} \bigcup_{k\geq n} \{X_k = 0\}$$

donc $A = \limsup_n \{X_n = 0\}$. Comme $\forall n \geq 1,\ \bigcup_{k\geq n} \{X_k = 0\} \in \mathcal{F}^n$ donc $A \in \mathcal{F}^\infty$.

$$B = \{(X_n)_{n\in\mathbb{N}} \text{ est bornée par } M = \{\forall n \geq 1,\ |X_n| \leq M\} = \bigcap_{n\geq 1}\{|X_n| \leq M\} \notin \mathcal{F}^\infty$$

car $\{|X_1| \leq M\} \notin \mathcal{F}^n,\ \forall n \geq 2$.

$$\begin{aligned} C &= \{\text{la suite } (X_n)_{n\geq 1} \text{ converge}\} = \{\text{la suite } (X_{n+q})_{n\geq 1} \text{ converge}\} \\ &= \bigcap_{\varepsilon\in\mathbb{Q}^*} \bigcup_{N\geq q} \bigcap_{\substack{n\geq N \\ p\geq 0}} \{|X_n - X_{n+p}| < \varepsilon\} \in \mathcal{F}^q \end{aligned}$$

donc $C \in \mathcal{F}^q$ pour tout $q \geq 1$ donc $C \in \mathcal{F}^\infty$.

**Théorème 5.14 (*Loi du 0-1 de Kolmogorov*)**

*Soit $(\mathcal{F}_n)_{n\geq 1}$ une suite de tribu indépendantes. Alors la tribu asymptotique est triviale*

$$\forall A \in \mathcal{F}^\infty, \qquad \mathbb{P}(A) = 0 \text{ ou } 1$$

**Exemple :** $(X_n)$ suite de variables indépendantes $\mathcal{F}_n = \sigma(X_n)$. On a $(\mathcal{F}_n)$ indépendantes et donc $\mathbb{P}(C) = \mathbb{P}((X_n) \text{ converge}) = 0$ ou 1.

$\triangleright$ On va montrer que $\mathcal{F}^\infty$ est indépendante d'elle même. En effet, si c'est le cas, pour $A, B \in \mathcal{F}^\infty$, $\mathbb{P}(A\cap B) = \mathbb{P}(A)\mathbb{P}(B)$. Par conséquent, si on prend $B = a$, $\mathbb{P}(A) = \mathbb{P}(A)^2$ donc $\mathbb{P}(A) = 0$ ou 1.

Par indépendance des tribus, $\mathcal{F}^n$ est indépendante de $\sigma\left(\bigcup_{k<n}\mathcal{F}_k\right)$. Comme $\mathcal{F}^\infty \subset \mathcal{F}^n$, $\mathcal{F}^\infty$ est indépendant de $\sigma\left(\bigcup_{k<n}\mathcal{F}_k\right)$. Ceci est vrai pour tout $n \in \mathbb{N}$. Par conséquent, $\mathcal{F}^\infty$ est indépendante de $\mathcal{A} = \bigcup_{n\in\mathbb{N}} \sigma\left(\bigcup_{k<n}\mathcal{F}_k\right)$.

Montrons que $\mathcal{A}$ est stable par intersection finie. Soient $A, B \in \mathcal{A}$. Il existe $n, m \geq 1$ tels que $A \in \sigma\left(\bigcup_{k<n}\mathcal{F}_k\right)$ et $B \in \sigma\left(\bigcup_{k<m}\mathcal{F}_k\right)$. Donc $A, B \in \sigma\left(\bigcup_{k<\max(m,n)}\mathcal{F}_k\right)$ qui est un tribu. Par conséquent, $A\cap B \in \sigma\left(\bigcup_{k<\max(n,m)}\mathcal{F}_k\right)$ donc $A \cap B \in \mathcal{A}$.

Ainsi, $\mathcal{F}^\infty$ est indépendante de $\mathcal{A}$, $\mathcal{F}^\infty$ et $\mathcal{A}$ sont stables par intersection finie, donc $\mathcal{F}^\infty$ et $\sigma(\mathcal{A})$ sont indépendantes. On remarque que

$$\forall k \geq 1, \qquad \mathcal{F}_k \subset \sigma(\mathcal{A})$$

donc

$$\mathcal{F}^\infty \subset \mathcal{F}^1 = \bigcup_{k\geq 1}\mathcal{F}_k \subset \sigma(\mathcal{A}).$$

Comme $\mathcal{F}^\infty$ et $\sigma(A)$ sont indépendantes, $\mathcal{F}^\infty$ est indépendante d'elle-même. $\square$

**Remarque :** Si $(\mathcal{F}_n)$ est une suite de tribus indépendantes, alors $\mathcal{F}^\infty$ est triviale. Les variables aléatoires $\mathcal{F}^\infty$-mesurables sont donc les variables constantes presque sûrement.

### 5.4.2 Lemmes de Borel-Cantelli

**Lemme 5.15 (*Borel-Cantelli 1*)**

*Soit $(A_n)_{n\geq 1}$ une suite d'événements telle que*

$$\sum_{n\geq 1} \mathbb{P}(A_n) < \infty.$$

*Alors*

$$\mathbb{P}(\limsup_n A_n) = 0.$$

*ie. seul un nombre fini de $A_n$ est réalisé.*

$\triangleright$ La suite $\left(\bigcup_{k\geq n} A_k\right)_{n\geq 1}$ est une suite décroissante donc, comme $\limsup_n = \bigcap_{n\geq 1}\bigcup_{k\geq n} A_k$,

$$\mathbb{P}(\limsup_n) = \lim_{n\to+\infty} \mathbb{P}\left(\bigcup_{k\geq n} A_k\right) \leq \lim_{n\to+\infty} \sum_{k\geq n} \mathbb{P}(A_k) = 0 \qquad \text{(reste d'une série convergente)}$$

donc $\mathbb{P}(\limsup A_n) = 0$. $\square$

**Lemme 5.16 (*Borel-Cantelli 2*)**

*Soit $(A_n)_{n\geq 1}$ une suite d'événements indépendants telle que $\sum \mathbb{P}(A_n)$ diverge. Alors*

$$\mathbb{P}(\limsup A_n) = 1.$$

*ie. , presque sûrement, il y a une infinité d'événements $A_n$ qui se réalise.*

$\triangleright$ $\mathbb{P}(\limsup A_n) = 1 - \mathbb{P}(\liminf A_n^c)$. Or $\mathbb{P}(\liminf A_n^c) = \mathbb{P}\left(\bigcup_{n\geq 1}\bigcap_{k\geq n} A_k\right)$. Or $\left(\bigcap_{k\geq n} A_k\right)$ est une suite croissante et $\left(\bigcap_{k=n}^{m} A_k^c\right)$ est une suite décroissante donc

$$\begin{aligned}
\mathbb{P}(\liminf A_n^c) &= \lim_{n\to+\infty} \mathbb{P}\left(\bigcap_{k\geq n} A_k^c\right) = \lim_{n\to+\infty}\lim_{m\to+\infty} \mathbb{P}\left(\bigcap_{k=n}^{m} A_k^c\right) \\
&= \lim_{n\to+\infty}\lim_{m\to+\infty} \prod_{k=n}^{m} \mathbb{P}(A_k^c) = \lim_{n\to+\infty}\lim_{m\to+\infty} \prod_{k=n}^{m} \underbrace{(1-\mathbb{P}(A_k))}_{\leq e^{-\mathbb{P}(A_k)}} \\
&= \lim_{n\to+\infty}\lim_{m\to+\infty} \exp\left(-\sum_{k=n}^{m} \mathbb{P}(A_k)\right) = 0
\end{aligned}$$

$\square$

## 5.5 Probabilité conditionnelle

**Rappel :** $A$ et $B$ sont indépendants si et seulement si

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

Que faire si $A$ et $B$ ne sont pas indépendants ?

**Exemple :** : On lance deux dés.

$$X = \text{le nombre de chiffres pairs}$$

$$Y = \text{le maximum des deux chiffres}$$

On prend

$$A = \{Y = 1\}$$

et

$$B = \{X = 2\}.$$

Sachant que $B$ est réalisé, l'événement $A$ est réalisé avec probabilité 0. Mais $\mathbb{P}(A) = \frac{1}{36} \neq 0$. La connaissance de $B$ modifie la probabilité de réalisation de $A$.

