

Université Pierre-et-Marie-Curie — 4 place Jussieu — 75252 Paris Cedex 05

PROBABILITÉS DE BASE

Année 2009 – 2010

Zhan SHI

<http://www.proba.jussieu.fr/pageperso/zhan/proba.html>

Table des matières

Chapitre 1. Espaces de probabilité	1
Chapitre 2. Variables aléatoires discrètes	9
Chapitre 3. Variables aléatoires	15
Chapitre 4. Vecteurs aléatoires	27
Chapitre 5. Indépendance	39
Chapitre 6. Fonctions caractéristiques	49
Chapitre 7. Suites et séries de variables aléatoires	61
Chapitre 8. Loi des grands nombres	73
Chapitre 9. Convergence en loi	81
Chapitre 10. Théorème central limite	91
Chapitre 11. Espérances conditionnelles	101
Chapitre 12. Vecteurs aléatoires gaussiennes	113
Chapitre 13. Exemples de processus aléatoires	123

Chapitre 1. Espaces de probabilité

1. Axiomatique de Kolmogorov

Définition 1.1. Un espace mesurable (Ω, \mathcal{F}) est un ensemble Ω non-vide muni d'une tribu \mathcal{F} qui vérifie les propriétés suivantes :

- (a) $\Omega \in \mathcal{F}$.
- (b) Si $A \in \mathcal{F}$, alors $A^c \in \mathcal{F}$, où A^c désigne le complémentaire de A relatif à Ω : $A^c = \Omega \setminus A$.
- (c) Si $(A_n)_{n \geq 1}$ est une suite d'éléments dans \mathcal{F} , alors $\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$.

Remarque. Par passage au complémentaire, on a aussi sous l'hypothèse de (c) que $\bigcap_{n=1}^{\infty} A_n \in \mathcal{F}$. □

Définition 1.2. On appelle **mesure de probabilité** (ou simplement : probabilité) sur (Ω, \mathcal{F}) une application \mathbb{P} définie sur \mathcal{F} qui vérifie les axiomes suivants :

- i. Si $A \in \mathcal{F}$, alors $0 \leq \mathbb{P}(A) \leq 1$.
- ii. $\mathbb{P}(\Omega) = 1$.
- iii (σ -additivité). Si $(A_n)_{n \geq 1}$ est une suite finie ou infinie dénombrable d'éléments de \mathcal{F} , deux à deux disjoints (c'est-à-dire tels que $A_i \cap A_j = \emptyset$ si $i \neq j$), alors $\mathbb{P}(\bigcup_n A_n) = \sum_n \mathbb{P}(A_n)$.

On appellera $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité (parfois : espace probabilisé).

Langage probabiliste : $A \in \mathcal{F}$ est un événement.

2. Propriétés élémentaires

Propriété 2.1. Pour tout $A \in \mathcal{F}$, on a $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$.

Preuve. Par iii et ii, $\mathbb{P}(A) + \mathbb{P}(A^c) = \mathbb{P}(A \cup A^c) = \mathbb{P}(\Omega) = 1$. D'où le résultat. \square

Propriété 2.2. $\mathbb{P}(\emptyset) = 0$.

Remarque. $A \in \mathcal{F}$, $\mathbb{P}(A) = 0$ n'implique pas forcément $A = \emptyset$. De même, $A \in \mathcal{F}$, $\mathbb{P}(A) = 1$ n'implique pas $A = \Omega$. \square

Propriété 2.3. Si $A \in \mathcal{F}$ et $B \in \mathcal{F}$, avec $B \subset A$, alors $\mathbb{P}(A \setminus B) = \mathbb{P}(A \cap B^c) = \mathbb{P}(A) - \mathbb{P}(B)$.

Propriété 2.4. Si $A \in \mathcal{F}$ et $B \in \mathcal{F}$, alors $\mathbb{P}(A \setminus B) = \mathbb{P}(A) - \mathbb{P}(A \cap B)$.

Propriété 2.5. Si $A \in \mathcal{F}$ et $B \in \mathcal{F}$, alors $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$.

Propriété 2.6. Si $(A_n)_{n \geq 1}$ est une suite croissante d'éléments dans \mathcal{F} , alors $\mathbb{P}(\cup_{n=1}^{\infty} A_n) = \lim_{n \rightarrow \infty} \mathbb{P}(A_n)$.

Preuve. Soient $B_1 = A_1$, $B_2 = A_2 \cap A_1^c$, $B_3 = A_3 \cap A_2^c$, \dots , $B_n = A_n \cap A_{n-1}^c$. Les événements $(B_n)_{n \geq 1}$ étant deux à deux disjoints, avec $\cup_{n=1}^{\infty} B_n = \cup_{n=1}^{\infty} A_n$, on a, d'après la propriété de σ -additivité que $\mathbb{P}(\cup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \mathbb{P}(B_n) = \lim_{k \rightarrow \infty} \sum_{n=1}^k \mathbb{P}(B_n) = \lim_{k \rightarrow \infty} \mathbb{P}(\cup_{n=1}^k B_n) = \lim_{k \rightarrow \infty} \mathbb{P}(A_k)$. \square

Propriété 2.7. Si $(A_n)_{n \geq 1}$ est une suite décroissante d'éléments dans \mathcal{F} , alors $\mathbb{P}(\cap_{n=1}^{\infty} A_n) = \lim_{n \rightarrow \infty} \mathbb{P}(A_n)$.

Preuve. Ceci est une conséquence directe de la propriété précédente par passage au complémentaire. \square

Propriété 2.8. Soit $\{A_n\}_{n \geq 1}$ une partition finie ou infinie dénombrable de (Ω, \mathcal{F}) (c'est-à-dire que les A_n sont deux à deux disjoints, et que leur union est Ω). Alors pour tout $B \in \mathcal{F}$,

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

Preuve. Soient $B_n = A_n \cap B$. Les (B_n) sont deux à deux disjoints, avec $\cup_{n=1}^{\infty} B_n = B$. Par σ -additivité,

$$\mathbb{P}(B) = \sum_n \mathbb{P}(B_n) = \sum_n \mathbb{P}(B \cap A_n). \quad \square$$

3. Exemple : équi-probabilité et dénombrement

On considère un espace fini $\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}$, muni de la tribu $\mathcal{F} = \mathcal{P}(\Omega) := \{A : A \subset \Omega\}$, et de la probabilité $\mathbb{P} : \mathcal{F} \rightarrow [0, 1]$ avec

$$\mathbb{P}(A) = \frac{\#(A)}{\#(\Omega)} = \frac{\#(A)}{n}, \quad \forall A \in \mathcal{F}.$$

Cette probabilité s'appelle "équi-probabilité" sur Ω , car il ne privilège aucun élément particulier de Ω .

Exemple 3.1. On jette trois fois une pièce de monnaie parfaite. On peut représenter l'espace Ω comme l'ensemble des applications de $\{1, 2, 3\}$ (trois jets) dans $\{P, F\}$ ($P =$ pile, $F =$ face). Autrement dit, $\Omega = \{P, F\}^3$, donc $n = \#(\Omega) = 2^3 = 8$.

Il est alors facile de voir par exemple que

$$\mathbb{P}(\text{on sort exactement une fois } P) = \frac{3}{8},$$

$$\mathbb{P}(\text{on sort au moins une fois } P) = 1 - \mathbb{P}(\text{on sort trois fois } F) = 1 - \frac{1}{8} = \frac{7}{8}. \quad \square$$

Exemple 3.2. On considère l'arrivée d'une course de chevaux, avec dix partants, numérotés de 1 à 10. On note l'ordre d'arrivée. On suppose que les concurrents sont de force égale et qu'il n'y a pas d'ex-aequo. L'espace Ω est l'ensemble des injections de $\{1, 2, \dots, 10\}$ dans lui-même. Donc $n = \#(\Omega) = 10!$. On a alors

$$\begin{aligned} \mathbb{P}(\text{le numéro 10 arrive dernier}) &= \frac{\#(\omega \in \Omega : \omega(10) = 10)}{10!} \\ &= \frac{\text{nombre d'injections de } \{1, 2, \dots, 9\} \text{ dans lui-même}}{10!} \\ &= \frac{9!}{10!} = \frac{1}{10}. \end{aligned}$$

Si l'on s'intéresse à l'événement $A = \{\text{le numéro 10 arrive dans les trois premiers}\}$, on peut considérer $A_k = \{\text{le numéro 10 arrive à la } k\text{-ième place}\}$ ($k = 1, 2$ ou 3). Les A_k sont deux à deux disjoints, on a

$$\mathbb{P}(A) = \mathbb{P}(A_1) + \mathbb{P}(A_2) + \mathbb{P}(A_3) = \frac{3}{10}. \quad \square$$

Exemple 3.3. Le jeu du loto, qui est un jeu de la Française des Jeux, consiste à choisir 6 numéros distincts parmi $\{1, 2, \dots, 49\}$. On suppose que les boules qui portent les 49 numéros sont toutes parfaites.

Ici on ne s'intéresse qu'aux résultats des 6 boules. L'espace Ω est

$$\Omega = \{ \{a_1, a_2, \dots, a_6\} : 1 \leq a_i \leq 49, \text{ les } a_i \text{ sont deux à deux différents} \}.$$

(Par exemple, $\{1, 2, \dots, 6\} = \{6, 5, \dots, 1\}$). On a $n = \#(\Omega) = \binom{49}{6}$. Par conséquent,

$$\mathbb{P}(\text{on gagne le premier prix avec un bulletin}) = \frac{1}{\binom{49}{6}} \approx 7 \times 10^{-8}. \quad \square$$

4. Événements indépendants

Définition 4.1. Soit A et B deux événements sur l'espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$. On dit qu'ils sont indépendants si $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.

Exemple 4.2. Deux événements A et B sont indépendants si et seulement si A^c et B sont indépendants.

En effet, $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B) \Leftrightarrow \mathbb{P}(B) - \mathbb{P}(A \cap B) = \mathbb{P}(B) - \mathbb{P}(A)\mathbb{P}(B) \Leftrightarrow \mathbb{P}(A^c \cap B) = \mathbb{P}(A^c)\mathbb{P}(B)$.

On peut appliquer le même argument pour voir que A et B sont indépendants si et seulement si A^c et B^c sont indépendants. \square

Warning. Il ne faut pas confondre deux événements indépendants [$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$] avec deux événements disjoints [$A \cap B = \emptyset$]. Ce sont des notions totalement différentes. \square

Définition 4.3. Les événements A_1, \dots, A_n sont indépendants si, pour tout $k \leq n$ et tout k -uplet $1 \leq i_1 < i_2 < \dots < i_k \leq n$,

$$(2.1) \quad \mathbb{P}(A_{i_1} \cap \dots \cap A_{i_k}) = \mathbb{P}(A_{i_1}) \dots \mathbb{P}(A_{i_k}).$$

Exercice 4.4. Soit $B_i = A_i$ ou A_i^c . Il est clair que A_1, \dots, A_n sont indépendants si et seulement si B_1, \dots, B_n sont indépendants.

On verra dans le Chapitre V une explication simple de cette propriété. \square

Exemple 4.5. On serait tenté de penser que si $(A_i)_{1 \leq i \leq n}$ sont deux-à-deux indépendants, alors ils sont aussi indépendants. Ceci est en fait faux, même pour $n = 3$. L'indépendance deux-à-deux dit que

$$\mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_1)\mathbb{P}(A_2),$$

$$\mathbb{P}(A_2 \cap A_3) = \mathbb{P}(A_2)\mathbb{P}(A_3),$$

$$\mathbb{P}(A_1 \cap A_3) = \mathbb{P}(A_1)\mathbb{P}(A_3),$$

alors que l'indépendance de A_1 , A_2 et A_3 demande une condition supplémentaire :

$$\mathbb{P}(A_1 \cap A_2 \cap A_3) = \mathbb{P}(A_1)\mathbb{P}(A_2)\mathbb{P}(A_3).$$

Cette dernière n'est pas du tout garantie par les trois conditions précédentes.

On jette deux fois une pièce de monnaie parfaite. Soient $A_1 = "P \text{ la première fois}"$, $A_2 = "P \text{ la deuxième fois}"$, et $A_3 = "deux fois le même résultat"$. Il est intuitivement clair que A_1 , A_2 et A_3 ne peuvent être indépendants, car A_3 est totalement déterminés par A_1 et A_2 (par exemple, si A_1 et A_2 sont réalisés, alors on sait que A_3 est aussi réalisé). Mathématiquement, on a $\mathbb{P}(A_1) = \mathbb{P}(A_2) = \mathbb{P}(A_3) = 1/2$, tandis que

$$\mathbb{P}(A_1 \cap A_2 \cap A_3) = \mathbb{P}(A_1 \cap A_2) = \frac{1}{4} \neq \frac{1}{8} = \mathbb{P}(A_1)\mathbb{P}(A_2)\mathbb{P}(A_3),$$

ce qui prouve que A_1 , A_2 et A_3 ne sont pas indépendants.

On a pourtant

$$\begin{aligned} \mathbb{P}(A_1 \cap A_2) &= \frac{1}{4} = \mathbb{P}(A_1)\mathbb{P}(A_2), \\ \mathbb{P}(A_2 \cap A_3) &= \mathbb{P}(A_2 \cap A_1) = \frac{1}{4} = \mathbb{P}(A_2)\mathbb{P}(A_3), \\ \mathbb{P}(A_1 \cap A_3) &= \mathbb{P}(A_1 \cap A_2) = \frac{1}{4} = \mathbb{P}(A_1)\mathbb{P}(A_3), \end{aligned}$$

d'où l'indépendance deux-à-deux. □

Exercice 4.6. (Cet exercice ne sera pas traité dans le cours.) Désignons par $\varphi(n)$ la fonction d'Euler de la théorie des nombres, c'est-à-dire, $\varphi(n)$ est le nombre des entiers plus petits que n et qui sont premiers avec n . Alors

$$(2.3) \quad \varphi(n) = n \prod_{p: p|n} \left(1 - \frac{1}{p}\right),$$

où le produit est sur tous les facteurs premiers p de n .

Pour (re)démontrer cette formule bien connue de la théorie des nombres, on considère le modèle probabiliste suivant : on choisit au hasard un nombre parmi $\{1, 2, \dots, n\}$ avec équi-probabilité. Pour tout nombre premier p , soit

$$A_p = \{ \text{le nombre choisi est divisible par } p \}.$$

Soient p_1, p_2, \dots, p_m les facteurs premiers de n . Montrons d'abord que $A_{p_1}, A_{p_2}, \dots, A_{p_m}$ sont des événements indépendants. D'après la Proposition 2.2, il suffit de montrer que $\mathbb{P}(A_{p_{i_1}} \cap \dots \cap A_{p_{i_k}}) = \mathbb{P}(A_{p_{i_1}}) \cdots \mathbb{P}(A_{p_{i_k}})$. Or, il est clair que

$$\begin{aligned} \mathbb{P}(A_{p_i}) &= \mathbb{P}\left(\text{le nombre est un élément de } \in \{p_i, 2p_i, 3p_i, \dots, \frac{n}{p_i}p_i\}\right) \\ &= \frac{n/p_i}{n} = \frac{1}{p_i}, \end{aligned}$$

tandis que

$$\begin{aligned} \mathbb{P}(A_{p_{i_1}} \cap \dots \cap A_{p_{i_k}}) &= \mathbb{P}\left(\text{le nombre choisi est un élément de } \{q, 2q, 3q, \dots, \frac{n}{q}q\}\right) \\ &= \frac{n/q}{n} = \frac{1}{p_{i_1} \cdots p_{i_k}}, \end{aligned}$$

où $q := p_{i_1} \cdots p_{i_k}$, ce qui donne $\mathbb{P}(A_{p_{i_1}} \cap \dots \cap A_{p_{i_k}}) = \mathbb{P}(A_{p_{i_1}}) \cdots \mathbb{P}(A_{p_{i_k}})$.

On sait donc que $A_{p_1}, A_{p_2}, \dots, A_{p_m}$ sont indépendants. Ainsi, $A_{p_1}^c, A_{p_2}^c, \dots, A_{p_m}^c$ sont aussi indépendants. On a,

$$\mathbb{P}(A_{p_1}^c \cap \dots \cap A_{p_k}^c) = \mathbb{P}(A_{p_1}^c) \cdots \mathbb{P}(A_{p_k}^c) = \prod_{i=1}^k \left(1 - \frac{1}{p_i}\right).$$

D'autre part, $\mathbb{P}(A_{p_1}^c \cap \dots \cap A_{p_k}^c) = \frac{\varphi(n)}{n}$. D'où l'identité (2.3). \square

5. Probabilité conditionnelle

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

Définition 5.1. Soient $A \in \mathcal{F}$ et $B \in \mathcal{F}$, avec $\mathbb{P}(B) > 0$. La probabilité conditionnelle de A sachant B est définie par

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

Exemple 5.2. On joue avec deux dés, un rouge et un blanc.

Soit $A = \{\text{on obtient deux fois } 6\}$. Alors $\mathbb{P}(A) = 1/36$.

Soit $B = \{\text{le dé blanc donne } 6\}$. On a $\mathbb{P}(B) = 1/6$. Donc la probabilité d'avoir deux 6, sachant que le dé blanc a donné 6, est

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

On remarque que $\mathbb{P}(A|B) \neq \mathbb{P}(A)$. □

Propriété 5.3. Si $\mathbb{P}(B) > 0$, alors A et B sont indépendants $\iff \mathbb{P}(A|B) = \mathbb{P}(A)$.

Propriété 5.4. Soit $\mathbb{P}(B) > 0$. L'application $A \mapsto \mathbb{P}(A|B)$ (pour $A \in \mathcal{F}$) vérifie les axiomes de Kolmogorov; elle définit donc une nouvelle probabilité sur l'espace mesurable (Ω, \mathcal{F}) .

Ainsi, toutes les propriétés pour les probabilités sont valables pour les probabilités conditionnelles. Par exemple, si $(A_n)_{n \geq 1}$ est une suite croissante d'éléments de \mathcal{F} , alors

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n | B\right) = \lim_{n \rightarrow \infty} \mathbb{P}(A_n | B).$$

La preuve de la Propriété 5.4, qui n'est pas présentée ici, est tout à fait triviale.

La propriété suivante permet de déduire $\mathbb{P}(B|A)$ à partir de $\mathbb{P}(A|B)$.

Propriété 5.5. Soient A et B deux événements tels que $\mathbb{P}(A) > 0$ et $\mathbb{P}(B) > 0$. On a alors

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

Preuve. Par définition, $\mathbb{P}(B|A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A|B)\mathbb{P}(B)}{\mathbb{P}(A)}$. □

Propriété 5.6. Soit $\{A_n\}_{n=1,2,\dots}$ une partition finie ou infinie dénombrable de (Ω, \mathcal{F}) , avec $\mathbb{P}(A_n) > 0$ pour tout n . Alors

$$\mathbb{P}(B) = \sum_n \mathbb{P}(A_n)\mathbb{P}(B|A_n), \quad \forall B \in \mathcal{F}.$$

Preuve. D'après la Propriété 2.8, $\mathbb{P}(B) = \sum_n \mathbb{P}(B \cap A_n)$. Le résultat découle alors de la relation $\mathbb{P}(B \cap A_n) = \mathbb{P}(A_n)\mathbb{P}(B|A_n)$. □

Exemple 5.7. On tire deux cartes d'un paquet de 52. Quelle est la probabilité que la deuxième carte soit "Dame" ?

La situation dépend du résultat de la première carte. Si celle-ci est Dame, alors la probabilité (conditionnelle) que la deuxième carte soit Dame est $3/51$. Si celle-ci ne l'est pas, alors la probabilité (conditionnelle) en question devient $4/51$. On est donc dans

le cadre de probabilités conditionnelles, et l'on fait une discussion sur le résultat de la première carte.

Soient

$$A_1 := \{\text{la première carte est Dame}\},$$

$$A_2 := \{\text{la première carte n'est pas Dame}\}.$$

On a $A_1 \cup A_2 = \Omega$ et $A_1 \cap A_2 = \emptyset$. Soit $B := \{\text{la deuxième carte est Dame}\}$. Par la Propriété 5.6,

$$\mathbb{P}(B) = \mathbb{P}(A_1)\mathbb{P}(B|A_1) + \mathbb{P}(A_2)\mathbb{P}(B|A_2) = \frac{4}{52} \times \frac{3}{51} + \frac{48}{52} \times \frac{4}{51} = \frac{1}{13}.$$

On constate que la deuxième carte a exactement la même chance d'être Dame que la première, ce qui est rassurant par exemple pour le tirage au sort dans un tournoi sportif. \square

Propriété 5.8 (Formule de Bays). Soit $\{A_n\}_{n=1,2,\dots}$ une partition finie ou infinie dénombrable de (Ω, \mathcal{F}) , avec $\mathbb{P}(A_n) > 0$ pour tout n . Alors pour tout $B \in \mathcal{F}$ tel que $\mathbb{P}(B) > 0$,

$$\mathbb{P}(A_n|B) = \frac{\mathbb{P}(A_n)\mathbb{P}(B|A_n)}{\sum_m \mathbb{P}(A_m)\mathbb{P}(B|A_m)}.$$

Preuve. Par définition,

$$\mathbb{P}(A_n|B) = \frac{\mathbb{P}(A_n \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(A_n)\mathbb{P}(B|A_n)}{\mathbb{P}(B)},$$

ce qui complète la preuve à l'aide de la propriété précédente. \square

Exemple 5.9. Dans un concours de pronostic sportif pour un match entre l'équipe E et l'équipe F (pas de match nul), les participants sont composés à 50% d'étudiants de Jussieu, à 20% d'étudiants de Dauphine, et à 30% d'étudiants d'Orsay. On constate que 60% des étudiants de Jussieu prévoient l'équipe E gagnante, ainsi que 30% des étudiants de Dauphine, et 90% des étudiants d'Orsay. On tire une personne au hasard parmi les participants : elle pronostique sur l'équipe F gagnante. Quelle est la probabilité qu'il s'agit de quelqu'un de Jussieu ?

On définit $A_1 := \{\text{Jussieu}\}$, $A_2 := \{\text{Dauphine}\}$, et $A_3 := \{\text{Orsay}\}$. On sait que ces trois événements forment une partition de Ω , avec $\mathbb{P}(A_1) = 0,5$, $\mathbb{P}(A_2) = 0,2$, et $\mathbb{P}(A_3) = 0,3$.

Soit $B := \{\text{supporteur de l'équipe F}\}$. Par hypothèse, $\mathbb{P}(B|A_1) = 0,4$, $\mathbb{P}(B|A_2) = 0,7$, et $\mathbb{P}(B|A_3) = 0,1$. D'après la formule de Bays,

$$\begin{aligned} \mathbb{P}(A_1|B) &= \frac{\mathbb{P}(A_1)\mathbb{P}(B|A_1)}{\mathbb{P}(A_1)\mathbb{P}(B|A_1) + \mathbb{P}(A_2)\mathbb{P}(B|A_2) + \mathbb{P}(A_3)\mathbb{P}(B|A_3)} = \\ &= \frac{0,5 \times 0,4}{0,5 \times 0,4 + 0,2 \times 0,7 + 0,3 \times 0,1} = \frac{20}{37} > \frac{1}{2}. \end{aligned} \quad \square$$

Chapitre 2. Variables aléatoires discrètes

1. Loi d'une variable aléatoire discrète

Dans tout le chapitre, $(\Omega, \mathcal{F}, \mathbb{P})$ désigne un espace de probabilité, tel que Ω soit au plus dénombrable.

Soit (E, \mathcal{E}) un espace mesurable. On appelle variable aléatoire (v.a.) discrète à valeurs dans E toute application mesurable $X : \Omega \rightarrow E$. Le plus souvent, $E = \mathbb{N}, \mathbb{Z}, \mathbb{R}$ ou \mathbb{R}^d .

Un exemple de v.a. discrète : le nombre de bons numéros au loto (voir l'Exemple 3.3 traité dans le Chapitre 1).

On note P_X la mesure image de \mathbb{P} par X . C'est-à-dire que, pour tout $A \in \mathcal{E}$,

$$P_X(A) = \mathbb{P}(\{\omega \in \Omega : X(\omega) \in A\}) = \mathbb{P}(\{X \in A\}) = \mathbb{P}(X \in A).$$

P_X , qui est une mesure de probabilité sur E (car $P_X(E) = \mathbb{P}(\Omega) = 1$), est appelée la **loi de la variable aléatoire** X ou encore sa distribution.

Soit $\{x_k, k = 1, 2, \dots\} := X(\Omega)$, qui représente l'ensemble des valeurs que X peut prendre. Cet ensemble est au plus infini dénombrable, car c'est le cas pour Ω . On suppose que $\{x_k\} \in \mathcal{E}, \forall k$. La loi de X est caractérisée par les données (x_k) et (p_k) , où :

$$p_k = P_X(\{x_k\}) = \mathbb{P}(\{X = x_k\}), \quad k = 1, 2, \dots$$

En effet, pour tout $A \in \mathcal{E}$,

$$P_X(A) = \sum_{k: x_k \in A} p_k.$$

2. Moments d'une v.a. discrète réelle

Lorsque $(E, \mathcal{E}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ (espace réel muni de la tribu borélienne), on dit que X est une v.a. réelle.

Définition 2.1. Soit X une v.a. réelle, et soit $\{x_k, k = 1, 2, \dots\}$ l'ensemble des valeurs prises par X . Si

$$\sum_k |x_k| p_k < \infty,$$

on dit que X admet un moment d'ordre 1, et on pose

$$\mathbb{E}(X) = \sum_k x_k p_k.$$

On appelle $\mathbb{E}(X)$ l'**espérance** — ou la moyenne — de X .

Plus généralement, pour $n \geq 1$, on dit que X admet un moment d'ordre n si X^n admet un moment d'ordre 1.

Remarque. Il est clair que si la v.a. X est bornée, c'est-à-dire s'il existe un réel $M > 0$ tel que $|X(\omega)| \leq M$ pour tout $\omega \in \Omega$, alors X admet des moments de tous ordres. C'est le cas en particulier lorsque $\#(\Omega) < \infty$. \square

Définition 2.2. Si X admet un moment d'ordre 2, alors elle admet un moment d'ordre 1. On pose alors

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

et on appelle cette quantité la **variance** de X .

Le fait que l'existence d'un moment d'ordre 2 implique celle d'un moment d'ordre 1, ainsi que la propriété suivante, seront prouvés dans le chapitre 3 dans un contexte beaucoup plus général.

Propriété 2.3. Si X admet un moment d'ordre 2, alors

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

Par conséquent, $\text{Var}(X) \geq 0$, et on appelle $\sqrt{\text{Var}(X)}$ l'écart-type de X .

Propriété 2.4. Soit g une fonction réelle telle que

$$\sum_k |g(x_k)| p_k < \infty.$$

La v.a. $g \circ X$, que l'on note souvent $g(X)$, admet un moment d'ordre 1, avec

$$\mathbb{E}(g(X)) = \sum_k g(x_k) p_k.$$

En particulier, si $\sum_k (x_k)^2 p_k < \infty$, alors X admet un moment d'ordre 2, et on a

$$\mathbb{E}(X^2) = \sum_k (x_k)^2 p_k.$$

Heuristiquement, $\mathbb{E}(X)$ mesure la “moyenne” de la v.a. X , tandis que $\text{Var}(X)$ mesure la dispersion par rapport à la moyenne. Par exemple, soient X et Y les variables qui désignent les notes de deux étudiants. On suppose :

$$\begin{aligned} \mathbb{P}(X = 10) &= 1/4, & \mathbb{P}(X = 11) &= 1/2, & \mathbb{P}(X = 12) &= 1/4, \\ \mathbb{P}(Y = 3) &= 1/8, & \mathbb{P}(Y = 9) &= 1/6, & \mathbb{P}(Y = 11) &= 3/8, \\ \mathbb{P}(Y = 14) &= 1/4, & \mathbb{P}(Y = 18) &= 1/12. \end{aligned}$$

Les variables X et Y étant bornées, elles admettent des moments de tous ordres.

$$\begin{aligned} \mathbb{E}(X) &= \frac{1}{4} \times 10 + \frac{1}{2} \times 11 + \frac{1}{4} \times 12 = 11, \\ \mathbb{E}(Y) &= \frac{1}{8} \times 3 + \frac{1}{6} \times 9 + \frac{3}{8} \times 11 + \frac{1}{4} \times 14 + \frac{1}{12} \times 18 = 11. \end{aligned}$$

Pourtant,

$$\begin{aligned} \mathbb{E}(X^2) &= \frac{1}{4} \times 10^2 + \frac{1}{2} \times 11^2 + \frac{1}{4} \times 12^2 = 121,5, \\ \mathbb{E}(Y^2) &= \frac{1}{8} \times 3^2 + \frac{1}{6} \times 9^2 + \frac{3}{8} \times 11^2 + \frac{1}{4} \times 14^2 + \frac{1}{12} \times 18^2 = 136. \end{aligned}$$

Donc

$$\begin{aligned} \text{Var}(X) &= \mathbb{E}(X^2) - (\mathbb{E}X)^2 = 121,5 - 121 = \frac{1}{2}, \\ \text{Var}(Y) &= \mathbb{E}(Y^2) - (\mathbb{E}Y)^2 = 136 - 121 = 15. \end{aligned}$$

Les résultats du deuxième étudiant sont donc beaucoup moins réguliers que ceux du premier, ce qui est, a priori, parfaitement clair.

3. Exemples de lois discrètes

Exemple (variable constante). $\mathbb{P}(X = c) = 1$ pour un réel c . $\mathbb{E}(X) = c$. $\text{Var}(X) = 0$. P_X est la masse de Dirac en c . □

Exemple (loi de Bernoulli). On dit que X est une variable de Bernoulli de paramètre $p \in [0, 1]$ si

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

On a $\mathbb{E}(X) = p$, $\text{Var}(X) = p(1 - p)$. □

Exemple (loi binomiale). On dit que X est une variable binomiale de paramètres (n, p) , où $n \in \mathbb{N}^*$ et $p \in [0, 1]$ si

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

On a $\mathbb{E}(X) = np$, $\text{Var}(X) = np(1 - p)$. Interprétation : si l'on jette n fois une pièce de monnaie avec $\mathbb{P}(P) = p$ et $\mathbb{P}(F) = 1 - p$. Le nombre de fois où l'on obtient P suit la loi binomiale de paramètres (n, p) .

Exemple (loi géométrique). On dit que X est une variable géométrique de paramètre $p \in]0, 1[$ si

$$\mathbb{P}(X = n) = (1 - p)p^n, \quad n = 0, 1, 2, \dots$$

On a $\mathbb{E}(X) = p/(1 - p)$, $\text{Var}(X) = p/(1 - p)^2$. Interprétation : si l'on jette une pièce de monnaie avec $\mathbb{P}(P) = p$ et $\mathbb{P}(F) = 1 - p$. Le nombre de fois où l'on obtient P avant d'avoir un F suit la loi géométrique de paramètre p .

Exemple (loi de Poisson). On dit que X est une variable de Poisson de paramètre $\theta > 0$ si

$$\mathbb{P}(X = n) = e^{-\theta} \frac{\theta^n}{n!}, \quad n = 0, 1, 2, \dots$$

On a $\mathbb{E}(X) = \theta$, $\text{Var}(X) = \theta$.

4. Fonction génératrice d'une v.a. à valeurs entières positives

Dans cette section, on suppose que X est une v.a. à valeurs dans \mathbb{N} .

Définition 4.1. La fonction génératrice de X est la fonction $G_X : [0, 1] \rightarrow [0, 1]$ définie par

$$G_X(s) = \mathbb{E}(s^X) = \sum_{n=0}^{\infty} s^n \mathbb{P}(X = n), \quad s \in [0, 1],$$

avec la convention $0^0 := 1$.

Propriété 4.2. La fonction génératrice détermine la loi de X . Plus précisément, pour tout $n \geq 0$,

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

où $G_X^{(n)}(0)$ désigne la dérivée n -ième de G_X en 0.

Preuve. La série entière $s \mapsto \sum_{n=0}^{\infty} s^n \mathbb{P}(X = n)$ a un rayon de convergence supérieur ou égal à 1 (car $\sum_{n=0}^{\infty} \mathbb{P}(X = n)$ converge). On peut donc récupérer $\mathbb{P}(X = n)$ en utilisant la formule qui relie les coefficients d'une série entière avec les dérivées en 0. \square

Exemple (loi de Bernoulli). Si X est une variable de Bernoulli de paramètre $p \in [0, 1]$, alors

$$G_X(s) = 1 - p + sp.$$

Exemple (loi binomiale). Si X suit la loi binomiale de paramètres (n, p) ,

$$G_X(s) = \sum_{i=0}^n \binom{n}{i} p^i (1-p)^{n-i} s^i = (sp + 1 - p)^n.$$

Exemple (loi géométrique). Soit X une variable géométrique de paramètre $p \in]0, 1[$,

$$G_X(s) = \sum_{n=0}^{\infty} (1-p)p^n s^n = \frac{1-p}{1-sp}.$$

Exemple (loi de Poisson). Si X est une variable de Poisson de paramètre $\theta > 0$,

$$G_X(s) = \sum_{n=0}^{\infty} e^{-\theta} \frac{\theta^n}{n!} s^n = e^{s\theta - \theta} = e^{\theta(s-1)}.$$

Chapitre 3. Variables aléatoires

1. Loi d'une variable aléatoire

Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité, et soit E un espace muni d'une tribu \mathcal{E} . On appelle variable aléatoire (v.a.) à valeurs dans E toute application mesurable $X : \Omega \rightarrow E$.

Souvent, on prend $E = \mathbb{R}$ muni de la tribu borélienne, et on appelle X v.a. réelle.

Soit X une v.a. à valeurs dans E , et soit \tilde{E} un espace muni d'une tribu $\tilde{\mathcal{E}}$. Si $g : (E, \mathcal{E}) \rightarrow (\tilde{E}, \tilde{\mathcal{E}})$ est une fonction mesurable, alors $g(X)$ est une v.a., à valeurs dans \tilde{E} .

Définition 1.1. Soit $X : (\Omega, \mathcal{F}) \rightarrow (E, \mathcal{E})$ une variable aléatoire. On appelle **loi** de X (ou encore sa distribution) la mesure de probabilité P_X sur (E, \mathcal{E}) donnée par

$$P_X(A) = \mathbb{P}(X^{-1}(A)) = \mathbb{P}(X \in A) = \mathbb{P}(\{\omega \in \Omega : X(\omega) \in A\}), \quad A \in \mathcal{E}.$$

[En termes de la théorie d'intégration, il s'agit de la mesure image de \mathbb{P} par l'application mesurable X .]

Définition 1.2. Si une v.a. réelle X est intégrable par rapport à \mathbb{P} , c'est-à-dire si $\int_{\Omega} |X| d\mathbb{P} < \infty$, on dit alors que X admet un moment d'ordre 1, et on pose

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

On appelle $\mathbb{E}(X)$ **espérance** de X (ou sa moyenne).

Théorème 1.3. P_X est l'unique mesure sur (E, \mathcal{E}) telle que pour toute fonction $g : E \rightarrow \mathbb{R}$ mesurable et bornée, on ait

$$\mathbb{E}(g(X)) = \int_E g(x) P_X(dx).$$

Preuve. L'unicité est évidente en prenant pour g l'indicatrice d'un ensemble mesurable générique $A \in \mathcal{F}$, puisque dans ce cas

$$\mathbb{E}(g(X)) = \mathbb{E}(\mathbf{1}_{X \in A}) = \mathbb{P}(X \in A) = P_X(A).$$

On montre maintenant l'identité pour toute fonction g mesurable et bornée. En fait, on prouve un résultat plus fort, en montrant l'identité pour toute fonction g mesurable telle que $\int_E |g(x)| P_X(dx) < \infty$.

L'identité est vraie pour toute fonction mesurable g étagée (c'est à dire, combinaison linéaire finie de fonctions indicatrices). Puisque toute fonction mesurable positive s'écrit comme limite simple d'une suite croissante de fonctions étagées positives, par passage à la limite à l'aide du théorème de convergence monotone, on voit que l'identité reste vraie pour g mesurable positive telle que $\int_E g(x) P_X(dx) < \infty$ (sans cette condition d'intégrabilité, l'identité reste néanmoins valable, sous la forme de $\infty = \infty$). Enfin, en considérant g^+ et g^- , on obtient l'identité pour g telle que $\int_E |g(x)| P_X(dx) < \infty$. \square

La preuve du théorème précédent nous donne le résultat suivant.

Théorème 1.4. Soit $g : E \rightarrow \mathbb{R}$ une fonction mesurable telle que $\int_E |g(x)| P_X(dx) < \infty$. Alors $g(X)$ admet un moment d'ordre 1, avec

$$\mathbb{E}(g(X)) = \int_E g(x) P_X(dx).$$

Remarque. Dans le cas d'une v.a. discrète, $P_X = \sum_k p_k \delta_{x_k}$ (notation : $p_k = \mathbb{P}(X = x_k)$, δ_{x_k} = masse de Dirac en x_k). Soit g une fonction mesurable telle que $\sum_k |g(x_k)| p_k < \infty$. On a, par le Théorème 1.4,

$$\mathbb{E}(g(X)) = \int_E g(x) P_X(dx) = \sum_k g(x_k) p_k.$$

C'est l'énoncé de la Propriété 2.4 du Chapitre 2. \square

2. Fonction de répartition d'une v.a. réelle

On suppose que X est une variable aléatoire **réelle**, c'est-à-dire que $E = \mathbb{R}$ (muni de sa tribu borélienne). Dans ce cas-là, la loi de X (c'est-à-dire P_X) est une mesure de probabilité sur \mathbb{R} .

Définition 2.1. On appelle **fonction de répartition** de X la fonction $F_X : \mathbb{R} \rightarrow [0, 1]$ donnée par

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

Propriété 2.2. (i) La fonction F_X est croissante, avec

$$\lim_{x \rightarrow -\infty} F_X(x) = 0 \quad \text{et} \quad \lim_{x \rightarrow \infty} F_X(x) = 1.$$

(ii) La fonction F_X est continue à droite, sa limite à gauche au point x vaut

$$F_X(x-) \left(= \lim_{y \rightarrow x-} F_X(y) \right) = \mathbb{P}(X < x) = P_X(]-\infty, x[).$$

Preuve. (i) Si $x \leq x'$, alors $]-\infty, x] \subset]-\infty, x']$ et on a bien $F_X(x) \leq F_X(x')$. Comme $]-\infty, \infty[= \bigcup_{n \geq 1}]-\infty, n]$, et comme la famille $(]-\infty, n])_{n \geq 1}$ est croissante, on a

$$\lim_{n \rightarrow \infty} F_X(n) = P_X(]-\infty, \infty[) = 1.$$

De même, $]-\infty, \infty[= \bigcup_{n \geq 1}]-n, \infty[$ et comme $F_X(-n) = 1 - P_X(]-n, \infty[)$, on a donc

$$1 - \lim_{n \rightarrow \infty} F_X(-n) = P_X(]-\infty, \infty[) = 1.$$

(ii) Soit $(x_n, n \geq 1)$ une suite réelle qui décroît vers x . La suite des événements $\{X \leq x_n\}$ est décroissante et $\{X \leq x\} = \bigcap_{n \geq 1} \{X \leq x_n\}$. On a donc $F_X(x) = \lim_{n \rightarrow \infty} F_X(x_n)$.

Supposons maintenant que $(y_n, n \geq 1)$ est une suite strictement croissante qui converge vers x . Alors la suite des événements $\{X \leq y_n\}$ est croissante et $\{X < x\} = \bigcup_{n \geq 1} \{X \leq y_n\}$. On a donc $\mathbb{P}(X < x) = \lim_{n \rightarrow \infty} F_X(y_n)$. \square

Propriété 2.3. Pour tout $x \in \mathbb{R}$, on a

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

(En particulier, si X est une v.a. discrète, la fonction de répartition F_X caractérise la loi de X .)

Preuve. Il suffit d'écrire $\{X \leq x\} = \{X < x\} \cup \{X = x\}$ et appliquer l'additivité d'une probabilité. \square

Propriété 2.4. La fonction de répartition F_X caractérise la loi P_X .

Preuve. Pour tout intervalle de type $]a, b]$, on a $F_X(b) - F_X(a) = P_X(]a, b])$. Une mesure sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ (tribu borélienne de \mathbb{R}) étant déterminée par la donnée des masses qu'elle attribue aux intervalles de ce type (théorème de classe monotone), on en déduit que F_X caractérise la loi de X . \square

Remarque. (i) Pour une variable aléatoire réelle X admettant un moment d'ordre 1, il y a deux façons d'écrire son espérance :

$$\mathbb{E}(X) = \int_{\Omega} X \, d\mathbb{P} = \int_{\mathbb{R}} x P_X(dx).$$

(ii) Dans la Section 5 du présent chapitre, on va se donner une condition nécessaire et suffisante pour qu'une fonction définie sur \mathbb{R} soit la fonction de répartition associée à une loi de probabilité sur \mathbb{R} . \square

Exemple 2.5. Soit X une variable aléatoire réelle dont la fonction de répartition est F_X . On s'intéresse à la fonction de répartition des variables aléatoires suivantes :

$$\begin{aligned} Y &:= X + a, \\ Z &:= X^2, \\ U &:= X^+ = \max(X, 0), \end{aligned}$$

où $a \in \mathbb{R}$ est un réel connu.

On a,

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

Pour calculer $F_Z(x)$, on remarque que $F_Z(x) = \mathbb{P}(Z \leq x) = 0$ lorsque $x < 0$. Si $x \geq 0$,

$$\begin{aligned} F_Z(x) &= \mathbb{P}(X^2 \leq x) = \mathbb{P}(-\sqrt{x} \leq X \leq \sqrt{x}) \\ &= \mathbb{P}(X \leq \sqrt{x}) - \mathbb{P}(X < -\sqrt{x}) \\ &= F_X(\sqrt{x}) - F_X(-\sqrt{x} - 0). \end{aligned}$$

Donc

$$F_Z(x) = \begin{cases} F_X(\sqrt{x}) - F_X(-\sqrt{x} - 0) & \text{si } x \geq 0, \\ 0 & \text{sinon.} \end{cases}$$

On calcule maintenant $F_U(x)$. Il est clair que $F_U(x) = 0$ pour $x < 0$. Lorsque $x \geq 0$,

$$F_U(x) = \mathbb{P}(\max(X, 0) \leq x) = \mathbb{P}(X \leq x) = F_X(x).$$

En conclusion,

$$F_U(x) = \begin{cases} F_X(x) & \text{si } x \geq 0, \\ 0 & \text{sinon.} \end{cases} \quad \square$$

3. Moments d'une v.a. réelle

Propriété 3.1. Soient X_1, \dots, X_n des v.a. réelles admettant toutes un moment d'ordre 1, et soient a_1, \dots, a_n des réels quelconques. Alors la v.a. $\sum_{i=1}^n a_i X_i$ admet également un moment d'ordre 1, et

$$\mathbb{E} \left(\sum_{i=1}^n a_i X_i \right) = \sum_{i=1}^n a_i \mathbb{E}(X_i).$$

Inégalité de Markov. Soit X une v.a. réelle qui admet un moment d'ordre 1. Pour tout $a > 0$, on a

$$\mathbb{P}(|X| \geq a) \leq \frac{\mathbb{E}(|X|)}{a}.$$

Preuve. On a

$$\mathbb{E}(|X|) = \int_{\Omega} |X| \, d\mathbb{P} \geq \int_{\Omega} |X| \mathbf{1}_{\{|X| \geq a\}} \, d\mathbb{P} \geq \int_{\Omega} a \mathbf{1}_{\{|X| \geq a\}} \, d\mathbb{P} = a \mathbb{P}(|X| \geq a).$$

D'où l'inégalité cherchée. □

Définition 3.2. On dit qu'une v.a. réelle X admet un moment d'ordre n si $\int_{\Omega} |X|^n \, d\mathbb{P} < \infty$.

Propriété 3.3. Soient $0 < p < q$. Si $\int_{\Omega} |X|^q \, d\mathbb{P} < \infty$, alors $\int_{\Omega} |X|^p \, d\mathbb{P} < \infty$. En particulier, si X admet un moment d'ordre n , alors elle admet des moments de tous ordre $m \leq n$.

Preuve. On peut supposer sans perte de généralité que $q > p = 1$ (sinon, on considère $|X|^p$ au lieu de $|X|$ dans le raisonnement ci-dessous).

Par l'inégalité de Hölder (valable, en réalité, pour les mesures qui ne sont pas nécessairement des mesures de probabilité), pour tout $(a, b) \in]1, \infty[^2$ tel que $a^{-1} + b^{-1} = 1$, on a

$$\int_{\Omega} |XY| \, d\mathbb{P} \leq \left(\int_{\Omega} |X|^a \, d\mathbb{P} \right)^{1/a} \left(\int_{\Omega} |Y|^b \, d\mathbb{P} \right)^{1/b}.$$

Appliquons l'inégalité à $a = q$, $b = q/(q - 1)$ et $Y = 1$, et obtient

$$\int_{\Omega} |X| \, d\mathbb{P} \leq \left(\int_{\Omega} |X|^q \, d\mathbb{P} \right)^{1/q}.$$

D'où le résultat. □

Définition et Propriété 3.4. Si X admet un moment d'ordre 2, alors elle admet un moment d'ordre 1. On pose alors

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

et on appelle cette quantité la **variance** de X . On a en plus

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

Par conséquent, $\text{Var}(X) \geq 0$, et on appelle $\sqrt{\text{Var}(X)}$ l'écart-type de X .

