Intégration et Probabilités

Étienne Pardoux

14 septembre 2015

Table des matières

Introduction

Chapitre 1

Mesure

Un espace mesuré est un triplet $(\Omega, \mathcal{F}, \mu)$ constitué de :
un espace (ou ensemble) Ω
une σ–algèbre (ou tribu) $\mathcal{F}$ de parties de Ω
une mesure μ sur l'espace mesurable $(\Omega, \mathcal{F})$.
Le but de ce chapitre est de définir et d'étudier cette notion.

1.1 Rappels sur les limites de réels

Etant donné $\{x_n, n \geq 1\}$ une suite quelconque de nombres réels, les quantités suivantes sont toujours définies dans $\bar{\mathbb{R}}$:

$$\limsup_{n \mapsto \infty} x_n = \overline{\lim}_{n \mapsto \infty} x_n = \lim_{n \to \infty} \downarrow [\sup_{p \geq n} x_p]$$
$$\liminf_{n \mapsto \infty} x_n = \underline{\lim}_{n \mapsto \infty} x_n = \lim_{n \mapsto \infty} \uparrow [\inf_{p \geq n} x_p]$$

Avec les notations $a \vee b = \sup(a, b)$ et $a \wedge b = \inf(a, b)$ on peut encore écrire :

$$\limsup_{n \mapsto \infty} x_n = \bigwedge_{n \geq 1} \bigvee_{p \geq n} x_p$$
$$\liminf_{n \mapsto \infty} x_n = \bigvee_{n \geq 1} \bigwedge_{p \geq n} x_p$$

On admettra la

Proposition 1.1.1. *Une suite* $\{x_n,\ n \in \mathbb{N}\}$ *de nombres réels converge dans* $\bar{\mathbb{R}} = \mathbb{R} \cup \{+\infty\} \cup \{-\infty\}$ *si et seulement si* $\overline{\lim}\ x_n = \underline{\lim}\ x_n$. *Dans ce cas,* $\lim\ x_n = \overline{\lim}\ x_n = \underline{\lim}\ x_n$.

Notons que l'on a toujours $\underline{\lim}\ x_n \leq \overline{\lim}\ x_n$.

Remarque 1.1.2. *1. Etant donné une suite* $\{x_n\}$ *de nombres réels, on n'a pas le droit d'écrire "*$\lim\ x_n$*" tant que l'on ne sait pas si cette limite existe, i.e. si* $\limsup\ x_n = \liminf\ x_n$. *La limite existe toujours lorsque la suite* $\{x_n\}$ *est soit croissante, soit décroissante.*

2. On dit qu'une suite de réels $\{x_n\}$ *converge (dans* $\mathbb{R}$*) si on a à la fois :*

a- $\liminf\ x_n = \limsup\ x_n = \lim\ x_n$ *et*

b- $\lim\ x_n \in \mathbb{R}$.

On dit qu'une suite de réels $\{x_n\}$ *diverge (dans* $\mathbb{R}$*) si elle ne converge pas, i.e. si :*

soit $\limsup\ x_n \neq \liminf\ x_n$

soit $\limsup\ x_n = \liminf\ x_n = +\infty$ *ou* $-\infty$

1.2 Opérations sur les ensembles

On note $A, B, C, \ldots$ les parties de Ω. On définit

$$\begin{aligned}
A^c &= \{\omega \in \Omega; \omega \notin A\} \\
A \cup B &= \{\omega \in \Omega; \omega \in A \text{ ou } \omega \in B\} \\
A \cap B &= \{\omega \in \Omega; \omega \in A \text{ et } \omega \in B\} \\
A \triangle B &= (A \cap B^c) \cup (A^c \cap B) \\
&= A \cup B \backslash A \cap B.
\end{aligned}$$

La notation $A \backslash B$ n'est utilisée que lorsque $A \supset B$, et elle désigne alors $A \cap B^c$.

Si $A_n \uparrow$ (i.e. $A_n \subset A_{n+1}$), $\lim\ A_n \triangleq \bigcup_n A_n$

Si $A_n \downarrow$ (i.e. $A_n \supset A_{n+1}$), $\lim\ A_n \triangleq \bigcap_n A_n$.

Etant donnée $\{A_n, n \in \mathbb{N}\}$ une suite de parties de Ω, on définit :

$$\begin{aligned}
\limsup_n A_n &= \lim_n \downarrow [\bigcup_{p \geq n} A_p] \\
&= \bigcap_n \bigcup_{p \geq n} A_p \\
&= \left\{\omega \in \Omega; \sum_1^\infty \mathbf{1}_{A_n}(\omega) = +\infty\right\} \\
\liminf_n A_n &= \lim_n \uparrow [\bigcap_{p \geq n} A_p] \\
&= \bigcup_n \bigcap_{p \geq n} A_p \\
&= \left\{\omega \in \Omega; \sum_1^\infty \mathbf{1}_{A_n^c}(\omega) < +\infty\right\}
\end{aligned}$$

$\limsup A_n$ est l'ensemble des ω qui appartiennent à une infinité de A_n.

$\liminf A_n$ est l'ensemble des ω qui appartiennent à tous les A_n sauf au plus un nombre fini. En particulier $\liminf A_n \subset \limsup A_n$.

1.3 Espace mesurable

On appelle espace mesurable un couple $(\Omega, \mathcal{F})$ formé d'un ensemble Ω, muni d'une σ–algèbre (cf. la Définition 1.3.2 ci–dessous) $\mathcal{F}$ de parties de cet ensemble.

Définition 1.3.1. *Une classe $\mathcal{F}_0$ de parties de Ω est appelée une algèbre si :*

$$A,\ B \in \mathcal{F}_0 \Rightarrow A^c,\ A \cup B \in \mathcal{F}_0$$

[donc aussi $A \cap B \in \mathcal{F}_0$].

On remarque que si $\mathcal{F}_0$ est une algèbre non vide, $\Omega, \emptyset \in \mathcal{F}_0$.

Définition 1.3.2. *Une algèbre $\mathcal{F}$ est appelée une σ–algèbre (ou tribu) si de plus :*

$$A_n \in \mathcal{F}; n = 1, 2, \ldots \ \Rightarrow \bigcup_1^\infty A_n \in \mathcal{F}.$$

On vérifie aisément qu'une intersection quelconque d'algèbres [resp. de σ–algèbres] est une algèbre [resp. une σ–algèbre].

Proposition 1.3.3. *Soit $\mathcal{G}$ une algèbre. Les propriétés suivantes sont équivalentes :*

(i) $\mathcal{G}$ est une σ–algèbre

(ii) $\mathcal{G}$ est stable par intersection dénombrable

(iii) $\mathcal{G}$ est stable par limite croissante

(iv) $\mathcal{G}$ est stable par limite décroissante.

PREUVE Par passage au complémentaire, (i)$\Leftrightarrow$ (ii) et (iii) $\Leftrightarrow$ (iv). (i) $\Rightarrow$ (iii) est évident [cf. définition de la limite croissante]. (iii) $\Rightarrow$ (i) résulte de $\displaystyle\bigcup_n A_n = \lim_n \uparrow (\bigcup_{p\leq n} A_p)$. □

Proposition 1.3.4. *Soit $\mathcal{C}$ une classe de parties de Ω. Alors il existe une plus petite algèbre contenant $\mathcal{C}$, et une plus petite σ–algèbre contenant $\mathcal{C}$, notée $\sigma(\mathcal{C})$.*

PREUVE Il existe au moins une σ–algèbre contenant $\mathcal{C}$, à savoir $\mathcal{P}(\Omega)$ [=classe de toutes les parties de Ω]. Or l'intersection de toutes les algèbres [resp. les σ–algèbres] contenant $\mathcal{C}$ est une algèbre [resp. une σ–algèbre]. □

Définition 1.3.5. *Soit Ω un espace topologique (i.e. un espace muni d'une famille d'ouverts, par exemple un espace métrique). On appelle σ–algèbre (ou tribu) borélienne sur Ω la plus petite σ–algèbre contenant tous les ouverts de Ω.*

Etant donnés $(\Omega_1,\ \mathcal{F}_1)$ et $(\Omega_2,\ \mathcal{F}_2)$ deux espaces mesurables, on note $(\Omega_1 \times \Omega_2, \mathcal{F}_1 \otimes \mathcal{F}_2)$ *l'espace mesurable produit* défini par :

$$\begin{aligned}&\Omega_1 \times \Omega_2 = \text{ensemble des couples } (\omega_1, \omega_2), \text{où } \omega_1 \in \Omega_1 \text{ et } \omega_2 \in \Omega_2\\ &\mathcal{F}_1 \otimes \mathcal{F}_2 = \sigma(\mathcal{F}_1 \times \mathcal{F}_2),\ \text{ où :}\end{aligned}$$

$$\mathcal{F}_1 \times \mathcal{F}_2 \stackrel{\triangle}{=} \{A_1 \times A_2; A_1 \in \mathcal{F}_1 \text{ et } A_2 \in \mathcal{F}_2\}$$

Définition 1.3.6. *Une classe $\mathcal{P}$ de parties de Ω est appelée un π–**système** si elle est stable par intersection finie.*

Définition 1.3.7. *Une classe des parties de Ω est appelée un λ–***système** *si elle vérifie :*

(λ_1) $\Omega \in \mathcal{L}$

(λ_2) $A, B \in \mathcal{L}$ *et* $B \subset A \Rightarrow A \backslash B \in \mathcal{L}$

(λ_3) $\{A_n, n \geq 1\} \subset \mathcal{L}$ *et* $A_n \uparrow A \ \Rightarrow A \in \mathcal{L}$

Théorème 1.3.8. *($\pi - \lambda$) Soit $\mathcal{P}$ un π–système, et $\mathcal{L}$ un λ–système, tels que $\mathcal{P} \subset \mathcal{L}$. Alors $\sigma(\mathcal{P}) \subset \mathcal{L}$.*

Preuve Soit $\lambda(\mathcal{P})$ le plus petit λ–système contenant $\mathcal{P}$. Bien sûr, $\lambda(\mathcal{P}) \subset \mathcal{L}$. Si $\lambda(\mathcal{P})$ est aussi un π–système, alors c'est une σ–algèbre (cf. l'exercice 1.8.1 (ii)), et dans ce cas $\sigma(\mathcal{P}) \ \subset \ \lambda(\mathcal{P}) \ \subset \ \mathcal{L}$.

Montrons donc que $\lambda(\mathcal{P})$ est un π–système. Pour tout $A \subset \Omega$, posons $\mathcal{G}_A = \{B; A \cap B \in \lambda(\mathcal{P})\}$. Si $A \in \lambda(\mathcal{P})$, alors $\mathcal{G}_A$ est un λ–système. Il n'est pas difficile de voir que si $A \in \mathcal{P}$, $\mathcal{P} \subset \mathcal{G}_A$, et a fortiori $\mathcal{G}_A$ est un λ–système. Donc si $A \in \mathcal{P}$, $\lambda(\mathcal{P}) \subset \mathcal{G}_A$, ce qui veut dire que si $A \in \mathcal{P}$, $B \in \lambda(\mathcal{P})$, alors $A \cap B \in \lambda(\mathcal{P})$. Donc si $B \in \lambda(\mathcal{P})$, $\mathcal{G}_B$ est un λ–système qui contient $\mathcal{P}$, donc $\lambda(\mathcal{P})$, c'est à dire : si $B, C \in \lambda(\mathcal{P})$, $B \ \cap \ C \in \lambda(\mathcal{P})$. □

Corollaire 1.3.9. *($\pi - \lambda$) Soit $\mathcal{P}$ un π–système. Alors $\lambda(\mathcal{P}) = \sigma(\mathcal{P})$.*

Preuve On a montré que $\sigma(\mathcal{P}) \subset \lambda(\mathcal{P})$. Mais $\sigma(\mathcal{P})$ est un λ–système. Donc $\lambda(\mathcal{P}) \subset \sigma(\mathcal{P})$. □

Autre énoncé du Corollaire $\lambda - \pi$:

Soit $\mathcal{P}$ une classe stable par intersection. La plus petite classe contenant $\mathcal{P}$ et Ω, stable par différence et limite croissante, est $\sigma(\mathcal{P})$.

1.4 Mesure

On suppose donné un espace mesurable $(\Omega, \mathcal{F})$.

Définition 1.4.1. *Une application $\mu : \mathcal{F} \to \bar{\mathbb{R}}_+$ est appelée une* **mesure** *si :*

(i) $\mu(\emptyset) = 0$;

(ii)Etant donné $\{A_n, n \geq 1\} \subset \mathcal{F}$, avec $A_k \cap A_\ell = \emptyset$, dès que $k \neq \ell$, alors

$$\mu\left(\bigcup_1^\infty A_n\right) = \sum_1^\infty \mu(A_n).$$

La propriété (ii) est appelée la σ–additivité. On définit de la même façon une mesure sur une algèbre $\mathcal{F}_0$ de parties de Ω, en rajoutant dans (ii) la condition “si $\bigcup\limits_1^\infty A_n \in \mathcal{F}_0$”. Une mesure μ est dite **finie** ou **infinie**, suivant que $\mu(\Omega) < \infty$ ou $\mu(\Omega) = +\infty$.

Définition 1.4.2. *Une mesure μ, définie sur un espace mesurable $(\Omega, \mathcal{F})$, sera dite σ–**finie** s'il existe $\{A_n, n \geq 1\} \subset \mathcal{F}$ telle que :*

(i) $\bigcup\limits_1^\infty A_n = \Omega$

(ii) $\mu(A_n) < \infty$, *pour tout* $n \geq 1$.

*Soit $\mathcal{C} \subset \mathcal{F}$. On dira que μ est σ–**finie le long de** $\mathcal{C}$, si (i) et (ii) sont satisfaites avec une suite $\{A_n, n \geq 1\} \subset \mathcal{C}$.*

Exemple 1.4.3. *Supposons que Ω est un ensemble fini ou dénombrable. On définit la mesure de comptage ν sur $\mathcal{P}(\Omega)$ par*

$$\nu(A) = Card(A), \quad \textit{pour tout } A \subset \Omega.$$

ν est finie si Ω est fini, σ–finie si Ω est dénombrable.

Dans toute la suite, nous nous limiterons à l'étude des mesures σ–finies.

Proposition 1.4.4. *Soit μ une mesure définie sur un espace mesurable $(\Omega, \mathcal{F})$.*

(i) Si $\{A_n, n \geq 1\} \subset \mathcal{F}$, $A_n \uparrow A$, alors $\mu(A_n) \uparrow \mu(A)$.

(ii) Si $\{A_n, n \geq 1\} \subset \mathcal{F}$, $A_n \downarrow A$, et $\mu(A_1) < \infty$, alors $\mu(A_n) \downarrow \mu(A)$.

(iii) Si $\{A_n, \geq 1\} \subset \mathcal{F}$, $\mu\left(\bigcup\limits_1^\infty A_n\right) \leq \sum\limits_1^\infty \mu(A_n)$.

(iv) μ étant σ–finie, $\mathcal{F}$ ne peut pas contenir une collection non dénombrable d'ensembles disjoints de mesure non nulle.

PREUVE

(i) Posons $B_1 = A_1$; $B_n = A_n \backslash A_{n-1}$, $n \geq 2$. Les B_n dont deux à deux disjoints, $A_n = \bigcup\limits_1^n B_k$, $A = \bigcup\limits_1^\infty B_k$. Donc (i) résulte de la σ–additivité.

(ii) On a $A_1 \setminus A_n \uparrow A_1 \setminus A$. Il résulte de (i) :

$$\mu(A_1) - \mu(A_n) = \mu(A_1 \setminus A_n) \uparrow \mu(A_1 \setminus A) = \mu(A_1) - \mu(A),$$

d'où la conclusion puisque $\mu(A_1) < \infty$.

(iii) $\forall n$, $\mu\left(\bigcup_1^n A_k\right) \leq \sum_1^n \mu(A_k)$ [cf. Exercice 1.8.4] donc $\mu\left(\bigcup_1^n A_k\right) \leq \sum_1^\infty \mu(A_k)$, et on applique (i) au membre de gauche.

(iv) Soit $\{B_\theta, \theta \in \Theta\} \subset \mathcal{F}$ une collection d'ensembles disjoints, avec $\mu(B_\theta) > 0$, $\forall\, \theta \in \Theta$. Etant donné $\{A_n, n \geq 1\} \subset \mathcal{F}$, avec $\bigcup_1^\infty A_n = \Omega$, et $\mu(A_n) < \infty$, pour tout n, on pose $\Theta_n = \{\theta \in \Theta; \mu(A_n \cap B_\theta) > 0\}$. Pour que $\Theta = \bigcup_1^\infty \Theta_n$ soit dénombrable, il suffit que Θ_n le soit pour tout $n \geq 1$. Mais $\Theta_n = \bigcup_{k=1}^\infty \Theta_n^k$, avec $\Theta_n^k = \{\theta \in \Theta; \mu(A_n \cap B_\theta) > \frac{1}{k}\}$. Or $\mathrm{card}(\Theta_n^k) \leq k \cdot \mu(A_n) < \infty$. D'où le résultat

$\square$

Remarque 1.4.5. *a) (ii) n'est pas vrai sans l'hypothèse $\mu(A_1) < \infty$. [ou du moins il existe n tel que $\mu(A_n) < \infty$]. Exemple : μ =mesure de comptage des entiers (sur $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$), $A_n = \mathbb{N} \cap [n, +\infty)$. $\mu(A_n) = +\infty$, pour tout n. Or $\mu(A) = 0$ puisque $A = \emptyset$.*

b) La proposition 1.4.4 est encore vraie si $\mathcal{F}$ est remplacée par une algèbre $\mathcal{F}_0$, à condition de supposer $A \in \mathcal{F}_0[(i), (ii)]$, $\bigcup_1^\infty A_n \in \mathcal{F}_0$ [(iii)], et μ est σ–finie le long de $\mathcal{F}_0$, [(iv)].

Théorème 1.4.6. *(d'extension de Carathéodory) Soit μ_0 une mesure définie sur une algèbre $\mathcal{F}_0$ de parties de Ω, qui est supposée σ–finie le long de $\mathcal{F}_0$. Il existe une unique mesure μ sur $\mathcal{F} = \sigma(\mathcal{F}_0)$, telle que $\mu|_{\mathcal{F}_0} = \mu_0$.*

Preuve L'unicité résulte du Théorème 1.4.11 qui suit. Nous allons démontrer l'existence. Commençons par définir l'application suivante de

$\mathcal{P}(\Omega)$ à valeurs dans $[0,+\infty]$:

$$\text{pour tout } A \subset \Omega, \quad \mu^*(A) = \inf \sum_n \mu_0(A_n),$$

où l'infimum est pris sur toutes les suites dénombrables $\{A_n\}$ telles que $A_n \in \mathcal{F}_0$ et $A \subset \bigcup_n A_n$.

Le lecteur vérifiera aisément que μ^* est une *mesure extérieure*, au sens où elle vérifie

$\mu^*(\emptyset) = 0$;

$\mu^*(A) \leq \mu^*(B)$, si $A \subset B$ (monotonie) ;

$\mu^*(\cup_n A_n) \leq \sum_n \mu^*(A_n)$ (sous–additivité dénombrable) ;

qui coïncide avec μ sur $\mathcal{F}_0$.

On dira qu'une partie A de Ω est μ^*–mesurable si pour tout $E \subset \Omega$,

$$\mu^*(A \cap E) + \mu^*(A^c \cap E) = \mu^*(E),$$

ou de façon équivalente, compte tenu de la sous–additivité,

$$\mu^*(A \cap E) + \mu^*(A^c \cap E) \leq \mu^*(E),$$

pour tout $E \subset \Omega$. On désigne par $\mathcal{M}(\mu^*)$ la classe des ensembles μ^*–mesurable. Le résultat d'existence du Théorème est donc une conséquence des Lemmes 1.4.9 et 1.4.10 ci–dessous.

Lemme 1.4.7. *$\mathcal{M}(\mu^*)$ est une algèbre.*

Preuve Il est parfaitement clair que $\mathcal{M}(\mu^*)$ contient Ω et est fermé par passage au complémentaire. Soit A et B dans $\mathcal{M}(\mu^*)$ et $E \subset \Omega$. En utilisant successivement les propriétés $B \in \mathcal{M}(\mu^*)$, $A \in \mathcal{M}(\mu^*)$ et la sous–additivité de μ^*, on obtient

$$\begin{aligned}
\mu^*(E) &= \mu^*(B \cap E) + \mu^*(B^c \cap E) \\
&= \mu^*(A \cap B \cap E) + \mu^*(A^c \cap B \cap E) \\
&\quad + \mu^*(A \cap B^c \cap E) + \mu^*(A^c \cap B^c \cap E) \\
&\geq \mu^*(A \cap B \cap E) + \mu^*((A^c \cap B \cap E) \cup (A \cap B^c \cap E) \cup (A^c \cap B^c \cap E)) \\
&= \mu^*((A \cap B) \cap E) + \mu^*((A \cap B)^c \cap E).
\end{aligned}$$

Le lemme est établi. □

Lemme 1.4.8. *Si* $\{A_1, A_2, \ldots\}$ *est une suite finie ou dénombrable d'éléments disjoints de* $\mathcal{M}(\mu^*)$, *et* $E \subset \Omega$, *alors*

$$(*) \quad \mu^*\left(E \cap (\cup_k A_k)\right) = \sum_k \mu^*(E \cap A_k).$$

Preuve Considérons tout d'abord le cas d'une suite finie $\{A_1, A_2, \ldots, A_n\}$. Pour $n = 1$, il n'y a rien à montrer. Dans le cas $n = 2$, soit $A_1 \cup A_2 = \Omega$, et dans ce cas $A_2 = A_1^c$, donc $(*)$ découle de ce que $A_1 \in \mathcal{M}(\mu^*)$. Si $A_1 \cup A_2 \neq \Omega$, $(*)$ s'obtient en écrivant que $A_1 \in \mathcal{M}(\mu^*)$, avec E remplacé par $E \cap (A_1 \cup A_2)$.

Supposons $(*)$ vraie pour une suite de longueur $n-1$. $(*)$ étant également vraie pour $n = 2$, on a

$$\begin{aligned}
\mu^*\left(E \cap (\cup_{k=1}^n A_k)\right) &= \mu^*\left(E \cap \left(\cup_{k=1}^{n-1} A_k\right)\right) + \mu^*\left(E \cap A_n\right) \\
&= \sum_{k=1}^n \mu^*\left(E \cap A_k\right).
\end{aligned}$$

$(*)$ est donc établie pour n fini. Dans le cas dénombrable, par monotonie,

$$\begin{aligned}
\mu^*\left(E \cap (\cup_{k=1}^\infty A_k)\right) &\geq \mu^*\left(E \cap (\cup_{k=1}^n A_k)\right) \\
&= \sum_{k=1}^n \mu^*\left(E \cap A_k\right).
\end{aligned}$$

Il reste à faire tendre $n \to \infty$ et à utiliser la sous–additivité pour terminer la preuve. $\square$

Lemme 1.4.9. $\mathcal{M}(\mu^*)$ *est une* σ*–algèbre et la restriction de* μ^* *à* $\mathcal{M}(\mu^*)$ *est* σ*–additive.*

Preuve Soit $A_1, A_2, \ldots$ des ensembles disjoints de $\mathcal{M}(\mu^*)$, tels que $A = \cup_n A_n$. Par le Lemme 1.4.7, $F_n = \cup_{k=1}^n A_k$ appartient à $\mathcal{M}(\mu^*)$, donc pour tout $E \subset \Omega$, en utilisant aussi le Lemme 1.4.8 et la monotonie, on déduit que

$$\begin{aligned}
\mu^*(E) &= \mu^*(E \cap F_n) + \mu^*(E \cap F_n^c) \\
&\geq \sum_{k=1}^n \mu^*(E \cap A_k) + \mu^*(E \cap A^c).
\end{aligned}$$

En faisant tendre $n \to \infty$ et en utilisant à nouveau le Lemme 1.4.8, on obtient

$$\begin{aligned}\mu^*(E) &\geq \sum_{k=1}^{\infty} \mu^*(E \cap A_k) + \mu^*(E \cap A^c) \\ &= \mu^*(E \cap A) + \mu^*(E \cap A^c).\end{aligned}$$

Donc $A \in \mathcal{M}(\mu^*)$. $\mathcal{M}(\mu^*)$ étant une algèbre fermée pour les réunions dénombrables disjointes, il est facile d'en déduire que c'est une σ–algèbre. Finalement, la σ–additivité de μ^* sur $\mathcal{M}(\mu^*)$ découle du Lemme 1.4.8 avec $E = \Omega$. □

Lemme 1.4.10. *La classe $\mathcal{M}(\mu^*)$ contient l'algèbre $\mathcal{F}_0$.*

Preuve Soit $A \in \mathcal{F}_0$. Soit $E \subset \Omega$ et $\varepsilon > 0$. Il existe une suite $\{A_n\} \subset \mathcal{F}_0$ telle que $E \subset \cup_n A_n$ et $\mu^*(E) \geq \sum_n \mu_0(A_n) - \varepsilon$. Pour tout $n \geq 1$, $B_n = A_n \cap A$ et $C_n = A_n \cap A^c$ sont dans l'algèbre $\mathcal{F}_0$. Donc par la définition de μ^* et l'additivité de μ_0,

$$\begin{aligned}\mu^*(E \cap A) + \mu^*(E \cap A^c) &\leq \sum_n \mu_0(B_n) + \sum_n \mu_0(C_n) \\ &= \sum_n \mu_0(A_n) \\ &\leq \mu^*(E) + \varepsilon.\end{aligned}$$

En prenant la limite quand $\varepsilon \to 0$ dans l'inégailité entre les termes extrêmes, on obtient que $A \in \mathcal{M}(\mu^*)$. □

Théorème 1.4.11. *Soit deux mesures μ_1 et μ_2 définies sur une σ–algèbre $\mathcal{F} = \sigma(\mathcal{P})$, où $\mathcal{P}$ est un π–système. On suppose :*

(i) μ_1 et μ_2 sont σ–finies le long de $\mathcal{P}$

(ii) $\mu_1(A) = \mu_2(A)$, pour tout $A \in \mathcal{P}$.

Alors $\mu_1 = \mu_2$.

Preuve

a) Soit $A \in \mathcal{P}$ tel que $\mu_1(A) = \mu_2(A) < \infty$. Posons $\mathcal{L}_A = \{B \in \sigma(\mathcal{P}); \mu_1(A \cap B) = \mu_2(A \cap B)\}$. Il est assez facile de voir que $\mathcal{L}_A$ est un λ–système, et il est clair que $\mathcal{L}_A$ contient $\mathcal{P}$. Donc, par le Théorème 1.3.8, $\mathcal{L}_A \supset \sigma(\mathcal{P})$, d'où $\mathcal{L}_A = \sigma(\mathcal{P})$, puisque, par définition $\mathcal{L}_A \subset \sigma(\mathcal{P})$. Finalement $\mu_1(A \cap B) = \mu_2(A \cap B)$, pour tout $B \in \mathcal{F}$, et pour tout $A \in \mathcal{P}$ t.q. $\mu_1(A) = \mu_2(A) < \infty$.

b) Soit $\{A_n, n \geq 1\} \subset \mathcal{P}$ t.q. $\bigcup_1^\infty A_n = \Omega$, et $\mu_\alpha(A_n) < \infty$, pour tout $n \geq 1$, $\alpha = 1, 2$. Soit $B \in \mathcal{F}$. Comme $\mathcal{P}$ est un π–système qui contient les A_i, il contient les $A_i \cap A_j, \ldots$. Il résulte alors de la formule démontrée à l'Exercice 1.8.5 et de la partie a) que pour tout $n \geq 1$,

$$\mu_1\left(\bigcup_1^n (B \cap A_i)\right) = \mu_2\left(\bigcup_1^n (B \cap A_i)\right).$$

En faisant tendre $n \to \infty$, on obtient :

$$\mu_1(B) = \mu_2(B)$$

$\square$

On a le résultat d'approximation :

Proposition 1.4.12. *Soit $(\Omega, \mathcal{F}, \mu)$ un espace mesuré, et $\mathcal{F}_0$ une algèbre, t.q. $\sigma(\mathcal{F}_0) = \mathcal{F}$. Pour tout $A \in \mathcal{F}$ tel que $\mu(A) < \infty$ et pour tout $\varepsilon > 0$, il existe $A_0 \in \mathcal{F}_0$ t.q. $\mu(A \triangle A_0) \leq \varepsilon$.*

Preuve On sait (voir la démonstration du théorème 1.4.6) que

$$\mu(A) = \inf \sum_n \mu(A_n),$$

où l'infimum est pris sur les suites $\{A_n,\ n \in \mathbb{N}\} \subset \mathcal{F}_0$ t. q. $A \subset \bigcup_n A_n$. Donc il existe une suite $\{\bar{A}_n\} \subset \mathcal{F}_0$ telle que :

$$A \subset \bigcup_n \bar{A}_n \ \text{ et } \sum_n \mu(\bar{A}_n) \leq \mu(A) + \frac{\varepsilon}{2},$$

donc $\mu\left(\bigcup_n \bar{A}_n \setminus A\right) \leq \frac{\varepsilon}{2}$, et $\sum_n \mu(\bar{A}_n) < \infty$. Alors il existe N t.q. $\mu\left(\bigcup_{n>N} \bar{A}_n\right) \leq \sum_{n>N} \mu(\bar{A}_n) \leq \frac{\varepsilon}{2}$

$$\text{Or } \left(\bigcup_{n \leq N} \bar{A}_n\right) \triangle A \subset \left(\bigcup_n \bar{A}_n \backslash A\right) \cup \left(\bigcup_{n>N} \bar{A}_n\right)$$

$$\text{et } A_0 = \bigcup_{n \leq N} \bar{A}_n \in \mathcal{F}_0$$

□

On ne peut pas en général étendre μ à la σ–algèbre de toutes les parties de Ω, comme le montre l'exercice 1.8.6.

Définition 1.4.13. *On dira que la mesure* μ **est portée par** $A \in \mathcal{F}$ *si* $\mu(A^c) = 0$.

1.5 Mesure de Lebesgue

Nous allons maintenant définir la mesure de Lebesgue λ sur $(\mathbb{R}^d, \mathcal{B}_d)$. Considérons tout d'abord le cas $d = 1$.

Soit $\mathcal{S}$ la classe des intervalles de la forme $]a, b]$ $(-\infty \leq a \leq b < +\infty$, ou $]a, +\infty[, -\infty \leq a < +\infty)$. Soit $\mathcal{F}_0$ la plus petite algèbre contenant $\mathcal{S}$. Alors $\mathcal{B}_1 = \sigma(\mathcal{F}_0)$. $\mathcal{F}_0$ est la classe des réunions finies d'intervalles disjoints de la forme ci–dessus. Soit $A \in \mathcal{F}_0$. A possède une décomposition canonique

$$A = \bigcup_1^n I_k$$

où $\{I_k, k \leq n\} \subset \mathcal{S}$ et pour tout $k \neq \ell$, $I_k \cup I_\ell \notin \mathcal{S}$. On définit une application $\lambda : \mathcal{S} \to \bar{\mathbb{R}}_+$ par :

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

On étend λ à $\mathcal{F}_0$ en posant :

$$\lambda(A) = \sum_1^n \lambda(I_k); A \in \mathcal{F}_0$$

où $\{I_k, k \leq n\}$ est la décomposition canonique de A.

Proposition 1.5.1. λ *est une mesure sur* $\mathcal{F}_0$.

Preuve Il est clair que $\lambda(\emptyset) = 0$. Soit $\{A_n, n \geq 1\} \subset \mathcal{F}_0$ t.q. $A_k \cap A_\ell = \emptyset$ si $k \neq \ell$, et $A = \bigcup_1^\infty A_n \in \mathcal{F}_0$. Soit $A = \bigcup_1^p I_k$ la décomposition canonique de A. Alors $I_k = \bigcup_1^\infty (A_n \cap I_k)$, et il suffit de montrer que $\lambda(I_k) = \sum_1^\infty \lambda(A_n \cap I_k)$.

Or chaque $A_n \cap I_k$ est une réunion finie d'éléments de $\mathcal{S}$, et il suffit en fait de montrer que λ est "σ–additive sur $\mathcal{S}$", i.e. si $\{I_p,\ p \geq 1\} \subset \mathcal{S}$, $I_p \cap I_q = \emptyset$ si $p \neq q$ et $I = \bigcup_1^\infty I_p \in \mathcal{S}$, alors $\lambda(I) = \sum_1^\infty \lambda(I_p)$. Cette propriété résulte des deux lemmes qui suivent.

Lemme 1.5.2. *Si $\{]a_k, b_k],\ k \geq 1\}$ est une suite d'intervalles disjoints t.q. $\bigcup_k]a_k, b_k] \subset]a, b]$, alors*

$$\sum_k (b_k - a_k) \leq b - a$$

Preuve Plaçons nous tout d'abord dans le cas d'une suite finie de n intervalles. Le résultat est vrai pour $n = 1$. Supposons le vrai pour toute suite de longueur $n - 1$, et supposons que la numérotation des (a_k, b_k) est telle que $a_n > a_k$, pour tout $k \leq n - 1$ Alors :

$$\bigcup_1^{n-1}]a_k, b_k] \subset]a, a_n]$$

Par l'hypothèse de récurrence,

$$\sum_1^{n-1} (b_k - a_k) \leq a_n - a$$

$$\text{Donc } \sum_1^{n} (b_k - a_k) \leq b_n - a \leq b - a$$

Dans le cas d'une suite infinie, on a, par la première partie de la preuve,

$$\sum_1^{n} (b_k - a_k) \leq b - a, \quad \forall n,$$

d'où le résultat se déduit en faisant tendre n vers l'infini.

□

Lemme 1.5.3. *Si $]a, b] \subset \bigcup_k]a_k, b_k]$, alors $b - a \leq \sum_k (b_k - a_k)$*

Preuve Plaçons nous tout d'abord dans le cas d'une suite finie de n intervalles. Le résultat est vrai pour $n = 1$. Supposons le vrai pour une suite de longueur $n-1$. Etant donné $]a, b] \subset \bigcup_{k=1}^{n}]a_k, b_k]$, on suppose, pour fixer les idées, que la numérotation des (a_k, b_k) est telle que $a_n < b \leq b_n$. Si $a_n \leq a$, le résultat est immédiat. Dans le cas contraire,

$$]a, a_n] \subset \bigcup_{1}^{n-1}]a_k, b_k],$$

d'où par l'hypothèse de récurrence,

$$\begin{aligned} a_n - a &\leq \sum_{k=1}^{n-1} (b_k - a_k), \\ b - a \leq b_n - a &= b_n - a_n + a_n - a \\ &\leq \sum_{1}^{n} (b_k - a_k). \end{aligned}$$

Le résultat est démontré dans le cas d'une suite finie. Dans le cas d'une suite dénombrable avec $]a, b] \subset \bigcup_{1}^{\infty}]a_k, b_k]$, soit $\varepsilon \in]0, b-a[$. Les intervalles $]a_k, b_k + \varepsilon 2^{-k}[$ forment un recouvrement ouvert du compact $[\bar{a} + \varepsilon, \bar{b}]$, où

$$\bar{a} = \begin{cases} a & \text{si } a \in \mathbb{R}, \\ M & \text{si } a = -\infty; \end{cases} \qquad \bar{b} = \begin{cases} b & \text{si } b \in \mathbb{R}, \\ M & \text{si } b = +\infty. \end{cases}$$

et M est choisi tel que $\bar{b} \geq \bar{a} + \varepsilon$. Donc il existe n tel que :

$$[\bar{a} + \varepsilon, \bar{b}] \subset \bigcup_{1}^{n}]a_k, b_k + \varepsilon 2^{-k}[$$

et a fortiori :

$$]\bar{a} + \varepsilon, \bar{b}] \subset \bigcup_{1}^{n}]a_k, b_k + \varepsilon 2^{-k}]$$

D'après le résultat dans le cas d'une suite finie, pour tout $\varepsilon > 0$, $M > 0$,

$$\bar{b} - \varepsilon - \bar{a} \leq \sum_{1}^{\infty} (b_k - a_k) + \varepsilon.$$

$$\text{D'où} \quad b - a \leq \sum_{1}^{\infty} (b_k - a_k).$$

□

Il résulte de la Proposition 1.5.1, grâce au Théorème 1.4.6, que λ s'étend en une unique mesure sur $\mathcal{B} = \sigma(\mathcal{S})$. Pour traiter le cas $d > 1$, on peut soit refaire le raisonnement ci–dessus en remplaçant les intervalles par des pavés, soit utiliser la théorie des mesures produit – cf. Chapitre 2 ci–dessous. Nous admettrons donc l'existence de la mesure de Lebesgue λ sur $(\mathbb{R}^d, \mathcal{B}_d)$, caractérisée par le fait que si $A = \{x; x_i \in]a_i, b_i], i = 1 \cdots d\}$ avec $a_i, b_i \in \bar{\mathbb{R}}$, $a_i \leq b_i$, pour tout i

$$\lambda(A) = \prod_{i=1}^{d} (b_i - a_i)$$

(avec la convention $0.\infty = 0$)

Théorème 1.5.4. *Si $A \in \mathcal{B}_d$, alors $A + x = \{x + a; a \in A\} \in \mathcal{B}_d$, et*

$$\lambda(A) = \lambda(A + x), \quad \forall x \in \mathbb{R}^d$$

PREUVE Soit $\mathcal{G} = \{A \subset \mathbb{R}^d;\ A + x \in \mathcal{B}_d,\ \forall x \in \mathbb{R}^d\}$.

$\mathcal{G}$ est une σ–algèbre qui contient les pavés, donc $\mathcal{G} \supset \mathcal{B}_d$, mais $\mathcal{G} \subset \mathcal{B}_d$, d'où l'égalité.

Soit $x \in \mathbb{R}^d$ fixé. Posons $\mu(A) \stackrel{\triangle}{=} \lambda(A + x)$, $A \in \mathcal{B}_d$. μ est une mesure, μ et λ coïncident sur le π–système des pavés, donc $\mu = \lambda$. □

Il résulte de ce Théorème et de 1.4.4 (iv) que la mesure de Lebesgue de tout sous espace vectoriel de $\mathbb{R}^d$ de dimension $\leq d - 1$ est 0. Soit en effet V un tel sous espace. Alors il exise $x \in \mathbb{R}^d \setminus V$, t.q. $\bigcup_{\alpha \in \mathbb{R}} V + \alpha x \subset \mathbb{R}^d$, et les $V + \alpha x$ sont disjoints et tous de même mesure de Lebesgue. On déduit alors de 1.4.4 (iv) que cette mesure commune est 0.

1.6 Mesures sur $(\mathbb{R}^d, \mathcal{B}_d)$

On a le :

Théorème 1.6.1. *Soit μ une mesure sur $(\mathbb{R}^d, \mathcal{B}_d)$ t.q. $\mu(A) < \infty$ pour tout $A \in \mathcal{B}_d$, A borné.*

(i) Pour tout $A \in \mathcal{B}_d$ tel que $\mu(A) < \infty$ et $\varepsilon > 0$, il existe un fermé F et un ouvert O tels que

$$F \subset A \subset O$$
$$\text{et} \;\; \mu(O \backslash F) < \varepsilon.$$

(ii) Si $A \in \mathcal{B}_d$, et $\mu(A) < \infty$, alors :

$$\mu(A) = \sup_{\substack{K \subset A \\ K \text{ compact}}} \mu(K).$$

Remarque 1.6.2. *On ne pourrait en général pas trouver O et F tels que $O \subset A \subset F$, et $\mu(F - O) < \varepsilon$. Exemple : $\mu = \delta_x$, définie par*

$$\delta_x(B) = \begin{cases} 1 & si\ x \in B, \\ 0 & si\ x \notin B\ ; \end{cases} \quad et\ A = \{x\}.$$

Preuve

a) Montrons que (i) $\Rightarrow$ (ii). Si $\mu(A) < \infty$, il existe $A_0 \in \mathcal{B}_d$, A_0 borné, tel que $A_0 \subset A$ et $\mu(A \backslash A_0) < \varepsilon/2$. Or d'après (i), il existe $K \subset A_0$, K fermé (donc compact puisque borné) t.q. $\mu(A_0 \backslash K) < \varepsilon/2$; donc $\mu(A \backslash K) < \varepsilon$.

b) Montrons (i). Il suffit de montrer qu'il existe O ouvert $\supset A$ t.q. $\mu(O \backslash A) < \varepsilon$. La 2ème partie du résultat s'en déduit par passage au complémentaire.

Supposons tout d'abord que A est un pavé borné de la forme : $A = \{x; a_i < x_i \leq b_i,\ i = 1, \ldots, d\}$. Soit $O_n = \{x; a_i < x_i < b_i + \frac{1}{n}\}$. $\mu(O_1) < \infty$, $O_n \downarrow A$, donc par la Proposition 1.4.4 (ii), il existe n t.q. $\mu(O_n \backslash A) < \varepsilon$.

Les réunions finies de pavés (éventuellement non bornés ; $-\infty \leq a_i,\ b_i \leq \infty$) forment une algèbre $\mathcal{F}_0$ et $\sigma(\mathcal{F}_0) = \mathcal{B}_d$. Or tout pavé est une réunion dénombrable de pavés bornés. Il résulte donc de la preuve du Théorème d'extension 1.4.6 que pour tout $A \in \mathcal{B}_d$, il existe une suite $\{A_n,\ n \geq 1\}$ de pavés bornés t.q. (on utilise ici l'hypothèse $\mu(A) < \infty$)

$$A \subset \bigcup_1^\infty A_n, \quad \mu\left(\bigcup_1^\infty A_n \backslash A\right) < \varepsilon/2.$$

Mais pour tout $n \geq 1$, il existe un ouvert $O_n \supset A_n$ t.q. $\mu(O_n \backslash A_n) < \varepsilon/2^n$. Or

$$\bigcup_1^\infty O_n \backslash A \subset \bigcup_1^\infty (O_n \backslash A_n) \bigcup \bigcup_1^\infty (A_n \backslash A)$$

Donc $\mu\left(\bigcup_1^\infty O_n \backslash A\right) < \varepsilon$, et $\bigcup_1^\infty O_n$ est un ouvert qui contient A.

$\square$

1.7 Applications mesurables

Soit $(\Omega, \mathcal{F})$ et $(E, \mathcal{E})$ deux espaces mesurables, f une application de Ω dans E.

Définition 1.7.1. *L'application $f : \Omega \to E$ est dite $\mathcal{F}/\mathcal{E}$ mesurable (ou tout simplement mesurable) si pour tout $A \in \mathcal{E}$, $f^{-1}(A) \in \mathcal{F}$.*

$(G, \mathcal{G})$ désigne un troisième espace mesurable.

Théorème 1.7.2. *(i) Soit $\mathcal{C}$ une classe de parties de E, t.q. $\sigma(\mathcal{C}) = \mathcal{E}$, et $f : \Omega \to E$. Si $f^{-1}(C) \in \mathcal{F}$, pour tout $C \in \mathcal{C}$, alors f est $\mathcal{F}/\mathcal{E}$ mesurable.*

(ii) Si $f : \Omega \to E$ est $\mathcal{F}/\mathcal{E}$ mesurable, et $g : E \to G$ est $\mathcal{E}/\mathcal{G}$ mesurable, alors $g \circ f : \Omega \to G$ est $\mathcal{F}/\mathcal{G}$ mesurable.

Preuve

(i) D'après l'Exercice 1.8.7, $\{A \subset E; f^{-1}(A) \in \mathcal{F}\}$ est une σ–algèbre de parties de E, qui contient $\mathcal{C}$, donc $\sigma(\mathcal{C}) = \mathcal{E}$.

(ii) est évident.

$\square$

Considérons maintenant le cas $(E, \mathcal{E}) = (\mathbb{R}^d, \mathcal{B}_d)$.

Proposition 1.7.3. *$f : \Omega \to \mathbb{R}^d$ est $\mathcal{F}/\mathcal{B}_d$ mesurable si et seulement si*

$$\text{pour tout } 1 \leq i \leq d, \ \ f_i : \Omega \to \mathbb{R} \text{ est } \mathcal{F}/\mathcal{B} \text{ mesurable.}$$

Preuve On remarque que : $\{\omega; f_1(\omega) < x_1, \ldots, f_d(\omega) < x_d\} = \bigcap_{i=1}^{d} \{\omega; f_i(\omega) < x_i\}$. On applique le Théorème 1.7.2 (i). □

Proposition 1.7.4. *Si $f : \mathbb{R}^k \to \mathbb{R}^d$ est continue, alors elle est $\mathcal{B}_k/\mathcal{B}_d$ mesurable.*

Preuve Il suffit d'appliquer le Théorème 1.7.2 (i) avec pour $\mathcal{C}$ la classe des ouverts de $\mathbb{R}^d$. □

On va maintenant considérer des applications mesurables $f : (\Omega, \mathcal{F}) \to (\bar{\mathbb{R}}, \bar{\mathcal{B}})$, où

$$\bar{\mathbb{R}} = \mathbb{R} \cup \{+\infty\} \cup \{-\infty\}, \ \ \bar{\mathcal{B}} = \sigma(\mathcal{B}, \{+\infty\}, \{-\infty\}).$$

Dans la suite, pour signifier que $f : \Omega \to \bar{\mathbb{R}}$ est $\mathcal{F}/\bar{\mathcal{B}}$ mesurable, on dira que f est $\mathcal{F}$ mesurable.

Remarquons que si f et g sont des applications $\mathcal{F}$ mesurables de Ω dans $\bar{\mathbb{R}}$, $f+g$, $f\,g$ (si elles sont définies !), $\sup(f,g)$ et $\inf(f,g)$ sont mesurables (appliquer 1.7.4 et 1.7.2 (ii)).

Théorème 1.7.5. *Soit $\{f_n, n \geq 1\}$ des applications $\mathcal{F}$–mesurables, de Ω à valeurs dans $\bar{\mathbb{R}}$.*

(i) $\sup\limits_n f_n$, $\inf\limits_n f_n$, $\limsup\limits_n f_n$, $\liminf\limits_n f_n$ sont $\mathcal{F}$–mesurables.

(ii) Si $f(\omega) = \lim\limits_n f_n(\omega)$ existe pour tout $\omega \in \Omega$, alors f est $\mathcal{F}$–mesurable.

(iii) $\{\omega; f_n(\omega)$ converge dans $\bar{\mathbb{R}}\} \in \mathcal{F}$; $\{\omega; f_n(\omega)$ converge dans $\mathbb{R}\} \in \mathcal{F}$.

(iv) si $f : \Omega \to \bar{\mathbb{R}}$ est $\mathcal{F}$–mesurable, alors $\{\omega; f_n(\omega) \to f(\omega)\} \in \mathcal{F}$.

Preuve

(i) La première affirmation découle de ce que pour tout $x \in \bar{\mathbb{R}}$, $\{\sup\limits_n f_n \leq x\} = \bigcap\limits_n \{f_n \leq x\}$. Les autres résultats s'en déduisent en remarquant que $\inf_n f_n = -\sup_n(-f_n)$, $\limsup_n f_n = \inf_n \sup_{k\geq n} f_k$, et $\liminf_n f_n = -\limsup(-f_n)$.

(ii) L'hypothèse faite entraîne que $f(\omega) = \limsup_n f_n(\omega)$, pour tout ω.

(iii) $\{f_n$ converge dans $\bar{\mathbb{R}}\} = \{\limsup f_n = \liminf f_n\}$.
$\{f_n$ converge dans $\mathbb{R}\} = \{f_n$ converge dans $\bar{\mathbb{R}}\} \cap \{\limsup f_n \in \mathbb{R}\}$.

(iv) $\{f_n \to f\} = \{\limsup f_n - \liminf f_n = 0\} \cap \{f - \limsup f_n = 0\}$.

□

On appelle fonction ($\mathcal{F}$ mesurable) **étagée** une fonction de la forme :

$$f = \sum_{i=1}^{n} \alpha_k \mathbf{1}_{A_k}, \;\; n \geq 1, \; \alpha_k \in \mathbb{R}, A_k \in \mathcal{F}$$

.

Proposition 1.7.6. *La classe des applications $\mathcal{F}$ mesurables de Ω dans $\bar{\mathbb{R}}$ est la plus petite classe de fonctions de Ω dans $\bar{\mathbb{R}}$ qui contient les fonctions $\mathcal{F}$ mesurables étagées, et est fermée pour la convergence ponctuelle.*

Preuve Le fait qu'une fonction $\mathcal{F}$ mesurable étagée est $\mathcal{F}$ mesurable est évident. De plus, la classe des fonctions $\mathcal{F}$–mesurables est fermée pour la convergence ponctuelle, d'après le Théorème 1.7.5 (ii). Il reste à montrer que si f est $\mathcal{F}$ mesurable, alors il existe f_n étagées t.q. $f_n(\omega) \to f(\omega)$ quand $n \to \infty$, pour tout ω. On pose :

$$A_k^n = f^{-1}\left(\left[\frac{k}{n}, \frac{k+1}{n}\right[\right), \; n \geq 1, \; k \in \mathbb{Z}$$

$$\begin{aligned} f_n(\omega) = & \sum_{k=-n^2}^{n^2} \frac{k}{n} \mathbf{1}_{A_k^n}(\omega) - n\, \mathbf{1}_{f^{-1}([-\infty,-n[)}(\omega) \\ & + n \mathbf{1}_{f^{-1}([n+\frac{1}{n},+\infty])}(\omega) \end{aligned}$$

□

Remarque 1.7.7. *Si f est $\mathcal{F}$–mesurable à valeurs dans $\mathbb{R}_+$, alors on peut construire une suite de fonctions étagées qui converge en croissant vers f. On pose*

$$B_k^n = f^{-1}\left(\left[\frac{k}{2^n}, \frac{k+1}{2^n}\right[\right), n \geq 1, \; k \geq 0, \; \textit{et}$$

$$f_n(\omega) = \sum_{k=0}^{2^{2n}} \frac{k}{2^n} \mathbf{1}_{B_k^n}(\omega) + 2^n \mathbf{1}_{\{f^{-1}([2^n+2^{-n},+\infty))\}}.$$

Théorème 1.7.8. *(des classes monotones)*
Soit $\mathcal{H}$ un π–système de parties de Ω tel que $\sigma(\mathcal{H}) = \mathcal{F}$, et $\mathcal{L}$ une classe d'applications $\mathcal{F}$–mesurables de Ω à valeurs dans $\bar{\mathbb{R}}$, qui vérifie :

(i) $1 \in \mathcal{L}$; $1_A \in \mathcal{L}$, pour tout $A \in \mathcal{H}$.

(ii) Si $f, g \in \mathcal{L}$ et $\alpha, \beta \in \mathbb{R}$, alors $\alpha f + \beta g \in \mathcal{L}$.

(iii) Si $f_n \in \mathcal{L}$ et $f_n \uparrow f$ quand $n \to \infty$, alors $f \in \mathcal{L}$.

[resp.(iii)' à condition que $f(\omega) \in \mathbb{R}$, pour tout $\omega \in \Omega$]
[resp.(iii)" à condition qu'il existe c t.q. $|f(\omega)| \leq c$, pour tout $\omega \in \Omega$]
Alors :
(C) $\mathcal{L}$ contient toutes les applications $\mathcal{F}$–mesurables à valeurs $\bar{\mathbb{R}}$.
[resp.(C)' $\mathcal{L}$ contient toutes les applications $\mathcal{F}$–mesurables à valeurs $\mathbb{R}$].
[resp.(C)" $\mathcal{L}$ contient toutes les applications $\mathcal{F}$–mesurables et bornées à valeurs $\mathbb{R}$].

Preuve

a) $\mathcal{J} = \{A; 1_A \in \mathcal{L}\}$. $\mathcal{J}$ est un λ–système qui contient $\mathcal{H}$, donc d'après le théorème $\pi - \lambda$, $\mathcal{F} \subset \mathcal{J}$

b) Soit f une application $\mathcal{F}$–mesurable à valeurs $\bar{\mathbb{R}}_+$. f_n étant la suite de la Remarque 1.7.7, $f_{2^n} \in \mathcal{L}$ et $f_{2^n} \uparrow f$. Donc $f \in \mathcal{L}$.

c) Si f est une application $\mathcal{F}$ mesurable à valeurs dans $\bar{\mathbb{R}}$, alors $f = f^+ - f^-$, et nous venons de voir que f^+ et $f^- \in \mathcal{L}$. Donc $f \in \mathcal{L}$, d'après (ii). Les deux autres cas se traitent de façon similaire.

$\square$

On démontre de façon analogue le :

Corollaire 1.7.9. *Soit $\mathcal{L}$ une classe d'applications $\mathcal{F}$–mesurables de Ω à valeurs dans $\bar{\mathbb{R}}_+$, qui vérifie :*

(i') $1_A \in \mathcal{L}$, pour tout $A \in \mathcal{F}$.

(ii') Si $f, g \in \mathcal{L}$ et $\alpha, \beta \geq 0$, alors $\alpha f + \beta g \in \mathcal{L}$.

(iii) Si $f_n \in \mathcal{L}$ et $f_n \uparrow f$ quand $n \to \infty$, alors $f \in \mathcal{L}$.

[resp.(iii)' à condition que $f(\omega) \in \mathbb{R}$, pour tout $\omega \in \Omega$]
[resp.(iii)" à condition qu'il existe c t.q. $|f(\omega)| \leq c$, pour tout $\omega \in \Omega$]
Alors :
(C) $\mathcal{L}$ contient toutes les applications $\mathcal{F}$–mesurables à valeurs $\bar{\mathbb{R}}_+$.
[resp.(C)' $\mathcal{L}$ contient toutes les applications $\mathcal{F}$–mesurables à valeurs $\mathbb{R}_+$].
[resp.(C)" $\mathcal{L}$ contient toutes les applications $\mathcal{F}$–mesurables et bornées à valeurs $\mathbb{R}_+$].

Définition 1.7.10. *Soit $(\Omega, \mathcal{F}, \mu)$ un espace mesuré, $f : \Omega \to E$ une application $\mathcal{F}/\mathcal{E}$ mesurable. On définit alors la mesure sur $(E, \mathcal{E})$, image de μ par f : μf^{-1} par*

$$\mu f^{-1}(A) = \mu\left[f^{-1}(A)\right], \quad A \in \mathcal{E}.$$

1.8 Exercices

Dans tous ces exercices, $(\Omega, \mathcal{F}, \mu)$ est un espace mesuré où la mesure μ est σ–finie.

Exercice 1.8.1. *(i) Une σ–algèbre est à la fois un π– et un λ–système.*

(ii) Une classe qui est à la fois un π– et un λ–système est une σ–algèbre.

Exercice 1.8.2. *Justifier l'affirmation "Si $A \in \lambda(\mathcal{P})$, alors $\mathcal{G}_A$ est un λ–système." dans la preuve du Théorème 1.3.8.*

Exercice 1.8.3. *Justifier l'affirmation "$\mathcal{L}_A$ est un λ–système" dans la partie a) de la preuve du Théorème 1.4.11.*

Exercice 1.8.4. *Montrer que si $A_i \in \mathcal{F}$, $1 \le i \le n$,*

$$\mu\left(\bigcup_{i=1}^{n} A_i\right) \le \sum_{i=1}^{n} \mu(A_i).$$

Exercice 1.8.5. *Formule du crible. Même situation qu'à l'exercice précédent. Sous l'hypothèse $\mu(A_i) < \infty$ pour tout $1 \le i \le n$, montrer que*

$$\mu\left(\bigcup_{1}^{n} A_i\right) = \sum_{1}^{n} \mu(A_i) - \sum_{1 \le i < j \le n} \mu(A_i \cap A_j) + \cdots$$
$$+ (-1)^{n+1}\mu(A_1 \cap \cdots \cap A_n)$$

Exercice 1.8.6. *(Ensemble non mesurable) Considérons $\Omega = [0, 1[$, muni de sa tribu borélienne $\mathcal{B} = \mathcal{B}([0, 1[)$. On définit sur $(\Omega, \mathcal{B})$ la mesure de Lebesgue λ (cf. ci–dessus section 1.5) t.q.*

$$\lambda([a, b[) = b - a, \quad 0 \le a < b \le 1.$$

Pour $x, y \in [0, 1[$, on définit

$$x \oplus y = \begin{cases} x + y, & \text{si } x + y \in [0, 1[\\ x + y - 1, & \text{si } x + y \ge 1 \end{cases}$$

si $A \subset [0,1[, x \in [0,1[$ *on pose* $A \oplus x = \{a \oplus x, a \in A\}$.

On dira que x *et* y *sont équivalents* ($x \sim y$) *si il existe* $r \in \mathbb{Q} \cap [0,1[$ *t.q.* $x \oplus r = y$. *On admettra (on a besoin pour cela de l'axiome du choix) qu'il existe un sous–ensemble* H *de* $[0,1[$ *qui contient exactement un représentant de chaque classe d'équivalence.*

1. *Montrer (en utilisant le Théorème* $\pi-\lambda$*) que pour tout* $A \in \mathcal{B}$, $A \oplus x \in \mathcal{B}$ *et* $\lambda(A \oplus x) = \lambda(A)$.
2. H *étant l'ensemble défini ci–dessus, montrer que la collection* $\{H \oplus r,\ r \in \mathbb{Q} \cap [0,1]\}$ *est une collection dénombrable d'ensembles disjoints, dont la réunion est* $[0,1[$.
3. *Montrer par l'absurde que* $H \notin \mathcal{B}$.

Exercice 1.8.7. *Soit* $\varphi : \Omega \to E$. *Si* $\mathcal{C}$ *est une classe de parties de* E, *on note* $\varphi^{-1}(\mathcal{C}) = \{\varphi^{-1}(C); C \in \mathcal{C}\}$. *Si* $\mathcal{A}$ *est une classe des parties de* Ω, *on note* $\varphi(\mathcal{A}) = \{B \subset E; \varphi^{-1}(B) \in \mathcal{A}\}$. *Montrer que*

1. *Si* $\mathcal{C}$ *est une* σ*–algèbre, alors* $\varphi^{-1}(\mathcal{C})$ *est une* σ*–algèbre.*
2. *Si* $\mathcal{A}$ *est une* σ*–algèbre, alors* $\varphi(\mathcal{A})$ *est une* σ*–algèbre.*
3. $\sigma(\varphi^{-1}(\mathcal{C})) = \varphi^{-1}(\sigma(\mathcal{C}))$.
4. $\varphi : \Omega \to \mathbb{R}$ *est mesurable ssi* $\{\varphi < x\} \in \mathcal{F}$, *pour tout* $x \in \mathbb{R}$. *Même énoncé avec* $\{\varphi < x\} \in \mathcal{F}$ *remplacé par* $\{\varphi \leq x\} \in \mathcal{F}$.