**Définition 5.17**

*Soient $A, B \in \mathcal{F}$ avec $\mathbb{P}(B) \neq 0$. On définit la probabilité de $A$ sachant $B$, notée $\mathbb{P}(A|B)$, par*

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

**Remarque :** Si $A$ et $B$ sont indépendants,

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

Le fait de savoir que $B$ est réalisé n'influence pas la probabilité de réalisation de $A$.

**Propriété 5.18**

*L'application* $\begin{array}{ccc} \mathcal{F} & \to & [0,1] \\ A & \mapsto & \mathbb{P}(A|B) \end{array}$ *est une mesure de probabilité.*

$\triangleright$
- $\mathbb{P}(\Omega|B) = \frac{\mathbb{P}(B)}{\mathbb{P}(B)} = 1.$
- Soit $(A_i)_{i \in I}$ une suite d'événements disjoints.

$$\mathbb{P}\left(\bigcup_{i\in I} A_i | B\right) = \frac{\mathbb{P}\left(\left(\bigcup_{i\in I} A_i\right) | B\right)}{\mathbb{P}(B)} = \frac{\mathbb{P}\left(\bigcup_{i\in I} (A_i \cap B)\right)}{\mathbb{P}(B)} = \frac{\sum_{i\in I} \mathbb{P}(A_i \cap B)}{\mathbb{P}(B)} = \sum_{i\in I} \mathbb{P}(A_i|B).$$

$\square$

**Conséquences :** – $\forall A \in \mathcal{F}$, $\mathbb{P}(A|B) = 1 - \mathbb{P}(A^c|B)$

– si $(A_i)$ est une suite croissante

$$\mathbb{P}\left(\bigcup A_i|B\right) = \lim \mathbb{P}(A_i|B).$$

etc.

⚠ $\mathbb{P}(A|B^c) \neq 1 - \mathbb{P}(A|B)$.

**Formule des probabilités totales :** Soient $A, B \in \mathcal{F}$ avec $\mathbb{P}(B) \neq 0$.

$$\mathbb{P}(A) = \mathbb{P}((A \cap B) \text{ ou } (A \cap B^c)) = \mathbb{P}(A \cap B) + \mathbb{P}(A \cap B^c) = \mathbb{P}(A|B)\mathbb{P}(B) + \mathbb{P}(A|B^c)\mathbb{P}(B^c)$$

donc

$$\mathbb{P}(A) = \mathbb{P}(A|B)\mathbb{P}(B) + \mathbb{P}(A|B^c)(1 - \mathbb{P}(B)).$$

De façon générale, si $(B_i)_{1\leq i\leq n}$ est une partition de $\Omega$, alors

$$\mathbb{P}(A) = \sum_{i=1}^{n} \mathbb{P}(A|B_i)\mathbb{P}(B_i)$$

**Formule de Bayes :** Soient $A, B \in \mathcal{F}$ avec $\mathbb{P}(A) \neq 0$ et $\mathbb{P}(B) \neq 0$.

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(A|B)\mathbb{P}(B)}{\mathbb{P}(A|B)\mathbb{P}(B) + \mathbb{P}(A|B^c)(1 - \mathbb{P}(B))}.$$

**Exemple :** Une maladie affecte 0,5% de la population. Un test T permet de dépister la maladie avec la fiabilité suivantes.

– T est positif pour 95% des personnes malades,

– T est négatif pour 99% des personnes saines.

Quelle est la probabilité qu'un individu ayant un test positif soit malade ?

On veut calculer $\mathbb{P}(M|P)$. On sait que $\mathbb{P}(P|M) = 0,95$, $\mathbb{P}(P^c|M^c) = 0,99$ et $\mathbb{P}(M) = 0,005$. Par la formule de Bayes,

$$\mathbb{P}(M|P) = \frac{\mathbb{P}(P|M)\mathbb{P}(M)}{\mathbb{P}(P|M)\mathbb{P}(M) + \mathbb{P}(P|M^c)(1 - \mathbb{P}(M))} \approx 0,323.$$

**Application :** Loi conditionnelle.

Soit $B \in \mathcal{F}, \mathbb{P}(B) \neq 0$. Soit $X$ une variable aléatoire. La loi conditionnelle de $X$ sachant l'événement $B$ est définie par $\mathbb{P}_X(\cdot|B)$, c'est-à-dire $\mathbb{P}(X \in A|B)$, $\forall A \in \mathcal{B}(\mathbb{R})$.

L'espérance conditionnelle (si elle existe) de $X$ sachant $B$, notée $\mathbb{E}[X|B]$ est l'espérance de la loi conditionnelle $\mathbb{P}_X(\cdot|B)$.

Dans le cas discret : si $X$ est à valeurs dans $\{x_i\}$, il suffit de calculer $\mathbb{P}(X = x_i|B)$ $\forall i \in I$ pour avoir $\mathbb{P}_X(\cdot|B)$.

**Exemple :** Soient $N$ une variable aléatoire à valeurs dans $\mathbb{N}$, $X_i$ des variables de loi $\mathcal{B}(p)$ avec $N, (X_i)_{i\in I}$ indépendants. On considère
$$S = \sum_{i=1}^{N} X_i.$$
Quelle est la loi de $S$ sachant que $N = n$ ?

$\sum_{i=1}^{n} X_i \in \{0, \ldots, n\}$ et $\mathbb{P}\left(\sum_{i=1}^{n} X_i = k\right) = \mathbb{P}(k$ variables parmi les $X_i$ sont égales à $1)=$ $\binom{n}{k} p^k(1-p)^{n-k}$. On a donc $\sum_{i=1}^{n} X_i \sim \mathcal{B}(n,p)$. Sachant que $N = n$, $S \sim \mathcal{B}(n,p)$. En effet, soit $k \geq 0$,

$$\begin{aligned}
\mathbb{P}(S = k|N = n) &= \mathbb{P}\left(\sum_{i=1}^{N} X_i = K|N = n\right) \\
&= \begin{cases} 0 \text{ si } k > n \\ \mathbb{P}\left(\sum_{i=1}^{n} X_i = k|N = n\right) = \mathbb{P}\left(\sum_{i=1}^{n} X_i = k\right) = \binom{n}{k} p^k(1-p)^{n-k} \text{ si } k \leq n. \end{cases}
\end{aligned}$$

**Remarque :** Si on connait la loi de $N$, on peut en déduire la loi du couple $(S, N)$. Pour $k, n \in \mathbb{N}$,
$$\mathbb{P}(S = k, N = n) = \mathbb{P}(S = k|N = n)\mathbb{P}(N = n).$$

## 5.6 Somme de variables aléatoires indépendantes

Soient $X$ et $Y$ deux variables aléatoires indépendantes.

**But :** Trouver la loi $\mathbb{P}_{X+Y}$ de $X + Y$.

On sait que $(X, Y)$ a pour loi $\mathbb{P}_X \otimes \mathbb{P}_Y$. Par conséquent, la fonction de répartition de $X+Y$ est
$$F_{X+Y}(t) = \mathbb{P}(X + Y \leq t) = \iint_{\mathbb{R}^2} \mathbf{1}_{x+y\leq t}\mathbb{P}_X(\mathrm{d}x)\mathbb{P}_Y(\mathrm{d}y)$$
Par le théorème de Fubini-Tonelli,
$$F_{X+Y}(t) = \int_{\mathbb{R}} \left(\int_{-\infty}^{t-y} \mathbb{P}_X(\mathrm{d}x)\right) \mathbb{P}_X(\mathrm{d}y) = \int_{\mathbb{R}} F_X(t-y)\mathbb{P}_Y(\mathrm{d}y)$$
et de même
$$F_{X+Y}(t) = \int_{\mathbb{R}} F_Y(t-x)\mathbb{P}_X(\mathrm{d}x).$$
On remarque que pour $A \in \mathcal{B}(\mathbb{R})$
$$\mathbb{P}(X + Y \in A) = \int \mathbf{1}_{x+y\in A}\mathbb{P}_X(\mathrm{d}x)\mathbb{P}_Y(\mathrm{d}y) = \int_{\mathbb{R}} \mathbb{P}_Y(A-x)\mathbb{P}_X(\mathrm{d}x) = \int_{\mathbb{R}} \mathbb{P}_X(A-y)\mathbb{P}_Y(\mathrm{d}y)$$
où $A - x = \{y - x,\ y \in A\}$.

### 5.6.1 Convolution de mesures

On considère $(E, \mathcal{E})$ un espace vectoriel mesuré.

