Probabilités (Rappels)

“Probabilités & Finance”

`www.master-finance.proba.jussieu.fr`

thématique de la spécialité

Probabilités & Applications

du Master 2

Sciences & Technologies

Université Pierre & Marie Curie (Paris 6)

GILLES PAGÈS

2006-07

Chapitre 1

Introduction

Nous avons tous des notions intuitives de probabilités parce que la vie courante nous met en présence de phénomènes aléatoires. Il s'agit de situations dont l'issue reste incertaine, mais au sujet desquelles nous disposons néanmoins d'une certaine information, par exemple grâce à la connaissance du passé.

Doit-on par exemple souscrire une assurance contre la foudre ? Oui, si l'on habite une maison isolée en haute-montagne, non si l'on habite à Paris près de la Tour Eiffel. Mais, entre ces deux situations extrêmes, on ne dispose que de peu de critères objectifs pour prendre une décision.

Cependant, on peut noter qu'aux temps préhistoriques une éclipse de soleil apparaissait comme un phénomène aléatoire alors qu'il n'en est rien maintenant pour tout détenteur de calendrier des Postes. De même, en France, aujourd'hui, le sexe d'un enfant à naître n'a plus rien d'aléatoire, échographie aidant, sauf volonté explicite des futurs parents.

Ces exemples illustrent le fait que, souvent, un phénomène n'apparaît aléatoire à un observateur qu'en raison d'un manque d'information. Cette information lacunaire résulte généralement de la complexité des causes régissant le phénomène.

Historiquement les premiers problèmes de probabilités dont on trouve trace datent du 17ème siècle et sont liés à des jeux de hasard. Un jeu de hasard est un jeu dans lequel l'incertitude est de nature purement probabiliste : jouer à pile ou face même avec une pièce truquée reste un vrai jeu de hasard si chacun des deux joueurs en est informé. Parmi ces problèmes, on peut citer :

• Le problème de Galilée (1554-1642) :

Le prince de Toscane demande à Galilée pourquoi, en lançant trois dés, on obtient plus souvent un total de 10 qu'un total de 9, alors qu'il y a six façons d'obtenir ces résultats.

• Le problème du Chevalier de Méré :

Le Chevalier de Méré demande à Pascal (1623-1662) s'il est plus facile d'obtenir un six au moins en lançant quatre fois de suite un seul dé, ou d'obtenir un double-six au moins en lançant vingt quatre fois de suite deux dés.

• Le pari de Pascal :

Le mathématicien et philosophe Blaise Pascal est, lui, le premier à avoir introduit la notion d'espérance mathématique dans son célèbre "pari". Son but, dans "Les Pensées", était de convaincre les tenants du septiscisme qu'il y avait tout à gagner en croyant à l'existence de Dieu.
Son raisonnement était le suivant :

- "Dieu est, ou il n'est pas".

- "Nous sommes incapables de savoir ni ce qu'il est ni si il est"

- "Nous n'avons pas le choix, nous sommes embarqués dans la vie, c'est la seule certitude, nous devons parier" sur cette question

- "La vie est un "jeu" qui ne comporte qu'une partie se présentant comme suit :
. Mise : la vie humaine (finie)
. Gain : la béatitude éternelle (infinie)

- Donc, en pariant sur la non-existence, de Dieu, il y a un risque non nul de perte infinie alors que le gain est forcément fini. Par contre, en pariant sur son existence on ne risque qu'une perte finie face à un gain éventuel infini.

Les solutions des problèmes de Galilée et du chevalier de Méré sont proposées au chap. I.3.

Chapitre 2

Introduction au modèle probabiliste. Variables aléatoires discrètes.

2.1 Espace probabilisé : définition et interprétation

Le formalisme et l'axiomatique modernes des Probabilités sont, pour une large part, l'oeuvre du mathématicien russe Kolmogorov dans les années 1930. Ils tirent l'essentiel de leur puissance de ce qu'ils (re)placent les Probabilités dans le cadre de la théorie de la mesure comme nous le verrons au chap. II.

Une *épreuve aléatoire* est caractérisée par l'ensemble des *résultats* ou *issues* possibles, généralement noté Ω (on parle aussi parfois "d'états du monde").

Exemples :

1) Jet d'une pièce de monnaie : $\Omega := \{\text{pile, face}\}$ (ou $\{P, F\}$),

2) Lancer d'un dé : $\Omega := \{1, ..., 6\}$,

3) Jet de n pièces de monnaie : $\Omega := \{P, F\}^n = \{(\alpha_1, ..., \alpha_n), \alpha_i \in \{P, F\}, 1 \leq i \leq 6\}$,

4) Nombre de pannes d'une machine : $\Omega := \mathbb{N}$,

5) Durée de vie d'une machine : $\Omega := \mathbb{R}_+$,

6) Lancer d'une fléchette sur une cible : $\Omega := D(0;1) \cup \{\delta\}$ où $D(0;1)$ désigne le disque unité représentant la cible et δ le point cimetière symbolisant le "ratage" de la cible.

On constate que selon les cas, Ω est fini, infini dénombrable ou non dénombrable.

Probabiliser l'ensemble Ω, c'est lui adjoindre une *tribu* $\mathcal{A} \subset \mathcal{P}(\Omega)$ de parties (ou *événements*) de Ω et une mesure de probabilité $\mathbb{P}$ attribuant à chaque événement $A \in \mathcal{A}$ une probabilité $\mathbb{P}(A) \in [0,1]$ caractérisant le degré de vraisemblance de la survenue de A.

Définition :

(a) Une famille de parties $\mathcal{A} \subset \mathcal{P}(\Omega)$ est une *tribu* si :

(i) $\phi \in \mathcal{A}$,

(ii) $A \in \mathcal{A} \Rightarrow {}^c\!A \in \mathcal{A}$ [donc $\Omega \in \mathcal{A}$],

(iii) Si $A_n \in \mathcal{A},\ n \geq 0$, alors $\bigcup_{n\in\mathbb{N}} A_n \in \mathcal{A}$ $\left[\text{et donc } \bigcap_{n\in\mathbb{N}} A_n \in \mathcal{A}\right]$.

On appelle *événement* tout élément de la tribu $\mathcal{A}$.

(b) On appelle *probabilité* sur $(\Omega, \mathcal{A})$ toute fonction $\mathbb{P} : \mathcal{A} \to [0,1]$ vérifiant :

(i) $\mathbb{P}(\Omega) = 1$,

(ii) Si $A_n \in \mathcal{A},\ n \geq 1$, avec $A_i \cap A_j = \phi$ dès que $i \neq j$ alors,

$$\mathbb{P}\left(\bigcup_n A_n\right) \overset{\sigma-\text{additivité}}{=} \sum_{n\geq 1} \mathbb{P}(A_n) \left(= \lim_n \sum_{k=1}^{n} \mathbb{P}(A_k)\right).$$

(c) Un triplet $(\Omega, \mathcal{A}, \mathbb{P})$ est appelé un *espace probabilisé*.

On remarque au vu de (b) qu'une probabilité n'est rien d'autre qu'une mesure positive de masse 1 (cf. ci-dessous : $\mathbb{P}(\phi) = 0$). L'intérêt d'une telle analogie est essentiel par exemple lorsque $\Omega = \mathbb{R}$, $\mathbb{R}_+$ ou $\mathbb{R}^d$ puisque la construction de mesures de probabilité sur de tels espaces s'appuiera alors sur les (très délicats) théorèmes d'existence et/ou de prolongement de mesures.

2.2 Premières propriétés

On se donne des événements (i.e. des éléments de la tribu $\mathcal{A}$) $A, B, A_n, n \geq 1$, sur un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$.

P1. $\mathbb{P}(\phi) = 0$ $\left(\mathbb{P}(\phi) = \lim_n n\mathbb{P}(\phi) \text{ d'après (b) } (ii) \text{ d'où le résultat}\right)$.

En particulier (b) (ii) reste donc vrai pour une *famille finie* $A_1, ..., A_n$ d'événements.

P2. $\mathbb{P}({}^c\!A) = 1 - \mathbb{P}(A)$ (où ${}^c\!A$ désigne le complémentaire de A).

P3. Si $A \subset B$, $\mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(B \setminus A) \geq \mathbb{P}(A)$ (où $B \setminus A := B \cap {}^c\!A \in \mathcal{A}$).

En effet $B = A \cup (B \cap {}^c\!A)$ avec $A \cap (B \cap {}^c\!A) = \phi$.

P4. $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$.

On écrit $A \cup B = (A \setminus A \cap B) \cup (B \setminus A \cap B) \cup (A \cap B)$ et on applique (b) (ii) et la propriété P3.

P5. Si $A_n \subset A_{n+1}$ $\left(\text{i.e. } (A_n)_{n\geq 1} \text{ croissante}\right)$ alors $\mathbb{P}\left(\bigcup_n A_n\right) = \lim_n\uparrow \mathbb{P}(A_n)$.

Si $A_{n+1} \subset A_n$ $\left(\text{i.e. } (A_n)_{n\geq 1} \text{ décroissante}\right)$ alors $\mathbb{P}\left(\bigcap_n A_n\right) = \lim_n\downarrow \mathbb{P}(A_n)$.

On pose $B_1 := A_1$ puis, par récurrence, $B_n := A_n \setminus \bigcup_{k=1}^{n-1} B_k, n \geq 2$. Les B_n sont clairement 2 à 2 disjoints et $\bigcup_{k=1}^{n} B_k = A_n$ puisque $A_n \subset \bigcup_{k=1}^{n} B_k$ et $B_n \subset A_n$. D'où $\mathbb{P}(A_n) = \sum_{k=1}^{n} \mathbb{P}(B_k) \to \sum_{k=1}^{+\infty} \mathbb{P}(B_k) = \mathbb{P}\left(\bigcup_{n\geq 1} B_n\right) = \mathbb{P}\left(\bigcup_{n\geq 1} A_n\right)$. Le résultat analogue dans le cas décroissant s'obtient par passage au complémentaire via les lois de Morgan.

P6. *σ-sous-additivité* : $\mathbb{P}\left(\bigcup_n A_n\right) \leq \sum_n \mathbb{P}(A_n)$.

On établit d'abord par récurrence que $\mathbb{P}\left(\bigcup_{k=1}^{n} A_k\right) \leq \sum_{k=1}^{n} \mathbb{P}(A_k)$ pour tout $n \geq 1$.

Si $n = 2$ cela découle de P4. Le passage de n à $n+1$ se fait par :

$$\mathbb{P}\left(\bigcup_{k=1}^{n+1} A_k\right) \leq \mathbb{P}\left(\bigcup_{k=1}^{n} A_k\right) + \mathbb{P}(A_{n+1}) \leq \sum_{k=1}^{n} \mathbb{P}(A_k) + \mathbb{P}(A_{n+1}).$$

On conclut en passant à la limite dans l'inégalité (via P5 "croissante").

Remarque : La formule P4 se généralise en *l'identité* dite de *Poincaré*

$$\mathbb{P}\left(\bigcup_{k=1}^{n} A_k\right) = \sum_{k=1}^{n} (-1)^{k-1} \sum_{\substack{I\subset\{1,\ldots,n\} \\ |I|=k}} \mathbb{P}\left(\bigcap_{i\in I} A_i\right) \qquad \text{(cf. exercice I.8).}$$

2.3 Probabilités sur un espace fini ou dénombrable

On munit canoniquement un ensemble E, supposé fini ou dénombrable, de la tribu de ses parties $\mathcal{P}(E)$. Il est clair par l'axiome de σ-additivité (b) (*ii*) qu'une probabilité π sur $(E, \mathcal{P}(E))$ est entièrement déterminée par la donnée des $\pi(e), e \in E$ où $\pi(e) := \pi(\{e\})$ puisqu'alors

$$\forall A \in \mathcal{P}(E), \qquad \pi(A) = \sum_{e\in A} \pi(e).$$

Dans ce cas on constate que l'on peut indifféremment assimiler π à une *probabilité* sur $\mathcal{P}(E)$ ou à une fonction sur E. Dans ce second cadre on parle parfois pour π de *densité discrète*.

La détermination, pour une épreuve aléatoire donnée, des probabilités ponctuelles $\pi(e), e \in E$, est une question généralement délicate, relevant de la théorie de l'estimation en Statistique. Cependant, lorsque E est fini, il arrive que, pour des raisons de

symétrie, tous les singletons $\{e\}, e \in E$, aient la même probabillité $\pi(e) = p$. On dit alors qu'il y a *équiprobabilité*. Comme

$$1 = \sum_{e\in E} \pi(e) = \sum_{e\in E} p = |E|\, p, \text{ il vient } p = \frac{1}{|E|} \text{ et partant, } \forall A \in \mathcal{P}(E), \ \pi(A) = \frac{|A|}{|E|}.$$

Dans ce cas très particulier, la détermination des probabilités des différents événements se ramène donc à des problèmes de dénombrement. C'est la situation que l'on rencontre dans la plupart des jeux de hasard (dés, cartes, roulettes,...).

Exemples :

• Paradoxe du chevalier de Méré (cf. introduction).

A_1 = "Obtenir au moins un 6 en 4 lancers de 1 dé".
A_2 = "Obtenir au moins un double-six en 24 lancers de 2 dés".

$$A_1 \in \mathcal{P}(\Omega_1), \text{ où } \Omega_1 = \{1,...,6\}^4 \text{ et } A_2 \in \mathcal{P}(\Omega_2) \text{ où } \Omega_2 = \left(\{1,...,6\}^2\right)^{24}.$$

Les dés étant supposés non pipés, on se place dans les deux cas sous l'hypothèse d'équiprobabilité. Donc, si $\mathbb{P}_1$ et $\mathbb{P}_2$ désignent respectivement les probabilités sur Ω_1 et Ω_2 :

$$\mathbb{P}_i(A_i) = \frac{|A_i|}{|\Omega_i|}, \quad i = 1, 2.$$

En fait, il est plus commode de passer aux complémentaires :
${}^cA_1 = \{\text{ne jamais obtenir 6 en 4 lancers}\}$; $|\Omega_1| = 6^4$ et $|{}^cA_1| = 5^4$ donc $\mathbb{P}(A_1) = 1 - \left(\frac{5}{6}\right)^4 = 1 - \frac{625}{1296} \approx 0,5177$.
${}^cA_2 = \{\text{ne jamais obtenir un double six en 24 lancers}\}$; $|\Omega_2| = 36^{24}$ et $|{}^cA_2| = 35^{24}$ donc $\mathbb{P}(A_2) = 1 - \left(\frac{35}{36}\right)^{24} \approx 0,4914$.

On constate donc, et là se nichait le paradoxe, que de disposer de 6 fois plus de tentatives pour obtenir un résultat 6 fois moins probable ne rétablit pas (tout à fait ...) l'équilibre.

• Problème du Prince de Toscane à Galilée :

$\Omega = \{1,...,6\}^3$ et hypothèse d'équiprobabilité. $|\Omega| = 6^3 = 216$.
A := "Obtenir une somme de 9" et B := "Obtenir une somme de 10".

9	$= 1+2+6$	soit 3 ! façons (selon l'ordre)	10	$= 1+3+6$	soit 3 ! façons
	$= 1+3+5$	soit 3 ! façons		$= 1+4+5$	soit 3 ! façons
	$= 1+4+4$	soit 3 façons		$= 2+2+6$	soit 3 façons
	$= 2+2+5$	soit 3 façons		$= 2+3+5$	soit 3 ! façons
	$= 2+3+4$	soit 3 ! façons		$= 2+4+4$	soit 3 façons
	$= 3+3+3$	soit 1 façon		$= 3+3+4$	soit 3 façons
		= 25 façons			= 27 façons

D'où $\mathbb{P}(A) = \dfrac{|A|}{|\Omega|} = \dfrac{25}{216}$ et $\mathbb{P}(B) = \dfrac{|B|}{|\Omega|} = \dfrac{27}{216}$.

2.4 Variables aléatoires à valeurs dans un espace discret

Définition 2 :

On appelle variable aléatoire (v.a.) discrète sur un espace $(\Omega, \mathcal{A}, \mathbb{P})$ à valeurs dans un ensemble E fini ou dénombrable, toute application $X : (\Omega, \mathcal{A}, \mathbb{P}) \to E$ vérifiant :

$$\forall e \in E, \;\; X^{-1}(\{e\}) \in \mathcal{A}.$$

Remarques :

• Par commodité on notera $\{X = e\} := X^{-1}(\{e\})$ l'image réciproque de $\{e\}$ par X ; plus généralement, on écrira $\{X \in B\} := X^{-1}(B)$.

• La condition ci-dessus exprime simplement que $X : (\Omega, \mathcal{A}) \to (E, \mathcal{P}(E))$ est une application *mesurable*.

• Pour toute partie A de Ω, $A \in \mathcal{A}$ ssi la fonction indicatrice de A, $\mathbf{1}_A$ est une variable aléatoire.

Propriété :

$\mathcal{X}_{\mathbb{K}} := \{X : (\Omega, \mathcal{A}) \longrightarrow E, E \subset \mathbb{K}, E$ fini ou dénombrable$\}$ est un $\mathbb{K}$-e.v. ($\mathbb{K} = \mathbb{R}$ ou $\mathbf{C}$).

Démonstration :

On considère $X : \Omega \to E,\; Y : \Omega \to E',\; \lambda \in \mathbb{K}$ et $Z := \lambda X + Y$.

$$\{Z = e\} = \bigcup_{\substack{(\varepsilon,\varepsilon') \in E \times E' \\ \lambda\varepsilon + \varepsilon' = e}} \{X = \varepsilon\} \cap \{X = \varepsilon'\}$$

or $\{(\varepsilon, \varepsilon') \in E \times E' / \lambda\varepsilon + \varepsilon' = e\} \subset E \times E'$, il est donc dénombrable. #

Définition 3 :

On appelle *loi* (ou parfois *densité discrète*, cf. Remarque p.5) de X la probabilité π sur $(E, \mathcal{P}(E))$ définie par : $\forall e \in E, \quad \pi(e) := \mathbb{P}(X = e)$.

Remarque : La probabilité π est souvent notée $\mathbb{P}_X$ puisque c'est en fait la mesure image de $\mathbb{P}$ par X. On vérifie que la v.a. X **transporte** la probabilité $\mathbb{P}$ sur $(\Omega, \mathcal{A})$ en une probabilité sur $(E, \mathcal{P}(E))$.

2.4.1 Lois discrètes usuelles

On désigne par $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé.

• Loi de Bernoulli $\mathcal{B}(p),\; p \in [0,1]$:

On considère un événement $C \in \mathcal{A}$ de probabilité $p := \mathbb{P}(C)$.
Soit $X := \mathbf{1}_C$ la fonction indicatrice de C ($X(\omega) = 1$ si $\omega \in C, = 0$ sinon).
On pose $\pi := \mathbb{P}_X$.
$\pi(\{1\}) = \mathbb{P}(X = 1) = \mathbb{P}(C) = p$ et $\pi(\{0\}) = 1 - p$.
π est appelée *loi de Bernoulli de paramètre* $p \in [0,1]$.

• Loi binomiale $\mathcal{B}(n;p), p \in [0,1]$:

Soit $X : (\Omega, \mathcal{A}, \mathbb{P}) \longrightarrow \{0, ..., n\}$ vérifiant

$$\forall k \in \{0, ..., n\}, \;\; \mathbb{P}(X = k) = C_n^k p^k (1-p)^{n-k} \in [0,1] \cdot$$

La loi $\pi := \mathbb{P}_X$ sur $\{0, ..., n\}$ est bien une probabilité car

$$\pi(k) \geq 0 \text{ et } \sum_{k=0}^{n} \pi(k) = \sum_{k=0}^{n} C_n^k p^k (1-p)^{n-k} = (p + 1 - p)^n = 1 \cdot$$

La loi de X est appelée *loi binomiale*, notée $\mathcal{B}(n;p)$, $n \geq 1$, $p \in [0,1]$.
Lorsque $p = \dfrac{N_1}{N} \in \mathbf{Q}^+ \cap [0,1]$, on peut la réaliser comme le nombre de boules blanches observées lors de n tirages *avec remise* dans une urne contenant N boules dont N_1 sont blanches. Plus généralement la loi $\mathcal{B}(n;p)$ s'obtient comme la loi de la somme de n v.a. indépendantes de loi $\mathcal{B}(p)$ (cf. section 7).

• Loi hypergéométrique $\mathcal{H}(n, N_1, N_2)$:

On procède cette fois à n tirages *sans remise* d'une boule dans une urne contenant initialement N boules dont N_1 sont blanches et N_2 sont noires ; on suppose que $n \leq \min(N_1, N_2)$. On modélise ce tirage sur l'espace :

$$\Omega := \{\text{parties à } n \text{ éléments d'un ensemble à } N_1 + N_2 \text{ éléments dont } N_1 \text{ sont blancs}\}$$

et l'on se place sous l'hypothèse d'équiprobabilité des tirages. On désigne par :

$$X : \begin{cases} \Omega \to \{0, ..., n\} \\ \omega \longmapsto X(\omega) := \text{card}\{x \in \omega / x \text{ est de type "blanc"}\} \end{cases}$$

$$\forall k \in \{0, ..., n\}, \;\; \mathbb{P}(X = k) = \frac{C_{N_1}^k C_{N_2}^{n-k}}{C_{N_1+N_2}^n}.$$

On vérifie que $\sum_{k=0}^{n} C_{N_1}^k C_{N_2}^{n-k} = C_{N_1+N_2}^n$ en identifiant les coefficients de Y^n dans les deux développements de $(1+Y)^{N_1+N_2} = (1+Y)^{N_1}(1+Y)^{N_2}$.

La loi de X a pour nom *loi hypergéométrique* de paramètres n, N_1, N_2.

• Loi géométrique (ou loi de Pascal) $\mathcal{G}(p), p \in]0,1[$:

On définit sur $\mathbb{N}^*$ la loi de probabilité : $\forall k \in \mathbb{N}^*, \quad \pi(k) := (1-p)^{k-1} p$. C'est bien une probabilité sur $\mathbb{N}^*$ puisque $\sum_{k \in \mathbb{N}^*} (1-p)^{k-1} p = p \dfrac{1-0}{1-(1-p)} = 1$. Cette loi s'obtient comme celle du "1er succès" lors de la répétition successive indéfinie d'épreuves de Bernoulli $\mathcal{B}(p)$ "indépendantes" (cf. section 5).

• Loi de Poisson $\mathcal{P}(\lambda), \lambda > 0$:

On définit sur $\mathbb{N}$ la probabilité : $\forall k \in \mathbb{N}, \quad \pi(k) = e^{-\lambda} \dfrac{\lambda^k}{k!}$. π est bien une probabilité puisque $\sum_{k \in \mathbb{N}} \pi(k) = e^{-\lambda} \sum_k \dfrac{\lambda^k}{k!} = e^{-\lambda+\lambda} = 1$. La loi de Poisson est une loi "d'événements rares" comme le montre la proposition suivante.

Proposition 1 :

Soit $X_n, n \geq 1$, une suite de v.a. de lois $\mathcal{B}(n; p_n)$ telle que $np_n \to \lambda > 0$ (*i.e.* $p_n \sim \dfrac{\lambda}{n}$). Alors :

$$\forall k \in \mathbb{N}, \quad \mathbb{P}(X_n = k) \to e^{-\lambda} \frac{\lambda^k}{k!}, \text{ quand } n \to +\infty.$$

Démonstration :

$$\begin{aligned} \text{Si } n \geq k, \quad \mathbb{P}(X_n = k) &= C_n^k p_n^k (1-p_n)^{n-k} = \frac{n!}{k!\,(n-k)!} p_n^k (1-p_n)^{n-k} \\ &= \frac{n(n-1) \times .. \times (n-k+1)}{n^k} \frac{(np_n)^k}{k!} \left(1 - \frac{np_n}{n}\right)^n (1-p_n)^{-k} \\ &\longrightarrow 1^k \frac{\lambda^k}{k!} e^{-\lambda} 1^{-k} = e^{-\lambda} \frac{\lambda^k}{k!}. \ \# \end{aligned}$$

On rencontre d'autres lois discrètes (cf. exercices) comme la loi binomiale négative, les lois multinomiales et hypergéométriques multi-dimensionnelles, ces deux dernières lois sont d'ailleurs des probabilités sur $\{0, ..., n\}^d, d \geq 2$.

2.4.2 Espérance (mathématique) d'une v.a. discrète ; applications

Définition 4 :

Soit $X : (\Omega, \mathcal{A}, \mathbb{P}) \longrightarrow E$, E *fini* ou *dénombrable*, $E \subset \mathbb{R}$ ou $\mathbf{C}$. Si

• $\sum_{x \in E} |x| \ \mathbb{P}(X = x) < +\infty$ alors X est *intégrable* et l'on définit alors *l'espérance* mathématique de X par :

$$\mathbb{E}(X) := \sum_{x \in E} x \ \mathbb{P}(X = x).$$

• Si $E \subset \overline{\mathbb{R}}_+$, on peut toujours définir $\mathbb{E}(X) := \sum_{x \in E} x \ \mathbb{P}(X = x) \in \overline{\mathbb{R}}_+$ (et l'on dira que X est intégrable si $\mathbb{E}(X) < +\infty$).

Remarques :

• Cette définition est justifiée par le fait qu'une série absolument convergente ou à termes positifs est commutativement convergente.

• Si $E \subset \mathbf{C}$ est $fini$, X est toujours intégrable.

• $\mathbb{E}(X)$ est entièrement déterminée par la loi de X ; cependant, si Ω lui-même est *fini ou dénombrable*, on vérifie aisément que $\mathbb{E}(X) = \sum_{\omega \in \Omega} X(\omega) \mathbb{P}(\{\omega\})$.

Proposition 2 :

Si $X, Y : (\Omega, \mathcal{A}, \mathbb{P}) \to E \subset \mathbb{R}$ ou $\mathbf{C}$ sont intégrables (resp. ≥ 0) alors : $\forall \lambda, \mu \in \mathbb{R}$ (resp. $\mathbb{R}_+$), $\lambda X + \mu Y$ est intégrable (resp. ≥ 0) et

$$\mathbb{E}(\lambda X + \mu Y) = \lambda \mathbb{E}(X) + \mu \mathbb{E}(Y).$$

Démonstration : Si $\lambda = \mu = 1$, $X + Y$ est à valeurs dans $F := E + E$ et

$$\begin{aligned} \sum_{z \in F} |z| \, \mathbb{P}(X + Y = z) &= \sum_{z \in F} \sum_{\substack{x, y \in E \\ x+y=z}} |x + y| \, \mathbb{P}(X = x, Y = y) \\ &\leq \sum_{x \in E} \sum_{y \in E} (|x| + |y|) \, \mathbb{P}(X = x, Y = y) \\ &\leq \sum_{x \in E} |x| \ \mathbb{P}(X = x) + \sum_{y \in E} |y| \, \mathbb{P}(Y = y) < +\infty. \end{aligned}$$

L'additivité de $\mathbb{E}(\cdot)$ s'établit via le même cheminement.

Si $\lambda \in \mathbb{R}^*$, $\sum_{z \in \lambda E} |z| \mathbb{P}(\lambda X = z) = \sum_{x \in E} |\lambda x| \mathbb{P}(\lambda X = \lambda x)$ car $x \longmapsto \lambda x$ est une bijection de E sur λE. Donc λX est intégrable, etc. Si $\lambda = 0$, $\mathbb{E}(\lambda X) = \mathbb{E}(0) = 0 = \lambda \mathbb{E}(X)$. #

Corollaire :

Si X et Y sont *intégrables* et $X \leq Y$ alors $\mathbb{E}(X) \leq \mathbb{E}(Y)$.

Proposition 3 :

Soit $\pi := (\pi(x))_{x \in E}$ une probabilité sur l'ensemble dénombrable E et $X : (\Omega, \mathcal{A}, \mathbb{P}) \longrightarrow (E, \mathcal{P}(E))$ une v.a.. Les assertions suivantes sont équivalentes :

(*i*) $\quad X$ a pour loi π $[$*i.e.* $\forall x \in E, \ \pi(x) = \mathbb{P}(X = x)]$,

(*ii*) $\quad \forall f : E \longrightarrow \mathbb{R}_+, \quad \mathbb{E}(f(X)) = \sum_{x \in E} f(x) \pi(x) \in \overline{\mathbb{R}}_+$,
$\left[\text{en particulier } \pi(x) = \mathbb{P}(X = x) = \mathbb{E}\left(\mathbf{1}_{\{x\}}(X)\right)\right]$,

(*iii*) $\quad \forall f : E \longrightarrow \mathbb{R}$, bornée, $f(X)$ est intégrable et $\mathbb{E}(f(X)) = \sum_{x \in E} f(x) \pi(x)$.

Démonstration :

$(i) \Longrightarrow (ii)$: Soit $Y := f(X)$ et $F := f(E) \subset \mathbb{R}_+$, $\mathbb{E}(Y) = \sum_{y\in F} y\mathbb{P}(Y=y)$.

Or $(Y=y) = \bigcup_{x\in f^{-1}(y)} (X=x)$ (union disjointe) d'où :

$$\begin{aligned}\mathbb{E}(Y) &= \sum_{y\in F} y \sum_{x\in f^{-1}(y)} \mathbb{P}(X=x) = \sum_{y\in F}\sum_{x\in f^{-1}(y)} f(x)\,\mathbb{P}(X=x)\\ &= \sum_{x\in E} f(x)\underbrace{\mathbb{P}(X=x)}_{=\pi(x)}.\end{aligned}$$

$(ii) \Rightarrow (iii)$ On décompose simplement $f = f^+ - f^-$ ($f^+ := \max(f,0)$ et $f^- := \max(-f,0)$); or $0 \le f^\pm \le |f| \le M := \sup_{x\in E} |f(x)|$, donc (cf. (ii))

$$\mathbb{E}\left(f^\pm(X)\right) = \sum_{x\in E} f^\pm(x)\,\pi(x) \le \sum_{x\in E} M\pi(x) = M < +\infty$$

i.e. $f^\pm(X)$ sont intégrables. D'autre part, en reprenant les calculs de $(i) \Rightarrow (ii)$ et en posant $F := f(E)$, il est immédiat que

$$\begin{aligned}\sum_{z\in F} |z|\mathbb{P}(f(X)=z) \le &= \sum_{x\in E} |f(x)|\mathbb{P}(X=x)\\ &= \sum_{x\in E} |f|(x)\pi(x) \le M < +\infty\end{aligned}$$

donc $f(X)$ est intégrable. Enfin

$$\begin{aligned}\sum_{x\in E} f(x)\,\pi(x) &= \sum_{x\in E}(f^+(x) - f^-(x))\,\pi(x) = \sum_{x\in E} f^+(x)\,\pi(x) - \sum_{x\in E} f^-(x)\,\pi(x)\\ &= \mathbb{E}(f^+(X)) - \mathbb{E}(f^-(X)) = \mathbb{E}(f(X)) \text{ d'après la proposition 2.}\end{aligned}$$

$(iii) \Longrightarrow (i)$: Il suffit de vérifier que $\mathbb{P}(X=x) = \pi(x), x \in E$, ce qui découle de l'assertion (iii) appliquée aux $f_x := \mathbf{1}_{\{x\}}$ et du fait que $\mathbb{P}(X=x) = \mathbb{E}\left(\mathbf{1}_{\{x\}}(X)\right)$ (cf. (ii)). #

Exemple : si $A \in \mathcal{P}(X)$, $\mathbb{E}(\mathbf{1}_A(X)) = \sum_{x\in E} \mathbf{1}_A(x)\,\pi(x) = \sum_{x\in A} \pi(x) = \pi(A) = \mathbb{P}(X\in A)$.

Applications :

• D'après (ii), X est intégrable ssi $\mathbb{E}|X| < +\infty$ et, plus généralement $f(X)$ est intégrable ssi $\mathbb{E}|f(X)| < +\infty$.

• Si $E \subset \mathbb{R}$ et $p \in \mathbb{N}^*$, dès que $\sum_{x\in E} |x|^p\mathbb{P}(X=x) < +\infty$, on définit le *moment d'ordre* p de X par $\mathbb{E}(X^p)$.

• Si $p < p'$ et si $\mathbb{E}\left(|X|^{p'}\right) < +\infty$, alors $\mathbb{E}(|X|^p) < +\infty$ puisque $\mathbb{E}(|X|^p) \le \mathbb{E}\left(|X|^{p'}\right) + \mathbb{P}(|X| \le 1)$. En particulier, X est intégrable dès que X^2 l'est. D'où la définition de la variance.

Définition 5 :

Si X admet un moment d'ordre 2, alors on pose : $Var(X) := \mathbb{E}\left((X - \mathbb{E}X)^2\right)$, *la variance* de X et $\sigma(X) := \sqrt{Var(X)}$, *l'écart-type* de X.

Proposition 4 :

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

(b) $\forall \lambda, \mu \in \mathbb{R}$, $Var(\lambda X + \mu) = \lambda^2 Var(X)$.

Démonstration :

$$\begin{aligned}\mathrm{Var}(X) &= \sum_{x\in E}(x - \mathbb{E}X)^2\,\mathbb{P}(X=x)\\ \text{(a)} \qquad &= \sum_{x\in E} x^2\mathbb{P}(X=x) - 2\mathbb{E}(X)\sum_{x\in E} x\mathbb{P}(X=x) + (\mathbb{E}X)^2\\ &= \mathbb{E}(X^2) - (\mathbb{E}X)^2\end{aligned}$$

(b) est immédiat par linéarité de l'espérance. #

Exemples : A partir des définitions ci-dessus, on calcule :

• $X \overset{\mathcal{L}}{\sim} \mathcal{B}(p)$, $\mathbb{E}X = p$ et $Var(X) = p(1-p)$.

• $X \overset{\mathcal{L}}{\sim} \mathcal{B}(n;p)$, $\mathbb{E}X = np$ et $VarX = np(1-p)$.

• $X \overset{\mathcal{L}}{\sim} \mathcal{H}(n;N_1;N_2)$, $\mathbb{E}(X) = n\frac{N_1}{N_1+N_2}$ et $Var(X) = \frac{N_1+N_2-n}{N_1+N_2-1} \times n \times \frac{N_1}{N_1+N_2}\left(1 - \frac{N_1}{N_1+N_2}\right)$
$\frac{N_1+N_2-n}{N_1+N_2-1}$ est appelé le facteur d'exhaustivité.

• $X \overset{\mathcal{L}}{\sim} \mathcal{P}(\lambda)$, $\mathbb{E}(X) = \lambda$ et $Var(X) = \lambda$.

• $X \overset{\mathcal{L}}{\sim} \mathcal{G}(p)$, $\mathbb{E}(X) = \dfrac{1}{p}$ et $Var(X) = \dfrac{1-p}{p^2}$.

2.5 Probabilités conditionnelles élémentaires. Evénements indépendants

2.5.1 Conditionnement par un événement non négligeable

Définition 6 :

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé et $B \in \mathcal{A}$ tel que $\mathbb{P}(B) \neq 0$. On définit la probabilité conditionnelle de A sachant (ou "quand") B par :

$$\mathbb{P}(A/B) := \frac{\mathbb{P}(A\cap B)}{\mathbb{P}(B)} \text{ (notée aussi } \mathbb{P}^B(A)).$$

Cette définition est justifiée par le fait que si Ω est fini et que l'on se place sous l'hypothèse d'*équiprobabilité*, la probabilité que l'on tire un élément de A sachant que l'on tire dans B est $\dfrac{|A\cap B|}{|B|} = \dfrac{\mathbb{P}(A\cap B)}{\mathbb{P}(B)}$ puisque $\mathbb{P}(C) = \dfrac{|C|}{|\Omega|}$.

On vérifie aussitôt que $\mathbb{P}(\,.\,|B)$ *est une nouvelle probabilité* sur $(\Omega, \mathcal{A}, \mathbb{P})$ avec toutes les propriétés (cf. section 2) que cela induit, auxquelles s'ajoutent quelques formules classiques spécifiques.

<u>Proposition 5</u> :

(a) $\forall A_1, ..., A_n \in \mathcal{A}$ tels que $\mathbb{P}(A_1\cap ...\cap A_n) > 0$,

$$\mathbb{P}(A_1\cap ...\cap A_n) = \mathbb{P}(A_1)\,\mathbb{P}(A_1/A_2)\,\mathbb{P}(A_3/A_1\cap A_2)\times ...\times \mathbb{P}(A_n/A_1\cap ...\cap A_{n-1})\,.$$

(b) $\mathbb{P}(A) = \mathbb{P}(A/B)\ \mathbb{P}(B) + \mathbb{P}(A/{}^cB)\ \mathbb{P}({}^cB)$

(c) $\mathbb{P}(A/B) = \mathbb{P}(B/A)\dfrac{\mathbb{P}(A)}{\mathbb{P}(B)}$ (formule de Bayes dite aussi "de probabilités des causes").

<u>Démonstration :</u>
(a) Raisonnons par récurrence sur n. Si $n = 2$, c'est la définition. D'autre part $\mathbb{P}(A_1\cap ...\cap A_{n+1}) = \mathbb{P}(A_1\cap ...\cap A_n)\ \mathbb{P}(A_{n+1}/A_1\cap ...\cap A_n)$ d'où le résultat d'après l'hypothèse de récurrence.
(b) et (c) sont évidents. #

2.5.2 Indépendance

<u>Définition 7</u> :

Deux événements $A,\ B \in \mathcal{A}$ sont **indépendants** si

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

On vérifie que, dès que $\mathbb{P}(B) = 0$ ou 1, A et B sont toujours indépendants ; d'autre part, si $\mathbb{P}(B) \neq 0, 1$, la définition ci-dessus est équivalente à $\mathbb{P}(A/B) = \mathbb{P}(A/{}^cB) = \mathbb{P}(A)$. D'où la justification de la terminologie "indépendant" : conditionner par B ou son complémentaire laisse la probabilité de A inchangée.

<u>Remarque</u> : Si A et B sont indépendants, A et ${}^cB$, ${}^cA$ et ${}^cB$ le sont aussi.

<u>Exemple</u> : Jet de 2 pièces équilibrées, $\Omega = \{P, F\}^2$, $\mathbb{P}(A) = \dfrac{|A|}{|\Omega|}$ (équiprobabilité).

Soit $A_1 := \{\text{les 2 pièces tombent sur le même côté}\} = \{PP, FF\}$,
$A_2 := \{\text{la 1ère pièce tombe sur pile}\} = \{PP, PF\}$,
$A_3 := \{\text{la 2ème pièce tombe sur face}\} = \{PF, FF\}$.

On vérifie aisément que $\mathbb{P}(A_i) = \dfrac{1}{2}, i = 1, 2, 3$. $\mathbb{P}(A_i\cap A_j) = \dfrac{1}{4}, 1 \leq i \neq j \leq 3$ donc les A_i sont *2 à 2 indépendants.*

<u>Définition 8</u> :

Soit $(A_i)_{i\in I}$ une famille finie d'événements de $\mathcal{A}$. Les A_i sont (mutuellement) indépendants si

$$\forall J \subset I,\ \mathbb{P}\Big(\bigcap_{j\in J} A_j\Big) = \prod_{j\in J} \mathbb{P}(A_j)\,.$$

<u>Remarques</u> :

- Il y a donc, selon cette définition, $2^{|I|}$ relations (dont $|I|$ sont triviales) à vérifier.
- On montre aisément par récurrence sur $|I|$ que l'on peut substituer son complémentaire à un nombre arbitraire d'événements $A_i,\ \ i \in I$.

<u>Contre-exemples</u> :

- Dans l'exemple précédent A_1, A_2, A_3 ne sont pas indépendants car $A_1\cap A_2\cap A_3 = \emptyset$ et $\prod_{i=1} \mathbb{P}(A_i) = \dfrac{1}{8} \neq 0$. En fait, sur $\Omega = \{P, F\}^2$, il n'existe pas 3 événements $A_i \neq \emptyset$, $A_i \neq \Omega$, qui soient indépendants car, quitte à passer au complémentaire, on peut supposer que $\mathbb{P}(A_i) \in \left\{\dfrac{1}{4}, \dfrac{1}{2}\right\}$ donc $\prod_{i=1} \mathbb{P}(A_i)$ est dans $\left[\dfrac{1}{4^3}, \dfrac{1}{2^3}\right]$; or $\bigcap_{i=1}^{3} A_i = \emptyset$ ou $\mathbb{P}\Big(\bigcap_{i=1}^{3} A_i\Big) \geq \dfrac{1}{4} > \dfrac{1}{2^3}$.
- Si $\Omega = \{1, ..., 6\}$, $\mathbb{P}(\{\omega\}) = \dfrac{1}{6}$, $A_1 = \{1, 2, 3\}$, $A_2 = \{2, 4, 6\}$ et $A_3 = \{1, 2, 4, 5\}$ vérifient $\mathbb{P}(A_1\cap A_2\cap A_3) = \dfrac{1}{6} = \prod_{i=1} \mathbb{P}(A_i)$, cependant $\mathbb{P}(A_1\cap A_2) = \dfrac{1}{6} \neq \mathbb{P}(A_1)\mathbb{P}(A_2)$.

2.6 V.a. discrètes indépendantes

<u>Définition 9</u> :

Soient $X_i : (\Omega, \mathcal{A}, \mathbb{P}) \to E_i, i = 1, ..., n,\ \ n$ v.a. discrètes définies sur un *même* espace probabilisé. Les $X_i, i = 1, ..., n$ sont *indépendantes* si et seulement si :

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

<u>Proposition 6</u> :

Les assertions suivantes sont équivalentes :

(i) Les $X_i, i = 1, ..., n$ sont indépendantes,

(ii) $\forall f_i : E_i \to \mathbb{R}$, bornée ou positive, $i = 1, ..., n$,

$$\mathbb{E}\left(\prod_{i=1} f_i(X_i)\right) = \prod_{i=1} \mathbb{E}(f_i(X_i)).$$

(iii) $\forall B_i \in \mathcal{P}(E_i), 1 \le i \le n$,

$$\mathbb{P}(X_1 \in B_1, ..., X_n \in B_n) = \mathbb{P}(X_1 \in B_1) ... \mathbb{P}(X_n \in B_n).$$

(iv) $\forall (x_1, ..., x_n) \in E_1 \times ... \times E_n$,

les événements $\{X_i = x_i\}, i = 1, ..., n,$ sont indépendants.

<u>Démonstration</u> :
$(i) \Rightarrow (ii)$ D'après la proposition 3 (ii) (section 4.2.) appliquée à $X = (X_1, ..., X_n)$, il vient :

$$\begin{aligned}\mathbb{E}\left(\prod_{i=1} f_i(X_i)\right) &= \sum_{(x_1,...x_n)\in E_1\times...\times E_n} \prod_{i=1}^{n} f_i(x_i)\ \mathbb{P}((X_1, ..., X_n) = (x_1, ..., x_n)) \\ &= \sum_{(x_1,...x_n)\in E_1\times...\times E_n} \prod_{i=1}^{n} (f_i(x_i)\ \mathbb{P}(X_i = x_i)) \\ &= \prod_{i=1}^{n} \left(\sum_{x_i\in E_i} f_i(x_i)\ \mathbb{P}(X_i = x_i)\right) = \prod_{i=1}^{n} \mathbb{E}(f_i(X_i)).\end{aligned}$$

$(ii) \Rightarrow (iii)$ Prendre $f_i := \mathbf{1}_{B_i}, 1 \le i \le n$, et noter que $\mathbb{P}(X_i \in B_i) = \mathbb{E}(\mathbf{1}_{B_i}(X_i))$.
$(iii) \Rightarrow (iv)$ Pour tout $J \subset \{1, ..., n\}$, prendre $B_j = E_j$ si $j \notin J$ et $B_j = \{x_j\}$ si $j \in J$.
$(iv) \Rightarrow (i)$ Evident via la définition de l'indépendance des événements. #

<u>Remarques</u> :

• On constate sur l'assertion (iii) que l'indépendance des X_i se traduit par la relation $\pi = \pi_1 \otimes ... \otimes \pi_n$ où π désigne la loi de $X = (X_1, ..., X_n)$ sur $E_1 \times ... \times E_n$, π_i, $1 \le i \le n$, les lois des X_i sur E_i et $\pi_1 \otimes ... \otimes \pi_n$ la mesure produit correspondante.

• Si l'on interprète π et π_i, $1 \le i \le n$, en termes de densités discrètes, la relation d'indépendance se traduit par (cf. (ii)) :

$$\forall (x_1, ..., x_n) \in E_1 \times ... \times E_n,\ \pi(x_1, ..., x_n) = \pi_1(x_1) \times ... \times \pi_n(x_n).$$

<u>Corollaires</u> :

Si les $X_i, 1 \le i \le n$, sont indépendantes, alors :
(a) $\forall I \subset \{1, ..., n\}$, $(X_i)_{i\in I}$ sont indépendantes,
(b) Si $I_1 \cup I_2 = \{1, ..., n\}$, $I_1 \cap I_2 = \phi$, alors $Y_1 = (X_i)_{i\in I_1}$ et $Y_2 = (X_i)_{i\in I_2}$ sont indépendantes,
(c) Si $g_i : E_i \to F_i$, $1 \le i \le n$, alors les $g_i(X_i)$, $1 \le i \le n$, sont indépendantes.
(d) Les événements $A_1, \cdots, A_n$ sont indépendants ssi les v.a. $\mathbf{1}_{A_1}, \cdots, \mathbf{1}_{A_n}$ sont indépendantes.

<u>Démonstration</u> :
(a) Prendre $f_i := \mathbf{1}$ si $i \notin I$.
(b) Supposons pour simplifier que $I_1 = \{1, ..., p\}$, $I_2 = {}^c I_1$; soient $y_1 = (x_1, ..., x_p)$ et $y_2 = (x_{p+1}, ..., x_n)$,

$$\begin{aligned}\mathbb{P}(Y_1 = y_1, Y_2 = y_2) &= \mathbb{P}((X_1, ..., X_n) = (x_1, ..., x_n)) \\ &= \prod_{i=1}^{n} \mathbb{P}(X_i = x_i) \\ &= \mathbb{P}(Y_1 = y_1)\, \mathbb{P}(Y_2 = y_2) \text{ via (a).}\end{aligned}$$

(c) Appliquer la caractérisation (ii) à $f_i := h_i o g_i, h_i : F_i \to \mathbb{R}_+$, $1 \le i \le n$.
(d) Le sens direct découle de la remarque sous la définition 8 et la réciproque du point (iv) de la proposition 6. #

<u>Proposition 7</u> :
(a) Si les $X_1, ..., X_n$ sont indépendantes à valeurs dans $E_i \subset \mathbb{R}$, intégrables (resp. positives) alors $X_1 ... X_n$ est intégrable (resp. positive) et

$$\mathbb{E}(X_1 ... X_n) = \mathbb{E}(X_1) ... \mathbb{E}(X_n) \text{ (la "réciproque" est fausse).}$$

(b) Si les $X_i, 1 \le i \le n$, sont indépendantes de carré intégrable, à valeurs dans $E_i \subset \mathbb{R}$, alors

$$\operatorname{Var}(X_1 + ... + X_n) = \operatorname{Var}(X_1) + ... + \operatorname{Var}(X_n).$$

<u>Démonstration</u> :
(a) On applique proposition 6 (ii) à $f_i(x_i) := |x_i|$ pour montrer que :

$$\mathbb{E}|X_1 ... X_n| = \mathbb{E}|X_1| ... \mathbb{E}|X_n| < +\infty.$$

L'égalité découle alors du théorème sur les produits de séries absolument convergentes.

$$\text{(b) } \mathbb{E}\left(\left(\sum_{i=1}^{n} X_i\right)^2\right) = \sum_{i,j} \mathbb{E}(X_i X_j) = \sum_i \mathbb{E}(X_i^2) - (\mathbb{E}X_i)^2 + \sum_{i,j} \mathbb{E}(X_i)\,\mathbb{E}(X_j)$$
$$= \sum_{i=1}^{n} Var(X_i) + \left[\mathbb{E}\left(\sum_{i=1}^{n} X_i\right)\right]^2 . \#$$

Remarque : Comme le montre la démonstration, l'additivité de la variance est en fait vérifiée dès que

$$\mathbb{E}(X_i X_j) = \mathbb{E}(X_i)\,\mathbb{E}(X_j),\ i \neq j.$$

Cette propriété est appelée *non corrélation* des v.a. X_i, $1 \leq i \leq n$.

Ainsi, d'après (a), les X_i sont non corrélées dès qu'elles sont 2 à 2 indépendantes. Or, l'indépendance 2 à 2 **n'entraîne pas** en général la mutuelle indépendance (considérer les indicatrices $\mathbf{1}_{A_i}$, $i = 1, 2, 3$ construites à partir des événements A_i explicités dans l'exemple du paragraphe 5.2). L'exemple ci-après fournit directement une situation où deux variables non corrélées ne sont pas indépendantes.

Exemple : Soit X une v.a. définie sur un espace $(\Omega, \mathcal{A}, \mathbb{P})$ à valeurs dans $\{-1, 0, 1\}$ et de loi uniforme *i.e.* $\mathbb{P}(X = -1) = \mathbb{P}(X = 0) = \mathbb{P}(X = 1) = 1/3$. On pose $Y := X^2$. Comme $XY = X^3 = X$, il est clair que $\mathbb{E}(XY) = 0$ et que $\mathbb{E}(X)\mathbb{E}(Y) = 0 \times \mathbb{E}(Y) = 0$. Les v.a. sont donc non corrélées. Pour autant, elles ne sont pas indépendantes puisque

$$\mathbb{E}(X^2 Y) = \mathbb{E}(X^4) = 1 \times (2/3) + 0 \times (1/3) = 2/3 \neq \mathbb{E}(X^2)\mathbb{E}(Y) = \mathbb{E}(X^2)^2 = (2/3)^2.$$

2.7 Fonction génératrice d'une v.a. à valeurs dans $\overline{\mathbb{N}}$. Applications

On prend comme conventions dans cette section :

$$0^0 = 1,\ 0 \times (+\infty) = 0 \text{ et : } \forall s \in [-1, 1],\ \ s^\infty = 0.$$

Définition 10 :

Soit $X : (\Omega, \mathcal{A}, \mathbb{P}) \to \overline{\mathbb{N}} = \mathbb{N} \cup \{+\infty\}$. On appelle *fonction génératrice* de X la fonction g_X définie sur $[-1, 1]$ par

$$g_X(s) := \mathbb{E}\left(s^X\right).$$

Cette définition a bien un sens puisque, si $s \in [-1, 1]$, $s^X \in \{1, s, ..., s^n, ...\} \cup \{0\} \subset [-1, 1]$ et, d'après la proposition 3 (ii) et la convention ci-dessus :

$$g_X(s) = \sum_{n \in \mathbb{N}} s^n \mathbb{P}(X = n).$$

Plusieurs démonstrations sur la fonction génératrice reposent sur le lemme (classique) suivant relatifs aux séries entières.

Lemme : (a) Si $a_n \geq 0, n \geq 0$ et $\sum_n a_n s^n$ de rayon (supérieur ou égal à) 1. Alors :

$$\lim_{s \to 1^-} \sum_n a_n s^n = \sum_n a_n \leq +\infty.$$

(b) Si $\sum_n |a_n| < +\infty$ alors $s \mapsto \sum_n a_n s^n$ est continue sur $[-1, 1]$.

Preuve : (a) $\forall n \geq 0$, $0 \leq a_n s^n \uparrow a_n$ quand $s \uparrow 1^-$ donc, d'après le théorème de Beppo-Lévi appliqué à la mesure de décompte $\mu := \sum_n \delta_n$,

$$\sum_{n \in \mathbb{N}} a_n s^n \uparrow \sum_{n \in \mathbb{N}} a_n \text{ quand } s \uparrow, 1^-.$$

(b) On procède de même en concluant par convergence dominée. #

Propriétés :

(a) • $g_X(0) = \mathbb{P}(X = 0)$ et $g_X(1) = \mathbb{P}(X < +\infty) \leq 1$ donc la série entière ci-dessus a un rayon de convergence $R_{g_X} \geq 1$. En particulier $g_X \in \mathcal{C}^\infty(]-1, 1[)$.

• En outre, $g_X \in \mathcal{C}([-1, 1])$, g_X est croissante sur $[0, 1]$ ainsi que toutes ses dérivées.

(b) $g_X = g_Y \Rightarrow X \overset{\mathcal{L}}{\sim} Y$ et $\mathbb{P}(X = n) = \dfrac{g_X^{(n)}(0)}{n!}$, $n \in \mathbb{N}$.

Démonstration :

(a) Le premier point est évident ainsi que la croissance des $g_X^{(p)}, p \in \mathbb{N}$ (où l'on a posé $g_X^{(0)} := g_X$). La continuité en ± 1 découle du lemme (b).

(b) Par dérivations successives sous le signe somme :

$$g_X^{(n)}(0) = n!\mathbb{P}(X = n),\ n \in \mathbb{N} \text{ et } \mathbb{P}(X = +\infty) = 1 - \sum_{n \in \mathbb{N}} \mathbb{P}(X = n).\ \#$$

Remarque : Si $\mathbb{P}(X = +\infty) > 0$, il est immédiat (cf. prop. 3) que $\mathbb{E}(X) = +\infty$.

Proposition 8 :

(a) On suppose que $\mathbb{P}(X = +\infty) = 0$. Alors

$$\mathbb{E}(X) = \lim_{s \to 1^-} g_X'(s) \leq +\infty.$$

En particulier X est intégrable ssi $\lim_{s \to 1^-} g_X'(s) < +\infty$ (soit encore ssi g_X est dérivable en $1-$).

(b) On suppose toujours que $\mathbb{P}(X = +\infty) = 0$. Alors, pour tout $p \in \mathbb{N}$,

$$\underbrace{\mathbb{E}\left(X\left(X-1\right)...\left(X-p+1\right)\right)}_{\text{moment factoriel d'ordre } p}=\lim_{s\to1^-} g_X^{(p)}\left(s\right)\leq+\infty.$$

X admet un moment d'ordre $p\in\mathbb{N}^*$ fini ssi X a un moment factoriel d'ordre p fini *i.e.* $\lim\limits_{s\to1^-} g_X^{(p)}\left(s\right)<+\infty$ (soit encore ssi g_X admet une dérivée d'ordre p en $1-$).

En particulier si X est de carré intégrable : $\mathrm{Var}\left(X\right)=\lim\limits_{s\to1^-} g_X''(s)-\lim\limits_{s\to1^-} g_X'(s)\times\left(\lim\limits_{s\to1^-} g_X'(s)-1\right)$.

(c) Si $\mathbb{P}\left(X=+\infty\right)=0$ et $R_{g_X}>1$ alors g_X est indéfiniment (continument) dérivable sur $]-R_{g_X},R_{g_X}[$, par conséquent X a des moments à tout ordre et

$$\mathbb{E}\left(X\left(X-1\right)...\left(X-p+1\right)\right)=g_X^{(p)}\left(1\right)\leq+\infty.$$

<u>Démonstration de la proposition 8</u> :
(a)-(b) Soit $p\in\mathbb{N}$. Il vient par dérivations successives de la série entière sous le signe somme sur $]-1,1[$ et le lemme (a) ci-dessus :

$$\lim_{1^-} g_X^{(p)}=\sum_{k\geq p}A_k^p\mathbb{P}(X=k)\leq+\infty.$$

La proposition 3, $\mathbb{P}(X=+\infty)=0$ et la convention $0\times+\infty$impliquent alors que

$$\mathbb{E}(X(X-1)\cdots(X-p+1))=\lim_{1^-} g_X^{(p)}.$$

L'équivalence entre la finitude des moments factoriels et naturels d'ordre p se déduit des inégalités

$$X(X-1)\cdots(X-p+1)\leq X^p\leq(2^p+1)(p-1)^p+2^pX(X-1)\cdots(X-p+1).$$

La seconde inégalité s'obtient en majorant X sur l'événement $\{X<p-1\}$ et en décomposant $X=X-(p-1)+p-1$ sur $\{X\geq p-1\}$ puis en utilisant $(a+b)^p\leq 2^p(a^p+b^p)$.

(c) Ce point est évident au vu des précédents. #

<u>Proposition 9</u> : Si $X_1,...,X_n$ sont indépendantes, alors, au moins pour $s\in[-1,1]$,

$$g_{X_1+...+X_n}\left(s\right)=g_{X_1}\left(s\right)...g_{X_n}\left(s\right).$$

<u>Démonstration</u> : Pour tout $s\in[-1,1]\,,n\to s^n$ est bornée, donc d'après la proposition 6 (*ii*),

$$g_{X_1+...+X_n}\left(s\right)=\mathbb{E}\left(s^{X_1+...+X_n}\right)=\mathbb{E}\left(s^{X_1}...s^{X_n}\right)=\mathbb{E}\left(s^{X_1}\right)...\mathbb{E}\left(s^{X_n}\right)=g_{X_1}\left(s\right)...g_{X_n}\left(s\right).$$

<u>Applications aux lois usuelles</u> :

- $\bullet\underline{X\overset{\mathcal{L}}{\sim}\mathcal{B}\left(p\right)}$: $g_X\left(s\right)=sp+1-p.$
- $\bullet\underline{X\overset{\mathcal{L}}{\sim}\mathcal{B}\left(n;p\right)}$: $g_X\left(s\right)=\sum\limits_{0\leq k\leq n}C_n^kp^k\left(1-p\right)^{n-k}s^k=\left(sp+1-p\right)^n.$

D'après la proposition 9 ci-dessus, on a donc :

$$X\overset{\mathcal{L}}{\sim}X_1+...+X_n \text{ où } X_i\overset{\mathcal{L}}{\sim}\mathcal{B}\left(p\right),1\leq i\leq n, \text{ sont indépendantes.}$$

$\mathbb{E}\left(X\right)=g_X'\left(1\right)=np,\quad Var\left(X\right)=\mathbb{E}\left(X\left(X-1\right)\right)-\mathbb{E}\left(X\right)\left(\mathbb{E}\left(X\right)-1\right)=np\left(1-p\right).$

$\bullet\underline{X\overset{\mathcal{L}}{\sim}\mathcal{G}\left(p\right)}$: $g_X\left(s\right)=\sum\limits_{k\in\mathbb{N}^*}p\left(1-p\right)^{k-1}s^k=\dfrac{ps}{1-s\left(1-p\right)}$

car $\sum\limits_{k\geq0}x^k=\dfrac{1}{1-x}$ pour $|x|<1$.

$\mathbb{E}(X)=g_X'\left(1\right)=\dfrac{1}{p}$ et $Var\left(X\right)=\dfrac{1-p}{p^2}$ (cf. formule ci-dessus).

