

Théorie des probabilités — Cours

Ivan Lejeune

15 novembre 2024

Table des matières

Chapitre 1 — Bases de la théorie des probabilités	2
1 Espaces probabilisés	2
1.1 Probabilité	2
1.4 Exemples d'espaces probabilisés	3
2 Variables aléatoires	3
2.1 Loi d'une variable aléatoire	3
2.8 Lois usuelles	5
3 Moments d'une variable aléatoire	7
3.1 Espérance	7
3.5 Moments d'ordre p	9
3.9 Moments de lois usuelles	10
4 Fonctions associées à une variable aléatoire	10
4.1 Fonction de répartition	10
4.3 Fonction caractéristique	12
4.5 Fonction génératrice	13
Chapitre 2 — Indépendance	15
5 Indépendances d'événements	15
5.1 Conditionnement	15
5.2 Quelques formules	15
6 Indépendance de variables aléatoires	17
6.2 Critères d'indépendance	18
Cas discret	18
Avec les fonctions de répartition	19
Avec les fonctions caractéristiques	19
Cas des variables aléatoires à densité	19
7 Résultats asymptotiques.	19
7.1 Lemme de Borel-Cantelli	19
7.2 Loi du 0 – 1 de Kolmogorov	22
Chapitre 3 — Loi des grands nombres	24
8 Différents modes de convergence.	24
8.1 Convergences presque sûre et probabilité	24
8.4 Convergence dans $\mathcal{L}^p = \mathcal{L}^p(\Omega, \mathcal{F}, \mathbb{P})$	26

Chapitre 1 — Bases de la théorie des probabilités

1 Espaces probabilisés

1.1 Probabilité

Définition 1.2. Soit (Ω, \mathcal{F}) un espace mesurable. Une **mesure** sur (Ω, \mathcal{F}) est une application

$$\begin{aligned}\mu: \mathcal{F} &\rightarrow [0, +\infty] \\ A &\mapsto \mu(A)\end{aligned}$$

qui vérifie les propriétés suivantes :

1. $\mu(\emptyset) = 0$
2. μ est σ -additive, c'est-à-dire que pour toute suite $(A_n)_{n \in \mathbb{N}}$ d'éléments 2 à 2 disjoints de \mathcal{F} , on a

$$\mu\left(\bigcup_{n \in \mathbb{N}} A_n\right) = \sum_{n \in \mathbb{N}} \mu(A_n)$$

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

Si de plus $\mu(\Omega) = 1$, on dit que $(\Omega, \mathcal{F}, \mu)$ est un **espace probabilisé** et μ est une **probabilité**.

On notera alors $\mu = \mathbb{P}$.

Remarque. Comme $\mathbb{P}(\Omega) = 1$, une mesure de probabilité est une mesure dans $[0, 1]$. Un événement A est dit **presque sûr** si $\mathbb{P}(A) = 1$.

Exemples 1.3.

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

$$\delta_\omega(A) = \begin{cases} 1 & \text{si } \omega \in A \\ 0 & \text{sinon} \end{cases} = \mathbb{1}_A(\omega)$$

On vérifie facilement que c'est bien une probabilité.

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

$$\mathbb{P} = \frac{\mu(\cdot)}{\mu(\Omega)}$$

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

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

1.4 Exemples d'espaces probabilisés

Suivent quelques exemples classiques d'espaces probabilisés.

Exemples 1.5. cours a completer

2 Variables aléatoires

2.1 Loi d'une variable aléatoire

Définition 2.2. Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé et (E, \mathcal{E}) un espace mesurable. Une **variable aléatoire** est une application

$$X: \Omega \rightarrow E$$

mesurable. C'est-à-dire

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

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

Si $E = \mathbb{R}^d$ et $\mathcal{E} = \mathcal{B}(\mathbb{R}^d)$, on parle de **variable aléatoire vectorielle**.

Exemples 2.3.

▷ Lancer de deux dés.

On considère l'expérience aléatoire qui consiste à lancer deux dés équilibrés. Alors

$$\Omega = \{1, \dots, 6\}^2, \quad \mathcal{F} = \mathcal{P}(\Omega).$$

On s'intéresse à la somme des résultats obtenus et on définit

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

On munit l'ensemble d'arrivée de la tribu pleine.

X est une variable aléatoire car l'espace de départ est muni de la tribu pleine.

▷ Infinité de lancers d'un dé.

On considère l'expérience aléatoire qui consiste à lancer un dé équilibré une infinité de fois. Alors

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

On considère la tribu \mathcal{F} la plus petite tribu contenant les A_{x_1, \dots, x_k} . On s'intéresse au nombre de lancers jusqu'à l'apparition du premier 6. On définit

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

avec la convention $\inf \emptyset = +\infty$. On munit l'ensemble d'arrivée de la tribu pleine.

Pour $k \geq 1$, on a

$$\begin{aligned} Y^{-1}(\{k\}) &= \{\omega = (\omega_n)_{n \in \mathbb{N}^*} \mid \omega_1 \neq 6, \dots, \omega_{k-1} \neq 6, \omega_k = 6\} \\ &= \bigcup_{x_1, \dots, x_{k-1} \in \{1, \dots, 5\}} A_{x_1, \dots, x_{k-1}, 6} \in \mathcal{F} \end{aligned}$$

Par ailleurs,

$$\begin{aligned} Y^{-1}(\{+\infty\}) &= \{\omega = (\omega_n)_{n \in \mathbb{N}^*} \mid \forall n \in \mathbb{N}^*, \omega_n \neq 6\} \\ &= \bigcap_{k=1}^{+\infty} \bigcup_{x_1, \dots, x_k \in \{1, \dots, 5\}} A_{x_1, \dots, x_k} \in \mathcal{F} \end{aligned}$$

Comme $\mathbb{N} \cup \{+\infty\}$ est dénombrable, on en déduit que Y est une variable aléatoire.

▷ Bouteille à la mer.

On considère l'expérience aléatoire qui consiste à observer la position d'une bouteille à la mer. Alors

$$\Omega = \mathcal{C}^0([0, 1], \mathbb{R}^2)$$

On considère la tribu \mathcal{F} la plus petite tribu rendant mesurables les applications coordonnées

$$\begin{aligned} f_t: \Omega &\rightarrow \mathbb{R}^2 \\ \omega &\mapsto \omega(t) \end{aligned}$$

On s'intéresse à la position de la bouteille au temps $t = 1$. On définit

$$\begin{aligned} Z: \Omega &\rightarrow \mathbb{R}^2 \\ \omega &\mapsto \omega(1) \end{aligned}$$

Alors, par construction de la tribu \mathcal{F} , on a que Z est une variable aléatoire.

Définition 2.4. Soit X une variable aléatoire de $(\Omega, \mathcal{F}, \mathbb{P})$ dans (E, \mathcal{E}) . La **loi** de X est la mesure image de X par \mathbb{P} , définie par

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

Exemples 2.5.

▷ Infinité de lancers d'un dé.

On considère

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

La loi \mathbb{P}_Y de Y est une mesure de probabilité sur $\mathbb{N}^* \cup \{+\infty\}$.

Soit $k \in \mathbb{N}^*$. On a

$$\begin{aligned} \mathbb{P}_Y(\{k\}) &= \mathbb{P}(Y^{-1}(\{k\})) \\ &= \mathbb{P}(Y = k) \\ &= \mathbb{P}\left(\bigcup_{x_1, \dots, x_{k-1} \in \{1, \dots, 5\}} A_{x_1, \dots, x_{k-1}, 6}\right) \\ &= \sum_{x_1, \dots, x_{k-1}} \underbrace{\mathbb{P}(A_{x_1, \dots, x_{k-1}, 6})}_{= \frac{1}{6^k}} \\ &= \frac{5^{k-1}}{6^k} = \left(\frac{5}{6}\right)^{k-1} \frac{1}{6} \end{aligned}$$

Par ailleurs, on a vu à la fin de la section précédente que la probabilité de ne jamais obtenir de 6 est nulle :

$$\mathbb{P}_Y(\{+\infty\}) = \mathbb{P}(Y^{-1}(\{+\infty\})) = \mathbb{P}(Y = +\infty) = 0$$

On en déduit que la loi de Y est

$$\sum_{k=1}^{+\infty} \left(\frac{5}{6}\right)^{k-1} \frac{1}{6} \delta_k$$

Cette loi est appelée **loi géométrique** de paramètre $\frac{5}{6}$.

