Fondements des probabilités
Licence 3 de Mathématiques, parcours Agregation-Recherche.

Ces notes de cours ont été construites à partir des notes de cours de Jean-Christophe Breton et celles de Philippe Briand.

Hélène Guérin
IRMAR, Université Rennes 1
Campus de Beaulieu, 35042 Rennes Cedex FRANCE.
helene.guerin@univ-rennes1.fr

Mai 2017.

Table des matières

Chapitre 1

Espaces de probabilité, Vocabulaire probabiliste

1 Espaces de probabilité

On ne peut pas toujours prédire avec certitude le résultat d'une expérience, on parle alors d'expérience aléatoire. Par exemple, lancer un dé ou une pièce de monnaie sont des expériences aléatoires. D'autres phénomènes ne sont peut-être pas aléatoires (comme la succession des nucléotides sur un brin d'ADN), mais n'en connaissant pas tous les mécanismes, il peut être utile pour pouvoir les étudier de les supposer aléatoires.
L'espace fondamental est un triplet généralement noté $(\Omega, \mathcal{F}, \mathbb{P})$. L'ensemble Ω est un ensemble non vide décrivant toutes des réalisations possibles d'une expérience. Un point $\omega \in \Omega$ est une réalisation particulière de cette expérience.

Exemple.
1. On lance un dé, on s'intéresse au chiffre obtenu. L'ensemble fondamentale est alors fini, plus précisément
$$\Omega = \{1, 2, 3, 4, 5, 6\}.$$
2. On jette une pièce autant de fois que nécessaire pour obtenir une fois "face". L'ensemble fondamentale est alors infini dénombrable. Si on note F pour "face" et P pour "pile", on a
$$\Omega = \{F, PF, PPF, PPPF, \ldots, PPP \cdots PPF, \ldots\}.$$
3. On s'intéresse à la durée de vie d'une ampoule. L'ensemble fondamental est alors infini, non dénombrable : $\Omega = [0, +\infty[$.

Au lieu de s'intéresser à toute l'expérience, on peut ne s'intéresser qu'à certains résultats. Un **événement** A est une combinaison de résultats possibles. C'est un sous-ensemble de l'ensemble fondamental, $A \subset \Omega$ et on dit que l'événement A se réalise si $\omega \in A$.

Exemple. On lance deux dés. L'ensemble fondamental est

$$\Omega = \{(1, 1), (1, 2), (1, 3) \ldots, (2, 1), (2, 2), \ldots, (6, 5), (6, 6)\}.$$

On peut considérer les événements suivants :
- la somme des points est 6 : $A = \{(1, 5), (5, 1), (2, 4), (4, 2), (3, 3)\}$,
- la somme des points est un multiple de 3 :
$B = \{(1, 2), (2, 1), (1, 5), (5, 1), (2, 4), (4, 2), (3, 3), (3, 6), (6, 3), (4, 5), (5, 4), (6, 6)\}$.

On aimerait associer à un événement A un nombre $\mathbb{P}(A)$ qui rende compte des "chances" de réalisation de A. Pour cela, on utilise la théorie de la mesure et on doit se limiter à un sous-ensemble $\mathcal{F}$ de parties de Ω appelé tribu des événements et on considérer une mesure de probabilité $\mathbb{P}$ sur $(\Omega, \mathcal{F})$.

Formalisons ces notions en utilisant ce que vous avez appris pendant votre cours de théorie de la mesure. Les probabilités reposent sur la théorie de la mesure mais son vocabulaire est différent.

Considérons Ω un ensemble et $\mathcal{P}(\Omega) = \{A : A \subset \Omega\}$ l'ensemble de ses parties.

Definition 1. Tribu
Une tribu $\mathcal{F}$ *sur* Ω *est une classe de partie de* Ω*, i.e.* $\mathcal{F} \subset \mathcal{P}(\Omega)$*, telle que*

1. $\Omega \in \mathcal{F}$,
2. *si* $A \in \mathcal{F}$*, alors* $A^c \in \mathcal{F}$*,* (stable par passage au complémentaire)
3. *si pout tout* $n \in \mathbb{N}$*,* $A_n \in \mathcal{F}$*, alors* $\cup_{n\in\mathbb{N}} A_n \in \mathcal{F}$*.* (stable par union dénombrable)

Dans ce cas, $(\Omega, \mathcal{F})$ s'appelle un espace mesurable et les éléments de $\mathcal{F}$ sont appelés événements.

Exemple. $\mathcal{F}_1 = \{\emptyset, \Omega\}$ est une tribu sur Ω, appelée tribu grossière, $\mathcal{F}_2 = \mathcal{P}(\Omega)$, $\mathcal{F}_3 = \{\emptyset, \Omega, A, A^c\}$, où $A \subset \Omega$, sont aussi des tribus.

On remarque qu'une tribu contient $\emptyset$, est stable par union finie, par intersection finie et dénombrable, par différence,...
Par ailleurs, une intersection quelconque de tribus est encore une tribu. Soit $\mathcal{A} \subset \mathcal{P}(\Omega)$, l'intersection de toutes les tribus contenant $\mathcal{A}$ est une tribu qui par construction est la plus petite tribu contenant $\mathcal{A}$. Cette tribu est appelée **tribu engendrée** par $\mathcal{A}$ et on la note généralemant $\sigma(\mathcal{A})$.
Lorsque Ω est un ensemble dénombrable on utilise souvent la tribu $\mathcal{P}(\Omega)$ et lorsque Ω est un espace topologique, on utilise souvent la tribu borélienne $\mathcal{B}(\Omega)$, qui correspond à la tribu engendrée par la classe des ouverts.

Exemple. La tribu $\mathcal{B}(\mathbb{R})$ est engendrée par les intervalles ouverts, et aussi par les intervalles du type $]s, t[$ et encore par les intervalles du type $]-\infty, t]$. La tribu $\mathcal{B}(\mathbb{R}^d)$ est engendrée par les pavés du type $\Pi_{i=1}^d]s_i, t_i[$.

Venons-en à présent à la notion de mesure de probabilité $\mathbb{P}$.

Definition 2. *Une* mesure probabilité*, ou plus simplement une probabilité, sur un espace mesurable* $(\Omega, \mathcal{F})$ *est une application* $\mathbb{P} : \mathcal{F} \to [0, 1]$ *telle que*

1. $\mathbb{P}(\emptyset) = 0$,
2. *si* $A_n \in \mathcal{F}$ *pour tout* $n \in \mathbb{N}$ *avec* $A_n \cap A_m = \emptyset$ *pour* $n \neq m$*, alors*
$$\mathbb{P}\left(\bigcup_{n\in\mathbb{N}} A_n\right) = \sum_{n\in\mathbb{N}} \mathbb{P}(A_n) \quad (\sigma\text{-}\textit{additivité}),$$
3. $\mathbb{P}(\Omega) = 1$

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

Exemple. 1. $([0, 1], \mathcal{B}([0, 1]), \lambda)$ où λ est la mesure de Lebesgue sur $[0, 1]$.

2. Mesure de Dirac sur $(\Omega, \mathcal{P}(\Omega))$: Soit $a \in \Omega$, pour tout $A \in \mathcal{F}$, on définit la mesure de Dirac au point a par
$$\delta_a(A) = \begin{cases} 1 \text{ si } a \in A \\ 0 \text{ si } a \notin A \end{cases}$$
3. Pour définir une probabilité sur $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$, il suffit de se donner une suite de réels positifs p_n tel que $\sum_{n\in\mathbb{N}} p_n = 1$. Alors $\mathbb{P} = \sum_{n\in\mathbb{N}} p_n\delta_n$, où δ_n est la masse de Dirac en n, est une mesure de probabilité sur $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$. On a alors $\mathbb{P}(A) = \sum_{n\in\mathbb{N}} p_n \mathbb{1}_A(n)$.

2 Propriétés élémentaires des mesures de probabilités

En terme probabiliste, une union $\bigcup$ s'interprète comme un "ou" et une intersection $\bigcap$ s'interprète comme un "et".

Propriété 3. *Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité et A, B et $(A_n)_{n\in\mathbb{N}}$ des événements, alors*

1. *Si $A \cap B = \emptyset$, alors $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$* (additivité)
2. *Si $B \subset A$, alors $\mathbb{P}(B) \leqslant \mathbb{P}(A)$* (croissance) *et $\mathbb{P}(A \setminus B) = \mathbb{P}(A) - \mathbb{P}(B)$.*
3. *$\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$*
4. *$\mathbb{P}(\cup_{n\in\mathbb{N}} A_n) \leqslant \sum_{n\in\mathbb{N}} \mathbb{P}(A_n)$*
5. *si $(A_n)_{n\in\mathbb{N}}$ suite croissante (i.e., $A_n \subset A_{n+1}$), alors*
$$\mathbb{P}\left(\bigcup_{n\in\mathbb{N}} A_n\right) = \lim_{n\to\infty} \mathbb{P}(A_n) = \sup_{n\in\mathbb{N}} \mathbb{P}(A_n)$$
6. *si $(A_n)_{n\in\mathbb{N}}$ suite décroissante (i.e., $A_{n+1} \subset A_n$), alors*
$$\mathbb{P}\left(\bigcap_{n\in\mathbb{N}} A_n\right) = \lim_{n\to\infty} \mathbb{P}(A_n) = \inf_{n\in\mathbb{N}} \mathbb{P}(A_n)$$

Remarque. *En général on ne peut pas calculer $\mathbb{P}(A \cup B)$ à partir de $\mathbb{P}(A)$ et $\mathbb{P}(B)$. En fait, on a*
$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$$
Cette formule se généralise pour plus de deux événements en utilisant une preuve par récurrence.

Propriété 4. *Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité et $A_1, A_2, \ldots, A_n$ des événements de $\mathcal{F}$, alors*
$$\mathbb{P}\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n \mathbb{P}(A_i) = \sum_{k=2}^n (-1)^{k+1} \sum_{1\leqslant i_1<i_2<\ldots<i_k\leqslant n} \mathbb{P}(A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_k}).$$
Par conséquent,
$$\mathbb{P}(A \cup B \cup C) = \mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(C) - \mathbb{P}(A \cap B) - \mathbb{P}(A \cap C) - \mathbb{P}(B \cap C) + \mathbb{P}(A \cap B \cap C).$$

Démonstration. (Propriété 3)

1. Suffit de prendre $A_1 = A, A_2 = B, A_n = \emptyset$ pour $n \geqslant 3$.
2. $A = B \cup (A \cap B^c)$ union disjointe et $A \cap B^c = A \setminus B$.

3. On le démontre par récurrence. On a $A_1 \cup A_2 = A_1 \cup (A_2 \cap A_1^c)$ et $A_2 \cap A_1^c \subset A_2$.
4. Il s'agit de convergence monotone... mais on peut aussi vouloir le redémontrer ! Comme union croissante, $A_n = \cup_{k=1}^n A_k$. On pose $B_n = A_n \setminus A_{n_1}$ (avec $A_0 = \emptyset$) événements disjoints et $\cup_{k=1}^n B_k = A_n$. Par conséquent,

$$\lim_{n\to\infty} \mathbb{P}(A_n) = \lim_{n\to\infty} \mathbb{P}\left(\bigcup_{k=1}^n B_k\right) \underset{\text{additivité}}{=} \lim_{n\to\infty} \sum_{k=1}^n \mathbb{P}(B_k) = \sum_{k=1}^\infty \mathbb{P}(B_k) \underset{\sigma-\text{additivité}}{=} \mathbb{P}\left(\bigcup_{k=1}^\infty B_k\right) = \mathbb{P}\left(\bigcup_{k=1}^\infty A_k\right).$$

5. Par passage au complémentaire.

□

3 Liminf et Limsup d'ensembles

On considère $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité et $(A_n)_{n\in\mathbb{N}}$ une suite d'événements. Deux événements complexes mais très utiles sont la limite inférieure et la limite supérieure.

Definition 5. *On définit*

$$\liminf_{n\to\infty} A_n = \bigcup_{n\geqslant 1} \bigcap_{k>n} A_k \quad \textit{et} \quad \limsup_{n\to\infty} A_n = \bigcap_{n\geqslant 1} \bigcup_{k>n} A_k.$$

On peut mettre en parallèle les notions équivalentes pour les suites : Soit $(u_n)_{n\in\mathbb{N}}$ une suite réelle, alors $\liminf_{n\to\infty} u_n = \sup_{n\geqslant 0} \inf_{k\geqslant n} u_k$ et $\limsup_{n\to\infty} u_n = \inf_{n\geqslant 0} \sup_{k\geqslant n} u_k$ qui correspondent respectivement aux plus petite et plus grande valeur d'adhérence de la suite.
On remarque que $\liminf_{n\to\infty} A_n \subset \limsup_{n\to\infty} A_n$ et $(\limsup_{n\to\infty} A_n)^c = \liminf_{n\to\infty} A_n^c$.

Démonstration. soit n fixé de façon arbitraire. On a $\bigcap_{k>n} A_k \subset A_p$ pour tout $p > n$, donc $\bigcap_{k>n} A_k \subset \bigcup_{p>m} A_p$ pour tout $m \in \mathbb{N}$ et donc $\bigcap_{k>n} A_k \subset \bigcap_{m\geqslant 1} \bigcup_{p>m} A_p = \limsup_{n\to\infty} A_n$ et on conclut facilement. □

Remarque. *Soit $(A_n)_{n\in\mathbb{N}}$ une suite infinie d'événements.*

1. *Si $\omega \in \liminf_{n\to\infty} A_n$, alors $\exists n \geqslant 0$, $\forall k \geqslant n$, $\omega \in A_k$. Par conséquent, $\omega \in \liminf_{n\to\infty} A_n$ signifie* à partir d'un certain rang, tous les A_n sont réalisés *(ou* tous les A_n sont réalisés sauf un nombre fini*)* .
2. *Si $\omega \in \limsup_{n\to\infty} A_n$, alors $\forall n \geqslant 0$, $\exists k \geqslant n$, $\omega \in A_k$. Par conséquent, $\omega \in \limsup_{n\to\infty} A_n$ signifie* une infinité de A_n sont réalisés.

Propriété 6. *Soit $(A_n)_{n\in\mathbb{N}}$ une suite d'événements. On a*

$$\liminf_{n\to\infty} \mathbb{1}_{A_n} = \mathbb{1}_{\liminf_{n\to\infty} A_n} \quad \textit{et} \quad \limsup_{n\to\infty} \mathbb{1}_{A_n} = \mathbb{1}_{\limsup_{n\to\infty} A_n}$$

Démonstration. Il suffit de le faire pour lim sup. On a $\omega \in \limsup_{n\to\infty} A_n$ si $\forall n$, $\exists k > n$, tel que $\omega \in A_k$, ie $\forall n$, $\exists k > n$ $\mathbb{1}_{A_k}(\omega) = 1$, soit $\inf_{n\geqslant 0} \sup_{k>n} \mathbb{1}_{A_k}(\omega) = 1$, soit $\limsup_{n\to\infty} \mathbb{1}_{A_n}(\omega) = 1$. L'équivalence est claire. □

Propriété 7. *Soit $(A_n)_{n\in\mathbb{N}}$ une suite d'événements. On a*

$$\mathbb{P}\left(\liminf_{n\to\infty} A_n\right) \leqslant \liminf_{n\to\infty} \mathbb{P}(A_n) \leqslant \limsup_{n\to\infty} \mathbb{P}(A_n) \leqslant \mathbb{P}\left(\limsup_{n\to\infty} A_n\right).$$

Démonstration. Soit $B_n = \bigcup_{k>n-1} A_k$ et $C_n = \bigcap_{k>n-1} A_k$. Alors B_n suite décroissante d'ensembles de limite $\limsup A_n$ et C_n suite croissante d'ensemble de limite $\liminf A_n$. Par convergence monotone, on a

$$\mathbb{P}(A_n) \leqslant \mathbb{P}(B_n) \to \mathbb{P}(\limsup A_n) \quad \text{et} \quad \mathbb{P}(A_n) \geqslant \mathbb{P}(C_n) \to \mathbb{P}(\liminf A_n)$$

Il suffit de prendre les limites sup et limites inf de $\mathbb{P}(A_n)$. □

4 Classe monotone

On rappelle dans cette section un procédé d'extension des définitions de certains objets sur les tribus après les avoir définis sur une classe restreinte d'ensemble.

Definition 8. *Une famille $\mathcal{M}$ de parties de Ω est appelée* **classe monotone** *(ou $\lambda-$système) si*

1. $\Omega \in \mathcal{M}$,
2. *$\mathcal{M}$ est stable par différence : lorsque $A, B \in \mathcal{M}$ et $B \subset A$ alors $A \backslash B \in \mathcal{M}$,*
3. *$\mathcal{M}$ est stable par union croissante (si $A_n \in \mathcal{M}$ et $A_n \subset A_{n+1}$ alors $\cup_{n\geqslant 0} A_n \in \mathcal{M}$).*

Remarque. — *Une intersection quelconque de classes monotones est encore une classe monotone.*
- — *Une tribu est une classe monotone, car $A \backslash B = A \cap B^c$.*
- — *Une classe monotone stable par intersection finie est une tribu.*

Definition 9. *Pour $\mathcal{E} \subset \mathcal{P}(\Omega)$, on appelle classe monotone engendrée par $\mathcal{E}$ la plus petite classe monotone contenant $\mathcal{E}$, i.e. l'intersection de toutes les classes monotones contenant $\mathcal{E}$. On la note $\mathcal{M}(\mathcal{E})$.*

Theorem 10. Théorème des classes monotones, Théorème de Dynkin
Soit $\mathcal{A}$ une famille de parties de Ω stable par intersection finie (appelé π-système). Alors $\mathcal{M}(\mathcal{A}) = \sigma(\mathcal{A})$.

(Voir votre cours d'intégration pour la preuve.)
Une application bien connue du théorème de Dynkin.

Theorem 11. *Si deux probabilités $\mathbb{P}_1$ et $\mathbb{P}_2$ coïncident sur $\mathcal{A}$ stable par intersections finies, alors elles coïncident sur $\sigma(\mathcal{A})$.*

Démonstration. Soit $\mathcal{M} = \{A \in \mathcal{F} : \mathbb{P}_1(A) = \mathbb{P}_2(A)\}$. Alors $\Omega \in \mathcal{M}$, si $A, B \in \mathcal{M}$ avec $B \subset A$ alors $A \setminus B \in \mathcal{M}$ (car $\mathbb{P}_1(A \setminus B) = \mathbb{P}_1(A) - \mathbb{P}_1(B) = \mathbb{P}_2(A \setminus B)$) et si $(A_n)_{n\geqslant 0}$ suite croissante de $\mathcal{M}$, $\bigcup A_n \in \mathcal{M}$ (car $\mathbb{P}_1(\bigcup A_n) = \lim_{n\to\infty} \mathbb{P}_1(A_n) = \mathbb{P}_2(\bigcup A_n)$). Par conséquent, $\mathcal{M}$ est une classe monotone, avec $\mathcal{A} \subset \mathcal{M}$ et donc $\mathcal{M}(\mathcal{A}) \subset \mathcal{M}$. Comme $\mathcal{A}$ est un π-système, par le théorème de la classe monotone $\mathcal{M}(\mathcal{A}) = \sigma(\mathcal{A})$, cqfd. □

Chapitre 2

Variables aléatoires

Les variables aléatoires sont l'objet de base de ce cours. Il s'agit de fonction du hasard dont nous présenteront dans ce chapitre les outils clefs pour leur étude.
Imaginons une personne jouant une partie de dés en 10 coups : à chaque lancer, il gagne 2 euros s'il obtient un 5 ou un 6, un euro pour un quatre, rien pour un 3, perd un euros pour un 2 et trois euros s'il obtient 1.
Si on cherche à modéliser l'expérience, on peut prendre $\Omega = \{1, \dots, 6\}^{10}$, $\mathcal{F} = \mathcal{P}(\Omega)$ et si le dé est bien équilibré chaque configuration a la même probabilité d'apparaître, il est donc naturel de choisir $\mathbb{P}(\{\omega\}) = 6^{-10}$ pour tout $\omega \in \Omega$. Pour le joueur, ce qui est intéressent, ce n'est pas la description de l'expérience, mais juste le gain final qu'il va avoir. Bien entendu le gain dépend du résultat de l'expérience. Le gain est une application sur Ω à valeurs dans $\{-30, \dots, 20\}$. On parle de variable aléatoire

1 Définition et propriétés

Soit $(\Omega, \mathcal{F})$ et $(E, \mathcal{E})$ deux espaces mesurables et X une application de Ω dans E. Pour $B \in \mathcal{E}$, on note $\{X \in B\}$ l'image réciproque de B par X :

$$\{X \in B\} := X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\}$$

Definition 12. *Soit* $(\Omega, \mathcal{F}, \mathbb{P})$ *un espace de probabilité et* $(E, \mathcal{E})$ *un espace mesurable.*
On appelle variable aléatoire *(v.a.) toute application* X *mesurable d'un espace de probabilité* $(\Omega, \mathcal{F}, \mathbb{P})$ *à valeurs dans un espace mesurable* $(E, \mathcal{E})$*, i.e.* $X^{-1}(\mathcal{E}) \subset \mathcal{F}$*.*

Souvent $(E, \mathcal{E}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, on parle alors de v.a. réelle. Si $(E, \mathcal{E}) = (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, on parle de vecteur aléatoire. On peut alors écrire $X = (X_1, \dots, X_d)$ avec $X_i : (\Omega, \mathcal{F}, \mathbb{P}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ variable aléatoire appelée $i^{\text{ème}}$ marginale de X. Si $(E, \mathcal{E}) = (\mathbb{R}^+, \mathcal{B}(\mathbb{R}^+))$ alors X est dite v.a. positive et si $(E, \mathcal{E}) = (\mathbb{N}, \mathcal{P}(\mathbb{N}))$, X est dite v.a. entière.

Exemple. Une fonction étagée est une fonction qui ne prend qu'un nombre fini de valeurs. Elle s'écrit sous la forme $X = \sum_{i=1}^{n} x_i \mathbb{1}_{A_i}$ avec $A_i \in \mathcal{F}$ et $x_1, \dots, x_n \in \mathbb{R}$ (il n'y a pas unicité de la représentation).

Pour rappel, toute v.a. est limite simple de fonction étagées. Et si de plus la v.a. est positive alors la limite peut-être choisie croissante. (Voir votre cours d'intégration ou le TD 1).

Propriété 13. *Si* $\mathcal{E} = \sigma(\mathcal{C})$*,* X *est une v.a. si et seulement si* $X^{-1}(\mathcal{C}) \subset \mathcal{F}$*.*

Par conséquent si $(E,\mathcal{E}) = (\mathbb{R},\mathcal{B}(\mathbb{R}))$, X est une v.a. si et seulement si $\{X \leqslant t\} \in \mathcal{F}$ pour tout $t \in \mathbb{R}$.
On rappelle par ailleurs que la mesurabilité est stable par composition. La somme, différence, quotient de v.a. réelles est encore une v.a. Le max et le min de v.a. réelles est encore une v.a., la partie imaginaire et la partie réelle d'un v.a complexe est une v.a. Si (X_n) est une suite de v.a., alors l'inf, le sup, les $\liminf$ et $\limsup$ sont des v.a.

2 Loi d'une variable aléatoire

Definition 14. *Soit $X : (\Omega,\mathcal{F},\mathbb{P}) \to (E,\mathcal{E})$ une v.a.. On appelle la loi de X la mesure image $\mathbb{P}_X$ de $\mathbb{P}$ sur $(E,\mathcal{E})$ par X : pour $A \in \mathcal{E}$*

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

Dans le cas d'un vecteur aléatoire, $X = (X_1,\ldots,X_n)$, la loi $\mathbb{P}_{(X_1,\ldots,X_n)}$ est appelée la loi jointe des variables aléatoires $X_1,\ldots,X_n$. Les lois des coordonnées de la variable s'appellent les lois marginales.

Propriété 15. *La loi image $\mathbb{P}_X$ est une mesure de probabilité sur $(E,\mathcal{E})$.*

Démonstration. Vous avez déjà vu en cours d'intégration qu'une mesure image est une mesure. Il suffit donc de vérifier que $\mathbb{P}_X$ est une probabilité, i.e. de masse 1 : $\mathbb{P}_X(E) = \mathbb{P}(\Omega) = 1$ car $\mathbb{P}$ est une probabilité sur Ω. $\square$

Par abus de notation, si $A = \{x\}$, on notera $\mathbb{P}(X = x)$ au lieu de $\mathbb{P}(X \in \{x\})$.

Exemple. On lance deux dés et on regarde X le minimum des deux chiffres obtenu. On a $\Omega = \{1,2,3,4,5,6\}^2$ et $X : \Omega \to \{1,2,3,4,5,6\}$, $(\omega_1,\omega_2) \mapsto \min(\omega_1,\omega_2)$ est bien une v.a. (car projection mesurable et min mesurable). Les dés n'étant pas truqués il est naturel de considérer la loi uniforme sur Ω, i.e. $\mathbb{P}(\{\omega_1,\omega_2\}) = 1/36$.
Par conséquent, on a

k	1	2	3	4	5	6
$\mathbb{P}(X = k)$	11/36	9/36	7/36	5/36	3/36	1/36

On vérifie rapidement qu'il s'agit d'une mesure de probabilité, car la somme vaut bien 1.

Definition 16. *On dit que deux variables X et Y à valeurs dans le même espace $(E,\mathcal{E})$ sont de même loi si $\mathbb{P}_X = \mathbb{P}_Y$. On notera alors $\mathcal{L}(X) = \mathcal{L}(Y)$ ou $X \overset{\mathcal{L}}{=} Y$.*

On remarque que deux variables peuvent avoir la même loi sans pour autant être définie sur le même espace de probabilité (on identifie seulement les mesures images).

Remarque. *D'après le théorème de la classe monotone, si $\mathcal{E} = \sigma(\mathcal{C})$ où $\mathcal{C}$ est stable par intersection finie (π-système), alors X et Y sont de même loi si et seulement si $\mathbb{P}(X \in B) = \mathbb{P}(Y \in B)$ pour tout $B \in \mathcal{C}$. De même, si μ mesure de probabilité sur $(E,\mathcal{E})$, dans ce cas, X suit la loi μ si et seulement si $\mathbb{P}(X \in B) = \mu(B)$ pour tout $B \in \mathcal{C}$.*

Dans la suite, on aura tendance à introduire des v.a. en omettant de définir l'espace $(\Omega,\mathcal{F},\mathbb{P})$ sur lequel on se place. Cependant, il ne faut jamais oublier qu'une v.a. est une fonction.

À partir de la loi d'un vecteur aléatoire, on obtient facilement les lois des marginales.

Propriété 17. *Si* $(X,Y) \in E^2$ *est un couple de loi* $\mathbb{P}_{(X,Y)}$, *alors les lois marginales* $\mathbb{P}_X$ *et* $\mathbb{P}_Y$ *de* X *et* Y *sont données par : pour* $A \in \mathcal{E}$

$$\mathbb{P}_X(A) = \mathbb{P}_{(X,Y)}(A \times E) \text{ et } \mathbb{P}_Y(A) = \mathbb{P}_{(X,Y)}(E \times A)$$

Plus généralement, si on considère un $n-$*uplet,*

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

Démonstration. La preuve est évidente car $\{X \in A\} = \{(X,Y) \in A \times E\}$. □

La loi d'un vecteur aléatoire détermine la loi de chacune des marginales, mais la réciproque est fausse. De la loi des marginales, on ne peut pas en déduire (en général) la loi du vecteur.

Exemple. Soit (X,Y) un couple aléatoire de loi

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

et (U,V) un autre couple de loi

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

On a $\mathbb{P}_{(X,Y)} \neq \mathbb{P}_{(U,V)}$ et pourtant

$$\mathbb{P}_X = \mathbb{P}_U = \frac{1}{2}\delta_0 + \frac{1}{2}\delta_1$$
$$\mathbb{P}_Y = \mathbb{P}_V = \frac{1}{4}\delta_0 + \frac{3}{4}\delta_1.$$

Definition 18. *Soit* $X : (\Omega, \mathcal{F}, \mathbb{P}) \to (E, \mathcal{E})$ *une variable aléatoire. On appelle atome de (la loi de)* X *tout point* $x \in E$ *tel que* $\mathbb{P}(X = x) \neq 0$.

On appellera support d'une variable aléatoire $X : (\Omega, \mathcal{F}, \mathbb{P}) \to (E, \mathcal{E})$ le plus petit événement S de E tel que $\mathbb{P}(X \in S) = 1$ (cet objet est mal défini dans le cas général, il faut juste en retenir le sens intuitif). Dans la suite de ce chapitre, nous allons considérer des variables à valeurs réelles : $(E, \mathcal{E}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

3 Fonction de répartition

Definition 19. *On appelle fonction de répartition d'une variable aléatoire* $X : (\Omega, \mathcal{F}, \mathbb{P}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ *la fonction* F_X *définie sur* $\mathbb{R}$ *à valeurs dans* $[0,1]$ *par*

$$F_X(x) = \mathbb{P}(X \leqslant x) = \mathbb{P}_X(]-\infty, x]).$$

Proposition 20. *La fonction de répartition caractérise la loi d'un variable aléatoire, i.e. si* X, Y *deux v.a. telles que* $F_X = F_Y$ *alors* $\mathbb{P}_X = \mathbb{P}_Y$.

Démonstration. La famille $\mathcal{C} = \{]-\infty, x], X \in \mathbb{R}\}$ est stable par intersection finie et engendre $\mathcal{B}(\mathbb{R})$, on conclut par le théorème de la classe monotone. □

Exemple. On lance deux dés et on regarde X le minimum des deux chiffres obtenu. Tracer la fonction de répartition de X.

Propriété 21. *Soit F_X la fonction de répartition d'une variable aléatoire réelle X.*

1. *F_X est une fonction croissante,*
2. $\lim_{x\to-\infty} F_X(x) = 0$ *et* $\lim_{x\to+\infty} F_X(x) = 1$,
3. *F_X est une fonction continue à droite et admet une limite à gauche,*
$$F_X(x^-) = \lim_{\substack{y\to x \\ y<x}} F_X(y) = \mathbb{P}(X < x),$$
4. $\mathbb{P}(X \in [a,b]) = F_X(b) - F_X(a^-)$,
5. *si x n'est pas un atome de X alors F_X est continue à gauche en x (donc continue). L'ensemble des points de discontinuité de F_X est l'ensemble des atomes de X.*

Démonstration. 1. Si $x \leqslant y$, on a $\{X \leqslant x\} \subset \{X \leqslant y\}$ et donc par croissance de la mesure de probabilité $\mathbb{P}$, $F_X(x) \leqslant F_X(y)$.

La fonction F_X étant monotone, elle admet des limites à gauche et à droite en tout point et l'ensemble de ses points de discontinuité est au plus dénombrable.

2. $\lim_{x\to-\infty} F_X(x) = \lim_{n\to+\infty} F_X(-n)$. On pose $A_n =]-\infty, -n]$ qui est une suite décroissante de boréliens avec $\bigcap_{n\geqslant 1} A_n = \emptyset$. Par conséquent, $\lim_{n\to+\infty} F_X(-n) = \mathbb{P}_X(\emptyset) = 0$. De même, $\lim_{x\to+\infty} F_X(x) = \lim_{n\to+\infty} F_X(n)$. On pose $A_n =]-\infty, n]$ qui est une suite croissante de boréliens avec $\bigcup_{n\geqslant 1} A_n = \mathbb{R}$. Par conséquent, $\lim_{n\to+\infty} F_X(n) = \mathbb{P}_X(\mathbb{R}) = 1$.
3. On introduit la suite décroissante $A_n =]-\infty, x+1/n]$. On a $\bigcap A_n =]-\infty, x]$ et donc $\lim_{y\to x^+} F_X(y) = \lim_{n\to\infty} F_X(x+1/n) = \lim_{n\to\infty} \mathbb{P}_X(A_n) = \mathbb{P}_X(]-\infty, x]) = F_X(x)$. La fonction est continue à droite.
On introduit maintenant la suite croissante $A_n =]-\infty, x-1/n]$. On a $\bigcup A_n =]-\infty, x[$ et donc $\lim_{y\to x^-} F_X(y) = \lim_{n\to\infty} \mathbb{P}_X(A_n) = \mathbb{P}_X(]-\infty, x[)$. La fonction est admet une limite à gauche qui vaut $\mathbb{P}(X < x)$.
4. $\mathbb{P}(X \in [a,b]) = \mathbb{P}(X \leqslant b) - \mathbb{P}(X < a) = F_X(b) - F_X(a^-)$ car $[a,b] =]\infty, b] \setminus]-\infty, a[$.
5. En général, $\mathbb{P}(X < x) \neq \mathbb{P}(X \leqslant x)$. En fait, $\mathbb{P}(X \leqslant x) = \mathbb{P}(X < x) + \mathbb{P}(X = x)$ car ensembles disjoints et donc la fonction est continue en x si et seulement si x n'est pas un atome.

□

Proposition 22. *La fonction de répartition admet au plus un nombre dénombrable de points de discontinuité. Une variable aléatoires réelle a par conséquent un nombre au plus dénombrable d'atomes.*

Démonstration. Ceci est lié à la monotonie de la fonction. On peut aussi le redémontrer pour la cas particulier des fonctions de répartition. Si on note $D_n = \{x : F_X(x) - F_X(x^-) \geqslant 1/n\}$ l'ensemble des points de discontinuité avec un saut d'amplitude supérieur à $1/n$, comme $F_X(x) \in [0,1]$, on a card $D_n \leqslant n$. Donc l'ensemble des points de discontinuité $D = \cup_{n\geqslant 1} D_n$ est dénombrable. □

Theorem 23. *Une fonction $F : \mathbb{R} \to [0,1]$ croissante càdlàg (continue à droite et limitée à gauche) avec* $\lim_{x\to-\infty} F(x) = 0$ *et* $\lim_{x\to+\infty} F(x) = 1$ *est la fonction de répartition d'une v.a. X. De plus, l'ensemble des points où la fonction a un saut est l'ensemble des atomes de X.*

Démonstration. (en exo)
La preuve est complexe si on souhaite construire X sur un espace de probabilité général $(\Omega, \mathcal{F}, \mathbb{P})$ et repose sur le théorème de Carathéodory. On se place dans le cas particulier $(]0,1[, \mathcal{B}(]0,1[), \lambda)$ où λ mesure de Lebesgue sur $]0,1[$. Soit $\omega \in]0,1[$. On note $X(\omega) = \inf\{t \in \mathbb{R} : F(t) \geqslant \omega\}$. Ceci est bien défini car F est croissante et $\overline{F(\mathbb{R})} = [0,1]$. Si F est strictement croissante, alors elle est inversible et $X(\omega) = F^{-1}(\omega)$.

X est une variable aléatoire car F étant càdlàg, elle est mesurable et donc $\{t \in \mathbb{R} : F(t) \geqslant \omega\}$ est un ensemble mesurable.
Montrons que $\{t \in \mathbb{R} : F(t) \geqslant \omega\} = [X(\omega), +\infty[$. On a de façon évidente $\{t \in \mathbb{R} : F(t) \geqslant \omega\} \subset [X(\omega), +\infty[$. Par ailleurs, soit $t \geqslant X(\omega)$, alors par croissance $F(t) \geqslant F(X(\omega))$ et par continuité à droite $F(X(\omega)) \geqslant \omega$, donc $F(t) \geqslant \omega$ et $t \in \{t \in \mathbb{R} : F(t) \geqslant \omega\}$. CQFD.
Par ailleurs, vérifions que la fonction de répartition de X est bien F. On a, en effet,

$$\begin{aligned}
\mathbb{P}(X \leqslant t) &= \mathbb{P}(\{\omega \in]0,1[: X(\omega) \leqslant t\}) \\
&= \mathbb{P}(\{\omega \in]0,1[: t \in [X(\omega), +\infty[)\}) \\
&= \mathbb{P}(\{\omega \in]0,1[: F(t) \geqslant \omega\}) \\
&= \int_0^{F(t)} d\lambda(\omega) = F(t).
\end{aligned}$$

□

Definition 24. *Soit $X = (X_1, \dots, X_d)$ un vecteur aléatoire de $\mathbb{R}^d$. On appelle fonction de répartition de X la fonction définie sur $\mathbb{R}^d$ par*

$$F_X(t) = F_X(t_1, \dots, t_d) = \mathbb{P}(]-\infty, t_1] \times \ldots \times]-\infty, t_d]) = \mathbb{P}(X_1 \leqslant t_1, \dots, X_d \leqslant t_d).$$

On remarque que la fonction de répartition de la $i^{\text{ème}}$ marginale vérifie $F_{X_i}(t_i) = \lim_{\substack{t_j \to +\infty \\ j \neq i}} F_X(t)$.

4 Exemples de variables aléatoires

4.1 Variables discrètes

Definition 25. *Une v.a. $X : (\Omega, \mathcal{F}, \mathbb{P}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ est dite discrète si son support est au plus dénombrable.*

Dans ce cas, son support est exactement l'ensemble $\mathcal{A}$ de ses atomes.
Rappel : La mesure de dirac δ_a en $a \in \mathbb{R}$ est définie pour tout borélien $A \in \mathcal{B}(\mathbb{R})$

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

Propriété 26. *Si X est une variable aléatoire discrète alors sa loi est la somme pondérée des mesures de Dirac en ses atomes*

$$\mathbb{P}_X = \sum_{x \in \mathcal{A}} p_x \delta_x,$$

avec $p_x = \mathbb{P}(X = x)$.

Démonstration. Comme $\mathcal{A}$ est au plus dénombrable et $\mathcal{A} = \bigcup_{x \in \mathcal{A}} \{x\}$, pour tout $A \in \mathcal{B}(\mathbb{R})$, on a

$$\mathbb{P}_X(A) = \mathbb{P}_X\left(A \cap \bigcup_{x \in \mathcal{A}} \{x\}\right) = \sum_{x \in \mathcal{A}} \mathbb{P}_X(A \cap \{x\}) = \sum_{x \in \mathcal{A}} \mathbb{P}(X = x, x \in A) = \sum_{x \in \mathcal{A}} \mathbb{P}(X = x)\delta_x(A).$$

□

Remarque. *Pour caractériser la loi d'une v.a. discrète, il suffit de donner l'ensemble de ses atomes et d'exprimer* $\mathbb{P}(X=x)$ *pour tout atome* x.

