Cours de Licence 3: Probabilités

Cours de: Marc Arnaudon
Rédigé par Hugo Clouet*

Année universitaire 2023 - 2024
Version: 9 novembre 2025

Table des matières

*email: hugo.clouet@etu.u-bordeaux.fr

Résumé

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.
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.

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 1.1.0.1. 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 1.1.0.2. (1) L'expérience $\mathcal{E}$ consiste à lancer un dé, on a alors $\Omega = \{1, \ldots, 6\}$ qui est fini.
(2) L'expérience $\mathcal{E}$ consiste à jouer à pile ou face jusqu'à l'obtention d'un pile, on a alors $\Omega = \{P, FP, FFP, FFFP, \ldots\}$ qui est infini dénombrable.
(3) L'expérience $\mathcal{E}$ consiste à évaluer la durée de vie d'une étoile dans notre galaxie, on a alors $\Omega =]0, +\infty[$ qui est infini non dénombrable.

Un élément ω de Ω sera appelé *réalisation* de l'expérience ou *évènement élémentaire*. Un sous-ensemble A de Ω sera appelé un *évènement*.

Exemple 1.1.0.3. Dans (1), si on pose l'évènement A : « on obtient un nombre pair », on a alors $A = \{2, 4, 6\}$. Dans (2), si on pose l'évènement B : « obtient pile en moins de trois lancer », on a alors $B = \{P, FP, FFP\}$.

L'ensemble des évènements sera soit $\mathcal{P}(\Omega)$ l'ensemble des parties de Ω, soit un sous-ensemble $\mathcal{A}$ de $\mathcal{P}(\Omega)$ appelé *tribus* (que l'on définira un peu après).

1.2 Rappels sur les ensembles

On associe les évènements à des sous-ensembles de Ω. Soit A, B des évènements :
- $A \cup B$: A *ou* B est réalisé.
- $A \cap B$: A *et* B sont réalisés.
- $A^{\complement}$ (ou $\overline{A}$) : A n'est pas réalisé.

Définition 1.2.0.1. On dit que A et B sont *disjoints* ou *incompatibles* si $A \cap B = \varnothing$. De plus, si $(A_n)_{n \in \mathbf{N}_{>0}}$ est une suite d'évènements, alors :
(i) $\bigcup_{n \geqslant 1} A_n$ = « au moins un des évènements est réalisé ».
(ii) $\bigcap_{n \geqslant 1} A_n$ = « tous les évènements sont réalisés ».

Proposition 1.2.0.2 (Règles de calcul)**.** Pour A, B, C des évènements, on a :
(i) $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$ et plus généralement on a

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

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

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

(iii) $(A \cap B)^{\complement} = A^{\complement} \cup B^{\complement}$ et plus généralement on a

$$\left(\bigcap A_n\right)^{\complement} = \bigcup A_n^{\complement}.$$

(iv) $(A \cup B)^{\complement} = A^{\complement} \cap B^{\complement}$ et plus généralement on a

$$\left(\bigcup A_n\right)^{\complement} = \bigcap A_n^{\complement}.$$

(v) $\left(A^{\complement}\right)^{\complement} = A$.

Démonstration. Il suffit de procéder par double inclusion (exercice). □

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 1.3.0.1. Soit $\mathcal{A}$ un sous-ensemble de $\mathcal{P}(\Omega)$. On dit que $\mathcal{A}$ est une *tribu* (ou *σ-algèbre*) si :
(i) $\Omega \in \mathcal{A}$;
(ii) $\mathcal{A}$ est stable par passage au complémentaire : si $A \in \mathcal{A}$, alors $\overline{A} \in \mathcal{A}$;
(iii) $\mathcal{A}$ est stable par réunion dénombrable : si $(A_n)_{n\in\mathbf{N}}$ est une famille dénombrable d'ensembles de $\mathcal{A}$, alors $\bigcup_{n\in\mathbf{N}} A_n \in \mathcal{A}$.
Le couple $(\Omega, \mathcal{A})$ est appelé un *espace mesurable* ou *espace probabilisable*.

Dans une tribu, toutes les opérations finies ou dénombrables d'union, d'intersection et par complémentaire sont autorisées. De plus, on a $\varnothing \in \mathcal{A}$ car $\varnothing = \Omega^{\complement}$. Si $A, B \in \mathcal{A}$, alors $A \cup B = A \cup B \cup \varnothing \cup \cdots \in \mathcal{A}$.

Dans la suite, l'ensemble des évènements $\mathcal{A}$ associés à une expérience sera toujours une tribu. En particulier, si $\mathcal{A} \neq \mathcal{P}(\Omega)$, alors certaines parties de Ω ne sont pas des évènements aléatoires et donc on ne pourra pas calculer leur probabilité.

Proposition 1.3.0.2. Une tribu $\mathcal{A}$ est stable par intersection dénombrable.

Démonstration. On a $(A \cap B)^{\complement} = A^{\complement} \cup B^{\complement}$, donc $A \cap B = \left(A^{\complement} \cup B^{\complement}\right)^{\complement}$. Si on considère une suite d'évènements aléatoires $(A_n)_{n\geqslant 1} \in \mathcal{A}$, alors

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

□

Exemple 1.3.0.3. (1) L'ensemble $\{\varnothing, \Omega\}$ est une tribu appelée *tribu triviale* ou *grossière*, c'est la plus petite tribu.
(2) L'ensemble $\mathcal{P}(\Omega)$ est toujours une tribu (appelé *tribu discrète*), c'est même la plus grande tribu, elle contient toutes les autres.
(3) Si $A \subset \Omega$, alors $\mathcal{A} = \{\varnothing, A, A^{\complement}, \Omega\}$ est la plus petite tribu contenant A.

Définition 1.3.0.4. Une *partition* $(B_i)_{i\in I}$ de Ω est un ensemble de parties de Ω tel que :
(i) pour tout $i \in I$, B_i est non vide ;
(ii) $i \neq j \Rightarrow B_i \cap B_j = \varnothing$;
(iii) $\bigcup_{i\in I} B_i = \Omega$.

On prendra toujours I finie ou dénombrable. Remarquons que $\mathcal{A} = \left\{ \bigcup_{i\in J} B_i ;\ J \subset I \right\}$ est une tribu sur Ω. Par convention, $\bigcup_{i\in\varnothing} B_i = \varnothing$.

On peut facilement montrer que $\mathcal{A}$ est une tribu et de plus :

$$\left(\bigcup_{i\in J} B_i\right)^{\complement} = \bigcup_{i\in(I\setminus J)} B_i.$$

On dira que $\mathcal{A}$ est la tribu engendrée (comme définie ci-après) par $\mathcal{B} = \{B_i ;\ i \in I\}$.

Proposition 1.3.0.5. Toute intersection de tribus est une tribu.

Démonstration. (exercice) Remarquer que $\mathcal{A}_1 \cap \mathcal{A}_2 = \{A \in \mathcal{P}(\Omega) \text{ tel que } A \in \mathcal{A}_1 \text{ et } A \in \mathcal{A}_2\}$. □

Définition 1.3.0.6. 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 $\mathcal{A}$.

On voit que $\sigma(\mathcal{A})$ existe toujours. 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 $\mathcal{A}$).

Remarque 1.3.0.7. Si Ω est fini ou dénombrable, toute tribu est engendrée par une partition.

Exemple 1.3.0.8. Si A est une partie de Ω, alors

$$\sigma(\{A\}) = \{\varnothing, A, \overline{A}, \Omega\} \quad \text{et} \quad \sigma(\{A, \overline{A}\}) = \{\varnothing, A, \overline{A}, \Omega\}.$$

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

Proposition 1.3.0.9. La *tribu borélienne* de $\mathbf{R}^d$, notée $\mathcal{B}(\mathbf{R}^d)$ est la tribu engendrée par les ensembles ouverts de $\mathbf{R}^d$. Elle coïncide avec :
• la tribu engendrée par les ensembles fermés de $\mathbf{R}^d$;
• la tribu engendrée par les pavés $[a_1, b_1] \times \cdots \times [a_d, b_d]$;
• la tribu engendrée par les pavés $[a_1, b_1[\times \cdots \times [a_d, b_d[$;
• la tribu engendrée par les pavés $]a_1, b_1] \times \cdots \times]a_d, b_d]$;
• la tribu engendrée par les pavés $]a_1, b_1[\times \cdots \times]a_d, b_d[$;
• la tribu engendrée par les ensembles $[a_1, +\infty[\times \cdots \times [a_d, +\infty[$.

Remarque 1.3.0.10. Bien sûr, $\mathcal{B}(\mathbf{R}^d) \neq \mathcal{P}(\mathbf{R}^d)$.

Définition 1.3.0.11. Soit $(\Omega, \mathcal{A})$ un espace mesurable ou probabilisable. On appelle *probabilité* sur $(\Omega, \mathcal{A})$ toute application $\mathbf{P} \colon \mathcal{A} \to [0, 1]$ telle que :
(i) $\mathbf{P}(\Omega) = 1$;
(ii) Pour toute famille $(A_n)_{n\in\mathbf{N}}$ d'ensembles de $\mathcal{A}$ deux à deux disjoints, on a :

$$\mathbf{P}\left(\bigcup_{n\in\mathbf{N}} A_n\right) = \sum_{n\in\mathbf{N}} \mathbf{P}(A_n).$$

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

Remarque 1.3.0.12. (1) Une probabilité est une mesure positive de masse totale égale à 1.
(2) Remarquons également que $\mathbf{P}(\Omega \cup \varnothing \cup \cdots \cup \varnothing) = \mathbf{P}(\Omega) = 1$ et

$$\mathbf{P}(\Omega) + \sum_{n\geqslant 2} \mathbf{P}(A_n) \Rightarrow (\forall n \geqslant 2)\ \ \mathbf{P}(A_n) = 0$$

donc $\mathbf{P}(\varnothing) = 0$.

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

Proposition 1.3.0.13. Soit $(\Omega, \mathcal{A}, \mathbf{P})$ un espace de probabilité, $A, B \in \mathcal{A}, (A_n)_{n\in\mathbf{N}}$ une famille dénombrable d'ensembles de $\mathcal{A}$. (i) On a $\mathbf{P}(A^\complement) = 1 - \mathbf{P}(A)$.
(ii) On a $\mathbf{P}(A) = \mathbf{P}(A \cap B) + \mathbf{P}(A \cap B^\complement)$.
(iii) Si $A \subset B$, alors $\mathbf{P}(A) \leqslant \mathbf{P}(B)$.
(iv) On a $\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(A \cap B)$.
(v) On a $\mathbf{P}\big(\bigcup_{n\in\mathbf{N}} A_n\big) \leqslant \sum_{n\in\mathbf{N}} \mathbf{P}(A_n)$.
(vi) Si $(A_n)_{n\in\mathbf{N}}$ est croissante, *i.e.* $A_n \subset A_{n+1}$, alors on a $\mathbf{P}\big(\bigcup_{n\in\mathbf{N}} A_n\big) = \lim_{n\to+\infty} \mathbf{P}(A_n)$.
(vii) Si $(A_n)_{n\in\mathbf{N}}$ est décroissante, alors on a $\mathbf{P}\big(\bigcap_{n\in\mathbf{N}} A_n\big) = \lim_{n\to+\infty} \mathbf{P}(A_n)$.

Démonstration. (i) On a $1 = \mathbf{P}(\Omega) = \mathbf{P}(A \cup A^\complement) = \mathbf{P}(A) + \mathbf{P}(A^\complement)$ car l'union est disjointe.
(ii) On a $A = A \cap (B \cup B^\complement) = (A \cap B) \cup (A \cap B^\complement)$ et l'union est disjointe.
(iii) On a $B = (B \cap A) \cup (B \cap A^\complement)$ et l'union est disjointe. On a donc

$$\mathbf{P}(B) = \mathbf{P}(B \cap A) + \mathbf{P}(B \cap A^\complement) = \mathbf{P}(A) + \mathbf{P}(B \cap A^\complement) \geqslant \mathbf{P}(A).$$

(iv) On a $A \cup B = A \cup (B \cap A^\complement)$, donc $\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B \cap A^\complement)$. De plus, on a également $B = (B \cap A) \cup (B \cap A^\complement)$, donc $\mathbf{P}(B \cap A^\complement) = \mathbf{P}(B) - \mathbf{P}(A \cap B)$.
(v) On suppose le point (vi) prouvé. On pose $B_n = \bigcup_{i=0}^{n} A_i$. Comme $B_n = B_{n-1} \cup A_n$, on a $\mathbf{P}(B_n) \leqslant \mathbf{P}(B_{n-1}) + \mathbf{P}(A_n)$ et par récurrence immédiate,

$$\mathbf{P}(B_n) \leqslant \sum_{i=0}^{n} \mathbf{P}\left(A_i\right) \leqslant \sum_{i=0}^{+\infty} \mathbf{P}(A_i).$$

Comme $(B_n)_{n\in\mathbf{N}}$ est croissante, le point (vi) donne également

$$\lim_{n\to+\infty} \mathbf{P}(B_n) = \mathbf{P}\left(\bigcup_{n\in\mathbf{N}} B_n\right) = \mathbf{P}\left(\bigcup_{n\in\mathbf{N}} A_n\right).$$

(vi) On définit une suite d'évènements $(C_n)_{n\in\mathbf{N}}$ par $C_0 = A_0$, $C_n = A_n \backslash A_{n-1} := A_n \cap (A_{n-1})^\complement$ pour $n \geqslant 1$. L'ensemble des $(C_n)_{n\in\mathbf{N}}$ est ainsi une suite d'évènements, deux à deux disjoints, vérifiant

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

On a alors

$$\mathbf{P}\left(\bigcup_{n\in\mathbf{N}} A_n\right) = \mathbf{P}\left(\bigcup_{n\in\mathbf{N}} C_n\right) = \sum_{n\in\mathbf{N}} \mathbf{P}(C_n) = \lim_{n\to+\infty} \sum_{i=0}^{n} \mathbf{P}(C_i) = \lim_{n\to+\infty} \mathbf{P}(A_n).$$

(vii) Il suffit de passer au complémentaire pour utiliser le point précédent : on pose $B_n = \overline{A_n}$ pour tout $n \in \mathbf{N}$. L'ensemble des $(B_n)_{n\in\mathbf{N}}$ est une suite croissante, donc on a $\mathbf{P}\left(\bigcup_{n\in\mathbf{N}} B_n\right) = \lim_{n\to+\infty} \mathbf{P}(B_n)$. Ainsi,

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

et

$$\lim_{n\to+\infty} \mathbf{P}(B_n) = \lim_{n\to+\infty} \mathbf{P}(\overline{A_n}) = 1 - \lim_{n\to+\infty} \mathbf{P}(A_n).$$

□

Une question naturelle est de savoir s'il est toujours possible de définir un espace de probabilité à partir d'un espace probabilisable, c'est-à-dire : peut-on toujours construire au moins une mesure de probabilité sur un espace probabilisable ? Si $\Omega = \varnothing$, alors cela n'est pas possible car on doit alors avoir $\mathbf{P}(\Omega) = 1$ et $\mathbf{P}(\varnothing) = 0$, ce qui n'est pas compatible. Par contre, si Ω est non vide, c'est effectivement toujours possible de créer une mesure de probabilité de type "Dirac". En effet, prenons $x \in \Omega$, alors la mesure suivante

$$\mathbf{P}_x \colon A \in \mathcal{A} \mapsto \begin{cases} 1 & \text{si } x \in A \\ 0 & \text{sinon} \end{cases}$$

est une mesure de probabilité.

1.4 Probabilité uniforme

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

Proposition 1.4.0.1. La formule

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

définit une probabilité $\mathbf{P}$ sur $(\Omega, \mathcal{P}(\Omega))$ dès que la famille de nombre réels $(p_\omega)_{\omega\in\Omega}$ vérifie :
(i) $p_\omega \in [0,1]$ pour tous les $\omega \in \Omega$;
(ii) $\sum_{\omega\in\Omega} p_\omega = 1$.
Inversement, toute probabilité $\mathbf{P}$ sur $(\Omega, \mathcal{P}(\Omega))$ est de cette forme : il suffit de poser $p_\omega := \mathbf{P}(\{\omega\})$ pour tout $\omega \in \Omega$.

Démonstration. • Le sens direct est laissé en exercice.
• Pour le sens réciproque, tout $A \in \mathcal{P}(\Omega)$ s'écrit $A = \bigcup_{\omega\in A} \{\omega\}$ (union disjointe finie ou dénombrable), donc on a directement $\mathbf{P}(A) = \sum_{\omega\in A} \mathbf{P}(\{\omega\})$. □

Remarque 1.4.0.2. La proposition 1.4.0.1 est fausse si Ω n'est pas dénombrable.

Définition 1.4.0.3. Soit $\mathbf{P}$ une probabilité sur $(\Omega, \mathcal{P}(\Omega))$ avec Ω fini. On dit que $\mathbf{P}$ est la *loi uniforme* sur Ω si toutes les épreuves $\omega \in \Omega$ sont *équiprobables* : *i.e.*

$$p_\omega = \frac{1}{|\Omega|}.$$

En particulier, on a

$$\mathbf{P}(A) = \frac{|A|}{|\Omega|} \quad (\forall A \subset \Omega).$$

Remarque 1.4.0.4. (1) Le calcul des probabilités se ramène ici à un simple calcul de dénombrement. On a

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

(2) Si Ω est dénombrable infini, il n'existe pas de probabilité uniforme.
(3) « au hasard » signifie qu'on munit $(\Omega, \mathcal{P}(\Omega), \mathbf{P})$ avec $\mathbf{P}$ uniforme.

2 Probabilités conditionnelles, indépendance

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

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 $\mathbf{P}$ pour tenir compte de cette nouvelle information.

Définition 2.1.0.1. Soient $A, B \in \mathcal{A}$ tels que $\mathbf{P}(B) > 0$. On appelle *probabilité conditionnelle de A sachant B*, le nombre réel

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

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

On sait que le résultat ω de notre expérience est dans B. Sachant cela, on cherche à savoir la probabilité que A se réalise.

Exemple 2.1.0.2. Si on pose $\Omega =$ « l'ensemble des jours de 2022 » (muni de la tribu $\mathcal{P}(\Omega)$ et $\mathbf{P}$ uniforme) et $A =$ « l'ensemble des jours où la température a été supérieur à 10 degrés à une heure donnée et à un point donné », on a donc $\mathbf{P}(A) = \frac{|A|}{|\Omega|}$ (proportion annuelle). Si de plus, on pose $B =$ « l'ensemble des jours du mois de janvier », alors

$$\mathbf{P}(A|B) = \frac{\mathbf{P}(A \cap B)}{\mathbf{P}(B)} = \frac{|A \cap B|}{|\Omega|} \times \frac{|\Omega|}{|B|} = \frac{|A \cap B|}{|B|}$$

(proportion mensuelle).

Proposition 2.1.0.3. Soit $B \in \mathcal{A}$ tel que $\mathbf{P}(B) > 0$. L'application

$$\begin{aligned} \mathbf{P}(\cdot|B)\colon \mathcal{A} &\to [0,1] \\ A &\mapsto \mathbf{P}(A|B) \end{aligned}$$

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

Démonstration. • Pour tout $A \in \mathcal{A}$, on a $A \cap B \subset B$, donc $\mathbf{P}(A|B) \in [0,1]$.
• De plus, $\mathbf{P}(\Omega|B) = \frac{\mathbf{P}(B)}{\mathbf{P}(B)} = 1$.
• Enfin, si $(A_n)_{n\in\mathbf{N}}$ est une famille d'évènements aléatoires deux à deux disjoints, on a

$$\mathbf{P}\Big(\bigcup_{n\in\mathbf{N}} A_n|B\Big) = \frac{\mathbf{P}\big(\big(\bigcup_{n\in\mathbf{N}} A_n\big) \cap B\big)}{\mathbf{P}(B)} = \frac{\mathbf{P}\big(\bigcup_{n\in\mathbf{N}} (A_n \cap B)\big)}{\mathbf{P}(B)} = \sum_{n\in\mathbf{N}} \frac{\mathbf{P}(A_n \cap B)}{\mathbf{P}(B)} = \sum_{n\in\mathbf{N}} \mathbf{P}(A_n|B).$$

□

Remarque 2.1.0.4. (1) On a $\mathbf{P}(B|B) = 1$, donc $\mathbf{P}(B^{\complement}|B) = 0$.
(2) $(A|B)$ ne veut rien dire. En particulier, ce n'est pas un évènement !
(3) Conséquences : $\mathbf{P}(A^{\complement}|B) = 1 - \mathbf{P}(A|B)$, $\mathbf{P}(A \cup C|B) = \mathbf{P}(A|B) + \mathbf{P}(C|B) - \mathbf{P}(A \cap C|B)$, etc...

2.2 Formules usuelles

2.2.1 Formule des probabilités composées

Théorème 2.2.1.1 (Formule des probabilités composées)**.** (i) Soient $A, B \in \mathcal{A}$ avec $\mathbf{P}(A) > 0$. On a la formule des probabilités composées suivante :

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

(ii) Plus généralement, soient $n \geqslant 2$ et $A_1, \ldots, A_n \in \mathcal{A}$ tels que $\mathbf{P}(A_1 \cap \cdots \cap A_{n-1}) > 0$. On a alors

$$\mathbf{P}(A_1 \cap \cdots \cap A_n) = \mathbf{P}(A_1)\,\mathbf{P}(A_2|A_1) \cdots \mathbf{P}(A_n|A_1 \cap \cdots \cap A_{n-1}).$$

Remarque 2.2.1.2. Comme pour tout $1 \leqslant k \leqslant n-1$, $(A_1 \cap \cdots \cap A_{n-1}) \subset (A_1 \cap \cdots \cap A_k)$, alors $\mathbf{P}(A_1 \cap \cdots \cap A_k) > 0$ et donc les probabilités conditionnelles précédentes sont bien définies.