Définition 2.6 Variable aléatoire discrète. Une variable aléatoire X est dite **discrète** si X est à valeurs dans un ensemble E au plus dénombrable. On prend alors $\mathcal{E} = \mathcal{P}(E)$ et si $A \in \mathcal{E}$, on a

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

La loi \mathbb{P}_X de X est alors entièrement déterminée par les quantités $p_x = \mathbb{P}(X = x)$ pour tout $x \in E$:

$$\mathbb{P}_X(A) = \sum_{x \in E} p_x \delta_x$$

Définition 2.7 Variable aléatoire à densité. Une variable aléatoire X à valeurs dans $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ est dite **à densité** par rapport à la mesure de Lebesgue λ_d si il existe une fonction mesurable

$$f: \mathbb{R}^d \rightarrow [0, +\infty[$$

telle que $\mathbb{P}_X = f \lambda_d$:

$$\forall A \in \mathcal{B}(\mathbb{R}^d), \mathbb{P}_X(A) = \int_A f(x) d\lambda_d(x)$$

Il faut que f vérifie

$$\int_{\mathbb{R}^d} f(x) d\lambda_d(x) = 1$$

Par exemple, si $d = 1$, on a

$$\mathbb{P}_X([a, b]) = \int_a^b f(x) d\lambda_1(x)$$

On notera souvent $f_X = f$ et on appelle cette fonction la **densité** de X .

2.8 Lois usuelles

▷ Lois discrètes :

Loi uniforme sur un ensemble fini $\{x_1, \dots, x_n\}$.

Soit $E = \{x_1, \dots, x_n\}$ un ensemble fini.

Une variable aléatoire X suit une loi uniforme sur E , notée $X \sim \mathcal{U}(E)$, si sa loi est

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

c'est-à-dire

$$\mathbb{P}_X(A) = \mathbb{P}(X \in A) = \frac{\text{card}(A)}{n}$$

▷ Loi de Bernoulli.

Une variable aléatoire X suit une loi de Bernoulli de paramètre $p \in [0, 1]$, notée $X \sim \mathcal{B}(p)$, si sa loi est

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

c'est-à-dire

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

▷ Loi binomiale.

Une variable aléatoire X suit une loi binomiale de paramètres $n \in \mathbb{N}^*$ et $p \in [0, 1]$, notée $X \sim \mathcal{B}(n, p)$, si sa loi est

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

cela correspond au nombre de succès dans n répétitions d'une expérience de Bernoulli de paramètre p (de manière indépendante).

▷ Loi géométrique.

Une variable aléatoire X suit une loi géométrique de paramètre $p \in]0, 1]$, notée $X \sim \mathcal{G}(p)$, si sa loi est

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

c'est-à-dire

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

cela correspond au nombre de répétitions d'une expérience de Bernoulli de paramètre p avant le premier succès.

▷ Loi de Poisson.

Une variable aléatoire X suit une loi de Poisson de paramètre $\theta > 0$, notée $X \sim \mathcal{P}(\theta)$, si sa loi est

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

c'est-à-dire

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

cela correspond au nombre d'événements rares dans un intervalle de temps donné.

▷ Lois à densité sur \mathbb{R}^d :

Loi uniforme sur un ensemble A de \mathbb{R}^d .

Soit $A \in \mathcal{B}(\mathbb{R}^d)$, telle que $0 < \lambda_d(A) < +\infty$. Une variable aléatoire X suit une loi uniforme sur A , notée $X \sim \mathcal{U}(A)$, si sa loi est

$$\mathbb{P}_X = \frac{1}{\lambda_d(A)} \mathbb{1}_A$$

c'est-à-dire \mathbb{P}_X admet la densité constante $\frac{1}{\lambda_d(A)} \mathbb{1}_A$. Autrement dit, si $B \in \mathcal{B}(\mathbb{R}^d)$, on a

$$\mathbb{P}_X(B) = \int_B \frac{1}{\lambda_d(A)} \mathbb{1}_A(x) d\lambda_d(x) = \frac{\lambda_d(A \cap B)}{\lambda_d(A)}$$

dans le cas $d = 1$ et $A = [a, b]$, la densité est $f(x) = \frac{\mathbb{1}_{[a,b]}(x)}{b-a}$.

▷ Loi exponentielle.

Une variable aléatoire X suit une loi exponentielle de paramètre $\theta > 0$, notée $X \sim \mathcal{E}(\theta)$, si sa loi est

$$\mathbb{P}_X = \theta e^{-\theta x} \lambda_1 \mathbb{1}_{\mathbb{R}^+}$$

c'est-à-dire \mathbb{P}_X admet la densité $f_X(x) \theta e^{-\theta x} \mathbb{1}_{\mathbb{R}^+}$ par rapport à la mesure de Lebesgue λ_1 . Autrement dit, si $A \in \mathcal{B}(\mathbb{R})$, on a

$$\mathbb{P}_X(A) = \int_A \theta e^{-\theta x} \mathbb{1}_{\mathbb{R}^+}(x) d\lambda_1(x)$$

dans le cas $d = 1$. Cette loi vérifie la propriété de l'absence de mémoire, c'est-à-dire

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

▷ Loi normale ou gaussienne.

Une variable aléatoire X suit une loi normale de paramètres $\mu \in \mathbb{R}$ et $\sigma > 0$, notée $X \sim \mathcal{N}(\mu, \sigma^2)$, si sa loi est

$$\mathbb{P}_X = f_X \lambda_1$$

où f_X est la densité de la loi normale, donnée par

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Autrement dit, si $A \in \mathcal{B}(\mathbb{R})$, on a

$$\mathbb{P}_X(A) = \int_A \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} d\lambda_1(x)$$

remarque, la loi $\mathcal{N}(., .)$, c'est-à-dire avec la densité

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

est appelée loi normale standard.

3 Moments d'une variable aléatoire

On considère dans la suite $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé

3.1 Espérance

Définition 3.2. Soit $X: \Omega \rightarrow \mathbb{R}$ une variable aléatoire réelle. On appelle **espérance** de X la quantité

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

qui est bien définie si X est positive ou si $X \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$.

On définit de même l'espérance d'une variable aléatoire $X: \Omega \rightarrow \mathbb{C}$. Si $X = (X_1, \dots, X_d)$ est un vecteur aléatoire, alors on pose

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

si $\mathbb{E}[X_i]$ existe pour tout $i \in \{1, \dots, d\}$.

Remarque.

1. On interprète $\mathbb{E}[X]$ comme la valeur moyenne de X .
2. Si $\mathbb{E}[X] = 0$, on dit que X est **centrée**.
3. Si $X = \mathbb{1}_A$, alors $\mathbb{E}[X] = \mathbb{P}(A)$.
4. On omettra le cas où X est positive et $\mathbb{E}[X] = +\infty$.

Proposition. Soient X et Y deux variables aléatoires admettant une espérance et $a, b \in \mathbb{R}$. Alors

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

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

Proposition. Soit $X: \Omega \rightarrow E$ une variable aléatoire à valeurs dans un espace mesurable (E, \mathcal{E}) . Soit $h: E \rightarrow \mathbb{R}$ une fonction mesurable positive. Alors, $h \circ X$ est une variable aléatoire et

$$\begin{aligned} \mathbb{E}[h(X)] &= \int_{\Omega} h(X(\omega)) d\mathbb{P}(\omega) \\ &= \int_E h(x) d\mathbb{P}_X(x) \end{aligned}$$

Si $h: E \rightarrow \mathbb{R}$ et mesurable (pas forcément positive), alors $h \in \mathcal{L}^1(E, \mathcal{E}, \mathbb{P}_X)$ si et seulement si $h \circ X \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ et l'égalité précédente reste vraie.

Démonstration. Vue en théorie de la mesure.
On vérifie l'égalité :

- pour les fonctions indicatrices,
- pour les fonctions étagées,
- pour les fonctions positives,
- pour les fonctions intégrables

□

Remarque. En particulier, si $E = \mathbb{R}$ et $h = \text{Id}_{\mathbb{R}}$, on obtient

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

à condition que cette quantité existe.

Exemples 3.3. Loi de Bernoulli

$$\begin{aligned}\mathbb{P}_X &= (1-p)\delta_0 + p\delta_1 \\ \mathbb{E}[X] &= \int_{\mathbb{R}} x d\mathbb{P}_X(x) \\ &= \int_{\mathbb{R}} x d((1-p)\delta_0 + p\delta_1)(x) \\ &= (1-p) \underbrace{\int_{\mathbb{R}} x d\delta_0(x)}_{=0} + p \underbrace{\int_{\mathbb{R}} x d\delta_1(x)}_{=1} \\ &= p\end{aligned}$$

Si X suit une loi discrète

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

alors son espérance vaut

$$\begin{aligned}\mathbb{E}[X] &= \int_{\mathbb{R}} x d\mathbb{P}_X \\ &= \sum_{k=1}^{\infty} p_k \underbrace{\int_{\mathbb{R}} x d\delta_{x_k}}_{=x_k} \\ &= \sum_{k=1}^{\infty} p_k x_k\end{aligned}$$

Proposition. Soit $X: \Omega \rightarrow E$ une variable aléatoire. La loi de X est caractérisée par les quantités $\mathbb{E}[h(X)]$ où $h: E \rightarrow \mathbb{R}$ décrit l'ensemble des fonctions mesurables bornées. C'est-à-dire, si X et X' sont deux variables aléatoires vérifiant

$$\mathbb{E}[h(X)] = \mathbb{E}[h(X')]$$

pour toute fonction h mesurable bornée, alors X et X' ont la même loi.