Preuve. L'existence d'un moment d'ordre 1 découle de la Propriété 3.3. On a

$$\begin{aligned} \mathbb{E} \left[(X - \mathbb{E}X)^2 \right] &= \mathbb{E} \left[X^2 + (\mathbb{E}X)^2 - 2X \mathbb{E}(X) \right] \\ &= \mathbb{E}(X^2) + (\mathbb{E}X)^2 - 2(\mathbb{E}X)^2 = \mathbb{E}(X^2) - (\mathbb{E}X)^2. \end{aligned} \quad \square$$

Propriété 3.5. Une variable aléatoire est constante avec probabilité 1 si et seulement si sa variance est nulle. La v.a. est alors égale à sa moyenne avec probabilité 1.

Preuve. Si X a pour variance nulle, alors $\mathbb{E}[(X - \mathbb{E}X)^2] = 0$. Donc $\mathbb{P}(X - \mathbb{E}X \neq 0) = 0$, c'est-à-dire $\mathbb{P}(X = \mathbb{E}(X)) = 1$.

La réciproque, qui est évidente, a déjà été démontrée dans le premier exemple de la Section 3 du Chapitre 2. □

Propriété 3.6. Soit X une v.a. réelle ayant un moment d'ordre 2, et soient a et b deux réels. Alors $aX + b$ admet aussi un moment d'ordre 2, avec

$$\text{Var}(aX + b) = a^2 \text{Var}(X).$$

Preuve. Si X^2 est intégrable par rapport à \mathbb{P} , alors $(aX + b)^2$ l'est également. On a

$$\begin{aligned} \text{Var}(aX + b) &= \mathbb{E} \left[(aX + b - \mathbb{E}(aX + b))^2 \right] = \mathbb{E} \left[(aX + b - a\mathbb{E}X - b)^2 \right] \\ &= \mathbb{E} \left[a^2(X - \mathbb{E}X)^2 \right] = a^2 \text{Var}(X). \end{aligned} \quad \square$$

Inégalité de Bienaymé–Tchebychev. Soit X une variable aléatoire qui admet un moment d'ordre 2. Pour tout réel $a > 0$, on a

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

Preuve. On a

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq a) = \mathbb{P}(|X - \mathbb{E}(X)|^2 \geq a^2),$$

et il ne reste qu'à appliquer l'inégalité de Markov. \square

4. Densité d'une v.a. réelle

Rappelons brièvement un résultat très important de la théorie des mesures. Soient μ_1 et μ_2 deux mesures σ -finies sur l'espace mesurable (E, \mathcal{E}) . On dit que μ_1 est absolument continue par rapport à μ_2 si, pour tout $A \in \mathcal{E}$, $\mu_2(A) = 0 \implies \mu_1(A) = 0$. Le théorème de Radon–Nikodym dit que ceci est le cas si et seulement s'il existe $f : E \rightarrow \mathbb{R}_+$ mesurable telle que $\mu_1 = f \bullet \mu_2$ (c'est-à-dire $\mu_1(A) = \int_A f d\mu_2$ pour tout $A \in \mathcal{E}$). Souvent, on écrit dans ce cas $\mu_1 \ll \mu_2$, et $f := \frac{d\mu_1}{d\mu_2}$.

Soit X une v.a. réelle. Si la mesure de probabilité P_X est absolument continue par rapport à la mesure de Lebesgue, c'est-à-dire si $P_X(dx) = f_X(x) dx$ (dx étant la mesure de Lebesgue), alors on dit que X admet une **fonction de densité** f_X . Ceci signifie que $P_X(A) = \int_A f_X(x) dx$ pour toute partie borélienne $A \subset \mathbb{R}$. On a alors

$$F_X(x) = \int_{-\infty}^x f_X(u) du, \quad x \in \mathbb{R}.$$

La fonction de répartition est dans ce cas-là une fonction continue sur \mathbb{R} .

Remarque. (i) Si la fonction de répartition F_X est continue sur \mathbb{R} , X n'admet pas nécessairement une densité.

(ii) Si F_X est continue et presque partout dérivable, telle que F_X soit une fonction primitive de sa dérivée, alors X a une densité qui est la dérivée de F_X .

(iii) Si f et g sont des fonctions de densité d'une même v.a. réelle, alors $f = g$ presque partout. \square

Théorème 4.1. Si X admet la densité f_X , et si g est une fonction mesurable sur \mathbb{R} telle que

$$\int_{\mathbb{R}} |g(x)| f_X(x) dx < \infty,$$

alors $g(X)$ admet un moment d'ordre 1, avec

$$\mathbb{E}(g(X)) = \int_{\mathbb{R}} g(x) f_X(x) dx.$$

En particulier, (i) si $\int_{\mathbb{R}} |x| f_X(x) dx < \infty$, alors X admet un moment 1, avec

$$\mathbb{E}(X) = \int_{\mathbb{R}} x f_X(x) dx.$$

(ii) Si A est un borélien de \mathbb{R} , alors

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

Preuve. On a $P_X(dx) = f_X(x) dx$. Le résultat découle alors du Théorème 1.4. \square

On se donne quelques exemples de densités de probabilité de v.a. réelles.

Exemple 4.2 (loi uniforme). La mesure de Lebesgue sur $[0, 1]$ (ou $(0, 1)$) est une mesure de probabilité, qu'on appelle la loi uniforme sur $[0, 1]$ (ou $(0, 1)$). La densité est

$$f_X(x) = \mathbf{1}_{(0,1)}(x).$$

En particulier,

$$\mathbb{E}(X) = \int_0^1 x dx = \frac{1}{2}, \quad \text{Var}(X) = \frac{1}{12}.$$

On peut aussi facilement calculer la fonction de répartition :

$$F_X(x) = \begin{cases} 0 & \text{si } x < 0, \\ x & \text{si } 0 \leq x \leq 1, \\ 1 & \text{si } x > 1. \end{cases}$$

Plus généralement, si $a < b$, on appelle loi uniforme sur $[a, b]$ la mesure de probabilité $(b - a)^{-1} \mathbf{1}_{[a,b]}(x) dx$. Donc pour la loi uniforme sur $[a, b]$, la fonction de densité est

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

On remarque que si X suit la loi uniforme sur $[0, 1]$, alors la v.a. $Y = (b - a)X + a$ suit la loi uniforme sur $[a, b]$. En effet, pour tout $x \in \mathbb{R}$, on a $F_Y(x) = \mathbb{P}(Y \leq x) = \mathbb{P}(X \leq \frac{x-a}{b-a}) = 0$ si $x < a$, et $F_Y(x) = 1$ si $x > b$; et si $x \in [a, b]$,

$$F_Y(x) = \int_0^{(x-a)/(b-a)} du = \frac{x-a}{b-a}.$$

Donc

$$F_Y(x) = \int_{-\infty}^x \frac{1}{b-a} \mathbf{1}_{]a,b[}(t) dt, \quad x \in \mathbb{R}.$$

L'expression à droite n'est autre que la fonction de répartition d'une v.a. suivant la loi uniforme sur $[a, b]$. Comme la fonction de répartition détermine la loi d'une v.a. réelle, on conclut que Y suit la loi uniforme sur $[a, b]$. \square

Exemple 4.3 (loi exponentielle). Pour tout $\theta > 0$, la mesure $\theta e^{-\theta x} \mathbf{1}_{\mathbb{R}_+}(x) dx$ a pour masse 1, c'est donc une mesure de probabilité que l'on appelle la loi exponentielle de paramètre θ . La densité de cette loi de probabilité vaut donc

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

En particulier,

$$\mathbb{E}(X) = \int_0^{\infty} x \theta e^{-\theta x} dx = \frac{1}{\theta}.$$

Plus généralement, pour $n \geq 1$,

$$\mathbb{E}(X^n) = \int_0^{\infty} x^n \theta e^{-\theta x} dx = \frac{\Gamma(n+1)}{\theta^n} = \frac{n!}{\theta^n}.$$

On peut aussi facilement calculer la fonction de répartition :

$$F_X(x) = \begin{cases} 0 & \text{si } x \leq 0, \\ 1 - e^{-\theta x} & \text{si } x > 0. \end{cases} \quad \square$$

Exemple 4.4 (loi de Cauchy standard). La mesure sur \mathbb{R} de densité $\frac{1}{\pi} \frac{1}{1+x^2} dx$ a bien pour masse 1, c'est une mesure de probabilité sur \mathbb{R} que l'on appelle loi de Cauchy standard. La densité de la loi de Cauchy standard est donc

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

Bien que $x \mapsto x f_X(x)$ soit une fonction impaire sur \mathbb{R} (on serait donc tenté de dire que $\mathbb{E}(X) = 0$), X n'admet pas de moment d'ordre 1, car $|x| f_X(x)$ n'est pas intégrable par rapport à la mesure de Lebesgue.

La fonction de répartition de la loi de Cauchy standard est

$$F_X(x) = \frac{1}{2} + \frac{1}{\pi} \text{Arctan}(x). \quad \square$$

Exemple 4.5 (loi gaussienne). En écrivant

$$\begin{aligned} \left(\int_{-\infty}^{\infty} e^{-x^2/2} dx \right)^2 &= \left(\int_{-\infty}^{\infty} e^{-x^2/2} dx \right) \left(\int_{-\infty}^{\infty} e^{-y^2/2} dy \right) \\ &= \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dy e^{-(x^2+y^2)/2} \\ &= 2\pi \int_0^{\infty} e^{-r^2/2} r dr \\ &= 2\pi, \end{aligned}$$

on obtient l'identité $\int_{-\infty}^{\infty} e^{-x^2/2} dx = \sqrt{2\pi}$. Ainsi, la mesure $\frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$ a pour masse 1, on l'appelle la loi gaussienne standard (ou la loi normale standard). La densité est donc

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

Plus généralement, pour tout $\mu \in \mathbb{R}$ et $\sigma^2 > 0$, la mesure sur \mathbb{R} ,

$$\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx$$

a également pour masse 1 (il s'agit d'un changement de variables simple), on l'appelle la loi gaussienne $\mathcal{N}(\mu, \sigma^2)$. On peut vérifier que

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

Il est facile de vérifier que si X suit la loi gaussienne $\mathcal{N}(\mu, \sigma^2)$, alors $Y := \frac{X-\mu}{\sigma}$ suit la loi gaussienne standard $\mathcal{N}(0, 1)$.

La loi gaussienne standard est souvent appelée la loi gaussienne centrée réduite, car elle a pour espérance 0 et pour variance 1.

La fonction de répartition de la loi gaussienne n'admet pas d'expression explicite simple. \square

Exemple 4.6. Soit X une v.a. uniforme sur $]-\frac{\pi}{2}, \frac{\pi}{2}[$, c'est-à-dire que la densité de X est

$$f_X(x) = \frac{1}{\pi} \mathbf{1}_{]-\frac{\pi}{2}, \frac{\pi}{2}[}(x).$$

Soit $Y = \tan(X)$. On s'intéresse à la loi de la v.a. Y . Pour tout $x \in \mathbb{R}$,

$$\mathbb{P}(Y \leq x) = \mathbb{P}(X \leq \text{Arctan}(x)) = \frac{1}{\pi} \int_{-\pi/2}^{\text{Arctan}(x)} du = \frac{1}{2} + \frac{1}{\pi} \text{Arctan}(x).$$

Y suit donc la loi de Cauchy standard.

Si X est une v.a. uniforme sur $(0, 1)$, on a vu dans l'Exemple 4.2 que $\pi(X - \frac{1}{2})$ suit la loi uniforme sur $] -\frac{\pi}{2}, \frac{\pi}{2}[$. Donc $Y = \tan(\pi(X - \frac{1}{2}))$ suit la loi de Cauchy standard. On verra dans la section suivante qu'il s'agit d'un phénomène très général. \square

5. Loi uniforme et loi réelle quelconque

Proposition 5.1. (i) Soit $F : \mathbb{R} \rightarrow [0, 1]$ une fonction croissante, continue à droite, telle que

$$\lim_{x \rightarrow -\infty} F(x) = 0, \quad \lim_{x \rightarrow \infty} F(x) = 1.$$

Alors elle est fonction de répartition associée à une loi de probabilité réelle.

(ii) Soit $F : \mathbb{R} \rightarrow [0, 1]$ la fonction de répartition associée à la loi de probabilité μ sur \mathbb{R} , c'est-à-dire $F(x) = \mu(]-\infty, x])$. Soit $F^{-1} : [0, 1] \rightarrow \mathbb{R}$ l'inverse continue-à-droite de F , c'est-à-dire

$$F^{-1}(x) = \inf\{y \in \mathbb{R} : F(y) > x\}, \quad x \in]0, 1[.$$

Si X suit la loi uniforme sur $[0, 1]$, alors la variable aléatoire $Y = F^{-1}(X)$ suit la loi μ .

Preuve. Soit $F : \mathbb{R} \rightarrow [0, 1]$ une fonction croissante, continue à droite, telle que

$$\lim_{x \rightarrow -\infty} F(x) = 0, \quad \lim_{x \rightarrow \infty} F(x) = 1.$$

Soit $Y = F^{-1}(X)$. Il s'agit de vérifier que Y a pour fonction de répartition F .

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

$$(*) \quad Y \leq x \iff F^{-1}(X) \leq x \iff \inf\{y : F(y) > X\} \leq x.$$

D'une part, la dernière assertion de $(*)$ est satisfaite dès que $F(x) > X$, autrement dit on a $\{X < F(x)\} \subset \{Y \leq x\}$. Comme X suit la loi uniforme, il en découle que

$$F(x) = \mathbb{P}(X < F(x)) \leq \mathbb{P}(Y \leq x).$$

D'autre part, si la dernière assertion de $(*)$ est vérifiée, alors $F(y) > X$ pour tout $y > x$, c'est-à-dire $\{Y \leq x\} \subset \{X < F(y)\}$. On a donc $\mathbb{P}(Y \leq x) \leq \mathbb{P}(X < F(y)) = F(y)$ (puisque X suit la loi uniforme). On fait tendre y vers x , et comme F est continue à droite, on obtient

$$\mathbb{P}(Y \leq x) \leq F(x).$$

En conclusion, on a donc vérifié que F est la fonction de répartition de Y . \square

La Proposition 5.1 a une application importante en informatique. La plupart des ordinateurs permettent de générer des pseudo “variables aléatoires” de “loi” approximativement uniforme sur $[0, 1]$, c’est-à-dire que l’ordinateur fournit à l’utilisateur un nombre réel “aléatoire” X tel que $X \in [0, x]$ avec fréquence d’environ x pour $x \in [0, 1]$. La Proposition 5.1 permet donc de générer des variables aléatoires réelles de lois arbitraires à partir d’une variable aléatoire X suivant la loi uniforme sur $[0, 1]$.

Chapitre 4. Vecteurs aléatoires

1. Fonction de répartition d'un vecteur aléatoire

Un vecteur aléatoire de dimension N est une variable aléatoire à valeurs dans \mathbb{R}^N (muni de sa tribu borélienne). On peut donc appliquer tous les résultats sur les v.a. générales.

On note $X = (X_1, \dots, X_N)$. La fonction de répartition de X est

$$\begin{aligned} F_X(x_1, \dots, x_N) &= P_X (] - \infty, x_1] \times \dots \times] - \infty, x_N]) \\ &= \mathbb{P} (X_1 \leq x_1, \dots, X_N \leq x_N), \quad (x_1, \dots, x_N) \in \mathbb{R}^N. \end{aligned}$$

On remarque que F_X est une fonction sur \mathbb{R}^N , à valeurs dans $[0, 1]$, croissante en chaque x_i .

2. Densité d'un vecteur aléatoire

Si P_X (loi du vecteur aléatoire X) est absolument continue par rapport à la mesure de Lebesgue sur \mathbb{R}^N , c'est-à-dire, si

$$P_X(dx_1 \cdots dx_N) = f_X(x_1, \dots, x_N) dx_1 \cdots dx_N,$$

alors f_X est appelée la densité de X . Relation entre densité et fonction de répartition :

$$F_X(x_1, \dots, x_N) = \int \cdots \int_{]-\infty, x_1] \times \cdots \times]-\infty, x_N]} f_X(u_1, \dots, u_N) du_1 \cdots du_N.$$

Plus généralement, si une fonction $g : \mathbb{R}^N \rightarrow \mathbb{R}$ est telle que

$$\int \cdots \int_{\mathbb{R}^N} |g(x_1, \dots, x_N)| f_X(x_1, \dots, x_N) dx_1 \cdots dx_N < \infty,$$

alors la v.a. réelle $g(X_1, \dots, X_N)$ admet un moment d'ordre 1, avec

$$\mathbb{E}(g(X_1, \dots, X_N)) = \int \cdots \int_{\mathbb{R}^N} g(x_1, \dots, x_N) f_X(x_1, \dots, x_N) dx_1 \cdots dx_N.$$

En particulier, pour tout borélien $A \subset \mathbb{R}^N$,

$$\mathbb{P}(X \in A) = \int \cdots \int_A f_X(x_1, \dots, x_N) dx_1 \cdots dx_N,$$

ce qui n'est pas surprenant, car par définition, $P_X(A) = \int \cdots \int_A f_X(x_1, \dots, x_N) dx_1 \cdots dx_N$.

Exemple 2.1. Soit (X, Y) de densité

$$f_{(X,Y)}(x, y) = \theta^2 e^{-\theta(x+y)} \mathbf{1}_{\mathbb{R}_+^*}(x) \mathbf{1}_{\mathbb{R}_+^*}(y),$$

où $\theta > 0$ est une constante fixée. Comme $\int \int_{\mathbb{R}^2} f_{(X,Y)}(x, y) dx dy = 1$, $f_{(X,Y)}$ est une fonction de densité sur \mathbb{R}^2 .

On s'intéresse à la fonction de répartition $F_{(X,Y)}$ du vecteur aléatoire (X, Y) . Par définition,

$$F_{(X,Y)}(x, y) = \int \int_{]-\infty, x] \times]-\infty, y]} f_{(X,Y)}(u, v) du dv.$$

Vu que $f_{(X,Y)} = 0$ dès que $u \leq 0$ ou $v \leq 0$, on déduit que $F_{(X,Y)}(x, y) \neq 0$ seulement si $x > 0$ et $y > 0$. Dans ce cas,

$$F_{(X,Y)}(x, y) = \int \int_{]0, x] \times]0, y]} \theta^2 e^{-\theta(u+v)} du dv = (1 - e^{-\theta x})(1 - e^{-\theta y}).$$

Par conséquent,

$$F_{(X,Y)}(x, y) = \begin{cases} (1 - e^{-\theta x})(1 - e^{-\theta y}), & \text{si } x > 0 \text{ et } y > 0, \\ 0, & \text{sinon.} \end{cases} \quad \square$$

Exemple 2.2. Soit (X, Y) de densité

$$f_{(X,Y)}(x, y) = \frac{1}{x^3} \mathbf{1}_A(x, y),$$

où $A = \{(x, y) \in \mathbb{R}^2 : x \geq 1, 0 \leq y \leq x\}$. On cherche à calculer la fonction de répartition.

Il est facile de vérifier que $\int \int_{\mathbb{R}^2} f_{(X,Y)}(x, y) dx dy = 1$. Donc $f_{(X,Y)}$ est bien une fonction de densité sur \mathbb{R}^2 .

On sait que pour tout $(x, y) \in \mathbb{R}^2$, la fonction de répartition de (X, Y) est

$$\begin{aligned} F_{(X,Y)}(x, y) &= \int \int_{]-\infty, x] \times]-\infty, y]} f_{(X,Y)}(u, v) \, du \, dv \\ &= \int \int_{A \cap (]-\infty, x] \times]-\infty, y])} \frac{1}{u^3} \, du \, dv. \end{aligned}$$

Si $x < 1$ ou si $y < 0$, on a $A \cap (]-\infty, x] \times]-\infty, y]) = \emptyset$, donc $F_{(X,Y)}(x, y) = 0$.

Si $x \geq 1$ et $0 \leq y \leq 1$,

$$F_{(X,Y)}(x, y) = \int_1^x \int_0^y \frac{1}{u^3} \, du \, dv = \frac{y}{2} \left(1 - \frac{1}{x^2} \right).$$

Si $x \geq 1$ et $1 < y < x$,

$$\begin{aligned} F_{(X,Y)}(x, y) &= \int_1^x \int_0^1 \frac{1}{u^3} \, du \, dv + \int_1^y \, dv \int_v^x \frac{1}{u^3} \, du \\ &= \frac{1}{2} \left(1 - \frac{1}{x^2} \right) + \frac{1}{2} \int_1^y \, dv \left(\frac{1}{v^2} - \frac{1}{x^2} \right) \\ &= \frac{1}{2} \left(1 - \frac{1}{x^2} \right) + \frac{1}{2} \left(1 - \frac{1}{y} \right) - \frac{y-1}{2x^2} \\ &= 1 - \frac{1}{2y} - \frac{y}{2x^2}. \end{aligned}$$

Finalement, si $x \geq 1$ et $y \geq x$,

$$F_{(X,Y)}(x, y) = \int_1^x \, du \int_0^u \, dv \frac{1}{u^3} = \int_1^x \, du \frac{1}{u^2} = 1 - \frac{1}{x}.$$

Si en plus, on veut par exemple calculer la probabilité $\mathbb{P}(2 \leq X \leq 3)$, on a

$$\begin{aligned} \mathbb{P}(2 \leq X \leq 3) &= \int \int_{[2, 3] \times \mathbb{R}} f_{(X,Y)}(u, v) \, du \, dv = \int_2^3 \, du \int_0^u \, dv \frac{1}{u^3} \\ &= \int_2^3 \, du \frac{1}{u^2} = \frac{1}{2} - \frac{1}{3} = \frac{1}{6}. \end{aligned} \quad \square$$

3. Moments

Définition 3.1. Soit $X = (X_1, \dots, X_N)$ un vecteur aléatoire. Si X_1, \dots, X_N admettent toutes un moment d'ordre 1, alors le vecteur

$$\mathbb{E}(X) = (\mathbb{E}(X_1), \dots, \mathbb{E}(X_N))$$

est appelé l'espérance (ou l'espérance-vecteur) de X .

Définition 3.2. Soient X et Y deux v.a. réelles admettant toutes des moments d'ordre 2. Alors

$$\text{Cov}(X, Y) = \mathbb{E}((X - \mathbb{E}X)(Y - \mathbb{E}Y))$$

est appelée la **covariance** entre X et Y .

On remarque que la covariance $\text{Cov}(X, Y)$ est bien définie tant que X et Y admettent des moments d'ordre 2. En effet, rappelons l'inégalité de Cauchy–Schwarz (qui est un cas spécial de l'inégalité de Hölder que l'on a vue au chapitre précédent) :

Inégalité de Cauchy–Schwarz. Si X et Y sont des v.a. réelles admettant des moments d'ordre 2, alors

$$\mathbb{E}(|XY|) \leq \sqrt{\mathbb{E}(X^2)} \sqrt{\mathbb{E}(Y^2)}.$$

La proposition suivante nous donne quelques propriétés élémentaires de la covariance.

Propriété 3.3. Si X et Y admettent un moment d'ordre 2, alors

- (i) $\text{Cov}(X, Y) = \text{Cov}(Y, X)$.
- (ii) $\text{Cov}(X, X) = \text{Var}(X)$.
- (iii) $\text{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X) \cdot \mathbb{E}(Y)$.
- (iv) $\text{Cov}(aX + b, cY + d) = ac \text{Cov}(X, Y)$ pour tous réels a, b, c et d .
- (v) $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y)$.

Preuve. (i), (ii) Evidente.

(iii) On a

$$\begin{aligned} \text{Cov}(X, Y) &= \mathbb{E}(XY - X\mathbb{E}(Y) - Y\mathbb{E}(X) + \mathbb{E}(X) \cdot \mathbb{E}(Y)) \\ &= \mathbb{E}(XY) - \mathbb{E}(X) \cdot \mathbb{E}(Y) - \mathbb{E}(X) \cdot \mathbb{E}(Y) + \mathbb{E}(X) \cdot \mathbb{E}(Y) \\ &= \mathbb{E}(XY) - \mathbb{E}(X) \cdot \mathbb{E}(Y). \end{aligned}$$

(iv) Par définition,

$$\begin{aligned} \text{Cov}(aX + b, cY + d) &= \mathbb{E}\left((aX + b - a\mathbb{E}X - b)(cY + d - c\mathbb{E}Y - d)\right) \\ &= ac \mathbb{E}((X - \mathbb{E}X)(Y - \mathbb{E}Y)) \\ &= ac \text{Cov}(X, Y). \end{aligned}$$

(v) On a

$$\begin{aligned}\operatorname{Var}(X + Y) &= \mathbb{E}((X + Y - \mathbb{E}(X + Y))^2) = \mathbb{E}(((X - \mathbb{E}X) + (Y - \mathbb{E}Y))^2) \\ &= \mathbb{E}((X - \mathbb{E}X)^2 + (Y - \mathbb{E}Y)^2 + 2(X - \mathbb{E}X)(Y - \mathbb{E}Y)) \\ &= \operatorname{Var}(X) + \operatorname{Var}(Y) + 2\operatorname{Cov}(X, Y).\end{aligned}\quad \square$$

Définition 3.4. Si X et Y admettent des moments d'ordre 2 telles que $\operatorname{Var}(X)\operatorname{Var}(Y) > 0$, alors

$$\rho(X, Y) := \frac{\operatorname{Cov}(X, Y)}{\sqrt{\operatorname{Var}(X)} \sqrt{\operatorname{Var}(Y)}}$$

est appelé le coefficient de corrélation entre X et Y .

Propriété 3.5. Si X et Y admettent des moments d'ordre 2 telles que $\operatorname{Var}(X)\operatorname{Var}(Y) > 0$, alors $\rho(X, Y) \in [-1, 1]$.

Preuve. Par l'inégalité de Cauchy–Schwarz,

$$\begin{aligned}|\operatorname{Cov}(X, Y)| &\leq \mathbb{E}(|(X - \mathbb{E}X)(Y - \mathbb{E}Y)|) \\ &\leq \sqrt{\mathbb{E}((X - \mathbb{E}X)^2)} \sqrt{\mathbb{E}((Y - \mathbb{E}Y)^2)} = \sqrt{\operatorname{Var}(X)} \sqrt{\operatorname{Var}(Y)}.\end{aligned}\quad \square$$

Définition 3.6. Soit $X = (X_1, \dots, X_N)$ un vecteur aléatoire tel que chaque composante admet un moment d'ordre 2. On appelle

$$D = \begin{pmatrix} \operatorname{Var}(X_1) & \operatorname{Cov}(X_1, X_2) & \cdots & \operatorname{Cov}(X_1, X_N) \\ \operatorname{Cov}(X_2, X_1) & \operatorname{Var}(X_2) & \cdots & \operatorname{Cov}(X_2, X_N) \\ \vdots & \vdots & \ddots & \vdots \\ \operatorname{Cov}(X_N, X_1) & \operatorname{Cov}(X_N, X_2) & \cdots & \operatorname{Var}(X_N) \end{pmatrix}$$

matrice de variances-covariances (ou: matrice de covariances, ou encore: matrice de dispersion) de X . Il s'agit de la matrice symétrique $D = (D_{ij})_{1 \leq i, j \leq N}$ avec $D_{ij} = \operatorname{Cov}(X_i, X_j)$.

Propriété 3.7. La matrice D est positive au sens que pour tout $a = (a_1, \dots, a_N) \in \mathbb{R}^N$,

$$a D a^t = \sum_{1 \leq i, j \leq N} D_{ij} a_i a_j \geq 0.$$

Preuve. On a

$$\begin{aligned}
 \sum_{1 \leq i, j \leq N} D_{ij} a_i a_j &= \sum_{1 \leq i, j \leq N} a_i a_j \mathbb{E}((X_i - \mathbb{E}X_i)(X_j - \mathbb{E}X_j)) \\
 &= \mathbb{E} \left(\sum_{1 \leq i, j \leq N} a_i (X_i - \mathbb{E}X_i) a_j (X_j - \mathbb{E}X_j) \right) \\
 &= \mathbb{E} \left(\left(\sum_{i=1}^N a_i (X_i - \mathbb{E}X_i) \right)^2 \right) \geq 0. \quad \square
 \end{aligned}$$

Il existe une extension de la Propriété 3.3 (iv)–(v) à plusieurs variables aléatoires réelles.

Propriété 3.8. (i) Si X_1, \dots, X_N sont des v.a. réelles admettant toutes un moment d'ordre 2, alors

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

(ii) Si $X_1, \dots, X_N, Y_1, \dots, Y_M$ sont des v.a. réelles admettant toutes un moment d'ordre 2, alors pour tous réels $a_1, \dots, a_N, b_1, \dots, b_M$,

$$\text{Cov} \left(\sum_{i=1}^N a_i X_i, \sum_{j=1}^M b_j Y_j \right) = \sum_{i=1}^N \sum_{j=1}^M a_i b_j \text{Cov}(X_i, Y_j).$$

4. Changement de variables

Soit X un vecteur aléatoire à valeurs dans un ouvert $\Delta \subset \mathbb{R}^N$ (souvent : $\Delta = \mathbb{R}^N$) qui admet comme densité $f_X = f_X \mathbf{1}_\Delta$. Soit

$$h : \Delta \rightarrow D$$

(avec $D \subset \mathbb{R}^N$) un C^1 -difféomorphisme de Δ dans D , c'est-à-dire que h est une application de Δ dans D qui est bijective et continûment différentiable, et dont l'application réciproque

h^{-1} (de D dans Δ) est aussi continûment différentiable. Alors le vecteur aléatoire $Y = h \circ X = h(X)$ a pour densité

$$\begin{aligned} f_Y(y) &= \frac{f_X(x)}{|(Dh)(x)|} \mathbf{1}_D(y), \quad y = h(x) \\ &= \frac{f_X(h^{-1}(y))}{|(Dh)(h^{-1}(y))|} \mathbf{1}_D(y), \end{aligned}$$

où $(Dh)(x)$ est le jacobien de h en x : soit $h = (h_1, \dots, h_N)$,

$$(Dh)(x) = \det \begin{pmatrix} \frac{\partial h_1(x_1, \dots, x_N)}{\partial x_1} & \dots & \frac{\partial h_1(x_1, \dots, x_N)}{\partial x_N} \\ \vdots & \ddots & \vdots \\ \frac{\partial h_N(x_1, \dots, x_N)}{\partial x_1} & \dots & \frac{\partial h_N(x_1, \dots, x_N)}{\partial x_N} \end{pmatrix}.$$

Ceci est appelé formule du changement de variables. N'oublions pas la **valeur absolue** pour le jacobien!

Exemple 4.1 (suite de l'Exemple 2.1). Soit (X, Y) un vecteur aléatoire de densité

$$f_{(X,Y)}(x, y) = \theta^2 e^{-\theta(x+y)} \mathbf{1}_{\mathbb{R}_+^* \times \mathbb{R}_+^*}(x, y).$$

On définit les v.a. $U = X + Y$ et $V = X - Y$. Quelle est la loi du vecteur aléatoire (U, V) ?

L'application $h(x, y) = (x + y, x - y)$ est un C^1 -difféomorphisme de $\mathbb{R}_+^* \times \mathbb{R}_+^*$ dans $D = \{(u, v) : u > 0, |v| < u\}$, avec $h^{-1}(u, v) = (\frac{u+v}{2}, \frac{u-v}{2})$. Le jacobien est

$$(Dh)(x) = \det \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} = -2.$$

Donc (U, V) a pour densité

$$f_{(U,V)}(u, v) = \frac{1}{2} \theta^2 e^{-\theta(x+y)} \mathbf{1}_{\mathbb{R}_+^* \times \mathbb{R}_+^*}(x, y) = \frac{1}{2} \theta^2 e^{-\theta u} \mathbf{1}_D(u, v), \quad \square$$

Preuve et Remarque. On démontre la formule du changement de variables. Notons qu'il est souvent beaucoup plus facile d'utiliser ce que l'on appelle la **méthode de la fonction muette** :

Soit $\varphi : \mathbb{R}^N \rightarrow \mathbb{R}$ une fonction borélienne bornée. On a

$$\mathbb{E}(\varphi(Y)) = \mathbb{E}(\varphi(h(X))) = \int_{\Delta} \varphi(h(x)) f_X(x) dx.$$

Par un changement de variables (usuel), l'intégrale vaut

$$\int_D \varphi(y) \frac{f_X(h^{-1}(y))}{|(Dh)(h^{-1}(y))|} dy.$$

(Rappelons le changement de variables usuel : si $g : \mathbb{R}^N \rightarrow \mathbb{R}$ est une fonction borélienne qui est intégrable par rapport à $\mathbf{1}_D(y) dy$, alors $\int_D g(y) dy = \int_\Delta g(h(x)) |(Dh)(x)| dx$.)

Ceci étant vrai pour toute fonction φ , on en déduit (par exemple en prenant des fonctions indicatrices) que $h(X)$ a pour densité

$$\frac{f_X(h^{-1}(y))}{|(Dh)(h^{-1}(y))|} \mathbf{1}_D(y). \quad \square$$

Exemple 4.2. Soit X une v.a. qui suit la loi uniforme sur $]0, 1[$, et soit $h :]0, 1[\rightarrow \mathbb{R}$ avec $h(x) = \tan(\pi(x - \frac{1}{2}))$. On cherche la loi de $Y = h(X)$. On a vu dans l'Exemple 4.6 du Chapitre 4 que Y suit la loi de Cauchy standard. On redémontre ce résultat ici en utilisant la méthode de la fonction muette.

Remarquons que

$$h^{-1}(y) = \frac{1}{2} + \frac{\text{Arctan}(y)}{\pi}, \quad y \in \mathbb{R}.$$

Soit $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ une fonction borélienne bornée. On a

$$\mathbb{E}(\varphi(Y)) = \mathbb{E}(\varphi(h(X))) = \int_{-1}^1 \varphi(h(x)) dx = \int_{\mathbb{R}} \varphi(y) \frac{dy}{\pi(1+y^2)}.$$

En particulier, soit $\varphi = \mathbf{1}_{]-\infty, a]}$, et on a

$$F_Y(a) = \int_{-\infty}^a \frac{dy}{\pi(1+y^2)},$$

où $F_Y(a) := \mathbb{P}(Y \leq a)$ est la fonction de répartition de Y en a . A droite, on reconnaît la fonction de répartition (en a) de la loi de Cauchy standard. Comme la fonction de répartition détermine une loi de probabilité, on déduit que Y suit la loi de Cauchy standard, et admet comme densité $f_Y(y) = \frac{1}{\pi(1+y^2)}$. \square

Exemple 4.3. Supposons que X suit la loi uniforme sur $]0, 1[$. On s'intéresse à la loi de

$$Y = \ln\left(\frac{1}{X}\right).$$

Pour nous, $h(x) = \ln(1/x)$, qui est un C^1 -difféomorphisme de $]0, 1[$ dans \mathbb{R}_+^* , avec $h^{-1}(y) = e^{-y}$, pour $y \in \mathbb{R}_+^*$.

Soit $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ une fonction borélienne bornée. On a

$$\mathbb{E}(\varphi(Y)) = \mathbb{E}(\varphi(h(X))) = \int_0^1 \varphi(h(x)) \, dx = \int_{\mathbb{R}} \varphi(y) \mathbf{1}_{\mathbb{R}_+^*}(y) e^{-y} \, dy.$$

En particulier, soit $\varphi = \mathbf{1}_{]-\infty, a]}$, et on a

$$\mathbb{P}(Y \leq a) = \int_{-\infty}^a \mathbf{1}_{\mathbb{R}_+^*}(y) e^{-y} \, dy.$$

Le terme à droite est la fonction de répartition en a de la loi exponentielle de paramètre 1. Donc $P_Y = \mathbf{1}_{\mathbb{R}_+^*}(x) e^{-x} \, dx$, autrement dit, Y suit la loi exponentielle de paramètre 1. \square

Exemple 4.4. Soit X une variable aléatoire suivant la loi de Cauchy standard. On cherche d'abord à déterminer la loi de $1/X$, et ensuite à montrer que $\mathbb{E}(|X|^{1-\varepsilon}) \geq 1$, quel que soit $\varepsilon \in]0, 1]$.

On calcule d'abord la loi de $Y = 1/X$. Soit $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ une fonction borélienne bornée. On a

$$\begin{aligned} \mathbb{E}(\varphi(Y)) &= \mathbb{E}(\varphi(1/X)) = \int_{\mathbb{R}} \varphi(1/x) \frac{1}{\pi} \frac{1}{1+x^2} \, dx \\ &= \int_{\mathbb{R}} \varphi(y) \frac{1}{\pi} \frac{1}{1+y^{-2}} \frac{1}{y^2} \, dy \\ &= \int_{\mathbb{R}} \varphi(y) \frac{1}{\pi} \frac{1}{1+y^2} \, dy. \end{aligned}$$

Donc Y admet comme densité $f_Y(y) = \frac{1}{\pi} \frac{1}{1+y^2}$. Autrement dit, $1/X$ suit également la loi de Cauchy standard.

Les variables X et $1/X$ suivent la même loi. Posons $Z = |X|^{1-\varepsilon}$. On a en particulier, $\mathbb{E}(Z) = \mathbb{E}(\frac{1}{Z}) \in]0, \infty[$. Or, d'après l'inégalité de Cauchy-Schwarz,

$$\mathbb{E}(Z) \cdot \mathbb{E}\left(\frac{1}{Z}\right) \geq \left[\mathbb{E}\left(\sqrt{Z} \cdot \sqrt{\frac{1}{Z}}\right) \right]^2 = 1.$$

On obtient alors $\mathbb{E}(|X|^{1-\varepsilon}) \geq 1$. \square

5. Lois marginales

Soit $X = (X_1, \dots, X_N)$ un vecteur aléatoire. On appelle loi marginale la loi de tout sous-vecteur $(X_{i_1}, \dots, X_{i_k})$ (avec $k < N$) extrait de X . Pour distinguer P_X des lois marginales, on l'appelle de temps en temps loi conjointe de X_1, \dots, X_N .

Afin de simplifier l'écriture, on ne s'intéresse qu'à la loi de (X_1, \dots, X_k) .

Proposition 5.1. Soit

$$\begin{aligned} F_{(X_1, \dots, X_N)}(x_1, \dots, x_N) &= P_{(X_1, \dots, X_N)} (] - \infty, x_1] \times \dots \times] - \infty, x_N]) \\ &= \mathbb{P}(X_1 \leq x_1, \dots, X_N \leq x_N), \end{aligned}$$

la fonction de répartition de (X_1, \dots, X_N) . Soit $1 \leq k < N$. Alors la fonction de répartition de (X_1, \dots, X_k) est

$$F_{(X_1, \dots, X_k)}(x_1, \dots, x_k) = \lim_{x_{k+1} \rightarrow +\infty, \dots, x_N \rightarrow +\infty} F_{(X_1, \dots, X_N)}(x_1, \dots, x_N).$$

On écrit souvent

$$F_{(X_1, \dots, X_k)}(x_1, \dots, x_k) = F_{(X_1, \dots, X_N)}(x_1, \dots, x_k, +\infty, \dots, +\infty).$$

En particulier, la loi (marginale) de X_i est donnée par sa fonction de répartition

$$F_{X_i}(x_i) = F_{(X_1, \dots, X_N)}(+\infty, \dots, +\infty, x_i, +\infty, \dots, +\infty).$$

Preuve. Fixons $(x_1, \dots, x_k) \in \mathbb{R}^k$. Soient $(x_{k+1}^{(n)})_{n \geq 1}, \dots, (x_N^{(n)})_{n \geq 1}$ des suites croissantes tendant toutes vers l'infini. La suite d'événements (indexée par n)

$$\{X_1 \leq x_1, \dots, X_k \leq x_k, X_{k+1} \leq x_{k+1}^{(n)}, \dots, X_N \leq x_N^{(n)}\}$$

croît vers $\{X_1 \leq x_1, \dots, X_k \leq x_k\}$. Donc

$$\begin{aligned} \mathbb{P}(X_1 \leq x_1, \dots, X_k \leq x_k) \\ = \lim_{n \rightarrow \infty} \mathbb{P}\left(X_1 \leq x_1, \dots, X_k \leq x_k, X_{k+1} \leq x_{k+1}^{(n)}, \dots, X_N \leq x_N^{(n)}\right), \end{aligned}$$

ou encore,

$$F_{(X_1, \dots, X_k)}(x_1, \dots, x_k) = \lim_{n \rightarrow \infty} F_{(X_1, \dots, X_N)}(x_1, \dots, x_k, x_{k+1}^{(n)}, \dots, x_N^{(n)}).$$

Ceci prouve le théorème. □

Proposition 5.2. Si $f_{(X_1, \dots, X_N)}$ est la densité du vecteur aléatoire $X = (X_1, \dots, X_N)$, alors le vecteur aléatoire (X_1, \dots, X_k) admet la densité

$$f_{(X_1, \dots, X_k)}(x_1, \dots, x_k) = \int \dots \int_{\mathbb{R}^{N-k}} f_{(X_1, \dots, X_N)}(x_1, \dots, x_N) dx_{k+1} \dots dx_N.$$

Preuve. D'après la Proposition 5.1 et le théorème de Fubini,

$$\begin{aligned} F_{(X_1, \dots, X_k)}(x_1, \dots, x_k) \\ = \int_{-\infty}^{x_1} du_1 \cdots \int_{-\infty}^{x_k} du_k \left(\int_{\mathbb{R}} \cdots \int_{\mathbb{R}} f_{(X_1, \dots, X_N)}(u_1, \dots, u_N) du_{k+1} \cdots du_N \right). \end{aligned}$$

Donc (X_1, \dots, X_k) a une loi qui est absolument continue par rapport à la mesure de Lebesgue sur \mathbb{R}^k , de densité

$$f_{(X_1, \dots, X_k)}(x_1, \dots, x_k) = \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} f_{(X_1, \dots, X_N)}(x_1, \dots, x_k, u_{k+1}, \dots, u_N) du_{k+1} \cdots du_N. \quad \square$$

Exemple 5.3. Soit (X, Y) un vecteur aléatoire de densité

$$f_{(X, Y)}(x, y) = \theta^2 e^{-\theta(x+y)} \mathbf{1}_{\mathbb{R}_+^* \times \mathbb{R}_+^*}(x, y).$$

Par la Proposition 5.2, on sait que la loi de Y est absolument continue par rapport à la mesure de Lebesgue. De plus, la densité de X est

$$\begin{aligned} f_X(x) &= \int_{\mathbb{R}} f_{(X, Y)}(x, y) dy = \int_0^{\infty} \theta^2 e^{-\theta(x+y)} \mathbf{1}_{\mathbb{R}_+^*}(x) dy \\ &= \theta^2 e^{-\theta x} \mathbf{1}_{\mathbb{R}_+^*}(x) \int_0^{\infty} e^{-\theta y} dy = \theta e^{-\theta x} \mathbf{1}_{\mathbb{R}_+^*}(x). \end{aligned}$$

Donc X suit la loi exponentielle de paramètre θ .

De même, on peut vérifier que Y suit également la loi exponentielle de paramètre θ .
□

La proposition ci-dessus a une analogue pour les vecteurs aléatoires discrets.

Proposition 5.4. Si (X_1, \dots, X_N) est un vecteur aléatoire à valeurs dans $E_1 \times \cdots \times E_N$. Alors pour tout $k < N$,

$$\mathbb{P}(X_1 = x_1, \dots, X_k = x_k) = \sum_{(x_{k+1}, \dots, x_N) \in E_{k+1} \times \cdots \times E_N} \mathbb{P}(X_1 = x_1, \dots, X_N = x_N).$$

Chapitre 5. Indépendance

1. Variables aléatoires indépendantes

Définition 1.1. Soient X et Y deux variables aléatoires à valeurs dans E et \tilde{E} , respectivement. On dit que X et Y sont indépendantes si pour tout $A \in \mathcal{E}$ et tout $B \in \tilde{\mathcal{E}}$,

$$\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A) \mathbb{P}(Y \in B).$$

Quand on exprime l'indépendance en termes des distributions, on obtient que deux v.a. X et Y sont indépendantes si et seulement si $P_{(X,Y)}$ — la loi du couple (X, Y) — vérifie

$$P_{(X,Y)}(A \times B) = P_X(A) P_Y(B), \quad \forall A \in \mathcal{E}, B \in \tilde{\mathcal{E}}.$$

Autrement dit, la loi $P_{(X,Y)}$ sur $E \times \tilde{E}$ est la loi produit $P_X \otimes P_Y$.

Proposition 1.2. Pour que deux v.a. X et Y soient indépendantes, il faut et il suffit que

$$\mathbb{E}(f(X)g(Y)) = \mathbb{E}(f(X)) \mathbb{E}(g(Y)),$$

pour toute fonction mesurable bornée $f : E \rightarrow \mathbb{R}$ et toute fonction mesurable bornée $g : \tilde{E} \rightarrow \mathbb{R}$.

Preuve. Si X et Y sont indépendantes, alors on a bien

$$\mathbb{E}(f(X)g(Y)) = \mathbb{E}(f(X)) \mathbb{E}(g(Y))$$

lorsque $f = \mathbf{1}_A$ et $g = \mathbf{1}_B$. Par linéarité, la formule reste valable pour f et g étagées. Le cas général en découle par approximation. \square

Corollaire 1.3. Si X et Y sont indépendantes, alors il en est de même pour $f(X)$ et $g(Y)$ pour toutes fonctions mesurables f et g .

