

Intégration et Probabilités

É. Pardoux

13 novembre 2009

Table des matières

1	Mesure	5
1.1	Rappels sur les limites de réels	5
1.2	Opérations sur les ensembles	8
1.3	Espace mesurable	9
1.4	Mesure	11
1.5	Mesure de Lebesgue	16
1.6	Mesures sur $(\mathbb{R}^d, \mathcal{B}_d)$	19
1.7	Applications mesurables	21
2	Intégration	25
2.1	Propriété vérifiée presque partout	25
2.2	Intégrale des fonctions non négatives	26
2.3	Intégrale des fonctions de signe quelconque.	31
2.4	Mesure produit et Théorème de Fubini	39
3	Espaces L^p	45
3.1	Définition des espaces L^p	45
3.2	Propriétés des espaces $L^p(\mu)$	48
3.3	Théorème de Radon–Nikodym et dualité des espaces L^p	52
3.4	Compléments sur la théorie de l’intégration	58
3.4.1	Théorème de représentations de Riesz	58
3.4.2	Intégrale de Stieltjes–Lebesgue	58
3.4.3	Théorème de différentiation de Lebesgue	60
4	Probabilité-Indépendance-Variable aléatoire	63
4.1	Probabilité	63
4.2	Probabilités conditionnelles	64
4.3	Événements indépendants	65

4.4	Variable aléatoire	68
4.5	Exemples de lois de probabilité	71
4.5.1	Lois discrètes	71
4.5.2	Lois à densité	72
4.6	Variables aléatoires indépendantes	74
4.7	Moments des variables aléatoires	76
5	Fonction caractéristique et loi de Gauss	81
5.1	Fonction caractéristique d'une loi de probabilité sur \mathbb{R}	81
5.2	Fonction caractéristique d'une loi de probabilité sur \mathbb{R}^d	87
5.3	Vecteurs aléatoires gaussiens	90
6	Convergence des variables aléatoires et loi des grands nombres	95
6.1	Théorème d'extension de Kolmogorov	95
6.2	Convergence des variables aléatoires	96
6.3	La loi faible des grands nombres	101
6.4	Loi 0 – 1 de Kolmogorov	101
6.5	Convergence des séries	103
6.6	Loi forte des grands nombres	105
7	La convergence en loi	109
7.1	Définition et premières propriétés	109
7.2	Relation avec les autres types de convergence, et propriétés supplémentaires.	115
7.3	Convergence en loi et fonctions caractéristiques	118
7.4	Le Théorème Central Limite	122
7.5	Uniforme intégrabilité	123
8	Espérance et Probabilité Conditionnelle	127
8.1	Introduction	127
8.2	Espérance conditionnelle par rapport à une σ -algèbre	128
8.2.1	Définition de l'espérance conditionnelle	129
8.2.2	Propriétés de l'espérance conditionnelle	131
8.3	Espérance conditionnelle par rapport à une variable aléatoire	135
8.4	Probabilité conditionnelle	137

Chapitre 1

Mesure

Un **espace mesuré** est un triplé $(\Omega, \mathcal{F}, \mu)$ 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 (Ω, \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 $\{x_n, n \in \mathbb{N}\}$ une suite quelconque de nombres réels, les quantités suivantes sont toujours définies dans $\overline{\mathbb{R}}$:

$$\begin{aligned}\limsup_{n \rightarrow \infty} x_n &= \limsup_{n \rightarrow \infty} x_n = \lim_{n \rightarrow \infty} \downarrow [\sup_{p \geq n} x_p] \\ \liminf_{n \rightarrow \infty} x_n &= \liminf_{n \rightarrow \infty} x_n = \lim_{n \rightarrow \infty} \uparrow [\inf_{p \geq n} x_p]\end{aligned}$$

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

$$\begin{aligned}\limsup_{n \rightarrow \infty} x_n &= \bigwedge_{n \in \mathbb{N}} \bigvee_{p \geq n} x_p \\ \liminf_{n \rightarrow \infty} x_n &= \bigvee_{n \in \mathbb{N}} \bigwedge_{p \geq n} x_p\end{aligned}$$

Proposition 1.1.1. Une suite $\{x_n, n \in \mathbb{N}\}$ de nombres réels converge dans $\overline{\mathbb{R}} = \mathbb{R} \cup \{+\infty\} \cup \{-\infty\}$ 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 $\underline{\lim} x_n \leq \overline{\lim} x_n$.

PREUVE

a) **Supposons que x_n converge dans $\overline{\mathbb{R}}$.**

– **1er cas :** $x_n \rightarrow +\infty$ i.e. $\forall M \in \mathbb{R}, \exists n_M$ tel que

$$x_n \geq M, \forall n \geq n_M.$$

d'où

$$\inf_{p \geq n} x_p \geq M, \forall n \geq n_M.$$

donc $\underline{\lim} x_n \geq M, \forall M \in \mathbb{R}$ i.e. $\overline{\lim} x_n \geq \underline{\lim} x_n = +\infty$.

– **2ème cas :** $x_n \rightarrow -\infty$ raisonnement analogue.

– **3ème cas :** $\{x_n\}$ est une suite de Cauchy, i.e.

$$\forall \varepsilon > 0, \exists n_\varepsilon \in \mathbb{N} \text{ t.q. } \forall n, p \geq n_\varepsilon, |x_n - x_p| \leq \varepsilon.$$

Fixons tout d'abord $p \geq n_\varepsilon$. On a, $\forall n \geq n_\varepsilon$:

$$x_n \in [x_p - \varepsilon, x_p + \varepsilon].$$

donc aussi

$$\sup_{q \geq n} x_q \in [x_p - \varepsilon, x_p + \varepsilon]$$

d'où

$$\overline{\lim} x_n \in [x_p - \varepsilon, x_p + \varepsilon]$$

donc

$$x_p \in [\overline{\lim} x_n - \varepsilon, \overline{\lim} x_n + \varepsilon], \forall p \geq n_\varepsilon,$$

d'où l'on déduit comme ci-dessus : $\underline{\lim} x_n \in [\overline{\lim} x_n - \varepsilon, \overline{\lim} x_n + \varepsilon]$.

soit

$$|\overline{\lim} x_n - \underline{\lim} x_n| \geq \varepsilon.$$

Cette dernière inégalité étant vraie $\forall \varepsilon > 0$, on a $\overline{\lim} x_n = \underline{\lim} x_n$.

b) **Supposons que $\overline{\lim} x_n = \underline{\lim} x_n$.**

– **1er cas :**

$$\overline{\lim} x_n = \underline{\lim} x_n = +\infty$$

Alors $\forall M, \exists n_M$ tel que $\forall n \geq n_M$

$$\inf_{p \geq n} x_p \geq M$$

i.e.

$$x_n \geq M, \forall n \geq n_M.$$

Cela suffit à démontrer que $x_n \rightarrow +\infty$

– **2ème cas :**

$$\overline{\lim} x_n = \underline{\lim} x_n = -\infty$$

Raisonnement analogue.

– **3ème cas :**

$$\overline{\lim} x_n = \underline{\lim} x_n = x \in \mathbb{R}$$

Alors $\forall \varepsilon > 0, \exists n_\varepsilon$ tel que $\forall n \geq n_\varepsilon,$

$$\left| \sup_{p \geq n} x_p - x \right| \leq \varepsilon \text{ et } \left| \inf_{p \geq n} x_p - x \right| \leq \varepsilon$$

d'où $\sup_{p \geq n} x_p \in [x - \varepsilon, x + \varepsilon]$, et $\inf_{p \geq n} x_p \in [x - \varepsilon, x + \varepsilon]$, donc $x_p \in [x - \varepsilon, x + \varepsilon] \forall p \geq n$, et en particulier $\forall p \geq n_\varepsilon$. Donc $x_n \rightarrow x$.

□

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, \dots les parties de Ω . On définit

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

La notation $A \setminus B$ 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 :

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

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

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

1.3 Espace mesurable

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

□

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

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

□

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 $\bigcup_n A_n = \lim_n \uparrow \left(\bigcup_{p \leq n} A_p \right)$. □

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 . \square

Espace mesurable produit

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 :

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

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

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 $B \subset A \Rightarrow A \setminus B \in \mathcal{L}$
- (λ_3) $\{A_n, n \in \mathbb{N}\} \subset \mathcal{L}$ et $A_n \uparrow A \Rightarrow A \in \mathcal{L}$

Exercice 1.3.8. (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.

Théorème 1.3.9. ($\pi - \lambda$) 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. 1.3.8 (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. $\forall A \subset \Omega$, posons $\mathcal{G}_A = \{B; A \cap B \in \lambda(\mathcal{P})\}$. $A \in \lambda(\mathcal{P}) \Rightarrow \mathcal{G}_A$ est un λ -système (exercice), $A \in \mathcal{P} \Rightarrow \mathcal{P} \subset \mathcal{G}_A$, et a fortiori \mathcal{G}_A est un λ -système. Donc $A \in \mathcal{P} \Rightarrow \lambda(\mathcal{P}) \subset \mathcal{G}_A$, 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 : $B, C \in \lambda(\mathcal{P}) \Rightarrow B \cap C \in \lambda(\mathcal{P})$.

Corollaire 1.3.10. $(\pi - \lambda)$ 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 mesuré (Ω, \mathcal{F}) .

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

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

(ii) Etant donné $\{A_n, n \in \mathbb{N}\} \subset \mathcal{F}$, 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 $\bigcup_1^\infty A_n \in \mathcal{F}_0$ ”. 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 (Ω, \mathcal{F}) , sera dite σ -**finie** si $\exists \{A_n, n \in \mathbb{N}\} \subset \mathcal{F}$ telle que :

$$(i) \bigcup_0^\infty A_n = \Omega$$

(ii) $\mu(A_n) < \infty, \forall n \in \mathbb{N}$.

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 $\{A_n, n \in \mathbb{N}\} \subset \mathcal{C}$.

□

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

$$\forall A \subset \Omega, \nu(A) = \text{Card}(A).$$

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 (Ω, \mathcal{F}) .

- (i) Si $\{A_n, n \in \mathbb{N}\} \subset \mathcal{F}$, $A_n \uparrow A$, alors $\mu(A_n) \uparrow \mu(A)$
- (ii) Si $\{A_n, n \in \mathbb{N}\} \subset \mathcal{F}$, $A_n \downarrow A$, $\mu(A_1) < \infty$, alors $\mu(A_n) \downarrow \mu(A)$
- (iii) Si $\{A_n, n \in \mathbb{N}\} \subset \mathcal{F}$, $\mu \left(\bigcup_1^\infty A_n \right) \leq \sum_1^\infty \mu(A_n)$
- (iv) Si μ est σ -finie, alors \mathcal{F} ne peut pas contenir une collection non dénombrable d'ensembles disjoints de mesure non nulle.

PREUVE

- (i) Posons $B_1 = A_1$; $B_n = A_n \setminus A_{n-1}$, $n \geq 2$. Les B_n sont deux à deux disjoints, $A_n = \bigcup_1^n B_k$, $A = \bigcup_1^\infty B_k$. Donc (i) résulte de la σ -additivité.

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

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

- (iii) $\forall n$, $\mu \left(\bigcup_1^n A_k \right) \leq \sum_1^n \mu(A_k)$ [exercice] 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 $\{A_n, n \in \mathbb{N}\} \subset \mathcal{F}$, avec $\bigcup_1^\infty A_n = \Omega$, et $\mu(A_n) < \infty$, $\forall n$. On pose $\Theta_n = \{\theta \in \Theta; \mu(A_n \cap B_\theta) > 0\}$.

$$\Theta_n = \bigcup_{p=1}^\infty \Theta_n^p, \text{ avec } \Theta_n^p = \{\theta \in \Theta; \mu(A_n \cap B_\theta) > \frac{1}{p}\}. \text{ Or } \text{card}(\Theta_n^p) \leq$$

$$p \mu(A_n) < \infty. \text{ Donc } \Theta_n \text{ est au plus dénombrable, ainsi que } \Theta = \bigcup_1^\infty \Theta_n.$$

□

Remarque 1.4.5. a) (ii) n'est pas vrai sans l'hypothèse $\mu(A_1) < \infty$. [ou du moins $\exists n$ tel que $\mu(A_n) < \infty$]. Exemple : $\mu = \lambda$ = mesure de Lebesgue sur (\mathbb{R}^2, B_2) . $A_n = \{x = (x_1, x_2); 0 \leq x_2 \leq \frac{1}{n}\}$. $\lambda(A_n) = +\infty, \forall n$. Or $\lambda(A) = 0$. [cf. ci-dessous].

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. Soit μ_0 une mesure définie sur une algèbre \mathcal{F}_0 de parties de Ω . Il existe une mesure μ sur $\mathcal{F} = \sigma(\mathcal{F}_0)$, telle que $\mu|_{\mathcal{F}_0} = \mu_0$. L'extension μ de μ_0 à \mathcal{F} est unique si μ_0 est σ -finie le long de \mathcal{F}_0 .

PREUVE Existence : On définit une "mesure extérieure" sur $\mathcal{P}(\Omega)$ par : $\forall A \subset \Omega, \mu^*(A) = \inf \sum_n \mu_0(A_n)$, où l'inf est pris sur toutes les suites dénombrables $\{A_n\}$ telles que $A_n \in \mathcal{F}_0$ et $A \subset \bigcup_n A_n$.

On démontre que la restriction de μ^* à \mathcal{F}_0 est μ_0 , et que la restriction de μ^* à \mathcal{F} est une mesure. Nous admettrons ces points.

Unicité : Elle résulte du Théorème suivant :

Théorème 1.4.7. 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) $\mu_1(P) = \mu_2(P), \forall P \in \mathcal{P}$.

Alors $\mu_1 = \mu_2$.

PREUVE

a) Soit $A \in \mathcal{P}$, t.q. $\mu_1(A) = \mu_2(A) < \infty$. Posons $\mathcal{L} = \{B \in \sigma(\mathcal{P}); \mu_1(A \cap B) = \mu_2(A \cap B)\}$. \mathcal{L} est un λ -système (exercice), et $\mathcal{L} \supset \mathcal{P}$. Donc -Théorème 1.3.9- $\mathcal{L} \supset \sigma(\mathcal{P})$, d'où $\mathcal{L} = \sigma(\mathcal{P})$. [puisque, par définition de \mathcal{L} , $\mathcal{L} \subset \sigma(\mathcal{P})$].

Donc $\mu_1(A \cap B) = \mu_2(A \cap B), \forall B \in \mathcal{F}$, et $\forall A \in \mathcal{P}$ 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 $\mu_\alpha(A_n) < \infty, \forall n \geq 1, \alpha = 1, 2$.
Soit $B \in \mathcal{F}$. Or pour $\alpha = 1, 2$, et $\forall n \geq 1$, on a l'égalité suivante (exercice)

$$(*) \quad \mu_\alpha \left(\bigcup_1^n (B \cap A_i) \right) = \sum_1^n \mu_\alpha(B \cap A_i) - \sum_{1 \leq i < j \leq n} \mu_\alpha(B \cap A_i \cap A_j) + \dots + (-1)^n \mu_\alpha(B \cap A_1 \cap \dots \cap A_n)$$

Comme \mathcal{P} est un π -système qui contient les A_i , il contient les $A_i \cap A_j, \dots$. Il résulte alors de la formule (*) et de la partie a) que

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

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

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

□

On a le résultat d'approximation :

Proposition 1.4.8. Soit $(\Omega, \mathcal{F}, \mu)$ un espace mesuré, et \mathcal{F}_0 une algèbre, t.q. $\sigma(\mathcal{F}_0) = \mathcal{F}$. $\forall A \in \mathcal{F}$ tel que $\mu(A) < \infty$ et $\forall \varepsilon > 0, \exists A_0 \in \mathcal{F}_0$ t.q. $\mu(A \Delta A_0) \leq \varepsilon$.

PREUVE D'après l'indication donnée pour la démonstration du théorème 1.4.6, $\mu(A) = \inf \sum_n \mu(A_n); A_n \in \mathcal{F}_0$ et $A \subset \bigcup_n A_n$, Donc \exists une suite $\{\bar{A}_n\} \subset \mathcal{F}_0$ telle que :

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

d'où $\mu \left(\bigcup_n \bar{A}_n - A \right) \leq \frac{\varepsilon}{2}$, et $\sum_n \mu(\bar{A}_n) < \infty$. Alors $\exists N$ t.q. $\mu \left(\bigcup_{n>N} \bar{A}_n \right) \leq \sum_{n>N} \mu(\bar{A}_n) \leq \frac{\varepsilon}{2}$

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

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

□

On ne peut pas en général étendre μ à la σ -algèbre de toutes les parties de Ω , comme le montre l'exemple suivant :

Exemple 1.4.9. (*Ensemble non mesurable*) Considérons $\Omega = [0, 1[$, muni de sa tribu borélienne $\mathcal{B} = \mathcal{B}([0, 1[)$. On définit sur (Ω, \mathcal{B}) la mesure de Lebesgue λ (cf. ci-dessous section 5) t.q.

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

Pour $x, y \in [0, 1[$, on définit

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

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

Montrons que $\forall A \in \mathcal{B}$, (*) $A \oplus x \in \mathcal{B}$ et $\lambda(A \oplus x) = \lambda(A)$.

Pour cela, posons $\mathcal{L} = \{A \in \mathcal{B} \text{ qui vérifient (*)}\}$. \mathcal{L} est un λ -système qui contient les intervalles, d'où d'après le Théorème $\pi - \lambda$, $\mathcal{L} = \mathcal{B}$.

