MASTER et MAGISTÈRE DE MATHÉMATIQUES 1ère année PROBABILITÉS DE BASE

Mihai Gradinaru

2010-2011

Table des matières

Chapitre 1

Espace de probabilité

1.1 Tribus

Une **expérience aléatoire** se décrit mathématiquement par la donnée d'un ensemble Ω (**univers**) dont les éléments notés ω sont les **résultats** (ou **issues)** possibles de l'expérience. Un **événement aléatoire** lié à l'expérience peut être représenté par une partie de Ω. Il sera toujours représenté par l'ensemble des résultats ω de l'expérience qui le réalisent. A priori il pourrait sembler naturel de considérer que toute partie de Ω représente un événement, mais cela n'est possible que si Ω est dénombrable. Pour des espaces plus grands, $\Omega = \mathbb{R}$ (ou $\mathbb{R}^d$ ou un espace métrique E) on a besoin de la notion de tribu.

La description d'un événement comme une partie de Ω est à l'origine de la notation ensembliste. Le **contraire** de l'événement A (qui est réalisé si A ne l'est pas) correspond au complémentaire d'un ensemble $A \subset \Omega$ qui sera noté $A^c = \Omega \setminus A$. L'événement A **ou** B (qui est réalisé si au moins un des deux événements A ou B est réalisé) correspond à la réunion $A \cup B$. L'événement A **et** B (qui est réalisé si les événements A et B sont réalisés à la fois) correspond à l'intersection $A \cap B$. L'événement A **implique** l'événement B si A ne peut être réalisé sans que B le soit aussi et on note $A \subset B$. L'événement **impossible** sera noté $\emptyset$ et l'événement **certain** sera noté Ω. A et B sont **incompatibles** si $A \cap B = \emptyset$.

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

Définition 1.1 *Un sous-ensemble $\mathcal{A}$ de $\mathcal{P}(\Omega)$ est une* **tribu** *(ou σ-***algèbre***) sur Ω si*

a) $\Omega \in \mathcal{A}$

b) $\mathcal{A}$ est stable par passage au complémentaire $A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$)

c) $\mathcal{A}$ est stable par réunion dénombrable ($A_j \in \mathcal{A}$, $j \in \mathbb{N} \Rightarrow \cup_{j\in\mathbb{N}} A_j \in \mathcal{A}$).

Le couple $(\Omega, \mathcal{A})$ *est un* **espace mesurable**.

On appelle **événement** tout élément de la tribu $\mathcal{A}$. Si $\{\omega\} \in \mathcal{A}$ on appelle $\{\omega\}$ **événement élémentaire**.

Remarques : i) Par passage au complémentaire une tribu est aussi stable par intersection dénombrable.

ii) En remplaçant l'axiome *(c)* par
(c') $\mathcal{A}$ est stable par réunion finie ($A_1, \ldots, A_n \in \mathcal{A} \Rightarrow A_1 \cup \ldots \cup A_n \in \mathcal{A}$)
on obtient la définition d'une **algèbre**. Toute tribu est une algèbre. □

Exemples : i) $\mathcal{P}(\Omega)$ est toujours une tribu.
ii) $\{\emptyset, \Omega\}$ est une tribu appelée **tribu triviale**.
iii) La famille des ouverts de $\mathbb{R}^d$ n'est pas une tribu car le complémentaire d'un ouvert n'est pas nécessairement un ouvert.
iv) Une réunion de deux tribus n'est pas une tribu en général. En effet, soit $\Omega = \{0,1,2\}$, $\mathcal{A}_1 = \{\emptyset, \{0\}, \{1,2\}, \Omega\}$ et $\mathcal{A}_2 = \{\emptyset, \{1\}, \{0,2\}, \Omega\}$. La réunion de $\{0\}$ et $\{1\}$ n'appartient pas à $\mathcal{A}_1 \cup \mathcal{A}_2$.
v) Une intersection d'un nombre quelconque de tribus est une tribu.

En général il n'est pas facile de décrire tous les éléments d'une tribu ; on utilise le plus souvent leurs générateurs.

Définition 1.2 *Soit $\mathcal{E}$ un sous-ensemble de $\mathcal{P}(\Omega)$. La* **tribu** $\sigma(\mathcal{E})$ **engendrée par** $\mathcal{E}$ *est l'intersection de toutes les tribus contenant $\mathcal{E}$; elle est donc la plus petite tribu contenant $\mathcal{E}$.*

REMARQUE : La tribu engendrée par deux tribus $\mathcal{A}_1$ et $\mathcal{A}_2$ est notée $\mathcal{A}_1 \vee \mathcal{A}_2 = \sigma(\mathcal{A}_1, \mathcal{A}_2) = \sigma(\mathcal{A}_1 \cup \mathcal{A}_2)$ qui est en général différente de $\mathcal{A}_1 \cup \mathcal{A}_2$. □

Exemple : Soit A un sous-ensemble strict de Ω qui n'est pas vide. La tribu $\sigma(\{A\}) = \{\emptyset, A, A^c, \Omega\}$.

Définition 1.3 *Si $\Omega = E$ est un espace métrique, on appelle* **tribu borélienne***, notée $\mathcal{B}(E)$, la tribu engendrée par les ouverts de E. Tout élément de cette tribu est appellé* **borélien***.*

REMARQUE : La tribu borélienne est aussi engendrée par les fermés. Sur $\mathbb{R}$ la tribu borélienne coïncide avec la tribu engendrée par les intervalles $]a,b[$ ou $[a,b]$, ou $]a,b]$, ou $[a,b[$, $-\infty \leqslant a < b \leqslant \infty$. □

Par la suite, lorsque Ω est $\mathbb{R}$ (ou $\mathbb{R}^d$ ou un espace métrique E), il sera toujours muni de sa tribu borélienne. Si Ω est discret, on le munira de la tribu de ses parties.

Définition 1.4 *Soient $(\Omega_i, \mathcal{A}_i)$, $i = 1,2$, deux espaces mesurables. On appelle* **ensemble élémentaire** *de $\Omega = \Omega_1 \times \Omega_2$ une réunion finie de* **pavés** *$A_1 \times A_2$, avec $A_i \in \mathcal{A}_i$, $i = 1,2$. La* **tribu produit** *$\mathcal{A}_1 \otimes \mathcal{A}_2$ sur Ω est la tribu engendrée par les ensembles élémentaires.*

REMARQUE* : En utilisant que tout ouvert de $\mathbb{R}^2$ peut s'écrire comme un réunion dénombrable de pavés d'intervalles ouverts, on montre que $\mathcal{B}(\mathbb{R}^2) = \mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R})$. □

1.2 Variables aléatoires

Une variable aléatoire sera définie par référence à une expérience aléatoire, comme une variable X dont la valeur dépend du résultat ω de cette expérience. Elle est donc une fonction définie sur l'univers Ω associé à l'expérience.

Pour définir une variable aléatoire on introduit d'abord quelques notations. Si X est une application de Ω dans E et si B est une partie de E on notera $X^{-1}(B) := \{\omega \in \Omega : X(\omega) \in B\}$. Si $\mathcal{B}$ est une famille de parties de E, on notera $X^{-1}(\mathcal{B}) := \{X^{-1}(B) : B \in \mathcal{B}\}$.

Définition 1.5 *a) Soient $(\Omega, \mathcal{A})$, $(E, \mathcal{B})$ deux espaces mesurables et $X : \Omega \to E$. On dit que X est* **mesurable** *(pour $\mathcal{A}$ et $\mathcal{B}$) si $X^{-1}(\mathcal{B}) \subset \mathcal{A}$ (c'est-à-dire, pour tout $B \in \mathcal{B}$, $X^{-1}(B) \in \mathcal{A}$). Lorsque les deux espaces de départ et d'arrivée sont des espaces métriques munis de leurs tribus boréliennes la fonction est appelée* **borélienne**.

b) Soient $(E, \mathcal{B})$ un espace mesurable et $X : \Omega \to E$. On appelle **tribu engendrée par** *X (sur Ω) la plus petite tribu qui rend X mesurable : $\sigma(X) := \{X^{-1}(B) : B \in \mathcal{B}\} = X^{-1}(\mathcal{B})$.*

c) Soit $\{X_t : t \in T\}$ une famille de fonctions de Ω dans un espace mesurable $(E, \mathcal{B})$. La **tribu engendrée par** *$\{X_t : t \in T\}$ est la plus petite tribu qui rend mesurable toute fonction de $\{X_t : t \in T\}$: $\sigma(X_t : t \in T) := \sigma(\cup_{t\in T}\sigma(X_t))$.*

REMARQUE : Dire que X est mesurable revient à dire que $\sigma(X) \subset \mathcal{A}$. □

Exemples : i) Pour A une partie de Ω, on définit la **fonction indicatrice** de A par

$$\mathbf{1}_A(\omega) := \begin{cases} 1, & \text{si } \omega \in A \\ 0, & \text{si } \omega \notin A. \end{cases}$$

La fonction $\mathbf{1}_A$ est mesurable pour $\mathcal{A}$ (en tant que fonction à valeurs dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$) si et seulement si $A \in \mathcal{A}$.
ii) La tribu $\mathcal{B}(\mathbb{R}^2)$ est engendrée par les projections Π_1 et Π_2 sur les coordonnées ($\Pi_1(x, y) = x$ et $\Pi_2(x, y) = y$).

Définition 1.6 *Une fonction mesurable X définie sur Ω muni de la tribu $\mathcal{A}$ à valeurs dans $\mathbb{R}$ (ou $\mathbb{R}^d$ ou $\mathbb{R}^\infty$) muni de sa tribu borélienne est appellée* **variable aléatoire réelle** *(ou* **vecteur aléatoire** *ou encore* **suite aléatoire***).*

En pratique, lorsque la tribu de l'espace d'arrivée est engendrée par un système de générateurs, pour vérifier qu'une fonction est mesurable (en particulier, variable aléatoire), il suffit de vérifier la propriété caractéristique sur les générateurs.

Proposition 1.1 *1) Soient X de $(\Omega, \mathcal{A})$ dans $(E, \mathcal{B})$, et $\mathcal{C} \subset \mathcal{P}(E)$. On suppose que la tribu $\mathcal{B}$ sur E est engendrée par $\mathcal{C}$, $\mathcal{B} = \sigma(\mathcal{C})$. Alors*

$$\sigma(X) = X^{-1}(\mathcal{B}) = X^{-1}(\sigma(\mathcal{C})) = \sigma(X^{-1}(\mathcal{C})) := \sigma(\{X^{-1}(C) : C \in \mathcal{C}\}).$$

2) *En particulier, pour qu'une fonction X de $(\Omega, \mathcal{A})$ dans $(E, \sigma(\mathcal{C}))$ soit mesurable, il suffit que $X^{-1}(\mathcal{C}) \subset \mathcal{A}$.*

Preuve : Il est clair que $X^{-1}(\sigma(\mathcal{C}))$ est une tribu contenant $X^{-1}(\mathcal{C})$, d'où $X^{-1}(\sigma(\mathcal{C})) \supset \sigma(X^{-1}(\mathcal{C})$. Pour l'autre inclusion, soit

$$\mathcal{T} = \{B \subset E : X^{-1}(B) \in \sigma(X^{-1}(\mathcal{C}))\}.$$

On peut vérifier facilement que $\mathcal{T}$ est une tribu. Par sa définition $X^{-1}(\mathcal{T}) \subset \sigma(X^{-1}(\mathcal{C}))$. De plus $\mathcal{C} \subset \mathcal{T}$ car $X^{-1}(\mathcal{C}) \subset \sigma(X^{-1}(\mathcal{C}))$, et donc $\sigma(\mathcal{C}) \subset \mathcal{T}$. On déduit que $X^{-1}(\sigma(\mathcal{C})) \subset X^{-1}(\mathcal{T}) \subset \sigma(X^{-1}(\mathcal{C}))$. Notons qu'on peut traiter de la même façon le cas d'une famille quelconque de fonctions.

Si $X^{-1}(\mathcal{C}) \subset \mathcal{A}$, alors $\sigma(X^{-1}(\mathcal{C})) \subset \mathcal{A}$. Comme $\sigma(X) = \sigma(X^{-1}(\mathcal{C}))$ par le premier point, la conclusion s'ensuit. □

Remarque : Une fonction $X : \Omega \to \mathbb{R}$ est une variable aléatoire réelle si et seulement si pour tout réel t, $\{X \leqslant t\} \in \mathcal{A}$. □

Proposition 1.2 *La composée de deux fonctions mesurables est mesurable. En particulier, la composée $Y = g(X)$ d'une fonction borélienne g avec une une variable aléatoire X est une variable aléatoire.*

Preuve : Soient $X_j : (\Omega_j, \mathcal{A}_j) \to (\Omega_{j+1}, \mathcal{A}_{j+1})$, $j = 1, 2$. Si $A \in \mathcal{A}_3$, $(X_2 \circ X_1)^{-1}(A) = X_1^{-1}(X_2^{-1}(A))$. Comme X_2 est mesurable, $X_2^{-1}(A) \in \mathcal{A}_2$, et comme X_1 est mesurable, $X_1^{-1}(X_2^{-1}(A)) \in \mathcal{A}_1$. □

Proposition 1.3 *Soient E_1 et E_2 deux espaces métriques munis de leurs tribus boréliennes. Toute fonction continue $f : E_1 \to E_2$ est borélienne.*

Preuve : On remarque que si O est un ouvert dans E_2 et f est une fonction continue, $f^{-1}(O)$ est un ouvert. Puis on applique la Proposition 1.1. □

Proposition 1.4 *Si X, Y sont deux variables aléatoires de Ω dans $\mathbb{R}$, alors l'application $\Omega \ni \omega \mapsto (X(\omega), Y(\omega)) \in \mathbb{R}^2$ est un vecteur aléatoire bi-dimensionnel. La reciproque est aussi vraie.*

Preuve : Soit $A \times B$ un pavé dans $\mathcal{B}(\mathbb{R}^2)$ et $Z(\omega) = (X(\omega), Y(\omega))$. Alors, $Z^{-1}(A \times B) = X^{-1}(A) \cap Y^{-1}(B) \in \mathcal{A}$. Comme les pavés engendrent $\mathcal{B}(\mathbb{R}^2)$, on conclut grâce à la Proposition 1.1.(voir aussi la Proposition 1.7). Pour la réciproque il suffit de voir que X et Y sont des composées des applications mesurables projections (continues) et Z. □

Proposition 1.5 *L'espace des variables aléatoires réelles est stable pour les opérations suivantes :* $(\alpha X)(\omega) = \alpha X(\omega)$ $(\alpha \in \mathbb{R})$*,* $(X+Y)(\omega) = X(\omega) + Y(\omega)$*,* $(XY)(\omega) = X(\omega)Y(\omega)$ *et* $(X \vee Y)(\omega) = X(\omega) \vee Y(\omega)$ *(le maximum de* X, Y*).*

Preuve : La fonction $\omega \mapsto \alpha X(\omega)$ est la composée de la variable aléatoire X et la fonction continue $x \mapsto \alpha x$. De même, $X+Y$ (respectivement XY, respectivement $X \vee Y$) est la composée du vecteur aléatoire (X, Y) (en vertu de la Proposition 1.3) et de la fonction continue $(x, y) \mapsto x + y$ (respectivement $(x, y) \mapsto xy$, respectivement $(x, y) \mapsto x \vee y$). □

Une limite ponctuelle de fonctions continues n'est pas nécessairement continue. Pour les fonctions mesurables (et en particulier pour les variables aléatoires) on peut montrer le suivant :

Théorème 1.1 *Soit* $(X_n)_{n\in\mathbb{N}}$ *une suite de variables aléatoires de* Ω *dans un espace métrique* E*. Si cette suite de fonctions converge ponctuellement vers* X *(c'est-à-dire, pour tout* $\omega \in \Omega$*,* $\lim_{n\to\infty} X_n(\omega) = X(\omega)$*), alors* X *est une variable aléatoire à valeurs dans* E*.*

Preuve : D'après la Proposition 1.1, il suffit de montrer que si O est un ouvert dans E, alors $X^{-1}(O) \in \mathcal{A}$. On pose

$$O_r := \left\{ x \in O : d(x, E \setminus O) > \frac{1}{r} \right\},\ r \in \mathbb{N}^*.$$

L'ensemble O_r est ouvert, donc un borélien de E. Ainsi,

$$X^{-1}(O) = \bigcup_{r,m\in\mathbb{N}^*} \bigcap_{n\geqslant m} X_n^{-1}(O_r)$$

est un événement de $\mathcal{A}$. □

Définition 1.7 *On appelle* **variable aléatoire étagée ou simple** *(à valeurs dans* $\mathbb{R}^d$*) une variable aléatoire définie sur* Ω *de la forme* $X = \sum_{j=1}^{r} a_j \mathbf{1}_{A_j}$ *où les* $A_1, \ldots, A_r$ *sont des événements disjoints dans* $\mathcal{A}$*, et où les coefficients* $a_1, \ldots, a_r \in \mathbb{R}^d$*.*

Proposition 1.6 *Toute variable aléatoire* X *est limite simple de variables aléatoires étagées. Si de plus* X *est une variable aléatoire positive, la limite peut être choisie croissante.*

Preuve : Soit d'abord X positive. On définit, pour $n, k \in \mathbb{N}^*$,

$$A_{n,k} := \left\{ \omega : \frac{k-1}{2^n} \leqslant X(\omega) < \frac{k}{2^n} \right\}$$

et

$$B_n := \{\omega : X(\omega) > n\}.$$

Les $A_{n,k}$ et les B_n sont éléments de $\mathcal{A}$ en tant qu'images réciproques par la variable aléatoire X d'intervalles. La suite

$$X_n(\omega) := \sum_{k=1}^{n2^n} \frac{k-1}{2^n} \mathbb{1}_{A_{n,k}}(\omega) + n\mathbb{1}_{B_n}(\omega)$$

converge en croissant vers $X(\omega)$.

Si X est quelconque, on écrit $X = X^+ - X^-$ avec $X^+ = X \vee 0$ et $X^- = (-X) \vee 0$, et on approxime les variables positives X^+ et X^- par la méthode précédente. □

1.3 Classes monotones

Dans cette section on va construire un procédé d'extension des définitions de certains objets sur les tribus après les avoir définis sur des classes resteintes d'ensembles.

Définition 1.8 *Une famille $\mathcal{M}$ de parties de Ω est appelée* **classe monotone** *si*

a) $\Omega \in \mathcal{M}$

b) lorsque $A, B \in \mathcal{M}$ et $B \subset A$, alors $A \setminus B \in \mathcal{M}$

c) $\mathcal{M}$ est stable par réunion monotone croissante ($A_j \in \mathcal{M}, j \in \mathbb{N}, A_j \subset A_{j+1} \Rightarrow \cup_{j\in\mathbb{N}} A_j \in \mathcal{M}$). Pour $\mathcal{E} \subset \mathcal{P}(\Omega)$, on note $\mathcal{M}(\mathcal{E})$ la **classe monotone** *engendrée par $\mathcal{E}$, c'est-à-dire l'intersection de toutes les classes monotones contenant $\mathcal{E}$.*

Remarque : Une intersection d'un nombre quelconque de classes monotones est une classe monotone. □

Exemples : i) Une tribu est une classe monotone. En effet, pour voir cela, il suffit de voir que $A \setminus B = A \cap B^c$.
ii) Une classe monotone stable par intersection finie est une tribu. En effet cette classe sera aussi stable par réunion finie en vertu de l'axiome (b) de la Définition 8, et toute réunion peut s'écrire comme une réunion croissante ($\cup_{j\in\mathbb{N}} A_j = \cup_{j\in\mathbb{N}}(\cup_{k\leqslant j} A_k)$, pour toute famille A_j, $j \in \mathbb{N}$).

Théorème 1.2 *(des classes monotones)*
Soit $\mathcal{E}$ une famille de parties de Ω stable par intersection finie. Alors $\mathcal{M}(\mathcal{E}) = \sigma(\mathcal{E})$.

Remarque : On pourrait enoncer ce résultat sous la forme suivante : si $\mathcal{M}_1$ est une classe monotone contenant la famille de parties $\mathcal{E}$ (stable par intersection finie), alors $\mathcal{M}_1 \supset \sigma(\mathcal{E})$.

Preuve : En vertu de l'exemple i) ci-dessus, $\sigma(\mathcal{E})$ est une classe monotone qui contient $\mathcal{E}$ et donc $\mathcal{M}(\mathcal{E}) \subset \sigma(\mathcal{E})$. Pour prouver l'inclusion inverse, on montre que $\mathcal{M}(\mathcal{E})$ est stable par intersection finie (car alors, d'après l'exemple ii), $\mathcal{M}(\mathcal{E})$ sera une tribu contenant $\mathcal{E}$, et donc $\sigma(\mathcal{E}) \subset \mathcal{M}(\mathcal{E})$). Il suffit de prouver que si $A, B \in \mathcal{M}(\mathcal{E})$, alors $A \cap B \in \mathcal{M}(\mathcal{E})$. Soit

$$\mathcal{M}_1 := \{A \in \mathcal{M}(\mathcal{E}) : \forall B \in \mathcal{E},\ A \cap B \in \mathcal{M}(\mathcal{E})\}.$$

L'ensemble $\mathcal{M}_1$ est une classe monotone qui contient $\mathcal{E}$, donc $\mathcal{M}(\mathcal{E})$. Soit

$$\mathcal{M}_2 := \{A \in \mathcal{M}(\mathcal{E}) : \forall B \in \mathcal{M}(\mathcal{E}),\, A \cap B \in \mathcal{M}(\mathcal{E})\}.$$

L'ensemble $\mathcal{M}_2$ est une classe monotone. De plus il contient $\mathcal{E}$: on doit pour cela montrer que si $B \in \mathcal{E}$, alors $\forall C \in \mathcal{M}(\mathcal{E})$, $B \cap C \in \mathcal{M}(\mathcal{E})$. Or $C \in \mathcal{M}(\mathcal{E}) \subset \mathcal{M}_1$, donc puisque $B \in \mathcal{E}$, $B \cap C = C \cap B \in \mathcal{M}(\mathcal{E})$. Ainsi, $\mathcal{M}_2 \supset \mathcal{E}$, donc $\mathcal{M}_2 \supset \mathcal{M}(\mathcal{E})$, ce qui montre que $\mathcal{M}(\mathcal{E})$ est stable par intersection finie. Le théorème est prouvé. □

Proposition 1.7 *Soit $X : \Omega \to \mathbb{R}^d$ une fonction vectorielle $X(\omega) = (X_1(\omega), \ldots, X_d(\omega))$ sur l'espace mesurable $(\Omega, \mathcal{A})$. X est un vecteur aléatoire si et seulement si chaque coordonnée X_j est une variable aléatoire réelle.*

Preuve : On va faire la preuve pour $d = 2$. On suppose d'abord que X est un vecteur aléatoire et soit $A \in \mathcal{B}(\mathbb{R})$. Alors

$$X_1^{-1}(A) = \{\omega \in \Omega : X_1(\omega) \in A\}$$
$$= \{\omega \in \Omega : (X_1(\omega), X_2(\omega)) \in A \times \mathbb{R}\} = X^{-1}(A \times \mathbb{R}),$$

donc X_1 est une variable aléatoire. De la même façon on montre que X_2 est une variable aléatoire.

Réciproquement, supposons que X_1 et X_2 sont deux variables aléatoires réelles. Soient A, $B \in \mathcal{B}(\mathbb{R})$. On a

$$X^{-1}(A \times B) = \{\omega \in \Omega : X_1(\omega) \in A,\, X_2(\omega) \in B\}$$

$$= X_1^{-1}(A) \cap X_2^{-1}(B) \in \mathcal{A}.$$

Cette propriété d'appartenance s'étend à l'ensemble $\mathcal{E}$ des réunions finies de pavés de $\mathbb{R}^2$, deux à deux disjoints, laquelle engendre la tribu borélienne produit $\mathcal{B}(\mathbb{R}^2)$. De plus $\mathcal{E}$ est stable par intersection finie. Donc $\mathcal{M}(\mathcal{E}) = \sigma(\mathcal{E}) = \mathcal{B}(\mathbb{R}^2)$. On introduit

$$\mathcal{M}_1 = \{C \in \mathcal{B}(\mathbb{R}^2) : X^{-1}(C) \in \mathcal{A}\}$$

Puisque, $X^{-1}(\cup_{j \in \mathbb{N}} C_j) = \cup_{j \in \mathbb{N}} X^{-1}(C_j)$ pour toute famille C_j, $j \in \mathbb{N}$, on peut montrer que $\mathcal{M}_1$ est une classe monotone. De plus $\mathcal{M}_1$ contient $\mathcal{E}$, donc $\mathcal{M}_1 \supset \mathcal{M}(\mathcal{E}) = \mathcal{B}(\mathbb{R}^2)$. Par conséquent X est un vecteur aléatoire (voir aussi la Proposition 1.4). □

1.4 Probabilités

Définition 1.9 *Soit $(\Omega, \mathcal{A})$ un espace mesurable. On appelle* **probabilité** *toute application $\mathrm{P} : \mathcal{A} \to [0, 1]$ telle que*

a) $\mathrm{P}(\Omega) = 1$

b) P *est σ-additive ($A_j \in \mathcal{A}$, $j \in \mathbb{N}$, $A_i \cap A_j = \emptyset$ si $i \neq j \Rightarrow \mathrm{P}(\cup_{j \in \mathbb{N}} A_j) = \sum_{j \in \mathbb{N}} \mathrm{P}(A_j)$).*

Un espace mesurable muni d'une probabilité $(\Omega, \mathcal{A}, \mathrm{P})$ est appelé **espace de probabilité**.

REMARQUE : Observons que $\emptyset$ est la réunion dénombrable et disjointe de l'ensemble vide, donc $\mathrm{P}(\emptyset) = \sum_{j\in\mathbb{N}} \mathrm{P}(\emptyset)$, d'où $\mathrm{P}(\emptyset) = 0$. Observons aussi que $\Omega = A \cup A^c \cup \emptyset \cup \ldots$, d'où $\mathrm{P}(A^c) = 1 - \mathrm{P}(A)$. De la même façon on peut prouver que P est (finie) additive. $\square$

Exemples : i) Soit $(\Omega, \mathcal{A})$ un espace mesurable et $\omega \in \Omega$. L'application

$$\delta_\omega : \mathcal{A} \to \{0,1\},\ \delta_\omega(A) := \mathbf{1}_A(\omega)$$

est une probabilité, appelée **masse de Dirac en** ω.
ii) Soit $\Omega = \{1,2,3,4,5,6\}$ muni de la tribu des parties. On voit que $\mathrm{P}(A) := \mathrm{card}(A)/6$ est une probabilité. On peut voir que $\mathrm{P} := \frac{1}{6}\sum_{j=1}^{6}\delta_j$. Cette probabilité sert à modéliser le jet du dé.
iii) Soit $0 < p < 1$ et $\Omega = \{0,1\}$. La probabilité $\mathrm{P} := p\delta_1 + (1-p)\delta_0$ est appelée **probabilité de Bernoulli de paramètre** p. Lorsque $p = 1/2$ cette probabilité sert à modéliser le jet d'une piece (jeu de pile ou face). Plus généralement, toute probabilité concentrée en deux points distincts sera appelée probabilité de Bernoulli.

Proposition 1.8 *Soit* $(\Omega, \mathcal{A}, \mathrm{P})$ *un espace de probabilité et* $(A_j)_{j\in J}$, $J \subset \mathbb{N}$ *une famille finie ou dénombrable d'événements.*

1) P *est croissante :* $A_1 \subset A_2 \Rightarrow \mathrm{P}(A_1) \leqslant \mathrm{P}(A_2)$ *;*

2) $\mathrm{P}(A_1 \cup A_2) = \mathrm{P}(A_1) + \mathrm{P}(A_2) - \mathrm{P}(A_1 \cap A_2)$ *;*

3) Si $A_j \subset A_{j+1}$ *pour tout* j*, alors* $\mathrm{P}(\cup_j A_j) = \lim_{j\to\infty} \mathrm{P}(A_j)$ *;*

4) Si $A_j \supset A_{j+1}$ *pour tout* j*, alors* $\mathrm{P}(\cap_j A_j) = \lim_{j\to\infty} \mathrm{P}(A_j)$ *;*

5) P *est sous-additive :* $\mathrm{P}(\cup_{j\in J} A_j) \leqslant \sum_{j\in J} \mathrm{P}(A_j)$.

Preuve : 1) A_2 est la réunion disjointe des événements A_1 et $A_2 \setminus A_1$ et la Définition 9 d'une probabilité fournit $\mathrm{P}(A_2) = \mathrm{P}(A_1) + \mathrm{P}(A_2 \setminus A_1) \geqslant \mathrm{P}(A_1)$.

2) De la même façon

$$A_1 \cup A_2 = A_1 \cup (A_2 \setminus A_1),\ A_2 = (A_2 \setminus A_1) \cup (A_1 \cap A_2),$$

et on en obtient l'égalité, puisque les paires d'événements A_1 et $A_2 \setminus A_1$, $A_2 \setminus A_1$ et $A_1 \cap A_2$ sont aussi disjoints.

3) Soit $B_k = A_{k+1} \setminus A_k$, $k \in \mathbb{N}$. Les ensembles B_k sont disjoints et comme $A_j = A_0 \cup \bigcup_{0\leqslant k\leqslant j-1} B_k$, $j \in \mathbb{N}$, on a

$$\bigcup_j A_j = A_0 \cup \bigcup_k B_k.$$

Par σ-additivité on obtient

$$\mathrm{P}\left(\bigcup_j A_j\right) = \mathrm{P}(A_0) + \sum_k \mathrm{P}(B_k) = \mathrm{P}(A_0) + \lim_{j\uparrow\infty} \sum_{k=0}^{j-1} \mathrm{P}(B_k)$$

$$= \lim_{j\uparrow\infty}\left(\mathrm{P}(A_0) + \sum_{k=0}^{j-1} \mathrm{P}(B_k)\right) = \lim_{j\uparrow\infty} \mathrm{P}(A_j).$$

4) Les $B_j = A_j^c$, forment une suite croissante, donc comme toute probabilité est bornée par 1, la suite $\{\mathrm{P}(B_j) : j \in \mathbb{N}\}$ est une suite croissante bornée. Donc la limite $\lim_{j\uparrow\infty} \mathrm{P}(B_j)$ existe et, d'après 3),

$$\lim_{j\uparrow\infty} \mathrm{P}(B_j) = \mathrm{P}\left(\bigcup_j B_j\right) = \mathrm{P}\left(\Omega \setminus \bigcap_j A_j\right)$$

$$= \mathrm{P}(\Omega) - \mathrm{P}\left(\bigcap_j A_j\right).$$

Donc,

$$\mathrm{P}\left(\bigcap_j A_j\right) = 1 - \lim_{j\uparrow\infty} \mathrm{P}(B_j) = \lim_{j\uparrow\infty} [1 - \mathrm{P}(B_j)] = \lim_{j\uparrow\infty} \mathrm{P}(A_j),$$

ce qui prouve l'assertion.

5) Si J est fini, par exemple $J = \{1, 2, \ldots, r\}$, on procède par récurrence en remarquant que, par 2),

$$\mathrm{P}(A_1 \cup A_2) = \mathrm{P}(A_1) + \mathrm{P}(A_2) - \mathrm{P}(A_1 \cap A_2) \leqslant \mathrm{P}(A_1) + \mathrm{P}(A_2).$$

Si J est infini on peut supposer $J = \mathbb{N}$. On a alors, pour tout $k \in \mathbb{N}^*$,

$$\mathrm{P}\left(\bigcup_{0\leqslant j\leqslant k} A_j\right) \leqslant \sum_{j\in\mathbb{N}} \mathrm{P}(A_j).$$

En considérant les ensembles croissants $B_k = \bigcup_{0\leqslant j\leqslant k} A_j$ le résultat se déduit de 3). □

Remarque : Soit $\Omega = \{\omega_0, \omega_1, \ldots\}$ un ensemble (au plus) dénombrable muni de la tribu de ses parties. La formule :

$$\mathrm{P}(A) := \sum_{\{j:\omega_j\in A\}} p_j, \; A \in \mathcal{P}(\Omega)$$

définit une probabilité pourvu que la suite de réels $(p_j)_{j\geqslant 0}$ vérifie les deux conditions suivantes :

$$p_j \geqslant 0, \; j \geqslant 0 \text{ et } \sum_{j\geqslant 0} p_j = 1.$$

Inversement, si P est une probabilité sur $(\Omega, \mathcal{P}(\Omega))$, en posant $p_j := \mathrm{P}(\{\omega_j\})$, on voit que la suite de réels positifs ou nuls $(p_j)_{j\geqslant 0}$ satisfait $\sum_{j\geqslant 0} p_j = \mathrm{P}(\Omega) = 1$, et que pour tout événement A, $\mathrm{P}(A) = \sum_{\omega_j\in A} p_j$.

En particulier, lorsque Ω est fini, si par symétrie (donnée par la nature de l'expérience) tous les p_j sont égaux entre eux, alors nécessairement $p_j = 1/\mathrm{card}(\Omega)$ (les résultats sont équiprobables) et la probabilité correspondante est la probabilité **uniforme** $\mathrm{P}(A) = \mathrm{card}(A)/\mathrm{card}(\Omega)$, $A \in \mathcal{P}(\Omega)$. Cette probabilité uniforme modélise la plupart de jeux de hasard. Les calcul de probabilités d'événements sont des calculs de cardinaux d'ensembles (dénombrement). □

Remarque : Soient $(\Omega, \mathcal{A}) = (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ et f une fonction borélienne $f : \mathbb{R}^d \to \mathbb{R}$. La formule

$$\mathrm{P}(A) := \int_A f(x)\,dx,\ A \in \mathcal{B}(\mathbb{R}^d)$$

(intégrale de Lebesgue) définit une probabilité pourvu que $f \geqslant 0$ et d'intégrale de Lebesgue $\int_{\mathbb{R}^d} f(x)\,dx = 1$. Inversement si la probabilité P sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ est **absolument continue par rapport à la mesure de Lebesgue** λ, c'est-à-dire si $\lambda(A) = 0 \Rightarrow \mathrm{P}(A) = 0$, alors (théorème de Radon-Nikodym)

$$\exists f \geqslant 0,\ \text{borélienne} : \mathrm{P}(A) = \int_A f(x)\,dx,\ \forall A \in \mathcal{B}(\mathbb{R}^d).$$

On dit que f est la **densité** de P par rapport à λ et elle est unique λ-p.p. □

Preuve du théorème de Radon-Nikodym (facultatif).
Sans perte de généralité on peut supposer $d = 1$. Il suffit de prouver le résultat suivant :

(E_1) Soit P *une probabilité sur* $(]0,1], \mathcal{B}(]0,1]))$ *telle qu'elle soit absolument continue par rapport à la mesure de Lebesgue* λ *sur* $]0,1]$*. Alors il existe une fonction (unique* λ*-p.p.) borélienne positive* f *telle que* $\mathrm{P}(A) = \int_A f(x)\lambda(dx)$*, pour tout* $A \in \mathcal{B}(]0,1])$*.*

En effet, on écrit

$$\mathbb{R} = \bigcup_{n\in\mathbb{Z}} I_n,\ \text{ avec }\ I_n =]n, n+1].$$

Alors $\mathrm{P}(\mathbb{R}) = \sum_{n\in\mathbb{Z}} \mathrm{P}(I_n)$. On note $p_n = \mathrm{P}(I_n)$, $n \in \mathbb{Z}$. Alors $\mathrm{P}_n = \frac{1}{p_n}\mathrm{P}$ est une proba sur I_n et elle est absolument continue par rapport à λ. Par (E_1) on déduit l'existence des fonctions $f_n \geqslant 0$ boréliennes telles que $\mathrm{P}_n(A) = \int_A f_n(x)\lambda(dx)$, pour tout $A \in \mathcal{B}(I_n)$, pour tout $n \in \mathbb{Z}$. Il suffit de prendre la fonction borélienne positive $f = \sum_{n\in\mathbb{Z}} \frac{1}{p_n} f_n$ pour avoir la conclusion.

On prouve donc (E_1) ; on commmence par introduire

$$\mathcal{G} := \{g :]0,1] \to]0,\infty[\ \text{ borélienne } : \mathrm{P}(A) \geqslant \int_A g(x)\lambda(dx),\ \forall A \in \mathcal{B}(]0,1])\}.$$

Il est clair que $g \equiv 0$ est dans $\mathcal{G}$. On montre que $\mathcal{G}$ est sup-stable, c'est-à-dire que si $g, h \in \mathcal{G}$ alors $\sup(g,h) \in \mathcal{G}$. En effet, on note $A_1 = \{g \geqslant h\}$ et $A_2 = A_1^c$ et soit $A \in \mathcal{B}(]0,1])$. Alors

$$\int_A \sup(g,h)\,d\lambda = \int_{A\cap A_1} g\,d\lambda + \int_{A\cap A_2} h\,d\lambda \leqslant \mathrm{P}(A\cap A_1) + \mathrm{P}(A\cap A_2) = \mathrm{P}(A).$$

On note

$$\gamma = \sup_{g\in\mathcal{G}} \int_{]0,1]} g\,d\lambda.$$

Il est clair que $\gamma \leqslant 1$, puisque $\int_{]0,1]} g\,d\lambda \leqslant \mathrm{P}(]0,1]) = 1$, pour tout $g \in \mathcal{G}$. Il existe une suite $\{g_n^* : n \geqslant 1\} \subset \mathcal{G}$ telle que $\lim_{n\to\infty} \int_{]0,1]} g_n^* d\lambda = \gamma$. Alors la suite $\{g_n : n \geqslant 1\}$, avec $g_n := \sup(g_1^*, \ldots, g_n^*)$ est aussi dans $\mathcal{G}$. Comme $g_n^* \leqslant g_n$, $\int_{]0,1]} g_n^* d\lambda \leqslant \int_{]0,1]} g_n d\lambda$, pour tout $n \geqslant 1$. On en déduit que $\lim_{n\to\infty} \int_{]0,1]} g_n d\lambda = \gamma$. Comme $\{g_n : n \geqslant 1\}$ est croissante donc par convergence monotone $\int_{]0,1]}(\lim_{n\to\infty} g_n)d\lambda = \gamma$. Notons $f = \sup_{n\geqslant 1} g_n \in \mathcal{G}$. Comme $\int_{]0,1]} f\,d\lambda = \gamma$, l'application $g \mapsto \int_{]0,1]} g\,d\lambda$ atteint son maximum sur $\mathcal{G}$ en f. Montrons que $\mathrm{P}(A) = \int_A f\,d\lambda$. Il est clair que $\mathrm{P}(A) \geqslant \int_A f\,d\lambda$, car $f \in \mathcal{G}$. On définit

$$\nu(A) = \mathrm{P}(A) - \int_A f\,d\lambda.$$

Si $\lambda(A) = 0$ alors $\mathrm{P}(A) = 0$, par hypothèse et aussi $\int_A f\,d\lambda = 0$, donc $\nu(A) = 0$. Montrons que $\nu \equiv 0$. Supposons $\nu(]0,1]) > 0$ et on note $0 < \beta = \frac{1}{2}\nu(]0,1])$. Alors $\nu(]0,1]) = 2\beta > \beta = \beta\lambda(]0,1])$.On a besoin du résultat suivant :

(E_2) Soient μ et ν deux mesures finies sur un espace mesurable $(\Omega,\mathcal{A})$ telles que $\mu(\Omega) < \nu(\Omega)$. Alors il existe $A' \in \mathcal{A}$ tel que $\mu(A') < \nu(A')$ et $\mu(A'') \leqslant \nu(A'')$ pour tous $A'' \in A' \cap \mathcal{A}$.

De ce résultat (avec ν et $\mu = \beta\lambda$) on déduit qu'il existe $A' \in \mathcal{B}(]0,1])$ tel que $\nu(A') > \beta\lambda(A')$ et $\nu(A'') \geqslant \beta\lambda(A'')$, pour tous $A'' \in A' \cap \mathcal{B}(]0,1])$. On note $f_0 = f + \beta \mathbb{1}_{A'}$. Alors

$$\int_A f_0\,d\lambda = \int_A f\,d\lambda + \beta\lambda(A\cap A') \leqslant \int_A f\,d\lambda + \nu(A) = \mathrm{P}(A), \forall A \in \mathcal{B}(]0,1]),$$

donc $f_0 \in \mathcal{G}$. De plus

$$\int_{]0,1]} f_0\,d\lambda = \int_{]0,1]} f\,d\lambda + \beta\lambda(A') = \gamma + \beta\lambda(A') > \gamma,$$

puisque $\lambda(A') > 0$ (par $\nu(A') > \beta\lambda(A')$ et absolue continuité de ν par rapport à λ). Mais l'inégalité $\int_{]0,1]} f_0\,d\lambda > \gamma$ contredit le choix de γ. L'hypothèse $\nu(]0,1]) > 0$ est donc fausse. Pour finir la preuve il faut justifier (E_2).

La fonction $\delta = \nu - \mu$ est bornée sur $\mathcal{A}$ puisque $-\mu(\Omega) \leqslant \delta(A) \leqslant \nu(\Omega)$, pour tous $A \in \mathcal{A}$. Par récurrence on définit $\{B_n : n \geqslant 0\}$, $\{A_n : n \geqslant 0\}$:

$$B_0 := \emptyset,\ A_0 := \Omega = \Omega \setminus B_0$$

Si $B_0, \ldots, B_n$ et $A_0, \ldots, A_n$ sont définis, on pose

$$\alpha_n := \inf_{A \in A_n \cap \mathcal{A}} \delta(A).$$

Ainsi $\alpha_n \leqslant \delta(\emptyset) = 0$. Dans le cas $\alpha_n = 0$, on pose $B_{n+1} := \emptyset$ et $A_{n+1} := A_n = A_n \setminus B_{n+1}$. Si $\alpha_n < 0$, on choisit $B_{n+1} \in A_n \cap \mathcal{A}$ tel que $\delta(B_{n+1}) \leqslant \frac{1}{2}\alpha_n$ et on pose $A_{n+1} := A_n \setminus B_{n+1}$. La suite $\{B_n : n \geqslant 0\}$ est disjointe, donc la série $\sum_{n\geqslant 0} \delta(B_n)$ est convergente (vers $\nu(\cup B_n) - \mu(\cup B_n)$). Ainsi $\lim_{n\to\infty} \delta(B_n) = 0$, d'où $\lim_{n\to\infty} \alpha_n = 0$. On pose

$$A' := \bigcup_{n\geqslant 0} A_n.$$

D'après les propriétés de continuité des mesures μ et ν similaires aux celles des probabilités, comme $\{A_n : n \geqslant 0\}$ est décroissante,

$$\delta(A') = \lim_{n\to\infty} \delta(A_n).$$

Montrons que A' satisfait aux propriétés. Comme $\delta(B_{n+1}) \leqslant 0$, pour tout n on a

$$\delta(A_{n+1}) = \delta(A_n) - \delta(B_{n+1}) \geqslant \delta(A_n) \geqslant \delta(A_{n-1}) \geqslant \ldots \geqslant \delta(A_0) = \delta(\Omega) > 0,$$

d'où $\nu(A') - \mu(A') = \delta(A') = \lim_{n\to\infty} \delta(A_n) > 0$. Soit $A'' \in A' \cap \mathcal{A}$. Alors $A'' \in A_n \cap \mathcal{A}$ et donc $\delta(A'') \geqslant \alpha_n$ pour tout n. Comme $\lim_{n\to\infty} \alpha_n = 0$, on trouve $\nu(A'') - \mu(A'') = \delta(A'') \geqslant 0$.

Passons à la preuve de l'unicité. Soient f et g deux fonctions boréliennes positives telles que $\mathrm{P}(A) = \int_A f(x)\lambda(dx) = \int_A g(x)\lambda(dx)$, pour tout $A \in \mathcal{B}(]0,1])$. Alors $\int_A (f-g)(x)\lambda(dx) = 0$, pour tout $A \in \mathcal{B}(]0,1])$. Soit $B = \{f \leqslant g\} \in \mathcal{B}(]0,1])$. Alors $\int_B (f-g)(x)\lambda(dx) = 0$ et $f - g \geqslant 0$ sur B. On en déduit que $f - g = 0$ p.p. sur B. De même, en considérant $B^c \in \mathcal{B}(]0,1])$, on déduit que $f - g = 0$ p.p. sur B^c. Alors $f - g = 0$ p.p. sur $]0,1]$. □

Définition 1.10 *Soient* $(\Omega, \mathcal{A}, \mathrm{P})$ *un espace de probabilité. On dit qu'un ensemble* N *est* **négligeable** *s'il existe un événement* $B \in \mathcal{A}$ *tel que* $N \subset B$ *et* $\mathrm{P}(B) = 0$.

On dit qu'une propriété $\mathrm{Prop}(\omega)$ *est vérifiée* P-**presque surement** *si l'ensemble* $\{\omega : \mathrm{Prop}(\omega)$ *: est fausse*$\}$ *est négligeable.*

Une tribu sur Ω *est dite* **complète** *si elle contient tous les ensembles négligeables.*

Remarque : On peut toujours supposer, sans perte de généralité, que l'espace de probabilité est complet. □

Exemples : i) Soit $X : \Omega = \{1,2,3\} \to \{0,1\}$ mesurable définie par $X(1) = X(2) = 1$, $X(3) = 0$ et soit P la probabilité définie par $\mathrm{P}(\{1\}) = \mathrm{P}(\{2\}) = 1/2$, $\mathrm{P}(\{3\}) = 0$. Alors X est constante et égale à 1 p.s.
ii) Soit l'intervalle $\Omega = [0,1]$ muni de la tribu borélienne et de la mesure de Lebesgue λ. Soit

$$X(\omega) := \begin{cases} 1, & \text{si } \omega \in \mathbb{Q} \cap [0,1] \\ 0, & \text{sinon .} \end{cases}$$

Alors $X = 0$, p.s. (car $\lambda(\mathbb{Q} \cap [0,1]) = 0$).
De même, la variable aléatoire

$$Y(\omega) := \mathrm{sign}(\omega - \frac{1}{2}) = \begin{cases} 1, & \text{si } \omega > 1/2 \\ 0, & \text{si } \omega = 1/2 \\ -1, & \text{si } \omega < 1/2. \end{cases}$$

Alors Y est continue p.s. (car son seul point de discontinuité est $\frac{1}{2}$ et $\lambda(\{1/2\}) = 0$.