Proposition 27. *Soit* X *une v.a. discrète et* $\mathcal{A}=\{x_i : i\in I\}$ *l'ensemble de ses atomes (supposés ordonnés :* $x_i<x_{i+1}$*). La fonction de répartition* F_X *de* X *est croissante, constante sur chaque intervalle* $[x_k,x_{k+1}[$ *avec un saut* $p_k=\mathbb{P}(X=x_k)$ *en chaque atome* x_k *:* $p_k=F_X(x_k)-F_X(x_{k-1})$

Démonstration. Si $x<\min(x_i)$, alors $F_X(x)=0$ et si $x>\sup(x_i)$, $F_X(x)=1$ car en dehors du support. Soient $s,t\in[x_k,x_{k+1}[$ avec $s<t$, $F_X(t)-F_X(s)=\mathbb{P}(X\in]s,t])=\sum_{i:s<x_i\leqslant t}\mathbb{P}(X=x_i)=0$ car la somme est vide. La fonction est bien constante par intervalle. $\square$

Exemple.
- Une variable aléatoire constante est discrète, $X=c$ a pour loi $\mathbb{P}_X=\delta_c$.
- X une variable aléatoire qui prend les valeurs $x_1\ldots,,x_n$ avec probabilité respectives $p_1,\ldots,p_n$ a pour loi $\mathbb{P}_X=p_1\delta_{x_1}+\ldots,p_n\delta_{x_n}$.
- Lorsque tous les atomes ont la même probabilité de se réaliser on parle de loi équirépartie (équiprobable ou encore uniforme). Si X est équirépartie sur $\{x_1\ldots,,x_n\}$ alors $\mathbb{P}_X=\frac{1}{n}\sum_{k=1}^n\delta_{x_k}$.
- Si $A\in\mathcal{F}$ et $X=\mathbb{1}_A$ alors X est discrète à valeurs dans $\{0,1\}$ et $\mathbb{P}_X=\mathbb{P}(A)\delta_1+(1-\mathbb{P}(A))\delta_0$. Cette loi est appelée loi de Bernoulli $\mathcal{B}(p)$ avec $p=\mathbb{P}(A)$.
  De manière générale, une variable qui ne prend que deux valeurs suit une loi de Bernoulli.
- On appelle loi Binomiale $\mathcal{B}(n,p)$avec $n\in\mathbb{N}^*$ et $p\in]0,1[$ la loi d'une v.a. X à valeurs dans $\{0,\ldots,n\}$ avec
$$\mathbb{P}_X=\sum_{k=0}^n\binom{k}{n}p^k(1-p)^{n-k}\delta_k.$$
  On trouve cette loi quand on considère le nombre de succès dans une suite de n épreuves identiques ayant chacune une probabilité de succès p.
- On appelle loi Géométrique $\mathcal{G}(p)$ avec $p\in]0,1[$ la loi d'une v.a. à valeurs dans $\mathbb{N}^*$ avec
$$\mathbb{P}_X=\sum_{k=1}^\infty p(1-p)^{k-1}\delta_k.$$
  On trouve cette loi quand on considère la position du premier succès dans une suite d'épreuves aléatoires identiques de probabilité de succès p.
- On appelle loi de Poisson $\mathcal{P}(\lambda)$ avec $\lambda>0$ la loi d'une v.a. à valeurs dans $\mathbb{N}$ avec
$$\mathbb{P}_X=\sum_{k=0}^\infty\frac{\lambda^k}{k!}e^{-\lambda}\delta_k.$$
  Cette loi est souvent utilisée pour modéliser des événements rares.

Remarque. *Soit* X *une v.a. discrète à valeurs dans* $\{x_i : i\in I\}$ *et on note* $p_i=\mathbb{P}(X=x_i)$*. Soit* $g:\mathbb{R}\to\mathbb{R}$ *une application mesurable. Alors* $Y=g(X)$ *est une v.a. discrète à valeurs dans* $\{y_j : j\in J\}=g(\{x_i : i\in I\})$ *et*
$$\mathbb{P}(Y=y_j)=\sum_{i\in I:g(x_i)=y_j}p_i.$$

Démonstration. Il est clair que Y est à valeurs dans $g(\{x_i : i\in I\})$. Par ailleurs, on a $g^{-1}(\{y_j\})=\{x_i : g(x_i)=y_j\}$ et donc
$$\mathbb{P}(Y=y_j)=\mathbb{P}(g\circ X=y_j)=\mathbb{P}\big(X^{-1}\big(g^{-1}(\{y_j\})\big)\big)=\mathbb{P}_X\big(g^{-1}(\{y_j\})\big)=\sum_{x_i\in g^{-1}(\{y_j\})}\mathbb{P}(X=x_i).$$
$\square$

Rappel : Loi d'un vecteur aléatoire discret. Soit (X, Y) à valeurs dans $\mathcal{A} = \{(x_i, y_j) : i \in I, j \in J\}$. Les lois des marginales sont obtenues par projection : pour $i \in I$, $\mathbb{P}_X(x_i) = \mathbb{P}(X = x_i) = \sum_{j \in J} \mathbb{P}(X = x_i, Y = y_j)$ et pour $j \in J$, $\mathbb{P}_Y(y_j) = \mathbb{P}(Y = y_j) = \sum_{i \in I} \mathbb{P}(X = x_i, Y = y_j)$.

Exemple. On lance deux dés équilibrés. On note X le nombre de chiffre pair et Y le minimum des deux chiffres. Donner la loi du couple (X, Y). Calculer la probabilité de ne pas avoir de chiffre pair.

4.2 Théorème de radon-Nikodym

Definition 28. *Soit μ et ν deux mesures définies sur un espace mesurable $(\Omega, \mathcal{F})$. On dit que μ est absolument continue par rapport à ν, noté $\mu << \nu$, si pour tout $A \in \mathcal{F}$,*

$$\nu(A) = 0 \textit{ implique } \mu(A) = 0.$$

Theorem 29 (Radon-Nikodym). *Soient μ et ν deux mesures $\sigma-$finies. Si $\mu << \nu$ alors il existe une fonction $f : (\Omega, \mathcal{F}) \to \mathbb{R}$ mesurable telle que*

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

La fonction f s'appelle la densité de μ par rapport à ν. Elle est positive si les mesures sont positives. Elle est souvent notée $f = \frac{d\mu}{d\nu}$. De plus, on a le lien suivant (sous condition que les intégrales soient bien définies)

$$\int g d\mu = \int g f d\nu.$$

Si μ est une mesure finie alors $f \in L^1(\nu)$.

Exemple.

1. soit $\mu = \sum_{k \geqslant 0} p_k \delta_k$ avec p_k des réels positifs, alors $\mu << \nu$ avec $\nu = \sum_{k \geqslant 0} \delta_k$ la mesure de comptage sur $\mathbb{N}$. La densité est donnée par $f(k) = p_k$.
2. Soit λ la mesure de Lebesgue. Aucune loi discrète n'est absolument continue par rapport à la mesure de Lebesgue car une loi discrète possède des atomes : $\mathbb{P}(X = x) > 0$ et $\lambda(\{x\}) = 0$. Il suffit d'avoir un atome pour qu'une loi ne soit pas absolument continue par rapport à la mesure de Lebesgue.

4.3 Variables à densité

On appelle variables à densité les variables dont la loi est absolument continue par rapport à la mesure de Lebesgue. On note λ la mesure de Lebesgue sur $\mathbb{R}$.

Definition 30. *Une variable aléatoire X est dite à densité f si $\mathbb{P}_X << \lambda$, i.e.*

$$\forall A \in \mathcal{B}(\mathbb{R}) \quad \mathbb{P}(X \in A) = \int_A f d\lambda.$$

On a alors $\mathbb{P}(X \in [a, b]) = \int_a^b f(x) dx$ pour tout $a, b \in \mathbb{R}$.

On note $f(x)dx$ la loi d'une v.a. X à densité.

Remarque. *La fonction f est mesurable positive et $\int_{\mathbb{R}} f(x) dx = 1$. De façon réciproque, une telle fonction est une densité de probabilité.*

Exemple.
1. On appelle loi uniforme sur $[a,b]$, notée $\mathcal{U}([a,b])$ la loi de densité $\frac{1}{b-a}\mathbb{1}_{[a,b]}$.
2. On appelle loi exponentielle de paramètre $\alpha > 0$, notée $\mathcal{E}(\alpha)$, la loi de densité $\alpha e^{-\alpha x}\mathbb{1}_{x>0}$. Cette loi est souvent utilisée pour modéliser des durées de vie.
3. On appelle loi normale centrée réduite, notée $\mathcal{N}(0,1)$, la loi de densité $\frac{1}{\sqrt{2\pi}}e^{\frac{-x^2}{2}}$.
4. On appelle loi normale de moyenne m et de variance σ^2, notée $\mathcal{N}(m,\sigma^2)$, la loi de densité $\frac{1}{\sigma\sqrt{2\pi}}e^{\frac{-(x-m)^2}{2\sigma^2}}$.
5. On appelle loi de Cauchy $\mathcal{C}$ de paramètre $a > 0$ la loi de densité $\frac{a}{\pi(a^2+x^2)}$.

Proposition 31 (Lien entre fonction de répartition et densité)**.** *Soit X une variable aléatoire à densité f. Alors sa fonction de répartition vérifie*

1. *$\forall x \in \mathbb{R}$, $F_X(x) = \int_{-\infty}^{x} f(t)dt$.*
2. *F_X est continue sur $\mathbb{R}$.*
3. *Si f est continue en x_0, alors F_X est dérivable en x_0 de dérivée $F_x'(x_0) = f(x_0)$.*
4. *Si X v.a. ayant une fonction de répartition F de la forme $F(x) = \int_{-\infty}^{x} f(t)dt$ où f fonction mesurable positive, alors X admet nécessaire f pour densité.*

Démonstration.
1. $\forall x \in \mathbb{R}$, $F_X(x) = \mathbb{P}_X(]-\infty,x]) = \int_{-\infty}^{x} f(t)dt$.
2. Ok, car pas d'atome.
3. Si f est continue en x_0, donc définie sur un voisinage en x_0. Soit $\varepsilon > 0$, il existe $\delta > 0$ tel que $|t-x_0| < \delta \Rightarrow |f(t)-f(x_0)| < \varepsilon$. Soit h tel que $|h| < \delta$. On a alors
$$\begin{aligned}|F_X(x_0+h) - F_X(x_0) - hf(x_0)| &= \left|\int_{x_0}^{x_0+h} [f(t)-f(x_0)]dt\right| \\ &\leqslant \varepsilon|h|\end{aligned}$$
Par conséquent, le taux d'accroissement $(F_X(x_0+h) - F_X(x_0))/h$ converge vers $f(x_0)$ quand $|h| \to 0$.
4. F étant une fonction de répartition, f est une densité car $\int_{-\infty}^{+\infty} f(t)dt = \lim_{x\to\infty} F(x) = 1$. On introduit μ la mesure de probabilité définie par $\mu(A) = \int_A f(t)dt$ pour $A \in \mathcal{B}(\mathbb{R})$. Cette mesure est absolument continue par rapport à Lebesgue de densité f. $\mathbb{P}_X$ et μ coïncident sur $\mathcal{C} = \{]-\infty,x], x \in \mathbb{R}\}$ qui engendre $\mathbb{B}(\mathbb{R})$ et stable par intersection finie, donc $\mathbb{P}_X = \mu$ et $\mathbb{P}_X$ admet f pour densité.

$\square$

Proposition 32. *Une variable aléatoire X suit la loi normale $\mathcal{N}(m,\sigma^2)$ si et seulement si $Z = \frac{X-m}{\sigma} \sim \mathcal{N}(0,1)$.*

Démonstration. Si $X \sim \mathcal{N}(m,\sigma^2)$, alors la fonction de répartition de Z vérifie
$$\mathbb{P}(Z \leqslant t) = \mathbb{P}(X \leqslant \sigma t + m) = \int_{-\infty}^{\sigma t+m} e^{\frac{-(x-m)^2}{2\sigma^2}} \frac{dx}{\sqrt{2\pi}\sigma}.$$
En faisant le changement de variable, $y = (x-m)/\sigma$, on obtient
$$\mathbb{P}(Z \leqslant t) = \int_{-\infty}^{t} e^{\frac{-y^2}{2}} \frac{dy}{\sqrt{2\pi}}.$$
Donc $Z \sim \mathcal{N}(0,1)$. La réciproque est évidente. $\square$

Par conséquent, si $X \sim \mathcal{N}(m, \sigma^2)$, alors on peut écrire $X = \sigma Z + m$ avec $Z \sim \mathcal{N}(0, 1)$.

Avoir une fonction de répartition continue n'implique pas que la v.a. soit à densité. Les lois à densité sont en fait absolument continues : on dit que la fonction F est absolument continue sur $[a, b]$ si, pour tout $\varepsilon > 0$, il existe $\delta > 0$ tel que, pour toute suite $([a_n, b_n])_{n\in\mathbb{N}}$ de sous-intervalles de $[a, b]$ d'intérieurs disjoints,

$$\sum_{n\geqslant 0} (b_n - a_n) < \delta \Rightarrow \sum_{n\geqslant 0} |F(a_n) - F(b_n)| < \varepsilon.$$

En fait, F est absolument continue sur $[a, b]$ si et seulement s'il existe une fonction f intégrable sur $[a, b]$ telle que pour tout $x \in [a, b]$,

$$F(x) - F(a) = \int_a^x f(t)dt.$$

Exemple. Une fonction continue, dérivable presque partout et de dérivée dans L^1 n'est pas forcément l'intégrale de sa dérivée. En effet, l'escalier de Cantor est une fonction $f : [0, 1] \to [0, 1]$ continue dérivable presque partout (croissante) et de dérivée nulle presque partout.
On note K_3 l'ensemble triadique de Cantor. K_3 est un sous-ensemble compact de $[0, 1]$ d'intérieur vide donc totalement discontinu. Il est non dénombrable mais négligeable pour la mesure de Lebesgue. $K_3 = \bigcap_{n\in\mathbb{N}} K_n$ avec $K_0 = [0, 1]$, $K_1 = K_0 \backslash]1/3, 2/3[$, $\ldots$, K_n est la réunion de 2^n ensemble disjoints de longueur $1/3^n$: $K_n = \bigcup I_{q_n}$ avec $I_{q_n} = [q_n/3^n, q_{n+1}/3^n]$ et $q_n = \sum_{i=1}^n a_i 3^{n-i}$ avec $a_i \in \{0, 2\}$. Par conséquent,

$$K_3 = \left\{ \sum_{n=1}^{\infty} \frac{x_n}{3^n} : x_n \in \{0, 2\} \right\}.$$

On considère la suite de fonctions affines par morceaux définies sur $[0, 1]$ de la façon suivante :

$$f_0(x) = x$$
$$f_1(0) = 0,\ f(1) = 1,\ f_1([1/3, 2/3]) = 1/2$$

On construire la fonction f_{n+1} à partir de f_n : sur chaque intervalle $[u, v]$ où f_n est non constante, on remplace f_n par la fonction linéaire par morceaux qui vaut $\frac{f_n(u)+f_n(v)}{2}$ sur $[(2u + v)/3, (2v + u)/3]$. On a par conséquent

$$|f_{n+1}(x) - f_n(x)| \leqslant 2^{-n}$$

et donc la série $\sum_{n\geqslant 1}(f_{n+1} - f_n)(x)$ converge uniformément, i.e. (f_n) converge uniformément. Sa limite f est par conséquent continue, croissante, avec $f(0) = 0$ et $f(1) = 1$. De plus $f' = 0$ sur K_3^c car celui-ci est une réunion d'intervalles sur lesquelles f est constante.

Cas des vecteurs aléatoires à densité

Definition 33 (Densité en dimension d)**.** *Une fonction $f : \mathbb{R}^d \to \mathbb{R}$ est une densité de probabilité (en dimension d) si*

- *f est positive : $\forall t \in \mathbb{R}^d,\ f(t) \geqslant 0$,*
- *f est intégrable d'intégrale 1 : $\int_{\mathbb{R}^d} f(t)dt = 1$.*

Un vecteur aléatoire $X = (X_1, \ldots, X_d)$ suit la loi de densité f si pour tous les pavés $\Pi_{i=1}^d [a_i, b_i]$ de $\mathbb{R}^d$,

$$\mathbb{P}\Big((X_1, \ldots, X_d) \in \Pi_{i=1}^d [a_i, b_i]\Big) = \int_{\Pi_{i=1}^d [a_i, b_i]} f(t)dt = \int_{a_1}^{b_1} \ldots \int_{a_d}^{b_d} f(t_1, \ldots, t_d)dt_1 \ldots dt_d.$$

Proposition 34 (Théorème de changement de variables, pas fait en cours)**.** *Soit X un vecteur aléatoire de $\mathbb{R}^d$ de densité f et $g : \mathbb{R}^d \to \mathbb{R}^d$ un difféomorphisme (bijection continument différentiable, ainsi que son inverse). Alors $Y = g(X)$ est un vecteur aléatoire de densité*

$$f_Y(y) = f_X(g^{-1}(y))\big|Jg^{-1}(y)\big|$$

où $Jg = det\Big((\partial_j g_i)_{1\leqslant i.j\leqslant d}\Big)$ est le jacobien de g.

Démonstration. $\mathbb{P}(g(X) \in B) = \mathbb{P}(X \in g^{-1}(B)) = \int_{\mathbb{R}^d} \mathbb{1}_{x\in g^{-1}(B)} f_X(x)dx = \int_{\mathbb{R}^d} \mathbb{1}_{g(x)\in B} f_X(x)dx$. Puis appliquer le théorème de changement de variables. □

Proposition 35 (Densités marginales)**.** *Si (X, Y) est un couple de densité f alors X et Y sont des v.a. de à densité*

$$f_X(x) = \int_{\mathbb{R}} f(x, y)dy, \quad f_Y(y) = \int_{\mathbb{R}} f(x, y)dx.$$

Démonstration. Étudions la loi de X, la preuve étant identique pour Y. Soit $A \in \mathcal{B}(\mathbb{R})$. Par Fubini-Tonelli, on a

$$\mathbb{P}_X(A) = \mathbb{P}(X \in A, Y \in \mathbb{R}) = \iint_{A\times\mathbb{R}} f(x, y)dxdy = \int_A \left[\int_{\mathbb{R}} f(x, y)dy\right]dx.$$

Par ailleurs, $\int_{\mathbb{R}} f(x, y)dy$ est bien une densité de probabilité car positive et d'intégrale 1. □

Exemple.

— Vérifier que $f(x, y) = \frac{1}{3}\mathbb{1}_{[0,1]\times[-1,2]}(x, y)$ est une densité de probabilité sur $\mathbb{R}^2$. Montrer que $f_X(x) = \mathbb{1}_{[0,1]}(x)$ et $f_Y(y) = \frac{1}{3}\mathbb{1}_{[-1,2]}(y)$.

— Les densités marginales de $f(x, y) = ye^{-xy}\mathbb{1}_{\mathbb{R}^+}(x)\mathbb{1}_{[0,1]}(y)$ sont $f_X(x) = (1-e^{-x}-xe^{-x})/x^2\mathbb{1}_{\mathbb{R}^+}(x)$ et $Y \sim \mathcal{U}[0, 1]$.

— Vérifier que les lois marginales de $f(x, y) = \frac{1}{3\pi}e^{-\frac{x^2+2xy+5y^2}{6}}$ sont des lois gaussiennes.

Chapitre 3

Espérance d'une variable aléatoire

1 Définition et propriétés

Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité et $X : (\Omega, \mathcal{F}, \mathbb{P}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ une variable aléatoire. X est dite intégrable si

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

Definition 36. *On définit l'espérance d'une v.a. positive ou intégrable X, notée $\mathbb{E}[X]$, comme l'intégrale de X par rapport à la mesure $\mathbb{P}$:*

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

L'espérance représente la valeur moyenne prise par la variable.
On remarque que comme les mesures de probabilités sont de masse finie, les constantes sont intégrables et de même les v.a. bornées sont intégrables.

- Si $X = \mathbb{1}_A$ avec $A \in \mathcal{F}$ une v.a. de Bernoulli, alors $\mathbb{E}[X] = \mathbb{P}(A)$.
- Si $X = (X_1, \ldots, X_d)$ vecteur aléatoire de $\mathbb{R}^d$ avec $\mathbb{E}[|X_i|] < \infty$ $\forall i \in \{1, \ldots, d\}$, alors on définit

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

Tous les théorèmes de convergence en intégration se traduisent en terme d'espérance :

- Convergence monotone : Si X_n suite croissante de v.a. positives ($0 \leqslant X_n \leqslant X_{n+1}$) alors

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

- Lemme de Fatou : Si X_n suite de v.a. positives ($X_n \geqslant 0$ $\forall n \geqslant 0$) alors

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

- Convergence dominée : Si X_n suite de v.a. qui converge p.s vers X ($\exists N \in \mathcal{F} : \mathbb{P}(N) = 0$ tel que $\forall \omega \notin N$, $X_n(\omega) \to X(\omega)$) et telle que il existe Y une v.a. positive d'espérance finie ($\mathbb{E}[Y] < \infty$) telle que $\forall n \in \mathbb{N}$, $|X_n| \leqslant Y$ p.s., alors

$$\lim_{n\to\infty} \mathbb{E}[|X_n - X|] = 0 \quad \text{et donc} \quad \lim_{n\to\infty} \mathbb{E}[X_n] = \mathbb{E}[X].$$

Definition 37. *Une variable aléatoire intégrable est dite centrée si $\mathbb{E}[X] = 0$.*

L'espérance étant l'intégrale de X par rapport à la mesure $\mathbb{P}$, l'espérance vérifie les propriétés classiques de l'intégrale. De même on peut utiliser les théorème de Fubini-Tonelli et de Fubini.

Propriété 38. *Soit X, Y deux v.a. intégrables et $a, b \in \mathbb{R}$.*

1. *L'espérance st croissante : si $X \leqslant Y$ p.s., alors $\mathbb{E}[X] \leqslant \mathbb{E}[Y]$,*
2. *l'espérance est linéaire : $\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]$,*
3. $|\mathbb{E}[X]| \leqslant \mathbb{E}[|X|]$,
4. *Si X est une v.a. positive avec $\mathbb{E}[X] = 0$ alors $X = 0$ p.s.,*
5. *Inégalité de Markov : si X v.a. positive, alors pour tout $t > 0$*

$$\mathbb{P}(X \leqslant t) \leqslant \frac{\mathbb{E}[X]}{t}$$

Démonstration. Si X positive alors $t\mathbb{1}_{X \geqslant t} \leqslant X$ et on conclut en prenant l'espérance. □

Pour le calcul effectif de l'espérance d'une v.a., on se ramène au calcul d'une intégrale sur $\mathbb{R}$ grâce au théorème de transfert. En effet, Ω est en général un espace abstrait et $\mathbb{P}$ une mesure pas explicite, par contre $\mathbb{P}_X$ est la mesure image de $\mathbb{P}$ par l'application X. Par conséquent,

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

Proposition 39. *Soit $X : (\Omega, \mathcal{F}, \mathbb{P}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ une v.a. de loi $\mathbb{P}_X$. Alors X est intégrable si et seulement si*

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

et alors $\mathbb{E}[X] = \int_\mathbb{R} x\mathbb{P}_X(dx)$.
Généralisation : si $h : (\mathbb{R}, \mathcal{B}(\mathbb{R})) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ une fonction borélienne. Alors $h(X)$ est une variable aléatoire. $h(X)$ est $\mathbb{P}-$intégrable si et seulement si h est $\mathbb{P}_X-$intégrable : $\int_\mathbb{R} |h(x)| \mathbb{P}_X(dx) < \infty$ et alors

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

Proposition 40. *Soit X une v.a. positive. Alors*

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

De plus, $\mathbb{E}[X] < \infty$ si et seulement s'il existe $\varepsilon > 0$ (ou pour tout $\varepsilon > 0$) tel que

$$\sum_{n \geqslant 0} \mathbb{P}(X > \varepsilon n) < \infty \text{ ou } \sum_{n \geqslant 0} 2^n \mathbb{P}(X > \varepsilon 2^n) < \infty$$

Démonstration. Par Fubini-Tonelli $\int_0^{+\infty} \mathbb{P}(X > t) dt = \int_0^{+\infty} \mathbb{E}[\mathbb{1}_{X > t}] dt$.
Pour tout $t \in [n, n+1]$, $\mathbb{P}(X > n+1) \leqslant \mathbb{P}(X > t) \leqslant \mathbb{P}(X > n)$. Donc $\mathbb{P}(X > n+1) \leqslant \int_n^{n+1} \mathbb{P}(X > t) dt \leqslant \mathbb{P}(X > n)$ et en prenant la somme sur n, on obtient

$$\sum_{n \geqslant 1} \mathbb{P}(X > n) \leqslant \int_0^\infty \mathbb{P}(X > t) dt \leqslant \sum_{n \geqslant 0} \mathbb{P}(X > n).$$

De même, on remarque que $[1,\infty[=\bigcup_{n=0}^{\infty}[2^n,2^{n+1}[$ et donc

$$\sum_{n\geqslant 1}2^n\mathbb{P}(X>2^{n+1})\leqslant\int_1^{\infty}\mathbb{P}(X>t)dt\leqslant\sum_{n\geqslant 0}2^n\mathbb{P}(X>2^n).$$

Par conséquent,

$$\sum_{n\geqslant 1}2^n\mathbb{P}(X>2^{n+1})\leqslant\int_0^{\infty}\mathbb{P}(X>t)dt\leqslant 1+\sum_{n\geqslant 0}2^n\mathbb{P}(X>2^n).$$

On conclut en remplaçant X par X/ε. □

Proposition 41 (Inégalité de Jensen)**.** *Soient $\varphi:\mathbb{R}\to\mathbb{R}$ une fonction convexe et X une v.a. réelle tels que X et $\varphi(X)$ soit intégrables, alors*

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

Démonstration. Une fonction convexe est le sup de fonctions affines :

$$\varphi(x)=\sup\{f(x):f\text{ affine telle que }f\leqslant\varphi\}.$$

Par linéarité et croissance de l'espérance, si f est affine avec $f\leqslant\varphi$, alors $f(\mathbb{E}[X])=\mathbb{E}[f(X)]\leqslant\mathbb{E}[\varphi(X)]$. On conclut en prenant le sup. □

Exemple. On a $\mathbb{E}[X]^+\leqslant\mathbb{E}[X^+]$, $\mathbb{E}[X]^2\leqslant\mathbb{E}[X^2]$, $\exp(\lambda\mathbb{E}[X])\leqslant\mathbb{E}[e^{\lambda X}]$ pour $\lambda>0$, $\mathbb{E}[X]\vee c\leqslant\mathbb{E}[X\vee c]$, $1/\mathbb{E}[X]\leqslant\mathbb{E}[1/X]$ si X positive.

On remarque que si φ est une fonction mesurable telle que $\varphi(\mathbb{E}[X])\leqslant\mathbb{E}[\varphi(X)]$ pour toute v.a. X, alors φ est convexe (il suffit de prendre X de loi $p\delta_x+(1-p)\delta_y$, avec $x,y\in\mathbb{R}$).

2 Exemples de calcul d'espérance

2.1 Espérance d'une v.a. discrète

Soit X une v.a. de loi $\mathbb{P}_X=\sum_{i\in I}p_i\delta_{x_i}$ où I ensemble au plus dénombrable et $x_i\in\mathbb{R}$. Alors X est intégrable si et seulement si $\sum_{i\in I}|x_i|p_i<\infty$ et dans ce cas

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

On remarque que si l'ensemble des atomes $\mathcal{A}$ de X est fini alors X est toujours intégrable.
Soit $h:\mathbb{R}\to\mathbb{R}$ mesurable, alors $h(X)$ intégrable si et seulement si $\sum_{i\in I}|h(x_i)|p_i<\infty$ et dans ce cas

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

Exemple.

1. Si $X=c$ est constance, alors $\mathbb{E}[X]=c$.
2. Si $X\sim\mathcal{B}(p)$, alors $\mathbb{E}[X]=p$.
3. Si X est équirépartie sur I de cardinal n, alors $\mathbb{E}[X]=\frac{1}{n}\sum_{i=1}^n x_i$.

4. Si $X \sim \mathcal{B}(n,p)$ suit la loi binomiale, alors $\mathbb{E}[X] = np$. En effet

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

5. Si $X \sim \mathcal{G}(p)$ de loi géométrique de paramètre $p \in]0,1[$, alors $\mathbb{E}[X] = 1/p$. En effet,

$$\mathbb{E}[X] = p\sum_{k=1}^{\infty} k(1-p)^{k-1} = p\sum_{k=0}^{\infty}\left(x^k\right)'(1-p).$$

La série entière $\sum_{k\geqslant 0} x^k$ a pour rayon de convergence $R = 1$ et donc est dérivable à l'intérieur de son disque de convergence. Comme $p \in]0,1[$,

$$\mathbb{E}[X] = p\left(\sum_{k=0}^{\infty} x^k\right)'(1-p) = p\left(\frac{1}{1-x}\right)'(1-p) = \frac{p}{(1-(1-p))^2} = \frac{1}{p}.$$

6. Si $X \sim \mathcal{P}(\lambda)$ de loi de Poisson de paramètre $\lambda > 0$, alors $\mathbb{E}[X] = \lambda$. En effet,

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

2.2 Espérance d'une variable à densité

Soit $X : (\Omega, \mathcal{F}, \mathbb{P}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ une v.a. de densité f, alors X est intégrable si et seulement si $\mathbb{E}[|X|] = \int_{\mathbb{R}} |x|f(x)dx < \infty$ et dans ce cas

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

De plus, si $h : \mathbb{R} \to \mathbb{R}$ mesurable, alors $h(X)$ intégrable si et seulement si $\int_{\mathbb{R}} |h(x)|f(x)dx < \infty$ et dans ce cas

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

Exemple.
1. Si $X \sim \mathcal{U}([a,b])$ de loi uniforme sur $[a,b]$, alors X intégrable car bornée et $\mathbb{E}[X] = \frac{a+b}{2}$.
2. Si $X \sim \mathcal{E}(\alpha)$ de loi exponentielle de paramètre $\alpha > 0$, alors $\mathbb{E}[X] = \frac{1}{\alpha}$.
3. Si $X \sim \mathcal{C}$ de loi de Cauchy, $\mathbb{E}[|X|] = \frac{2}{\pi}\int_0^{\infty} \frac{x}{1+x^2}dx = \infty$. Non intégrable !
4. Si $X \sim \mathcal{N}(0,1)$ de loi normale centrée réduite, alors X est intégrable et $\mathbb{E}[X] = 0$.
5. Si $X \sim \mathcal{N}(m,\sigma^2)$ de loi normale, alors par changement de variable $X = \sigma Z + m$ avec $Z \sim \mathcal{N}(0,1)$, on en déduit que X est intégrable et $\mathbb{E}[X] = m$.

2.3 Quelques autres lois

Supposons que la loi de X soit la combinaison convexe de deux mesures de probabilité μ et ν : il existe $\alpha \in [0,1]$ tel que

$$\mathbb{P}_X = \alpha\mu + (1-\alpha)\nu$$

alors

$$\mathbb{E}[X] = \alpha \int x\mu(dx) + (1-\alpha) \int x\nu(dx).$$

Par exemple, $\mathbb{P}_X = \frac{1}{2}\delta_2 + \frac{1}{2}\mathbb{1}_{[0,1]}(x)dx$. X est une v.a. positive et

$$\mathbb{E}[X] = \int_{\mathbb{R}} x\mathbb{P}_X(dx) = \frac{1}{2}2 + \frac{1}{2}\int_{\mathbb{R}} x\mathbb{1}_{[0,1]}(x)dx = \frac{1}{4}.$$

De manière plus générale, si la loi de X est une combinaison convexe d'une loi discrète et d'une loi à densité,

$$\mathbb{P}_X = \sum_{i=1}^n p_i\delta_{x_i} + f(x)dx$$

où $p_i > 0$ et h fonction mesurable positive, tels que $\sum_{i=1}^n p_i + \int_{\mathbb{R}} f(x)dx = 1$, alors la variable est intégrable si $\sum_{i=1}^n p_i|x_i| + \int_{\mathbb{R}} |x|f(x)dx$ et dans ce cas

$$\mathbb{E}[X] = \sum_{i=1}^n p_i x_i + \int_{\mathbb{R}} xf(x)dx.$$

3 Identification de la loi d'une variable aléatoire par l'espérance

L'espérance d'une variable aléatoire, représentant juste la valeur moyenne de la variable, ne peut pas par conséquent caractériser la loi de la variable (il est facile d'imaginer que deux variables aient la même moyenne sans pour autant avoir la même loi). On peut cependant utiliser l'espérance pour caractériser la loi d'une variable.

Theorem 42. *Soit $X : (\Omega, \mathcal{F}, \mathbb{P}) \to (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ un vecteur aléatoire et μ une mesure de probabilité sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. Alors la loi de X est μ si et seulement si l'un des conditions suivantes est réalisée*

1. *Pour tout $A \in \mathcal{B}(\mathbb{R}^d)$, $\mathbb{P}_X(A) = \mathbb{P}(X \in A) = \mu(A)$,*
2. *la fonction de répartition satisfait pour tout $t = (t_1, \ldots, t_d) \in \mathbb{R}^d$, en notant $]-\infty, t] = \Pi_{i=1}^d]-\infty, t_i]$, $F_X(t) = \mathbb{P}(X_1 \leqslant t_1, \ldots, X_d \leqslant t_d) = \mu(]-\infty, t])$,*
3. *pour toute fonction $h : \mathbb{R}^d \to \mathbb{R}$ continue à support compact positive (C^∞ à support compact positive suffit),*
$$\mathbb{E}[h(X)] = \int_{\mathbb{R}^d} h(x)\mu(dx).$$

Démonstration. Il suffit de vérifier la dernière équivalence. Notons ν la loi de X. Supposons que

$$\mathbb{E}[h(X)] = \int_{\mathbb{R}^d} h(x)\nu(dx) = \int_{\mathbb{R}} h(x)\mu(dx)$$

pour tout fonction h continue à support compact positive. Soit $K \in \mathcal{B}(\mathbb{R}^d)$ un ensemble compact de $\mathbb{R}^d$. Il existe un ouvert O tel que $K \subset O$ et l'adhérence $\overline{O}$ compact. On introduit la fonction

$$h(x) = \frac{d(x, O^c)}{d(x, K) + d(x, O^c)}.$$

La fonction h est positive, bornée par 1, continue sur $\mathbb{R}^d$, à support compact car le complémentaire $O^c = \{x : h(x) = 0\}$ et qui vaut 1 si et seulement si $x \in K$. Par conséquent,

$$\mathbb{1}_K \leqslant h(x) \leqslant \mathbb{1}_O.$$

On construit une suite décroissante d'ouverts O_n telle que $\lim_{n\to\infty} O_n = K$ (par exemple $O_n = \{x : d(x,K) < 1/n\}$ ce qui donne en dimension 1, pour $K = [a,b]$, $O_n =]a - 1/n, b + 1/n[$).
Comme $n \to d(., O_n^c)$ est décroissante et $u \to u/(a+u)$ est croissante, la suite h_n est décroissante vers $\mathbb{1}_K$. Par convergence dominée (ou convergence monotone car $(1-h_n)_{n\geqslant 0}$ est positive croissante),

$$\lim_{n\to\infty} \nu(h_n) = \nu(K) \quad \text{ et } \lim_{n\to\infty} \mu(h_n) = \mu(K).$$

Par hypothèse, $\forall n \geqslant 0$, $\nu(h_n) = \mu(h_n)$. Donc $\mu(K) = \nu(K)$ pour tout compact K, donc les mesures sont égales.
Réciproquement, Il est clair que si les mesures sont égales, on a égalité des intégrales.
On remarque qu'on peut se restreindre aux fonctions $\mathcal{C}^\infty$ positives à support compact via le théorème de convergence dominée, car les fonctions $\mathcal{C}_c^\infty$ sont denses dans l'espace des fonctions continues pour la norme infinie. □

Le théorème précédent est très utile lorsqu'on cherche la loi d'une variable Y de la forme $Y = g(X)$ où X est une variable aléatoire de loi connue. En effet, on peut exprimer l'espérance d'une fonctionnelle de Y via la loi de X : si h est mesurable bornée (ou mesurable positive, peu importe), alors

$$\int h(y)\mathbb{P}_Y(dy) = \mathbb{E}[h(Y)] = \mathbb{E}[h(g(X))] = \int h \circ g(x)\mathbb{P}_X(dx).$$

C'est notamment très utile si X est à densité sur $\mathbb{R}^d$.

Exemple. 1. Supposons que $X \sim \mathcal{U}(]0,1[)$ et cherchons la loi de $Y = -\ln(X)$. Soit h une fonction mesurable positive, alors

$$\mathbb{E}[h(Y)] = \mathbb{E}[h(-\ln(X))] = \int_0^1 h(-\ln(x))dx.$$

En utilisant le changement de variables $y = -\ln(x)$, on a donc pour toute fonction h mesurable positive

$$\mathbb{E}[h(Y)] = \int_0^\infty h(y)e^{-y}dy.$$

Y suit la loi exponentielle de paramètre 1.

2. Soit X une variable de loi de Cauchy. Cherchons la loi de $Y = X^+$.
Soit h une fonction mesurable positive, alors

$$\begin{aligned}\mathbb{E}[h(Y)] = \mathbb{E}[h(X^+)] &= \int_{\mathbb{R}} h(x^+)\frac{1}{\pi(1+x^2)}dx \\ &= h(0)\mathbb{P}(X \leqslant 0) + \int_0^\infty h(x)\frac{1}{\pi(1+x^2)}dx \\ &= h(0)\frac{1}{2} + \int_{\mathbb{R}} h(x)\frac{1}{\pi(1+x^2)}\mathbb{1}_{x>0}dx.\end{aligned}$$

Par conséquent, $\mathbb{P}_Y = \frac{1}{2}\delta_0 + \frac{1}{\pi(1+x^2)}\mathbb{1}_{x>0}dx$.

4 Moments d'une variable aléatoire

Definition 43. *Une variable aléatoire $X : (\Omega, \mathcal{F}, \mathbb{P}) \to (E, \mathcal{E})$, avec $E = \mathbb{R}$, $\mathbb{C}$ ou $\mathbb{R}^d$, admet un moment d'ordre $p > 0$ si*

$$\mathbb{E}[|X|^p] = \int_\Omega |X(\omega)|^p \mathbb{P}(d\omega) = \int_E |x|^p \mathbb{P}_X(dx) < \infty.$$

On définit l'espace des variables ayant un moment d'ordre $p > 0$:

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

Pour $p = \infty$, on définit l'ensemble des variables p.s. bornées,

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

Pour $X \in L^p$, on introduit $||X||_p = \mathbb{E}[|X|^p]^{1/p}$ et pour $X \in L^\infty$, $||X||_\infty = \inf\{c > 0 : \mathbb{P}(|X| > c) = 0\}$.

Du fait de la convexité de $x \mapsto x^p$ pour $p \geqslant 1$, on déduit facilement que pour $p \in [1, +\infty]$, L^p est un espace vectoriel.

Exemple. Soit X une v.a. de loi $\mathcal{N}(0, 1)$, alors X a des moments finis de tous ordres. En effet, pour tout $p > 0$, $x^p e^{-x^2/2}$ est intégrable sur $\mathbb{R}$. De plus, par parité $\mathbb{E}[X^{2p+1}] = 0$ et $\mathbb{E}[X^{2p}] = \frac{(2p)!}{2^k k!}$ (Cf TD).