On dira que x et y sont équivalents ($x \sim y$) si $\exists r$ rationnel $\in [0, 1[$ t.q. $x \oplus r = y$.

Grâce à l'axiome du choix, \exists une partie H de $[0, 1[$ qui contient exactement un représentant de chaque classe d'équivalence. Considérons les ensembles $H \oplus r$, r parcourant l'ensemble des rationnels de $[0, 1[$. Il s'agit d'une collection dénombrable d'ensembles disjoints, dont la réunion est $[0, 1[$. Donc si $H \in \mathcal{B}$,

$$1 = \sum_{\substack{r \in [0, 1[\\ r \text{ rationnel}}} \lambda(H \oplus r).$$

Mais $\lambda(H \oplus r) = \lambda(H)$. Soit $\lambda(H) = 0$ et $1 = 0$, soit $\lambda(H) > 0$, et $1 = +\infty$!. Donc $H \notin \mathcal{B}$.

□

Définition 1.4.10. 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 $[a, b[$ ($-\infty < a \leq b \leq +\infty$) { ou $] - \infty, b]$ ($-\infty < b \leq +\infty$). 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 $\forall k \neq \ell, I_k \cap I_\ell = \emptyset$. On définit une application $\lambda : \mathcal{S} \rightarrow \bar{\mathbb{R}}_+$ par :

$$\lambda([a, b]) = b - a \quad [\text{ et } (\lambda(] - \infty, b]) = +\infty]$$

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 $\{[a_k, b_k[, k \in \mathbb{N}\}$ 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 une suite de longueur $n - 1$, et supposons que la numérotation des (a_k, b_k) est telle que $a_n > a_k, \forall k \leq n - 1$ Alors :

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

Par l'hypothèse de récurrence,

$$\begin{aligned} \sum_1^{n-1} (b_k - a_k) &\leq a_n - a \\ \text{Donc } \sum_1^n (b_k - a_k) &\leq b_n - a \leq b - a \end{aligned}$$

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

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

ce qui suffit à établir le résultat. □

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,

$$b - a \leq b_n - a \leq \sum_1^n (b_k - a_k)$$

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 $]a_k - \varepsilon 2^{-k}, b_k[$ forment un recouvrement ouvert du compact $[\bar{a}, \bar{b} - \varepsilon]$, où

$$\bar{a} = \begin{cases} a & \text{si } a \in \mathbb{R} \\ M & \text{si } a = -\infty \end{cases}$$

$$\bar{b} = \begin{cases} b & \text{si } b \in \mathbb{R} \\ M & \text{si } b = +\infty \end{cases}$$

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

$$[\bar{a}, \bar{b} - \varepsilon] \subset \bigcup_1^n]a_k - \varepsilon 2^{-k}, b_k[$$

et a fortiori :

$$[\bar{a}, \bar{b} - \varepsilon[\subset \bigcup_1^n]a_k - \varepsilon 2^{-k}, b_k[$$

D'après le résultat dans le cas d'une suite finie,

$$\bar{b} - \varepsilon - \bar{a} \leq \sum_1^\infty (b_k - a_k) + \varepsilon, \quad \forall \varepsilon > 0, \forall M$$

$$\text{D'où } b - a \leq \sum_1^\infty (b_k - a_k)$$

□

Il résulte de la proposition 1.5.1, grâce à 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 II ci-dessous. Nous

admettrons donc provisoirement 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 $a_i, b_i \in \mathbb{R}, a_i \leq b_i, \forall i$

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

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

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 $\mu(A) \triangleq \lambda(A + x), A \in \mathcal{B}_d$. μ et λ coïncident sur le π -système des pavés, donc $\mu = \lambda$. \square

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 $\exists x \in \mathbb{R}^d - V$, 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. Il résulte 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. $\mu(A) < \infty, \forall A \in \mathcal{B}_d$, A borné.*

(i) $\forall A \in \mathcal{B}_d$ et $\varepsilon > 0, \exists$ un fermé F et un ouvert O tels que

$$F \subset A \subset O$$

$$\text{et } \mu(O - F) < \varepsilon.$$

(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*

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

PREUVE

- a) Montrons que (i) \Rightarrow (ii). Si $\mu(A) < \infty$, $\exists A_0 \in \mathcal{B}_d$, A_0 borné, tel que $A_0 \subset A$ et $\mu(A - A_0) < \varepsilon/2$. Or (i) $\Rightarrow \exists K \subset A_0$, K fermé (donc compact puisque borné) t.q. $\mu(A_0 - K) < \varepsilon/2$; donc $\mu(A - K) < \varepsilon$.
- b) Montrons (i). Il suffit de montrer que $\exists O$ ouvert $\supset A$ t.q. $\mu(O - A) < \varepsilon$. 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 : $A = \{x; a_i \leq x_i < b_i : , i = 1, \dots, d\}$. Soit $O_n = \{x; a_i - \frac{1}{n} < x_i < b_i\}$. $\mu(O_1) < \infty$, $O_n \downarrow A$, donc (cf.1.4.4 (ii)) $\exists n$ t.q. $\mu(O_n - A) < \varepsilon$.

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 $\forall A \in \mathcal{B}_d$, \exists une suite $\{A_n, n \in \mathbb{N}^*\}$ de pavés bornés t.q.

$$A \subset \bigcup_1^{\infty} A_n$$

$$\mu \left(\bigcup_1^{\infty} A_n - A \right) < \varepsilon/2.$$

et $\forall n$, \exists un ouvert $O_n \supset A_n$ t.q. $\mu(O_n - A_n) < \varepsilon/2^n$. Or

$$\bigcup_1^{\infty} O_n - A \subset \bigcup_1^{\infty} (O_n - A_n) \cup \left(\bigcup_1^{\infty} A_n - A \right)$$

Donc $\mu \left(\bigcup_1^{\infty} O_n - A \right) < \varepsilon$, et $\bigcup_1^{\infty} O_n$ est un ouvert $\supset A$.

□

1.7 Applications mesurables

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

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

Exercice 1.7.2. *Soit $\varphi : \Omega \rightarrow E$. 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 :*

- (i) \mathcal{C} est une σ -algèbre $\Rightarrow \varphi^{-1}(\mathcal{C})$ est une σ -algèbre.
- (ii) \mathcal{A} est une σ -algèbre $\Rightarrow \varphi(\mathcal{A})$ est une σ -algèbre.
- (iii) $\sigma(\varphi^{-1}(\mathcal{C})) = \varphi^{-1}(\sigma(\mathcal{C}))$.
- (iv) $\varphi : \Omega \rightarrow \mathbb{R}$ est mesurable ssi $\{\varphi < x\} \in \mathcal{F}, \forall x \in \mathbb{R}$. Même énoncé avec $\{\varphi < x\} \in \mathcal{F}$ remplacé par $\{\varphi \leq x\} \in \mathcal{F}$

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

Théorème 1.7.3. (i) *Soit \mathcal{C} une classe de parties de E , t.q. $\sigma(\mathcal{C}) = \mathcal{E}$, et $f : \Omega \rightarrow E$. Si $f^{-1}(C) \in \mathcal{F}, \forall C \in \mathcal{C}$, alors f est \mathcal{F}/\mathcal{E} mesurable.*

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

PREUVE

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

□

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

Proposition 1.7.4. *$f : \Omega \rightarrow \mathbb{R}^d$ est $\mathcal{F}/\mathcal{B}_d$ mesurable si et seulement si*

$\forall i \in \{1, \dots, d\}, f_i : \Omega \rightarrow \mathbb{R}$ est \mathcal{F}/\mathcal{B} mesurable

$$\left(\text{où } f(\omega) = \begin{pmatrix} f_1(\omega) \\ \vdots \\ f_d(\omega) \end{pmatrix} \right)$$

PREUVE On remarque que : $\{\omega; f_1(\omega) < x_1, \dots, f_d(\omega) < x_d\} = \bigcap_{i=1}^d \{\omega; f_i(\omega) < x_i\}$. On applique 1.7.3 (i). \square

Proposition 1.7.5. *Si $f : \mathbb{R}^k \rightarrow \mathbb{R}^d$ est continue, alors elle est $\mathcal{B}_k/\mathcal{B}_d$ mesurable.*

PREUVE Appliquer 1.7.3 (i) avec \mathcal{C} =classe des ouverts de \mathbb{R}^d . \square

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

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

Dans la suite, pour signifier que $f : \Omega \rightarrow \bar{\mathbb{R}}$ 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 1.7.5 et 1.7.3 (ii)).

Théorème 1.7.6. *Soit $\{f_n, n \in \mathbb{N}\}$ des applications \mathcal{F} -mesurables, de Ω à valeurs dans $\bar{\mathbb{R}}$.*

- (i) $\sup_n f_n, \inf_n f_n, \limsup_n f_n, \liminf_n f_n$ sont \mathcal{F} -mesurables.
- (ii) Si $f(\omega) = \lim_n f_n(\omega) \exists, \forall \omega$, alors f est \mathcal{F} -mesurable.
- (iii) $\{\omega; f_n(\omega) \text{ converge}\} \in \mathcal{F}$.
- (iv) si $f : \Omega \rightarrow \bar{\mathbb{R}}$ est \mathcal{F} -mesurable, alors $\{\omega; f_n(\omega) \rightarrow f(\omega)\} \in \mathcal{F}$.

PREUVE

$$(i) \forall x \in \bar{\mathbb{R}}, \{\sup_n f_n \leq x\} = \bigcap_n \{f_n \leq x\}$$

$$\inf_n f_n = -\sup_n (-f_n)$$

$$\limsup_n f_n = \inf_n \sup_{k \geq n} f_k$$

$$\liminf_n f_n = -\limsup_n (-f_n)$$

- (ii) L'hypothèse faite entraîne que $f(\omega) = \limsup_n f_n(\omega); \forall \omega$.

$$(iii) \{f_n \text{ converge}\} = \{\limsup f_n - \liminf f_n = 0\}$$

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

□

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

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

Proposition 1.7.7. *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 (exercice!). De plus, la classe des fonctions \mathcal{F} -mesurables est fermée pour la convergence ponctuelle. cf. 1.7.6 (ii). Il reste à montrer que si f est \mathcal{F} mesurable, alors $\exists f_n$ étagées t.q. $f_n(\omega) \rightarrow f(\omega), \forall \omega$. On pose :

$$A_k^n = f^{-1} \left(\left[\frac{k}{n}, \frac{k+1}{n} \right] \right), \quad n \in \mathbb{N}^*, k \in \mathbb{Z}$$

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

□

Théorème 1.7.8. *(des classes monotones)*

Soit \mathcal{H} un π -système de parties de Ω . $\sigma(\mathcal{H}) = \mathcal{F}$ et \mathcal{L} une classe d'applications \mathcal{F} -mesurables à valeurs dans $\bar{\mathbb{R}}$, qui vérifie :

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

$$(ii) f, g \in \mathcal{L} \text{ et } \alpha, \beta \in \mathbb{R} \Rightarrow \alpha f + \beta g \in \mathcal{L}$$

$$(iii) f_n \in \mathcal{L} \text{ et } f_n \uparrow f \Rightarrow f \in \mathcal{L}.$$

[resp. (iii)' à condition que $f(\omega) \in \mathbb{R}, \forall \omega \in \Omega$]

[resp. (iii)'' à condition que $\exists c \in \mathbb{R}_+ \text{ t.q. } |f(\omega)| \leq c, \forall \omega \in \Omega$]

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 définie comme à la Proposition 1.7.7, $f_{2^n} \uparrow f$, et $f_{2^n} \in \mathcal{L}$. 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.

□

La démonstration ci-dessus contient celle du :

Corollaire 1.7.9. *Mêmes hypothèses qu'au théorème 1.7.8, en remplaçant (i) et (ii) par :*

(i) $1_A \in \mathcal{L}, \forall A \in \mathcal{F}$

(ii) $f, g \in \mathcal{L}$ et $\alpha, \beta \geq 0 \Rightarrow \alpha f + \beta g \in \mathcal{L}$

On a alors les mêmes conclusions, en remplaçant $\bar{\mathbb{R}}$ par $\bar{\mathbb{R}}_+$ et \mathbb{R} par \mathbb{R}_+ .

□

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

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

□

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 $\exists N \in \mathcal{F}$ avec $\mu(N) = 0$, t.q. $\forall \omega \notin N$, $P(\omega)$ est vérifiée.

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

Définition 2.1.1. f et g sont dites égales presque partout (en abrégé p.p.) si $\exists N \in \mathcal{F}$, avec $\mu(N) = 0$, t.q. $\forall \omega \notin N$, $f(\omega) = g(\omega)$.

On écrira

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

ou plus simplement

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

“ $f = g$ p.p.” est une relation d'équivalence. (cf. chapitre III).

□

Définition 2.1.2. On dit que f_n converge vers f p.p. si $\exists N \in \mathcal{F}$ avec $\mu(N) = 0$ t.q.

$$\forall \omega \notin N, f_n(\omega) \rightarrow f(\omega)$$

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

Théorème 2.1.3. Soit $\mathcal{G} \subset \mathcal{F}$ une tribu, et $\{f_n, n \in \mathbb{N}\}$ une suite d'applications \mathcal{G} -mesurables, et f une application de Ω dans $\overline{\mathbb{R}}$ t.q.

$$f_n \rightarrow f \text{ p.p.}$$

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

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

PREUVE $\Gamma = \{\omega; f_n(\omega) \text{ converge}\} \in \mathcal{G}$, d'après le Théorème 1.7.6 (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

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

$$\forall x \leq 0, \{\bar{f} < x\} = \Gamma \cap \{\overline{\lim} f_n < x\} \in \mathcal{G}$$

□

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

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

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 $\overline{\mathbb{R}}_+$.

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 :

$$\bigcup_{i=1}^n A_i = \Omega$$

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

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

$$\Omega, \{x_i, 1 \leq i \leq n\} \subset \mathbb{R}_+, \text{ si } f = \sum_{i=1}^n x_i \mathbf{1}_{A_i},$$

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

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

(iii) Si $0 \leq f_n(\omega) \uparrow f(\omega), \forall \omega$, 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 $y_j \leq x_i$.

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

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

- (ii) Trivial d'après 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, $\forall \{A_i, 1 \leq i \leq k\}$ partition de Ω ,

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

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

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

Si $\omega \in A_i$, soit $f_n(\omega) \uparrow f(\omega) \geq v_i = u_i = 0$
soit $f_n(\omega) \uparrow f(\omega) \geq v_i > u_i > 0$

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 dans la preuve du Théorème 1.7.8, en utilisant (iii). Soit $f = \sum_1^n x_i \mathbf{1}_{A_i}$, $g = \sum_1^n y_j \mathbf{1}_{B_j}$. Sans restreindre la généralité, on peut supposer que $\{A_i; i \leq n\}$ et $\{B_j; j \leq m\}$ forment des partitions de Ω .

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

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

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

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

□

Théorème 2.2.2. (i) Si $f = 0$ μ p.p., alors $\int f d\mu = 0$.

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

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

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

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

PREUVE

(i) résulte de ce que $\forall A_i \in \mathcal{F}$,

soit $A_i \cap \{f = 0\} \neq \emptyset$, et alors $\inf_{\omega \in A_i} f(\omega) = 0$

soit $A_i \cap \{f = 0\} = \emptyset$, et alors $\mu(A_i) = 0$.

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

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

(iii) $+\infty > \int f d\mu \geq +\infty \cdot \mu(\{f = +\infty\}) \Rightarrow \mu(f = +\infty) = 0$.

(iv) Soit

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

$\forall \omega \in \Omega$, $f(\omega) \leq g(\omega) + h(\omega)$. D'après 2.2.2 (ii) et (iv), (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 $\int g d\mu \leq \int f d\mu$.

(v) On pose

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

$\forall \omega \in \Omega$, $f(\omega) \leq g(\omega) + h(\omega)$. On a ensuite les mêmes inégalités qu'en (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 $\int f^+ d\mu$ et $\int f^- d\mu$. Trois cas se présentent :

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

b) L'une au moins des deux quantités $\int f^+ d\mu$ et $\int f^- d\mu$ 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 $\int f^+ d\mu$ et $\int f^- d\mu$ sont finies. f est alors dite **intégrable**, et on définit $\int f d\mu$ comme dans le cas quasi-intégrable.

Alors $\int f d\mu \in \mathbb{R}$.

Remarquons que $|f| = f^+ + f^-$, donc d'après le Théorème 2.2.1 (iv), f est intégrable si et seulement si $\int |f| d\mu < \infty$.

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 $\alpha = \beta = 1$.

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

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

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

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 $f = \frac{1}{\alpha}(\alpha f + \beta g) - \frac{\beta}{\alpha}g$ le serait, et $(\alpha f + \beta g)^- \leq \alpha f^- + |\beta| - |g|$, qui est intégrable.

$$\text{Donc } \int (\alpha f + \beta g) d\mu = +\infty.$$

□

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 du Théorème 2.3.2 (i), car $-|f| \leq f \leq |f|$.

(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) Soit f_n, f des applications mesurables à valeurs dans $\bar{\mathbb{R}}$.*

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

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

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 $\tilde{g}_n(\omega) = \sup_{k \leq n} g_k(\omega)$, $\tilde{g}(\omega) = \lim_{n \rightarrow \infty} \tilde{g}_n(\omega)$. Il résulte des Théorèmes 2.2.1 (iii) et 2.2.2 (iv) : $\int \tilde{g}_n d\mu \uparrow \int \tilde{g} d\mu$. 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 que $\int g_n d\mu \uparrow \int g d\mu$. Le résultat découle alors du Corollaire 2.3.3. \square

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 $\{f_n, n \in \mathbb{N}\}$ une suite d'applications mesurables à valeurs dans $\bar{\mathbb{R}}$.