Démonstration. Soient X et X' telles que $\mathbb{E}[h(X)] = \mathbb{E}[h(X')]$ pour toute fonction h mesurable bornée.

Soit $A \in \mathcal{E}$. On pose $h = \mathbb{1}_A$ et alors

$$\mathbb{P}(X \in A) = \mathbb{E}[\mathbb{1}_A(X)] = \mathbb{E}[h(X)] = \mathbb{E}[h(X')] = \mathbb{E}[\mathbb{1}_A(X')] = \mathbb{P}(X' \in A)$$

Donc X et X' ont la même loi. □

Exemples 3.4. On considère une variable aléatoire X de loi uniforme sur $]0, 1[$ et on pose $Y = -\ln(X)$. Déterminons la loi de Y .

Soit h une fonction mesurable bornée. Alors

$$\begin{aligned}\mathbb{E}[h(Y)] &= \mathbb{E}[h(-\ln(X))] \\ &= \int_{\mathbb{R}} h(-\ln(x)) d\mathbb{P}_X(x) \\ &= \int_{\mathbb{R}} h(-\ln(x)) \underbrace{\mathbb{1}_{]0,1[}(x) d\lambda_1(x)}_{=\mathbb{P}_X} \\ &= \int_{\mathbb{R}} h(y) \underbrace{\mathbb{1}_{]0,\infty[}(y) e^{-y} d\lambda_1(y)}_{=\mathbb{P}_X} \quad \text{avec } y = -\ln(x)\end{aligned}$$

D'après la proposition précédente, la loi de Y est

$$\mathbb{1}_{]0, \infty[}(y) e^{-y} \lambda_1(y)$$

c'est-à-dire que Y suit une loi exponentielle de paramètre 1.

3.5 Moments d'ordre p

Pour $p \in [1, \infty[$, on définit l'espace $\mathcal{L}^p(\Omega, \mathcal{F}, \mathbb{P})$ comme l'ensemble des variables aléatoires vérifiant

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

On a bien que $\|\cdot\|_p$ est une norme qui rend $\mathcal{L}^p(\Omega, \mathcal{F}, \mathbb{P})$ complet (toute suite de Cauchy converge). Dans le cas $p = \infty$, on pose

$$\|X\|_{\infty} = \sup_{\omega \in \Omega} |X(\omega)| = \lim_{p \rightarrow \infty} \|X\|_p$$

Donc $\mathcal{L}^{\infty}(\Omega, \mathcal{F}, \mathbb{P})$ est l'ensemble des variables aléatoires bornées.

Définition 3.6. Soit $p \in [1, \infty[$ et X une variable aléatoire dans $\mathcal{L}^p(\Omega, \mathcal{F}, \mathbb{P})$. Le **moment d'ordre p** de X est

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

Remarque.

1. Le moment d'ordre 1 correspond à l'espérance.
2. Si $p \leq q$ alors $\mathcal{L}^q(\Omega, \mathcal{F}, \mathbb{P}) \subset \mathcal{L}^p(\Omega, \mathcal{F}, \mathbb{P})$
En particulier, si X admet un moment d'ordre q alors X admet un moment d'ordre p pour tout $p \leq q$

Proposition. Soit X une variable aléatoire réelle positive et $p \in [1, \infty[$. Alors, pour tout $t > 0$, on a

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

Cette inégalité permet de contrôler le comportement à l'infini de X .

Démonstration. On part de l'inégalité

$$t^p \mathbb{1}_{\{X \geq t\}} \leq X^p$$

On applique l'espérance

$$\begin{aligned} t^p \mathbb{E}[\mathbb{1}_{\{X \geq t\}}] &\leq \mathbb{E}[X^p] \\ \mathbb{P}(X \geq t) &\leq \frac{\mathbb{E}[X^p]}{t^p} \end{aligned}$$

□

Le cas $\mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$:

On rappelle que $\mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$ est muni d'un produit scalaire

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

La norme associée est

$$\|X\|_2 = \sqrt{\langle X, X \rangle} = \sqrt{\int_{\Omega} X(\omega)^2 d\mathbb{P}(\omega)}$$

Définition 3.7. La **variance** d'une variable aléatoire réelle X est le moment d'ordre 2 de $X - \mathbb{E}[X]$, soit

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

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

On définit **l'écart-type** par

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

Remarque.

1. En développant le carré dans la définition de la variance, on obtient

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

2. On a $\text{Var}(X) = 0$ si et seulement si X est constante.

Proposition. Soit X une variable aléatoire dans $\mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$. Alors, pour tout $t > 0$, on a

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

Définition 3.8. Soient X et Y deux variables aléatoires dans $\mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$. On définit la **covariance** de X et Y par

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

où on rappelle que

$$\langle U, V \rangle = \mathbb{E}[UV] = \int_{\Omega} U(\omega)V(\omega) d\mathbb{P}(\omega)$$

Par les propriétés du produit scalaire dans $\mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$, on en déduit que la covariance est symétrique, bilinéaire et positive :

$$\text{COV}(X, X) = \text{Var}(X) \geq 0$$

Par ailleurs, on a aussi la relation de Pythagore :

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

3.9 Moments de lois usuelles

Sous forme de tableau :

Loi	Espérance	Variance
$\mathcal{U}(\{1, \dots, n\})$	$\frac{n+1}{2}$	$\frac{n^2-1}{12}$
$\mathcal{B}(p)$	p	$p(1-p)$
$\mathcal{B}(n, p)$	np	$np(1-p)$
$\mathcal{G}(p)$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
$\mathcal{P}(\theta)$	θ	θ
$\mathcal{U}([a, b])$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
$\mathcal{E}(\theta)$	$\frac{1}{\theta}$	$\frac{1}{\theta^2}$
$\mathcal{N}(\mu, \sigma^2)$	μ	σ^2

4 Fonctions associées à une variable aléatoire

4.1 Fonction de répartition

Définition 4.2. Soit X une variable aléatoire réelle. La **fonction de répartition** de X est la fonction

$$F_X: \mathbb{R} \rightarrow [0, 1], \quad t \mapsto \mathbb{P}(X \leq t)$$

Exemple. Si X est une variable réelle de densité f_X par rapport à la mesure de Lebesgue ($\mathbb{P}_X = f_X \lambda_1$), alors

$$\mathbb{P}_X(A) = \int_A f_X(t) d\lambda_1(t)$$

Donc, la fonction de répartition de X est donnée par

$$F_X(t) = \mathbb{P}_X([-\infty, t]) = \int_{-\infty}^t f_X(u) d\lambda_1(u)$$

Par exemple, si $X \sim \mathcal{E}(\theta)$, c'est-à-dire

$$f_X(t) = \theta e^{-\theta t} \mathbb{1}_{\mathbb{R}_+}(t)$$

alors

$$\begin{aligned} F_X(t) &= \int_{-\infty}^t \theta e^{-\theta u} \mathbb{1}_{\mathbb{R}_+}(u) d\lambda_1(u) \\ &= \begin{cases} 0 & \text{si } t < 0 \\ [-e^{-\theta u}]_0^t & \text{si } t \geq 0 \end{cases} \\ &= (1 - e^{-\theta t}) \mathbb{1}_{\mathbb{R}_+}(t) \end{aligned}$$

De manière générale, l'égalité $F_X(t) = \int_{-\infty}^t f_X(u) d\lambda_1(u)$ assure que F_X est dérivable λ_1 -p.p, avec $F'_X(t) = f_X(t)$ pour λ_1 -presque tout $t \in \mathbb{R}$.

