# Probabilités

Pierron Théo

ENS Ker Lann

# Table des matières

# Chapitre 1

# Notions de base

## 1.1 Espaces de probabilité, langage probabiliste

### 1.1.1 Introduction

**Définition 1.1** Soit $\Omega$ l'ensemble des résultats possibles d'une expérience.

On appelle résultats les éléments de $\Omega$.

Les sous-ensembles de $\Omega$ sont des évènements.

$\mathcal{P}(\Omega)$ est l'ensemble des évènements si $\Omega$ est au plus dénombrable. Sinon, on va prendre comme famille d'évènements une tribu $\mathcal{A}$ sur $\Omega$.

À tout évènement $A$, on va associer un nombre $P(A)$ où $P$ est une mesure finie de masse 1.

On appelle alors espace de probabilité le triplet $(\Omega, \mathcal{A}, P)$.

### 1.1.2 Rappels

**Définition 1.2** Une classe $A \in \mathcal{P}(\mathcal{P}(E))$ est dite algèbre (de Boole) ssi :

- $\varnothing \in A$
- $a \in A \Rightarrow a^c \in A$
- $a_1, a_2 \in A^2 \Rightarrow a_1 \cup a_2 \in A$

**Définition 1.3** Une classe de parties $A \subset \mathcal{P}(E)$ est dite une tribu ($\sigma$-algèbre de Boole) ssi :

- $\varnothing \in A$
- $a \in A \Rightarrow a^c \in A$
- $(a_i)_i \in A^{\mathbb{N}} \Rightarrow \bigcup\limits_{i \geqslant 0} a_i \in A$

**<u>Définition 1.4</u>** Une classe de parties $A \subset \mathcal{P}(E)$ est dite classe monotone ssi :

- $\varnothing \in A$
- $(a_i)_i \in A^{\mathbb{N}}$ croissante $\Rightarrow \bigcup\limits_{i \geqslant 0} a_i \in A$
- $(a_i)_i \in A^{\mathbb{N}}$ décroissante $\Rightarrow \bigcap\limits_{i \geqslant 0} a_i \in A$

**Proposition 1.1**

- Une algèbre est une tribu ssi c'est aussi une classe monotone
- Soit $\mathcal{E} \subset \mathcal{P}(\Omega)$. Il existe une unique tribu (resp. classe monotone) minimale pour l'inclusion qui contient $\mathcal{E}$.
  On l'appelle tribu (resp. classe monotone) engendrée par $\mathcal{E}$ notée $\sigma(\mathcal{E})$ (resp. $M(\mathcal{E})$).
- Si $A \subset \Omega$, $\sigma(A) = \{\varnothing, A, A^c, \Omega\}$.
- Si $\mathcal{E}$ est une algèbre, $\sigma(\mathcal{E}) = M(\mathcal{E})$.
- Si $\mathcal{E}$ est une algèbre et $\mathcal{C} \subset \mathcal{P}(\Omega)$ contenant $\mathcal{E}$ et stable par union croissante et intersection décroissante, alors $\mathcal{C} \supset \Sigma(\mathcal{E})$.

**<u>Définition 1.5</u>** Soit $\mathcal{A}$ une tribu sur $\Omega$ et $P : A \to \mathbb{R}$.

$P$ est une probabilité ssi :

- $\forall A \in \mathcal{A}$, $P(A) \geqslant 0$.
- Si $(A_j)_j \in \mathcal{A}^{\mathbb{N}}$ sont disjoints (ie incompatibles), alors $P\left(\bigcup\limits_{j \geqslant 0} A_j\right) = \sum\limits_{j=0}^{\infty} P(A_j)$.
- $P(\Omega) = 1$

**Proposition 1.2**

- $P(\varnothing) = 0$.
- $P(A) \leqslant 1$ pour tout $A$.
- $P(A^c) = 1 - P(A)$.
- $P(A \cup B) + P(A \cap B) = P(A) + P(B)$.
- Si $A \subset B$, $P(A) = P(B) - P(B \setminus A)$.
- Si $A = \bigcup\limits_{j \geqslant 0} A_j$ avec $(A_j)_j$ croissante, ou $A = \bigcap\limits_{j \geqslant 0} A_j$ avec $(A_j)_j$ décroissante, alors $\lim\limits_{n \to +\infty} P(A_n) = P(A)$.
- $P\left(\bigcup\limits_{j \geqslant 0} A_j\right) \leqslant \sum\limits_{j=0}^{\infty} P(A_j)$.
- La $\sigma$-additivité est équivalente à l'additivité et (si $(A_n)_n$ est décroissante, $\bigcap\limits_{n \geqslant 0} A_n = \varnothing \Rightarrow \lim\limits_{n \to +\infty} P(A_n) = 0$).

 

- Soit $\mathcal{E}$ une algèbre et $p : \mathcal{E} \to \mathbb{R}$ positive, de masse 1, additive et vérifiant si $(A_n)_n$ est décroissante, $\bigcap_{n \geqslant 0} A_n = \varnothing \Rightarrow \lim_{n \to +\infty} p(A_n) = 0$, alors il existe une unique extension de $p$ à une probabilité $P$ sur $\sigma(\mathcal{E})$.

*Démonstration.*

- $\varnothing = \bigcup_{j \geqslant 0} \varnothing$ donc $P(\varnothing) = \sum_{j=0}^{\infty} P(\varnothing)$ donc $P(\varnothing) = 0$.
- Clair
- $\Omega = A \cup A^c$ donc $1 = P(A) + P(A^c)$.
- $A \cup B = (A \setminus (A \cap B)) \cup B$ donc $P(A \cup B) = P(A) - P(A \cap B) + P(B)$.
- Si $A \subset B$, $B = A \cup (B \setminus A)$ et $A \cap (B \setminus A) = \varnothing$ donc $P(B) = P(A) + P(B \setminus A)$ d'où le résultat.
- cf. INTP
- idem
- Un sens est clair via l'INTP. Soit $(A_j)_j$ une suite de parties disjointes de $\mathcal{A}$.

  $\bigcup_{k \geqslant n+1} A_k$ décroit vers $\varnothing$ donc, par hypothèse, $P\left(\bigcup_{k \geqslant n+1} A_k\right)$ tend vers 0.
  On a donc :

$$\begin{aligned} P\left(\bigcup_{j \geqslant 0} A_j\right) &= P\left(\left(\bigcup_{j=1}^{n} A_j\right) \cup \left(\bigcup_{k \geqslant n+1} A_k\right)\right) \\ &= P\left(\bigcup_{j=1}^{n} A_j\right) + P\left(\bigcup_{k \geqslant n+1} A_k\right) \\ &= \sum_{j=1}^{n} P(A_j) + P\left(\bigcup_{k \geqslant n+1} A_k\right) \end{aligned}$$

  Avec $n \to +\infty$, on a bien la $\sigma$-additivité.
- Admis ∎

**Exemple 1.1**

- $\Omega = \{\omega_j, j \geqslant 1\}$ dénombrable, $\mathcal{A} = \mathcal{P}(\Omega)$.
  Soit $(p_j)_j$ une suite de réels tels que $p_j \geqslant 0$, $\sum_{j \geqslant 1} p_j = 1$. On définit, pour $A \in \mathcal{P}(\Omega)$, $P(A) = \sum_{\{j, \omega_j \in A\}} p_j$.
  C'est une probabilité sur $(\Omega, \mathcal{A})$.
- Soit $f$ positive intégrable pour la mesure de Lebesgue et d'intégrale 1. Pour tout borélien $B$, on pose $P(B) = \int_B f(x)\lambda(\mathrm{d}x)$.

 

$P$ est une probabilité sur $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$.

- Soit $F : \mathbb{R} \to [0,1]$ croissante telle que $F(+\infty) = 1$ et $F(-\infty) = 0$. On définit une probabilité sur les unions finies disjointes d'intervalles semi-ouverts par :

$$P_F(]a,b]) = F(b) - F(a) \text{ et } P(\varnothing) = 0$$

$$P_F(]-\infty, b]) = F(b) \text{ et } P_F(]a,+\infty[) = A - F(a)$$

On a de plus :

<u>Théorème 1.1</u> de Lebesgue-Stieltjes *$P_F$ est positive, additive et de masse 1. De plus, pour tout $\forall (A_n)_n$ décroissant vers $\varnothing$, $\lim\limits_{n\to+\infty} P_F(A_n) = 0$ ssi $F$ est continue à droite.*

*Ainsi, si $F$ est une fonction continue à droite, croissante vérifiant les conditions $F(+\infty) = 1$ et $F(-\infty) = 0$, il existe une unique probabilité $P_F$ sur $\mathscr{B}(\mathbb{R})$ telle que $F(x) = P_F(]-\infty, x])$. Inversement, chaque probabilité sur $\mathscr{B}(\mathbb{R})$ définit une telle fonction. Elle est appelée fonction de répartition.*

*Démonstration.*

$\Rightarrow$ On a vu que si $t_n \to t^+$, alors $\bigcap\limits_{n\geqslant 1} ]-\infty, t_n] = ]-\infty, t]$ donc $F(t_n) = P_F(]-\infty, t_n[) \to P_F(]-\infty, t]) = F(t)$.

$\Leftarrow$ Dans l'autre sens, on remarque que les trois premières propriétés sont vraies par définition de $P_F$. Soit donc une suite décroissante $(A_n)_n$ de limite $\varnothing$.

Notons $A_n = \bigcup\limits_{j=1}^{k_n} ]a_{j,n}, b_{j,n}]$. Soit $\delta > 0$.

Il existe $R > 0$ tel que $F(-R) \leqslant \frac{\delta}{4}$ et $1 - F(R) \leqslant \frac{\delta}{4}$.

Par continuité à droite de $F$ pour tout $n, j$, il existe $c_{j,n} \in ]a_{j,n}, b_{j,n}]$ tel que $F(c_{j,n}) - F(a_{j,n}) \leqslant \frac{\delta}{2^{j+n+1}}$.

Posons de plus $B_n = \bigcup\limits_{i=1}^{k_n} (]c_{j,n}, b_{j,n}] \cap ]-R, R])$ et $C_n = \bigcap\limits_{m=1}^{n} B_m$.

On a $B_n \in \mathcal{A}$, $B_n \subset A_n$, $C_n \in \mathcal{A}$ et $C_n \subset A_n$.

De plus, $A_n \setminus C_n \subset \bigcup\limits_{m=1}^{n} (A_m \setminus B_m)$ donc :

$$A_n \cap \left( \bigcup_{n=1}^{m} B_m^c \right) = \bigcup_{m=1}^{n} (A_n \cap B_n^c)$$

 

D'où :

$$\begin{aligned}
P_F(A_n) - P_F(C_n) &\leqslant P_F(]-R,R]^c) + \sum_{m=1}^{n} P_F((A_m \setminus B_m) \cap ]-R,R]) \\
&\leqslant P_F(]-R,R]^c) + \sum_{m=1}^{n}\sum_{j=1}^{k_m} P_F(]c_{j,m}, b_{j,m}]) \\
&= F(-R) + 1 - F(R) + \sum_{m=1}^{n}\sum_{j=1}^{k_m} (F(b_{j,m}) - F(c_{j,m})) \\
&\leqslant \frac{\delta}{2} + \sum_{m=1}^{\infty}\sum_{j=1}^{k_m} \frac{\delta}{2^{m+j+1}} = \delta
\end{aligned}$$

$\overline{C_n}$ est borné donc compact. Il est de plus inclus dans $A_n$ donc la suite des $\overline{C_n}$ converge vers $\varnothing$ donc il existe $m$ tel que pour $n \geqslant m$, $\overline{C_n} = \varnothing$ donc $\forall n \geqslant m, C_n = \varnothing$. Donc $P_F(C_n) = 0$ pour tout $n \geqslant m$.
Donc il existe $m \geqslant 1$ tel que pour tout $n \geqslant m$, $P_F(A_n) = P_F(A_n) - P_F(C_n) \leqslant \delta$ donc $\lim\limits_{n \to +\infty} P_F(A_n) = 0$. ■

*Remarque 1.1* *Si $\Omega$ est au plus dénombrable, pour construire une proba, on doit disposer d'une suite $p \in [0,1]^{\mathbb{N}}$ tels que $\sum_{n=1}^{\infty} p_n = 1$. Par exemple, pour un lancer de dé à $n$ faces non pipé, on pose $p_1 = \ldots = p_n = \frac{1}{n}$ et $p_i = 0$ pour $i > n$.*

*Sur $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$, on peut définir une probabilité à l'aide d'une fonction borélienne d'intégrale 1.*

*Soit $(\Omega, \mathcal{A})$ un espace mesurable. Si $\Omega = \{\omega_i, i \in \mathbb{N}\}$ et si $p$ est une suite positive telle que $\sum_{i=0}^{\infty} p_i = 1$, alors une probabilité sur $\Omega$ est donnée par : $P = J \mapsto \sum_{j \in J} p_j \delta_{\omega_j}(J)$.*

**Définition 1.6** Soit $(\Omega, \mathcal{A}, \mathcal{P})$ un espace de probabilité.

On dit que $N \subset \Omega$ est négligeable ssi il existe $B \in \mathcal{A}$ tel que $N \subset B$ et $P(B) = 0$.

On note $\mathcal{N}$ la famille des négligeables et $\mathcal{A}^C = \sigma(\mathcal{A} \cup \mathcal{N})$ la tribu complétée.

THÉORÈME 1.2 *Il existe une unique extension de $P$ à $P^C$ sur $\mathcal{A}^C$. L'espace $(\Omega, \mathcal{A}^C, P^C)$ est alors appelé espace complet. Par la suite, tous les espaces de probabilités seront complets.*

**Définition 1.7** Un évènement $A$ est dit presque sûr ssi $A^c$ est négligeable.

 

## 1.2 Variables aléatoires, lois

### 1.2.1 Définitions

Soit $(\Omega, \mathcal{A}, P)$ un espace de probabilités et $X : \Omega \to \mathbb{R}^d$ muni de $\mathscr{B}(\mathbb{R}^d)$.

**Définition 1.8** $X$ est une variable aléatoire $d$-dimesionnelle ssi $X$ est mesurable.

On notera $\{X \in B\} = X^{-1}(B)$.

**Exemple 1.2** Pour $A \subset \Omega$, $1_A$ est une variable aléatoire ssi $A \in \mathcal{A}$.

**Proposition 1.3** $X$ est une variable aléatoire réelle ssi $\forall t, \{X \leqslant t\} \in \mathcal{A}$.

*Démonstration.* $\{B \in \mathscr{B}(\mathbb{R}), X^{-1}(B) \in \mathcal{A}\}$ est une tribu qui contient une famille de générateurs de $\mathscr{B}(\mathbb{R})$ donc elle contient $\mathscr{B}(\mathbb{R})$. ■

**Proposition 1.4** La tribu engendrée par une variable aléatoire réelle $X : \Omega \to \mathbb{R}^d$ est la plus petite tribu sur $\Omega$ qui rendre mesurable l'application $X$.

**Proposition 1.5** Chaque variable aléatoire réelle sur $(\Omega, \mathcal{A}, P)$ induit un espace de probabilité avec la probabilité image $Q = X(P)$ définie par $Q(B) = P(X \in B)$.