(i) Si $\exists f$ intégrable t.q. $f_n \geq f$ p.p., $\forall n$, alors :

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

(ii) Si $\exists f$ intégrable t.q. $f_n \leq f$ p.p., $\forall n$, 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 $g_n(\omega) = \inf_{m \geq n} f_m(\omega)$. $f \leq g_n \leq f_n$ p.p., $g_n \uparrow \liminf_n f_n$. Donc $g_1^- \leq f^-$, g_1^- est intégrable, et d'après le Théorème 2.3.5 (i),

$$\begin{aligned} \int (\liminf_n f_n) d\mu &= \lim_n \int g_n d\mu \\ \int (\liminf_n f_n) d\mu &= \lim_n \int g_n d\mu \\ &\leq \liminf_n \int f_n d\mu. \end{aligned}$$

La dernière inégalité s'obtient en prenant la liminf dans l'inégalité $\int g_n d\mu \leq \int f_n d\mu$. \square

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

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

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

(ii) $\exists g$ **intégrable** t.q. $|f_n(\omega)| \leq g(\omega)$ p.p., $\forall n \in \mathbb{N}$. Alors f_n et f sont intégrables, et

$$\int f_n d\mu \rightarrow \int f d\mu$$

PREUVE Posons $N_n = \{|f_n| > g\}$, $N = \{f_n \not\rightarrow f\}$. Alors $M = N \cup \left(\bigcup_1^\infty N_n\right)$ vérifie $\mu(M) = 0$, et $\forall \omega \notin M$, $|f(\omega)| \leq g(\omega)$. 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.

Grâce au théorème 2.3.7 (ii)

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

Donc $\int g_n d\mu \rightarrow 0$. 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 $\{f_n, n \in \mathbb{N}\}$ des applications mesurables à valeurs dans $\bar{\mathbb{R}}$.

(i) si $f_n \geq 0$ p.p., alors $\int \left(\sum_n f_n \right) d\mu = \sum_n \int f_n d\mu$

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

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

(iii) Si $\sum_n \int |f_n| d\mu < \infty$, alors $\sum_n f_n$ 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) $\int (\sum_1^N f_n) d\mu = \sum_1^N \int f_n d\mu$, et grâce au Théorème de convergence monotone, on peut faire tendre $N \rightarrow \infty$.

(ii) résulte immédiatement du Théorème de convergence dominée.

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

□

Proposition 2.3.11. Soit $f : \Omega \times]a, b[\rightarrow \bar{\mathbb{R}}$ t.q. $\forall t \in]a, b[, \omega \rightarrow f(\omega, t)$ soit mesurable.

(i) Supposons que ω p.p., $t \rightarrow f(\omega, t)$ est continue au oint $t_0 \in]a, b[$ et que $\exists g$ intégrable t.q. $|f(\omega, t)| \leq g(\omega)$ p.p., $\forall t \in]a, b[$, alors $t \rightarrow \int f(t, \omega) d\mu(\omega)$ est continue au point t_0 .

(ii) Supposons que p.p., $t \rightarrow f(\omega, t)$ est dérivable en tout $t \in]a, b[$, que $\exists g$ intégrable t.q. $|f'(\omega, t)| \leq g(\omega)$ p.p. On suppose en outre que $f(t)$ est intégrable, $\forall t \in]a, b[$. Alors $t \rightarrow \int f(\omega, t) d\mu(\omega)$ est dérivable en tout $t \in]a, b[$, et $\frac{d}{dt} \int f(\omega, t) d\mu(\omega) = \int f'(\omega, t) d\mu(\omega)$.

PREUVE

(i) résulte directement du Théorème de convergence dominée.

(ii)

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

$$\text{Or } \frac{f(\omega, t+h) - f(\omega, t)}{h} \rightarrow f'(\omega, t) \text{ p.p., quand } h \rightarrow 0,$$

$$\text{et } \left| \frac{f(\omega, t+h) - f(\omega, t)}{h} \right| = |f'(\omega, t + \theta h)| \leq g(\omega) \text{ p.p.}$$

par le théorème des accroissements finis.

□

Etant donnés 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 $\int_{[a,b]} f(x) dx$, 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 $\int_a^b f(x) dx = \lim_{n \rightarrow \infty} \sum_{i=0}^{n-1} f(x_i^n) (x_{i+1}^n - x_i^n)$, où $a = x_0^n < x_1^n < \dots < x_n^n = b$ vérifie $\sup_i (x_{i+1}^n - x_i^n) \rightarrow 0$ quand $n \rightarrow \infty$.

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 λ -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 $\int_{]a,b]}$, $\int_{]a,b[}$, \dots

□

Si $f \geq 0$, $\nu(A) = \int_A f d\mu$, $A \in \mathcal{F}$, définit une mesure ν sur (Ω, \mathcal{F}) (exercice).

On dit alors que ν admet la densité f par rapport à μ , et on a la propriété (*) $\forall A \in \mathcal{F}, \mu(A) = 0 \Rightarrow \nu(A) = 0$. Si deux mesures ν et μ vérifient la propriété (*), on dit que ν est **absolument continue** par rapport à μ -noté $\nu \ll \mu$.

Proposition 2.3.13. *Soit $f \geq 0$, ν la mesure sur (Ω, \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. fg 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 (Ω', \mathcal{F}') un deuxième espace mesurable, $T : \Omega \rightarrow \Omega'$ 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. \square

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 $f \mathbf{1}_{TV}$ est λ -quasi-intégrable si et seulement si $f \circ T \mathbf{1}_V |J|$ est λ quasi intégrable.*

PREUVE On va se limiter au cas $d = 1$, $V =]a, b[$. L'égalité sur l'énoncé peut alors se réécrire :

$$\int_a^b f(T(x)) |T'(x)| dx = \int_{Ta}^{Tb} f(y) dy \left(\text{si } Ta \leq Tb, \int_{Tb}^{Ta} \text{ sinon} \right)$$

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 $]a, b[$, on a

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

(traiter séparément les cas $T' \geq 0$ et $T' \leq 0$; T' a un signe constant, puisque T est injective).

Donc $\nu T^{-1} = \lambda$, mesure de Lebesgue. Il résulte alors du Théorème 2.3.14 :

$$\int_{T^{-1}B} f \circ T(x) \nu(dx) = \int_B f(y) dy.$$

Il suffit de choisir $B =]Ta, Tb[$. □

2.4 Mesure produit et Théorème de Fubini

Dans cette section, on se donne deux espaces mesurés (X, \mathcal{X}, μ) et (Y, \mathcal{Y}, ν) , 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 $E \in \mathcal{X} \otimes \mathcal{Y}$, $\forall x \in X$, $\{y; (x, y) \in E\} \in \mathcal{Y}$ et $\forall y \in Y$, $\{x; (x, y) \in E\} \in \mathcal{X}$.

(ii) Si $f : X \times Y \rightarrow \overline{\mathbb{R}}$ est $\mathcal{X} \otimes \mathcal{Y}$ mesurable, alors $\forall x \in X$, $y \rightarrow f(x, y)$ est \mathcal{Y} mesurable et $\forall y \in Y$, $x \rightarrow f(x, y)$ est \mathcal{X} mesurable.

PREUVE

(i) $\forall x \in X$, on définit $T_x : X \rightarrow X \times Y$ 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 1.7.3 (i) que T_x est $\mathcal{X}/\mathcal{X} \otimes \mathcal{Y}$ mesurable. Ceci démontre la première partie de (i). La deuxième partie se démontre de façon analogue.

(ii) On remarque que l'application $y \rightarrow f(x, y)$ coïncide avec $f \circ T_x$, et on applique le Théorème 1.7.3 (ii).

□

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

- (i) $x \rightarrow \nu(\{y; (x, y) \in E\})$ est \mathcal{X} -mesurable.
- (ii) $y \rightarrow \mu(\{x; (x, y) \in E\})$ 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 $x \rightarrow \nu(\{y; (x, y) \in E\})$ 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 $\pi - \lambda$, 1.3.9, $E \supset \sigma(\mathcal{X} \times \mathcal{Y}) = \mathcal{X} \otimes \mathcal{Y}$. □

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

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

$$\pi(A \times B) = \mu(A)\nu(B)$$

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

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

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. □

Théorème 2.4.4. (Fubini) *Soit $f : X \times Y \rightarrow \bar{\mathbb{R}}$ 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)

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

b) Si f est π -intégrable, alors :

(i)' Pour μ -presque tout x , $y \rightarrow f(x, y)$ est ν -intégrable. Pour ν -presque tout y , $x \rightarrow f(x, y)$ 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'invertir 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 $0 = \alpha_1 < \alpha_2 < \dots < \alpha_n, \alpha_n \rightarrow 1$, quand $n \rightarrow \infty$. $\forall n$, 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 :

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

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

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

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

$$\forall y \in [0, 1] \int_0^1 f(x, y) dx = 0; \int_0^1 dy \int_0^1 f(x, y) dx = 0$$

Et 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 2.4.2 et 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).

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 :

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

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 $\forall x \notin A (\mu(A) = 0)$, $\int_Y f(x, y) \nu(dy) = g(x)$, 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 $f = f^+ - f^-$. \square

Exercice 2.4.6. Dédurre (i) et (iii) de la Proposition 2.3.10 comme corollaire du Théorème de Fubini. [On choisit $(X, \mathcal{X}, \mu) = (\Omega, \mathcal{F}, \mu)$, $(Y, \mathcal{Y}, \nu) = (\mathbb{N}, \mathcal{P}(\mathbb{N}), \sum_1^\infty \delta_n)$].

\square

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 $\overline{\mathbb{R}}_+$.*

(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)

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

PREUVE

- (i) Si l'un des deux facteurs du second membre de l'inégalité est nul, le membre de gauche est nul, et l'inégalité est satisfaite. De même, l'inégalité est satisfaite lorsque l'un des deux facteurs du second membre est égal à $+\infty$, l'autre étant 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 $\exists s, t \in \mathbb{R}$ 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 $\forall \omega \in \Omega$. Donc en intégrant.

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

- (ii) A nouveau, il suffit de faire la preuve dans le cas où le second membre de l'inégalité est strictement positif et fini, et $f + g$ n'est pas p.p. égal à 0.

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

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

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 $t \rightarrow t^p$ 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

$$\text{par } \left(\int (f + g)^p d\mu \right)^{\frac{1}{p} - 1}$$

□

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 :

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

$\forall \lambda \in \mathbb{R}$, donc le discriminant du trinôme est ≤ 0 . Le discriminant est nul (i.e. on a l'égalité dans l'inégalité de Cauchy–Schwarz) si et seulement si $\exists \lambda \in \mathbb{R}$ 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

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

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 $p \in]1, +\infty[$.

Exercice 3.1.4. Montrer que si $\|f\|_{\infty}$, $\mu(\{\omega; |f(\omega)| > \|f\|_{\infty}\}) = 0$.

□

$\|\cdot\|_p$ 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 :

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

L'ensemble des classes d'équivalences d'éléments de $\mathcal{L}^p(\mu)$ forme 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 3.1.5. 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 $f_n \sim f'_n$, $n \in \mathbb{N}$, alors $\sup_{n \in \mathbb{N}} f_n \sim \sup_{n \in \mathbb{N}} f'_n$, $\inf_{n \in \mathbb{N}} f_n \sim \inf_{n \in \mathbb{N}} f'_n$, 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'équivalences 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 !.

Exercice 3.1.6. a) Montrer que si μ est une mesure finie, $p < r$, alors $L^r(\mu) \subset L^p(\mu)$ [appliquer Hölder à $1 \cdot |f|^p$], et que dans le cas où μ est une probabilité, $p \rightarrow \|f\|_p$ 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, $p < r$, $\exists f \in L^p(\lambda)$ t.q. $f \notin L^r(\lambda)$ et $\exists g \in L^r(\lambda)$ tel que $g \notin L^p(\lambda)$.

3.2 Propriétés des espaces $L^p(\mu)$

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

Muni du produit scalaire

$$(f, g) = \int f g d\mu$$

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 $\{f_n, n \in \mathbb{N}\} \subset L^p(\mu)$ t.q. $\forall \varepsilon > 0, \exists N_\varepsilon$ tel que $n, m \geq N_\varepsilon \Rightarrow \|f_n - f_m\|_p \leq \varepsilon$.

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

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

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 $\forall \omega \in A$, 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 $f_{n_k} \rightarrow f$ p.p.. Il reste à montrer que $f \in L^p(\mu)$, et $f_n \rightarrow f$ dans $L^p(\mu)$, i.e. $\|f_n - f\|_p \rightarrow 0$.

Soit $\varepsilon > 0$, et N_ε t.q. $\|f_n - f_m\|_p \leq N_\varepsilon, \forall n, m \geq N$. 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 $E = \left(\bigcup_{k=1}^\infty A_k \right) \cup \left(\bigcup_{n,m=1}^\infty B_{n,m} \right)$. 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 $\omega \in E$.

Alors $f \in L^\infty(\mu)$, et $\|f_n - f\|_\infty \rightarrow 0$

□

On notera $f_n \xrightarrow{L^p(\mu)} f$ pour $f_n \rightarrow f$ dans $L^p(\mu)$, i.e. $f_n, f \in L^p(\mu)$ et $\|f - f_n\|_p \rightarrow 0$. 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 \exists une sous-suite $\{f_{n_k}\}$ t.q.*

$$f_{n_k} \longrightarrow f p.p.$$

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

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

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

Remarque 3.2.4. *S n'est en général pas dense dans $L^\infty(\mu)$ (ex : $\Omega = \mathbb{R}_+$, $\mathcal{F} = \mathcal{B}(\mathbb{R}_+)$, $\mu = \lambda$, $f \sum_0^\infty \mathbf{1}_{[2n, 2n+1[}$ 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, \exists une suite $\{f_n\}$ de fonctions étagées t.q. $0 \leq f_n \uparrow f$. Alors $f_n \in S$, et $\|f - f_n\|_p \rightarrow 0$, d'après le théorème de convergence monotone. Pour f de signe quelconque, on décompose $f = f^+ - f^-$. □

Théorème 3.2.5. *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 $C_K(\mathbb{R}^d)$, espace des fonctions continues à support compact de \mathbb{R}^d à valeurs dans \mathbb{R} , est dense dans $L^p(\mu)$, $\forall p \in [1, +\infty[$.*

Exercice 3.2.6. *Donner un contre-exemple dans le cas $p = +\infty$.* □

On va d'abord établir le :

Lemme 3.2.7. *Soit K un compact et O un ouvert de \mathbb{R}^d , avec $K \subset O$. Alors il existe $g \in C_K(\mathbb{R}^d)$ avec $0 \leq g(x) \leq 1$, $\forall x$, $g(x) = 1$, $\forall x \in K$; $g(x) = 0$, $\forall x \in O^c$.*

PREUVE Il suffit de traiter le cas O borné. Alors $g(x) = \frac{d(x, O^c)}{d(x, K) + d(x, O^c)}$ 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, il suffit de montrer que $\forall f \in S$ et $\forall \eta > 0$, $\exists g \in C_K(\mathbb{R}^d)$ t.q.

$$\|f - g\|_p \leq \eta$$

Il suffit donc en fait de montrer que $\forall A \in \mathcal{F}$, avec $\mu(A) < \infty$, $\forall \varepsilon > 0$, $\exists g \in C_K(\mathbb{R}^d)$ t.q.

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

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

$$\text{Donc : } K \subset A \subset O, \quad \mu(O - K) \leq \varepsilon,$$

et si g désigne la fonction construite au lemme 3.2.7,

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

\square

Théorème 3.2.8. *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

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

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

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

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

$$\exists B \in \mathcal{F}_0 \text{ t.q. } \mu(A \Delta B) \leq \theta.$$

Puisque $\mu(A) < \infty$ et \mathcal{F}_0 est une algèbre t.q. $\sigma(\mathcal{F}_0) = \mathcal{F}$, cette dernière propriété est une conséquence immédiate de la Proposition 1.4.8 \square

3.3 Théorème de Radon–Nikodym et dualité des espaces L^p

Soit λ et μ deux mesures σ -finies sur (Ω, \mathcal{F}) . On dit que λ est **absolument continue** par rapport à μ , noté $\lambda \ll \mu$, si $\forall A \in \mathcal{F}$, $\mu(A) = 0 \Rightarrow \lambda(A) = 0$. Elles sont dites **équivalentes** si l'on a à la fois $\lambda \ll \mu$ et $\mu \ll \lambda$. Elles sont dites **étrangères** si $\exists E \in \mathcal{F}$ t.q. $\lambda(E) = \mu(E^c) = 0$.

