

Notes de cours - Probabilités

Adrien Richou

Introduction

L'objet de la théorie des probabilités est de fournir des modèles mathématiques permettant l'étude d'expériences dont le résultat n'est pas connu ou ne peut pas être prévu avec une totale certitude. Voici quelques exemples de la vie courante :

Expérience	Résultat observable
Lancer d'un dé	Un entier $k \in \{1, \dots, 6\}$
Lancer d'une pièce jusqu'à l'obtention d'un pile	Un entier $k \in \mathbb{N}$: le temps d'attente du premier succès
Mise en service d'une ampoule	Durée de vie $T \in \mathbb{R}^+$
Lancer d'une flèche sur une cible	Point d'impact $M \in \mathbb{R}^2$
Évolution temporelle de la température d'une pièce pendant une journée	Une courbe réelle continue

Le résultat précis de ces expériences n'est en général pas prévisible. Toutefois, l'observation et/ou l'intuition amènent souvent à prévoir certains comportements. Par exemple, si on jette 6000 fois un dé à 6 faces, on s'attend à ce que le nombre total de 4 soit voisin de 1000. La théorie des probabilités permet de donner un sens mathématique rigoureux à ces constatations empiriques.

En aval des probabilités se trouve la statistique qui permet de confronter les modèles probabilistes à la réalité observée.

Quelques références

- *Probabilités 1 & 2*, J.-Y. Oувrard
- *De l'intégration aux probabilités*, O. Garet et A. Kutzmann
- *Probabilités*, P. Barbé et M. Ledoux

1 Introduction et rappels, espaces de probabilités

1.1 Vocabulaire et exemples

On s'intéresse à une expérience aléatoire, c'est-à-dire dont on ne connaît pas le résultat de manière certaine.

Définition 1 On appelle expérience aléatoire toute expérience \mathcal{E} conduisant, selon le hasard, à plusieurs résultats possibles. On appelle univers associé à \mathcal{E} , l'ensemble Ω de tous les résultats possibles de \mathcal{E} .

Exemple 1

- L'expérience \mathcal{E} consiste à lancer un dé. $\Omega_1 = \{1, \dots, 6\}$ est fini.
- L'expérience \mathcal{E} consiste à jouer à pile ou face jusqu'à l'obtention d'un pile. $\Omega_2 = \{P, FP, FFP, FFFP, \dots\}$ est infini dénombrable.
- L'expérience \mathcal{E} consiste à évaluer la durée de vie d'une étoile dans notre galaxie. $\Omega_3 =]0, +\infty[$ est infini non dénombrable.

Exemple du lancer de dé L'univers est donné par $\Omega = \{1, \dots, 6\}$.

$\omega \in \Omega$ est appelé une réalisation de l'expérience (ou évènement élémentaire).

Exemples d'évènements (aléatoires) :

- Le dé vaut 6, $\{6\}$
- Le dé est paire, $\{2, 4, 6\}$

Donc un évènement correspond à un sous-ensemble de Ω . Pour cet exemple, on peut prendre $\mathcal{P}(\Omega)$ (i.e. l'ensemble des parties de Ω) pour l'ensemble des évènements aléatoires. Dans le cas général, l'ensemble des évènements aléatoires est une tribu.

Une probabilité \mathbb{P} est une mesure qui, à chaque évènement aléatoire, associe une note entre 0 et 1, de façon "cohérente".

Ici, si on lance le dé N fois et si on note $N(i)$ le nombre d'apparition de la face i , alors empiriquement on a

$$\frac{N(i)}{N} \xrightarrow{N \rightarrow +\infty} \frac{1}{6}.$$

On pose

$$\mathbb{P}(\{1\}) = \mathbb{P}(\{2\}) = \dots = \mathbb{P}(\{6\}) = \frac{1}{6}.$$

1.2 Rappels sur les ensembles

Les évènements aléatoires étant des ensembles, on peut considérer des opérations sur ces évènements aléatoires.

- $A \cup B$: A ou B réalisé.

- $A \cap B$: A et B réalisés.

- A^c ou \bar{A} : A n'est pas réalisé.

Définition 2 A et B sont disjoints ou incompatibles si $A \cap B = \emptyset$.

Si $(A_n)_{n \in \mathbb{N}^*}$ est une suite d'évènements, alors

- $\bigcup_{n \geq 1} A_n =$ "Au moins un des évènements est réalisé"
- $\bigcap_{n \geq 1} A_n =$ "Tous les évènements sont réalisés"

Proposition 1 (règles de calcul)

- $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$ et plus généralement on a

$$A \cap \left(\bigcup_{n \geq 1} B_n \right) = \bigcup_{n \geq 1} (A \cap B_n).$$

- $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$ et plus généralement on a

$$A \cup \left(\bigcap_{n \geq 1} B_n \right) = \bigcap_{n \geq 1} (A \cup B_n).$$

- $(A \cap B)^c = A^c \cup B^c$ et plus généralement

$$\left(\bigcap A_n \right)^c = \bigcup A_n^c$$

- $(A \cup B)^c = A^c \cap B^c$ et plus généralement

$$\left(\bigcup A_n\right)^c = \bigcap A_n^c$$

- $(A^c)^c$.

1.3 tribus, mesures de probabilités

On note $\mathcal{P}(\Omega)$ l'ensemble des parties de Ω . Un sous-ensemble \mathcal{A} de $\mathcal{P}(\Omega)$ est un ensemble de parties de Ω .

Définition 3 Soit \mathcal{A} un sous ensemble de $\mathcal{P}(\Omega)$. On dit que \mathcal{A} est une σ -algèbre ou tribu si

1. $\Omega \in \mathcal{A}$,
2. \mathcal{A} est stable par passage au complémentaire : si $A \in \mathcal{A}$, alors ${}^c A = \bar{A} \in \mathcal{A}$,
3. \mathcal{A} est stable par réunion dénombrable : si $(A_n)_{n \in \mathbb{N}}$ est une famille dénombrable d'ensembles de \mathcal{A} , alors $\bigcup_{n \in \mathbb{N}} A_n \in \mathcal{A}$.

Définition 4 Soient Ω un univers et \mathcal{A} une tribu associée. Le couple (Ω, \mathcal{A}) est appelé un espace mesurable ou espace probabilisable.

Remarque 1

- Un espace probabilisable est un espace sur lequel on va pouvoir considérer une mesure de probabilité.
- Dans la suite, l'ensemble des événements aléatoires associés à Ω est donné par une tribu \mathcal{A} . En particulier, si $\mathcal{A} \neq \mathcal{P}(\Omega)$, alors certaines parties de Ω ne sont pas des événements aléatoires et donc on ne pourra pas calculer leur probabilité.

Remarque 2 Si \mathcal{A} est une tribu, alors $\Omega^c = \emptyset \in \mathcal{A}$.

Proposition 2 Une tribu \mathcal{A} est stable par intersection dénombrable.

Preuve. $(A \cap B)^c = A^c \cup B^c$ donc $A \cap B = (A^c \cup B^c)^c$. Si on considère une suite d'évènements aléatoires $A_n \in \mathcal{A}, \forall n \geq 1$, alors

$$\bigcap_{n \geq 1} A_n = \left(\bigcup_{n \geq 1} A_n^c \right)^c \in \mathcal{A}.$$

□

Remarque 3 (Exemples de tribus)

- $\{\emptyset, \Omega\}$ est une tribu appelée tribu triviale, c'est la plus petite tribu.
- $\mathcal{P}(\Omega)$ est toujours une tribu. C'est la plus grosse tribu, elle contient toutes les autres.
- Si $A \subset \Omega$, $\mathcal{A} = \{\emptyset, A, A^c, \Omega\}$ est la plus petite tribu contenant A .

Proposition 3 Une intersection de tribus est encore une tribu.

Preuve. Exercice

□

Définition 5 Soit \mathcal{A} un sous-ensemble de $\mathcal{P}(\Omega)$. On appelle tribu engendrée par \mathcal{A} , notée $\sigma(\mathcal{A})$, la plus petite tribu contenant \mathcal{A} .

$\sigma(\mathcal{A})$ existe, c'est l'intersection de toutes les tribus contenant \mathcal{A} (l'intersection se fait sur au moins un élément car $\mathcal{P}(\Omega)$ est une tribu contenant \mathcal{A}).

Exemple 2 Si A est une partie de Ω , alors

$$\sigma(\{A\}) = \{\emptyset, A, \bar{A}, \Omega\}, \quad \sigma(\{A, \bar{A}\}) = \{\emptyset, A, \bar{A}, \Omega\}.$$

Si Ω est discret (fini ou dénombrable), on pourra toujours prendre $\mathcal{A} = \mathcal{P}(\Omega)$. Si $\Omega = \mathbb{R}^d$, il faut faire plus attention.

Définition 6 (et proposition) La tribu borélienne de \mathbb{R}^d , notée $\mathcal{B}(\mathbb{R}^d)$ est la tribu engendrée par les ensembles ouverts de \mathbb{R}^d . Elle coïncide avec

- la tribu engendrée par les ensembles fermés de \mathbb{R}^d ,
- la tribu engendrée par les pavés $[a_1, b_1] \times \dots \times [a_d, b_d]$,
- la tribu engendrée par les pavés $]a_1, b_1[\times \dots \times]a_d, b_d[$,
- la tribu engendrée par les pavés $]a_1, b_1] \times \dots \times]a_d, b_d]$,
- la tribu engendrée par les pavés $]a_1, b_1[\times \dots \times]a_d, b_d[$,
- la tribu engendrée par les ensembles $[a_1, +\infty[\times \dots \times [a_d, +\infty[$.

Remarque 4 $\mathcal{B}(\mathbb{R}^d) \neq \mathcal{P}(\mathbb{R}^d)$

Définition 7 Soit (Ω, \mathcal{A}) un espace mesurable ou probabilisable. On appelle probabilité sur (Ω, \mathcal{A}) toute application $\mathbb{P} : \mathcal{A} \rightarrow [0, 1]$ telle que

1. $\mathbb{P}(\Omega) = 1$,
2. Pour toute famille $(A_n)_{n \in \mathbb{N}}$ d'ensembles de \mathcal{A} deux à deux disjoints, on a

$$\mathbb{P}\left(\bigcup_{n \in \mathbb{N}} A_n\right) = \sum_{n \in \mathbb{N}} \mathbb{P}(A_n) \quad (\sigma\text{-additivité}).$$

Remarque 5 Une probabilité est une mesure positive de masse totale égale à 1.

Définition 8 Un triplet $(\Omega, \mathcal{A}, \mathbb{P})$ est appelé espace probabilisé ou espace de probabilité.

Dans toute la suite on se place dans un espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$ donné.

Proposition 4 Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité, $A, B \in \mathcal{A}$, $(A_n)_{n \in \mathbb{N}}$ une famille dénombrable d'ensembles de \mathcal{A} .

1. $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$,
2. $\mathbb{P}(A) = \mathbb{P}(A \cap B) + \mathbb{P}(A \cap B^c)$,
3. si $A \subset B$, $\mathbb{P}(A) \leq \mathbb{P}(B)$,
4. $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$,
- 5.

$$\mathbb{P}\left(\bigcup_{n \in \mathbb{N}} A_n\right) \leq \sum_{n \in \mathbb{N}} \mathbb{P}(A_n),$$

6. si $(A_n)_{n \in \mathbb{N}}$ est croissante, i.e. $A_n \subset A_{n+1}$, alors on a

$$\mathbb{P}\left(\bigcup_{n \in \mathbb{N}} A_n\right) = \lim_{n \in \mathbb{N}} \mathbb{P}(A_n),$$

7. si $(A_n)_{n \in \mathbb{N}}$ est décroissante, alors on a

$$\mathbb{P}\left(\bigcap_{n \in \mathbb{N}} A_n\right) = \lim_{n \in \mathbb{N}} \mathbb{P}(A_n).$$

Preuve.

1. $1 = \mathbb{P}(\Omega) = \mathbb{P}(A \cup A^c) = \mathbb{P}(A) + \mathbb{P}(A^c)$ car l'union est disjointe.
2. $A = A \cap (B \cup B^c) = (A \cap B) \cup (A \cap B^c)$ et l'union est disjointe.
3. $B = (B \cap A) \cup (B \cap A^c)$ et l'union est disjointe. Donc

$$\mathbb{P}(B) = \mathbb{P}(A \cap B) + \mathbb{P}(B \cap A^c) = \mathbb{P}(A) + \mathbb{P}(B \cap A^c) \geq \mathbb{P}(A).$$

4. $A \cup B = A \cup (B \cap A^c)$ donc $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B \cap A^c)$.
De plus on a également $B = (B \cap A) \cup (B \cap A^c)$ donc $\mathbb{P}(B \cap A^c) = \mathbb{P}(B) - \mathbb{P}(A \cap B)$.
5. On suppose le point 6) prouvé. On pose $B_n = \bigcup_{i=0}^n A_i$. Comme $B_n = B_{n-1} \cup A_n$ on a $\mathbb{P}(B_n) \leq \mathbb{P}(B_{n-1}) + \mathbb{P}(A_n)$ et par récurrence immédiate,

$$\mathbb{P}(B_n) \leq \sum_{i=0}^n \mathbb{P}(A_i) \leq \sum_{i=0}^{+\infty} \mathbb{P}(A_i).$$

Comme $(B_n)_{n \in \mathbb{N}}$ est croissante, le point 6) nous donne également

$$\lim_{n \in \mathbb{N}} \mathbb{P}(B_n) = \mathbb{P}\left(\bigcup_{n \in \mathbb{N}} B_n\right) = \mathbb{P}\left(\bigcup_{n \in \mathbb{N}} A_n\right).$$

6. On définit une suite d'évènements $(C_n)_{n \in \mathbb{N}}$ par

$$C_0 = A_0, \quad C_n = A_n \setminus A_{n-1} := A_n \cap (A_{n-1})^c, \quad \forall n \geq 1.$$

$(C_n)_{n \in \mathbb{N}}$ est ainsi une suite d'évènements, deux à deux disjoints, vérifiant

$$A_n = \bigcup_{i=0}^n C_i \quad \text{et} \quad \mathbb{P}(A_n) = \sum_{i=0}^n \mathbb{P}(C_i).$$

On a alors

$$\begin{aligned} \mathbb{P}\left(\bigcup_{n \in \mathbb{N}} A_n\right) &= \mathbb{P}\left(\bigcup_{n \in \mathbb{N}} C_n\right) = \sum_{n \in \mathbb{N}} \mathbb{P}(C_n) = \lim_{n \rightarrow +\infty} \sum_{i=0}^n \mathbb{P}(C_i) \\ &= \lim_{n \rightarrow +\infty} \mathbb{P}(A_n). \end{aligned}$$

7. il suffit de passer au complémentaire pour utiliser le point précédent : On pose $B_n = \overline{A_n}$ pour tout $n \in \mathbb{N}$. $(B_n)_{n \in \mathbb{N}}$ est une suite croissante donc on a $\mathbb{P}(\bigcup_{n \in \mathbb{N}} B_n) = \lim_{n \in \mathbb{N}} \mathbb{P}(B_n)$. Or

$$\mathbb{P}\left(\bigcup_{n \in \mathbb{N}} B_n\right) = \mathbb{P}\left(\bigcup_{n \in \mathbb{N}} \overline{A_n}\right) = \mathbb{P}\left(\overline{\bigcap_{n \in \mathbb{N}} A_n}\right) = 1 - \mathbb{P}\left(\bigcap_{n \in \mathbb{N}} A_n\right)$$

et

$$\lim_{n \in \mathbb{N}} \mathbb{P}(B_n) = \lim_{n \in \mathbb{N}} \mathbb{P}(\overline{A_n}) = 1 - \lim_{n \in \mathbb{N}} \mathbb{P}(A_n).$$

□

1.4 Probabilité uniforme

On suppose dans cette partie que Ω est fini ou dénombrable. On choisit alors pour tribu des événements \mathcal{A} l'ensemble $\mathcal{P}(\Omega)$.

Proposition 5

- La formule

$$\mathbb{P}(A) := \sum_{\omega \in A} p_\omega, \quad \forall A \in \mathcal{P}(\Omega),$$

définit une probabilité \mathbb{P} sur $(\Omega, \mathcal{P}(\Omega))$ dès que la famille de nombres réels $(p_\omega)_{\omega \in \Omega}$ vérifie

1. $p_\omega \in [0, 1]$ pour tous les $\omega \in \Omega$,
 2. $\sum_{\omega \in \Omega} p_\omega = 1$.
- Inversement, toute probabilité \mathbb{P} sur $(\Omega, \mathcal{P}(\Omega))$ est de cette forme : il suffit de poser

$$p_\omega := \mathbb{P}(\{\omega\}), \quad \forall \omega \in \Omega.$$

Définition 9 Soit \mathbb{P} une probabilité sur $(\Omega, \mathcal{P}(\Omega))$ avec Ω fini. On dit que \mathbb{P} est la loi uniforme sur Ω si toutes les épreuves $\omega \in \Omega$ sont équiprobables : i.e.

$$p_\omega = \frac{1}{\text{Card}(\Omega)}.$$

En particulier on a

$$\mathbb{P}(A) = \frac{\text{Card}(A)}{\text{Card}(\Omega)}, \quad \forall A \subset \Omega.$$

Remarque 6 *Le calcul des probabilités se ramène ici à un simple calcul de dénombrement. On a*

$$\mathbb{P}(A) = \frac{\text{nombre de cas favorables}}{\text{nombre de cas possibles}}.$$

Remarque 7 *Si Ω est dénombrable infini, il n'existe pas de probabilité uniforme.*

2 Probabilités conditionnelles, indépendance

Dans toute la suite on se place dans un espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$.

2.1 Définition

On suppose que l'on a l'information supplémentaire suivante : l'évènement B est réalisé. On souhaite alors modifier la mesure de probabilité \mathbb{P} pour tenir compte de cette nouvelle information.

Définition 10 Soient $A, B \in \mathcal{A}$ tels que $\mathbb{P}(B) > 0$. On appelle probabilité conditionnelle de A sachant B , le nombre réel

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

On peut également utiliser la notation $\mathbb{P}_B(A)$.

Proposition 6 Soit $B \in \mathcal{A}$ tel que $\mathbb{P}(B) > 0$. L'application

$$\mathbb{P}(\cdot|B) : \begin{cases} \mathcal{A} & \rightarrow [0, 1] \\ A & \mapsto \mathbb{P}(A|B). \end{cases}$$

est une mesure de probabilité sur (Ω, \mathcal{A}) .

Preuve.

- Pour tout $A \in \mathcal{A}$, $A \cap B \subset B$ donc $\mathbb{P}(A|B) \in [0, 1]$.
- $\mathbb{P}(\Omega|B) = 1$.

- Si $(A_n)_{n \in \mathbb{N}}$ est une famille d'événements aléatoires deux à deux disjoints, on a

$$\begin{aligned} \mathbb{P}\left(\bigcup_{n \in \mathbb{N}} A_n \mid B\right) &= \frac{\mathbb{P}\left(\left(\bigcup_{n \in \mathbb{N}} A_n\right) \cap B\right)}{\mathbb{P}(B)} = \frac{\mathbb{P}\left(\bigcup_{n \in \mathbb{N}} (A_n \cap B)\right)}{\mathbb{P}(B)} \\ &= \sum_{n \in \mathbb{N}} \frac{\mathbb{P}(A_n \cap B)}{\mathbb{P}(B)} = \sum_{n \in \mathbb{N}} \mathbb{P}(A_n \mid B). \end{aligned}$$

□

Remarque 8

- On a $\mathbb{P}(B|B) = 1$.
- $(A|B)$ ne veut rien dire. En particulier ce n'est pas un évènement !

Conséquences :

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

2.2 Formules usuelles

2.2.1 Formule des probabilités composées

Théorème 1 1. Soient $A, B \in \mathcal{A}$ avec $\mathbb{P}(A) > 0$. On a la formule des probabilités composées suivante :

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

2. Plus généralement, soient $n \geq 2$ et $A_1, \dots, A_n \in \mathcal{A}$ tels que

$$\mathbb{P}(A_1 \cap \dots \cap A_{n-1}) > 0.$$

Alors on a

$$\mathbb{P}(A_1 \cap \dots \cap A_n) = \mathbb{P}(A_1)\mathbb{P}(A_2|A_1)\dots\mathbb{P}(A_n|A_1 \cap \dots \cap A_{n-1}).$$

Remarque 9 Comme pour tout $1 \leq k \leq n-1$, $(A_1 \cap \dots \cap A_{n-1}) \subset (A_1 \cap \dots \cap A_k)$, alors

$$\mathbb{P}(A_1 \cap \dots \cap A_k) > 0$$

et donc les probabilités conditionnelles précédentes sont bien définies.

Preuve. Montrons le second point par récurrence sur n , le premier point correspondant à $n = 2$.

Pour $n = 2$ le résultat découle directement de la formule des probabilités conditionnelles.

Supposons vrai le résultat pour $n \geq 2$. Alors on a

$$\mathbb{P}(A_1 \cap \dots \cap A_{n+1}) = \mathbb{P}(A_1 \cap \dots \cap A_n) \mathbb{P}(A_{n+1} | A_1 \cap \dots \cap A_n)$$

puis on conclut grâce à l'hypothèse de récurrence. \square

Le premier point du théorème peut sembler à première vue inutile car il s'agit d'une simple réécriture de la définition d'une probabilité conditionnelle. En fait il n'en est rien, les deux formules ont leur intérêt en pratique. Dans certaines situations, on connaît la probabilité de l'intersection de deux événements et on calcule la probabilité conditionnelle, tandis que dans d'autres situations on connaît la probabilité conditionnelle et on en déduit la probabilité de l'intersection de deux événements.

2.2.2 Formule des probabilités totales

Définition 11 Une partition de Ω (ou système complet d'évènements) est une famille $(A_i)_{1 \leq i < N}$ (avec $N \in \mathbb{N}^*$ ou $N = +\infty$) d'évènements tels que

- $A_i \cap A_j = \emptyset$ si $i \neq j$,
- $\bigcup_{i < N} A_i = \Omega$.

Théorème 2 1. Soient $A, B \in \mathcal{A}$ tels que $0 < \mathbb{P}(A) < 1$, on a

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

2. Plus généralement, si $(A_n)_{n < N}$ est une partition de Ω avec $\mathbb{P}(A_n) > 0$ pour tout $n < N$, on a

$$\mathbb{P}(B) = \sum_{n < N} \mathbb{P}(A_n) \mathbb{P}(B|A_n).$$

Preuve. On a $\bigcup_{n < N} A_n = \Omega$ et la réunion est disjointe, donc

$$\mathbb{P}(B) = \mathbb{P}(B \cap (\bigcup_{n < N} A_n)) = \mathbb{P}(\bigcup_{n < N} (B \cap A_n)) = \sum_{n < N} \mathbb{P}(B \cap A_n).$$

Il suffit alors d'utiliser la formule des probabilités composées pour conclure.

Le premier point du théorème se traite de la même façon car $\{A, A^c\}$ est une partition de Ω . \square

2.2.3 Formule de Bayes

Théorème 3 1. Soient $A, B \in \mathcal{A}$ tels que $0 < \mathbb{P}(A) < 1$ et $\mathbb{P}(B) > 0$, on a

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

2. Plus généralement, si $(A_n)_{n < N}$ est une partition de Ω avec $\mathbb{P}(A_n) > 0$ pour tout $n < N$, et $\mathbb{P}(B) > 0$, on a

$$\mathbb{P}(A_k|B) = \frac{\mathbb{P}(A_k)\mathbb{P}(B|A_k)}{\sum_{n < N} \mathbb{P}(A_n)\mathbb{P}(B|A_n)}, \quad \forall k < N.$$

Preuve. Pour tout $k < N$ on a

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

et on applique la formule des probabilités totales pour conclure.

Encore une fois, le premier point du théorème se traite de la même façon. \square

2.3 Évènements indépendants

Heuristique : A est indépendant de B si $\mathbb{P}(A) = \mathbb{P}(A|B)$.