*Démonstration.* $Q(B) \geqslant 0$, $Q(\mathbb{R}^d) = P(\Omega) = 1$.

Si les $B_j$ sont disjoints,

$$\begin{aligned} Q\left(\bigcup_{j=1}^{\infty} B_j\right) &= P\left(X^{-1}(\left(\bigcup_{j=1}^{\infty} B_j\right)\right) \\ &= P\left(\bigcup_{j=1}^{\infty} X^{-1}(B_j)\right) \\ &= \sum_{j=1}^{\infty} P(X^{-1}(B_j)) \\ &= \sum_{j=1}^{\infty} Q(B_j) \end{aligned}$$

■

**Définition 1.9** Soit $X$ une variable aléatoire réelle sur $(\Omega, \mathcal{A}, P)$. La loi de $X$ est la probabilité $Q$ de la proposition précédente qui sera notée $P_X : \mathscr{B}(\mathbb{R}^d) \to [0, 1]$.

**Définition 1.10** La fonction de répartition de la variable aléatoire réelle $X$ est la fonction de répartition associée à la probabilité $P_X$ :

$$F_X : \begin{cases} \mathbb{R} & \to & [0,1] \\ t & \mapsto & P_X(]-\infty, t]) \end{cases}$$

 

*Remarque 1.2* *Si on connaît $X$, on peut déterminer $P_X$ donc $F_X$. En revanche, si on connait $F_X$, on ne peut pas trouver $X$.*

<u>Théorème 1.3</u> *La fonction de répartiton caractérise la loi de la variable aléatoire réelle, ie $F_X = F_Y \Rightarrow P_X = P_Y$.*

*Démonstration.* $F_X = F_Y$ donc $P_X(]-\infty, t]) = P_Y(]-\infty, t])$ donc $P_X(]a, b]) = P_Y(]a, b])$ pour tous $a < b$.

Donc $P_X$ et $P_Y$ sont égales sur les intervalles.

$$\mathcal{M}_1 = \{B \in \mathscr{B}(\mathbb{R}), P_X(B) = P_Y(B)\}$$

est une classe monotone (par continuité de $P_X$ et $P_Y$) contenant les intervalles dont vaut $\mathscr{B}(\mathbb{R})$. ∎

**Proposition 1.6** Soit $X$ une variable aléatoire réelle et soit $g : \mathbb{R}^d \to \mathbb{R}^k$ borélienne. Alors $g(X)$ est une variable aléatoire $k$-dimensionnelle.

Corollaire 1.1

- *Soit $X$ une variable aléatoire réelle et $g : \mathbb{R} \to \mathbb{R}$ continue. $g(X)$ est une variable aléatoire réelle donc on a :*
  $X^r$, $|X|^r$, $\mathrm{e}^{-\lambda X}$, $\mathrm{e}^{itX}$.
- $X \vee Y = \max\{X, Y\}$, $X \wedge Y$, $X + Y$, $X - Y$, $XY$, $\frac{X}{Y}$ *sont des variables aléatoires avec $Y$ une autre variable aléatoire réelle*
- *Soit $(X_n)_n$ une suite de variables aléatoires. Alors :*

  $$\sup_n X_n, \quad \inf_n X_n, \quad \limsup_n X_n, \quad \liminf X_n$$

  *sont des variables aléatoires et si $\lim\limits_n X_n$ existe, alors, c'est une variable aléatoire réelle*

<u>**Définition 1.11**</u> Une variable aléatoire réelle est dite simple ssi elle est étagée.

**Proposition 1.7** Toute variable aléatoire réelle $X$ est limite simple de d'uine suite de variable aléatoire réelle simples. Si $X$ est positive, alors les termes de la suite sont positives et la suite peut être choisie croissante.

**Exemple 1.3** Une variable aléatoire réelle est discrète ssi il existe $B \subset \mathbb{R}^d$ au plus dénombrable telle que $P(X \in B) = 1$. Cela signifie que $X$ prend un nombre au plus dénombrable de valeurs avec une probabilité positive.

On a donc $X : \Omega \to \{x_i, i \in \mathbb{N}\}$.

Posons $p_j = P(X = x_j) = P_X(\{x_j\})$. On a donc $P_X = \sum\limits_{j \geqslant 1} p_j \delta_{x_j}$.

Si on suppose les $x_i$ croissants, le graphe de la fonction est un escalier croissant nul jusqu'à $x_1$ et qui est constant sur tous les $[x_i, x_{i+1}[$.

 

**Proposition 1.8** Soit $g : \{x_i\} \to \{y_j\}$ surjective.
Soit $Y = g(X)$ une variable aléatoire discrète.

$$P(Y = y_l) = \sum_{j, g(x_j) = y_l} P(X = x_j)$$

*Démonstration.*

$$\begin{aligned} P(Y = y_l) &= P(\{\omega, Y(\omega) = y_l\}) \\ &= P(\{\omega, g(X(\omega)) = y_l\}) \\ &= P\left(\bigcup_{j, g(x_j) = y_l} \{\omega, X(\omega) = x_j\}\right) \\ &= \sum_{j, g(x_j) = y_l} P(X = x_j) \end{aligned}$$

**Exemple 1.4**

- Loi de Bernoulli ($B(1, p)$) :

$$\begin{pmatrix} 0 & 1 \\ 1 - p & p \end{pmatrix}$$

(On tire 1 avec une probabilité $p$ et 0 avec $1 - p$)

- Loi binômiale ($B(n, p)$)

$$\begin{pmatrix} 0 & 1 & \dots & k & \dots & n \\ (1 - p)^n & np(1 - p)^{n-1} & \dots & \binom{n}{k} p^k (1 - p)^{n-k} & \dots & p^n \end{pmatrix}$$

- Loi géométrique $\mathcal{G}(p)$ :

$$\begin{pmatrix} 1 & \dots & k & \dots \\ p & \dots & p(1 - p)^{k-1} & \dots \end{pmatrix}$$

- Loi uniforme :

$$\begin{pmatrix} 1 & \dots & n \\ \frac{1}{n} & \dots & \frac{1}{n} \end{pmatrix}$$

- Loi de Poisson :

$$\begin{pmatrix} 0 & 1 & \dots & k & \dots \\ e^{-\lambda} & \lambda e^{-\lambda} & \dots & e^{-\lambda} \frac{\lambda^k}{k!} & \dots \end{pmatrix}$$

 

### 1.2.2 Variables aléatoires à densité

**Définition 1.12** Soit $X$ un variable aléatoire.

On note $P_X(B) = \int_B f_X(x)\,\mathrm{d}x$. Quand $f_X \geqslant 0$ est borélienne telle que $\int_{\mathbb{R}} f_X(x)\,\mathrm{d}x$, $P_X$ est une probabilité.

$f_X$ est la densité de probabilité de la loi $P_X$ (par rapport à la mesure de Lebesgue).

On a $P_X(]-\infty, t]) = F_X(t) = \int_{-\infty}^t f_X(x)\,\mathrm{d}x$.

**Proposition 1.9** $F_X$ est dérivable pp et $F_X' = f_X$.

**Exemple 1.5**

- Variable aléatoire de la loi gaussienne standard $X \sim \mathcal{N}(0,1)$.
  La densité de $X$ est $f_X(x) = \frac{\mathrm{e}^{-\frac{x^2}{2}}}{\sqrt{2\pi}}$.
  On notera $\Phi(t) = F_X(t) = \int_{-\infty}^t f_X(x)\,\mathrm{d}x$ la fonction de Laplace.
  On a $\Phi(-t) = 1 - \Phi(t)$.
  De plus,

$$\begin{aligned} P(-X \leqslant t) &= P(X > -t) = 1 - P(X \leqslant -t) \\ &= 1 - \Phi(-t) = \Phi(t) = P(X \leqslant t) \end{aligned}$$

  Donc $X$ et $-X$ ont même loi : $X \sim -X$.
- Variable aléatoire de loi gaussienne générale $Y \sim \mathcal{N}(m, \sigma^2)$.
  On a $Y = \sigma X + m$ avec $X \sim \mathcal{N}(0,1)$, $m \in \mathbb{R}$ et $\sigma \neq 0$.
  Si $\sigma > 0$,

$$F_Y(t) = P(Y \leqslant t) = P(\sigma X \leqslant t - m) = \Phi\left(\frac{t-m}{\sigma}\right)$$

  Donc $F_Y(t)$ est dérivable et sa dérivée vaut $f_Y(t) = \frac{\mathrm{e}^{-\frac{(t-m)^2}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}}$.
- Variable aléatoire de la loi de Laplace :

$$f_X(t) = \frac{\mathrm{e}^{-|t|}}{2}$$

- Variable aléatoire de loi exponentielle (modélisation de la durée de vie d'un composant) :

$$f_X(x) = \lambda \mathrm{e}^{-\lambda x} 1_{\{x \geqslant 0\}}$$

  On a $F_X(t) = 1 - \mathrm{e}^{-\lambda t}$.

 

- Loi uniforme sur $[a,b]$ :

$$f_U = \frac{1}{b-a} 1_{[a,b]}$$

On a $F_U(t) = 0$ si $t \leqslant a$, $F_U(t) = 1$ si $i \geqslant b$ et $F_U(t) = \frac{t}{b-a}$ sinon.
- Loi de Cauchy :

$$f_X(x) = \frac{1}{\pi(1+x^2)}$$

- Loi $\Gamma$ :

$$f_X(x) = \frac{\lambda^p}{\Gamma(p)} x^{p-1} \mathrm{e}^{-\lambda x} 1_{\{x>0\}}$$

## 1.3 Vecteurs aléatoires

**Définition 1.13** Soit $X : \Omega \to \mathbb{R}^d$.

$X$ est un vecteur aléatoire ssi $X_1, \ldots, X_d$ sont des variables aléatoires réelles.

**Définition 1.14** Soit $P_X : \mathbb{R}^d \to [0,1]$ la loi d'un vecteur aléatoire $X = (X_1, \ldots, X_d)$.

On définit la fonction de répartition jointe par :

$$F_X(t_1, \ldots, t_d) = P_X(]-\infty, t_1] \times \ldots \times ]-\infty, t_d])$$

La loi de la marginale $X_j$ est donnée par la fonction de répartition

$$F_{X_j}(t) = \lim_{t_i \to +\infty, i \neq j} F_X(t_1, \ldots, t_d)$$

### 1.3.1 Variables discrètes

$P(X = x_j) = P\left(\bigcup_{i=0}^{\infty} \{X = x_j, Y = y_i\}\right) = \sum_{i=0}^{\infty} P(X = x_j, Y = y_i)$.

La loi marginale de $X$ est donnée par :

$$\begin{pmatrix} x_j \\ p_j = \sum_{i=0}^{\infty} p_{j,i} \end{pmatrix}$$

 

### 1.3.2 Variables à densité

$P_X(B) = \int_B f_X(x)\,\mathrm{d}x$.

$$F_X(t_1,\dots,t_d) = \int_{-\infty}^{t_1}\dots\int_{-\infty}^{t_d} f_X(x)\,\mathrm{d}x$$

Dans le cas $(X,Y) \in \mathbb{R}^d \times \mathbb{R}^k$,

$$P_X(B) = P_{X,Y}(B \times \mathbb{R}^k) = \iint_{B\times\mathbb{R}^k} f_{X,Y}(x,y)\,\mathrm{d}x\mathrm{d}y = \int_B \underbrace{\int_{\mathbb{R}^k} f_{X,Y}(x,y)\,\mathrm{d}y}_{f_X(x)}\,\mathrm{d}x$$

On vérifie que $f_X$ est positive, borélienne d'intégrale 1. C'est donc la densité marginale de $X$.

**Proposition 1.10** Soit $X$ un vecteur aléatoire $d$-dimensionnel de densité $f_X$ et $g : \mathbb{R}^d \to \mathbb{R}^d$ un $C^1$-difféomorphisme.

$Y = g(X)$ est à densité $f_Y(y) = F_X(g^{-1}(y))\det(J_y(g^{-1}))$.

**Exemple 1.6** Soit $(X,Y)$ deux vecteurs bidimensionnels à densité. On cherche la loi de $X+Y$.

On pose $g : (x,y) \mapsto (x+y, y)$. C'est un difféomorphisme et le déterminant de la jacobienne est 1.

On a donc $f_{X+Y,Y}(u,v) = f_{X,Y}(u-v,v)$.

D'où :

$$f_{X+Y}(u) = \int_{\mathbb{R}} f_{X+Y,Y}(u,v)\,\mathrm{d}v = \int_{\mathbb{R}} f_{X,Y}(u-v,v)\,\mathrm{d}v$$

## 1.4 Simuler des variables aléatoires

L'ordinateur sait choisir aléatoirement un nombre entre 0 et 1 : ça revient à simuler une variable aléatoire uniforme sur $[0,1]$.

Pour simuler un lancer de pile ou face avec une probabilité de pile $p$, on pose $X = 1_{\{1-p\leqslant U\leqslant 1\}}$ et ça marche.

**Proposition 1.11** Soit $X$ une variable aléatoire réelle de répartition $F$.

Posons $G(u) = \inf\{t \in \mathbb{R}, F(t) \geqslant u\}$ pour $u \in ]0,1]$.

$G(U)$ a la répartition $F$.

*Démonstration.* On vérifie que $G(u) \leqslant t \quad \text{ssi} \quad u \leqslant F(t)$ :

Soit $A(u) = \{t \in \mathbb{R}, F(t) \geqslant u\}$.

Comme $F$ est continue à droite, $A$ est fermé.

Si $G(u) \leqslant t$, alors comme $A(u)$ est fermé, $t \in A(u)$ donc $F(t) \geqslant u$.

De plus, si $G(u) > t$, $\inf A(u) > t$ donc $t \notin A(u)$ donc $F(t) < u$

Donc $P(G(u) \leqslant t) = P(U \leqslant F(t)) = F(t)$. ∎

 

*Remarque 1.3*

- *Si $F$ est continue et strictement croissante, $G = F^{-1}$.*
- *Si $A(u)$ est fermé, $\inf A(u) \in A(u)$ donc $F(G(u)) \geqslant u$. Si $F$ est continue à gauche, $F(G(u)) = u$.*

## 1.5 Espérance, variance, co-variances et inégalités

**Définition 1.15** On appelle espérance de $X$ l'intégrale par rapport à $P$ de $X$ :

$$E(X) = \int_\Omega X(\omega)P(\mathrm{d}\omega)$$

*Remarque 1.4*

- $E(c1_A) = cP(A)$
- *$E$ est linéaire sur les variables aléatoires étagées*
- *Si $(X_n)_n$ croît vers $X \geqslant 0$, $E(X) = \lim\limits_{n\to+\infty} E(X_n)$*
- *Si $X$ est quelconque, on décompose en $X^+$ et $X^-$ et on dit que $X$ est intégrable ssi $E(|X|)$ est finie. On a alors $E(X) = E(X^+) - E(X^-)$.*
- *$E$ est linéaire tout court.*

**Proposition 1.12**

- $|E(X)| \leqslant E(|X|)$.
- Toute variable aléatoire bornée est intégrable
- Si $(X_n)_n$ positive croît vers $X$ alors $E(X_n)$ tend vers $E(X)$ et

$$E\left(\sum_{n\geqslant 1} X_n\right) = \sum_{n\geqslant 1} E(X_n)$$

- Si $(X_n)_n$ est minorée par $Z$ intégrable, $E(\liminf\limits_{n\to+\infty} X_n) \leqslant \liminf\limits_{n\to+\infty} E(X_n)$.
- Si $(X_n)_n$ est dominée par $Z$ intégrable et converge simplement vers $X$ alors $E(X_n)$ converge vers $E(X)$ et $E(|X_n - X|)$ tend vers 0.
- Si $\sum\limits_{n\geqslant 1} E(|X_n|)$ est finie alors $\sum\limits_{n\geqslant 1} X_n$ est finie presque sûrement et

$$E\left(\sum_{n\geqslant 1} X_n\right) = \sum_{n\geqslant 1} E(X_n)$$

*Démonstration.* Soit $S_n = \sum\limits_{k=1}^{n} |X_k|$.