1.5 La construction d'une probabilité sur $(0,1]$*

i) Soit

$$\Omega = (0,1],\ \mathcal{B} = \mathcal{B}_{(0,1]} \ \text{ et } \ \mathcal{S} = \{(a,b] : 0 \leqslant a \leqslant b \leqslant 1\}.$$

Il est clair que $\emptyset,\ \Omega \in \mathcal{S}$.
Si $I_1, I_2 \in \mathcal{S}$, alors $I_1 \cap I_2 \in \mathcal{S}$. Donc $\mathcal{S}$ est stable par intersection finie.
Enfin, si $I \in \mathcal{S}$, alors I^c est une union disjointe de deux intervalles de $\mathcal{S}$ (faire un dessin).
De plus on sait que $\sigma(\mathcal{S}) = \mathcal{B}$ la tribu borélienne sur Ω.

ii) On définit sur $\mathcal{S}$ la fonction $\lambda : \mathcal{S} \to [0,1]$, par :

$$\lambda(\emptyset) = 0 \ \text{ et } \ \lambda((a,b]) = b - a.$$

On a $\lambda(\Omega) = \lambda((0,1]) = 1$.
Montrons que λ est une fonction additive sur $\mathcal{S}$. Soit $(a,b] \in \mathcal{S}$ et supposons que :

$$(a,b] = \bigcup_{i=1}^{r} (a_i, b_i],$$

où les intervalles dans le membre droit sont disjoints. Supposons aussi que ces intervalles ont été numérotées convenablement :

$$a_1 = a,\ b_r = b,\ b_i = a_{i+1},\ i = 1, \ldots, r-1.$$

Alors $\lambda((a,b]) = b - a$ et

$$\sum_{i=1}^{r} \lambda((a_i,b_i]) = \sum_{i=1}^{r} (b_i - a_i) = b_1 - a_1 + b_2 - a_2 + \ldots + b_r - a_r = b_r - a_1 = b - a.$$

iii) Montrons maintenant que λ est σ-additive. Soit $(a,b] \in \mathcal{S}$ et supposons que :

$$(a,b] = \bigcup_{i=1}^{\infty} (a_i, b_i],$$

où les intervalles dans le membre droit sont à nouveau disjoints.

iii)-a) Montrons d'abord que

$$\lambda((a,b]) = b - a \leqslant \sum_{i=1}^{\infty} (b_i - a_i) = \sum_{i=1}^{\infty} \lambda((a_i,b_i])\,. \tag{1.1}$$

On choisit $\varepsilon < b - a$ et on observe que

$$[a+\varepsilon, b] \subset \bigcup_{i=1}^{\infty} \left(a_i, b_i + \frac{\varepsilon}{2^i} \right)\ .$$

Le membre droit est un recouvrement ouvert du compact $[a+\varepsilon, b]$, donc on peut en extraire un recouvrement fini : il existe un entier N tel que

$$[a+\varepsilon, b] \subset \bigcup_{i=1}^{N} \left(a_i, b_i + \frac{\varepsilon}{2^i} \right)\ . \tag{1.2}$$

Pour terminer la preuve de (1.1) il suffit de prouver que

$$b - a - \varepsilon \leqslant \sum_{i=1}^{N} (b_i - a_i + \frac{\varepsilon}{2^i})\,. \tag{1.3}$$

En effet on aurait alors

$$b - a - \varepsilon \leqslant \sum_{i=1}^{N} (b_i - a_i + \frac{\varepsilon}{2^i}) \leqslant \sum_{i=1}^{\infty} (b_i - a_i) + \varepsilon$$

d'où

$$b - a \leqslant \sum_{i=1}^{\infty} (b_i - a_i) + 2\varepsilon$$

et comme ε est choisit arbitraire on obtient (1.1). Il reste à prouver que (1.2) implique (1.3).

iii)-b) Avec un petit changement de notations, on va montrer que

$$[c, d] \subset \bigcup_{i=1}^{N} (c_i, d_i) \,. \tag{1.4}$$

implique

$$d - c \leqslant \sum_{i=1}^{N} (d_i - c_i) \,. \tag{1.5}$$

On fait une récurrence : c'est clair pour $N = 1$. Supposons que (1.4) pour $N - 1$ implique (1.5) pour $N - 1$ et on vérifie l'implication pour N. Supposons que

$$c_N = \max_{i=1,\ldots,N} c_i \text{ et } c_N < d \leqslant d_N .$$

On considère deux cas (faire un dessin) :

1. Supposons $c_N \leqslant c$. Alors $d - c \leqslant d_N - c_N \leqslant \sum_{i=1}^{N} (d_i - c_i)$.
2. Supposons $c < c_N$. Alors par (1.4) on a

$$[c, c_N] \subset \bigcup_{i=1}^{N-1} (c_i, d_i),$$

et par l'hypothèse de récurrence

$$c_N - c \leqslant \sum_{i=1}^{N-1} (d_i - c_i),$$

donc

$$d - c = d - c_N + c_N - c \leqslant d - c_N + \sum_{i=1}^{N-1} (d_i - c_i) \leqslant d_N - c_N + \sum_{i=1}^{N-1} (d_i - c_i) = \sum_{i=1}^{N} (d_i - c_i).$$

Cela achève la preuve de (1.1).

iii)-c) Pour terminer la preuve de la σ-additivité on va vérifier l'inégalité inverse. On reprend $(a, b] = \bigcup_{i=1}^{\infty} (a_i, b_i]$ avec une union d'intervalles disjoints. Comme pour tout n, $\bigcup_{i=1}^{n} (a_i, b_i]$ est une union d'intervalles disjoints la même chose est vrai pour

$$(a, b] \setminus \bigcup_{i=1}^{n} (a_i, b_i] =: \bigcup_{j=1}^{m} I_j .$$

Comme λ est additive on peut écrire

$$\lambda((a,b]) = \lambda(\bigcup_{i=1}^{n}(a_i,b_i] \cup \bigcup_{j=1}^{m} I_j) = \sum_{i=1}^{n}\lambda((a_i,b_i]) + \sum_{j=1}^{m}\lambda(I_j) \geqslant \sum_{i=1}^{n}\lambda((a_i,b_i])$$

Lorsque $n \to \infty$ on trouve

$$\lambda((a,b]) \geqslant \sum_{i=1}^{\infty}\lambda((a_i,b_i])$$

donc λ est σ-additive sur $\mathcal{S}$.

iv) Une définition et le resultat clé en cadre général :

Définition 1.11 *(semi-algèbre)*
Soit Ω un ensemble non-vide quelconque. Une famille de sous-ensembles $\mathcal{S}$ de Ω est une **semi-algèbre** *si*

- $\emptyset, \Omega \in \mathcal{S}$ *;*
- $\mathcal{S}$ *est stable par intersections finies ;*
- *si $A \in \mathcal{S}$, alors il existe un n fini et des ensembles $C_1, \ldots, C_n$ disjoints dans $\mathcal{S}$ tels que $A^c = C_1 \cup \ldots \cup C_n$.*

Théorème 1.3 *(d'extension)*
Supposons que $\mathcal{S}$ est une semi-algèbre sur un ensemble non-vide quelconque Ω et que P est une fonction σ-additive d'ensembles définie sur $\mathcal{S}$, à valeurs dans $[0,1]$, telle que $\mathrm{P}(\Omega) = 1$. Alors il existe une unique mesure de probabilité $\mathbb{P}$ sur $\sigma(\mathcal{S})$ qui prolonge P (c'est-à-dire que $\mathbb{P}_{|\mathcal{S}} = \mathrm{P}$).

Ce résultat s'applique pour $\Omega = (0,1]$ avec $\mathcal{S}$ définie en *i)* et λ définie en *ii)*. La probabilité ainsi obtenue est la **mesure de Lebesgue sur** $(0,1]$.
v) Donnons les idées de preuve du résultat clé. On considère Ω un ensemble non-vide quelconque. Le Théorème 1.3 est une conséquence des trois lemmes suivants. Avant de les énoncer on a besoin d'une définition (rappel) :

Définition 1.12 *(algèbre et algèbre engéndrée)*
1. Une famille de sous-ensembles $\mathcal{A}$ de Ω est une **algèbre** *si*

- $\Omega \in \mathcal{A}$ *;*
- $A \in \mathcal{A}$ *implique* $A^c \in \mathcal{A}$ *;*
- $A, B \in \mathcal{A}$ *implique* $A \cup B \in \mathcal{A}$.

2. **L'algèbre engendrée** *par une famille de sous-ensembles $\mathcal{S}$ de Ω est la plus petite algèbre contenant $\mathcal{S}$ (on note $\mathcal{A}(\mathcal{S})$).*

Lemme 1.1 *(algèbre engéndrée par une semi-algèbre)*
Supposons que $\mathcal{S}$ est une semi-algèbre sur Ω. Alors, l'algèbre engendrée par $\mathcal{S}$ est la famille des unions finies d'éléments disjoints de $\mathcal{S}$:

$$\mathcal{A}(\mathcal{S}) := \{\bigcup_{i\in I} S_i : I \textit{ fini}, \ S_i \in \mathcal{S} \textit{ disjoints } \}. \tag{1.6}$$

Lemme 1.2 *(première extension)*
Supposons que $\mathcal{S}$ est une semi-algèbre sur Ω et que P est une fonction σ-additive définie sur $\mathcal{S}$, à valeurs dans $[0,1]$, telle que $\mathrm{P}(\Omega)=1$. Alors il existe une unique extension P' de P sur $\mathcal{A}(\mathcal{S})$, telle que P' est σ-additive sur $\mathcal{A}(\mathcal{S})$ et $\mathrm{P}'(\Omega)=1$. Cette extension est définie par

$$\mathrm{P}'(\bigcup_{i\in I} S_i) = \sum_{i\in I} \mathrm{P}'(S_i). \tag{1.7}$$

Lemme 1.3 *(deuxième extension)*
Supposons que $\mathcal{A}$ est une algèbre sur Ω et que P' est une fonction σ-additive définie sur $\mathcal{A}$, à valeurs dans $[0,1]$, telle que $\mathrm{P}'(\Omega)=1$. Alors il existe une unique mesure de probabilité $\mathbb{P}$ sur $\sigma(\mathcal{A})$ (tribu engendrée par $\mathcal{A}$) qui prolonge P'.

Preuve du Lemme 1.1. On note Λ l'ensemble du membre droit de (1.6). Il est clair que $\Lambda \supset \mathcal{S}$ et on montre que Λ est une algèbre. On vérifie les trois axiomes de la Définition 1.12.

1. Définition 1.11 implique $\Omega \in \mathcal{S}$, donc $\Omega \in \Lambda$.
2. Si $\bigcup_{i\in I} S_i$ et $\bigcup_{j\in J} S'_j$ sont deux éléments de Λ, alors
$$\left(\bigcup_{i\in I} S_i\right) \bigcap \left(\bigcup_{j\in J} S'_j\right) = \bigcup_{(i,j)\in I\times J} S_i \cap S'_j \in \Lambda$$
puisque $\{S_i \cap S'_j : (i,j) \in I\times J\}$ est une famille finie disjointe d'éléments de $\mathcal{S}$ (qui est stable par intersection finie).
3. Enfin, vérifions la stabilité au passage au complémentaire. Soit $\bigcup_{i\in I} S_i \in \Lambda$ dont le complémentaire est $\bigcap_{i\in I} S_i^c$. D'après la troisième axiome d'une semi-algèbre, comme $S_i \in \mathcal{S}$, on a
$$S_i^c = \bigcup_{j\in J_i} S_{ij},$$
avec une famille finie disjointe $\{S_{ij} : j \in J_i\}$ d'éléments de $\mathcal{S}$. Par le point précédent on conclut que $\bigcap_{i\in I} S_i^c \in \Lambda$.

Cela montre que Λ est une algèbre contenant $\mathcal{S}$, donc $\Lambda \supset \mathcal{A}(\mathcal{S})$. Par ailleurs, $\bigcup_{i\in I} S_i \in \Lambda$ implique $\bigcup_{i\in I} S_i \in \mathcal{A}(\mathcal{S})$, d'où $\Lambda \subset \mathcal{A}(\mathcal{S})$. $\square$

Preuve du Lemme 1.2. On commence par vérifier que P' est bien définie par (1.7), ensuite que P' est σ-additive sur $\mathcal{A}(\mathcal{S})$ et enfin que l'extension est unique.

1. Supposons que $A \in \mathcal{A}(\mathcal{S})$ admet deux représentations différentes $A = \bigcup_{i\in I} S_i = \bigcup_{j\in J} S'_j$ et on a besoin de vérifier que $\sum_{i\in I} \mathrm{P}(S_i) = \sum_{j\in J} \mathrm{P}(S'_j)$ pour que P' ait une unique valeur en A. Comme $S_i \in \mathcal{A}$,
$$\sum_{i\in I} \mathrm{P}(S_i) = \sum_{i\in I} \mathrm{P}(S_i \cap A) = \sum_{i\in I} \mathrm{P}(S_i \cap \bigcup_{j\in J} S'_j) = \sum_{i\in I} \mathrm{P}(\bigcup_{j\in J} S_i \cap S'_j)$$
et comme $S_i = \bigcup_{j\in J} S_i \cap S'_j \in \mathcal{S}$ on peut utiliser l'additivité de P. Ainsi
$$\sum_{i\in I} \mathrm{P}(S_i) = \sum_{i\in I}\sum_{j\in J} \mathrm{P}(S_i \cap S'_j) = \sum_{j\in J}\sum_{i\in I} \mathrm{P}(S_i \cap S'_j) = \sum_{j\in J} \mathrm{P}(S'_j).$$
La dernière égalité s'obtient en faisant le chemin inverse.

2. Supposons que pour $i \geqslant 1$ les ensembles disjoints sont $A_i = \bigcup_{j\in J_i} S_{ij} \in \mathcal{A}(\mathcal{S})$, $S_{ij} \in \mathcal{S}$ et $A = \bigcup_{i=1}^{\infty} A_i \in \mathcal{A}(\mathcal{S})$. Par ailleurs, comme $A \in \mathcal{A}(\mathcal{S})$, A admet la représentation $A = \bigcup_{k\in K} S_k$, $S_k \in \mathcal{S}$, $k \in K$ fini. Alors
$$\mathcal{S} \ni S_k = S_k \cap A = \bigcup_{i=1}^{\infty} S_k \cap A_i = \bigcup_{i=1}^{\infty} \bigcup_{j\in J_i} S_k \cap S_{ij} \text{ avec } S_{ij} \in \mathcal{S}.$$
Par (1.7)
$$\mathrm{P}'(A) = \sum_{k\in K} \mathrm{P}(S_k) = \sum_{k\in K} \sum_{i=1}^{\infty} \sum_{j\in J_i} \mathrm{P}(S_k \cap S_{ij}) = \sum_{i=1}^{\infty} \sum_{j\in J_i} \sum_{k\in K} \mathrm{P}(S_k \cap S_{ij})$$
$$= \sum_{i=1}^{\infty} \sum_{j\in J_i} \mathrm{P}(S_{ij}) = \sum_{i=1}^{\infty} \mathrm{P}(\bigcup_{j\in J_i} S_{ij}) = \sum_{i=1}^{\infty} \mathrm{P}'(A_i),$$
puisque $\sum_{k\in K} S_k \cap S_{ij} = A \cap S_{ij} = S_{ij} \in \mathcal{S}$.
3. Soient P'_1 et P'_2 deux extensions additives. Alors, pour tout $A = \bigcup_{i\in I} S_i \in \mathcal{A}(\mathcal{S})$ on a $\mathrm{P}'_1(A) = \sum_{i\in I} \mathrm{P}(S_i) = \mathrm{P}'_2(A)$.

On ainsi construit l'extension de P à une algèbre. □

Preuve du Lemme 1.3. On divise la preuve en trois parties : en 1ère partie on étend P' à une fonction d'ensemble Π σ-additive sur une famille $\mathcal{G} \supset \mathcal{A}$. En 2ème partie on étend Π à une fonction d'ensemble Π^* sur $\mathcal{P}(\Omega) \supset \sigma(\mathcal{A})$ et en 3ème partie on fait la restriction de Π^* à $\sigma(A)$ et on obtient la probabilité recherchée.

1. On définit d'abord la famille $\mathcal{G}$:
$$\mathcal{G} = \{\bigcup_{j=1}^{\infty} A_j : A_j \in \mathcal{A}\} = \{\lim_n \uparrow B_n : B_n \in \mathcal{A}, B_n \subset B_{n+1}, \forall n\}$$
et ensuite la fonction $\Pi : \mathcal{G} \to [0,1]$, par : si $G = \lim_n \uparrow B_n \in \mathcal{G}$, où $B_n \in \mathcal{A}$, alors
$$\Pi(G) = \lim_{n\to\infty} \mathrm{P}'(B_n). \tag{1.8}$$
Cette dernière définition est bien justifiée car P' est σ-additive donc la propriété de continuité sur des suites croissantes est vraie. On dit que $\{B_n\}$ est une **suite approchante** de G. Il reste à voir que Π est bien définie, c'est-à-dire

Fait 1 *Si G admet deux suites approchantes $\{B_n\}$ et $\{B'_n\}$,*
$$G = \lim_n \uparrow B_n = \lim_n \uparrow B'_n \text{ alors } \lim_{n\to\infty} \mathrm{P}'(B_n) = \lim_{n\to\infty} \mathrm{P}'(B'_n). \tag{1.9}$$
Listons quelques propriétés de Π et $\mathcal{G}$:

Fait 2 *On a*
$$\Omega, \emptyset \in \mathcal{G} \text{ et } \Pi(\omega) = 1, \Pi(\emptyset) = 0,$$
et pour $G \in \mathcal{G}$
$$0 \leqslant \Pi(G) \leqslant 1. \tag{1.10}$$
Enfin, on a $\mathcal{A} \subset \mathcal{G}$ et $\Pi_{|\mathcal{A}} = \mathrm{P}'$.

Fait 3 *Si $G_i \in \mathcal{G}$ pour $i = 1, 2$, alors $G_1 \cup G_2 \in \mathcal{G}$, $G_1 \cap G_2 \in \mathcal{G}$ et Π est additive :*

$$\Pi(G_1 \cup G_2) + \Pi(G_1 \cap G_2) = \Pi(G_1) + \Pi(G_2). \tag{1.11}$$

Fait 4 *Π est monotone sur $\mathcal{G}$: si $G_i \in \mathcal{G}$ pour $i = 1, 2$ et $G_1 \subset G_2$, alors $\Pi(G_1) \leqslant \Pi(G_2)$.*

Fait 5 *$\mathcal{G}$ est stable par des limites des suites croissantes et Π est continue sur des suites croissantes : si $G_n \in \mathcal{G}$ et $G_n \uparrow G$, alors $G \in \mathcal{G}$ et $\Pi(G) = \lim_{n\to\infty} \Pi(G_n)$.*

Par les Faits 3 et 5 on déduit que Π est σ-additive sur $\mathcal{G}$ donc la 1ère partie est vérifiée.

2. On définit $\Pi^* : \mathcal{P}(\Omega) \to [0, 1]$ par

$$\forall A \in \mathcal{P}(\Omega) : \Pi^*(A) = \inf\{\Pi(G) : A \subset G \in \mathcal{G}\}. \tag{1.12}$$

$\Pi^*(A)$ est le plus petit majorant des valeurs $\Pi(G)$ sur des ensembles $G \in \mathcal{G}$ contenant A. C'est **la mesure extérieure de** A. Comme pour Π, on va lister les propriétés de Π^* :

Fait 6 *On a*

$$\Pi^*_{|\mathcal{G}} = \Pi \tag{1.13}$$

et $0 \leqslant \Pi^*(A) \leqslant 1$, pour tout $A \in \mathcal{P}(\Omega)$. En particulier, $\Pi^*(\Omega) = \Pi(\Omega) = 1$ et $\Pi^*(\emptyset) = \Pi(\emptyset) = 0$.

Fait 7 *Π^* est sous-additive : on a pour $A_1, A_2 \in \mathcal{P}(\Omega)$*

$$\Pi^*(A_1 \cup A_2) + \Pi^*(A_1 \cap A_2) \leqslant \Pi^*(A_1) + \Pi^*(A_2). \tag{1.14}$$

En particulier

$$1 = \Pi^*(\Omega) \leqslant \Pi^*(A) + \Pi^*(A^c). \tag{1.15}$$

Fait 8 *Π^* est monotone sur $\mathcal{P}(\Omega)$.*

Fait 9 *Π^* est continue sur des suites croissantes : si $A_n \uparrow A$, alors $\Pi^*(A_n) \uparrow \Pi^*(A)$.*

3. On introduit une sous-famille $\mathcal{D}$ de $\mathcal{P}(\Omega)$:

$$\mathcal{D} := \{D \in \mathcal{P}(\Omega) : \Pi^*(D) + \Pi^*(D^c) = 1\}. \tag{1.16}$$

Fait 10 *$\mathcal{D}$ est une tribu et $\Pi^*_{|\mathcal{D}}$ est une probabilité sur $(\Omega, \mathcal{D})$.*

Fait 11 *$\mathcal{D} \supset \mathcal{A}$, donc $\mathcal{D} \supset \sigma(\mathcal{A})$, et alors $\Pi^*_{|\sigma(\mathcal{A})}$ est la probabilité (unique par un argument de classe monotone) désirée.*

Preuve du Fait 1 : Il suffit de montrer que :

$$\bigcup_{n=1}^{\infty} B_n \subset \bigcup_{n=1}^{\infty} B'_n \text{ implique } \lim_{n\to\infty} \mathrm{P}'(B_n) \leqslant \lim_{n\to\infty} \mathrm{P}'(B'_n). \tag{1.17}$$

Pour m fixé $\lim_n \uparrow (B_m \cap B'_n) = B_m$ et on a aussi $B_m \cap B'_n \subset B'_n$. On sait que la σ-additivité de P implique la continuité sur des suites croissantes. On en déduit :

$$\lim_{n\to\infty} \mathrm{P}'(B'_n) \geqslant \lim_{n\to\infty} \mathrm{P}'(B_m \cap B'_n) = \mathrm{P}'(B_m).$$

Comme cette inégalité a lieu pour tout m, on en déduit que

$$\lim_{n\to\infty} \mathrm{P}'(B_n') \geqslant \lim_{m\to\infty} \mathrm{P}'(B_m).$$

□

Preuve du Fait 2 : Si on pose $B_n = \Omega$ pour tout n, alors

$$\mathcal{A} \ni B_n = \Omega \uparrow \Omega \text{ et } \Pi(\Omega) = \lim_{n\to\infty} \mathrm{P}'(\Omega) = 1.$$

Le même argument marche pour $\emptyset$. (1.10) s'obtient par le fait que $0 \leqslant \mathrm{P}'(B_n) \leqslant 1$, pour toute suite approchante $\{B_n\}$ de $\mathcal{A}$. Enfin, pour montrer que $\Pi(A) = \mathrm{P}'(A)$ pour tout $A \in \mathcal{A}$, on prend la suite approchante identiquement égale à A. □

Preuve du Fait 3 : Soient les suites approchantes $B_{n1}, B_{n2} \in \mathcal{A}$, telles que $B_{ni} \uparrow G_i$ pour $i = 1, 2$. Comme $\mathcal{A}$ est une algèbre, on voit que

$$\mathcal{A} \ni B_{n1} \cup B_{n2} \uparrow G_1 \cup G_2,\ \mathcal{A} \ni B_{n1} \cap B_{n2} \uparrow G_1 \cap G_2,$$

qui montrent que $G_1 \cup G_2, G_1 \cap G_2 \in \mathcal{G}$. De plus

$$\mathrm{P}'(B_{n1} \cup B_{n2}) + \mathrm{P}'(B_{n1} \cap B_{n2}) = \mathrm{P}'(B_{n1}) + \mathrm{P}'(B_{n2}),$$

et pour $n \to \infty$ on trouve (1.11). □

Preuve du Fait 4 : C'est une conséquence directe de (1.17). □

Preuve du Fait 5 : Pour chaque n, G_n admet une suite approchante $B_{m,n} \in \mathcal{A}$ telle que $\lim_m \uparrow B_{m,n} = G_n$. On définit $D_m = \cup_{n=1}^m B_{m,n} \in \mathcal{A}$ (car $\mathcal{A}$ est stable par union finie). On va montrer que

$$\lim_m \uparrow D_m = G \tag{1.18}$$

et alors G admet une suite approchante croissante d'éléments de $\mathcal{A}$, donc $G \in \mathcal{G}$.
Montrons d'abord que $\{D_m\}$ est croissante :

$$D_m = \bigcup_{n=1}^{m} B_{m,n} \subset \bigcup_{n=1}^{m} B_{m+1,n} \subset \bigcup_{n=1}^{m+1} B_{m+1,n} = D_{m+1}$$

Calculons la limite de $\{D_m\}$. Si $n \geqslant m$, on a par la définition de D_m :

$$B_{m,n} \subset D_m = \bigcup_{j=1}^{m} B_{m,j} \subset \bigcup_{j=1}^{m} G_j = G_m.$$

donc $B_{m,n} \subset D_m \subset G_m$. On prend la limite en m :

$$G_n = \lim_m \uparrow B_{m,n} \subset \lim_m \uparrow D_m \subset \lim_m \uparrow G_m = G$$

et ensuite la limite en n :

$$G = \lim_n \uparrow G_n \subset \lim_m \uparrow D_m \subset \lim_m \uparrow G_m = G.$$

Donc $G \in \mathcal{G}$ et par la définition de Π, on sait que $\Pi(G) = \lim_{m\to\infty} \Pi(D_m)$. Il reste à prouver que $\Pi(G_n) \uparrow \Pi(G)$. Par les trois inclusions précédentes :

$$\Pi(B_{m,n}) \leqslant \Pi(D_m) \leqslant \Pi(G_m).$$

On fait $m \to \infty$ et comme $G_n = \lim_m \uparrow B_{m,n}$,

$$\Pi(G_n) \leqslant \lim_{m\to\infty} \Pi(D_m) \leqslant \lim_{m\to\infty} \Pi(G_m),\ \forall n.$$

On fait $n \to \infty$:

$$\lim_{n\to\infty} \Pi(G_n) \leqslant \lim_{m\to\infty} \Pi(D_m) \leqslant \lim_{m\to\infty} \Pi(G_m)$$

donc $\lim_{n\to\infty} \Pi(G_n) = \lim_{m\to\infty} \Pi(D_m) = \Pi(G)$. □

Preuve du Fait 6 : Il est clair que si $A \in \mathcal{G}$, alors $A \in \{G : A \subset G \in \mathcal{G}\}$ et donc l'infimum est atteint en A. □

Preuve du Fait 7 : Pour vérifier (1.14), on fixe $\varepsilon > 0$ et on trouve $G_i \in \mathcal{G}$ tels que $G_i \supset A_i$ et pour $i = 1, 2$,

$$\Pi^*(A_i) + \frac{\varepsilon}{2} \geqslant \Pi(G_i).$$

On somme ces deux inégalités et on trouve

$$\Pi^*(A_1) + \Pi^*(A_2) + \varepsilon \geqslant \Pi(G_1) + \Pi(G_2) = \Pi(G_1 \cup G_2) + \Pi(G_1 \cap G_2).$$

par le Fait 3 pour Π. Comme $G_1 \cup G_2 \supset A_1 \cup A_2$ et $G_1 \cap G_2 \supset A_1 \cap A_2$, par la définition de Π^* qu'on peut encore minorer par

$$\Pi^*(A_1 \cup A_2) + \Pi^*(A_1 \cap A_2).$$

□

Preuve du Fait 8 : Cette propriété est une conséquence du fait que Π est monotone sur $\mathcal{G}$ (Fait 4). □

Preuve du Fait 9 : On fixe $\varepsilon > 0$. Pour chaque $n \geqslant 1$ on trouve $G_n \in \mathcal{G}$ tels que $G_n \supset A_n$ et

$$\Pi^*(A_n) + \frac{\varepsilon}{2^n} \geqslant \Pi(G_n).$$

On pose $G'_n = \cup_{m=1}^n G_m$. Comme $\mathcal{G}$ est stable par réunion finie, $G'_n \in \mathcal{G}$ et $\{G'_n\}$ est croissante. On montre par récurrence

$$\Pi^*(A_n) + \varepsilon \sum_{m=1}^n 2^{-m} \geqslant \Pi(G'_n). \tag{1.19}$$

Pour $n = 1$ c'est le choix de G_n. On montre que "$n \Rightarrow n+1$". On a

$$A_n \subset G_n \subset G'_n \text{ et } A_{n+1} \subset G_{n+1} \subset G'_{n+1}$$

et donc $A_n \subset G'_n$ et $A_{n+1} \subset G'_{n+1}$, donc $A_n \subset G'_n \cap G_{n+1} \in \mathcal{G}$. Ainsi

$$\Pi(G'_{n+1}) = \Pi(G'_n \cup G_{n+1}) = \Pi(G'_n) + \Pi(G_{n+1}) - \Pi(G'_n \cap G_{n+1})$$

par (1.11). On peut alors majorer le membre de droite de l'égalité précédente par

$$\leqslant \left(\Pi^*(A_n) + \varepsilon \sum_{m=1}^{n} 2^{-m}\right) + \Pi^*(A_{n+1}) + \frac{\varepsilon}{2^{n+1}} - \Pi^*(A_n) = \varepsilon \sum_{m=1}^{n+1} 2^{-m} + \Pi^*(A_{n+1})$$

qui est (1.19) pour $n+1$. On fait $n \to \infty$ dans (1.19). D'après la monotonie de Π sur $\mathcal{G}$ et celle de Π^* sur $\mathcal{P}(\Omega)$, et comme $\mathcal{G}$ est stable par des unions croissantes, on obtient

$$\lim_{n\to\infty} \Pi^*(A_n) + \varepsilon \geqslant \lim_{n\to\infty} \Pi(G'_n) = \Pi(\bigcup_{j=1}^{\infty} G'_j).$$

Comme $A = \lim_n \uparrow A_n \subset \bigcup_{j=1}^{\infty} G'_j \in \mathcal{G}$, on déduit que $\lim_{n\to\infty} \Pi^*(A_n) \geqslant \Pi^*(A)$. Par ailleurs, la monotonie donne $\Pi^*(A_n) \leqslant \Pi^*(A)$, d'où $\lim_{n\to\infty} \Pi^*(A_n) \leqslant \Pi^*(A)$. □

Preuve du Fait 10 : D'abord on prouve que $\mathcal{D}$ est une algèbre. Il est clair que $\Omega \in \mathcal{D}$, puisque $\Pi^*(\Omega) = 1$ et $\Pi^*(\emptyset) = 0$. Le passage au complémentaire est évident donc il reste à vérifier la stabilité aux unions et intersections finies. Si $A_1, A_2 \in \mathcal{D}$, alors par (1.14) on trouve :

$$\Pi^*(D_1 \cup D_2) + \Pi^*(D_1 \cap D_2) \leqslant \Pi^*(D_1) + \Pi^*(D_2) \tag{1.20}$$

$$\Pi^*((D_1 \cup D_2)^c) + \Pi^*((D_1 \cap D_2)^c) \leqslant \Pi^*(D_1^c) + \Pi^*(D_2^c). \tag{1.21}$$

On additionne (1.20) et (1.21) pour obtenir

$$\Pi^*(D_1 \cup D_2) + \Pi^*((D_1 \cup D_2)^c) + \Pi^*(D_1 \cap D_2) + \Pi^*((D_1 \cap D_2)^c) \leqslant 2 \tag{1.22}$$

où le membre de droite est obtenu parce que $D_1, D_2 \in \mathcal{D}$. Par (1.15) le membre de gauche est $\geqslant 2$ donc en (1.22) on a égalité. En combinant cette égalité avec (1.15) on trouve

$$\Pi^*(D_1 \cup D_2) + \Pi^*((D_1 \cup D_2)^c) = 1$$

$$\Pi^*(D_1 \cap D_2) + \Pi^*((D_1 \cap D_2)^c) = 1,$$

donc $D_1 \cup D_2, D_1 \cap D_2 \in \mathcal{D}$ et $\mathcal{D}$ est une algèbre. De plus on obtient des égalités en (1.20) et (1.21) (sinon on contredit (1.22)), donc Π^* est additive sur $\mathcal{D}$.

Pour montrer que $\mathcal{D}$ est une tribu il suffit de vérifier que $\mathcal{D}$ est une classe monotone (et utiliser ensuite le théorème de classe monotone). Comme $\mathcal{D}$ est stable par passage au complémentaire il suffit de montrer que $D_n \in \mathcal{D}$, $D_n \uparrow D$, implique $D \in \mathcal{D}$. Par le Fait 9

$$\lim_{n\to\infty} \Pi^*(D_n) = \Pi^*(\bigcup_{n=1}^{\infty} D_n) = \Pi^*(D).$$

Par ailleurs, pour tout $m \geqslant 1$,

$$\Pi^*((\bigcup_{n=1}^{\infty} D_n)^c) = \Pi^*(\bigcap_{n=1}^{\infty} D_n^c) \leqslant \Pi^*(D_m^c),$$

d'où, par (1.15),

$$1 \leqslant \Pi^*(\bigcup_{n=1}^{\infty} D_n) + \Pi^*((\bigcup_{n=1}^{\infty} D_n)^c) \leqslant \lim_{n\to\infty} \Pi^*(D_n) + \Pi^*(D_m^c). \tag{1.23}$$

Si on fait $m \to \infty$, comme $D_n \in \mathcal{D}$,

$$1 \leqslant\leqslant \lim_{n\to\infty} \Pi^*(D_n) + \lim_{m\to\infty} \Pi^*(D_m^c) = \lim_{n\to\infty} (\Pi^*(D_n) + \Pi^*(D_n^c)) = 1,$$

donc (1.23) est une égalité et donc $D = \bigcup_{n=1}^{\infty} D_n \in \mathcal{D}$. Ainsi $\mathcal{D}$ est une algèbre et une classe monotone, donc une tribu.

Montrons que $\Pi^*_{|\mathcal{D}}$ est σ-additive. Soit D_n une suite disjointe dans $\mathcal{D}$. Comme $\mathcal{D}$ est une algèbre et par le Fait 9 et l'additivité de Π^* on a

$$\Pi^*(\bigcup_{n=1}^{\infty} D_n) = \Pi^*(\lim_n \bigcup_{i=1}^{n} D_i) = \lim_{n\to\infty} \Pi^*(\bigcup_{i=1}^{n} D_i) = \lim_{n\to\infty} \sum_{i=1}^{n} \Pi^*(D_i) = \sum_{i=1}^{\infty} \Pi^*(D_i).$$

□

Preuve du Fait 11. Tout élément $A \in \mathcal{A}$ est élément de $\mathcal{G}$ (suite approchante constante) et alors $\Pi^*(A) = \Pi(A) = \mathrm{P}'(A)$ et la même chose pour A^c. Mais alors, par (1.15) $1 \leqslant \Pi^*(A) + \Pi^*(A^c) = \mathrm{P}'(A) + \mathrm{P}'(A^c) = 1$, d'où $A \in \mathcal{D}$. Ainsi, $\mathcal{D} \supset \mathcal{A}$, donc la tribu $\mathcal{D}$ contient $\sigma(\mathcal{A})$. La restriction $\Pi^*_{|\sigma\mathcal{A}}$ est la probabilité désirée. L'unicité de cette extension de $\mathcal{A}$ à $\sigma(\mathcal{A})$ s'obtient par un argument de classe monotone. □

1.6 Loi d'une variable aléatoire

Définition 1.13 *Soient* $(\Omega, \mathcal{A}, \mathrm{P})$ *un espace de probabilité et* X *une variable aléatoire* Ω *dans un espace mesurable* $(E, \mathcal{B})$. *L'application* $\mathrm{P}_X : \mathcal{B} \to [0,1]$ *définie par* $\mathrm{P}_X(B) := \mathrm{P}(X^{-1}(B))$, *définit une probabilité sur* $(E, \mathcal{B})$ *appelée* **loi** *de* X.

Soit une autre variable aléatoire $Y : (\Omega', \mathcal{A}', \mathrm{P}') \to (E, \mathcal{B})$. *On dit que* X *et* Y **ont la même loi** *lorsque* $\mathrm{P}_X = \mathrm{P}_Y \Leftrightarrow \mathrm{P}_X(B) = \mathrm{P}_Y(B)$, $\forall B \in \mathcal{B}$.

Remarque : On peut alléger les notations :

$$\mathrm{P}_X(B) = \mathrm{P}(\{\omega \in \Omega : X(\omega) \in B\}) = \mathrm{P}(X \in B).$$

En ce qui concerne la deuxième partie, remarquer que les deux variables aléatoires ne sont pas nécessairemant définies sur le même espace de probabilité. □

Exemple : Sur $\Omega = \{1, 2, 3, 4, 5, 6\}$ muni de la tribu de ses parties et de la probabilité $\mathrm{P}(A) = \mathrm{card}(A)/6$ on définit X à valeurs dans $\{0, 1\}$, par $X(\omega) = 1$, si ω est pair et $X(\omega) = 0$ si ω est impair. On vérifie que $\mathrm{P}_X(\{0\}) = \mathrm{P}_X(\{1\}) = 1/2$.

Remarque : Si l'on se donne une probabilité sur $(E, \mathcal{B})$ (une loi) on peut toujours l'écrire comme la loi d'une variable aléatoire (prendre l'identité pour la variable aléatoire). Pour les applications, en général, seule compte la loi et on explicite plus rarement la variable aléatoire et l'espace de probabilité $(\Omega, \mathcal{A}, \mathrm{P})$. Par exemple, on dit que X est de loi de Bernoulli si $\mathrm{P}(X = 1) = 1 - \mathrm{P}(X = 0) = p$ au lieu de dire que $X : (\Omega, \mathcal{A}, \mathrm{P}) \to \{0, 1\}$ avec $\mathrm{P}_X(\{1\}) = 1 - \mathrm{P}_X(\{0\}) = p$. Aussi la représentation d'une loi par une variable aléatoire n'est pas unique. En reprenant l'exemple de la loi de Bernoulli, on peut choisir

$$\Omega = \{0, 1\},\ \mathcal{A} = \mathcal{P}(\Omega),\ \mathrm{P} = p\delta_1 + (1 - p)\delta_0,\ X(\omega) = \omega$$

ou

$$\Omega' = [0,1],\ \mathcal{A}' = \mathcal{B}([0,1]),\ \mathrm{P}' = \lambda,\ Y(\omega) = \mathbb{1}_{[0,p]}(\omega).$$

□

Définition 1.14 *On dit qu'une loi* Q *sur* $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ *est* **discrète** *si c'est une combinaison linéaire finie ou dénombrable à coefficients positifs ou nuls de masses de Dirac*

$$\mathrm{Q} = \sum_{j\in J} p_j \delta_{x_j},\ x_j \in \mathbb{R}^d,\ j \in J \subset \mathbb{N}.$$

Une variable aléatoire X *est* **discrète** *si sa loi est* $\mathrm{P}_X = \sum_{j\in J} p_j\delta_{x_j}$. X *ne prend (presque sûrement) qu'un nombre fini ou dénombrable de valeurs.*

X *est discrète si et seulement si* $X = \sum_{j\in J} x_j \mathbb{1}_{A_j}$ *avec* $A_j = X^{-1}(\{x_j\}) = \{\omega \in \Omega : X(\omega) = x_j\}$ *disjoints deux à deux et dont la réunion est* Ω.

On dit qu'une loi Q *sur* $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ **admet une densité** *s'il existe une fonction* f *borélienne (presque sûrement) positive ou nulle, telle que*

$$\mathrm{Q}(B) = \int_{\mathbb{R}^d} f(x)\mathbb{1}_B(x)\,dx,\ B \in \mathcal{B}(\mathbb{R}^d).$$

On notera la densité de la loi d'une variable aléatoire X*, par* f_X *et on dira qu'elle est la* **densité** *de* X *:*

$$\mathrm{P}_X(B) = \mathrm{P}(X \in B) = \int_{\mathbb{R}^d} f_X(x)\mathbb{1}_B(x)\,dx,\ B \in \mathcal{B}(\mathbb{R}^d).$$

Remarque : (cas discret) On peut écrire que $p_j = \mathrm{P}(X = x_j) = p_X(x_j)$ pour tous $j \in J$. Les coefficients p_j sont positifs ou nuls et $\sum_{j\in J} p_j = 1$. □

Remarque : (cas à densité) Si f est la densité d'une loi alors elle est presque sûrement unique (c'est-à-dire, si g est aussi une densité de la même loi alors $f = g$ p.p.). De plus

$$\int_{\mathbb{R}^d} f(x)\,dx = 1.$$

□

Remarque : On peut montrer que si f est une fonction borélienne positive p.p. sur $\mathbb{R}^d$ et telle que son intégrale sur tout l'espace est égale à 1, alors il existe un espace de probabilité $(\Omega, \mathcal{A}, \mathrm{P})$ et une variable aléatoire X dont la loi admet f pour densité.

Une condition suffisante (et nécessaire) pour que la loi de X admette une densité est la condition d'absolue continuité de la probabilité P_X par rapport à la mesure de Lebesgue. □

Remarque : Soit X une variable aléatoire dont la loi est discrète donnée par $p_X(x_j) = \mathrm{P}(X = x_j)$, $j \in J$ et soit $g : J \to I$ une application mesurable de l'ensemble dénombrable J dans l'ensemble dénombrable I. Alors $Y = g(X)$ est une variable aléatoire dont la loi est discrète donnée par

$$p_Y(y_i) = \mathrm{P}(Y = y_i) = \sum_{j\in J : g(x_j) = y_i} p_X(j).$$

On verra que ce resultat ne se généralise pas tel quel à des fonctions de plusieurs variables dans le cas discret. □

Proposition 1.9 *Soit X un vecteur aléatoire à valeurs dans $\mathbb{R}^d$ de densité f_X et soit $g : \mathbb{R}^d \to \mathbb{R}^d$ un difféomorphisme de $\mathbb{R}^d$, c'est-à-dire, une bijection continûment différentiable ainsi que son inverse. Alors $Y = g(X)$ est une variable aléatoire à valeurs dans $\mathbb{R}^d$ qui admet pour densité la fonction*

$$f_Y(y) = f_X(g^{-1}(y))|\mathrm{Jac}\, g^{-1}(y)|,\ y \in \mathbb{R}^d,$$

où $\mathrm{Jac}\, g(x) = \det\left(\frac{\partial g_i}{\partial x_j} : 1 \leqslant i, j \leqslant d\right)$ *désigne le jacobien de g.*

Preuve : Soit $B \in \mathcal{B}(\mathbb{R}^d)$ et on calcule

$$\mathrm{P}(g(X) \in B) = \mathrm{P}(X \in g^{-1}(B)) = \int_{\mathbb{R}^d} \mathbb{1}_{\{x \in g^{-1}(B)\}} f_X(x)dx = \int_{\mathbb{R}^d} \mathbb{1}_B(g(x)) f_X(x)dx.$$

On pose $y = g(x) \Leftrightarrow x = g^{-1}(y)$ et on applique le théorème de changement de variable

$$\mathrm{P}(g(X) \in B) = \int_{\mathbb{R}^d} \mathbb{1}_B(y) f_X(g^{-1}(y))|\mathrm{Jac}\, g^{-1}(y)|\, dy.$$

□

Définition 1.15 *Soit X une variable aléatoire réelle définie sur un espace de probabilité $(\Omega, \mathcal{A}, \mathrm{P})$. On appelle* **fonction de répartition** *de X ou de sa loi P_X et on note F_X, la fonction sur $\mathbb{R}$ définie par*

$$F_X(t) = \mathrm{P}_X(]-\infty, t]) = \mathrm{P}\left(\{\omega \in \Omega : X(\omega) \leqslant t\}\right) = \mathrm{P}(X \leqslant t),\ t \in \mathbb{R}.$$

Proposition 1.10 *La fonction de répartition F_X de la variable aléatoire réelle X vérifie les propriétés suivantes :*

1) $0 \leqslant F_X \leqslant 1$ *;*

2) F_X *est croissante, continue à droite avec une limite à gauche en tout point ;*

3) $\lim_{t\to-\infty} F_X(t) = 0$ *et* $\lim_{t\to\infty} F_X(t) = 1$.

Réciproquement, une fonction F vérifiant 1)-3) est la fonction de répartition d'une variable aléatoire réelle.

Preuve : 1) vient du fait que P est à valeurs dans [0,1]. La croissance découle de la croissance de P.

La continuité à droite peut être vue comme une conséquence de la Proposition 8, 4) en remarquant que

$$\{X \leqslant t\} = \bigcap_{n \in \mathbb{N}^*} \{X \leqslant t + \frac{1}{n}\}$$

et que la croissance de F_X implique

$$\lim_{h\downarrow 0} F_X(t+h) = \lim_{n\uparrow\infty} F_X(t+\frac{1}{n}) = F_X(t).$$

La limite à gauche est également une conséquence de la croissance de F_X.

La propriété 3) vient encore de la Proposition 8, 4) en remarquant que $\emptyset = \cap_{n\in\mathbb{N}^*}\{X \leqslant -n\}$ et donc

$$0 = \mathrm{P}(\emptyset) = \lim_{n\uparrow\infty} \mathrm{P}(X \leqslant -n) = \lim_{n\uparrow\infty} F_X(-n),$$

tandis que

$$1 = \mathrm{P}(\Omega) = \lim_{n\uparrow\infty} \mathrm{P}(X \leqslant n),$$

d'après la Proposition 8, 3).