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

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

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

PREUVE

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

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

3.3. THÉORÈME DE RADON–NIKODYM ET DUALITÉ DES ESPACES L^p 53

Donc l'application $f \rightarrow \int f d\lambda$ est une fonctionnelle linéaire continue sur l'espace de Hilbert $l^2(\nu)$, d'où (cf. cours L3) $\exists g \in L^2(\nu)$ t.q. :

$$\int f d\lambda = \int f g d\nu, \quad \forall f \in L^2(\nu)$$

Si $\{A_n, n \in \mathbb{N}\}$ vérifie $\nu(A_n) < \infty, \forall n$ et $\bigcup_1^\infty A_n = \Omega$, alors en choisissant $f = \mathbf{1}_{A_n \cap \{g < 0\}}, n = 0, 1, \dots$ dans l'égalité ci-dessus, on montre que $g \geq 0 \nu$ 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 :

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

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

i.e. $\exists g$ 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

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

Mais puisque $\lambda \ll \mu, \lambda(g=1) = 0$, et $\mathbf{1}_{\{g=1\}} \cdot \lambda = 0$. D'où le résultat cherché avec $h = \mathbf{1}_{\{g < 1\}} \frac{g}{1-g}$

b) Le cas général : $\lambda \sigma$ -finie ;

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

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

Alors les λ_n sont des mesures finies, et :

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

D'après la partie a) de la démonstration, $\forall n, \exists h_n$ 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, $\forall \omega \in \Omega$, la série $h(\omega) = \sum_1^{\infty} h_n(\omega)$ converge, et

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

□

Théorème 3.3.2. (*Décomposition de Lebesgue*)

Soit λ et μ deux mesures σ -finies sur (Ω, \mathcal{F}) . λ s'écrit de façon unique comme la somme d'une mesure absolument continue 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 de 3.3.1 cf. égalité (*), dans le cas $\lambda(\Omega) < \infty$. Le cas général est laissé en exercice. □

Définition 3.3.3. On appellera mesure signée sur (Ω, \mathcal{F}) une application μ de \mathcal{F} dans \mathbb{R} t.q. :

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

$$(ii) \quad \begin{cases} \forall \{A_n, n \in \mathbb{N}\} \subset \mathcal{F} \text{ avec } A_k \cap A_\ell = \emptyset \text{ si } k \neq \ell, \\ \mu \left(\bigcup_{n=1}^{\infty} A_n \right) = \sum_{n=1}^{\infty} \mu(A_n) \end{cases}$$

Nous admettrons la :

Proposition 3.3.4. (*Jordan-Hahn*) Si μ est une mesure signée définie sur (Ω, \mathcal{F}) , alors les formules :

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

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

3.3. THÉORÈME DE RADON–NIKODYM ET DUALITÉ DES ESPACES L^p 55

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

(ii) μ^+ et μ^- sont étrangères.

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

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

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

Corollaire 3.3.5. Soit λ une mesure signée, μ une mesure σ -finie, définies sur (Ω, \mathcal{F}) . Si $\forall N \in \mathcal{F}, \mu(N) = 0 \Rightarrow \lambda(N) = 0$, alors λ^+ et λ^- sont absolument continues par rapport à μ , et $\exists h \in L^1(\mu)$ t.q. :

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

PREUVE D'après la proposition 3.3.4, $\exists E \in \mathcal{F}$ 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 : $\Phi(f) := \int f g d\mu$ 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 ? i.e. : 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 3.3.6. Soit μ une mesure σ -finie sur (Ω, \mathcal{F}) , $p \in [1, +\infty[$, et Φ une forme linéaire continue sur $L^p(\mu)$. Alors $\exists g \in L^q(\mu)$ (où $\frac{1}{q} + \frac{1}{p} = 1$) t.q. :

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

De plus, $\|\Phi\| := \sup_{\|f\|_p=1} |\Phi(f)| = \|g\|_q$; i.e.

$L^p(\mu)$ est isomorphe et isométrique au dual de $L^p(\mu)$.

PREUVE

a) **l'unicité** (de la classe d'équivalence) est claire : si g et g' satisfont les propriétés de l'énoncé, alors, $\forall A$ 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 !).

b) **Le cas** $\mu(\Omega) < \infty$. $\forall A \in \mathcal{F}$, $\mathbf{1}_A \in L^p(\mu)$, 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) $\lambda\left(\bigcup_1^\infty A_n\right) = \sum_1^\infty \lambda(A_n)$, 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 $\mu(A) = 0 \Rightarrow \mathbf{1}_A \sim 0 \Rightarrow \lambda(A) = 0$.

Il résulte alors du Corollaire 3.3.5 que $\exists g \in L^1(\mu)$ t.q.

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

$$\Phi(f) = \int f g d\mu, \quad \forall f \text{ étagée, i.e. de la forme}$$

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

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

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

Donc $\|h_n\|_q \leq \|\Phi\|$ d'où par convergence monotone,

$$\|g\|_q \leq \|\Phi\| \text{ donc } g \in L^q(\mu)$$

3.3. THÉORÈME DE RADON–NIKODYM ET DUALITÉ DES ESPACES L^p 57

cas $p = 1$ Supposons que $\|g\|_\infty > \|\Phi\|$. Alors : $h = \varepsilon(g)\mathbf{1}_{\{|g| > \|\Phi\|\}}$ est étagée et $\neq 0$.

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

ce qui est impossible, car $\Phi|h| \leq \|\Phi\| \int (h) d\mu$, pour $h \neq 0$. Donc $\|g\|_\infty \leq \|\Phi\|$. Donc, dans tous les cas, Φ et $f \rightarrow \int f g d\mu$ sont deux formes linéaires continues sur $L^p(\mu)$, qui coïncident sur le sous ensemble dense des fonctions étagées (cf. Théorème 3.2.3), donc sont identiques. De plus, on vient de montrer que $\|g\|_q \leq \|\Phi\|$, mais par Hölder, $\|\Phi\| \leq \|g\|_q$, d'où l'égalité.

c) **Le cas général** : μ σ -finie $\exists \{A_n, n \geq 1\}$ t.q. $\bigcup_1^\infty A_n = \Omega$, les A_n sont disjoints, et $\mu(A_n) < \infty, \forall n$. Posons

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

Alors $h \in L^1(\mu)$. Posons $\tilde{\mu} = h \cdot \mu$. $\tilde{\mu}$ est finie, et on peut lui appliquer le résultat de b).

Or $f \rightarrow h^{1/q} f$ 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

$$\exists \tilde{g} \in L^q(\tilde{\mu}) \text{ t.q. } \psi(f) = \int f \tilde{g} d\tilde{\mu}$$

.

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

- si $p > 1$, $\int |g|^q d\mu = \int |\tilde{g}|^q d\tilde{\mu} = \|\Psi\|^q = \|\Phi\|^q$
- si $p = +\infty$, $\|g\|_\infty = \|\tilde{g}\|_\infty = \|\Psi\| = \|\Phi\|$. Finalement, comme $\tilde{g} \cdot \tilde{\mu} = h^{1/q} g \cdot \mu$,

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

□

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 : Analyse réelle et complexe, Masson.

3.4.1 Théorème de représentations 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. $f(x) \rightarrow 0$, quand $\|x\| \rightarrow 0$).

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) $\forall \Phi$ forme linéaire continue sur $C_0(\mathbb{R}^d)$, \exists une unique mesure signée μ définie sur $(\mathbb{R}^d, \mathcal{B}_d)$, telle que :

$$\Phi(f) = \int f d\mu, \quad \forall f \in C_0(\mathbb{R}^d)$$

De plus, on a :

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

où μ désigne la variation totale de μ .

Exercice 3.4.2. Montrer l'unicité

3.4.2 Intégrale de Stieltjes–Lebesgue

Plaçons-nous dans le cas (Ω, \mathcal{F}) . Si \mathbb{P} est une probabilité, i.e. une mesure \mathbb{P} t.q. $\mathbb{P}(\Omega) = 1$, (cf. ci-dessous Chapitre V), on peut lui associer sa fonction de répartition $F : \mathbb{R} \rightarrow [0, 1]$, 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 :

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

Si $f \in C_K(\mathbb{R})$, $\int f dF = \lim_{n \rightarrow \infty} \sum_{i=-\infty}^{+\infty} f(y_i) \left[F\left(\frac{i+1}{n}\right) - F\left(\frac{i}{n}\right) \right]$

avec $y_i \in [\frac{i}{n}, \frac{i+1}{n}]$, $\forall i$.

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

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

Réciproquement, si $F = \mathbb{R} \rightarrow \mathbb{R}$ 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 gauche. $\bar{F}(x) = \lim_{\substack{y \uparrow x \\ y < x}} F(y)$.

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 $F = F^+ - F^-$, $F = [0, 1] \rightarrow \mathbb{R}$, qui est la différence de deux fonctions croissantes continues à gauche.

La classe des fonctions $F = [0, 1] \rightarrow \mathbb{R}$ 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. $0 = x_0 < x_1 < x_2 < \dots < x_n = 1$ (Décomposition de Jordan - analogue à la Proposition 3.3.4).

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 gauche 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 $\int f dF$ l'intégrale $\int f d\mu$, 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 $F = \mathbb{R} \rightarrow \mathbb{R}$ définie par :

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

est de classe C^1 , et que $F'(x) = f(x)$, $\forall x \in \mathbb{R}$. (cf. Théorie de l'intégrale de Riemann).

Soit maintenant $f \in L^1(\lambda) - \lambda$ 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 3.4.3. (de différentiation de Lebesgue). Soit $f \in \mathcal{L}^1(\lambda)$.

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

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

□

On a une généralisation en dimension d :

Théorème 3.4.4. Soit λ_d la mesure de Lebesgue sur $(\mathbb{R}^d, \mathcal{B}_d)$. $\forall f \in \mathcal{L}^1(\lambda_d)$, $\exists N_f \in \mathcal{B}_d$ avec $\lambda_d(N) = 0$ tel que $\forall x \notin N_f$,

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

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

□

Remarque 3.4.5. ($d = 1$)

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

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

– Ex 2 : $\exists F$ monotone croissante : $[0, 1] \rightarrow [0, 1]$, avec $F(0) = 0$ et $F(1) = 1$, et telle que $\forall x \notin C$ (C : ensemble triadique de Cantor), $F'(x)$ existe et est nulle. La mesure associée à F (cf. 3.4.2) est étrangère par rapport à la mesure de Lebesgue.

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

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

Ceci entraîne (cf. 3.4.2) que F est à variation bornée, et que la mesure μ qui lui est associée 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 **fonction absolument continue**. Une fonction $F : [0, 1] \rightarrow \mathbb{R}$ est absolument continue si et seulement si $\forall \varepsilon > 0, \exists \delta > 0$ t.q. $\forall \{[a_n, b_n], n \in \mathbb{N}\}$ 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.

Chapitre 4

Probabilité-Indépendance- Variable aléatoire

4.1 Probabilité

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

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

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

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

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

Remarque 4.1.2. Une probabilité est une mesure μ de masse totale 1, i.e. t.q. $\mu(\Omega) = 1$.

□

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 Probabilités conditionnelles

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 (Ω, \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 $A_k \in \mathcal{F}$; $k = 1, \dots, n$. Alors, avec la convention ci-dessus,

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

PREUVE

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

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

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

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

Supposons que l'on ait :

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

Alors en multipliant l'égalité ci-dessus par :

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

on obtient le résultat.

□

4.3 Evénements indépendants

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

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

□

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 :

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

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

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

Exercice 4.3.3. 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 : $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A)\mathbb{P}(B)$.

Définition 4.3.4. Une suite $\{A_1, \dots, A_n\} \subset \mathcal{F}$ est dite indépendante si :

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

pour toute suite d'indices tous distincts i_1, \dots, i_k compris entre 1 et n .

□

Remarque 4.3.5. Le fait que les A_n soient 2 à 2 indépendants n'entraîne pas que la suite $\{A_1, \dots, A_n\}$ soit indépendante, comme le montre l'exemple suivant : on joue 3 fois à Pile ou Face avec une pièce non truquée. On pose :

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

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

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

Proposition 4.3.6. *Si la suite $\{A_1, \dots, A_n\}$ est indépendante, les 2^n suites $\{A'_1, \dots, A'_n\}$ 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 $\{A_1, \dots, A_n\}$ entraîne celle de $\{A_1, \dots, A_{\ell-1}, A_\ell^c, A_{\ell+1}, \dots, A_n\}, \forall \ell \in \{1, \dots, n\}$, 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 $\{A_1^c, A_2, \dots, A_n\}$, il suffit d'établir que $\forall i_1, \dots, i_k$ indices tous distincts dans $\{2, \dots, n\}$,

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

Or le membre de gauche est égal à :

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

□

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

On déduit alors aisément de la Proposition 4.3.6 :

Corollaire 4.3.7. *La suite $\{A_1, \dots, A_n\}$ est indépendante si et seulement si l'égalité*

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

est satisfaite $\forall B_i \in \mathcal{F}_i; i = 1 \dots n$.

Définition 4.3.8. *Une suite infinie d'événements $\{A_n, n \in \mathbb{N}\}$ est dite indépendante si toutes les sous-suites finies de $\{A_n, n \in \mathbb{N}\}$ sont indépendantes, i.e. si $\forall k, \forall i_1, \dots, i_k$ entiers tous distincts,*

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

□

Lemme 4.3.9. (Borel–Cantelli)

[i] Pour toute suite infinie $\{A_n, n \in \mathbb{N}\} \subset \mathcal{F}$,

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

[ii] Si la suite $\{A_n, n \in \mathbb{N}\}$ est indépendante,

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

Remarque 4.3.10. 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 :

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

$$\limsup A_n = A$$

PREUVE

(i)

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

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 $\log(1+x) \leq x$, $\forall x \geq -1$, 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 : $X : (\Omega, \mathcal{F}) \rightarrow (\mathbb{R}, \mathcal{B})$.

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

□

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 : $X : \Omega \rightarrow \mathbb{R}$ est une v.a.r., il suffit par exemple de vérifier que $\forall x \in \mathbb{R}$, $\{\omega; X(\omega) \leq x\} \in \mathcal{F}$.

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

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

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, $\mathcal{G}_1 \vee \mathcal{G}_2 \vee \dots \vee \mathcal{G}_n = \sigma(\mathcal{G}_1 \cup \mathcal{G}_2 \cup \dots \cup \mathcal{G}_n)$.

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 $\sigma(X) = \sigma(X_1) \vee \dots \vee \sigma(X_n)$.

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

$$\mathcal{C} = \{A_1 \times \mathbb{R}, A_1 \in \mathcal{B}_1\} \cup \{\mathbb{R} \times A_2, A_2 \in \mathcal{B}_2\}$$

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

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 :*

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

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 :

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

(iii)

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

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 $x = (x_i, \dots, x_d)$, où chaque x_i est égal soit à a_i , soit à b_i .

On pose $\Delta_A F = \sum_{\text{sommets de } A} \text{sign}_A(x) F(x)$, où $\text{sign}_A(x) = \pm 1$, suivant que le nombre de x_i égaux à a_i est pair ou impair.

- (i) $\Delta_A F \geq 0, \forall A = \prod_1^d]a_i, b_i]$, $a < b$.
- (ii) Si $x^{(k)} \downarrow x$ [i.e. $x_i^{(k)} \downarrow x_i, \forall i$], alors $F(x^{(k)}) \rightarrow F(x)$.
- (iii) Si $\exists j$ t.q. $x_j \rightarrow -\infty$ $F(x) \rightarrow 0$ Si $x_j \rightarrow +\infty, j = 1, \dots, d$, $F(x) \rightarrow 1$.

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.

Unicité Elle résulte du Théorème 1.4.7, 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{P} sur $(\mathbb{R}^d, \mathcal{B}_d)$ admet la densité f (par rapport à la mesure de Lebesgue) si $\forall B \in \mathcal{B}_d, \mathbb{P}(B) = \int_B f(x)dx$; 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 $B(p)$:

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

C'est la loi du gain à l'issue d'un jet de Pile ou Face.

2. **Loi binômiale** Soit Y_1, \dots, Y_n des v.a.r. **i.i.d.** (indépendantes et identiquement distribuées), de loi commune $B(p)$. Alors la loi de $X = Y_1 + \dots + Y_n$ est not'ee $B(n,p)$. Elle est caractérisée par :

$$\mathbb{P}(X = k) = C_n^k p^k (1-p)^{n-k}; \quad k = 0, 1, \dots, n \text{ (exercice)}$$

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 probabilité p d'obtenir Pile à chaque coup, notée $G(p)$.

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

En fait on peut aussi la caractériser par

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

Exercice 4.5.1. Soit X une v. a. à valeurs dans \mathbb{N} t. q. $\mathbb{P}(X > k) > 0$, $\forall k \in \mathbb{N}$. Alors il existe $p \in (0, 1)$ tel que la loi de X est la loi $G(p)$ ssi $\forall k, \ell \in \mathbb{N}$, $\mathbb{P}(X > k + \ell | X > k) = \mathbb{P}(X > \ell)$.

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 $k \geq m$,

$$\begin{aligned}\mathbb{P}(X = k) &= C_{k-1}^{m-1} p^{m-1} (1-p)^{k-m} p \\ &= C_{k-1}^{m-1} p^m (1-p)^{k-m}\end{aligned}$$

Exercice 4.5.2. 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 .

5. **Loi de Poisson** de paramètre $\lambda > 0$ notée $\mathcal{P}(\lambda)$

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

Exercice 4.5.3. La loi $\mathcal{P}(\lambda)$ est la limite des lois $B(n, p_n)$ lorsque $n \rightarrow \infty$, à condition que $np_n \rightarrow \lambda > 0$ quand $n \rightarrow \infty$.

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 $M \rightarrow \infty$ des lois uniformes sur

$$\left\{ \frac{1}{M}, \frac{2}{M}, \dots, 1 \right\}.$$

On peut considérer plus généralement la loi uniforme sur $[a, b]$, $\forall a < b$. 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 $P(X > x) = e^{-\lambda x}$, $\forall x > 0$. Cette caractérisation est celle qu'il faut toujours utiliser!