**Définition 5.19**

*Soient $\mu$ et $\nu$ deux mesures sur $(E, \mathcal{E})$. On définit la convolée de $\mu$ avec $\nu$, notée $\mu * \nu$, par*

$$\forall A \in \mathcal{E}, \qquad \mu * \nu(A) = \int_E \mu(A - x)\nu(dx)$$

**Propriété 5.20**

- *$\mu * \nu$ est une mesure positive sur $(E, \mathcal{E})$.*
- *$\mu * \nu = \nu * \mu$.*
- *Si $\mu$ et $\nu$ sont des mesures de probabilité alors $\mu * \nu$ est une mesure de probabilité.*
- *$\mu * \delta_0 = \mu$.*
- *Le produit de convolution est associatif.*

$\triangleright$ – $\mu * \nu(\varnothing) = 0$ car $\mu(\varnothing - x) = \mu(\varnothing) = 0$.

– $(A_i)$ ensembles disjoints. Pour tout $x \in E$, les $(A_i - x)$ sont disjoints. En effet, si $z \in (A_i - x) \cap (A_j - x)$ alors $z = y_i - x = y_j - x$ donc $y_i = y_j$ donc $i = j$. Alors, par le théorème de convergence monotone,

$$\mu * \nu\left(\bigcup_{i \in I} A_i\right) = \int_E \sum_{i \in I} \mu(A_i - x)\nu(\mathrm{d}x) = \sum_{i \in I} \int_E \mu(A_i - x)\nu(\mathrm{d}x) = \sum_{i \in I} \mu * \nu(A_i).$$

Donc $\mu * \nu$ est bien une mesure.

– Si $\mu$ et $\nu$ sont des mesures de probabilité,

$$\mu * \nu(E) = \int_E \mu(\underbrace{E - x}_{=E})\nu(\mathrm{d}x) = \mu(E)\nu(E) = 1$$

donc $\mu * \nu$ est une mesure de probabilité.

– Par le théorème de Fubini-Tonelli

$$\mu * \nu(A) = \int_E \mu(A - x)\nu(\mathrm{d}x) = \int_{E \times E} \mathbf{1}_{x+y \in A} \mu(\mathrm{d}y)\nu(\mathrm{d}x) = \nu * \mu(A).$$

– $\mu * \delta_0(A) = \displaystyle\int_{E \times E} \mathbf{1}_{x+y \in A} \delta_0(\mathrm{d}y)\mu(\mathrm{d}x) = \int_E \mathbf{1}_{x \in A} \mu(\mathrm{d}x) = \mu(A).$ $\square$

**Conséquences :** – Si $X$ et $Y$ sont des variables aléatoires indépendantes alors $\mathbb{P}_{X+Y} = \mathbb{P}_X * \mathbb{P}_Y$.

– Si $X_1, \ldots, X_n$ sont (mutuellement) indépendantes alors

$$\mathbb{P}_{X_1 + \cdots + X_n} = \mathbb{P}_{X_1} * \cdots * \mathbb{P}_{X_n}.$$

– La fonction caractéristique de $\displaystyle\sum_{i=1}^n X_i$ est

$$\varphi_{X_1 + \cdots + X_n} = \mathbb{E}\left[e^{it(X_1 + \cdots + X_n)}\right] = \prod_{i=1}^n \varphi_{X_i(t)}$$

car les variables sont indépendantes.

### 5.6.2 Exemples

**Cas discret :** Soient $X$ une variable aléatoire à valeurs dans $\{x_i\}_{i\in I}$ et $Y$ une variable aléatoire à valeurs dans $\{y_j\}_{j\in J}$ telles que $X$ et $Y$ sont indépendantes. On définit $Z = X+Y$. $Z$ est à valeurs dans $\{x_i + y_j,\ i \in I,\ j \in J\}$.

$$\mathbb{P}(Z = z) = \mathbb{P}(X+Y = z) = \sum_{i\in I} \mathbb{P}(Y = z - x_i)\mathbb{P}(X = x_i) = \sum_{j\in J} \mathbb{P}(X = z - y_j)\mathbb{P}(Y = y_j)$$

– Si $X_1, \dots, X_n$ sont indépendantes de loi $\mathcal{B}(p)$ alors $S = X_1 + \cdots + X_n$ suit la loi $\mathcal{B}(n,p)$
Conséquence : $\mathbb{E}[S] = \mathbb{E}[X_1] + \cdots + \mathbb{E}[X_n] = np$ car l'espérance est linéaire. $\mathrm{Var}(S) = \mathrm{Var}(X_1) + \cdots + \mathrm{Var}(X_n) = np(1-p)$ car les variables sont indépendantes.

**Cas à densité**

**Proposition 5.21**

*Si $X$ est une variable aléatoire de densité $f_X$, $Y$ est une variable de densité $f_Y$ et $X$ et $Y$ sont indépendantes alors $X+Y$ est une variable à densité de densité*

$$f_{X+Y}(z) = \int_{\mathbb{R}} f_X(z-y) f_Y(y) \mathrm{d}y = \int_{\mathbb{R}} f_Y(z-x) f_X(x) \mathrm{d}x = f_X * f_Y(z).$$

$\triangleright$ Soit $A \in \mathcal{B}(\mathbb{R})$

$$\mathbb{P}(X+Y \in A) = \int_{\mathbb{R}^2} \mathbf{1}_{x+y\in A} \mathbb{P}_X(\mathrm{d}x)\mathbb{P}_Y(\mathrm{d}y) = \int_{\mathbb{R}^2} \mathbf{1}_{x+y\in A} f_X(x) f_Y(y) \mathrm{d}x\mathrm{d}y$$

On pose $z = x+y$ et $t = x$. $x \in \mathbb{R}, y \in \mathbb{R}, x+y \in A \iff z \in A, t \in \mathbb{R}$ donc par changement de variables

$$\mathbb{P}(X+Y \in A) = \int_{A\times\mathbb{R}} f_X(t) f_Y(z-t) \mathrm{d}z\mathrm{d}t = \int_A \left( \int_{\mathbb{R}} f_X(t) f_Y(z-t) \mathrm{d}t \right) \mathrm{d}z.$$

$\square$

**Exemple :** Soient $X \sim \mathcal{N}(m_X, \sigma_X^2)$ et $Y \sim \mathcal{N}(m_Y, \sigma_Y^2)$ des variables aléatoires indépendantes. Quelle est la loi de $X+Y$ ?

$$\varphi_{X+Y}(t) = \varphi_X(t)\varphi_Y(t) = e^{itm_X} e^{\frac{t^2\sigma_X^2}{2}} e^{itm_Y} e^{\frac{t^2\sigma_Y^2}{2}} = e^{it(m_X+m_Y)} e^{\frac{t^2}{2}(\sigma_X^2+\sigma_Y^2)}$$

donc $X+Y \sim \mathcal{N}(m_X + m_Y, \sigma_X^2 + \sigma_Y^2)$.

**Remarque :** L'utilisation de la fonction caractéristique est en général bien utile pour trouver la loi d'une somme de variables aléatoires indépendantes.

# Chapitre 6

# Convergence d'une suite de variables aléatoires

## 6.1 Convergence trajectorielle

Soient $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité, $X$ une variable aléatoire et $(X_n)_{n\in\mathbb{N}}$ une suite de variable aléatoire sur $(\Omega, \mathcal{F}, \mathbb{P})$ à valeurs dans $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. $\mathbb{R}^d$ est muni de la norme euclidienne $\|.\|$.

### 6.1.1 Convergence presque sûre

Elle correspond à la convergence simple des suites de fonctions :

$$\forall \omega \in \Omega, \qquad X_n(\omega) \xrightarrow[n\to+\infty]{} X(\omega) \qquad \text{(convergence simple)}$$

Cette notion est trop forte pour les probabilités.