Soit maintenant F une fonction vérifiant 1)-3). On définit, pour $a < b$ la fonction d'ensemble $\mu(]a,b]) = F(b) - F(a)$. La définition de μ s'étend à l'ensemble $\mathcal{C}$ des réunions finies d'intervalles. On peut aussi montrer que pour toute suite croissante $(A_n)_{n\in\mathbb{N}}$ d'éléments de $\mathcal{C}$, de réunion A, on ait $\lim_{n\uparrow\infty}\mu(A_n) = \mu(A)$. Alors, par un théorème de prolongement (admis), μ se prolonge en une probabilité sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. □

Théorème 1.4 *La fonction de répartition caractérise la loi, c'est-à-dire,* $F_X = F_Y$ *si et seulement si* $\mathrm{P}_X = \mathrm{P}_Y$.

Preuve : En effet, si $F_X = F_Y$, alors P_X et P_Y coïncident sur les intervalles, donc sur la tribu engendrée par l'ensemble $\mathcal{E}$ des intervalles ; cette tribu est la tribu borélienne. De plus $\mathcal{E}$ est stable par intersection finie. Donc, par le théorème de classe monotone $\mathcal{M}(\mathcal{E}) = \sigma(\mathcal{E}) = \mathcal{B}(\mathbb{R})$. Soit

$$\mathcal{M}_1 = \{B \in \mathcal{B}(\mathbb{R}) : \mathrm{P}_X(B) = \mathrm{P}_Y(B)\}.$$

D'après la propriété de continuité monotone de P on prouve que $\mathcal{M}_1$ est une classe monotone contenant $\mathcal{E}$. Le résultat s'ensuit. □

Proposition 1.11 *La fonction de répartition admet au plus un nombre dénombrable de points de discontinuité.*

Preuve : Soit D_n l'ensemble des points de discontinuité avec un saut d'amplitude plus grande que $1/n$:

$$D_n := \{t \in \mathbb{R} : F(t) - F(t-) \geqslant \frac{1}{n}\}.$$

Comme $0 \leqslant F \leqslant 1$ on a nécessairement $\mathrm{card}(D_n) \leqslant n$. L'ensemble des points de discontinuité est $\cup_{n\in\mathbb{N}^*} D_n$, et donc est dénombrable. □

Remarque : Supposons que la loi de X est discrète et que X ne prend qu'un nombre fini de valeurs $X(\Omega) = \{x_1, \ldots, x_r\}$, avec $x_1 < \ldots < x_r$. Alors

$$F_X(t) = \begin{cases} 0, & \text{si } t < x_1 \\ p_1 + \ldots + p_j, & \text{si } x_j \leqslant t < x_{j+1} \text{ avec } j < n, \\ 1, & \text{si } t \geqslant x_r \end{cases} \qquad t \in \mathbb{R},$$

où $p_j = \mathrm{P}(X = x_j)$, $j = 1, \ldots, r$; on peut voir que $p_j = F_X(x_j) - F_X(x_j-)$.

REMARQUE : Supposons que la loi de X admet une densité f_X. Alors

$$F_X(t) = \int_{-\infty}^{t} f_X(x)\, dx, t \in \mathbb{R}.$$

On peut voir que F_X est continue : $F_X(t-) = F_X(t)$ et en particulier

$$\mathrm{P}(X = t) = 0, \forall t \in \mathbb{R}.$$

De plus F_X est dérivable p.p. sur $\mathbb{R}$. Ainsi il y a une bijection entre l'ensemble des fonctions boréliennes positives p.p. sur $\mathbb{R}$ et d'intégrale 1 et l'ensemble des fonctions de répartition continues sur $\mathbb{R}$ et dérivables p.p. sur $\mathbb{R}$. Si de plus f_X est continue, alors F_X est dérivable.□

1.7 Dérivabilité des fonctions monotones*

Rappel : si f est une fonction continue et si F est une fonction continûment différentiable on a

$$\frac{d}{dx}\int_a^x f(t)dt = f(x) \text{ et } \int_a^b F'(t)dt = F(b) - F(a).$$

Questions : La première égalité est-elle vraie pour des fonctions intégrables au sens de Lebesgue ? Quelle est la classe (aussi vaste que possible) des fonctions pour lesquelles la deuxième égalité a lieu ?

Le but est d'étudier les propriétés de l'intégrale de Lebesgue $F(x) := \int_a^x f(t)dt$, comme fonction de sa borne supérieure. Si $f \geqslant 0$ alors F est monotone non décroissante. On sait que toute fonction intégrable f est différence de deux fonctions intégrables non négatives $f = f^+ - f^-$ donc F se décompose en une différence de deux fonctions monotones non décroissantes. Ainsi l'étude de F peut être réduit à l'étude des fonctions monotones du même type. Les propriétés **5** et **7** ci-dessous, donnent la réponse à la première question. La réponse à la deuxième est contenue dans les propriétés **8, 11** et **12**.

Définition 1.16 *Une fonction g sur un intervalle $[a, b]$ est dite monotone non décroissante si $t \leqslant t'$ implique $g(t) \leqslant g(t')$. La limite $\lim_{h\downarrow 0} g(t_0 + h)$ (lorsqu'elle existe) s'appelle limite à droite de g au point t_0 et se note $g(t_0+)$. De la même façon on définit la limite à gauche de g en t_0 notée $g(t_0-)$. Le point où ces deux limites existent, mais sont inégales, s'appelle point de discontinuité de première espèce. La différence $g(t_0+) - g(t_0-)$ s'appelle saut de g en t_0. Si $g(t_0) = g(t_0+)$ on dit que la fonction est continue à droite en t_0.*

Propriété 1 *Une fonction monotone non décroissante sur un intervalle $[a, b]$ est borélienne et bornée, donc intégrable.*

En effet si g est monotone non décroissante sur $[a, b]$, on a $g(a) \leqslant g(t) \leqslant g(b)$ sur $[a, b]$. D'autre part $\{t : g(t) < c\}$ est soit un segment, soit un intervalle semi-ouvert (soit l'ensemble vide). En effet, supposons qu'il existe des points t tels que $g(t) < c$. On note α la borne supérieure de ces points. Alors $\{t : g(t) < c\}$ est soit $[a, \alpha]$, soit $[a, \alpha)$.

Propriété 2 *Une fonction monotone ne peut avoir que des discontinuités de première espèce.*

En effet, soit t_0 un point quelconque de $[a, b]$ et soit $t_n \uparrow t_0$. La suite $\{g(t_n)\}$ est bornée par $g(a)$ et $g(b)$, donc elle a au moins un point d'accumulation. L'existence de plusieurs points d'accumulation pour une telle suite est en contradiction avec la monotonie de g. Ainsi $g(t_0-)$ existe et on fait de la même façon pour $g(t_0+)$.

Propriété 3 *L'ensemble des points de discontinuité d'une fonction monotone est au plus dénombrable.*

En effet, la somme de tout nombre fini de sauts de g sur $[a, b]$ ne dépasse pas $g(b) - g(a)$. Par conséquent, pour chaque n, le nombre de sauts plus grand que $\frac{1}{n}$ est fini. En faisant leur somme pour tous les $n = 1, 2, \ldots$, on trouve que le nombre total de sauts est fini ou dénombrable.

Propriété 4 *Toute fonction monotone continue à droite peut être représentée de façon unique comme la somme d'une fonction continue monotone et d'une fonction de sauts (continue à droite).*

En effet, soient $t_1, t_2, \ldots$ les points de discontinuité de la fonction non décroissante g et soient $p_1, p_2, \ldots$ ses sauts en ces points. On pose $p(t) = \sum_{\{n:t_n \leqslant t\}} p_n$. Soit $\gamma(t) = g(t) - p(t)$. Cette fonction est non décroissante car, pour $t' < t''$

$$\gamma(t'') - \gamma(t') = [g(t'') - g(t')] - [p(t'') - p(t')] \geqslant 0$$

comme somme de l'accroissement de g sur $[t', t'']$ et la somme de ses sauts sur le même intervalle. Pour t^* un point arbitraire on a (voir !)

$$\gamma(t^*+) - \gamma(t^*-) = [g(t^*+) - g(t^*-)] - p^* = 0$$

(où p^* est le saut de p au point t^*), par la continuité à droite de g et de p.

Propriété 5 *Une fonction monotone sur un intervalle* $[a.b]$ *admet presque partout sur cet intervalle une dérivée finie.*

On pose $R(t) := \frac{g(t)-g(t_0)}{t-t_0}$ et on introduit les notations suivantes :

$$\Lambda_{dr}(t_0) := \limsup_{t \downarrow t_0} R(t),\ \lambda_{dr}(t_0) := \liminf_{t \downarrow t_0} R(t),\ \Lambda_{ga}(t_0) := \limsup_{t \uparrow t_0} R(t),\ \lambda_{ga}(t_0) := \liminf_{t \uparrow t_0} R(t).$$

On a toujours $\lambda_{dr} \leqslant \Lambda_{dr}$ et $\lambda_{ga} \leqslant \Lambda_{ga}$. (Pour simplifier on ne fait pas apparaître la dépendance en t_0.)

Il faut prouver que $-\infty < \lambda_{ga} = \lambda_{dr} = \Lambda_{ga} = \Lambda_{dr} < \infty$ a lieu en presque tous $t_0 \in [a, b]$.

On commence par prouver le résultat pour g continue non décroissante. Il suffit de prouver que, presque partout, on a

$$\Lambda_{dr} < \infty \ \text{ et } \ \lambda_{ga} \geqslant \Lambda_{dr}.$$

En effet, si l'on pose $\hat{g}(t) = -g(-t)$, g sera continue non décroissante sur $[-b, -a]$. Alors, on peut voir que pour tout $t_0 \in (a, b)$, on a

$$\Lambda_{ga}(t_0) = \hat{\Lambda}_{dr}(-t_0),\ \lambda_{dr}(t_0) = \hat{\lambda}_{ga}(-t_0).$$

Donc en appliquant la deuxième inégalité à $\hat{g}$ on obtient $\lambda_{dr} \geqslant \Lambda_{ga}$, d'où, presque partout, $\Lambda_{dr} \leqslant \lambda_{ga} \leqslant \Lambda_{ga} \leqslant \lambda_{dr} \leqslant \Lambda_{dr}$.

Montrons d'abord que $\Lambda_{dr} < \infty$ presque partout. Supposons que $\Lambda_{dr} = \infty$ en un point t_0. Alors, pour toute constante C il existe $\tau > t_0$ tel que $R(\tau) > C$ ou

$$g(\tau) - g(t_0) > C(\tau - t_0) \Leftrightarrow g(\tau) - C\tau > g(t_0) - Ct_0.$$

Soit la fonction continue $h(t) = g(t) - Ct$. On voit que t_0 est un point invisible à droite pour h (voir le lemme suivant admis) :

Lemme 1 *(Riesz)*
Soit h une fonction continue sur $[a, b]$. Un point t_0 est dit invisible à droite pour h s'il existe un point τ, $t_0 < \tau \leqslant b$ tel que $h(t_0) < h(\tau)$. Alors, l'ensemble des points invisibles à droite pour h est un ouvert de $[a, b]$ et, par conséquent est une réunion finie ou dénombrable d'intervalles ouverts deux à deux disjoints (a_k, b_k) (et peut être un intervalle semi-ouvert d'extrémité a). Pour chacun de ces intervalles on a $h(a_k) \leqslant h(b_k)$.

Par le lemme on déduit que

$$g(a_k) - Ca_k \leqslant g(b_k) - Cb_k \Leftrightarrow g(b_k) - g(a_k) \geqslant C(b_k - a_k),$$

d'où

$$\sum_k (b_k - a_k) \leqslant \sum_k \frac{g(b_k) - g(a_k)}{C} \leqslant \frac{g(b) - g(a)}{C}.$$

Comme C peut être aussi grande que l'on veut, on en déduit que l'ensemble des points pour lesquels Λ_{dr} peut être recouvert par une famille d'intervalles dont la somme des longueurs est aussi petite que l'on veut. Par conséquent, cet ensemble est de mesure de Lebesgue nulle.

Montrons maintenant que $\lambda_{ga} \geqslant \Lambda_{dr}$ presque partout. On considère deux rationnels $0 < c < C < \infty$ et on pose $\rho = c/C$. On désigne par $E_{c,C}$ l'ensemble des t pour lesquels $\Lambda_{dr} > C$ et $\lambda_{ga} < c$. On va montrer que $E_{c,C}$ est de mesure de Lebesgue nulle et comme l'ensemble des points pour lesquels $\lambda_{ga} < \Lambda_{dr}$ est une réunion finie ou dénombrable d'ensembles de la forme $E_{c,C}$, on aura la conclusion. Pour montrer que $\ell(E_{c,C}) = 0$ (la mesure de Lebesgue sur $\mathbb{R}$ sera notée ici ℓ) on a besoin du lemme suivant (admis) :

Lemme 2 *Soit un sous ensemble E de l'intervalle $[a, b]$ tel que pour tout intervalle $(\alpha, \beta) \subset [a, b]$ on a $\ell\,(E \cap (\alpha, \beta)) \leqslant \rho(\beta - \alpha)$, où $0 < \rho < 1$. Alors $\ell(E) = 0$.*

Soit l'ensemble des $t \in (\alpha, \beta)$ pour lesquels $\lambda_{ga} < C$. À tout point de cet ensemble on peut associer un $\tau < t$ tel que

$$\frac{g(\tau) - g(t)}{\tau - t} < c \Leftrightarrow g(\tau) - c\tau > g(t) - c\,t.$$

Ainsi t est invisible à gauche pour $g(t) - ct$ (définition identique), donc d'après un résultat similaire au Lemme 1, l'ensemble de tous ces t est une réunion finie ou dénombrable d'intervalles $(\alpha_k, \beta_k) \subset (\alpha, \beta)$ et $g(\beta_k) - c\beta_k \leqslant g(\alpha_k) - c\alpha_k$, c'est-à-dire

$$g(\beta_k) - g(\alpha_k) \leqslant c(\beta_k - \alpha_k).$$

Sur chacun des intervalles (α_k, β_k) on considère l'ensemble G_k des points t pour lesquels $\Lambda_{dr} > C$. On reprend un raisonnement identique et on déduit que G_k est une réunion finie ou dénombrable d'intervalles $(\alpha_{kj}, \beta_{kj})$ et que

$$\beta_{kj} - \alpha_{kj} \leqslant \frac{1}{C}[g(\beta_{kj}) - g(\alpha_{kj})].$$

Il est clair qu'on peut recouvrir $E_{c,C} \cap (\alpha, \beta)$ par des intervalles $(\alpha_{kj}, \beta_{kj})$ et on a

$$\sum_{k,j}(\beta_{kj} - \alpha_{kj}) \leqslant \frac{1}{C}\sum_{k,j}[g(\beta_{kj}) - g(\alpha_{kj})] \leqslant \frac{1}{C}\sum_{k}[g(\beta_k) - g(\alpha_k)] \leqslant \frac{c}{C}\sum_{k}(\beta_k - \alpha_k) \leqslant \rho(\beta - \alpha).$$

Le Lemme 2 s'applique et le résultat est prouvé dans le cas d'une fonction g continue non décroissante (à l'exception des Lemmes 1 et 2). Si g est monotone discontinue on peut utiliser un résultat identique au Lemme 1 à condition de bien définir la notion de point t_0 invisible à droite dans ce cas : il existe $\tau > t_0$ tel que $\max\{g(t_0-), g(t_0), g(t_0+)\} < g(\tau)$.

Propriété 6 *La fonction de sauts d'une fonction monotone a une dérivée nulle presque partout.*

En effet, une telle fonction est la somme d'une série convergente de fonctions non décroissantes de la forme $q_n(t) = p_n \mathbb{1}_{\{t>t_n\}}$ dont chacune a une dérivée presque partout nulle. Il reste à utiliser le théorème de dérivation terme à terme d'une série de fonctions monotones (admis).

Propriété 7 *Pour toute fonction intégrable au sens de Lebesgue f on a presque partout l'égalité suivante :* $\frac{d}{dx}\int_a^x f(t)dt = f(x)$.

On note $F(x) = \int_a^x f(t)dt$. Montrons d'abord que $f(x) \geqslant F'(x)$ presque partout. En effet, si $f(x) < F'(x)$, il existe deux rationnels tels que $f(x) < \alpha < \beta < F'(x)$. Soit $E_{\alpha,\beta}$ l'ensemble des x pour lesquels a lieu cette inégalité. Cet ensemble est borélien car f et F' le sont. On va montrer que la mesure de Lebesgue de chacun des ensembles $E_{\alpha,\beta}$ est nulle. Comme leur nombre est au plus dénombrable on en déduit que $\ell(\{x : f(x) < F'(x)\}) = 0$.

Soit $\varepsilon > 0$ et soit $\delta > 0$ tel que $|\int_A f(t)dt| < \varepsilon$, dès que $\ell(A) < \delta$. On choisit l'ouvert $G \subset [a, b]$ tel que $G \supset E_{\alpha,\beta}$ et $\ell(G) < \ell(E_{\alpha,\beta}) + \delta$. Si $x \in E_{\alpha,\beta}$, alors pour tous les $\xi > x$ voisins de x,

$$\frac{F(\xi) - F(x)}{\xi - x} > \beta \Leftrightarrow F(\xi) - \beta\xi > F(x) - \beta x.$$

Donc x est un point invisible à droite pour la fonction $F(x) - \beta x$ sur chacun des intervalles composant G. Par le Lemme 1, on peut indiquer un ouvert $S = \cup_k (a_k, b_k)$ tel que $E_{\alpha,\beta} \subset S \subset G$ et

$$F(b_k) - \beta b_k \geqslant F(a_k) - \beta a_k \Leftrightarrow F(b_k) - F(a_k) \geqslant \beta(b_k - a_k) \Leftrightarrow \int_{a_k}^{b_k} f(t)dt \geqslant \beta(b_k - a_k),$$

d'où

$$\int_S f(t)dt \geqslant \beta\ell(S).$$

D'autre part

$$\int_S f(t)dt = \int_{E_{\alpha,\beta}} f(t)dt + \int_{S\setminus E_{\alpha,\beta}} f(t)dt \leqslant \alpha\ell(E_{\alpha,\beta}) + \varepsilon \leqslant \alpha\ell(S) + \varepsilon.$$

On en déduit

$$\alpha\ell(S) + \varepsilon \geqslant \beta\ell(S) \Leftrightarrow \ell(S) \leqslant \frac{\varepsilon}{\beta - \alpha}.$$

Ainsi, l'ensemble $E_{\alpha,\beta}$ peut être enfermé dans un ensemble ouvert de mesure arbitrairement petite donc $\ell(E_{\alpha,\beta}) = 0$.

Nous avons prouvé que $f(x) \geqslant F'(x)$ presque partout. En remplaçant $f(x)$ par $-f(x)$ on peut montrer que $-f(x) \geqslant -F'(x)$ donc $f(x) \leqslant F'(x)$ presque partout. L'égalité est prouvée.

Propriété 8 *La dérivée F' d'une fonction monotone non décroissante F est intégrable au sens de Lebesgue et $\int_a^b F'(x)dx \leqslant F(b) - F(a)$.*

Par définition, la dérivée de F au point x est la limite de $\Phi_h(x) = \frac{1}{h}(F(x+h) - F(x))$ quand $h \to 0$. La monotonie de F entraîne son intégrabilité donc l'intégrabilité de chacune des fonctions Φ_h. On peut donc intégrer :

$$\int_a^b \Phi_h(x)dx = \frac{1}{h}\int_a^b F(x+h)dx - \frac{1}{h}\int_a^b F(x)dx = \frac{1}{h}\int_b^{b+h} F(x)dx - \frac{1}{h}\int_a^{a+h} F(x)dx.$$

Le membre de droite tend vers $F(b) - F(a+)$ quand $h \downarrow 0$. Par le lemme de Fatou on obtient

$$\int_a^b F'(x)dx \leqslant \lim_{h\to 0}\int_a^b \Phi_h(x)dx = F(b) - F(a+) \leqslant F(b) - F(a)$$

(l'existence de l'intégrale de F' est assurée également par le lemme de Fatou). Il est facile de voir que $F : [0,1] \to \mathbb{R}$ donnée par $F(x) = \mathbb{1}_{(\frac{1}{2},1]}(x)$ satisfait l'inégalité stricte !

Définition 1.17 *Une fonction F sur un intervalle $[a,b]$ est dite à variation bornée s'il existe une constante C telle que pour toute subdivision $a = x_0 < x_1 < \ldots < x_n = b$ on ait la variation totale $\mathbb{V}_a^b[F] := \sum_k |F(x_k) - F(x_{k-1})| \leqslant C$.*

Propriété 9 *(admise)*
Toute fonction monotone est à variation bornée. Toute fonction à variation bornée est différence de deux fonctions monotones non décroissantes.

Définition 1.18 *Une fonction F sur un intervalle $[a,b]$ est dite absolument continue sur $[a,b]$, si pour tout $\varepsilon > 0$ il existe $\delta > 0$ tel que pour toute famille au plus dénombrable d'intervalles deux à deux disjoints (a_k, b_k) satisfaisant $\sum_k(b_k - a_k) < \delta$, on a $\sum_k |F(b_k) - F(a_k)| < \varepsilon$.*

Propriété 10 *(admise)*
Toute fonction absolument continue est uniformément continue. Toute fonction absolument continue est à variation bornée. Toute fonction absolument continue est différence de deux fonctions absolument continues monotones non décroissantes.

Propriété 11 *Si f est une fonction intégrable au sens de Lebesgue alors $F(x) = \int_a^x f(t)dt$ est une fonction absolument continue.*

Si $\{(a_k, b_k)\}$ est une famille quelconque d'intervalles deux à deux disjoints alors

$$\sum_k |F(b_k) - F(a_k)| = \sum_k \left| \int_{a_k}^{b_k} f(t)dt \right| \leqslant \sum_k \int_{a_k}^{b_k} |f(t)|dt = \int_{\cup_k (a_k, b_k)} |f(t)|dt.$$

En vertu de la continuité absolue de l'intégrale de Lebesgue, la dernière expression tend vers zéro quand la longueur totale des intervalles (a_k, b_k) tend vers zéro.

Propriété 12 *La dérivée F' d'une fonction absolument continue sur un intervalle $[a, b]$ est intégrable au sens de Lebesgue sur cet intervalle et pour tout x, $a \leqslant x \leqslant b$ on a $\int_a^x F'(t)dt = F(x) - F(a)$.*

Il suffit de prouver le résultat pour le cas où F est monotone non décroissante. Alors $\Phi(x) = F(x) - \int_a^x F'(t)dt$ est aussi monotone non décroissante : si $x'' > x'$ on a

$$\Phi(x'') - \Phi(x') = F(x'') - F(x') - \int_{x'}^{x''} F'(t)dt \geqslant 0.$$

D'autre part Φ est absolument continue comme différence de deux fonctions absolument continues et $\Phi'(x) = 0$ presque partout (par **7**). On a besoin d'un lemme (admis) :

Lemme 3 *Si la dérivée d'une fonction absolument continue monotone non décroissante est nulle presque partout, alors cette fonction est une constante.*

On en déduit que Φ est une constante. En faisant $x = a$ dans sa définition on trouve que cette constante est égale à $F(a)$.

Cela termine les idées sur les fonctions monotones. □

Exemples : i) Soit $\lambda > 0$ et $F(t) = (1 - e^{-\lambda t})\mathbb{1}_{[0,\infty[}(t)$. C'est une fonction de répartition dérivable et la densité est $f(t) = \lambda e^{-\lambda t}\mathbb{1}_{[0,\infty[}(t)$ (loi exponentielle).
ii) $F = \mathbb{1}_{[x,\infty[}$ est la fonction de répartition de la masse de Dirac en x, δ_x.

Exemple : (lois gaussiennes)
Soit la fonction borélienne positive

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}},\ x \in \mathbb{R}.$$

On vérifie que $\int_{\mathbb{R}} f(x)dx = 1$, donc que f est une densité de probabilité. Le graphe de cette fonction est la “cloche de Gauss”, symétrique par rapport à 0. C'est la densité de la loi **gaussienne** (ou **normale**) **centrée réduite**, notée $\mathcal{N}(0, 1)$.

Sa fonction de répartition est notée

$$\Phi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-\frac{x^2}{2}} dx,\ t \in \mathbb{R}.$$

Cette fonction n'est pas connue explicitement, mais les valeurs sont tabulées. On peut écrire, pour tout $t \in \mathbb{R}$,

$$\Phi(-t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-t} e^{-\frac{x^2}{2}} dx = \frac{1}{\sqrt{2\pi}} \int_{t}^{+\infty} e^{-\frac{x^2}{2}} dx$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{x^2}{2}} dx - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-\frac{x^2}{2}} dx = 1 - \Phi(t).$$

En particulier $\Phi(0) = \frac{1}{2}$.

Si X a la fonction de répartition Φ alors $X \sim \mathcal{N}(0,1)$. De plus, par le calcul précédent $-X \sim \mathcal{N}(0,1)$, donc X et $-X$ ont la même loi (loi symétrique).

Soit X une variable aléatoire gaussienne centrée réduite. Y est une **variable aléatoire gaussienne** s'il existe deux réels m, σ tels que $Y = \sigma X + m$. Si $\sigma = 0$, la gaussienne Y est dite dégénérée. Si $\sigma > 0$, la fonction de répartition de Y est

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

C'est une fonction dérivable et donc la densité de Y est

$$f_Y(y) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(y-m)^2}{2\sigma^2}\right),\ y \in \mathbb{R}.$$

On écrit $Y \sim \mathcal{N}(m, \sigma^2)$.

Proposition 1.12 *Soit F une fonction de répartition. On appelle* **fonction de quantile** *la fonction*

$$G(u) = \inf\{t : F(t) > u\},\ u \in]0,1[.$$

Soit U une variable aléatoire uniforme sur $]0,1[$, c'est-à-dire de densité $f_U(x) = \mathbb{1}_{]0,1[}(x)$. Alors $G(U)$ a pour fonction de répartition F.

Preuve : On voit que pour tout $t \in \mathbb{R}$

$$G(U) \leqslant t \Leftrightarrow \inf\{s : F(s) > U\} \leqslant t.$$

Cette dernière assertion est satisfaite dès que $F(t) > U$. Autrement dit on a $\{U < F(t)\} \subset \{G(U) \leqslant t\}$. D'autre part si l'inégalité $\inf\{s : F(s) > U\} \leqslant t$ est vérifiée, alors $F(s) > U$ pour tout $s > t$, c'est-à-dire on a $\{G(U) \leqslant t\} \subset \{U < F(s)\}$. Comme U suit la loi uniforme, il en découle des deux inclusions que, pour tout $s > t$

$$F(t) = \mathrm{P}(U < F(t)) \leqslant \mathrm{P}(G(U) \leqslant t) \leqslant \mathrm{P}(U \leqslant F(s)) = F(s).$$

Comme F est continue à droite, en faisant tendre s vers t, on obtient $\mathrm{P}(G(U) \leqslant t) = F(t)$ pour tout $t \in \mathbb{R}$, ce qui est le résultat. □

Remarque : On peut voir que G est croissante, continue à droite et admet une limite à gauche en tout point de $]0,1[$. Pour voir la continuité à droite il suffit de remarquer que $\{t : F(t) > u\} = \cup_{n\in\mathbb{N}^*}\{t : F(t) > u+\frac{1}{n}\}$. On peut aussi prouver que $F(t) = \inf\{u : G(u) > t\}$ (d'ici aussi le nom, **inverse continue à droite de** F). On peut vérifier que si F est inversible, G est l'inverse de F au sens usuel. La proposition précédente permet de générer des variables aléatoires de loi arbitraire à partir d'une variable uniforme (fournie par l'ordinateur sous forme des nombres pseudo-aléatoires). □

Définition 1.19 *Soit $X = (X_1, \ldots, X_d)$ un vecteur aléatoire à valeurs dans $\mathbb{R}^d$. On appelle* **fonction de répartition** *de X ou de sa loi P_X, la fonction sur $\mathbb{R}^d$ définie par*

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

La loi de la variable aléatoire réelle X_j est appelée la **j-ème marginale** *de X. Elle est donnée par sa fonction de répartition :*

$$F_{X_j}(t_j) = \lim_{t_1, \ldots, t_{j-1}, t_{j+1}, \ldots, t_d \to \infty} F_X(t).$$

Remarque : La loi d'un vecteur aléatoire détermine chacune des lois marginales, mais la réciproque est fausse en général. Voici un exemple : soit deux couples aléatoires (X, Y) et (U, V) dont les lois discrètes sont données par

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

et

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

Les lois marginales sont données par

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

On produit deux couples aléatoires de lois différentes mais ayant les mêmes lois marginales.□

Remarque : Supposons que la loi du couple (X, Y) est discrète à valeurs dans $\{x_i : i \in I\} \times \{y_j : j \in J\}$:

$$p_{i,j} = \mathrm{P}(X = x_i, Y = y_j),\, i \in I, j \in J.$$

Alors les lois de X et de Y sont données par :

$$p_i = \mathrm{P}(X = x_i) = \sum_{j \in J} p_{i,j} \text{ et } q_j = \mathrm{P}(Y = y_j) = \sum_{i \in I} p_{i,j}.$$

En effet

$$p_i = \mathrm{P}(X = x_i) = \mathrm{P}(X = x_i, Y \in \{y_j : j \in J\})$$
$$= \mathrm{P}\left(\bigcup_{j \in J}(X = x_i, Y = y_j)\right) = \sum_{j \in J} \mathrm{P}(X = x_i, Y = y_j).$$

□

Remarque : Supposons que la loi du couple (X, Y) admet une densité $f_{X,Y}$. Alors les deux lois marginales admettent des densités :

$$f_X(x) = \int_{\mathbb{R}} f_{X,Y}(x, y)\, dy \text{ et } f_Y(y) = \int_{\mathbb{R}} f_{X,Y}(x, y)\, dx.$$

En effet,

$$\mathrm{P}_X(A) = \mathrm{P}(X \in A) = \mathrm{P}((X, Y) \in A \times \mathbb{R})$$
$$= \iint_{\mathbb{R}^2} \mathbb{1}_A(x) f_{X,Y}(x, y)\, dx\, dy = \int_{\mathbb{R}} \mathbb{1}_A(x) \left(\int_{\mathbb{R}} f_{X,Y}(x, y)\, dy\right) dx.$$

□

1.8 Lois de probabilités usuelles

1. Loi de Bernoulli.

$$\mathrm{Q} = (1-p)\delta_0 + p\delta_1,\ p \in [0,1].$$

X suit la loi de Bernoulli, notée $\mathcal{B}(1,p)$ si

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

Si $p=0$ (respectivement $p=1$) X est presque sûrement égale à 0 (respectivement 1).

Il s'agit de modéliser une expérience à deux issues possibles (succès et échec) dont probabilité de succès vaut p. Le résultat X de cette expérience suit la loi de Bernoulli de paramètre p.

2. Loi binomiale.

$$\mathrm{Q} = \sum_{k=0}^{n} C_n^k p^k (1-p)^{n-k}\delta_k,\ p \in]0,1[.$$

X suit la loi binomiale, notée $\mathcal{B}(n,p)$ si

$$\mathrm{P}(X=k) = C_n^k p^k (1-p)^{n-k},\ \text{ si } \ k=0,\dots,n$$

et $\mathrm{P}(X=k)=0$ sinon.

Il s'agit de répeter n fois l'expérience à deux issues possibles dont probabilité de succès vaut p, en assurant chaque fois les mêmes conditions initiales (indépendance des répétitions). Le nombre X de succès obtenus suit la loi binomiale.

3. Loi uniforme.

$$\mathrm{Q} = \frac{1}{r}\sum_{j=1}^{r}\delta_{x_j},\ r \in \mathbb{N}^*$$

X suit la loi uniforme, notée $\mathcal{U}(x_1,\dots,x_r)$ si

$$\mathrm{P}(X=x_j) = \frac{1}{r},\ j=1,\dots,r.$$

On modélise une expérience à r issues possibles dont chacune a la même probabilité $1/r$. Le résultat X de cette expérience suit la loi uniforme.

4. Loi géométrique.

$$\mathrm{Q} = \sum_{k=1}^{\infty} p(1-p)^{k-1}\delta_k,\ p \in]0,1[.$$

X suit la loi gémétrique, notée $\mathcal{G}(p)$ si

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

On reprend la modélisation de la loi binomiale, mais cette fois-ci X note la première fois où le succès est obtenu. Alors X suit la loi géométrique.

5. Loi de Poisson.

$$\mathrm{Q} = \sum_{k=0}^{\infty} e^{-\lambda}\frac{\lambda^k}{k!}\delta_k,\ \lambda > 0.$$

X suit la loi de Poisson, notée $\mathcal{P}(\lambda)$ si

$$\mathrm{P}(X = k) = e^{-\lambda}\frac{\lambda^k}{k!},\ k \in \mathbb{N}.$$

6. Loi hypergéométrique.
X suit la loi hypergéométrique, notée $\mathcal{H}(N, n, p)$, $N \in \mathbb{N}^*$, $1 \leqslant n \leqslant N$, $p \in]0, 1[$, si

$$\mathrm{P}(X = k) = \frac{C_{Np}^{k} C_{N(1-p)}^{n-k}}{C_N^n},\ \text{ si }\ \max\{0, n - N(1-p)\} \leqslant k \leqslant \min\{n, Np\}$$

et $= 0$ sinon.

Si on tire n boules en une seule fois, sans remise, dans une urne contenant N boules dont Np sont blanches et $N(1-p)$ sont rouges, le nombre X de boules blanches tirées suit la loi hypergéométrique.

7. Loi multinomiale.
$(X_1, \ldots, X_d)$ à valeurs dans $\mathbb{N}^d$, suit la loi multinomiale, notée $\mathcal{M}(n, p_1, \ldots, p_n)$, $n \in \mathbb{N}$, $p_1 + \ldots + p_d = 1$, $p_1, \ldots, p_d \in [0, 1]$:

$$\mathrm{P}((X_1, \ldots, X_d) = (n_1, \ldots, n_d)) = \frac{n!}{n_1! \ldots n_d!} p_1^{n_1} \ldots p_d^{n_d},\ \text{ si }\ n_1 + \ldots n_d = n$$

et $= 0$ sinon.

Si l'on dispose de n boules que l'on jette une par une aléatoirement dans d boîtes différentes, chaque boule ayant la probabilité p_i d'être jetée dans la i-ème boîte, les nombres $(X_1, \ldots, X_d)$ de boules dans les boîtes suivent la loi multinomiale.

8. Loi uniforme.
Une variable aléatoire réelle X suit une loi uniforme sur $[a, b]$, $a < b$, notée $\mathcal{U}_{[a,b]}$, si sa densité par rapport à la mesure de Lebesgue est

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

9. Loi exponentielle.
Une variable aléatoire réelle X suit une loi exponentielle de paramètre $\lambda > 0$, notée $\mathcal{E}(\lambda)$, si sa densité par rapport à la mesure de Lebesgue est

$$f(x) := \lambda e^{-\lambda x}\mathbf{1}_{[0,\infty[}(x).$$

10. Loi gamma.
Une variable aléatoire réelle X suit une loi gamma de paramètres $p > 0$ et $\lambda > 0$, notée $\gamma(p, \lambda)$, si sa densité par rapport à la mesure de Lebesgue est

$$f(x) := \frac{\lambda^p}{\Gamma(p)} x^{p-1} e^{-\lambda x}\mathbf{1}_{]0,\infty[}(x),$$

où $\Gamma(p) := \int_0^\infty x^{p-1}e^{-x}dx$ est la fonction d'Euler de seconde espèce.

10. Loi de chi-deux.
Une variable aléatoire réelle X suit une loi de chi-deux de paramètres $n \in \mathbb{N}^*$, notée $\chi^2(n)$, si sa densité par rapport à la mesure de Lebesgue est

$$f(x) := \frac{\left(\frac{1}{2}\right)^{\frac{n}{2}}}{\Gamma\left(\frac{n}{2}\right)} x^{\frac{n}{2}-1}e^{-\frac{x}{2}}\mathbf{1}_{]0,\infty[}(x)$$

11. Loi beta.
Une variable aléatoire réelle X suit une loi beta de paramètres $p > 0$ et $q > 0$, notée $\beta(p,q)$, si sa densité par rapport à la mesure de Lebesgue est

$$f(x) := \frac{x^{p-1}(1-x)^{q-1}}{B(p,q)}\mathbf{1}_{[0,1]}(x),$$

où $B(p,q) := \int_0^1 x^{p-1}(1-x)^{q-1}dx$ est la fonction d'Euler de première espèece.

11. Loi de Laplace.
Une variable aléatoire réelle X suit une loi de Laplace si sa densité par rapport à la mesure de Lebesgue est

$$f(x) := \frac{1}{2}e^{-|x|}.$$

12. Loi de Cauchy.
Une variable aléatoire réelle X suit une loi de Cauchy si sa densité par rapport à la mesure de Lebesgue est

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

13. Loi de Gauss (ou normale).
Une variable aléatoire réelle X suit une loi de Gauss (ou normale) de paramètres $m \in \mathbb{R}$ et σ^2, $\sigma > 0$, notée $\mathcal{N}(m,\sigma^2)$, si sa densité par rapport à la mesure de Lebesgue est

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

13. Loi de Gauss (ou normale) multidimensionnelle.
Une variable aléatoire $(X_1,\ldots,X_d)$ à valeurs dans $\mathbb{R}^d$ suit une loi de Gauss (ou normale) de paramètres $m \in \mathbb{R}^d$ et Σ, matrice inversible, notée $\mathcal{N}(m,\Sigma)$, si sa densité par rapport à la mesure de Lebesgue sur $\mathbb{R}^d$ est

$$f(x) := \frac{1}{(2\pi)^{\frac{d}{2}}\sqrt{\det\Sigma}}\exp\left(-\frac{1}{2}(x-m)^*\Sigma^{-1}(x-m)\right), \quad x \in \mathbb{R}^d.$$

Chapitre 2

Espérance des variables aléatoires

2.1 Définitions et théorèmes de convergence

Dans tout ce chapitre nous considérons des variables aléatoires d'un espace de probabilité complet $(\Omega, \mathcal{A}, \mathrm{P})$ à valeurs dans $\mathbb{R}$ ou $\mathbb{R}^d$.

Définition 2.1 *Soit $A \in \mathcal{A}$ un événement. La variable aléatoire $X = \mathbb{1}_A$ est* **intégrable** *et son* **espérance** *est donnée par :*

$$\mathrm{E}(X) = \mathrm{E}(\mathbb{1}_A) = \int_\Omega X(\omega)\mathrm{P}(d\omega) := \mathrm{P}(A).$$

Définition 2.2 *Si X est une variable aléatoire étagée positive $X = \sum_{j=1}^r a_j \mathbb{1}_{A_j}$ avec $a_1, \ldots, a_r \geqslant 0$ et les $A_1, \ldots, A_r$ événements disjoints, alors X est* **intégrable** *et son espérance est donnée par :*

$$\mathrm{E}(X) := \sum_{j=1}^r a_j \mathrm{P}(A_j) = \sum_{j=1}^r a_j \int_\Omega \mathbb{1}_{A_j} d\mathrm{P}.$$

Remarque : On peut vérifier la linéarité de l'espérance sur les variables aléatoires étagées positives. □

Définition 2.3 *Soit X une variable aléatoire positive sur $(\Omega, \mathcal{A}, \mathrm{P})$. On définit et on note son espérance par :*

$$\mathrm{E}(X) = \int_\Omega X\, d\mathrm{P} = \int_\Omega X(\omega)\mathrm{P}(d\omega)$$
$$:= \sup\left\{\mathrm{E}(Y) : Y \text{ variable aléatoire étagée positive },\ Y \leqslant X\right\}.$$

Remarque : On peut voir que l'espérance d'une variable aléatoire positive peut être infinie. □

Proposition 2.1 *(i) Si $0 \leqslant X \leqslant Y$, alors $0 \leqslant \mathrm{E}(X) \leqslant \mathrm{E}(Y)$.*

(ii) Si $A \subset B$ et $X \geqslant 0$, alors $\mathrm{E}(X\mathbb{1}_A) \leqslant \mathrm{E}(X\mathbb{1}_B)$.

(iii) Si $X \geqslant 0$ et $\alpha \geqslant 0$, alors $\mathrm{E}(\alpha X) = \alpha \mathrm{E}(X)$.

(iv) Si $X = 0$, alors $\mathrm{E}(X) = 0$.

(v) Si $\mathrm{P}(B) = 0$*, alors* $\mathrm{E}(X\mathbb{1}_B) = 0$.

Preuve : L'idée est d'établir la proposition sur des variables aléatoires étagées, puis passer au supremum pour les variables positives. Prouvons, suivant ce schéma, *(iv)*. On voit que si $X = \sum_{j=1}^r a_j \mathbb{1}_{A_j}$, alors

$$\mathrm{E}(\alpha X) = \sum_{j=1}^r \alpha a_j \mathrm{P}(A_j) = \alpha \sum_{j=1}^r a_j \mathrm{P}(A_j) = \alpha \mathrm{E}(X).$$

□

Théorème 2.1 *(convergence monotone)*

1) Soit $(X_n)_{n\in\mathbb{N}}$ *une suite croissante de variables aléatoires positives sur* $(\Omega, \mathcal{A}, \mathrm{P})$*, convergeant ponctuellement vers* X*. Alors l'espérance de la variable aléatoire positive* X *satisfait :*

$$\mathrm{E}(X) = \lim_{n\to\infty} \mathrm{E}(X_n).$$

2) Soit $(Y_n)_{n\in\mathbb{N}}$ *une suite de variables aléatoires positives. Si on note* $Y = \sum_{n\geqslant 0} Y_n < \infty$*, alors*

$$\mathrm{E}(Y) = \sum_{n\geqslant 0} \mathrm{E}(Y_n).$$

Preuve : X est une variable aléatoire par le Théorème 1.1. Comme X_n est croissante positive, d'après la Proposition 2.1 $\mathrm{E}(X_n)$ est une suite croissante et positive, donc admet une limite $\ell \geqslant 0$ (éventuellement $+\infty$). Comme $X_n \leqslant X$, par la Proposition 2.1 on voit que $\ell \leqslant \mathrm{E}(X)$. Soit

$$0 \leqslant Y = \sum_{j=1}^r b_j \mathbb{1}_{B_j} \leqslant X$$

une variable aléatoire positive étagée. Soit $0 \leqslant c < 1$. On a

$$\mathrm{E}(X_n) \geqslant \mathrm{E}\left[\mathbb{1}_{\{X_n \geqslant cY\}} X_n\right] \leqslant c\mathrm{E}\left[Y\mathbb{1}_{\{X_n\geqslant cY\}}\right] = c\sum_{j=1}^r b_j \mathrm{P}\left(B_j \cap \{X_n \geqslant cY\}\right)$$

d'après la Proposition 2.1 et la Définition 2.2. Quand $n \to \infty$,

$$\ell \geqslant c \sum_{j=1}^r b_j \lim_{n\to\infty} \mathrm{P}\left(B_j \cap \{X_n \geqslant cY\}\right) = c\sum_{j=1}^r b_j \mathrm{P}(B_j) = c\mathrm{E}(Y),$$

la seconde égalité résultant de la Proposition 1.8 3) et du fait que $\cup_n \{X_n \geqslant cY\} = \Omega$. Mais c était arbitraire dans $[0, 1[$, donc $\ell \geqslant \mathrm{E}(Y)$ pour toute variable étagée $0 \leqslant Y \leqslant X$. Donc, par définition $\ell \geqslant \mathrm{E}(X)$ et le théorème est prouvé. □

Théorème 2.2 *(lemme de Fatou)*
Soit $(X_n)_{n\in\mathbb{N}}$ *une suite de variables aléatoires positives. Alors*

$$\mathrm{E}\left(\liminf_{n\to\infty} X_n\right) \leqslant \liminf_{n\to\infty} \mathrm{E}(X_n).$$

Preuve : On note $Y_n = \inf_{m \geqslant n} X_m$. Cette suite est croissante et converge simplement vers $\liminf_{n \to \infty} X_n$. De plus $Y_n \leqslant X_n$. Il suffit d'appliquer le Théorème 2.1 et d'utiliser la Proposition 2.1 (i). □

Soit X une variable aléatoire réelle. On note $X^+ = X \vee 0$ sa partie positive et $X^- = (-X) \vee 0$ sa partie négative. Alors $X = X^+ - X^-$ et $|X| = X^+ + X^-$. On sait que X^+ et X^- sont aussi des variables aléatoires.

Définition 2.4 *Soit* $X = X^+ - X^-$ *une variable aléatoire réelle. On dit que* X *est* **intégrable** *si* $\mathrm{E}(X^+) < \infty$ *et* $\mathrm{E}(X^-) < \infty$, *ou, equivalent, si* $\mathrm{E}(|X|) < \infty$. *Dans ce cas, son* **espérance** *est :*

$$\mathrm{E}(X) := \mathrm{E}(X^+) - \mathrm{E}(X^-).$$

Une variable aléatoire d'espérance 0 *est dite* **centrée**.

Exemple : Soit l'espace de probabilité $(\Omega, \mathcal{A}, \delta_{\omega_0})$ où δ_{ω_0} est la masse de Dirac dans un point (quelconque) fixé $\omega_0 \in \Omega$. Soit X une variable aléatoire quelconque. Alors X est intégrable par rapport à la masse de Dirac et

$$\int X(\omega)\, \delta_{\omega_0}(d\omega) = X(\omega_0).$$

Plus généralement, si $\mathrm{P} = \sum_{j=1}^r a_j \delta_{\omega_j}$ avec $a_j \geqslant 0$, $\sum_{j=1}^r a_j = 1$ et $\omega_j \in \Omega$, alors $\mathrm{E}(X) = \sum_{j=1}^r a_j X(\omega_j)$.