Propriété 44 (Voir le cours d'intégration)**.**

1. *Inégalité de Cauchy-Schwartz : Soient $X, Y \in L^2$ alors*

   $$\mathbb{E}[|XY|] = ||XY||_1 \leqslant ||X||_2 ||Y||_2.$$

   Il y a égalité si et seulement si les variables X et Y sont liés, i.e $\exists a, b \in \mathbb{R}$ tels que $aX + bY$ est nul p.s

2. *Inégalité de Hölder : Soient $p, q \in [1, \infty]$ deux réels conjugués, i.e. tels que $\frac{1}{p} + \frac{1}{q} = 1$. Si X, Y deux variables sur $(\Omega, \mathcal{F}, \mathbb{P})$ avec $X \in L^p$ et $Y \in L^q$, alors $XY \in L^1$ avec*

   $$||XY||_1 \leqslant ||X||_p ||Y||_q.$$

3. *Inégalité de Minkowski : Soient $p \in [1, +\infty]$ et X, Y deux v.a. dans L^p, alors*

   $$||X + Y||_p \leqslant ||X||_p + ||Y||_p.$$

4. *Pour $p \in [1, +\infty]$, $(L^p, ||.||_p)$ est un espace vectoriel normé.*
5. *Théorème de Riesz-Fischer : $(L^p, \|.\|_p)$ est un espace vectoriel complet, i.e. espace de Banach et L^2 est un espace de Hilbert pour le produit scalaire $< X, Y >= \mathbb{E}[X\overline{Y}]$.*
6. *Si $1 \leqslant p \leqslant q \leqslant +\infty$, alors $L^\infty \subset L^q \subset L^p \subset L^1$. De plus $p \to \|.\|_p$ est croissante et $\lim_{p\to\infty} \|.\|_p = \|.\|_\infty$.*

5 Variance, Covariance

Definition 45. *Soit $X \in L^2(\Omega, \mathcal{F}, \mathbb{P})$ à valeurs réelles. On définit la variance de X par*

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

et l'écart-type par $\sigma_X = \sqrt{Var(X)}$.

L'espérance correspond à la valeur moyenne d'une variable aléatoire. L'écart-type correspond à la distance moyenne entre la variable et sa moyenne. C'est une mesure de dispersion. Par exemple, les mesures $\mathbb{P}_1 = \frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_1$ et $\mathbb{P}_2 = \frac{1}{2}\delta_{-1000} + \frac{1}{2}\delta_{1000}$ ont la même moyenne égale à 0, mais la dispersion de la seconde est bien plus grande que celle de la première.

Proposition 46.
1. *$Var(X) \geqslant 0$,*
2. *en développant le carrée, on a $Var(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$ (Formule de Koenig-Huygens),*
3. *$Var(aX) = a^2 Var(X)$,*
4. *$Var(X + b) = Var(X)$,*
5. *$Var(X) = 0$ si et seulement si X est constant p.s.*

Dans le cas d'une loi discrète, $\mathbb{P}_X = \sum_{i \in I} p_i \delta_{x_i}$ alors

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

et dans le cas d'une loi à densité, $\mathbb{P}_X = f(x)dx$,

$$Var(X) = \int_{\mathbb{R}} (x - \mathbb{E}[X])^2 f(x)dx = \int_{\mathbb{R}} x^2 f(x)dx - \mathbb{E}[X]^2.$$

Exemple.
1. Si $X \sim \mathcal{B}(p)$, $Var(X) = p(1-p)$,
2. Si $X \sim \mathcal{P}(\lambda)$, $Var(X) = \lambda$,
3. Si $X \sim \mathcal{E}(\alpha)$, $Var(X) = \frac{1}{\alpha^2}$,
4. Si $X \sim \mathcal{N}(0,1)$, alors $Var(X) = \mathbb{E}[X^2] = 2\int_0^\infty x^2 e^{-x^2/2} dx / \sqrt{2\pi}$ et par I.P.P., $Var(X) = 1$.
5. Si $X \sim \mathcal{N}(m, \sigma^2)$, alors $X = m + \sigma Z$ avec $Z \sim \mathcal{N}(0,1)$, par conséquent $Var(X) = \sigma^2$.

Definition 47. *La covariance entre deux variables X et Y dans L^2 est définie par*

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

On remarque que la covariance est une application bilinéaire, $Cov(X,X) = Var(X)$, $Cov(X,a) = 0$. Par ailleurs, en développant le carré de la variance d'une somme, on a

Propriété 48. *Soient $X, Y \in L^2$, alors*

$$Var(X + Y) = Var(X) + Var(Y) + 2Cov(X,Y),$$

et en utilisant l'inégalité de Cauchy-Schwartz,

$$|Cov(X,Y)| \leqslant \sqrt{Var(X)Var(Y)}.$$

Definition 49. *Soient X, Y deux variables dans L^2 non constantes, alors on appelle coefficient de corrélation la quantité*

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

On remarque facilement que $\rho(X,Y) \in [-1,1]$ et $|\rho(X,Y)| = 1$ si et seulement si X et Y sont liés, i.e. il existe $a, b \in \mathbb{R}$ avec $(a,b) \neq (0,0)$ tels que $X = aY + b$ (cas d'égalité dans Cauchy-Schwarz). En fait, $a = Cov(X,Y)/Var(Y)$ et $b = \mathbb{E}[X] - a\mathbb{E}[Y]$.
Par ailleurs ρ ne dépend pas de l'unité de mesure des variables X et Y.
Les variables sont dites non corrélées si $\rho(X,Y) = 0$.

Propriété 50 (Inégalité de Tchebychev)**.** *Soit X une variable dans L^2, alors pour tout $t > 0$*

$$\mathbb{P}(|X - \mathbb{E}[X]| > t) \leqslant \frac{Var(X)}{t^2}.$$

Démonstration. Utiliser l'inégalité de Markov. □

Exemple. Combien de fois faut-il lancer une pièce équilibrée pour que la moyenne du nombre de pile soit dans $[1/2 - 0.1, 1/2 + 0.1]$ avec probabilité supérieure à 0.96 ?
Soit S_n le nombre de pile obtenu sur n lancers. On cherche n tel que

$$\mathbb{P}\left(\frac{S_n}{n} \in [1/2 - 0.1, 1/2 + 0.1]\right) \geqslant 0.96.$$

On remarque que S_n suit la loi $\mathcal{B}(n, 1/2)$. La variance d'une loi Binomiale $X \sim \mathcal{B}(n,p)$ est $Var(X) = np$. Par ailleurs, $\mathbb{P}\big(\frac{S_n}{n} \in [1/2 - 0.1, 1/2 + 0.1]\big) = 1 - \mathbb{P}(|S_n/n - 1/2| > 0.1)$. Donc par l'inégalité de Tchebychev,

$$\mathbb{P}(|S_n/n - 1/2| > 0.1) \leqslant \frac{1}{4n(0.1)^2}$$

Par conséquent pour n tel que $n > \frac{1}{0.16 \times 0.01}$, soit $n \geqslant 625$.
(Vérifier sur un tableur que n est très surestimé avec Tchebychev.)

Definition 51. *Soit $X = (X_1, \ldots, X_d)^t$ un vecteur aléatoire colonne) dans L^2. On appelle matrice de covariance la matrice $d \times d$*

$$\Sigma_X = (Cov(X_i, X_j))_{1 \leqslant i,j \leqslant d}.$$

On remarque que $\Sigma_X = \mathbb{E}[XX^t] - \mathbb{E}[X]\mathbb{E}[X]^t = \mathbb{E}\big[(X - \mathbb{E}[X])(X - \mathbb{E}[X])^t\big]$
La matrice Σ_X est positive, symétrique et

$$a^t \Sigma_X a = \mathbb{E}\big[a^t (X - \mathbb{E}[X])(X - \mathbb{E}[X])^t a\big] = \mathbb{E}\Big[\big(a^t (X - \mathbb{E}[X])\big)^2\Big]$$

où a^t est le vecteur transposé. La matrice Σ_X est définie si aucune combinaison linéaire des marginales est p.s. constante.

6 Fonction caractéristique

Definition 52. *Soit $X : (\Omega, \mathcal{F}, \mathbb{P}) \to (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ une variable aléatoire. On appelle fonction caractéristique $\varphi_X : \mathbb{R}^d \to \mathbb{C}$ la fonction définie pour $t = (t_1, \dots, t_d) \in \mathbb{R}^d$*

$$\varphi_X(t) = \mathbb{E}\big[e^{i<t,X>}\big] = \mathbb{E}\Big[e^{i\sum_{k=1}^d t_k X_k}\Big],$$

où $< .,. >$ est le produit scalaire euclidien sur $\mathbb{R}^d$.

On remarque que φ_X est bien définie pour tout $t \in \mathbb{R}^d$ car $|e^{i<t,X>}| = 1$ donc intégrable.
Par ailleurs, si X est une v.a. réelle à densité, on retrouve la transformée de Fourier vue dans votre cours d'analyse,

$$\mathbb{E}[e^{itX}] = \int_{\mathbb{R}} e^{itx} f_X(x)dx = \mathcal{F}(f)\left(\frac{t}{2\pi}\right).$$

Theorem 53. *La fonction caractéristique caractérise la loi, i.e. si X, Y deux v.a. alors $\varphi_X = \varphi_Y$ si et seulement si $\mathbb{P}_X = \mathbb{P}_Y$.*

Démonstration. φ_X est la transformée de Fourier de la loi $\mathbb{P}_X$, le théorème d'inversion de Fourier s'applique car $|\varphi_X| \leqslant 1$ borné donc intégrable. Par conséquent, $\mathbb{P}_X = \mathcal{F}^{-1}(\varphi_X)$.
Preuve en dimension 1. Supposons que $\varphi_X = \varphi_Y$.
Soit $h \in \mathcal{C}_c(\mathbb{R}, \mathbb{R})$ continue à support compact. On remarque, en utilisant une approximation de l'unité, que par convergence dominée

$$h(x) = \frac{1}{2\pi} \lim_{\lambda \searrow 0} \int_{\mathbb{R}} h(x - \lambda t) \frac{2}{1+t^2} dt.$$

En effet,

$$\left| h(x-\lambda t)\frac{2}{1+t^2} \right| \leqslant \|h\|_\infty \frac{2}{1+t^2} \in L^1(\mathbb{R}).$$

Exprimons maintenant $\mathbb{E}[h(X)]$ en terme de φ_X :

$$\begin{aligned}
\mathbb{E}[h(X)] &= \int_{\mathbb{R}} h(x)\mathbb{P}_X(dx) \\
&= \int_{\mathbb{R}} \left(\frac{1}{2\pi} \lim_{\lambda \searrow 0} \int_{\mathbb{R}} h(x-\lambda t)\frac{2}{1+t^2}dt \right) \mathbb{P}_X(dx) \\
&= \frac{1}{2\pi} \lim_{\lambda \searrow 0} \int_{\mathbb{R}} \left(\int_{\mathbb{R}} h(x-\lambda t)\frac{2}{1+t^2}dt \right) \mathbb{P}_X(dx) \text{ par CV dominée par } ||h||_\infty \frac{2}{1+t^2} \in L^1(dt \otimes \mathbb{P}_X) \\
&= \frac{1}{2\pi} \lim_{\lambda \searrow 0} \int_{\mathbb{R}} \left(\int_{\mathbb{R}} h(s)\frac{2\lambda}{\lambda^2 + (x-s)^2}ds \right) \mathbb{P}_X(dx) \text{ par chgt de variable } s = x - \lambda t \\
&= \frac{1}{2\pi} \lim_{\lambda \searrow 0} \int_{\mathbb{R}} \left(\int_{\mathbb{R}} \frac{2\lambda}{\lambda^2 + (x-s)^2} \mathbb{P}_X(dx) \right) h(s)ds \text{ par Fubini.}
\end{aligned}$$

Par ailleurs, en décomposant en éléments simples $2\lambda / \big(\lambda^2 + (x-s)^2\big)$, on a

$$\begin{aligned}
\frac{2\lambda}{\lambda^2 + (x-s)^2} &= \frac{1}{\lambda - i(x-s)} + \frac{1}{\lambda + i(x-s)} \\
&= \int_0^\infty e^{-(\lambda - i(x-s))t}dt + \int_0^\infty e^{-(\lambda + i(x-s))t}dt \\
&= \int_0^\infty e^{i(x-s)t}e^{-\lambda t}dt + \int_0^\infty e^{-i(x-s)t}e^{-\lambda t}dt \\
&= \int_{\mathbb{R}} e^{i(x-s)t}e^{-\lambda|t|}dt.
\end{aligned}$$

Par conséquent, par Fubini (voir justification plus bas),

$$\begin{aligned}
\mathbb{E}[h(X)] &= \frac{1}{2\pi}\lim_{\lambda\searrow 0}\int_{\mathbb{R}}\left(\int_{\mathbb{R}}\left(\int_{\mathbb{R}} e^{i(x-s)t}e^{-\lambda|t|}dt\right)\mathbb{P}_X(dx)\right)h(s)ds \\
&= \frac{1}{2\pi}\lim_{\lambda\searrow 0}\int_{\mathbb{R}}\int_{\mathbb{R}}\left(\int_{\mathbb{R}} e^{ixt}\mathbb{P}_X(dx)\right)e^{-ist}e^{-\lambda|t|}dt\, h(s)ds \\
\mathbb{E}[h(X)] &= \frac{1}{2\pi}\lim_{\lambda\searrow 0}\int_{\mathbb{R}} h(s)\int_{\mathbb{R}}\varphi_X(t)e^{-\lambda|t|}e^{-ist}dtds.
\end{aligned}$$

Par conséquent, $\mathbb{E}[h(X)]$ ne dépend que de h et de φ_X. Si $\varphi_X = \varphi_Y$, on en déduit que $\mathbb{E}[h(X)] = \mathbb{E}[h(Y)]$ pour tout $h \in \mathcal{C}_c(\mathbb{R},\mathbb{R})$ et donc $\mathbb{P}_X = \mathbb{P}_Y$.
Justification de Fubini,

$$\int_{\mathbb{R}^3}\left|e^{i(x-s)t}e^{-\lambda|t|}h(s)\right|\mathbb{P}_X(dx)dtds = \int_{\mathbb{R}} e^{-\lambda|t|}dt\int_{\mathbb{R}}|h(s)|ds < \infty$$

car h est à support compact.
La réciproque est évidente. □

Proposition 54. *Soit X un vecteur aléatoire de $\mathbb{R}^d$ alors sa fonction caractéristique vérifie*

1. $|\varphi_X(t)| \leqslant 1$,
2. $\varphi_X(-t) = \overline{\varphi_X(t)}$,
3. $\varphi_X(0) = 1$,
4. φ_X *est uniformément continue.*

Démonstration. Soient $t, h \in \mathbb{R}^d$, on a

$$\begin{aligned}
|\varphi_X(t+h) - \varphi_X(t)| &= \left|\mathbb{E}\left[e^{i<t+h/2,X>}2i\sin(<h/2,X>)\right]\right| \\
&\leqslant 2\mathbb{E}[|\sin(<h/2,X>)|] \\
&\leqslant 2\mathbb{E}[|<h/2,X>|\wedge 1].
\end{aligned}$$

Par convergence dominée $\mathbb{E}[|<h/2,X>|\wedge 1] \to 0$ quand $|h| \to 0$ de façon indépendante de t. Donc φ_X est uniformément continue. □

Exemple.

Si X est discrète à valeurs dans $\mathcal{A} = \{x_k, k \in K\}$, alors, en notant $p_k = \mathbb{P}(X = x_k)$,

$$\varphi_X(t) = \sum_{k\in K} e^{itx_k}p_k.$$

Si $X \sim \mathcal{B}(n,p)$, on a alors, par la formule du binôme de Newton,

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

Si X à densité alors

$$\varphi_X(t) = \int_{\mathbb{R}} e^{itx}f_X(x)dx.$$

Par exemple, si $X \sim \mathcal{N}(0,1)$ alors

$$\varphi_X(t) = \int e^{\frac{-x^2}{2}} e^{ixt} \frac{dx}{\sqrt{2\pi}} = e^{\frac{-t^2}{2}} \int e^{\frac{-(x-it)^2}{2}} \frac{dx}{\sqrt{2\pi}}.$$

On voit apparaître l'intégrale de la densité de $Y = Z + it$ où $Z \sim \mathcal{N}(0,1)$ et donc $\varphi_X(t) = e^{-\frac{t^2}{2}}$.
On en déduit que si $X \sim \mathcal{N}(m, \sigma^2)$,

$$\varphi_X(t) = e^{itm} e^{-\frac{t^2\sigma^2}{2}}.$$

car X peut s'écrire sous la forme $X = \sigma Z + m$ avec $Z \sim \mathcal{N}(0,1)$.
Si $X \sim \mathcal{E}(\alpha)$ avec $\alpha > 0$, alors

$$\begin{aligned}\varphi_X(t) &= \int_0^\infty \alpha e^{-\alpha x} e^{ixt} dx \\ &= \alpha \left[\frac{-1}{\alpha - it} e^{-(\alpha - it)x} \right]_0^\infty \\ &= \frac{\alpha}{\alpha - it}.\end{aligned}$$

Theorem 55 (Inversion de Fourier)**.** *Soit X une v.a. de fonction caractéristique φ_X. Si φ_X est intégrable sur $\mathbb{R}^d$ par rapport à la mesure de Lebesgue, alors X admet une densité continue bornée donnée par*

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

Démonstration. Preuve en dimension $d = 1$. Soit $h \in \mathcal{C}_c(\mathbb{R})$. Alors

$$\begin{aligned}\mathbb{E}[h(X)] &= \frac{1}{2\pi} \lim_{\lambda \searrow 0} \int_{\mathbb{R}} h(s) \int_{\mathbb{R}} \varphi_X(t) e^{-\lambda|t|} e^{-ist} dt ds \\ &= \frac{1}{2\pi} \int_{\mathbb{R}} h(s) \left(\int_{\mathbb{R}} \varphi_X(t) e^{-ist} dt \right) ds,\end{aligned}$$

par convergence dominée et Fubini. En effet,

$$\left| h(s) \varphi_X(t) e^{-\lambda|t|} e^{-ist} \right| \leqslant |h(s)||\varphi_X(t)| \in L^1(ds \otimes dt).$$

□

Exemple. Si X de loi de Cauchy de paramètre $a > 0$, on a

$$\varphi_X(t) = \frac{1}{\pi} \int_{\mathbb{R}} \frac{a}{a^2 + x^2} e^{ixt} dx.$$

En décomposant $a/(a^2 + x^2)$ en éléments simples on a

$$\frac{1}{\pi} \frac{a}{a^2 + x^2} = \frac{1}{2\pi} \left(\frac{1}{a + ix} + \frac{1}{a - ix} \right) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-a|s|} e^{-ixs} ds \in L^1(dx).$$

Par conséquent, par le théorème d'inversion de Fourier $\varphi_X(t) = e^{-a|t|}$. On remarque que cette fonction est continue en 0, mais n'est pas dérivable.

Propriété 56. *Soit X une v.a. réelle de fonction caractéristique φ_X. Si X admet un moment d'ordre $p \in \mathbb{N}^*$ alors φ_X est de classe $\mathcal{C}^p$ et pour tout $k \leqslant p$,*

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

Réciproquement, si φ_X est p fois dérivable en 0, alors X admet des moments d'ordres inférieur ou égal à $2[p/2]$.

La proposition est encore vraie en dimension quelconque.

Démonstration. Supposons que $X \in L^p$. Soit $\omega \in \Omega$, alors

$h : t \mapsto e^{itX(\omega)}$ est de classe $\mathcal{C}^\infty$,

$h^{(k)}(t) = (iX(\omega))^k e^{itX(\omega)}$,

$\left|h^{(k)}(t)\right| \leqslant |X(\omega)|^k$.

Par ailleurs, pour tout $k \leqslant p$, $|X|^k \in L^1$. On conclut par récurrence en utilisant le théorème de dérivation d'une intégrale dépendant d'un paramètre.
Supposons φ_X p fois dérivable en 0. Montrons par récurrence que $X \in L^{2k}$ avec $k \leqslant [p/2]$. Pour $k = 0$, c'est évident.
Supposons que pour un $k < [p/2]$, on a $X \in L^{2k}$ et donc d'après la première partie du théorème, on a $\varphi_X^{(2k)}(t) = (-1)^p \mathbb{E}[X^{2k} e^{itX}]$.
Comme $2(k+1) = 2k+2 \leqslant 2[p/2] \leqslant p$, on a $\varphi_X \in \mathcal{C}^{2k+2}$, d'après la formule de Taylor à l'ordre 2,

$$\begin{aligned}
\varphi_X^{(2k+2)}(0) &= \lim_{h\to 0} \frac{\varphi_X^{(2k)}(h) - 2\varphi_X^{(2k)}(0) + \varphi_X^{(2k)}(-h)}{h^2} \\
&= (-1)^k \lim_{h\to 0} \mathbb{E}\left[\frac{e^{ihX} - 2 + e^{-ihX}}{h^2} X^{2k}\right] \\
&= (-1)^k \lim_{h\to 0} \mathbb{E}\left[\left(\frac{e^{ihX/2} - e^{-ihX/2}}{h}\right)^2 X^{2k}\right] \\
&= (-1)^{k+1} \lim_{h\to 0} \mathbb{E}\left[\left(\frac{\sin(hX/2)}{hX/2}\right)^2 X^{2k+2}\right]
\end{aligned}$$

Par le lemme de Fatou, on a

$$\begin{aligned}
\mathbb{E}[X^{2k+2}] &= \mathbb{E}\left[\lim_{h\to 0}\left(\frac{\sin(hX/2)}{hX/2}\right)^2 X^{2k+2}\right] \\
&\leqslant \liminf_{h\to 0} \mathbb{E}\left[\left(\frac{\sin(hX/2)}{hX/2}\right)^2 X^{2k+2}\right] = (-1)^{k+1}\varphi_X^{(2k+2)}(0) < \infty.
\end{aligned}$$

On en déduit que $X \in L^{2(k+1)}$. On conclut par récurrence. $\square$

Par conséquent, si φ_X est suffisamment régulière, on a

$$\mathbb{E}[X] = -i\varphi'_X(0),\ \mathbb{E}[X^2] = -\varphi_X^{(2)}(0),\ \mathbb{E}[X^3] = i\varphi_X^{(3)}(0),\ \ldots$$

Exemple. La loi Binomiale $X \sim \mathcal{B}(n,p)$ admet des moments de tous ordres car à support fini. Par ailleurs, on a

$$\varphi_X(t) = \left(pe^{it} + (1-p)\right)^n$$

donc en dérivant successivement, on obtient

$$\mathbb{E}[X] = np, \quad \mathbb{E}[X]^2 = np + n(n-1)p^2 \quad \Rightarrow \quad Var(X) = np(1-p).$$

7 Autres fonctionnelles caractérisant la loi d'une v.a.

7.1 Fonction génératrice pour les v.a. entières

Soit X une v.a. à valeurs dans $\mathbb{N}$. On note $p_k = \mathbb{P}(X = k)$.
On appelle fonction génératrice des moments la fonction définie et continue sur $[0,1]$ par

$$M_X(t) = \mathbb{E}[t^X] = \sum_{p \geqslant 0} p_k t^k.$$

Comme le rayon de convergence de la série entière M_X est supérieur à 1, la fonction est $\mathcal{C}^\infty$ sur $[0,1[$. De plus

$$M_X^{(n)}(0) = k! p_k.$$

En utilisant le théorème de dérivation sous le signe intégrale, on remarque que M_X est $n-$fois dérivable en 1 si et seulement si $\mathbb{E}[X^n] < \infty$ et alors

$$M_X^{(n)}(1) = \mathbb{E}[X(X-1)..(X-n+1)].$$

Par exemple,
- si $X \sim \mathcal{B}(p)$ alors $M_X(t) = (1-p) + pt$,
- si $X \sim \mathcal{B}(n,p)$ alors $M_X(t) = ((1-p) + pt)^n$,
- si $X \sim \mathcal{G}(p)$ alors $M_X(t) = \frac{pt}{1-(1-p)t}$.

7.2 Transformée de Laplace

Soit X une v.a. à valeurs dans $\mathbb{R}^d$. On appelle transformée de Laplace de X la fonction définie par

$$L_X(t) = \mathbb{E}[e^{<t,X>}]$$

pour les valeurs de $t \in \mathbb{R}^d$ tel que $e^{<t,X>}$ soit intégrable.

Proposition 57. *1.* $L_X(0) = 1$.

2. Si L_X est définie sur un voisinage de 0, alors L_X caractérise la loi de X.

3. Soit X une v.a. réelle telle que L_X est définie sur un voisinage de 0, alors L_X est analytique sur ce voisinage et on a

$$L_X(t) = \sum_{n \geqslant 0} \frac{t^n}{n!} \mathbb{E}[X^n] \quad et \quad \mathbb{E}[X^n] = L_X^{(n)}(0).$$

Ceci implique notamment que $X \in L^p$ pour tout $p \geqslant 1$.

4. Si X est positive, alors $L_X(t) < \infty$ pour $t \leqslant 0$.

Exemple. Si $X \sim \mathcal{E}(\alpha)$, avec $\alpha > 0$ alors la transformée de Laplace est définie sur $]-\infty, \alpha[$ et pour $t < \alpha$,

$$L_X(t) = \int_0^\infty \alpha e^{-(\alpha - t)x} dx = \frac{\alpha}{\alpha - t}.$$

8 Lois classiques

• Lois discrètes

Loi	**Probabilité** $\mathbb{P}(X=k)$	**Moments**	$\varphi_X(t)$
Bernoulli $\mathcal{B}(p)$	$\mathbb{P}(X=1)=p$ $\mathbb{P}(X=0)=1-p$	$\mathbb{E}[X]=p$, $Var(X)=p(1-p)$	$1-p+pe^{it}$
Binomiale $\mathcal{B}(n,p)$	$\binom{n}{k}p^k\,(1-p)^{n-k}$, $0\leqslant k\leqslant n$	$\mathbb{E}[X]=np$, $Var(X)=np(1-p)$	$\left(1-p+pe^{it}\right)^n$
Géométrique $\mathcal{G}(p)$	$p\,(1-p)^{k-1}$, $k\geqslant 1$	$\mathbb{E}[X]=\frac{1}{p}$, $Var(X)=\frac{1-p}{p^2}$	$\frac{pe^{it}}{1-(1-p)e^{it}}$
Poisson $\mathcal{P}(\lambda)$	$\frac{\lambda^k}{k!}e^{-\lambda}$, $k\geqslant 0$	$\mathbb{E}[X]=\lambda$, $Var(X)=\lambda$	$e^{\lambda(e^{it}-1)}$

• Lois à densité

Loi	**Densité**	**Moments**	$\varphi_X(t)$
Uniforme $\mathcal{U}([a,b])$	$\dfrac{1}{b-a}\mathbb{1}_{x\in[a,b]}$	$\mathbb{E}[X]=\frac{a+b}{2}$, $Var(X)=\frac{(b-a)^2}{12}$	$\frac{e^{ibt}-e^{iat}}{it(b-a)}$
Exponentielle $\mathcal{E}(\lambda)$, $\lambda>0$	$\lambda e^{-\lambda x}\mathbb{1}_{x>0}$	$\mathbb{E}[X]=\dfrac{1}{\lambda}$, $Var(X)=\dfrac{1}{\lambda^2}$	$\frac{\lambda}{\lambda-it}$
Normale $\mathcal{N}\left(\mu,\sigma^2\right)$, $\mu\in\mathbb{R}$, $\sigma>0$	$\frac{1}{\sigma\sqrt{2\pi}}e^{\frac{-(x-\mu)^2}{2\sigma^2}}$	$\mathbb{E}[X]=\mu$, $Var(X)=\sigma^2$	$e^{i\mu t}e^{\frac{-\sigma^2t^2}{2}}$
Gamma $\Gamma\,(a,\lambda)$, $a,\lambda>0$	$\frac{\lambda^a}{\Gamma(a)}x^{a-1}e^{-\lambda x}\mathbb{1}_{x>0}$	$\mathbb{E}[X]=\dfrac{a}{\lambda}$, $Var(X)=\dfrac{a}{\lambda^2}$	$\left(\frac{\lambda}{\lambda-it}\right)^a$
Cauchy $\mathcal{C}(a)$, $a>0$	$\frac{a}{\pi(x^2+a^2)}$	non définis	$e^{-a\|t\|}$

Chapitre 4

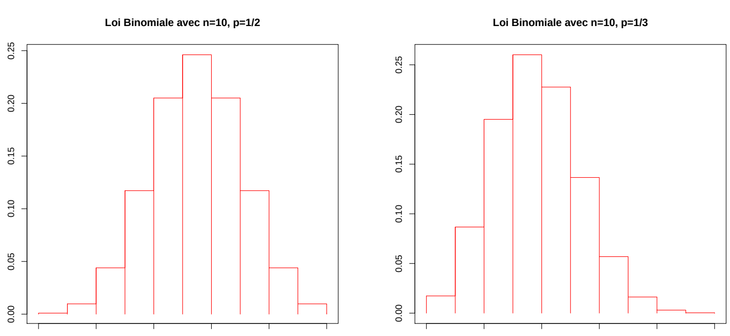
Lois usuelles

1 Lois Discrètes

Loi Uniforme sur un ensemble fini : la loi uniforme (équirépartie) sur l'ensemble à n éléments $\{x_1, \ldots, x_n\}$ est telle que tous les éléments sont équiprobables et donc définie par $\mathbb{P}(\{x_i\}) = 1/n$ pour tout $i \in \{1, \ldots, n\}$.

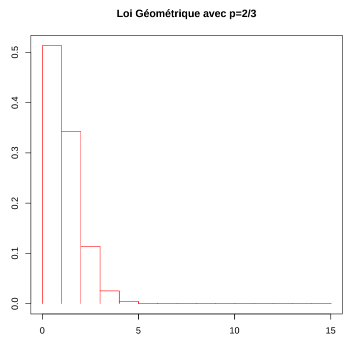
Loi de Bernoulli, $\mathcal{B}(p)$, avec $p \in]0,1[$. La loi de Bernoulli modélise une expérience aléatoire qui n'a que deux résultats possibles : succès et échec, auquel on attribue en général les valeurs 1 et 0. Une variable aléatoire X ne prend que deux valeurs.
Loi : $\mathbb{P}(X = 1) = p \quad \mathbb{P}(X = 0) = 1 - p$.
Espérance : p, Variance : $p(1-p)$.
Cette expérience est appelée **épreuve de Bernoulli** de paramètre p.

Loi Binomiale, $\mathcal{B}(n,p)$, avec $p \in]0,1[$, $n \in \mathbb{N}$, $n \geqslant 1$. On renouvelle n fois de manières indépendantes une épreuve de Bernoulli de paramètre p. La loi du nombre X de succès obtenus à l'issue des n épreuves s'appelle loi Binomiale de paramètres n et p. Ce nombre est à valeurs dans $\{0, 1, \ldots, n\}$. Cette loi est notamment utilisée pour des sondages **avec remise**.
Loi : $\mathbb{P}(X = k) = \begin{pmatrix} n \\ k \end{pmatrix} p^k (1-p)^{n-k}$ avec $k \in \{0, 1, .., n\}$.
Espérance : np, Variance : $np(1-p)$.

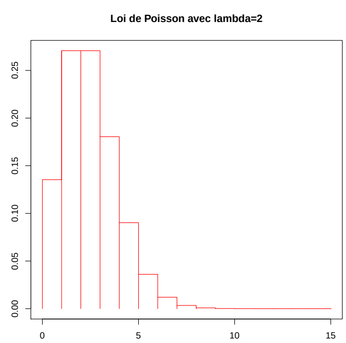


Loi Géométrique, $\mathcal{G}(p)$, avec $p \in]0,1[$. La loi géométrique est la loi du premier succès : on note X le nombre de réalisations (indépendantes) nécessaires d'une épreuve de Bernoulli jusqu'à obtenir un succès.

Loi : $\mathbb{P}(X = k) = p(1-p)^{k-1}$ avec $k \geqslant 1$.
Espérance : $\dfrac{1}{p}$, Variance : $\dfrac{1-p}{p^2}$.
Fonction de répartition : $F(k) = 1 - (1-p)^k$ avec $k \geqslant 1$.



Loi de Poisson, $\mathcal{P}(\lambda)$, avec $\lambda > 0$. Cette loi est souvent utilisée pour dénombrer des événements "rares", comme des pannes, des sinistres, des clients entrant dans une boutique pendant un intervalle de temps donné.
Loi : $\mathbb{P}(X = k) = \dfrac{\lambda^k}{k!} e^{-\lambda}$ avec $k \in \mathbb{N}$.
Espérance : λ, Variance : λ



Loi Binomiale Négative, avec $r \in \mathbb{N}$, $r \geqslant 1$, $p \in]0,1[$. Il s'agit d'une généralisation de la loi géométrique. On réalise plusieurs fois une épreuve de Bernoulli de façon indépendante jusqu'à obtenir r succès exactement. On note X le nombre de réalisations nécessaires.
Loi : $\mathbb{P}(X = k) = \begin{pmatrix} k-1 \\ r-1 \end{pmatrix} p^r (1-p)^{k-r}$ si $k \geqslant r$.

Espérance : $\dfrac{r}{p}$, Variance : $\dfrac{r(1-p)}{p^2}$.

Loi Hypergéométrique $H(N, m, n)$, avec $N \geqslant 1$, $(m, n) \in \{1, .., N\}^2$. On considère N éléments dont m sont marqués. On tire au hasard et **sans remise** un échantillon de n éléments et on compte le nombre X d'éléments marqués sur notre échantillon.

Loi : $\mathbb{P}(X=k)=\dfrac{\begin{pmatrix} m \\ k \end{pmatrix}\begin{pmatrix} N-m \\ n-k \end{pmatrix}}{\begin{pmatrix} N \\ n \end{pmatrix}}$ si $k\in\{0,..,\min(m,n)\}$.

Espérance : np, Variance : $\dfrac{N-n}{N-1}np(1-p)$ avec $p=\frac{m}{N}$.

Loi Multinomiale $\mathcal{M}_d(n,p)$, avec $n\in\mathbb{N}^*$, $n\geqslant 1$, $p\in]0,1[^d$ tel que $p_1+p_2+..+p_d=1$. Il s'agit d'une généralisation de la loi Binomiale. On considère une expérience dont d résultats sont possibles et la probabilité d'obtenir le résultat i est p_i. On réalise n fois, de façon indépendante, cette expérience et on dénombre le nombre X_1 de fois où on a obtenu 1, X_2 le nombre de fois où on a obtenu 2, ..., et X_d le nombre de fois où on a obtenu d parmi les n expériences. Par exemple, on lance n fois un dé à 6 faces.

Loi : $\mathbb{P}(X_1=k_1,X_2=k_2,\ldots,X_d=k_d)=\dfrac{n!}{k_1!k_2!...k_d!}p_1^{k_1}p_2^{k_2}...p_d^{k_d}$ avec $\displaystyle\sum_{i=i}^{d}k_i=n$.

Espérance : $(np_1,\cdots,np_d)$.

2 Lois sur $\mathbb{R}$ à densité

Loi Uniforme, $\mathcal{U}([a,b])$, avec $a,b\in\mathbb{R}$, $a<b$.
Densité :

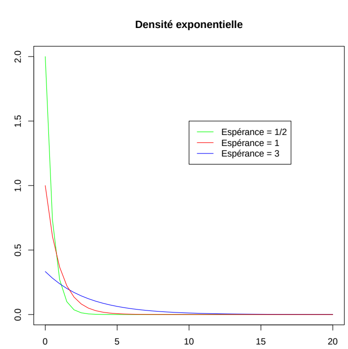
$$f(x)=\frac{1}{b-a}\mathbb{1}_{[a,b]}(x).$$

Espérance : $\dfrac{a+b}{2}$, Variance : $\dfrac{(b-a)^2}{12}$.
Fonction de répartition :

$$\begin{aligned} F(x) &= 0 && \text{si } x<a \\ &= \frac{x-a}{b-a} && \text{si } x\in[a,b] \\ &= 1 && \text{si } x>b \end{aligned}$$

Loi Exponentielle, $\mathcal{E}(\lambda)$, avec $\lambda>0$. Cette loi est en général utilisée pour modéliser des durées de vie sans vieillissement ou des temps d'attente.
Densité :

$$f(x)=\lambda e^{-\lambda x}\mathbb{1}_{x>0}.$$



Densité de la loi exponentielle

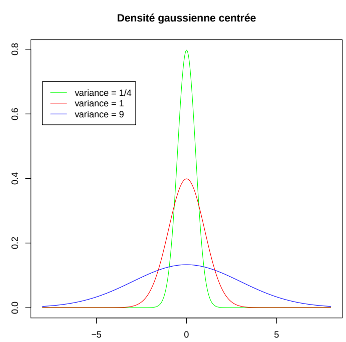
Espérance : $\frac{1}{\lambda}$, Variance : $\frac{1}{\lambda^2}$.
Fonction de répartition :

$$\begin{aligned} F(x) &= 0 & \text{si } x < 0 \\ &= 1 - e^{-\lambda x} & \text{si } x \geqslant 0 \end{aligned}$$

La loi exponentielle est la seule loi continue qui vérifie la propriété d'absence de mémoire : Si $X \sim \mathcal{E}(\lambda)$, alors pour tout $s, t > 0$ $P(X > t + s | X > t) = P(X > s)$.

Loi Normale (ou loi Gaussienne), $\mathcal{N}\left(m, \sigma^2\right)$, avec $m \in \mathbb{R}$, $\sigma > 0$. La loi Normale est une loi centrale dans la théorie des probabilités. Elle est notamment très utilisée en statistique.
Densité :

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}} \quad \text{avec } x \in \mathbb{R}.$$



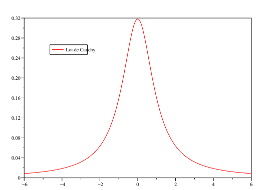
Densité de loi normale

Espérance : m, Variance : σ^2.

Loi de Cauchy
Densité :

$$f(x) = \frac{1}{\pi} \frac{1}{1 + x^2} \text{ avec } x \in \mathbb{R}.$$

Son espérance et sa variance ne sont pas définies (intégrales divergentes sur $\mathbb{R}$).



Densité de la loi de Cauchy

Loi Gamma, $G(a, \lambda)$, avec $a > 0$ et $\lambda > 0$.
Densité :

$$f(x) = \frac{\lambda^a}{\Gamma(a)} x^{a-1} e^{-\lambda x} \mathbb{1}_{x>0} \text{ avec } \Gamma(a) = \int_0^\infty t^{a-1} e^{-t} dt.$$

Espérance : $\dfrac{a}{\lambda}$, Variance : $\dfrac{a}{\lambda^2}$.
Si $a \in \mathbb{N}$, on a $\Gamma(a) = (a-1)!$

Loi Beta, $\mathcal{B}\,(a,b)$, avec $a > 0$ et $b > 0$.
Densité :

$$f\,(x) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1}\mathbb{1}_{x\in]0,1[}.$$

Espérance : $\dfrac{a}{a+b}$, Variance : $\dfrac{ab}{(a+b)^2(a+b+1)}$.