Exemple. On considère $X \sim \mathcal{B}(p)$ et $\mathbb{P}_X = (1-p)\delta_0 + p\delta_1$. Alors, la fonction de répartition de X est donnée par

$$\begin{aligned} F_X(t) &= \mathbb{P}_X([-\infty, t]) \\ &= (1-p)\delta_0([-\infty, t]) + p\delta_1([-\infty, t]) \\ &= \begin{cases} 0 & \text{si } t < 0 \\ 1-p & \text{si } 0 \leq t < 1 \\ 1 & \text{si } t \geq 1 \end{cases} \end{aligned}$$

Plus généralement, si X est une variable aléatoire discrète de loi $\mathbb{P}_X = \sum_{k \in E} p_k \delta_{x_k}$, avec les x_k ordonnés, alors

$$F_X(t) = \sum_{k=0}^{\infty} p_k \delta_{x_k}([-\infty, t]) = \sum_{k=0}^{\infty} p_k \mathbb{1}_{[x_k, +\infty[}(t)$$

Proposition. Soit X une variable aléatoire réelle de fonction de répartition F_X . Alors

1. F_X est croissante, continue à droite et admet des limites à gauche (càdlàg)
2. Les limites de F_X en $-\infty$ et $+\infty$ sont

$$\lim_{t \rightarrow \infty} F_X(t) = 1, \quad \lim_{t \rightarrow -\infty} F_X(t) = 0$$

3. En posant

$$F_X(t-) = \lim_{\substack{\varepsilon \rightarrow 0 \\ \varepsilon > 0}} F_X(t - \varepsilon)$$

On a

$$F_X(t-) = \mathbb{P}(X < t) = \mathbb{P}(X \leq t)$$

et donc

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

Démonstration. Admise (ou a faire en exercice)

□

Proposition. La fonction de répartition d'une variable aléatoire X caractérise sa loi. C'est à dire que deux variables aléatoires ayant la même fonction de répartition ont la même loi.

Démonstration. Si $F_X = F_Y$ alors les mesures de probabilités \mathbb{P}_X et \mathbb{P}_Y associées à X et Y coïncident sur les ensembles de la forme $[-\infty, t]$. Ces ensembles engendrent la tribu borélienne de \mathbb{R} et sont stables par intersection finie, donc le lemme de classes monotones assure que $\mathbb{P}_X = \mathbb{P}_Y$. \square

4.3 Fonction caractéristique

Définition 4.4. Soit X une variable aléatoire réelle. La **fonction caractéristique** de X est la fonction

$$\begin{aligned}\varphi_X: \mathbb{R} &\rightarrow \mathbb{C} \\ t &\mapsto \mathbb{E}[e^{itX}] = \int_{\Omega} e^{itX(\omega)} d\mathbb{P}(\omega)\end{aligned}$$

Remarque.

1. La fonction caractéristique est bien définie car

$$\int_{\Omega} |e^{itX(\omega)}| d\mathbb{P}(\omega) = \int_{\Omega} d\mathbb{P}(\omega) = 1$$

2. La fonction caractéristique correspond à un signe près à la transformée de Fourier de la variable aléatoire X .

Exemple.

- Soit $X \sim \mathcal{B}(p)$. Alors

$$\varphi_X(t) = \mathbb{E}[e^{itX}] = (1-p)e^{it0} + pe^{it1} = 1-p+pe^{it}$$

- Soit X une variable aléatoire discrète de loi $\mathbb{P}_X = \sum_{k \in E} p_k \delta_{x_k}$. Alors

$$\varphi_X(t) = \mathbb{E}[e^{itX}] = \sum_{k \in E} p_k e^{itx_k}$$

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

$$\begin{aligned}\varphi_X(t) &= \mathbb{E}[e^{itX}] \\ &= \int_{\mathbb{R}} e^{itx} \frac{1}{b-a} \mathbb{1}_{[a,b]}(x) d\lambda_1(x) \\ &= \left[\frac{e^{itx}}{it} \right]_a^b \\ &= \frac{e^{itb} - e^{ita}}{it(b-a)} \quad \text{si } t \neq 0\end{aligned}$$

et pour $t = 0$, on a

$$\varphi_X(0) = \mathbb{E}[e^{i0X}] = \mathbb{E}[1] = 1$$

- Soit $X \sim \mathcal{N}(0, 1)$. Alors

$$\varphi_X(t) = \mathbb{E}[e^{itX}] = \int_{\mathbb{R}} e^{itx} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} d\lambda_1(x) = e^{-\frac{t^2}{2}}$$

cf. TD4 pour le calcul de l'intégrale.

Proposition. Soit X une variable aléatoire réelle. Si X admet un moment d'ordre $p \in \mathbb{N}$, alors

la fonction caractéristique φ_X est de classe \mathcal{C}^p et

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

Démonstration. La fonction $g : (t, \omega) \mapsto e^{itX(\omega)}$ est intégrable par rapport à ω , dérivable par rapport à t et vérifie

$$\left| \frac{\partial g}{\partial t}(t, \omega) \right| = |iX(\omega)e^{itX(\omega)}| = |X(\omega)|$$

Donc si X admet un moment d'ordre 1, alors

$$\left| \frac{\partial g}{\partial t}(t, \omega) \right| = |X(\omega)|$$

qji est intégrable, donc le théorème de dérivation sous le signe intégral assure que

$$\varphi'_X(t) = \int_{\Omega} iX(\omega)e^{itX(\omega)} d\mathbb{P}(\omega)$$

en particulier, en $t = 0$, on a

$$\varphi'_X(0) = i\mathbb{E}[X]$$

On raisonne ensuite par récurrence sur p pour les autres moments. \square

Théorème. La fonction caractéristique φ_X caractérise la loi de X . C'est-à-dire que si $\varphi_X = \varphi_Y$ alors X et Y ont la même loi.

Démonstration. Admise, elle est basée sur l'injectivité de la transformée de Fourier de mesures. \square

4.5 Fonction génératrice

On considère ici des variables aléatoires à valeurs dans \mathbb{N} .

Définition 4.6. Soit X une variable aléatoire à valeurs dans \mathbb{N} . La **fonction génératrice** de X est la série entière

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

Comme $\sum_{n=0}^{\infty} \mathbb{P}(X = n) = 1$, on en déduit que le rayon de convergence de cette série entière est au moins 1, donc G_X est définie sur $[-1, 1]$ et continue sur $[-1, 1]$. Par ailleurs, G_X est de classe \mathcal{C}^∞ sur $] -1, 1[$ avec

$$\mathbb{P}(X = n) = \frac{G_X^{(n)}(0)}{n!}$$

Donc G_X caractérise la loi de X .

En dérivant G_X terme à terme, on obtient

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

On prend une suite $(z_k)_{k \in \mathbb{N}}$ qui croît vers 1. Alors, la suite des fonctions $(X z_k^{X-1})_k$ est croissante, positive et donc par le théorème de convergence monotone, on a

$$\mathbb{E}\left[\lim_{k \rightarrow \infty} X z_k^{X-1}\right] = \lim_{k \rightarrow \infty} \underbrace{\mathbb{E}[X z_k^{X-1}]}_{G'_X(z_k)}$$

d'où

$$\mathbb{E}[X] = \lim_{\substack{z \rightarrow 1 \\ z < 1}} G'_X(z)$$

On peut noter $G'_X(1-) = \lim_{\substack{z \rightarrow 1 \\ z < 1}} G'_X(z)$.

Bref, si X admet un moment d'ordre 1, alors $G'_X(1-) = \mathbb{E}[X]$.

Plus généralement, on peut montrer que

$$G_X^{(k)}(1-) = \mathbb{E}[X(X-1)\dots(X-k+1)]$$

pour tout $k \in \mathbb{N}$. Cela permet de déterminer les moments de X .

Exemple. \triangleright Soit $X \sim \mathcal{B}(p)$. Alors

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

et donc

$$G'_X(z) = p$$

donc

$$\mathbb{E}[X] = G'_X(1) = p$$

\triangleright Soit $X \sim \mathcal{B}(n, p)$. Alors

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

donc

$$G'_X(z) = np(pz + 1 - p)^{n-1}$$

et donc

$$\mathbb{E}[X] = G'_X(1) = np$$

\triangleright Soit $X \sim \mathcal{G}(p)$. Alors

$$\begin{aligned} G_X(z) = \mathbb{E}[z^X] &= \sum_{k=1}^{\infty} p(1-p)^{k-1} z^k \\ &= zp \sum_{k=1}^{\infty} (1-p)^{k-1} z^{k-1} \\ &= \frac{zp}{1 - (1-p)z} \end{aligned}$$

et

$$\mathbb{E}[X] = G'_X(1) \rightarrow \text{exercice}$$

\triangleright Soit $X \sim \mathcal{P}(\theta)$. Alors

$$\begin{aligned} G_X(z) = \mathbb{E}[z^X] &= \sum_{k=0}^{\infty} e^{-\theta} \frac{\theta^k}{k!} z^k \\ &= e^{-\theta} \sum_{k=0}^{\infty} \frac{(\theta z)^k}{k!} \\ &= e^{-\theta} e^{\theta z} \\ &= e^{\theta(z-1)} \end{aligned}$$

donc

$$G'_X(z) = \theta e^{\theta(z-1)}$$

et donc

$$\mathbb{E}[X] = G'_X(1) = \theta$$

Chapitre 2 — Indépendance

On considère ici $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé.