Définition 2.5 *Soit* X *un vecteur aléatoire à valeurs dans* $\mathbb{R}^d$, $X = (X_1, \ldots, X_d)$. X *est intégrable si et seulement si* X_j *sont intégrables,* $j = 1, \ldots, d$, *et*

$$\mathrm{E}(X) = (\mathrm{E}(X_1), \ldots, \mathrm{E}(X_d)).$$

Remarque : En particulier une variable aléatoire à valeurs complexes est intégrable si et seulement si les variables aléatoires parties réelle et imaginaire sont intégrables. □

Proposition 2.2 *Si* X, Y *sont intégrables et si* $\alpha, \beta \in \mathbb{R}$, *alors*

$$\mathrm{E}(\alpha X + \beta Y) = \alpha \mathrm{E}(X) + \beta \mathrm{E}(Y).$$

De plus, si $X \leqslant Y$, *alors* $\mathrm{E}(X) \leqslant \mathrm{E}(Y)$.

Preuve : On suppose d'abord $X, Y \geqslant 0$ et $\alpha, \beta \geqslant 0$. D'après la Proposition 1.6 il existe des suites $\{X_n\}_{n \in \mathbb{N}}$ et $\{Y_n\}_{n \in \mathbb{N}}$ de variables aléatoires étagées qui convergent en croissant vers X et Y respectivement. Alors la suite $\alpha X_n + \beta Y_n$ converge en croissant vers $\alpha X + \beta Y$, et le résultat se déduit du théorème de convergence monotone. En général on sépare les parties positive et négative et on distinge selon les signes de α et β. Si $X \leqslant Y$, alors $Y - X \geqslant 0$, donc d'après la Propostion 2.1, $\mathrm{E}(Y - X) \geqslant 0$ et la conclusion s'ensuit par linéarité. □

Proposition 2.3 *Soit Y intégrable et soit $\{X_n\}_{n\in\mathbb{N}}$ une suite de variables aléatoires intégrables.*

i) Si $Y \leqslant X_n$, alors $\mathrm{E}(\liminf_{n\to\infty} X_n) \leqslant \liminf_{n\to\infty} \mathrm{E}(X_n)$.

ii) Si $X_n \leqslant Y$, alors $\limsup_{n\to\infty} \mathrm{E}(X_n) \leqslant \mathrm{E}(\limsup_{n\to\infty} X_n)$.

Preuve : Par le lemme de Fatou on a

$$\mathrm{E}[\liminf_{n\to\infty}(X_n - Y)] \leqslant \liminf_{n\to\infty} \mathrm{E}(X_n - Y)$$

ce qui prouve (i) par linéarité de l'espérance. De même

$$\mathrm{E}[\liminf_{n\to\infty}(Y - X_n)] \leqslant \liminf_{n\to\infty} \mathrm{E}(Y - X_n).$$

□

Théorème 2.3 *(convergence dominée de Lebesgue)*
Soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires telles que $|X_n| \leqslant Y$, où Y est intégrable et la suite converge simplement vers X. Alors

$$\lim_{n\to\infty} \mathrm{E}(X_n) = \mathrm{E}(X).$$

Preuve : Comme $\lim_{n\to\infty} X_n = X$ et $-Y \leqslant X_n \leqslant Y$, la Proposition 2.3 donne

$$\limsup_{n\to\infty} \mathrm{E}(X_n) \leqslant \mathrm{E}(\limsup_{n\to\infty} X_n) = \mathrm{E}(X) = \mathrm{E}(\liminf_{n\to\infty} X_n) \leqslant \liminf_{n\to\infty} \mathrm{E}(X_n).$$

□

Proposition 2.4 *(inégalité de Jensen)*
Si $\varphi : \mathbb{R} \to \mathbb{R}$ est une fonction convexe et si $X : \Omega \to \mathbb{R}$ est une variable aléatoire, telle que X et $\varphi(X)$ soient intégrables, alors

$$\varphi\left(\mathrm{E}(X)\right) \leqslant \mathrm{E}\left(\varphi(X)\right).$$

Preuve : La convexité de φ assure qu'en tout point son graphe est au-dessus de sa tangente : pour tout $t \in \mathbb{R}$, il existe δ (on peut prendre pour δ la dérivée à gauche ou à droite de φ en t) tel que

$$\varphi(x) \geqslant \varphi(t) + \delta(x - t).$$

On applique cette inégalité à $t = \mathrm{E}(X)$ et $x = X(\omega)$, pour tout ω, et on intégre les deux membres. D'après la Proposition 2.2 on déduit la conclusion. □

Remarque : Si φ est strictement convexe, l'égalité $\varphi\left(\mathrm{E}(X)\right) = \mathrm{E}\left(\varphi(X)\right)$ n'a lieu que si X est p.s. constante. De plus, si l'égalité a lieu pour toute variable aléatoire X, alors φ est linéaire.□

Remarque : Dans la pratique, l'inégalité de Jensen est le plus souvent utilisée pour les fonctions $\varphi(x) = |x|$, x^2, et $1/x$ lorsque $x > 0$. En particulier, une variable aléatoire dont le carré est intégrable est intégrable, et si X est à valeurs strictement positives,

$$\mathrm{E}\left(\frac{1}{X}\right) \geqslant \frac{1}{\mathrm{E}(X)}.$$

□

2.2 Théorème de transport

Théorème 2.4 *(de transport)*
Soit X un vecteur aléatoire défini sur $(\Omega, \mathcal{A}, \mathrm{P})$ à valeurs dans $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ et soit ϕ une fonction borélienne de $\mathbb{R}^d$ dans $\mathbb{R}$. Alors $\phi(X)$ est intégrable si et seulement si ϕ est intégrable sur $\mathbb{R}^d$ par rapport à la loi de X, c'est-à-dire

$$\mathrm{E}(|\phi(X)|) < \infty \rightleftharpoons \int_{\mathbb{R}^d} |\phi(t)| \mathrm{P}_X(dt) < \infty.$$

Si tel est le cas,

$$\mathrm{E}\left(\phi(X)\right) = \int_{\Omega} \phi(X(\omega)) \mathrm{P}(d\omega) = \int_{\mathbb{R}^d} \phi(t) \mathrm{P}_X(dt).$$

En particulier, lorsque X est réelle intégrable,

$$\mathrm{E}(X) = \int_{\mathbb{R}} t \mathrm{P}_X(dt).$$

Preuve : On reprend l'idée générale de construction de l'espérance. On suppose $d = 1$. Si $\phi = \mathbb{1}_B$ pour un borélien B de $\mathbb{R}$,

$$\int_{\mathbb{R}} \mathbb{1}_B(t) \mathrm{P}_X(dt) = \mathrm{P}_X(B) = \mathrm{P}(X \in B) = \mathrm{E}\left(\mathbb{1}_B \circ X\right)$$

et la formule est vraie dans ce cas. Si ϕ est une fonction étagée, la formule est valide par linéarité par rapport à P. Si ϕ est positive, soit $(\phi_n)_{n \in \mathbb{N}}$ une suite de fonctions positives étagées convergeant en croissant vers ϕ (Proposition 1.6). Alors $\phi_n \circ X$ est une suite de variables aléatoires étagées qui converge ponctuellement en croissant vers $\phi \circ X$. En utilisant le théorème de convergence monotone pour l'intégrale par rapport à P_X et pour l'espérance,

$$\int_{\mathbb{R}} \phi(t) \mathrm{P}_X(dt) = \lim_{n \to \infty} \int_{\mathbb{R}} \phi_n(t) \mathrm{P}_X(dt) = \lim_{n \to \infty} \mathrm{E}(\phi_n \circ X) = \mathrm{E}(\phi \circ X).$$

Dans le cas général, on voit que

$$\int_{\mathbb{R}} |\phi(t)| \mathrm{P}_X(dt) = \mathrm{E}(|\phi \circ X|)$$

et donc $\phi \circ X$ est intégrable par rapport à P si et seulement si ϕ est intégrable par rapport à P_X. En posant $\phi = \phi^+ - \phi^-$, on conclut que

$$\int_{\mathbb{R}} \phi(t) \mathrm{P}_X(dt) = \int_{\mathbb{R}} \phi^+(t) \mathrm{P}_X(dt) - \int_{\mathbb{R}} \phi^-(t) \mathrm{P}_X(dt) = \mathrm{E}\left(\phi^+ \circ X\right) - \mathrm{E}\left(\phi^- \circ X\right) = \mathrm{E}(\phi \circ X).$$

Le théorème est démontré. □

Dans la pratique, souvent, la loi de X est discrète ou à densité et le théorème de transport peut s'écrire comme dans les deux propositions suivantes :

Proposition 2.5 *Si X est une variable aléatoire discrète de loi donnée par :*

$$\mathrm{P}_X = \sum_{j \in J} p_j \delta_{x_j},$$

alors,

$$\mathrm{E}(|\phi(X)|) < \infty \rightleftharpoons \sum_{j \in J} |\phi(x_j)| p_j < \infty,$$

et, sous cette condition,

$$\mathrm{E}\left(\phi(X)\right) = \int_{\mathbb{R}^d} \phi(t) \mathrm{P}_X(dt) = \sum_{j \in J} \phi(x_j) p_j = \sum_{j \in J} \phi(x_j) \mathrm{P}(X = x_j).$$

Proposition 2.6 *Si X est une variable aléatoire à densité f_X par rapport à la mesure de Lebesgue sur $\mathbb{R}^d$:*

$$\mathrm{P}_X(B) = \int_B f_X(t)\, dt,\ B \in \mathcal{B}(\mathbb{R}^d),$$

alors,

$$\mathrm{E}(|\phi(X)|) < \infty \rightleftharpoons \int_{\mathbb{R}^d} |\phi(t)| f_X(t) dt < \infty,$$

et, sous cette condition,

$$\mathrm{E}\left(\phi(X)\right) = \int_{\mathbb{R}^d} \phi(t) \mathrm{P}_X(dt) = \int_{\mathbb{R}^d} \phi(t) f_X(t) dt.$$

On peut trouver la densité d'une variable aléatoire à l'aide du résultat suivant :

Proposition 2.7 *Supposons qu'il existe une fonction borélienne positive f telle que pour toute fonction ϕ borélienne positive (ou borélienne bornée) on ait*

$$\mathrm{E}\left(\phi(X)\right) = \int_{\mathbb{R}^d} \phi(t) f(t) dt.$$

Alors f est la densité de X par rapport à la mesure de Lebesgue.

Remarque : Soit X un vecteur aléatoire à valeurs dans $\mathbb{R}^d$ admettant une densité f_X. Soit g un difféomorphisme sur $\mathbb{R}^d$, de jacobien $\mathrm{Jac}\, g\,(x)$. Alors le vecteur $Y = g(X)$ a pour densité

$$f_Y(y) = (f_X \circ g^{-1})(y)\, |\mathrm{Jac}\, g^{-1}(y)|.$$

En effet, si ϕ est une fonction borélienne bornée (par exemple une indicatrice d'un borélien), par la Proposition 2.6 et la formule de changement de variables pour des intégrales de Lebesgue, on a

$$\mathrm{E}\left((\phi \circ g)(X)\right) = \int_{\mathbb{R}^d} (\phi \circ g)(t) f_X(t) dt = \int_{\mathbb{R}^d} \phi(y)(f_X \circ g^{-1})(y)\, |\mathrm{Jac}\, g^{-1}(y)| dy,$$

et il reste à appliquer la Proposition 2.7. $\square$

Exemples : i) Soit X de loi $\frac{1}{2}\delta_0 + \frac{1}{2}\delta_1$. Alors $\mathrm{E}(X) = \frac{1}{2}$. Dans un jeu de pile ou face, on obtient en moyenne une fois sur deux pile ($X = 1$) et une fois sur deux face ($X = 0$).
ii) Soient $x_1, \ldots, x_r$ des réels et soit $\mathrm{P}_X = \frac{1}{r}\sum_{j=1}^{r}\delta_{x_j}$. Alors $\mathrm{E}(X) = \frac{1}{r}\sum_{j=1}^{r} x_j$ est la moyenne des x_j.
iii) Soit $X \sim \mathcal{E}(\lambda)$ Alors,

$$\mathrm{E}(X) = \int_{\mathbb{R}} t f_X(t)dt = \int_{\mathbb{R}} t\lambda e^{-\lambda t}\mathbb{1}_{[0,\infty[}(t)dt = \int_0^{\infty} \lambda t e^{-\lambda t}dt = \frac{1}{\lambda}.$$

iv) (loi gaussienne) Si $X \sim \mathcal{N}(0,1)$ alors, par symétrie

$$\mathrm{E}(X) = \int_{\mathbb{R}} t f_X(t)dt = \int_{\mathbb{R}} t e^{-\frac{t^2}{2}}\frac{dt}{\sqrt{2\pi}} = 0.$$

v) Si X est une variable aléatoire réelle, intégrable, la linéarité de l'espérance implique

$$\mathrm{E}(aX + b) = a\mathrm{E}(X) + b, \ \forall a, b \in \mathbb{R}.$$

Donc si $Y \sim \mathcal{N}(m, \sigma^2)$ alors $\mathrm{E}(Y) = m$.

Il est parfois plus commode de calculer l'espérance à partir de la fonction de répartition :

Proposition 2.8 *Si X est une variable aléatoire réelle positive de fonction de répartition F_X, alors*

$$\mathrm{E}(X) = \int_0^{\infty} \mathrm{P}(X > t)dt = \int_0^{\infty} (1 - F_X(t))\, dt.$$

De plus, $\mathrm{E}(X) < \infty$ *si et seulement si, pour un ou tout* $\varepsilon > 0$,

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

Preuve : D'après le théorème de Fubini on peut écrire

$$\int_0^{\infty} \mathrm{P}(X > t)dt = \int_0^{\infty} \mathrm{E}\left(\mathbb{1}_{]t,\infty[}(X)\right) dt = \mathrm{E}\left(\int_0^X dt\right) = \mathrm{E}(X).$$

Pour la deuxième partie, on voit que

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

en divisant $[0, \infty[$ en intervalles $[n, n+1[$. De la même façon

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

On conclut en remplaçant X par $\frac{X}{\varepsilon}$. $\square$

Proposition 2.9 *(inégalité de Markov)*
Si X est intégrable et $t > 0$, alors

$$\mathrm{P}(X \geqslant t) \leqslant \frac{\mathrm{E}(X^+)}{t} \leqslant \frac{\mathrm{E}(|X|)}{t}.$$

Preuve : On observe que

$$\mathbb{1}_{[t,\infty[}(X) \leqslant \frac{X}{t}\mathbb{1}_{[t,\infty[}(X) \leqslant \frac{X^+}{t} \leqslant \frac{|X|}{t}$$

et on intégre cette inégalité par rapport à P. □

Remarque : Cette inégalité est utilisée généralement soit pour X positive, soit pour $|X|$. Elle n'est intéressante que si le second membre est plus petit que 1. □
Remarque : On peut reformuler le point *(iv)* de la Proposition 2.1 :
(iv') $\mathrm{E}(X) = 0 \rightleftharpoons X = 0$ *p.s.* □

2.3 Moments, variance et covariance

Pour définir les moments et la variance d'une variable aléatoire réelle X il suffit de considérer dans le théorème de transport $\phi(t) = t^k$, $k \in \mathbb{N}^*$ ou $\phi(t) = (t - \mathrm{E}(X))^2$.

Définition 2.6 *Soit X une variable aléatoire réelle. On dit que X* **admet un moment d'ordre** k, *où $k \in \mathbb{N}^*$, si $\mathrm{E}(|X|^k) < \infty$. Sous cette condition* **le moment d'ordre** *k de X est*

$$\mu_k(X) = \mathrm{E}(X^k),\ k \in \mathbb{N}^*.$$

En particulier, $\mu_1(X) = \mathrm{E}(X)$.

Proposition 2.10 *Supposons que la variable aléatoire réelle X admet un moment d'ordre k, $k \in \mathbb{N}^*$. Alors*

$$\mu_k(X) = \int_{\mathbb{R}} t^k \mathrm{P}_X(dt).$$

Si X est réelle discrète

$$\mu_k(X) = \sum_{j\in J} x_j^k p_j$$

et si X admet la densité f_X par rapport à la mesure de Lebesgue sur $\mathbb{R}$

$$\mu_k(X) = \int_{\mathbb{R}} t^k f_X(t)dt.$$

Lorsque X est positive ayant la fonction de répartition F_X, alors

$$\mu_k(X) = k\int_0^\infty t^{k-1}\mathrm{P}(X > t)dt = k\int_0^\infty t^{k-1}\left(1 - F_X(t)\right)dt.$$

Définition 2.7 *Soit X une variable aléatoire réelle dont le carré est intégrable (ayant un moment d'ordre 2). On appelle* **variance** *de X, ou de sa loi, et on note $\mathrm{Var}(X)$, la quantité*

$$\mathrm{Var}(X) = \mathrm{E}\left((X - \mathrm{E}(X))^2\right).$$

La racine $\sqrt{\mathrm{Var}(X)}$ est appelée **écart type**, *parfois noté $\sigma(X)$. Une variable aléatoire de variance 1 est dite* **réduite**.

Remarque : Une expression équivalente de la variance est

$$\mathrm{Var}(X) = \mathrm{E}(X^2) - \mathrm{E}(X)^2 = \mu_2(X) - \mu_1(X)^2,$$

car grâce à la linéarité de l'espérance

$$\mathrm{Var}(X) = \mathrm{E}\left(X^2 - 2X\mathrm{E}(X) + \mathrm{E}(X)^2\right) = \mathrm{E}(X^2) - 2\mathrm{E}(X)^2 + \mathrm{E}(X)^2.$$

□

Remarque : Si $\mathrm{Var}(X) = 0$ alors $X = \mathrm{E}(X)$ p.s. □

Proposition 2.11 *Supposons que la variable aléatoire réelle X a le carré intégrable. Alors*

$$\mathrm{Var}(aX + b) = a^2\mathrm{Var}(X),\ a, b \in \mathbb{R}.$$

Exemple : (loi gaussienne) Soit $X \sim \mathcal{N}(0, 1)$. Alors X admet des moments de tous ordres ; tous les moments impairs sont nuls et

$$\mu_{2k}(X) = \frac{(2k)!}{2^k k!},\ k \in \mathbb{N}^*$$

par intégration par parties. Par exemple, le moment d'ordre deux est

$$\mu_2(X) = 1,$$

et donc $\mathrm{Var}(X) = 1$ (d'ici la terminologie gaussienne reduite). Si $Y \sim \mathcal{N}(m, \sigma^2)$, alors $\mathrm{Var}(Y) = \sigma^2$ et $\mu_2(Y) = m^2 + \sigma^2$. Ainsi les paramètres de la loi $\mathcal{N}(m, \sigma^2)$ sont l'espérance et la variance de cette loi.

Proposition 2.12 *(inégalité de Tchebytchev)*
Si X est une variable aléatoire réelle dont le carré est intégrable, et $t > 0$, alors

$$\mathrm{P}(|X - \mathrm{E}(X)| \geqslant t) \leqslant \frac{\mathrm{Var}(X)}{t^2}.$$

Remarque : Si X admet un moment d'ordre k, $k \in \mathbb{N}^*$, et $t > 0$, alors,

$$\mathrm{P}(X \geqslant t) \leqslant \frac{\mathrm{E}(|X|^k)}{t^k}.$$

□

Preuve : Si X est une variable aléatoire réelle dont le carré est intégrable, alors par l'inégalité de Markov
$$\mathrm{P}(X \geqslant t) \leqslant \frac{\mathrm{E}(|X|^2)}{t^2},$$
pour tout $t > 0$, puisque $\{X \geqslant t\} \subset \{|X|^2 \geqslant t^2\}$. De la même façon on vérifie la remarque ci-dessus.

Il suffit maintenant d'appliquer cette inégalité à la variable $|X - \mathrm{E}(X)|$ pour obtenir l'inégalité de Tchebytchev. □

Exemple : (**fonction génératrice d'une v.a. à valeurs entières**)
La **fonction génératrice** d'une variable aléatoire à valeurs dans $\mathbb{N}$ est définie sur l'intervalle réel $[-1, 1]$ (soit sur le disque unité $\{|z| \leqslant q\}$ du plan complexe) par

$$G_X(s) = \mathrm{E}(s^X) = \sum_{n \geqslant 0} s^n p_n = \sum_{n \geqslant 0} s^n \mathrm{P}(X = n),\ |s| \leqslant 1.$$

La fonction génératrice est une fonction continue sur $\{|s| \leqslant 1\}$, indéfiniment différentiable sur $\{|s| < 1\}$ et elle détermine la loi qui a servi à la définir, puisque, pour tout $n \in \mathbb{N}$,

$$\left[\frac{d^n}{ds^n} G_X(s)\right]_{s=0} = n!\,\mathrm{P}(X = n).$$

Ainsi G_X caractérise la loi de X. La terminologie génératrice vient du fait qu'elle doit permettre de calculer les moments de X.

Ainsi, X est intégrable si et seulement si G_X est dérivable à gauche en $s = 1$ et dans ce cas (la dérivée à gauche ici) :
$$\mu_1(X) = \mathrm{E}(X) = G'_X(1).$$

De la même façon, X est de carré intégrable si et seulement si G_X est deux fois dérivable à gauche en $s = 1$ et dans ce cas :

$$\mathrm{E}[X(X-1)] = G''_X(1), \text{ d'où } \mu_2(X) = G'_X(1) + G''_X(1).$$

Plus généralement,

$$\mathrm{E}[X(X-1)\dots(X-k+1)] = \left[\frac{d^k}{ds^k} G_X(s)\right]_{s \uparrow 1}.$$

□

Définition 2.8 *Soient X et Y deux variables aléatoires réelles dont les carrés sont intégrables. On appelle* **la covariance** *de X et Y la quantité*

$$\mathrm{Cov}(X, Y) = \mathrm{E}\left((X - \mathrm{E}(X))(Y - \mathrm{E}(Y))\right) = \mathrm{E}(XY) - \mathrm{E}(X)\mathrm{E}(Y).$$

En particulier, si $X = Y$, $\mathrm{Cov}(X, X) = \mathrm{Var}(X)$.
On appelle **coefficient de corrélation linéaire** *la quantité*

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

lorsque les deux variables ne sont pas presque sûrement constantes.

Remarque : Comme $|xy| \leqslant \frac{1}{2}(x^2 + y^2)$, la variable XY est intégrable et la covariance est bien définie. □

Remarque : Par l'inégalité Cauchy-Schwarz (voir §2.4) on voit que

$$|\rho(X, Y)| \leqslant 1$$

avec égalité (respectivement 1 ou -1) si et seulement si Y est une fonction affine de X (respectivement $aX + b$ ou $-aX + b$ avec $a > 0$). □

Définition 2.9 *Soit X un vecteur aléatoire à valeurs dans $\mathbb{R}^d$. Sa* **matrice (carrée) de covariance** *est la matrice :*

$$K_X = (\mathrm{Cov}(X_j, X_k))_{1 \leqslant j,k \leqslant d}$$

Remarque : A la variance se substitue à présent une matrice. Il est facile de voir que :

$$K_X = \mathrm{E}(XX^*) - \mathrm{E}(X)\mathrm{E}(X)^*,$$

(ici, si $x \in \mathbb{R}^d$, xx^* est une matrice $d \times d$).

Cette matrice est symétrique (par la symétrie de la covariance) et semi-définie positive, puisque pour tous réels $\alpha_1, \ldots, \alpha_d$,

$$\sum_{1 \leqslant j,k \leqslant d} K_X(j,k)\alpha_j\alpha_k = \mathrm{E}\left[\left(\sum_{j=1}^{d} \alpha_j(X_j - \mathrm{E}(Xj))\right)^2\right] \geqslant 0.$$

Elle est définie positive si aucune combinaison linéaire des composantes X_j n'est p.s. constante. □

2.4 Espaces L^p

Nous avons défini la classe des variables aléatoires (réelles) intégrables sur un espace de probabilité $(\Omega, \mathcal{A}, \mathrm{P})$, notée dorénavant par $\mathcal{L}^1(\Omega, \mathcal{A}, \mathrm{P})$. Pour $0 < p < \infty$, on note $\mathcal{L}^p(\Omega, \mathcal{A}, \mathrm{P})$ (ou simplement $\mathcal{L}^p$ si le contexte est clair) l'ensemble des variables aléatoires X telles que $\mathrm{E}(|X|^p) < \infty$. $\mathcal{L}^0$ est simplement l'ensemble des variables aléatoires. Enfin, on définit $\mathcal{L}^\infty$ l'ensembles des variables aléatoires pour lesquelles il existe $c > 0$ avec

$$\mathrm{P}(\{\omega : |X(\omega)| > c\}) = 0.$$

C'est l'ensemble des variables aléatoires presque sûrement bornées.

Si $X \in \mathcal{L}^p$, $0 < p < \infty$, on pose

$$\|X\|_p := (\mathrm{E}(|X|^p))^{1/p}.$$

Pour $X \in \mathcal{L}^\infty$, on pose

$$\|X\|_\infty := \inf\{c > 0 : \mathrm{P}(|X| > c) = 0\}$$

(**supremum essentiel** de X).

Deux réels $p, q \geqslant 1$ sont **conjugués** si $\frac{1}{p} + \frac{1}{q} = 1$. On convient que 1 et ∞ sont conjugués.

Proposition 2.13 *(inégalité de Hölder)*
Soient p, q conjugués avec $1 \leqslant p \leqslant \infty$, $X \in \mathcal{L}^p$, $Y \in \mathcal{L}^q$. Alors $XY \in \mathcal{L}^1$ et

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

Preuve : Si $p = 1$ ou $p = \infty$ l'inégalité est évidente. Si $\|X\|_p\|Y\|_q = 0$, alors $XY = 0$ p.s. et l'inégalité de Hölder est triviale. Supposons donc $\|X\|_p\|Y\|_q \neq 0$. Par homogénéité, on peut supposer que $\|Y\|_q = 1$. Il suffit donc de prouver que

$$(\mathrm{E}(|XY|))^p \leqslant \mathrm{E}\left(|X|^p\right).$$

On considère la mesure de probabilité Q ayant la densité $|Y|^q$ par rapport à la mesure de probabilité P,

$$\mathrm{Q}(A) = \int_A |Y(\omega)|^q \mathrm{P}(d\omega) = \mathrm{E}\left(\mathbb{1}_A |Y|^q\right), \; A \in \mathcal{A}.$$

L'inégalité à prouver devient

$$\left(\int_\Omega |X(\omega)|\,|Y(\omega)|^{1-q} \mathrm{Q}(d\omega)\right)^p \leqslant \int_\Omega |X(\omega)|^p |Y(\omega)|^{-q} \mathrm{Q}(d\omega)$$

ou encore, avec une notation évidente

$$\left(\mathrm{E}^{\mathrm{Q}}(|X|\,|Y|^{1-q})\right)^p \leqslant \mathrm{E}^{\mathrm{Q}}\left(|X|^p |Y|^{-q}\right)$$

qui est une conséquence de l'inégalité de Jensen pour la fonction convexe $\varphi(x) = x^p$, car $(1-q)p = -q$.
Autre preuve. On suppose encore une fois que $\|X\|_p\|Y\|_q \neq 0$ et soient $a, b > 0$. Alors il existent $s, t \in \mathbb{R}$ tels que $a = e^{s/p}$, $b = e^{t/q}$. D'après la convexité de l'exponentielle

$$\exp(\frac{s}{p} + \frac{t}{q}) \leqslant \frac{e^s}{p} + \frac{e^t}{q}, \qquad \text{d'où} \quad ab \leqslant \frac{a^p}{p} + \frac{b^q}{q}.$$

Prendre ensuite dans cette dernière inégalité $a = {}^{|X|}/_{\|X\|_p}$ et $b = {}^{|Y|}/_{\|Y\|_q}$. On trouve

$$\frac{|XY|}{\|X\|_p\|Y\|_q} \leqslant \frac{1}{p}\frac{|X|^p}{\mathrm{E}(|X|^p)} + \frac{1}{q}\frac{|Y|^q}{\mathrm{E}(|Y|^q)}.$$

Il suffit ensuite de prendre l'espérance dans cette inégalité. □

Remarque : Pour $p = q = 2$ on obtient **l'inégalité de Cauchy-Schwarz** :

$$\|XY\|_1 \leqslant \|X\|_2\|Y\|_2 \rightleftharpoons \mathrm{E}(|XY|) \leqslant \sqrt{\mathrm{E}(|X|^2)}\sqrt{\mathrm{E}(|Y|^2)}.$$

□

Remarque : Les variables aléatoires p.s. constantes appartiennent à tous les espace $\mathcal{L}^p$, $1 \leqslant p \leqslant \infty$ et (par l'inégalité de Jensen ou de Hölder),

$$\mathcal{L}^{p'}(\Omega, \mathcal{A}, \mathrm{P}) \subset \mathcal{L}^p(\Omega, \mathcal{A}, \mathrm{P}), \;\text{ lorsque }\; 1 \leqslant p \leqslant p' \leqslant \infty.$$

Autrement dit, l'application $p \mapsto \|\cdot\|_p$ est croissante. En effet, soit $p < p'$ et on applique l'inégalité de Hölder à $|X|^p$ et 1 avec les exposants $\frac{p'}{p}$ et $\frac{p'}{p'-p}$:

$$\|X\|_p^p = \mathrm{E}(|X|^p \cdot 1) \leqslant \left(\mathrm{E}(|X|^{p\frac{p'}{p}})\right)^{\frac{p}{p'}} = \|X\|_{p'}^p.$$

De plus

$$\lim_{p\to\infty} \|X\|_p = \|X\|_\infty.$$

En effet, on voit d'abord que $|X| \leqslant \|X\|_\infty$, d'où $\|X\|_p \leqslant \|X\|_\infty$. D'apr'es la croissance de $p \mapsto \|\cdot\|_p$ on déduit que $\lim_{p\to\infty} \|X\|_p \leqslant \|X\|_\infty$.

Soit ensuite a arbitraire tel que $a < \|X\|_\infty$. Alors l'événement $A := \{|X| > a\}$ est de probabilité strictement positive. En particulier $\mathrm{P}(A)^{\frac{1}{p}} \to 1$, lorsque $p \to \infty$. On a $|X|^p \geqslant a^p \mathbb{1}_A$, d'où $\|X\|_p \geqslant a\mathrm{P}(A)^{\frac{1}{p}}$, donc $\lim_{p\to\infty} \|X\|_p \geqslant a$, pour tout $a < \|X\|_\infty$. Comme a était arbitraire, on obtient $\lim_{p\to\infty} \|X\|_p \geqslant \|X\|_\infty$. □

Proposition 2.14 *(inégalité de Minkowski)*
Soit $p \geqslant 1$. Si $X, Y \in \mathcal{L}^p$, alors $X + Y \in \mathcal{L}^p$ et

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

Preuve : On observe que p et $\frac{p}{p-1}$ sont conjugués. En utilisant l'inégalité triangulaire et l'inégalité de Hölder,

$$\|X + Y\|_p^p = \mathrm{E}\left(|X + Y|^p\right) \leqslant \mathrm{E}\left(|X|\,|X + Y|^{p-1} + |Y|\,|X + Y|^{p-1}\right)$$

$$\leqslant (\|X\|_p + \|Y\|_p)\left\||X + Y|^{p-1}\right\|_{\frac{p}{p-1}} = (\|X\|_p + \|Y\|_p)\,\|X + Y\|_p^{p-1}.$$

C'est le résultat si $\|X + Y\|_p \neq 0$. L'inégalité est triviale si $\|X + Y\|_p = 0$. □

Remarque : De l'inégalité de Minkowski on déduit que $\|\cdot\|_p$ est une semi-norme (pas norme) sur $\mathcal{L}^p$ (en effet $\|X\|_p = 0$ n'implique pas $X = 0$, mais seulement $X = 0$ p.s.). □

On introduit sur $\mathcal{L}^0$ la relation suivante :

$$X \equiv Y \rightleftharpoons X = Y \text{ p.s.} \rightleftharpoons \{X \neq Y\} \text{ est négligeable.}$$

On peut montrer que cette relation est une relation d'équivalence.

On notera $\mathrm{L}^p = \mathrm{L}^p(\Omega, \mathcal{A}, \mathrm{P})$ l'espace quotient de $\mathcal{L}^p$ par cette relation d'équivalence. Ainsi un élément X de L^p s'identifie à un représentant de la classe de tous les éléments $Y \in \mathcal{L}^p$ tels que $X = Y$ p.s.

Proposition 2.15 *Pour tout $p \geqslant 1$, l'espace L^p est un espace vectoriel normé complet (de Banach).*

Nous démontrerons ce résultat au §4.3.

REMARQUE : *Pour p et q conjugués et $1 \leqslant p < \infty$ le dual de L^p est L^q. Autrement dit, les formes linéaires continues sur L^p sont les variables aléatoires de la forme $\mathrm{L}^p \ni X \mapsto \mathrm{E}(XY) \in \mathbb{R}$, pour $Y \in \mathrm{L}^q$. De plus $\|X\|_p = \sup\{\mathrm{E}(XY) : \|Y\|_q \leqslant 1\}$.* On prendra garde au fait que L^1 n'est pas en général le dual de L^∞. □

On pourra utiliser des arguments géométriques dans les espaces L^2 par le suivant :

Proposition 2.16 *L'espace L^2 est un espace de Hilbert pour le produit scalaire $< X, Y >= \mathrm{E}(XY)$.*

2.5 Fonctions caractéristiques

On a déjà vu que la fonction de répartition d'une variable aléatoire X caractérise sa loi. Autrement dit, sur $\mathbb{R}$ par exemple, la donnée de

$$F_X(t) = \mathrm{E}\left(\mathbb{1}_{]-\infty,t]}(X)\right),\ t \in \mathbb{R}$$

détermine la loi de X. Mais comme les indicatrices sont des fonctions boréliennes bornées, la donnée de $\mathrm{E}(\phi(X))$ pour toute fonction borélienne bornée ϕ caractérise la loi P_X. Par exemple $\mathbb{1}_{]-\infty,t]}$ peut être approchée par

$$\phi_n(x) = \begin{cases} 1 \text{ si } x \leqslant t \\ 1 + n(t-x) \text{ si } t \leqslant x \leqslant t + \frac{1}{n} \\ 0 \text{ si } x > t + \frac{1}{n}. \end{cases}$$

Il s'ensuit (par convergence dominée) que la donnée de $\mathrm{E}(\phi(X))$ pour toute fonction continue bornée ϕ sur $\mathbb{R}$ caractérise P_X. On peut approcher les fonctions indicatrices par des fonctions C^∞ bornées ; donc la donnée de $\mathrm{E}(\phi(X))$ pour toute fonction infiniment dérivable bornée caractérise P_X. Dans ce paragraphe on prendra la famille des fonctions sinus et cosinus.

Définition 2.10 *Soit X un vecteur aléatoire sur $(\Omega, \mathcal{A}, \mathrm{P})$ à valeurs dans $\mathbb{R}^d$. On appelle* **fonction caractéristique** *de X ou* **transformée de Fourier** *de sa loi, et on note φ_X, la fonction à valeurs complexes*

$$\begin{aligned} \mathbb{R}^d \ni t \mapsto \varphi_X(t) = \mathrm{E}\left(e^{i<t,X>}\right) &= \int_{\mathbb{R}^d} e^{i<t,x>} \mathrm{P}_X(dx) \\ &= \mathrm{E}\left(\cos < t, X >\right) + i\mathrm{E}\left(\sin < t, X >\right). \end{aligned}$$

REMARQUE : La fonction caractéristique est bien définie car $x \mapsto e^{i<t,x>}$ est bornée donc intégrable par rapport à P_X. □

REMARQUE : Si X est à densité f par rapport à la mesure de Lebesgue sur $\mathbb{R}^d$, alors

$$\varphi_X(t) = \int_{\mathbb{R}^d} e^{i<t,x>} f(x) dx$$

est aussi appelée la transformée de Fourier de f. □

Exemple : (loi gaussienne)
Soit $X \sim \mathcal{N}(0,1)$. Alors

$$\varphi_X(t) = \mathrm{E}\left(e^{itX}\right) = \int_{\mathbb{R}} \frac{e^{itx-x^2/2}}{\sqrt{2\pi}} dx = e^{-t^2/2}.$$

En effet, soit $z = a + ib \in \mathbb{C}$. Comme $|xe^{zx-x^2/2}| \leqslant |x|e^{ax-x^2/2}$ avec $\int_{\mathbb{R}} |x|e^{ax-x^2/2}dx < \infty$, la fonction complexe

$$\mathbb{C} \ni z \mapsto \int_{\mathbb{R}} \frac{e^{zx-x^2/2}}{\sqrt{2\pi}} dx \in \mathbb{C}$$

est holomorphe (on peut dériver sous le signe intégrale). En calculant cette intégrale pour z réel, par exemple $z = a$ on trouve

$$\int_{\mathbb{R}} \frac{e^{ax-x^2/2}}{\sqrt{2\pi}} dx = e^{a^2/2}.$$

Comme la fonction complexe

$$\mathbb{C} \ni z \mapsto e^{z^2/2} \in \mathbb{C}$$

est aussi holomorphe et coïncidant sur $\mathbb{R}$ à celle introduite ci-dessus, elles coïncident sur tout $\mathbb{C}$:

$$\int_{\mathbb{R}} \frac{e^{zx-x^2/2}}{\sqrt{2\pi}} dx = e^{z^2/2}, \forall z \in \mathbb{C}.$$

En particulier, pour $z = it$ on trouve la fonction caractéristique de la loi gaussienne standard. Si $Y \sim \mathcal{N}(m, \sigma^2)$, alors $Y = \sigma X + m$, où $X \sim \mathcal{N}(0,1)$, et on trouve

$$\varphi_Y(t) = \mathrm{E}\left(e^{itY}\right) = e^{itm}\mathrm{E}\left(e^{it\sigma X}\right) = e^{itm-\frac{t^2\sigma^2}{2}}.$$

Théorème 2.5 *La fonction caractéristique caractérise la loi, c'est-à-dire,* $\varphi_X = \varphi_Y$ *si et seulement si* $\mathrm{P}_X = \mathrm{P}_Y$.

Preuve : Supposons, pour simplifier, que X et Y sont deux variables aléatoires réelles. On va noter, pour $\sigma > 0$,

$$f_\sigma(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-x^2/2\sigma^2}, \quad \hat{f}_\sigma(t) = e^{-\frac{t^2\sigma^2}{2}}.$$

D'après l'exemple précédent

$$\int_{\mathbb{R}} f_\sigma(x)e^{itx}dx = \hat{f}_\sigma(t),$$

donc

$$f_\sigma(t-s) = \frac{1}{\sqrt{2\pi\sigma^2}} \hat{f}_\sigma\left(\frac{t-s}{\sigma^2}\right) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{\mathbb{R}} f_\sigma(x)e^{i(t-s)x/\sigma^2}dx.$$

Supposons maintenant que $\varphi_X = \varphi_Y$. En admettant qu'on puisse échanger l'intégration en dx et l'espérance, on peut écrire

$$\mathrm{E}\left[f_\sigma(X-s)\right] = \mathrm{E}\left[\frac{1}{\sqrt{2\pi\sigma^2}} \int_{\mathbb{R}} f_\sigma(x)e^{i(X-s)x/\sigma^2}dx\right] = \int_{\mathbb{R}} f_\sigma(x)\frac{1}{\sqrt{2\pi\sigma^2}}\varphi_X\left(\frac{x}{\sigma^2}\right)e^{-isx/\sigma^2}dx.$$

La même égalité est vraie pour Y. Donc

$$\mathrm{E}\left[f(X)\right] = \mathrm{E}\left[f(Y)\right],$$

pour toute fonction de la forme $f(t) = f_\sigma(t-s)$ pour $s \in \mathbb{R}$ et $\sigma > 0$ arbitraires. Par conséquent aussi pour toute fonction f dans l'espace vectoriel engendré par ces fonctions. D'après le théorème Stone-Weierstrass, cet espace est dense dans l'espace C_0 des fonctions continues sur $\mathbb{R}$ et ayant une limite nulle à l'infini, pour la topologie de la convergence uniforme. Ainsi

$$\mathrm{E}\left[f(X)\right] = \mathrm{E}\left[f(Y)\right], \forall f \in \mathrm{C}_0.$$

Mais si O est un ouvert, $\mathbb{1}_O$ est limite croissante de fonctions de C_0. On en déduit que

$$\mathrm{P}_X(O) = \mathrm{E}\left[\mathbb{1}_O(X)\right] = \mathrm{E}\left[\mathbb{1}_O(Y)\right] = \mathrm{P}_Y(O), \forall O \text{ ouvert}$$

ce qui implique $\mathrm{P}_X = \mathrm{P}_Y$. □

Une autre preuve : On suppose toujours que X et Y sont deux variables aléatoires réelles. Par calcul direct (voir aussi la loi de Laplace)

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

On peut écrire

$$\mathrm{E}\left[\frac{2\lambda}{\lambda^2 + (X-s)^2}\right] = \int_{\mathbb{R}} \frac{2\lambda}{\lambda^2 + (y-s)^2} \mathrm{P}_X(dy)$$

$$= \int_{\mathbb{R}} \mathrm{P}_X(dy) \int_{\mathbb{R}} dx\, e^{i(y-s)x} e^{-\lambda|x|} = \int_{\mathbb{R}} e^{-isx} e^{-\lambda|x|} \varphi_X(x) dx$$

où dans la dernière égalité on a utilisé le théorème de Fubini. Soit f une fonction continue à support compact. En appliquant à nouveau le théorème de Fubini et un changement de variable, on en déduit

$$\int_{\mathbb{R}} ds\, f(s) \int_{\mathbb{R}} \mathrm{P}_X(dy) \frac{2\lambda}{\lambda^2 + (y-s)^2} = \int_{\mathbb{R}} \mathrm{P}_X(dy) \int_{\mathbb{R}} du\, f(y - \lambda u) \frac{2}{1+u^2}$$

Pour $\lambda \downarrow 0$, l'intégrant $\int_{\mathbb{R}} f(y - \lambda u) \frac{2}{1+u^2} du$ converge vers $2\pi f(y)$ en restant dominé par $2\pi \|f\|_\infty$. Ainsi, par convergence dominée

$$\mathrm{E}\left[f(X)\right] = \int_{\mathbb{R}} f(y) \mathrm{P}_X(dy) = \lim_{\lambda \downarrow 0} \int_{\mathbb{R}} ds\, f(s) \int_{\mathbb{R}} e^{-isx} e^{-\lambda|x|} \varphi_X(x) \frac{dx}{2\pi}.$$

Le terme de droite est une fonction de f et de φ_X, et donc φ_X caractérise bien la loi P_X. □

Exemples : i) Si $X = a$ p.s., $a \in \mathbb{R}^d$, alors $\mathrm{P}_X = \delta_a$, donc $\varphi_X(t) = e^{i<t,a>}$.
ii) Si $X \sim \gamma(p, \lambda)$ alors $\varphi_X(t) = \left(\frac{\lambda}{\lambda - it}\right)^p$.

Proposition 2.17 *Soit X un vecteur aléatoire à valeurs dans $\mathbb{R}^d$. Alors :*

i) $|\varphi_X(t)| \leqslant 1$.

ii) $\varphi_X(-t) = \overline{\varphi_X(t)}$.

iii) $\varphi_X(0) = 1$.

iv) φ_X *est uniformément continue.*

v) φ_X *est "de type positif" :*

$$\forall n \in \mathbb{N}^*, \, \forall t_1, \ldots, t_n \in \mathbb{R}^d, \, \forall z_1, \ldots, z_n \in \mathbb{C}\,, \; \sum_{j,k=1}^{n} \varphi_X(t_k - t_j) z_k \bar{z}_j \geqslant 0.$$

Preuve : On a $|\varphi_X(t)| \leqslant \mathrm{E}\left(|e^{i<t,X>}|\right) = \mathrm{E}(1) = 1$ et $\varphi_X(0) = \mathrm{E}(1) = 1$.
Pour prouver (ii) on écrit

$$\bar{\varphi}_X(t) = \overline{\mathrm{E}\left(e^{i<t,X>}\right)} = \mathrm{E}\left(\overline{e^{i<t,X>}}\right) = \mathrm{E}\left(e^{-i<t,X>}\right) = \mathrm{E}\left(e^{i<-t,X>}\right) = \varphi_X(-t).$$

Pour montrer (iv) on calcule, pour t, $h \in \mathbb{R}^d$

$$\varphi_X(t+h) - \varphi_X(t) = \mathrm{E}\left(e^{i<t+h,X>} - e^{i<t,X>}\right)$$

$$= \mathrm{E}\left[e^{i<t,X>} e^{\frac{i}{2}<h,X>} \left(e^{\frac{i}{2}<h,X>} - e^{-\frac{i}{2}<h,X>}\right)\right] = \mathrm{E}\left[e^{i<t,X>} e^{\frac{i}{2}<h,X>} 2i \sin \frac{<h,X>}{2}\right],$$

d'où

$$|\varphi_X(t+h) - \varphi_X(t)| \leqslant \mathrm{E}\left[2|\sin \frac{<h,X>}{2}|\right] \leqslant \mathrm{E}\left[(|h|\,|X|) \wedge 2\right].$$

Le membre de droite de cette dernière inégalité tend vers zéro quand $|h|$ tend vers zéro indépendament de t.
Enfin, pour (v),

$$\sum_{j,k=1}^{n} \varphi_X(t_k - t_j) z_k \bar{z}_j = \sum_{j,k=1}^{n} \mathrm{E}\left(e^{i<t_k - t_j, X>}\right) z_k \bar{z}_j = \mathrm{E}\left[\left(\sum_{k=1}^{n} e^{i<t_k,X>} z_k\right)\left(\sum_{j=1}^{n} e^{i<-t_j,X>} \bar{z}_j\right)\right]$$

$$= \mathrm{E}\left[\left(\sum_{k=1}^{n} e^{i<t_k,X>} z_k\right)\overline{\left(\sum_{j=1}^{n} e^{i<t_j,X>} z_j\right)}\right] = \mathrm{E}\left(\left|\sum_{k=1}^{n} e^{i<t_k,X>} z_k\right|^2\right) \geqslant 0.$$

□

Remarque : L'intérêt de la proposition précédente réside surtout dans le fait qu'elle admet la réciproque suivante (théorème de Bochner) : *une fonction φ vérifiant les conditions iii)-v) est la fonction caractéristique d'un vecteur aléatoire.* □

Remarque : Il existe plusieurs formules dites d'inversion permettant d'obtenir effectivement la loi à partir de la fonction caractéristique. En voici une possible

Proposition 2.18 *(théorème d'inversion de Fourier)*
Soit φ une fonction caractéristique intégrable par rapport à la mesure de Lebesgue sur $\mathbb{R}^d$. Alors φ est la fonction caractéristique d'un vecteur aléatoire X admettant la densité continue bornée donnée par :

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

Preuve : Nous allons démontrer cette proposition à l'aide des idées de la deuxième preuve du Théorème 2.5. Pour simplifier nous allons supposer encore une fois que $d = 1$. Puisque $\varphi = \varphi_X$ est intégrable, on peut écrire

$$2\pi \mathrm{E}[g(X)] = \lim_{\lambda \downarrow 0} \int_{\mathbb{R}} \mathrm{P}_X(dy) \int_{\mathbb{R}} g(y - \lambda u) \frac{2}{1+u^2} du = \lim_{\lambda \downarrow 0} \int_{\mathbb{R}} g(s) ds \int_{\mathbb{R}} \frac{2\lambda}{\lambda^2 + (y-s)^2} \mathrm{P}_X(dy)$$

$$= \lim_{\lambda \downarrow 0} \int_{\mathbb{R}} g(s) ds \int_{\mathbb{R}} e^{-isx} e^{-\lambda |x|} \varphi(x) dx = \int_{\mathbb{R}} g(s) ds \int_{\mathbb{R}} e^{-isx} \varphi(x) dx$$

Il reste à utiliser la Proposition 2.7 pour conclure. □

Proposition 2.19 *Soit X une variable aléatoire réelle de fonction caractéristique φ_X.*
i) Si X admet un moment d'ordre $k \in \mathbb{N}^*$, alors, φ_X est k fois dérivable, de dérivée ℓ-ième ($\ell \leqslant k$)

$$\varphi_X^{(\ell)}(t) = i^\ell \mathrm{E}\left(X^\ell e^{itX}\right).$$

En particulier, $\varphi_X^{(\ell)}(0) = i^\ell \mathrm{E}(X^\ell)$.
ii) Si φ_X est k fois dérivable en 0, $k \in \mathbb{N}^*$, alors X admet des moments d'ordres plus petits ou égaux à $2\ell \leqslant k$.