On a $E(S_n) = \sum\limits_{k=1}^{n} E(|X_k|)$.

 

Or $S_n$ croît vers $S = \sum_{n\geqslant 1} |X_n|$.

Par convergence monotone, $E(S) = \sum_{n\geqslant 1} E(|X_n|)$. Donc $S$ est intégrable.

Soit $\delta > 0$.

$1_{\{S=\infty\}} \leqslant \delta S$ donc $E(1_{\{S=\infty\}}) \leqslant \delta E(S)$ donc $P(\{S=\infty\}) \leqslant \delta E(S)$.

Avec $\delta \to 0$, $P(\{S = \infty\}) = 0$ donc $S$ est finie presque sûrement. Or $\left|\sum_{k=1}^{n} X_k\right| \leqslant S_n \leqslant S$ vérifie les hypothèses de la convergence dominée.

On a donc le résultat. ■

<u>Théorème 1.4</u> de transport ou tranfert *Soit $X : \Omega \to \mathbb{R}^d$ une variable aléatoire, $g : \mathbb{R}^d \to \mathbb{R}$ borélienne. Notons $Y = g(X)$.*

*$Y \in \mathscr{L}^1(\mu)$ ssi $g \in \mathscr{L}^1(P_X)$.*

*$E(Y) = E(|g(X)|)$ ssi $\int_{\mathbb{R}^+} |g(t)| P_X(\mathrm{d}t)$ est finie. Dans ce cas $E(g(X)) = \int_{\mathbb{R}} g(t) P_X(\mathrm{d}t)$.*

*Démonstration.* C'est vrai pour $g = 1_B$. Par linéarité, on a le résultat pour les fonctions étagées.

Par convergence dominée, on a le résultat pour les fonctions mesurables positives puis pour les mesurables quelconques. ■

**Exemple 1.7**

- Si $X$ suit une loi de Bernoulli ($X \sim \mathcal{B}(1,p)$) alors $E(X) = p$
- Si $X$ suit une loi binômiale ($X \sim \mathcal{B}(n,p)$) alors $E(X) = np$
- Si $X$ suit une loi de Poisson ($X \sim \mathcal{P}(\lambda)$) alors $E(X) = \lambda$
- Si $X$ suit une loi normale ($X \sim \mathcal{N}(0,1)$) alors $E(X) = 0$
- Si $X$ suit une loi normale ($X \sim \mathcal{N}(m,\sigma^2)$) alors $E(X) = m$

<u>**Définition 1.16**</u> ($d = 1$) Si $|X|^k \in \mathscr{L}^1$ avec $X$ une variable aléatoire réelle, alors :

$$E(X^k) = \int_{\mathbb{R}} t^k P_X(\mathrm{d}t) = \mu_k(t)$$

On appelle $\mu_k(X)$ le moment d'ordre $k$ de $X$.

Si $X^2 \in \mathscr{L}^1$ alors la variance de $X$ est $V(X) = E((X - E(X))^2) = E(X^2) - (E(X))^2$.

*Remarque 1.5*

- $0 \leqslant V(X)$
- *$X \in \mathscr{L}^2$ ssi $X \in \mathscr{L}^1$ et $V(X)$ finie.*
- $V(cX) = c^2 V(X)$
- $V(X + c) = V(X)$
- *Si $X \sim \mathcal{N}(m,\sigma^2)$, $V(X) = \sigma^2$.*

 

- *Si $X$ est à densité, $E(g(X)) = \int_{\mathbb{R}} g(t) f_X(t)\,\mathrm{d}t$.*

**Proposition 1.13** Soit $X$ une variable aléatoire réelle.

$$E(X) = \int_0^{\infty} P(X > t)\,\mathrm{d}t = \int_0^{\infty} (1 - F_X(t))\,\mathrm{d}t$$

De plus, $E(X) - 1 \leqslant \sum_{n \geqslant 1} P(X > n) \leqslant E(X)$.

Ainsi, $X \in \mathscr{L}^1$ ssi $\sum_{n \geqslant 1} P(X > n) < \infty$.

*Démonstration.*

$$\begin{aligned}
\int_0^{\infty} P(X > t)\,\mathrm{d}t &= \int_0^{\infty} E(1_{\{X > t\}})\,\mathrm{d}t \\
&= E\left(\int_0^{\infty} 1_{\{X > t\}}\,\mathrm{d}t\right) \\
&= E\left(\int_0^{\infty} 1_{[0, X[}(t)\,\mathrm{d}t\right) \\
&= E\left(\int_0^{X} \mathrm{d}t\right) \\
&= E(X)
\end{aligned}$$

Donc :

$$\begin{aligned}
E(X) &= \sum_{n \geqslant 1} \int_{n-1}^{n} P(X > t)\,\mathrm{d}t \\
&\leqslant \sum_{n \geqslant 1} P(X > n - 1) \\
&= P(X > 0) + \sum_{n \geqslant 2} P(X > n - 1) \\
&\leqslant 1 + \sum_{n \geqslant 1} P(X > n) \\
&\leqslant 1 + \sum_{n \geqslant 1} \int_{n-1}^{n} P(X > t)\,\mathrm{d}t \\
&= 1 + E(X)
\end{aligned}$$

∎

**Proposition 1.14 (Inégalités de Markov et Tchebytchev)** Soit $X$ une variable aléatoire réelle.

Si $X \in \mathscr{L}^1$ alors pour tout $t$, $P(X \geqslant t) \leqslant \frac{E(X^+)}{t} \leqslant \frac{E(|X|)}{t}$.

Si $X \in \mathscr{L}^2$ alors pour tout $t > 0$, $P(|X - E(X)| \geqslant t) \leqslant \frac{V(X)}{t^2}$.

 

*Démonstration.*

- $1_{[t,+\infty[} \leqslant \frac{X}{t} 1_{[t,+\infty[}(X) \leqslant \frac{X^+}{t} \leqslant \frac{|X|}{t}$
  Donc $P(X \geqslant t) = E(1_{[t,+\infty[}) \leqslant \frac{E(|X|)}{t}$.
- $Y = X - EX$.
  On a $t^2 1_{\{|Y| \geqslant t\}} \leqslant Y^2$ et on prend l'espérance. ■

*Remarque 1.6 Par Markov,* $E(|X|) = 0$ ssi $X = 0$ *presque sûrement.*

*L'inégalité pour* $Y$ *se la preuve de Tchebychev marche avec* $p \geqslant 1$ *à la place de* 2 *et on a* $P(|Y| \geqslant t) \leqslant \frac{E(|Y|)^p}{t^p}$.

**Proposition 1.15 (Inégalité de Jensen)** Soit $X \in \mathscr{L}^1$ à valeurs dans $I \subset \mathbb{R}$. Soit $\varphi$ convexe.

Alors $\varphi(X)$ est une variable aléatoire réelle $\mathscr{L}^1$ et $\varphi(E(X)) \leqslant E(\varphi(X))$.

**Lemme 1.4.1**
Si $\varphi$ est convexe, pour tout $x < t < y$, on a $S(x,t) \leqslant S(x,y) \leqslant S(t,y)$ avec $S(x,y) = \frac{\varphi(x)-\varphi(y)}{x-y}$.

**Lemme 1.4.2**
Si $\varphi$ est convexe, $\varphi$ est dérivable à gauche et à droite en tout point.

De plus la dérivée à droite est croissante et $\varphi(y) = \sup\limits_{x \in I} \varphi(x) + \varphi'_d(x)(y-x)$.

*Remarque 1.7* $|E(X)|^p \leqslant E(|X|^p)$ *quand* $X^p \in \mathscr{L}^1$.

**<u>Définition 1.17</u>** Soient $X, Y$ des variables aléatoires réelles de $\mathscr{L}^2$.

Alors $XY \in \mathscr{L}^1$ et la covariance de $X$ et $Y$ notée $\mathrm{Cov}(X,Y) = E((X - E(X))(Y - E(Y))) = E(XY) - E(X)E(Y)$ est bien définie.

Le cœfficient de corrélation de $X$ et $Y$ est défini par $\rho(X,Y) = \frac{\mathrm{Cov}(X,Y)}{V(X)V(Y)}$ si $X$ et $Y$ ne sont pas constantes presque sûrement.

Corollaire 1.2 *(Inégalité de Hölder)*

*Soit* $X \in \mathscr{L}^p$, $Y \in \mathscr{L}^q$ *avec* $\frac{1}{p} + \frac{1}{q} = 1$.

*Alors* $XY \in \mathscr{L}^1$ *et* $E(|XY|) \leqslant E(|X|^p)^{\frac{1}{p}} E(|X|^q)^{\frac{1}{q}}$.

*Démonstration.* On peut supposer que $E(|Y|)^q = 1$.

Posons $Q(A) = E(|Y|^q 1_A)$. C'est une probabilité. On applique alors Jensen pour $\varphi = x \mapsto x^p$, la variable aléatoire $|X||Y|^{1-q}$ et la probabilité $Q$.

On a $E_Q(Z) = E(Z|Y|^q)$ donc $\varphi(E_Q(|X||Y|^{1-q})) \leqslant E_Q(\varphi(|X||Y|^{1-q}))$.

Après simplifications, on a le résultat. ■

*Remarque 1.8 Avec* $p = q = 2$, *on obtient Cauchy-Schwartz.*

Corollaire 1.3 (Inégalité de Minkowski) $X \mapsto E(|X|^p)^{\frac{1}{p}}$ *vérifie l'inégalité triangulaire.*

**<u>Définition 1.18</u>** $L^p = \mathscr{L}^p/\mathrm{ps}$.

 

**Proposition 1.16** $L^p$ est un Banach pour $\|\cdot\|_p$. De plus $L^2$ est un Hilbert.

**Proposition 1.17** Soit $r > p \geqslant 1$.
Alors $L^r \subset L^p$ et $\|\cdot\|_p \leqslant \|\cdot\|_r$ sur $L^r$.

*Démonstration.* On peut supposer $X \geqslant 0$ car l'inégalité fait apparaître seulement $|X|$.

De plus on peut approcher $X$ par $(X_n)_n$ avec $X_n = \min(X, n)$ et $0 \leqslant X_n \leqslant X$.

On peut supposer $X$ bornée. On utilise alors Jensen avec $\varphi(x) = |x|^s$ avec $s = \frac{r}{p}$.

On a $\|X\|_p^r = (E(|X|^p))^s \leqslant E(|X|^s) = \|X\|_r^r$. ∎

**Définition 1.19** Soit $X$ un vecteur aléatoire tel que $E(|X|^2)$ soit finie.
La matrice de covariance de $X$ est la matrice $K_X = (\mathrm{Cov}(X_i, X_j))_{i,j}$.

**Proposition 1.18** $K_X \in S_n^{++}$.
Pour toute matrice $A$ et vecteur aléatoire $X$, $E(AX) = AE(X)$ et $K_{AX} = AK_XA^*$.

*Démonstration.* On a :

$$\sum_{j=1}^{d}\sum_{k=1}^{d} K_X(j,k) u_j u_k = V(u^* X) \geqslant 0$$

Le deuxième point est débile. ∎

 

# Chapitre 2

# Indépendance

## 2.1 Définitions et premières propriétés

**Définition 2.1** Une famille $(A_t)_{t\in T}$ d'évènements est dite indépendante ssi pour tout $J \subset T$ fini,

$$P\left(\bigcap_{i\in J} A_i\right) = \prod_{i\in J} P(A_i)$$

**Définition 2.2** Une famille $(\mathcal{A}_t)_{t\in T}$ de tribus est dite indépendante ssi pour tout $J \subset T$ fini et pour tout choix d'évènements $A_j \in \mathcal{A}_j$,

$$P\left(\bigcap_{i\in J} A_i\right) = \prod_{i\in J} P(A_i)$$

**Définition 2.3** Une famille $(X_t)_{t\in T}$ de variables aléatoires est dite indépendante ssi pour tout $J \subset T$ fini et pour tout choix $(B_j)_{j\in J}$ de boréliens,

$$P\left(\bigcap_{i\in J} \{X_j \in B_j\}\right) = \prod_{i\in J} P(X_j \in B_j)$$

*Remarque 2.1 On note l'indépendance $A \perp B$.*

*$A \perp\!\!\!\perp B$ ssi $\sigma A \perp\!\!\!\perp \sigma(B)$.*

*De manière générale, une famille de variable aléatoires est indépendantes ssi la famille des tribus images est indépendante.*

**Proposition 2.1** Soient $E$ et $F$ deux algèbres de parties de $\Omega$.

On suppose $E \perp\!\!\!\perp F$. Alors $\sigma(E) \perp\!\!\!\perp \sigma(F)$.

*Démonstration.* Soit $A \in E$ fixé.

Soit $\mathcal{M}_1 = \{B \in \sigma(F), P(A \cap B) = P(A)P(B)\}$.

$\mathcal{M}_1$ est une classe monotone qui contient $F$. En effet, si $(B_n)_n$ en est une suite croissante,

$$\begin{aligned} P(A \cap \lim_{n\to+\infty} B_n) &= P(\lim_{n\to+\infty}(A \cap B_n)) \\ &= \lim_{n\to+\infty} P(A \cap B_n) \\ &= \lim_{n\to+\infty} P(A)P(B_n) \\ &= P(A)P(\lim_{n\to+\infty} B_n) \end{aligned}$$

Par le théorème de classe monotone, $\sigma(F) = \mathcal{M}_1$ donc tout élément de $\sigma(F)$ est indépendant de tout élément de $E$.

De même, pour $B \in \sigma(F)$, $\mathcal{M}_2 = \{A \in \sigma(E), P(A \cap B) = P(A)P(B)\}$ est une classe monotone qui contient $E$.

Donc tout élément de $\sigma(E)$ est indépendant de tout élément de $\sigma(F)$. ■

Corollaire 2.1 *Soit $(X_t)_{t\in T}$ une famille de variables aléatoires réelles.*

*C'est une famille indépendante ssi pour tout $J \subset T$ finie et pour tout choix $(s_j)_{j\in J}$ réels, on a :*

$$P\left(\bigcap_{j\in J}\{X_j \leqslant s_j\}\right) = \prod_{j\in J} P(\{X_j \leqslant s_j\})$$

*Démonstration.* Il suffit de vérifier que $\{X_1^{-1}(]-\infty, s]), s \in \mathbb{R}\} \perp\!\!\!\perp \{X_2^{-1}(]-\infty, s]), s \in \mathbb{R}\}$. ■

Corollaire 2.2 $X_1 \perp\!\!\!\perp X_2$ *ssi* $P_{X_1,X_2} = P_{X_1} \otimes P_{X_2}$.