$\bullet\underline{X\overset{\mathcal{L}}{\sim}\mathcal{P}\left(\lambda\right)}$: $g_X\left(s\right)=\sum\limits_{n\in\mathbb{N}}e^{-\lambda}\dfrac{\lambda^n}{n!}s^n=e^{-\lambda(1-s)}=e^{\lambda(s-1)}.$

D'où $\mathbb{E}(X)=\lambda$ et $Var\left(X\right)=\lambda$ et, pour tout $p\geq1$, $\mathbb{E}\left[X(X-1)\cdots(X-p+1)\right]=\lambda^p$.

<u>Remarques</u> :

- La réciproque de la proposition 9 est fausse.
- On peut définir de façon analogue la fonction génératrice d'un vecteur aléatoire à valeurs dans $\mathbb{N}^d\cup\{\infty\}$ par

$$g_{(X_1,...,X_d)}\left(s_1,...,s_d\right)=\mathbb{E}\left(s_1^{X_1}\times...\times s_d^{X_d}\right).$$

Plus généralement, la fonction génératrice est un outil essentiel de l'Analyse combinatoire.

2.8 Compléments sur l'indépendance

<u>Définition 11</u> : Une famille quelconque de v.a. $(X_i)_{i\in I}$ définies sur un espace probabilisé $(\Omega,\mathcal{A},\mathbb{P})$ est constitué de v.a. indépendantes si et seulement si $\forall J\in\mathcal{P}\left(I\right)$, J finie, les v.a. $\left(X_j\right)_{j\in J}$ sont indépendantes.

Ainsi une suite $\left(X_n\right)_{n\geq1}$ de v.a. est constituée de v.a. indépendantes si et seulement si $\forall n\geq1$, $X_1,...,X_n$ sont indépendantes (cf. corollaire (a) de la proposition 6).

La construction de telles suites dépasse largement le cadre de ce chapitre introductif. Vue leur importance en Probabilités, c'est l'une des motivations pour recourir à la théorie de la mesure (cf. chap. IV, VIII).

Chapitre 3

La probabilité comme mesure

La présentation des Probabilités que nous avons adoptée dans le chapitre I était essentiellement fondée sur une approche intuitive et mathématiquement élémentaire dans laquelle n'interviennent que des espaces probabilisés et des variables aléatoires discrètes (*i.e.* ne faisant intervenir qu'un nombre fini ou dénombrable de valeurs). Les concepts mathématiques qu'elle utilise étaient pour la plupart connus dès 1830 et certains résultats importants que nous n'avons pas encore évoqués (inégalité de Bienaymé-Tchebycheff, loi faible des grands nombres) avaient déjà été établis dans ce cadre.

C'est au début du 20è siècle, avec la génèse de la théorie de la mesure due à Borel et Lebesgue, qu'interviendra la mutation des Probabilités qui aboutira, sous l'égide de Kolmogorov (1930), à son axiomatique moderne.

3.1 Les limites de l'approche élémentaire : probabilités sur un ensemble non dénombrable

En dépit de la définition proposée dans le chapitre I, nous n'avons jusqu'à maintenant jamais construit de probabilités sur un ensemble Ω *non dénombrable*. Or cette question se pose très rapidement dans des problèmes de modélisation probabiliste pourtant anodins a priori. Ainsi, la modélisation d'un nombre infini de parties de Pile ou Face est hors d'atteinte des seules méthodes développées précédemment.

En effet, pour un nombre fini n de parties on peut considérer $\Omega = \{P, F\}^n$, $\mathcal{A} = \mathcal{P}(\Omega)$, $\mathbb{P}(\{\omega\}) := \dfrac{1}{2^n}$, dans le cas d'une pièce non truquée, et

$$\forall \omega = (\omega_1, ..., \omega_n) \in \Omega,\ \mathbb{P}(\{\omega\}) := p^{|\{i/\omega_i = P\}|} \times (1-p)^{|\{i/\omega_i = F\}|}$$

lorsque la pièce est truquée. En revanche, la même démarche est une impasse lorsqu'on se place sur $\Omega = \{P, F\}^{\mathbb{N}}$: Ω est infini - non dénombrable - puisque card Ω = card $\mathbb{R}$. Or, la construction d'un modèle infini de parties de pile ou face est indispensable dès que l'on s'intéresse à des questions aussi naturelles que, par exemple, la

loi de première apparition de Pile lors d'une partie indéfinie de Pile ou Face. Plus généralement, l'étude de problèmes asymptotiques (loi des grands nombres, théorème Central-Limite,..), même à partir d'une modélisation discrète, impose l'introduction d'espaces Ω de grande taille et de v.a. prenant surjectivement leurs valeurs dans (un intervalle de) $\mathbb{R}$.

3.2 Rappels de théorie de la mesure

3.2.1 Tribus, espaces mesurables, boréliens

Définition :

Soit E un ensemble et $\mathcal{E} \subset \mathcal{P}(E)$. $\mathcal{E}$ est une *tribu* (ou σ - algèbre) si :

$$\begin{array}{ll} (i) & \phi \in \mathcal{E}, \\ (ii) & A \in \mathcal{E} \Longrightarrow {}^cA \in \mathcal{E}, \\ (iii) & A_n \in \mathcal{E},\ n \geq 1, \text{ alors } \bigcup_n A_n \in \mathcal{E}. \end{array}$$

En particulier une tribu est stable par intersection dénombrable (lois de Morgan). On appelle *espace mesurable* un espace $(E, \mathcal{E})$ muni d'une tribu.

Exemples : • $\mathcal{P}(E)$, $\{\phi, E\}$, $\{\phi, A, {}^cA, E\}$, $A \subset E$ fixé.
• $\{A \subset E$, A ou ${}^cA$ dénombrable$\}$ ($\neq \mathcal{P}(E)$ ssi E est non dénombrable).
• $\mathcal{E}_i$, $i \in I$, $I \neq \phi$, tribus, alors $\bigcap_{i \in I} \mathcal{E}_i$ est une tribu.

Du dernier exemple, on déduit l'existence de la plus petite tribu, notée $\sigma(\mathcal{C})$, contenant une famille $\mathcal{C} \subset \mathcal{P}(E)$ de parties de E : $\sigma(\mathcal{C}) = \bigcap_{\substack{\mathcal{E} \supset \mathcal{C} \\ \mathcal{E}\ tribu}} \mathcal{E}$.

Exemple fondamental : Si E est un espace topologique dont les ouverts forment la famille $\mathcal{O}_E$, on appelle *tribu borélienne de* E la tribu $\mathcal{B}(E) := \sigma(\mathcal{O}_E)$ ($\mathcal{B}(E)$ est évidemment aussi engendrée par les fermés de E).

Dès que E est *métrique séparable*, on vérifie que $\mathcal{B}(E) := \sigma\Big(B(x_n, r),\ n \in \mathbb{N}, r \in \mathbf{Q}_+^*\Big)$ où $(x_n)_{n \in \mathbb{N}}$ est une suite dense dans E. Dans la pratique on rencontre surtout les cas où $E := \mathbb{R}$, $\mathbf{C}$, $\mathbb{R}^d$, $\mathbb{R}_+$, $\overline{\mathbb{R}}$, $\overline{\mathbb{R}}_+$, etc.

Ainsi $\mathcal{B}\left(\mathbb{R}^d\right) = \sigma\left(\prod_{i=1}^d [a_i, b_i],\ \ a_i, b_i \in \mathbf{Q}\right) = \sigma\left(\prod_{i=1}^d]a_i, +\infty[\,,\ a_i \in \mathbf{Q}\right)$.

3.2.2 Applications mesurables

Rappelons d'abord que, si $f : E \to F$ et $B \subset F$ on note souvent $\{f \in B\}$ pour :

$$f^{-1}(B) := \{x / f(x) \in B\}.$$

Définition :

Soient $(E,\ \mathcal{E})$ et $(F,\ \mathcal{F})$ deux espaces mesurables et $f : E \to F$.
On dira que f est *mesurable* si $\forall B \in \mathcal{F}, f^{-1}(B) = \{f \in B\} \in \mathcal{E}$.
En Probabilités, f est une *variable aléatoire*.

On vérifie immédiatement par les formules de Hausdorff que :

$$\sigma(f) := \{f^{-1}(B), B \in \mathcal{F}\} \text{ est une tribu sur } E.$$

On l'appelle *tribu engendrée par* f car c'est la plus petite tribu sur E rendant f mesurable. On la note aussi parfois $f^{-1}(\mathcal{F})$.

Proposition :

(a) $f : (E,\mathcal{E}) \to (F,\mathcal{F})$ est une variable aléatoire (*i.e.* mesurable) ssi $\sigma(f) \subset \mathcal{E}$

(b) Si $\mathcal{F} = \sigma(\mathcal{C})$ où $\mathcal{C} \subset \mathcal{P}(F)$, alors f est une v.a. ssi $\forall C \in \mathcal{C},\ \{f \in C\} \in \mathcal{E}$.

Démonstration : (a) est évident.
(b) : $\mathcal{T} := \{B \subset F \ /\ f^{-1}(B) \in \mathcal{E}\}$ est une tribu ; or, $\mathcal{T} \supset \mathcal{C}$, donc $\mathcal{T} \supset \sigma(\mathcal{C}) = \mathcal{F}$ donc $f^{-1}(\mathcal{F}) \subset \mathcal{E}$. Le sens direct est immédiat. #

Corollaire (important) :

Soit $f : (E,\mathcal{E}) \to (F,\ \mathcal{B}(F))$ (F espace topologique muni de ses boréliens).

$$f \text{ est mesurable ssi } \forall O \in \mathcal{O}_F,\ f^{-1}(O) \in \mathcal{E}.$$

En particulier si E est aussi un espace topologique et $\mathcal{E} = \mathcal{B}(E)$ alors :

$$f \text{ continue} \Longrightarrow f \text{ mesurable.}$$

Applications :

(a) $i : \mathbb{R} \hookrightarrow \overline{\mathbb{R}}$ est continue donc mesurable (borélienne).
Donc $i^{-1}\left(\mathcal{B}\left(\overline{\mathbb{R}}\right)\right) \subset \mathcal{B}(\mathbb{R})$. Or $i^{-1}\left(\mathcal{B}\left(\overline{\mathbb{R}}\right)\right) = \left\{B \cap \mathbb{R},\ B \in \mathcal{B}\left(\overline{\mathbb{R}}\right)\right\}$. D'autre part $\mathbb{R} \in \mathcal{O}_{\overline{\mathbb{R}}} \subset \mathcal{B}(\overline{\mathbb{R}})$. Comme l'ensemble des boréliens de $\mathbb{R}$ qui sont des boréliens de $\overline{\mathbb{R}}$ est une tribu et que cette tribu contient donc tous les ouverts de $\mathbb{R}$, il ne peut s'agir que de la tribu borélienne *i.e.* $\mathcal{B}(\mathbb{R}) \subset \mathcal{B}\left(\overline{\mathbb{R}}\right)$. En outre, $\{+\infty\}$ et $\{-\infty\}$ sont des fermés de $\overline{\mathbb{R}}$. D'où

$$\mathcal{B}\left(\overline{\mathbb{R}}\right) = \{A,\ A \cup \{\pm\infty\},\ A \cup \{+\infty\},\ A \cup \{-\infty\},\ A \in \mathcal{B}(\mathbb{R})\}.$$

On a donc aussi :

$$\mathcal{B}\left(\overline{\mathbb{R}}\right) = \sigma(]a, +\infty], a \in \mathbb{R}) = \sigma([a, +\infty],\ a \in \mathbb{R}).$$

(b) $f : (E,\mathcal{E}) \longrightarrow \mathbb{R}$ ou $\overline{\mathbb{R}}$ est mesurable ssi $\forall a \in \mathbb{R},\ \{f \geq a\} \in \mathcal{E}$.

Propriétés (admises ici) :

P1 $f = (f_1,\ f_2, ..., f_d) : (E,\mathcal{E}) \longrightarrow \left(\mathbb{R}^d,\ \mathcal{B}\left(\mathbb{R}^d\right)\right)$ est mesurable si et seulement si toutes les f_i le sont.

Par suite, *sommes* et *produits* de fonctions mesurables le sont (par continuité de l'addition et de la multiplication de $\mathbb{R}^2$ dans $\mathbb{R}$).

P2 Si les $f_n : (E,\mathcal{E}) \longrightarrow (\overline{\mathbb{R}},\ \mathcal{B}\left(\overline{\mathbb{R}}\right))$, $n \geq 1$, sont mesurables alors $\sup\limits_n f_n$ et $\inf\limits_n f_n$ sont mesurables.

P3 Soit $f_n : (E,\mathcal{E}) \longrightarrow (\mathbb{K}, \mathcal{B}(\mathbb{K}))$, $n \geq 1$, une suite de fonctions mesurables $\left(\mathbb{K} = \mathbb{R} \text{ ou } \overline{\mathbb{R}}\right)$, alors :

(a) $\overline{\lim\limits_n}\ f_n := \inf\limits_n \sup\limits_{k\geq n} f_k$ et $\underline{\lim\limits_n}\ f_n := \sup\limits_n \inf\limits_{k \geq n} f_k$ sont mesurables de E dans $\overline{\mathbb{R}}$.

(b) $\{f_n \longrightarrow .\} := \left\{x \in E / f_n(x) \text{ converge dans } \overline{\mathbb{R}}\right\} = \left\{\overline{\lim\limits_n}\ f_n = \underline{\lim\limits_n}\ f_n\right\} \in \mathcal{E}$.

(c) L'application $f := \overline{\lim\limits_n}\ f_n = \underline{\lim\limits_n}\ f_n = \lim\limits_n f_n$ définie de $\{f_n \longrightarrow .\}$, muni de la tribu trace $\{A \cap \{f_n \longrightarrow\ .\ \},\ A \in \mathcal{E}\}$, dans $\left(\overline{\mathbb{R}}, \mathcal{B}\left(\overline{\mathbb{R}}\right)\right)$ est mesurable.

Liens entre ensembles et applications mesurables :

On définit pour $A \subset E$, $\mathbf{1}_A(x) = \begin{cases} 1 \text{ si } x \in A \\ 0 \text{ si } x \notin A \end{cases}$. On rappelle que

$$\mathbf{1}_{\cap_n A_n} = \inf_n \mathbf{1}_{A_n},\quad \mathbf{1}_{\cup_n A_n} = \sup_n \mathbf{1}_{A_n},$$
$$\mathbf{1}_{\overline{\lim}_n A_n} = \overline{\lim_n}\ \mathbf{1}_{A_n} \text{ et } \mathbf{1}_{\underline{\lim}_n A_n} = \underline{\lim_n}\ \mathbf{1}_{A_n}.$$

Enfin, on a évidemment

$$A \in \mathcal{E} \quad \text{ssi} \quad \mathbf{1}_A \text{ est mesurable.}$$

Fonctions étagées sur $(E,\mathcal{E})$:

Définition :

$f : (E,\mathcal{E}) \longrightarrow \mathbb{R}$ ou $\mathbf{C}$ est dite étagée si elle est mesurable et ne prend qu'un nombre fini de valeurs (finies) ; f peut donc s'écrire :
$f := \sum\limits_{i \in I} \alpha_i \mathbf{1}_{A_i},\ A_i \neq \phi,\ A_i \in \mathcal{E},\ A_i \cap A_j = \phi,\ i \neq j,\ \cup_{i\in I} A_i = E,\ \alpha_i \in \mathbb{R}$ ou $\mathbf{C}$, I fini.

En d'autres termes, $(A_i)_{i\in I}$ est une partition $\mathcal{E}$-mesurable de E.

L'ensemble des fonctions étagées sur $(E,\mathcal{E})$ est une algèbre (sur $\mathbb{R}$ ou $\mathbf{C}$) réticulée (*i.e.* stable par min et max).

Lemme fondamental :

Soit $g : (E,\mathcal{E}) \longrightarrow \overline{\mathbb{R}}_+$. Il existe une suite croissante de fonctions positives étagées convergeant vers g. Si g est bornée la convergence est uniforme.

Démonstration : On vérifie que $g_n = \sum_{k=0}^{n2^n-1} \frac{k}{2^n}\mathbf{1}_{\left\{\frac{k}{2^n}\le g<\frac{k+1}{2^n}\right\}} + n\mathbf{1}_{\{g\,\ge\, n\}}$ convient.
Si $g < N$, alors, dès que $n \ge N$, $0 \le g - g_n \le \frac{1}{2^n}$. #

En décomposant g en $g = g^+ - g^-$, on montre que toute fonction mesurable à valeurs dans $\overline{\mathbb{R}}$ est limite de fonctions étagées. On étend aux fonctions complexes en décomposant $g = g_1 + ig_2$.

Application : Soit $X : (E, \sigma(X)) \longrightarrow (F, \mathcal{F})$, $f : (E, \sigma(X)) \longrightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$.
Si f est $\sigma(X)$-mesurable alors il existe une fonction $g\,(\mathcal{F}, \mathcal{B}(\mathbb{R}))$-mesurable telle que $f = g(X)$.

Démonstration :
Comme $\sigma(X) = \{(X\in A),\ A\in\mathcal{B}(\mathbb{R})\}$ $[(X\in A) := X^{-1}(A)]$, si f est une indicatrice $f = \mathbf{1}_{(X\in A)} = \mathbf{1}_A \circ X = g\circ X$.

Si f est étagée positive $f = \left(\sum_i \alpha_i \mathbf{1}_{A_i}\right)\circ X$. Puis, si $f_n \nearrow f$ avec $f_n = g_n \circ X$, on a $f = g\circ X$ avec $g := \overline{\lim_n}\, g_n$, etc. #

3.3 Mesures positives, mesures finies, probabilités

Définition :
• On appelle mesure positive sur un espace mesurable $(E,\ \mathcal{E})$ toute application :
$\mu : \begin{cases} \mathcal{E} \longrightarrow \overline{\mathbb{R}}_+ \\ A \longrightarrow \mu(A) \end{cases}$ vérifiant :

(i) $\mu(\phi) = 0$,
(ii) Si $A_n \in \mathcal{E}, n\ge 1$, avec $A_i \cap A_j = \phi$ si $i \ne j$, alors $\mu\left(\bigcup_{n\ge1} A_n\right) = \sum_{n\ge1} \mu(A_n)$.

• Si $\mu(E) < +\infty$, μ est dite *finie* (ou *bornée*)

• Si $\mu(E) = 1$, μ est appelée (mesure de) *probabilité*.

Exemples : • $\mu = 0$ (mesure nulle).
• $\mu(A) = +\infty$ si $A \ne \phi$, $\mu(\phi) = 0$ (mesure grossière).
• Pour $a\in E$ on définit la mesure de Dirac, notée δ_a, par $\delta_a(A) = 1$ si $a\in A$, $\delta_a(A) = 0$ si $a \notin A$.

• Mesure de décompte sur $(E, \mathcal{P}(E))$: $\forall A \in \mathcal{P}(E)$, $\mu(A) = |A|$.
• Mesure de Lebesgue sur $\left(\mathbb{R}^d, \mathcal{B}\left(\mathbb{R}^d\right)\right)$: Il existe une unique mesure sur $\left(\mathbb{R}^d, \mathcal{B}\left(\mathbb{R}^d\right)\right)$, dite mesure de Lebesgue et notée λ_d, vérifiant :

(i) $\lambda_d\left([0,1]^d\right) = 1$,
(ii) $\forall a\in\mathbb{R}^d, \forall A\in\mathcal{B}\left(\mathbb{R}^d\right), \lambda_d(a+A) = \lambda_d(A)$ où $a + A := \{a+x, x\in A\}$.

Pour tout hypercube on montre alors que $\lambda_d\left(\prod_{i=1}^d (a_i, b_i)\right) = \prod_{i=1}^d (b_i - a_i)$.

Définition :
Une partie $N \subset E$ est μ-négligeable si et seulement si il existe $A \in \mathcal{E}$ vérifiant $N \subset A$ et $\mu(A) = 0$. Une propriété vraie sauf sur un ensemble négligeable est vraie μ-presque partout (μ-*p.p.*). Si μ est une *probabilité*, on parle alors de μ-*presque sûrement* (μ-*p.s.*).

Propriétés :
P1 Si $A \subset B$, alors $\mu(A) \le \mu(B)$ et $\mu(B\setminus A) = \mu(B) - \mu(A)$ si $\mu(A) < +\infty$.
P2 $A, B \in \mathcal{E}$, $\mu(A\cup B) + \mu(A\cap B) = \mu(A) + \mu(B)$.
P3 $A_n \in \mathcal{E}, n\ge1, A_n \subset A_{n+1}$, alors $\mu\left(\bigcup_n A_n\right) = \lim_n\uparrow \mu(A_n)$.
P4 $A_n \in \mathcal{E}, n\ge1, A_{n+1}\subset A_n$ et $\underline{\mu(A_1) < +\infty}$ alors $\mu\left(\bigcap_n A_n\right) = \lim_n\downarrow \mu(A_n)$
P5 $A_n \in \mathcal{E}, n\ge 1, \mu\left(\bigcup_n A_n\right) \le \sum_n \mu(A_n) \le +\infty$.

3.3.1 Caractérisation d'une mesure, unicité

Les tribus sont généralement trop "riches" pour être décrites exhaustivement. On est donc contraint de procéder par des voies détournées pour établir, par exemple, que deux mesures sont égales.

Définition :
On appelle λ-système toute famille Λ de parties de E vérifiant :

(i) $\phi \in \Lambda$
(ii) $A_n \in \Lambda, n\ge1, A_n \subset A_{n+1}$, alors $\bigcup_n^{\uparrow} A_n \in \Lambda$
(iii) $A, B \in \Lambda, A\subset B$ alors $B\setminus A \in \Lambda$ (stabilité par *différence propre*)

On vérifie aussitôt que si $\mathcal{C} \subset \mathcal{P}(E)$, il existe un plus petit λ-système $\Lambda(\mathcal{C})$ contenant $\mathcal{C}$. En effet, $\mathcal{P}(E)$ est un λ-système contenant $\mathcal{C}$ donc $\bigcap_{\substack{\Lambda\supset\mathcal{C} \\ \Lambda\ \lambda\text{- système}}} \Lambda$ existe et c'est évidemment un λ-système contenant $\mathcal{C}$. C'est forcément le plus petit.

Lemme : Si $E \in \Lambda$ et Λ est stable par intersection finie alors Λ est une tribu.

Démonstration : On vérifie d'abord que Λ est stable par différence simple : si $A, B \in \Lambda$, $B \setminus A = B \setminus (A \cap B) \in \Lambda$. Comme $E \in \Lambda$, Λ est stable par complémentaire. Comme $\bigcup_n A_n = \bigcup_n \uparrow \left(\bigcup_{k=1}^n A_k \right)$, il suffit d'établir la stabilité par réunion finie qui découle elle-même de la stabilité par complémentaire et intersection finie. #

Proposition :
Soit $\mathcal{C} \subset \mathcal{P}(E)$ vérifiant : $\begin{cases} (i)\, E \in \mathcal{C} \\ (ii)\, A, B \in \mathcal{C} \Rightarrow A \cap B \in \mathcal{C} \end{cases}$. Alors : $\Lambda(\mathcal{C}) \supset \sigma(\mathcal{C})$.

Démonstration : Il suffit de vérifier, au vu du lemme, que $\Lambda(\mathcal{C})$ est stable par intersection finie (puisque $E \in \mathcal{C} \subset \Lambda(\mathcal{C})$). Soit $C \in \mathcal{C}$ et $\Lambda_C := \{A \in \Lambda(\mathcal{C}) / A \cap C \in \Lambda(\mathcal{C})\}$. On vérifie que Λ_C est un λ-système contenant $\mathcal{C}$ donc $\Lambda_C = \Lambda(\mathcal{C})$.

Soit maintenant $D \in \Lambda(\mathcal{C})$. Λ_D est également un λ-système et $\Lambda_D \supset \mathcal{C}$ d'après ce qui précède donc $\Lambda_D = \Lambda(\mathcal{C})$. Finalement, $\Lambda(\mathcal{C})$ est stable par intersection finie et contient E, c'est donc une tribu qui contient $\mathcal{C}$ et par conséquent $\sigma(\mathcal{C})$. #

Application :
Si μ_1 et μ_2, *probabilités* sur $(E, \mathcal{E})$, coïncident sur une famille de parties $\mathcal{C}$ stable par intersection telle que $\mathcal{E} = \sigma(\mathcal{C})$, alors $\mu_1 = \mu_2$.

Démonstration : On peut supposer que $E \in \mathcal{C}$ puisque $\mu_1(E) = \mu_2(E) = 1$ et on vérifie que $\Lambda := \{A \in \mathcal{E} / \mu_1(A) = \mu_2(A)\}$ est un λ-système. #

Exemple : Si $(E, \mathcal{B}(E))$ est un espace topologique muni de ses boréliens, μ_1 et μ_2 deux mesures finies. Alors :

$$\mu_1 = \mu_2 \text{ sur } \mathcal{O}_E \Rightarrow \mu_1 = \mu_2.$$

Remarque : L'application ci-dessus s'étend au cas de mesures "σ-finies", *i.e.* telles qu'il existe $E_n \in \mathcal{C}$, $n \geq 1$, vérifiant $E = \bigcup_{n \geq 1} E_n$ et $\mu_1(E_n) = \mu_2(E_n) < +\infty$ (il suffit d'introduire les $\mu_i^{(n)}(.) := \mu_i(\cdot \cap E_n)$).

Exemple : Il existe au plus une mesure λ_d sur $\left(\mathbb{R}^d, \mathcal{B}\left(\mathbb{R}^d\right)\right)$ vérifiant

$$\lambda_d\left(\prod_{i=1}^d]a_i, b_i[\right) = \prod_{i=1}^d (b_i - a_i)$$

car $\mathcal{C} = \left\{\prod_{i=1}^d]a_i, b_i[\,, a_i,\ b_i \in \mathbf{Q}\right\}$ est stable par intersection finie, engendre $\mathcal{B}\left(\mathbb{R}^d\right)$ et contient les $E_n :=]-n, n[^d$, $n \geq 1$.

3.3.2 Un théorème de prolongement

On rappelle qu'une famille $\mathcal{C}$ de parties de E est une algèbre de Boole sur E si elle contient E et si elle est stable par complémentaire et par réunion *finie*.

Théorème (de Carathéodory, admis) :
Soit μ une fonction définie sur une algèbre de Boole $\mathcal{C} \subset \mathcal{P}(E)$, à valeurs dans $\mathbb{R}_+$, vérifiant :

$(i)\ \mu(\phi) = 0,$
$(ii)\ A, B \in \mathcal{C},\ A \cap B = \phi \Rightarrow \mu(A \cup B) = \mu(A) + \mu(B),$
$(iii)\ A_n \in \mathcal{C},\ n \geq 1,\ A_{n+1} \subset A_n,\ \bigcap_n \downarrow A_n = \phi \Rightarrow \mu(A_n) \to 0.$

Alors, *il existe une unique mesure finie* $\tilde{\mu}$ sur $\sigma(\mathcal{C})$ prolongeant μ.

Exemple $(d = 1)$:
On considère $\mathcal{C} := \{I_1 \cup ... \cup I_n,\ n \geq 1,\ I_k$ intervalle de $[0, 1]$, 2 à 2 disjoints$\}$ et $\lambda(I_1 \cup ... \cup I_n) = \sum_{1 \leq k \leq n} long(I_k)$. $\mathcal{C}$ est une algèbre, $\sigma(\mathcal{C}) = \mathcal{B}([0, 1])$ et λ se prolonge en une mesure sur $([0, 1], \mathcal{B}([0, 1]))$: la mesure de Lebesgue sur $[0, 1]$.

3.4 Rappels de théorie de l'intégration

3.4.1 Construction (ébauche)

Soit $(E, \mathcal{E}, \mu)$ un espace mesurable.

• Soit $f = \sum_{i=1}^n \alpha_i \mathbf{1}_{A_i}$ une fonction étagée positive $\left(\alpha_i \geq 0, (A_i)_{1 \leq i \leq n}\right.$ partition $\mathcal{E}$-mesurable de $\left.E\right)$.
Par définition on pose :

$$\int f d\mu := \sum_{\alpha \in f(E)} \alpha \mu(\{f = \alpha\}) = \sum_{i=1}^n \alpha_i \mu(A_i)$$

avec la convention $0 \times (\pm\infty) = 0$ (utile si $\mu(\{f = 0\}) = +\infty$).

• Pour une fonction mesurable positive $f : (E, \mathcal{E}) \longrightarrow \left(\overline{\mathbb{R}}_+, \mathcal{B}\left(\overline{\mathbb{R}}_+\right)\right)$, on pose :

$$\int f d\mu := \sup_{\substack{g \leq f \\ g \text{ étagée}}} \int g d\mu \in \overline{\mathbb{R}}_+.$$

On vérifie que, si $\alpha,\ \beta \in \mathbb{R}_+$ et f, g sont mesurables positives,

$$\int (\alpha f + \beta g)\, d\mu = \alpha \int f d\mu + \beta \int g d\mu$$

• Si f est à valeurs dans $\overline{\mathbb{R}}$, on dira que f est μ-intégrable si et seulement si $\int |f|\, d\mu < +\infty$. Dans ce cas $\int f^{\pm} d\mu < +\infty$ et l'on pose

$$\int f d\mu := \int f^{+} d\mu - \int f^{-} d\mu.$$

• Si f est à valeurs complexes, on pose $\int f d\mu := \int Re\,(f)\, d\mu + i \int Im\,(f)\, d\mu$ dès que $Re\,(f)$ et $Im\,(f)$ sont intégrables (équivalent à $\int |f|\, d\mu < +\infty$).

On note $\mathcal{L}^1_{\mathbb{K}}(\mu) := \{f : (E, \mathcal{E}) \longrightarrow \mathbb{K},\ \mu\text{-intégrable}\}$, ($\mathbb{K} = \mathbb{R}$, $\overline{\mathbb{R}}$, $\mathbb{R}_+$, $\mathbf{C}$).

Proposition : (*a*) On vérifie que si f est intégrable alors $\mu(\{|f| = +\infty\}) = 0$ (Attention ! La réciproque est *fausse*).

(*b*) D'autre part, si $f = g$ μ-*p.p.* (*i.e.* $\mu(\{f \neq g\}) = 0$), $\int f d\mu = \int g d\mu$ dès que l'une des deux intégrales a un sens.

(*c*) L'application $f \mapsto \int f d\mu$ est une forme linéaire de $\mathcal{L}^1_{\mathbb{K}}(\mu)$.

(*d*) Inégalité triangulaire : $\left|\int f d\mu\right| \leq \int |f|\, d\mu$ [égalité ssi $|f(x)| = \lambda f(x)$ μ-*p.p.* pour un $\lambda \in \mathbb{R}$ ou $\mathbf{C}$]

Notation : On note aussi $\int f d\mu = \int f(x)\, d\mu(x) = \int f(x)\, \mu(dx)$.

Définition :

Si $\mu(E) = 1$ on note $\mathbb{E}(f)$ au lieu de $\int f d\mu$, $\mathbb{E}$ pour *espérance* mathématique.

Rappel des théorèmes essentiels :

• Théorème de Beppo-Levi (ou convergence croissante) :

$f_n : (E, \mathcal{E}) \to \overline{\mathbb{R}}_+$, $n \geq 1$, $f_n \nearrow f$, alors $\int f_n d\mu \nearrow \int f d\mu \leq +\infty$ quand $n \to +\infty$.

• Lemme de Fatou :

$f_n : (E, \mathcal{E}) \to \overline{\mathbb{R}}_+$, $n \geq 1$. Alors $\int \underline{\lim}_n f_n d\mu \leq \underline{\lim}_n \int f_n d\mu$.

• Théorème de Lebesgue (convergence dominée) :

$f_n : (E, \mathcal{E}) \to \mathbb{K}$, $n \geq 1$ (avec $\mathbb{K} = \mathbb{R}$ ou $\mathbf{C}$), vérifiant :

$$\begin{cases} (i) & \mu(dx)\text{-}p.p. \quad f_n(x) \to .\,, \\ (ii) & \mu(dx)\text{-}p.p. \quad |f_n(x)| \leq g(x)\,,\ g \in \mathcal{L}^1_{\mathbb{R}_+}(\mu)\,, \end{cases}$$

$$\text{Alors : } \begin{cases} (\alpha) & \exists f \in \mathcal{L}^1_{\mathbb{K}}(\mu) \text{ tq } f_n \to f\ \mu\text{-}p.p., \\ (\beta) & \int |f_n - f|\, d\mu \underset{n \to +\infty}{\longrightarrow} 0. \end{cases}$$

3.4.2 Inégalités de Hölder et Minkowski, espaces $\mathcal{L}^p$ et $\mathbf{L}^p, 1 \leq p \leq +\infty$.

• Si $1 \leq p < +\infty$, $\mathcal{L}^p_{\mathbb{K}}(\mu) := \{f : (E, \mathcal{E}) \to \mathbb{K} \text{ v.a. } / \int |f|^p d\mu < +\infty\}$ est un $\mathbb{K}$-e.v. semi-normé par $\|f\|_p = (\int |f|^p d\mu)^{1/p}$ ($\mathbb{K} = \mathbb{R}$ ou $\mathbf{C}$).

• Si $p = +\infty$ (et $\mu \neq 0$), on définit $\|f\|_\infty := \inf\{\alpha / \mu(\{|f| > \alpha\}) = 0\}$ et $\mathcal{L}^\infty_{\mathbb{K}}(\mu) := \{f : (E, \mathcal{E}) \to \mathbb{K} \ /\ \|f\|_\infty < +\infty\}$. $\mathcal{L}^\infty_{\mathbb{K}}(\mu)$ est aussi un $\mathbb{K}$-e.v. semi-normé.

On quotiente classiquement ces espaces par le "noyau" de la semi-norme qui correspond à chaque fois à la relation d'équivalence : $f \sim g$ si et seulement si $f - g = 0$ μ-*p.p.*.

On note ces quotients $L^p_{\mathbb{K}}(\mu) := \{$classes de v.a. modulo l'égalité μ-*p.p.*$\}$

Théorème :

(a) $1 \leq p \leq +\infty$, $L^p_{\mathbb{K}}(\mu)$ est un $\mathbb{K}$-espace de Banach.

(b) Si $p = 2$, $L^p_{\mathbb{K}}(\mu)$ est un $\mathbb{K}$-espace de Hilbert pour le produit scalaire

$$(f, g) := \int f \overline{g} d\mu.$$

(c) Si μ *est une probabilité*, on a : $\mathcal{L}^\infty_{\mathbb{K}} \subset \mathcal{L}^q_{\mathbb{K}} \subset \mathcal{L}^p_{\mathbb{K}} \subset \mathcal{L}^1_{\mathbb{K}}$ si $1 \leq p \leq q \leq +\infty$.

Le fait que les $\|\cdot\|_p$ soient des (semi-) normes repose essentiellement sur les inégalités importantes suivantes (valables tant dans les $\mathcal{L}^p(\mu)$ que dans les $L^p(\mu)$) :

• $\forall p, q \in [1, +\infty]$, $\dfrac{1}{p} + \dfrac{1}{q} = 1$, $\forall f \in \mathcal{L}^p_{\mathbb{K}}(\mu), \forall g \in \mathcal{L}^q_{\mathbb{K}}(\mu)$

$$fg \in \mathcal{L}^1_{\mathbb{K}}(\mu) \text{ et } \|fg\|_1 \leq \|f\|_p \|g\|_q$$

[égalité si et seulement si $\alpha |f|^p(x) = \beta |g|^q(x)\, \mu(dx)$-*p.p.*, $(\alpha, \beta) \in \mathbb{R}^2_+ \setminus \{(0,0)\}$].

• $\forall p \in [1, +\infty], \forall f, g \in \mathcal{L}^p_{\mathbb{K}}$ $\|f + g\|_p \leq \|f\|_p + \|g\|_p$.

Complément : inégalité de Jensen.

Soit $(E, \mathcal{E}, \mu)$, μ *probabilité*, $\varphi : \mathbb{R} \to \mathbb{R}$ convexe et $f \in \mathcal{L}^1_{\mathbb{R}}(\mu)$ telle que $\varphi(f) \in \mathcal{L}^1_{\mathbb{R}}(\mu)$. Alors :

$$\varphi(\mathbb{E}_\mu(f)) \leq \mathbb{E}_\mu(\varphi(f))$$

Ex : $\varphi(x) = |x|$, x^2, e^x, $-\ln(x)$, ...

3.5 Compléments de théorie de la mesure (caractérisation, mesure image...)

3.5.1 Retour sur la caractérisation d'une mesure

Proposition :
Soit (E,d) un espace *métrique* et μ_1,μ_2 deux mesures sur $\mathcal{B}(E)$.
(a) Si μ_1 et μ_2 sont *finies*,

$$\left(\forall f\in\mathcal{C}_b(E):=\{f:E\to\mathbb{R}\text{ continues bornées}\}\int fd\mu_1=\int fd\mu_2\right)\Rightarrow\mu_1=\mu_2.$$

(b) Soit E localement compact (tout point admet un voisinage compact), dénombrable à l'infini $\left(E=\bigcup_{n\geq 1}K_n, K_n\text{ compact}\right)$. Si μ_1 et μ_2 sont *finies sur les compacts* et

$$\forall f\in\mathcal{C}_K(E):=\{g:E\to\mathbb{R},\text{ continues à support compact}\},\ \int fd\mu_1=\int fd\mu_2,$$

alors $\mu_1=\mu_2$.

Remarque : L'exemple typique d'espace localement compact dénombrable à l'infini est évidemment $\mathbb{R}^d$ (en revanche $\mathbb{R}^{\mathbb{N}}$ n'est pas localement compact).

Démonstration :
(a) Soit $O\in\mathcal{O}_E$. On pose $O_\varepsilon:=\{x\in E/d(x,{}^cO)>\varepsilon\}$ et $f_\varepsilon(x):=\dfrac{d(x,{}^cO)}{d(x,{}^cO)+d(x,O_\varepsilon)}$; $f_\varepsilon\in\mathcal{C}(E,[0,1])$ et $f_\varepsilon=0$ sur ${}^cO$ et $f_\varepsilon=1$ sur $O_\varepsilon\subset O$. Quand $\varepsilon\downarrow 0, f_\varepsilon(x)\uparrow\mathbf{1}_O(x)$.
D'après le théorème de Beppo-Levi, on a donc $\mu_1(O)=\mu_2(O)$, d'où (cf. section 3.a) $\mu_1=\mu_2$.

(b) C'est un petit exercice classique de topologie de montrer l'existence d'une suite de compacts $K_n, n\geq 1$, vérifiant $E=\bigcup_{n\geq 1}K_n$ et $K_n\subset\overset{o}{K}_{n+1}$. On considère $\mu_i^{(n)}:=\mu_i\left(\cdot\cap\overset{o}{K}_{n+1}\right)$. Soit $O\in\mathcal{O}_E$ et $O^{(n)}:=O\cap\overset{o}{K}_{n+1}$. Il est clair que les mesures $\mu_i^{(n)}$ sont finies et que $\mu_i^{(n)}(O)=\mu_i\left(O^{(n)}\right)$. Or les $f_\varepsilon^{(n)}$ relatives à $O^{(n)}$ sont à support dans le compact $\overline{O^{(n)}}\subset K_{n+1}$, donc, procédant comme en (a), on obtient que $\mu_1^{(n)}(O)=\mu_1\left(O^{(n)}\right)=\mu_2\left(O^{(n)}\right)=\mu_2^{(n)}(O)$.
Donc $\mu_1^{(n)}=\mu_2^{(n)}$. On conclut en notant que $\mu_i(A)=\lim_n\uparrow\mu_i^{(n)}(A)$. #

Corollaire : Si μ_1 et μ_2 sont deux mesures sur $(\mathbb{R}^d,\mathcal{B}(\mathbb{R}^d))$, finies sur les compacts, alors :

$$\forall f\in\mathcal{C}_K(\mathbb{R}^d,\mathbb{R}),\int fd\mu_1=\int fd\mu_2\implies\mu_1=\mu_2.$$

3.5.2 Mesure image

Nous allons décrire une méthode permettant de transporter une mesure d'un espace sur un autre à l'aide d'une application.
Nous avons vu dans le cadre élémentaire du chapitre 1 que cette idée est à la base de la notion de loi d'une v.a. lorsque la mesure initiale est une probabilité.

Proposition :
Soit $(E,\mathcal{E})$ et $(F,\mathcal{F})$ deux espaces mesurables, $h:(E,\mathcal{E})\to(F,\mathcal{F})$ une application *mesurable* et μ une mesure sur $(E,\mathcal{E})$. L'application $\nu:\begin{cases}\mathcal{F}\to\overline{\mathbb{R}}_+\\ B\to\nu(B):=\mu(h^{-1}(B))\end{cases}$ est une mesure sur $(F,\mathcal{F})$ de même masse totale que μ. En particulier ν est une probabilité si μ l'est. On note $\underline{\nu:=h(\mu)}$ ou $\underline{\mu_h}$ selon les cas.

Démonstration : $\nu(\phi)=\mu(h^{-1}(\phi))=\mu(\phi)=0$. Soient $B_n\in\mathcal{F},\ n\geq 1$, avec $B_i\cap B_j=\phi$ si $i\neq j$. Les $h^{-1}(B_n)$ sont évidemment 2 à 2 disjoints et $\bigcup_n h^{-1}(B_n)=h^{-1}\left(\bigcup_n B_n\right)$ donc :

$$\nu\left(\bigcup_n B_n\right)=\mu\left(h^{-1}\left(\bigcup_n B_n\right)\right)=\mu\left(\bigcup_n h^{-1}(B_n)\right)=\sum_n\mu\left(h^{-1}(B_n)\right)=\sum_n\nu(B_n).$$

Enfin : $\nu(F)=\mu(h^{-1}(F))=\mu(E)$. #

Remarques : • En notation probabiliste on écrit $\nu(B):=\mu(\{h\in B\})$.
• On peut noter que l'on peut définir ν sur une tribu a priori plus grande que $\mathcal{F}$: *la tribu image* de $\mathcal{E}$ par h définie par $\{B\in\mathcal{F}/h^{-1}(B)\in\mathcal{E}\}$. Cette tribu est en fait la plus grande tribu sur F rendant h mesurable comme fonction sur $(E,\mathcal{E})$.

Définition :
La mesure μ_h est appelée la *mesure image* de μ par h. Si μ est une probabilité, μ_h est appelée *loi de* h (sous μ).

Théorème :
Avec les notations de la proposition précédente, $f:(F,\mathcal{F})\to\mathbb{K}$ est μ_h-intégrable si et seulement si $f\circ h:(E,\mathcal{E})\to\mathbb{K}$ est μ-intégrable et $\int fd\mu_h=\int f\circ hd\mu$.
L'égalité a toujours lieu si f est positive.

Démonstration :
Si $f=\mathbf{1}_B, B\in\mathcal{F}$, $\int fd\mu_h=\mu_h(B)=\mu\left(h^{-1}(B)\right)=\int\mathbf{1}_{h^{-1}(B)}d\mu$.
Or $\mathbf{1}_{h^{-1}(B)}=\mathbf{1}_B\circ h$ donc $\int fd\mu_h=\int\mathbf{1}_B\circ hd\mu=\int f\circ hd\mu$.

L' égalité s'étend aux fonctions étagées positives par linéarité, aux fonctions mesurables positives par le théorème de Beppo-Levi (et le lemme fondamental d'approximation) puis aux fonctions réelles et complexes par décomposition *ad hoc*. #

<u>Remarque</u> : Les rappels concernant les théorèmes de Fubini et de changement de variables seront faits ultérieurement.

Chapitre 4

Variables aléatoires, loi d'une variable aléatoire

L'une des difficultés principales de ce chapitre est due au changement de notation et de terminologie qu'il introduit.

4.1 Définitions et généralités

En comparant les définitions d'espace probabilisé donné au chap. I et d'espace mesuré au chap. II, on vérifie immédiatement qu'un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$ n'est autre qu'un espace mesuré dont la mesure $\mathbb{P}$ est une probabilité.

Définitions :

Soient $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé et $(E, \mathcal{E})$ un espace mesurable.

(a) Une application mesurable $X : (\Omega, \mathcal{A}) \to (E, \mathcal{E})$ est appelée une *variable aléatoire* (v.a.).

(b) On appelle *loi d'une v.a.* X la probabilité image de $\mathbb{P}$ par X, généralement notée $\mathbb{P}_X$ (ou parfois aussi μ_X).

• Si $E = \mathbb{R}$ on parle de v.a. réelle (v.a.r.), si $E = \mathbf{C}$ de v.a. complexe, si $E = \mathbb{R}^d$ on parle de vecteur aléatoire.

Les projections $\pi_i : \left(\mathbb{R}, \mathcal{B}\left(\mathbb{R}^d\right)\right) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ sont continues donc mesurables. Les $X_i = \pi_i \circ X$ sont donc des v.a.r. appelées *marginales* de $X = (X_1, ..., X_d)$.

Les lois $\mathbb{P}_{X_i}$, $1 \leq i \leq d$, des X_i sont les *lois marginales* de X. On peut évidemment remplacer $\mathbb{R}$ par $\mathbf{C}$ dans ces définitions. Il est à noter qu'en général, la connaissance des lois marginales $\mathbb{P}_{X_i}, 1 \leq i \leq d$, ne détermine pas la loi $\mathbb{P}_X$ du vecteur X.

Exemple : On considère les quatre points de $\mathbb{R}^2$ $a = (0,0)$, $b = (1,0)$, $c = (1,1)$ et $d = (0,1)$. Soit $\Omega = \{a, b, c, d\}$, $\mathcal{A} = \mathcal{P}(\Omega)$ et les deux probabilités $\mathbb{P} = \frac{1}{4}(\delta_a + \delta_b + \delta_c + \delta_d)$, $\mathbb{Q} = \frac{1}{6}(\delta_a + \delta_d) + \frac{1}{3}(\delta_b + \delta_c)$. On définit la v.a. X comme l'injection canonique de Ω dans $\mathbb{R}^2$ *i.e.* $X(a) = (0,0)$, etc. Il est évident que X n'a pas la même loi sous $\mathbb{P}$ et $\mathbb{Q}$ car $\mathbb{P}(X = (0,0)) = 1/4$ et $\mathbb{Q}(X = (0,0)) = 1/6$. Les marginales X_1 et X_2 sont à valeurs dans $\{0,1\}$. On constate que X_1 a la même loi sous $\mathbb{P}$ que sous $\mathbb{Q}$, puisque *e.g.*

$$\mathbb{P}(X_1 = 1) = \mathbb{P}(\{b, c\}) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2} \ \text{ et } \ \mathbb{P}(X_1 = 0) = \mathbb{P}(\{a, d\}) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2},$$

$$\mathbb{Q}(X_1 = 1) = \mathbb{Q}(\{b, c\}) = \frac{1}{3} + \frac{1}{6} = \frac{1}{2} \ \text{ et } \ \mathbb{P}(X_1 = 0) = \mathbb{P}(\{a, d\}) = \frac{1}{6} + \frac{1}{3} = \frac{1}{2}.$$

Le même phénomène s'observe sur X_2.

• Il résulte de la définition de la loi $\mathbb{P}_X$ d'une v.a. X que :

$$\forall B \in \mathcal{E}, \ \ \mathbb{P}_X(B) = \mathbb{P}\left(X^{-1}(B)\right) = \mathbb{P}(\{X \in B\})$$

et plus généralement

$$f \in \mathcal{L}^1_{\mathbb{K}}(E, \mathcal{E}, \ \mathbb{P}_X) \ \text{ si et seulement si } \ f \circ X \in \mathcal{L}^1_{\mathbb{K}}(\Omega, \mathcal{A}, \mathbb{P})$$

avec

$$(*) \quad \int_E f(x)\, \mathbb{P}_X(dx) = \int_\Omega f(X)\, d\mathbb{P} \ \text{(toujours vraie si } f \geq 0).$$

Définition :

Soit $Y : (\Omega, \mathcal{A}, \mathbb{P}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ une v.a. intégrable ou positive. On appelle espérance de Y, notée $\mathbb{E}(Y)$, la quantité $\mathbb{E}(Y) = \int Y d\mathbb{P}$.

Ainsi la relation $(*)$ ci-dessus s'écrit-elle $\mathbb{E}(f(X)) = \int f(x)\, \mathbb{P}_X(dx)$ (ou même $\mathbb{E}(f(X)) = \int f d\mathbb{P}_X$) dès qu'elle a un sens.

Remarque : Dans le chapitre I la définition de l'espérance d'une v.a. discrète était bien donnée par la relation $(*)$ puisque : $\mathbb{E}(f(X)) = \int_E f(e) d\pi(e) = \sum_{e \in E} f(e)\pi(e)$ où $\pi(e) = \mathbb{P}(X = e), e \in E$, est la loi de X.

• Si X est une *v.a. réelle*, on définit, lorsqu'ils existent,

$\mathbb{E}(X^n)$ le moment d'ordre $n \in \mathbb{N}^*$ (si $\mathbb{E}|X|^n < +\infty$),
$\mathbb{E}((X - \mathbb{E}(X))^n)$ le moment centré d'ordre $n \in \mathbb{N}^*$ (si $X \in \mathcal{L}^n \subset \mathcal{L}^1$),
$\mathbb{E}(|X|^\alpha)$ le moment absolu d'ordre $\alpha \in \mathbb{R}^*_+$ (existe toujours).

Lorsque $\mathbb{E}(X) = 0$, X est dite *centrée* . Si $n = 2$, le moment centré est appelé *la variance de* X : $Var(X) := \mathbb{E}\left((X - \mathbb{E}(X))^2\right)$. La variance de X existe dès que

$X \in \mathcal{L}^2(\mathbb{P})$ et elle est toujours *positive*. On définit enfin *l'écart-type* par $\sigma(X) := \sqrt{Var(X)}$.

On vérifie aussitôt que $Var(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$.

• Si X et Y sont deux v.a.r. de carré intégrable sur un même espace probabilisé, alors $(X - \mathbb{E}X)(Y - \mathbb{E}Y)$ est intégrable d'après l'inégalité de Cauchy-Schwarz et l'on définit la *covariance* entre X et Y par

$$cov(X,Y) := \mathbb{E}((X - \mathbb{E}X)(Y - \mathbb{E}Y)).$$

Il est immédiat par linéarité de l'espérance que $cov(X,Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$.

Toujours par linéarité de l'espérance, on vérifie les identités suivantes

$$\begin{array}{ll} \forall\lambda \in \mathbb{R}, & Var(\lambda X) = \lambda^2 Var(X) \text{ et } \sigma(\lambda X) = |\lambda|\,\sigma(X), \\ \forall\lambda,\mu \in \mathbb{R}, & Var(\lambda X + \mu Y) = \lambda^2 Var(X) + 2\lambda\mu\, cov(X,Y) + \mu^2 Var(Y). \end{array}$$

• On définit le *coefficient de corrélation* entre X et Y, v.a. de carré intégrables supposées non *p.s.* constantes de façon que $Var(X)$ et $Var(Y) \neq 0$, par

$$\rho(X,Y) := \frac{cov(X,Y)}{\sigma(X)\,\sigma(Y)}.$$

L'inégalité de Cauchy-Schwarz et son cas d'égalité entraînent que

$$\rho(X,Y) \in [-1,1] \text{ et } |\rho(X,Y)| = 1 \text{ ssi } \exists\lambda \in \mathbb{R}^*,\ \mu \in \mathbb{R} \text{ tq } X = \lambda Y + \mu\ \mathbb{P}\text{-}p.s.$$

• On définit également *la médiane* de X comme étant tout nombre λ vérifiant à la fois $\mathbb{P}(X \leq \lambda)$ et $\mathbb{P}(X \geq \lambda) \geq \frac{1}{2}$. Il existe toujours au moins une médiane et il peut en exister plusieurs (cf. exercices).

• Soit X un *vecteur* aléatoire à valeurs dans $\mathbb{R}^d$ (*i.e.* mesurable de $(\Omega,\mathcal{A})$ dans $\left(\mathbb{R}^d, \mathcal{B}\left(\mathbb{R}^d\right)\right)$. On dira que X est intégrable si et seulement si chacune de ses composantes X_i l'est. On vérifie aussitôt que X est intégrable si et seulement si $\|X\|$ l'est pour une norme quelconque sur $\mathbb{R}^d$ et l'on pose $\mathbb{E}(X) := \left|\begin{array}{c} \mathbb{E}(X_1) \\ \vdots \\ \mathbb{E}(X_d) \end{array}\right. \in \mathbb{R}^d$.

Si $\mathbb{E}\left(\|X\|^2\right) < +\infty$, chacune des $X_i \in \mathcal{L}^2(\mathbb{P})$ et l'on peut définir *la matrice de dispersion* de X par

$$D(X) := [cov(X_i, X_j)]_{1\leq i,j\leq d} = [\mathbb{E}((X_i - \mathbb{E}X_i)(X_j - \mathbb{E}X_j))]_{1\leq i,j\leq d}.$$

Pour tout $\lambda = \left|\begin{array}{c} \lambda_1 \\ \vdots \\ \lambda_d \end{array}\right. \in \mathbb{R}^d$, ${}^t\lambda D(X)\lambda = \sum_{i,j} \lambda_i\lambda_j cov(X_i,X_j) = Var\left(\sum_{i=1}^d \lambda_i X_i\right) \geq 0.$

La matrice $D(X)$ est donc toujours *symétrique positive.*

4.2 Lois de probabilités usuelles (à densité)

Dans la pratique, ce sont souvent les lois qui sont données, l'espace probabilisé $(\Omega,\mathcal{A},\mathbb{P})$ et la v.a. X étant construits "ensuite". Ainsi une probabilité μ étant donnée sur $\mathbb{R}$, on peut construire $\Omega = \mathbb{R}$, $\mathcal{A} = \mathcal{B}(\mathbb{R})$, $\mathbb{P} = \mu$ et poser $X = Id_{\mathbb{R}}$. X sera alors une v.a. de loi μ. Il est en effet souvent commode de raisonner sur des v.a..

Pour autant, on se gardera bien d'identifier en toutes circonstances une v.a. et sa loi : en effet on rencontre souvent des v.a. X et Y définies sur un même espace probabilisé $(\Omega,\mathcal{A},\mathbb{P})$ ayant même loi sans être égales *p.s.*. Ainsi, si $X \overset{\mathcal{L}}{\sim} \mathcal{B}(\{0,1\},1/2)$, il est immédiat que $Y := 1 - X \overset{\mathcal{L}}{\sim} \mathcal{B}(\{0,1\},1/2)$ alors que $\{X \neq Y\} = \Omega$!

Avant de passer aux exemples, un peu de terminologie. Si $(E,\mathcal{E},m)$ est un espace mesuré et $h \in \mathcal{L}^1_{\mathbb{R}^+}(m)$ avec $\int hdm = 1$, il est aisé de vérifier que $\mu(A) := \int_A hdm$ est elle-même une probabilité. On dit que μ est *m-absolument continue* (ou absolument continue par rapport à m) de *densité h*. Si $m = \lambda$ (mesure de Lebesgue) on parle de *loi absolument continue* et si m est la mesure de décompte sur E on parle de *loi discrète.* En général une loi de probabilité n'est ni absolument continue, ni discrète...

4.2.1 Loi uniforme $U([a,b])\,; a < b$:

C'est la loi de densité $\frac{1}{b-a}\mathbf{1}_{[a,b]}(x)$ (par rapport à la mesure de Lebesgue). La loi $U([0,1])$ joue un rôle essentiel en simulation puisqu'elle est le point de départ de toute la théorie (cf. chap.V) .

Si $X \overset{\mathcal{L}}{\sim} U([a,b])$, $\mathbb{E}(X) = \frac{1}{b-a}\int_a^b x dx = \frac{b+a}{2}$ et

$$Var(X) = \frac{1}{b-a}\int_a^b \left(x - \frac{b+a}{2}\right)^2 dx = \frac{(b-a)^2}{12}.$$

4.2.2 La loi gamma, $\gamma(\alpha,\beta)\,, \alpha,\beta > 0$:

On rappelle que la fonction gamma est définie sur $\mathbb{R}_+^*$ par :

$$\forall t > 0,\ \Gamma(t) = \int_0^{+\infty} e^{-x} x^{t-1} dx$$

et vérifie $\Gamma(t+1) = t\Gamma(t)\,, \Gamma(n) = (n-1)!$ si $n \in \mathbb{N}^*$.

$X \overset{\mathcal{L}}{\sim} \gamma(\alpha,\beta)$ si la loi de X a pour densité par rapport à la mesure de Lebesgue

$$f_{\alpha,\beta}(x) := \frac{1}{\beta^\alpha \Gamma(\alpha)} e^{-\frac{x}{\beta}} x^{\alpha-1}\mathbf{1}_{\mathbb{R}_+^*}(x)$$

C'est bien une densité de probabilité car $f_{\alpha,\beta} \geq 0$ et $\int_0^{+\infty} f_{\alpha,\beta}(x)\,dx = 1$. A partir de l'identité $\int_0^{+\infty} x^\rho f_{\alpha,\beta}(x)dx = \frac{\beta^\rho \Gamma(\alpha+\rho)}{\Gamma(\alpha)}$ pour tout $\rho > 0$, il vient immédiatement

$$\mathbb{E}(X) = \alpha\beta \text{ et } Var\,(X) = \alpha\beta^2.$$

Lorsque $\alpha = 1$ et $\beta = \frac{1}{\lambda}$, $\lambda > 0$, on parle de *loi exponentielle* $\mathcal{E}\,(\lambda)$.
Lorsque $\alpha = \frac{n}{2}$ et $\beta = 2$, on parle de *loi du chi*^{2} *à* n *degrés de liberté.*

4.2.3 Loi normale $\mathcal{N}\left(m, \sigma^2\right)$ sur $\mathbb{R}$

$X \overset{\mathcal{L}}{\sim} \mathcal{N}\,(m, \sigma^2)$ si la loi de X a pour densité

$$\forall x \in \mathbb{R}, \quad f\,(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-m)^2}{2\sigma^2}\right].$$

où $m \in \mathbb{R}$ et $\sigma > 0$.