Définition 12 Soient $A, B \in \mathcal{A}$. On dit que A et B sont indépendants si

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

Remarque 10 La notion d'indépendance est différente de la notion d'incompatibilité. En particulier, si A et B sont incompatibles (i.e. disjoints), alors $\mathbb{P}(A \cap B) = 0$.

Remarque 11 Si $A, B \in \mathcal{A}$ sont indépendants avec $\mathbb{P}(B) > 0$, alors on a

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

Proposition 7 Soient $A, B \in \mathcal{A}$. Si A et B sont indépendants, alors A et B^c , A^c et B , A^c et B^c sont également indépendants.

Preuve. Par symétrie, il suffit de montrer que A et B^c sont indépendants. On a $\mathbb{P}(A) = \mathbb{P}(A \cap B) + \mathbb{P}(A \cap B^c)$. Donc

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

□

Proposition 8 Soient $(A_n)_{n < N}$ une famille d'événements de \mathcal{A} deux à deux disjoints et soit $B \in \mathcal{A}$. Si, pour tout $n < N$, A_n et B sont indépendants, alors $\bigcup_{n < N} A_n$ et B sont également indépendants.

Preuve.

$$\begin{aligned} \mathbb{P}\left(\left(\bigcup_{n < N} A_n\right) \cap B\right) &= \mathbb{P}\left(\bigcup_{n < N} (A_n \cap B)\right) = \sum_{n < N} \mathbb{P}(A_n \cap B) \\ &= \sum_{n < N} \mathbb{P}(A_n)\mathbb{P}(B) = \mathbb{P}\left(\bigcup_{n < N} A_n\right)\mathbb{P}(B). \end{aligned}$$

□

On va maintenant prolonger la notion d'indépendance de deux événements au cas des familles finies ou dénombrables d'événements.

Définition 13 Soient $(A_n)_{n < N}$ des événements aléatoires (donc des éléments de \mathcal{A}).

1. $(A_n)_{n < N}$ est une famille d'événements 2 à 2 indépendants si pour tout $n, m < N$ avec $n \neq m$, on a

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

2. $(A_n)_{n < N}$ est une famille d'événements indépendants dans leur ensemble (ou mutuellement indépendants ou bien encore indépendants) si, pour tout $2 \leq k < N$ et pour toute sous-suite finie $(A_{i_1}, A_{i_2}, \dots, A_{i_k})$ d'événements distincts, on a

$$\mathbb{P}(A_{i_1} \cap \dots \cap A_{i_k}) = \prod_{j=1}^k \mathbb{P}(A_{i_j}).$$

Remarque 12

- L'indépendance mutuelle implique l'indépendance 2 à 2. La réciproque est fausse, c.f. exemple suivant.
- Pour une famille finie d'événements aléatoires, l'indépendance 2 à 2 impose $C_n^2 = \frac{n(n-1)}{2}$ conditions tandis que l'indépendance mutuelle impose

$$C_n^2 + C_n^3 + \dots + C_n^n = 2^n - C_n^0 - C_n^1 = 2^n - n - 1$$

conditions.

Exemple 3 On lance deux dés à six faces. On considère les événements suivants : $A = \ll \text{le dé 1 est pair} \gg$, $B = \ll \text{le dé 2 est impair} \gg$, $C = \ll \text{les deux dés ont même parité} \gg$.

On a $\Omega = \{\omega = (\omega_1, \omega_2) | \omega_1, \omega_2 \in \{1, \dots, 6\}\} = \{1, \dots, 6\}^2$. On considère la probabilité uniforme \mathbb{P} . On a

$$A = \{2, 4, 6\} \times \{1, \dots, 6\}, \quad \mathbb{P}(A) = 1/2,$$

$$B = \{1, \dots, 6\} \times \{1, 3, 5\}, \quad \mathbb{P}(B) = 1/2,$$

$$C = \{2, 4, 6\} \times \{2, 4, 6\} \cup \{1, 3, 5\} \times \{1, 3, 5\}, \quad \mathbb{P}(C) = 1/2,$$

$$A \cap B = \{2, 4, 6\} \times \{1, 3, 5\}, \quad \mathbb{P}(A \cap B) = 1/4,$$

$$A \cap C = \{2, 4, 6\} \times \{2, 4, 6\}, \quad \mathbb{P}(A \cap C) = 1/4,$$

$$B \cap C = \{1, 3, 5\} \times \{1, 3, 5\}, \quad \mathbb{P}(B \cap C) = 1/4,$$

$$A \cap B \cap C = \emptyset, \quad \mathbb{P}(A \cap B \cap C) = 0.$$

Donc A , B et C sont indépendants deux à deux mais non indépendants dans leur ensemble.

3 Variables aléatoires, variables aléatoires discrètes

3.1 Rappels et Définition

Commençons par quelques rappels de théorie de la mesure.

Définition 14 Soit (E, \mathcal{A}) et (F, \mathcal{B}) deux espaces mesurables. On dit qu'une fonction $f : (E, \mathcal{A}) \rightarrow (F, \mathcal{B})$ est mesurable si pour tout $B \in \mathcal{B}$, $f^{-1}(B) \in \mathcal{A}$, en rappelant que

$$f^{-1}(B) = \{a \in E, f(a) \in B\}.$$

Remarque 13 Si $\mathcal{A} = \mathcal{P}(E)$, toute application $f : (E, \mathcal{A}) \rightarrow (F, \mathcal{B})$ est mesurable.

Soient $B_1, B_2 \subset F$, $A_1, A_2 \subset E$.

•

$$\begin{aligned} f^{-1}(B_1 \cap B_2) &= \{a \in E, f(a) \in B_1 \cap B_2\} \\ &= \{a \in E, f(a) \in B_1\} \cap \{a \in E, f(a) \in B_2\} \\ &= f^{-1}(B_1) \cap f^{-1}(B_2). \end{aligned}$$

•

$$\begin{aligned} f^{-1}(B_1 \cup B_2) &= \{a \in E, f(a) \in B_1 \cup B_2\} \\ &= \{a \in E, f(a) \in B_1\} \cup \{a \in E, f(a) \in B_2\} \\ &= f^{-1}(B_1) \cup f^{-1}(B_2). \end{aligned}$$

•

$$f(A_1) = \{b \in F, \exists a \in A_1, f(a) = b\}.$$

•

$$f(A_1 \cap A_2) \subsetneq f(A_1) \cap f(A_2).$$

Exemple de non égalité : $A_1 = \{1\}$, $A_2 = \{2\}$, $f(1) = f(2)$.

•

$$f(A_1 \cup A_2) = f(A_1) \cup f(A_2).$$

On rappelle que l'on se place dans un espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$.

Définition 15 Soit (F, \mathcal{B}) un espace mesurable. On appelle variable aléatoire (v.a. en abrégé) définie sur $(\Omega, \mathcal{A}, \mathbb{P})$ et à valeurs dans F , toute application mesurable $X : \Omega \rightarrow F$, c'est à dire satisfaisant pour tout $B \in \mathcal{B}$, $X^{-1}(B) \in \mathcal{A}$ avec

$$X^{-1}(B) = \{\omega \in \Omega \text{ tels que } X(\omega) \in B\}.$$

On appelle variable aléatoire réelle une variable aléatoire à valeurs dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. On appelle vecteur aléatoire réel une variable aléatoire à valeurs dans $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$.

Remarque 14 Soit X une variable aléatoire réelle. On utilisera les notations suivantes :

- $\{X = a\} := X^{-1}(\{a\}) = \{\omega \in \Omega \text{ tels que } X(\omega) = a\}$,
- $\{X \in [a, b]\} := \{a \leq X \leq b\} = X^{-1}([a, b]) := \{\omega \in \Omega \text{ tels que } X(\omega) \in [a, b]\}$,
- $\{X \in]-\infty, b]\} := \{X \leq b\} = X^{-1}(]-\infty, b]) = \{\omega \in \Omega \text{ tels que } X(\omega) \in]-\infty, b]\}$,
- ...

Plus généralement, si X est une variable aléatoire, on utilise la notation $\{X \in B\} := X^{-1}(B) = \{\omega \in \Omega \text{ tels que } X(\omega) \in B\}$ pour tout $B \in \mathcal{B}$.

Proposition 9

- Soit X est un vecteur aléatoire réel et $h : \mathbb{R}^d \rightarrow \mathbb{R}^d$ une fonction mesurable. Alors $h(X)$ est un vecteur aléatoire réel.
- $X = (X_1, \dots, X_n)$ est un vecteur aléatoire réel si et seulement si chacune de ses composantes X_i est une variable aléatoire réelle.

Remarque 15 On a vu que $\mathcal{B}(\mathbb{R})$ coïncide avec la tribu engendrée par les intervalles $[a, b]$ avec $a, b \in \mathbb{R}$ et $a < b$. Ainsi X est une variable aléatoire réelle si pour tout intervalle $[a, b]$, $\{X \in [a, b]\} \in \mathcal{A}$, i.e. $\{X \in [a, b]\}$ est un événement aléatoire.

3.2 Loi de probabilité

3.2.1 Cadre général

Proposition et définition 1 Soit X une variable aléatoire. On appelle loi de probabilité de X la mesure de probabilité \mathbb{P}_X définie sur (F, \mathcal{B}) par

$$\mathbb{P}_X(B) := \mathbb{P}(X^{-1}(B)) = \mathbb{P}(X \in B), \quad \forall B \in \mathcal{B}.$$

C'est la mesure image de \mathbb{P} par X .

Preuve.

- Tout d'abord, \mathbb{P}_X est bien définie : en effet, si $B \in \mathcal{B}$, alors $(X \in B) \in \mathcal{A}$, c'est à dire que c'est un évènement aléatoire, car X est une fonction mesurable. En particulier, $\mathbb{P}(X \in B)$ a un bien un sens.
- $\mathbb{P}_X(B) = \mathbb{P}(X \in B) \in [0, 1]$.
- $\mathbb{P}_X(F) = \mathbb{P}(X \in F) = \mathbb{P}(\Omega) = 1$.
- Soit $(B_n)_{n \geq 1}$ une famille disjointes d'éléments de \mathcal{B} . Alors on a

$$\mathbb{P}_X\left(\bigcup_{n \geq 1} B_n\right) = \mathbb{P}(X^{-1}\left(\bigcup_{n \geq 1} B_n\right)) = \mathbb{P}\left(\bigcup_{n \geq 1} X^{-1}(B_n)\right).$$

La dernière union est disjointe car si $\omega \in X^{-1}(B_i) \cap X^{-1}(B_j)$, alors $X(\omega) \in B_i \cap B_j = \emptyset$. Donc

$$\mathbb{P}_X\left(\bigcup_{n \geq 1} X^{-1}(B_n)\right) = \sup_{n \geq 1} \mathbb{P}_X(X^{-1}(B_n)) = \sum_{n \geq 1} \mathbb{P}_X(B_n).$$

□

Remarque 16 *Puisque $\mathcal{B}(\mathbb{R})$ est engendré par les intervalles (ouverts, fermés,...) et les ensembles du type $] -\infty, t]$, la loi d'une variable aléatoire réelle est caractérisée par la donnée des valeurs*

- $\mathbb{P}(X \in [a, b])$, pour tous $a < b$,
- ou $\mathbb{P}(X \in]a, b])$, pour tous $a < b$,
- ou $\mathbb{P}(X \leq t)$, pour tous $t \in \mathbb{R}$,
- ...

Lorsque l'on considère la loi d'une variable aléatoire on s'intéresse aux valeurs que prend la variable aléatoire et aux probabilités associées. Par contre on oublie l'univers Ω : Pour toute variable aléatoire réelle on peut calculer sa loi, par contre on ne peut pas reconstruire une variable aléatoire à partir de sa loi. En particulier, deux variables aléatoires réelles qui ne sont même pas nécessairement définies sur le même univers, peuvent avoir une même loi! (c.f. exemples plus loin).

3.2.2 Le cas particulier des variables aléatoires discrètes

Définition 16 *On dit que X est une variable aléatoire discrète si son voisinage $X(\Omega)$ est fini ou infini dénombrable.*

Plus généralement, s'il existe un ensemble discret (fini ou infini dénombrable) $B \subset X(\Omega)$ tel que $\mathbb{P}(X \in B) = 1$, on dit également que X est une variable aléatoire

discrète. On appelle alors support de X le plus petit ensemble discret \tilde{B} vérifiant $\mathbb{P}(X \in \tilde{B}) = 1$. En pratique on notera encore $X(\Omega)$ le support de X ce qui est clairement un abus de notation.

On peut alors prendre $F = X(\Omega)$. On a

$$F = \{x_i, 0 \leq i < N\} \quad \text{avec} \quad N \in \mathbb{N}^* \text{ ou } N = +\infty.$$

Comme F est discret, on peut prendre $\mathcal{B} = \mathcal{P}(F)$. La variable aléatoire X définit une probabilité \mathbb{P}_X sur l'ensemble mesurable discret (F, \mathcal{B}) . On a vu au chapitre 1 comment caractériser une probabilité sur un ensemble discret :

Pour connaître \mathbb{P}_X il suffit de connaître les

$$\mathbb{P}_X(\{x_i\}) = \mathbb{P}(X = x_i), \quad 0 \leq i < N.$$

Proposition 10 *Soit X une variable aléatoire discrète. La loi de X est caractérisée par*

- $X(\Omega) = \{x_i, 0 \leq i < N\}$ avec $N \in \mathbb{N}^*$ ou $N = +\infty$.
- $\mathbb{P}_X(\{x_i\}) = \mathbb{P}(X = x_i)$, $0 \leq i < N$.

En particulier, pour tout $A \subset \mathbb{R}$, on a

$$\mathbb{P}(X \in A) = \sum_{x_i \in A} \mathbb{P}(X = x_i).$$

Exemple 1. $\Omega = \{1, \dots, 6\}$ et \mathbb{P} la probabilité uniforme. On pose

$$X(\omega) = (\omega - 3)^2.$$

X est une variable aléatoire nécessairement discrète car Ω est discret.

$$X(\Omega) = \{0, 1, 4, 9\}.$$

De plus

$$\begin{aligned} \mathbb{P}_X(\{0\}) &= \mathbb{P}(X = 0) = \mathbb{P}(\{3\}) = 1/6 \\ \mathbb{P}_X(\{1\}) &= \mathbb{P}(X = 1) = \mathbb{P}(\{2, 4\}) = 1/3 \\ \mathbb{P}_X(\{4\}) &= \mathbb{P}(X = 4) = \mathbb{P}(\{1, 5\}) = 1/3 \\ \mathbb{P}_X(\{9\}) &= \mathbb{P}(X = 9) = \mathbb{P}(\{6\}) = 1/6. \end{aligned}$$

Exemple 2. $\Omega = [0, 1]$ et \mathbb{P} la mesure de Lebesgue sur Ω . On pose

$$X = \lfloor 2\omega \rfloor.$$

$X(\Omega) = \{0, 1, 2\}$ donc X est une variable discrète.

$$\mathbb{P}(X = 0) = \mathbb{P}([0, 1/2[) = 1/2$$

$$\mathbb{P}(X = 1) = \mathbb{P}([1/2, 1]) = 1/2$$

$$\mathbb{P}(X = 2) = \mathbb{P}(\{1\}) = 0.$$

Exemple 3. On lance deux pièces équilibrées.

$$\Omega = \{\omega = (\omega_1, \omega_2), \omega_1, \omega_2 \in \{0, 1\}\}.$$

On considère la probabilité uniforme sur Ω . On note

- X : résultat du premier lancer de pièce. $X((\omega_1, \omega_2)) = \omega_1$,
- Y : résultat du premier lancer de pièce. $Y((\omega_1, \omega_2)) = \omega_2$.

$X(\Omega) = \{0, 1\}$ et $Y(\Omega) = \{0, 1\}$. De plus

$$\mathbb{P}(X = 0) = \mathbb{P}(X = 1) = \mathbb{P}(Y = 0) = \mathbb{P}(Y = 1) = 1/2.$$

Ainsi X et Y ont même loi mais $X \neq Y$. En effet

$$(X = Y) = \text{« les deux pièces ont même résultat »},$$

et $\mathbb{P}(X = Y) = 1/2$.

Remarque 17 *On peut prendre Ω bien plus grand sans pour autant changer la loi de X et Y . Par exemple*

$$\Omega = \{\text{Résultats de tous les lancers de pièces depuis la nuit des temps}\}.$$

3.3 Espérance et variance d'une variable aléatoire

3.3.1 Espérance d'une variable aléatoire

Proposition et définition 2 1. *Soit X une variable aléatoire réelle positive, alors l'espérance de X est définie par la quantité, éventuellement infinie,*

$$\mathbb{E}[X] := \int_{\Omega} X(\omega) d\mathbb{P}(\omega) \in \mathbb{R}^+ \cup \{+\infty\}.$$

2. *Soit X une variable aléatoire réelle non nécessairement positive, alors l'espérance de X est définie uniquement si*

$$\mathbb{E}[|X|] = \int_{\omega \in \Omega} |X(\omega)| d\mathbb{P}(\omega) < +\infty.$$

Dans ce cas on dit qu'elle est intégrable et elle vaut

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

Remarque 18 L'espérance de X n'existe pas toujours lorsque la v.a. n'a pas de signe.

L'espérance de X correspond à la moyenne des valeurs que peut prendre X pondérées par la loi de probabilité \mathbb{P}_X .

Proposition 11 (formule de changement de variable) Soit X une v.a. réelle telle que $\mathbb{E}[X]$ existe. Alors

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

Proposition 12 On a les propriétés suivantes qui découlent des propriétés de l'intégrale de Lebesgues.

- *Positivité.* Si $X \geq 0$, alors $\mathbb{E}[X] \geq 0$. Donc si $X \geq Y$, alors $\mathbb{E}[X] \geq \mathbb{E}[Y]$.
- *Linéarité.* Soient $a, b \in \mathbb{R}$, X et Y deux variables aléatoires définies sur le même espace de probabilité, alors

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

- $\mathbb{E}[1] = \mathbb{P}(\Omega) = 1$.

De plus l'espérance d'une variable aléatoire ne dépend que de sa loi : deux v.a. de même loi, ont même espérance (si elle existe).

Proposition 13 (Cas des v.a. discrètes.) Soit X une v.a. discrète.

- Si X est positive, alors

$$\mathbb{E}[X] = \sum_{x_i \in X(\Omega)} x_i \mathbb{P}(X = x_i).$$

- Si X est de signe quelconque et

$$\sum_{x_i \in X(\Omega)} |x_i| \mathbb{P}(X = x_i) < +\infty,$$

alors $\mathbb{E}[X]$ existe et

$$\mathbb{E}[X] = \sum_{x_i \in X(\Omega)} x_i \mathbb{P}(X = x_i).$$

Exemple du lancer de dé. $\Omega = \{1, \dots, 6\}$, \mathbb{P} la probabilité uniforme. On pose $X(\omega) = (\omega - 3)^2$. On a

$$\mathbb{E}[X] = 0 \times \mathbb{P}(X = 0) + 1 \times \mathbb{P}(X = 1) + 4 \times \mathbb{P}(X = 4) + 9 \times \mathbb{P}(X = 9) = \frac{19}{16}.$$

3.3.2 Formule de transfert

Proposition 14 (Formule de transfert) *Soit $g : \mathbb{R} \rightarrow \mathbb{R}$ une fonction mesurable et X une variable aléatoire réelle. On pose $Y = g(X)$. Alors Y est une variable aléatoire réelle. Si $\int_{\mathbb{R}} |g(x)| d\mathbb{P}_X(x) < +\infty$ alors Y est intégrable et*

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

Remarque 19 *Si Y est positive, alors la formule précédente est vraie même si l'intégrale vaut $+\infty$.*

La formule de transfert est très utile en pratique car elle permet de calculer l'espérance de Y sans avoir à déterminer sa loi!

Preuve. On commence par supposer que Y est à valeurs positive. Comme on a $Y = g(X)$, la loi de Y (i.e. \mathbb{P}_Y) est juste la mesure image de la loi de X (i.e. \mathbb{P}_X) par la fonction g . Ainsi, on peut appliquer la formule de changement de variable pour obtenir

$$\int_{\mathbb{R}} g(x) d\mathbb{P}_X(x) = \int_{\mathbb{R}} y d\mathbb{P}_Y(y) = \mathbb{E}[Y].$$

Dans le cas général, on considère d'abord $|Y| = |g(X)| = \tilde{g}(X)$ avec $\tilde{g} = |g|$, puis on applique le résultat précédent :

$$\int_{\mathbb{R}} |g(x)| d\mathbb{P}_X(x) = \int_{\mathbb{R}} y d\mathbb{P}_{|Y|}(y) = \mathbb{E}[|Y|].$$

Donc, si $\int_{\mathbb{R}} |g(x)| d\mathbb{P}_X(x) < +\infty$, alors Y est intégrable et ainsi la formule est encore vérifiée. \square

Regardons maintenant le cas particulier des v.a. discrètes.

Proposition 15 (Formule de transfert, cas discret) *Soit $g : \mathbb{R} \rightarrow \mathbb{R}$ une fonction mesurable et X une variable aléatoire réelle discrète. On pose $Y = g(X)$. Alors Y est une variable aléatoire réelle discrète.*

- Si Y est positive, alors

$$\mathbb{E}[Y] = \sum_{x_i \in X(\Omega)} g(x_i) \mathbb{P}(X = x_i).$$

- Si Y est de signe quelconque et que

$$\sum_{x_i \in X(\Omega)} |g(x_i)| \mathbb{P}(X = x_i) < +\infty,$$

alors on a

$$\mathbb{E}[Y] = \sum_{x_i \in X(\Omega)} g(x_i) \mathbb{P}(X = x_i).$$

Preuve. Nous avons déjà démontré la proposition dans le cas général, mais d'un point de vue pédagogique il peut être intéressant de la redémontrer dans le cas discret.

Soient X et Y deux v.a. réelles discrètes positives telles que $X(\Omega) = \{x_i, 0 \leq i < N\}$ et $Y(\Omega) = \{y_j, 0 \leq j < M\}$. Alors

$$\mathbb{E}[Y] = \sum_{y_j \in Y(\Omega)} y_j \mathbb{P}(Y = y_j).$$

Or $\mathbb{P}(Y = y_j) = \mathbb{P}(X \in g^{-1}(\{y_j\}))$ avec

$$g^{-1}(\{y_j\}) = \{x_k \in X(\Omega), g(x_k) = y_j\}.$$

Donc

$$\begin{aligned} \mathbb{E}[Y] &= \sum_{y_j \in Y(\Omega)} y_j \mathbb{P}(X \in g^{-1}(\{y_j\})) \\ &= \sum_{y_j \in Y(\Omega)} y_j \sum_{x_k \in X(\Omega), g(x_k) = y_j} \mathbb{P}(X = x_k) \\ &= \sum_{y_j \in Y(\Omega)} \sum_{x_k \in X(\Omega), g(x_k) = y_j} g(x_k) \mathbb{P}(X = x_k) \\ &= \sum_{x_k \in X(\Omega)} g(x_k) \mathbb{P}(X = x_k). \end{aligned}$$

□

Exemple du dé. Reprenons l'exemple $X(\omega) = (\omega - 3)^2$. On pose D la variable aléatoire donnant la valeur du dé. D est à valeur dans $\{1, \dots, 6\}$ et de loi uniforme : $\mathbb{P}(D = k) = 1/6$ pour tout $k \in \{1, \dots, 6\}$.

$$\mathbb{E}[X] = \mathbb{E}[(D-3)^2] = (1-3)^2 \mathbb{P}(D = 1) + (2-3)^2 \mathbb{P}(D = 2) + \dots + (6-3)^2 \mathbb{P}(D = 6) = \frac{19}{6}.$$

Application importante.

Définition 17 Si $\mathbb{E}[X^k]$ existe, $\mathbb{E}[X^k]$ s'appelle le moment d'ordre k de X .