**Définition 6.1**

*La suite $(X_n)_{n\in\mathbb{N}}$ converge presque surement vers $X$ s'il existe $N \in \mathcal{F}$ tel que $\mathbb{P}(N) = 0$ et $\forall \omega \in N,\ X_n(\omega) \xrightarrow[n\to+\infty]{} X(\omega)$.*

*On note alors $X_n \xrightarrow[n\to+\infty]{p.s} X$.*

**Remarque :** 1. $\{X_n \xrightarrow[n\to+\infty]{} X\} = \{\omega \in \Omega,\ X_n(\omega) \xrightarrow[n\to+\infty]{} X(\omega)\} = \bigcap_{p\in\mathbb{N}^*} \bigcup_{n\in\mathbb{N}} \bigcap_{k\geq n} \left\{|X_n - X| < \frac{1}{p}\right\}$ donc $\{X_n \xrightarrow[n\to+\infty]{} X\} \in \mathcal{F}$. Ainsi,

$$X_n \xrightarrow[n\to+\infty]{\text{p.s}} X \qquad \Longleftrightarrow \qquad \mathbb{P}(X_n \xrightarrow[n\to+\infty]{} X) = 1.$$

2. Si $X_n = (X_n^1, \ldots, X_n^d)$

$$X_n \xrightarrow[n\to+\infty]{\text{p.s}} X = (X^1, \ldots, X^d) \qquad \Longleftrightarrow \qquad \forall i \in \{1, \ldots, d\},\ X_n^i \xrightarrow[n\to+\infty]{\text{p.s}} X^i.$$

**Proposition 6.2**

$$X_n \xrightarrow[n\to+\infty]{p.s} X \qquad \Longleftrightarrow \qquad \forall \varepsilon > 0,\ \mathbb{P}(\limsup\{|X_n - X| > \varepsilon\}) = 0$$

▷ $\Rightarrow$ : Soit $\varepsilon > 0$. Soit $\omega \in \limsup\{|X_n - X| > \varepsilon\}$.

$$\forall n \geq 0,\ \exists k \geq n, \qquad |X_k(\omega) - X(\omega)| > \varepsilon.$$

Il y a une infinité de $k$ telle que $|X_k(\omega) - X(\omega)| > \varepsilon$. Donc $X_n(\Omega)$ ne converge pas vers $X(\Omega)$. Ainsi, $\Omega \in \{X_n \xrightarrow[n\to+\infty]{} X\}^c$. Ainsi, $\mathbb{P}(\limsup\{|X_n - X| > \varepsilon\}) \leq \mathbb{P}(\{X_n \xrightarrow[n\to+\infty]{} X\}^c) = 0$.

$\Leftarrow$ : $N = \bigcup_{p\in\mathbb{N}^*} \limsup\left\{|X_n - X| > \frac{1}{p}\right\}$ est réunion dénombrable d'ensembles négligeables donc $\mathbb{P}(N) = 0$. Si $\omega \notin N$ alors

$$\omega \in \bigcap_{p\in\mathbb{N}^*} \liminf\left\{|X_n - X| \leq \frac{1}{p}\right\}$$

donc

$$\forall p \in \mathbb{N}^*,\ \exists n = n(\omega, p),\ \forall k \geq n, \qquad |X_k(\omega) - X(\omega)| \leq \frac{1}{p}$$

*ie.* $X_n(\omega) \xrightarrow[n\to+\infty]{} X(\omega)$. Ainsi $X_n \xrightarrow[n\to+\infty]{\text{p.s}} X$. □

**Corollaire 6.3**

*Si* $\forall \varepsilon > 0 \sum_{n\geq 0} \mathbb{P}(|X_n - X| > \varepsilon)$ *converge alors* $X_n \xrightarrow[n\to+\infty]{p.s} X$.

▷ Soit $\varepsilon > 0$. Comme $\sum \mathbb{P}(|X_n - X| > \varepsilon)$ converge, par le premier lemme de Borel-Cantelli on a $\mathbb{P}(\limsup\{|X_n - X| > \varepsilon\}) = 0$ et on conclut par la proposition précédente. □

**Proposition 6.4**

*Si les* $(X_n)$ *sont mutuellement indépendantes, on a*

$$X_n \xrightarrow[n\to+\infty]{p.s} 0 \qquad \Longleftrightarrow \qquad \forall \varepsilon > 0,\ \sum_{n\geq 0} \mathbb{P}(|X_n| > \varepsilon) \text{ converge.}$$

▷ $\Leftarrow$ : fait dans le corollaire précédent.

$\Rightarrow$ : Posons $A_n = \{|X_n| > \varepsilon\}$. Les $(A_n)$ sont des événements indépendants. Si $\sum_{n\geq 0} \mathbb{P}(A_n)$ diverge alors par le second lemme de Borel-Cantelli, $\mathbb{P}(\limsup A_n) = 1$ ce qui contredit la convergence presque sure. Donc $\sum_{n\geq 0} \mathbb{P}(|X_n| > \varepsilon)$ converge. □

**Lemme 6.5**

*Soit* $(\varepsilon_n)$ *suite de réels positifs telle que* $\sum_{n=0}^{\infty} \varepsilon_n < +\infty$. *Si* $\sum_{n\geq 0} \mathbb{P}(|X_{n+1} - X_n| > \varepsilon_n)$ *est convergente alors* $(X_n)$ *converge presque sûrement.*

▷ $X_n = X_0 + \sum_{k=0}^{n-1}(X_{k+1} - X_k)$. Par le lemme de Borel-Cantelli, comme $\sum_{n\geq 0} \mathbb{P}(|X_{n+1} - X_n| > \varepsilon_n)$ converge, $\mathbb{P}(\limsup\{X_{n+1} - X_n| > \varepsilon\}) = 0$. Soit $\omega \notin \limsup\{|X_{n+1} - X_n| > \varepsilon\}$. Il existe $n(\omega) \geq 0$

tel que $\forall k \geq n$ $|X_{k+1}(\omega) - X_k(\omega)| \leq \varepsilon_k$. Donc $\sum_{k\geq 0} |X_{k+1}(\omega) - X_k(\omega)|$ est convergente donc $(X_n)$ converge. On a bien $(X_n)$ converge presque sûrement. $\square$

**Propriétés 6.6**

- *Si* $X_n \xrightarrow[n\to+\infty]{p.s} X$ *et* $X_n \xrightarrow[n\to+\infty]{p.s} Y$ *alors* $X = Y$ *presque sûrement.*
- *Si* $X_n \xrightarrow[n\to+\infty]{p.s} X$ *et* $f : \mathbb{R}^d \to \mathbb{R}^d$ *est continue alors* $f(X_n) \xrightarrow[n\to+\infty]{p.s} f(X)$.
- *Si* $\forall n \in \mathbb{N},\ X_n \overset{p.s}{=} Y_n$,

$$X_n \xrightarrow[n\to+\infty]{p.s} X \qquad \Longleftrightarrow \qquad Y_n \xrightarrow[n\to+\infty]{p.s} Y.$$

### 6.1.2 Convergence dans $L^p$

$L^p$ est l'espace quotient de $\mathcal{L}^p$ (ensemble des variables aléatoires $X$ telles que $|X|^p$ est intégrable) avec la relation d'équivalence $X \sim Y \Leftrightarrow X = Y$ p.s.

**Définition 6.7**

*Soit* $(X_n)$ *une suite de variables aléatoires de* $L^p$. $(X_n)$ *converge vers* $X$ *dans* $L^p$ *(en moyenne d'ordre p) si* $\mathbb{E}[|X_n - X|^p] \xrightarrow[n\to+\infty]{} 0$. *On note* $X_n \xrightarrow[n\to+\infty]{L^p} X$.

**Proposition 6.8**

*Si* $(X_n)$ *est une suite de variable de* $L^p$ *telle que* $X_n \xrightarrow[n\to+\infty]{L^p} X$ *alors pour tout* $q \leq p$, $X_n \xrightarrow[n\to+\infty]{L^p} X$.

▷ Par l'inégalité de Hölder,

$$\mathbb{E}[|X_n - X|^q] \leq \mathbb{E}[|X_n - X|^p]^{\frac{q}{p}} \xrightarrow[n\to+\infty]{} 0.$$

$\square$

**Théorème 6.9 (*de convergence dominée*)**

*Soit* $(X_n)$ *une suite de variables aléatoires telle que*
- *il existe une variable aléatoire* $Y$ *positive dans* $L^p$ *telle que* $|X_n| \leq Y$ *p.s*
- $X_n \xrightarrow[n\to+\infty]{p.s} X$.

*Alors* $X_n \in L^p$ *et* $X_n \xrightarrow[n\to+\infty]{L^p} X$.

**Exemple :** Soit $(X_n)$ une suite de variables aléatoires de loi

$$\mathbb{P}(X_n = 0) = 1 - p_n \qquad \mathbb{P}(X_n = x_n) = p_n$$

avec $p_n \in ]0, 1[$ et $x_n \geq 1$.

Les $(X_n)$ sont indépendantes donc $X_n \xrightarrow[n\to+\infty]{\text{p.s}}$ si et seulement si pour tout $\varepsilon > 0$ $\sum_{n\geq 0} \mathbb{P}(|X_n| > \varepsilon)$

converge, *ie.* $\sum_{n\geq 0} p_n < \infty$.