Si $m = 0$ et $\sigma = 1$ on parle de *loi normale centrée réduite.* Dans ce cas on vérifie que f est une densité car

$$\int_{-\infty}^{+\infty} e^{-\frac{x^2}{2}} \frac{dx}{\sqrt{2\pi}} = \frac{2}{\sqrt{2\pi}} \int_0^{+\infty} e^{-\frac{x^2}{2}} dx = \sqrt{\frac{2}{\pi}} \int_0^{\infty} u^{-\frac{1}{2}} e^{-u} \frac{1}{\sqrt{2}} du = \frac{\Gamma\left(\frac{1}{2}\right)}{\sqrt{\pi}} = 1.$$

D'autre part

$$\mathbb{E}(X) = \int_{-\infty}^{+\infty} \underbrace{x e^{-\frac{x^2}{2}}}_{\in \mathcal{L}^1(\lambda)} \frac{dx}{\sqrt{2\pi}} = \frac{1}{\sqrt{2\pi}} [e^{-\frac{x^2}{2}}]_{-\infty}^{+\infty} = 0,$$

$$Var(X) = \mathbb{E}(X^2) = \int_{-\infty}^{\infty} x^2 e^{-\frac{x^2}{2}} \frac{dx}{\sqrt{2\pi}} = \frac{\Gamma\left(\frac{1}{2}\right)}{\sqrt{\pi}} = 1.$$

Dans le cas général on procède au changement de variable $x := \sigma u + m$ et l'on vérifie de la même façon que $\mathbb{E}(X) = m$ et $Var\,(X) = \mathbb{E}((X-m)^2) = \sigma^2$.

4.2.4 Loi normale centrée réduite sur $\mathbb{R}^d$, $d \geq 2$

X, vecteur aléatoire à valeurs dans $\mathbb{R}^d$, suit une $\mathcal{N}\,(0, I_d)$ si elle a pour densité par rapport à la mesure de Lebesgue sur $\mathbb{R}^d$ la fonction définie par :

$$\forall x \in \mathbb{R}^d, \quad f\,(x) := (2\pi)^{-\frac{d}{2}} e^{-\frac{\|x\|^2}{2}} \text{ où } \|x\|^2 = \sum_{i=1}^d x_i^2.$$

On vérifie par le théorème de Fubini (cf. chap.IV) que f est bien une densité et que $\mathbb{E}(X) = \overrightarrow{0}_{\mathbb{R}^d}$ et $D(X) = I_d$.
Cette loi sera généralisée lors de l'étude des vecteurs gaussiens.

Remarque : On prendra garde, lors de la description d'une loi à densité, de préciser les parties de $\mathbb{R}$ ou $\mathbb{R}^d$ sur lesquelles la densité est nulle notamment par un recours soigneux à des fonctions indicatrices.

4.3 Caractérisation d'une loi. Application au calcul de lois

Lorsque l'on doit déterminer la loi d'une v.a. X définie comme fonction d'autres v.a. : $X := \varphi\,(Y, Z, ...)$, on s'appuie sur la proposition suivante, application directe du théorème sur l'intégration des fonctions par rapport à la mesure image (cf. section 5.2 du chap. II) et des résultats de caractérisation de mesure (cf. section 5.1).

Proposition (caractérisation de la loi) :
Soit $X : (\Omega, \mathcal{A}, \mathbb{P}) \to (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ et μ une probabilité sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$.
Il y a équivalence entre les assertions suivantes :

(i) $\quad \mu = \mathbb{P}_X \quad \left[i.e.\ \forall B \in \mathbb{R}^d, \quad \mu\,(B) = \mathbb{P}\,(X \in B)\right]$,
(ii) $\quad \forall f \in \mathcal{C}_K(\mathbb{R}^d, \mathbb{R}), \quad \mathbb{E}\,(f\,(X)) = \int_{\mathbb{R}^d} f\,(x)\,\mu\,(dx)$,
(iii) $\quad \forall f : (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d)) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, borélienne bornée (ou positive) $\mathbb{E}(f(X)) = \int_{\mathbb{R}^d} f(x)\mu(dx)$.

Remarques : • Le résultat important est $(i) \iff (iii)$. C'est aussi le plus facile. La caractérisation (ii) est moins importante. La proposition s'étend à toute v.a. à valeurs dans un espace E métrique localement compact et σ-compact (*i.e.* réunion dénombrable de compacts).
• Si E est un espace métrique quelconque, l'assertion (ii) subsiste en remplaçant $\mathcal{C}_K\,(E, \mathbb{R})$ par $\mathcal{C}_b\,(E, \mathbb{R})$.
Démonstration :

$(i) \Rightarrow (iii)$: cf. section 5.2. du chap. II sur la mesure image.
$(iii) \Rightarrow (ii)$: évident.
$(ii) \Rightarrow (i)$: par définition de la loi de X, $\mathbb{E}(f(X)) = \int f(x)\mathbb{P}_X(dx)$ pour toute f borélienne bornée, donc $\forall f \in \mathcal{C}_K(\mathbb{R}^d, \mathbb{R}) \quad \int f(x)\mathbb{P}_X\,(dx) = \int f\,(x)\,\mu\,(dx)$.

D'où $\mathbb{P}_X = \mu$ d'après section 5.a. de II. #

Exemples d'application $(d = 1)$ **:**
(1) Soit $X \overset{\mathcal{L}}{\sim} \mathcal{N}(0,1)\,, \sigma > 0$ et $Y := \sigma X + m$. Quelle est la loi de Y ?
Soit f continue à support compact. $\mathbb{E}(f(Y)) = \mathbb{E}(f(\sigma X + m)) = \int f(\sigma x + m)\mathbb{P}_X(dx)$
d'où $\mathbb{E}(f(Y)) = \int_{-\infty}^{+\infty} f(\sigma x + m) e^{-\frac{x^2}{2}} \frac{dx}{\sqrt{2\pi}}$. On pose $x := \frac{y-m}{\sigma} = \varphi(y)$. Il vient
$\mathbb{E}(f(Y)) = \int_{-\infty}^{+\infty} f(y) \exp\left[-\frac{(y-m)}{2\sigma^2}\right] \frac{dy}{\sigma\sqrt{2\pi}}$ donc la loi de Y est donnée par

$$\mathbb{P}_Y(dy) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(y-m)^2}{2\sigma^2}\right] dy$$

i.e. $Y \overset{\mathcal{L}}{\sim} \mathcal{N}(m, \sigma^2)$.

(2) Plus généralement si X a une loi de densité g sur $\mathbb{R}$ et $Y := aX + b, a \neq 0$, on montre de même que

$$\mathbb{P}_Y(dy) = h(y)dy \text{ avec } h(y) := \frac{1}{|a|} g\left(\frac{y-b}{a}\right).$$

(3) Soit X une v.a.r. de densité g. Déterminer la loi de $Y := X^2$.
Soit f continue à support compact (ou simplement borélienne bornée) sur $\mathbb{R}$.

$$\mathbb{E}(f(Y)) = \mathbb{E}(f(X^2)) = \int_{-\infty}^{+\infty} f(x^2) \underbrace{g(x)dx}_{\mathbb{P}_X(dx)} = \int_{-\infty}^{0} f(x^2)g(x)dx + \int_{0}^{+\infty} f(x^2)g(x)dx.$$

On procède par changement de variable dans chacune des intégrales :

$$\int_0^{+\infty} f(x^2)g(x)dx = \int_0^{+\infty} f(y)g(\sqrt{y})\frac{dy}{2\sqrt{y}} \text{ et } \int_{-\infty}^{0} f(x^2)g(x)dx = \int_0^{+\infty} f(y)g(-\sqrt{y})\frac{dy}{2\sqrt{y}}.$$

D'où $\mathbb{E}(f(Y)) = \int f(y)\left(\frac{g(\sqrt{y}) + g(-\sqrt{y})}{2\sqrt{y}}\right) \mathbf{1}_{\mathbb{R}_+^*}(y)\,dy$ *i.e.*

$$\mathbb{P}_Y(dy) = \frac{g(\sqrt{y}) + g(-\sqrt{y})}{2\sqrt{y}} \mathbf{1}_{\mathbb{R}_+^*}(y)\,dy.$$

Ainsi, lorsque g est paire, Y a pour densité $h(y) = \frac{g(\sqrt{y})}{\sqrt{y}} \mathbf{1}_{\mathbb{R}_+^*}(y)$. Par exemple, si $X \overset{\mathcal{L}}{\sim} \mathcal{N}(0,1)\,,\ X^2$ a pour densité $y^{-\frac{1}{2}} e^{-y/2} \frac{1}{\sqrt{2\pi}} \mathbf{1}_{\mathbb{R}_+^*}(y)$ *i.e.*

$$X \overset{\mathcal{L}}{\sim} \mathcal{N}(0,1) \Rightarrow X^2 \overset{\mathcal{L}}{\sim} \gamma\left(\frac{1}{2}, 2\right) = \chi^2(1).$$

Dans les exemples précédents on s'est cantonné à des v.a. réelles. Lorsque X est un vecteur aléatoire, l'outil de base est le théorème de changement de variables pour les intégrales multiples que nous allons rappeler ici.

Théorème de changement de variables :
Soient U, V deux ouverts de $\mathbb{R}^d$ et $\varphi : U \to V$ un difféomorphisme (φ bijective, C^1, de réciproque φ^{-1} également C^1). Alors

$$\forall f \in \mathcal{L}^1(V, \mathcal{B}(V), dv),\ \int_V f(v)\,dv = \int_U f(\varphi(u))\,|J_\varphi(u)|\,du \quad (**)$$

où $J_\varphi(u) = \det\left(\left[\frac{\partial \varphi_i}{\partial u_j}(u)\right]_{1 \le i,j \le d}\right)$. L'égalité $(**)$ s'étend aux fonctions $f \geq 0$.

Rappel (théorème d'inversion locale) : Pour que $\varphi : U \longrightarrow V$ soit un C^1-difféomorphisme, il suffit que φ soit bijective de U dans V, C^1 sur U et que l'application linéaire $\varphi'(u)$ soit inversible en tout point $u \in U$ (*i.e.* $J_\varphi(u) \neq 0$).

Application :
Soit $X : (\Omega, \mathcal{A}, \mathbb{P}) \to (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ tel que $\mathbb{P}(X \in D) = 1$ où D est un ouvert de $\mathbb{R}^d$ (on dit que $X \in D$ $\mathbb{P}$-*p.s.*). On définit $Y := \Psi(X)$ où $\Psi : D \to \Delta \subset \mathbb{R}^d$ est un *difféomorphisme*. Alors, si la loi de X a pour densité g,

(*i*) $g = 0$ dx-*p.s.* sur ${}^cD$,

(*ii*) Y a une densité h sur $\mathbb{R}^d$ donnée par :

$$h(y) := \begin{cases} g(\Psi^{-1}(y))\,|J_{\Psi^{-1}}(y)| & \text{si } y \in \Delta \\ 0 & \text{si } y \notin \Delta \end{cases}.$$

Démonstration :
(*i*) $0 = \mathbb{P}(X \in {}^cD) = \int \mathbf{1}_{{}^cD}(X)d\mathbb{P} = \mathbb{P}_X({}^cD) = \int_{{}^cD} g(x)dx$ donc $g = 0$ *p.s.* sur ${}^cD$.

(*ii*) Soit $f : \mathbb{R}^d \to \mathbb{R}$ borélienne positive. Appliquant la proposition précédente au difféomorphisme $\Psi^{-1} : \Delta \to D$, on obtient :

$$\begin{aligned} \mathbb{E}(f(Y)) &= \mathbb{E}(f \circ \Psi(X)) = \int_D f \circ \Psi(x)g(x)dx \\ &= \int_\Delta f(y)g(\Psi^{-1}(y))\,|J_{\Psi^{-1}}(y)|\,dy \\ \text{d'où} \quad h(y) &= g(\Psi^{-1}(y))\,|J_{\Psi^{-1}}(y)|\,\mathbf{1}_\Delta(y). \quad \# \end{aligned}$$

Remarques :

• Comme $(\Psi^{-1})'(y) = (\Psi')^{-1}(\Psi^{-1}(y))$, il vient $J_{\Psi^{-1}}(y) = (J_\Psi)^{-1}(\Psi^{-1}(y))$. On en déduit aussitôt que $h = \left(\frac{g}{|J_\Psi|}\mathbf{1}_D\right) \circ \Psi^{-1}$ puisque $\mathbf{1}_D \circ \Psi^{-1} = \mathbf{1}_\Delta$.

• Généralement, dans les applications, Ψ n'est pas un difféomorphisme. On découpe alors le domaine D de façon à se ramener à la situation décrite ci-dessus. Dans tous les cas, il est préférable de refaire le calcul plutôt que d'utiliser les formules synthétiques ci-dessus.

Exemples $(d \geq 2)$ **:**

(1) Loi d'une marginale (n'utilise pas le changement de variables)
Soit $X = (X_1, X_2) : (\Omega, \mathcal{A}, \mathbb{P}) \to \mathbb{R}^2$ de densité g. Quelles sont les lois des marginales ?

$$\mathbb{E}\left(f\left(X_1\right)\right) = \int\!\!\int f(x_1)g(x_1,x_2)dx_1dx_2 = \int f(x_1)\left(\int g(x_1,x_2)dx_2\right)dx_1,$$

d'après le théorème de Fubini (cf. chap.IV ci-après) donc X_1 a pour densité

$$g_1\left(x_1\right) = \int g\left(x_1, x_2\right)dx_2.$$

La généralisation à des vecteurs aléatoires généraux est évidente.

(2) Soit $X = (X_1, X_2) \stackrel{\mathcal{L}}{\sim} \mathcal{N}(0, I_2)$. Quelle est la loi de la v.a.r. définie par $Y := \dfrac{X_1}{X_2}$ sur $\{X_2 \neq 0\}$ et $Y := 0$ sur $\{X_2 = 0\}$?
Soit f borélienne positive.

$$\mathbb{E}\left(f\left(Y\right)\right) = \int\!\!\int\left(f\left(\frac{x_1}{x_2}\right)\mathbf{1}_{\{x_2 \neq 0\}} + f(0)\mathbf{1}_{\{x_2=0\}}\right)e^{-\frac{x_1^2+x_2^2}{2}}\frac{dx_1dx_2}{2\pi}$$

Or

$$\int\!\!\int \mathbf{1}_{\{x_2=0\}}e^{-\frac{x_1^2+x_2^2}{2}}\frac{dx_1dx_2}{2\pi} = \int \mathbf{1}_{\{x_2=0\}} \times 1 \times e^{-\frac{x_2^2}{2}}\frac{dx_2}{\sqrt{2\pi}} = 0.$$

Donc

$$\mathbb{E}\left(f\left(Y\right)\right) = \int\!\!\int f\left(\frac{x_1}{x_2}\right)\mathbf{1}_{\{x_2\neq 0\}}e^{-\frac{x_1^2+x_2^2}{2}}\frac{dx_1dx_2}{2\pi}.$$

On pose $(x_1, x_2) = \varphi(u,v)$ avec $\varphi(u,v) := (uv, v)$. φ est une bijection C^1 de $\mathbb{R}\times\mathbb{R}^*$ sur $\mathbb{R}\times\mathbb{R}^*$ (de réciproque $\varphi^{-1}(x_1,x_2) = (x_1/x_2, x_2)$) et $J_\varphi(u,v) = \begin{vmatrix} v & u \\ 0 & 1 \end{vmatrix} = v \neq 0$.
On peut donc écrire :

$$\begin{aligned}\mathbb{E}\left(f\left(Y\right)\right) &= \int\!\!\int f\left(u\right)\mathbf{1}_{\{v\neq 0\}}e^{-\frac{v^2}{2}(u^2+1)}\left|v\right|\frac{dudv}{2\pi} \\ &= \int f(u)\left(\int_{\{v\neq 0\}}e^{-\frac{v^2}{2}(u^2+1)}\left|v\right|dv\right)\frac{du}{2\pi}.\end{aligned}$$

Or

$$\begin{aligned}\int_{\{v\neq 0\}}\exp\left(-\frac{v^2}{2}\left(u^2+1\right)\right)\left|v\right|dv &= 2\int_0^{+\infty}e^{-\frac{v^2}{2}(u^2+1)}vdv = 2\int_0^{+\infty}e^{-w(u^2+1)}dw \\ &= \frac{2}{u^2+1}.\end{aligned}$$

Donc

$$\mathbb{E}\left(f\left(Y\right)\right) = \int f\left(u\right)\frac{du}{\pi\left(u^2+1\right)} \quad i.e.\ \mathbb{P}_Y\left(dy\right) = \frac{dy}{\pi\left(y^2+1\right)}.$$

La loi de Y est appelée *distribution de Cauchy de paramètre 1.*
On notera au passage que

$$\mathbb{E}\left(|Y|\right) = 2\int_0^{+\infty}y\frac{dy}{\pi\left(y^2+1\right)} = +\infty \quad i.e. \quad Y \notin \mathcal{L}^1(\Omega, \mathcal{A}, \mathbb{P}).$$

(3) Soit X une v.a.r. de loi μ. Quelle est la loi de $Y := X^+$?
Soit f borélienne bornée.

$$\begin{aligned}\mathbb{E}(f(Y)) &= \mathbb{E}(f(0)\mathbf{1}_{\{X\leq 0\}}) + \mathbb{E}(f(X)\mathbf{1}_{\{X>0\}}) \\ &= f(0)\mathbb{P}(\{X \leq 0\}) + \int f(x)\mathbf{1}_{\{x>0\}}dx \\ &= f(0)\mu(\mathbb{R}_-) + \int f(x)\mathbf{1}_{\{x>0\}}dx\end{aligned}$$

d'où finalement [1]

$$\mathbb{P}_Y(dy) = \mu(\mathbb{R}_-)\delta_0(dy) + \mathbf{1}_{\mathbb{R}_+^*}\mu(dy).$$

4.4 Fonction de répartition d'une v.a.r.

Dans le cadre des v.a.réelles, il est parfois commode d'introduire la notion de fonction de répartition :

Définition :
Soit $X : (\Omega, \mathcal{A}, \mathbb{P}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$. La *fonction de répartition* de X est définie par :

$$\forall t \in \mathbb{R}, \quad F_X\left(t\right) := \mathbb{P}\left(X \leq t\right) = \mathbb{P}_X\left(]-\infty, t]\right).$$

Propriétés :
P1. F_X est croissante, continue à droite, limitée à gauche $(F_X(t-) = \mathbb{P}(X < t))$.
P2. $\lim_{+\infty} F_X = 1$ et $\lim_{-\infty} F_X = 0$.

[1] Rappelons que si $\nu := \delta_a$, alors $\int f d\nu = f(a)$ dès que $f(a)$ est fini.

P3. Si $F_X = F_Y$ alors X et Y ont même loi *i.e.* $\mathbb{P}_X = \mathbb{P}_Y$.
P4. Soit $F : \mathbb{R} \to [0,1]$ croissante, continue à droite, limitée à gauche et vérifiant $\lim\limits_{-\infty} F = 0, \lim\limits_{+\infty} F = 1$. Il existe alors une unique mesure de probabilité μ sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ vérifiant :

$$\forall a, b \in \mathbb{R}, \quad \mu(]a,b]) = F(b) - F(a).$$

Remarque : On peut affaiblir **P3** en supposant que F_X et F_Y coïncident *a priori* seulement sur une partie D dense dans $\mathbb{R}$.

Démonstration : P1. Si $t_n \downarrow t$, $\bigcap\limits_n^{\downarrow} \{X \le t_n\} = \{X \le t\}$ donc $F_X(t_n) \downarrow F_X(t)$. Si $t_n \uparrow t$, $t_n \neq t$, $\bigcup\limits_n^{\uparrow}]-\infty, t_n] =]-\infty, t[$ donc $F_X(t_n) \uparrow \mathbb{P}(X < t)$.
P2. Si $t_n \downarrow -\infty$, $\bigcap\limits_n^{\downarrow} \{X \le t_n\} = \phi$ et si $t_n \uparrow +\infty$, $\bigcup\limits_n^{\uparrow} \{X \le t_n\} = \Omega$.
P3. Il suffit de remarquer que $\mathcal{C} := \{]-\infty, t], t \in \mathbb{R}\}$ est stable par intersection finie et que $\sigma(\mathcal{C}) = \mathcal{B}(\mathbb{R})$ puis d'appliquer le théorème de caractérisation (cf. Application du II.3.1) aux lois $\mathbb{P}_X$ et $\mathbb{P}_Y$.
P4. L'unicité découle de **P3**. L'existence de μ, que nous admettrons, s'établit comme celle de la mesure de Lebesgue sur $[0,1]$ – qui correspond d'ailleurs au cas $F(x) = x$ – à l'aide du théorème de prolongement d'une mesure finie définie sur une algèbre de Boole (cf. II.3.2. théorème et exemple). #

Notation et terminologie : L'intégrale $\int f d\mu$ est parfois notée $\int f dF$ par référence à *l'intégrale de Stieltjes* dont la construction ci-dessus n'est qu'un cas particulier : l'intégrale de Stièljès peut être définie par rapport à une fonction F à variation finie quelconque.

Le recours aux fonctions de répartition est notamment intéressant pour l'étude des lois de minimum et de maximum de v.a. indépendantes : $\min\limits_{1 \le i \le n} X_i, \max\limits_{1 \le i \le n} X_i$ (cf. chap.IV.4.2.).

Applications : • Si une v.a.r. X a pour densité f, alors $F_X(t) = \int_{-\infty}^{t} f(u) du$.
• Réciproquement dès que F_X admet une telle représentation intégrale, X a pour densité f (par rapport à la mesure de Lebesgue). Ainsi, en pratique, si F_X est **continûment dérivable, sauf éventuellement en un nombre fini de points**, X a pour densité F'_X.

La fonction de répartition joue également un rôle important en simulation et en statistique.

Exemple : Soit $X \overset{\mathcal{L}}{\sim} \mathcal{E}(\lambda)$ *i.e.* de densité $\lambda e^{-\lambda x} \mathbf{1}_{\mathbb{R}_+}(x)$, $\lambda > 0$. Alors

$$\forall t \in \mathbb{R}_- \quad F_X(t) = \int_{-\infty}^{t} 0 \, du = 0$$
$$\forall t \in \mathbb{R}_+ \quad F_X(t) = \int_0^t \lambda e^{-\lambda u} du = 1 - e^{-\lambda t}.$$

4.5 Inégalités de Bienaymé-Tchebycheff

Ces célèbres inégalités sont essentiellement triviales dans le cadre de la théorie de la mesure. Leur intérêt est d'ordre théorique.

Proposition :
Soit $X : (\Omega, \mathcal{A}, \mathbb{P}) \to \mathbb{R}$, $\alpha > 0$, $\lambda > 0$. Alors

$$\mathbb{P}(|X| \ge \lambda) \le \frac{\mathbb{E}(|X|^\alpha)}{\lambda^\alpha} \quad (\alpha = 1 \text{ Markov}, \alpha = 2 \text{ B.T.}).$$

En particulier, si $X \in L^2_{\mathbb{R}}(\mathbb{P})$,

$$\mathbb{P}(|X - \mathbb{E}(X)| \ge \lambda) \le \frac{Var(X)}{\lambda^2}$$

Démonstration :

$$\mathbb{E}(|X|^\alpha) = \int |X|^\alpha d\mathbb{P} \ge \int_{\{|X| \ge \lambda\}} |X|^\alpha d\,\mathbb{P} \ge \lambda^\alpha \int \mathbf{1}_{\{|X| \ge \lambda\}} d\mathbb{P} = \lambda^\alpha \, \mathbb{P}(|X| \ge \lambda). \ \#$$

Exemple :
Si $X \overset{\mathcal{L}}{\sim} \mathcal{E}(\mu)$, l'inégalité de B.T. s'écrit, pour tout $x > 0$

$$\mathbb{P}(X \ge x) = \int_x^{+\infty} \mu e^{-\mu u} du = e^{-\mu x} \le \frac{\mathbb{E}(X^2)}{x^2} = \frac{Var(X) + (\mathbb{E}X)^2}{x^2} = \frac{2}{\mu^2 x^2}$$

ce qui illustre le caractère généralement grossier de cette majoration de la "queue de distribution" d'une v.a..

Chapitre 5

Variables aléatoires indépendantes

La notion d'indépendance joue un rôle fondamental en Probabilités. A bien des égards, c'est elle qui marque la spécifité des probabilités par rapport à la théorie de la mesure. Comme précédemment, nous allons nous appuyer sur l'approche intuitive développée dans le **cadre discret** où nous avions vu que deux v.a. discrètes X et Y sont indépendantes si et seulement si

$$\forall x, y, \quad \mathbb{P}(X = x \text{ et } Y = y) = \mathbb{P}(X = x)\, \mathbb{P}(Y = y)$$

en encore, en termes de lois :

$$\forall x, y, \ \pi_{(X,Y)}(x, y) = \pi_X(x)\pi_Y(y) \ \ i.e. \ \ \pi_{(X,Y)} = \pi_X \otimes \pi_Y.$$

C'est la raison pour laquelle nous allons commencer par des rappels sur les tribus et les mesures produits.

5.1 Rappels sur les tribus et les mesures produits. Théorèmes de Fubini

5.1.1 Définition

Définition :
Soient $(E_i, \mathcal{E}_i)$, $i = 1, 2$, deux espaces mesurables. La tribu produit $\mathcal{E}_1 \otimes \mathcal{E}_2$ sur $E_1 \times E_2$ est définie par : $\mathcal{E}_1 \otimes \mathcal{E}_2 : = \sigma\left(A_1 \times A_2, A_i \in \mathcal{E}_i, \ i = 1, 2\right)$.

C'est donc la tribu engendrée par les rectangles de parties mesurables.

Si l'on note $\pi_i : (E, \mathcal{E}_1 \otimes \mathcal{E}_2) \to (E_i, \mathcal{E}_i)$ la projection canonique sur E_i, on montre aisément que π_1 et π_2 sont mesurables et que toute tribu $\mathcal{T}$ sur E rendant les π_i, $i = 1, 2$, mesurables contient $\mathcal{E}_1 \otimes \mathcal{E}_2$.

D'autre part $f = (f_1, f_2) : (F, \mathcal{F}) \to (E_1 \times E_2, \ \mathcal{E}_1 \otimes \mathcal{E}_2)$ est mesurable ssi f_1 et f_2 le sont.

Un cas particulier essentiel est celui où les E_i sont des espaces topologiques munis de leurs boréliens $\mathcal{B}(E_i)$. En effet deux tribus cohabitent naturellement sur $E = E_1 \times E_2$: la tribu $\mathcal{B}(E_1 \times E_2)$ des boréliens de la topologie produit sur E et la tribu produit $\mathcal{B}(E_1) \otimes \mathcal{B}(E_2)$.
Dans les cas qui nous intéressent : $E_i = \mathbb{R}$, $\mathbf{C}$, $\overline{\mathbb{R}}$, $\mathbb{R}_+$, $\overline{\mathbb{R}}_+$, $\mathbb{R}^d$, on a toujours égalité entre ces 2 tribus *i.e.*

$$\mathcal{B}(\mathbb{R}^2) = \mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R}), \ \ \mathcal{B}(\overline{\mathbb{R}} \times \mathbb{R}_+) = \mathcal{B}\left(\overline{\mathbb{R}}\right) \otimes \mathcal{B}(\mathbb{R}_+),$$

$$\mathcal{B}\left(\mathbb{R}^{p+q}\right) = \mathcal{B}(\mathbb{R}^p) \otimes \mathcal{B}(\mathbb{R}^q), \text{ etc.}$$

Remarques : • En fait on a toujours $\mathcal{B}(E_1) \otimes \mathcal{B}(E_2) \subset \mathcal{B}(E_1 \times E_2)$ car les projections π_i sont continues de $E_1 \times E_2 \longrightarrow E_i$ donc $(\mathcal{B}(E_1 \times E_2), \mathcal{B}(E_i))$-mesurables. L'égalité a lieu si les espaces E_1 et E_2 sont à base dénombrable d'ouverts ce qui est évidemment le cas de tous les espaces ci-dessus et plus généralement de tout espace métrique contenant une famille dénombrable dense (espace séparable).
• Les résultats ci-dessus s'étendent au produit d'un nombre fini d'espaces E_i, $i = 1, \ldots, n$, et même, en fait, à un produit quelconque $E = \prod_{i \in I} E_i$ si l'on définit :

$$\begin{aligned} \bigotimes_{i \in I} \mathcal{E}_i \ &:= \ \sigma\left\{\prod_{i \in I} A_i, A_i \in \mathcal{E}_i, A_i = E_i, \ \text{ sauf pour un nombre fini d'indices } i\right\} \\ &= \ \sigma\left(\pi_i, \ i \in I, \ \pi_i \text{ projection de } E \text{ sur } E_i\right). \end{aligned}$$

• Ce sont ces résultats qui permettent de montrer par exemple que la somme de deux fonctions mesurables réelles est mesurable. En effet si $f_i : (F, \mathcal{F}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, $i = 1, 2$ alors $f = (f_1, f_2) : (F, \mathcal{F}) \to (\mathbb{R}^2, \mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R}))$ est mesurable. D'autre part l'addition $\left\{\begin{array}{l} \mathbb{R}^2 \to \mathbb{R} \\ (x, y) \mapsto x + y \end{array}\right.$ est continue donc $(\mathcal{B}(\mathbb{R}^2), \mathcal{B}(\mathbb{R}))$-mesurable. Mais comme $\mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R}) = \mathcal{B}(\mathbb{R}^2)$, par composition, $f_1 + f_2$ l'est également. Idem pour le produit.

5.1.2 Mesure produit, théorèmes de Fubini

On rappelle qu'une mesure μ sur $(E, \mathcal{E})$ est σ-*finie* s'il existe $E_n \in \mathcal{E}, \ n \geq 1$, telle que : $E = \bigcup_n E_n$ et $\mu(E_n) < +\infty$.
Théorème :
Soient $(E_i, \mathcal{E}_i, \mu_i)$, $i = 1, 2$, deux espaces mesurables, μ_i, $i = 1, 2$, deux mesures σ-finies. Il existe une *unique* mesure (σ-finie) μ sur $(E_1 \times E_2, \mathcal{E}_1 \otimes \mathcal{E}_2)$ vérifiant :

$$\forall A_1 \in \mathcal{E}_1, \ \forall A_2 \in \mathcal{E}_2, \ \mu(A_1 \times A_2) = \mu_1(A_1)\, \mu_2(A_2).$$

Cette mesure, appelée mesure produit de μ_1 et μ_2, est notée $\mu_1 \otimes \mu_2$. Si les μ_i sont des probabilités, il en est de même de $\mu_1 \otimes \mu_2$.

Application :
L'application essentielle (en théorie de la mesure) est évidemment la construction de la mesure de Lebesgue λ_d sur $\mathbb{R}^d$ à partir de la mesure de Lebesgue λ sur $\mathbb{R}$ puisque $\lambda_d := \lambda \otimes \ldots \otimes \lambda$.

Théorème de Fubini-Tonelli :
On se place sous les hypothèses du théorème précédent. Soit $f : (E_1 \times E_2, \mathcal{E}_1 \otimes \mathcal{E}_2) \to \left(\overline{\mathbb{R}}_+, \mathcal{B}\left(\overline{\mathbb{R}}_+\right)\right)$, mesurable.

(a) $\forall x_1 \in E_1,\ x_2 \mapsto f(x_1, x_2)$ est $\mathcal{E}_2$-mesurable,
$\forall x_2 \in E_2,\ x_1 \mapsto f(x_1, x_2)$ est $\mathcal{E}_1$-mesurable.

(b) $x_1 \mapsto \int f(x_1, x_2)\mu_2(dx_2)$ est $\mathcal{E}_1$ - mesurable,
$x_2 \mapsto \int f(x_1, x_2)\mu_1(dx_1)$ est $\mathcal{E}_2$ - mesurable.

(c)
$$\begin{aligned}\int f(x_1, x_2)\, \mu_1 \otimes \mu_2 (dx_1, dx_2) &= \int \left[\int f(x_1, x_2)\, \mu_1(dx_1)\right] \mu_2(dx_2) \leq +\infty \\ &= \int \left[\int f(x_1, x_2)\, \mu_2(dx_2)\right] \mu_1(dx_1) \leq +\infty.\end{aligned}$$

Théorème de Fubini-Lebesgue : Si $f \in \mathcal{L}^1_{\mathbb{K}}(E_1 \times E_2, \mathcal{E}_1 \otimes \mathcal{E}_2, \mu_1 \otimes \mu_2)$, $\mathbb{K} = \mathbb{R}$ ou $\mathbf{C}$. Alors

$$\mu_1(dx_1)\text{-p.p.},\ x_2 \mapsto f(x_1, x_2) \in \mathcal{L}^1_{\mathbb{K}}(\mu_2) \text{ et } \mu_2(dx_2)\text{-p.p.},\ x_1 \mapsto f(x_1, x_2) \in \mathcal{L}^1_{\mathbb{K}}(\mu_1),$$

$$x_1 \mapsto \int f(x_1, x_2)\, \mu_2(dx_2) \in \mathcal{L}^1_{\mathbb{K}}(\mu_1) \text{ et } x_2 \mapsto \int f(x_1, x_2)\, \mu_1(dx_1) \in \mathcal{L}^1_{\mathbb{K}}(\mu_2)$$

et, enfin, la relation (c) du théorème de Fubini-Tonelli est vérifiée.

Application au calcul d'espérance : Pour toute v.a.r. *positive* X sur $(\Omega, \mathcal{A}, \mathbb{P})$,

$$\mathbb{E}(X) = \int_0^{+\infty} \mathbb{P}(X \geq x)dx.$$

On se place sur $(\Omega \times \mathbb{R}_+, \mathcal{A} \otimes \mathcal{B}(\mathbb{R}_+), \mathbb{P} \otimes \lambda_{\mathbb{R}_+})$. On considère l'ensemble $A := \{(\omega, u) \,/\, X(\omega) \geq u\}$. $A \in \mathcal{A} \otimes \mathcal{B}(\mathbb{R}_+$ car son complémentaire se décompose en

$${}^cA = \bigcup_{r \in \mathbf{Q}_+^*} \{X < r\} \times [r, +\infty[.$$

Le théorème de Fubini-Tonelli (c) appliqué à $f := \mathbf{1}_A$ conduit à l'identité car

$$\mathbb{E}(X) = \int_\Omega \int_0^X (\omega)\, du d\mathbb{P} = \int \mathbf{1}_A(\omega, u)\, d\mathbb{P} \otimes \lambda_{\mathbb{R}_+} = \int_0^{+\infty} \mathbb{P}(X \geq x)dx.$$

5.2 Variables aléatoires indépendantes

Si l'on considère n v.a. $X_i : (\Omega, \mathcal{A}, \mathbb{P}) \longrightarrow (E_i, \mathcal{E}_i)$, $1 \leq i \leq n$, le vecteur $X := (X_1, \cdots, X_n) : \Omega \longrightarrow \prod_{i=1}^n E_i$ est donc (cf. section 1.1) $(\mathcal{A}, \otimes_{i=1}^n \mathcal{E}_i)$-mesurable. Par suite, la loi $\mathbb{P}_{(X_1, \cdots, X_n)}$ est une probabilité sur $\left(\prod_{i=1}^n E_i, \otimes_{i=1}^n \mathcal{E}_i\right)$, tout comme $\mathbb{P}_{X_1} \otimes \ldots \otimes \mathbb{P}_{X_n}$.

Définitions :
(a) Soient $X_i : (\Omega, \mathcal{A}, \mathbb{P}) \to (E_i, \mathcal{E}_i)$, $i = 1, \ldots, n$. Les v.a. X_i sont indépendantes si $\mathbb{P}_{(X_1, \ldots, X_n)} = \mathbb{P}_{X_1} \otimes \ldots \otimes \mathbb{P}_{X_n}$ *i.e.* la loi du n-uplet $(X_1, \ldots, X_n)$ est le produit des lois marginales des X_i.

(b) Une famille quelconque $(X_i)_{i \in I}$ de v.a. est composée de v.a. indépendantes si, pour toute partie $J \subset I$, J finie, les X_j, $j \in J$, sont indépendantes.

Il est à noter que, dans ce cas, la loi du n-uplet $(X_1, ..., X_n)$ est entièrement déterminée par les lois marginales $\mathbb{P}_{X_i}$. D'autre part, la caractérisation de la me- sure produit sur les rectangles (cf. théorème du 1.2), implique que les $X_1, ..., X_n$ sont indépendantes si et seulement si :

$$\forall A_1 \in \mathcal{E}_1, ..., \forall A_n \in \mathcal{E}_n,\ \mathbb{P}_{(X_1, ..., X_n)}(A_1 \times \ldots \times A_n) = \mathbb{P}_{X_1}(A_1) \ldots \mathbb{P}_{X_n}(A_n).$$

Soit encore : $X_1, ..., X_n$ sont indépendantes si et seulement si :

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

Remarques : • Il est clair au vu de $(*)$ que toute sous-famille d'une famille de v.a. indépendantes est constituée de v.a. indépendantes.

• Soit $X := (X_1, \cdots, X_n)$ un vecteur aléatoire à valeurs dans $\mathbb{R}^n$ ayant pour densité $g_{(X_1, \cdots, X_n)}(x_1, \cdots, x_n)$ et dont les lois marginales $\mathbb{P}_{X_i}(dx_i)$ ont des densités g_i par rapport à la mesure de Lebesgue (cf. chap.III section 3). Alors il découle instantanément de la définition que :

$$X_1, .., X_n \text{ sont indépendantes ssi } g_{(X_1, ..., X_n)}(x_1, ..., x_n) = g_1(x_1) \ldots g_n(x_n).$$

sauf éventuellement sur un ensemble Lebesgue *négligeable* de $\mathbb{R}^n$.

• Plus généralement, si la loi de $(X_1, \cdots, X_n)$ a une densité de la forme $g_{(X_1, ..., X_n)}(x_1, ..., x_n) := \gamma_1(x_1) \ldots \gamma_n(x_n)$, alors les v.a. X_i sont indépendantes et il existe des constantes $c_1, \cdots, c_n$ telles que $\mathbb{P}_{X_i}(dx_i) = c_i \gamma_i(x_i) dx_i$.

Exemple : Si $X = (X_1, ..., X_n) \overset{\mathcal{L}}{\sim} \mathcal{N}(0; I_n)$ *i.e.* X a pour densité $f(x_1, ..., x_n) = (2\pi)^{-\frac{n}{2}} \exp\left[-\frac{1}{2}(x_1^2 + ... + x_n^2)\right] = \prod_{i=1}^n \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{x_i^2}{2}}\right)$, les X_i sont indépendantes de même loi $\mathcal{N}(0; 1)$.

On peut légèrement affiner la caractérisation $(*)$ comme le montrent les caractérisations ci-après.

Proposition :

Il y a équivalence entre

(i) $X_1, ..., X_n$ sont indépendantes,

(ii) Si $\mathcal{E}_i = \sigma(\mathcal{C}_i)$, $\mathcal{C}_i$ stable par intersection finie, $i = 1, ..., n$

$$\forall A_1 \in \mathcal{C}_1, ..., \forall A_n \in \mathcal{C}_n, \quad \mathbb{P}\left(\bigcap_{i=1}^n \{X_i \in A_i\}\right) = \prod_{i=1}^n \mathbb{P}(X_i \in A_i).$$

(iii) $\forall f_i : (E_i, \mathcal{E}_i) \to \mathbb{R}$, borélienne bornée ou positive, $i = 1, ..., n$,

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

Démonstration : $(i) \Rightarrow (iii)$: d'après la proposition de caractérisation III, section 3,

$$\begin{aligned}
\mathbb{E}\left(\prod_{i=1}^n f_i(X_i)\right) &= \int \prod_{i=1}^n f_i(x_i)\ \mathbb{P}_{(X_1,...,X_n)}(dx_1, ..., dx_n) \\
&= \int \prod_{i=1}^n f_i(x_i) \mathbb{P}_{X_1}(dx_1) ... \mathbb{P}_{X_n}(dx_n) \text{ par hypothèse,} \\
&= \prod_{i=1}^n \int f_i(x_i)\, \mathbb{P}_{X_i}(dx_i) \text{ d'après le théorème de Fubini.} \\
&= \prod_{i=1}^n \mathbb{E}(f_i(X_i)).
\end{aligned}$$

$(iii) \Rightarrow (ii)$: prendre $f_i = \mathbf{1}_{A_i}$, $A_i \in \mathcal{C}_i$, $i = 1, ..., n$.

$(ii) \Rightarrow (i)$: On fixe $C_i \in \mathcal{C}_i$, $i = 2, .., n$ et l'on considère les mesures finies sur E_1 définies par $\mu_1(A_1) = \mathbb{P}(X_1 \in A_1,\ X_i \in C_i, ; 2 \leq i \leq n)$ et $\mu_1'(A_1) = \mathbb{P}(X_1 \in A_1) \prod_{i=2}^n \mathbb{P}(X_i \in C_i)$. μ_1 et μ_1' coïncident sur $\mathcal{C}_1$ donc sur $\sigma(\mathcal{C}_1) = \mathcal{E}_1$. Puis on fixe $A_1 \in \mathcal{E}_1$ et $C_i \in \mathcal{C}_i$, $3 \leq i \leq n$ et on recommence la procédure précédente. De proche en proche, on finit par établir la relation $(*)$ ci-avant. #

Application ($n = 2$ pour simplifier) : Deux v.a.r. X, Y sur $(\Omega, \mathcal{A}, \mathbb{P})$ sont indépendantes

ssi $\forall u, v \in \mathbb{R}, \mathbb{P}(X \leq u, Y \leq v) = \mathbb{P}(X \leq u)\mathbb{P}(Y \leq v)$

ssi $\forall u, v \in \mathbb{R}, \mathbb{P}_{(X,Y)}(]-\infty, u] \times]-\infty, v]) = F_X(u) F_Y(v)$.

Corollaires :

(a) Si les $X_i : (\Omega, \mathcal{A}) \to (E_i, \mathcal{E}_i)$, $i = 1, \ldots, n$, sont indépendantes et les $h_i : (E_i, \mathcal{E}_i) \to (F_i, \mathcal{F}_i)$, $i = 1, \ldots, n$, sont mesurables, alors les $h_i(X_i)$, $i = 1, \ldots, n$, sont indépendantes

(b) Si les $X_1, ..., X_n$ sont indépendantes et intégrables alors

$$\prod_{i=1}^n X_i \in \mathcal{L}^1(\mathbb{P}) \text{ et } \quad \mathbb{E}\left(\prod_{i=1}^n X_i\right) = \prod_{i=1}^n \mathbb{E}(X_i).$$

(c) Si U et V sont indépendantes (de carré intégrable), alors $\mathrm{Cov}(U, V) = 0$ et $\rho(U, V) = 0$ (la réciproque est fausse, cf. contre-exemple VIII.2).

Démonstration : (a) On applique la caractérisation (iii) à $f_i \circ h_i$, $1 \leq i \leq n$.

(b) D'après la caractérisation (iii) ci-avant donc $\mathbb{E}|X_1 \cdots X_n| = \mathbb{E}|X_1| \cdots \mathbb{E}|X_n| < +\infty$. On conclut *via* le théorème de Fubini-Lebesgue en reprenant la preuve de $(i) \Rightarrow (iii)$.

(c) $\mathrm{Cov}(U, V) = \mathbb{E}((U - \mathbb{E}U)(V - \mathbb{E}V)) = \mathbb{E}(UV) - \mathbb{E}U\ \mathbb{E}V = 0$. #

Remarque : Lorsque n v.a.r. $X_1, ..., X_n$, vérifient que $\mathrm{Cov}(X_i, X_j) = 0$ pour tout $i \neq j$, on dit que les X_i, $i = 1, .., n$ sont *non corrélées.* Il est clair qu'un vecteur $X = (X_1, ..., X_n)$ est composé de v.a. non corrélées ssi sa matrice de dispersion $D(X) = [\mathrm{Cov}(X_i, X_j)]_{1 \leq i,j \leq n}$ est diagonale.

On a par ailleurs les implications suivantes dont *toutes* les réciproques sont fausses

$$\begin{aligned}
(X_1, \cdots, X_n \text{ sont indépendantes}) &\Longrightarrow (X_1, \cdots, X_n \text{ sont 2 à 2 indépendantes}) \\
(X_1, \cdots, X_n \text{ sont 2 à 2 indépendantes}) &\Longrightarrow (X_1, \cdots, X_n \text{ sont 2 à 2 non corrélées})
\end{aligned}$$

Proposition :

Soient $X_1, ..., X_n$, n v.a.r. sur $(\Omega, \mathcal{A}, \mathbb{P})$ 2 à 2 non corrélées ; alors

$$\mathrm{Var}(X_1 + ... + X_n) = \mathrm{Var}(X_1) + ... + \mathrm{Var}(X_n).$$

Démonstration : Cela découle de la formule plus générale

$$\mathrm{Var}(X_1 + ... + X_n) = \sum_{i=1}^n \mathrm{Var}(X_i) + 2 \sum_{i<j} \mathrm{Cov}(X_i, X_j). \ \#$$

Construction de v.a. indépendantes : Soient $\mu_1, \ldots, \mu_n$, n probabilités définies sur des espaces $(E_1, \mathcal{E}_1), ..., (E_n, \mathcal{E}_n)$. La question qui se pose est : comment construire n v.a. indépendantes $X_1, ..., X_n$, de façon que X_i ait pour loi μ_i? La réponse est en fait presque immédiate : on pose $\Omega := E_1 \times \ldots \times E_n$, $\mathcal{A} := \mathcal{E}_1 \otimes ... \otimes \mathcal{E}_n$, $\mathbb{P} := \mu_1 \otimes ... \otimes \mu_n$ et $X_i := \pi_i$, $i^{\text{è}}$ projection canonique de Ω sur E_i.

5.3 Evénements et tribus indépendants

Définition :
Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé et $A_1, ..., A_n \in \mathcal{A}$. Les événements A_i, $i = 1, ..., n$, sont indépendants si les $\mathbf{1}_{A_i}$ sont des v.a. indépendantes.

Dans la pratique, comme les $\mathbf{1}_{A_i} \in \{0, 1\}$, cela revient à vérifier les relations :

$$\mathbb{P}(\cap_{i=1}^n A_i') = \prod_{i=1}^n \mathbb{P}(A_i') \text{ où } A_i' = A_i,\ {}^cA_i \text{ ou } \Omega.$$

On montre par récurrence sur n qu'il suffit de vérifier que :

$$\forall I \subset \{1, ..., n\},\ \mathbb{P}(\bigcap_{i\in I} A_i) = \prod_{i\in I} \mathbb{P}(A_i).$$

On retrouve donc là la définition du chapitre I.5.

Définition :
n sous-tribus $\mathcal{A}_1, ..., \mathcal{A}_n$ sur $(\Omega, \mathcal{A}, \mathbb{P})$ sont indépendantes si :

$$\forall A_1 \in \mathcal{A}_1, ..., \forall A_n \in \mathcal{A}_n, \quad A_1, ..., A_n \text{ sont indépendants.}$$

Remarque : On vérifie sans difficulté, comme dans la caractérisation de l'indépendance des v.a., que si $\mathcal{A}_i = \sigma(\mathcal{C}_i)$, $i = 1, ..., n$, $\mathcal{C}_i$ stable par intersection finie, les $\mathcal{A}_i$ sont indépendantes ssi $\forall A_1 \in \mathcal{C}_1, ..., \forall A_n \in \mathcal{C}_n,\ A_1, ..., A_n$ sont indépendants.

Exemple : Si les v.a. $X_1, ..., X_n$ sont indépendantes $(\Omega, \mathcal{A}, \mathbb{P})$, les sous-tribus $\sigma(X_1), ..., \sigma(X_n)$ sont indépendantes.

Application : Loi du 0-1 (dite aussi du *tout ou rien*)

Introduisons d'abord quelques notations relatives à une suite de v.a. $X_1, ..., X_n, ...$ Soit $\mathcal{A}_n^m := \sigma(X_n, ..., X_m)$ la *plus petite sous-tribu* sur $(\Omega, \mathcal{A})$ rendant mesurables les $X_n, ..., X_m$. Il est clair que $\mathcal{A}_n^m = \sigma(\mathcal{C}_n^m)$ où

$$\mathcal{C}_n^m := \left\{ \bigcap_{i=n}^m \{X_i \in A_i\},\ A_i \in \mathcal{E}_i,\ i = n, ..., m \right\}$$

est stable par intersection finie.

En général une réunion, même croissante, de tribus n'étant pas une tribu, on définit $\mathcal{A}_n^\infty := \sigma(\mathcal{A}_n^m,\ m \geq n)$ la plus petite tribu contenant $\bigcup_{m\geq n}^{\uparrow} \mathcal{A}_n^m$ (famille stable par intersection finie). Enfin, on définit la *tribu asymptotique* relative à la suite $(X_n)_{n\geq 1}$ par $\mathcal{A}^\infty := \bigcap_{n\geq 1}^{\downarrow} \mathcal{A}_n^\infty$.

Théorème :
Si la suite $X_1, ..., X_n, ...$ est constituées de variables aléatoires indépendantes, alors :

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

Démonstration : Il est clair au vu de la remarque ci-dessus que, pour $m \geq n+1$, $\mathcal{A}_1^n = \sigma(\mathcal{C}_1^n)$ et $\mathcal{A}_{n+1}^m = \sigma(\mathcal{C}_{n+1}^m)$ sont indépendantes. Donc $\mathcal{A}_1^n$ et $\mathcal{A}_{n+1}^\infty$ le sont aussi car $\bigcup_{m\geq n+1}^{\uparrow} \mathcal{A}_{n+1}^m$ est stable par intersection finie. Comme $\mathcal{A}^\infty \subset \mathcal{A}_{n+1}^\infty$, $\mathcal{A}^\infty$ est indépendante de $\mathcal{A}_1^n$ pour tout $n \geq 1$ et, partant, $\mathcal{A}^\infty$ est indépendante de $\sigma(\mathcal{A}_1^n, n \geq 1) = \mathcal{A}_1^\infty$ (cf. Remarque).

Or, par définition, $\mathcal{A}^\infty \subset \mathcal{A}_1^\infty$. Mais alors $\mathcal{A}^\infty$ est indépendante de $\mathcal{A}^\infty$ donc si $A \in \mathcal{A}^\infty$, $\mathbb{P}(A \cap A) = \mathbb{P}(A)^2$ *i.e.* $\mathbb{P}(A) = \mathbb{P}(A)^2 \Leftrightarrow \mathbb{P}(A) = 0$ ou 1. #

Exemples : • Soit $(c_n)_{n\geq 1}$ une suite de réels et $X_n, n \geq 1$, une suite de v.a. indépendantes. Alors

$$\sum_{n\geq 1} c_n X_n \text{ converge dans } \mathbb{R}\ \mathbb{P}\text{-}p.s. \text{ ou diverge } \mathbb{P}\text{-}p.s..$$

Ceci découle directement du théorème, une fois montré que l'événement $C_\infty := \left\{ \sum_{k\geq 1} c_k X_k \text{ converge dans } \mathbb{R} \right\} \in \mathcal{A}^\infty$ et que la limite $p.s.$ (éventuelle) est $\mathcal{A}^\infty$-mesurable donc $p.s.$ constante. Or $C_\infty \in \mathcal{A}_1^\infty$ car pour tout $n \geq 1$, $\sum_{k=1}^n c_k X_k$ est $\mathcal{A}_1^\infty$-mesurable. De même, pour tout $n \geq 1$, $C_\infty = \left\{ \sum_{k\geq n}^{n} c_k X_k \text{ converge dans } \mathbb{R} \right\}$ est $\mathcal{A}_n^\infty$-mesurable car la convergence d'une série ne dépend pas de ses premiers termes. D'où $C_\infty \in \mathcal{A}^\infty = \cap_n \mathcal{A}_n^\infty$.

• Sous les mêmes hypothèses sur les X_n, on montre que

$$\frac{X_1 + \cdots + X_n}{n} \text{ converge dans } \mathbb{R}\ \mathbb{P}\text{-}p.s. \text{ ou diverge } \mathbb{P}\text{-}p.s..$$

En outre, en cas de convergence $p.s.$, il existe $\ell \in \mathbb{R}$ tq $\sum_{n\geq 1} c_n X_n = \ell$ $\mathbb{P}$-$p.s.$ car la limite au sens de Césaro ne dépend pas des premiers termes de la suite.

Remarque : L'existence d'une suite infinie de v.a. indépendantes a été ici implicitement admise.

5.4 Exemples d'application au calcul de lois

5.4.1 Loi de la somme de v.a. indépendantes à valeurs dans $\mathbb{R}^d$ (ou $\mathbf{C}^d$)

Nous allons commencer par définir la convolution des mesures.

Définition :
Soient μ_1 et μ_2 deux probabilités sur $\left(\mathbb{K}^d, \mathcal{B}\left(\mathbb{K}^d\right)\right)$ ($\mathbb{K}$=$\mathbb{R}$ ou $\mathbf{C}$). La convolution de μ_1 et μ_2, notée $\mu_1 * \mu_2$, est définie comme la mesure image de la mesure produit $\mu_1 \otimes \mu_2$ sur $\left(\mathbb{K}^d \times \mathbb{K}^d, \mathbb{B}\left(\mathbb{K}^d\right)^{\otimes 2}\right)$ par l'application (continue donc mesurable)

$$\begin{array}{ccc} \mathbb{K}^d \times \mathbb{K}^d & \to & \mathbb{K}^d \\ (x,y) & \mapsto & x+y \end{array}.$$

Proposition :
Soit $f : \mathbb{K}^d \to \mathbb{R}$ borélienne bornée ou positive. Alors :

$$\int_{\mathbb{K}^d} f(z)\, \mu_1 * \mu_2 (dz) = \iint_{\mathbb{K}^d \times \mathbb{K}^d} f(x_1 + x_2)\, \mu_1 \otimes \mu_2 (dx_1, dx_2)$$

et l'égalité s'étend à toute f telle que l'une des deux intégrales ait un sens.

Démonstration : Il suffit d'appliquer le théorème de caractérisation de la mesure image (cf. II.5.2). #

A ce stade le théorème de Fubini fournit pour toute $f \in \mathcal{L}^1(\mu_1 * \mu_2)$, l'identité

$$(*) \quad \int f(z)\, \mu_1 * \mu_2 (dz) = \int \left[\int f(x_1+x_2)\, \mu_1(dx_1)\right] \mu_2(dx_2) = \int \left[\int f(x_1+x_2)\, \mu_2(dx_2)\right] \mu_1(dx_1)$$

Application : Si $\mu_1(dx_1) = f_1(x_1)dx_1$ et $\mu_2(dx_2) = f_2(x_2)\,dx_2$, alors :

$$\mu_1 * \mu_2 (dz) = f_1 * f_2(z)\, dz \text{ où } f_1 * f_2(z) := \int f_1(z-x) f_2(x)\, dx = \int f_1(x) f_2(z-x)\, dx.$$

Démonstration : Soit f borélienne bornée ou positive ; d'après l'identité $(*)$

$$\begin{aligned} \int f(z)\, \mu_1 * \mu_2(dz) &= \int \left[\int f(x_1+x_2) f_1(x_1) dx_1\right] f_2(x_2)\, dx_2 \\ &= \int \left[\int f(z) f_1(z-x_2)\, dz\right] f_2(x_2)\, dx_2 \text{ (en posant } x_1 := z - x_2) \\ &= \int f(z) \left[\int f_1(z-x_2)\, f_2\right] dz \text{ (d'après le théorème de Fubini),} \end{aligned}$$

d'où $\mu_1 * \mu_2(dz) = f_1 * f_2(z)\, dz$. #

Théorème :
Si X, Y sont deux v.a. sur $(\Omega, \mathcal{A}, \mathbb{P})$ indépendantes, à valeurs dans $\mathbb{K}^d$, alors $\mathbb{P}_{X+Y} = \mathbb{P}_X * \mathbb{P}_Y$.

Démonstration : Soit f une fonction borélienne bornée ;

$$\begin{aligned} \mathbb{E}(f(X+Y)) &= \int f(x+y) \mathbb{P}_{(X,Y)}(dx,dy) \text{ d'après le théorème de caractérisation,} \\ &= \int f(x+y) \mathbb{P}_X \otimes \mathbb{P}_Y (dx,dy) \text{ car } \mathbb{P}_{(X,Y)} = \mathbb{P}_X \otimes \mathbb{P}_Y \text{ par indépendance,} \\ &= \int f(z) \mathbb{P}_X * \mathbb{P}_Y(dz) \text{ d'après la proposition ci-avant.} \end{aligned}$$

D'où, finalement, $\mathbb{P}_{X+Y} = \mathbb{P}_X * \mathbb{P}_Y$ d'après le théorème de caractérisation. #

Exemple d'application : Soient X et Y deux v.a.r. *indépendantes* sur un espace $(\Omega, \mathcal{A}, \mathbb{P})$ de lois respectives $X \overset{\mathcal{L}}{\sim} \gamma(\alpha, \beta)$ et $Y \overset{\mathcal{L}}{\sim} \gamma(\alpha', \beta)$. Alors

$$X + Y \overset{\mathcal{L}}{\sim} \gamma(\alpha + \alpha', \beta).$$

Par définition, les densités respectives de X et Y sont

$$f(x) = \frac{1}{\beta^\alpha \Gamma(\alpha)} e^{-x/\beta} x^{\alpha-1} \mathbf{1}_{\mathbb{R}_+^*}(x) \text{ et } g(y) = \frac{1}{\beta^{\alpha'} \Gamma(\alpha')} e^{-y/\beta} y^{\alpha'-1} \mathbf{1}_{\mathbb{R}_+^*}(y).$$

$$\text{D'où } f * g(z) = C^{te} \times \int e^{-\frac{z-x}{\beta}} (z-x)^{\alpha-1} e^{-\frac{x}{\beta}} x^{\alpha'-1} \mathbf{1}_{\mathbb{R}_+^*}(z-x)\, \mathbf{1}_{\mathbb{R}_+^*}(x)\, dx.$$

Or

$$\mathbf{1}_{\mathbb{R}_+^*}(z-x)\, \mathbf{1}_{\mathbb{R}_+^*}(x) = \mathbf{1}_{]0,z[}(x) \text{ si } z > 0 \text{ et } 0 \text{ si } z \leq 0.$$

Donc, si $z > 0$,

$$\begin{aligned} f * g(z) &= C^{te} e^{-z/\beta} \int_0^z (z-x)^{\alpha-1} x^{\alpha'-1} dx \\ &\overset{x:=zu}{\underset{\uparrow}{=}} C^{te} e^{-z/\beta} z^{\alpha-1+\alpha'-1+1} \int_0^1 (1-u)^{\alpha-1} u^{\alpha'-1} du \\ &= C^{te} e^{-z/\beta} z^{\alpha+\alpha'-1} \text{ si } z > 0 \end{aligned}$$

et $f * g(z) = 0$ si $z \leq 0$.
Finalement, comme l'on *sait* que $f * g$ est une *densité de probabilité*, il vient :

$$f * g(z) = \frac{1}{\beta^{\alpha+\alpha'} \Gamma(\alpha+\alpha')} e^{-z/\beta} z^{\alpha+\alpha'-1} \mathbf{1}_{\mathbb{R}_+^*}(z) \text{ i.e. } X + Y \overset{\mathcal{L}}{\sim} \gamma(\alpha+\alpha', \beta). \ \#$$

Remarque : Outre le résultat recherché, on obtient, en explicitant les constantes rencontrées dans les calculs, l'identité :

$$\frac{1}{\beta^\alpha}\frac{1}{\Gamma(\alpha)}\frac{1}{\beta^{\alpha'}}\frac{1}{\Gamma(\alpha')}\times\underbrace{\int_0^1(1-u)^{\alpha-1}u^{\alpha'-1}du}_{:=B(\alpha,\alpha')}=\frac{1}{\beta^{\alpha+\alpha'}}\frac{1}{\Gamma(\alpha+\alpha')}$$

$$i.e.\ \ \frac{\Gamma(\alpha)\Gamma(\alpha')}{\Gamma(\alpha+\alpha')}=B(\alpha,\alpha').$$

Cette identité relie la fonction Γ à la fonction "bêta" de 1ère espèce B. La loi de densité $\frac{1}{B(\alpha,\alpha')}x^\alpha(1-x)^{\alpha'-1}\mathbf{1}_{]0,1[}(x)$ sur $[0,1]$ est appelée *loi β de 1ère espèce.*

Corollaire (de l'exemple) :

Rappelons que la loi du $\chi^2(n)$ est définie comme une loi $\gamma\left(\frac{n}{2};2\right)$. On a donc le résultat suivant :

$X_i \overset{\mathcal{L}}{\sim} \chi^2(n_i), i=1,2,\ X_1$ et X_2 indépendantes, alors $X_1+X_2 \overset{\mathcal{L}}{\sim} \chi^2(n_1+n_2)$.