Proposition 16 Si X est une variable aléatoire discrète à valeurs dans $\{x_i, 0 \leq i < N\}$ avec $N \in \mathbb{N}$ ou $N = +\infty$ et telle que

$$\mathbb{E}[|X|^k] = \sum_{0 \leq i < N} |x_i|^k \mathbb{P}(X = x_i) < +\infty$$

alors

$$\mathbb{E}[X^k] = \sum_{0 \leq i < N} x_i^k \mathbb{P}(X = x_i).$$

3.3.3 Variance d'une variable aléatoire réelle

Définition 18 Soit X une variable aléatoire de carré intégrable, i.e. $\mathbb{E}[X^2] < +\infty$. On appelle variance de X , notée $\text{Var}(X)$, le nombre positif

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

On appelle alors écart-type de X le nombre positif

$$\sigma := \sqrt{\text{Var}(X)}.$$

Remarque 20

- $\text{Var}(X) \geq 0$.
- La variance est une mesure de l'écart de la variable aléatoire à sa moyenne : plus la variance est petite, plus la v.a. à des valeurs concentrées près de sa moyenne.
- $\text{Var}(X) = 0$ si et seulement si $X = \mathbb{E}[X]$ \mathbb{P} -p.s., c'est-à-dire X constante \mathbb{P} -p.s. : il existe $c \in \mathbb{R}$ telle que $\mathbb{P}(X = c) = 1$.

Proposition 17 Soit X une variable aléatoire de carré intégrable et $a, b \in \mathbb{R}$. On a

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

Preuve.

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

□

Proposition 18 Soit X une variable aléatoire de carré intégrable. Alors on a

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

Preuve.

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

Proposition 19 Si X est une variable aléatoire discrète à valeurs dans $\{x_i, 0 \leq i < N\}$ avec $N \in \mathbb{N}$ ou $N = +\infty$ et telle que

$$\mathbb{E}[|X|^2] = \sum_{0 \leq i < N} |x_i|^2 \mathbb{P}(X = x_i) < +\infty$$

alors

$$\text{Var}(X) = \sum_{0 \leq i < N} x_i^2 \mathbb{P}(X = x_i) - \left(\sum_{0 \leq i < N} x_i \mathbb{P}(X = x_i) \right)^2.$$

□

3.4 Variables aléatoires indépendantes

Soient X et Y deux v.a. définies sur le même espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$.

$$X : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow (E_1, \mathcal{B}_1)$$

$$Y : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow (E_2, \mathcal{B}_2).$$

On peut considérer le vecteur aléatoire

$$(X, Y) : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow (E_1 \times E_2, \mathcal{B}_1 \times \mathcal{B}_2)$$

avec $\mathcal{B}_1 \times \mathcal{B}_2$ la tribu produit.

Définition 19 On dit que X et Y sont indépendantes si $\mathbb{P}_{(X,Y)}$ est la mesure produit $\mathbb{P}_X \times \mathbb{P}_Y$.

Rappel : Si $B_1 \in \mathcal{B}_1$ et $B_2 \in \mathcal{B}_2$ alors

$$\mathbb{P}_X \times \mathbb{P}_Y(B_1 \times B_2) = \mathbb{P}_X(B_1)\mathbb{P}_Y(B_2).$$

Regardons ce que cela donne concrètement pour les v.a. discrètes.

Proposition 20 *Si X et Y sont deux v.a. discrètes à valeurs dans $\{x_i, 0 \leq i < N\}$ et $\{y_j, 0 \leq j < M\}$. Alors X et Y sont indépendantes si et seulement si, pour tout $0 \leq i < N$ et $0 \leq j < M$ on a*

$$\mathbb{P}(X = x_i, Y = y_j) = \mathbb{P}(X = x_i)\mathbb{P}(Y = y_j),$$

c'est-à-dire que $(X = x_i)$ et $(Y = y_j)$ sont des évènements indépendants.

Plus généralement, on peut également une notion d'indépendance pour plus de deux v.a.

Définition 20 *Soit $(X_n)_{n \geq 1}$ une suite de v.a. définies sur le même espace de probabilité. On dit que ces v.a. sont indépendantes si, pour tout $n \geq 1$, on a*

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

Dans le cadre discret cela se traduit par : pour tout $n \geq 1$, pour tout $x_1 \in X_1(\Omega), \dots, x_n \in X_n(\Omega)$, on a

$$\mathbb{P}(X_1 = x_1, \dots, X_n = x_n) = \prod_{i=1}^n \mathbb{P}(X_i = x_i).$$

Proposition 21 *Soient X et Y deux v.a. indépendantes. Si XY est intégrable alors $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$. Plus généralement, soient $f : \mathbb{R} \rightarrow \mathbb{R}$ et $g : \mathbb{R} \rightarrow \mathbb{R}$ deux fonctions mesurables. Si $f(X)g(Y)$ est intégrable, alors $\mathbb{E}[f(X)g(Y)] = \mathbb{E}[f(X)]\mathbb{E}[g(Y)]$.*

Preuve. Cas général. On a

$$\mathbb{E}[f(X)g(Y)] = \int_{\mathbb{R}^2} f(x)g(y)d\mathbb{P}_{(X,Y)}(x, y)$$

d'après le théorème de la mesure image. En utilisant l'indépendance puis Fubini on obtient alors

$$\begin{aligned} \int_{\mathbb{R}^2} f(x)g(y)d\mathbb{P}_{(X,Y)}(x, y) &= \int_{\mathbb{R}^2} f(x)g(y)d\mathbb{P}_X(x)d\mathbb{P}_Y(y) \\ &= \int_{\mathbb{R}} f(x)d\mathbb{P}_X(x) \int_{\mathbb{R}} g(y)d\mathbb{P}_Y(y) \\ &= \mathbb{E}[f(X)]\mathbb{E}[g(Y)]. \end{aligned}$$

Cas discret. $Z = XY$ donc $Z(\Omega) = \{z \in \mathbb{R}, \exists x \in X(\Omega), \exists y \in Y(\Omega), z = xy\}$.
Soit $z \in Z(\Omega)$, alors

$$\begin{aligned}\mathbb{P}(Z = z) &= \mathbb{P}(XY = z) = \sum_{x \in X(\Omega), y \in Y(\Omega), z=xy} \mathbb{P}(X = x, Y = y) \\ &= \sum_{x \in X(\Omega), y \in Y(\Omega), z=xy} \mathbb{P}(X = x)\mathbb{P}(Y = y).\end{aligned}$$

Avec la loi de Z on peut alors calculer son espérance.

$$\begin{aligned}\mathbb{E}[XY] &= \mathbb{E}[Z] = \sum_{z \in Z(\omega)} z\mathbb{P}(Z = z) \\ &= \sum_{z \in Z(\omega)} \sum_{x \in X(\Omega), y \in Y(\Omega), z=xy} xy\mathbb{P}(X = x)\mathbb{P}(Y = y) \\ &= \sum_{x \in X(\Omega), y \in Y(\Omega)} xy\mathbb{P}(X = x)\mathbb{P}(Y = y) \\ &= \sum_{x \in X(\Omega)} x\mathbb{P}(X = x) \sum_{y \in Y(\Omega)} y\mathbb{P}(Y = y) = \mathbb{E}[X]\mathbb{E}[Y].\end{aligned}$$

□

Proposition 22 Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires. Pour tout $n \geq 2$, X_1, \dots, X_n sont indépendantes si et seulement si pour toutes fonctions mesurables $h_1, \dots, h_n : \mathbb{R} \rightarrow \mathbb{R}$ positives ou bien bornées on a

$$\mathbb{E}\left[\prod_{i=1}^n h_i(X_i)\right] = \prod_{i=1}^n \mathbb{E}[h_i(X_i)].$$

Remarque 21 La proposition précédente implique que si $(X_n)_{n \in \mathbb{N}}$ est une suite de variables aléatoires indépendantes et si $(h_n)_{n \in \mathbb{N}}$ est une suite de fonctions mesurables réelles, alors $(h_n(X_n))_{n \in \mathbb{N}}$ est une suite de variables aléatoires indépendantes

Proposition 23 Si X et Y sont deux v.a. indépendantes, de carré intégrable, alors

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

Preuve.

$$\begin{aligned}\text{Var}(X + Y) &= \mathbb{E}[(X + Y)^2] - \mathbb{E}[X + Y]^2 \\ &= \text{Var}(X) + \text{Var}(Y) + 2\mathbb{E}[XY] - 2\mathbb{E}[X]\mathbb{E}[Y] \\ &= \text{Var}(X) + \text{Var}(Y).\end{aligned}$$

□

3.5 Lois discrètes usuelles

3.5.1 Loi uniforme sur $\{1, \dots, n\}$

Définition 21 Soit $n \in \mathbb{N}^*$. On dit que X suit la loi uniforme sur $\{1, \dots, n\}$ si $X(\Omega) = \{1, \dots, n\}$ et

$$\mathbb{P}(X = k) = \frac{1}{n}, \quad \forall k \in \{1, \dots, n\}.$$

On note $X \sim \mathcal{U}(\{1, \dots, n\})$.

C'est la loi du tirage d'un dé équilibré à n faces. On a

$$\mathbb{E}[X] = \sum_{k=1}^n k\mathbb{P}(X = k) = \frac{1}{n} \sum_{k=1}^n k = \frac{1}{n} \frac{n(n+1)}{2}.$$

De plus,

$$\mathbb{E}[X^2] = \sum_{k=1}^n k^2\mathbb{P}(X = k) = \frac{1}{n} \frac{n(n+1)(2n+1)}{6},$$

donc $\text{Var}(X) = \frac{n^2-1}{12}$.

3.5.2 Loi de Bernoulli

Définition 22 Soit $p \in [0, 1]$. On dit que X suit la loi de Bernoulli de paramètre p si $X(\Omega) = \{0, 1\}$ et

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

On note $X \sim \mathcal{B}(p)$.

Une v.a. de Bernoulli permet de modéliser une expérience aléatoire avec deux issues : un succès ou un échec.

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

De plus

$$\mathbb{E}[X^2] = 1^2 \times \mathbb{P}(X = 1) + 0^2 \times \mathbb{P}(X = 0) = p.$$

Donc $\text{Var}(X) = p(1 - p)$. Remarquons au passage que X^2 suit la même loi que X !

3.5.3 Loi binomiale

Définition 23 Soient $p \in [0, 1]$ et $n \in \mathbb{N}^*$. On dit que X suit la loi binomiale de paramètres (n, p) si X suit la loi du nombre de succès lorsque l'on réalise n expériences aléatoires indépendantes ayant chacune une probabilité p de succès. On note $X \sim \mathcal{B}(n, p)$.

Remarque 22 $\mathcal{B}(1, p) = \mathcal{B}(p)$.

Proposition 24 Soient X_1, \dots, X_n n v.a. indépendantes de loi $\mathcal{B}(p)$. Alors

$$\sum_{i=1}^n X_i \sim \mathcal{B}(n, p).$$

Proposition 25 Soit $X \sim \mathcal{B}(n, p)$. Alors $X(\Omega) = \{0, \dots, n\}$ et

$$\mathbb{P}(X = k) = C_n^k p^k (1 - p)^{n-k}, \quad 0 \leq k \leq n.$$

Remarque 23 La formule du binôme de Newton nous donne bien que

$$\sum_{k=0}^n \mathbb{P}(X = k) = (p + (1 - p))^n = 1.$$

Preuve. Soit $X \sim \mathcal{B}(n, p)$. Comme X représente un nombre de succès et qu'il y a n expériences, on a bien $X(\Omega) = \{0, \dots, n\}$. On note X_i le résultat de l'expérience i ($X_i \sim \mathcal{B}(p)$).

Soit $k \in \{0, \dots, n\}$. L'évènement $(X = k)$ correspond à l'évènement " k succès et $n - k$ échecs". Il y a C_n^k positions possibles pour les k succès parmi les n expériences. De plus, une fois fixées les positions des k succès et des $n - k$ échecs, la probabilité d'obtenir ces k succès dans ces positions fixées vaut

$$\mathbb{P}(\cap_{i \in I} (X_i = 1) \cap_{j \in J} (X_j = 0)) = \prod_{i \in I} \mathbb{P}(X_i = 1) \prod_{j \in J} \mathbb{P}(X_j = 0) = p^k (1 - p)^{n-k}$$

par indépendance. □

Proposition 26 Soient $p \in [0, 1]$, $n \in \mathbb{N}^*$ et $X \sim \mathcal{B}(n, p)$. Alors $\mathbb{E}[X] = np$ et $\text{Var}(X) = np(1 - p)$.

Preuve. L'espérance et la variance d'une v.a. ne dépendent que de la loi de cette variable aléatoire, il suffit donc de prendre $X = \sum_{i=1}^n X_i$ avec X_1, \dots, X_n indépendantes et de loi $\mathcal{B}(p)$. On a alors

$$\mathbb{E} \left[\sum_{i=1}^n X_i \right] = \sum_{i=1}^n \mathbb{E}[X_i] = np$$

par linéarité et

$$\text{Var} \left(\sum_{i=1}^n X_i \right) = \sum_{i=1}^n \text{Var}(X_i) = np(1 - p)$$

par indépendance. □

3.5.4 Loi géométrique

Définition 24 Soit $p \in]0, 1]$. On dit que X suit la loi géométrique de paramètre p si X suit la loi du rang du premier succès lorsque l'on réalise une infinité d'expériences aléatoires indépendantes, chaque expérience ayant une probabilité p de succès. On note $X \sim \mathcal{G}(p)$.

Proposition 27 Soit $X \sim \mathcal{G}(p)$. Alors $X(\Omega) = \mathbb{N}^*$ et

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

Preuve. On considère $(X_i)_{i \in \mathbb{N}^*}$ des v.a. indépendantes de loi $\mathcal{B}(p)$: X_i représente le résultat de la i -ème expérience. On note N l'instant du premier succès. Alors $N(\Omega) = \mathbb{N}^* \cup \{+\infty\}$. Pour $k \in \mathbb{N}^*$,

$$\mathbb{P}(N = k) = \mathbb{P}(X_1 = 0, \dots, X_{k-1} = 0, X_k = 1) = \mathbb{P}(X_1 = 0) \dots \mathbb{P}(X_{k-1} = 0) \mathbb{P}(X_k = 1) = (1-p)^{k-1} p$$

par indépendance. De plus

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

car $|1 - p| < 1$, donc $\mathbb{P}(N = +\infty) = 1 - \sum_{k=1}^{+\infty} \mathbb{P}(N = k) = 0$. □

Remarque 24 Si $p = 1$, alors $\mathbb{P}(N = 1) = 1$. Si $p = 0$ alors $\mathbb{P}(N = +\infty) = 1$.

Proposition 28 On a $\mathbb{E}[X] = \frac{1}{p}$ et $\text{Var}(X) = \frac{1-p}{p^2}$.

3.5.5 Loi de Poisson

Définition 25 Soit $\lambda > 0$. On dit que X suit la loi de Poisson si $X(\Omega) = \mathbb{N}$ et si $\mathbb{P}(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}$, pour tout $k \in \mathbb{N}$. On note $X \sim \mathcal{P}(\lambda)$.

Proposition 29 Soit $X \sim \mathcal{P}(\lambda)$. Alors $\mathbb{E}[X] = \lambda$ et $\text{Var}(X) = \lambda$.

3.6 Fonction de répartition

Soit X une variable aléatoire réelle. On rappelle que \mathbb{P}_X est définie par

$$\mathbb{P}_X(B) = \mathbb{P}(X \in B), \quad \forall B \in \mathcal{B}(\mathbb{R}).$$

Or la tribu $\mathcal{B}(\mathbb{R})$ est engendrée par les intervalles $] - \infty, t]$ pour tout $t \in \mathbb{R}$. Donc pour connaître \mathbb{P}_X il suffit de connaître $\mathbb{P}(X \in] - \infty, t]) = \mathbb{P}(X \leq t)$ pour tout $t \in \mathbb{R}$.

Définition 26 Soit X une variable aléatoire réelle. On appelle fonction de répartition de X , l'application

$$F_X : \begin{cases} \mathbb{R} & \rightarrow [0, 1] \\ t & \mapsto \mathbb{P}(X \leq t) = \mathbb{P}_X(] - \infty, t]). \end{cases}$$

Proposition 30 La fonction de répartition d'une variable aléatoire réelle X caractérise sa loi.

Cette proposition découle directement des remarques précédentes. En particulier, on peut en déduire la probabilité pour une v.a. d'être dans un intervalle.

Proposition 31

- $\mathbb{P}(X \in]a, b]) = F_X(b) - F_X(a)$ car

$$F_X(a) + \mathbb{P}(X \in]a, b]) = \mathbb{P}((X \leq a) \cup (X \in]a, b])) = \mathbb{P}(X \leq b) = F_X(b).$$

- $\mathbb{P}(X \in]a, b]) = \mathbb{P}(X \in \cup_{n \geq 1}]a, b - 1/n]) = \mathbb{P}(\cup_{n \geq 1} (X \in]a, b - 1/n])) = \lim_{n \rightarrow +\infty} F_X(b - 1/n) - F_X(a)$ car l'union est croissante.
- $\mathbb{P}(X \in [a, b]) = \mathbb{P}(X \in \cap_{n \geq 1}]a - 1/n, b]) = \mathbb{P}(\cap_{n \geq 1} (X \in]a - 1/n, b])) = \lim_{n \rightarrow +\infty} F_X(b) - F_X(a - 1/n)$ car l'intersection est décroissante.
- $\mathbb{P}(X \in [a, b]) = \mathbb{P}(X \in \cup_{n \geq 1}]a, b - 1/n]) = \mathbb{P}(\cup_{n \geq 1} (X \in]a, b - 1/n])) = \lim_{n \rightarrow +\infty} F_X(b - 1/n) - \lim_{k \rightarrow +\infty} F_X(a - 1/k)$ en utilisant le point 3) et la croissance de l'union.

Conséquence : Si deux v.a. ont même fonction de répartition, elles ont même loi!

Remarque 25 Si X est une v.a. discrète à valeurs dans $E \subset \mathbb{R}$, alors X est une v.a. réelle, on peut donc calculer la fonction de répartition.

Exemple 4 Soit $X \sim \mathcal{B}(p)$. Alors

$$\mathbb{P}(X \leq t) = \begin{cases} 0 & \text{si } t < 0 \\ 1 - p & \text{si } 0 \leq t < 1 \\ 1 & \text{sinon.} \end{cases}$$

Proposition 32 Soit X une variable aléatoire réelle de fonction de répartition F_X . Alors on a

1. F_X est une fonction croissante.
2. $\lim_{t \rightarrow -\infty} F_X(t) = 0$ et $\lim_{t \rightarrow +\infty} F_X(t) = 1$.
3. F_X est continue à droite et possède une limite à gauche en tout point (càdlàg) : Si $t_n \searrow t$ alors $F_X(t_n) \rightarrow F_X(t)$, si $t_n \nearrow t$ alors $F_X(t_n)$ a une limite notée $F_X(t^-)$.

Preuve.

- Soit $s \leq t$, alors $] -\infty, s] \subset] -\infty, t]$ donc $F_X(s) \leq F_X(t)$.
- Soit $t_n \searrow +\infty$, alors $\Omega = \cup_{n \in \mathbb{N}} (X \leq t_n)$. Comme l'union est croissante on a

$$1 = \mathbb{P}(\Omega) = \lim_{n \rightarrow +\infty} \mathbb{P}(X \leq t_n) = \lim_{n \rightarrow +\infty} F_X(t_n).$$

- Soit $t_n \nearrow -\infty$, alors $\emptyset = \cap_{n \in \mathbb{N}} (X \leq t_n)$. Comme l'intersection est décroissante on a

$$0 = \mathbb{P}(\emptyset) = \lim_{n \rightarrow +\infty} \mathbb{P}(X \leq t_n) = \lim_{n \rightarrow +\infty} F_X(t_n).$$

- Soit $t \in \mathbb{R}$ et $t_n \searrow t$, alors $\cup_{n \in \mathbb{N}} (X \leq t_n) = (X \leq t)$. Donc $\lim_{n \rightarrow +\infty} F_X(t_n) = F_X(t)$.
- Soit $t \in \mathbb{R}$ et $t_n \nearrow t$, alors $\cup_{n \in \mathbb{N}} (X \leq t_n) = (X < t)$. Donc $\lim_{n \rightarrow +\infty} F_X(t_n)$ existe et vaut $\mathbb{P}(X < t)$.

□

Remarque 26 Soit $t \in \mathbb{R}$. $(X = t) \cup (X < t) = (X \leq t)$ donc $\mathbb{P}(X = t) = F_X(t) - F_X(t^-)$. En particulier, $\mathbb{P}(X = t) > 0$ ssi F_X admet un saut en t de taille $\mathbb{P}(X = t)$.

Proposition 33 Soit X une variable aléatoire discrète réelle. Alors F_X est une fonction “en escalier”. Plus précisément, elle est croissante, càdlàg, constante par morceaux et “saute” uniquement aux points $t \in X(\Omega)$, la valeur du saut valant $\mathbb{P}(X = t)$.

Preuve. Soit $t \in \mathbb{R}$, on a $(X \in] - \infty, t]) = (X \in] - \infty, t] \cap X(\Omega))$ donc

$$F_X(t) = \mathbb{P}(X \in] - \infty, t] \cap X(\Omega)) = \sum_{x_i \leq t, x_i \in X(\Omega)} \mathbb{P}(X = x_i).$$

Prenons (si elles existent) deux valeurs $x_k, x_j \in X(\Omega)$ telles que $x_k < x_j$ et $]x_k, x_j[\cap X(\Omega) = \emptyset$. Alors, pour tout $t \in [x_k, x_j[$, on a

$$F_X(t) = \sum_{x_i \leq t, x_i \in X(\Omega)} \mathbb{P}(X = x_i) = \sum_{x_i \leq x_k, x_i \in X(\Omega)} \mathbb{P}(X = x_i) = F_X(x_k).$$

De plus,

$$\begin{aligned} F_X(x_j) &= \sum_{x_i \leq x_j, x_i \in X(\Omega)} \mathbb{P}(X = x_i) = \left(\sum_{x_i \leq x_k, x_i \in X(\Omega)} \mathbb{P}(X = x_i) + \mathbb{P}(X = x_j) \right) \\ &= F_X(x_k) + \mathbb{P}(X = x_k). \end{aligned}$$

□

Proposition 34 Soit $F : \mathbb{R} \rightarrow [0, 1]$ une fonction croissante, càdlàg, qui tend vers 0 en $-\infty$ et vers 1 en $+\infty$. Alors il existe une v.a. de fonction de répartition F .

Preuve. On considère $\Omega =]0, 1[$ muni de la tribu \mathcal{A} des boréliens sur Ω et de \mathbb{P} la restriction de la mesure de Lebesgue à $]0, 1[$ (c’est bien une mesure telle que $\mathbb{P}(\Omega) = 1$ donc c’est une mesure de probabilité). Pour tout $\omega \in \Omega$ on pose

$$X(\omega) = \inf\{x, F(x) \geq \omega\}.$$

Remarquons que, comme $\lim_{x \rightarrow -\infty} F(x) = 0$ et $\lim_{x \rightarrow +\infty} F(x) = 1$, $X(\omega) \in \mathbb{R}$. On va montrer que X a pour fonction de répartition F ce qui démontre le résultat.

Tout d’abord, on peut remarquer que si F est continue et strictement croissante, alors F est inversible et $X(\omega) = F^{-1}(\omega)$. Dans ce cas on a, pour tout $t \in \mathbb{R}$

$$\mathbb{P}(X \leq t) = \mathbb{P}(\{\omega, F^{-1}(\omega) \leq t\}) = \mathbb{P}(\{\omega, \omega \leq F(t)\}) = \mathbb{P}(]0, F(t)]) = F(t).$$

Dans le cas général, F n'est pas nécessairement inversible, par contre on peut montrer que pour tout $t \in \mathbb{R}$

$$(X \leq t) = \{\omega, \omega \leq F(t)\} \tag{1}$$

ce qui permet de conclure par le même calcul que précédemment. On montre (1) par double inclusion. On fixe $t \in \mathbb{R}$.