Définition 1.4. On dit que n v.a. X_1, \dots, X_n sont indépendantes, si

$$\mathbb{P}(X_1 \in A_1, \dots, X_n \in A_n) = \mathbb{P}(X_1 \in A_1) \cdots \mathbb{P}(X_n \in A_n), \quad \forall A_1 \in \mathcal{E}_1, \dots, \forall A_n \in \mathcal{E}_n.$$

Ceci équivaut à :

$$\mathbb{E}(f_1(X_1) \cdots f_n(X_n)) = \mathbb{E}(f_1(X_1)) \cdots \mathbb{E}(f_n(X_n))$$

pour toutes les fonctions mesurables bornées $f_k : E_k \rightarrow \mathbb{R}$.

Warning. L'indépendance deux-à-deux n'implique pas l'indépendance de plusieurs variables. Par exemple, si X et Y sont indépendantes, ainsi que Y et Z , et X et Z , on n'a pas forcément que X , Y et Z soient indépendantes. \square

Remarque. Si X_1, \dots, X_n sont indépendantes, et si $1 \leq i_1 < i_2 < \dots < i_k \leq n$, alors les nouvelles v.a. (X_1, \dots, X_{i_1}) , $(X_{i_1+1}, \dots, X_{i_2})$, \dots , $(X_{i_{k-1}+1}, \dots, X_n)$ sont indépendantes. Par conséquent, si f_1, \dots, f_{i_k} sont des fonctions mesurables sur $E_1 \times \dots \times E_{i_1}$, \dots , $E_{i_{k-1}+1} \times \dots \times E_n$ respectivement, alors les v.a.

$$f_1(X_1, \dots, X_{i_1}), f_2(X_{i_1+1}, \dots, X_{i_2}), \dots, f_{i_k}(X_{i_{k-1}+1}, \dots, X_n)$$

sont indépendantes. \square

Remarque. Si X_1, \dots, X_n sont toutes des v.a. réelles **discrètes**, à valeurs dans E_1, \dots, E_n , respectivement, alors leur indépendance équivaut à :

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

pour tout $(x_1, \dots, x_n) \in E_1 \times \dots \times E_n$. \square

Exemple 1.5. Soient X_1, \dots, X_n des v.a. indépendantes telles que X_i suivent la loi de Poisson de paramètre $\theta_i > 0$ ($1 \leq i \leq n$). On s'intéresse à la loi de $Y = \sum_{i=1}^n X_i$.

Considérons la fonction génératrice de Y : pour $s \in [0, 1]$,

$$G_Y(s) = \mathbb{E}(s^Y) = \mathbb{E}\left(\prod_{i=1}^n s^{X_i}\right) = \prod_{i=1}^n G_{X_i}(s) = \prod_{i=1}^n e^{-\theta_i(1-s)} = e^{-\theta(1-s)},$$

avec $\theta := \sum_{i=1}^n \theta_i$. Donc Y suit la loi de Poisson de paramètre θ . \square

Théorème 1.6. Les v.a. réelles X_1, \dots, X_n sont indépendantes si et seulement si

$$(*) \quad F_{(X_1, \dots, X_n)}(x_1, \dots, x_n) = F_{X_1}(x_1) \cdots F_{X_n}(x_n), \quad \forall (x_1, \dots, x_n) \in \mathbb{R}^n.$$

Preuve. La partie “seulement si” est évidente. Il suffit de prendre $A_i =]-\infty, x_i]$ dans la Définition 1.4.

Pour vérifier la partie “si”, on suppose l’identité (*). Fixons $(x_2, \dots, x_n) \in \mathbb{R}^{n-1}$. Considérons deux applications $\mu_1 : \mathcal{B}(\mathbb{R}) \rightarrow [0, 1]$ et $\tilde{\mu}_1 : \mathcal{B}(\mathbb{R}) \rightarrow [0, 1]$ ($\mathcal{B}(\mathbb{R})$ étant la tribu borélienne de \mathbb{R}) définies par

$$\begin{aligned} \mu_1(A) &:= \mathbb{P}(X_1 \in A, X_2 \leq x_2, \dots, X_n \leq x_n), \\ \tilde{\mu}_1(A) &:= \mathbb{P}(X_1 \in A) \mathbb{P}(X_2 \leq x_2) \cdots \mathbb{P}(X_n \leq x_n), \quad A \in \mathcal{B}(\mathbb{R}). \end{aligned}$$

Il est clair que μ_1 et $\tilde{\mu}_1$ sont deux mesures (finies) sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. De plus, il résulte de (*) que $\mu_1(]-\infty, x]) = \tilde{\mu}_1(]-\infty, x])$ quel que soit $x \in \mathbb{R}$. Donc pour tous $a < b$ réels, $\mu_1(]a, b]) = \tilde{\mu}_1(]a, b])$. Une mesure sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ étant déterminée par les valeurs qu’elle attribue aux intervalles du type de $]a, b]$, on déduit que μ_1 et $\tilde{\mu}_1$ sont deux mesures identiques sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. En d’autres mots, pour tout $A \in \mathcal{B}(\mathbb{R})$ et tout $(x_2, \dots, x_n) \in \mathbb{R}^{n-1}$,

$$\mathbb{P}(X_1 \in A, X_2 \leq x_2, \dots, X_n \leq x_n) = \mathbb{P}(X_1 \in A) \mathbb{P}(X_2 \leq x_2) \cdots \mathbb{P}(X_n \leq x_n).$$

Fixons maintenant $A_1 \in \mathcal{B}(\mathbb{R})$ et $(x_3, \dots, x_n) \in \mathbb{R}^{n-2}$, et considérons

$$\begin{aligned} \mu_2(A) &:= \mathbb{P}(X_1 \in A_1, X_2 \in A, X_3 \leq x_3, \dots, X_n \leq x_n), \\ \tilde{\mu}_2(A) &:= \mathbb{P}(X_1 \in A_1) \mathbb{P}(X_2 \in A) \mathbb{P}(X_3 \leq x_3) \cdots \mathbb{P}(X_n \leq x_n), \quad A \in \mathcal{B}(\mathbb{R}). \end{aligned}$$

De nouveau, il s’agit de mesures sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ qui attribuent la même masse sur l’intervalle $]-\infty, x]$ (quel que soit $x \in \mathbb{R}$) et qui sont donc identiques sur $\mathcal{B}(\mathbb{R})$. En itérant la procédure, on obtient : pour $A_i \in \mathcal{B}(\mathbb{R})$ ($1 \leq i \leq n$),

$$\mathbb{P}(X_1 \in A_1, X_2 \in A_2, \dots, X_n \in A_n) = \mathbb{P}(X_1 \in A_1) \mathbb{P}(X_2 \in A_2) \cdots \mathbb{P}(X_n \in A_n).$$

D’où l’indépendance des variables aléatoires X_1, \dots, X_n . \square

Théorème 1.7. Soit (X_1, \dots, X_n) un vecteur aléatoire dont la loi est absolument continue par rapport à la mesure de Lebesgue. Alors X_1, \dots, X_n sont indépendantes si et seulement si

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

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

Preuve. La partie “seulement si” découle de (*) en dérivant. La partie “si” se démontre en intégrant dans les deux côtés de (**). \square

Exemple 1.8 (suite des Exemples 2.1, 4.1 et 5.3 du Chapitre 4). Soit (X, Y) un vecteur aléatoire de densité

$$f_{(X,Y)}(x, y) = \theta^2 e^{-\theta(x+y)} \mathbf{1}_{\mathbb{R}_+^* \times \mathbb{R}_+^*}(x, y).$$

On a vu dans l’Exemple 5.3 que

$$f_X(x) = \theta e^{-\theta x} \mathbf{1}_{\mathbb{R}_+^*}(x), \quad f_Y(y) = \theta e^{-\theta y} \mathbf{1}_{\mathbb{R}_+^*}(y).$$

Donc $f_{(X,Y)}(x, y) = f_X(x)f_Y(y)$. Par la Proposition 1.7, on conclut que X et Y sont indépendantes. \square

Exemple 1.9. Soient N_1 et N_2 deux variables indépendantes suivant la même loi gaussienne centrée réduite. On s’intéresse à la variable aléatoire

$$X = \frac{N_1}{N_2}.$$

Soit $u \in \mathbb{R}$. On a, par le théorème de Fubini,

$$F_X(u) = \mathbb{P}\left(\frac{N_1}{N_2} \leq u\right) = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} \mathbf{1}_{\{x/y \leq u\}} \frac{1}{2\pi} e^{-(x^2+y^2)/2} dx \right) dy.$$

Faisons un changement de variables $z = x/y$ dans l’intégrale entre parenthèses, et appliquons ensuite de nouveau le théorème de Fubini :

$$\begin{aligned} F_X(u) &= \int_{\mathbb{R}} \left(\int_{-\infty}^u \frac{|y|}{2\pi} e^{-(z^2+1)y^2/2} dz \right) dy \\ &= \int_{-\infty}^u \left(\int_{\mathbb{R}} \frac{|y|}{2\pi} e^{-(z^2+1)y^2/2} dy \right) dz \\ &= \int_{-\infty}^u \frac{1}{\pi} \frac{1}{1+z^2} dz. \end{aligned}$$

Par conséquent, $\frac{N_1}{N_2}$ suit la loi de Cauchy standard.

On remarque que cet exemple est en accord avec l’Exemple 4.4 du Chapitre 4 où l’on a démontré que l’inverse de la loi de Cauchy standard est encore la loi de Cauchy standard. \square

Exemple 1.10. Soient N_1 et N_2 deux variables indépendantes suivant la même loi gaussienne centrée réduite. On s'intéresse à la variable aléatoire

$$X = N_1^2 + N_2^2.$$

Il est clair que X est à valeurs dans \mathbb{R}_+^* . Soit $u > 0$. On a

$$F_X(u) = \mathbb{P}(N_1^2 + N_2^2 \leq u) = \iint_{\{x^2+y^2 \leq u\}} \frac{1}{2\pi} e^{-(x^2+y^2)/2} dx dy.$$

Avec un passage aux coordonnées polaires, on obtient

$$F_X(u) = \int_0^{\sqrt{u}} r e^{-r^2/2} dr = 1 - e^{-u/2}.$$

Autrement dit, $(N_1^2 + N_2^2)$ suit la loi exponentielle de paramètre $1/2$ (donc de moyenne 2). \square

Propriété 1.11. Si la densité du vecteur aléatoire (X_1, \dots, X_n) se factorise :

$$f_{(X_1, \dots, X_n)}(x_1, \dots, x_n) = g_1(x_1) \cdots g_n(x_n),$$

alors X_1, \dots, X_n sont indépendantes.

Preuve. D'après la Proposition 5.2 du chapitre précédent, X_1 admet la densité

$$f_{X_1}(x_1) = c_1 g_1(x_1), \quad x_1 \in \mathbb{R},$$

où la constante

$$c_1 = \int \cdots \int_{\mathbb{R}^{n-1}} g_2(x_2) \cdots g_n(x_n) dx_2 \cdots dx_n.$$

De même, pour tout $2 \leq j \leq n$, la densité de X_j est

$$f_{X_j}(x_j) = c_j g_j(x_j), \quad x_j \in \mathbb{R},$$

avec une certaine constante c_j . Donc

$$f_{X_1}(x_1) \cdots f_{X_n}(x_n) = c_1 \cdots c_n g_1(x_1) \cdots g_n(x_n).$$

Or,

$$\begin{aligned} 1 &= \int \cdots \int_{\mathbb{R}^n} f_{X_1}(x_1) \cdots f_{X_n}(x_n) dx_1 \cdots dx_n \\ &= c_1 \cdots c_n \int \cdots \int_{\mathbb{R}^n} g_1(x_1) \cdots g_n(x_n) dx_1 \cdots dx_n \\ &= c_1 \cdots c_n \int \cdots \int_{\mathbb{R}^n} f_{(X_1, \dots, X_n)}(x_1, \dots, x_n) dx_1 \cdots dx_n \\ &= c_1 \cdots c_n, \end{aligned}$$

ce qui implique que

$$f_{(X_1, \dots, X_n)}(x_1, \dots, x_n) = g_1(x_1) \cdots g_n(x_n) = f_{X_1}(x_1) \cdots f_{X_n}(x_n).$$

D'après la Proposition 1.7, les v.a. X_1, \dots, X_n sont indépendantes. \square

Exemple 1.12. Soient X et Y deux v.a. réelles indépendantes, de densités respectives

$$f_X(x) = \frac{\theta^a}{\Gamma(a)} e^{-\theta x} x^{a-1} \mathbf{1}_{\mathbb{R}_+^*}(x),$$

$$f_Y(y) = \frac{\theta^b}{\Gamma(b)} e^{-\theta y} y^{b-1} \mathbf{1}_{\mathbb{R}_+^*}(y),$$

où $a > 0$, $b > 0$ et $\theta > 0$ sont des constantes connues. (On dit que X et Y suivent les lois gamma de paramètres (a, θ) et (b, θ) respectivement). On s'intéresse à la loi de $X + Y$.

Faisons un changement de variables $U = X + Y$ et $V = X/(X + Y)$. On a

$$h(x, y) = \left(x + y, \frac{x}{x + y} \right) = (u, v),$$

qui est une bijection de $\mathbb{R}_+^* \times \mathbb{R}_+^*$ dans $\mathbb{R}_+^* \times]0, 1[$ avec $h^{-1}(u, v) = (uv, u(1 - v))$. Il s'agit d'un difféomorphisme, dont le jacobien est

$$(Dh)(x, y) = \det \begin{pmatrix} \frac{1}{y} & \frac{1}{x+y} \\ -\frac{1}{(x+y)^2} & -\frac{1}{(x+y)^2} \end{pmatrix} = -\frac{1}{x+y} = -\frac{1}{u}.$$

Soit $\varphi : \mathbb{R}^2 \rightarrow \mathbb{R}$ une fonction mesurable bornée. On a

$$\begin{aligned} \mathbb{E}(\varphi(U, V)) &= \mathbb{E}(\varphi(h(X, Y))) \\ &= \iint_{\mathbb{R}_+^* \times \mathbb{R}_+^*} \varphi(h(x, y)) \frac{\theta^{a+b}}{\Gamma(a)\Gamma(b)} e^{-\theta(x+y)} x^{a-1} y^{b-1} dx dy \\ &= \iint_{\mathbb{R}_+^* \times]0, 1[} \varphi(u, v) \frac{\theta^{a+b}}{\Gamma(a)\Gamma(b)} e^{-\theta u} (uv)^{a-1} (u(1-v))^{b-1} \frac{1}{|-1/u|} du dv \\ &= \iint_{\mathbb{R}^2} \varphi(u, v) \frac{\theta^{a+b}}{\Gamma(a)\Gamma(b)} e^{-\theta u} \mathbf{1}_{\mathbb{R}_+^*}(u) u^{a+b-1} \mathbf{1}_{]0, 1[}(v) v^{a-1} (1-v)^{b-1} du dv. \end{aligned}$$

Donc (U, V) a pour densité

$$f_{(U, V)}(u, v) = \frac{\theta^{a+b}}{\Gamma(a)\Gamma(b)} e^{-\theta u} \mathbf{1}_{\mathbb{R}_+^*}(u) u^{a+b-1} \mathbf{1}_{]0, 1[}(v) v^{a-1} (1-v)^{b-1}.$$

D'après la Proposition 1.11, U et V sont indépendantes. Comme $\frac{\theta^{a+b}}{\Gamma(a+b)}e^{-\theta u}u^{a+b-1}\mathbf{1}_{\mathbb{R}_+^*}(u)$ est la densité de la loi gamma de paramètre $(a+b, \theta)$, on conclut que

$$f_U(u) = \frac{\theta^{a+b}}{\Gamma(a+b)}e^{-\theta u}u^{a+b-1}\mathbf{1}_{\mathbb{R}_+^*}(u).$$

Donc la somme de deux variables aléatoires gamma indépendantes suit encore une loi gamma.

On peut également voir que

$$\begin{aligned} f_V(v) &= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}v^{a-1}(1-v)^{b-1}\mathbf{1}_{]0,1[}(v) \\ &= \frac{v^{a-1}(1-v)^{b-1}\mathbf{1}_{]0,1[}(v)}{B(a,b)}. \end{aligned}$$

Cette loi est connue sous le nom de bêta. □

Remarque. (i) Si les v.a. réelles intégrables X et Y sont indépendantes, alors

$$\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y);$$

il suffit d'étudier le cas où X et Y sont toutes positives (sinon, on considèrera séparément leur partie positive et partie négative), l'identité ci-dessus est alors une conséquence de la Proposition 1.2 appliquée aux v.a. $\min(X, n)$ et $\min(Y, n)$, et du théorème de convergence monotone en faisant $n \rightarrow \infty$.

Donc, si X et Y admettent des moments d'ordre 2, l'indépendance entre les v.a. intégrables X et Y implique que $\text{Cov}(X, Y) = 0$. On a ainsi

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y).$$

(Rappelons que $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$ si et seulement si $\text{Cov}(X, Y) = 0$; on dira que X et Y sont **non-corrélées**).

Attention : **la réciproque est fautive!** Si $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$ (ceci équivaut à : $\text{Cov}(X, Y) = 0$), on n'aura pas forcément l'indépendance entre X et Y . Voir l'Exemple 1.13 ci-dessous pour un exemple simple.

(ii) Plus généralement, si les v.a. réelles X_1, \dots, X_n sont indépendantes, alors

$$\mathbb{E}(X_1 \cdots X_n) = \mathbb{E}(X_1) \cdots \mathbb{E}(X_n),$$

et $\mathbb{E}(X_i X_j) = \mathbb{E}(X_i)\mathbb{E}(X_j)$ si $i \neq j$. Donc, si X et Y admettent des moments d'ordre 2, alors $\text{Cov}(X_i, X_j) = 0$, $i \neq j$. Par conséquent,

$$\text{Var}(X_1 + \cdots + X_n) = \text{Var}(X_1) + \cdots + \text{Var}(X_n). \quad \square$$

Récapitulation. Pour un couple de v.a. réelles admettant des moments d'ordre 2,

$$\begin{aligned} X \text{ et } Y \text{ indépendantes} &\implies \text{Cov}(X, Y) = 0 \iff \mathbb{E}(XY) = \mathbb{E}(X) \mathbb{E}(Y) \\ &\iff \text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y). \end{aligned}$$

Pour un vecteur aléatoire dont toute composante admet un moment d'ordre 2,

$$\begin{aligned} X_1, \dots, X_n \text{ indépendantes} &\implies X_1, \dots, X_n \text{ deux-à-deux indépendantes} \\ &\implies \text{Cov}(X_i, X_j) = 0 \quad \forall i \neq j \\ &\implies \text{Var}(X_1 + \dots + X_n) = \text{Var}(X_1) + \dots + \text{Var}(X_n). \quad \square \end{aligned}$$

Exemple 1.13 (exemple de v.a. réelles non-corrélées qui ne sont pas indépendantes). Soit X une v.a. réelle discrète telle que

$$\mathbb{P}(X = 1) = \mathbb{P}(X = 2) = \mathbb{P}(X = -1) = \mathbb{P}(X = -2) = \frac{1}{4}.$$

On pose $Y = X^2$. Alors $\mathbb{E}(X) = 0$ et $\mathbb{E}(XY) = \mathbb{E}(X^3) = 0$. On a donc

$$\text{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X) \mathbb{E}(Y) = 0,$$

c'est-à-dire que X et Y sont non-corrélées.

Il est clair que X et Y ne sont pas indépendantes, car

$$\mathbb{P}(X = 1, Y = 4) = 0 \neq \frac{1}{4} \times \frac{1}{2} = \mathbb{P}(X = 1) \mathbb{P}(Y = 4). \quad \square$$

2. Événements indépendants

Proposition 2.1. Soient A_1, \dots, A_n des événements sur un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$. Ils sont (mutuellement) indépendants si et seulement si les variables aléatoires $\mathbf{1}_{A_1}, \dots, \mathbf{1}_{A_n}$ sont indépendantes.

Preuve. Partie “seulement si”. Supposons que A_1, \dots, A_n sont des événements indépendants. Par définition, pour tout $k \leq n$ et tout k -uplet $1 \leq i_1 < i_2 < \dots < i_k \leq n$,

$$\mathbb{P}(A_{i_1} \cap \dots \cap A_{i_k}) = \mathbb{P}(A_{i_1}) \dots \mathbb{P}(A_{i_k}). \quad (*)$$

Soient $(x_1, \dots, x_n) \in \mathbb{R}^n$. D'après le Théorème 1.6, il suffit de vérifier que

$$(**) \quad \mathbb{P}(\mathbf{1}_{A_1} \geq x_1, \dots, \mathbf{1}_{A_n} \geq x_n) = \mathbb{P}(\mathbf{1}_{A_1} \geq x_1) \cdots \mathbb{P}(\mathbf{1}_{A_n} \geq x_n).$$

S'il existe un $x_j > 1$, alors les deux côtés de (**) sont nuls, donc l'identité est vraie. On suppose donc que $x_j \leq 1$, pour tout $1 \leq j \leq n$.

Soit $E := \{1 \leq j \leq n : x_j > 0\}$. Si $E = \emptyset$, c'est-à-dire $x_j \leq 0$ pour tous les j , alors les deux côtés de (**) valent 1, donc l'identité est vraie. Si en revanche $E \neq \emptyset$, soit $E = \{i_1, \dots, i_k\}$. Dans ce cas, $x_j \leq 0$ si $j \notin E$ donc $\mathbf{1}_{A_j} \geq x_j$, tandis que si $j \in E$, alors $x_j \in]0, 1]$, et donc $\mathbf{1}_{A_j}(\omega) \geq x_j$ équivaut à $\omega \in A_j$. Par conséquent, $\{\mathbf{1}_{A_1} \leq x_1, \dots, \mathbf{1}_{A_n} \leq x_n\} = A_{i_1} \cap \dots \cap A_{i_k}$, et $\mathbb{P}(\mathbf{1}_{A_1} \geq x_1) \cdots \mathbb{P}(\mathbf{1}_{A_n} \geq x_n) = \prod_{j \in E} \mathbb{P}(A_j)$. Par (*), on a $\mathbb{P}(\mathbf{1}_{A_1} \geq x_1, \dots, \mathbf{1}_{A_n} \geq x_n) = \mathbb{P}(\mathbf{1}_{A_1} \geq x_1) \cdots \mathbb{P}(\mathbf{1}_{A_n} \geq x_n)$, c'est à dire (**).

Partie "si". Supposons que $\mathbf{1}_{A_1}, \dots, \mathbf{1}_{A_n}$ sont indépendantes. Donc pour tout $k \leq n$ et tout k -uplet $1 \leq i_1 < i_2 < \dots < i_k \leq n$, $\mathbf{1}_{A_{i_1}}, \dots, \mathbf{1}_{A_{i_k}}$ sont indépendantes. En particulier, $\mathbb{P}(\mathbf{1}_{A_{i_1}} \in B, \dots, \mathbf{1}_{A_{i_k}} \in B) = \mathbb{P}(\mathbf{1}_{A_{i_1}} \in B) \cdots \mathbb{P}(\mathbf{1}_{A_{i_k}} \in B)$ pour tout borélien $B \subset \mathbb{R}$. Prenons $B = \{1\}$, et on obtient $\mathbb{P}(A_{i_1} \cap \dots \cap A_{i_k}) = \mathbb{P}(A_{i_1}) \cdots \mathbb{P}(A_{i_k})$. \square

On jette un rapide coup d'oeil sur la probabilité conditionnelle.

Exemple 2.2. Soit Y une variable aléatoire suivant la loi de Poisson de paramètre $\theta > 0$: $\mathbb{P}(Y = n) = e^{-\theta} \frac{\theta^n}{n!}$ pour $n = 0, 1, \dots$ Soit X telle que pour chaque n ,

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

(Certains disent que "sachant $Y = n$, X suit la loi binomiale de paramètre (n, p) .")

On s'intéresse à la loi de X . Pour tout $k \geq 0$,

$$\begin{aligned} \mathbb{P}(X = k) &= \sum_{n=0}^{\infty} \mathbb{P}(X = k | Y = n) \mathbb{P}(Y = n) \\ &= \sum_{n=k}^{\infty} \binom{n}{k} p^k (1-p)^{n-k} e^{-\theta} \frac{\theta^n}{n!} \\ &= \frac{e^{-\theta} (\theta p)^k}{k!} \sum_{n=k}^{\infty} \frac{[\theta(1-p)]^{n-k}}{(n-k)!} \\ &= \frac{e^{-\theta p} (\theta p)^k}{k!}. \end{aligned}$$

Autrement dit, X suit la loi de Poisson de paramètre (θp) . \square

Chapitre 6. Fonctions caractéristiques

1. Définition et exemples

Dans tout ce chapitre, sauf mention contraire, X désigne une v.a. à valeurs dans \mathbb{R}^N (donc un vecteur aléatoire de dimension N) et P_X sa loi. Notons que P_X est une mesure de probabilité sur \mathbb{R}^N .

Définition 1.1. L'application $\varphi_X : \mathbb{R}^N \rightarrow \mathbb{C}$ donnée par

$$\varphi_X(t) = \mathbb{E} \left(e^{i\langle t, X \rangle} \right) = \int_{\mathbb{R}^N} e^{i\langle t, x \rangle} P_X(dx)$$

s'appelle la fonction caractéristique de X .

Autrement dit, la fonction caractéristique d'une variable aléatoire est la transformée de Fourier de sa loi, au sens où $\varphi_X(t) = \widehat{P}_X(-\frac{t}{2\pi})$, avec

$$\widehat{P}_X(t) = \int_{\mathbb{R}^N} e^{-2i\pi\langle t, x \rangle} P_X(dx),$$

qui la transformée de Fourier de P_X .

Il est clair que

- on a toujours $\varphi_X(0) = 1$;
- la fonction caractéristique prend ses valeurs dans le disque unité, c'est-à-dire $|\varphi_X(t)| \leq 1$ pour tout $t \in \mathbb{R}^N$. En effet, $|\mathbb{E}(e^{i\langle t, X \rangle})| \leq \mathbb{E}(|e^{i\langle t, X \rangle}|) = 1$;
- $\varphi_{\lambda X + a}(t) = e^{i\langle a, t \rangle} \varphi_X(\lambda t)$, pour tout $\lambda \in \mathbb{R}$ et tout $a \in \mathbb{R}^N$;
- si P_X admet la densité f_X (par rapport à la mesure de Lebesgue), alors $\varphi_X(t) = \int_{\mathbb{R}^N} e^{i\langle t, x \rangle} f_X(x) dx$;

- φ_X est une fonction semi-positive dans le sens que pour tous $n \geq 1$, $t_i \in \mathbb{R}^N$, $z_i \in \mathbb{C}$ ($1 \leq i \leq n$), on a $\sum_{k=1}^n \sum_{\ell=1}^n \varphi_X(t_k - t_\ell) z_k \bar{z}_\ell \geq 0$. En effet,

$$\begin{aligned} \sum_{k=1}^n \sum_{\ell=1}^n \varphi_X(t_k - t_\ell) z_k \bar{z}_\ell &= \mathbb{E} \left(\sum_{k=1}^n \sum_{\ell=1}^n e^{i\langle t_k, X \rangle} z_k e^{-i\langle t_\ell, X \rangle} \bar{z}_\ell \right) \\ &= \mathbb{E} \left(\left| \sum_{k=1}^n e^{i\langle t_k, X \rangle} z_k \right|^2 \right) \geq 0. \end{aligned}$$

- Théorème de Bochner (admis) : toute fonction φ semi-positive satisfaisant $\varphi(0) = 1$ est la fonction caractéristique associée à une loi de probabilité.

On regarde maintenant quelques exemples importants en dimension $N = 1$.

Exemple 1.2 (loi uniforme). Si $P_X = \frac{1}{b-a} \mathbf{1}_{]a,b[}(x) dx$ (loi uniforme sur $[a, b]$), alors

$$\varphi_X(t) = \frac{e^{ibt} - e^{iat}}{i(b-a)t}. \quad \square$$

Exemple 1.3 (loi exponentielle). Si P_X est la loi exponentielle de paramètre $\theta > 0$, $P_X = \theta e^{-\theta x} \mathbf{1}_{\mathbb{R}_+}(x) dx$, alors

$$\varphi_X(t) = \int_0^\infty \theta e^{-\theta x} e^{itx} dx = \frac{\theta}{\theta - it}. \quad \square$$

2. Propriétés des fonctions caractéristiques

Proposition 2.1. La fonction caractéristique est continue sur \mathbb{R}^N .

Preuve. C'est une conséquence immédiate de la continuité de l'application $t \rightarrow e^{i\langle t, x \rangle}$ pour chaque $x \in \mathbb{R}^N$, et du théorème de convergence dominée.

Plus concrètement, fixons $t \in \mathbb{R}^N$, et soit $(t_n)_{n \geq 1}$ une suite arbitraire tendant vers t . On a, pour tout $x \in \mathbb{R}^N$, $e^{i\langle t_n, x \rangle} \rightarrow e^{i\langle t, x \rangle}$ ($n \rightarrow \infty$). D'autre part, pour tout n et tout x , $|e^{i\langle t_n, x \rangle}| \leq 1$ qui est intégrable par rapport à la mesure P_X . Par le théorème de convergence dominée,

$$\varphi_X(t_n) = \int_{\mathbb{R}^N} e^{i\langle t_n, x \rangle} P_X(dx) \rightarrow \int_{\mathbb{R}^N} e^{i\langle t, x \rangle} P_X(dx) = \varphi_X(t), \quad n \rightarrow \infty.$$

La suite (t_n) étant quelconque, on en déduit que φ_X est continue en t . \square

Proposition 2.2. Si P_X est absolument continue par rapport à la mesure de Lebesgue, alors

$$\lim_{|t| \rightarrow \infty} \varphi_X(t) = 0.$$

Remarque. Ceci est un cas particulier d'un résultat général (Théorème de Riemann–Lebesgue) sur la transformée de Fourier : Si g est une fonction mesurable intégrable sur \mathbb{R}^N , alors $\widehat{g}(t) \rightarrow 0$ quand $|t| \rightarrow \infty$.

Attention : Ici la condition que la loi soit absolument continue est importante. Sinon, on ne peut pas espérer en général que $\lim_{|t| \rightarrow \infty} \varphi_X(t) = 0$. Par exemple, si $X = c$ (constante; nous sommes en dimension $N = 1$), on a $\varphi_X(t) = e^{ict}$ qui ne tend pas vers 0. \square

Théorème 2.3. Deux variables aléatoires X et Y ont même loi si et seulement si $\varphi_X = \varphi_Y$. Autrement dit, comme son nom l'indique, la fonction caractéristique φ_X caractérise la loi de X .

Preuve. Si $P_X = P_Y$, on a en particulier pour tout $t \in \mathbb{R}^N$,

$$\varphi_X(t) = \int_{\mathbb{R}^N} e^{i\langle t, x \rangle} P_X(dx) = \int_{\mathbb{R}^N} e^{i\langle t, x \rangle} P_Y(dx) = \varphi_Y(t).$$

Pour la réciproque, on va supposer pour simplifier l'écriture que la dimension est $N = 1$. Le cas des dimensions supérieures est similaire, mais avec des notations plus lourdes. On remarque d'abord que, pour $k \geq 1$ et $\theta \in \mathbb{R}$,

$$\int_{-\infty}^{\infty} e^{it\theta} e^{-|t|/k} dt = \frac{1}{\frac{1}{k} - i\theta} + \frac{1}{\frac{1}{k} + i\theta} = \frac{2k}{1 + \theta^2 k^2}.$$

On a donc

$$\int_{\mathbb{R}} \frac{2k}{1 + (y - x)^2 k^2} P_X(dy) = \int_{\mathbb{R}} P_X(dy) \int_{-\infty}^{\infty} dt e^{it(y-x)} e^{-|t|/k};$$

puis en appliquant le théorème de Fubini

$$\int_{\mathbb{R}} \frac{2k}{1 + (y - x)^2 k^2} P_X(dy) = \int_{-\infty}^{\infty} e^{-itx} e^{-|t|/k} \varphi_X(t) dt.$$

Considérons ensuite une fonction h continue, positive, bornée et intégrable. En appliquant le théorème de Fubini et un changement de variables, on tire

$$\int_{-\infty}^{\infty} h(x) \left(\int_{\mathbb{R}} \frac{2k}{1 + (y-x)^2 k^2} P_X(dy) \right) dx = \int_{\mathbb{R}} \int_{-\infty}^{\infty} \left(h \left(y - \frac{t}{k} \right) \frac{2}{1+t^2} \right) dt P_X(dy).$$

Quand on fait tendre k vers ∞ , le terme entre parenthèses dans le membre de droite converge vers $h(y) \frac{2}{1+t^2}$ en restant dominé par $\|h\|_{\infty} \frac{2}{1+t^2}$ (qui est intégrable par rapport à la mesure produit $dt \otimes P_X$). En appliquant le théorème de convergence dominée, on voit donc que

$$\int_{\mathbb{R}} h(y) P_X(dy) = \frac{1}{2\pi} \lim_{k \rightarrow \infty} \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} e^{-itx} e^{-|t|/k} \varphi_X(t) dt \right) h(x) dx.$$

Donc si $\varphi_X = \varphi_Y$, on a $\int_{\mathbb{R}} h(y) P_X(dy) = \int_{\mathbb{R}} h(y) P_Y(dy)$ pour toute fonction h continue, positive, bornée et intégrable. Soit $a < b$. Il existe une suite de fonctions continues, positives, bornées et intégrables (h_k) telle que h_k converge simplement vers $\mathbf{1}_{]a,b[}$, et que $|h_k| \leq 1$. Par le théorème de convergence dominée, $\int_{\mathbb{R}} h_k(y) P_X(dy) \rightarrow \int_{\mathbb{R}} \mathbf{1}_{]a,b[}(y) P_X(dy) = \mathbb{P}(a < X < b)$. Donc $\mathbb{P}(a < X < b) = \mathbb{P}(a < Y < b)$. Faisant b tendre vers ∞ , on a $1 - F_X(a) = 1 - F_Y(a)$. Comme la fonction de répartition détermine la loi d'une v.a., on en déduit que $P_X = P_Y$. \square

Le Théorème 2.3 a beaucoup d'applications, car, il est pour certaines v.a. beaucoup plus facile de calculer la fonction caractéristique que de calculer par exemple la fonction de répartition.

Une conséquence immédiate du Théorème 2.3 est la suivante, parfois appelée **formule d'inversion**, que l'on a déjà vue sous forme de transformée de Fourier inverse.

Théorème 2.4. Si φ_X est intégrable par rapport à la mesure de Lebesgue sur \mathbb{R}^N , alors X admet la densité

$$f_X(x) = \frac{1}{(2\pi)^N} \int_{\mathbb{R}^N} e^{-i\langle t, x \rangle} \varphi_X(t) dt, \quad x \in \mathbb{R}^N.$$

En plus f_X est bornée sur \mathbb{R}^N .

Exemple 2.5. Soit Y une variable aléatoire réelle de densité $f_Y(x) := \frac{1}{2}e^{-|x|}$. On a

$$\begin{aligned} \varphi_Y(t) &= \frac{1}{2} \int_{-\infty}^{\infty} e^{itx-|x|} dx \\ &= \frac{1}{2} \int_0^{\infty} e^{-(1-it)x} dx + \frac{1}{2} \int_{-\infty}^0 e^{(1+it)x} dx \\ &= \frac{1}{2} \left(\frac{1}{1-it} + \frac{1}{1+it} \right) \\ &= \frac{1}{1+t^2}. \end{aligned}$$

Il s'agit d'une fonction intégrable sur \mathbb{R} . Par le Théorème 2.4,

$$\frac{1}{2\pi} \int_{\mathbb{R}} \frac{e^{-itx}}{1+t^2} dt = f_Y(x) = \frac{1}{2}e^{-|x|}.$$

ce qui équivaut à :

$$\int_{\mathbb{R}} \frac{1}{\pi} \frac{e^{itx}}{1+x^2} dx = e^{-|t|}.$$

Autrement dit, si X suit la loi de Cauchy standard, alors $\varphi_X(t) = e^{-|t|}$.

On peut également démontrer ce résultat en utilisant le théorème des résidus en analyse complexe. \square

Théorème 2.6. Les v.a. réelles X_1, \dots, X_n sont indépendantes si et seulement si

$$\varphi_X(t) = \prod_{k=1}^n \varphi_{X_k}(t_k), \quad \forall t = (t_1, \dots, t_n) \in \mathbb{R}^n,$$

où $X := (X_1, \dots, X_n)$.

Preuve. Soit $P_{X_1} \otimes \dots \otimes P_{X_n}$ la mesure produit sur espace $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$. On a, par définition, $\int_{\mathbb{R}^n} e^{i\langle t, x \rangle} (P_{X_1} \otimes \dots \otimes P_{X_n})(dx) = \prod_{k=1}^n \varphi_{X_k}(t_k)$. Donc $\varphi_X(t) = \prod_{k=1}^n \varphi_{X_k}(t_k)$ si et seulement si les fonctions caractéristiques associées à $P_{X_1} \otimes \dots \otimes P_{X_n}$ et à P_X sont identiques, ce qui équivaut à $P_{X_1} \otimes \dots \otimes P_{X_n} = P_{(X_1, \dots, X_n)}$ (car la fonction caractéristique caractérise une loi de probabilité), c'est à dire, X_1, \dots, X_n sont indépendantes. \square

Remarque. Si X_1, \dots, X_n sont des vecteurs aléatoires (de dimension $N \geq 1$) indépendantes, alors

$$\varphi_{X_1 + \dots + X_n}(t) = \varphi_{X_1}(t) \times \dots \times \varphi_{X_n}(t), \quad t \in \mathbb{R}^N.$$

En particulier, si X_1, \dots, X_n sont des v.a. réelles indépendantes, alors

$$(*) \quad \varphi_{X_1+\dots+X_n}(t) = \varphi_{X_1}(t) \times \dots \times \varphi_{X_n}(t), \quad t \in \mathbb{R}.$$

Attention à ne pas confondre ceci avec le Théorème 2.6 ! En fait, la condition (*) n'est pas suffisante pour assurer l'indépendance entre X_1, \dots, X_n : elle dit simplement que $\varphi_{(X_1, \dots, X_n)}(t, \dots, t) = \varphi_{X_1}(t) \dots \varphi_{X_n}(t)$ pour tout $t \in \mathbb{R}$, tandis que la condition dans le Théorème 2.6 dit que $\varphi_{(X_1, \dots, X_n)}(t_1, \dots, t_n) = \varphi_{X_1}(t_1) \dots \varphi_{X_n}(t_n)$, pour tout $(t_1, \dots, t_n) \in \mathbb{R}^n$. Cette dernière est beaucoup plus forte que la condition (*). \square

Exemple 2.7. Soit (X, Y) un vecteur aléatoire de densité

$$f_{(X,Y)}(x, y) = 2 \mathbf{1}_D(x, y),$$

où

$$D := \{(x, y) : 0 \leq x \leq 1/2, x \leq y \leq x + 1/2\} \\ \cup \{(x, y) : 1/2 \leq x \leq 1, x \leq y \leq 1 \text{ ou } 0 \leq y \leq x - 1/2\}.$$

(Il convient de faire un dessin.) On peut facilement déterminer les lois (marginales) de X et de Y . En effet, pour tout $x \in [0, 1]$, $f_X(x) = \int_0^1 f_{(X,Y)}(x, y) dy = 1$, et bien sûr $f_X(x) = 0$ si $x \notin [0, 1]$. De même, $f_Y(y) = \mathbf{1}_{[0,1]}(y)$. Autrement dit, X et Y suivent la même loi uniforme sur $[0, 1]$: donc

$$\varphi_X(t) = \varphi_Y(t) = \frac{e^{it} - 1}{it}.$$

Il est clair que X et Y ne sont pas indépendantes car $f_{(X,Y)}(x, y) \neq f_X(x)f_Y(y)$ (ou plus simplement : parce que $f_{(X,Y)}$ ne se factorise pas).

On s'intéresse à la loi de $X + Y$. On a,

$$\varphi_{X+Y}(t) = 2 \iint_{(x,y) \in D} e^{it(x+y)} dx dy \\ = \iint_{(x,y) \in D} e^{it(x+y)} dx dy + \iint_{(y,x) \in D} e^{it(x+y)} dx dy.$$

Or, si l'on écrit $\tilde{D} := \{(x, y) : (y, x) \in D\}$, on voit que $D \cup \tilde{D} = [0, 1]^2$, et que $D \cap \tilde{D} = \{(x, y) \in [0, 1]^2 : y = x, \text{ ou } |y - x| = 1/2\}$. Ce dernier ensemble est de mesure de Lebesgue (sur \mathbb{R}^2) nulle. Donc

$$\varphi_{X+Y}(t) = \iint_{(x,y) \in [0,1]^2} e^{it(x+y)} dx dy = \left(\int_0^1 e^{itx} dx \right)^2 = \varphi_X(t)\varphi_Y(t).$$

On a donc un exemple de couple de variables aléatoires X et Y qui ne sont pas indépendantes, mais qui satisfait la condition que $\varphi_{X+Y}(t) = \varphi_X(t)\varphi_Y(t)$, $\forall t \in \mathbb{R}$. \square

Exemple 2.8. Si X et Y sont deux v.a. de Cauchy standard indépendantes, alors

$$\varphi_{X+Y}(t) = e^{-|t|} \times e^{-|t|} = e^{-2|t|}.$$

D'autre part,

$$\varphi_{2X}(t) = \varphi_X(2t) = e^{-2|t|}.$$

Donc la somme de deux v.a. de Cauchy indépendantes a même loi que deux fois une v.a. de Cauchy standard. \square

Soient maintenant μ et ν deux mesures finies sur \mathbb{R}^N . On rappelle que le produit de convolution de μ et ν , noté par $\mu * \nu$, est la mesure image de $\mu \otimes \nu$ par l'application de $\mathbb{R}^N \times \mathbb{R}^N$ dans \mathbb{R}^N définie par $(x, y) \mapsto x + y$. Autrement dit, $\mu * \nu$ est une mesure sur \mathbb{R}^N définie par

$$(\mu * \nu)(A) = \iint_{\mathbb{R}^N \times \mathbb{R}^N} \mathbf{1}_A(x + y) \mu(dx) \nu(dy), \quad A \in \mathcal{B}(\mathbb{R}^N).$$

Donc pour toute fonction borélienne bornée h ,

$$(\dagger) \quad \int_{\mathbb{R}^N} h(x) (\mu * \nu)(dx) = \iint_{\mathbb{R}^N \times \mathbb{R}^N} h(x + y) \mu(dx) \nu(dy).$$

En prenant $h(x) = e^{-2i\pi\langle t, x \rangle}$ dans (\dagger) et en appliquant le théorème de Fubini, on tire que la transformée de Fourier de la mesure finie $\mu * \nu$ est le produit des transformées de Fourier de μ et de ν .

Si on suppose maintenant que μ et ν sont des mesures de probabilités. La mesure produit $\mu \otimes \nu$ sur $\mathbb{R}^N \times \mathbb{R}^N$ est donc la loi de (X, Y) , où X et Y sont **indépendantes** et de loi respectives μ et ν . La formule (\dagger) montre que $\mu * \nu$ est la loi de la somme $X + Y$. On peut bien sûr itérer le procédé. On retiendra le résultat suivant :

Proposition 2.8. Soient X_1, \dots, X_n des vecteurs aléatoires indépendants à valeurs dans \mathbb{R}^N , de lois respectives P_{X_1}, \dots, P_{X_n} . La loi de $X_1 + \dots + X_n$ est le produit de convolution $P_{X_1} * \dots * P_{X_n}$.

3. Application au calcul des moments

La fonction caractéristique d'une v.a. réelle X permet de calculer très facilement les moments de X , pourvu qu'ils existent.

Proposition 3.1. Supposons que la v.a. réelle X admet un moment d'ordre $n \geq 1$. Alors φ_X est de classe \mathcal{C}^n et

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

En particulier,

$$\mathbb{E}(X^n) = \frac{\varphi_X^{(n)}(0)}{i^n}.$$

Preuve. Nous allons prouver la formule pour $n = 1$; le cas général en découle par récurrence. On remarque tout d'abord que pour tout $\theta \in \mathbb{R}$,

$$|\sin(\theta)| \leq |\theta|, \quad 1 - \cos(\theta) = 2 \sin^2(\theta/2) \leq |\theta|^2/2.$$

On a pour tout $t \in \mathbb{R}$,

$$\frac{\varphi_X(t + \varepsilon) - \varphi_X(t)}{\varepsilon} = \int_{\mathbb{R}} \left(\frac{e^{i(t+\varepsilon)x} - e^{itx}}{\varepsilon} \right) P_X(dx) = \int_{\mathbb{R}} \left(\frac{e^{i\varepsilon x} - 1}{\varepsilon} \right) e^{itx} P_X(dx).$$

L'intégrand dans le terme de droite est dominé par $2|x|$ et converge vers ixe^{itx} quand $\varepsilon \rightarrow 0$. On applique le théorème de convergence dominée pour obtenir la formule

$$\varphi_X'(t) = i \mathbb{E}(X e^{itX}).$$

L'application $t \mapsto xe^{itx}$ est continue pour chaque x , et est dominée par $|x|$ indépendamment de t . Par convergence dominée, on a

$$\lim_{t \rightarrow t_0} \int_{\mathbb{R}} x e^{itx} P_X(dx) = \int_{\mathbb{R}} x e^{it_0x} P_X(dx),$$

ce qui établit la continuité de φ_X' . □

Remarque. On se gardera de croire que si φ_X admet une dérivée d'ordre n à l'origine, alors $\mathbb{E}(X^n) = \varphi_X^{(n)}(0)$; il existe des v.a. dont la fonction caractéristique est continûment dérivable, et qui n'ont pas de moment d'ordre 1. Par exemple, supposons que

$$P_X(dx) = c \mathbf{1}_{\{|x|>2\}} \frac{dx}{|x|^2 \ln|x|},$$

où c est la constante de normalisation. D'une part, on sait que $\int_{-\infty}^{\infty} |x| P_X(dx) = \infty$ (intégrale de Bertrand). D'autre part, par symétrie, on a

$$1 - \varphi_X(\varepsilon) = c \int_{|x|>2} (1 - e^{i\varepsilon x}) \frac{dx}{|x|^2 \ln|x|} = 2c \int_2^{\infty} (1 - \cos(\varepsilon x)) \frac{dx}{x^2 \ln x}.$$

Montrons maintenant que $\varphi'_X(0) = 0$: après un changement de variables $y = \varepsilon x$, il s'agit de montrer que $\int_{2\varepsilon}^{\infty} \frac{1 - \cos(y)}{y^2 \ln(y/\varepsilon)} dy \rightarrow 0$ quand $\varepsilon \rightarrow 0$.