5 Indépendances d'événements

5.1 Conditionnement

Si $A \in \mathcal{F}$ est un événement tel que $\mathbb{P}(A) > 0$, alors la **probabilité conditionnelle de $B \in \mathcal{F}$ sachant A** est définie par

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

alors, l'application

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

est une probabilité sur (Ω, \mathcal{F}) . Intuitivement, l'espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P}_A)$ correspond à une expérience aléatoire où l'on sait a priori que l'événement A est vérifié.

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

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(A|B)\mathbb{P}(B)}{\mathbb{P}(A)} \quad \text{formule de Bayes.}$$

5.2 Quelques formules

Une partition $(A_i)_{i \in I}$ avec $I \subset \mathbb{N}$ formée d'événements A_i est une famille d'événements vérifiant

- $\forall i, j \in I, i \neq j \implies A_i \cap A_j = \emptyset$,
- $\bigcup_{i \in I} A_i = \Omega$.

Dans ce cas, on a la **formule des probabilités totales** :

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

pour tout événement $B \in \mathcal{F}$ et $\mathbb{P}(A_i) > 0$ pour tout $i \in I$.

Cas particulier : avec la partition (A, \bar{A}) , on a

$$\mathbb{P}(B) = \mathbb{P}(A) \mathbb{P}(B|A) + \mathbb{P}(A^C) \mathbb{P}(B|A^C).$$

où $A^C = \{\omega \in \Omega \mid \omega \notin A\}$ est le complémentaire de A .

On peut alors étendre la formule de Bayes à une partition $(A_i)_{i \in I}$ de Ω :

$$\mathbb{P}(A_i|B) = \frac{\mathbb{P}(B|A_i) \mathbb{P}(A_i)}{\sum_{j \in I} \mathbb{P}(B|A_j) \mathbb{P}(A_j)}.$$

Cas particulier : avec la partition (A, \bar{A}) , on a

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

Exemple. On se place dans la cas d'une maladie qui touche une personne sur 100. Si une personne est malade (noté M), alors le test est positif dans 99% des cas. Si une personne n'est pas malade (noté M^C), alors le test est positif dans 1% des cas.

Notons P l'événement "le test est positif". Alors, si un test est positif, la probabilité que la

personne soit malade est donnée par

$$\begin{aligned}\mathbb{P}(M|P) &= \frac{\mathbb{P}(P|M)\mathbb{P}(M)}{\mathbb{P}(P|M)\mathbb{P}(M) + \mathbb{P}(P|M^C)\mathbb{P}(M^C)} \\ &= \frac{0.99 \times 0.01}{0.99 \times 0.01 + 0.01 \times 0.99} \\ &= \frac{1}{2}.\end{aligned}$$

Définition 5.3. Deux événements A et B sont **indépendants** si

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

En particulier,

$$\mathbb{P}(B|A) = \mathbb{P}(B) \quad \text{et} \quad \mathbb{P}(A|B) = \mathbb{P}(A).$$

Exemple. On lance 2 dés. On a alors

$$\Omega = \{1, 2, \dots, 6\}^2, \quad \mathcal{F} = \mathcal{P}(\Omega), \quad \mathbb{P} = \text{probabilité uniforme}.$$

On considère les événements

$$A = \{6\} \times \{1, \dots, 6\}, \quad B = \{1, \dots, 6\} \times \{6\}.$$

Alors,

$$\mathbb{P}(A) = \mathbb{P}(B) = \frac{1}{6},$$

et

$$\mathbb{P}(A \cap B) = \mathbb{P}(\{6\} \times \{6\}) = \frac{1}{36} = \mathbb{P}(A)\mathbb{P}(B).$$

Donc A et B sont indépendants

Définition 5.4. Des événements $A_1, \dots, A_n \in \mathcal{F}$ sont **indépendants** si

$$\forall I \subset \{1, \dots, n\}, I \neq \emptyset, \quad \mathbb{P}\left(\bigcap_{i \in I} A_i\right) = \prod_{i \in I} \mathbb{P}(A_i).$$

Plus généralement, une famille d'événements $(A_i)_{i \in I}$ est indépendante si toute sous-famille finie est constituée d'événements indépendants.

Exemple. On lance 2 dés. On considère les événements

A = le premier dé est pair,
 B = le deuxième dé est pair,
 C = la somme des dés est paire.

Alors,

$$\begin{aligned}\mathbb{P}(A) &= \frac{1}{2}, \\ \mathbb{P}(B) &= \frac{1}{2}, \\ \mathbb{P}(A \cap B) &= \frac{1}{4}, \\ \mathbb{P}(C) &= \frac{1}{2}, \\ \mathbb{P}(A \cap C) &= \frac{1}{4}, \\ \mathbb{P}(B \cap C) &= \frac{1}{4},\end{aligned}$$

Donc les événements A et B , A et C , B et C sont indépendants respectivement. Par ailleurs, on a

$$\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A \cap B) = \frac{1}{4} \neq \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C) = \frac{1}{8}.$$

Donc les événements A , B et C ne sont pas indépendants.

Remarque. L'indépendance d'événements A_1, \dots, A_n implique l'indépendance de A_1^c, A_2, \dots, A_n (et de même en passant au complémentaire sur d'autres indices). En particulier, si A et B sont indépendants, alors A et B^c sont indépendants.

6 Indépendance de variables aléatoires

Définition 6.1. Soient X_1, \dots, X_n des variables aléatoires à valeurs dans $(E_1, \mathcal{E}_1), \dots, (E_n, \mathcal{E}_n)$. On dit que ces variables aléatoires sont **indépendantes** si pour tout $B_1 \in \mathcal{E}_1, \dots, B_n \in \mathcal{E}_n$, les événements $\{X_1 \in B_1\}, \dots, \{X_n \in B_n\}$ sont indépendants :

$$\mathbb{P}(X_1 \in B_1, \dots, X_n \in B_n) = \prod_{i=1}^n \mathbb{P}(X_i \in B_i).$$

Cette égalité se réécrit

$$\begin{aligned}\mathbb{P}((X_1, \dots, X_n) \in B_1 \times \dots \times B_n) &= \prod_{i=1}^n \mathbb{P}(X_i \in B_i) \\ \mathbb{P}_{(X_1, \dots, X_n)}(B_1 \times \dots \times B_n) &= \prod_{i=1}^n \mathbb{P}_{X_i}(B_i).\end{aligned}$$

Cela signifie que

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

Plus généralement, une famille de variables aléatoires $(X_i)_{i \in I}$ est indépendante si toute sous-famille finie l'est.

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

1. Les variables aléatoires X_1, \dots, X_n sont indépendantes.
2. Pour toutes fonctions $h_1 : E_1 \rightarrow \mathbb{R}, \dots, h_n : E_n \rightarrow \mathbb{R}$ positives ou intégrables pour $(\mathcal{L}^1(E_k, \mathcal{E}_k, \mathbb{P}_{X_k}))$, on a

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

Démonstration. \triangleright On montre que 1. implique 2.

On considère le vecteur aléatoire

$$X = (X_1, \dots, X_n) : \Omega \rightarrow E_1 \times \dots \times E_n, \\ \omega \mapsto (X_1(\omega), \dots, X_n(\omega)).$$

Alors

$$\begin{aligned} \mathbb{E}[h_1(X_1) \cdots h_n(X_n)] &= \int_{\Omega} h_1(X_1(\omega)) \cdots h_n(X_n(\omega)) d\mathbb{P}(\omega) \\ &\stackrel{\text{formule de transfert}}{=} \int_{E_1 \times \dots \times E_n} h_1(x_1) \cdots h_n(x_n) d\mathbb{P}_{(X_1, \dots, X_n)}(x_1, \dots, x_n) \\ &= \int_{E_1 \times \dots \times E_n} h_1(x_1) \cdots h_n(x_n) d(\mathbb{P}_{X_1} \otimes \dots \otimes \mathbb{P}_{X_n})(x_1, \dots, x_n) \\ &\stackrel{\text{Fubini Tonelli}}{=} \int_{E_n} \int_{E_{n-1}} \cdots \int_{E_1} h_1(x_1) \cdots h_n(x_n) d\mathbb{P}_{X_1}(x_1) \cdots d\mathbb{P}_{X_n}(x_n) \\ &= \int_{E_1} h_1(x_1) d\mathbb{P}_{X_1}(x_1) \cdots \int_{E_n} h_n(x_n) d\mathbb{P}_{X_n}(x_n) \\ &= \mathbb{E}[h_1(X_1)] \cdots \mathbb{E}[h_n(X_n)]. \end{aligned}$$

▷ On montre que 2. implique 1.