Exercice 4.5.4. Soit X une v. a. à valeurs dans \mathbb{R}_+ , telle que $\mathbb{P}(X > x) > 0$, $\forall x > 0$. Alors il existe $\lambda > 0$ t. q. $X \simeq \mathcal{E}(\lambda)$ ssi

$$\forall x, y > 0, \quad \mathbb{P}(X > x + y | X > x) = \mathbb{P}(X > y).$$

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 $\Gamma(p) = p\Gamma(p-1)$, $\Gamma(k) = (k-1)!$, $\Gamma(1) = 1$, $\Gamma(1/2) = \sqrt{\pi}$. Pour $p, \lambda > 0$, on définit la loi $\Gamma(p, \lambda)$ comme la loi de probabilité sur \mathbb{R}_+ de densité

$$\frac{\lambda^p}{\Gamma(p)} e^{-\lambda x} x^{p-1}.$$

Exercice 4.5.5. Si X et Y sont indépendantes, $X \simeq \Gamma(p, \lambda)$, $Y \simeq \Gamma(q, \lambda)$, alors $X + Y \simeq \Gamma(p+q, \lambda)$.

Exercice 4.5.6. Si $Y \simeq N(0, 1)$, $X = Y^2 \simeq \Gamma(1/2, 1/2)$.

Il résulte des deux exercices précédents que si Y_1, \dots, Y_n sont des v. a. r. i. i. d., de loi commune la loi $N(0, 1)$, alors

$$X = Y_1^2 + \dots + Y_n^2 \simeq \Gamma\left(\frac{n}{2}, \frac{1}{2}\right).$$

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).

7. **Loi beta** La loi beta(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

$$B(r, s) = \int_0^1 x^{r-1} (1-x)^{s-1} dx = \frac{\Gamma(r)\Gamma(s)}{\Gamma(r+s)}.$$

Exercice 4.5.7. Soit X_1, \dots, X_n des v. a. r. i. i. d., toutes de loi $\mathcal{U}([0, 1])$. Préciser la loi de $\min(X_1, \dots, X_n)$ et celle de $\max(X_1, \dots, X_n)$.

4.6 Variables aléatoires indépendantes

Définition 4.6.1. – La suite finie $\mathcal{C}_1 \dots \mathcal{C}_n$ de sous classes de \mathcal{F} est dite indépendante si l'égalité :

$$\mathbb{P}(B_1 \cap B_2 \cap \dots \cap B_n) = \mathbb{P}(B_1)\mathbb{P}(B_2) \dots \mathbb{P}(B_n)$$

est vraie $\forall B_i \in \mathcal{C}_i \cup \{\Omega\}$, $i = 1, \dots, n$.

– 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.

□

Nous allons tout d'abord relier cette notion d'indépendance avec celle de la Section 3.

Proposition 4.6.2. Soit $(A_1, \dots, A_n) \subset \mathcal{F}$. $\forall i \leq n$, 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. $(\mathbf{1}_{A_1}, \dots, \mathbf{1}_{A_n})$ est indépendante.

PREUVE (i) \Leftrightarrow (ii) a été établi au Corollaire 4.3.4. (ii) \Rightarrow (iii) résulte de ce que $\mathcal{F}_i = \sigma(\mathbf{1}_{A_i})$ □

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 $(\sigma(\mathcal{C}_1), \dots, \sigma(\mathcal{C}_n))$ 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 $Q : \mathcal{F} \rightarrow [0, 1]$ 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 1.4.7].

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 $\forall A_1 \in \sigma(\mathcal{C}_1), A_2 \in \mathcal{C}_2 \cup \{\Omega\}, \dots, A_n \in \mathcal{C}_n \cup \{\Omega\}$. Par récurrence, on démontre que cette égalité est vraie $\forall A_i \in \sigma(\mathcal{C}_i), i \leq n$. \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.*

$$\mathbb{P}_X(dx_1, \dots, dx_n) = \mathbb{P}_{X_1}(dx_1) \times \cdots \times \mathbb{P}_{X_n}(dx_n)$$

PREUVE Soit $B_1 \in \mathcal{B}_{d_1}, \dots, B_n \in \mathcal{B}_{d_n}$.

$$\begin{aligned} \mathbb{P}_X(B_1 \times \cdots \times B_n) &= \mathbb{P}((X_1, \dots, X_n) \in B_1 \times \cdots \times B_n) \\ &= \mathbb{P}(\{X_1 \in B_1\} \cap \cdots \cap \{X_n \in B_n\}) \\ &= \mathbb{P}_{X_1}(B_1) \dots \mathbb{P}_{X_n}(B_n), \end{aligned}$$

si la suite X_1, \dots, X_n est indépendante. Réciproquement, si cette égalité est satisfaite, alors $\forall B_1 \in \mathcal{B}_{d_1}, \dots, B_n \in \mathcal{B}_{d_n}$,

$$\mathbb{P}(\{X_1 \in B_1\} \cap \dots \cap \{X_n \in B_n\}) = \mathbb{P}(X_1 \in B_1) \dots \mathbb{P}(X_n \in B_n)$$

i.e.

$$\mathbb{P}(A_1 \cap \dots \cap A_n) = \mathbb{P}(A_1) \dots \mathbb{P}(A_n), \forall A_i \in \sigma(X_i).$$

\square

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 \mathbb{P}_X , admet la densité :*

$$f(x_1, \dots, x_n) = f_1(x_1) \cdots f_n(x_n)$$

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 $n_k \leq n$ sont indépendants).

PREUVE Utiliser 4.6.4, et l'associativité du produit de mesures. \square

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 $f_i : \mathbb{R}^{d_i} \rightarrow \mathbb{R}^{d'_i}$, $i = 1, \dots, n$, 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 $\sigma(f_i(X_i)) \subset \sigma(X_i)$ \square

Proposition 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 (Ω, \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 $\mathbb{E}(X_i)$, $i = 1 \dots d$.

\square

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 , $\varphi : \mathbb{R}^d \rightarrow \bar{\mathbb{R}}$ 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 :*

$$\int_{\Omega} \varphi \circ X(\omega) \mathbb{P}(d\omega) = \int_{\mathbb{R}^d} \varphi(x) \mathbb{P}_X(dx)$$

□

Cette proposition est essentielle pour le calcul pratique des espérances.

Etant donnée 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.

Proposition 4.7.3. *(Inégalité de Jensen).*

Soit X un vecteur aléatoire de dimension n , et $\varphi : \mathbb{R}^n \rightarrow \mathbb{R}$ 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 : $\forall x_0 \in \mathbb{R}^n \exists a \in \mathbb{R}^n$ t.q. :

$$(a, x - x_0) + \varphi(x_0) \leq \varphi(x), \forall x \in \mathbb{R}^n$$

$$(a, X - \mathbb{E}(X)) + \varphi(\mathbb{E}(X)) \leq \varphi(X) \text{ p.s.}$$

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 $1 \cdot |X|$].

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 :

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

Si X et Y sont deux v.a.r. de carré intégrable, on appelle covariance de X et Y la quantité :

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

Exercice 4.7.5. Calculer l'espérance et la variance de toutes les lois de probabilité présentées à la section 4.5.

Définition 4.7.6. 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 :

$$(4.1) \quad \text{Cov}(X) = \mathbb{E}((X - \mathbb{E}(X))(X - \mathbb{E}(X))')$$

$$(4.2) \quad = \mathbb{E}(XX') - \mathbb{E}(X)\mathbb{E}(X)', \text{ i. e. } \text{Cov}(X)_{ij} = \text{Cov}(X_i, X_j), \quad 1 \leq i, j \leq d.$$

Ces deux matrices sont autoadjointes et ≥ 0 .

Le caractère semidéfini positif de la matrice $\text{Cov}(X)$ se déduit aisément du fait que pour tout $u \in \mathbb{R}^d$, $\text{Var}(\langle u, X \rangle) = u' \text{Cov}(X) u$.

Proposition 4.7.7. (Inégalité de Bienaymé-Tchebicheff) Soit X une v.a.r., et $p > 0$. Alors :

$$\mathbb{P}(|X| \geq c) \leq \mathbb{E}(|X|^p)/c^p, \quad \forall c > 0$$

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

$$(ii) \mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$$

PREUVE

a) Supposons tout d'abord $X \geq 0, Y \geq 0$. On utilise 4.7.2, 4.6.4 et Fubini :

$$\begin{aligned} \mathbb{E}(XY) &= \int_{\mathbb{R}^2} xy \, d\mathbb{P}_{(X,Y)} \\ &= \int \int xy \, d\mathbb{P}_X(x) d\mathbb{P}_Y(y) \\ &= \left(\int x \, d\mathbb{P}_X(x) \right) \left(\int y \, d\mathbb{P}_Y(y) \right) \\ &= \mathbb{E}(X)\mathbb{E}(Y) \end{aligned}$$

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, \dots, X_d une suite de v. a. r. On a les implications suivantes*

$$\begin{aligned} & \text{La suite de v. a. } X_1, \dots, X_d \text{ est indépendante} \\ & \Downarrow \\ & \text{Les v. a. } X_1, \dots, X_d \text{ sont 2 à 2 indépendantes} \\ & \Downarrow \\ & \text{La matrice } \text{Cov}(X) \text{ est diagonale.} \end{aligned}$$

Lorsque $\text{Cov}(X_i, X_j) = 0, \forall 1 \leq i \neq j \leq n$, on dit que les v. a. X_1, \dots, X_d sont non corréliées.

Exercice 4.7.10. *Montrer par deux contre-exemples que les réciproques des deux implications du Corollaire sont fausses.*

Notons $\mathbf{1}$ le vecteur de \mathbb{R}^d dont toutes les coordonnées sont égales à 1. On a $\text{Var}(X_1 + \dots + X_n) = \text{Var}(\langle \mathbf{1}, X \rangle) = \mathbf{1}'\text{Cov}(X)\mathbf{1}$, dont on déduit la

Proposition 4.7.11. *Soit X_1, \dots, X_n des v. a. r. de carré intégrable.*

1. Alors

$$\text{Var}(X_1 + \dots + X_n) = \sum_{k=1}^n \text{Var}(X_k) + \sum_{1 \leq k \neq \ell \leq n} \text{Cov}(X_k, X_\ell).$$

2. Si les v. a. X_1, \dots, X_n sont deux à deux indépendantes, alors

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

On a la

Proposition 4.7.12. Soit X_k un v. a. de dimension d_k , $1 \leq k \leq n$. La suite X_1, \dots, X_n est indépendante si et seulement si pour toutes les fonctions boréliennes $f_k : \mathbb{R}^{d_k} \rightarrow \mathbb{R}_+$,

$$\mathbb{E} \left[\prod_{k=1}^n f_k(X_k) \right] = \prod_{k=1}^n \mathbb{E}[f_k(X_k)].$$

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 résultat au Théorème 5.2.5.

Chapitre 5

Fonction caractéristique et loi de Gauss

5.1 Fonction caractéristique d'une loi de probabilité sur \mathbb{R}

Soit \mathcal{K} une classe de fonctions de $C_b(\mathbb{R}; \mathbb{C})$.

Définition 5.1.1. On dit que la classe \mathcal{K} est séparante si $\forall \mathbb{P}$ et Q probabilités sur $(\mathbb{R}, \mathcal{B})$,

$$\left\{ \int f(x) \mathbb{P}(dx) = \int f(x) Q(dx), \forall f \in \mathcal{K} \right\} \Rightarrow \mathbb{P} = Q.$$

Lemme 5.1.2. L'espace $C_K(\mathbb{R})$ des fonctions continues à support compact est séparant.

PREUVE Soit $a \leq b \in \mathbb{R}$. on définit :

$$f_n(x) = \begin{cases} 0 & \text{si } x \notin]a - \frac{1}{n}, b + \frac{1}{n}[\\ n(x - a) + 1 & \text{si } x \in [a - \frac{1}{n}, a] \\ 1 & \text{si } x \in [a, b] \\ n(b - x) + 1 & \text{si } x \in [b, b + \frac{1}{n}] \end{cases}$$

Alors si \mathbb{P} est une probabilité sur $(\mathbb{R}, \mathcal{B})$,

$$\mathbb{P}([a, b]) = \lim_n \int_{\mathbb{R}} f_n(x) \mathbb{P}(dx)$$

Donc si $\int f d\mathbb{P} = \int f dQ, \forall f \in C_K(\mathbb{R})$, on a aussi $\mathbb{P}(I) = Q(I), \forall I$ intervalle fermé borné. Or la classe des intervalles fermés bornés est stable par intersection finie et engendre \mathcal{B} .

D'où $\mathbb{P} = Q$, d'après le Théorème 1.4.7. □

Théorème 5.1.3. *La classe des fonctions $\{e^{iux}, u \in \mathbb{R}\}$ est séparante.*

On va utiliser la : □

Proposition 5.1.4. *Soit $p \in \mathbb{N}$ 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 $f_m(x) \rightarrow f(x)$, uniformément sur $[-p, p]$.

PREUVE On va montrer le résultat pour f à valeurs dans \mathbb{C} . Soit g l'application de $(-p, +p]$ à valeurs dans 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 : $f = h \circ g$, où $h \in C(E; \mathbb{C})$. 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 de 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. $h_m(z) \rightarrow h(z)$ uniformément. Alors si $f_m = h_m \circ g$, la suite f_m a les propriétés de l'énoncé. □

PREUVE DU THÉORÈME 5.1.3 : Soit \mathbb{P} et Q deux probabilités sur $(\mathbb{R}, \mathcal{B})$ t.q.

$$\int e^{iux} \mathbb{P}(dx) = \int e^{iux} Q(dx), \quad \forall u \in \mathbb{R}.$$

Alors $\forall \ell \in \mathbb{N} \forall \alpha_1, \dots, \alpha_\ell \in \mathbb{C}, u_1, \dots, u_\ell \in \mathbb{R}$,

5.1. FONCTION CARACTÉRISTIQUE D'UNE LOI DE PROBABILITÉ SUR \mathbb{R} 83

(*)
$$\int_{\mathbb{R}} \left(\sum_1^{\ell} \alpha_k e^{iu_k x} \right) \mathbb{P}(dx) = \int_{\mathbb{R}} \left(\sum_1^{\ell} \alpha_k e^{iu_k x} \right) Q(dx).$$
 On va maintenant déduire de (*) et de 5.1.4 que

$$\int f d\mathbb{P} = \int f dQ, \forall f \in C_K(\mathbb{R}), \text{ d'où } \mathbb{P} = Q \text{ d'après le Lemme 5.1.2.}$$

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.

(**)
$$\int f_n d\mathbb{P} \rightarrow \int f d\mathbb{P} \text{ et } \int f_n dQ \rightarrow \int f dQ.$$

 $\exists n_0$ t.q. $\forall n \geq n_0, f(-n) = f(n) = 0$. Fixons $n \geq n_0$.
 $\exists N \in \mathbb{N}$ et $\alpha_k \in \mathbb{C}, k = -N, \dots, 0, 1, \dots, N$ t.q. si

$$f_n(x) = \sum_{-N}^N \alpha_k e^{i\pi k x/n}$$

$$|f(x) - f_n(x)| \leq \frac{1}{n}, \forall x \in [-n, n]$$

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. \square

Définition 5.1.5. – Etant donné une probabilité \mathbb{P} sur $(\mathbb{R}, \mathcal{B})$, on appelle **fonction caractéristique** de P la fonction φ définie sur \mathbb{R} , à valeurs dans \mathbb{C} , définie par :

$$\varphi(u) = \int_{\mathbb{R}} e^{iu x} \mathbb{P}(dx)$$

(= transformée de Fourier de la mesure \mathbb{P} ; d'après le Théorème 5.1.3 elle mérite bien son nom!).

– Etant donnée 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.

$$\varphi_X(u) = \int_{\mathbb{R}} e^{iu x} \mathbb{P}_X(dx) = \mathbb{E}[e^{iu X}]$$

\square

Proposition 5.1.6. Si $\varphi(u)$ est la f.c. d'une mesure de probabilité \mathbb{P} , alors :

- (i) $\varphi(0) = 1$
- (ii) $|\varphi(u)| \leq 1, \forall u \in \mathbb{R}$
- (iii) φ est uniformément continue sur \mathbb{R}
- (iv) $\varphi(-u) = \bar{\varphi}(u)$.

PREUVE ($d = 1$) (i) est trivial.

$$(ii) \left| \int e^{iux} \mathbb{P}(dx) \right| \leq \int |e^{iux}| \mathbb{P}(dx) = 1$$

(iii)

$$\begin{aligned} |\varphi(u+h) - \varphi(u)| &= \left| \int e^{iux} (e^{ihx} - 1) \mathbb{P}(dx) \right| \\ &\leq \int |e^{ihx} - 1| \mathbb{P}(dx) = \delta(h), \end{aligned}$$

où $\delta(h) \rightarrow 0$ quand $h \rightarrow 0$, par le Théorème de convergence dominée.

(iv) est trivial. □

Théorème 5.1.7. (Transformée de Fourier inverse) Soit $\varphi(u)$ la f.c. d'une probabilité \mathbb{P} sur $(\mathbb{R}, \mathcal{B})$, t.q. $\int_{\mathbb{R}} |\varphi(u)| du < \infty$.