Chapitre 2

Intégration

Dans ce chapitre, on suppose donné un espace mesuré $(\Omega, \mathcal{F}, \mu)$.

2.1 Propriété vérifiée presque partout

On dira qu'une propriété $P(\omega)$ est vérifiée presque partout si il existe $N \in \mathcal{F}$ avec $\mu(N) = 0$, t.q. pour tout $\omega \notin N$, $P(\omega)$ est vérifiée.

Par exemple, si $\{f_n, n \geq 1\}$, f et g sont des applications de Ω à valeurs dans $\bar{\mathbb{R}}$ ou $\mathbb{R}^d$, on pose la

Définition 2.1.1. *f et g sont dites égales presque partout (en abrégé p.p.) s'il existe $N \in \mathcal{F}$, avec $\mu(N) = 0$, t.q. pour tout $\omega \notin N$, $f(\omega) = g(\omega)$.*

On écrira

$$f(\omega) = g(\omega) \quad p.p.$$

ou plus simplement

$$f = g \quad p.p.$$

Notons que "$f = g$ p.p." est une relation d'équivalence (cf. Chapitre 3).

Définition 2.1.2. *On dit que f_n converge vers f p.p. s'il existe $N \in \mathcal{F}$ avec $\mu(N) = 0$ t.q.*

$$\text{pour tout } \omega \notin N, \; f_n(\omega) \to f(\omega)$$

On écrira $f_n(\omega) \to f(\omega)$ p.p. ou plus simplement $f_n \to f$ p.p.

Théorème 2.1.3. *Soit $\mathcal{G} \subset \mathcal{F}$ une tribu, $\{f_n, n \geq 1\}$ une suite d'applications $\mathcal{G}$–mesurables, et f une application de Ω dans $\bar{\mathbb{R}}$ t.q.*

$$f_n \to f \;\; p.p.$$

Alors il existe une application $\mathcal{G}$–mesurable $\bar{f}$ t.q.

$$f = \bar{f} \;\; p.p.$$

PREUVE $\Gamma = \{\omega; f_n(\omega) \text{ converge }\} \in \mathcal{G}$, d'après le Théorème 1.7.5 (iii).

$$\text{Posons } \bar{f}(\omega) = \begin{cases} \lim \; f_n(\omega), & \text{si } \omega \in \Gamma; \\ 0, & \text{si } \omega \notin \Gamma. \end{cases}$$

Alors $(\bar{f}) = f$ p.p., et $\bar{f}$ est $\mathcal{G}$–mesurable, puisque

$$\forall x > 0, \{\bar{f} < x\} = \Gamma^c \bigcup (\Gamma \cap \{\overline{\lim} f_n < x\}) \in \mathcal{G};$$
$$\forall x \leq 0, \{\bar{f} < x\} = \Gamma \bigcap \{\overline{\lim} f_n < x\} \in \mathcal{G}.$$

□

Remarque 2.1.4. *Presque partout veut dire en dehors d'un ensemble $N \in \mathcal{F}$, t.q. $\mu(N) = 0$. Etant donné N un tel ensemble mesurable de mesure nulle, et $N' \subset N$, $N' \notin \mathcal{F}$, peut–on dire que N' est de mesure nulle ? Autrement dit, avec la notation $\mathcal{F} \vee \mathcal{G} = \sigma(\mathcal{F} \cup \mathcal{G})$, et étant donnée*

$$\bar{\mathcal{F}} = \mathcal{F} \vee \sigma(N \subset \Omega; \exists \; A \in \mathcal{F}, N \subset A \;\; et \;\; \mu(A) = 0)$$

peut–on étendre μ à $\bar{\mathcal{F}}$? La réponse est oui : il existe une unique extension de μ à $\bar{\mathcal{F}}$, que l'on note encore μ. L'espace mesurable $(\Omega, \bar{\mathcal{F}}, \mu)$ est dit **complet**. *Noter que la tribu complétée $\bar{\mathcal{F}}$ dépend de μ.*

2.2 Intégrale des fonctions non négatives

Dans cette section, f, g désignent des applications $\mathcal{F}$–mesurables à valeurs dans $\bar{\mathbb{R}}_+$.

On va définir :

$$\int f d\mu = \int_\Omega f(\omega) d\mu(\omega) = \int_\Omega f(\omega) \mu(d\omega)$$

On appelle **partition finie** de Ω une collection finie $\{A_i; 1 \leq i \leq n\} \subset \mathcal{F}$ telle que :

$$\bigcup_1^n A_i = \Omega$$
$$A_i \cap A_j = \emptyset, \text{ dès que } i \neq j.$$

on définit :

$$\int f d\mu = \sup \left[\sum_{i=1}^{n} \left(\inf_{\omega \in A_i} f(\omega) \right) \cdot \mu(A_i) \right],$$

où :

(i) la somme est évaluée avec la convention : $0 \cdot \infty = \infty \cdot 0 = 0$;

(ii) le sup est pris sur toutes les partitions finies $\{A_i, 1 \leq i \leq n\}$ de Ω.

Théorème 2.2.1. *(i) Etant donnée $\{A_i; 1 \leq i \leq n\}$ une partition finie de Ω, $\{x_i, 1 \leq i \leq n\} \subset \mathbb{R}_+$, si $f = \sum_1^n x_i \mathbf{1}_{A_i}$,*

$$\textit{Alors} \int f d\mu = \sum_{i=1}^{n} x_i \mu(A_i)$$

(ii) Si $0 \leq f(\omega) \leq g(\omega)$ pour tout ω, alors $\int f d\mu \leq \int g d\mu$.

(iii) Si $0 \leq f_n(\omega) \uparrow f(\omega)$ pour tout ω, alors

$$\int f_n d\mu \uparrow \int f d\mu.$$

(iv) Si $\alpha, \beta \geq 0$, f et g sont non négatives,

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

Preuve

(i) Soit $\{B_j; 1 \leq j \leq m\}$ une partition finie de Ω, $y_j = \inf_{\omega \in B_j} f(\omega)$. Si $A_i \cap B_j \neq \emptyset$, alors $y_j \leq x_i$.

$$\text{Donc } \sum_j y_j \mu(B_j) = \sum_{j,i} y_j \mu(A_i \cap B_j) \leq \sum_{j,i} x_i \mu(A_i \cap B_j) = \sum_i x_i \mu(A_i)$$

Donc le sup est atteint avec la partition $\{A_i; i \leq n\}$

(ii) Résulte directement de la définition.

(iii) D'après (ii), $\int f_n d\mu \uparrow$, $\int f_n d\mu \leq \int f d\mu$. Il suffit donc de montrer que $\int f d\mu \leq \lim_n \int f_n d\mu$, ou encore, pour toute partition $\{A_i, 1 \leq i \leq k\}$ de Ω,

$$(*) \qquad \sum_1^k v_i \mu(A_i) \leq \lim_n \int f_n d\mu, \quad \text{où } v_i = \inf_{\omega \in A_i} f(\omega).$$

Pour établir $(*)$, il suffit de considérer les partitions telles que

$$\sum_1^k v_i \mu(A_i) > 0,$$

et de montrer que :

$$x < \sum_1^k v_i \mu(A_i) \ \Rightarrow \ x < \lim_n \int f_n d\mu.$$

Soit donc $x < \sum_1^k v_i \mu(A_i)$. Alors il existe $\{u_i; i \leq k\}$ tels que

$$\text{soit } 0 = u_i = v_i, \quad \text{soit } 0 < u_i < v_i; \quad \text{et } x < \sum_1^k u_i \mu(A_i)$$

$$\begin{aligned} \text{Si } \omega \in A_i, \text{ soit } & f_n(\omega) \uparrow f(\omega) \geq v_i = u_i = 0 \\ \text{soit } & f_n(\omega) \uparrow f(\omega) \geq v_i > u_i > 0 \end{aligned}$$

Donc si $A_{in} = \{\omega \in A_i; f_n(\omega) \geq u_i\}$, $A_{in} \uparrow A_i$, d'où $\mu(A_{in}) \uparrow \mu(A_i)$.

$$\begin{aligned}
\int f_n d\mu &\geq \sum_1^k \left(\inf_{\omega \in A_{in}} f_n(\omega) \right) \mu(A_{in}) + 0 \\
&\geq \sum_1^k u_i \mu(A_{in}), \text{ donc :} \\
\lim_n \int f_n d\mu &\geq \sum_1^k u_i \mu(A_i) > x.
\end{aligned}$$

(iv) Il suffit de montrer l'égalité avec f et g étagées. On passe ensuite à la limite avec les suites construites à la Remarque 1.7.7, en utilisant (iii). Soit $f = \sum_{i=1}^n x_i \mathbf{1}_{A_i}$, $g = \sum_{j=1}^n y_j \mathbf{1}_{B_j}$. Sans restreindre la généralité, on peut supposer que $\{A_i; 1 \leq i \leq n\}$ et $\{B_j; 1 \leq j \leq m\}$ forment des partitions de Ω.

$$\alpha f + \beta g = \sum_{i=1}^n \sum_{j=1}^m (\alpha x_i + \beta y_j) \mathbf{1}_{A_i \cap B_j}$$

$\{A_i \cap B_j; 1 \leq i \leq n, 1 \leq j \leq m\}$ est encore une partition de Ω, et

$$\sum_{j=1}^m \mu(A_i \cap B_j) = \mu(A_i), \; \sum_{i=1}^n \mu(A_i \cap B_j) = \mu(B_j), \text{ d'où :}$$

$$\begin{aligned}
\int (\alpha f + \beta g) d\mu &= \sum_{i=1}^n \sum_{j=1}^m (\alpha x_i + \beta y_j) \mu(A_i \cap B_j) \\
&= \alpha \sum_{i=1}^n x_i \mu(A_i) + \beta \sum_{j=1}^m y_j \mu(B_j) \\
&= \alpha \int f d\mu + \beta \int g d\mu.
\end{aligned}$$

□

Théorème 2.2.2. *Soit f et g deux fonctions mesurables de Ω à valeurs dans $\mathbb{R}_+$.*

(i) Si $f = 0$ μ *p.p. , alors* $\displaystyle\int f d\mu = 0$.

(ii) Si $\mu(\{\omega; f(\omega) > 0\}) > 0$, *alors* $\displaystyle\int f d\mu > 0$.

(iii) Si $\displaystyle\int f d\mu < \infty$, *alors* $f < \infty$ μ *p.p.*

(iv) Si $f = g$ μ *p.p., alors* $\displaystyle\int f d\mu = \int g d\mu$

(v) Si $f \leq g$ μ *p.p., alors* $\displaystyle\int f d\mu \leq \int g d\mu$.

Preuve

(i) résulte de ce que pour tout $A \in \mathcal{F}$,

$$\text{soit } A \cap \{f = 0\} \neq \emptyset, \text{ et alors } \inf_{\omega \in A} f(\omega) = 0;$$
$$\text{soit } A \cap \{f = 0\} = \emptyset, \text{ et alors } \mu(A) = 0.$$

(ii) Posons $A_n = \{f > \frac{1}{n}\}$, $A_n \uparrow \{f > 0\}$. Puisque $\mu(\{f > 0\}) > 0$, il existe n t.q. $\mu(A_n) > 0$,

$$\int f d\mu \geq \frac{1}{n}\mu(A_n) + 0\mu(A_n^c) > 0.$$

(iii) $+\infty > \displaystyle\int f d\mu \geq +\infty \times \mu(f = +\infty)$, donc $\mu(f = +\infty) = 0$.

(iv) Soit

$$h(\omega) = \begin{cases} 0, & \text{si } f(\omega) = g(\omega)\,; \\ +\infty, & \text{sinon.} \end{cases}$$

pour tout $\omega \in \Omega$, $f(\omega) \leq g(\omega) + h(\omega)$. D'après 2.2.1 (ii), (iv), et (i) ci–dessus,

$$\int f d\mu \leq \int (g + h) d\mu = \int g d\mu + \int h d\mu = \int g d\mu.$$

On montre de même $\displaystyle\int g d\mu \leq \int f d\mu$.

(v) On définit maintenant h par

$$h(\omega) = \begin{cases} 0, & \text{si } f(\omega) \leq g(\omega)\,; \\ +\infty, & \text{sinon.} \end{cases}$$

On a alors la première série d'inégalités de (iv).

□

2.3 Intégrale des fonctions de signe quelconque.

Etant donnée f une application mesurable à valeurs dans $\bar{\mathbb{R}}$, on utilise la décomposition :

$$f = f^+ - f^-.$$

On définit $\displaystyle\int f^+ d\mu$ et $\displaystyle\int f^- d\mu$. Trois cas se présentent :

a) $\displaystyle\int f^+ d\mu = \int f^- d\mu = +\infty$, et alors l'intégrale de f ne peut pas être définie.

b) L'une au moins des deux quantités $\displaystyle\int f^+ d\mu$ et $\displaystyle\int f^- d\mu$ est finie. f est alors dite **quasi intégrable**, et on pose

$$\int f d\mu = \int f^+ d\mu - \int f^- d\mu \quad (\in \bar{\mathbb{R}})$$

c) Les deux quantités $\displaystyle\int f^+ d\mu$ **et** $\displaystyle\int f^- d\mu$ sont finies. f est alors dite **intégrable**, et on définit $\displaystyle\int f d\mu$ comme dans le cas quasi–intégrable. Alors $\displaystyle\int f d\mu$ est un nombre réel (fini).

Remarquons que $|f| = f^+ + f^-$, donc le Théorème 2.2.1 (iv) justifie que l'on pose la

Définition 2.3.1. *Une application mesurable f à valeurs dans $\bar{\mathbb{R}}$ est dite μ–**intégrable** si :*

$$\int |f| d\mu < \infty.$$

On peut remarquer que **si** f est quasi–intégrable, et si $|\int f d\mu| < \infty$, alors f est intégrable. Mais **on n'a pas le droit d'écrire** $\int f d\mu$ **avant d'avoir vérifié que** f **est intégrable (ou au moins quasi–intégrable)**.

Théorème 2.3.2. *(i)* **Monotonie** *Si f et g sont quasi–intégrables, et $f \leq g$ p.p., alors* $\int f d\mu \leq \int g d\mu$.

(ii) **Linéarité** *Si f et g sont intégrables, $\alpha, \beta \in \mathbb{R}$, alors $\alpha f + \beta g$ est intégrable, et*

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

Preuve

(i) résulte du Théorème 2.2.2 (v), car $f \leq g$ p.p. $\Rightarrow f^+ \leq g^+$ p.p. et $f^- \geq g^-$ p.p.

(ii) Il résulte du Théorème 2.2.1

$$\begin{aligned} \int |\alpha f + \beta g| d\mu &\leq \int (|\alpha| \cdot |f| + |\beta| \cdot |g|)\, d\mu \\ &= |\alpha| \int |f| d\mu + |\beta| \int |g| d\mu < \infty. \end{aligned}$$

Donc $\alpha f + \beta g$ est intégrable. De plus,

$$\alpha \int f d\mu = \int \alpha f d\mu.$$

Il reste à établir l'égalité avec $\alpha = \beta = 1$.

$$\begin{gathered} (f+g)^+ - (f+g)^- = f + g = f^+ - f^- + g^+ - g^- \\ (f+g)^+ + f^- + g^- = f^+ + g^+ + (f+g)^- \end{gathered}$$

$$\int (f+g)^+ d\mu + \int f^- d\mu + \int g^- d\mu = \int f^+ d\mu + \int g^+ d\mu + \int (f+g)^- d\mu,$$

d'après le Théorème 2.2.1 (iv). Le résultat s'obtient en regroupant convenablement les termes. □

Corollaire 2.3.3. *Si f est quasi–intégrable, g intégrable, $\alpha, \beta \in \mathbb{R}$, alors $\alpha f + \beta g$ est quasi–intégrable, et*

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

PREUVE Si $|\int f d\mu| < \infty$, alors f est intégrable, et le résultat découle du Théorème 2.3.2. Supposons donc que $|\int f d\mu| = +\infty$. Pour fixer les idées, supposons que $\int f d\mu = +\infty$ et $\alpha > 0$. (les autres cas se traitent de façon analogue). Le second membre de l'égalité à établir vaut $+\infty$. Or $\alpha f + \beta g$ n'est pas intégrable, sinon $f = \dfrac{1}{\alpha}(\alpha f + \beta g) - \dfrac{\beta}{\alpha} g$ le serait, et $(\alpha f + \beta g)^- \leq \alpha f^- + |\beta||g|$, qui est intégrable. Donc $\alpha f + \beta g$ est quasi–intégrable, et $\int (\alpha f + \beta g) d\mu = +\infty$, d'où le résultat. □

Corollaire 2.3.4. *f et g désignent des applications mesurables à valeurs dans $\mathbb{R}$.*

(i) Si f est intégrable, $|\int f d\mu| \leq \int |f| d\mu$. En particulier si f et g sont intégrables, $|\int f d\mu - \int g d\mu| \leq \int |f - g| d\mu$.

(ii) Si $f = 0$ p.p., alors f est intégrable et $\int f d\mu = 0$

(iii) Si f est intégrable et $g = f$ p.p., alors g est intégrable et $\int f d\mu = \int g d\mu$.

PREUVE

(i) résulte de $|\int f d\mu| \leq \int f^+ d\mu + \int f^- d\mu = \int |f| d\mu$.

(ii) Il résulte de l'hypothèse que $f^+ = 0$ et $f^- = 0$ p. p. La suite du raisonnement est facile.

(iii) D'après (ii) $g - f$ est intégrable et $\int (g - f) d\mu = 0$. Il reste à appliquer le Théorème 2.3.2 (ii).

□

Théorème 2.3.5. *(de convergence monotone – Beppo–Levi) Soit f_n, f des applications mesurables à valeurs dans $\bar{\mathbb{R}}$.*

(i) Si $\displaystyle\int f_1^- d\mu < \infty$, et $f_n \uparrow f$ p.p., alors $\displaystyle\int f_n d\mu \uparrow \int f d\mu$.

(ii) Si $\displaystyle\int f_1^+ d\mu < \infty$, et $f_n \downarrow f$ p.p., alors $\displaystyle\int f_n d\mu \downarrow \int f d\mu$.

Preuve (ii) résulte de (i) par multiplication par -1. Montrons (i). On pose $g_n = f_n + f_1^-$, $g_n \geq 0$ p.p., $g_n \uparrow g$ p.p., avec $g = f + f_1^-$. On pose alors $\tilde{g}_n(\omega) = \sup_{k\leq n} g_k(\omega)$, $\tilde{g}(\omega) = \lim_{n\to\infty} \tilde{g}_n(\omega)$. Il résulte du Théorèmes 2.2.1 (iii) : $\displaystyle\int \tilde{g}_n d\mu \uparrow \int \tilde{g} d\mu$. Mais il n'est pas difficile de montrer que $g_n = \tilde{g}_n$ p. p., $g = \tilde{g}$ p. p. Il résulte alors du Corollaire 2.3.4 (iii) que $\displaystyle\int g_n d\mu \uparrow \int g d\mu$. Le résultat découle alors du Corollaire 2.3.3. □

Remarque 2.3.6. *(i) s'applique en particulier lorsque $f_n \geq 0$, $f_n \uparrow f$.*

(ii) s'applique en particulier lorsque $f_n \leq 0$, $f_n \downarrow f$.

Théorème 2.3.7. *("Lemme de Fatou") Soit $\{f_n, n \geq 1\}$ une suite d'applications mesurables à valeurs dans $\bar{\mathbb{R}}$.*

(i) S'il existe g intégrable t.q. $f_n \geq g$ p.p., pour tout n, alors :

$$\int (\liminf_n f_n) d\mu \leq \liminf_n \int f_n d\mu.$$

(ii) S'il existe g intégrable t.q. $f_n \leq g$ p.p., pour tout n, alors :

$$\int (\limsup_n f_n) d\mu \geq \limsup_n \int f_n d\mu.$$

Tous les intégrands qui apparaissent dans les intégrales ci–dessus sont quasi–intégrables.

Preuve (ii) résulte de (i) par multiplication par -1. Montrons (i). Posons $g_n(\omega) = \inf\limits_{m\geq n} f_m(\omega)$. On a $g \leq g_n \leq f_n$ p.p., et $g_n \uparrow \liminf\limits_n f_n$. Donc $g_1^- \leq g^-$,

g_1^- est intégrable, et d'après le Théorème 2.3.5 (i),

$$\int (\lim_n g_n) d\mu = \lim_n \int g_n d\mu$$
$$\int (\liminf_n f_n) d\mu = \lim_n \int g_n d\mu$$
$$\leq \liminf_n \int f_n d\mu,$$

où la dernière inégalité s'obtient en prenant la lim inf dans l'inégalité $\int g_n d\mu \leq \int f_n d\mu$. □

Remarque 2.3.8. *(i) est vrai en particulier dès que $f_n \geq 0$ p.p., pour tout n, et (ii) est vrai en particulier dès que $f_n \leq 0$ p.p., pour tout n.*

Théorème 2.3.9. *(de convergence dominée de Lebesgue) Soit $\{f_n, n \geq 1\}$ et f des applications mesurables de Ω à valeurs dans $\bar{\mathbb{R}}$, t.q.*

(i) $f_n \to f$ p.p.

(ii) Il existe g **intégrable** *t.q. $|f_n(\omega)| \leq g(\omega)$ p.p., pour tout $n \geq 1$. Alors f_n et f sont intégrables, et*

$$\int f_n d\mu \to \int f d\mu$$

Preuve Posons $N_n = \{|f_n| > g\}$, $N = \{f_n \not\to f\}$. Alors $M = N \cup (\bigcup_1^\infty N_n)$ vérifie $\mu(M) = 0$, et pour $\omega \notin M$, $|f(\omega)| \leq g(\omega)$. D'après le Théorème 2.2.2 (v), f_n et f sont intégrables.

Posons $g_n = |f - f_n|$; $g_n \leq 2g$ p.p., et grâce au Théorème 2.3.7 (ii)

$$0 \leq \liminf_n \int g_n d\mu \leq \limsup_n \int g_n d\mu \leq \int (\limsup g_n) d\mu = 0.$$

Donc $\int g_n d\mu \to 0$. Or $\left| \int f d\mu - \int f_n d\mu \right| \leq \int g_n d\mu$. □

! Attention ! g doit être indépendante de n !

On déduit des théorèmes de convergence ci–dessus les propositions :

Proposition 2.3.10. *Soit* $\{f_n,\ n \geq 1\}$ *des applications mesurables à valeurs dans* $\bar{\mathbb{R}}$.

(i) si $f_n \geq 0$ *p.p. pour tout* $n \geq 1$, *alors* $\displaystyle\int \left(\sum_n f_n\right) d\mu = \sum_n \int f_n d\mu$

(ii) Si $\displaystyle\sum_n f_n$ *converge p.p., et* $\displaystyle|\sum_1^n f_k| \leq g$ *p.p., pour tout* n, *où* g *est* **intégrable**, *alors* $\displaystyle\sum_n f_n$ *est intégrable, et*

$$\int \sum_n f_n d\mu = \sum_n \int f_n d\mu.$$

(iii) Si $\displaystyle\sum_n \int |f_n| d\mu < \infty$, *alors* $\displaystyle\sum_n f_n$ *converge p.p., la somme de la série est intégrable et*

$$\int \sum_n f_n d\mu = \sum_n \int f_n d\mu.$$

Preuve

(i) $\displaystyle\int (\sum_1^N f_n) d\mu = \sum_1^N \int f_n d\mu$, et grâce au Théorème de convergence monotone, on peut faire tendre $N \to \infty$.

(ii) résulte immédiatement du Théorème de convergence dominée.

(iii) l'hypothèse entraîne que $\displaystyle\sum_1^\infty |f_n| < \infty$ p.p., donc $\displaystyle\sum_n f_n$ converge p.p., et $\displaystyle|\sum_1^n f_k| \leq \sum_1^\infty |f_k|$ qui est intégrable. On applique alors (ii)

□

Proposition 2.3.11. *Soit* $f : \Omega \times]a, b[\to \bar{\mathbb{R}}$ *t.q. pour tout* $t \in]a, b[$, $\omega \to f(\omega, t)$ *est mesurable.*

(i) Supposons que ω *p.p.,* $t \to f(\omega, t)$ *est continue au point* $t_0 \in]a, b[$ *et qu'il existe* g *intégrable t.q.* $|f(\omega, t)| \leq g(\omega)$ *p.p., pour tout* $t \in]a, b[$, *alors* $t \to \displaystyle\int f(t, \omega) d\mu(\omega)$ *est continue au point* t_0.

(ii) Supposons que ω p.p., $t \to f(\omega,t)$ est dérivable en tout $t \in]a,b[$, que pour tout $t \in]a,b[$, $f(\cdot,t)$ est intégrable et $\left|\frac{\partial f}{\partial t}(\omega,t)\right| \leq g(\omega)$ p.p., où g est intégrable. Alors $t \to \displaystyle\int f(\omega,t)d\mu(\omega)$ est dérivable en tout $t \in]a,b[$, et

$$\frac{d}{dt}\int f(\omega,t)d\mu(\omega) = \int \frac{\partial f}{\partial t}(\omega,t)d\mu(\omega).$$

Preuve

(i) résulte directement du Théorème de convergence dominée.

(ii)

$$\begin{aligned}\frac{1}{h}\left[\int f(\omega,t+h)d\mu(\omega) - \int f(\omega,t)d\mu(\omega)\right]&\\ = \int \frac{f(\omega,t+h)-f(\omega,t)}{h}d\mu(\omega)&\end{aligned}$$

$$\text{Or } \frac{f(\omega,t+h)-f(\omega,t)}{h} \to \frac{\partial f}{\partial t}(\omega,t) \text{ p.p., quand } h \to 0,$$

$$\text{et } \left|\frac{f(\omega,t+h)-f(\omega,t)}{h}\right| \leq g(\omega) \text{ p.p.}$$

par le théorème des accroissements finis. On conclut à nouveau à l'aide du Théorème de convergence dominée.

$\square$

Etant donné f mesurable et intégrable et $A \in \mathcal{F}$ on définit :

$$\int_A f d\mu := \int \mathbf{1}_A f d\mu.$$

Remarque 2.3.12. *(lien avec l'intégrale de Riemann)*

Soit $a,b \in \mathbb{R}$, $a < b$, et $f \in C(\mathbb{R})$. Alors f est mesurable et $|\mathbf{1}_{]a,b]}f| \leq \sup_{x\in[a,b]} |f(x)| \times \mathbf{1}_{]a,b]}$, qui est intégrable, donc la construction ci–dessus permet de définir $\displaystyle\int_{]a,b]} f(x)dx$, intégrale de $\mathbf{1}_{]a,b]}f$ par rapport à la mesure de Lebesgue λ sur $(\mathbb{R},\mathcal{B})$.

Par ailleurs, on sait dans ce cas définir l'intégrale de Riemann $\int_a^b f(x)dx = \lim_{n\to\infty} \sum_{i=0}^{n-1} f(x_i^n)(x_{i+1}^n - x_i^n)$ *,où* $a = x_0^n < x_1^n < \cdots < x_n^n = b$ *vérifie* $\sup_i (x_{i+1}^n - x_i^n) \to 0$ *quand* $n \to \infty$.

On vérifie aisément que les deux intégrales coïncident. Remarquons que les fonctions $\mathbf{1}_{[a,b[}\ f$, $\mathbf{1}_{]a,b]}\ f$, $\mathbf{1}_{]a,b[}\ f$ *et* $\mathbf{1}_{[a,b]} f$ *sont* $\lambda-$ *p.p. égales. Donc on peut adopter, pour la valeur commune de leurs intégrales, la notation unique* $\int_a^b f(x)dx$. *Il n'en serait pas de même si l'on intégrait par rapport à une mesure* μ *sur* $(\mathbb{R}, \mathcal{B})$ *qui ne vérifie pas* $\mu(\{x\}) = 0, \forall x \in \mathbb{R}$. *Alors la notation* $\int_a^b$ *n'aurait aucun sens, et serait à proscrire : il faudrait absolument préciser s'il s'agit de* $\int_{]a,b]}$, $\int_{[a,b[}$, $\ldots$

On montre à l'exercice 2.5.2 que si $f \geq 0$, $\nu(A) = \int_A f d\mu$, $A \in \mathcal{F}$, définit une mesure ν sur $(\Omega, \mathcal{F})$. On dit alors que ν admet la densité f par rapport à μ, et on a la propriété

$$(*) \quad \text{pour tout } A \in \mathcal{F},\ \mu(A) = 0 \quad \Rightarrow \nu(A) = 0.$$

Si deux mesures ν et μ vérifient la propriété $(*)$, on dit que ν est **absolument continue** par rapport à μ, noté $\nu \ll \mu$. En fait l'absolue continuité $\nu \ll \mu$ entraîne l'existence d'une fonction mesurable f telle que $d\nu = f d\mu$, c'est le Théorème de Radon–Nikodym 3.3.1 ci–dessous.

Proposition 2.3.13. *Soit* $f \geq 0$, ν *la mesure sur* $(\Omega, \mathcal{F})$ *définie par* $\nu(A) = \int_A f d\mu$. *Si* g *est une application mesurable de* Ω *dans* $\bar{\mathbb{R}}$, *t.q.* $f g$ *est* μ*–quasi–intégrable, alors* g *est* ν *quasi–intégrable, et :*

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

PREUVE Il suffit de démontrer l'égalité ci–dessus lorsque $g \geq 0$. Soit $\mathcal{L}$ la classe des applications mesurables à valeurs dans $\bar{\mathbb{R}}_+$ qui vérifient l'égalité de l'énoncé. $\mathcal{L}$ contient 1_A, $A \in \mathcal{F}$ par définition de ν. Le résultat découle aisément du Théorème des classes monotones (Corollaire 1.7.9) $\square$

Soit $(\Omega', \mathcal{F}')$ un deuxième espace mesurable, $T : \Omega \to \Omega'$ une application $\mathcal{F}/\mathcal{F}'$ mesurable.

Théorème 2.3.14. *Soit f une application mesurable de Ω' dans $\bar{\mathbb{R}}$. f est μT^{-1} quasi–intégrable si et seulement si $f\circ T$ est μ quasi–intégrable, et alors :*

$$\int_{\Omega} f\circ T(\omega)\mu(d\omega) = \int_{\Omega'} f(\omega')\mu T^{-1}(d\,\omega').$$

Preuve Identique à celle de 2.3.13. □

Le Corollaire qui suit indique comment on effectue un changement de variable dans une intégrale par rapport à la mesure de Lebesgue.

Corollaire 2.3.15. *Soit T une bijection d'un ouvert $V\subset\mathbb{R}^d$ sur un ouvert $T(V)$ de $\mathbb{R}^d$. On suppose que T est continue et admet des dérivées partielles du 1er ordre continues et on note $J(x) = \det\left(\frac{\partial T_i}{\partial x_j}(x)\right)$. Alors $f\mathbf{1}_{T(V)}$ est λ–quasi–intégrable si et seulement si $f\circ T\mathbf{1}_V|J|$ est λ quasi intégrable, et*

$$\int_V f\circ T(x)|J(x)|dx = \int_{T(V)} f(y)dy.$$

Preuve On va se limiter au cas $d=1$, $V=]a,b[$. L'égalité sur l'énoncé peut alors se réécrire (dans le cas $T(a)<T(b)$ auquel on se restreint dans cette preuve, l'autre cas se traite de façon analogue) :

$$\int_a^b f(T(x))|T'(x)|dx = \int_{T(a)}^{T(b)} f(y)dy,$$

qu'il suffit d'établir pour $f\geq 0$. On définit sur $\mathcal{B}(]a,b[)$ la mesure : $\nu(A) = \int_A |T'(x)|dx$.

Si B est un sous–intervalle de $T(]a,b[)$, on a

$$\nu T^{-1}(B) = \int_{T^{-1}(B)} T'(x)dx = \int_B dy,$$

où l'on a utilisé le fait que, T' ayant un signe constant puisque T est injective, notre hypothèse $T(a)<T(b)$ veut dire que $T'(x)\geq 0$.

Donc $\nu\, T^{-1} = \lambda$, mesure de Lebesgue sur $(T(]a,b[),\mathcal{B}(T(]a,b[)))$. Le résultat est une conséquence du Théorème 2.3.14 avec $\mu=\nu$, $\Omega=]a,b[$, $\Omega'=T(]a,b[)$. □

2.4 Mesure produit et Théorème de Fubini

Dans cette section, on se donne deux espaces mesurés $(X, \mathcal{X}, \mu)$ et $(Y, \mathcal{Y}, \nu)$, où μ et ν sont des mesures σ–**finies** (cette hypothèse est cruciale !). On notera x le point générique de X, y le point générique de Y.

On définit l'espace mesurable produit de $(X \times Y, \mathcal{X} \otimes \mathcal{Y})$ comme au Chapitre 1, section 3.

Proposition 2.4.1. *(i) Si $E \in \mathcal{X} \otimes \mathcal{Y}$, alors pour tout $x \in X$, $\{y; (x,y) \in E\} \in \mathcal{Y}$ et pour tout $y \in Y$, $\{x; (x,y) \in E\} \in \mathcal{X}$.*

(ii) Si $f : X \times Y \to \bar{\mathbb{R}}$ est $\mathcal{X} \otimes \mathcal{Y}$ mesurable, alors pour tout $x \in X$, $y \to f(x,y)$ est $\mathcal{Y}$ mesurable et pour tout $y \in Y$, $x \to f(x,y)$ est $\mathcal{X}$ mesurable.

PREUVE

(i) Pour tout $x \in X$, on définit $T_x : Y \to X \times Y$ par $T_x(y) = (x,y)$

$$\forall\, A \in \mathcal{X}, B \in \mathcal{Y}, \quad T_x^{-1}(A \times B) = \begin{cases} B, & \text{si } x \in A; \\ \emptyset, & \text{si } x \notin A. \end{cases}$$

Donc $T_x^{-1}(A \times B) \in \mathcal{Y}$. Il résulte du Théorème 1.7.2 (i) que T_x est $\mathcal{Y}/\mathcal{X} \otimes \mathcal{Y}$ mesurable. Ceci démontre la première partie de (i). La deuxième partie se démontre de façon analogue.

(ii) On remarque que l'application $y \to f(x,y)$ coïncide avec $f \circ T_x$, et on applique le Théorème 1.7.2 (ii).

□

Proposition 2.4.2. *Soit $E \in \mathcal{X} \otimes \mathcal{Y}$.*

(i) $x \to \nu(\{y; (x,y) \in E\})$ est $\mathcal{X}$–mesurable.

(ii) $y \to \mu(\{x; (x,y) \in E\})$ est $\mathcal{Y}$–mesurable.

PREUVE Les deux propriétés se démontrent de la même façon. Montrons (i). Soit $\mathcal{C}$ la classe des $E \subset X \times Y$ qui sont tels que $x \to \nu(\{y; (x,y) \in E\})$ est $\mathcal{X}$ mesurable. $\mathcal{C}$ est un λ–système qui contient le π–système des rectangles $A \times B$, $A \in \mathcal{X}$, $B \in \mathcal{Y}$, puisque

$$\nu(\{y; (x,y) \in A \times B\}) = \mathbf{1}_A(x)\nu(B)$$

est $\mathcal{X}$ mesurable, à valeurs dans $\bar{\mathbb{R}}_+$. Donc, d'après le Théorème 1.3.8 ($\pi - \lambda$), $\mathcal{C}$ contient $\sigma(\mathcal{X} \times \mathcal{Y}) = \mathcal{X} \otimes \mathcal{Y}$. □

Théorème 2.4.3. *Il existe une unique mesure σ finie*

$$\pi = \mu \times \nu \text{ sur } (X \times Y, \mathcal{X} \otimes \mathcal{Y}) \text{ t. q. } :$$

$$\pi(A \times B) = \mu(A)\nu(B), \ \forall A \in \mathcal{X}, B \in \mathcal{Y}.$$

De plus, pour tout $E \in \mathcal{X} \otimes \mathcal{Y}$,

$$\pi(E) = \int_X \nu(\{y; (x,y) \in E\})\mu(dx) = \int_Y \mu(\{x; (x,y) \in E\}\nu(dy).$$

Preuve Les deux formules :

$$\pi'(E) = \int_X \nu(\{y; (x,y) \in E\})\mu(dx)$$
$$\pi''(E) = \int_Y \mu(\{x; (x,y) \in E\})\nu(dy)$$

définissent deux mesures sur $(X \times Y, \mathcal{X} \otimes \mathcal{Y})$ qui coïncident sur $\mathcal{X} \times \mathcal{Y} = \{A \times B; A \in \mathcal{X}, B \in \mathcal{Y}\}$. En effet $\pi'(A \times B) = \pi''(A \times B) = \mu(A)\nu(B)$. Or $\mathcal{X} \times \mathcal{Y}$ est un π–système, et $\sigma(\mathcal{X} \times \mathcal{Y}) = \mathcal{X} \otimes \mathcal{Y}$. Et il résulte aisément de ce que μ et ν sont σ–finies, que π' et π'' sont σ–finies le long de $\mathcal{X} \times \mathcal{Y}$. Donc π' et π'' coïncident. L'unicité se démontre de la même façon. $\square$

Théorème 2.4.4. *(Fubini) Soit $f : X \times Y \to \bar{\mathbb{R}}$ une application $\mathcal{X} \otimes \mathcal{Y}$ mesurable.*

a) Si $f \geq 0$ π p.p., alors :

(i) $\int_Y f(x,y)\nu(dy)$ est $\mathcal{X}$ mesurable, $\int_X f(x,y)\mu(dx)$ est $\mathcal{Y}$ mesurable.

(ii)

$$\int_{X \times Y} f(x,y)\pi(dx,dy) = \int_X \left(\int_Y f(x,y)\nu(dy)\right)\mu(dx)$$
$$= \int_Y \left(\int_X f(x,y)\mu(dx)\right)\nu(dy).$$

b) Si f est π–intégrable, alors :

(i)' Pour μ–presque tout x, $y \to f(x,y)$ est ν–intégrable Pour ν–presque tout y, $x \to f(x,y)$ est μ–intégrable

(ii)' $\int_Y f(x,y)\nu(dy)$ *est p.p. égale à une fonction* $\mathcal{X}$ *mesurable et* μ *intégrable.* $\int_X f(x,y)\mu(dx)$ *est p.p. égale à une fonction* $\mathcal{Y}$ *mesurable et* ν *intégrable.*

(iii)' Les égalités (ii) sont satisfaites.

Remarque 2.4.5. *(i)Le théorème de Fubini permet d'intervertir les ordres d'intégration par rapport à* μ *et* ν*, lorsque* f *est* ≥ 0*, ou* $\pi = \mu \times \nu$*–intégrable. En fait, le résultat est encore vrai si* f *est* π*–quasi–intégrable. Mais lorsque cette hypothèse n'est pas satisfaite,il se peut que les quantités* $\int_X \left(\int_Y f(x,y)\nu(dy) \right) \mu(dx)$ *et* $\int_Y \left(\int_X f(x,y)\mu(dx) \right) \nu(dy)$ *soient toutes deux définies, mais soient différentes. Voici un exemple d'une telle situation :* $X = Y = [0,1]$*,* $\mathcal{X} = \mathcal{Y} = \mathcal{B}([0,1])$*,* $\mu = \nu = \lambda$*, mesure de Lebesgue. Soit* $0 = \alpha_1 < \alpha_2 \cdots < \alpha_n, \alpha_n \to 1$*, quand* $n \to \infty$*. Pour tout* $n \geq 1$*, on suppose donnée une fonction* $g_n \in C([0,1])$*, telle que* $g_n(t) = 0$*, si* $t \notin]\alpha_n, \alpha_{n+1}[$*, et* $\int_0^1 g_n(t)dt = 1$*. On pose :*

$$f(x,y) = \sum_{n=1}^{\infty} [g_n(x) - g_{n+1}(x)] g_n(y)$$

Pour tout $(x,y) \in [0,1]^2$*, un terme au plus de la somme ci–dessus est non nul, donc la série converge.*

$$\text{Pour tout } x \in [0,1], \quad \int_0^1 f(x,y)dy = \sum_{n=1}^{\infty} [g_n(x) - g_{n+1}(x)] = g_1(x)$$

donc $\int_0^1 dx \int_0^1 f(x,y)dy = 1$*. Mais pour tout* $y \in [0,1]$*,*

$$\int_0^1 f(x,y)dx = 0, \quad \text{donc } \int_0^1 dy \int_0^1 f(x,y)dx = 0.$$

Notons que nous n'avons bien intégré que des fonctions intégrables.

(ii) En pratique si $f \geq 0$, on peut intervertir l'ordre des intégrations. Si f est de signe quelconque, il faut s'assurer que $\int_{X \times Y} |f| d\pi < \infty$, ce que l'on fait en calculant soit $\int_X d\mu \int |f| d\nu$, soit $\int_Y d\nu \int_X |f| d\mu$.

PREUVE

a) Si $f = \mathbf{1}_E$, $E \in \mathcal{X} \otimes \mathcal{Y}$, le résultat découle de la Proposition 2.4.2 et du Théorème 2.4.3. Le fait que (i) et (ii) sont satisfaites par toutes les applications $\mathcal{X} \otimes \mathcal{Y}$ mesurables à valeurs dans $\bar{\mathbb{R}}_+$ résulte du Théorème des classes monotones (Corollaire 1.7.9) puisque si $f, g \geq 0$ vérifient (i) et (ii), $\alpha, \beta > 0$, alors $\alpha f + \beta g$ vérifie (i) et (ii).

b) D'après a) et la π–intégrabilité de f,

$$\int_X \left(\int_Y |f(x,y)| \nu(dy) \right) \mu(dx) = \int_Y \left(\int_X |f(x,y)| \mu(dx) \right) \nu(dy) < \infty.$$

Il en résulte, d'après le Théorème 2.2.2 (iii), que

$$\int_Y |f(x,y)| \nu(dy) < \infty \; \mu \text{ p.p. et } \int_X |f(x,y)| \mu(dx) < \infty \; \nu \text{ p.p.}$$

On a montré (i)'. Montrons la première partie de (ii)'.
Si $A = \left\{ x; \int_Y |f(x,y)| \nu(dy) = +\infty \right\}$, $\mu(A) = 0$, et on pose :

$$g(x) = \begin{cases} 0 & \text{si } x \in A \\ \int_X f(x,y) \nu(dy), & \text{si } x \in A^c \end{cases}$$

Il résulte aisément de a) (i) que g est $\mathcal{X}$ mesurable. g est μ–intégrable, puisque $|g(x)| \leq \int_X |f(x,y)| \nu(dy)$. Bien que $\int f(x,y) \nu(dy)$ ne soit éventuellement pas défini pour $x \in A$, puisque $\mu(A) = 0$ et que pour tout $x \notin A$, $\int_Y f(x,y) \nu(dy) = g(x)$, on pose par convention

$$\int_X \left(\int_Y f(x,y) \nu(dy) \right) \mu(dx) = \int_X g(x) \mu(dx).$$

(iii)' résulte de (ii) en utilisant la décomposition $f = f^+ - f^-$. □

2.5 Exercices

Exercice 2.5.1. *Soit* $\{A_i,\ 1 \le i \le n\} \subset \mathcal{F}$, *et* $x_i > 0$, $1 \le i \le n$. *On pose*

$$f(\omega) = \sum_{i=1}^{n} x_i \mathbf{1}_{A_i}(\omega).$$

Montrer que $\int f d\mu = \sum_{i=1}^{n} x_i \mu(A_i)$. *Dans le cas où les* x_i *sont de signe quelconque, quelle condition faut–il imposer pour que* f *soit intégrable ? A–t–on alors encore la même formule que dans le cas des* $x_i \ge 0$ *?*

Exercice 2.5.2. *Si* f *est une application mesurable de* Ω *à valeurs dans* $\mathbb{R}_+$, $\nu(A) = \displaystyle\int_A f d\mu$, $A \in \mathcal{F}$, *définit une mesure* ν *sur* $(\Omega, \mathcal{F})$.

Exercice 2.5.3. *Déduire (i) et (iii) de la Proposition 2.3.10 dans le cas* μ σ*–finie comme corollaire du Théorème de Fubini. [On choisit* $(X, \mathcal{X}, \mu) = (\Omega, \mathcal{F}, \mu), (Y, \mathcal{Y}, \nu) = (\mathbb{N}, \mathcal{P}(\mathbb{N}), \displaystyle\sum_{1}^{\infty} \delta_n)$*].*

Exercice 2.5.4. *Montrer que si* f *est* μ*–intégrable, alors quand* $n \to \infty$,

$$\int_{\{f>n\}} f d\mu \to 0, \qquad \int_{\{|f| \le 1/n\}} f d\mu \to 0.$$

Exercice 2.5.5. *Dans cet exercice, on se place sur un espace mesuré muni d'une mesure* μ **finie**. *On se donne une suite* $\{f_n,\ n \ge 1\}$ *d'applications mesurables :* $\Omega \to \mathbb{R}$ *telle que* $f \to 0$ *p. p. On suppose en outre satisfaite l'hypothèse*

$$(*) \quad C_2 := \sup_n \int f_n^2 d\mu < \infty.$$

1. *Montrer que* f_n *est* μ*–integrable, pour tout* $n \ge 1$.
2. *Montrer que pour tout* $n \ge 1$, $M > 0$,

$$\int |f_n| d\mu \le \int (|f_n| \wedge M)\, d\mu + \frac{1}{M} \int f_n^2 d\mu.$$

3. *En déduire que*

$$\limsup_{n\to\infty} \int |f_n| d\mu \le \frac{C_2}{M}.$$

4. *Conclure que* $\int |f_n|d\mu \to 0$ *et* $\int f_n d\mu \to 0$ *quand* $n \to \infty$.
5. *Montrer que les mêmes résultats sont encore vrais si l'on remplace dans l'hypothèse* $(*)$ f_n^2 *par* $|f_n|^p$, *avec un certain* $p > 1$.
6. *Donner un exemple (avec* $\Omega = [0,1], \mathcal{F} = \mathcal{B}([0,1], \mu =$ *mesure de Lebesgue) d'une suite de fonctions à valeurs positives qui vérifie* $(*)$, *mais telle que la fonction* $g(x) := \sup_{n\geq 1} f_n(x)$ *n'est pas intégrable (donc le résultat ne découle pas directement du Théorème de convergence dominée). Indication : on pourra condidérer la suite* f_n *définie par*

$$f_n(x) = \begin{cases} 0, & si\ 0 \leq x < \frac{1}{\sqrt{n}} - \frac{1}{n}\ ; \\ \sqrt{n}, & si\ \frac{1}{\sqrt{n}} - \frac{1}{n} \leq x \leq \frac{1}{\sqrt{n}}\ ; \\ 0, & si\ \frac{1}{\sqrt{n}} < x \leq 1. \end{cases}$$

Exercice 2.5.6. *On se place dans le cadre de l'exercice 2.5.5, mais au lieu de supposer que* $f_n \to 0$, *on suppose maintenant que* $f_n \to f$ *p. p. On suppose toujours* $(*)$ *satisfaite.*

1. *Montrer que* $\int f^2 d\mu \leq C_2$.
2. *En déduire que l'on peut appliquer le résultat de l'exercice 2.5.5 à la suite* $g_n := f_n - f$.
3. *Montrer que* $\int f_n d\mu \to \int f d\mu$, *quand* $n \to \infty$.

Exercice 2.5.7. *On se place à nouveau dans le cadre de l'exercice 2.5.5, mais on ne suppose plus que la mesure* μ *est finie (mais on la suppose* σ*–finie). On remplace alors l'hypothèse* $(*)$ *par l'hypothèse qu'il existe une fonction intégrable* g *à valeurs dans* $(0, +\infty)$ *telle que*

$$\sup_n \int \frac{f_n^2}{g} d\mu < \infty.$$

Montrer que les conclusions de l'exercice 2.5.5 sont encore valides.

Exercice 2.5.8. *Dans cet exercice, toutes les fonctions sont supposées définies sur un espace mesuré* $(\Omega, \mathcal{F}, \mu)$, *à valeurs dans* $\mathbb{R}$. *On rappelle que* $a^+ = \sup(a, 0)$ *et pourra utiliser le fait que si* $a, b \geq 0$, *alors* $(a - b)^+ \leq a$.

1. *Soit* f_n, $n \geq 1$ *des fonctions mesurables, et* f *une fonction mesurable et* μ*-intégrable, telles que*

   (i) $f_n \geq 0$ *p. p., quelque soit* $n \geq 1$;

(ii) $f_n \to f$ *p. p. quand* $n \to \infty$ *;*

(iii) $\int_\Omega f_n d\mu \to \int_\Omega f d\mu$, *quand* $n \to \infty$.

Montrer que $\int_\Omega |f - f_n| d\mu \to 0$, *quand* $n \to \infty$.
Indication : on pourra remarquer que $|f - f_n| = 2(f - f_n)^+ - (f - f_n)$.

2. *Soit* f_n, g_n, $n \geq 1$ *des fonctions mesurables telles que*

   (i) $0 \leq f_n \leq g_n$ *p. p., quelque soit* $n \geq 1$ *;*

   (ii) $f_n \to f$, $g_n \to g$ *p. p. quand* $n \to \infty$ *;*

   (iii) g *et toutes les* g_n *sont intégrables, et* $\int_\Omega |g - g_n| d\mu \to 0$, *quand* $n \to \infty$.

   En déduire que f_n *et* f *sont intégrables, et que* $\int_\Omega f_n d\mu \to \int_\Omega f d\mu$, *quand* $n \to \infty$.
   Indication : on pourra utiliser l'identité

$$\int_\Omega f_n d\mu = \int_\Omega \inf(f_n, g) d\mu + \int_\Omega (f_n - g)^+ d\mu,$$

$$\text{et } \int_\Omega (f_n - g)^+ d\mu \leq \int_\Omega (g_n - g)^+ d\mu \leq \int_\Omega |g_n - g| d\mu.$$

3. *Déduire des questions 1 et 2 que si* ρ_n, φ_n, ψ_n, $n \geq 1$, *sont des applications mesurables, ainsi que* ρ, φ *et* ψ, *telles que*

   (i) $\rho_n \leq \varphi_n \leq \psi_n$, *p. p., quelque soit* $n \geq 1$ *;*

   (ii) $\rho_n \to \rho$, $\varphi_n \to \varphi$, $\psi_n \to \psi$ *p. p. quand* $n \to \infty$ *;*

   (iii) ρ_n, ρ, ψ_n, ψ *sont intégrables, pour tout* $n \geq 1$ *;*

   (iv) $\int_\Omega \rho_n d\mu \to \int_\Omega \rho d\mu$, $\int_\Omega \psi_n d\mu \to \int_\Omega \psi d\mu$ *quand* $n \to \infty$ *;*

   alors pour tout $n \geq 1$, φ_n *et* φ *sont intégrables, et* $\int_\Omega \varphi_n d\mu \to \int_\Omega \varphi d\mu$ *quand* $n \to \infty$.

Chapitre 3

Espaces L^p

On suppose à nouveau donné un espace mesuré $(\Omega, \mathcal{F}, \mu)$, avec μ mesure σ–finie.

3.1 Définition des espaces L^p.

Deux exposants $p, q \in [1, +\infty]$ seront dits **conjugués** si $\frac{1}{p} + \frac{1}{q} = 1$ (d'où soit $p, q \in]1, +\infty[$, soit l'un vaut 1 et l'autre $+\infty$).

Proposition 3.1.1. *Soit $p, q \in]1, +\infty[$ deux exposants conjugués, f et g deux applications mesurables à valeurs dans $\bar{\mathbb{R}}_+$, dont aucune des deux n'est p.p. nulle.*

(i) (Inégalité de Hölder)

$$\int_\Omega fg \; d\mu \leq \left(\int_\Omega f^p d\mu \right)^{1/p} \left(\int_\Omega g^q d\mu \right)^{1/q}.$$

(ii) (Inégalité de Minkowski)

$$\left(\int_\Omega (f+g)^p d\mu \right)^{1/p} \leq \left(\int_\Omega f^p d\mu \right)^{1/p} + \left(\int_\Omega g^p d\mu \right)^{1/p}.$$

Preuve

(i) L'inégalité est satisfaite lorsque l'un des deux facteurs du second membre est égal à $+\infty$, puisque l'autre est non nul. Donc si l'on

pose $A = \left(\int f^p d\mu\right)^{1/p}$, $B = \left(\int g^q d\mu\right)^{1/q}$, il suffit de faire la démonstration dans le cas $A, B \in]0, +\infty[$.
Posons $F = f/A$, $G = g/B$. Soit $\omega \in \Omega$ tel que $F(\omega), G(\omega) \in]0, +\infty[$. Alors il existe $s, t \in \mathbb{R}$ t.q. $F(\omega) = e^{s/p}$ et $G(\omega) = e^{t/q}$, et la convexité de l'application exponentielle entraîne :

$$e^{\frac{s}{p}+\frac{t}{q}} \leq \frac{1}{p}e^s + \frac{1}{q}e^t,$$

d'où $F(\omega)G(\omega) \leq \frac{1}{p}F(\omega)^p + \frac{1}{q}G(\omega)^q$ inégalité qui est en fait vraie pour tout $\omega \in \Omega$. Donc en intégrant.

$$\int FG d\mu \leq \frac{1}{p} + \frac{1}{q} = 1, \text{ d'où le résultat.}$$

(ii) Il suffit de faire la preuve dans le cas où le second membre de l'inégalité est fini.

$$(f+g)^p = f(f+g)^{p-1} + g(f+g)^{p-1}, \text{ et d'après (i),}$$

$$\int f(f+g)^{p-1} d\mu \leq \left(\int f^p d\mu\right)^{1/p} \left(\int (f+g)^{(p-1)q} d\mu\right)^{1/q}$$
$$\int g(f+g)^{p-1} d\mu \leq \left(\int g^p d\mu\right)^{1/p} \left(\int (f+g)^{(p-1)q} d\mu\right)^{1/q}$$

Par addition

$$\int (f+g)^p d\mu \leq \left[\left(\int f^p d\mu\right)^{1/p} + \left(\int g^p d\mu\right)^{1/p}\right] \left(\int (f+g)^p d\mu\right)^{1-\frac{1}{p}}$$

D'après la convexité de $t \to t^p$ de $\mathbb{R}_+$ dans $\mathbb{R}_+$, $p > 1$, $2^{-p}(f+g)^p \leq 2^{-1}(f^p + g^p)$. Il résulte alors des restrictions faites ci–dessus que $\int (f+g)^p d\mu \in]0, +\infty[$, et on peut multiplier la dernière inégalité par $\left(\int (f+g)^p d\mu\right)^{\frac{1}{p}-1}$, d'où le résultat.

□

Remarque 3.1.2. *Dans le cas $p = q = 2$, l'inégalité (i) est aussi appelée inégalité de Cauchy–Schwarz. On peut la redémontrer de la façon suivante :*

$$\int (f + \lambda g)^2 d\mu = \int f^2 d\mu + 2\lambda \int fg d\mu + \lambda^2 \int g^2 d\mu \geq 0,$$

pour tout $\lambda \in \mathbb{R}$, donc le discriminant du trinome est ≤ 0. Le discriminant est nul (i.e. on a l'égalité dans l'inégalité de Cauchy–Schwarz) si et seulement si il existe $\lambda \in \mathbb{R}$ tel que $f + \lambda g = 0$ μ p.p.

□

Définition 3.1.3. *Soit f une application mesurable à valeurs dans $\bar{\mathbb{R}}$.*

Si $p \in [1, +\infty[$, on pose $\|f\|_p = \left(\int_\Omega |f|^p d\mu\right)^{1/p}$

On pose

$$\|f\|_\infty = \begin{cases} +\infty, & \text{si } \mu(|f| > \alpha) > 0, \text{ pour tout } \alpha \geq 0\,; \\ \inf\{\alpha \geq 0;\ \mu(|f| > \alpha) = 0\}, & \text{sinon.} \end{cases}$$

Pour $p \in [1, +\infty]$, on désigne par $\mathcal{L}^p(\Omega, \mathcal{F}, \mu)$ - ou plus simplement $\mathcal{L}^p(\mu)$- l'espace vectoriel des f t.q. $\|f\|_p < \infty$.

Le fait que $\mathcal{L}^p(\mu)$ est un espace vectoriel est immédiat pour $p = 1$ ou $+\infty$, et résulte de l'inégalité de Minkowski, pour $p \in]1, +\infty[$.

$\|\cdot\|_p$ est une semi–norme sur $\mathcal{L}^p(\mu)$:

$$\begin{aligned} \|f\|_p &\geq 0 \quad (\|f\|_p = 0 \Leftrightarrow f = 0 \text{ p.p.}) \\ \|\lambda f\|_p &= |\lambda| \cdot \|f\|_p \\ \|f + g\|_p &\leq \|f\|_p + \|g\|_p. \end{aligned}$$

Considérons sur $\mathcal{L}^p(\mu)$ la relation d'équivalence $f \sim f' \Leftrightarrow f = f'$ μ p.p.. Cette relation d'équivalence est compatible avec la structure vectorielle d'espace semi–normé de $\mathcal{L}^p(\mu)$, puisque :

$$\begin{gathered} f \sim f' \Rightarrow \|f\|_p = \|f'\|_p \\ f \sim f',\ g \sim g' \Rightarrow f + g \sim f' + g' \\ f \sim f' \Rightarrow \lambda f \sim \lambda f' \end{gathered}$$

L'ensemble des classes d'équivalence d'éléments de $\mathcal{L}^p(\mu)$ forme donc un espace vectoriel, noté $L^p(\Omega, \mathcal{F}, \mu)$ - ou $L^p(\mu)$, et $\|\cdot\|_p$ est une norme sur $L^p(\Omega, \mathcal{F}, \mu)$.

Remarque 3.1.4. *En sus des propriétés ci–dessus, on a que si $f \sim f'$, f' est (quasi)–intégrable si et seulement si f l'est, et alors* $\int f d\mu = \int f' d\mu$. *De plus, si $f_n \sim f'_n$, $n \geq 1$, alors $\sup_{n\geq 1} f_n \sim \sup_{n\geq 1} f'_n$, $\inf_{n\geq 1} f_n \sim \inf_{n\geq 1} f'_n$, il en est de même avec les* $\limsup$ *et les* $\liminf$*, et les limites lorsqu'elles existent. Ces propriétés ne seraient plus vraies pour des familles non dénombrables d'applications mesurables.*

Etant donnée une application mesurable f, désignons par $\tilde{f}$ sa classe d'équivalence, i.e. $\tilde{f} = \{f' \text{ mesurable}\,; f' \sim f\}$. Les propriétés ci–dessus permettent d'opérer sur les classes d'équivalence d'applications mesurables comme sur les applications mesurables elles–mêmes, à condition de ne considérer qu'une quantité dénombrable d'applications mesurables à la fois.