*Démonstration.*

$\Rightarrow$ $P_{(X,Y)}(A \times B) = P(X \in A, Y \in B) = P(X \in A)P(Y \in B) = P_X(A)P_Y(B)$

$\Leftarrow$ On a

$$\begin{aligned} P(X \in A, Y \in B) &= P_{(X,Y)}(A \times B) = P_X \otimes P_Y(A \times B) \\ &= P_X(A)P_Y(B) = P(X \in A)P(Y \in B) \end{aligned}$$ ■

Corollaire 2.3 $X_1, \ldots, X_n$ *sont indépendantes ssi* $F_{X_1,\ldots,X_n}(t_1, \ldots, t_n) = F_{X_1}(t_1) \ldots F_{X_n}(t_n)$ *pour tous* $(t_1, \ldots, t_n)$.

Corollaire 2.4 *Soient $(X_1, \ldots, X_n)$ des variables indépendantes et $n_1 + \ldots + n_r = n$ des entiers positifs.*

*Les variables aléatoires*

$$(X_1, \ldots, X_{n_1}), (X_{n_1+1}, \ldots, X_{n_2}), \ldots, (X_{n_1+\ldots+n_{r-1}+1}, \ldots, X_n)$$

*sont indépendantes.*

 

**Proposition 2.2** Soient $X, Y$ deux variables aléatoires de dimension $d$ et $k$.

$X \perp\!\!\!\perp Y$ ssi une des conditions suivantes est vérifiée :

1. $P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$ pour $A \in \mathscr{B}(\mathbb{R}^d)$, $B \in \mathscr{B}(\mathbb{R}^k)$
2. $P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$ pour $A, B$ générateurs de $\mathscr{B}(\mathbb{R}^d)$ et $\mathscr{B}(\mathbb{R}^k)$
3. $f(X) \perp\!\!\!\perp g(Y)$ pour tout $f : \mathbb{R}^d \to \mathbb{R}^{d'}$ et $g : \mathbb{R}^k \to \mathbb{R}^{k'}$ boréliennes
4. $E(f(X)g(Y)) = E(f(X))E(g(Y))$ pour chaque paire de fonctions boréliennes bornées (ou positives) $\mathbb{R}^d \to \mathbb{R}$ et $\mathbb{R}^k \to \mathbb{R}$

*Démonstration.*

$1 \Leftrightarrow 2$ Déjà vu

$3 \Rightarrow 1$ Il suffit de prendre $f = \mathrm{Id}$ et $g = \mathrm{Id}$.

$1 \Rightarrow 3$ $f(X)^{-1}(\mathscr{B}(\mathbb{R}^d)) = X^{-1}(f^{-1}(\mathscr{B}(\mathbb{R}^d))) \subset X^{-1}(\mathscr{B}(\mathbb{R}^d))$.
De même $g(Y)^{-1}(\mathscr{B}(\mathbb{R}^k)) \subset Y^{-1}(\mathscr{B}(\mathbb{R}^k))$.
Donc $X \perp\!\!\!\perp Y$ ssi $X^{-1}(\mathscr{B}(\mathbb{R}^d)) \perp\!\!\!\perp Y^{-1}(\mathscr{B}(\mathbb{R}^k))$. Il y a donc indépendance des familles $X^{-1}(\mathscr{B}(\mathbb{R}^d))$ et $Y^{-1}(\mathscr{B}(\mathbb{R}^k))$.

$4 \Rightarrow 1$ Il suffit de prendre $f = 1_A$ et $g = 1_B$.

$$E(1_A(X)1_B(Y)) = E(1_{A\times B}(X, Y)) = P(X \in A)P(Y \in B)$$

$1 \Rightarrow 4$ Par 1, c'est vrai pour les indicatrices. On conclut par convergence monotone et décomposition de Han-Jordan. ■

**Exemple 2.1**

- Soient $X, Y$ discrètes.
  On note $p_{j,k} = P((X, Y) = (p_j, p_k))$. $X \perp\!\!\!\perp Y$ ssi $p_{j,k} = p_j p_k$.
- Soit $(X, Y)$ à densité sur $\mathbb{R}^2$ de densité $f_{X,Y}$.
  Alors $X$ et $Y$ sont à densité de densités $f_X(x) = \int_{\mathbb{R}} f_{X,Y}(x, y)\,\mathrm{d}y$ et $f_Y(y) = \int_{\mathbb{R}}^{f} {}_{X,Y}(x, y)\,\mathrm{d}x$.
  $X \perp\!\!\!\perp Y$ ssi $f_{X,Y}(x, y) = f_X(x)f_Y(y)$ pour presque tout $x, y$.
  De gauche à droite, on utilise le théorème de classe monotone sur

$$\left\{C \in \mathscr{B}(\mathbb{R}^2), \iint_C f_{X,Y}(x, y)\,\mathrm{d}x\,\mathrm{d}y = \iint_C f_X(x)f_Y(y)\,\mathrm{d}x\,\mathrm{d}y\right\}$$

  Pour la réciproque, on écrit des intégrales.

 

## 2.2 Somme de variables aléatoires indépendantes

**Proposition 2.3** Soient $X, Y$ deux variables aléatoires réelles indépendantes.

Si $X$, $Y$ et $XY$ sont positives ou intégrables alors $E(XY) = E(X)E(Y)$.

*Démonstration.*

$$E(|XY|) = \iint |xy| P_{(X,Y)}(\mathrm{d}x, \mathrm{d}y) = \iint |x||y| P_X(\mathrm{d}x) P_Y(\mathrm{d}y) < \infty$$

Ensuite on refait pareil sans valeurs abolues. ■

*Remarque 2.2 Si $X \perp\!\!\!\perp Y$,* $\mathrm{Cov}(X, Y) = 0$. *Mais la réciproque est fausse ($X \sim \mathcal{N}(0,1)$ et $Y = X^2$).*

**<u>Définition 2.4</u>** On dit que $X$ et $Y$ sont non-corrélées ssi $\mathrm{Cov}(X, Y) = 0$.

**Proposition 2.4** Si $X_1, \ldots, X_n \in L^2$ sont non corrélées deux à deux, alors :

$$\mathrm{Var}(X_1 + \ldots + X_n) = \sum_{j=1}^{n} \mathrm{Var}(X_j)$$

$$P\left(\left|\sum_{j=1}^{n} (X_j - E(X_j))\right| \geqslant t\right) \leqslant \frac{1}{t^2} \sum_{j=1}^{n} \mathrm{Var}(X_j)$$

*Démonstration.* L'inégalité découle directement de Bienaymé-Tchebychev.

Pour le reste, OPS que $X_1, \ldots, X_n$ sont centrées (d'espérance nulle).

On développe alors $(X_1 + \ldots + X_n)^2$ et on passe à l'espérance pour obtenir le résultat. ■

**Proposition 2.5** Soit $X, Y$ indépendantes.

$X + Y$ a la loi $P_X * P_Y$ (convolution).

Pour toute fonction mesurable $h$,

$$\int f(t)(P_X * P_Y)(\mathrm{d}t) = \iint h(x + y) P_X(\mathrm{d}x) P_Y(\mathrm{d}y)$$

*Remarque 2.3 Si $X$ et $Y$ sont de plus à densité, alors la densité de $X + Y$ est $f_X * f_Y$.*

 

*Démonstration.*

$$\begin{aligned}\int h(t)P_{X+Y}(\mathrm{d}t) &= E(h(X+Y)) \\ &= \iint h(x+y)P_{(X,Y)}(\mathrm{d}x,\mathrm{d}y) \\ &= \iint h(x+y)P_X \otimes P_Y(\mathrm{d}x,\mathrm{d}y) \\ &= \iint h(x+y)P_X(\,\mathrm{d}x)P_Y(\,\mathrm{d}y)\end{aligned}$$

**Exemple 2.2**

- Si $X \sim B(n,p)$, $Y \sim B(m,p)$ et $X \perp\!\!\!\perp Y$, alors $X+Y \sim B(n+m,p)$.
- Si $X \sim \mathcal{P}(\lambda)$, $Y \sim \mathcal{P}(\mu)$ et $X \perp\!\!\!\perp Y$, alors $X+Y \sim \mathcal{P}(\lambda+\mu)$.
- Si $X \sim \mathcal{N}(m,\sigma^2)$, $Y \sim \mathcal{N}(n,\tau^2)$ et $X \perp\!\!\!\perp Y$ alors $X+Y \sim \mathcal{N}(m+n,\tau^2+\sigma^2)$.

## 2.3 Suite d'évènements et de variables aléatoires indépendantes

### 2.3.1 Construction d'un produit infini de probabilités

On note $\mathbb{R}^\infty$ l'ensemble des suites réelles.

Soient $(\mathbb{R},\mathscr{B}(\mathbb{R}),P_1),(\mathbb{R},\mathscr{B}(\mathbb{R}),P_2),\ldots$ des espaces de probabilités.

Les cylindre ouverts ($\prod_{i=1}^{\infty} A_i$ avec $A_i=\mathbb{R}$ sauf un nombre fini) donnent la topologie produit sur $\mathbb{R}^\infty$ : $\mathscr{B}(\mathbb{R}^\infty)$.

On pose $P^{(n)} = P_1\otimes\ldots\otimes P_n$. $(P^{(n)})_n$ est consistante : $P^{(n)}(A_1\times\ldots\times A_n) = P^{(n+1)}(A_1\times\ldots\times A_n\times\mathbb{R})$.

<u>THÉORÈME 2.1</u> DE KOLMOGOROV *Si une famille $(P^{(n)})_n$ de probabilités sur $\mathbb{R}^\infty$ est consistante, alors il existe une unique probabilité $P^{(\infty)}$ sur $\mathbb{R}^\infty$ tel que $\pi_n(P^\infty) = P^{(n)}$ avec $\pi_n : x \mapsto (x_1,\ldots,x_n)$.*

### 2.3.2 Construction explicite d'une suite de variables aléatoires indépendantes

On se place sur $(]0,1],\mathscr{B}(]0,1]),\lambda)$.

Pour $\omega \in ]0,1]$, sa représentation diadique la suite $(d_n(\omega))_n$ telle que :

$$\omega = \sum_{n=1}^{\infty} \frac{d_n(\omega)}{2^n}$$

 

Il y a unicité sauf pour les réels de la forme $\frac{k}{2^n}$ (en nombre dénombrable donc de mesure nulle).

$(d_n)_n$ est une famille de variables aléatoires indépendantes de loi $B(1, \frac{1}{2})$.

On pose $A_1 = ]\frac{1}{2}, 1]$, $A_2 = ]\frac{1}{4}, \frac{1}{2}] \cup ]\frac{3}{4}, 1]$ et :

$$A_n = \bigcup_{u_1,\dots,u_{n-1}\in\{0,1\}} ]0.u_1u_2\dots u_{n-1}100\dots, 0.u_1u_2\dots u_{n-1}11\dots]$$

$(A_n)_n$ est une suite d'évènements indépendants.

### 2.3.3 Lois de 0–1

**<u>Définition 2.5</u>** Soit $(A_n)_n$ une suite de sous-tribus de $\mathcal{A}$. On pose $T_n = \sigma\left(\bigcup_{m\geqslant n} \mathcal{A}_m\right)$.

La tribu $T_\infty = \bigcap_{n\geqslant 1} T_n$ est appellée tribu terminale ou asymptotique de $A_n$.

<u>Théorème 2.2</u> (Loi de 0-1 de Kolmogorov) *Soit $(\mathcal{A}_n)_n$ une suite de tribus indépendantes.*

*Chaque évènement de $T_\infty$ est de probabilité 0 ou 1 (La tribu terminale est presque sûrement triviale).*

*Démonstration.* Soit $A \in T_\infty$. On note $\mathcal{D}$ l'ensemble des évènements indépendants de $A$ : $D \in \mathcal{D}$ ssi $P(A \cap D) = P(A)P(D)$.

Montrons que $\mathcal{D}$ est une classe monotone, on aura alors $P(A)^2 = P(A)$ et le résultat.

Pour tout $n$, $A \in T_{n+1}$ donc $A \perp\!\!\!\perp \underbrace{\sigma(\mathcal{A}_1 \cup \dots \cup \mathcal{A}_n)}_{=\mathcal{S}_n}$ car $T_{n+1} \perp\!\!\!\perp \mathcal{S}_n$.

Donc pour tout $B \in \bigcup_{n=1}^{\infty} \mathcal{S}_n$, $A \perp\!\!\!\perp B$ donc $\mathcal{D}$ est une classe monotone.

Or $\mathcal{A}_n \subset \bigcup_{n=1}^{\infty} \mathcal{S}_n$ donc $T_n \subset \sigma\left(\bigcup_{n=1}^{\infty} \mathcal{S}_n\right)$.

Donc $T_\infty \subset \mathcal{D}$ donc $A \in \mathcal{D}$. ∎

Corollaire 2.5 (Loi 0-1 de Borel) *Pour toute suite $(A_n)_n$ d'évènements indépendants, on a $P(\limsup_n A_n) = 0$ ou 1.*

*Démonstration.* Soit $\mathcal{A}_n = \sigma(A_n)$. C'est une suite de tribus indépendantes.

On pose $B_n = \bigcup_{m\geqslant n} B_m$. Ce sont des éléments de $T_n$.

 

$\limsup A_n = \bigcap_{k=1}^{\infty} B_k \in T_j$ pour tout $j$. Donc $\limsup A_n \in T_\infty$ et Kolmogorov conclut. ■

<u>THÉORÈME 2.3</u> (BOREL-CANTELLI) *Soit $(A_n)_n$ une suite d'évènements de $\mathcal{A}$.*

$$\sum_{n=1}^{\infty} A_n < \infty \Rightarrow P(\limsup_n A_n) = 0$$

*Et si les évènements sont indépendants alors la réciproque est vraie.*

*Démonstration.* Soit $A = \limsup A_n$. $A \subset \bigcup_{j \geqslant n} A_j$ pour tout $n$.

Donc $P(A) \leqslant P\left(\bigcup_{j \geqslant n} A_j\right) \leqslant \sum_{j \geqslant n} P(A_j)$ qui est un reste dde série convergente.

Donc $P(A) = 0$.

Notons $I_n = 1_{A_n}$, $S_n = \sum_{j=1}^{n} I_j$ et $S = \sum_{m=1}^{\infty} I_m$.

On a $I_j^2 = I_j$ par Bienaymé et :

$$\begin{aligned} \mathrm{Var}(S_n) &= \sum_{j=1}^{n} \mathrm{Var}(I_j) \\ &= \sum_{j=1}^{n} (E(I_j^2) - E(I_j)^2) \\ &= \sum_{j=1}^{n} (E(I_j) - E(I_j)^2) \\ &= E(S_n) - \sum_{j=1}^{n} E(I_j)^2 \\ &\leqslant E(S_n) \end{aligned}$$

Par hypothèse, $\sum_{n \geqslant 1} E(I_n) = +\infty$.