Par le théorème de convergence dominée, on voit que $\int_1^{\infty} \frac{1 - \cos(y)}{y^2 \ln(y/\varepsilon)} dy \rightarrow 0$. Pour l'intégrale sur $[2\varepsilon, 1]$, on remarque que $\int_{2\varepsilon}^1 \frac{1 - \cos(y)}{y^2 \ln(y/\varepsilon)} dy \leq K \int_{2\varepsilon}^1 \frac{1}{\ln(y/\varepsilon)} dy$ pour une certaine constante $K < \infty$, car la fonction $\frac{1 - \cos(y)}{y^2}$ est bornée. Par un changement de variables, $\int_{2\varepsilon}^1 \frac{1}{\ln(y/\varepsilon)} dy = \varepsilon \int_2^{1/\varepsilon} \frac{1}{\ln x} dx$. Comme $\int_2^3 \frac{1}{\ln x} dx$ est une constante finie, il reste seulement de vérifier que $\frac{1}{A} \int_3^A \frac{1}{\ln x} dx \rightarrow 0$ quand $A \rightarrow \infty$.

Or, par intégration par parties,

$$\int_3^A \frac{1}{\ln x} dx = \left[\frac{x}{\ln x} \right]_3^A + \int_3^A \frac{1}{(\ln x)^2} dx \leq \frac{A}{\ln A} - \frac{3}{\ln 3} + c_* \int_3^A \frac{1}{\ln x} dx,$$

où l'on a utilisé dans la dernière inégalité le fait que $\frac{1}{(\ln x)^2} \leq \frac{c_*}{\ln x}$ pour tout $x \in [3, \infty[$, avec la constante $c_* := \frac{1}{\ln 3} < 1$. Donc $\int_3^A \frac{1}{\ln x} dx \leq \frac{1}{1-c_*} \left(\frac{A}{\ln A} - \frac{3}{\ln 3} \right)$, ce qui implique que $\frac{1}{A} \int_3^A \frac{1}{\ln x} dx \rightarrow 0$, $A \rightarrow \infty$. \square

Exemple 3.2. (loi gaussienne). Si P_X est la loi gaussienne standard $\mathcal{N}(0, 1)$, alors

$$\varphi_X(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\frac{x^2}{2}\right) e^{itx} dx$$

est à valeurs réelles (par symétrie). Une dérivation sous le signe somme donne

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

A l'aide d'une intégration par parties (dériver $x \rightarrow e^{itx}$ et intégrer $x \rightarrow xe^{-x^2/2}$), on obtient

$$\varphi'_X(t) = -\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} t \exp\left(-\frac{x^2}{2}\right) e^{itx} dx = -t \varphi_X(t).$$

La résolution de cette équation différentielle donne $\ln(|\varphi_X(t)|) = -t^2/2 + c$. Comme $\varphi_X(0) = 1$, on conclut que

$$\varphi_X(t) = e^{-t^2/2}.$$

Plus généralement, on peut utiliser la représentation $Y = \sigma X + \mu$ pour Y de loi $\mathcal{N}(\mu, \sigma^2)$ en fonction d'une variable normale standard X afin d'obtenir la fonction caractéristique associée à la loi $\mathcal{N}(\mu, \sigma^2)$:

$$\varphi_Y(t) = \exp\left(it\mu - \frac{\sigma^2 t^2}{2}\right).$$

Si $P_Y = \delta_\mu$ (masse de Dirac en μ), on a $\varphi_Y(t) = e^{it\mu}$, ce qui correspond formellement à la formule précédente en prenant $\sigma = 0$. Donc la loi constante est souvent considérée comme un cas dégénéré de loi gaussienne. \square

Exemple 3.3. Si X a pour loi $\mathcal{N}(\mu, \sigma^2)$ et Y a pour loi $\mathcal{N}(\nu, \tau^2)$, avec X et Y indépendantes, alors

$$\begin{aligned}\varphi_{X+Y}(t) &= \varphi_X(t)\varphi_Y(t) = \exp\left(it\mu - \frac{\sigma^2 t^2}{2}\right) \times \exp\left(it\nu - \frac{\tau^2 t^2}{2}\right) \\ &= \exp\left(it(\mu + \nu) - \frac{(\sigma^2 + \tau^2)t^2}{2}\right).\end{aligned}$$

Donc $X + Y$ a pour loi $\mathcal{N}(\mu + \nu, \sigma^2 + \tau^2)$. \square

Exemple 3.4. Si X a pour loi $\mathcal{N}(0, 1)$, indépendante de la v.a. ε qui a pour loi $\mathbb{P}(\varepsilon = 1) = \mathbb{P}(\varepsilon = -1) = 1/2$. Soit $Y = \varepsilon X$. On a

$$\begin{aligned}\varphi_Y(t) &= \mathbb{E}(e^{it\varepsilon X}) = \int_{\mathbb{R}} \int_{\mathbb{R}} e^{itux} P_X(dx) P_\varepsilon(du) = \int_{\mathbb{R}} \varphi_X(tu) P_\varepsilon(du) \\ &= \int_{\mathbb{R}} \exp\left(-\frac{t^2 u^2}{2}\right) P_\varepsilon(du) = \exp\left(-\frac{t^2}{2}\right).\end{aligned}$$

Donc Y suit la loi $\mathcal{N}(0, 1)$.

On a $\mathbb{E}(X) = 0 = \mathbb{E}(Y)$. De plus, $\mathbb{E}(XY) = \mathbb{E}(\varepsilon X^2) = \mathbb{E}(\varepsilon)\mathbb{E}(X^2) = 0$. Donc $\text{Cov}(X, Y) = 0$. D'autre part, comme $|X| = |Y|$, les v.a. X et Y ne sont évidemment pas indépendantes. On a donc un exemple de deux v.a. réelles gaussiennes non-corrélées, qui ne sont pas indépendantes. \square

Exemple 3.5. Plus généralement, si X est une v.a. réelle dont la loi est symétrique sur \mathbb{R} (c'est-à-dire que $P_X = P_{-X}$), et si ε est une v.a. réelle indépendante de X telle que $\varepsilon^2 = 1$, alors $P_{\varepsilon X} = P_X$, ce qui est intuitivement clair.

En effet, pour tout $t \in \mathbb{R}$,

$$\begin{aligned}\varphi_{\varepsilon X}(t) &= \mathbb{E}(e^{it\varepsilon X}) = \int \int e^{ituX} P_\varepsilon(du) P_X(dx) \\ &= \int \mathbb{E}(e^{it u X}) P_\varepsilon(du) \\ &= \mathbb{E}(e^{itX}) \mathbb{P}(\varepsilon = 1) + \mathbb{E}(e^{-itX}) \mathbb{P}(\varepsilon = -1) \\ &= \mathbb{E}(e^{itX}) = \varphi_X(t).\end{aligned}$$

Notons que $P_X = P_{-X}$ si et seulement si φ_X est une fonction paire sur \mathbb{R} , ce qui est le cas si par exemple X admet une fonction de densité qui est symétrique sur \mathbb{R} . \square

Exemple 3.6. Si X et Y sont des v.a. réelles indépendantes avec $P_X = P_Y = \mathcal{N}(\mu, \sigma^2)$, alors $X + Y$ et $X - Y$ sont indépendantes.

En effet, pour $(t_1, t_2) \in \mathbb{R}^2$,

$$\begin{aligned}\varphi_{(X+Y, X-Y)}(t_1, t_2) &= \mathbb{E}\left(e^{i(t_1+t_2)X+i(t_1-t_2)Y}\right) \\ &= \varphi_X(t_1+t_2)\varphi_Y(t_1-t_2) \\ &= \exp\left(i\mu(t_1+t_2) - \frac{\sigma^2(t_1+t_2)^2}{2} + i\mu(t_1-t_2) - \frac{\sigma^2(t_1-t_2)^2}{2}\right) \\ &= \exp(i(2\mu)t_1 - \sigma^2(t_1^2 + t_2^2)).\end{aligned}$$

D'autre part, d'après l'Exemple 3.3, $P_{X+Y} = \mathcal{N}(2\mu, 2\sigma^2)$ et $P_{X-Y} = \mathcal{N}(0, 2\sigma^2)$. Donc $\varphi_{X+Y}(t_1) = \exp(i(2\mu)t_1 - \sigma^2 t_1^2)$, et $\varphi_{X-Y}(t_2) = \exp(-\sigma^2 t_2^2)$. Par conséquent,

$$\varphi_{(X+Y, X-Y)}(t_1, t_2) = \varphi_{X+Y}(t_1)\varphi_{X-Y}(t_2), \quad \forall (t_1, t_2) \in \mathbb{R}^2.$$

La Proposition 2.5 nous confirme alors que $X + Y$ et $X - Y$ sont indépendantes. \square

Chapitre 7. Suites et séries de variables aléatoires

On considère une suite X_1, \dots, X_n, \dots de v.a. réelles et on s'intéresse au comportement asymptotique de cette suite aléatoire. L'étude repose en partie sur un résultat élémentaire très utile.

1. Lemme de Borel–Cantelli

On considère une suite d'événements $\{A_n, n \geq 1\}$ dans (Ω, \mathcal{F}) . La suite des événements $\bigcup_{k \geq n} A_k, n \geq 1$ est décroissante, son intersection

$$A = \limsup_{n \rightarrow \infty} A_n = \bigcap_{n \geq 1} \bigcup_{k \geq n} A_k$$

est encore un événement dans \mathcal{F} . On remarquera que A représente l'ensemble des aléas ω qui appartiennent à une infinité d'événements A_n , autrement dit $\mathbf{1}_A = \limsup_{n \rightarrow \infty} \mathbf{1}_{A_n}$; ce qui justifie la notation. On écrira souvent $\{A_n \text{ infiniment souvent}\}$ à la place de l'ensemble $\limsup_{n \rightarrow \infty} A_n$.

On peut également définir $\liminf_{n \rightarrow \infty} A_n = (\limsup_{n \rightarrow \infty} A_n^c)^c$. Autrement dit,

$$\liminf_{n \rightarrow \infty} A_n = \bigcup_{n \geq 1} \bigcap_{k \geq n} A_k.$$

Il est clair que $\liminf_{n \rightarrow \infty} A_n = \{\omega \in \Omega : \omega \in A_n \text{ sauf un nombre fini d'indice } n\}$, ou encore $\mathbf{1}_{\liminf_{n \rightarrow \infty} A_n} = \liminf_{n \rightarrow \infty} \mathbf{1}_{A_n}$.

Exemple 1.1. Soient $\Omega = \mathbb{R}$, $\mathcal{F} = \mathcal{B}(\mathbb{R})$ (tribu borélienne), $A_1 = \emptyset$, $A_2 = [-3, 3]$ et

$$A_{2k-1} = [-1, 2[, \quad A_{2k} = [-2, 1], \quad k \geq 1.$$

On a

$$\limsup_{n \rightarrow \infty} A_n = [-2, 2[, \quad \liminf_{n \rightarrow \infty} A_n = [-1, 1].$$

Théorème 1.2 (lemme de Borel–Cantelli). Soit $\{A_n\}_{n \geq 1}$ une suite d'événements.

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

$$\mathbb{P}(A_n \text{ infiniment souvent}) = 0.$$

(ii) Si $\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty$, et si les événements $\{A_n\}_{n \geq 1}$ sont indépendants (c'est à dire, pour tout n , A_1, \dots, A_n sont indépendants), alors

$$\mathbb{P}(A_n \text{ infiniment souvent}) = 1.$$

Remarque. Dans les applications, on utilise souvent la version suivante de la conclusion dans la partie (i) : il existe un événement B avec $\mathbb{P}(B) = 1$, tel que pour tout $\omega \in B$, on puisse trouver $n_0 = n_0(\omega) < \infty$ tel que $\omega \in A_n^c$ lorsque $n \geq n_0$. \square

Preuve du lemme de Borel–Cantelli. (i) Fixons $\varepsilon > 0$. Il existe un entier N tel que $\sum_{n=N}^{\infty} \mathbb{P}(A_n) < \varepsilon$, et en conséquence $\mathbb{P}\left(\bigcup_{n \geq N} A_n\right) < \varepsilon$. De la définition de l'événement $\limsup_{n \rightarrow \infty} A_n$, on tire a fortiori $\mathbb{P}(\limsup_{n \rightarrow \infty} A_n) < \varepsilon$.

(ii) Comme les A_n sont indépendants, il en est de même pour leurs complémentaires, et on a

$$\mathbb{P}\left(\bigcup_{j=k}^m A_j\right) = 1 - \mathbb{P}\left(\bigcap_{j=k}^m A_j^c\right) = 1 - \prod_{j=k}^m \mathbb{P}(A_j^c) \geq 1 - \exp\left[-\sum_{j=k}^m \mathbb{P}(A_j)\right].$$

On fait tendre m vers ∞ , et on tire

$$\mathbb{P}\left(\bigcup_{j=k}^{\infty} A_j\right) \geq 1 - \exp\left[-\sum_{j=k}^{\infty} \mathbb{P}(A_j)\right] = 1.$$

Comme les événements $\bigcup_{j=k}^{\infty} A_j$ décroissent vers $\limsup_{n \rightarrow \infty} A_n$ quand $k \rightarrow \infty$, on a bien $\mathbb{P}(\limsup_{n \rightarrow \infty} A_n) \geq 1$ (et donc = 1). \square

2. Divers modes de convergence

Définition 2.1 (convergence p.s.). On dit que la suite de v.a. $\{X_n\}_{n \geq 1}$ converge presque sûrement vers une v.a. X s'il existe un événement A avec $\mathbb{P}(A) = 1$, tel que

$$\lim_{n \rightarrow \infty} X_n(\omega) = X(\omega) \quad \text{pour tout } \omega \in A.$$

Autrement dit,

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

On dit souvent que X_n converge vers X avec probabilité 1.

Exemple 2.2. Soit (X_n) une suite de v.a. réelles indépendantes suivant la même loi gaussienne $\mathcal{N}(0, 1)$. Soit $S_n = X_1 + \dots + X_n$. On sait que S_n suit la loi $\mathcal{N}(0, n)$. Donc par l'inégalité de Markov, pour tout $\varepsilon > 0$,

$$\mathbb{P}(|S_n| > n\varepsilon) = \mathbb{P}(|S_n|^3 > \varepsilon^3 n^3) \leq \frac{\mathbb{E}(|S_n|^3)}{\varepsilon^3 n^3} = \frac{\mathbb{E}(|X_1|^3)}{\varepsilon^3 n^{3/2}}.$$

On a $\sum_n \mathbb{P}(|S_n| > n\varepsilon) < \infty$. D'après la première partie du lemme de Borel–Cantelli,

$$\mathbb{P}(|S_n| > n\varepsilon, \text{ pour une infinité de } n) = 0.$$

De façon équivalente, il existe $A \in \mathcal{F}$ avec $\mathbb{P}(A) = 1$ tel que

$$\forall \omega \in A, \exists n_0 = n_0(\omega, \varepsilon) < \infty \text{ tel que } |S_n(\omega)| \leq n\varepsilon, \forall n \geq n_0.$$

A fortiori, pour tout $\varepsilon > 0$,

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

La v.a. $\limsup_{n \rightarrow \infty} \frac{|S_n|}{n}$ vaut donc 0 avec probabilité 1, autrement dit, $\frac{S_n}{n} \rightarrow 0$ p.s.

On a utilisé le fait que si X est une v.a. réelle telle que $\mathbb{P}(0 \leq X \leq \varepsilon) = 1$ quel que soit $\varepsilon > 0$, alors $\mathbb{P}(X = 0) = 1$. Ceci se démontre par exemple en remarquant que $\{X = 0\} = \bigcap_{k \geq 1} \{0 \leq X \leq \frac{1}{k}\}$. \square

Exemple 2.3. Soient X, X_1, X_2, \dots des v.a. réelles, et soit $\{\varepsilon_n\}$ une suite de réels strictement positifs tendant vers 0. Si $\sum \mathbb{P}(|X_n - X| \geq \varepsilon_n)$ est convergente, alors $X_n \rightarrow X$ p.s.

En effet, d'après la première partie du lemme de Borel–Cantelli, il existe $A \in \mathcal{F}$ avec $\mathbb{P}(A) = 1$ tel que

$$\forall \omega \in A, \exists n_0 = n_0(\omega) < +\infty \text{ tel que } |X_n(\omega) - X(\omega)| < \varepsilon_n, \forall n \geq n_0.$$

A fortiori, si $\omega \in A$, alors $X_n(\omega) \rightarrow X(\omega)$ ($n \rightarrow +\infty$). Donc X_n converge presque sûrement vers X . \square

Exemple 2.4. Soient X, X_1, X_2, \dots des v.a. réelles telles que $\sum \mathbb{P}(|X_n - X| \geq \varepsilon)$ soit convergente pour tout $\varepsilon > 0$, alors $X_n \rightarrow X$ p.s.

Ceci est un renforcement de l'exemple précédent, car si $\varepsilon_n \rightarrow 0$ ($n \rightarrow +\infty$), alors $\mathbb{P}(|X_n - X| \geq \varepsilon_n) \geq \mathbb{P}(|X_n - X| \geq \varepsilon)$ lorsque n est suffisamment grand.

Fixons $\varepsilon > 0$. Par hypothèse et le lemme de Borel–Cantelli, il existe $A \in \mathcal{F}$ avec $\mathbb{P}(A) = 1$ tel que

$$\forall \omega \in A, \exists n_0 = n_0(\omega) < +\infty \text{ tel que } |X_n(\omega) - X(\omega)| < \varepsilon, \forall n \geq n_0.$$

Donc pour tout $\omega \in A$, on a $\limsup_{n \rightarrow \infty} |X_n(\omega) - X(\omega)| \leq \varepsilon$. D'où : $\mathbb{P}(\limsup_{n \rightarrow \infty} |X_n - X| \leq \varepsilon) = 1$. Le réel $\varepsilon > 0$ étant quelconque, on déduit que $\mathbb{P}(\lim_{n \rightarrow \infty} (X_n - X) = 0) = 1$; autrement dit, X_n converge presque sûrement vers X . \square

Exemple 2.5. On se donne une suite de v.a. $\{X_n\}_{n \geq 1}$ indépendantes, suivant toutes la même loi. Si $\mathbb{E}(|X_1|) = \infty$, alors $\limsup_{n \rightarrow \infty} n^{-1}|X_n| = \infty$ p.s.

En effet, fixons un réel $a > 0$. On a, par le théorème de Fubini

$$\sum_{n=1}^{\infty} \mathbb{P}(|X_n/n| \geq a) = \sum_{n=1}^{\infty} \mathbb{P}(a^{-1}|X_1| \geq n) = \mathbb{E}(\lfloor a^{-1}|X_1| \rfloor) \geq \mathbb{E}(a^{-1}|X_1|) - 1 = \infty.$$

On applique la deuxième partie du lemme de Borel–Cantelli.

On a utilisé l'identité importante qui dit que si Y est une variable aléatoire à valeurs dans \mathbb{N} , alors $\mathbb{E}(Y) = \sum_{n=1}^{\infty} \mathbb{P}(Y \geq n)$ (les deux termes convergent ou divergent en même temps). En effet, par le théorème de Fubini, $\sum_{n=1}^{\infty} \mathbb{P}(Y \geq n) = \sum_{n=1}^{\infty} \sum_{j=n}^{\infty} \mathbb{P}(Y = j) = \sum_{j=1}^{\infty} \sum_{n=1}^j \mathbb{P}(Y = j) = \sum_{j=1}^{\infty} j \mathbb{P}(Y = j) = \mathbb{E}(Y)$. \square

Définition 2.6 (convergence en probabilité). On dit que la suite $\{X_n\}_{n \geq 1}$ converge en probabilité vers une v.a. X si pour tout $\varepsilon > 0$

$$\lim_{n \rightarrow \infty} \mathbb{P}(|X_n - X| > \varepsilon) = 0.$$

Remarque. Il est facile de voir que les résultats usuels sur les limites (unicité de la limite, linéarité, etc) sont valables pour les convergences p.s. et en probabilité. \square

Exemple 2.7. Soit $\{X_n\}$ une suite de v.a. réelles telle que $\mathbb{E}(X_n) \rightarrow a$ et $\text{Var}(X_n) \rightarrow 0$. Alors, par l'inégalité de Markov, pour tout $\varepsilon > 0$,

$$\begin{aligned} \mathbb{P}(|X_n - a| > \varepsilon) &= \mathbb{P}((X_n - a)^2 > \varepsilon^2) \leq \frac{\mathbb{E}[(X_n - a)^2]}{\varepsilon^2} \\ &= \frac{\text{Var}(X_n) + (a - \mathbb{E}(X_n))^2}{\varepsilon^2} \rightarrow 0. \end{aligned}$$

Donc X_n converge en probabilité vers a . □

Exemple 2.8. Si les v.a. réelles X_n suivent la même loi, alors

$$\mathbb{P}\left(\left|\frac{X_n}{n}\right| > \varepsilon\right) = \mathbb{P}(|X_n| > \varepsilon n) \rightarrow 0.$$

Donc $\frac{X_n}{n}$ converge en probabilité vers 0. Dans les deux exemples, il n'y a rien qui nous garantit la convergence p.s. □

Exemple 2.9. $\{X_n\}_{n \geq 1}$ converge en probabilité vers X si et seulement si, pour tout $\varepsilon > 0$, il existe $N \in \mathbb{N}$ tel que $\mathbb{P}(|X_n - X| > \varepsilon) \leq \varepsilon$ pour tout $n \geq N$.

Il suffit de prouver la partie "si". Fixons $\varepsilon_0 > 0$. Par hypothèse, pour tout $\varepsilon \in]0, \varepsilon_0]$, il existe $N \in \mathbb{N}$ tel que : si $n \geq N$, alors $\mathbb{P}(|X_n - X| > \varepsilon) \leq \varepsilon$. Donc $\mathbb{P}(|X_n - X| > \varepsilon_0) \leq \mathbb{P}(|X_n - X| > \varepsilon) \leq \varepsilon$. Par définition, $\lim_{n \rightarrow \infty} \mathbb{P}(|X_n - X| > \varepsilon_0) = 0$. Autrement dit, $X_n \rightarrow X$ en probabilité. □

Proposition 2.10. Si $X_n \rightarrow X$ p.s., alors la convergence a lieu également en probabilité.

Preuve. Soit $A \in \mathcal{F}$ tel que $\mathbb{P}(A) = 1$ et que pour tout $\omega \in A$, $X_n(\omega) \rightarrow X(\omega)$, $n \rightarrow \infty$.

On fixe $\varepsilon > 0$. Pour chaque entier n , on considère l'événement

$$B_n = \{\omega \in A : \sup_{k \geq n} |X_k(\omega) - X(\omega)| > \varepsilon\}.$$

La suite des B_n est décroissante et $\bigcap_{n \geq 1} B_n = \emptyset$ (puisque X_n converge vers X sur A). On a donc $\lim_{n \rightarrow \infty} \mathbb{P}(B_n) = 0$. Comme

$$\{|X_n - X| > \varepsilon\} \subset B_n \cup A^c,$$

on a a fortiori $\lim_{n \rightarrow \infty} \mathbb{P}(|X_n - X| > \varepsilon) = 0$. □

Exemple 2.11. La réciproque de la Proposition 2.10 est fautive; voici un contre exemple. On prend $\Omega = [1, 2]$ que l'on munit de la tribu borélienne et de la mesure de Lebesgue.

Pour chaque n , on note k l'unique entier tel que $2^k \leq n < 2^{k+1}$, et on prend $X_n(\omega) = \mathbf{1}_{[n2^{-k}, (n+1)2^{-k}]}(\omega)$. Il est clair que pour tout $\varepsilon \in]0, 1[$, $\mathbb{P}(|X_n| > \varepsilon) = 2^{-k}$, de sorte que $X_n \rightarrow 0$ en probabilité. Néanmoins, pour tout $\omega \in [1, 2]$, il existe une infinité d'entiers n pour lesquels $X_n(\omega) = 1$, et $X_n(\omega)$ ne converge pas vers 0. \square

En revanche on a une réciproque partielle.

Proposition 2.12. Si $X_n \rightarrow X$ en probabilité, alors il existe une sous-suite extraite $X_{N(n)}$ qui converge vers X p.s.

Preuve. Comme $\mathbb{P}(|X_k - X| > \varepsilon) \rightarrow 0$ ($k \rightarrow \infty$) pour tout $\varepsilon > 0$, on peut trouver pour tout $n \geq 1$ un entier $N(n)$ tel que $\mathbb{P}(|X_{N(n)} - X| > 1/n) \leq 2^{-n}$. De plus, la suite $\{N(n)\}_{n \geq 1}$ peut être choisie de sorte qu'elle soit strictement croissante. La série $\sum \mathbb{P}(|X_{N(n)} - X| > 1/n)$ étant convergente, on peut appliquer l'Exemple 2.3 pour voir que $X_{N(n)}$ converge presque sûrement vers X . \square

Proposition 2.13. $X_n \rightarrow X$ en probabilité si et seulement si pour toute sous-suite $(X_{N(n)})$ on peut en extraire une sous-sous-suite $(X_{N(M(n))})$ qui converge vers X presque sûrement.

Preuve. Si $X_n \rightarrow X$ en probabilité, alors $X_{N(n)} \rightarrow X$ en probabilité. Par la proposition précédente, on peut en extraire une sous-sous-suite $(X_{N(M(n))})$ qui converge vers X p.s.

Réciproquement, supposons que pour toute sous-suite $(X_{N(n)})$ on peut en extraire une sous-sous-suite $(X_{N(M(n))})$ qui converge vers X presque sûrement. Fixons $\varepsilon > 0$, et on étudie la suite numérique $a_n := \mathbb{P}(|X_n - X| > \varepsilon)$. Pour toute sous-suite $(a_{N(n)})$ on peut en extraire une sous-sous-suite $(a_{N(M(n))})$ qui converge vers 0. Ceci équivaut à $a_n \rightarrow 0$ ($n \rightarrow \infty$). Autrement dit, $X_n \rightarrow X$ en probabilité. \square

Remarque. On se gardera de croire que la condition que “pour toute sous-suite $(X_{N(n)})$ on puisse en extraire une sous-sous-suite $(X_{N(M(n))})$ qui converge vers X presque sûrement” soit équivalente à la convergence p.s. de (X_n) vers X . D'après la proposition précédente, ceci signifie simplement la convergence en probabilité de (X_n) vers X , ce qui est plus faible que la convergence p.s. En réalité, la convergence p.s. de (X_n) vers X est équivalente à la condition suivante : il existe $A \in \mathcal{F}$ avec $\mathbb{P}(A) = 1$ tel que tout $\omega \in A$ vérifie la propriété : pour toute sous-suite $X_{N(n,\omega)}(\omega)$ on puisse en extraire une sous-sous-suite $X_{N(M(n,\omega),\omega)}(\omega)$ qui converge vers $X(\omega)$. Mais attention, cette dernière propriété concerne toute sous-suite $X_{N(n,\omega)}(\omega)$ où $N(n,\omega)$ dépend de ω , ce qui n'est pas le cas de la condition du départ qui était un peu plus faible. \square

La Proposition 2.13 a surtout des applications théoriques. En voici deux.

Proposition 2.14 (lemme de Fatou). Si $X_n \geq 0$ (pour tout $n \geq 1$) et si $X_n \rightarrow X$ en probabilité, alors $\liminf_{n \rightarrow \infty} \mathbb{E}(X_n) \geq \mathbb{E}(X)$.

Preuve. Soit $(X_{N(n)})$ une sous-suite quelconque. Par la proposition précédente, il existe une sous-sous-suite $(X_{N(M(n))})$ qui converge vers X p.s. Par le lemme de Fatou usuel, $\liminf_{n \rightarrow \infty} \mathbb{E}(X_{N(M(n))}) \geq \mathbb{E}(X)$. Donc $\liminf_{n \rightarrow \infty} \mathbb{E}(X_n) \geq \mathbb{E}(X)$. \square

Proposition 2.15. Si $X_n \rightarrow X$ en probabilité, et si f est une fonction continue sur \mathbb{R} , alors $f(X_n) \rightarrow f(X)$ en probabilité.

Preuve. Soit $(f(X_{N(n)}))$ une sous-suite quelconque. Par la Proposition 2.13, il existe une sous-sous-suite $(X_{N(M(n))})$ qui converge vers X p.s., donc $(f(X_{N(M(n))}))$ converge vers $f(X)$ p.s. En utilisant de nouveau la Proposition 2.13, $f(X_n) \rightarrow f(X)$ en probabilité. \square

En théorie de l'intégration, on a vu d'autres modes de convergence que l'on rappelle.

Définition 2.16 (convergence dans L^p). Pour tout $p \geq 1$, on dit qu'une suite $\{X_n\}_{n \geq 1}$ de v.a. converge dans $L^p(\Omega, \mathcal{F}, \mathbb{P})$ vers une v.a. $X \in L^p(\Omega, \mathcal{F}, \mathbb{P})$ si

$$\lim_{n \rightarrow \infty} \mathbb{E}(|X_n - X|^p) = 0.$$

Lorsque $p = 2$, on dit également que X_n converge vers X en moyenne quadratique.

Remarque. (i) Par l'inégalité de Hölder, si $X_n \rightarrow X$ dans L^p , et si $q \in [1, p[$, alors la convergence a lieu dans L^q .

(ii) Si $X_n \rightarrow X$ dans L^p , alors $|X_n| \rightarrow |X|$ dans L^p (car $||X_n| - |X|| \leq |X_n - X|$).

(iii) Si $X_n \rightarrow X$ dans L^p , et si $p \in \mathbb{N}^*$, alors $\mathbb{E}(X_n^p) \rightarrow \mathbb{E}(X^p)$ (exercice à domicile). \square

Exemple 2.17. On se donne une suite de v.a. $\{X_n\}_{n \geq 1}$, suivant toutes la même loi. Si $X_1 \in L^1(\Omega, \mathcal{F}, \mathbb{P})$, alors la suite X_n/n converge vers 0 p.s. et dans $L^1(\mathbb{P})$.

En effet, $\mathbb{E}(|X_n/n|) = n^{-1}\mathbb{E}(|X_1|) \rightarrow 0$, donc $X_n/n \rightarrow 0$ dans $L^1(\mathbb{P})$.

Pour la convergence p.s., on fixe un réel $k > 0$. On a, par le théorème de Fubini,

$$\sum_{n=1}^{\infty} \mathbb{P}(|X_n/n| \geq 1/k) = \sum_{n=1}^{\infty} \mathbb{P}(k|X_1| \geq n) = \mathbb{E}(\lfloor k|X_1| \rfloor) < \infty.$$

On applique la partie directe de Borel–Cantelli. L'événement

$$A_k = \{\limsup_{n \rightarrow \infty} n^{-1} |X_n| < 1/k\}.$$

a pour probabilité 1, il en est de même pour l'intersection de ces événements. D'où la convergence p.s. \square

Comparons la notion de convergence L^p avec les précédentes.

Proposition 2.18. Si $X_n \rightarrow X$ dans L^p , alors la convergence a encore lieu en probabilité.

Preuve. Fixons $\varepsilon > 0$. Par l'inégalité de Markov,

$$\mathbb{P}(|X - X_n| > \varepsilon) \leq \frac{\mathbb{E}(|X - X_n|^p)}{\varepsilon^p} \rightarrow 0, \quad n \rightarrow \infty. \quad \square$$

Exemple 2.19. La réciproque est en général fautive. Soit $\{X_n\}_{n \geq 3}$ une suite de v.a. réelles telle que $\mathbb{P}(X_n = n) = \frac{1}{\ln n}$ et $\mathbb{P}(X_n = 0) = 1 - \frac{1}{\ln n}$. Pour tout $\varepsilon > 0$,

$$\mathbb{P}(|X_n| > \varepsilon) \leq \frac{1}{\ln n} \rightarrow 0.$$

Donc $X_n \rightarrow 0$ en probabilité. D'autre part,

$$\mathbb{E}(|X_n|^p) = \frac{n^p}{\ln n} \rightarrow \infty.$$

Donc X_n ne converge pas vers 0 dans L^p . \square

Proposition 2.20. Si $X_n \rightarrow X$ en probabilité et s'il existe une v.a. réelle $Y \in L^p$ telle que $|X_n| \leq Y$ pour tout n , alors $X_n \rightarrow X$ dans L^p .

Preuve. Si la suite $\{X_n\}_{n \geq 1}$ converge vers X en probabilité, c'est encore le cas pour toute suite extraite, disons $\{X_{N(n)}\}_{n \geq 1}$. On sait qu'on peut extraire de cette dernière une sous-suite, disons $\{X_{N(M(n))}\}_{n \geq 1}$ qui converge p.s. vers X . Par hypothèse, on peut appliquer le théorème de convergence dominée, i.e., $X_{N(M(n))} \rightarrow X$ dans $L^p(\Omega, \mathcal{F}, \mathbb{P})$. Ainsi, de toute suite extraite de $\{X_n\}_{n \geq 1}$, on a su extraire une sous-sous-suite qui converge vers X dans L^p . Donc $X_n \rightarrow X$ dans L^p . \square

Corollaire 2.21 (convergence dominée). Si $X_n \rightarrow X$ en probabilité et s'il existe une v.a. réelle Y admettant un moment d'ordre 1 telle que $|X_n| \leq Y$ pour tout n , alors $\mathbb{E}(X_n) \rightarrow \mathbb{E}(X)$.

Exemple 2.22. On se donne Y une variable aléatoire réelle définie sur l'espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$, uniformément distribuée sur l'intervalle $[3, 6]$. Pour tout $n \geq 1$ et tout $\omega \in \Omega$, on pose

$$X_n(\omega) = \begin{cases} 5n^2 & \text{si } 3 \leq Y(\omega) \leq 3 + (4/n^2), \\ 0 & \text{sinon.} \end{cases}$$

- (a) Déterminer $\mathbb{E}(X_n)$ et $\mathbb{E}(X_n^2)$, ($n \geq 1$).
- (b) Calculer $\mathbb{E}(X_{n+1}X_{n+2})$, ($n \geq 1$).
- (c) La suite (X_n) converge-t-elle presque sûrement vers une limite?
- (d) La suite (X_n) converge-t-elle en probabilité vers une limite?
- (e) La suite (X_n) converge-t-elle dans L^1 vers une limite?

Solution. (a) On peut écrire $X_n = 5n^2 \mathbf{1}_{[3, 3+4n^{-2}]}(Y)$. Donc $\mathbb{E}(X_n) = 5n^2 \mathbb{P}(3 \leq Y \leq 3 + 4n^{-2})$ et $\mathbb{E}(X_n^2) = 25n^4 \mathbb{P}(3 \leq Y \leq 3 + 4n^{-2})$. Si $n = 1$, on a $\mathbb{P}(3 \leq Y \leq 7) = 1$, donc $\mathbb{E}(X_1) = 5$ et $\mathbb{E}(X_1^2) = 25$. Si $n \geq 2$, on a $\mathbb{P}(3 \leq Y \leq 3 + 4n^{-2}) = 4n^{-2}/3$, donc $\mathbb{E}(X_n) = \frac{20}{3}$ et $\mathbb{E}(X_n^2) = \frac{100}{3}n^2$ ($n \geq 2$).

(b) Par définition, $X_{n+1}X_{n+2} = 25(n+1)^2(n+2)^2 \mathbf{1}_{[3, 3+4(n+2)^{-2}]}(Y)$, ce qui implique que $\mathbb{E}(X_{n+1}X_{n+2}) = 25(n+1)^2(n+2)^2 \mathbb{P}(3 \leq Y \leq 3 + 4(n+2)^{-2}) = \frac{100}{3}(n+1)^2$.

(c) Soit $\Omega_0 = \{\omega \in \Omega : 3 < Y(\omega) \leq 6\}$. On a $\mathbb{P}(\Omega_0) = 1$. Pour tout $\omega \in \Omega_0$, posons $n_0 = n_0(\omega) = 2/\sqrt{Y(\omega) - 3}$, alors $Y(\omega) > 3 + 4/n^2$ (donc $X_n(\omega) = 0$) dès que $n > n_0$. Par conséquent, $\mathbb{P}\{\omega \in \Omega : \lim_{n \rightarrow \infty} X_n(\omega) = 0\} = 1$, c'est-à-dire que X_n converge presque sûrement vers (la variable constante) 0.

(d) Convergence presque sûre impliquant convergence en probabilité, on déduit que X_n converge en probabilité vers 0.

(e) Si X_n convergerait dans L^1 vers une limite disons X , alors elle convergerait également en probabilité vers X (Proposition 2.17). Or, dans la question précédente, on a vu que X_n converge en probabilité vers 0, ce qui impliquerait $\mathbb{P}(X = 0) = 1$. Autrement dit, X_n convergerait dans L^1 vers 0, ce qui entraînerait $\lim_{n \rightarrow \infty} \mathbb{E}(X_n) = \mathbb{E}(0) = 0$. Ceci contredit le fait que $\mathbb{E}(X_n) = \frac{20}{3}$ ($n \geq 2$). En conclusion, X_n ne peut converger dans L^1 . \square

Exemple 2.23. Il est à noter qu'une suite de v.a. réelles converge p.s., en probabilité ou dans L^p ($p \geq 1$) si et seulement si elle est de Cauchy. Ceci est clair pour la convergence dans L^p (car dans la théorie d'intégration on a vu que L^p , $p \geq 1$, est un espace de Banach) et pour la convergence p.s. (par vérification directe). On montre maintenant qu'il est vrai pour la convergence en probabilité.

Soient $(X_n)_{n \geq 1}$ des v.a. réelles qui forment une suite de Cauchy dans le sens que quel que soit $\varepsilon > 0$, $\mathbb{P}(|X_p - X_q| > \varepsilon) \rightarrow 0$, $p \rightarrow \infty$, $q \rightarrow \infty$ (c'est-à-dire que pour

tout $\varepsilon > 0$ et tout $\delta > 0$, il existe $N \in \mathbb{N}$ vérifiant la propriété suivante : $p \geq N$, $q \geq N \Rightarrow \mathbb{P}(|X_p - X_q| > \varepsilon) \leq \delta$. On peut trouver une suite strictement croissante $(N(k))_{k \geq 1}$ telle que pour chaque $k \geq 1$, $\mathbb{P}(|X_p - X_q| > 2^{-k}) \leq 2^{-k}$ si $p \geq N(k)$, $q \geq N(k)$. En particulier, $\mathbb{P}(|X_{N(k+1)} - X_{N(k)}| > 2^{-k}) \leq 2^{-k}$, quel que soit $k \geq 1$. Il résulte du lemme de Borel–Cantelli qu’il existe $A \in \mathcal{F}$ avec $\mathbb{P}(A) = 1$ tel que si $\omega \in A$, alors on puisse trouver $k_0 = k_0(\omega) < \infty$ tel que $|X_{N(k+1)}(\omega) - X_{N(k)}(\omega)| \leq 2^{-k}$ lorsque $k \geq k_0$. La série géométrique étant convergente, on déduit que la série numérique $\sum_{k=1}^{\infty} (X_{N(k+1)}(\omega) - X_{N(k)}(\omega))$ converge pour tout $\omega \in A$. Donc $X_{N(k)}$ converge (lorsque $k \rightarrow \infty$) presque sûrement (et a fortiori en probabilité), et soit X la limite. Alors pour $\varepsilon > 0$, $\mathbb{P}(|X_n - X| > \varepsilon) \leq \mathbb{P}(|X_n - X_{N(k)}| > \varepsilon/2) + \mathbb{P}(|X_{N(k)} - X| > \varepsilon/2) \rightarrow 0$ (lorsque $k \rightarrow \infty$, $n \rightarrow \infty$). Ceci prouve que X_n converge en probabilité vers X .

Réciproquement, si $X_n \rightarrow X$ en probabilité, alors pour tout $\varepsilon > 0$, $\mathbb{P}(|X_n - X_m| > \varepsilon) \leq \mathbb{P}(|X_n - X| > \varepsilon/2) + \mathbb{P}(|X_m - X| > \varepsilon/2) \rightarrow 0$ ($n \rightarrow \infty$, $m \rightarrow \infty$). \square

3. Séries de variables aléatoires indépendantes

Il est connu que la série numérique $\sum \frac{1}{n}$ diverge (série harmonique), tandis que la série numérique $\sum \frac{(-1)^n}{n}$ converge (théorème des séries alternées). Une question naturelle est de savoir ce qui se passe lorsque l’on met un signe aléatoire sur chacun des termes $\frac{1}{n}$. Plus précisément, soit $(\varepsilon_n)_{n \geq 1}$ une suite de v.a. indépendantes telle que $\mathbb{P}(\varepsilon_n = 1) = \mathbb{P}(\varepsilon_n = -1) = 1/2$ quel que soit n , et on cherche à étudier la convergence (dans quel sens?) de la série $\sum \frac{\varepsilon_n}{n}$.

L’étude de ce problème se situe dans un contexte plus général. On se donne une suite de v.a. réelles $\{X_n\}_{n \geq 1}$ que l’on suppose indépendantes (c’est-à-dire que pour tout n , X_1, \dots, X_n sont des v.a. indépendantes). On note $S_n = X_1 + \dots + X_n$ la somme partielle jusqu’à l’ordre n , et on s’intéresse à la convergence de la suite des v.a. S_n .

Théorème 3.1. Si S_n converge en probabilité, alors S_n converge également p.s.

La preuve de ce théorème difficile ne fait pas partie du programme du cours. Elle repose sur un lemme technique.

Lemme 3.2. Soient $\{U_n\}_{1 \leq n \leq M}$ et $\{V_n\}_{1 \leq n \leq M}$ deux suites finies de v.a. positives telles que pour chaque $n \leq M$, V_n est indépendant de $(U_n, U_{n+1}, \dots, U_M)$. On a alors pour tout $x > 0$,

$$\mathbb{P}\left(\max_{1 \leq n \leq M} (U_n - V_n) > x\right) \geq \left(\min_{1 \leq n \leq M} \mathbb{P}(V_n < x)\right) \times \mathbb{P}\left(\max_{1 \leq n \leq M} U_n > 2x\right).$$

Preuve du Lemme 3.2. Définissons les événements suivants :

$$A_n = \{U_n > 2x, U_{n+1} \leq 2x, \dots, U_M \leq 2x\}, \quad 1 \leq n \leq M.$$

(En fait, $A_M = \{U_M > 2x\}$). Il est clair que les A_n sont deux-à-deux disjoints, et que $\bigcup_{n=1}^M A_n = \{\max_{1 \leq n \leq M} U_n > 2x\}$. Considérons l'expression :

$$I = \sum_{n=1}^M \mathbb{P}(\{V_n < x\} \cap A_n).$$

On a, d'une part,

$$\begin{aligned} I &= \mathbb{P}\left(\bigcup_{n=1}^M (\{V_n < x\} \cap A_n)\right) \leq \sum_{n=1}^M \mathbb{P}(V_n < x, U_n > 2x) \\ &\leq \mathbb{P}\left(\max_{1 \leq n \leq M} (U_n - V_n) > x\right). \end{aligned}$$

D'autre part, grâce à l'indépendance,

$$\begin{aligned} I &= \sum_{n=1}^M \mathbb{P}(V_n < x) \mathbb{P}(U_n > 2x, U_{n+1} \leq 2x, \dots, U_M \leq 2x) \\ &\geq \left(\min_{1 \leq n \leq M} \mathbb{P}(V_n < x)\right) \times \sum_{n=1}^M \mathbb{P}(U_n > 2x, U_{n+1} \leq 2x, \dots, U_M \leq 2x) \\ &= \left(\min_{n=1, \dots, M} \mathbb{P}(V_n < x)\right) \times \mathbb{P}\left(\bigcup_{n=1}^M \{U_n > 2x, U_{n+1} \leq 2x, \dots, U_M \leq 2x\}\right) \\ &= \left(\min_{1 \leq n \leq M} \mathbb{P}(V_n < x)\right) \times \mathbb{P}\left(\max_{1 \leq n \leq M} U_n > 2x\right). \quad \square \end{aligned}$$

Preuve du Théorème 3.1. Soit S la limite des S_n (en probabilité). On sait qu'il existe une suite extraite $S_{N(n)}$ qui converge vers S p.s. (ici, $N : \mathbb{N}^* \rightarrow \mathbb{N}^*$ est une certaine application strictement croissante). Pour chaque entier $n \geq 1$, il existe un unique entier ℓ tel que $N(\ell - 1) \leq n < N(\ell)$, et on pose $\alpha(n) = N(\ell - 1)$. L'application α est croissante sur \mathbb{N}^* . On introduit

$$U_n = |S - S_n| \quad , \quad V_n = |S_{\alpha(n)} - S_n| \quad , \quad W_n = |S - S_{\alpha(n)}|.$$

Par construction, on sait que $W_n \rightarrow 0$ p.s. et $V_n \rightarrow 0$ en probabilité (critère de Cauchy). De plus, pour chaque n fixé, la v.a. V_n est indépendante de (U_n, U_{n+1}, \dots) .