D'autre part, on a vu (cf. III.3. exemple (3)) que si $X \overset{\mathcal{L}}{\sim} \mathcal{N}(0;1)$, alors $X^2 \overset{\mathcal{L}}{\sim} \chi^2(1)$. En conséquence, par une récurrence immédiate :

$$X=(X_1,...,X_n)\overset{\mathcal{L}}{\sim}\mathcal{N}(0;I_n)\Rightarrow X_1^2+...+X_n^2\overset{\mathcal{L}}{\sim}\chi^2(n).$$

C'est cette propriété qui motive la mise en exergue des lois du χ^2, notamment pour leur utilité en statistique. D'autres méthodes de calcul de lois de somme seront étudiées plus loin.

5.4.2 Lois du minimum et du maximum de n v.a.r indépendantes de même loi

Soient $X_1,...,X_n$, n v.a.r. indépendantes de même loi μ. On notera F leur fonction de répartition. On pose

$$M_n := \max_{1\le i\le n} X_i \quad \text{et} \quad m_n := \min_{1\le i\le n} X_i.$$

Proposition :
(a) M_n a pour fonction de répartition $x\mapsto F^n(x)$ et
m_n a pour fonction de répartition $x\mapsto 1-(1-F(x))^n$.
(b) Si $\mu(dx)=f(x)dx$ alors M_n a pour densité $nF^{n-1}(x)f(x)$ et m_n a pour densité $n(1-F(x))^{n-1}f(x)$.

Démonstration : (a) $\mathbb{P}(M_n\le x) = \mathbb{P}\left(\max_{1\le i\le n}X_i\le x\right) = \mathbb{P}\left(\bigcap_{i=1}^n\{X_i\le x\}\right)$
$= \prod_{i=1}^n \mathbb{P}(X_i\le n) = F(x)^n.$

$\mathbb{P}(m_n\le x)=1-\mathbb{P}(m_n>x)=1-\mathbb{P}\left(\bigcap_{i=1}^n\{X_i>x\}\right)=1-(1-F(x))^n.$

(b) Si f est continue – sauf éventuellement en un nombre fini de points – il suffit de procéder par dérivation. Sinon on a recours au lemme suivant généralisant la formule d'intégration par parties.

Lemme :

Soient μ et ν deux probabilités sur $\mathbb{R}$ de densités respectives f et g et de répartitions F et G. Alors :

$$\forall x\in\mathbb{R},\ \int_{-\infty}^x G(u)f(u)du+\int_{-\infty}^x g(u)F(u)du=F(x)G(x)$$

Démonstration : D'après le théorème de Fubini, on a les égalités :

$$\int_{-\infty}^x G(u)f(u)du=\int\mathbf{1}_{\{v\le u\le x\}}g(v)f(u)dudv,$$
$$\int_{-\infty}^x g(u)F(u)du=\int\mathbf{1}_{\{v\le u\le x\}}g(u)f(v)dudv=\int\mathbf{1}_{\{u\le v\le x\}}g(v)f(u)dudv.$$

D'où, en sommant et en notant que $\iint\mathbf{1}_{(u=v)}dudv=0$,

$$\int_{-\infty}^x G(u)f(u)du+\int_{-\infty}^x g(u)F(u)du=\int f(u)g(v)\mathbf{1}_{]-\infty,x]}(u)\mathbf{1}_{]-\infty,x]}(v)dudv=F(x)G(x).\ \#$$

On déduit du lemme que $2\int_{-\infty}^x F(u)f(u)du=F^2(u)$, puis par récurrence sur n, que $n\int_{-\infty}^x F^{n-1}(u)f(u)du=F^n(x)$. M_n a donc pour densité (cf.III.4) : $nF^{n-1}f$.

On procède de même pour m_n. #

Exemple : Si les X_i sont indépendantes de loi $\mathcal{E}(\lambda),\lambda>0, i=1,...,n$ alors $m_n\overset{\mathcal{L}}{\sim}\mathcal{E}(n\lambda)$; en effet $F(x)=\left(1-e^{-\lambda x}\right)\mathbf{1}_{\mathbb{R}_+}(x)$, donc $\mathbb{P}(m_n\le x)=(1-e^{-n\lambda x})\mathbf{1}_{\mathbb{R}_+}(x)$.

Remarque : Ce résultat est un cas particulier des problèmes de statistique d'ordre, *i.e.* de l'étude de la loi du réordonnement croissant $\left(X^{(1)},...,X^{(n)}\right)$ d'un vecteur aléatoire $(X_1,...,X_n)$ défini par :

$$\left(X^{(1)},...,X^{(n)}\right):=\begin{cases}\left(X_{\sigma(1)},...,X_{\sigma(n)}\right) & \text{sur } \left\{X_{\sigma(1)}<...<X_{\sigma(n)}\right\}\\ (0,...,0) & \text{sur } \{\omega/\exists i\neq j X_i(\omega)=X_j(\omega)\}\end{cases}.$$

5.5 Loi faible des grands nombres

C'est la première forme historiquement apparue de la loi des grands nombres (due à Bernoulli dans le cadre du jeu de pile ou face équilibré). Elle exprime que lorsque l'on fait la moyenne empirique $\overline{X}_n := \dfrac{X_1 + ... + X_n}{n}$ d'un échantillon de v.a. indépendantes de même loi celle-ci a tendance à se rapprocher, en un certain sens, de la moyenne $\mathbb{E}(X_1)$ de X_1.

<u>**Théorème**</u> **:**
Soit $X_n, n \geq 1$, une suite de v.a. 2 à 2 *non corrélées*, de carré intégrable et de même loi. Alors

$$\forall \varepsilon > 0 \quad \mathbb{P}\left(\left|\overline{X}_n - \mathbb{E}\left(X_1\right)\right| \geq \varepsilon\right) \leq \frac{Var\left(X_1\right)}{n\varepsilon^2} \underset{n\to+\infty}{\longrightarrow} 0.$$

<u>Démonstration</u> : On vérifie par linéarité que $\mathbb{E}\left(\overline{X}_n\right) = \mathbb{E}(X_1)$ puis, via l'inégalité de Bienaymé-Tchebycheff que, pour tout $\varepsilon > 0$,

$$\begin{aligned}\mathbb{P}\left(\left|\overline{X}_n - \mathbb{E}\left(X_1\right)\right| \geq \varepsilon\right) \leq \frac{Var\left(\overline{X}_n\right)}{\varepsilon^2} &= \frac{1}{n^2\varepsilon^2} Var\left(X_1 + ... + X_n\right) \\ &= \frac{nVar\left(X_1\right)}{n^2\varepsilon^2} \text{ par non corrélation,} \\ &= \frac{Var(X_1)}{n\varepsilon^2}. \quad \#\end{aligned}$$

Chapitre 6

Simulation de variables aléatoires

Dans nombre de modèles probabilistes, indépendamment de l'estimation (statistique) de leurs paramètres, le ou les résultats recherchés se ramènent à des calculs d'espérance. Généralement le calcul analytique direct ou par approximation de ces espérances est inaccessible. On a alors recours à des méthodes de simulation consistant, pour l'essentiel, à mettre en œuvre informatiquement la loi des grands nombres. Cette approche a pour nom méthode de Monte-Carlo. Elle doit l'essentiel de son succès à l'émergence de moyens de calcul puissants, les ordinateurs, permettant la répétition "virtuelle" d'un phénomène aléatoire un nombre élevé de fois. Signalons cependant pour mémoire le calcul de π par la méthode de l'aiguille de Buffon qui montre que l'idée d'utiliser la loi des grands nombres pour le calcul d'espérance est bien antérieure à l'électronique !

Après des débuts sous l'égide des militaires, à la fin de la seconde guerre mondiale, pour la résolution de problèmes de particules liés au développement des armes nucléaires (le centre de Los Alamos aux Etats-Unis, le C.E.A. en France, la recherche militaire en ex-U.R.S.S., ...), les applications de la simulation probabiliste sont aujourd'hui omniprésentes, de l'ingénierie financière (options exotiques, américaines) aux télécommunications (dimensionnement de centraux téléphoniques, etc) en passant, sur un plan plus théorique, par la résolution d'équations intégro-différentielles (équation de Boltzmann).

On distingue en simulation deux champs d'investigation relativement distincts. Le premier concerne la définition, la génération et la validation (statistique) de **suites de nombres au hasard** à valeurs dans $\{0,1\}$ ou $[0,1]$. Le second, de nature plus directement probabiliste, consiste, à partir de ces nombres au hasard, supposés "parfaits", à développer des méthodes de simulation de nombres distribués selon des lois diverses (normale, gamma, ...). Pour notre part, nous aborderons surtout le second problème, partant du principe, malheureusement erroné en général, que les suites de nombres au hasard disponibles sur les ordinateurs (instruction "RANDOM") sont satisfaisantes.

6.1 Nombres aléatoires

Sur un plan strictement mathématique, la définition d'une suite de nombres aléatoires (uniformément répartis) dans $[0,1]$ doit évidemment être :

"**Définition**" :
Une suite $x_n, n \geq 1$, de nombres de $[0,1]$ est une suite de nombres aléatoires s'il existe un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$, une suite $U_n, n \geq 1$, de v.a. indépendantes de loi $\mathcal{U}([0,1])$ et $\omega \in \Omega$ tels que $x_n = U_n(\omega)$ pour tout $n \geq 1$.

Mais cette même définition naïve - et abstraite - n'est pas satisfaisante car l'état du monde $\omega \in \Omega$ en question est-il "bon"*i.e.* "générique"? Nombre de propriétés probabilistes (cf. loi des grands nombres) ne sont vérifiées que $\mathbb{P}$-*p.s.*. Ainsi, si ω est précisément dans un ensemble négligeable ne vérifiant pas la propriété requise, il n'est pas générique.

Quoi qu'il en soit, on ne dispose en général pas de suite (U_n) de v.a. *a priori* indépendantes, de loi $\mathcal{U}([0,1])$! Parallèlement, les travaux de logiciens comme Martin-Löf ont conduit à considérer que toute suite (x_n) générable par un algorithme ne saurait être considérée comme aléatoire. Définition embarrassant pour des objets dont l'usage requiert une génération presque instantanée sur ordinateur ! Ainsi les décimales de π ne sont pas aléatoires au sens de Martin-Löf puisque précisément décimales de π.

L'approche des informaticiens théoriques, imposant à l'algorithme générateur d'avoir une complexité proportionnelle à la longueur de la suite générée, ne résoud guère la question. Parallèlement, le besoin sans cesse croissant de telles suites de nombres, à la fois les plus longues possibles et les plus rapides à fabriquer, a conduit à imposer une approche plus pragmatique – nous dirons heuristique – fondée sur la recherche de suites vérifiant un certain nombre de critères statistiques d'équirépartition et d'indépendance (tests d'équirépartition par paquets, tests de non corrélation, tests de rang, etc).

Dans la pratique, de telles suites étant, comme l'est la précision des ordinateurs, finies, on considère des suites (x_n) de nombres que nous appellerons pseudo-aléatoires de la forme :

$$x_n = \frac{y_n}{N}, \;\; y_n \in \{0, \ldots, N-1\}.$$

L'un des procédés classiques consiste à obtenir les y_n par une récurrence congruente du type :

$$y_{n+1} \equiv a y_n + b \mod. N$$

où $\mathrm{pgcd}(a, N) = 1$, de façon à ce que $\overline{a}$ soit inversible dans le groupe multiplicatif $(\mathbf{Z}/N\mathbf{Z})^*$ des inversibles de $\mathbf{Z}/N\mathbf{Z}$ pour la multiplication.

Si $b = 0$ (cas le plus courant), on parle de *générateur homogène*.

On choisira N aussi grand que possible en fonction des capacités de calcul en nombres entiers de la machine ($N = 2^{31} - 1$ pour une machine 32 *bits*, etc).

Toujours si $b = 0$, la longueur de la suite sera determinée par la période $\tau := \min\{t \;/\; a^t \equiv 1 \mod. N\}$ de a dans $((\mathbf{Z}/N\mathbf{Z})^*, +)$.

Comme $\mathrm{card}(\mathbf{Z}/N\mathbf{Z})^* = \varphi(N)$ où $\varphi(N) := \mathrm{card}\{1 \le k \le N-1,\ \text{tq pgcd}(k,N)=1\}$ est l'indicatrice d'Euler, il vient, d'après le théorème de Lagrange :

$$\tau = \mathrm{card}(<\overline{a}>) \mid \varphi(N)$$

(où $<\overline{a}> :=$ sous-groupe multiplicatif de $(\mathbf{Z}/N\mathbf{Z})^*$ engendré par $\overline{a}$). Rappelons que :

$$\varphi(N) = N \prod_{p|N,\ p \text{ premier}} \left(1 - \frac{1}{p}\right).$$

L'étude (difficile) du groupe $((\mathbf{Z}/N\mathbf{Z})^*, \times)$ lorsque N est un entier primaire conduit au théorème suivant :

Théorème :
Soit $N = p^\alpha$, p premier, $\alpha \in \mathbb{N}^*$.
(a) Si $\alpha = 1$ (i.e. $N = p$ premier), alors $((\mathbf{Z}/N\mathbf{Z})^*, \times)$ (dont le cardinal est $p-1$) est un groupe cyclique. En d'autres termes, il existe $\overline{a} \in \{1, \ldots, p-1\}$ t.q. $(\mathbf{Z}/p\mathbf{Z})^* = <\overline{a}>$. La période maximale est donc $\tau = p - 1$.

(b) Si $p = 2$, $\alpha \ge 3$, $(\mathbf{Z}/N\mathbf{Z})^*$, dont le cardinal est $2^{\alpha-1} = \dfrac{N}{2}$, n'est pas cyclique. La période maximale pouvant être atteinte est alors $\tau = 2^{\alpha-2}$ avec $a \equiv \pm 3$ mod. 8.
(c) Si $p \ne 2$, alors $(\mathbf{Z}/N\mathbf{Z})^*$, dont le cardinal est $p^{\alpha-1}(p-1)$, est cyclique, donc $\tau = p^{\alpha-1}(p-1)$. Il est engendré par tout élément a dont la classe $\tilde{a}$ dans $(\mathbf{Z}/p\mathbf{Z})$ engendre le groupe cyclique $((\mathbf{Z}/p\mathbf{Z})^*, \times)$.

Cependant, on prendra garde que la longueur d'une suite, si elle est une qualité indispensable, ne garantit en rien sa qualité lors des tests statistiques ! Ainsi le générateur de la bibliothèque FORTRAN IMSL n'entre pas dans les cadres précédents : on prend $N := 2^{31} - 1$ (qui est un nombre premier), $a := 7^5, b := 0$ ($a \not\equiv 0$ mod.8).

En fin de chapitre, est proposé un générateur programmé en Pascal, dû à Knuth. Ses paramètres sont $N := 10^8$ et $a := 31452821$. Il ne rentre pas dans le cadre du théorème puisque N n'est pas primaire ($10 = 2 \times 5$).

6.2 Simulation de lois non uniformes : quelques exemples importants

Dans cette section, on fait l'hypothèse que l'on dispose d'une suite "idéale" de nombres pseudo-aléatoires. On souhaite, à partir de ceux-ci, engendrer des nombres distribués selon une loi μ sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. Le problème se ramène donc à trouver des fonctions $\varphi : \mathbb{R}^m \to \mathbb{R}^d$ (simples) telles que $X = \varphi(U_1, \ldots, U_m)$ ait pour loi μ si les $U_1, \ldots, U_m$ sont indépendantes de loi $\mathcal{U}([0,1])$. En théorie, on peut même toujours choisir $m = 1$.

En pratique, pour assurer la simplicité de φ, on prend $m \ge 2$ ou d, voire $m = +\infty$ (méthode du rejet).
On introduit alors la notion de rendement d'une méthode de simulation par

$$\rho := \frac{\text{Quantité de nombres distribués selon } \mu}{\text{Quantité de nombres (pseudo-)aléatoires employés}}.$$

Comme nous le verrons plus bas, cette quantité est généralement une v.a. (non constante). Lorsque tel sera le cas on s'attachera à calculer l'espérance de ce rendement aléatoire.

6.2.1 Simulation d'une v.a. à valeurs dans un ensemble fini

Soit μ une probabilité sur un espace fini $E = \{x_1, \ldots, x_n\}$ définie par :

$$\mu(\{x_i\}) = p_i \in [0,1],\ 1 \le i \le n.$$

On simule μ à partir d'une v.a.r. U de loi $\mathcal{U}([0,1])$ en posant :

$$X := x_1 \mathbf{1}_{\{U \le p_1\}} + x_2 \mathbf{1}_{\{p_1 < U \le p_1+p_2\}} + \ldots + x_k \mathbf{1}_{\{p_1+\ldots+p_{k-1} < U \le p_1+\ldots+p_k\}} + \ldots + x_n \mathbf{1}_{\{p_1+\ldots+p_{n-1} < U \le 1\}}.$$

Le rendement de la simulation est ici clairement $\rho = 1$.

6.2.2 Loi binomiale $\mathcal{B}(n,p)$, $p \in [0,1]$

On peut utiliser la méthode générale précédente, mais les risques d'erreur d'arrondi la pénalisent fortement (calcul des $C_n^k p^k (1-p)^{n-k}$).
En revanche, à partir de n v.a. $U_1, \ldots, U_n$ indépendantes de loi $\mathcal{U}([0,1])$, on peut poser :

$$X := \sum_{i=1}^n \mathbf{1}_{\{U_i \le p\}}.$$

Les U_i étant indépendantes, les $\mathbf{1}_{\{U_i \le p\}}$ le sont aussi, de loi $\mathcal{B}(p)$, donc $\mathcal{L}(X) = \mathcal{B}(n;p)$.
Le rendement de la simulation est cependant médiocre si n est grand : $\rho = \frac{1}{n}$.

6.2.3 Loi géométrique $\mathcal{G}(p)$, $p \in]0,1]$

On pose simplement :

$$X := \min\{\ell \ge 1 \ /\ U_\ell \le p\}.$$

On vérifie que $\{X = k\} = \left(\bigcap_{i=1}^{k-1} \{U_i > p\}\right) \cap \{U_k \le p\}$ donc, ces événements étant indépendants, on vérifie immédiatement que $\mathbb{P}(X = k) = (1-p)^{k-1} p,\ k \ge 1$.
La quantité de nombres aléatoires nécessaire - qui n'est autre que X - est *p.s.* finie mais non bornée.

Le rendement moyen de la simulation est

$$\rho = \mathbb{E}\left(\frac{1}{X}\right) = \sum_{k\geq 1}\frac{1}{k}p(1-p)^{k-1} = \frac{p}{1-p}\sum_{k\geq 1}\frac{(1-p)^k}{k} = \frac{-p\ln p}{1-p}.$$

On notera que l'inégalité de Jensen entraîne immédiatement $\rho \geq \dfrac{1}{\mathbb{E}(X)} = p$.

6.2.4 Loi exponentielle $\mathcal{E}(\lambda)$, $\lambda > 0$

On pose, à partir d'une v.a. U de loi $\mathcal{U}([0,1])$,

$$X := -\frac{1}{\lambda}\ln(U).$$

X est *p.s.* définie car $\mathbb{P}(X=0)=0$. D'autre part la v.a. X est positive et, pour tout $x \geq 0$, $\mathbb{P}(X \leq x) = \mathbb{P}(U \geq e^{-\lambda x}) = 1 - e^{-\lambda x}$. Donc $\mathcal{L}(X) = \mathcal{E}(\lambda)$.

Le rendement de la simulation est $\rho = 1$. En fait, cette méthode est un cas particulier d'une méthode plus générale (cf. paragraphe 3.1).

6.2.5 Loi normale centrée réduite sur $\mathbb{R}$ (et $\mathbb{R}^2$)

La méthode la plus efficace pour générer des nombres selon une loi $\mathcal{N}(0,1)$ s'appuie sur la proposition suivante :

Proposition :
Soient R^2 et $\Theta : (\Omega, \mathcal{A}, \mathbb{P}) \to \mathbb{R}$ deux v.a. indépendantes et de lois respectives $\mathcal{L}(R^2) = \mathcal{E}(\frac{1}{2})$ et $\mathcal{L}(\Theta) = \mathcal{U}([0,2\pi])$. Alors $X := (R\cos\Theta, R\sin\Theta)$ suit une loi $\mathcal{N}(0, I_2)$ (où $R := \sqrt{R^2}$).

Démonstration : Soit f borélienne bornée. On va procéder "à l'envers" :

$$\iint_{\mathbb{R}^2} f(x_1,x_2)\exp(-\frac{x_1^2+x_2^2}{2})\frac{dx_1dx_2}{2\pi} = \iint f(\rho\cos\theta, \rho\sin\theta)e^{-\frac{\rho^2}{2}}\mathbf{1}_{\mathbb{R}_+^*}(\rho)\mathbf{1}_{]0,2\pi[}(\theta)\rho\frac{d\rho d\theta}{2\pi}$$

par le changement de variables standard : $x_1 = \rho\cos\theta, x_2 = \rho\sin\theta$. Si l'on pose ensuite $\rho = \sqrt{r}$, il vient :

$$\iint_{\mathbb{R}^2} f(x_1,x_2)\exp(-\frac{x_1^2+x_2^2}{2})\frac{dx_1dx_2}{2\pi} = \iint f(\sqrt{r}\cos\theta, \sqrt{r}\sin\theta)\frac{e^{-\frac{r}{2}}}{2}\mathbf{1}_{\mathbb{R}_*^+}(\rho)\mathbf{1}_{]0,2\pi[}(\theta)\frac{drd\theta}{2\pi}$$
$$= \mathbb{E}\left(f(\sqrt{R^2}\cos\Theta, \sqrt{R^2}\sin\Theta)\right) = \mathbb{E}(f(X)). \ \#$$

Corollaire :

On peut simuler la loi $\mathcal{N}(0;I_2)$ à partir d'un couple (U_1,U_2) de v.a. indépendantes de loi $\mathcal{U}([0,1])$ en posant :

$$X := \left(\sqrt{-2\ln(U_1)}\cos(2\pi U_2), \sqrt{-2\ln(U_1)}\sin(2\pi U_2)\right).$$

Le rendement de la simulation est $\rho = 1$.

Démonstration : Il suffit d'appliquer la méthode de simulation du paragraphe précécent (exponentielle) et de remarquer que si $U \sim \mathcal{U}([0,1])$, $\mathcal{L}(2\pi U) = \mathcal{U}([0,2\pi])$. #

6.3 Méthodes générales de simulation de v.a.r.

6.3.1 Méthode(s) de la fonction de répartition

Soit μ une loi de probabilité sur $\mathbb{R}$ ayant une fonction de répartition F strictement croissante et continue. F admet alors une fonction réciproque F^{-1} définie sur $]0,1[$.

Proposition :
Si $\mathcal{L}(U) = \mathcal{U}([0,1])$, alors $X := F^{-1}(U)$ a pour loi μ.

Démonstration : Soit $x \in \mathbb{R}$, $\mathbb{P}(X \leq x) = \mathbb{P}(F^{-1}(U) \leq x)$. Or F^{-1} est strictement croissante, donc $\{F^{-1}(U) \leq x\} = \{U \leq F(x)\}$. D'où $\mathbb{P}(X \leq x) = F(x)$. #

Remarques : • Si μ a une densité f vérifiant $\{f = 0\}$ est d'intérieur vide, alors $F(x) = \int_{-\infty}^{x} f(u)du$ est continue, strictement croissante.
• On peut remplacer $\mathbb{R}$ par un intervalle $[a,b] \subset \overline{\mathbb{R}}$.

• Lorsque F n'est pas strictement croissante (ou pas continue), on définit la réciproque canonique continue à droite par :

$$\forall t \in]0,1[, \quad F_d^{-1} = \inf\{s/F(s) > t\}.$$

On montre que F_d^{-1} est croissante, continue à droite et que

$$F_d^{-1}(u) \leq v \Rightarrow F(v) \geq u \text{ et } F(v) > u \Rightarrow F_d^{-1}(u) \leq v.$$

Par suite, si $X = F_d^{-1}(U)$,

$$\mathbb{P}(X \leq x) = \mathbb{P}(F_d^{-1}(U) \leq x)\left\{\begin{array}{l} \leq \mathbb{P}(F(x) \geq U) = F(x) \\ \geq \mathbb{P}(F(x) > U) = F(x) \end{array}\right\} = F(x).$$

Si X est discrète à valeurs dans E fini, on retrouve ainsi la méthode de simulation du 2.1..

Exemple (Simulation d'une loi de Cauchy) :
Rappelons que $\mathcal{L}(X) = \text{Cauchy}(c)$ ssi $\mathbb{P}_X(dx) = \dfrac{c}{\pi(x^2+c^2)}dx$.

$$\forall x \in \mathbb{R},\quad F_X(x) = \int_{-\infty}^{x} \frac{c}{u^2+c^2}\frac{du}{\pi} = \frac{1}{\pi}\left(\text{Arctg}(\frac{x}{c}) + \frac{\pi}{2}\right)$$

d'où $F_X^{-1}(x) = c\,\text{tg}(\pi(u-\frac{1}{2}))$. Par suite,

$$\mathcal{L}(U) = \mathcal{U}([0,1]) \Rightarrow \mathcal{L}\left(c\,\text{tg}(\pi(U-\frac{1}{2}))\right) = \text{Cauchy}(c).$$

6.3.2 Méthode du rejet

La méthode du rejet est une méthode de simulation nécessitant de connaître la densité de la loi μ que l'on cherche à simuler. Soit f cette densité.

Proposition :
On suppose que f est bornée à support compact $[a,b]$. Soit $U_n, n\geq 1$, une suite de v.a. indépendantes de loi $\mathcal{U}([0,1])$. On pose, pour tout $n \geq 1$:

$$\begin{cases} V_n := a + (b-a)U_{2n-1} \\ W_n := MU_{2n} \end{cases} \qquad (\text{où } M = \|f\|_\infty := \sup_{x\in[a,b]} f(x)).$$

Soit enfin $\theta := \min\{n \,/\, W_n < f(V_n)\}$.

(a) $\theta \overset{\mathcal{L}}{\sim} G(\frac{1}{M(b-a)})$ (en particulier $\theta < +\infty$ $\mathbb{P}$-$p.s.$).

(b) $V_\theta \overset{\mathcal{L}}{\sim} f(x)dx$ et le rendement de la simulation est $\dfrac{1}{2\theta}$.

Démonstration : (a) On vérifie immédiatement que les v.a. $(V_n, W_n)_{n\geq 1}$ sont i.i.d. de loi $\mathcal{U}([a,b]\times[0,M])$. D'où il vient, pour tout $n \geq 1$,

$$\mathbb{P}(\theta \geq n) = \mathbb{P}(W_k \geq f(V_k), 1\leq k\leq n-1) = (1-\mathbb{P}(W_1 < f(V_1)))^{n-1} = (1-p)^{n-1}$$

avec $p := \mathbb{P}(W_1 < f(V_1)) = \displaystyle\int_a^b\int_0^M \mathbf{1}_{\{w<f(v)\}}\frac{dvdw}{M(b-a)} = \frac{\int_a^b f}{M\times(b-a)} = \frac{1}{M\times(b-a)} > 0.$

(b) Déterminons maintenant la loi de V_θ. Soit $\alpha \in \mathbb{R}$ (on peut évidemment supposer $\alpha \in [a,b]$).

$$\begin{aligned}\mathbb{P}(V_\theta \leq \alpha) = \sum_{n\geq 1}\mathbb{P}(\theta = n; V_n \leq \alpha) &= \sum_{n\geq 1}\mathbb{P}(W_k \geq f(V_k),\, 1\leq k\leq n-1,\, W_n < f(V_n),\, V_n \leq \alpha) \\ &= \sum_{n\geq 1}\mathbb{P}(W_1 \geq f(V_1))^{n-1}\mathbb{P}(W_1 < f(V_1),\, V_1 \leq \alpha) \\ &= \frac{\mathbb{P}(W_1 < f(V_1),\, V_1\leq\alpha)}{1-\mathbb{P}(W_1 \geq f(V_1))} = \frac{\mathbb{P}(W_1 < f(V_1),\, V_1\leq\alpha)}{\mathbb{P}(W_1 < f(V_1))} \\ &= M(b-a)\frac{\int_a^\alpha f(x)dx}{M(b-a)} = \int_{-\infty}^{\alpha} f(x)dx.\end{aligned}$$

L'identité des fonctions de répartition entraîne $\mathcal{L}(V_\theta) = f(x)dx$. #

Généralisation :
Lorsque f n'est pas à support compact mais cependant bornée, on procède comme suit : on considère une fonction $g : \mathbb{R} \to \mathbb{R}_+^*$ bornée par M, intégrable, telle que :

(i) $0 \leq f \leq g$, g continue,

(ii) $G(x) := \dfrac{\int_{-\infty}^x g(u)du}{\int_{-\infty}^{+\infty} g(u)du}$ admet une réciproque $H := G^{-1}$ "explicite" (i.e. telle que le calcul informatique de $H(x)$ soit peu "coûteux").

On pose alors : $\begin{cases} V_n := H(U_{2n+1}) \\ W_n := U_{2(n+1)} \end{cases}$ et $\theta := \min\{n \;/\; g(V_n)W_n < f(V_n)\}$. Alors

$$\mathcal{L}(V_\theta) = f(x)dx.$$

Démonstration : La démonstration est analogue à celle de la proposition précédente.

$$\mathbb{P}\left(g(V_1)W_1 < f(V_1)\right) = \mathbb{P}\left(U_2 < \frac{f}{g}(H(U_1))\right) = \int_{[0,1]^2}\mathbf{1}_{\{u_2<\frac{f}{g}(H(u_1))\}}du_1du_2 = \int_0^1 \frac{f}{g}(H(u))du.$$

Il vient, en considérant le changement de variables : $u = G(x)$ (G difféomorphisme de $\mathbb{R}$ sur $]0,1[$) :

$$\begin{aligned}\mathbb{P}\left(g(V_1)W_1 < f(V_1)\right) &= \int_{\mathbb{R}}\frac{f}{g}(x)G'(x)dx \\ &= \frac{1}{\int_{-\infty}^{+\infty} g(u)du}\int_{-\infty}^{+\infty} f(u)du \\ &= \frac{1}{\int_{-\infty}^{+\infty} g(u)du} \in [0,1[\end{aligned}$$

Or $\mathbb{P}(\theta \geq n) = (1-\mathbb{P}(g(V_1)W_1 < f(V_1)))^{n-1}$. En conséquence, θ suit une loi géométrique de paramètre $\frac{1}{\int_{\mathbb{R}} g(u)du}$ (donc $\mathbb{P}(\theta = +\infty) = 0$). Ensuite, si $\alpha \in \mathbb{R}$,

$$\mathbb{P}(V_\theta \leq \alpha) = \sum_{n\geq 1}\mathbb{P}\left(g(V_k)W_k \geq f(V_k), 1\leq k\leq n-1,\, g(V_n)W_n < f(V_n), V_n \leq \alpha\right).$$

et l'on conclut comme précédemment. #

Chapitre 7

Fonction caractéristique

La fonction caractéristique, définie comme la transformée de Fourier de la loi d'une v.a., est un outil puissant tant pour le calcul de loi que pour les théorèmes limites.

Le produit scalaire canonique sur $\mathbb{R}^d$ sera noté $(x|y) := \sum_{i=1}^{d} x_i y_i$ et la norme : $\|x\| := \sqrt{(x|x)}$.

7.1 Transformée de Fourier d'une probabilité, fonction caractéristique.

Définition :
(a) Soit μ une probabilité sur $\left(\mathbb{R}^d, \mathcal{B}\left(\mathbb{R}^d\right)\right)$. On appelle transformée de Fourier de μ, la fonction à valeurs complexes définie sur $\mathbb{R}^d$ par :

$$\forall u \in \mathbb{R}^d, \ \ \widehat{\mu}(u) := \int_{\mathbb{R}^d} e^{i(u|x)} \mu(dx) \ \ \left(\forall u \in \mathbb{R}^d \ e^{i(u|.)} \in \mathcal{L}^1_{\mathbf{C}}(\mu)!\right).$$

(b) Soit X : $(\Omega, \mathcal{A}, \mathbb{P}) \longrightarrow \mathbb{R}^d$ un vecteur aléatoire. On appelle fonction caractéristique X, notée Φ_X, la transformée de Fourier de la loi $\mathbb{P}_X$ de X.

Remarques : • Si $\mu = f.\lambda_d, f \in \mathcal{L}^1_{\mathbb{R}_+}(\lambda_d)$, alors $\widehat{\mu} = \widehat{f}$, $\widehat{f}$ transformée de Fourier de f. La définition ci-dessus est donc une généralisation de la transformée usuelle sur les fonctions de $\mathcal{L}^1(\lambda_d)$ $\left(\lambda_d \text{ désigne la mesure de Lebesgue sur } \mathbb{R}^d\right)$.
• La définition (a) s'étend immédiatement à une mesure positive finie voire une mesure signée.

D'après le théorème de la mesure image, la fonction caractéristique s'écrit :

$$\forall u \in \mathbb{R}^d, \ \ \Phi_X(u) := \int_{\mathbb{R}^d} e^{i(u|x)} \mathbb{P}_X(dx) = \mathbb{E}\left(e^{i(u|X)}\right).$$

Propriétés :
P1. $\widehat{\mu}(0) = 1$ et $\forall u \in \mathbb{R}^d, \ |\widehat{\mu}(u)| \le 1$.

P2. $\forall u, v \in \mathbb{R}^d, \ \ |\widehat{\mu}(u) - \widehat{\mu}(v)| \le \int_{\mathbb{R}^d} 2 \wedge (\|u-v\| \, \|x\|) \, \mu(dx)$. Par suite $\widehat{\mu}$ est uniformément continue sur $\mathbb{R}^d$.

P3. Si $\mu = f.\lambda_d$ alors $\lim_{\|u\| \to +\infty} \widehat{\mu}(u) = 0$ *i.e.* $\widehat{\mu} \in \mathcal{C}_0\left(\mathbb{R}^d, \mathbf{C}\right)$.

Démonstration : P1 est évidente.

P2 :
$$\begin{aligned}
|\widehat{\mu}(u) - \widehat{\mu}(v)| &\le \int_{\mathbb{R}^d} \left| e^{i(u|x)} - e^{i(v|x)} \right| \mu(dx) = \int_{\mathbb{R}^d} \left| e^{i\left(\frac{u-v}{2}|x\right)} - e^{i\left(\frac{v-u}{2}|x\right)} \right| \mu(dx) \\
&\le \int_{\mathbb{R}^d} 2 \left| \sin\left(\frac{u-v}{2}|x\right) \right| \mu(dx) \\
&\le \int_{\mathbb{R}^d} \min\left(2, 2\left|\left(\frac{u-v}{2}|x\right)\right|\right) \mu(dx) \quad \text{car } |\sin x| \le |x| \wedge 1, \\
&\le \int_{\mathbb{R}^d} \min(2, \|u-v\| \, \|x\|) \, \mu(dx) \quad (\text{inégalité de C.S.}).
\end{aligned}$$

L'uniforme continuité découle de cette inégalité par convergence dominée.
P3. Si $f = \mathbf{1}_{\prod_{k=1}^d [a_k, b_k]}$, (indicatrice de pavé),

$$\widehat{\mu}(u) = \prod_{k=1}^{d} \int_{a_k}^{b_k} e^{iu_k x_k} dx_k = \prod_{1 \le k \le d, \, u_k \ne 0} \frac{e^{iu_k b_k} - e^{iu_k a_k}}{iu_k} \times \prod_{1 \le k \le d, \, u_k = 0} (b_k - a_k).$$

D'où $|\widehat{\mu}(u)| \le \prod_{k=1}^{d} \left(\frac{2}{|u_k|} \wedge |a_k - b_k|\right) \underset{\|u\| \to +\infty}{\longrightarrow} 0$.

Le résultat s'étend aux combinaisons linéaires de telles indicatrices. Or l'ensemble de ces fonctions (en "escalier") est dense dans $\mathcal{L}^1\left(\mathbb{R}^d, \lambda_d\right)$.
Soit donc $f_n \xrightarrow{\mathcal{L}^1} f$, $|\widehat{\mu}(u)| \le |\widehat{\mu}_n(u)| + \|f - f_n\|$ donc $\overline{\lim}_{\|u\| \to +\infty} |\widehat{\mu}(u)| \le \|f - f_n\|$ pour tout n, et partant $\overline{\lim}_{\|u\| \to +\infty} |\widehat{\mu}(u)| = 0$. #

La propriété P2 s'étend immédiatement aux mesures finies. La propriété P3, connue sous le nom de *théorème de Riemann-Lebesgue*, s'étend, elle, comme le montre la démonstration à toute fonction f λ_d-intégrable. Malheureusement la réciproque de ce résultat est fausse. Ce qui en fait partiellement office est le *théorème d'inversion* de la transformée de Fourier pour la démonstration duquel nous renvoyons à la section 3.

<u>Théorème</u> (inversion) :
Si $\widehat{\mu} \in L^1(\lambda_d)$, alors μ a une densité *continue* g par rapport à la mesure de Lebesgue λ_d donnée par :

$$\forall x \in \mathbb{R}^d, \quad g(x) := \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} e^{i(u|x)} \widehat{\mu}(-u)\, du = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} e^{-i(u|x)} \widehat{\mu}(u)\, du.$$

La fonction caractéristique tire son nom (et son utilité) du théorème de caractérisation suivant (démontré dans la section 3) :

<u>Théorème</u> (injectivité de la transformée de Fourier, caractérisation) :
(a) Soient μ_1, μ_2 deux probabilités sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$.

$$\widehat{\mu}_1 = \widehat{\mu}_2 \Longrightarrow \mu_1 = \mu_2.$$

(b) Soient X et Y deux vecteurs aléatoires d-dimensionnels,

$$\Phi_X = \Phi_Y \iff X \overset{\mathcal{L}}{\sim} Y.$$

<u>Application importante</u> : La v.a. X est symétrique (*i.e.* $X \overset{\mathcal{L}}{=} -X$ ou encore la loi $\mathbb{P}_X$ de X est invariante par $x \mapsto -x$) si et seulement si Φ_X est à valeurs réelles. En effet :

$$\overline{\Phi_X(u)} = \overline{\mathbb{E}(e^{iuX})} = \mathbb{E}\left(e^{-iuX}\right) = \overline{\mathbb{E}(e^{-iuX})} = \Phi_{-X}(u).$$

<u>Remarque</u> : On peut avoir l'égalité $\widehat{\mu}_1 = \widehat{\mu}_2$ sur un intervalle $[-a, a]$ sans que $\mu_1 = \mu_2$ (cf. exercices).

7.2 Fonctions caractéristiques des lois usuelles

Dans la suite, Φ désigne la fonction caractéristique d'une v.a. X ayant la loi étudiée.

7.2.1 Loi binomiale $\mathcal{B}(n; p)$

$$\Phi(u) = \sum_{k=0}^{n} e^{iuk} C_n^k p^k (1-p)^{n-k} = \left(1 - p + pe^{iu}\right)^n.$$

7.2.2 Loi géométrique $\mathcal{G}(p)$

$$\Phi(u) = \sum_{k \geq 1} e^{iuk} (1-p)^{k-1} p = \frac{pe^{iu}}{1 - (1-p)e^{iu}}.$$

7.2.3 Loi de Poisson $\mathcal{P}(\lambda)$

$$\Phi(u) = \sum_{k \geq 0} e^{-\lambda} \frac{\lambda^k e^{iuk}}{k!} = e^{\lambda\left(e^{iu} - 1\right)}.$$

7.2.4 Loi uniforme $\mathcal{U}([a, b])$

$$\begin{aligned} \Phi(u) &= \int_a^b e^{iux} \frac{dx}{b-a} = \frac{e^{iub} - e^{iua}}{i(b-a)u}. \\ \text{Si } a = -b < 0, \quad \Phi(u) &= \frac{\sin(ub)}{ub}. \end{aligned}$$

7.2.5 Loi gamma $\gamma(\alpha, \beta)$, $\alpha, \beta > 0$

$$\Phi(u) = \int_0^{+\infty} e^{iux} x^{\alpha-1} e^{-\frac{x}{\beta}} \frac{dx}{\beta^\alpha \Gamma(\alpha)} = \int_0^{+\infty} e^{\left(iu - \frac{1}{\beta}\right)x} x^{\alpha-1} \frac{dx}{\beta^\alpha \Gamma(\alpha)}.$$

Vérifions les hypothèses du théorème de dérivation sous le signe $\int$ (cf. section 5) :

$$\forall x \in \mathbb{R}_+^*, \quad \frac{\partial}{\partial u}\left(e^{\left(iu - \frac{1}{\beta}\right)x} x^{\alpha-1}\right) = ixe^{\left(iu - \frac{1}{\beta}\right)x} x^{\alpha-1}.$$

$$\forall u \in \mathbb{R}, \quad \left|\frac{\partial}{\partial u}\left(e^{\left(iu - \frac{1}{\beta}\right)x} x^{\alpha-1}\right)\right| \leq e^{-\frac{1}{\beta}x} x^\alpha \mathbf{1}_{\mathbb{R}_+^*}(x) \in \mathcal{L}^1(dx).$$

Donc $\Phi'(u) = i \int_0^{+\infty} e^{\left(iu - \frac{1}{\beta}\right)x} x^\alpha \frac{dx}{\beta^\alpha \Gamma(\alpha)}$. On intègre par parties :

$$\begin{cases} x^\alpha & \overset{'}{\longrightarrow} \alpha x^{\alpha-1} \\ e^{\left(iu - \frac{1}{\beta}\right)x} & \overset{\int}{\longrightarrow} \dfrac{e^{\left(iu - \frac{1}{\beta}\right)x}}{iu - \frac{1}{\beta}} \end{cases}$$

$$\Phi'(u) = -\frac{i\alpha}{iu - \frac{1}{\beta}} \Phi(u) = -\alpha \frac{-i\beta}{1 - iu\beta} \Phi(u).$$

Finalement $\Phi(0) = 1 \Rightarrow \Phi(u) = (1 - iu\beta)^{-\alpha} = e^{-\alpha \ln(1 - iu\beta)}$ où $u \mapsto \ln(1 - iu\beta)$ est la détermination continue du logarithme de $u \mapsto 1 - iu\beta$ valant 0 en $u = 0$ (*i.e.* $\ln(1) = 0$).

En particulier si $X \overset{\mathcal{L}}{\sim} \mathcal{E}(\lambda)$, $\Phi_X(u) = \dfrac{1}{1 - iu/\lambda} = \dfrac{\lambda}{\lambda - iu}$

et si $X \overset{\mathcal{L}}{\sim} \chi^2(n)$, $\Phi_X(u) = (1 - 2iu)^{-\frac{n}{2}}$.

7.2.6 Lois $\mathcal{N}(0;1)$, $\mathcal{N}(m;\sigma^2)$ et $\mathcal{N}(0,I_d)$

(a) Si $X \overset{\mathcal{L}}{\sim} \mathcal{N}(0;1)$, $\Phi(u) = \int e^{iux} e^{-\frac{x^2}{2}} \frac{dx}{\sqrt{2\pi}}$.

Par dérivation sous le signe somme (voir section 5 pour un rappel), il vient :

$$\begin{aligned}\Phi'(u) &= i\int_{-\infty}^{+\infty} e^{iux} e^{-\frac{x^2}{2}} x \frac{dx}{\sqrt{2\pi}}. \quad \text{Comme} \begin{cases} e^{iux} & \xrightarrow{\ '\ } iue^{iux} \\ xe^{-\frac{x^2}{2}} & \xrightarrow{\int} -e^{-\frac{x^2}{2}} \end{cases} \\ \Phi'(u) &= i^2 u\int_{-\infty}^{+\infty} e^{iux} e^{-\frac{x^2}{2}} x \frac{dx}{\sqrt{2\pi}} = -u\Phi(u).\end{aligned}$$

d'où

$$\Phi(u) = \Phi(0) e^{-\frac{u^2}{2}} = e^{-\frac{u^2}{2}}.$$

(b) On peut passer à la loi $\mathcal{N}(0;I_d)$ sur $\mathbb{R}^d$ en utilisant le théorème de Fubini ; en effet

$$\Phi(u) = \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} e^{i(u|x)} e^{-\frac{1}{2}\|x\|^2} dx_1 \times \ldots \times dx_d = e^{-\frac{u_1^2}{2}} \times \ldots \times e^{-\frac{u_d^2}{2}} = e^{-\frac{\|u\|^2}{2}}.$$

Ensuite, on en déduit la fonction caractéristique de $aX+b$; $a \in \mathbb{R}$, $b \in \mathbb{R}^d$:

$$\Phi_{aX+b}(u) = e^{i(u|b)} \Phi_X(au).$$

(c) D'où le résultat pour la loi $\mathcal{N}(m;\sigma^2)$ sur $\mathbb{R}$:

$$\Phi(u) = e^{ium - \sigma^2 \frac{u^2}{2}}.$$

7.2.7 Loi de Cauchy

Le calcul repose sur la formule d'inversion (cf. section 1) et sur la détermination de la transformée de Fourier d'une loi de Laplace $\nu(dx) = \frac{a}{2} e^{-a|x|} dx$, $a > 0$.

$$\begin{aligned}\Phi_\nu(u) &= \frac{a}{2}\int \exp(iux - a|x|)\, dx = \frac{a}{2}\Big(\int_{-\infty}^0 e^{(iu+a)x} dx + \int_0^{+\infty} e^{(iu-a)x} dx\Big). \\ \Phi_\nu(u) &= \frac{a}{2}\left(\frac{1}{iu+a} + \frac{1}{a-iu}\right) = \frac{a^2}{a^2+u^2}.\end{aligned}$$

$\Phi_\nu \in \mathcal{L}^1(du)$ donc (théorème d'inversion) :

$$\forall x \in \mathbb{R} \quad \frac{a}{2} e^{-a|x|} = \frac{1}{2\pi} \int e^{-iux} \frac{a^2}{a^2+u^2} du.$$

D'où finalement si $\mu(du) = \frac{a}{\pi(u^2+a^2)} du$ (*loi de Cauchy* de paramètre $a > 0$) alors :

$$\Phi(x) = e^{-a|x|}.$$

7.3 Démonstrations des théorèmes de caractérisation et d'inversion

L'idée de ces démonstrations est de régulariser les mesures concernées par convolution de façon à faire apparaître des densités de probabilités. L'ensemble de la preuve, très technique, peut être omis lors d'une première lecture.

Dans un premier temps, on introduit des fonctions connues en Analyse comme approximation de l'unité et définies par :

$$\forall \sigma > 0, \quad \varphi_\sigma(x) := \left(\sigma\sqrt{2\pi}\right)^{-d} \exp\left(-\frac{\|x\|^2}{2\sigma^2}\right) \text{ (les } \varphi_\sigma \text{ sont paires).}$$

<u>**Lemme 1**</u> **:**
(a) $\varphi_\sigma = \frac{1}{\sigma^d}\varphi_1\left(\frac{\cdot}{\sigma}\right)$ et φ_σ est une densité de probabilité (celle de σZ, $Z \overset{\mathcal{L}}{\sim} \mathcal{N}(0,I_d)$).

(b) $\varphi_1 = (2\pi)^{-\frac{d}{2}} \widehat{\mu}_0$ où $\mu_0 \overset{\mathcal{L}}{\sim} \mathcal{N}(0,I_d)$ $\left(\text{sur } \mathbb{R}^d\right)$.

(c) On pose $m_\sigma := \varphi_\sigma.\lambda_d$ (λ_d mesure de Lebesgue). Soit μ une probabilité sur $\left(\mathbb{R}^d, \mathcal{B}\left(\mathbb{R}^d\right)\right)$.

$$y \mapsto \int_{\mathbb{R}^d} \varphi_\sigma(y-x)\mu(dx) \text{ est la densité de } \mu * m_\sigma.$$

<u>Démonstration</u> : (a) Evident car φ_1 est la densité de $\mathcal{N}(0,I_d)$.

(b) Découle trivialement du 2.6.

(c) Soit $f : \mathbb{R}^d \to \mathbb{R}_+$ borélienne positive.

$$\begin{aligned}\int_{\mathbb{R}^d} f(z)\, \mu * m_\sigma(dz) &= \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} f(x+y)\,\varphi_\sigma(y)\, dy\mu(dx), \\ &= \int_{\mathbb{R}^d}\left[\int_{\mathbb{R}^d} f(x+y)\,\varphi_\sigma(y)\, dy\right]\mu(dx) \text{ (Fubini)}, \\ &= \int_{\mathbb{R}^d}\left[\int_{\mathbb{R}^d} f(z)\,\varphi_\sigma(z-x)\, dz\right]\mu(dx) \text{ par changement de variable,} \\ &= \int_{\mathbb{R}^d} f(z)\left[\int_{\mathbb{R}^d} \varphi_\sigma(z-x)\,\mu(dx)\right] dz \text{ via le théorème de Fubini. } \#\end{aligned}$$

Nous allons maintenant exprimer cette densité à l'aide de $\widehat{\mu}$.

<u>**Lemme 2**</u> **:** $\forall y \in \mathbb{R}^d$, $\int_{\mathbb{R}^d} \varphi_\sigma(y-x)\,\mu(dx) = (2\pi)^{-d/2} \int_{\mathbb{R}^d} e^{-i(u|y)} \widehat{\mu}(u)\, \varphi_1(\sigma u)\, du$.

<u>Démonstration</u> : D'après les assertions (a) et (b) du lemme 1 et la parité de φ_σ,

$$\int_{\mathbb{R}^d} \varphi_\sigma(y-x)\mu(dx) = \int_{\mathbb{R}^d} \varphi_\sigma(x-y)\mu(dx) = \left(\sigma\sqrt{2\pi}\right)^{-d} \int_{\mathbb{R}^d} \widehat{\mu}_0\left(\frac{x-y}{\sigma}\right)\mu(dx).$$

Afin d'intervertir les signes d'intégrales, on vérifie que

$$\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\left|e^{i\left(\frac{x-y}{\sigma}|u\right)}e^{-\frac{\|u\|^2}{2}}\right|du\mu(dx)\le\mu\left(\mathbb{R}^d\right)\int_{\mathbb{R}^d}e^{-\frac{\|u\|^2}{2}}du<+\infty.$$

D'où

$$\begin{aligned}\int_{\mathbb{R}^d}\varphi_\sigma(y-x)\,\mu(dx) &= \left(\sigma\sqrt{2\pi}\right)^{-d}\int_{\mathbb{R}^d}\left[\int_{\mathbb{R}^d}e^{i\left(x|\frac{u}{\sigma}\right)}\mu(dx)\right]e^{-i\left(y|\frac{u}{\sigma}\right)}e^{-\frac{\|u\|^2}{2}}\frac{du}{(2\pi)^{d/2}},\\ &= \left(\sigma\sqrt{2\pi}\right)^{-d}\int_{\mathbb{R}^d}\widehat{\mu}\left(\frac{u}{\sigma}\right)\varphi_1(u)\,e^{-i\left(y|\frac{u}{\sigma}\right)}du.\end{aligned}$$

On pose $u:=\sigma v$ sur $\mathbb{R}^d$, d'où :

$$\int_{\mathbb{R}^d}\varphi_\sigma(y-x)\,\mu(dx)=\left(\sigma\sqrt{2\pi}\right)^{-d}\int_{\mathbb{R}^d}e^{-i(y|v)}\widehat{\mu}(v)\,\varphi_1(\sigma v)\,\sigma^d dv.\quad\#$$

<u>Démonstration du théorème de caractérisation</u> : Si $\widehat{\mu}_1=\widehat{\mu}_2$, il découle de l'identité du lemme 2 ci-avant que :

$$\forall y\in\mathbb{R}^d,\ \forall\sigma>0,\ \int_{\mathbb{R}^d}\varphi_\sigma(y-x)\,\mu_1(dx)=\int_{\mathbb{R}^d}\varphi_\sigma(y-x)\,\mu_2(dx).$$

En conséquence, pour tout $f:\mathbb{R}^d\to\mathbb{R}^+$, continue bornée :

$$(*)\int_{\mathbb{R}^d}f(y)\left[\int_{\mathbb{R}^d}\varphi_\sigma(y-x)\,\mu_1(dx)\right]dy=\int_{\mathbb{R}^d}f(y)\left[\int_{\mathbb{R}^d}\varphi_\sigma(y-x)\,\mu_2(dx)\right]dy.$$

Or, *via* le théorème de Fubini-Tonelli et deux changements de variables :

$$\begin{aligned}\int_{\mathbb{R}^d}f(y)\int_{\mathbb{R}^d}\varphi_\sigma(y-x)\,\mu_1(dx)\,dy &= \int_{\mathbb{R}^d}\left[\int_{\mathbb{R}^d}f(x+y)\,\mu_1(dx)\right]\varphi_\sigma(y)\,dy\\ &= \int_{\mathbb{R}^d}\left[\int_{\mathbb{R}^d}f(x+\sigma z)\,\mu_1(dx)\right]\varphi_1(z)\,dz.\end{aligned}$$

Quand $\sigma\to0$, f étant continue, il est clair par convergence dominée que :

$$\int_{\mathbb{R}^d}\left[\int_{\mathbb{R}^d}f(x+\sigma z)\,\mu_1(dx)\right]\varphi_1(z)\,dz\overset{\sigma\to0}{\longrightarrow}\int_{\mathbb{R}^d}f(x)\,\mu_1(dx)\underbrace{\int_{\mathbb{R}^d}\varphi_1(z)\,dz}_{=1}.$$

Finalement, de cette convergence et de l'égalité $(*)$, on déduit que :

$$\forall f\in\mathcal{C}_b\left(\mathbb{R}^d,\mathbb{R}\right),\ \int fd\mu_1=\int fd\mu_2 \text{ et partant } \mu_1=\mu_2 \text{ (cf. II.5).}\quad\#$$

<u>Démonstration du théorème d'inversion</u> :

On remarque que $\varphi_1(\sigma u)\underset{\sigma\to0}{\to}\varphi_1(0)=(2\pi)^{-d/2}$ et que $|\varphi_1(\sigma.)|\le(2\pi)^{-\frac{d}{2}}$ puis, comme $\widehat{\mu}\in\mathcal{L}^1(du)$, on peut passer à la limite dans le second terme de l'égalité du lemme 2 par convergence dominée *i.e.* :

$$(2\pi)^{-d/2}\int_{\mathbb{R}^d}e^{-i(u|y)}\widehat{\mu}(u)\,\varphi_1(\sigma u)\,du\underset{\sigma\to0}{\longrightarrow}(2\pi)^{-d}\int_{\mathbb{R}^d}e^{-i(u|y)}\widehat{\mu}(u)\,du.$$

À nouveau par convergence dominée, on en déduit que, si $f\in\mathcal{C}_K\left(\mathbb{R}^d,\ \mathbb{R}\right)$,

$$\int_{\mathbb{R}^d}f(y)\left[(2\pi)^{-\frac{d}{2}}\int_{\mathbb{R}^d}e^{-i(u|y)}\widehat{\mu}(u)\,\varphi_1(\sigma u)\,du\right]dy\underset{\sigma\to0}{\longrightarrow}\int_{\mathbb{R}^d}f(y)\left[(2\pi)^{-d}\int_{\mathbb{R}^d}e^{-i(u|y)}\widehat{\mu}(u)\,du\right]dy.$$

Parallèlement, au vu de la démonstration du théorème ci-avant, il vient :

$$\int_{\mathbb{R}^d}f(y)\left[\int_{\mathbb{R}^d}\varphi_\sigma(y-x)\,\mu(dx)\right]dy\underset{\sigma\to0}{\longrightarrow}\int_{\mathbb{R}^d}fd\mu.$$

D'où :

$$\forall f\in\mathcal{C}_K\left(\mathbb{R}^d,\mathbb{R}\right),\ \int_{\mathbb{R}^d}fd\mu=\int_{\mathbb{R}^d}f(y)\underbrace{\left(\frac{1}{(2\pi)^d}\int_{\mathbb{R}^d}e^{-i(u|y)}\widehat{\mu}(u)\,du\right)}_{:=g(y)}dy.$$

g est à valeurs réelles (changement de variable $u:=-v$) et continue (convergence dominée). Soit $g:=g^+-g^-$ la décomposition canonique de g en différence de fonctions positives et $\gamma_p\in\mathcal{C}_K\left(\mathbb{R}^d,\ \mathbb{R}\right),p\ge0$, vérifiant $0\le\gamma_p\le g^-,\gamma_p\nearrow g^-$.

Il vient, pour tout $p\ge1$,

$$0\le\int_{\mathbb{R}^d}\gamma_pd\mu=\int_{\mathbb{R}^d}\underbrace{\gamma_pg^+}_{=0}d\lambda_d-\int_{\mathbb{R}^d}\gamma_pg^-d\lambda_d=-\int_{\mathbb{R}^d}\gamma_pg^-d\lambda_d\le0.$$

Donc $\int_{\mathbb{R}^d}\gamma_pd\mu=0$ et, par convergence monotone, $\int_{\mathbb{R}^d}g^-d\mu=0$ et $\int_{\mathbb{R}^d}(g^-)^2d\lambda_d=0$ donc $g^-=0$, μ-*p.s.* et λ_d-*p.p.*.