Preuve : i) On peut voir que, pour tout $t \in \mathbb{R}$,

$$\left| e^{iu} - 1 - \frac{iu}{1!} - \ldots - \frac{(iu)^{n-1}}{(n-1)!} \right| \leqslant \frac{|u|^n}{n!}.$$

En effet, $f_1(u) = i \int_0^u e^{ix} dx$ est de module plus petit que $|u|$ et par récurrence, $f_n(u) = i \int_0^u f_{n-1}(x) dx$ est de module plus petit que $|u|^n/(n!)$. On suppose que $\mathrm{E}(|X|) < \infty$ et on prouve que φ_X est dérivable en tout point $t \in \mathbb{R}$. Pour tout $h \neq 0$

$$\frac{\varphi_X(t+h) - \varphi_X(t)}{h} = \int_{\mathbb{R}} e^{itx} \frac{e^{ihx} - 1}{h} \mathrm{P}_X(dx).$$

D'après l'inégalité ci-dessus pour $n = 1$,

$$\left| e^{itx} \frac{e^{ihx} - 1}{h} \right| \leqslant |x|$$

fonction qui est intégrable par rapport à P_X indépendament de h. Par le théorème de convergence dominée

$$\varphi_X'(t) = \lim_{h \to 0} \int_{\mathbb{R}} e^{itx} \frac{e^{ihx} - 1}{h} \mathrm{P}_X(dx) = \int_{\mathbb{R}} x e^{itx} \mathrm{P}_X(dx) = \mathrm{E}\left(X e^{itX}\right).$$

Les dérivées d'ordres supérieurs se calculent de la même façon.
ii) D'après le fait que $p \mapsto \| \cdot \|_p$ est croissante, il suffit de prouver l'existence des moments d'ordre pair inférieur à 2ℓ pour assurer celle des moments d'ordre inférieur à 2ℓ. Supposons montrée l'existence du moment d'ordre $2r$, ce qui est vrai pour $r = 0$. D'après la première partie de la proposition,

$$\varphi_X^{(2r)}(0) = (\operatorname{Re} \varphi_X)^{(2r)}(0) = (-1)^r \mathrm{E}(X^{2r}),$$

puisque la partie imaginaire $\operatorname{Im}\varphi_X$ est impaire. On en déduit

$$(-1)^{r+1}\varphi_X^{(2r+2)}(0) = \lim_{h\to 0}\frac{(\operatorname{Re}\varphi_X)^{(2r)}(h) - (\operatorname{Re}\varphi_X)^{(2r)}(0)}{h} = \lim_{h\to 0}\int_{\mathbb{R}} x^{2r}\frac{1-\cos hx}{2h^2}\mathrm{P}_X(dx).$$

En effet, par calcul direct on voit que, par exemple pour $k = 1$

$$\varphi_X''(0) = \lim_{h\to 0}\frac{1}{2}\left[\frac{\varphi_X'(2h) - \varphi_X'(0)}{2h} + \frac{\varphi_X'(0) - \varphi_X'(-2h)}{2h}\right]$$

$$= \lim_{h\to 0}\frac{2\varphi_X'(2h) - 2\varphi_X'(-2h)}{8h} = \lim_{h\to 0}\frac{1}{4h^2}[\varphi_X(2h) - 2\varphi_X(0) + \varphi_X(-2h)]$$

$$= \lim_{h\to 0}\int_{\mathbb{R}}\left(\frac{e^{ihx} - e^{-ihx}}{2h}\right)^2\mathrm{P}_X(dx) = -\lim_{h\to 0}\int_{\mathbb{R}} x^2\left(\frac{\sin hx}{hx}\right)^2\mathrm{P}_X(dx).$$

Par le lemme de Fatou on conclut que

$$\int_{\mathbb{R}} x^{2r+2}\mathrm{P}_X(dx) \leqslant (-1)^{r+1}\varphi_X^{(2r+2)}(0) < \infty.$$

□

Remarque : En général, une loi n'est pas caractérisée par ses moments. □

Définition 2.11 *Soit X un vecteur aléatoire sur $(\Omega, \mathcal{A}, \mathrm{P})$ à valeurs dans $\mathbb{R}^d$. On appelle* **fonction génératrice des moments** *de X ou* **transformée de Laplace** *de sa loi, et on note*

$$\ell_X(s) = \mathrm{E}\left(e^{<s,X>}\right),$$

la fonction définie pour les valeurs de s pour lesquelles $\mathrm{e}^{<s,X>}$ est intégrable.

Remarque : La transformée de Laplace, si elle est finie sur un voisinage de 0, caractérise la loi, comme la transformée de Fourier. □

Proposition 2.20 *Soit X une variable aléatoire réelle telle que e^{sX} est intégrable pour s dans un intervalle contenant 0. Alors la transformée de Laplace ℓ_X est définie sur un intervalle contenant 0. De plus elle est analytique sur un voisinage de 0 et*

$$\ell_X(s) = \sum_{n\geqslant 0}\frac{s^n}{n!}\mathrm{E}(X^n),$$

pour tout s dans ce voisinage. En particulier,

$$\mu_n(X) = \mathrm{E}(X^n) = (\ell_X)^{(n)}(0).$$

2.6 Lois de probabilités usuelles

1. Loi de Bernoulli.
Si $X \sim \mathcal{B}(1,p)$, alors
$$\mathrm{E}(X) = p,\ \mathrm{Var}(X) = p(1-p)$$
et
$$\varphi_X(t) = 1 - p + pe^{it}$$

2. Loi binomiale.
Si $X \sim \mathcal{B}(n,p)$, alors
$$\mathrm{E}(X) = np,\ \mathrm{Var}(X) = np(1-p)$$
et
$$\varphi_X(t) = (1 - p + pe^{it})^n$$

3. Loi uniforme.
Si $X \sim \mathcal{U}(1,\ldots,r)$, alors
$$\mathrm{E}(X) = \frac{r+1}{2},\ \mathrm{Var}(X) = \frac{(r+1)(2r+1)}{6}$$
et
$$\varphi_X(t) = \frac{1}{r}e^{it}\frac{1-e^{irt}}{1-e^{it}}$$

4. Loi géométrique.
Si $X \sim \mathcal{G}(p)$, alors
$$\mathrm{E}(X) = \frac{1}{p},\ \mathrm{Var}(X) = \frac{1-p}{p^2}$$
et
$$\varphi_X(t) = \frac{pe^{it}}{1-(1-p)e^{it}}$$

5. Loi de Poisson.
Si $X \sim \mathcal{P}(\lambda)$, alors
$$\mathrm{E}(X) = \lambda,\ \mathrm{Var}(X) = \lambda$$
et
$$\varphi_X(t) = \exp\big(\lambda(e^{it}-1)\big)$$

6. Loi hypergéométrique.
Si $X \sim \mathcal{H}(N,n,p)$, alors
$$\mathrm{E}(X) = np,\ \mathrm{Var}(X) = \frac{N-n}{N-1}np(1-p).$$

7. Loi multinomiale.
Si $(X_1,\ldots,X_d) \sim \mathcal{M}(n,p_1,\ldots,p_n)$, $n \in \mathbb{N}$, alors
$$\mathrm{E}(X_1,\ldots,X_d) = (np_1,\ldots,np_d),$$

$$\mathrm{Var}(X_j) = np_j(1-p_j),\ \mathrm{Cov}(X_j, X_k) = -np_jp_k$$

et

$$\varphi_X(t_1,\dots,t_d) = \left(\sum_{j=1}^{d} p_j e^{it_jp_j}\right)^n.$$

8. Loi uniforme.
Si $X \sim \mathcal{U}_{[a,b]}$, alors

$$\mathrm{E}(X) = \frac{a+b}{2},\ \mathrm{Var}(X) = \frac{(b-a)^2}{12}$$

et

$$\varphi_X(t) = \frac{e^{itb} - e^{ita}}{it(b-a)}$$

9. Loi exponentielle.
Si $X \sim \mathcal{E}(\lambda)$, alors

$$\mathrm{E}(X) = \frac{1}{\lambda},\ \mathrm{Var}(X) = \frac{1}{\lambda^2}$$

et

$$\varphi_X(t) = \frac{\lambda}{\lambda - it}$$

10. Loi gamma.
Si $X \sim \gamma(p, \lambda)$, alors

$$\mathrm{E}(X) = \frac{p}{\lambda},\ \mathrm{Var}(X) = \frac{p}{\lambda^2}$$

et

$$\varphi_X(t) = \left(\frac{\lambda}{\lambda - it}\right)^p$$

10. Loi de chi-deux.
Si $X \sim \chi^2(n)$, alors

$$\mathrm{E}(X) = n,\ \mathrm{Var}(X) = 2n$$

et

$$\varphi_X(t) = \frac{1}{(1-2it)^{n/2}}$$

11. Loi beta.
Si $X \sim \beta(p, q)$, alors

$$\mathrm{E}(X) = \frac{B(p+1,q)}{B(p,q)},\ \mathrm{Var}(X) = \frac{pq}{(p+q)^2(p+q+1)}.$$

11. Loi de Laplace.
Si X suit une loi de Laplace, alors

$$\mathrm{E}(X) = 0,\ \mathrm{Var}(X) = 1$$

et

$$\varphi_X(t) = \frac{1}{1+t^2}$$

12. Loi de Cauchy.
Si X suit une loi de Cauchy, alors

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

13. Loi de Gauss (ou normale).
Si $X \sim \mathcal{N}(m, \sigma^2)$, alors

$$\mathrm{E}(X) = m,\ \mathrm{Var}(X) = \sigma^2$$

et

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

13. Loi de Gauss (ou normale) multidimensionnelle.
Si $(X_1, \ldots, X_d) \sim \mathcal{N}(m, \Sigma$, alors

$$\mathrm{E}(X_1, \ldots, X_d) = m,\ K_X = \Sigma$$

et

$$\varphi_X(t_1, \ldots, t_d) = e^{i<m,t> - \frac{1}{2} t^* \Sigma t}.$$

Chapitre 3

Indépendance

3.1 Indépendance

Définition 3.1 *Sur un espace de probabilité* $(\Omega, \mathcal{A}, \mathrm{P})$*, deux événements* A, B *sont dits* **indépendants** *si*

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

Exemple. On jette deux dés, un rouge et un noir. Les événements

$$A = \{ \text{ le dé rouge montre un nombre strictement supérieur à } 4 \}$$

et

$$B = \{ \text{ le dé noir montre un } 4\}.$$

sont indépendants, car les deux dé le sont. Pour modéliser on considère

$$(\Omega, \mathcal{A}, \mathrm{P}) = (\{(i,j) : i,j = 1,2,3,4,5,6\}, \mathcal{P}(\Omega), \mathrm{P}), \ \mathrm{P} \ \text{ la probabilité uniforme.}$$

Alors,

$$A = \{(5,1), \ldots, (5,6), (6,1), \ldots, (6,6)\}, \ \mathrm{P}(A) = \frac{1}{3}$$

$$B = \{(1,4), \ldots, (6,4)\}, \ \mathrm{P}(B) = \frac{1}{6}$$

donc

$$A \cap B = \{(5,4), (6,4)\}, \ \mathrm{P}(A \cap B) = \frac{1}{18} = \mathrm{P}(A)\mathrm{P}(B).$$

□

Remarque. Si A, B sont deux événements indépendants, alors A, B^c, A^c, B et A^c, B^c sont des paires d'événements indépendants. Ainsi on peut dire que les tribus $\sigma(\{A\})$ et $\sigma(\{B\})$ sont indépendantes. □

Définition 3.2 *Soit* $(\Omega, \mathcal{A}, \mathrm{P})$ *un espace de probabilité.*

a) Une famille quelconque d'événements $A_i \in \mathcal{A}$, $i \in I$, est **mutuellement indépendante** *si pour tout $J \subset I$ fini*

$$\mathrm{P}\left(\bigcap_{j\in J} A_j\right) = \prod_{j\in J} \mathrm{P}(A_j).$$

b) Une famille quelconque de sous-tribus $\mathcal{A}_i \subset \mathcal{A}$, $i \in I$, est **mutuellement indépendante** *si toute famille d'événements $A_i \in \mathcal{A}$, $i \in I$, est mutuellement indépendante.*

Exemples : i) Soit $(\Omega, \mathcal{A}, \mathrm{P}) = ([0,1], \mathcal{B}([0,1]), \lambda)$ et soit, pour tout $n \in \mathbb{N}^*$

$$A_n = \bigcup_{1\leqslant j\leqslant 2^{n-1}} \left]\frac{2(j-1)}{2^n}, \frac{2j-1}{2^n}\right].$$

La famille $\{A_n : n \in \mathbb{N}^*\}$ est mutuellement indépendante.
ii) On jette deux dés, un rouge et un noir. Les événements

$$A_1 = \{ \text{ le dé rouge montre un nombre impair } \},$$

$$A_2 = \{ \text{ le dé noir montre un nombre impair } \},$$

$$A_3 = \{ \text{ la somme des deux dés est un nombre impair } \}$$

sont trois évévenements indépendants deux à deux (c'est-à-dire, A_1, A_2 sont indépendants, A_2, A_3 sont indépendants et A_1, A_3 sont indépendants), mais ne sont pas indépendants mutuellement : pour $i, j \in \{1,2,3\}$, $i \neq j$,

$$\mathrm{P}(A_i) = \frac{1}{2},\ \mathrm{P}(A_i \cap A_j) = \frac{1}{4},\ \mathrm{P}(A_1 \cap A_2 \cap A_3) = 0.$$

□

Par la suite par le mot "indépendance" on désigne "indépendance mutuelle" (et toute autre forme d'indépendance, plus faible, sera précisée explicitement).

Proposition 3.1 *Si $\mathcal{E}_1, \mathcal{E}_2$ sont deux familles d'ensembles stables par intersection finie (ou deux algèbres) indépendantes, dans l'espace de probabilité $(\Omega, \mathcal{A}, \mathrm{P})$, alors les tribus $\sigma(\mathcal{E}_1)$, $\sigma(\mathcal{E}_2)$ sont indépendantes.*

Preuve. Soit $A_1 \in \mathcal{E}_1$. La classe monotone

$$\mathcal{M}_1 := \{A_2 \in \sigma(\mathcal{E}_2) : \mathrm{P}(A_1 \cap A_2) = \mathrm{P}(A_1)\mathrm{P}(A_2)\}$$

des événements indépendants de A_1 contient $\mathcal{E}_2$. Elle contient donc la classe monotone engendrée par $\mathcal{E}_2$ et par le théorème de classe monotone, elle contient $\sigma(\mathcal{E}_2)$. Soit maintenant $A_2 \in \sigma(\mathcal{E}_2)$. La classe monotone

$$\mathcal{M}_2 := \{A_1 \in \sigma(\mathcal{E}_1) : \mathrm{P}(A_1 \cap A_2) = \mathrm{P}(A_1)\mathrm{P}(A_2)\}$$

des événements indépendants de A_2 contient $\mathcal{E}_1$ d'après ce qui précéde et donc $\sigma(\mathcal{E}_1)$. □.

Définition 3.3 *Une famille quelconque de variables aléatoires X_i, $i \in I$, sur $(\Omega, \mathcal{A}, \mathrm{P})$ à valeurs dans $(E, \mathcal{B})$ est (mutuellement)* **indépendante** *si la famille des tribus engendrées par les X_i est (mutuellement) indépendante, c'est-à-dire, pour tout $J \subset I$ fini et tous les ensembles mesurables B_j, $j \in J$,*

$$\mathrm{P}(X_j \in B_j : j \in J) = \mathrm{P}\left(\bigcap_{j \in J} \{X_j \in B_j\}\right) = \prod_{j \in J} \mathrm{P}(X_j \in B_j).$$

Exemples : i) On reprend l'exemple du jet de dés rouge et noir. Soient

$$X_1(i,j) = i,\ X_2(i,j) = j,$$

les projections de Ω sur la première et la seconde composante. Alors

$$\begin{aligned}\mathrm{P}(A \cap B) &= \mathrm{P}\left(X_1^{-1}(\{5,6\}) \cap X_2^{-1}(\{4\})\right) = \mathrm{P}(\{(5,4),(6,4)\}) = \frac{1}{18} \\ &= \mathrm{P}(A)\mathrm{P}(B),\end{aligned}$$

donc A et B sont indépendants. On vérifie que X_1 et X_2 le sont aussi. La tribu $\sigma(X_1)$ est formée des ensembles $A_1 \times \{1,2,3,4,5,6\}$ avec $A_1 \subset \{1,2,3,4,5,6\}$ et la tribu $\sigma(X_2)$ est formée des ensembles $\{1,2,3,4,5,6\} \times B_2$ avec $B_2 \subset \{1,2,3,4,5,6\}$. Donc si $A \in \sigma(X_1)$ et $B \in \sigma(X_2)$ sont non vides, alors

$$A = A_1 \times \{1,2,3,4,5,6\},\ B = \{1,2,3,4,5,6\} \times B_2$$

et

$$\begin{aligned}\mathrm{P}(A \cap B) = \mathrm{P}(A_1 \times B_2) &= \sum_{(i,j) \in A_1 \times B_2} \frac{1}{36} \\ = \sum_{i \in A_1} \frac{1}{6} \sum_{j \in B_2} \frac{1}{6} &= \mathrm{P}(A_1)\mathrm{P}(B_2) = \mathrm{P}(A)\mathrm{P}(B).\end{aligned}$$

ii) La famille de variables aléatoires

$$X_n = \mathbb{1}_{A_n} = \sum_{1 \leqslant j \leqslant 2^{n-1}} \mathbb{1}_{\left]\frac{2(j-1)}{2^n}, \frac{2j-1}{2^n}\right]},\ n \in \mathbb{N}$$

de $[0,1]$ dans $\{0,1\}$ est indépendante. La loi commune des X_n est une loi de Bernoulli de paramètre $\frac{1}{2}$.

Il est utile de reformuler l'indépendance des variables aléatoires en terme de lois de ces variables. On a vu que lorsqu'on donne un vecteur aléatoire $X = (X_1, \ldots, X_d)$, sa loi détermine toutes les lois marginales, c'est-à-dire la loi de chacune des X_j. On a aussi vu que la réciproque est fausse en général ; néanmoins si les coordonnées sont (mutuellement) indépendantes, les lois marginales déterminent la loi du vecteur.

Proposition 3.2 *Soit $(X_1, \ldots, X_d)$ une famille finie de variables aléatoires indépendantes sur $(\Omega, \mathcal{A}, \mathrm{P})$. La loi $\mathrm{P}_{(X_1,\ldots,X_d)}$ du vecteur aléatoire sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ est égale au produit des lois marginales $\mathrm{P}_{X_1} \otimes \ldots \otimes \mathrm{P}_{X_d}$. Réciproquement, si la loi du vecteur est égale au produit des marginales, alors les variables sont indépendantes.*

Preuve. Pour simplifier on va supposer $d = 2$. Si $C = A \times B$ est un pavé dans $\mathcal{B}(\mathbb{R}^2)$, par l'hypothèse d'indépendance

$$\mathrm{P}_{(X_1,X_2)}(C) = \mathrm{P}(X_1^{-1}(A) \cap X_2^{-1}(B)) = \mathrm{P}(X_1^{-1}(A))\mathrm{P}(X_2^{-1}(B)) = \mathrm{P}_{X_1}(A)\mathrm{P}_{X_2}(B).$$

L'égalité s'étend à la famille des réunions finies disjointes de pavés (fermée à l'intersection finie) laquelle engendre $\mathcal{B}(\mathbb{R}^2)$. La réciproque se déduit de l'identité précédente et de la définition de la loi, car

$$\mathrm{P}((X_1, X_2) \in A \times B) = \mathrm{P}_{(X_1,X_2)}(C) = \mathrm{P}_{X_1}(A)\mathrm{P}_{X_2}(B) = \mathrm{P}(X_1 \in A)\mathrm{P}(X_2 \in B).$$

□

Exemple. Soit $(X, Y) \in \mathbb{R}^2$ un couple aléatoire de densité $f(x)g(y)$ par rapport à la mesure de Lebesgue sur $\mathbb{R}^2$, avec f, g boréliennes positives. Alors X et Y sont indépendantes, et de densités proportionnelles à f et respectivement g.

Proposition 3.3 *Une famille quelconque de variables aléatoires réelles* $\{X_i : i \in I\}$ *sur* $(\Omega, \mathcal{A}, \mathrm{P})$ *est indépendante si et seulement si pour toute famille finie* $J \subset I$ *et toute famille finie de fonctions boréliennes* ϕ_j $j \in J$*, telles que* $\phi_j(X_j) \in \mathrm{L}^1$*,* $j \in J$*,*

$$\mathrm{E}\left(\prod_{j\in J} \phi_j(X_j)\right) = \prod_{j\in J} \mathrm{E}(\phi_j(X_j)).$$

Preuve. Supposons la famille $\{X_i : i \in I\}$ indépendante et soit $J \subset I$ une partie finie, par exemple $J = \{1, \dots, n\}$. Par le théorème de Fubini

$$\mathrm{E}\left(\prod_{j=1}^{n} \phi_j(X_j)\right) = \int \prod_{j=1}^{n} \phi_j(x_j) d\mathrm{P}_{(X_1,\dots,X_n)}(x_1, \dots, x_n)$$

$$= \int \prod_{j=1}^{n} \phi_j(x_j) d\mathrm{P}_{X_1}(x_1) \otimes \dots d\mathrm{P}_{X_n}(x_n) = \prod_{j=1}^{n} \int \phi_j(x_j) d\mathrm{P}_{X_j}(x_j) = \prod_{j=1}^{n} \mathrm{E}\left(\phi_j(X_j)\right).$$

La réciproque s'obtient en considérant pour les fonctions ϕ_j des indicatrices de boréliens. □

Proposition 3.4 *La famille* $(X_1, \dots, X_d)$ *de variables aléatoires réelles est indépendante si et seulement si pour tout* $(t_1, \dots, t_d) \in \mathbb{R}^d$*,*

$$\varphi_{(X_1,\dots,X_d)}(t_1, \dots, t_d) = \varphi_{X_1}(t_1) \dots \varphi_{X_d}(t_d).$$

Preuve. Le produit $\varphi_{X_1} \dots \varphi_{X_d}$ est la fonction caractéristique de la loi produit $\mathrm{P}_{X_1} \otimes \dots \otimes \mathrm{P}_{X_d}$ et le résultat s'ensuit car la fonction caractéristique caractérise la loi. □

Remarque. Un cas particulier de la Proposition 3 est que si $X_1, \dots, X_d$ sont indépendantes et intégrables alors

$$\mathrm{E}(X_1 \dots X_d) = \mathrm{E}(X_1) \dots \mathrm{E}(X_d).$$

Cette propriété ne caractérise pas l'indépendance. Elle décrit une propriété plus faible :

Définition 3.4 *Deux variables aléatoires réelles* $X, Y \in \mathrm{L}^2((\Omega, \mathcal{A}, \mathrm{P}))$ *sont dites* **non corrélées** *si*

$$\mathrm{E}(XY) = \mathrm{E}(X)\mathrm{E}(Y) \Leftrightarrow \mathrm{E}((X - \mathrm{E}(X)))(Y - \mathrm{E}(Y))) = 0.$$

On dit aussi que les variables centrées $X - \mathrm{E}(X)$ *et* $Y - \mathrm{E}(Y)$ *sont orthogonales dans* L^2 *(pour son produit scalaire).*

Exemples. i) Deux variable aléatoires indépendantes de carré intégrable sont non corrélées.
ii) Soit $X \sim \mathcal{N}(0,1)$ et on note $Y = X^2$. Alors X et Y sont non corrélées. En effet, elles sont de carré intégrable et

$$\mathrm{E}(XY) = \mathrm{E}(X^3) = 0 = \mathrm{E}(X)\mathrm{E}(Y),$$

par les calculs des moments des variables gaussiennes centrées. Toutefois X, Y ne sont pas indépendantes car

$$\mathrm{P}(X \geqslant 1, Y \geqslant 1) = \mathrm{P}(X \geqslant 1) \neq \mathrm{P}(X \geqslant 1)\mathrm{P}(Y \geqslant 1),$$

puisque $\mathrm{P}(Y \geqslant 1) \lneqq 1$.

Proposition 3.5 *(identité de Bienaymé et inégalité de Bienaymé-Tchebytchev).*
Si $X_1, \ldots, X_d$ *sont deux-à-deux non corrélées, alors*

$$\mathrm{Var}\left(\sum_{j=1}^{d} X_j\right) = \sum_{j=1}^{d} \mathrm{Var}(X_j).$$

On peut en déduire l'inégalité

$$\mathrm{P}\left(\left|\sum_{j=1}^{d}(X_j - \mathrm{E}(X_j))\right| \geqslant t\right) \leqslant \frac{1}{t^2}\sum_{j=1}^{d} \mathrm{Var}(X_j),\ t > 0.$$

Preuve. Les variables $X_j - \mathrm{E}(X_j)$ et $X_k - \mathrm{E}(X_k)$, $j \neq k$, sont orthogonales dans L^2, donc

$$\mathrm{Var}\left(\sum_{j=1}^{d} X_j\right) = \mathrm{E}\left(\left(\sum_{j=1}^{d}(X_j - \mathrm{E}(X_j))\right)^2\right)$$

$$= \sum_{1 \leqslant j,k \leqslant d} \mathrm{E}\left((X_j - \mathrm{E}(X_j))(X_k - \mathrm{E}(X_k))\right) = \sum_{j=1}^{d} \mathrm{E}\left((X_j - \mathrm{E}(X_j))^2\right) = \sum_{j=1}^{d} \mathrm{Var}(X_j).$$

L'inégalité Bienaymé-Tchebytchev est une conséquence de l'inégalité de Tchebytchev. $\square$

3.2 Sommes de variables aléatoires indépendantes

On a vu que pour des variables aléatoires $X_1, \dots, X_n$ indépendantes de même loi et de carré intégrable, l'inégalité de Bienaymé-Tchebytchev montre que si $t > 0$,

$$\mathrm{P}\left(\left|\sum_{j=1}^{d}(X_j - \mathrm{E}(X_1))\right| \geqslant t\sqrt{n}\right) \leqslant \frac{\mathrm{Var}(X_1)}{t^2}.$$

Ainsi l'ordre de grandeur de la somme $\sum_{j=1}^{d}(X_j - \mathrm{E}(X_1))$ est au plus $\sqrt{n}$ (on peut dire que $\sum_{j=1}^{d} X_j$ ressemble à un terme déterministe $n\mathrm{E}(X_1)$ de l'ordre n si $\mathrm{E}(X_1) \neq 0$ plus un terme aléatoire de l'ordre au plus $\sqrt{n}$). On va évaluer la loi de $\sum_{j=1}^{d} X_j$ On précisera la loi du terme aléatoire au chapitre IV. La somme de variables aléatoires indépendantes joue un rôle essentiel en probabilités et donc on calculera sa loi.

Proposition 3.6 *Soient X et Y deux variables aléatoires réelles indépendantes sur $(\Omega, \mathcal{A}, \mathrm{P})$. La loi de la somme $X+Y$ est donnée par le* **produit de convolution** $\mathrm{P}_X \star \mathrm{P}_Y$ *des lois P_X et P_Y, défini pour toute fonction borélienne bornée $\phi : \mathbb{R} \to \mathbb{R}$, par*

$$\int_{\mathbb{R}} \phi(t)(\mathrm{P}_X \star \mathrm{P}_Y)(dt) = \int_{\mathbb{R}^2} \phi(x+y)\mathrm{P}_X(dx)\mathrm{P}_Y(dy).$$

Preuve. Soit $g(x,y) = x+y$ et on écrit le théorème de transport :

$$\int_{\mathbb{R}} \phi(t)\mathrm{P}_{X+Y}(dt) = \mathrm{E}\left(\phi(X+Y)\right) = \mathrm{E}\left(\phi \circ g(X,Y)\right)$$

$$= \iint_{\mathbb{R}^2} \phi \circ g(x,y) d\mathrm{P}_{(X,Y)}(x,y) = \iint_{\mathbb{R}^2} \phi \circ g(x,y) d(\mathrm{P}_X \otimes \mathrm{P}_Y)(x,y)$$

$$= \iint_{\mathbb{R}^2} \phi(x+y)\mathrm{P}_X(dx)\mathrm{P}_Y(dy).$$

$\square$.

Remarque : Le produit de convolution vérifie un certain nombre de propriétés algébriques issues de sa description en terme de variables aléatoires, mais qui ne suffisent pas à le caractériser :

i) (élément neutre) $\mathrm{P}_X \star \delta_0 = \mathrm{P}_X$ (car $X + 0 = X$) ;
ii) (commutativité) $\mathrm{P}_X \star \mathrm{P}_Y = \mathrm{P}_Y \star \mathrm{P}_X$ (car $X+Y = Y+X$) ;
iii) (associativité) $(\mathrm{P}_X \star \mathrm{P}_Y) \star \mathrm{P}_Z = \mathrm{P}_X \star (\mathrm{P}_Y \star \mathrm{P}_Z)$ (car $(X+Y)+Z = X+(Y+Z)$) ;
iv) (distributivité) $\mathrm{P}_X \star (\lambda \mathrm{P}_Y + (1-\lambda)\mathrm{P}_Z) = \lambda(\mathrm{P}_X \star \mathrm{P}_Y) + (1-\lambda)(\mathrm{P}_X \star \mathrm{P}_Z)$, pour tout $\lambda \in [0,1]$. $\square$

Remarque : Soient X et Y deux variables aléatoires indépendantes à valeurs dans $\mathbb{N}$. Alors la loi de $X+Y$ est donnée par

$$\mathrm{P}(X+Y=n) = \sum_{j=0}^{n} \mathrm{P}(X=j)\mathrm{P}(Y=n-j).$$

En effet, $\{X+Y=n\} = \cup_{j=0}^{n}\{X=j, Y=n-j\}$ et il suffit de remarquer que les événements sont incompatibles et d'utiliser l'indépendance de X et Y. $\square$

REMARQUE : Soient X et Y deux variables aléatoires réelles indépendantes ayant les densités par rapport à la mesure de Lebesgue f et g. Alors la somme $X+Y$ a une densité h par rapport à la mesure de Lebesgue, donnée par le **produit de convolution** des fonctions f, g

$$h(x) := (f \star g)(x) = \int_{\mathbb{R}} f(x-y)g(y)dy = \int_{BBr} g(x-y)f(y)dy = (g \star f)(x), \, x \in \mathbb{R}.$$

En effet, si ϕ est une fonction borélienne bornée

$$\int_{\mathbb{R}} \phi(t)(\mathrm{P}_X \star \mathrm{P}_Y)(dt) = \int_{\mathbb{R}^2} \phi(x+y)\mathrm{P}_X(dx)\mathrm{P}_Y(dy)$$

$$= \int_{\mathbb{R}^2} \phi(x+y)f(x)g(y)dxdy = \int_{\mathbb{R}^2} \phi(z)f(z-y)g(y)dzdy = \int_{\mathbb{R}} \phi(z)h(z)dz.$$

Ce produit de convolution a un effet "régularisant" car la fonction h est continue bien que f et g sont seulement mesurables (et positives d'intégrale 1). $\square$

Les fonctions caractéristiques fournissent un autre moyen de déterminer la loi de la somme de deux variables aléatoires indépendantes.

Proposition 3.7 *Si X et Y sont deux variables aléatoires réelles indépendantes sur $(\Omega, \mathcal{A}, \mathrm{P})$, la fonction caractéristique de leur somme est donnée par le produit des fonctions caractéristiques*

$$\varphi_{X+Y}(t) = \varphi_X(t)\varphi_Y(t), \, t \in \mathbb{R}.$$

Preuve. Il s'agit d'une conséquence de la Proposition 3.3, car pour tout $t \in \mathbb{R}$,

$$\varphi_{X+Y}(t) = \mathrm{E}\left(e^{it(X+Y)}\right) = \mathrm{E}\left(e^{itX}e^{itY}\right) = \mathrm{E}\left(e^{itX}\right)\mathrm{E}\left(e^{itY}\right) = \varphi_X(t)\varphi_Y(t).$$

$\square$

REMARQUE : Il ne faut pas confondre la fonction caractéristique d'un couple aléatoire dont les coordonnées sont indépendantes

$$\varphi_{(X,Y)}(s,t) = \varphi_X(s)\varphi_Y(t), \, (s,t) \in \mathbb{R}^2$$

avec la fonction caractéristique de la somme

$$\varphi_{X+Y}(t) = \varphi_X(t)\varphi_Y(t), \, t \in \mathbb{R}.$$

$\square$

Exemples. i) Si $X = a$ p.s. et $Y = b$ p.s., $a \neq b$, alors X et Y sont indépendantes et

$$X + Y = a + b, \text{ p.s. } \Leftrightarrow \delta_a \star \delta_b = \delta_{a+b}.$$

ii) Si $X \sim \mathcal{P}(\lambda)$ et $X \sim \mathcal{P}(\mu)$ sont indépendantes, alors $X + Y \sim \mathcal{P}(\lambda + \mu)$.
iii) Si $X \sim \mathcal{B}(n,p)$ et $X \sim \mathcal{B}(m,p)$ sont indépendantes, alors $X + Y \sim \mathcal{B}(n+m,p)$.
iv) On jette une pièce n fois et on veut savoir la loi du nombre de pile. On modèlise les n jets par n variables aléatoires indépendantes $X_1, \ldots, X_n$ chacune de loi de Bernoulli $\mathcal{B}(1,p)$ (avec $p = \frac{1}{2}$ si la pièce n'est pas truquée), où $X_j = 1$ si au j-ème lancer on obtient pile et $= 0$ sinon. Le nombre de pile est donc $S_n = X_1 + \ldots + X_n$ et par iii) on voit que $S_n \sim \mathcal{B}(n,p)$. En particulier $\frac{S_n}{n}$ est le nombre moyen de pile sur n lancers. De plus par l'inégalité de Bienaymé-Tchebitchev, pour tout $\varepsilon > 0$

$$\mathrm{P}\left(\left|\frac{S_n}{n} - p\right| \geqslant \varepsilon\right) \leqslant \frac{1}{\varepsilon^2}\mathrm{Var}\left(\frac{S_n}{n}\right) = \frac{p(1-p)}{\varepsilon^2 n}.$$

Donc la probabilité que $\frac{S_n}{n}$ s'écarte de sa valeur moyenne tend vers 0 lorsque n tend vers l'infini. Cela signifie que si nous lançons un grand nombre de fois une pièce non truquée la proportion de pile sera avec un grande probabilité $\frac{1}{2}$.

3.3 Applications de l'indépendance

Étant données des probabilités P_j, par exemple sur $\mathbb{R}$ peut-on trouver des variables aléatoires X_j sur un certain espace de probabilité $(\Omega, \mathcal{A}, \mathrm{P})$ qui soient indépendantes et telles que $\mathrm{P}_{X_j} = \mathrm{P}_j$? Lorqu'il y a un nombre fini $(\mathrm{P}_1, \ldots, \mathrm{P}_d)$ cela ne pose pas de problème : on prend $\Omega = \mathbb{R}^d$ muni de sa tribu borélienne, $\mathrm{P} = \mathrm{P}_1 \otimes \ldots \otimes \mathrm{P}_d$ et on considère $X_j : \mathbb{R}^d \to \mathbb{R}$ les applications coordonnées.

On va décrire à présent le cas dénombrable $\{\mathrm{P}_j : j \in \mathbb{N}\}$. On cherche à construire un espace de probabilité $(\Omega, \mathcal{A}, \mathrm{P})$ et une famille de variables aléatoires indépendantes $\{X_j : j \in \mathbb{N}\}$ sur $(\Omega, \mathcal{A}, \mathrm{P})$ à valeurs dans $(E_j, \mathcal{B}_j)$, $j \in \mathbb{N}$, telles que $\mathrm{P}_{X_j} = \mathrm{P}_j$ (en fait en pratique $E_j = \mathbb{R}$ ou $\mathbb{R}^d$).

Soit $\Omega = \prod_{j \in \mathbb{N}} E_j$ et on pose X_j la projection sur la j-ème coordonnée. Soit $\mathcal{A}$ la tribu produit des $\mathcal{B}_j$, $j \in \mathbb{N}$ ou équivalent, la tribu engedrée par les X_j. Cette tribu est aussi engendrée par la famille $\mathcal{C}$ des **cylindres** qui sont les ensembles A de la forme

$$A = C_n \times E_{n+1} \times E_{n+2} \times \ldots,\ C_n \in \mathcal{B}_1 \otimes \ldots \otimes \mathcal{B}_n$$

(C_n est la **base** de A). Sur $\mathcal{C}$ on peut définir la fonction $\mathrm{Q} : \mathcal{C} \to [0,1]$, par

$$\mathrm{Q}(A) := \mathrm{P}_1 \otimes \ldots \otimes \mathrm{P}_n(C_n),\ A \in \mathcal{C}\ \text{ (de base } C_n).$$

Théorème 3.1 *(de Kolmogorov).*
La fonction d'ensemble Q *se prolonge en une unique probabilité* P *sur* $(\Omega, \mathcal{A})$*. Sous* P*, les variables aléatoires* X_j *sont indépendantes et de lois* P_j*.*

De ce théorème (admis) on peut parler librement d'une suite $\{X_n : n \in \mathbb{N}\}$ de variables aléatoires réelles indépendantes sur un espace de probabilité $(\Omega, \mathcal{A}, \mathrm{P})$.

Définition 3.5 *Soit* $\{\mathcal{T}_n : n \in \mathbb{N}\}$ *une famille indépendante de tribus sur* $(\Omega, \mathcal{A}, \mathrm{P})$ *(par exemple, on peut prendre* $\mathcal{T}_n = \sigma(X_n)$ *avec* $\{X_n : n \in \mathbb{N}\}$ *variables aléatoires indépendantes).*

On désigne par $\mathcal{A}_n$ la tribu engendrée par $\mathcal{T}_n, \mathcal{T}_{n+1}, \ldots$ (dans l'exemple la tribu est engendrée par $X_n, X_{n+1}, \ldots$) et on note **la tribu asymptotique** *ou* **terminale** *:*

$$\mathcal{A}_\infty := \bigcap_{n \in \mathbb{N}} \mathcal{A}_n$$

Théorème 3.2 *(loi du tout ou rien ou loi de 0-1).*
Pour tout $A \in \mathcal{A}_\infty$, $\mathrm{P}(A) = 0$ ou $= 1$ (on dit que $\mathcal{A}_\infty$ est p.s. triviale).

Preuve. Soit $A \in \mathcal{A}_\infty$ fixé. On considère la classe monotone des événements indépendants de A,

$$\mathcal{M} := \{B \in \mathcal{A} : \mathrm{P}(A \cap B) = \mathrm{P}(A)\mathrm{P}(B)\}.$$

Si on montre que $\mathcal{M} \supset \mathcal{A}_\infty$, alors $A \in \mathcal{M}$ et donc $\mathrm{P}(A) = \mathrm{P}(A)^2$, c'est-à-dire $\mathrm{P}(A) = 0$ ou $= 1$.
Soit, pour tout n, $\mathcal{B}_n = \sigma(\mathcal{T}_0, \ldots, \mathcal{T}_n)$ et on pose $\mathcal{B}_\infty = \cup_n \mathcal{B}_n$. En tant que réunion croissante, $\mathcal{B}_\infty$ est une algèbre. On peut voir que les tribus $\mathcal{B}_n$ et $\mathcal{A}_{n+1}$ sont indépendantes. Ainsi, tout élément de $\mathcal{B}_n$ est indépendant de $A \in \mathcal{A}_\infty \subset \mathcal{A}_{n+1}$. Ainsi $\mathcal{B}_\infty \subset \mathcal{M}$. Par le théorème de classe monotone, $\sigma(\mathcal{B}_\infty) = \mathcal{M}(\mathcal{B}_\infty) \subset \mathcal{M}$. Il reste à prouver que $\sigma(\mathcal{B}_\infty) \supset \mathcal{A}_\infty$. En effet, pour tout k,

$$\mathcal{T}_k \subset \mathcal{B}_k \subset \mathcal{B}_\infty \subset \sigma(\mathcal{B}_\infty).$$

Donc, pour tout n, $\mathcal{A}_n = \sigma(\mathcal{T}_k : k \geqslant n) \subset \sigma(\mathcal{B}_\infty)$, d'où le résultat. □

Exemples. i) Soit $\{A_n : n \in \mathbb{N}\}$ une suite d'événements indépendants de $(\Omega, \mathcal{A}, \mathrm{P})$. Alors

$$A = \limsup_n A_n = \{A_n \text{ se réalise une infinité de fois }\}$$

est un événement terminal pour la suite de tribus $\mathcal{T}_n = \sigma(A_n) = \{\emptyset, A_n, A_n^c, \Omega\}$. Donc $\mathrm{P}(A) = 0$ ou 1.

On abrège "A_n se réalise une infinité de fois" par "A_n infiniment souvent" ou "A_n i.s.". Notons que $\mathrm{P}(A_n \text{ i.s.}) = 0$ signifie que p.s. seulement un nombre fini d'événements A_n se réalisent : pour presque tous $\omega \in \Omega$, il existe $n(\omega) \in \mathbb{N}$ (dépend de ω) tel que si $n \geqslant n(\omega)$, alors $\omega \notin A_n$, c'est-à-dire A_n ne se réalise pas.
ii) Soient $\{X_n : n \in \mathbb{N}\}$ variables aléatoires réelles indépendantes, $\mathcal{T}_n = \sigma(X_n)$ et soit $\{a_n : n \in \mathbb{N}\}$ une suite de réels positifs telle que $\lim_{n\to\infty} a_n = \infty$. On note

$$A = \left\{\omega \in \Omega : \frac{1}{a_n}(X_1(\omega) + \ldots + X_n(\omega)) \text{ converge }\right\}.$$

Alors $A \in \mathcal{A}_\infty$ car pour tout $m \in \mathbb{N}$,

$$A = \left\{\omega \in \Omega : \frac{1}{a_n}(X_m(\omega) + \ldots + X_n(\omega)) \text{ converge }\right\}.$$

Donc $\mathrm{P}(A) = 0$ ou 1.