Dans la suite on confondra par abus de langage une classe d'équivalence avec l'un quelconque de ses représentants. Par exemple, on écrira $f \in L^p(\mu)$. Cet abus de langage est sans danger tant que l'on ne considère qu'une quantité dénombrable d'applications mesurables à la fois.

3.2 Propriétés des espaces $L^p(\mu)$

Théorème 3.2.1. *Pour tout $p \in [1, +\infty]$, l'espace $L^p(\mu)$, muni de la norme $\|\cdot\|_p$, est un espace de Banach.*

Muni du produit scalaire

$$\langle f, g\rangle = \int fg d\mu$$

l'espace $L^2(\mu)$ est un espace de Hilbert.

Preuve Il suffit de montrer que $L^p(\mu)$ est complet, i.e. que toute suite de Cauchy converge.

a) **Cas** $p < \infty$

Soit $\{f_n, n \geq 1\} \subset L^p(\mu)$ t.q. pour tout $\varepsilon > 0$, il existe N_ε tel que $n, m \geq N_\varepsilon \Rightarrow \|f_n - f_m\|_p \leq \varepsilon$.

Définissons une sous–suite f_{n_k} en choisissant : $n_0 = 0$; pour tout $k \geq 1$, n_k vérifie : $n_k > n_{k-1}$ pour tous $n, m \geq n_k$, $\|f_n - f_m\|_p \leq 2^{-k}$. Alors pour tout $k \geq 1$, $\|f_{n_{k+1}} - f_{n_k}\|_p \leq 2^{-k}$. On pose :

$$g_k = \sum_1^k |f_{n_{i+1}} - f_{n_i}|,\ g = \sum_1^\infty |f_{n_{i+1}} - f_{n_i}|.$$

Il résulte de l'inégalité de Minkowski que $\|g_k\|_p \leq 1$, donc par convergence monotone $\|g\|_p \leq 1$, et en particulier si $A = \{\omega; g(\omega) < \infty\}$, $\mu(A^c) = 0$. Mais pour tout $\omega \in A$, la série

$$f_{n_1}(\omega) + \sum_1^\infty (f_{n_{i+1}}(\omega) - f_{n_i}(\omega))$$

converge absolument. On définit :

$$f(\omega) = \begin{cases} f_{n_1}(\omega) + \sum_1^\infty (f_{n_{i+1}}(\omega) - f_{n_i}(\omega)), & \text{si } \omega \in A; \\ 0, & \text{sinon.} \end{cases}$$

Il est clair que $f_{n_k} \to f$ p.p.. Il reste à montrer que $f \in L^p(\mu)$, et $f_n \to f$ dans $L^p(\mu)$, i.e. $\|f_n - f\|_p \to 0$.
Soit $\varepsilon > 0$, et N_ε t.q. $||f_n - f_m||_p \leq \varepsilon$, pour tout $n, m \geq N_\varepsilon$. Alors si $m \geq N_\varepsilon$, par le lemme de Fatou,

$$||f - f_m||_p \leq \liminf_k ||f_{n_k} - f_m||_p \leq \varepsilon.$$

b) **Cas** $p = +\infty$ Posons $A_k = \{\omega; |f_k(\omega)| > ||f_k||_\infty\}$, $B_{n,m} = \{\omega; |f_n(\omega) - f_m(\omega)| > \|f_n - f_m\|_\infty\}$. et $E = \left(\bigcup_{k=1}^\infty A_k\right) \bigcup \left(\bigcup_{n,m=1}^\infty B_{n,m}\right)$. Alors $\mu(E) = 0$. Sur E^c, f_n converge uniformément vers une limite bornée f. On définit f sur Ω en posant $f(\omega) = 0$ si $\omega \in E$.

$$\text{Alors } f \in L^\infty(\mu), \text{ et } \|f_n - f\|_\infty \to 0$$

□

On notera $f_n \xrightarrow{L^p(\mu)} f$ pour $f_n \to f$ dans $L^p(\mu)$, i.e. $f_n, f \in L^p(\mu)$ et $\|f - f_n\|_p \to 0$. On vient de démontrer au passage le :

Théorème 3.2.2. *Si $p \in [1, +\infty]$ et si $f_n \xrightarrow{L^p(\mu)} f$, alors il existe une sous–suite $\{f_{n_k}\}$ t. q.*

$$f_{n_k} \longrightarrow f \text{ p.p., quand } k \to \infty.$$

Théorème 3.2.3. *Soit S la classe des fonctions de la forme*

$$f = \sum_1^n \alpha_i \mathbf{1}_{A_i},$$

avec $n \in \mathbb{N}$ *;* $\alpha_i \in \mathbb{R}$*,* $A_i \in \mathcal{F}$*,* $i = 1, \dots, n$ *et* $\mu\left(\bigcup_1^n A_i\right) < \infty$*. Alors pour tout* $1 \leq p < \infty$*,* $S \subset L^p(\mu)$ *et* S *est dense dans* $L^p(\mu)$*.*

Remarque 3.2.4. *S n'est en général pas dense dans $L^\infty(\mu)$ (ex : $\Omega = \mathbb{R}_+, \mathcal{F} = \mathcal{B}(\mathbb{R}_+), \mu = \lambda$, $f = \sum_0^\infty 2^{-n}\mathbf{1}_{[2n,2n+1[}$ ne peut pas être approchée en norme $\|\cdot\|_\infty$ par une suite de S.*

Preuve $S \subset L^p(\mu)$ résulte de $\mu\left(\bigcup_1^\infty A_n\right) < \infty$. Soit $f \in L^p(\mu)$, $f \geq 0$. Alors par la Remarque 1.7.7, il existe une suite $\{f_n\}$ de fonctions étagées t.q. $0 \leq f_n \uparrow f$. Alors $f_n \in S$, et $\|f - f_n\|_p \to 0$, d'après le théorème de convergence monotone. Pour f de signe quelconque, on décompose $f = f^+ - f^-$. $\square$

Théorème 3.2.5. *Soit μ une mesure sur $(\mathbb{R}^d, \mathcal{B}_d)$ t.q. $\mu(A) < \infty$, pour tout $A \in \mathcal{B}_d$, A borné. Alors $C_K(\mathbb{R}^d)$, espace des fonctions continues à support compact de $\mathbb{R}^d$ à valeurs dans $\mathbb{R}$, est dense dans $L^p(\mu)$, pour tout $p \in [1, +\infty[$.*

On va d'abord établir le :

Lemme 3.2.6. *Soit K un compact et O un ouvert de $\mathbb{R}^d$, avec $K \subset O$. Alors il existe $g \in C_K(\mathbb{R}^d)$ qui vérifie $0 \leq g(x) \leq 1$ pour tout x, $g(x) = 1$ pour tout $x \in K$ et $g(x) = 0$ pour tout $x \in O^c$.*

Preuve Il suffit de traiter le cas O borné. Alors $g(x) = \dfrac{d(x, O^c)}{d(x, K) + d(x, O^c)}$ répond à la question. $\square$

Preuve du Théorème 3.2.5 : Il est clair que $C_K(\mathbb{R}^d) \subset L^p(\mu)$. Pour montrer la densité, au vu du Théorème 3.2.3, il suffit de montrer que pour tout $A \in \mathcal{F}$, avec $\mu(A) < \infty$, pour tout $\varepsilon > 0$, il existe $g \in C_K(\mathbb{R}^d)$ t.q.

$$\|\mathbf{1}_A - g\|_p \leq \varepsilon$$

D'après le théorème 1.6.1 (i), il existe O ouvert tel que $A \subset O$ et $\mu(O \setminus A) \leq \varepsilon^p/2$, et d'après 1.6.1 (ii), il existe K compact t.q. $K \subset A$ et $\mu(A \setminus K) \leq \varepsilon^p/2$.

$$\text{Donc : } K \subset A \subset O, \ \ \mu(O \setminus K) \leq \varepsilon^p,$$

et si g désigne la fonction construite au lemme 3.2.6,

$$\|\mathbf{1}_A - g\|_p \leq \varepsilon$$

$\square$

Théorème 3.2.7. *Soit μ une mesure sur $(\mathbb{R}^d, \mathcal{B}_d)$ t.q. $\mu(A) < \infty$, $\forall A \in \mathcal{B}_d$, A borné. Alors, si $p \in [1, +\infty[$, $L^p(\mu)$ est séparable.*

PREUVE Soit $\mathcal{F}_0$ la classe des réunions finies de pavés de la forme $\prod_i^d [a_i, b_i[$ de $\mathbb{R}^d$, dont les sommets ont leurs coordonnées dans $\mathbb{Q} \cup \{+\infty\} \cup \{-\infty\}$. ($\mathbb{Q}$= ensemble des rationnels). Alors $\mathcal{F}_0$ est dénombrable, car le produit de deux ensembles dénombrables est dénombrable et l'ensemble des parties finies d'un ensemble dénombrable est dénombrable.

$\mathcal{F}_0$ est une algèbre, et $\sigma(\mathcal{F}_0) = \mathcal{B}_d$. Soit

$$\mathcal{L} = \left\{ \sum_1^n \alpha_i \mathbf{1}_{A_i}; n \geq 1,\ \alpha_i \in \mathbb{Q},\ \ A_i \in \mathcal{F}_0,\ \ \mu(A_i) < \infty \right\}.$$

A nouveau, $\mathcal{L}$ est dénombrable. Pour montrer que $\mathcal{L}$ est dense dans $L^p(\mu)$, il suffit de montrer que pour tout $f \in S, \eta > 0$, il existe $g \in \mathcal{L}$ t.q. $\|f - g\|_p < \eta$. En fait, il suffit pour cela de montrer que pour tout $\alpha \in \mathbb{R}$, $A \in \mathcal{F}$ avec $\mu(A) < \infty$, pour tout $\varepsilon > 0$,

$$\text{il existe } \beta \in \mathbb{Q},\ B \in \mathcal{F}_0 \text{ t. q. } \|\alpha \mathbf{1}_A - \beta \mathbf{1}_B\|_p \leq \varepsilon.$$

Mais $\|\alpha \mathbf{1}_A - \beta \mathbf{1}_B\|_p \leq |\alpha - \beta| \cdot (\mu(A))^{1/p} + |\beta| \cdot (\mu(A \triangle B))^{1/p}$. Il reste donc à montrer que quelque soit $\theta > 0$,

$$\text{il existe } B \in \mathcal{F}_0 \text{ t.q. } \mu(A \triangle B) \leq \theta,$$

ce qui résulte de la Proposition 1.4.12 $\square$

3.3 Théorème de Radon–Nikodym et dualité

Soit λ et μ deux mesures σ–finies sur $(\Omega, \mathcal{F})$. On dit que λ est **absolument continue** par rapport à μ, noté $\lambda \ll \mu$, si pour tout $A \in \mathcal{F}$, $\mu(A) = 0$ implique $\lambda(A) = 0$. Les deux mesures sont dites **équivalentes** si l'on a à la fois $\lambda \ll \mu$ et $\mu \ll \lambda$. Elles sont dites **étrangères** s'il existe $E \in \mathcal{F}$ t. q. $\lambda(E) = \mu(E^c) = 0$.

Théorème 3.3.1. *(Radon–Nikodym) Soit λ et μ deux mesures σ–finies sur $(\Omega, \mathcal{F})$, t.q. $\lambda \ll \mu$. Alors il existe h, application mesurable à valeurs dans $\mathbb{R}_+$, t.q. :*

$$\lambda = h \cdot \mu$$
$$\textit{i.e. pour tout } A \in \mathcal{F}, \lambda(A) = \int_A h d\mu.$$

PREUVE

a) Supposons tout d'abord que $\lambda(\Omega) < \infty$. Posons $\nu = \lambda + \mu$. Alors $\lambda(A) \leq \nu(A)$, $A \in \mathcal{F}$. Donc si $f \in L^2(\nu)$, $f \in L^2(\lambda) \subset L^1(\lambda)$ (car $\lambda(\Omega) < \infty$), et :

$$\left| \int f \, d\lambda \right| \leq \sqrt{\lambda(\Omega)} \sqrt{\int f^2 d\lambda} \leq \sqrt{\lambda(\Omega)} \sqrt{\int f^2 d\nu}.$$

Donc l'application $f \to \int f d\lambda$ est une fonctionnelle linéaire continue sur l'espace de Hilbert $L^2(\nu)$, donc par le Théorème 3.3.2 ci–dessous, il existe $g \in L^2(\nu)$ t. q. :

$$\int f \, d\lambda = \int fg d\nu, \text{ pour tout } f \in L^2(\nu).$$

Si $\{A_n, n \geq 1\}$ vérifie $\nu(A_n) < \infty$ pour tout n, et $\bigcup_1^\infty A_n = \Omega$, alors en choisissant $f = \mathbf{1}_{A_n \cap \{g<0\}}$, $n = 0, 1, \ldots$ dans l'égalité ci–dessus, on montre que $g \geq 0$ ν p.p. Alors grâce au théorème de convergence monotone, l'égalité est encore vraie avec $f = \mathbf{1}_{\{g>1\}}$, d'où l'on tire :

$$\nu(g > 1) \geq \lambda(g > 1) = \int_{\{g>1\}} g d\nu$$

$$\text{Donc } \int_{\{g-1>0\}} (g - 1) \, d\,\nu \leq 0 \Rightarrow g \leq 1 \; \nu \text{ p.p.},$$

donc il existe g mesurable à valeurs dans $[0, 1]$ t.q.

$$\lambda = g \cdot \nu, \text{ soit}$$
$$(1 - g) \cdot \lambda = g \cdot \mu$$

Donc

$$\mu(g=1) = \int_{\{g=1\}} (1-g)d\lambda = 0$$

Or

(∗) $$\lambda = \mathbf{1}_{\{g<1\}} \frac{g}{1-g} \cdot \mu + \mathbf{1}_{\{g=1\}} \cdot \lambda.$$

Mais puisque $\lambda \ll \mu$, $\mathbf{1}_{\{g=1\}} \cdot \lambda = 0$, d'où le résultat cherché avec $h = 1_{\{g<1\}} \dfrac{g}{1-g}$

b) Le cas général : λ σ–finie ;

Soit $\{C_n, n \geq 1\} \subset \mathcal{F}$ t.q. $C_0 = \emptyset$, $C_n \subset C_{n+1}, \bigcup_1^\infty C_n = \Omega$, et $\lambda(C_n) < \infty$ pour tout n. Définissons, pour $n \geq 1$:

$$\lambda_n(A) = \lambda(A \cap (C_n - C_{n-1})), \;\; A \in \mathcal{F}.$$

Alors les λ_n sont des mesures finies, et :

$$\lambda(A) = \sum_1^\infty \lambda_n(A), \; \forall A \in \mathcal{F}.$$

D'après la partie a) de la démonstration, pour tout n il existe h_n t.q. $\lambda_n = h_n \cdot \mu$, avec $h_n \geq 0$ partout, et on peut le choisir t.q. $h_n = 0$ sur $(C_n - C_{n-1})^c$.

Alors pour tout $\omega \in \Omega$, la série $h(\omega) = \sum_1^\infty h_n(\omega)$ converge, et

$$\lambda = h \cdot \mu.$$

□

Il nous reste à montrer le

Théorème 3.3.2. *Soit L est une forme linéaire continue sur un espace de Hilbert H. Alors il existe un unique $y \in H$ tel que*

$$Lx = \langle x, y \rangle, \quad \textit{pour tout } x \in H.$$

Preuve L'unicité est facile à montrer. Si $Lx = 0$ pour tout $x \in H$, alors $y = 0$ répond à la question. On peut donc supposer que $M := \{x,\ Lx = 0\} \neq H$. Donc il existe $z \in M^{\perp}$, $z \neq 0$. Posons

$$y = \frac{Lz}{\langle z, z\rangle} z.$$

Alors il est facile de vérifier que $\langle y, y\rangle = Ly$, et $y \neq 0$. A tout $x \in H$, on associe

$$x' = \frac{Lx}{\langle y, y\rangle} y.$$

Il est clair que $L(x - x') = 0$, donc $x - x' \in M$, et $x - x' \perp y$, d'où

$$\langle x, y\rangle = Lx,$$

ceci pour tout $x \in H$, ce qui établit le résultat. □

Théorème 3.3.3. *(Décomposition de Lebesgue)*
Soit λ et μ deux mesures σ–finies sur $(\Omega, \mathcal{F})$. λ s'écrit de façon unique comme la somme d'une mesure absolument coninue par rapport à μ, et d'une mesure étrangère par rapport à μ ;

Preuve L'unicité est facile. L'existence de la décomposition a été écrite dans la preuve du Théorème 3.3.1 cf. égalité $(*)$, dans le cas $\lambda(\Omega) < \infty$. Le cas général est laissé en exercice. □

Définition 3.3.4. *On appellera mesure signée sur $(\Omega, \mathcal{F})$ une application μ de $\mathcal{F}$ dans $\mathbb{R}$ t.q. :*

(i) $\mu(\emptyset) = 0$;

(ii) $\begin{cases} \text{pour tout } \{A_n, n \geq 1\} \subset \mathcal{F} \text{ avec } A_k \cap A_\ell = \emptyset \text{ si } k \neq \ell, \\ \mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n). \end{cases}$

Nous admettrons la :

Proposition 3.3.5. *(Jordan–Hahn) Si μ est une mesure signée définie sur $(\Omega, \mathcal{F})$, alors les formules :*

$$\mu^+(A) = \sup_{\substack{B \in \mathcal{F} \\ B \subset A}} \mu(B) \ \text{ et } \ \mu^-(A) = \sup_{\substack{B \in \mathcal{F} \\ B \subset A}} (-\mu(B))$$

définissent deux mesures finies sur $(\Omega, \mathcal{F})$ t.q.

(i) $\mu = \mu^+ - \mu^-$

(ii) μ^+ *et* μ^- *sont étrangères.*

La mesure $|\mu| = \mu^+ + \mu^-$*, appelée variation totale de* μ*, vérifie :*

$$\forall\, A \in \mathcal{F},\ \ |\mu|(A) = \sup \sum_1^\infty |\mu(A_n)|,$$

où le sup *est pris sur toutes les partitions* $\{A_n, n \geq 1\}$ *de* A*.*

Corollaire 3.3.6. *Soit* λ *une mesure signée,* μ *une mesure* σ*–finie, définies sur* $(\Omega, \mathcal{F})$*. Si pour tout* $N \in \mathcal{F}$*,* $\mu(N) = 0$ *implique* $\lambda(N) = 0$*, alors* λ^+ *et* λ^- *sont absolument continues par rapport à* μ*, et il existe* $h \in L^1(\mu)$ *t.q :*

$$\lambda(A) = \int_A h d\mu,\ \ \forall A \in \mathcal{F}.$$

Preuve D'après la Proposition 3.3.5, il existe $E \in \mathcal{F}$ t.q. :

$$\begin{aligned} \lambda^+(A) &= \lambda(A \cap E) \\ \lambda^-(A) &= -\lambda(A \cap E^c) \end{aligned}$$

Soit $N \in \mathcal{F}$ t.q. $\mu(N) = 0$. Alors $\mu(N \cap E) = 0$, donc $\lambda(N \cap E) = \lambda^+(N) = 0$. De même, $\lambda^-(N) = 0$. Il reste à appliquer le Théorème de Radon–Nikodym à λ^+ et à λ^- et à utiliser le fait que ce sont des mesures finies. $\square$

Soit $(\Omega, \mathcal{F}, \mu)$ un espace mesuré, $p \in [1, +\infty]$ et q l'exposant conjugué de p. Soit $g \in L^q(\mu)$. Il résulte de l'inégalité de Hölder que $\Phi(\cdot)$ définie par : $\Phi(f) := \int fg d\mu$ est une forme linéaire continue sur $L^p(\mu)$, et que $\|\Phi\| := \sup_{\|f\|_p = 1} |\Phi(f)| \leq \|g\|_q$. Nous allons nous poser la question réciproque : toutes les formes linéaires continues sur $L^p(\mu)$ sont–elles de la forme ci–dessus, autrement dit peut–on identifier le dual de $L^p(\mu)$ à $L^q(\mu)$?

La réponse est (en général) non pour $p = +\infty$; et oui pour $p \in [1, +\infty[$.

Théorème 3.3.7. *Soit* μ *une mesure* σ*–finie sur* $(\Omega, \mathcal{F})$*,* $p \in [1, +\infty[$*, et* Φ *une forme linéaire continue sur* $L^p(\mu)$*. Alors il existe* $g \in L^q(\mu)$ *(où* $\frac{1}{q} + \frac{1}{p} = 1$*) t.q :*

$$\Phi(f) = \int fg d\mu,\ \ \forall\, f \in L^p(\mu)$$

De plus, $\|\Phi\| := \sup_{\|f\|_p=1} |\Phi(f)| = \|g\|_q$ *; i. e.* $L^q(\mu)$ *est isomorphe et isométrique au dual de* $L^p(\mu)$.

PREUVE ETAPE 1 : UNICITÉ (de la classe d'équivalence de g). Si g et g' satisfont les propriétés de l'énoncé, alors pour tout A avec $\mu(A) < \infty$, $\int_A (g - g')d\mu = 0$. Cela suffit à entraîner $g - g' = 0$ μ p.p. (utiliser la σ–finitude de μ !).
ETAPE 2 : EXISTENCE DE g DANS LE CAS $\mu(\Omega) < \infty$. Pour tout $A \in \mathcal{F}$, $\mathbf{1}_A \in L^p(\mu)$, donc on peut poser $\lambda(A) = \Phi(\mathbf{1}_A)$. λ est une application de $\mathcal{F}$ dans $\mathbb{R}$ qui vérifie :

(i) $\lambda(\emptyset) = 0$

(ii) $\lambda(\bigcup_1^\infty A_n) = \sum_1^\infty \lambda(A_n)$, si $A_k \cap A_\ell = \emptyset$, dès que $k \neq \ell$. En effet,

$$\mathbf{1}_{\{\bigcup_1^\infty A_n\}} = \lim_n \mathbf{1}_{\{\bigcup_1^n A_p\}} = \lim \sum_1^n \mathbf{1}_{A_p} = \sum_1^\infty \mathbf{1}_{A_p}$$

et la convergence a lieu dans $L^p(\mu)$.

Donc λ est une mesure signée. Or si $\mu(A) = 0$, $\mathbf{1}_A \sim 0$, donc $\lambda(A) = 0$.

Il résulte alors du Corollaire 3.3.6 qu'il existe $g \in L^1(\mu)$ t. q.

$$\lambda(A) = \int_A g d\mu \text{ , et par linéarité :}$$

$$\Phi(f) = \int fg d\mu \text{ pour toute } f \text{ étagée.}$$

cas $p > 1$ Soit h_n une suite croissante de fonctions étagées, t.q. $h_n \uparrow |g|$ On définit l'application sign de $\mathbb{R}$ dans $\mathbb{R}$ par : $\text{sign}(x) = +1$ si $x \geq 0$, -1 si $x < 0$. Alors $\text{sign}(g)(h_n)^{q-1}$ est étagée, et $h_n^q \leq h_n^{q-1}\text{sign}(g)g$, donc

$$\int h_n^q d\mu \leq \Phi(\text{sign}(g)h_n^{q-1}) \leq \|\Phi\|\|h_n\|_q^{q-1}$$

Donc $\|h_n\|_q \leq \|\Phi\|$, d'où par convergence monotone, $\|g\|_q \leq \|\Phi\|$, et $g \in L^q(\mu)$.

cas $p = 1$ Supposons que $\|g\|_\infty > \|\Phi\|$. Alors $h = \text{sign}(g)\mathbf{1}_{\{|g|>\|\Phi\|\}}$ est étagée et

$$\begin{aligned}\Phi(h) &= \int_{\{|g|>\|\Phi\|\}} |g| d\mu \\ &> \|\Phi\|\mu(|g| > \|\Phi\|) \\ &= \|\Phi\| \int |h| d\mu,\end{aligned}$$

ce qui est impossible, car $\Phi(h) \leq \|\Phi\| \displaystyle\int |h|d\mu$. Donc $\|g\|_\infty \leq \|\Phi\|$.

Donc, dans tous les cas, Φ et $f \to \displaystyle\int fgd\mu$ sont deux formes linéaires continues sur $L^p(\mu)$, qui coïncident sur le sous–ensemble dense des fonctions étagées, donc par le Théorème 3.2.3 elles sont identiques. De plus, on vient de montrer que $\|g\|_q \leq \|\Phi\|$, mais par Hölder, $\|\Phi\| \leq \|g\|_q$, d'où l'égalité.

Etape 3 : Existence de g dans le cas général : μ σ–finie

Il existe $\{A_n, n \geq 1\}$ t. q. $\displaystyle\bigcup_1^\infty A_n = \Omega$, les A_n sont disjoints, et $\mu(A_n) < \infty$, pour tout $n \geq 1$. Posons

$$h(\omega) = \sum_1^\infty [n^2\mu(A_n)]^{-1}\mathbf{1}_{A_n}(\omega)$$

Alors $h \in L^1(\mu)$. Donc la mesure $\tilde{\mu} = h.\mu$ est finie, et on peut lui appliquer le résultat de l'Etape 2.

Or $f \to h^{1/q}f$ est une isométrie de $L^p(\tilde{\mu})$ sur $L^p(\mu)$, et $\Psi(f) := \Phi(h^{1/p}f)$ définit une forme linéaire continue sur $L^p(\tilde{\mu})$, donc il existe $\tilde{g} \in L^q(\tilde{\mu})$ t. q. $\Psi(f) = \int f\tilde{g}d\tilde{\mu}$.

Posons $g = h^{1/q}\ \tilde{g}$ ($= \tilde{g}$ si $p = 1$). Alors :

si $p > 1$, $\displaystyle\int |g|^q d\mu = \int |\tilde{g}|^q d\tilde{\mu} = \|\Psi\|^q = \|\Phi\|^q$;

si $p = 1$, $\|g\|_\infty = \|\tilde{g}\|_\infty = \|\Psi\| = \|\Phi\|$.

Finalement, comme $\tilde{g} \cdot \tilde{\mu} = h^{1/q}g \cdot \mu$,

$$\Phi(f) = \Psi(h^{-1/q}f) = \int (h^{-1/q}f\tilde{g}d\tilde{\mu} = \int fgd\mu.$$

□

3.4 Compléments sur la théorie de l'intégration

Nous allons indiquer trois compléments à la théorie de l'intégration, pour l'essentiel sans démonstration. Les démonstrations manquantes peuvent se trouver par exemple dans W. Rudin [3].

3.4.1 Théorème de représentation de Riesz

Notons $C_K(\mathbb{R}^d)$ l'espace des fonctions continues à support compact, de $\mathbb{R}^d$ à valeurs dans $\mathbb{R}$, et $C_0(\mathbb{R}^d)$ l'espace des fonctions continues de $\mathbb{R}^d$ à valeurs dans $\mathbb{R}$ qui sont nulles à l'infini (i.e. t.q. $f(x) \to 0$, quand $\|x\| \to 0$).

Muni de la norme sup, $C_0(\mathbb{R}^d)$ est un espace de **Banach**, et $C_K(\mathbb{R}^d)$ est dense dans $C_0(\mathbb{R}^d)$. Nous allons caractériser le dual de $C_0(\mathbb{R}^d)$.

Théorème 3.4.1. *(Riesz) Pour toute forme linéaire continue Φ sur $C_0(\mathbb{R}^d)$, il existe une unique mesure signée μ définie sur $(\mathbb{R}^d, \mathcal{B}_d)$, telle que*

$$\Phi(f) = \int f d\mu, \text{ pour tout } f \in C_0(\mathbb{R}^d).$$

De plus

$$\|\Phi\| \stackrel{\triangle}{=} \sup_{\sup |f|=1} |\Phi(f)| = |\mu|(\mathbb{R}^d),$$

où $|\mu|$ désigne la variation totale de μ.

3.4.2 Intégrale de Stieltjes–Lebesgue

Plaçons–nous dans le cas $(\mathbb{R}, \mathcal{B})$. Si $\mathbb{P}$ est une probabilité, i.e. une mesure t. q. $\mathbb{P}(\Omega) = 1$, on peut lui associer sa fonction de répartition $F : \mathbb{R} \to [0,1]$ par $F(x) = \mathbb{P}((-\infty, x])$, et si $f \in L^1(\mathbb{P})$, on peut noter :

$$\int f dF = \int f d\mathbb{P}.$$

Cette notation est assez naturelle, puisque :

$$\text{Si } F \in C^1(\mathbb{R}), \int f d\mathbb{P} = \int_{\mathbb{R}} f(x) F'(x) dx.$$

Si $f \in C_K(\mathbb{R})$, $\displaystyle\int f d\mathbb{P} = \lim_{n\to\infty} \sum_{i=-\infty}^{+\infty} f(y_i)\left[F\left(\frac{i+1}{n}\right) - F\left(\frac{i}{n}\right)\right]$, avec $y_i \in [\frac{i}{n}, \frac{i+1}{n}]$, pour tout i.

Soit maintenant μ une mesure σ–finie sur $(\mathbb{R}, \mathcal{B})$, t.q. $\mu(A) < \infty$, pour tout $A \in \mathcal{B}$, A borné. On peut encore lui associer une fonction croissante et continue à droitee F, unique à une constante additive près, telle que

$$\text{pour tout } a, b \in \mathbb{R}, a < b,\ \ \mu(]a, b]) = F(b) - F(a).$$

Réciproquement, si $F = \mathbb{R} \to \mathbb{R}$ est monotone croissante, elle est continue sauf au plus en un ensemble dénombrable de points, et on peut choisir un représentant de F (dans la classe d'équivalence pour l'égalité λ p.p.) qui soit continu à droite : $\bar{F}(x) = \lim\limits_{\substack{y\downarrow x\\ y>x}} F(y)$.

La relation $(*)$ $\mu(]a, b]) = \bar{F}(b) - \bar{F}(a)$ définit une mesure sur la semi algèbre des intervalles semi–ouverts $]a, b]$, qui s'étend en une mesure σ–finie sur $\mathcal{B}$.

Soit maintenant une mesure signée μ sur $([0, 1], \mathcal{B}([0, 1]))$; Alors $\mu = \mu^+ - \mu^-$, et on lui associe une fonction $F = F^+ - F^-$, $F = [0, 1] \to \mathbb{R}$, qui est la différence de deux fonctions croissantes continues à droite.

La classe des fonctions $F = [0, 1] \to \mathbb{R}$ qui sont différences de deux fonctions croissantes coïncide avec la classe des fonctions à variations bornées F, i.e. t.q.

$$V(f) = \sup \sum_{0}^{n-1} |F(x_{i+1}) - F(x_i)| < \infty,$$

où le sup est étendu à toutes les suites. $0 = x_0 < x_1 < x_2 < \cdots < x_n = 1$ (décomposition de Jordan – analogue à la Proposition 3.3.5).

Soit donc F à variation bornée, $F = F^+ - F^-$ où F^+ et F^- sont deux fonctions croissantes. Soit $\bar{F}^+$ et $\bar{F}^-$ les représentants continus à droite de F^+ et F^-. On leur associe par $(*)$, deux mesures μ^+ et μ^- sur $([0, 1], \mathcal{B}([0, 1]))$, et la mesure signée $\mu = \mu^+ - \mu^-$.

Alors, si $f \in L^1(\mu)$, on notera $\displaystyle\int f dF$ l'intégrale $\displaystyle\int f d\mu$, et on l'appelle intégrale de Stieltjes–Lebesgue.

3.4.3 Théorème de différentiation de Lebesgue

Soit $f \in C_K(\mathbb{R})$. On sait que la fonction $F = \mathbb{R} \to \mathbb{R}$ définie par :

$$F(x) = \int_{\infty}^{x} f(y)dy$$

est de classe C^1, et que $F'(x) = f(x)$, pour tout $x \in \mathbb{R}$.

Soit maintenant $f \in L^1(\lambda)$, λ mesure de Lebesgue sur $(\mathbb{R}, \mathcal{B})$, et $F(x) = \int_{-\infty}^{x} f(y)dy$, $x \in \mathbb{R}$. On se pose la question : F est–elle dérivable, et si oui, sa dérivée coïncide–t–elle avec f ? Il est clair que l'on ne peut pas avoir un résultat aussi fort que ci–dessus. (ex : $f(x) = \mathbf{1}_{\{x \geq 0\}}, F(x) = x^+$).

Théorème 3.4.2. *(de différentiation de Lebesgue). Soit $f \in \mathcal{L}^1(\lambda)$.*

On pose $F(x) \stackrel{\triangle}{=} \int_{-\infty}^{x} f(y)dy$.

Alors il existe $N_f \in \mathcal{B}$, tel que $\lambda(N_f) = 0$ et sur $\mathbb{R} \backslash N_f$, F est dérivable et sa dérivée coïncide avec f.

□

On a une généralisation en dimension d :

Théorème 3.4.3. *Soit λ_d la mesure de Lebesgue sur $(\mathbb{R}^d, \mathcal{B}_d)$. Pour tout $f \in \mathcal{L}^1(\lambda_d)$, il existe $N_f \in \mathcal{B}_d$ avec $\lambda_d(N) = 0$ tel que pour tout $x \notin N_f$,*

$$\frac{1}{\lambda_d(Q_n)} \int_{Q_n} f(y)dy \to f(x),$$

où $\{Q_n\}$ est une suite décroissante d'hypercubes centrés en x, avec $\bigcap_1^\infty Q_n = \{x\}$.

□

Remarque 3.4.4. $(d = 1)$

(i) Une fonction dérivable p.p. n'est pas nécessairement l'intégrale de sa dérivée.

Exemple 1 : $F(x) = \mathbf{1}_{\{x \geq 0\}}$, $F'(x)$ est définie et nulle p.p.

Exemple 2 : Il existe F monotone croissante : $[0,1] \to [0,1]$, avec $F(0) = 0$ et $F(1) = 1$, et telle que pour tout $x \notin C$ (C ensemble triadique de Cantor), $F'(x)$ existe et est nulle. La mesure associée à F (cf. la section 3.4.2) est étrangère par rapport à la mesure de Lebesgue.

(ii) Soit $F : [0,1] \to \mathbb{R}$, t.q. F soit p.p. dérivable, et coïncide avec l'intégrale de sa dérivée. Alors il existe $f \in L^1([0,1], \mathcal{B}([0,1]), \lambda)$ t.q. pour tout $x \in [0,1]$,

$$F(x) = \int_0^x f(y)dy.$$

Ceci entraîne que F est à variation bornée, et que la mesure μ qui lui est associée (cf. section 3.4.2) est telle que μ^+ et μ^- sont absolument continues par rapport à la mesure de Lebesgue. Une fonction F qui possède ces propriétés est dite absolument continue. Une fonction F : $[0,1] \to \mathbb{R}$ est absolument continue si et seulement si pour tout $\varepsilon > 0$, il existe $\delta > 0$ t.q. pour toute collection $\{[a_n, b_n], n \geq 1\}$ d'intervalles disjoints de $[0,1]$,

$$\sum_0^\infty (b_n - a_n) < \delta \Rightarrow \sum_0^\infty |F(b_n) - F(a_n)| < \varepsilon.$$

(iii) Il convient de rapprocher ce que nous venons de discuter de la théorie de la dérivation au sens des distributions.

3.5 Exercices

Exercice 3.5.1. *Montrer que si $\|f\|_\infty < \infty$, $\mu(\{\omega; |f(\omega)| > \|f\|_\infty\}) = 0$.*

Exercice 3.5.2. *a) Soit $1 \leq p < r \leq +\infty$. Montrer que $L^r(\mu) \subset L^p(\mu)$ si et seulement si μ est une mesure finie, et que dans le cas où μ est une probabilité, $p \to \|f\|_p$ est croissante.*

b) Si $\Omega = \mathbb{N}$, $\mu = \sum_1^\infty \delta_n$, on note ℓ^p l'espace $L^p(\mu)$. Montrer que si $p < r$, $\ell^p \subset \ell^r$.

c) Montrer que si $\Omega = \mathbb{R}$, $\mathcal{F} = \mathcal{B}$, $\mu = \lambda$, mesure de Lebesgue, $p < r$, il existe $f \in L^p(\lambda)$ t.q. $f \notin L^r(\lambda)$ et il existe $g \in L^r(\lambda)$ tel que $g \notin L^p(\lambda)$.

Exercice 3.5.3. *Soit* $1 \leq p < q < \infty$. *Montrer que si* $f \in L^p(\mu) \cap L^q(\mu)$, *alors* $f \in L^r(\mu)$, *pour tout* $p \leq r \leq q$.

Exercice 3.5.4. *On suppose la mesure* μ *finie. Montrer que* $L^\infty(\mu) \subset \cap_{p\geq 1} L^p(\mu)$ *et que pour tout* $f \in L^\infty(\mu)$, $\|f\|_p \to \|f\|_\infty$, *quand* $p \to \infty$.

Exercice 3.5.5. *Soit* $f \in L^p(\mu)$ *et* $g \in L^q(\mu)$. *Montrer que si* $\frac{1}{p} + \frac{1}{q} = \frac{1}{r}$, *alors* $fg \in L^r(\mu)$.

Exercice 3.5.6. *Soit* $f \in L^p(\mathbb{R}^d, \mathcal{B}_d, \mu)$, *où* μ *désigne la mesure de Lebesgue sur* $(\mathbb{R}^d, \mathcal{B}_d)$ *et* $1 \leq p < \infty$. *Pour tout* $h \in \mathbb{R}^d$, *on pose* $\tau_h f = f(\cdot - h)$. *Montrer que* $\tau_h f \to f$ *dans* $L^p(\mathbb{R}^d, \mathcal{B}_d, \mu)$.

Exercice 3.5.7. *On se donne* $\phi \in C_b(\mathbb{R}^d; \mathbb{R}_+)$ *(donc en particulier bornée) qui vérifie* $\int_{\mathbb{R}^d} \phi(x)dx = 1$. *Pour tout* $\varepsilon > 0$, *on pose alors* $\phi_\varepsilon = \varepsilon^{-d}\phi(\cdot/\varepsilon)$. *Soit* $p \geq 1$. *Etant donné* $f \in L^p(\mathbb{R}^d)$, *on définit le produit de convolution*

$$(\phi_\varepsilon * f)(x) = \int_{\mathbb{R}^d} \phi_\varepsilon(x-y) f(y) dy.$$

Montrer que

1. *Si* f *est continue à support compact,* $\phi_\varepsilon * f \to f$ *uniformément dans* $\mathbb{R}^d$ *quand* $\varepsilon \to 0$.
2. *Si* $f \in L^p(\mathbb{R}^d)$, $\phi_\varepsilon * f \to f$ *dans* $L^p(\mathbb{R}^d)$.

Exercice 3.5.8. *Donner un contre–exemple au résultat du Théorème 3.2.5 dans le cas* $p = +\infty$.

Exercice 3.5.9. *Montrer l'unicité dans le Théorème 3.4.1*

Chapitre 4

Probabilité, Indépendance, Variables aléatoires

4.1 Probabilité

Soit $(\Omega, \mathcal{F})$ un espace mesurable (on dit aussi probabilisable).

Définition 4.1.1. *On appelle (mesure de) probabilité sur* $(\Omega, \mathcal{F})$ *une application* $\mathbb{P} : \mathcal{F} \to [0, 1]$ *qui vérifie :*

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

(ii) Pour tout suite $\{A_n, n \geq 1\} \subset \mathcal{F}$*, t.q.* $A_k \cap A_\ell = \emptyset$ *si* $k \neq \ell$

$$\mathbb{P}(\bigcup_1^\infty A_n) = \sum_1^\infty \mathbb{P}(A_n).$$

Remarque 4.1.2. *Une probabilité est une mesure* μ *de masse totale* 1*, i.e. t.q.* $\mu(\Omega) = 1$*. On a* $\mathbb{P}(A_n) \to \mathbb{P}(A)$ *dès que* $A_n \uparrow A$ *ou* $A_n \downarrow A$*, cf. Proposition 1.4.4.*

□

Remarque 4.1.3. *Dans le cas d'une mesure de probabilité on dit "presque sûrement" – en abrégé p.s. – au lieu de presque partout.*

□

Un triplet $(\Omega, \mathcal{F}, \mathbb{P})$, où $\mathbb{P}$ est une mesure de probabilité, est appelé espace de probabilité. Un tel espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$ sera supposé donné dans toute la suite de ce chapitre.

4.2 Probabilité conditionnelle

Définition 4.2.1. *Soit $A, B \in \mathcal{F}$, avec $P(A) > 0$. La probabilité de l'événement B, conditionné par A (ou "sachant A", ou "si A") est donnée par :*

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

A étant fixé tel que $\mathbb{P}(A) > 0$, on vérifie aisément que $\mathbb{P}(\cdot|A)$ est une probabilité sur $(\Omega, \mathcal{F})$. L'égalité de la définition se réécrit encore :

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

Cette dernière égalité peut s'écrire même lorsque $\mathbb{P}(A) = 0$, à condition de convenir que

$$0 \ \times \ \text{terme indéterminé} \ = 0$$

Proposition 4.2.2. *Soit $A_k \in \mathcal{F}$; $k = 1, \ldots, n$. Alors, avec la convention ci–dessus,*

$$\begin{aligned}\mathbb{P}(A_1 \cap A_2 \cap \cdots \cap A_n) = \mathbb{P}(A_1)\mathbb{P}(A_2|A_1)\mathbb{P}(A_3|A_1 \cap A_2) \cdots \\ \cdots \mathbb{P}(A_n|A_1 \cap A_2 \cap \cdots \cap A_{n-1})\end{aligned}$$

Preuve

a) Cas où $\mathbb{P}(A_1 \cap A_2 \cap \cdots \cap A_n) = 0$. Si $\mathbb{P}(A_1) = 0$, l'égalité est satisfaite. Si $\mathbb{P}(A_1) > 0$, il existe $m \in \{2, \cdots, n\}$ t.q.

$$\mathbb{P}(A_1 \cap \cdots \cap A_m) = 0 < \mathbb{P}(A_1 \cap \cdots \cap A_{m-1})$$

D'où $\mathbb{P}(A_m|A_1 \cap \cdots \cap A_{m-1}) = 0$ et l'égalité est satisfaite.

b) Cas où $\mathbb{P}(A_1 \cap A_2 \cap \cdots \cap A_n) > 0$. On démontre l'égalité par récurrence. Elle est satisfaite pour $n = 2$.
Supposons que l'on ait :

$$\mathbb{P}(A_1 \cap \cdots \cap A_{n-1}) = \mathbb{P}(A_1)P(A_2|A_1) \cdots \mathbb{P}(A_{n-1}|A_1 \cap \cdots \cap A_{n-2}).$$

Alors en multipliant l'égalité ci–dessus par :

$$\frac{\mathbb{P}(A_1 \cap \cdots \cap A_n)}{\mathbb{P}(A_1 \cap \cdots \cap A_{n-1})} = \mathbb{P}(A_n|A_1 \cap \cdots \cap A_{n-1})$$

on obtient le résultat.

$\square$

4.3 Evénements indépendants

Définition 4.3.1. *Deux événements A et $B \in \mathcal{F}$ sont dits indépendants si*

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

$\square$

Lorsque $\mathbb{P}(A) \cdot \mathbb{P}(B) \neq 0$, A et B sont indépendants si et seulement si $\mathbb{P}(B) = \mathbb{P}(B|A)$, ou $\mathbb{P}(A) = \mathbb{P}(A|B)$; i.e. A et B sont indépendants si et seulement si la probabilité de B n'est pas modifiée par l'information "A est réalisé".

Exemple 4.3.2. *Probabilité produit :*

$$(\Omega, \mathcal{F}, \mathbb{P}) = (\Omega_1 \times \Omega_2, \mathcal{F}_1 \otimes \mathcal{F}_2, \mathbb{P}_1 \times \mathbb{P}_2)$$

Soit $A = A_1 \times \Omega_2$, $B = \Omega_1 \times B_2$

$$\begin{aligned} A \cap B &= A_1 \times B_2 \\ \mathbb{P}(A \cap B) &= \mathbb{P}_1(A_1)\mathbb{P}_2(B_2) = \mathbb{P}(A)\mathbb{P}(B) \end{aligned}$$

Définition 4.3.3. *Une suite $\{A_1, \ldots, A_n\} \subset \mathcal{F}$ est dite indépendante si :*

$$\mathbb{P}(A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_k}) = \mathbb{P}(A_{i_1})\mathbb{P}(A_{i_2}) \ldots P(A_{i_k})$$

pour tout suite d'indices tous distincts $i_1, \ldots, i_k$ compris entre 1 et n.

$\square$

Remarque 4.3.4. Le fait que les A_n soient 2 à 2 indépendants n'entraîne pas que la suite $\{A_1, \ldots, A_n\}$ soit indépendante, *comme le montre l'exemple suivant : on joue 3 fois à Pile ou Face avec une pièce non truquée. On pose :*

$$X_i = \begin{cases} 1 & \text{si on obtient Pile au ième coup} \\ 0 & \text{si on obtient Face au ième coup} \end{cases}$$

On suppose la suite $\{X_1 = 1\}$, $\{X_2 = 1\}$, $\{X_3 = 1\}$ indépendante, chacun de ces événements étant de probabilité $1/2$.

On pose $A_{ij} = \{X_i = X_j\}$. Les événements A_{12}, A_{23}, A_{31} sont 2 à 2 indépendants, mais la suite n'est pas indépendante, puisque

$$\mathbb{P}(A_{12}) = \mathbb{P}(A_{23}) = \mathbb{P}(A_{31}) = \frac{1}{2},$$
$$\mathbb{P}(A_{12} \cap A_{23}) = \mathbb{P}(A_{23} \cap A_{31}) = \mathbb{P}(A_{12} \cap A_{31}) = \mathbb{P}(A_{12} \cap A_{23} \cap A_{31}) = \frac{1}{4}.$$

Proposition 4.3.5. *Si la suite $\{A_1, \ldots, A_n\}$ est indépendante, les 2^n suites $\{A'_1, \ldots, A'_n\}$ obtenues en choisissant $A'_i = A_i$ ou A_i^c sont encore indépendantes.*

Preuve Il suffit d'établir que l'indépendance de la suite $\{A_1, \ldots, A_n\}$ entraîne celle de $\{A_1, \ldots, A_{\ell-1}, A_\ell^c, A_{\ell+1}, \ldots, A_n\}$, pour tout $\ell \in \{1, \ldots, n\}$, et d'itérer ce résultat. La notion d'indépendance d'une suite ne faisant pas intervenir l'ordre, il suffit en fait d'établir le résultat pour $\ell = 1$. Pour établir l'indépendance de $\{A_1^c, A_2, \ldots, A_n\}$, il suffit d'établir que pour tout $i_1, \ldots, i_k$ indices tous distincts dans $\{2, \ldots, n\}$,

$$\mathbb{P}(A_1^c \cap A_{i_1} \cap \ldots \cap A_{i_k}) = \mathbb{P}(A_1^c)\mathbb{P}(A_{i_1}) \cdots \mathbb{P}(A_{i_k})$$

Or le membre de gauche est égal à :

$$\begin{aligned}\mathbb{P}(A_{i_1} \cap \ldots \cap A_{i_k}) - \mathbb{P}(A_1 \cap A_{i_1} \cap \ldots \cap (A_{i_k}) &= \\ = (1 - \mathbb{P}(A_1))\mathbb{P}(A_{i_1}) \ldots \mathbb{P}(A_{i_k}).&\end{aligned}$$

□

Etant donnée une suite $\{A_1, \ldots, A_n\}$, définissons les tribus $\mathcal{F}_i = \{\emptyset, A_i, A_i^c, \Omega\}$.

On déduit alors aisément de la Proposition 4.3.5 :

Corollaire 4.3.6. *La suite $\{A_1, \ldots, A_n\}$ est indépendante si et seulement si l'égalité*

$$\mathbb{P}(B_1 \cap \ldots \cap B_n) = \mathbb{P}(B_1) \ldots \mathbb{P}(B_n)$$

est satisfaite pour tout choix des $B_i \in \mathcal{F}_i$; $i = 1 \ldots n$.

Définition 4.3.7. *Une suite infinie d'événements $\{A_n, n \geq 1\}$ est dite indépendante si toutes les sous–suites finies de $\{A_n, n \geq 1\}$ sont indépendantes, i.e. si pour tout k, pour tout $i_1, \ldots, i_k$ entiers tous distincts,*

$$\mathbb{P}(A_{i_1} \cap \cdots \cap A_{i_k}) = \mathbb{P}(A_{i_1}) \cdots \mathbb{P}(A_{i_k})$$

□

Lemme 4.3.8. *(Borel–Cantelli)*

(i) Pour toute suite infinie $\{A_n, n \geq 1\} \subset \mathcal{F}$*,*

$$\sum_1^\infty \mathbb{P}(A_n) < \infty \Rightarrow \mathbb{P}(\limsup A_n) = 0.$$

(ii) Si la suite $\{A_n, n \geq 1\}$ *est indépendante,*

$$\sum_1^\infty \mathbb{P}(A_n) = +\infty \Rightarrow \mathbb{P}(\limsup A_n) = 1.$$

Remarque 4.3.9. 1. *Dans le cas d'une suite indépendante, on a en particulier une loi* $0-1$ *:* $\mathbb{P}(\limsup A_n)$ *ne prend que les valeurs* 0 *et* 1*.*

2. *(ii) est faux sans l'hypothèse d'indépendance, comme le montre l'exemple suivant :*

$$A_n \equiv A,\ \ 0 < \mathbb{P}(A) < 1$$

$$\limsup A_n = A$$

Preuve

(i)

$$\begin{aligned}
\mathbb{P}(\limsup A_n) &= \mathbb{P}\left[\bigcap_m \left(\bigcup_{n=m}^\infty A_n\right)\right] \\
&= \lim_m \mathbb{P}\left(\bigcup_{n=m}^\infty A_n\right) \\
&\leq \lim_m \sum_m^\infty \mathbb{P}\left(A_n\right) \\
&= 0,
\end{aligned}$$

puisque par hypothèse la série $\sum_1^\infty \mathbb{P}(A_n)$ converge.

(ii)

$$\mathbb{P}\left(\bigcup_m^\infty A_n\right) = 1 - \mathbb{P}\left(\bigcap_m^\infty A_n^c\right)$$

Il suffit donc de montrer que $\mathbb{P}\left(\bigcap_m^\infty A_n^c\right) = 0, \forall m$. D'après l'indépendance,

$$\mathbb{P}\left(\bigcap_m^\infty A_n^c\right) = \prod_m^\infty (1 - \mathbb{P}(A_n))$$

En posant $\log 0 = -\infty$, et en utilisant l'inégalité $\log(1+x) \leq x$, pour tout $x \geq -1$, et l'hypothèse, on obtient :

$$\begin{aligned} \log\left[\prod_m^\infty (1 - \mathbb{P}(A_n))\right] &= \sum_m^\infty \log(1 - \mathbb{P}(A_n)) \\ &\leq -\sum_m^\infty \mathbb{P}(A_n) = -\infty. \end{aligned}$$

□

4.4 Variable aléatoire

Définition 4.4.1. *On appelle* **variable aléatoire réelle** *une application mesurable $X : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B})$.*

On appelle **vecteur aléatoire (de dimension d)** *une application mesurable $X : (\Omega, \mathcal{F}) \to (\mathbb{R}^d, \mathcal{B}_d)$.*

□

On emploiera souvent “variable aléatoire” (en abrégé v.a.) pour désigner indifféremment une variable aléatoire réelle (en abrégé v.a.r.) ou un vecteur aléatoire.

Remarque 4.4.2. *Pour vérifier que : $X : \Omega \to \mathbb{R}$ est une v.a.r., il suffit par exemple de vérifier que pour tout $x \in \mathbb{R}$, $\{\omega; X(\omega) \leq x\} \in \mathcal{F}$.*

Définition 4.4.3. *On appelle* **tribu naturelle** *du vecteur aléatoire X (de dimension d) la tribu*

$$\sigma(X) \stackrel{\triangle}{=} X^{-1}(\mathcal{B}_d) = \{X^{-1}(B), B \in \mathcal{B}_d\}$$

Si $\mathcal{G}$ et $\mathcal{H}$ sont deux tribus de parties de Ω, on note $\mathcal{G} \vee \mathcal{H} = \sigma(\mathcal{G} \cup \mathcal{H})$ la plus petite tribu qui contient $\mathcal{G}$ et $\mathcal{H}$. De même, $\mathcal{G}_1 \vee \mathcal{G}_2 \vee \cdots \vee \mathcal{G}_n = \sigma(\mathcal{G}_1 \cup \mathcal{G}_2 \cup \cdots \cup \mathcal{G}_n)$.

Proposition 4.4.4. *Soit $\{X_i, i = 1, \dots, n\}$ des vecteurs aléatoires de dimension respective $d_1, \dots, d_n$. On pose :*

$$X = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}$$

Alors X est un v.a. de dimension $\sum_{i=1}^n d_i$ et $\sigma(X) = \sigma(X_1) \vee \cdots \vee \sigma(X_n)$.

Preuve (dans le cas $n = 2$, $d_1 = d_2 = 1$).

$$\begin{aligned} \sigma(X_1) &= X_1^{-1}(\mathcal{B}_1) = X^{-1}(\{A_1 \times \mathbb{R}; A_1 \in \mathcal{B}_1\}) \\ \sigma(X_2) &= X^{-1}(\{\mathbb{R} \times A_2; A_2 \in \mathcal{B}_1\}) \\ \sigma(X_1) \vee \sigma(X_2) &= \sigma(X^{-1}(\mathcal{C})) \end{aligned}$$

où (attention, $\mathcal{C}$ n'est pas un π–système !)

$$\mathcal{C} = \{A_1 \times \mathbb{R},\ A_1 \in \mathcal{B}_1\} \cup \{\mathbb{R} \times A_2, A_2 \in \mathcal{B}_1\}$$

Or $\sigma(\mathcal{C}) = \mathcal{B}_2$, et $\sigma(X^{-1}(\mathcal{C})) = X^{-1}(\sigma(\mathcal{C})) = \sigma(X)$. □

Définition 4.4.5. *Etant donné un vecteur aléatoire X de dimension d, on appelle loi de probabilité de X la mesure de probabilité $\mathbb{P}_X$ sur $(\mathbb{R}^d, \mathcal{B}_d)$ définie par : $\mathbb{P}_X = \mathbb{P}X^{-1}$.*

Etant donné une mesure de probabilité Q sur $(\mathbb{R}^d, \mathcal{B}_d)$, il existe un espace de probabilité, et un v.a. défini sur cet espace, de dimension d, de loi Q. Il suffit de choisir $\Omega = \mathbb{R}^d$, $\mathcal{F} = \mathcal{B}_d$, $\mathbb{P} = Q$, $X(\omega) = \omega$.

Définition 4.4.6. *Soit X une v.a.r. [resp. un vecteur aléatoire de dimension d]. On appelle fonction de répartition de X la fonction :*

$$F_X : \mathbb{R}[resp.\mathbb{R}^d] \to [0, 1]$$

définie par $F_X(x) = \mathbb{P}(X \leq x) = \mathbb{P}_X(] - \infty, x])$. $[resp. F_X(x_1, \dots, x_d) = \mathbb{P}(X_1 \leq x_1, \dots, X_d \leq x_d)]$.

Plus généralement à une probabilité Q sur $(\mathbb{R}^d, \mathcal{B}_d)$, on associe sa fonction de répartition.

Propriétés des fonctions de répartition.

a) **Cas** $d = 1$

(i) F est croissante

(ii) F est continue à droite :

$$x_n \downarrow x \Rightarrow F(x_n) \to F(x)$$

(iii) Comportement à l'infini :

$$\begin{aligned} F(x) &\to 0, \text{ si } x \to -\infty \\ F(x) &\to 1, \text{ si } x \to +\infty. \end{aligned}$$

b) **cas** $d > 1$ Soit $a, b, \in \mathbb{R}^d$, avec $a < b$. Soit A le pavé $\prod_1^d]a_i, b_i]$. Ce pavé a 2^d sommets, qui sont les $x = (x_i, \ldots, x_d)$, où chaque x_i est égal soit à a_i, soit à b_i.
On pose $\Delta_A F = \sum_{x \text{ sommet de } A} \text{sign}_A(x)\ F(x)$, où $\text{sign}_A(x) = +1$, si le nombre de x_i égaux à a_i est pair, $\text{sign}_A(x) = -1$ si ce nombre est impair.

(i) $\Delta_A F \geq 0$, pour tout $A = \prod_1^d]a_i, b_i] \ \ , a < b$.

(ii) Si $x(k) \downarrow x$ [i.e. $x_i^{(k)} \downarrow x_i$, $\forall i$], alors $F(x^{(k)}) \to F(x)$.

(iii) Si il existe j t.q. $x_j \to -\infty$ $F(x) \to 0$ Si $x_j \to +\infty$, $j = 1, \ldots, d$, $F(x) \to 1$.

Proposition 4.4.7. *Soit F une application de $\mathbb{R}^d$ dans $\mathbb{R}$ qui vérifie (i), (ii) et (iii). Alors il existe une et une seule mesure de probabilité $\mathbb{P}$ sur $(\mathbb{R}^d, \mathcal{B}_d)$, dont F est la fonction de répartition.*

Preuve **Existence** (cas $d = 1$). On définit $\mathbb{P}$ sur la semi–algèbre $\mathcal{J}$ des intervalles de la forme $]a, b]$, par $\mathbb{P}(]a, b]) = F(b) - F(a)$. Grâce à la continuité à droite de F, on démontre que $\mathbb{P}$ est σ–additive sur $\mathcal{J}$, comme à la Proposition 1.5.1. L'existence découle alors du Théorème 1.4.6.

Unicité Elle résulte du Théorème 1.4.11, en remarquant que $\mathcal{P} = (\{x \leq x_0\}, x_0 \in \mathbb{R}^d)$ est un π–système, t.q. $\sigma(\mathcal{P}) = \mathcal{B}_d$. $\square$

Définition 4.4.8. *Soit f une application mesurable de $\mathbb{R}^d$ à valeurs dans $\mathbb{R}_+$. On dit qu'une probabilité $\mathbb{Q}$ sur $(\mathbb{R}^d, \mathcal{B}_d)$ admet la densité f (par rapport à la mesure de Lebesgue) si pour tout $B \in \mathcal{B}_d$, $\mathbb{Q}(B) = \int_B f(x)dx$; alors nécessairement $\int_{\mathbb{R}^d} f(x)dx = 1$.*

4.5 Exemples de lois de probabilité

4.5.1 Lois discrètes

La **mesure de Dirac** au point x ; $\delta_x(A) = \begin{cases} 1 & \text{si } x \in A \\ 0 & \text{si } x \notin A \end{cases}$ est une mesure de probabilité. C'est la loi d'une v.a. X t.q. $\mathbb{P}(X = x) = 1$, donc d'une "variable aléatoire" dont le comportement n'est pas du tout aléatoire !