- Si $\omega \leq F(t)$ alors $X(\omega) \leq t$ par définition de l'inf.
- Inversement, on suppose $X(\omega) \leq t$. Il existe donc une suite décroissante $(x_k)_{k \in \mathbb{N}}$ telle que $\lim_{k \rightarrow +\infty} x_k = X(\omega)$ et $F(x_k) \geq \omega$ pour tout $k \in \mathbb{N}$. Par continuité à droite de F , on a $\lim_{k \rightarrow +\infty} F(x_k) = F(X(\omega)) \geq \omega$. Comme F est croissante, on en déduit que $\omega \leq F(X(\omega)) \leq F(t)$.

□

4 Variables aléatoires à densité

4.1 Définitions, propriétés

Soit X une variable aléatoire réelle.

Définition 27 On dit que X est une v.a. à densité (ou absolument continue) si sa loi \mathbb{P}_X est absolument continue par rapport à la mesure de Lebesgue, c'est-à-dire que : pour tout $A \in \mathcal{B}(\mathbb{R})$ tel que $\lambda(A) = 0$, on a $\mathbb{P}_X(A) = 0$.

Théorème 4 (Théorème de Radon-Nikodym) X est une v.a. à densité ssi il existe une fonction mesurable $f \geq 0$ telle que pour tout borélien $B \in \mathcal{B}(\mathbb{R})$ et donc pour tout intervalle I de \mathbb{R} , on a

$$\mathbb{P}(X \in B) = \int_B f(u)du, \quad \mathbb{P}(X \in I) = \int_I f(u)du.$$

Cette fonction f est appelée densité de la variable X .

Rappelons que la fonction de répartition caractérise la loi d'une v.a. réelle X et que l'on a

$$F_X(t) = \mathbb{P}(X \leq t) = \mathbb{P}(X \in]-\infty, t]).$$

Ceci nous amène donc au résultat suivant :

Proposition 35 X est une variable aléatoire à densité si et seulement si il existe une fonction mesurable $f : \mathbb{R} \rightarrow \mathbb{R}^+$ telle que la fonction de répartition de X s'écrit

$$F_X(t) = \int_{-\infty}^t f(u)du. \quad (2)$$

Remarque 27 (Importante !) Pour montrer que F_X s'écrit comme (2), il suffit d'avoir

1. F_X continue,
2. F_X C^1 par morceaux.

Dans ce cas, on a $f(t) = F'_X(t)$ en tout point t où F_X est dérivable.

Obtenir une condition nécessaire et suffisante sur F_X pour quelle s'écrive comme (2) est une question difficile. La bonne classe à considérer est la classe des fonctions absolument continues. Pour se convaincre de la difficulté de la question, on peut considérer la fonction « escalier de Cantor » (ou escalier du diable), notée ϕ qui a la propriété d'être continue, croissante, de valeur 0 en 0, 1 en 1 et de dérivée nulle λ -p.p. (avec λ la mesure de Lebesgue) : On a alors $\phi(1) - \phi(0) = 1$ et $\int_0^1 \phi'(u)du = 0$...

Proposition 36 Soit X une v.a. à densité, de densité f .

1. f caractérise la loi de X .
2. $\mathbb{P}(X = a) = \int_a^a f(u)du = 0$ pour tout $a \in \mathbb{R}$. En particulier, une v.a. discrète n'est pas à densité et une v.a. à densité n'est pas discrète !
3. Pour tout $a < b$ on a

$$\mathbb{P}(X \in [a, b]) = \mathbb{P}(X \in [a, b[) = \mathbb{P}(X \in]a, b]) = \mathbb{P}(X \in]a, b[) = \int_a^b f(u)du.$$

4. $\int_{\mathbb{R}} f(u)du = \mathbb{P}(X \in \mathbb{R}) = 1$.

Remarque 28 Si f est continue en x et $\varepsilon > 0$ « petit », alors

$$\mathbb{P}(X \in [x - \varepsilon/2, x + \varepsilon/2]) = \int_{x-\varepsilon/2}^{x+\varepsilon/2} f(u)du \simeq \varepsilon f(x).$$

Donc la densité est grande près de x si la probabilité de prendre des valeurs proche de x est « grande ».

Proposition 37 Si $f : \mathbb{R} \rightarrow \mathbb{R}$ est une fonction mesurable telle que

1. $f \geq 0$ p.p.
2. $\int_{\mathbb{R}} f(x)dx = 1$,

alors f est la densité d'une certaine v.a. X .

Preuve. Il suffit de remarquer que $t \mapsto \int_{-\infty}^t f(u)du$ est une fonction croissante, continue, qui tend vers 0 en $-\infty$ et 1 en $+\infty$, puis appliquer la proposition 34. \square

4.2 Espérance et variance

Rappelons que si $\mathbb{E}[|X|] < +\infty$ alors

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

Proposition 38 Soit X une variable aléatoire à densité, de densité f . X est intégrable si et seulement si

$$\mathbb{E}[|X|] = \int_{\mathbb{R}} |x|f(x)dx < +\infty.$$

Dans ce cas, l'espérance de X est alors égale à

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

Proposition 39 Soit X une variable aléatoire à densité, de densité f , et h une fonction mesurable réelle. On suppose que

$$\int_{\mathbb{R}} |h(x)|f(x)dx < +\infty.$$

Alors, si on pose $Y = h(X)$, Y est une variable aléatoire intégrable et

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

Attention, la v.a. Y n'est pas nécessairement à densité! Prenons l'exemple de la fonction $h = 0$: alors $Y = 0$, c'est-à-dire que Y est une v.a. constante nulle, elle est donc discrète et ne peut pas être à densité.

Exemple 5 On a

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

Ainsi, sous réserve d'existence, on obtient

$$\text{Var}(X) = \int_{\mathbb{R}} x^2 f(x)dx - \left(\int_{\mathbb{R}} x f(x)dx \right)^2.$$

4.3 Indépendance

Soient X et Y deux v.a. réelles. Rappelons que X et Y sont indépendantes si la loi de (X, Y) est la mesure produit $\mathbb{P}_X \times \mathbb{P}_Y$. Cela peut se réécrire également sous la forme suivante pour des v.a. réelles :

Proposition 40 X et Y sont indépendantes si et seulement si pour tous intervalles réels I et J on a

$$\mathbb{P}(X \in I, Y \in J) = \mathbb{P}(X \in I)\mathbb{P}(Y \in J).$$

4.4 Quelques lois usuelles

4.4.1 Loi uniforme

Définition 28 Soient $a, b \in \mathbb{R}$ avec $a < b$. X suit la loi uniforme sur l'intervalle $[a, b]$ si sa densité vaut

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

On note $X \sim \mathcal{U}([a, b])$.

Proposition 41 Si $[c, d] \subset [a, b]$ alors $\mathbb{P}(X \in [c, d]) = \frac{d-c}{b-a}$.

Proposition 42 $X \sim \mathcal{U}([a, b])$. Alors on a

$$F_X(t) = \begin{cases} 0 & \text{si } t \leq a, \\ \frac{t-a}{b-a} & \text{si } t \in [a, b], \\ 1 & \text{si } t \geq b. \end{cases}$$

Proposition 43 Si $X \sim \mathcal{U}([a, b])$,

$$\mathbb{E}[X] = \frac{a+b}{2}, \quad \text{Var}(X) = \frac{(b-a)^2}{12}.$$

4.4.2 Loi exponentielle

Définition 29 Soient $\lambda > 0$. X suit la loi exponentielle de paramètre λ si

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

On note $X \sim \mathcal{E}(\lambda)$.

Proposition 44 $X \sim \mathcal{E}(\lambda)$. Alors on a

$$F_X(t) = \begin{cases} 0 & \text{si } t \leq 0, \\ 1 - e^{-\lambda t} & \text{si } t \geq 0. \end{cases}$$

Proposition 45 Si $X \sim \mathcal{E}(\lambda)$,

$$\mathbb{E}[X] = \frac{1}{\lambda}, \quad \text{Var}(X) = \frac{1}{\lambda^2}.$$

4.4.3 Loi de Cauchy

Définition 30 Soient $c > 0$. X suit la loi de Cauchy de paramètre c si

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

On note $X \sim \mathcal{C}(c)$.

Proposition 46 $X \sim \mathcal{C}(c)$. Alors on a

$$F_X(t) = \frac{1}{\pi} \arctan\left(\frac{t}{c}\right) + \frac{1}{2}.$$

$\mathbb{E}[X]$ n'existe pas! En effet X prend ses valeurs dans tout \mathbb{R} et

$$\mathbb{E}[|X|] = \int_{\mathbb{R}} \frac{c|x|}{\pi(c^2 + x^2)} dx = \frac{2c}{\pi} \int_0^{+\infty} \frac{x}{c^2 + x^2} dx = +\infty.$$

4.4.4 Loi normale centrée réduite

Définition 31 X suit la loi normale centrée réduite si

$$f_X(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right).$$

On note $X \sim \mathcal{N}(0, 1)$.

Notons qu'il n'existe pas de formule explicite pour la fonction de répartition de la loi normale centrée réduite.

Proposition 47 $X \sim \mathcal{N}(0, 1)$. On a

$$\mathbb{E}[X] = 0, \quad \text{Var}(X) = 1.$$

4.5 Quelques exemples de calculs de lois

On est souvent confronté au problème suivant : Soient X une v.a. dont on connaît la loi et $h : \mathbb{R} \rightarrow \mathbb{R}$ une fonction mesurable. On pose $Y = h(X)$. Peut-on déterminer la loi de Y ? Un outil possible pour répondre à cette question est le calcul de la fonction de répartition de Y .

4.5.1 Exemple 1

On pose $Y = X^2$ avec $X \sim \mathcal{U}([0, 1])$. On va calculer F_Y la fonction de répartition de Y .

Tout d'abord, on a $X(\Omega) = [0, 1]$ donc $Y(\Omega) = [0, 1]$. En particulier, $F_Y(t) = 0$ si $t < 0$ et $F_Y(t) = 1$ si $t \geq 1$.

Si $t \in [0, 1]$, alors

$$\begin{aligned} F_Y(t) &= \mathbb{P}(Y \leq t) = \mathbb{P}(X^2 \leq t) = \mathbb{P}(-\sqrt{t} \leq X \leq \sqrt{t}) \\ &= \mathbb{P}(X \leq \sqrt{t}) = F_X(\sqrt{t}). \end{aligned}$$

Comme on a $t \in [0, 1] \Leftrightarrow \sqrt{t} \in [0, 1]$, on en déduit que $F_Y(t) = \sqrt{t}$.

En conclusion, F_Y est C^0 sur \mathbb{R} et C^1 sur $\mathbb{R} \setminus \{0, 1\}$. Donc Y est une v.a. à densité, de densité g donnée par

$$g(y) = \frac{1}{2\sqrt{y}} \mathbb{1}_{]0,1[}(y).$$

4.5.2 Exemple 2

On considère $X \sim \mathcal{E}(\lambda)$ et $Y \sim \mathcal{E}(\mu)$ deux v.a. indépendantes. On pose $Z = \min(X, Y)$. Z est une v.a. à valeurs dans \mathbb{R}^+ . En particulier, $\mathbb{P}(Z > t) = 1$ si $t < 0$. Si $t \geq 0$, on a

$$\mathbb{P}(Z > t) = \mathbb{P}(X > t, Y > t) = \mathbb{P}(X > t)\mathbb{P}(Y > t)$$

par indépendance de X et Y . Or

$$\mathbb{P}(X > t) = 1 - \mathbb{P}(X \leq t) = e^{-\lambda t}$$

donc $\mathbb{P}(Z > t) = e^{-(\lambda+\mu)t}$. Au final, on trouve

$$F_Z(t) = 1 - \mathbb{P}(Z > t) = \begin{cases} 0 & \text{si } t < 0, \\ 1 - e^{-(\lambda+\mu)t} & \text{sinon,} \end{cases}$$

et donc $Z \sim \mathcal{E}(\lambda + \mu)$ car on reconnaît la fonction de répartition.

4.5.3 Exemple 3

Soit X une v.a. à densité, de densité f continue par morceaux. Soient $m \in \mathbb{R}$ et $\sigma > 0$ deux paramètres. On pose $Y = m + \sigma X$. Alors

$$F_Y(t) = \mathbb{P}(\sigma X + m \leq t) = \mathbb{P}\left(X \leq \frac{t - m}{\sigma}\right) = F_X\left(\frac{t - m}{\sigma}\right).$$

F_X est continue sur \mathbb{R} et dérivable par morceaux, donc Y est une v.a. à densité et sa densité est donnée par

$$g(t) = F'_Y(t) = \frac{1}{\sigma} f\left(\frac{t - m}{\sigma}\right).$$

Voici un exemple d'application pour la loi normale.

Proposition 48 Si $X \sim \mathcal{N}(0, 1)$, $m \in \mathbb{R}$, $\sigma > 0$ et $Y := m + \sigma X$, alors Y est une v.a. à densité, de densité

$$f_Y(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x - m)^2}{2\sigma^2}\right).$$

On note $Y \sim \mathcal{N}(m, \sigma^2)$. Inversement, si $Y \sim \mathcal{N}(m, \sigma^2)$, alors $\frac{Y - m}{\sigma} \sim \mathcal{N}(0, 1)$.

On en déduit facilement l'espérance et la variance d'une loi normale quelconque.

Proposition 49 Si $X \sim \mathcal{N}(m, \sigma^2)$ avec $m \in \mathbb{R}$ et $\sigma > 0$, alors

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

5 Couples et vecteurs aléatoires : cas discret

Définition 32 (X_1, \dots, X_n) est un vecteur aléatoire défini sur $(\Omega, \mathcal{A}, \mathbb{P})$ si chaque X_i est une variable aléatoire définie sur $(\Omega, \mathcal{A}, \mathbb{P})$. Lorsque $n = 2$ on parle de couple de variables aléatoires.

$$\begin{aligned} X_i : \Omega &\rightarrow E_i := X_i(\Omega) \\ \omega &\mapsto X_i(\omega) \end{aligned}$$

$$\begin{aligned} (X_1, \dots, X_n) : \Omega &\rightarrow E_1 \times \dots \times E_n \\ \omega &\mapsto (X_1(\omega), \dots, X_n(\omega)) \end{aligned}$$

Proposition et définition 3 On dit que (X_1, \dots, X_n) est un vecteur discret si $E_1 \times \dots \times E_n$ est un ensemble discret. En particulier, (X_1, \dots, X_n) est discret si et seulement si X_1, \dots, X_n sont des v.a. discrètes.

5.1 Loi d'un vecteur aléatoire discret

Dans le cas discret, la loi du vecteur aléatoire est déterminée par la donnée de E_1, \dots, E_n et par l'ensemble des valeurs

$$\mathbb{P}(X_1 = x_1, \dots, X_n = x_n), \quad x_1 \in E_1, \dots, x_n \in E_n.$$

On parle parfois de loi jointe pour la loi du vecteur (X_1, \dots, X_n) .

Exemple : On considère $X_1 \sim \mathcal{B}(p)$ et $X_2 \sim \mathcal{B}(p)$ avec X_1, X_2 indépendantes et $p = 1/4$. On pose

$$X = \max(X_1, X_2), \quad Y = \min(X_1, X_2).$$

Calculons la loi de (X, Y) . Tout d'abord, nous avons

$$E_1 = X(\Omega) = \{0, 1\}, \quad E_2 = Y(\Omega) = \{0, 1\}.$$

Ensuite, nous pouvons calculer les probabilités

$$\begin{aligned} \mathbb{P}(X = 0, Y = 0) &= \mathbb{P}(X_1 = 0, X_2 = 0) = \mathbb{P}(X_1 = 0)\mathbb{P}(X_2 = 0) \quad \text{par indépendance} \\ &= (1 - p)^2 = 9/16, \end{aligned}$$

$$\mathbb{P}(X = 0, Y = 1) = 0,$$

$$\mathbb{P}(X = 1, Y = 0) = \mathbb{P}(X_1 = 1, X_2 = 0) + \mathbb{P}(X_1 = 0, X_2 = 1) = 2p(1 - p) = 6/16,$$

$$\mathbb{P}(X = 1, Y = 1) = \mathbb{P}(X_1 = 1)\mathbb{P}(X_2 = 1) = p^2 = 1/16.$$

Ainsi la loi jointe de (X, Y) est résumée par le tableau suivant :

$Y \setminus X$	0	1
0	9/16	6/16
1	0	1/16

Remarquons que nécessairement, la somme des éléments du tableau doit faire 1. Plus généralement on a la propriété suivante.

Proposition 50 *Soit (X_1, \dots, X_n) un vecteur aléatoire discret, alors on a*

$$\sum_{(x_1, \dots, x_n) \in E_1 \times \dots \times E_n} \mathbb{P}(X_1 = x_1, \dots, X_n = x_n) = 1.$$

5.2 Lois marginales

Définition 33 *On appelle lois marginales d'un vecteur aléatoire (X_1, \dots, X_n) , les lois des variables aléatoires X_1, \dots, X_n .*

n = 2 : On considère deux variables aléatoires discrètes

$$\begin{aligned} X : \Omega &\rightarrow E_1 = \{x_i, 0 \leq i < M\} \\ Y : \Omega &\rightarrow E_2 = \{y_j, 0 \leq j < N\}. \end{aligned}$$

On suppose que l'on connaît la loi jointe de (X, Y) et on cherche à calculer la loi de X et de Y . Soit $x_i \in E_1$, alors

$$\begin{aligned} \mathbb{P}(X = x_i) &= \mathbb{P}((X = x_i) \cap (\cup_{y_j \in E_2} (Y = y_j))) = \mathbb{P}(\cup_{y_j \in E_2} (X = x_i, Y = y_j)) \\ &= \sum_{y_j \in E_2} \mathbb{P}(X = x_i, Y = y_j). \end{aligned}$$

De la même façon on a, lorsque $y_j \in E_2$,

$$\mathbb{P}(Y = y_j) = \sum_{x_i \in E_1} \mathbb{P}(X = x_i, Y = y_j).$$

Cas général :

Proposition 51 *Soient X_1, \dots, X_n des variables aléatoires discrètes. On note*

$$E_i = \{x_k, 0 \leq k < N_i\}, \quad 1 \leq i \leq n.$$

Alors on a

$$\mathbb{P}(X_i = x) = \sum_{(x_1, \dots, x_n) \in E_1 \times \dots \times E_n, x_i = x} \mathbb{P}(X_1 = x_1, \dots, X_n = x_n).$$

En particulier, la loi jointe d'un vecteur aléatoire discret permet de calculer ses lois marginales.

Revenons au premier exemple de ce chapitre. On peut calculer les lois de X et de Y .

$$\begin{aligned} \mathbb{P}(X = 0) &= \mathbb{P}(X = 0, Y = 0) + \mathbb{P}(X = 0, Y = 1) = 9/16, \\ \mathbb{P}(X = 1) &= \mathbb{P}(X = 1, Y = 0) + \mathbb{P}(X = 1, Y = 1) = 7/16, \\ \mathbb{P}(Y = 0) &= \mathbb{P}(X = 0, Y = 0) + \mathbb{P}(X = 1, Y = 0) = 15/16, \\ \mathbb{P}(Y = 1) &= \mathbb{P}(X = 0, Y = 1) + \mathbb{P}(X = 1, Y = 1) = 1/16. \end{aligned}$$

En particulier, $X \sim \mathcal{B}(7/16)$ et $Y \sim \mathcal{B}(1/16)$. Ces calculs reviennent à faire les sommes sur les lignes et les colonnes du tableau.

$Y \backslash X$	0	1	
0	9/16	6/16	15/16
1	0	1/16	1/16
	9/16	7/16	

Remarque 29 (Importante)

- Les lois marginales d'un vecteur aléatoire ne permettent pas de déterminer la loi jointe de ce vecteur.
- Si l'on connaît les lois marginales d'un vecteur (X_1, \dots, X_n) et que ses composantes sont indépendantes, alors on peut calculer la loi jointe de ce vecteur.

Exemple 1 : Il existe des couples de variables aléatoires qui ont mêmes lois marginales mais qui n'ont pas la même loi jointe.

$Y \backslash X$	0	1	
0	1/4	1/4	1/2
0	1/4	1/4	1/2
	1/2	1/2	

$Y \backslash X$	0	1	
0	1/2	0	1/2
1	0	1/2	1/2
	1/2	1/2	

Exemple 2 : Soient $X \sim \mathcal{B}(7/16)$ et $Y \sim \mathcal{B}(1/16)$ indépendantes. On peut calculer la loi du couple (X, Y) .

$Y \backslash X$	0	1	
0	$9/16 * 15/16$	$7/16 * 15/16$	$15/16$
1	$9/16 * 1/16$	$7/16 * 1/16$	$1/16$
	$9/16$	$7/16$	

5.3 Covariance

Définition 34 Soient X et Y deux variables aléatoires de carré intégrable (pas nécessairement discrètes). La covariance de X et de Y est donnée par

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

Proposition 52

- $\text{Cov}(X, X) = \text{Var}(X)$,
- $\text{Cov}(X, Y) = \text{Cov}(Y, X)$,
- $\text{Cov}(\alpha X_1 + \beta X_2, Y) = \alpha \text{Cov}(X_1, Y) + \beta \text{Cov}(X_2, Y)$.

En particulier, la covariance est une forme bilinéaire symétrique positive.

Attention : la covariance est une forme bilinéaire symétrique positive qui n'est pas définie positive. En effet, $\text{Cov}(X, X) = \text{Var}(X) = 0$ si et seulement si X est une v.a. constante. Ce n'est donc pas un produit scalaire... On peut néanmoins voir la covariance comme un produit scalaire sur le sous-espace des variables aléatoires centrées (i.e. d'espérance nulle).

Proposition 53

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

Preuve. On applique juste l'inégalité de Cauchy-Schwartz :

$$\begin{aligned} |\mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]| &\leq \mathbb{E}[|X - \mathbb{E}[X]| |Y - \mathbb{E}[Y]|] \\ &\leq \mathbb{E}[(X - \mathbb{E}[X])^2]^{1/2} \mathbb{E}[(Y - \mathbb{E}[Y])^2]^{1/2}. \end{aligned}$$

□

Définition 35 On peut définir le coefficient de corrélation de X et Y par

$$r(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X) \text{Var}(Y)}} \in [-1, 1],$$

si $\text{Var}(X) \text{Var}(Y) > 0$.

Une covariance positive indique que les deux variables aléatoires varient dans le « même sens » tandis qu'une covariance négative indique au contraire des variations dans des « sens opposés ». On peut par exemple regarder les cas extrêmes donnés par $|r(X, Y)| = 1$. Cela correspond à un cas d'égalité dans l'inégalité de Cauchy-Schwartz ce qui signifie que $X - \mathbb{E}[X]$ et $Y - \mathbb{E}[Y]$ sont proportionnelles, ou dit autrement, $Y = aX + b$. De plus $r(X, Y) = 1$ correspond à $a > 0$ tandis que $r(X, Y) = -1$ lorsque $a < 0$.

Proposition 54

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

Preuve.

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

□

Proposition 55 Soient X et Y deux v.a. discrètes, alors

$$\mathbb{E}[XY] = \sum_{x \in X(\Omega), y \in Y(\Omega)} xy \mathbb{P}(X = x, Y = y).$$

Exemple : Calculons la covariance des variables X et Y du premier exemple de ce chapitre.

$$\mathbb{E}[XY] = \sum_{x \in X(\Omega), y \in Y(\Omega)} \mathbb{P}(X = x, Y = y) = 1 \times \mathbb{P}(X = 1, Y = 1) = 1/16,$$

et donc

$$Cov(X, Y) = 1/16 - 1/16 \times 7/16 = 9/16^2 > 0.$$

Proposition 56

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

En particulier, si X et Y sont indépendantes alors $Cov(X, Y) = 0$.