Théorème 3.3 *(Lemme de Borel-Cantelli).*
Soit $\{A_n : n \in \mathbb{N}\}$ une suite d'événements de $(\Omega, \mathcal{A}, \mathrm{P})$.

a) $\sum_{n\in\mathbb{N}} \mathrm{P}(A_n) < \infty \Longrightarrow \mathrm{P}(A_n \text{ } i.s. \text{ }) = 0.$

b) Si la suite $\{A_n : n \in \mathbb{N}\}$ *est indépendante, alors la réciproque est vraie :* $\sum_{n\in\mathbb{N}} \mathrm{P}(A_n) = \infty \Longrightarrow \mathrm{P}(A_n \text{ } i.s. \text{ }) = 1.$

Preuve. La partie a) est évidente, car pour tout n

$$A = \bigcap_{n\in\mathbb{N}} \bigcup_{m\geqslant n} A_m \subset \bigcup_{m\geqslant n} A_m,$$

et donc, $\mathrm{P}(A_n \text{ i.s.}) = \mathrm{P}(A) \leqslant \sum_{m\geqslant n} \mathrm{P}(A_m)$ qui tend vers 0, lorsque $n \to \infty$ si la série converge.
Pour prouver b) on remarque d'abord que pour tout n et tout $N \geqslant n$,

$$\mathrm{P}\left(\bigcup_{n\leqslant m\leqslant N} A_m\right) = 1 - \mathrm{P}\left(\bigcap_{n\leqslant m\leqslant N} A_m^c\right) = 1 - \prod_{n\leqslant m\leqslant N} (1 - \mathrm{P}(A_m)).$$

Comme $1 - x \leqslant e^{-x}$, pour tout $x \geqslant 0$, on obtient

$$\mathrm{P}\left(\bigcup_{n\leqslant m\leqslant N} A_m\right) \geqslant 1 - \exp\left(-\sum_{n\leqslant m\leqslant N} \mathrm{P}(A_m)\right).$$

Par l'hypothèse, lorsque $N \to \infty$, $\sum_{n\leqslant m\leqslant N} \mathrm{P}(A_m) \to \infty$, pour tout n. Donc $\mathrm{P}(\cup_{m\geqslant n} A_m) = 1$ et il suffit de remarquer que $\mathrm{P}(A) = \lim_{n\to\infty} \mathrm{P}(\cup_{m\geqslant n} A_m)$. □

Exemples. i) Soit $\{X_n : n \in \mathbb{N}\}$ une suite de variables aléatoires réelles telles que, pour un $M \in \mathbb{R}$, $\sum_{n\in\mathbb{N}} \mathrm{P}(X_n \geqslant M) < \infty$. Alors

$$\mathrm{P}(X_n \geqslant M \text{ i.s. }) = 0 \Leftrightarrow \mathrm{P}(\liminf_n \{X_n < M\}) = 1.$$

Donc $\limsup_{n\uparrow\infty} X_n \leqslant M$ p.s.

De la même façon, si $\sum_{n\in\mathbb{N}} \mathrm{P}(X_n \leqslant M) < \infty$, alors $\liminf_{n\uparrow\infty} X_n \geqslant M$ p.s.

ii) On lance une pièce équilibrée une infinité de fois. Quelle est la probabilité d'obtenir une infinité de fois deux pile consécutifs ? Si $\{X_n : n \in \mathbb{N}\}$ une suite de variables aléatoires indépendantes de Bernoulli de paramètre $\frac{1}{2}$ et si on pose $A_n = \{X_n = X_{n+1} = 1\}$ on s'intéresse à $\mathrm{P}(A_n \text{ i.s. })$. $\{A_n : n \in \mathbb{N}\}$ n'est pas une suite indépendante car X_{n+1} détermine à la fois A_n et A_{n+1}, mais $\{A_{2n} : n \in \mathbb{N}\}$ est indépendante, avec $\mathrm{P}(A_{2n}) = \frac{1}{4}$, pour tout n. On peut appliquer la deuxième partie du lemme de Borel-Cantelli, donc $\mathrm{P}(A_{2n} \text{ i.s. }) = 1$. Comme $\{A_{2n} \text{ i.s. }\} \subset \{A_n \text{ i.s. }\}$, on conclut que $\mathrm{P}(A_n \text{ i.s. }) = 1$.

3.4 Vecteurs gaussiens et indépendance

$X : (\Omega, \mathcal{A}, \mathrm{P}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ suit la loi gaussienne $\mathcal{N}(m, \sigma^2)$ si sa densité par rapport à la mesure de Lebesgue sur $\mathbb{R}$ est

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

ou, équivalent sa fonction caractéristique est

$$\varphi_X(t) = \exp\left(imt - \frac{1}{2}\sigma^2 t^2\right),\ t \in \mathbb{R}.$$

Si $Y \sim \mathcal{N}(0,1)$, alors $X = m + \sigma Y \sim \mathcal{N}(m, \sigma^2)$. Les paramètres d'espérance m et de variance σ^2 caractérisent la loi.

Définition 3.6 *Le vecteur aléatoire* $X = (X_1, \ldots, X_d) : (\Omega, \mathcal{A}, \mathrm{P}) \to (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ **est de loi gaussienne** *si pour tout* $\alpha = (\alpha_1, \ldots, \alpha_d) \in \mathbb{R}^d$*, la variable aléatoire réelle*

$$< \alpha, X >= \sum_{j=1}^{d} \alpha_j X_j$$

est gaussienne.

Dans cette définition la variable $< \alpha, X >$ est caractérisée par son espérance

$$\mathrm{E}\left(\sum_{j=1}^{d} \alpha_j X_j\right) = \sum_{j=1}^{d} \alpha_j \mathrm{E}(X_j)$$

et par sa variance

$$\mathrm{Var}\left(\sum_{j=1}^{d} \alpha_j X_j\right) = \sum_{1 \leqslant j,k \leqslant d} \alpha_j \alpha_k \mathrm{E}\left((X_j - \mathrm{E}(X_j))(X_k - \mathrm{E}(X_k))\right).$$

Ainsi le vecteur gaussien $X = (X_1, \ldots, X_d)$ est caractérisé par son vecteur espérance

$$m = \mathrm{E}(X) = (\mathrm{E}(X_1), \ldots, \mathrm{E}(X_d))^*$$

et sa matrice covariance

$$\Sigma = \left(\mathrm{E}\left((X_j - \mathrm{E}(X_j))(X_k - \mathrm{E}(X_k))\right)\right)_{1 \leqslant j,k \leqslant d}.$$

Un vecteur gaussien centré a la matrice de covariance $\left(\mathrm{E}\left(X_j X_k\right)\right)_{1 \leqslant j,k \leqslant d}$. En termes de fonction caractéristique :

$$\varphi_X(u) = \exp\left(i < u, m > -\frac{1}{2} u^* \Sigma u\right),\ u \in \mathbb{R}^d.$$

Notons que si X est un vecteur gaussien ses coordonnées sont des variables aléatoires gaussiennes mais la réciproque est fausse en général.

Un exemple fondamental est le vecteur $G = (G_1, \ldots, G_d)$ dont les composantes sont indépendantes de même loi $\mathcal{N}(0,1)$. Ce vecteur est centré et sa matrice de covariance est la matrice identité. Sa densité par rapport à la mesure de Lebesgue sur $\mathbb{R}^d$ est

$$x \mapsto (2\pi)^{-d/2} \exp(-\|x\|^2/2) \quad \text{où} \quad \|x\|^2 = x_1^2 + \ldots + x_d^2.$$

On notera $\mathcal{N}(0, \mathrm{Id})$ la loi de G.

Proposition 3.8 *Soit* $X = (X_1, \ldots, X_d)$ *un vecteur gaussien centré de matrice de covariance* $\sigma = A^*A$, *où* A *est une matrice carrée. Alors* X *a la même loi que* AG, *où* $G \sim \mathcal{N}(0, \mathrm{Id})$.

Par cette proposition, on voit que si A est inversible, pour tout borélien $B \in \mathcal{B}(\mathbb{R}^d)$,

$$\mathrm{P}(X \in B) = \frac{1}{(2\pi)^{d/2}|\det A|} \int_B \exp\left(-\frac{1}{2} < A^{-1}x, A^{-1}x >\right) dx,$$

car $\mathrm{P}(X \in B) = \mathrm{P}(G \in A^{-1}(B))$ qui se calcule à l'aide de la densité de G.

On peut être encore plus précis : on peut toujours écrire $\Sigma = Q\Delta Q^*$, où Q est une matrice orthogonale ($Q^{-1} = Q^*$) et Δ est une matrice diagonale positive (avec éventuellement des zéros sur la diagonale). Par changement de base on voit que le vecteur gaussien Q^*X a pour matrice de covariance la matrice diagonale Δ, donc la même loi que $\sqrt{\Delta}G$. La diagonalisation de la matrice de covariance de X nous a donc permis de trouver une nouvelle base dans laquelle les composantes de X sont orthogonales :

Proposition 3.9 *Soit* X *un vecteur gaussien centré de matrice de covariance* Σ. *Si le composantes de* X *sont deux à deux non corrélées (c'est-à-dire* Σ *est diagonale), alors la famille* $(X_1, \ldots, X_d)$ *est indépendante.*

3.5 Probabilité (et espérance) conditionnelle (facultatif)

Reprenons l'exemple du lancer d'un dé. Cette expérience aléatoire peut être modélisée en donnant $\Omega = \{1, 2, 3, 4, 5, 6\}$ muni de la tribu de ses parties $\mathcal{P}(\Omega)$ et la probabilité uniforme $\mathrm{P}(\{\omega\}) = \frac{1}{6}$, pour tout $\omega \in \Omega$. La variable aléatoire résultat du lancer est l'identité sur Ω. Supposons maintenant que le lancer s'effectue sans regarder le dé et qu'un spectateur nous dise que nous avons obtenu un chiffre impair. Avec cette information on peut recalculer les chances d'avoir obtenu un certain résultat. Par exemple il est clair que la chance d'obtenir un résultat pair est nulle et celle d'obtenir un résultat parmi 1, 3 et 5 est $\frac{1}{3}$. On note $\Omega_{\text{impair}} = \{1, 3, 5\}$. Nous calculons la probabilité de $\{\omega\}$ sachant que $\omega \in \Omega_{\text{impair}}$ en calculant

$$\frac{\mathrm{P}(\{\omega\} \cap \Omega_{\text{impair}})}{\mathrm{P}(\Omega_{\text{impair}})}$$

Définition 3.7 *Soit* $(\Omega, \mathcal{A}, \mathrm{P})$ *un espace de probabilité et soit* $B \in \mathcal{A}$ *un événement tel que* $\mathrm{P}(B) > 0$.

i) On appelle **probabilité conditionnelle de** A **sachant** B *le nombre*

$$\mathrm{P}(A \mid B) := \frac{\mathrm{P}(A \cap B)}{\mathrm{P}(B)}.$$

ii) On appelle **probabilité conditionnelle sachant** B *la probabilité notée* $\mathrm{P}(\cdot \mid B)$

$$\mathcal{A} \ni A \mapsto \frac{\mathrm{P}(A \cap B)}{\mathrm{P}(B)}.$$

Remarque :
Si A et B sont indépendants $\mathrm{P}(A \mid B) = \mathrm{P}(A)$, c'est-à-dire que dans cette situation, la connaissance de B n'apporte aucune information quant à la réalisation de A. □

Remarque :
Les égalités suivantes sont vraies pour tous les événements A et B (de probabilité positive ou pas) :

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

□

Remarque :
Soit X une variable aléatoire réelle intégrable. On peut calculer son espérance par rapport à la probabilité (conditionnelle) $\mathrm{P}(\cdot \mid B)$, nommée **espérance conditionnelle sachant** B :

$$\mathrm{E}(X \mid B) := \int_\Omega X(\omega)\mathrm{P}(\cdot \mid B)(d\omega) = \frac{1}{\mathrm{P}(B)}\int_B X(\omega)\mathrm{P}(d\omega) = \frac{1}{\mathrm{P}(B)}\mathrm{E}(X\mathbb{1}_B).$$

De la même façon on peut parler de fonction de répartion conditionnelle ou de fonction caractéristique conditionnelle sachant B :

$$\mathbb{R} \ni t \mapsto \mathrm{P}(X \leqslant t \mid B),\ \mathbb{R} \ni t \mapsto \mathrm{E}(e^{itX} \mid B)$$

lesquelles caractérisent la loi de X sachant B. □

Exemples. i) Soit $X \sim \mathcal{E}(\lambda)$. Par calcul direct on voit que, pour tout $s, t > 0$ $\mathrm{P}(X > s + t \mid X > s) = \mathrm{P}(X > t)$. On dit que la loi exponentielle n'a pas de mémoire. On peut vérifier la même chose pour la loi géométrique.
ii) Soient $U_1, \ldots, U_n$ variables aléatoires indépendantes de même loi uniforme sur [0,1]. On note $M_n = \max_{1 \leqslant j \leqslant n} U_j$ et $m_n = \min_{1 \leqslant j \leqslant n} U_j$. Par calcul direct on peut vérifier que pour tous $u_1, \ldots, u_d \in [0, 1]$ et tous $0 \leqslant a < b \leqslant 1$,

$$\mathrm{P}(U_1 \leqslant u_1, \ldots, U_n \leqslant u_n \mid a \leqslant m_n \leqslant M_n \leqslant b) = \mathrm{P}(V_1 \leqslant u_1, \ldots, V_n \leqslant u_n),$$

où $V_1, \ldots, V_n$ sont variables aléatoires indépendantes de même loi uniforme sur [a,b].

Définition 3.8 *Soit* $(\Omega, \mathcal{A}, \mathrm{P})$ *un espace de probabilité. Une famille d'événements* $(B_j)_{j \in J}$*,* $J \subset \mathbb{N}$*, est* **système complet d'événements** *si les* B_j *sont disjoints et* $\mathrm{P}(\cup_{j \in J} B_j) = 1$*. Ainsi* $(B_j)_{j \in J}$ *forme une partition de* Ω *quitte à ajouter un événement négligeable.*

Proposition 3.10 *Soit* $(B_j)_{j \in J}$ *un système complet d'événements sur* $(\Omega, \mathcal{A}, \mathrm{P})$*.*

1. *Pour tout* $A \in \mathcal{A}$ *la* formule des probabilités totales *a lieu :*

$$\mathrm{P}(A) = \sum_{j \in J} \mathrm{P}(A \cap B_j) = \sum_{j \in J^*} \mathrm{P}(A \mid B_j)\mathrm{P}(B_j),$$

   où $J^* = \{j \in J : \mathrm{P}(B_j) > 0\}$*.*

2. *Si de plus* $\mathrm{P}(A) > 0$*, pour tout* $k \in J^*$ *la* formule de Bayes *a lieu :*

$$\mathrm{P}(B_k \mid A) = \frac{\mathrm{P}(A \mid B_k)\mathrm{P}(B_k)}{\sum_{j \in J^*} \mathrm{P}(A \mid B_j)\mathrm{P}(B_j)}.$$

Preuve. C'est immédiat car $\mathrm{P}(A) = \sum_{j \in J^*} \mathrm{P}(A \cap B_j)$ et pour tout k,

$$\mathrm{P}(A)\mathrm{P}(B_k \mid A) = \mathrm{P}(B_k \cap A) = \mathrm{P}(A \mid B_k)\mathrm{P}(B_k).$$

□

REMARQUE :
On observe que la tribu engendrée par une partition $(B_j)_{j \in J}$ est décrite comme la collection de toutes les unions possibles d'événements B_j et de leurs complémentaires. Ainsi tout événement peut être fractionné sur l'ensembles élémentaires B_j. □

Définition 3.9 *Un événement B d'une tribu $\mathcal{B}$ est un* **atome** *si pour tout événement $C \in \mathcal{B}$ tel que $C \subset B$, soit $C = \emptyset$ soit $C = B$.*

REMARQUE :
Ainsi si $(B_j)_{j \in J}$ est une partition mesurable de Ω, alors B_j sont les atomes de la tribu engendrée par les B_j, $\mathcal{B} = \sigma(B_j : j \in J)$. □

Définition 3.10 *Soit $\mathcal{B} = \sigma(B_j : j \in J)$, où $(B_j)_{j \in J}$ est une partition mesurable sur $(\Omega, \mathcal{A}, \mathrm{P})$ et soit $J^* = \{j \in J : \mathrm{P}(B_j) > 0\}$. On appelle* **probabilité conditionnelle de** $A \in \mathcal{A}$ **sachant** $\mathcal{B}$*, la variable aléatoire*

$$\mathrm{P}(A \mid \mathcal{B}) = \sum_{j \in J^*} \mathrm{P}(A \mid B_j) \mathbb{1}_{B_j}.$$

REMARQUE :
La probabilité conditionnelle de A sachant $\mathcal{B}$ est donc une variable aléatoire constante sur les atomes de $\mathcal{B}$, donc mesurable par rapport à $\mathcal{B}$. Pour tout $\omega \in \Omega$ l'application $\mathcal{A} \ni A \mapsto \mathrm{P}(A \mid \mathcal{B})(\omega)$ est une probabilité telle que $\mathrm{P}(B_j \mid \mathcal{B})(\omega) = 1$ si $\omega \in B_j$ et $\mathrm{P}(A \mid \mathcal{B})(\omega) = 0$, si $\omega \in B_j$ et $\mathrm{P}(A \cap B_j) = 0$. On peut aussi remarquer que, pour tout $B \in \mathcal{B}$ tel que $\mathrm{P}(B) > 0$,

$$\mathrm{E}\left(\mathrm{P}(A \mid \mathcal{B}) \mathbb{1}_B\right) = \mathrm{P}(A \cap B) = \mathrm{E}\left(\mathbb{1}_A \mathbb{1}_B\right).$$

On pourrait aussi écrire

$$\mathrm{P}(A \mid \mathcal{B}) = \sum_{j \in J^*} \left(\frac{1}{\mathrm{P}(B_j)} \mathrm{E}\left(\mathbb{1}_A \mathbb{1}_{B_j}\right) \right) \mathbb{1}_{B_j}.$$

□

Définition 3.11 *Soit $\mathcal{B} = \sigma(B_j : j \in J)$, où $(B_j)_{j \in J}$ est une partition mesurable sur $(\Omega, \mathcal{A}, \mathrm{P})$ et soit $J^* = \{j \in J : \mathrm{P}(B_j) > 0\}$. Soit X une variable aléatoire intégrable. On appelle* **espérance conditionnelle de** X **sachant** $\mathcal{B}$ *la variable aléatoire $\mathcal{B}$-mesurable, définie* P-*p.s.*

$$\mathrm{E}(X \mid \mathcal{B}) = \sum_{j \in J^*} \left(\frac{1}{\mathrm{P}(B_j)} \mathrm{E}\left(X \mathbb{1}_{B_j}\right) \right) \mathbb{1}_{B_j}.$$

Si $\mathcal{B}$ est engendrée par une variable aléatoire discrète Z on note $\mathrm{E}(X \mid \mathcal{B}) = \mathrm{E}(X \mid Z)$.

Chapitre 4

Convergence des suites de variables aléatoires

Les differentes notions de convergence de variables aléatoires sont essentielles pour les applications. Voici un exemple : la fréquence observée du nombre de faces pile obtenu au cours d'un jeu de pile ou face, après n lancers est "proche" de la probabilité p d'obtenir pile, pourvu que n soit grand. Donc, si p est inconnue (on ne sait pas si la pièce est truquée), nous avons un moyen de l'approximer.

Dans tout ce chapitre les suites de variables aléatoires sont supposées construites sur un espace de probabilité $(\Omega, \mathcal{A}, \mathrm{P})$. Pour simplifier on ne considère que des variables aléatoires réelles, mais les énoncés et les résultats sont vrais pour des vecteurs aléatoires à valeurs dans $\mathbb{R}^d$.

4.1 Convergence presque sûre

Définition 4.1 *Une suite de variables aléatoires réelles* $(X_n)_{n\in\mathbb{N}}$ **converge presque sûrement (p.s.)** *vers la variable aléatoire réelle* X *si*

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

Dans ce cas on note $\lim_{n\to\infty} X_n = X$ *p.s. ou* $X_n \xrightarrow{p.s.} X$*, lorsque* $n \to \infty$*.*

Remarque : On voit que l'ensemble $\{\lim_{n\to\infty} X_n = X\}$ est bien un événement, car

$$\{\lim_{n\to\infty} X_n = X\} = \bigcap_{r\geqslant 1} \bigcup_{m\in\mathbb{N}} \bigcap_{n\geqslant m} \left\{|X_n - X| < \frac{1}{r}\right\}.$$

On voit que pour une suite d'événements $(A_r)_{r\geqslant 1}$, $\mathrm{P}(\cap_{r\geqslant 1} A_r) = 1$ si et seulement si $\mathrm{P}(A_r) = 1$ pour tout r. Donc X_n converge vers X p.s. si et seulement si

$$\forall\varepsilon > 0,\ \mathrm{P}\left(\bigcup_{m\in\mathbb{N}} \bigcap_{n\geqslant m} \{|X_n - X| < \varepsilon\}\right) = 1,$$

ou, par passage au complémentaire,

$$\forall\varepsilon > 0,\ \mathrm{P}\left(\bigcap_{m\in\mathbb{N}}\bigcup_{n\geqslant m}\{|X_n - X| \geqslant \varepsilon\}\right) = 0,$$

ou, équivalent à

$$(*) \qquad \forall\varepsilon > 0,\ \mathrm{P}\left(|X_n - X| \geqslant \varepsilon \text{ i.s. }\right) = 0.$$

Par convergence monotone, c'est encore équivalent à

$$(**) \qquad \forall\varepsilon > 0,\ \lim_{m\to\infty} \mathrm{P}\left(\sup_{n\geqslant m}|X_n - X| \geqslant \varepsilon\right) = 0.$$

On peut encore caractériser la convergence presque sûre à l'aide du critère de Cauchy : X_n converge p.s. si et seulement si

$$\forall\varepsilon > 0,\ \mathrm{P}\left(\bigcup_{n\in\mathbb{N}}\bigcap_{m\geqslant n}\{|X_n - X_m| < \varepsilon\}\right) = 1.$$

Enfin on peut aussi dire que X_n converge vers X p.s. si X_n converge ponctuellement vers X, quitte à enlever un ensemble de mesure nulle (celui pour lequel $X_n(\omega)$ ne converge pas vers $X(\omega)$).

Proposition 4.1 *(lemme de Borel-Cantelli de convergence p.s.) Soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires réelles.*

1) *Si pour tout $\varepsilon > 0$, $\sum_{n\in\mathbb{N}} \mathrm{P}(|X_n - X| \geqslant \varepsilon) < \infty$, alors $X_n \xrightarrow{p.s.} X$.*
2) *Si les variables $(X_n)_{n\in\mathbb{N}}$ sont mutuellement indépendantes, alors $X_n \xrightarrow{p.s.} 0$ si et seulement si, pour tout $\varepsilon > 0$, $\sum_{n\in\mathbb{N}} \mathrm{P}(|X_n| \geqslant \varepsilon) < \infty$.*

Preuve. Pour la première partie on considère $\varepsilon > 0$ et les événements

$$A_n = \{|X_n - X| \geqslant \varepsilon\},\ n \in \mathbb{N}.$$

On applique le lemme de Borel-Cantelli 3.3 aux A_n. On déduit $\mathrm{P}(A_n \text{ i.s. }) = 0$ ce qui fournit le résultat d'après (*). La deuxième partie se démontre de façon analogue à partir de la partie avec indépendance du lemme de Borel-Cantelli 3.3. Remarquer qu'il convient de supposer X nulle ou constante, sans quoi les événements A_n ne sont pas nécessairement indépendants.□

Exemple : Soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires indépendantes et de même loi de Bernoulli $\mathcal{B}(1,p)$. On montre que $Y_n = \sum_{j=1}^n 2^{-j}X_j$ converge presque sûrement en vérifiant le critère de Cauchy : pour, $m > n$, $|Y_n - Y_m| \leqslant \sum_{j=n+1}^m 2^{-j} \leqslant 2^{-n}$. Donc

$$\bigcup_{n\in\mathbb{N}}\bigcap_{m\geqslant n}\{\omega\in\Omega : |Y_n(\omega) - Y_m(\omega)| < \varepsilon\} \supset \bigcup_{n\in\mathbb{N}}\bigcap_{m\geqslant n}\{\omega\in\Omega : 2^{-n} < \varepsilon\}$$

$$= \bigcup_{n\in\mathbb{N}}\{\omega\in\Omega : 2^{-n} < \varepsilon\} = \Omega$$

4.2 Convergence en probabilité

Définition 4.2 *Une suite de variables aléatoires réelles* $(X_n)_{n\in\mathbb{N}}$ **converge en probabilité** *vers la variable aléatoire réelle* X *si*

$$\forall \varepsilon > 0, \lim_{n\to\infty} \mathrm{P}\left(|X_n - X| \geqslant \varepsilon\right) = 0.$$

Dans ce cas on note $\lim_{n\to\infty} X_n = X$ *en probabilité ou* $X_n \xrightarrow{prob} X$*, lorsque* $n \to \infty$*.*

Remarque : Autrement dit, $\lim_{n\to\infty} X_n = X$ en probabilité si

$$\forall \varepsilon > 0, \forall \delta > 0 \, \exists n_0(\varepsilon, \delta) \text{ tel que pour } n \geqslant n_0, \ \mathrm{P}\left(|X_n - X| \geqslant \varepsilon\right) < \delta,$$

ou équivalent

$$\forall \varepsilon > 0, \exists n_0(\varepsilon) \text{ tel que pour } n \geqslant n_0, \ \mathrm{P}\left(|X_n - X| \geqslant \varepsilon\right) < \varepsilon.$$

Il est clair que la première implique la deuxième ; pour voir l'autre implication, si $\varepsilon \leqslant \delta$ il suffit de prendre $n_0(\varepsilon, \delta) = n_0(\varepsilon)$; si $\delta < \varepsilon$, on prend $n_0(\varepsilon, \delta) = n_0(\delta)$ et par la deuxième on a alors $\mathrm{P}\left(|X_n - X| \geqslant \delta\right) \leqslant \delta$ et il suffit de voir que $\{|X_n - X| \geqslant \varepsilon\} \subset \{|X_n - X| \geqslant \delta\}$.

Remarque : A nouveau on peut prendre des vecteurs aléatoires dans $\mathbb{R}^d$ avec une norme quelconque à la place de $|\cdot|$ et une fois de plus il y a équivalence avec la convergence des coordonnées grâce aux inégalités suivantes :

$$\mathrm{P}\left(\max_{1\leqslant j\leqslant d} |X_n^j - X^j| \geqslant \varepsilon\right) \leqslant \mathrm{P}\left(\cup_{1\leqslant j\leqslant d}\{|X_n^j - X^j| \geqslant \varepsilon\}\right) \leqslant \sum_{1\leqslant j\leqslant d} \mathrm{P}\left(|X_n^j - X^j| \geqslant \varepsilon\right).$$

Proposition 4.2 *La convergence presque sûre implique la convergence en probabilité.*

Preuve. On voit la différence par rapport à la convergence p.s. qui exige un supremum (**). Plus précisement, soit A l'événement de probabilité 1 qui apparaît dans la définition de la convergence presque sûre. On fixe $\varepsilon > 0$ et pour chaque entier n on considère

$$B_n = \{\omega \in A : \sup_{m\geqslant n} |X_m(\omega) - X(\omega)| \geqslant \varepsilon\}.$$

La suite (B_n) est décroissante et $\cap_n B_n = \emptyset$, car X_n converge vers X sur A. On a donc, $\lim_{n\to\infty} \mathrm{P}(B_n) = 0$. Comme

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

on tire la convergence en probabilité.

Une autre manière de prouver : si $X_n \xrightarrow{\text{p.s.}} X$, pour tout $\varepsilon > 0$,

$$\{\lim_{n\to\infty} X_n = X\} \subset \liminf_n \{|X_n - X| < \varepsilon\},$$

donc par le lemme de Fatou (ou exercice 1.14)

$$0 \leqslant \limsup_{n\to\infty} \mathrm{P}(|X_n - X| \geqslant \varepsilon) \leqslant \mathrm{P}\left(\limsup_n \{|X_n - X| \geqslant \varepsilon\}\right) \leqslant 1 - \mathrm{P}(\lim_{n\to\infty} X_n = X) = 0.$$

□

Remarque : La convergence en probabilité n'implique pas la convergence presque sûre, comme le montre l'exemple suivant : soit l'espace de probabilité $([0,1], \mathcal{B}([0,1]), \lambda)$ et on définit

$$Y_{2^n,j} = \mathbb{1}_{]\frac{j}{2^n}, \frac{j+1}{2^n}]}(\omega),\ 0 \leqslant j \leqslant 2^n - 1,\ n \geqslant 1$$

et

$$X_{2^n+j} = Y_{2^n,j},\ 0 \leqslant j \leqslant 2^n - 1,\ n \geqslant 1.$$

Alors, pour tout $\omega \in [0,1]$, $\liminf_{n\to\infty} X_n(\omega) = 0$ et $\limsup_{n\to\infty} X_n(\omega) = 1$, donc X_n ne converge pas presque sûrement. Cependant, pour tout $\varepsilon \in]0,1[$ et $m = 2^n + j$, $0 \leqslant j \leqslant 2^n - 1$, $\mathrm{P}(|X_m| \geqslant \varepsilon) = 2^{-n}$. Donc X_n converge en probabilité vers 0.

Exemples : i) Soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires réelles non-corrélées, centrées, telles que $\mathrm{Var}(X_n) = \sigma^2$. Alors leurs moyennes partielles $\frac{1}{n}\sum_{1\leqslant j\leqslant n} X_j$ convergent en probabilité vers 0. En effet, par l'inégalité de Tchebytchev,

$$\mathrm{P}\left(\left|\frac{1}{n}\sum_{1\leqslant j\leqslant n} X_j\right| \geqslant \varepsilon\right) \leqslant \frac{1}{n^2\varepsilon^2}\mathrm{Var}\left(\sum_{1\leqslant j\leqslant n} X_j\right) = \frac{\sigma^2}{n\varepsilon^2}$$

De point de vue modélisation, il suffit de lancer une cinquantaine de fois une pièce non truquée pour voir que la proportion de pile se stabilise vers 0,5.
ii) Soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires indépendantes telle que $X_n \sim \mathcal{B}(1, p_n)$. Alors

$$X_n \xrightarrow{\text{prob}} 0 \Longleftrightarrow \lim_{n\to\infty} p_n = 0.$$

Mais par le lemme de Borel-Cantelli (Proposition 4.1) on a

$$X_n \xrightarrow{\text{p.s.}} 0 \Longleftrightarrow \sum_n \mathrm{P}(|X_n| \geqslant \varepsilon) < \infty,\ \forall \varepsilon > 0,$$

ce qui est équivalent à $\sum_n p_n < \infty$, puisque $\mathrm{P}(|X_n| \geqslant \varepsilon) = p_n$ si $0 < \varepsilon < 1$.

Remarque : Il est possible de définir une métrique sur $\mathrm{L}^0(\Omega, \mathcal{A}, \mathrm{P})$ définissant la convergence en probabilité des variables aléatoires :

$$d(X,Y) := \mathrm{E}(|X - Y| \wedge 1),\ X, Y \in \mathrm{L}^0(\Omega, \mathcal{A}, \mathrm{P}).$$

En effet, $d(X,Y) = 0$ si et seulement si $X = Y$ p.s. et on vérifie que $d(X,Y) \geqslant 0$ et que $d(X,Z) \leqslant d(X,Y) + d(Y,Z)$.

Proposition 4.3 *La suite $(X_n)_{n\in\mathbb{N}}$ converge en probabilité vers X si et seulement si*

$$\lim_{n\to\infty} d(X_n, X) = 0.$$

Preuve : Par l'inégalité de Markov, pour tout $\varepsilon \in]0,1]$,

$$P(|X_n - X| \geqslant \varepsilon) = P(|X_n - X| \wedge 1 \geqslant \varepsilon) \leqslant \frac{d(X_n, X)}{\varepsilon}.$$

Par ailleurs,

$$\begin{aligned} d(X_n, X) &= E\left[(|X_n - X| \wedge 1)\mathbb{1}_{\{|X_n - X| \geqslant \varepsilon\}}\right] + E\left[(|X_n - X| \wedge 1)\mathbb{1}_{\{|X_n - X| < \varepsilon\}}\right] \\ &\leqslant P(|X_n - X| \geqslant \varepsilon) + \varepsilon. \end{aligned}$$

Si $X_n \xrightarrow{\text{prob}} X$, il existe n_0 tel que si $n \geqslant n_0$, $P(|X_n - X| \geqslant \varepsilon) \leqslant \varepsilon$. Donc $d(X_n, X) \leqslant 2\varepsilon$, pour $n \geqslant n_0$, d'où la conclusion. $\square$

Remarque : Dans tout ce qui précéde on pourrait aussi utiliser la distance

$$d'(X,Y) = E\left(\frac{|X - Y|}{1 + |X - Y|}\right).$$

Remarque : Par unicité de la limite dans tout espace métrique on déduit que la limite en probabilité est unique p.s.

Proposition 4.4 *Soit X et $(X_n)_{n \in \mathbb{N}}$ des variables aléatoires réelles. Alors $X_n \xrightarrow{prob} X$ si et seulement si de toute suite croissante d'entiers (n') on peut extraire une sous-suite (n'_k) telle que $X_{n'_k} \xrightarrow{p.s.} X$.*

Preuve :
"$\Longrightarrow$" : par la convergence en probabilité, pour tout $k \in \mathbb{N}^*$, soit n'_k le plus petit entier tel que

$$P(|X_{n'_k} - X| \geqslant \frac{1}{k}) < \frac{1}{2^k}.$$

Alors,

$$\sum_{k \in \mathbb{N}^*} P(|X_{n'_k} - X| \geqslant \frac{1}{k}) < \infty.$$

En particulier, pour tout $\varepsilon > 0$, il existe un k tel que $P(|X_{n'_k} - X| \geqslant \varepsilon) \leqslant P(|X_{n'_k} - X| \geqslant \frac{1}{k})$, d'où

$$\sum_{k \in \mathbb{N}^*} P(|X_{n'_k} - X| \geqslant \varepsilon) < \infty,$$

et donc, par le lemme de Borel-Cantelli (Proposition 4.1), $\lim_{k\to\infty} X_{n'_k} = X$ p.s.
"$\Longleftarrow$" : supposons le contraire, c'est-à-dire que X_n ne converge pas vers X en probabilité. Alors il existe une suite croissante (n'), il existe $\delta > 0$ et $\varepsilon > 0$ tels que $P(|X_{n'} - X| \geqslant \varepsilon) \geqslant \delta$. Mais, par l'hypothèse, on peut extraire une sous-suite (n'_k) telle que $X_{n'_k}$ converge p.s. vers X, et par la Proposition 4.3 en probabilité. Mais ceci est absurde car $P(|X_{n'_k} - X| \geqslant \varepsilon) \geqslant \delta$.
$\square$

Remarque : De cette proposition on voit que la convergence presque sûre n'est pas une convergence métrique, car si c'etait le cas, elle coïnciderait avec la convergence en probabilité.

Proposition 4.5 *Soient $(X_n)_{n\in\mathbb{N}}$ et $(Y_n)_{n\in\mathbb{N}}$ deux suites de variables aléatoires réelles telles que X_n converge en probabilité vers une variable aléatoire X et Y_n converge en probabilité vers la variable aléatoire Y.*

1) Si $g : \mathbb{R} \to \mathbb{R}$ est continue, alors $g(X_n)$ converge en probabilité vers $g(X)$.

2) Pour tous $\alpha, \beta \in \mathbb{R}$, $\alpha X_n + \beta Y_n$ converge en probabilité vers $\alpha X + \beta Y$.

3) De plus $X_n Y_n$ converge en probabilité vers XY.

Preuve : Prouvons 3) : soit (n') une suite extraite croissante et on extrait une sous-suite (n'') telle que $X_{n''} \to X$ p.s. De (n'') on peut extraire une sous-suite (n''') telle que $Y_{n'''} \to Y$ p.s. Alors $X_{n'''}Y_{n'''} \to XY$ p.s. On conclut à l'aide des Proposition 4.2 et 4.5. □

Proposition 4.6 *Soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires réelles telle qu'elle vérifie le critère de Cauchy en probabilité, c'est-à-dire que*

$$\forall \varepsilon > 0,\ \exists n_0,\ \forall n \geqslant m \geqslant n_0\ \mathrm{P}(|X_n - X_m| \geqslant \varepsilon) < \varepsilon,$$

ou de façon équivalente, que

$$\forall \varepsilon > 0,\ \exists n_0,\ \forall n \geqslant m \geqslant n_0\ d(X_n, X_m) < \varepsilon.$$

Alors X_n converge en probabilité si et seulement si X_n est de Cauchy en probabilité. Autrement dit, l'espace $\mathrm{L}^0(\Omega, \mathcal{A}, \mathrm{P})$ est complet pour la distance d métrisant la convergence en probabilité.

Preuve : Supposons que X_n est convergente en probabilité vers X. Ainsi, pour tout $\varepsilon > 0$, il existe n_0 tel que dès que $n \geqslant n_0$ on ait $\mathrm{P}(|X_n - X| \geqslant \varepsilon) < \varepsilon$. Par l'inégalité triangulaire on peut écrire, pour tout $n \geqslant m \geqslant n_0$

$$\{|X_n - X_m| \geqslant \varepsilon\} \subset \{|X_n - X| \geqslant \varepsilon\} \cup \{|X_m - X| \geqslant \varepsilon\}.$$

D'ici on déduit le critère de Cauchy en probabilité. Réciproquement, soit $\varepsilon = 2^{-k}$ dans la condition de Cauchy en probabilité. On pose $n_1 = 1$ et soit n_k le plus petit entier tel que $\mathrm{P}(|X_n - X_m| \geqslant 2^{-k}) < 2^{-k}$:

$$n_k = \min\{N > n_{k-1} : \mathrm{P}(|X_n - X_m| \geqslant 2^{-k}) < 2^{-k}\, \forall n, m \geqslant N\}$$

Ainsi $n_k \leqslant n_{k+1}$ et

$$\mathrm{P}(|X_{n_{k+1}} - X_{n_k}| \geqslant 2^{-k}) < 2^{-k}.$$

Par le lemme de Borel-Cantelli 3.3, pour presque tout ω, il existe un entier $k(\omega) < \infty$ tel que si $m \geqslant k(\omega)$, $|X_{n_{m+1}}(\omega) - X_{n_m}(\omega)| \leqslant 2^{-m}$. Alors la suite de réels $X_{n_m}(\omega)$ est de Cauchy. En effet, soit $\varepsilon > 0$ et $p > l$; on a

$$|X_{n_p}(\omega) - X_{n_l}(\omega)| \leqslant \sum_{l \leqslant m \leqslant p-1} |X_{n_{m+1}}(\omega) - X_{n_m}(\omega)| \leqslant \sum_{l \leqslant m \leqslant p-1} 2^{-m} \leqslant \sum_{m \geqslant l} 2^{-m} \leqslant 2^{-l+1}.$$

Puisque pour tout $p > l > l_0$, $2^{-l+1} \leqslant \varepsilon$, donc $|X_{n_p}(\omega) - X_{n_l}(\omega)| \leqslant \varepsilon$. Ainsi X_{n_k} converge p.s. vers une limite X et en particulier cette sous-suite converge en probabilité vers X. Pour conclure on écrit :

$$\mathrm{P}(|X_n - X| \geqslant \varepsilon) \leqslant \mathrm{P}\left(|X_n - X_{n_k}| \geqslant \frac{\varepsilon}{2}\right) + \mathrm{P}\left(|X_{n_k} - X| \geqslant \frac{\varepsilon}{2}\right)$$

et on fait petit chaque terme en utilisant la condition de Cauchy et la précédente convergence en proba de la sous-suite X_{n_k}. □

4.3 Convergence dans L^p

Définition 4.3 *Une suite de variables aléatoires réelles p-intégrables, $0 < p < \infty$, $(X_n)_{n\in\mathbb{N}}$* **converge dans** L^p *vers une variable p-intégrable X si*

$$\lim_{n\to\infty} \|X_n - X\|_p = 0 \Leftrightarrow \lim_{n\to\infty} \mathrm{E}\left(|X_n - X|^p\right) = 0.$$

Dans ce cas on note $X_n \xrightarrow{\mathrm{L}^p} X$, lorsque $n \to \infty$.

Proposition 4.7 *La convergence dans L^p implique la convergence en probabilité.*

Preuve. Par l'inégalité de Markov, pour tout $\varepsilon > 0$,

$$\mathrm{P}(|X_n - X| \geqslant \varepsilon) \leqslant \frac{\mathrm{E}(|X_n - X|^p)}{\varepsilon},\ p > 0,$$

d'où la conclusion. □

Remarque : La Proposition 4.4 justifie la terminologie pour la convergence L^0 pour la convergence en probabilité.

Remarque : La convergence en probabilité n'implique pas la convergence L^p comme le montre l'exemple suivant : soit l'espace de probabilité $(]0,1], \mathcal{B}(]0,1]), \lambda)$, soit $\alpha > 0$ et on considère

$$X_n(\omega) = \omega^{-\alpha} \mathbb{1}_{]-\infty, \frac{1}{n}]}(\omega),\ n \geqslant 1.$$

Pour tout $\varepsilon \in]0,1[$, $\mathrm{P}(|X_n| \geqslant \varepsilon) = \frac{1}{n}$, d'où $X_n \to 0$ en probabilité. Or, $X_n \notin \mathrm{L}^p$, dès que $\alpha p \geqslant 1$ car

$$\mathrm{E}(X_n^p) = \int_0^{\frac{1}{n}} \omega^{-\alpha p} d\omega.$$

Remarque : La convergence presque sûre n'implique pas la convergence L^p comme le montre l'exemple suivant : soit l'espace de probabilité $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathrm{P})$, avec P une probabilité, et soit X_n de loi $(1 - n^{-p})\delta_0 + n^{-p}\delta_n$, $p > 1$. Soit $\varepsilon > 0$; pour $n \geqslant \varepsilon^{-p}$, $\mathrm{P}(|X_n| \geqslant \varepsilon) = n^{-p}$. Par le lemme de Borel-Cantelli (Proposition 4.1) X_n converge presque sûrement vers 0, puisque $p > 1$. Toutefois, $\mathrm{E}(|X_n|^p) = 1$, pour tout n.

Définition 4.4 *Une suite de variables aléatoires intégrables* $(X_n)_{n\in\mathbb{N}}$ *est dite* **uniformément intégrable** *si*

$$\lim_{c\to\infty}\sup_{n\in\mathbb{N}} \mathrm{E}\left(|X_n|\mathbb{1}_{\{|X_n|>c\}}\right) = 0.$$

Remarque : Une famille de variables aléatoires avec un seul élément est uniformément intégrable par convergence dominée.
Une suite de variables aléatoires dominée par une variable Y intégrable, c'est-à-dire, si $|X_n| \leqslant Y$ pour tout n, alors la suite est uniformément intégrable :

$$\lim_{c\to\infty}\sup_{n\in\mathbb{N}} \mathrm{E}\left(|X_n|\mathbb{1}_{\{|X_n|>c\}}\right) \leqslant \lim_{c\to\infty} \mathrm{E}\left(Y\mathbb{1}_{\{Y>c\}}\right) = 0.$$

Enfin, toute suite finie de variables aléatoires intégrables est uniformément intégrable car elle dominée par la variable intégrable $\sum_{1\leqslant j\leqslant n}|X_j|$.
Soient $(X_n)_{n\in\mathbb{N}}$ et $(Y_n)_{n\in\mathbb{N}}$ deux suites de variables intégrables telles que $|X_n| \leqslant |Y_n|$. Alors si la suite Y_n est uniformément intégrable, alors la suite X_n est uniformément intégrable.
Si les variables X_n sont p-intégrables alors la suite $(|X_n|^p)_{n\in\mathbb{N}}$ est uniformément intégrable si cette suite est **bornée dans** $\mathrm{L}^{p+\delta}$:

$$\sup_n \mathrm{E}\left(|X_n|^{p+\delta}\right) < \infty.$$

En effet

$$\sup_n \mathrm{E}\left(|X_n|^p\mathbb{1}_{\{|X_n|^p>c\}}\right) = \sup_n \mathrm{E}\left(|X_n|^p\mathbb{1}_{\{|\frac{X_n}{c^{1/p}}|>1\}}\right) = \sup_n \mathrm{E}\left(|X_n|^p\cdot 1\cdot \mathbb{1}_{\{\frac{|X_n|^\delta}{c^{\delta/p}}>1\}}\right)$$

$$\leqslant \sup_n \mathrm{E}\left(|X_n|^p\frac{|X_n|^\delta}{c^{\delta/p}}\right) \leqslant c^{-\delta/p}\sup_n \mathrm{E}\left(|X_n|^{p+\delta}\right).$$

Remarque : Montrons que :

$$X \text{ intégrable } \Longrightarrow \forall\varepsilon>0,\ \exists\eta>0,\ \text{ t.q. si } A\in\mathcal{A} \text{ avec } \mathrm{P}(A)<\eta \text{ alors } \mathrm{E}\left(|X|\mathbb{1}_A\right)\leqslant\varepsilon.$$

D'une part, par le théorème de convergence dominée, pour c suffisament grand,

$$\mathrm{E}\left(|X|\mathbb{1}_{\{|X|>c\}}\right) \leqslant \frac{\varepsilon}{2}.$$

D'autre part, pour $\eta \leqslant \frac{\varepsilon}{2c}$,

$$\mathrm{E}\left(|X|\mathbb{1}_A\right) = \mathrm{E}\left(|X|\mathbb{1}_{A\cap\{|X|>c\}}\right) + \mathrm{E}\left(|X|\mathbb{1}_{A\cap\{|X|\leqslant c\}}\right) \leqslant \frac{\varepsilon}{2} + c\mathrm{P}(A) \leqslant \frac{\varepsilon}{2} + c\eta \leqslant \varepsilon.$$

Proposition 4.8 *La suite* $(X_n)_{n\in\mathbb{N}}$ *de variables aléatoires réelles intégrables est uniformément intégrable si et seulement si :*

a) $\forall\varepsilon>0,\ \exists\eta>0,$ *t.q. si* $A\in\mathcal{A}$ *avec* $\mathrm{P}(A)<\eta$ *alors* $\forall n\in\mathbb{N}\ \mathrm{E}\left(|X_n|\mathbb{1}_A\right)\leqslant\varepsilon$

et

b) $\sup_{n\in\mathbb{N}}\mathrm{E}(|X_n|)<\infty$.