Alors \mathbb{P} admet une densité $f(x)$ uniformément continue et bornée, donnée par :

$$(*) \quad f(x) = \frac{1}{2\pi} \int e^{-iux} \varphi(u) du.$$

On va utiliser le Lemme (qui par ailleurs est important en lui-même) :

Lemme 5.1.8. Soit X et Y deux v.a.r. indépendantes X de loi \mathbb{P} , et Y de loi Q admettant une densité bornée $g(x)$.

Alors la loi de la v.a.r. $X + Y$ admet la densité $h(x) = \int_{\mathbb{R}} g(x-y) \mathbb{P}(dy)$.

PREUVE Posons $Z = X + Y$. D'après le Lemme 5.1.2, il suffit de montrer que $\forall f \in C_K(\mathbb{R})$,

$$\mathbb{E}[f(Z)] = \int_{\mathbb{R}} f(z) \left(\int_{\mathbb{R}} g(z-x) \mathbb{P}(dx) \right) dz$$

5.1. FONCTION CARACTÉRISTIQUE D'UNE LOI DE PROBABILITÉ SUR \mathbb{R}^85

On va utiliser à deux reprises le Théorème de Fubini (vérifier en exercice qu'on a le droit de le faire).

$$\begin{aligned}
 \mathbb{E}[f(Z)] &= \mathbb{E}[f(X + Y)] \\
 &= \int \int f(x + y) \mathbb{P}(dx) g(y) dy \\
 &= \int \left[\int f(x + y) g(y) dy \right] \mathbb{P}(dx) \\
 &= \int \left[\int f(z) g(z - x) dz \right] \mathbb{P}(dx) \\
 &= \int f(z) \left[\int g(z - x) \mathbb{P}(dx) \right] dz.
 \end{aligned}$$

□

PREUVE du Théorème 5.1.7. On va admettre pour l'instant – on le vérifiera directement ci-dessous à la section 3 – qu'il existe au moins une loi Q sur \mathbb{R} , de densité bornée $g(x)$, 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

$$\begin{aligned}
 \frac{1}{2\pi} \int \varphi(u) \psi(u) e^{-iuy} du &= \frac{1}{2\pi} \int \int e^{iux} e^{-iuy} \psi(u) \mathbb{P}(dx) du \\
 &= \int \left[\frac{1}{2\pi} \int e^{-iu(y-x)} \psi(u) du \right] \mathbb{P}(dx) \\
 &= \int g(y - x) \mathbb{P}(dx).
 \end{aligned}$$

Cette fonction de y est, d'après le Lemme 5.1.8, la densité de la loi de $X + Y$, si X et Y sont indépendants X de loi \mathbb{P} et Y de loi Q .

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 (*).

Donc, par le raisonnement ci-dessus,

$$(**) \quad \frac{1}{2\pi} \int \varphi(u) \psi(\varepsilon u) e^{-iuy} du = \rho_\varepsilon(y)$$

= 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 :

$$(***) \quad \frac{1}{2\pi} \int \int F(y) \varphi(u) \psi(\varepsilon u) e^{-iuy} du dy = \mathbb{E}[F(X + \varepsilon Y)]$$

Quand $\varepsilon \rightarrow 0$

$$\begin{aligned} X + \varepsilon Y &\rightarrow X \text{ p.s.} \\ \psi(\varepsilon u) &\rightarrow 1, \forall u \in \mathbb{R} \end{aligned}$$

Faisant tendre $\varepsilon \rightarrow 0$ dans (***) à l'aide du Théorème de convergence dominée, on remarque que $|F(y)| \times |\varphi(u)|$ est $dy \times du$ intégrable, on obtient

$$\frac{1}{2\pi} \int \int F(y) \varphi(u) e^{-iuy} du dy = \mathbb{E}[F(X)]$$

Soit, en utilisant (*),

$$\int F(y) f(y) dy = \int F(y) \mathbb{P}(dy), \forall F \in C_K(\mathbb{R}).$$

f est donc la densité de \mathbb{P} , d'après le Lemme 5.1.2. Il est clair que f est bornée (car φ est intégrable). En outre

$$|f(x+h) - f(x)| \leq \frac{1}{2\pi} \int |\varphi(u)| \cdot |e^{-iuh} - 1| du,$$

qui tend vers 0 quand $h \rightarrow 0$, par le théorème de convergence dominée. Donc f est uniformément continue. \square

Théorème 5.1.9. 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 $|\delta(u)| \leq 2\mathbb{E}(|X|^n)$, $\forall u \in \mathbb{R}$; et $\delta(u) \rightarrow 0$, quand $u \rightarrow 0$.

PREUVE Si $f(y)$, définie sur \mathbb{R} , est continûment dérivable jusqu'à l'ordre n , on a :

$$\begin{aligned} f(y) &= f(0) + \int_0^y f'(u) du \\ &= f(0) + yf'(0) + \int_0^y \int_0^{u_1} f''(u_2) du_2 du_1 \\ &= \sum_{k=0}^n \frac{y^k}{k!} f^{(k)}(0) + \int_0^y \int_0^{u_1} \cdots \int_0^{u_{n-1}} (f^{(n)}(u_n) - f^{(n)}(0)) du_n \cdots du_1, \end{aligned}$$

5.2. FONCTION CARACTÉRISTIQUE D'UNE LOI DE PROBABILITÉ SUR \mathbb{R}^D 87

d'où :

$$(*) \quad e^{iuX} = \sum_0^n \frac{(iuX)^k}{k!} + \frac{(iu)^n}{n!} R(u, X), \text{ avec}$$

$$R(u, X) = n! \int_0^X \int_0^{v_1} \cdots \int_0^{v_{n-1}} (e^{iuv_n} - 1) dv_n \cdots dv_1, \text{ d'où } |R(u, X)| \leq 2|X|^n$$

On peut donc prendre l'espérance dans (*), d'où l'égalité de l'énoncé, avec

$$\begin{aligned} \delta(u) &= \mathbb{E}[R(u, X)] \\ \text{donc } |\delta(u)| &\leq 2 \mathbb{E}[|X|^n], \end{aligned}$$

et $\delta(u) \rightarrow 0$ quand $u \rightarrow 0$, par le Théorème de convergence dominée. \square

Corollaire 5.1.10. *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*

$$\varphi_X^{(k)}(0) = i^k \mathbb{E}(X^k); \quad k = 1, 2, \dots, n$$

PREUVE

$$\mathbb{E} \left[\frac{e^{i(u+h)X} - e^{iuX}}{h} \right] = \mathbb{E} [iX e^{i(u+\theta h)X}]$$

où θ est une v.a. à valeurs dans $]0, 1[$. Mais :

$$\begin{aligned} iX e^{i(u+\theta h)X} &\rightarrow iX e^{iuX} \text{ p.s., quand } h \rightarrow 0 \\ |X e^{i(u+\theta h)X}| &\leq |X|, \text{ qui est intégrable.} \end{aligned}$$

Donc, par le Théorème de convergence dominée,

$$\varphi'_X(u) = \mathbb{E}[iX e^{iuX}]$$

en particulier $\varphi'_X(0) = i\mathbb{E}(X)$. Le résultat pour les dérivées suivantes s'établit en itérant l'argument ci-dessus. \square

5.2 Fonction caractéristique d'une loi de probabilité sur \mathbb{R}^d

On vérifie comme dans le cas $d = 1$ que $C_K(\mathbb{R}^d)$ est séparant. On en déduit alors le :

Théorème 5.2.1. *La classe des fonctions.*

$$\left\{ \exp \left(i \sum_1^d u_k x_k \right), (u_1, \dots, u_d) \in \mathbb{R} \right\} \text{ est séparante.}$$

PREUVE Comme au Théorème 5.1.3, il suffit d'approcher toute fonction $f \in C_K(\mathbb{R}^d)$ par des combinaisons linéaires de fonctions de cette classe. Si $f \in C_K(\mathbb{R}^d)$, $\exists n_0$ t.q. $\forall n \geq 0$,

$$\text{supp}(f) \subset [-n, n]^d = R_n$$

Grâce au Théorème de Stone–Weierstrass la Proposition 5.1.4 se généralise en remplaçant le cercle par un tore de dimension d], $\exists f_n$, combinaison linéaire de

$$\exp[i \pi (m_1 x_1 + \dots + m_d x_d) n]$$

(les m_k entiers), qui approche 1 à $\frac{1}{n}$ près sur \mathbb{R}_n . La démonstration se termine comme pour le Théorème 5.1.3. \square

Définition 5.2.2. *Si \mathbb{P} est une probabilité sur \mathbb{R}^d , on définit sa fonction caractéristique comme la fonction de \mathbb{R}^d dans \mathbb{C} :*

$$\varphi(u) = \int_{\mathbb{R}^d} e^{i(u,x)} \mathbb{P}(dx).$$

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 : $\varphi_X(u) = \int e^{i(u,x)} \mathbb{P}_X(dx) = \mathbb{E}[e^{i(u,X)}]$

\square

Proposition 5.2.3. *Soit X un v.a. de dimension d . Si $\mathbb{E}(|X|^n) < \infty$, alors*

$$\varphi_X(u) = \sum_{k!} \frac{i^k}{k!} \mathbb{E}[(u, X)^k] + \delta(u) \frac{|u|^n}{n!}$$

$$\text{où } |\delta(u)| \leq 2\mathbb{E}(|X|^n)$$

$$\text{et } \delta(u) \rightarrow 0 \text{ quand } u \rightarrow 0.$$

PREUVE La proposition résulte du Théorème 5.1.9, en remarquant que

$$\varphi_X(u) = \varphi_{(u,X)}(1)$$

\square

Corollaire 5.2.4. *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*

$$\begin{aligned} \nabla \varphi_X(0) &= i \mathbb{E}(X) \\ \left(\frac{\partial^2 \varphi_X}{\partial u_i \partial u_j}(0) \right) &= -\mathbb{E}(X X'). \end{aligned}$$

Théorème 5.2.5. *Soit $X = (X_1, \dots, X_d)$ un v.a. de dimension d . La suite de v.a.r. (X_1, \dots, X_d) est indépendante si et seulement si $\forall u \in \mathbb{R}^d$,*

$$\varphi_X(u) = \prod_{k=1}^d \varphi_{X_k}(u_k)$$

On utilisera le résultat suivant (immédiat à partir de la Proposition 4.7.8).

Lemme 5.2.6. *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 5.2.5

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 :

$$\int_{\mathbb{R}^d} e^{i \sum u_k x_k} \mathbb{P}_X(dx) = \int_{\mathbb{R}^d} e^{i \sum u_k x_k} \mathbb{P}_{X_1}(dx_1) \cdots \mathbb{P}_{X_d}(dx_d)$$

Donc d'après le Théorème 5.2.1, $\mathbb{P}_X = \mathbb{P}_{X_1} \cdots \mathbb{P}_{X_d}$. Le résultat découle alors du Théorème 4.6.4.

□

Corollaire 5.2.7. *Si X_1, \dots, X_d sont des v.a.r. indépendantes, et si $S_d = X_1 + \dots + X_d$, alors*

$$\varphi_{S_d}(u) = \prod_1^d \varphi_{X_k}(u) \quad u \in \mathbb{R}$$

□

Attention : la réciproque du Corollaire 5.2.7 est fautive

Exercice 5.2.8. *Soit X un v.a. de dimension d . Sa loi P_X est uniquement déterminée par la donnée des lois de toutes les v.a.r. $\lambda \cdot X$, $\lambda \in \mathbb{R}^d$, $|\lambda| = 1$.*

5.3 Vecteurs aléatoires gaussiens

On désigne par $N(\mu, \sigma^2)$ la loi de Gauss sur \mathbb{R} , de densité $(\sigma\sqrt{2\pi})^{-1} \exp[-(x - \mu)^2/2\sigma^2]$ ($\mu \in \mathbb{R}$, $\sigma > 0$).

Une v.a.r. X est une v.a.r. gaussienne si soit $\exists \mu \in \mathbb{R}$, $\sigma > 0$, 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 $\mu = E(X)$, $\sigma^2 = Var(X)$.

Définition 5.3.1. *Un vecteur aléatoire X de dimension d est un vecteur aléatoire gaussien si et seulement si pour tout $c = (c_1, \dots, c_d) \in \mathbb{R}^d$, $c \cdot X \stackrel{\Delta}{=} \sum_1^d c_i X_i$ 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 5.3.2. *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 $\forall z \in \mathbb{C}$,

$$\frac{1}{\sqrt{2\pi}} \int e^{-\frac{x^2}{2} + zx} dx = e^{z^2/2}$$

Cette égalité est vraie $\forall z \in \mathbb{R}$, puisque :

$$\frac{1}{\sqrt{2\pi}} \int e^{-\frac{x^2}{2}+zx} dx = e^{z^2/2} \frac{1}{\sqrt{2\pi}} \int e^{-\frac{1}{2}(x-z)^2} dx = e^{z^2/2}$$

En utilisant $\int \sum_1^\infty |f_n(x)| dx < \infty \Rightarrow \int \sum_1^\infty f_n(x) dx = \sum_1^\infty \int f_n(x) dx$, on a :

$$\frac{1}{\sqrt{2\pi}} \int e^{-\frac{x^2}{2}+zx} dx = \sum_1^\infty \frac{z^n}{n! \sqrt{2\pi}} \int x^n e^{-\frac{x^2}{2}} dx.$$

$$\text{et } e^{z^2/2} = \sum_1^\infty \frac{z^{2n}}{2^n n!}$$

Ces deux fonctions coïncidant sur \mathbb{R} , les coefficients des séries coïncident. Donc les fonctions coïncident sur \mathbb{C} . \square

Remarque 5.3.3. Si l'on pose $g(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$, $\psi(u) = e^{-u^2/2}$, on a bien $g(x) = \frac{1}{2\pi} \int e^{-u^2/2-iux} du$, 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 5.1.7. \square

Si maintenant X est un v.a. gaussien d'espérance μ et de matrice de covariance Γ alors $\forall u \in \mathbb{R}^d$, $(u, X) \simeq N((u, \mu), (\Gamma u, u))$ et donc il résulte du Lemme :

$$\begin{aligned} \varphi_X(u) &= \mathbb{E}[e^{i(u, X)}] \\ &= e^{i(u, \mu) - \frac{1}{2}(\Gamma u, u)} \end{aligned}$$

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 $\mathbf{N}(0, \sigma^2)$. Sa f.c. est $\varphi_X(u) = \exp\left(-\frac{\sigma^2 u^2}{2}\right)$. En développant l'exponentielle, on obtient

$$e^{-\frac{\sigma^2 u^2}{2}} = \sum_1^\infty (-1)^n \frac{(\sigma u)^{2n}}{2^n n!}$$

D'où d'après le Théorème 5.1.9,

$$\mathbb{E}(X^{2n}) = \frac{\sigma^{2n} (2n)!}{2^n n!} \quad \mathbb{E}(X^{2n+1}) = 0.$$

Proposition 5.3.4. Soit $X = (X_1, \dots, X_d)$ un vecteur aléatoire gaussien. Alors la suite X_1, \dots, 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 $\Gamma = Cov(X)$ est diagonale, alors avec $\mu = \mathbb{E}(X)$,

$$\begin{aligned}\varphi_X(u) &= \exp[i(u, \mu) - \frac{1}{2}(\Gamma u, u)] \\ &= \prod_1^d \exp[iu_k \mu_k - \frac{1}{2}\Gamma_{kk} u_k^2] \\ &= \prod_1^d \varphi_{X_k}(u_k).\end{aligned}$$

Le résultat découle alors du Théorème 5.2.5 □

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. $cov(X, Y) = 0$) X et Y ne sont pas forcément indépendantes !

Exemple 5.3.5. $X \simeq N(0, 1)$. Soit β t.q. $\{P(|X| \leq \beta)\} = \frac{1}{2}$.

$$\text{On définit } Y(\omega) = \begin{cases} X(\omega) & \text{si } |X(\omega)| \leq \beta \\ -X(\omega) & \text{si } |X(\omega)| > \beta \end{cases}$$

On vérifie (exercice) que $Y \simeq N(0, 1)$, $E(XY) = 0$. Mais X et Y ne sont pas indépendants !

$$\begin{aligned}\mathbb{P}(|X| \leq \beta) &= \mathbb{P}(\{|X| \leq \beta\} \cap \{|Y| \leq \beta\}) \\ &\neq \mathbb{P}(|X| \leq \beta)\mathbb{P}(|Y| \leq \beta) \\ &= (\mathbb{P}(|X| \leq \beta))^2\end{aligned}$$

Par contre, si X_1, \dots, X_d sont des v.a.r. gaussiennes indépendantes, le vecteur aléatoire $X = (X_1, \dots, X_d)$ est un v.a. gaussien !

Proposition 5.3.6. *Soit Γ une matrice $n \times n$. Alors Γ est la matrice de covariance d'un v.a. gaussien, si et seulement si :*

- (i) $\Gamma = \Gamma'$
- (ii) $\Gamma \geq 0$

PREUVE **C.N** : déjà vu.