Preuve.

$$\begin{aligned}\text{Var}(X + Y) &= \mathbb{E}[(X + Y)^2] - \mathbb{E}[X + Y]^2 \\ &= \mathbb{E}[X^2 + Y^2 + 2XY] - (\mathbb{E}[X]^2 + \mathbb{E}[Y]^2 + 2\mathbb{E}[X]\mathbb{E}[Y]) \\ &= \mathbb{E}[X^2] - \mathbb{E}[X]^2 + \mathbb{E}[Y^2] - \mathbb{E}[Y]^2 + 2(\mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]).\end{aligned}$$

□

Remarque 30 *La réciproque est fautive : si X et Y sont non corrélées, elles ne sont pas nécessairement indépendantes. Prenons par exemple, $X \sim \mathcal{U}(\{-1, 0, 1\})$ et $Y = \mathbb{1}_{X=0}$. Alors on a $\mathbb{E}[X] = 0$, $Y \sim \mathcal{B}(1/3)$, $\mathbb{E}[Y] = 1/3$, $XY = 0$, $\mathbb{E}[XY] = 0$ et $\text{Cov}(XY) = 0$. Par contre X et Y ne sont pas indépendantes : en effet on a*

$$\mathbb{P}(X = 0, Y = 0) = 0, \quad \mathbb{P}(X = 0) = 1/3, \quad \mathbb{P}(Y = 0) = 2/3.$$

5.4 Variables aléatoires à valeurs dans \mathbb{N}

Définition 36 *Soit X une variable aléatoire à valeurs dans \mathbb{N} . On définit sa fonction génératrice (ou série génératrice) par*

$$G_X(s) = \mathbb{E}[s^X] = \sum_{k \geq 0} s^k \mathbb{P}(X = k).$$

Proposition 57

- G_X est une série entière,
- $G_X(1) = 1$, donc en particulier son rayon de convergence R est supérieur ou égal à 1,
-

$$\mathbb{P}(X = k) = \frac{1}{k!} G_X^{(k)}(0), \quad k \in \mathbb{N}.$$

La proposition suivante découle directement du troisième point précédent.

Proposition 58 *La série génératrice d'une v.a X à valeurs dans \mathbb{N} caractérise sa loi.*

Proposition 59 *Soient X et Y deux v.a. indépendantes à valeurs dans \mathbb{N} . Alors $X + Y$ est une v.a. à valeurs dans \mathbb{N} et*

$$G_{X+Y}(s) = G_X(s)G_Y(s), \quad |s| < 1.$$

Preuve. Par indépendance on a

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

□

Remarquons qu'il est également possible de faire directement le calcul : Soit $n \in \mathbb{N}$,

$$\mathbb{P}(X + Y = n) = \mathbb{P}\left(\bigcup_{0 \leq k \leq n} ((X = k) \cap (Y = n - k))\right) = \sum_{k=0}^n \mathbb{P}(X = k) \mathbb{P}(Y = n - k).$$

Exercice 1 Soient $X \sim \mathcal{P}(\lambda)$ et $Y \sim \mathcal{P}(\mu)$ indépendantes. Montrer que $X + Y \sim \mathcal{P}(\lambda + \mu)$.

Proposition 60 Si $|s| < R$ alors on peut dériver termes à termes pour obtenir

$$G'_X(s) = \sum_{k=1}^{+\infty} k s^{k-1} \mathbb{P}(X = k).$$

En particulier, si $R > 1$, $G'(1) = \mathbb{E}[X] < +\infty$. Inversement, si $\mathbb{E}[X] < +\infty$, le théorème radial d'Abel nous donne $G'_X(s) \rightarrow \mathbb{E}[X]$ pour $s \rightarrow 1^-$.

De même, si $\mathbb{E}[X^2] < +\infty$,

$$\begin{aligned} G''_X(s) &= \sum_{k \geq 0} k(k-1) \mathbb{P}(X = k) s^{k-2} = \sum_{k \geq 0} k^2 \mathbb{P}(X = k) s^{k-2} - \sum_{k \geq 0} k \mathbb{P}(X = k) s^{k-2} \\ &\rightarrow \mathbb{E}[X^2] - \mathbb{E}[X] \quad \text{Lorsque } s \rightarrow 1^-. \end{aligned}$$

Exercice 2 Calculer la fonction génératrice de la loi $\mathcal{G}(p)$. En déduire son espérance et sa variance.

6 Caractérisation de la loi d'une variable aléatoire réelle

Dans ce chapitre nous allons lister tous les outils disponibles dans ce cours pour pouvoir caractériser la loi d'une variable réelle. On considère donc X une variable aléatoire réelle définie sur un espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$.

Proposition 61 (Caractérisation 1, fonction de répartition) *La fonction de répartition de X caractérise sa loi.*

Proposition 62 (Caractérisation 2, loi discrète) *Si X est une v.a. discrète, sa loi est caractérisée par la donnée de $X(\Omega)$ et $\mathbb{P}(X = x)$ pour tous les $x \in X(\Omega)$.*

Proposition 63 (Caractérisation 3, méthode de la fonction muette) *La connaissance de $\mathbb{E}[h(X)]$ pour toute fonction bornée continue h caractérise la loi de X .*

Remarquons que l'on peut considérer également la classe des fonctions h mesurables positives, ou bien encore celle des fonctions h mesurables bornées.

Preuve. Pour tout $t \in \mathbb{R}$, on considère la suite de fonctions $(h_{t,n})_{n \in \mathbb{N}^*}$ donnée par

$$h_{t,n}(x) = \begin{cases} 1 & \text{si } x < t \\ 1 - n(x - t) & \text{si } x \in [t, t + 1/n] \\ 0 & \text{sinon.} \end{cases}$$

$h_{t,n}$ est continue et bornée pour tout $n \in \mathbb{N}$. De plus, $(h_{t,n})_{n \in \mathbb{N}^*}$ converge simplement, en décroissant, vers la fonction $\mathbb{1}_{x \leq t}$. Ainsi, le théorème de convergence monotone nous donne

$$\mathbb{E}[h_{t,n}(X)] \xrightarrow{n \rightarrow +\infty} \mathbb{E}[\mathbb{1}_{X \leq t}] = \mathbb{P}(X \leq t) = F_X(t).$$

En particulier, si on connaît $\mathbb{E}[h_{t,n}(X)]$ pour tous $n \in \mathbb{N}^*$ et tous $t \in \mathbb{R}$, alors on connaît la fonction de répartition de X , qui caractérise la loi de X \square

Concrètement cette caractérisation est utilisée à l'aide de la proposition suivante.

Proposition 64 (méthode de la fonction muette, cas pratiques)

- *S'il existe des réels $\{x_i, 0 \leq i < N\}$ et $\{\alpha_i, 0 \leq i < N\}$ tels que pour toute fonction h continue bornée on a*

$$\mathbb{E}[h(X)] = \sum_{0 \leq i < N} h(x_i) \alpha_i$$

alors X est une variable aléatoire discrète, telle que $X(\Omega) = \{x_i, 0 \leq i < N\}$ et $\mathbb{P}(X = x_i) = \alpha_i$ pour tout $0 \leq i < N$.

- *S'il existe une fonction réelle mesurable g telle que pour toute fonction h continue bornée on a*

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

alors X est une variable aléatoire à densité, de densité g .

On va maintenant définir un nouvel outil qui permet également de caractériser la loi d'une v.a.

Définition 37 *On définit la fonction caractéristique de la v.a. réelle X par*

$$\varphi_X(t) = \mathbb{E}[e^{itX}], \quad \forall t \in \mathbb{R}.$$

Remarquons que φ_X est bien définie pour tout $t \in \mathbb{R}$ car $|e^{itX}| = 1$ donc $\mathbb{E}[|e^{itX}|] < +\infty$. Par le théorème de transfert, on peut préciser la fonction caractéristique lorsque X est discrète ou à densité.

Proposition 65

- *Si X est discrète, alors*

$$\varphi_X(t) = \sum_{x \in X(\Omega)} e^{itx} \mathbb{P}(X = x).$$

- *Si X est une variable à densité, de densité f , alors*

$$\varphi_X(t) = \int_{\mathbb{R}} e^{itx} f(x)dx = \hat{f}(-t)$$

avec \hat{f} la transformée de Fourier de f (à une constante près).

Proposition 66 (Caractérisation 4, fonction caractéristique) *La fonction caractéristique de X caractérise sa loi.*

La preuve est admise, mais remarquons que pour les v.a. à densité cela découle de l'injectivité de la transformée de Fourier.

La fonction caractéristique est un outil pratique pour calculer les moments d'une variable aléatoire ou étudier la loi de sommes de v.a. indépendantes.

Proposition 67

- *Si $\mathbb{E}[|X|^n] < +\infty$ alors φ_X est dérivable n fois et $\varphi_X^{(n)}(0) = (i)^n \mathbb{E}[X^n]$.*

- Si X et Y sont indépendantes, alors $\varphi_{X+Y}(t) = \varphi_X(t)\varphi_Y(t)$ pour tous $t \in \mathbb{R}$.

Preuve. Pour le premier point, il suffit d'appliquer le théorème de dérivation sous le signe intégrale. Pour cela remarquons que $t \mapsto e^{itx}$ est une fonction de classe C^n pour tout $x \in \mathbb{R}$. De plus, pour tout $1 \leq k \leq n$, on a

$$\frac{d^k}{dt^k}(e^{itx}) = (ix)^k e^{itx}$$

donc

$$\left| \frac{d^k}{dt^k}(e^{itX}) \right| \leq |X|^k$$

qui est intégrable : en effet $\mathbb{E}[|X|^n] < +\infty$ donc $\mathbb{E}[|X|^k] < +\infty$ pour tout $1 \leq k \leq n$ en utilisant l'inégalité de Hölder par exemple.

Pour le second point, on a, en utilisant l'hypothèse d'indépendance :

$$\mathbb{E}[e^{it(X+Y)}] = \mathbb{E}[e^{itX} e^{itY}] = \mathbb{E}[e^{itX}] \mathbb{E}[e^{itY}].$$

□

A l'aide de cet outil il est également possible de montrer une autre caractérisation possible de la loi de X dans le cas particulier où X est bornée.

Proposition 68 *Si X est une v.a. bornée, i.e. il existe une constante M tel que $|X| \leq M$, alors les moments de X caractérisent la loi de X .*

Le résultat précédent devient faux sans l'hypothèse de bornitude de X . En effet, on peut trouver deux variables aléatoires réelles X et Y n'ayant pas même loi et telles que $\mathbb{E}[X^k] = \mathbb{E}[Y^k]$ pour tout $k \in \mathbb{N}$. On renvoie le lecteur assidu à un contre-exemple du TD.

Preuve. On suppose qu'il existe une constante M tel que $|X| \leq M$. Commençons par remarquer que

$$\varphi_X(t) = \mathbb{E}[e^{itX}] = \mathbb{E} \left[\sum_{n \geq 0} \frac{(itX)^n}{n!} \right].$$

Donc si on a le droit d'intervertir l'espérance et la somme, on aurait alors

$$\varphi_X(t) = \sum_{n \geq 0} \frac{(it)^n}{n!} \mathbb{E}[X^n],$$

c'est à dire que la connaissance de tous les moments permettrait de connaître la fonction caractéristique et donc de caractériser la loi de X . Reste donc à justifier l'interversion précédente. On a

$$\left| \frac{(itX)^n}{n!} \right| \leq \frac{t^n |X|^n}{n!} \leq \frac{t^n M^n}{n!}$$

qui est le terme général d'une série sommable donc on peut appliquer le théorème de Fubini par exemple (le fait d'avoir des fonctions à valeurs complexes ne pose pas de problème supplémentaire). \square

7 Couples et vecteurs aléatoires à densité

7.1 Vecteurs à densité

On considère $X = (X_1, \dots, X_n) : \Omega \rightarrow \mathbb{R}^n$ un vecteur aléatoire.

Définition 38 On dit que X est un vecteur aléatoire à densité si sa loi est absolument continue par rapport à la mesure de Lebesgue sur \mathbb{R}^n .

En pratique on utilise plutôt la définition équivalente suivante qui, comme dans le cas des variables aléatoires à densité, provient du théorème de Radon-Nikodym.

Proposition et définition 4 $X = (X_1, \dots, X_n)$ admet une densité si et seulement s'il existe une fonction $f : \mathbb{R}^n \rightarrow \mathbb{R}$ mesurable positive telle que pour tout $A \in \mathcal{B}(\mathbb{R}^n)$,

$$\mathbb{P}(X \in A) = \mathbb{P}((X_1, \dots, X_n) \in A) = \int_A f(x_1, \dots, x_n) dx_1 \dots dx_n.$$

f est appelée densité de X .

Remarque 31

- Il suffit de le vérifier pour tout pavé $A = I_1 \times \dots \times I_n$ avec I_1, \dots, I_n des intervalles de \mathbb{R} . On peut même se restreindre à des intervalles de la forme $I_j =] - \infty, t_j]$.
- On a nécessairement $\int_{\mathbb{R}^n} f(x_1, \dots, x_n) dx_1 \dots dx_n = 1$. Inversement, on peut montrer que si $f : \mathbb{R}^n \rightarrow \mathbb{R}^+$ est une fonction mesurable d'intégrale 1 sur \mathbb{R}^n alors c'est une densité pour un certain vecteur aléatoire.

Proposition 69 (X_1, \dots, X_n) est un vecteur à densité, de densité f , si et seulement si pour toute fonction continue bornée (ou mesurable bornée, ou mesurable positive) on a

$$\mathbb{E}[h(X_1, \dots, X_n)] = \int_{\mathbb{R}^n} h(x_1, \dots, x_n) f(x_1, \dots, x_n) dx_1 \dots dx_n.$$

Preuve. Le sens direct est juste une application du théorème de transfert. La réciproque se démontre de la même façon que dans le cas unidimensionnel, en approchant les indicatrices de pavés par des fonctions continues bornées. \square

7.2 Lois marginales

Proposition 70 Soit (X_1, \dots, X_n) un vecteur aléatoire à densité, de densité f . Alors pour tout $1 \leq i \leq n$, X_i est une variable aléatoire à densité, de densité f_{X_i} donnée par

$$f_{X_i}(x) = \int_{\mathbb{R}} \dots \int_{\mathbb{R}} f(x_1, \dots, x_{i-1}, x, x_{i+1}, \dots, x_n) dx_1 \dots dx_{i-1} dx_{i+1} \dots dx_n \quad x \in \mathbb{R}.$$

Par exemple pour $n = 2$, si (X, Y) est un couple aléatoire à densité, de densité f , alors

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

Preuve. Calculons la fonction de répartition de X_i . Soit $x \in \mathbb{R}$,

$$\begin{aligned} F_{X_i}(x) &= \mathbb{P}(X_i \leq x) = \mathbb{P}((X_1, \dots, X_n) \in \mathbb{R} \times \dots \times]-\infty, x] \times \dots \times \mathbb{R}) \\ &= \int_{x_1 \in \mathbb{R}} \dots \int_{x_i \leq x} \dots \int_{x_n \in \mathbb{R}} f(x_1, \dots, x_n) dx_1 \dots dx_n \\ &= \int_{-\infty}^x \left(\int_{x_1 \in \mathbb{R}} \dots \int_{x_n \in \mathbb{R}} f(x_1, \dots, x_n) dx_1 \dots dx_{i-1} dx_{i+1} \dots dx_n \right) dx_i \quad (\text{Fubini}). \end{aligned}$$

Donc X_i est à densité et on obtient la formule annoncée pour sa densité. \square

Remarque 32 Attention, si X et Y sont des variables aléatoires à densité, (X, Y) n'est pas nécessairement à densité ! Par exemple, (X, X) prend ses valeurs dans la première bissectrice de \mathbb{R}^2 donc ne peut pas être à densité, même si X est à densité.

Exemple 6 Soit (X, Y) un couple aléatoire de loi uniforme sur le disque unité

$$\mathcal{D} = \{(x, y) | x^2 + y^2 \leq 1\}.$$

La densité de (X, Y) est donnée par

$$f_{(X,Y)}(x, y) = \frac{1}{\pi} \mathbb{1}_{\mathcal{D}}(x, y).$$

$f_X(x) = \int_{\mathbb{R}} f(x, y) dy$. En particulier, $f_X(x) = 0$ si $x \notin [-1, 1]$. Si $x \in [-1, 1]$,

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

Donc, finalement on a

$$f_X(x) = \frac{2}{\pi} \sqrt{1-x^2} \mathbb{1}_{[-1,1]}(x)$$

et par symétrie,

$$f_Y(y) = \frac{2}{\pi} \sqrt{1-y^2} \mathbb{1}_{[-1,1]}(y).$$

En particulier, X et Y ont même loi.

Proposition 71 Soit (X, Y) un couple à densité, de densité f . Sous réserve d'existence, on a

- $\mathbb{E}[X] = \int_{\mathbb{R}} x f_X(x) dx = \int_{\mathbb{R}^2} x f(x, y) dx dy,$
- $\mathbb{E}[Y] = \int_{\mathbb{R}} y f_Y(y) dy = \int_{\mathbb{R}^2} y f(x, y) dx dy,$
- $\mathbb{E}[h(X, Y)] = \int_{\mathbb{R}^2} h(x, y) f(x, y) dx dy.$ En particulier, pour le calcul de la covariance on a $\mathbb{E}[XY] = \int_{\mathbb{R}^2} xy f(x, y) dx dy.$

7.3 Indépendance

Proposition 72 Soit (X_1, \dots, X_n) un vecteur aléatoire de densité $f_{(X_1, \dots, X_n)}$ et de densités marginales f_{X_1}, \dots, f_{X_n} . Les v.a. X_1, \dots, X_n sont indépendantes si et seulement si

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

Preuve. On va faire la preuve pour $n = 2$, le cas général se traitant de la même façon. Soient I et J deux intervalles quelconques de \mathbb{R} .

Si X et Y sont indépendantes, alors

$$\begin{aligned} \mathbb{P}((X, Y) \in I \times J) &= \mathbb{P}(X \in I, Y \in J) = \mathbb{P}(X \in I, Y \in J) \\ &= \int_I f_X(x) dx \int_J f_Y(y) dy = \iint_{I \times J} f_X(x) f_Y(y) dx dy \end{aligned}$$

en appliquant le théorème de Fubini pour les fonctions positives. Comme c'est vrai pour tous les intervalles réels I et J , on a que $f_{(X, Y)} = f_X f_Y$.

Inversement, si $f_{(X, Y)} = f_X f_Y$ alors

$$\begin{aligned} \mathbb{P}((X, Y) \in I \times J) &= \iint_{I \times J} f_X(x) f_Y(y) dx dy = \int_I f_X(x) dx \int_J f_Y(y) dy \\ &= \mathbb{P}(X \in I) \mathbb{P}(Y \in J) \end{aligned}$$

en appliquant une nouvelle fois le théorème de Fubini pour les fonctions positives. On a donc l'indépendance de X et Y en appliquant la Proposition 40. \square

Remarque 33 (Importante) Pour montrer que X_1, \dots, X_n sont indépendantes, il suffit de montrer que $f_{(X_1, \dots, X_n)}(x_1, \dots, x_n)$ est à variables séparables, c'est à dire qu'il existe des fonctions mesurables positives q_1, \dots, q_n telles que

$$f_{(X_1, \dots, X_n)}(x_1, \dots, x_n) = \prod_{i=1}^n q_i(x_i).$$

En effet, dans ce cas on a nécessairement, grâce au théorème de Fubini,

$$\begin{aligned} f_{X_i}(x_i) &= \int_{\mathbb{R}^{n-1}} f_{(X_1, \dots, X_n)}(x_1, \dots, x_n) dx_1 \dots dx_{i-1} dx_{i+1} \dots dx_n \\ &= \left(\prod_{j \neq i} \int_{\mathbb{R}} q_j(x_j) dx_j \right) q_i(x_i) := C_i q_i(x_i) \end{aligned}$$

avec C_1, \dots, C_n des constantes positives. De plus,

$$1 = \int_{\mathbb{R}} f_{X_i}(x_i) dx_i = C_i \int_{\mathbb{R}} q_i(x_i) dx_i$$

donc $C_i = \left(\int_{\mathbb{R}} q_i(x_i) dx_i \right)^{-1}$ et nécessairement $\prod_{i=1}^n C_i = 1$.

7.4 Quelques exemples de calculs de lois

7.4.1 Transformation d'un vecteur aléatoire

On considère (X_1, \dots, X_n) un vecteur aléatoire de densité f , à valeurs dans un ouvert $U \subset \mathbb{R}^n$. Soient $V \subset \mathbb{R}^n$ un ouvert et $\phi : U \rightarrow V$ une fonction mesurable. On pose $Y = (Y_1, \dots, Y_n) = \phi(X_1, \dots, X_n)$ et comme d'habitude on cherche à déterminer la loi de Y . En particulier, dans ce cs de figure, on cherche à déterminer des conditions sur ϕ pour que notre vecteur aléatoire Y soit encore à densité. Pour cela nous allons utiliser la méthode de la fonction muette.

Soit $h : \mathbb{R}^n \rightarrow \mathbb{R}$ une fonction continue bornée quelconque. Alors, d'après le théorème de transfert, on a

$$\begin{aligned} \mathbb{E}[h(Y_1, \dots, Y_n)] &= \mathbb{E}[h(\phi(X_1, \dots, X_n))] \\ &= \int_U h(\phi(x_1, \dots, x_n)) f(x_1, \dots, x_n) dx_1 \dots dx_n. \end{aligned} \quad (3)$$

Or, pour pouvoir montrer que (Y_1, \dots, Y_n) est de densité g , il faut obtenir :

$$\mathbb{E}[h(Y_1, \dots, Y_n)] = \int_V h(y_1, \dots, y_n) g(y_1, \dots, y_n) dy_1 \dots dy_n. \quad (4)$$

Pour passer de (3) à (4), il est naturel d'appliquer un changement de variable dans l'intégrale.

Théorème 5 (Rappel) Soit $\varphi : U \rightarrow V$ un C^1 -diffeomorphisme entre deux ouverts de \mathbb{R}^n (φ est une bijection C^1 et φ^{-1} est également C^1). Alors, sous réserve d'intégrabilité, on a

$$\int_U h(\varphi(x_1, \dots, x_n)) dx_1 \dots dx_n = \int_V h(y_1, \dots, y_n) |Jac \varphi^{-1}(y_1, \dots, y_n)| dy_1 \dots dy_n,$$

avec

$$Jac \varphi^{-1}(y_1, \dots, y_n) = \left| \left(\frac{\partial(\varphi^{-1})_i}{\partial y_j}(y_1, \dots, y_n) \right)_{1 \leq i, j \leq n} \right|.$$

Un bon exemple valant mieux qu'un long discours, appliquons tout cela au problème suivant : Soient X et Y deux v.a. indépendantes de loi $\mathcal{N}(0, 1)$, quelle est la loi de X/Y ?