1. **Loi de Bernoulli** notée $B(p)$, $(0 < p < 1)$:

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

   C'est la loi du gain à l'issue d'un jet de Pile ou Face.

2. **Loi binômiale** Soit $Y_1, \ldots, Y_n$ des v.a.r.. **i.i.d.** (indépendantes et identiquement distribuées), de loi commune $B(p)$. Alors la loi de $X = Y_1 + \cdots + Y_n$ est notée $B(n, p)$. Elle est caractérisée par :

$$\mathbb{P}(X = k) = C_n^k p^k (1 - p)^{n-k}; \quad k = 0, 1, \ldots, n \text{ (exercice)}$$

   C'est la loi du gain cumulé à l'issue de n coups de Pile ou Face mutuellement indépendants, avec la probabilité p de gagner à chaque coup.

3. **Loi géométrique** C'est la loi du nombre de coups nécessaires pour obtenir un Pile, en jouant à Pile ou Face, avec la probabilité p d'obtenir Pile à chaque coup, notée $G(p)$.

$$\mathbb{P}(X = k) = (1 - p)^{k-1} p, \ k \in \mathbb{N}^*$$

   En fait on peut aussi la caractériser par

$$\mathbb{P}(X > k) = (1 - p)^k, \ k \in \mathbb{N}.$$

4. **Loi binômiale négative** C'est la loi du nombre de coups nécessaires pour obtenir m Piles, en jouant à Pile ou Face, avec probabilité p d'obtenir Pile à chaque coup. Pour $k \geq m$,

$$\begin{aligned} \mathbb{P}(X = k) &= C_{k-1}^{m-1} p^{m-1}(1-p)^{k-m} p \\ &= C_{k-1}^{m-1} p^{m}(1-p)^{k-m} \end{aligned}$$

5. **Loi de Poisson** de paramètre $\lambda > 0$ notée $\mathcal{P}(\lambda)$

$$\mathbb{P}(X = k) = e^{-\lambda}\frac{\lambda^k}{k!}, k \geq 1.$$

La loi de Poisson est la loi limite de lois binômiales, cf. l'exercice 4.8.6 ci–dessous

4.5.2 Lois à densité

1. **Loi uniforme** sur $[0, 1]$, notée $\mathcal{U}([0, 1])$. C'est la loi sur $(\mathbb{R}, \mathcal{B})$ de densité $\mathbf{1}_{[0,1]}$. C'est la limite quand $M \to \infty$ des lois uniformes sur

$$\{\frac{1}{M}, \frac{2}{M}, \ldots, 1\}.$$

On peut considérer plus généralement la loi uniforme sur $[a, b]$, pour tout $a < b$. Sa densité est la fonction $(b-a)^{-1}\mathbf{1}_{[a,b]}$.

2. **Loi exponentielle** $\mathcal{E}(\lambda), \lambda > 0$. C'est la loi sur $(\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+))$ de densité $\lambda e^{-\lambda x}$. $X \simeq \mathcal{E}(\lambda)$ ssi $P(X > x) = e^{-\lambda x}$, pour tout $x > 0$. Cette caractérisation est celle qu'il faut toujours utiliser !

3. **Loi de Laplace** (ou exponentielle symétrique) sur $\mathbb{R}$ de paramètre $\lambda > 0$. C'est la loi de densité

$$\frac{\lambda}{2}\exp(-\lambda|x|).$$

4. **Loi gaussienne** (ou normale) sur $\mathbb{R}$

$$N(\mu, \sigma^2) = \begin{cases} \text{loi de densité } \frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) & \text{, si } \sigma > 0; \\ \delta_\mu & \text{, si } \sigma = 0. \end{cases}$$

5. **Loi de Cauchy** : loi sur $\mathbb{R}$ de densité

$$\frac{1}{\pi}\,\frac{1}{1+x^2}.$$

6. **Loi Gamma** (p, λ) Pour $p > 0$, on pose $\Gamma(p) = \int_0^\infty e^{-x} x^{p-1} dx$. Notons que $\Gamma(p) = (p-1)\Gamma(p-1)$, $\Gamma(k) = (k-1)!$, $\Gamma(1) = 1$, $\Gamma(1/2) = \sqrt{\pi}$. Pour $p, \lambda > 0$, on définit la loi $\Gamma(p, \lambda)$ comme la loi de probabilité sur $\mathbb{R}_+$ de densité

$$\frac{\lambda^p}{\Gamma(p)} e^{-\lambda x} x^{p-1}.$$

Il résulte des deux exercices 4.8.17 et 4.8.18 que si $Y_1, \ldots, Y_n$ sont des v. a. r. i. i. d., de loi commune la loi $N(0,1)$, alors

$$X = Y_1^2 + \cdots + Y_n^2 \simeq \Gamma\left(\frac{n}{2}, \frac{1}{2}\right).$$

Cette loi est appelée la loi **du χ^2 à n degrés de liberté**. Elle joue un rôle très important en statistique (test du Chi-deux, cf. section 7.6 ci–dessous).

7. **Loi bêta** La loi bêta(r, s), $r, s > 0$, est la loi sur $[0,1]$ de densité

$$\frac{1}{B(r,s)} x^{r-1}(1-x)^{s-1}.$$

Le coefficient de normalisation vaut (voir l'exercice 4.8.17)

$$B(r,s) = \int_0^1 x^{r-1}(1-x)^{s-1} dx = \frac{\Gamma(r)\Gamma(s)}{\Gamma(r+s)}.$$

4.6 Variables aléatoires indépendantes

Définition 4.6.1. — *La suite finie $\mathcal{C}_1 \ldots \mathcal{C}_n$ de sous classes de $\mathcal{F}$ est dite indépendante si l'égalité :*

$$\mathbb{P}(B_1 \cap B_2 \cap \cdots \cap B_n) = \mathbb{P}(B_1)\mathbb{P}(B_2) \times \cdots \times \mathbb{P}(B_n)$$

est vraie pour tout $B_i \in \mathcal{C}_i \cup \{\Omega\}$, $i = 1, \ldots, n$.

— *La suite des vecteurs aléatoires* $X_1, \dots, X_n$ *(de dimension respective* $d_1, \dots, d_n$*) est dite indépendante si la suite des tribus* $\sigma(X_1), \dots, \sigma(X_n)$ *est indépendante.*

$\square$

Nous allons tout d'abord relier cette notion d'indépendance avec celle de la Section 3.

Proposition 4.6.2. *Soit* $(A_1, \dots, A_n) \subset \mathcal{F}$*. Pour tout* $i \leq n$*, on note* $\mathcal{F}_i = \{\emptyset, A_i, A_i^c, \Omega\}$*. Les trois conditions suivantes sont équivalentes :*

(i) La suite d'événements $(A_1, \dots, A_n)$ *est indépendante (au sens de la Définition 4.3.2).*

(ii) La suite de tribus $(\mathcal{F}_1, \dots, \mathcal{F}_n)$ *est indépendante.*

(iii) La suite de v.a.r. $(1_{A_1}, \dots, 1_{A_n})$ *est indépendante.*

Preuve (i) $\Leftrightarrow$ (ii) a été établi au Corollaire 4.3.3. (ii) $\Leftrightarrow$ (iii) résulte de ce que $\mathcal{F}_i = \sigma(\mathbf{1}_{A_i})$, $1 \leq i \leq n$. $\square$

Théorème 4.6.3. *Soit* $\mathcal{C}_1, \dots, \mathcal{C}_n$ *des* π*–systèmes inclus dans* $\mathcal{F}$*. Si la suite* $\mathcal{C}_1, \dots, \mathcal{C}_n$ *est indépendante, alors la suite de tribus* $(\sigma(\mathcal{C}_1), \dots, (\mathcal{C}_n))$ *est indépendante.*

Preuve Soit $A_2, \dots, A_n$ avec $A_i \in \mathcal{C}_i \cup \{\Omega\}$, et $P(A_2 \cap \dots \cap A_n) > 0$. Soit $Q : \mathcal{F} \to [0, 1]$ défini par :

$$Q(A) = \frac{\mathbb{P}(A \cap A_2 \cap \dots \cap A_n)}{\mathbb{P}(A_2 \cap \dots \cap A_n)}$$

Q est une probabilité, qui coïncide avec $\mathbb{P}$ sur $\mathcal{C}_1$, donc sur $\sigma(\mathcal{C}_1)$ [cf. Théorème 1.4.11].

L'égalité :

$$\mathbb{P}(A_1 \cap A_2 \cap \dots \cap A_n) = \mathbb{P}(A_1)\mathbb{P}(A_2) \cdots \mathbb{P}(A_n)$$

est donc vraie pour tout $A_1 \in \sigma(\mathcal{C}_1), A_2 \in \mathcal{C}_2 \cup \{\Omega\}, \dots A_n \in \mathcal{C}_n \cup \{\Omega\}$. Par récurrence, on démontre que cette égalité est vraie pour tout $A_i \in \sigma(\mathcal{C}_i)$, $i \leq n$. $\square$

Théorème 4.6.4. *Pour que la suite de v.a.* $X_1, \dots, X_n$ *(de* dim *resp.* $d_1, \dots, d_n$*) soit indépendante, il faut et il suffit que la loi du vecteur aléatoire* $X = (X_1, \dots, X_n)$ *sur* $\mathbb{R}^{\sum_1^n d_i}$ *soit le produit des lois des* X_i*, i.e.*

$$\mathbb{P}_X(dx_1, \dots, dx_n) = \mathbb{P}_{X_1}(dx_1) \times \dots \times \mathbb{P}_{X_n}(dx_n)$$

Preuve Soit $B_1 \in \mathcal{B}_{d_1}, \dots, B_n \in \mathcal{B}_{d_n}$. Si la suite $X_1, \dots, X_n$ est indépendante,

$$\begin{aligned}\mathbb{P}_X(B_1 \times \cdots \times B_n) &= \mathbb{P}((X_1, \dots, X_n) \in B_1 \times \cdots \times B_n) \\ &= \mathbb{P}(\{X_1 \in B_1\} \cap \cdots \cap \{X_n \in B_n\}) \\ &= \mathbb{P}_{X_1}(B_1) \dots \mathbb{P}_{X_n}(B_n),\end{aligned}$$

et alors $\mathbb{P}_X$ et $\mathbb{P}_{X_1} \times \cdots \times \mathbb{P}_{X_n}$ coïncident sur le π–système des “rectangles”, qui engendre la tribu $\mathcal{B}_{\sum_{i=1}^n d_i}$, d'où l'égalité des probabilités.

Réciproquement, si cette égalité est satisfaite, alors pour tout $B_1 \in \mathcal{B}_{d_1}, \dots, B_n \in \mathcal{B}_{d_n}$,

$$\mathbb{P}(\{X_1 \in B_1\} \cap \cdots \cap \{X_n \in B_n\}) = \mathbb{P}(X_1 \in B_1) \cdots \mathbb{P}(X_n \in B_n)$$

i.e.

$$\mathbb{P}(A_1 \cap \cdots \cap A_n) = \mathbb{P}(A_1) \cdots \mathbb{P}(A_n), \forall\, A_i \in \sigma(X_i),\ 1 \leq i \leq n.$$

□

Corollaire 4.6.5. *Si la suite des v.a. $X_1 \dots X_n$ (de dimension respective $d_1, \dots, d_n$) est indépendante, et si chaque $\mathbb{P}_{X_i}$ admet une densité f_i par rapport à la mesure de Lebesgue sur $\mathbb{R}^{d_i}$, alors la probabilité $\mathbb{P}_X$ admet la densité :*

$$f(x_1, \dots, x_n) = f_1(x_1) \times \cdots \times f_n(x_n)$$

par rapport à la mesure de Lebesgue sur $\mathbb{R}^{\sum_1^n d_i}$.

Proposition 4.6.6. *Soit $X_1, \dots, X_n$ une suite de vecteurs aléatoires indépendants, $(X'_1, \dots, X'_n)$ une permutation de $(X_1, \dots, X_n)$. Alors les k vecteurs aléatoires :*

$$(X'_1, X'_2, \dots, X'_{n_1})', (X'_{n_1+1}, \dots, X'_{n_2})', \dots, (X'_{n_{k-1}+1}, \dots, X'_{n_k})'$$

(où $1 \leq n_1 < n_2 < \cdots < n_k \leq n$) sont indépendants.

Preuve Utiliser le fait que l'indépendance se conserve par permutation, 4.6.4, et l'associativité du produit de mesures. □

Proposition 4.6.7. *Soit $X_1, \dots, X_n$ une suite de v.a. indépendants (de dimension respective $d_1, \dots, d_n$), et $f_i : \mathbb{R}^{d_i} \to \mathbb{R}^{d'_i}$, $i = 1, \dots, n$, des applications mesurables. Alors $f_1(X_1), \dots, f_n(X_n)$ est une suite de vecteurs aléatoires indépendants.*

PREUVE Il suffit de remarquer que $\sigma(f_i(X_i)) \subset \sigma(X_i)$ □

Définition 4.6.8. *Une suite infinie de sous classes de $\mathcal{F}$ est indépendante si toutes les sous suites finies sont indépendantes.*

Une suite infinie de v.a. est dite indépendante si toutes les sous suites finies sont indépendantes.

4.7 Moments des variables aléatoires

Soit X une v.a.r. définie sur l'espace de probabilité $(\Omega, \mathcal{F}, P)$. X est dite **quasi–intégrable** si au moins l'une des deux quantités $\int_\Omega X^+ d\mathbb{P}$ ou $\int_\Omega X^- d\mathbb{P}$ est finie, **intégrable** si $\int_\Omega |X| d\mathbb{P} = \int_\Omega X^+ d\mathbb{P} + \int_\Omega X^- d\mathbb{P} < \infty$.

Définition 4.7.1. *Soit X une v.a.r. intégrable (ou seulement quasi–intégrable). on définit alors* **l'espérance mathématique** *de X par :*

$$\mathbb{E}(X) := \int_\Omega X(\omega)\mathbb{P}(d\omega)$$

Soit X un vecteur aléatoire de dimension d, $X = (X_1, \dots, X_d)'$ tel que chaque $X_i (i = 1, \dots, d)$ soit une v.a.r. intégrable. On définit alors l'espérance de X, $\mathbb{E}(X)$, comme le vecteur de coordonnées $\mathbb{E}(X_i), i = 1 \cdots d$.

□

Tous les Théorèmes que nous avons démontrés pour les intégrales s'appliquent aux espérances, avec la particularité que $\mathbb{P}$ est une mesure finie. Cela entraîne par exemple que les constantes sont $\mathbb{P}$–intégrables.

On déduit du Théorème 2.3.14 :

Proposition 4.7.2. *Soit X un v.a. de dimension d, $\varphi : \mathbb{R}^d \to \bar{\mathbb{R}}$ une application mesurable. Alors $\varphi \circ X$ est $\mathbb{P}$ (quasi) intégrable si et seulement si φ est $\mathbb{P}_X$ (quasi) intégrable, et dans ce cas :*

$$\int_\Omega \varphi \circ X(\omega)\mathbb{P}(d\omega) = \int_{\mathbb{R}^d} \varphi(x)\mathbb{P}_X(dx)$$

En particulier, si $d = 1$ et X est (quasi) intégrable,

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

□

Cette proposition est essentielle pour le calcul pratique des espérances.

Etant donné Q une probabilité sur $(\mathbb{R}, B)$, si $\int_{\mathbb{R}} |x| dQ(x) < \infty$, alors on peut définir **l'espérance de la probabilité** Q comme la valeur commune des espérances des v.a.r. de loi Q, grâce à 4.7.2 avec $\varphi(x) = x$.

Proposition 4.7.3. *(Inégalité de Jensen).*

Soit X un vecteur aléatoire de dimension d, et $\varphi : \mathbb{R}^d \to \mathbb{R}$ une fonction **convexe**. *Si X est intégrable, alors $\varphi(X)$ est quasi–intégrable, et*

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

Preuve On va utiliser la propriété suivante des fonctions convexes : pour tout $x_0 \in \mathbb{R}^d$, il existe $a \in \mathbb{R}^d$ t.q. :

$$(a, x - x_0) + \varphi(x_0) \leq \varphi(x), \forall\, x \in \mathbb{R}^d,$$

$$\text{d'où avec } x_0 = \mathbb{E}(X), \quad (a, X - \mathbb{E}(X)) + \varphi(\mathbb{E}(X)) \leq \varphi(X) \text{ p.s.}$$

Donc $\varphi(X)$ est quasi–intégrable, et il reste à prendre l'espérance dans cette inégalité. □

L'inégalité de Jensen ne s'étend pas au cas où l'on remplace l'espérance par une intégrale par rapport à une mesure quelconque.

Soit X une v.a.r. Si X^2 est intégrable (on dit que X est de carré intégrable), alors X est intégrable [appliquer l'inégalité de Cauchy–Schwarz à $1 \cdot |X|$ ou encore l'inégalité de Jensen avec la fonction convexe $\varphi(x) = x^2$].

Si X et Y sont de carré intégrable, alors $X + Y$ est de carré intégrable et XY est intégrable (utiliser $(x + y)^2 \leq 2x^2 + 2y^2$, $|xy| \leq \frac{1}{2}(x^2 + y^2)$).

On peut alors poser la :

Définition 4.7.4. *Soit X une v.a.r. de puissance p ième intégrable ($p > 0$). On appelle moment d'ordre p de X la quantité :* $\mathbb{E}(X^p)$.

Si X est de carré intégrable, on appelle variance de X la quantité :

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

Si X et Y sont deux v.a.r. de carré intégrable, on appelle covariance de X et Y la quantité :

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

Il est clair que la covariance de deux v.a.r. est symmétrique, au sens où $\mathrm{Cov}\,(X,Y) = \mathrm{Cov}\,(Y,X)$. Il est aussi facile de vérifier que l'application $X, Y \to \mathrm{Cov}\,(X,Y)$ est bilinéaire. En particulier $\mathrm{Cov}\,(X, aY + bZ) = a\,\mathrm{Cov}\,(X,Y) + b\,\mathrm{Cov}\,(X,Z)$.

Définition 4.7.5. *Soit X un v.a. de dimension d, t.q. chaque X_i $(i = 1 \cdots d)$ soit de carré intégrable. Alors on appelle*

— *matrice des moments d'ordre 2 de X la matrice :*

$$\mathbb{E}(XX')$$

— *matrice de covariance de X la matrice :*

$$\begin{aligned} Cov(X) &= \mathbb{E}\left((X - \mathbb{E}(X))(X - \mathbb{E}(X))'\right) \\ &= \mathbb{E}(XX') - \mathbb{E}(X)\mathbb{E}(X)', \\ i.\ e.\ Cov(X)_{ij} &= Cov(X_i, X_j),\ 1 \le i, j \le d. \end{aligned}$$

Ces deux matrices sont autoadjointes et ≥ 0.

Le caractère semidéfini positif de la matrice $\mathrm{Cov}(X)$ se déduit aisément du fait que pour tout $u \in \mathbb{R}^d$, $\mathrm{Var}(\langle u, X\rangle) = u'\mathrm{Cov}\,(X)u$, cf. Exercice 4.8.23. On pose aussi la définition suivante, qui généralise à la fois la covariance de deux v.a.r., et la notion de matrice de covariance d'un vecteur aléatoire.

Définition 4.7.6. *Soit X et Y deux v.a. de dimension resp. d et k, dont toutes les coordonnées sont de carré intégrable. On définit la matrice de covariance de X et Y comme la matrice $d \times k$ définie par*

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

Cette covariance de deux vecteurs aléatoires n'est pas symétrique dans le cas où $\inf(d,k) > 1$, puisque $\mathrm{Cov}\,(Y,X) = \mathrm{Cov}\,(X,Y)'$. Elle est encore bilinéaire. Clairement $\mathrm{Cov}\,(X) = \mathrm{Cov}\,(X,X)$, où le memre de gauche de cette égalité est pris au sens de la définition 4.7.5, et celui de droite au sens de la définition 4.7.6.

Proposition 4.7.7. *(Inégalité de Bienaymé–Tchebicheff) Soit X une v.a.r., et $c, p > 0$. Alors :*

$$\mathbb{P}(|X| \geq c) \leq \mathbb{E}(|X|^p)/c^p.$$

PREUVE

$$\mathbb{E}(|X|^p) \geq \int_{\{|X| \geq c\}} |X|^p d\mathbb{P} \geq c^p \mathbb{P}(|X| \geq c)$$

□

Proposition 4.7.8. *Soit X et Y deux v.a.r. intégrables et indépendantes. Alors :*

(i) XY est intégrable

(ii) $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$

La seconde affirmation est encore vraie si X et Y sont non–négatives et indépendantes.

PREUVE

a) Supposons tout d'abord $X \geq 0$, $Y \geq 0$. On utilise 4.7.2, 4.6.4 et Fubini :

$$\begin{aligned}\mathbb{E}(XY) &= \int_{\mathbb{R}^2} xy\, \mathbb{P}_{(X,Y)}(dx, dy) \\ &= \int\int xy\, \mathbb{P}_X(dx)\mathbb{P}_Y(dy) \\ &= \left(\int x\, \mathbb{P}_X(dx)\right)\left(\int y\, \mathbb{P}_Y(dy)\right) \\ &= \mathbb{E}(X)\mathbb{E}(Y)\end{aligned}$$

b) Cas général : $|X|$ et $|Y|$ sont indépendantes. Donc (i) résulte de a). (ii) se démontre alors comme en a).

□

Corollaire 4.7.9. *Soit $X_1, \ldots, X_d$ une suite de v. a. r. On a les implications suivantes*

La suite de v. a. $X_1, \ldots, X_d$ est indépendante

$\Downarrow$

Les v. a. $X_1, \ldots, X_d$ sont 2 à 2 indépendantes

$\Downarrow$

La matrice $Cov(X)$ est diagonale.

Lorsque $Cov(X_i, X_j) = 0$, pour tout $1 \leq i \neq j \leq n$, on dit que les v. a. $X_1, \ldots, X_d$ sont non corrélées.

Notons $\mathbf{1}$ le vecteur de $\mathbb{R}^d$ dont toutes les coordonnées sont égales à 1. On a $\mathrm{Var}(X_1 + \cdots + X_d) = \mathrm{Var}\,(\langle \mathbf{1}, X\rangle) = \mathbf{1}'\mathrm{Cov}\,(X)\mathbf{1}$, dont on déduit la

Proposition 4.7.10. *Soit $X_1, \ldots, X_n$ des v. a. r. de carré intégrable.*

1. *Alors*

$$Var\,(X_1 + \cdots + X_n) = \sum_{k=1}^{n} Var\,(X_k) + \sum_{1 \le k \ne \ell \le n} Cov\,(X_k, X_\ell).$$

2. *Si les v. a. $X_1, \ldots X_n$ sont deux à deux indépendantes, alors*

$$Var\,(X_1 + \cdots + X_n) = Var\,(X_1) + \cdots + Var\,(X_n).$$

On a la

Proposition 4.7.11. *Soit X_k un v. a. de dimension d_k, $1 \le k \le n$. La suite $X_1, \ldots, X_n$ est indépendante si et seulement si pour toutes les fonctions boréliennes $f_k : \mathbb{R}^{d_k} \to \mathbb{R}_+$,*

$$\mathbb{E}\left[\prod_{k=1}^{n} f_k(X_k)\right] = \prod_{k=1}^{n} \mathbb{E}[f_k(X_k)].$$

Preuve La CN s'obtient en itérant la Proposition 4.7.8. La CS s'obtient en choisissant $f_k = \mathbf{1}_{B_k}$, B_k arbitraire dans la tribu borélienne $\mathcal{B}_{d_k}$. $\square$

Notons que le résultat ci–dessus serait vrai en remplaçant les fonctions boréliennes positives par les fonctions continues bornées par exemple. On verra une autre variante de ce réultat au Théorème 5.3.5.

4.8 Exercices

Exercice 4.8.1. *Soit $\mathbb{P}$ une application de $\mathcal{F}$ à valeurs dans $[0, 1]$, qui vérifie*

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

(ii) Pour tout $n \ge 2$, $A_1, \ldots, A_n \in \mathcal{F}$ deux à deux disjoints, $\mathbb{P}(\cup_{k=1}^n A_k) = \sum_{k=1}^n \mathbb{P}(A_k)$.

(iii) Si $A_n \in \mathcal{F}$, $n \ge 1$ et $A_n \downarrow \emptyset$, $\mathbb{P}(A_n) \downarrow 0$.

Montrer que $\mathbb{P}$ est une mesure de probabilité.

Indication : On considère $B = \cup_n B_n$, où les B_n sont disjoints, on pose $C_n = \cup_{k>n} B_k$, et on applique (ii) à $B = B_1 \cup \cdots \cup B_n \cup C_n$, (iii) à $\mathbb{P}(C_n)$.

Exercice 4.8.2. *Soit A et $B \in \mathcal{F}$ deux événements indépendants.*

(i) Montrer que A^c et B^c [resp. A^c et B, A et B^c] sont indépendants.

(ii) Etablir la relation : $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A)\mathbb{P}(B)$.

Exercice 4.8.3. *On dit qu'une probabilité $\mathbb{Q}$ sur $(\mathbb{R}^d, \mathcal{B}_d)$ est diffuse si elle n'a pas d'atome, c'est à dire si $\mathbb{Q}(\{x\}) = 0$, pour tout $x \in \mathbb{R}^d$. Soit X et Y deux vecteurs aléatoires indépendants de dimension d. On suppose que la loi $\mathbb{P}_X$ de X est diffuse. Montrer que $\mathbb{P}(X = Y) = 0$.*

Exercice 4.8.4. *Soit X une v. a. à valeurs dans $\mathbb{N}$ t. q. $\mathbb{P}(X > k) > 0$, pour tout $k \in \mathbb{N}$. Alors il existe $p \in (0,1)$ tel que la loi de X est la loi $G(p)$ ssi pour tout $k, \ell \in \mathbb{N}$, $\mathbb{P}(X > k + \ell | X > k) = \mathbb{P}(X > \ell)$.*

Exercice 4.8.5. *Montrer qu'une v. a. suit la loi binômiale négative de paramètres m et p ssi elle est la somme de m géométriques indépendantes de paramètre p.*

Exercice 4.8.6. *Montrer que la loi $\mathcal{P}(\lambda)$ est la limite des lois $B(n, p_n)$ lorsque $n \to \infty$, à condition que $np_n \to \lambda > 0$ quand $n \to \infty$.*

Exercice 4.8.7. *Soit $\{X_n, n \geq 1\}$ une suite de variables i. i. d. à valeurs dans $\mathbb{N}$, de loi $(p(k), k \in \mathbb{N})$, telle que*

$$p(k) > 0, \ \forall k \in \mathbb{N}, \qquad \sum_{k \in \mathbb{N}} kp(k) < \infty.$$

1. *Montrer que pour tout $a \in \mathbb{N}$, $\mathbb{P}\left(\limsup_n \{X_n \geq a\}\right) = 1$.*
2. *Montrer que $\mathbb{P}\left(\limsup_n \{X_n \geq n\}\right) = 0$.*

Exercice 4.8.8. *Soit $(X_n)_{n \in \mathbb{N}}$ une suite de v.a.i.i.d de loi $\mathcal{E}(1)$. Montrer que*

$$\limsup_{n \to \infty} \frac{X_n}{\log(n)} = 1 \quad p. \ s.$$

Exercice 4.8.9. *Soit $(U_n)_{n \in \mathbb{N}}$ une suite de v.a.r.i.i.d. de loi $B(1/2)$. Montrer que V définie par*

$$V = \sum_{n \geq 1} \frac{U_n}{2^n}$$

est une v.a. de loi $\mathcal{U}([0,1])$.

Exercice 4.8.10. *Soit U une v.a.r de loi $\mathcal{U}([0,1])$. Le but de l'exercice est de construire à partir de U une v.a.r X de fonction de répartition F donnée.*

1. *On suppose que F est continue et strictement croissante. Montrer que $X = F^{-1}(U)$ a pour fonction de répartition F.*
2. *On suppose que F est une fonction de répartition quelconque. Définir une application G sur $[0;1]$ telle que*

$$G(t) \le x \Leftrightarrow t \le F(x)\,.$$

On notera $G = F^{-1}$ (inverse généralisé de F). Montrer que $X = F^{-1}(U)$ a pour fonction de répartition F.

Exercice 4.8.11. *Soit X une v. a. à valeurs dans $\mathbb{R}_+$, telle que $\mathbb{P}(X > x) > 0$, pour tout $x > 0$. Alors il existe $\lambda > 0$ t. q. $X \simeq \mathcal{E}(\lambda)$ ssi*

$$\textit{Pour tout } x, y > 0, \quad \mathbb{P}(X > x + y | X > x) = \mathbb{P}(X > y).$$

Exercice 4.8.12. *Soit U une v. a. r. de loi $\mathcal{U}([0,1])$, et $\lambda > 0$. Quelle est la loi de la v. a. r. $-\log(U)/\lambda$?*

Exercice 4.8.13. *Soit U et V deux v. a. r. indépendantes, toutes les deux de loi $\mathcal{U}([0,1])$. Montrer que les deux v. a.*

$$\sqrt{-2\log(U)}\cos(2\pi V) \quad \textit{et} \ \sqrt{-2\log(U)}\sin(2\pi V)$$

sont indépendantes, toutes deux de loi $N(0,1)$.

Exercice 4.8.14. *Soit f et g deux densités de probabilité sur $\mathbb{R}^d$. On suppose que pour tout $x \in \mathbb{R}^d$,*

$$f(x) \le kg(x), \quad \textit{avec un certain } k > 0.$$

Soit $\{(X_n, U_n),\ n \ge 1\}$ une suite i. i. d., telle que X_n et U_n sont indépendants, X_n est un v. a. de dimension d de densité g, U_n est une v. a. r. de loi $\mathcal{U}([0,1])$. Soit

$$N = \inf\{j \ge 1,\ kU_j g(X_j) \le f(X_j)\} \quad \textit{et } X = X_N.$$

Montrer que la loi de X est la loi sur $\mathbb{R}^d$ de densité f.

Exercice 4.8.15. *A tout réel positif x, on associe le nombre entier $\lceil x \rceil := \inf\{k$ entier t. q. $x \le k\}$. Soit $X \simeq \mathcal{E}(\lambda)$ une v. a. à valeurs dans $\mathbb{R}_+$ (on suppose $\lambda > 0$). Quelle est la loi de la v. a. $Y = \lceil X \rceil$?*

Exercice 4.8.16. *Soit X et Y deux v. a. indépendantes, $X \simeq \mathcal{E}(\lambda)$ et $Y \simeq \mathcal{E}(\mu)$ $(\lambda, \mu > 0)$. Quelle est la loi de la v. a. $Z = \inf(X, Y)$? Calculer les probabilités des événements $\{X < Y\}$, $\{X = Y\}$, $\{X > Y\}$.*

Exercice 4.8.17. *Soit X et Y des v.a. indépendantes, $X \simeq \Gamma(r, \lambda)$, $Y \simeq \Gamma(s, \lambda)$.*

1. *Quelle est la loi de la v.a. $X + Y$?*
2. *Montrer que les deux v.a. $X + Y$ et $(X + Y)^{-1}X$ sont indépendantes et que la loi de $(X + Y)^{-1}X$ est une loi bêta.*
3. *Etablir la formule du cours pour la constante de normalisation de la loi bêta.*

Exercice 4.8.18. *Si $Y \simeq N(0, 1)$, $X = Y^2 \simeq \Gamma(1/2, 1/2)$.*

Exercice 4.8.19. *Soit $X_1, \ldots, X_n$ des v. a. r. i. i. d., toutes de loi $\mathcal{U}([0, 1])$. Préciser la loi de $\min(X_1, \ldots, X_n)$ et celle de $\max(X_1, \ldots, X_n)$.*

Exercice 4.8.20. *Calculer l'espérance et la variance de toutes les lois de probabilité présentées à la section 4.5.*

Exercice 4.8.21. *Soit X une v.a.r. à valeurs positives. Montrer que*

1. *Si X est à valeurs dans $\mathbb{N}$, $\mathbb{E}[X] = \sum_{i=0}^{\infty} \mathbb{P}(X > i) = \sum_{i=1}^{\infty} \mathbb{P}(X \geq i)$.*
2. *Si X est à valeurs dans $\mathbb{R}_+$, $\mathbb{E}[X] = \int_0^{\infty} \mathbb{P}(X > x)dx = \int_0^{\infty} \mathbb{P}(X \geq x)dx$.*

Exercice 4.8.22. *Montrer par deux contre–exemples que les réciproques des deux implications du Corollaire 4.7.9 sont fausses en général.*

Exercice 4.8.23. *Soit X un vecteur aléatoire de dimension d, dont toutes les coordonnées sont de carré intégrable, et*

1. *Soit $u \in \mathbb{R}^d$. Montrer l'identité*

$$Var\left(\sum_{k=1}^{d} u_k X_k\right) = \sum_{k=1}^{d} \sum_{\ell=1}^{d} Cov\,(X_k, X_\ell) u_k u_\ell = u' Cov\,(X) u.$$

2. *Soit A une matrice $k \times d$. On pose $Y = AX$. Montrer que $Cov(Y) = ACov(X)A'$ (on pourra par exemple déduire ce résultat d'une double application du résultat précédent, qui en est un cas particulier).*

Exercice 4.8.24. *(Formules de Wald) Soit N une v.a. à valeurs dans $\mathbb{N}$, et les X_n des v.a. i.i.d. à valeurs dans $\mathbb{R}$. On pose*

$$U = \sum_{n=1}^{N} X_n = \sum_{n=1}^{\infty} \mathbf{1}_{\{n \leq N\}} X_n.$$

On suppose en outre que les v.a. $N, X_1, X_2, \ldots$ sont indépendantes.

1. *Montrer que si N et X_1 sont intégrables, alors U l'est aussi et en outre $\mathbb{E}[U] = \mathbb{E}[N]\mathbb{E}[X_1]$.*
2. *On suppose maintenant N et X_1 de carré intégrable. Montrer que U est de carré intégrable et que $Var[U] = \mathbb{E}[N]\,Var[X_1] + Var[N](\mathbb{E}[X_1])^2$.*

Exercice 4.8.25. *Soit $T_1, \ldots, T_k$ des v. a. r. indépendantes à valeurs dans $\mathbb{R}_+$, la loi de chacune des T_i étant à densité strictement positive sur $\mathbb{R}_+$ (donc en particulier telles que $\mathbb{P}(T_i > t) > 0$, pour tout $t > 0$). On définit les v. a. $\Lambda = \inf_{1 \leq i \leq k} T_i$, et ξ à valeurs dans $\{1, \ldots, k\}$ par*

$$\xi = \inf\{1 \leq j \leq k;\ T_j = \Lambda\}.$$

On suppose que T_i suit la loi exponentielle de paramètre $\lambda_i > 0$, $1 \leq i \leq k$. On a donc que pour tout $t \geq 0$, $\mathbb{P}(T_i > t) = \exp(-\lambda_i t)$, $1 \leq i \leq k$.

1. *Montrer que que $\mathbb{P}(\{\xi = i\} \backslash \{T_i < \min_{j \neq i} T_j\}) = 0$.*
2. *Calculer $\mathbb{P}(\Lambda > t)$. En déduire que la v. a. Λ suit une loi exponentielle de paramètre $\sum_{1 \leq i \leq k} \lambda_i$.*
3. *Montrer que T_i et $\inf_{j \neq i} T_j$ sont deux v. a. indépendantes, toutes les deux de loi exponentielle dont on précisera les paramètres.*
4. *Calculer $\mathbb{P}(T_i < \inf_{j \neq i} T_j)$ et $\mathbb{P}(t < T_i < \inf_{j \neq i} T_j)$.*
5. *Préciser la loi de ξ (on calculera $\mathbb{P}(\xi = i)$, $1 \leq i \leq k$), et montrer que les deux v. a. Λ et ξ sont indépendantes (on montrera – pourquoi est–ce suffisant ? – que pour tout $1 \leq i \leq k$, $t > 0$, les événements $\{\xi = i\}$ et $\{\Lambda > t\}$ sont indépendants).*

Chapitre 5

Fonction caractéristique et loi de Gauss

5.1 Préliminaires

Dans ce chapitre nous allons manipuler des intégrales de fonctions à valeurs complexes. On peut toujours écrire une telle fonction comme sa partie réelle plus i fois sa partie imaginaire. Une telle fonction est alors mesurable et intégrable si sa partie réelle et sa partie imaginaire le sont toutes les deux, et dans ce cas son intégrale est égale à celle de sa partie réelle plus i fois celle de sa partie imaginaire. Puisque l'on peut approcher partie réelle et partie imaginaire par des fonctions étagées, on peut aussi approcher toute fonction mesurable à valeurs complexes par une suite de fonctions étagées. L'inégalité suivante entre modules

$$\left| \int_\Omega f(\omega)\mu(d\omega) \right| \leq \int_\Omega |f(\omega)|\mu(d\omega)$$

est immédiate si f est étagée. Elle s'étend assez facilement à toutes les fonctions f $\mathcal{F}$–mesurables et intégrables à valeurs dans $\mathbb{C}$.

Introduisons maintenant une notion qui va nous servir dans ce chapitre et le chapitre 7. Soit $\mathcal{K}$ une classe de fonctions de $C_b(\mathbb{R}^d; \mathbb{C})$.

Définition 5.1.1. *On dit que la classe $\mathcal{K}$ est séparante si pour toutes $\mathbb{Q}$ et $\mathbb{Q}'$ probabilités sur $(\mathbb{R}^d, \mathcal{B}_d)$,*

$$\left\{ \int f(x)\mathbb{Q}(dx) = \int f(x)\mathbb{Q}'(dx), \text{ pour toute } f \in \mathcal{K} \right\} \Rightarrow \mathbb{Q} = \mathbb{Q}'.$$

Remarquons tout d'abord que

Lemme 5.1.2. *Soit $\mathbb{Q}$ et $\mathbb{Q}'$ deux probabilités sur $(\mathbb{R}^d, \mathcal{B}_d)$, telles que $\mathbb{Q}(F) = \mathbb{Q}'(F)$, pour tout F fermé borné de $\mathbb{R}^d$. Alors $\mathbb{Q} = \mathbb{Q}'$.*

PREUVE La classe des fermés bornés est stable par intersection, et engendre $\mathcal{B}_d$. □

Lemme 5.1.3. *L'ensemble $C_K(\mathbb{R}^d; \mathbb{R}_+)$ des fonctions continues positives à support compact est séparant.*

PREUVE Soit $\mathbb{Q}$ et $\mathbb{Q}'$ deux probabilités sur $(\mathbb{R}^d, \mathcal{B}_d)$ telles que

$$\int f d\mathbb{Q} = \int f d\mathbb{Q}', \quad \text{pour toute } f \text{ continue positive à support compact.}$$

Soit en outre F un fermé borné $\subset \mathbb{R}^d$. Posons $f_\varepsilon(x) = \varphi\left(\frac{1}{\varepsilon}\rho(x, F)\right)$, avec $\varphi(x) = \mathbf{1}_{\{x \leq 0\}} + (1-x)^+ \mathbf{1}_{\{x \geq 0\}}$ et $\rho(x, F) = \inf_{y \in F} \|x - y\|$. f_ε est continue et positive. Comme en outre F est borné, f_ε est à support compact. Donc pour tout $\varepsilon > 0$,

$$\int f_\varepsilon(x) \mathbb{Q}(dx) = \int f_\varepsilon(x) \mathbb{Q}'(dx).$$

Comme en outre f_ε est majorée par 1 et décroit quand $\varepsilon \downarrow 0$ vers l'indicatrice de F,

$$\mathbb{Q}(F) = \mathbb{Q}'(F), \quad \text{pour tout } F \text{ fermé borné,}$$

d'où le résultat grâce au Lemme précédent. □

5.2 Fonction caractéristique sur $\mathbb{R}$

On va avoir besoin du

Théorème 5.2.1. *La classe des fonctions $\{e^{iux}, u \in \mathbb{R}\}$ est séparante.*

Démontrons tout d'abord la

Proposition 5.2.2. *Soit $p \geq 1$ et f une fonction continue de $[-p, +p]$ à valeurs dans $\mathbb{R}$, telle que $f(-p) = f(p)$. Alors il existe une suite f_m de la forme :*

$$f_m(x) = \sum_{-m \leq k \leq m} a_k^m e^{i\pi kx/p} \quad (a_k^m \in \mathbb{C}).$$

telle que $f_m(x) \to f(x)$, uniformément sur $[-p, p]$.

PREUVE On va montrer le résultat pour f à valeurs dans $\mathbb{C}$. Soit g la bijection bicontinue de $(-p, +p]$ sur le cercle unité E du plan complexe définie par :

$$g(x) = e^{i\pi x/p}.$$

La fonction f de l'énoncé peut se factoriser sous la forme : $f = h \circ g$, avec $h = f \circ g^{-1} \in C(E; \mathbb{C})$. Soit $A \subset C(E; \mathbb{C})$ l'algèbre engendrée par les fonctions 1, z et $\bar{z}$. A sépare les points de E. Rappelons le théorème de Stone–Weierstrass :

> Toute algèbre A de fonctions de $C(E; \mathbb{C})$ (E compact), contenant les constantes, qui sépare les points de E et qui vérifie $f \in A \Rightarrow \bar{f} \in A$, est dense dans $C(E; \mathbb{C})$ [cet espace étant muni de la topologie de la convergence uniforme]

Soit donc h_m une suite dans A, t.q. $h_m(z) \to h(z)$ uniformément. Alors si $f_m = h_m \circ g$, la suite f_m a les propriétés de l'énoncé. □

PREUVE DU THÉORÈME 5.2.1 : Soit $\mathbb{Q}$ et $\mathbb{Q}'$ deux probabilités sur $(\mathbb{R}, \mathcal{B})$ t.q.

$$\int e^{iux} \mathbb{Q}(dx) = \int e^{iux} \mathbb{Q}'(dx), \quad \forall\, u \in \mathbb{R}.$$

Alors pour tout $\ell \geq 1\ \forall\, \alpha_1, \ldots, \alpha_\ell \in \mathbb{C}$, $u_1, \ldots, u_\ell \in \mathbb{R}$,

$$(*) \qquad \int_{\mathbb{R}} \left(\sum_1^\ell \alpha_k e^{iu_k x} \right) \mathbb{Q}(dx) = \int_{\mathbb{R}} \left(\sum_1^\ell \alpha_k e^{iu_k x} \right) \mathbb{Q}'(dx).$$

On va maintenant déduire de $(*)$ et de la Proposition 5.2.2 que

$$\int f d\mathbb{Q} = \int f d\mathbb{Q}', \ \forall\, f \in C_K(\mathbb{R}), \text{ d'où } \mathbb{Q} = \mathbb{Q}' \text{ d'après le Lemme 5.1.3.}$$

Fixons donc $f \in C_K(\mathbb{R})$. On va construire une suite f_n de la forme $f_n = h_n \circ g, h_n \in A$, t.q.

$(**)$ $\qquad \int f_n d\mathbb{Q} \to \int f d\mathbb{Q}$ et $\int f_n d\mathbb{Q}' \to \int f d\mathbb{Q}'$.

Il existe n_0 t.q. pour tout $n \geq n_0$, $f(-n) = f(n) = 0$. Fixons $n \geq n_0$.

Il existe $N \geq 1$ et $\alpha_k \in \mathbb{C}$, $k = -N, \ldots, 0, 1, \ldots, N$ t.q. si

$$f_n(x) = \sum_{-N}^{N} \alpha_k e^{i\pi kx/n}$$

$$|f(x) - f_n(x)| \leq \frac{1}{n}, \ \forall\, x \in [-n, n]$$

ce qui entraîne $|f_n(x)| \leq \sup_x |f(x)| + 1$ car f_n est périodique de période $2n$. $(**)$ résulte des deux dernières inégalités, par le Théorème de convergence dominée. □

Définition 5.2.3. — *Etant donné une probabilité $\mathbb{Q}$ sur $(\mathbb{R}, \mathcal{B})$, on appelle fonction caractéristique de $\mathbb{Q}$ la fonction $\varphi_{\mathbb{Q}}$ définie sur $\mathbb{R}$ à valeurs dans $\mathbb{C}$ par :*

$$\varphi_{\mathbb{Q}}(u) = \int_{\mathbb{R}} e^{iux} \mathbb{Q}(dx)$$

— *Etant donné une v.a.r. X, on définit sa fonction caractéristique φ_X comme étant la f.c. de la mesure $\mathbb{P}_X$, i.e.*

$$\varphi_X(u) = \int_{\mathbb{R}} e^{iux} \mathbb{P}_X(dx) = \mathbb{E}[e^{iuX}]$$

Remarquons que la fonction caractéristique de la mesure $\mathbb{Q}$ est sa transformée de Fourier ; d'après le Théorème 5.2.1 elle mérite bien son nom !.

Proposition 5.2.4. *Si $\varphi(u)$ est la f.c. d'une mesure de probabilité $\mathbb{Q}$, alors :*

(i) $\varphi(0) = 1$

(ii) $|\varphi(u)| \leq 1$, *pour tout* $u \in \mathbb{R}$

(iii) φ *est uniformément continue sur* $\mathbb{R}$

(iv) $\varphi(-u) = \bar{\varphi}(u)$.

Preuve (i) est élémentaire.

(ii) $\left| \int e^{iux} \mathbb{Q}(dx) \right| \leq \int \left| e^{iux} \right| \mathbb{Q}(dx) = 1$

(iii)

$$\begin{aligned} |\varphi(u+h) - \varphi(u)| &= \left| \int e^{iux}(e^{ihx} - 1) \mathbb{Q}(dx) \right| \\ &\leq \int \left| e^{ihx} - 1 \right| \mathbb{Q}(dx) = \delta(h), \end{aligned}$$

où $\delta(h) \to 0$ quand $h \to 0$, par le Théorème de convergence dominée.

(iv) est élémentaire. □

Théorème 5.2.5. *(Transformée de Fourier inverse) Soit $\varphi(u)$ la f.c. d'une probabilité $\mathbb{Q}$ sur $(\mathbb{R}, \mathcal{B})$, t.q. $\int_{\mathbb{R}} |\varphi(u)| du < \infty$.*

Alors $\mathbb{Q}$ admet une densité $f(x)$ uniformément continue et bornée, donnée par :

$$(*) \qquad f(x) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-iux} \varphi(u) du.$$

On va utiliser le Lemme (qui par ailleurs est important en lui-même) :

Lemme 5.2.6. *Soit X et Y deux v.a.r. indépendantes X de loi $\mathbb{Q}$, et Y admettant une densité $g(x)$.*

Alors la loi de la v.a.r. $X+Y$ admet la densité $h(x) = \int_{\mathbb{R}} g(x-y)\mathbb{Q}(dy)$.

Preuve Posons $Z = X + Y$. D'après le Lemme 5.1.3, il suffit de montrer que pour toute $f \in C_K(\mathbb{R}; \mathbb{R}_+)$,

$$\mathbb{E}[f(Z)] = \int_{\mathbb{R}} f(z) \left(\int_{\mathbb{R}} g(z-x)\mathbb{Q}(dx) \right) dz$$

On va utiliser à deux reprises le Théorème de Fubini (que l'on peut utiliser puisque l'on intègre que des fonctions positives).

$$\begin{aligned}
\mathbb{E}[f(Z)] &= \mathbb{E}[f(X+Y)] \\
&= \int \int f(x+y)\mathbb{Q}(dx) g(y) dy \\
&= \int \left[\int f(x+y) g(y) dy \right] \mathbb{Q}(dx) \\
&= \int \left[\int f(z) g(z-x) dz \right] \mathbb{Q}(dx) \\
&= \int f(z) \left[\int g(z-x)\mathbb{Q}(dx) \right] dz.
\end{aligned}$$

□

Remarque 5.2.7. *Ce Lemme est encore vrai pour des v.a. X et Y de dimension $d > 1$. On en déduit que si X et Y sont deux vecteurs aléatoires de dimension d indépendants tels que la loi de X admet une densité, alors la* *loi de $X - Y$ admet une densité, donc en particulier $\mathbb{P}(X = Y) = 0$. Mais cette dernière propriété est vraie sous une hypothèse plus faible, cf. l'Exercice 4.8.3*

Preuve du Théorème 5.2.5. On va admettre pour l'instant – on le vérifiera directement ci–dessous à la section 3 – qu'il existe au moins une probabilité sur $(\mathbb{R}, \mathcal{B})$, de densité $g(x)$, et de f.c. intégrable $\psi(u)$, qui vérifient $(*)$.

Par définition de φ, on a, en utilisant successivement le Théorème de Fubini, et l'hypothèse que nous venons de faire

$$\begin{aligned}\frac{1}{2\pi}\int_{\mathbb{R}} \varphi(u)\psi(u)e^{-iuy}du &= \frac{1}{2\pi}\int_{\mathbb{R}}\int_{\mathbb{R}} e^{iux}e^{-iuy}\psi(u)\mathbb{Q}(dx)du \\ &= \int_{\mathbb{R}}[\frac{1}{2\pi}\int_{\mathbb{R}} e^{-iu(y-x)}\psi(u)du]\mathbb{Q}(dx) \\ &= \int_{\mathbb{R}} g(y-x)\mathbb{Q}(dx).\end{aligned}$$

Cette fonction de y est, d'après le Lemme 5.2.6, la densité de la loi de $X + Y$, si X et Y sont indépendants X de loi $\mathbb{Q}$ et Y de densité g.

Soit $\varepsilon > 0$. La densité de la loi de εY est $\frac{1}{\varepsilon}g\left(\frac{x}{\varepsilon}\right)$; sa f.c. est $\varphi_{\varepsilon Y}(u) = \psi(\varepsilon u)$. εY vérifie encore $(*)$ puisque

$$\frac{1}{\varepsilon}g\left(\frac{x}{\varepsilon}\right) = \frac{1}{2\pi}\int_{\mathbb{R}} e^{-i\frac{u}{\varepsilon}x}\psi(u)\frac{du}{\varepsilon} = \frac{1}{2\pi}\int_{\mathbb{R}} e^{-ivx}\psi(\varepsilon v)dv.$$

Donc il résulte du calcul ci–dessus que

$$(**) \quad \rho_\varepsilon(y) = \frac{1}{2\pi}\int_{\mathbb{R}} \varphi(u)\psi(\varepsilon u)e^{-iuy}du$$

est la densité de la loi de $X + \varepsilon Y$.

Soit $F \in C_K(\mathbb{R})$. Intégrant $F(y)$ par rapport aux deux membres de $(**)$, on obtient :

$$(***) \quad \mathbb{E}[F(X + \varepsilon Y)] = \frac{1}{2\pi}\int\int F(y)\varphi(u)\psi(\varepsilon u)e^{-iuy}\,du\,dy.$$

Quand $\varepsilon \to 0$

$$X + \varepsilon Y \to X \text{ p.s.}, \quad \psi(\varepsilon u) \to 1, \forall\, u \in \mathbb{R}.$$

Les hypothèses faites sur φ et F entraînent que $|F(y)| \times |\varphi(u)|$ est $dy \times du$ intégrable sur $\mathbb{R}^2$. On peut donc faire tendre $\varepsilon \to 0$ dans $(***)$ à l'aide du Théorème de convergence dominée, d'où

$$\mathbb{E}[F(X)] = \frac{1}{2\pi}\int\int F(y)\varphi(u)e^{-iuy}dudy.$$

Ceci entraîne que f donnée par $(*)$ est la densité bornée de la loi $\mathbb{Q}$. En outre

$$\sup_{x\in\mathbb{R}} |f(x+h) - f(x)| \leq \frac{1}{2\pi}\int |\varphi(u)| \cdot |e^{-iuh} - 1|du,$$

qui tend vers 0 quand $h \to 0$, par le théorème de convergence dominée. Donc f est uniformément continue. □

Théorème 5.2.8. *Si $\mathbb{E}(|X|^n) < \infty$, alors*

$$\varphi_X(u) = \sum_{k=0}^{n} \frac{(iu)^k}{k!}\mathbb{E}(X^k) + \frac{(iu)^n}{n!}\delta(u)$$

avec $|\delta(u)| \leq 2\mathbb{E}\,(|X|^n)$, pour tout $u \in \mathbb{R}$; et $\delta(u) \to 0$, quand $u \to 0$.

Preuve Si $f \in C^n(\mathbb{R}; \mathbb{C})$, on a :

$$\begin{aligned}
f(y) &= f(0) + \int_0^y f'(u)du \\
&= f(0) + yf'(0) + \int_0^y \int_0^{u_1} f''(u_2)du_2du_1 \\
&= \sum_{k=0}^{n} \frac{y^k}{k!} f^{(k)}(0) + \int_0^y \int_0^{u_1} \cdots \int_0^{u_{n-1}} (f^{(n)}(u_n) - f^{(n)}(0))du_n \cdots du_1,
\end{aligned}$$

d'où :

$$(*) \qquad e^{iuX} = \sum_0^n \frac{(iuX)^k}{k!} + \frac{(iu)^n}{n!}R(u,X), \text{ avec}$$

$$R(u,X) = n!\int_0^X \int_0^{v_1} \cdots \int_0^{v_{n-1}} (e^{iuv_n} - 1)dv_n \cdots dv_1, \text{ d'où } |R(u,X)| \leq 2|X|^n$$

On peut donc prendre l'espérance dans $(*)$, d'où l'égalité de l'énoncé, avec

$$\delta(u) = \mathbb{E}[R(u,X)]$$
$$\text{donc } |\delta(u)| \leq 2\,\mathbb{E}[|X|^n],$$

et $\delta(u) \to 0$ quand $u \to 0$, par le Théorème de convergence dominée. □

Corollaire 5.2.9. *Soit X une v.a.r. t.q. $\mathbb{E}(|X|^n) < \infty$. Alors sa f.c. $\varphi_X(u)$ est n fois continûment dérivable, et*

$$\varphi_X^{(k)}(0) = i^k \mathbb{E}(X^k); \;\; k = 1, 2, \ldots, n$$

Preuve

$$\mathbb{E}\left[\frac{e^{i(u+h)X} - e^{iuX}}{h}\right] = \mathbb{E}\left[iXe^{i(u+\theta\, h)X}\right]$$

où θ est une v.a. à valeurs dans $]0,1[$. Mais :

$$iX\, e^{i(u+\theta\, h)X} \to i\, X\, e^{iuX} \text{ p.s., quand } h \to 0$$
$$|X\, e^{i(u+\theta\, h)X}| \leq |X|, \text{ qui est intégrable.}$$

Donc, par le Théorème de convergence dominée,

$$\varphi'_X(u) = \mathbb{E}[i\, X\, e^{iu} X]$$

en particulier $\varphi'_X(0) = i\mathbb{E}(X)$. Le résultat pour les dérivées suivantes s'établit en itérant l'argument ci–dessus. □

5.3 Fonction caractéristique sur $\mathbb{R}^d$

Indiquons les grandes lignes de la preuve du

Théorème 5.3.1. *La classe des fonctions.*

$$\left\{ \exp\left(i \sum_1^d u_k x_k \right), (u_1, \ldots, u_d) \in \mathbb{R}^d \right\} \;\; \textit{est séparante.}$$

Preuve Comme au Théorème 5.2.1, il suffit d'approcher toute fonction $f \in C_K(\mathbb{R}^d)$ par des combinaisons linéaires de fonctions de cette classe. Si $f \in C_K(\mathbb{R}^d)$, il existe n_0 t.q. pour tout $n \geq 0$,

$$\sup(f) \subset [-n, n]^d = R_n$$

Grâce au Théorème de Stone–Weierstrass (la Proposition 5.2.2 se généralise en remplaçant le cercle par un tore de dimension d), il existe f_n, combinaison linéaire de fonctions de la forme

$$\exp[i\, \pi(m_1 x_1 + \cdots + m_d x_d) n]$$

(où les m_k sont des entiers), qui approche f à $\frac{1}{n}$ près sur R_n. La démonstration se termine comme pour le Théorème 5.2.1. □

Définition 5.3.2. *Si $\mathbb{Q}$ est une probabilité sur $\mathbb{R}^d$, on définit sa fonction caractéristique comme la fonction de $\mathbb{R}^d$ dans $\mathbb{C}$:*

$$\varphi_{\mathbb{Q}}(u) = \int_{\mathbb{R}^d} e^{i(u,x)}\mathbb{Q}(dx).$$

Si X est un vecteur aléatoire de dimension d, on définit sa f.c. φ_X comme la f.c. de la probabilité $\mathbb{P}_X$: $\varphi_X(u) = \int e^{i(u,x)}\mathbb{P}_X(dx) = \mathbb{E}[e^{i(u,X)}]$

□

Proposition 5.3.3. *Soit X un v.a. de dimension d. Si $\mathbb{E}(|X|^n) < \infty$, alors*

$$\varphi_X(u) = \sum_{k=0}^{n} \frac{i^k}{k!}\mathbb{E}\left[(u,X)^k\right] + \delta(u)\frac{|u|^n}{n!},$$
$$\text{avec } |\delta(u)| \leq 2\mathbb{E}(|X|^n) \quad \text{et } \delta(u) \to 0 \text{ quand } u \to 0.$$

Preuve La proposition résulte du Théorème 5.2.8, en remarquant que

$$\varphi_X(u) = \varphi_{(u,X)}(1) = \varphi_{(\frac{u}{|u|},X)}(|u|).$$

□

Corollaire 5.3.4. *Si X est un v.a. de dimension d, t.q. $E(|X|^2) < \infty$, alors $\varphi_X(u)$ admet des dérivées partielles jusqu'à l'ordre 2, et*

$$\nabla\varphi_X(0) = i\,\mathbb{E}(X)$$
$$\left(\frac{\partial^2\varphi_X}{\partial u_i \partial u_j}(0)\right) = -\mathbb{E}(XX').$$

Théorème 5.3.5. *Soit $X = (X_1, \ldots, X_d)'$ un v.a. de dimension d. La suite de v.a.r. $(X_1, \ldots, X_d)$ est indépendante si et seulement si pour tout $u \in \mathbb{R}^d$,*

$$\varphi_X(u) = \prod_{k=1}^{d} \varphi_{X_k}(u_k)$$

Ce Théorème se généralise comme suit (nous omettons la preuve, qui est essentiellement la même qu'au Théorème 5.3.5).