Démonstration. Montrons le second point par récurrence sur n, le premier point correspond à $n = 2$. Pour $n = 2$, le résultat découle directement de la formule des probabilités conditionnelles. Supposons le résultat vrai pour $n \geqslant 2$. On a alors

$$\mathbf{P}(A_1 \cap \cdots \cap A_{n+1}) = \mathbf{P}(A_1 \cap \cdots \cap A_n)\,\mathbf{P}(A_{n+1}|A_1 \cap \cdots \cap A_n)$$

puis on conclut grâce à l'hypothèse de récurrence. □

Le premier point du théorème 2.2.1.1 peut sembler à première vue inutile car il s'agit d'une simple réécriture de la définition 2.1.0.1. 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 évènements 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 évènements.

Exemple 2.2.1.3. On prend 2 urnes, l'une avec 2 boules blanches et 1 rouge, l'autre avec 2 boules rouges et 1 blanche. On choisit au hasard une urne. Une fois l'urne choisie, on tire au hasard une boule. On pose les évènements :
• A_1 : on a choisi l'urne 1 ;
• A_2 : on a tiré une boule rouge.
On a donc $\mathbf{P}(A_1 \cap A_2) = \mathbf{P}(A_2|A_1)\,\mathbf{P}(A_1) = \frac{1}{3} \times \frac{1}{2} = \frac{1}{6}$.

2.2.2 Formule des probabilités totales

Théorème 2.2.2.1 (Formule des probabilités totales)**.** (i) Soient $A, B \in \mathcal{A}$ tels que $0 < \mathbf{P}(A) < 1$, on a

$$\mathbf{P}(B) = \mathbf{P}(A)\,\mathbf{P}(B|A) + \mathbf{P}(A^{\complement})\,\mathbf{P}(B|A^{\complement}).$$

(ii) Plus généralement, si $(A_n)_{1\leqslant n<N}$ est une partition de Ω avec $\mathbf{P}(A_n) > 0$ pour tout $n < N$, on a

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

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

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

Il suffit alors d'utiliser la formule des probabilités composées pour conclure. Le premier point se traite de la même façon car $\{A, A^{\complement}\}$ est une partition de Ω. □

2.2.3 Formule de Bayes

Théorème 2.2.3.1 (Formule de Bayes)**.** (i) Soient $A, B \in \mathcal{A}$ tels que $0 < \mathbf{P}(A) < 1$ et $\mathbf{P}(B) > 0$, on a

$$\mathbf{P}(A|B) = \frac{\mathbf{P}(A)\,\mathbf{P}(B|A)}{\mathbf{P}(A)\,\mathbf{P}(B|A) + \mathbf{P}(A^{\complement})\,\mathbf{P}(B|A^{\complement})}.$$

(ii) Plus généralement, si $(A_n)_{1\leqslant n<N}$ est une partition de Ω avec $\mathbf{P}(A_n) > 0$ pour tout $n < N$ et $\mathbf{P}(B) > 0$, on a

$$\mathbf{P}(A_k|B) = \frac{\mathbf{P}(A_k)\,\mathbf{P}(B|A_k)}{\sum\limits_{n<N} \mathbf{P}(A_n)\,\mathbf{P}(B|A_n)} \qquad (\forall k < N).$$

Démonstration. Pour tout $k < N$, on a

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

et on applique la formule des probabilités totales pour conclure. Encore une fois, le premier point se traite de la même façon. □

Exemple 2.2.3.2. On reprend notre expérience des 2 urnes. Sachant que l'on a tiré une boule rouge, quelle est la probabilité que l'on ait choisi la première urne ? On a :

$$\mathbf{P}(A_1|A_2) = \frac{\mathbf{P}(A_1)\,\mathbf{P}(A_2|A_1)}{\mathbf{P}(A_1)\,\mathbf{P}(A_2|A_1) + \mathbf{P}(A_1^{\complement})\,\mathbf{P}(A_2|A_1^{\complement})} = \frac{1/2 \times 1/3}{1/2 \times 1/3 + 1/2 \times 2/3} = \frac{1}{3}.$$

2.3 Évènements indépendants

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

Définition 2.3.0.1. Soient $A, B \in \mathcal{A}$.
(i) A et B sont *indépendants* si

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

(ii) $A_1, \ldots, A_n$ sont deux à deux indépendants si pour tout $i \neq j$,

$$\mathbf{P}(A_i \cap A_j) = \mathbf{P}(A_i)\,\mathbf{P}(A_j).$$

(iii) $A_1, \ldots, A_n$ sont *mutuellement indépendants* si pour tout $J \subset [\![1, n]\!]$ avec $|J| \geqslant 2$,

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

Remarque 2.3.0.2. (1) La notion d'indépendance est différente de la notion d'incompatibilité. En particulier, si A et B sont incompatibles (*i.e.* disjoints), alors $\mathbf{P}(A \cap B) = 0$.
(2) Si $A, B \in \mathcal{A}$ sont indépendants avec $\mathbf{P}(B) > 0$, alors on a

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

(3) On peut trouver A_1, A_2, A_3 deux à deux indépendants mais pas mutuellement indépendants.

Proposition 2.3.0.3. Soient $A, B \in \mathcal{A}$. Si A et B sont indépendants, alors A et $B^{\complement}$, $A^{\complement}$ et B, $A^{\complement}$ et $B^{\complement}$ sont également indépendants.

Démonstration. Par symétrie, il suffit de montrer que A et $B^{\complement}$ sont indépendants. On a $\mathbf{P}(A) = \mathbf{P}(A \cap B) + \mathbf{P}(A \cap B^{\complement})$, donc

$$\mathbf{P}(A \cap B^{\complement}) = \mathbf{P}(A) - \mathbf{P}(A)\,\mathbf{P}(B) = \mathbf{P}(A)(1 - \mathbf{P}(B)) = \mathbf{P}(A)\,\mathbf{P}(B^{\complement}).$$

□

Proposition 2.3.0.4. Soient $(A_n)_{n<N}$ une famille d'évènements 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.

Démonstration. On a

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

□

Exemple 2.3.0.5. Soit $\Omega = \{(0,0), (0,1), (1,0), (1,1)\}$, $\mathcal{A} = \mathcal{P}(\Omega)$ et $\mathbf{P}$ une probabilité uniforme. Posons

$$A_1 = \{(0,0), (0,1)\} \quad A_2 = \{(0,0), (1,0)\} \quad A_3 = \{(0,0), (1,1)\}.$$

On a donc

$$\begin{aligned}\mathbf{P}(A_1) = \mathbf{P}(A_2) = \mathbf{P}(A_3) &= \frac{2}{4} = \frac{1}{2} \\ \mathbf{P}(A_1 \cap A_2) = \mathbf{P}(A_1 \cap A_3) = \mathbf{P}(A_2 \cap A_3) &= \frac{1}{4} \\ \mathbf{P}(A_1 \cap A_2 \cap A_3) = \frac{1}{4} \neq \frac{1}{8} &= \mathbf{P}(A_1)\,\mathbf{P}(A_2)\,\mathbf{P}(A_3)\end{aligned}$$

donc A_1, A_2, A_3 ne sont pas mutuellement indépendants. Si on prend $A_4 = \{(0,1), (1,0)\}$, on voit bien que A_1, A_2, A_4 sont deux à deux indépendants. De plus, $A_1 \cap A_2 \cap A_4 = \varnothing$, donc ils ne sont également pas mutuellement indépendants.

3 Variables aléatoires

3.1 Rappels et définitions

Commençons par quelques rappels de théorie de la mesure.

Définition 3.1.0.1. Soit $(E, \mathcal{A})$ et $(F, \mathcal{B})$ deux espaces mesurables. On dit qu'une fonction $f\colon (E, \mathcal{A}) \to (F, \mathcal{B})$ est *mesurable* si pour tout $B \in \mathcal{B}$, on a $f^{-1}(B) \in \mathcal{A}$, en rappelant que

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

Remarque 3.1.0.2. (1) Si $\mathcal{A} = \mathcal{P}(E)$, toute application $f\colon (E, \mathcal{A}) \to (F, \mathcal{B})$ est mesurable.
(2) On a stabilité de l'intersection et de l'union par image réciproque :

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

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

Cependant, par image direct, il y a une légère subtilité. Rappelons d'abord que

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

donc on a

$$f(A_1 \cup A_2) = f(A_1) \cup f(A_2) \quad \text{et} \quad f(A_1 \cap A_2) \subsetneq f(A_1) \cap f(A_2).$$

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

Définition 3.1.0.3. Soit $(F, \mathcal{B})$ un espace mesurable. On appelle *variable aléatoire* (v.a. en abrégé) définie sur $(\Omega, \mathcal{A}, \mathbf{P})$ et à valeurs dans F, toute application mesurable $X\colon \Omega \to F$, c'est-à-dire que pour tout $B \in \mathcal{B}$, on a $X^{-1}(B) \in \mathcal{A}$. On appelle *variable aléatoire réelle* une variable aléatoire à valeurs dans $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$. On appelle *vecteur aléatoire réelle* une variable aléatoire à valeurs dans $(\mathbf{R}^d, \mathcal{B}(\mathbf{R}^d))$.

Remarque 3.1.0.4. Soit X une variable aléatoire réelle. On utilisera les notations suivantes :
- $\{X = a\} := X^{-1}(\{a\}) = \{\omega \in \Omega;\ X(\omega) = a\}$;
- $\{X \in [a, b]\} := \{a \leqslant X \leqslant b\} = X^{-1}([a, b]) := \{\omega\ \in \Omega;\ X(\omega) \in [a, b]\}$;
- $\{X \leqslant t\} = X^{-1}(]-\infty, t])$;
- $\{X \in B\} = X^{-1}(B)$.

Proposition 3.1.0.5. (i) Soit X un vecteur aléatoire réel et $h\colon \mathbf{R}^d \to \mathbf{R}^{d'}$ une fonction mesurable. L'ensemble $h(X)$ est alors un vecteur aléatoire réel.
(ii) $X = (X_1, \ldots, X_n)$ est un vecteur aléatoire réel si et seulement si chacune de ses composantes X_i est une variable aléatoire réelle.

Remarque 3.1.0.6. On a vu que $\mathcal{B}(\mathbf{R})$ coïncide avec la tribu engendrée par les intervalles $[a, b]$ avec $a, b \in \mathbf{R}$ 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 évènement aléatoire.

3.2 Loi de probabilité

3.2.1 Cadre général

Proposition-Définition 3.2.1.1. Soit X une variable aléatoire. On appelle *loi de probabilité* de X la mesure de probabilité $\mathbf{P}_X$ définie sur $(F, \mathcal{B})$ par $\mathbf{P}_X(B) := \mathbf{P}(X^{-1}(B)) = \mathbf{P}(X \in B)$ pour tout $B \in \mathcal{B}$. C'est la mesure image de $\mathbf{P}$ par X.

Démonstration. • Tout d'abord, $\mathbf{P}_X$ 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, $\mathbf{P}(X \in B)$ a bien un sens.
• $\mathbf{P}_X(B) = \mathbf{P}(X \in B) \in [0, 1]$.

• $\mathbf{P}_X(F) = \mathbf{P}(X \in F) = \mathbf{P}(\Omega) = 1$.
• Soit $(B_n)_{n\geqslant 1}$ une famille disjointes d'éléments de $\mathcal{B}$. On a alors

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

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

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

□

Remarque 3.2.1.2. Puisque $\mathcal{B}(\mathbf{R})$ 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 : • $\mathbf{P}(X \in [a, b])$, pour tous $a < b$;
• $\mathbf{P}(X \in]a, b[)$, pour tous $a < b$;
• $\mathbf{P}(X \leqslant t)$, pour tous $t \in \mathbf{R}$.

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. Cependant, 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 !

3.2.2 Cas particulier des variables aléatoires discrètes

Définition 3.2.2.1. 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 $\mathbf{P}(X \in B) = 1$, on dit également que X est une variable discrète. On appelle alors *support* de X le plus petit ensemble discret $\widetilde{B}$ vérifiant $\mathbf{P}(X \in \widetilde{B}) = 1$. 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 $F = \{x_i;\ 0 \leqslant i < N\}$, avec $N \in \mathbf{N}_{>0}$ ou $N = +\infty$. Comme F est discret, on peut prendre $\mathcal{B} = \mathcal{P}(F)$. La variable aléatoire X définit une probabilité $\mathbf{P}_X$ sur l'ensemble mesurable discret $(F, \mathcal{B})$. On a vu au chapitre 1 comment caractériser une probabilité sur un ensemble discret :

Pour connaître $\mathbf{P}_X$, il suffit de connaître les

$$\mathbf{P}_X(\{x_i\}) = \mathbf{P}(X = x_i) \quad \text{pour} \quad 0 \leqslant i < N.$$

Proposition 3.2.2.2. Soit X une variable aléatoire discrète. La loi de X est caractérisée par : (i) $X(\Omega) = \{x_i;\ 0 \leqslant i < N\}$ avec $N \in \mathbf{N}_{>0}$ ou $N = +\infty$.
(ii) $\mathbf{P}_X(\{x_i\}) = \mathbf{P}(X = x_i)$ pour $0 \leqslant i < N$.
En particulier, pour $A \subset \mathbf{R}$, on a

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

Exemple 3.2.2.3. (1) Soit $\Omega = \{1, \ldots, 6\}$ et $\mathbf{P}$ la probabilité uniforme. On pose $X(\omega) = (\omega - 3)^2$. L'application X est une variable aléatoire nécessairement discrète car Ω est discret. De plus, $X(\Omega) = \{0, 1, 4, 9\}$. On a alors

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

(2) Soit $\Omega = [0, 1]$ et $\mathbf{P}$ la mesure de Lebesgue sur Ω. On pose $X(\Omega) = \lfloor 2\omega \rfloor$. On a $X(\Omega) = \{0, 1, 2\}$, donc X est une variable aléatoire discrète. On a alors

$$\begin{aligned}
&\mathbf{P}(X = 0) = \mathbf{P}([0, 1/2[) = 1/2,\\
&\mathbf{P}(X = 1) = \mathbf{P}([1/2, 1[) = 1/2,\\
&\mathbf{P}(X = 2) = \mathbf{P}(\{1\}) = 0
\end{aligned}$$

(3) On lance deux pièces équilibrées. On a $\Omega = \{\omega = (\omega_1, \omega_2);\ \omega_1, \omega_2 \in \{0,1\}\}$. 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 : $Y((\omega_1, \omega_2)) = \omega_2$.

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

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

Ainsi, X et Y ont même loi mais $X \neq Y$. En effet, $(X = Y)$ = « les deux pièces ont même résultat » et $\mathbf{P}(X = Y) = 1/2$.

Remarque 3.2.2.4. 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-Définition 3.3.1.1. (i) Si X est une variable aléatoire réelle positive, alors l'*espérance* de X est définie par la quantité, éventuellement infinie,

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

(ii) Si X est une variable aléatoire réelle non nécessairement positive, alors l'espérance de X est définie uniquement si

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

Dans ce cas, on dit qu'elle est *intégrable* et elle vaut $\mathbf{E}[X] \in \mathbf{R}$.

Remarque 3.3.1.2. L'espérance de X n'existe pas toujours lorsque la variable aléatoire 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 $\mathbf{P}_X$.

Proposition 3.3.1.3 (Formule de changement de variable)**.** Si X est une variable aléatoire réelle telle que $\mathbf{E}[X]$ existe, alors

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

Proposition 3.3.1.4. On a les propriétés suivantes qui découlent des propriétés de l'intégrale de Lebesgue.
(i) *(Positivité).* Si $X \geqslant 0$, alors $\mathbf{E}[X] \geqslant 0$, donc si $X \geqslant Y$, alors $\mathbf{E}[X] \geqslant \mathbf{E}[Y]$.
(ii) *(Linéarité).* Soient $a, b \in \mathbf{R}$, X et Y deux variables aléatoires définies sur le même espace de probabilité, alors $\mathbf{E}[aX + bY] = a\,\mathbf{E}[X] + b\,\mathbf{E}[Y]$.
(iii) $\mathbf{E}[1] = \mathbf{P}(\Omega) = 1$.
De plus, l'espérance d'une variable aléatoire ne dépend que de sa loi : deux variables aléatoires de même loi, ont même espérance (si elle existe).

Proposition 3.3.1.5 (Cas des variables aléatoires discrètes)**.** Soit X une variable aléatoire discrète. (i) Si X est positive, alors

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

(ii) Si X est de signe quelconque et

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

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

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

Démonstration. La mesure $\mathbf{P}_X$ est totalement déterminée par les $\mathbf{P}_X(\{x_i\}) = \mathbf{P}(X = x_i)$. On note $p(x_i) = \mathbf{P}_X(\{x_i\})$. On a alors

$$\mathbf{P}_X = \sum_{x_i \in X(\Omega)} p(x_i)\delta_{x_i}$$

avec δ_{x_i} la mesure de Dirac. Ainsi, on obtient

$$\begin{aligned}\mathbf{E}[X] &= \int_{\mathbf{R}} x_i \,\mathrm{d}\mathbf{P}_X(x_i) = \sum_{x_i \in X(\Omega)} x_i p(x_i) \\ &= \sum_{x_i \in X(\Omega)} x_i \,\mathbf{P}(X = x_i).\end{aligned}$$

□

Exemple 3.3.1.6. Soit $\Omega = \{1, \dots, 6\}$ et $\mathbf{P}$ la probabilité uniforme. On pose $X(\omega) = (\omega - 3)^2$. On a alors

$$\mathbf{E}[X] = 0 \times \mathbf{P}(X = 0) + 1 \times \mathbf{P}(X = 1) + 4 \times \mathbf{P}(X = 4) + 9 \times \mathbf{P}(X = 9) = \frac{19}{6}.$$

3.3.2 Formule de transfert

Proposition 3.3.2.1 (Formule de transfert)**.** Soit $g\colon \mathbf{R} \to \mathbf{R}$ une fonction mesurable et X une variable aléatoire réelle. On pose $Y = g(X)$. L'ensemble Y est alors une variable aléatoire réelle. Si $\int_{\mathbf{R}} |g(x)| \,\mathrm{d}\mathbf{P}_X(x) < +\infty$, alors Y est intégrable et

$$\mathbf{E}[Y] = \int_{\mathbf{R}} g(x) \,\mathrm{d}\mathbf{P}_X(x).$$

Remarque 3.3.2.2. 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.

Démonstration. On commence par supposer que $g(x) = \mathbf{1}_A(x)$ avec $A \in \mathcal{B}(\mathbf{R})$. On a alors

$$\begin{aligned}\mathbf{E}[g(X)] &= \int_{\Omega} \mathbf{1}_A(X(\omega)) \,\mathrm{d}\mathbf{P}(\omega) = \mathbf{P}(X \in A) \\ &= \mathbf{P}_X(A) = \int_{\mathbf{R}} \mathbf{1}_A(x) \,\mathrm{d}\mathbf{P}_X(x).\end{aligned}$$

Ensuite, on prend

$$g(x) = \sum_{i=1}^{n} \alpha_i \mathbf{1}_{A_i}(x)$$

avec $\alpha_i \in \mathbf{R}$ et $A_i \in \mathcal{B}(\mathbf{R})$ pour tout $i \in [\![1, n]\!]$. Par linéarité à droite et à gauche, on a

$$\begin{aligned}\mathbf{E}[g(X)] &= \int_{\Omega} \sum_{i=1}^{n} \alpha_i \mathbf{1}_{A_i}(X(\omega)) \,\mathrm{d}\mathbf{P}(\omega) \\ &= \int_{\mathbf{R}} \sum_{i=1}^{n} \alpha_i \mathbf{1}_{A_i}(x) \,\mathrm{d}\mathbf{P}_X(x).\end{aligned}$$

De plus, si on prend maintenant $g \geqslant 0$, alors g est limite d'une suite croissante de fonctions étagées toutes positives. On obtient donc

$$\mathbf{E}[g(X)] = \int_{\mathbf{R}} g(x) \,\mathrm{d}\mathbf{P}_X(x)$$

par le théorème de convergence monotone.
Enfin, si g est réelle quelconque, on écrit $g = g_+ - g_-$ avec $g_+ = \max(g, 0) \geqslant 0$ et $g_- = -\min(g, 0) \geqslant 0$ puis on procède comme précédemment. □

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

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

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

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

Si $\int_{\mathbf{R}} |g(x)| \,\mathrm{d}\mathbf{P}_X(x) < +\infty$, alors Y est intégrable et ainsi la formule est encore vérifiée. □

Regardons maintenant le cas particulier des variables aléatoire discrètes.

Proposition 3.3.2.3 (Formule de transfert, cas discret)**.** Soit $g\colon \mathbf{R} \to \mathbf{R}$ une fonction mesurable et X une variable aléatoire réelle discrète. On pose $Y = g(X)$. L'ensemble Y est alors une variable aléatoire réelle discrète.
(i) Si Y est positive, alors

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

(ii) Si Y est de signe quelconque et que $\sum_{x_i \in X(\Omega)} |g(x_i)|\ \mathbf{P}(X = x_i) < +\infty$, alors on a

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

Démonstration. 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 variables aléatoires réelles discrètes positives telles que $X(\Omega) = \{x_i;\ 0 \leqslant i < N\}$ et $Y(\Omega) = \{y_i;\ 0 \leqslant j < M\}$. On a alors

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

Néanmoins, $\mathbf{P}(Y = y_j) = \mathbf{P}(X \in g^{-1}(\{y_j\}))$ avec $g^{-1}(\{y_j\}) = \{x_k \in X(\Omega);\ g(x_k) = y_j\}$, donc

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

□

Exemple 3.3.2.4. Reprenons l'exemple $X(\omega) = (\omega - 3)^2$. On pose D la variable aléatoire donnant la valeur du dé. L'application D est à valeurs dans $\{1, \ldots, 6\}$ et de loi uniforme : $\mathbf{P}(D = k) = 1/6$ pour tout $k \in \{1, \ldots, 6\}$.