Commençons par remarquer que X/Y est une v.a. bien définie car $\mathbb{P}(Y = 0) = 0$, Y étant une v.a. à densité. Pour répondre à la question nous allons chercher la loi de $(U, V) = (X/Y, Y)$ puis prendre la première marginale. Tout d'abord, (X, Y) est un couple à densité car X et Y sont à densité et indépendantes. Sa densité est donnée par

$$f_{(X,Y)}(x, y) = f_X(x)f_Y(y) = \frac{1}{2\pi} e^{-x^2/2 - y^2/2}.$$

Soit $h : \mathbb{R}^2 \rightarrow \mathbb{R}$ une fonction continue bornée quelconque. Alors

$$\mathbb{E}[h(U, V)] = \mathbb{E}[h(X/Y, Y)] = \iint_{\mathbb{R} \times \mathbb{R}^*} h(x/y, y) \frac{1}{2\pi} \exp\left(-\frac{x^2 + y^2}{2}\right) dx dy.$$

On pose

$$\varphi : \begin{cases} \mathbb{R} \times \mathbb{R}^* & \rightarrow \mathbb{R} \times \mathbb{R}^* \\ (x, y) & \mapsto (x/y, y) = (u, v) \end{cases}.$$

On a

$$\begin{cases} u = x/y \\ v = y \end{cases} \Leftrightarrow \begin{cases} x = uv \\ y = v \end{cases}$$

donc

$$\varphi^{-1} : \begin{cases} \mathbb{R} \times \mathbb{R}^* & \rightarrow \mathbb{R} \times \mathbb{R}^* \\ (u, v) & \mapsto (uv, v) \end{cases}.$$

De plus φ et φ^{-1} sont C^1 , donc φ est un C^1 difféomorphisme.

$$\text{Jac } \varphi^{-1}(u, v) = \begin{vmatrix} v & u \\ 0 & 1 \end{vmatrix} = v$$

ce qui nous donne, par le théorème de changement de variable,

$$\begin{aligned} \mathbb{E}[h(U, V)] &= \iint_{\mathbb{R} \times \mathbb{R}^*} h(u, v) |v| \frac{1}{2\pi} \exp\left(-\frac{u^2 v^2 + v^2}{2}\right) dudv \\ &= \iint_{\mathbb{R}^2} h(u, v) |v| \frac{1}{2\pi} \exp\left(-\frac{u^2 v^2 + v^2}{2}\right) dudv. \end{aligned}$$

Ainsi, (U, V) est une variable aléatoire à densité, de densité

$$f_{(U, V)}(u, v) = |v| \frac{1}{2\pi} \exp\left(-\frac{u^2 v^2 + v^2}{2}\right), \quad (u, v) \in \mathbb{R}^2.$$

Il en découle que U est une variable aléatoire à densité, de densité

$$\begin{aligned} f_U(u) &= \int_{\mathbb{R}} |v| \frac{1}{2\pi} \exp\left(-\frac{u^2 v^2 + v^2}{2}\right) dv = \frac{1}{\pi} \int_0^{+\infty} v \exp\left(-\frac{v^2(1+u^2)}{2}\right) dv \\ &= \frac{1}{\pi} \left[\frac{\exp\left(-\frac{v^2(1+u^2)}{2}\right)}{1+u^2} \right]_0^{+\infty} = \frac{1}{\pi} \frac{1}{1+u^2}. \end{aligned}$$

Ainsi, $X/Y \sim \mathcal{C}(1)$.

7.4.2 Loi de la somme de deux variables aléatoires indépendantes à densité

Proposition 73 Soient X et Y deux v.a. indépendantes, à densité, de densités respectives f_X et f_Y . Alors $X + Y$ est une v.a. à densité de densité

$$f_{X+Y}(t) = f_X \star f_Y(t) = \int_{\mathbb{R}} f_X(x) f_Y(t-x) dx = \int_{\mathbb{R}} f_X(t-y) f_Y(y) dy.$$

\star est appelé produit de convolution.

Remarque 34 Si X et Y sont positives, alors

$$f_{X+Y}(t) = \int_0^t f_X(s) f_Y(t-s) ds \mathbb{1}_{\mathbb{R}^+}(t).$$

On peut faire le parallèle avec la formule obtenue lorsque l'on considère la somme de deux v.a. discrètes à valeurs dans \mathbb{N} :

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

Preuve. (X, Y) est un vecteur aléatoire à densité, de densité $(x, y) \mapsto f_X(x)f_Y(y)$. Calculons la fonction de répartition de $X + Y$. Soit $t \in \mathbb{R}$,

$$F_{X+Y}(t) = \mathbb{P}(X + Y \leq t) = \mathbb{P}((X, Y) \in A_t) = \iint_{A_t} f_X(x)f_Y(y)dx dy$$

avec $A_t = \{(x, y) \in \mathbb{R}^2 | x + y \leq t\}$.

$$\begin{aligned} F_{X+Y}(t) &= \int_{\mathbb{R}} \left(\int_{-\infty}^{t-y} f_X(x)dx \right) f_Y(y)dy \\ &= \int_{\mathbb{R}} \left(\int_{-\infty}^t f_X(s - y)ds \right) f_Y(y)dy \quad (s = x + y) \\ &= \int_{-\infty}^t \left(\int_{\mathbb{R}} f_X(s - y)f_Y(y)dy \right) ds \quad (\text{Fubini}). \end{aligned}$$

On en déduit ainsi le résultat. □

Exemple d'application Soient X et Y indépendantes, de loi $\mathcal{E}(\lambda)$. Alors

$$\begin{aligned} f_{X+Y}(t) &= \int_0^t f(t-s)f(s)ds \mathbb{1}_{\mathbb{R}^+}(t) = \int_0^t \lambda^2 e^{-\lambda(t-s)} e^{-\lambda s} ds \mathbb{1}_{\mathbb{R}^+}(t) \\ &= \lambda^2 \int_0^t e^{-\lambda t} ds \mathbb{1}_{\mathbb{R}^+}(t) = \lambda^2 t e^{-\lambda t} \mathbb{1}_{\mathbb{R}^+}(t). \end{aligned}$$

On peut généraliser le résultat à la somme de n v.a. indépendantes de même loi exponentielle.

Définition 39 Soient $\lambda > 0$ et $a > 0$ deux paramètres. On appelle loi gamma de paramètres (a, λ) , et on note $\Gamma(a, \lambda)$ la loi de densité

$$f(x) = \frac{\lambda^a}{\Gamma(a)} x^{a-1} e^{-\lambda x} \mathbb{1}_{\mathbb{R}^+}(x).$$

avec $\Gamma(a) = \int_0^{+\infty} x^{a-1} e^{-x} dx$. La loi $\Gamma(1, \lambda)$ correspond à la loi $\mathcal{E}(\lambda)$.

Proposition 74 Si X_1, \dots, X_n sont des v.a. indépendantes de loi $\mathcal{E}(\lambda)$ alors $X_1 + \dots + X_n \sim \Gamma(n, \lambda)$.

La proposition peut se démontrer par récurrence.

8 Convergences

8.1 Quelques inégalités

On considère dans cette partie des v.a. réelles.

8.1.1 Inégalité de Markov

Proposition 75 *Soit X une v.a. positive et $a > 0$. Alors*

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

Preuve.

$$\begin{aligned} \mathbb{E}[X] &= \mathbb{E}[X\mathbb{1}_{X \geq a} + X\mathbb{1}_{X < a}] = \mathbb{E}[X\mathbb{1}_{X \geq a}] + \mathbb{E}[X\mathbb{1}_{X < a}] \\ &\geq \mathbb{E}[a\mathbb{1}_{X \geq a}] + \mathbb{E}[0\mathbb{1}_{X < a}] \geq a\mathbb{E}[\mathbb{1}_{X \geq a}]. \end{aligned}$$

$Y := \mathbb{1}_{X \geq a}$ est une v.a. à valeurs dans $\{0, 1\}$ donc elle suit une loi de Bernoulli. Comme

$$\{Y = 1\} \Leftrightarrow \{X \geq a\}$$

alors $\mathbb{E}[Y] = \mathbb{P}(X \geq a)$ ce qui nous donne le résultat. \square

8.1.2 Inégalité de Bienaymé-Tchebychev

Proposition 76 *Soit X une v.a. réelle de carré intégrable, $a > 0$, alors*

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

Preuve. Il suffit d'appliquer l'inégalité de Markov à la variable aléatoire positive $|X - \mathbb{E}[X]|^2$:

$$\begin{aligned} \mathbb{P}(|X - \mathbb{E}[X]| \geq a) &= \mathbb{P}(|X - \mathbb{E}[X]|^2 \geq a^2) \\ &\leq \frac{\mathbb{E}[|X - \mathbb{E}[X]|^2]}{a^2} = \frac{\text{Var}(X)}{a^2}. \end{aligned}$$

\square

8.1.3 Inégalité de Jensen

Rappelons la définition d'une fonction convexe réelle.

Définition 40 $f : \mathbb{R} \rightarrow \mathbb{R}$ est une fonction convexe si pour tous $(x, y) \in \mathbb{R}^2$, $p \in [0, 1]$ on a

$$f(px + (1 - p)y) \leq pf(x) + (1 - p)f(y).$$

Proposition 77 Soit $f : \mathbb{R} \rightarrow \mathbb{R}$ une fonction C^1 . Alors f est convexe si et seulement si le graphe de f est au dessus de toutes ses tangentes. Également, f est convexe si et seulement si f' est croissante.

Proposition 78 Soit $\phi : \mathbb{R} \rightarrow \mathbb{R}$ une fonction convexe. Alors $\phi(\mathbb{E}[X]) \leq \mathbb{E}[\phi(X)]$ sous réserve d'existence des espérances.

Preuve. Si ϕ est C^1 alors le graphe de ϕ est au dessus de ses tangentes, ainsi on a, pour tout $t \in \mathbb{R}$,

$$\phi(x) \geq \phi(t) + \phi'(t)(x - t), \quad \forall x \in \mathbb{R}.$$

Dans le cas général, on peut montrer que cette inégalité s'étend de la façon suivante : pour tout $t \in \mathbb{R}$, il existe $m(t)$ tel que

$$\phi(x) \geq \phi(t) + m(t)(x - t), \quad \forall x \in \mathbb{R}.$$

Ainsi on a

$$\phi(X) \geq \phi(t) + m(t)(X - t)$$

et

$$\mathbb{E}[\phi(X)] \geq \phi(t) + m(t)(\mathbb{E}[X] - t).$$

En prenant $t = \mathbb{E}[X]$ dans l'inégalité précédente, on prouve le résultat. □

Exemple 7 Pour tout $r \geq 1$,

$$|\mathbb{E}[X]|^r \leq \mathbb{E}[|X|^r].$$

En particulier, si $\mathbb{E}[|X|^r] < +\infty$ alors $\mathbb{E}[|X|^s] < +\infty$ pour tout $1 \leq s \leq r$.

8.1.4 Inégalité de Hölder

Proposition 79 Soient $1 < p < +\infty$ et $1 < q < +\infty$ tels que $1/p + 1/q = 1$ (q est appelé exposant conjugué de p). Alors

$$\mathbb{E}[|XY|] \leq \mathbb{E}[|X|^p]^{1/p} \mathbb{E}[|Y|^q]^{1/q}.$$

Remarquons que le cas particulier $p = 2 = q$ correspond à l'inégalité de Cauchy-Schwarz.

Preuve. Une preuve possible repose sur l'inégalité de Young que nous allons commencer par prouver, et qui dit ceci : pour tous $x, y \in \mathbb{R}^2$, et p, q conjugués, alors

$$|xy| \leq \frac{|x|^p}{p} + \frac{|y|^q}{q}.$$

Cette inégalité est évidente lorsque $x = 0$ ou $y = 0$. Sinon, la concavité de la fonction \ln nous donne

$$\ln |xy| = \frac{1}{p} \ln |x|^p + \frac{1}{q} \ln |y|^q \leq \ln \left(\frac{|x|^p}{p} + \frac{|y|^q}{q} \right)$$

et la croissance de la fonction \ln nous permet de conclure.

Montrons maintenant l'inégalité de Hölder. Si $\mathbb{E}[|X|^p]^{1/p} \mathbb{E}[|Y|^q]^{1/q} = 0$ le résultat est évident. Sinon, on a, d'après l'inégalité de Young :

$$\begin{aligned} \frac{\mathbb{E}[|XY|]}{\mathbb{E}[|X|^p]^{1/p} \mathbb{E}[|Y|^q]^{1/q}} &= \mathbb{E} \left[\left| \frac{X}{\mathbb{E}[|X|^p]^{1/p}} \frac{Y}{\mathbb{E}[|Y|^q]^{1/q}} \right| \right] \leq \mathbb{E} \left[\frac{|X|^p / \mathbb{E}[|X|^p]}{p} + \frac{|Y|^q / \mathbb{E}[|Y|^q]}{q} \right] \\ &\leq \frac{1}{p} + \frac{1}{q} = 1. \end{aligned}$$

□

8.2 Lemme de Borel-Cantelli

Définition 41 Soit $(A_n)_{n \geq 1}$ des évènements aléatoires (de la tribu \mathcal{A}).

•

$$\begin{aligned} \limsup_{n \rightarrow +\infty} A_n &= \bigcap_{n \in \mathbb{N}^*} \bigcup_{p \geq n} A_p \in \mathcal{A} \\ &= \{\omega \in \Omega, \forall n \in \mathbb{N}^*, \exists p \geq n, \omega \in A_p\} \\ &= \{\omega \in \Omega, \omega \text{ appartient à une infinité de } A_n\} \\ &= \text{« une infinité de } A_n \text{ sont réalisés »}. \end{aligned}$$

•

$$\begin{aligned}
\liminf_{n \rightarrow +\infty} A_n &= \bigcup_{n \in \mathbb{N}^*} \bigcap_{p \geq n} A_p \in \mathcal{A} \\
&= \{\omega \in \Omega, \exists n \in \mathbb{N}^*, \forall p \geq n, \omega \in A_p\} \\
&= \{\omega \in \Omega, \omega \text{ appartient à tous les } A_n \text{ à partir d'un certain rang}\} \\
&= \text{« tous les } A_n \text{ sont réalisés sauf un nombre fini »}.
\end{aligned}$$

Remarque 35

- $\liminf_{n \rightarrow +\infty} A_n \subset \limsup_{n \rightarrow +\infty} A_n$
- $(\limsup_{n \rightarrow +\infty} A_n)^c = \liminf_{n \rightarrow +\infty} A_n^c$

Exemple 8

- $A_n = [0, 2 + (-1)^n] : \liminf_{n \rightarrow +\infty} A_n = [0, 1], \limsup_{n \rightarrow +\infty} A_n = [0, 3].$
- $A_n = [0, 2 + (-1)^n + (-1/2)^n] : \liminf_{n \rightarrow +\infty} A_n = [0, 1[, \limsup_{n \rightarrow +\infty} A_n = [0, 3].$

Théorème 6 (Lemme de Borel-Cantelli) Soit $(A_n)_{n \in \mathbb{N}}$ une suite d'évènements aléatoires.

- i) Si $\sum_{n \in \mathbb{N}} \mathbb{P}(A_n) < +\infty$, alors $\mathbb{P}(\limsup_{n \rightarrow +\infty} A_n) = 0$.
- ii) Si les $(A_n)_{n \in \mathbb{N}}$ sont indépendants et si $\sum_{n \in \mathbb{N}} \mathbb{P}(A_n) = +\infty$, alors

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

Preuve.

- i) $A = \limsup_{n \rightarrow +\infty} A_n = \bigcap_{n \geq 0} \bigcup_{p \geq n} A_p \subset \bigcup_{p \geq n} A_p$ pour tout n . Donc

$$\mathbb{P}(\limsup_{n \rightarrow +\infty} A_n) \leq \mathbb{P}(\bigcup_{p \geq n} A_p) \leq \sum_{p \geq n} \mathbb{P}(A_p) \xrightarrow{n \rightarrow +\infty} 0,$$

car $\sum_{p \geq n} \mathbb{P}(A_p)$ est le reste d'une série convergente.

- ii) On a $\mathbb{P}(\limsup_{n \rightarrow +\infty} A_n) = \mathbb{P}(\bigcap_{n \geq 1} (\bigcup_{p \geq n} A_p)) = \lim_{n \rightarrow +\infty} \mathbb{P}(\bigcup_{p \geq n} A_p)$ car $(\bigcup_{p \geq n} A_p)_{n \geq 1}$ est décroissante. Pour conclure, il suffit de montrer que $\mathbb{P}(\bigcup_{p \geq n} A_p) = 1$ pour tout $n \geq 1$. On a

$$\begin{aligned}
\mathbb{P}(\bigcup_{p=n}^N A_p) &= 1 - \mathbb{P}(\bigcap_{p=n}^N A_p^c) = 1 - \prod_{p=n}^N \mathbb{P}(A_p^c) \quad \text{par indépendance} \\
&= 1 - \prod_{p=n}^N (1 - \mathbb{P}(A_p)).
\end{aligned}$$

En utilisant l'inégalité de convexité $1 - x \leq e^{-x}$ pour tout $x \in \mathbb{R}$, on obtient

$$\mathbb{P}(\cup_{p=n}^N A_p) \geq 1 - \prod_{p=n}^N \exp(-\mathbb{P}(A_p)) = 1 - \exp\left(-\sum_{p=n}^N \mathbb{P}(A_p)\right) \xrightarrow{N \rightarrow +\infty} 1$$

car la série diverge. Ainsi, la suite $(\cup_{p=n}^N A_p)_{N \geq n}$ étant croissante, on a

$$\mathbb{P}(\cup_{p \geq n} A_p) = \lim_{N \rightarrow +\infty} \mathbb{P}(\cup_{p=n}^N A_p) \geq 1$$

ce qui permet de conclure.

□

8.3 Définitions des différentes notions de convergences

On considère $(X_n)_{n \in \mathbb{N}}$ une suite de v.a. réelles définies sur un même espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$ et X une v.a. également définie sur cet espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$. Il est également possible d'étendre ce qui suit à des vecteurs aléatoires réels en remplaçant toutes les valeurs absolues par la norme euclidienne. On veut donner un sens mathématique à « X_n converge vers X ». Il s'avère qu'il existe en théorie des probabilité plusieurs notions possibles de convergence. Nous allons les lister ci-dessous.

Définition 42 On dit que $(X_n)_{n \in \mathbb{N}}$ converge presque sûrement vers X si :

$$\mathbb{P}(\{\omega \in \Omega, X_n(\omega) \rightarrow X(\omega)\}) = 1.$$

On note alors $X_n \xrightarrow[n \rightarrow +\infty]{p.s.} X$.

Définition 43 On dit que $(X_n)_{n \in \mathbb{N}}$ converge en probabilité vers X si :

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

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

Remarquons que si $\varepsilon < \varepsilon'$ alors $\mathbb{P}(|X_n - X| \geq \varepsilon') \leq \mathbb{P}(|X_n - X| \geq \varepsilon)$, ce qui implique que si

$$\mathbb{P}(|X_n - X| \geq \varepsilon) \xrightarrow{n \rightarrow +\infty} 0$$

alors

$$\mathbb{P}(|X_n - X| \geq \varepsilon') \xrightarrow{n \rightarrow +\infty} 0$$

pour tous les $\varepsilon' \geq \varepsilon$. Donc pour montrer une convergence en probabilité, on peut se contenter de montrer que

$$\mathbb{P}(|X_n - X| \geq \varepsilon) \xrightarrow{n \rightarrow +\infty} 0$$

uniquement pour les $\varepsilon > 0$ plus petits que n'importe quelle borne $A > 0$ fixée.

Définition 44 On dit que $(X_n)_{n \in \mathbb{N}}$ converge dans L^p ($p \geq 1$) vers X si : $X_n \in L^p$ (i.e. $\mathbb{E}[|X_n|^p] < +\infty$) pour tous $n \in \mathbb{N}$, $X \in L^p$ ($\mathbb{E}[|X|^p] < +\infty$) et

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

On note alors $X_n \xrightarrow[n \rightarrow +\infty]{L^p} X$.

Définition 45 On dit que $(X_n)_{n \in \mathbb{N}}$ converge en loi vers X si : $F_{X_n}(t) \rightarrow F_X(t)$ en tout point de continuité de F_X . On note alors $X_n \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} X$.

Remarque 36

- Si F_X est continue, alors pour tous $a < b \in \mathbb{R}$,

$$\mathbb{P}(X_n \in [a, b]) \xrightarrow{n \rightarrow +\infty} \mathbb{P}(X \in [a, b]).$$

- Pour la convergence en loi, on a pas besoin que les v.a. soient définies sur le même espace de probabilité ! De plus, concernant la limite X , seule sa loi importe. En particulier, on peut par exemple écrire $X_n \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, 1)$ au lieu de $X_n \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} X$ avec $X \sim \mathcal{N}(0, 1)$. Par contre, pour les autres convergences tout ceci est faux.

Concernant la convergence en loi, on peut se poser la question de l'utilité de regarder la convergence de la fonction de répartition uniquement aux points de continuité de la fonction de répartition limite. Il est néanmoins possible de voir sur des exemples simples que cette restriction est tout à fait naturelle. Prenons par exemple la suite de v.a. constantes $X_n = 1/n$. Il semblerait naturelle d'avoir $X_n \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} 0$. Regardons ce qu'il en est de la fonction de répartition :

$$F_{X_n}(t) = \mathbb{P}(X_n \leq t) = \begin{cases} 0 & \text{si } t < 1/n \\ 1 & \text{sinon} \end{cases} \xrightarrow{n \rightarrow +\infty} \begin{cases} 0 & \text{si } t \leq 0 \\ 1 & \text{sinon} \end{cases}.$$

Or

$$F_0(t) = \mathbb{P}(0 \leq t) = \begin{cases} 0 & \text{si } t < 0 \\ 1 & \text{sinon} \end{cases}.$$

Donc on a bien $X_n \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} 0$ même si la fonction de répartition ne converge pas au point de discontinuité 0.

8.4 Autres caractérisations et liens entre les convergences

Proposition 80

$$X_n \xrightarrow[n \rightarrow +\infty]{p.s.} X \Leftrightarrow \forall \varepsilon > 0, \mathbb{P} \left(\limsup_{n \rightarrow +\infty} \{|X_n - X| \geq \varepsilon\} \right) = 0.$$

Remarque 37 Si pour tout $\varepsilon > 0$, on a

$$\sum_{n \in \mathbb{N}} \mathbb{P}(|X_n - X| \geq \varepsilon) < +\infty,$$

alors le lemme de Borel-Cantelli associé à la proposition précédente permet de montrer la convergence presque sûre de $(X_n)_{n \in \mathbb{N}}$ vers X .

Preuve. Soit $\omega \in \Omega$.

$$\begin{aligned} X_n(\omega) \xrightarrow[n \rightarrow +\infty]{} X(\omega) &\Leftrightarrow \forall \varepsilon > 0, \exists N \in \mathbb{N}, \forall n \geq N, |X_n(\omega) - X(\omega)| < \varepsilon \\ &\Leftrightarrow \omega \in \bigcap_{\varepsilon > 0} \bigcup_{N \in \mathbb{N}} \bigcap_{n \geq N} A_n^\varepsilon \end{aligned}$$

avec $A_n^\varepsilon = \{|X_n - X| < \varepsilon\}$. Donc, en passant à la contraposée, on obtient

$$\begin{aligned} X_n(\omega) \not\xrightarrow[n \rightarrow +\infty]{} X(\omega) &\Leftrightarrow \omega \in \bigcup_{\varepsilon > 0} \bigcap_{N \in \mathbb{N}} \bigcup_{n \geq N} (A_n^\varepsilon)^c \\ &\Leftrightarrow \omega \in \bigcup_{\varepsilon > 0} \limsup_{n \rightarrow +\infty} \{|X_n - X| \geq \varepsilon\}. \end{aligned}$$

Ainsi on a

$$\mathbb{P}(\limsup_{n \rightarrow +\infty} \{|X_n - X| \geq \varepsilon\}) \leq \mathbb{P}(X_n(\omega) \not\xrightarrow[n \rightarrow +\infty]{} X(\omega)), \quad \forall \varepsilon > 0,$$

et, comme l'union est décroissante en ε ,

$$\lim_{\varepsilon \rightarrow 0} \mathbb{P}(\limsup_{n \rightarrow +\infty} \{|X_n - X| \geq \varepsilon\}) = \mathbb{P}\left(\bigcup_{\varepsilon > 0} \limsup_{n \rightarrow +\infty} \{|X_n - X| \geq \varepsilon\}\right) = \mathbb{P}(X_n(\omega) \not\rightarrow X(\omega)).$$

Donc si $X_n \xrightarrow[n \rightarrow +\infty]{p.s.} X$ alors $\mathbb{P}(X_n(\omega) \not\rightarrow X(\omega)) = 0$ ce qui implique que

$$\mathbb{P}(\limsup_{n \rightarrow +\infty} \{|X_n - X| \geq \varepsilon\}) = 0, \quad \forall \varepsilon > 0.$$

Inversement, si

$$\mathbb{P}(\limsup_{n \rightarrow +\infty} \{|X_n - X| \geq \varepsilon\}) = 0, \quad \forall \varepsilon > 0,$$

alors

$$\lim_{\varepsilon \rightarrow 0} \mathbb{P}(\limsup_{n \rightarrow +\infty} \{|X_n - X| \geq \varepsilon\}) = 0 = \mathbb{P}(X_n(\omega) \not\rightarrow X(\omega)),$$

ce qui prouve le résultat. \square

La proposition suivante est admise : elle donne des caractérisation équivalente de la convergence en loi.