Soit $A_1 \in \mathcal{C}_1, \dots, A_n \in \mathcal{C}_n$. On pose $h_i = \mathbb{1}_{A_i}(x_i)$. Alors

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

se réécrit

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

et donc

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

□

Exemple. Si X_1 et X_2 sont des variables aléatoires intégrables et indépendantes, alors

$$\mathbb{E}[X_1 X_2] = \mathbb{E}[X_1] \mathbb{E}[X_2].$$

En particulier,

$$\text{Cov}(X_1, X_2) = \mathbb{E}[X_1 X_2] - \mathbb{E}[X_1] \mathbb{E}[X_2] = 0.$$

et donc

$$\text{Var}(X_1 + X_2) = \text{Var}(X_1) + \text{Var}(X_2) + 2 \text{Cov}(X_1, X_2) = \text{Var}(X_1) + \text{Var}(X_2).$$

Exemple. Soit $t \in \mathbb{R}$ fixé et X_1, X_2 deux variables aléatoires réelles indépendantes. Avec $h_j(x_j) = e^{itx_j}, j = 1, 2$, on obtient

$$\mathbb{E}[e^{itX_1} e^{itX_2}] = \mathbb{E}[e^{itX_1}] \mathbb{E}[e^{itX_2}].$$

c'est-à-dire

$$\varphi_{X_1+X_2}(t) = \varphi_{X_1}(t) \varphi_{X_2}(t).$$

6.2 Critères d'indépendance

Cas discret

Soit X et Y deux variables aléatoires à valeurs dans des espaces discrets E et F . Elles sont indépendantes si et seulement si

$$\mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x) \mathbb{P}(Y = y)$$

pour tout $x \in E$ et $y \in F$.

En effet, si $A \subset E$ et $B \subset F$, alors

$$\begin{aligned} \mathbb{P}(X \in A, Y \in B) &= \mathbb{P}(X \in \cup_{x \in A} \{x\}, Y \in \cup_{y \in B} \{y\}) \\ &= \sum_{\substack{x \in A \\ y \in B}} \mathbb{P}(X = x, Y = y) \end{aligned}$$

Avec les fonctions de répartition

Soient X et Y deux variables aléatoires à valeurs réelles. Elles sont indépendantes si et seulement si pour tout $x, y \in \mathbb{R}$,

$$\mathbb{P}(X \leq x, Y \leq y) = \mathbb{P}(X \leq x)\mathbb{P}(Y \leq y).$$

c'est-à-dire

$$F_{X,Y}(x, y) = F_X(x)F_Y(y).$$

Avec les fonctions caractéristiques

Soient X_1, X_2 deux variables aléatoires à valeurs réelles. Elles sont indépendantes si et seulement si

$$\begin{aligned}\mathbb{E}[e^{it_1 X_1} e^{it_2 X_2}] &= \mathbb{E}[e^{it_1 X_1}] \mathbb{E}[e^{it_2 X_2}] \\ \mathbb{E}[e^{i(t_1 X_1 + t_2 X_2)}] &= \mathbb{E}[e^{it_1 X_1}] \mathbb{E}[e^{it_2 X_2}].\end{aligned}$$

c'est-à-dire

$$\varphi_{(X_1+X_2)}(t_1 + t_2) = \varphi_{X_1}(t_1)\varphi_{X_2}(t_2).$$

Cas des variables aléatoires à densité

Soit (X_1, X_2) un vecteur aléatoire de densité $f_{(X_1, X_2)}$ par rapport à λ_2 (mesure de Lebesgue sur \mathbb{R}^2). On suppose que

$$f_{(X_1, X_2)}(x_1, x_2) = f_1(x_1)f_2(x_2)$$

avec $f_j : \mathbb{R} \rightarrow \mathbb{R}_+$ positive et vérifiant

$$\int_{\mathbb{R}} f_j(x) d\lambda_1(x) = 1.$$

Alors les variables aléatoires X_1 et X_2 sont indépendantes de densité respectives f_1 et f_2 .

Exemple. Soit (X, Y) un vecteur aléatoire de densité

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

Alors,

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

donc les variables aléatoires X et Y sont indépendantes et suivent toutes les deux une loi normale centrée réduite.

7 Résultats asymptotiques

7.1 Lemme de Borel-Cantelli

Si $(A_n)_{n \in \mathbb{N}}$ est une suite d'événements et $A_n \in \mathcal{F}$, $\forall n \in \mathbb{N}$, alors on pose

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

et

$$\liminf_{n \rightarrow \infty} A_n = \bigcup_{n=0}^{\infty} \bigcap_{k=n}^{\infty} A_k = \{\omega \in \Omega \mid \omega \in A_k, \text{ pour tout } k \text{ à partir d'un certain rang}\}.$$

On remarque que

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

et

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

Proposition. Soit $(A_n)_{n \in \mathbb{N}}$ une suite d'événements, alors

1. Si $\sum_{n=0}^{\infty} \mathbb{P}(A_n) < \infty$, alors

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

ou de manière équivalente

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

2. Si $\sum_{n=0}^{\infty} \mathbb{P}(A_n) = \infty$ et les événements A_n sont indépendants, alors

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

ou de manière équivalente

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

Remarque. L'indépendance est nécessaire. Prendre $A_n = A$ avec $\mathbb{P}(A) \in]0, 1[$. Alors $\sum_{n=0}^{\infty} \mathbb{P}(A) = \infty$ et $\mathbb{P}(\limsup_{n \rightarrow \infty} A_n) = \mathbb{P}(A) \neq 1$.

Démonstration. \triangleright On montre 1.

On pose $B_n = \bigcup_{k=n}^{\infty} A_k$. Comme les $(B_n)_{n \in \mathbb{N}}$ sont décroissants, on a

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

et

$$\begin{aligned} \mathbb{P}(B_n) &= \mathbb{P}\left(\bigcup_{k=n}^{\infty} A_k\right) \\ &\leq \sum_{k=n}^{\infty} \mathbb{P}(A_k) \\ &\xrightarrow{n \rightarrow \infty} 0. \end{aligned}$$

donc

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

\triangleright On montre 2.

Pour $n \in \mathbb{N}$, on a

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

Par théorème de convergence dominée (par 1), on a

$$\begin{aligned}
\mathbb{P}\left(\bigcap_{k=n}^{\infty} A_k^c\right) &= \lim_{N \rightarrow \infty} \mathbb{E} \left[\prod_{k=n}^N \mathbb{1}_{A_k^c} \right] \\
&= \lim_{N \rightarrow \infty} \mathbb{P} \left(\bigcap_{k=n}^N A_k^c \right) \\
\text{par indep des } A_k &= \lim_{N \rightarrow \infty} \prod_{k=n}^N \mathbb{P}(A_k^c) \\
&= \lim_{N \rightarrow \infty} \prod_{k=n}^N (1 - \mathbb{P}(A_k)) \\
&\leq \lim_{N \rightarrow \infty} \prod_{k=n}^N e^{-\mathbb{P}(A_k)} \\
&= \lim_{N \rightarrow \infty} e^{-\sum_{k=n}^N \mathbb{P}(A_k)} \\
&= e^{-\sum_{k=n}^{\infty} \mathbb{P}(A_k)} \\
&= 0.
\end{aligned}$$

On conclut en écrivant

$$\mathbb{P}(\liminf_{n \rightarrow \infty} A_k^c) = \mathbb{P}\left(\bigcup_{n=0}^{\infty} \bigcap_{k=n}^{\infty} A_k^c\right) = 0$$

□

Exemple. Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires avec $X_n \sim \mathcal{B}(p_n)$.

On considère les événements $A_n = \{X_n = 1\}$.

de sorte que $p_n = \mathbb{P}(X_n = 1) = \mathbb{P}(A_n)$.

Borel-Cantelli donne alors

1. Si $\sum_{n=0}^{\infty} p_n < \infty$ alors la suite (X_n) prend la valeur 1 un nombre fini de fois, \mathbb{P} -presque sûrement.
2. Si $\sum_{n=0}^{\infty} p_n = \infty$ et que les (X_n) sont indépendantes, alors la suite (X_n) prend une infinité de fois la valeur 1, \mathbb{P} -presque sûrement.