Loi de Weibull, $W\,(\alpha,\beta,\nu)$, avec $\alpha > 0$ (paramètre d'échelle), $\beta > 0$ (paramètre de forme), $\nu \in \mathbb{R}$ (paramètre de position). Cette loi permet en démographie de modéliser le vieillissement, en épidémiologie la durée d'incubation d'une maladie infectieuse. Pour $\beta = 1$ et $\nu = 0$ on retrouve la loi exponentielle, i.e. la loi des organismes qui ne sont pas soumis au vieillissement. Plus β est grand, plus le vieillissement se fait pesant (la mortalité augmente avec l'âge). Pour $\beta < 1$, plus on vieillit moins forte est la probabilité de mourir dans l'unité de temps qui vient (penser à la mortalité infantile).
Densité :

$$f(x) = \frac{\beta}{\alpha}\left(\frac{x-\nu}{\alpha}\right)^{\beta-1} e^{-\left(\frac{x-\nu}{\alpha}\right)^{\beta}}\mathbb{1}_{x\geqslant\nu}$$

Fonction de répartition :

$$\begin{aligned} F\,(x) &= 0 && \text{si } x < \nu \\ &= 1 - e^{-\left(\frac{x-\nu}{\alpha}\right)^{\beta}} && \text{si } x \geqslant \nu \end{aligned}$$

Loi de Pareto avec $a > 0$ (paramètre de forme) et $b > 0$ (paramètre de position). Cette loi est notamment utilisée en assurance pour modéliser les très gros sinistres.
Densité :

$$f(x) = a\frac{b^a}{x^{a+1}}\mathbb{1}_{x\geqslant b}.$$

Espérance : $\dfrac{ab}{a-1}$ pour $a > 1$, Variance : $\dfrac{ab^2}{(a-1)^2(a-2)}$ pour $a > 2$.
Fonction de répartition :

$$\begin{aligned} F\,(x) &= 0 && \text{si } x < b \\ &= 1 - \left(\frac{b}{x}\right)^{a} && \text{si } x \geqslant b \end{aligned}$$

Loi du Chi-deux $\chi^2(n)$ avec $n \in \mathbb{N}^*$ (nombre de degré de liberté). Cette loi est utilisée en statistique pour tester des hypothèses. C'est la loi de la somme des carrés de n variables gaussiennes $\mathcal{N}(0,1)$ indépendantes. On remarque que $\chi^2(n) = G(\frac{n}{2},\frac{1}{2})$.
Densité :

$$f(x) = \frac{\left(\frac{1}{2}\right)^{n/2}}{\Gamma\left(\frac{n}{2}\right)} x^{\frac{n}{2}-1} e^{\frac{-x}{2}}\mathbb{1}_{x>0}.$$

Espérance : n, Variance : $2n$.

3 Fonctions caractéristiques

• Lois discrètes

Loi	**Probabilité** $\mathbb{P}(X=k)$	**Moments**	$\varphi_X(t)$
Bernoulli	$\mathbb{P}(X=1)=p$ $\mathbb{P}(X=0)=1-p$	$\mathbb{E}[X]=p$, $Var(X)=p(1-p)$	$1-p+pe^{it}$
Binomiale	$\binom{n}{k}p^k\,(1-p)^{n-k}$, $0\leqslant k\leqslant n$	$\mathbb{E}[X]=np$, $Var(X)=np(1-p)$	$\left(1-p+pe^{it}\right)^n$
Géométrique	$p\,(1-p)^{k-1}$, $k\geqslant 1$	$\mathbb{E}[X]=\frac{1}{p}$, $Var(X)=\frac{1-p}{p^2}$	$\frac{pe^{it}}{1-(1-p)e^{it}}$
Poisson	$\frac{\lambda^k}{k!}e^{-\lambda}$, $k\geqslant 0$	$\mathbb{E}[X]=\lambda$, $Var(X)=\lambda$	$e^{\lambda(e^{it}-1)}$

• Lois à densité

Loi	**Densité**	**Moments**	$\varphi_X(t)$
Uniforme $\mathcal{U}([a,b])$	$\dfrac{1}{b-a}\mathbb{1}_{x\in[a,b]}$	$\mathbb{E}[X]=\frac{a+b}{2}$, $Var(X)=\frac{(b-a)^2}{12}$	$\frac{e^{ibt}-e^{iat}}{it(b-a)}$
Exponentielle $\mathcal{E}(\lambda)$, $\lambda>0$	$\lambda e^{-\lambda x}\mathbb{1}_{x>0}$	$\mathbb{E}[X]=\dfrac{1}{\lambda}$, $Var(X)=\dfrac{1}{\lambda^2}$	$\frac{\lambda}{\lambda-it}$
Normale $\mathcal{N}\left(\mu,\sigma^2\right)$, $\mu\in\mathbb{R}$, $\sigma>0$	$\frac{1}{\sigma\sqrt{2\pi}}e^{\frac{-(x-\mu)^2}{2\sigma^2}}$	$\mathbb{E}[X]=\mu$, $Var(X)=\sigma^2$	$e^{i\mu t}e^{\frac{-\sigma^2t^2}{2}}$
Gamma $G\,(a,\lambda)$, $a,\lambda>0$	$\frac{\lambda^a}{\Gamma(a)}x^{a-1}e^{-\lambda x}\mathbb{1}_{x>0}$	$\mathbb{E}[X]=\dfrac{a}{\lambda}$, $Var(X)=\dfrac{a}{\lambda^2}$	$\left(\frac{\lambda}{\lambda-it}\right)^a$
Cauchy $\mathcal{C}(a)$, $a>0$	$\frac{a}{\pi(x^2+a^2)}$	non définis	$e^{-a\|t\|}$

Chapitre 5

Notion d'indépendance en probabilité

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

1 Probabilité conditionnelle

On jette deux dés distincts bien équilibrés. On a $(\Omega, \mathcal{F}, \mathbb{P})$ avec $\Omega = \{(1,1),(1,2),\dots,(5,6),(6,6)\}$, $\mathcal{F} = \mathcal{P}(\Omega)$ et $\mathbb{P}$ la loi uniforme sur Ω.
On veut connaitre la probabilité de l'événement A : "la somme des dés vaut 8". On remarque que $A = \{(2,6),(6,2),(3,5),(5,3),(4,4)\}$, d'où $\mathbb{P}(A) = 5/36$.
Maintenant on cherche toujours la même probabilité, mais on sait déjà que le premier dé donne un 3. On note B l'événement : "le premier dé vaut 3".
Que vaut alors la probabilité que la somme vaut 8 sachant que le premier dé vaut 3 ? La probabilité recherchée est appelée probabilité conditionnelle de A sachant B et est notée $\mathbb{P}(A|B)$.
Sachant que le premier dé vaut 3, on ne regarde plus la même expérience, on regarde finalement une expérience sur un sous-espace de Ω (c'est comme-ci on ne lançait plus qu'un dé). Il y a alors 6 résultats dans cette expérience : $\{(3,1),(3,2),(3,3),(3,4),(3,5),(3,6)\}$ et chacune a la même probabilité d'apparaître, soit $1/6$. Par conséquent la probabilité que la somme des deux dés soit égale à 8 sachant que le premier dé vaut 3 est $\mathbb{P}(A|B) = 1/6$ (il n'y a qu'une possibilité, le deuxième dé doit valoir 5).
Le fait d'avoir l'information B change la probabilité de A. On remarque par ailleurs que l'événement $A \cap B$ est l'événement "le premier dé donne 3 et la somme des chiffres vaut 8 lorsqu'on lance deux dés", on a alors $A \cap B = \{(3,5)\}$ dans l'ensemble $\Omega = \{(1,1),(1,2),\dots,(5,6),(6,6)\}$ de cardinal 36.
D'autre part, on a $B = \{(3,1),(3,2),(3,3),(3,4),(3,5),(3,6)\}$.
D'où $\mathbb{P}(A \cap B) = 1/36$ et $\mathbb{P}(B) = 6/36 = 1/6$. On remarque que

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

Definition 58. *Soient A et B deux événements avec $\mathbb{P}(B) \neq 0$. La* **probabilité conditionnelle** *$P(A|B)$ de A sachant B est définie par :*

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

Propriété 59. *Soit $B \in \mathcal{F}$ avec $\mathbb{P}(B) \neq 0$, alors l'application $A \in \mathcal{F} \mapsto \mathbb{P}(A|B)$ est une mesure de probabilité sur $(\Omega, \mathcal{F})$. Elle vérifie donc toutes les propriétés des probabilités.*

Démonstration. L'application est bien à valeurs dans $[0,1]$ car $\mathbb{P}(A\cap B)\leqslant \mathbb{P}(B)$. On a $\mathbb{P}(\Omega|B)=1$, et si $(A_n)_{n\geqslant 0}$ est une suite d'événements disjoints alors $(A_n\cap B)_{n\geqslant 0}$ sont disjoints et donc on en déduit que $\mathbb{P}(\cup A_n|B)=\sum \mathbb{P}(A_n|B)$. $\square$

Propriété 60 (Formule des probabilités totales)**.** *On considère deux événements A et B quelconques, avec $\mathbb{P}(B)\in]0,1[$.*
Alors

$$\mathbb{P}(A)=\mathbb{P}(A|B)\mathbb{P}(B)+\mathbb{P}(A|\overline{B})(1-\mathbb{P}(B))$$

Démonstration. En effet

$$\begin{aligned}\mathbb{P}(A)&=\mathbb{P}(A\cap B)+\mathbb{P}(A\cap \overline{B})\\&=\mathbb{P}(A|B)\mathbb{P}(B)+\mathbb{P}(A|\overline{B})\mathbb{P}(\overline{B}).\end{aligned}$$

$\square$

La formule de Bayes permet de calculer les probabilités a postériori d'un événement en fonction des probabilités a priori de cet événement, i.e. connaitre $\mathbb{P}(B|A)$ quand on connait $\mathbb{P}(A|B)$ et $\mathbb{P}(A|\overline{B})$.

Proposition 61 (Formule de Bayes)**.** *Soient A et B deux événements avec $\mathbb{P}(B)\in]0,1[$ et $\mathbb{P}(A)>0$.*

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

Démonstration. Il suffit d'écrire $\mathbb{P}(B|A)=\dfrac{\mathbb{P}(A\cap B)}{\mathbb{P}(A)}=\dfrac{\mathbb{P}(A|B)\mathbb{P}(B)}{\mathbb{P}(A)}$ et d'utiliser la formule des probabilités totales.

$\square$

$\mathbb{P}(A|\overline{B})\neq 1-\mathbb{P}(A|B)$. Par contre, on a bien $\mathbb{P}(\overline{A}|B)=1-\mathbb{P}(A|B)$.

Les formules des probabilités totales et de Bayes peuvent être généralisées de la façon suivante : soit A, $B_1,\dots,B_n$ des événements tels que les B_i soient disjoints et $\cup_{i=1}^n B_i=\Omega$ (partition de l'espace fondamental), alors

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

Exemple. Une maladie affecte 0.5% de la population. Un test T permet de dépister cette maladie avec la fiabilité suivante :

T est positif pour 95% des personnes affectées par la maladie
T est négatif pour 99% des personnes non affectées par la maldie.

Quelle la probabilité qu'un individu ayant un test positif soit affecté par la maladie ?

$$\mathbb{P}(M|P)=\frac{0.95\times 0.005}{0.95\times 0.005+0.01\times 0.995}=95/294\simeq 0.323.$$

En effet, sur 200 personnes, en moyenne 1 personne sera porteuse de la maladie et le test décèlera ce cas avec un proba de 0.95.Donc en moyenne sur 200 personnes, on détectera $0,95$ des cas. Par contre, sur les 200 personnes (dont 199 saines), le test détectera en moyenne $199.01=1.99$ faux positifs. La proportion de résultats correct est donc : $0.95/(0,9+1.99)=0.323$!

2 Notion d'indépendance

Definition 62 (Indépendance d'événements)**.** *Deux événements $A, B \in \mathcal{F}$ sont dits indépendants si*

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

Une famille d'événements $(A_i)_{i\in I}$ est dite (mutuellement) indépendante si $\forall i_1, \ldots, i_p \in I$

$$\mathbb{P}(A_{i_1} \cap \ldots \cap A_{i_p}) = \prod_{k=1}^{p} \mathbb{P}(A_{i_k}).$$

On remarque que si A et B sont deux événements, avec $\mathbb{P}(B) > 0$, alors A et B sont indépendants si et seulement si $\mathbb{P}(A|B) = \mathbb{P}(A)$.

Indépendance et disjoint sont deux notions bien distinctes.
Si A et B disjoints alors $A \cap B = \emptyset$ et donc $\mathbb{P}(A \cap B) = 0$. A et B ne peuvent alors être indépendants sauf si l'un des deux est négligeable.

L'indépendance mutuelle implique l'indépendance 2 à 2, mais la réciproque est fausse.

Exemple. Un joueur joue à Pile ou Face. Il gagne 1 euro si Pile et perd 1 euro si Face.
On considère les événements suivants

$$\begin{aligned} A &= \{\text{gagner 1 euro au premier lancer}\}, \\ B &= \{\text{gagner 1 euro au second lancer}\}, \\ C &= \{\text{gagner ou perdre 2 euros sur les deux premiers lancers}\}. \end{aligned}$$

On a $\mathbb{P}(A) = \mathbb{P}(B) = 1/2$. Les lancers étant indépendants, $\mathbb{P}(C) = 1/2$. Par ailleurs, $A \cap C = B \cap C = \{\text{gagner 2 euros sur les deux premiers lancers}\}$. Par conséquent, $\mathbb{P}(A \cap C) = 1/4 = \mathbb{P}(A)\mathbb{P}(C)$ et de même $\mathbb{P}(B \cap C) = \mathbb{P}(B)\mathbb{P}(C)$. Les événements A, B, C sont indépendants deux à deux.
Par ailleurs, $A \cap B \cap C = \{\text{gagner 2 euros sur les deux premiers lancers}\}$ qui est de probabilité $1/4 \neq (1/2)^3$. Les événements ne sont pas mutuellement indépendants.

Definition 63 (Indépendance de tribus)**.** *Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité. Deux familles $\mathcal{M}_1$ et $\mathcal{M}_2$ d'ensemble mesurables sont dites indépendantes si $\forall A \in \mathcal{M}_1$ et $\forall B \in \mathcal{M}_2$*

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

Une famille quelconque de tribus $(\mathcal{F}_i)_{i\in I}$ est dite indépendante si pour tout $A_i \in \mathcal{F}_i$, $i \in I$, la famille $(A_i)_{i\in I}$ est mutuellement indépendante.

Proposition 64. *Soient $\mathcal{A}_1$ et $\mathcal{A}_2$ deux sous-ensembles de $\mathcal{F}$ stables par intersection finie ($\prod -$systèmes) et indépendants, alors $\sigma(\mathcal{A}_1)$ et $\sigma(\mathcal{A}_2)$ sont deux tribus indépendantes.*

Démonstration. Soit $A_1 \in \mathcal{A}_1$. On note

$$\mathcal{M} = \{A_2 \in \sigma(\mathcal{A}_2) : \mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_1)\mathbb{P}(A_2)\}.$$

On a $\mathcal{A}_2 \subset \mathcal{M}$. Par ailleurs, $\mathcal{M}$ contient Ω, est stable par différence et union croissante, donc c'est une classe monotone. $\mathcal{M}$ contient la classe monotone engendrée par $\mathcal{A}_2$ qui est égale à $\sigma(\mathcal{A}_2)$ car $\mathcal{A}_2$ est stable par intersection finie (voir théorème des classes monotones). D'où $\mathcal{M} = \sigma(\mathcal{A}_2)$.
De même, pour $A_2 \in \sigma(\mathcal{A}_2)$, on définit

$$\mathcal{M}' = \{A_1 \in \sigma(\mathcal{A}_1) : \mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_1)\mathbb{P}(A_2)\}.$$

Par la même raisonnement, $\mathcal{M}'$ est une classe monotone contenant $\mathcal{A}_1$ et donc $\sigma(\mathcal{A}_1)$, d'où $\mathcal{M}' = \sigma(\mathcal{A}_1)$. On en déduit que $\sigma(\mathcal{A}_1)$ et $\sigma(\mathcal{A}_2)$ sont indépendantes (car A_2 était choisi de façon quelconque). $\square$

En généralisant, on a la proposition suivante,

Propriété 65. *Soit $(C_i)_{i\in I}$ une familles de $\prod$-systèmes, alors $(C_i)_{i\in I}$ sont indépendants si et seulement si $(\sigma(C_i))_{i\in I}$ sont indépendantes.*

Démonstration. S'il y a indépendance des tribus, il est évident qu'il y a indépendance de $(C_i)_{i\in I}$.
Réciproquement, on fixe $A_{i_k} \in \mathcal{C}_{i_k}$ pour $2 \leqslant k \leqslant p$ et on considère

$$\mathcal{M}_1 = \left\{ A_{i_1} \in \sigma(C_{i_1}) : \mathbb{P}(A_{i_1} \cap \ldots \cap A_{i_p}) = \prod_{k=1}^{p} \mathbb{P}(A_{i_k}) \right\}$$

Par classe monotone, $\mathcal{M}_1 = \sigma(C_{i_1})$.
Soit $A_{i_1} \in \sigma(C_{i_1})$ et on fixe $A_{i_k} \in \mathcal{C}_{i_k}$ pour $3 \leqslant k \leqslant p$. On considère alors la famille

$$\mathcal{M}_2 = \left\{ A_{i_2} \in \sigma(C_{i_1}) : \mathbb{P}(A_{i_1} \cap \ldots \cap A_{i_p}) = \prod_{k=1}^{p} \mathbb{P}(A_{i_k}) \right\}$$

Par classe monotone, $\mathcal{M}_2 = \sigma(C_{i_2})$. On continue par récurrence et on montre alors que

$$\mathbb{P}(A_{i_1} \cap \ldots \cap A_{i_p}) = \prod_{k=1}^{p} \mathbb{P}(A_{i_k})$$

pour tout $p \geqslant 1$ et $A_{i_k} \in \sigma(C_{i_k})$ pour $1 \leqslant k \leqslant p$. Cqfd. $\square$

Theorem 66. *Soit $(\mathcal{F}_i)_{i\in I}$ une famille de sous-tribus indépendantes de $\mathcal{F}$. Soit $I = \bigcup_{a\in A} I_a$ une partition de I ($I_a \cap I_b = \emptyset$ pour $a \neq b$) et on note $\mathcal{G}_a = \sigma(\cup_{i\in I_a}\mathcal{F}_i)$, alors $(\mathcal{G}_a)_{a\in A}$ sont des tribus indépendantes.*

Démonstration. La tribu $\mathcal{F}_i$ contient les informations portées par les événements constituant la tribu.
On note

$$\mathcal{C}_a = \{B \in \mathcal{F} : \exists J \text{ fini } \subset I_a : \exists A_j \in \mathcal{F}_j,\ j \in J,\ B = \cap_{j\in J} A_j\}$$

Pour $i \in I_a$, on a $\mathcal{F}_i \subset \mathcal{C}_a$, par conséquent $\cup_{i\in I_a}\mathcal{F}_i \subset \mathcal{C}_a \subset \mathcal{G}_a$. Par ailleurs, $\mathcal{C}_a$ est stable par intersection finie. En effet, si $B, B' \in \mathcal{C}_a$, alors on peut écrire

$$B \cap B' = \bigcap_{j\in J} A_j \cap \bigcap_{j\in J'} A'_j = \bigcap_{j\in J\cup J'} A''_j$$

avec $A''_j = A_j$ ou A'_j selon que $j \in J$ ou $j \in J'$. Par conséquent, $\sigma(\mathcal{C}_a) = \mathcal{G}_a$ et par hypothèse les $\mathcal{C}_a$ sont indépendants. $\square$

On rappelle qu'on appelle tribu engendré par une variable aléatoire X la plus petite tribu rendant l'application $X : \Omega \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ mesurable, i.e.

$$\sigma(X) = \left\{X^{-1}(A) : A \in \mathcal{B}(\mathbb{R})\right\}$$

La tribu $\sigma(X)$ est la tribu qui contient toutes les informations liées à la v.a. X.

Definition 67 (Indépendance de variables aléatoires)**.** *Deux variables aléatoires X et Y sont dites indépendants si $\sigma(X)$ et $\sigma(Y)$ sont des tribus indépendantes, i.e. si $\forall A, B \in \mathcal{B}(\mathbb{R})$,*

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

Une famille de variables aléatoires $(X_i)_{i\in I}$ est (mutuellement) indépendantes si leur tribus associées sont indépendantes, i.e. $\forall p \geqslant 1$, $\forall i_1, \dots, i_p \in I$, $A_{i_1}, \dots, A_{i_p} \in \mathcal{B}(\mathbb{R})$,

$$\mathbb{P}(X_{i_1} \in A_{i_1}, \dots, X_{i_p} \in A_{i_p}) = \prod_{k=1}^{p} \mathbb{P}(X_{i_k} \in A_{i_k}).$$

3 Critères d'indépendance et exemples

Proposition 68. *Soit $X = (X_1, \dots, X_d)$ un vecteur aléatoire. Les propositions suivantes sont équivalentes*

1. *les variables $X_1, \dots, X_d$ sont mutuellement indépendantes,*
2. *la loi du vecteur $X = (X_1, \dots, X_d)$ est la loi produit des marginales :*

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

3. *$F_X(x) = \prod_{i=1}^d F_{X_i}(x_i)$ $\forall x \in \mathbb{R}^d$ où $F_X(x) = \mathbb{P}(X_1 \leqslant x_1, \dots, X_d \leqslant x_d)$,*
4. *pour toute fonction $(h_i)_{1\leqslant i\leqslant d}$ mesurables bornées (ou positives, ou dans L^1)*

$$\mathbb{E}[h_1(X_1) \dots, h_d(X_d)] = \prod_{i-1}^{d} \mathbb{E}[h_i(X_i)],$$

5. *$\varphi_X(t) = \prod_{i=1}^d \varphi_{X_i}(t_i)$ $\forall t \in \mathbb{R}^d$ où $\varphi_X(t) = \mathbb{E}\left[e^{i<t,X>}\right]$ avec $t = (t_1, \dots, t_d)$,*
6. *si la transformée de Laplace de X est définie sur un voisinage de 0, $L_X(t) = \mathbb{E}[e^{<t,X>}] = \prod_{i=1}^d L_{X_i}(t_i)$ $\forall t \in \mathbb{R}^d$,*
7. *(cas discret) $\mathbb{P}(X_1 = x_1, \dots, X_d = x_d) = \prod_{i=1}^d \mathbb{P}(X_i = x_i)$ $\forall x \in \mathbb{R}^d$,*
8. *(cas à densité) si la densité de $X = (X_1, \dots, X_d)$ est à variables séparables, i.e.*

$$f_X(x_1, \dots, x_d) = f_{X_1}(x_1) \dots f_{X_d}(x_d).$$

Démonstration. 1. $\Leftrightarrow$ 2. : Comme l'ensemble des pavés engendrent $\mathcal{B}(\mathbb{R}^d) = \mathcal{B}(\mathbb{R})^{\otimes d}$, et est stable par intersection finie, les marginales sont indépendantes si et seulement si pour tout pavé $A = A_1 \times \dots A_d$

$$\mathbb{P}(X_1 \in A_1, \dots, X_d \in A_d) = \prod_{i=1}^{d} \mathbb{P}(X_i \in A_i)$$

Or $\mathbb{P}(X_1 \in A_1, \dots, X_d \in A_d) = \mathbb{P}_X(A)$ et $\prod_{i=1}^d \mathbb{P}(X_i \in A_i) = \mathbb{P}_{X_1} \otimes \dots \otimes \mathbb{P}_{X_d}(A)$.

Par conséquent, les marginales sont indépendantes si et seulement si $\mathbb{P}_X$ et $\mathbb{P}_{X_1} \otimes \ldots \otimes \mathbb{P}_{X_d}$ coincident sur les pavés (i.e sur $\mathcal{B}(\mathbb{R}^d)$).

2. $\Leftrightarrow$ 3. : L'ensemble $\{B =]-\infty, x_1] \times]-\infty, x_d], x_i \in \mathbb{R}\}$ est stable par intersection finie et engendre $\mathcal{B}(\mathbb{R}^d)$, donc on conclut par le théorème de classe monotone.

2. $\Leftrightarrow$ 4. On suppose les variables indépendantes, par le théorème de transfert, puis indépendance + théorème de Fubini,

$$\mathbb{E}[h_1(X_1)\ldots h_d(X_d)] = \int_{\mathbb{R}^d} h_1(x_1)\ldots h_d(x_d)\mathbb{P}_X(x_1,\ldots,x_d) = \prod_{i=1}^{d} \int h_i(x_i)\mathbb{P}_{X_i}(x_i) = \prod_{i-1}^{d} \mathbb{E}[h_i(X_i)].$$

Réciproquement, il suffit de prendre $h_i = \mathbb{1}_{A_i}$ avec $A_i \in \mathcal{B}(\mathbb{R})$.

1. $\Leftrightarrow 5., 6.$: les fonctions s'expriment comme des espérances de fonctionnelles de X et caractérisent la loi.

1. $\Leftrightarrow$ 7. : (en dimension 2) $\mathbb{P}(X \in A, Y \in B) = \sum_{x\in A, y\in B} \mathbb{P}(X = x, Y = y)$ et par Fubini-Tonelli $\mathbb{P}(X \in A)\mathbb{P}(Y \in B) = \sum_{x\in A, y\in B} \mathbb{P}(X = x)\mathbb{P}(Y = y)$.

1. $\Leftrightarrow$ 8. : $\mathbb{P}(X \in A, Y \in B) = \int_{A\times B} f_{(X,Y)}(x,y)dxdy$ et $\mathbb{P}(X \in A)\mathbb{P}(Y \in B) = \int_A f_X(x)dx \int_B f_Y(y)dy$. $\square$

Remarque. *Si X et Y sont indépendantes, connaitre les lois marginales suffit à connaitre la loi du couple (X,Y), ce qui est faux en général.*

Exemple.

1. Deux v.a. discrètes X et Y sont indépendantes ssi $\mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x)\mathbb{P}(Y = y)$ pour tout x, y. Il est plus simple de montrer que deux v.a. ne sont pas indépendantes car il suffit d'exhiber un couple (x,y) tel que $\mathbb{P}(X = x, Y = y) \neq \mathbb{P}(X = x)\mathbb{P}(Y = y)$. Par exemple, on lance deux dés et X est le nombre de chiffres pairs obtenus et Y est le max des chiffres, alors $\mathbb{P}(X = 0, Y = 2) = 0 \neq \frac{1}{4}\frac{1}{18} = \mathbb{P}(X = 0)\mathbb{P}(Y = 2)$. Les varaibles ne sont pas indépendantes.
2. Soit (X,Y) un couple de densité $f_{(X,Y)}(x,y) = \frac{1}{3}\mathbb{1}_{[0,1]\times[-1,2]}(x,y)$. La fonction est à variables séparables, il y a donc indépendance.
3. Soit (X,Y) un couple de densité $f_{(X,Y)}(x,y) = \frac{1}{3\prod} e^{-\frac{x^2+2xy+5y^2}{6}}$. Il n'y a pas indépendance, car $f_{(X,Y)}(x,y) \neq f_X(x)f_Y(y)$ où $f_X(x) = e^{-4x^2/30}/\sqrt{15\prod/2}$ et $f_Y(y) = e^{-4x^2/6}/\sqrt{3\prod/2}$.
4. Soit (X,Y) un couple de densité $f_{(X,Y)}(x,y) = \sqrt{x/y}$ si $0 < y \leqslant x < 1$ et 0 sinon. Les variables ne sont pas indépendantes.

Propriété 69. *Soit $(X_i)_{i\in I}$ une famille de v.a.. Les variables sont mutuellement indépendantes si et seulement si pour toute sous-partie finie J de I les variables $(X_j)_{j\in J}$ sont mutuellement indépendantes.*

Démonstration. Si $(X_i)_{i\in I}$ indépendantes. Soit $J \subset I$ finie. Alors par le théorème de transfert pour tout h_j borélienne telle que $h_j(X_j) \in L^1$,

$$\mathbb{E}\left[\prod_{i\in J} h_j(X_j)\right] = \int \prod_{i\in J} h_j(x_j)\mathbb{P}_{X_1},$$

$\square$

4 Corrélation et indépendance

Rappels :
Soient X et Y deux v.a. dans L^2. La covariance est définie par

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

Si X et Y sont deux vecteurs aléatoires (ligne) dans L^2, alors

$$Cov(X,Y) = \mathbb{E}\big[(X-\mathbb{E}[X])^t(Y-\mathbb{E}[Y])\big] = \mathbb{E}[X^tY] - \mathbb{E}[X]^t\mathbb{E}[Y].$$

Definition 70. *Deux variables aléatoires X et Y dans L^2 sont dites non corrélées si $Cov(X,Y)=0$.*

Remarque. *Si X et Y sont indépendantes dans L^2 alors $Cov(X,Y)=0$, elles sont non corrélées. Mais la réciproque est fausse.*

Exemple. Soit $U \sim \mathcal{U}[-1,1]$ et $V=U^2$, alors par parité $\mathbb{E}[UV]=0$ et $\mathbb{E}[U]=0$. Les variables ne sont pas corrélées. Par contre elles ne sont pas indépendantes : $\mathbb{P}(|U|<1/2, V>1/2) = 0 \neq \mathbb{P}(|U|<1/2)\mathbb{P}(V>1/2)$.

Propriété 71. *Soient X, Y deux variables non corrélées alors*

$$Var(X+Y) = Var(X) + Var(Y).$$

Si $X_1,\ldots,X_n$ v.a. dans L^2 deux à deux non corrélées, alors

$$Var(X_1+\ldots+X_n) = \sum_{i=1}^n Var(X_i).$$

Démonstration. On rappelle que l'application covariance est bilinéaire, par conséquent,

$$\begin{aligned} Var(\sum_{i=1}^n X_i) &= Cov(\sum_{i=1}^n X_i, \sum_{j=1}^n X_j) \\ &= \sum_{i,j=1}^n Cov(X_i,X_j) = \sum_{i=1}^n Var(X_i) + \sum_{i\neq j} Cov(X_i,X_j). \end{aligned}$$

On conclut en remarquant que $Cov(X_i,X_j)=0$ pour $i\neq j$. □

5 Événements asymptotiques

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

5.1 Tribus du futur et tribu asymptotique

Definition 72. *Soit $(\mathcal{F}_n)_{n\geqslant 0}$ une suite de sous-tribus de $\mathcal{F}$.*
La suite $(\mathcal{F}_n)_{n\geqslant 0}$ est appelée filtration si la suite est croissante, i.e. $\forall n\geqslant 0$, $\mathcal{F}_n \subset \mathcal{F}_{n+1}$.

Exemple. Souvent quand $(X_n)_{n\geqslant 1}$ est une suite de v.a., on lui associe sa filtration naturelle :

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

On a bien $\mathcal{F}_n \subset \mathcal{F}_{n+1}$.
Par exemple, on lance une infinité de fois une pièce et X_n est le résultat du $n^{\text{ième}}$ lancer. n représente le temps. $\mathcal{F}_n$ contient alors toutes les informations de ce qui s'est passé jusqu'à l'instant n : le passé.

Definition 73. *Soit $(\mathcal{F}_n)_{n\geqslant 0}$ une suite de sous-tribus de $\mathcal{F}$. On appelle tribus du futur, pour $n \in \mathbb{N}$,*

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

On appelle tribu asymptotique la tribu

$$\mathcal{F}^\infty = \bigcap_{n\geqslant 0} \mathcal{F}^n.$$

Exemple. Si on lance une infinité de fois une pièce et X_n est le résultat du $n^{\text{ième}}$ lancer. Si $\mathcal{F}_n = \sigma(X_n)$ représente le présent (pas une filtration en général !), alors $\mathcal{F}^n = \sigma(X_k : k \geqslant n)$ contient toutes les informations de ce qui se passe après l'instant n : le futur, et $\mathcal{F}^\infty$ contient les événements asymptotiques.

Exemple. Soit $(X_n)_{n\geqslant 1}$ une suite de variables aléatoires. On pose $\mathcal{F}_n = \sigma(X_n)$.

— On considère l'événement $A = \{X_n \text{ passe une infinité de fois par } 0\}$. On a

$$A = \{\forall n \geqslant 1, \exists k \geqslant n : X_k = 0\} = \bigcap_{n\geqslant 0}\bigcup_{k\geqslant n} \{X_k = 0\} = \limsup \{X_n = 0\}.$$

On va montrer que $a \in \mathcal{F}^p$ pour tout $p \geqslant 1$. Soit $p \geqslant 1$ fixé. Pour tout $n \geqslant p$, on a $\bigcup_{k\geqslant n} \{X_k = 0\} \in \mathcal{F}^p$. Comme la suite $\left(\bigcup_{k\geqslant n} \{X_k = 0\}\right)_{n\geqslant 0}$ est décroisante, on a $\bigcap_{n\geqslant 1}\bigcup_{k\geqslant n} \{X_k = 0\} = \bigcap_{n\geqslant p}\bigcup_{k\geqslant n} \{X_k = 0\}$. Par conséquent, pour tout $p \geqslant 0$,

$$\limsup \{X_n = 0\} \in \mathcal{F}^p,$$

i.e. $\limsup \{X_n = 0\} \in \mathcal{F}^\infty$. L'événement A est un événement asymptotique.

— On considère maintenant l'événement $B = \{\text{les } X_n \text{ sont bornés par } M\}$. On a

$$B = \{\forall n \geqslant 1, |X_n| \leqslant M\} = \bigcap_{n\geqslant 1} \{|X_n| \leqslant M\} = \{|X_1| \leqslant M\} \cap \bigcap_{n\geqslant 2} \{|X_n| \leqslant M\}.$$

Cet événement n'appartient pas à la tribu asymptotique en général car $\{|X_1| \leqslant M\} \notin \mathcal{F}^n$ pour $n \geqslant 2$, car $\mathcal{F}^n = \sigma(X_n, X_{n+1}, \ldots)$.

— On considère l'événement $C = \{\text{la suite } (X_n)_{n\geqslant 1} \text{ converge}\}$. On a

$$\begin{aligned} C &= \{\text{la suite } (X_n)_{n\geqslant 1} \text{ est de Cauchy}\} \\ &= \left\{\forall \varepsilon \in \mathbb{Q}^+, \exists N \geqslant 1, \forall n \geqslant N, q \geqslant 0 : |X_n - X_{n+q}| < \varepsilon\right\} \\ &= \bigcap_{\varepsilon\in\mathbb{Q}^+}\bigcup_{N\geqslant 1}\bigcap_{\substack{n\geqslant N\\ q\geqslant 0}} \{|X_n - X_{n+q}| < \varepsilon\}. \end{aligned}$$

Soit $p \geqslant 1$ fixé. On a $\{\text{la suite } (X_n)_{n\geqslant 1} \text{ converge}\} = \{\text{la suite } (X_{n+p})_{n\geqslant 0} \text{ converge}\}$, par conséquent

$$C = \bigcap_{\varepsilon\in\mathbb{Q}^+}\bigcup_{N\geqslant p}\bigcap_{\substack{n\geqslant N\\ q\geqslant 0}} \{|X_n - X_{n+q}| < \varepsilon\}.$$

Cependant, $\bigcap_{n\geqslant N, q\geqslant 0} \{|X_n - X_{n+q}| < \varepsilon\} \in\subset \mathcal{F}^p$ pour tout $N \geqslant p$, donc

$$\bigcup_{N\geqslant p}\bigcap_{\substack{n\geqslant N\\ q\geqslant 0}} \{|X_n - X_{n+q}| < \varepsilon\} \in \mathcal{F}^p.$$

Par stabilité par intersection dénombrable, $C \in \mathcal{F}^p$ pour tout $p \geqslant 0$, et donc $C \in \mathcal{F}^\infty$.

Theorem 74 (Loi du 0-1 de Kolmogorov (Tout ou Rien))**.** *Soit $(\mathcal{F}_n)_{n\geqslant 1}$ une suite de sous-tribus indépendantes. Alors la tribu asymptotique est triviale, i.e.*

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

Conséquence 75. Soit $(X_n)_{n\geqslant 1}$ une suite de variables (mutuellement) indépendantes. On pose $\mathcal{F}_n = \sigma(X_n)$. Les tribus sont indépendantes. Alors comme A et C sont asymptotiques, on a $\mathbb{P}(A) = 0$ ou 1 et $\mathbb{P}(C) = 0$ ou 1. Par conséquent, soit p.s. la suite passe une infinité de fois par 0 soit p.s. la suite passe un nombre fini de fois par 0. De même, soit p.s. la suite $(X_n)_{n\geqslant 1}$ converge, soit p.s. la suite ne converge pas.

Démonstration. On va montrer que $\mathcal{F}^\infty$ est indépendante d'elle-même. En effet, si $\mathcal{F}^\infty$ est indépendante de $\mathcal{F}^\infty$, alors tout événement $A \in \mathcal{F}^\infty$ est indépendant de lui-même : $\mathbb{P}(A) = \mathbb{P}(A \cap A) = \mathbb{P}(A)^2$ et donc $\mathbb{P}(A) = 0$ ou 1.
Soit $n \geqslant 1$.
On rappelle que $\mathcal{F}^n = \sigma\big(\bigcup_{k\geqslant n} \mathcal{F}_k\big)$, donc $\mathcal{F}^n$ est indépendant de $\sigma\Big(\bigcup_{p<n} \mathcal{F}_p\Big)$.
Comme $\mathcal{F}^\infty = \bigcap_{m\geqslant 1} \mathcal{F}^m$, on a $\mathcal{F}^\infty \subset \mathcal{F}^n$. $\mathcal{F}^\infty$ est donc indépendant de $\sigma\Big(\bigcup_{p<n} \mathcal{F}_p\Big)$ pour tout $n \geqslant 1$.
Par conséquent, $\mathcal{F}^\infty$ est donc indépendant de $\bigcup_{n\geqslant 1} \sigma\Big(\bigcup_{p<n} \mathcal{F}_p\Big)$.
L'ensemble $\mathcal{C} = \bigcup_{n\geqslant 1} \sigma\Big(\bigcup_{p<n} \mathcal{F}_p\Big)$ est stable par intersection finie. En effet si $A, B \in \mathcal{C}$, il existe n, m tels que $A \in \sigma\Big(\bigcup_{p<n} \mathcal{F}_p\Big)$ et $B \in \sigma\Big(\bigcup_{p<m} \mathcal{F}_p\Big)$, par conséquent $A \cap B \in \sigma\Big(\bigcup_{p<\max(n,m)} \mathcal{F}_p\Big) \subset \mathcal{C}$.
$\mathcal{F}^\infty$ étant stable par intersection finie, on en déduit que $\mathcal{F}^\infty$ et $\sigma(\mathcal{C})$ sont indépendantes.
Par ailleurs, pour tout $p \geqslant 1$, $\mathcal{F}_p \subset \mathcal{C}$ et donc $\bigcup_{p\geqslant 1} \mathcal{F}_p \subset \mathcal{C}$.
Comme $\mathcal{F}^\infty \subset \mathcal{F}^1 = \sigma\Big(\bigcup_{p\geqslant 1} \mathcal{F}_p\Big) \subset \sigma(\mathcal{C})$, on en déduit que la tribu est indépendante d'elle-même. □

5.2 Lemmes de Borel-Cantelli

Ce sont des lemmes très important en probabilité. L'un est trivial et l'autre est plus subtile.