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

Définition 3.3.2.5. Si $\mathbf{E}[X^k]$ existe, $\mathbf{E}[X^k]$ s'appelle le *moment d'ordre k* de X.

Remarque 3.3.2.6. $\mathbf{E}[X^k]$ existe si $X^k \geqslant 0$ ou X^k est intégrable.

Proposition 3.3.2.7. Si X est une variable aléatoire discrète à valeurs dans $\{x_i;\ 0 \leqslant i < N\}$ avec $N \in \mathbf{N}$ ou $N = +\infty$ et telle que $\mathbf{E}[|X^k|] = \sum_{0 \leqslant i < N} |x_i|^k\, \mathbf{P}(X = x_i) < +\infty$ alors

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

3.3.3 Variance d'une variable aléatoire réelle

Définition 3.3.3.1. Soit X une variable aléatoire de carré intégrable, *i.e.* $\mathbf{E}[X^2] < +\infty$. On appelle *variance* de X, notée $\mathbf{V}(X)$, le nombre positif

$$\mathbf{V}(X) = \mathbf{E}[(X - \mathbf{E}[X])^2].$$

On appelle alors *écart-type* de X le nombre positif

$$\sigma := \sqrt{\mathbf{V}(X)}.$$

Remarque 3.3.3.2. (1) $\mathbf{V}(X) \geqslant 0$.
(2) La variance est une mesure de l'écart de la variable aléatoire à sa moyenne : plus la variance est petite, plus la variable aléatoire a des valeurs concentrées près de sa moyenne.
(3) $\mathbf{V}(X) = 0$ si et seulement si $X = \mathbf{E}[X]$ $\mathbf{P}$-presque sûrement, *i.e.* X constante $\mathbf{P}$-presque sûrement, il existe $c \in \mathbf{R}$ telle que $\mathbf{P}(X = c) = 1$.
(4) Si $\mathbf{E}[X^2] < \infty$, alors $\mathbf{E}[X]$ existe. En effet, d'après l'inégalité de Hölder :

$$\mathbf{E}[|X| \times 1] \leqslant \mathbf{E}[|X|^2]^{1/2} \times \mathbf{E}[1]^{1/2} \Rightarrow \mathbf{E}[|X|]^2 \leqslant \mathbf{E}[X^2]$$

donc $\mathbf{E}[X^2] < +\infty \Rightarrow \mathbf{E}[|X|] < +\infty \Rightarrow \mathbf{E}[X]$ existe. De manière générale, si $1 \leqslant p \leqslant q$, $L^1(\Omega, \mathcal{A}, \mathbf{P}) \subset L^p(\Omega, \mathcal{A}, \mathbf{P})$.

Proposition 3.3.3.3. Soit X une variable aléatoire de carré intégrable et $a, b \in \mathbf{R}$. On a

$$\mathbf{V}(aX + b) = a^2 \, \mathbf{V}(X).$$

Démonstration. On a

$$\begin{aligned} \mathbf{V}(aX + b) &= \mathbf{E}[(aX + b - \mathbf{E}[aX + b])^2] \\ &= \mathbf{E}[a^2 (X - \mathbf{E}[X])^2] \\ &= a^2 \, \mathbf{V}(X). \end{aligned}$$

□

Proposition 3.3.3.4. Soit X une variable aléatoire de carré intégrable. On a alors

$$\mathbf{V}(X) = \mathbf{E}[X^2] - \mathbf{E}[X]^2.$$

Démonstration. On a

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

□

Proposition 3.3.3.5. Si X est une variable aléatoire discrète à valeurs dans $\{x_i;\ 0 \leqslant i < N\}$ avec $N \in \mathbf{N}$ ou $N = +\infty$ et telle que $\mathbf{E}[|X|^2] = \sum_{0 \leqslant i < N} |x_i|^2 \, \mathbf{P}(X = x_i) < +\infty$, alors

$$\mathbf{V}(X) = \sum_{0 \leqslant i < N} x_i^2 \, \mathbf{P}(X = x_i) - \left(\sum_{0 \leqslant i < N} x_i \, \mathbf{P}(X = x_i) \right)^2.$$

3.4 Variables aléatoires indépendantes

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

$$X \colon (\Omega, \mathcal{A}, \mathbf{P}) \to (E_1, \mathcal{B}_1) \quad \text{et} \quad Y \colon (\Omega, \mathcal{A}, \mathbf{P}) \to (E_2, \mathcal{B}_2).$$

On peut considérer le vecteur aléatoire $(X, Y) \colon (\Omega, \mathcal{A}, \mathbf{P}) \to (E_1 \times E_2, \mathcal{B}_1 \times \mathcal{B}_2)$ avec $\mathcal{B}_1 \times \mathcal{B}_2$ la tribu produit.

Définition 3.4.0.1. On dit que X et Y sont *indépendantes* si $\mathbf{P}_{(X,Y)}$ est la mesure produit $\mathbf{P}_X \times \mathbf{P}_Y$.

Rappel. Si $B_1 \in \mathcal{B}_1$ et $B_2 \in \mathcal{B}_2$, alors $\mathbf{P}_X \times \mathbf{P}_Y(B_1 \times B_2) = \mathbf{P}_X(B_1) \, \mathbf{P}_Y(B_2)$.

Regardons ce que cela donne concrètement pour les variables aléatoires discrètes.

Proposition 3.4.0.2. Soient X et Y deux variables aléatoires discrètes à valeurs dans $\{x_i;\ 0 \leqslant i < N\}$ et $\{y_j;\ 0 \leqslant j < M\}$. Les variables X et Y sont indépendantes si et seulement si, pour tout $0 \leqslant i < N$ et $0 \leqslant j < M$ on a

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

i.e. que $(X = x_i)$ et $(Y = y_j)$ sont des évènements indépendants.

Plus généralement, on peut avoir également une notion d'indépendance pour plus de deux variables aléatoires.

Définition 3.4.0.3. Soit $(X_n)_{n \geqslant 1}$ une suite de variables aléatoires définies sur le même espace de probabilité. On dit que ces variables aléatoires sont indépendantes si, pour tout $n \geqslant 1$, on a

$$\mathbf{P}_{(X_1, \ldots, X_n)} = \mathbf{P}_{X_1} \times \cdots \times \mathbf{P}_{X_n}.$$

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

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

Proposition 3.4.0.4. Soient X et Y deux variables aléatoires indépendantes. Si XY est intégrable, alors

$$\mathbf{E}[XY] = \mathbf{E}[X] \, \mathbf{E}[Y].$$

Plus généralement, soient $f \colon \mathbf{R} \to \mathbf{R}$ et $g \colon \mathbf{R} \to \mathbf{R}$ deux fonctions mesurables. Si $f(X)$ et $g(Y)$ sont intégrables, alors $\mathbf{E}[f(X)g(Y)] = \mathbf{E}[f(X)] \, \mathbf{E}[g(Y)]$.

Démonstration. Cas général : On a

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

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

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

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

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

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

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

□

Proposition 3.4.0.5. Soit $(X_n)_{n\in\mathbf{N}}$ une suite de variables aléatoires. Pour tout $n \geqslant 2$, $X_1, \ldots, X_n$ sont indépendantes si et seulement si pour toutes fonctions mesurables $h_1, \ldots, h_n : \mathbf{R} \to \mathbf{R}$ positives ou bien bornées, on a

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

Remarque 3.4.0.6. La proposition précédente implique que si $(X_n)_{n\in\mathbf{N}}$ est une suite de variables aléatoires indépendantes et si $(h_n)_{n\in\mathbf{N}}$ est une suite de fonctions mesurables réelles, alors
$(h_n(X_n))_{n\in\mathbf{N}}$ est une suite de variables aléatoires indépendantes.

Démonstration. • Le sens direct est évident.
• Réciproquement, $\mathbf{P}_{(X_1,\ldots,X_n)}$ est caractérisée par les $\mathbf{P}_{(X_1,\ldots,X_n)}(A_1 \times \cdots \times A_n)$. On prend $h_i(x) = \mathbf{1}_{A_i}(x)$ avec $A_i \in \mathcal{B}(\mathbf{R})$:

$$\mathbf{P}_{(X_1,\ldots,X_n)}(A_1 \times \cdots \times A_n) = \prod_{i=1}^{n} \mathbf{P}_{X_i}(A_i) = \mathbf{P}_{X_1} \otimes \cdots \otimes \mathbf{P}_{X_n}(A_1, \ldots, A_n).$$

□

Proposition 3.4.0.7. Si X et Y sont deux variables aléatoires indépendantes, de carré intégrable, alors

$$\mathbf{V}(X + Y) = \mathbf{V}(X) + \mathbf{V}(Y).$$

Démonstration. On a

$$\begin{aligned}\mathbf{V}(X + Y) &= \mathbf{E}[(X + Y)^2] - \mathbf{E}[X + Y]^2\\ &= \mathbf{V}(X) + \mathbf{V}(Y) + 2\,\mathbf{E}[XY] - 2\,\mathbf{E}[X]\,\mathbf{E}[Y]\\ &= \mathbf{V}(X) + \mathbf{V}(Y).\end{aligned}$$

□

3.5 Lois discrètes usuelles

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

Définition 3.5.1.1. Soit $n \in \mathbf{N}_{>0}$. On dit que X suit la *loi uniforme* sur $\{1,\dots,n\}$ si $X(\Omega) = \{1,\dots,n\}$ et

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

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

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

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

De plus,

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

donc

$$\mathbf{V}(X) = \frac{n^2-1}{12}.$$

3.5.2 Loi de Bernoulli

Définition 3.5.2.1. Soit $p \in [0,1]$. On dit que X suit la *loi de Bernoulli* de paramètre p si $X(\Omega) = \{0,1\}$ et

$$\mathbf{P}(X = 1) = p \qquad \mathbf{P}(X = 0) = 1 - p.$$

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

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

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

De plus

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

On a donc

$$\mathbf{V}(X) = p(1-p).$$

Remarquons au passage que X^2 suit la même loi que X !

Définition 3.5.2.2. Soit $p \in [0,1]$. On dit que X suit la *loi de Rademacher* si $X(\Omega) = \{-1,1\}$ et

$$\mathbf{P}(X = 1) = p \qquad \mathbf{P}(X = -1) = 1 - p.$$

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

Remarque 3.5.2.3. Si $X \sim \mathcal{B}(p)$, alors $Y = 2X - 1 \sim \mathcal{R}(p)$. En effet,

$$\mathbf{P}(Y = 1) = \mathbf{P}(2X - 1 = 1) = \mathbf{P}(X = 1) = p$$

et

$$\mathbf{P}(Y = -1) = \mathbf{P}(2X - 1 = -1) = \mathbf{P}(X = 0) = 1 - p.$$

De plus,

$$\mathbf{E}[Y] = \mathbf{E}[2X - 1] = 2\,\mathbf{E}[X] - 1 = 2p - 1$$

et

$$\mathbf{V}(Y) = \mathbf{V}(2X - 1) = \mathbf{V}(2X) = 4\,\mathbf{V}(X) = 4p(1-p).$$

3.5.3 Loi binomiale

Définition 3.5.3.1. Soient $p \in [0,1]$ et $n \in \mathbf{N}_{>0}$. 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 3.5.3.2. Il est clair que $\mathcal{B}(1,p) = \mathcal{B}(p)$.

Proposition 3.5.3.3. Soient $X_1, \dots, X_n$ n variables aléatoires indépendantes de loi $\mathcal{B}(p)$. On a alors

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

Démonstration. Posons $Y_n = \sum_{i=1}^{n} X_i$. Tout d'abord, il est clair que $Y_n(\Omega) = \{0,\dots,n\}$. Soit $j \in \{0,\dots,n\}$, on a

$$\begin{aligned}
\mathbf{P}(Y_n = j) &= \mathbf{P}\left(\sum_{k=1}^{n} X_k = j\right) \\
&= \binom{n}{j} \mathbf{P}\left(\bigcap_{\substack{J \subset [\![1,n]\!] \\ |J|=j}} \{X_i = 1\} \cap \bigcap_{i \in J} \{X_i = 0\}\right) \\
&= \binom{n}{j} \prod_{i \in J} \mathbf{P}(X_i = 1) \prod_{i \in J} \mathbf{P}(X_i = 0) \\
&= \binom{n}{j} p^j (1-p)^{n-j}
\end{aligned}$$

donc $Y_n \sim \mathcal{B}(n,p)$ (en utilisant la proposition ci-dessous). □

Proposition 3.5.3.4. Si $X \sim \mathcal{B}(n,p)$, alors $X(\Omega) = \{0,\dots,n\}$ et

$$\mathbf{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k} \quad (0 \leqslant k \leqslant n).$$

Remarque 3.5.3.5. La formule du binôme de Newton nous donne bien que

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

Démonstration. 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 (puisque $X_i \sim \mathcal{B}(p)$). Soit $k \in \{0,\dots,n\}$. L'évènement $(X = k)$ correspond à l'évènement « k succès et $n-k$ échecs ». Il y a $\binom{n}{k}$ 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

$$\mathbf{P}\left(\bigcap_{i \in I} (X_i = 1) \bigcap_{j \in J} (X_j = 0)\right) = \prod_{i \in I} \mathbf{P}(X_i = 1) \prod_{j \in J} \mathbf{P}(X_j = 0) = p^k (1-p)^{n-k}$$

par indépendance. □

Proposition 3.5.3.6. Si $p \in [0,1]$, $n \in \mathbf{N}_{>0}$ et $X \sim \mathcal{B}(n,p)$, alors

$$\mathbf{E}[X] = np \qquad \mathbf{V}(X) = np(1-p).$$

Démonstration. L'espérance et la variance d'une variable aléatoire 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

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

par linéarité et

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

par indépendance. □

3.5.4 Loi géométrique

Définition 3.5.4.1. 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 3.5.4.2. Si $X \sim \mathcal{G}(p)$, alors $X(\Omega) = \mathbf{N}_{>0}$ et

$$\mathbf{P}(X = k) = p(1-p)^{k-1} \quad (k \geqslant 1).$$

Démonstration. On considère $(X_i)_{i \in \mathbf{N}_{>0}}$ des variables aléatoires indépendantes de loi $\mathcal{B}(p)$: X_i représente le résultat de la i-ème expérience. On note N l'instant du premier succès. On a alors $N(\Omega) = \mathbf{N}_{>0} \cup \{+\infty\}$. Pour $k \in \mathbf{N}_{>0}$,

$$\begin{aligned}\mathbf{P}(N = k) &= \mathbf{P}(X_1 = 0, \dots, X_{k-1} = 0, X_k = 1) \\ &= \mathbf{P}(X_1 = 0) \cdots \mathbf{P}(X_{k-1} = 0)\, \mathbf{P}(X_k = 1) = (1-p)^{k-1} p\end{aligned}$$

par indépendance. De plus,

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

car $|1-p| < 1$, donc

$$\mathbf{P}(N = +\infty) = 1 - \sum_{k=1}^{+\infty} \mathbf{P}(N = k) = 0.$$

□

Remarque 3.5.4.3. Si $p = 1$, alors $\mathbf{P}(N = 1) = 1$. Si $p = 0$, alors $\mathbf{P}(N = +\infty) = 1$.

Proposition 3.5.4.4. On a

$$\mathbf{E}[X] = \frac{1}{p} \qquad \mathbf{V}(X) = \frac{1-p}{p^2}.$$

Démonstration. Commençons par l'espérance. On a

$$\mathbf{E}[X] = \sum_{k \geqslant 1} kp(1-p)^{k-1} = p \sum_{k \geqslant 1} k(1-p)^{k-1}.$$

On pose $f_k(p) = -(1-p)^k$, alors $f_k'(p) = k(1-p)^{k-1}$. Ainsi,

$$\begin{aligned}\mathbf{E}[X] &= p \sum_{k \geqslant 1} f_k'(p) = p\Big(\sum_{k \geqslant 1} f_k(p) \Big)' \\ &= p\left(\frac{-(1-p)}{p} \right)' = p\left(\frac{p-1}{p} \right)' = \frac{1}{p}.\end{aligned}$$

Calculons maintenant $\mathbf{E}[X^2]$:

$$\mathbf{E}[X^2] = \sum_{k \geqslant 1} k^2 p(1-p)^{k-1} = p(1-p) \sum_{k \geqslant 1} k^2 (1-p)^{k-2}.$$

Comme précédemment, on pose $g_k(p) = (1-p)^{k-1}$, donc $g_k'(p) = -k(1-p)^{k-1}$ et $g_k''(p) = k(k-1)(1-p)^{k-2} = k^2(1-p)^{k-2} - k(1-p)^{k-2}$. On a alors

$$\begin{aligned}\mathbf{E}[X^2] &= p(1-p) \sum_{k \geqslant 1} g_k''(p) + k(1-p)^{k-2} = p(1-p) \sum_{k \geqslant 1} g_k''(p) + p \sum_{k \geqslant 1} k(1-p)^{k-1} \\ &= p(1-p)\Big(\sum_{k \geqslant 1} g_k(p) \Big)'' + \frac{1}{p} = p(1-p)\left(\frac{1-p}{p} \right)'' = (1-p)\frac{2}{p^2} + \frac{1}{p}.\end{aligned}$$

Enfin,

$$\mathbf{V}(X) = (1-p)\frac{2}{p^2} + \frac{1}{p} - \frac{1}{p^2} = \frac{1-p}{p^2}.$$

□

3.5.5 Loi de Poisson

Définition 3.5.5.1. Soit $\lambda > 0$. On dit que X suit la *loi de Poisson* si $X(\Omega) = \mathbf{N}$ et si

$$\mathbf{P}(X = k) = e^{-\lambda}\frac{\lambda^k}{k!} \quad (\forall k \in \mathbf{N}).$$

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

Proposition 3.5.5.2. Si $X \sim \mathcal{P}(\lambda)$, alors

$$\mathbf{E}[X] = \lambda \qquad \mathbf{V}(X) = \lambda.$$

Démonstration. On a

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

La variance est laissé en exercice. □

3.6 Fonction de répartition

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

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

Néanmoins, la tribu $\mathcal{B}(\mathbf{R})$ est engendrée par les intervalles $]-\infty, t]$ pour tout $t \in \mathbf{R}$. Pour connaître $\mathbf{P}_X$, il suffit donc de connaître $\mathbf{P}(X \in]-\infty, t]) = \mathbf{P}(X \leqslant t)$ pour tout $t \in \mathbf{R}$.

Définition 3.6.0.1. Soit X une variable aléatoires réelle. On appelle *fonction de répartition* de X, l'application

$$\begin{aligned} F_X\colon \mathbf{R} &\to [0,1] \\ t &\mapsto \mathbf{P}(X \leqslant t) = \mathbf{P}_X(]-\infty, t]) \end{aligned}$$

Proposition 3.6.0.2. 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 variable aléatoire d'être dans un intervalle.

Proposition 3.6.0.3. On a :
(i) $\mathbf{P}(X \in]a, b]) = F_X(b) - F_X(a)$ car

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

(ii) $\mathbf{P}(X \in]a, b[) = \mathbf{P}\left(X \in \bigcup_{n\geqslant 1}]a, b - 1/n]\right) = \mathbf{P}\left(\bigcup_{n\geqslant 1}(X \in]a, b - 1/n])\right) = \lim_{n\to+\infty} F_X(b - 1/n) - F_X(a)$ car l'union est croissante.
(iii) $\mathbf{P}(X \in [a, b]) = \mathbf{P}\left(X \in \bigcap_{n\geqslant 1}]a - 1/n, b]\right) = \mathbf{P}\left(\bigcap_{n\geqslant 1}(X \in]a - 1/n, b])\right) = \lim_{n\to+\infty} F_X(b) - F_X(a - 1/n)$ car l'intersection est décroissante.
(iv) $\mathbf{P}(X \in [a, b[) = \mathbf{P}\left(X \in \bigcup_{n\geqslant 1} [a, b-1/n]\right) = \mathbf{P}\left(\bigcup_{n\geqslant 1}(X \in [a, b-1/n])\right) = \lim_{n\to+\infty} F_X(b-1/n) - \lim_{k\to+\infty} F_X(a-1/k)$ en utilisant le point (iii) et la croissance de l'union.

Si deux variables aléatoires ont même fonction de répartition, elles ont même loi !

Remarque 3.6.0.4. Si X est une variable aléatoire discrète à valeurs dans $E \subset \mathbf{R}$, alors X est une variable aléatoire réelle, on peut donc calculer la fonction de répartition.

Exemple 3.6.0.5. Si $X \sim \mathcal{B}(p)$, alors

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

Proposition 3.6.0.6. Si X est une variable aléatoire réelle de fonction de répartition F_X, alors on a :
(i) F_X est une fonction croissante.
(ii) $\lim_{t\to-\infty} F_X(t) = 0$ et $\lim_{t\to+\infty} F_X(t) = 1$.
(iii) F_X est continue à droite et possède une limite à gauche en tout point (« càdlàg ») : si $t_n \searrow t$ (*i.e.* décroît et tend vers t), alors $F_X(t_n) \to F_X(t)$, si $t_n \nearrow t$, alors $F_X(t_n)$ a une limite notée $F_X(t^-)$.