C.S. si Γ vérifie (i) et (ii), on peut définir $A = \Gamma^{1/2}$, on a $AA' = \Gamma$, et si $Y = (Y_1, \dots, Y_d)$ avec les Y_i indépendants, tous de loi $N(0; 1)$, $X = AY$ répond à la question. \square

Proposition 5.3.7. *Soit X un v.a. de dimension d , de loi $N(\mu, \Gamma)$ [i.e. gaussien avec $\mathbb{E}(X) = \mu$, $\text{cov}(X) = \Gamma$]. Si $\text{rang}(\Gamma) = d$, alors la loi de X admet la densité :*

$$f_X(x) = \frac{1}{(2\pi)^{d/2} \det(\Gamma)} \exp \left[-\frac{1}{2} (\Gamma^{-1}(x - \mu), x - \mu) \right]$$

PREUVE Posons $Y = \Gamma^{-1/2}(X - \mu)$. Alors la loi de Y est $N(0, I)$. Donc (cf. la Proposition 5.3.4) Y_1, \dots, 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 :

$$f_Y(y) = (2\pi)^{-d/2} \exp(-|y|^2/2)$$

Soit $\varphi \in C_K(\mathbb{R}^d)$.

$$\begin{aligned} \mathbb{E}[\varphi(X)] &= \mathbb{E}[\varphi(\mu + \Gamma^{1/2}Y)] \\ &= (2\pi)^{-d/2} \int_{\mathbb{R}^d} e^{-|y|^2/2} \varphi(\mu + \Gamma^{1/2}y) dy \\ &= (2\pi)^{-d/2} \int_{\mathbb{R}^d} e^{-|\Gamma^{-1/2}(x-\mu)|^2/2} \varphi(x) (\det \Gamma^{-1/2}) dx \end{aligned}$$

Le résultat découle alors de ce que $C_K(\mathbb{R}^d)$ est une classe séparante. \square

Exercice 5.3.8. *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 , $\varphi : S \rightarrow \mathbb{R}^d$ une application injective qui possède des dérivés partielles continues sur S , et dont le jacobien ne s'annule pas sur S .

94 CHAPITRE 5. FONCTION CARACTÉRISTIQUE ET LOI DE GAUSS

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 :

$$f_Y(x) = \begin{cases} \frac{f_X(\varphi^{-1}(x))}{|J_\varphi(\varphi^{-1}(x))|}, & \text{si } x \in \varphi(S) \\ 0 & \text{sinon} \end{cases}$$

où J_φ , jacobien de φ , est défini par :

$$J_\varphi(x) = \det \begin{pmatrix} \frac{\partial \varphi_1}{\partial x_1}(x) & \cdots & \frac{\partial \varphi_d}{\partial x_1}(x) \\ \vdots & & \vdots \\ \frac{\partial \varphi_1}{\partial x_d}(x) & \cdots & \frac{\partial \varphi_d}{\partial x_d}(x) \end{pmatrix}$$

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. $X_1, X_2, \dots, X_n, \dots$, telle que $\forall n$, la loi du v.a. (X_1, \dots, 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 :

(*) $\mathbb{P}_{n+1} \pi_{n+1,n}^{-1} = \mathbb{P}_n, \forall n \in \mathbb{N}$, où $\pi_{n+1,n}$ est la projection de \mathbb{R}^{n+1} sur \mathbb{R}^n définie par :

$$\pi_{n+1,n}(x_1, \dots, x_{n+1}) = (x_1, \dots, x_n)$$

Soit \mathbb{R}^∞ l’espace des suites infinies de nombres réels, $x = (x_1, \dots, x_n, \dots)$. On définit l’application $\pi_n : \mathbb{R}^\infty \rightarrow \mathbb{R}^n$ par :

$$\pi_n(x) = (x_1, \dots, x_n).$$

On appelle **σ -algèbre des ensembles cylindriques (de \mathbb{R}^∞) de dimension n** la σ -algèbre :

$$\mathcal{C}_n = \pi_n^{-1}(\mathcal{B}_n)$$

On appelle **algèbre des ensembles cylindriques** : $\mathcal{C} = \bigcup_{n \in \mathbb{N}} \mathcal{C}_n$ [pourquoi est-ce une algèbre ? et pas une σ -algèbre ?]

Se donner une famille \mathbb{P}_n de probabilités sur $(\mathbb{R}^n, \mathcal{B}_n)$, $\forall n \in \mathbb{N}$, qui vérifie la relation de compatibilité (*) est équivalent à se donner une fonction additive d'ensemble \mathbb{P} sur \mathcal{C} , telle que $\forall n \in \mathbb{N}$, $\mathbb{P}\pi_n^{-1} = \mathbb{P}_n$ [(*) permet de définir sans ambiguïté $\mathbb{P}(A)$, $\forall A \in \mathcal{C}$]. 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 $\forall n \in \mathbb{N}$, $\mathbb{P}\pi_n^{-1}$ soit une mesure de probabilité sur $(\mathbb{R}^n, \mathcal{B}_n)$, \exists une et une seule mesure de probabilité $\hat{\mathbb{P}}$ sur $(\mathbb{R}^\infty, \mathcal{B}_\infty)$, t.q. $\hat{\mathbb{P}}|_{\mathcal{C}} = \mathbb{P}$.*

Exemple d'application du Théorème 6.1.1

\exists 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. $\{X_n\}_{n \in \mathbb{N}}$ 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 $\mathbb{P}\pi_n^{-1} = Q^n$, $\forall n \in \mathbb{N}$. Enfin on pose

$$X_n(\omega) = \omega_n [\omega = (\omega_1, \omega_2, \dots)]$$

□

L'unicité dans le Théorème est évidente. Nous admettrons l'existence.

6.2 Convergence des variables aléatoires

Définition 6.2.1. *Soit $\{X_n, n \in \mathbb{N}\}$ 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) $X_n \rightarrow X$ p.s. si $\mathbb{P}(\{\omega, X_n(\omega) \xrightarrow[n \rightarrow \infty]{} X(\omega)\}) = 1$.
- (ii) $X_n \rightarrow X$ dans $L^p(\Omega, \mathcal{F}, P)$, ($p \geq 1$) si $X_n, X \in L^p(\Omega)$, et $\mathbb{E}(|X_n - X|^p) \rightarrow 0$, quand $n \rightarrow \infty$.
- (iii) $X_n \rightarrow X$ en probabilité (noté $X_n \xrightarrow{p} X$) si $\forall \varepsilon > 0$, $\mathbb{P}(|X_n - X| > \varepsilon) \rightarrow 0$, quand $n \rightarrow \infty$. □

Exercice 6.2.2. – Montrer que si $1 \leq p < q$, $X_n \rightarrow X$ dans $L^q(\Omega)$ entraîne que $X_n \rightarrow X$ dans $L^p(\Omega)$.

– Montrer par un contre-exemple que $X_n \xrightarrow{p} X$ n'entraîne pas $X_n \xrightarrow{p.s.} X$. (on pourra choisir $(\Omega, \mathcal{F}, \mathbb{P}) = ([0, 1], \mathcal{B}([0, 1]), \lambda)$, $X = 0$, $X_n = \mathbf{1}_{B_n}$, les B_n étant des boréliens de $[0, 1]$ "bien choisis").

Proposition 6.2.3. Supposons que $\exists p \geq 1$ t.q. $X_n \rightarrow X$ dans $L^p(\Omega)$. Alors $X_n \xrightarrow{p} X$.

PREUVE Fixons $\varepsilon > 0$. D'après l'inégalité de Bienaymé Tchebycheff (Proposition 4.7.7),

$$P(|X_n - X| > \varepsilon) \leq \varepsilon^{-p} E(|X_n - X|^p).$$

□

Proposition 6.2.4. Si $X_n \rightarrow X$ p.s., alors $X_n \xrightarrow{p} X$.

PREUVE Sous l'hypothèse, si $Y_n = \inf(|X_n - X|, 1)$, $Y_n \rightarrow 0$ p.s., et $|Y_n| \leq 1$ p.s., $\forall n$. Donc, d'après le Théorème de convergence dominée, $E(Y_n) = E(|Y_n|) \rightarrow 0$, i.e. $Y_n \rightarrow 0$ dans $L^1(\Omega)$, donc en probabilité d'après la Proposition 6.2.3. □

Proposition 6.2.5. Si $X_n \rightarrow X$ p.s., et si $\exists Z$ v.a.r. \mathbb{P} -intégrable t.q. $|X_n|^p \leq Z$ p.s., $\forall n$ (avec $p \geq 1$), alors

$$X_n \rightarrow X \text{ dans } L^p(\Omega).$$

PREUVE

$$\mathbb{P}(|X|^p > Z) \leq \mathbb{P}\left(\bigcup_{n=1}^{\infty} \{|X_n|^p > Z\} \cup \{X_n \not\rightarrow X\}\right) = 0$$

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 $|X_n - X|^p \rightarrow 0$ p.s. Le résultat découle donc du Théorème de convergence dominée. □

La Proposition 6.2.5 est aussi un corollaire de 6.2.4, et de :

Théorème 6.2.6. Si $X_n \xrightarrow{p} X$, et si $\exists Z$ v.a.r. P -intégrable t.q. $|X_n|^p \leq Z$ p.s., $\forall n$ (avec $p \geq 1$), alors

$$X_n \rightarrow X \text{ dans } L^p(\Omega)$$

PREUVE

$$\forall k \in \mathbb{N}, \left\{ |X| > Z^{1/p} + \frac{1}{k} \right\} \subset \{|X_n|^p > Z\} \cup \left\{ |X - X_n| > \frac{1}{k} \right\}$$

$$\text{d'où } \mathbb{P} \left(|X| > Z^{1/p} + \frac{1}{k} \right) \leq \lim_{n \rightarrow \infty} \mathbb{P} \left(|X - X_n| > \frac{1}{k} \right) = 0$$

ceci $\forall k > 0$, donc $|X|^p \leq Z$ p.s., donc $X_n, X \in L^p(\Omega)$,

$$\text{et } |X_n - X|^p \leq 2^p Z = Y$$

$$\begin{aligned} \mathbb{E}(|X_n - X|^p) &\leq \int_{\{|X_n - X|^p > M\}} |X_n - X|^p d\mathbb{P} + \int_{\{\varepsilon < |X_n - X|^p \leq M\}} |X_n - X|^p d\mathbb{P} + \varepsilon \\ &\leq \int_{\{Y > M\}} Y d\mathbb{P} + M \mathbb{P}(|X_n - X|^p > \varepsilon) + \varepsilon \end{aligned}$$

$$\text{Donc } \limsup_{n \rightarrow \infty} \mathbb{E}(|X_n - X|^p) \leq \int_{\{Y > M\}} Y d\mathbb{P} + \varepsilon, \forall 0 < \varepsilon < M$$

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.

Donc par le Théorème de convergence monotone, $\int_{\{Y > M\}} Y d\mathbb{P} \rightarrow 0$.

Finalement $0 \leq \liminf_n \mathbb{E}(|X_n - X|^p) \leq \limsup_n \mathbb{E}(|X_n - X|^p) \leq 0$. D'où $\mathbb{E}(|X_n - X|^p) \rightarrow 0$ quand $n \rightarrow \infty$. \square

Corollaire 6.2.7. Si $X_n \xrightarrow{p} X$, et si $\exists Y$ v.a.r. \mathbb{P} -intégrable t.q. $|X_n| \leq Y$ p.s. $\forall n$, alors X_n et X sont \mathbb{P} -intégrables, et $\mathbb{E}(X_n) \rightarrow \mathbb{E}(X)$.

PREUVE On applique le Théorème avec $p = 1$. Donc

$$|\mathbb{E}(X_n) - \mathbb{E}(X)| \leq \mathbb{E}(|X_n - X|) \rightarrow 0.$$

\square

Proposition 6.2.8. Si $X_n \xrightarrow{p} X$, alors il existe une sous-suite $\{X_{n_k}\}$ t.q. $X_{n_k} \rightarrow X$ p.s.

PREUVE Posons $Z_n = \inf(|X_n - X|, 1)$. $Z_n \xrightarrow{p} 0$, et $|Z_n| \leq 1$. Donc, d'après le Théorème 6.2.6, $Z_n \rightarrow 0$ dans $L^1(\Omega)$. Donc, d'après le Théorème 3.2.1, \exists une sous-suite Z_{n_k} t.q. $Z_{n_k} \rightarrow 0$ p.s., d'où $X_{n_k} - X \rightarrow 0$ p.s. \square

Proposition 6.2.9. $X_n \xrightarrow{p} X$ 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 6.2.8.

C.S. Si la suite $\{X_n\}$ ne converge pas en probabilité vers X , alors $\exists \varepsilon, \eta > 0$ et une sous-suite $\{X_{n_k}\}$ t.q. $\mathbb{P}(|X - X_{n_k}| \geq \varepsilon) > \eta, \forall k$. Alors aucune sous-suite extraite de la sous-suite $\{X_{n_k}\}$ ne converge p.s. vers X . \square

Une conséquence immédiate de la Proposition 6.2.9 est

Corollaire 6.2.10. 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

$$g(X_n) \xrightarrow{p} g(X).$$

\square

Proposition 6.2.11. Soit $\{X_n, n \in \mathbb{N}\}$ et X des v.a.r. (ou des vecteurs aléatoires de dimension d). Les assertions suivantes sont équivalentes :

- (i) $X_n \rightarrow X$ p.s.
- (ii) $\sup_{j \geq n} |X_j - X| \xrightarrow{p} 0$.

PREUVE

$$\begin{aligned} \mathbb{P}(\limsup_n \{|X_n - X| > \varepsilon\}) &= \mathbb{P}\left(\bigcap_{n=1}^{\infty} \bigcup_{j=n}^{\infty} \{|X_j - X| > \varepsilon\}\right) \\ &= \lim_n \mathbb{P}\left(\sup_{j \geq n} |X_j - X| > \varepsilon\right) \end{aligned}$$

Donc :

$$\begin{aligned} X_n \rightarrow X \text{ p. s.} &\Leftrightarrow \forall \varepsilon > 0, \mathbb{P}(\limsup_n \{|X_n - X| > \varepsilon\}) = 0 \\ &\Leftrightarrow \forall \varepsilon > 0, \mathbb{P}(\sup_{j \geq n} |X_j - X| > \varepsilon) \rightarrow 0, \text{ quand } n \rightarrow \infty. \end{aligned}$$

\square

Proposition 6.2.12. Soit $\{X_n, n \in \mathbb{N}\}$ 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 $n \rightarrow \infty$.

(ii) \exists une v.a.r. X t.q. $X_n \rightarrow X$ p.s. [resp. un v.a. de dimension d]

PREUVE

– (ii) \Rightarrow (i) :

$$\begin{aligned} \sup_{j \geq n} |X_j - X_n| &\leq \sup_{j \geq n} |X_j - X| + |X - X_n| \\ \mathbb{P} \left(\sup_{j \geq n} |X_j - X_n| > \varepsilon \right) &\leq \mathbb{P} \left(\sup_{j \geq n} |X_j - X| > \varepsilon/2 \right) + \mathbb{P}(|X - X_n| > \varepsilon/2) \\ &\rightarrow 0. \end{aligned}$$

– (i) \Rightarrow (ii) Posons $Y_n = \sup_{j \geq n} |X_j - X_n|$. D'après la Proposition 6.2.8, \exists une sous-suite $\{Y_{n_k}\}$ t.q. $Y_{n_k} \rightarrow 0$ p.s. Alors a fortiori $\sup_{\ell \geq k} |X_{n_\ell} - X_{n_k}| \rightarrow 0$ p.s., donc la suite $\{X_{n_k}(\omega)\}$ est p.s. une suite de Cauchy, qui converge dans \mathbb{R} [resp. \mathbb{R}^d]. Donc, d'après le Théorème 4.1.3, \exists une v.a.r. [resp. un v.a. de dimension d] X t.q. $X_{n_k} \rightarrow X$ 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 $n_k \rightarrow +\infty$, 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 :

$$0 \leq \liminf_{n \rightarrow \infty} \mathbb{P} \left(\sup_{j \geq n} |X_j - X| > \varepsilon \right) \leq \limsup_{n \rightarrow \infty} \mathbb{P} \left(\sup_{j \geq n} |X_j - X| > \varepsilon \right) \leq 0$$

Le résultat découle alors de la Proposition 6.2.11. □

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*

$$\frac{X_1 + \cdots + X_n}{n} \xrightarrow{p} \mathbb{E}(X_1), \text{ quand } n \rightarrow \infty.$$

PREUVE Posons $\mu = \mathbb{E}(X_1)$, $\sigma^2 = \text{Var}(X_1)$, $S_n = X_1 + \cdots + X_n$. $\mathbb{E}(S_n) = n\mu$, $\text{Var}(S_n) = n\sigma^2$. Appliquons l'inégalité de Bienaymé–Tchebycheff (4.7.7) :

$$\mathbb{P}\left(\left|\frac{S_n}{n} - \mu\right| > \varepsilon\right) = \mathbb{P}(|S_n - n\mu| > n\varepsilon) \leq \frac{n\sigma^2}{n^2\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2} \rightarrow 0.$$

□

L'indépendance de $\{X_n\}$ a servi uniquement à établir assurer que

$$\text{Var}(S_n) = \sum_1^n \text{Var}(X_k).$$

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. $\text{Cov}(X_k, X_\ell) = 0$, si $k \neq \ell$. On pose :*

$$\mu_i = \mathbb{E}(X_i), \sigma_i^2 = \text{Var}(X_i^2), \quad i \geq 1. \text{ Alors}$$

$$\text{si } \frac{1}{n^2} \sum_1^n \sigma_i^2 \rightarrow 0, \text{ quand } n \rightarrow \infty,$$

$$\mathbb{P}\left(\left|\frac{X_1 + \cdots + X_n}{n} - \frac{\mu_1 + \cdots + \mu_n}{n}\right| \geq \varepsilon\right) \rightarrow 0, \quad \forall \varepsilon > 0.$$

□

6.4 Loi 0 – 1 de Kolmogorov

Etant donnée une suite de v.a. $\{X_n, n \geq 1\}$, on note $\sigma(X_n, X_{n+1}, \dots)$ la plus petite σ -algèbre qui contient $\sigma(X_{n+p}), \forall p \geq 0$.