Theorem 76 (Lemme de Borel-Cantelli 1)**.** *Soit $(A_n)_{n\geqslant 1}$ une suite d'événements telle que*

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

alors $\mathbb{P}(\limsup_{n\to\infty} A_n) = 0$. Par conséquent, seul un nombre fini d'événements A_n sont réalisés.

Démonstration. On rappelle que $\limsup A_n = \bigcap_{n\geqslant 1} \bigcup_{k\geqslant n} A_k$. La suite $\big(\bigcup_{k\geqslant n} A_k\big)_{n\geqslant 1}$ est décroissante, donc

$$\mathbb{P}(\limsup_{n\to\infty} A_n) = \lim \mathbb{P}\left(\bigcup_{k\geqslant n} A_k\right).$$

Par ailleurs, $\mathbb{P}\big(\bigcup_{k\geqslant n} A_k\big) \leqslant \sum_{k\geqslant n} \mathbb{P}(A_k)$. Comme la série $\sum_{n=1}^{\infty} \mathbb{P}(A_n)$ converge, son reste converge vers 0 et donc $\mathbb{P}(\limsup A_n) = 0$. □

Theorem 77 (Lemme de Borel-Cantelli 2)**.** *Soit $(A_n)_{n\geqslant 1}$ une suite d'événements* **indépendants** *telle que*

$$\sum_{n=1}^{\infty} \mathbb{P}(A_n) = +\infty,$$

alors $\mathbb{P}(\limsup_{n\to\infty} A_n) = 1$. Par conséquent, ps. une infinité d'événements A_n sont réalisés.

Démonstration. On a $\mathbb{P}(\limsup A_n) = 1 - \mathbb{P}(\liminf \overline{A_n})$. Comme la suite $\left(\bigcap_{k\geqslant n} \overline{A_k}\right)_{n\geqslant 1}$ est croissante, on a

$$\begin{aligned}
\mathbb{P}\big(\liminf \overline{A_n}\big) &= \mathbb{P}\left(\bigcup_{n\geqslant 0}\bigcap_{k\geqslant n} \overline{A_k}\right)\\
&= \lim_{n\to\infty} \mathbb{P}\left(\bigcap_{k\geqslant n} \overline{A_k}\right)\\
&= \lim_{n\to\infty}\lim_{m\to\infty} \mathbb{P}\left(\bigcap_{k=n}^{m} \overline{A_k}\right) \text{ car } \bigcap_{k\geqslant n} \overline{A_k} = \bigcap_{m\geqslant 0}\bigcap_{k=n}^{m} \overline{A_k} \text{ (intersection décroissante)}\\
&= \lim_{n\to\infty}\lim_{m\to\infty} \prod_{k=n}^{m} \mathbb{P}\big(\overline{A_k}\big) \text{ par indépendance}\\
&= \lim_{n\to\infty}\lim_{m\to\infty} \prod_{k=n}^{m} (1-\mathbb{P}(A_k)).
\end{aligned}$$

Or comme $0 \leqslant 1-x \leqslant e^{-x}$ pour $x\in[0,1]$, on a $\prod_{k=n}^{m}(1-\mathbb{P}(A_k)) \leqslant \exp\left(-\sum_{k=n}^{m}\mathbb{P}(A_k)\right)$. Comme $\sum_{n=1}^{\infty}\mathbb{P}(A_n) = +\infty$, on en déduit que

$$\mathbb{P}\Big(\liminf_{n\to\infty} \overline{A_n}\Big) = 0.$$

□

6 Somme de deux variables aléatoires indépendantes

Soient X et Y deux variables aléatoires indépendantes. On cherche la loi de $X+Y$. Comme X et Y sont indépendantes, la loi de (X,Y) est $\mathbb{P}_X\otimes\mathbb{P}_Y$.
Par conséquent, soit $z\in\mathbb{R}$, on a

$$\begin{aligned}
F_{X+Y}(z) = \mathbb{P}(X+Y\leqslant z) &= \iint_{\mathbb{R}^2} \mathbb{1}_{x+y\leqslant z}\mathbb{P}_X(dx)\mathbb{P}_Y(dy)\\
&= \int_{\mathbb{R}} \mathbb{P}_X(dx)\left(\int_{-\infty}^{z-x}\mathbb{P}_Y(dy)\right) \text{ par Fubini}\\
&= \int_{\mathbb{R}} F_Y(z-x)\mathbb{P}_X(dx)\\
&= \int_{\mathbb{R}} F_X(z-y)\mathbb{P}_Y(dy) \text{ de façon similaire.}
\end{aligned}$$

6.1 Convolution de mesures

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

Definition 78. *Soient μ et ν deux mesures positives sur $(E,\mathcal{E})$ $\sigma-$finies. On définit le produit de convolution $\mu*\nu$ entre deux mesures comme la mesure définie sur $(E,\mathcal{E})$ par*

$$\mu*\nu(A) = \iint_{E^2}\mathbb{1}_{x+y\in A}\nu(dx)\mu(dy) = \int_E \mu(A-x)\nu(dx),$$

où $A-x = \{y-x : y\in A\}$.

Il est facile de vérifier que $\mu * \nu$ soit une mesure : $\mu * \nu(\emptyset) = 0$, si $(A_n)_{n\geqslant 0}$ disjoints alors à x fixé, $(A_n - x)$ sont disjoints et par convergence monotone,

$$\mu * \nu\Big(\bigcup A_n\Big) = \int_E \sum_n \mu(A_n - x)\nu(dx) = \sum_n \mu * \nu(A_n).$$

Proposition 79. *1. La convolution est commutative : $\mu * \nu = \nu * \mu$,*

2. δ_0 est élément neutre : $\mu * \delta_0 = \mu$,

3. si μ et ν deux mesures de probabilités, alors $\mu * \nu$ aussi.

Démonstration. Soit $A \in \mathcal{E}$, alors

1. $\mu * \nu(A) = \iint \mathbb{1}_{x+y\in A}\mu(dx)\nu(dy) = \nu * \mu(A)$.
2. $\mu * \delta_0(A) = \int_E \delta_0(A - x)\mu(dx) = \int_E \mathbb{1}_{x\in A}\mu(dx) = \mu(A)$.
3. $\mu * \nu(E) = \iint \mathbb{1}_{x+y\in E}\mu(dx)\nu(dy) = \iint_{E^2} \mu(dx)\nu(dy) = \mu(E)\nu(E) = 1$.

□

Par conséquent si X et Y sont deux variables indépendantes, on a pour tout borélien $A \in \mathcal{B}(\mathbb{R})$,

$$\begin{aligned}\mathbb{P}(X + Y \in A) &= \iint \mathbb{1}_{x+y\in A}\mathbb{P}_X(dx)\mathbb{P}_Y(dy) \\ &= \int \mathbb{P}_Y(A - x)\mathbb{P}_X(dx).\end{aligned}$$

D'où

$$\mathbb{P}_{X+Y} = \mathbb{P}_X * \mathbb{P}_Y,$$

et par récurrence si $X_1, \ldots, X_n$ indépendantes, alors

$$\mathbb{P}_{\sum_{k=1}^n X_k} = \mathbb{P}_{X_1} * \ldots * \mathbb{P}_{X_n}.$$

Remarque. *Si $X_1, \ldots, X_n$ indépendantes alors la fonction caractéristique φ de $\sum_{k=1}^n X_k$ vérifie*

$$\varphi(t) = \prod_{k=1}^n \varphi_{X_k}(t).$$

6.2 Cas des variables discrètes

Soit X et Y deux variables discrètes à valeurs respectivement dans $S(X) = \{x_i\}_{i\in I}$ et $S(Y) = \{y_j\}_{j\in J}$. Alors $X + Y$ est à valeurs dans $S(X + Y) = \{z_k\}_{k\in K}$ avec $z_k \in \{x_i + y_j : i \in I, j \in J\}$.
Si $S(X) = S(Y) = \mathbb{N}$, alors $S(X + Y) = \mathbb{N}$ et pour $n \in \mathbb{N}$, on a par indépendance

$$\mathbb{P}(X + Y = n) = \sum_{k=0}^n \mathbb{P}(X = k, Y = n - k) = \sum_{k=0}^n \mathbb{P}(X = k)\mathbb{P}(Y = n - k).$$

Exemple. La loi Binomiale $\mathcal{B}(n, p)$ est la loi de la somme de n variables de Bernoulli $\mathcal{B}(p)$ indépendantes.

6.3 Cas des variables à densité

Proposition 80. *Soit X et Y deux variables indépendantes de densité respective f_X et f_Y. Alors $X+Y$ est une variable à densité de densité $f_X * f_Y$ où*

$$f_X * f_Y(z) = \int_{\mathbb{R}} f_X(x) f_Y(z-x) dx = \int_{\mathbb{R}} f_X(z-x) f_Y(x) dx.$$

Démonstration. Soit $A \in \mathcal{B}(\mathbb{R})$, alors

$$\begin{aligned}\mathbb{P}_{X+Y}(A) = \mathbb{P}_X * \mathbb{P}_Y(A) &= \iint \mathbb{1}_A(x+y) \mathbb{P}_X(dx) \mathbb{P}_Y(dy) \\ &= \iint \mathbb{1}_A(x+y) f_X(x) f_Y(y) dx dy.\end{aligned}$$

Par changement de variables $u = x$ et $v = x+y$ et Fubini-Tonelli, on a

$$\mathbb{P}_{X+Y}(A) = \iint \mathbb{1}_A(v) f_X(u) f_Y(v-u) du dv = \int_A \left(\int_{\mathbb{R}} f_X(u) f_Y(v-u) du \right) dv.$$

Ceci est vrai $\forall A \in \mathcal{B}(A)$, donc $X+Y$ est à densité de densité $f_X * f_Y$. $\square$

Exemple. Soit X et Y deux variables gaussiennes indépendantes de loi respective $\mathcal{N}(m_1, \sigma_1^2)$ et $\mathcal{N}(m_2, \sigma_2^2)$. Alors $X+Y$ suit la loi gaussienne $\mathcal{N}(m_1+m_2, \sigma_1^2+\sigma_2^2)$.
On pourrait démontrer le résultat en utilisant la proposition précédente, mais les calculs sont fastidieux (voir p.104 du polycopié de Jean-Christophe Breton). Regardons la fonction caractéristique φ de $X+Y$. Par indépendance,

$$\varphi(t) = \varphi_X(t) \varphi_Y(t) = e^{im_1 + \frac{\sigma_1^2}{2} t^2} e^{im_2 + \frac{\sigma_2^2}{2} t^2} = e^{i(m_1+m_2) + \frac{\sigma_1^2 + \sigma_2^2}{2} t^2}.$$

On reconnaît la fonction caractéristique de $\mathcal{N}(m_1+m_2, \sigma_1^2+\sigma_2^2)$.

Chapitre 6

Convergence d'une suite de variables aléatoires

On se place sur un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$. On considère une suite de v.a. $(X_n)_{n\geqslant 0}$ et X une variable aléatoire à valeurs dans $\left(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d)\right)$. On note $|.|$ la norme euclidienne dans $\mathbb{R}^d$.

1 Convergences trajectorielles

1.1 Convergence presque-sûre

Cette notion correspond à la convergence simple pour les fonctions, cependant exiger d'avoir $\forall \omega \in \Omega$, $X_n(\omega) \to X(\omega)$ est trop restrictif (les v.a. sont souvent définies à un ensemble négligeable près).

Definition 81 (Convergence p.s.)**.** *Une suite de v.a. $(X_n)_{n\geqslant 0}$ converge vers X p.s. s'il existe $N \in \mathcal{F}$ tel que $\mathbb{P}(N) = 0$ et*

$$\forall \omega \notin N, X_n(\omega) \to X(\omega).$$

On note $X_n \xrightarrow{p.s} X$.

Remarque. 1. *Si $X_n \xrightarrow{p.s} X$, alors*

$$\mathbb{P}(X_n \to X) = 1,$$

car, en reprenant les notations de la définition, $\overline{N} \subset \{\omega \in \Omega : X_n(\omega) \to X(\omega)\}$ et $\mathbb{P}(\overline{N}) = 1$.

2. *$X_n = (X_n^1, \ldots, X_n^d) \xrightarrow{p.s} X = (X^1, \ldots, X^d)$ si et seulement si $\forall k \in \{1, \ldots, d\}$, $X_n^k \xrightarrow{p.s} X^k$.*

Proposition 82. *$X_n \xrightarrow{p.s} X$ si et seulement si $\forall \varepsilon > 0$ $\mathbb{P}(\limsup_n \{|X_n - X| > \varepsilon\}) = 0$.*

Démonstration. Supposons que $X_n \xrightarrow{p.s} X$. Il existe $N \in \mathcal{F}$ tel que $\mathbb{P}(N) = 0$ et $\forall \omega \notin N$, $X_n(\omega) \to X(\omega)$.
Soit $\varepsilon > 0$.
Si $\omega \in \limsup_n \{|X_n - X| > \varepsilon\} = \bigcap_{n\geqslant 0} \bigcup_{k\geqslant n} \{|X_n - X| > \varepsilon\}$, alors il existe une infinité de n tels que $|X_n(\omega) - X(\omega)| > \varepsilon$, ce qui implique que la suite $X_n(\omega)$ ne converge pas vers $X(\omega)$. Par conséquent $\limsup_n \{|X_n - X| > \varepsilon\} \subset N$, et donc $\mathbb{P}(\limsup_n \{|X_n - X| > \varepsilon\}) = 0$.

Supposons que $\forall \varepsilon > 0$, $\mathbb{P}(\limsup_n \{|X_n - X| > \varepsilon\}) = 0$.
On pose $N = \bigcup_{p\geqslant 0} \limsup_n \{|X_n - X| > 2^{-p}\}$. N est un ensemble négligeable comme union dénombrable d'ensemble négligeable.
Soit $\omega \notin N$. Alors $\forall p \geqslant 0$ $\omega \in \liminf_n \{|X_n - X| \leqslant 2^{-p}\}$. Il existe n_ω tel que $\forall k \geqslant n_\omega$, $|X_k(\omega) - X(\omega)| \leqslant 2^{-p}$. Donc si $\omega \notin N$, $X_n(\omega) \to X_(\omega)$. □

Corollary 83. *Si pour tout $\varepsilon > 0$ $\sum_{n\geqslant 0} \mathbb{P}(|X_n - X| > \varepsilon) < +\infty$, alors la suite $(X_n)_{n\geqslant 0}$ converge vers X p.s.*

Démonstration. Il suffit d'appliquer le premier lemme de Borel-Cantelli pour avoir $\forall \varepsilon > 0, \mathbb{P}(\limsup_n \{|X_n - X| > \varepsilon\}) = 0$. □

Proposition 84. *On suppose les variables $(X_n)_{n\geqslant 0}$ mutuellement indépendantes.*
La suite $(X_n)_{n\geqslant 0}$ converge presque surement vers 0 si et seulement si $\forall \varepsilon > 0$, $\sum_{n\geqslant 0} \mathbb{P}(|X_n| > \varepsilon) < +\infty$.

Démonstration. Il suffit de montrer l'équivalence dans un sens. Supposons qu'il existe $\varepsilon > 0$ tel que $\sum_{n\geqslant 0} \mathbb{P}(|X_n| > \varepsilon) = +\infty$. Les événements $A_n = \{|X_n| > \varepsilon\}$ étant indépendants, par le second lemme de Borel-Cantelli on a $\mathbb{P}(\limsup_n \{|X_n| > \varepsilon\}) = 1$. Donc $(X_n)_{n\geqslant 0}$ ne converge pas vers 0. Contradiction.
Rque : La preuve ne fonctionne que si la limite est nulle pour avoir l'indépendance des A_n. □

Proposition 85. *Soit $(\varepsilon_n)_{n\geqslant 0}$ une suite de réels positifs telle que $\sum_{n\geqslant 0} \varepsilon_n < +\infty$. Si*

$$\sum_{n\geqslant 0} \mathbb{P}(|X_{n+1} - X_n| > \varepsilon_n) < +\infty,$$

alors $(X_n)_{n\geqslant 0}$ converge p.s.

Démonstration. D'après le premier lemme de Borel-Cantelli, on a $\mathbb{P}(\limsup \{|X_{n+1} - X_n| > \varepsilon_n\}) = 0$.
On remarque que $X_n = X_0 + \sum_{k=0}^{n-1} (X_{k+1} - X_k)$.
Soit $\omega \in \liminf \{|X_{n+1} - X_n| \leqslant \varepsilon_n\}$, alors il existe $n_\omega \geqslant 0$ tel que pour tout $k \geqslant n_\omega$, $|X_{k+1}(\omega) - X_k(\omega)| \leqslant \varepsilon_k$.
Comme la série de $(\varepsilon_n)_{n\geqslant 0}$ converge, la série $\sum_{k=0}^{n-1} (X_{k+1}(\omega) - X_k(\omega))$ est absolument convergente. Donc $(X_n(\omega))$ converge pour tout $\omega \in \liminf \{|X_{n+1} - X_n| \leqslant \varepsilon_n\}$, qui est de probabilité 1. La suite converge p.s. □

On énonce maintenant quelques propriétés immédiates :

Propriété 86. *Soit $(X_n)_{n\geqslant 0}$ une suite de v.a..*

1. *Si $X_n \xrightarrow{p.s} X$ et $X_n \xrightarrow{p.s} Y$, alors $X = Y$ p.s. (unicité p.s. de la limite).*
2. *Si $X_n \xrightarrow{p.s} X$ et $f : \mathbb{R}^d \to \mathbb{R}^q$ continue, alors $f(X_n) \xrightarrow{p.s} f(X)$.*
3. *Si $X_n = Y_n$ p.s. pour tout $n \geqslant 0$, alors $X_n \xrightarrow{p.s} X$ si et seulement si $Y_n \xrightarrow{p.s} X$.*

Démonstration. En exo. □

1.2 Convergence dans L^p

Soit $p \in [1, \infty[$.
On rappelle que l'espace L^p est l'espace quotient de l'ensemble $\mathcal{L}^p$, des v.a. X tel que $|X|^p$ intégrable, par la relation d'équivalence $\sim_{ps}$, "être égal presque sûrement".

Definition 87. *Soit $(X_n)_{n\geqslant 0}$ une suite de v.a. de L^p. On dit que la suite converge dans L^p vers X, noté $X_n \xrightarrow{L^p} X$, si*

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

Remarque. 1. *$X_n = (X_n^1, \dots, X_n^d) \xrightarrow{L^p} X = (X^1, \dots, X^d)$ ssi $X_n^k \xrightarrow{L^p} X^k$, $\forall k \in \{1, \dots, d\}$.*

2. *Quand $p = 1$, on parle de convergence en moyenne et quand $p = 2$ on parle de convergence quadratique.*
3. *Une conséquence de l'inégalité triangulaire est l'unicité de la limite, au sens p.s..*

Proposition 88. *Soit $(X_n)_{n\geqslant 0}$ une suite de v.a. de L^q. Si $X_n \xrightarrow{L^q} X$, alors $X_n \xrightarrow{L^p} X$ pour tout $p \leqslant q$.*

Démonstration. La preuve repose sur l'inégalité de Hölder. En effet, si $p \leqslant q$,

$$\mathbb{E}[|X_n - X|^p] \leqslant \mathbb{E}[|X_n - X|^q]^{p/q}.$$

□

Le théorème de convergence dominée permet d'obtenir la convergence dans L^p à partir de la convergence ps.

Theorem 89 (Convergence ps vers convergence dans L^p)**.**
Soient $(X_n)_{n\geqslant 0}$ et X des v.a. avec $X_n \xrightarrow{p.s} X$. S'il existe Y va. positive telle que

$$\forall n \geqslant 0, |X_n| \leqslant Y \quad \mathbb{P} - ps \text{ et } Y \in L^p,$$

alors $X_n \in L^p$ et $(X_n)_{n\geqslant 0}$ converge vers X dans L^p.

Remarque. *Si $p = \infty$, alors*

$$\begin{aligned} ||X||_\infty &= essup|X| \\ &= \inf\{c \geqslant 0 : \mathbb{P}(|X| > c) = 0\}. \end{aligned}$$

L'espace $(L^\infty, ||.||_\infty)$ est un espace de Banach. Si $X_n \xrightarrow{L^\infty} X$ alors pour tout $p \geqslant 1$, $X_n \xrightarrow{L^p} X$ et $X_n \xrightarrow{p.s} X$.

Exemple. Soit $(X_n)_{n\geqslant 0}$ une suite de v.a. indépendantes de loi

$$\mathbb{P}(X_n = x_n) = p_n, \quad \mathbb{P}(X_n = 0) = 1 - p_n,$$

avec $p_n \in]0,1[$ et $x_n \geqslant 1$.
La variables étant indépendantes, la suite $(X_n)_{n\geqslant 0}$ converge p.s. vers 0 si et seulement si $\forall \varepsilon > 0$, $\sum_{n\geqslant 0} \mathbb{P}(|X_n| > \varepsilon) < \infty$, soit $\sum_{n\geqslant 0} p_n < \infty$.
La suite $(X_n)_{n\geqslant 0}$ converge dans L^p vers 0 si et seulement si $\mathbb{E}[|X_n|^p] = p_n x_n^p \underset{n\to\infty}{\longrightarrow} 0$.
Par conséquent,
- si $p_n = 2^{-n}$ et $x_n = 2^n$, on a $X_n \xrightarrow{p.s} 0$ et $(X_n)_{n\geqslant 0}$ ne converge pas dans L^1,
- si $p_n = 1/n$ et $x_n = 1$, on a $X_n \xrightarrow{L^1} 0$ et $(X_n)_{n\geqslant 0}$ ne converge pas ps vers 0,
- si $p_n = 1/n^2$ et $x_n = n$, on a $X_n \xrightarrow{p.s} 0$, $X_n \xrightarrow{L^1} 0$, mais $(X_n)_{n\geqslant 0}$ ne converge pas dans L^2.

1.3 Convergence en probabilité

Definition 90. *Une suite de v.a. $(X_n)_{n\geqslant 0}$ converge en probabilité vers X, noté $X_n \xrightarrow{Proba} X$ ou $X_n \xrightarrow{\mathbb{P}} X$, si pour tout $\varepsilon > 0$,*

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

Proposition 91. *$X_n = (X_n^1, \ldots, X_n^d) \xrightarrow{Proba} X = (X^1, \ldots, X^d)$ ssi $X_n^k \xrightarrow{Proba} X^k$, $\forall k \in \{1, \ldots, d\}$.*

Démonstration. Si $X_n \overset{Proba}{\longrightarrow} X$, alors pour tout $\varepsilon > 0$,

$$\mathbb{P}(|X_n^k - X^k| > \varepsilon) \leqslant \mathbb{P}(|X_n - X| > \varepsilon)$$

car $|X_n^k - X^k| \leqslant |X_n - X| = \left(\sum_{k=1}^d |X_n^k - X^k|^2\right)^{1/2}$.

Supposons maintenant que $X_n^k \overset{Proba}{\longrightarrow} X^k$, $\forall k \in \{1, \dots, d\}$. On a

$$\{|X_n - X| > \varepsilon\} \subset \bigcup_{1 \leqslant k \leqslant d} \left\{|X_n^k - X^k| > \varepsilon/\sqrt{d}\right\}$$

En effet, si pour tout $k \in \{1, \dots, d\}$, $|X_n^k - X^k| \leqslant \varepsilon/\sqrt{d}$ alors $|X_n - X| \leqslant \varepsilon$.
Par conséquent,

$$\begin{aligned}\mathbb{P}(\{|X_n - X| > \varepsilon\}) &\leqslant \mathbb{P}\left(\bigcup_{1 \leqslant k \leqslant d} \left\{|X_n^k - X^k| > \varepsilon/\sqrt{d}\right\}\right) \\ &\leqslant \sum_{k=1}^d \mathbb{P}\Big(|X_n^k - X^k| > \varepsilon/\sqrt{d}\Big).\end{aligned}$$

Par hypothèse, on en déduit donc que $\lim_{n\to\infty} \mathbb{P}(|X_n - X| > \varepsilon) = 0$. □

Remarque. *Si $X_n \overset{Proba}{\longrightarrow} X$ et $X_n \overset{Proba}{\longrightarrow} Y$, alors $X = Y$ ps.*

Démonstration. Soit $\varepsilon > 0$, $|X - Y| \leqslant |X - X_n| + |X_n - Y|$ et donc on a

$$\mathbb{P}(|X - Y| > \varepsilon) \leqslant \lim_{n\to\infty} [\mathbb{P}(|X_n - X| > \varepsilon/2) + \mathbb{P}(|X_n - Y| > \varepsilon/2)] = 0.$$

Comme $\mathbb{P}(X \neq Y) = \mathbb{P}\big(\bigcup_{k\geqslant 1} |X - Y| > 1/k\big) = 0$, on en déduit $X = Y$ ps. □

Proposition 92. *Soit $f : \mathbb{R}^d \to \mathbb{R}^q$ continue. Si $X_n \overset{Proba}{\longrightarrow} X$, alors $f(X_n) \overset{Proba}{\longrightarrow} f(X)$.*

Par conséquent, si $X_n \overset{Proba}{\longrightarrow} X$ et $Y_n \overset{Proba}{\longrightarrow} Y$, alors $X_n + Y_n \overset{Proba}{\longrightarrow} X + Y$ et $X_nY_n \overset{Proba}{\longrightarrow} XY$, si $(\alpha_n)_{n\geqslant 0}$ suite de nombres réels tels que $\alpha_n \to \alpha$, alors $X_n + \alpha_n Y_n \overset{Proba}{\longrightarrow} X + \alpha Y$, etc....

Démonstration. On utilise l'uniforme continuité de f sur les compacts, notamment sur les boules de type $\mathcal{B}(0, a)$. Soit $\varepsilon > 0$ et $a > 0$ fixés.
Il existe $\eta = \eta(\varepsilon, a) > 0$ tel que $\forall x, y$ avec $|x| \leqslant a$, $|x - y| \leqslant \eta$, on a $|f(x) - f(y)| \leqslant \varepsilon$. Par conséquent

$$\{|X| \leqslant a\} \cap \{|X_n - X| \leqslant \eta\} \subset \{|f(X_n) - f(X)| \leqslant \varepsilon\}.$$

Par passage au complémentaire, on a

$$\{|f(X_n) - f(X)| > \varepsilon\} \subset \{|X| > a\} \cup \{|X_n - X| > \eta\}.$$

Par conséquent

$$\limsup_{n\to\infty} \mathbb{P}(|f(X_n) - f(X)| > \varepsilon) \leqslant \mathbb{P}(|X| > a) + \lim_{n\to\infty} \mathbb{P}(|X_n - X| > \eta) = \mathbb{P}(|X| > a).$$

Comme X est une va à valeurs dans $\mathbb{R}^d$ (finie), par convergence monotone $\lim_{a\to\infty} \mathbb{P}(|X| > a) = 0$ et donc $\mathbb{P}(|f(X_n) - f(X)| > \varepsilon) \to 0$ quand $n \to \infty$. □

Proposition 93 (Convergence ps ou L^1 vers convergence en probabilité)**.**
Si $(X_n)_{n\geqslant 0}$ converge vers X p.s ou dans L^1 alors $(X_n)_{n\geqslant 0}$ converge en probabilité vers X.

Démonstration. Soit $\varepsilon > 0$.
Supposons que $X_n \xrightarrow{p.s} X$. On a

$$\mathbb{P}(|X_n - X| > \varepsilon) = \mathbb{E}\big[\mathbb{1}_{|X_n - X| > \varepsilon}\big].$$

La variable $\mathbb{1}_{|X_n - X| > \varepsilon}$ est dominée par 1 et converge ps. vers 0, on conclut par convergence dominée.
Supposons que $X_n \xrightarrow{L^1} X$. Alors par l'inégalité de Markov,

$$\mathbb{P}(|X_n - X| > \varepsilon) \leqslant \frac{\mathbb{E}[|X_n - X|]}{\varepsilon},$$

qui tend vers 0 quand $n \to +\infty$. □

Exemple. Soit $(X_n)_{n\geqslant 0}$ une suite de v.a. indépendantes de loi

$$\mathbb{P}(X_n = x_n) = p_n, \quad \mathbb{P}(X_n = 0) = 1 - p_n,$$

avec $p_n \in]0,1[$ et $x_n \geqslant 1$. On a $\mathbb{P}(|X_n| > \varepsilon) = p_n \; \forall \varepsilon \in]0,1[$.
Pour $p_n = 1/n$ et $x_n = n$, on a $X_n \xrightarrow{Proba} 0$ mais X_n ne converge ni ps ni dans L^1.

Proposition 94 (Convergence en probabilité vers convergence ps)**.**
Si $(X_n)_{n\geqslant 0}$ converge en probabilité vers X alors il existe une sous-suite $(X_{n_k})_{k\geqslant 0}$ qui converge ps vers X.

Démonstration. Pour tout $k \geqslant n$, on a $\lim_{n\to\infty} \mathbb{P}(|X_n - X| > 2^{-(k+1)}) = 0$. Il existe donc $n_k \in \mathbb{N}$ tel que pour tout $n \geqslant n_k$, $\mathbb{P}(|X_n - X| > 2^{-(k+1)}) \leqslant 2^{-k+1}$.
On peut supposer que la suite est strictement croissante, car si n_k convient, alors $n_k + 1$ convient aussi.
Par conséquent, pour tout $k \in \mathbb{N}$,

$$\mathbb{P}\Big(|X_{n_{k+1}} - X_{n_k}| > 2^{-k}\Big) \leqslant \mathbb{P}\Big(|X_{n_{k+1}} - X| > 2^{-(k+1)}\Big) + \mathbb{P}\Big(|X_{n_k} - X| > 2^{-(k+1)}\Big) \leqslant 2^{-k}.$$

La série $\sum \mathbb{P}\big(|X_{n_{k+1}} - X_{n_k}| > 2^{-k}\big)$ converge, donc d'après la Proposition 85, la suite $(X_{n_k})_{k\geqslant 0}$ converge ps.
On note Y sa limite. Comme $X_{n_k} \xrightarrow{p.s} Y$ on a $X_{n_k} \xrightarrow{Proba} Y$. Or $X_{n_k} \xrightarrow{Proba} X$ car sous-suite de $(X_n)_{n\geqslant 0}$.
Par conséquent, $Y = X$ ps. □

Remarque. $X_n \xrightarrow{Proba} X$ *si et seulement si de toute sous-suite on peut extraire une sous-suite qui converge ps vers X.*

Démonstration. Si $X_n \xrightarrow{Proba} X$, il suffit d'utiliser la proposition précédente. Réciproquement, on raisonne par l'absurde. Si X_n ne converge pas en proba vers X, alors il existe $\varepsilon > 0$ et $\alpha > 0$ tel que $\forall n \geqslant 0$ $\exists n_k > n$ tel que $\mathbb{P}(|X_{n_k} - X| > \varepsilon) \geqslant \alpha$. On construit ainsi une suite strictement croissante d'entiers telle que $\mathbb{P}(|X_{n_k} - X| > \varepsilon) \geqslant \alpha$. On peut alors extraire une sous-suite $X_{n'_k}$ qui converge ps vers X et donc en probabilité. Mais c'est incompatible avec $\forall k \in \mathbb{N}$ $\mathbb{P}(|X_{n_k} - X| > \varepsilon) \geqslant \alpha$. □

1.4 Uniforme intégrabilité

Definition 95 (Uniforme intégrabilité)**.** *Une suite de v.a. $(X_n)_{n\geqslant 1}$ intégrables est dite uniformément intégrable (ou équi-intégrable) si*

$$\lim_{a\to\infty}\sup_{n\geqslant 1}\mathbb{E}\big[|X_n|\mathbb{1}_{|X_n|>a}\big]=0.$$

Remarque.
- *Une suite $(X_n)_{n\geqslant 1}$ bornée par une v.a positive Z intégrable est uniformément intégrable.*
- *Un $n-$uplet de v.a. $(X_k)_{k\in\{1,\dots,n\}}$ intégrables est uniformément intégrable.*
- *Considérons deux suites $(X_n)_{n\in\mathbb{N}}$ et $(Y_n)_{n\in\mathbb{N}}$ avec $|X_n|\leqslant|Y_n|$ pour tout $n\in\mathbb{N}$. Si $(Y_n)_{n\in\mathbb{N}}$ est uniformément intégrable, alors $(X_n)_{n\in\mathbb{N}}$ est uniformément intégrable.*

Démonstration.
- Supposons que $|X_n|\leqslant Z$, alors $|X_n|\mathbb{1}_{|X_n|>a}\leqslant Z\mathbb{1}_{Z>a}$ et donc

$$\sup_{n\geqslant 0}\mathbb{E}\big[|X_n|\mathbb{1}_{|X_n|>a}\big]\leqslant\mathbb{E}[Z\mathbb{1}_{Z>a}].$$

  On conclut en utilisant le théorème de convergence dominée.
- Le n-uplet est dominée par la variable $Z=\sum_{k=1}^n|X_k|$ qui est intégrable.
- On utilise le fait que $|X_n|\mathbb{1}_{|X_n|>a}\leqslant|Y_n|\mathbb{1}_{|Y_n|>a}$ pour tout $n\in\mathbb{N}$.

□

Proposition 96. *Si $(X_n)_{n\geqslant 1}$ est une suite de variable de L^p telle qu'elle soit bornée dans $L^{p+\delta}$ pour un $\delta>0$, alors $(X_n^p)_{n\geqslant 1}$ est uniforméement intégrable.*

Démonstration. On remarque que $|X_n|^p\mathbb{1}_{|X_n|^p>a}\leqslant|X_n|^p\frac{|X_n|^\delta}{a^{\delta/p}}\mathbb{1}_{|X_n|^p>a}\leqslant\frac{1}{a^{\delta/p}}|X_n|^{p+\delta}$. Par conséquent,

$$\sup_{n\in\mathbb{N}}\mathbb{E}\big[|X_n|^p\mathbb{1}_{|X_n|^p>a}\big]\leqslant\frac{1}{a^{\delta/p}}\sup_{n\in\mathbb{N}}\mathbb{E}\Big[|X_n|^{p+\delta}\Big],$$

qui converge bien vers 0 quand $a\to\infty$. □

L'hypothèse borné dans L^1 ne suffit pas pour avoir $(X_n)_{n\geqslant 1}$ uniformément intégrable (voir l'exemple où $X_n\xrightarrow{Proba}0$ et $X_n\overset{L^1}{\nrightarrow}0$).
On remarque que si $X\in L^1$, alors $\forall\varepsilon>0$, $\exists\eta>0$ tel que si $A\in\mathcal{F}$ avec $\mathbb{P}(A)<\eta$ alors $\mathbb{E}[|X|\mathbb{1}_A]<\varepsilon$. En effet, par convergence dominée, à $\varepsilon>0$ fixé, on a pour a assez grand $\mathbb{E}\big[|X|\mathbb{1}_{|X|>a}\big]<\varepsilon/2$ et donc

$$\begin{aligned}\mathbb{E}[|X|\mathbb{1}_A]&=\mathbb{E}\big[|X|\mathbb{1}_{A\cap\{|X|\leqslant a\}}\big]+\mathbb{E}\big[|X|\mathbb{1}_{A\cap\{|X|>a\}}\big]\\&\leqslant a\mathbb{P}(A)+\varepsilon/2.\end{aligned}$$

Par conséquent, pour tout $A\in\mathcal{F}$ avec $\mathbb{P}(A)<\varepsilon/(2a)$, on a $\mathbb{E}[|X|\mathbb{1}_A])<\varepsilon$. Ce résultat est généralisé pour les suites uniformément intégrables.

Proposition 97 (Critère d'uniforme intégrabilité)**.** *Une suite de v.a. $(X_n)_{n\geqslant 1}$ est uniformémement intégrable ssi*

- $\forall\varepsilon>0$, $\exists\eta>0$ *tel que pour $A\in\mathcal{F}$ avec $\mathbb{P}(A)<\eta$ on a $\mathbb{E}[|X_n|\mathbb{1}_A]<\varepsilon$ pour tout $n\geqslant 1$*
et
- $\sup_{n\geqslant 1}\mathbb{E}[|X_n|]<\infty$.

Démonstration. $\Rightarrow$] Supposons $(X_n)_{n\geqslant 1}$ est uniformémement intégrable.
Soit $\varepsilon > 0$, il existe $a > 0$ tel que $\sup_{n\geqslant 1} \mathbb{E}\big[|X_n|\mathbb{1}_{|X_n|>a}\big] < \varepsilon/2$.
Soit $A \in \mathcal{F}$, on a pour tout $n \geqslant 1$

$$\begin{aligned}\mathbb{E}[|X_n|\mathbb{1}_A] &= \mathbb{E}\big[|X_n|\mathbb{1}_{A\cap\{|X_n|\leqslant a\}}\big] + \mathbb{E}\big[|X_n|\mathbb{1}_{A\cap\{|X_n|>a\}}\big] \\ &\leqslant a\mathbb{P}(A) + \varepsilon/2.\end{aligned}$$

Il suffit de choisir $\eta = \varepsilon/(2a)$ et la suite est bien bornée dans L^1 si on considère $A = \Omega$.
[$\Leftarrow$ Soit $\varepsilon > 0$ et η tel que pour tout $A \in \mathcal{F}$ avec $\mathbb{P}(A) < \eta$ on a $\mathbb{E}[|X_n|\mathbb{1}_A] < \varepsilon$ $\forall n \geqslant 1$.
On pose $M = \sup_{n\geqslant 1} \mathbb{E}[|X_n|]$. D'après l'inégalité de Markov, on a

$$\mathbb{P}(|X_n| > a) \leqslant \frac{M}{a}.$$

Par conséquent, pour a assez grand on a $\mathbb{P}(|X_n| > a) < \eta$ et donc pour tout $n \geqslant 1$

$$\mathbb{E}\big[|X_n|\mathbb{1}_{|X_n|>a}\big] < \varepsilon.$$

La suite est bien uniformément intégrable.