Démonstration. (i) Soit $s \leqslant t$, alors $]-\infty, s] \subset]-\infty, t]$, donc $F_X(s) \leqslant F_X(t)$.
(ii) Soit $t_n \searrow +\infty$, alors $\Omega = \bigcup_{n\in\mathbf{N}} (X \leqslant t_n)$. Comme l'union est croissante, on a

$$1 = \mathbf{P}(\Omega) = \lim_{n\to+\infty} \mathbf{P}(X \leqslant t_n) = \lim_{n\to+\infty} F_X(t_n).$$

Soit $t_n \nearrow -\infty$, alors $\varnothing = \bigcap_{n\in\mathbf{N}} (X \leqslant t_n)$. Comme l'intersection est décroissante, on a

$$0 = \mathbf{P}(\varnothing) = \lim_{n\to+\infty} \mathbf{P}(X \leqslant t_n) = \lim_{n\to+\infty} F_X(t_n).$$

(iii) Soit $t \in \mathbf{R}$ et $t_n \searrow t$, alors $\bigcap_{n\in\mathbf{N}} (X \leqslant t_n) = (X \leqslant t)$. On a donc $\lim_{n\to+\infty} F_X(t_n) = F_X(t)$.
Soit $t \in \mathbf{R}$ et $t_n \nearrow t$, alors $\bigcup_{n\in\mathbf{N}} (X \leqslant t_n) = (X < t)$. On a donc que $\lim_{n\to+\infty} F_X(t_n)$ existe et vaut $\mathbf{P}(X < t)$. □

Remarque 3.6.0.7. Soit $t \in \mathbf{R}$. On a $(X = t) \cup (X < t) = (X \leqslant t)$, donc $\mathbf{P}(X = t) = F_X(t) - F_X(t^-)$. En particulier, $\mathbf{P}(X = t) > 0$ si et seulement si F_X admet un saut en t de taille $\mathbf{P}(X = t)$.

Proposition 3.6.0.8. Si X une variable aléatoire discrète réelle, alors F_X est une fonction « en escalier ». Plus précisément, elle est croissante, continue à droite avec une limite à gauche, constante par morceaux et « saute » uniquement aux points $t \in X(\Omega)$, la valeur du saut valant $\mathbf{P}(X = t)$.

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

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

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

$$F_X(t) = \sum_{\substack{x_i \leqslant t \\ x_i \in X(\Omega)}} \mathbf{P}(X = x_i) = \sum_{\substack{x_i \leqslant x_k \\ x_i \in X(\Omega)}} \mathbf{P}(X = x_i) = F_X(x_k).$$

De plus,

$$\begin{aligned} F_X(x_j) = \sum_{\substack{x_i \leqslant x_j \\ x_i \in X(\Omega)}} \mathbf{P}(X = x_i) &= \Bigg(\sum_{\substack{x_i \leqslant x_k \\ x_i \in X(\Omega)}} \mathbf{P}(X = x_i) + \mathbf{P}(X = x_j) \Bigg) \\ &= F_X(x_k) + \mathbf{P}(X = x_k). \end{aligned}$$

□

Proposition 3.6.0.9. Soit $F\colon \mathbf{R} \to [0,1]$ une fonction croissante, continue à droite avec une limite à gauche, qui tend vers 0 en $-\infty$ et vers 1 en $+\infty$. Il existe alors une variable aléatoire de fonction de répartition F.

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

$$X(\omega) = \inf\{x;\ F(x) \geqslant \omega\}.$$

Remarquons que, comme $\lim_{x\to-\infty} F(x) = 0$ et $\lim_{x\to+\infty} F(x) = 1$, $X(\omega) \in \mathbf{R}$. 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 $t \in \mathbf{R}$,

$$\mathbf{P}(X \leqslant t) = \mathbf{P}(\{\omega;\ F^{-1}(\omega) \leqslant t\}) = \mathbf{P}(\{\omega;\ \omega \leqslant F(t)\}) = \mathbf{P}(]0, F(t)]) = F(t).$$

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

$$(X \leqslant t) = \{\omega;\ \omega \leqslant F(t)\}$$

ce qui permet de conclure par le même calcul que précédemment. On montre cette égalité par double inclusion. On fixe $t \in \mathbf{R}$.
• Si $\omega \leqslant F(t)$, alors $X(\omega) \leqslant t$ par définition de l'inf.
• Inversement, on suppose $X(\omega) \leqslant t$. Il existe donc une suite décroissante $(x_k)_{k\in\mathbf{N}}$ telle que $\lim_{k\to+\infty} x_k = X(\omega)$ et $F(x_k) \geqslant \omega$ pour tout $k \in \mathbf{N}$. Par continuité à droite de F, on a $\lim_{k\to+\infty} F(x_k) = F(X(\omega)) \geqslant \omega$. Comme F est croissante, on en déduit que $\omega \leqslant F(X(\omega)) \leqslant F(t)$. □

4 Variables aléatoires à densité

4.1 Définitions, propriétés

Soit X une variable aléatoire réelle.

Définition 4.1.0.1. On dit que X est une variable aléatoire à *densité* (ou *absolument continue*) si sa loi $\mathbf{P}_X$ est absolument continue par rapport à la mesure de Lebesgue, c'est-à-dire que : pour tout $A \in \mathcal{B}(\mathbf{R})$ tel que $\lambda(A) = 0$, on a $\mathbf{P}_X(A) = 0$.

Théorème 4.1.0.2 (Radon-Nikodym)**.** X est une variable aléatoire à densité si et seulement si il existe une fonction mesurable $f \geqslant 0$ telle que pour tout borélien $B \in \mathcal{B}(\mathbf{R})$ et donc pour tout intervalle I de $\mathbf{R}$, on a

$$\mathbf{P}(X \in B) = \int_B f(u)\,\mathrm{d}u \quad \text{et} \quad \mathbf{P}(X \in I) = \int_I f(u)\,\mathrm{d}u.$$

Cette fonction f est appelée *densité* de la variable X.

Proposition 4.1.0.3. X est une variable aléatoire à densité si et seulement si il existe une fonction mesurable $f\colon \mathbf{R} \to \mathbf{R}_+$ telle que la fonction de répartition de X s'écrit

$$F_X(t) = \int_{-\infty}^{t} f(u)\,\mathrm{d}u.$$

Démonstration. Il suffit de voir que X est caractérisé par F_X et que $\mathbf{P}_X$ est caractérisée par $\mathbf{P}_X(]-\infty, t]) = F_X(t)$. □

Remarque 4.1.0.4. Pour montrer que F_X s'écrit comme dans la proposition 4.1.0.3, il suffit d'avoir :
(1) F_X continue ;
(2) F_X est $\mathcal{C}^1$ 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 dans la proposition 4.1.0.3 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-presque partout (avec λ la mesure de Lebesgue). On a alors $\phi(1) - \phi(0) = 1$ et $\int_0^1 \phi'(u)\,\mathrm{d}u = 0\ldots$

Proposition 4.1.0.5. Soit X une variable à densité, de densité f.
(i) f caractérise la loi de X.
(ii) $\mathbf{P}(X = a) = \int_a^a f(u)\,\mathrm{d}u = 0$ pour tout $a \in \mathbf{R}$. En particulier, une variable aléatoire discrète n'est pas à densité et une variable aléatoire à densité n'est pas discrète !
(iii) Pour tout $a < b$, on a

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

(iv) On a

$$\int_{\mathbf{R}} f(u)\,\mathrm{d}u = \mathbf{P}(X \in \mathbf{R}) = 1.$$

Remarque 4.1.0.6. Si f est continue en x et $\varepsilon > 0$ « petit », alors

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

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

Proposition 4.1.0.7. Si $f\colon \mathbf{R} \to \mathbf{R}$ est une fonction mesurable telle que :
(i) $f \geqslant 0$ presque partout ;
(ii) $\int_{\mathbf{R}} f(x)\,\mathrm{d}x = 1$;
alors f est la densité d'une certaine variable aléatoire X.

Démonstration. Il suffit de remarquer que $t \mapsto \int_{-\infty}^t f(u)\,\mathrm{d}u$ est une fonction croissante, continue, qui tend vers 0 en $-\infty$ et 1 en $+\infty$, puis appliquer la proposition 3.6.0.9. □

4.2 Espérance et variance

Proposition 4.2.0.1. Soit X une variable aléatoire à densité, de densité f. La variable X est intégrable si et seulement si

$$\mathbf{E}[|X|] = \int_{\mathbf{R}} |x| f(x)\,\mathrm{d}x < +\infty.$$

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

$$\mathbf{E}[X] = \int_{\mathbf{R}} x f(x)\,\mathrm{d}x.$$

Proposition 4.2.0.2. Soit X une variable aléatoire à densité, de densité f et h une fonction mesurable réelle. On suppose que

$$\int_{\mathbf{R}} |h(x)| f(x)\,\mathrm{d}x < +\infty.$$

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

$$\mathbf{E}[Y] = \mathbf{E}[h(X)] = \int_{\mathbf{R}} h(x) f(x)\,\mathrm{d}x.$$

Remarque 4.2.0.3. Attention ! La variable aléatoire 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 variable aléatoire constante nulle, elle est donc discrète et ne peut pas être à densité.

Exemple 4.2.0.4. On a

$$\mathbf{E}[X^2] = \int_{\mathbf{R}} x^2 f(x)\,\mathrm{d}x.$$

Ainsi, sous réserve d'existence, on obtient

$$\mathbf{V}(X) = \int_{\mathbf{R}} x^2 f(x)\,\mathrm{d}x - \left(\int_{\mathbf{R}} x f(x)\,\mathrm{d}x \right)^2.$$

4.3 Indépendance

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

Proposition 4.3.0.1. X et Y sont indépendantes si et seulement si pour tous intervalles réels I et J, on a

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

4.4 Quelques lois usuelles

4.4.1 Loi uniforme

Définition 4.4.1.1. Soient $a, b \in \mathbf{R}$ avec $a < b$. La variable aléatoire X suit la *loi uniforme* sur l'intervalle $[a, b]$ si sa densité vaut

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

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

Proposition 4.4.1.2. Si $[c, d] \subset [a, b]$ alors $\mathbf{P}(X \in [c, d]) = \frac{d-c}{b-a}$.

Proposition 4.4.1.3. Si $X \sim \mathcal{U}([a, b])$, alors on a

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

Démonstration. Soit $t \in \mathbf{R}$. Par définition, on a

$$\begin{aligned} F_X(t) &= \int_{-\infty}^{t} f_X(x)\,\mathrm{d}x = \int_{-\infty}^{t} \frac{1}{b-a} \mathbf{1}_{[a,b]}(x)\,\mathrm{d}x \\ &= \begin{cases} 0 & \text{si } t \leqslant a \\ \int_a^t \frac{1}{b-a}\,\mathrm{d}x & \text{si } t \in [a, b] \\ 1 & \text{si } t \geqslant b \end{cases} = \begin{cases} 0 & \text{si } t \leqslant a \\ \frac{t-a}{b-a} & \text{si } t \in [a, b] \\ 1 & \text{si } t \geqslant b \end{cases} \end{aligned}$$

□

Proposition 4.4.1.4. Si $X \sim \mathcal{U}([a,b])$,

$$\mathbf{E}[X] = \frac{a+b}{2} \qquad \mathbf{V}(X) = \frac{(b-a)^2}{12}.$$

Démonstration. D'une part, on a

$$\begin{aligned}\mathbf{E}[X] &= \int_{\mathbf{R}} \frac{x}{b-a}\mathbf{1}_{[a,b]}(x)\,\mathrm{d}x = \frac{1}{b-a}\int_a^b x\,\mathrm{d}x \\ &= \frac{1}{b-a}\left[\frac{x^2}{2}\right]_a^b = \frac{1}{b-a}\frac{b^2-a^2}{2} = \frac{a+b}{2}.\end{aligned}$$

D'autre part, on a

$$\begin{aligned}\mathbf{E}[X^2] &= \int_{\mathbf{R}} \frac{x^2}{b-a}\mathbf{1}_{[a,b]}(x)\,\mathrm{d}x = \frac{1}{b-a}\int_a^b x^2\,\mathrm{d}x \\ &= \frac{1}{b-a}\left[\frac{x^3}{3}\right]_a^b = \frac{1}{b-a}\frac{(b-a)(a^2+ab+b^2)}{3} = \frac{a^2+ab+b^2}{3}.\end{aligned}$$

Enfin, on a

$$\begin{aligned}\mathbf{V}(X) &= \frac{a^2+ab+b^2}{3} - \frac{a^2+2ab+b^2}{4} = \frac{a^2-2ab+b^2}{12} \\ &= \frac{(b-a)^2}{12}.\end{aligned}$$

□

4.4.2 Loi exponentielle

Définition 4.4.2.1. Soient $\lambda > 0$. X suit la *loi exponentielle* de paramètre λ si

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

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

Proposition 4.4.2.2. Si $X \sim \mathcal{E}(\lambda)$, alors on a

$$F_X(t) = \begin{cases} 0 & \text{si} \quad t \leqslant 0 \\ 1 - e^{-\lambda t} & \text{si} \quad t \geqslant 0 \end{cases}$$

Démonstration. Par définition, pour $t \in \mathbf{R}$, on a

$$\begin{aligned}F_X(t) &= \int_{-\infty}^t f(x)\,\mathrm{d}x = \int_{-\infty}^t \lambda e^{-\lambda x}\mathbf{1}_{\mathbf{R}_+}(x)\,\mathrm{d}x \\ &= \lambda\int_0^t e^{-\lambda x}\,\mathrm{d}x = \lambda\left[-\frac{1}{\lambda}e^{-\lambda x}\right]_0^t \qquad (\text{si } t \geqslant 0) \\ &= 1 - e^{-\lambda t}.\end{aligned}$$

Si $t \leqslant 0$, alors il est clair que $F_X(t) = 0$. □

Proposition 4.4.2.3. Si $X \sim \mathcal{E}(\lambda)$,

$$\mathbf{E}[X] = \frac{1}{\lambda} \qquad \mathbf{V}(X) = \frac{1}{\lambda^2}.$$

Démonstration. D'une part, on a

$$\begin{aligned}\mathbf{E}[X] &= \int_{\mathbf{R}} xf(x)\,\mathrm{d}x = \int_{\mathbf{R}} x\lambda e^{-\lambda x}\mathbf{1}_{\mathbf{R}_+}(x)\,\mathrm{d}x \\ &= \lambda\int_0^{+\infty} xe^{-\lambda x}\,\mathrm{d}x = \lambda\Big(-\frac{1}{\lambda}\big[xe^{-\lambda x}\big]_0^{+\infty} + \frac{1}{\lambda}\int_0^{+\infty} e^{-\lambda x}\,\mathrm{d}x\Big) \\ &= \int_0^{+\infty} e^{-\lambda x}\,\mathrm{d}x = \frac{1}{\lambda}\big[-e^{-\lambda x}\big]_0^{+\infty} \\ &= \frac{1}{\lambda}.\end{aligned}$$

D'autre, part, on a

$$\begin{aligned}\mathbf{E}[X^2] &= \int_{\mathbf{R}} x^2 f(x)\,\mathrm{d}x = \lambda \int_0^{+\infty} x^2 e^{-\lambda x}\,\mathrm{d}x \\ &= \lambda\Big(-\frac{1}{\lambda}\big[x^2 e^{-\lambda x}\big]_0^{+\infty} + \frac{2}{\lambda}\underbrace{\int_0^{+\infty} x e^{-\lambda x}\,\mathrm{d}x}_{=1/\lambda^2}\Big) = \frac{2}{\lambda^2}\end{aligned}$$

et ainsi,

$$\mathbf{V}(X) = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}.$$

□

4.4.3 Loi de Cauchy

Définition 4.4.3.1. 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 4.4.3.2. Si $X \sim \mathcal{C}(c)$, alors on a

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

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

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

4.4.4 Loi normale centrée réduite

Définition 4.4.4.1. 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 4.4.4.2. Si $X \sim \mathcal{N}(0,1)$, on a

$$\mathbf{E}[X] = 0 \qquad \mathbf{V}(X) = 1.$$

4.5 Quelques exemples de calculs de lois

On est souvent confronté au problème suivant : soient X une variable aléatoire dont on connaît la loi et $h\colon \mathbf{R} \to \mathbf{R}$ 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 $t \geqslant 1$. Si $t \in [0,1]$, alors

$$F_Y(t) = \mathbf{P}(Y \leqslant t) = \mathbf{P}(X^2 \leqslant t) = \mathbf{P}(-\sqrt{t} \leqslant X \leqslant \sqrt{t}) = \mathbf{P}(X \leqslant \sqrt{t}) = F_X(\sqrt{t}).$$

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 $\mathcal{C}^0$ sur $\mathbf{R}$ et $\mathcal{C}^1$ sur $\mathbf{R}\backslash\{0,1\}$. Ainsi, Y est une variable aléatoire à densité, de densité g donnée par

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

4.5.2 Exemple 2

On considère $X \sim \mathcal{E}(\lambda)$ et $Y \sim \mathcal{E}(\mu)$ deux variables aléatoires indépendantes. On pose $Z = \min(X, Y)$. L'application Z est une variable aléatoire à valeurs dans $\mathbf{R}^+$. En particulier, $\mathbf{P}(Z > t) = 1$ si $t < 0$. Si $t \geqslant 0$, on a

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

par indépendance de X et Y. Cependant,

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

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

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

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

4.5.3 Exemple 3

Soit X une variable aléatoire à densité f continue par morceaux. Soit $m \in \mathbf{R}$ et $\sigma > 0$ deux paramètres. On pose $Y = m + \sigma X$. On a alors

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

F_X est continue sur $\mathbf{R}$ et dérivable par morceaux, donc Y est une variable aléatoires 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 4.5.3.1. Si $X \sim \mathcal{N}(0,1)$, $m \in \mathbf{R}$, $\sigma > 0$ et $Y := m + \sigma X$, alors Y est une variable aléatoire 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 4.5.3.2. Si $X \sim \mathcal{N}(m, \sigma^2)$ avec $m \in \mathbf{R}$ et $\sigma > 0$, alors

$$\mathbf{E}[X] = m \qquad \mathbf{V}(X) = \sigma^2.$$

5 Couples et vecteurs aléatoires : cas discret

Définition 5.0.0.1. $(X_1, \dots, X_n)$ est un *vecteur aléatoire* défini sur $(\Omega, \mathcal{A}, \mathbf{P})$ si chaque X_i est une variable aléatoire définie sur $(\Omega, \mathcal{A}, \mathbf{P})$. Lorsque $n = 2$ on parle de *couple de variables aléatoires*.

Proposition-Définition 5.0.0.2. On dit que $(X_1, \dots, X_n)$ est un *vecteur discret* si $E_1 \times \cdots \times E_n$ est un ensemble discret. En particulier, $(X_1, \dots, X_n)$ est discret si et seulement si $X_1, \dots, X_n$ sont des variables aléatoires 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 $\mathbf{P}(X_1 = x_1, \dots, X_n = x_n)$, $x_1 \in E_1, \dots, x_n \in E_n$. On parle parfois de *loi jointe* pour la loi du vecteur $(X_1, \dots, X_n)$.

Exemple 5.1.0.1. 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

$$X = \max(X_1, X_2) \quad \text{et} \quad Y = \min(X_1, X_2).$$

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

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

Ensuite, nous pouvons calculer les probabilités :

$$\begin{aligned}
&\mathbf{P}(X = 0, Y = 0) = \mathbf{P}(X_1 = 0, X_2 = 0) = \mathbf{P}(X_1 = 0)\,\mathbf{P}(X_2 = 0) = (1-p)^2 = 9/16\\
&\mathbf{P}(X = 0, Y = 1) = 0,\\
&\mathbf{P}(X = 1, Y = 0) = \mathbf{P}(X_1 = 1, X_2 = 0) + \mathbf{P}(X_1 = 0, X_2 = 1) = 2p(1-p) = 6/16\\
&\mathbf{P}(X = 1, Y = 1) = \mathbf{P}(X_1 = 1)\,\mathbf{P}(X_2 = 1) = p^2 = 1/16
\end{aligned}$$

Remarquons que nécessairement, la somme des probabilités dois faire 1.

Plus généralement, on a la propriété suivante.

Proposition 5.1.0.2. Soit $(X_1, \dots, X_n)$ un vecteur aléatoire discret, alors on a

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

5.2 Lois marginales

Définition 5.2.0.1. On appelle *lois marginales* d'un vecteur aléatoire $(X_1, \dots, X_n)$, les lois des variables aléatoires $X_1, \dots, X_n$.

Si $n = 2$: on considère deux variables aléatoires discrètes

$$X : \Omega \to E_1 = \{x_i;\ 0 \leqslant i < M\} \quad \text{et} \quad Y : \Omega \to E_2 = \{y_j;\ 0 \leqslant j < N\}.$$

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

$$\begin{aligned}
\mathbf{P}(X = x_i) &= \mathbf{P}\left((X = x_i) \cap \left(\bigcup_{y_j \in E_2} (Y = y_j)\right)\right) = \mathbf{P}\left(\bigcup_{y_j \in E_2} (X = x_i, Y = y_j)\right)\\
&= \sum_{y_j \in E_2} \mathbf{P}(X = x_i, Y = y_j).
\end{aligned}$$

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

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

Dans le cas général :

Proposition 5.2.0.2. Soient $X_1, \dots, X_n$ des variables aléatoires discrètes. On note

$$E_i = \{x_k;\ 0 \leqslant k < N_i\} \quad (1 \leqslant i \leqslant n).$$

On a alors

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

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.