Enfin, en approchant 1 par une suite croissante de fonctions de $\mathcal{C}_K\left(\mathbb{R}^d,\mathbb{R}_+\right)$ on vérifie que $\int_{\mathbb{R}^d}gd\lambda_d=1$ *i.e.* que g est une densité de probabilité. Finalement $\mu=g.\lambda_d$ grâce au théorème de caractérisation (cf II.5) pour les mesures finies. #

7.4 Fonctions caractéristiques et indépendance

<u>Proposition :</u>
(a) Soient $X_k:\ (\Omega,\mathcal{A},\mathbb{P})\to\mathbb{R}^{d_k},1\le k\le n$. Les $X_k,1\le k\le n$, sont indépendants si et seulement si :

$$\forall(u_1,...,u_n)\in\mathbb{R}^{d_1}\times...\times\mathbb{R}^{d_n},\ \ \Phi_{(X_1,...,X_n)}(u_1,...,u_n)=\prod_{k=1}^n\Phi_{X_k}(u_k).$$

(b) Soient $X_k:(\Omega,\mathcal{A},\mathbb{P})\to\mathbb{R}^d,\ 1\le k\le n$, des vecteurs aléatoires indépendants. Alors :

$$\Phi_{X_1+...+X_n}=\Phi_{X_1}\times...\times\Phi_{X_n}.$$

<u>Démonstration</u> : (a) On vérifie immédiatement à partir de l'identité :

$$\forall\, u{:=}(u_1,\cdots,u_n) \in \mathbb{R}^{d_1}\times\cdots\times\ \mathbb{R}^{d_n},\quad e^{i(u,x)} = e^{i(u_1,x_1)}\times\cdots\times e^{i(u_n,x_n)}$$

et du théorème de Fubini que

$$\widehat{\mu_1\otimes ...\otimes \mu_n}\,(u_1,...,u_n) = \widehat{\mu}_1(u_1)\times ...\times\widehat{\mu}_n(u_n)\,.$$

Donc, d'après le théorème d'injectivité de la transformée de Fourier :

$$\Phi_{(X_1,...,X_n)}(u_1,...,u_n) = \Phi_{X_1}(u_1)...\Phi_{X_n}(u_n) \Longleftrightarrow \mathbb{P}_{(X_1,...,X_n)} = \mathbb{P}_{X_1}\otimes ...\otimes \mathbb{P}_{X_n}.$$

(b) $\Phi_{X_1+\cdots+X_n}(u) = \mathbb{E}\left(e^{i(u,X_1+...+X_n)}\right) = \mathbb{E}\left(e^{i(u,X_1)}\times ...\times e^{i(u,X_n)}\right) = \prod_{1\le k\le n}\mathbb{E}\left(e^{i(u,X_k)}\right)$ par indépendance. #

Applications au calcul de loi : Soient X et Y deux v.a. *indépendantes*. Alors :

(a) $X \overset{\mathcal{L}}{\sim} B(n_1;p)$ et $Y \overset{\mathcal{L}}{\sim} B(n_2;p) \Rightarrow X+Y \overset{\mathcal{L}}{\sim} B(n_1+n_2;p)$.

(b) $X \overset{\mathcal{L}}{\sim} \mathcal{P}(\lambda)$ et $Y \overset{\mathcal{L}}{\sim} \mathcal{P}(\mu) \Rightarrow X+Y \overset{\mathcal{L}}{\sim} \mathcal{P}(\lambda+\mu)$.

(c) $X \overset{\mathcal{L}}{\sim} \gamma(a;b)$ et $Y \overset{\mathcal{L}}{\sim} \gamma(a';b) \Rightarrow X+Y \overset{\mathcal{L}}{\sim} \gamma(a+a',b)$.

(d) $X \overset{\mathcal{L}}{\sim} \mathcal{N}(m;\sigma^2)$ et $Y \overset{\mathcal{L}}{\sim} \mathcal{N}\left(m';\sigma'^2\right) \Rightarrow \lambda X+\lambda' Y \overset{\mathcal{L}}{\sim} \mathcal{N}\left(\lambda m+\lambda' m';\lambda^2\sigma^2+\lambda'^2\sigma'^2\right)$.
où $\mathcal{N}(m;0) = \delta_m$.

(e) $X \overset{\mathcal{L}}{\sim}$ Cauchy(c) et $Y \overset{\mathcal{L}}{\sim}$ Cauchy$(c') \Rightarrow X+Y \overset{\mathcal{L}}{\sim}$ Cauchy $(c+c')$.

Des applications importantes de l'assertion (d) au cadre vectoriel seront développées dans le chapitre VII sur les vecteurs gaussiens.

7.5 Fonctions caractéristiques et moments

Rappel du théorème de dérivation sous le signe somme :

Soit I⊂ℝ un intervalle ouvert non vide, $(X,\mathcal{A},\mu)$ un espace mesuré et $f: I\times X \to$ ℝ ou ℂ une fonction vérifiant :

(*i*) $\forall u\in I,\ x\mapsto f(u,x)$ est (mesurable et) μ-intégrable,

(*ii*) $\mu(dx)$-*p.p.*, $u\to f(u,x)$ est *dérivable en tout point de* I,

(*iii*) $\mu(dx)$-*p.p.*, $\forall u\in I$, $\left|\frac{\partial f}{\partial t}(u,x)\right| \le g(x)$, $g\in\mathcal{L}^1_{\mathbb{R}_+}(X,\mathcal{A},\mu)$.

Alors : $\dfrac{d}{du}\displaystyle\int_X f(u,x)\,\mu(dx) = \int_X \frac{\partial}{\partial u} f(u,x)\,\mu(dx)$ en tout point $u\in I$.

Démonstration : Soit $\varphi_h(u,x) = \dfrac{1}{h}\left(f(u+h,x)-f(u,x)\right)$.
$\mu(dx)$-*p.p.*, $\forall\, u\in I$, $\varphi_h(u,.)\to\dfrac{\partial f}{\partial u}(u,.)$ quand $h\to 0$.
Pour h assez petit, il vient par application des accroissements finis :

$$|\varphi_h(u,x)| \le \sup_{v\in I}\left|\frac{\partial f}{\partial u}(v,x)\right| \le g(x)\quad \mu(dx)\text{-}p.p..$$

On conclut par convergence dominée. #

Proposition :
Soit $X:(\Omega,\mathcal{A},\mathbb{P})\to\mathbb{R}^d$ un vecteur aléatoire de fonction caractéristique Φ_X.
(a) $\mathbb{E}(\|X\|) < +\infty \Rightarrow \Phi_X\in\mathcal{C}^1\left(\mathbb{R}^d\right)$ et $\dfrac{\partial\Phi_X}{\partial u_k}(u) = i\mathbb{E}\left(X_k e^{i(u|X)}\right)$.
En particulier $\mathbb{E}(X) = -i\nabla\Phi_X(0)$.
(b) Si $\mathbb{E}(\|X\|^n) < +\infty$ alors $\Phi_X\in\mathcal{C}^n\left(\mathbb{R}^d,\mathbf{C}\right)$ et

$$\frac{\partial^n\Phi_X}{\partial u_1^{n_1}\dots\partial u_d^{n_d}}(u) = i^n\mathbb{E}\left(X_1^{n_1}\dots X_d^{n_d}e^{i(u|X)}\right)$$

où $n_1+\ldots+n_d = n$.
En particulier $D(X) = -\nabla^2\Phi_X(0)+\nabla\Phi_X(0)^t\,\nabla\Phi_X(0)$ (${}^t u$ = vecteur ligne).

Démonstration : C'est une application immédiate du théorème de dérivation ci-avant. #

Malheureusement, la réciproque de cette proposition est fausse en général comme le montre le contre-exemple suivant :
Contre-exemple : Soit $\mu(dx) = \dfrac{c}{x^2\ln x}\mathbf{1}_{\{|x|\ge 2\}}\,dx$ où la constante c est choisie de façon que la mesure positive μ soit une probabilité. Il est clair qu'une variable aléatoire X ayant μ pour loi n'a pas de moment d'ordre 1 puisque

$$\int_{\mathbb{R}}|x|\mu(dx) = 2c\int_2^{+\infty}\frac{dx}{x\ln x} = +\infty.$$

D'autre part

$$\hat{\mu}(u) = \int_{\mathbb{R}} e^{iux}\frac{c}{x^2\ln x}\mathbf{1}_{\{|x|\ge 2\}}\,dx = 2c\int_2^{+\infty}\frac{\cos(ux)}{x^2\ln x}\,dx$$

$$\text{D'où }\frac{\hat{\mu}(u)-1}{u} = -\frac{2c}{u}\int_2^{+\infty}\frac{1-\cos(ux)}{x^2\ln x}\,dx = -c\int_0^{+\infty}\frac{\sin^2 y}{y^2\ln(2y/u)}\mathbf{1}_{\{y\ge u\}}\,dy.$$

(on a posé $x := 2y/u$). Or

$$\left|\frac{\sin^2 y}{y^2\ln(2y/u)}\mathbf{1}_{\{y\ge u\}}\right| \le \frac{\sin^2 y}{y^2\ln(2)} \in\mathcal{L}^1(\lambda)\quad\text{et}\quad \lim_{u\to+\infty}\frac{\sin^2 y}{y^2\ln(2y/u)}\mathbf{1}_{\{y\ge u\}} = 0$$

donc, d'après le théorème de convergence dominée,

$$\lim_{u\to 0}\frac{\hat{\mu}(u)-1}{u}=0$$

i.e. $\hat{\mu}$ est dérivable en 0 de dérivée nulle. On montre de façon similaire que $\hat{\mu}$ est en fait $\mathcal{C}^1$ sur $\mathbb{R}$ avec $\hat{\mu}'(u)=-2c\int_2^{+\infty}\frac{\sin(ux)}{x^2\ln x}\,dx$.

On a cependant la proposition suivante,énoncée dans le cas réel pour simplifier, qui montre que cette réciproque est presque vraie :

Proposition $(d=1)$ **:**
Si Φ_X est $2n$ fois dérivable en 0 alors $\mathbb{E}(X^{2n})<+\infty$ (et donc X admet des moments de tous ordres jusqu'à $2n$).

Démonstration : $n=1$. D'après la formule de Taylor-Young appliquée à $\Phi_X(h)$ et $\Phi_X(-h)$,

$$\frac{\Phi_X(h)+\Phi_X(-h)-2\Phi_X(0)}{h^2}\underset{h\to 0}{\longrightarrow}\Phi_X''(0).$$

Donc

$$\begin{aligned}-\Phi_X''(0) &= \lim_{h\to 0}\mathbb{E}\left(\frac{2-\left(e^{ihX}+e^{-ihX}\right)}{h^2}\right)=2\lim_{h\to 0}\mathbb{E}\left(\frac{1-\cos(hX)}{h^2}\right)\\ &\geq 2\mathbb{E}\left(\lim_{h\to 0}\frac{1-\cos(hX)}{h^2}\right)=2\mathbb{E}\left(\frac{X^2}{2}\right)=\mathbb{E}\left(X^2\right)\quad (\textit{via}\text{ le lemme de Fatou}).\end{aligned}$$

Pour passer du rang n au rang $n+1$, on montre d'abord que $\mathbb{E}|X|^{2n}<+\infty$ (Φ_X est particulier $\mathcal{C}^{2n}$ et on peut appliquer l'hypothèse de récurrence). On a alors $\Phi_X^{(2n)}(u)=(-1)^n\mathbb{E}\left(X^{2n}e^{iuX}\right)$ et on peut appliquer la méthode développée pour $n=1$ à $\Phi_X^{(2n+2)}(0)=(\Phi_X^{(2n)})''(0)$. #

Remarque : Moralement, il y a donc correspondance entre la régularité de la fonction caractéristique et l'existence de moments d'ordre supérieur pour la v.a. X ou, ce qui revient au même, la vitesse de décroissance à l'infini de (la densité de) sa loi.

Ces résultats constituent en pratique un moyen commode pour (re-)trouver les moments d'une v.a..

Exemple : Si $X\overset{\mathcal{L}}{\sim}\mathcal{N}(0,\sigma^2)$ alors $\forall n\in\mathbb{N}$, $\mathbb{E}(X^{2n})=\frac{\sigma^{2n}}{2^n}\frac{(2n)!}{n!}$.

En effet $\Phi_X(t)=\exp\left(-\frac{\sigma^2}{2}t^2\right)=\sum_{k\geq 0}(-1)^k\frac{\sigma^{2k}}{2^k}\cdot\frac{t^{2k}}{k!}\in\mathcal{C}^\infty(\mathbb{R})$ donc :

$$\begin{aligned}\mathbb{E}(X^{2n}) &= (-1)^n\Phi_X^{(2n)}(0)\\ &= (-1)^n(-1)^n\frac{\sigma^{2n}}{2^n}\times\frac{(2n)!}{n!}=\frac{\sigma^{2n}}{2^n}\frac{(2n)!}{n!}\end{aligned}$$

au vu du développement ci-dessus.

7.6 Autres transformations.

7.6.1 Fonction génératrice

Déjà définie et étudiée au chapitre I.7, elle joue un rôle similaire à la fonction caractéristique pour les v.a. à valeurs dans $\mathbb{N}$.

7.6.2 Transformée de Laplace d'une v.a. positive

Définition :
(a) Soit μ une mesure finie sur $(\mathbb{R}_+,\mathcal{B}(\mathbb{R}_+))$. La transformée de Laplace de μ est définie sur $\mathbb{R}_+$ par :

$$\forall s\in\mathbb{R}_+,\ \ \widetilde{\mu}(s):=\int_0^{+\infty}e^{-sx}\mu(dx)$$

(b) La transformée de Laplace d'une v.a. positive X est celle de sa loi $\widetilde{\mathbb{P}}_X$, notée souvent $\widetilde{\mu}_X$ dans la suite. C'est donc la fonction

$$\widetilde{\mu}_X: s\mapsto\mathbb{E}\left(e^{-sX}\right).$$

Proposition :

$$\widetilde{\mu}_1=\widetilde{\mu}_2\Rightarrow\mu_1=\mu_2.$$

Démonstration : L'algèbre $\mathcal{E}$ de fonctions de $\overline{\mathbb{R}}_+$ dans $\mathbb{R}$ engendrée par $\{e^{-s.},\ s\geq 0\}$ sépare les points et contient les constantes. D'après le théorème de Stone-Weierstrass, elle est donc dense dans $\mathcal{C}\left(\overline{\mathbb{R}}_+,\mathbb{R}\right)$ (fonctions continues sur $\mathbb{R}_+$ ayant une limite finie en $+\infty$) pour la convergence uniforme. μ_1 et μ_2 étant finies de même masse coïncident sur $\mathcal{E}$ et partant sur $\mathcal{C}\left(\overline{\mathbb{R}}_+,\mathbb{R}\right)$. En particulier $\int f d\mu_1=\int f d\mu_2$ pour toute $f\in\mathcal{C}_K(\mathbb{R}_+,\mathbb{R})$. Donc $\mu_1=\mu_2$. #

Proposition :
(a) $\widetilde{\mu}_X$ est une fonction décroissante, $\widetilde{\mu}_X(0)=\mathbb{P}(X\in\mathbb{R}_+)$, $\lim_{s\to+\infty}\widetilde{\mu}_X(s)=\mathbb{P}(X=0)$.
(b) $\widetilde{\mu}_X\in\mathcal{C}^\infty(\mathbb{R}_+^*)$ et pour tout $s>0$, $\widetilde{\mu}_X^{(n)}(s)=(-1)^n\mathbb{E}(X^ne^{-sX})$.
(c) $\forall n\in\mathbb{N},\ \mathbb{E}(X^n)=(-1)^n\lim_{s\to 0}\widetilde{\mu}_X^{(n)}(s)\leq+\infty$

Démonstration : (a) est essentiellement évident. La limite en $+\infty$ découle du théorème de convergence dominée puisque $e^{-sX} \to \mathbf{1}_{\{X=0\}}$ quand $s \to +\infty$ en restant dominée par 1.
(b) Application du théorème de dérivation en remarquant que $x \mapsto x^n e^{-sx}$ est une fonction bornée dès que $s > 0$.

(c) D'après le théorème de convergence monotone (Beppo-Levi) $X^n e^{-sX} \uparrow X^n$ lorsque $s \downarrow 0$, donc

$$\mathbb{E}(X^n) = \lim_{s\to 0} \mathbb{E}\left(X^n e^{-sX}\right). \quad \#$$

Proposition :
Si X et Y sont indépendantes, alors $\widetilde{\mu}_{X+Y} = \widetilde{\mu}_X . \widetilde{\mu}_Y$

Démonstration :

$$\mathbb{E}\left(e^{-s(X+Y)}\right) = \mathbb{E}\left(e^{-sX}e^{-sY}\right) = \mathbb{E}\left(e^{-sX}\right)\mathbb{E}\left(e^{-sY}\right). \quad \#$$

Exemple : Transformée de Laplace d'une loi $\gamma(\alpha;\beta)$

$$\begin{aligned}\widetilde{\mu}(s) &:= \frac{1}{\beta^\alpha \Gamma(\alpha)}\int_0^{+\infty} e^{-sx}e^{-\frac{x}{\beta}}x^{\alpha-1}dx = \frac{1}{\beta^\alpha\Gamma(\alpha)}\int_0^{+\infty} e^{-\left(s+\frac{1}{\beta}\right)x}x^{\alpha-1}dx \\ &= \frac{\left(s+\frac{1}{\beta}\right)^{-\alpha}\Gamma(\alpha)}{\beta^\alpha\Gamma(\alpha)} = (\beta s+1)^{-\alpha}.\end{aligned}$$

Ceci permet de retrouver le résultat sur la somme de v.a. indépendantes de lois $\gamma(\alpha;\beta)$ et $\gamma(\alpha';\beta)$.

Remarques :
• L'usage de la transformée de Laplace de préférence à la transformée de Fourier est recommandé pour les v.a. positives pour d'évidentes raisons de simplicité (pas de nombres complexes, bonne caractérisation des moments, etc).

• On peut évidemment définir la transformée de Laplace d'un vecteur aléatoire à valeurs $\mathbb{R}_+^d$. Les propriétés espérées subsistent. Enfin, il est évidemment possible de définir la transformée de Laplace de v.a. non positives dans certains cas :

$$X \overset{\mathcal{L}}{\sim} \mathcal{N}(0;1), \quad \forall s \in \mathbb{R}, \quad \widetilde{\mu}_X(s) = e^{s^2/2}.$$

Chapitre 8

Vecteurs gaussiens

8.1 Compléments sur les vecteurs aléatoires

Soit $X : (\Omega, \mathcal{A}, \mathbb{P}) \longrightarrow \mathbb{R}^d$. On rappelle que $X_i \in \mathcal{L}^1(\mathbb{P})$, $1 \le i \le d$, si et seulement si $\|X\| \in \mathcal{L}^1_{\mathbb{R}_+}(\mathbb{P})$ et $X_i \in \mathcal{L}^2(\mathbb{P})$ si et seulement si $\|X\| \in \mathcal{L}^2_{\mathbb{R}_+}(\mathbb{P})$. On définit :

$$\mathbb{E}(X) := \begin{pmatrix} \mathbb{E}(X_1) \\ \vdots \\ \mathbb{E}(X_d) \end{pmatrix} \quad \text{si } \|X\| \in \mathcal{L}^1_{\mathbb{R}_+}(\mathbb{P}) \text{ (on dit que } X \text{ est intégrable),}$$

$$D(X) = \Big[\underbrace{\mathbb{E}\left((X_i - \mathbb{E}X_i)(X_j - \mathbb{E}X_j)\right)}_{:=\mathrm{Cov}(X_i,X_j)}\Big]_{1\le i,j\le d} \quad \text{si } \|X\| \in \mathcal{L}^2_{\mathbb{R}_+}(\mathbb{P}) \text{ (on dit que } X$$

est de carré intégrable).

Notations : • L'exposant ^{t} désigne la transposition.

• Les vecteurs standard $u \in \mathbb{R}^d$ sont en colonne, ${}^t u$ est un vecteur ligne (le produit $u^t v$ est donc une matrice $d \times d$).

Ainsi, on vérifie immédiatement que :

$$D(X) = \mathbb{E}\Big((X - \mathbb{E}X)^t (X - \mathbb{E}X)\Big) = \mathbb{E}(X^t X) - \mathbb{E}(X)^t \mathbb{E}(X).$$

• Le produit scalaire canonique sur $\mathbb{R}^d$ sera noté selon les cas $(.|.)$ ou à l'aide d'un simple ".".

Propriétés :

P1. $\forall u \in \mathbb{R}^d$, $\mathrm{Var}\left(\sum_{i=1}^d u_i X_i\right) = {}^t u D(X) u = \sum_{i,j} u_i u_j \mathrm{Cov}(X_i, X_j)$.

P2. $D(X)$ est une matrice (symétrique) positive.

P3. Si $A \in \mathcal{M}(m \times d)$ alors AX est une v.a. à valeurs dans $\mathbb{R}^m$ et

$$\mathbb{E}(AX) = A\mathbb{E}(X),\ D(AX) = AD(X)\,{}^t A.$$

Démonstration : On vérifie P1 à la main ; P2 découle trivialement de P1 ($\mathrm{Var}(Y) \ge 0$!). La linéarité de l'espérance entraîne immédiatement que $\mathbb{E}(AX) = A\,\mathbb{E}(X)$. Enfin

$$\begin{aligned} D(AX) &= \mathbb{E}\left((AX - \mathbb{E}(AX))\,{}^t(AX - \mathbb{E}(AX))\right) = \mathbb{E}\Big(A(X - \mathbb{E}(X))\,{}^t(A(X - \mathbb{E}(X)))\Big) \\ &= \mathbb{E}\left(A(X - \mathbb{E}(X))\,{}^t(X - \mathbb{E}(X))\,{}^t A\right) = AD(X)\,{}^t A. \end{aligned}$$

Remarque : Si $D(X)$ est dégénérée (*i.e.* de rang $\le d-1$), il existe $u \in \mathbb{R}^d$ tel que ${}^t u D(X) u = 0$ soit encore $\mathrm{Var}\left(\sum_{k=1}^n u_k X_k\right) = 0$, soit encore $\sum_{k=1}^n u_k X_k = \sum_{k=1}^n u_k \mathbb{E}(X_k)$ *p.s.*.
En d'autres termes :

$$D(X) \text{ dégénérée } \Leftrightarrow \exists H \text{ hyperplan tel que } \mathbb{P}(X \in H) = 1.$$

8.2 Vecteurs gaussiens

On se place sur $\left(\mathbb{R}^d, \mathcal{B}(\mathbb{R})^d\right)$ (on rappelle que $\mathcal{B}\left(\mathbb{R}^d\right) = \mathcal{B}(\mathbb{R})^{\otimes d}$). Par convention on pose $\delta_m := \mathcal{N}(m; 0)$ (δ_m masse de Dirac en m).

Définition :
$X : (\Omega, \mathcal{A}, \mathbb{P}) \longrightarrow \left(\mathbb{R}^d, \mathcal{B}\left(\mathbb{R}^d\right)\right)$ est un vecteur gaussien si et seulement si

$$\forall u \in \mathbb{R}^d,\ \sum_{i=1}^d u_i X_i \text{ suit une loi normale.}$$

Interprétation : Au vu de la définition :

$$\begin{aligned} X \text{ vecteur gaussien } &\Longleftrightarrow \forall\, \vec{u} \in \mathbb{R}^d \setminus \{0\},\ (\vec{u}\,|X) \text{ v.a. gaussienne,} \\ &\Longleftrightarrow \forall\, \vec{u} \in \mathbb{R}^d \setminus \{0\},\ \pi^{\perp}_{\mathbb{R}\vec{u}}(X) \text{ v.a. gaussienne,} \\ &\Longleftrightarrow X \text{ gaussienne } \textit{dans toutes les directions.} \end{aligned}$$

On en déduit que, lorsque u parcourt la base canonique de $\mathbb{R}^d$, chacune des composantes X_i, $1 \le i \le d$, est gaussienne. On prendra garde que *la réciproque est fausse* en général comme le montre l'exemple suivant :

Contre-exemple : Soient $Z : (\Omega, \mathcal{A}, \mathbb{P}) \longrightarrow \mathbb{R}$ et $\varepsilon : (\Omega, \mathcal{A}, \mathbb{P}) \longrightarrow \{-1, 1\}$ indépendantes, $Z \overset{\mathcal{L}}{\sim} \mathcal{N}(0;1)$ et $\varepsilon \overset{\mathcal{L}}{\sim} \mathcal{B}\left(\{\pm 1\}; \frac{1}{2}\right)$. On pose $X := (X_1, X_2)$ où $X_1 := Z$ et $X_2 := \varepsilon Z$. Alors : $X_2 \overset{\mathcal{L}}{\sim} \mathcal{N}(0;1)$ car :

$$\begin{aligned} \Phi_{\varepsilon Z}(u) &= \mathbb{E}\left(e^{iuZ}\mathbf{1}_{(\varepsilon=1)}\right) + \mathbb{E}\left(e^{-iuZ}\mathbf{1}_{\{\varepsilon=-1\}}\right) \\ &= \mathbb{P}\{\varepsilon = 1\}\,\Phi_Z(u) + \mathbb{P}(\varepsilon = 1)\,\Phi_Z(-u) = \Phi_Z(u). \qquad (\Phi_Z \text{ est paire !}) \end{aligned}$$

Cependant le vecteur X, dont les deux composantes sont normales, n'est pas gaussien car $X_1 + X_2 = Z(1 + \varepsilon)$ ne l'est pas. En effet :

$$\begin{aligned}\mathbb{P}(X_1+X_2=0) &= \mathbb{P}(Z=0 \text{ ou } \varepsilon=-1)\\ &= \underbrace{\mathbb{P}(Z=0)}_{=0}+\mathbb{P}(\varepsilon=-1)-\underbrace{\mathbb{P}(\varepsilon=-1 \text{ et } Z=0)}_{=0}\\ &= \tfrac{1}{2}.\end{aligned}$$

Or, si X_1+X_2 suivait une loi normale, cette probabilité vaudrait soit 0 (loi normale à densité) soit 1 masse de Dirac).

On remarque en outre que $\mathrm{Cov}(X_1,X_2)=\mathbb{E}(\varepsilon Z^2)=\mathbb{E}(\varepsilon)\mathbb{E}(Z^2)=0$ bien que X_1 et X_2 ne soient pas indépendantes (sinon leur somme aurait une loi normale, cf. VI.4).

Proposition : (*a*) Si les composantes d'un vecteur aléatoire X sont indépendantes et de loi normales alors X est un vecteur gaussien.

(*b*) Si X est un vecteur gaussien d-diemensionnel et $A\in\mathcal{M}(p\times d)$, alors $Y:=AX$ est un vecteur gaussien p-dimensionnel.

Démonstration : C'est une conséquence immédiate du fait que toute combinaison linéaire de v.a.r. indépendantes de loi normale est normale (cf. VI.4). #

Chaque composante X_i d'un vecteur gaussien étant gaussienne, elle est dans $\mathcal{L}^2_{\mathbb{R}}(\mathbb{P})$; X admet donc une *matrice de dispersion* $D(X)$ que l'on notera dans ce cadre $\Sigma^{(2)}\in\mathcal{S}^+(\mathbb{R}^d)$. De même, on notera $m:=\mathbb{E}(X)$.

Définition :

Un vecteur gaussien X est *non dégénéré* si

$$\forall\vec{u}\in\mathbb{R}^d,\quad \mathrm{Var}\left(\vec{u}\,.X\right)=0\Longleftrightarrow \vec{u}=0$$

La matrice $\Sigma^{(2)}$ étant symétrique positive, ceci équivaut à dire, d'après la propriété P1 ci-avant, que la matrice $\Sigma^{(2)}$ est non-dégénérée (au sens inversible).

Proposition :

Si X est un vecteur gaussien de moyenne m et de dispersion $\Sigma^{(2)}$, alors :

$$\forall u\in\mathbb{R}^d\quad \Phi_X(u)=e^{i(m|u)-\frac{1}{2}\,{}^tu\Sigma^{(2)}u}.$$

Notation : Le vecteur moyenne m et la matrice de dispersion $\Sigma^{(2)}$ caractérisent donc entièrement la loi de X. On désigne alors la loi d'un tel vecteur gaussien par $\mathcal{N}\left(m;\Sigma^{(2)}\right)$.

Démonstration : On sait que si $Z\overset{\mathcal{L}}{\sim}\mathcal{N}(\mu;\sigma^2)$ (v.a.r.) alors $\Phi_Z(s)=e^{i\mu s-\frac{\sigma^2}{2}s^2}$; or $\Phi_X(u)=\mathbb{E}\left(e^{i(u|X)}\right)=\Phi_{(u|X)}(1)$. Par hypothèse $(u|X)=\sum_{k=1}^{d}u_kX_k$ est gaussienne d'où nécessairement, $(u|X)\overset{\mathcal{L}}{\sim}\mathcal{N}\left((u|\mathbb{E}X),\ {}^tu\Sigma^{(2)}u\right)$. Donc $\Phi_X(u)=e^{i(u|m)-\frac{1}{2}\,{}^tu\Sigma^{(2)}u}$. #

• Application à la caractérisation de l'indépendance des composantes d'un vecteur gaussien.

Soit X un vecteur gaussien de loi $\mathcal{N}\left(m,\Sigma^{(2)}\right)$. On divise X en p paquets $X=(X_1,...,X_p)$ avec X_k à valeur dans $\mathbb{R}^{d_k}$, $1\le k\le p$, $(d_1+...+d_p=d)$. Chaque X_k est un vecteur gaussien de moyenne m_k et de dispersion $\Sigma^{(2)}_k, 1\le k\le p$.

Proposition :

Les $X_1,...,X_p$ sont indépendantes si et seulement si $\Sigma^{(2)}$ est de la forme :

$$\Sigma^{(2)}=\begin{bmatrix}\Sigma^{(2)}_1 & 0 & 0\\ 0 & \ddots & 0\\ 0 & 0 & \Sigma^{(2)}_p\end{bmatrix}\qquad (p \text{ blocs diagonaux}).$$

Démonstration : $\Longleftarrow$: Si $\Sigma^{(2)}$ est de la forme ci-dessus, on vérifie immédiatement que :

$$\begin{aligned}\Phi_X(u) &= e^{i(u|m)-\frac{1}{2}\sum_{k=1}^{p}{}^tu_k\Sigma^{(2)}_ku_k}\qquad \text{où } u=(u_1,...,u_p),\, u_k\in\mathbb{R}^{d_k},\\ &=\Phi_{X_1}(u_1)...\Phi_{X_p}(u_p)\end{aligned}$$

donc (cf. proposition (a) de VI 3.) $X_1,...,X_p$ sont indépendantes.

$\Longrightarrow$: Le sens direct est évident . . . et reste vrai en dehors du cadre gaussien. #

Corollaire : En faisant $d_k=1$, $1\le k\le d=p$ dans l'énoncé ci-dessus on constate que les composantes d'un vecteur gaussien sont indépendantes si et seulement si elles sont non corrélées (*i.e.* $\Sigma^{(2)}$ diagonale).

Ce résultat tombe en défaut si les X_i sont gaussiennes sans que le vecteur $X=(X_1,\ldots,X_d)$ le soit (cf. contre-exemple ci-avant).

• Application à la simulation d'un vecteur gaussien $\mathcal{N}\left(m;\Sigma^{(2)}\right)$.

Si $d=1$, $\sigma>0$ et $X\overset{\mathcal{L}}{\sim}\mathcal{N}(m;\sigma^2)$, on peut simuler X en remarquant que $Z=\frac{X-m}{\sigma}\overset{\mathcal{L}}{\sim}\mathcal{N}(0;1)$ *i.e.* $X=m+\sigma Z, Z\overset{\mathcal{L}}{\sim}\mathcal{N}(0;1)$ (cf. IV.2.e.).

Si $d\ge 2$, on exploite la même idée. Cela passe par un petit lemme d'algèbre linéaire *autour* de la racine carrée d'une matrice symétrique positive.

Lemme :

(a) Si $\Sigma^{(2)}\in\mathcal{S}^+(d,\mathbb{R})$ (resp. inversible), $\exists\,\Sigma\in\mathcal{S}^+(d,\mathbb{R})$ (resp. inversible). tq $\Sigma^{(2)}=\Sigma^2=\Sigma\,{}^t\Sigma$.

(b) Si $rg\left(\Sigma^{(2)}\right)=p$, $\exists\,\Sigma\in\mathcal{M}(d\times p)$ tq $\Sigma^{(2)}=\Sigma\,{}^t\Sigma$.

Preuve : On diagonalise $\Sigma^{(2)}$ dans le groupe orthogonal *i.e.* $\Sigma^{(2)}=PD\,{}^tP$ où $D=Diag\,[\lambda_1...,\lambda_d]$, $\lambda_i\in\mathbb{R}_+$ et ${}^tP\,P=I_d$.

(a) On pose $\Sigma := PDiag\left[\sqrt{\lambda_1}, ..., \sqrt{\lambda_d}\right]^t P$ qui convient manifestement.

(b) Comme $rg\left(\Sigma^{(2)}\right) = p$, on peut, à une permutation des vecteurs propres près, supposer que $\lambda_{p+1} = ... = \lambda_d = 0$ et $\lambda_1, ..., \lambda_p > 0$. Soit alors :

$$\Delta_p = \underbrace{\left[\begin{array}{cccc} \sqrt{\lambda_1} & 0 & \dots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \sqrt{\lambda_p} \\ \hline 0 & \dots & \dots & 0 \\ \vdots & & & \vdots \\ 0 & \dots & \dots & 0 \end{array}\right]}_{p} \updownarrow d$$

et $P_p \in M\,(p \times d)$ constituée des p lignes supérieures de P. On pose alors $\Sigma := P\Delta_p\, {}^tP_p$.

Il est clair que $\Sigma\, {}^t\Sigma = P\Delta_p \underbrace{{}^tP_p P_p}_{=I_p} {}^t\Delta_p\, {}^tP = PDiag\left[\lambda_1, ..., \lambda_d\right]\, {}^tP = \Sigma^{(2)}$. #

Remarque : Si l'on impose dans l'assertion (a) à Σ de commuter avec $\Sigma^{(2)}$, la solution proposée dans la preuve est la seule possible : il y a unicité de la racine carrée dans $\mathcal{S}^+(d, \mathbb{R})$ (les deux matrices étant symétriques positives sont diagonalisables à valeurs propres réelles positives et comme elles commutent, elles ont une base de diagonalisation commune dans laquelle le résultat est évident).

Proposition :

Soit $\Sigma \in \mathcal{M}\,(d \times p)$ vérifiant $\Sigma^t\Sigma = \Sigma^{(2)}$ et si $Z \overset{\mathcal{L}}{\sim} \mathcal{N}\,(0; I_p)$. Alors :

$$\forall m \in \mathbb{R}^d,\ m + \Sigma Z \overset{\mathcal{L}}{\sim} \mathcal{N}\left(m; \Sigma^{(2)}\right).$$

Démonstration : $m+\Sigma Z$ est clairement un vecteur gaussien d-dimensionnel de moyenne $\mathbb{E}(m + \Sigma Z) = m + \Sigma\mathbb{E}(Z) = m$ et de dispersion $D(m + \Sigma Z) = D(\Sigma Z) = \Sigma D(Z)\, {}^t\Sigma = \Sigma I_p\, {}^t\Sigma = \Sigma\, {}^t\Sigma = \Sigma^{(2)}$ ce qui caractérise entièrement la loi $\mathcal{N}\left(m; \Sigma^{(2)}\right)$ d'après la forme de la fonction caractéristique d'un vecteur gaussien. #

Remarque : Si $rg\left(\Sigma^{(2)}\right) = p \leq d - 1$, on a clairement intérêt pour économiser le générateur aléatoire à considérer une matrice Σ ayant le moins de colonnes possibles (idéalement p). On utilisera donc préférentiellement la matrice Σ proposée dans le lemme (b) plutôt que la racine carrée de $\Sigma^{(2)}$. Ceci ne pose pas de problème numérique particulier en pratique, une fois $\Sigma^{(2)}$ diagonalisée : il suffit justement de suivre la procédure du lemme (b).

Inversement, comme $\mathrm{rg}(\Sigma\, {}^t\Sigma) = \mathrm{rg}(\Sigma)$, on ne peut simuler un vecteur gaussien dont la dispersion $\Sigma^{(2)}$ est de rang p à partir d'un vecteur gaussien $\mathcal{N}(0, I_r)$, $r < p$.

• Densité de la loi d'un vecteur gaussien $\mathcal{N}\left(m; \Sigma^{(2)}\right)$ non dégénéré.

Proposition :

Si $\Sigma^{(2)} \in \mathcal{S}^+(d, \mathbb{R})$ est *inversible*, $m \in \mathbb{R}^d$ et $X \overset{\mathcal{L}}{\sim} \mathcal{N}\left(m; \Sigma^{(2)}\right)$, alors :

$$\mathbb{P}_X(dx) = \frac{1}{(2\pi)^{d/2}} \frac{1}{\sqrt{\text{dét}\Sigma^{(2)}}} \exp\left[-\tfrac{1}{2}{}^t(x - m)\left(\Sigma^{(2)}\right)^{-1}(x - m)\right] dx$$

où $dx = dx_1 \dots dx_d$ désigne la mesure de Lebesgue sur $\mathbb{R}^d$.

Démonstration : Soit $Z \overset{\mathcal{L}}{\sim} \mathcal{N}\,(0; I_d)$. Comme $\Phi_Z\,(u) = \prod_{k=1}^{d} e^{-\frac{u_k^2}{2}}$, les Z_k sont indépendantes ; d'où la densité produit donnée par $\mathbb{P}_Z(dz_1, \cdots, dz_d) = \dfrac{1}{(2\pi)^{d/2}} e^{-\frac{1}{2}\left(z_1^2 + ... + z_d^2\right)} dz_1 ... dz_d$.

Soit Σ la racine carrée de $\Sigma^{(2)}$ donnée par le lemme (a). Les vecteurs aléatoires X et $m + \Sigma Z$ ont même loi car même fonction caractéristique donc, si f est borélienne positive,

$$\mathbb{E}(f\,(X)) = \mathbb{E}(f\,(m + \Sigma Z)) = \int f\,(m + \Sigma z)\, \frac{1}{(2\pi)^{d/2}} e^{-\frac{1}{2}\left(z_1^2 + ... + z_d^2\right)} dz_1 ... dz_d$$

On pose $z := \Sigma^{-1}\,(x - m) = \varphi\,(x)$, φ automorphisme de $\mathbb{R}^d$ et $J_\varphi(x) = dét\Sigma^{-1} \neq 0$.
D'où, par changement de variable,

$$\begin{aligned} \mathbb{E}\,(f\,(X)) &= \int f\,(x)\, \frac{1}{(2\pi)^{d/2}} e^{-\frac{1}{2}\,{}^t(x-m)\,{}^t\Sigma^{-1}\Sigma^{-1}(x-m)} \left|dét\Sigma^{-1}\right| dx_1 ... dx_d \\ &= \int f\,(x)\, e^{-\frac{1}{2}\,{}^t(x-m)\left(\Sigma^{(2)}\right)^{-1}(x-m)} \frac{1}{(2\pi)^{d/2}\sqrt{dét\Sigma^{(2)}}}\, dx_1 ... dx_d. \quad \# \end{aligned}$$

Chapitre 9

Convergence de variables aléatoires

Tout comme dans le cadre déterministe, il est souvent important d'étudier des comportements asymptotiques liés à des phénomènes aléatoires. A cette fin, nous allons donner un sens à la notion de **convergence d'une suite de v.a.r.**. Malheureusement, à la différence des suites déterministes, il sera nécessaire, pour des raisons techniques, de définir plusieurs modes de convergence.

Auparavant, une section sera consacrée au lemme de Borel-Cantelli, outil important dans l'étude des théorèmes-limites.

9.1 Lemme de Borel-Cantelli, exemples d'utilisation

Proposition ("lemme de Borel-Cantelli") :

Soit $(A_n)_{n\in\mathbb{N}}$, une suite d'événements d'un espace $(\Omega, \mathcal{A}, \mathbb{P})$. On rappelle que :

$$\overline{\lim_n}\, A_n = \bigcap_n \bigcup_{k\geq n} A_k \text{ (et } \underline{\lim_n}\, A_n = \bigcup_n \bigcap_{k\geq n} A_k).$$

(a) $\sum_n \mathbb{P}(A_n) < +\infty \Rightarrow \mathbb{P}\left(\overline{\lim_n}\, A_n\right) = 0$ $\left[\text{On utilise souvent } \mathbb{P}\left(\underline{\lim_n}\ {}^c A_n\right) = 1\right]$.

(b) Si les A_n sont indépendants et si $\sum_n \mathbb{P}(A_n) = +\infty$ alors $\mathbb{P}\left(\overline{\lim_n}\, A_n\right) = 1$.

Démonstration :

(a) $\sum_n \mathbb{P}(A_n) = \sum_n \mathbb{E}(\mathbf{1}_{A_n}) = \mathbb{E}\left(\sum_n \mathbf{1}_{A_n}\right) < +\infty$ (convergence monotone). Une v.a. intégrable étant $p.s.$ finie, $\sum_n \mathbf{1}_{A_n} < +\infty$ $\mathbb{P}$-$p.s.$ i.e. :

$$\mathbb{P}(\{\omega / \omega \in A_n \text{ pour un nombre infini de } n \geq 1\}) = 0$$

i.e. $\mathbb{P}\left(\overline{\lim_n} A_n\right) = 0$.

(b) On note d'abord que ${}^c\left(\overline{\lim_n} A_n\right) = \bigcup_n^{\uparrow} \bigcap_{k\geq n} {}^c A_k$.

Par suite $\mathbb{P}\left(\overline{\lim_n} A_n\right) = 1 - \lim_n \mathbb{P}\left(\bigcap_{k\geq n} {}^c A_k\right)$. D'autre part

$\mathbb{P}\left(\bigcap_{k\geq n} {}^c A_k\right) = \lim_m \downarrow \mathbb{P}\left(\bigcap_{k=n}^m {}^c A_k\right)$. Or

$$\forall\, m \geq n, \quad \mathbb{P}\left(\bigcap_{k=n}^m {}^c A_k\right) = \prod_{k=n}^m (1 - \mathbb{P}(A_k)) \leq e^{-\sum_{k=n}^m \mathbb{P}(A_k)}$$

puisque les ${}^c A_k$, $n \leq k \leq m$, sont indépendants et que $e^{-u} \geq 1 - u$.

Finalement pour tout $n \geq 0$, $\mathbb{P}\left(\bigcap_{k\geq n} {}^c A_k\right) \leq \lim_m \exp\left(-\sum_{k=n}^m \mathbb{P}(A_k)\right) = 0$. #

Exemples d'application :

(1) Convergence d'une série de v.a. positives : Soit $(X_n)_{n\geq 1}$ une suite de v.a. positives $p.s.$ finies. Il existe une suite $c_n > 0, n \geq 1$, de réels t.q. $\sum_n c_n X_n < +\infty$, $p.s.$

Démonstration : Comme $X_n < +\infty$ $\mathbb{P}$-$p.s.$, $\lim_{x\to+\infty} \mathbb{P}(X_n > x) = 0$, donc il existe $\alpha_n > 0$ t.q. $\mathbb{P}(X_n > \alpha_n) \leq \dfrac{1}{n^2}$. D'après le lemme de B.C., $\mathbb{P}\left(\overline{\lim_n}\{X_n > \alpha_n\}\right) = 0$ et, par passage au complémentaire, $\mathbb{P}\left(\bigcup_{n\geq 1}\bigcap_{k\geq n} \{X_n \leq \alpha_n\}\right) = 1$ i.e.

$$\mathbb{P}(d\omega)\text{-}p.s.\ \exists n(\omega) \text{ t.q. } \forall n \geq n(\omega),\ X_n(\omega) \leq \alpha_n.$$

Mais alors, pour de tels $\omega \in \Omega$, $\sum_{n\geq n(\omega)} c_n X_n(\omega) \leq \sum_{n\geq n(\omega)} c_n \alpha_n$. Choisir $c_n = \dfrac{1}{n^2 \alpha_n}$ assure alors la convergence $p.s.$ de la série. #

(2) Le théorème du singe dactylographe : Soit E un ensemble fini (l'alphabet) de cardinal $|E|$. On suppose qu'un singe frappe au hasard sur un clavier figurant cet alphabet et ce de façon indéfinie. Cela revient à considérer une suite $X_n, n \geq 1$, de v.a. indépendantes à valeurs dans E de même loi, uniforme sur E : $\mathbb{P}(X_1 = e) = \frac{1}{|E|}$, $e \in E$. La question de la construction d'une telle suite sera évoquée à la section 3. On considère un texte $T := (e_1, \cdots, e_N)$ de longueur N défini comme un élément de $E^N, N \geq 1$. Le lemme de B.C. permet de montrer que le singe dactylographe va taper ce texte une infinité de fois. En effet soit $A_k := \left\{X_{(k-1)N+1} = e_1, ..., X_{kN} = e_N\right\}$, $k \geq 1$; les $A_k, k \geq 1$, sont des événements indépendants, de même probabilité $\mathbb{P}(A_k) = \frac{1}{|E|^N} > 0$.

Donc $\sum_k \mathbb{P}(A_k) = +\infty$ et partant $\mathbb{P}\left(\overline{\lim_n} A_n\right) = 1$, i.e.

$$\mathbb{P}\left(\left\{\omega / \text{card}\left\{k \,/\, \left(X_{(k-1)N+1}, ..., X_{kN}\right) = T\right\} = \infty\right\}\right) = 1.$$

9.2 Divers modes de convergence

Soit X_n, $n \geq 1$, une suite de v.a. à valeurs dans $\mathbb{R}^d$ définies sur un espace $(\Omega, \mathcal{A}, \mathbb{P})$ et X une v.a. à valeurs dans $\mathbb{R}^d$ définie sur le même espace.

Définitions :

(a) $X_n \xrightarrow{\mathbb{P}\text{-}p.s.} X$ ($\mathbb{P}$-presque sûrement) si et seulement si $\exists N, \mathbb{P}$-négligeable (i.e. $N \subset A$, $A \in \mathcal{A}, \mathbb{P}(A) = 0$) tel que $\forall \omega \notin N, X_n(\omega) \to X(\omega)$.

(b) Si $X_n \in L^p(\Omega, \mathcal{A}, \mathbb{P})$ et $X \in L^p(\Omega, \mathcal{A}, \mathbb{P})$,

$$X_n \xrightarrow{L^p} X \text{ si } X_n \to X \text{ dans } L^p(\Omega, \mathcal{A}, \mathbb{P}) \text{ i.e. si } \mathbb{E}(\|X_n - X\|^p) \to 0$$

Si $p = 2$ on parle de convergence en moyenne quadratique.

(c) $X_n \xrightarrow{P} X$ (en probabilité) si et seulement si :

$$\forall \varepsilon > 0, \lim_n \mathbb{P}(\|X_n - X\| > \varepsilon) \to 0.$$

Commentaire : La convergence $\mathbb{P}$-*p.s.* est le mode le plus intuitif, c'est aussi le plus délicat à obtenir. Comme nous allons le voir, le mode le plus faible est la convergence en probabilité. Pour l'instant c'est le seul dont nous ayons un exemple intéressant : la loi faible des grands nombres (cf. chap. IV.5), qui montre que si $(X_n)_{n\geq 1}$ est une suite de v.a. L^2, 2 à 2 non corrélées et de même loi, alors $\dfrac{X_1 + ... + X_n}{n} \xrightarrow{P} \mathbb{E}(X_1)$. Nous allons voir à la section suivante que ce résultat peut être sensiblement amélioré.

Les liens entre ces différents modes de convergence sont explicités dans le théorème suivant sous forme de tableau :

Théorème :

Démonstration :

- $\mathbb{P}$-p.s. $\to$ P : Soit $\varepsilon > 0$ fixé.

$$\mathbb{P}(\|X_n - X\| > \varepsilon) = \int \mathbf{1}_{\{\|X_n - X\| > \varepsilon\}} d\mathbb{P} \to 0 \text{ quand } n \to +\infty$$

(convergence dominée).

- $L^1 \Longrightarrow$ P : $\mathbb{P}(\|X_n - X\| > \varepsilon) \leq \dfrac{\mathbb{E}(\|X_n - X\|)}{\varepsilon} \to 0$ quand $n \to +\infty$.
- $L^q \Longrightarrow L^p$: $\mathbb{E}(\|X_n - X\|^p) \leq \mathbb{E}(\|X_n - X\|^q)^{\frac{p}{q}}$ (inégalité de Hölder), $q \geq p \geq 1$.
- $L^\infty \Longrightarrow \mathbb{P}$-*p.s.* et L^q : évident car, d'une part, $\mathbb{P}(d\omega)$-*p.s.* $\|X_n(\omega) - X(\omega)\| \leq \|X_n - X\|_{L^\infty(\mathbb{P})}$ et d'autre part $\mathbb{E}(\|X_n - X\|^q) \leq \|X_n - X\|^q_{L^\infty(\mathbb{P})}$.
- P $\Rightarrow$ (suite extraite converge *p.s.*) :

Pour tout $k \geq 1$, $\mathbb{P}\left(\|X_n - X\| > \dfrac{1}{2^k}\right) \underset{n\to+\infty}{\longrightarrow} 0$. Donc, en particulier,

$$\forall k \geq 1, \exists n_k \;/\; \forall n \geq n_k, \mathbb{P}\left(\|X_{n_k} - X\| > \frac{1}{2^k}\right) \leq \frac{1}{2^k}.$$

De là, on construit de proche en proche une suite extraite $\varphi(k)$, $k \geq 1$, vérifiant :

$$\sum_k \mathbb{P}\left(\left\|X_{\varphi(k)} - X\right\| > \frac{1}{2^k}\right) \leq \sum_{k\geq 1} \frac{1}{2^k} = 1 < +\infty.$$

Donc, d'après le lemme de Borel-Cantelli,

$$\mathbb{P}\left(\bigcap_n \bigcup_{k\geq n} \left\{\left\|X_{\varphi(k)} - X\right\| > \frac{1}{2^k}\right\}\right) = 0,$$

soit encore $\mathbb{P}\left(\bigcup_n \bigcap_{k\geq n} \left\{\left\|X_{\varphi(k)} - X\right\| \leq \frac{1}{2^k}\right\}\right) = 1$. Or il est immédiat que :

$$\begin{aligned}\bigcup_n \bigcap_{k\geq n} \left\{\left\|X_{\varphi(k)} - X\right\| \leq \frac{1}{2^k}\right\} &= \left\{\omega \;/\; \exists n(\omega) \text{ t.q. } \forall n \geq n(\omega), \left\|X_{\varphi(n)}(\omega) - X(\omega)\right\| \leq \frac{1}{2^n}\right\} \\ &\subset \left\{\omega \;/\; X_{\varphi(n)}(\omega) \to X(\omega)\right\}.\end{aligned}$$

- P + (domination L^p) $\Longrightarrow L^p$ *i.e.* $\left.\begin{array}{c}\mathbb{P}\text{-}ps,\ \|X_n\| \leq Y \in L^p_{\mathbb{R}^+}(\mathbb{P}) \\ X_n \xrightarrow{P} X\end{array}\right\} \Longrightarrow X_n \xrightarrow{L^p} X$. Au vu du point précédent, il existe une suite extraite $(X_{\varphi(n)})_{n\geq 0}$ convergeant $\mathbb{P}$-*ps* vers X. Par suite, $\|X\| \leq Y$ $\mathbb{P}$-*ps*. D'où, il vient, en s'appuyant sur l'inégalité entre réels positifs $(u+v)^p \leq 2^{p-1}(u^p + v^p)$,

$$\begin{aligned}\|X_n - X\|^p &\leq \varepsilon^p \mathbf{1}_{\{\|X_n - X\| \leq \varepsilon\}} + \|X_n - X\|^p \mathbf{1}_{\{\|X_n - X\| > \varepsilon\}} \\ &\leq \varepsilon^p + 2^{p-1} Y^p \mathbf{1}_{\{\|X_n - X\| > \varepsilon\}} \ \mathbb{P}\text{-}_{ps}.\end{aligned}$$

Pour tout réel $A > 0$, on obtient alors la majoration

$$\|X_n - X\|^p \leq \varepsilon^p + 2^{p-1} Y^p \mathbf{1}_{\{Y^p \geq A\}} + 2^{p-1} A \mathbf{1}_{\{\|X_n - X\| > \varepsilon\}}$$

D'où, en intégrant,

$$\mathbb{E}\left(\|X_n - X\|^p\right) \leq \varepsilon^p + 2^{p-1}\mathbb{E}\left(Y^p \mathbf{1}_{\{Y^p \geq A\}}\right) + 2^{p-1} A \mathbb{P}\left(\|X_n - X\| > \varepsilon\right).$$

Partant

$$\overline{\lim_n}\, \mathbb{E}\left(\|X_n - X\|^p\right) \leq \varepsilon^p + 2^{p-1}\mathbb{E}\left(Y^p \mathbf{1}_{\{Y^p \geq A\}}\right)$$

puisque $\mathbb{P}\left(\|X_n - X\| > \varepsilon\right) \overset{n\to+\infty}{\to} 0$. On conclut en faisant tendre $\varepsilon \to 0$ et $A \to +\infty$. #

Remarque : • L'avant-dernier point assure l'unicité de la limite en probabilité.

• On peut affiner le lien entre convergence en probabilités et en norme L^1 en montrant que

$$X_n \xrightarrow{\mathbb{P}} X \quad \text{si et seulement si} \quad \mathbb{E}\left(\|X_n - X\| \wedge 1\right) \overset{n\to+\infty}{\longrightarrow} 0.$$

En outre, $\delta(X, Y) := \mathbb{E}\left(\|X - Y\| \wedge 1\right)$ est une distance sur l'ensemble des v.a. définies sur l'espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$ à valeurs dans $\mathbb{R}^d$. Cet ensemble est parfois noté $L^0_{\mathbb{R}^d}(\mathbb{P})$. On montre sans difficulté (exercice !) que $(L^0_{\mathbb{R}^d}(\mathbb{P}), \delta)$ est un espace métrique complet.

Contre-exemples : On se place sur l'espace probabilisé $([0,1], \mathcal{B}([0,1]), dx)$

(1) Soit $X_n(\omega) = n^\alpha \mathbf{1}_{[0,\frac{1}{n}]}(\omega)\,(\alpha \geq 0)$. $\forall \alpha \geq 0$, $X_n \xrightarrow{\mathrm{P}} 0$ (et même *p.s.*) et $X_n \xrightarrow{L^p} 0$ si et seulement si $\mathbb{E}(X_n^p) = n^{\alpha p - 1} \to 0$ *i.e.* $\alpha < \dfrac{1}{p}$.

(2) Soit $\varphi = (\varphi_1, \varphi_2) : \mathbb{N} \to \{(n_1, n_2) \in \mathbb{N}^2 \,/\, n_2 \leq n_1\}$ la bijection définie par le diagramme ci-dessous.

On pose $Y_{(k,\ell)} := \sqrt{k}\,\mathbf{1}_{\left[\frac{\ell}{k}, \frac{\ell+1}{k}\right]}$ et $X_n := Y_{\varphi(n)}$ (noter que $Y_{(k,\ell)} \equiv 0$ si $\ell > k$). On vérifie que $\lim_n \varphi_1(n) = +\infty$. Or $\mathbb{E}(Y_{k,\ell}) = \dfrac{\sqrt{k}}{k}\mathbf{1}_{\{\ell < k\}} = \dfrac{1}{\sqrt{k}}\mathbf{1}_{\{\ell<k\}} \to 0$ lorsque $k \to +\infty$ donc $X_n \xrightarrow{L^1} 0$ (donc en P). En revanche si $\omega \in [0,1[$ il existe $0 \leq \ell_k \leq k-1$ t.q. $\dfrac{\ell_k}{k} \leq \omega < \dfrac{\ell_k + 1}{k}$; d'où $Y_{(k,\ell_k)}(\omega) = \sqrt{k} \to +\infty$. Finalement $X_n(\omega) \not\to 0$ car $X_{\varphi^{-1}(k,\ell_k)} \to +\infty$.

9.2.1 Critères de convergence p.s.

Critère 1 :

(a) Si, $\forall \varepsilon > 0$, $\sum_n \mathbb{P}\left(\|X_n - X\| > \varepsilon\right) < +\infty$ alors, $X_n \overset{\mathbb{P}\text{-}p.s.}{\longrightarrow} X$.

(b) Si $\sum_n \mathbb{P}\left(\|X_{n+1} - X_n\| > \varepsilon_n\right) < +\infty$ et $\sum_n \varepsilon_n < +\infty$ alors $X_n \overset{\mathbb{P}\text{-}p.s.}{\longrightarrow}$ . .

Démonstration : (a) D'après le lemme de Borel-Cantelli, pour tout $p \geq 1$,

$$\mathbb{P}\left(\bigcap_n \bigcup_{k \geq n} \left\{\|X_k - X\| > \frac{1}{p}\right\}\right) = 0.$$

Par σ-sous-additivité, il vient alors

$$\mathbb{P}\left(\bigcup_p \bigcap_n \bigcup_{k \geq n} \left\{\|X_k - X\| > \frac{1}{p}\right\}\right) = 0.$$

d'où, par simple passage au complémentaire,

$$\mathbb{P}\left(\bigcap_p \bigcup_n \bigcap_{k \geq n} \left\{\|X_k - X\| \leq \frac{1}{p}\right\}\right) = 1.$$

Cette égalité se traduit immédiatement par

$$\mathbb{P}(d\omega)\text{-}p.s.,\ \forall p \in \mathbb{N}^*,\ \exists n_p(\omega) \ \text{ tel que } \forall n \geq n_p(\omega),\ \|X_n(\omega) - X(\omega)\| \leq \frac{1}{p}$$

i.e. $X_n(\omega) \to X(\omega)\, \mathbb{P}(d\omega)$-*p.s.*.

(b) Toujours d'après le lemme de Borel-Cantelli, il vient :

$$\mathbb{P}(d\omega)\text{-}p.s.,\ \exists n(\omega) \,/\, \forall n \geq n(\omega),\ \|X_{n+1}(\omega) - X_n(\omega)\| \leq \varepsilon_n$$

d'où, $\mathbb{P}(d\omega)$-*p.s.*, $\exists n(\omega) \,/ \sum_{n \geq n(\omega)} \|X_{n+1}(\omega) - X_n(\omega)\| < +\infty$, *i.e.* $\mathbb{P}(d\omega)$-*p.s.*, la série $\sum_n \|X_{n+1}(\omega) - X_n(\omega)\|$ est absolument convergente, donc convergente.