□

Proposition 98 (Critère de Vitali)**.** *Soit $(X_n)_{n\geqslant 1}$ une suite de v.a. intégrables. Il y a équivalence entre*

1. $(X_n)_{n\geqslant 1}$ *converge dans* L^1,
2. $(X_n)_{n\geqslant 1}$ *est uniformément intégrable et* $(X_n)_{n\geqslant 1}$ *converge en probabilité.*

Démonstration. 1. $\Rightarrow$ 2. On sait déjà que la convergence dans L^1 implique la convergence en proba. On va utiliser la proposition précédente pour montrer que la suite est uniformément intégrable. Notons X_0 sa limite qui est dans L^1. Soit $\varepsilon > 0$, il existe $n_0 \geqslant 1$ tel que pour tous $n \geqslant n_0$, $\mathbb{E}[|X_n - X_0|] < \varepsilon/2$.
Soit $A \in \mathcal{F}$, alors pour $n \geqslant n_0$, on a

$$\begin{aligned}\mathbb{E}[|X_n|\mathbb{1}_A] &\leqslant \mathbb{E}[|X_n - X_0|\mathbb{1}_A] + \mathbb{E}[|X_0|\mathbb{1}_A] \\ &\leqslant \frac{\varepsilon}{2} + \mathbb{E}[|X_0|\mathbb{1}_A]\end{aligned}$$

Par conséquent, $\sup_{n\geqslant n_0} \mathbb{E}[|X_n|\mathbb{1}_A] \leqslant \frac{\varepsilon}{2} + \mathbb{E}[|X_0|\mathbb{1}_A]$ et donc

$$\sup_{n\geqslant 1} \mathbb{E}[|X_n|\mathbb{1}_A] \leqslant \frac{\varepsilon}{2} + \sup_{0\leqslant k\leqslant n_0} \mathbb{E}[|X_k|\mathbb{1}_A]$$

En prenant $A = \Omega$, on obtient $\sup_{n\geqslant 1} \mathbb{E}[|X_n|] < \infty$ et comme $(|X_k|)_{0\leqslant k\leqslant n_0}$ est uniformément intégrable, il existe $\eta > 0$ tel que si $\mathbb{P}(A) < \eta$ alors $\sup_{0\leqslant k\leqslant n_0} \mathbb{E}[|X_k|\mathbb{1}_A] < \varepsilon/2$. On en déduit que (X_n) est uniformément intégrable.
2. $\Rightarrow$ 1. Comme (X_n) converge en probabilité vers X, on peut extraire une sous-suite (n') qui converge p.s.. D'après le lemme de Fatou, on a

$$\mathbb{E}[|X|] = \mathbb{E}[\liminf |X_{n'}|] \leqslant \liminf \mathbb{E}[|X_{n'}|] \leqslant \sup_n \mathbb{E}[|X_n|] < \infty.$$

Donc $X \in L^1$.
Soit $\varepsilon > 0$. On a

$$\begin{aligned}\mathbb{E}[|X_n - X|] &\leqslant \varepsilon + \mathbb{E}\big[|X_n - X|\mathbb{1}_{|X_n-X|>\varepsilon}\big] \\ &\leqslant \varepsilon + \mathbb{E}\big[|X_n|\mathbb{1}_{|X_n-X|>\varepsilon}\big] + \mathbb{E}\big[|X|\mathbb{1}_{|X_n-X|>\varepsilon}\big]\end{aligned}$$

Comme $(X, X_n)_{n\geqslant 1}$ est uniformément intégrable, il existe $\eta > 0$ tel que si $\mathbb{P}(A) < \eta$, on a

$$\mathbb{E}[|X_n|\mathbb{1}_A] < \varepsilon \quad \mathbb{E}[|X|\mathbb{1}_A] < \varepsilon.$$

Comme la suite converge en probabilité, il existe $n_0 \geqslant 1$ tel que pour tout $n \geqslant n_0$, $\mathbb{P}(|X_n - X| > \varepsilon) < \eta$. Par conséquent pour tout $n \geqslant n_0$, $\mathbb{E}[|X_n - X|] < 3\varepsilon$. La suite converge bien dans L^1. □

1.5 Critères de convergence

Les critères de type Cauchy sont parfois bien utiles pour établir la convergence d'une suite sans avoir à expliciter la valeur de la limite.

Proposition 99. *Soit $(X_n)_{n\geqslant 1}$ une suite de v.a. à valeurs dans $\mathbb{R}^d$. Alors*

1. *$(X_n)_{n\geqslant 1}$ converge presque sûrement ssi $\forall\varepsilon > 0$,*

$$\lim_{n\to\infty} \mathbb{P}\left(\sup_{k\geqslant 1} |X_{n+k} - X_n| > \varepsilon\right) = 0.$$

2. *$(X_n)_{n\geqslant 1}$ converge en probabilité ssi $\forall\varepsilon > 0$,*

$$\lim_{n\to\infty} \sup_{k\geqslant 1} \mathbb{P}(|X_{n+k} - X_n| > \varepsilon) = 0.$$

3. *$(X_n)_{n\geqslant 1}$ une famille de v.a. de L^p converge dans L^p ssi*

$$\lim_{n\to\infty} \sup_{k\geqslant 1} \mathbb{E}(|X_{n+k} - X_n|^p) = 0.$$

Démonstration. 3. Le dernier point est juste le théorème de Riesz-Fisher : l'espace L^p est complet (voir cours d'intégration).

1. Regardons la convergence p.s.. Si $(X_n)_{n\geqslant 1}$ converge p.s., la suite est ps. de Cauchy. Par conséquent, $\sup_{k\geqslant 1} |X_{n+k} - X_n|$ converge vers 0 p.s. quand $n \to +\infty$ et donc en probabilité, d'où le résultat. Réciproquement, on pose $Y_n = \sup_{p,q\geqslant n} |X_p - X_q|$ qui est décroissante en n et minorée par 0, donc converge ps. vers une v.a. notée Y. Par ailleurs d'après l'inégalité triangulaire $Y_n \leqslant 2\sup_{k\geqslant 1} |X_{n+k} - X_n|$ ce qui implique que $(Y_n)_{n\geqslant 1}$ converge en probabilité vers 0 et donc $Y = 0$ ps. La suite $(X_n)_{n\geqslant 1}$ est par conséquent p.s. de Cauchy et donc converge p.s. par complétude de $\mathbb{R}^d$.

2. Regardons maintenant la convergence en probabilité. Si $(X_n)_{n\geqslant 1}$ converge en probabilité, comme $|X_{n+k} - X_n| \leqslant |X_{n+k} - X| + |X_n - X|$, on a

$$\mathbb{P}(|X_{n+k} - X_n| > \varepsilon) \leqslant \mathbb{P}(|X_{n+k} - X| > \varepsilon/2) + \mathbb{P}(|X_n - X| > \varepsilon/2) \leqslant 2\sup_{k\geqslant n} \mathbb{P}(|X_k - X| > \varepsilon/2)$$

Le majorant convergeant vers 0, on a la première implication. Supposons la réciproque, on construit une suite strictement croissante (n_q) d'entiers telle que

$$\sup_{k\geqslant 1} \mathbb{P}\big(|X_{n_q+k} - X_{n_q}| > 2^{-q}\big) \leqslant 2^{-q}.$$

En particulier, $\mathbb{P}\big(|X_{n_q+1} - X_{n_q}| > 2^{-q}\big) \leqslant 2^{-q}$ et par conséquent la sous-suite $\big(X_{n_q}\big)$ converge p.s. et donc en probabilité vers X. On a alors,

$$\mathbb{P}(|X_q - X| > \varepsilon) \leqslant \mathbb{P}\big(|X_q - X_{n_q}| > \varepsilon/2\big) + \mathbb{P}\big(|X_{n_q} - X| > \varepsilon/2\big) \leqslant \sup_{k\geqslant 1} \mathbb{P}(|X_{q+k} - X_q| > \varepsilon/2) + \mathbb{P}\big(|X_{n_q} - X| > \varepsilon/2\big).$$

Les deux termes de droite convergent vers 0 et donc la suite converge en probabilité.

□

Remarque. *La convergence en probabilité est métrisable par la métrique*

$$d_{\mathbb{P}}(X,Y) = \mathbb{E}[\min(|X-Y|,1)].$$

mais la convergence ps n'est pas métrisable.

2 Convergence étroite et convergence en loi

La convergence en loi d'une suite de variables aléatoires $(X_n)_{n\geqslant 0}$ correspond à la convergence étroite des mesures de probabilité $\mathbb{P}_{X_n}$.

2.1 Convergence étroite

Soit $(\mu_n)_{n\in\mathbb{N}}$ une suite de probabilité sur $\mathbb{R}^d$ et μ une probabilité sur $\mathbb{R}^d$.
On note $\mathcal{C}_b$ l'ensemble des fonctions continues bornées $\mathbb{R}^d \to \mathbb{R}$ et $\mathcal{C}_c$ l'ensemble des fonctions continues à support compact $\mathbb{R}^d \to \mathbb{R}$.

Definition 100. *La suite de mesures de probabilité $(\mu_n)_{n\in\mathbb{N}}$ converge étroitement vers μ si*

$$\forall f \in \mathcal{C}_b, \quad \lim_{n\to\infty} \int_{\mathbb{R}^d} f(x)\mu_n(dx) = \int_{\mathbb{R}^d} f(x)\mu(dx).$$

Proposition 101. *Soit $\mathcal{H} \subset \mathcal{C}_b$ tel que $\mathcal{C}_c \subset \overline{\mathcal{H}}$ (adhérence pour la norme $||.||_\infty$).*
Une suite de mesures de probabilité $(\mu_n)_{n\in\mathbb{N}}$ converge étroitement vers μ si et seulement si

$$\forall f \in \mathcal{H}, \quad \lim_{n\to\infty} \int_{\mathbb{R}^d} f(x)\mu_n(dx) = \int_{\mathbb{R}^d} f(x)\mu(dx).$$

Démonstration. Si $\mathcal{H} = \mathcal{C}_c$. On utilise un argument de troncature. Pour $p > 0$, on introduit la fonction θ_p linéaire par morceaux telle que $\theta_p : \mathbb{R}^+ \to [0,1]$, avec $\theta_p = 1$ sur $[0,p]$ et $\theta_p = 0$ sur $[2p, +\infty[$.
Soit $f \in \mathcal{C}_b$. On a $f(x) = f(x)\theta_p(|x|) + f(x)(1-\theta_p(|x|))$.
Par conséquent

$$\int f d\mu_n = \int f\theta_p d\mu_n + \int f(1-\theta_p)d\mu_n.$$

Par ailleurs,

$$\int f(1-\theta_p)d\mu_n \leqslant ||f||_\infty \int (1-\theta_p(|x|)\mu_n(dx)).$$

On a donc

$$\left|\int f d\mu_n - \int f d\mu\right| \leqslant \left|\int f\theta_p d\mu_n - \int f\theta_p d\mu\right| + ||f||_\infty \left(2 - \int \theta_p(|x|)\mu_n(dx) - \int \theta_p(|x|)\mu(dx)\right).$$

Comme $f\theta_p \in \mathcal{C}_c$, le premier terme converge vers 0 quand $n\to\infty$.
Par ailleurs, $\theta_p \in \mathcal{C}_c$, donc $\int \theta_p(|x|)\mu_n(dx)$ converge vers $\int \theta_p(|x|)\mu(dx)$ quand $n\to\infty$.
Par conséquent,

$$\limsup_{n\to\infty} \left|\int f d\mu_n - \int f d\mu\right| \leqslant 2||f||_\infty \left(1 - \int \theta_p(|x|)\mu(dx)\right).$$

Or, pour tout $x\in\mathbb{R}^d$, $\theta_p(|x|) \to 1$ quand $p\to\infty$ et $\theta_p(|x|) \leqslant 1$, donc par convergence dominée, en faisant tendre $p \to +\infty$, on obtient

$$\lim_{n\to\infty} \left|\int f d\mu_n - \int f d\mu\right| = 0.$$

Cas général : Supposons que

$$\forall h \in \mathcal{H} \quad \lim_{n\to\infty} \int_{\mathbb{R}^d} h(x)\mu_n(dx) = \int_{\mathbb{R}^d} h(x)\mu(dx).$$

Soit $f \in \mathcal{C}_c$ et $\varepsilon > 0$. Il existe $h_\varepsilon \in \mathcal{H}$ tel que $||f - h_\varepsilon||_\infty \leqslant \varepsilon$.
On a $f = h_\varepsilon + f - h_\varepsilon$ et donc

$$\left|\int f d\mu_n - \int f d\mu\right| \leqslant \left|\int h_\varepsilon d\mu_n - \int h_\varepsilon d\mu\right| + 2||f - h_\varepsilon||_\infty$$

et donc, pour tout $\varepsilon > 0$,

$$\limsup_{n\to\infty} \left|\int f d\mu_n - \int f d\mu\right| \leqslant 2\varepsilon.$$

CQFD. □

Remarque. *La continuité de f est essentielle. Si $(\mu_n)_{n\in\mathbb{N}}$ converge étroitement vers μ, en général, on n'a pas $\mu_n(A) \underset{n\to\infty}{\longrightarrow} \mu(A)$ pour $A \in \mathcal{B}(\mathbb{R}^d)$.*
Par exemple, $\mu_n = \delta_{1/n}$ converge étroitement vers δ_0, car $\int f\delta_{1/n} = f(1/n) \to f(0)$ par continuité de f. Par ailleurs, $\delta_{1/n}(\{0\}) = 0$ pour tout $n \geqslant 1$ et $\delta_0(\{0\}) = 1$.

Theorem 102. *On a équivalence entre*

1. *$(\mu_n)_{n\in\mathbb{N}}$ converge étroitement vers μ,*
2. *Pour tout fermé F, $\limsup \mu_n(F) \leqslant \mu(F)$,*
3. *Pour tout ouvert O, $\mu(O) \leqslant \liminf \mu_n(O)$,*
4. *Pour tout borélien $B \in \mathcal{B}(\mathbb{R}^d)$ tel que $\mu\left(\overline{B} \setminus \mathring{B}\right) = 0$,*
$$\lim_{n\to\infty} \mu_n(B) = \mu(B).$$
5. *Pour toute fonction f mesurable bornée tel que $\mu(D_f) = 0$ où D_f ensemble des points de discontinuité de f,*
$$\lim_{n\to\infty} \int f d\mu_n = \int f d\mu.$$

Démonstration. $1 \Rightarrow 2$ Soit F un fermé de $\mathbb{R}^d$. Pour $k \geqslant 1$, on considère les fonctions

$$f_k(x) = \frac{1}{(1 + d(x, F))^k}$$

Ces fonctions sont continues, bornées par 1. La suite décroît vers $\mathbb{1}_F(x)$. Par conséquent, pour $k \geqslant 1$,

$$\limsup_{n\to\infty} \mu_n(F) = \limsup_{n\to\infty} \int \mathbb{1}_F d\mu_n \leqslant \limsup_{n\to\infty} \int f_k d\mu_n = \int f_k d\mu.$$

Par ailleurs, par convergence dominée, $\int f_k d\mu$ converge vers $\mu(F)$ quand $k \to \infty$. On en déduit que $\limsup \mu_n(F) \leqslant \mu(F)$.

$2 \Leftrightarrow 3$ Par passage au complémentaire.

$2+3 \Rightarrow 4$ On a $\mathring{B} \subset B \subset \overline{B}$, alors

$$\mu(\mathring{B}) \quad \leqslant \quad \liminf \mu_n(\mathring{B}) \quad \leqslant \quad \liminf \mu_n(B) \quad \leqslant \quad \limsup \mu_n(B) \quad \leqslant \quad \limsup \mu_n(\overline{B}) \quad \leqslant \quad \mu(\overline{B})$$

Dans le cas où $\mu\Big(\overline{B} \setminus \mathring{B}\Big) = 0$, on en déduit que $\mu(B) = \liminf \mu_n(B) = \limsup \mu_n(B)$.

$5 \Rightarrow 1$ Si f est continue bornée alors $D_f = \emptyset$, ok.

$4 \Rightarrow 5$ Soit $f : \mathbb{R}^d \to \mathbb{R}$ mesurable bornée. Alors l'ensemble $\big\{t \in \mathbb{R} : \mu\big(\{x \in \mathbb{R}^d : f(x) = t\}\big) > 0\big\}$ est un ensemble au plus dénombrable, car c'est l'ensemble des points de discontinuité de la fonction de répartition de la v.a. f définie sur $\big(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \mu\big)$. Son complémentaire, noté D, est donc dense dans $\mathbb{R}^d$.

Soit $M > 0$ tel que $||f||_\infty \leqslant M$. Soit $\varepsilon > 0$. Il existe $t_1 < t_2 < \ldots < t_k$ de D tels que $t_1 < -M$ et $t_k > M$ et $\max_{i \leqslant k} |t_i - t_{i-1}| \leqslant \varepsilon$. On pose

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

Alors pour tout $x \in \mathbb{R}^d$, $|f(x) - g(x)| \leqslant \max_{i \leqslant k} |t_i - t_{i-1}| \leqslant \varepsilon$. On a donc,

$$\left| \int f d\mu_n - \int f d\mu \right| \leqslant \left| \int g d\mu_n - \int g d\mu \right| + 2\varepsilon.$$

Or

$$\int g d\mu_n = \sum_{i=1}^{k-1} t_i \mu_n \Big(\Big\{ x \in \mathbb{R}^d : f(x) \in [t_i, t_{i+1}[\Big\} \Big)$$

et

$$\int g d\mu = \sum_{i=1}^{k-1} t_i \mu \Big(\Big\{ x \in \mathbb{R}^d : f(x) \in [t_i, t_{i+1}[\Big\} \Big).$$

Soit $B = \big\{x \in \mathbb{R}^d : f(x) \in [t_i, t_{i+1}[\big\}$. Alors

$$\begin{aligned} &\big\{x \in D_f^c : t_i < f(x) < t_{i+1}\big\} \subset \mathring{B} \qquad \text{et} \\ &\overline{B} \subset \big\{x \in D_f^c : t_i \leqslant f(x) \leqslant t_{i+1}\big\} \cup D_f \end{aligned}$$

Par conséquent, $\overline{B} \setminus \mathring{B} \subset \bigcup_{i=1}^k \big\{x \in \mathbb{R}^d : f(x) = t_i\big\} \cup D_f$. Comme $t_i \in D^c$ et $\mu(D_f) = 0$, l'ensemble est négligeable ce qui permet de conclure.

□

2.2 Convergence en loi

Soit $(X_n)_{n \geqslant 0}$ une suite de v.a et X une v.a. à valeurs dans $\mathbb{R}^d$

Definition 103. *La suite $(X_n)_{n \geqslant 0}$ converge en loi vers X si $(\mathbb{P}_{X_n})_{n \geqslant 0}$ converge étroitement vers $\mathbb{P}_X$. On note alors $X_n \xrightarrow{Loi} X$.*

Propriété 104. *Par conséquent, on a $X_n \xrightarrow{Loi} X$*
ssi $\forall f \in \mathcal{C}_b$, $\mathbb{E}[f(X_n)] \underset{n \to \infty}{\longrightarrow} \mathbb{E}[f(X)]$,

ssi $\forall f \in \mathcal{H}$, *avec* $\mathcal{H} \subset \mathcal{C}_b$ *et* $\mathcal{C}_c \subset \overline{\mathcal{H}}$, $\mathbb{E}[f(X_n)] \underset{n\to\infty}{\longrightarrow} \mathbb{E}[f(X)]$,
ssi pour tout fermé F, $\limsup \mathbb{P}(X_n \in F) \leqslant \mathbb{P}(X \in F)$,
ssi pour tout ouvert O, $\mathbb{P}(X \in O) \leqslant \liminf \mathbb{P}(X_n \in O)$,
ssi pour tout borélien $B \in \mathcal{B}(\mathbb{R}^d)$ *tel que* $\mathbb{P}\Big(X \in \overline{B} \setminus \mathring{B}\Big) = 0$,

$$\lim_{n\to\infty} \mathbb{P}(X_n \in B) = \mathbb{P}(X \in B)$$

ssi $\forall t \in \mathbb{R}$ *point de continuité de* F_X, $\lim_{n\to\infty} F_{X_n}(t) = F_X(t)$.

Démonstration. Il suffit de vérifier la dernière assertion, pour cela on considère $B = \mathbb{1}_{]-\infty,t]}$. Si F_X est continue en t, alors $\mathbb{P}\Big(X \in \overline{B} \setminus \mathring{B}\Big) = \mathbb{P}(X = t) = 0$. La convergence en loi implique donc la convergence des fonctions de répartition au point t.
Réciproquement, on considère $f \in \mathcal{C}_c^\infty$, alors pour tout $x \in \mathbb{R}$,

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

Par Fubini (vu que f est à support compact), on a

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

Comme l'ensemble des points de discontinuité de F_X est au plus dénombrable, $F_{X_n} \to F_X$ $\lambda-$pp sur $\mathbb{R}$. Par le théorème de convergence dominée (en effet, $|f'(t)(1 - F_{X_n}(t))| \leqslant |f'(t)| \in L^1(\mathbb{R}, dt)$ car $f' \in \mathcal{C}_c^\infty$),

$$\lim_{n\to\infty} \mathbb{E}[f(X_n)] = \int_{\mathbb{R}} f'(t) \lim_{n\to\infty} (1 - F_{X_n}(t))dt = \mathbb{E}[f(X)]$$

On conclut en utilisant le fait que $\mathcal{C}_c^\infty$ est dense dans $\mathcal{C}_c$. □

Remarque. *On peut définir la convergence en loi de* $(X_n)_{n\geqslant 0}$ *vers* X *sans que les variables soient définies sur le même espace de probabilité.*

Proposition 105. *Si* $X_n \xrightarrow{Loi} X$ *et* $g : \mathbb{R}^d \to \mathbb{R}^q$ *une fonction continue, alors* $g(X_n) \xrightarrow{Loi} g(X)$.

Démonstration. Soit $f : \mathbb{R}^q \to \mathbb{R}$ une fonction continue bornée, alors $f \circ g : \mathbb{R}^d \to \mathbb{R}$ est continue bornée. Par ailleurs, $\mathbb{E}[f(g(X_n))] = \mathbb{E}[f \circ g(X_n)]$. On conclut en utilisant le fait que $X_n \xrightarrow{Loi} X$. □

Corollary 106. *Si* $X_n = (X_n^1, \ldots, X_n^d) \xrightarrow{Loi} X = (X^1, \ldots, X^d)$, *alors pour tout* $k \in \{1, \ldots, d\}$, $X_n^k \xrightarrow{Loi} X^k$.

Attention : La réciproque est fausse !

Exemple. Soit X de loi $\mathbb{P}(X = 1) = \mathbb{P}(X = -1) = 1/2$. On pose $X_n = X$ si n est pair et $X_n = -X$ si n impair et $Y_n = X$ pour tout n. Comme $X \overset{Loi}{=} -X$, on a $X_n \overset{Loi}{=} X$ et $Y_n \overset{Loi}{=} X$ pour tout $n \geqslant 1$, et donc $X_n \xrightarrow{Loi} X$ et $Y_n \xrightarrow{Loi} X$. Par contre (X_n, Y_n) ne converge pas en loi vers (X, X), car sinon $X_n + Y_n$ convergerait en loi vers $2X$. Or $X_n + Y_n = 2X$ si n est pair et 0 si n impair (de même $X_nY_n = 1$ ou -1 selon la parité de n).

Theorem 107. *La convergence loi* $X_n \xrightarrow{Loi} X$ *est équivalente à la convergence simple des fonctions caractéristiques.*

Démonstration. (en dimension 1.)
Si $X_n \xrightarrow{Loi} X$. Comme $f_t = e^{itx}$ est $\mathcal{C}_b$, on a de façon évidente la convergence simple des fonctions caractéristiques.
Si les fonctions caractéristiques convergent simplement. Soit $f : \mathbb{R} \to \mathbb{R}$ une fonction $\mathcal{C}_c^\infty$. Alors sa transformée de Fourier $\hat{f}(t) = \int_{\mathbb{R}} e^{itx} f(x)dx \in L^1(\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$ car $\widehat{f''}(t) = \int e^{itx} f''(x)dx = -t^2 \hat{f}(t)$ (par I.P.P. du fait que la fonction est à support compact) et $||\widehat{f''}||_\infty \leqslant ||f''||_1$, ce qui implique $\hat{f}(t) \underset{\infty}{=} O(1/t^2) \in L^1(\mathbb{R}, dx)$.
Par le théorème d'inversion de Fourier,

$$f(x) = \frac{1}{2\pi} \int e^{-itx} \hat{f}(t) dt.$$

Par conséquent, par Fubini,

$$\mathbb{E}[f(X_n)] = \frac{1}{2\pi} \int \mathbb{E}[e^{-itX_n}] \hat{f}(t) dt$$

et on conclut par convergence dominée. □

Theorem 108 (Paul Lévy)**.** *On considère une suite de fonctions caractéristiques* $(\varphi_n)_{n \geqslant 0}$ *de v.a.* X_n *réelles. On suppose que pour tout* $t \in \mathbb{R}$, $\lim_{n\to\infty} \varphi_n(t) = \varphi(t)$ *où* φ *est une fonction continue en* $t = 0$. *Alors il existe une v.a.* X *telle que* φ *soit sa fonction caractéristique et* $X_n \xrightarrow{Loi} X$.

2.3 Liens avec les autres convergences

Proposition 109. *Si* $X_n \xrightarrow{Proba} X$ *alors* $X_n \xrightarrow{Loi} X$.

Attention, la réciproque de la proposition est fausse. Reprenons l'exemple précédent. Soit X de loi $\mathbb{P}(X = 1) = \mathbb{P}(X = -1) = 1/2$ et considérons la suite $X_n = X$ si n est pair et $X_n = -X$ si n impair. Alors $\mathbb{P}(|X_{n+1} - X_n| > \varepsilon) = \mathbb{P}(2|X| > \varepsilon) = 1$ pour tout $\varepsilon < 1$.

Démonstration. Soit $t \in \mathbb{R}^d$. On a

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

En effet, $\left|e^{ix} - 1\right| \leqslant 2$ et $\left|e^{ix} - 1\right| = \left|i \int_0^x e^{is} ds\right| \leqslant |x|$.
Soit $\varepsilon > 0$, comme $1 = \mathbb{1}_{|X_n - X| \leqslant \varepsilon} + \mathbb{1}_{|X_n - X| > \varepsilon}$, on a

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

Par conséquent, $\limsup_{n\to\infty} |\varphi_{X_n}(t) - \varphi_X(t)| \leqslant \varepsilon |t|$ pour tout $\varepsilon > 0$. Il suffit de faire tendre ε vers 0 pour conclure. □

Proposition 110. *Soit* $c \in \mathbb{R}^d$. *Supposons que les variables* X_n *sont construites sur le même espace de probabilité et que* $X_n \xrightarrow{Loi} c$. *Alors* $X_n \xrightarrow{Proba} c$.

Démonstration. (en dimension 1). Soit $\varepsilon > 0$.

$$\begin{aligned} \mathbb{P}(|X_n - c| > \varepsilon) &= \mathbb{P}(X_n < c - \varepsilon) + \mathbb{P}(X_n > c + \varepsilon) \\ &\leqslant F_{X_n}(c - \varepsilon) + 1 - F_{X_n}(c + \varepsilon). \end{aligned}$$

Comme $\varepsilon > 0$, la fonction de répartition F_c de la constante c est continue en $c+\varepsilon$ et $c-\varepsilon$. Comme $X_n \xrightarrow{Loi} c$, alors $F_{X_n}(c-\varepsilon) \to F_c(c-\varepsilon) = 0$ et $F_{X_n}(c=\varepsilon) \to F_c(c+\varepsilon) = 1$ quand $n \to \infty$. Donc $\lim_{n\to\infty} \mathbb{P}(|X_n - X| > \varepsilon) = 0$ pour tout $\varepsilon > 0$. □

Lemma 111 (Lemme de Slutsky)**.** *On considère deux suites $(X_n)_{n\geqslant 0}$ et $(Y_n)_{n\geqslant 0}$ de variables aléatoires avec*

$$X_n \xrightarrow{Loi} X \quad et \quad Y_n \xrightarrow{Proba} c.$$

Alors on a la convergence en loi du couple $(X_n, Y_n) \xrightarrow{Loi} (X, c)$.

Démonstration. La fonction caractéristique de (X,c) est $\varphi_{(X,c)}(t,s) = \varphi_X(t)e^{isc}$. Soit $\varepsilon > 0$. On a

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

Comme $Y_n \xrightarrow{Proba} c$ et $X_n \xrightarrow{Loi} X$, on a $\limsup \left|\varphi_{(X_n,Y_n)}(t,s) - \varphi_{(X,c)}(t,s)\right| \leqslant \varepsilon|s|$ pour tout $\varepsilon > 0$. On conclut en faisant $\varepsilon \to 0$. □

Conséquence 112. Si $X_n \xrightarrow{Loi} X$ et $X_n - Y_n \xrightarrow{Proba} 0$, alors $Y_n \xrightarrow{Loi} X$.
En effet, d'après le lemme précédent $(X_n, X_n - Y_n) \xrightarrow{Loi} (X, 0)$ et $Y_n = f(X_n, X_n - Y_n)$ avec $f(u,v) = u - v$ fonction continue.

Proposition 113 (Généralisation du Lemme de Fatou)**.** *Si $X_n \xrightarrow{Loi} X$, alors $\mathbb{E}[|X|] \leqslant \liminf_{n\to\infty} \mathbb{E}[|X_n|]$.*

Démonstration. Soit $k > 0$. Comme $x \mapsto \min(|x|, k)$ est continue bornée, on a

$$\begin{aligned}
\mathbb{E}[\min(|X|,k)] &= \lim_{n\to\infty} \mathbb{E}[\min(|X_n|,k)] = \liminf \mathbb{E}[\min(|X_n|,k)] \\
&\leqslant \liminf \mathbb{E}[|X_n|].
\end{aligned}$$

On conclut par le théorème de convergence monotone en faisant $k \to +\infty$. □

Exemple. Soit X_n des v.a. de loi géométrique de paramètre respectif $\mathcal{G}(\lambda/n)$ avec $\lambda > 0$. La suite X_n/n converge en loi vers la loi exponentielle $\mathcal{E}(\lambda)$.
En effet, la fonction caractéristique de X_n est

$$\varphi_{X_n}(t) = \frac{\frac{\lambda}{n}e^{it}}{1 - \left(1 - \frac{\lambda}{n}\right)e^{it}}.$$

Par conséquent,

$$\begin{aligned}
\varphi_{X_n/n}(t) = \varphi_{X_n}(t/n) &= \frac{\frac{\lambda}{n}e^{it/n}}{1 - \left(1 - \frac{\lambda}{n}\right)e^{it/n}} \\
&= \frac{\lambda}{n(e^{-it/n} - 1) + \lambda}.
\end{aligned}$$

Or

$$n(e^{-it/n} - 1) = n(\cos(t/n) - 1) - in\sin(t/n) \to -it \quad \text{quand } n \to \infty.$$

Par conséquent, $\varphi_{X_n/n}(t) \to \lambda/(\lambda - it)$. X_n/n converge vers la loi exponentielle quand $n \to \infty$.

3 Conclusion

$$
\begin{array}{ccccc}
 & & \text{Convergence dans } L^\infty & & \\
 & & \downarrow & & \\
 & & \text{Convergence dans } L^p & & \\
 & & \downarrow & & \\
 & & \text{Convergence dans } L^1 & & \\
 & & \downarrow & & \\
\text{Convergence p.s.} & \rightarrow & \text{Convergence en probabilité} & \rightarrow & \text{Convergence en loi}
\end{array}
$$

Chapitre 7

Théorèmes Limites

On se place sur un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$. On considère une suite de v.a. réelles $(X_n)_{n\geqslant 0}$ indépendantes et de même loi : i.i.d. (indépendantes et identiquement distribuées)

1 Loi des grands nombres : LGN

1.1 Version faible de la loi des grands nombres

Theorem 114. *Soit $(X_n)_{n\geqslant 0}$ une suite de v.a. i.i.d. dans L^1 de moyenne $m = \mathbb{E}[X_1]$. Alors*

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow[n\to\infty]{Proba} m.$$

Démonstration.
Cas L^2 : on suppose que les variables sont dans L^2.
On note $m = \mathbb{E}[X_1]$ et $\sigma^2 = Var(X_1)$ et on pose $M_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. On remarque que du fait que les variables sont indépendantes et de même loi, on a $\mathbb{E}[M_n] = m$ et $Var(M_n) = \sigma^2/n$.
Soit $\varepsilon > 0$. Alors par l'inégalité de Tchebychev, on a

$$\mathbb{P}(|M_n - m| > \varepsilon) \leqslant \frac{Var(M_n)}{\varepsilon^2} = \frac{1}{n}\frac{\sigma^2}{\epsilon^2} \underset{n\to\infty}{\longrightarrow} 0.$$

Cas L^1 : on suppose que les variables sont dans L^1.

Lemma 115. *Soit $(a_i)_{i\geqslant 0}$ et $(b_i)_{i\geqslant 0}$ deux suites de nombres complexes avec $|a_i| \leqslant 1$ et $|b_i| \leqslant 1$ pour tout $i \geqslant 0$, alors*

$$\left|\prod_{i=1}^{n} a_i - \prod_{i=1}^{n} b_i\right| \leqslant \sum_{i=1}^{n} |a_i - b_i|$$

Démonstration. Preuve par récurrence. Pour $n = 1$ c'est trivial. Supposons la propriété vraie pour le rang n, alors

$$\begin{aligned}\left|\prod_{i=1}^{n+1} a_i - \prod_{i=1}^{n+1} b_i\right| &\leqslant |a_{n+1}|\left|\prod_{i=1}^{n} a_i - \prod_{i=1}^{n} b_i\right| + |a_{n+1} - b_{n_1}|\left|\prod_{i=1}^{n} b_i\right| \\ &\leqslant \sum_{i=1}^{n} |a_i - b_i| + |a_{n+1} - b_{n+1}|.\end{aligned}$$

Cqfd. □

Retour à la preuve de la loi des grands nombres.
Soit $\varepsilon > 0$. Quitte à écrire $X_i = m + X_i'$ et $S_n/n = m + S_n'/n$, on peut supposer $m = 0$. Il suffit de montrer la convergence en loi de S_n/n vers 0 via les fonctions caractéristiques.
Comme les $X_i \in L^1$, la fonction caractéristique commune φ est dérivable en 0. On a $\varphi(0) = 1$ et $\varphi'(0) = i\mathbb{E}[X] = 0$.
Comme les variables sont i.i.d, on a $\varphi_{S_n}(t) = \varphi(t)^n$ et $\varphi_{S_n/n}(t) = \varphi(t/n)^n$.
En posant $a_i = \varphi_X(t/n)$ et $b_i = 1$ pour $i \geqslant 0$, d'après le lemme, on a

$$\left|\varphi_{S_n/n}(t) - 1\right| \leqslant n|\varphi(t/n) - 1| = t \times \frac{n}{t}|\varphi(t/n) - 1|.$$

Comme φ est dérivable en 0, on a $\lim_{n\to\infty} \left|\varphi_{S_n/n}(t) - 1\right| \leqslant t|\varphi'(0)| = 0$. Par conséquent, S_n/n converge en loi et donc en probabilité vers 0. $\square$

1.2 Version forte de la loi des grands nombres

Theorem 116 (Kolmogorov)**.** *Soit $(X_n)_{n\geqslant 0}$ une suite de v.a. i.i.d. dans L^1 de moyenne $m = \mathbb{E}[X_1]$. Alors*

$$\frac{1}{n}\sum_{i=1}^{n} X_i \underset{n\to\infty}{\longrightarrow} m \quad \textit{p.s. et dans } L^1.$$

Remarque. *Si $(X_n)_{n\geqslant 0}$ une suite de v.a. i.i.d. telle que il existe $c \in \mathbb{R}$ avec*

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow[n\to\infty]{p.s} c,$$

alors $\mathbb{E}[|X_1|] < \infty$ et $c = \mathbb{E}[X_1]$.
Il y a donc équivalence entre la convergence presque sûre de la moyenne empirique et l'existence d'un moment d'ordre 1 pour les suites de v.a. i.i.d.

Démonstration. Si on a convergence, alors $X_n/n = S_n/n - S_{n-1}/(n-1) \times \frac{n-1}{n} \xrightarrow{p.s} c - c = 0$. D'après Borel Cantelli, la suite étant i.i.d., on a $\forall\varepsilon > 0$, $\sum \mathbb{P}\left(\frac{|X_n|}{n} > \varepsilon\right) < \infty$. Par conséquent,

$$\sum \mathbb{P}\left(\frac{|X_n|}{n} > \varepsilon\right) = \sum \mathbb{P}(|X_n| > n\varepsilon) = \sum \mathbb{P}(|X_1| > n\varepsilon)$$

On conclut en remarquant que $\sum_{n\geqslant 1} \mathbb{P}(|X_1| > n) \leqslant \mathbb{E}[X_1] \leqslant \sum_{n\geqslant 0} \mathbb{P}(|X_1| > n)$. $\square$

Démonstration. (Loi forte des Grands Nombres dans le cas L^2)
On suppose que $X_1 \in L^2$. On note $x^+ = sup(x, 0)$ et $x^- = sup(-x, 0)$.
On peut écrire $X_n = X_n^+ - X_n^-$ avec X_n^+ et X_n^- satisfaisants les mêmes hypothèses que X_n. Par ailleurs, on peut écrire $M_n = \frac{1}{n}\sum_{i=1}^n X_i = M_n^+ - M_n^-$ où $M_n^+ = \frac{1}{n}\sum_{i=1}^n X_i^+$, $M_n^- = \frac{1}{n}\sum_{i=1}^n X_i^-$ et $m = \mathbb{E}[X_1^+] - \mathbb{E}[X_1^-]$.
On peut donc supposer sans restriction que $X_n \geqslant 0$.
On pose $Z_n = M_{n^2}$. On a $\mathbb{E}[Z_n] = m$ et $Var(Z_n) = \sigma^2/n^2$
On souhaite montrer que $Z_n \xrightarrow{p.s} m$. Soit $\varepsilon > 0$, par l'inégalité de Bienaymé-Chebychev,

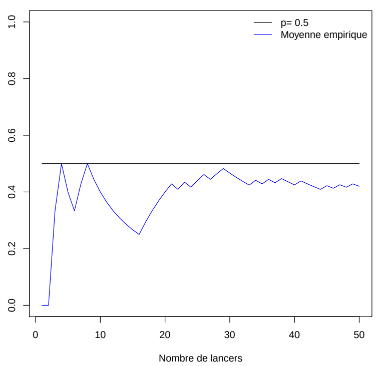
$$\sum \mathbb{P}(|Z_n - m| > \varepsilon) \leqslant \sum \frac{\sigma^2}{\varepsilon^2 n^2} < \infty.$$

Donc $Z_n \xrightarrow{p.s} m$.
Montrons maintenant que $M_n \xrightarrow{p.s} m$. On pose $q_n = [\sqrt{n}]$. On a $q_n \leqslant \sqrt{n} \leqslant q_n + 1$, par conséquent, $q_n^2 \leqslant n \leqslant (q_n+1)^2$.
Comme les variables sont positives, en posant $S_n = \sum_{i=1}^n X_i$, on a

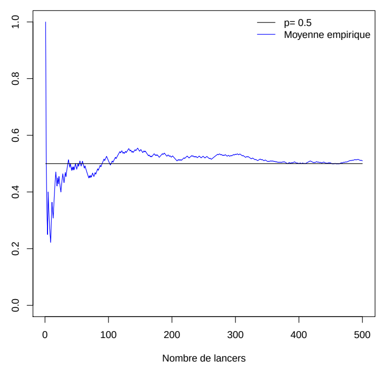
$$\begin{aligned} \frac{S_{q_n^2}}{n} &\leqslant \frac{S_n}{n} &\leqslant \frac{S_{(q_n+1)^2}}{n} \\ \frac{q_n^2}{n} Z_{q_n} &\leqslant M_n &\leqslant \frac{(q_n+1)^2}{n} Z_{q_n+1} \end{aligned}$$

Comme $\sqrt{n} - 1 \leqslant q_n \leqslant \sqrt{n}$, $q_n^2/n \to 1$ quand $n \to \infty$. Comme $Z_n \xrightarrow{p.s} m$, par encadrement $M_n \xrightarrow{p.s} m$. □

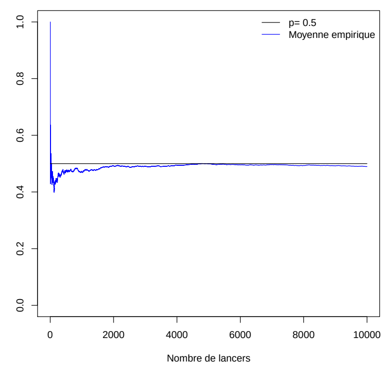
Simulations Considérons l'exemple du jeu "Pile ou Face" avec une pièce bien équilibrée. La probabilité de tomber sur "Pile" est alors $p = 1/2$. Le joueur mise sur "Pile". On répète plusieurs fois l'expérience et on observe l'évolution de la fréquence du nombre de succès en fonction du nombre de lancers :



pour $n = 50$



pour $n = 500$



pour $n = 10\,000$.