Exemple. On observe les motifs dans une suite indépendante et identiquement distribuée.

Soit $(X_k)_{k \geq 1}$ une suite de variables aléatoires indépendantes et de même loi $\mathcal{B}(\frac{1}{2})$.

Soit $n \geq 1$ fixé. Soit $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)$ une suite déterministe fixée dans $\{0, 1\}^n$.

Pour $k \in \mathbb{N}$, on définit

$$B_k(\varepsilon) = \left\{ (x_i)_{i \geq 1} \in \{0, 1\}^{\mathbb{N}^*} \mid x_{k+1} = \varepsilon_1, \dots, x_{k+n} = \varepsilon_n \right\}.$$

On considère alors l'événement

$$A_k(\varepsilon) = \left\{ (X_i)_{i \geq 1} \in B_k(\varepsilon) \right\} = \left\{ X_{k+i-1} = \varepsilon_1, \dots, X_{k+n} = \varepsilon_n \right\}.$$

l'événement $A_k(\varepsilon)$ est réalisé si X_{k+1} prend la valeur ε_1 , X_{k+2} prend la valeur ε_2 et ainsi de suite avec X_{k+n} qui prend la valeur ε_n . On pose

$$A(\varepsilon) = \limsup_{k \rightarrow \infty} A_k(\varepsilon)$$

qui correspond à l'événement "le motif $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)$ apparaît une infinité de fois dans la suite $(X_k)_{k \geq 1}$ ".

Les événements $(A_k(\varepsilon))_{k \geq 0}$ ne sont pas indépendants :

$$\begin{aligned}
A_0(\varepsilon) &= \{X_1 = \varepsilon_1, \dots, X_n = \varepsilon_n\}, \\
A_1(\varepsilon) &= \{X_2 = \varepsilon_1, \dots, X_{n+1} = \varepsilon_n\}. A_2(\varepsilon) = \{X_3 = \varepsilon_1, \dots, X_{n+2} = \varepsilon_n\}.
\end{aligned}$$

les événements $(A_{k \times n}(\varepsilon))_{k \in \mathbb{N}}$ sont indépendants

$$\begin{aligned} A_{0 \times n}(\varepsilon) &= \{X_1 = \varepsilon_1, \dots, X_n = \varepsilon_n\}, \\ A_{1 \times n}(\varepsilon) &= \{X_{n+1} = \varepsilon_1, \dots, X_{2n} = \varepsilon_n\} \cdot A_{2 \times n}(\varepsilon) = \{X_{2n+1} = \varepsilon_1, \dots, X_{3n} = \varepsilon_n\}. \end{aligned}$$

Par ailleurs, on a

$$\begin{aligned} \mathbb{P}(A_{k \times n}(\varepsilon)) &= \mathbb{P}(X_{k \times n+1} = \varepsilon_1, \dots, X_{(k+1) \times n} = \varepsilon_n) \\ &= \mathbb{P}(X_{k \times n+1} = \varepsilon_1) \cdots \mathbb{P}(X_{(k+1) \times n} = \varepsilon_n) \\ &= \frac{1}{2^n}. \end{aligned}$$

qui est le terme général d'une série DIVERGENTE (car on somme sur k).

Par Borel-Cantelli, on a

$$\mathbb{P}(\limsup_{k \rightarrow \infty} A_{k \times n}(\varepsilon)) = 1.$$

Il en suit

$$\begin{aligned} \mathbb{P}(A(\varepsilon)) &= \mathbb{P}(\limsup_{k \rightarrow \infty} A_k(\varepsilon)) \\ &= \mathbb{P}(\limsup_{k \rightarrow \infty} A_{k \times n}(\varepsilon)) \\ &= 1. \end{aligned}$$

Donc si on tape au hasard sur un clavier binaire, on fera apparaitre n'importe quel texte une infinité de fois avec probabilité 1.

7.2 Loi du 0 – 1 de Kolmogorov

Soit $\mathcal{F}_1, \mathcal{F}_2, \dots$ des sous-tribus de \mathcal{F} . On pose

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

la tribu engendrée par les $\mathcal{F}_k, k \geq n$.

On définit alors la tribu asymptotique (ou de queue) par

$$\mathcal{F}_\infty = \bigcap_{n \geq 1} \overline{\mathcal{F}_n}.$$

Exemple. Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires et $\mathcal{F}_n = \sigma(X_n) = \{X_n^{-1}(B), B \in \mathcal{B}\}$. Montrons que l'événement " X_n converge" est dans \mathcal{F}_∞ . L'événement " X_n converge" est réalisé si et seulement si l'événement " $(X_n)_{n \geq k}$ converge" est réalisé pour tout $k \geq 1$. Bref,

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

car $(X_n)_{n \geq k}$ ne dépend que de X_k, X_{k+1}, \dots

Ceci est vrai pour tout k donc

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

Proposition. Si les tribus $\mathcal{F}_1, \mathcal{F}_2, \dots$ sont indépendantes (un événement d'une tribu est indépendant de tout événement d'une autre tribu), alors la tribu asymptotique \mathcal{F}_∞ ne contient que des événements de probabilité 0 ou 1 :

$$\forall A \in \mathcal{F}_\infty, \mathbb{P}(A) \in \{0, 1\}.$$

Exemple. Dans l'exemple précédent, on a donc

$$\mathbb{P}((X_n)_{n \geq 1} \text{ converge}) \in \{0, 1\}.$$

■ Donc soit la suite $(X_n)_{n \geq 1}$ converge presque sûrement, soit elle diverge presque sûrement.

Chapitre 3 — Loi des grands nombres

Dans tout ce chapitre, on considère $(X_n)_{n \geq 1}$ une suite de variables aléatoires réelles (ou complexes) définies sur un espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$ et une variable aléatoire X réelle (ou complexe).

8 Différents modes de convergence

8.1 Convergences presque sûre et probabilité

Définition 8.2. On dit que X_n converge **presque sûrement** vers X et on note $X_n \xrightarrow[n \rightarrow \infty]{\text{p.s.}} X$, si

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

Remarque.

1. Cela correspond à la convergence presque partout pour la mesure de probabilité \mathbb{P} .
2. Si $X_n \xrightarrow[n \rightarrow \infty]{\text{p.s.}} X$ et si $h : \mathbb{R} \rightarrow \mathbb{R}$ est une fonction continue, alors $h(X_n) \xrightarrow[n \rightarrow \infty]{\text{p.s.}} h(X)$. En particulier, si $X_n \xrightarrow[n \rightarrow \infty]{\text{p.s.}} X$ et $Y_n \xrightarrow[n \rightarrow \infty]{\text{p.s.}} Y$, alors

$$aX_n + bY_n \xrightarrow[n \rightarrow \infty]{\text{p.s.}} aX + bY, \quad a, b \in \mathbb{R}.$$

et

$$X_n Y_n \xrightarrow[n \rightarrow \infty]{\text{p.s.}} XY.$$

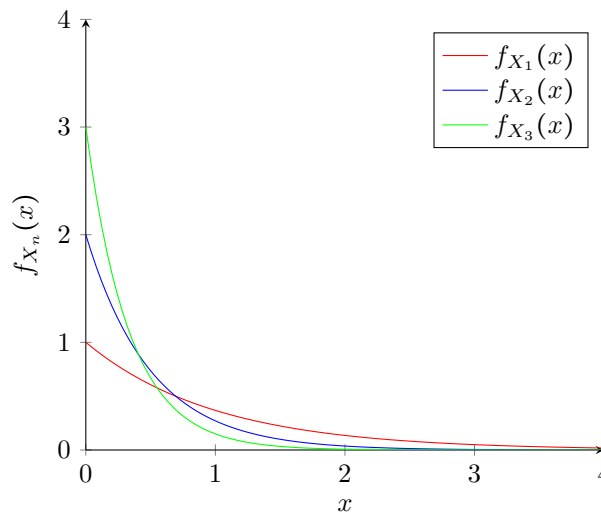
Définition 8.3. On dit que $(X_n)_{n \geq 1}$ converge **en probabilité** vers X si

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

On note alors $X_n \xrightarrow[n \rightarrow \infty]{\mathbb{P}} X$.

Cela signifie que la probabilité que X_n s'écarte de X d'un écart supérieur à ε tend vers 0 lorsque n tend vers l'infini.

Exemple. \triangleright On considère $X_n \sim \mathcal{E}(n)$



Regardons $\mathbb{P}(|X_n| \geq \varepsilon)$:

$$\begin{aligned}\mathbb{P}(|X_n| \geq \varepsilon) &= \mathbb{P}(X_n \geq \varepsilon) \\ &= \int_{\varepsilon}^{+\infty} ne^{-nx} dx \\ &= [-e^{-nx}]_{\varepsilon}^{+\infty} \\ &= e^{-n\varepsilon}.\end{aligned}$$

Donc $X_n \xrightarrow[n \rightarrow \infty]{\mathbb{P}} 0$.