En d'autres termes, $\mathbb{P}(d\omega)$-*p.s.*, la suite $X_n(\omega)$ est convergente. #

Remarque : Le critère 1.(a), bien que n'étant que suffisant, illustre la différence entre convergence *p.s.* et en probabilité. Lorsqu'une suite X_n, $n \geq 1$, vérifie le critère 1(a), on parle parfois de *convergence complète*.

Critère 2 :

Si, $\forall \varepsilon > 0$, $\lim_n \mathbb{P}\left(\sup_{k \geq 1} \|X_{n+k} - X_n\| > \varepsilon\right) = 0$ alors $X_n \overset{\mathbb{P}\text{-}p.s.}{\to}$ . .

Démonstration : On construit par récurrence une suite $(n_p)_{p\geq 1}$ telle que

$$\mathbb{P}\left(\sup_{k\geq 1}\left\|X_{n_p+k}-X_{n_p}\right\|>\frac{1}{p}\right)\leq\frac{1}{2^p}.$$

D'après le lemme de Borel-Cantelli,

$$\mathbb{P}(d\omega)\text{-}p.s. \qquad \lim_p \sup_{k\geq 1}\left\|X_{n_p+k}(\omega)-X_{n_p}(\omega)\right\|=0.$$

i.e. $\mathbb{P}(d\omega)$-*p.s.* $\forall\, p\in\mathbb{N}$, $\exists\, n_p(\omega)\in\mathbb{N}$ t.q. $\forall k,\ \ell\geq n^p(\omega)$,

$$\|X_\ell(\omega)-X_m(\omega)\|\leq\left\|X_\ell(\omega)-X_{n_p}(\omega)\right\|+\left\|X_{n_p}(\omega)-X_m(\omega)\right\|\leq\frac{2}{p},$$

soit encore, $\mathbb{P}(d\omega)$-*p.s.* $X_n(\omega)$ est de Cauchy et, partant, convergente. #

9.2.2 Un critère de convergence en probabilité

Nous n'allons en proposer qu'un, dont l'intérêt est essentiellement théorique.

Proposition :

$X_n \xrightarrow{P} X$ si et seulement si de toute suite extraite on peut extraire une suite $\mathbb{P}$-*p.s.* convergeant vers X.

Démonstration :

$\Longrightarrow$: Si $X_n \xrightarrow{P} X$, alors $X_{\varphi(n)} \xrightarrow{P} X$ et il existe $\varphi(\psi(n))$ extraite de φ tel que $X_{\varphi\circ\psi(n)} \xrightarrow{\mathbb{P}\text{-}p.s.} X$ (cf. théorème de la section 2).

$\Longleftarrow$: On raisonne par contraposition. Si $X_n \overset{P}{\not\to} X$, alors il existe $\varepsilon_0>0$ tel que $\mathbb{P}(\|X_n-X\|>\varepsilon_0)\not\to 0$. On extrait donc $X_{\varphi(n)}$ de X_n vérifiant pour tout $n\geq 1$,

$$\mathbb{P}\left(\left\|X_{\varphi(n)}-X\right\|>\varepsilon_0\right)>\alpha>0.$$

Il est alors clair qu'aucune suite extraite de $X_{\varphi(n)}$ ne peut converger *p.s.* vers X sinon elle convergerait en probabilité ce qui est clairement impossible. #

On en déduit immédiatement les corollaires suivants

Corollaire :

Soit $f:\mathbb{R}^d\to\mathbb{R}^d$, *continue.* Alors $X_n \xrightarrow{P} X \Rightarrow f(X_n) \xrightarrow{P} f(X)$.

Corollaire : La convergence *p.s.* n'est pas une notion métrique.

Démonstration : Sur un espace métrique une suite $(x_n)_{n\geq 0}$ converge vers une limite x_∞ si et seulement si de toute suite extraite on peut extraire une suite extraite convergeant vers x_∞. #

9.2.3 Familles uniformément intégrables

La notion d'équi-intégrabilité va permettre de généraliser la réciproque de l'implication $L^1\Rightarrow P$ dans le diagramme des modes de convergence en affaiblissant l'hypothèse de domination (cf. le second théorème du paragraphe).

Définition :

Une famille $\mathcal{F}$ de v.a. est "uniformément intégrable" ou "équi-intégrable" si : $\sup_{X\in\mathcal{F}} \int_{\{\|X\|>a\}}\|X\|\,d\mathbb{P}\to 0$ quand $a\to+\infty$.

Exemple 1 : Le cas "dominé".

S'il existe $Y\in L^1(\Omega,\mathcal{A},\mathbb{P})$ tel que $\forall X\in\mathcal{F},\ \|X\|\leq Y$, alors $\mathcal{F}$ est équi-intégrable (en particulier la famille réduite à $\{Y\}$ est équi-intégrable).

Exemple 2 : Bornitude des moments.

S'il existe $\alpha\in\,]1,+\infty[$ tel que $\sup_{X\in\mathcal{F}}\mathbb{E}\|X\|^\alpha<+\infty$, alors $\mathcal{F}$ est équi-intégrable.

Démonstration :

$\mathbb{E}\|X\|^\alpha\geq a^{\alpha-1}\int_{\|X\|>a}\|X\|\,d\mathbb{P}$ et donc $\sup_{X\in\mathcal{F}}\int_{\|X\|>a}\|X\|\,d\mathbb{P}\leq\frac{1}{a^{\alpha-1}}\sup_{X\in\mathcal{F}}\mathbb{E}\|X\|^\alpha$. #

Théorème :

$\mathcal{F}$ équi-intégrable si et seulement si :

$$\begin{cases} (i) & \sup_{X\in\mathcal{F}}\mathbb{E}\|X\|<+\infty \\ (ii) & \forall\varepsilon>0,\ \exists\eta>0,\ \text{tel que } \forall B\in\mathcal{A}, \mathbb{P}(B)\leq\eta\ \Rightarrow\ \sup_{X\in\mathcal{F}}\int_B\|X\|\,d\mathbb{P}\leq\varepsilon. \end{cases}$$

Démonstration : Pour $B\in\mathcal{A}$ et $a\in\mathbb{R}_+^*$,

$$\int_B\|X\|\,d\mathbb{P}=\int_{B\cap\{\|X\|>a\}}\|X\|\,d\mathbb{P}+\int_{B\cap\{\|X\|\leq a\}}\|X\|\,d\mathbb{P},$$

et donc :

$$\int_B\|X\|\,d\mathbb{P}\leq\int_{\{\|X\|>a\}}\|X\|\,d\mathbb{P}+a\mathbb{P}(B)$$

$\Rightarrow$: Supposons donc $\mathcal{F}$ équi-intégrable. Prenons $B=\Omega$ et a_0 tel que :

$$\sup_{X\in\mathcal{F}}\int_{\{\|X\|>a_0\}}\|X\|\,d\mathbb{P}\leq 1.$$

Alors : $\forall X\in\mathcal{F},\mathbb{E}\|X\|\leq 1+a_0$, d'où (i).

Soit $\varepsilon>0$. Choisissons $a_1>0$ tel que $\sup_{X\in\mathcal{F}}\int_{\{\|X\|>a_1\}}\|X\|\,d\mathbb{P}\leq\frac{\varepsilon}{2}$ et $\eta=\frac{\varepsilon}{2a_1}$. Si $\mathbb{P}(B)<\eta$, on a :

$$\forall X\in\mathcal{F},\int_B\|X\|\,d\mathbb{P}\leq\sup_{X\in\mathcal{F}}\int_{\{\|X\|>a_1\}}\|X\|\,d\mathbb{P}+a_1\eta\leq\varepsilon.$$

<u>$\Leftarrow$:</u> Réciproquement, supposons (i) et (ii). Soient $\varepsilon > 0$ et η associé par (ii). D'après l'inégalité de Markov, il vient

$$\forall a > 0,\ \sup_{X\in\mathcal{F}} \mathbb{P}\left(\|X\| > a\right) \leq \frac{1}{a} \sup_{X\in\mathcal{F}} \mathbb{E}\,\|X\| < +\infty \text{ d'après } (i)$$

Soit $a_0 = \frac{1}{\eta} \sup_{X\in\mathcal{F}} \mathbb{E}\,\|X\|$ et $a > a_0$

$$X \in \mathcal{F} \Rightarrow \mathbb{P}\left(\|X\| > a\right) \leq \mathbb{P}\left(\|X\| > a_0\right) \leq \frac{1}{a_0} \sup_{X\in\mathcal{F}} \mathbb{E}\,\|X\| \leq \eta.$$

D'où par (ii), $\int_{\{\|X\|>a\}} \|X\|\, d\mathbb{P} \leq \varepsilon$. En résumé

$$\forall\varepsilon > 0, \exists a_0 \text{ tel que } a > a_0 \Rightarrow \sup_{X\in\mathcal{F}} \int_{\{\|X\|>a\}} \|X\|\, d\mathbb{P} \leq \varepsilon,$$

d'où l'équi-intégrabilité de $\mathcal{F}$. #

<u>Théorème</u> (généralisant la convergence dominée de Lebesgue) :

$$\left.\begin{array}{r}(X_n)_{n\geq 1} \text{ est uniformément intégrable}\\ X_n \xrightarrow{\mathrm{P}} X\end{array}\right\} \Longrightarrow X_n \xrightarrow{L^1} X.$$

<u>Démonstration</u> : Il existe une sous-suite X_{n_k} telle que $X_{n_k} \xrightarrow{p.s.} X$ et, par Fatou,

$$\mathbb{E}\,\|X\| \leq \varliminf_{k\to+\infty} \mathbb{E}\,\|X_{n_k}\| < +\infty$$

par (i) de l'équi-intégrabilité. Donc X est intégrable.

Soit $\varepsilon > 0$. (X_n) et X étant équi-intégrables, $\exists\, \eta > 0$ tel que, pour tout $B \in \mathcal{A}$,

$$\begin{aligned}\mathbb{P}\left(B\right) \leq \eta \quad &\Rightarrow \sup_n \int_B \|X_n\|\, d\mathbb{P} \leq \frac{\varepsilon}{4} \text{ et } \int_B \|X\|\, d\mathbb{P} \leq \frac{\varepsilon}{4}\\ &\Rightarrow \sup_n \int_B \|X_n - X\|\, d\mathbb{P} \leq \frac{\varepsilon}{2}.\end{aligned}$$

Mais $\mathbb{P}\left(\|X_n - X\| > \frac{\varepsilon}{2}\right) \underset{n\to+\infty}{\rightarrow} 0$, donc

$$\exists n_0 \text{ tel que } n \geq n_0 \Rightarrow \mathbb{P}\left(\|X_n - X\| > \frac{\varepsilon}{2}\right) \leq \eta.$$

Finalement, si $n \geq n_0$,

$$\mathbb{E}\,\|X_n - X\| \leq \int_{\left\{\|X_n - X\|\leq\frac{\varepsilon}{2}\right\}} \|X_n - X\|\, d\mathbb{P} + \int_{\left\{\|X_n - X\|>\frac{\varepsilon}{2}\right\}} \|X_n - X\|\, d\mathbb{P} \leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

D'où : $\mathbb{E}\,\|X_n - X\| \underset{n\to\infty}{\rightarrow} 0$. #

9.3 Loi forte des grands nombres

9.3.1 Construction d'une suite (infinie) de variables aléatoires indépendantes

<u>Théorème</u> :
Soit $(E_n, \mathcal{E}_n, \mu_n)$, $n \geq 1$, une suite d'espaces probabilisés. Alors il existe sur $E = \prod_{n\geq 1} E_n$ muni de la tribu

$$\mathcal{E} := \bigotimes_{n\in\mathbb{N}^*} \mathcal{E}_n := \sigma\left(\prod_{n\in\mathbb{N}^*} A_n, A_n \in \mathcal{E}_n, A_n = E_n \text{ sauf un nombre fini d'indices } n\right)$$

une unique mesure de probabilité μ vérifiant :

$$(*)\, \forall A_i \in \mathcal{E}_i, A_i = E_i \text{ sauf pour un nombre fini de i, } \mu\left(\prod_{i\geq 1} A_i\right) := \prod_{i/A_i\neq E_i} \mu_i\left(A_i\right).$$

<u>Démonstration</u> : On considère $\mathcal{C} = \left\{A\times \prod_{i\geq n+1} E_i, A \in \mathcal{E}_1 \otimes \ldots \otimes \mathcal{E}_n, n \geq 1\right\}$. $\mathcal{C}$ est une algèbre de Boole et $\sigma\left(\mathcal{C}\right) = \mathcal{E}$. On définit sur $\mathcal{C}$ la mesure μ par $\mu\left(A\times \prod_{i\geq n+1} E_i\right) = \mu_1 \otimes \ldots \otimes \mu_n\left(A\right)$. On vérifie que cette définition est cohérente *i.e.* $A\times \prod_{i\geq n+1} E_i = B\times \prod_{i\geq n+2} E_i \Rightarrow \mu\left(A\times \prod_{i\geq n+1} E_i\right) = \mu\left(B\times \prod_{i\geq n+2} E_i\right)$. On vérifie ensuite que μ est additive sur deux parties disjointes de $\mathcal{C}$ et enfin que si $A_k \downarrow \phi, A_k \in \mathcal{C}, \mu\left(A_k\right) \downarrow 0$ (ce sont des petits exercices simples). Le théorème de prolongement (cf. II.3.2) assure alors que μ se prolonge en une mesure de probabilité sur tout $\mathcal{E}$. μ vérifie évidemment (*) de par sa définition même. #

<u>Application</u> : On considère $\Omega = E, \mathcal{A} = \mathcal{E}, \mathbb{P} = \mu$ et $X_k := \pi_{E_k}$ projection canonique de E sur E_k. On vérifie immédiatement que les $X_1, \ldots, X_n$ sont indépendantes pour tout n, *i.e.* les $X_n, n \geq 1$, sont indépendantes.

9.3.2 Loi forte des grands nombres : le cas L^2

Ce cadre, sans être optimal, est un bon compromis entre la difficulté de la démonstration et son champ d'application. La situation la plus générale – le cadre L^1 – sera étudié au paragraphe suivant.

<u>Théorème</u> :

Soit $X_n, n \geq 1$, une suite de v.a.r. de même loi, $X_1 \in L^2$, 2 à 2 non corrélées *i.e.* $\forall i \neq j,\ \mathrm{cov}(X_i, X_j) = 0$. Alors :

$$\overline{X}_n := \frac{X_1 + ... + X_n}{n} \underset{\mathbb{P}\text{-}p.s.}{\overset{L^2}{\longrightarrow}} \mathbb{E}(X_1) \text{ quand } n \to +\infty.$$

<u>Démonstration</u> : Partant de l'identité

$$\frac{X_1 + ... + X_n}{n} - \mathbb{E}(X_1) = \frac{(X_1 - \mathbb{E}(X_1)) + ... + (X_n - \mathbb{E}(X_n))}{n},$$

on peut supposer, quitte à remplacer les X_k par les v.a. 2 à 2 non corrélées $X_k - \mathbb{E}(X_1)$, que $\mathbb{E}(X_1) = 0$, ce que nous ferons dans la suite. Les variables X_k étant (centrées) et 2 à 2 non corrélées, il vient alors

$$\mathbb{E}\left(\overline{X}_n^2\right) = \frac{\mathrm{Var}(\overline{X}_n)}{n} = \frac{\mathbb{E}(X_1^2)}{n} \underset{n\to+\infty}{\longrightarrow} 0$$

donc $\dfrac{X_1 + \ldots + X_n}{n} \overset{L^2}{\to} 0.$

On pose $S_n := X_1 + \ldots + X_n$. On a $\mathbb{E}(S_n^2) = n\mathbb{E}(X_1^2)$ (cf. ci-avant). L'inégalité de Bienaymé-Tchebychev conduit à

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

D'où il ressort que :

$$\forall \varepsilon > 0, \quad \sum_{n\geq 1} \mathbb{P}\left(\left|\frac{S_{n^2}}{n^2}\right| > \varepsilon\right) \leq \frac{\mathbb{E}(X_1^2)}{\varepsilon^2} \sum_{n\geq 1} \frac{1}{n^2} < +\infty.$$

donc (cf. critère 1.(a), section 2.1) $\dfrac{S_{n^2}}{n^2} \overset{\mathbb{P}\text{-}p.s.}{\longrightarrow} 0$ quand $n \to +\infty$.

On pose alors $D_n := \max\limits_{n^2 \leq k < (n+1)^2} |S_k - S_{n^2}|$. Montrons que $\dfrac{D_n}{n^2} \overset{\mathbb{P}\text{-}p.s.}{\longrightarrow} 0.$

$$D_n^2 = \max_{n^2 \leq k < (n+1)^2} (S_k - S_{n^2})^2 \leq \sum_{k \in \left\{n^2, \ldots, (n+1)^2 - 1\right\}} (S_k - S_{n^2})^2.$$

Or $\mathbb{E}\left((S_k - S_{n^2})^2\right) = (k - n^2)\,\mathbb{E}(X_1^2)$ car les X_i sont 2 à 2 non corrélées. Donc :

$$\mathbb{E}\left(D_n^2\right) \leq \mathbb{E}\left(X_1^2\right) \sum_{n^2 \leq k < (n+1)^2} \left(k - n^2\right) = \mathbb{E}\left(X_1^2\right) \sum_{k=0}^{(n+1)^2 - n^2 - 1} k = \mathbb{E}\left(X_1^2\right) \frac{2n(2n+1)}{2}.$$

D'où $\sum\limits_{n\geq 1} \mathbb{P}\left(\dfrac{D_n}{n^2} > \varepsilon\right) \leq \sum\limits_{n\geq 1} \dfrac{\mathbb{E}(D_n^2)}{n^4\varepsilon^2} \leq \dfrac{\mathbb{E}(X_1^2)}{\varepsilon^2} \sum\limits_{n\geq 1} \dfrac{n(2n+1)}{n^4} < +\infty.$

Or $\left|\dfrac{S_n}{n}\right| \leq \dfrac{\left|S_{[\sqrt{n}]^2}\right| + D_{[\sqrt{n}]}}{n} \leq \left|\dfrac{S_{[\sqrt{n}]^2}}{[\sqrt{n}]^2}\right| + \dfrac{D_{[\sqrt{n}]}}{[\sqrt{n}]^2}$ où $[x]$ désigne la partie entière de x. On conclut en remarquant que $[\sqrt{n}] \to +\infty$ quand $n \to +\infty$. #

<u>Remarque</u> : Le théorème s'applique en particulier lorsque les X_n sont *indépendantes* de même loi. En fait dans ce cas, on a un résultat plus précis, mais incomparablement plus compliqué, appelé "loi du logarithme itéré", (où $\sigma(X_1)$ désigne l'écart-type de X_1) :

$$\overline{\lim_n} \frac{X_1 + \ldots + X_n - n\mathbb{E}(X_1)}{\sqrt{2n\ln(\ln n)}} \overset{p.s.}{=} \sigma(X_1) \text{ et } \underline{\lim_n} \frac{X_1 + \ldots + X_n - n\mathbb{E}(X_1)}{\sqrt{2n\ln(\ln n)}} \overset{p.s.}{=} -\sigma(X_1).$$

Ce résultat se met plus explicitement sous forme d'une vitesse de convergence dans la loi forte des grands nombres

$$\overline{\lim_n} \sqrt{\frac{n}{2\ln(\ln n)}} \left(\overline{X}_n - \mathbb{E}(X_1)\right) \overset{p.s.}{=} \sigma(X_1) \text{ et } \underline{\lim_n} \sqrt{\frac{n}{2\ln(\ln n)}} \left(\overline{X}_n - \mathbb{E}(X_1)\right) \overset{p.s.}{=} -\sigma(X_1).$$

Nous allons maintenant établir un résultat permettant de lever l'hypothèse d'identité en loi des X_n. Ce théorème en tant que tel est moins important que le précédent. En outre, il requiert que les X_n soient indépendantes.

<u>Théorème</u> :
Soit $(X_n)_{n\geq 1}$, une suite de v.a. indépendantes et dans L^2. Alors :

$$\sum_{n\geq 1} \frac{\mathrm{Var}(X_n)}{n^2} < +\infty \Rightarrow \frac{1}{n} \sum_{k=1}^{n} (X_k - \mathbb{E}(X_k)) \overset{\mathbb{P}\text{-}p.s.}{\longrightarrow} 0.$$

<u>Corollaire</u> :
Si de plus $\frac{1}{n} \sum\limits_{k=1}^{n} \mathbb{E}(X_k) \to m$ (par exemple si $\mathbb{E}(X_n) \to m$), alors

$$\frac{X_1 + \ldots + X_n}{n} \overset{\mathbb{P}\text{-}p.s.}{\longrightarrow} m.$$

Le théorème repose essentiellement sur l'inégalité de Kolmogorov :

<u>Proposition</u> : (inégalité de Kolmogorov)
Soit $X_1, \cdots, X_n$ des variables aléatoires indépendantes, $X_i \in L^2, 1 \leq i \leq n$ alors

$$\forall \varepsilon > 0, \mathbb{P}\left(\max_{1\leq k\leq n} |X_1 + \ldots + X_k - \mathbb{E}(X_1 + \ldots + X_k)| \geq \varepsilon\right) \leq \frac{1}{\varepsilon^2} \sum_{k=1}^{n} \mathrm{Var}(X_k).$$

<u>Démonstration</u> : Soit $S_k := X_1 + \ldots + X_k$. On peut supposer que $\mathbb{E}(X_i) = 0$, $1 \leq i \leq n$. On pose $A := \left\{\max\limits_{1\leq k\leq n} |S_k| \geq \varepsilon\right\}$ et $A_k = \{|S_1| < \varepsilon, \ldots, |S_{k-1}| < \varepsilon, |S_k| \geq \varepsilon\}$.

$A = \bigcup_{k=1}^{n} A_k$ et $A_k \cap A_\ell = \phi$ dès que $k \neq \ell$. D'autre part,

$$S_n^2 = S_k^2 + (S_n - S_k)^2 + 2S_k(S_n - S_k)$$

Or $\mathbb{E}(\mathbf{1}_{A_k} S_k (S_n - S_k)) = \mathbb{E}(\mathbf{1}_{A_k} S_k)\,\mathbb{E}(S_n - S_k) = 0$ car $S_n - S_k$ est indépendante de $\mathbf{1}_{A_k} S_k$. Par conséquent $\mathbb{E}(S_n^2 \mathbf{1}_{A_k}) \geq \mathbb{E}(S_k^2 \mathbf{1}_{A_k})$.

Finalement $\mathbb{E}(\mathbf{1}_A S_n^2) \geq \sum_{k=1}^{n} \mathbb{E}(S_k^2 \mathbf{1}_{A_k}) \geq \varepsilon^2 \mathbb{P}(A)$ et, comme les X_i sont indépendantes $\mathbb{E}(\mathbf{1}_A S_n^2) \leq \mathbb{E}(S_n^2) = \sum_{k=1}^{n} \mathrm{Var}(X_k)$. #

Lemme (Kronecker) :

Soit $(x_n)_{n\geq 1}$ une suite de réels. Alors

(i) $x_n \to x \Rightarrow \frac{1}{n}(x_1 + \ldots + x_n) \to x$, (Césaro)

(ii) $\sum_{k=1}^{n} x_k \to x \Rightarrow \frac{1}{n} \sum_{k=1}^{n} k x_k \to 0$ (Kronecker).

Démonstration : (i) est classique et
$\frac{1}{n} \sum_{k=1}^{n} k x_k = x_1 + \ldots + x_n - \frac{x_1 + (x_1 + x_2) + \ldots + (x_1 + \ldots + x_n)}{n} \to x - x = 0$. #

Lemme :

Soit (Y_n) une suite de v.a.r. indépendantes vérifiant $\sum_n \mathrm{Var}(Y_n) < +\infty$. Alors : $\sum_n (Y_n - \mathbb{E}(Y_n))$ converge $p.s.$ et dans L^2.

Preuve : D'après le critère 2 de convergence $p.s.$, il suffit de montrer que

$$\mathbb{P}\left(\sup_{k\geq 1} |Y_{n+1} + \ldots + Y_{n+k} - \mathbb{E}(Y_{n+1} + \ldots + Y_{n+k})| > \varepsilon\right) \to 0 \text{ quand } n \to +\infty.$$

Or, grâce à l'inégalité de Kolmogorov, par passage à la limite croissante,

$$\mathbb{P}\left(\sup_{k\geq 1} |Y_{n+1} + \ldots + Y_{n+k} - \mathbb{E}(Y_{n+1} + \ldots + Y_{n+k})| > \varepsilon\right) \leq \frac{1}{\varepsilon^2} \sum_{k=n+1}^{+\infty} \mathrm{Var}(X_k) \underset{n\to+\infty}{\to} 0$$

d'après l'hypothèse sur la série des variances. #

Démonstration du théorème : On applique le lemme ci-dessus à $Y_n = \frac{X_n}{n}$. D'où la convergence $p.s.$ (et L^2) de la série $\sum_k \frac{X_k - \mathbb{E}(X_k)}{k}$. Le le résultat en découle *via* le lemme de Kronecker. #

9.3.3 Loi des grands nombres : le cas L^1

Cette fois nous allons lever l'hypothèse L^2 (mais en supposant les v.a.r. de même loi dans le théorème final).

Rappel : Soit $(X_n)_{n\geq 1}$ une suite de v.a. indépendantes (intégrables). D'après la loi du 0-1 de Kolmogorov, il est clair que l'événement $\left\{\frac{X_1 + \ldots + X_n}{n} \text{ converge}\right\}$ est dans la tribu asymptotique donc de probabilité 0 ou 1 et qu'en outre la limite $p.s.$, si elle existe, est $p.s.$ constante.

Ceci sera utile à la "réciproque" du théorème.

Théorème :

Soit $(X_n)_{n\geq 1}$ une suite de v.a.r. indépendantes et de *même loi*. Alors :

$$\overline{X}_n := \frac{X_1 + \ldots + X_n}{n} \text{ converge } p.s. \text{ si et seulement si } \mathbb{E}|X_1| < +\infty$$

et dans ce cas $\frac{X_1 + \ldots + X_n}{n} \overset{\mathbb{P}\text{-}p.s.}{\to} \mathbb{E}(X_1)$.

La démonstration s'appuie sur un lemme général de comparaison.

Lemme :

Soit X une v.a. positive. Alors

(a) $\sum_{k\geq 1} \mathbb{P}(X \geq k) \leq \mathbb{E}X \leq 1 + \sum_{k\geq 1} \mathbb{P}(X \geq k)$,

(b) $\sum_{k\geq 1} \frac{1}{k^2} \mathbb{E}\left(X^2 \mathbf{1}_{\{X\leq k\}}\right) \leq 1 + 4\mathbb{E}(X)$.

Démonstration : (a) Du théorème de Fubini on déduit que :

$$\mathbb{E}(X) = \int \left(\int_0^{+\infty} \mathbf{1}_{\{X\geq t\}} dt d\mathbb{P}\right) = \int_0^{+\infty} \mathbb{P}(X \geq t)\, dt$$

$$= \sum_{k=0}^{+\infty} \int_k^{k+1} \mathbb{P}(X \geq t)\, dt \geq \sum_{k\geq 0} \mathbb{P}(X \geq k+1) \text{ et } \leq \sum_{k\geq 0} \mathbb{P}(X \geq k).$$

(b) Toujours d'après le théorème du Fubini, on a :

$$\forall k \in \mathbb{N}, \mathbb{E}\left(X^2 \mathbf{1}_{\{X\leq k\}}\right) = 2\int_0^{+\infty} x \mathbb{P}(x \leq X \leq k)\, dx \leq 2\int_0^{k} x \mathbb{P}(X \geq x)\, dx$$

donc :

$$\sum_{k=2}^{+\infty} \frac{1}{k^2} \mathbb{E}\left(X^2 \mathbf{1}_{\{X\leq k\}}\right) \leq 2\int_0^{+\infty} x \left(\sum_{k=2}^{+\infty} \frac{1}{k^2} \mathbf{1}_{\{x\leq k\}}\right) \mathbb{P}(X \geq x)\, dx$$

$$
\begin{aligned}
&\leq 2\int_0^{+\infty} x \int_{1\vee(x-1)}^{+\infty} \frac{du}{u^2} \mathbb{P}\left(X \geq x\right) dx \\
&\leq 2\int_0^{+\infty} \underbrace{\frac{x}{1\vee(x-1)}}_{\leq 2} \mathbb{P}\left(X \geq x\right) dx \\
&\leq 4\int_0^{+\infty} \mathbb{P}\left(X \geq x\right) dx = 4\mathbb{E}\left(X\right). \quad \#
\end{aligned}
$$

<u>Démonstration du théorème</u> : $\underline{\Leftarrow}$: On pose $Y_k = X_k \mathbf{1}_{\{|X_k| \leq k\}}$. D'après le lemme (b), $\sum_{k\geq 1} \frac{\operatorname{Var}(Y_k)}{k^2} < +\infty$ donc, d'après le second théorème du cas L^2, il vient :

$$\frac{1}{n}\sum_{k=1}^{n} \left(Y_k - \mathbb{E}\left(Y_k\right)\right) \overset{\mathbb{P}\text{-}p.s.}{\longrightarrow} 0.$$

D'autre part $Y_k(\omega) = X_k(\omega)$ pour k suffisamment grand donc $\frac{S_n}{n} - \frac{1}{n}\sum_{k=1}^{n} Y_k \overset{\mathbb{P}\text{-}p.s.}{\longrightarrow} 0$. Enfin $\mathbb{E}\left(Y_k\right) = \int x \mathbf{1}_{\{|x|\leq k\}} \mathbb{P}_{X_1}(dx) \to \mathbb{E}\left(X_1\right)$ par convergence dominée et donc $\frac{1}{n}\sum_{k=1}^{n} \mathbb{E}\left(Y_k\right) \to \mathbb{E}\left(X_1\right)$.

$\underline{\Rightarrow}$: Si $\frac{S_n}{n} \overset{p.s.}{\to} c$, d'après la loi du 0-1, la v.a. c est $p.s.$ constante. D'autre part $\frac{X_n}{n} = \frac{S_n}{n} - \frac{n-1}{n}\frac{S_{n-1}}{n-1} \overset{p.s.}{\to} 0$ donc $\mathbb{P}\left(\bigcap_n \bigcup_{k\geq 1} \{|X_k| \geq k\}\right) = 0$ et, comme les événements $\{|X_k| \geq k\}$ sont indépendants, il vient par le lemme de Borel-Cantelli que, nécessairement, $\sum_n \mathbb{P}\left(\{|X_n| \geq n\}\right) < +\infty$ (contraposée !) *i.e.*, comme les X_i ont même loi, $\sum_n \mathbb{P}\left(\{|X_1| \geq n\}\right) < +\infty$ donc $\mathbb{E}|X_1| < +\infty$. Et donc finalement $c = \mathbb{E}\left(X_1\right)$ d'après le sens direct. #

Chapitre 10

Convergence en loi

Jusqu'à maintenant tous les modes de convergence que nous avons rencontrés faisaient intervenir l'espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$ sur lequel étaient définies les v.a.. La convergence en loi, comme son nom l'indique, ne porte que sur les lois de v.a. ce qui en fait une notion particulièrement adaptée aux problèmes de nature probabiliste.

10.1 Convergence étroite de mesures de probabilité

Définition :
Soient $\mu_n, n \geq 1$, et μ des probabilités sur $\left(\mathbb{R}^d, \mathcal{B}\left(\mathbb{R}^d\right)\right)$. On dira que μ_n *converge étroitement vers* μ (notée $\mu_n \Rightarrow \mu$) si :

$$\forall f \in \mathcal{C}_b\left(\mathbb{R}^d, \mathbb{R}\right) := \left\{f : \mathbb{R}^d \to \mathbb{R} \text{ continue bornée}\right\}, \int f d\mu_n \overset{n \to +\infty}{\longrightarrow} \int f d\mu.$$

Remarque : On peut étendre cette définition à des mesures finies.

Justification : Pour parler de convergence il faut que la limite soit unique. C'est le cas puisque si $\mu_n \Rightarrow \mu$ et μ', alors μ et μ' coïncident sur $\mathcal{C}_b\left(\mathbb{R}^d, \mathbb{R}\right)$ donc sont égales.

Exemple : Si $\mu_n := \delta_{a_n}$ et $\mu := \delta_a$, $\mu_n \Rightarrow \mu$ si et seulement si $a_n \to a$.
En effet si $a_n \to a$ il est évident que $f(a_n) \to f(a)$ pour toute fonction f continue (en a !).
Réciproquement, pour tout $i \in \{1, \ldots, d\}$ fixé, $x := (x^1, \cdots, x^d) \mapsto \mathrm{Arctg}\,(x^i)$ est dans $\mathcal{C}_b\left(\mathbb{R}^d, \mathbb{R}\right)$, $\mathrm{Arctg}(a_n^i) \to \mathrm{Arctg}(a^i)$ quand $n \to +\infty$ et donc $a_n^i \to a^i, 1 \leq i \leq n$, i.e. $a_n = \left(a_n^1, \ldots, a_n^d\right) \to a = \left(a^1, \ldots, a^d\right)$.

En fait on peut prendre un ensemble de "fonctions-test" beaucoup plus petit que celui de la définition.

Proposition 1 : Soient $(\mu_n)_{n \geq 1}$ une suite de probabilités sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ et μ une probabilité sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$.

$$\mu_n \Rightarrow \mu \text{ si et seulement si } \forall f \in \mathcal{C}_K\left(\mathbb{R}^d, \mathbb{R}\right) \int f d\mu_n \to \int f d\mu.$$

Démonstration : Seule la réciproque est à montrer. Soit, pour tout $p \geq 1$, $\varphi_p : \mathbb{R}_+ \to [0,1]$, continue, valant 1 sur $[0,p]$, nulle en dehors de $[0, p+1]$.
Par suite, $\varphi_p(\|.\|) \in \mathcal{C}_K\left(\mathbb{R}^d, \mathbb{R}\right)$ et $f_p := f.\varphi_p(\|.\|) \in \mathcal{C}_K\left(\mathbb{R}^d, \mathbb{R}\right)$ pour tout $p \geq 1$.
Donc $\int \varphi_p(\|x\|)\mu_n(dx) \to \int \varphi_p(\|x\|)\mu(dx)$ et $\int f_p d\mu_n \to \int f_p d\mu$ quand $n \to +\infty$.
Comme $f = f_p + f\left(1 - \varphi_p(\|.\|)\right)$ et $\|f\|_\infty = \sup_{x \in \mathbb{R}^d} |f(x)| < +\infty$, il vient :

$$\left|\int f d\mu_n - \int f d\mu\right|$$

$$\leq \left|\int f_p d\mu_n - \int f_p d\mu\right| + \|f\|_\infty \left(\int \left(1 - \varphi_p(\|x\|)\right)\mu(dx) + \int \left(1 - \varphi_p(\|x\|)\right)\mu_n(dx)\right),$$

$$\leq \left|\int f_p d\mu_n - \int f_p d\mu\right| + \|f\|_\infty \left(2 - \int \varphi_p(\|x\|)\mu(dx) - \int \varphi_p(\|x\|)\mu_n(dx)\right),$$

d'où $\overline{\lim_n}\left|\int f d\mu_n - \int f d\mu\right| \leq 0 + 2\|f\|_\infty \left(1 - \int \varphi_p(\|x\|)\mu(dx)\right)$ pour tout $p \geq 1$.
Or $\int \varphi_p(\|x\|)\mu(dx) \underset{p \to +\infty}{\to} \int 1 d\mu = 1$ par convergence dominée car $\varphi_p \uparrow 1$. Cela achève la démonstration. #

Remarques : (1) Si μ_n et μ sont seulement *finies*, le résultat subsiste sous réserve d'ajouter l'hypothèse : $\mu_n(\mathbb{R}^d) \to \mu(\mathbb{R}^d)$.

(2) Si l'on a seulement $\mu_n(\mathbb{R}^d)_{n \geq 1}$, *bornée*, il subsiste le résultat plus faible :

$$\forall f \in \mathcal{C}_K\left(\mathbb{R}^d, \mathbb{R}\right) \int f d\mu_n \to \int f d\mu \Rightarrow \begin{cases} (i) & \mu\left(\mathbb{R}^d\right) \leq \underline{\lim}_n \mu_n\left(\mathbb{R}^d\right) < +\infty, \\ (ii) & \forall f \in \mathcal{C}_0\left(\mathbb{R}^d, \mathbb{R}\right) \int f d\mu_n \to \int f d\mu. \end{cases}$$

En effet, $\forall p \geq 1$, $\int \varphi_p(\|x\|)\mu(dx) = \lim_n \int \varphi_p(\|x\|) d\mu_n \leq 1 \times \underline{\lim}_n \mu_n\left(\mathbb{R}^d\right)$. D'où, par le lemme de Fatou :

$$\mu\left(\mathbb{R}^d\right) = \int \underline{\lim}_p \varphi_p(\|x\|)\mu(dx) \leq \underline{\lim}_p \int \varphi_p(\|x\|)\mu(dx) \leq \underline{\lim}_n \mu_n\left(\mathbb{R}^d\right).$$

Soit maintenant $f \in \mathcal{C}_0\left(\mathbb{R}^d, \mathbb{R}\right)$. Avec les notations de la preuve de la proposition 1, on a $f = f_p + f\mathbf{1}_{\{\|x\| \geq p\}}\left(1 - \varphi_p(\|.\|)\right)$.

D'où il vient : $\forall p \geq 1$,

$$\overline{\lim_n}\left|\int f d\mu_n - \int f d\mu\right| \leq \sup_{\|x\|\geq p} |f(x)| \left(\mu\left(\mathbb{R}^d\right) + \overline{\lim_n}\mu_n\left(\mathbb{R}^d\right) - 2\int \varphi_p(\|x\|)\,\mu(dx)\right) \underset{p\to+\infty}{\longrightarrow} 0.$$

<u>Exemple</u> (sur les remarques) : $\mu_n = \frac{1}{2}\left(\delta_{\frac{1}{n}} + \delta_n\right)$ et $\mu = \frac{1}{2}\delta_0$. On constate que $\mu_n(\mathbb{R})$ ne tend pas vers $\mu(\mathbb{R})$, mais que $\int f d\mu_n \to \int f d\mu$ si $f \in \mathcal{C}_0(\mathbb{R},\mathbb{R})$. En revanche, $\int \sin(x)\,\mu_n(dx) = \dfrac{\sin\frac{1}{n} + \sin n}{2} \not\to \frac{1}{2}\sin 0 = 0$ bien que $\sin(.) \in \mathcal{C}_b(\mathbb{R},\mathbb{R})$.

<u>Application à la convergence étroite des probabilités discrètes.</u>

Soit E une partie de $\mathbb{R}^d$ *localement finie i.e.* telle que, pour tout compact K de $\mathbb{R}^d$, $E\cap K$ soit fini (typiquement E est fini ou $E := \mathbb{N}^d, \mathbf{Z}^d, \ldots$). Un tel ensemble est nécessairement dénombrable car $E = \cup_{n\geq 1}(E\cap B(o;n))$. Soient μ_n et μ des probabilités vérifiant $\mu_n(E) = \mu(E) = 1$. Alors :

$$\mu_n \Rightarrow \mu \text{ si et seulement si } \forall e\in E\ \mu_n(\{e\}) \to \mu(\{e\}).$$

<u>Démonstration</u> : Soit $e_0 \in E$. Comme $E\cap\overline{B}(e_0,1)$ est fini, il existe r_0 tel que $E\cap\overline{B}(e_0,r_0) = \{e_0\}$ (prendre $r_0 := \min\{|e_0 - e|,\ e\in\overline{B}(e_0,1)\cap E\}$). On pose $F_0(x) := \max(r_0 - |x - e_0|, 0)$, $x\in\mathbb{R}^d$. La fonction f_0 est clairement continue et bornée donc $\int f_e d\mu_n \to \int f_e d\mu$.

Or $\int f_0 d\mu_n = r_0\mu_n(\{e_0\})$ et $\int f_0 d\mu = r_0\mu(\{e_0\})$, d'où le résultat.

Réciproquement, pour toute fonction $f\in\mathcal{C}_K\left(\mathbb{R}^d,\mathbb{R}\right)$,

$$\int f\,d\mu_n = \sum_{x\in E\cap\mathrm{supp}(f)} f(x)\mu_n(\{x\}) \longrightarrow n\to+\infty \sum_{x\in E\mathrm{supp}(f)} f(x)\mu(\{x\})$$

($E\cap\mathrm{supp}(f)$ est fini puisque f est à support compact). D'où le résultat d'après la proposition 1. #

<u>Exemples</u> :

(1) <u>Convergence d'une binômiale</u> $\mathcal{B}(n;p_n)$, $np_n \to \lambda > 0$.
Nous avons vu (cf. chap. I.4) que sous l'hypothèse ci-dessus, si $\mu_n := \mathcal{B}(n;p_n)$, $\forall k\in\mathbb{N}$, $\mu_n(\{k\}) \underset{n\to+\infty}{\longrightarrow} e^{-\lambda}\frac{\lambda^k}{k!}$. On en déduit donc (avec $E = \mathbb{N}$) que : $\mathcal{B}(n;p_n) \Rightarrow \mathcal{P}(\lambda)$.

(2) <u>Convergence d'une loi hypergéométrique vers une binomiale.</u>

Soit $\mu_{(N_1,N_2)} = \mathcal{H}(n;N_1;N_2)$. Alors :

$$\left.\begin{array}{ll} N_1 \text{ et } N_2 & \to +\infty \\ \dfrac{N_1}{N_1+N_2} & \to p\in\,]0,1[\end{array}\right\} \Rightarrow \left(\mu_{(N_1,N_2)} \Rightarrow \mathcal{B}(n;p)\right)$$

En effet, si $k\in\{0,\ldots,N\}$, $\mathbb{P}(X_n = k) = \dfrac{C_{N_1}^k C_{N_2}^{n-k}}{C_{N_1+N_2}^n}$ soit encore

$$\begin{aligned}\mathbb{P}(X_N = k) &= \frac{\dfrac{N_1!}{k!(N_1-k)!}\dfrac{N_2!}{(n-k)!(N_2-n+k)!}}{\dfrac{(N_1+N_2)!}{n!(N_1+N_2-n)!}} \\ &= \frac{n!}{k!(n-k)!}\times\frac{N_1}{N_1+N_2}\times\frac{N_1-1}{N_1+N_2-1}\times\ldots\times\frac{N_1-k+1}{N_1+N_2-k+1}\times \\ &\quad \frac{N_2}{N_1+N_2-k}\times\frac{N_2-1}{N_1+N_2-k-1}\times\ldots\times\frac{N_2-n+k+1}{N_1+N_2-n+1} \\ &\underset{\frac{N_1}{N_1+N_2}\to p\in]0,1[}{\overset{N_1,N_2\to+\infty}{\longrightarrow}} C_n^k p^k(1-p)^{n-k}.\end{aligned}$$

On notera le caractère intuitif de ce résultat.

Revenons maintenant au cas général. En fait on peut affiner la proposition 1 en :

<u>Proposition 2</u> :
Soient $\mu_n, n\geq 1$, et μ des probabilités et

$$\mathcal{C}\subset\mathcal{C}_0\left(\mathbb{R}^d,\mathbb{R}\right) := \left\{f : \mathbb{R}^d\to\mathbb{R} \text{ continue}, \lim_{\|x\|\to+\infty} f(x) = 0\right\}.$$

Alors :

$$\left.\begin{array}{ll} (i) & \mathcal{C} \text{ total dans } \left(\mathcal{C}_0\left(\mathbb{R}^d,\mathbb{R}\right), \|.\|_\infty\right) \\ (ii) & \forall f\in\mathcal{C}, \int f d\mu_n \to \int f d\mu \end{array}\right\} \Rightarrow (\mu_n \Rightarrow \mu).$$

<u>Démonstration</u> : Soit $f\in\mathcal{C}_K\left(\mathbb{R}^d,\mathbb{R}\right)\subset\mathcal{C}_0\left(\mathbb{R}^d,\mathbb{R}\right)$ et $f_p\in vect(\mathcal{C})$, $p\geq 1$, telles que $\|f - f_p\|_\infty \to 0$.

$$\left|\int f d\mu - \int f d\mu_n\right| \leq \left|\int f_p d\mu - \int f_p d\mu_n\right| + 2\|f - f_p\|_\infty.$$

Or $\int f_p d\mu_n \underset{n\to+\infty}{\longrightarrow} \int f_p d\mu$ par linéarité via (ii), donc

$$\overline{\lim_n}\left|\int f d\mu - \int f d\mu_n\right| \leq 2\|f - f_p\|_\infty \text{ pour tout } p\geq 1.$$

D'où $\int f d\mu_n \to \int f d\mu$ dès que $f \in \mathcal{C}_K\left(\mathbb{R}^d, \mathbb{R}\right)$. Le résultat final découle de la proposition 1. #

Remarques : • Ce résultat s'étend aux mesures finies en ajoutant (iii) : $\mu_n\left(\mathbb{R}^d\right) \to \mu\left(\mathbb{R}^d\right)$.

• L'énoncé contient évidemment la proposition 1 puisque $\mathcal{C}_K(\mathbb{R}^d, \mathbb{R})$ est dense dans $\mathcal{C}_0(\mathbb{R}^d, \mathbb{R})$. En fait, $\mathcal{C}_0(\mathbb{R}^d, \mathbb{R})$ est l'adhérence de $\mathcal{C}_K(\mathbb{R}^d, \mathbb{R})$ dans $\mathcal{C}_b\left(\mathbb{R}^d, \mathbb{R}\right)$ pour la topologie de la norme uniforme sur $\mathbb{R}^d$ et c'est la raison de l'apparition de $\mathcal{C}_0\left(\mathbb{R}^d, \mathbb{R}\right)$ dans cet énoncé.

Application : Caractérisation de la *convergence étroite* par la *transformée de Laplace*. On se place sur $(\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+))$. Soit $\mu_n, n \geq 1$, et μ des probabilités. Alors :

$$\mu_n \Rightarrow \mu \text{ si et seulement si } \widetilde{\mu}_n \to \widetilde{\mu}.$$

En effet $\mathcal{C} := \{e^{-s.}, s > 0\}$ est total dans $\left(\mathcal{C}_0(\mathbb{R}_+, \mathbb{R}), \|.\|_\infty\right)$ (cf. VI.6).

Nous allons donner une caractérisation de la convergence étroite sur les boréliens de $\mathbb{R}^d$, connue sous le nom de théorème Porte-manteau.

Théorème 1 :
Les assertions suivantes sont équivalentes :

(i) $\mu_n \Rightarrow \mu$,

(ii) $\forall O \subset \mathbb{R}^d, O$ ouvert, $\mu(O) \leq \underline{\lim}_n \mu_n(O)$,

(iii) $\forall F \subset \mathbb{R}^d, F$ fermé, $\mu(F) \geq \overline{\lim}_n \mu_n(F)$,

(iv) $\forall A \in \mathcal{B}\left(\mathbb{R}^d\right)$, $\mu(\partial A) = 0 \Rightarrow \mu_n(A) \to \mu(A)$ $\left[\text{où } \partial A := \overline{A} \setminus \overset{o}{A}\right]$,

Démonstration :
$(i) \Rightarrow (ii)$: On considère, pour tout entier $p \geq 1$, la fonction f_p définie par $f_p(x) := \min(1, p\, d(x, {}^cO))$. Il est clair que les fonctions f_p sont continues, à valeurs dans $[0,1]$ et que $f_p \leq f_{p+1}$.

D'autre part, ${}^cO$ étant fermé, $d(x, {}^cO) = 0$ si et seulement si $x \notin O$. Par suite $f_p \uparrow \mathbf{1}_O$. Par suite, pour tout $p \geq 1$ $\int f_p d\mu_n \to \int f_p d\mu$. D'où

$$\forall\, p \geq 1, \qquad \underline{\lim}_n \mu_n(O) \geq \lim_n \int f_p d\mu_n = \int f_p d\mu.$$

Enfin $\int f_p d\mu \uparrow \mu(O)$ quand $p \to= \infty$ par convergence monotone.

$(ii) \Longleftrightarrow (iii)$: Evident par passage au complémentaire.

$(iii) \Rightarrow (iv)$: $\overset{o}{A} \subset A \subset \overline{A}$ donc

$$\mu\left(\overset{o}{A}\right) \leq \underline{\lim}_n \mu_n\left(\overset{o}{A}\right) \leq \underline{\lim}_n \mu_n(A) \leq \overline{\lim}_n \mu_n(A) \leq \overline{\lim}_n \mu_n\left(\overline{A}\right) \leq \mu\left(\overline{A}\right).$$

Or $\mu\left(\overset{o}{A}\right) = \mu(\overline{A}) = \mu(A)$ par hypothèse, donc

$$\mu(A) = \underline{\lim}_n \mu_n(A) = \overline{\lim}_n \mu_n(A) = \lim_n \mu_n(A).$$

$(iv) \Rightarrow (ii)$: $\partial O_\varepsilon = \{x / d(x, {}^cO) = \varepsilon\}$. Il est clair que, μ étant une probabilité, $I_n := \{\varepsilon / \mu(\partial O_\varepsilon) \geq 1/n\}$ a au plus n éléments donc $\{\varepsilon / \mu(\partial O_\varepsilon) > 0\} = \bigcup_{n \geq 1} I_n$ est au plus dénombrable.

Par conséquent, comme $\mathbf{1}_{O_\varepsilon} \uparrow \mathbf{1}_O$ quand $\varepsilon \downarrow 0$, il existe ε_p, $p \geq 1$, tel que $\mathbf{1}_{O_{\varepsilon_p}} \uparrow \mathbf{1}_O$ et $\mu\left(\partial O_{\varepsilon_p}\right) = 0$. D'où

$$\forall p \geq 1,\ \mu\left(O_{\varepsilon_p}\right) = \lim_n \mu_n\left(O_{\varepsilon_p}\right) \leq \underline{\lim}_n \mu_n(O).$$

On conclut en remarquant que $\mu(O_{\varepsilon_p}) \uparrow \mu(O)$ quand $p \uparrow +\infty$.

$(ii) \Rightarrow (i)$: Soit $f \in \mathcal{C}_b\left(\mathbb{R}^d, \mathbb{R}\right)$. A une transformation affine près, on peut supposer que : $0 \leq f < M$. Pour tout $p \geq 1$, on pose $f_p := M \sum_{0 \leq k \leq p-1} \frac{k}{p} \mathbf{1}_{\left\{\frac{k}{p} M \leq f < \frac{k+1}{p} M\right\}}$ et $A_k^{(p)} := \left\{f < \frac{k}{p} M\right\}$, $0 \leq k \leq p$. On vérifie aussitôt que : $0 \leq f - f_p \leq \frac{M}{p}$ et, en intégrant par rapport à μ,

$$\int f_p d\mu = \frac{M}{p} \sum_{k=0}^{p-1} k\left(\mu\left(A_{k+1}^{(p)}\right) - \mu\left(A_k^{(p)}\right)\right) = \frac{M}{p}\left((p-1)\,\mu\left(A_p^{(p)}\right) - \sum_{k=1}^{p-1} \mu\left(A_k^{(p)}\right)\right).$$

Cette identité est aussi vérifiée en remplaçant μ par μ_n.
Remarquant alors que $A_p^{(p)} = \mathbb{R}^d$, $A_0^{(p)} = \phi$ et que les $A_k^{(p)}, k \leq p-1$, sont ouverts, il vient, pour tout $p \geq 1$:

$$\begin{aligned}
\overline{\lim}_n \int f_p d\mu_n &\leq \frac{M}{p}\left((p-1)\,\overline{\lim}_n \mu_n\left(A_p^{(p)}\right) + \overline{\lim}_n\left(-\sum_{k=1}^{p-1} \mu_n\left(A_k^{(p)}\right)\right)\right) \\
&\leq \frac{M}{p}\left((p-1)\,\overline{\lim}_n \mu_n\left(A_p^{(p)}\right) - \sum_{k=1}^{p-1} \underline{\lim}_n \mu_n\left(A_k^{(p)}\right)\right) \\
&\leq \frac{M}{p}\left((p-1)\,\mu\left(A_p^{(p)}\right) - \sum_{k=1}^{p-1} \mu\left(A_k^{(p)}\right)\right) = \int f_p d\mu \leq \int f d\mu.
\end{aligned}$$

Partant,

$$\overline{\lim}_n \int f d\mu_n \leq \frac{M}{p} + \overline{\lim}_n \int f_p d\mu_n \leq \frac{M}{p} + \int f d\mu \text{ pour tout } p \geq 1.$$

D'où $\overline{\lim}_n \int f d\mu_n \leq \int f d\mu$.

Comme $0 < M - f \leq M < 2M$ et $M - f \in \mathcal{C}\left(\mathbb{R}^d, \mathbb{R}\right)$, on a de même

$$\overline{\lim_n} \int (M - f)\, d\mu_n \leq \int (M - f)\, d\mu$$

d'où, immédiatement, $\int f d\mu \leq \underline{\lim}_n \int f d\mu_n$.

Finalement : $\int f d\mu = \underline{\lim}_n \int f d\mu_n = \overline{\lim}_n \int f d\mu_n = \lim_n \int f d\mu_n$. #

<u>Corollaire 1</u> :
(a) Soit $h : \mathbb{R}^d \to \mathbb{R}^p$ borélienne et μ_n et μ des probabilités sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. Alors :

$$\left.\begin{array}{ll} (i) & \mu\left(\{x / h \text{ n'est pas continue en } x\}\right) = 0 \\ (ii) & \mu_n \Rightarrow \mu \end{array}\right\} \Rightarrow \left(h\left(\mu_n\right) \Rightarrow h\left(\mu\right)\right).$$

(b) Si h est en outre à valeurs réelles et bornée alors $\int h d\mu_n \to \int h d\mu$.

<u>Démonstration</u> :
(a) Soit F un fermé de $\mathbb{R}^p$ et $x \in \overline{h^{-1}(F)}$. Soit $x_n \in h^{-1}(F)$, $x_n \to x$. Si h est continue en x, $h(x) = \lim_n h(x_n) \in \overline{F} = F$ donc

$$\overline{h^{-1}(F)} \subset h^{-1}(F) \cup \mathrm{Disc}(h)$$

où $\mathrm{Disc}(f) := \left\{x \in \mathbb{R}^d / f \text{ est discontinue en } x\right\}$. Par suite, il vient d'après le (iii) de la proposition ci-dessus :

$$\begin{aligned} \overline{\lim_n}\, h(\mu_n)(F) = \overline{\lim_n}\, \mu_n\left(h^{-1}(F)\right) & \leq \overline{\lim_n}\, \mu_n\left(\overline{h^{-1}(F)}\right) \leq \mu\left(\overline{h^{-1}(F)}\right) \\ & \leq \mu\left(h^{-1}(F)\right) + \underbrace{\mu\left(\mathrm{Disc}(h)\right)}_{=0} = \mu(h^{-1}(F)) = h(\mu)(F). \end{aligned}$$

(b) Comme $h(\mu_n) \Rightarrow h(\mu)$, $\int f \circ h d\mu_n \to \int f \circ h d\mu$, $\forall f \in \mathcal{C}_b(\mathbb{R}^p, \mathbb{R})$, donc en particulier pour les $f_k(x) = x\mathbf{1}_{\{|x| \leq k\}} + k\frac{x}{|x|}\mathbf{1}_{\{|x|>k\}}$, $k > 0$. Or pour k suffisamment grand $f_k \circ h = h$. #

Nous allons maintenant revenir aux v.a. en introduisant la convergence en loi.

10.2 Convergence en loi de variables aléatoires

<u>Définition</u> :
Soient $X_n, n \geq 1$, et X des variables aléatoires à valeurs dans $\mathbb{R}^d$, $d \geq 1$. On dira que X_n converge en loi vers X, (noté "$X_n \xrightarrow{\mathcal{L}} X$") si et seulement si $\mathbb{P}_{X_n} \Rightarrow \mathbb{P}_X$ (les lois convergent étroitement).

Sur un plan pratique, cela se traduit à l'aide du théorème de la mesure image par :

$$X_n \xrightarrow{\mathcal{L}} X \text{ si et seulement si } \forall f \in \mathcal{C}_b\left(\mathbb{R}^d, \mathbb{R}\right) \quad \mathbb{E}\left(f\left(X_n\right)\right) \to \mathbb{E}\left(f\left(X\right)\right).$$

Au vu des résultats précédents on peut affirmer que la loi limite est unique, que l'ensemble des fonctions-test peut-être réduit à $\mathcal{C}_K\left(\mathbb{R}^d, \mathbb{R}\right)$, voire à une partie totale (pour la norme uniforme) $\mathcal{C} \subset \mathcal{C}_0\left(\mathbb{R}^d, \mathbb{R}\right)$. En particulier, si X_n et X sont réelles *positives*

$$X_n \xrightarrow{\mathcal{L}} X \text{ si et seulement si } \widetilde{\mu}_{X_n} \to \widetilde{\mu}_X \text{ (transformée de Laplace).}$$

La traduction du théorème porte-manteau et de son corollaire est immédiate :

<u>Proposition 3</u> :
Il y a équivalence entre :
(i) $X_n \xrightarrow{\mathcal{L}} X$,
(ii) $\forall O \subset \mathbb{R}^d, O$ ouvert, $\mathbb{P}(X \in O) \leq \underline{\lim}_n \mathbb{P}(X_n \in O)$,
(iii) $\forall F \subset \mathbb{R}^d, F$ fermé, $\mathbb{P}(X \in F) \geq \overline{\lim}_n \mathbb{P}(X_n \in F)$,
(iv) $\forall A \in \mathcal{B}\left(\mathbb{R}^d\right), \mathbb{P}(X \in \partial A) = 0 \Rightarrow \mathbb{P}(X \in A) = \lim_n \mathbb{P}(X_n \in A)$.

<u>Corollaire</u> :
Soit $h : \mathbb{R}^d \to \mathbb{R}^p$, borélienne et $X_n \xrightarrow{\mathcal{L}} X$ sur $\mathbb{R}^d$. Alors :

$$\mathbb{P}\left(X \in \mathrm{Disc}(h)\right) = 0 \Rightarrow h\left(X_n\right) \xrightarrow{\mathcal{L}} h\left(X\right)$$

(et $\mathbb{E}\left(h\left(X_n\right)\right) \to \mathbb{E}\left(h\left(X\right)\right)$ si h est *bornée*)

<u>Application</u> (caractérisation de la convergence en loi des v.a.r. par les fonctions de répartitions) :

<u>Théorème 2</u> :
Soient $X_n, n \geq 1$, et X des v.a.r. de fonctions de répartition respectives F_n et F. Alors :

$$X_n \xrightarrow{\mathcal{L}} X \text{ si et seulement si } \forall x \in \mathrm{Cont}(F), \; F_n(x) \underset{n \to +\infty}{\longrightarrow} F(x)$$

où $\mathrm{Cont}(F) := \{x \in \mathbb{R} \,/ F \text{ est continue en } x\}$.