Théorème 5.3.6. *Soit $X_1, X_2, \ldots, X_d$ des vecteurs aléatoires de dimension respective $\ell_1, \ell_2, \ldots, \ell_d$, et $X = (X_1, X_2, \ldots, X_d)'$ le v.a. de dimension $\sum_{k=1}^d \ell_k$ obtenu en concaténant les X_k. La suite de v.a. $(X_1, \ldots, X_d)$ est indépendante si et seulement si pour tout $u = (u_1, \ldots, u_d) \in \mathbb{R}^{\sum_{k=1}^d \ell_k}$, avec u_k de dimension ℓ_k, $1 \le k \le d$,*

$$\varphi_X(u) = \prod_{k=1}^d \varphi_{X_k}(u_k).$$

On utilisera le résultat suivant (immédiat à partir de la Proposition 4.7.8).

Lemme 5.3.7. *Soit X et Y deux v.a.* **indépendants**, *de dimension respective d_1 et d_2, et f, g des applications mesurables de $\mathbb{R}^{d_1}$ et $\mathbb{R}^{d_2}$ respectivement, à valeurs dans $\mathbb{C}$. Si $\mathbb{E}|f(X)| < \infty$ et $\mathbb{E}|g(Y)| < \infty$, alors*

$$\begin{aligned} &\mathbb{E}|f(X)g(Y)| < \infty \text{ et} \\ &\mathbb{E}[f(X)g(Y)] = \mathbb{E}[f(X)]\mathbb{E}[g(Y)]. \end{aligned}$$

Preuve du Théorème 5.3.5

C.N. Supposons les X_k indépendantes. En utilisant $d-1$ fois le résultat du lemme, on obtient :

$$\begin{aligned} \varphi_X(u) &= \mathbb{E}\left(\prod_1^d e^{iu_k X_k}\right) \\ &= \prod_{k=1}^d \varphi_{X_k}(u_k) \end{aligned}$$

C.S. L'égalité de l'énoncé s'écrit :

$$\int_{\mathbb{R}^d} e^{i\sum u_k x_k} \mathbb{P}_X(dx) = \int_{\mathbb{R}^d} e^{i\sum u_k x_k} \mathbb{P}_{X_1}(dx_1) \times \cdots \times \mathbb{P}_{X_d}(dx_d)$$

Donc d'après le Théorème 5.3.1, $\mathbb{P}_X = \mathbb{P}_{X_1} \times \cdots \times \mathbb{P}_{X_d}$. Le résultat découle alors du Théorème 4.6.4.

□

Corollaire 5.3.8. *Si $X_1, \ldots, X_d$ sont des v.a.r. indépendantes, et si $S_d = X_1 + \cdots + X_d$, alors*

$$\varphi_{S_d}(u) = \prod_1^d \varphi_{X_k}(u), \quad \forall u \in \mathbb{R}$$

□

Bien sûr, la réciproque du Corollaire 5.3.8 est fausse.

5.4 Vecteurs aléatoires gaussiens

Si $\mu \in \mathbb{R}$ et $\sigma > 0$, on désigne par $N(\mu, \sigma^2)$ la loi de Gauss sur $\mathbb{R}$, de densité $(\sigma\sqrt{2\pi})^{-1}\exp[-(x-\mu)^2/2\sigma^2]$.

Une v.a.r. X est une v.a.r. gaussienne si soit il existe $\mu \in \mathbb{R}$, $\sigma > 0$, t.q. $X \simeq N(\mu, \sigma^2)$, soit la loi de X est la masse de Dirac en un point $\mu \in \mathbb{R}$ (cas $\sigma = 0$). Alors $\mu = \mathbb{E}(X)$, $\sigma^2 = \mathrm{Var}\,(X)$.

Définition 5.4.1. *Un vecteur aléatoire X de dimension d est un vecteur aléatoire gaussien si et seulement si pour tout $c = (c_1, \ldots, c_d) \in \mathbb{R}^d$, $c \cdot X \stackrel{\triangle}{=} \sum_1^d c_i X_i$ est une v.a.r. gaussienne.*

Il résulte de la définition que l'image par une application affine d'un v.a. gaussien est un v.a. gaussien.

Lemme 5.4.2. *La fonction caractéristique d'une v.a.r. X de loi $N(\mu, \sigma^2)$ est la fonction*

$$\varphi_X(u) = \exp\left[iu\mu - \frac{\sigma^2}{2}u^2\right]$$

Preuve Il suffit de traiter le cas $\mu = 0$, $\sigma = 1$. Il faut montrer : $\frac{1}{\sqrt{2\pi}}\int e^{-\frac{x^2}{2}+iux}dx = e^{-u^2/2}$, $u \in \mathbb{R}$. On va en fait montrer que pour tout $z \in \mathbb{C}$,

$$\frac{1}{\sqrt{2\pi}}\int e^{-\frac{x^2}{2}+zx}dx = e^{z^2/2}$$

Cette égalité est vraie pour tout $z \in \mathbb{R}$, puisque :

$$\frac{1}{\sqrt{2\pi}}\int e^{-\frac{x^2}{2}+zx}dx = e^{z^2/2}\frac{1}{\sqrt{2\pi}}\int e^{-\frac{1}{2}(x-z)^2}dx = e^{z^2/2}$$

Par la Proposition 2.3.10 (iii),

$$\int \sum_1^\infty |f_n(x)|dx < \infty \Rightarrow \int \sum_1^\infty f_n(x)dx = \sum_1^\infty \int f_n(x)dx, \quad \text{donc :}$$

$$\frac{1}{\sqrt{2\pi}}\int e^{-\frac{x^2}{2}+zx}dx = \sum_1^\infty \frac{z^n}{n!\sqrt{2\pi}}\int x^n e^{-\frac{x^2}{2}}dx,$$
$$\text{et } e^{z^2/2} = \sum_1^\infty \frac{z^{2n}}{2^n n!}$$

Ces deux fonctions coïncidant sur $\mathbb{R}$, les coefficients des séries coïncident. Donc les fonctions coïncident sur $\mathbb{C}$. □

Remarque 5.4.3. *Si l'on pose* $g(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$, $\psi(u) = e^{-u^2/2}$, *on a bien* $g(x) = \dfrac{1}{2\pi}\displaystyle\int e^{-u^2/2-iux}du$, *d'après ce qui vient d'être démontré. Ceci établit l'existence du couple* (g, ψ) *dont l'existence avait été admise dans la preuve du Théorème 5.2.5.*

□

Si maintenant X est un v.a. gaussien d'espérance μ et de matrice de covariance Σ alors pour tout $u \in \mathbb{R}^d$, $(u, X) \simeq N((u,\mu),(\Sigma u, u))$ et donc il résulte du Lemme :

$$\begin{aligned}\varphi_X(u) &= \mathbb{E}[e^{i(u,X)}]\\ &= \exp\left[i(u,\mu) - \frac{1}{2}(\Sigma u, u)\right]\end{aligned}$$

On remarque que l'espérance et la matrice de covariance suffisent à caractériser la loi d'un vecteur aléatoire gaussien. Soit X une v.a.r. de loi $\mathbb{N}(0,\sigma^2)$. Sa f.c. est $\varphi_X(u) = \exp\left(-\frac{\sigma^2u^2}{2}\right)$. En développant l'exponentiele, on obtient

$$e^{-\frac{\sigma^2u^2}{2}} = \sum_1^\infty (-1)^n \frac{(\sigma u)^{2n}}{2^n n!}$$

D'où d'après le Théorème 5.2.8,

$$\mathbb{E}(X^{2n}) = \frac{\sigma^{2n}(2n)!}{2^n n!}, \qquad \mathbb{E}(X^{2n+1}) = 0.$$

Théorème 5.4.4. *Soit* $X = (X_1, \ldots, X_d)$ *un vecteur aléatoire gaussien de dimension d. Alors la suite de v.a.r.* $X_1, \ldots, X_d$ *est indépendante si et seulement si les* X_k *sont* 2 *à* 2 *non corrélées, i.e.* $Cov(X)$ *est une matrice diagonale.*

PREUVE La non corrélation est toujours une conséquence de l'indépendance.

La réciproque n'est vraie que dans le cas gaussien. Si $\Sigma = Cov(X)$ est diagonale, alors avec $\mu = \mathbb{E}(X)$,

$$\begin{aligned}\varphi_X(u) &= \exp[i(u,\mu) - \frac{1}{2}(\Sigma u, u)] \\ &= \prod_1^d \exp[iu_k\mu_k - \frac{1}{2}\Sigma_{kk}u_k^2] \\ &= \prod_1^d \varphi_{X_k}(u_k).\end{aligned}$$

Le résultat découle alors du Théorème 5.3.5 □

Ce Théorème se généralise comme suit (nous omettons la preuve, qui est essentiellement la même qu'au Théorème 5.4.4, en utilisant le Théorème 5.3.6 au lieu du Théorème 5.3.5).

Théorème 5.4.5. *Soit $X_1, X_2, \ldots, X_d$ des vecteurs aléatoires de dimension respective $k_1, k_2, \ldots, k_d$, et $X = (X_1, X_2, \ldots, X_d)$ le v.a. de dimension $\sum_{i=1}^d k_i$ obtenu en concaténant les X_i. Si le vecteur aléatoire X est un va. gaussien, alors la suite de v.a. $(X_1, \ldots, X_d)$ est indépendante si et seulement si la matrice $Cov(X)$ est bloc–diagonale, c'est à dire qu'elle s'écrit*

$$Cov(X) = \begin{pmatrix} Cov(X_1) & 0 & \ldots & 0 & 0 \\ 0 & Cov(X_2) & 0 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & 0 & Cov(X_d) \end{pmatrix}.$$

Attention ! Si X et Y sont 2 v.a.r. gaussiennes, $\begin{pmatrix} X \\ Y \end{pmatrix}$ n'est pas forcément un vecteur aléatoire gaussien. Donc si X et Y sont 2 v.a.r. gaussiennes non corrélées (i.e. t.q. $\mathrm{cov}(X,Y) = 0$) X et Y ne sont pas forcément indépendantes !

Exemple 5.4.6. *Soit X et V indépendantes, $X \simeq N(0,1)$, $\mathbb{P}(V = -1) = \mathbb{P}(V = 1) = 1/2$. On pose $Y = V \times X$. Alors*

$$\varphi_Y(u) = \mathbb{E}[e^{iuVX}] = \frac{1}{2}\mathbb{E}[e^{iuX}] + \frac{1}{2}\mathbb{E}[e^{-iuX}] = e^{-u^2/2}.$$

Donc $Y \simeq N(0,1)$. *En outre* $\mathbb{E}(XY) = \mathbb{E}(X^2V) = \mathbb{E}(X^2)\mathbb{E}(V) = 0$. *Mais* X *et* Y *ne sont pas indépendantes ! Pour tout* $a > 0$,

$$\begin{aligned}
\mathbb{P}(|X| \leq a) &= \mathbb{P}\left(\{|X| \leq a\} \cap \{|Y| \leq a\}\right) \\
&\neq \mathbb{P}(|X| \leq a)\mathbb{P}(|Y| \leq a) \\
&= (\mathbb{P}(|X| \leq a))^2
\end{aligned}$$

On peut vérifier directement que (X,Y) *n'est pas un v. a. gaussien.* $\mathbb{P}(X+Y=0) = 1/2$, *donc* $X+Y$ *n'est pas une v. a. r. gausienne.*

On a par ailleurs la

Proposition 5.4.7. *Soit* $X_1, \dots, X_d$ *des v.a.r. gaussiennes indépendantes. Alors le vecteur aléatoire* $X = (X_1, \dots, X_d)'$ *est un v.a. gaussien.*

Preuve Supposons que chaque X_k est de loi $N(\mu_k, \sigma_k^2)$. Alors par l'indépendance (cf. Théorème 5.3.5)

$$\varphi_x(u_1, \dots, u_d) = \prod_{k=1}^{d} \varphi_{X_k}(u_k) = \exp\left(i \sum_k u_k \mu_k - \frac{1}{2} \sum_k \sigma_k^2 u_k^2\right),$$

qui est bien la fonction caractéristique de la loi $N(\mu, \Sigma)$, où

$$\mu = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_d \end{pmatrix}, \quad \begin{pmatrix} \sigma_1^2 & 0 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & \sigma_d^2 \end{pmatrix}.$$

□

Soit μ un vecteur de $\mathbb{R}^d$ et Σ une matrice $d \times d$ auto–adjointe et semi–définie positive. Si $Y = (Y_1, \dots, Y_d)$ est tel que la suite $Y_1, \dots, Y_d$ est i. i. d., de loi commune la loi $N(0,1)$, alors la loi de Y est la loi $N(0, I)$ sur $\mathbb{R}^d$, et celle de $X = \mu + \Sigma^{1/2}Y$ est la loi $N(\mu, \Sigma)$. Notons que ce n'est pas cette formule qu'il faut utiliser en général pour simuler un v. a. de loi $N(\mu, \Sigma)$. On utilisera plutôt la formule $X = \mu + \Gamma Y$, où Γ une matrice à d lignes (et un nombre de colonnes qui peut être différent) telle que $\Gamma\Gamma' = \Sigma$. Il est en particulier assez facile de calculer les coefficients de la matrice Γ en fonction de ceux de Σ si l'on impose que Γ est une matrice $d \times d$ triangulaire inférieure. C'est la factorisation de Cholesky. La loi d'un v.a. gaussien de dimension d

n'admet pas forcément une densité par rapport à la mesure de Lebesgue sur $\mathbb{R}^d$. C'est le cas si et seulement si sa matrice de covariance est inversible, et la formule pour la densité est alors donnée par la

Proposition 5.4.8. *Soit X un v.a. de dimension d, de loi $N(\mu, \Sigma)$ [i.e. gaussien avec $\mathbb{E}(X) = \mu$, $cov(X) = \Sigma$].Si rang $(\Sigma) = d$, alors la loi de X admet la densité :*

$$f_X(x) = \frac{1}{\sqrt{(2\pi)^d dét\ (\Sigma)}} \exp\left[-\frac{1}{2}(\Sigma^{-1}(x-\mu), x-\mu)\right]$$

PREUVE Posons $Y = \Sigma^{-1/2}(X-\mu)$. Alors la loi de Y est $N(0, I)$. D'après la Proposition 5.4.4, $Y_1, \ldots, Y_d$ est une suite de v.a.r. i.i.d., de loi commune $N(0,1)$. Donc la loi de Y admet la densité :

$$f_Y(y) = (2\pi)^{-d/2} \exp\left(-|y|^2/2\right)$$

Soit $\varphi \in C_K(\mathbb{R}^d)$.

$$\begin{aligned}
\mathbb{E}[\varphi(X)] &= \mathbb{E}[\varphi(\mu + \Sigma^{1/2}Y)] \\
&= (2\pi)^{-d/2} \int_{\mathbb{R}^d} e^{-|y|^2/2} \varphi(\mu + \Sigma^{1/2}y) dy \\
&= (2\pi)^{-d/2} \int_{\mathbb{R}^d} e^{-|\Sigma^{-1/2}(x-\mu)|^2/2} \varphi(x)(\text{dét } \Sigma^{-1/2}) dx
\end{aligned}$$

Le résultat découle alors de ce que $C_K(\mathbb{R}^d)$ est une classe séparante.

□

5.5 Exercices

Exercice 5.5.1. *Soit X un v.a. de dimension d. Sa loi $\mathbb{P}_X$ est uniquement déterminée par la donnée des lois de toutes les v.a.r. $\lambda \cdot X$, $\lambda \in \mathbb{R}^d$, $|\lambda| = 1$.*

Exercice 5.5.2. *Soit X un vecteur aléatoire de dimension d tel que $\mathbb{E}[|X|^2] < \infty$. Montrer qu'une CN pour que la loi de X admette une densité par rapport à la mesure de Lebesgue sur $\mathbb{R}^d$ est que sa matrice de covariance soit inversible (on montrera que si cette matrice n'est pas inversible, le support de la loi de X est inclu dans un sous–espace affine de dimension au plus $d-1$). La Proposition 5.4.8 montre que dans le cas gaussien cette CN est aussi une CS. Montrer par un contre–exemple que ce n'est pas une CS en général.*

Exercice 5.5.3. *Déduire du Corollaire 2.3.15, en utilisant le raisonnement de la Proposition précédente, le résultat suivant :*

Soit S *un ouvert de* $\mathbb{R}^d$, $\varphi : S \to \mathbb{R}^d$ *une application injective qui possède des dérivés partielles continues sur* S, *et dont le jacobien ne s'annule pas sur* S.

Soit X *un v.a. de dimension* d, *t.q.* $\mathbb{P}(X \in S) = 1$, *dont la loi admet une densité* f_X. *Alors la loi du v.a.* $Y = \varphi(X)$ *admet la densité* f_Y *définie par :*

$$f_Y(x) = \begin{cases} \frac{f_X(\varphi^{-1}(x))}{|J_\varphi(\varphi^{-1}(x))|}, & si\ x \in \varphi(S) \\ 0 & sinon \end{cases}$$

où J_φ, *jacobien de* φ, *est défini par :*

$$J_\varphi(x) = d\acute{e}t \begin{pmatrix} \frac{\partial \varphi_1}{\partial x_1}(x) \cdots & \frac{\partial \varphi_d}{\partial x_1}(x) \\ \vdots & \vdots \\ \frac{\partial \varphi_1}{\partial x_d}(x) \cdots & \frac{\partial \varphi_d}{\partial x_d}(x) \end{pmatrix}$$

Exercice 5.5.4. *Soit* X *et* Y *deux v.a.r. indépendantes. On suppose que la loi de* Y *admet une densité par rapport à la mesure de Lebesgue. En déduire que* $\mathbb{P}(X = Y) = 0$. *Montrer par un contre–exemple que le résultat est faux si les lois de* X *et* Y *sont toutes deux discrètes, à supports non disjoints.*

Indication : on pourra utiliser le Lemme 5.2.6.

Exercice 5.5.5. *Soit* X *et* Y *deux v.a. indépendantes de lois respectives* $\mathcal{P}(\lambda)$ *et* $\mathcal{P}(\mu)$. *Calculer les fonctions caractéristiques de* X, Y *et* $X + Y$. *En déduire la loi de* $X + Y$.

Exercice 5.5.6. *Calculer les fonctions caractéristiques des lois suivantes.*

1. *La loi exponentielle* $\mathcal{E}(\lambda)$.
2. *La loi exponentielle symétrique de densité* $\frac{1}{2}\exp(-|x|)$.
3. *La loi de Cauchy de densité* $(\pi(1 + x^2))^{-1}$.
4. *La loi uniforme sur* $[-a, a]$.
5. *La loi* $N(0, 1)$.

Exercice 5.5.7. *Soit* (X, Y) *un vecteur gaussien centré de matrice de covariance* $\begin{pmatrix} 1 & \sqrt{3}/2 \\ \sqrt{3}/2 & 1 \end{pmatrix}$.

1. *Est-ce-que les variables X et Y sont indépendantes ?*
2. *Quelles sont les lois des v. a. r. $X+Y$ et $X-Y$?*
3. *Quelle est la loi du couple $(X+Y, X-Y)$? Les variables $X+Y$ et $X-Y$ sont-elles indépendantes ?*

Exercice 5.5.8. *On considère un vecteur aléatoire X de loi $N(0,\Sigma)$ sur $\mathbb{R}^3$ avec*

$$\Sigma = \begin{pmatrix} 3 & -1 & 0 \\ -1 & 3 & 0 \\ 0 & 0 & 2 \end{pmatrix}$$

Trouver une application linéaire A de $\mathbb{R}^3$ dans $\mathbb{R}^3$ telle que les coordonnées du vecteur AX soient indépendantes.

Exercice 5.5.9. *Soit X, Y et Z trois v.a.r. indépendantes de loi $N(0,1)$, et $S = X+Y+Z$. Montrer que S et $T = (X-Y)^2 + (Y-Z)^2 + (Z-X)^2$ sont indépendantes.*

Exercice 5.5.10. *On appelle loi du χ^2 à k degrés de liberté et on note χ^2_k la loi de $N_1^2 + \cdots + N_k^2$, si $N_1, \ldots, N_k$ sont des v.a.r. indépendantes de loi $N(0,1)$.*

Soit $(X_1, ..., X_n)$ des v.a.i.i.d telles que $\mathbb{E}(X_1) = 0$ et $\mathbb{E}(X_1^2) = 1$. On pose $\bar{S} = \frac{1}{n}\sum_{i=1}^n X_i$ et $\bar{V} = \frac{1}{n}\sum_{i=1}^n (X_i - \bar{S})^2$

1. *On suppose que $X \sim N(0,1)$. Montrer que $\bar{S}$ et $\bar{V}$ sont indépendantes, et que $n\bar{V} \sim \chi^2_{n-1}$.*
2. *On suppose maintenant que $\bar{S}$ et $\bar{V}$ sont indépendantes. On se propose de montrer que alors X suit une loi normale. Soit φ la fonction caractéristique de X. En calculant $E(n\bar{V}e^{it\bar{S}})$ de deux manières différentes, montrer que $\varphi''\varphi - (\varphi')^2 + \varphi^2 = 0$. Conclure.*
   On pourra noter que $n\bar{V} = \sum_{i=1}^n X_i^2 - n\bar{S}^2$.

Chapitre 6

Convergence des variables aléatoires et loi des grands nombres

6.1 Théorème d'extension de Kolmogorov

Dans la suite, on sera amené à dire : "Soit la suite infinie de v.a.r. $X_1, X_2, \ldots, X_n, \ldots$, telle que pour tout $n \geq 1$, la loi du v.a. $(X_1, \ldots, X_n)$ sur $(\mathbb{R}^n, \mathcal{B}_n)$ soit une loi donnée $\mathbb{P}_n$". Encore faut–il vérifier qu'il existe un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$ sur lequel on puisse définir une telle suite de v.a.r.

Il est clair que la famille des lois $\mathbb{P}_n$ doit vérifier la condition de compatibilité :

$(*)$ $\quad \mathbb{P}_{n+1}\pi_{n+1,n}^{-1} = \mathbb{P}_n, \ \forall\, n \geq 1$, où $\pi_{n+1,n}$ est la projection de $\mathbb{R}^{n+1}$ sur $\mathbb{R}^n$ définie par :

$$\pi_{n+1,n}(x_1, \ldots, x_{n+1}) = (x_1, \ldots, x_n)$$

Soit $\mathbb{R}^\infty$ l'espace des suites infinies de nombres réels, $x = (x_1, \ldots, x_n, \ldots)$. On définit l'application $\pi_n : \mathbb{R}^\infty \to \mathbb{R}^n$ par :

$$\pi_n(x) = (x_1, \ldots, x_n).$$

On appelle σ–**algèbre des ensembles cylindriques (de** $\mathbb{R}^\infty$**) de dimension** n la σ–algèbre :

$$\mathcal{C}_n = \pi_n^{-1}(\mathcal{B}_n)$$

On appelle **algèbre des ensembles cylindriques** : $\mathcal{C} = \bigcup_{n\geq 1} \mathcal{C}_n$ [$\mathcal{C}$ n'est pas une σ–algèbre, mais c'est une algèbre, comme tout réunion croissante d'algèbres.]

Se donner une famille $\mathbb{P}_n$ de probabilités sur $(\mathbb{R}^n, \mathcal{B}_n)$, pour tout $n \geq 1$, qui vérifie la relation de compatibilité $(*)$ est équivalent à se donner une fonction additive d'ensemble $\mathbb{P}$ sur $\mathcal{C}$, telle que pour tout $n \geq 1$, $\mathbb{P}\pi_n^{-1} = \mathbb{P}_n$ [$(*)$ permet de définir sans ambiguïté $\mathbb{P}(A)$, pour tout $A \in \mathcal{C}$]. La question qui se pose est alors : peut–on étendre $\mathbb{P}$ en une probabilité $\hat{\mathbb{P}}$ sur $(\mathbb{R}^\infty, \mathcal{B}_\infty)$, où par définition $\mathcal{B}_\infty = \sigma(\mathcal{C})$? La réponse est positive :

Théorème 6.1.1. *(d'extension de Kolmogorov). Etant donnée une fonction additive d'ensemble $\mathbb{P}$ définie sur $\mathcal{C}$, telle que pour tout $n \geq 1$, $\mathbb{P}\pi_n^{-1}$ soit une mesure de probabilité sur $(\mathbb{R}^n, \mathcal{B}_n)$, il existe une et une seule mesure de probabilité $\hat{\mathbb{P}}$ sur $(\mathbb{R}^\infty, \mathcal{B}_\infty)$, t.q. $\hat{\mathbb{P}}|_{\mathcal{C}} = \mathbb{P}$.*

Remarque 6.1.2. *Exemple d'application du Théorème 6.1.1 Il existe un espace de probabilité $(\Omega, \mathcal{F}, \hat{\mathbb{P}})$ tel que l'on puisse définir sur cet espace une suite infinie de v.a.r. $\{X_n\}_{n\geq 1}$ i.i.d. (indépendantes et identiquement distribuées), de loi commune une probabilité Q sur $(\mathbb{R}, \mathcal{B})$. On choisit $\Omega = \mathbb{R}^\infty$, $\mathcal{F} = \mathcal{B}_\infty$, $\hat{\mathbb{P}} = Q^\infty =$ extension de la fonction additive d'ensemble $\mathbb{P}$ sur $\mathcal{C}$, qui vérifie $\mathbb{P}\pi_n^{-1} = Q^n, \ \forall\, n \geq 1$. Enfin on pose*

$$X_n(\omega) = \omega_n \, [\omega = (\omega_1, \omega_2, \dots)]$$

Preuve L'unicité dans le Théorème est évidente. L'existence résulte du Théorème 1.4.6 et de l'Exercice 4.8.1 si l'on montre que pour toute suite $\{A_n,\ n \geq 1\} \subset \mathcal{C}$ telle que $A_n \downarrow \emptyset$, $\mathbb{P}(A_n) \downarrow 0$. Il suffit donc de montrer que si $\{A_n\} \subset \mathcal{C}$ est une suite décroissante, $\varepsilon > 0$ et $\mathbb{P}(A_n) \geq \varepsilon$ pour tout n, alors $\cap A_n \neq \emptyset$. Sans restreindre la généralité, on peut supposer que $A_n \in \mathcal{C}_n$ pour tout $n \geq 1$, ce qui veut dire qu'il existe $B_n \in \mathcal{B}_n$ tel que $A_n = \pi_n^{-1}(B_n)$, et $\mathbb{P}(A_n) = \mathbb{P}_n(B_n)$. Le théorème d'approximation 1.6.1 permet d'affirmer que pour tout $n \geq 1$, il existe un compact $K_n \subset B_n$ tel que $\mathbb{P}(A_n \backslash \pi_n^{-1}(K_n)) \leq \varepsilon/2^{n+1}$. Alors $C_n := \cap_{j=1}^n \pi_j^{-1}(K_j)$ vérifie $C_n \subset A_n$ et $\mathbb{P}(A_n \backslash C_n) \leq \varepsilon/2$, donc C_n est non vide. Choisissons un point $x^{(n)} \in C_n$. Pour tout $k \leq n$, $(x_1^{(n)}, \dots x_k^{(n)}) \in K_k$. Pour tout k fixé, la suite $(x_k^{(1)}, \dots, x_k^{(n)}, \dots)$ appartient donc au compact K_k à partir d'un certain rang, donc on peut en extraire une sous–suite convergente. Par le procédé diagonal, on fabrique alors une suite $x^{(n_i)}$ telle que $x_k^{(n_i)}$ converge pour tout k dans $\mathbb{N}$ vers un certain $\overline{x}_k$. Désignons par $\overline{x}$ le point de $\mathbb{R}^\infty$ dont la k–ième coordonnée est précisément $\overline{x}_k$. Par construction, $(\overline{x}_1, \ldots, \overline{x}_k) \in K_k \subset B_k$, donc $\overline{x}$ est dans $C_k \subset A_k$ pour tout $k \geq 1$, donc dans $\cap A_n$. $\square$

6.2 Convergence des variables aléatoires

Définition 6.2.1. *Soit $\{X_n, n \geq 1\}$ une suite de v.a.r. [resp. de vecteurs aléatoires de dimension d], et X une v.a.r. [resp. un vecteur aléatoire de dimension d]*

(i) $X_n \to X$ p.s. si $\mathbb{P}(\{\omega, X_n(\omega) \underset{n\to\infty}{\to} X(\omega)\}) = 1$.

(ii) $X_n \to X$ en moyenne d'ordre p, $(p \geq 1)$ si $\mathbb{E}(|X_n|^p + |X|^p) < \infty$ et $\mathbb{E}(|X_n - X|^p) \to 0$, quand $n \to \infty$.

(iii) $X_n \to X$ en probabilité (noté $X_n \overset{p}{\to} X$) si pour tout $\varepsilon > 0$, $\mathbb{P}(|X_n - X| > \varepsilon) \to 0$, quand $n \to \infty$. $\square$

On dira de façon équivalente soit "converge en moyenne d'ordre p", soit "converge dans $L^p(\Omega)$".

Proposition 6.2.2. *Supposons qu'il existe $p \geq 1$ t.q. $X_n \to X$ en moyenne d'ordre p. Alors $X_n \overset{p}{\to} X$.*

Preuve Fixons $\varepsilon > 0$. D'après l'inégalité de Bienaymé Tchebycheff (Proposition 4.7.7),

$$\mathbb{P}(|X_n - X| > \varepsilon) \leq \varepsilon^{-p}\mathbb{E}(|X_n - X|^p).$$

$\square$

Proposition 6.2.3. *Si $X_n \to X$ p.s., alors $X_n \overset{p}{\to} X$.*

Preuve Sous l'hypothèse, si $Y_n = \inf(|X_n - X|, 1)$, $Y_n \to 0$ p.s., et $|Y_n| \leq 1$ p.s., pour tout n. Donc, d'après le Théorème de convergence dominée, $\mathbb{E}(Y_n) = \mathbb{E}(|Y_n|) \to 0$, i.e. $Y_n \to 0$ dans $L^1(\Omega)$, donc en probabilité d'après la Proposition 6.2.2. $\square$

Proposition 6.2.4. *Si $X_n \to X$ p.s., et s'il existe Z v.a.r.* $\mathbb{P}$**–intégrable** *t.q. $|X_n|^p \leq Z$ p.s., pour tout n (avec $p \geq 1$), alors $\mathbb{E}(|X|^p) < \infty$ et*

$$X_n \to X \quad \textit{dans } L^p(\Omega).$$

PREUVE La première affirmation découle du Lemme de Fatou. Donc $X_n, X \in L^P(\Omega)$ et $|X_n - X|^p \leq 2^{p-1}(|X_n|^p + |X|^p) \leq 2^p Z$, et $|X_n - X|^p \to 0$ p.s. Le résultat découle donc du Théorème de convergence dominée. $\square$

On a aussi le

Théorème 6.2.5. *Si $X_n \xrightarrow{p} X$, et s'il existe Z v.a.r. P–**intégrable** et $p \geq 1$ t.q. $[X_n|^p \leq Z$ p.s. pour tout $n \geq 1$, alors*

$$X_n \to X \textit{ dans } L^p(\Omega).$$

PREUVE Pour tout $k \geq 1$,

$$\left\{|X| > Z^{1/p} + \frac{1}{k}\right\} \subset \{|X_n|^p > Z\} \bigcup \left\{|X - X_n| > \frac{1}{k}\right\},$$

$$\text{d'où } \mathbb{P}\left(|X| > Z^{1/p} + \frac{1}{k}\right) \leq \lim_{n\to\infty} \mathbb{P}\left(|X - X_n| > \frac{1}{k}\right) = 0,$$

ceci pour tout $k > 0$, donc $|X|^p \leq Z$ p.s., d'où $X_n, X \in L^p(\Omega)$,

$$\text{et } |X_n - X|^p \leq 2^p Z = Y.$$

$$\begin{aligned}
\mathbb{E}(|X_n - X|^p) &\leq \int_{\{|X_n - X|^p > M\}} |X_n - X|^p d\mathbb{P} + \int_{\{\varepsilon < |X_n - X|^p \leq M\}} |X_n - X|^p d\mathbb{P} + \varepsilon \\
&\leq \int_{\{Y > M\}} Y d\mathbb{P} + M\mathbb{P}(|X_n - X|^p > \varepsilon) + \varepsilon.
\end{aligned}$$

$$\text{Donc } \limsup_{n\to\infty} \mathbb{E}(|X_n - X|^p) \leq \int_{\{Y>M\}} Y d\mathbb{P} + \varepsilon, \forall\, 0 < \varepsilon < M.$$

Or $\mathbf{1}_{\{Y>M\}}Y \leq Y$, Y est intégrable ; et $\mathbf{1}_{\{Y>M\}}Y \downarrow \mathbf{1}_{\{Y=\infty\}}Y = 0$ p.s. quand $M \to \infty$. Donc par le Théorème de convergence monotone, $\int_{\{Y>M\}} Y d\mathbb{P} \to 0$.

Finalement $0 \leq \liminf_n \mathbb{E}([X_n - X|^p) \leq \limsup_n \mathbb{E}(|X_n - X|^p) \leq 0$. D'où $\mathbb{E}(|X_n - X|^p) \to 0$ quand $n \to \infty$. $\square$

Corollaire 6.2.6. *Si $X_n \xrightarrow{p} X$, et s'il existe Y v.a.r. $\mathbb{P}$–**intégrable** t.q. $|X_n| \leq Y$ p.s. pour tout $n \geq 1$, alors X_n et X sont $\mathbb{P}$–intégrables, et $\mathbb{E}(X_n) \to \mathbb{E}(X)$.*

Preuve On applique le Théorème avec $p = 1$. Donc

$$|\mathbb{E}(X_n) - \mathbb{E}(X)| \leq \mathbb{E}(|X_n - X|) \to 0.$$

□

Notons que l'on a généralisé le théorème de convergence dominée en remplaçant la convergence p.s. par la convergence en probabilité.

Proposition 6.2.7. *Si $X_n \xrightarrow{p} X$, alors il existe une sous–suite $\{X_{n_k}\}$ t.q. $X_{n_k} \to X$ p.s.*

Preuve Posons $Y_n = \inf(|X_n - X|, 1)$. $Y_n \xrightarrow{p} 0$, et $|Y_n| \leq 1$. Donc d'après le Théorème 6.2.5, $Y_n \to 0$ dans $L^1(\Omega)$. D'où d'après le Théorème 3.2.2, il existe une sous–suite Y_{n_k} t.q. $Y_{n_k} \to 0$ p.s., d'où $X_{n_k} \to X$ p.s. □

Proposition 6.2.8. *$X_n \xrightarrow{p} X$ si et seulement si de toute sous–suite de la suite $\{X_n\}$ on peut extraire une sous–suite qui converge p.s. vers X.*

Preuve La C.N. résulte de la Proposition 6.2.7.

C.S. Si la suite $\{X_n\}$ ne converge pas en probabilité vers X, alors il existe $\varepsilon, \eta > 0$ et une sous–suite $\{X_{n_k}\}$ t.q. $\mathbb{P}(|X - X_{n_k}| \geq \varepsilon) > \eta$, pour tout $k \geq 1$. Alors aucune sous–suite extraite de la sous–suite $\{X_{n_k}\}$ ne converge p.s. vers X. □

Une conséquence immédiate de la Proposition 6.2.8 est le

Corollaire 6.2.9. *Si $\{X_n\}$ et X sont des v.a. de dimension d, t.q. $X_n \xrightarrow{p} X$, et $g \in C(\mathbb{R}^d; \mathbb{R}^k)$, alors*

$$g(X_n) \xrightarrow{p} g(X).$$

□

Proposition 6.2.10. *Soit $\{X_n, n \geq 1\}$ et X des v.a.r. (ou des vecteurs aléatoires de dimension d). Les assertions suivantes sont équivalentes :*

(i) $X_n \to X$ p.s.

(ii) $\sup_{j \geq n} |X_j - X| \xrightarrow{p} 0$.

Preuve

$$\begin{aligned}\mathbb{P}(\limsup_n \{[X_n - X| > \varepsilon\}) &= \mathbb{P}\left(\bigcap_{n=1}^{\infty} \bigcup_{j=n}^{\infty} \{|X_j - X| > \varepsilon\}\right) \\ &= \lim_n \mathbb{P}\left(\sup_{j \geq n} |X_j - X| > \varepsilon\right)\end{aligned}$$

Mais on a les inclusions (qui mèlent les limsup de réels et d'événements !)
$\{\limsup_n |X_n - X| > \varepsilon\} \subset \limsup_n \{|X_n - X| > \varepsilon\} \subset \{\limsup_n |X_n - X| \geq \varepsilon\}$.
On déduit des égalités et des inclusions ci–dessus

$$\begin{aligned} X_n \to X \text{ p. s } &\Rightarrow \forall\, \varepsilon > 0,\ \mathbb{P}(\limsup_n |X_n - X| \geq \varepsilon) = 0 \\ &\Rightarrow \sup_{j \geq n} |X_j - X| \to 0 \text{ en probabilité} \\ &\Rightarrow \forall\, \varepsilon > 0,\ \mathbb{P}(\limsup_n |X_n - X| > \varepsilon) = 0 \\ &\Rightarrow X_n \to X \text{ p.s.,} \end{aligned}$$

ce qui établit l'équivalence de l'énoncé. □

Proposition 6.2.11. *Soit $\{X_n, n \geq 1\}$ une suite de v.a.r. (ou de vecteurs aléatoires de dimension d). Les deux assertions suivantes sont équivalentes :*

(i) $\sup_{j \geq n} |X_j - X_n| \xrightarrow{p} 0$*, quand* $\to \infty$.

(ii) il existe une v.a.r. X t.q. $X_n \to X$ p.s. [resp. un v.a. de dimension d]

Preuve

— (ii) ⇒ (i) résulte de la Proposition précédente, puisque

$$\sup_{j \geq n} |X_j - X_n| \leq \sup_{j \geq n} |X_j - X| + |X - X_n|$$

$$\mathbb{P}\left(\sup_{j \geq n} |X_j - X_n| > \varepsilon\right) \leq \mathbb{P}\left(\sup_{j \geq n} |X_j - X| > \varepsilon/2\right) + \mathbb{P}(|X - X_n| > \varepsilon/2) \to 0.$$

— (i) ⇒ (ii) Posons $Y_n = \sup_{j \geq n} |X_j - X_n|$. D'après la Proposition 6.2.7, il existe une sous–suite $\{Y_{n_k}\}$ t.q. $Y_{n_k} \to 0$ p.s. Alors a fortiori $\sup_{\ell \geq k} |X_{n_\ell} - X_{n_k}| \to 0$ p.s., donc la suite $\{X_{n_k}(\omega)\}$ est p.s. une suite de Cauchy, qui converge dans $\mathbb{R}$ [ou $\mathbb{R}^d$]. Donc, d'après le Théorème 2.1.3, il existe une v.a.r. [ou un v.a. de dimension d] X t.q. $X_{n_k} \to X$ p.s. Et pour tout k tel que $n_k \geq n$,

$$\begin{aligned} \sup_{j \geq n} |X_j - X| &\leq \sup_{j \geq n} |X_j - X_n| + |X_n - X_{n_k}| + |X_{n_k} - X| \\ &\leq 2 \sup_{j \geq n} |X_j - X_n| + |X_{n_k} - X|. \end{aligned}$$

$$\mathbb{P}(\sup_{j\geq n}|X_j - X| > \varepsilon) \leq \mathbb{P}(\sup_{j\geq n}|X_j - X_n| > \varepsilon/4) + \mathbb{P}(|X_{n_k} - X| > \varepsilon/2).$$

Faisant tendre $n_k \to +\infty$, on obtient :

$$\mathbb{P}(\sup_{j\geq n}|X_j - X| > \varepsilon) \leq \mathbb{P}(\sup_{j\geq n}|X_j - X_n| > \varepsilon/4)$$

D'où :

$$0 \leq \liminf_{n\to\infty} \mathbb{P}\left(\sup_{j\geq n}|X_j - X| > \varepsilon\right) \leq \limsup_{n\to\infty} \mathbb{P}\left(\sup_{j\geq n}|X_j - X| > \varepsilon\right) \leq 0$$

Le résultat découle alors à nouveau de la Proposition 6.2.10.

□

6.3 La loi faible des grands nombres

Théorème 6.3.1. *Soit* $\{X_n, n \geq 1\}$ *une suite de v.a.r. i.i.d.(indépendantes et identiquement distribuées), t.q.* $\mathbb{E}(X_1^2) < \infty$. *Alors*

$$\frac{X_1 + \cdots + X_n}{n} \xrightarrow{p} \mathbb{E}(X_1), \ \textit{quand } n \to \infty.$$

Preuve Posons $\mu = \mathbb{E}(X_1), \sigma^2 = Var(X_1), S_n = X_1 + \cdots + X_n$. $\mathbb{E}(S_n) = n\mu$, $Var(S_n) = n\,\sigma^2$. Appliquons l'inégalité de Bienaymé– Tchebycheff (4.7.7) :

$$\mathbb{P}\left(\left|\frac{S_n}{n} - \mu\right| > \varepsilon\right) = \mathbb{P}(|S_n - n\mu| > n\varepsilon) \leq \frac{n\sigma^2}{n^2\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2} \to 0.$$

□

L'indépendance de $\{X_n\}$ a servi uniquement à assurer que

$$Var(S_n) = \sum_1^n Var(X_k).$$

On démontre aisément la généralisation suivante du Théorème précédent :

Théorème 6.3.2. *Soit $\{X_n, n \geq 1\}$ une suite de v.a.r. de carré intégrable t.q. $Cov(X_k, X_\ell) = 0$, si $k \neq \ell$. On pose :*

$$\mu_i = \mathbb{E}(X_i), \sigma_i^2 = Var(X_i^2),\ \ i \geq 1.\ \textit{Alors}$$

si $\displaystyle\frac{1}{n^2}\sum_1^n \sigma_i^2 \to 0$, *quand* $n \to \infty$,

$$\mathbb{P}\left(\left|\frac{X_1 + \cdots + X_n}{n} - \frac{\mu_1 + \cdots + \mu_n}{n}\right| \geq \varepsilon\right) \to 0,\ \forall\, \varepsilon > 0.$$

□

6.4 Loi $0-1$ de Kolmogorov

Etant donné une suite de v.a. $\{X_n, n \geq 1\}$, on note $\sigma(X_n, X_{n+1}, \ldots)$ la plus petite σ–algèbre qui contient $\sigma(X_{n+p})$, pour tout $p \geq 0$.

Définition 6.4.1. *On appelle tribu asymptotique de la suite $\{X_n, n \geq 1\}$ la tribu* $\mathcal{G} = \displaystyle\bigcap_{n=1}^{\infty} \sigma(X_n, X_{n+1}, \ldots)$.

□

Exemple 6.4.2.
- $E = \{\omega; \dfrac{X_1(\omega) + \cdots + X_n(\omega)}{n}$ *converge* $\} \in \mathcal{G}$
- *Si* $A_n = \{X_n \in B_n\}$ $(B_n \in \mathcal{B}, \forall\, n)$ $\limsup_n A_n \in \mathcal{G}$.

Proposition 6.4.3. *Soit $\{X_n, n \geq 1\}$ une suite de v.a.r.* **indépendantes**,*$(i_1, i_2, \ldots)$ et $(j_1, j_2, \ldots)$ deux sous–suites* **disjointes** *d'indices. Alors les σ–algèbres : $\mathcal{F}_1 = \sigma(X_{i_1}, X_{i_2}, \ldots)$ et $\mathcal{F}_2 = \sigma(X_{j_1}, X_{j_2}, \ldots)$ sont indépendantes.*

PREUVE Définissons les algèbres :

$$\mathcal{F}_1^0 = \bigcup_{n=1}^{\infty} \sigma(X_{i_1}, \ldots, X_{i_n}) \text{ et } \mathcal{F}_2^0 = \bigcup_{n=1}^{\infty} \sigma(X_{j_1}, \ldots X_{j_n})$$

Si $A \in \mathcal{F}_1^0$ et $B \in \mathcal{F}_2^0$, alors d'après la Proposition 4.6.6,

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

Il nous reste à montrer que $(*)$ est vraie, pour tout $A \in \mathcal{F}_1$, $B \in \mathcal{F}_2$. L'égalité est élémentairee dès que $\mathbb{P}(A) = 0$ ou $\mathbb{P}(B) = 0$.

Fixons $B \in \mathcal{F}_2^0$, t.q. $\mathbb{P}(B) \neq 0$. On pose : $Q_1(A) \triangleq \dfrac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$, $A \in \mathcal{F}_1$. Q_1 est une probabilité sur $\mathcal{F}_1$, qui coïncide avec $\mathbb{P}$ sur $\mathcal{F}_1^0$. Donc $Q_1 = \mathbb{P}$, d'où on a $(*)$ pour tout $A \in \mathcal{F}_1$ $B \in \mathcal{F}_2^0$.

Soit maintenant $A \in \mathcal{F}_1$, t.q. $\mathbb{P}(A) \neq 0$. On définit $Q_2(B) = \dfrac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$. Alors Q_2 et $\mathbb{P}$ sont deux probabilités sur $\mathcal{F}_2$, qui coïncident sur $\mathcal{F}_2^0$, donc sur $\mathcal{F}_2$. □

Théorème 6.4.4. *(Loi $0-1$ de Kolmogorov) Soit $\{X_n\}_{n\geq 1}$ une suite infinie de v.a.r. indépendantes. Alors, si $E \in \mathcal{G}$, la tribu asymptotique associée, $\mathbb{P}(E) = 0$ ou 1.*

Preuve Posons $\mathcal{F}_0 = \displaystyle\bigcup_{n=1}^{\infty} \sigma(X_1, \ldots, X_n)$. Alors $\mathcal{F}_0$ est une algèbre, et si $\mathcal{F} = \sigma(\mathcal{F}_0)$, $\mathcal{G} \subset \mathcal{F}$. Soit $E \in \mathcal{G}$. Alors $E \in \mathcal{F}$. Donc d'après la Proposition 1.4.12, il existe une suite $\{E_j\} \subset \mathcal{F}_0$ t.q. $\mathbb{P}(E \triangle E_j) \to 0$, $j \to \infty$. Donc $\mathbb{P}(E_j) \to \mathbb{P}(E)$ et $\mathbb{P}(E \cap E_j) \to \mathbb{P}(E)$.

Mais il existe n_j t.q. $E_j \in \sigma(X_1, \ldots, X_{n_j})$ et $E \in \sigma(X_{n_j+1}, \ldots)$, et d'après la Proposition 6.4.3,

$$\mathbb{P}(E \cap E_j) = \mathbb{P}(E)\mathbb{P}(E_j).$$

Donc en faisant tendre $j \to \infty$

$$\mathbb{P}(E) = (\mathbb{P}(E)^2,$$

d'où le résultat. □

On sait maintenant que beaucoup d'événements qui nous intéressent sont de probabilité soit 0, soit 1. Comment distinguer les deux cas ? Un outil est le lemme de Borel–Cantelli. Il permet par exemple de montrer la

Proposition 6.4.5. *Soit $\{X_n\}_{n\geq 1}$ une suite infinie de v.a.r. indépendantes et de même loi. On définit :*

$$A_n = \{|X_n[> n\}$$

$$\textit{Alors } \mathbb{E}|X_1| < \infty \Leftrightarrow \mathbb{P}[\limsup A_n] = 0$$

PREUVE Il suffit essentiellement de prendre l'espérance dans $\sum_1^\infty \mathbf{1}_{\{|X_1|>n\}} \leq |X_1| \leq \sum_0^\infty \mathbf{1}_{\{|X_1|>n\}}$ p.s., et d'utiliser le Lemme de Borel–Cantelli. □

6.5 Convergence des séries

On va donner une condition suffisante pour qu'une série de v.a.r. converge p.s. On a d'abord la généralisation suivante de l'inégalité de Bienaymé Tchebycheff :

Théorème 6.5.1. *(Inégalité de Kolmogorov). Soit $X_1, \ldots, X_n$ des v.a.r. indépendantes, toutes de carré intégrable et d'espérance nulle. On pose $S_k = X_1 + \cdots + X_k$. Alors pour tout $\alpha > 0$,*

$$\mathbb{P}\left(\max_{1\leq k\leq n} |S_k| \geq \alpha\right) \leq \frac{1}{\alpha^2}\mathbb{E}(S_n^2)$$

PREUVE On pose $A = \{\max_{1\leq k\leq n} [S_k| \geq \alpha\}$,

$$A_k = \{[S_k| \geq \alpha\} \bigcap \{\max_{1\leq j\leq k-1} |S_j| < \alpha\}$$

$$A = \bigcup_1^n A_k \text{ et } A_k \cap A_\ell = \emptyset, \text{ si } k \neq \ell$$

$$\begin{aligned}
\mathbb{E}(S_n^2) &\geq \int_A S_n^2 d\mathbb{P} \\
&= \sum_{k=1}^n \int_{A_k} S_n^2 d\mathbb{P} \\
&= \sum_{k=1}^n \int_{A_k} \left[(S_k^2 + 2S_k(S_n - S_k) + (S_n - S_k)^2\right] d\mathbb{P} \\
&\geq \sum_{k=1}^n \left[\int_{A_k} S_k^2 d\mathbb{P} + 2\mathbb{E}(\mathbf{1}_{A_k} S_k)\mathbb{E}(S_n - S_k)\right] \\
&\geq \alpha^2 \sum_1^n \mathbb{P}(A_k) = \alpha^2 \mathbb{P}(A).
\end{aligned}$$

On a utilisé à l'avant–dernière ligne le fait que $\mathbf{1}_{A_k} S_k$ et $S_n - S_k$ sont indépendantes, ce qui se déduit de ce que $A_k \in \sigma(X_1, \dots, X_k)$, $S_n - S_k$ est $\sigma(X_{k+1}, \dots, X_n)$–mesurable. □

Théorème 6.5.2. *Soit $\{X_n\}_{n\geq 1}$ une suite de v.a.r. indépendantes et de carré intégrable, t.q.*

(i) $\mathbb{E}(X_n) = 0$, *pour tout* $n \geq 1$;

(ii) $\sum_1^\infty \mathbb{E}(X_n^2) < \infty$

Alors $S_n = X_1 + \cdots + X_n$ *converge p.s. vers une limite* S *quand* $n \to \infty$.

Preuve D'après le Théorème 6.5.1,

$$\mathbb{P}[\max_{1\leq k\leq r} |S_{n+k} - S_n| > \varepsilon] \leq \frac{1}{\varepsilon^2} \sum_{k=1}^{r} \mathbb{E}(X_{n+k}^2)$$

Par la continuité monotone séquentielle des probabilités,

$$\mathbb{P}[\sup_{1\leq k} |S_{n+k} - S_n| > \varepsilon] \leq \frac{1}{\varepsilon^2} \sum_{n+1}^{\infty} \mathbb{E}(X_i^2),$$

soit

$$\mathbb{P}[\sup_{j\geq n} |S_j - S_n| > \varepsilon] \leq \frac{1}{\varepsilon^2} \sum_{n+1}^{\infty} \mathbb{E}(X_i^2)$$

Donc, par (ii), $\mathbb{P}[\sup_{j\geq n} |S_j - S_n| > \varepsilon] \underset{n\to\infty}{\to} 0$, pour tout $\varepsilon > 0$

Soit $\sup_{j\geq n} |S_j - S_n| \overset{p}{\to} 0$, donc d'après la Proposition 6.2.11, il existe une v.a.r. S t.q. $S_n \to S$ p.s. □

Remarque 6.5.3. *Sous les hypothèses du théorème, on a aussi que* S_n *converge dans* $L^2(\Omega, \mathcal{F}, \mathbb{P})$

□

6.6 Loi forte des grands nombres

Théorème 6.6.1. *Soit* $\{X_n\}_{n\geq 1}$ *une suite de v.a.r.* **indépendantes** *t.q.*

(i) $\mathbb{E}(X_n) = 0$

(ii) $\sum_n \mathbb{E}(X_n^2)/n^2 < \infty$

Alors $\frac{1}{n}\sum_1^n X_k \to 0$ *p.s.*

Preuve Posons $S_n = X_1 + \cdots + X_n$, $\sigma_n^2 = \mathbb{E}(X_n^2)$.

Il suffit de montrer que pour tout $\varepsilon > 0$, si l'on pose

$$B_k = \{\omega;\ \text{il existe } n \in [2^k, 2^{k+1}[\text{ t. q. } |S_n(\omega)| > n\varepsilon\},$$

$$\text{alors } \mathbb{P}(\limsup B_k) = 0.$$

Pour cela, d'après le Lemme de Borel–Cantelli, il suffit de vérifier que $\sum_{k=1}^{\infty} \mathbb{P}(B_k) < \infty$. Si $\omega \in B_k$, alors $|S_n(\omega)| > \varepsilon 2^k$ pour un $n < 2^{k+1}$.

Donc d'après le Théorème 6.5.1,

$$\mathbb{P}(B_k) \leq \mathbb{P}\left(\sup_{n\leq 2^{k+1}} |S_n| > \varepsilon 2^k\right)$$
$$\leq \frac{1}{\varepsilon^2 2^{2k}} \sum_{n=1}^{2^{k+1}} \sigma_n^2.$$

$$\sum_0^{\infty} \mathbb{P}(B_k) \leq \sum_{k=0}^{\infty}\sum_{n=1}^{2^{k+1}} \frac{\sigma_n^2}{\varepsilon^2 2^{2k}}$$
$$= \frac{1}{\varepsilon^2}\sum_{n=1}^{\infty}\sigma_n^2 \sum_{k;2^{k+1}\geq n} 2^{-2k}.$$

Il reste à montrer que $\sum_{k;2^{k+1}\geq n} 2^{-2k} \leq c/n^2$. Or

$$\sum_{k_n}^{\infty} 2^{-2k} = 2^{-2k_n}\sum_0^{\infty} 2^{-2k}$$
$$= \frac{4}{3}2^{-2k_n}.$$

Mais

$$2^{k_n+1} \geq n \Rightarrow 2^{-2k_n} \times 2^{-2} \leq \frac{1}{n^2}$$

Donc

$$\frac{4}{3} 2^{-2k_n} \leq \frac{16}{3n^2}$$

□

Dans le cas où les X_n sont i.i.d., on n'a besoin que d'une condition sur le premier moment :

Théorème 6.6.2. *Soit $\{X_n, n \geq 1\}$ une suite de v.a.r. i.i.d. (indépendantes et identiquement distribuées).*

(i) Si $\mathbb{E}(|X_1|) < \infty$, alors $\dfrac{X_1 + \cdots + X_n}{n} \to \mathbb{E}(X_1)$ p.s.

(ii) Si $\mathbb{E}(|X_1|) = +\infty$, alors $\dfrac{X_1 + \cdots + X_n}{n}$ diverge p.s.

Preuve (i) On va se ramener au Théorème 6.6.1. On pose

$$\tilde{X}_n = X_n \mathbf{1}_{\{|X_n| \leq n\}},\ A_n = \{|X_n| > n\} = \{X_n \neq \tilde{X}_n\}$$

D'après la Proposition 6.4.5, puisque $\mathbb{E}(|X_1|) < \infty$, $\mathbb{P}(\limsup A_n) = 0$. D'où grâce au Lemme de Borel–Cantelli, p.s., $\tilde{X}_n(\omega) = X_n(\omega)$ à partir d'un certain rang $N(\omega)$. La conclusion de (i) est donc équivalente à :

$$\frac{\tilde{X}_1 + \cdots + \tilde{X}_n}{n} \to \mathbb{E}(X_1) \text{ p.s.}$$

Mais comme $\mathbb{E}(\tilde{X}_n) \to \mathbb{E}(X_1)$ [convergence dominée], il suffit de montrer que

$$\frac{1}{n} \sum_1^n (\tilde{X}_k - \mathbb{E}(\tilde{X}_k)) \to 0 \text{ p.s.}$$

Ceci résulte du Théorème 6.6.1, à condition que $\displaystyle\sum_1^\infty \frac{Var(\tilde{X}_n)}{n^2} < \infty$.

Or

$$Var(\tilde{X}_n) \leq \mathbb{E}(\tilde{X}_n^2)$$

Donc il nous reste à montrer que

$$\sum_1^\infty \mathbb{E}(\tilde{X}_n^2)/n^2 < \infty.$$

Mais :

$$\begin{aligned}
\sum_1^\infty \mathbb{E}(\tilde{X}_n^2)/n^2 &= \sum_1^\infty \frac{1}{n^2} \int_{\{|x|\le n\}} x^2 \mathbb{P}_{X_1}(dx) \\
&= \sum_{n=1}^\infty \sum_{k=1}^n \frac{1}{n^2} \int_{\{k-1<|x|\le k\}} x^2 \mathbb{P}_{X_1}(dx) \\
&= \sum_{k=1}^\infty \left(\sum_{n=k}^\infty \frac{1}{n^2} \right) \int_{\{k-1<|x|\le k\}} x^2 \mathbb{P}_{X_1}(dx) \\
&< \sum_{k=1}^\infty \frac{2}{k} \int_{\{k-1<|x|\le k\}} x^2 \mathbb{P}_{X_1}(dx) \\
&\le 2\mathbb{E}(|X_1|),
\end{aligned}$$

où on a utilisé l'inégalité :

$$\sum_k^\infty \frac{1}{n^2} < \frac{1}{k^2} + \int_k^\infty \frac{dx}{x^2} < \frac{2}{k}$$

(ii) Posons $S_n = X_1 + \cdots + X_n$. Supposons, par l'absurde, que $\frac{S_n}{n}$ converge sur un ensemble de mesure > 0. Alors, d'après le Théorème 6.4.4, S_n/n converge p.s., d'où :

$$\frac{X_n}{n} = \frac{S_n}{n} - \left(\frac{n-1}{n}\right) \frac{S_{n-1}}{n-1} \to 0 \text{ p.s.}$$

Donc, si $A_n = \{|X_n| > n)$, $\mathbb{P}(\limsup A_n) = 0$, ce qui entraîne, d'après la Proposition 6.4.5, $\mathbb{E}(|X_1|) < \infty$, ce qui est contraire à l'hypothèse. □

6.7 Exercices

Exercice 6.7.1. Convergence en probabilité et convergence p.s.
Soit $(X_n, n \ge 1)$ une suite de variables aléatoires réelles définies sur $(\Omega, \mathcal{F}, P)$. Montrer que si pour tout $\epsilon > 0$, $\sum_{n=1}^{+\infty} \mathbb{P}(|X_n - X| \ge \epsilon) < +\infty$, alors X_n converge p.s. vers X.