Comme $\lim_{n \to +\infty} S_n = S$, on a $\lim_{n \to +\infty} E(S_n) = E(S) = +\infty$.

Or $\omega \in A$ ssi $\omega \in A_n$ pour une infinité de $n$ donc $S(\omega) = +\infty$.

Donc $A = \{S = +\infty\}$ et on montre que $P(S = +\infty) = 1$ :

$$P(|S_n - E(S_n)| \leqslant t) \geqslant 1 - \frac{\mathrm{Var}(S_n)}{t^2}$$

 

Comme $\lim\limits_{n\to+\infty} S_n = +\infty$, on peut supposer $E(S_n) > 0$. Alors :

$$P\left(S_n \geqslant \frac{E(S_n)}{2}\right) \geqslant P\left(|S_n - E(S_n)| \leqslant \frac{E(S_n)}{2}\right) \geqslant 1 - 4\frac{\mathrm{Var}(S_n)}{E(S_n)^2}$$

Mais ce terme tend vers 1 donc pour tout $\varepsilon > 0$, $P(S_n \geqslant \frac{E(S_n)}{2}) \geqslant 1-\varepsilon$.
Comme $S \geqslant S_n$, $P(S \geqslant \frac{E(S_n)}{2}) \geqslant 1-\varepsilon$ pour tout $n$ sauf un nombre fini.
Donc, avec $n \to +\infty$ et $\varepsilon \to 0$, $P(S = +\infty) \geqslant 1$. ∎

Corollaire 2.6 *Si $(A_n)_n$ contient une sous-suite d'évènements indépendants dont la somme des probabilités diverge alors $P(\limsup A_n) = 1$.*

*Démonstration.* $\limsup A_{\varphi(n)} \subset \limsup A_n$. ∎

<u>Théorème 2.4</u> *Soit $(X_n)_n$ une suite de variables aléatoires indépendantes. On note $\mathcal{A}_n = \sigma(X_n)$.*

*Alors chaque évènement asymptotique pour $(\mathcal{A}_n)_n$ est de probabilité 0 ou 1.*

*Remarque 2.4 Si les remarques spnt positives et indépendantes, est-ce que $\left\{\sum\limits_{n=1}^{\infty} X_n = +\infty\right\}$ est un évènement asymptotique ?*

Corollaire 2.7 *Soit $X : \Omega \to \mathbb{R}$ est une variable aléatoire $T$-mesurable avec $T$ presque sûrement triviale.*

*Alors il existe $c \in \mathbb{R}$ tel que $X = c$ presque sûrement.*

*Démonstration.* $F_X(t) = P(X \leqslant t) = P(X^{-1}(]-\infty, t]))$.

Donc $F_X(t) \in \{0,1\}$. Il y a donc un saut de taille 1. On pose $c = \inf\{t, F_X(t) = 1\}$.

Il existe $t_n$ qui décroît vers $c$. On a $\{X \leqslant t_n\}$ qui tend vers $\{X \leqslant c\}$.

Donc $P(X \leqslant t_n)$ tend vers $P(X \leqslant c)$ donc $c \in \{t, F_X(t) = 1\}$.

∎

 

# Chapitre 3

# Fonctions caractéristiques

## 3.1 Définitions et premières propriétés

**Définition 3.1** Soit $X : \Omega \to \mathbb{R}^d$ une variable aléatoire de loi $P_X$. La fonction caractéristique de la variable aléatoire $X$ ou de sa loi $P_X$ est $\varphi_X : \mathbb{R}^d \to \mathbb{C}$ définie par :

$$\varphi_X(t) = E(\mathrm{e}^{i\langle t,X\rangle}) = E(\cos(\langle t,X\rangle)) + iE(\sin(\langle t,X\rangle))$$

*Remarque 3.1*
- $\varphi_x(0) = 1$
- $|\varphi_X(t)| = |E(\mathrm{e}^{i\langle t,X\rangle})| \leqslant E(1) = 1$
- $\overline{\varphi_X(t)} = \varphi_X(-t) = \varphi_{-X}(t)$ *donc si X est de loi symétrique, ($X \sim -X$), alors $\varphi_{-X}(t) = \varphi_X(t)$ donc $\overline{\varphi_X(t)} = \varphi_X(t)$.*

Calculs :
- Si $X$ est discrète, $\varphi_X(t) = \sum_{j\geqslant 1} p_j \mathrm{e}^{i\langle t,x_j\rangle}$.
- Si $X \sim \mathcal{P}(\lambda)$, $P(X = n) = \mathrm{e}^{\lambda(\mathrm{e}^{it}-1)}$.
- Si $X$ est à densité, $\varphi_X(t) = \int_{\mathbb{R}^d} \mathrm{e}^{i\langle t,x\rangle} f_X(x)\,\mathrm{d}x$.
- Si $X \sim \mathcal{E}(\lambda)$, $\varphi_X(t) = \frac{\lambda}{\lambda - it}$.

**Proposition 3.1** $\varphi_X$ est uniformément continue, vaut 0 en 1 et est de type positif :

$$\forall n \geqslant 1, \forall z_1, \ldots, z_n \in \mathbb{C}, \forall t_1, \ldots, t_n \in \mathbb{R}^d, \sum_{j=1}^{n}\sum_{k=1}^{n} \varphi_X(t_k - t_j) z_k \overline{z_j} \geqslant 0$$

*Démonstration.* On a :

$$\begin{aligned}|\varphi_X(t+h)-\varphi_X(t)| &= |E(\mathrm{e}^{i\langle t+h,X\rangle}-\mathrm{e}^{i\langle t,X\rangle})| \\ &= |E(\mathrm{e}^{i\langle t,X\rangle}(\mathrm{e}^{i\langle h,X\rangle}-1))| \\ &\leqslant E(|\mathrm{e}^{i\langle t,X\rangle}(\mathrm{e}^{i\langle h,X\rangle}-1)|) \\ &= E(|\mathrm{e}^{i\langle h,X\rangle}-1|) \to 0\end{aligned}$$

D'où l'uniforme continuité. De plus,

$$\begin{aligned}\sum_{j=1}^{n}\sum_{k=1}^{n}\varphi_X(t_k-t_j)z_k\overline{z_j} &= \sum_{j=1}^{n}\sum_{k=1}^{n}z_k\overline{z_j}E(\mathrm{e}^{i\langle t_k-t_j,X\rangle}) \\ &= E\left(\sum_{j=1}^{n}\sum_{k=1}^{n}z_k\mathrm{e}^{i\langle t_k,X\rangle}\overline{z_j\mathrm{e}^{i\langle t_j,X\rangle}}\right) \\ &= E\left(\left(\sum_{k=1}^{n}z_k\mathrm{e}^{i\langle t_k,X\rangle}\right)\overline{\left(\sum_{j=1}^{n}z_j\mathrm{e}^{i\langle t_j,X\rangle}\right)}\right) \\ &= E\left(\left|\sum_{k=1}^{n}z_k\mathrm{e}^{i\langle t_k,X\rangle}\right|\right) \geqslant 0\end{aligned}$$

<u>Théorème 3.1</u> (Bochner) *Si $g : \mathbb{R}^d \to \mathbb{C}$ est continue en $0$, de type positif et vaut $1$ en $0$ alors il existe une variable aléatoire $X$ et que $g=\varphi_X$.*

**Proposition 3.2** Pour tout $t\in\mathbb{R}$, $\varphi_X'(t)=E(iX\mathrm{e}^{itX})$.

*Démonstration.* Soit $h\neq 0$.

$$\frac{\varphi_X(t+h)-\varphi_X(t)}{h}-E(iX\mathrm{e}^{tX})=E\left(\mathrm{e}^{itX}\frac{\mathrm{e}^{ihX}-1-ihX}{h}\right)$$

On a de plus $|\mathrm{e}^{ix}-1-ix|\leqslant\min\{\frac{|x|^2}{2},2|x|\}$.
Donc $\frac{\mathrm{e}^{ihX}-1-ihX}{h}\leqslant 2|X|\in L^1$ et $\frac{\mathrm{e}^{ihX}-1-ihX}{h}\leqslant\frac{|h||X|^2}{2}\to 0$ quand $h\to 0$.
Par convergence dominée, on a le résultat. ∎

*Remarque 3.2 En particulier, $\varphi_X'(0)=iE(X)$.*

**Proposition 3.3** Soit $X\in L^p$ une variable aléatoire de fonction caractéristique $\varphi_X$.

Alors $\varphi_X$ est $p$ fois dérivable (avec des dérivées continues).

Pour tout $k\in[\![1,p]\!]$, $\varphi_X^{(k)}(t)=i^kE(X^k\mathrm{e}^{itX})$.

*Remarque 3.3*

- *De plus, $\varphi_X^{(k)}(0)=i^kE(X^k)$.*

 

- *En particulier, si $p = 2$, $E(X) = \frac{1}{i}\varphi_X'(0)$ et $E(X^2) = -\varphi_X''(0)$.*
- *La réciproque est fausse : on peut avoir $\varphi_X$ dérivable mais pas l'existence de $E(X)$.*
  *Par exemple : on prend $X$ discrète tel que $P(X = \pm j) = \frac{c}{j^2 \ln(j)}$ avec $c > 0$ pour que ce soit un loi de probabilité.*
  $E(X) = 2\sum_{j=2}^{\infty} \frac{c}{j\ln(j)} = +\infty.$
  $\varphi_X(t) = 2c\sum_{j=2}^{\infty} \frac{\cos(tj)}{j^2 \ln(j)}$ *et* $\sum_{j=2}^{\infty} \frac{\sin(tj)}{j\ln(j)} < \infty$ *et converge uniformément.*
  *Donc $\varphi_X$ est dérivable et vaut $0$ en $0$.*

**Proposition 3.4** Soit $X$ un variable aléatoire réelle de fonction caractéristique $\varphi_X$.

Si $\varphi_X$ est $p$ fois dérivable en $0$ avec $p \geqslant 2$ alors $X$ admet des moments d'ordres inférieurs à $2r$ avec $r = \lfloor \frac{p}{2} \rfloor$.

*Démonstration.* On le fait dans le cas $\varphi_X$ deux fois dérivable en 0.

$$\begin{aligned}
\varphi_X''(0) &= \lim_{h \to 0} \frac{\varphi_X(h) - \varphi_X(-h) - 2\varphi_X(0)}{h^2} \\
&= \lim_{h \to 0} \frac{E(\mathrm{e}^{ihX} + \mathrm{e}^{-ihX}) - 2}{h^2} \\
&= \lim_{h \to 0} \frac{2}{h^2}(E(\cos(hX)) - 1)
\end{aligned}$$

On peut montrer que $x^2 = \lim\limits_{h \to 0} \frac{2(1-\cos(hX))}{h^2}$.

$$\begin{aligned}
E(X^2) &= E\left(\lim_{h \to 0} \frac{2(1 - \cos(hX))}{h^2}\right) \\
&\leqslant \liminf_{h \to 0} E\left(2\frac{1 - \cos(hX)}{h^2}\right) = -\varphi_X''(0) < \infty
\end{aligned}$$

via Fatou. ■

*Remarque 3.4 Si $X \in L^2$ et $E(X) = 0$, alors $\varphi_X(u) = 1 - \frac{\sigma^2 u^2}{2} + o(u)$ avec $\sigma^2 = E(X^2) = \mathrm{Var}(X)$.*

 

**Calcul de la fonction caractéristique de $\mathcal{N}(0,1)$**

- Première méthode :

$$\begin{aligned}\varphi_X(t) &= \int_{-\infty}^{\infty} \mathrm{e}^{itx}\frac{\mathrm{e}^{-\frac{x^2}{2}}}{\sqrt{2\pi}}\,\mathrm{d}x \\ &= \int_{-\infty}^{\infty} \cos(tx)\frac{\mathrm{e}^{-\frac{x^2}{2}}}{\sqrt{2\pi}}\,\mathrm{d}x + i\underbrace{\int_{-\infty}^{\infty} \sin(tx)\frac{\mathrm{e}^{-\frac{x^2}{2}}}{\sqrt{2\pi}}\,\mathrm{d}x}_{=0} \\ &= \int_{-\infty}^{\infty} \cos(tx)\mathrm{e}^{-\frac{x^2}{2}}\sqrt{2\pi}\,\mathrm{d}x\end{aligned}$$

On dérive sous l'intégrale (oui oui, on peut) :

$$\begin{aligned}\varphi_X'(t) &= \frac{1}{\sqrt{2\pi}} \\ &= \int_{-\infty}^{\infty} -x\sin(tx)\mathrm{e}^{-\frac{x^2}{2}}\,\mathrm{d}x \\ &\overset{IPP}{=} -\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} t\cos(tx)\mathrm{e}^{-\frac{x^2}{2}}\,\mathrm{d}x \\ &= -t\varphi_X(t)\end{aligned}$$

Donc $\varphi_X(t) = \mathrm{e}^{-\frac{t^2}{2}}\varphi_X(0) = \mathrm{e}^{-\frac{t^2}{2}}$.
- Deuxième méthode :
Pour $a \in \mathbb{R}$, on a :

$$\begin{aligned}\int_{-\infty}^{\infty} \mathrm{e}^{ax}\frac{\mathrm{e}^{-\frac{x^2}{2}}}{\sqrt{2\pi}}\,\mathrm{d}x &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} \mathrm{e}^{\frac{a^2}{2}}\mathrm{e}^{-\frac{(x-a)^2}{2}}\,\mathrm{d}x \\ &= \mathrm{e}^{\frac{a^2}{2}}\underbrace{\int_{-\infty}^{\infty}\frac{\mathrm{e}^{-\frac{(x-a)^2}{2}}}{\sqrt{2\pi}}\,\mathrm{d}x}_{=1}\end{aligned}$$

On pose $h_1 = z \mapsto \mathrm{e}^{\frac{z^2}{2}}$ et $h_2 : z \mapsto \int_{-\infty}^{\infty} \mathrm{e}^{zx}\frac{\mathrm{e}^{-\frac{x^2}{2}}}{\sqrt{2\pi}}\,\mathrm{d}x$.
$h_1$ et $h_2$ sont holomorphes et égales sur $\mathbb{R}$ donc par prolongement analytique (voir CDHO), $h_1 = h_2$ sur $\mathbb{C}$.
Donc $h_1(it) = h_2(it)$ donc $\varphi_X(t) = \mathrm{e}^{-\frac{t^2}{2}}$.

**Proposition 3.5** Soit $X$ un vecteur aléatoire de dimension $d$ vérifiant que $E(|X|^p) < \infty$ où $p \geqslant 1$. Alors $\varphi_X$ admet des dérivées partielles continues jusqu'à l'ordre $p$ et :

$$\frac{\partial^p}{\partial t_{j_1}\dots\partial t_{j_p}}\varphi_X(t) = i^p E(X_{j_1}\dots X_{j_p}\mathrm{e}^{i\langle t,X\rangle})$$

 

*Remarque 3.5* *En particulier,* $E(X_{j_1}\dots X_{j_p}) = \frac{\partial^p}{\partial t_{j_1}\dots\partial t_{j_p}}\varphi_X(t)|_{t=0}$.