<u>Démonstration</u> (partielle) : <u>$\Rightarrow$:</u> F est continue en x si et seulement si $\mathbb{P}_X(\{x\}) = 0$ i.e. si et seulement si $\mathbb{P}(X \in \partial\,]-\infty, x]) = 0$. Donc si $X_n \xrightarrow{\mathcal{L}} X$ et $x \in \mathrm{Cont}(F)$, $\mathbb{P}_{X_n}(]-\infty, x]) \to \mathbb{P}_X(]-\infty, x])$ i.e. $F_n(x) \to F(x)$.

<u>$\Leftarrow$:</u> Supposons F continue sur $\mathbb{R}$. Alors, $\forall I :=]a,b[\,,\ a,b \in \overline{\mathbb{R}}, a \leq b$, on a

$$\mathbb{P}_{X_n}(]a,b[) = F_n(b) - F_n(a) \to F(b) - F(a) = \mathbb{P}_X(]a,b[)\,.$$

Soit O un ouvert de $\mathbb{R}$; O s'écrit comme réunion dénombrable disjointe d'intervalles ouverts $I_p, p \geq 1$, d'où $\mathbb{P}_X(O) = \sum_{p\geq 1} \mathbb{P}_X(I_p)$. Par suite

$$\mathbb{P}_X(O) = \sum_{p\geq 1} \lim_n \mathbb{P}_{X_n}(I_p) \leq \underline{\lim}_n \sum_{p\geq 1} \mathbb{P}_{X_n}(I_p) = \underline{\lim}_n \mathbb{P}_{X_n}(O)\,.$$

On conclut que $X_n \xrightarrow{\mathcal{L}} X$ via (ii) de la proposition 3. #

10.3 Convergence en loi et fonction caractéristique

Nous allons voir que les fonctions caractéristiques sont un outil particulièrement bien adapté à l'étude de la convergence en loi des v.a. d-dimensionnelles.

On rappelle que si $X\ :\ (\Omega, \mathcal{A}, \mathbb{P}) \longrightarrow (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ est une v.a.,

$$\Phi_X(u) := \int e^{i(u,x)} \mathbb{P}_X(dx) = \mathbb{E}(e^{i(u,X)})$$

où $(u,x) = \sum_{i=1}^{d} u_i x_i$ désigne le produit scalaire canonique sur $\mathbb{R}^d$.

10.3.1 Caractérisation

<u>Théorème 3</u> :

$X_n \xrightarrow{\mathcal{L}} X$ si et seulement si $\Phi_{X_n} \to \Phi_X$ (fonctions caractéristiques).

<u>Démonstration</u> :

<u>$\Rightarrow$:</u> C'est évident puisque $e^{i(u,x)} = \underbrace{\cos(u,x)}_{\in \mathcal{C}_b(\mathbb{R}^d,\mathbb{R})} + i \underbrace{\sin(u,x)}_{\in \mathcal{C}_b(\mathbb{R}^d,\mathbb{R})}$

<u>$\Leftarrow$:</u> Rappelons que si μ est une probabilité, de transformée de Fourrier $\widehat{\mu}$, on a l'identité (cf. VI.3. lemme 2) : $\forall y \in \mathbb{R}^d$,

$$\int \varphi_\sigma(y-x)\,\mu(dx) = (2\pi)^{-\frac{d}{2}} \int e^{-i(u,y)} \widehat{\mu}(u)\,\varphi_1(\sigma u)\,du \quad \text{où} \quad \varphi_\sigma(u) = \frac{e^{-\frac{\|u\|^2}{2\sigma^2}}}{\left(\sigma\sqrt{2\pi}\right)^d}.$$

Posons $\mu := \mathbb{P}_X$ et $\mu_n := \mathbb{P}_{X_n}$. D'après le théorème de convergence dominée appliqué à l'intégrale du second membre, il vient :

$$\forall \sigma > 0,\ \ \forall y \in \mathbb{R}^d,\quad \int \varphi_\sigma(y-x)\,\mu_n(dx) \overset{n\to+\infty}{\longrightarrow} \int \varphi_\sigma(y-x)\,\mu(dx)\,.$$

A ce stade, on reprend le fil de la démonstration de l'injectivité de la transformation de Fourier : soit $f \in \mathcal{C}_K\left(\mathbb{R}^d, \mathbb{R}\right)$; par convergence dominée,

$$\forall \sigma > 0, \int f(y) \left[\int \varphi_\sigma(y-x)\,\mu_n(dx)\right] dy \underset{n\to+\infty}{\to} \int f(y) \left[\int \varphi_\sigma(y-x)\,\mu(dx)\right] dy.$$

Via le théorème de Fubini et le changement de variable $y := x + \sigma u$, il vient :

$$\forall \sigma > 0, \int \left(\int f(x+\sigma u)\,\varphi_1(u)\,du\right) \mu_n(dx) \underset{n\to+\infty}{\to} \int \left[\int f(x+\sigma u)\,\varphi_1(u)\,du\right] \mu(dx)\,.$$

Soit $\mathcal{C} := \left\{x \mapsto \int f(x+\sigma u)\,\varphi_1(u)\,du, f \in \mathcal{C}_K\left(\mathbb{R}^d, \mathbb{R}\right), \sigma > 0\right\}$. $\mathcal{C} \subset \mathcal{C}_0\left(\mathbb{R}^d, \mathbb{R}\right)$ (évident par convergence dominée) et $\mathcal{C}$ est en fait dense ; pour démontrer ce dernier point il nous suffit d'établir que toute fonction $f \in \mathcal{C}_K\left(\mathbb{R}^d, \mathbb{R}\right)$ est <u>uniformément</u> approchable par des éléments de $\mathcal{C}$.

$$\begin{aligned}\left|\int f(x+\sigma u)\,\varphi_1(u)\,du - f(x)\right| &\leq \int |f(x+\sigma u) - f(x)|\,\varphi_1(u)\,du \\ &\quad (\varphi_1(u) \text{ est une densité de probabilité}) \\ &\leq \sup_{|v|\leq\delta} |f(x+v) - f(x)| \times 1 + 2\,\|f\|_\infty \int_{\{\|u\|>\frac{\delta}{\sigma}\}} \varphi_1(u)\,du\end{aligned}$$

d'où il vient immédiatement :

$$\forall \delta > 0,\ \overline{\lim_{\sigma\to 0}} \sup_{x\in\mathbb{R}^d} \left|\int f(x+\sigma u)\,\varphi_1(u)\,du - f(x)\right| \leq \sup_{x\in\mathbb{R}^d, |v|\leq\delta} |f(x+v) - f(x)|$$

Or f étant uniformément continue (car dans $\mathcal{C}_K\left(\mathbb{R}^d, \mathbb{R}\right)$), le second membre de l'inégalité tend vers 0 quand $\delta \to 0$. $\mathcal{C}$ est donc bien un sous-ensemble total de $\mathcal{C}_0(\mathbb{R}^d, \mathbb{R})$. On conclut par la proposition 2. #

<u>Exemples d'application</u> :

(1) <u>Convergence d'une loi binomiale vers une loi de Poisson</u> :

$$\text{Si } X_n \overset{\mathcal{L}}{\sim} \mathcal{B}(n; p_n),\ np_n \to \lambda > 0 \text{ alors } X_n \xrightarrow{\mathcal{L}} \mathcal{P}(\lambda).$$

En effet $\Phi_{X_n}(u) = ((1-p_n) + p_n e^{iu})^n = \left(1 + \dfrac{np_n(e^{iu}-1)}{n}\right)^n$. Or, comme $z \mapsto \left(1 + \dfrac{z}{n}\right)^n$ converge uniformément sur les compacts vers e^z (cf. appendice), si $z_n \to z$, $\left(1 + \dfrac{z_n}{n}\right)^n \to e^z$. Dans le cas qui nous occupe, on a donc $\Phi_{X_n}(u) \underset{n\to+\infty}{\to} e^{\lambda\left(e^{iu}-1\right)} = \Phi_{\mathcal{P}(\lambda)}(u)$.

On retrouve donc ainsi le résultat établi directement dans l'exemple 1 de la section 1 et au chapitre I).

(2) Premier contact avec le théorème central-limite :

Soit $X_n, n \geq 1$, suite de v.a. indépendantes et de même loi $\mathbb{P}(X_1 = \pm 1) = \frac{1}{2}$. Alors

$$Y_n := \frac{X_1 + ... + X_n}{\sqrt{n}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1).$$

En effet, $\Phi_{Y_n}(u) = \Phi_{X_1}\left(\frac{u}{\sqrt{n}}\right)^n$

$$\Phi_{Y_n}(u) = \left[\cos\left(\frac{u}{\sqrt{n}}\right)\right]^n = \left(1 - \frac{u^2}{2n} + o\left(\frac{u^2}{n}\right)\right)^n = \left(1 - \frac{u^2\left(1+\varepsilon\left(\frac{u^2}{n}\right)\right)}{2n}\right)^n \underset{n\to+\infty}{\longrightarrow} e^{-\frac{u^2}{2}}.$$

Ce résultat très important sera généralisé à la section 5.

10.3.2 Théorème de Paul Lévy (et théorème de compacité faible)

Théorème 4 (Prohorov) :
(a) Soit $\mu_n, n \geq 1$, une suite de probabilités sur $\left(\mathbb{R}^d, \mathcal{B}\left(\mathbb{R}^d\right)\right)$. Alors il existe une mesure μ sur $\left(\mathbb{R}^d, \mathcal{B}\left(\mathbb{R}^d\right)\right)$ avec $\mu\left(\mathbb{R}^d\right) \leq 1$ et une suite extraite $\mu_{\varphi(n)}$ de μ_n vérifiant :

$$\forall f \in \mathcal{C}_0\left(\mathbb{R}^d, \mathbb{R}\right), \int f d\mu_{\varphi(n)} \to \int f d\mu \quad (\textit{convergence dite ``vague''}).$$

(b) Si la suite $(\mu_n)_{n\geq 1}$ est *tendue*, *i.e.* si

$$\forall \varepsilon > 0,\ \exists K_\varepsilon,\ \text{compact},\ \exists n_\varepsilon \in \mathbb{N} \text{ tel que } n \geq n_\varepsilon \Longrightarrow \mu_n({}^cK_\varepsilon) \leq \varepsilon,$$

alors il existe une *probabilité* μ sur $\left(\mathbb{R}^d, \mathcal{B}\left(\mathbb{R}^d\right)\right)$ et une suite extraite $\mu_{\varphi(n)}$ de μ_n telles que $\mu_n \Longrightarrow \mu$.

Démonstration (partielle) : (a) On suppose que $d = 1$. $\forall r \in \mathbf{Q}, (F_n(r))_{n\geq 1}$ est bornée donc, par le procédé diagonal, il existe une sous-suite $\varphi(n) \uparrow +\infty$ telle que :

$$\forall r \in \mathbf{Q}, F_{\varphi(n)}(r) \to F(r).$$

On pose alors : $\forall x \in \mathbb{R}, F(x) := \lim_{r\to x^+} F(r)$. On vérifie (exercice facile) que F ainsi définie est croissante, continue à droite. Soit μ la mesure de Stieltjès relative à F i.e. l'unique mesure sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ vérifiant $\mu(]a,b]) = F(b) - F(a)$. Partant du fait que $\forall r \in \mathbf{Q}, F_{\varphi(n)}(r) \to F(r)$ et de la densité de $\mathbf{Q}$, on établit aisément que la convergence a en fait lieu pour tout $r \in \text{Cont}(F)$ puis, par une légère adaptation de la preuve du

théorème 2 sur les fonctions de répartition que $\mu(O) \leq \underline{\lim}_n \mu_n(O)$ pour tout ouvert O de $\mathbb{R}$. En particulier $\mu(\mathbb{R}) = 1$.

Ensuite, adaptant cette fois le preuve $(ii) \Rightarrow (i)$ du théorème 2, on en déduit que $\int f d\mu_n \to \int f d\mu$ pour toute $f \in \mathcal{C}_K(\mathbb{R}, \mathbb{R})$ (l'adaptation consiste à remplacer $A_p^{(p)} = \mathbb{R}$ par un intervalle compact $[a,b]$ tel que $F(a) = F(a-)$, $F(b) = F(b-)$) et $\text{supp} f = [a,b]$.

L'extension à $\mathcal{C}_0\left(\mathbb{R}^d, \mathbb{R}\right)$ découle de la remarque de la section 1.

(b) Il suffit de montrer que la mesure μ ci-desssus est en fait une probabilité dès que la suite (μ_n) est tendue. Soit $\varepsilon \in]0,1[$ et K_ε un compact vérifiant l'hypothèse. On peut quitte à l'agrandir, supposer que $K_\varepsilon = [-M_\varepsilon, M_\varepsilon]$. On considère une fonction continue φ_ε à valeurs dans $[0,1]$, valant 1 sur $[-M_\varepsilon, M_\varepsilon]$ et 0 sur ${}^c[-(M_\varepsilon+1), M_\varepsilon+1]$. La fonction φ_ε est continue à support compact donc

$$\mu(\mathbb{R}) \geq \int \varphi_\varepsilon d\mu = \lim_n \int \varphi_\varepsilon d\mu_n \geq \overline{\lim}_n \mu_n(K_\varepsilon) \geq 1 - \varepsilon.$$

Finalement $1 \geq \mu(\mathbb{R}) \geq 1 - \varepsilon$ pour tout $\varepsilon \in]0,1[$. #

Exemple : La suite de probabilités $\mu_n := \delta_n$, $n \geq 1$ converge vaguement vers la mesure nulle et n'est évidemment pas tendue.

Théorème 5 (Paul Lévy) :
Soit $X_n, n \geq 1$, une suite de v.a. à valeurs dans $\mathbb{R}^d$ et $\Phi : \mathbb{R}^d \to \mathbb{R}$ une fonction *continue en 0*. Alors :

$$\left(\forall u \in \mathbb{R}^d, \Phi_{X_n}(u) \to \Phi(u)\right) \Rightarrow \left(\exists \text{ une probabilité } \mu \text{ sur } \mathbb{R}^d \text{ telle que } \Phi = \widehat{\mu}\right)$$

(ou encore il existe une v.a. Y telle que $\Phi = \Phi_Y$).

Démonstration : Soit $\mu_n = \mathbb{P}_{X_n}$. D'après le théorème de Prohorov, il existe μ, mesure sur $\mathbb{R}^d$, telle que $\mu\left(\mathbb{R}^d\right) \leq 1$ et $\forall f \in \mathcal{C}_0\left(\mathbb{R}^d, \mathbb{R}\right) \int f d\mu_n \to \int f d\mu$. D'autre part, comme dans la démonstration du théorème 3, il vient :

$$\forall y \in \mathbb{R}^d, \forall \sigma > 0, \int \varphi_\sigma(y-x)\mu_n(dx) \underset{n\to+\infty}{\to} (2\pi)^{-\frac{d}{2}} \int e^{-i(u,y)}\Phi(u)\varphi_1(\sigma u)\,du$$

et, parallèlement,

$$\int \varphi_\sigma(y-x)\mu_n(dx) \to \int \varphi_\sigma(y-x)\mu(dx) \ \text{ car } \varphi_\sigma(y-.) \in \mathcal{C}_0\left(\mathbb{R}^d, \mathbb{R}\right).$$

Donc

$$\forall y \in \mathbb{R}^d, \forall \sigma > 0,\ \int e^{-i(u,y)}\widehat{\mu}(u)\varphi_1(\sigma u)\,du = \int e^{-i(u,y)}\Phi(u)\varphi_1(\sigma u)\,du$$

où $\widehat{\mu}(u) := \int e^{i(u,x)}\mu(dx)$ est la transformée de Fourier de la mesure finie μ (le fait que l'identité reste vraie dans ce cas est évident). Il vient par injectivité de la transformation de Fourier

$$\widehat{\mu}\varphi_1(\sigma.) = \Phi\varphi_1(\sigma.)\ \lambda_d\text{-}p.p..$$

Soit $u_p \to 0$ telle que $\widehat{\mu}(u_p)\varphi_1(\sigma u_p) = \Phi(u_p)\varphi_1(\sigma u_p)$. Φ et $\widehat{\mu}$ étant continues en 0, $\widehat{\mu}(0) = \Phi(0) = \lim_n \Phi_{X_n}(0) = 1$. Donc μ est une probabilité.

Mézalor, $\mu_n \Rightarrow \mu$ puisque $\int f d\mu_n \to \int f d\mu$ pour toute $f \in \mathcal{C}_K\left(\mathbb{R}^d, \mathbb{R}\right)$.#

<u>Remarque</u> : Le théorème de Paul Lévy est d'un usage essentiellement théorique.

10.4 Lien avec les autres modes de convergence

La convergence en loi est le plus faible des modes de convergence définis jusqu'à présent.

<u>Proposition 4</u> :
(a) Si $X_n \xrightarrow{\mathrm{P}} X$ alors $X_n \xrightarrow{\mathcal{L}} X$.
(b) Si $X_n \xrightarrow{\mathcal{L}} c \in \mathbb{R}^d$ alors $X_n \xrightarrow{\mathrm{P}} c$.

<u>Démonstration</u> :
(a) Soit $f : \mathbb{R}^d \to \mathbb{R}$ continue bornée. f étant continue $f(X_n) \xrightarrow{\mathrm{P}} f(X)$. D'autre part $|f(X_n)| \leq M < +\infty$ donc d'après la proposition du chapitre VIII.2, on en déduit que $f(X_n)$ converge vers $f(X)$ dans L^1, ce qui entraîne :

$$\mathbb{E}(f(X_n)) \to \mathbb{E}(f(X)).$$

(b) Soit $\varphi \in \mathcal{C}(\mathbb{R}_+, [0,1])$, $\varphi \equiv 0$ sur $\left[0, \frac{1}{2}\right]$, $\varphi \equiv 1$ sur $[1, +\infty]$.

Pour tout $\varepsilon > 0$, $\mathbf{1}_{\{\|x\| \geq \varepsilon\}} \leq \varphi\left(\frac{\|x\|}{\varepsilon}\right)$, et $x \mapsto \varphi\left(\frac{\|x-c\|}{\varepsilon}\right) \in \mathcal{C}_b(\mathbb{R}^d, \mathbb{R})$ donc

$$\mathbb{P}(\|X_n - c\| \geq \varepsilon) \leq \mathbb{E}\left(\varphi\left(\frac{\|X_n - c\|}{\varepsilon}\right)\right) \to \mathbb{E}\left(\varphi\left(\frac{\|c - c\|}{\varepsilon}\right)\right) = 0,$$

donc $\mathbb{P}(\|X_n - c\| > \varepsilon) \underset{n\to+\infty}{\longrightarrow} 0$. #

On peut par ailleurs se demander quel est le lien entre la convergence en loi d'une suite de variables aléatoires et la convergence d'éventuelles densités, s'il en existe. C'est l'objet du théorème suivant :

<u>Théorème 6</u> (de Scheffé) :
Soit (μ_n) une suite de probabilités sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, toutes absolument continues par rapport à une mesure ν. Il existe donc une famille de densités f_n telles que :

$$\forall n \in \mathbb{N},\ \forall A \in \mathcal{B}(\mathbb{R}^d),\ \mu_n(A) = \int_A f_n(x)\nu(dx).$$

S'il existe une application $g : \mathbb{R}^d \to \mathbb{R}_+$ mesurable telle que $\int g d\nu = 1$ et si $f_n \underset{n\to+\infty}{\to} g$, ν-$p.s.$, alors μ_n converge étroitement vers la probabilité $g d\nu$ de densité g par rapport à ν.

<u>Démonstration</u> : Soit $A \in \mathcal{B}(\mathbb{R}^d)$,

$$\left|\mu_n(A) - \int_A g d\nu\right| = \left|\int_A (f_n - g)\, d\nu\right| \leq \|g - f_n\|_{L^1(\nu)}$$

or $\int (g - f_n) d\nu = 1 - 1 = 0 = \int (g - f_n)^+ d\nu - \int (g - f_n)^- d\nu$
d'où $\int |g - f_n| d\nu = \int (g - f_n)^+ d\nu + \int (g - f_n)^- d\nu = 2\int (g - f_n)^+ d\nu$.
Donc comme $0 \leq (g - f_n)^+ \leq g \in L^1(\nu)$ (les fonctions sont positives) et :

$$(g - f_n)^+ \underset{n\longrightarrow+\infty}{\overset{\nu\text{-}p.s.}{\to}} 0,$$

on peut appliquer le théorème de convergence dominée, qui nous donne :

$$\int (g - f_n)^+ d\nu \underset{n\to+\infty}{\to} 0.$$

On en conclut : $\|g - f_n\|_{L^1(\nu)} \underset{n\to+\infty}{\to} 0$.
En résumé, on a prouvé :

$$\sup_{A \in \mathcal{B}(\mathbb{R}^d)} \left|\mu_n(A) - \int_A g d\nu\right| \underset{n\to+\infty}{\to} 0$$

ce qui est bien plus fort que le résultat annoncé. #

<u>Contre-exemple</u> : La réciproque est fausse. Considérons en effet les densités $f_n(x) = (1 - \cos(2\pi n x))\mathbf{1}_{[0,1]}(x)$. Elles ne convergent en aucun point de $]0,1[$. Néanmoins, la suite des fonctions de répartition de μ_n, F_n,

$$F_n(x) = (x - \frac{\sin 2\pi n x}{2\pi n})\mathbf{1}_{[0,1]}(x) + \mathbf{1}_{]1,+\infty[}(x),$$

tend, quand n tend vers $+\infty$, vers $F(x) = x\mathbf{1}_{[0,1]}(x) + \mathbf{1}_{]1,\infty[}(x)$. En d'autres termes, μ_n converge étroitement vers la loi uniforme sur $[0,1]$.

10.5 Théorème Central-limite

Le théorème de la limite centrale fournit une vitesse de convergence dans la loi des grands nombres sous réserve que les v.a. soient indépendantes, de même loi et de carré intégrable.

Théorème 7 :
Soit $(X_n)_{n\geq 1}$ une suite de v.a. indépendantes à valeurs dans $\mathbb{R}^d$, de même loi. On suppose que $\mathbb{E}\|X_1\|^2 < +\infty$ et l'on note $m := \mathbb{E}(X_1)$ et $\Sigma^{(2)} := \mathrm{Var}(X)$. Alors :

$$\sqrt{n}\left(\frac{X_1+...+X_n}{n}-m\right) = \frac{X_1+...+X_n-nm}{\sqrt{n}} \xrightarrow{\mathcal{L}} \mathcal{N}\left(0;\Sigma^{(2)}\right).$$

Démonstration : Soit $\Phi := \Phi_{X_1-\mathbb{E}X_1}$. D'après les hypothèses de l'énoncé et les résultats sur les fonctions caractéristiques (cf. VI.5), Φ est deux fois dérivable en 0. Par suite

$$\Phi(u) = 1 - \frac{1}{2}\,{}^t u\Sigma^{(2)}u + \|u\|^2\varepsilon(u)$$

(formule de Taylor-Young). La fonction caractéristique de $X_1+...+X_n-nm$ est $\Phi(u)^n$ et donc celle de $\dfrac{X_1+...+X_n-nm}{\sqrt{n}}$ est $\Phi\left(\dfrac{u}{\sqrt{n}}\right)^n$.

D'où, pour tout $u \in \mathbb{R}^d$,

$$\Phi\left(\frac{u}{\sqrt{n}}\right)^n = \left(1 - \frac{{}^t u\Sigma^{(2)}u + \|u\|^2\varepsilon\left(\frac{u}{\sqrt{n}}\right)}{2n}\right)^n.$$

La convergence uniforme sur les compacts de $\mathbf{C}$ de $\left(1+\frac{z}{n}\right)^n$ vers e^z (cf. Appendice) entraîne que $\Phi\left(\dfrac{u}{\sqrt{n}}\right)^n \to \exp\left(-\dfrac{{}^t u\Sigma^{(2)}u}{2}\right) = \Phi_{\mathcal{N}(0;\Sigma^{(2)})}(u)$. #

Corollaire 2 :
Soit $X_n, n\geq 1$, une suite de v.a.r. de même loi, indépendantes. $\mathbb{E}(X_1) = m \in \mathbb{R}$ et $\sigma^2 = Var(X_1) < +\infty$. Alors, si $\sigma > 0, \forall a, b \in \overline{\mathbb{R}}$

$$\mathbb{P}\left(\frac{X_1+...+X_n-nm}{\sigma\sqrt{n}} \in [a,b]\right) \to \mathbb{P}(\mathcal{N}(0,1)\in[a,b]) = \int_a^b e^{-\frac{u^2}{2}}\frac{du}{\sqrt{2\pi}}.$$

Démonstration : Ceci découle du théorème porte-manteau (iv) puisque la loi $\mathcal{N}(0,1)$ a une densité par rapport à la mesure de Lebesgue, donc ne charge pas $\partial[a,b] = \{a,b\}$. #

Ce dernier résultat est très utilisé en Statistique ; il est à la base des techniques de construction des intervalles de confiance.

Exemple : Si $X_n, n \geq 1$, est une suite de v.a. indépendantes de loi $\mathcal{E}xp(1)$ alors, pour tout $n \geq 1, X_1+\cdots+X_n \overset{\mathcal{L}}{\sim} \gamma(n,1)$. D'où, d'après le corollaire précédent ($\mathbb{E}(X_1) = Var(X_1) = 1$),

$$\mathbb{P}(X_1 = \cdots + X_n - n \leq 0) = \mathbb{P}\left(\frac{X_1+...+X_n-n}{\sqrt{n}} \leq 0\right) \underset{n\to+\infty}{\longrightarrow} \mathbb{P}(\mathcal{N}(0,1)<0) = \frac{1}{2}$$

d'une part et vaut $\displaystyle\int_0^n e^{-x}\frac{x^{n-1}}{(n-1)!}dx$ d'autre part. Donc $\displaystyle\int_0^n e^{-x}x^{n-1}dx \sim \frac{(n-1)!}{2}$.

Appendice : Convergence uniforme sur les compacts de $\mathbf{C}$ de $\left(1+\dfrac{z}{n}\right)^n$ vers e^z.

$$e^z - \left(1+\frac{z}{n}\right)^n = \sum_{k=0}^{n}\underbrace{\left(\frac{1}{k!}-\frac{C_n^k}{n^k}\right)}_{\geq 0}z^k + \sum_{k=n+1}^{+\infty}\frac{z^k}{k!}.\text{ Donc :}$$

$$\left|e^z - \left(1+\frac{z}{n}\right)^n\right| \leq \sum_{k=0}^{n}\left(\frac{1}{k!}-\frac{C_n^k}{n^k}\right)|z|^k + \sum_{k=n+1}^{+\infty}\frac{|z|^k}{k!}.$$

D'où, si $|z| \leq M$,

$$\begin{aligned}\left|e^z - \left(1+\frac{z}{n}\right)^n\right| &\leq \sum_{k=0}^{n}\left(\frac{1}{k!}-\frac{C_n^k}{n^k}\right)M^k + \sum_{k=n+1}^{+\infty}\frac{M^k}{k!} = e^M - \left(1+\frac{M}{n}\right)^n \\ &\leq e^M - e^{n\ln\left(1+\frac{M}{n}\right)} \to 0\end{aligned}$$

quand $n \to +\infty$, uniformément sur $B(0;M)$. #

Chapitre 11

Espérance conditionnelle. Loi conditionnelle

11.1 Préliminaires

Le but de ce chapitre est de généraliser les notions de probabilité et de loi conditionnelles élémentaires introduites au chapitre I. En effet ces notions, appliquées à des variables aléatoires, nécessitaient de se placer dans un cadre résolument discret.

11.1.1 Loi conditionnelle : l'approche élémentaire

Rappelons que, sur un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$, si $B \in \mathcal{A}$ et $\mathbb{P}(B) \neq 0$, on définit la probabilité conditionnelle de (tout événement) $A \in \mathcal{A}$ sachant B par :

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \text{ [on note parfois } \mathbb{P}^B(A)].$$

Il est immédiat que $\mathbb{P}(.|B)$ est une probabilité sur $(\Omega, \mathcal{A})$, et qu'à ce titre, on peut définir pour toute v.a. $X : (\Omega, \mathcal{A}, \mathbb{P}(.|B)) \to (E, \mathcal{E})$ la loi conditionnelle de X sachant B comme la probabilité image (sur $(E, \mathcal{E})$) de $\mathbb{P}(.|B)$. On note généralement cette loi conditionnelle $\mathbb{P}_X(.|B)$. Il vient alors :

$$\forall C \in \mathcal{E}, \quad \mathbb{P}_X(C|B) := \frac{\mathbb{P}(\{X \in C\} \cap B)}{\mathbb{P}(B)}$$

et plus généralement, pour toute fonction borélienne $f : (E, \mathcal{E}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ bornée ou positive,

$$\int_E f(x)\mathbb{P}_X(dx|B) = \frac{\int_B f(X)d\mathbb{P}}{\mathbb{P}(B)} = \frac{\mathbb{E}(\mathbf{1}_B f(X))}{\mathbb{P}(B)}.$$

Remarques : • Notons au passage que la loi conditionnelle ne peut en général pas s'exprimer à partir de la loi de X.

• Si X est une v.a. réelle intégrable, *i.e.* telle que $\mathbb{E}|X| < +\infty$, alors $X \in L^1(\mathbb{P}(.|B))$ et

$$\mathbb{E}(X|B) = \int_B X \frac{d\mathbb{P}}{\mathbb{P}(B)} = \frac{\mathbb{E}(\mathbf{1}_B X)}{\mathbb{P}(B)}.$$

En fait, c'est cette dernière notion, dite "espérance conditionnelle", que nous allons privilégier dans un premier temps car elle est plus maniable, dans un cadre général, que celle de loi conditionnelle.

11.1.2 Espérance conditionnelle d'une v.a. : l'approche élémentaire

Considérons maintenant, outre la variable aléatoire réelle X (supposée réelle positive ou bornée dans tous ces préliminaires), une v.a. $Y : (\Omega, \mathcal{A}) \to (E, \mathcal{P}(E))$ définie sur le même espace probabilisé, où E désigne fini ou dénombrable. Pour simplifier, supposons également que $\mathbb{P}(\{Y = e\}) > 0$ pour tout $e \in E$.

Il paraît naturel de définir la notion d'espérance conditionnelle de la v.a.r. X sachant Y par :

$$\mathbb{E}(X|Y) := \sum_{e \in E} \mathbb{E}(X|\{Y = e\})\mathbf{1}_{\{Y=e\}} = \sum_{e \in E} \frac{\int_{\{Y=e\}} X d\mathbb{P}}{\mathbb{P}(\{Y = e\})} \mathbf{1}_{\{Y=e\}}.$$

Ainsi définie, on peut considérer $\mathbb{E}(X|Y)$ comme une fonction $\varphi(Y)$ de la v.a. Y : $\varphi(y) = \dfrac{\int_{\{Y=y\}} X d\mathbb{P}}{\mathbb{P}(\{Y = y\})}$, $y \in E$.

Exemple : $\Omega := \{1, 2, \ldots, 6\}, \mathbb{P}(A) := \dfrac{|A|}{6}, X(\omega) := \omega, Y(\omega) = \begin{cases} 1 \text{ si } \omega \text{ impair} \\ 0 \text{ si } \omega \text{ pair} \end{cases}$.

Alors $\mathbb{E}(X|Y) = \begin{cases} 3 \text{ si } Y = 1 \\ 4 \text{ si } Y = 0 \end{cases}$.

Une fois remarqué que $\sigma(Y) = \{\bigcup_{e \in C} \{Y = e\}, C \subset E\}$, il est naturel d'étendre à nouveau la notion d'espérance conditionnelle en conditionnant par une sous-tribu $\mathcal{B}$ de $\mathcal{A}$, au moins lorsque $\mathcal{B}$ est purement atomique. Une sous-tribu $\mathcal{B}$ est dite *atomique* si elle est engendrée par un partition finie ou dénombrable $(B_i)_{i \in I}$ de Ω, constituée d'éléments de $\mathcal{A}$. On vérifie immédiatement qu'alors

$$\mathcal{B} = \{\bigcup_{i \in J} B_i, \quad J \subset I\}$$

et l'on pose, par analogie avec la définition précédente :

$$\mathbb{E}(X \,|\, \mathcal{B}) = \sum_{i \in I, \mathbb{P}(B_i) \neq 0} \frac{\int_{B_i} X d\mathbb{P}}{\mathbb{P}(B_i)} \mathbf{1}_{B_i} = \sum_{i \in I, \mathbb{P}(B_i) \neq 0} \mathbb{E}(X|B_i)\mathbf{1}_{B_i}.$$

Remarque : • Si $\mathcal{B} = \{\emptyset, \Omega\}$, $\mathbb{E}(X \mid \mathcal{B}) \equiv \mathbb{E}(X)$ n'est autre que la variable aléatoire constante égale à l'espérance.

• $\mathcal{B} = \mathcal{A}$ (tribu originelle sur $(\Omega, \mathcal{A}, \mathbb{P})$, alors $\mathbb{E}(X \mid \mathcal{B}) = X$. C'est par exemple le cas lorsque Ω est fini et que $\mathcal{A} = \mathcal{P}(\Omega)$.

A ce stade, il est instructif d'examiner les propriétés de l'objet que nous avons fabriqué.

Proposition : Supposons pour simplifier que I est fini et que $\mathbb{P}(B_i) \neq 0$ pour tout $i \in I$. L'espace vectoriel engendré par les $\mathbf{1}_{B_i}$, $i \in I$ n'est autre que $\mathcal{L}^2(\Omega, \mathcal{B}, \mathbb{P})$ *i.e.*

$$\mathcal{L}^2(\Omega, \mathcal{B}, \mathbb{P}) := \{Z = \sum_{i \in I} \lambda_i \mathbf{1}_{B_i},\ \lambda_i \in \mathbb{R}\}.$$

Ce sous-espace de $\mathcal{L}^2_{\mathbb{R}}(\Omega, \mathcal{A}, \mathbb{P})$, une fois muni du produit scalaire usuel $(U \mid V) := \mathbb{E}(UV)$, est un $\mathbb{R}$-espace de Hilbert (de dimension finie $|I|$). La variable aléatoire $\mathbb{E}(X \mid \mathcal{B})$ vérifie alors que

(a) $\mathbb{E}(X \mid \mathcal{B}) \in \mathcal{L}^2_{\mathbb{R}}(\Omega, \mathcal{B}, \mathbb{P})$,

(b) $\|X - \mathbb{E}(X \mid \mathcal{B})\|_2 = \min\{\|X - Z\|_2,\ Z \in \mathcal{L}^2(\Omega, \mathcal{B}, \mathbb{P})\}$,

(c) pour tout $Z \in \mathcal{L}^2(\Omega, \mathcal{B}\mathbb{P})$, $\mathbb{E}(X\,Z) = \mathbb{E}(\mathbb{E}(X \mid \mathcal{B})\,Z)$.

Démonstration de la proposition : le point (a) est immédiat au vu de le définition. Montrons par exemple le point (b), le point (c) s'établissant de façon analogue.

Soit $(\lambda_i)_{i \in I} \in \mathbb{R}^I$. Comme $B_i \cap B_j = \phi$ si $i \neq j$,

$$\|X - \sum_{i \in I} \lambda_i \mathbf{1}_{B_i}\|_2^2 = \mathbb{E}(X^2) - 2\sum_{i \in I} \lambda_i \mathbb{E}(\mathbf{1}_{B_i} X) + \sum_{i \in I} \lambda_i^2 \mathbb{P}(B_i).$$

La fonction φ de $(\lambda_i)_{i \in I}$ ainsi définie est clairement strictement convexe (les $\mathbb{P}(B_i) \neq 0$) ; elle atteint donc son minimum en le seul $|I|$-uplet annulant son gradient $\nabla\varphi$. Or

$$\forall i \in I, \quad \frac{\partial \varphi}{\partial \lambda_i} = 2\lambda_i \mathbb{P}(B_i) - 2\mathbb{E}(\mathbf{1}_{B_i} X) = 0 \Leftrightarrow \lambda_i = \frac{\mathbb{E}(\mathbf{1}_{B_i} X)}{\mathbb{P}(B_i)}.$$

Donc $\|X - Z\|_2$ est donc bien minimum en $\displaystyle\sum_{i \in I} \frac{\int_{B_i} X\, d\mathbb{P}}{\mathbb{P}(B_i)} \mathbf{1}_{B_i} = \mathbb{E}(X \mid \mathcal{B})$ parmi toutes le varaibles aléatoires $\mathcal{B}$-mesurables.#

L'analogie entre ces propriétés de l'espérance conditionnelle et la définition de la projection orthogonale sur un sous-espace (fermé) d'un espace de Hilbert saute aux yeux comme l'illustre le rappel suivant :

Rappel : Soit $(H, (.|.))$ un $\mathbb{R}$-espace de Hilbert et F un s.e.v. fermé de H. Pour tout $h \in H$, il existe un unique vecteur $\pi_F(h)$ dans F vérifiant l'une des deux assertions équivalentes suivantes :

(i) $\|h - \pi_F(h)\| = \min_{h' \in F} \|h - h'\|$,

(ii) $\forall h' \in F,\ (h|h') = (\pi_F(h)|h')$.

L'espérance conditionnelle $\mathbb{E}(X \mid \mathcal{B})$ apparaît donc *la projection orthogonale de* $\mathcal{L}^2_{\mathbb{R}}(\Omega, \mathcal{A}, \mathbb{P})$ *sur son sous-espace des variables aleátoires* $\mathcal{B}$*-mesurables* $\mathcal{L}^2_{\mathbb{R}}(\Omega, \mathcal{B}, \mathbb{P})$. Sauf que $\ldots \mathcal{L}^2_{\mathbb{R}}(\Omega, \mathcal{A}, \mathbb{P})$ n'est pas un espace de Hilbert puisque $(U, V) \mapsto \mathbb{E}(UV)$ n'est pas un produit scalaire (c'est une forme bilinéaire positive mais non définie dès que l'un des $\mathbb{P}(B_i)$ est non nul). Ce problème technique peut être surmonté.

Cependant, la construction précédente souffre d'autres défauts importants. En effet, elle ne permet pas de conditionner par un événement de probabilité nulle, ce qui rend impossible toute définition de $\mathbb{E}(X|Y)$ par ce biais lorsque Y n'est pas une v.a. purement discrète. Dans certains cas simples, on peut contourner l'obstacle par une astuce d'approximation. Le résultat ainsi obtenu pourra nous servir de repère lors du passage au cas général.

Supposons en effet que le couple $(X, Y) : (\Omega, \mathcal{A}, \mathbb{P}) \to \mathbb{R}^2$ admette une densité $f(x, y)$ (que nous supposerons continue bornée pour simplifier). Il est "tentant" de raisonner comme suit.

$$\begin{aligned}
\mathbb{E}(X|\{Y = y\}) &:= \lim_{\varepsilon \to 0} \mathbb{E}(X|\{Y \in [y, y+\varepsilon]\}) = \lim_{\varepsilon \to 0} \frac{\int X \mathbf{1}_{\{y \le Y \le y+\varepsilon\}} d\mathbb{P}}{\mathbb{P}(\{y \le Y \le y+\varepsilon\})} \\
&= \lim_{\varepsilon \to 0} \frac{\frac{1}{\varepsilon}\int_y^{y+\varepsilon} dv \left(\int x f(x, v) dx\right)}{\frac{1}{\varepsilon}\int_y^{y+\varepsilon} dv \left(\int f(x, v) dx\right)} \\
\mathbb{E}(X|\{Y = y\}) &:= \frac{\int x f(x, y) dx}{\int f(x, y) dx}.
\end{aligned}$$

La dernière égalité apparaît donc comme une définition naturelle.

Si l'on substitue $\varphi(X)$ à X dans les calculs précédents, il vient :

$$\mathbb{E}(\varphi(X)|\{Y = y\}) = \int \varphi(x) \frac{f(x, y)}{\int f(u, y) du} dx \ \ i.e.\ \ \mathcal{L}(X|\{Y = y\}) = \frac{f(x, y)}{\int f(u, y) du} dx.$$

Cette approche soulève de nombreuses questions, notamment quant à la régularité de f (on a utilisé ici que $y \mapsto \int f(x, y) dx$ est continue). En outre, ceci ne résoud en rien la question du conditionnement par des variables "mixtes", voire non discrètes et non absolument continues.

11.2 Espérance conditionnelle sachant une sous-tribu

Soit $\mathcal{B}$ une sous-tribu de $\mathcal{A}$ sur un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$.

Rappels : $\mathcal{L}^2_{\mathbb{R}}(\Omega,\mathcal{A},\mathbb{P}) := \{X : (\Omega,\mathcal{A},\mathbb{P}) \to \mathbb{R}$ t.q. $\int X^2 d\mathbb{P} < +\infty\}$, muni de $\|X\|_2$, est un $\mathbb{R}$-e.v. semi-normé que l'on transforme canoniquement en un $\mathbb{R}$-e.v. normé en le quotientant par la relation d'equivalence :

$$X \equiv Y \Leftrightarrow \|X-Y\|_2 = 0 \Leftrightarrow \mathbb{P}(\{X \neq Y\}) = 0 \quad (\text{en outre } \|X\|_2 = \|Y\|_2).$$

L'ensemble quotient n'est autre que $L^2_{\mathbb{R}}(\Omega,\mathcal{A},\mathbb{P})$.

L'idée de départ est de définir **l'espérance conditionnelle** de X sachant $\mathcal{B}$ en généralisant la proposition 1 des préliminaires *i.e.* comme la **projection orthogonale de** $X\in L^2_{\mathbb{R}}(\Omega,\mathcal{A},\mathbb{P})$ **sur** $L^2_{\mathbb{R}}(\Omega,\mathcal{B},\mathbb{P})$.

Le problème est que, si $\mathcal{L}^2_{\mathbb{R}}(\Omega,\mathcal{B},\mathbb{P}) \subset \mathcal{L}^2_{\mathbb{R}}(\Omega,\mathcal{A},\mathbb{P})$, il n'en est pas (tout à fait) de même des espaces $L^2_{\mathbb{R}}$ correspondants.

En effet, si $X \in \mathcal{L}^2_{\mathbb{R}}(\Omega,\mathcal{B},\mathbb{P})$, considérons les classes d'équivalence :

$$\begin{cases} \dot{X}_{L^2(\Omega,\mathcal{A},\mathbb{P})} := \{Z,\ \text{v.a. } \mathcal{A}\text{-mesurables, t.q. } \mathbb{P}(\{Z \neq X\}) = 0\} \text{ et} \\ \dot{X}_{L^2(\Omega,\mathcal{B},\mathbb{P})} := \{Z,\ \text{v.a. } \mathcal{B}\text{-mesurables, t.q. } \mathbb{P}(\{Z \neq X\}) = 0\}. \end{cases}$$

On a donc toujours $\dot{X}_{L^2(\Omega,\mathcal{B},\mathbb{P})} \subset \dot{X}_{L^2(\Omega,\mathcal{A},\mathbb{P})}$ mais, en général, $\dot{X}_{L^2(\Omega,\mathcal{B},\mathbb{P})} \subsetneq \dot{X}_{L^2(\Omega,\mathcal{A},\mathbb{P})}$.

Remarque : Si $\{N\in \mathcal{A} \,/\, \mathbb{P}(N)=0\} \subset \mathcal{B}$, alors il y a égalité : si Z est $\mathcal{A}$-mesurable et $\mathbb{P}(Z \neq X) = 0$, alors Z est $\mathcal{B}$-mesurable. En effet, si $C \in \mathcal{B}(\mathbb{R})$,

$$\{Z\in C\} = ({}^c\{Z \neq X\} \cap \{X \in C\}) \cup (\{Z\neq X\} \cap \{Z \in C\}) \in \mathcal{B}$$

puisque

- $\{Z \neq X\} \cap \{Z \in C\} \in \mathcal{A}$ et $\mathbb{P}(\{Z\neq X\} \cap \{Z \in C\}) \le \mathbb{P}(\{Z \neq X\}) = 0$ donc $\{Z \neq X\} \cap \{Z\in C\} \in \mathcal{B}$.
- $\{Z \neq X\} \in \mathcal{B}$ pour les mêmes raisons et donc ${}^c\{Z \neq X\} \cap \{X \in C\} \in \mathcal{B}$.

Quoi qu'il en soit, il est toujours possible grâce à l'inclusion des classes d'équivalence, d'identifier $L^2_{\mathbb{R}}(\Omega,\mathcal{B},\mathbb{P})$ à un s.e.v. fermé de $L^2_{\mathbb{R}}(\Omega,\mathcal{A},\mathbb{P})$:

Proposition : $i : \begin{cases} L^2_{\mathbb{R}}(\Omega,\mathcal{B},\mathbb{P}) & \to & L^2_{\mathbb{R}}(\Omega,\mathcal{A},\mathbb{P}) \\ \dot{X}_{L^2(\mathcal{B})} & \mapsto & \dot{X}_{L^2(\mathcal{A})} \end{cases}$ est une isométrie.

En particulier, $i(L^2_{\mathbb{R}}(\Omega,\mathcal{B},\mathbb{P}))$ est complet donc fermé dans $L^2_{\mathbb{R}}(\Omega,\mathcal{A},\mathbb{P})$.

Démonstration : Tout d'abord notons que l'application i est bien définie car $\dot{X}_{L^2(\mathcal{B})} \subset \dot{X}_{L^2(\mathcal{A})}$. Soit $Z \in \dot{X}_{L^2(\mathcal{B})}$. Vu qu'alors $Z \in \dot{X}_{L^2(\mathcal{A})} = i\left(\dot{X}_{L^2(\mathcal{B})}\right)$, il vient :

$$\|i(\dot{X}_{L^2(\Omega,\mathcal{B},\mathbb{P})})\|_2^2 = \|\dot{X}_{L^2(\Omega,\mathcal{A},\mathbb{P})}\|_2^2 = \int Z^2 d\mathbb{P} = \|\dot{X}_{L^2(\Omega,\mathcal{B},\mathbb{P})}\|_2^2. \ \#$$

Avertissement : Dans toute la suite, on identifiera $i(L^2_{\mathbb{R}}(\Omega,\mathcal{B},\mathbb{P}))$ et $L^2_{\mathbb{R}}(\Omega,\mathcal{B},\mathbb{P})$.

11.2.1 Espérance conditionnelle d'une v.a. de carré intégrable

Définition : Soit $X \in L^2(\Omega,\mathcal{A},\mathbb{P})$ et $\mathcal{B}$ une sous-tribu de $\mathcal{A}$. On appelle *espérance conditionnelle* de X sachant $\mathcal{B}$ la projection orthogonale $\pi^{\perp}_{L^2(\Omega,\mathcal{B},\mathbb{P})}(X)$ de X sur $L^2(\Omega,\mathcal{B},\mathbb{P})$. On la note $\mathbb{E}(X\,|\,\mathcal{B})$ (ou aussi $\mathbb{E}^{\mathcal{B}}(X)$).

De la caractérisation de la projection orthogonale, on déduit immédiatement le théorème de caractérisation :

Théorème 1 : Soit $X \in L^2_{\mathbb{R}}(\Omega,\mathcal{A},\mathbb{P})$ et $Y \in L^2_{\mathbb{R}}(\Omega,\mathcal{B},\mathbb{P})$. Les trois assertions suivantes sont équivalentes :

(*i*) $Y = \mathbb{E}(X\,|\,\mathcal{B})$,

(*ii*) $\|X-Y\|_2 = \min_{Z\in L^2(\Omega,\mathcal{B},\mathbb{P})} \|X-Z\|_2$,

(*iii*) $\forall Z\in L^2(\Omega,\mathcal{B},\mathbb{P}),\ \ \mathbb{E}(ZX) = \mathbb{E}(ZY)$.

Exemple : Si $\mathcal{B} = \{\phi,\Omega\}$, alors $\mathbb{E}(X\,|\,\mathcal{B}) = \mathbb{E}(X)$.

De cette caractérisation, on déduit les premières propriétés importantes de l'espérance conditionnelle.

Propriétés : Soit $X \in L^2_{\mathbb{R}}(\Omega,\mathcal{A},\mathbb{P})$.

P1. $\mathbb{E}(.\,|\,\mathcal{B}) : L^2_{\mathbb{R}}(\Omega,\mathcal{A},\mathbb{P}) \to L^2_{\mathbb{R}}(\Omega,\mathcal{B},\mathbb{P})$ est une application linéaire de norme 1 ; en particulier, on a : $\|\mathbb{E}(X\,|\,\mathcal{B})\|_2 \le \|X\|_2$. En outre, il y a égalité si et seulement si $X\in L^2_{\mathbb{R}}(\Omega,\mathcal{B},\mathbb{P})$ auquel cas $\mathbb{E}(X\,|\,\mathcal{B}) = X$.

P2. Si $X\in L^2_{\mathbb{R}}(\Omega,\mathcal{B},\mathbb{P})$, $\ \mathbb{E}(X\,|\,\mathcal{B}) = X$.

P3. Si $X \ge 0$ $\mathbb{P}$-*p.s.*, alors $\mathbb{E}(X\,|\,\mathcal{B}) \ge 0$, $\mathbb{P}$-*p.s.*. En conséquence $\mathbb{E}(\,.\,|\,\mathcal{B})$ est un opérateur croissant sur $L^2_{\mathbb{R}}(\Omega,\mathcal{A},\mathbb{P})$.

P4. $\mathbb{E}(\mathbb{E}(X|\mathcal{B})) = \mathbb{E}(X)$.

P5. $|\mathbb{E}(X\,|\,\mathcal{B})| \le \mathbb{E}(|X|\,|\,\mathcal{B})$ (et donc $\mathbb{E}(|\mathbb{E}(X\,|\,\mathcal{B})|) \le \mathbb{E}(|X|)$).

P6. Si W est $\mathcal{B}$-mesurable et bornée, alors $\mathbb{E}(WX\,|\,\mathcal{B}) = W\mathbb{E}(X\,|\,\mathcal{B})$.

Démonstration : • **P1** et **P2** découlent immédiatement du fait que $\mathbb{E}(.\,|\,\mathcal{B})$ est un projecteur orthogonal.

• **P3** : Soit $B_n := \{\mathbb{E}(X\,|\,\mathcal{B}) < -\frac{1}{n}\}$; $\mathbf{1}_{B_n} \in L^\infty_{\mathbb{R}}(\Omega,\mathcal{B},\mathbb{P}) \subset L^2_{\mathbb{R}}(\Omega,\mathcal{B},\mathbb{P})$, d'où

$$0 \le \mathbb{E}(\mathbf{1}_{B_n}X) = \mathbb{E}(\mathbf{1}_{B_n}\mathbb{E}(X\,|\,\mathcal{B})) \le -\frac{1}{n}\mathbb{P}(B_n).$$

Par suite, pour tout $n\ge 1$, $\mathbb{P}(B_n) = 0$ et partant $\mathbb{P}(\{\mathbb{E}(X\,|\,\mathcal{B}) < 0\}) = 0$ car $\{\mathbb{E}(X\,|\,\mathcal{B}) < 0\} = \bigcup_{n\ge 1}^{\uparrow} B_n$.

• **P4** : Evident une fois noté que $Z \equiv \mathbf{1} \in L^2_{\mathbb{R}}(\Omega, \mathcal{B}, \mathbb{P})$.

• **P5** : On décompose $X := X^+ - X^-$ avec les définitions habituelles.

$$\begin{aligned} |\mathbb{E}(X\,|\,\mathcal{B})| &= |\mathbb{E}(X^+\,|\,\mathcal{B}) - \mathbb{E}(X^-\,|\,\mathcal{B})| \ \text{ par linéarité de } \mathbb{E}(.\,|\,\mathcal{B}),\\ &\le \mathbb{E}(X^+\,|\,\mathcal{B}) + \mathbb{E}(X^-\,|\,\mathcal{B}) \ \text{ d'après } \mathbf{P3},\\ &\le \mathbb{E}(X^+ + X^-\,|\,\mathcal{B}) = \mathbb{E}(|X|\,|\,\mathcal{B}). \end{aligned}$$

• **P6** : Comme W est $\mathcal{B}$-mesurable et $\mathbb{P}$-essentiellement bornée, $W\mathbb{E}(X\,|\,\mathcal{B}) \in L^2_{\mathbb{R}}(\Omega, \mathcal{B}, \mathbb{P})$. Pour tout $Z \in L^2_{\mathbb{R}}(\Omega, \mathcal{B}, \mathbb{P})$, on a aussi $ZW \in L^2_{\mathbb{R}}(\Omega, \mathcal{B}, \mathbb{P})$. D'où

$$\mathbb{E}(ZWX) = \mathbb{E}(ZW\mathbb{E}(X\,|\,\mathcal{B})).$$

On en déduit *via* la caractérisation (iii) du théorème 1 que $\mathbb{E}(WX\,|\,\mathcal{B}) = W\mathbb{E}(X\,|\,\mathcal{B})$ #.

<u>Proposition</u> : Soient $\mathcal{B}$ et $\mathcal{C}$ deux sous-tribus emboîtées de $\mathcal{A}$, *i.e.* vérifiant $\mathcal{C} \subset \mathcal{B} \subset \mathcal{A}$. Alors, pour toute v.a. $X \in L^2_{\mathbb{R}}(\Omega, \mathcal{A}, \mathbb{P})$, on a :

$$\mathbb{E}(X|\mathcal{C}) = \mathbb{E}(\mathbb{E}(X\,|\,\mathcal{B})|\mathcal{C}).$$

<u>Démonstration</u> : Ce résultat n'est autre qu'une traduction en termes d'espérance conditionnelle du théorème de double projection.

On peut cependant à titre d'exercice établir le résultat "directement" à partir de la caractérisation (iii) du théorème 1. #

11.2.2 Espérance conditionnelle d'une v.a. intégrable

Le problème est maintenant d'étendre la notion d'espérance conditionnelle au cas où X est seulement intégrable. Pour ce faire, nous allons nous appuyer sur un théorème de prolongement par densité des applications uniformément continues à valeurs dans un espace métrique complet.

<u>Rappel</u> : Soient (E, d) et (F, δ) deux espaces métriques, (F, δ) étant supposé **complet**.

Soit $\left\{ \begin{array}{l} \bullet\ D \subset E,\ d\text{-dense dans } E \\ \bullet\ \varphi : (D, d) \to (F, \delta),\ (d, \delta)\text{-}\textbf{uniformément} \text{ continue.} \end{array} \right.$

Alors il existe une unique application $\Phi : (E, d) \to (F, \delta)$ vérifiant :

$$\Phi \text{ est } (d, \delta)\text{-uniformément continue et } \Phi_{|D} = \varphi.$$

Ce résultat s'applique notamment lorsque E et F sont des e.v.n. , $(F, \|.\|_F)$ étant un espace de Banach et φ une application linéaire continue (donc uniformément continue) d'un sous-espace vectoriel D de E dense dans F. Le prolongement ϕ est alors linéaire et $|\!|\!|\Phi|\!|\!| = |\!|\!|\varphi|\!|\!|$.

<u>Théorème 2</u> : L'opérateur linéaire d'espérance conditionnelle $\mathbb{E}(.\,|\,\mathcal{B})$ de $L^2_{\mathbb{R}}(\Omega, \mathcal{A}, \mathbb{P})$ dans $L^2_{\mathbb{R}}(\Omega, \mathcal{B}, \mathbb{P})$ se prolonge de façon unique en un opérateur de $L^1_{\mathbb{R}}(\Omega, \mathcal{A}, \mathbb{R})$ dans $L^1_{\mathbb{R}}(\Omega, \mathcal{B}, \mathbb{P})$, de norme 1.

<u>Démonstration</u> : **Étape 1** : $L^2_{\mathbb{R}}(\Omega, \mathcal{A}, \mathbb{P})$ est $\|.\|_1$-dense dans $L^1_{\mathbb{R}}(\Omega, \mathcal{A}, \mathbb{P})$ (et il en est de même des espaces $L^p_{\mathbb{R}}(\Omega, \mathcal{B}, \mathbb{P})$, $p \ge 2$!).
Ceci est immédiat : si $X \in L^1_{\mathbb{R}}(\Omega, \mathcal{A}, \mathbb{P})$, pour tout $n \ge 1$, $X_n := X\mathbf{1}_{\{|X| \le n\}} \in L^\infty_{\mathbb{R}}(\Omega, \mathcal{A}, \mathbb{P}) \subset L^2_{\mathbb{R}}(\Omega, \mathcal{A}, \mathbb{P})$ et

$$\|X - X_n\|_1 = \int |X|\mathbf{1}_{\{|X| \ge n\}} d\mathbb{P} \underset{n\to+\infty}{\longrightarrow} 0 \quad \text{par convergence dominée.}$$

Étape 2 : L'opérateur $\mathbb{E}(.\,|\,\mathcal{B}) : (L^2_{\mathbb{R}}(\Omega, \mathcal{A}, \mathbb{P}), \|.\|_1) \longrightarrow (L^1_{\mathbb{R}}(\Omega, \mathcal{B}, \mathbb{P}), \|.\|_1)$ est continu. En effet, si $X \in L^2_{\mathbb{R}}(\Omega, \mathcal{A}, \mathbb{P})$, la propriété **P5** entraîne :

$$\|\mathbb{E}(X\,|\,\mathcal{B})\|_1 = \mathbb{E}(|\mathbb{E}(X\,|\,\mathcal{B})|) \le \mathbb{E}|X| = \|X\|_1.$$

Donc $\mathbb{E}(.\,|\,\mathcal{B})$ est un opérateur linéaire continu de norme 1 (car si $X \in L^2(\Omega, \mathcal{B}, \mathbb{P})$, $\mathbb{E}(X\,|\,\mathcal{B}) = X$).
On conclut par le théorème de prolongement rappelé ci-avant. #

Comme sous-produit du théorème, on obtient une définition "pratique" de l'espérance conditionnelle dans L^1 puisque si $X \in L^1(\Omega, \mathcal{A}, \mathbb{P})$,

$$X_n := X\mathbf{1}_{\{|X| \le n\}} \overset{L^1}{\to} X \text{ entraîne } \mathbb{E}(X\,|\,\mathcal{B}) = L^1\text{-}\lim_n \mathbb{E}(X_n\,|\,\mathcal{B}).$$

Nous allons voir que l'on peut "oublier" cette définition par prolongement d'un opérateur défini sur L^2, en étendant les caractérisations du théorème 1 :

<u>Théorème 3</u> : Soit $X \in L^1_{\mathbb{R}}(\Omega, \mathcal{A}, \mathbb{P})$ et $Y \in L^1_{\mathbb{R}}(\Omega, \mathcal{B}, \mathbb{P})$. Les trois assertions suivantes sont équivalentes :
(i) $Y = \mathbb{E}(X\,|\,\mathcal{B})$,
(ii) $\forall B \in \mathcal{B}$, $\mathbb{E}(\mathbf{1}_B Y) = \mathbb{E}(\mathbf{1}_B X)$,
(iii) $\forall Z \in L^\infty_{\mathbb{R}}(\Omega, \mathcal{B}, \mathbb{P})$, $\mathbb{E}(ZY) = \mathbb{E}(ZX)$.