Preuve : Supposons l'uniforme intégrabilité :

$$\forall \varepsilon > 0, \exists c > 0 \text{ t.q. } \sup_{n\in\mathbb{N}} \mathrm{E}\left(|X_n|\mathbb{1}_{\{|X_n|>c\}}\right) \leqslant \frac{\varepsilon}{2}.$$

Pour $A \in \mathcal{A}$, pour tout $n \in \mathbb{N}$,

$$\mathrm{E}(|X_n|\mathbb{1}_A) = \mathrm{E}\left(|X_n|\mathbb{1}_{A\cap\{|X_n|>c\}}\right) + \mathrm{E}\left(|X_n|\mathbb{1}_{A\cap\{|X_n|\leqslant c\}}\right) \leqslant \frac{\varepsilon}{2} + c\mathrm{P}(A).$$

On obtient *a)* en prenant $\eta = \frac{\varepsilon}{2c}$ et *b)* en prenant $A = \Omega$. Réciproquement, on fixe $\varepsilon > 0$ et $\eta > 0$ fournis par *a)*, on note $M = \sup_{n\in\mathbb{N}} \mathrm{E}(|X_n|)$ qui est fini par *b)* et on note $c_0 = \frac{M}{\eta}$. Par l'inégalité de Markov, pour tout $c \geqslant c_0$ et tout n, $\mathrm{P}(|X_n| > c) \leqslant \eta$. On applique *a)* à $A = \{|X_n| > c\}$ pour chaque n et on trouve $\mathrm{E}\left(|X_n|\mathbb{1}_{\{|X_n|>c\}}\right) \leqslant \varepsilon$, d'où la conclusion. □

Proposition 4.9 *Soit* $(X_n)_{n\in\mathbb{N}}$ *une suite de variables aléatoires intégrables. Il y a équivalence entre :*

1) $(X_n)_{n\in\mathbb{N}}$ *converge dans* L^1 *;*

2) $(X_n)_{n\in\mathbb{N}}$ *est une suite de Cauchy dans* L^1*, c'est-à-dire* $\mathrm{E}(|X_n - X_m|) \to 0$*, lorsque* $n, m \to \infty$ *;*

3) $(X_n)_{n\in\mathbb{N}}$ *est une suite uniformément intégrable et* $(X_n)_{n\in\mathbb{N}}$ *converge en probabilité.*

Remarque : Par ce résultat on voit que l'uniforme intégrabilité et la convergence en probabilité (ou la convergence p.s.) impliquent la convergence dans L^1. En particulier comme $|\mathrm{E}(X_n) - \mathrm{E}(X)| \leqslant \mathrm{E}(|X_n - X|)$, on déduit la convergence de la suite premier moment. On a ainsi, pour une suite dominée, un résultat de convergence dominée pour la convergence en probabilité. Enfin cette proposition montre que L^1 est un espace complet (voir Proposition 2.15).

Preuve :
1) $\Longrightarrow$ 2) :
La convergence dans L^1 implique la condition de Cauchy dans L^1 par l'inégalité triangulaire.
2) $\Longrightarrow$ 3) :
Par la condition de Cauchy, pour tout $\varepsilon > 0$, il existe $n_0(\varepsilon)$ tel que si $n \geqslant m \geqslant n_0$, $\mathrm{E}(|X_n - X_m|) \leqslant \frac{\varepsilon}{2}$. Alors, pour tout $A \in \mathcal{A}$ et tout $n \geqslant n_0$

$$\mathrm{E}\left(|X_n|\mathbb{1}_A\right) = \mathrm{E}\left(|X_n - X_{n_0} + X_{n_0}|\mathbb{1}_A\right) \leqslant \mathrm{E}\left(|X_{n_0}|\mathbb{1}_A\right) + \mathrm{E}(|X_n - X_{n_0}|) \leqslant \mathrm{E}\left(|X_{n_0}|\mathbb{1}_A\right) + \frac{\varepsilon}{2}.$$

Donc

$$\sup_{n\geqslant n_0} \mathrm{E}\left(|X_n|\mathbb{1}_A\right) \leqslant \mathrm{E}\left(|X_{n_0}|\mathbb{1}_A\right) + \frac{\varepsilon}{2}, \text{ d'où } \sup_{n\geqslant n_0} \mathrm{E}\left(|X_n|\mathbb{1}_A\right) \leqslant \sup_{j\leqslant n_0} \mathrm{E}\left(|X_j|\mathbb{1}_A\right) + \frac{\varepsilon}{2}.$$

Si $A = \Omega$,

$$\sup_{n\geqslant n_0} \mathrm{E}\left(|X_n|\mathbb{1}_A\right) \leqslant \sup_{j\leqslant n_0} \mathrm{E}\left(|X_j|\mathbb{1}_A\right) + \frac{\varepsilon}{2} < \infty,$$

donc la partie *b)* de la Proposition 4.9 est vérifiée. Par ailleurs la famille finie $\{X_j : j \leqslant n_0\}$ est uniformément intégrable, donc, pour tout $\varepsilon > 0$, il existe un $\eta > 0$ tel que si $\mathrm{P}(A) < \eta$, on ait $\sup_{j \leqslant n_0} \mathrm{E}(|X_j| \mathbb{1}_A) \leqslant \frac{\varepsilon}{2}$. On déduit alors

$$\sup_n \mathrm{E}(|X_n| \mathbb{1}_A) \leqslant \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon,$$

c'est-à-dire la partie *a)* de la Proposition 4.9. Donc la suite est uniformément intégrable. Enfin, par l'inégalité de Markov

$$\mathrm{P}(|X_n - X_m| \geqslant \varepsilon) \leqslant \frac{\mathrm{E}(|X_n - X_m|)}{\varepsilon} \to 0,\ n, m \to \infty,$$

donc la suite est de Cauchy en probabilité. Par la Proposition 4.7 la suite est convergente en probabilité.
3) $\Longrightarrow$ 1) :
Comme X_n converge en probabilité vers une variable X, de toute suite croissante (n') on peut extraire une sous-suite (n'_k) telle que $X_{n'_k}$ converge p.s. vers X (Proposition 4.5). Alors, par le lemme de Fatou

$$\mathrm{E}(|X|) = \mathrm{E}(\liminf_{k \to \infty} |X_{n'_k}|) \leqslant \liminf_{k \to \infty} \mathrm{E}(|X_{n'_k}|) \leqslant \sup_{n \in \mathbb{N}} \mathrm{E}(|X_n|) < \infty,$$

par le point *b)* de la Proposition 4.9, donc X est intégrable.
Par ailleurs, pour tout $\varepsilon > 0$,

$$\begin{aligned} \mathrm{E}(|X_n - X|) &\leqslant \mathrm{E}\left(|X_n - X| \mathbb{1}_{\{|X_n - X| < \varepsilon\}}\right) + \mathrm{E}\left(|X_n| \mathbb{1}_{\{|X_n - X| \geqslant \varepsilon\}}\right) + \mathrm{E}\left(|X| \mathbb{1}_{\{|X_n - X| \geqslant \varepsilon\}}\right) \\ &\leqslant \varepsilon + \mathrm{E}\left(|X_n| \mathbb{1}_{\{|X_n - X| \geqslant \varepsilon\}}\right) + \mathrm{E}\left(|X| \mathbb{1}_{\{|X_n - X| \geqslant \varepsilon\}}\right). \end{aligned}$$

La suite de variables aléatoires intégrables $\{X, X_1, X_2, \ldots, X_n, \ldots\}$ est encore uniformément intégrable. On fixe $\eta > 0$ et par la convergence en probabilité de la suite X_n, pour n suffisamment grand on a $\mathrm{P}(|X_n - X| \geqslant \varepsilon) \leqslant \eta$. En prenant dans le *a)* de la Proposition 4.9 $A = \{|X_n - X| \geqslant \varepsilon\}$, on voit que pour tout n suffisamment grand

$$\mathrm{E}\left(|X_n| \mathbb{1}_{\{|X_n - X| \geqslant \varepsilon\}}\right) \leqslant \varepsilon \ \text{ et } \ \mathrm{E}\left(|X| \mathbb{1}_{\{|X_n - X| \geqslant \varepsilon\}}\right) \leqslant \varepsilon.$$

On a obtenu $\mathrm{E}(|X_n - X|) \leqslant 3\varepsilon$, d'où la convergence dans L^1. □

Proposition 4.10 *Soit $p \geqslant 1$ et soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires dans L^p. Il y a équivalence entre :*

1) $(X_n)_{n \in \mathbb{N}}$ converge dans L^p ;

2) $(X_n)_{n \in \mathbb{N}}$ est une suite de Cauchy dans L^p, c'est-à-dire $\|X_n - X_m\|_p \to 0$, lorsque $n, m \to \infty$;

3) $(|X_n|^p)_{n \in \mathbb{N}}$ est une suite uniformément intégrable et $(X_n)_{n \in \mathbb{N}}$ converge en probabilité.

Remarque : Par ce résultat la preuve de la Proposition 2.15 est complète. Par ailleurs, pour assurer l'uniforme intégrabilité de $(|X_n|^p)_{n \in \mathbb{N}}$ il suffit de vérifier que la suite $(X_n)_{n \in \mathbb{N}}$ est bornée dans $\mathrm{L}^{p+\delta}$, $\delta > 0$.

Preuve. Cette preuve est similaire à celle de la Proposition 4.10. On donne les idées.
1) $\Longrightarrow$ 2) :
La condition de Cauchy dans L^p s'obtient de la convergence dans L^p par l'inégalité de Minkowski (Proposition 2.14).
2) $\Longrightarrow$ 3) :
Si $(X_n)_{n\in\mathbb{N}}$ est de Cauchy dans L^p elle est de Cauchy en probabilité (par l'inégalité de Markov), donc convergente en probabilité vers une variable aléatoire X. A nouveau par l'inégalité de Minkowski

$$|\|X_n\|_p - \|X_m\|_p| \leqslant \|X_n - X_m\|_p \to 0,\ n, m \to \infty.$$

Donc la suite réelle $(\|X_n\|_p)_{n\in\mathbb{N}}$ est de Cauchy, donc convergente, donc $\sup_n \|X_n\|_p < \infty$ (condition *b)* dans la Proposition 4.9). De plus par le lemme de Fatou on peut déduire aussi que la limite en probabilité satisfait aussi $\|X\|_p < \infty$, donc $X \in \mathrm{L}^p$. Enfin

$$\mathrm{E}\left(|X_n|^p \mathbb{1}_A\right) = \mathrm{E}\left(|X_n - X_m + X_m|^p \mathbb{1}_A\right) \leqslant 2^p \|X_n - X_m\|_p^p + 2^p \mathrm{E}\left(|X_m|^p \mathbb{1}_A\right).$$

Par la condition de Cauchy on fixe n_0 et on prend $n \geqslant n_0$ et par l'uniforme intégrabilité de $\{|X_{n_0}|\}$ on peut faire petite la quantité $2^p\mathrm{E}\left(|X_{n_0}|^p \mathbb{1}_A\right)$, dès que $\mathrm{P}(A)$ est petite. On trouve ainsi la condition *a)* dans la Proposition 4.9 et donc l'uniforme intégrabilité.
3) $\Longrightarrow$ 1) :
De toute suite croissante (n') on peut extraire une sous-suite (n'_k) telle que $X_{n'_k}$ converge p.s. vers une variable X. Par l'uniforme intégrabilité et le lemme de Fatou

$$\mathrm{E}(|X|^p) \leqslant \liminf_{k\to\infty} \mathrm{E}(|X_{n'_k}|^p) \leqslant \sup_n \mathrm{E}(|X_n|^p) < \infty,$$

donc $X \in \mathrm{L}^p$. Ensuite on fait comme dans la proposition précédente. □

4.4 Convergence en loi

On a vu que deux variables aléatoires X, Y ont la même loi lorsque $\mathrm{P}_X = \mathrm{P}_Y$, ou si et seulement si $F_X = F_Y$, ou si et seulement si, pour toute fonction continue bornée $g : \mathbb{R} \to \mathbb{R}$, $\mathrm{E}[g(X)] = \mathrm{E}[g(Y)]$, ou si et seulement si $\varphi_X = \varphi_Y$. Ces diverses égalités donnent lieu à diverses définitions de la convergence en loi.

Définition 4.5 (et théorème)
Une suite de variables aléatoires réelles $(X_n)_{n\in\mathbb{N}}$ **converge en loi** *vers* X *ou que la suite de lois* $(\mathrm{P}_{X_n})_{n\in\mathbb{N}}$ **converge étroitement** *vers* P_X *si, lorsque* $n \to \infty$ *:*

1) $\mathrm{E}[g(X_n)] \to \mathrm{E}[g(X)] \Leftrightarrow \int_{\mathbb{R}} g(x)\mathrm{P}_{X_n}(dx) \to \int_{\mathbb{R}} g(x)\mathrm{P}_X(dx)$ *pour toute fonction continue bornée* $g : \mathbb{R} \to \mathbb{R}$ *;*

2) $F_{X_n}(t) \to F_X(t)$ *en tout point de continuité* t *de* F_X *;*

3) $\varphi_{X_n}(t) \to \varphi_X(t)$ *pour tout* $t \in \mathbb{R}$.

On note alors

$$X_n \xrightarrow{loi} X \Longleftrightarrow \mathrm{P}_{X_n} \xrightarrow{e} \mathrm{P}_X$$

Proposition 4.11 *La convergence presque sûre implique la convergence en loi.*

Preuve. Ceci est une conséquence du théorème de convergence dominée et de la partie 1) de la définition. □

Proposition 4.12 *La convergence en probabilité implique la convergence en loi.*

Preuve. Soit g continue et bornée. Si X_n converge en proba vers X alors $g(X_n)$ converge en proba vers $g(X)$. De plus la suite $(g(X_n))_{n\in\mathbb{N}}$ est bornée donc u.i. D'ici on déduit la convergence dans L^1 de $g(X_n)$ vers $g(X)$ et enfin on obtient 1). □

Proposition 4.13 *La convergence dans* L^p *implique la convergence en loi.*

Preuve. On rappelle que la convergence dans L^p, entraîne la convergence en probabilité. Donc si X_n converge dans L^p, $p > 0$, vers X, alors X_n converge en loi vers X. □

Remarque : La convergence en loi n'implique ni la convergence presque sûre, ni la convergence en probabilité, comme le montre l'exemple suivant : soit $X \sim \mathcal{N}(0,1)$ et $X_n = (-1)^n X$. Alors, par la symétrie de la loi gaussienne standard, X_n à même loi que X. Donc X_n converge en loi vers X, mais ne converge ni presque sûrement vers X, ni en probabilité vers X.

Proposition 4.14 *Si la suite X_n converge en loi vers une constante c si et seulement si X_n converge en probabilité vers c.*

Preuve : Par 2) on a que, lorsque $n \to \infty$,

$$F_{X_n}(t) \to \begin{cases} 0 & \text{si } t < c \\ 1 & \text{si } t \geqslant c, \end{cases} \quad \forall t \neq c.$$

Pour tout $\varepsilon > 0$, on a

$$\mathrm{P}(|X_n - c| \leqslant \varepsilon) = \mathrm{P}(c - \varepsilon \leqslant X_n \leqslant c + \varepsilon)$$

$$\geqslant \mathrm{P}(c - \varepsilon < X_n \leqslant c + \varepsilon) = F_{X_n}(c+\varepsilon) - F_{X_n}(c-\varepsilon) \to 1 - 0 = 1$$

car $c + \varepsilon, c - \varepsilon \neq c$, d'où la conclusion. □

Remarque : Si X_n et Y_n convergent en loi vers X et Y, on ne peut rien dire sur la convergence du couple (X_n, Y_n). Dans l'exemple ci-dessus avec les lois gaussiennes, le couple (X, X_n), ne converge pas en loi. Toutefois on peut montrer que si si Y_n converge en loi vers une constante c, alors (X_n, Y_n) converge en loi vers (X, c) et en particulier $X_n + Y_n$ converge en loi vers $X + c$ et $X_n Y_n$ converge en loi vers cX.

Remarque : **(convergence en loi des variables aléatoires à valeurs entières)**
Soient X_n,$n \in \mathbb{N}$, et X à valeurs entières. Alors, lorsque $n \to \infty$

$$X_n \xrightarrow{\text{loi}} X \overset{(i)}{\Longleftrightarrow} \mathrm{P}(X_n = k) \longrightarrow \mathrm{P}(X = k), \forall\, k \in \mathbb{N} \overset{(ii)}{\Longleftrightarrow} G_{X_n}(s) \longrightarrow G_X(s) \;\text{ pour }\; s \in [0,1].$$

$\overset{(i)}{\Longrightarrow}$: en effet, pour $k \in \mathbb{N}$, soient s, t des points de continuité de F_X, tels que $k-1 < s < k < t < k+1$, d'où, lorsque $n \to \infty$,

$$\mathrm{P}(X_n = k) = F_{X_n}(t) - F_{X_n}(s) \longrightarrow F_X(t) - F_X(s) = \mathrm{P}(X = k).$$

$\overset{(i)}{\Longleftarrow}$: soit g une fonction continue à support compact dans $[0, N]$. On a

$$\mathrm{E}[g(X_n)] = \sum_{k=0}^{N} g(k)\mathrm{P}(X_n = k) \to \sum_{k=0}^{N} g(k)\mathrm{P}(X = k) = \mathrm{E}[g(X)]$$

$\overset{(ii)}{\Longrightarrow}$: on utilise le même argument que ci-dessus en ajoutant la convergence dominée.
$\overset{(ii)}{\Longleftarrow}$: on a besoin d'un résultat classique sur la convergence des fonctions holomorphes :
Soit une suite des fonctions holomorphes sur le disque unité $(f_n)_{n\in\mathbb{N}}$, ayant, pour chaque n, tous les coefficients de chaque série entière positifs, et qui converge simplement sur $[0,1]$; alors la suite converge simplement sur tout le disque unité et la limite f est une fonction holomorphe sur le disque unité ; de plus pour tout entier k, la k-ième dérivée de f_n converge simplement vers la k-ième dérivée de f (sur le disque ouvert).
On prend ici $f_n = G_{X_n}$ dont (la limite de) la k-ième dérivée au point 1 est $(k!)\mathrm{P}(X_n = k)$. On obtient la conclusion.

Preuve du Théorème 4.5 :
1) $\Longrightarrow$ 2) :

Pour tout $\varepsilon > 0$ il existe une fonction continue, $g_\varepsilon : \mathbb{R} \to [0,1]$ telle que $g_\varepsilon(x) = 1$ si $x \leqslant t$ et $g_\varepsilon(x) = 0$ si $x \geqslant t + \varepsilon$. On a

$$\lim_{n\to\infty} \mathrm{E}[g_\varepsilon(X_n)] = \mathrm{E}[g_\varepsilon(X)] \leqslant \mathrm{E}[\mathbb{1}_{]-\infty,t+\varepsilon]}(X)] = \mathrm{P}(X \leqslant t+\varepsilon) = F_X(t+\varepsilon)$$

et

$$\mathrm{E}[g_\varepsilon(X_n)] \geqslant \mathrm{E}[\mathbb{1}_{]-\infty,t]}(X_n)] = \mathrm{P}(X_n \leqslant t) = F_{X_n}(t),$$

donc $\limsup_{n\to\infty} F_{X_n}(t) \leqslant F_X(t+\varepsilon)$ et par continuité à droite, $\limsup_{n\to\infty} F_{X_n}(t) \leqslant F_X(t)$ en prenant ε arbitrairement petit. Soit ensuite $h_\varepsilon : \mathbb{R} \to [0,1]$ telle que $h_\varepsilon(x) = 1$ si $x \leqslant t-\varepsilon$ et $h_\varepsilon(x) = 0$ si $x \geqslant t$. On a

$$\lim_{n\to\infty} \mathrm{E}[h_\varepsilon(X_n)] = \mathrm{E}[h_\varepsilon(X)] \geqslant \mathrm{E}[\mathbb{1}_{]-\infty,t-\varepsilon]}(X)] = \mathrm{P}(X \leqslant t-\varepsilon) = F_X(t-\varepsilon)$$

et

$$\mathrm{E}[h_\varepsilon(X_n)] \leqslant \mathrm{E}[\mathbb{1}_{]-\infty,t]}(X_n)] = \mathrm{P}(X_n \leqslant t) = F_{X_n}(t),$$

donc $\liminf_{n\to\infty} F_{X_n}(t) \geqslant F_X(t-\varepsilon)$. Par hypothèse $F_X(t-) = F_X(t)$ et pour ε arbitairement petit on trouve $\liminf_{n\to\infty} F_{X_n}(t) \geqslant F_X(t)$.
2) $\Longrightarrow$ 1) :

Pas a)

Soit d'abord $g : \mathbb{R} \to \mathbb{R}$ de classe C^1 à support compact ; en particulier g' est bornée à support compact. Par le théorème de Fubini, on a

$$\mathrm{E}[g(X_n)] = \int_{\mathbb{R}} \mathrm{P}_{X_n}(dx) \int_{-\infty}^{x} g'(y)dy = \int_{\mathbb{R}} (1 - F_{X_n}(y))g'(y)dy.$$

On sait que l'ensemble des points de discontinuité de F_X est au plus dénombrable, donc de mesure de Lebesgue nulle ; hors de ces points $1 - F_{X_n}$ converge vers $1 - F_X$ tout en restant bornée par 1, car toute fonction de répartition est à valeurs dans [0,1]. La mesure positive $|g'(y)|dy$ a sa masse totale finie, c'est-à-dire :

$$\int_{\mathbb{R}} |g'(y)|dy \leqslant \sup_{y\in\mathrm{supp}(g')} |g'(y)| \times \lambda_{Leb}\left(\mathrm{supp}(g')\right) < \infty,$$

donc par convergence dominée,

$$\lim_{n\to\infty} \mathrm{E}[g(X_n)] = \lim_{n\to\infty} \int_{\mathbb{R}} (1 - F_{X_n}(y))g'(y)dy = \int_{\mathbb{R}} (1 - F_X(y))g'(y)dy = \mathrm{E}[g(X)].$$

Pas b)

Soit ensuite g continue à support compact, mais pas nécéssairement dérivable. Pour tout $\varepsilon > 0$, il existe une fonction h de classe C^1 à support compact telle que $|g(x) - h(x)| \leqslant \frac{\varepsilon}{4}$, pour tout $t \in \mathbb{R}$ (par exemple la convolée de g par l'approximation de l'unité de classe C^1 à support compact). Par ce qu'on vient de voir, on sait trouver un entier n_0, tel que $|\mathrm{E}[h(X_n)] - \mathrm{E}[h(X)]| \leqslant \frac{\varepsilon}{2}$ dès que $n \geqslant n_0$. Donc, pour $n \geqslant n_0$,

$$|\mathrm{E}[g(X_n)] - \mathrm{E}[g(X)]| \leqslant \mathrm{E}(|g(X_n) - h(X_n)|) + |\mathrm{E}[h(X_n)] - \mathrm{E}[h(X)]| + \mathrm{E}(|g(X) - h(X)|) \leqslant \varepsilon.$$

Pas c)

Enfin soit g continue bornée et $\varepsilon > 0$. On peut supposer que $|g| \leqslant 1$. Il existe une fonction $\breve{g}$ continue à support compact $0 \leqslant \breve{g} \leqslant 1$, telle que $\mathrm{E}[\breve{g}(X)] > 1 - \frac{\varepsilon}{5}$. On déduit qu'il existe un entier n_1 tel que $\mathrm{E}[\breve{g}(X_n)] > 1 - \frac{\varepsilon}{4}$ dès que $n \geqslant n_1$. La fonction $g\breve{g}$ est continue à support compact, donc il existe un entier n_2 tel que $|\mathrm{E}[(g\breve{g})(X_n)] - \mathrm{E}[(g\breve{g})(X)]| \leqslant \frac{\varepsilon}{2}$ dès que $n \geqslant n_2$. On a alors, pour tout $n \geqslant \max\{n_1, n_2\}$,

$$\begin{aligned}|\mathrm{E}[g(X_n)] - \mathrm{E}[g(X)]| &\leqslant |\mathrm{E}[(g\breve{g})(X_n)] - \mathrm{E}[(g\breve{g})(X)]| + |\mathrm{E}[g(1-\breve{g})(X_n)]| + |\mathrm{E}[g(1-\breve{g})(X)]| \\ &\leqslant \frac{\varepsilon}{2} + \mathrm{E}[(1-\breve{g})(X_n)] + \mathrm{E}[(1-\breve{g})(X)] \leqslant \varepsilon.\end{aligned}$$

Ceci termine la preuve de cette implication.

1) $\Longrightarrow$ 3) :

Il suffit de prendre $g_1(x) = \cos(tx)$ et $g_2(x) = \sin(tx)$:

$$\varphi_{X_n}(t) = \mathrm{E}[g_1(X_n)] + i\mathrm{E}[g_2(X_n)] \to \mathrm{E}[g_1(X)] + i\mathrm{E}[g_2(X)] = \varphi_X(t).$$

3) $\Longrightarrow$ 1) :

Pas a)

Supposons d'abord que $g : \mathbb{R} \to \mathbb{R}$ est continue à support compact, avec sa transformée de Fourier $\hat{g}(t) = \int_{\mathbb{R}} e^{itx} g(x)dx$ est dans $\mathrm{L}^1(\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$. On sait, par inversion de Fourier que $g(x) = \frac{1}{2\pi}\int_{\mathbb{R}} e^{-itx}\hat{g}(t)dt$. Par le théorème de Fubini,

$$\mathrm{E}[g(X_n)] = \frac{1}{2\pi}\mathrm{E}\left[\int_{\mathbb{R}} e^{-itX_n}\hat{g}(t)dt\right] = \frac{1}{2\pi}\int_{\mathbb{R}} \mathrm{E}\left[e^{-itX_n}\right]\hat{g}(t)dt = \frac{1}{2\pi}\int_{\mathbb{R}} \varphi_{X_n}(-t)\hat{g}(t)dt.$$

Mais, $\varphi_{X_n}(-t)$ converge vers $\varphi_X(-t)$ pour chaque t et de plus, on sait que $|\varphi_{X_n}(-t)| \leqslant 1$, pour tout $t \in \mathbb{R}$. Comme $\hat{g}$ est intégrable, par convergence dominée on trouve

$$\lim_{n\to\infty} \mathrm{E}[g(X_n)] = \frac{1}{2\pi}\int_{\mathbb{R}} \varphi_X(-t)\hat{g}(t)dt = \mathrm{E}[g(X)].$$

Pas b)
On peut voir que la condition $\hat{g} \in \mathrm{L}^1(\mathbb{R})$ est satisfaite pour toute fonction g de classe C^2 à support compact. En effet, la transformée de Fourier $\hat{g}$ est bornée et la transformée de Fourier de la fonction à support compact g'' est bornée et est égale à $t^2\hat{g}(t)$. D'où pour $|t| \to \infty$, $\hat{g}(t) = O(\frac{1}{t^2})$. Comme $\hat{g}$ est bornée, elle est donc intégrable sur $\mathbb{R}$. Ainsi l'affirmation est vérifiée pour g de classe C^2 à support compact.

Pas c)
Pour g juste continue à support compact et ensuite g continue bornée on peut répéter les mêmes arguments que dans la preuve 2) $\Longrightarrow$ 1) (les pas *b)* et *c)*).

La preuve du théorème est complète. $\square$

REMARQUE : *(facultatif)*
En fait l'implication 3) $\Longrightarrow$ 1) peut être énoncée comme suit :

Théorème 4.1 (de Lévy) *Pour chaque $n \geqslant 1$, φ_n est la fonction caractéristique d'une loi de probabilité sur $\mathbb{R}$, Q_n (ou d'une v.a.r. X_n). Supposons que $\lim_{n\to\infty} \varphi_n(t) =: \varphi(t)$ existe pour chaque $t \in \mathbb{R}$, et que φ est une fonction continue en $t = 0$. Alors φ est la fonction caractéristique d'une loi de probabilité Q sur $\mathbb{R}$ (ou d'une v.a.r. X) et Q_n converge étroitement vers Q (ou X_n converge en loi vers X) quand $n \to \infty$.*

Preuve :
1er pas. Soit $r_1, r_2, \ldots$ une énumération de $\mathbb{Q}$ et pour chaque $j \geqslant 1$ on considère la suite $\{F_n(r_j)\}_{n\geqslant 1}$ où F_n est la fonction de répartition de la loi Q_n, $F_n(t) = Q_n((-\infty, t])$. La suite $\{F_n(r_j)\}_n$ est bornée par 1, donc elle contient une sous-suite convergente. Par la procédure d'extraction diagonale on peut trouver une sous-suite $F_{n_k} =: G_k$ telle que la limite $\lim_{k\to\infty} G_k(r) =: b_r$ existe pour tout $r \in \mathbb{Q}$. Remarquons que si $r < r'$, alors $b_r \leqslant b_{r'}$, puisque F_{n_k} est croissante.

2ème pas. Nous allons reconstruire une fonction à partir du "squelette" b_r. On note $G(t) = \inf_{r>t} b_r$. Il est clair que si $t < t'$, alors $G(t) \leqslant G(t')$ donc G est croissante. Si $t_n \downarrow t$ et si $r > t$ alors à partir d'un certain rang $r > t_n$. On en déduit que $G(t) = \inf_n G(t_n)$ pour toute suite $t_n \downarrow t$ et ainsi G est continue à droite.

3ème pas. Montrons qu'en tout point de continuité t de G on a $\lim_{n\to\infty} G_n(t) = G(t)$. Soit $r \in \mathbb{Q}$, $r > t$. Alors $G_n(t) \leqslant G_n(r)$ et $\lim_{n\to\infty} G_n(r) = b_r$. On déduit que $\limsup_{n\to\infty} G_n(t) \leqslant b_r$ et comme cela est vrai pour tout $r \in \mathbb{Q}$ avec $r > t$, si on prend l'infimum pour $r > t$ on trouve $\limsup_{n\to\infty} G_n(t) \leqslant G(t)$. Soit maintenant $s < t$. On peut trouver un $r \in \mathbb{Q}$ tel que $s < r < t$. Alors $\liminf_{n\to\infty} G_n(t) \geqslant \liminf_{n\to\infty} G_n(r) = b_r \geqslant G(s)$. Comme cela est vrai pour chaque $s < t$, on en déduit $\liminf_{n\to\infty} G_n(t) \geqslant \sup_{s<t} G(s) = G(t-) = G(t)$ puisque t est un point de continuité de G. Remarquons qu'à ce stade G n'est pas nécéssairement une fonction de répartition. Exemple : soit $F_n(t) = \delta_n((-\infty, t])$ dont la limite ponctuelle $G(t)$ existe et est nulle pour tout t, mais $G \equiv 0$ n'est pas une fonction de répartition.

Lemme 4.1
Soient ψ et G respectivement la fonction caractéristique et la fonction de répartition d'une probabilité sur $\mathbb{R}$. Alors :

$$1 - G(\frac{2}{T}) + G(-\frac{2}{T}) \leqslant 2\left(1 - \frac{1}{2T}\int_{-T}^{T} \psi(t)dt\right), \quad \forall T > 0.$$

4ème pas. D'après le lemme, pour tout $k \geqslant 1$

$$1 - G_k(\frac{2}{T}) + G_k(-\frac{2}{T}) = 1 - F_{n_k}(\frac{2}{T}) + F_{n_k}(-\frac{2}{T}) \leqslant 2\left(1 - \frac{1}{2T}\int_{-T}^{T}\varphi_{n_k}(t)dt\right).$$

Il est possible de choisir T pour que $\pm\frac{2}{T}$ soient des points de continuité de G. On fait $k \to \infty$ et par convergence dominée au membre de droite de l'inégalité précédente on trouve

$$1 - G(\frac{2}{T}) + G(-\frac{2}{T}) \leqslant 2\left(1 - \frac{1}{2T}\int_{-T}^{T}\varphi(t)dt\right).$$

Comme $\varphi_n(0) = 1$, $\varphi(0) = 1$ étant continue en ce point. On laisse $T \to 0$ de telle façon que $\pm\frac{2}{T}$ restent des points de continuité de G. On déduit que $1 - G(\infty) + G(-\infty) = 0$. Avec les propriétés précédentes de G cela implique que G est une fonction de répartition.

5ème pas. On a donc trouvé que $\lim_{k\to\infty} G_k = \lim_{k\to\infty} F_{n_k} = G$ en tout point de continuité de G avec G fonction de répartition. Soit Q la loi (unique, par caractérisation à l'aide des fonctions de répartition) ayant G comme fonction de répartition. D'après la convergence en loi Q_{n_k} converge étroitement vers Q et φ_{n_k} converge ponctuellement vers φ_Q, lorsque $k \to \infty$. Ainsi $\varphi_Q = \varphi$. Mais cela reste vrai pour toute sous-suite de départ qui converge vers une fonction de répartition. Ainsi φ sera limite ponctuelle de toute sous-suite de φ_n. D'après la caractérisation des lois avec fonction caractéristique on conclut que Q_n converge étroitement vers Q et φ_n converge ponctuellement vers $\varphi = \varphi_Q$ lorsque $n \to \infty$. □

Preuve du Lemme : Soit X une v.a.r. de fonction caractéristique ψ et soit $T > 0$. D'après le théorème de Fubini :

$$\begin{aligned}\frac{1}{2T}\int_{-T}^{T}\psi(t)dt &= \mathrm{E}\left[\frac{1}{2T}\int_{-T}^{T}e^{itX}dt\right] = \mathrm{E}\left[\frac{\sin(TX)}{TX}\right] \leqslant \mathrm{E}\left|\frac{\sin(TX)}{TX}\right| \\ &\leqslant \mathrm{E}\left[\left|\frac{\sin(TX)}{TX}\right|\mathbb{1}_{|X|<\ell}\right] + \mathrm{E}\left[\left|\frac{\sin(TX)}{TX}\right|\mathbb{1}_{|X|\geqslant\ell}\right] \leqslant \mathrm{E}\left[\mathbb{1}_{|X|<\ell}\right] + \frac{1}{T\ell}\mathrm{E}\left[\mathbb{1}_{|X|\geqslant\ell}\right],\end{aligned}$$

où à la dernière inégalité on a utilisé les inégalités $|\sin x| \leqslant 1$ et $|\sin x| \leqslant x$. On déduit

$$\begin{aligned}1 - \frac{1}{2T}\int_{-T}^{T}\psi(t)dt &\geqslant 1 - \mathrm{E}\left[\mathbb{1}_{|X|<\ell}\right] - \frac{1}{T\ell}\mathrm{E}\left[\mathbb{1}_{|X|\geqslant\ell}\right] \\ &= (1 - \frac{1}{T\ell})\mathrm{E}\left[\mathbb{1}_{|X|\geqslant\ell}\right] \geqslant (1 - \frac{1}{T\ell})[1 - G(\ell) + G(-\ell)].\end{aligned}$$

Il suffit de prendre $\ell = \frac{2}{T}$ pour conclure. □

A présent on peut démontrer le théorème de Bochner (cf. §2.5, p.53) :

Théorème 4.2 (de Bochner) *Soit $\varphi : \mathbb{R} \to \mathbb{C}$ une fonction continue en 0 et telle que $\varphi(0) = 1$. On suppose que φ est de type positif :*

$$\forall n \geqslant 1, \forall z_1 \dots, z_n \in \mathbb{C}, \forall t_1, \dots t_n \in \mathbb{R}, \quad \sum_{j,k=1}^{n}\varphi(t_j - t_k)z_j\overline{z_k} \geqslant 0.$$

Alors φ est la fonction caractéristique d'une probabilité (ou d'une v.a.) sur $\mathbb{R}$.

Preuve : L'idée est de construire des approximations φ_n de φ qui sont des fonctions caractéristiques et telles que $\lim_{n\to\infty} \varphi_n(t) = \varphi(t)$, pour tout $t \in \mathbb{R}$. Ensuite on applique le théorème de Lévy précédent.

1er pas. Nous établissons d'abord quelques propriétés simples des fonctions de type positif :

a) Si φ est de type positif, la fonction $t \mapsto \varphi(t)e^{iat}$ l'est aussi pour tout réel a. Il s'agit d'une vérification directe et élémentaire de la définition.

b) Si une famille φ_j de fonctions sont de type positif, alors toute combinaison linéaire à coefficients positifs $a_j \geqslant 0$, $\varphi = \sum_j a_j \varphi_j$ est de type positif. Si chaque fonction satisfait $\varphi_j(0) = 1$ et si $\sum_j a_j = 1$ alors $\varphi(0) = 1$ aussi (combinaison convexe).

c) Si φ est de type positif, alors $\varphi(0) \geqslant 0$, $\varphi(-t) = \overline{\varphi(t)}$ et $|\varphi(t)| \leqslant \varphi(0)$ pour tout t. Nous allons utiliser le fait que la matrice $(\varphi(t_j - t_k))_{1\leqslant j,k\leqslant n}$ est hermitienne, positive $\forall t_1, \ldots, t_n$ réels. La première assertion s'obtient à partir de la positivité de $\varphi(0)|z|^2$, la deuxième est une conséquence du caractère hermitien et enfin la troisième s'obtient en prenant $n = 2$, $t_1 = t$ et $t_2 = 0$ comme conséquence de la positivité de la matrice : $|\varphi(t)|^2 \leqslant |\varphi(0)|^2$.

d) On a $\forall s, t \in \mathbb{R}$ $|\varphi(t) - \varphi(s)|^2 \leqslant 4\varphi(0)|\varphi(0) - \varphi(t-s)|$. On prend la matrice positive $(\varphi(t_j - t_k))_{1\leqslant j,k\leqslant n}$ avec $n = 3$, $t_1 = t$, $t_2 = s$ et $t_3 = 0$:

$$\begin{pmatrix} 1 & \varphi(t-s) & \varphi(t) \\ \overline{\varphi(t-s)} & 1 & \varphi(s) \\ \overline{\varphi(t)} & \overline{\varphi(s)} & 1 \end{pmatrix}.$$

En particulier son déterminant est positif donc

$$\begin{aligned} 0 &\leqslant 1 + \varphi(s)\varphi(t-s)\overline{\varphi(t)} + \overline{\varphi(s)\varphi(t-s)}\varphi(t) - |\varphi(s)|^2 - |\varphi(t)|^2 - |\varphi(t-s)|^2 \\ &= 1 - |\varphi(s) - \varphi(t)|^2 - |\varphi(t-s)|^2 - \varphi(t)\overline{\varphi(s)}(1 - \varphi(t-s)) - \overline{\varphi(t)}\varphi(s)(1 - \overline{\varphi(t-s)}) \\ &\leqslant 1 - |\varphi(s) - \varphi(t)|^2 - |\varphi(t-s)|^2 + 2|1 - \varphi(t-s)| \end{aligned}$$

On déduit

$$|\varphi(s) - \varphi(t)|^2 \leqslant 1 - |\varphi(t-s)|^2 + 2|1 - \varphi(t-s)| \leqslant 4|1 - \varphi(t-s)|.$$

e) D'après le point précédent on déduit qu'une fonction de type positif continue en $t = 0$ est continue partout (et en fait uniformément continue).

2ème pas. Montrons que si φ est de type positif, continue sur $\mathbb{R}$ et intégrable, alors

$$f(x) := \frac{1}{2\pi} \int_{\mathbb{R}} e^{-itx} \varphi(t) dt \geqslant 0$$

est continue et $\int_{\mathbb{R}} f(x)dx = 1$. De plus $F(t) := \int_{-\infty}^{t} f(x)dx$ est une fonction de répartition de fonction caractéristique :

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

Comme φ est intégrable, alors f est clairement bornée et continue. Montrons qu'elle est positive. D'après la convergence dominée et un changement de variables on a

$$\begin{aligned} f(x) = \lim_{T\to\infty} \frac{1}{2\pi}\int_{-T}^{T}(1-\frac{|t|}{T})e^{-itx}\varphi(t)dt &= \lim_{T\to\infty}\frac{1}{2\pi T}\int_0^T\int_0^T e^{-i(t-s)x}\varphi(t-s)dt\,ds \\ &= \lim_{T\to\infty}\frac{1}{2\pi T}\int_0^T\int_0^T e^{-itx}e^{isx}\varphi(t-s)dt\,ds \geqslant 0. \end{aligned}$$

où la dernière inégalité peut s'obtenir en approchant les intégrales avec des sommes de Riemann et en utilisant le fait que φ est de type positif. Il reste à montrer la formule d'inversion. On définit

$$f_\sigma(x) := f(x)\exp(-\frac{\sigma^2x^2}{2}).$$

D'après le théorème de Fubini, on a, pour tout $t\in\mathbb{R}$,

$$\begin{aligned} \int_{\mathbb{R}} e^{itx}f_\sigma(x)dx = \int_{\mathbb{R}} e^{itx}f(x)\exp(-\frac{\sigma^2x^2}{2}) &= \frac{1}{2\pi}\int_{\mathbb{R}}\int_{\mathbb{R}} e^{itx}\varphi(s)e^{-isx}\exp(-\frac{\sigma^2x^2}{2})ds\,dx \\ &= \int_{\mathbb{R}}\varphi(s)\frac{1}{\sqrt{2\pi\sigma^2}}\exp(-\frac{(t-s)^2}{2\sigma^2})ds. \end{aligned}$$

Si on prend $t=0$ dans cette dernière égalité on trouve :

$$\int_{\mathbb{R}} f_\sigma(x)dx = \int_{\mathbb{R}}\varphi(s)\frac{1}{\sqrt{2\pi\sigma^2}}\exp(-\frac{s^2}{2\sigma^2})ds \leqslant 1.$$

Nous allons maintenant faire $\sigma\to 0$. On a $f_\sigma \geqslant 0$ et $\lim_{\sigma\to 0} f_\sigma = f$. D'après le lemme de Fatou utilisée dans l'inégalité précédente on déduit que f est intégrable et $\int_{\mathbb{R}} f(x)dx \leqslant 1$. On va faire maintenant $\sigma\to 0$ dans une des égalités ci-dessus :

$$\int_{\mathbb{R}} e^{itx}f_\sigma(x)dx = \int_{\mathbb{R}}\varphi(s)\frac{1}{\sqrt{2\pi\sigma^2}}\exp(-\frac{(t-s)^2}{2\sigma^2})ds.$$

Au membre de gauche $x\mapsto e^{it}f_\sigma(x)$ est dominée par la fonction intégrable f. Au membre de droite on fait d'abord un changement de variable et ensuite la limite se calcule aisément :

$$\int_{\mathbb{R}} e^{itx}f(x)dx = \lim_{\sigma\to 0}\int_{\mathbb{R}}\varphi(t+\sigma s)\frac{1}{\sqrt{2\pi}}\exp(-\frac{s^2}{2})ds = \varphi(t)$$

c'est-à-dire la formule d'inversion.

3ème pas. Soit enfin φ une fonction de type positif continue en 0, $\varphi(0)=1$. Elle est donc continue sur $\mathbb{R}$ et ensuite $t\mapsto\varphi(t)e^{itx}$ est aussi de type positif et continue, pour tout $x\in\mathbb{R}$. Si $\sigma>0$ cela sera encore vrai pour la combinaison convexe suivante :

$$\varphi_\sigma(t) = \int_{\mathbb{R}}\varphi(t)e^{itx}\frac{1}{\sqrt{2\pi\sigma^2}}\exp(-\frac{x^2}{2\sigma^2})dx = \varphi(t)\exp(-\frac{\sigma^2t^2}{2}).$$

Il n'est pas difficile de voir que la fonction φ_σ est intégrable. Le pas précédent s'applique donc à la fonction φ_σ. Si on fait $\sigma\to 0$ on déduit par le théorème de Lévy précédent que φ est aussi une fonction caractéristique. $\square$

REMARQUE : Le théorème de Lévy repose sur une question intéressante : savoir si d'une suite de variables aléatoires on peut extraire une sous suite convergente en loi (on dira que la suite est **relativement compacte** pour la convergence étroite).

Ce n'est pas toujours le cas : soit la suite de variables aléatoires $X_n = n$. Pour toute fonction g continue tendant vers 0 à l'infini on a $\mathrm{E}[g(X_n)] = g(n) \to 0$, donc on ne peut pas extraire une sous suite convergente en loi. Il s'agit du fait que "la masse est partie à l'infini". Lorsqu'on évite ce phénomène alors la suite est relativement compacte et réciproquement.

La formulation du phénomène de ne pas perdre de masse est la suivante :
Une suite de lois P_n *sur* $\mathbb{R}$ *est dite* **tendue** *si*

$$\lim_{M\to\infty} \sup_{n\in\mathbb{N}} \mathrm{P}_n \left(\{x \in \mathbb{R} : |x| \geqslant M\}\right) = 0,$$

ou encore, pour tout $\varepsilon > 0$, *il existe un* $M > 0$ *tel que*

$$\mathrm{P}_n \left(\{x \in \mathbb{R} : |x| \geqslant M\}\right) \leqslant \varepsilon, \ \forall n \in \mathbb{N}.$$

Le résultat important est donné par le **théorème de Prohorov** :
Une suite de lois de probabilités est relativement compacte (pour la convergence étroite) si et seulement si elle est tendue. □

4.5 Les lois des grands nombres et le théorème central limite

Soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires réelles indépendantes et de même loi (i.i.d.) qu'une variable notée X. Pour tout entier $n \geqslant 1$, on $S_n = \sum_{j=1}^{n} X_j$ et on va s'intéresser aux propriétes asymptotiques de S_n, ou plus précisement de $\frac{S_n}{n}$, la **moyenne empirique** des X_j. De point de vue pratique, par exemple, si X_j modélise le nombre de pile montre au lancer d'une pièce, $\frac{S_n}{n}$ est la proportion de lancers montrant pile dans n essais. On peut voir sur cet exemple que $\frac{S_n}{n}$ converge vers $\mathrm{E}(X) = \mathrm{P}(X = 1)$ et que la loi de $(S_n - \mathrm{E}(S_n))/\sqrt{n}$ ressemble à une loi normale lorsque n est assez grand.

Le premier résultat montre que l'intuition est correcte : si la pièce est non truquée la proportion de pile tend à se stabiliser vers 0,5 ; donc la théorie construite constitue une modélisation raisonable du réel et qu'une certaine régularité apparaît dans les phénomènes aléatoires.