$(X_n)$ converge vers 0 dans $L^p$ si et seulement si $\mathbb{E}[|X_n|^p] = x_n^p p_n \xrightarrow[n\to+\infty]{} 0$.

- Si $p_n = \frac{1}{n}$ et $x_n = 1$. Alors $X_n$ ne converge pas ps mais $X_n$ converge dans tous les $L^p$.
- Si $p_n = 2^{-n}$ et $x_n = 2^n$ alors $X_n \xrightarrow[n\to+\infty]{\text{p.s}} 0$ mais $X_n$ ne converge dans aucun $L^p$.
- Si $p_n = \frac{1}{n^2}$ et $x_n = n$ alors $X_n \xrightarrow[n\to+\infty]{L^1} 0$ mais ne converge pas dans $L^2$.

### 6.1.3 Convergence en probabilité

Soient $(X_n)$ une suite de variables aléatoires sur $(\Omega, \mathcal{F}, \mathbb{P})$ et $X$ une variable aléatoire.

**Définition 6.10**

*La suite $(X_n)$ converge en probabilité vers $X$, noté $X_n \xrightarrow[n\to+\infty]{\mathbb{P}} X$, si*

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

*ou encore*

$$\forall \varepsilon > 0, \qquad \lim_{n\to+\infty} \mathbb{P}(|X_n - X| < \varepsilon) = 1.$$

**Propriété 6.11**

*$X_n$ est à valeurs dans $\mathbb{R}^d$.*

$$X_n = (X_n^1, \dots, X_n^d) \xrightarrow[n\to+\infty]{\mathbb{P}} X = (X^1, \dots, X^d) \qquad \Longleftrightarrow \qquad \forall i \in \{1, \dots, d\}, \quad X_n^i \xrightarrow[n\to+\infty]{\mathbb{P}} X^i.$$

▷ On a

$$|X_n - X|^2 = \sum_{i=1}^{d} |X_n^i - X^i|^2.$$

$\Rightarrow$ : $\forall i \in \{1, \dots, d\}$, $\mathbb{P}(|X_n^i - X^i| > \varepsilon) \leq \mathbb{P}(|X_n - X| > \varepsilon)$ d'où le résultat.

$\Leftarrow$ : $\mathbb{P}(|X_n - X| > \varepsilon) \leq \mathbb{P}\left(\bigcup_{i=1}^{d} \left\{|X_n^i - X^i| > \frac{\varepsilon}{\sqrt{d}}\right\}\right) \leq \sum_{i=1}^{d} \mathbb{P}\left(|X_n^i - X^i| > \frac{\varepsilon}{\sqrt{d}}\right) \xrightarrow[n\to+\infty]{} 0$
d'où le résultat. □

**Remarque :** Si $X_n \xrightarrow[n\to+\infty]{\mathbb{P}} X$ et $X_n \xrightarrow[n\to+\infty]{\mathbb{P}} Y$ alors $X = Y$ ps (unicité ps de la limite).

▷ Soit $\varepsilon > 0$,

$$\mathbb{P}(|X - Y| > \varepsilon) \leq \mathbb{P}\left(|X - X_n| > \frac{\varepsilon}{2}\right) + \mathbb{P}\left(|X_n - Y| > \frac{\varepsilon}{2}\right) \xrightarrow[n\to+\infty]{} 0$$

donc

$$\forall \varepsilon > 0, \qquad \mathbb{P}(|X - Y| > \varepsilon) = 0$$

*ie.*

$$\mathbb{P}(X \neq Y) = \mathbb{P}\left(\bigcup_{k\geq 1} |X - Y| > \frac{1}{k}\right) = 0$$

*ie.* $X = Y$ ps. □

**Proposition 6.12**

*Soit* $f : \mathbb{R}^d \to \mathbb{R}^p$ *continue. Si* $X_n \xrightarrow[n\to+\infty]{\mathbb{P}} X$ *alors* $f(X_n) \xrightarrow[n\to+\infty]{\mathbb{P}} f(X)$.

**Conséquence :** Si $X_n \xrightarrow[n\to+\infty]{\mathbb{P}} X$, $Y_n \xrightarrow[n\to+\infty]{\mathbb{P}} Y$, $\alpha_n \in \mathbb{R}$ $\alpha_n \xrightarrow[n\to+\infty]{} \alpha$ alors $\alpha_n X_n + Y_n \xrightarrow[n\to+\infty]{\mathbb{P}} \alpha X + Y$.
Si $X_n, Y_n \in \mathbb{R}$, $X_n Y_n \xrightarrow[n\to+\infty]{\mathbb{P}} XY$.

▷ On utilise l'uniforme continuité de $f$ sur les compacts. Soient $a > 0, \varepsilon > 0$. Il existe $\delta > 0$ tel que

$$\forall x, y \in \mathbb{R}^d, \qquad |x| \leq a, |x - y| \leq \delta \quad \Longrightarrow \quad |f(x) - f(y)| \leq \varepsilon.$$

donc

$$\{|X| \leq a\} \cap \{|X_n - X| \leq \delta\} \subset \{|f(X_n) - f(X)| \leq \varepsilon\}$$

d'où, par passage au complémentaire,

$$\{|f(X_n) - f(X)| > \varepsilon\} \subset \{|X| > a\} \cup \{|X_n - X| > \delta\}$$

donc

$$\mathbb{P}(|f(X_n) - f(X)| > \varepsilon) \leq \mathbb{P}(|X| > a) + \underbrace{\mathbb{P}(|X_n - X| > \delta)}_{\xrightarrow[n\to+\infty]{} 0}.$$

Alors,

$$\limsup_{n\to+\infty} \mathbb{P}(|f(X_n) - f(X)| > \varepsilon) \leq \mathbb{P}(|X| > a) \xrightarrow[a\to+\infty]{} 0$$

par le théorème de convergence monotone (car $X$ est à valeurs dans $\mathbb{R}^d$). Ainsi, $\mathbb{P}(|f(X_n) - f(X)| > \varepsilon) \xrightarrow[n\to+\infty]{} 0$ pour tout $\varepsilon > 0$ donc $f(X_n) \xrightarrow[n\to+\infty]{\mathbb{P}} f(X)$. □

**Proposition 6.13**

*Si* $X_n \to X$ *ps ou dans* $L^1$ *alors* $X_n \xrightarrow[n\to+\infty]{\mathbb{P}} X$.

▷ $(i)$ Supposons que $X_n \xrightarrow[n\to+\infty]{\text{p.s}} X$. Soit $\varepsilon > 0$.

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

Or $\mathbf{1}_{\{|X_n - X| > \varepsilon\}} \xrightarrow[n\to+\infty]{\text{p.s}} 0$ car $X_n \xrightarrow[n\to+\infty]{\text{p.s}} X$, $\mathbf{1}_{\{|X_n - X| > \varepsilon\}} \leq 1$ donc par le théorème de convergence dominée $\mathbb{P}(|X_n - X| > \varepsilon) \xrightarrow[n\to+\infty]{} 0$ pour tout $\varepsilon$ *ie.* $X_n \xrightarrow[n\to+\infty]{\mathbb{P}} X$.

$(ii)$ Supposons que $X_n \xrightarrow[n\to+\infty]{L^1} X$. Soit $\varepsilon > 0$. On sait que $\mathbb{E}[|X_n - X|] \xrightarrow[n\to+\infty]{} 0$. Par l'inégalité de Markov,

$$\mathbb{P}(|X_n - X| > \varepsilon) \leq \frac{\mathbb{E}[|X_n - X|]}{\varepsilon} \xrightarrow[n\to+\infty]{} 0$$

donc $X_n \xrightarrow[n\to+\infty]{\mathbb{P}} X$. □

**Exemple :** $X_n \sim \mathcal{B}(p_n)$ indépendantes, $X_n \in \{0, x_n\}$ avec $x_n \geq 1$. $X_n \xrightarrow[n\to+\infty]{\mathbb{P}} 0$ si $\forall \varepsilon > 0$, $\mathbb{P}(|X| > \varepsilon) \xrightarrow[n\to+\infty]{} 0$. Or $\varepsilon < 1$, $\mathbb{P}(|X_n| > \varepsilon) = p_n$. Prenons $p_n = \frac{1}{n}$. Alors $X_n \xrightarrow[n\to+\infty]{\mathbb{P}} 0$.

Avec $x_n = n$, $X_n$ ne converge pas ps et $X_n$ ne converge pas dans $L^1$ vers 0.