Exercice 6.7.2. Contrexemples

1. **La convergence en probabilité n'implique pas la convergence presque sûre**.
On munit $(\Omega = [0,1], \mathcal{F} = \mathcal{B}([0,1]))$ de la mesure de Lebesgue λ. On rappelle que tout entier naturel n peut s'écrire de manière unique sous la forme $n = 2^m + k$, où $m, k \in \mathbb{N}$ avec $k \leq 2^m - 1$. On définit la suite d'évènements $A_n = [k2^{-m}, (k+1)2^{-m}[$, et on pose $X_n = \mathbf{1}_{A_n}$. Montrer que $X_n \xrightarrow{p} 0$, quand $n \to \infty$, mais pas presque sûrement.

2. **La convergence en probabilité n'implique pas la convergence dans L^p**.
Soit $(X_n)_{n\geq 1}$ une suite de v.a. telles que
$$\mathbb{P}(X_n = n) = \frac{1}{\ln n} = 1 - P(X_n = 0)$$
Montrer que $X_n \xrightarrow{p} 0$, mais que X_n ne converge vers 0 dans aucun des L^p, $p > 0$.

3. **La convergence L^p n'implique pas la convergence presque sûre**. *Soit $(X_n)_{n\geq 1}$ une suite de v.a. indépendantes telles que*
$$\mathbb{P}(X_n = 0) = 1 - \frac{1}{n^{1/4}}; \quad \mathbb{P}(X_n = \pm 1) = \frac{1}{2n^{1/4}}$$
Montrer que $X_n \xrightarrow{L^2} 0$, mais que X_n ne converge pas presque sûrement vers 0.

4. **La convergence presque sûre n'implique pas la convergence dans L^p**. *Soit $(X_n)_{n\geq 1}$ une suite de v.a. telles que*
$$\mathbb{P}(X_n = 0) = 1 - \frac{1}{n^\alpha}; \quad \mathbb{P}(X_n = \pm n) = \frac{1}{2n^\alpha}$$
Montrer que pour tout $\alpha \in]1,2]$ $X_n \to 0$ p. s., mais que X_n ne converge pas vers 0 dans L^2.

Exercice 6.7.3. *Soit $(X_n, n \geq 1)$ une suite de variables aléatoires réelles indépendantes équidistribueés, telles que*
$$\forall n \geq 1, \forall \epsilon > 0,\ \mathbb{P}(X_n \geq 0) = 1;\ \mathbb{P}(X_n < \epsilon) > 0.$$
Soit $Y_n = \inf(X_1, \cdots, X_n)$.

1. *Montrer que Y_n converge en probabilité vers* 0.
2. *Est-ce que Y_n converge presque sûrement vers* 0 *?*

Exercice 6.7.4. — *Montrer que si $1 \leq p < q$, $X_n \to X$ dans $L^q(\Omega)$ entraîne que $X_n \to X$ en moyenne d'ordre p.*

— *Montrer par un contre–exemple que $X_n \xrightarrow{p} X$ n'entraîne pas $X_n \xrightarrow{p.s.} X$. (on pourra choisir $(\Omega, \mathcal{F}, \mathbb{P}) = ([0,1], B([0,1]), \lambda), X = 0, X_n = \mathbf{1}_{B_n}$, les B_n étant des boréliens de $[0,1]$ "bien choisis").*

Exercice 6.7.5. *Soit $(X_n)_{n\geq 1}$ une suite de v.a., et soit $\mathcal{G} = \cap_{n=1}^{\infty} \sigma(X_n, X_{n+1}, \cdots)$ la tribu asymptotique associée.*

1. *Parmi les évènements suivants, quels sont ceux qui appartiennent à $\mathcal{G}$?*
 $A = \{X_n = 0$ *pour tout* $n \geq 1\}$;
 $B = \{\sum_{n=1}^{\infty} X_n$ *converge* $\}$;
 $C = \{\sum_{n=1}^{\infty} X_n$ *converge vers une limite* $> 0\}$;
 $D = \left\{\lim_{n\to\infty} \frac{1}{c_n} \sum_{k=1}^{n} X_k = \mu\right\}$ *où c_n est une suite tendant vers $+\infty$;*
 $E = \left\{\limsup_{n\to\infty} \frac{X_{3n+2}-5}{2^n} \leq 10\right\}$.
2. *Si on suppose les X_i indépendantes, que dire de la probabilité de tels évènements ?*

Exercice 6.7.6. *Soit $(X_n, n \geq 1)$ une suite de v.a. indépendantes.*

1. *Montrer que toute variable aléatoire asymptotique, i.e. toute v.a. sur $(\Omega, \mathcal{G})$ est presque sûrement constante.*
2. *Est ce que $\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} X_k$ est presque sûrement constante ?*
3. *On suppose que $X_n \xrightarrow{p} Y$ quand $n \to \infty$. Montrer que Y est une v.a. presque sûrement constante.*

Exercice 6.7.7. *Soit $(X_n, n \geq 1)$ une suite de v.a. i.i.d de loi de Bernoulli de paramètre p. On pose pour tout n $Y_n = X_n X_{n+1}$.*

1. *Quelle est la loi de Y_n ?*
2. *Calculer Cov(Y_n, Y_{n+k}) pour tout n et tout k.*
3. *Montrer que $\bar{Y}_n = \frac{1}{n} \sum_{k=1}^{n} Y_k \xrightarrow{p} p^2$.*

Exercice 6.7.8. *Soit $(\epsilon_n)_{n\geq 1}$ une suite de v.a. i.i.d telles que $\mathbb{P}(\epsilon_n = \pm 1) = 1/2$. Soit $0 < a < 1$. On pose $X_n = 1 + a\epsilon_n$. Montrer que le produit $X_1 \times \cdots \times X_n$ converge presque sûrement vers 0.*

Chapitre 7

Convergence en loi et TCL

7.1 Définition et premières propriétés

Tous les v.a. seront supposés construits sur un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$. Rappelons (cf. Lemme 5.1.3) que l'ensemble $C_b(\mathbb{R}^d)$ est séparant. On pose les définitions.

Définition 7.1.1. *Soit $\{\mathbb{Q}_n, n \geq 1\}$ et $\mathbb{Q}$ des probabilités sur $(\mathbb{R}^d, \mathcal{B}_d)$. On dit que $\mathbb{Q}_n$* **converge étroitement** *vers $\mathbb{Q}$, et on note $\mathbb{Q}_n \Rightarrow \mathbb{Q}$ si*

$$\textit{pour tout } f \in C_b(\mathbb{R}^d), \ \int f(x)\mathbb{Q}_n(dx) \to \int f(x)\mathbb{Q}(dx), \ \textit{quand } n \to \infty.$$

Définition 7.1.2. *Soit $\{X_n, n \geq 1\}$ et X des v.a. de dimension d. On dit que X_n* **converge en loi** *vers X, et on note $X_n \xrightarrow{\mathcal{L}} X$, si $\mathbb{P}_{X_n} \Rightarrow \mathbb{P}_X$, i.e.*

$$\textit{pour tout } f \in C_b(\mathbb{R}^d), \quad \mathbb{E}[f(X_n)] \to \mathbb{E}[f(X)], \ \textit{quand } n \to \infty.$$

Définition 7.1.3. *Soit $\mathbb{Q}$ une probabilité sur $(\mathbb{R}^d, \mathcal{B}_d)$. On dit que $A \in \mathcal{B}_d$ est un* **ensemble de $\mathbb{Q}$–continuité** *si $\mathbb{Q}(\partial A) = 0$, où $\partial A = \bar{A} - \overset{\circ}{A}$.*

Théorème 7.1.4. *Soit $\{\mathbb{Q}_n, n \geq 1\}$ et $\mathbb{Q}$ des probabilités sur $(\mathbb{R}^d, \mathcal{B}_d)$. Il y a équivalence entre :*

(i) $\mathbb{Q}_n \Rightarrow \mathbb{Q}$ *;*

(ii) $\limsup_{n\to\infty} \mathbb{Q}_n(F) \leq \mathbb{Q}(F)$, *pour tout F fermé ;*

(iii) $\liminf_{n\to\infty} \mathbb{Q}_n(G) \geq \mathbb{Q}(G)$, *pour tout G ouvert ;*

(iv) $\lim_{n\to\infty} \mathbb{Q}_n(A) = \mathbb{Q}(A)$, *pour tout A ensemble de $\mathbb{Q}$–continuité.*

Théorème 7.1.5. *Soit $\{X_n, n \geq 1\}$ et X des v.a. de dimension d. Il y a équivalence entre :*

(i) $X_n \xrightarrow{\mathcal{L}} X$ *;*

(ii) $\limsup_{n\to\infty} \mathbb{P}(X_n \in F) \leq \mathbb{P}(X \in F)$, *pour tout F fermé ;*

(iii) $\liminf_{n\to\infty} \mathbb{P}(X_n \in G) \geq \mathbb{P}(X \in G)$, *pour tout G ouvert ;*

(iv) $\lim_{n\to\infty} \mathbb{P}(X_n \in A) = \mathbb{P}(X \in A)$, *pour tout A ensemble de $\mathbb{P}_X$–continuité.*

Preuve du Théorème 7.1.4 (i) $\Rightarrow$ (ii) Soit F fermé et $f_\varepsilon(x)$ comme au Lemme 5.1.3. Comme $\mathbf{1}_F \leq f_\varepsilon$,

$$\begin{aligned}\limsup_n \mathbb{Q}_n(F) &\leq \lim_n \int f_\varepsilon(x)\mathbb{Q}_n(dx) \\ &= \int f_\varepsilon(x)\mathbb{Q}(dx) \to \mathbb{Q}(F) \text{ quand } \varepsilon \to 0.\end{aligned}$$

(ii) $\Rightarrow$ (iii) Soit G un ouvert, $F = G^c$.

$$\begin{aligned}\liminf \mathbb{Q}_n(G) &= 1 - \limsup \mathbb{Q}_n(F) \\ &\geq 1 - \mathbb{Q}(F) \\ &= \mathbb{Q}(G).\end{aligned}$$

(ii)+(iii)$\Rightarrow$ (iv)

$$\begin{aligned}\mathbb{Q}(\mathring{A}) &\leq \liminf \mathbb{Q}_n(\mathring{A}) \\ &\leq \liminf \mathbb{Q}_n(A) \\ &\leq \limsup \mathbb{Q}_n(A) \\ &\leq \limsup \mathbb{Q}_n(\bar{A}) \\ &\leq \mathbb{Q}(\bar{A})\end{aligned}$$

Donc si $\mathbb{Q}(\mathring{A}) = \mathbb{Q}(\bar{A})$, $\mathbb{Q}_n(A) \to \mathbb{Q}(A)$

(iv) $\Rightarrow$ (i) Soit $f \in C_b(\mathbb{R}^d)$. Alors il existe K t.q. $|f(x)| \leq K$, pour tout x. Soit $\varepsilon > 0$, et $\alpha_0 < \alpha_1 < \cdots < \alpha_\ell$ t.q.

a) $\alpha_0 < -K < K < \alpha_\ell$,

b) $\alpha_i - \alpha_{i-1} < \varepsilon$,

c) $\mathbb{Q}(\{x; f(x) = \alpha_i\}) = 0,\ i = 0, \ldots, \ell$.

L'existence de tels α_i résultera du Lemme 7.1.6 Posons

$$A_i = \{x; \alpha_{i-1} < f(x) \leq \alpha_i\}$$

f étant continue,

$$\bar{A}_i \subset \{x; \alpha_{i-1} \leq f(x) \leq \alpha_i\},$$
$$\mathring{A}_i \supset \{x; \alpha_{i-1} < f(x) < \alpha_i\},$$

d'où

$$\partial\, A_i \subset \{x; f(x) = \alpha_{i-1}\} \cup \{x; f(x) = \alpha_i\}.$$

Donc

$$\mathbb{Q}(\partial A_i) = 0,\ \ i = 1, \ldots, \ell.$$

Or

$$|\int f\, d\mathbb{Q}_n - \sum_1^\ell \alpha_i \mathbb{Q}_n(A_i)| \leq \varepsilon,$$
$$|\int f\, d\mathbb{Q} - \sum_1^\ell \alpha_i \mathbb{Q}(A_i)| \leq \varepsilon,$$
$$\sum_1^\ell \alpha_i\, \mathbb{Q}_n(A_i) \to \sum_1^\ell \alpha_i\, \mathbb{Q}(A_i).$$

Donc

$$\begin{aligned}
\int f d\mathbb{Q} - 2\varepsilon &\leq \sum_1^\ell \alpha_i \mathbb{Q}(A_i) - \varepsilon \\
&= \lim_n \sum_1^\ell \alpha_i \mathbb{Q}_n(A_i) - \varepsilon \\
&\leq \liminf_n \int f d\mathbb{Q}_n \\
&\leq \limsup_n \int f d\mathbb{Q}_n \\
&\leq \lim_n \sum_1^\ell \alpha_i \mathbb{Q}_n(A_i) + \varepsilon \\
&= \sum_1^\ell \alpha_i \mathbb{Q}(A_i) + \varepsilon \\
&\leq \int f d\mathbb{Q} + 2\varepsilon,
\end{aligned}$$

ceci pour tout $\varepsilon > 0$. □

Lemme 7.1.6. *$E = \{\alpha \in \mathbb{R}; \mathbb{Q}f^{-1}(\alpha) > 0\}$ est au plus dénombrable.*

Preuve Soit F la fonction de répartition de la loi de probabilité $\mathbb{Q}f^{-1}$ sur $(\mathbb{R}, \mathcal{B})$. E est l'ensemble des points de discontinuité de F. Posons

$$E_n = \{x; F(x) - F(x^-) \geq \frac{1}{n}\}$$
$$\text{card}E_n \leq n, \text{ et } E_n \uparrow E,$$

donc E est au plus dénombrable. □

Théorème 7.1.7. Cas $d = 1$ *Soit $\{X_n, n \geq 1\}$ et X des variables aléatoires réelles. Soit F_n la fonction de répartition de X_n, F celle de X. Il y a équivalence entre :*

(i) $X_n \xrightarrow{\mathcal{L}} X$;

(ii) $F_n(x) \to F(x)$, pour tout $x \in C(F)$, ensemble des points de continuité de F.

Remarque 7.1.8. *D'après le Lemme 7.1.6, le complémentaire de $C(F)$ est au plus dénombrable.*

Preuve (i) $\Rightarrow$ (ii) résulte du Théorème 7.1.5 (iv), avec $A =]-\infty, x]$.

(ii) $\Rightarrow$ (i) $C(F)$ est dense dans $\mathbb{R}$. Soit $\mathcal{A}$ la classe des intervalles $]a, b]$, avec $a < b$, et $a, b \in C(F)$. (ii) entraîne que

$$\mathbb{P}(X_n \in]a, b]) \to \mathbb{P}(X \in]a, b])$$

Soit G un ouvert de $\mathbb{R}$. G est une réunion dénombrable d'éléments disjoints de $\mathcal{A}$, $G = \bigcup_1^\infty A_m$.

$$\begin{aligned}\mathbb{P}\left(X \in \bigcup_{m \le M} A_m\right) &= \lim_n \mathbb{P}\left(X_n \in \bigcup_{m \le M} A_m\right) \\ &\le \liminf \mathbb{P}(X_n \in G)\end{aligned}$$

D'où en faisant tendre $M \to \infty$,

$$\mathbb{P}(X \in G) \le \liminf \mathbb{P}(X_n \in G)$$

Il suffit alors d'utiliser le Théorème 7.1.5 □

Un autre résultat en dimension 1 est le

Lemme 7.1.9. *(Fatou) Soit $\{X_n,\ n \ge 1\}$ et X des v. a. r., telles que $X_n \xrightarrow{\mathcal{L}} X$. Supposons que $X_n \ge 0$ p. s. pour tout $n \ge 0$. Alors $X \ge 0$ p. s. et*

$$\mathbb{E}(X) \le \liminf_n \mathbb{E}(X_n).$$

Preuve D'après le Théorème 7.1.5 (iii) avec $G =]-\infty, 0[$, $\mathbb{P}(X < 0) \le \liminf_n \mathbb{P}(X_n < 0) = 0$. En outre, si $f(x) = x^+ \wedge K$, $f \in C_b(\mathbb{R})$, donc pour tout $K > 0$,

$$\begin{aligned}\mathbb{E}(X \wedge K) &= \lim_n \mathbb{E}(X_n \wedge K) \\ &\le \liminf_n \mathbb{E}(X_n).\end{aligned}$$

Il reste à faire tendre $K \to \infty$ pour conclure. □

Définition 7.1.10. *Une suite $\{\mathbb{Q}_n, n \geq 1\}$ de probabilités sur $(\mathbb{R}^d, \mathcal{B}_d)$ est dite* **tendue** *si pour tout $\varepsilon > 0$, il existe un compact K de $\mathbb{R}^d$ t.q.*

$$\mathbb{Q}_n(K) \geq 1 - \varepsilon, \quad \textit{pour tout } n \geq 1.$$

Une suite $\{X_n, n \geq 1\}$ de v.a. de dimension d est dite **tendue** *si la famille $\{\mathbb{P}_{X_n}, n \geq 1\}$ est tendue, i.e. si pour tout $\varepsilon > 0$, il existe un compact K de $\mathbb{R}^d$ t.q.*

$$\mathbb{P}(X_n \in K) \geq 1 - \varepsilon, \quad \textit{pour tout } n \geq 1.$$

Si Q est une probabilité sur $(\mathbb{R}^d, \mathcal{B}_d)$, pour tout $\varepsilon > 0$, il existe un compact $K \subset \mathbb{R}^d$ t.q. $Q(K) \geq 1-\varepsilon$. Une réunion finie de compacts étant un compact, il en résulte que toute famille finie de lois de probabilité est tendue.

Théorème 7.1.11. *Si $X_n \xrightarrow{\mathcal{L}} X$, alors la famille $\{X_n, n \geq 1\}$ de vecteurs aléatoires de dimension d est tendue.*

Preuve Soit $\varepsilon > 0$, $M > 0$ t.q. $\mathbb{P}(|X| \geq M) \leq \varepsilon/2$. Alors, puisque $X_n \xrightarrow{\mathcal{L}} X$, d'après le Théorème 7.1.5 (ii),

$$\limsup_n \mathbb{P}(|X_n| \geq M) \leq \mathbb{P}(|X| \geq M) \leq \varepsilon/2.$$

Donc il existe n_0 tel que pour tout $n \geq n_0$,

$$\mathbb{P}(|X_n| \geq M) \leq \varepsilon.$$

D'après la remarque ci–dessus, il existe un compact K de $\mathbb{R}^d$, tel que

$$\mathbb{P}(X_n \in K) \geq 1 - \varepsilon, \forall\, n < n_0.$$

Posons $\bar{K} = K \cup \{x; |x| \leq M\}$. $\bar{K}$ est un compact, et

$$\mathbb{P}(X_n \in \bar{K}) \geq 1 - \varepsilon, \, \forall\, n \geq 1$$

□

On a la réciproque partielle du Théorème 7.1.11 :

Théorème 7.1.12. *Soit $\{X_n, n \geq 1\}$ une suite* **tendue** *de vecteurs aléatoires de dimension d. Alors il existe une sous–suite $\{X_{n_k},\ k \geq 1\}$ et un vecteur aléatoire X de dimension d, t.q. quand $k \to \infty$,*

$$X_{n_k} \xrightarrow{\mathcal{L}} X.$$

PREUVE On va se contenter de faire la preuve dans le cas $d = 1$. Pour tout $n \geq 1$, désignons par F_n la fonction de répartition de la v. a. r. X_n. Notons que pour tout $x \in \mathbb{R}$, la suite de nombres réels $\{F_n(x),\ n \geq 1\}$ est dans le compact $[0,1]$, donc contient une sous–suite convergente. Soit maintenant une suite $\{x_k,\ k \geq 1\}$ dénombrable dense dans $\mathbb{R}$. On choisit une sous–suite $\{n^1_\ell,\ \ell \geq 1\}$ telle que $F_{n^1_\ell}(x_1)$ converge dans $[0,1]$. On choisit une seconde sous–suite $\{n^2_\ell,\ \ell \geq 1\} \subset \{n^1_\ell,\ \ell \geq 1\}$ telle que $F_{n^2_\ell}(x_2)$ converge dans $[0,1]$, et ainsi de suite. Il résulte de ce choix de sous–suites emboîtées que pour tout $k \geq 1$, la suite $F_{n^\ell_\ell}(x_k)$ converge, vers une limite que nous appellerons $G(x_k)$. Notons que G hérite des fonctions F_n la propriété de monotonie suivante : si $x_j < x_k$, alors $0 \leq G(x_j) \leq G(x_k) \leq 1$.

Définissons la fonction $\overline{G} : \mathbb{R} \to [0,1]$ par $\overline{G}(x) = \inf\{G(x_k),\ x_k > x\}$. $\overline{G}$ (qui vérifie $G(x_k) \leq \overline{G}(x_k) \leq G(x_\ell)$ pour tout k, ℓ tels que $x_k < x_\ell$) est clairement croissante et continue à droite. Montrons maintenant que $F_{n^\ell_\ell}(x) \to \overline{G}(x)$, pour tout x point de continuité de $\overline{G}$. Soit x un tel point de continuité et i, j et k tels que $x_i < x_j < x < x_k$. On a

$$G(x_j) = \lim_\ell F_{n^\ell_\ell}(x_j) \leq \liminf_\ell F_{n^\ell_\ell}(x) \leq \limsup_\ell F_{n^\ell_\ell}(x) \leq \lim_\ell F_{n^\ell_\ell}(x_k) = G(x_k).$$

En outre

$$\overline{G}(x_i) \leq G(x_j) \leq G(x_k) \leq \overline{G}(x_k).$$

Donc

$$\overline{G}(x_i) \leq \liminf_\ell F_{n^\ell_\ell}(x) \leq \limsup_\ell F_{n^\ell_\ell}(x) \leq \overline{G}(x_k).$$

Il reste à faire tendre $x_i \uparrow x$ et $x_k \downarrow x$, en exploitant la continuité de $\overline{G}$ au point x.

Il reste à montrer que $\overline{G}$ est la fonction de répartition d'une probabilité sur $\mathbb{R}$, et pour cela il reste à vérifier que $\overline{G}(x) \to 0$ si $x \to -\infty$, et $\to 1$ si $x \to +\infty$. Ceci résulte de ce que $\overline{G}$ est croissante à valeurs dans $[0,1]$ et vérifie que pour tout $\varepsilon > 0$, il existe $x_\varepsilon < y_\varepsilon$ tels que $\overline{G}(y_\varepsilon) - \overline{G}(x_\varepsilon) \geq 1 - \varepsilon$. Cette dernière propriété résulte clairement de la tension de la suite $\{X_n\}$ et de la convergence de la sous–suite $\{F_{n^\ell_\ell}\}$ vers $\overline{G}$. □

Corollaire 7.1.13. *Soit $\mathcal{K} \subset C_b(\mathbb{R}^d;\mathbb{C})$ une classe séparante, et $\{X_n, n \geq 1\}$ une suite* **tendue** *de vecteurs aléatoires de dimension d.*

Alors il existe un vecteur aléatoire X de dimension d tel que $X_n \xrightarrow{\mathcal{L}} X$, si et seulement si pour tout $f \in \mathcal{K}$, la suite de nombres réels $\{\mathbb{E}[f(X_n)], n \geq 1\}$ converge (dans $\mathbb{R}$).

Preuve La C.N. est évidente. Démontrons la C.S.

D'après le Théorème 7.1.12, il existe une sous–suite $\{X_{n_k}\}$ et X t.q. $X_{n_k} \xrightarrow{\mathcal{L}} X$.

Supposons par l'absurde que la suite X_n toute entière ne converge pas en loi vers X. Alors il existe $g \in C_b(\mathbb{R}^d)$, $\varepsilon > 0$, et une sous–suite $\{X_{n_\ell};\ \ell \geq 1\}$ t.q. :

$$(*) \quad |\mathbb{E}[g(X_{n_\ell})] - \mathbb{E}[g(X)]| \geq \varepsilon, \;\text{ pour tout } \ell \geq 1.$$

A nouveau d'après le Théorème 7.1.12, on peut extraire de la suite $\{X_{n_\ell}\}$ une sous–suite $\{X_{n_{\ell'}}\}$ qui converge en loi vers un v.a. X'.

Mais il résulte de l'hypothèse de la C.S. que $\mathbb{E}[f(X)] = \mathbb{E}[f(X')]$, quelque soit $f \in \mathcal{K}$, ce qui entraîne que $\mathbb{P}_X = \mathbb{P}_{X'}$, donc $\mathbb{E}[g(X)] = \mathbb{E}[g(X')]$, pour tout $g \in C_b(\mathbb{R}^d)$, ce qui contredit $(*)$. □

7.2 Relation avec les autres types de convergence, et propriétés supplémentaires

Théorème 7.2.1. *Soit $\{X_n, n \geq 1\}$ et X des v.a. de dimension d, définis sur un même espace de probabilité $(\Omega, \mathcal{F}, P)$. Alors*

$$(i) \quad X_n \xrightarrow{p} X \Rightarrow X_n \xrightarrow{\mathcal{L}} X$$

$$(ii) \quad \left.\begin{array}{c} X_n \xrightarrow{\mathcal{L}} X \\ \mathbb{P}_X = \delta_x \end{array}\right\} \Rightarrow X_n \xrightarrow{p} X.$$

Remarque 7.2.2.

1. *Il est clair que l'on peut avoir $X_n \xrightarrow{\mathcal{L}} X$, avec les X_n définis sur des espaces de probabilité différents. Ce n'est pas le cas pour la convergence en probabilité.*
2. *En général, $\mathbb{P}_X$ ne détermine pas X (même à une classe d'équivalence près) i.e. il existe Y t.q. $\mathbb{P}(X \neq Y) > 0$, et $\mathbb{P}_X = \mathbb{P}_Y$. Il est alors clair que la convergence en loi ne peut pas entraîner la convergence en probabilité. Mais si $\mathbb{P}_X = \mathbb{P}_Y = \delta_x$, $X = Y$ p.s.*
3. *Il résulte de (i) que la convergence p.s., la convergence dans L^p et la convergence en probabilité, entraînent la convergence en loi.*

Preuve

(i) Si $X_n \xrightarrow{p} X$, $f \in C_b(\mathbb{R}^d)$, alors $f(X_n) \xrightarrow{p} f(X)$, et puisque f est bornée, $\mathbb{E}[f(X_n)] \to \mathbb{E}[f(X)]$ par le Corollaire 6.2.6.

(ii)

$$\begin{aligned}\mathbb{P}(|X_n - X| > \varepsilon) &\leq \mathbb{P}(|X_n - x| \geq \varepsilon)\\ \limsup \mathbb{P}(|X_n - X| > \varepsilon) &\leq \limsup \mathbb{P}(X_n \in B^c(x;\varepsilon))\\ &\leq \mathbb{P}(X \in B^c(x;\varepsilon)) = 0,\end{aligned}$$

où $B(x;\varepsilon) = \{y \in \mathbb{R}^d; |y - x| < \varepsilon\}$.

□

Contrairement aux autres types de convergence, la convergence en loi n'est pas "additive", au sens où

$$\left.\begin{aligned}X_n \xrightarrow{\mathcal{L}} X\\ Y_n \xrightarrow{\mathcal{L}} Y\end{aligned}\right\} \not\Rightarrow X_n + Y_n \xrightarrow{\mathcal{L}} X + Y$$

Cela vient de ce que la donnée de $\mathbb{P}_X$ et de $\mathbb{P}_Y$ ne détermine pas $\mathbb{P}_{X+Y}$. On a cependant des résultats lorsque les données déterminent $\mathbb{P}_{X+Y}$. En particulier, le résultat suivant est classique et très utile.

Théorème 7.2.3. *Soit $\{X_n, Y_n; n \geq 1\}$ et X des v.a. de dimension d, $y \in \mathbb{R}^d$. On suppose :*

$$\begin{aligned}X_n &\xrightarrow{\mathcal{L}} X\\ Y_n &\xrightarrow{\mathcal{L}} y\end{aligned}$$

Alors

(i) $X_n + Y_n \xrightarrow{\mathcal{L}} X + y$

(ii) **si** $d = 1$; $X_n Y_n \xrightarrow{\mathcal{L}} yX$

Le théorème 7.2.3 est en fait un corollaire des deux théorèmes qui suivent

Théorème 7.2.4. *Soit $\{X_n, : n \geq 1,\ X\}$ des vecteurs aléatoires de dimension d, $\{Y_n,\ n \geq 1\}$ des vecteurs aléatoires de dimension k, $y \in \mathbb{R}^k$. Si*

$$X_n \xrightarrow{\mathcal{L}} X, \quad Y_n \xrightarrow{\mathcal{L}} y$$

quand $n \to \infty$, alors

$$\begin{pmatrix}X_n\\ Y_n\end{pmatrix} \xrightarrow{\mathcal{L}} \begin{pmatrix}X\\ y\end{pmatrix}.$$

Preuve Il suffit de montrer que pour toute $\Phi \in C_b(\mathbb{R}^d \times \mathbb{R}^k)$,

$$\mathbb{E}[\Phi(X_n, Y_n)] \to \mathbb{E}[\Phi(X, y)].$$

Comme évidemment $\mathbb{E}[\Phi(X_n, y)] \to \mathbb{E}[\Phi(X, y)]$, il suffit de vérifier que

$$\mathbb{E}[\Phi(X_n, Y_n)] - \mathbb{E}[\Phi(X_n, y)] \to 0,$$

lorsque $n \to \infty$. Il est facile de vérifier que la tension de la suite $\{X_n,\ n \geq 1\}$ et celle de la suite $\{Y_n,\ n \geq 1\}$ entraînent celle de la suite $\{(X_n, Y_n),\ n \geq 1\}$ de vecteurs aléatoires de dimension $d + k$. Donc pour tout $\varepsilon > 0$, on peut trouver un compact $K \subset \mathbb{R}^{d+k}$ tel que $\mathbb{P}[(X_n, Y_n) \notin K] \leq \varepsilon$. Notons $c := \sup_{(x,y)\in\mathbb{R}^{d+k}} |\Phi(x, y)|$. Il résulte de l'uniforme continuité de Φ sur le compact K et de la convergence en probabilité de Y_n vers y que

$$(*) \quad [\Phi(X_n, Y_n) - \Phi(X_n, y)]\, \mathbf{1}_{\{(X_n,Y_n)\in K\}} \to 0$$

en probabilité quand $n \to \infty$. En effet, de l'uniforme continuité de Φ sur K, on déduit que pour tout $\varepsilon > 0$, il existe $\delta > 0$ tel que si $(x, y) \in K$, $(x, y') \in K$ et $|y - y'| \leq \delta$, alors $|\Phi(x, y) - \Phi(x, y')| \leq \varepsilon$. Donc

$$\mathbb{P}(|\Phi(X_n, Y_n) - \Phi(X_n, y)|\, \mathbf{1}_{\{(X_n,Y_n)\in K\}} > \varepsilon) \leq \mathbb{P}(|Y_n - y| > \delta) \to 0$$

quand $n \to \infty$. Comme en plus la v. a. r. à gauche de $(*)$ est bornée en valeur absolue par la constante $2c$,

$$\mathbb{E}\left(|\Phi(X_n, Y_n) - \Phi(X_n, y)|\, \mathbf{1}_{\{(X_n,Y_n)\in K\}}\right) \to 0.$$

Donc par le choix de K,

$$\limsup_{n\to\infty} |\mathbb{E}[\Phi(X_n, Y_n)] - \mathbb{E}[\Phi(X_n, y)]| \leq 2c\varepsilon.$$

Le résultat s'en déduit en faisant tendre $\varepsilon \to 0$. □

Théorème 7.2.5. *Soit $\{X_n, Y_n,\ n \geq 1\}$, X et Y des v.a. de dimension d. On suppose que $\begin{pmatrix} X_n \\ Y_n \end{pmatrix} \xrightarrow{\mathcal{L}} \begin{pmatrix} X \\ Y \end{pmatrix}$. Alors*

(i) $X_n + Y_n \xrightarrow{\mathcal{L}} X + Y$.

(ii) $X_n Y_n \xrightarrow{\mathcal{L}} XY$ si $d = 1$

Ce théorème est une conséquence de la proposition suivante, laquelle résulte directement des définitions.

Proposition 7.2.6. *Soit $\{X_n, n \geq 1\}$ et X des v.a. de dimension d, et $\varphi : \mathbb{R}^d \to \mathbb{R}^k$ une application continue. On suppose que $X_n \xrightarrow{\mathcal{L}} X$. Alors $\varphi(X_n) \xrightarrow{\mathcal{L}} \varphi(X)$.*

7.3 Convergence en loi et fonctions caractéristiques

D'après ce qui précède, si φ_n est la fonction caractéristique de X_n, et φ la fonction caractéristique de X, alors d'une part

$$X_n \xrightarrow{\mathcal{L}} X \Rightarrow \varphi_n(u) \to \varphi(u), \ \ \forall\, u \in \mathbb{R}^d,$$

et d'autre part par le Corollaire 7.1.13 et le Théorème 5.3.1

$$\left.\begin{array}{l} \varphi_n(u) \to \varphi(u), \forall\, u \in \mathbb{R} \\ \{X_n\} \text{ est tendue} \end{array}\right\} \Rightarrow X_n \xrightarrow{\mathcal{L}} X$$

Nous allons améliorer ces deux résultats. On va tout d'abord établir un lemme. Etant donné une fonction caractéristique φ d'une probabilité sur $\mathbb{R}$, on définit la fonction

$$T\varphi(u) = 1 - \frac{1}{2}\varphi(u) - \frac{1}{2}\varphi(-u) = 1 - \mathbb{E}[\cos(uX)].$$

Lemme 7.3.1. *Il existe une constante $\alpha > 0$ telle que pour toute v. a. r. X de fonction caractéristique φ, et pour tout $\theta > 0$,*

$$\mathbb{P}\left(|X| > 1/\theta\right) \leq \frac{\alpha}{\theta}\int_0^\theta T\varphi(u)du.$$

Preuve Par le théorème de Fubini,

$$\begin{aligned} \frac{1}{\theta}\int_0^\theta T\varphi(u)du &= \frac{1}{\theta}\mathbb{E}\int_0^\theta [1-\cos(uX)]du \\ &= \mathbb{E}\left(1 - \frac{\sin(\theta X)}{\theta X}\right) \\ &\geq \mathbb{E}\left[\left(1 - \frac{\sin(\theta X)}{\theta X}\right)\mathbf{1}_{\{|X\theta|>1\}}\right] \\ &\geq \inf_{|t|>1}\left(1 - \frac{\sin t}{t}\right)\mathbb{P}(|X| > 1/\theta). \end{aligned}$$

Le lemme est démontré, avec

$$\alpha = \left[\inf_{|t|>1}\left(1 - \frac{\sin t}{t}\right)\right]^{-1}$$

□

Théorème 7.3.2. *(P. Lévy) Soit $\{X_n, n \geq 1\}$ une suite de v.a. de dimension d. Notons φ_n la fonction caractéristique de X_n. On suppose :*

(i) $\lim\limits_{n\to\infty} \varphi_n(u)$ existe, pour tout $u \in \mathbb{R}^d$.

(ii) $\varphi(u) \stackrel{\triangle}{=} \lim_n \varphi_n(u)$ est continue en $u = 0$.

Alors il existe un v.a. X de dimension d tel que :

$$X_n \stackrel{\mathcal{L}}{\to} X,$$

et φ est la fonction caractéristique de X.

Preuve Il suffit, pour pouvoir appliquer le Corollaire 7.1.13 combiné au Théorème 5.3.1, de montrer que la suite $\{X_n, n \geq 1\}$ est tendue. Mais la tension de cette suite est équivalente à celle des d suites $\{X_n^i, n \geq 1\}$, $1 \leq i \leq d$, où X_n^i désigne la i–ème coordonnée du vecteur aléatoire X_n. Donc il suffit de démontrer que dans le cas $d = 1$, les conditions du théorème entraînent la tension de la suite $\{X_n, n \geq 1\}$.

Nous supposons donc dorénavant que $d = 1$. Posons $\mathbb{Q}_n = \mathbb{P}_{X_n}$. D'après le Lemme 7.3.1, pour tout $\theta > 0$, si l'on pose

$$S_\theta = \{x; |x| \leq 1/\theta\},$$

on a

$$\mathbb{Q}_n(S_\theta^c) \leq \frac{\alpha}{\theta} \int_0^\theta T\varphi_n(u)du.$$

D'après (i) et le Théorème de convergence dominée,

$$\limsup \mathbb{Q}_n(S_\theta^c) \leq \frac{\alpha}{\theta} \int_0^\theta T\varphi(u)du.$$

Mais grâce à (ii), $T\varphi(u)$ est continue en $u = 0$, et est nulle pour $u = 0$ par définition de T.

Donc

$$\frac{1}{\theta} \int_0^\theta T\varphi(u)du \to 0, \text{ quand } \theta \to 0.$$

Fixons $\varepsilon > 0$. Soit alors $a > 0$ tel que

$$\limsup_n \mathbb{Q}_n(S_a^c) \leq \varepsilon/2$$

Donc il existe n_0 t.q. pour tout $n \geq n_0$, $\mathbb{Q}_n(S_a^c) \leq \varepsilon$. Soit K un compact tel que $K \supset S_a$, et $\mathbb{Q}_n(K) \geq 1 - \varepsilon$, pour tout $n < n_0$. Alors $\mathbb{Q}_n(K) \geq 1 - \varepsilon$, pour tout $n \geq 1$. $\{\mathbb{Q}_n;\ n \geq 1\}$ est donc tendue. $\square$

Corollaire 7.3.3. *Soit $\{X_n, n \in \mathbb{N}\}$ une suite de vecteurs aléatoires de dimension d. On note φ_n la fonction caractéristique de X_n.*

Supposons qu'il existe un vecteur aléatoire X de dimension d, de fonction caractéristique φ, t.q.

$$\varphi_n(u) \to \varphi(u), \ \ \forall\, u \in \mathbb{R}^d$$

Alors

$$X_n \xrightarrow{\mathcal{L}} X.$$

□

Théorème 7.3.4. *Soit $\{X_n, n \geq 1\}$ et X des vecteurs aléatoires de dimension d, et $\{\varphi_n, n \geq 1\}$ et φ les fonctions caractéristiaues correspondantes.*

Supposons que $X_n \xrightarrow{\mathcal{L}} X$.

Alors $\varphi_n(u) \to \varphi(u)$ uniformément sur tout compact.

Preuve

a) Montrons tout d'abord que $X_n \xrightarrow{\mathcal{L}} X$ entraîne que les $\{\varphi_n, n \geq 0\}$ sont équicontinues (on pose $\varphi = \varphi_0$) i.e. $\sup\limits_{u \in \mathbb{R}^d} \sup\limits_{n \geq 0} |\varphi_n(u+h) - \varphi_n(u)| \to 0$, quand $h \to 0$.

Soit K un compact de $\mathbb{R}^d$. On pose $\mathbb{Q}_n = \mathbb{P}_{X_n}$.

$$\begin{aligned}|\varphi_n(u+h) - \varphi_n(u)| &\leq \int_K |e^{i(u+h,x)} - e^{i(u,x)}| \mathbb{Q}_n(dx) + 2\mathbb{Q}_n(K^c) \\ &\leq \sup_{x \in K} |e^{i(h,x)} - 1| + 2\mathbb{Q}_n(K^c)\end{aligned}$$

Fixons $\varepsilon > 0$. D'après le Théorème 7.1.11, il existe K_ε compact tel que $\mathbb{Q}_n(K_\varepsilon^c) \leq \varepsilon/4$. Alors il existe ρ_ε t.q. pour tout $|h| \leq \rho_\varepsilon$, $\sup\limits_{x \in K_\varepsilon} |e^{i(h,x)} - 1| \leq \varepsilon/2$. d'où

$$\sup_{u \in \mathbb{R}^d} \sup_{n \geq 0} |\varphi_n(u+h) - \varphi_n(u)| \leq \varepsilon, \ \ \forall\, |h| \leq \rho_\varepsilon.$$

b) Soit maintenant K un compact quelconque de $\mathbb{R}^d$. Soit $u_1, u_2, \ldots, u_m \in K$ tels que les boules de centre u_i et de rayon ρ_ε forment un recouvrement de K.

Etant donné $u \in K$, soit u_k t.q. $|u - u_k| \leq \rho_\varepsilon$.

$$
\begin{aligned}
|\varphi_n(u) - \varphi(u)| &\leq |\varphi_n(u) - \varphi_n(u_k)| + |\varphi_n(u_k) - \varphi(u_k)| + |\varphi(u_k) - \varphi(u)| \\
\sup_{u \in K} |\varphi_n(u) - \varphi(u)| &\leq 2\varepsilon + \sup_{k \leq m} |\varphi_n(u_k) - \varphi(u_k)|
\end{aligned}
$$

et $\sup_{k \leq m} |\varphi_n(u_k) - \varphi(u_k)| \to 0$, quand $n \to \infty$. Donc

$$\limsup_{n \to \infty} \sup_{u \in K} |\varphi_n(u) - \varphi(u)| \leq 2\varepsilon,$$

avec $\varepsilon > 0$ arbitraire. D'où le résultat.

□

7.4 Le Théorème Central Limite

Théorème 7.4.1. *Soit $\{X_n, n \geq 1\}$ une suite i.i.d. de v.a. de dimension d, dont la loi commune admet un moment d'ordre 2 fini. On note $\mu = E(X_1)$, $\Sigma = Cov(X_1)$.*

Alors

$$\frac{X_1 + \cdots + X_n - n\mu}{\sqrt{n}} \xrightarrow{\mathcal{L}} N(0, \Sigma).$$

Etablissons tout d'abord le :

Lemme 7.4.2. *Soit $z_1, \ldots, z_m$ et $z'_1 \ldots, z'_m$ des nombres complexes de module ≤ 1. Alors :*

$$|z_1 \times \cdots \times z_m - z'_1 \times \cdots \times z'_m| \leq \sum_{k=1}^{m} |z_k - z'_k|$$

Preuve

$$
\begin{aligned}
z_1 \cdots z_m - z'_1 \cdots z'_m &= (z_1 - z'_1) z_2 \cdots z_m + z'_1 (z_2 \cdots z_m - z'_2 \cdots z'_m) \\
|z_1 \cdots z_m - z'_1 \cdots z'_m| &\leq |z_1 - z'_1| + |z_2 \cdots z_m - z'_2 \cdots z'_m|,
\end{aligned}
$$

et on réutilise $m - 1$ fois le même argument.
Preuve du Théorème On suppose qu'on s'est ramené au cas $\mu = 0$. Posons

$$S_n = X_1 + \cdots + X_n$$

$$\varphi_{S_{n/\sqrt{n}}}(u) = \varphi_{S_n}\left(\frac{u}{\sqrt{n}}\right) = \varphi^n_{X_1}\left(\frac{u}{\sqrt{n}}\right).$$

Par hypothèse, $\mathbb{E}(|X_1|^2) < \infty$. Donc, d'après la Proposition 5.3.3

$$\varphi_{X_1}(u) = 1 - \frac{1}{2}\mathbb{E}[(u, X_1)^2] + \delta(u)|u[^2$$

avec $\delta(u) \to 0$, si $u \to 0$, et $|\delta(u)| \leq 2\mathbb{E}(|X_1|^2)$.

Donc

$(*)$ $\qquad \varphi_{X_1}\left(\frac{u}{\sqrt{n}}\right) = 1 - \frac{1}{2n}\mathbb{E}[(u, X_1)^2] + \delta\left(\frac{u}{\sqrt{n}}\right)\frac{|u|^2}{n}.$

Il résulte du Lemme 7.4.2 que pour $n > \mathbb{E}[(u, X_1)^2]/2$,

$$\left|\varphi^n_{X_1}\left(\frac{u}{\sqrt{n}}\right) - \left(1 - \frac{\mathbb{E}[(u, X_1)^2]}{2n}\right)^n\right| \leq n\left|\varphi_{X_{n_1}}\left(\frac{u}{\sqrt{n}}\right) - \left(1 - \frac{\mathbb{E}[(u, X_1)^2]}{2n}\right)\right|$$

Et d'après $(*)$ le membre de droite de cette inégalité tend vers 0, quand $n \to +\infty$. De plus :

$$\left(1 - \frac{t^2}{2n}\right)^n \to e^{-t^2/2}, \ \forall\, t \in \mathbb{R}$$

Donc

$$\varphi_{S_{n/\sqrt{n}}}(u) \to \exp\left[-\frac{1}{2}(\Sigma u, u)\right], \ \forall u \in \mathbb{R}^d.$$

Le Théorème découle alors du Corollaire 7.3.3. □

7.5 Tableaux triangulaires

Nous allons maintenant généraliser le résultat de la section précédente. Commençons par un Lemme technique qui nous sera utile.

Lemme 7.5.1. *Pour tout $n \geq 0$, x réel,*

$$\left|e^{ix} - \sum_{k=0}^{n}\frac{(ix)^k}{k!}\right| \leq \min\left\{\frac{|x|^{n+1}}{(n+1)!}, \frac{2|x|^n}{n!}\right\}.$$

Nous utiliserons essentiellement l'inégalité suivante, qui se déduit du lemme dans le cas $n = 2$:

(7.1) $$|e^{ix} - (1 + ix - x^2/2)| \leq \min\{|x|^3, x^2\}.$$

Preuve En écrivant l'accroissement de la fonction $s \to (x-s)^{n+1}e^{is}$ entre $s=0$ et $s=x$ comme l'intégrale de sa dérivée, on obtient l'identité

$$\int_0^x (x-s)^n e^{is} ds = \frac{x^{n+1}}{n+1} + \frac{i}{n+1}\int_0^x (x-s)^{n+1}e^{is}ds.$$

En partant de l'identité $e^{ix} = 1 + i\int_0^x e^{is}ds$, et en appliquant successivement la première identité avec $n=0$, puis avec $n=1$,..., on obtient un premier développement limité de la fonction e^{ix} :

$$e^{ix} = \sum_{k=0}^{n} \frac{(ix)^k}{k!} + \frac{i^{n+1}}{n!}\int_0^x (x-s)^n e^{is}ds.$$

On déduit aisément de la première identité avec n remplacé par $n-1$ que

$$\frac{i}{n}\int_0^x (x-s)^n e^{is}ds = \int_0^x (x-s)^{n-1}(e^{is}-1)ds.$$

Une combinaison des deux dernières identité fournit un second développement limité de la fonction e^{ix} :

$$e^{ix} = \sum_{k=0}^{n} \frac{(ix)^k}{k!} + \frac{i^n}{(n-1)!}\int_0^x (x-s)^{n-1}(e^{is}-1)ds.$$

Le résultat se déduit aisément en combinant les deux développements limités et des estimations élémentaires des deux termes de reste. □

Supposons maintenant que pour chaque $n \geq 1$, on se donne une suite

$$X_{n1}, \ldots, X_{nr_n}$$

de variables aléatoires réelles indépendantes. On a là ce que l'on appelle un tableau triangulaire. Les différentes suites pour des n différents peuvent être définies sur des espaces de probabilités différents. Dans le paragraphe précédent on avait $r_n = n$ et $X_{nk} = X_k$ pour tout n. Il s'agit donc ici d'une généralisation de la situation précédente (mais limitée à la dimension 1 pour simplifier les notations).

On suppose que

$$\mathbb{E}[X_{nk}] = 0, \quad \sigma_{nk}^2 = \mathbb{E}[X_{nk}^2], \quad s_n^2 = \sum_{k=1}^{r_n} \sigma_{nk}^2. \tag{7.2}$$

On supposera que $s_n > 0$, au moins pour n suffisamment grand. On suppose en outre satisfaite la *condition de Lindeberg* suivante : pour tout $\varepsilon > 0$,

$$\lim_{n\to\infty} \sum_{k=1}^{r_n} \frac{1}{s_n^2} \int_{|X_{nk}|\geq \varepsilon s_n} X_{nk}^2 d\mathbb{P} = 0. \tag{7.3}$$

On pose $S_n = X_{n1} + \cdots + X_{nr_n}$. On a le

Théorème 7.5.2. *Supposons que pour chaque $n \geq 1$ les v. a. $X_{n1}, \ldots, X_{nr_n}$ sont mutuellemenbt indépendantes, et qu'elles satisfont les conditions* (7.2) *et* (7.3)*. Alors, quand $n \to \infty$,*

$$\frac{S_n}{s_n} \xrightarrow{\mathcal{L}} N(0,1).$$

Dans la situation de la section précédente, puisque les X_n sont i.i.d., la condition de Lindeberg se ramène à

$$\lim_{n\to\infty} \frac{1}{\sigma} \int_{|X_1|\geq \varepsilon\sigma\sqrt{n}} X_1^2 dP = 0,$$

ce qui est immédiat par convergence dominée, puisque $\mathbb{E}[X_1^2] < \infty$ et $\{|X_1| \geq \varepsilon\sigma\sqrt{n}\} \downarrow \emptyset$, as $n \to \infty$. Donc le théorème 7.5.2 est bien une généralisation du TCL.

Preuve Quite à templacer X_{nk} par X_{nk}/s_n, on peut supposer que $s_n = 1$, c'est à dire

$$\sum_{k=1}^{r_n} \sigma_{nk}^2 = 1.$$

Désignons par φ_{nk} la fonction caractéristique de X_{nk}. En combinant l'inégalité élémentaire

$$|\varphi_{nk}(u) - (1 - u^2\sigma_{nk}^2/2)| \leq \mathbb{E}\left[\left|e^{iuX_{nk}} - (1 + iuX_{nk} - u^2X_{nk}^2/2)\right|\right]$$

avec (7.1), on obtient

$$|\varphi_{nk}(u) - (1 - u^2\sigma_{nk}^2/2)| \leq \mathbb{E}\left[\min\{|uX_{nk}|^2, |uX_{nk}|^3\}\right].$$

L'espérance dans le membre de droite est finie, puisque l'un des deux termes du min est intégrable. En fait pour tout $\varepsilon > 0$, ce membre de droite est majoré par

$$\int_{|X_{nk}|<\varepsilon} |uX_{nk}|^3 d\mathbb{P} + \int_{|X_{nk}|\geq\varepsilon} |uX_{nk}|^2 d\mathbb{P} \leq \varepsilon|u|^3\sigma_{nk}^2 + u^2 \int_{|X_{nk}\geq\varepsilon} X_{nk}^2 d\mathbb{P}.$$

Puisque $\sum_{k=1}^{r_n} \sigma_{nk}^2 = 1$ et ε est arbitraire, il résulte des calculs qui précèdent et de la condition de Lindeberg (7.3) que pour tout réel fixé u, lorsque $n \to \infty$,

$$\sum_{k=1}^{r_n} |\varphi_{nk}(u) - (1 - u^2\sigma_{nk}^2/2)| \to 0. \tag{7.4}$$

Notons que

$$\sigma_{nk}^2 \leq \varepsilon^2 + \int_{|X_{nk} \geq \varepsilon} X_{nk}^2 d\mathbb{P},$$

et donc il résulte de la condition de Lindeberg (7.3) (rappelons que $s_n = 1$) que lorsque $n \to \infty$,

$$\max_{1 \leq k \leq r_n} \sigma_{nk}^2 \to 0. \tag{7.5}$$

Donc pour u fixé et n suffisamment grand, $0 \leq 1 - u^2\sigma_{nk}^2/2 \leq 1$ pour tout $1 \leq k \leq r_n$. Il résulte donc de (7.4) et du Lemme 7.4.2 que lorsque $n \to \infty$,

$$\left| \prod_{k=1}^{r_n} \varphi_{nk}(u) - \prod_{k=1}^{r_n} (1 - u^2\sigma_{nk}^2/2) \right| \to 0. \tag{7.6}$$

A nouveau grâce au Lemme 7.4.2,

$$\begin{aligned} \left| \prod_{k=1}^{r_n} e^{-u^2\sigma_{nk}^2/2} - \prod_{k=1}^{r_n} (1 - u^2\sigma_{nk}^2/2) \right| &\leq \sum_{k=1}^{r_n} \left| e^{-u^2\sigma_{nk}^2/2} - 1 + u^2\sigma_{nk}^2/2 \right| \\ &\leq u^4 e^{u^2} \sum_{k=1}^{r_n} \sigma_{nk}^4 \\ &\leq u^4 e^{u^2} \max_{1 \leq k \leq r_n} \sigma_{nk}^2 \\ &\to 0, \end{aligned} \tag{7.7}$$

quand $n \to \infty$, où on a utilisé à la seconde inégalité le fait que pour tout réel z,

$$|e^z - 1 - z| \leq z^2 \sum_{k=2}^{\infty} \frac{|z|^{k-2}}{k!} \leq z^2 e^{|z|},$$

et (7.5) à la dernière ligne.

Le résultat découle maintenant des convergences (7.6), (7.7) et du Théorème de Paul Lévy (Corollaire 7.3.3). □

On a aussi le

Théorème 7.5.3. *Supposons que pour chaque $n \geq 1$ les v. a. $X_{n1}, \ldots, X_{nr_n}$ sont mutuellemenbt indépendantes, et qu'elles satisfont* (7.2) *et la condition de Lyapounov*

$$\lim_n \sum_{k=1}^{r_n} \frac{1}{s_n^{2+\delta}} \mathbb{E}\left[|X_{nk}^{2+\delta}\right] = 0,$$

pour un certain $\delta > 0$. Alors, quand $n \to \infty$,

$$\frac{S_n}{s_n} \xrightarrow{\mathcal{L}} N(0,1).$$

Preuve Il suffit de montrer que la condition de Lyapounov entraîne celle de Lindeberg. En effet

$$\begin{aligned}\sum_{k=1}^{r_n} \frac{1}{s_n^2} \int_{|X_{nk} \geq \varepsilon s_n} X_{nk}^2 d\mathbb{P} &\leq \sum_{k=1}^{r_n} \frac{1}{s_n^2} \int_{|X_{nk} \geq \varepsilon s_n} \frac{X_{nk}^{2+\delta}}{\varepsilon^\delta s_n^\delta} d\mathbb{P} \\ &\leq \varepsilon^{-\delta} \sum_{k=1}^{r_n} \frac{1}{s_n^{2+\delta}} \mathbb{E}\left[|X_{nk}^{2+\delta}\right].\end{aligned}$$

□

On a aussi le

Corollaire 7.5.4. *Soit $X_1, X_2, \ldots$ des v.a. indépendantes uniformément bornées en valeur absolue par une constante K et centrées. Si $s_n^2 = Var(X_1 + \cdots + X_n) \to \infty$ quand $n \to \infty$, alors*

$$\frac{X_1 + \cdots + X_n}{s_n} \xrightarrow{\mathcal{L}} N(0,1).$$

Preuve Il suffit de montrer que la condition de Lyapounov est vérifiée avec $\delta = 1$, $r_n = n$, $X_{nk} = X_k$. Or

$$\sum_{k=1}^{n} \frac{1}{s_n^3} \mathbb{E}\left[|X_k|^3\right] \leq K \sum_{k=1}^{n} \frac{E[X_k^2]}{s_n^3} = \frac{K}{s_n} \to 0,$$

quand $n \to \infty$. □

7.6 Test du χ^2

Le test du χ^2 est un test statistique qui est universellement utilisé. Il est basé sur une application du théorème de la limite centrale comme nous allons le voir. On considère un échantillon de taille n : $X = (X_1, \dots, X_n)$ d'une loi sur un ensemble fini $E = \{x_1, \dots, x_d\}$. C'est à dire que la suite $X_1, \dots, X_n$ est i. i. d. On veut tester si l'on peut admettre oui ou non que la loi commune des X_k est la loi de probabilité $\mathbb{Q}$ sur E, caractérisée par $\mathbb{Q}(\{x_j\}) = p_j > 0$, $1 \le j \le d$. Aux observations $X_1, \dots, X_n$ on associe les fréquences empiriques N_j/n, où

$$N_j = \sum_{k=1}^{n} \mathbf{1}_{\{X_k = x_j\}}, \ 1 \le j \le d.$$

K. Pearson a introduit la statistique

$$T_n = \sum_{j=1}^{d} \frac{1}{np_j} (N_j - np_j)^2 = \sum_{j=1}^{d} \frac{n}{p_j} \left(\frac{N_j}{n} - p_j \right)^2,$$

dont on s'attend à ce qu'elle prenne des valeurs nettement différentes, suivant que la loi commune des X_k est la loi $\mathbb{Q}$, ou bien une autre loi. C'est en tout cas vrai pour l'espérance de T_n. Notons que N_j suit la loi binômilae $B(n, p_j)$ (resp. $B(n, p'_j)$) si la loi commune des X_k est la loi $\mathbb{Q}$ (resp. la loi $\mathbb{Q}'$ caractérisée par $\mathbb{Q}'(\{x_j\}) = p'_j$, $1 \le j \le d$). Donc

$$\begin{aligned}
\mathbb{E}_{\mathbb{Q}}[T_n] &= \sum_{j=1}^{d} (1 - p_j) = d - 1, \\
\mathbb{E}_{\mathbb{Q}'}[T_n] &= \sum j = 1^d \frac{1}{np_j} \left[n^2 (p'_j - p_j)^2 + np'_j (1 - p'_j) \right] \\
&\ge n \sum_{j=1}^{d} \frac{(p'_j - p_j)^2}{p_j}.
\end{aligned}$$

Bien sûr, le teste du χ^2 est utilisé dans des cas où $n >> p$.

Le test du χ^2 consiste à rejeter l'hypothèse que la loi des X_k est $\mathbb{Q}$ ssi $T_n \ge a$. Il faut alors choisir a tel que la probabilité de rejeter à tort l'hypothse̊ que la loi des X_k est $\mathbb{Q}$ soit petite (typiquement égale à 0,95 ou 0,99). Pour cela il faudrait connaître la loi de la v. a. T_n. En fait on connaît seulement sa limite en loi quand $n \to \infty$ sous $\mathbb{Q}$ (c'est à dire lorsque la loi commune des X_k est la loi $\mathbb{Q}$).