Fixons $x > 0$ arbitrairement petit. Puisque $W_n \geq U_n - V_n$, on a d'après le Lemme 3.2 que pour tous les entiers p, q avec $p \leq q$

$$\begin{aligned} \mathbb{P}\left(\max_{p \leq n \leq q} U_n > 2x\right) &\leq \frac{\mathbb{P}(\max_{p \leq n \leq q} (U_n - V_n) > x)}{\min_{p \leq n \leq q} \mathbb{P}(V_n < x)} \\ &\leq \frac{\mathbb{P}(\max_{p \leq n \leq q} W_n > x)}{\min_{p \leq n \leq q} \mathbb{P}(V_n < x)}. \end{aligned}$$

Fixons $\varepsilon > 0$. Puisque $V_n \rightarrow 0$ en probabilité, on peut trouver un entier n_0 tel que $\mathbb{P}(V_n < x) > 1 - \varepsilon$ dès que $n > n_0$.

Puisque $W_n \rightarrow 0$ p.s., on peut trouver un entier n_1 tel que $\mathbb{P}(\max_{p \leq n \leq q} W_n > x) < \varepsilon$ pour tout $n_1 \leq p \leq q$.

En conséquence, pour $p \geq n_0 + n_1$,

$$\mathbb{P}\left(\max_{p \leq n \leq q} U_n > 2x\right) \leq \frac{\varepsilon}{1 - \varepsilon}, \quad \forall q \geq p,$$

et donc également, en faisant tendre d'abord q vers ∞ puis p vers ∞

$$\mathbb{P}\left(\limsup_{n \rightarrow \infty} U_n > 2x\right) \leq \frac{\varepsilon}{1 - \varepsilon}.$$

Comme ε et x sont arbitraires, on a bien $\limsup_{n \rightarrow \infty} U_n = 0$ p.s., ce qui termine la preuve du Théorème 3.1. \square

Proposition 3.3. Supposons que les v.a. réelles indépendantes X_n sont centrées et admettent toutes un moment d'ordre 2, i.e., $\mathbb{E}(X_n) = 0$ et $\mathbb{E}(X_n^2) < \infty$. Alors S_n converge dans $L^2(\Omega, \mathcal{F}, \mathbb{P})$ si et seulement si la série des moyennes quadratiques converge, $\sum_{n=1}^{\infty} \mathbb{E}(X_n^2) < \infty$. Dans ce cas, la convergence a lieu p.s. également, et la limite, S_{∞} , est une v.a. centrée dont la variance vaut $\sum_{n=1}^{\infty} \mathbb{E}(X_n^2)$.

Preuve. Par le critère de Cauchy, S_n converge dans $L^2(\Omega, \mathcal{F}, \mathbb{P})$ si et seulement si $\mathbb{E}[(S_n - S_m)^2] \rightarrow 0$ ($n \rightarrow \infty, m \rightarrow \infty$). Or, pour $m < n$, $\mathbb{E}[(S_n - S_m)^2] = \text{Var}(S_n - S_m) = \sum_{i=m}^n \text{Var}(X_i) = \sum_{i=m}^n \mathbb{E}(X_i^2)$, qui converge vers 0 si et seulement si la série $\sum_i \mathbb{E}(X_i^2)$ converge. Donc S_n converge dans $L^2(\Omega, \mathcal{F}, \mathbb{P})$ si et seulement si $\sum_{n=1}^{\infty} \mathbb{E}(X_n^2) < \infty$.

Si c'est le cas, la convergence a encore lieu en probabilité; il ne reste qu'à appliquer le Théorème 3.1. \square

Exemple 3.4. On prend $X_n = \frac{\varepsilon_n}{n}$, avec $\{\varepsilon_n\}_{n \geq 1}$ une suite de v.a. indépendantes et équidistribuées telle que $\mathbb{P}(\varepsilon_1 = 1) = \mathbb{P}(\varepsilon_1 = -1) = 1/2$. La série $\sum_{n=1}^{\infty} |X_n|$ diverge p.s., alors que la série $\sum_{n=1}^{\infty} X_n$ converge p.s. (semi-convergence). \square

Chapitre 8. Loi des grands nombres

On se donne X_1, \dots, X_n, \dots une suite de v.a. réelles indépendantes et identiquement distribuées (en abbréviation : i.i.d.), et on s'intéresse à la convergence de la suite au sens de Césaro.

1. Loi des grands nombres

Pour ne pas passer trop de temps sur des questions techniques un peu délicates, on admettra que si P_X est une loi de probabilité sur \mathbb{R} , on peut construire un espace de probabilité adéquat et une suite de v.a. X_1, \dots, X_n, \dots indépendantes et toutes de loi P_X . Il est un cas simple dans lequel une telle construction est explicite.

Exemple 1.1. (Construction de Lebesgue d'une suite de v.a. de Bernoulli indépendantes). On prend $\Omega = [0, 1)$, muni de la tribu borélienne et de la mesure de Lebesgue. On écrit $\omega \in \Omega$ sous forme dyadique, $\omega = \sum_{n=1}^{\infty} X_n(\omega)2^{-n}$ (suivant la convention usuelle, l'écriture où tous les $X_n(\omega)$ valent 1 à partir d'un certain rang est exclue). Alors la suite $\{X_n\}_{n \geq 1}$ est une suite de v.a. indépendantes, chacune suit une loi de Bernoulli de paramètre $1/2$. En effet, si on se donne une suite finie (x_1, \dots, x_n) de 0 et 1 et si on pose $y = \sum_{j=1}^n x_j 2^{-j}$, alors $\{X_1 = x_1, \dots, X_n = x_n\} = [y, y + 2^{-n}[$ et donc

$$\mathbb{P}(X_1 = x_1, \dots, X_n = x_n) = 2^{-n} = \prod_{j=1}^n \mathbb{P}(X_j = x_j). \quad \square$$

La loi des grands nombres a été découverte par Jacques Bernoulli, précisément pour une suite de v.a. de Bernoulli indépendantes. C'est l'un des résultats les plus importants de la théorie des probabilités.

Théorème 1.2 (loi des grands nombres forte). (i) Si $\mathbb{E}(|X|) < \infty$ disons $\mathbb{E}(X) = \mu$, alors

$$\lim_{n \rightarrow \infty} \frac{X_1 + \cdots + X_n}{n} = \mu,$$

presque sûrement.

(ii) Si $\mathbb{E}(|X|) = \infty$, alors la suite $\{n^{-1}(X_1 + \cdots + X_n)\}_{n \geq 1}$ diverge p.s.

Preuve. (i) Nous ne démontrerons ici la loi que sous l'hypothèse plus restrictive que X admet un moment d'ordre 2. Une preuve complète est pourvue dans la Section 3 du présent chapitre.

Tout d'abord, en considérant séparément la partie positive de X_n et sa partie négative, on remarque qu'il suffit de démontrer le résultat pour des v.a. *positives*, ce que l'on suppose désormais.

On note $S(n) = X_1 + \cdots + X_n$. On a $\mathbb{E}(S(n)) = n\mu$ et $\text{Var}(S(n)) = n\text{Var}(X)$. En particulier,

$$\text{Var}\left(\frac{S(n^4) - n^4\mu}{n^4}\right) = \frac{\text{Var}(X)}{n^4}.$$

D'après l'inégalité de Bienaymé–Tchebychev,

$$\mathbb{P}\left(|n^{-4}S(n^4) - \mu| > \frac{1}{n}\right) \leq \frac{\text{Var}(X)}{n^2}.$$

On déduit que la série $\sum_{n=1}^{\infty} \mathbb{P}(|n^{-4}S(n^4) - \mu| > 1/n)$ converge. D'après le lemme de Borel–Cantelli, on a donc presque-sûrement,

$$|n^{-4}S(n^4) - \mu| \leq \frac{1}{n}, \quad \text{sauf pour un nombre fini d'entiers } n.$$

Fixons un aléas ω pour lequel l'événement ci-dessus est réalisé et prenons n suffisamment grand. Pour tout entier k tel que $n^4 \leq k \leq (n+1)^4$, on a donc (puisque la suite $(S_j, j = 1, \dots)$ est croissante)

$$\frac{S(k)}{k} \leq \frac{S((n+1)^4)}{n^4} = \left(\frac{n+1}{n}\right)^4 \frac{S((n+1)^4)}{(n+1)^4} \leq \left(\frac{n+1}{n}\right)^4 (\mu + 1/n).$$

Comme $\left(\frac{n+1}{n}\right)^4$ tend vers 1 quand $n \rightarrow \infty$, on a donc établi que pour cet aléas ω

$$\limsup_{k \rightarrow \infty} \frac{S(k)}{k} \leq \mu.$$

De même, on a

$$\frac{S(k)}{k} \geq \frac{S(n^4)}{(n+1)^4} = \left(\frac{n}{n+1}\right)^4 \frac{S(n^4)}{n^4} \geq \left(\frac{n}{n+1}\right)^4 (\mu - 1/n).$$

Comme $\left(\frac{n}{n+1}\right)^4$ tend vers 1 quand $n \rightarrow \infty$, on a donc établi que pour cet aléas ω

$$\liminf_{k \rightarrow \infty} \frac{S(k)}{k} \geq \mu.$$

En conclusion, on a bien pour presque tout ω : $\lim_{k \rightarrow \infty} \frac{S(k)}{k} = \mu$.

(ii) On a déjà vu dans l'Exemple 2.5 du chapitre précédent que si les v.a. X_1, \dots, X_n, \dots sont indépendantes et de même loi et que $\mathbb{E}(|X|) = \infty$, alors p.s., $\limsup_{n \rightarrow \infty} n^{-1}|X_n| = \infty$. Dans ces conditions, la suite des moyennes de Césaro ne peut pas converger. \square

Remarque. On fait souvent référence au théorème comme la loi **forte** des grands nombres. Une conséquence immédiate du théorème est que sous les hypothèses (i), la suite $n^{-1}S(n)$ converge *en probabilité* vers la moyenne μ . On parle dans ce dernier cas de loi **faible** des grands nombres, par opposition à la loi forte. \square

Remarque. La loi des grands nombres justifie l'interprétation fréquentielle de la notion de probabilité. Typiquement, on réalise un grand nombre d'expériences dans des conditions identiques, dont le résultat dépend de l'aléas (exemple, durée de vie d'une ampoule électrique > 1000 heures). On note $X_n = 1$ si la n -ième expérience réussit, $X_n = 0$ sinon. La suite X_1, \dots, X_n, \dots est une suite de v.a. de Bernoulli indépendantes, de moyenne $\mathbb{E}(X) = \mathbb{P}(X = 1)$. La fréquence de succès de l'expérience converge donc p.s. vers la probabilité de succès. \square

Exemple 1.3. Soit $\{U_j\}_{j \geq 1}$ une suite i.i.d. de v.a. suivant la loi uniforme sur $[0, 1]$. Par la loi forte des grands nombres, lorsque n tend vers l'infini,

$$\frac{1}{n} \sum_{j=1}^n U_j \rightarrow \mathbb{E}(U_1) = \frac{1}{2}, \quad \text{p.s.}$$

Ceci correspond à l'intuition que la moyenne (arithmétique) de n nombres réels "choisis par hasard" entre 0 et 1 devrait être proche de $1/2$.

Que se passe-t-il pour la moyenne géométrique? Soit

$$X_n = \left(\prod_{j=1}^n U_j \right)^{1/n}.$$

On a

$$\ln X_n = \frac{1}{n} \sum_{j=1}^n \ln U_j = -\frac{1}{n} \sum_{j=1}^n \ln \left(\frac{1}{U_j} \right).$$

La suite de v.a. $\{\ln(1/U_j)\}_{j \geq 1}$ est de nouveau i.i.d. Pour savoir l'existence du moment d'ordre 1, soit $h : \mathbb{R} \rightarrow \mathbb{R}$ une fonction borélienne bornée, on a

$$\mathbb{E}\left(h\left(\ln \frac{1}{U_1}\right)\right) = \int_0^1 h\left(\ln \frac{1}{x}\right) dx = \int_0^\infty h(y)e^{-y} dy = \int_{\mathbb{R}} h(y)e^{-y} \mathbf{1}_{\mathbb{R}_+^*}(y) dy.$$

Ceci montre que $\ln(1/U_1)$ suit la loi exponentielle de paramètre 1. En particulier, elle admet un moment d'ordre 1, avec $\mathbb{E}(\ln(1/U_1)) = 1$. Donc $\ln X_n \rightarrow -1$, p.s., c'est-à-dire $X_n \rightarrow 1/e$, p.s. \square

Exemple 1.4. Soient f et g deux fonctions continues sur $[0, 1]$, telles que $0 \leq f(x) < Cg(x)$ pour $x \in [0, 1]$, où $C > 0$ est une constante. On montre

$$\lim_{n \rightarrow \infty} \int_0^1 \cdots \int_0^1 \frac{f(x_1) + \cdots + f(x_n)}{g(x_1) + \cdots + g(x_n)} dx_1 \cdots dx_n = \frac{\int_0^1 f(x) dx}{\int_0^1 g(x) dx}.$$

Pour prouver cette convergence, soit $\{U_j\}_{j \geq 1}$ une suite i.i.d. de v.a. suivant la loi uniforme sur $[0, 1]$. Posons

$$X_n = \frac{1}{n} \sum_{j=1}^n f(U_j), \quad Y_n = \frac{1}{n} \sum_{j=1}^n g(U_j).$$

D'après la loi forte des grands nombres, $X_n \rightarrow \mathbb{E}(f(U_1)) = \int_0^1 f(x) dx$, p.s., et $Y_n \rightarrow \int_0^1 g(x) dx$, p.s. Comme $\int_0^1 g(x) dx > 0$, on a $X_n/Y_n \rightarrow \int_0^1 f(x) dx / \int_0^1 g(x) dx$, p.s. Comme $|X_n/Y_n| \leq C$ qui est intégrable, il résulte du théorème de convergence dominée que

$$\mathbb{E}\left(\frac{X_n}{Y_n}\right) \rightarrow \frac{\int_0^1 f(x) dx}{\int_0^1 g(x) dx}.$$

Or,

$$\mathbb{E}\left(\frac{X_n}{Y_n}\right) = \int_0^1 \cdots \int_0^1 \frac{f(x_1) + \cdots + f(x_n)}{g(x_1) + \cdots + g(x_n)} dx_1 \cdots dx_n,$$

ce qui donne le résultat voulu. \square

2. Application : théorème de Glivenko–Cantelli

Nous démontrons le théorème de Glivenko–Cantelli, théorème d'une importance fondamentale en statistique mathématique (souvent appelé "théorème fondamental de statistique").

Théorème 2.1 (Glivenko–Cantelli). Supposons que les v.a. réelles X_1, \dots, X_n soient les éléments d'un échantillon provenant d'une population statistique, c'est-à-dire que ce sont des v.a. indépendantes et identiquement distribuées. Soit $F(\cdot)$ leur fonction de répartition commune, et soit $F_n(\cdot)$ la fonction de répartition empirique de l'échantillon, c'est-à-dire que

$$F_n(x) = \frac{1}{n} \#\{1 \leq i \leq n : X_i \leq x\}, \quad x \in \mathbb{R}.$$

Alors,

$$\sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \rightarrow 0, \quad \text{p.s.}$$

Preuve. Soit $M \geq 2$ un entier. Pour chaque $1 \leq k \leq M - 1$, on peut trouver un nombre réel x_k tel que $F(x_k - 0) \leq \frac{k}{M} \leq F(x_k)$. Par exemple, on peut prendre $x_k = \inf\{x \in \mathbb{R} : F(x) \geq \frac{k}{M}\}$. Posons en plus $x_0 := -\infty$ et $x_M := +\infty$. Pour chaque x fixé, la loi forte des grands nombres nous confirme que $F_n(x)$ converge p.s. vers $F(x)$, et que $F_n(x - 0)$ converge p.s. vers $F(x - 0)$. Donc il existe $A \in \mathcal{F}$ avec $\mathbb{P}(A) = 1$ tel que

$$\begin{aligned} \forall \omega \in A, \quad \exists n_0 = n_0(\omega, M) < \infty \text{ tel que pour } 0 \leq k \leq M \text{ et } n \geq n_0, \\ |F_n(x_k, \omega) - F(x_k)| \leq \frac{1}{M}, \quad |F_n(x_k - 0, \omega) - F(x_k - 0)| \leq \frac{1}{M}. \end{aligned}$$

Soit $x \in]x_{k-1}, x_k[$. Si $\omega \in A$ et $n \geq n_0$, on a

$$\begin{aligned} F_n(x) - F(x) &\leq F_n(x_k - 0) - F(x_{k-1}) \leq F(x_k - 0) + \frac{1}{M} - F(x_{k-1}) \\ &\leq \frac{k}{M} + \frac{1}{M} - \frac{k-1}{M} = \frac{2}{M}, \end{aligned}$$

et

$$\begin{aligned} F_n(x) - F(x) &\geq F_n(x_{k-1}) - F(x_k - 0) \geq F(x_{k-1}) - \frac{1}{M} - F(x_k - 0) \\ &\geq \frac{k-1}{M} - \frac{1}{M} - \frac{k}{M} = -\frac{2}{M}. \end{aligned}$$

D'où : $\sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \leq 2/M$ pour $\omega \in A$ et $n \geq n_0$, ce qui implique que

$$\mathbb{P} \left(\limsup_{n \rightarrow \infty} \sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \leq \frac{2}{M} \right) = 1.$$

La valeur de $M \geq 2$ étant quelconque, on obtient le résultat. □

3. Preuve de la loi des grands nombres

(Cette preuve ne fait pas partie du programme de l'examen.)

Soient X_1, X_2, \dots des v.a. réelles i.i.d. telles que $\mathbb{E}(|X_1|) < +\infty$. On montre que $\frac{S(n)}{n} \rightarrow \mathbb{E}(X_1)$, p.s., où $S(n) := \sum_{i=1}^n X_i$.

Il suffit de démontrer le résultat pour les v.a. positives (sinon on considèrera les parties positives et les parties négatives, séparément). On suppose donc $X_n \geq 0$.

La démonstration se fait en plusieurs étapes.

Lemme 3.1. Soient $Y_k := X_k \mathbf{1}_{\{X_k \leq k\}}$ ($k \geq 1$) et $T(n) := \sum_{k=1}^n Y_k$. Alors $\frac{S(n) - T(n)}{n} \rightarrow 0$, p.s.

Preuve. On a $\sum_{n=1}^{\infty} \mathbb{P}(X_n \geq n) = \sum_{n=1}^{\infty} \mathbb{P}(X_1 \geq n) = \sum_{n=1}^{\infty} \mathbb{P}(\lfloor X_1 \rfloor \geq n) = \mathbb{E}(\lfloor X_1 \rfloor) < \infty$. Par la première partie du lemme de Borel–Cantelli, il existe $A \in \mathcal{F}$ avec $\mathbb{P}(A) = 1$ tel que pour tout $\omega \in A$, il existe $n_0 = n_0(\omega) < \infty$ satisfaisant $X_n(\omega) < n$ ($n \geq n_0$). Donc $Y_n(\omega) = X_n(\omega)$. Ceci implique que $S(n) - T(n) = S(n_0) - T(n_0)$ pour tout $\omega \in A$ et tout $n \geq n_0$. \square

Lemme 3.2. $\frac{\mathbb{E}(T(n))}{n} \rightarrow \mathbb{E}(X_1)$, $n \rightarrow \infty$.

Preuve. Par le théorème de convergence dominée (ou par le théorème de convergence monotone), $\mathbb{E}(Y_k) = \mathbb{E}(X_1 \mathbf{1}_{\{X_1 \leq k\}}) \rightarrow \mathbb{E}(X_1)$ ($k \rightarrow \infty$). Donc $\frac{\mathbb{E}(T(n))}{n} = \frac{\sum_{k=1}^n \mathbb{E}(Y_k)}{n} \rightarrow \mathbb{E}(X_1)$, $n \rightarrow \infty$. \square

Lemme 3.3.

$$\sum_{k=1}^{\infty} \frac{\text{Var}(Y_k)}{k^2} \leq 8\mathbb{E}(X_1).$$

Preuve. Remarquons d'abord que $\mathbb{E}(Z^p) = \int_0^{\infty} \mathbb{P}(Z > x) p x^{p-1} dx$, pour toute v.a. positive Z et tout $p > 0$. En effet, par le théorème de Fubini, $\int_0^{\infty} \mathbb{P}(Z > x) p x^{p-1} dx = \int_0^{\infty} \int_{\Omega} \mathbf{1}_{\{Z(\omega) > y^{1/p}\}} d\mathbb{P}(\omega) dy = \int_{\Omega} \int_0^{\infty} \mathbf{1}_{\{Z(\omega) > y^{1/p}\}} dy d\mathbb{P}(\omega) = \int_{\Omega} Z^p(\omega) d\mathbb{P}(\omega) = \mathbb{E}(Z^p)$.

Donc

$$\begin{aligned} \text{Var}(Y_k) &\leq \mathbb{E}(Y_k^2) = 2 \int_0^{\infty} \mathbb{P}(Y_k > x) x dx \\ &= 2 \int_0^k \mathbb{P}(Y_k > x) x dx = 2 \int_0^k \mathbb{P}(X_k > x) x dx \\ &= 2 \int_0^k \mathbb{P}(X_1 > x) x dx. \end{aligned}$$

Par le théorème de Fubini,

$$\begin{aligned} \sum_{k=1}^{\infty} \frac{\text{Var}(Y_k)}{k^2} &\leq 2 \sum_{k=1}^{\infty} \int_0^{\infty} \frac{\mathbb{P}(X_1 > x)x \mathbf{1}_{\{x \leq k\}}}{k^2} dx \\ &= 2 \int_0^{\infty} \left(\sum_{k=1}^{\infty} \frac{\mathbb{P}(X_1 > x)x \mathbf{1}_{\{x \leq k\}}}{k^2} \right) dx \\ &\leq 2 \int_0^1 \mathbb{P}(X_1 > x)x \left(\sum_{k=1}^{\infty} \frac{1}{k^2} \right) dx + 2 \int_1^{\infty} \mathbb{P}(X_1 > x)x \left(\sum_{k: k \geq \lfloor x \rfloor} \frac{1}{k^2} \right) dx. \end{aligned}$$

Pour $j \geq 1$,

$$\begin{aligned} \sum_{k=j}^{\infty} \frac{1}{k^2} &= \frac{1}{j^2} + \sum_{k=j+1}^{\infty} \frac{1}{k^2} \leq \frac{1}{j} + \sum_{k=j+1}^{\infty} \frac{1}{k(k-1)} \\ &= \frac{1}{j} + \sum_{k=j+1}^{\infty} \left(\frac{1}{k-1} - \frac{1}{k} \right) = \frac{1}{j} + \frac{1}{j} = \frac{2}{j}. \end{aligned}$$

D'où :

$$\begin{aligned} \sum_{k=1}^{\infty} \frac{\text{Var}(Y_k)}{k^2} &\leq 4 \int_0^1 \mathbb{P}(X_1 > x)x dx + 4 \int_1^{\infty} \frac{\mathbb{P}(X_1 > x)x}{\lfloor x \rfloor} dx \\ &\leq 4 \int_0^1 \mathbb{P}(X_1 > x)x dx + 8 \int_1^{\infty} \mathbb{P}(X_1 > x) dx \\ &\leq 8 \int_0^{\infty} \mathbb{P}(X_1 > x) dx = 8\mathbb{E}(X_1). \quad \square \end{aligned}$$

Lemme 3.4. Fixons $\theta > 1$, et soit $n_j := \lfloor \theta^j \rfloor$. Alors $\frac{T(n_j)}{n_j} \rightarrow \mathbb{E}(X_1)$, p.s.

Preuve. Si $\varepsilon > 0$ et $A_j := \{|T(n_j) - \mathbb{E}(T(n_j))| > \varepsilon n_j\}$, alors par l'inégalité de Bienaymé–Tchebychev,

$$\mathbb{P}(A_j) \leq \frac{\text{Var}(T(n_j))}{\varepsilon^2 n_j^2} = \frac{\sum_{k=1}^{n_j} \text{Var}(Y_k)}{\varepsilon^2 n_j^2}.$$

D'après le théorème de Fubini,

$$\sum_{j=1}^{\infty} \mathbb{P}(A_j) = \sum_{k=1}^{\infty} \sum_{j \geq 1: n_j \geq k} \frac{\text{Var}(Y_k)}{\varepsilon^2 n_j^2}.$$

Puisque $n_j = \theta^j/2$, on a

$$\sum_{j \geq 1: n_j \geq k} \frac{1}{n_j^2} \leq 4 \sum_{j \in \mathbb{Z}, \theta^j \geq k} \frac{1}{\theta^{2j}} \leq \frac{4k^{-2}}{1 - \theta^{-2}},$$

car il s'agit d'une série géométrique dont le premier terme est majoré par k^{-2} . Par conséquent,

$$\sum_{j=1}^{\infty} \mathbb{P}(A_j) \leq \frac{4}{1-\theta^{-2}} \sum_{k=1}^{\infty} \frac{\text{Var}(Y_k)}{k^2} \leq \frac{32\mathbb{E}(X_1)}{1-\theta^{-2}} < \infty,$$

(voir le Lemme 3.3). D'après la première partie du lemme de Borel–Cantelli, il existe $B \in \mathcal{F}$ avec $\mathbb{P}(B) = 1$ tel que si $\omega \in B$, alors on puisse trouver $j_0 = j_0(\omega) < \infty$ tel que $|T(n_j) - \mathbb{E}(T_{n_j})| \leq \varepsilon n_j$ pour tout $j \geq j_0$. Donc $\limsup_{j \rightarrow \infty} \frac{|T(n_j) - \mathbb{E}(T_{n_j})|}{n_j} \leq \varepsilon$, p.s. Comme $\varepsilon > 0$ est quelconque, on a $\frac{T(n_j) - \mathbb{E}(T_{n_j})}{n_j} \rightarrow 0$, p.s. On peut conclure en utilisant le fait que $\frac{\mathbb{E}(T_{n_j})}{n_j} \rightarrow \mathbb{E}(X_1)$ (voir le Lemme 3.2). \square

Lemme 3.5. $\frac{T(n)}{n} \rightarrow \mathbb{E}(X_1)$, p.s.

Preuve. Pour chaque n , on peut trouver j tel que $n_{j-1} \leq n < n_j$. La suite $(T(n))_{n \geq 1}$ étant croissante (somme de variables positives), on a

$$\frac{T(n_{j-1})}{n_j} \leq \frac{T(n)}{n} \leq \frac{T(n_j)}{n_{j-1}}.$$

En faisant n (donc j) tendre vers $+\infty$, et à l'aide du fait que $\frac{n_j}{n_{j-1}} \rightarrow \theta$ ($j \rightarrow \infty$), on obtient, grâce au Lemme 3.4,

$$\mathbb{P}\left(\frac{\mathbb{E}(X_1)}{n} \leq \liminf_{n \rightarrow \infty} \frac{T(n)}{n} \leq \limsup_{n \rightarrow \infty} \frac{T(n)}{n} \leq \theta \mathbb{E}(X_1)\right) = 1.$$

La constante $\theta > 1$ étant quelconque, ceci implique que $\frac{T(n)}{n} \rightarrow \mathbb{E}(X_1)$, p.s. \square

La loi forte des grands nombres est une conséquence des Lemmes 3.1 et 3.5. On remarque que l'indépendance des v.a. $\{X_n\}_{n \geq 1}$ n'a pas vraiment été utilisée : elle peut être remplacée par l'indépendance deux-à-deux. \square

Chapitre 9. Convergence en loi

On introduit un nouveau concept de convergence pour une suite de v.a., qui ne dépend cette fois que de la loi de chacune de ces v.a.

1. Convergence en loi d'une suite de v.a.

Définition 1.1. On dit qu'une suite de v.a. X_1, \dots, X_n, \dots à valeurs dans \mathbb{R}^N (donc une suite de vecteurs aléatoires) converge en loi vers une v.a. X (à valeurs dans \mathbb{R}^N) si, pour toute fonction f continue bornée sur \mathbb{R}^N ,

$$\mathbb{E}(f(X_n)) \rightarrow \mathbb{E}(f(X)).$$

Remarque. (i) Les v.a. X_1, \dots, X_n, \dots ne sont pas nécessairement définies sur un même espace de probabilité.

(ii) Même si les v.a. X_1, \dots, X_n, \dots sont définies sur un même espace de probabilité, la convergence en loi a cette particularité qu'elle a encore lieu si l'on remplace chaque X_n par Y_n qui a la même loi que X_n (c'est-à-dire $P_{X_n} = P_{Y_n}$ quel que soit n). Ceci n'est pas vrai pour les autres convergences (p.s., en probabilité ou dans L^p). \square

Propriété 1.2. $\{X_n\}_{n \geq 1}$ converge en loi vers X si et seulement si pour toute fonction f continue à support compact,

$$(*) \quad \mathbb{E}(f(X_n)) \rightarrow \mathbb{E}(f(X)).$$

Preuve. Supposons que (*) est vérifiée par toute fonction continue à support compact. Soit $f \geq 0$ une fonction continue bornée et $\varepsilon > 0$ fixé. Donc $|f(x)| \leq C$ pour une constante $C > 0$ et tout $x \in \mathbb{R}^N$.

Il existe une fonction φ continue à support compact avec $\varphi \leq 1$ et $\mathbb{E}(\varphi(X)) > 1 - \frac{\varepsilon}{5C}$. On déduit qu'il existe donc un entier n_0 tel que $\mathbb{E}(\varphi(X_n)) > 1 - \frac{\varepsilon}{4C}$ dès que $n \geq n_0$. D'autre part, la fonction $f\varphi$ est continue et à support compact; on sait donc trouver un entier n_1 tel que

$$|\mathbb{E}(\varphi(X_n)f(X_n)) - \mathbb{E}(\varphi(X)f(X))| < \varepsilon/2 \quad \text{dès que } n \geq n_1.$$

On a alors pour tout $n \geq n_0 + n_1$,

$$\begin{aligned} |\mathbb{E}(f(X_n)) - \mathbb{E}(f(X))| &\leq |\mathbb{E}(\varphi(X)f(X)) - \mathbb{E}(\varphi(X_n)f(X_n))| \\ &\quad + |\mathbb{E}[(1 - \varphi(X))f(X)]| + |\mathbb{E}[(1 - \varphi(X_n))f(X_n)]| \\ &\leq \frac{\varepsilon}{2} + C \mathbb{E}(1 - \varphi(X)) + C \mathbb{E}(1 - \varphi(X_n)) \\ &\leq \frac{\varepsilon}{2} + \frac{\varepsilon}{5} + \frac{\varepsilon}{4} < \varepsilon. \end{aligned}$$

Ceci montre la convergence en loi. □

Comparons maintenant la convergence en loi avec les autres types de convergence pour des suites de v.a.

Propriété 1.3. Si X_1, \dots, X_n, \dots est une suite de v.a. réelles qui converge en probabilité vers une v.a. X (en particulier, si la convergence a lieu dans $L^p(\Omega, \mathcal{F}, \mathbb{P})$, ou p.s.), alors X_n converge également vers X en loi.

Preuve. On sait que de toute sous-suite, on peut extraire une sous-sous-suite qui converge p.s. vers X . En appliquant le théorème de convergence dominée ($|f(X_n)|$ reste dominé par $\sup_{\mathbb{R}} |f(x)| < \infty$), on voit que $\mathbb{E}(f(X_n)) \rightarrow \mathbb{E}(f(X))$ le long de cette sous-sous-suite. Comme ce résultat est valable pour toute suite extraite, on a bien $\lim_{n \rightarrow \infty} \mathbb{E}(f(X_n)) = \mathbb{E}(f(X))$. □

La réciproque est bien sûr fautive (si X_1, \dots, X_n, \dots est une suite de v.a. i.i.d., par exemple de Bernoulli, il est facile de vérifier que X_n ne converge pas en probabilités, puisque le critère de Cauchy ne peut pas être vérifié). Cependant, on a une réciproque très partielle.

Propriété 1.4. Si une suite de v.a. réelles $\{X_n\}_{n \geq 1}$ définies sur un même espace probabilisé converge en loi vers X qui est une constante p.s., alors la convergence a également lieu en probabilité.

Preuve. Soit $X_n \rightarrow c$ en loi (avec $c \in \mathbb{R}$). Soit $\varepsilon > 0$. On peut trouver une fonction φ continue sur \mathbb{R} telle que $\varphi(c) = 1$ et que $\varphi(x) \leq \mathbf{1}_{\{|x-c| \leq \varepsilon\}}$. Donc

$$\mathbb{P}(|X_n - c| \leq \varepsilon) \geq \mathbb{E}(\varphi(X_n)) \rightarrow \mathbb{E}(\varphi(c)) = 1.$$

Par conséquent, $X_n \rightarrow c$ en probabilité. \square

2. Cas des v.a. à valeurs entières

Étudions la convergence en loi pour des v.a. entières, où les résultats sont très simples. Rappelons que pour toute v.a. X à valeurs dans \mathbb{N} , la fonction génératrice G_X de X est la fonction sur $[0, 1]$ définie par $G_X(s) := \mathbb{E}(s^X) = \sum_{n=0}^{\infty} s^n \mathbb{P}(X = n)$.

Propriété 2.1. Soient X_1, \dots, X_n, \dots des v.a. à valeurs dans \mathbb{N} . Les assertions suivantes sont équivalentes :

- (i) X_n converge vers X en loi.
- (ii) Pour chaque $k \in \mathbb{N}$, on a $\lim_{n \rightarrow \infty} \mathbb{P}(X_n = k) = \mathbb{P}(X = k)$.
- (iii) G_{X_n} converge simplement vers G_X sur $[0, 1]$.

Preuve. (i) \Rightarrow (ii) : Prendre pour f une fonction continue bornée qui vaut 1 en k et 0 pour les autres entiers.

(ii) \Rightarrow (i) : Soit f une fonction continue à support compact, disons dans $[0, M]$ (où $M \in \mathbb{N}$). On a

$$\mathbb{E}(f(X_n)) = \sum_{k=0}^M f(k) \mathbb{P}(X_n = k) \rightarrow \sum_{k=0}^M f(k) \mathbb{P}(X = k) = \mathbb{E}(f(X)).$$

D'après la Propriété 1.2, on a la convergence en loi.

- (ii) \Rightarrow (iii) : Même argument que ci-dessus en ajoutant la convergence dominée.
- (iii) \Rightarrow (ii) : Admise. \square

3. Convergence en loi et fonctions de répartition

On va ensuite étudier la convergence en loi des v.a. réelles à l'aide de leurs fonctions de répartition F_{X_n} . On dira qu'un réel x est un point de continuité d'une fonction de répartition F_X si $F_X(x) = F_X(x-)$, et que c'est un point de discontinuité sinon. Autrement

dit, x est un point de discontinuité pour F_X si et seulement si la loi de X , P_X , a un atome en x .

Théorème 3.1. Soient X_1, \dots, X_n, \dots une suite de v.a. réelles. Soit F_X la fonction de répartition d'une v.a. réelle X . Les assertions suivantes sont équivalentes :

- (1) X_n converge en loi vers X .
- (2) Si x est un point de continuité de F_X , alors $\lim_{n \rightarrow \infty} F_{X_n}(x) = F_X(x)$.

Preuve. La preuve se fait en deux étapes distinctes.

Première étape : montrons que (1) \Rightarrow (2). Pour tout $\varepsilon > 0$, il existe une fonction $\varphi_\varepsilon : \mathbb{R} \rightarrow [0, 1]$ continue telle que $\varphi_\varepsilon(t) = 1$ si $t \leq x$ et $\varphi_\varepsilon(t) = 0$ si $t \geq x + \varepsilon$. On a alors d'une part

$$\lim_{n \rightarrow \infty} \mathbb{E}(\varphi_\varepsilon(X_n)) = \mathbb{E}(\varphi_\varepsilon(X)) \leq \mathbb{P}(X \leq x + \varepsilon) = F_X(x + \varepsilon).$$

Or $\mathbb{E}(\varphi_\varepsilon(X_n)) \geq \mathbb{P}(X_n \leq x) = F_{X_n}(x)$, et donc $\limsup_{n \rightarrow \infty} F_{X_n}(x) \leq F_X(x + \varepsilon)$. Comme $\varepsilon > 0$ peut être arbitrairement petit et que F_X est continue à droite, on déduit que

$$\limsup_{n \rightarrow \infty} F_{X_n}(x) \leq F_X(x).$$

Pour établir l'inégalité réciproque, on prend une fonction continue $\psi_\varepsilon : \mathbb{R} \rightarrow [0, 1]$ telle que $\psi_\varepsilon(t) = 1$ si $t \leq x - \varepsilon$ et $\psi_\varepsilon(t) = 0$ si $t \geq x$. Comme précédemment, on tire cette fois dans un premier temps

$$\lim_{n \rightarrow \infty} \mathbb{E}(\psi_\varepsilon(X_n)) = \mathbb{E}(\psi_\varepsilon(X)) \geq \mathbb{P}(X \leq x - \varepsilon) = F_X(x - \varepsilon),$$

puis $\liminf_{n \rightarrow \infty} F_{X_n}(x) \geq F_X(x - \varepsilon)$. Comme, par hypothèse, $F_X(x-) = F_X(x)$ et $\varepsilon > 0$ peut être pris arbitrairement petit, on a finalement $\liminf_{n \rightarrow \infty} F_{X_n}(x) \geq F_X(x)$. D'où $F_{X_n}(x) \rightarrow F_X(x)$.

Deuxième étape : (2) \Rightarrow (1). Supposons maintenant que $\lim_{n \rightarrow \infty} F_{X_n}(x) = F_X(x)$ pour tout point de continuité x de F_X .

Admettons pour l'instant qu'il existe des variables aléatoires réelles $(Y_n)_{n \geq 1}$ et Y telles que $F_{X_n} = F_{Y_n}$ (pour tout $n \geq 1$), $F_X = F_Y$ et que $Y_n \rightarrow Y$ p.s. (Ceci est un cas spécial de ce qui porte le nom de la "**construction de Skorokhod**").

Puisque $Y_n \rightarrow Y$ p.s., par convergence dominée, on a, pour toute fonction f continue et bornée, on a $\mathbb{E}(f(Y_n)) \rightarrow \mathbb{E}(f(Y))$ ($n \rightarrow \infty$). Or, $\mathbb{E}(f(Y_n)) = \mathbb{E}(f(X_n))$ (pour tout n), et $\mathbb{E}(f(Y)) = \mathbb{E}(f(X))$, on obtient la convergence en loi de (X_n) vers X .

Il reste donc de montrer l'existence des v.a. $(Y_n)_{n \geq 1}$ et Y vérifiant les propriétés ci-dessus. [Cette preuve ne fait pas partie du programme de l'examen.] Soient $\Omega =]0, 1[$,

\mathcal{F} la tribu borélienne de Ω , et soit \mathbb{P} la mesure de Lebesgue sur Ω (attention : les v.a. “invisibles” $(X_n)_{n \geq 1}$ et X n’ont rien à voir avec notre espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$). Définissons $Y_n(\omega) := \inf\{y \in \mathbb{R} : F_{X_n}(y) > \omega\}$, $Y(\omega) := \inf\{y \in \mathbb{R} : F_X(y) > \omega\}$, et $U(\omega) := \omega$, $\omega \in \Omega$. Il est clair que U suit la loi uniforme sur $[0, 1]$ (on peut le vérifier par exemple en calculant sa fonction de répartition.) D’après la Proposition 5.1 du Chapitre 4, $F_{Y_n} = F_{X_n}$, pour chaque n , et $F_Y = F_X$. Il suffit alors de montrer que $Y_n(\omega) \rightarrow Y(\omega)$, en tout $\omega \in \Omega$ sauf sur un ensemble au plus infini dénombrable.

Commençons par identifier l’ensemble exceptionnel. Soient pour tout $\omega \in \Omega$, $a(\omega) := \inf\{y \in \mathbb{R} : F_X(y) \geq \omega\}$ et $b(\omega) := \inf\{y \in \mathbb{R} : F_X(y) > \omega\} = Y(\omega)$. Il est clair que $a(\omega) \leq b(\omega)$ et $b(\omega) \leq a(\omega')$, quels que soient $\omega \in \Omega$ et $\omega' \in \Omega$ avec $\omega' > \omega$. Posons $\Omega_0 := \{\omega \in \Omega : a(\omega) = b(\omega)\}$. L’ensemble $\Omega - \Omega_0$ est au plus infini dénombrable, car $\{]a(\omega), b(\omega)[; \omega \in \Omega - \Omega_0\}$ est une famille d’intervalles ouverts disjoints dont chacun contient un rationnel différent.

On montre maintenant que $Y_n(\omega) \rightarrow Y(\omega)$ pour tout $\omega \in \Omega_0$.

Fixons $\omega \in \Omega_0$. Soit $z < Y(\omega) = b(\omega)$ tel que F_X soit continue en z . Puisque $z < a(\omega) = \inf\{y \in \mathbb{R} : F_X(y) \geq \omega\}$, on a $F_X(z) < \omega$. Par hypothèse, $F_{X_n}(z) \rightarrow F_X(z)$ ($n \rightarrow \infty$), on a $F_{X_n}(z) < \omega$ lorsque n est suffisamment grand, disons $n \geq n_0$. Donc $z \leq Y_n(\omega)$, $n \geq n_0$. D’où $\liminf_{n \rightarrow \infty} Y_n(\omega) \geq z$. Comme z peut être aussi proche de $Y(\omega)$ que l’on veut, on obtient $\liminf_{n \rightarrow \infty} Y_n(\omega) \geq Y(\omega)$.

Réciproquement, fixons $\omega \in \Omega_0$, et soit $z > Y(\omega) = \inf\{y \in \mathbb{R} : F_X(y) > \omega\}$ tel que F_X soit continue en z . On a $F_X(z) > \omega$. Puisque $F_{X_n}(z) \rightarrow F_X(z)$, on a $F_{X_n}(z) > \omega$ si $n \geq n_0$, ce qui nous donne $Y_n(\omega) \leq z$ ($n \geq n_0$), d’où $\limsup_{n \rightarrow \infty} Y_n(\omega) \leq z$. Faisant z tendre vers $Y(\omega)$ le long d’une suite de points de continuité de F_X , on obtient $\limsup_{n \rightarrow \infty} Y_n(\omega) \leq Y(\omega)$. Par conséquent, $Y_n(\omega) \rightarrow Y(\omega)$ pour tout $\omega \in \Omega_0$. \square

La construction de Skorokhod nous confirme que $X_n \rightarrow X$ en loi équivaut à l’existence des variables aléatoires réelles $(Y_n)_{n \geq 1}$ et Y telles que $P_{X_n} = P_{Y_n}$ (pour tout $n \geq 1$), $P_X = P_Y$ et que $Y_n \rightarrow Y$ p.s. Ceci aide à affaiblir les conditions dans certains résultats. Voici un exemple.

Proposition 3.2 (lemme de Fatou). Si $X_n \geq 0$ (pour tout $n \geq 1$) et si $X_n \rightarrow X$ en loi, alors $\liminf_{n \rightarrow \infty} \mathbb{E}(X_n) \geq \mathbb{E}(X)$.

Preuve. Quitte à changer l’espace, il existe des variables aléatoires réelles $(Y_n)_{n \geq 1}$ et Y telles que $P_{X_n} = P_{Y_n}$ (pour tout $n \geq 1$), $P_X = P_Y$ et que $Y_n \rightarrow Y$ p.s. Comme $Y_n \geq 0$, le lemme de Fatou usuel nous donne $\liminf_{n \rightarrow \infty} \mathbb{E}(Y_n) \geq \mathbb{E}(Y)$, ce qui signifie alors $\liminf_{n \rightarrow \infty} \mathbb{E}(X_n) \geq \mathbb{E}(X)$. \square

Exemple 3.3. Le Théorème 3.1 nous dit que si F_{X_n} converge simplement vers F_X (fonction de répartition de la v.a. X), alors X_n converge en loi vers X . En général, si F_{X_n} converge simplement vers une fonction disons F , F n'est pas nécessairement la fonction de répartition associée à une loi de probabilité sur \mathbb{R} . Par exemple, soit pour $n \geq 1$,

$$F_{X_n}(x) = \begin{cases} 0 & \text{si } x \leq -n, \\ (x+n)/(2n), & \text{si } -n < x \leq n, \\ 1 & \text{si } x > n. \end{cases}$$

F_{X_n} est une suite de fonctions continues croissantes sur \mathbb{R} , avec $\lim_{x \rightarrow -\infty} F_{X_n}(x) = 0$ et $\lim_{x \rightarrow \infty} F_{X_n}(x) = 1$, donc une suite de fonctions de répartition (en fait, F_{X_n} est la fonction de répartition de la loi uniforme sur l'intervalle $[-n, n]$). Pourtant, F_{X_n} converge simplement vers $1/2$, qui n'est pas une fonction de répartition. \square

Exemple 3.4. Il est à remarquer que la convergence en loi de X_n vers X n'implique pas nécessairement $F_n(x) \rightarrow F(x)$ en tout $x \in \mathbb{R}$. Par exemple, soit $X_n = X + \frac{1}{n}$. On a $X_n \rightarrow X$ p.s., donc on a a fortiori la convergence en loi. Pourtant, $F_{X_n}(x) = \mathbb{P}(X_n \leq x) = \mathbb{P}(X \leq x - \frac{1}{n}) = F_X(x - \frac{1}{n}) \rightarrow F_X(x-)$. Donc on a $F_n(x) \rightarrow F(x)$ seulement si x est un point de continuité de F_X . \square

Théorème 3.5 (théorème de Slutsky). Soient $\{X_n\}_{n \geq 1}$ et $\{Y_n\}_{n \geq 1}$ deux suites de v.a. réelles définies sur un même espace de probabilité. Si X_n converge en loi vers X , et si Y_n converge en loi vers $c \in \mathbb{R}$, alors

- (1) $X_n + Y_n \rightarrow X + c$ en loi;
- (2) $X_n Y_n \rightarrow c X$ en loi;
- (3) $\frac{X_n}{Y_n} \rightarrow \frac{X}{c}$ en loi, si $c \neq 0$.