1.3 Application de la loi des grands nombres

Estimation d'une proportion inconnue

Une entreprise reçoit un lot de composants électroniques. Elle n'accepte le lot que si la proportion de pièces défectueuses est inférieur à 2%.
Ce lot étant important, il est impossible de vérifier l'état de chaque composant. On note p la probabilité qu'un composant soit défectueux. La quantité p est inconnue.
On choisit un échantillon de n composants au hasard dans le lot. On introduit les variables X_i où X_i vaut 1 si le $i^{\text{ème}}$ composant est défectueux et vaut 0 sinon. Les X_i sont i.i.d de loi $\mathcal{B}(p)$.
D'après la loi des grands nombres, la fréquence $M_n = \frac{1}{n}\sum_{i=1}^n X_i$ converge p.s. vers p. Il est donc légitime d'estimer p par M_n pour une grande valeur de n.
On peut se demander si M_n est une bonne approximation. Par l'inégalité de Chebychev, pour tout $p \in [0,1]$,

$$\mathbb{P}(|M_n - p| > t) \leqslant \frac{p(1-p)}{nt^2} \leqslant \frac{1}{4nt^2}$$

Donc $\mathbb{P}(p \in [M_n - t, M_n + t]) \geqslant 1 - 1/(4nt^2)$.

Méthode de Monte-Carlo

Soit $(X_n)_{n\geqslant 0}$ une suite de v.a. i.i.d à valeurs dans $\mathbb{R}^d$. et soit $g : \mathbb{R}^d \to \mathbb{R}$ une fonction mesurable telle que $\mathbb{E}[|g(X_1)|] < \infty$.

D'après la loi des grands nombres $\frac{1}{n}\sum_{i=1}^n g(X_i) \xrightarrow{p.s} \mathbb{E}[g(X_1)]$.

Si X_1 est une v.a. à densité f, alors $\mathbb{E}[g(X_1)] = \int g(x)f(x)dx$. Par conséquent, $\frac{1}{n}\sum_{i=1}^n g(X_i)$ est une approximation de l'intégrale $\int g(x)f(x)dx$.

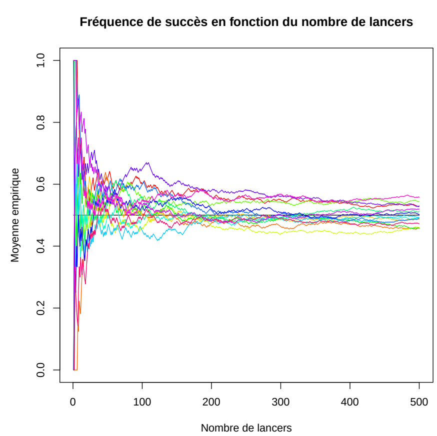
Par exemple, si $X_i \sim \mathcal{U}([a,b])$, alors $\frac{1}{n}\sum_{i=1}^n g(X_i)$ est une approximation de l'intégrale $\frac{1}{b-1}\int_a^b g(x)dx$.

La méthode de Monte-Carlo permet d'approcher numériquement une intégrale que l'on ne sait pas calculer. Le point fort de cette méthode par rapport aux méthodes numériques (rectangles, trapèze, ...) est que l'on n'a pas besoin de régularité sur la fonction g et elle est très facile à mettre en oeuvre quelque soit la dimension $d \geqslant 1$.

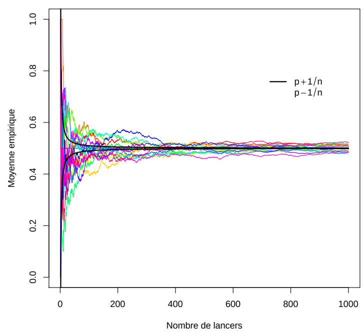
On peut alors se poser la question de la vitesse de convergence.

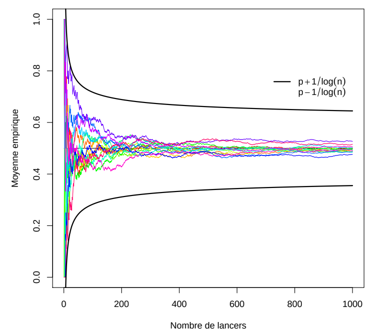
Simulations Observons la vitesse de convergence à l'aide de simulations.

On reprend notre exemple et on trace plusieurs réalisations de la trajectoire *aléatoire* $n \mapsto M_n$ (chaque réalisation ayant une couleur différente sur le graphique). On obtient le résultat suivant

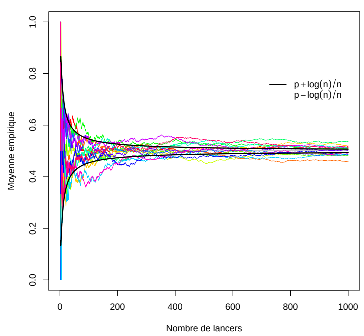


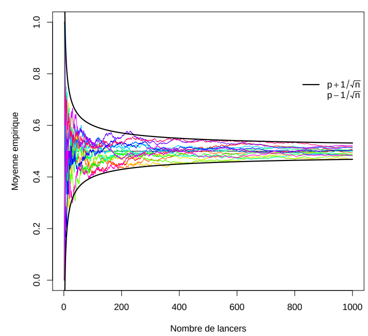
On observe que la convergence est plutôt lente. Essayons différentes fonctionnelles pour évaluer la vitesse de convergence (courbe tracée en noir).





la vitesse semble moins rapide que n — la vitesse semble plus rapide que $\ln(n)$





la vitesse semble moins rapide que $\frac{n}{\ln(n)}$ — la vitesse semble de l'ordre de $\sqrt{n}$.

Le théorème central limite va nous fournir la vitesse de convergence de la loi des grands nombres.

2 Théorème Central Limite (TCL)

Theorem 117. *Soit $(X_n)_{n\geqslant 0}$ une suite de v.a. réelles i.i.d. dans L^2 d'espérance $m=\mathbb{E}[X_1]$ et de variance $\sigma^2 = Var(X_1)>0$. On pose $S_n=\sum_{i=1}^n X_i$. Alors*

$$\frac{S_n - nm}{\sigma\sqrt{n}} \xrightarrow[n\to\infty]{Loi} \mathcal{N}(0,1)$$

ce qu'on peut réécrire sous la forme,

$$\frac{\sqrt{n}}{\sigma}\left(\frac{1}{n}\sum_{i=1}^n X_i - m\right) \xrightarrow[n\to\infty]{Loi} \mathcal{N}(0,1).$$

Conséquence 118. Pour tout $a,b\in\overline{\mathbb{R}}$, sous les hypothèses du théorème, on a

$$\mathbb{P}\left(\frac{\sqrt{n}}{\sigma}\left(\frac{1}{n}\sum_{i=1}^n X_i - m\right)\in[a,b]\right) \xrightarrow[n\to\infty]{} \mathbb{P}(Z\in[a,b]), \qquad \text{avec } Z\sim\mathcal{N}(0,1).$$

Démonstration. En remplaçant X_i par $\frac{X_i-m}{\sigma}$ on peut se ramener à $m=0$ et $\sigma=1$.

Lemma 119. *Pour tout $x\in\mathbb{R}$,*

$$\left|e^{ix}-\sum_{k=0}^{p}\frac{(ix)^k}{k!}\right| \leqslant \min\left(\frac{|x|^{p+1}}{(p+1)!},\frac{2|x|^p}{p!}\right).$$

Démonstration. On utilise la formule de Taylor avec reste intégral ;

$$e^{ix} = \sum_{k=0}^{p} \frac{(ix)^k}{k!} + \frac{i^{p+1}}{p!} \int_0^x (x-s)^p e^{is} ds.$$

Par conséquent,

$$\left| e^{ix} - \sum_{k=0}^{p} \frac{(ix)^k}{k!} \right| \leqslant \frac{1}{p!} \left| \int_0^x (x-s)^p ds \right| = \frac{|x|^{p+1}}{(p+1)!}.$$

Par ailleurs, par intégration par parties (en posant $u = (x-s)^p$ et $dv = e^{is} ds$),

$$\left| e^{ix} - \sum_{k=0}^{p} \frac{(ix)^k}{k!} \right| = \left| \frac{i^p}{(p-1)!} \int_0^x (x-s)^{p-1} e^{is} ds - \frac{(ix)^p}{p!} \right|$$

□

Comme les variables sont i.i.d, $\varphi_{s_n/n}(t) = \varphi(t/n)^n$. D'après le lemme 115, en posant $a_i = \varphi(t/n)$ et $b_i = (1 - t^2/2n)$, on a

$$\left| \varphi(t/\sqrt{n})^n - (1 - t^2/2n)^n \right| \leqslant n \left| \varphi(t/\sqrt{n}) - (1 - t^2/2n) \right|$$

Par ailleurs, d'après le lemme 119, et en remarquant que $\mathbb{E}[X] = 0$,

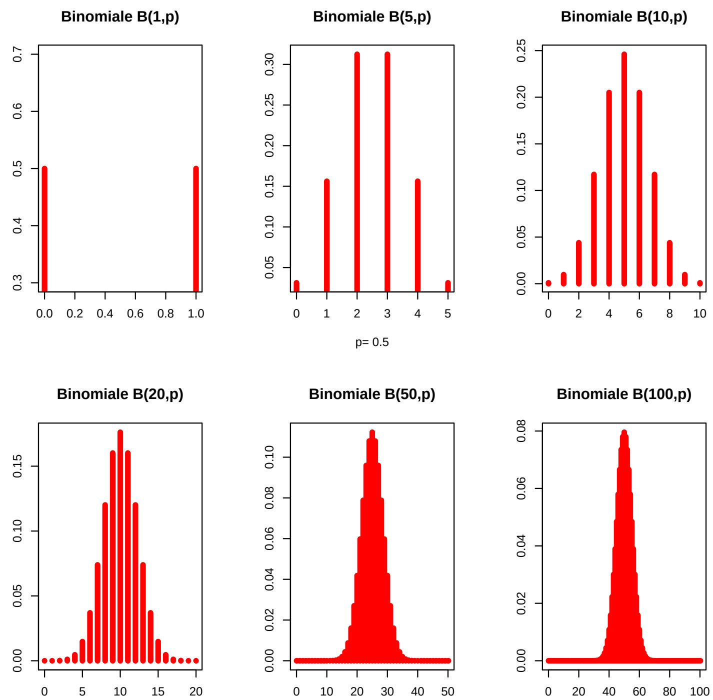
$$n \left| \varphi(t/\sqrt{n}) - (1 - t^2/2n) \right| \leqslant n\mathbb{E}\left[\min\left(\frac{t^3|X|^3}{6n^{3/2}}, \frac{t^2|X|^2}{n} \right) \right] = \mathbb{E}\left[\min\left(\frac{t^3|X|^3}{6n^{1/2}}, t^2|X|^2 \right) \right].$$

Par le théorème de convergence dominée, on en déduit que $\lim_{n\to\infty} \left| \varphi(t/\sqrt{n})^n - (1 - t^2/2n)^n \right| = 0$.
Par ailleurs, $(1 - t^2/2n)^n = e^{n \ln\left(1 - \frac{t^2}{2n}\right)} \underset{n\to\infty}{\longrightarrow} e^{-t^2/2}$ qui est la fonction caractéristique de la loi $\mathcal{N}(0,1)$. On conclut en utilisant le théorème de Paul Lévy.

□

Application
On considère des variables X_i i.i.d. dans L^2. On peut toujours approcher la loi de $S_n = \sum_{i=1}^n X_i$ par la loi normale quelque soit la loi initiale des X_i.
Par conséquent, si $X_n \sim \mathcal{B}(p)$, on sait que $S_n \sim \mathcal{B}(n,p)$. Comme $\frac{S_n - np}{\sqrt{np(1-p)}}$ est proche de la loi $\mathcal{N}(0,1)$, on dit que la loi $\mathcal{B}(n,p)$ est proche de $\mathcal{N}(np, np(1-p))$ pour n grand.
Par conséquent, si $X_n \sim \mathcal{P}(\lambda)$, on sait que $S_n \sim \mathcal{P}(n\lambda)$. Comme $\frac{S_n - n\lambda}{\sqrt{n\lambda}}$ est proche de la loi $\mathcal{N}(0,1)$, on dit que la loi $\mathcal{P}(n\lambda)$ est proche de $\mathcal{N}(n\lambda, n\lambda)$ pour n grand.

Simulations Observons l'évolution de la loi Binomiale $\mathcal{B}(n,p)$ en fonction de n vers la "cloche gaussienne", pour $p = 1/2$:



Exemple. *Un joueur qui doit choisir au hasard une carte dans un jeu de 32 cartes obtient certains avantages s'il découvre un roi. On constate qu'il a retourné 11 fois un roi sur 50 essais. Peut-on présumer, au risque de* 5%*, que ce joueur est un tricheur ?*

La probabilité de tirer un roi dans un jeu de 32 cartes est $p = 4/32 = 1/8$. On introduit les variables aléatoire X_i qui représentent le résultat du i$^{\text{ème}}$ jeu. On a $X_i = 1$ si le joueur obtient un roi et $X_i = 0$ si le joueur obtient une autre carte qu'un roi. On suppose que le jeu est bien mélangée entre chaque tirage, et donc de façon naturelle les variables X_i sont indépendantes, de loi de Bernoulli $\mathcal{B}(1/8)$.
On va exprimer un intervalle de fluctuation de la fréquence de succès. D'après le théorème central limite, on a

$$\mathbb{P}\left(M_n \in \left[p - t\frac{\sqrt{p(1-p)}}{\sqrt{n}}, p + t\frac{\sqrt{p(1-p)}}{\sqrt{n}}\right]\right) = \mathbb{P}\left(\frac{\sqrt{n}}{\sqrt{p(1-p)}}|M_n - p| \leqslant t\right) \underset{n\to\infty}{\longrightarrow} \mathbb{P}(|Z| \leqslant t),$$

avec $Z \sim \mathcal{N}(0,1)$.
Ici $n = 50$ que l'on va supposer assez grand pour utiliser l'approximation gaussienne. On choisit $t > 0$ tel que $\mathbb{P}(|Z| \leqslant t) = 2\mathbb{P}(Z \leqslant t) - 1 = 95\%$. D'après la table gaussienne, on a $t = 1,96$. Par conséquent, pour $p = 1/8$ et $n = 50$,

$$\mathbb{P}(M_{50} \in I) \simeq 95\%$$

avec $I = \left[\frac{1}{8} - 1.96\frac{\sqrt{7}}{8\sqrt{50}}, \frac{1}{8} + 1.96\frac{\sqrt{7}}{8\sqrt{50}}\right] = [0.033, 0.217]$.
Le joueur a obtenu une fréquence de succès de $11/50 = 0.22$ qui est hors de l'intervalle de fluctuations. Par conséquent, au niveau de risque 5%, on peut mettre en doute l'honnêteté du joueur.

Application Construction d'intervalles de confiance.
Reprenons un exemple précéent. Une entreprise reçoit un lot de composants électroniques. Elle n'accepte le lot que si la proportion de pièces défectueuses est inférieur à 2%.

On note p la probabilité qu'un composant soit défectueux. La quantité p est inconnue.
On choisit un échantillon de n composants au hasard dans le lot. On introduit les variables X_i où X_i vaut 1 si le $i^{\text{ème}}$ composant est défectueux et vaut 0 sinon. Les X_i sont i.i.d de loi $\mathcal{B}(p)$.
En statistique, la fréquence $M_n = \frac{1}{n}\sum_{i=1}^n X_i$ est appelée moyenne empirique et est notée $\overline{X}_n$. D'après la loi forte des grands nombres, la fréquence $\overline{X}_n$ converge p.s. vers p. Il est donc légitime d'estimer p par $\overline{X}_n$ pour une grande valeur de n. On parle d'estimation ponctuelle de p.
Par ailleurs, comme les variables sont centrées en p, d'après le T.C.L, on a pour tout $t > 0$,

$$\mathbb{P}\bigg(\frac{\sqrt{n}}{\sigma}\big|\overline{X}_n - p\big| \leqslant t\bigg) = \mathbb{P}\bigg(p \in \bigg[\overline{X}_n - \sigma\frac{t}{\sqrt{n}}, \overline{X}_n + \sigma\frac{t}{\sqrt{n}}\bigg]\bigg) \underset{n\to\infty}{\longrightarrow} \mathbb{P}(|Z| \leqslant t), \qquad \text{avec } Z \sim \mathcal{N}(0,1).$$

On peut par exemple choisir $t = 1.96$, et pour n assez grand, on a, $\forall p \in]0,1[$,

$$\mathbb{P}\bigg(p \in \bigg[\overline{X}_n - \sigma\frac{1.96}{\sqrt{n}}, \overline{X}_n + \sigma\frac{1.96}{\sqrt{n}}\bigg]\bigg) \simeq \mathbb{P}(|Z| \leqslant 1.96) = 95\%.$$

Dans le cas de la loi de Bernoulli, on a $\sigma^2 = p(1-p)$ qui est inconnu, par conséquent l'intervalle $\Big[\overline{X}_n - t\sqrt{\frac{p(1-p)}{n}}, \overline{X}_n + t\sqrt{\frac{p(1-p)}{n}}\Big]$ ne peut pas être calculé juste avec la connaissance de l'échantillon $(X_1, \ldots, X_n)$. On va alors introduire un estimateur de $\sigma^2 = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$.

On considère une suite de variables indépendantes $(X_n)_{n\geqslant 1}$ i.i.d. dans L^2 d'espérance m et de variance σ^2. Un estimateur naturel de la variance est la variance empirique

$$S_n^2 = \frac{1}{n}\sum_{i=1}^n \left(X_i - \overline{X}_n\right)^2 = \frac{1}{n}\sum_{i=1}^n X_i^2 - \overline{X}_n^2$$

On remarque que $\mathbb{E}[S_n^2] = \frac{n-1}{n}\sigma^2$ et d'après la loi des grands nombres $S_n^2 \xrightarrow{p.s} \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \sigma^2$. En utilisant le Lemme de Slutsky et le théorème central limite, on en déduit que

$$\frac{\sqrt{n}}{S_n}\left(\frac{1}{n}\sum_{i=1}^n X_i - m\right) \underset{n\to\infty}{\xrightarrow{Loi}} \mathcal{N}(0,1).$$

Dans le cas, où les v.a. X_i suivent la loi $\mathcal{B}(p)$, on remarque que $S_n^2 = \overline{X}_n(1 - \overline{X}_n)$, car $X_i^2 = X_i$. Par conséquent, pour tout $t > 0$, pour tout $p \in]0,1[$,

$$\mathbb{P}\left(p \in \left[\overline{X}_n - t\sqrt{\frac{\overline{X}_n(1-\overline{X}_n)}{n}}, \overline{X}_n + t\sqrt{\frac{\overline{X}_n(1-\overline{X}_n)}{n}}\right]\right) \underset{n\to\infty}{\longrightarrow} \mathbb{P}(|Z| \leqslant t), \qquad \text{avec } Z \sim \mathcal{N}(0,1).$$

En prenant $t = 1,96$, quelque soit $p \in]0,1[$, pour n assez grand, on a

$$\mathbb{P}\left(p \in \left[\overline{X}_n - 1,96\sqrt{\frac{\overline{X}_n(1-\overline{X}_n)}{n}}, \overline{X}_n + 1,96\sqrt{\frac{\overline{X}_n(1-\overline{X}_n)}{n}}\right]\right) \simeq 95\%.$$

<u>Conclusion</u>
Quelque soit $p \in]0,1[$, l'inconnue p est dans l'intervalle $\Big[\overline{X}_n - 1,96\sqrt{\frac{\overline{X}_n(1-\overline{X}_n)}{n}}, \overline{X}_n + 1,96\sqrt{\frac{\overline{X}_n(1-\overline{X}_n)}{n}}\Big]$ avec probabilité 95% pour n assez grand. Cet intervalle est calculable à partir de l'échantillon, qui est connu. On appelle cet intervalle *intervalle de confiance asymptotique de p au niveau de confiance* 95%.

3 Preuve de la loi forte des grands nombres

Theorem 120 (Théorème des séries centrées)**.** *Soit $(X_n)_{n\geqslant 1}$ une suite de variables réelles indépendantes telle que $\forall n \geqslant 1$ $X_n \in L^2$ et $\mathbb{E}[X_n] = 0$.*
Si $\sum_{n\geqslant 1} \mathbb{E}[X_n^2] < \infty$, alors $(S_n)_{n\geqslant 1}$ converge p.s. et dans L^2 vers une v.a. réelle.

Démonstration. On montre tout d'abord le lemme suivant :

Lemma 121. *Soit $(X_n)_{n\geqslant 1}$ une suite de variables réelles indépendantes telle que $\forall n \geqslant 1$ $X_n \in L^2$ et $\mathbb{E}[X_n] = 0$. Alors pour tout $\varepsilon > 0$, on a*

$$\mathbb{P}(\sup_{1\leqslant k\leqslant n} |S_k| \geqslant \varepsilon) \leqslant \frac{Var(S_n)}{\varepsilon^2}.$$

Démonstration. On introduit les événements disjoints

$$A_k = \{|S_1| < \varepsilon, \ldots, |S_{k-1}| < \varepsilon, |S_k| \geqslant \varepsilon\}.$$

On a

$$\begin{aligned}\mathbb{E}\big[S_n^2\big] &\geqslant \sum_{k=1}^n \mathbb{E}\big[S_n^2 \mathbb{1}_{A_k}\big] = \sum_{k=1}^n \mathbb{E}\big[(S_n - S_k + S_k)^2 \mathbb{1}_{A_k}\big] \\ &\geqslant \sum_{k=1}^n \mathbb{E}\big[\big(2(S_n - S_k)S_k + S_k^2\big)\mathbb{1}_{A_k}\big].\end{aligned}$$

Comme $A_k \in \sigma(X_1, \ldots, X_k)$, S_k est $\sigma(X_1, \ldots, X_k)$-mesurable et $S_n - S_k = \sum_{i=k+1}^n X_i$ est indépendant de $\sigma(X_1, \ldots, X_k)$, on a $\mathbb{E}[2(S_n - S_k)S_k \mathbb{1}_{A_k}] = 2\mathbb{E}[S_n - S_k]\mathbb{E}[S_k \mathbb{1}_{A_k}] = 0$ car les X_i sont centrées. Par conséquent,

$$\mathbb{E}\big[S_n^2\big] \geqslant \sum_{k=1}^n \mathbb{E}\big[S_k^2 \mathbb{1}_{A_k}\big] \geqslant \varepsilon^2 \sum_{k=1}^n \mathbb{P}(A_k) = \varepsilon^2 \mathbb{P}(\sup_{1\leqslant k\leqslant n} |S_k| \geqslant \varepsilon).$$

□

Démontrons maintenant le théorème des séries centrées.
D'après les critères de convergence, on a $(X_n)_{n\geqslant 1}$ converge presque sûrement ssi $\forall \varepsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}\left(\sup_{k\geqslant 1} |S_{n+k} - S_n| > \varepsilon\right) = 0.$$

Soit $\varepsilon > 0$, par monotonie on a

$$\mathbb{P}\left(\sup_{k\geqslant 1} |S_{n+k} - S_n| > \varepsilon\right) = \lim_{p\to\infty} \mathbb{P}\left(\sup_{1\leqslant k\leqslant p} |S_{n+k} - S_n| > \varepsilon\right).$$

Par ailleurs,

$$S_{n+k} - S_n = \sum_{i=n+1}^{n+k} X_i = \sum_{i=1}^k X_{n+i}$$

Par conséquent, en utilisant le lemme et l'indépendance des variables centrées X_i,

$$\mathbb{P}\left(\sup_{1\leqslant k\leqslant p} |S_{n+k} - S_n| > \varepsilon\right) \leqslant \frac{1}{\varepsilon^2} Var\left(\sum_{i=1}^p X_{n+i}\right) = \frac{1}{\varepsilon^2} \sum_{i=n+1}^{n+p} \mathbb{E}[X_i^2].$$

On obtient ainsi,

$$\mathbb{P}\left(\sup_{k\geqslant 1}|S_{n+k}-S_n|>\varepsilon\right)\leqslant\frac{1}{\varepsilon^2}\sum_{i=n+1}^{\infty}\mathbb{E}[X_i^2].$$

Comme la série $\sum\mathbb{E}[X_n^2]$ converge, on obtient que la suite $(S_n)_{n\geqslant 1}$ converge ps.

□

Dans le cas des sommes de v.a. indépendantes, on a un résultat remarquable sur les différents types de convergence.

Theorem 122. *Soit $(X_n)_{n\geqslant 1}$ une suite de variables réelles indépendantes. Pour $n\geqslant 1$, on pose $S_n=\sum_{k=1}^n X_k$. On a équivalence entre*

1. *$(S_n)_{n\geqslant 1}$ converge p.s. vers une v.a. réelle,*
2. *$(S_n)_{n\geqslant 1}$ converge en proba vers une v.a. réelle,*
3. *$(S_n)_{n\geqslant 1}$ converge en loi vers une v.a. réelle.*

Démonstration. Voir Probability and Measure de P. Billingsley. □

Donnons maintenant une preuve de la loi forte des grands nombres. On rappelle tout d'abord quelques lemmes bien utiles en analyse.

Lemma 123. *Soit $(b_n)_{n\geqslant 1}$ une suite croissante de réels strictement positifs telle que $\lim\limits_{n\to+\infty}b_n=+\infty$. On pose $b_0=0$.*
Soit $(x_n)_{n\geqslant 1}$ une suite de réels.

<u>*Cesaro*</u> *: Si la suite $(x_n)_{n\geqslant 1}$ converge vers x alors $\displaystyle\lim_{n\to+\infty}\frac{1}{b_n}\sum_{k=1}^n(b_k-b_{k-1})x_k=x$.*

<u>*Kronecker*</u> *: Si la série $\displaystyle\sum_{n\geqslant 1}\frac{1}{b_n}x_n$ converge dans $\mathbb{R}$ alors $\displaystyle\lim_{n\to+\infty}\frac{1}{n}\sum_{k=1}^n x_k=0$.*

Lemma 124. *Pour $\alpha>1$, $k\geqslant 1$, $\displaystyle\sum_{n\geqslant k+1}n^{-\alpha}\leqslant\frac{1}{(\alpha-1)k^{\alpha-1}}$.*

Démonstration. Comme la finction $x\mapsto x^{-\alpha}$ est décroissante sur $]0,\infty[$, on a

$$\sum_{n\geqslant k+1}n^{-\alpha}\leqslant\sum_{n\geqslant k+1}\int_{n-1}^{n}x^{-\alpha}dx=\int_k^\infty x^{-\alpha}dx=\frac{k^{1-\alpha}}{(\alpha-1)}.$$

□

Démonstration. (Loi des grands nombres)
On peut supposer sans perte de généralité que $m=\mathbb{E}[X_1]=0$ et on souhaite montrer que $M_n=\frac{1}{n}\sum_{k=1}^n X_n$ converge p.s. vers 0.
On pose

$$\widehat{X}_n=X_n\mathbb{1}_{|X_n|<n},\qquad\widehat{M}_n=\tfrac{1}{n}\textstyle\sum_{k=1}^n\widehat{X}_k;$$
$$\widetilde{X}_n=\widehat{X}_n-\mathbb{E}\left[\widehat{X}_n\right],\quad\widetilde{M}_n=\tfrac{1}{n}\textstyle\sum_{k=1}^n\widetilde{X}_k.$$

Les suites $\left(\widehat{X}_n\right)_{n\geqslant 1}$ et $\left(\widetilde{X}_n\right)_{n\geqslant 1}$ sont toutes les deux des suites de v.a. indépendantes. On va procéder en deux étapes :

(a) On a les équivalences suivantes

1. $M_n \xrightarrow{p.s} 0$ si et seulement si $\widehat{M}_n \xrightarrow{p.s} 0$.
2. $\widehat{M}_n \xrightarrow{p.s} 0$ si et seulement si $\widetilde{M}_n \xrightarrow{p.s} 0$.

(b) On montre que $\widetilde{M}_n \xrightarrow{p.s} 0$.

Montrons le premier point.

(a) 1. On montre que $M_n - \widehat{M}_n \xrightarrow{p.s} 0$.

On a $\forall n \geqslant 1$, $M_n - \widehat{M}_n = \frac{1}{n}\sum_{k=1}^{n} X_k \mathbb{1}_{|X_k|\geqslant k}$. Comme les variables sont de même loi, on a

$$\sum_{n\geqslant 1}\mathbb{P}(|X_n|\geqslant n) = \sum_{n\geqslant 1}\mathbb{P}(|X_1|\geqslant n) \leqslant 1 + \mathbb{E}[|X_1|] < \infty.$$

Donc d'après le premier lemme de Borel-Cantelli,

$$\mathbb{P}(\limsup\{|X_n|\geqslant n\}) = 0.$$

Soit $\omega \in \liminf\{|X_n| < n\}$, il existe alors $n_\omega \geqslant 1$ tel que $\forall n \geqslant n_\omega$, $|X_n(\omega)| < n$, i.e. $\widehat{X}_n(\omega) = X_n(\omega)$. Par conséquent, $M_n(\omega) - \widehat{M}_n(\omega) = \frac{1}{n}\sum_{k=1}^{n_\omega} X_k(\omega)\mathbb{1}_{|X_k(\omega)|\geqslant k}$. On en déduit que pour tout $\omega \in \liminf\{|X_n| < n\}$, événement de probabilité 1, $M_n(\omega) - \widehat{M}_n(\omega) \to 0$ quand $n \to +\infty$.

2. On va maintenant montrer que $\widehat{M}_n - \widetilde{M}_n \xrightarrow{p.s} 0$. Du fait que les variables soient de même loi, on remarque que

$$\widehat{M}_n - \widetilde{M}_n = \frac{1}{n}\sum_{k=1}^{n}\mathbb{E}\big[X_k \mathbb{1}_{|X_k|\geqslant k}\big] = \frac{1}{n}\sum_{k=1}^{n}\mathbb{E}\big[X_1 \mathbb{1}_{|X_1|\geqslant k}\big].$$

Comme $X_1 \in L^1$, par convergence dominée, on a $\lim_{n\to\infty}\mathbb{E}\big[X_1\mathbb{1}_{|X_1|\geqslant n}\big] = 0$. En utilisant le lemme de Cesaro avec $b_n = n$ et $x_n = \mathbb{E}\big[X_1\mathbb{1}_{|X_1|\geqslant n}\big]$, on en déduit que $\widehat{M}_n - \widetilde{M}_n$ converge vers 0.

(b) On étudie la convergence p.s. de $\left(\widetilde{M}_n\right)_{n\geqslant 1}$.

Comme $\widetilde{M}_n = \frac{1}{n}\sum_{k=1}^{n}\widetilde{X}_k$, d'après le Lemme de Kronecker, il suffit de montrer que $\sum_{k=1}^{n}\widetilde{X}_k/k$ converge dans $\mathbb{R}$ p.s.
D'après le théorème des séries centrées, il suffit de montrer que $\sum_{n\geqslant 1}\mathbb{E}[\widetilde{X}_n^2]/n^2 < \infty$. On a

$$\sum_{n\geqslant 1}\frac{1}{n^2}\mathbb{E}[\widetilde{X}_n^2] = \sum_{n\geqslant 1}\frac{1}{n^2}\mathbb{E}\left[\left(\widehat{X}_n - \mathbb{E}\left[\widehat{X}_n\right]\right)^2\right] = \sum_{n\geqslant 1}\frac{1}{n^2}Var\left(\widehat{X}_n\right) \leqslant \sum_{n\geqslant 1}\frac{1}{n^2}\mathbb{E}\left[\widehat{X}_n^2\right]$$

Comme les variables sont de même loi, on a pour tout $n \geqslant 1$,

$$\mathbb{E}\left[\widehat{X}_n^2\right] = \mathbb{E}\big[X_n^2\mathbb{1}_{|X_n|<n}\big] = \mathbb{E}\big[X_1^2\mathbb{1}_{|X_1|<n}\big].$$

Par conséquent, par Fubini-Tonelli et en utilisant le lemme 124,

$$\begin{aligned}\sum_{n\geqslant 1}\frac{1}{n^2}\mathbb{E}[\widetilde{X}_n^2] &\leqslant \mathbb{E}\left[X_1^2\sum_{n\geqslant 1}\frac{1}{n^2}\mathbb{1}_{|X_1|<n}\right] \leqslant \mathbb{E}\left[X_1^2\sum_{n\geqslant [|X_1|]+1}\frac{1}{n^2}\right]\\ &\leqslant \mathbb{E}\left[\frac{X_1^2}{[|X_1|]+1}\right] \leqslant \mathbb{E}[|X_1|] < \infty.\end{aligned}$$

Par conséquent, $\widetilde{M}_n \xrightarrow{p.s} 0$ et donc $M_n \xrightarrow{p.s} 0$.

Montrons maintenant la convergence de $(M_n)_{n\geqslant 1}$ vers 0 dans L^1. Soit $p \geqslant 1$, on a

$$|M_n| = \min(|M_n|, p) + (|M_n| - p)^+,$$

où $(x-p)^+ = \sup(x-p, 0)$.Comme la fonction $x \mapsto (x-p)^+$ est convexe croissante, on a

$$(|M_n| - p)^+ \leqslant \frac{1}{n}\sum_{k=1}^{n}(|X_k| - p)^+.$$

Par convergence dominée, à p fixé, on a $\mathbb{E}[\min(|M_n|, p)] \to 0$ quand $n \to \infty$ et comme les variables sont de même loi

$$\mathbb{E}\big[(|M_n| - p)^+\big] \leqslant \frac{1}{n}\sum_{k=1}^{n}\mathbb{E}\big[(|X_k| - p)^+\big] = \mathbb{E}\big[(|X_1| - p)^+\big].$$

Par convergence dominée par $|X_1|$, on a $\mathbb{E}[(|X_1| - p)+] \to 0$ quand $p \to \infty$. On en déduit la convergence vers 0 dans L^1 de $(M_n)_{n\geqslant 1}$. $\square$

4 Extensions

On peut se demander s'il existe des extensions à la loi des grands nombres ou au TCL pour des suites de v.a. non i.i.d. Le théorème des séries centrées permet de travailler avec des suites de v.a. indépendantes mais pas nécessairement de même loi. Si les variables sont faiblement dépendantes, dans le cas des chaînes de Markov, on a alors un théorème similaire qui s'appelle le théorème ergodique.

Dans le cas du théorème central limite, on peut affaiblir les conditions. Le théorème reste notamment vrai sous les conditions de Lindeberg : supposons que les v.a. (X_n) sont indépendantes dans L^2, mais pas nécessairement de même loi. En notant $m_n = \mathbb{E}[X_n]$, $\sigma_n^2 = Var(X_n)$ et $s_n^2 = \sum_{k=1}^n \sigma_k^2$, si pour tout $\varepsilon > 0$,

$$\lim_{n\to\infty} \frac{1}{s_n^2}\sum_{k=1}^{n}\mathbb{E}\big[(X_n - m_n)^2 \mathbb{1}_{|X_n - m_n| > \varepsilon s_n}\big] = 0,$$

alors

$$\frac{1}{s_n}\sum_{k=1}^{n}(X_n - m_n) \xrightarrow{Loi} \mathcal{N}(0,1).$$

Il existe par ailleurs un théorème de Loi des grands nombres et un théorème Central Limite pour certains types de processus $(X_n)_{n\geqslant 1}$, appelés martingales. Pour ces processus, la dépendance entre X_{n+1} et $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$ est contrôlée par la condition suivante : connaissant le passé $\mathcal{F}_n$, la loi de X_{n+1} est centrée en X_n.

5 Exercices

Exercice 1. Une compagnie aérienne gère des vols de 200 passagers avec les données suivantes :

- le poids en kilogrammes d'un passager est une v.a.r. d'espérance 65 et d'écart type 15 ;
- quand le poids maximal de bagages autorisé par passager est de Mkg, le poids de bagages par passager en kilogrammes est une v.a.r. d'espérance $0.7M$ et d'écart type $0.2M$;
- les 400 v.a.r. "poids des passagers" et "poids des bagages" sont indépendantes.

Quelle valeur de M maximale permet d'avoir 95 chances sur 100 de faire rentrer 200 passagers dans l'avion sans dépasser la charge maximale autorisée, égale à 18 tonnes ?

Corrigé : On note X_i le poids du $i^{\text{ème}}$ passager et B_i le poids de son bagage. Les variables sont supposées indépendantes. Les X_i suivent la même loi et les B_i suivent la même loi. Le poids des 200 passagers et de leurs bagages est $\sum_{i=1}^{200} Z_i$, avec $Z_i = X_i + B_i$. On cherche par conséquent M tel que

$$\mathbb{P}\left(\sum_{i=1}^{200} Z_i \leqslant 18\,000\right) \geqslant 0.95.$$

Les variables Z_i sont des v.a. i.i.d d'espérance $65+0.7M$ et de variance $225+(0.2M)^2$. On sait que d'après le théorème central limite, on a

$$\frac{\sqrt{n}}{\sqrt{225+(0.2M)^2}}\left(\frac{1}{n}\sum_{i=1}^{n} Z_i - (65+0.7M)\right) \xrightarrow{Loi} Z \sim \mathcal{N}(0,1).$$

Par conséquent,

$$\begin{aligned}
&\mathbb{P}\left(\sum_{i=1}^{200} Z_i \leqslant 18\,000\right)\\
&= \mathbb{P}\left(\frac{\sqrt{200}}{\sqrt{225+(0.2M)^2}}\left(\frac{1}{200}\sum_{i=1}^{200} Z_i - (65+0.7M)\right) \leqslant \frac{\sqrt{200}}{\sqrt{225+(0.2M)^2}}(90-(65+0.7M))\right)\\
&\simeq \mathbb{P}\left(Z \leqslant \frac{\sqrt{200}}{\sqrt{225+(0.2M)^2}}(25-0.7M)\right).
\end{aligned}$$

D'après la table gaussienne, il faut

$$\frac{\sqrt{200}}{\sqrt{225+(0.2M)^2}}(25-0.7M) \geqslant 1.65$$

On a par conséquent besoin de résoudre l'équation de degré 2

$$\begin{aligned}
&(25-0.7M)^2 - \frac{(1.65)^2}{200}\big(225+(0.2M)^2\big) = 0\\
&\left((0.7)^2 - \left(\frac{1.65\times 0.2}{200}\right)^2\right)M^2 - 1.4\times 25M + 625 - 225\frac{(1.65)^2}{200}\\
&0.49M^2 - 35M + 621,94 = 0
\end{aligned}$$

On a $\Delta = 6$, les solution sont $M_1 = 33.21$kg et $M_2 = 38.21$kg. La compagnie d'avion choisi par conséquent $M = 33$kg. $\triangle$

Chapitre 8

Vecteurs Gaussiens

1 Variables gaussiennes

Definition 125. *Une variable X suit la loi normale (gaussienne) $\mathcal{N}(m, \sigma^2)$ avec $m \in \mathbb{R}$ et $\sigma > 0$ si sa loi est à densité donnée par*

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

Remarque : Si $\sigma = 0$, la loi de la variable est dégénérée et la v.a. X est constante égale à m. On étend ainsi la dénomination de v.a. gaussienne aux constantes.