Proposition 7.6.1. *Sous $\mathbb{Q}$, la loi de la v. a. T_n converge, quand $n \to \infty$, vers la loi χ^2_{d-1}.*

Preuve Soit $e_1, \ldots, e_d$ une base orthonormée de $\mathbb{R}^d$. On définit les v. a. $(Y_1, \ldots, Y_n)$ de dimension d par

$$Y_k = e_j \text{ ssi } X_k = x_j,\ 1 \leq j \leq d,\ 1 \leq k \leq n.$$

Alors les Y_k sont i. i. d., avec

$$\mathbb{E}[Y_1] = \begin{pmatrix} p_1 \\ \vdots \\ p_d \end{pmatrix}, \quad (\mathrm{Cov}\,(Y_1))_{ij} = p_i \delta_{ij} - p_i p_j.$$

Or

$$\sum_{k=1}^n Y_k = \begin{pmatrix} N_1 \\ \vdots \\ N_d \end{pmatrix}, \quad Z_n = \begin{pmatrix} \frac{N_1 - np_1}{\sqrt{n}} \\ \vdots \\ \frac{N_d - np_d}{\sqrt{n}} \end{pmatrix} = \frac{1}{\sqrt{n}} \left(\sum_{k=1}^n Y_k - n\mathbb{E}(Y_1) \right),$$

$$\text{donc } Z_n \xrightarrow{\mathcal{L}} N(0, \mathrm{Cov}\,(Y_1)).$$

Mais $T_n = f(Z_n)$, avec $f \in C(\mathbb{R}^d)$ défionie par $f(z) = \sum_{j=1}^d z_j^2 / p_j$. Donc

$$T_n \xrightarrow{\mathcal{L}} f(V), \text{ où } V \simeq N(0, \mathrm{Cov}\,(Y_1)).$$

Soit S une matrice orthogonale, telle que

$$S \begin{pmatrix} \sqrt{p_1} \\ \vdots \\ \sqrt{p_d} \end{pmatrix} = e_d,$$

ce qui revient à supposer que la dernière ligne de la matrice S est la ligne $(\sqrt{p_1}, \ldots, \sqrt{p_d})$, et que les autres lignes sont choisies de telle sorte que la collection de ces vecteurs ligne cosntitue une base orthonormée de $\mathbb{R}^d$. On pose

$$\xi = S \begin{pmatrix} V_1/\sqrt{p_1} \\ \vdots \\ V_d/\sqrt{p_d} \end{pmatrix}.$$

Alors ξ est un vecteur gaussien centré de matrice de covariance

$$S(\delta_{ij} - \sqrt{p_i p_j})S' = \begin{pmatrix} 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \\ 0 & 0 & \dots & 0 & 0 \end{pmatrix},$$

autrement dit $\xi_1, \dots, \xi_{d-1}$ sont i. i. d. de loi $N(0,1)$, et $\xi_d = 0$. Le théorème résulte de ce que

$$f(V) = \sum_{j=1}^{d} V_j^2/p_j = \sum_{j=1}^{d} \xi_j^2 = \xi_1^2 + \dots + \xi_{d-1}^2.$$

$\square$

7.7 Méthode de Monte Carlo

La méthode de Monte Carlo est une méthode numérique qui sert à calculer des valeurs approchées d'espérances. Supposons que l'on cherche à calculer la quantité $\mathbb{E}[X]$, où évidemment la v. a. r. X est supposée intégrable. On considère un cas où on ne connaît pas la valeur exacte de cette espérance, mais on sait *simuler sur ordinateur* des réalisations de la v. a. r. X. Par exemple, X peut être de la forme $X = f(Y)$, où la loi de Y est une loi facile à simuler, et où la fonction f, quoique compliquée, est soit connue analytiquement, soit calculable avec une grande précision sur ordinateur.

Expliquons ce que nous entendons par *simuler sur ordinateur* des réalisations d'une v. a.. Un programme informatique calcule des quantités tout à fait déterministes. Cependant, les spécialistes de systèmes dynamiques ont mis au point des algorithmes qui fabriquent des suites de nombres qui *ressemblent étrangement* à des suites de v. a. i. i. d., de loi commune la loi $\mathcal{U}(0,1)$, au sens où des tests statistiques perfectionnés ne font guère la différence. Il existe par ailleurs de nombreux algorithmes qui construisent des suites de v. a. de loi connue (Bernoulli, binômiale, Poisson, géométrique, exponentielle, gaussienne, etc...) à partir d'une suite i. i. d. de loi $\mathcal{U}([0,1])$.

Le principe de la méthode de Monte Carlo est le suivant. Puisque la loi forte des grands nombres nous dit que $n^{-1}(X_1 + \dots + X_n) \to \mathbb{E}[X]$ p. s., du moment que les $(X_i,\ i \geq 1)$ sont i. i. d. de loi celle de X, on simule un *grand* nombre de v. a. $X_1, \ldots, X_n$ i. i. d. de loi celle de X, et on produit comme valeur approchée de $\mathbb{E}[X]$ la quantité

$$n^{-1}(X_1 + \cdots + X_n).$$

La question importante est bien sûr d'avoir une idée de l'erreur

$$\varepsilon_n = \mathbb{E}[X] - n^{-1}(X_1 + \cdots + X_n$$

commise. La force de la méthode de Monte Carlo est précisément, grâce au théorème central limite, on a une idée assez précise de cette erreur, du moins si $\mathbb{E}[X^2] < \infty$, ce que nous supposerons dorénavant. Le TCL nous dit en effet que, sous cette hypothèse,

$$\sqrt{n}\ \varepsilon_n \xrightarrow{\mathcal{L}} \sigma Z,$$

où $\sigma^2 = \mathrm{Var}\,(X)$ et Z suit la loi $N(0,1)$. Compte tenu de la vitesse de convergence dans le TCL et des valeurs de n couramment utilisées dans la méthode de Monte Carlo, on pourra admettre que l'erreur ε_n est quasiment égale à $n^{-1/2}\sigma$ que multiplie une v. a. de loi $N(0,1)$, c'est à dire que

$$\mathbb{P}\left(\frac{X_1 + \cdots + X_n}{n} - 1,96\frac{\sigma}{\sqrt{n}} \leq \mathbb{E}[X] \leq \frac{X_1 + \cdots + X_n}{n} + 1,96\frac{\sigma}{\sqrt{n}}\right) \simeq 0,95,$$

$$\mathbb{P}\left(\frac{X_1 + \cdots + X_n}{n} - 2,6\frac{\sigma}{\sqrt{n}} \leq \mathbb{E}[X] \leq \frac{X_1 + \cdots + X_n}{n} + 2,6\frac{\sigma}{\sqrt{n}}\right) \simeq 0,95,$$

ce qui donne une idée assez précise de l'erreur commise, *à condition de connaître* σ ! Evidemment, en pratique ce n'est pas le cas. L'espérance de X étant inconnue, on s'attend à ce que sa variance soit a fortiori inconnue. Mais les simulations que l'on a affectuées pour approximer l'espérance nous permettent de calculer l'approximation suivante de la variance :

$$\sigma_n^2 = n^{-1}(X_1^2 + \cdots + X_n^2) - \left[n^{-1}(X_1 + \cdots + X_n)\right]^2.$$

Le fait que $\sigma_n^2 \to \sigma^2$ résulte à nouveau de la loi forte des grands nombres. Bien sûr, on ne contrôle pas la qualité de cette dernière approximation. Mais en pratique, on veut l'ordre de grandeur de l'erreur, et donc un ordre de grandeur de σ^2 suffit, et il est fourni par σ_n^2.

Remarque 7.7.1. *Concernant la vitesse de convergence, $n^{-1/2}$ est très lent, par rapport aux méthodes classiques d'analyse numérique. Mais le contrôle de la vitesse de convergence dans les méthodes numériques classiques de calcul d'intégrale supposent une certaine régularité de la fonction que l'on intègre, ce qui n'est pas du tout nécessaire ici. En outre, les méthodes d'analyse numérique classique s'effondrent en grande dimension, ce qui n'est pas le cas de la méthode de Monet Carlo (du moins pas dans les mêmes proportions).*

La méthode de Monte Carlo est surtout utilisée pour les problèmes pour lesquels les méthodes d'analyse numérique classique ne marchent pas bien. La vitesse des ordinateurs actuels fait que la lenteur de la convergence n'est pas forcément un inconvénient majeur. La méthode de Monte Carlo est surtout appréciée pour sa simplicité de programmation.

La possibilité de se faire assez facilement une idée de l'erreur du calcul est un atout important.

L'analyse faite ci–dessus de l'erreur nous indique que plus la variance de X est petite, plus l'algorithme converge vite. Donc à chaque fois que c'est possible, on a intérêt à se débrouiller pour se ramener à simuler des v. a. de variance la plus petite possible. Il existe de nombreuses méthodes dites "de réduction de variance" pour accélérer la méthode de Monte Carlo. Nous allons en présenter deux.

7.7.1 Réduction de variance : variables antithétiques

Cette méthode s'applique à chaque fois que l'on veut calculer $\mathbb{E}[X]$, avec $X = f(Y)$, où la fonction $f : \mathbb{R} \to \mathbb{R}$ est monotone, soit croissante soit décroissante, et Y est soit une v. a. de loi $U([0,1])$, donc telle que Y et $1-Y$ ont même loi, soit telle que Y et $-Y$ ont même loi (c'est le cas en particulier si Y est gaussienne centrée, ou exponentielle symétrique). Nous allons nous placer dans le second cas, donc nous supposerons dans la suite que Y et $-Y$ ont même loi. La méthode des variables antithétiques est basée sur le résultat suivant.

Lemme 7.7.2. *Si f est monotone (croissante ou décroissante), non p. p. constante, et $Var(Y) > 0$, alors*

$$Cov\ (f(Y), f(-Y)) < 0.$$

Preuve Soit Z une v. a. indépendante de Y, de même loi que Y. Alors

$$\begin{gathered}[f(Y)-f(Z)][f(-Y)-f(-Z)] \leq 0, \\ \mathbb{P}\left([f(Y)-f(Z)][f(-Y)-f(-Z)]<0\right)>0.\end{gathered}$$

Donc

$$\mathbb{E}\left([f(Y)-f(Z)][f(-Y)-f(-Z)]\right)<0.$$

Le résultat résulte de l'identité suivante, dont la vérification est laissée au lecteur :

$$\mathbb{E}\left([f(Y)-f(Z)][f(-Y)-f(-Z)]\right)=2\mathrm{Cov}\left(f(Y), f(-Y)\right).$$

□

La méthode des variables antithétique est basée sur l'algorithme suivant. On calcule

$$\frac{X_1+X_2+\cdots+X_n}{n},$$

mais au lieu de prendre $X_k=f(Y_k)$, on choisit

$$X_k=\frac{f(Y_k)+f(-Y_k)}{2},$$

pour $1 \leq k \leq n$, où les Y_k sont i. i. d. de loi celle de Y. L'intérêt de cette variante vient de ce que

$$\begin{aligned}\mathrm{Var}\left(\frac{f(Y_k)+f(-Y_k)}{2}\right) &= \frac{\mathrm{Var}\left(f(Y_k)\right)}{4}+\frac{\mathrm{Cov}\left(f(Y_k), f(-Y_k)\right)}{2}+\frac{\mathrm{Var}\left(f(Y_k)\right)}{4} \\ &< \frac{\mathrm{Var}\left(f(Y_k)\right)}{2}.\end{aligned}$$

Donc on fait mieux que diviser la variance par deux, c'est à dire que l'on fait mieux que si l'on multipliait le nombre de simulations par deux. Autrement dit, on peut s'attendre à ce que l'erreur dans la méthode modifiée comme ci dessus au bout de n simulations soit inférieure à celle de la méthode standard au bout de $2n$ simulations.

Remarquons que dans le cas où les Y_k sont de loi $U([0,1])$, on choisit

$$X_k=\frac{f(Y_k)+f(1-Y_k)}{2},$$

et on a le même résultat que ci–dessus.

7.7.2 Réduction de variance : variables de contrôle

Cette méthode s'applique à chaque fois que $X = f(Y)$ avec $f = g + h$, tels que l'on ait une formule explicite pour $\mathbb{E}[h(Y)]$, et que $\operatorname{Var}(g(Y)) < \operatorname{Var}|(f(Y))$. Alors on approche $\mathbb{E}[X]$ par

$$\frac{g(Y_1) + \cdots + g(Y_n)}{n} + \mathbb{E}[h(Y)].$$

Décrivons un exemple d'application de cette méthode, à savoir le calcul du prix d'une "option panier européenne" en mathématique financière. Dans cet exemple, on a besoin de calculer $\mathbb{E}[f(Y)]$, où Y et un vecteur aléatoire de dimension d tel que pour $1 \leq k \leq d$, $Y_k = \exp(Z_k)$, le vecteur $Z = (Z_1, \ldots, Z_d)$ étant un v. a. gaussien, et $f(y) = \left(\sum_{k=1}^d a_k y_k - K\right)^+$, avec K et les a_k des réels positifs. Posons $g(y) = \left(K - \sum_{k=1}^d a_k y_k\right)^+$. Alors puisque pour tout nombre réel z, $z = z^+ - z^- = z^+ - (-z)^+$, $z^+ = z + (-z)^+$, soit

$$\left(\sum_{k=1}^d a_k y_k - K\right)^+ = \left(K - \sum_{k=1}^d a_k y_k\right)^+ + \sum_{k=1}^d a_k y_k - K,$$

soit $f(y) = g(y) + h(y)$, avec $h(y) = \sum_{k=1}^d a_k y_k - K$. Mais d'une part

$$\mathbb{E}[h(Y)] = \sum_{k=1}^d a_k \mathbb{E}[Y_k] - K,$$

qui se calcule explicitement, puisque l'on sait calculer l'espérance de l'exponentielle d'une v. a. r. gaussienne, et d'autre part puisque $g(y) \leq K$, $\operatorname{Var}[g(Y)] \leq K^2$, alors que dans la plupart des cas intéressants, $\operatorname{Var}[f(Y)]$ est nettement plus grande. En fait celle-ci est bien souvent du même ordre de grandeur que $\operatorname{Var}[\sum_{k=1}^d a_k Y_k]$, que l'on peut calculer explicitement.

7.8 Intégrabilité uniforme

Soit X une v. a. r. p. s. finie. Alors

$$|X| \mathbf{1}_{|X| \geq K} \to 0 \quad \text{p. s. quand } K \to \infty.$$

Si en outre X est intégrable, alors par convergence dominée,

$$\mathbb{E}\left(|X|\mathbf{1}_{|X|\geq K}\right) \to 0, \quad \text{quand } K \to \infty.$$

On pose la

Définition 7.8.1. *La suite de v. a. r. $\{X_n,\ n \geq 1\}$ est dite uniformément intégrable si*

$$\lim_{K\to\infty} \sup_{n\geq 1} \mathbb{E}\left(|X_n|\mathbf{1}_{|X_n|\geq K}\right) = 0.$$

Bien sûr, une suite uniformément intégrable est formée de v. a. intégrables, puisque il existe K' t. q.

$$\sup_{n\geq 1} \mathbb{E}\left(|X_n|\mathbf{1}_{|X_n|\geq K'}\right) \leq 1, \quad \text{donc } \mathbb{E}|X_n| \leq 1 + K'.$$

On a le

Théorème 7.8.2. *Soit X_n et X des v. a. r. telles que $X_n \xrightarrow{\mathcal{L}} X$. Alors*

1. *Si la suite $\{X_n,\ n \geq 1\}$ est uniformément intégrable, alors X est intégrable et*

$$\mathbb{E}(X_n) \to \mathbb{E}(X), \text{ quand } n \to \infty. \tag{7.1}$$

2. *Si $X_n \geq 0$ p. s. pour tout $n \geq 1$, les X_n et X sont intégrables, et* (7.1) *a lieu, alors la suite $\{X_n,\ n \geq 1\}$ est uniformément intégrable.*

Preuve de *1.* L'uniforme intégrabilité entraîne l'existence de K tel que $\mathbb{E}|X_n| \leq 1 + K$, donc par Fatou (Lemme 7.1.9) l'intégrabilité de X. Pour tout $K > 0$, soit $\varphi_K \in C_b(\mathbb{R})$ telle que

$$\varphi_K(x) = \begin{cases} x, & \text{si } |x| \leq K; \\ 0, & \text{si } |x| \geq K+1. \end{cases},$$

et en outre

$$|x - \varphi_K(x)| \leq |x|\mathbf{1}_{\{|x|>K\}}, \ \forall x \in \mathbb{R}.$$

On a alors

$$\begin{gathered} \mathbb{E}\varphi_K(X_n) \to \mathbb{E}\varphi_K(X), \\ |\mathbb{E}X_n - \mathbb{E}\varphi_K(X_n)| \leq \mathbb{E}\left(|X_n|\mathbf{1}_{\{|X_n|>K\}}\right), \\ |\mathbb{E}X - \mathbb{E}\varphi_K(X)| \leq \mathbb{E}\left(|X|\mathbf{1}_{\{|X|>K\}}\right), \end{gathered}$$

Donc

$$\limsup_n |\mathbb{E}X - \mathbb{E}X_n| \leq \sup_n \mathbb{E}\left(|X_n|\mathbf{1}_{\{|X_n|>K\}}\right) + \mathbb{E}\left(|X|\mathbf{1}_{\{|X|>K\}}\right),$$

et le membre de droite tend vers 0 quand $K \to \infty$. Donc le membre de gauche est nul.
Preuve de *2*. Il n'est pas difficile de montrer que la convergence en loi entraîne que pour tout K tel que $\mathbb{P}(X = K) = 0$, quand $n \to \infty$,

$$\mathbb{E}\left(X_n\mathbf{1}_{\{X_n \leq K\}}\right) \to \mathbb{E}\left(X\mathbf{1}_{\{X \leq K\}}\right).$$

Donc si $\mathbb{E}X_n \to \mathbb{E}X$, par différence,

$$\mathbb{E}\left(X_n\mathbf{1}_{\{X_n > K\}}\right) \to \mathbb{E}\left(X\mathbf{1}_{\{X > K\}}\right).$$

Puisque X est intégrable, la limite ci–dessus peut être rendue plus petite que $\varepsilon/2$, en choisissant K assez grand. Donc il existe K et n_0 tels que

$$\mathbb{E}\left(X_n\mathbf{1}_{\{X_n > K\}}\right) \leq \varepsilon, \ \forall n \geq n_0.$$

Puisque tous les X_n sont intégrables, la même chose est vraie pour les X_n, $1 \leq n \leq n_0$ à condition peut–être de prendre K un peu plus grand. □

Dans l'énoncé qui suit, $\|X\|_1$ désigne la norme de la v. a. r. X dans $L^1(\Omega)$, i. e. $\|X\|_1 = \mathbb{E}(|X|)$.

Corollaire 7.8.3. *Si X_n et X des v. a. r. intégrables, et $X_n \xrightarrow{\mathcal{L}} X$, alors il y a équivalence entre les deux affirmations*

1. *La suite $\{X_n,\ n \geq 1\}$ est uniformément intégrable.*
2. *$\|X_n\|_1 \to \|X\|_1$, quand $n \to \infty$.*

Si en outre $X_n \to X$ en probabilité, chacune des deux conditions ci–dessus est équivalente à

$$X_n \to X \quad \textit{dans } L^1(\Omega).$$

Preuve Le Théorème, appliqué à la suite de v. a. $\{|X_n|,\ n \geq 1\}$, implique la première partie du Corollaire. Si $X_n \to X$ en probabilité, alors $|X - X_n| \to 0$ en probabilité, et l'uniforme intégrabilité de la suite $\{X_n\}$ entraîne clairement celle de la suite $\{|X - X_n|\}$, donc le Théorème entraîne que $\mathbb{E}|X - X_n| \to 0$. Réciproquement, si $X_n \to X$ dans $L^1(\Omega)$, $\|X_n\|_1 \to \|X\|_1$. □

Concluons ce chapitre en donnant des critères de tension et d'uniforme intégrabilité.

Proposition 7.8.4. *Une suite $\{X_n,\ n \geq 1\}$ de v. a. r. est* **tendue** *si et seulement si il existe une application $G : \mathbb{R}_+ \to \mathbb{R}_+$ telle que $G(x) \to \infty$, quand $x \to \infty$, et*

$$\sup_n \mathbb{E}G(|X_n|) < \infty.$$

Une suite $\{X_n,\ n \geq 1\}$ de v. a. r. est **uniformément intégrable** *si et seulement si il existe une application $H : \mathbb{R}_+ \to \mathbb{R}_+$ telle que $H(x)/x \to \infty$, quand $x \to \infty$, et*

$$\sup_n \mathbb{E}H(|X_n|) < \infty.$$

En outre, la fonction G peut être choisie continue croissante, et la fonction H peut être choisie croissante et convexe.

Preuve Le fait que la première (resp. seconde) condition est une condition suffisante de tension (resp. d'uniforme intégrabilité) est facile à vérifier. Etant donné une suite $\{k_n,\ n \geq 1\}$ de réels positifs tendant vers $+\infty$, posons

$$\begin{aligned} G(r) &= \sum_{n=0}^{\infty} \left(n + \frac{r - k_n}{k_{n+1} - k_n} \right) \mathbf{1}_{[k_n, k_{n+1}[}(r), \\ H(r) &= \int_0^r G(s)ds. \end{aligned}$$

G et H sont continues croissantes et H est convexe. En outre $G(x) \to \infty$, quand $x \to \infty$, et $H(x)/x \to \infty$, quand $x \to \infty$.

Soit $\{X_n\}$ une suite tendue. On choisit $k_0 = 0$ et pour $n \geq 1$, k_n est le plus petit réel tel que $k_n \geq k_{n-1} + 1$ et

$$\sup_{n \geq 1} \mathbb{P}(|X_n| \geq k_n) \leq \frac{1}{(n+1)^3}.$$

On vérifie alors aisément que $\sup_n \mathbb{E}G(|X_n|) < \infty$.

Soit $\{X_n,\ n \geq 1\}$ une suite de v. a. r. uniformément intégrable. On choisit $k_0 = 0$ et pour $n \geq 1$, k_n est le plus petit réel tel que $k_n \geq k_{n-1} + 1$ et

$$\sup_{n \geq 1} \mathbb{E}(|X_n|; |X_n| \geq k_n) \leq \frac{1}{(n+1)^3}.$$

On vérifie alors aisément que $\sup_n \mathbb{E}H(|X_n|) < \infty$. $\square$

7.9 Exercices

Exercice 7.9.1. *Montrer par un contre exemple que si $Q_n \Rightarrow Q$, on peut avoir $Q_n(A) \not\longrightarrow Q(A)$ pour certains $A \in \mathcal{F}$.*

Exercice 7.9.2. *Soit X et (pour chaque $n \geq 1$) X_n des v. a. r., qui possèdent les densités f et f_n.*

1. *Montrer que*
$$\int_{\mathbb{R}} (f(x) - f_n(x))^+ dx = \int_{\mathbb{R}} (f(x) - f_n(x))^- dx = \frac{1}{2} \int_{\mathbb{R}} |f(x) - f_n(x)| dx.$$
2. *En déduire (grâce au théorème de convergence dominée) que si $f_n(x) \to f(x)$ p. p. (pour la mesure de Lebesgue), alors $f_n \to f$ dans $L^1(\mathbb{R}, \mathcal{B}, dx)$.*
3. *Montrer que à nouveau si $f_n(x) \to f(x)$ p. p., X_n converge en loi vers X, quand $n \to \infty$ (on pourra montrer que $\mathbb{P}(X_n \leq x) \to \mathbb{P}(X \leq x)$, pour tout $x \in \mathbb{R}$).*

Exercice 7.9.3. *Soit X_n et Y_n, $n \geq 1$ des v. a. r. Si $X_n \to 0$ en probabilité, et la suite $\{Y_n\}$ est tendue, alors $X_n Y_n \to 0$ en probabilité.*

Exercice 7.9.4. *Soit $(X_n, n \geq 1)$ une suite de v.a avec $X_n \sim B(n, p_n)$ et $np_n \to \lambda > 0$. Montrer que $X_n \xrightarrow{\mathcal{L}} \mathcal{P}(\lambda)$.*

Exercice 7.9.5. *On suppose que les $X_n, n \geq 1$ sont indépendantes et que pour tout n, $X_n \sim \mathcal{E}(\alpha)$. On note $M_n = \max(X_k, k \leq n)$. Montrer que quand $n \to \infty$, $\alpha M_n - \log n$ converge en loi et préciser sa loi limite.*

Exercice 7.9.6. *Soit $X_\lambda \sim \mathcal{P}(\lambda)$ (loi de Poisson) et $Y_\lambda = \frac{X_\lambda - \lambda}{\sqrt{\lambda}}$.*

1. *Montrer que Y_λ converge en loi lorsque $\lambda \to \infty$ et déterminer sa loi limite.*
2. *Retrouver le résultat comme conséquence du Théorème Central Limite, en supposant que $\lambda \to \infty$ dans $\mathbb{N}$.*
3. *Montrer que*
$$e^{-n} \sum_{k=0}^{n} \frac{n^k}{k!} \to 1/2, \quad \text{quand } n \to \infty.$$

Exercice 7.9.7. *On suppose que les* $X_n, n \in \mathbb{N}$ *sont indépendantes et que pour tout* n, $X_n \sim N(0,1)$. *On pose* $U_n = \sum_{i=0}^{n-1} \frac{X_{2i+1}}{X_{2i+2}}$ *et* $V_n = \sum_{i=1}^{n} X_i^2$. *Montrer que* $\frac{U_n}{V_n}$ *converge en loi lorsque* $n \to \infty$ *et préciser sa loi limite.*

Exercice 7.9.8. *Soit* $(X_n, n \geq 1)$ *une suite de v.a.i.i.d. de loi de Cauchy, et* $S_n = \sum_{i=1}^{n} X_i$. *Montrer que* $\frac{S_n}{n}$ *converge en loi, mais pas en probabilité.*

Exercice 7.9.9. Convergence en loi et convergence des moments.

1. *Montrer que pour tout* $t \in \mathbb{R}$ *et pour tout* $l \in \mathbb{N}^*$, $\left| e^{it} - \sum_{k=0}^{l-1} \frac{(it)^k}{k!} \right| \leq \frac{|t|^l}{l!}$
2. *Soit* X *une variable aléatoire telle que* $E(e^{\lambda |X|}) < \infty$ *pour tout* $\lambda > 0$. *Soit* $(X_n, n \geq 1)$ *une suite de v.a ayant des moments de tous ordres. On suppose que pour tout* $k \in \mathbb{N}$, $E(X_n^k) \to E(X^k)$. *Montrer que* $X_n \xrightarrow{\mathcal{L}} X$. *On pourra d'abord montrer que si* ℓ *est pair,*

$$\overline{lim} \left| E(e^{itX_n}) - \sum_{k=0}^{\ell-1} \frac{(it)^k}{k!} E(X^k) \right| \leq \frac{|t|^\ell}{\ell!} E(X^\ell)$$

Chapitre 8

Espérance Conditionnelle

8.1 Introduction

Soit X et Y deux v.a.r. Supposons que l'on ait seulement accès à l'observation de la réalisation de Y. On veut en déduire des informations sur X. Par exemple, on veut calculer la probabilité conditionnelle :

$$\mathbb{P}_X(A|Y=y) \stackrel{\triangle}{=} \mathbb{P}(X \in A|Y=y), \;\; A \in \mathcal{B}.$$

Supposons pour l'instant que $\mathbb{P}(Y=y) > 0$. [le but de ce chapitre est précisément de traiter le cas où $\mathbb{P}(Y=y)=0$]. Alors,

$$\mathbb{P}_X(A|Y=y) = \frac{\mathbb{P}(\{X \in A\} \cap \{Y=y\})}{\mathbb{P}(Y=y)}$$

$A \to P_X(A|Y=y)$ est une probabilité, et on peut aussi définir, par exemple pour φ mesurable bornée,

$$\mathbb{E}[\varphi(X)|Y=y] \stackrel{\triangle}{=} \int \varphi(x)\mathbb{P}_X(dx|Y=y),$$

espérance conditionnelle de $\varphi(X)$, sachant que $Y=y$.

Réciproquement, la probabilité conditionnelle se définit à partir de l'espérance conditionnelle :

$$\mathbb{P}(X \in A|Y=y) = \mathbb{E}[\mathbf{1}_A(X)|Y=y].$$

Pour des raisons qui apparaitront à la fin de ce chapitre, il vaut mieux définir d'abord l'espérance conditionnelle. Pour motiver la façon dont nous allons procéder, considérons le cas très simple où Y prend un nombre fini de valeurs distinctes $y_1, \ldots, y_n$, avec $\mathbb{P}(Y = y_i) > 0$, $i = 1 \cdots n$. Supposons de plus X de carré intégrable. Il résulte de ce qui précède :

$$\mathbb{E}[X|Y = y_i] = \mathbb{P}(Y = y_i)^{-1} \int_{\{Y=y_i\}} X(\omega)\mathbb{P}(d\omega).$$

Donc il existe une fonction $g : \mathbb{R} \to \mathbb{R}$ t.q.

$$\mathbb{E}[X|Y = y_i] = g(y_i)$$

En composant les applications Y et g, on obtient une nouvelle v.a. $g(Y)$, et on pose par définition :

$$\mathbb{E}(X|Y) = g(Y)$$

On verra ci–dessous que $\mathbb{E}(X|Y)$ est en un certain sens la meilleure approximation de X, par une fonction de Y.

Soit maintenant une autre v.a.r. Z, qui prend également n valeurs distinctes, $z_1, \ldots, z_n$, et t.q. $\{Z = z_i\} = \{Y = y_i\}$, $i = 1, \cdots, n$, alors si h est telle que $\mathbb{E}(X|Z = z_i) = h(z_i), i = 1, \ldots, n$, on a :

$$g(y_i) = h(z_i), \;\; i = 1; \ldots, n, \text{ soit}$$

$$g(Y(\omega)) = h(Z(\omega)), \;\; \forall\, \omega \in \Omega$$

$$\text{donc } E(X|Y) = E(X|Z)$$

Y et Z prennent des valeurs différentes. Qu'est–ce qui est commun à Y et Z ? les tribus associées : $\sigma(Y) = \sigma(Z)$. Cette égalité peut s'interpréter intuitivement en disant que l'information apportée par l'observation de Y est exactement la même que celle apportée par l'observation de Z. Posons $\mathcal{G} = \sigma(Y)$. On note $E(X|\mathcal{G})$ pour $E(X|Y) = E(X|Z)$.

Nous allons maintenant définir et étudier ces notions dans le cas général, dans l'ordre inverse de cette introduction.

8.2 Par rapport à une σ–algèbre

Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité et $\mathcal{G} \subset \mathcal{F}$ une tribu.

8.2.1 Définition de l'espérance conditionnelle d'une v. a. de carré intégrable

Soit X une v.a.r. de carré intégrable. Si l'on identifie X et sa classe d'équivalence [i.e. la classe des v.a. qui lui sont p.s. égales], on peut écrire $X \in L^2(\Omega, \mathcal{F}, \mathbb{P})$.

Pour pouvoir considérer $L^2(\Omega, \mathcal{G}, \mathbb{P})$ comme un sous espace vectoriel de $L^2(\Omega, \mathcal{F}, \mathbb{P})$, on va supposer pour tout le reste de ce chapitre que $\mathcal{G}$ contient tous les ensembles de $\mathbb{P}$ mesure nulle de $\mathcal{F}$.

Remarque 8.2.1. *L'hypothèse que nous venons de faire est en réalité inutile. En effet, si l'on note $L^2(\mathcal{G}, \mathbb{P}_\mathcal{F})$ le sous espace vectoriel de $L^2(\Omega, \mathcal{F}, \mathbb{P})$ formé des classes d'équivalence qui contiennent au moins un représentant $\mathcal{G}$–mesurable, alors $L^2(\mathcal{G}, \mathbb{P}_\mathcal{F})$ est complet, comme image par une application isométrique de $L^2(\Omega, \mathcal{G}, \mathbb{P})$ (l'application qui à $u \in L^2(\Omega, \mathcal{G}, \mathbb{P})$ associe l'élément de $L^2(\mathcal{G}, \mathbb{P}_\mathcal{F})$ qui contient u). C'est donc un sous espace vectoriel fermé de $L^2(\Omega, \mathcal{F}, \mathbb{P})$. Donc on peut se dispenser de l'hypothèse que $\mathcal{G}$ contient tous les ensembles de $\mathbb{P}$–mesure nulle de $\mathcal{F}$. Il suffirait de remplacer ci–dessous $L^2(\Omega, \mathcal{G}, \mathbb{P})$ par $L^2(\mathcal{G}, \mathbb{P}_\mathcal{F})$.*

Définition 8.2.2. *On appelle* **espérance conditionnelle par rapport à** *$\mathcal{G}$ l'opérateur de projection orthogonale dans $L^2(\Omega, \mathcal{F}, \mathbb{P})$ sur $L^2(\Omega, \mathcal{G}, \mathbb{P})$.*

X étant une v.a.r. de carré intégrable, on appelle espérance conditionnelle de X sachant $\mathcal{G}$ - notée $\mathbb{E}^\mathcal{G}(X)$ ou $\mathbb{E}(X|\mathcal{G})$ - l'image de la classe d'équivalence de X par l'opérateur défini ci–dessus. $\mathbb{E}^\mathcal{G}(X)$ est donc un élément de $L^2(\Omega, \mathcal{G}, \mathbb{P})$, i.e. une classe d'équivalence. On considérera souvent $\mathbb{E}^\mathcal{G}(X)$ comme étant une variable aléatoire $\mathcal{G}$–mesurable (son choix à l'intérieur d'une classe d'équivalence est arbitraire !). Par définition de la projection orthogonale,

$(*)$ $\quad \mathbb{E}(YX) = \mathbb{E}(Y\mathbb{E}^\mathcal{G}(X))$, $\forall\, Y$ v.a.r. $\mathcal{G}$–mesurable et de carré intégrable. Notons que les propriétés de la projection orthogonale entraînent que

$$\mathbb{E}\left(|X - \mathbb{E}(X|\mathcal{G})|^2\right) = \min_{Y \in L^2(\Omega, \mathcal{G}, \mathbb{P})} \mathbb{E}(|X - Y|^2),$$

autrement dit $\mathbb{E}(X|\mathcal{G})$ est, parmi les (classes de) v. a. $\mathcal{G}$–mesurables et de carré intégrable, celle qui minimise la distance dans L^2 à X.

En choisissant $Y = 1$ dans $(*)$, on obtient :

$$\mathbb{E}(\mathbb{E}^\mathcal{G}(X)) = \mathbb{E}(X).$$

Lemme 8.2.3. *Soit X une v.a.r. de carré intégrable. Si $X \geq 0$ p.s., alors $\mathbb{E}^{\mathcal{G}}(X) \geq 0$ p.s.*

Preuve En appliquant $(*)$ avec $Y = \mathbf{1}_{\{\mathbb{E}^{\mathcal{G}}(X)<0\}}$, on obtient :

$$0 \leq \mathbb{E}\left[\mathbf{1}_{\{\mathbb{E}^{\mathcal{G}}(X)<0\}} X\right] = \mathbb{E}\left[\mathbf{1}_{\{\mathbb{E}^{\mathcal{G}}(X)<0\}} \mathbb{E}^{\mathcal{G}}(X)\right] \leq 0$$

D'où

$$\mathbb{P}(\mathbb{E}^{\mathcal{G}}(X) < 0) = 0.$$

□

Soit à nouveau $X \in L^2(\Omega, \mathcal{F}, \mathbb{P})$.

$$X = X^+ - X^-$$

D'aprés la linéarité de la projection orthogonale,

$$\mathbb{E}^{\mathcal{G}}(X) = \mathbb{E}^{\mathcal{G}}(X^+) - \mathbb{E}^{\mathcal{G}}(X^-)$$

Le Lemme 8.2.3 entraîne que $\mathbb{E}^{\mathcal{G}}(X^+) \geq 0$, $\mathbb{E}^{\mathcal{G}}(X^-) \geq 0$ p.s. Donc :

$$|\mathbb{E}^{\mathcal{G}}(X)| \leq \mathbb{E}^{\mathcal{G}}(X^+) + \mathbb{E}^{\mathcal{G}}(X^-) = \mathbb{E}^{\mathcal{G}}(|X|) \text{ p.s.},$$

d'où, par la formule établie juste avant le Lemme 8.2.3,

$$(**) \quad \mathbb{E}[|\mathbb{E}^{\mathcal{G}}(X)|] \leq \mathbb{E}(|X|).$$

8.2.2 Définition de l'espérance conditionnelle d'une v. a. intégrable

Le Théorème qui suit contient la définition de l'espérance conditionnelle $\mathbb{E}^{\mathcal{G}}(X)$ dans le cas d'une v.a.r. X intégrable.

Théorème 8.2.4. *Si X est une v.a.r.* **intégrable**, *$\mathbb{E}^{\mathcal{G}}(X)$ est l'unique classe d'équivalence de v.a.r. intégrables qui vérifie :*

(i) $\mathbb{E}^{\mathcal{G}}(X)$ est $\mathcal{G}$-mesurable et

(ii) $\mathbb{E}(ZX) = \mathbb{E}(Z\mathbb{E}^{\mathcal{G}}(X))$, pour tout Z $\mathcal{G}$ mesurable et bornée

Preuve Etant donné X intégrable, posons $X_n(\omega) = X(\omega)\mathbf{1}_{\{|X(\omega)|\le n\}}$. X_n est de carré intégrable et on peut définir $\mathbb{E}(X_n|\mathcal{G})$. En outre la suite $\{X_n,\ n \ge 1\}$ est de Cauchy dans $L^1(\mathbb{P})$. Il résulte alors de $(**)$ que la suite $\{\mathbb{E}(X_n|\mathcal{G})\}$ est aussi de Cauchy dans $L^1(\mathbb{P})$. Appelons $\mathbb{E}^{\mathcal{G}}(X)$ sa limite. Clairement (i) est satisfait. En outre, on sait que

$$\mathbb{E}(ZX^n) = \mathbb{E}(Z\mathbb{E}^{\mathcal{G}}[X^n]),\ \forall\, Z\ \mathcal{G}\text{–mesurable et bornée.}$$

(ii) s'obtient alors par passage à la limite dans $L^1(\Omega)$ dans cette identité.

L'unicité se déduit aisément du fait qui suit. Si Y est $\mathcal{G}$–mesurable et intégrable, et vérifie $\int_G Y d\mathbb{P} = 0$ pour tout $G \in \mathcal{G}$, alors

$$\int_{\{Y>0\}} Y d\mathbb{P} = 0 \text{ et } \int_{\{Y<0\}} Y d\mathbb{P} = 0.\ \text{ Donc } Y = 0 \text{ p. s.}$$

On déduit du Théorème des classes monotones 1.7.8 le

Corollaire 8.2.5. *Si X est une v.a.r. intégrable, $\mathbb{E}^{\mathcal{G}}[X]$ est l'unique (classe d'équivalence de) v.a.r. intégrables qui vérifie :*

(i) $\mathbb{E}^{\mathcal{G}}[X]$ est $\mathcal{G}$–mesurable.

(ii)' Pour tout $G \in \mathcal{G}$, $\int_G X d\mathbb{P} = \int_G \mathbb{E}^{\mathcal{G}}[X] d\mathbb{P}$.

On a en outre le

Corollaire 8.2.6. *Si X est une v.a.r. intégrable et $\mathcal{C}$ est un π–système qui contient Ω, avec $\sigma(\mathcal{C}) = \mathcal{G}$, alors $\mathbb{E}^{\mathcal{G}}[X]$ est l'unique (classe d'équivalence de) v.a.r. intégrables qui vérifie :*

(i) $\mathbb{E}^{\mathcal{G}}[X]$ est $\mathcal{G}$–mesurable.

(ii)" Pour tout $C \in \mathcal{C}$, $\int_C X d\mathbb{P} = \int_C \mathbb{E}^{\mathcal{G}}[X] d\mathbb{P}$.

Preuve Il suffit de montrer que $(i) + (ii)'' \Rightarrow (ii)'$ du Corollaire précédent. Ceci résulte du Théorème $\pi - \lambda$ 1.3.8, en remarquant que

$$\mathcal{L} = \left\{ B; \int_B X d\mathbb{P} = \int_B \mathbb{E}^{\mathcal{G}}[X] d\mathbb{P} \right\} \quad \text{est un } \lambda\text{–système.}$$

□

Exemple 8.2.7. *Supposons que $\mathcal{G}$ est engendrée par une partition finie $(B_1, \dots, B_n)$ de Ω, avec $\mathbb{P}(B_i) > 0$ pour $1 \le i \le n$. Alors $\mathbb{E}^{\mathcal{G}}[X]$ est la v.a.r. constante sur chaque B_i donnée par :*

$$\mathbb{E}^{\mathcal{G}}[X](\omega_i) = \frac{1}{\mathbb{P}(B_i)} \int_{B_i} X(\omega)\mathbb{P}(d\omega),\ \ \omega_i \in B_i.$$

En particulier, si $\mathcal{G} = \{\emptyset, \Omega\}$, $\mathbb{E}^{\mathcal{G}}[X] = \mathbb{E}[X]$.

8.2.3 Définition de l'espérance conditionnelle d'une v. a. non négative

Soit X une v.a.r, t. q. $X \geq 0$ p.s. Alors si $X_n \stackrel{\triangle}{=} \inf(X, n)$, X_n est intégrable pour tout n, et $X_n \uparrow X$ p.s. Il résulte du Lemme 8.2.3 que $\{\mathbb{E}^{\mathcal{G}}(X_n)\}$ est une suite p.s. croissante, et on peut définir $\mathbb{E}^{\mathcal{G}}(X)$ comme la limite p.s. de la suite $\mathbb{E}^{\mathcal{G}}(X_n)$. On pourrait définir plus généralement l'espérance conditionnelle d'une v. a. quasi–intégrable.

8.3 Propriétés de l'espérance conditionnelle

Proposition 8.3.1. *Si X,X_1,X_2 sont des v.a.r. intégrables,*

a) $\mathbb{E}^{\mathcal{G}}[aX_1 + bX_2] = a\mathbb{E}^{\mathcal{G}}[X_1] + b\mathbb{E}^{\mathcal{G}}[X_2]$

b) $X \geq 0$ *p. s.* $\Rightarrow \mathbb{E}^{\mathcal{G}}[X] \geq 0$ *p. s.*
$X > 0$ *p. s.* $\Rightarrow \mathbb{E}^{\mathcal{G}}[X] > 0$ *p. s.*

c)Si $\mathcal{H}$ et $\mathcal{G} \vee \sigma(X)$ sont indépendantes,

$$\mathbb{E}[X|\mathcal{G} \vee \mathcal{H}] = \mathbb{E}[X|\mathcal{G}]$$

en particulier, si $\mathcal{H}$ et X sont indépendantes,

$$\mathbb{E}[X|\mathcal{H}] = \mathbb{E}[X]$$

d) Si $\mathcal{H} \subset \mathcal{G}$, $\mathbb{E}^{\mathcal{H}}[\mathbb{E}^{\mathcal{G}}[X]] = \mathbb{E}^{\mathcal{H}}[X]$ *en particulier,* $\mathbb{E}[\mathbb{E}^{\mathcal{G}}[X]] = \mathbb{E}[X]$.

e) Si X est $\mathcal{G}$–mesurable, Y et XY sont intégrables, alors

$$\mathbb{E}^{\mathcal{G}}[YX] = X\mathbb{E}^{\mathcal{G}}[Y]$$

Remarque 8.3.2. *Une conséquence de cette Proposition est la double assertion suivante, qui donne les deux cas où $\mathbb{E}^{\mathcal{G}}(X)$ s'explicite aisément. Soit X une v.a.r. intégrable.*

Si X est $\mathcal{G}$–mesurable, alors $\mathbb{E}^{\mathcal{G}}(X) = X$.

Si X et $\mathcal{G}$ sont indépendantes, alors $\mathbb{E}^{\mathcal{G}}(X) = \mathbb{E}(X)$.

Preuve a) découle de ce que la projection orthogonale dans L^2 est linéaire, propriété qui se conserve par passage à la limite dans L^1.

b) la 1ère partie se démontre comme au Lemme 8.2.3. La deuxième en observant que $\int_{\{\mathbb{E}^{\mathcal{G}}(X)\leq 0\}} X d\mathbb{P} = 0 \Rightarrow \mathbb{P}(\mathbb{E}^{\mathcal{G}}(X) \leq 0) = 0$, dès que $X > 0$ p.s.

c) On utilise le Corollaire 8.2.6, avec

$$\mathcal{C} = \{G \cap H; G \in \mathcal{G}, H \in \mathcal{H}\}$$

La deuxième partie résulte de la première en posant $\mathcal{G} = \{\emptyset, \Omega\}$.

d) Pour tout $H \in \mathcal{H}$, $\int_H X d\mathbb{P} = \int_H \mathbb{E}^{\mathcal{G}}(X) d\mathbb{P} = \int_H \mathbb{E}^{\mathcal{H}}[\mathbb{E}^{\mathcal{G}}(X)] d\mathbb{P}$. La deuxième partie s'obtient alors en choisissant $\mathcal{H} = \{\emptyset, \Omega\}$.

e) Supposons tout d'abord X $\mathcal{G}$–mesurable et bornée. Pour tout Z $\mathcal{G}$–mesurable et bornée,

$$\mathbb{E}[Z\mathbb{E}^{\mathcal{G}}(XY)] = \mathbb{E}[ZXY] = \mathbb{E}[ZX\mathbb{E}^{\mathcal{G}}(Y)]$$

Pour passer au cas général, on approche X par $X_n = X\mathbf{1}_{\{|X|\leq n\}}$, et on remarque que $X_n \to X$ p.s. et $X_n Y \to XY$ dans $L^1(\Omega)$. $\square$

On a un Théorème de convergence dominée pour les espérances conditionnelles :

Proposition 8.3.3. *Si $X_n \to X$ p.s. et s'il existe Z intégrable t.q. $|X_n| \leq Z$ p.s. pour tout $n \geq 1$, alors X_n et X sont intégrables et*

$$\mathbb{E}^{\mathcal{G}}[X_n] \to \mathbb{E}^{\mathcal{G}}[X] \textit{ p.s. et dans } L^1(\Omega)$$

Montrons d'abord le :

Lemme 8.3.4. *Soit $\{X_n\}_{n\geq 1}$ une suite croissante [resp. décroissante] de v.a.r. intégrables, et X une v.a.r. intégrable. Alors les deux assertions suivantes sont équivalentes :*

(i) $X_n \to X$ p.s.

(ii) $X_n \to X$ dans $L^1(\Omega)$.

De plus, si l'une des deux assertions est vraie, $\mathbb{E}^{\mathcal{G}}[X_n] \to \mathbb{E}^{\mathcal{G}}[X]$ p.s. et dans $L^1(\Omega)$.

Preuve Si $X_n \uparrow X$ p.s. ou $X_n \downarrow X$ p.s., alors $|X - X_n| \downarrow 0$ p.s., et $|X - X_n| \leq |X - X_0|$ intégrable, donc $\mathbb{E}(|X - X_n|) \to 0$. Réciproquement, $X_n \to X$ dans $L^1(\Omega)$ entraîne $X_n \leq X$ p.s. [resp. $X_n \geq X$ p.s.]. Donc il

existe Z intégrable t.q. $X_n \uparrow Z$ p.s. [resp. $X_n \downarrow Z$ p.s.]. Alors par unicité de la limite en probabilités, $Z = X$ p.s. Enfin, (ii) $\Rightarrow \mathbb{E}^{\mathcal{G}}(X_n) \to \mathbb{E}^{\mathcal{G}}(X)$ dans $L^1(\Omega)$, donc aussi p.s. par ce qui précède, car $\mathbb{E}^{\mathcal{G}}(X_n) \uparrow$ ou $\downarrow$.
Preuve de la Proposition 8.3.3

$$\mathbb{E}^{\mathcal{G}}\left(\inf_{m\geq n} X_m\right) \leq \inf_{m\geq n} \mathbb{E}^{\mathcal{G}}(X_m) \leq \sup_{m\geq n} \mathbb{E}^{\mathcal{G}}(X_m) \leq \mathbb{E}^{\mathcal{G}}\left(\sup_{m\geq n} X_m\right)$$

Or $\inf_{m\geq n} X_m \uparrow X$, $\sup_{m\geq n} X_m \downarrow X$ p.s., et aussi dans $L^1(\Omega)$ d'après le lemme, que l'on peut utiliser grâce à l'hypothèse de domination, d'où :

$$\mathbb{E}^{\mathcal{G}}(X) \leq \liminf \mathbb{E}^{\mathcal{G}}(X_n) \leq \limsup \mathbb{E}^{\mathcal{G}}(X_n) \leq \mathbb{E}^{\mathcal{G}}(X)$$

d'où

$$\mathbb{E}^{\mathcal{G}}(X_n) \to \mathbb{E}^{\mathcal{G}}(X) \text{ p.s.}$$

□

Proposition 8.3.5. *Pour tout $p \geq 1$,*

(i) Si X est de puissance p–ième intégrable, $\mathbb{E}^{\mathcal{G}}[X]$ est de puissance p–ième intégrable.

(ii) Si $X_n \to X$ dans $L^p(\Omega)$,

$$\mathbb{E}^{\mathcal{G}}[X_n] \to \mathbb{E}^{\mathcal{G}}[X] \text{ dans } L^p(\Omega).$$

Preuve Dans les cas $p = 1$ et $p = 2$, ces propriétés résultent de la construction. Dans le cas général, ces propriétés résultent de :

$$(***) \quad |\mathbb{E}^{\mathcal{G}}(X)|^p \leq \mathbb{E}^{\mathcal{G}}[|X|^p],$$

qui entraîne $\mathbb{E}[|\mathbb{E}^{\mathcal{G}}(X)|^p] \leq \mathbb{E}[|X|^p]$. (∗∗∗) est un cas particulier de l'inégalité de Jensen pour les espérances conditionnelles :

Proposition 8.3.6. *Soit $\varphi : \mathbb{R} \to \mathbb{R}$ une fonction convexe. Alors pour toute v.a.r. intégrable X t.q. $\varphi(X)$ est également intégrable,*

$$\varphi[\mathbb{E}^{\mathcal{G}}(X)] \leq \mathbb{E}^{\mathcal{G}}[\varphi(X)].$$

Preuve On va utiliser la propriété suivante des fonctions convexes (qui est facile à vérifier dans le cas où φ est de classe C^1) : il existe une suite (a_n, b_n) dans $\mathbb{R}^2$ t.q. :

$$\varphi(x) = \sup_n (a_n x + b_n), \quad \forall\, x \in \mathbb{R}$$

Donc

$$\varphi(X) \geq a_n X + b_n \text{ p.s.}, \forall\, n \geq 1,$$
$$\text{et } \mathbb{E}^{\mathcal{G}}[\varphi(X)] \text{ p.s.} \geq \sup_n (a_n \mathbb{E}^{\mathcal{G}}(X) + b_n) = \varphi(\mathbb{E}^{\mathcal{G}}[X])$$

□

Proposition 8.3.7. *Soit X un v.a. de dimension d, Y un v.a. de dimension k, $\mathcal{G}$ une σ–algèbre tels que X est $\mathcal{G}$–mesurable, Y et $\mathcal{G}$ sont indépendantes. Alors si f une application mesurable de $\mathbb{R}^{d\times k}$ dans $\mathbb{R}$ t.q. $f(X,Y)$ est intégrable. Alors*

$$\begin{aligned}\mathbb{E}[f(X,Y)|\mathcal{G}] &= \int_{\mathbb{R}^d} f(X,y)\mathbb{P}_Y(dy) \\ &= \mathbb{E}[f(x,Y)]\Big|_{x=X}\end{aligned}$$

Preuve Supposons tout d'abord que f est de la forme $f(x,y) = g(x)\times h(y)$, avec g (resp. h) mesurable bornée de $\mathbb{R}^d$ (resp. $\mathbb{R}^k$) dans $\mathbb{R}$. Alors

$$\mathbb{E}[h(X)g(Y)|\mathcal{G}] = h(X)\mathbb{E}[g(Y)],$$

donc dans ce cas les deux formules de l'énoncé sont correctes. Le résultat est bien sûr encore vrai pour $f(x,y) = \sum_{i=1}^n a_i h_i(x) g_i(y)$, si les h_i et les g_i sont mesurables et bornées. Le résultat général est alors une conséquence du Théorème des classes monotones 1.7.8. □

8.4 Par rapport à une variable aléatoire

Rappelons qu'à tout v.a. Y, on associe sa "tribu naturelle" $\sigma(Y)$, qui est la plus petite sous–tribu de $\mathcal{F}$ qui rend Y mesurable. On pose :

Définition 8.4.1. *Si X est une v.a.r. intégrable, on définit l'espérance conditionnelle de X, sachant Y, par :*

$$\mathbb{E}(X|Y) \stackrel{\triangle}{=} \mathbb{E}(X|\sigma(Y))$$

□

Dans cette définition, Y peut être un vecteur aléatoire de dimension quelconque. Remarquons que si $\sigma(Y) = \sigma(Y')$

$$\mathbb{E}(X|Y) = \mathbb{E}(X|Y') \text{ p.s.}$$

Dans la pratique, on observe une réalisation y de Y, et on voudrait définir $\mathbb{E}(X|Y = y)$. Pour cela, on va utiliser la :

Proposition 8.4.2. *Soit Y un v.a. de dimension d, et Z une v.a.r., tous deux définis sur $(\Omega, \mathcal{F})$. Si Z est $\sigma(Y)$–mesurable, alors il existe une application φ borélienne de $\mathbb{R}^d$ dans $\mathbb{R}$, telle que :*

$$Z = \varphi(Y) \ p.s.$$

Preuve Considérons la classe des v.a.r. sur $(\Omega, \mathcal{F})$: $\mathcal{C} = \{\varphi(Y);\ \varphi \text{ mesurable } : \mathbb{R}^d \to \mathbb{R}\}$.

Il suffit de montrer que $\mathcal{C}$ contient toutes les v.a.r. $\sigma(Y)$–mesurables. Pour cela, on va appliquer le Théorème des classes monotones 1.7.8 ; $\mathcal{C}$ est un espace vectoriel, et pour tout $A \in \sigma(Y)$, il existe $B \in \mathcal{B}_d$ t.q. $A = \{Y \in B\}$, donc $\mathcal{C}$ contient $\mathbf{1}_A = \mathbf{1}_{\{Y \in B\}}$.

Il nous reste à vérifier que si $X_n \in \mathcal{C}$, $X_n(\omega) \to X(\omega)$, pour tout $\omega \in \Omega$, alors $X \in \mathcal{C}$.

Soit donc φ_n une suite d'applications mesurables de $\mathbb{R}^d$ dans $\mathbb{R}$, t.q. $\varphi_n(Y(\omega)) \to X(\omega)$, pour tout $\omega \in \Omega$. Posons $B \in \{y \in \mathbb{R}^d; \lim \varphi_n(y) \text{ existe }\}$, $B \in \mathcal{B}_d$, car puisqu'une suite de réels converge ssi elle est de Cauchy,

$$B = \bigcap_{k=1}^{\infty} \bigcup_{n=1}^{\infty} \bigcap_{i=1}^{\infty} \left\{ |\varphi_{n+i}(y) - \varphi_n(y)| < \frac{1}{k} \right\}.$$

On pose

$$\varphi(y) = \begin{cases} \lim \varphi_n(y) & \text{si } y \in B \\ 0 & \text{si } y \in B^c \end{cases}$$

Alors φ est mesurable, et $X = \varphi(Y)$ p.s. □

D'après la Proposition 8.4.2, il existe une application mesurable $\varphi : \mathbb{R}^d \to \mathbb{R}$, t.q. :

$$\mathbb{E}(X|Y) = \varphi(Y) \text{ p.s.}$$

φ peut être considérée comme une v.a.r. définie sur $(\mathbb{R}^d, \mathcal{B}_d, \mathbb{P}_Y)$. On vérifie qu'elle est unique à classe de $\mathbb{P}_Y$ -équivalence près. On pose alors, si X est une v.a.r. intégrable :

Définition 8.4.3. $\mathbb{E}(X|Y=y)$ *est l'unique classe d'équivalence* $\varphi(y)$ *de v.a.r. définies sur* $(\mathbb{R}^d, \mathcal{B}_d, \mathbb{P}_Y)$*, t.q.* $\mathbb{E}(X|Y)$ *et* $\varphi(Y)$ *soient p.s. égales.*

$\square$

Remarque 8.4.4. *Fixons* y_0 *t.q.* $\mathbb{P}_Y(\{y_0\}) = 0$*. On ne peut pas définir* $\mathbb{E}(X|Y=y_0)$*, car cette quantité dépend du choix arbitraire du représentant dans la classe de fonctions* $y \to \mathbb{E}(X|Y=y)$*.*

8.5 Probabilité conditionnelle

Définition 8.5.1. *On appelle probabilité conditionnelle de l'événement* A*, sachant* $\mathcal{G}$*, la (classe d'équivalence de) v.a.r. :*

$$\mathbb{P}(A|\mathcal{G}) \stackrel{\triangle}{=} \mathbb{E}[\mathbf{1}_A|\mathcal{G}]$$

On appelle probabilité conditionnelle de A*, sachant* X *:*

$$\mathbb{P}(A|X] \stackrel{\triangle}{=} \mathbb{E}[\mathbf{1}_A|X]$$

On appelle probabilité conditionnelle de A *sachant que* $X=x$ *:*

$$\mathbb{P}(A|X=x) \stackrel{\triangle}{=} \mathbb{E}[\mathbf{1}_A|X=x]$$

$\square$

On peut alors se poser la question suivante : l'espérance conditionnelle peut–elle être définie comme une intégrale par rapport à la probabilité conditionnelle ? Ceci n'est pas toujours possible. Le problème est de choisir pour tout $A \in \mathcal{F}$ un représentant de $\mathbb{P}(A|\mathcal{G})$, en faisant tous ces choix de façon "cohérente", pour que à ω fixé, $A \to \mathbb{P}(A|\mathcal{G})(\omega)$ soit une probabilité.