*De plus, si* $X \in L^2$, $\varphi_X(t) = 1 + i\langle t, E(X)\rangle - \frac{1}{2}t^*E(XX^*)t + o(|t|^2)$.

*Démonstration.* On dérive sous l'espérance, ce qui est possible car on domine aisément $\frac{\partial^p}{\partial t_{j_1}\dots\partial t_{j_p}}\mathrm{e}^{i\langle t,x\rangle}$ par $|x_{j_1}|\dots|x_{j_p}| \in L^1$. ∎

## 3.2 Propriété fondamentale de la fonction caractéristique

<u>THÉORÈME 3.2</u> $\varphi_X = \varphi_Y$ ssi $P_X = P_Y$.

*Démonstration dans le cas* $d = 1$.

- On a, pour tout $t$, $\displaystyle\int_{\mathbb{R}} \mathrm{e}^{itu}\mathrm{e}^{-\lambda|u|}\,\mathrm{d}u = \frac{2\lambda}{\lambda^2+t^2}$.

  Avec $t = X(\omega) - s$ et $s \in \mathbb{R}$, on a :

$$E\left(\int_{\mathbb{R}} \mathrm{e}^{i(X-s)u}\right) = E\left(\frac{2\lambda}{\lambda^2+(X-s)^2}\right)$$

$$\begin{aligned} E\left(\frac{2\lambda}{\lambda^2+(X-s)^2}\right) &\overset{\text{Fubini}}{=} \int_{\mathbb{R}} E(\mathrm{e}^{i(X-s)u}\mathrm{e}^{-\lambda|u|})\,\mathrm{d}u \\ &= \int_{\mathbb{R}} \mathrm{e}^{-isu-\lambda|u|} \underbrace{\int_{\mathbb{R}} \mathrm{e}^{ixu}P_X(\mathrm{d}x)}_{=\varphi_X(u)}\,\mathrm{d}u \end{aligned}$$

  Par hypothèse, $\varphi_X = \varphi_Y$ donc, en reprenant les calculs avec $Y$ au lieu de $X$, on a :

$$E\left(\frac{2\lambda}{\lambda^2+(X-s)^2}\right) = E\left(\frac{2\lambda}{\lambda^2+(Y-s)^2}\right)$$

  Soit $g$ continue à support compact.

$$\begin{aligned} \int_{\mathbb{R}} g(s)\int_{\mathbb{R}} \frac{2\lambda}{\lambda^2+(x-s)^2}P_X(\mathrm{d}x)\,\mathrm{d}s &\overset{\text{Fubini}}{=} \int_{\mathbb{R}}\int_{\mathbb{R}} \frac{2\lambda g(s)}{\lambda^2+(x-s)^2}\,\mathrm{d}sP_X(\mathrm{d}x) \\ &= 2\int_{\mathbb{R}} \frac{g(x-\lambda t)}{1+t^2}\,\mathrm{d}tP_X(\mathrm{d}x) \end{aligned}$$

  Ainsi, on a :

$$\int_{\mathbb{R}} P_X(\mathrm{d}x)\int_{\mathbb{R}} \frac{g(x-\lambda t)}{1+t^2}\,\mathrm{d}t = int_{\mathbb{R}}P_Y(\mathrm{d}x)\int_{\mathbb{R}} \frac{g(x-\lambda t)}{1+t^2}\,\mathrm{d}t$$

 

- On fait tendre $\lambda$ vers 0 et on obtient $E(g(X)) = E(g(Y))$ via la convergence dominée.
  En effet, $\int_{\mathbb{R}} \frac{g(x-\lambda t)}{1+t^2} \, \mathrm{d}t \to \pi g(x)$ quand $\lambda \to 0$. On a donc $\pi E(g(X)) = \pi E(g(Y))$.
- Soit $a, b \in \mathbb{R}$ et $g_n$ qui approche $1_{]a,b]}$ en étant non nulle sur $[a, b+\frac{1}{n}]$. On a $E(g_n(X)) = E(g_n(Y))$ donc par convergence dominée,

$$E(1_{]a,b]}(X)) = E(1_{]a,b]}(Y))$$

  Donc $P_X = P_Y$ sur tout intervalle $]a, b]$ donc, par classe monotone, $P_X = P_Y$. ■

**Proposition 3.6 (inversion de Fourier)** Soit $\varphi \in L^1$ la fonction caractéristique d'une variable $X$.

Alors $X$ est à densité continue, bornée :

$$f(x) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \mathrm{e}^{-i\langle t,x\rangle} \varphi(t) \, \mathrm{d}t$$

*Démonstration dans le cas $d = 1$.* Soit $\lambda > 0$.

$$\begin{aligned} 2\pi E(g(X)) &= \lim_{\lambda \to 0} \int_{\mathbb{R}} P_X(\mathrm{d}x) \int_{\mathbb{R}} \frac{2g(x-\lambda t)}{1+t^2} \, \mathrm{d}t \\ &= \lim_{\lambda \to 0} \int_{\mathbb{R}} g(s) \, \mathrm{d}s \int_{\mathbb{R}} \frac{2\lambda}{\lambda^2 + (x-s)^2} P_X(\mathrm{d}x) \\ &= \lim_{\lambda \to 0} \int_{\mathbb{R}} g(s) \, \mathrm{d}s \int_{\mathbb{R}} \mathrm{e}^{-isu-\lambda|u|} \varphi_X(u) \, \mathrm{d}u \\ &= \int_{\mathbb{R}} g(s) \, \mathrm{d}s \int_{\mathbb{R}} \mathrm{e}^{-isu} \varphi_X(u) \, \mathrm{d}u \text{ Par convergence dominée} \end{aligned}$$

Donc $f(s)$ est la densité de $X$ ■

**Fonction caractéristique de la loi de Laplace**

On a $f_X(x) = \frac{1}{2}\mathrm{e}^{-|x|}$.

 

$$\begin{aligned}
\varphi_X(t) &= E(\mathrm{e}^{itX}) \\
&= \int_{-\infty}^{\infty} \mathrm{e}^{itX} \frac{\mathrm{e}^{-|x|}}{2} \,\mathrm{d}x \\
&= \frac{1}{2}\int_{-\infty}^{0} \mathrm{e}^{(it+1)x} \,\mathrm{d}x + \frac{1}{2}\int_{0}^{\infty} \mathrm{e}^{(it-1)x} \,\mathrm{d}x \\
&= \frac{1}{2(it+1)} - \frac{1}{2(it-1)} \\
&= \frac{1}{1+t^2}
\end{aligned}$$

**Fonction caractéristique de la loi de Cauchy**

On a $f_Y(y) = \frac{1}{\pi(1+y^2)}$.

$$\varphi_Y(t) = E(\mathrm{e}^{itY}) = \int_{-\infty}^{\infty} \frac{1}{\pi(1+y^2)} \mathrm{e}^{ity} \,\mathrm{d}y$$

$\varphi_X \in L^1$ donc par l'inversion de Fourier,

$$\frac{\mathrm{e}^{-|x|}}{2} = \frac{1}{2\pi}\int_{\mathbb{R}} \mathrm{e}^{-itx} \varphi_X(t) \,\mathrm{d}t = \frac{1}{2} \underbrace{\int_{\mathbb{R}} \mathrm{e}^{it(-x)} f_Y(t) \,\mathrm{d}t}_{=\varphi_Y(-x)}$$

Donc $\varphi_Y(x) = \mathrm{e}^{-|x|}$.

**Proposition 3.7** Soit $(X, Y)$ un vecteur aléatoire.
$X \perp\!\!\!\perp Y$ ssi $\varphi_{(X,Y)}(s,t) = \varphi_X(s)\varphi_Y(t)$.

*Démonstration.*
$\Rightarrow$ Si $X \perp\!\!\!\perp Y$, on a :

$$\begin{aligned}
\varphi_{(X,Y)}(s,t) &= E\left(\exp\left(i\left\langle \begin{pmatrix} s \\ t \end{pmatrix}, \begin{pmatrix} X \\ Y \end{pmatrix} \right\rangle\right)\right) \\
&= E(\mathrm{e}^{i(sX+tY)}) \\
&= E(\mathrm{e}^{isX}\mathrm{e}^{itY}) \\
&= E(\mathrm{e}^{isX})E(\mathrm{e}^{itY}) \\
&= \varphi_X(s)\varphi_Y(t)
\end{aligned}$$

 

$\Leftarrow$ On a :

$$\begin{aligned}\iint e^{isx+ity}P_{(X,Y)}(\mathrm{d}x,\mathrm{d}y) &= \int e^{isx}P_X(\mathrm{d}x)\int e^{ity}P_Y(\mathrm{d}y)\\ &= \iint e^{i(sx+ty)}P_X(\mathrm{d}x)P_Y(\mathrm{d}y)\\ &= \iint e^{isx+ity}(P_X\otimes P_Y)(\mathrm{d}x,\mathrm{d}y)\end{aligned}$$

Donc $P_{(X,Y)} = P_X \otimes P_Y$. Le théorème de caractérisation assure que $X \perp\!\!\!\perp Y$. ∎

**Proposition 3.8** Si $X \perp\!\!\!\perp Y$, $\varphi_{X+Y} = \varphi_X\varphi_Y$.

*Démonstration.*

$$\begin{aligned}E(e^{it(X+Y)}) &= E(e^{itX}e^{itY})\\ &= E(e^{itX})E(e^{itY})\\ &= \varphi_X(t)\varphi_Y(t)\end{aligned}$$

## 3.3 Vecteurs gaussiens

**Proposition 3.9** $Y$ est une variable aléatoire gaussienne ssi il existe $a, b$ et $X \sim \mathcal{N}(0,1)$ tel que $Y = aX + b$.

**Définition 3.2** On note $\mathcal{N}_1 = \{\mathcal{N}(m,\sigma^2), m, \sigma\}$.

**Proposition 3.10** $\mathcal{N}_1$ est stable par multiplication par un scalaire.

**Définition 3.3** Un vecteur aléatoire $X$ de dimension $d$ est dit vecteur gaussien ssi pour tout $u \in \mathbb{R}^d$,

$$\langle u, X\rangle = \sum_{j=1}^{d} u_j X_j$$

est une variable aléatoire gaussienne.

Autrement dit, $X$ est un vecteur aléatoire gaussien ssi toute combinaison linéaire de ses coordonnées est une variable aléatoire gaussienne.

*Remarque 3.6*

- *Si $X$ est un vecteur gaussien alors $X_1, \ldots, X_d$ sont gaussiennes et dans $L^p$.*
- *Si $X_1, \ldots, X_d$ sont des variables gaussiennes indépendantes, alors le vecteur $(X_1, \ldots, X_d)$ est gaussien.*
- *On dit que le vecteur $X$ est gaussien ssi pour toute forme linéaire $h$, le loi de $h(X)$ appartient à $\mathcal{N}_1$.*
  *En effet, il existe $u$ tel que $h = \langle u, \cdot\rangle$.*

 

**Proposition 3.11** Si $X$ est un vecteur gaussien d'espérance $m \in \mathbb{R}^d$ et de matrice de covariance $K$, alors sa fonction caractéristique est donnée par :

$$\varphi_X(t) = \mathrm{e}^{i\langle t,m\rangle - \frac{t^*Kt}{2}}$$

*Démonstration.*

$$\varphi_X(t) = E\mathrm{e}^{i\langle t,X\rangle}) = \varphi_{\langle t,X\rangle}(1) = \mathrm{e}^{iE(\langle t,X\rangle) - \frac{1}{2}\operatorname{Var}(\langle t,X\rangle)}$$

De plus,

$$E(t^*X) = t^*E(X) = \langle t, E(X)\rangle = \langle t, m\rangle$$

et

$$\operatorname{Var}(t^*X) = \operatorname{Cov}(t^*X) = t^*\operatorname{Cov}(X)t = t^*Kt$$

D'où le résultat. ■

**Proposition 3.12** Soit $X$ un vecteur gaussien $d$-dimensionnel d'espérance $m$ et de matrice de covariance $K$.

Soit $A \in \mathfrak{M}_{k,d}(\mathbb{R})$ et $b \in \mathbb{R}^k$.

On note $Y = b + AX$. Alors $Y$ est un vecteur gaussien d'espérance $b + Am$ et de matrice de covariance $AKA^*$.

*Démonstration.* On montre que $Y$ est un vecteur gaussien.

Soit $s \in \mathbb{R}^k$.

$d^*Y = s^*(b + AX) = s^*b + (A^*s)^*X$ et $A^*s \in \mathbb{R}^d$ donc $(A^*s)^*X$ est gaussienne, ce qui conclut. ■

**Proposition 3.13** Soit $m \in \mathbb{R}^d$ et $K$ une matrice $d \times d$ symétrique positive. Alors il existe un vecteur gaussien d'espérance $m$ et de variance $K$.

*Démonstration.* Il existe $A \in \mathfrak{M}_d(\mathbb{R})$ tel que $K = AA^*$.

Soient $Y_1, \ldots, Y_d$ $d$ variables aléatoires indépendantes de même loi $\mathcal{N}(0,1)$.

$Y = (Y_1, \ldots, Y_d)$ est un vecteur gaussien d'espérance 0 et de matrice de covariance Id.

On pose $Y = m + AY$. Alors $X$ est le vecteur gaussien recherché. ■

**Proposition 3.14** Soit $X$ un vecteur gaussien de matrice de covariance $K$ et d'espérance $m$.

Les variables aléatoires $X_1, \ldots, X_d$ sont indépendantes ssi $K$ est diagonale.

*Démonstration.*

$\Rightarrow$ trivial

 

$\Leftarrow$ Si $K$ est diagonale, on a :

$$\begin{aligned}\varphi_X(t) &= E(\mathrm{e}^{i\langle t,X\rangle}) \\ &= \mathrm{e}^{i\langle t,m\rangle - \frac{t^*Kt}{2}} \\ &= \exp\left(i\sum_{j=1}^d t_j m_j - \frac{1}{2}\sum_{j=1}^d t_j^2\sigma_j^2\right)\end{aligned}$$

Soit $t^{(l)} = (0,\dots,0,t_l,0,\dots,0) \in \mathbb{R}^d$.

$$E(\mathrm{e}^{it_lX_l}) = \mathrm{e}^{it_lm_l - \frac{t_l^2\sigma_l^2}{2}}$$

Donc $\varphi_{X_l}(t_l) = \mathrm{e}^{it_lm_l - \frac{t_l^2\sigma_l^2}{2}}$ pour tout $l$.
D'où :

$$\begin{aligned}\varphi_X(t) &= \exp\left(i\sum_{j=1}^d t_j m_j - \frac{1}{2}\sum_{j=1}^d t_j^2\sigma_j^2\right) \\ &= \prod_{j=1}^d \mathrm{e}^{it_jm_j - \frac{t_j^2\sigma_j^2}{2}} \\ &= \prod_{j=1}^d \varphi_{X_j}(t_j)\end{aligned}$$

Donc $X_1,\dots,X_d$ sont indépendantes. ∎

**Proposition 3.15** Soit $X$ un vecteur gaussien d'espérance $m$ et de matrice de covariance $K$.

$X$ admet une densité ssi $K$ est inversible. Dans ce cas,

$$f_X(x) = \frac{1}{(2\pi)^{\frac{d}{2}}\sqrt{\det(K)}}\mathrm{e}^{-\frac{(x-m)^*K^{-1}(x-m)}{2}}$$

*Démonstration.*

**Lemme 3.2.1**
Soit $Z$ un vecteur aléatoire $d$-dimensionnel ayant une densité.

Soit $H \subset \mathbb{R}^d$ un sous-espace vectoriel tel que $\dim(H) < d$. Alors $P(Z \in H) = 0$.