On peut montrer que $X_n \xrightarrow[n \rightarrow \infty]{\text{p.s.}} 0$ (en TD).

Proposition. Si $X_n \xrightarrow[n \rightarrow \infty]{\text{p.s.}} X$, alors $X_n \xrightarrow{\mathbb{P}} X$.

Démonstration. Soit $\varepsilon > 0$. On a

$$\mathbb{P}(|X_n - X| \geq \varepsilon) = \mathbb{E}[\underbrace{\mathbb{1}_{|X_n - X| \geq \varepsilon}}_{f_n \xrightarrow[n \rightarrow \infty]{0\text{-p.p.}}}]$$

Comme $|f_n| \leq 1$ qui est intégrable, le théorème de convergence dominée donne

$$\mathbb{P}(|X_n - X| \geq \varepsilon) = \mathbb{E}[\lim_{n \rightarrow \infty} \mathbb{1}_{|X_n - X| \geq \varepsilon}] = 0.$$

□

Exemple. On considère $(X_n)_{n \geq 1}$ une suite de variables aléatoires de loi

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

Soit $\varepsilon > 0$. Alors

$$\mathbb{P}(|X_n| \geq \varepsilon) \leq \frac{1}{n} \xrightarrow[n \rightarrow \infty]{} 0.$$

donc $X_n \xrightarrow[n \rightarrow \infty]{\mathbb{P}} 0$.

Supposons que les $(X_n)_{n \geq 1}$ sont indépendantes. Les événements $A_n = \{X_n = n\}$ sont indépendants et vérifient

$$\begin{aligned}\sum_{n=1}^{+\infty} \mathbb{P}(A_n) &= \sum_{n=1}^{+\infty} \mathbb{P}(X_n = n) \\ &= \sum_{n=1}^{+\infty} \frac{1}{n} \\ &= +\infty.\end{aligned}$$

Donc par Borel-Cantelli, on a

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

Donc avec probabilité 1, X_n prend la valeur n une infinité de fois. Donc avec probabilité 1, $(X_n)_{n \geq 1}$ ne converge pas vers 0. Donc $(X_n)_{n \geq 1}$ ne converge pas presque sûrement.

Proposition. Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires réelles dans $\mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$. On suppose que les X_n sont indépendantes et identiquement distribuées (c'est-à-dire de même loi, i.i.d. en abrégé).

On note μ leur espérance : $\mu = \mathbb{E}[X_1]$. On pose $S_n = X_1 + \dots + X_n$. Alors

$$\overline{X_n} = \frac{S_n}{n} = \frac{X_1 + \dots + X_n}{n} \xrightarrow[n \rightarrow \infty]{\mathbb{P}} \mu.$$

On appelle $\overline{X_n} = \frac{S_n}{n}$ la **moyenne empirique**

Démonstration.

▷ On calcule d'abord l'espérance de $\overline{X_n}$:

$$\mathbb{E}[\overline{X_n}] = \frac{\mathbb{E}[X_1] + \dots + \mathbb{E}[X_n]}{n} = \mu.$$

▷ On calcule ensuite la variance de $\overline{X_n}$:

$$\begin{aligned} \text{Var}(\overline{X_n}) &= \text{Var}\left(\frac{X_1 + \dots + X_n}{n}\right) \\ &= \frac{1}{n^2} \text{Var}(X_1 + \dots + X_n) \\ &= \frac{\text{Var}(X_1)}{n} \end{aligned}$$

▷ On applique ensuite l'inégalité de Bienaymé-Tchebychev :

Soit $\varepsilon > 0$. On a

$$\begin{aligned} \mathbb{P}(|\overline{X_n} - \mu| \geq \varepsilon) &= \mathbb{P}(|\overline{X_n} - \mathbb{E}[\overline{X_n}]| \geq \varepsilon) \\ &\stackrel{\text{B-T}}{\leq} \frac{\text{Var}(\overline{X_n})}{\varepsilon^2} \\ &= \frac{\text{Var}(X_1)}{n\varepsilon^2} \xrightarrow{n \rightarrow \infty} 0. \end{aligned}$$

Donc $\overline{X_n} \xrightarrow[n \rightarrow \infty]{\mathbb{P}} \mu$. □

Exemple. Si les X_n sont i.i.d. de loi $\mathcal{B}(p)$, alors

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

8.4 Convergence dans $\mathcal{L}^p = \mathcal{L}^p(\Omega, \mathcal{F}, \mathbb{P})$

Définition 8.5. Soit $p \in [1, +\infty]$. On dit que $(X_n)_{n \geq 1}$ converge vers X dans \mathcal{L}^p , notée $X_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}^p} X$, si

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

Remarque.

1. On rappelle que \mathcal{L}^p est un espace vectoriel muni de la norme

$$\|\cdot\|_p = (\mathbb{E}[|X|^p])^{\frac{1}{p}} = \left(\int_{\Omega} |X(\omega)|^p d\mathbb{P}(\omega) \right)^{\frac{1}{p}}.$$

2. Les espaces \mathcal{L}^p sont décroissants pour l'inclusion : si $p \leq q$, alors $\mathcal{L}^q \subset \mathcal{L}^p$ avec l'inégalité $\|\cdot\|_p \leq \|\cdot\|_q$. Donc la convergence dans \mathcal{L}^q implique la convergence dans \mathcal{L}^p pour $p \leq q$.

Proposition. Si $X_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}^p} X$, alors $X_n \xrightarrow[n \rightarrow \infty]{\mathbb{P}} X$.

Démonstration. Soit $\varepsilon > 0$. On a

$$\mathbb{P}(|X_n - X| \geq \varepsilon) \leq \frac{\mathbb{E}[|X_n - X|^p]}{\varepsilon^p} \xrightarrow{n \rightarrow \infty} 0.$$

Donc $X_n \xrightarrow[n \rightarrow \infty]{\mathbb{P}} X$. □

Exemple. On se fixe $p \in [1, +\infty[$ et on modifie l'exemple précédent en définissant la suite $(X_n)_{n \geq 1}$ via

$$\mathbb{P}(X_n = n^{\frac{1}{p}}) = \frac{1}{n}, \quad \text{et} \quad \mathbb{P}(X_n = 0) = 1 - \frac{1}{n}.$$

Soit $\varepsilon > 0$. Alors

$$\mathbb{P}(|X_n| \geq \varepsilon) \leq \frac{1}{n} \longrightarrow 0.$$

Donc $X_n \xrightarrow[n \rightarrow \infty]{\mathbb{P}} 0$.

Par ailleurs

$$\begin{aligned} \mathbb{E}[|X_n|^p] &= \left(n^{\frac{1}{p}}\right)^p \times \frac{1}{n} + 0^p \times \left(1 - \frac{1}{n}\right) \\ &= 1 \not\xrightarrow[n \rightarrow \infty]{} 0 \end{aligned}$$

Donc $X_n \not\xrightarrow[n \rightarrow \infty]{\mathcal{L}^p} 0$.

Cependant, si $q < p$, alors

$$\begin{aligned} \mathbb{E}[|X_n|^q] &= n^{\frac{q}{p}} \times \frac{1}{n} + 0^q \times \left(1 - \frac{1}{n}\right) \\ &= n^{\frac{q}{p}-1} \xrightarrow[n \rightarrow \infty]{} 0. \end{aligned}$$

Donc $X_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}^q} 0$ pour $q < p$.

Remarque.

1. Si $X_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}^p} X$, alors il existe une suite extraite $(n_k)_{k \geq 1}$ telle que $X_{n_k} \xrightarrow[n \rightarrow \infty]{\text{p.s.}} X$.
2. Si $X_n \xrightarrow[n \rightarrow \infty]{\text{p.s.}} X$ et si $|X_n| \leq Y \in \mathcal{L}^p$, alors $X_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}^p} X$ (par convergence dominée).
3. Si $X_n \xrightarrow[n \rightarrow \infty]{\mathbb{P}} X$ alors il existe une suite extraite $(n_k)_{k \geq 1}$ telle que $X_{n_k} \xrightarrow[n \rightarrow \infty]{\text{p.s.}} X$.
4. Si $X_n \xrightarrow[n \rightarrow \infty]{\mathbb{P}} X$ et s'il existe $M > 0$ tel que $\mathbb{E}[|X_n|^p] \leq M$, alors $X_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}^q} X$ pour $q \in [1, p[$.