Preuve. On vérifie seulement (1), la preuve de (2) et (3) étant tout à fait dans le même esprit.

Fixons $x \in \mathbb{R}$ un point de continuité de F_{X+c} (ce qui revient à dire que $x - c$ est un point de continuité de F_X). On peut trouver une suite $\{\varepsilon_k\}_{k \geq 0}$ de positifs qui décroissent vers 0 tel que $x - c + \varepsilon_k$ et $x - c - \varepsilon_k$ soient tous des points de continuité de F_X (ce qui est possible car l'ensemble des points de discontinuité d'une fonction monotone est dénombrable). On a, pour chaque $k \geq 1$,

$$\begin{aligned} F_{X_n+Y_n}(x) &= \mathbb{P}(X_n + Y_n \leq x) \\ &\leq \mathbb{P}(X_n + Y_n \leq x, |Y_n - c| \leq \varepsilon_k) + \mathbb{P}(|Y_n - c| > \varepsilon_k) \\ &\leq \mathbb{P}(X_n \leq x - c + \varepsilon_k) + \mathbb{P}(|Y_n - c| > \varepsilon_k). \end{aligned}$$

En faisant n tendre vers l'infini, ceci donne

$$\limsup_{n \rightarrow \infty} F_{X_n + Y_n}(x) \leq \limsup_{n \rightarrow \infty} F_{X_n}(x - c + \varepsilon_k) = F_X(x - c + \varepsilon_k).$$

Comme ε_k peut être arbitrairement proche de 0, on obtient,

$$\limsup_{n \rightarrow \infty} F_{X_n + Y_n}(x) \leq F_X(x - c).$$

De même, on a, pour tout $k \geq 1$,

$$\mathbb{P}(X_n \leq x - c - \varepsilon_k) \leq \mathbb{P}(X_n + Y_n \leq x) + \mathbb{P}(|Y_n - c| > \varepsilon_k),$$

et donc

$$F_X(x - c - \varepsilon_k) \leq \liminf_{n \rightarrow \infty} F_{X_n + Y_n}(x), \quad \forall k \geq 1,$$

ce qui implique que

$$F_X(x - c) \leq \liminf_{n \rightarrow \infty} F_{X_n + Y_n}(x).$$

Donc

$$\lim_{n \rightarrow \infty} F_{X_n + Y_n}(x) = F_X(x - c) = F_{X+c}(x),$$

pour tout x point de continuité de F_{X+c} . □

4. Convergence en loi et fonctions caractéristiques

Pour conclure ce chapitre, on présente un résultat extrêmement utile pour savoir si une suite de v.a. converge en loi, et déterminer sa limite. On rappelle qu'on note φ_Y la fonction caractéristique d'une v.a. Y .

Théorème 4.1. On considère comme à l'accoutumée une suite de v.a. $X, X_1, \dots, X_n, \dots$ à valeurs dans \mathbb{R}^N . Les conditions suivantes sont équivalentes :

- (1) X_n converge en loi vers X .
- (2) φ_{X_n} converge simplement vers φ_X .

Preuve. (1) \Rightarrow (2) : Pour tout $t \in \mathbb{R}^N$, la fonction $x \rightarrow e^{i\langle t, x \rangle}$ est continue et bornée. On a donc

$$\varphi_{X_n}(t) = \mathbb{E} \left(e^{i\langle t, X_n \rangle} \right) \rightarrow \mathbb{E} \left(e^{i\langle t, X \rangle} \right) = \varphi_X(t).$$

(2) \Rightarrow (1) : Pour simplifier, nous allons supposer que la dimension est $N = 1$; le cas général est similaire, mais avec des notations plus lourdes. Supposons tout d'abord que

f est une fonction continue à support compact, dont la transformée de Fourier $\widehat{f}(t) = \int_{\mathbb{R}} e^{-2\pi itx} f(x) dx$ est dans $L^1(\mathbb{R})$. On sait par Fourier inverse que

$$f(x) = \int_{\mathbb{R}} e^{2\pi itx} \widehat{f}(t) dt.$$

On a donc en appliquant le théorème de Fubini (justifié)

$$\begin{aligned} \mathbb{E}(f(X_n)) &= \mathbb{E}\left(\int_{\mathbb{R}} e^{2\pi itX_n} \widehat{f}(t) dt\right) \\ &= \int_{\mathbb{R}} \mathbb{E}(e^{2\pi itX_n}) \widehat{f}(t) dt = \int_{\mathbb{R}} \varphi_{X_n}(2\pi t) \widehat{f}(t) dt. \end{aligned}$$

Par hypothèse, on sait que $\varphi_{X_n}(2\pi t)$ converge vers $\varphi_X(2\pi t)$ pour chaque t . D'autre part, $|\varphi_{X_n}(2\pi t)| \leq 1$ pour tout $t \in \mathbb{R}$. Comme \widehat{f} est intégrable, le théorème de convergence dominée s'applique, et on tire

$$(*) \quad \lim_{n \rightarrow \infty} \mathbb{E}(f(X_n)) = \int_{\mathbb{R}} \varphi_X(2\pi t) \widehat{f}(t) dt = \mathbb{E}(f(X)).$$

Ensuite on remarque que la condition $\widehat{f} \in L^1(\mathbb{R})$ est satisfaite pour toute fonction f de classe \mathcal{C}^2 à support compact. En effet, la transformée de Fourier de f'' est $t^2 \widehat{f}(t)$ (intégration par parties), et on sait que c'est une fonction bornée. On a donc $\widehat{f}(t) = O(t^{-2})$ à l'infini, et comme \widehat{f} est elle aussi bornée, c'est bien une fonction intégrable. Ainsi, on sait que (*) est vérifiée pour toute fonction f de classe \mathcal{C}^2 à support compact. L'extension de (*) pour toute fonction continue à support compact est omise. \square

Exemple 4.2. Supposons que pour chaque $n \geq 1$, X_n et Y_n sont indépendantes, et que X et Y sont indépendantes. Si X_n converge en loi vers X , et si Y_n converge en loi vers Y , alors $X_n + Y_n$ converge en loi vers $X + Y$.

En effet, par le Théorème 4.1, pour tout $t \in \mathbb{R}$, $\varphi_{X_n}(t) \rightarrow \varphi_X(t)$ et $\varphi_{Y_n}(t) \rightarrow \varphi_Y(t)$. Grâce à l'indépendance, $\varphi_{X_n+Y_n}(t) = \varphi_{X_n}(t)\varphi_{Y_n}(t) \rightarrow \varphi_X(t)\varphi_Y(t) = \varphi_{X+Y}(t)$. Donc $X_n + Y_n$ converge en loi vers $X + Y$. \square

Exemple 4.3. Supposons que pour chaque $n \geq 1$, $P_{X_n} = \mathcal{N}(\mu_n, \sigma_n^2)$. Si $\mu_n \rightarrow \mu$, et $\sigma_n^2 \rightarrow \sigma^2$, alors X_n converge en loi vers $\mathcal{N}(\mu, \sigma^2)$.

En effet, $\varphi_{X_n}(t) = \exp(i\mu_n t - \sigma_n^2 t^2/2) \rightarrow \exp(i\mu t - \sigma^2 t^2/2)$, qui est la fonction caractéristique en t de la loi $\mathcal{N}(\mu, \sigma^2)$. Par le Théorème 4.1, X_n converge en loi vers $\mathcal{N}(\mu, \sigma^2)$.

\square

Exemple 4.4. Soient W_1, W_2, \dots des variables aléatoires indépendantes suivant la même loi uniforme sur $[-1, 1]$. On cherche à étudier la convergence en loi de $\frac{1}{n} \sum_{j=1}^n \frac{1}{W_j}$. (Rappel : $\int_0^\infty \frac{1-\cos(y)}{y^2} dy = \frac{\pi}{2}$.)

On remarque que la loi forte des grands nombres appliquée à la suite de variables aléatoires iid $(\frac{1}{W_j})_{j \geq 1}$ ne nous donne rien d'intéressant, car $\frac{1}{W_j}$ n'admet pas de moment d'ordre 1. On essaie donc d'appliquer le Théorème 4.1.

La fonction caractéristique de $1/W_1$ vaut

$$\varphi_{1/W_1}(t) = \frac{1}{2} \int_0^1 (e^{it/x} + e^{-it/x}) dx = \int_0^1 \cos(t/x) dx = |t| \int_{|t|}^\infty \frac{\cos(y)}{y^2} dy,$$

avec un changement de variables $x = |t|/y$. D'après le rappel, $\int_0^\infty \frac{1-\cos(y)}{y^2} dy = \frac{\pi}{2}$. Donc on peut écrire que quand $t \rightarrow 0$,

$$\begin{aligned} \varphi_{1/W_1}(t) &= 1 - |t| \int_{|t|}^\infty \frac{1 - \cos(y)}{y^2} dy \\ &= 1 - \frac{\pi |t|}{2} + |t| \int_0^{|t|} \frac{1 - \cos(y)}{y^2} dy \\ &= 1 - \frac{\pi |t|}{2} + o(|t|). \end{aligned}$$

Soit $T_n = \frac{1}{n} \sum_{j=1}^n \frac{1}{W_j}$. Alors pour tout $t \in \mathbb{R}$ fixé,

$$\varphi_{T_n}(t) = [\varphi_{1/W_1}(t/n)]^n = \left[1 - \frac{\pi |t|}{2n} + o(n^{-1}) \right]^n \rightarrow e^{-\pi |t|/2}.$$

Donc T_n converge en loi vers $\pi C/2$, où C est une variable aléatoire suivant la loi de Cauchy standard. \square

Une difficulté possible quand on cherche à appliquer le Théorème 4.1, est que l'on doit savoir a priori que la limite des fonctions caractéristiques φ_{X_n} est elle aussi une fonction caractéristique. Cela n'a rien d'automatique.

Exemple 4.5. Soit $P_{X_n} = \mathcal{N}(0, n)$, alors $\varphi_{X_n}(t) = \exp(-nt^2/2) \rightarrow 0$ pour $t \neq 0$ et $\varphi_{X_n}(0) = 1$ pour tout $n \geq 1$.

Pourtant il n'y a pas de convergence en loi pour P_{X_n} . En effet, la fonction indicatrice du singleton 0 (qui est la limite simple de φ_{X_n}) ne peut être une fonction caractéristique associée à une loi de probabilité sur \mathbb{R} car elle n'est pas continue sur \mathbb{R} . \square

Même si le théorème de Bochner énoncé dans le Chapitre 6 donne une condition nécessaire et suffisante pour qu'une fonction soit la fonction caractéristique d'une loi de

probabilité, cette condition n'est pas toujours facile à vérifier. On admet le critère suivant très simple qui porte souvent le nom de Paul Lévy :

Théorème 4.6. Si pour chaque $n \geq 1$, φ_n est la fonction caractéristique d'une loi de probabilité sur \mathbb{R}^N , et si φ_n converge simplement vers une fonction φ qui est continue en 0, alors φ est la fonction caractéristique d'une loi de probabilité sur \mathbb{R}^N .

Autrement dit, si l'on sait que la suite des fonctions caractéristiques φ_{X_n} converge simplement vers une fonction φ qui est continue en 0, alors X_n converge en loi vers une v.a. X dont la fonction caractéristique est φ .

Exemple 4.7. Soit $\{U_n\}_{n \geq 1}$ une suite i.i.d. de v.a. telle que $\mathbb{P}(U_1 = 1) = \mathbb{P}(U_1 = -1) = 1/2$. Soit

$$X_n := \sum_{j=1}^n \frac{U_j}{2^j}.$$

On a, pour tout $t \in \mathbb{R}$, $\varphi_{U_j}(t) = \mathbb{E}(e^{itU_j}) = (e^{it} + e^{-it})/2 = \cos(t)$. Donc

$$\varphi_{X_n}(t) = \prod_{j=1}^n \varphi_{U_j} \left(\frac{t}{2^j} \right) = \prod_{j=1}^n \cos \left(\frac{t}{2^j} \right) = \frac{\sin(t)}{2^n \sin(t/2^n)} \rightarrow \frac{\sin(t)}{t},$$

qui est une fonction continue en 0, et est donc d'après le Théorème 4.6 la fonction caractéristique d'une loi de probabilité. En fait, il s'agit de la fonction caractéristique de la loi uniforme sur $(-1, 1)$: soit X une v.a. suivant la loi uniforme sur $(-1, 1)$, alors $\varphi_X(t) = \frac{1}{2} \int_{-1}^1 e^{itx} dx = (e^{it} - e^{-it})/(2it) = \frac{\sin(t)}{t}$.

Par conséquent, X_n converge en loi vers X . □

Chapitre 10. Théorème central limite

1. Retour sur la loi faible des grands nombres

Nous allons tout d'abord établir la loi faible des grands nombres (i.e., pour la convergence en probabilité, et non pas pour la convergence presque sûre) sous la seule hypothèse de finitude du moment d'ordre 1. Rappelons ce dont il s'agit :

Théorème 1.1. (Loi faible des grands nombres). On se donne X_1, \dots, X_n, \dots une suite de v.a. réelles i.i.d., avec $\mathbb{E}(|X_1|) < \infty$; on note $S_n = X_1 + \dots + X_n$ la somme partielle au rang n . Alors S_n/n converge en probabilité vers la moyenne $\mathbb{E}(X_1) \in \mathbb{R}$.

Preuve. Établissons d'abord la stratégie. Pour cela, on sait qu'il suffit de montrer la convergence en loi (puisque la limite est une constante). A cette fin, on va calculer la fonction caractéristique de S_n/n et vérifier qu'elle converge bien vers la fonction caractéristique de la constante $\mathbb{E}(X_1)$.

Notons $\varphi_X(t) = \mathbb{E}(e^{itX})$ ($t \in \mathbb{R}$) la fonction caractéristique de X . Comme les v.a. X_1, \dots, X_n sont indépendantes et toutes de même loi que X , la fonction caractéristique de S_n est $\varphi_{S_n} = \varphi_X^n$. On a donc

$$\varphi_{S_n/n}(t) = \mathbb{E}\left(e^{itS_n/n}\right) = \varphi_{S_n}(t/n) = \varphi_X(t/n)^n.$$

On veut étudier le comportement du terme de droite quand $n \rightarrow \infty$. Pour cela, on le réécrit comme $(1 - (1 - \varphi_X(t/n)))^n$, et on se souvient (voir la Proposition 3.1 du Chap. 6) de ce que

$$\varphi_X(0) = 1 \quad , \quad \varphi'_X(0) = i\mathbb{E}(X).$$

On sait donc que $1 - \varphi_X(t/n) \sim -it\mathbb{E}(X)/n$ quand $n \rightarrow \infty$. Il en découle que

$$\ln \varphi_{S_n/n}(t) = n \ln(1 - (1 - \varphi_X(t/n))) \sim it\mathbb{E}(X).$$

En prenant l'exponentielle, on tire finalement

$$\varphi_{S_n/n}(t) \sim e^{it\mathbb{E}(X)}.$$

Or le terme de droite est la fonction caractéristique de la v.a. constante $\mathbb{E}(X)$; c'est précisément ce qu'on cherchait à établir. \square

C'est cette même idée qui va servir à établir un résultat du second ordre pour la moyenne de Césaro, le théorème central limite.

2. Théorème central limite

Le théorème central limite est l'un des résultats les plus importants de la théorie des probabilités. De façon informelle, ce théorème donne une estimation très précise de l'erreur qu'on commet en approchant la moyenne mathématique par la moyenne empirique (i.e. la moyenne de Césaro).

Il a d'abord été observé par Gauss, qui l'appelait la loi des erreurs; mais ce dernier n'en a pas donné de démonstration rigoureuse. La preuve en a été apportée par Moivre et Laplace, le théorème porte parfois leurs noms. La dénomination actuelle est apparue vers 1950 (en anglais : central limit theorem, ce qu'on a parfois traduit à tort par "le théorème de la limite centrale").

Théorème 2.1 (Théorème Central Limite). Soient X_1, \dots, X_n, \dots une suite de v.a. réelles i.i.d., avec $\mathbb{E}(X^2) < \infty$. On note $S_n = X_1 + \dots + X_n$ la somme partielle au rang n , et $\sigma^2 = \text{Var}(X)$. Alors

$$\lim_{n \rightarrow \infty} \frac{S_n - n\mathbb{E}(X)}{\sqrt{n}} = \mathcal{N}(0, \sigma^2) \quad \text{en loi,}$$

où l'on a noté $\mathcal{N}(0, \sigma^2)$ la loi de Gauss centrée et de variance σ^2 , i.e. la loi sur \mathbb{R} :

$$\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{x^2}{2\sigma^2}\right) dx.$$

Remarque. Sous les conditions du théorème, pour tous réels $a < b$, lorsque n tend vers l'infini,

$$\mathbb{P}\left(\frac{S_n - n\mathbb{E}(X)}{\sigma\sqrt{n}} < a\right) \rightarrow \Phi(a),$$

$$\begin{aligned} \mathbb{P}\left(\frac{S_n - n\mathbb{E}(X)}{\sigma\sqrt{n}} \leq a\right) &\rightarrow \Phi(a), \\ \mathbb{P}\left(a < \frac{S_n - n\mathbb{E}(X)}{\sigma\sqrt{n}} < b\right) &\rightarrow \Phi(b) - \Phi(a), \\ \mathbb{P}\left(a \leq \frac{S_n - n\mathbb{E}(X)}{\sigma\sqrt{n}} \leq b\right) &\rightarrow \Phi(b) - \Phi(a). \end{aligned} \quad \square$$

Preuve du Théorème 2.1. On va adopter la même approche que dans la preuve de la loi faible des grands nombres, mais au second ordre au lieu du premier ordre. Quitte à remplacer X par $X - \mathbb{E}(X)$, on peut supposer que $\mathbb{E}(X) = 0$.

Rappelons que la fonction caractéristique de S_n/\sqrt{n} est donnée par

$$\varphi_{S_n/\sqrt{n}}(t) = \varphi_X(t/\sqrt{n})^n.$$

D'autre part, comme X a un moment d'ordre 2, sa fonction caractéristique est de classe \mathcal{C}^2 , et son développement de Taylor à l'origine est donné par

$$\varphi_X(t) = 1 + t\varphi'_X(0) + \frac{t^2}{2}\varphi''_X(0) + o(t^2), \quad t \rightarrow 0.$$

Le fait que $\mathbb{E}(X) = 0$ et $\mathbb{E}(X^2) = \sigma^2$ entraîne que $\varphi'_X(0) = 0$ et $\varphi''_X(0) = -\sigma^2$. Autrement dit, on a

$$\varphi_X(t) = 1 - \frac{\sigma^2 t^2}{2} + o(t^2), \quad t \rightarrow 0.$$

Il en découle que pour chaque $t \in \mathbb{R}$ fixé

$$n \ln \varphi_X(t/\sqrt{n}) \sim -n \left(\frac{\sigma^2 t^2}{2n} \right) = -\frac{\sigma^2 t^2}{2}, \quad n \rightarrow \infty,$$

et donc, en passant à l'exponentielle

$$\varphi_{S_n/\sqrt{n}}(t) \sim \exp\left\{-\frac{\sigma^2 t^2}{2}\right\}.$$

Le terme de droite est la fonction caractéristique de la loi $\mathcal{N}(0, \sigma^2)$, et le théorème est établi. □

Exemple 2.2. On fait une suite indépendante de jets d'une pièce de monnaie parfaite. On note

$$X_k = \begin{cases} 1 & \text{si le } k\text{-ième jet est "Pile",} \\ 0 & \text{sinon.} \end{cases}$$

Dans ce modèle simple, $\{X_k\}_{k \geq 1}$ est une suite iid de v.a. de Bernoulli (de paramètre $1/2$), avec

$$\mathbb{E}(X_k) = \frac{1}{2}, \quad \text{Var}(X_k) = \frac{1}{4}.$$

Par la loi forte des grands nombres, S_n/n converge presque sûrement vers $1/2$, alors que le théorème central limite confirme que

$$\frac{S_n - n/2}{\sqrt{n/4}}$$

converge en loi vers une gaussienne centrée réduite. Autrement dit, lorsque n est suffisamment grand,

$$\mathbb{P}\left(|S_n - \frac{n}{2}| \leq \alpha \sqrt{\frac{n}{4}}\right) \approx \mathbb{P}(|\mathcal{N}| \leq \alpha) = \Phi(\alpha) - \Phi(-\alpha) = 2\Phi(\alpha) - 1.$$

Regardons l'exemple de $n = 10,000$ et $\alpha = 1,96$. On a $\Phi(\alpha) \approx 0,975$ et donc $2\Phi(\alpha) - 1 \approx 0,95$. On a aussi $\alpha \sqrt{n/4} = 98$. L'identité ci-dessus nous dit que

$$\mathbb{P}(4,902 \leq S_n \leq 5,098) \approx 0,95. \quad \square$$

Exemple 2.3. On fait une suite indépendante d'essais de dés, et on note X_k le nombre obtenu au k -ième essai. On a

$$\begin{aligned} \mathbb{E}(X_k) &= (1 + 2 + 3 + \dots + 6)/6 = 7/2, \\ \text{Var}(X_k) &= (1^2 + 2^2 + \dots + 6^2)/6 - (7/2)^2 = 35/12. \end{aligned}$$

La loi forte des grands nombres nous dit que lorsque n est grand, le score moyen S_n/n devrait être proche de $3,5$, tandis que d'après le théorème central limite,

$$\mathbb{P}\left(|S_n - \frac{7n}{2}| \leq \alpha \sqrt{\frac{35n}{12}}\right) \approx \mathbb{P}(|\mathcal{N}| \leq \alpha) = \Phi(\alpha) - \Phi(-\alpha) = 2\Phi(\alpha) - 1.$$

Prenons par exemple $n = 10,000$ et $\alpha = 1,96$. On a $2\Phi(\alpha) - 1 \approx 0,95$, $1,96 \sqrt{35n/12} \approx 334,8$ et que

$$\mathbb{P}(34,666 \leq S_n \leq 35,334) \approx 0,95.$$

Donc, avec 95% de chance, le score moyen se situe dans l'intervalle $[34666, 35334]$.

Si l'on prend $\alpha = 0,6744$, on a $\Phi(\alpha) = 0,75$ donc $2\Phi(\alpha) - 1 = 0,5$. Dans ce cas-là, si l'on prend toujours $n = 10,000$, on obtient $0,6744 \sqrt{35n/12} \approx 115,2$. Donc

$$\mathbb{P}(34,885 \leq S_n \leq 35,115) \approx 0,50.$$

Donc le score moyen a à peu près autant de chance de tomber dans [34885, 35115] que de tomber à l'extérieur. \square

Exemple 2.4. Soit μ une loi de probabilité sur \mathbb{R} , avec $\int_{\mathbb{R}} x^2 \mu(dx) < \infty$, ayant la propriété suivante : si X et Y sont des v.a. indépendantes suivant la loi μ , alors $X + Y$ et aX suivent la même loi, pour certaine constante $a \in \mathbb{R}$.

Pour éviter la situation triviale, on suppose que $\text{Var}(X) > 0$ (sinon, X sera presque sûrement une constante).

En comparant les espérance et variance de $X + Y$ et de aX , on obtient : $2\mathbb{E}(X) = a\mathbb{E}(X)$ et $2\text{Var}(X) = a^2\text{Var}(X)$. Comme $\text{Var}(X) > 0$, ceci n'est possible que si $a^2 = 2$ et $\mathbb{E}(X) = 0$.

Soit maintenant X_1, X_2, \dots une suite iid de v.a. suivant la loi μ . Par hypothèse,

$$\frac{X_1 + X_2 + \dots + X_{4^n}}{2^n}$$

a la même loi que X , donc converge en loi vers X . D'autre part, par la théorème central limite,

$$\frac{X_1 + X_2 + \dots + X_{4^n}}{2^n} \rightarrow \mathcal{N}(0, \text{Var}(X)), \quad \text{en loi,}$$

ce qui implique que X suit la loi gaussienne $\mathcal{N}(0, \text{Var}(X))$.

Par conséquent, μ est une mesure gaussienne centrée. \square

Exemple 2.5. Soient X_1, X_2, \dots des v.a. réelles i.i.d. telles que $X_1 \geq 0$, $\mathbb{E}(X_1) = 1$ et $\text{Var}(X_1) = \sigma^2 \in]0, +\infty[$. Alors

$$\frac{2(\sqrt{S_n} - \sqrt{n})}{\sigma} \rightarrow \mathcal{N}(0, 1), \quad \text{en loi.}$$

En effet, on peut écrire $\frac{2(\sqrt{S_n} - \sqrt{n})}{\sigma} = \frac{X_n}{Y_n}$, où

$$X_n := \frac{S_n - n}{\sqrt{n\sigma^2}}, \quad Y_n := \frac{1}{2} \left(\sqrt{\frac{S_n}{n}} + 1 \right).$$

Par le théorème central limite, $X_n \rightarrow \mathcal{N}(0, 1)$ en loi, tandis que la loi forte des grands nombres confirme que $Y_n \rightarrow 1$ p.s. On peut donc appliquer le théorème de Slutsky (Théorème 3.5 du Chapitre 9) pour conclure que $\frac{X_n}{Y_n} \rightarrow \mathcal{N}(0, 1)$ en loi. \square

Exemple 2.6. Soit f une fonction continue et bornée sur \mathbb{R} . Alors, lorsque $n \rightarrow \infty$,

$$e^{-n} \sum_{k=0}^{\infty} f\left(\frac{n-k}{\sqrt{n}}\right) \frac{n^k}{k!} \rightarrow \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-x^2/2} f(x) dx,$$

$$e^{-n} \sum_{k=0}^n \frac{n^k}{k!} \rightarrow \frac{1}{2}.$$

On montre ces convergences en utilisant le théorème central limite. Soit X_1, X_2, \dots une suite iid de v.a. de Poisson de paramètre 1. On sait que, pour chaque n fixé, $S_n = X_1 + \dots + X_n$ suit la loi de Poisson de paramètre n (puisque la fonction génératrice de la loi de Poisson de paramètre n est $s \rightarrow e^{-n(1-s)}$). Donc

$$e^{-n} \sum_{k=0}^{\infty} f\left(\frac{n-k}{\sqrt{n}}\right) \frac{n^k}{k!} = \mathbb{E}\left(f\left(\frac{n-S_n}{\sqrt{n}}\right)\right).$$

D'autre part, d'après le théorème central limite,

$$\frac{n-S_n}{\sqrt{n}} \rightarrow \mathcal{N}(0,1), \quad \text{en loi.}$$

Comme f est continue et bornée sur \mathbb{R} , par définition,

$$\mathbb{E}\left(f\left(\frac{n-S_n}{\sqrt{n}}\right)\right) \rightarrow \mathbb{E}(f(\mathcal{N}(0,1))),$$

ce qui donne la première convergence.

Pour la seconde, il suffit de remarquer que

$$e^{-n} \sum_{k=0}^n \frac{n^k}{k!} = \mathbb{P}(S_n \leq n) = \mathbb{P}\left(\frac{S_n-n}{\sqrt{n}} \leq 0\right) \rightarrow \mathbb{P}(\mathcal{N}(0,1) \leq 0) = \frac{1}{2}. \quad \square$$

Exemple 2.7. Nous allons démontrer la formule de Stirling à l'aide du théorème central limite. Prenons comme dans l'exercice précédent S_n la somme partielle d'une suite i.i.d. de variables aléatoires de loi de Poisson de paramètre 1. On note $T_n := \frac{S_n-n}{\sqrt{n}}$, et \mathcal{N} désigne une v.a. normale standard. On sait donc d'après le théorème central limite que T_n converge en loi vers \mathcal{N} . Pour chaque $a > 0$ fixé, la fonction $f_a(x) = \min(|x|, a)$ est continue et bornée sur \mathbb{R} , donc

$$(*) \quad \mathbb{E}(f_a(T_n)) \rightarrow \mathbb{E}(f_a(\mathcal{N})), \quad n \rightarrow \infty.$$

D'autre part, pour une v.a. X quelconque avec $\mathbb{E}(X^2) < \infty$,

$$\begin{aligned} 0 \leq \mathbb{E}(|X| - f_a(X)) &= \mathbb{E}((|X| - a)\mathbf{1}_{\{|X|>a\}}) \\ &\leq \mathbb{E}(|X|\mathbf{1}_{\{|X|>a\}}) \leq \sqrt{\mathbb{E}(X^2)\mathbb{P}(|X|>a)} \\ &\leq \frac{\mathbb{E}(X^2)}{a}, \end{aligned}$$

à l'aide des inégalités de Cauchy–Schwarz et de Markov. Remarquons que $\mathbb{E}(T_n^2) = \text{Var}(T_n) = 1$ et que $\mathbb{E}(\mathcal{N}^2) = 1$. La convergence (*) implique donc que

$$\begin{aligned} \limsup_{n \rightarrow \infty} \mathbb{E}(|T_n|) &\leq \mathbb{E}(|\mathcal{N}|) + \frac{2}{a}, \\ \liminf_{n \rightarrow \infty} \mathbb{E}(|T_n|) &\geq \mathbb{E}(|\mathcal{N}|) - \frac{2}{a}. \end{aligned}$$

Ceci signifie que $\mathbb{E}(|T_n|) \rightarrow \mathbb{E}(|\mathcal{N}|)$. Puisque $|x| = (x + x^+)/2$ (où x^+ est la partie positive de x), et que $\mathbb{E}(T_n) = 0 = \mathbb{E}(\mathcal{N})$, on a montré que

$$\mathbb{E}(T_n^+) \rightarrow \mathbb{E}(\mathcal{N}^+).$$

Le terme de droite vaut

$$\frac{1}{\sqrt{2\pi}} \int_0^\infty x \exp\{-x^2/2\} dx = \frac{1}{\sqrt{2\pi}}.$$

Quand au terme de gauche, on peut l'écrire comme

$$\begin{aligned} e^{-n} \sum_{j=n+1}^\infty \frac{n^j}{j!} \left(\frac{j-n}{\sqrt{n}} \right) &= \frac{e^{-n}}{\sqrt{n}} \lim_{k \rightarrow \infty} \sum_{j=n+1}^k \left[\frac{n^j}{(j-1)!} - \frac{n^{j+1}}{j!} \right] \\ &= \frac{e^{-n}}{\sqrt{n}} \lim_{k \rightarrow \infty} \left[\frac{n^{n+1}}{n!} - \frac{n^{k+1}}{k!} \right] \\ &= \frac{e^{-n}}{\sqrt{n}} \frac{n^{n+1}}{n!} = \frac{e^{-n} n^n \sqrt{n}}{n!}. \end{aligned}$$

En remettant les pièces en place, on a donc démontré la formule de Stirling

$$\frac{e^{-n} n^n \sqrt{n}}{n!} \sim \frac{1}{\sqrt{2\pi}}. \quad \square$$

Exemple 2.8. Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires i.i.d. à valeurs dans E (muni de la tribu \mathcal{E}). Soit $A \in \mathcal{E}$. On s'intéresse à

$$R_n(\omega) := \#\{1 \leq i \leq n : X_i(\omega) \in A\}.$$

En écrivant $R_n = \sum_{i=1}^n \mathbf{1}_{\{X_i \in A\}}$, il est clair que R_n est une variable aléatoire réelle. Les v.a. $(\mathbf{1}_{\{X_i \in A\}})_{i \geq 1}$ sont i.i.d., telles que $\mathbb{E}(\mathbf{1}_{\{X_i \in A\}}) = \mathbb{P}(X_1 \in A)$ et $\text{Var}(\mathbf{1}_{\{X_i \in A\}}) = \mathbb{P}(X_1 \in A)\mathbb{P}(X_1 \notin A)$. Le théorème central limite confirme alors que

$$\frac{R_n - n\mathbb{P}(X_1 \in A)}{\sqrt{n}}$$

converge en loi vers $\mathcal{N}(0, \sigma^2)$, où $\sigma^2 := \mathbb{P}(X_1 \in A)\mathbb{P}(X_1 \notin A)$. \square

Exemple 2.9. Soient Y_1, Y_2, \dots des variables aléatoires réelles indépendantes, suivant toutes la loi gaussienne centrée réduite $\mathcal{N}(0, 1)$.

(i) Montrer que $(\sum_{j=1}^n Y_j)/\sqrt{\sum_{j=1}^n Y_j^2}$ converge en loi et préciser la loi limite.

(ii) Soit $c > 0$. Déterminer si la série $\sum_n \mathbb{P}(Y_n > c\sqrt{2 \ln n})$ est convergente. On pourra utiliser les inégalités suivantes : pour tout $x > 1$, $(\frac{1}{x} - \frac{1}{x^3})e^{-x^2/2} \leq \int_x^\infty e^{-u^2/2} du \leq \frac{1}{x}e^{-x^2/2}$.

(iii) Montrer que

$$\limsup_{n \rightarrow \infty} \frac{Y_n}{\sqrt{\ln n}} = \sqrt{2}, \quad \text{p.s.}$$

(iv) Que peut-on dire de $\liminf_{n \rightarrow \infty} \frac{Y_n}{\sqrt{\ln n}}$?

(v) Soit R_n la variable aléatoire définie de la façon suivante : $R_n(\omega)$ désigne le nombre d'éléments parmi $Y_1(\omega), Y_2(\omega), \dots, Y_n(\omega)$ qui sont strictement positifs. Montrer que $\frac{2R_n - n}{\sqrt{n}}$ converge en loi et préciser la loi limite.

(vi) Définissons N par $N(\omega) = \inf\{n \geq 1 : Y_n(\omega) > 0\}$ (convention : $\inf \emptyset = \infty$). Calculer $\mathbb{P}(N < \infty)$. Soit $\alpha > 0$ un réel. Déterminer les valeurs de α telles que $\mathbb{E}(e^{\alpha N}) < \infty$.

Solution. (i) Soient $S_n := \sum_{j=1}^n Y_j$ et $T_n := \sum_{j=1}^n Y_j^2$. Par la loi des grands nombres, T_n/n converge p.s. (donc a fortiori, en loi) vers 1. Donc $\sqrt{T_n/n}$ converge en loi vers 1. D'autre part, S_n/\sqrt{n} suit la loi $\mathcal{N}(0, 1)$, donc a fortiori converge en loi vers $\mathcal{N}(0, 1)$. Il résulte du théorème de Slutsky (Théorème 3.5 du chapitre précédent) que $\frac{S_n}{\sqrt{T_n}} = \frac{S_n/\sqrt{n}}{\sqrt{T_n/n}}$ converge en loi vers $\mathcal{N}(0, 1)$.

Remarquons que dans cette question, on n'a pas vraiment utilisé l'hypothèse que les variables aléatoires Y_i suivent la loi gaussienne $\mathcal{N}(0, 1)$: il suffit que $\mathbb{E}(Y_1) = 0$ et $\mathbb{E}(Y_1^2) = 1$.

(ii) Soit d'abord $c > 1$. Soit n tellement grand que $c\sqrt{2 \ln n} > 1$. Par l'indication fournie dans l'exercice,

$$\mathbb{P}(Y_n > c\sqrt{2 \ln n}) \leq \frac{1}{c\sqrt{2 \ln n} n e^2},$$

ce qui donne le terme général d'une série convergente (séries de Bertrand). Donc la série $\sum_n \mathbb{P}(Y_n > c\sqrt{2 \ln n})$ converge.

D'autre part,

$$\mathbb{P}(Y_n > \sqrt{2 \ln n}) \geq \left(\frac{1}{(2 \ln n)^{1/2}} - \frac{1}{(2 \ln n)^{3/2}} \right) \frac{1}{n} \sim \frac{1}{(2 \ln n)^{1/2} n},$$

ce qui donne le terme général d'une série divergente. Donc la série $\sum_n \mathbb{P}(Y_n > \sqrt{2 \ln n})$ diverge. A fortiori, $\sum_n \mathbb{P}(Y_n > c\sqrt{2 \ln n})$ diverge si $c \leq 1$.

Conclusion : $\sum_n \mathbb{P}(Y_n > c\sqrt{2 \ln n})$ converge si et seulement si $c > 1$.

(iii) Soit $c > 1$. On peut appliquer le lemme de Borel–Cantelli pour arriver à : il existe un ensemble mesurable $A \in \mathcal{F}$ avec $\mathbb{P}(A) = 1$ tel que, pour $\omega \in A$, on peut trouver $n_0 = n_0(\omega) < \infty$ satisfaisant : $n \geq n_0 \implies Y_n(\omega) \leq c\sqrt{2 \ln n}$. Par conséquent, presque sûrement, $\limsup_{n \rightarrow \infty} \frac{Y_n}{\sqrt{2 \ln n}} \leq c$. Le réel $c > 1$ étant quelconque, on déduit que

$$\limsup_{n \rightarrow \infty} \frac{Y_n}{\sqrt{\ln n}} \leq \sqrt{2}, \quad \text{p.s.}$$

Pour vérifier la borne inférieure, on utilise la deuxième partie du lemme de Borel–Cantelli (rappelons que les variables (Y_n) sont indépendantes et que $\sum_n \mathbb{P}(Y_n > \sqrt{2 \ln n})$ diverge), pour voir que l'on peut trouver $B \in \mathcal{F}$ avec $\mathbb{P}(B) = 1$ tel que, $\forall \omega \in B$, \exists une infinité de n satisfaisant $Y_n(\omega) > \sqrt{2 \ln n}$. Donc $\mathbb{P}(\limsup_{n \rightarrow \infty} \frac{Y_n}{\sqrt{2 \ln n}} \geq 1) = 1$.

En conclusion, $\limsup_{n \rightarrow \infty} \frac{Y_n}{\sqrt{\ln n}} = \sqrt{2}$ p.s.

(iv) Les variables aléatoires $(-Y_n)_{n \geq 1}$ sont aussi indépendantes et suivent la même loi $\mathcal{N}(0, 1)$. On peut donc appliquer le résultat dans la question précédente pour voir que, p.s., $\limsup_{n \rightarrow \infty} \frac{-Y_n}{\sqrt{\ln n}} = \sqrt{2}$. Autrement dit, $\liminf_{n \rightarrow \infty} \frac{Y_n}{\sqrt{\ln n}} = -\sqrt{2}$, presque sûrement.

(v) Par définition, $R_n = \sum_{j=1}^n \mathbf{1}_{A_j}$, où $A_j = \{\omega \in \Omega : Y_j(\omega) > 0\}$. Les variables aléatoires $(\mathbf{1}_{A_j})_{j \geq 1}$ sont iid, avec $\mathbb{E}(\mathbf{1}_{A_1}) = \mathbb{P}(A_1) = 1/2$ et $\text{Var}(\mathbf{1}_{A_1}) = \mathbb{P}(A_1) - (\mathbb{P}(A_1))^2 = 1/4$. D'après le théorème central limite, $\frac{2R_n - n}{\sqrt{n}}$ converge en loi vers $\mathcal{N}(0, 1)$.

(vi) D'après la question (iii), $\mathbb{P}(\{\omega \in \Omega : \limsup_{n \rightarrow \infty} Y_n(\omega) = \infty\}) = 1$. A fortiori, $\mathbb{P}(\{\omega : \exists n \text{ tel que } Y_n(\omega) > 0\}) = 1$. Donc $\mathbb{P}(N < \infty) = 1$.

Pour chaque n , $\mathbb{P}(T = n) = \mathbb{P}(Y_1 \leq 0) \cdots \mathbb{P}(Y_{n-1} \leq 0) \mathbb{P}(Y_n > 0) = 2^{-n}$. Or, $\sum_n e^{\alpha n} 2^{-n} < \infty$ si et seulement si $\alpha < \ln 2$, on déduit que $\mathbb{E}(e^{\alpha N}) < \infty \iff \alpha < \ln 2$. \square

3. Application à des tests statistiques

Nous allons voir sur deux exemples comment le théorème central limite peut être appliqué pour confirmer ou infirmer des hypothèses statistiques.

Exemple 3.1. Sur 53 680 familles de lapins ayant 8 petits chacune (soit 429 440 lapereaux au total), il y a 221 023 mâles et 208 417 femelles. La question est de savoir si le nombre de mâles est significativement plus élevé que celui de femelles. Autrement dit, on voudrait tester l'hypothèse : Les chances sont égales d'avoir un mâle ou une femelle.

Notons $X_i = 1$ si le i -ème lapereau est un mâle, et 0 si c'est une femelle. Si l'hypothèse est vraie, X_i suit une loi de Bernoulli de paramètre $1/2$, de moyenne $1/2$ et de variance $1/4$. Dans ce cas, en notant $S_{429440} = X_1 + \dots + X_{429440}$ le nombre total de mâles, le théorème central limite nous dit que pour tout $a < b$

$$\mathbb{P} \left(\frac{a\sqrt{429440} + 429440}{2} \leq S_{429440} \leq \frac{b\sqrt{429440} + 429440}{2} \right) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2} dx.$$

Pour $a = -1,96$ et $b = 1,96$, la quantité de droite vaut 0,95. Autrement dit, avec 5% d'erreur, S_{429440} devrait être compris entre

$$-2328 + 214720 = 214064 \quad \text{et} \quad 2328 + 214720 = 215376.$$

Or le nombre observé de S_{429440} est bien au delà de ces bornes, on va donc rejeter l'hypothèse avec un risque d'erreur de 5%. \square

Exemple 3.2. On veut tester l'hypothèse mendelienne sur la couleur des yeux : bleu récessif et marron dominant. Si l'hypothèse est vraie, une personne prise au hasard a une chance sur quatre d'avoir les yeux bleus. Si l'hypothèse est valide, combien de personnes doit-on observer pour être certain avec une probabilité de 99,8% que la proportion de personnes aux yeux marrons sera comprise entre 0,7 et 0,8?

C'est en quelque sorte le problème inverse du précédent : il s'agit de déterminer la taille de l'échantillon. Sous l'hypothèse de Mendel, pour un échantillon de n personnes, en notant S_n le nombre de celles qui ont les yeux marrons, on a avec $\mu = 0,75$ (espérance) et $\sigma^2 = 0,75 \times 0,25 = 0,1875$ (variance),

$$\mathbb{P}(\sqrt{n}\sigma a + n\mu \leq S_n \leq \sqrt{n}\sigma b + n\mu) \approx \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2} dx.$$

L'intégrale vaut environs 0,998 pour $a = -3$ et $b = 3$. La probabilité peut s'écrire

$$\mathbb{P} \left(\frac{\sigma a}{\sqrt{n}} + \mu \leq \frac{S_n}{n} \leq \frac{\sigma b}{\sqrt{n}} + \mu \right).$$

On doit donc choisir n tel que

$$\frac{\sigma a}{\sqrt{n}} + \mu = 0,75 - \frac{3 \times \sqrt{0,1875}}{n} \geq 0,7, \quad \frac{\sigma b}{\sqrt{n}} + \mu = 0,75 + \frac{3 \times \sqrt{0,1875}}{n} \leq 0,8,$$

et il suffit de prendre $n = 675$.

Ainsi, si ayant observé 675 personnes, on trouve que la proportion de gens ayant yeux marrons est "très éloignée" de l'intervalle $[0,7, 0,8]$, il faudra revoir l'hypothèse de Mendel. \square

Chapitre 11. Espérances conditionnelles

1. Existence et unicité

On se donne un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$, une v.a. X intégrable : $\mathbb{E}(|X|) < +\infty$, et une sous-tribu $\mathcal{G} \subset \mathcal{F}$.

Définition 1.1. On appelle espérance conditionnelle de X sachant \mathcal{G} , notée par $\mathbb{E}(X | \mathcal{G})$, toute variable aléatoire Y intégrable vérifiant les deux conditions suivantes :

- (i) Y est \mathcal{G} -mesurable;
- (ii) Pour tout $A \in \mathcal{G}$, $\int_A X \, d\mathbb{P} = \int_A Y \, d\mathbb{P}$ (c'est-à-dire $\mathbb{E}(X \mathbf{1}_A) = \mathbb{E}(Y \mathbf{1}_A)$.)

On écrira probabilité conditionnelle $\mathbb{P}(A | \mathcal{G}) := \mathbb{E}(\mathbf{1}_A | \mathcal{G})$.