**Proposition 6.14**

*Si $(X_n)$ converge en probabilité vers $X$ alors il existe une sous-suite $(X_{n_k})$ qui converge ps vers $X$.*

$\triangleright$ Rappel : si $(\varepsilon_k)$ est une suite telle que $\sum \varepsilon_k < +\infty$ et $\sum \mathbb{P}(|X_{k+1} - X_k| > \varepsilon_k) < +\infty$ alors $X_k$ converge ps.

Pour tout $k \geq 0$, en prenant $\varepsilon = 2^{-k-1}$,

$$\lim_{n\to+\infty} \mathbb{P}(|X_n - X| > 2^{-k-1}) = 0$$

donc il existe $n_k$ tel que

$$\forall n \geq n_k, \qquad \mathbb{P}(|X_n - X| > 2^{-k-1}) \leq 2^{-k-1} \qquad (*).$$

On construit ainsi une suite $(n_k)$ que l'on peut supposer croissante car $(*)$ est vraie pour $n_k$ mais aussi pour $n_k + 1, \ldots$ Prenons $\varepsilon_k = 2^{-k}$. On a bien $\sum \varepsilon_k < +\infty$ et

$$\mathbb{P}(|X_{n+1} - X_{n_k}| > 2^{-k}) \leq \mathbb{P}(|X_{n_k+1} - X| > 2^{-k-1}) + \mathbb{P}(|X_{n_k} - X| > 2^{-k-1}) \leq 2^{-k}$$

qui est le terme général d'une série convergente. Donc $(X_{n_k})_{k\geq 0}$ converge presque sûrement.

Soit $Y$ la limite de $(X_{n_k})$. $X_{n_k} \xrightarrow[n\to+\infty]{\text{p.s}} Y$ donc $X_{n_k} \xrightarrow[n\to+\infty]{\mathbb{P}} Y$. Or $(X_{n_k})$ est une sous-suite de $(X_n)$ donc $X_{n_k} \xrightarrow[n\to+\infty]{\mathbb{P}} X$. Donc $Y = X$ ps. $\square$

**Remarque :** $X_n \xrightarrow[n\to+\infty]{\mathbb{P}} X$ si et seulement si de toute sous-suite $(X_{n_k})$ on peut extraire une sous-suite qui converge ps.

### 6.1.4 Critères de convergence (de type Cauchy)

**Proposition 6.15**

*Soit $(X_n)$ une suite de variables aléatoires à valeurs dans $\mathbb{R}^d$.*

*$(i)$ $(X_n)$ converge ps si et seulement si $\forall \varepsilon > 0$, $\lim_{n\to+\infty} \mathbb{P}(\sup_{k\geq 0} |X_{n+k} - X_n| > \varepsilon) = 0$.*

*$(ii)$ $(X_n)$ converge en probabilité si et seulement si $\forall \varepsilon > 0$, $\lim_{n\to+\infty} \sup_{k\geq 0} \mathbb{P}(|X_{n+k} - X_n| > \varepsilon) = 0$.*

*$(iii)$ $X_n \in L^p$. $(X_n)$ converge dans $L^p$ si et seulement si $\lim_{n\to+\infty} \sup_{k\geq 0} \mathbb{E}[|X_{n+k} - X_n|^p] = 0$.*

$\triangleright$ $(iii)$ Il s'agit de la complétude de $L^p$.

$(i) \Rightarrow$ : Supposons que $(X_n)$ converge presque sûrement. Presque sûrement, $(X_n)$ est une suite de Cauchy dans $\mathbb{R}^d$ : $\sup_{k\geq 0} |X_{n+k} - X_n| \xrightarrow[n\to+\infty]{} 0$ presque sûrement et donc en probabilité.

$\Leftarrow$ : On note $Y_n = \sup_{p,q\geq n} |X_p - X_q|$. $Y_n$ est une suite décroissante minorée par 0 donc $(Y_n)$ converge presque sûrement vers $Y$. Par ailleurs, $Y_n \leq 2\sup_{k\geq 0} |X_{n+k} - X_n|$ car $|X_p - X_q| \leq |X_p - X_n| + |X_n - X_q|$. Comme $\sup_{k\geq 0} |X_{n+k} - X_n| \xrightarrow[n\to+\infty]{\mathbb{P}} 0$ et $Y_n \xrightarrow[n\to+\infty]{\text{p.s}} Y$, $Y = 0$ ps. Ceci implique que $(X_n)$ est de Cauchy ps et donc $(X_n)$ converge ps ($\mathbb{R}^d$ complet).

$(ii) \Rightarrow \mathbb{P}(|X_{n+k} - X_n| > \varepsilon) \leq \mathbb{P}(|X_{n+k} - X| > \frac{\varepsilon}{2}) + \mathbb{P}(|X_n - X| > \frac{\varepsilon}{2})$ donc

$$\sup_{k\geq 0} \mathbb{P}(|X_{n+k} - X_n| > \varepsilon) \leq 2\sup_{k\geq 0} \mathbb{P}\left(|X_{n+k} - X| > \frac{\varepsilon}{2}\right).$$

Comme $\lim_{n\to+\infty} \mathbb{P}(|X_n - X| > \frac{\varepsilon}{2}) = 0$, on a

$$\lim_{n\to+\infty} \sup_{k\geq 0} \mathbb{P}\left(|X_{n+k} - X| > \frac{\varepsilon}{2}\right) = 0,$$

d'où le résultat.

$\Leftarrow$ Il existe une sous-suite $(X_{n_p})$ telle que

$$\sup_{k\geq 0} \mathbb{P}(|X_{n_p+k} - X_{n_p}| > 2^{-p}) \leq 2^{-p}.$$

En particulier,

$$\mathbb{P}(|X_{n_{p+1}} - X_{n_p}| > 2^{-p}) \leq 2^{-p}$$

donc $(X_{n_p})$ converge ps. Soit $X$ sa limite

$$\begin{aligned}
\mathbb{P}(|X_n - X| > \varepsilon) &\leq \mathbb{P}(|X_n - X_{n_p}| > \tfrac{\varepsilon}{2}) + \mathbb{P}(|X_{n_p} - X| > \tfrac{\varepsilon}{2}) \\
&\leq \sup_{k\geq 0} \mathbb{P}(|X_n - X_{n+k}| > \tfrac{\varepsilon}{2}) + \mathbb{P}(|X_{n_p} - X| > \tfrac{\varepsilon}{2}) \xrightarrow[n\to+\infty]{} 0.
\end{aligned}$$

□

**Remarque :** Notons $\mathcal{L}^0$ l'ensemble des variables aléatoires sur $(\Omega, \mathcal{F}, \mathbb{P})$ et $L^0 = \mathcal{L}^0/\sim$.

$$d(X,Y) = \inf\{c > 0,\ \mathbb{P}(|X - Y| > c) \leq c\}.$$

$d$ est une distance sur $L^0$ : la convergence pour la distance $d$ est la convergence en probabilité.

## 6.2 Convergence étroite et convergence en loi

### 6.2.1 Convergence étroite

Soit $(\mu_n)$ une suite de mesures de probabilité sur $\mathbb{R}^d$. Soit $\mu$ une probabilité sur $\mathbb{R}^d$. On note $\mathcal{C}_b$ l'ensemble des fonctions continues bornées $f : \mathbb{R}^d \to \mathbb{R}$ et $\mathcal{C}_c$ l'ensemble des fonctions continues à support compact $f : \mathbb{R}^d \to \mathbb{R}$.

**Définition 6.16**

*La suite $(\mu_n)$ converge étroitement vers $\mu$ si*

$$\forall f \in \mathcal{C}_b, \qquad \int_{\mathbb{R}^d} f(x)\mu_n(dx) \xrightarrow[n\to+\infty]{} \int_{\mathbb{R}^d} f(x)\mu(dx).$$

**Remarque :** La continuité de $f$ est essentielle. Par conséquent, on n'a pas forcément $\mu_n(B) \xrightarrow[n\to+\infty]{} \mu(B)$ pour $B \in \mathcal{B}(\mathbb{R}^d)$.

**Exemple :** $\mu = \delta_{\frac{1}{n}}$.

$$\int f \mathrm{d}\mu_n = f\left(\frac{1}{n}\right) \xrightarrow[n\to+\infty]{} f(0) = \int f\delta_0.$$

$B = \{0\}$ $\delta_n(B) = 0$ mais $\delta_0(B) = 1$.