*Démonstration.* Soit $H'$ un hyperplan contenant $H$.

Quitte à changer les coordonnées, on peut supposer que

$$H' = \{(x_1,\dots,x_d), x_d = 0\}$$

 

$$P(Z \in H) \leqslant P(Z \in H') = \int f_Z(x_1, \dots, x_d) 1_{\{x_d = 0\}} \, \mathrm{d}x_1 \dots \mathrm{d}x_d = 0$$

■

On a vu que $X$ et $m + AY$ ont la même loi avec $Y = (Y_1, \dots, Y_d)$ et $Y_1, \dots, Y_d$ indépendantes de même loi et avec $A$ telle que $K = AA^*$.

On a $\det(K) = \det^2(A)$ donc $A$ et $K$ sont inversibles ou non en même temps.

Si $A$ est non inversible, $\mathrm{Im}(A)$ est de dimension $< d$ donc, si $X$ est à densité, $X - m$ aussi et $P(X - m \in H) = 0$.

On a donc $P((m + AY) - m \in H) = 0$ donc $0 = P(AY \in H) = 1$. Donc $X$ n'est pas à densité.

Si $A$ est inversible, $m + AY$ est un $C^1$ difféomorphisme donc à densité donc $X$ est à densité.

De plus,

$$f_Y(y) = \prod_{j=1}^{d} \left( \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{y_j^2}{2}} \right) = \frac{1}{(2\pi)^{\frac{d}{2}}} \mathrm{e}^{-\frac{\langle y, y \rangle}{2}}$$

Notons $X' = m + AY$. On a $Y = A^{-1}(X' - m)$ et $x \mapsto A^{-1}(x - m)$ est de jacobienne $A^{-1}$ donc de jacobien $\frac{1}{\sqrt{\det(K)}}$.

On a aussi :

$$\begin{aligned} \langle y, y \rangle &= \langle A^{-1}(x - m), A^{-1}(x - m) \rangle \\ &= (A^{-1}(x - m))^* A^{-1}(x - m) \\ &= (x - m)^* (A^{-1})^* A^{-1} (x - m) \\ &= (x - m)^* K^{-1} (x - m) \end{aligned}$$

D'où la densité de $X$ dans le cas inversible. ■

 

# Chapitre 4

# Loi des grands nombres – Convergence

## 4.1 Loi des grands nombres

On a différents types de convergence :

- La convergence presque partout.
- La convergence en norme $L^p$.
- La convergence en probabilité :

$$X_n \to X \quad \text{ssi} \quad \forall \varepsilon > 0, \lim_{n \to +\infty} P(|X_n - X| \geqslant \varepsilon) = 0$$

*Remarque 4.1 La convergence presque sûre est équivalente à*

$$P(\limsup_{n \to +\infty} \{|X_n - X| \geqslant \varepsilon\} = 0$$

**Proposition 4.1** La convergence presque sûre implique la convergence en probabilité.

*Démonstration.* Si $X_n$ converge presque sûrement vers $X$ alors pour tout $\varepsilon > 0$,

$$\lim_{n \to +\infty} P\left(\bigcup_{k \geqslant n} \{|X_k - X| \geqslant \varepsilon\}\right) = 0$$

Or $\{|X_n - X| \geqslant \varepsilon\} \subset \bigcup_{k \geqslant n} \{|X_k - X| \geqslant \varepsilon\}$ donc $X_n$ converge en probabilité vers $X$. ∎

*Remarque 4.2 La réciproque est fausse.*

*On considère* $X_1 = 1_{[0,1]}$, $X_2 = 1_{[0,\frac{1}{2}]}$, $X_3 = 1_{[\frac{1}{2},1]}$, ...

*Si* $X_n = 1_{[\frac{j_n-1}{k_n},\frac{j_n}{k_n}]}$ *alors* $\lambda(|X_n| \geqslant \varepsilon) = \frac{1}{k_n} \to 0$ *donc* $X_n$ *converge en probabilité vers* $0$.

*Or pour tout* $\omega \in [0,1]$ *et* $k$ *entier, il existe* $j$ *tel que* $\omega \in [\frac{j-1}{k},\frac{j}{k}]$ *donc pour une infinité de* $n$, $X_n(\omega) \neq 0$.

*Donc* $X_n$ *ne converge pas presque sûrement vers* $0$.

**Proposition 4.2** Soit $X_n$ une suite de variables aléatoires réelles. Si $X_n$ converge presque sûrement (resp. en probabilité) vers $X$.

Soit $g : \mathbb{R} \to \mathbb{R}$ continue.

Alors $g(X_n)$ converge presque sûrement (resp. en probabilité) vers $g(X)$.

**Proposition 4.3 Lemme de Borel-Cantelli de convergence presque sûre**

(i) Si $\forall \varepsilon > 0, \sum_{n=1}^{\infty} P(|X_n - X| \geqslant \varepsilon) < \infty$ alors $X_n$ converge presque sûrement vers $X$.

(ii) Si $(X_n)_n$ est une suite de variables indépendantes, alors $X_n$ converge presque sûrement vers $c$ ssi $\forall \varepsilon > 0, \sum_{n=1}^{\infty} P(|X_n - c| \geqslant \varepsilon) < \infty$.

*Démonstration.*

(i) On pose $A_n = \{|X_n - X| \geqslant \varepsilon\}$.

On a $\sum_{n=1}^{\infty} P(A_n) < \infty$ donc par Borel-Cantelli, $P(\limsup A_n) = 0$.

Donc on a la convergence presque sûre de $X_n$ vers $X$.

(ii) Si $\sum_{n=1}^{\infty} P(A_n) = +\infty$ avec $A_n = \{|X_n - c| \geqslant \varepsilon\} = X_n^{-1}(]c-\varepsilon, c+\varepsilon[)^c$ des évènements indépendants.

Par Borel-Cantelli, $P(\limsup A_n) = 1$, d'où une contradiction car la convergence presque sûre implique $P(\limsup A_n) = 0$. ∎

**Proposition 4.4** $X_n$ converge en probabilité vers $X$ ssi :

$$\lim_{n\to+\infty} E\left(\frac{|X_n - X|}{1 + |X_n - X|}\right) = 0$$

**Proposition 4.5** $X_n$ est une suite de variables convergeant en probabilité vers $X$ ssi elle est de Cauchy en probabilité :

$$\forall \varepsilon > 0, \lim_{n,m\to+\infty} P(|X_n - X_m| \geqslant \varepsilon) = 0$$

*Démonstration.*

 

$\Rightarrow$ On a :

$$|X_n - X_m| \leqslant |X_n - X| + |X_m - X| \leqslant 2\frac{\varepsilon}{2}$$

Donc, $\{|X_n - X| < \frac{\varepsilon}{2}\} \cap \{|X_m - X| < \frac{\varepsilon}{2}\} \subset \{|X_n - X_m| < \varepsilon\}$.
Donc en passant au complémentaire,

$$P(|X_n - X_m| \geqslant \varepsilon) \leqslant P\left(|X_n - X| \geqslant \frac{\varepsilon}{2}\right) + P\left(|X_m - X| \geqslant \frac{\varepsilon}{2}\right)$$

Donc $\lim\limits_{n,m\to+\infty} P(|X_n - X_m| \geqslant \varepsilon) = 0$.

$\Leftarrow$ Nous allons montrer que sous cette hypothèse, il existe une sous suite $(X_{n_j})$ qui converge presque sûrement vers $X$.
La condition de Cauchy pour $\varepsilon = \frac{1}{2^j}$ est : pour tout $j$,

$$\lim_{n,m\to+\infty} P(|X_n - X_m| \geqslant \frac{1}{2^j}) = 0$$

On va construire la suite $n_j$ : $n_1 = 1$, et :

$$n_{j+1} = \min\left\{n > n_j, P\left(|X_r - X_s| \geqslant \frac{1}{2^j}\right) \leqslant \frac{1}{2^j}, \forall r, s \geqslant n\right\}$$

$n_j$ est croissante et $P(|X_{n_{j+1}} - X_{n_j}| \geqslant \frac{1}{2^j}) \leqslant \frac{1}{2^j}$.
Comme $\sum\limits_{j=1}^{\infty} P(|X_{n_{j+1}} - X_{n_j}| \geqslant \dfrac{1}{2^j}) < \infty$, on peut appliquer le lemme de Borel-Cantelli :

$$P\left(\liminf\left\{|X_{n_{j+1}} - X_{n_j}| < \frac{1}{2^j}\right\}\right) = 1$$

On pose $\Omega' = \liminf\{|X_{n_{j+1}} - X_{n_j}| < \frac{1}{2^j}\}$.
$\omega \in \Omega'$ ssi à partir d'un certain rang, $|X_{n_{j+1}} - X_{n_j}| < \frac{1}{2^j}$.
Autrement dit, à partir d'un certain rang, presque sûrement, $|X_{n_{j+1}} - X_{n_j}| < \frac{1}{2^j}$.
Soient $k > l$ deux entiers suffisamment grands,

$$|X_{n_k}(\omega) - X_{n_l}(\omega)| \leqslant \sum_{j=l}^{k-1} |X_{n_{j+1}}(\omega) - X_{n_j}(\omega)| \leqslant \sum_{j=l}^{\infty} \frac{1}{2^j} \to 0$$

quand $l \to +\infty$.
Donc $X_{n_j}$ satisfait une condition de Cauchy pour presque tout $\omega$.
Donc $X_{n_j}$ converge presque sûrement, donc en probabilité vers $X$.
On a $P(|X_n - X| \geqslant \varepsilon) \leqslant P(|X_{n_j} - X| \geqslant \frac{\varepsilon}{2}) + P(|X_n - X_{n_j}| \geqslant \frac{\varepsilon}{2}) \to 0$. ■

 

**Proposition 4.6** Soit $X_n$ une suite.

$X_n$ converge en probabilité vers $X$ ssi chaque sous-suite $(X_{n_j})_j$ contient une sous-suite $(X_{n_{j_l}})_l$ qui converge presque sûrement vers $X$.

*Démonstration.*

$\Rightarrow$ Supposons que $X_n$ converge vers $X$ en probabilités.
Soit $n_j$ une suite quelconque.
$X_{n_j}$ converge vers $X$ en probabilité donc $X_{n_j}$ est de Cauchy en probabilité donc il existe $n_{j_l}$ une sous-suite de $n_j$ telle que $X_{n_{j_l}}$ converge presque sûrement vers $X$.

$\Leftarrow$ Supposons par l'absurde que $X_n$ ne converge pas en probabilité vers $X$.
Alors il existe $X_{n_j}$ une suite extraite telle que $X_{n_j}$ ne converge pas en probabilité vers $X$.
Mais par hypothèse, $X_{n_j}$ contient une sous-suite $X_{n_{j_l}}$ qui converge presque sûrement donc en probabilité.

■

<u>Théorème 4.1</u> *Soit $(X_n)_n$ une suite de variables aléatoires indépendantes et identiquement distribuées intégrables.*

*Alors :*

$$\frac{X_1 + \ldots + X_n}{n} \to E(X_1)$$

*en probabilité (faible) et presque sûrement (forte).*

*Remarque 4.3 On a une réciproque :*

*Soit $X_n$ une suite de variables iid telles que $(\frac{S_n}{n})_n$ admet une limite presque sûre.*

*Alors les variables sont intégrables et la limite de $\frac{S_n}{n}$ est $E(X_1)$.*

*Preuve de la loi faible dans le cas $L^2$.* On note $m$ et $\sigma^2$ l'espérance et la variance communes.

On pose aussi $\overline{X_n} = \frac{X_1+\ldots+X_n}{n}$. On remarque que $E(\overline{X_n}) = m$.

Soit $\varepsilon > 0$ quelconque.

$$P(|\overline{X_n} - m| \geqslant \varepsilon) \leqslant \frac{\mathrm{Var}(\overline{X_n})}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2} \to 0$$

Donc on a la convergence en probabilité de $\overline{X_n}$ vers $m$. ■

*Preuve de la loi forte dans le cas $L^2$.* On a $\mathrm{Var}(S_n) = n\sigma^2$.

Donc $\mathrm{Var}(S_{n^4}) = n^4\sigma^2$.

Donc $\mathrm{Var}(\overline{X_{n^4}}) = \frac{\sigma^2}{n^4}$

On a alors pour $n > 1$, $P(|\overline{X_{n^4}} - m| \geqslant \frac{1}{n}) \leqslant n^2 \,\mathrm{Var}(\overline{X_{n^4}}) = \frac{\sigma^2}{n^2}$.

 

Donc c'est une série sommable et par Borel-Cantelli, $P(\liminf_n \{|\overline{X_{n^4}} - m| < \frac{1}{n}\}) = 1$

Donc presque sûrement, il existe $n_0 > 1$ tel que pour $n \geqslant n_0$, $|\overline{X_{n^4}} - m| < \frac{1}{n}$.

Il est loisible de supposer $X_n$ positive.

$(S_n)_n$ est une suite croissante et si $k \in [\![n^4, (n+1)^4]\!]$, on a :

$$\frac{S_k}{k} \leqslant \frac{S_{(n+1)^4}}{n^4} = \frac{S_{(n+1)^4}}{(n+1)^4}\left(\frac{n+1}{n}\right)^4 < \left(m + \frac{1}{n}\right)\left(\frac{n+1}{n}\right)^4$$

pour $n > n_0$.

Donc, presque sûrement, $\limsup \overline{X_k} \leqslant m$.

De même on minore et on trouve finalement $\limsup \overline{X_k} = m$ presque sûrement. ∎

## 4.2 Convergences de suites de variables aléatoires

**Définition 4.1** Soit $(X_n)_n$ une suite de $L^p$ et $X \in L^p$.

On dit que $X_n$ converge vers $X$ dans $L^p$ ssi $\lim_{n\to+\infty} E(|X_n - X|^p) = 0$.

**Proposition 4.7** Converger dans $L^p$ implique converger en probabilité.

*Démonstration.* Soit $\varepsilon > 0$.

$$P(|X_n - X| \geqslant \varepsilon) \leqslant P(|X_n - X|^p \geqslant \varepsilon^p) \leqslant \frac{E(|X_n - X|^p)}{\varepsilon^p} \to 0$$

D'où le résultat. ∎

*Remarque 4.4 La réciproque est fausse : prendre $X_n = 2^n 1_{]0,\frac{1}{n}[}$. On a convergence en probabilité mais pas dans $L^p$.*

**Définition 4.2** Soit $(X_n)_n$ une suite de variables aléatoires réelles intégrables est dite uniformément intégrable ssi

$$\lim_{R\to+\infty} \sup_{n\geqslant 1} E(|X_n| 1_{|X_n|>R}) = 0$$

**Exemple 4.1**

- Une suite constante est uniformément intégrable.
- Une suite dominée par $Y \in L^1$ est uniformément intégrable.
- Une famille finie de variables intégrables est uniformément intégrable.