Montrons l'existence et l'unicité de l'espérance conditionnelle.

Unicité. Si Y et Y' sont deux v.a. intégrables vérifiant (i) et (ii). On a, pour tout $A \in \mathcal{G}$, $\mathbb{E}(Y \mathbf{1}_A) = \mathbb{E}(Y' \mathbf{1}_A)$. Soit $\varepsilon > 0$ et soit $A := \{\omega \in \Omega : Y(\omega) - Y'(\omega) \geq \varepsilon\} \in \mathcal{G}$. On a

$$0 = \mathbb{E}(Y \mathbf{1}_A) - \mathbb{E}(Y' \mathbf{1}_A) = \mathbb{E}((Y - Y') \mathbf{1}_A) \geq \varepsilon \mathbb{P}(A),$$

ce qui signifie que $\mathbb{P}(A) = 0$. Le choix de $\varepsilon > 0$ étant quelconque, on obtient $Y \leq Y'$ p.s. De même, $Y' \leq Y$ p.s. Donc $Y = Y'$ p.s. Ceci montre l'unicité.

Strictement dit, on doit écrire $Y = \mathbb{E}(X | \mathcal{G})$ p.s., mais l'expression “presque sûr” est souvent omise. □

Existence. La preuve s'appuie sur le Théorème de Radon–Nikodym : si μ et ν sont deux mesures σ -finies sur (Ω, \mathcal{G}) telles que $\nu \ll \mu$ (c'est-à-dire pour tout $A \in \mathcal{G}$, $\mu(A) = 0 \implies \nu(A) = 0$), alors il existe une fonction mesurable $f \geq 0$ telle que $\int_A f \, d\mu = \nu(A)$, quel que soit $A \in \mathcal{G}$. La fonction f est notée par $\frac{d\nu}{d\mu}$.

On suppose sans perte de généralité que $X \geq 0$ (sinon, on considèrera séparément X^+ et X^-). Soit μ la mesure sur \mathcal{G} définie par $\mu(A) = \int_A X \, d\mathbb{P}$, $A \in \mathcal{G}$ (c'est-à-dire $\mu = X \bullet \mathbb{P}$; l'intégrabilité de X garantit que μ est une mesure finie). On a donc deux mesures \mathbb{P} (ou plutôt la restriction de \mathbb{P} sur \mathcal{G}) et μ sur \mathcal{G} telles que $\mu \ll \mathbb{P}$. Par le théorème de Radon–Nikodym, $\frac{d\mu}{d\mathbb{P}}$ est bien définie, qui est intégrable car μ est une mesure finie.

Or, $\frac{d\mu}{d\mathbb{P}}$ est \mathcal{G} -mesurable, telle que pour tout $A \in \mathcal{G}$, $\int_A X \, d\mathbb{P} = \mu(A) = \int_A \frac{d\mu}{d\mathbb{P}} \, d\mathbb{P}$, ce qui prouve que $\frac{d\mu}{d\mathbb{P}}$ vérifie les conditions (i) et (ii). \square

Remarque. Heuristiquement on peut interpréter $\mathbb{E}(X | \mathcal{G})$ comme l'espérance de X en tenant compte des informations apportées par la tribu \mathcal{G} : pour chaque $A \in \mathcal{G}$, on sait si A est réalisé ou non. Voici quelques exemples simples pour illustrer cette idée.

Exemple 1.2. Soit X une v.a. intégrable et \mathcal{G} -mesurable. Sachant la tribu \mathcal{G} , X qui est \mathcal{G} -mesurable est considérée comme une constante : $\mathbb{E}(X | \mathcal{G}) = X$.

En effet, il est clair dans cette situation que X satisfait les conditions (i) et (ii). \square

Exemple 1.3. L'exemple précédent concerne la situation particulière où toute information sur X est déjà contenue dans la tribu \mathcal{G} . On étudie maintenant l'autre situation extrême : X est indépendante de \mathcal{G} , c'est-à-dire X et $\mathbf{1}_A$ sont indépendantes quel que soit $A \in \mathcal{G}$. Dans ce cas, le fait de savoir \mathcal{G} n'apporte strictement rien sur X : $\mathbb{E}(X | \mathcal{G}) = \mathbb{E}(X)$.

En effet, $\mathbb{E}(X)$ est une v.a. (dégénérée) intégrable et \mathcal{G} -mesurable. Il nous suffit de vérifier la condition (ii). Soit $A \in \mathcal{G}$. Comme X et $\mathbf{1}_A$ sont indépendantes, on a $\mathbb{E}(X \mathbf{1}_A) = \mathbb{E}(X) \mathbb{E}(\mathbf{1}_A) = \mathbb{E}(\mathbb{E}(X) \mathbf{1}_A)$, ce qui prouve (ii). \square

L'espérance conditionnelle jouit de la plupart des propriétés de l'espérance.

Propriété 1.4. (1) Si X et Y sont deux v.a. intégrables, alors $\mathbb{E}(aX + bY | \mathcal{G}) = a\mathbb{E}(X | \mathcal{G}) + b\mathbb{E}(Y | \mathcal{G})$.

(2) Si en plus $X \geq Y$, alors $\mathbb{E}(X | \mathcal{G}) \geq \mathbb{E}(Y | \mathcal{G})$. En particulier, $X \geq 0$ implique $\mathbb{E}(X | \mathcal{G}) \geq 0$.

(3) Si \mathcal{H} est une sous-tribu de \mathcal{F} telle que $\mathcal{G} \subset \mathcal{H}$, alors $\mathbb{E}[\mathbb{E}(X | \mathcal{H}) | \mathcal{G}] = \mathbb{E}(X | \mathcal{G})$.

Preuve. (1) et (2) se vérifient par définition. Montrons (3). Il est clair que $\mathbb{E}[\mathbb{E}(X | \mathcal{H}) | \mathcal{G}]$ est bien définie, car par définition, $\mathbb{E}(X | \mathcal{H})$ est intégrable. Remarquons que $\mathbb{E}[\mathbb{E}(X | \mathcal{H}) | \mathcal{G}]$ est \mathcal{G} -mesurable, ce qui donne la propriété (i). Pour (ii), soit $A \in \mathcal{G}$, et on a

$$\mathbb{E}[\mathbb{E}[\mathbb{E}(X | \mathcal{H}) | \mathcal{G}] \mathbf{1}_A] = \mathbb{E}[\mathbb{E}(X | \mathcal{H}) \mathbf{1}_A] = \mathbb{E}(X \mathbf{1}_A),$$

car $A \in \mathcal{H}$. Ceci donne la propriété (ii), et confirme donc que $\mathbb{E}[\mathbb{E}(X | \mathcal{H}) | \mathcal{G}] = \mathbb{E}(X | \mathcal{G})$.
□

Théorème 1.5. Si X est une v.a. intégrable, alors $\mathbb{E}(\mathbb{E}(X | \mathcal{G})) = \mathbb{E}(X)$.

Preuve. Ceci est une conséquence de la propriété (ii), en prenant $A = \Omega$. □

Théorème 1.6. Si X et Y sont des v.a. telles que Y et XY soient intégrables, et si X est \mathcal{G} -mesurable, alors $\mathbb{E}(XY | \mathcal{G}) = X \mathbb{E}(Y | \mathcal{G})$.

Preuve. Sans perte de généralité, on suppose que X et Y sont positives (sinon, on considèrera leurs parties positive et négative, séparément).

L'identité est vraie si X est une fonction indicatrice : soit $A \in \mathcal{G}$, alors $\mathbf{1}_A \mathbb{E}(Y | \mathcal{G})$ est une v.a. intégrable et \mathcal{G} -mesurable, telle que pour tout $B \in \mathcal{G}$,

$$\mathbb{E}[\mathbf{1}_A \mathbb{E}(Y | \mathcal{G}) \mathbf{1}_B] = \mathbb{E}(Y \mathbf{1}_{A \cap B}) = \mathbb{E}(\mathbf{1}_A Y \mathbf{1}_B),$$

ce qui implique que $\mathbf{1}_A \mathbb{E}(Y | \mathcal{G}) = \mathbb{E}(\mathbf{1}_A Y | \mathcal{G})$.

Par combinaison linéaire, l'identité reste vraie pour toute fonction étagée X qui est \mathcal{G} -mesurable : pour tout $B \in \mathcal{G}$,

$$\int_B X \mathbb{E}(Y | \mathcal{G}) \, d\mathbb{P} = \int_B XY \, d\mathbb{P}.$$

Le théorème de convergence monotone confirme alors que l'identité fonctionne encore si X est positive et \mathcal{G} -mesurable. En particulier, en prenant $B = \Omega$ on déduit que $X \mathbb{E}(Y | \mathcal{G})$ est intégrable. Par définition, elle vaut $\mathbb{E}(XY | \mathcal{G})$. □

2. Espérance conditionnelle par rapport à une v.a.

Lorsque la tribu est engendrée par une variable aléatoire (pas nécessairement à valeurs réelles), l'espérance conditionnelle a des propriétés particulières. Cela aide surtout à simplifier le calcul dans des cas pratiques.

Définition 2.1. Si \mathcal{G} est engendrée par une variable Z (à valeurs dans un espace mesurable quelconque), et si X est une v.a. réelle intégrable, alors on écrira $\mathbb{E}(X | Z)$ au lieu de $\mathbb{E}(X | \mathcal{G})$.

Si $Z = (Z_1, Z_2, \dots)$, on écrira aussi $\mathbb{E}(X | Z_1, Z_2, \dots)$.

Soit Z une v.a. à valeurs dans (E, \mathcal{E}) . Soit $A \in \sigma(Z)$. Par définition, il existe $B \in \mathcal{E}$ telle que $A = \{\omega \in \Omega : Z(\omega) \in B\}$; autrement dit, $\mathbf{1}_A = \mathbf{1}_B(Z)$. Donc toute fonction étagée qui est $\sigma(Z)$ -mesurable s'écrit comme une fonction étagée (\mathcal{E} -mesurable) de Z . Par un passage à la limite, si ξ est une v.a. réelle intégrable et $\sigma(Z)$ -mesurable, elle s'écrit comme $\xi = h(Z)$, où $h : E \rightarrow \mathbb{R}$ est une fonction borélienne. Prenons $\xi = \mathbb{E}(X | Z)$ et on obtient

$$\mathbb{E}(X | Z) = h(Z).$$

L'unicité de l'espérance conditionnelle nous assure que si h_1 est une fonction borélienne telle que $\mathbb{E}(X | Z) = h_1(Z)$, alors $h_1 = h$, P_Z -p.p.

Exemple 2.2. Soient X et Y deux v.a. réelles indépendantes et identiquement distribuées, telles que $\mathbb{E}(|X|) < \infty$. Alors $\mathbb{E}(X | X + Y) = (X + Y)/2$.

En fait, si $P_{(X,Z)} = P_{(Y,Z)}$, alors $\mathbb{E}(X | Z) = \mathbb{E}(Y | Z)$. Pour voir cela, on observe d'abord que $\mathbb{E}(X | Z)$ est une v.a. intégrable et $\sigma(Z)$ -mesurable. De plus, pour tout $A \in \sigma(Z)$, $\mathbb{E}(\mathbb{E}(X | Z) \mathbf{1}_A) = \mathbb{E}(X \mathbf{1}_A)$. On sait qu'il existe une fonction borélienne h telle que $\mathbf{1}_A = h(Z)$, ce qui nous donne $\mathbb{E}(X \mathbf{1}_A) = \mathbb{E}(X h(Z)) = \mathbb{E}(Y h(Z)) = \mathbb{E}(Y \mathbf{1}_A)$. D'où $\mathbb{E}(\mathbb{E}(X | Z) \mathbf{1}_A) = \mathbb{E}(Y \mathbf{1}_A)$. Par conséquent, $\mathbb{E}(X | Z) = \mathbb{E}(Y | Z)$.

Pour revenir à l'exemple $Z := X + Y$, il suffit de remarquer que $P_{(X, X+Y)} = P_{(Y, X+Y)}$, car $P_{(X,Y)} = P_{(Y,X)}$. Donc $\mathbb{E}(X | X + Y) = \mathbb{E}(Y | X + Y)$. Or, $\mathbb{E}(X + Y | X + Y) = X + Y$, on obtient $\mathbb{E}(X | X + Y) = (X + Y)/2$. \square

Théorème 2.3. Soient X et Z deux v.a. *indépendantes*, et soit φ une fonction mesurable telle que $\mathbb{E}[|\varphi(X, Z)|] < \infty$. Soit $h(z) = \mathbb{E}(\varphi(X, z))$. Alors

$$h(Z) = \mathbb{E}(\varphi(X, Z) | Z).$$

Remarque. Le sens du théorème est intuitivement clair. Si l'on cherche la valeur de $\mathbb{E}(\varphi(X, Z) | Z)$, comme Z est mesurable par rapport à $\sigma(Z)$, on peut la considérer comme une constante, disons z . Dans ce cas, $\mathbb{E}(\varphi(X, z) | Z)$ devient $\mathbb{E}(\varphi(X, z))$ (car X et Z sont indépendantes), qui n'est autre que $h(z)$. Et bien sûr, z est en réalité Z . En remplaçant z par Z dans $h(z)$, on obtient $h(Z)$ comme la valeur de $\mathbb{E}(\varphi(X, Z) | Z)$. \square

Preuve du Théorème 2.3. Le théorème de Fubini nous dit que $h(z)$ est bien définie, et que $h(Z)$ est intégrable. Soit $A \in \sigma(Z)$. On vérifie que $\mathbb{E}[h(Z) \mathbf{1}_A] = \mathbb{E}[\varphi(X, Z) \mathbf{1}_A]$.

Comme $A \in \sigma(Z)$, on peut écrire $A = \{\omega : Z(\omega) \in B\}$. Par l'indépendance et le théorème de Fubini,

$$\begin{aligned} \mathbb{E}[\varphi(X, Z)\mathbf{1}_A] &= \mathbb{E}[\varphi(X, Z)\mathbf{1}_B(Z)] \\ &= \int \int \varphi(x, z)\mathbf{1}_B(z) P_X(dx)P_Z(dz) \\ &= \int h(z)\mathbf{1}_B(z) P_Z(dz) \\ &= \mathbb{E}[h(Z)\mathbf{1}_B(Z)]. \end{aligned}$$

Il suffit alors de remarquer que $\mathbf{1}_B(Z(\omega)) = \mathbf{1}_A(\omega)$. \square

Voici une recette générale pour calculer $\mathbb{E}(X | Z)$, lorsque (X, Z) est un vecteur aléatoire à valeurs dans \mathbb{R}^2 dont la loi est absolument continue par rapport à la mesure de Lebesgue sur \mathbb{R}^2 .

Théorème 2.4. Soit (X, Z) un vecteur aléatoire à valeurs dans \mathbb{R}^2 qui admet une densité $f_{(X,Z)}$. On suppose que $\mathbb{E}(|X|) < \infty$ et que $f_Z(z) > 0$ pour tout z . On pose

$$\begin{aligned} f_{X|Z=z}(x) &:= \frac{f_{(X,Z)}(x, z)}{f_Z(z)}, & x \in \mathbb{R}, \\ h(z) &:= \int_{\mathbb{R}} x f_{X|Z=z}(x) dx, & z \in \mathbb{R}. \end{aligned}$$

Alors

$$\mathbb{E}(X | Z) = h(Z).$$

Preuve. Il est clair d'après le théorème de Fubini que $h(Z)$ est une variable aléatoire intégrable. Soit $A \in \sigma(Z)$. Il s'agit de prouver que $\mathbb{E}[h(Z)\mathbf{1}_A] = \mathbb{E}[X\mathbf{1}_A]$.

Comme $A \in \sigma(Z)$, on peut écrire $A = \{\omega : Z(\omega) \in B\}$, où $B \subset \mathbb{R}$ est une partie borélienne. On a alors

$$\begin{aligned} \mathbb{E}[X\mathbf{1}_A] &= \mathbb{E}[X\mathbf{1}_B(Z)] \\ &= \int_{\mathbb{R}} dx \int_B dz x f_{(X,Z)}(x, z) \\ &= \int_B dz h(z)f_Z(z) \\ &= \mathbb{E}[h(Z)\mathbf{1}_B(Z)]. \end{aligned}$$

D'où le résultat désiré. \square

Remarque. On remarque que $f_{X|Z=z}(x)$ est la limite de

$$\begin{aligned} & \frac{\mathbb{P}(X \in [x, x + dx], Z \in [z, z + dz]) / (dx dz)}{\mathbb{P}(Z \in [z, z + dz]) / dz} \\ &= \mathbb{P}(X \in [x, x + dx] | Z \in [z, z + dz]) / dx, \end{aligned}$$

lorsque dx et dz tendent vers 0. C'est pour cette raison que $f_{X|Z=z}$ est souvent appelée “densité conditionnelle de X sachant $Z = z$ ” (il s’agit bien sûr d’une fonction de densité associée à une loi de probabilité sur \mathbb{R}). Dans la littérature, on trouve de temps en temps l’écriture informelle $\mathbb{E}(X | Z = z)$ pour désigner $h(z)$. \square

Exemple 2.5. Soient X et Y deux v.a. réelles indépendantes suivant la même loi gaussienne $\mathcal{N}(0, 1)$. Soit $Z = X + Y$.

Il est facile de calculer la loi de (X, Z) . Pour toute fonction borélienne bornée $\varphi : \mathbb{R}^2 \rightarrow \mathbb{R}$, on a

$$\begin{aligned} \mathbb{E}(h(X, Z)) &= \int_{\mathbb{R}} \int_{\mathbb{R}} h(x, x + y) \frac{1}{2\pi} e^{-(x^2 + y^2)/2} dx dy \\ &= \int_{\mathbb{R}} \int_{\mathbb{R}} h(x, z) \frac{1}{2\pi} e^{-(x^2 + (z-x)^2)/2} dx dy. \end{aligned}$$

Donc (X, Z) admet une densité qui vaut

$$f_{(X, Z)}(x, z) = \frac{1}{2\pi} e^{-(x^2 + (z-x)^2)/2}, \quad (x, z) \in \mathbb{R}^2.$$

D’autre part, $P_Z = \mathcal{N}(0, 2)$, donc $f_Z(z) = \frac{1}{2\sqrt{\pi}} e^{-z^2/4}$. Posons, pour chaque $z \in \mathbb{R}$,

$$f_{X|Z=z}(x) := \frac{f_{(X, Z)}(x, z)}{f_Z(z)} = \frac{1}{\sqrt{\pi}} e^{-(x^2 + (z-x)^2)/2 + z^2/4} = \frac{1}{\sqrt{\pi}} e^{-(x-z/2)^2}.$$

La fonction $f_{X|Z=z}$ n’est autre que la densité de la loi gaussienne $\mathcal{N}(z/2, 1/2)$ (d’où vient une formulation du genre “sachant $Z = z$, X suit la loi gaussienne $\mathcal{N}(z/2, 1/2)$ ”). Posons

$$h(z) = \int_{\mathbb{R}} x f_{X|Z=z}(x) dx = z/2.$$

D’après le Théorème 2.4, $\mathbb{E}(X | Z) = h(Z) = Z/2$. On arrive donc à la même conclusion que dans l’Exemple 2.2. \square

Exemple 2.6. Soient N, X_1, X_2, \dots des v.a. réelles indépendantes, admettant toutes des moments d’ordre 1. On suppose que les X_i suivent la même loi, et que N est à valeurs dans \mathbb{N} . Définissons

$$Y(\omega) := \sum_{i=1}^{N(\omega)} X_i(\omega),$$

(convention : $\sum_{i=1}^0 := 0$). On cherche à déterminer le moment d'ordre 1 de Y .

Il est clair que Y est une v.a.; il suffit d'écrire $Y(\omega) := \sum_{n=1}^{+\infty} \sum_{i=1}^n X_i(\omega) \mathbf{1}_{\{N(\omega)=n\}}$, qui est représentée comme somme de v.a. réelles. Il est également clair que Y admet un moment d'ordre 1, car par le théorème de Fubini,

$$\begin{aligned} \mathbb{E} \left(\sum_{n=1}^{+\infty} \sum_{i=1}^n |X_i| \mathbf{1}_{\{N=n\}} \right) &= \sum_{n=1}^{+\infty} \sum_{i=1}^n \mathbb{E} (|X_i| \mathbf{1}_{\{N=n\}}) \\ &= \sum_{n=1}^{+\infty} \sum_{i=1}^n \mathbb{E} (|X_i|) \mathbb{P} (N = n) \\ &= \mathbb{E} (|X_1|) \sum_{n=1}^{+\infty} n \mathbb{P} (N = n) = \mathbb{E} (|X_1|) \mathbb{E} (N) < +\infty. \end{aligned}$$

Le même argument nous donne en fait $\mathbb{E}(Y) = \mathbb{E}(X_1)\mathbb{E}(N)$.

On peut calculer $\mathbb{E}(Y)$ à l'aide du Théorème 2.3. En effet, soit

$$h(n) = \mathbb{E} \left(\sum_{i=1}^n X_i \right) = n\mathbb{E}(X_1),$$

alors le Théorème 2.3 nous dit que $\mathbb{E}(Y | N) = h(N) = N\mathbb{E}(X_1)$. D'après le Théorème 1.5, $\mathbb{E}(Y) = \mathbb{E}(N\mathbb{E}(X_1)) = \mathbb{E}(N) \mathbb{E}(X_1)$. \square

On termine cette section avec une inégalité célèbre.

Théorème 2.7 (Inégalité de Jensen). Soit X une variable aléatoire réelle, et soit φ une fonction convexe. Si X et $\varphi(X)$ sont intégrables, alors

$$\varphi(\mathbb{E}[X | \mathcal{G}]) \leq \mathbb{E}[\varphi(X) | \mathcal{G}], \quad \text{p.s.}$$

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

Preuve. Soit φ' la dérivée à droite de φ . On a, pour tous réels x et y ,

$$\varphi'(x)(y - x) \leq \varphi(y) - \varphi(x).$$

Fixons $n \geq 1$, et soit $\xi := \mathbb{E}[X | \mathcal{G}]$. On remplace dans l'inégalité précédente x et y par $\xi \mathbf{1}_{\{|\xi| \leq n\}}$ et X , respectivement, pour arriver à :

$$\varphi'(\xi \mathbf{1}_{\{|\xi| \leq n\}}) (X - \xi \mathbf{1}_{\{|\xi| \leq n\}}) \leq \varphi(X) - \varphi(\xi \mathbf{1}_{\{|\xi| \leq n\}}).$$

On prend l'espérance conditionnelle par rapport à \mathcal{G} dans les deux côtés. Comme ξ est \mathcal{G} -mesurable, il résulte du Théorème 1.6 que

$$\varphi'(\xi \mathbf{1}_{\{|\xi| \leq n\}}) (\xi - \xi \mathbf{1}_{\{|\xi| \leq n\}}) \leq \mathbb{E}[\varphi(X) | \mathcal{G}] - \varphi(\xi \mathbf{1}_{\{|\xi| \leq n\}}).$$

On fait $n \rightarrow \infty$. En remarquant que $\xi \mathbf{1}_{\{|\xi| \leq n\}} \rightarrow \xi$, on arrive à : $0 \leq \mathbb{E}[\varphi(X) | \mathcal{G}] - \varphi(\xi)$. C'est ce qu'il fallait démontrer. \square

3. Espérance conditionnelle et projection orthogonale

Lorsque X est une v.a. réelle admettant un moment d'ordre 2, il y a une interprétation géométrique pour représenter $\mathbb{E}(X | \mathcal{G})$.

Théorème 3.1. Si $\mathbb{E}(X^2) < \infty$, alors $\mathbb{E}(X | \mathcal{G})$ est la variable Y , \mathcal{G} -mesurable et admettant un moment d'ordre 2, qui minimise "l'erreur quadratique" $\mathbb{E}[(X - Y)^2]$.

Remarque. (i) Soit $L^2(\mathcal{F}) := \{Y : \mathbb{E}(Y^2) < \infty\}$ qui, muni du produit scalaire $\langle Y_1, Y_2 \rangle := \mathbb{E}(Y_1 Y_2)$, est un espace de Hilbert. Il est facile de voir que $L^2(\mathcal{G}) := \{Y : Y \text{ est } \mathcal{G}\text{-mesurable, } \mathbb{E}(Y^2) < \infty\}$ est un sous-espace fermé. Sous la condition $\mathbb{E}(X^2) < \infty$, le Théorème 3.1 nous confirme que $\mathbb{E}(X | \mathcal{G})$ est la projection orthogonale de X sur $L^2(\mathcal{G})$, c'est-à-dire l'élément dans $L^2(\mathcal{G})$ qui est le plus proche de X .

(ii) Dans le cas particulier où $\mathcal{G} = \sigma(Z)$, le Théorème 3.1 s'énonce comme suit : la fonction mesurable f minimisant la quantité $\mathbb{E}[(X - f(Z))^2]$ est telle que $f(Z) = \mathbb{E}(X | Z)$. Si Z est une constante, on retombe sur le fait que la constante c minimisant $\mathbb{E}[(X - c)^2]$ est $c = \mathbb{E}(X)$.

Preuve. Puisque $\mathbb{E}(X^2) < \infty$, l'inégalité de Jensen (Théorème 2.7) nous dit que $\mathbb{E}(X | \mathcal{G})$ est un élément de $L^2(\mathcal{G})$. Il suffit donc de montrer que $X - \mathbb{E}(X | \mathcal{G})$ est orthogonale de $L^2(\mathcal{G})$. Soit $Z \in L^2(\mathcal{G})$. On sait que XZ admet un moment d'ordre 1, et par le Théorème 1.6, $\mathbb{E}(ZX | \mathcal{G}) = Z \mathbb{E}(X | \mathcal{G})$. Donc

$$\mathbb{E}(ZX) = \mathbb{E}[\mathbb{E}(ZX | \mathcal{G})] = \mathbb{E}[Z \mathbb{E}(X | \mathcal{G})].$$

Autrement dit, $\mathbb{E}[Z(X - \mathbb{E}(X | \mathcal{G}))] = 0$. \square

4. Introduction aux martingales

On aborde la notion de martingale (à temps discret), et étudie quelques propriétés de base des martingales.

Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité. Une filtration $(\mathcal{F}_n)_{n \geq 0}$ sur cet espace est une famille croissante de sous-tribus de \mathcal{F} :

$$\mathcal{F}_0 \subset \mathcal{F}_1 \subset \mathcal{F}_2 \subset \cdots \subset \mathcal{F}.$$

Une suite de variables aléatoires $(X_n, n \geq 0)$ est dite adaptée par rapport à (\mathcal{F}_n) si pour tout n , X_n est \mathcal{F}_n -mesurable.

Définition 4.1. On dit que (X_n) est une martingale [resp. surmartingale; sous-martingale] si

- (i) (X_n) est adapté;
- (ii) $\forall n, \mathbb{E}(|X_n|) < \infty$;
- (iii) $\forall n, \mathbb{E}(X_{n+1} | \mathcal{F}_n) = X_n$, p.s. [resp., $\mathbb{E}(X_{n+1} | \mathcal{F}_n) \leq X_n$; $\mathbb{E}(X_{n+1} | \mathcal{F}_n) \geq X_n$].

Remarquons que si X est une surmartingale, alors $n \mapsto \mathbb{E}(X_n)$ est décroissante.

Exemple 4.2. Soit ξ une variable aléatoire intégrable. Soit $X_n := \mathbb{E}(\xi | \mathcal{F}_n)$. Alors (X_n) est une martingale. □

Exemple 4.3. Soit $(\xi_i)_{i \geq 0}$ une suite de variables aléatoires intégrables et indépendantes. Soient

$$X_n := \sum_{i=0}^n \xi_i, \quad \mathcal{F}_n := \sigma\{X_i, 0 \leq i \leq n\}.$$

Il est clair que (X_n) est adapté, et est intégrable. De plus, $\mathbb{E}(X_{n+1} | \mathcal{F}_n) = X_n + \mathbb{E}(\xi_{n+1})$. Donc X est une martingale si $\mathbb{E}(\xi_n) = 0, \forall n$; est une sous-martingale si $\mathbb{E}(\xi_n) \geq 0, \forall n$; et est une surmartingale si $\mathbb{E}(\xi_n) \leq 0, \forall n$. □

Il est clair que X est une sous-martingale si et seulement si $-X$ est une surmartingale, et que X est une martingale si et seulement si elle est à la fois une sous-martingale et une surmartingale. En voici quelques autres propriétés.

Proposition 4.4. Si (X_n) est une sous-martingale, alors $\forall m > n, \mathbb{E}(X_m | \mathcal{F}_n) \geq X_n$, p.s.

Preuve. On a

$$\mathbb{E}(X_m | \mathcal{F}_n) = \mathbb{E}[\mathbb{E}(X_m | \mathcal{F}_{m-1}) | \mathcal{F}_n] \geq \mathbb{E}[X_{m-1} | \mathcal{F}_n].$$

Ceci démontre le résultat cherché avec un argument par récurrence. □

Proposition 4.5. (i) Si (X_n) et (Y_n) sont des martingales (resp. sous-martingales; sur-martingales), alors $(X_n + Y_n)$ l'est également.

(ii) Si (X_n) et (Y_n) sont des sous-martingales, alors $\max(X_n, Y_n)$ est une sous-martingale.

Preuve. Évidente. □

Proposition 4.6. (i) Si (X_n) est une martingale et si φ est une fonction convexe telle que pour tout n , $\varphi(X_n)$ soit intégrable. Alors $(\varphi(X_n))$ est une sous-martingale.

(ii) Si (X_n) est une sous-martingale et si φ est une fonction convexe croissante telle que pour tout n , $\varphi(X_n)$ soit intégrable. Alors $(\varphi(X_n))$ est une sous-martingale.

Preuve. (i) Il est clair que $(\varphi(X_n))$ est adapté. D'autre part, d'après l'inégalité de Jensen (Théorème 2.7),

$$\mathbb{E}[\varphi(X_{n+1}) \mid \mathcal{F}_n] \geq \varphi(\mathbb{E}[X_{n+1} \mid \mathcal{F}_n]) = \varphi(X_n).$$

Donc $(\varphi(X_n))$ est une sous-martingale.

(ii) Comme φ est croissante et convexe, on a

$$\mathbb{E}[\varphi(X_{n+1}) \mid \mathcal{F}_n] \geq \varphi(\mathbb{E}[X_{n+1} \mid \mathcal{F}_n]) \geq \varphi(X_n).$$

D'où la conclusion. □

Proposition 4.7. (i) Soit (X_n) une martingale et soit $p \geq 1$ un réel. Si $\mathbb{E}(|X_n|^p) < \infty$ pour tout n , alors $(|X_n|^p)$ est une sous-martingale.

En particulier, pour tout $a \in \mathbb{R}$, $(|X_n - a|)$ est une sous-martingale.

(ii) Si (X_n) est une sous-martingale, et si $a \in \mathbb{R}$, alors $(X_n - a)^+$ est une sous-martingale.

Preuve. Il s'agit d'appliquer la Proposition 4.6 à $\varphi(x) := |x|^p$ (fonction convexe) et à $\varphi(x) := (x - a)^+$ (fonction convexe et croissante). □

On note désormais

$$\mathcal{F}_\infty := \bigvee_{n=0}^{\infty} \mathcal{F}_n,$$

la plus petite sous-tribu de \mathcal{F} qui contient toutes les \mathcal{F}_n .

Définition 4.8. Une application $T : \Omega \rightarrow \{0, 1, 2, \dots; \infty\}$ est appelée un temps d'arrêt si pour tout $n \geq 0$, $\{T \leq n\} \in \mathcal{F}_n$.

Remarque. On peut facilement vérifier que T est un temps d'arrêt si et seulement si $\forall n \leq \infty, \{T = n\} \in \mathcal{F}_n$. \square

Définition 4.9. La tribu \mathcal{F}_T est définie par

$$\begin{aligned} \mathcal{F}_T &= \{A \in \mathcal{F}_\infty : \forall n, A \cap \{T = n\} \in \mathcal{F}_n\} \\ &= \{A \in \mathcal{F}_\infty : \forall n, A \cap \{T \leq n\} \in \mathcal{F}_n\}. \end{aligned}$$

Propriété 4.10. (i) Si S et T sont deux temps d'arrêt, alors $S \vee T$ et $S \wedge T$ sont aussi des temps d'arrêt.

(ii) Si S et T sont des temps d'arrêt tels que $S \leq T$, p.s., alors $\mathcal{F}_S \subset \mathcal{F}_T$.

(iii) Si (X_n) est adapté et si T est un temps d'arrêt, alors $X_T \mathbf{1}_{\{T < \infty\}}$ est \mathcal{F}_T -mesurable.

Preuve. (i) On a $\{S \vee T \leq n\} = \{S \leq n\} \cap \{T \leq n\}$, et $\{S \wedge T \leq n\} = \{S \leq n\} \cup \{T \leq n\}$.

(ii) Si $A \in \mathcal{F}_S$, alors $A \cap \{T \leq n\} = (A \cap \{S \leq n\}) \cap \{T \leq n\} \in \mathcal{F}_n$.

(iii) Il suffit de remarquer que $X_T \mathbf{1}_{\{T < \infty\}} = \sum_{n=0}^{\infty} X_n \mathbf{1}_{\{T=n\}}$ est \mathcal{F}_∞ -mesurable, et que pour tout $A \subset \mathbb{R}$ borélien et n , $\{X_T \mathbf{1}_{\{T < \infty\}} \in A\} \cap \{T = n\} = \{X_n \in A\} \cap \{T = n\} \in \mathcal{F}_n$. \square

On s'intéresse maintenant au processus $(X_{T \wedge n})$ arrêté à un temps d'arrêt T .

Théorème 4.11. Si T est un temps d'arrêt, et si (X_n) est une sous-martingale (resp. surmartingale), alors $(X_{T \wedge n})$ est une sous-martingale (resp. surmartingale).

Preuve. Comme $|X_{T \wedge n}| \leq |X_0| + |X_1| + \dots + |X_n|$, $X_{T \wedge n}$ est intégrable. Il est clair que $(X_{T \wedge n})$ est adapté, car $X_{T \wedge n}$ est $\mathcal{F}_{T \wedge n}$ -mesurable, et cette dernière est une sous-tribu de \mathcal{F}_n . Enfin, puisque $\{T \geq n+1\} = \{T \leq n\}^c \in \mathcal{F}_n$,

$$\begin{aligned} \mathbb{E}[(X_{T \wedge (n+1)} - X_{T \wedge n}) | \mathcal{F}_n] &= \mathbb{E}[(X_{n+1} - X_n) \mathbf{1}_{\{T \geq n+1\}} | \mathcal{F}_n] \\ &= \mathbf{1}_{\{T \geq n+1\}} \mathbb{E}[(X_{n+1} - X_n) | \mathcal{F}_n] \geq 0. \end{aligned} \quad \square$$

Avant d'énoncer le résultat principal de cette section, on prouve un résultat préliminaire.

Lemme 4.12. Soient X et Y deux variables aléatoires intégrables, et soit \mathcal{G} une sous-tribu de \mathcal{F} . On a $\mathbb{E}(X | \mathcal{G}) \leq \mathbb{E}(Y | \mathcal{G})$, p.s., si et seulement si $\mathbb{E}(X \mathbf{1}_A) \leq \mathbb{E}(Y \mathbf{1}_A)$ pour tout $A \in \mathcal{G}$.

Preuve. Pour simplifier l'écriture, on pose $\tilde{X} := \mathbb{E}(X | \mathcal{G})$ et $\tilde{Y} := \mathbb{E}(Y | \mathcal{G})$. Pour tout $A \in \mathcal{G}$, on a $\mathbb{E}[X \mathbf{1}_A] - \mathbb{E}[Y \mathbf{1}_A] = \mathbb{E}\{\mathbb{E}[(X - Y) | \mathcal{G}] \mathbf{1}_A\} = \mathbb{E}\{[\tilde{X} - \tilde{Y}] \mathbf{1}_A\}$.

Partie “seulement si”. Supposons que $\tilde{X} \leq \tilde{Y}$ p.s., alors pour tout $A \in \mathcal{G}$, $\mathbb{E}[X \mathbf{1}_A] - \mathbb{E}[Y \mathbf{1}_A] = \mathbb{E}\{[\tilde{X} - \tilde{Y}] \mathbf{1}_A\} \leq 0$.

Partie “si”. Réciproquement, on suppose que $A \in \mathcal{G}$, $\mathbb{E}[X \mathbf{1}_A] \leq \mathbb{E}[Y \mathbf{1}_A]$ pour tout $A \in \mathcal{G}$. On considère $A := \{\omega : \tilde{X} > \tilde{Y}\} \in \mathcal{G}$ (car \tilde{X} et \tilde{Y} sont toutes deux \mathcal{G} -mesurables). On a alors $0 \geq \mathbb{E}[X \mathbf{1}_A] - \mathbb{E}[Y \mathbf{1}_A] = \mathbb{E}\{[\tilde{X} - \tilde{Y}] \mathbf{1}_A\} \geq 0$. Donc $\mathbb{E}\{[\tilde{X} - \tilde{Y}] \mathbf{1}_A\} = 0$. Comme $[\tilde{X} - \tilde{Y}] \mathbf{1}_A \geq 0$ p.s., ceci n'est possible que si $[\tilde{X} - \tilde{Y}] \mathbf{1}_A = 0$ p.s. ; autrement dit, $\tilde{X} \leq \tilde{Y}$ p.s. \square

Théorème 4.13 (théorème d'arrêt). Soient S et T deux temps d'arrêt bornés tels que $S \leq T$. Si (X_n) est une sous-martingale (resp. surmartingale), alors

$$\mathbb{E}(X_T | \mathcal{F}_S) \geq X_S \quad [\text{resp. } \leq X_S] \quad \text{p.s.}$$

Preuve. Supposons que $\mathbb{P}(S \leq T \leq k) = 1$. Alors $|X_T| \leq |X_0| + |X_1| + \dots + |X_k|$ qui est intégrable. Soit $A \in \mathcal{F}_S$. On a

$$\mathbb{E}\{[X_T - X_S] \mathbf{1}_A\} = \sum_{n=0}^k \mathbb{E}\{[X_{T \wedge k} - X_n] \mathbf{1}_{A \cap \{S=n\}}\}.$$

Comme $A \cap \{S = n\} \in \mathcal{F}_n$, et $(X_{T \wedge n})$ est une sous-martingale (Théorème 4.11), on a, d'après le Lemma 4.12,

$$\mathbb{E}\{[X_{T \wedge k} - X_n] \mathbf{1}_{A \cap \{S=n\}}\} \geq \mathbb{E}\{[X_{T \wedge n} - X_n] \mathbf{1}_{A \cap \{S=n\}}\} = 0,$$

la dernière identité provenant du fait que $X_{T \wedge n} = X_n$ sur $\{S = n\}$. On a donc démontré que $\mathbb{E}\{[X_T - X_S] \mathbf{1}_A\} \geq 0$. \square

Corollaire 4.14. Soit T un temps d'arrêt tel que $\mathbb{P}(T \leq k) = 1$. Si (X_n) est une sous martingale, alors

$$\mathbb{E}(X_0) \leq \mathbb{E}(X_T) \leq \mathbb{E}(X_k).$$

En particulier, si (X_n) est une martingale, alors $\mathbb{E}(X_T) = \mathbb{E}(X_0)$.

Chapitre 12. Vecteurs aléatoires gaussiens

1. Définition et propriétés

On rappelle que la loi gaussienne de moyenne $\mu \in \mathbb{R}$ et de variance $\sigma^2 > 0$ a pour densité $\frac{1}{\sigma\sqrt{2\pi}} \exp(-\frac{(x-\mu)^2}{2\sigma^2})$, $x \in \mathbb{R}$. Sa fonction caractéristique vaut $\exp(i\mu t - \frac{\sigma^2}{2}t^2)$, $t \in \mathbb{R}$. Il est commode de convenir qu'une masse de Dirac δ_μ est la loi gaussienne de moyenne μ et de variance nulle.

Définition 1.1. Soit $X = \begin{pmatrix} X_1 \\ \vdots \\ X_N \end{pmatrix}$ une v.a. à valeurs dans \mathbb{R}^N (donc un vecteur aléatoire de dimension N). On dit que X est un vecteur gaussien si toute combinaison linéaire de ses coordonnées (c'est-à-dire $\sum_{j=1}^N \lambda_j X_j = \lambda_1 X_1 + \dots + \lambda_N X_N$ pour $\lambda = \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_N \end{pmatrix} \in \mathbb{R}^N$) suit une loi gaussienne.

Remarque. (i) Si $X = \begin{pmatrix} X_1 \\ \vdots \\ X_N \end{pmatrix}$ est un vecteur gaussien, alors chaque coordonnée est une v.a. gaussienne réelle.

(ii) Attention, la réciproque est **fausse**. Par exemple, soient Y et ε deux variables aléatoires indépendantes telles que Y suive la loi gaussienne centrée réduite $\mathcal{N}(0, 1)$ et que $\mathbb{P}(\varepsilon = 1) = \mathbb{P}(\varepsilon = -1) = 1/2$. Soit $W = \varepsilon Y$. Alors (Y, W) n'est pas un vecteur gaussien car $Y + W$ n'est pas une v.a. gaussienne (en effet, $\mathbb{P}(Y + W = 0) = 1/2$). Pourtant chacune des composantes est une v.a. gaussienne (il a été démontré dans l'Exemple 3.4 du Chapitre 6 que W est une v.a. gaussienne). \square

Exemple 1.2. Si X_1, \dots, X_N sont des v.a. gaussiennes (réelles) indépendantes, alors (X_1, \dots, X_N) est un vecteur gaussien, car on sait que la somme de v.a. gaussiennes indépendantes est gaussienne. \square

Exemple 1.3. Si $X = \begin{pmatrix} X_1 \\ \vdots \\ X_N \end{pmatrix}$ est un vecteur gaussien, A est une matrice $M \times N$ déterministe, et $B \in \mathbb{R}^M$, alors $AX + B$ est un vecteur gaussien, car les combinaisons linéaires des composantes de $AX + B$ sont des combinaisons linéaires (plus des constantes) de X_1, \dots, X_N . \square

Propriété 1.4. La fonction caractéristique d'un vecteur gaussien X est donnée par

$$\varphi_X(t) = \exp \left(i \sum_{j=1}^N \mu_j t_j - \frac{1}{2} \sum_{1 \leq j, k \leq N} D_{jk} t_j t_k \right), \quad t \in \mathbb{R}^N,$$

où $D = (D_{jk})_{N \times N}$ est la matrice de covariances de X . En conséquence, la loi d'un vecteur gaussien est complètement déterminée par sa moyenne et sa matrice de covariances. On

notera $\mathcal{N}(\mu, D)$, où $\mu := \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_N \end{pmatrix}$.

Preuve. Soit $t = \begin{pmatrix} t_1 \\ \vdots \\ t_N \end{pmatrix} \in \mathbb{R}^N$, et soit $Y := \sum_{j=1}^N t_j X_j$. On a

$$\varphi_X(t) = \mathbb{E}(e^{iY}) = \varphi_Y(1).$$

Or, Y est une v.a. gaussienne réelle, $\mathbb{E}(Y) = \sum_{j=1}^N t_j \mu_j$ et $\text{Var}(Y) = \sum_{1 \leq j, k \leq N} t_j t_k D_{jk}$, ce qui nous donne

$$\varphi_X(t) = \exp \left(i \sum_{j=1}^N t_j \mu_j - \frac{1}{2} \sum_{1 \leq j, k \leq N} t_j t_k D_{jk} \right). \quad \square$$

Nous montrons maintenant comment construire un vecteur gaussien de moyenne et de matrice de covariances données.

Propriété 1.5. Soit $\mu \in \mathbb{R}^N$, et soit D une matrice $N \times N$ symétrique positive (c'est-à-dire que $\sum_{1 \leq j, k \leq N} \lambda_j \lambda_k D_{jk} \geq 0$ pour tout $\lambda \in \mathbb{R}^N$). Alors, il existe un vecteur gaussien N -dimensionnel de moyenne μ et de matrice de covariances D .

Preuve. On sait construire N variables gaussiennes centrées réduites indépendantes, $\mathcal{N}_1, \dots, \mathcal{N}_N$. On note dans la preuve

$$\mu = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_N \end{pmatrix}, \quad \mathcal{N} = \begin{pmatrix} \mathcal{N}_1 \\ \vdots \\ \mathcal{N}_N \end{pmatrix}.$$

Toute matrice symétrique positive admettant une racine carrée, on peut trouver une matrice symétrique $C = (C_{jk})_{N \times N}$ telle que $C^2 = D$. Posons

$$X = C\mathcal{N} + \mu.$$

(Attention à la notation : X est en colonne.) D'après les Exemples 1.2 et 1.3, X est un vecteur gaussien, de moyenne μ . Pour déterminer sa matrice de covariances, remarquons que

$$\begin{aligned} \text{Cov}(X_j, X_k) &= \mathbb{E}((X_j - \mu_j)(X_k - \mu_k)) \\ &= \mathbb{E} \left[\left(\sum_{\ell=1}^N C_{j\ell} \mathcal{N}_\ell \right) \left(\sum_{m=1}^N C_{km} \mathcal{N}_m \right) \right] \\ &= \sum_{1 \leq \ell, m \leq N} C_{j\ell} C_{km} \mathbb{E}(\mathcal{N}_\ell \mathcal{N}_m). \end{aligned}$$

Or, $\mathbb{E}(\mathcal{N}_\ell \mathcal{N}_m)$ vaut 1 si $\ell = m$ et vaut 0 sinon, ce qui implique que $\text{Cov}(X_j, X_k) = \sum_{\ell=1}^N C_{j\ell} C_{k\ell} = (CC^t)_{jk}$, où C^t est le transposé de C . Donc la matrice de covariances de X est CC^t .