Proposition 81 *les assertions suivantes sont équivalentes :*

i) $X_n \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} X,$

ii) *pour toute fonction h continue bornée on a $\mathbb{E}[h(X_n)] \xrightarrow[n \rightarrow +\infty]{} \mathbb{E}[h(X)],$*

iii) *pour tout $t \in \mathbb{R}, \varphi_{X_n}(t) \xrightarrow[n \rightarrow +\infty]{} \varphi_X(t).$*

La dernière caractérisation suppose que la limite de φ_{X_n} soit une fonction caractéristique. Il existe un théorème plus fort qui se passe de cette hypothèse.

Théorème 7 (Théorème de Levy) *S'il existe une fonction $\phi : \mathbb{R} \rightarrow \mathbb{C}$ continue en 0 telle que $\varphi_{X_n}(t) \xrightarrow[n \rightarrow +\infty]{} \phi(t)$ pour tout $t \in \mathbb{R}$. Alors X_n converge en loi et ϕ est la fonction caractéristique de la limite.*

Dans le cas des v.a. discrètes il existe encore une autre caractérisation possible de la convergence en loi.

Proposition 82 *Soit $(X_n)_{n \in \mathbb{N}}$ et X des v.a. à valeurs dans \mathbb{N} . On a $X_n \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} X$ si et seulement si*

$$\mathbb{P}(X_n = k) \xrightarrow[n \rightarrow +\infty]{} \mathbb{P}(X = k), \quad \forall k \in \mathbb{N}.$$

Preuve.

- Supposons que $(X_n)_{n \in \mathbb{N}}$ converge en loi vers X . Pour $k \in \mathbb{N}$ on considère une fonction h_k continue telle que $h_k(x) = 0$ pour tous $x \notin]k-1, k+1[$ et $h_k(k) = 1$. On a alors

$$\mathbb{P}(X_n = k) = \mathbb{E}[h_k(X_n)] \xrightarrow{n \rightarrow +\infty} \mathbb{E}[h_k(X)] = \mathbb{P}(X = k).$$

- Inversement, on suppose que

$$\mathbb{P}(X_n = k) \xrightarrow{n \rightarrow +\infty} \mathbb{P}(X = k), \quad \forall k \in \mathbb{N}.$$

Soit $t \in \mathbb{R}$ quelconque.

$$F_{X_n}(t) = \mathbb{P}(X_n \leq t) = \sum_{k=0}^{\lfloor t \rfloor} \mathbb{P}(X_n = k) \xrightarrow{n \rightarrow +\infty} \sum_{k=0}^{\lfloor t \rfloor} \mathbb{P}(X = k) = F_X(t),$$

donc $X_n \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} X$.

□

Il existe de forts liens entre les différentes notions de convergences, comme l'illustre la proposition suivante.

Proposition 83

- i) La convergence L^q implique la convergence L^p lorsque $q > p \geq 1$,*
- ii) La convergence L^p ($p \geq 1$) implique la convergence en probabilité,*
- iii) La convergence presque-sûre implique la convergence en probabilité,*
- iv) La convergence en probabilité implique la convergence en loi.*

Preuve.

- i) En utilisant l'inégalité de Hölder, ou l'inégalité de Jensen, on a

$$\mathbb{E}[|X_n - X|^p] \leq \mathbb{E}[|X_n - X|^{pq/p}]^{p/q} \times 1 = \mathbb{E}[|X_n - X|^q]^{p/q}$$

car $q/p \geq 1$. Ainsi, la membre de gauche de l'inégalité tend vers 0 lorsque celui de droite tend vers 0.

- ii) Soit $\varepsilon > 0$ quelconque. D'après l'inégalité de Markov, on a

$$\mathbb{P}(|X_n - X| \geq \varepsilon) = \mathbb{P}(|X_n - X|^p \geq \varepsilon^p) \leq \frac{\mathbb{E}[|X_n - X|^p]}{\varepsilon^p}.$$

Ainsi, pour tout ε , $\mathbb{P}(|X_n - X| \geq \varepsilon)$ tend vers 0 si $\mathbb{E}[|X_n - X|^p]$ tend vers 0.

iii) Soit $\varepsilon > 0$ quelconque.

$$\mathbb{P}(|X_n - X| \geq \varepsilon) \leq \mathbb{P}\left(\bigcup_{p \geq n} \{|X_p - X| \geq \varepsilon\}\right) \xrightarrow{n \rightarrow +\infty} \mathbb{P}\left(\bigcap_{n \geq 0} \bigcup_{p \geq n} \{|X_p - X| \geq \varepsilon\}\right).$$

Donc si

$$\mathbb{P}(\limsup_{n \rightarrow +\infty} \{|X_n - X| \geq \varepsilon\}) = \mathbb{P}\left(\bigcap_{n \geq 0} \bigcup_{p \geq n} \{|X_p - X| \geq \varepsilon\}\right) \xrightarrow{n \rightarrow +\infty} 0$$

alors $\mathbb{P}(|X_n - X| \geq \varepsilon) \xrightarrow{n \rightarrow +\infty} 0$.

iv) Soit t un point de continuité de F_X .

$$\begin{aligned} F_{X_n}(t) &= \mathbb{P}(X_n \leq t) \\ &= \mathbb{P}\left(\underbrace{\{X_n \leq t\} \cap \{|X_n - X| < \varepsilon\}}_{\subset \{X \leq t + \varepsilon\}}\right) + \mathbb{P}(\{X_n \leq t\} \cap \{|X_n - X| \geq \varepsilon\}) \\ &\leq \mathbb{P}(X \leq t + \varepsilon) + \mathbb{P}(|X_n - X| \geq \varepsilon) \xrightarrow{n \rightarrow +\infty} F_X(t + \varepsilon). \end{aligned}$$

Donc $\limsup_{n \rightarrow +\infty} F_{X_n}(t) \leq F_X(t + \varepsilon)$ et comme le membre de gauche ne dépend par de ε , on peut faire tendre ε vers 0 dans le membre de droite pour obtenir : $\limsup_{n \rightarrow +\infty} F_{X_n}(t) \leq F_X(t)$.

De la même façon,

$$\begin{aligned} 1 - F_{X_n}(t) &= \mathbb{P}(X_n > t) \\ &= \mathbb{P}\left(\underbrace{\{X_n > t\} \cap \{|X_n - X| < \varepsilon\}}_{\subset \{X > t - \varepsilon\}}\right) + \mathbb{P}(\{X_n > t\} \cap \{|X_n - X| \geq \varepsilon\}) \\ &\leq 1 - \mathbb{P}(X \leq t - \varepsilon) + \mathbb{P}(|X_n - X| \geq \varepsilon) \xrightarrow{n \rightarrow +\infty} 1 - F_X(t - \varepsilon). \end{aligned}$$

Donc $\liminf_{n \rightarrow +\infty} F_{X_n}(t) \geq F_X(t - \varepsilon)$ et comme le membre de gauche ne dépend par de ε , on peut faire tendre ε vers 0 dans le membre de droite pour obtenir : $\liminf_{n \rightarrow +\infty} F_{X_n}(t) \geq F_X(t)$ en utilisant la continuité de F_X au point t .

On en conclut que

$$\liminf_{n \rightarrow +\infty} F_{X_n}(t) = \limsup_{n \rightarrow +\infty} F_{X_n}(t) = \lim_{n \rightarrow +\infty} F_{X_n}(t) = F_X(t).$$

□

Il est bon de remarquer que les réciproques sont fausses, nous renvoyons au TD pour des contre-exemples. Néanmoins il existe des réciproques partielles.

Proposition 84

- i) Si X_n converge vers X en probabilité, alors X_n converge vers X presque-sûrement à une sous suite près.
- ii) Si X_n converge en loi vers une constante c alors X_n converge en probabilité vers c .

Preuve.

- i) On pose $n_0 = 0$ et grâce à l'hypothèse de convergence en probabilité, on construit par récurrence une suite strictement croissante $(n_k)_{k \in \mathbb{N}}$ telle que pour tout $k \in \mathbb{N}^*$, on a

$$\mathbb{P}(|X_{n_k} - X| \geq 1/k) \leq \frac{1}{k^2}.$$

Ainsi, pour tout $\varepsilon > 0$ on a

$$\begin{aligned} \sum_{k \in \mathbb{N}} \mathbb{P}(|X_{n_k} - X| \geq \varepsilon) &= \sum_{k < 1/\varepsilon} \mathbb{P}(|X_{n_k} - X| \geq \varepsilon) + \sum_{k \geq 1/\varepsilon} \mathbb{P}(|X_{n_k} - X| \geq \varepsilon) \\ &\leq \sum_{k < 1/\varepsilon} \mathbb{P}(|X_{n_k} - X| \geq \varepsilon) + \sum_{k \geq 1/\varepsilon} \mathbb{P}(|X_{n_k} - X| \geq 1/k) \\ &\leq [1/\varepsilon] + 1 + \sum_{k \geq 1/\varepsilon} \frac{1}{k^2} < +\infty. \end{aligned}$$

Le lemme de Borel-Cantelli permet alors de conclure à la convergence presque-sûre de $(X_{n_k})_{k \in \mathbb{N}}$ vers X .

- ii) Tout d'abord $F_c(t) = \mathbb{1}_{t > c}$ qui est continue sur $\mathbb{R} \setminus \{c\}$. Donc

$$F_{X_n}(t) \xrightarrow{n \rightarrow +\infty} \begin{cases} F_c(t) = 0 & \text{si } t < c \\ F_c(t) = 1 & \text{si } t > c. \end{cases}$$

On a

$$\mathbb{P}(|X_n - c| \geq \varepsilon) = \mathbb{P}(X_n \leq c - \varepsilon) + \mathbb{P}(X_n \geq c + \varepsilon) \xrightarrow{n \rightarrow +\infty} F_c(c - \varepsilon) + (1 - F_c(c + \varepsilon)) = 0,$$

car F_c est continue en $c + \varepsilon$.

□

Proposition 85 *Si X_n converge vers X et Y_n converge vers Y presque-sûrement (resp. en probabilité, resp. dans L^p) alors (X_n, Y_n) converge vers (X, Y) presque-sûrement (resp. en probabilité, resp. dans L^p). Cela implique en particulier que $X_n + Y_n$ converge vers $X + Y$ et $X_n Y_n$ converge vers XY (pour les mêmes types de convergences).*

Attention, la proposition précédente est fautive pour la convergence en loi. Prenons par exemple $X \sim \mathcal{N}(0, 1)$ et posons $X_n = X$. Comme $-X$ à même loi que X , alors $X_n \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} X$ et $X_n \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} -X$. Pourtant $2X_n = 2X$ ne converge pas en loi vers 0!

Proposition 86 *Soit $h : \mathbb{R} \rightarrow \mathbb{R}$ une fonction continue. Alors si X_n converge vers X presque-sûrement ou en probabilité ou en loi, alors $h(X_n)$ converge vers $h(X)$ presque-sûrement ou en probabilité ou en loi.*

Preuve. Exercice. □

8.5 Quelques exemples de convergences

8.5.1 Lois de Bernoulli

On considère $X_n \sim \mathcal{B}(p_n)$ avec $p_n \xrightarrow[n \rightarrow +\infty]{} 0$. Pour tout $0 < \varepsilon < 1$,

$$\mathbb{P}(|X_n - 0| \geq \varepsilon) = \mathbb{P}(X_n = 1) = p_n \xrightarrow[n \rightarrow +\infty]{} 0$$

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

De plus, pour tout $p \geq 1$,

$$\mathbb{E}[|X_n|^p] = \mathbb{E}[X_n] = p_n \xrightarrow[n \rightarrow +\infty]{} 0,$$

donc $X_n \xrightarrow[n \rightarrow +\infty]{L^p} 0$.

Enfin, si on suppose les X_n indépendantes, pour tout $\varepsilon > 0$ on a, d'après le lemme de Borel-Cantelli,

$$\mathbb{P}(\limsup_{n \rightarrow +\infty} |X_n - 0| \geq \varepsilon) = 0 \text{ ssi } \sum_{n \geq 1} \mathbb{P}(|X_n - 0| \geq \varepsilon) < +\infty.$$

Donc $X_n \xrightarrow[n \rightarrow +\infty]{p.s.} 0$ si et seulement si $\sum_{n \geq 1} p_n < +\infty$.

Attention, si on a plus l'indépendance, alors $\sum_{n \geq 1} p_n < +\infty$ implique la convergence presque-sûre mais la réciproque est pas nécessairement vraie. En effet prenons $\Omega = [0, 1]$ muni de la mesure de Lebesgue et $X_n(\omega) = \mathbb{1}_{\omega \in [0, 1/n]}$. Alors $X_n \sim \mathcal{B}(1/n)$, $\sum_{n \geq 1} 1/n = +\infty$ et pourtant $X_n \xrightarrow[n \rightarrow +\infty]{p.s.} 0$!

On pose maintenant $Y_n = nX_n$ avec $X_n \sim \mathcal{B}(1/n)$. Pour tout $0 < \varepsilon < 1$,

$$\mathbb{P}(|Y_n - 0| \geq \varepsilon) = \mathbb{P}(X_n = 1) = 1/n \xrightarrow[n \rightarrow +\infty]{} 0$$

donc $Y_n \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} 0$. Par contre

$$\mathbb{E}[|Y_n|] = n \times \frac{1}{n} = 1,$$

donc Y_n ne converge pas vers 0 dans L^1 (et donc ne converge pas tout court dans L^1 car on rappelle que la convergence dans L^1 vers une limite Y implique la convergence en probabilité vers Y et donc nécessairement on devrait avoir $Y = 0$).

8.5.2 Approximation Binomiale-Poisson

Proposition 87 Soit $X_n \sim \mathcal{B}(n, \lambda/n)$. Alors $X_n \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{P}(\lambda)$.

Preuve. Exercice. Comme on a des v.a. à valeurs dans \mathbb{N} , il suffit de regarder la convergence de $\mathbb{P}(X_n = k)$ pour tous les $k \in \mathbb{N}$. Soit $k \in \mathbb{N}$,

$$\begin{aligned} \mathbb{P}(X_n = k) &= C_n^k \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} \quad (\text{Pour } n \geq k) \\ &= \frac{n!}{k!(n-k)!} \frac{\lambda^k}{n^k} \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-k} \\ &= \frac{\lambda^k}{k!} \underbrace{\frac{n(n-1)\dots(n-k+1)}{nn\dots n}}_{\xrightarrow[n \rightarrow +\infty]{} 1} \underbrace{\left(1 - \frac{\lambda}{n}\right)^n}_{\xrightarrow[n \rightarrow +\infty]{} e^{-\lambda}} \underbrace{\left(1 - \frac{\lambda}{n}\right)^{-k}}_{\xrightarrow[n \rightarrow +\infty]{} 1} \\ &\xrightarrow[n \rightarrow +\infty]{} e^{-\lambda} \frac{\lambda^k}{k!}. \end{aligned}$$

Il est également possible de démontrer le résultat en regardant la convergence de la fonction de répartition. \square

8.5.3 Maximum de lois exponentielles

Soit $(X_{n \in \mathbb{N}})_{n \geq 0}$ des v.a. i.i.d. de loi $\mathbb{E}(\lambda)$. On pose $M_n = \max\{X_1, \dots, X_n\}$ et $Y_n = \lambda M_n - \ln n$. On s'intéresse à la convergence en loi de Y_n et pour cela, on étudie la convergence de la fonction de répartition. Soit $t \in \mathbb{R}$,

$$\begin{aligned} F_{Y_n}(t) &= \mathbb{P}((\lambda M_n - \ln n) \leq t) = \mathbb{P}\left(M_n \leq \frac{t + \ln n}{\lambda}\right) \\ &= \prod_{i=1}^n \mathbb{P}\left(X_i \leq \frac{t + \ln n}{\lambda}\right) \\ &= \left(1 - \exp\left(-\frac{\lambda(t + \ln n)}{\lambda}\right)\right)^n \quad \text{car } \frac{t + \ln n}{\lambda} \geq 0 \text{ pour } n \text{ assez grand} \\ &= \left(1 - \frac{e^{-t}}{n}\right)^n = \exp\left(n \ln\left(1 - \frac{e^{-t}}{n}\right)\right) \xrightarrow{n \rightarrow +\infty} e^{-e^{-t}}. \end{aligned}$$

Comme $t \mapsto e^{-e^{-t}}$ est une fonction de répartition (croissante, continue, tendant vers 0 en $-\infty$ et 1 en $+\infty$), Y converge en loi vers Y de fonction de répartition $F_Y(t) = e^{-e^{-t}}$. Remarquons que cette fonction de répartition est C^1 , donc Y est une v.a. à densité.

9 Théorèmes limites en probabilité

9.1 Lois des grands nombres

Soit $(X_n)_{n \in \mathbb{N}}$ une suite de v.a. i.i.d. On pose $S_n = X_1 + \dots + X_n$ et on s'intéresse à la moyenne empirique.

Donnons un exemple concret pour motiver cette étude. Si on lance plusieurs fois une pièce équilibrée, les lancers étant indépendants, on peut noter X_n le résultat du n -ième lancer ($X_n = 1$ si on obtient « pile » et $X_n = 0$ si on obtient « face »). Alors S_n/n représente la proportion de pile après n lancers. Empiriquement, on sait que cette proportion tend vers $1/2$ quand le nombre de lancers tend vers l'infini. Nous allons donc donner un sens mathématique à cette constatation empirique.

Proposition 88 (Loi faible des grands nombres) *Soit $(X_n)_{n \in \mathbb{N}}$ une suite de v.a. i.i.d. de même loi que X . On suppose que $\mathbb{E}[|X|^2] < +\infty$. Alors*

$$\frac{S_n}{n} \xrightarrow[n \rightarrow +\infty]{L^2} \mathbb{E}[X] \quad \text{et donc} \quad \frac{S_n}{n} \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} \mathbb{E}[X].$$

Preuve. $\mathbb{E}[S_n/n] = \mathbb{E}[X]$ et donc

$$\mathbb{E} \left[\left| \frac{S_n}{n} - \mathbb{E}[X] \right|^2 \right] = \text{Var} \left(\frac{S_n}{n} \right) = \frac{1}{n} \text{Var}(X) \xrightarrow[n \rightarrow +\infty]{} 0.$$

□

Théorème 8 (Loi forte des grands nombres) *Soit $(X_n)_{n \in \mathbb{N}}$ une suite de v.a. i.i.d. de même loi que X . On suppose que $\mathbb{E}[|X|] < +\infty$. Alors*

$$\frac{S_n}{n} \xrightarrow[n \rightarrow +\infty]{p.s.} \mathbb{E}[X].$$

Preuve. Nous allons prouver le résultat uniquement avec l'hypothèse supplémentaire $\mathbb{E}[|X|^4] < +\infty$, le cas général étant admis.

On pose $m_1 = \mathbb{E}[X]$, $m_2 = \mathbb{E}[X^2]$ et $m_4 = \mathbb{E}[X^4]$. On peut supposer que $m_1 = 0$ car il suffit de remplacer X_i par $Y_i := X_i - m_1$ qui est bien d'espérance nulle. Montrer que $\frac{S_n}{n} \xrightarrow[n \rightarrow +\infty]{p.s.} 0$ est équivalent à montrer que pour tout $\varepsilon > 0$, $\mathbb{P}(\limsup_{n \rightarrow +\infty} \{|S_n/n| > \varepsilon\}) = 0$. D'après le lemme de Borel-Cantelli, il suffit de montrer que

$$\sum_{n \geq 1} \mathbb{P}(|S_n/n| > \varepsilon) < +\infty.$$

Soit $\varepsilon > 0$ fixé. D'après l'inégalité de Markov on a

$$\mathbb{P}(|S_n/n| > \varepsilon) = \mathbb{P}(|S_n|^2 > \varepsilon^4 n^4) \leq \frac{\mathbb{E}[S_n^4]}{\varepsilon^4 n^4},$$

et par linéarité de l'espérance,

$$\mathbb{E}[S_n^4] = \mathbb{E}[(X_1 + \dots + X_n)^4] = \sum_{1 \leq i, j, k, l \leq n} \mathbb{E}[X_i X_j X_k X_l].$$

Lorsque $i \neq j \neq k \neq l$, l'indépendance nous donne

$$\mathbb{E}[X_i X_j X_k X_l] = \mathbb{E}[X_i] \mathbb{E}[X_j] \mathbb{E}[X_k] \mathbb{E}[X_l] = 0.$$

De la même façon, dès qu'un indice est seul, on a $\mathbb{E}[X_i X_j X_k X_l] = 0$. Les seuls termes non nuls sont :

- $\mathbb{E}[X_i^4] = m_4$, il y en a n .
- $\mathbb{E}[X_i^2 X_j^2] = \mathbb{E}[X_i^2] \mathbb{E}[X_j^2] = m_2^2$, pour $1 \leq i \neq j \leq n$. Il y en a $C_n^2 C_4^2 = 3n(n-1)$ (choix de i et j , puis place des i).

Donc $\mathbb{E}[(S_n)^4] \leq Cn^2$ et $\mathbb{P}(|S_n/n| > \varepsilon) \leq \frac{C}{\varepsilon^2 n^2}$ qui est le terme général d'une série convergente. \square

Reprenons l'exemple de départ : si on lance n fois une pièce, la proportion de piles tend vers p la probabilité d'obtenir pile avec cette pièce.

9.2 Théorème centrale limite

On sait maintenant que, sous réserve d'existence de $\mathbb{E}[X]$, $\frac{S_n}{n} \xrightarrow[n \rightarrow +\infty]{p.s.} \mathbb{E}[X]$ et donc $\frac{S_n}{n} - \mathbb{E}[X] \xrightarrow[n \rightarrow +\infty]{p.s.} 0$. Afin de quantifier la vitesse de cette convergence, on cherche α tel que $n^\alpha \left(\frac{S_n}{n} - \mathbb{E}[X] \right)$ tend vers quelque chose de non trivial.

Théorème 9 (théorème de la limite centrale ou théorème central limite) *Soit $(X_n)_{n \in \mathbb{N}}$ une suite de v.a. i.i.d. de même loi que X . On suppose que $\mathbb{E}[|X|^2] < +\infty$. On note $\mathbb{E}[X] = m$ et $\text{Var}(X) = \sigma^2$ (supposée non nulle). Alors*

$$\frac{\sqrt{n}}{\sigma} \left(\frac{S_n}{n} - m \right) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, 1).$$

Remarquons que si on pose $Y_n = \frac{\sqrt{n}}{\sigma} \left(\frac{S_n}{n} - m \right)$ alors $\mathbb{E}[Y_n] = 0$ et $\text{Var}(Y_n) = 1$.