Propriété 126.

1. *$X \sim \mathcal{N}(m, \sigma^2)$ si et seulement si $\frac{X-m}{\sigma} \sim \mathcal{N}(0, 1)$.*
2. *$\mathbb{E}[X] = m$ et $Var(X) = \sigma^2$.*
3. *Fonction caractéristique : $\varphi_X(t) = e^{imt} e^{\frac{-\sigma^2 t^t}{2}}$.*
4. *Si $X \sim \mathcal{N}(m_X, \sigma_X^2)$ et $Y \sim \mathcal{N}(m_Y, \sigma_Y^2)$ indépendantes, alors $X + Y \sim \mathcal{N}(m_X + m_Y, \sigma_X^2 + \sigma_Y^2)$.*

2 Vecteurs Gaussiens

On se place dans $\mathbb{R}^d$. On note A^t la transposée d'une matrice A. Un vecteur X à valeurs dans $\mathbb{R}^d$ sera toujours considéré comme un vecteur colonne :

$$X = \begin{pmatrix} X_1 \\ \vdots \\ X_d \end{pmatrix} = (X_1, \ldots, X_d)^t$$

On note le produit scalaire $< x, y >= \sum_{i=1}^d x_i y_i = x^t\, y$.

Definition 127. *Un vecteur aléatoire X de $\mathbb{R}^d$ est gaussien si toutes les combinaisons linéaires de ses coordonnées suivent un loi gaussienne : $\forall a \in \mathbb{R}^d < a, X >= \sum_{i=1}^d a_i X_i$ est de loi gaussienne.*

Attention : Savoir que pour tout $i \geqslant 1$, X_i est une variable gaussienne ne suffit pas pour que $X = (X, \ldots, X_d)$ soit un vecteur gaussien.
Par contre, $X = (X_1, \ldots, X_d)$, avec les X_i indépendantes de loi normale, est un vecteur gaussien.

Proposition 128. *L'image par une application linéaire d'un vecteur gaussien est encore un vecteur gaussien, i.e. si X est un vecteur gaussien et $A \in \mathcal{M}_{p,d}(\mathbb{R})$, alors AX est un vecteur gaussien de $\mathbb{R}^p$.*

Démonstration. Soit $a \in \mathbb{R}^p$. On a $< a, AX >= a^t\, AX = \left(A^t\, a\right)^t X =< A^t\, a, X >$ qui suit une loi gaussienne car combinaison linéaire du vecteur X. $\square$

2.1 Loi d'un vecteur gaussien

Rappel : un vecteur aléatoire X de $\mathbb{R}^d$ est dans L^p si et seulement si $\forall i \in \{1, \ldots, d\}$ X_i est dans L^p. Par conséquent, un vecteur gaussien est dans L^p pour tout $p \geqslant 1$.

Definition 129. *Soit $X = (X_1, \ldots, X_d)^t$ un vecteur aléatoire de $\mathbb{R}^d$.*
Si $X \in L^1$, $\mathbb{E}[X] = (\mathbb{E}[X_1], \ldots, \mathbb{E}[X_d])^t$. Si $\mathbb{E}[X] = 0$ le vecteur est dit centré.
La matrice de covariance d'un vecteur X de L^2 est la matrice carré symétrique positive

$$\begin{aligned} Cov(X) &= (Cov(X_i, X_j))_{1 \leqslant i,j \leqslant d} = (\mathbb{E}[X_i X_j] - \mathbb{E}[X_i]\mathbb{E}[X_j])_{1 \leqslant i,j \leqslant d} \\ &= \mathbb{E}[XX^t] - \mathbb{E}[X]\mathbb{E}[X]^t = \mathbb{E}\big[(X - \mathbb{E}[X])(X - \mathbb{E}[X])^t\big]. \end{aligned}$$

Proposition 130. *Soit X un vecteur aléatoire de $\mathbb{R}^d$. X est un vecteur gaussien si et seulement si sa fonction caractéristique est de la forme*

$$z \mapsto \exp\left(i < z, m > -\frac{1}{2} < Kz, z >\right) = \exp\left(iz^t m - \frac{1}{2} z^t K z\right)$$

où $m \in \mathbb{R}^d$ et K une matrice réelle symétrique positive.
Dans ce cas $m = \mathbb{E}[X]$ et $K = Cov(X)$.

Démonstration. Supposons que X est un vecteur gaussien. La fonction caractéristique de X est définie par, $z \in \mathbb{R}^d$.

$$\varphi_X(z) = \mathbb{E}\big[e^{i<z,X>}\big] = \mathbb{E}\Big[e^{i \sum_{k=1}^d z_k X_k}\Big]$$

où $< z, X >$ est une variable gaussienne d'espérance $\mathbb{E}[< z, X >] =< z, \mathbb{E}[X] >$ et de variance

$$Var(< z, X >) = \mathbb{E}\Big[\big(z^t X\big)^2\Big] - \big(z^t\, \mathbb{E}[X]\big)^2 = z^t\, \mathbb{E}[XX^t] z - z^t\, \mathbb{E}[X]\mathbb{E}[X]^t\, z = z^t\, Cov(X) z,$$

en remarquant que $< z, X >= z^t X = X^t z$. Par conséquent,

$$\varphi_X(z) = \varphi_{<z,X>}(1) = e^{i<z,\mathbb{E}[X]>} e^{-\frac{1}{2} z^t\, Cov(X) z} = e^{i<z,\mathbb{E}[X]>} e^{-\frac{1}{2} <z, Cov(X) z>}.$$

Supposons maintenant que X est un vecteur aléatoire de fonction caractéristique

$$\varphi_X(z) = \exp\left(i < z, m > -\frac{1}{2} < Kz, z >\right) = \exp\left(iz^t m - \frac{1}{2} z^t K z\right)$$

Soit $t \in \mathbb{R}$, alors

$$\varphi_{<z,X>}(t) = \varphi_X(tz) = \exp\left(i < z, m > t - \frac{1}{2} < Kz, z > t^2\right).$$

On en déduit que pour tout $z \in \mathbb{R}^d$, $< z, X >$ est une variable gaussienne d'espérance $< z, m >$ et de variance $< Kz, z >$. X est donc un vecteur gaussien.
Identifions maintenant m et K.
Par linéarité $\mathbb{E}[< z, X >] =< z, \mathbb{E}[X] >=< z, m >$ pour tout $z \in \mathbb{R}^d$ et notamment pour les vecteurs de la base canonique $(e_i)_{1 \leqslant i \leqslant d}$. On en déduit que $m = \mathbb{E}[X]$.
De même, on a $Var(< z, X >) =< z, Cov(X) z >=< Kz, z >$ pour tout $z \in \mathbb{R}^d$. En prenant $z = e_i + \alpha e_j$ et du fait que les matrices sont symétrique, on obtient $Cov(X)_{ii} + 2\alpha Cov(X)_{ij} + \alpha^2 Cov(X)_{jj} = K_{ii} + 2\alpha K_{ij} + \alpha^2 K_{jj}$ pour tout $\alpha \in \mathbb{R}$. Un polynôme qui a un infinité de racines est nul et donc $Cov(X)_{ii} = K_{ii}$, $Cov(X)_{ij} = K_{ij}$, $Cov(X)_{jj} = K_{jj}$. On obtient ainsi de suite $K = Cov(X)$. □

La loi d'un vecteur gaussien est entièrement déterminée à partir de son espérance et de sa matrice de covariance.

Definition 131. *La loi d'un vecteur gaussien de moyenne $m \in \mathbb{R}^d$ et de matrice de covariance K (matrice symétrique positive) est notée $\mathcal{N}_d(m, K)$.*

Remarque. — *Si $m = 0$ et $K = Id$, alors $\varphi_X(a) = e^{-||a||^2/2}$.*
— *Si $m = 0$, $Cov(X) = \mathbb{E}[XX^t]$ et donc $\varphi_X(a) = e^{-\mathbb{E}[<a,X>^2]/2}$.*

Proposition 132. *Si $X \sim \mathcal{N}_d(m, K)$ et $A \in \mathcal{M}_{p,d}(\mathbb{R})$, alors $AX \sim \mathcal{N}_p\big(Am, AKA^t\big)$.*

Démonstration. On sait que AX est un vecteur gaussien. Comme l'espérance est linéaire et que A est une application linéaire, $\mathbb{E}[AX] = A\mathbb{E}[X]$. Par ailleurs,

$$Cov(AX) = \mathbb{E}\big[A(X-m)(A(X-m))^t\big] = A\mathbb{E}\big[(X-m)(X-m)^t\big]A^t = ACov(X)A^t.$$

□

Definition 133. *Un vecteur gaussien $X \sim \mathcal{N}_d(m, K)$ est dit dégénéré si K est non inversible.*

Si X est dégénéré alors il existe $a \in \mathbb{R}^d$, $a \neq 0$ tel que $Ka = 0$. Par conséquent, $Var(< a, X >) = a^t\, K\, a = 0$ et donc $< a, X >= \mathbb{E}[< a, X >]$ p.s.
On remarque que X prend ses valeurs dans l'hyperplan d'équation $< a, x >= b$, avec $b = \mathbb{E}[< a, X >]$ et X vit dans un sous-espace affine de dimension inférieur ou égal à $d - 1$.
Par exemple, si $N = \mathcal{N}(m, \sigma^2)$ est une variable gaussienne en dimension 1, alors le vecteur gaussien $X = (N, \ldots, N)^t$ vit sur la droite $x_1 = x_2 = \ldots = x_d$.

Proposition 134. *Si $X \sim \mathcal{N}_d(m, K)$ est un vecteur gaussien non dégénéré, alors $\sqrt{K}^{-1}(X - m) \sim \mathcal{N}_d(0, Id)$.*

Démonstration. Comme X est non dégénéré, K est une matrice symétrique définie positive. Il existe $A = \sqrt{K}$ telle que $K = AA^t$. En effet comme K est diagonalisable à valeurs propres positives dans une base orthonormale, on a $K = P^t\, DP$ et il suffit de prendre $A = P^t\, D^{1/2}$. On a $K = \sqrt{K}\left(\sqrt{K}\right)^t$.
$Y = \sqrt{K}^{-1}(X - m)$ est un vecteur gaussien (comme image par une application affine de X) d'espérance 0 et de covariance $Cov(Y) = \sqrt{K}^{-1}K\left(\sqrt{K}^{-1}\right)^t = \sqrt{K}^{-1}\sqrt{K}\left(\sqrt{K}\right)^t\left(\sqrt{K}^{-1}\right)^t = Id.$ □

Remarque. *Si $X \sim \mathcal{N}_d(m, K)$ avec K inversible, alors on peut écrire $X = \sqrt{K}Z + m$ avec $Z \sim \mathcal{N}_d(0, Id)$.*

2.2 Indépendance

Proposition 135. *Soit (X, Y) un couple gaussien. Alors X et Y sont indépendantes si et seulement si $Cov(X, Y) = 0$.*

Démonstration. Si les variables sont indépendantes, il est clair que $Cov(X, Y) = 0$.
Inversement, si $Cov(X, Y) = 0$. Alors $(X, Y) \sim \mathcal{N}_2\left(m, \begin{bmatrix} \sigma_X^2 & 0 \\ 0 & \sigma_Y^2 \end{bmatrix}\right)$. La fonction caractéristique de (X, Y) est donc

$$\varphi_{(X,Y)}(t_1, t_2) = e^{im_X t_1} e^{im_Y t_2} e^{\frac{-1}{2}\sigma_X^2 t_1} e^{\frac{-1}{2}\sigma_Y^2 t_2} = \varphi_X(t_1)\varphi_Y(t_2).$$

Les variables sont indépendantes. □

Corollary 136. *Soit X un vecteur gaussien. Les coordonnées de X sont indépendantes si et seulement si la matrice de covariance $Cov(X)$ est diagonale.*

Corollary 137. *Soit $(X_1, \ldots, X_d, Y_1, \ldots, Y_p)^t$ un vecteur gaussien. Les vecteur $X = (X_1, \ldots, X_d)^t$ et $Y = (Y_1, \ldots, Y_p)^t$ sont indépendants si et seulement si $Cov(X_i, Y_j) = 0$ pour tous $i \neq j$.*

2.3 Densité gaussienne en dimension d

Soit X un vecteur gaussien de loi $\mathcal{N}_d(0, Id)$. Les coordonnées sont indépendantes donc $\mathbb{P}_X = \mathbb{P}_{X_1} \otimes \ldots \otimes \mathbb{P}_{X_d}$. Par conséquent,

$$\begin{aligned} f_X(x_1, \ldots, x_d) &= \frac{1}{(\sqrt{2\pi})^d} e^{\frac{-x_1^2}{2}} \ldots e^{\frac{-x_d^2}{2}} \\ &= \frac{1}{(\sqrt{2\pi})^d} e^{-\frac{1}{2}\sum_{i=1}^d x_i^2} = \frac{1}{(\sqrt{2\pi})^d} e^{-\frac{||x||^2}{2}}. \end{aligned}$$

Dans le cas général, $X \sim \mathcal{N}_d(m, K)$, si K est dégénéré, alors X ne peut pas avoir de densité par rapport à la mesure de Lebesgue sur $\mathbb{R}^d$ car il vit dans un sous-espace de dimension strictement inférieure à d (qui est donc de mesure nulle pour Lebesgue sur $\mathbb{R}^d$). On pourrait cependant étudier le vecteur dnas le sous-espace de dimension inférieure dans lequel il vit.

Proposition 138. *Soit $X \sim \mathcal{N}_d(m, K)$ un vecteur gaussien.*
Le vecteur X admet une densité par rapport à la mesure de Lebesgue sur $\mathbb{R}^d$ si et seulement si K est inversible.
Lorsque K est inversible, la densité sur $\mathbb{R}^d$ par rapport à la mesure de Lebesgue, donnée par

$$f_X(x) = \frac{1}{\left(\sqrt{2\pi}\right)^d \sqrt{\det(K)}} e^{-\frac{(x-m)^t K^{-1}(x-m)}{2}}.$$

Démonstration. Comme K inversible, on peut écrire $X = \sqrt{K}Z + m$ avec $Z \sim \mathcal{N}_d(0, Id)$. Soit $h : \mathbb{R}^d \to \mathbb{R}$ une fonction mesurable bornée. En effectuant le changement de variables $x = \sqrt{K}z + m$ qui est bien un C^1-difféomorphisme car K est non dégénérée et en remarquant que $\det(\sqrt{K}) = \sqrt{\det(K)}$, on obtient

$$\begin{aligned} \mathbb{E}[h(X)] = \mathbb{E}\Big[h(\sqrt{K}Z + m)\Big] &= \frac{1}{\left(\sqrt{2\pi}\right)^d} \int_{\mathbb{R}^d} h(\sqrt{K}z + m) e^{-\frac{||z||^2}{2}} dz \\ &= \frac{1}{\left(\sqrt{2\pi}\right)^d \sqrt{\det(K)}} \int_{\mathbb{R}^d} h(x) e^{-\frac{||\sqrt{K}^{-1}(x-m)||^2}{2}} dx \\ &= \frac{1}{\left(\sqrt{2\pi}\right)^d \sqrt{\det(K)}} \int_{\mathbb{R}^d} h(x) e^{-\frac{(x-m)^t K^{-1}(x-m)}{2}} dx. \end{aligned}$$

□

Exemple. Toutes les coordonnées d'un vecteur gaussien sont gaussiennes, mais la réciproque est fausse. Considérons un vecteur à coordonnées gaussiennes. Soit X une v.a. de loi $\mathcal{N}(0,1)$ et ε v.a. discrète avec $\mathbb{P}(\varepsilon = 1) = \mathbb{P}(\varepsilon = -1) = 1/2$. On pose $Y = \varepsilon X$. La fonction caractéristique de Y est $\varphi_Y(t) = \varphi_X(t)/2 + \varphi_X(-t)/2 = \varphi_X(t)$ et donc $Y \sim \mathcal{N}(0,1)$. Par contre le vecteur (X, Y) n'est pas gaussien car $X + Y$ a un atome en 0 (en effet $\mathbb{P}(X + Y = 0) = \mathbb{P}(\varepsilon = -1) = 1/2$). De même on peut remarquer que X, Y ne sont pas indépendants, car $\mathbb{P}(|X| < 1/4, |Y| > 3/4) = 0 \neq \mathbb{P}(|X| < 1/4)\mathbb{P}(|Y| > 3/4)$ et $Cov(X,Y) = \mathbb{E}[\varepsilon]\mathbb{E}[X^2] = 0$.

3 Théorème Central Limite multidimensionnel

Soit X un vecteur aléatoire de $\mathbb{R}^d$, dans L^1 de moyenne m. On considère $(X_n)_{n\geqslant 1}$ une suite de vecteurs aléatoires de même loi que X. Comme la convergence p.s. d'un vecteur est équivalente à la convergence p.s. des marginales, d'après la loi forte des grands nombres on a

$$\frac{X_1+\ldots+X_n}{n} \xrightarrow{p.s} m.$$

On ne peut pas faire la même chose avec le théorème central limite car la convergence en loi des marginales n'implique pas la convergence en loi du vecteur. Il existe cependant un théorème central limite multidimensionnel.

Theorem 139. *Soit X un vecteur aléatoire de $\mathbb{R}^d$, dans L^2. On considère $(X_n)_{n\geqslant 1}$ une suite de vecteurs aléatoires indépendants de même loi que X. On note $m=\mathbb{E}[X]$ et K la matrice de Covariance de X. On suppose K non dégénérée. Alors*

$$\sqrt{n}\sqrt{K}^{-1}\left(\frac{X_1+\ldots+X_n}{n}-m\right) \xrightarrow{Loi} \mathcal{N}(0,Id),$$

ce qui peut se réécrire $\sqrt{n}\big(\frac{X_1+\ldots+X_n}{n}-m\big) \xrightarrow{Loi} \mathcal{N}(0,K)$.

Démonstration. On pose $Z_n=\sqrt{n}\sqrt{K}^{-1}\big(\frac{X_1+\ldots+X_n}{n}-m\big)$. La fonction caractéristique de Z_n au point $t\in\mathbb{R}^d$ est définie par

$$\varphi_{Z_n}(z)=\mathbb{E}[e^{<Z_n,z>}].$$

En posant $u=(\sqrt{K}^{-1})^t z$, on remarque qu'on a

$$\begin{aligned}< Z_n,z > &= \sqrt{n} < \sqrt{K}^{-1}\left(\frac{X_1+\ldots+X_n}{n}-m\right), z >= \sqrt{n}z^t\sqrt{K}^{-1}\left(\frac{X_1+\ldots+X_n}{n}-m\right)\\ &= \sqrt{n}^{-1}\sum_{i=1}^{n} < X_i-m,u > .\end{aligned}$$

Les variables $< X_i-m,u >$. sont i.i.d d'espérance $\mathbb{E}[< X_i-m,u >] =< \mathbb{E}[X_i]-m,u >= 0$ et de variance

$$\begin{aligned}Var(< X_i-m,u >) = VAR(< X_i,u >) &= \mathbb{E}\left[\left(z^t\sqrt{K}^{-1}X\right)^2\right]\\ &= \mathbb{E}\left[z^t\sqrt{K}^{-1}XX^t\left(\sqrt{K}^{-1}\right)^t z\right] = \mathbb{E}\left[z^t\sqrt{K}^{-1}K\left(\sqrt{K}^{-1}\right)^t z\right]\\ &= z^t z = ||z||^2.\end{aligned}$$

D'après le théorème central limite en dimension 1, on a

$$< Z_n,t >\xrightarrow{Loi} \mathcal{N}(0,||z||^2).$$

Par conséquent,

$$\varphi_{<Z_n,z>}(1)=\mathbb{E}\big[e^{<Z_n,z>}\big] \underset{n\to+\infty}{\longrightarrow} e^{-\frac{||z||^2}{2}}.$$

$z \mapsto e^{-\frac{||z||^2}{2}}$ est la fonction caractéristique de $\mathcal{N}(0,Id)$ et donc d'après le théorème de Paul Lévy, $\sqrt{n}\sqrt{K}^{-1}\big(\frac{X_1+\ldots+X_n}{n}-m\big) \xrightarrow{Loi} \mathcal{N}(0,Id)$. ☐

Application Loi multinomiale

On considère une urne qui contient des boules numérotées de 1 à d. La proportion de boule de numéro i est p_i. On note $p = (p_1, \ldots, p_d)^t$ le vecteur de probabilité. On a $0 \leqslant p_i \leqslant 1$ et $\sum_{i=1}^{d} p_i = 1$.
On tire une boule au hasard, on note son numéro et on la remet dans l'urne. On répète n fois l'expérience. On comptabilise le nombre de boules tirées de chaque numéro dans un vecteur

$$N = \begin{pmatrix} N_1 & = & \text{nombre de fois qu'on a tiré la boule } 1 \\ & \vdots & \\ N_i & = & \text{nombre de fois qu'on a tiré la boule } i \\ & \vdots & \\ N_d & = & \text{nombre de fois qu'on a tiré la boule } d \end{pmatrix}$$

Les variables $N_1, \ldots, N_d$ ne sont pas indépendantes car $N_1 + \ldots + N_d = n$.
<u>But</u> : Étudier le comportement de $\frac{N}{n}$ quand $n \to \infty$.
On remarque que pour tout $1 \leqslant i \leqslant d$, N_i suit la loi binomiale $\mathcal{B}(n, p_i)$ car N_i représente le nombre de boules i tirées sur n tirages indépendants de la même expérience.
Le vecteur N suit la loi multinomiale $\mathcal{M}(n, p)$. En effet, pour $(k_1, \ldots, k_d) \in \mathbb{N}^d$ avec $\sum_{i=1}^{d} k_i = n$, on a

$$\begin{aligned} \mathbb{P}((N_1, \ldots, N_d) = (k_1, \ldots, k_d)) &= \binom{n}{k_2}\binom{n-k_1}{k_2} \ldots \binom{n-(k_1+\ldots+k_{d_1})}{k_d} p_1^{k_1} \ldots p_d^{k_d} \\ &= \frac{n!}{k_1! \ldots k_d!} p_1^{k_1} \ldots p_d^{k_d}. \end{aligned}$$

On introduit la variable X^n représentant le numéro de la boule tirée au $n^{\text{ème}}$ tirage. X^n est à valeurs dans $\{1, \ldots, d\}$. On pose $Y^n = (Y_1^n, \ldots, Y_d^n)$ avec $Y_i^n = \mathbb{1}_{X^n = i}$. Les vecteurs Y^n sont indépendants de loi multinomiale $\mathcal{M}(p, 1)$.
Par ailleurs, $\frac{N_i}{n} = \frac{1}{n} \sum_{k=1}^{n} \mathbb{1}_{X^k = i}$ qui converge p.s. vers p_i pour tout $i \in \{1, \ldots, d\}$ d'après la loi des grands nombres. On en déduit que $\frac{N}{n} \xrightarrow{p.s} p$.
D'après le théorème central limite en dimension 1, on a $\sqrt{n}\Big(\frac{N_i}{n} - p_i\Big) \xrightarrow{Loi} \mathcal{N}(0, p_i(1-p_i))$ pour tout $i \in \{1, \ldots, d\}$. Cependant la converge en loi des marginales ne permet pas de conclure sur la convergence en loi du vecteur $\sqrt{n}\big(\frac{N}{n} - p\big)$. On va avoir besoin du théorème central limite multidimensionnel.

On remarque que, $\mathbb{E}[Y^n] = p$ et $Cov(Y_i^n, Y_j^n) = \mathbb{E}[\mathbb{1}_{X^n=i}\mathbb{1}_{X^n=j}] - \mathbb{E}[\mathbb{1}_{X^n=i}]\mathbb{E}[\mathbb{1}_{X^n=j}] = -p_i p_j$ si $i \neq j$ et $Var(Y_i^n) = p_i(1-p_i)$. La matrice de covariance de Y^n est donc

$$K = \begin{pmatrix} p_1(1-p_1) & & & & \\ & \ddots & & & \\ & -p_i p_j & pi(1-p_i) & -p_i p_j & \\ & & & \ddots & \\ & & & & p_d(1-p_d) \end{pmatrix}.$$

D'après le théorème central limite multidimensionnel, on a

$$\sqrt{n}\left(\frac{1}{n}\sum_{k=1}^{n} Y^k - p\right) = \sqrt{n}\left(\frac{N}{n} - p\right) \xrightarrow{Loi} \mathcal{N}(0, K).$$

4 Loi du Chi-deux et théorème de Cochran

Definition 140. *Soit $X = (X_1, \dots, X_d)$ un vecteur de variables $\mathcal{N}(0,1)$ indépendantes. On définit la loi du Chi-deux, notée χ^2_d ou $\chi^2(d)$, à $d-$degrés de liberté la loi de $||X||^2 = X_1^2 + \ldots + X_d^2$.*

Proposition 141. *La loi du Chi-deux à $d-$degrés de liberté est la loi Gamma$\big(\frac{d}{2}, \frac{1}{2}\big)$ de densité*

$$f(x) = \frac{2^{\frac{-d}{2}}}{\Gamma\big(\frac{d}{2}\big)} x^{\frac{d}{2}-1} e^{-\frac{x}{2}} \mathbb{1}_{\mathbb{R}^+}(x).$$

Démonstration. Soit Z une variable $\mathcal{N}(0,1)$. Cherchons la loi de Z^2. Soit h une fonction mesurable positive, alors

$$\begin{aligned}\mathbb{E}[h(Z^2)] &= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} h(z^2) e^{-\frac{z^2}{2}} dz = \frac{2}{\sqrt{2\pi}} \int_0^\infty h(z^2) e^{-\frac{z^2}{2}} dz \\ &= \frac{2}{\sqrt{2\pi}} \int_0^\infty h(x) e^{-\frac{x}{2}} \frac{1}{2\sqrt{x}} dx.\end{aligned}$$

On remarque que Z^2 suit la loi Gamma$\big(\frac{1}{2}, \frac{1}{2}\big)$. Par conséquent, par indépendance des variables, on a déduit que $||X||^2 = X_1^2 + \ldots + X_d^2$ suit la loi Gamma$\big(\frac{d}{2}, \frac{1}{2}\big)$. □

Theorem 142 (Théorème de Cochran)**.** *Soit X un vecteur gaussien de loi $\mathcal{N}_d(0, \sigma^2 Id)$ avec $\sigma > 0$. On considère une décomposition orthogonale de $\mathbb{R}^d$:*

$$\mathbb{R}^d = V_1 \oplus \ldots . \oplus V_k \text{ avec } dim(V_i) = d_i \text{ et } V_i \perp V_j \text{ pour } i \neq j.$$

On note Π_V la projection orthogonale sur le sous-espace V.
Alors les vecteurs $\Pi_{V_1}(X), \dots, \Pi_{V_k}(X)$ sont gaussiens indépendants et

$$\frac{1}{\sigma^2} ||\Pi_{V_i}(X)||^2 \sim \chi^2_{d_i}, \text{ pour } i = 1, \dots, k.$$

Démonstration. Soit $(e_{i,j})_{i,j}$ une base orthonormée de $\mathbb{R}^d$ adaptée à la décomposition $\mathbb{R}^d = V_1 \oplus \ldots . \oplus V_k$, avec $\{e_{i,j} : j = 1, \dots, d_i\}$ est une base de V_i.
Le vecteur X se décompose dans la base $(e_{i,j})_{i,j}$: $X = \sum_{i,j} < X, e_{i,j} > e_{i,j}$.
On a $X = \sum_{i,j} < X, e_{i,j} > e_{i,j} \sim \mathcal{N}_d(0, \sigma^2 Id)$, par conséquent les vecteurs $\Pi_{V_i}(X) = \sum_{j=1}^{d_i} < X, e_{i,j} > e_{i,j}$ sont indépendants car les projections sont construites à partir d'une partition de la base $(e_{i,j})_{i,j}$ et

$$\frac{1}{\sigma^2} ||\Pi_{V_i}(X)||^2 = \sum_{j=1}^{d_i} < X/\sigma, e_{i,j} >^2 \sim \chi^2_{d_i},$$

□

Application Reprenons l'exemple de la loi multinomiale. On a

$$\sqrt{n} \left(\frac{N}{n} - p \right) \xrightarrow{Loi} \mathcal{N}(0, K)$$

On introduit la matrice diagonale $D = diag\big(1/\sqrt{p_1}, \dots, 1/\sqrt{p_d}\big)$. On a

$$\sqrt{n} D \left(\frac{N}{n} - p \right) \xrightarrow{Loi} \mathcal{N}(0, \Sigma)$$

avec

$$\Sigma = DKD = \begin{pmatrix} 1-p_1 & & & & \\ & \ddots & & & \\ & -\sqrt{p_i p_j} & 1-p_i & & -\sqrt{p_i p_j} \\ & & & \ddots & \\ & & & & 1-p_d \end{pmatrix} = Id - \sqrt{p}\sqrt{p}^t \text{ où } \sqrt{p} = \begin{pmatrix} \sqrt{p_1} \\ \vdots \\ \sqrt{p_d} \end{pmatrix}$$

Comme $||\sqrt{p}||^2 = 1$, la matrice Σ est une matrice de la projection orthogonale sur $Vect(\sqrt{p})^\perp$ ($\Sigma^2 = \Sigma$ et $\Sigma^t = \Sigma$). Par conséquent, si $X \sim \mathcal{N}(0, Id)$, alors $\Sigma X \sim \mathcal{N}(0, \Sigma)$.
Par ailleurs, $\sqrt{n} D\big(\frac{N}{n} - p\big) \xrightarrow{Loi} \Sigma X = \Pi_{Vect(\sqrt{p})^\perp}(X)$. D'après le théorème de Cochran, on en déduit que

$$n||D\Big(\frac{N}{n} - p\Big)||^2 = \sum_{k=1}^d \frac{(N_k - np_k)^2}{np_k} \xrightarrow{Loi} \chi^2_{d-1}.$$

Theorem 143. *On considère deux vecteurs de probabilité $\pi = (\pi_1, \ldots, \pi_d)^t$, $p = (p_1, \ldots, p_d)^t$ ($0 < \pi_k < 1$ avec $\sum_{k=1}^d \pi_i = 1$ et $0 < p_k < 1$ vec $\sum_{k=1}^d p_i = 1$).*
Soit X une variable discrète à valeurs dans $\{a_1, \ldots, a_d\}$ de loi π : $\mathbb{P}(X = a_k) = \pi_k$. On considère $(X_n)_{n\geqslant 1}$ une suite de v.a. indépendantes suivant la loi de X. On introduit la variable

$$T_n = \sum_{k=1}^d \frac{(N_k - np_k)^2}{np_k},$$

où $N_k = \sum_{i=1}^n \mathbb{1}_{X_i = a_k}$.
Sous l'hypothèse $\pi = p$, on a $T_n \xrightarrow{Loi} \chi^2_{d-1}$.
Sous l'hypothèse $\pi \neq p$, on a $T_n \xrightarrow{p.s} +\infty$.

Ce résultat est très utile en statistique pour évaluer si une variable aléatoire suit une loi donnée. On parle de test du χ^2 d'adéquation à une loi. On se fixe une loi finie p et on considère un échantillon $(X_1, \ldots, X_n)$ du caractère que l'on souhaite étudier. On évalue la quantité T_n. Si la valeur observée de T_n est grande, on aura tendance à rejeter l'hypothèse que notre échantillon est issu de la loi p.

Démonstration. Sous l'hypothèse $\pi = p$, on a déjà montrer juste au dessus que $T_n \xrightarrow{Loi} \chi^2_{d-1}$
Sous l'hypothèse $\pi \neq p$. Comme $(X_n)_{n\geqslant 1}$ une suite de v.a. indépendantes de loi π, d'après la loi forte des grands nombres, on a

$$\frac{N_k}{n} \xrightarrow{p.s} \pi_k.$$

Par ailleurs, comme

$$T_n = n \sum_{k=1}^d \frac{\left(\frac{N_k}{n} - p_k\right)^2}{p_k},$$

on a

$$\sum_{k=1}^d \frac{\left(\frac{N_k}{n} - p_k\right)^2}{p_k} \xrightarrow{p.s} \sum_{k=1}^d \frac{(\pi_k - p_k)^2}{p_k}$$

On a supposé $\pi \neq p$, ce qui implique que la limite précédente est strictement positive et donc

$$T_n \xrightarrow{p.s} +\infty.$$

$\square$

5 Application en statistique : moyenne et variance empirique

On considère une suite $(X_n)_{n\geqslant 1}$ de variables aléatoires réelles i.i.d. dans L^4, de moyenne commune m et de variance σ^2.
La moyenne empirique sont respectivement définies par

$$\overline{X}_n = \frac{X_1 + \ldots + X_n}{n},$$
$$S_n^2 = \frac{1}{n}\sum_{k=1}^n (X_k - \overline{X}_n)^2 = \frac{1}{n}\sum_{k=1}^n X_k^2 - \overline{X}_n^2.$$

On sait que ce sont de bons estimateurs de m et de σ^2, car $\overline{X}_n \xrightarrow{p.s} m$, $\mathbb{E}\big[\overline{X}_n\big] = m$ et $S_n^2 \xrightarrow{p.s} \sigma^2$, $\mathbb{E}\big[S_n^2\big] = \frac{n-1}{n}\sigma^2$. Par ailleurs, ces estimateurs sont asymptotiquement normaux, dans le sens suivant :

$$\sqrt{n}\big(\overline{X}_n - m\big) \xrightarrow{Loi} \mathcal{N}(0,\sigma^2),$$
$$\sqrt{n}\big(S_n^2 - \sigma^2\big) \xrightarrow{Loi} \mathcal{N}(0,\alpha),$$

avec $\alpha = Var((X-m)^2) = \mathbb{E}[(X-m)^4] - \sigma^4$.

Démonstration. La première convergence est une application directe du théorème central limite. Pour la seconde, on remarque tout d'abord que

$$S_n^2 = \frac{1}{n}\sum_{k=1}^n \Big((X_k - m) - \overline{(X-m)}_n\Big)^2 = \frac{1}{n}\sum_{k=1}^n (X_k - m)^2 - \overline{(X-m)}_n^2,$$

on peut donc supposer que $m = 0$. On a

$$\sqrt{n}\big(S_n^2 - \sigma^2\big) = \sqrt{n}\left(\frac{1}{n}\sum_{k=1}^n X_k^2 - \sigma^2\right) - \sqrt{n}\overline{X}_n^2.$$

Comme les variables sont dans L^4, par la théorème central limite, on sait que

$$\sqrt{n}\left(\frac{1}{n}\sum_{k=1}^n X_k^2 - \sigma^2\right) \xrightarrow{Loi} \mathcal{N}(0, Var(X^2)),$$

avec $Var(X^2) = \mathbb{E}[X^4] - \sigma^4 = \alpha$.
Par ailleurs, comme $\overline{X}_n \xrightarrow{Proba} 0$ et $\sqrt{n}\overline{X}_n \xrightarrow{Loi} \mathcal{N}(0,\sigma^2)$ d'après la L.G.N. et le T.C.L., on obtient, en utilisant le théorème de Slutsky, que $\sqrt{n}\overline{X}_n^2 \xrightarrow{Loi} 0$ et donc $\sqrt{n}\overline{X}_n^2 \xrightarrow{Proba} 0$.
On obtient le résultat en appliquant à nouveau le théorème de Slutsky. □

Dans le cas où l'échantillon $(X_n)_{n\geqslant 1}$ est gaussien, on a une résultat plus précis.

Theorem 144. *Soit $(X_n)_{n\geqslant 1}$ une suite de variables i.i.d. de loi $\mathcal{N}(m,\sigma^2)$. Alors*
- *$\overline{X}_n$ et S_n^2 sont indépendantes,*
- *$\overline{X}_n \sim \mathcal{N}(m, \sigma^2/n)$ et $\frac{n}{\sigma^2}S_n^2 \sim \chi^2_{n-1}$.*

Démonstration. On peut se ramener à $m = 0$ et $\sigma^2 = 1$ car $X'_n = (X_n - m)/\sigma$ i.i.d. de loi $\mathcal{N}(0,1)$ et $\overline{X}_n = \sigma \overline{X'}_n + m$ et $S_n^2 = \sigma^2 S_n'^2$.
On note $X = (X_1, \dots, X_n)^t$. On a donc $X \sim \mathcal{N}(0, Id)$.
On considère le vecteur $u_1 = (\frac{1}{\sqrt{n}}, \dots, \frac{1}{\sqrt{n}}) \in \mathbb{R}^n$ et $(u_1, \dots, u_n)$ une base orthonormée de $\mathbb{R}^n$.
On note $Z_1 = \Pi_{Vect(u_1)}(X) =< u_1, X > u_1$ et $Z_2 = \Pi_{Vect(u_1)^\perp}(X)$. D'après le théorème de Cochran Z_1 et Z_2 sont des variables gaussiennes indépendantes. Par ailleurs $||Z_2||_2^2 = ||X||^2 - ||Z_1||^2 = \sum_{k=1}^n X_k^2 - n\overline{X}_n^2 = nS_n^2$.
On en déduit que $\overline{X}_n = (1, \dots, 1).Z_1$ et $S_n^2 = ||Z_2||_2^2/n$ sont indépendantes et $nS_n^2 \sim \chi^2(n-1)$. $\square$

Bibliographie

[1] Barbe, Philippe ; Ledoux, Michel. TASSI, *Probabilité*, EDP Sciences, 2007

[2] Billingsley, Patrick. *Probability and measure. Third edition.* Wiley Series in Probability and Mathematical Statistics. A Wiley-Interscience Publication. John Wiley & Sons, Inc., New York, 1995.

[3] Breton, Jean-Christophe. https://perso.univ-rennes1.fr/jean-christophe.breton/Fichiers/proba_base.pdf, 2014.

[4] Briand, Philippe. https://www.lama.univ-savoie.fr/ briand/proba/g12_cours.pdf, 2004.

[5] Durrett, Rick. *Probability : theory and examples. Fourth edition.* Cambridge Series in Statistical and Probabilistic Mathematics, 31. Cambridge University Press, Cambridge, 2010.

[6] Jacod, Jean ; Protter, Philip. *Probability essentials. Second edition.* Universitext. Springer-Verlag, Berlin, 2003.