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

En particulier, $X \sim \mathcal{B}(7/16)$ et $Y \sim \mathcal{B}(1/16)$.

Remarque 5.2.0.3. (1) Les lois marginales d'un vecteur aléatoire ne permettent pas de déterminer la loi jointe de ce vecteur.
(2) 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.

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

5.3 Covariance

Définition 5.3.0.1. Soient X et Y deux variables aléatoires de carré intégrable (pas nécessairement discrètes). La *covariance* de X et de Y est donnée par

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

Proposition 5.3.0.2. (i) $\mathsf{Cov}(X, X) = \mathbf{V}(X)$.
(ii) $\mathsf{Cov}(X, Y) = \mathsf{Cov}(Y, X)$.
(iii) $\mathsf{Cov}(\alpha X_1 + \beta X_2, Y) = \alpha\,\mathsf{Cov}(X_1, Y) + \beta\,\mathsf{Cov}(X_2, Y)$.
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, $\mathsf{Cov}(X, X) = \mathbf{V}(X) = 0$ si et seulement si X est une variable aléatoire 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 5.3.0.3. On a

$$|\,\mathsf{Cov}(X, Y)| \leqslant \sqrt{\mathbf{V}(X)}\sqrt{\mathbf{V}(Y)}.$$

Démonstration. On applique juste l'inégalité de Cauchy-Schwarz :

$$\begin{aligned}
|\,\mathbf{E}[(X - \mathbf{E}[X])(Y - \mathbf{E}[Y])]| &\leqslant \mathbf{E}[|X - \mathbf{E}[X]| \cdot |Y - \mathbf{E}[Y]|] \\
&\leqslant \mathbf{E}[(X - \mathbf{E}[X])^2]^{1/2}\,\mathbf{E}[(Y - \mathbf{E}[Y])^2]^{1/2}.
\end{aligned}$$

□

Définition 5.3.0.4. On peut définir le *coefficient de corrélation* de X et Y par

$$r(X, Y) = \frac{\mathsf{Cov}(X, Y)}{\sqrt{\mathbf{V}(X)\,\mathbf{V}(Y)}} \in [-1, 1]$$

si $\mathbf{V}(X)\,\mathbf{V}(Y) > 0$.

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-Schwarz, ce qui signifie que $X - \mathbf{E}[X]$ et $Y - \mathbf{E}[Y]$ 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 5.3.0.5. On a

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

Démonstration.

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

□

Proposition 5.3.0.6. Soient X et Y deux variables aléatoires discrètes, alors

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

Démonstration. Il suffit de voir $\mathbf{E}[XY]$ comme $\mathbf{E}[h(X,Y)]$ avec $h : (x,y) \mapsto xy$ et d'appliquer le théorème de transfert. □

Exemple 5.3.0.7. Calculons la covariance des variables X et Y du premier exemple de ce chapitre.

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

et donc

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

Proposition 5.3.0.8. On a

$$\mathbf{V}(X+Y) = \mathbf{V}(X) + \mathbf{V}(Y) + 2\,\mathsf{Cov}(X,Y).$$

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

Démonstration.

$$\begin{aligned}\mathbf{V}(X+Y) &= \mathbf{E}[(X+Y)^2] - \mathbf{E}[X+Y]^2\\ &= \mathbf{E}[X^2+Y^2+2XY] - (\mathbf{E}[X]^2 + \mathbf{E}[Y]^2 + 2\,\mathbf{E}[X]\,\mathbf{E}[Y])\\ &= \mathbf{E}[X^2] - \mathbf{E}[X]^2 + \mathbf{E}[Y^2] - \mathbf{E}[Y]^2 + 2(\mathbf{E}[XY] - \mathbf{E}[X]\,\mathbf{E}[Y]).\end{aligned}$$

□

Remarque 5.3.0.9. La réciproque est fausse : si X et Y sont non corrélées, elles ne sont pas nécessairement indépendantes. Prenons par exemple, $X \sim \mathcal{U}(\{-1,0,1\})$ et $Y = \mathbf{1}_{\{X=0\}}$. On a alors $\mathbf{E}[X] = 0$, $Y \sim \mathcal{B}(1/3)$, $\mathbf{E}[Y] = 1/3$, $XY = 0$, $\mathbf{E}[XY] = 0$ et $\mathsf{Cov}(XY) = 0$. Par contre, X et Y ne sont pas indépendantes : en effet on a

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

5.4 Variables aléatoires à valeurs dans N

Définition 5.4.0.1. Soit X une variable aléatoire à valeurs dans $\mathbf{N}$. On définit sa *fonction génératrice* (ou *série génératrice*) par

$$G_X(s) = \mathbf{E}[s^X] = \sum_{k\geqslant 0} s^k\,\mathbf{P}(X=k).$$

Proposition 5.4.0.2. (i) G_X est une série entière ;
(ii) $G_X(1) = 1$, donc en particulier son rayon de convergence R est supérieur ou égal à 1 ;
(iii) On a

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

La proposition suivante découle directement du troisième point précédent.

Proposition 5.4.0.3. La série génératrice d'une variable aléatoire X à valeurs dans $\mathbf{N}$ caractérise sa loi.

Proposition 5.4.0.4. Si X et Y sont deux variables aléatoires indépendantes à valeurs dans $\mathbf{N}$, alors $X+Y$ est une variables aléatoires à valeurs dans $\mathbf{N}$ et

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

Démonstration. Par indépendance, on a

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

□

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

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

Exercice 5.4.0.5. Soient $X \sim \mathcal{P}(\lambda)$ et $Y \sim \mathcal{P}(\mu)$ indépendantes. Montrer que $X+Y \sim \mathcal{P}(\lambda+\mu)$.

Proposition 5.4.0.6. Si $|s| < R$, alors on peut dériver termes à termes pour obtenir

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

En particulier, si $R > 1$, $G'(1) = \mathbf{E}[X] < +\infty$. Inversement, si $\mathbf{E}[X] < +\infty$, le théorème radial d'Abel nous donne $G'_X(s) \to \mathbf{E}[X]$ pour $s \to 1^-$.
De même, si $\mathbf{E}[X^2] < +\infty$,

$$\begin{aligned} G''_X(s) &= \sum_{k\geqslant 0} k(k-1)\,\mathbf{P}(X=k)s^{k-2} = \sum_{k\geqslant 0} k^2\,\mathbf{P}(X=k)s^{k-2} - \sum_{k\geqslant 0} k\,\mathbf{P}(X=k)s^{k-2} \\ &\to \mathbf{E}[X^2] - \mathbf{E}[X] \end{aligned}$$

lorsque $s \to 1^-$.

Exercice 5.4.0.7. 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 $(\Omega, \mathcal{A}, \mathbf{P})$.

Proposition 6.0.0.1 (Caractérisation 1 : fonction de répartition)**.** La fonction de répartition de X caractérise sa loi.

Proposition 6.0.0.2 (Caractérisation 2 : loi discrète)**.** Si X est une variable aléatoire discrète, sa loi est caractérisé par la donnée de $X(\Omega)$ et $\mathbf{P}(X = x)$ pour tous les $x \in X(\Omega)$.

Proposition 6.0.0.3 (Caractérisation 3 : méthode de la fonction muette)**.** La connaissance de $\mathbf{E}[h(X)]$ 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.

Démonstration. Pour tout $t \in \mathbf{R}$, on considère la suite de fonctions $(h_{t,n})_{n\in\mathbf{N}_{>0}}$ 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 $n \in \mathbf{N}$. De plus, $(h_{t,n})_{n\in\mathbf{N}_{>0}}$ converge simplement, en décroissant, vers la fonction $\mathbf{1}_{\{x\leqslant t\}}$. Ainsi, le théorème de convergence monotone nous donne

$$\mathbf{E}[h_{t,n}(X)] \xrightarrow{n\to+\infty} \mathbf{E}[\mathbf{1}_{\{X\leqslant t\}}] = \mathbf{P}(X \leqslant t) = F_X(t).$$

En particulier, si on connaît $\mathbf{E}[h_{t,n}(X)]$ pour tous $n \in \mathbf{N}_{>0}$ et tous $t \in \mathbf{R}$, alors on connaît la fonction de répartition de X, qui caractérise la loi de X. □

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

Proposition 6.0.0.4 (Méthode de la fonction muette, cas pratiques)**.** (i) S'il existe des réels $\{x_i;\ 0 \leqslant i < N\}$ et $\{\alpha_i;\ 0 \leqslant i < N\}$ tels que pour toute fonction h continue bornée, on a

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

alors X est une variable aléatoire discrète, telle que $X(\Omega) = \{x_i;\ 0 \leqslant i < N\}$ et $\mathbf{P}(X = x_i) = \alpha_i$, pour tout $0 \leqslant i < N$.
(ii) S'il existe une fonction réelle mesurable g telle que pour toute fonction h continue bornée on a

$$\mathbf{E}[h(X)] = \int_{\mathbf{R}} h(x)g(x)\,\mathrm{d}x$$

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 variable aléatoire.

Définition 6.0.0.5. On définit la *fonction caractéristique* de la variable aléatoire réelle X par

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

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

Proposition 6.0.0.6. (i) Si X est discrète, alors

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

(ii) Si X est une variable à densité, de densité f, alors

$$\varphi_X(t) = \int_{\mathbf{R}} e^{itx} f(x)\,\mathrm{d}x = \widehat{f}(-t)$$

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

Proposition 6.0.0.7 (Caractérisation 4 : fonction caractéristique)**.** La fonction caractéristique de X caractérise sa loi.

La preuve est admise, mais remarquons que pour les variables aléatoires 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 variables aléatoires indépendantes.

Proposition 6.0.0.8. (i) Si $\mathbf{E}[|X|^n] < +\infty$, alors φ_X est dérivable n fois et $\varphi^{(n)}(0) = (i)^n \mathbf{E}[X^n]$.
(ii) Si X et Y sont indépendantes, alors $\varphi_{X+Y}(t) = \varphi_X(t)\varphi_Y(t)$ pour tout $t \in \mathbf{R}$.

Démonstration. • 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 $\mathcal{C}^n$ pour tout $x \in \mathbf{R}$. De plus, pour tout $1 \leqslant k \leqslant n$, on a

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

donc

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

qui est intégrable : en effet, $\mathbf{E}[|X|^n] < +\infty$, donc $\mathbf{E}[|X|^k] < +\infty$ pour tout $1 \leqslant k \leqslant n$ en utilisant l'inégalité de Hölder par exemple.
• Pour le second point, on a en utilisant l'hypothèse d'indépendance :

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

□

À 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 6.0.0.9. Si X est une variable aléatoire bornée, *i.e.* il existe une constante M tel que $|X| \leqslant M$, 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 de même loi et telles que $\mathbf{E}[X^k] = \mathbf{E}[Y^k]$ pour tout $k \in \mathbf{N}$. On renvoie le lecteur assidu à un contre-exemple du TD.

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

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

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

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

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

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

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). □

7 Couples et vecteurs aléatoires à densité

7.1 Vecteurs à densité

On considère $X = (X_1, \dots, X_n) \colon \Omega \to \mathbf{R}^n$ un vecteur aléatoire.

Définition 7.1.0.1. On dit que X est un vecteur aléatoire à *densité* si sa loi est absolument continue par rapport à la mesure de Lebesgue sur $\mathbf{R}^n$.

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-Définition 7.1.0.2. $X = (X_1, \dots, X_n)$ admet une densité si et seulement s'il existe une fonction $f \colon \mathbf{R}^n \to \mathbf{R}$ mesurable positive telle que pour tout $A \in \mathcal{B}(\mathbf{R}^n)$,

$$\mathbf{P}(X \in A) = \mathbf{P}((X_1, \dots, X_n) \in A) = \int_A f(x_1, \dots, x_n)\, \mathrm{d}x_1 \cdots \mathrm{d}x_n.$$

La fonction f est appelée densité de X.

Remarque 7.1.0.3. (1) Il suffit de le vérifier pour tout pavé $A = I_1 \times \cdots \times I_n$ avec $I_1, \dots, I_n$ des intervalles de $\mathbf{R}$. On peut même se restreindre à des intervalles de la forme $I_j =]-\infty, t_j]$.
(2) On a nécessairement $\int_{\mathbf{R}^n} f(x_1, \dots, x_n)\, \mathrm{d}x_1 \cdots \mathrm{d}x_n = 1$. Inversement, on peut montrer que si $f \colon \mathbf{R}^n \to \mathbf{R}^+$ est une fonction mesurable d'intégrale 1 sur $\mathbf{R}^n$, alors c'est une densité pour un certain vecteur aléatoire.

Proposition 7.1.0.4. $(X_1, \dots, X_n)$ 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

$$\mathbf{E}[h(X_1, \dots, X_n)] = \int_{\mathbf{R}^n} h(x_1, \dots, x_n) f(x_1, \dots, x_n)\, \mathrm{d}x_1 \cdots \mathrm{d}x_n.$$

Démonstration. 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. □

7.2 Lois marginales

Proposition 7.2.0.1. Si $(X_1, \dots, X_n)$ est un vecteur aléatoire à densité, de densité f, alors pour tout $1 \leqslant i \leqslant n$, X_i est une variable aléatoire à densité, de densité f_{X_i} donnée par

$$f_{X_i}(x) = \underbrace{\int_{\mathbf{R}} \cdots \int_{\mathbf{R}}}_{n-1 \text{ intégrales}} f(x_1, \dots, x_{i-1}, x, x_{i+1}, \dots, x_n)\, \mathrm{d}x_1 \cdots \mathrm{d}x_{i-1}\, \mathrm{d}x_{i+1} \cdots \mathrm{d}x_n \quad \text{avec} \quad x \in \mathbf{R}\,.$$

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

$$f_X(x) = \int_{\mathbf{R}} f(x, y)\, \mathrm{d}y \qquad f_Y(y) = \int_{\mathbf{R}} f(x, y)\, \mathrm{d}x.$$

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

$$\begin{aligned} F_{X_i}(x) &= \mathbf{P}(X_i \leqslant x) = \mathbf{P}((X_1, \dots, X_n) \in \mathbf{R} \times \cdots \times\,]-\infty, x] \times \cdots \times \mathbf{R}) \\ &= \int_{x_1 \in \mathbf{R}} \cdots \int_{x_i \leqslant x} \cdots \int_{x_n \in \mathbf{R}} f(x_1, \dots, x_n)\ \mathrm{d}x_1 \cdots \mathrm{d}x_n \\ &= \int_{-\infty}^{x} \left(\int_{x_1 \in \mathbf{R}} \cdots \int_{x_n \in \mathbf{R}} f(x_1, \dots, x_n)\ \mathrm{d}x_1 \cdots \mathrm{d}x_{i-1}\, \mathrm{d}x_{i+1} \cdots \mathrm{d}x_n \right) \mathrm{d}x_i \end{aligned}$$

par Fubini. La variable X_i est donc à densité et on obtient la formule annoncée pour sa densité. □

Remarque 7.2.0.2. 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 $\mathbf{R}^2$ donc ne peut pas être à densité, même si X est à densité.

Exemple 7.2.0.3. Soit (X,Y) un couple aléatoire de loi uniforme sur le disque unité

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

La densité de (X,Y) est donnée par

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

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

$$f_X(x) = \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} \frac{1}{\pi}\,\mathrm{d}y = \frac{2}{\pi}\sqrt{1-x^2}$$

Finalement, on a

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

et par symétrie,

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

En particulier, X et Y ont même loi.

Proposition 7.2.0.4. Soit (X,Y) un couple à densité f. Sous réserve d'existence, on a : (i) $\mathbf{E}[X] = \int_{\mathbf{R}} x f_X(x)\,\mathrm{d}x = \int_{\mathbf{R}^2} x f(x,y)\,\mathrm{d}x\,\mathrm{d}y$;
(ii) $\mathbf{E}[Y] = \int_{\mathbf{R}} y f_Y(y)\,\mathrm{d}y = \int_{\mathbf{R}^2} y f(x,y)\,\mathrm{d}x\,\mathrm{d}y$;
(iii) $\mathbf{E}[h(X,Y)] = \int_{\mathbf{R}^2} h(x,y) f(x,y)\,\mathrm{d}x\,\mathrm{d}y$. En particulier, pour le calcul de la covariance, on a $\mathbf{E}[XY] = \int_{\mathbf{R}^2} xy f(x,y)\,\mathrm{d}x\,\mathrm{d}y$.

7.3 Indépendance

Proposition 7.3.0.1. Soit $(X_1,\ldots,X_n)$ un vecteur aléatoire de densité $f_{(X_1,\ldots,X_n)}$ et de densités marginales $f_{X_1},\ldots,f_{X_n}$. Les variables aléatoires $X_1,\ldots,X_n$ sont indépendantes si et seulement si

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

Démonstration. 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 $\mathbf{R}$.
• Si X et Y sont indépendantes, alors

$$\begin{aligned}\mathbf{P}((X,Y)\in I\times J) &= \mathbf{P}(X\in I, Y\in J) = \mathbf{P}(X\in I)\,\mathbf{P}(Y\in J)\\ &= \int_I f_X(x)\,\mathrm{d}x \int_J f_Y(y)\,\mathrm{d}y = \int\int_{I\times J} f_X(x) f_Y(y)\,\mathrm{d}x\,\mathrm{d}y\end{aligned}$$

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 $f_{(X,Y)} = f_X f_Y$.
• Réciproquement, si $f_{(X,Y)} = f_X f_Y$, alors

$$\begin{aligned}\mathbf{P}((X,Y)\in I\times J) &= \int\int_{I\times J} f_X(x) f_Y(y)\,\mathrm{d}x\,\mathrm{d}y = \int_I f_X(x)\,\mathrm{d}x \int_J f_Y(y)\,\mathrm{d}y\\ &= \mathbf{P}(X\in I)\,\mathbf{P}(Y\in J)\end{aligned}$$

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 4.3.0.1. □

Remarque 7.3.0.2. (Importante). Pour montrer que $X_1,\ldots,X_n$ sont indépendantes, il suffit de montrer que $f_{(X_1,\ldots,X_n)}(x_1,\ldots,x_n)$ est à variables séparables, c'est-à-dire qu'il existe des fonctions mesurables positives $q_1,\ldots,q_n$ telles que

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

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

$$\begin{aligned}f_{X_i}(x_i) &= \int_{\mathbf{R}^{n-1}} f_{(X_1,\ldots,X_n)}(x_1,\ldots,x_n)\,\mathrm{d}x_1\cdots\mathrm{d}x_{i-1}\,\mathrm{d}x_{i+1}\cdots\mathrm{d}x_n\\ &= \left(\prod_{j\neq i}\int_{\mathbf{R}} q_j(x_j)\ \mathrm{d}x_j\right) q_i(x_i) := C_i q_i(x_i)\end{aligned}$$

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

$$1 = \int_{\mathbf{R}} f_{X_i}(x_i)\,\mathrm{d}x_i = C_i \int_{\mathbf{R}} q_i(x_i)\,\mathrm{d}x_i$$

donc $C_i = (\int_{\mathbf{R}} q_i(x_i)\,\mathrm{d}x_i)^{-1}$ 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 $U \subset \mathbf{R}^n$. Soient $V \subset \mathbf{R}^n$ un ouvert et $\phi\colon U \to V$ 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 cas 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 $h\colon \mathbf{R}^n \to \mathbf{R}$ une fonction continue bornée quelconque. D'après le théorème de transfert, on a alors

$$\begin{aligned}\mathbf{E}[h(Y_1, \dots, Y_n)] &= \mathbf{E}[h(\phi(X_1, \dots, X_n))] \\ &= \int_U h(\phi(x_1, \dots, x_n)) f(x_1, \dots, x_n)\,\mathrm{d}x_1 \cdots \mathrm{d}x_n\end{aligned}$$

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

$$\mathbf{E}[h(Y_1, \dots, Y_n)] = \int_V h(y_1, \dots, y_n) g(y_1, \dots, y_n)\,\mathrm{d}y_1 \cdots \mathrm{d}y_n.$$

Pour passer de $\int_U$ à $\int_V$ ci-dessus, il est naturel d'appliquer un changement de variable dans l'intégrale.

Théorème 7.4.1.1 (Rappel)**.** Soit $\phi\colon U \to V$ un C^1-difféomorphisme entre deux ouvert de $\mathbf{R}^n$ (ϕ est une bijection C^1 et ϕ^{-1} est également C^1). Sous réserve d'intégrabilité, on a

$$\int_U h(\phi(x_1, \dots, x_n))\,\mathrm{d}x_1 \cdots \mathrm{d}x_n = \int_V h(y_1, \dots, y_n) |\operatorname{Jac}(\phi^{-1}(y_1, \dots, y_n))|\,\mathrm{d}y_1 \cdots \mathrm{d}y_n$$

avec

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

Remarque 7.4.1.2. Si $\phi\colon U \to V$ est un difféomorphisme et X est à valeurs dans U de densité f_X, alors $\phi(V)$ est à valeurs dans V et a pour densité

$$f_Y(y) = \mathbf{1}_V(y) f_X(\phi^{-1}(y)) |\operatorname{Jac}(\varphi^{-1}(y))|$$

avec $y = (y_1, \dots, y_n)$.

Démonstration. Il suffit d'appliquer le théorème précédent dans

$$\int_U h(\phi(x_1, \dots, x_n)) f(x_1, \dots, x_n)\,\mathrm{d}x_1 \cdots \mathrm{d}x_n.$$

□

Appliquons tout cela au problème suivant. Soient X et Y deux variables aléatoires indépendantes de loi $\mathcal{N}(0,1)$, quelle est la loi de X/Y ?