Preuve. Comme pour la preuve de la loi forte des grands nombres, on peut supposer que $m = 0$, quitte à remplacer X_i par $X_i - m$. On va regarder la convergence de la fonction caractéristique de Y_n . Soit $t \in \mathbb{R}$ fixé

$$\begin{aligned}\phi_{Y_n}(t) &= \mathbb{E}[e^{itY_n}] = \mathbb{E}[e^{it\frac{S_n}{\sigma\sqrt{n}}}] = \mathbb{E}[e^{it\frac{X_1+\dots+X_n}{\sigma\sqrt{n}}}] \\ &= \mathbb{E}\left[e^{it\frac{X_1}{\sigma\sqrt{n}}}\right]^n = \left[\phi_{X_1}\left(\frac{t}{\sigma\sqrt{n}}\right)\right]^n.\end{aligned}$$

On fait un développement limité de ϕ_{X_1} en 0 : $\mathbb{E}[X_1^2] < +\infty$ donc ϕ_{X_1} est C^2 et

$$\begin{aligned}\phi_{X_1}(h) &= \phi_{X_1}(0) + h\phi'_{X_1}(0) + \frac{h^2}{2}\phi''_{X_1}(0) + o(h^2) \\ &= 1 + hi\mathbb{E}[X_1] - \frac{h^2}{2}\mathbb{E}[X_1^2] + o(h^2).\end{aligned}$$

Ainsi,

$$\phi_{X_1}\left(\frac{t}{\sigma\sqrt{n}}\right) = 1 - \frac{t^2\sigma^2}{2\sigma^2n} + o(1/n)$$

et

$$\phi_{Y_n}(t) = \left(1 - \frac{t^2}{2n} + o(1/n)\right)^n = \left(1 - \frac{t^2}{2n}\right)^n (1 + o(1/n))^n.$$

De manière classique

$$\left(1 - \frac{t^2}{2n}\right)^n = \exp\left(n \ln\left(1 - \frac{t^2}{2n}\right)\right) \xrightarrow{n \rightarrow +\infty} e^{-t^2/2}.$$

On ne peut pas faire la même chose pour le terme $(1 + o(1/n))^n$ car $o(1/n)$ est complexe ! Néanmoins on a

$$\begin{aligned}|(1 + o(1/n))^n - 1| &= \left|\sum_{k=1}^n C_n^k (o(1/n))^k\right| \leq \sum_{k=1}^n C_n^k |o(1/n)|^k = (1 + |o(1/n)|)^n - 1 \\ &= e^{n \ln(1 + |o(1/n)|)} - 1 \xrightarrow{n \rightarrow +\infty} 0.\end{aligned}$$

Donc $\phi_{Y_n}(t) \xrightarrow{n \rightarrow +\infty} e^{-t^2/2}$ qui est la fonction caractéristique de la loi $\mathcal{N}(0, 1)$.

□

10 Vecteurs gaussiens

Le but de ce dernier chapitre est de généraliser la loi gaussienne au cadre multidimensionnel.

10.1 Introduction (loi normale réelle)

Définition 46 Soient $m \in \mathbb{R}$ et $\sigma^2 > 0$. $X \sim \mathcal{N}(m, \sigma^2)$ si X est une v.a. à densité, de densité

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right), \quad x \in \mathbb{R}.$$

On définit également $X \sim \mathcal{N}(m, 0)$ si $X = m$ p.s.

On remarque que la loi $\mathcal{N}(m, \sigma^2)$ est à densité si $\sigma > 0$ et est discrète si $\sigma = 0$.

Proposition 89 Si $X \sim \mathcal{N}(m, \sigma^2)$ avec $\sigma > 0$ alors

$$\frac{X - m}{\sigma} \sim \mathcal{N}(0, 1).$$

Preuve. Il suffit de calculer la fonction de répartition de la v.a. $\frac{X-m}{\sigma}$. c.f. la partie 4.5.3. \square

Une conséquence de la proposition précédente est que tout calcul sur la loi $\mathcal{N}(m, \sigma^2)$, lorsque $\sigma > 0$, se ramène à un calcul sur la loi $\mathcal{N}(0, 1)$.

Proposition 90 Soit $X \sim \mathcal{N}(m, \sigma^2)$ de densité f_{m, σ^2} lorsque $\sigma > 0$. Alors on a

- $\int_{\mathbb{R}} f_{m, \sigma^2}(x) dx = 1$.
- $\mathbb{E}[X] = m$, $\text{Var}(X) = \sigma^2$.

Preuve. En appliquant un changement de variable de variable $u = (x - m)/\sigma$, on a

$$\int_{\mathbb{R}} f_{m, \sigma^2}(x) dx = \int_{\mathbb{R}} f_{0,1}(u) du = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-u^2/2} du := \frac{1}{\sqrt{2\pi}} I$$

avec $I = \int_{\mathbb{R}} e^{-u^2/2} du$. En appliquant le théorème de Fubini (pour les fonctions positives) ainsi qu'un changement de variable en coordonnées polaires, on obtient

$$\begin{aligned} I^2 &= \iint_{\mathbb{R}^2} \exp\left(-\frac{x^2 + y^2}{2}\right) dx dy = \int_{r>0} \int_{\theta=0}^{2\pi} \exp\left(-\frac{r^2}{2}\right) r dr d\theta = 2\pi \left[-e^{-r^2/2}\right]_0^{+\infty} \\ &= 2\pi \end{aligned}$$

ce qui montre le premier point.

Pour le second point, le résultat est immédiat lorsque $\sigma = 0$, sinon, on pose $Y = (X - m)/\sigma \sim \mathcal{N}(0, 1)$. Il suffit donc d'obtenir le résultat pour Y (i.e. pour la loi $\mathcal{N}(0, 1)$) puisque l'on a alors

$$\mathbb{E}[X] = \mathbb{E}[m + \sigma Y] = m, \quad \text{Var}(Y) = \text{Var}(m + \sigma X) = \sigma^2.$$

Remarquons que pour tout $r \geq 0$, on a $x^r e^{-x^2/2} = o(1/x^2)$ en $+\infty$. Donc d'après le critère de Riemann, on a

$$\mathbb{E}[|Y|^r] = 2 \int_0^{+\infty} x^r e^{-x^2/2} dx < +\infty.$$

En particulier, Y possède une espérance et une variance (finies).

$$\mathbb{E}[Y] = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} x e^{-x^2/2} dx = 0$$

car la fonction intégrée est intégrable et impaire. De plus, une intégration par parties nous donne

$$\begin{aligned} \mathbb{E}[X^2] &= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} x^2 e^{-x^2/2} dx = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} x \cdot x e^{-x^2/2} dx \\ &= \frac{1}{\sqrt{2\pi}} \left[x(-e^{-x^2/2}) \right]_{-\infty}^{+\infty} + \int_{\mathbb{R}} \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx = 0 + 1. \end{aligned}$$

□

Proposition 91 Soit $X \sim \mathcal{N}(m, \sigma^2)$. Alors

$$\phi_X(t) = \exp\left(itm - \frac{\sigma^2 t^2}{2}\right).$$

Preuve. Si $\sigma = 0$ le résultat est évident. Sinon, encore une fois on pose $Y = (X - m)/\sigma \sim \mathcal{N}(0, 1)$ et on commence par prouver le résultat pour Y (i.e. pour la loi $\mathcal{N}(0, 1)$). Comme Y est intégrable, ϕ_Y est dérivable et on a

$$\phi'_Y(t) = \mathbb{E}[iX e^{itX}] = \int_{\mathbb{R}} i x e^{itx} \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx \quad \forall t \in \mathbb{R}.$$

Une intégration par parties nous donne

$$\phi'_Y(t) = \left[i e^{itx} \left(-\frac{e^{-x^2/2}}{\sqrt{2\pi}} \right) \right]_{-\infty}^{+\infty} + \int_{\mathbb{R}} i i t e^{itx} \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx = -t \phi_Y(t).$$

Φ_Y est solution d'une équation différentielle linéaire du premier ordre, sa résolution nous permet d'obtenir

$$\phi_Y(t) = Ce^{-t^2/2} \quad \forall t \in \mathbb{R}$$

et nécessairement $C = 1$ car $\phi_Y(0) = 1$. Revenons maintenant à la v.a. X : pour tout $t \in \mathbb{R}$,

$$\phi_X(t) = \mathbb{E}[e^{it(m+\sigma Y)}] = e^{itm} \mathbb{E}[e^{it\sigma Y}] = e^{itm} \phi_Y(\sigma t) = e^{itm - \sigma^2 t^2 / 2}.$$

□

Proposition 92 Si $X \sim \mathcal{N}(m, \sigma^2)$ et $Y \sim \mathcal{N}(m', \sigma'^2)$ sont indépendantes, alors

$$X + Y \sim \mathcal{N}(m + m', \sigma^2 + \sigma'^2).$$

Preuve. Il suffit de remarquer que

$$\phi_{X+Y}(t) = \phi_X(t)\phi_Y(t), \quad \forall t \in \mathbb{R},$$

et la fonction caractéristique obtenue est celle de la loi $\mathcal{N}(m + m', \sigma^2 + \sigma'^2)$. □

10.2 Vecteurs gaussiens

Définition 47 Un vecteur aléatoire X de \mathbb{R}^n est dit gaussien si toute combinaison linéaire de ses composantes est une v.a. réelle gaussienne : pour tout $(\lambda_1, \dots, \lambda_n) \in \mathbb{R}^n$, $\sum_{i=1}^n \lambda_i X_i$ est une v.a. réelle gaussienne.

Remarque 38 Si on prend $(\lambda_1, \dots, \lambda_n) = (0, \dots, 0)$ alors $\sum_{i=1}^n \lambda_i X_i = 0$ ce qui explique pourquoi on a élargi la notion de v.a. gaussiennes en incluant les constantes.

Corollaire 1 Si $X = (X_1, \dots, X_n)$ est un vecteur gaussien alors chaque composante X_i est une v.a. gaussienne.

Attention, la réciproque du Corollaire est fautive ! Si on considère des v.a. X_1, \dots, X_n qui sont des gaussiennes (réelles) alors (X_1, \dots, X_n) n'est pas nécessairement un vecteur gaussien. Considérons le contre-exemple suivant. Soient $X \sim \mathcal{N}(0, 1)$ et $\varepsilon \sim \mathcal{R}(1/2)$ (i.e. $\mathbb{P}(\varepsilon = 1) = \mathbb{P}(\varepsilon = -1) = 1/2$) indépendantes. On pose $Y = \varepsilon X$. On peut montrer¹ que Y et X ont même fonction de répartition et donc même loi. Ainsi, $Y \sim \mathcal{N}(0, 1)$. Par contre

$$\mathbb{P}(X + Y = 0) = \mathbb{P}((1 + \varepsilon)X = 0) = \mathbb{P}(\{X = 0\} \cup \{\varepsilon = -1\}) = \mathbb{P}(\varepsilon = -1) = 1/2$$

1. Laissé en exercice au lecteur sérieux.

car

$$\mathbb{P}(\{X = 0\} \cup \{\varepsilon = -1\}) = \mathbb{P}(X = 0) + \mathbb{P}(\varepsilon = -1) - \mathbb{P}(X = 0)\mathbb{P}(\varepsilon = -1)$$

et $\mathbb{P}(X = 0) = 0$ (X est à densité). donc $X + Y$ n'est pas une v.a. à densité et n'est pas une v.a. constante : $X + Y$ n'est pas une v.a. gaussienne.

Par contre, si on ajoute une hypothèse d'indépendance, la réciproque du corollaire devient valable.

Proposition 93 *Si X_1, \dots, X_n sont des v.a. gaussiennes indépendantes, alors (X_1, \dots, X_n) est un vecteur gaussien.*

Preuve. On note $X_i \sim \mathcal{N}(m_i, \sigma_i^2)$ pour $1 \leq i \leq n$. On considère $(\lambda_1, \dots, \lambda_n) \in \mathbb{R}^n$. On a alors $\lambda_i X_i \sim \mathcal{N}(\lambda_i m_i, \lambda_i^2 \sigma_i^2)$ pour $1 \leq i \leq n$ et la Proposition 92 nous donne

$$\sum_{i=1}^n \lambda_i X_i \sim \mathcal{N}\left(\sum_{i=1}^n \lambda_i m_i, \sum_{i=1}^n \lambda_i^2 \sigma_i^2\right).$$

□

10.3 Complément sur les vecteurs aléatoires réels

Dans cette partie, nous considérons des vecteurs aléatoires réels non nécessairement gaussiens. On va généraliser les notions d'espérance, de variance et de fonction caractéristique au cadre vectoriel. Soit $X = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}$ un vecteur aléatoire de \mathbb{R}^n .

Espérance. L'espérance de X est définie comme

$$\mathbb{E}[X] = \begin{pmatrix} \mathbb{E}[X_1] \\ \vdots \\ \mathbb{E}[X_n] \end{pmatrix}.$$

Matrice de covariance. Dans le cadre vectoriel, la variance devient une matrice, appelée matrice de covariance et donnée par

$$\begin{aligned} \Gamma &= \mathbb{E}[(X - \mathbb{E}[X])(X - \mathbb{E}[X])^\top] = \mathbb{E}\left[\left((X_i - \mathbb{E}[X_i])(X_j - \mathbb{E}[X_j])\right)_{1 \leq i, j \leq n}\right] \\ &= (\text{Cov}(X_i, X_j))_{1 \leq i, j \leq n}. \end{aligned}$$

En utilisant la linéarité de l'espérance, on obtient une autre formulation possible de la matrice de covariance :

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

Proposition 94 *Toute matrice de covariance est symétrique semi-définie positive.*

Preuve. La symétrie est claire. On note $m = \mathbb{E}[X]$. Soit $u \in \mathbb{R}^n$ quelconque : par linéarité de l'espérance on a

$$u^\top \Gamma u = \mathbb{E}[u^\top (X - m)(X - m)^\top u] = \mathbb{E}[(X - m)^\top u]^2 \geq 0,$$

d'où la positivité. Notons qu'elle n'est pas nécessairement définie positive car, par exemple, $\Gamma = 0$ si $X = \mathbb{E}[X]$. \square

Fonction caractéristique. La fonction caractéristique d'un vecteur aléatoire de \mathbb{R}^n est donnée par

$$\phi_X : \begin{cases} \mathbb{R}^n & \rightarrow \mathbb{C} \\ t = \begin{pmatrix} t_1 \\ \vdots \\ t_n \end{pmatrix} & \mapsto \mathbb{E}[e^{i\langle t, X \rangle}] = \mathbb{E}[e^{i\sum_{j=1}^n t_j X_j}]. \end{cases}$$

Proposition 95 *La fonction caractéristique d'un vecteur aléatoire caractérise sa loi.*

Comme dans le cas réel, cette proposition est admise.

10.4 Transformation affine d'un vecteur gaussien

Proposition 96 *Si X est un vecteur gaussien, alors $Y = AX + B$ avec $A \in \mathcal{M}_{p \times n}(\mathbb{R})$, $B \in \mathbb{R}^p$, est un vecteur gaussien. On a $\mathbb{E}[Y] = A\mathbb{E}[X] + B$ et la matrice de covariance de Y , notée Γ_Y est donnée par $\Gamma_Y = A\Gamma_X A^\top$.*

Preuve. Toute combinaison linéaire des composantes de Y , $\sum_{i=1}^n \lambda_i Y_i$, est une combinaison affine des composantes de X , $\sum_{i=1}^n \lambda_i \left(\sum_{j=1}^n a_{ij} X_j + b_i \right)$, donc est une variable aléatoire gaussienne.

Par linéarité de l'espérance, on a $\mathbb{E}[Y] = A\mathbb{E}[X] + B$ et

$$\begin{aligned} \Gamma_Y &= \mathbb{E}[(Y - \mathbb{E}[Y])(Y - \mathbb{E}[Y])^\top] = \mathbb{E}[A(X - \mathbb{E}[X])(X - \mathbb{E}[X])^\top A^\top] \\ &= A\mathbb{E}[(X - \mathbb{E}[X])(X - \mathbb{E}[X])^\top]A^\top = A\Gamma_X A^\top. \end{aligned}$$

\square

Théorème 10 Soit X un vecteur gaussien d'espérance m et de matrice de covariance Γ . Alors

$$\phi_X(t) = \mathbb{E}[e^{i\langle t, X \rangle}] = \exp\left(i\langle t, m \rangle - \frac{1}{2}t^\top \Gamma t\right) \quad t \in \mathbb{R}^n.$$

Corollaire 2 La loi d'un vecteur gaussien X est entièrement caractérisée par son espérance m et sa matrice de covariance Γ . On note $X \sim \mathcal{N}(m, \Gamma)$, notation cohérente avec la dimension 1.

Preuve. Soit $t \in \mathbb{R}^n$ quelconque. On a

$$\phi_X(t) = \mathbb{E}[e^{i\langle t, X \rangle}] = \mathbb{E}[e^{i1Y}] = \phi_Y(1)$$

en posant $Y = \langle t, X \rangle$. Remarquons que $\langle t, X \rangle = AX + 0$ avec $A = t^\top$. Donc d'après la proposition 96 Y est un vecteur gaussien, donc une v.a. gaussienne car de dimension 1, d'espérance $Am = \langle t, m \rangle$ et de variance $A\Gamma A^\top = t^\top \Gamma t$. Donc

$$\phi_Y(u) = \exp\left(iu\langle t, m \rangle - \frac{u^2}{2}t^\top \Gamma t\right).$$

Il suffit alors de prendre $u = 1$ pour conclure. □

10.5 Indépendance

Contrairement au cadre général, l'indépendance est équivalente à la non-corrélation pour les vecteurs gaussiens.

Proposition 97 Soit $X = (X_1, \dots, X_n)$ un vecteur gaussien. On a équivalence entre

- i) X_1, \dots, X_n sont indépendantes,
- ii) X_1, \dots, X_n sont deux à deux indépendantes,
- iii) X_1, \dots, X_n sont non corrélées,
- iv) La matrice de covariance Γ de X est diagonale.

Preuve. Pour n'importe quel vecteur aléatoire on a $i) \Rightarrow ii) \Rightarrow iii) \Rightarrow iv)$. Montrons donc que $iv) \Rightarrow i)$. Soit $t \in \mathbb{R}^n$ quelconque,

$$\begin{aligned} \phi_X(t) &= \mathbb{E}[e^{i\langle t, X \rangle}] = \exp\left(i\langle t, m \rangle - \frac{1}{2}t^\top \Gamma t\right) \\ &= \exp\left(\sum_{k=1}^n \left(it_k m_k - \frac{1}{2}\Gamma_{kk} t_k^2\right)\right) = \prod_{k=1}^n \exp\left(it_k m_k - \frac{1}{2}\Gamma_{kk} t_k^2\right) \\ &= \prod_{k=1}^n \phi_{X_k}(t_k). \end{aligned}$$

Or $\prod_{k=1}^n \phi_{X_k}(t_k)$ correspond à la fonction caractéristique d'un vecteur aléatoire dont les composantes sont indépendantes et de lois gaussiennes $\mathcal{N}(m_k, \Gamma_{kk})$. Comme la fonction caractéristique caractérise la loi, X_1, \dots, X_k sont indépendantes. \square

Attention, la non corrélation de deux v.a. gaussienne n'implique pas l'indépendance lorsque l'on a pas un vecteur gaussien. Reprenons l'exemple précédent : $X \sim \mathcal{N}(0, 1)$ et $\varepsilon \sim \mathcal{R}(1/2)$ indépendantes. On pose $Y = \varepsilon X : Y \sim \mathcal{N}(0, 1)$.

On a $\mathbb{E}[X] = \mathbb{E}[Y] = 0$, $\mathbb{E}[XY] = \mathbb{E}[\varepsilon]\mathbb{E}[X^2] = 0$ et donc $\text{Cov}(X, Y) = 0$. Néanmoins X et Y ne sont pas indépendantes : par exemple on a

$$\mathbb{P}(X \in [0, 1], Y \in [2, 3]) = 0 \neq \mathbb{P}(X \in [0, 1])\mathbb{P}(Y \in [2, 3]) > 0.$$

10.6 Densité

Proposition 98 Si $Z \sim \mathcal{N}(0, I_n)$ alors Z est un vecteur à densité, de densité

$$f_Z(x) = \frac{1}{(2\pi)^{n/2}} \exp\left(-\frac{1}{2}\|x\|^2\right) \quad x \in \mathbb{R}^n.$$

Preuve. $Z \sim \mathcal{N}(0, I_n)$ donc Z_1, \dots, Z_n sont indépendantes et de même loi $\mathcal{N}(0, 1)$. Ainsi Z est un vecteur à densité et sa densité vaut

$$f_Z(x_1, \dots, x_n) = \prod_{i=1}^n f_{Z_i}(x_i) = \frac{1}{(2\pi)^{n/2}} \exp\left(-\frac{1}{2}(x_1^2 + \dots + x_n^2)\right).$$

\square

Dans le cas général, il est naturel de se poser la question suivante : est-ce qu'un vecteur gaussien est à densité et, si oui, que vaut cette densité? En dimension 1 on s'est ramené au cas centré réduit en posant $Y = \frac{X-m}{\sigma}$. Dans le cas vectoriel, on souhaiterait faire la même chose mais pour cela il faut donner un sens à la racine carrée de Γ (i.e. σ en dimension 1).

Proposition 99 Soit Γ une matrice de covariance. Il existe une matrice A de même taille telle que $\Gamma = AA^\top$.

Preuve. Γ est symétrique semi-définie positive donc il existe une matrice orthogonale P et une matrice diagonale

$$D = \begin{pmatrix} d_1 & & 0 \\ & \ddots & \\ 0 & & d_n \end{pmatrix}$$

telle que les coefficients diagonaux d_1, \dots, d_n sont positifs et $\Gamma = PDP^\top$. On pose

$$D^{1/2} = \begin{pmatrix} \sqrt{d_1} & & 0 \\ & \ddots & \\ 0 & & \sqrt{d_n} \end{pmatrix}$$

et $A = PD^{1/2}P^\top$ de telle sorte que $\Gamma = AA^\top$. □

Proposition 100 *Soit $X \sim \mathcal{N}(m, \Gamma)$, $Z \sim \mathcal{N}(0, I_n)$ et A une matrice de même taille que Γ et telle que $AA^\top = \Gamma$. Alors $AZ + m$ a même loi que X .*

Preuve. D'après la proposition 96, $AZ + m$ est un vecteur gaussien de moyenne $A0 + m = m$ et de matrice de covariance $AI_nA^\top = \Gamma$. □

Théorème 11 *Soit $X \sim \mathcal{N}(m, \Gamma)$. Alors X est un vecteur à densité si et seulement si Γ est inversible. Dans ce cas, sa densité f s'écrit*

$$f(x) = \frac{1}{(2\pi)^{n/2} \sqrt{\det \Gamma}} \exp\left(-\frac{1}{2}(x - m)^\top \Gamma^{-1}(x - m)\right), \quad x \in \mathbb{R}^n.$$

Preuve. D'après la proposition précédente, il suffit d'étudier le vecteur aléatoire $Y = AZ + m$ avec $AA^\top = \Gamma$ et $Z \sim \mathcal{N}(0, I_n)$. Z est à densité (sur \mathbb{R}^n) et Y est une transformation affine de Z . En particulier, Y prend ses valeurs dans le sous espace affine $E = \{Az + m, z \in \mathbb{R}^n\}$ de \mathbb{R}^n qui est de dimension $\text{rg}(A)$. On remarque que $\text{rg}(A) = \text{rg}(\Gamma)$.

Si Γ n'est pas inversible, A n'est pas inversible et f n'admet pas de densité.

Si Γ est inversible, A est inversible. On applique la méthode de la fonction muette : soit $h : \mathbb{R}^n \rightarrow \mathbb{R}$ continue bornée quelconque.

$$\mathbb{E}[h(X)] = \mathbb{E}[h(AZ + m)] = \int_{\mathbb{R}^n} h(az + m) \frac{1}{(2\pi)^{n/2}} \exp\left(-\frac{z^\top z}{2}\right) dz_1 \dots dz_n.$$

On fait le changement de variable suivant

$$\phi : \begin{cases} \mathbb{R}^n & \rightarrow \mathbb{R}^n \\ z & \mapsto Az + m \end{cases},$$

$$\phi^{-1} : \begin{cases} \mathbb{R}^n & \rightarrow \mathbb{R}^n \\ y & \mapsto A^{-1}(y - m) \end{cases}.$$

ϕ et ϕ^{-1} sont C^1 , donc ϕ est un C^1 difféomorphisme. De plus

$$\text{Jac}(\phi^{-1}) = \det(A^{-1}) = (\det(\Gamma^{-1}))^{1/2} = (\det(\Gamma))^{-n/2}$$

et $z^\top z = (y - m)^\top (A^{-1})^\top A^{-1} (y - m) = (y - m)^\top (AA^\top)^{-1} (y - m) = (y - m)^\top \Gamma^{-1} (y - m)$. Ainsi

$$\mathbb{E}[h(X)] = \int_{\mathbb{R}^n} h(y) \frac{1}{(2\pi)^{n/2}} \exp\left(-\frac{(y - m)^\top \Gamma^{-1} (y - m)}{2}\right) \frac{1}{(\det \Gamma)^{1/2}} dy_1 \dots dy_n,$$

ce qui permet de conclure. □