**Proposition 6.17**

*Soit $H \subset \mathcal{C}_b$ tel que $\mathcal{C}_c \subset \overline{H}^{\|\cdot\|_\infty}$. Alors une suite de mesures de probabilité $(\mu_n)$ converge étroitement vers $\mu$ si et seulement si*

$$\forall f \in H, \qquad \int f \mathrm{d}\mu_n \xrightarrow[n\to+\infty]{} \int f \mathrm{d}\mu.$$

$\triangleright$ Si $\mu_n \to \mu$ alors $\int f \mathrm{d}\mu_n \to \int f \mathrm{d}\mu$ pour tout $f \in H$ car $f$ est bornée.

$(i)$ Si $H = \mathcal{C}_c$, on utilise un argument de troncature. Soit $\theta_k : \mathbb{R}^+ \to [0,1]$ telle que $\theta_k \equiv 1$ sur $[0,k]$, $\theta_k$ linéaire sur $[k,2k]$ et $\theta_k \equiv 0$ sur $[2k,+\infty[$. Soit $f \in \mathcal{C}_b$.

$$f(x) = f(x)\theta_k(|x|) + f(x)(1-\theta_k(|x|))$$

donc

$$\int f \mathrm{d}\mu_n = \int f\theta_k \mathrm{d}\mu_n + \int f(1-\theta_k)\mathrm{d}\mu_n$$

et

$$\int f \mathrm{d}\mu = \int f\theta_k \mathrm{d}\mu + \int f(1-\theta_k)\mathrm{d}\mu.$$

$$\begin{aligned}
\left|\int f \mathrm{d}\mu_n - \int f \mathrm{d}\mu\right| &\leq \left|\int f\theta_k \mathrm{d}\mu_n - \int f\theta_k \mathrm{d}\mu\right| + \left|\int f(1-\theta_k)\mathrm{d}\mu_n\right| + \left|\int f(1-\theta_k)\mathrm{d}\mu\right| \\
&\leq \left|\int f\theta_k \mathrm{d}\mu_n - \int f\theta_k \mathrm{d}\mu\right| + \|f\|_\infty \left(1 - \int \theta_k \mathrm{d}\mu_n\right) + \|f\|_\infty \left(1 - \int \theta_k \mathrm{d}\mu\right).
\end{aligned}$$

donc, comme $\theta_k, f\theta_k \in \mathcal{C}_c$,

$$\limsup_n \left|\int f \mathrm{d}\mu_n - \int f \mathrm{d}\mu\right| \leq 0 + 2\|f\|_\infty (1 - \theta_k \mathrm{d}\mu).$$

Par ailleurs, $\theta_k(x) \xrightarrow[k\to+\infty]{} 1$ et $|\theta_k(x)| \leq 1$ donc par le théorème de convergence dominée $\int \theta_k \mathrm{d}\mu \xrightarrow[k\to+\infty]{} 1$ donc

$$\forall f \in \mathcal{C}_b, \qquad \left|\int_f \mathrm{d}\mu_n - \int f \mathrm{d}\mu\right| \xrightarrow[n\to+\infty]{} 0$$

donc $\mu_n \to \mu$ étroitement.

$(ii)$ Cas général. Soit $f \in \mathcal{C}_c$. Soit $\varepsilon > 0$. Comme $\mathcal{C}_c \in \overline{H}$, il existe $h_\varepsilon \in H$ tel que $\|f - h_\varepsilon\| < \varepsilon$. Alors, en écrivant $f = h_\varepsilon + (f - h_\varepsilon)$, on a :

$$\left|\int f \mathrm{d}\mu_n - \int f \mathrm{d}\mu\right| \leq \left|\int h_\varepsilon \mathrm{d}\mu_n - \int h_\varepsilon \mathrm{d}\mu\right| + 2\varepsilon$$

Donc pour tout $f \in \mathcal{C}_c$ $\int f \mathrm{d}\mu_n \to \int f \mathrm{d}\mu$. Donc $\mu_n \to \mu$ étroitement. $\square$

cours 15 mars

**Lemme 6.18 (*Slutsky*)**

*Si $X_n \xrightarrow[n\to+\infty]{\mathcal{L}} X$ et $Y_n \xrightarrow[n\to+\infty]{\mathbb{P}} c$ où $c$ est une constante alors $(X_n, Y_n) \xrightarrow[n\to+\infty]{\mathcal{L}} (X, c)$.*

**Application :** Si $X_n \xrightarrow[n\to+\infty]{\mathcal{L}} X$ et $X_n - Y_n \xrightarrow[n\to+\infty]{\mathbb{P}} 0$ alors $Y_n \xrightarrow[n\to+\infty]{\mathcal{L}} X$. En effet, $Y_n = X_n + (Y - n - X_n)$ comme on a $(X_n, Y_n - X_n) \xrightarrow[n\to+\infty]{\mathcal{L}} (X, 0)$, par le lemme de Slutsky on a bien $Y_n \xrightarrow[n\to+\infty]{\mathcal{L}} X + 0 = X$.

▷ On a

$$\varphi_{(X,c)}(s,t) = \mathbb{E}[e^{i(sX+tc)}] = e^{itc}\varphi_X(s).$$

Donc

$$\begin{aligned}
|\varphi_{(X_n,Y_n)}(s,t) - \varphi_{(X,c)}(s,t)| &= \left|\mathbb{E}[e^{isX_n}e^{itY_n}] - e^{itc}\mathbb{E}[e^{isX_n}] - e^{itc}\mathbb{E}[e^{isX_n}] - e^{itc}\mathbb{E}[e^{isX}]\right| \\
&\le \left|\mathbb{E}[e^{isX_n}(e^{itY_n} - e^{itc})]\right| + \left|e^{itc}(\mathbb{E}[e^{isX_n}] - \mathbb{E}[e^{isX}])\right| \\
&\le \underbrace{\mathbb{E}[|e^{itY_n} - e^{itc}|]}_{\le \varepsilon|t| + 2\mathbb{P}(|Y_n - c| > \varepsilon)\ \forall\varepsilon \text{ cf preuve prec.}} + \underbrace{|\varphi_{X_n}(s) - \varphi_X(s)|}_{\xrightarrow[n\to+\infty]{} 0 \text{ car } X_n \xrightarrow{L} X}.
\end{aligned}$$

On en déduit que

$$\forall s,t \in \mathbb{R}, \qquad \varphi_{(X_n,Y_n)}(s,t) \xrightarrow[n\to+\infty]{} \varphi_{(X,c)}(s,t).$$

Donc par le théorème de Paul-Lévy

$$(X_n, Y_n) \xrightarrow[n\to+\infty]{\mathcal{L}} (X, c).$$

$\square$

**Exemple :** Soit $(X_n)$ une suite de variables aléatoires telle que $X_n$ suit la loi géométrique $\mathcal{G}(\frac{\lambda}{n})$ avec $\lambda > 0$. Convergence en loi de $X_n$ de $n \to +\infty$ ? Pour $t > 0$,

$$\varphi_{X_n}(t) = \frac{\frac{\lambda}{n}e^{it}}{1 - (1 - \frac{\lambda}{n})e^{it}} \xrightarrow[n\to+\infty]{} 0$$

mais $\varphi_{X_n}(0) = 1$ donc $X_n$ ne converge pas en loi.

Regardons $Y_n = \frac{X_n}{n}$

$$\varphi_{Y_n}(t) = \mathbb{E}[e^{itY_n}] = \mathbb{E}[e^{i\frac{t}{n}X_n}] = \varphi_{X_n}\left(\frac{t}{n}\right) = \frac{\frac{\lambda}{n}e^{i\frac{t}{n}}}{1 - (1 - \frac{\lambda}{n})e^{i\frac{t}{n}}} = \frac{\lambda e^{i\frac{t}{n}}}{n - (n - \lambda)e^{i\frac{t}{n}}} = \frac{\lambda e^{i\frac{t}{n}}}{n(1 - e^{i\frac{t}{n}}) + \lambda e^{i\frac{t}{n}}}$$

donc

$$\varphi_{Y_n}(t) \xrightarrow[n\to+\infty]{} \frac{\lambda}{\lambda - it}$$

(fonction caractéristique de la loi exponentielle $\mathcal{E}(\lambda)$) Ainsi,

$$Y_n = \frac{X_n}{n} \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{E}(\lambda).$$

## 6.3 Conclusion

$$\begin{array}{ccccccc}
 & & \text{cv dans } L^1 & \Leftarrow & \text{cv dans } L^p,\ p\geq 1 & \Leftarrow & \text{cv dans } L^\infty \\
 & & \Downarrow & & & & \\
\text{cv ps} & \begin{array}{c}\Rightarrow \\ \underset{\exists \text{ss suite qui cv ps}}{\Leftarrow}\end{array} & \text{cv en proba} & \begin{array}{c}\Rightarrow \\ \underset{\text{si limite=cste}}{\Leftarrow}\end{array} & \text{cv en loi} & &
\end{array}$$

Si on a convergence en probabilité en uniforme intégrabilité *ie.*

$$\lim_{n\to+\infty} \sup_{a>0} \mathbb{E}[|X_n|\mathbf{1}_{|X_n|>a}] = 0$$

alors on a convergence dans $L^1$.