Commençons par remarquer que X/Y est une variable aléatoire bien définie car $\mathbf{P}(Y = 0) = 0$, Y étant une variable aléatoire 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 $h\colon \mathbf{R}^2 \to \mathbf{R}$ une fonction continue bornée quelconque. On a alors

$$\mathbf{E}[h(U, V)] = \mathbf{E}[h(X/Y, Y)] = \int\int_{\mathbf{R} \times \mathbf{R}^*} h(x/y, y) \frac{1}{2\pi} e^{-\frac{x^2+y^2}{2}}\,\mathrm{d}x\,\mathrm{d}y.$$

On pose

$$\begin{aligned}\varphi\colon \mathbf{R}\times\mathbf{R}^* &\to \mathbf{R}\times\mathbf{R}^*\\ (x,y) &\mapsto (x/y,y)=(u,v)\end{aligned}$$

On a

$$\left\{\begin{array}{ll} u &= x/y\\ v &= y\end{array}\right. \Leftrightarrow \left\{\begin{array}{ll} x &= uv\\ y &= v\end{array}\right.$$

donc

$$\begin{aligned}\varphi^{-1}\colon \mathbf{R}\times\mathbf{R}^* &\to \mathbf{R}\times\mathbf{R}^*\\ (u,v) &\mapsto (uv,v)\end{aligned}$$

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

$$\mathrm{Jac}(\varphi^{-1}(u,v)) = \left|\begin{smallmatrix} v & u\\ 0 & 1\end{smallmatrix}\right| = v$$

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

$$\begin{aligned}\mathbf{E}[h(U,V)] &= \int\int_{\mathbf{R}\times\mathbf{R}^*} h(u,v)|v|\frac{1}{2\pi}e^{-\frac{u^2v^2+v^2}{2}}\,\mathrm{d}u\,\mathrm{d}v\\ &= \int\int_{\mathbf{R}^2} h(u,v)|v|\frac{1}{2\pi}e^{-\frac{u^2v^2+v^2}{2}}\,\mathrm{d}u\,\mathrm{d}v.\end{aligned}$$

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

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

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

$$\begin{aligned}f_U(u) &= \int_{\mathbf{R}} |v|\frac{1}{2\pi}e^{-\frac{u^2v^2+v^2}{2}}\,\mathrm{d}v = \frac{1}{\pi}\int_0^{+\infty} ve^{-\frac{v^2(1+u^2)}{2}}\,\mathrm{d}v\\ &= \frac{1}{\pi}\left[\frac{e^{-\frac{v^2(1+u^2)}{2}}}{1+u^2}\right]_0^{+\infty} = \frac{1}{\pi}\frac{1}{1+u^2}.\end{aligned}$$

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

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

Proposition 7.4.2.1. Soient X et Y deux variables aléatoires indépendantes, à densité, de densités respectives f_X et f_Y. $X+Y$ est alors une variable aléatoire à densité, de densité

$$f_{X+Y}(t) = f_X * f_Y(t) = \int_{\mathbf{R}} f_X(x)f_Y(t-x)\,\mathrm{d}x = \int_{\mathbf{R}} f_X(t-y)f_Y(y)\,\mathrm{d}y.$$

L'opération $*$ est appelé *produit de convolution.*

Remarque 7.4.2.2. Si X et Y sont positives, alors

$$f_{X+Y}(t) = \int_0^t f_X(s)f_Y(t-s)\,\mathrm{d}s\mathbf{1}_{\mathbf{R}^+}(t).$$

On peut faire le parallèle avec la formule obtenue lorsque l'on considère la somme de deux variables aléatoires discrètes à valeurs dans $\mathbf{N}$:

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

Démonstration. (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 $t\in\mathbf{R}$,

$$F_{X+Y}(t) = \mathbf{P}(X+Y\leqslant t) = \mathbf{P}((X,Y)\in A_t) = \int\int_{A_t} f_X(x)f_Y(y)\,\mathrm{d}x\,\mathrm{d}y$$

avec $A_t = \{(x,y) \in \mathbf{R}^2;\ x+y \leqslant t\}$.

$$\begin{aligned} F_{X+Y}(t) &= \int_{\mathbf{R}} \left(\int_{-\infty}^{t-y} f_X(x)\,\mathrm{d}x \right) f_Y(y)\,\mathrm{d}y = \int_{\mathbf{R}} \left(\int_{-\infty}^{t} f_X(s-y)\,\mathrm{d}s \right) f_Y(y)\,\mathrm{d}y \\ &= \int_{-\infty}^{t} \left(\int_{\mathbf{R}} f_X(s-y) f_Y(y)\,\mathrm{d}y \right) \mathrm{d}s. \end{aligned}$$

On en déduit ainsi le résultat. □

Exemple 7.4.2.3. Soient X et Y indépendantes, de loi $\mathcal{E}(\lambda)$. On a alors

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

On peut généraliser le résultat à la somme de n variables aléatoires indépendantes de même loi exponentielle.

Définition 7.4.2.4. 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

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

avec

$$\Gamma(a) = \int_0^{+\infty} x^{a-1} e^{-x}\,\mathrm{d}x.$$

La loi $\Gamma(1,\lambda)$ correspond à la loi $\mathcal{E}(\lambda)$.

Proposition 7.4.2.5. Si $X_1, \ldots, X_n$ sont des variables aléatoires indépendantes de loi $\mathcal{E}(\lambda)$, alors $X_1 + \cdots + X_n \sim \Gamma(n,\lambda)$.

La proposition peut se démontrer par récurrence.

8 Convergences

8.1 Quelques inégalités

On considère dans cette partie des variables aléatoires réelles.

8.1.1 Inégalité de Markov

Proposition 8.1.1.1 (Inégalité de Markov)**.** Soit X une variable aléatoire positive et $a > 0$. On a alors

$$\mathbf{P}(X \geqslant a) \leqslant \frac{\mathbf{E}[X]}{a}.$$

Démonstration. On a

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

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

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

alors $\mathbf{E}[Y] = \mathbf{P}(X \geqslant a)$ ce qui nous donne le résultat. □

8.1.2 Inégalité de Bienaymé-Tchebychev

Proposition 8.1.2.1 (Inégalité de Bienaymé-Tchebychev)**.** Soit X une v.a. réelle de carré intégrable, $a > 0$, alors

$$\mathbf{P}(|X - \mathbf{E}[X]| \geqslant a) \leqslant \frac{\mathbf{V}(X)}{a^2}.$$

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

$$\begin{aligned}\mathbf{P}(|X - \mathbf{E}[X]| \geqslant a) &= \mathbf{P}(|X - \mathbf{E}[X]|^2 \geqslant a^2) \\ &\leqslant \frac{\mathbf{E}[|X - \mathbf{E}[X]|^2]}{a^2} = \frac{\mathbf{V}(X)}{a^2}.\end{aligned}$$

□

8.1.3 Inégalité de Jensen

Rappelons la définition d'une fonction convexe réelle.

Définition 8.1.3.1. On dit que $f\colon \mathbf{R} \to \mathbf{R}$ est une fonction *convexe* si pour tous $(x, y) \in \mathbf{R}^2$, $p \in [0, 1]$, on a

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

Proposition 8.1.3.2. Si $f\colon \mathbf{R} \to \mathbf{R}$ est une fonction $\mathcal{C}^1$, 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 8.1.3.3. Si $\phi\colon \mathbf{R} \to \mathbf{R}$ est une fonction convexe, alors

$$\phi(\mathbf{E}[X]) \leqslant \mathbf{E}[\phi(X)]$$

sous réserve d'existence des espérances.

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

$$\phi(x) \geqslant \phi(t) + \phi'(t)(x - t) \quad (\forall x \in \mathbf{R}).$$

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

$$\phi(x) \geqslant \phi(t) + m(t)(x - t) \quad (\forall x \in \mathbf{R}).$$

Ainsi, on a

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

et

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

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

Exemple 8.1.3.4. Pour tout $r \geqslant 1$,

$$|\,\mathbf{E}[X]|^r \leqslant \mathbf{E}[|X|^r]$$

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

8.1.4 Inégalité de Hölder

Proposition 8.1.4.1 (Inégalité de Hölder)**.** Soient $1 < p < +\infty$ et $1 < q < +\infty$ tels que $1/p + 1/q = 1$ (q est appelé l'exposant conjugué de p). On a alors

$$\mathbf{E}[|XY|] \leqslant \mathbf{E}[|X|^p]^{1/p}\,\mathbf{E}[|Y|^q]^{1/q}.$$

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

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

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

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

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

et la croissance de la fonction ln nous permet de conclure.
Montrons maintenant l'inégalité de Hölder. Si $\mathbf{E}[|X|^p]^{1/p}\,\mathbf{E}[|Y|^q]^{1/q} = 0$, le résultat est évident. Sinon, on a d'après l'inégalité de Young :

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

d'où l'inégalité. □

8.2 Lemme de Borel-Cantelli

Définition 8.2.0.1. Soit $(A_n)_{n\geqslant 1}$ des évènements aléatoires (de la tribu $\mathcal{A}$).
• On a :

$$\begin{aligned}\limsup_{n\to+\infty} A_n &= \bigcap_{n\in\mathbf{N}_{>0}}\bigcup_{p\geqslant n} A_p \quad \in \mathcal{A}\\ &= \{\omega\in\Omega;\ (\forall n\in\mathbf{N}_{>0})\ (\exists p\geqslant n)\ \omega\in A_p\}\\ &= \{\omega\in\Omega;\ \omega \text{ appartient à une infinité de } A_n\}\\ &= \text{« une infinité de } A_n \text{ sont réalisés »}.\end{aligned}$$

• On a :

$$\begin{aligned}\liminf_{n\to+\infty} A_n &= \bigcup_{n\in\mathbf{N}_{>0}}\bigcap_{p\geqslant n} A_p \quad \in \mathcal{A}\\ &= \{\omega\in\Omega;\ (\exists n\in\mathbf{N}_{>0})\ (\forall p\geqslant n)\ \omega\in A_p\}\\ &= \{\omega\in\Omega;\ \omega \text{ appartient à tous les éléments de } A_n \text{ à partir d'un certain rang}\}\\ &= \text{« tous les } A_n \text{ sont réalisés sauf un nombre fini »}.\end{aligned}$$

Remarque 8.2.0.2. (1) On a $\liminf_{n\to+\infty} A_n \subset \limsup_{n\to+\infty} A_n$.
(2) On a $\left(\limsup_{n\to+\infty} A_n\right)^{\complement} = \liminf_{n\to+\infty} A_n^{\complement}$.

Exemple 8.2.0.3. (1) Si $A_n = [0, 2 + (-1)^n]$: $\liminf_{n\to+\infty} A_n = [0,1]$, $\limsup_{n\to+\infty} A_n = [0,3]$.
(2) Si $A_n = [0, 2 + (-1)^n + (-1/2)^n]$: $\liminf_{n\to+\infty} A_n = [0,1[$, $\limsup_{n\to+\infty} A_n = [0,3]$.

Théorème 8.2.0.4 (Lemme de Borel-Cantelli)**.** Soit $(A_n)_{n\in\mathbf{N}}$ une suite d'évènements aléatoires.
(i) Si $\sum_{n\in\mathbf{N}} \mathbf{P}(A_n) < +\infty$, alors $\mathbf{P}\left(\limsup_{n\to+\infty} A_n\right) = 0$.
(ii) Si les $(A_n)_{n\in\mathbf{N}}$ sont indépendants et si $\sum_{n\in\mathbf{N}} \mathbf{P}(A_n) = +\infty$, alors

$$\mathbf{P}\left(\limsup_{n\to+\infty} A_n\right) = 1.$$

Démonstration. (i) Si $A = \limsup_{n\to+\infty} A_n = \bigcap_{n\geqslant 0}\bigcup_{p\geqslant n} A_p \subset \bigcup_{p\geqslant n} A_p$ pour tout n, alors

$$\mathbf{P}\big(\limsup_{n\to+\infty} A_n\big) \leqslant \mathbf{P}\Big(\bigcup_{p\geqslant n} A_p\Big) \leqslant \sum_{p\geqslant n} \mathbf{P}(A_p) \overset{n\to+\infty}{\longrightarrow} 0$$

car $\sum_{p\geqslant n} \mathbf{P}(A_p)$ est le reste d'une série convergente.
(ii) On a $\mathbf{P}\big(\limsup_{n\to+\infty} A_n\big) = \mathbf{P}\big(\bigcap_{n\geqslant 1}(\bigcup_{p\geqslant n} A_p)\big) = \lim_{n\to+\infty} \mathbf{P}\big(\bigcup_{p\geqslant n} A_p\big)$ car $\big(\bigcup_{p\geqslant n} A_p\big)_{n\geqslant 1}$ est décroissante. Pour conclure, il suffit de montrer que $\mathbf{P}\big(\bigcup_{p\geqslant n} A_p\big) = 1$ pour tout $n \geqslant 1$. On a

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

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

$$\mathbf{P}\Big(\bigcup_{p=n}^{N} A_p\Big) \geqslant 1 - \prod_{p=n}^{N} \exp(-\mathbf{P}(A_p)) = 1 - \exp\Big(-\sum_{p=n}^{N} \mathbf{P}(A_p)\Big) \xrightarrow{N\to+\infty} 1$$

car la série diverge. Ainsi, la suite $(\bigcup_{p\geqslant n} A_p)_{N\geqslant n}$ étant croissante, on a

$$\mathbf{P}\Big(\bigcup_{p\geqslant n} A_p\Big) = \lim_{N\to+\infty} \mathbf{P}\Big(\bigcup_{p=n}^{N} A_p\Big) \geqslant 1.$$

□

8.3 Définitions des différentes notions de convergences

On considère $(X_n)_{n\in\mathbf{N}}$ une suite de v.a. réelles définies sur un même espace de probabilité $(\Omega, \mathcal{A}, \mathbf{P})$ et X une v.a. également définie sur cet espace de probabilité $(\Omega, \mathcal{A}, \mathbf{P})$. 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 8.3.0.1. On dit que $(X_n)_{n\in\mathbf{N}}$ *converge presque sûrement* vers X si :

$$\mathbf{P}(\{\omega \in \Omega;\ X_n(\omega) \to X(\omega)\}) = 1.$$

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

Définition 8.3.0.2. On dit que $(X_n)_{n\in\mathbf{N}}$ *converge en probabilité* vers X si :

$$(\forall \varepsilon > 0) \qquad \mathbf{P}(|X_n - X| \geqslant \varepsilon) \xrightarrow{n\to+\infty} 0.$$

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

Remarquons que si $\varepsilon < \varepsilon'$, alors $\mathbf{P}(|X_n - X| \geqslant \varepsilon') \leqslant \mathbf{P}(|X_n - X| \geqslant \varepsilon)$, ce qui implique que si

$$\mathbf{P}(|X_n - X| \geqslant \varepsilon) \xrightarrow{n\to+\infty} 0$$

alors

$$\mathbf{P}(|X_n - X| \geqslant \varepsilon') \xrightarrow{n\to+\infty} 0$$

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

$$\mathbf{P}(|X_n - X| \geqslant \varepsilon) \xrightarrow{n\to+\infty} 0$$

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

Définition 8.3.0.3. On dit que $(X_n)_{n\in\mathbf{N}}$ *converge dans* L^p pour $p \geqslant 1$ vers X si : $X_n \in L^p$ (*i.e.* $\mathbf{E}[|X_n|^p] < +\infty$) pour tous $n \in \mathbf{N}$, $X \in L^p$ ($\mathbf{E}[|X|^p] < +\infty$) et

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

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

Définition 8.3.0.4. On dit que $(X_n)_{n\in\mathbf{N}}$ *converge en loi* vers X si : $F_{X_n}(t) \to F_X(t)$ en tout point de continuité de F_X. On note alors $X_n \xrightarrow[n\to+\infty]{\mathcal{L}} X$.

Remarque 8.3.0.5. (1) Si F_X est continue, alors pour tous $a < b \in \mathbf{R}$,

$$\mathbf{P}(X_n \in [a,b]) \xrightarrow{n\to+\infty} \mathbf{P}(X \in [a,b])$$

(2) Pour la convergence en loi, on a pas besoin 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 $X_n \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}(0,1)$ au lieu de $X_n \xrightarrow[n\to+\infty]{\mathcal{L}} X$ 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 $X_n \xrightarrow[n\to+\infty]{\mathcal{L}} 0$. Regardons ce qu'il en est de la fonction de répartition :

$$F_{X_n}(t) = \mathbf{P}(X_n \leqslant t) = \begin{cases} 0 & \text{si } t < 1/n \\ 1 & \text{sinon} \end{cases} \xrightarrow{n\to+\infty} \begin{cases} 0 & \text{si } t \leqslant 0 \\ 1 & \text{sinon} \end{cases}$$

Or

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

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

8.4 Autres caractérisations et liens entre les convergences

Proposition 8.4.0.1. On a

$$X_n \xrightarrow[n\to+\infty]{\text{p.s.}} X \Leftrightarrow (\forall \varepsilon > 0)\ \mathbf{P}\left(\limsup_{n\to+\infty}\{|X_n - X| \geqslant \varepsilon\}\right) = 0.$$

Remarque 8.4.0.2. Si pour tout $\varepsilon > 0$, on a

$$\sum_{n\in\mathbf{N}} \mathbf{P}(|X_n - X| \geqslant \varepsilon) < +\infty$$

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

Démonstration. Soit $\omega \in \Omega$.

$$\begin{aligned} X_n(\omega) \xrightarrow[n\to+\infty]{\mathcal{L}} X(\omega) &\Leftrightarrow (\forall \varepsilon > 0)\ (\exists N \in \mathbf{N})\ (\forall n \geqslant N)\ |X_n(\omega) - X(\omega)| < \varepsilon \\ &\Leftrightarrow \omega \in \bigcap_{\varepsilon>0} \bigcup_{N\in\mathbf{N}} \bigcap_{n\geqslant N} A_n^\varepsilon \end{aligned}$$

avec $A_n^\varepsilon = \{|X_n - X| < \varepsilon\}$. La dernière intersection est non dénombrable, néanmoins, comme elle est croissante en ε (lorsque ε croit), on peut se restreindre aux valeurs $\varepsilon = 1/k$ pour $k \in \mathbf{N}_{>0}$. On obtient alors

$$X_n(\omega) \xrightarrow[n\to+\infty]{\mathcal{L}} X(\omega) \Leftrightarrow \omega \in \bigcap_{k\in\mathbf{N}_{>0}} \bigcup_{N\in\mathbf{N}} \bigcap_{n\geqslant N} A_n^{1/k}. \tag{8.1}$$

En passant à la contraposée, on obtient donc

$$\begin{aligned} X_n(\omega) \not\to X(\omega) &\Leftrightarrow \omega \in \bigcup_{k\in\mathbf{N}_{>0}} \bigcap_{N\in\mathbf{N}} \bigcup_{n\geqslant N} (A_n^{1/k})^{\complement} \\ &\Leftrightarrow \omega \in \bigcup_{k\in\mathbf{N}_{>0}} \limsup_{n\to+\infty}\{|X_n - X| \geqslant 1/k\}. \end{aligned}$$

Ainsi, on a

$$\mathbf{P}\left(\limsup_{n\to+\infty}\{|X_n - X| \geqslant 1/k\}\right) \leqslant \mathbf{P}(X_n(\omega) \not\to X(\omega)) \quad (\forall k \in \mathbf{N}_{>0})$$

et comme l'union est croissante en k,

$$\lim_{k\to+\infty} \mathbf{P}\left(\limsup_{n\to+\infty}\{|X_n - X| \geqslant 1/k\}\right) = \mathbf{P}\left(\bigcup_{k\in\mathbf{N}_{>0}} \limsup_{n\to+\infty}\{|X_n - X| \geqslant 1/k\}\right) = \mathbf{P}(X_n(\omega) \not\to X(\omega)).$$

En utilisant à nouveau la croissance en k de l'union, on obtient

$$\mathbf{P}(\limsup_{n\to+\infty}\{|X_n - X| \geqslant \varepsilon\}) \leqslant \mathbf{P}(X_n(\omega) \not\to X(\omega)), \quad \forall \varepsilon > 0$$

et

$$\lim_{\varepsilon\to 0} \mathbf{P}\left(\limsup_{n\to+\infty}\{|X_n - X| \geqslant \varepsilon\}\right) = \mathbf{P}(X_n(\omega) \not\to X(\omega))$$

Si $X_n \xrightarrow[n\to+\infty]{\text{p.s.}} X$, alors $\mathbf{P}(X_n(\omega) \not\to X(\omega)) = 0$, ce qui implique que

$$\mathbf{P}(\limsup_{n\to+\infty}\{|X_n - X| \geqslant \varepsilon\}) = 0 \quad (\forall \varepsilon > 0).$$

Inversement, si

$$\mathbf{P}\left(\limsup_{n\to+\infty}\{|X_n - X| \geqslant \varepsilon\right) = 0 \quad (\forall \varepsilon > 0)$$

alors

$$\lim_{\varepsilon\to 0} \mathbf{P}\left(\limsup_{n\to+\infty}\{|X_n - X| \geqslant \varepsilon\}\right) = 0 = \mathbf{P}(X_n(\omega) \not\to X(\omega))$$

ce qui prouve le résultat. □

Remarque 8.4.0.3. Remarquons que (8.1) permet de justifier que $\{X_n \to X\}$ est bien un évènement aléatoire et donc on a le droit de calculer sa probabilité. En particulier, cela implique que la définition de la convergence presque sûre à bien un sens.