Théorème 4.3 *(loi faible de grands nombres)*
Soit $(X_n)_{n\in\mathbb{N}}$ *une suite de variables aléatoires i.i.d. avec* $\mathrm{E}(|X|) < \infty$ *et soit* $S_n = X_1 + \ldots + X_n$. *Alors, lorsque* $n \to \infty$,

$$\frac{S_n}{n} \xrightarrow{prob} \mathrm{E}(X).$$

Preuve : Comme $X \in \mathrm{L}^1$, la fonction caractéristique φ_X est dérivable et on sait que $\varphi_X(0) = 1$ et $\varphi'_X(0) = i\mathrm{E}(X)$, d'où son développement,

$$\varphi_X(t) = 1 + it\mathrm{E}(X) + o(t), \ \text{ lorsque } \ t \to 0.$$

Comme les variables $X_1, \dots, X_n$ sont indépendantes et de même loi que X, la fonction caractéristique de S_n est $\varphi_{S_n} = \varphi_X^n$. Donc lorsque $n \to \infty$,

$$\varphi_{\frac{S_n}{n}}(t) = \varphi_{S_n}(\frac{t}{n}) = \varphi_X(\frac{t}{n})^n = \left(1 + it\mathrm{E}(X)\frac{1}{n} + o(\frac{t}{n})\right)^n.$$

Par convergence dominée, on voit que, lorsque $n \to \infty$,

$$n\,|\varphi_X(\frac{t}{n}) - (1 + i\frac{t}{n}\mathrm{E}(X))| \to 0.$$

En effet on a[1]

$$|\varphi_X(\frac{t}{n}) - (1 + i\frac{t}{n}\mathrm{E}(X))| \leqslant \mathrm{E}\left[\frac{t^2|X|^2}{2n^2} \wedge \frac{2t|X|}{n}\right],$$

avec

$$n\left(\frac{t^2|X|^2}{2n^2} \wedge \frac{2t|X|}{n}\right) \leqslant 2t|X| \in \mathrm{L}^1 \text{ et } n\left(\frac{t^2|X|^2}{2n^2} \wedge \frac{2t|X|}{n}\right) \leqslant \frac{t^2|X|^2}{2n} \to 0.$$

Mais alors[2], lorsque $n \to \infty$,

$$|\varphi_X(\frac{t}{n})^n - (1 + \frac{it\mathrm{E}(X)}{n})^n| \leqslant n|\varphi_X(\frac{t}{n}) - (1 + i\frac{t}{n}\mathrm{E}(X))| \to 0.$$

On en déduit

$$\lim_{n\to\infty} \varphi_{\frac{S_n}{n}}(t) = e^{it\mathrm{E}(X)},$$

---

[1] Par la formule de Taylor au reste integral on a

$$e^{ix} = \sum_{k=0}^{n} \frac{(ix)^k}{k!} + \frac{i^{n+1}}{n!}\int_0^x (x-s)^n e^{is} ds,$$

d'où

$$|e^{ix} - \sum_{k=0}^{n} \frac{(ix)^k}{k!}| \leqslant \frac{x^{n+1}}{(n+1)!}.$$

Par ailleurs, un calcul simple montre que

$$e^{ix} - \sum_{k=0}^{n} \frac{(ix)^k}{k!} = \frac{i^n}{(n-1)!}\int_0^x (x-s)^{n-1} e^{is} ds - \frac{(ix)^n}{n!},$$

d'où

$$|e^{ix} - \sum_{k=0}^{n} \frac{(ix)^k}{k!}| \leqslant \frac{|x|^n}{n!} + \frac{|x|^n}{n!} = \frac{2|x|^n}{n!}.$$

Ainsi

$$|e^{ix} - \sum_{k=0}^{n} \frac{(ix)^k}{k!}| \leqslant \frac{x^{n+1}}{(n+1)!} \wedge \frac{2|x|^n}{n!}.$$

Enfin on déduit

$$|\varphi_X(t) - \sum_{k=0}^{n} \frac{(it)^k}{k!}\mathrm{E}(X^k)| \leqslant \mathrm{E}\left[\frac{|tX|^{n+1}}{(n+1)!} \wedge \frac{2|tX|^n}{n!}\right].$$

Il suffit de prendre n=1 et n=2 (pour la preuve du Théorème 4.3 ci-dessous).

[2] On a pour $i = 1, \dots, n$ et $a_i, b_i \in \mathbb{C}$, $|a_i|, |b_i| \leqslant 1$

$$\left|\prod_{i=1}^n a_i - \prod_{i=1}^n b_i\right| \leqslant \sum_{i=1}^n |a_i - b_i|$$

(preuve par récurrence).

où le membre droit est la fonction caractéristique de la variable aléatoire constante $\mathrm{E}(X)$. Par la Définition 4.5 on déduit que $\frac{S_n}{n}$ converge en loi vers la constante $\mathrm{E}(X)$, et par la Proposition 4.14 on tire la conclusion. $\square$

Théorème 4.4 *(loi forte de grands nombres)*
Soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires i.i.d. et soit $S_n = X_1 + \ldots + X_n$. Les deux conditions suivantes sont équivalentes :

1) $\mathrm{E}(|X|) < \infty$ *;*

2) $\frac{S_n}{n} \xrightarrow{p.s.} \mathrm{E}(X)$*, lorsque $n \to \infty$.*

Preuve :
2) $\Longrightarrow$ 1) :

Si la suite $\frac{S_n}{n}$ converge p.s., alors la suite $\frac{X_n}{n}$ converge p.s. vers 0. Par le lemme de Borel-Cantelli (Proposition 4.1, 2)), puisque les X_j sont indépendantes et toutes de même loi que X, pour tout (ou seulement un) $\varepsilon > 0$,

$$\sum_{n\geqslant 1} \mathrm{P}(|\frac{X_n}{n}| > \varepsilon) = \sum_{n\geqslant 1} \mathrm{P}(|X_n| > \varepsilon n) = \sum_{n\geqslant 1} \mathrm{P}(|X| > \varepsilon n) < \infty.$$

On conclut à l'aide de la Proposition 2.8.
1) $\Longrightarrow$ 2) : Notons $\mathrm{E}(X) = m$.

Pas a)

On prouve dans un premier temps le résultat sous l'hypothèse plus restrictive que X admet un moment d'ordre 2 : $\mathrm{E}(|X|^2) < \infty$. En considérant séparément la partie positive et sa partie négative X_n, on remarque qu'il suffit de démontrer le résultat pour des variables aléatoires positives, ce qu'on suppose désormais. On rappelle que $\mathrm{E}(S_n) = nm$ et $\mathrm{Var}(S_n) = n\mathrm{Var}(X)$. D'où,

$$\mathrm{Var}(\frac{S_{n^4}}{n^4} - m) = \frac{\mathrm{Var}(X)}{n^4},$$

et donc, par l'inégalité de Tchebytchev,

$$\mathrm{P}(|\frac{S_{n^4}}{n^4} - m| \geqslant \frac{1}{n}) \leqslant \frac{\mathrm{Var}(X)}{n^2}.$$

On déduit que la série $\sum_{n\geqslant 1} \mathrm{P}(|\frac{S_{n^4}}{n^4} - m| \geqslant \frac{1}{n})$ converge. Par le lemme de Borel-Cantelli on a presque sûrement,

$$|\frac{S_{n^4}}{n^4} - m| \leqslant \frac{1}{n}, \text{ sauf pour un nombre fini d'entiers } n.$$

On fixe un ω pour lequel l'événement ci-dessus est réalisé et on prend n suffisament grand. Comme la suite S_n est croissante, pour tout entier k tel que $n^4 \leqslant k \leqslant (n+1)^4$, on a

$$\frac{S_k(\omega)}{k} \leqslant \frac{S_{(n+1)^4}(\omega)}{n^4} = \left(\frac{n+1}{n}\right)^4 \frac{S_{(n+1)^4}(\omega)}{(n+1)^4} \leqslant \left(\frac{n+1}{n}\right)^4 (m + \frac{1}{n+1}).$$

Ainsi,

$$\limsup_{k\to\infty} \frac{S_k(\omega)}{k} \leqslant m.$$

De même
$$\frac{S_k(\omega)}{k} \geqslant \frac{S_{n^4}(\omega)}{(n+1)^4} = \left(\frac{n}{n+1}\right)^4 \frac{S_{n^4}(\omega)}{n^4} \geqslant \left(\frac{n}{n+1}\right)^4 (m - \frac{1}{n}).$$
On déduit
$$\liminf_{k\to\infty} \frac{S_k(\omega)}{k} \geqslant m.$$
En conséquence, on a pour presque tout ω, $\lim_{k\to\infty} \frac{S_k(\omega)}{k} = m$.

Pas b)

Supposons maintenant X intégrable. Quitte à remplacer les variables X_j par $X_j - \mathrm{E}(X_j)$, on peut supposer que $m = \mathrm{E}(X) = 0$. On fixe $\varepsilon > 0$. Comme $\mathrm{E}(|X|) < \infty$, il existe des variables Y_j, bornées, centrées indépendantes et de même loi, telles que $\mathrm{E}(|X_j - Y_j|) \leqslant \varepsilon$. [3]

On note $T_n = Y_1 + \ldots + Y_n$ et on a
$$\frac{|S_n|}{n} \leqslant \frac{1}{n}\sum_{j=1}^n |X_j - Y_j| + \frac{|T_n|}{n}.$$

D'après le pas *a)*, $\frac{|T_n|}{n}$ converge presque sûrement vers 0. On va montrer que pour ε suffisament petit,

$$(*) \qquad \limsup_{n\to\infty} \frac{1}{n}\sum_{j=1}^n |X_j - Y_j| \leqslant 2\mathrm{E}(|X_1 - Y_1|),$$

d'où on tire, par le choix des variables Y_j,

$$\limsup_{n\to\infty} \frac{|S_n|}{n} \leqslant \limsup_{n\to\infty} \frac{1}{n}\sum_{j=1}^n |X_j - Y_j| + \limsup_{n\to\infty} \frac{|T_n|}{n} \leqslant 2\varepsilon$$

presque sûrement. Comme ε est arbitraire, on conclut. On note $Z_j = |X_j - Y_j|$ variables aléatoires positives, indépendantes, de même loi qu'une variable Z intégrable et on veut borner
$$\limsup_{n\to\infty} \frac{1}{n}\sum_{j=1}^n Z_j.$$
On a pour tout entier k et tout $\delta > 0$,

$$\mathrm{P}\left(\max_{2^k < n \leqslant 2^{k+1}} \frac{1}{n}\sum_{j=1}^n Z_j \geqslant 2\mathrm{E}(Z) + \delta\right)$$

$$\leqslant \mathrm{P}\left(\exists j \in \{1, \ldots, 2^{k+1}\} : Z_j > 2^k\right) + \mathrm{P}\left(\max_{2^k < n \leqslant 2^{k+1}} \frac{1}{n}\sum_{j=1}^n Z_j \mathbb{1}_{[0,2^k]}(Z_j) \geqslant 2\mathrm{E}(Z) + \delta\right)$$

$$\leqslant 2^{k+1}\mathrm{P}(Z > 2^k) + \mathrm{P}\left(\sum_{j=1}^{2^{k+1}} Z_j \mathbb{1}_{[0,2^k]}(Z_j) \geqslant 2^{k+1}\mathrm{E}(Z) + \delta 2^k\right)$$

[3]Il existe $c > 0$ assez grand tel que $\mathrm{E}(|X|\mathbb{1}_{\{|X|>c\}}) \leqslant \varepsilon$. On prend $Y' = X\mathbb{1}_{\{|X|\leqslant c\}}$ et $Y = Y' - \mathrm{E}(Y')$ qui est centrée bornée par $2c$. Alors $\mathrm{E}(|X - Y|) \leqslant \mathrm{E}(|X - Y'|) + \mathrm{E}(Y') = \mathrm{E}(|X - Y'|) + |\mathrm{E}(X) - \mathrm{E}(Y')| \leqslant 2\mathrm{E}(|X - Y'|) \leqslant 2\varepsilon$. On prend $\varepsilon/2$ au lieu de ε et ensuite on fait des copies indépendantes de même loi de Y.

$$\leqslant 2^{k+1}\mathrm{P}(Z > 2^k) + \mathrm{P}\left(\sum_{j=1}^{2^{k+1}} \left(Z_j \mathbb{1}_{[0,2^k]}(Z_j) - \mathrm{E}(Z_j \mathbb{1}_{[0,2^k]}(Z_j))\right) \geqslant \delta 2^k\right).$$

On applique l'inégalité de Tchebytchev au deuxième terme du majorant précédent et on trouve

$$\mathrm{P}\left(\max_{2^k < n \leqslant 2^{k+1}} \frac{1}{n}\sum_{j=1}^{n} Z_j \geqslant 2\mathrm{E}(Z) + \delta\right) \leqslant 2^{k+1}\mathrm{P}(Z > 2^k) + \frac{1}{\delta^2 2^{2k}} 2^{k+1}\mathrm{E}\left(Z^2 \mathbb{1}_{[0,2^k]}(Z)\right)$$

$$\leqslant 2^{k+1}\mathrm{P}(Z > 2^k) + \frac{2}{\delta^2 2^k}\mathrm{E}\left(Z^2 \mathbb{1}_{[0,2^k]}(Z)\right)$$

Pour tout entier k,

$$\int_{2^k}^{2^{k+1}} \mathrm{P}(Z > t)dt \geqslant 2^k \mathrm{P}(Z > 2^{k+1}),$$

donc la Proposition 2.8 implique

$$\sum_{k \geqslant 0} 2^{k+1}\mathrm{P}(Z > 2^k) \leqslant 4\mathrm{E}(Z) < \infty.$$

De plus, si $2^l < Z \leqslant 2^{l+1}$, alors

$$Z^2 \sum_{k \geqslant 1} 2^{-k} \mathbb{1}_{[0,2^k]}(Z) \leqslant 2^{2l+2} \sum_{k \geqslant l+1} 2^{-k} \leqslant 4Z,$$

d'où

$$\sum_{k \geqslant 0} 2^{-k}\mathrm{E}\left(Z^2 \mathbb{1}_{[0,2^k]}(Z)\right) = \mathrm{E}\left(Z^2 \sum_{k \geqslant 0} 2^{-k} \mathbb{1}_{[0,2^k]}(Z)\right) \leqslant 4\mathrm{E}(Z).$$

Donc, finalement,

$$\sum_{k \geqslant 0} \mathrm{P}\left(\max_{2^k < n \leqslant 2^{k+1}} \frac{1}{n}\sum_{j=1}^{n} Z_j \geqslant 2\mathrm{E}(Z) + \delta\right) \leqslant 4(1 + \frac{2}{\delta^2})\mathrm{E}(Z) < \infty.$$

D'après le lemme de Borel-Cantelli (Théorème 3.3 a)), presque sûrement pour tout k assez grand

$$\max_{2^k < n \leqslant 2^{k+1}} \frac{1}{n}\sum_{j=1}^{n} Z_j < 2\mathrm{E}(Z) + \delta,$$

et comme δ est arbitraire on trouve l'inégalité (*). □

Remarque : Une autre façon d'énoncer la loi forte des grands nombres est de dire que si $\mathrm{E}(|X|) < \infty$ alors $\frac{S_n}{n} = \mathrm{E}(X) + o(1)$ p.s. lorsque $n \to \infty$. Le théorème limite central donne, en un certain sens, un terme de plus dans le développement asymptotique, précisant le comportement limite en loi du terme $o(1)$, sous une hypothèse suplémentaire sur la loi commune des X_j. On a ainsi une estimation trés précise de l'erreur commise en approchant l'espérance mathématique par la moyenne empirique. Le résultat permet d'approximer la loi de $\frac{S_n}{n}$ lorsque n est grand.

Théorème 4.5 *(central limite)*
Soit $(X_n)_{n\in\mathbb{N}}$ *une suite de variables aléatoires réelles i.i.d. avec* $\mathrm{E}(X^2) < \infty$ *et soit* $S_n = X_1 + \ldots + X_n$*. Alors, lorsque* $n \to \infty$,

$$\frac{S_n - n\mathrm{E}(X)}{\sqrt{n\mathrm{Var}(X)}} \xrightarrow{loi} \mathcal{N}(0,1).$$

Preuve : En remplaçant X_j par $(X_j - \mathrm{E}(X_j))/\sqrt{\mathrm{Var}(X_j)}$ on peut suposer que $\mathrm{E}(X) = 0$ et $\mathrm{Var}(X) = 1$. Comme $X \in \mathrm{L}^2$ sa fonction caractéristique est deux fois dérivable et son développement est

$$\varphi_X(t) = 1 + t\varphi'_X(0) + \frac{t^2}{2}\varphi''_X(0) + o(t^2) = 1 - \frac{t^2}{2} + o(t^2), \text{ lorsque } t \to 0.$$

On rappelle que par indépendance et équidistribution on a

$$\varphi_{S_n/\sqrt{n}}(t) = \varphi_X(t/\sqrt{n})^n = \left(1 - \frac{t^2}{2n} + o(\frac{t^2}{n})\right)^n, \text{ lorsque } n \to \infty.$$

Par convergence dominée, on voit que, lorsque $n \to \infty$,

$$|\varphi_X(\frac{t}{\sqrt{n}}) - (1 + i\frac{t}{\sqrt{n}}\mathrm{E}(X) - \frac{t^2}{n}\mathrm{E}(X^2))| \to 0.$$

En effet on a

$$|\varphi_X(\frac{t}{\sqrt{n}}) - (1 + i\frac{t}{\sqrt{n}}\mathrm{E}(X) - \frac{t^2}{n}\mathrm{E}(X^2))| \leqslant \mathrm{E}\left[\frac{t^3|X|^3}{6n^{3/2}} \wedge \frac{t^2|X|^2}{n}\right],$$

avec

$$n\left(\frac{t^3|X|^3}{6n^{3/2}} \wedge \frac{t^2|X|^2}{n}\right) \leqslant t^2|X|^2 \in \mathrm{L}^1 \text{ et } n\left(\frac{t^3|X|^3}{6n^{3/2}} \wedge \frac{t^2|X|^2}{n}\right) \leqslant \frac{t^3|X|^3}{6\sqrt{n}} \to 0.$$

A nouveau, lorsque $n \to \infty$

$$|\varphi_X(t/\sqrt{n})^n - (1 - \frac{t^2}{2n})^n| \leqslant n|\varphi_X(\frac{t}{\sqrt{n}}) - (1 + i\frac{t}{\sqrt{n}}\mathrm{E}(X) - \frac{t^2}{n}\mathrm{E}(X^2))| \to 0.$$

On en déduit

$$\lim_{n\to\infty} \varphi_{S_n/\sqrt{n}}(t) = e^{-t^2/2}.$$

Le membre de droite est la fonction caractéristique de la loi $\mathcal{N}(0,1)$ et le théorème est établi.
□

Proposition 4.15 *(vitesse de convergence en théorème central limite)*
Soit $(X_n)_{n\in\mathbb{N}}$ *une suite de variables aléatoires réelles i.i.d. avec* $\mathrm{E}(|X|^3) < \infty$*. On note* $\mathrm{E}(X) = m$, $\mathrm{Var}(X) = \sigma^2$, $S_n = X_1 + \ldots + X_n$ *et* $T_n = (S_n - nm)/\sqrt{n\sigma^2}$ *et* $N \sim \mathcal{N}(0,1)$*. Soit* $g : \mathbb{R} \to \mathbb{R}$ *une fonction de classe* C^3 *dont la dérivée troisième est bornée. Alors, lorsque* $n \to \infty$,

$$\mathrm{E}[g(T_n)] - \mathrm{E}[g(N)] = O(\frac{1}{\sqrt{n}}).$$

Preuve. Sans perte de généralité on peut supposer $m = 0$. On introduit une suite $(N_n)_{n\in\mathbb{N}}$ de variables aléatoires i.i.d. de loi $\mathcal{N}(0,1)$ et on pose $\bar{T}_n = (N_1 + \ldots + N_n)/\sqrt{n}$ de sorte que pour chaque n, $\bar{T}_n$ a encore pour loi $\mathcal{N}(0,1)$. Pour chaque $k \in \{1,\ldots,n\}$, on considère

$$T_n^k = \frac{X_1 + \ldots + X_k + N_{k+1} + \ldots + N_n}{\sqrt{n}} \quad \text{et} \quad \hat{T}_n^k = \frac{X_1 + \ldots + X_{k-1} + N_{k+1} + \ldots + N_n}{\sqrt{n}}.$$

On cherche donc à estimer

$$|\mathrm{E}[g(T_n)] - \mathrm{E}[g(\bar{T}_n)]| = \left|\sum_{k=0}^{n-1} \mathrm{E}[g(T_n^k)] - \mathrm{E}[g(T_n^{k+1})]\right|$$

$$\leqslant \sum_{k=0}^{n-1} \left|\mathrm{E}\left(g(T_n^k) - g(\hat{T}_n^k) + g(\hat{T}_n^k) - g(T_n^{k+1})\right)\right|.$$

Pour chaque différence on écrit le développement de Taylor de g au point $\hat{T}_n^k$. On voit que $T_n^k = \hat{T}_n^k + X_k/\sqrt{n}$ et $T_n^{k+1} = \hat{T}_n^k + N_k/\sqrt{n}$:

$$g(T_n^k) - g(\hat{T}_n^k) = g'(\hat{T}_n^k)\frac{X_k}{\sqrt{n}} + \frac{1}{2}g''(\hat{T}_n^k)\left(\frac{X_k}{\sqrt{n}}\right)^2 + R_n^k$$

et

$$g(T_n^{k+1}) - g(\hat{T}_n^k) = g'(\hat{T}_n^k)\frac{N_k}{\sqrt{n}} + \frac{1}{2}g''(\hat{T}_n^k)\left(\frac{N_k}{\sqrt{n}}\right)^2 + \hat{R}_n^k.$$

On sait que $\hat{T}_n^k$ est indépendante de N_k et aussi de X_k et que ces deux dernieres variables sont centrées et ont la même variance. Quand on prend la somme des deux développements et l'espérance, tout s'annule sauf les restes, $\mathrm{E}(R_n^k)$ et $\mathrm{E}(\hat{R}_n^k)$. Or

$$|R_n^k| \leqslant (\sup_{x\in\mathbb{R}} |g'''(x)|)\frac{|X_k|^3}{n^{3/2}} \quad \text{et} \quad |\hat{R}_n^k| \leqslant (\sup_{x\in\mathbb{R}} |g'''(x)|)\frac{|N_k|^3}{n^{3/2}}.$$

On trouve

$$|\mathrm{E}[g(T_n)] - \mathrm{E}[g(\bar{T}_n)]| \leqslant c\sum_{k=0}^{n-1} \frac{1}{n^{3/2}} = c\frac{1}{n^{1/2}}.$$

□

4.6 Quelques applications

Théorème 4.6 *(d'approximation de Weierstrass avec les polynômes de Bernstein)* *Soit* $f : [0,1] \to \mathbb{R}$ *une fonction continue et on définit les polynômes de Bernstein*

$$B_n(x) = \sum_{k=0}^{n} f\left(\frac{k}{n}\right)\mathrm{C}_n^k x^k (1-x)^{n-k}, \quad 0 \leqslant x \leqslant 1.$$

Alors la suite $(B_n)_{n\in\mathbb{N}}$ *converge uniformément vers* f*, lorsque* $n \to \infty$*.*

Preuve. Soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires i.i.d. de loi $\mathcal{B}(1,x)$, $x \in [0,1]$, et on pose $S_n = \sum_{j=1}^n X_j \sim \mathcal{B}(n,x)$. Il est alors facile de voir que

$$\mathrm{E}[f(\frac{S_n}{n})] = B_n(x).$$

D'après la loi faible des grands nombres

$$\frac{S_n}{n} \stackrel{\text{prob}}{\longrightarrow} \mathrm{E}(X) = x$$

d'où, comme f est continue

$$f(\frac{S_n}{n}) \stackrel{\text{prob}}{\longrightarrow} f(x)$$

et, par convergence dominée, car f est bornée,

$$B_n(x) = \mathrm{E}[f(\frac{S_n}{n})] \to f(x), \forall x \in [0,1].$$

Pour obtenir la convergence uniforme nous avons besoin d'introduire le module de continuité de f

$$\rho(\delta) = \sup_{|x-y|<\delta, 0\leqslant x,y\leqslant 1} |f(x) - f(y)|.$$

Il est connu que f est uniformément continue si et seulement si $\lim_{\delta\to 0}\rho(\delta) = 0$, ce qui est le cas dans la situation présente, puisque f est continue sur le compact $[0,1]$. Nous allons noter la norme uniforme :

$$\|f\|_u := \sup_{0\leqslant x\leqslant 1} |f(x)|.$$

On peut écrire

$$\|B -_n - f\|_u = \sup_{0\leqslant x\leqslant 1} |\mathrm{E}[f(\frac{S_n}{n}) - f(x)]| \leqslant \sup_{0\leqslant x\leqslant 1} \mathrm{E}[|f(\frac{S_n}{n}) - f(x)|]$$

$$\leqslant \sup_{0\leqslant x\leqslant 1} \mathrm{E}[|f(\frac{S_n}{n}) - f(x)|\mathbb{1}_{|\frac{S_n}{n}-x|\leqslant\varepsilon}] + \sup_{0\leqslant x\leqslant 1} \mathrm{E}[|f(\frac{S_n}{n}) - f(x)|\mathbb{1}_{|\frac{S_n}{n}-x|>\varepsilon}]$$

$$\leqslant \rho(\varepsilon)\mathrm{P}(|\frac{S_n}{n} - x| \leqslant \varepsilon) + 2\|f\|_u \sup_{0\leqslant x\leqslant 1} \mathrm{P}(|\frac{S_n}{n} - x| > \varepsilon) \leqslant \rho(\varepsilon) + 2\|f\|_u \sup_{0\leqslant x\leqslant 1} \frac{\mathrm{Var}(\frac{S_n}{n})}{\varepsilon^2}$$

$$\leqslant \rho(\varepsilon) + 2\|f\|_u \sup_{0\leqslant x\leqslant 1} \frac{x(1-x)}{n\varepsilon^2} \leqslant \rho(\varepsilon) + \frac{\|f\|_u}{2n\varepsilon^2}.$$

On conclut en prenant les limites supérieures lorsque $n \to \infty$ et ensuite lorsque $\varepsilon \to 0$. $\square$

Théorème 4.7 *(Glivenko-Cantelli)*
Soit $\{X_n : n \geqslant 1\}$ une suite de variables aléatoires réelles indépendantes et de même loi Q et de fonction de répartition

$$F(t) = Q(]-\infty, t]),\ t \in \mathbb{R}.$$

Pour tout $n \geqslant 1$ on introduit la fonction de répartition empirique

$$\hat{F}_n(t) :== \frac{1}{n}\mathrm{card}\{i : 1 \leqslant i \leqslant n, X_i \leqslant t\} = \frac{1}{n}\sum_{i=1}^n \mathbb{1}_{]-\infty,t]}(X_i),\ t \in \mathbb{R}.$$

Alors, lorsque $n \to \infty$,

$$\sup_{t\in\mathbb{R}} |\hat{F}_n(t) - F(t)| \xrightarrow{p.s.} 0.$$

Preuve. Les variables aléatoires $\mathbb{1}_{]-\infty,t]}(X_i) \sim \mathcal{B}(1, F(t))$ sont indépendantes de même loi. Alors, d'après la loi forte des grands nombres

$$\hat{F}_n(t) \xrightarrow{\text{p.s.}} F(t), \forall t \in \mathbb{R}.$$

Il s'agit d'une convergence ponctuelle, pour l'instant. Remarquons aussi que, toujours d'après la loi forte des grands nombres

$$\hat{F}_n(t-) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{]-\infty,t[}(X_i) \xrightarrow{\text{p.s.}} Q(]-\infty, t[) = F(t-), \forall t \in \mathbb{R}.$$

On trouve ainsi Ω'_t, Ω''_t événements de probabilité 1, tels que si $n \to \infty$ on a

$$\hat{F}_n(t, \omega) \to F(t),\ \forall \omega \in \Omega'_t$$

et

$$\hat{F}_n(t-, \omega) \to F(t-)\,\forall \omega \in \Omega''_t.$$

Notons $G(u) = \inf\{t : F(t) \geqslant u\}$, pour $0 < u < 1$, l'inverse généralisée à droite de F. On a

$$F(G(u)-) \leqslant u \leqslant F(G(u)).$$

Pour $m \geqslant 1$ et $1 \leqslant k \leqslant m$, on pose $t_k^m := G(\frac{k}{m})$. Si $t_{k-1}^m \leqslant t < t_k^m$, alors, par monotonie,

$$F(t_{k-1}^m) \leqslant F(t) \leqslant F(t_k^m-) \text{ et } \hat{F}_n(t_{k-1}^m) \leqslant \hat{F}_n(t) \leqslant \hat{F}_n(t_k^m-).$$

Pour des tels t,

$$\hat{F}_n(t_{k-1}^m, \omega) - F(t_k^m-) \leqslant \hat{F}_n(t, \omega) - F(t) \leqslant \hat{F}_n(t_k^m-, \omega) - F(t_{k-1}^m).$$

Mais

$$F(t_k^m-) - F(t_{k-1}^m) \leqslant \frac{1}{m},$$

ainsi que

$$F(t_1^m-) \leqslant \frac{1}{m},\ F(t_m^m) \geqslant 1 - \frac{1}{m}.$$

Donc

$$\hat{F}_n(t_{k-1}^m, \omega) - F(t_{k-1}^m) - \frac{1}{m} \leqslant \hat{F}_n(t, \omega) - F(t) \leqslant \hat{F}_n(t_k^m-, \omega) - F(t_k^m-) + \frac{1}{m},$$

ou encore

$$\sup_{t\in[t_{k-1}^m, t_k^m[} |\hat{F}_n(t, \omega) - F(t)| \leqslant \max\{|\hat{F}_n(t_{k-1}^m, \omega) - F(t_{k-1}^m)|, |\hat{F}_n(t_k^m-, \omega) - F(t_k^m-)|\} + \frac{1}{m}.$$

On note $D_n(\omega) = \sup_{t\in\mathbb{R}} |\hat{F}_n(t, \omega) - F(t)|$ et

$$D_n^m(\omega) = \max_{k=1,\ldots,m} \max\{|\hat{F}_n(t_k^m, \omega) - F(t_k^m)|, |\hat{F}_n(t_k^m-, \omega) - F(t_k^m-)|\}.$$

Alors

$$\max_{k=2,\dots,m} \sup_{t\in[t^m_{k-1},t^m_k[} |\hat{F}_n(t,\omega) - F(t)| \leqslant D^m_n(\omega) + \frac{1}{m},$$

Donc

$$\sup_{t\in[t^m_1,t^m_m[} |\hat{F}_n(t,\omega) - F(t)| \leqslant D^m_n(\omega) + \frac{1}{m}$$

Par des arguments similaires pour $t < t^m_1$ et $t \geqslant t^m_m$ on trouve

$$D_n(\omega) \leqslant D^m_n(\omega) + \frac{1}{m}.$$

Lorsque $\omega \in \Omega_m := \bigcap_k \Omega'_{t^m_k} \cap \bigcap_k \Omega''_{t^m_k}$, on a $\lim_{n\to\infty} D^m_n(\omega) = 0$, d'où on déduit que $\limsup_{n\to\infty} D_n(\omega) \leqslant \frac{1}{m}$. De plus il est facile de voir que Ω_m est de probabilité 1. Enfin, pour $\omega \in \bar{\Omega} = \bigcap_m \Omega_m$ on déduit $\lim_{n\to\infty} D_n(\omega) = 0$ avec $\bar{\Omega}$ de probabilité 1, d'où la conclusion. □

Théorème 4.8 *Soit $\{X_n : n \geqslant 0\}$ une suite de variables aléatoires réelles indépendantes et de même loi Q. Supposons que X est une variable aléatoire de loi Q qui admet des moments d'ordre deux. Lorsque $n \to \infty$,*

$$\bar{X} \xrightarrow{p.s.} \mathrm{E}(X),\ S^2 \xrightarrow{p.s.} \mathrm{Var}(X),$$

où

$$\bar{X} := \frac{1}{n}\sum_{i=1}^n X_i$$

est la moyenne empirique et

$$S^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})^2 = \frac{1}{n}\sum_{i=1}^n X_i^2 - \bar{X}^2$$

est la variance empirique, et où

$$\mathrm{E}(X) = \int xQ(dx),\ \mathrm{Var}(X) = \int x^2 Q(dx) - \left(\int xQ(dx)\right)^2.$$

Preuve. C'est une application directe de la loi forte des grands nombres. □

Théorème 4.9 *Soient X_1, X_2,...,X_n, n v.a. réelles, indépendantes, de même loi gaussienne $\mathcal{N}(m,\sigma^2)$. Alors*

1. *$\bar{X}_n$ et S^2 sont indépendantes.*
2. *$\bar{X}_n$ suit une loi $\mathcal{N}(m, \frac{\sigma^2}{n})$ et $\frac{n}{\sigma^2} S^2$ a pour loi $\chi^2(n-1)$.*

Preuve. 1) La première étape consiste à vérifier que l'on peut se ramener au cas où $m = 0$ et $\sigma = 1$. On pose

$$X'_i = \frac{X_i - m}{\sigma} \Longleftrightarrow X_i = \sigma X'_i + m \quad 1 \leqslant i \leqslant n. \tag{4.1}$$

Les v.a. X'_1, X'_2,...,X'_n sont indépendantes, chacune de loi $\mathcal{N}_1(0,1)$. On note $\bar{X}'_n$ et S'^2_n la moyenne et la variance empirique de cette nouvelle suite. On déduit de (4.1),

$$\bar{X}_n = \sigma\bar{X}'_n + m. \tag{4.2}$$

Par conséquent, $X_i - \bar{X}_n = \sigma(X'_i - \bar{X}'_n)$. On en déduit

$$S^2 = \sigma^2 S'^2. \tag{4.3}$$

Donc si le théorème est réalisé lorsque $m = 0$ et $\sigma = 1$, d'après (4.2) et (4.3) ce théorème l'est encore dans le cas général.

2) On suppose que chaque v.a. X_i est de loi $\mathcal{N}_1(0,1)$. Soit u_1 le vecteur de $\mathbb{R}^n$ de coordonnées $\frac{1}{\sqrt{n}},\frac{1}{\sqrt{n}},...,\frac{1}{\sqrt{n}}$. Ce vecteur est de norme 1, il existe donc u_2,u_3,...,u_n de $\mathbb{R}^n$ tels que $\mathcal{B} = \{u_1, u_2, ..., u_n\}$ soit une base orthonormée de $\mathbb{R}^n$. Soit A la matrice carrée d'ordre n, dont les colonnes sont u_1,u_2,...,u_n. Par construction A est une matrice orthogonale : $AA^* = A^*A = \mathrm{Id}$. On note $Y = A^*X$, et $Y^* = (Y_1, Y_2, ..., Y_n)$. Il est alors clair que Y est un vecteur gaussien centré de matrice de covariance K_Y :

$$K_Y = A^*K_X(A^*)^* = A^*\mathrm{Id}A = A^*A = \mathrm{Id},$$

les v.a. X_1,X_2,...,X_n étant indépendantes, réduites et centrées la matrice de covariance K_X de X est l'identité. Puisque la première ligne de A^* est $u_1^* = \left(\frac{1}{\sqrt{n}}, \frac{1}{\sqrt{n}}, , ..., \frac{1}{\sqrt{n}}\right)$, on a :

$$Y_1 = \frac{1}{\sqrt{n}}(X_1 + X_2 + ... + X_n) = \sqrt{n}\bar{X}_n. \tag{4.4}$$

On va exprimer S^2 à l'aide de Y.

$$\begin{aligned} nS^2 &= \sum_{k=1}^{n}(X_k - \bar{X}_n)^2 = \sum_{k=1}^{n}(X_k^2 - 2X_k\bar{X}_n + \bar{X}_n^2) \\ &= \left(\sum_{k=1}^{n} X_k^2\right) - 2\bar{X}_n\left(\sum_{k=1}^{n} X_k\right) + n\bar{X}_n^2. \end{aligned}$$

Par conséquent,

$$nS^2 = \left(\sum_{k=1}^{n} X_k^2\right) - 2\bar{X}_n(n\bar{X}_n) + n\bar{X}_n^2 = \left(\sum_{k=1}^{n} X_k^2\right) - n\bar{X}_n^2. \tag{4.5}$$

Mais $Y = A^*X$, A^* est une matrice orthogonale, elle conserve la norme : $\sum_{k=1}^n Y_k^2 = \sum_{k=1}^n X_k^2$. On déduit de (4.4) et (4.5) : $nS^2 = \sum_{k=1}^n Y_k^2 - Y_1^2 = \sum_{k=2}^n Y_k^2$. Y est un vecteur gaussien de matrice de covariance identité, les v.a. Y_1,Y_2,...,Y_n sont donc indépendantes et ont pour loi $\mathcal{N}_1(0,1)$, on en déduit l'indépendance de $\bar{X}_n$ et S^2. De plus $\bar{X}_n \sim \mathcal{N}_1(0, 1/n)$, et d'après un calcul de loi simple la loi de nS^2 est $\chi^2(n-1)$. □

Théorème 4.10 *(de Cochran)*
Soit G un vecteur gaussien $G \sim \mathcal{N}_d(0, \sigma^2\mathrm{I}_d)$. Soit une décomposition orthogonale de $\mathbb{R}^d = V_1 \oplus \ldots \oplus V_k$, $dimV_i = d_i$, $i = 1, \ldots, k$. Alors les vecteurs $\Pi_{V_1}(G), \ldots, \Pi_{V_k}(G)$ sont gaussiens indépendants et

$$\frac{\|\Pi_{V_i}(G)\|^2}{\sigma^2} \sim \chi^2(d_i),\ i = 1, \ldots, k.$$

Preuve. Soit (e_{ij}) une base orthonormée de $\mathbb{R}^d$ adaptée à la décomposition $\mathbb{R}^d = V_1 \oplus \ldots \oplus V_k$, c'est-à-dire $(e_{ij} : j = 1, \ldots, d_i)$ est une base orthonormée de V_i. Soit $G = \sum_{i,j} (G, e_{ij}) e_{ij}$ la décomposition de G dans cette nouvelle base. On pose $\eta_{ij} = (G, e_{ij})/\sigma$ et on obtient un vecteur gaussien $\eta = (\eta_{ij})_{i,j} \sim \mathcal{N}_d(0, \mathrm{I}_d)$. Mais

$$\Pi_{V_i}(G) = \sum_{j=1}^{d_i} (G, e_{ij}) e_{ij} = \sigma \sum_{j=1}^{d_i} \eta_{ij} e_{ij}$$

ne dépend que des η_{ij} pour $j \in \{1, \ldots, d_i\}$. On en déduit que les vecteurs $\Pi_{V_i}(G)$ sont indépendants et que

$$\frac{\|\Pi_{V_i}(G)\|^2}{\sigma^2} = \sum_{j=1}^{n_i} \eta_{ij}^2$$

suit la loi $\chi^2(d_i)$. □

On observe une variable aléatoire discrète X susceptible de prendre r valeurs $a_1, \ldots, a_r \in \mathbb{R}^d$ (le support de la loi est donc supposé fini et connu). On notera $\mathbf{p} = (p_1, \ldots, p_r)$ le vecteur des probabilités de chaque valeur possible :

$$p_j = \mathrm{P}(X = a_j) = Q(\{a_j\}) \in]0, 1[,\ j = 1, \ldots, r.$$

Soit par ailleurs une probabilité $Q_0 = \sum_{j=1}^r \pi_j \delta_{a_j}$ de même support mais avec les probabilités $\pi = (\pi_1, \ldots, \pi_r)$ connues. La question est $Q = Q_0$? En d'autres termes $\mathbf{p} = \pi$ vus comme vecteurs dans $\mathbb{R}^r$? Ainsi on veut tester

$$(\mathrm{H}) : Q = Q_0 \ \text{ contre } \ (\mathrm{A}) : Q \neq Q_0.$$

On observe un n-échantillon $\mathbf{X} = (X_1, \ldots, X_n)$ de loi Q (n v.a. i.i.d.) et on compte le nombre de fois qu'on rencontre la valeur a_j, $j = 1, \ldots, r$:

$$N_j = \sum_{i=1}^n \mathbb{1}_{\{X_i = a_j\}} \text{ le nombre des } X_i \text{ égaux à } a_j.$$

On peut voir que $\mathbf{N} = (N_1, \ldots, N_r)^*$ a la loi multimomiale $\mathcal{M}(n, p_1, \ldots, p_r)$:

$$\mathrm{P}(\mathbf{N} = \mathbf{n}) = \mathrm{P}((N_1, \ldots, N_r) = (n_1, \ldots, n_r)) = \frac{n!}{n_1! \ldots n_r!} p_1^{n_1} \ldots p_r^{n_r},\ n_1 + \ldots + n_r = n.$$

Soient

$$\hat{p}_j = \frac{N_j}{n},\ j = 1, \ldots, r \quad \text{(les fréquences empiriques).}$$

Par la loi forte des grands nombres on sait que lorsque $n \to \infty$,

$$\hat{p}_j \to p_j = \mathrm{E}(\mathbb{1}_{\{X = a_j\}}) = \mathrm{P}(X = a_j), \ \text{ p.s.}$$

Pour dire si $\mathbf{p} = \pi$ il faut regarder la distance entre ces deux vecteurs. Si on prend n grand il suffira de regarder la distance entre le vecteur des fréquences $\mathbf{f} = (f_1, \ldots, f_r) = (\frac{n_1}{n}, \ldots, \frac{n_r}{n})$ et $(\pi_1, \ldots, \pi_r)$.

Pearson a introduit la “distance” suivante (pas symétrique)

$$\rho(\mathbf{f}, \pi) = \sum_{j=1}^{r} \frac{(f_j - \pi_j)^2}{\pi_j}.$$

Nous avons à étudier la quantité :

$$T_n = n\rho(\hat{\mathbf{p}}, \pi) = n \sum_{j=1}^{r} \frac{(\hat{p}_j - \pi_j)^2}{\pi_j} = \sum_{j=1}^{r} \frac{(N_j - n\pi_j)^2}{n\pi_j}$$

qui doit être proche de zéro sous l'hypothèse (H).

Il est clair que sous l'hypothèse (A), c'est-à-dire si $\mathbf{p} \neq \pi$, $\rho(\hat{\mathbf{p}}, \pi) \to \rho(\mathbf{p}, \pi)$ p.s. donc dans ce cas $T_n \to \infty$ p.s.

Théorème 4.11 *Soient $\pi_1, \dots, \pi_r > 0$. Alors, sous l'hypothèse* (H) *la suite T_n converge en loi, quand $n \to \infty$ vers la loi $\chi^2(r-1)$. Sous l'hypothèse* (A) *la suite T_n tend presque sûrement vers ∞.*

Preuve. On voit que $\mathbf{N} = \sum_{i=1}^n \mathbf{Z}_i$ où

$$\mathbf{Z}_i = \begin{pmatrix} Z_{i,1} \\ \vdots \\ Z_{i,r} \end{pmatrix} = \begin{pmatrix} \mathbb{1}_{\{X_i = a_1\}} \\ \vdots \\ \mathbb{1}_{\{X_i = a_r\}} \end{pmatrix}.$$

sont i.i.d.

Sous (H) l'espérance commune est π et la matrice de covariance est Γ, avec

$$\Gamma_{jk} = \mathrm{Cov}(\mathbf{Z}_1)_{jk} = \mathrm{Cov}(Z_{1,j}, Z_{1,k}) = \delta_{jk}\pi_j - \pi_j\pi_k = \begin{cases} -\pi_j\pi_k, & \text{si } j \neq k \\ \pi_j(1-\pi_j), & \text{si } j = k. \end{cases}$$

Alors, par le théorème central limite, lorsque $n \to \infty$,

$$\frac{1}{\sqrt{n}}(\mathbf{N} - n\pi) \to G \sim \mathcal{N}_r(0, \Gamma) \text{ en loi.}$$

On note D la matrice diagonale $\mathrm{diag}(\frac{1}{\sqrt{\pi_1}}, \dots, \frac{1}{\sqrt{\pi_r}})$, donc

$$D\frac{1}{\sqrt{n}}(\mathbf{N} - n\pi) \to DG \text{ en loi.}$$

Par calcul

$$D\frac{1}{\sqrt{n}}(\mathbf{N} - n\pi) = \begin{pmatrix} \frac{N_1 - n\pi_1}{\sqrt{n\pi_1}} \\ \vdots \\ \frac{N_r - n\pi_r}{\sqrt{n\pi_r}} \end{pmatrix}$$

et

$$DG \sim \mathcal{N}_r(0, D\Gamma D),$$

où la matrice

$$\Sigma = D\Gamma D = (\delta_{jk} - \sqrt{\pi_j \pi_k})_{jk} = I_r - (\sqrt{\pi})(\sqrt{\pi})^*, \text{ et } (\sqrt{\pi}) = \begin{pmatrix} \sqrt{\pi_1} \\ \vdots \\ \sqrt{\pi_r} \end{pmatrix}.$$

On peut constater que $\|(\sqrt{\pi})\|^2 = 1$, d'où $\Sigma^2 = \Sigma^* = \Sigma$ est une matrice orthogonale et il s'agit de la matrice de la projection orthogonale sur $V := \text{Vect}(\sqrt{\pi})^{\perp}$. De plus si $\mathbf{Y} \sim \mathcal{N}_r(0, \mathrm{I}_r)$, alors $\Sigma\mathbf{Y} \sim \mathcal{N}_r(0, \mathrm{I}_r)$.

Ainsi, sous (H), lorsque $n \to \infty$,

$$\begin{pmatrix} \frac{N_1 - n\pi_1}{\sqrt{n\pi_1}} \\ \vdots \\ \frac{N_r - n\pi_r}{\sqrt{n\pi_r}} \end{pmatrix} \to \Sigma\mathbf{Y} \text{ en loi,}$$

d'où, par le théorème de Cochran,

$$n\rho(\hat{\mathbf{p}}, \pi) \to \|\Sigma\mathbf{Y}\|^2 = \|\Pi_V(\mathbf{Y})\|^2 \sim \chi^2(r-1) \text{ en loi,}$$

car $\dim V = r - 1$. $\square$