Comme C est symétrique, $CC^t = C^2 = D$, on a montré que X est bien un vecteur gaussien de moyenne μ et de matrice de covariances D . \square

Propriété 1.6 (Densité gaussienne). Soit μ un vecteur quelconque de \mathbb{R}^N et D une matrice $N \times N$ symétrique positive. Si $\det(D) \neq 0$, et si $D^{-1} = (D_{jk}^{-1})_{N \times N}$ désigne la matrice inverse de D , alors la loi gaussienne N -dimensionnelle $\mathcal{N}(\mu, D)$ est absolument continue par rapport à la mesure de Lebesgue sur \mathbb{R}^N , et a pour densité

$$\frac{1}{(2\pi)^{N/2} \sqrt{\det(D)}} \exp \left(-\frac{1}{2} \sum_{1 \leq j, k \leq N} D_{jk}^{-1} (x_j - \mu_j)(x_k - \mu_k) \right).$$

Reparque. Attention, $D_{jk}^{-1} \neq 1/D_{jk}$ en général.

Preuve. Soit $X = CN + \mu$ comme dans la construction précédente. On connaît la densité de \mathcal{N} , car $\mathcal{N}_1, \dots, \mathcal{N}_N$ sont des v.a. gaussiennes indépendantes. On obtient alors la densité de X en effectuant un changement de variables. \square

Théorème 1.7. Soit (X_1, \dots, X_N) un vecteur gaussien. Pour que les variables aléatoires X_1, \dots, X_N soient indépendantes, il faut et il suffit que la matrice de covariances de X soit diagonale.

Preuve. La condition est trivialement nécessaire. Pour la réciproque, on utilise de nouveau la construction $X = CN + \mu$. Si D (la matrice de covariances de X) est diagonale, alors C est aussi diagonale : $C = \text{diag}(C_{11}, \dots, C_{NN})$. On a alors $X_j = C_{jj}\mathcal{N}_j + \mu_j$ pour $1 \leq j \leq N$. Comme $\mathcal{N}_1, \dots, \mathcal{N}_N$ sont des v.a. (gaussiennes) indépendantes, on déduit l'indépendance entre les v.a. X_1, \dots, X_N . \square

Remarque. La preuve du théorème montre que si $(X_1, \dots, X_N, Y_1, \dots, Y_M, Z_1, \dots, Z_L)$ est un vecteur gaussien, alors (X_1, \dots, X_N) et (Y_1, \dots, Y_M) sont indépendants si et seulement si $\text{Cov}(X_i, Y_j) = 0$ quels que soient $1 \leq i \leq N$ et $1 \leq j \leq M$. \square

Exemple 1.8. Si un couple (X, Y) de v.a. réelles est gaussien, alors X et Y sont indépendantes si et seulement si $\text{Cov}(X, Y) = 0$. On se gardera bien de croire que si **individuellement** X et Y sont des v.a. réelles gaussiennes telles que $\text{Cov}(X, Y) = 0$, alors X et Y sont indépendantes (on pensera à l'Exemple 3.4 du Chapitre 6). Pour qu'il en soit ainsi, il est nécessaire que le **couple** (X, Y) soit un vecteur gaussien, ce qui n'a rien d'automatique. \square

Exemple 1.9. Soient X et Y deux variables aléatoires indépendantes telles que $P_X = P_Y = \mathcal{N}(\mu, \sigma^2)$. Il a été démontré dans l'Exemple 3.6 du Chapitre 6, à l'aide de fonction caractéristique, que $X + Y$ et $X - Y$ sont indépendantes. On redémontre ce résultat maintenant, à l'aide de propriétés de vecteurs gaussiens.

Soient a, b, c et d des réels quelconques. D'après les Exemples 1.2 et 1.3, $(aX + bY, cX + dY)$ est un vecteur gaussien. Comme $\text{Cov}(aX + bY, cX + dY) = (ac + bd)\sigma^2$, le Théorème 1.7 nous dit que $aX + bY$ et $cX + dY$ sont indépendantes si et seulement si $ac + bd = 0$, ce qui est le cas par exemple lorsque $a = b = c = 1$ et $d = -1$. \square

Exemple 1.10. Soient U_1, U_2, \dots des variables aléatoires indépendantes suivant la même loi gaussienne $\mathcal{N}(0, 1)$. Soient a_0, a_1, a_2, \dots des réels tels que $a_j a_{j+1} = 0$ pour tout $j \geq 0$,

et que la série $\sum a_n^2$ converge. On pose

$$V_n = \sum_{i=1}^n a_{n-i} U_i, \quad n = 1, 2, \dots$$

- (i) Étudier la convergence en loi de V_n .
- (ii) Les variables V_n et V_{n+1} sont-elles indépendantes?
- (iii) Étudier la convergence dans $L^2(\Omega, \mathcal{F}, \mathbb{P})$ de V_n .

Solution. (i) La variable V_n suit une loi gaussienne (somme de variables gaussiennes indépendantes), $\mathbb{E}(V_n) = 0$ et $\text{Var}(V_n) = \sum_{i=1}^n a_{n-i}^2 = \sum_{j=0}^{n-1} a_j^2$. Donc sa fonction caractéristique vaut $\varphi_{V_n}(t) = \exp(-\frac{t^2}{2} \sum_{j=0}^{n-1} a_j^2) \rightarrow \exp(-\frac{t^2}{2} \sum_{j=0}^{\infty} a_j^2)$, $n \rightarrow \infty$. Donc V_n converge en loi vers la gaussienne $\mathcal{N}(0, \sum_{j=0}^{\infty} a_j^2)$. (Cas dégénéré : V_n converge en loi vers 0 si $a_j = 0, \forall j \geq 0$).

(ii) Toute combinaison linéaire de V_n et de V_{n+1} donne une combinaison linéaire de U_1, U_2, \dots, U_{n+1} qui est une variable aléatoire gaussienne (somme de variables gaussiennes indépendantes). Par définition, (V_n, V_{n+1}) est un vecteur gaussien, et V_n et V_{n+1} sont indépendantes si et seulement si leur covariance s'annule. Or, $\text{Cov}(V_n, V_{n+1}) = \sum_{i=1}^n a_{n-i} a_{n+1-i} = 0$, on déduit que V_n est indépendante de V_{n+1} .

(iii) Si V_n converge dans L^2 , alors $\mathbb{E}[(V_n - V_{n+1})^2] \rightarrow 0$ ($n \rightarrow \infty$). Or, V_n et V_{n+1} sont indépendantes, on a $\mathbb{E}[(V_n - V_{n+1})^2] = \text{Var}(V_n) + \text{Var}(V_{n+1}) = \sum_{j=0}^{n-1} a_j^2 + \sum_{j=0}^n a_j^2 \rightarrow 2 \sum_{j=0}^{\infty} a_j^2$, ce qui n'est possible que si les a_j sont identiquement nuls.

En conclusion, V_n converge dans L^2 si et seulement si $a_j = 0, \forall j \geq 0$ (dans ce cas, V_n converge dans L^2 vers 0). □

Exemple 1.11. Soit un vecteur aléatoire gaussien X dont les $n + 1$ composantes Z, X_1, \dots, X_n sont des variables centrées réduites telles que (α étant une constante connue)

$$\begin{aligned} \rho(Z, X_i) &= \alpha, & 1 \leq i \leq n, \\ \rho(X_i, X_j) &= \alpha^2, & 1 \leq i \neq j \leq n. \end{aligned}$$

- (i) On pose

$$U_i = \frac{X_i - \alpha Z}{\sqrt{1 - \alpha^2}}, \quad 1 \leq i \leq n.$$

Déterminer la loi du vecteur (Z, U_1, \dots, U_n) et en déduire celle du vecteur (Z, X_1, \dots, X_n) .

- (ii) Loi de $S = \sum_{i=1}^n X_i$? Loi de $T = \frac{S}{Z}$?

Solution. (i) Le vecteur aléatoires (Z, U_1, \dots, U_n) étant une transformation linéaire du vecteur gaussien X , est un vecteur gaussien, centré. Pour déterminer sa loi, il suffit de déterminer la matrice de covariances. Pour $1 \leq i \leq n$,

$$\text{Var}(U_i) = \frac{\text{Var}(X_i) + \alpha^2 \text{Var}(Z) - 2\alpha \text{Cov}(X_i, Z)}{1 - \alpha^2} = 1,$$

et

$$\text{Cov}(Z, U_i) = \frac{\text{Cov}(Z, X_i) - \alpha \text{Var}(Z)}{\sqrt{1 - \alpha^2}} = 0.$$

Si $1 \leq i \neq j \leq n$,

$$\text{Cov}(U_i, U_j) = \frac{\text{Cov}(X_i, X_j) + \alpha^2 \text{Var}(Z) - \alpha \text{Cov}(X_i, Z) - \alpha \text{Cov}(X_j, Z)}{1 - \alpha^2} = 0.$$

Donc (Z, U_1, \dots, U_n) suit la loi gaussienne $\mathcal{N}(0, \text{Id})$, où Id désigne la matrice $(n+1) \times (n+1)$ identité. Autrement dit, Z, U_1, \dots, U_n sont des variables indépendantes suivant la même loi gaussienne centrée réduite.

Le vecteur aléatoire (Z, X_1, \dots, X_n) étant une transformation linéaire du vecteur gaussien (Z, U_1, \dots, U_n) , est donc aussi un vecteur gaussien. On a (en écrivant $\beta := \sqrt{1 - \alpha^2}$)

$$\begin{pmatrix} Z \\ X_1 \\ \vdots \\ X_n \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ \alpha & \beta & 0 & \cdots & 0 \\ \alpha & 0 & \beta & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \alpha & 0 & 0 & \cdots & \beta \end{pmatrix} \begin{pmatrix} Z \\ U_1 \\ \vdots \\ U_n \end{pmatrix} := A \cdot \begin{pmatrix} Z \\ U_1 \\ \vdots \\ U_n \end{pmatrix}.$$

Donc la matrice de covariances de $\begin{pmatrix} Z \\ X_1 \\ \vdots \\ X_n \end{pmatrix}$ est

$$\begin{aligned} D := AA^t &= \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ \alpha & \beta & 0 & \cdots & 0 \\ \alpha & 0 & \beta & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \alpha & 0 & 0 & \cdots & \beta \end{pmatrix} \begin{pmatrix} 1 & \alpha & \alpha & \cdots & \alpha \\ 0 & \beta & 0 & \cdots & 0 \\ 0 & 0 & \beta & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \beta \end{pmatrix} \\ &= \begin{pmatrix} 1 & \alpha & \alpha & \cdots & \alpha \\ \alpha & 1 & \alpha^2 & \cdots & \alpha^2 \\ \alpha & \alpha^2 & 1 & \cdots & \alpha^2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \alpha & \alpha^2 & \alpha^2 & \cdots & 1 \end{pmatrix}. \end{aligned}$$

Autrement dit, le vecteur aléatoire (Z, X_1, \dots, X_n) suit la loi gaussienne $\mathcal{N}(0, D)$.

(ii) Par définition,

$$S = n\alpha Z + \sqrt{1 - \alpha^2} \sum_{i=1}^n U_i,$$

qui suit la loi gaussienne $\mathcal{N}(0, n^2\alpha^2 + n(1 - \alpha^2))$.

D'autre part,

$$T = \frac{S}{Z} = n\alpha + \sqrt{n(1 - \alpha^2)} \frac{n^{-1/2} \sum_{i=1}^n U_i}{Z}.$$

La variable T a donc la même loi que $n\alpha + \sqrt{n(1 - \alpha^2)} C$, où C désigne une variable aléatoire suivant la loi de Cauchy standard. La densité de T vaut

$$f_T(x) = \frac{1}{\pi} \frac{\sqrt{n(1 - \alpha^2)}}{n(1 - \alpha^2) + (x - n\alpha)^2}, \quad x \in \mathbb{R}. \quad \square$$

2. Théorème central limite multi-dimensionnel

Soit $X = (X^{(1)}, \dots, X^{(N)})$ une v.a. à valeurs dans \mathbb{R}^N . On suppose que $\mathbb{E}(\|X\|^2) < \infty$. Soit μ le vecteur-espérance de X , et soit D sa matrice de covariances. On sait que D est une matrice symétrique positive. En conséquence, il existe une unique loi sur \mathbb{R}^N qui soit gaussienne de moyenne nulle et de matrice de covariances D ; on la notera $\mathcal{N}(0, D)$.

On sait d'après le théorème central limite uni-dimensionnel que pour chaque coordonnée j , si on se donne une suite de v.a. $X_1^{(j)}, \dots, X_n^{(j)}, \dots$ indépendantes et de même loi que $X^{(j)}$, alors

$$\frac{X_1^{(j)} + \dots + X_n^{(j)} - n\mu_j}{\sqrt{n}} \rightarrow \mathcal{N}(0, D_{jj}) \quad (\text{en loi}).$$

En revanche, le théorème central limite uni-dimensionnel ne permet pas de conclure quant à la convergence des vecteurs aléatoires (convergence conjointe des coordonnées). Ce point fait l'objet la version multi-dimensionnelle du théorème central limite :

Théorème 2.1. Soient X_1, \dots, X_n, \dots des vecteurs aléatoires indépendants, ayant tous la même loi que X . Alors

$$\frac{X_1 + \dots + X_n - n\mu}{\sqrt{n}}$$

converge en loi quand $n \rightarrow \infty$ vers $\mathcal{N}(0, D)$.

La preuve du théorème est très proche de celle que nous avons donnée en dimension 1, en montrant la convergence des fonctions caractéristiques. L'écriture est plus longue.

3. Espérances conditionnelles

Théorème 3.1. Soit (X, Y) un vecteur aléatoire gaussien à valeurs dans \mathbb{R}^2 , alors $\mathbb{E}(X | Y) = aY + b$, où a et b sont deux réels.

Preuve. Puisque X est une v.a. gaussienne réelle, elle admet des moments de tous ordres. D'après le Théorème 3.1 du Chapitre 11, $\mathbb{E}(X | Y)$ est la projection orthogonale de X sur $L^2(Y)$.

Si Y est dégénérée (P_Y est une masse de Dirac), il n'y a rien à démontrer, car $\sigma(Y) = \{\Omega, \emptyset\}$, et $\mathbb{E}(X | Y)$, qui est $\sigma(Y)$ -mesurable, est une v.a. constante.

Supposons maintenant $\text{Var}(Y) > 0$. Soit $a \in \mathbb{R}$. Le vecteur $(X - aY, Y)$ est encore gaussien (car toute combinaison linéaire de $X - aY$ et de Y donne une v.a. gaussienne). Si

$$a := \frac{\text{Cov}(X, Y)}{\text{Var}(Y)},$$

alors $\text{Cov}(X - aY, Y) = \text{Cov}(X, Y) - a\text{Var}(Y) = 0$. D'après le Théorème 1.6, $X - aY$ et Y sont indépendantes. Donc $\mathbb{E}(X - aY | Y) = \mathbb{E}(X - aY) := b$. On a,

$$\mathbb{E}(X | Y) = \mathbb{E}(aY | Y) + \mathbb{E}(X - aY | Y) = aY + b,$$

ce qu'il fallait démontrer. □

Remarque. On peut facilement généraliser le théorème pour le cas où Y est à valeurs dans \mathbb{R}^N dont la matrice de covariances est inversible. □

Exemple 3.2. La preuve du Théorème 3.1 donne la valeur exacte de a et b : $a = \text{Cov}(X, Y)/\text{Var}(Y)$ et $b = \mathbb{E}(X) - a\mathbb{E}(Y)$. Dans la pratique, il est facile de retrouver ces valeurs “à la main”. Par exemple, soit (X, Y) un vecteur aléatoire gaussien tel que $\mathbb{E}(X) = 3$, $\mathbb{E}(Y) = 0$, $\text{Var}(X) = 9$, $\text{Var}(Y) = 1$ et $\text{Cov}(X, Y) = -2$. Par le Théorème 2.1, on peut écrire $\mathbb{E}(X | Y) = aY + b$. En prenant l'espérance dans les deux côtés, on obtient $\mathbb{E}(X) = a\mathbb{E}(Y) + b$, c'est-à-dire $b = 3$.

D'autre part, $\mathbb{E}(XY | Y) = Y\mathbb{E}(X | Y) = aY^2 + bY$. D'où $\mathbb{E}(XY) = a\mathbb{E}(Y^2) + b\mathbb{E}(Y)$, c'est-à-dire $a = -2$. Donc

$$\mathbb{E}(X | Y) = -2Y + 3. \quad \square$$

Exemple 3.3. Dans l'exemple précédent, si l'on cherche les réels a , b , c et d qui minimisent la quantité $\mathbb{E}[(X - a - bY - cY^2 - dY^3)^2]$, on sait que $c = d = 0$, $a = 3$ et $b = -2$. La

valeur minimale de la quantité vaut alors

$$\begin{aligned}\mathbb{E}[(X - 3 + 2Y)^2] &= \mathbb{E} [((X - \mathbb{E}X) + 2(Y - \mathbb{E}Y))^2] \\ &= \text{Var}(X) + 4\text{Var}(Y) + 4\text{Cov}(X, Y) = 5.\end{aligned}$$

On peut retrouver ce résultat en faisant un calcul direct. \square

Exemple 3.4. Soit (X, Y) un vecteur aléatoire gaussien tel que $\mathbb{E}(X) = \mathbb{E}(Y) = 0$, $\mathbb{E}(X^2) = \mathbb{E}(Y^2) = 1$ et $\mathbb{E}(XY) = \varrho \in]-1, 1[$. Montrer que

$$\mathbb{P}(X > 0, Y > 0) = \frac{1}{4} + \frac{1}{2\pi} \arcsin(\varrho).$$

Il est possible de prouver l'identité en utilisant la densité exacte de (X, Y) , mais il est plus commode d'utiliser l'espérance conditionnelle. En effet, $Z := X - \varrho Y$ est une v.a. gaussienne indépendante de Y , car (Z, Y) est un vecteur gaussien tel que $\text{Cov}(Z, Y) = 0$. De plus, $P_Z = \mathcal{N}(0, 1 - \varrho^2)$. Donc par les Théorèmes 1.6 et 2.3 du chapitre précédent,

$$\begin{aligned}\mathbb{P}(X > 0, Y > 0 | Y) &= \mathbf{1}_{\{Y > 0\}} \mathbb{E}(\mathbf{1}_{\{Z > -\varrho Y\}} | Y) \\ &= \mathbf{1}_{\{Y > 0\}} \left\{ 1 - \Phi\left(-\frac{\varrho Y}{\sqrt{1 - \varrho^2}}\right) \right\},\end{aligned}$$

où $\Phi(x) := \frac{1}{(2\pi)^{1/2}} \int_{-\infty}^x e^{-u^2/2} du$ désigne la fonction de répartition de la gaussienne centrée réduite $\mathcal{N}(0, 1)$. Par symétrie, $1 - \Phi(-x) = \Phi(x)$. Donc

$$\begin{aligned}\mathbb{P}(X > 0, Y > 0) &= \mathbb{E} \left\{ \mathbf{1}_{\{Y > 0\}} \Phi\left(\frac{\varrho Y}{\sqrt{1 - \varrho^2}}\right) \right\} \\ &= \int_0^\infty \Phi\left(\frac{\varrho u}{\sqrt{1 - \varrho^2}}\right) \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du.\end{aligned}$$

Étudions la fonction $h(a) := \int_0^\infty \Phi(au) \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du$, $a \in \mathbb{R}$. Par convergence dominée,

$$h'(a) = \int_0^\infty \frac{u}{\sqrt{2\pi}} e^{-(au)^2/2} \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du = \frac{1}{2\pi(a^2 + 1)}, \quad a \in \mathbb{R}.$$

Comme $h(0) = \int_0^\infty \frac{1}{2} \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du = \frac{1}{4}$, on obtient

$$h(a) = \frac{1}{2\pi} \arctan(a) + \frac{1}{4}, \quad a \in \mathbb{R}.$$

Par conséquent,

$$\begin{aligned}\mathbb{P}(X > 0, Y > 0) &= h\left(\frac{\varrho}{\sqrt{1 - \varrho^2}}\right) = \frac{1}{2\pi} \arctan\left(\frac{\varrho}{\sqrt{1 - \varrho^2}}\right) + \frac{1}{4} \\ &= \frac{1}{2\pi} \arcsin(\varrho) + \frac{1}{4},\end{aligned}$$

ce qui donne l'identité cherchée. \square

Chapitre 13. Exemples de processus aléatoires

Un processus aléatoire (ou : processus stochastique) à temps discret est une suite de variables aléatoires $X_0, X_1, \dots, X_n, \dots$. Il décrit l'évolution d'un phénomène aléatoire si l'on considère les indices $0, 1, 2, \dots$ comme du temps.

Ce chapitre ne fait pas partie du programme de l'examen.

1. Marche aléatoire simple

Soient ξ_1, ξ_2, \dots des v.a. iid qui prennent les valeurs $+1$ ou -1 avec probabilité $1/2$. Définissons $X_0 = 0$ et

$$X_n = \sum_{j=1}^n \xi_j, \quad n = 1, 2, \dots$$

On peut considérer X_n comme la position à l'instant n d'une particule qui se déplace de manière aléatoire sur \mathbb{Z} .

On peut facilement étendre ce modèle en dimension quelconque. Soient ξ_1, ξ_2, \dots des v.a. iid à valeurs dans \mathbb{Z}^d , avec

$$\begin{aligned} \mathbb{P}(\xi_1 = (1, 0, \dots, 0)) &= \mathbb{P}(\xi_1 = (-1, 0, \dots, 0)) \\ &= \mathbb{P}(\xi_1 = (0, 1, \dots, 0)) = \mathbb{P}(\xi_1 = (0, -1, \dots, 0)) \\ &= \dots \\ &= \mathbb{P}(\xi_1 = (0, 0, \dots, 1)) = \mathbb{P}(\xi_1 = (0, 0, \dots, -1)) \\ &= \frac{1}{2d}. \end{aligned}$$

Soit $X_0 = 0$ et soit

$$X_n = \sum_{j=1}^n \xi_j, \quad n = 1, 2, \dots$$

Le processus stochastique $\{X_0, X_1, \dots, X_n, \dots\}$ est appelé “marche aléatoire simple (et symétrique)” en dimension d .

Le théorème central limite multi-dimensionnel nous confirme que

$$\frac{X_n}{\sqrt{n}} \rightarrow \mathcal{N}(0, D), \quad \text{en loi,}$$

avec

$$D = \begin{pmatrix} 1/d & 0 & \cdots & 0 \\ 0 & 1/d & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1/d \end{pmatrix}.$$

Comme D est une matrice diagonale, X_n/\sqrt{n} converge en loi vers un vecteur gaussien dont les composantes sont indépendantes.

On se pose maintenant la question si la marche aléatoire revient infiniment souvent à l'origine. Soit

$$p_n = \mathbb{P}(X_n = 0).$$

Il est évident que $p_n = 0$ si n est impair. En revanche, si n est pair, soit $n = 2k$,

$$\begin{aligned} p_{2k} &= \left(\frac{1}{2d}\right)^{2k} \sum_{k_1 + \dots + k_d = k, k_i \geq 0} \frac{(2k)!}{(k_1! \cdots k_d!)^2} \\ &= \left(\frac{1}{2d}\right)^{2k} \binom{2k}{k} \sum_{k_1 + \dots + k_d = k, k_i \geq 0} \left(\frac{k!}{k_1! \cdots k_d!}\right)^2. \end{aligned}$$

En particulier, on a

$$\begin{aligned} p_{2k} &= 2^{-2k} \binom{2k}{k}, & \text{si } d = 1, \\ p_{2k} &= 4^{-2k} \binom{2k}{k}^2, & \text{si } d = 2. \end{aligned}$$

A l'aide de la formule de Stirling, on obtient : lorsque $k \rightarrow \infty$,

$$\begin{aligned} p_{2k} &\sim \frac{1}{\sqrt{\pi k}}, & \text{si } d = 1, \\ p_{2k} &\sim \frac{1}{\pi k}, & \text{si } d = 2. \end{aligned}$$

En général, on a pour tout $d \geq 1$,

$$p_{2k} \sim \frac{1}{2^{d-1}} \left(\frac{d}{\pi k}\right)^{d/2}, \quad k \rightarrow \infty,$$

(pour cette dernière équivalence, voir par exemple le livre de Rényi, “*Calcul des Probabilités*”, Dunod, pp. 471–472). Donc si $d \geq 3$, on a

$$\sum_n p_n < \infty.$$

Par le lemme de Borel–Cantelli, presque sûrement, lorsque n est suffisamment grand, $X_n \neq 0$. Donc, avec probabilité 1, la marche aléatoire ne visite l’origine qu’un nombre fini de fois.

Lorsque $d = 1$ ou 2 , $\sum_n p_n = \infty$, on ne peut pas directement conclure, car les événements $\{X_n = 0\}$ ne sont pas indépendants. On va faire autrement. Soit

$$q_n = \mathbb{P}(X_1 \neq 0, X_2 \neq 0, \dots, X_{n-1} \neq 0, X_n = 0).$$

(En mots, q_n représente la probabilité que, à l’instant n , la marche aléatoire revienne à l’origine pour la première fois). Il est clair que $q_n = 0$ si n est impair. On a

$$\begin{aligned} p_{2k} &= \mathbb{P}\left(\bigcup_{j=1}^k \{X_1 \neq 0, X_2 \neq 0, \dots, X_{2j-1} \neq 0, X_{2j} = 0, X_{2k} = 0\}\right) \\ &= q_{2k} + \sum_{j=1}^{k-1} \mathbb{P}\{X_1 \neq 0, X_2 \neq 0, \dots, X_{2j-1} \neq 0, X_{2j} = 0, \xi_{2j+1} + \dots + \xi_{2k} = 0\} \\ &= q_{2k} + \sum_{j=1}^{k-1} q_{2j} p_{2k-2j} \\ (*) \quad &= q_{2k} + \sum_{\ell=1}^{k-1} q_{2k-2\ell} p_{2\ell}. \end{aligned}$$

Posons deux séries entières

$$\begin{aligned} f(x) &= \sum_{k=1}^{\infty} p_{2k} x^k, \\ g(x) &= \sum_{k=1}^{\infty} q_{2k} x^k, \quad |x| < 1. \end{aligned}$$

On a donc par (*),

$$f(x) = g(x) + f(x)g(x),$$

ou alors

$$g(x) = \frac{f(x)}{1 + f(x)}, \quad |x| < 1.$$

On sait que

$$\lim_{x \rightarrow 1^-} f(x) = \sum_{k=1}^{\infty} p_{2k}, \quad \lim_{x \rightarrow 1^-} g(x) = \sum_{k=1}^{\infty} q_{2k}.$$

Puisque $\sum_{k=1}^{\infty} p_{2k} = \infty$ (si $d = 1$ ou 2), on a $\lim_{x \rightarrow 1^-} f(x) = \infty$, donc $\lim_{x \rightarrow 1^-} g(x) = 1$. C'est-à-dire que $\sum_{k=1}^{\infty} q_{2k} = 1$. Ceci signifie que lorsque $d = 1$ ou 2 , la marche aléatoire revient à l'origine avec probabilité 1. Donc elle revient infiniment souvent à l'origine. On peut maintenant énoncer le résultat suivant.

Théorème 1.1 (Pólya 1921). (i) En dimension $d = 1$ ou 2 , avec probabilité 1 la marche aléatoire visite infiniment souvent l'origine.

(ii) En dimension $d \geq 3$, avec probabilité 1, la marche aléatoire ne visite l'origine qu'un nombre fini de fois. D'une façon équivalente, la probabilité que la marche aléatoire revienne à l'origine est strictement inférieure à 1.

2. Chaînes de Markov

Dans toute la section, X_0, X_1, \dots désigne une suite de variables aléatoires à valeurs dans E . On suppose que l'espace d'état E est au plus dénombrable.

Définition 2.1. On dit que X_0, \dots, X_n, \dots est une chaîne de Markov (homogène) s'il existe des nombres réels positifs $(P_{ij})_{(i,j) \in E^2}$ avec $\sum_{j \in E} P_{ij} = 1$ pour tout $i \in E$, tels que pour tout entier $n \geq 0$, tous $i, j, x_0, \dots, x_n \in E$:

$$\mathbb{P}(X_{n+1} = j \mid X_0 = x_0, \dots, X_{n-1} = x_{n-1}, X_n = i) = P_{ij},$$

pourvu que la probabilité conditionnelle soit bien définie. On appelle $(P_{ij})_{j \in E}$ la probabilité de transition pour l'état i . Lorsque l'espace d'état E est fini, $P = (P_{ij})_{(i,j) \in E^2}$ est une matrice qu'on appelle la matrice de transition de la chaîne.

Il est intuitivement clair, et facile de vérifier directement, que pour une chaîne de Markov, on a

$$\mathbb{P}(X_{n+1} = j \mid X_0 = x_0, \dots, X_{n-1} = x_{n-1}, X_n = i) = \mathbb{P}(X_{n+1} = j \mid X_n = i).$$

On dit que l'évolution de la chaîne après la date n ne dépend du passé X_0, \dots, X_n qu'à travers sa position au temps n (et non pas du trajet qu'elle a suivi pour atteindre cet état). Cette propriété est appelée **propriété de Markov**.

Voyons maintenant des exemples concrets.

Exemple 2.2 (marche aléatoire simple). Soit X_0, X_1, \dots une marche aléatoire simple et symétrique sur \mathbb{Z} . Lorsque l'événement $\{X_0 = x_0, \dots, X_{n-1} = x_{n-1}, X_n = i\}$ est non-vide (c'est-à-dire que $|x_k - x_{k-1}| = 1$ et $|x_{n-1} - i| = 1$), on a

$$\begin{aligned} & \mathbb{P}(X_{n+1} = j \mid X_0 = x_0, \dots, X_{n-1} = x_{n-1}, X_n = i) \\ &= \mathbb{P}(X_n + \xi_{n+1} = j \mid X_0 = x_0, \dots, X_{n-1} = x_{n-1}, X_n = i) \\ &= \mathbb{P}(\xi_{n+1} = j - i \mid X_0 = x_0, \dots, X_{n-1} = x_{n-1}, X_n = i) \\ &= \mathbb{P}(\xi_{n+1} = j - i) \\ &= \begin{cases} 1/2, & \text{si } |j - i| = 1, \\ 0, & \text{sinon.} \end{cases} \end{aligned}$$

Donc X_0, X_1, \dots forme une chaîne de Markov (avec $E = \mathbb{Z}$), avec pour tout $i \in \mathbb{Z}$,

$$P_{ij} = \begin{cases} 1/2, & \text{si } |j - i| = 1, \\ 0, & \text{sinon.} \end{cases} \quad \square$$

Exemple 2.3 (marche aléatoire avec réflexion au bord). On appelle X_0, X_1, \dots une marche aléatoire simple et symétrique sur $\{0, 1, 2, \dots, N\}$, avec réflexion au bord, si la marche aléatoire visite 0 (resp. N) à un certain moment, elle est forcée de visiter 1 (resp. $N - 1$) à l'étape suivante. On est exactement dans la situation précédente, sauf pour $i = 0$ et $i = N$: $P_{01} = 1$, $P_{0j} = 0$ si $j \neq 1$; $P_{N,N-1} = 1$, $P_{Nj} = 0$ si $j \neq N - 1$. Il s'agit d'une chaîne de Markov avec $E = \{0, 1, 2, \dots, N\}$. La matrice de transition est

$$P = \begin{pmatrix} 0 & 1 & 0 & 0 & \dots & 0 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 & \dots & 0 & 0 & 0 \\ 0 & 1/2 & 0 & 1/2 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 1/2 & 0 & 1/2 \\ 0 & 0 & 0 & 0 & \dots & 0 & 1 & 0 \end{pmatrix}. \quad \square$$

Exemple 2.4 (marche aléatoire avec absorption au bord). On appelle X_0, X_1, \dots une marche aléatoire simple et symétrique sur $\{0, 1, 2, \dots, N\}$, avec absorption au bord, au sens que si la marche aléatoire visite 0 ou N , elle y restera pour le reste du temps. On est encore dans la situation précédente, sauf pour $i = 0$ et $i = N$: $P_{00} = 1$, $P_{0j} = 0$ si $j \neq 0$;

$P_{NN} = 1$, $P_{Nj} = 0$ si $j \neq N$. Il s'agit d'une chaîne de Markov avec $E = \{0, 1, 2, \dots, N\}$. La matrice de transition est

$$P = \begin{pmatrix} 1 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 1/2 & 0 & 1/2 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1/2 & 0 & 1/2 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 1 \end{pmatrix}. \quad \square$$

Exemple 2.5 (chaîne à deux états). Considérons l'état d'une ligne téléphonique $X_n = 0$ si la ligne est libre à l'instant n , et $X_n = 1$ si la ligne est occupée. Supposons que sur chaque intervalle de temps, il y a une probabilité p qu'un appel arrive (un appel au plus). Si la ligne est déjà occupée, l'appel est perdu. Supposons également que si la ligne est occupée au temps n , il y a une probabilité q qu'elle se libère au temps $n + 1$. On peut modéliser ainsi une chaîne de Markov à valeurs dans $E = \{0, 1\}$, avec matrice de transition

$$P = \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix}. \quad \square$$

Exemple 2.6 (file d'attente simple). On modifie l'exemple précédent en supposant qu'on peut mettre un appel en attente. Les appels arrivent et la ligne se libère comme avant. Si un appel arrive pendant que la ligne est occupée et si le système n'est pas saturé, l'appel est mis en attente. Si un appel arrive alors qu'il y a déjà un en attente, il est perdu. Cette fois l'espace d'état est $E = \{0, 1, 2\}$. On a

$$P_{00} = 1-p, \quad P_{01} = p, \quad P_{02} = 0.$$

De même

$$P_{20} = 0, \quad P_{21} = q, \quad P_{22} = 1-q.$$

Le cas où il y a exactement un appel retenu au temps n est un peu plus délicat. On a $P_{10} = q(1-p)$ (l'appel se termine et pas d'appel nouveau arrive) et $P_{12} = p(1-q)$ (un appel nouveau arrive et celui en cours continue). Comme la somme $P_{10} + P_{11} + P_{12}$ doit valoir 1, on a donc $P_{11} = 1 - q(1-p) - p(1-q)$. Finalement, la matrice de transition de la chaîne est

$$P = \begin{pmatrix} 1-p & p & 0 \\ q(1-p) & 1 - q(1-p) - p(1-q) & p(1-q) \\ 0 & q & 1-q \end{pmatrix}. \quad \square$$

On décrit maintenant l'état d'une chaîne de Markov.

Propriété 2.7. Soit X_0, X_1, \dots une chaîne de Markov à valeurs dans E , avec pour probabilités de transition $(P_{ij})_{(i,j) \in E^2}$. Alors, pour $(x_0, \dots, x_n) \in E^{n+1}$, on a

$$\mathbb{P}(X_1 = x_1, \dots, X_n = x_n \mid X_0 = x_0) = P_{x_0, x_1} \cdots P_{x_{n-1}, x_n}.$$

En particulier, si l'espace d'état E est fini, disons $E = \{1, \dots, N\}$,

$$\mathbb{P}(X_n = j \mid X_0 = i) = P_{ij}^n,$$

où $P^n = P \times \cdots \times P$ au sens des produits de matrices et $P^n = (P_{ij}^n)_{i,j \in E}$.

Preuve. La première formule s'obtient par récurrence. Plus précisément, on a

$$\begin{aligned} & \mathbb{P}(X_0 = x_0, \dots, X_n = x_n \mid X_0 = x_0) \\ &= \mathbb{P}(X_1 = x_1, \dots, X_{n-1} = x_{n-1} \mid X_0 = x_0) \mathbb{P}(X_n = x_n \mid X_0 = x_0, \dots, X_{n-1} = x_{n-1}) \\ &= \mathbb{P}(X_1 = x_1, \dots, X_{n-1} = x_{n-1} \mid X_0 = x_0) P_{x_{n-1}, x_n} \\ &= P_{x_0, x_1} \cdots P_{x_{n-1}, x_n}. \end{aligned}$$

La seconde formule en découle de la première si $E = \{1, \dots, N\}$:

$$\begin{aligned} \mathbb{P}(X_n = j \mid X_0 = i) &= \sum_{x_1=1}^N \cdots \sum_{x_{n-1}=1}^N \mathbb{P}(X_1 = x_1, \dots, X_{n-1} = x_{n-1}, X_n = j \mid X_0 = i) \\ &= \sum_{x_1=1}^N \cdots \sum_{x_{n-1}=1}^N P_{i, x_1} \cdots P_{x_{n-1}, j} \\ &= P_{ij}^n. \end{aligned} \quad \square$$

Remarque. Une application importante de la Propriété 2.7 est la suivante : fixons des entiers n et k , et soient $x_0, \dots, x_n, \dots, x_{n+k}$ des points de E . On a alors

$$\begin{aligned} & \mathbb{P}(X_{n+1} = x_{n+1}, \dots, X_{n+k} = x_{n+k} \mid X_0 = x_0, \dots, X_n = x_n) \\ &= \frac{\mathbb{P}(X_0 = x_0, \dots, X_n = x_n, X_{n+1} = x_{n+1}, \dots, X_{n+k} = x_{n+k})}{\mathbb{P}(X_0 = x_0, \dots, X_n = x_n)} \\ &= \frac{P_{x_0, x_1} \cdots P_{x_{n-1}, x_n} P_{x_n, x_{n+1}} \cdots P_{x_{n+k-1}, x_{n+k}}}{P_{x_0, x_1} \cdots P_{x_{n-1}, x_n}} \\ &= P_{x_n, x_{n+1}} \cdots P_{x_{n+k-1}, x_{n+k}} \\ &= \mathbb{P}(X_1 = x_{n+1}, \dots, X_k = x_{n+k} \mid X_0 = x_n). \end{aligned}$$

En mots, si on sait qu'à l'instant n la chaîne est en x_n , alors la chaîne $\tilde{X}_0 = X_n, \dots$, $\tilde{X}_k = X_{n+k}, \dots$, obtenue par translation temporelle, ne dépend pas de la trajectoire suivie pour arriver en x_n au temps n , et a même loi que la chaîne initiale quand cette dernière est issue de x_n . C'est une propriété conforme à l'intuition que l'on peut avoir. \square

3. Propriétés des chaînes de Markov sur un espace fini

Dans cette section on suppose que X_0, X_1, \dots est une chaîne de Markov à valeurs dans l'espace d'état E qui est fini, $E = \{1, \dots, N\}$, et on note P la matrice de transition.

On s'intéresse à la distribution de X_n quand $n \rightarrow \infty$. D'après la Propriété 2.7, ceci revient à étudier la suite de matrices P^n quand $n \rightarrow \infty$.

Considérons d'abord l'exemple simple de la chaîne à deux états

$$P = \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix},$$

avec $0 < p, q < 1$. Pour calculer P^n , on peut diagonaliser P . Les valeurs propres sont 1 et $1-p-q$, et on diagonalise $D = Q^{-1}PQ$ avec

$$Q = \begin{pmatrix} 1 & -p \\ 1 & q \end{pmatrix}, \quad Q^{-1} = \begin{pmatrix} \frac{q}{p+q} & \frac{p}{p+q} \\ -\frac{1}{p+q} & \frac{1}{p+q} \end{pmatrix}, \quad D = \begin{pmatrix} 1 & 0 \\ 0 & 1-p-q \end{pmatrix}.$$

La diagonale de D est constituée des valeurs propres. Les colonnes de Q sont les vecteurs propres à droite de P , et les lignes de Q^{-1} les vecteurs propres à gauche. Les vecteurs propres sont uniques à une constante multiplicative près; on a choisi la constante pour la valeur propre 1 de sorte que la première ligne de Q^{-1} soit un vecteur probabilité (on verra plus tard pourquoi). On a alors

$$\begin{aligned} P^n &= (QDQ^{-1})^n \\ &= QD^nQ^{-1} \\ &= \begin{pmatrix} \frac{q+p(1-p-q)^n}{p+q} & \frac{p-p(1-p-q)^n}{p+q} \\ \frac{q-q(1-p-q)^n}{p+q} & \frac{p+q(1-p-q)^n}{p+q} \end{pmatrix}. \end{aligned}$$

Comme $|1-p-q| < 1$, on voit que

$$\lim_{n \rightarrow \infty} P^n = \begin{pmatrix} \frac{q}{p+q} & \frac{p}{p+q} \\ \frac{q}{p+q} & \frac{p}{p+q} \end{pmatrix}.$$

Le fait que la seconde valeur propre, $1 - p - q$ soit en module strictement inférieure à 1 est primordial dans le calcul; la matrice limite est constituée de deux vecteurs lignes identiques, qui est le vecteur propre à gauche de P , normalisé de sorte à être une probabilité.

Supposons maintenant que l'état initial de la chaîne est donné par une probabilité $\mathbb{P}(X_0 = 0) = a$, $\mathbb{P}(X_0 = 1) = 1 - a$. Nous avons vu que l'état de la chaîne au temps n suit la loi caractérisée par

$$\mathbb{P}(X_n = j \mid X_0 = i) = P_{ij}^n \quad i, j = 0 \text{ ou } 1,$$

et on déduit que quand $n \rightarrow \infty$, quelque soit l'état initial i de la chaîne,

$$\lim_{n \rightarrow \infty} \mathbb{P}(X_n = 0 \mid X_0 = i) = \frac{q}{p+q}, \quad \lim_{n \rightarrow \infty} \mathbb{P}(X_n = 1 \mid X_0 = i) = \frac{p}{p+q}.$$

On notera que la limite en loi qu'on a obtenue est donnée par le vecteur propre associé à la valeur propre 1, normalisé pour être une probabilité.

Passons au cas général.

Convention. Un vecteur $v = (v_1, \dots, v_N)$ est dit “vecteur probabilité” sur $E = \{1, \dots, N\}$ si les coefficients de v sont positifs et de somme 1. On dit qu'une v.a. Y à valeurs dans E suit la loi v si $\mathbb{P}(Y = i) = v_i$ pour tout $i \in E$.

Propriété 3.1. Si l'état initial de la chaîne, X_0 , suit la loi v , alors X_n suit la loi vP^n .

Preuve. D'après la Propriété 2.7, pour tout $j \in E$,

$$\mathbb{P}(X_n = j) = \sum_{i \in E} \mathbb{P}(X_0 = i) \mathbb{P}(X_n = j \mid X_0 = i) = \sum_{i \in E} v_i P_{ij}^n.$$

Donc X_n suit la loi vP^n . □

Définition 3.2. Un vecteur probabilité $v = (v_1, \dots, v_N)$ sur E est dit **probabilité invariante** (ou probabilité d'équilibre), pour la chaîne de matrice de transition P si v est un vecteur propre à gauche de valeur propre 1, i.e. $v = vP$.

La propriété suivante justifie la terminologie :

Propriété 3.3. Si v est une probabilité invariante et si l'état initial de la chaîne X_0 suit la loi v , alors X_n suit également la loi v pour tout n .

Preuve. Si v une probabilité invariante, alors $vP^n = v$. D'après la Propriété 3.1, X_n suit la loi v . \square

Il existe toujours une probabilité invariante lorsque la chaîne de Markov est à valeurs dans un espace d'état fini. Les questions naturelles qui se posent alors sont : (i) s'il y a unicité pour probabilité invariante; (ii) en admettant que la réponse est oui, si quelle que soit la distribution initiale de la chaîne, l'état converge quand le temps tend vers l'infini, vers cette unique probabilité invariante.

On peut observer que si, pour une distribution initiale donnée $m = (m_1, \dots, m_N)$, X_n converge en loi vers un vecteur probabilité v , autrement dit $\lim_{n \rightarrow \infty} mP^n = v$, alors $vP = v$, et v est nécessairement une probabilité invariante.

On a le résultat suivant dont on ne présente pas la preuve.

Théorème 3.4. Soit X_0, \dots, X_n, \dots une chaîne de Markov sur un espace fini, avec pour matrice de transition P . On suppose que les coefficients de P sont tous strictement positifs. Alors il existe une unique probabilité invariante, v , et quelque que soit la distribution initiale de la chaîne, X_n converge en loi vers v quand n tend vers ∞ .

Remarque. On se gardera de croire que le résultat précédent est valable en toute généralité. Il est facile de construire des chaînes de Markov possédant plusieurs probabilités invariantes. \square

Si tous les coefficients de P ne sont pas strictement positifs, on ne peut pas appliquer le Théorème 3.4. Il est possible d'appliquer un autre résultat.

Définition 3.5. On dit qu'une chaîne de Markov de probabilités de transition $(P_{ij})_{(i,j) \in E^2}$ est irréductible si, pour chaque couple d'états (i, j) , il existe des entiers m et n tels que $P_{ij}^m > 0$ et $P_{ji}^n > 0$.

On admet le résultat suivant qui généralise le Théorème 3.4.

Théorème 3.6. Une chaîne de Markov irréductible sur un espace fini a une unique probabilité invariante. C'est le vecteur propre à gauche associé à la valeur propre 1, qui est simple.

Remarque. Il est clair que, si E est fini, et si tous les coefficients de P^k sont strictement positifs pour un certain entier k , alors la chaîne est irréductible. Donc d'après le Théorème 3.6, il existe une unique probabilité invariante. \square