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

Proposition 8.4.0.4. Les assertions suivantes sont équivalentes :
(i) $X_n(\omega) \xrightarrow[n\to+\infty]{\mathcal{L}} X(\omega)$;
(ii) pour toute fonction h continue bornée on a $\mathbf{E}[h(X_n)] \xrightarrow{n\to+\infty} \mathbf{E}[h(X)]$;
(iii) pour tout $t \in \mathbf{R}$, $\varphi_{X_n}(t) \xrightarrow{n\to+\infty} \varphi_X(t)$.

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 8.4.0.5 (Levy)**.** S'il existe une fonction $\phi\colon \mathbf{R} \to \mathbf{C}$ continue en 0 telle que $\varphi_{X_n}(t) \xrightarrow{n\to+\infty} \phi(t)$ pour tout $t \in \mathbf{R}$, 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 8.4.0.6. Soit $(X_n)_{n\in\mathbf{N}}$ et X des v.a. à valeurs dans $\mathbf{N}$. On a $X_n \xrightarrow[n\to+\infty]{\mathcal{L}} X$ si et seulement si

$$\mathbf{P}(X_n = k) \xrightarrow{n\to+\infty} \mathbf{P}(X = k) \quad (\forall k \in \mathbf{N})$$

Démonstration. • Supposons que $(X_n)_{n\in\mathbf{N}}$ converge en loi vers X. Pour $k \in \mathbf{N}$, 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

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

• Inversement, on suppose que

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

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

$$F_{X_n}(t) = \mathbf{P}(X_n \leqslant t) = \sum_{k=0}^{\lfloor t\rfloor} \mathbf{P}(X_n = k) \xrightarrow{n\to+\infty} \sum_{k=0}^{\lfloor t\rfloor} \mathbf{P}(X = k) = F_X(t)$$

donc $X_n(\omega) \xrightarrow[n\to+\infty]{\mathcal{L}} X(\omega)$. □

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

Proposition 8.4.0.7. (i) La convergence L^q implique la convergence L^p lorsque $q > p \geqslant 1$;
(ii) La convergence L^p $(p \geqslant 1)$ implique la convergence en probabilité ;
(iii) La convergence presque-sûre implique la convergence en probabilité ;
(iv) La convergence en probabilité implique la converge en loi.

Démonstration. (i) En utilisant l'inégalité de Hölder, ou l'inégalité de Jensen, on a

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

car $q/p \geqslant 1$. Ainsi, le 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

$$\mathbf{P}(|X_n - X| \geqslant \varepsilon) = \mathbf{P}(|X_n - X|^p \geqslant \varepsilon^p) \leqslant \frac{\mathbf{E}[|X_n - X|^p]}{\varepsilon^p}.$$

Ainsi, pour tout ε, $\mathbf{P}(|X_n - X| \geqslant \varepsilon)$ tend vers 0 si $\mathbf{E}[|X_n - X|^p]$ tend vers 0.
(iii) Soit $\varepsilon > 0$ quelconque.

$$\mathbf{P}(|X_n - X| \geqslant \varepsilon) \leqslant \mathbf{P}\left(\bigcup_{p \geqslant n} \{|X_p - X| \geqslant \varepsilon\}\right) \xrightarrow{n \to +\infty} \mathbf{P}\left(\bigcap_{n \geqslant 0} \bigcup_{p \geqslant n} \{|X_p - X| \geqslant \varepsilon\}\right).$$

donc si

$$\mathbf{P}\left(\limsup_{n \to +\infty} \{|X_n - X| \geqslant \varepsilon\}\right) = \mathbf{P}\left(\bigcap_{n \geqslant 0} \bigcup_{p \geqslant n} \{|X_p - X| \geqslant \varepsilon\}\right) \xrightarrow{n \to +\infty} 0$$

alors $\mathbf{P}(|X_n - X| \geqslant \varepsilon) \xrightarrow{n \to +\infty} 0$.
(iv) Soit t un point de continuité de F_X.

$$\begin{aligned} F_{X_n}(t) &= \mathbf{P}(X_n \leqslant t) \\ &= \mathbf{P}(\{X_n \leqslant t\} \cap \{|X_n - X| < \varepsilon\}) + \mathbf{P}(\{X_n \leqslant t\} \cap \{|X_n - X| \geqslant \varepsilon\}) \\ &\leqslant \mathbf{P}(X \leqslant t + \varepsilon) + \mathbf{P}(|X_n - X| \geqslant \varepsilon) \xrightarrow{n \to +\infty} F_X(t + \varepsilon). \end{aligned}$$

On a donc $\limsup_{n \to +\infty} F_{X_n}(t) \leqslant F_X(t + \varepsilon)$ 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 : $\limsup_{n \to +\infty} F_{X_n}(t) \leqslant F_X(t)$.

De la même façon,

$$\begin{aligned} 1 - F_{X_n}(t) &= \mathbf{P}(X_n > t) \\ &= \mathbf{P}(\{X_n > t\} \cap \{|X_n - X| < \varepsilon\}) + \mathbf{P}(\{X_n > t\} \cap \{|X_n - X| \geqslant \varepsilon\}) \\ &\leqslant 1 - \mathbf{P}(X \leqslant t - \varepsilon) + \mathbf{P}(|X_n - X| \geqslant \varepsilon) \xrightarrow{n \to +\infty} 1 - F_X(t - \varepsilon). \end{aligned}$$

On a donc $\liminf_{n \to +\infty} F_{X_n}(t) \geqslant F_X(t - \varepsilon)$ et comme le membre de gauche ne dépend pas de ε, on peut faire tendre ε vers 0 dans le membre de droite pour obtenir : $\liminf_{n \to +\infty} F_{X_n}(t) \geqslant F_X(t)$ en utilisant la continuité de F_X au point t. On en conclut que

$$\liminf_{n \to +\infty} F_{X_n}(t) = \limsup_{n \to +\infty} F_{X_n}(t) = \lim_{n \to +\infty} F_{X_n}(t) = F_X(t).$$

□

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 8.4.0.8. (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.

Démonstration. (i) On pose $n_0 = 0$ et grâce à l'hypothèse de convergence en probabilité, on construit par récurrence une suite strictement croissante $(n_k)_{k \in \mathbf{N}}$ telle que pour tout $k \in \mathbf{N}_{>0}$, on a

$$\mathbf{P}(|X_{n_k} - X| \geqslant 1/k) \leqslant \frac{1}{k^2}.$$

Ainsi, pour tout $\varepsilon > 0$, on a

$$\begin{aligned}\sum_{k\in\mathbf{N}} \mathbf{P}(|X_{n_k} - X| \geqslant \varepsilon) &= \sum_{k<1/\varepsilon} \mathbf{P}(|X_{n_k} - X| \geqslant \varepsilon) + \sum_{k\geqslant 1/\varepsilon} \mathbf{P}(|X_{n_k} - X| \geqslant \varepsilon) \\ &\leqslant \sum_{k<1/\varepsilon} \mathbf{P}(|X_{n_k} - X| \geqslant \varepsilon) + \sum_{k\geqslant 1/\varepsilon} \mathbf{P}(|X_{n_k} - X| \geqslant 1/k) \\ &\leqslant \lfloor 1/\varepsilon \rfloor + 1 + \sum_{k\geqslant 1/\varepsilon} \frac{1}{k^2} < +\infty.\end{aligned}$$

Le lemme de Borel-Cantelli permet alors de conclure à la convergence presque-sûre de $(X_{n_k})_{k\in\mathbf{N}}$ vers X.
(ii) Tout d'abord, $F_c(t) = \mathbf{1}_{t\geqslant c}$ qui est continue sur $\mathbf{R}\backslash\{c\}$, donc

$$F_{X_n}(t) \xrightarrow{n\to+\infty} \begin{cases} F_c(t) = 0 & \text{si } t < c \\ F_c(t) = 1 & \text{si } t > c \end{cases}$$

On a

$$\mathbf{P}(|X_n - c| \geqslant \varepsilon) = \mathbf{P}(X_n \leqslant c - \varepsilon) + \mathbf{P}(X_n \geqslant c + \varepsilon) \xrightarrow{n\to+\infty} F_c(c - \varepsilon) + (1 - F_c(c + \varepsilon)) = 0$$

car F_c est continue en $c + \varepsilon$. □

Proposition 8.4.0.9. 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_nY_n converge vers XY (pour les mêmes types de convergences).

Remarque 8.4.0.10. Attention ! La proposition précédente est fausse 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 $X_n \xrightarrow[n\to+\infty]{\mathcal{L}} X$ et $X_n \xrightarrow[n\to+\infty]{\mathcal{L}} -X$. Pourtant $2X_n = 2X$ ne converge pas en loi vers 0 !

Proposition 8.4.0.11. Soit $h\colon \mathbf{R} \to \mathbf{R}$ est une fonction continue. 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.

Démonstration. Exercice. □

8.5 Quelques exemples de convergences

8.5.1 Lois de Bernoulli

On considère $X_n \sim \mathcal{B}(p_n)$ avec $p_n \xrightarrow{n\to+\infty} 0$. Pour tout $0 < \varepsilon < 1$,

$$\mathbf{P}(|X_n - 0| \geqslant \varepsilon) = \mathbf{P}(X_n = 1) = p_n \xrightarrow{n\to+\infty} 0$$

donc $X_n \xrightarrow[n\to+\infty]{\mathbf{P}} 0$. De plus, pour tout $p \geqslant 1$,

$$\mathbf{E}[|X_n|^p] = \mathbf{E}[X_n] = p_n \xrightarrow{n\to+\infty} 0$$

donc $X_n \xrightarrow[n\to+\infty]{L^p} 0$. Enfin, si on suppose les X_n indépendantes, pour tout $\varepsilon > 0$, on a d'après le lemme de Borel-Cantelli,

$$\mathbf{P}\left(\limsup_{n\to+\infty} |X_n - 0| \geqslant \varepsilon\right) = 0 \Leftrightarrow \sum_{n\geqslant 1} \mathbf{P}(|X_n - 0| \geqslant \varepsilon) < +\infty.$$

On a donc $X_n \xrightarrow[n\to+\infty]{\text{p.s.}} 0$ si et seulement si $\sum_{n\geqslant 1} p_n < +\infty$. Attention, si on a plus l'indépendance, alors $\sum_{n\geqslant 1} p_n < +\infty$ 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 $X_n(\omega) = \mathbf{1}_{\omega\in[0,1/n]}$. On a alors $X_n \sim \mathcal{B}(1/n)$, $\sum_{n\geqslant 1} 1/n = +\infty$ et pourtant $X_n \xrightarrow[n\to+\infty]{\text{p.s.}} 0$! On pose maintenant $Y_n = nX_n$ avec $X_n \sim B(1/n)$. Pour tout $0 < \varepsilon < 1$,

$$\mathbf{P}(|Y_n - 0| \geqslant \varepsilon) = \mathbf{P}(X_n = 1) = 1/n \xrightarrow{n\to+\infty} 0$$

donc $Y_n \xrightarrow[n\to+\infty]{\mathbf{P}} 0$. Par contre

$$\mathbf{E}[|Y_n|] = n \times \frac{1}{n} = 1,$$

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 8.5.2.1. Si $X_n \sim \mathcal{B}(n, \lambda/n)$, alors $X_n \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{P}(\lambda)$.

Démonstration. Comme on a des v.a. à valeurs dans $\mathbf{N}$, il suffit de regarder la convergence de $\mathbf{P}(X_n = k)$ pour tous les $k \in \mathbf{N}$. Soit $k \in \mathbf{N}$,

$$\begin{aligned}
\mathbf{P}(X_n = k) &= \binom{n}{k}\left(\frac{\lambda}{n}\right)^k\left(1-\frac{\lambda}{k}\right)^{n-k} \qquad (\text{pour } n \geqslant k)\\
&= \frac{n!}{k!(n-k)!}\frac{\lambda^k}{n^k}\left(1-\frac{\lambda}{n}\right)^n\left(1-\frac{\lambda}{n}\right)^{-k}\\
&= \frac{\lambda^k}{k!}\frac{n(n-1)\cdots(n-k+1)}{nn\cdots n}\left(1-\frac{\lambda}{n}\right)^n\left(1-\frac{\lambda}{n}\right)^{-k}\\
&\xrightarrow{n\to+\infty} e^{-\lambda}\frac{\lambda^k}{k!}.
\end{aligned}$$

Il est également possible de démontrer le résultat en regardant la convergence de la fonction de répartition. □

8.5.3 Maximum de lois exponentielles

Soit $(X_n)_{n\in\mathbf{N}}$ des v.a. i.i.d de loi $\mathcal{E}(\lambda)$. On pose $M_n = \max\{X_1, \ldots, X_n\}$ 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 $t \in \mathbf{R}$,

$$\begin{aligned}
F_{Y_n}(t) &= \mathbf{P}((\lambda M_n - \ln(n)) \leqslant t) = \mathbf{P}\left(M_n \leqslant \frac{t+\ln(n)}{\lambda}\right)\\
&= \prod_{i=1}^{n}\mathbf{P}\left(X_i \leqslant \frac{t+\ln(n)}{\lambda}\right)\\
&= \left(1-\exp\left(-\frac{\lambda(t+\ln(n))}{\lambda}\right)\right)^n \text{ car } \frac{t+\ln(n)}{\lambda} \geqslant 0 \text{ pour } n \text{ assez grand}\\
&= \left(1-\frac{e^{-t}}{n}\right)^n = \exp\left(n\ln\left(1-\frac{e^{-t}}{n}\right)\right) \xrightarrow{n\to+\infty} e^{-e^{-t}}.
\end{aligned}$$

Comme $t \mapsto e^{-e^{-t}}$ est une fonction de répartition (croissante, continue, tendant vers 0 en $-\infty$ et 1 en $+\infty$), Y_n converge en loi vers Y de fonction de répartition $F_Y(t) = e^{-e^{-t}}$. Remarquons que cette fonction de répartition est $\mathcal{C}^1$, donc Y est une v.a. à densité.

9 Théorèmes limites en probabilité

9.1 Lois des grands nombres

Soit $(X_n)_{n\in\mathbf{N}}$ une suite de v.a. i.i.d. On pose $S_n = X_1 + \cdots + X_n$ 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 9.1.0.1 (Loi faible des grands nombres)**.** Soit $(X_n)_{n\in\mathbf{N}}$ une suite de v.a. i.i.d de même loi que X. On suppose que $\mathbf{E}[|X|^2] < +\infty$. On a alors

$$\frac{S_n}{n} \xrightarrow[n\to+\infty]{L^2} \mathbf{E}[X] \quad \text{et donc} \quad \frac{S_n}{n} \xrightarrow[n\to+\infty]{\mathbf{P}} \mathbf{E}[X]\,\mathbf{E}[X].$$

Démonstration. $\mathbf{E}[S_n/n] = \mathbf{E}[X]$ et donc

$$\mathbf{E}\left[\left|\frac{S_n}{n} - \mathbf{E}[X]\right|^2\right] = \mathbf{V}\left(\frac{S_n}{n}\right) = \frac{1}{n}\,\mathbf{V}(X) \xrightarrow{n\to+\infty} \mathbf{E}[X].$$

□

Théorème 9.1.0.2 (Loi forte des grands nombres)**.** Soit $(X_n)_{n\in\mathbf{N}}$ une suite de v.a. i.i.d de même loi que X. On suppose que $\mathbf{E}[|X|] < +\infty$. On a alors

$$\frac{S_n}{n} \xrightarrow[n\to+\infty]{\text{p.s.}} \mathbf{E}[X].$$

Démonstration. Nous allons prouver le résultat uniquement avec l'hypothèse supplémentaire $\mathbf{E}[|X|^4] < +\infty$, le cas général étant admis.
On pose $m_1 = \mathbf{E}[X]$, $m_2 = \mathbf{E}[X^2]$ et $m_4 = \mathbf{E}[X^4]$. 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 $\frac{S_n}{n} \xrightarrow[n\to+\infty]{\text{p.s.}} 0$ est équivalent à montrer que pour tout $\varepsilon > 0$, $\mathbf{P}\left(\limsup_{n\to+\infty}\{|S_n/n| > \varepsilon\}\right) = 0$. D'après le lemme de Borel-Cantelli, il suffit de montrer que

$$\sum_{n\geqslant 1} \mathbf{P}(|S_n/n| > \varepsilon) < +\infty.$$

Soit $\varepsilon > 0$ fixé. D'après l'inégalité de Markov on a

$$\mathbf{P}(|S_n/n| > \varepsilon) = \mathbf{P}(|S_n|^2 > \varepsilon^4 n^4) \leqslant \frac{\mathbf{E}[S_n^4]}{\varepsilon^4 n^4}$$

et par linéarité de l'espérance,

$$\mathbf{E}[S_n^4] = \mathbf{E}[(X_1 + \cdots + X_n)^4] = \sum_{1\leqslant i,j,k,\ell\leqslant n} \mathbf{E}[X_i X_j X_k X_\ell].$$

Lorsque $i \neq j \neq k \neq \ell$, l'indépendance nous donne

$$\mathbf{E}[X_i X_j X_k X_\ell] = \mathbf{E}[X_i]\,\mathbf{E}[X_j]\,\mathbf{E}[X_k]\,\mathbf{E}[X_\ell] = 0.$$

De la même façon, dès qu'un indice est seul, on a $\mathbf{E}[X_i X_j X_k X_\ell] = 0$. Les seuls termes non nuls sont :
• $\mathbf{E}[X_i^4] = m_4$, il y en a n.
• $\mathbf{E}[X_i^2 X_j^2] = \mathbf{E}[X_i^2]\,\mathbf{E}[X_j^2] = m_2^2$, pour $1 \leqslant i \neq j \leqslant n$.
Il y en a $C_n^2 C_4^2 = 3n(n-1)$ (choix de i et j, puis place des i). On a donc $\mathbf{E}[(S_n)^4] \leqslant Cn^2$ et $\mathbf{P}(|S_n/n| > \varepsilon) \leqslant \frac{C}{\varepsilon^2 n^2}$ qui est le terme général d'une série convergente. □

Reprenons l'exemple de départ : si on lance n fois une pièce, la proportion de pile tend vers p.

9.2 Théorème centrale limite

On sait maintenant que, sous réserve d'existence de $\mathbf{E}[X]$, $\frac{S_n}{n} \xrightarrow[n\to+\infty]{\text{p.s.}} \mathbf{E}[X]$ et donc $\frac{S_n}{n} - \mathbf{E}[X] \xrightarrow[n\to+\infty]{\text{p.s.}} 0$. Afin de quantifier la vitesse de cette convergence, on cherche α tel que $n^\alpha\big(\frac{S_n}{n} - \mathbf{E}[X]\big)$ tend vers quelque chose de non trivial.

Théorème 9.2.0.1 (Théorème centrale limite)**.** Soit $(X_n)_{n\in\mathbf{N}}$ une suite de v.a. i.i.d de même loi que X. On suppose que $\mathbf{E}[|X|^2] < +\infty$. On note $\mathbf{E}[X] = m$ et $\mathbf{V}(X) = \sigma^2$ (supposée non nulle). On a alors

$$\frac{\sqrt{n}}{\sigma}\left(\frac{S_n}{n} - m\right) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}(0,1).$$

Remarquons que si on pose $Y_n = \frac{\sqrt{n}}{\sigma}(\frac{S_n}{n} - m)$, alors $\mathbf{E}[Y_n] = 0$ et $\mathbf{V}(Y_n) = 1$.

Démonstration. 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 $t \in \mathbf{R}$ fixé,

$$\begin{aligned}\phi_{Y_n}(t) &= \mathbf{E}[e^{itY_n}] = \mathbf{E}[e^{it\frac{S_n}{\sigma\sqrt{n}}}] = \mathbf{E}[e^{it\frac{X_1+\cdots+X_n}{\sigma\sqrt{n}}}] \\ &= \mathbf{E}[e^{it\frac{X_1}{\sigma\sqrt{n}}}]^n = \left[\phi_{X_1}\left(\frac{t}{\sigma\sqrt{n}}\right)\right]^n.\end{aligned}$$

On fait un développement limité de ϕ_{X_1} en 0 : $\mathbf{E}[X_1^2] < +\infty$ donc ϕ_{X_1} est $\mathcal{C}^2$ et

$$\begin{aligned}\phi_{X_1}(h) &= \phi_{X_1}(0) + h\phi'_{X_1}(0) + \frac{h^2}{2}\phi''_{X_1}(0) + \mathsf{o}(h^2) \\ &= 1 + hi\,\mathbf{E}[X_1] - \frac{h^2}{2}\,\mathbf{E}[X_1^2] + \mathsf{o}(h^2).\end{aligned}$$

Ainsi,

$$\phi_{X_1}\left(\frac{t}{\sigma\sqrt{n}}\right) = 1 - \frac{t^2\sigma^2}{2\sigma^2 n} + \mathsf{o}(1/n)$$

et

$$\phi_{Y_n}(t) = \left(1 - \frac{t^2}{2n} + \mathsf{o}(1/n)\right)^n = \left(1 - \frac{t^2}{2n}\right)^n (1 + \mathsf{o}(1/n))^n.$$

De manière classique,

$$\left(1 - \frac{t^2}{2n}\right)^n = \exp\left(n\ln\left(1 - \frac{t^2}{2n}\right)\right) \xrightarrow{n\to+\infty} e^{-t^2/2}.$$

On ne peut pas faire la même chose pour le terme $(1 + \mathsf{o}(1/n))^n$ car $\mathsf{o}(1/n)$ est complexe ! Néanmoins, on a

$$\begin{aligned}|(1 + \mathsf{o}(1/n))^n - 1| &= \left|\sum_{k=1}^n C_n^k(\mathsf{o}(1/n))^k\right| \leqslant \sum_{k=1}^n C_n^k|\,\mathsf{o}(1/n)|^k = (1 + |\,\mathsf{o}(1/n)|)^n - 1 \\ &= e^{n\ln(1)+|\,\mathsf{o}(1/n)|)} - 1 \xrightarrow{n\to+\infty} 0\end{aligned}$$

On a donc $\phi_{Y_n}(t) \xrightarrow{n\to+\infty} e^{-t^2/2}$ 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 10.1.0.1. Soient $m \in \mathbf{R}$ et $\sigma^2 > 0$. $X \sim \mathcal{N}(m, \sigma^2)$ si X est une v.a. à densité, de densité

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(- \frac{(x-m)^2}{2\sigma^2} \right) \quad \text{avec} \quad x \in \mathbf{R}\,.$$

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 10.1.0.2. Si $X \sim \mathcal{N}(m, \sigma^2)$ avec $\sigma > 0$, alors

$$\frac{X - m}{\sigma} \sim \mathcal{N}(0, 1).$$

Démonstration. Il suffit de calculer la fonction de répartition de la v.a. $\frac{X-m}{\sigma}$. □

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 10.1.0.3. Soit $X \sim \mathcal{N}(m, \sigma^2)$ de densité f_{m,σ^2} lorsque $\sigma > 0$. On a alors :
• $\int_{\mathbf{R}} f_{m,\sigma^2}(x)\,\mathrm{d}x = 1$.
• $\mathbf{E}[X] = m$ et $\mathbf{V}(X) = \sigma^2$.