# Chapitre 7

# Théorèmes limites

On va considérer une suite $(X_n)_{n\geq 1}$ de variables aléatoires indépendantes identiquement distribuées (de même loi).

## 7.1 Loi des grands nombres

### 7.1.1 Version faible de la loi des grands nombres

**Théorème 7.1**

*Soit $(X_n)$ une suite de variables aléatoires indépendantes identiquement distribuées ayant un moment d'ordre 1 fini : $\mathbb{E}[|X_1|] < +\infty$. Alors,*

$$\frac{X_1 + \cdots + X_n}{n} \xrightarrow[n\to+\infty]{\mathbb{P}} \mathbb{E}[X_1].$$

**Application :** Si les variables aléatoires $X_n$ suivent la loi $\mathcal{B}(p)$, alors $\frac{X_1+\cdots+X_n}{n}$ représente le nombre moyen de succès sur $n$ expériences. D'après la loi des grands nombres, on a

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

$\triangleright$ $(i)$ Preuve dans le cas $L^2$. On suppose les $X_n$ dans $L^2$. On note $m = \mathbb{E}[X_1]$ et $\sigma^2 = \mathrm{Var}(X_1)$. On veut montrer que

$$\forall \varepsilon > 0, \qquad \mathbb{P}(|M_n - m| > \varepsilon) \xrightarrow[n\to+\infty]{} 0$$

où $M_n = \frac{X_1+\cdots+X_n}{n}$. Or

$$\mathbb{E}[M_n] = m$$

et, comme les variables sont indépendantes

$$\mathrm{Var}(M_n) = \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}(X_i) = \frac{\sigma^2}{n}.$$

Soit $\varepsilon > 0$. Par le théorème de Bienaymé-Tchebychev

$$\mathbb{P}(|M_n - m| > \varepsilon) \leq \frac{\mathrm{Var}(M_n)}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2} \xrightarrow[n\to+\infty]{} 0$$

donc $M_n \xrightarrow[n\to+\infty]{\mathbb{P}} m$.

$(ii)$ Preuve dans le cas $L^1$. Il suffit de montrer que $M_n \xrightarrow[n\to+\infty]{\mathcal{L}} m$. On étudie la fonction caractéristique de $M_n$.

$$M_n = \frac{X_1 + \cdots + X_n}{n} \xrightarrow[n\to+\infty]{} m \qquad \Longleftrightarrow \qquad \frac{(X_1 - m) + \cdots + (X_n - m)}{n} \xrightarrow[n\to+\infty]{} 0$$

donc on peut supposer que les variables $X_i$ sont centrées. Notre but de alors de montrer que

$$\varphi_{M_n}(t) \xrightarrow[n\to+\infty]{} 1.$$

Comme les variables sont indépendantes,

$$\varphi_{M_n}(t) = \prod_{k=1}^{n} \mathbb{E}\left[e^{i\frac{X_k}{n}t}\right] = \prod_{k=1}^{n} \varphi_{X_1}\left(\frac{t}{n}\right) = \varphi_{X_1}\left(\frac{t}{n}\right)^n.$$

**Lemme 7.2**

*Pour tout $x \in \mathbb{R}$,*

$$\left|e^{ix} - \sum_{k=0}^{p} \frac{(ix)^k}{k!}\right| \leq \min\left(\frac{|x|^{p+1}}{(p+1)!}, \frac{2|x|^p}{p!}\right).$$

On en déduit que

$$\left|\mathbb{E}\left[e^{itX}\right] - \sum_{k=0}^{p} \frac{i^k t^k \mathbb{E}[X^k]}{k!}\right| \leq \mathbb{E}\left[\min\left(\frac{|t|^{p+1}|X|^{p+1}}{(p+1)!}, \frac{2|t|^p|X|^p}{p!}\right)\right].$$

**Lemme 7.3**

*$\forall a_i, b_i \in \mathbb{C}$ avec $\forall i$, $|a_i| \leq 1$ et $|b_i| \leq 1$*

$$\left|\prod_{i=1}^{n} a_i - \prod_{i=1}^{n} b_i\right| \leq \sum_{i=1}^{n} |a_i - b_i|$$

▷ (premier lemme) – En utilisant le développement de Taylor avec reste intégral,

$$e^{ix} - \sum_{k=0}^{p} \frac{(ix)^k}{k!} = \frac{i^{p+1}}{p!} \int_0^x (x-s)^p e^{is} \mathrm{d}s$$

donc

$$\left|e^{ix} - \sum_{k=0}^{p} \frac{(ix)^k}{k!}\right| \leq \frac{1}{p!} \left|\int_0^x (x-s)^p \mathrm{d}s\right| = \frac{|x|^{p+1}}{(p+1)!}.$$

– De même,

$$e^{ix} - \sum_{k=0}^{p} \frac{(ix)^k}{k!} = \frac{i^p}{(p-1)!} \int_0^x (x-s)^{p-1} e^{is} \mathrm{d}s - \frac{i^p x^p}{p!}$$

donc

$$\left|e^{ix} - \sum_{k=0}^{p} \frac{(ix)^k}{k!}\right| \leq \frac{1}{(p-1)!}\left|\int_0^x (x-s)\mathrm{d}s\right| + \frac{|x|^p}{p!} \leq \frac{2|x|^p}{p!}$$

□

▷ (second lemme) Preuve par récurrence.
- pour $n = 1$ : trivial
- si c'est vrai au rang $n$

$$\prod_{i=1}^{n+1} a_i - \prod_{i=1}^{n+1} b_i = a_{n+1}\left(\prod_{i=1}^{n} a_i - \prod_{i=1}^{n} b_i\right) + \prod_{i=1}^{n} b_i(a_{n+1} - b_{n+1})$$

donc

$$\left|\prod_{i=1}^{n+1} a_i - \prod_{i=1}^{n+1} b_i\right| \leq \left|\prod_{i=1}^{n} a_i - \prod_{i=1}^{n} b_i\right| + |a_{n+1} - b_{n+1}|$$

et on conclut par hypothèse de récurrence. □

On peut achever la preuve du théorème.

$$|\varphi_{M_n}(t) - 1| = \left|\varphi_X\left(\frac{t}{n}\right)^n - 1\right| \leq \sum_{i=1}^{n}\left|\varphi_X\left(\frac{t}{n}\right) - 1\right| = n\left|\varphi_X\left(\frac{t}{n}\right) - 1\right|$$

Or, comme $\mathbb{E}[X] = 0$,

$$\begin{aligned}
n\left|\varphi_X\left(\tfrac{t}{n}\right) - 1\right| &= n\left|\mathbb{E}\left[e^{iX\frac{t}{n}}\right] - 1\right| \\
&= n\left|\mathbb{E}\left[e^{iX\frac{t}{n}}\right] - 1 - \tfrac{it}{n}\mathbb{E}[X]\right| \\
&\leq n\mathbb{E}\left[\min\left(\tfrac{|t|^2|X|^2}{2n^2}, \tfrac{2|t||X|}{n}\right)\right] \\
&\leq \mathbb{E}\left[\min\left(\tfrac{|t|^2|X|^2}{2n}, 2|t||X|\right)\right]
\end{aligned}$$

Par le théorème de convergence dominée, on a

$$n\left|\varphi_X\left(\frac{t}{n}\right) - 1\right| \xrightarrow[n\to+\infty]{} 0$$

donc $\varphi_{M_n}(t) \xrightarrow[n\to+\infty]{} 1$. Ainsi $M_n \xrightarrow[n\to+\infty]{\mathcal{L}} 0$ et donc $M_n \xrightarrow[n\to+\infty]{\mathbb{P}} 0$. □

### 7.1.2 Version forte de la loi des grands nombres

**Théorème 7.4**

*Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires indépendantes identiquement distribuées dans*

*$L^1$. Alors*

$$\frac{X_1 + \cdots + X_n}{n} \xrightarrow[n \to +\infty]{p.s} \mathbb{E}[X_1].$$