 

**Proposition 4.8** Soit $(X_n)_n$ une suite de variables aléatoires intégrables.

$X_n$ converge dans $L^1$ ssi $X_n$ est de Cauchy dans $L^1$ ssi $X_n$ est uniformément intégrable et converge en probabilité.

**<u>Définition 4.3</u>** On dit qu'une suite $(X_n)_n$ converge en loi vers $X$ ssi pour tout $g$ continue bornée,

$$E(g(X_n)) \to E(g(X))$$

On dit aussi que la suite des probabilités $P_{X_n}$ converge faiblement (ou étroitement) vers $P_X$.

$$\begin{tikzcd}
\text{cv p.s.} \arrow[r, Rightarrow] & \text{cv proba} \arrow[r, Rightarrow] \arrow[l, dashed, bend left, "\text{sous-suites}"] \arrow[d, dashed, bend left, "\text{u.i.}"] & \text{cv loi} \arrow[l, dashed, bend right, "\text{cas particulier}"'] \\
& \text{cv } L^p \arrow[u, Rightarrow] &
\end{tikzcd}$$

**Proposition 4.9** Si $X_n$ converge en loi vers $X$ alors $h(X_n)$ converge en loi vers $h(X)$ pour tout $h$ continue.

**Proposition 4.10** Si $X_n$ converge en probabilité vers $X$ alors $X_n$ converge en loi vers $X$.

*Démonstration.*

$$\begin{aligned}
&|E(g(X_n)) - E(g(X))| \leqslant E(|g(X_n) - g(X)|) \\
&= E(|g(X_n) - g(X)|1_{|X_n - X| \leqslant \varepsilon}) + E(|g(X_n) - g(X)|1_{|X_n - X| > \varepsilon}) \\
&\leqslant \delta(\varepsilon) + 2\,\|g\|_\infty P(|X_n - X| > \varepsilon) \to 0
\end{aligned}$$

si $g$ est uniformément continue.

On a donc $\limsup |E(g(X_n)) - E(g(X))| \leqslant \delta(\varepsilon) \to 0$ quand $\delta \to 0$.

Si $g$ est continue bornée quelconque,

On pose $I_R = ]-R, R[$ il existe $R_\varepsilon$ tel que $1 - P_X(I_{R_\varepsilon}) < \varepsilon$.

On pose $u_\varepsilon$ la fonction qui vaut 1 sur $I_{R_\varepsilon}$, nulle sur $I^c_{R_\varepsilon+1}$ et affine sur les autres intervalles.

On pose $\widetilde{g} = u_\varepsilon g$. Elle est uniformément continue et bornée.

Donc on a $E(\widetilde{g}(X_n)) \to E(\widetilde{g}(X))$ et $E(u_\varepsilon(X_n)) \to E(u_\varepsilon(X))/$

Ainsi,

$$\begin{aligned}
|E(g(X_n)) - E(g(X))| \leqslant{}& E(|g(X_n) - \widetilde{g}(X_n)|) \\
&+ |E(\widetilde{g}(X_n)) - E(\widetilde{g}(X))| + E(|\widetilde{g}(X) - g(X)|)
\end{aligned}$$

Le deuxième terme tend vers 0 quand $n \to +\infty$.

 

On remarque que $\int (1-u_\varepsilon)(x)P_X(\mathrm{d}x) \leqslant 1 - P_X(R_\varepsilon) < \varepsilon$.
Donc

$$E(|g(1-u_\varepsilon)(X)|) \leqslant \|g\|_\infty \int (1-u_\varepsilon)(x)P_X(\mathrm{d}x) < \varepsilon \|g\|_\infty$$

De plus, on a $\int_{\mathbb{R}} (1-u_\varepsilon)(x)P_{X_n}(\mathrm{d}x) \to \int_{\mathbb{R}} (1-u_\varepsilon)(x)P_X(\mathrm{d}x)$.
On a ainsi pour $n$ assez grand :

$$\int (1-u_\varepsilon)(x)P_{X_n}(\mathrm{d}x) < 2\varepsilon$$

Ainsi, pour tout $\varepsilon$ et $n$ assez grand,

$$|E(g(X_n)) - E(g(X))| \leqslant \varepsilon(3\|g\|_\infty + 1)$$ ■

**Proposition 4.11** Si $X_n$ converge en loi vers une constante $c$ alors $X_n$ converge en probabilité vers $c$.

*Démonstration.* La convergence en probabilité est la convergence dans la métrique :

$$d(X,Y) = E\left(\frac{|X-Y|}{1+|X-Y|}\right)$$

Or $g_c = x \mapsto \frac{|x-c|}{1+|x-c|}$ est continue bornée donc, comme $X_n$ converge en loi, $E(g_c(X_n)) \to E(g_c(X)) = 0$ donc $X_n$ converge dans cette métrique. ■

Théorème 4.2 de convergence en loi *Soit $(X_n)_n$ une suite de variables aléatoires. Les conditions suivantes sont équivalentes :*

1. *$X_n \to X$ en loi*
2. *$\varphi_{X_n}(t) \to \varphi_X(t)$ sur $\mathbb{R}$*
3. *$F_{X_n}(t) \to F_X(t)$ pour tout point de continuité de $F_X$.*

*Remarque 4.5 $F_X$ est monotone donc le nombre de points de discontinuité est dénombrable, donc de mesure de Lebesgue nulle.*

**Proposition 4.12 Slutsky** Si $X_n \to X$ et $Y_n \to c$ en loi alors $(X_n, Y_n)$ converge en loi vers $(X,Y)$.

*Démonstration.*

$$\begin{aligned}
&|\varphi_{(X_n,Y_n)}(s,t) - \varphi_{(X,c)}(s,t)| = |E(\mathrm{e}^{isX_n}\mathrm{e}^{itY_n}) - E(\mathrm{e}^{isX}\mathrm{e}^{itc})| \\
&\leqslant |E(\mathrm{e}^{isX_n+itY_n}) - E(\mathrm{e}^{isX_n+itc})| + |E(\mathrm{e}^{isX_n+itc}) - E(\mathrm{e}^{isX+itc})| \\
&\leqslant E(|\mathrm{e}^{isX_n}(\mathrm{e}^{itY_n} - \mathrm{e}^{itc})|) + |\mathrm{e}^{itc}(E(\mathrm{e}^{isX_n}) - E(\mathrm{e}^{isX}))| \\
&= E(|\mathrm{e}^{itY_n} - \mathrm{e}^{itc}|) + |\varphi_{X_n}(s) - \varphi_X(s)|
\end{aligned}$$

 

On a :

$$\begin{aligned} E(\mathrm{e}^{itY_n} - \mathrm{e}^{itc}) &\leqslant |t|\varepsilon P(|Y_n - c| \leqslant \varepsilon) + 2P(|Y_n - c| > \varepsilon) \\ &\leqslant |t|\varepsilon + 2P(|Y_n - c| > \varepsilon) \to 0 \end{aligned}$$

Donc on a le résultat. ■

**Exemple 4.2** Soit $G$ gaussienne, $X_n = G$ et $Y_n = (-1)^n G$.

On a $X_n \to G$ et $Y_n \to G$ mais pas $(X_n, Y_n) \to (G, G)$ (en loi).

**Proposition 4.13** Si $X_n$ et $X$ sont des variables aléatoire discrètes à valeurs dans $\mathbb{Z}$.

Alors $X_n$ converge en loi vers $X$ ssi $P(X_n = k) \to P(X = k)$ pour tout $k \in \mathbb{Z}$.

**Proposition 4.14 Scheffé** Soit $(X_n)_n$ une suite de variables aléatoires ayant de densités $f_n$.

Si $f_n \to f$ presque partout, avec $\displaystyle\int_{\mathbb{R}} f(x)\,\mathrm{d}x = 1$, alors $X8n \to X$ en loi avec $X$ de densité $f$.

*Démonstration.*

$$\begin{aligned} |\varphi_{X_n}(t) - \varphi_X(t)| &\leqslant \left| \int \mathrm{e}^{itx} f_n(x)\,\mathrm{d}x - \int \mathrm{e}^{itx} f(x)\,\mathrm{d}x \right| \\ &= \left| \int \mathrm{e}^{itx}(f_n(x) - f(x))\,\mathrm{d}x \right| \\ &\leqslant \int |f_n(x) - f(x)|\,\mathrm{d}x \to 0 \end{aligned}$$

par convergence dominée. ■

*Démonstration du théorème de convergence en loi.*

1 ⇒ 2 $X_n$ converge en loi vers $X$ et $x \mapsto \cos(tx)$, $x \mapsto \sin(tx)$ sont continues bornées donc :

$$E(\cos(tX_n)) \to E(\cos(tX)) \text{ et } E(\sin(tX_n)) \to E(\sin(tX))$$

Donc $\varphi_{X_n}(t) \to \varphi_X(t)$

2 ⇒ 1 On suppose $\varphi_{X_n}(t) \to \varphi_X(t)$.

Si $g$ est continue à support compact telle que $g \in L^1$. On a :

$$g(x) = \frac{1}{2\pi} \int_{\mathbb{R}} \mathrm{e}^{-itx} \widehat{g}(t)\,\mathrm{d}t$$

 

On a donc :

$$\begin{aligned} E(g(X_n)) &= E\left(\frac{1}{2\pi}\int_{\mathbb{R}} \mathrm{e}^{-itX_n}\widehat{g}(t)\,\mathrm{d}t\right) \\ &= \frac{1}{2\pi}\int_{\mathbb{R}} \varphi_{X_n}(-t)\widehat{g}(t)\,\mathrm{d}t \\ &\to \frac{1}{2\pi}\int_{\mathbb{R}} \varphi_X(t)\widehat{g}(t)\,\mathrm{d}t = E(g(X)) \end{aligned}$$

On montre aussi que si $g$ est $C^2$ à support compact alors elle vérifie la condition précédente.
Il faut alors approcher les fonctions continues bornées par des fonctions $C^2$ à support compact.

$1 \Rightarrow 3$ Si $X_n \to X$ en loi, soit $\varepsilon > 0$ et $t$ un point de continuité de $F_X$.
On pose $g_\varepsilon : \mathbb{R} \to [0,1]$ valant 1 jusqu'à $t$, nulle après $t+\varepsilon$ et $C^2$ sur $[t, t+\varepsilon]$.
On a $E(g_\varepsilon(X_n)) \to E(g_\varepsilon(X))$ et :

$$E(g_\varepsilon(X)) \leqslant E(1_{]-\infty,t+\varepsilon]}(X)) = P(X \leqslant t+\varepsilon) = F_X(t+\varepsilon)$$

De plus, $E(g_\varepsilon(X_n)) \geqslant E(1_{]-\infty,t]}(X_n)) = F_{X_n}(t)$.
Donc $\limsup F_{X_n}(t) \leqslant \lim E(g_\varepsilon(X_n)) = E(g_\varepsilon(X)) \leqslant F_X(t+\varepsilon)$.
Donc $\limsup F_{X_n}(t) \leqslant F_X(t)$ (car $F_X$ continue en $t$).
On fait de même avec $h_\varepsilon$ définie comme $g$ avec $t-\varepsilon$ et $t$.
On obtient $\liminf F_{X_n}(t) \geqslant F_X(t-\varepsilon)$.
Avec $\varepsilon \to 0$, on a $F_{X_n}(t) \to F_X(t)$.

$3 \Rightarrow 1$ On suppose $g$ $C^1$ à support compact. On note $Q_n$ la loi de $X_n$.

$$\begin{aligned} E(g(X_n)) &= \int_{\mathbb{R}} g(x) Q_n(\mathrm{d}x) \\ &= \int_{\mathbb{R}} Q_n(\mathrm{d}x) \int_{-\infty}^{x} g'(t)\,\mathrm{d}t \\ &= \int_{\mathbb{R}} (1 - F_{X_n}(t)) g'(t)\,\mathrm{d}t \to \int_{\mathbb{R}} (1 - F_X(t)) g'(t)\,\mathrm{d}t \end{aligned}$$

On approche $g$ par des fonctions $C^1$ à support compact et ça marche. ∎

<u>Théorème 4.3</u> Lévy *Soit $(\varphi_n)_n$ une suite de fonctions caractéristiques associées à des variables aléatoires $X_n$ (ou à des lois $Q_n$).*

*Si $\varphi_n \to \varphi$ ponctuellement avec $\varphi$ continue en 0 alors $\varphi$ est la fonction caractéristique d'une variable aléatoire $X$ avec $X_n \to X$.*

**<u>Définition 4.4</u>** Une suite de variables $X_n$ est dite tendue ssi

$$\forall \varepsilon > 0, \exists R > 0, \sup_{n \geqslant 1} P(|X_n| > R) < \varepsilon$$

*Remarque 4.6 Si $X_n$ est tendue, il existe une sous-suite qui converge en loi.*

 

## 4.3 Théorème central limite

<u>THÉORÈME 4.4</u> CENTRAL LIMITE *Soit $X_n$ une suite de variables aléatoires iid de carrés intégrables/*

$$\frac{\sqrt{n}(\overline{X_n} - m)}{\sigma} \to G \sim \mathcal{N}(0,1)$$

*avec*

$$\overline{X_n} = \frac{X_1 + \ldots + X_n}{n} \to m$$

*Démonstration.* Il est loisible de supposer $m = 0$.

On note $\varphi$ la fonction caractéristique commune.

$\frac{\sqrt{n}\overline{X_n}}{\sigma} = \frac{S_n}{\sqrt{n}\sigma}$ et :

$$\begin{aligned}
\varphi_{\frac{S_n}{\sqrt{n}\sigma}}(t) &= E(\mathrm{e}^{it\frac{S_n}{\sqrt{n}\sigma}}) \\
&= E\left(\exp\left(i\frac{t}{\sqrt{n}\sigma}\sum_{k=1}^{n} X_k\right)\right) \\
&= \prod_{k=1}^{n} E(\mathrm{e}^{i\frac{t}{\sqrt{n}\sigma}X_k}) \\
&= E(\mathrm{e}^{i\frac{t}{\sqrt{n}\sigma}X_1})^n \\
&= \varphi\left(\frac{t}{\sqrt{n}\sigma}\right)^n
\end{aligned}$$

Comme $X_1 \in L^2$, $\varphi(u) = 1 - \frac{\sigma^2 u^2}{2} + o(u^2)$ donc

$$\varphi_{\frac{S_n}{\sqrt{n}\sigma}}(t) = \left(1 - \frac{t^2}{2n} + o(t^2)\right)^n$$

Donc pour $n \to +\infty$, on a $\varphi_{\frac{S_n}{\sqrt{n}\sigma}}(t) \to \mathrm{e}^{-\frac{t^2}{2}} = \varphi(t)$. ∎

 