Démonstration. • En appliquant un changement le variable $u = (x-m)/\sigma$, on a

$$\int_{\mathbf{R}} f_{m,\sigma^2}(x)\,\mathrm{d}x = \int_{\mathbf{R}} f_{0,1}(u)\,\mathrm{d}u = \frac{1}{\sqrt{2\pi}} \int_{\mathbf{R}} e^{-u^2/2}\,\mathrm{d}u := \frac{1}{\sqrt{2\pi}} I$$

avec $I = \int_{\mathbf{R}} e^{-u^2/2}\,\mathrm{d}u$. En appliquant le théorème de Fubini (pour les fonctions positives) ainsi qu'un changement de variable en coordonnées polaires, on obtient

$$\begin{aligned} I^2 &= \int\int_{\mathbf{R}^2} \exp\left(- \frac{x^2+y^2}{2} \right) \mathrm{d}x\,\mathrm{d}y = \int_{r>0} \int_{\theta=0}^{2\pi} \exp\left(- \frac{r^2}{2} \right) r\,\mathrm{d}r\,\mathrm{d}\theta \\ &= 2\pi \big[- e^{-r^2/2} \big]_0^{+\infty} \end{aligned}$$

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

$$\mathbf{E}[X] = \mathbf{E}[m + \sigma Y] = m, \quad \mathbf{V}(X) = \mathbf{V}(m + \sigma Y) = \sigma^2.$$

Remarquons que pour tout $r \geqslant 0$, on a $x^r e^{-x^2/2} = \mathsf{o}(1/x^2)$ en $+\infty$, donc d'après le critère de Riemann, on a

$$\mathbf{E}[|Y|^r] = 2 \int_0^{+\infty} x^r e^{-x^2/2}\,\mathrm{d}x < +\infty.$$

En particulier, Y possède une espérance et une variance (finies).

$$\mathbf{E}[Y] = \frac{1}{\sqrt{2\pi}} \int_{\mathbf{R}} x e^{-x^2/2}\,\mathrm{d}x = 0$$

car la fonction intégrée est intégrable et impaire. De plus, une intégration par parties nous donne

$$\begin{aligned} \mathbf{E}[X^2] &= \frac{1}{\sqrt{2\pi}} \int_{\mathbf{R}} x^2 e^{-x^2/2}\,\mathrm{d}x = \frac{1}{\sqrt{2\pi}} \int_{\mathbf{R}} x \times x e^{-x^2/2}\,\mathrm{d}x \\ &= \frac{1}{\sqrt{2\pi}} \big[x(-e^{-x^2/2}) \big]_{-\infty}^{+\infty} + \int_{\mathbf{R}} \frac{e^{-x^2/2}}{\sqrt{2\pi}}\,\mathrm{d}x = 0 + 1. \end{aligned}$$

□

Proposition 10.1.0.4. Si $X \sim \mathcal{N}(m, \sigma^2)$, alors

$$\phi_X(t) = \exp\left(itm - \frac{\sigma^2 t^2}{2}\right).$$

Démonstration. 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

$$\phi'_Y(t) = \mathbf{E}[iXe^{itX}] = \int_{\mathbf{R}} ixe^{itx} \frac{e^{-x^2/2}}{\sqrt{2\pi}}\,\mathrm{d}x \quad (\forall t \in \mathbf{R}).$$

Une intégration par parties nous donne

$$\phi'_Y(t) = \left[ie^{itx}\left(-\frac{e^{-x^2/2}}{\sqrt{2\pi}}\right)\right]_{-\infty}^{+\infty} + \int_{\mathbf{R}} i^2 te^{itx}\frac{e^{-x^2/2}}{\sqrt{2\pi}}\,\mathrm{d}x = -t\phi_Y(t).$$

ϕ_Y est solution d'une équation différentielle linéaire du premier ordre, sa résolution nous permet d'obtenir

$$\phi_Y(t) = Ce^{-t^2/2} \quad (\forall t \in \mathbf{R})$$

et nécessairement $C = 1$ car $\phi_Y(0) = 1$. Revenons maintenant à la v.a. X : pour tout $t \in \mathbf{R}$,

$$\phi_X(t) = \mathbf{E}[e^{it(m+\sigma Y)}] = e^{itm}\,\mathbf{E}[e^{it\sigma Y}] = e^{itm}\phi_Y(\sigma t) = e^{itm-\sigma^2 t^2/2}.$$

□

Proposition 10.1.0.5. Si $X \sim \mathcal{N}(m, \sigma^2)$ et $Y \sim \mathcal{N}(m', \sigma'^2)$ sont indépendantes, alors

$$X + Y \sim \mathcal{N}(m + m', \sigma^2 + \sigma'^2).$$

Démonstration. Il suffit de remarquer que

$$\phi_{X+Y}(t) = \phi_X(t)\phi_Y(t) \quad (\forall t \in \mathbf{R})$$

et la fonction caractéristique obtenue est celle de la loi $\mathcal{N}(m + m', \sigma^2 + \sigma'^2)$. □

10.2 Vecteurs gaussiens

Définition 10.2.0.1. Un vecteur aléatoire X de $\mathbf{R}^n$ est dit *gaussien* si toute combinaison linéaires de ses composante est une v.a. réelle gaussienne : pour tout $(\lambda_1, \ldots, \lambda_n) \in \mathbf{R}^n$, $\sum_{i=1}^n \lambda_i X_i$ est une v.a. réelle gaussienne.

Remarque 10.2.0.2. Si on prend $(\lambda_1, \ldots, \lambda_n) = (0, \ldots, 0)$, alors $\sum_{i=1}^n \lambda_i X_i = 0$ ce qui explique pourquoi on a élargi la notion de v.a. gaussienne en incluant les constantes.

Corollaire 10.2.0.3. Si $X = (X_1, \ldots, X_n)$ est un vecteur gaussien, alors chaque composante X_i est une v.a. gaussienne.

Remarque 10.2.0.4. Attention ! La réciproque du corollaire est fausse ! Si on considère des v.a. $X_1, \ldots, X_n$ qui sont des gausiennes (réelles), alors $(X_1, \ldots, X_n)$ 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.* $\mathbf{P}(\varepsilon = 1) = \mathbf{P}(\varepsilon = -1) = 1/2$) indépendantes. On pose $Y = \varepsilon X$. On peut montrer que Y et X ont même fonction de répartition et donc même loi. Ainsi, $Y \sim \mathcal{N}(0,1)$. Par contre,

$$\mathbf{P}(X + Y = 0) = \mathbf{P}((1+\varepsilon)X = 0) = \mathbf{P}(\{X = 0\} \cup \{\varepsilon = -1\}) = \mathbf{P}(\varepsilon = -1) = 1/2$$

car

$$\mathbf{P}(\{X = 0\} \cup \{\varepsilon = -1\}) = \mathbf{P}(X = 0) + \mathbf{P}(\varepsilon = -1) - \mathbf{P}(X + 0)\,\mathbf{P}(\varepsilon = -1)$$

et $\mathbf{P}(X = 0) = 0$ (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 on ajoute une hypothèse d'indépendance, la réciproque du corollaire devient valable.

Proposition 10.2.0.5. Si $X_1, \ldots, X_n$ sont des v.a. gaussiennes indépendantes, alors $(X_1, \ldots, X_n)$ est un vecteur gaussien.

Démonstration. On note $X_i \sim \mathcal{N}(m_i, \sigma_i^2)$ pour $1 \leqslant i \leqslant n$. On considère $(\lambda_1, \ldots, \lambda_n) \in \mathbf{R}^n$. On a alors $\lambda_i X_i \sim \mathcal{N}(\lambda_i m_i, \lambda_i^2 \sigma_i^2)$ pour $1 \leqslant i \leqslant n$ et la proposition 10.1.0.5 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 $\mathbf{R}^n$.

Espérance : L'espérance de X est définie comme

$$\mathbf{E}[X] = \begin{pmatrix} \mathbf{E}[X_1] \\ \vdots \\ \mathbf{E}[X_n] \end{pmatrix}.$$

Matrice de covariance : Dans le cadre vectoriel, la variance devient une matrice, appelée matrice de covariance et donnée par

$$\begin{aligned}\Gamma &= \mathbf{E}[(X - \mathbf{E}[X])^{\mathrm{t}}(X - \mathbf{E}[X])] = \mathbf{E}\left[((X_i - \mathbf{E}[X_i])(X_j - \mathbf{E}[X_j]))_{1 \leqslant i,j \leqslant n}\right] \\ &= (\mathsf{Cov}(X_i, X_j))_{1 \leqslant i,j \leqslant n}.\end{aligned}$$

En utilisant la linéarité de l'espérance, on obtient une autre formulation possible de la matrice de covariance :

$$\Gamma = \mathbf{E}[(X - \mathbf{E}[X])^{\mathrm{t}}(X - \mathbf{E}[X])] = \mathbf{E}[X^{\mathrm{t}}X] - \mathbf{E}[X]^{\mathrm{t}}\,\mathbf{E}[X].$$

Proposition 10.3.0.1. Toute matrice de covariance est symétrique semi-définie positive.

Démonstration. La symétrie est claire. On note $m = \mathbf{E}[X]$. Soit $u \in \mathbf{R}^n$ quelconque : par linéarité de l'espérance on a

$${}^{\mathrm{t}}u\Gamma u = \mathbf{E}[{}^{\mathrm{t}}u(X - m)^{\mathrm{t}}(X - m)u] = \mathbf{E}[({}^{\mathrm{t}}(X - m)u)^2] \geqslant 0$$

d'où la positivité. Notons qu'elle n'est pas nécessairement définie positive car, par exemple, $\Gamma = 0$ si $X = \mathbf{E}[X]$. □

Fonction caractéristique : La fonction caractéristique d'un vecteur aléatoire de $\mathbf{R}^n$ est donnée par

$$\begin{aligned}\phi_X \colon \mathbf{R}^n &\to \mathbf{C} \\ t = \begin{pmatrix} t_1 \\ \vdots \\ t_n \end{pmatrix} &\mapsto \mathbf{E}[e^{i\langle t, X\rangle}] = \mathbf{E}[e^{i\sum_{j=1}^n t_j X_j}]\end{aligned}$$

Proposition 10.3.0.2. 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 10.4.0.1. Si X est un vecteur gaussien, alors $Y = AX + B$ avec $A \in \mathsf{M}_{p\times n}(\mathbf{R})$, $B \in \mathbf{R}^p$, est un vecteur gaussien. On a $\mathbf{E}[Y] = A\,\mathbf{E}[X] + B$ et la matrice de covariance de Y, notée Γ_Y est donnée par $\Gamma_Y = A\Gamma_X{}^{\mathrm{t}}A$;

Démonstration. Toute combinaison linéaire des composantes de Y, $\sum_{i=1}^n \lambda_i Y_i$, est une combinaison affine des composantes de X, $\sum_{i=1}^n \lambda_i \left(\sum_{j=1}^n a_{i,j}X_j + b_i \right)$, donc est une variable aléatoire gaussienne.

Par linéarité de l'espérance, on a $\mathbf{E}[Y] = A\,\mathbf{E}[X] + B$ et

$$\begin{aligned}\Gamma_Y &= \mathbf{E}[(Y - \mathbf{E}[Y])^{\mathrm{t}}(Y - \mathbf{E}[Y])] = \mathbf{E}[A(X - \mathbf{E}[X])^{\mathrm{t}}(X - \mathbf{E}[X])^{\mathrm{t}}A] \\ &= A\,\mathbf{E}[(X - \mathbf{E}[X])^{\mathrm{t}}(X - \mathbf{E}[X])]^{\mathrm{t}}A = A\Gamma_X{}^{\mathrm{t}}A.\end{aligned}$$

□

Théorème 10.4.0.2. Soit X un vecteur gaussien d'espérance m et de matrice de covariance Γ. On a alors

$$\phi_X(t) = \mathbf{E}[e^{i\langle t, X\rangle}] = \exp\left(i\langle t, m\rangle - \frac{1}{2}t^{\mathrm{t}}\Gamma t \right) \quad \text{pour} \quad t \in \mathbf{R}^n\,.$$

Corollaire 10.4.0.3. 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.

Démonstration. Soit $t \in \mathbf{R}^n$ quelconque. On a

$$\phi_X(t) = \mathbf{E}[e^{i\langle t,X\rangle}] = \mathbf{E}[e^{i1Y}] = \phi_Y(1)$$

en posant $Y = \langle t, X\rangle$. Remarquons que $\langle t, X\rangle = AX + 0$ avec $A = {}^{\mathrm{t}}t$, donc d'après la proposition 10.4.0.1, Y est un vecteur gaussien, donc une v.a. gaussienne car de dimension 1, d'espérance $Am = \langle t, m\rangle$ et de variance $A\Gamma^{\mathrm{t}}A = t^{\mathrm{t}}\Gamma t$. On a donc

$$\phi_Y(u) = \exp\left(iu\langle t, m\rangle - \frac{u^2}{2}t^{\mathrm{t}}\Gamma t\right).$$

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 10.5.0.1. Soit $X = (X_1, \ldots, X_n)$ un vecteur gaussien. On a équivalence entre :
(i) $X_1, \ldots, X_n$ sont indépendantes ;
(ii) $X_1, \ldots, X_n$ sont deux à deux indépendantes ;
(iii) $X_1, \ldots, X_n$ sont non corrélés ;
(iv) La matrice de covariance Γ de X est diagonale.

Démonstration. Pour n'importe quel vecteur aléatoire, on a (i) ⇒ (ii) ⇒ (iii) ⇒ (iv). Montrons donc que (iv) ⇒ (i). Soit $t \in \mathbf{R}^n$ quelconque,

$$\begin{aligned}\phi_X(t) = \mathbf{E}[e^{i\langle t,X\rangle}] &= \exp\left(i\langle t, m\rangle - \frac{1}{2}t^{\mathrm{t}}\Gamma t\right)\\ &= \exp\left(\sum_{k=1}^{n}\left(it_k m_k - \frac{1}{2}\Gamma_{k,k}t_k^2\right)\right) = \prod_{k=1}^{n}\exp\left(it_k m_k - \frac{1}{2}\Gamma_{k,k}t_k^2\right)\\ &= \prod_{k=1}^{n}\phi_{X_k}(t_k).\end{aligned}$$

Cependant, $\prod_{k=1}^{n}\phi_{X_k}(t_k)$ correspond à la fonction caractéristique d'un vecteur aléatoire dont les composantes sont indépendantes et de lois gaussiennes $\mathcal{N}(m_k, \Gamma_{k,k})$. Comme la fonction caractéristique caractérise la loi, $X_1, \ldots, X_k$ sont indépendantes. □

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 $\mathbf{E}[X] = \mathbf{E}[Y] = 0$, $\mathbf{E}[XY] = \mathbf{E}[\varepsilon]\,\mathbf{E}[X^2] = 0$ et donc $\mathsf{Cov}(X, Y) = 0$. Néanmoins, X et Y ne sont pas indépendantes : par exemple on a

$$\mathbf{P}(X \in [0,1], Y \in [2,3]) = 0 \neq \mathbf{P}(X \in [0,1])\,\mathbf{P}(Y \in [2,3]) > 0.$$

10.6 Densité

Proposition 10.6.0.1. Si $Z \sim \mathcal{N}(0, I_n)$ alors Z est un vecteur à densité, de densité

$$f_Z(x) = \frac{1}{(2\pi)^{n/2}}\exp\left(-\frac{1}{2}\|x\|^2\right) \quad \text{avec} \quad x \in \mathbf{R}^n\,.$$

Démonstration. $Z \sim \mathcal{N}(0, I_n)$, donc $Z_1, \ldots, Z_n$ sont indépendantes et de même loi $\mathcal{N}(0,1)$. Ainsi, Z est un vecteur à densité et sa densité vaut

$$f_Z(x_1, \ldots, x_n) = \prod_{i=1}^{n} f_{Z_i}(x_i) = \frac{1}{(2\pi)^{n/2}}\exp\left(-\frac{1}{2}(x_1^2 + \cdots + x_n^2)\right).$$

□

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 10.6.0.2. Soit Γ une matrice de covariance. Il existe une matrice A de même taille telle que $\Gamma = A\,{}^{\mathrm{t}}A$.

Démonstration. 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 $\Gamma = PD\,{}^{\mathrm{t}}P$. On pose

$$D^{1/2} = \begin{pmatrix} \sqrt{d_1} & & 0 \\ & \ddots & \\ 0 & & \sqrt{d_n} \end{pmatrix}$$

et $A = PD^{1/2}\,{}^{\mathrm{t}}P$ de telle sorte que $\Gamma = A\,{}^{\mathrm{t}}A$. □

Proposition 10.6.0.3. Si $X \sim \mathcal{N}(m, \Gamma)$, $Z \sim \mathcal{N}(O, \mathrm{I}_n)$ et A est une matrice de même taille que Γ telle que $A\,{}^{\mathrm{t}}A = \Gamma$, alors $AZ + m$ a même loi que X.

Démonstration. D'après la proposition 10.4.0.1, $AZ + m$ est un vecteur gaussien de moyenne $A0 + m = m$ et de matrice de covariance $A\,\mathrm{I}_n\,{}^{\mathrm{t}}A = \Gamma$. □

Théorème 10.6.0.4. Si $X \sim \mathcal{N}(m, \Gamma)$, alors X est un vecteur à densité si et seulement si Γ est inversible. Dans ce cas, sa densité f s'écrit

$$f(x) = \frac{1}{(2\pi)^{n/2}\sqrt{\mathsf{det}(\Gamma)}} \exp\left(-\frac{1}{2}\,{}^{\mathrm{t}}(x-m)\Gamma^{-1}(x-m)\right) \quad \text{avec} \quad x \in \mathbf{R}^n\,.$$

Démonstration. D'après la proposition précédente, il suffit d'étudier le vecteur aléatoire $Y = AZ + m$ avec $A\,{}^{\mathrm{t}}A = \Gamma$ et $Z \sim \mathcal{N}(0, \mathrm{I}_n)$. La variable Z est à densité (sur $\mathbf{R}^n$) et Y est une transformation affine de Z. En particulier, Y prend ses valeurs dans le sous-espace affine $E = \{Az + m;\ z \in \mathbf{R}^n\}$ de $\mathbf{R}^n$ qui est de dimension $\mathrm{rg}(A)$. On remarque que $\mathrm{rg}(A) = \mathrm{rg}(\Gamma)$. 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 $h\colon \mathbf{R}^n \to \mathbf{R}$ continue bornée quelconque.

$$\mathbf{E}[h(X)] = \mathbf{E}[h(AZ+m)] = \int_{\mathbf{R}^n} h(az+m)\frac{1}{(2\pi)^{n/2}} \exp\left(-\frac{{}^{\mathrm{t}}zz}{2}\right) \mathrm{d}z_1 \cdots \mathrm{d}z_n.$$

On fait le changement de variable suivant

$$\begin{aligned} \phi\colon \mathbf{R}^n &\to \mathbf{R}^n \\ z &\mapsto Az + m \end{aligned}$$

$$\begin{aligned} \phi^{-1}\colon \mathbf{R}^n &\to \mathbf{R}^n \\ y &\mapsto A^{-1}(y-m) \end{aligned}$$

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

$$\mathrm{Jac}(\phi^{-1}) = \mathsf{det}(A^{-1}) = (\mathsf{det}(\Gamma^{-1}))^{1/2} = (\mathsf{det}(\Gamma))^{-n/2}$$

et ${}^{\mathrm{t}}zz = {}^{\mathrm{t}}(y-m)\,{}^{\mathrm{t}}(A^{-1})A^{-1}(y-m) = {}^{\mathrm{t}}(y-m)(A\,{}^{\mathrm{t}}A)^{-1}(y-m) = {}^{\mathrm{t}}(y-m)\Gamma^{-1}(y-m)$. Ainsi,

$$\mathbf{E}[h(X)] = \int_{\mathbf{R}^n} h(y)\frac{1}{(2\pi)^{n/2}} \exp\left(-\frac{{}^{\mathrm{t}}(y-m)\Gamma^{-1}(y-m)}{2}\right)\frac{1}{(\mathsf{det}(\Gamma))^{1/2}}\,\mathrm{d}y_1 \cdots \mathrm{d}y_n$$

ce qui permet de conclure. □