Soit $\begin{pmatrix} X \\ Y \end{pmatrix}$ un vecteur aléatoire à valeurs dans $\mathbb{R}^{d+k}$; Posons la :

Définition 8.5.2. *a)* $\mathbb{P}_Y(\cdot|X)$ *est appelée loi de probabilité conditionnelle régulière de* Y *sachant* X *si :*

(i) Pour tout $B \in \mathcal{B}_k$*,* $\mathbb{P}_Y(B|X) = \mathbb{P}(Y \in B|X)$ *p.s.*

[ii) Pour tout $\omega \in \Omega$*,* $B \to \mathbb{P}_Y(B|X)(\omega)$ *est une mesure de probabilité sur* $(\mathbb{R}^k, \mathcal{B}_k)$*.*

b) $\mathbb{P}_Y(\cdot|X=x)$ *est appelée loi de probabilité conditionnelle régulière de* Y*, sachant que* $X=x$ *si :*

(i) Pour tout $B \in \mathcal{B}_k$*,* $\mathbb{P}_Y(B|X=x) = \mathbb{P}(Y \in B|X=x)$ $\mathbb{P}_X$ *p.s.*

(ii) Pour tout $x \in \mathbb{R}^d$*,* $B \to \mathbb{P}_Y(B|X=x)$ *est une mesure de probabilité sur* $(\mathbb{R}^k, \mathcal{B}_k)$*.*

□

On démontre aisément à l'aide du Théorème des classes monotones la :

Proposition 8.5.3. *Si* $\mathbb{P}_Y(\cdot|X)$ *[resp.* $\mathbb{P}_Y(\cdot|X=x)$*] est une loi de probabilité conditionnelle régulière de* Y *sachant* X *[resp. sachant que* $X=x$*], alors pour toute fonction* φ *mesurable de* $\mathbb{R}^k$ *dans* $\mathbb{R}$*, t.q.*

$$\begin{aligned} \mathbb{E}(|\varphi(Y)|) &< \infty \\ \mathbb{E}[\varphi(Y)|X] &= \int_{\mathbb{R}^k} \varphi(y)\mathbb{P}_Y(dy|X) p.s. \\ \mathbb{E}[\varphi(Y)|X=x] &= \int_{\mathbb{R}^k} \varphi(y)\mathbb{P}_Y(dy|X=x)\mathbb{P}_X p.s. \end{aligned}$$

□

On démontre qu'une loi de probabilité conditionnelle régulière existe toujours. Nous allons la calculer dans deux cas :

1. celui où la loi du couple $\begin{pmatrix} X \\ Y \end{pmatrix}$ admet une densité,
2. celui où le couple $\begin{pmatrix} X \\ Y \end{pmatrix}$ est gaussien.

Proposition 8.5.4. *Soit* $\begin{pmatrix} X \\ Y \end{pmatrix}$ *un v.a. de dimension* $d+k$*, dont la loi admet la densité* $f(x,y)$*. On pose* $g(x) = \int f(x,y)dy$*,* $Q = \{x \in \mathbb{R}^d; 0 < g(x) < \infty\}$

Alors

a) $\mathbb{P}(X \in Q) = 1$

b) Pour tout $x \in Q$*,* $f(y|x) \triangleq \dfrac{f(x,y)}{g(x)}$ *est la densité d'une loi de probabilité sur* $\mathbb{R}^k$*.*

c) $f(y|X)dy$ est p.s. égale à une loi de probabilité conditionnelle régulière de Y, sachant X.

d) $f(y|x)dy$ est $\mathbb{P}_Y$ p.s. égale à une loi de probabilité conditionnelle régulière de Y sachant que $X = x$

Preuve a) $\mathbb{P}(X \notin Q) = \int_{\{g=0\}} g(x)dx + \int_{\{g=+\infty\}} g(x)dx$. Or

$$\int g(x)dx < \infty, \text{ donc } \int_{\{g=+\infty\}} dx = 0 \text{ donc } \int_{\{g=+\infty\}} g(x)dx = 0.$$

b) est élémentaire.

c) Soit $h(y)$ la densité d'une probabilité sur $(\mathbb{R}^k, \mathcal{B}_k)$. Posons

$$\bar{f}(y|X)(\omega) = \begin{cases} h(y) & \text{si } X(\omega) \notin Q \\ f(y|X)(\omega) & \text{si } X(\omega) \in Q \end{cases}$$

Grâce à a b , pour tout ω, $\bar{f}(y|X)(\omega)$ est la densité d'une probabilité sur $(\mathbb{R}^k, \mathcal{B}_k)$.

Il reste à vérifier que pour tout $B \in \mathcal{B}_k$,

$$\int_B \bar{f}(y|X)dy = P(Y \in B|X) \text{ p.s.}$$

D'après 8.4.2 et 8.2.4, ceci est vrai si pour tout φ mesurable et bornée de $\mathbb{R}^d$ dans $\mathbb{R}$,

$$\mathbb{E}[\varphi(X) \int_B \bar{f}(y|X)dy] = \mathbb{E}[\varphi(X)\mathbf{1}_B(Y)]$$

Or

$$\begin{aligned} \mathbb{E}[\varphi(X) \int_B \bar{f}(y|X)dy] &= \int_{\mathbb{R}^k} \int_B \varphi(x) f(y|x) dy dx \\ &= \int_{\mathbb{R}^k} \int_{\mathbb{R}^d} \mathbf{1}_B(y)\varphi(x) f(x,y) g(x) dy dx \end{aligned}$$

d) se démontre comme c). □

Montrons finalement le

Théorème 8.5.5. *Soit $\begin{pmatrix} X \\ Y \end{pmatrix}$ un vecteur aléatoire gaussien d'espérance $\begin{pmatrix} \mu \\ \nu \end{pmatrix}$ et de matrice de covariance $\begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$, tel que la matrice Σ_{22} est inversible. Posons $\hat{X} = \mathbb{E}(X|Y)$ et $\hat{\Sigma} = Cov(X - \hat{X})$. Alors*

1. *$\hat{X} = \mu + \Sigma_{12}\Sigma_{22}^{-1}(Y - \nu)$,*
2. *$\hat{\Sigma} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$,*
3. *$N(\hat{X}, \hat{\Sigma})$ est une loi de probabilité conditionnelle régulière de X, sachant Y.*

Preuve Notons d la dimension de X, k celle de Y. Cherchons tout d'abord A matrice $d \times k$ et b vecteur de $\mathbb{R}^d$ tels que $Z := AY + b$ satisfasse

$$\mathbb{E}(Z) = \mathbb{E}(X), \quad \mathrm{Cov}\,(Z, Y) = \mathrm{Cov}\,(X, Y),$$

où par définition $\mathrm{Cov}(X, Y) = \mathbb{E}(XY') - \mathbb{E}(X)\mathbb{E}(Y)' = \Sigma_{12}$. Les deux conditions ci–dessus imposent que

$$\mu = A\nu + b, \quad A\Sigma_{22} = \Sigma_{12},$$

d'où

$$\begin{aligned} &A = \Sigma_{12}\Sigma_{22}^{-1}, \quad b = \mu - \Sigma_{12}\Sigma_{22}^{-1}\nu, \\ &Z = \mu + \Sigma_{12}\Sigma_{22}^{-1}(Y - \nu). \end{aligned}$$

Il reste à montrer que $Z = \mathbb{E}(X|Y)$, et que le théorème est satisfait. La preuve de ces faits est basée sur la remarque suivante : les vecteurs aléatoires $\tilde{X} := X - Z$ et Y sont indépendants. En effet le vecteur aléatoire $\begin{pmatrix} \tilde{X} \\ Y \end{pmatrix}$ est un vecteur aléatoire gaussien (parce qu'il est l'image par une application affine du vecteur aléatoire gaussien $\begin{pmatrix} X \\ Y \end{pmatrix}$), et

$$\mathrm{Cov}\,(\tilde{X}, Y) = \mathrm{Cov}\,(X, Y) - \mathrm{Cov}\,(Z, Y) = 0.$$

On utilise ici la Proposition 8.5.6 ci–dessous, qui est un cas particulier du Théorème 5.4.5. Il est maintenant clair que

$$\mathbb{E}(X|Y) = \mathbb{E}(Z|Y) + \mathbb{E}(\tilde{X}|Y) = \mathbb{E}(Z|Y) + \mathbb{E}(\tilde{X}) = Z.$$

On note maintenant $\hat{X} := Z$. Notons que la loi de probabilité de $\tilde{X}$ est la loi $N(0, \hat{\Sigma})$ car clairement

$$\mathrm{Cov}\,(\hat{X}) = \Sigma_{12}\Sigma_{22}^{-1}\mathrm{Cov}\,(Y)\Sigma_{22}^{-1}\Sigma_{21} = \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21},$$

et par l'indépendance de $\hat{X}$ et $\tilde{X}$,

$$\mathrm{Cov}\,(X) = \mathrm{Cov}\,(\hat{X}) + \mathrm{Cov}\,(\tilde{X}).$$

Or si $f \in C_K(\mathbb{R}^d)$, on déduit de la Proposition 8.3.7

$$\begin{aligned}\mathbb{E}[f(X)|Y] &= \mathbb{E}[f(\hat{X}+\tilde{X})|Y] \\ &= \int_{\mathbb{R}^d} f(\hat{X}+x)\mathbb{P}_{\tilde{X}}(dx) \\ &= \int_{\mathbb{R}^d} f(x)N_{(\hat{X},\hat{\Sigma})}(dx),\end{aligned}$$

puisque la loi de probabilité de $\tilde{X}$ est la probabilité gaussienne $N(0,\hat{\Sigma})$, donc cette loi translatée de $\hat{X}$ est la probabilité $N(\hat{X},\hat{\Sigma})$. □

On a utilisé le cas particulier suivant du Théorème 5.4.5 :

Proposition 8.5.6. *Si* $\begin{pmatrix} U \\ V \end{pmatrix}$ *est un vecteur aléatoire gaussien, où* U *est de dimension* d *et* V *de dimension* k*, alors les vecteurs aléatoires* U *et* V *sont indépendants ssi la matrice de covariance du v. a.* $\begin{pmatrix} U \\ V \end{pmatrix}$ *est bloc–diagonale, i. e. ssi* $Cov(U,V) = 0$.

Finissons ce chapitre par la

Remarque 8.5.7. *Si* X *et* Y *sont deux vecteurs aléatoires, de dimensions respectives* d *et* k*, de carré intégrable, l'espérance conditionnelle* $\mathbb{E}(X|Y)$ *est la (classe d'équivalence de) v. a. de dimension* d $\varphi(Y)$ *caratérisée par le fait que*

$$\mathbb{E}\left[|X-\varphi(Y)|^2\right] = \min_{\psi} \mathbb{E}\left[|X-\psi(Y)|^2\right],$$

où le minimum est pris sur toutes les applications boréliennes $\psi : \mathbb{R}^k \to \mathbb{R}^d$*, telles que* $\mathbb{E}[|\psi(Y)|^2] < \infty$.

On pourrait se poser la question d'approcher au mieux X *dans* $L^2(\Omega,\mathcal{F},\mathbb{P};\mathbb{R}^d)$ *par une fonction affine de* Y*. On trouverait que le minimum est atteint par l'application* $y \to Ay+b$ *où* A *et* b *sont choisis tels que*

$$\mathbb{E}(AY+b) = \mathbb{E}(X), \quad Cov(AY+b,Y) = Cov(X,Y).$$

C'est exactement comme cela que nous avons caractérisé Z *au début de la preuve du Théorème 8.5.5. Dans le cas où le couple* (X,Y) *est gaussien,* *la meilleure approximation en moyenne quadratique de X par une fonction affine de Y coïncide avec la meilleure approximation en moyenne quadratique de X par une fonction arbitraire de Y.*

La formule qui définit Z est celle de la meilleure approximation en moyenne quadratique de X par une fonction affine de Y dans le cas général (non nécessairement gaussien).

8.6 Exercices

Exercice 8.6.1. *Soit $(\Omega, \mathcal{F}, P)$ un espace probabilisé et $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$. Soit $(A_i, i \in \mathbb{N})$ un système exhaustif d'évènements tel que pour tout $i \in \mathbb{N}$, $\mathbb{P}(A_i) > 0$. Soit $\mathcal{G}$ la tribu engendrée par les $(A_i, i \in \mathbb{N})$. Expliciter $\mathbb{E}(X|\mathcal{G})$.*

Soit X et Y deux variables aléatoires telles que X prend ses valeurs dans un ensemble dénombrable $\{x_i, i \in \mathbb{N}\}$, avec pour tout $i \in \mathbb{N}$, $\mathbb{P}(X = x_i) > 0$. Expliciter $\mathbb{E}(Y|X)$.

Exercice 8.6.2. *Soit X_1 et X_2 deux v.a indépendantes. Calculer $\mathbb{E}(X_1|X_1 + X_2)$ lorsque*

1. $X_i \sim \mathcal{B}(n_i, p); p \in [0; 1]; n_1, n_2 \in \mathbb{N}$;
2. $X_i \sim \mathcal{P}(\lambda_i), \lambda_i > 0.$

Exercice 8.6.3. *Soit X et Y deux variables aléatoires sur $(\Omega, \mathcal{F}, \mathbb{P})$. On suppose que X est à valeurs dans $\mathbb{N}$, et que Y suit une loi exponentielle de paramètre 1. On suppose de plus que la loi de X conditionnelle sachant Y est une loi de Poisson de paramètre Y, i.e*

$$\forall k \geq 0,\ \mathbb{P}\ p.s.,\ \ \mathbb{P}[X = k|Y] = \exp(-Y)\frac{Y^k}{k!}.$$

Déterminer la loi de (X, Y), la loi de X, la loi conditionnelle de Y sachant que $X = k$.

Exercice 8.6.4. *Soit X une v. a. r. de loi exponentielle de paramètre $\alpha > 0$. Soit $(X_n)_{n\geq 1}$ une suite de v.a.i.i.d de même loi que X. Soit d'autre part N une variable aléatoire indépendante de la suite $(X_n)_{n\geq 1}$ et de loi géométrique de paramètre $p \in]0, 1[$: $\forall n \in \mathbb{N}^*, \mathbb{P}(N = n) = p(1 - p)^{n-1}$. On pose $S = X_1 + \cdots + X_N$. Calculer $\mathbb{E}(S|N)$, la loi de S et $\mathbb{E}(N|S)$.*

Exercice 8.6.5. *Soit (X, Y) de loi uniforme sur le triangle $\Delta := \{(x, y) \in \mathbb{R}^2, 0 < x < y < 1\}$. Calculer $\mathbb{E}(Y|X)$, $\mathbb{E}(X|Y)$, et $\mathbb{E}(X^2|Y)$.*

Exercice 8.6.6. *Soit $\mathcal{G}, \mathcal{F}, \mathcal{H}$ des tribus sur Ω, telles $\mathcal{G} \subset \mathcal{H} \subset \mathcal{F}$. Soit $X \in L^2(\Omega, \mathcal{F}, \mathbb{P})$. Montrer l'égalité*

$$\mathbb{E}[(X - E(X|\mathcal{H}))^2] + \mathbb{E}[(E(X|\mathcal{H}) - E(X|\mathcal{G}))^2] = \mathbb{E}[(X - E(X|\mathcal{G}))^2].$$

Exercice 8.6.7. *Soit $p \geq 1$ et X et Y deux v.a.r. telles que $\|X\|_p = (E|X|^p)^{1/p} < +\infty$ et $\|Y\|_p < +\infty$. Montrer que $\|X+Y\|_p \geq \|X+\mathbb{E}(Y|X)\|_p$. En déduire que, si X et Y sont indépendantes, alors $\|X+Y\|_p \geq \|X + \mathbb{E}(Y)\|_p$.*

Exercice 8.6.8. *Soit X et Y deux éléments de $L^2(\Omega, \mathcal{F}, \mathbb{P})$ tels que $\mathbb{E}(X|Y) = Y$ et $\mathbb{E}(Y|X) = X$. Montrer que $X = Y$ p.s.*

Exercice 8.6.9. *On reprend les notations de l'exercice 4.8.25, mais on ne suppose plus les v. a. T_i de loi exponentielle. On définit la fonction de risque h_i de la v. a. T_i comme la fonction de $\mathbb{R}_+$ dans $\mathbb{R}_+$ donnée par*

$$h_i(t) = \frac{f_i(t)}{S_i(t)}, \text{ où } f_i \text{ est la densité de la loi de } T_i,\ S_i(t) = \mathbb{P}(T_i > t).$$

1. *Montrer que T_i suit la loi exponentielle de paramètre λ_i ssi $h_i(t) \equiv \lambda_i$.*
2. *Calculer $\mathbb{P}(\Lambda > t)$ en fonction des quantités $S_i(t)$, $1 \leq i \leq k$. En déduire que la v. a. Λ admet la densité $f_\Lambda(t) = \left(\prod_{j=1}^k S_j(t)\right) \sum_{i=1}^k h_i(t)$.*
3. *Montrer que $\mathbb{P}(T_i < \min_{j\neq i} T_j | T_i) = \prod_{j \neq i} S_j(T_i)$.*
4. *Montrer que pour toute $\varphi : \mathbb{R}_+ \to \mathbb{R}$ borélienne bornée,*
$$\mathbb{E}\left[\mathbf{1}_{\{T_i < \min_{j\neq i} T_j\}} \varphi(\Lambda)\right] = \int_0^\infty \prod_{j\neq i} S_j(t)\varphi(t) f_i(t) dt.$$
5. *Déterminer la fonction $\psi_i : \mathbb{R}_+ \to \mathbb{R}_+$ (en fonction des h_j) telle que*
$$\int_0^\infty \prod_{j\neq i} S_j(t)\varphi(t) f_i(t) dt = \int_0^\infty \psi_i(t)\varphi(t) f_\Lambda(t) dt.$$
6. *Déduire de ce qui précède que*
$$\mathbb{P}(\xi = i|\Lambda) = \psi_i(\Lambda).$$
Que devient ce résultat dans le cas étudié à l'exercice 4.8.24 (cas des T_i de loi exponentielle) ? Comment retrouve–t–on le fait que dans ce cas là ξ et Λ sont indépendantes, et comment voit–on qu'elles ne le sont pas dans le cas général ?

Chapitre 9

Martingales

9.1 Définition et exemples

On se donne maintenant un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$ muni d'une filtration $\{\mathcal{F}_n,\ n \geq 0\}$, i. e. une suite croissante de sous–tribus de $\mathcal{F}$. Autrement dit chaque $\mathcal{F}_n$ est une tribu, et pour tout $n \geq 0$, $\mathcal{F}_n \subset \mathcal{F}_{n+1} \subset \mathcal{F}$.

Définition 9.1.1. *Une suite $\{X_n,\ n \geq 0\}$ de v. a. r. est une martingale si*

1. *Pour tout $n \geq 0$, X_n est $\mathcal{F}_n$–mesurable et intégrable,*
2. *Pour tout $n \geq 0$, $\mathbb{E}(X_{n+1}|\mathcal{F}_n) = X_n$ p. s.*

Une sousmartingale est une suite qui vérifie la 1ère condition, et $\mathbb{E}(X_{n+1}|\mathcal{F}_n) \geq X_n$. Une surmartingale est une suite qui vérifie la premère condition et $\mathbb{E}(X_{n+1}|\mathcal{F}_n) \leq X_n$.

Remarque 9.1.2. *On devrait dire dans la définition $(\mathcal{F}_n)$–martingale. Notons que si $\{X_n\}$ est une $(\mathcal{F}_n)$–martingale, c'est à fortiori une $(\mathcal{G}_n)$–martingale, avec*

$$\mathcal{G}_n = \sigma(X_0, X_1, \ldots, X_n),\ n \geq 0.$$

Exemple 1 : Marche aléatoire Soit X_0, $\{Y_k,\ k \geq 1\}$ des v. a. r. indépendantes, les Y_k étant centrées. Alors

$$X_n = X_0 + Y_1 + \cdots + Y_n,\ n \geq 1$$

définit une $(\mathcal{G}_n)$–martingale, avec $\mathcal{G}_n = \sigma(X_0, Y_1, \ldots, Y_n)$.

Si au lieu d'être centrées, les Y_k vérifient $\mathbb{E}(Y_k) \geq 0$ (resp. $\mathbb{E}(Y_k) \leq 0$), alors $\{X_n\}$ est une sous– (resp. une sur–) martingale.

Exemple 2 : Procesus de branchement On considère le processus de branchement dit de Bienaymé–Galton–Watson, défini comme suit. $\{Z_n,\ n \geq 0\}$ désigne la taille d'une population aux générations $n = 0$, 1,.. On suppose que

$$Z_{n+1} = \sum_{k=1}^{Z_n} \xi_{n,k},$$

où $\xi_{n,k}$ désigne le nombre d'enfants à la génération $n+1$ du k–ième inidvidu de la n–ième génération. On suppose que les v. a. $\{\xi_{n,k},\ n \geq 1, k \geq 1\}$ sont i. i. d. On pose $m = \mathbb{E}(\xi)$ le nombre moyen d'enfant de chaque individu. Alors le processus $\{X_n,\ n \geq 0\}$ défini par $X_n = m^{-n} Z_n$ est une martingale.

Exemple 3 : Stratégie de jeu A l'instant $n-1$, au vu des résultats des coups déjà joués, le joueur mise une somme W_n. Si le tirage le désigne comme vainqueur, il gagne W_n, si le tirage le désigne comme perdant, il perd W_n. Autrement dit, si $\Delta_n = +1$ si le joueur gagne au n–ième coup, -1 si le joueur perd au n–ième coup, la richesse du joueur après n coups est

$$X_n = X_0 + W_1\Delta_1 + \cdots W_n\Delta_n,$$

où X_0 est la richesse initiale, et les Δ_n sont i. i. d., globalement indépendantes de X_0. On pose $\mathcal{F}_n = \sigma(X_0, Y_1, \dots, Y_n)$ et on suppose que pour tout $n \geq 1$, W_n est $\mathcal{F}_{n-1}$ mesurable, alors que les hypothèses ci–dessus font que Δ_n est indépendant de $\mathcal{F}_{n-1}$.

Si le jeu est juste, i. e. si $\mathbb{P}(\Delta_n = 1) = 1/2$, alors $\mathbb{E}(X_n) = 0$, et dans ce cas

$$\begin{aligned}\mathbb{E}(X_n|\mathcal{F}_{n-1}) &= X_{n-1} + W_n\mathbb{E}(\Delta_n)\\ &= X_{n-1},\end{aligned}$$

donc $\{X_n,\ n \geq 0\}$ est une martingale.

Si le jeu n'est pas juste (par exemple, la roulette au casino a 18 cases rouges, 18 noires et 2 vertes, donc en pariant sur rouge ou noir, on a 18/38¡1/2 chances de gagner), $\mathbb{E}(\Delta_n) < 0$, et alors $\{X_n,\ n \geq 0\}$ est une surmartingale.

Notons que si $\{X_n,\ n \geq 0\}$ est une martingale,

$$\begin{aligned}\mathbb{E}(X_{n+k}|\mathcal{F}_n) &= X_n,\ \forall n \geq 0, k \geq 1,\\ \mathbb{E}(X_n) &= \mathbb{E}(X_0).\end{aligned}$$

Il résulte de l'inégalité de Jensen pour l'espérance conditionnelle que

Proposition 9.1.3. *Si $\{X_n,\ n \geq 0\}$ est une martingale, $\varphi : \mathbb{R} \to \mathbb{R}$ une fonction convexe telle que $\varphi(X_n)$ soit intégrable pour tout $n \geq 0$, alors $\{\varphi(X_n),\ n \geq 0\}$ est une sousmartingale. La même conclusion est encore vraie lorque l'on suppose seulement que $\{X_n,\ n \geq 0\}$ est une sousmartingale, à condition d'ajouter l'hypothèse que la fonction φ est croissante.*

9.2 Temps d'arrêt

Définition 9.2.1. *On appelle temps d'arrêt une v. a. τ à valeurs dans $\mathbb{N} \cup \{+\infty\}$ qui vérifie $\{\tau = n\} \in \mathcal{F}_n$, pour tout $n \geq 0$.*

Un v. a. constante $\tau \equiv n$ est un temps d'arrêt. Si $\{X_n,\ n \geq 0\}$ est une suite adaptée à la filtration $\{\mathcal{F}_n\}$, i. e. telle que X_n est $\mathcal{F}_n$–mesurable pour tout $n \geq 0$, et si $A \in \mathcal{B}$, alors

$$\tau = \begin{cases} \inf\{n,\ X_n \in A\}, & \text{si un tel } n \text{ existe,} \\ +\infty, & \text{sinon,} \end{cases}$$

est un temps d'arrêt, puisque alors

$$\{\tau = n\} = \{X_0 \notin A, X_1 \notin A, \dots, X_{n-1} \notin A, X_n \in A\}.$$

Notons qu'avec la convention que l'inf d'un ensemble vide vaut $+\infty$, ce temps d'arrêt se définit plus simplement par

$$\tau = \inf\{n,\ X_n \in A\}.$$

Si τ est un temps d'arrêt, on définit la tribu

$$\mathcal{F}_\tau = \{A \in \mathcal{F},\ A \cap \{\tau = n\} \in \mathcal{F}_n, \forall n \in \mathbb{N}\}.$$

Si τ_1 et τ_2 sont deux temps d'arrêt, alors $\inf(\tau_1, \tau_2)$ est encore un temps d'arrêt. En outre, si $\tau_1 \leq \tau_2$ p. s., alors $\mathcal{F}_{\tau_1} \subset \mathcal{F}_{\tau_2}$ (exercice).

On a le théorème d'arrêt :

Théorème 9.2.2. *Si $\{X_n,\ n \geq 0\}$ est une martingale (resp. une sousmartingale), et τ_1 et τ_2 sont deux temps d'arrêt tels que $\tau_1 \leq \tau_2 \leq N$ p. s., alors X_{τ_i} est $\mathcal{F}_{\tau_i}$ mesurable et intégrable, $i = 1, 2$ et en outre*

$$\begin{aligned} \mathbb{E}(X_{\tau_2} | \mathcal{F}_{\tau_1}) &= X_{\tau_1} \\ (\textit{resp.}\ \ \mathbb{E}(X_{\tau_2} | \mathcal{F}_{\tau_1}) &\geq X_{\tau_1}). \end{aligned}$$

PREUVE Pour tout $A \in \mathcal{B}$, $n \geq 0$,

$$\{X_{\tau_i} \in A\} \cap \{\tau_i = n\} = \{X_n \in A\} \cap \{\tau_i = n\} \in \mathcal{F}_n,$$

et en outre

$$|X_{\tau_i}| \leq \sum_{k=1}^{N} |X_k|,$$

donc la première partie de l'énoncé est établie.

Soit $A \in \mathcal{F}_{\tau_1}$. Alors

$$A \cap \{\tau_1 < k \leq \tau_2\} = A \cap \{\tau_1 \leq k-1\} \cap \{\tau_2 \leq k-1\}^c \in \mathcal{F}_{k-1}.$$

En effet, d'une part

$$A \cap \{\tau_1 \leq k-1\} = \cup_{j=1}^{k-1} A \cap \{\tau_1 = j\} \ \in \mathcal{F}_{k-1}$$

$$\text{et aussi } \ \{\tau_2 \leq k-1\}^c \in \mathcal{F}_{k-1}.$$

Posons $\Delta_k = X_k - X_{k-1}$. On a

$$\begin{aligned}
\int_A (X_{\tau_2} - X_{\tau_1}) d\mathbb{P} &= \int_A \sum_{k=1}^{n} \mathbf{1}_{\{\tau_1 < k \leq \tau_2\}} \Delta_k d\mathbb{P} \\
&= \sum_{k=1}^{n} \int_{A \cap \{\tau_1 < k \leq \tau_2\}} \Delta_k d\mathbb{P} \\
&= 0
\end{aligned}$$

ou bien ≥ 0, suivant que $\{X_n, \ n \geq 0\}$ est une martingale ou une sousmartingale. □

9.3 La ruine du joueur

On reprend l'exemple 3 ci–dessus. Supposons que le joueur parie la même somme à chaque coup, disons 1 Euro, jusqu'à ce qu'il arrête de jouer, soit

$$W_n = \mathbf{1}_{\{n \leq \tau\}}, \quad \text{où } \tau \ \text{ est un temps d'arrêt.}$$

Alors

$$\begin{aligned}
X_n &= X_0 + W_1 \Delta_1 + \cdots + W_n \Delta_n \\
&= Y_{n \wedge \tau},
\end{aligned}$$

si $\{Y_n,\ n \geq 0\}$ désigne la marche aléatoire $Y_n = X_0 + \Delta_1 + \cdots + \Delta_n$. Typiquement, le joueur dispose d'une richesse initiale $X_0 = a > 0$, et il cherche à atteindre une richesse donée $c > a$, mais il est obligé de s'arrêter s'il n'a plus d'argent, autrement dit

$$\tau = \inf\{n,\ X_n = 0 \text{ ou } X_n = c\}.$$

On s'intéresse à calculer la probabilité que le joueur atteigne son objectif, i. e.

$$s(a,c) = \mathbb{P}(\text{succès}) = \mathbb{P}(X_\tau = c) = 1 - \mathbb{P}(\text{ruine}) = 1 - \mathbb{P}(X_\tau = 0).$$

Dans le cas d'un jeu juste, $\mathbb{E}\Delta_n = 0$, et $\{Y_n,\ n \geq 0\}$, donc aussi par le théorème d'arrêt $\{X_n = Y_{n\wedge\tau},\ n \geq 0\}$ est une martingale, donc

$$\mathbb{E}(Y_{n\wedge\tau}) = \mathbb{E}(X_0).$$

Mais $X_n \to X_\tau$ quand $n \to \infty$, et X_n est bornée, puisque $0 \leq X_n \leq c$. Donc par convergence dominée,

$$\mathbb{E}(X_\tau) = cs(a,c) = a,$$

soit

$$s(a,c) = a/c.$$

Dans le cas où $\mathbb{P}(\Delta_n = 1) = 1 - \mathbb{P}(\Delta_n = -1) = p < 1/2$,

$$s(a,c) = \frac{\rho^a - 1}{\rho^c - 1}, \quad \text{où } \rho = (1-p)/p.$$

Exemple 9.3.1. *Un joueur veut gagner 1 000 Euros à partir de 900 Euros.*

$$\textit{Si } p = 1/2, \quad s(900, 1000) = 0,9,$$
$$\textit{Si } p = 18/38 \textit{ (roulette)}, \ s(900, 1000) = 3 \times 10^{-5}.$$

Exemple 9.3.2. *Un joueur veut gagner 20 000 Euros à partir de 100 Euros.*

$$\textit{Si } p = 1/2, \quad s(100, 20000) = 5 \times 10^{-3},$$
$$\textit{Si } p = 18/38 \textit{ (roulette)}, \ s(100, 20000) = 3 \times 10^{-911}.$$

9.4 Inégalités

On a l'inégalité de Doob :

Proposition 9.4.1. *Si $X_1, \ldots, X_n$ est une sous–martingale, pour tout $\alpha > 0$,*

$$\mathbb{P}(\max_{1\le i\le n} X_i \ge \alpha) \le \frac{1}{\alpha}\mathbb{E}(X_n^+).$$

PREUVE On définit le temps d'arrêt $\tau = \inf\{0 \le k \le n,\ X_k \ge \alpha\}$ et on pose $M_k = \max_{1\le i\le k} X_i$. On a

$$\{M_n \ge \alpha\} \cap \{\tau \le k\} = \{M_k \ge \alpha\} \in \mathcal{F}_k.$$

Donc $\{M_n \ge \alpha\} \in \mathcal{F}_\tau$. Par le théorème d'arrêt,

$$\begin{aligned}\alpha\mathbb{P}(M_n \ge \alpha) &\le \int_{\{M_n\ge\alpha\}} X_\tau d\mathbb{P} \\ &\le \int_{\{M_n\ge\alpha\}} X_n d\mathbb{P} \\ &\le \int_{\{M_n\ge\alpha\}} X_n^+ d\mathbb{P} \\ &\le \mathbb{E}(X_n^+).\end{aligned}$$

□

Soit maintenant $\alpha < \beta$ deux réels, et $X_1, \ldots, X_n$ des v. a. r. On définit par récurrrence les temps d'arrêt suivants. $\tau_0 = 0$ et

$$\begin{aligned}\text{Pour } k \text{ impair, } \tau_k &= \inf\{\tau_{k-1} < j \le n,\ X_j \le \alpha\}; \\ \text{Pour } k \text{ pair, } \tau_k &= \inf\{\tau_{k-1} < j \le n,\ X_j \ge \beta\}.\end{aligned}$$

Le nombre U de montées de la suite $X_1, \ldots, X_n$ à travers $[\alpha, \beta]$ est défini par

$$U = \max\{i,\ X_{\tau_{2i-1}} \le \alpha < \beta \le X_{\tau_{2i}}\}.$$

On a le

Théorème 9.4.2. *Si $\{X_1, \ldots, X_n\}$ est une sousmartingale,*

$$\mathbb{E}(U) \le \frac{\mathbb{E}[(X_n-\alpha)^+]}{\beta-\alpha} \le \frac{\mathbb{E}(X_n^+)+|\alpha|}{\beta-\alpha}.$$

PREUVE Posons $Y_k = (X_k - \alpha)^+$, $\theta = \beta - \alpha$. $\{Y_1, \dots, Y_n\}$ est encore une sousmartingale. Les temps d'arrêt τ_k sont inchangés si l'on remplace $\{X_j \leq \alpha\}$ par $\{Y_j = 0\}$, $\{X_j \geq \beta\}$ par $\{Y_j \geq \theta\}$. Donc U est le nombre de montées de la suite $(Y_1, \dots, Y_n)$ à travers $[0, \theta]$. Puisque les τ_k sont strictment croissants tant qu'ils n'atteignent pas n, clairement $\tau_n = n$ p. s. Donc

$$\begin{aligned}
Y_n &= Y_{\tau_n} \\
&\geq Y_{\tau_n} - Y_{\tau_1} \\
&= \sum_{2 \leq k \leq n,\ k \text{ pair}} (Y_{\tau_k} - Y_{\tau_{k-1}}) + \sum_{2 \leq k \leq n,\ k \text{ impair}} (Y_{\tau_k} - Y_{\tau_{k-1}})
\end{aligned}$$

Le premier terme du membre de droite ci–dessus est minoré par θU, tandis que par le théorème d'arrêt, l'espérance du second terme est non négative. Donc $\mathbb{E} Y_n \geq \theta \mathbb{E} U$, soit encore

$$(\beta - \alpha)\mathbb{E}U \leq \mathbb{E}[(X_n - \alpha)^+].$$

□

9.5 Théorème de convergence

Théorème 9.5.1. *Soit $\{X_n,\ n \geq 1\}$ une sousmartingale. Si $K = \sup_n \mathbb{E}(X_n^+) < \infty$, alors il existe une v. a. r. intégrable X telle que $X_n \to X$ p. s.*

PREUVE Pour tout $-\infty < \alpha < \beta < \infty$, désignons par $U_n(\alpha, \beta)$ (resp. $U(\alpha, \beta)$) le nombre de montées de la suite $(X_1, \dots, X_n)$ (resp. de la suite $\{X_n,\ n \geq 1\}$) à travers $[\alpha, \beta]$. On a $U_n(\alpha, \beta) \uparrow U(\alpha, \beta)$ quand $n \to \infty$. Mais d'après le théorème 9.4.2 et l'hypothèse,

$$\mathbb{E}(U_n(\alpha, \beta) \leq \frac{\mathbb{E}(X_n^+) + |\alpha|}{\beta - \alpha} \leq \frac{K + |\alpha|}{\beta - \alpha},\ \forall n \geq 1,$$

donc $U(\alpha, \beta) < \infty$ p. s. Posons

$$X^* = \limsup_{n \to \infty} X_n, \quad X_* = \liminf_{n \to \infty} X_n.$$

On a clairement

$$\{X_* < X^*\} = \cup_{\alpha<\beta,\ \alpha,\beta\in\mathbb{Q}}\{U(\alpha,\beta) = +\infty\}$$

Donc $X_* = X^*$ p. s., et $X_n \to X$ p. s. dans $\overline{\mathbb{R}}$. Mais

$$\begin{aligned} \mathbb{E}|X_n| &= 2\mathbb{E}(X_n^+) - \mathbb{E}(X_n) \\ &\leq 2\mathbb{E}(X_n^+) - \mathbb{E}(X_0), \\ \text{donc} \quad \sup_n \mathbb{E}|X_n| &< \infty, \end{aligned}$$

et alors par le Lemme de Fatou $\mathbb{E}|X| < \infty$. Donc $X \in \mathbb{R}$ p. s., et en outre X est intégrable. □

Corollaire 9.5.2. *Soit $\{X_n,\ n \geq 1\}$ une surmartingale à valeurs dans $\mathbb{R}_+$ ou une martingale bornée soit par en dessus soit par en dessous. Alors X_n converge p. s. vers une limite intégrable X, quand $n \to \infty$.*

Preuve Dans le cas d'une surmartingale positive $\{X_n\}$, $\{-X_n\}$ est une sousmartingale dont la partie positive est nulle, donc le Théorème s'applique. Dans le cas d'une martingale $\{X_n\}$ bornée par en dessous, il existe c tel que $\{X_n + c\}$ est une martingale positive, à laquelle la première partie de ce Corollaire s'applique. Le dernier cas se ramène à celui–ci par multiplication par -1. □

Lemme 9.5.3. *Si Z est une v. a. r. intégrable, et $\{\mathcal{F}_n,\ n \geq 1\}$ une collection de tribus, alors la suite de v. a. $\{\mathbb{E}(Z|\mathcal{F}_n),\ n \geq 1\}$ est uniformément intégrable.*

Preuve D'après la Proposition 7.8.4, il existe $H : \mathbb{R}_+ \to \mathbb{R}_+$ convexe, vérifiant $H(x)/x \to +\infty$, quand $x \to +\infty$, telle que $\mathbb{E}[H(Z)] < \infty$. Par l'inégalité de Jensen pour l'espérance conditionnelle,

$$\begin{aligned} H(\mathbb{E}(Z|\mathcal{F}_n)) &\leq \mathbb{E}(H(Z)|\mathcal{F}_n), \text{ donc} \\ \sup_n \mathbb{E}[H(\mathbb{E}(Z|\mathcal{F}_n))] &\leq \mathbb{E}[H(Z)] < \infty, \end{aligned}$$

et le résultat se déduit de la Proposition 7.8.4. □

Soit $\{\mathcal{F}_n,\ n \geq 1\}$ une filtration. Posons $\mathcal{F}_\infty = \sigma(\cup_n \mathcal{F}_n)$. On a alors la

Proposition 9.5.4. *Si Z est une v. a. r. intégrable,*

$$\mathbb{E}(Z|\mathcal{F}_n) \to \mathbb{E}(Z|\mathcal{F}_\infty) \text{ p. s. et dans } L^1, \text{ quand } n \to \infty.$$

Preuve La suite $\{X_n,\ n \geq 1\}$ définie par $X_n = \mathbb{E}(Z|\mathcal{F}_n)$ est une martingale, qui est uniformément intégrable d'après le Lemme. En particulier, $\mathbb{E}|X_n| \leq \mathbb{E}|Z|$, et d'après le Théorème 9.5.1, il existe X v. a. r. $\mathcal{F}_\infty$–mesurable et intégrable telle que $X_n \to X$ p. s. Fixons $k \geq 1$. Quelque soit $A \in \mathcal{F}_\infty$, par l'uniforme intégrabilité,

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

Mais si $A \in \mathcal{F}_k$ et $n \geq k$,

$$\mathbb{E}[X_n; A] = \mathbb{E}[Z; A].$$

Donc $\mathbb{E}[Z; A] = \mathbb{E}[X; A]$, pour tout $A \in \cup_k \mathcal{F}_k$. Or $\cup_k \mathcal{F}_k$ est une algèbre qui engendre la tribu $\mathcal{F}_\infty$. Donc $X = \mathbb{E}(Z|\mathcal{F}_\infty)$. □

9.6 Convergence des martingales à l'envers

Le but de cette section est de montrer le

Théorème **9.6.1.** *Soit $\{\mathcal{G}_n,\ n \geq 1\}$ une suite décroissante de sous–tribus de $\mathcal{F}$. On pose $\mathcal{G} = \cap_n \mathcal{G}_n$. Alors pour toute v. a. intégrable Z, $\mathbb{E}(Z|\mathcal{G}_n) \to \mathbb{E}(Z|\mathcal{G})$ p. s.*

Posons tout d'abord la définition

Définition 9.6.2. *Etant donné une suite croissante $\{\mathcal{F}_n,\ n \in \mathbb{Z}_-\}$ de sous–tribus de $\mathcal{F}$, la suite $\{X_n,\ n \in \mathbb{Z}_-\}$ est appelée "martingale à l'envers" si X_n est $\mathcal{F}_n$–mesurable pour tout $n \in \mathbb{Z}_-$ et pour $n < -1$,*

$$\mathbb{E}(X_{n+1}|\mathcal{F}_n) = X_n.$$

On a le

Théorème **9.6.3.** *Si $\{X_n,\ n \in \mathbb{Z}_-\}$ est une martingale à l'envers, il existe une v. a. intégrable X telle que $X_n \to X$ p. s. et dans L^1 quand $n \to -\infty$.*

Preuve On raisonne de façon analogue à ce que nous avons fait dans le cas des martingales usuelles. On pose

$$X_* = \liminf_{n\to-\infty} X_n, \quad X^* = \limsup_{n\to-\infty} X_n,$$

et on note que

$$\{X_* < X^*\} = \cup_{\alpha<\beta\ \alpha,\beta\in\mathbb{Q}}\{X_* < \alpha < \beta < X^*\}.$$

Soit U_n le nombre de montées de la suite $X_{-n}, \dots, X_{-1}$ à travers $[\alpha, \beta]$. On a que

$$\mathbb{E}[U_n] \le \frac{\mathbb{E}|X_{-1}| + |\alpha|}{\beta - \alpha}.$$

Donc si $U := \lim_{n\to\infty} U_n$, $\mathbb{E}(U) < \infty$, donc $U < \infty$ p. s. et il s'ensuit que $X_* = X^*$ p. s. En outre pour tout $n > 1$,

$$X_{-n} = \mathbb{E}[X_{-1}|\mathcal{F}_{-n}],$$

donc par le Lemme 9.5.3 la suite $\{X_n,\ n \in \mathbb{Z}_-\}$ est uniformément intégrable, d'où la convergence dans L^1. □

On peut maintenant passer à la

Peuve du Théorème 9.6.1 Le théorème 9.6.3 montre que

$$X_{-n} := \mathbb{E}[Z|\mathcal{G}_n] \to X \text{ p. s. et dans } L^1.$$

En outre X est $\mathcal{G}_n$–mesurable pour tout n, donc X est $\mathcal{G}$–mesurable et pour tout $A \in \mathcal{G}$,

$$\begin{aligned}\mathbb{E}[X; A] &= \lim_{n\to\infty} \mathbb{E}[\mathbb{E}(Z|\mathcal{G}_n); A] \\ &= \mathbb{E}[\mathbb{E}(Z|\mathcal{G}); A],\end{aligned}$$

donc $X = \mathbb{E}[Z|\mathcal{G}]$ p. s. □

Preuve de la loi forte des grands nombres comme conséquence du théorème de convergence des martingales à l'envers

Soit $\{X_n,\ n \ge 1\}$ des v. a. i. i. d. intégrables. On pose $S_n := X_1 + \cdots + X_n$, $\mathcal{F}_{-n} := \sigma\{S_n, S_{n+1}, S_{n+2}, \dots\}$, $M_n := \mathbb{E}(X_1|\mathcal{F}_{-n})$. Alors $\{M_n,\ n \ge 1\}$ est une martingale à l'envers, et donc $M_n \to X$ p. s. et dans L^1. En outre

$$\begin{aligned}M_n &= \mathbb{E}(X_1|S_n) \\ &= \mathbb{E}(X_2|S_n) \\ &= \dots \\ &= \mathbb{E}(X_n|S_n) \\ &= \frac{S_n}{n}.\end{aligned}$$

On a montré que S_n/n converge p. s. et dans L^1 vers une v. a. X telle que $\mathbb{E}(X) = \mathbb{E}(X_1)$. Mais X est $\mathcal{G}$–mesurable, où $\mathcal{G}$ est la tribu asymptotique de la suite $\{X_n\}$, donc par la loi 0–1 de Kolmogorov, $\mathcal{G}$ est triviale, et $X = \mathbb{E}(X_1)$ p. s. □

9.7 Théorème de de Finetti

Une permutation π de l'ensemble $\{1, 2, \ldots\}$ est dite finie si $|\{i,\ \pi(i) \neq i\}| < \infty$. Posons la

Définition 9.7.1. *La suite dénombrable $\{X_n,\ n \geq 1\}$ est dite échangeable si pour toute permutation finie π of $\{1, 2, \ldots\}$,*

$$(X_1, X_2, \ldots) \stackrel{\mathcal{L}}{=} (X_{\pi(1)}, X_{\pi(2)}, \ldots).$$

Il n'est pas très difficile de montrer le

Lemme 9.7.2. *Etant donné la suite dénombrable de v. a. $\{X_1, X_2, \ldots\}$, les trois propriétés suivantes sont équivalentes*

1. *La suite $\{X_1, X_2, \ldots\}$ est échangeable.*
2. *Pour tout $n > 1$,*
$$(X_1, \ldots, X_{n-1}, X_n, X_{n+1}, \ldots) \stackrel{\mathcal{L}}{=} (X_n, \ldots, X_{n-1}, X_1, X_{n+1}, \ldots).$$
3. *Pour toute suite $\{n_i,\ i \geq 1\}$ d'entiers distincts,*
$$(X_1, X_2, X_3, \ldots) \stackrel{\mathcal{L}}{=} (X_{n_1}, X_{n_2}, X_{n_3}, \ldots).$$

Commençons par établir le

Lemme 9.7.3. *Soit Y une v. a. bornée, et $\mathcal{H} \subset \mathcal{G}$ deux sous–tribus de $\mathcal{F}$. Alors $\mathbb{E}\left[\mathbb{E}(Y|\mathcal{H})^2\right] = \mathbb{E}\left[\mathbb{E}(Y|\mathcal{G})^2\right]$ (ou a fortiori $\mathbb{E}(Y|\mathcal{H}) \stackrel{\mathcal{L}}{=} \mathbb{E}(Y|\mathcal{G})$) entraîne que $\mathbb{E}(Y|\mathcal{H}) = \mathbb{E}(Y|\mathcal{G})$ p. s.*

Preuve Le résultat découle de l'identité

$$\mathbb{E}\left[(\mathbb{E}(Y|\mathcal{G}) - \mathbb{E}(Y|\mathcal{H}))^2\right] = \mathbb{E}\left[(\mathbb{E}(Y|\mathcal{G}))^2\right] - \mathbb{E}\left[(\mathbb{E}(Y|\mathcal{H}))^2\right].$$

□

Nous pouvons maintenant énoncer le théorème de de Finetti

Théorème **9.7.4.** *Une suite infinie dénombrable et échangeable de v. a. $\{X_n,\ n \geq 1\}$ est un mélange de v. a. i. i. d., au sens oùconditionnellement en la tribu asymptotique de la suite $\{X_n\}$, les X_n sont i. i. d.*

Preuve Pour tout $n \geq 0$, soit $\mathcal{G}_n := \sigma(X_{n+1}, X_{n+2}, \ldots)$, et $\mathcal{G} := \cap_n \mathcal{G}_n$ la tribu asymptotique. Par échangeabilité, pour tout $n \geq 2$,

$$(X_1, X_2, X_3, \ldots) \stackrel{\mathcal{L}}{=} (X_1, X_{n+1}, X_{n+2}, \ldots).$$

Donc pour toute function borélienne bornée $\varphi : \mathbb{R} \to \mathbb{R}$, $n \geq 2$,

$$\mathbb{E}(\varphi(X_1)|\mathcal{G}_1) \stackrel{\mathcal{L}}{=} \mathbb{E}(\varphi(X_1)|\mathcal{G}_n).$$

Le Théorème 9.6.1 entraîne que

$$\mathbb{E}(\varphi(X_1)|\mathcal{G}_n) \to \mathbb{E}(\varphi(X_1)|\mathcal{G}) \quad \text{p. s., quand } n \to \infty.$$

On en déduit que

$$\mathbb{E}(\varphi(X_1)|\mathcal{G}_1) \stackrel{\mathcal{L}}{=} \mathbb{E}(\varphi(X_1)|\mathcal{G}).$$

Mais le Lemme 9.7.3 entraîne que l'égalité est vraie p. s. Il résulte que

$$X_1 \quad \text{et} \quad \mathcal{G}_1 \quad \text{sont conditionnellement indépendants sachant } \mathcal{G}.$$

Le même argument appliqué à $(X_n, X_{n+1}, \ldots)$ nous indique que pour tout $n \geq 1$,

$$X_n \quad \text{et} \quad \mathcal{G}_n \quad \text{sont conditionnellement indépendants sachant } \mathcal{G}.$$

Ceci entraîne que toute la suite $\{X_n,\ n \geq 1\}$ est conditionnellement indépendante sachant $\mathcal{G}$. Maintenant l'échangeabilité entraîne que pour tout $n \geq 1$,

$$(X_1, X_{n+1}, X_{n+2}, \ldots) \stackrel{\mathcal{L}}{=} (X_n, X_{n+1}, X_{n+2}, \ldots).$$

Donc pour les même φ que ci–dessus,

$$\mathbb{E}(\varphi(X_1)|\mathcal{G}_n) = \mathbb{E}(\varphi(X_n)|\mathcal{G}_n) \quad \text{p. s..}$$

Prenant l'espérance conditionnelle sachant $\mathcal{G}$ nous déduisons que

$$\mathbb{E}(\varphi(X_1)|\mathcal{G}) = \mathbb{E}(\varphi(X_n)|\mathcal{G}) \quad \text{p. s..}$$

Donc les lois conditionnelles sachant $\mathcal{G}$ des v. a. X_n sont identiques. $\square$

Il résulte du Théorème de de Finetti que si les X_k sont en outre intégrables, $n^{-1}\sum_{k=1}^n X_k$ converge p. s. quand $n \to \infty$. En effet

$$\mathbb{P}\left(n^{-1}\sum_{k=1}^n X_k \text{ converge}\right) = \mathbb{E}\left[\mathbb{P}\left(n^{-1}\sum_{k=1}^n X_k \text{ converge}\middle| \mathcal{G}\right)\right] = 1,$$

puisque, les $\{X_n\}$ étant i. i. d. et intégrables sachant $\mathcal{G}$, la loi forte des grands nombres s'applique conditionnellement en $\mathcal{G}$.

On déduit alors de ce qui précède le

Corollaire 9.7.5. *Soit $\{X_n,\ n \geq 1\}$ une suite dénombrable échangeable de v. a. à valeurs dans $\{0,1\}$. Conditionnellement en*

$$\lim_n n^{-1}\sum_{k=1}^n X_k = x \ \textit{p. s.},$$

les X_n sont i. i. d. Bernoulli de paramètre x.

9.8 Théorème central limite pour les martingales

9.9 Exercices

Exercice 9.9.1. *Soit $\Delta_1, \Delta_2, \ldots$ des v. a. indépendantes, intégrables et de moyenne nulle. Posons $X_1 = \Delta_1$, et pour $n \geq 1$, $X_{n+1} = X_n + \Delta_n f_n(X_1, \ldots, X_n)$. Supposons que les fonctions f_n sont mesurables, et telles que chaque X_n est intégrable. Montrer que $\{X_n,\ n \geq 1\}$ est une martingale.*

Exercice 9.9.2. *Les Δ_n étant comme à l'exercice précédent, et de plus toutes de variance σ^2, on pose cette fois $X_n = \left(\sum_{k=1}^n \Delta_k\right)^2 - n\sigma^2$ pour $n \geq 1$. Montrer que $\{X_n,\ n \geq 1\}$ est une martingale.*

Exercice 9.9.3. *On suppose maintenant que les Δ_k sont i.i.d., avec $\mathbb{P}(\Delta_k = 1) = \mathbb{P}(\Delta_k = -1) = 1/2$, et on pose $X_0 = 0$, $X_n = \sum_{k=1}^n \Delta_k$ pour $n \geq 1$, et pour tout $a \in \mathbb{Z}$, $T_a = \inf\{n \geq 1,\ X_n = a\}$. Pour $a, b \in \mathbb{Z}$, $a < 0 < b$, on pose enfin $S_{a,b} = T_a \wedge T_b$.*

1. *Montrer que* $S_{a,b} < \infty$ *p.s. (Indication : on pourra remarquer que* $\{S_{a,b} > n(b-a)\} \subset \cap_{k=1}^{n}\{|\Delta_{(k-1)(b-a)+1} + \cdots + \Delta_{k(b-a)}| < b-a\}$*).*
2. *Montrer que* $\mathbb{E}[X_{S_{a,b}\wedge n}] = 0$ *pour tout* $n \geq 1$*, et que* $\mathbb{E}[X_{S_{a,b}}] = 0$*. En déduire les valeurs de* $\mathbb{P}(T_a < T_b)$ *et* $\mathbb{P}(T_b < T_a)$*.*
3. *Montrer que* $\mathbb{E}[X^2_{S_{a,b}\wedge n}] = \mathbb{E}[S_{a,b} \wedge n]$ *(on pourra utiliser l'exercice précédent), et calculer* $\mathbb{E}[S_{a,b}]$*.*
4. *Montrer que* $\mathbb{P}(T_a < T_n) \to 1$ *lorsque* $n \to \infty$*. En déduire que* $\mathbb{P}(T_a < \infty) = 1$*. Montrer que* $\mathbb{E}[T_a] = +\infty$*.*

Exercice 9.9.4. *Soit* $\{X_n,\ n \geq 1\}$ *une martingale telle que* $|X_1(\omega)| \leq c$ *et* $|X_n(\omega) - X_{n-1}(\omega)| \leq c$ *pour tout* $n \geq 2$ *et pour tout* ω*. Soit* T *un temps d'arrêt d'espérance finie. Montrer que* X_T *est intégrable,* $\{k \leq T\} \in \mathcal{F}_{k-1}$ *et* $\mathbb{E}[X_T] = \mathbb{E}[X_1]$*. En déduire une nouvelle démonstration du fait que, dans l'exercice précédent,* $\mathbb{E}[T_a] = +\infty$*.*

Exercice 9.9.5. *Soit* $\{X_n,\ n \geq 1\}$ *une martingale qui est bornée soit par en dessus soit par en dessous par une constante. En déduire que* $\sup_{n\geq 1} \mathbb{E}[|X_n|] < \infty$*.*

Exercice 9.9.6. *Soit* $\{Y_n,\ n \geq 1\}$ *des v.a. positives i.i.d., qui vérifient* $\mathbb{E}[Y_n] = 1$*. On pose, pour* $n \geq 1$*,* $X_n = Y_1 Y_2 \times \cdots \times Y_n$*.*

1. *Montrer que* X_n *est une martingale qui converge p.s. vers une variable aléatoire* X *qui vérifie* $\mathbb{E}[X] \leq \prod_{n=1}^{\infty} \mathbb{E}[Y_n]$*.*
2. *On suppose maintenant que chaque* Y_n *prend les deux valeurs* $1/2$ *et* $3/2$*, chacune avec la probabilité* $1/2$*. Montrer qu'alors* $X = 0$ *p.s.*

Exercice 9.9.7. *Soit* $\{X_n,\ n \geq 1\}$ *une martingale telle que* $\mathbb{E}[X_1] = 0$ *et* $\mathbb{E}[X_n^2] < \infty$ *pour tout* $n \geq 1$*.*

1. *Montrer que*

$$\mathbb{E}\left[(X_{n+k} - X_n)^2\right] = \sum_{\ell=1}^{k} \mathbb{E}\left[(X_{n+\ell} - X_{n+\ell-1})^2\right].$$

2. *On suppose en outre que* $\sum_{n\geq 1} \mathbb{E}\left[(X_n - X_{n-1})^2\right] < \infty$*. Montrer que* X_n *converge vers une limite* X *p.s. et en moyenne quadratique quand* $n \to \infty$*.*

Exercice 9.9.8. *(Formules de Wald – suite) On reprend les notations et hypothèses de l'exercice 4.8.24, mais on ne suppose plus que N et les v.a. X_n sont indépendantes, mais plutôt que N est un temps d'arrêt de la filtration $\mathcal{F}_n$ définie par $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$, $n \geq 1$. Montrer que les formules établies à l'exercice 4.8.24 restent vraies. On commencera par montrer que $\{N \geq n\} \in \mathcal{F}_{n-1}$, et donc que cet événement est indépendant de X_n.*

On pourra en particulier en déduire que dès que les X_n sont intégrables, le processus

$$V_n = \sum_{k=1}^{n} \mathbf{1}_{\{k \leq N\}}[X_k - \mathbb{E}(X_1)], \quad n \geq 1$$

est une martingale.

Exercice 9.9.9. *Inégalité de Khinchine*

1. *Soit $Y \simeq N(0,1)$ et $X = sgn(Y)$. Montrer que $X = c\,\mathbb{E}(Y|sign(Y))$, avec $c = \sqrt{2/\pi}$.*
2. *Soit $\{Y_n,\ n \geq 1\}$ des v.a.r. i.i.d., toutes de loi $N(0,1)$. On pose $\mathcal{G} = \sigma\{sgn(Y_n),\ n \geq 1\}$. Montrer que les v.a.r.*

$$X_n = c\,\mathbb{E}(Y_n|\mathcal{G}),\ n \geq 1$$

sont i.i.d., avec

$$\mathbb{P}(X_n = 1) = \mathbb{P}(X_n = -1) = 1/2.$$

3. *Soit $\{a_n,\ n \geq 1\}$ des nombres réels tels que $\sum_{j=1}^{\infty} a_j^2 < \infty$. Montrer que la série $\sum_{j=1}^{\infty} a_j Y_j$ converge p.s. et dans tous les $L^p(\Omega)$, et préciser la loi de cette somme.*
4. *En déduire que la série $\sum_{j=1}^{\infty} a_j X_j$ converge dans tous les $L^p(\Omega)$ et que*

$$\mathbb{E}\left(\left|\sum_{j=1}^{\infty} a_j X_j\right|^p\right) \leq c^p \mathbb{E}\left(\left|\sum_{j=1}^{\infty} a_j Y_j\right|^p\right).$$

5. *Montrer l'inégalité de Khinchine*

$$\mathbb{E}\left(\left|\sum_{j=1}^{\infty} a_j X_j\right|^p\right) \leq c^p \left(\sum_{j=1}^{\infty} a_j^2\right)^{p/2} \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} |t|^p e^{-t^2/2} dt.$$

6. *Montrer que si* $2\lambda c^2 \sum_{j=1}^{\infty} a_j^2 < 1$, *alors*

$$\mathbb{E}\left[\exp\left(\lambda\left|\sum_{j=1}^{\infty} a_j X_j\right|^2\right)\right] < \infty.$$

Bibliographie

[1] Patrick Billingsley : *Probability and measure*, third ed., Wiley Series in Probability and Mathematical Statistics, John Wiley & Sons Inc., New York, 1995.

[2] Jean Jacod, Philip Protter : *L'essentiel en théorie des probabilités*, Cassini, 2002.

[3] Walter Rudin : *Analyse réelle et complexe*, Masson.

Index