<u>Démonstration</u> : On considère la suite $X_n = X\mathbf{1}_{\{|X| \le n\}}$, $n \ge 1$.
$(i) \Rightarrow (ii)$: Soit $B \in \mathcal{B}$, $\mathbf{1}_B \in L^2(\Omega, \mathcal{B}, \mathbb{P})$, donc, pour tout $n \ge 1$,

$$\mathbb{E}(\mathbf{1}_B X_n) = \mathbb{E}(\mathbf{1}_B \mathbb{E}(X_n\,|\,\mathcal{B})).$$

Or $|\mathbb{E}(\mathbf{1}_B X_n) - \mathbb{E}(\mathbf{1}_B X)| \le \mathbb{E}(\mathbf{1}_B |X_n - X|) \le \|X_n - X\|_1 \underset{n\to+\infty}{\longrightarrow} 0$, et
$|\mathbb{E}(\mathbf{1}_B \mathbb{E}(X_n|\mathcal{B})) - \mathbb{E}(\mathbf{1}_B \mathbb{E}(X|\mathcal{B}))| \le \|\mathbb{E}(X_n\,|\,\mathcal{B}) - \mathbb{E}(X\,|\,\mathcal{B})\|_1 \underset{n\to+\infty}{\longrightarrow} 0$.

D'où $\mathbb{E}(\mathbf{1}_B X) = \mathbb{E}(\mathbf{1}_B \mathbb{E}(X \mid \mathcal{B}))$.

$(ii) \Rightarrow (iii)$: Il suffit d'établir l'égalité pour toute v.a. $Z \in L^\infty_{\mathbb{R}}(\Omega, \mathcal{B}, \mathbb{P})$ positive. D'après le lemme fondamental d'approximation, il existe une suite de fonctions étagées positives, $\mathcal{B}$-mesurables $(Z_p)_{p\geq 1}$, vérifiant $Z_p \uparrow Z$ uniformément quand $p \to +\infty$. D'après (ii), $\mathbb{E}(Z_p Y) = \mathbb{E}(Z_p X)$ pour tout $p \geq 1$. D'autre part

$$|\mathbb{E}(Z_p Y) - \mathbb{E}(ZY)| \leq \|Z_p - Z\|_\infty \|Y\|_1 \text{ et } |\mathbb{E}(Z_p X) - \mathbb{E}(ZX)| \leq \|Z_p - Z\|_\infty \|X\|_1.$$

$(iii) \Rightarrow (i)$ D'après $(i) \Rightarrow (iii)$, on a aussi, $\mathbb{E}(Z\mathbb{E}(X \mid \mathcal{B})) = \mathbb{E}(XZ)$ donc, pour tout $Z \in L^\infty_{\mathbb{R}}(\Omega, \mathcal{B}, \mathbb{P})$,

$$\mathbb{E}(Z(\mathbb{E}(X \mid \mathcal{B}) - Y)) = 0.$$

En particulier, il vient, pour $Z := \mathrm{sign}(\mathbb{E}(X \mid \mathcal{B}) - Y)$, (où $\mathrm{sign}(x) = 1$ si $x > 0$, $\mathrm{sign}(x) = -1$ si $x < 0$, $\mathrm{sign}(0) = 0$) :

$$\mathbb{E}(|\mathbb{E}(X \mid \mathcal{B}) - Y|) = 0 \; i.e. \; Y = \mathbb{E}(X \mid \mathcal{B}) \; \mathbb{P}\text{-}p.s..\#$$

Remarque : Contrairement aux apparences, il est souvent plus commode d'utiliser la caractérisation (iii) pour calculer $\mathbb{E}(X \mid \mathcal{B})$.

Les propriétés **P1-P6** et le théorème de "double-projection" s'étendent immédiatement au cadre L^1 (les preuves sont identiques).

Propriétés :

$\tilde{\mathbf{P}}$**1.** $\mathbb{E}(\,.\mid \mathcal{B}) : L^1_{\mathbb{R}}(\Omega, \mathcal{A}, \mathbb{P}) \to L^1_{\mathbb{R}}(\Omega, \mathcal{B}, \mathbb{P})$ est une application linéaire de norme 1 ($i.e.$ $\|\mathbb{E}(X \mid \mathcal{B})\|_1 \leq \|X\|_1$).

Soit $X \in L^1_{\mathbb{R}}(\Omega, \mathcal{A}, \mathbb{P})$.

$\tilde{\mathbf{P}}$**2.** Si $X \in L^1_{\mathbb{R}}(\Omega, \mathcal{B}, \mathbb{P})$, $\mathbb{E}(X \mid \mathcal{B}) = X$, $\mathbb{P}$-$p.s.$.

$\tilde{\mathbf{P}}$**3.** Si $X \geq 0$, $\mathbb{P}$-$p.s.$, alors $\mathbb{E}(X \mid \mathcal{B}) \geq 0$, $\mathbb{P}$-$p.s.$. En conséquence $\mathbb{E}(\,.\mid \mathcal{B})$ est un opérateur croissant.

$\tilde{\mathbf{P}}$**4.** $\mathbb{E}(\mathbb{E}(X|\mathcal{B})) = \mathbb{E}(X)$.

$\tilde{\mathbf{P}}$**5.** $|\mathbb{E}(X \mid \mathcal{B})| \leq \mathbb{E}(|X| \mid \mathcal{B})$,

$\tilde{\mathbf{P}}$**6.** Si W est $\mathcal{B}$-mesurable et bornée, alors $\mathbb{E}(WX \mid \mathcal{B}) = W\mathbb{E}(X \mid \mathcal{B})$.

$\tilde{\mathbf{P}}$**6'.** Si $W \in L^2_{\mathbb{R}}(\Omega, \mathcal{B}, \mathbb{P})$ et $X \in L^2_{\mathbb{R}}(\Omega, \mathcal{A}, \mathbb{P})$, alors $WX \in L^1_{\mathbb{R}}(\Omega, \mathcal{A}, \mathbb{P})$ et

$$\mathbb{E}(WX \mid \mathcal{B}) = W\mathbb{E}(X \mid \mathcal{B}).$$

$\tilde{\mathbf{P}}$**7.** Si $\mathcal{C} \subset \mathcal{B} \subset \mathcal{A}$ sont des sous-tribus et $X \in L^1_{\mathbb{R}}(\Omega, \mathcal{A}, \mathbb{P})$, on a :

$$\mathbb{E}(X|\mathcal{C}) = \mathbb{E}(\mathbb{E}(X \mid \mathcal{B})|\mathcal{C}).$$

De plus, on peut établir une inégalité de Jensen conditionnelle, souvent utile :

Proposition (inégalité de Jensen conditionnelle) : Soit $X \in L^1_{\mathbb{R}}(\Omega, \mathcal{A}, \mathbb{P})$ et φ une fonction convexe telle que $\varphi(X) \in L^1_{\mathbb{R}}(\Omega, \mathcal{A}, \mathbb{P})$, alors :

$$\varphi(\mathbb{E}\,(X \mid \mathcal{B})) \leq \mathbb{E}(\varphi(X) \mid \mathcal{B}), \; \mathbb{P}\text{-}p.s..$$

Démonstration : Comme dans le cas non conditionnel, on utilise la caractérisation des fonctions convexes : φ convexe ssi $\varphi = \sup\{g,\; g \text{ affine}, g \leq \varphi\}$. Or, les fonctions affines sont de la forme $g_{a,b}(x) = ax + b, a, b \in \mathbb{R}$. Il est donc clair que

$$\varphi = \sup\{g,\; g_{a,b} \leq \varphi,\; a,\, b \in \mathbf{Q}\}.$$

Par suite, si $a,\, b \in \mathbf{Q}$,

$$g_{a,b}\,(\mathbb{E}(X \mid \mathcal{B})) = a\mathbb{E}(X \mid \mathcal{B}) + b = \mathbb{E}(aX + b \mid \mathcal{B}) = \mathbb{E}(g_{a,b}(X) \mid \mathcal{B}).$$

Comme $g_{a,b} \leq \varphi$, $g_{a,b}(\mathbb{E}(X \mid \mathcal{B})) \leq \mathbb{E}(\varphi(X) \mid \mathcal{B})$ $p.s.$. En passant au supremum sur les fonctions affines, $g_{a,b}$, $g_{a,b} \leq \varphi$, $a\, b \in \mathbf{Q}^2$ (dénombrable car $\mathbf{Q}^2$ est dénombrable), on obtient l'inégalité annoncée. #

Aplications : (a) Pour tout $p \geq 1$, $\mathbb{E}(\,.\mid \mathcal{B})$ est un opérateur linéaire continu de $L^p_{\mathbb{K}}(\mathcal{A}, \mathbb{P})$ dans $L^p_{\mathbb{K}}(\mathcal{A}, \mathbb{P})$ de norme 1. En effet, pour toute v.a. $X \in L^p_{\mathbb{K}}(\mathcal{A}, \mathbb{P})$,

$$\mathbb{E}\,(|\mathbb{E}(X \mid \mathcal{B})|^p) \leq \mathbb{E}\,(\mathbb{E}(|X|^p \mid \mathcal{B})) = \mathbb{E}|X|^p \quad (\text{égalité si } X \in L^p_{\mathbb{K}}(\mathcal{B}, \mathbb{P})).$$

(b) Les propriétés **P6** et **P6'** s'étendent au cas où $X \in L^p_{\mathbb{R}}(\Omega, \mathcal{A}, \mathbb{P})$ et $W \in L^q_{\mathbb{R}}(\Omega, \mathcal{B}, \mathbb{P})$, $p,\, q \geq 1$ at $1/p + 1/q = 1$.

Une parenthèse pour prendre du recul : Pour les variables intégrables, l'espérance conditionnelle n'est *a priori* plus la meilleure approximation au sens de la norme L^1 de X par une variable aléatoire $\mathcal{B}$-mesurable. Pour s'en convaincre il suffit de considérer le cas où $\mathcal{B} = \{\emptyset, \Omega\}$ (tribu triviale) pour lequel l'espérance conditionnelle coïncide avec l'espérance standard. En effet, dans ce cas l'application $a \mapsto \mathbb{E}|X - a|$ est une fonction convexe tendant vers l'infini quand $|a| \to +\infty$ qui atteint donc un minimum. Si la loi de X n'a pas d'atome, on montre que ce ou ces minima sont en fait la ou les *médiane(s)* de la loi de X ($i.e.$ des racines de l'équation $F_X(a) = 1/2$ où F_X désigne la fonction de répartition de la loi de X). Or, sauf cas particulier, médiane et espérance ne coïncident pas.

Si l'on privilégie l'espérance conditionnelle, y compris dans un cadre L^1 ou plus généralement L^p, c'est simplement qu'elle possède une propriété inestimable : la linéarité.

Pour autant que perd-t-on à privilégier ainsi l'espérance conditionnelle ? Soit $p \geq 1$, $X \in L^p_{\mathbb{R}}(\Omega, \mathcal{A}, \mathbb{P})$. Pour toute variable aléatoire et $Z \in L^p_{\mathbb{R}}(\Omega, \mathcal{B}, \mathbb{P})$, on a

$$\begin{aligned} \|X - \mathbb{E}(X \mid \mathcal{B})\|_p &\leq \|X - Z\|_p + \|Z - \mathbb{E}(X \mid \mathcal{B})\|_p \\ &= \|X - Z\|_p + \|\mathbb{E}(Z - X \mid \mathcal{B})\|_p \\ &\leq \|X - Z\|_p + \|Z - X\|_p \end{aligned}$$

d'après l'application (a) ci-avant. D'où

$$\|X - \mathbb{E}(X \mid \mathcal{B})\|_p \leq 2 \inf \left\{ \|X - Z\|_p, \ Z \in L^p_{\mathbb{R}}(\Omega, \mathcal{B}, \mathbb{P}) \right\}$$

ce qui signifie qu'en adoptant l'espérance conditionnelle dans un cadre L^p $(p \neq 2)$ on perd au pire un facteur 2 par rapport à l'éventuelle approximation optimale.

Exemple 1 : Si $\mathcal{B} = \{ \bigcup_{j \in J} B_j, J \subset I \}$, $(B_i)_{i \in I}$ *partition* finie ou dénombrable et $\mathcal{A}$-mesurable de Ω. Alors si $X \in L^1_{\mathbb{R}}(\Omega, \mathcal{A}, \mathbb{P})$,

$$\mathbb{E}(X \mid \mathcal{B}) = \sum_{i \in I, \mathbb{P}(B_i) \neq 0} \frac{\int_{B_i} X d\mathbb{P}}{\mathbb{P}(B_i)} \mathbf{1}_{B_i}.$$

Il est immédiat qu'une v.a.r. est $\mathcal{B}$-mesurable ssi il existe $(\lambda_i)_{i \in I}$ telle que $Y = \sum_{i \in I} \lambda_i \mathbf{1}_{B_i}$. En utilisant alors la caractérisation (ii) avec $B = B_i$, on obtient

$$\lambda_i \mathbb{P}(B_i) = \mathbb{E}(Y \mathbf{1}_{B_i}) = \mathbb{E}(X \mathbf{1}_{B_i}),$$

d'où, dès que $\mathbb{P}(B_i) \neq 0$, $\lambda_i = \dfrac{\int_{B_i} X d\mathbb{P}}{\mathbb{P}(B_i)}$. Sinon, on pose par exemple $\lambda_i = 0$. On vérifie que la formule obtenue convient. On retrouve ici la définition intuitive proposée dans l'introduction.

11.2.3 Espérance conditionnelle d'une v.a. positive

Si X est une v.a.r. **positive** sur $(\Omega, \mathcal{A}, \mathbb{P})$, on peut toujours définir son espérance conditionnelle $\mathbb{E}(X \mid \mathcal{B})$ en posant :

$$\mathbb{E}(X \mid \mathcal{B}) = \mathbb{P}\text{-}p.s. \lim_n \uparrow \mathbb{E}(X \mathbf{1}_{\{X \wedge n\}} \mid \mathcal{B}) \in \overline{\mathbb{R}}_+.$$

On vérifie sans peine qu'ainsi définie $\mathbb{E}(X \mid \mathcal{B})$ vérifie les caractérisations du théorème 3, sous réserve de considérer des v.a. test $Z \geq 0$ dans (iii). On s'appuie pour ce faire sur le théorème de Beppo Levi. Enfin, on a toujours : $\mathbb{E}(\mathbb{E}(X \mid \mathcal{B})) = \mathbb{E}(X) \leq +\infty$, $\mathbb{E}(X \mid \mathcal{B}) = X$ ssi X est $\mathcal{B}$-mesurable et la croissance de l'espérance conditionnelle. Plus généralement les propriétés **P1-P7** s'etendent à ce cadre dès qu'ils y ont un sens.

11.2.4 Théorèmes limites conditionnels

On peut obtenir aussi un théorème de convergence dominée conditionnel qu'il est intéressant de signaler, en procédant exactement comme dans le cadre non conditionnel.

Proposition : Convergence monotone conditionnelle.

Soit $(X_n)_{n \geq 1}$ une suite de v.a. de $L^1_{\mathbb{R}}(\Omega, \mathcal{A}, \mathbb{P})$ telle que :

$$\begin{cases} \forall n \geq 1, \ X_n \geq 0 \ \mathbb{P}\text{-}p.s., \\ \forall n \geq m, \ X_n \geq X_m \ \mathbb{P}\text{-}p.s.. \end{cases}$$

Alors $\lim_{n \to +\infty} \mathbb{E}(X_n \mid \mathcal{B}) = \mathbb{E}(\lim_{n \to +\infty} X_n \mid \mathcal{B})$, $\mathbb{P}$-$p.s.$

Démonstration : On remarque déjà que $\mathbb{E}(X_n \mid \mathcal{B})$ est croissante positive $p.s.$, donc converge $p.s.$ dans $\overline{\mathbb{R}}$. Notons U sa limite. U est $\mathcal{B}$-mesurable. De plus,

$$\forall B \in \mathcal{B}, \quad \mathbb{E}(\mathbf{1}_B \mathbb{E}(X_n \mid \mathcal{B})) = \mathbb{E}(\mathbf{1}_B X_n),$$

par définition de l'espérance conditionnelle. Donc, en utilisant le théorème de convergence monotone (non conditionnel) appliqué à cette égalité, on a :

$$\forall B \in \mathcal{B}, \quad \mathbb{E}(\mathbf{1}_B U) = \mathbb{E}(\mathbf{1}_B \lim_{n \to +\infty} X_n).$$

On en déduit le résultat annoncé. #

Proposition : Lemme de Fatou conditionnel.
Soit $(X_n)_{n \geq 1}$ une suite de v.a. positives, alors

$$\mathbb{E}\left(\underline{\lim}_{n \to +\infty} X_n \mid \mathcal{B} \right) \leq \underline{\lim}_{n \to +\infty} \mathbb{E}(X_n \mid \mathcal{B}) \quad \mathbb{P}\text{-}p.s.$$

Démonstration : Soit $Y_n := \inf_{k \geq n} X_k$. Alors $Y_n \leq X_n$ d'où l'on déduit *via* la propriété de croissance $\tilde{\mathbf{P3}}$ étendue aux v.a. positives :

$$\mathbb{E}(Y_n \mid \mathcal{B}) \leq \mathbb{E}(X_n \mid \mathcal{B}) \ \mathbb{P}\text{-}p.s. \text{ et partant } \underline{\lim}_n \mathbb{E}(Y_n \mid \mathcal{B}) \leq \underline{\lim}_n \mathbb{E}(X_n \mid \mathcal{B}).$$

Comme de plus la suite $(Y_n)_{n \geq 1}$ est croissante positive, le théorème de convergence monotone entraîne :

$$\lim_{n \to +\infty} \mathbb{E}(Y_n \mid \mathcal{B}) = \mathbb{E}(\lim_n Y_n \mid \mathcal{B}) = \mathbb{E}\left(\underline{\lim}_{n \to +\infty} X_n \mid \mathcal{B} \right), \ \mathbb{P}\text{-}p.s..$$

En combinant ces résultats, on obtient alors l'inégalité annoncée. #

Proposition : Convergence dominée conditionnelle.
Soit $(X_n)_{n \geq 1}$ une suite de v.a. de $L^1_{\mathbb{R}}(\Omega, \mathcal{A}, \mathbb{P})$ telle que :

- $X_n \underset{n \to +\infty}{\longrightarrow} \tilde{X}$ $\mathbb{P}$-$p.s.$,
- il existe $Y \in L^1_{\mathbb{R}}(\Omega, \mathcal{A}, \mathbb{P})$ telle que $\forall n \geq 1$, $|X_n| \leq Y$ $\mathbb{P}$-p.s..

Alors $\lim_{n\to+\infty} \mathbb{E}(X_n \,|\, \mathcal{B}) = \mathbb{E}(\tilde{X} \,|\, \mathcal{B})$, $\mathbb{P}$-*p.s.*

Démonstration : Quitte à changer X_n en $X_n - \tilde{X}$, encore intégrable, on peut supposer $\tilde{X} = 0$. Alors $Y - |X_n| \geq 0$, et donc en appliquant le lemme de Fatou conditionnel, on a :

$$\mathbb{E}\left(\underline{\lim}_{n\to+\infty} (Y - |X_n|) \,|\, \mathcal{B}\right) \leq \underline{\lim}_{n\to+\infty} \mathbb{E}\left(Y - |X_n| \,|\, \mathcal{B}\right), \ \mathbb{P}\text{-}p.s..$$

Le terme de gauche vaut $\mathbb{E}(Y \,|\, \mathcal{B})$ et celui de droite $\mathbb{E}(Y \,|\, \mathcal{B}) - \overline{\lim}_n \mathbb{E}(|X_n| \,|\, \mathcal{B})$. On a donc $\overline{\lim}_{n\to+\infty} \mathbb{E}(|X_n| \,|\, \mathcal{B}) = 0$. Comme $\mathbb{E}(|X_n| \,|\, \mathcal{B}) \geq |\mathbb{E}(X_n \,|\, \mathcal{B})|$, on en déduit que $\mathbb{E}(X_n \,|\, \mathcal{B}) \to 0$ quand $n \to +\infty$, $\mathbb{P}$-*p.s.*. #

11.3 Espérance conditionnelle et indépendance

Si deux événements A et B de $\mathcal{A}$ sont indépendants, alors $\mathbb{P}(A|B) = \mathbb{P}(A)$. Ceci caractérise même l'indépendance des deux événements. Nous allons généraliser cette remarque dans le cadre de l'espérance conditionnelle.

Définition : Une variable aléatoire $X \in L^1(\Omega, \mathcal{A}, \mathbb{P})$ (à valeurs dans un espace $(E, \mathcal{E})$) et une sous-tribu $\mathcal{B}$ sont indépendante si les tribus $\mathcal{B}$ et $\sigma(X)$ sont indépendantes, soit encore

$$\forall\, B \in \mathcal{B},\ \forall\, f : (E, \mathcal{E}) \to \mathbb{R} \text{ mesurable bornée}, \qquad \mathbb{E}(f(X)\mathbf{1}_B) = \mathbb{E}(f(X))\mathbb{P}(B).$$

La proposition suivante est une conséquence immédiate de la caractérisation (*ii*) de l'espérance conditionnelle dans le théorème 3 de la section 11.2.2.

Proposition : La variable aléatoire X et la sous-tribu $\mathcal{B}$ sont indépendantes si et seuelement si, pour toute $f : (E, \mathcal{E}) \to \mathbb{R}$ mesurable bornée,

$$\mathbb{E}(f(X) \,|\, \mathcal{B}) = \mathbb{E}(f(X)) \quad \mathbb{P}\text{-}p.s..$$

Dans ce cas, l'identité ci-desus reste vraie pour tout fonction $f \in \mathcal{L}^1_{\mathbb{R}}(\mathbb{P}_X)$. En particulier si $X \in \mathcal{L}^1_{\mathbb{R}}(\mathbb{P})$,

$$\mathbb{E}(X \,|\, \mathcal{B}) = \mathbb{E}(X) \quad \mathbb{P}\text{-}p.s..$$

Remarque : On peut aussi avoir $\mathbb{E}(X \,|\, \mathcal{B}) = \mathbb{E}(X)$ sans que X et $\mathcal{B}$ soient indépendantes. Ainsi si $X \overset{\mathcal{L}}{\sim} \mathcal{N}(0,1)$, $\varepsilon \overset{\mathcal{L}}{\sim} B(\{\pm 1\}, \frac{1}{2})$, ε indépendante de X et $\mathcal{B} = \sigma(\varepsilon X)$, il vient : $\mathbb{E}(X \,|\, \mathcal{B}) = 0 = \mathbb{E}(X)$ car si $C \in \mathcal{B}(\mathbb{R})$,

$$\begin{aligned}
\mathbb{E}(\mathbf{1}_{\{\varepsilon X \in C\}}X) &= \mathbb{E}(\mathbf{1}_{\{\varepsilon=1\}\cap\{X\in C\}}X) + \mathbb{E}(\mathbf{1}_{\{\varepsilon=-1\}\cap\{-X\in C\}}X) \\
&= \frac{1}{2}\left(\mathbb{E}(X\mathbf{1}_{\{X\in C\}}) + \mathbb{E}(X\mathbf{1}_{\{-X\in C\}})\right) = \frac{1}{2}\left(\mathbb{E}(X\mathbf{1}_{\{X\in C\}}) - \mathbb{E}(X\mathbf{1}_{\{X\in C\}})\right) \\
&= 0 \ \text{ car } X \overset{\mathcal{L}}{\sim} -X,
\end{aligned}$$

or X et εX ne sont pas indépendantes (cf. chapitre VII.2). Une approche directe consiste à noter que

$$\mathbb{E}((\varepsilon X)^2 \,|\, \mathcal{B}) = \mathbb{E}(X^2 \,|\, \mathcal{B}) = X^2 \neq \mathbb{E}(X^2).$$

Corollaire : Soit $X : (\Omega, \mathcal{A}, \mathbb{P}) \to \mathbb{R}^d$ est indépendante de la sous-tribu $\mathcal{B}$ si et seulement si

$$\forall\, u \in \mathbb{R}^d, \qquad \mathbb{E}\left(e^{i(u,X)} \,|\, \mathcal{B}\right) = \mathbb{E}\left(e^{i(u,X)}\right).$$

Démonstration : Le sens direct découle de la proposition précédente en posant $f_u(x) := e^{i(u,x)}$ (en fait ses parties imaginaires et réelles).

Pour la réciproque, on considère une v.a.r. Z $\mathcal{B}$-mesurable et l'on note que, pour tout $(u, v) \in \mathbb{R}^d \times \mathbb{R}$,

$$\begin{aligned}
\mathbb{E}\left(e^{i(u,X)+ivZ}\right) &= \mathbb{E}\left(\mathbb{E}(e^{i(u,X)} \,|\, \mathcal{B})e^{ivZ}\right) \\
&= \mathbb{E}\left(\mathbb{E}(e^{i(u,X)})e^{ivZ}\right) \\
&= \mathbb{E}(e^{i(u,X)})\mathbb{E}\left(e^{ivZ}\right).
\end{aligned}$$

cette dernière identité entraîne que X et Z sont indépendantes. On conclut en posant $Z = \mathbf{1}_B$, $B \in \mathcal{B}$. #

Proposition : Soient $X : (\Omega, \mathcal{A}, \mathbb{P}) \to (E, \mathcal{E})$, $Y : (\Omega, \mathcal{A}, \mathbb{P}) \to (F, \mathcal{F})$ et $\Phi : (E \times F, \mathcal{E} \otimes \mathcal{F}) \to \mathbb{R}$ bornée ou positive (voire élément de $\mathcal{L}^1(\mathbb{P}_{(X,Y)})$). Soit $\mathcal{B}$ une sous-tribu de $\mathcal{A}$. On suppose que X est indépendante de $\mathcal{B}$ et que Y est $\mathcal{B}$-mesurable. Alors

$$\mathbb{E}\left(\Phi(X, Y) \,|\, \mathcal{B}\right) = \varphi(Y) \quad \mathbb{P}\text{-}p.s.$$

où

$$\varphi(y) = \mathbb{E}\left(\Phi(X, y)\right) = \int_E \Phi(x, y)\mathbb{P}_X(dx).$$

Démonstration : Soit $Z : (\Omega, \mathcal{B}) \to \mathbb{R}$ bornée. On passe sur l'espace d'états *via* le théorème de transfert

$$\mathbb{E}\left(Z\,\Phi(X, Y)\right) = \int_{E\times F\times\mathbb{R}} z\Phi(x, y)\mathbb{P}_{(X,Y,Z)}(dx, dy, dz).$$

Le couple (Y, Z) est $\mathcal{B}$-mesurable donc indépendant de X, partant $\mathbb{P}_{(X,Y,Z)} = \mathbb{P}_X \otimes \mathbb{P}_{(Y,Z)}$ d'où, *via* le théorème de Fubini,

$$\begin{aligned}
\mathbb{E}\left(Z\,\Phi(X, Y)\right) &= \int_{E\times\mathbb{R}} z\left[\int_F \Phi(x, y)\mathbb{P}_X(dx)\right]\mathbb{P}_{(Y,Z)}(dy, dz) \\
&= \int_{E\times\mathbb{R}} z\,\varphi(y)\mathbb{P}_{(Y,Z)}(dy, dz) \\
&= \mathbb{E}\left(Z\,\varphi(Y)\right).
\end{aligned}$$

On conlut en remarquant que la v.a. $\varphi(Y)$ est clairement $\mathcal{B}$-mesurable puisque Y l'est. #

11.4 Espérance conditionnelle sachant une v.a.

11.4.1 Définition et propriétés

Pour conditionner par une variable aléatoire réelle, on s'appuie sur le résultat classique suivant (cf. chapitre II.2.1) :

Rappel : Si Y est une v.a. $Y : (\Omega, \mathcal{A}, \mathbb{P}) \to (E, \mathcal{E})$, alors Z est une v.a.r. $\sigma(Y)$-mesurable ssi il existe $\varphi : (E, \mathcal{E}) \to \mathbb{R}$ mesurable telle que $Z = \varphi(Y)$.

Définition : Soit $X \in L^1_{\mathbb{R}}(\Omega, \mathcal{A}, \mathbb{P})$ et $Y : (\Omega, \mathcal{A}) \to (E, \mathcal{E})$ deux v.a. sur $(\Omega, \mathcal{A})$. On définit l'espérance conditionnelle de X sachant Y par :

$$\mathbb{E}(X|Y) := \mathbb{E}(X|\sigma(Y)) \text{ (on note parfois } \mathbb{E}^Y(X)).$$

Au vu du rappel, $\mathbb{E}(X|Y)$ étant une v.a.r. $\sigma(Y)$-mesurable, il existe $\varphi : (E, \mathcal{E}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ telle que $\mathbb{E}(X|Y) = \varphi(Y)$. Il arrive que la terminologie espérance conditionnelle sachant Y s'applique à la fonction φ elle-même. On écrit d'ailleurs dans cet esprit :

$$\varphi(y) := \mathbb{E}(X|Y = y), \quad y \in \mathbb{R}.$$

Exemple 2 : Si $Y : (\Omega, \mathcal{A}) \to (E, \mathcal{E})$, E dénombrable, $\mathcal{E} := \mathcal{P}(E)$, alors (cf. exemple 1) pour toute v.a. $X \in L^1_{\mathbb{R}}(\Omega, \mathcal{A}, \mathbb{P})$,

$$\mathbb{E}(X|Y) = \sum_{e \in E, \mathbb{P}(Y=e)\neq 0} \frac{\int_{\{Y=e\}} X d\mathbb{P}}{\mathbb{P}(\{Y = e\})} \mathbf{1}_{\{Y=e\}},$$

d'où il vient : $\varphi(e) = \begin{cases} \dfrac{\int_{\{Y=e\}} X d\mathbb{P}}{\mathbb{P}(Y = e)} & \text{si } \mathbb{P}(\{Y = e\}) \neq 0 \\ 0 \text{ par exemple} & \text{si } \mathbb{P}(\{Y = e\}) = 0. \end{cases}$

Exemple 3 : Soit $X \in L^1_{\mathbb{R}}(\Omega, \mathcal{A}, \mathbb{P})$ et $X^+ := \max(X, 0)$. On suppose que $\mathbb{P}(X \leq 0) > 0$. On veut calculer $\mathbb{E}(X|X^+)$. On sait qu'il existe $\varphi : \mathbb{R} \to \mathbb{R}$ borélienne telle que $\mathbb{E}(X|X^+) = \varphi(X^+)$. La question se ramène donc à déterminer φ. Pour la même raison, on peut spécifier les v.a. test Z sous la forme $Z = \psi(X^+)$. Il vient alors :

$$\begin{aligned}
\mathbb{E}(\psi(X^+)X) &= \mathbb{E}(\psi(X^+)X\mathbf{1}_{\{X>0\}}) + \mathbb{E}(\psi(X^+)X\mathbf{1}_{\{X\leq 0\}}) \\
&= \mathbb{E}(\psi(X^+)X^+\mathbf{1}_{\{X>0\}}) + \mathbb{E}(\psi(0)X\mathbf{1}_{\{X\leq 0\}}) \\
&= \mathbb{E}(\psi(X^+)X^+\mathbf{1}_{\{X>0\}}) + \psi(0)\int_{\{X\leq 0\}} X d\mathbb{P} \\
&= \mathbb{E}(\psi(X^+)X^+\mathbf{1}_{\{X^+>0\}}) + \mathbb{E}\left(\psi(X^+)\mathbf{1}_{\{X^+=0\}}\left(\frac{\int_{\{X\leq 0\}} X d\mathbb{P}}{\mathbb{P}(X \leq 0)}\right)\right).
\end{aligned}$$

Si l'on pose $\varphi(y) = y\mathbf{1}_{\{y>0\}} + \dfrac{\int_{\{X\leq 0\}} X d\mathbb{P}}{\mathbb{P}(X \leq 0)}\mathbf{1}_{\{y=0\}}$, on vérifie bien : $\mathbb{E}(\psi(X^+)X) = \mathbb{E}(\psi(X^+)\varphi(X^+))$, d'où le résultat. #

Exemple 4 : On veut exprimer $\mathbb{E}(X||X|)$ à l'aide de $\mathbb{E}(\text{sign}(X)||X|)$.

Il suffit de remarquer que $X = \text{sign}(X)|X|$ et que $|X|$ est $\sigma(|X|)$-mesurable ! On conclut que : $\mathbb{E}(X||X|) = |X|\mathbb{E}(\text{sign}(X)||X|)$. Cette dernière espérance conditionnelle n'est généralement pas simplifiable en l'absence d'autres hypothèses. Cependant, si $X \overset{\mathcal{L}}{\sim} -X$, et $\mathbb{P}(X=0)=0$, on montre que $\text{sign}(X)$ et $|X|$ sont indépendantes et que $\text{sign}(X)$ est centrée. Soient f et g boréliennes bornées :

$$\begin{aligned}
\mathbb{E}(f(\text{sign}(X))g(|X|)) &= \mathbb{E}(f(-\text{sign}(X))g(|X|)) \text{ car } \text{sign}(-x) = -\text{sign}(x) \\
&= \mathbb{E}\left(g(|X|)\frac{f(\text{sign}(X)) + f(\text{sign}(-X))}{2}\right) \\
&= \frac{f(1) + f(-1)}{2}\mathbb{E}(g(|X|))
\end{aligned}$$

d'où (avec $g \equiv 1$), $\text{sign}(X) \overset{\mathcal{L}}{\sim} \frac{1}{2}(\delta_1 + \delta_{-1}) = Ber(\{\pm 1\}, \frac{1}{2})$ et, avec f et g à nouveau quelconques, $\mathbb{E}(f(\text{sign}(X))g(|X|)) = \mathbb{E}(f(\text{sign}(X))\mathbb{E}(g(|X|))$ *i.e.* $\text{sign}(X)$ et $|X|$ sont des v.a. indépendantes. En conséquence, $\mathbb{E}(\text{sign}(X)||X|) = \mathbb{E}(\text{sign}(X)) = 0$.

Nous allons maintenant montrer que le résultat "heuristique" obtenu pour des couples de v.a. à densité rentre bien dans notre cadre général.

Théorème 4 : Soit $(X_1, \ldots, X_d, Y) : (\Omega, \mathcal{A}, \mathbb{P}) \to \mathbb{R}^{d+1}$ un $d+1$-uplet de v.a.r. admettant une densité : $(x_1, \ldots, x_d, y) \mapsto h(x_1, \ldots, x_d, y)$ sur $\mathbb{R}^{d+1}$. On suppose de plus que $Y \in L^1(\Omega, \mathcal{A}, \mathbb{P})$, *i.e.* que $\mathbb{E}(|Y|) = \int |y|h(x_1, \ldots, x_d, y)dydx_1 \ldots dx_d < +\infty$. Alors :

$$\mathbb{E}(Y|(X_1, \ldots, X_d)) = \frac{\int yh(X_1, \ldots, X_d, y)dy}{\int h(X_1, \ldots, X_d, y)dy} \quad \mathbb{P}\text{-}p.s..$$

Remarque préliminaire : le terme de droite de l'égalité ci-dessus n'est défini que $\mathbb{P}$-*p.s.* d'après le théorème de Fubini-Tonelli. En effet, le d-uplet $(X_1, \ldots, X_d)$ a pour densité : $(x_1, \ldots, x_d) \mapsto \int h(x_1, \ldots, x_d, y)dy$, donc :

$$\mathbb{P}\left(\left\{\int h(X_1, \ldots, X_d, y)dy = 0\right\}\right) = \int\!\!\!\!-\!\!\int \mathbf{1}_{\{\int h(x_1,\ldots,x_d,y)dy=0\}}\left(\int h(x_1, \ldots, x_d, yv)dv\right)dx_1 \ldots dx_d = 0.$$

Démonstration : Soit $\psi : \mathbb{R}^d \to \mathbb{R}$ borélienne bornée. D'après le théorème de transfert,

$$\begin{aligned}
\mathbb{E}(\psi(X_1, \ldots, X_d)Y) &= \underbrace{\int \ldots \int}_{d+1} \psi(x_1, \ldots, x_d)yh(x_1, \ldots, x_d, y)dx_1 \ldots dx_d dy \\
&= \underbrace{\int \ldots \int}_{d} \psi(x_1, \ldots, x_d)\left(\int yh(x_1, \ldots, x_d, y)dy\right)dx_1 \ldots dx_d
\end{aligned}$$

d'après le théorème de Fubini. A ce stade, on réintroduit la densité du $d+1$-uplet $(X_1,\dots,X_d,Y)$:

$$\begin{aligned}\mathbb{E}(\psi(X_1,\dots,X_d)Y) &= \underbrace{\int ... \int}_{d+1} \psi(x_1,\dots,x_d)\left(\frac{\int yh(x_1,\dots,x_d,y)dy}{\int h(x_1,\dots,x_d,y)dy}\right) h(x_1,\dots,x_d,y)dx_1\dots dx_d dy \\ &= \mathbb{E}(\psi(X_1,\dots,X_d)\varphi(X_1,\dots,X_d))\end{aligned}$$

où $\varphi(x_1,\dots,x_d) = \begin{cases} \dfrac{\int yh(x_1,\dots,x_d,y)dy}{\int h(x_1,\dots,x_d,y)dy} & \text{si} \int h(x_1,\dots,x_d,y)dy \neq 0 \\ 0 \ \text{(par exemple)} & \text{si} \int h(x_1,\dots,x_d,y)dy = 0. \end{cases}$

Conformément à la remarque préliminaire, on note que

$$\mathbb{P}_{(X_1,\cdots,X_d)}(\{\int h(x_1,\dots,x_d,y)dy = 0\}) = 0,$$

ce qui justifie à nouveau l'écriture adoptée dans l'énoncé. #

Remarque : On aurait pu tout aussi bien calculer plus généralement $\mathbb{E}(f(Y)|(X_1,\dots,X_d))$ pour toute fonction f borélienne bornée, ce qui aurait conduit à :

$$\mathbb{E}(f(Y)|(X_1,\dots,X_d)) = \int f(y)\frac{h(X_1,\dots,X_d,y)}{\int h(X_1,\dots,X_d,v)dv}dy \ \mathbb{P}\text{-}p.s..$$

Cela "incite" à affirmer que la loi conditionnelle de Y sachant $(X_1,\dots,X_d)$, notée $\mathcal{L}(Y|(X_1,\dots,X_d))$ ou $\mathbb{P}_Y(dy|(X_1,\dots,X_d))$ est donnée par :

$$\mathbb{P}_Y(dy|(X_1,\dots,X_d)) = \frac{h(X_1,\dots,X_d,y)}{\int h(X_1,\dots,X_d,v)dv}dy.$$

Ce point sera développé plus loin.

Exemple 5 : Soit (X,Y) un vecteur gaussien centré non dégénéré de matrice de covariance $Q := \begin{pmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{pmatrix}$, *i.e.* $\sigma_X^2 = Var(X)$, $\sigma_Y^2 = Var(Y)$ et $\rho = corr(X,Y) \in]-1,1[$. On veut calculer $\mathbb{E}(Y|X)$ et plus généralement déterminer $\mathcal{L}(Y|X)$. Il vient classiquement :

$$Q^{-1} = \frac{1}{1-\rho^2}\begin{pmatrix} \frac{1}{\sigma_X^2} & -\frac{\rho}{\sigma_X\sigma_Y} \\ -\frac{\rho}{\sigma_X\sigma_Y} & \frac{1}{\sigma_Y^2}\end{pmatrix} := \begin{pmatrix} \alpha & -\gamma \\ -\gamma & \beta \end{pmatrix} \ \text{(pour simplifier).}$$

La densité s'écrit : $h(x,y) = C^{te} \times e^{-\frac{1}{2}{}^t\begin{pmatrix} x \\ y\end{pmatrix} Q^{-1}\begin{pmatrix} x \\ y\end{pmatrix}} dxdy$, donc

$$\mathcal{L}(Y|X) = \frac{e^{-\frac{1}{2}(\alpha X^2 - 2\gamma Xy + \beta y^2)}}{\int e^{-\frac{1}{2}(\alpha X^2 - 2\gamma Xv + \beta v^2)}dv}dy = C(X) \times e^{-\frac{\beta}{2}(y - \frac{\gamma X}{\beta})^2}dy.$$

Comme $\beta = \dfrac{1}{(1-\rho^2)\sigma_Y^2}$ et $\dfrac{\gamma X}{\beta} = \rho\frac{\sigma_X}{\sigma_Y}X$, on en déduit aussitôt que :

$$\mathcal{L}(Y|X) = \frac{1}{\sqrt{2\pi}\sigma_Y\sqrt{1-\rho^2}}\exp\left[-\frac{1}{2(1-\rho^2)\sigma_Y^2}\left(y - \rho\frac{\sigma_X}{\sigma_Y}X\right)^2\right]dy = \mathcal{N}\left(\rho\frac{\sigma_Y}{\sigma_X}X, (1-\rho^2)\sigma_Y^2\right).$$

D'où en particulier, $\mathbb{E}(Y|X) = \rho\dfrac{\sigma_Y}{\sigma_X}X$.

Revenons maintenant au cadre général pour étudier les liens entre indépendance et espérance conditionnelle.

Proposition : Soient X et Y deux v.a. indépendantes sur $(\Omega,\mathcal{A},\mathbb{P})$. Soit $f : \mathbb{R} \to \mathbb{R}$ une fonction borélienne bornée. Alors

$$\mathbb{E}(f(X,Y)|X) = \int f(X,y)\mathbb{P}_Y(dy).$$

Démonstration : Soit $\psi(X)$ une fonction-test. Il vient :

$$\begin{aligned}\mathbb{E}(\psi(X)f(X,Y)) &= \iint \psi(x)f(x,y)\mathbb{P}_{(X,Y)}(dx,dy) \\ &= \iint \psi(x)f(x,y)\mathbb{P}_X(dx)\mathbb{P}_Y(dy) \ \text{ par indépendance,} \\ &= \int \psi(x)\left[\int f(x,y)\mathbb{P}_Y(dy)\right]\mathbb{P}_X(dx) \ \text{d'après le théorème de Fubini.}\end{aligned}$$

On pose $\varphi(x) := \int f(x,y)\mathbb{P}_Y(dy)$, et on en déduit :

$$\mathbb{E}(\psi(X)f(X,Y)) = \int \psi(x)\varphi(x)\mathbb{P}_X(dx) = \mathbb{E}(\psi(X)\varphi(X)),$$

d'où $\mathbb{E}(f(X,Y)|X) = \int f(X,y)\mathbb{P}_Y(dy)$. #

Cette propriété, d'un usage très courant en calcul conditionnel, sera illustrée à la sous-section suivante.

11.4.2 Conditionnement par un vecteur gaussien

En fait l'exemple 5 relève d'un cadre beaucoup plus général, dans lequel il est facile de conditionner : les espaces gaussiens. Au niveau élémentaire qui est le nôtre, nous allons nous cantonner aux vecteurs gaussiens $(X_1,\dots,X_d,Y)$, que nous supposerons non-dégénérés.

$$\text{On pose } \Sigma^{(2)} := \left[\begin{array}{c|c} \mathrm{Cov}(X_i,X_j)_{1\le i,j\le d} & \mathrm{Cov}(X_i,Y)_{1\le i\le d} \\ \hline \mathrm{Cov}(Y,X_j)_{1\le j\le d} & \mathrm{Var}(Y^2) \end{array}\right] = \left[\begin{array}{c|c} \Sigma^{(2)}_X & \Sigma^{(2)}_{XY} \\ \hline \Sigma^{(2)}_{YX} & \Sigma^{(2)}_Y \end{array}\right]$$

la matrice de covariance (ou de dispersion) du vecteur gaussien $(X_1,\dots,X_d,Y)$. La matrice de covariance du vecteur X est notée $\Sigma^{(2)}_X$. Elle est symétrique non dégénérée ; le vecteur-ligne ${}^t\Sigma^{(2)}_{YX} = \Sigma^{(2)}_{XY}$ représente la matrice de covariance entre la v.a. Y et le vecteur X.

A ce stade, on munit $\mathbb{R}^{d+1}$ du produit scalaire euclidien $< u,v > := {}^t u\Sigma^{(2)}v$. $(e_1,\cdots,e_{d+1})$ désigne la base canonique de $\mathbb{R}^{d+1}$.

Soit $b_1,\dots,b_d$ les coordonnées sur $(e_1,\dots,e_d)$ de la projection "$\Sigma^{(2)}$-orthogonale" de e_{d+1} sur $\mathrm{Vect}\{e_1,\dots,e_d\}$.

Si l'on pose $\tilde X_i := X_i - \mathbb{E}(X_i)$, $1\le i\le d$, et $\tilde Y := Y - \mathbb{E}(Y) - \sum_{i=1}^d b_i\tilde X_i$, il est immédiat que $\tilde X := (\tilde X_1,\dots,\tilde X_d,\tilde Y)$ est un vecteur gaussien vérifiant, pour tout $i\in\{1,\dots,d\}$:

$$(*)\qquad \begin{aligned} \mathrm{Cov}(\tilde X_i\tilde Y) = \mathbb{E}(\tilde X_i\tilde Y) &= \mathrm{Cov}(X_i,Y) - \sum_{j=1}^d b_j\mathrm{Cov}(X_i,X_j) \\ &= {}^t e_i\Sigma^{(2)}\left(e_{d+1} - \sum_{j=1}^d b_je_j\right) = 0. \end{aligned}$$

En conséquence (cf. chapitre VII.2), $\tilde Y$ est indépendant de $(\tilde X_1,\dots,\tilde X_d)$ *i.e.* de $(X_1,\dots,X_d)$, et donc :

$$\mathbb{E}(\tilde Y|\sigma(X_1,\dots,X_d)) = \mathbb{E}(\tilde Y).$$

Cette identité s'écrit à son tour

$$\mathbb{E}(Y|(X_1,\dots,X_d)) = \mathbb{E}(Y) + \sum_{i=1}^d b_i(X_i - \mathbb{E}(X_i)).$$

Le calcul d'espérance conditionnelle se ramène donc ici à un problème d'algèbre linéaire, en l'espèce la détermination des b_i.

Notons pour simplifier ${}^tB := (b_1,\dots,b_d)$ (B est donc un vecteur colonne). Les égalités $(*)$ $\mathbb{E}(\tilde X_i\tilde Y)=0$, $1\le i\le d$, s'expriment à l'aide de $\Sigma^{(2)}$ en

$$\Sigma^{(2)}_{X_iY} = \Sigma^{(2)}_{XX_i}B,\ i=1,\dots,d \textit{ i.e. } \Sigma^{(2)}_{XY} = \Sigma^{(2)}_X B.$$

Finalement $B = (\Sigma^{(2)}_X)^{-1}\Sigma^{(2)}_{XY}$.

On peut alors exprimer l'espérance conditionnelle de façon synthétique sous la forme

$$\mathbb{E}(Y|(X_1,\dots,X_d)) = \mathbb{E}(Y) + {}^t\Sigma^{(2)}_{XY}(\Sigma^{(2)}_X)^{-1}(X - \mathbb{E}(X)).$$

<u>Théorème 5</u> : Si $(X_1,\dots,X_d,Y)$ est un vecteur gaussien de dispersion $\Sigma^{(2)}$, $\Sigma^{(2)}_X$ inversible. Alors,

$$\mathcal{L}(Y|(X_1,\dots,X_d)) = \mathcal{N}\left(\mathbb{E}(Y) + {}^t\Sigma^{(2)}_{XY}(\Sigma^{(2)}_X)^{-1}(X - \mathbb{E}(X));\ \Sigma^{(2)}_Y - {}^t\Sigma^{(2)}_{XY}(\Sigma^{(2)}_X)^{-1}\Sigma^{(2)}_{XY}\right).$$

<u>Remarque</u> : La formule se généralise immédiatement – sous la même forme – à un vecteur Y multidimensionnel, avec les mêmes notations.

<u>Démonstration</u> : Quitte à remplacer X par $X-\mathbb{E}X$ et Y par $Y-\mathbb{E}Y$, on peut toujours supposer les vecteurs centrés. En reprenant le raisonnement fait pour le calcul de l'espérance conditionnelle, on a

$$Y = \left(Y - \sum_{i=1}^d b_iX_i\right) + \sum_{i=1}^d b_iX_i = \tilde Y + \sum_{i=1}^d b_iX_i,$$

où $\tilde Y$ est indépendante du d-uplet $(X_1,\dots,X_d)$.

Par suite, si $f:\mathbb{R}\to\mathbb{R}$ est borélienne bornée, il vient :

$$\mathbb{E}(f(Y)|(X_1,\dots,X_d)) = \int f(\sum_{i=1}^d b_iX_i + y)\mathbb{P}_{\tilde Y}(dy).$$

Or $\tilde Y$ suit une loi normale centrée de variance vérifiant

$$\mathrm{Var}(\tilde Y) + \mathrm{Var}(\sum_{i=1}^d b_iX_i) = \mathrm{Var}(Y)$$

par additivité de la variance en cas d'indépendance.

$$\mathbb{E}(f(Y)|(X_1,\dots,X_d)) = \int f(\sum_{i=1}^d b_iX_i + \sqrt{\mathrm{Var}(\tilde Y)}z)\frac{e^{-\frac{z^2}{2}}}{\sqrt{2\pi}}dz.$$

Reste donc à exprimer $\mathrm{Var}(\tilde Y)$ en fonction de $\Sigma^{(2)}$. Or $\mathrm{Var}(Y) = \Sigma^{(2)}_Y$ et $\mathrm{Var}({}^tBX) = {}^tB\Sigma^{(2)}_XB$. On trouve le résulat annoncé en remplaçant B par son expression en fonction de $\Sigma^{(2)}$:

$$\mathrm{Var}(\tilde Y) = \Sigma^{(2)}_Y - {}^t\Sigma^{(2)}_{XY}(\Sigma^{(2)}_X)^{-1}\Sigma^{(2)}_X(\Sigma^{(2)}_X)^{-1}\Sigma^{(2)}_{XY}.\#$$

11.5 Notion générale de loi conditionnelle

Le théorème ci-dessous généralise la notion de loi conditionnelle en montrant, pour tout couple de vecteurs aléatoires (X,Y), l'existence d'une famille "régulière" (et *p.s.* unique) de probabilités $\mathbb{P}_X(.|Y)$ telle que l'on ait simultanément :

$$\mathbb{E}(f(X)|Y) \overset{\mathbb{P}\text{-}p.s.}{=} \int f(x)\mathbb{P}_X(dx|Y) \text{ et } y\mapsto \int f(x)\mathbb{P}_X(dx|y) \text{ borélienne.}$$

Théorème 6 : Soit $(X,Y):(\Omega,\mathcal{A},\mathbb{P})\to\mathbb{R}^d\times\mathbb{R}^{d'}$.
• Il existe une famille de probabilités sur $(\mathbb{R}^d,\mathcal{B}(\mathbb{R}^d))$ indexée par $y\in\mathbb{R}^{d'}$, notée $(\mathbb{P}_X(dx|y))_{y\in\mathbb{R}^{d'}}$ et vérifiant :

(*i*) $\forall f:\mathbb{R}^d\to\mathbb{R}$ borélienne bornée,

$$\mathbb{E}(f(X)|Y) \overset{\mathbb{P}\text{-}p.s.}{=} \int f(x)\mathbb{P}_X(dx|Y)$$

où $\int f(x)\mathbb{P}_X(dx|Y) := \int f(x)\mathbb{P}_X(dx|y)_{|y=Y}$.

(*ii*) $\forall f:\mathbb{R}^d\to\mathbb{R}$ borélienne bornée, $y\mapsto\int f(x)\mathbb{P}_X(dx|y)$ est borélienne bornée de $\mathbb{R}^{d'}$ dans $\mathbb{R}$.

• Si $(\mathbf{Q}(dx|y))_{y\in\mathbb{R}^{d'}}$ est une famille de probabilités sur $(\mathbb{R}^d,\mathcal{B}(\mathbb{R}^d))$ vérifiant (*i*) et (*ii*), alors $\mathbb{P}$-*p.s.*, $\mathbb{P}_X(dx|Y)=\mathbf{Q}(dx|Y)$, *i.e.*

$$\mathbb{P}\left(\{\omega\in\Omega \;/\; \mathbb{P}_X(dx|Y(\omega))\neq\mathbf{Q}(dx|Y(\omega))\}\right)=0.$$

Nous admettrons ce résultat dont la démonstration est très technique. Dans la plupart des applications courantes, la version régulière de la loi conditionnelle telle que définie ci-dessus apparaît naturellement à travers le calcul des espérances conditionnelles $\mathbb{E}(f(X)|Y)$.

C'est ainsi le cas dans le théorème 4 et la remarque qui le suit ; d'où il ressort que :

Proposition : Si la loi de (X,Y) a une densité h définie sur $\mathbb{R}^d\times\mathbb{R}^{d'}$, alors

$$\mathbb{P}_X(dx|Y=y)=\frac{h(x,y)}{\int_{\mathbb{R}^{d'}}h(u,y)du}dx$$

est une version régulière de la loi conditionnelle de X sachant Y.

Le théorème suivant peut être vu comme une généralisation de la formule de Bayes bien connue dans le cadre élémentaire. Sa démonstration s'apparente à celle du théorème de Fubini.

Théorème 7 : Avec les hypothèses et les notations du théorème 6, on a, pour toute fonction $f:\mathbb{R}^d\times\mathbb{R}^{d'}\to\mathbb{R}$ borélienne bornée :

(*i*) $\mathbb{E}(f(X,Y))=\int\mathbb{P}_Y(dy)\int f(x,y)\mathbb{P}_X(dx|y)$ *i.e.* $\mathbb{P}_{(X,Y)}(dx,dy)=\mathbb{P}_X(dx|y)\mathbb{P}_Y(dy)$.

(*ii*) $\mathbb{E}(f(X,Y)|Y)=\int f(x,Y)\mathbb{P}_X(dx|Y)$.

Démonstration : (*i*) Il suffit de vérifier que si $C\in\mathcal{B}(\mathbb{R}^d\times\mathbb{R}^{d'})$ et $C^y:=\{x\in\mathbb{R}^d \;/\; (x,y)\in C\}$, alors $C\mapsto\int\mathbb{P}_Y(dy)\mathbb{P}_X(C^y|y)$ est une probabilité sur $\mathcal{B}(\mathbb{R}^d\times\mathbb{R}^{d'})$ égale à $\mathbb{P}_{(X,Y)}$.

La première assertion est un corollaire immédiat du théorème de Beppo Levi pour les séries de fonctions et de la régularité (au sens du théorème 4) de $\mathbb{P}_X(dx|y)$.

D'autre part si $C=A\times B$, $A\in\mathcal{B}(\mathbb{R}^d)$, $B\in\mathcal{B}(\mathbb{R}^{d'})$, $C^y=\mathbf{1}_B(y)A$,

$$\begin{aligned}\int\mathbb{P}_Y(dy)\mathbb{P}_X(C^y|y) &= \int_B\mathbb{P}_X(A|y)\mathbb{P}_Y(dy)=\mathbb{E}(\mathbf{1}_{\{Y\in B\}}\mathbb{P}_X(A|Y))\\ &= \mathbb{E}(\mathbf{1}_{\{Y\in B\}}\mathbb{E}(\mathbf{1}_{\{X\in A\}}|Y))=\mathbb{E}(\mathbf{1}_{\{X\in A\}\times\{Y\in B\}})\\ &= \mathbb{P}((X,Y)\in A\times B)=\mathbb{P}_{(X,Y)}(C)\end{aligned}$$

D'où le résultat puisque, par définition, $\sigma(A\times B, A\in\mathcal{B}(\mathbb{R}^d), B\in\mathcal{B}(\mathbb{R}^{d'}))=\mathcal{B}(\mathbb{R}^d)\otimes\mathcal{B}(\mathbb{R}^{d'})=\mathcal{B}(\mathbb{R}^d\times\mathbb{R}^{d'})$ et $\{A\times B, A\in\mathcal{B}(\mathbb{R}^d), B\in\mathcal{B}(\mathbb{R}^{d'})\}$ est stable par intersection finie.

(*ii*) On déduit du point (*i*) que, si g est borélienne bornée sur $\mathbb{R}^{d'}$,

$$\begin{aligned}\mathbb{E}(g(Y)f(X,Y)) &= \int\mathbb{P}_Y(dy)\int g(y)f(x,y)\mathbb{P}_X(dx|y)=\int\mathbb{P}_Y(dy)g(y)\int f(x,y)\mathbb{P}_X(dx|y)\\ &= \mathbb{E}\left(g(Y)\underbrace{\left(\int f(x,Y)\mathbb{P}_Y(dx|Y)\right)}_{\sigma(Y)\text{-mesurable}}\right)\end{aligned}$$

d'où le résultat. #

Corollaire : X et Y sont indépendantes si et seulement si $\mathbb{P}_X(dx|Y)=\mathbb{P}_X(dx)$ $\mathbb{P}$-*p.s.*.

Démonstration : $\Rightarrow$: Soit f borélienne bornée sur $\mathbb{R}^d\times\mathbb{R}^{d'}$, alors d'après Fubini :

$$\mathbb{E}(f(X,Y))=\int\mathbb{P}_Y(dy)\left(\int\mathbb{P}_X(dx)f(x,y)\right),$$

car $\mathbb{P}_{(X,Y)}=\mathbb{P}_X\otimes\mathbb{P}_Y$. $y\mapsto\mathbb{P}_X(dx)$ est donc une version régulière de la loi conditionnelle de X sachant Y.

$\Leftarrow$: Soient f et g boréliennes bornées sur $\mathbb{R}^d$ et $\mathbb{R}^{d'}$,

$$\begin{aligned}\mathbb{E}(f(X)g(Y)) &= \int\mathbb{P}_Y(dy)\int f(x)g(y)\mathbb{P}_X(dx|y)=\int\mathbb{P}_Y(dy)\int f(x)g(y)\mathbb{P}_X(dx)\\ &= \left(\int\mathbb{P}_Y(dy)g(y)\right)\times\left(\int\mathbb{P}_X(dx)f(x)\right)=\mathbb{E}(g(Y))\mathbb{E}(f(X)).\quad\#\end{aligned}$$

Table des matières