Définition 6.4.1. On appelle tribu asymptotique de la suite $\{X_n, n \geq 1\}$ la tribu $\mathcal{G} = \bigcap_{n=1}^{\infty} \sigma(X_n, X_{n+1}, \dots)$.

□

Exemple 6.4.2. – $E = \left\{ \omega; \frac{X_1(\omega) + \dots + X_n(\omega)}{n} \text{ converge} \right\} \in \mathcal{G}$
 – 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, \dots) et (j_1, j_2, \dots) deux sous-suites **disjointes** d'indices. Alors les σ -algèbres : $\mathcal{F}_1 = \sigma(X_{i_1}, X_{i_2}, \dots)$ et $\mathcal{F}_2 = \sigma(X_{j_1}, X_{j_2}, \dots)$ sont **indépendantes**.

PREUVE Définissons les algèbres :

$$\mathcal{F}_1^0 = \bigcup_{n=1}^{\infty} \sigma(X_{i_1}, \dots, X_{i_n}) \text{ et } \mathcal{F}_2^0 = \bigcup_{n=1}^{\infty} \sigma(X_{j_1}, \dots, X_{j_n})$$

Si $A \in \mathcal{F}_1^0$ et $B \in \mathcal{F}_2^0$, alors d'après la Proposition 4.6.6,

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

Il nous reste à montrer que (*) est vraie, $\forall A \in \mathcal{F}_1, B \in \mathcal{F}_2$. L'égalité est triviale 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 : $Q_1(A) \triangleq \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$, $A \in \mathcal{F}_1$.

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 (*) $\forall A \in \mathcal{F}_1 \forall B \in \mathcal{F}_2^0$.

Soit maintenant $A \in \mathcal{F}_1$, t.q. $\mathbb{P}(A) \neq 0$. On définit $Q_2(B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)}$.

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 $\mathcal{F}_0 = \bigcup_{n=1}^{\infty} \sigma(X_1, \dots, X_n)$. 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 1.4.8,

\exists une suite $\{E_j\} \subset \mathcal{F}_0$ t.q. $\mathbb{P}(E \Delta E_j) \rightarrow 0, j \rightarrow \infty$. Donc $\mathbb{P}(E_j) \rightarrow \mathbb{P}(E)$ et $\mathbb{P}(E \cap E_j) \rightarrow \mathbb{P}(E)$.

Mais $\exists n_j$ t.q. $E_j \in \sigma(X_1, \dots, X_{n_j})$ et $E \in \sigma(X_{n_j+1}, \dots)$, 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 $j \rightarrow \infty$

$$\mathbb{P}(E) = (\mathbb{P}(E))^2$$

$$\text{Donc } \mathbb{P}(E) = 0 \text{ ou } 1.$$

□

On sait maintenant que beaucoup d'événements qui nous intéressent sont de probabilité soit 0, soit 1. Comment distinguer en 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 :*

$$A_n = \{|X_n| > n\}$$

$$\text{Alors } \mathbb{E}|X_1| < \infty \Leftrightarrow \mathbb{P}[\limsup A_n] = 0$$

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, \dots, X_n une suite de v.a.r. indépendantes, toutes de carré intégrable et d'espérance nulle. On pose $S_k = X_1 + \dots + X_k. \forall \alpha > 0,$*

$$\mathbb{P}\left(\max_{1 \leq k \leq n} |S_k| \geq \alpha\right) \leq \frac{1}{\alpha^2} \mathbb{E}(S_n^2)$$

PREUVE On pose $A = \{\max_{1 \leq k \leq n} |S_k| > \alpha\}$,

$$A_k = \{|S_k| \geq \alpha\} \cap \{|S_j| < \alpha; j = 1, 2, \dots, k-1\}$$

$$A = \bigcup_1^n A_k \text{ et } A_k \cap A_\ell = \emptyset, \text{ si } k \neq \ell$$

$$\begin{aligned} \mathbb{E}(S_n^2) &\geq \int_A S_n^2 d\mathbb{P} \\ &= \sum_{k=1}^n \int_{A_k} S_n^2 d\mathbb{P} \\ &= \sum_{k=1}^n \int_{A_k} [(S_k^2 + 2S_k(S_n - S_k) + (S_n - S_k)^2)] d\mathbb{P} \\ &\geq \sum_{k=1}^n \left[\int_{A_k} S_k^2 d\mathbb{P} + 2\mathbb{E}(\mathbf{1}_{A_k} S_k) \mathbb{E}(S_n - S_k) \right] \\ &\geq \alpha^2 \sum_1^n \mathbb{P}(A_k) = \alpha^2 \mathbb{P}(A). \end{aligned}$$

On a utilisé à l'avant-dernière ligne : $A_k \in \sigma(X_1, \dots, X_k)$, $S_n - S_k \in \sigma(X_{k+1}, \dots, X_n)$, $\mathbf{1}_{A_k} S_k$ et $S_n - S_k$ sont indépendantes. \square

Théorème 6.5.2. Soit $\{X_n\}_{n \geq 1}$ une suite de v.a.r. indépendantes, t.q.

(i) $\mathbb{E}(X_n) = 0, \forall n$

(ii) $\sum_1^\infty \mathbb{E}(X_n^2) < \infty$

Alors $S_n = X_1 + \dots + X_n$ converge p.s.

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), $\mathbb{P}[\sup_{j \geq n} |S_j - S_n| > \varepsilon] \xrightarrow{n \rightarrow \infty} 0, \forall \varepsilon > 0$

Soit $\sup_{j \geq n} |S_j - S_n| \xrightarrow{p} 0$, donc d'après la Proposition 6.2.12, \exists une v.a.r. S t.q. $S_n \rightarrow S$ 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 $\frac{1}{n} \sum_1^n X_k \rightarrow 0$ p.s.

PREUVE Posons $S_n = X_1 + \dots + X_n, \sigma_n^2 = \mathbb{E}(X_n^2)$.

Il suffit de montrer que $\forall \varepsilon > 0$, si l'on pose

$$B_k = \{\omega \exists n, 2^k \leq n < 2^{k+1}, |S_n(\omega)| > n \varepsilon\}$$

$$\mathbb{P}(\limsup B_k) = 0$$

Pour cela (cf. Lemme de Borel–Cantelli), il suffit de vérifier que $\sum_{k=1}^{\infty} \mathbb{P}(B_k) < \infty$. $\omega \in B_k \Rightarrow |S_n(\omega)| > \varepsilon 2^k$ pour un $n < 2^{k+1}$.

Donc d'après le Théorème 6.5.1,

$$\begin{aligned} \mathbb{P}(B_k) &\leq \mathbb{P}\left(\sup_{n \leq 2^{k+1}} |S_n| > \varepsilon 2^k\right) \\ &\leq \frac{1}{\varepsilon^2 2^{2k}} \sum_{n=1}^{2^{k+1}} \sigma_n^2. \end{aligned}$$

$$\begin{aligned} \sum_0^\infty \mathbb{P}(B_k) &\leq \sum_{k=0}^\infty \sum_{n=1}^{2^{k+1}} \frac{\sigma_n^2}{\varepsilon^2 2^{2k}} \\ &= \frac{1}{\varepsilon^2} \sum_{n=1}^\infty \sigma_n^2 \sum_{k: 2^{k+1} \geq n} 2^{-2k}. \end{aligned}$$

Il reste à montrer que $\sum_{k: 2^{k+1} \geq n} 2^{-2k} \leq c/n^2$. Or

$$\begin{aligned} \sum_{k_n}^\infty 2^{-2k} &= 2^{-2k_n} \sum_0^\infty 2^{-2k} \\ &= \frac{4}{3} 2^{-2k_n}. \end{aligned}$$

Mais

$$2^{k+1} \geq n \Rightarrow 2^{-2k_n} \times 2^{-2} \leq \frac{1}{n^2}$$

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 $\{X_n, n \in \mathbb{N}\}$ une suite de v.a.r. i.i.d. (indépendantes et identiquement distribuées).*

- (i) Si $\mathbb{E}(|X_1|) < \infty$, alors $\frac{X_1 + \cdots + X_n}{n} \rightarrow \mathbb{E}(X_1)$ p.s.
- (ii) Si $\mathbb{E}(|X_1|) = +\infty$, alors $\frac{X_1 + \cdots + X_n}{n}$ diverge p.s.

PREUVE (i) On va se ramener au Théorème 6.6.1. On pose

$$\tilde{X}_n = X_n \mathbf{1}_{\{|X_n| \leq n\}}, \quad A_n = \{|X_n| > n\} = \{X_n \neq \tilde{X}_n\}$$

D'après la Proposition 6.4.5, puisque $\mathbb{E}(|X_1|) < \infty$, $\mathbb{P}(\limsup A_n) = 0$. D'où (Cf 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 :

$$\frac{\tilde{X}_1 + \cdots + \tilde{X}_n}{n} \rightarrow \mathbb{E}(X_1) \text{ p.s.}$$

Mais comme $\mathbb{E}(\tilde{X}_n) \rightarrow \mathbb{E}(X_1)$ [convergence dominée], il suffit de montrer que

$$\frac{1}{n} \sum_1^n (\tilde{X}_k - \mathbb{E}(\tilde{X}_k)) \rightarrow 0 \text{ p.s.}$$

Ceci résulte du Théorème 6.6.1, à condition que $\sum_1^\infty \frac{\text{Var}(\tilde{X}_n)}{n^2} < \infty$.

Or

$$\text{Var}(\tilde{X}_n) \leq \mathbb{E}(\tilde{X}_n^2)$$

Donc il nous reste à montrer que

$$\sum_1^\infty \mathbb{E}(\tilde{X}_n^2)/n^2 < \infty.$$

Mais :

$$\begin{aligned} \sum_1^\infty \mathbb{E}(\tilde{X}_n^2)/n^2 &= \sum_1^\infty \frac{1}{n^2} \int_{\{|x| \leq n\}} x^2 \mathbb{P}_{X_1}(dx) \\ &= \sum_{n=1}^\infty \sum_{k=1}^n \frac{1}{n^2} \int_{\{k-1 < |x| \leq k\}} x^2 \mathbb{P}_{X_1}(dx) \\ &= \sum_{k=1}^\infty \left(\sum_{n=k}^\infty \frac{1}{n^2} \right) \int_{\{k-1 < |x| \leq k\}} x^2 \mathbb{P}_{X_1}(dx) \\ &< \sum_{k=1}^\infty \frac{2}{k} \int_{\{k-1 < |x| \leq k\}} x^2 \mathbb{P}_{X_1}(dx) \\ &\leq 2\mathbb{E}(|X_1|), \end{aligned}$$

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 $S_n = X_1 + \dots + X_n$. 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 :

$$\frac{X_n}{n} = \frac{S_n}{n} - \left(\frac{n-1}{n}\right) \frac{S_{n-1}}{n-1} \rightarrow 0 \text{ p.s.}$$

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. \square

Chapitre 7

La convergence en loi

7.1 Définition et premières propriétés

Tous les v.a. seront supposés construits sur un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$. Vérifions que comme indiqué au Chapitre 4, $C_b(\mathbb{R}^d)$ est séparable.

Lemme 7.1.1. *Soit R et Q deux probabilités sur $(\mathbb{R}^d, \mathcal{B}_d)$, telles que $R(F) = Q(F)$, $\forall F$ fermé $\subset \mathbb{R}^d$. Alors $R = Q$.*

PREUVE La classe des fermés est stable par intersection, et engendre \mathcal{B}_d .
 \square

Lemme 7.1.2. *Soit R et Q deux probabilités sur $(\mathbb{R}^d, \mathcal{B}_d)$, telles que $\int f(x)R(dx) = \int f(x)Q(dx)$, $\forall f \in C_b(\mathbb{R}^d)$. Alors $R = Q$.*

PREUVE Soit F fermé $\subset \mathbb{R}^d$. Posons $f_\varepsilon(x) = \varphi\left(\frac{1}{\varepsilon}\rho(x, F)\right)$, avec $\varphi(x) = \mathbf{1}_{\{x \leq 0\}} + (1-x)^+ \mathbf{1}_{\{x \geq 0\}}$. $f_\varepsilon \in C_b(\mathbb{R}^d)$ donc

$$\int f_\varepsilon(x)dR(x) = \int f_\varepsilon(x)dQ(x),$$

d'où par convergence monotone,

$$R(F) = Q(F), \forall F \text{ fermé}$$

\square

On peut remplacer l'hypothèse du Lemme 7.1.1 par la condition $R(F) = Q(F)$, $\forall F$ fermé borné. On en déduit alors que $C_K(\mathbb{R}^d)$ est séparable.

Définition 7.1.3. Soit $\{Q_n, n \in \mathbb{N}\}$ et Q des probabilités sur $(\mathbb{R}^d, \mathcal{B}_d)$. On dit que Q_n **converge étroitement** vers Q , et on note

$$Q_n \Rightarrow Q$$

$$\int f(x)Q_n(dx) \rightarrow \int f(x)Q(dx) \text{ si } \forall f \in C_b(\mathbb{R}^d)$$

□

Définition 7.1.4. Soit $\{X_n, n \in \mathbb{N}\}$ 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

$$\mathbb{P}_{X_n} \Rightarrow \mathbb{P}_X$$

$$\text{i.e. } \forall f \in C_b(\mathbb{R}^d), \quad \mathbb{E}[f(X_n)] \rightarrow \mathbb{E}[f(X)]$$

□

Définition 7.1.5. Soit Q une probabilité sur $(\mathbb{R}^d, \mathcal{B}_d)$. On dit que $A \in \mathcal{B}_d$ est un **ensemble de Q -continuité** si $Q(\partial A) = 0$, où $\partial A = \bar{A} - \overset{\circ}{A}$.

□

Théorème 7.1.6. Soit $\{Q_n, n \in \mathbb{N}\}$ et Q des probabilités sur $(\mathbb{R}^d, \mathcal{B}_d)$. Il y a équivalence entre :

- (i) $Q_n \Rightarrow Q$
- (ii) $\limsup Q_n(F) \leq Q(F), \forall F$ fermé
- (iii) $\liminf Q_n(G) \geq Q(G), \forall G$ ouvert
- (iv) $\lim Q_n(A) = Q(A), \forall A$ ensemble de Q -continuité.

Théorème 7.1.7. Soit $\{X_n, n \in \mathbb{N}\}$ et X des v.a. de dimension d . Il y a équivalence entre :

- (i) $X_n \xrightarrow{\mathcal{L}} X$
- (ii) $\limsup \mathbb{P}(X_n \in F) \leq \mathbb{P}(X \in F), \forall F$ fermé
- (iii) $\liminf \mathbb{P}(X_n \in G) \geq \mathbb{P}(X \in G), \forall G$ ouvert
- (iv) $\lim \mathbb{P}(X_n \in A) = \mathbb{P}(X \in A), \forall A$ ensemble de \mathbb{P}_X -continuité.

Exercice 7.1.8. Montrer par un contre exemple que si $Q_n \Rightarrow Q$, on peut avoir $Q_n(A) \not\rightarrow Q(A)$.

PREUVE DU THÉORÈME 7.1.6 (i) \Rightarrow (ii) Soit F fermé, $f_\varepsilon(x)$ comme au Lemme 7.1.1

$$\begin{aligned} \limsup Q_n(F) &\leq \lim_n \int f_\varepsilon(x) Q_n(dx) \\ &= \int f_\varepsilon(x) Q(dx) \rightarrow Q(F) \text{ quand } \varepsilon \rightarrow 0. \end{aligned}$$

(ii) \Rightarrow (iii) Soit G un ouvert, $F = G^c$.

$$\begin{aligned} \liminf Q_n(G) &= 1 - \limsup Q_n(F) \\ &\geq 1 - Q(F) \\ &= Q(G). \end{aligned}$$

(ii)+(iii) \Rightarrow (iv)

$$\begin{aligned} Q(\overset{\circ}{A}) &\leq \liminf Q_n(\overset{\circ}{A}) \\ &\leq \liminf Q_n(A) \\ &\leq \limsup Q_n(A) \\ &\leq \limsup Q_n(\bar{A}) \\ &\leq Q(\bar{A}) \end{aligned}$$

Donc si $Q(\overset{\circ}{A}) = Q(\bar{A})$, $Q_n(A) \rightarrow Q(A)$

(iv) \Rightarrow (i) Soit $f \in C_b(\mathbb{R}^d)$. $\exists K$ t.q. $|f(x)| \leq K$, $\forall x$. Soit $\varepsilon > 0$, et $\alpha_0 < \alpha_1 < \dots < \alpha_\ell$ t.q.

a) $\alpha_0 < -K < K < \alpha_\ell$

b) $\alpha_i - \alpha_{i-1} < \varepsilon$

c) $Q(\{x; f(x) = \alpha_i\}) = 0$, $i = 0, \dots, \ell$

L'existence de tels α_i résultera du Lemme 7.1.9 Posons

$$A_i = \{x; \alpha_{i-1} < f(x) \leq \alpha_i\}$$

f étant continue,

$$\bar{A}_i \subset \{x; \alpha_{i-1} \leq f(x) \leq \alpha_i\}$$

$$\overset{\circ}{A}_i \supset \{x; \alpha_{i-1} < f(x) < \alpha_i\}$$

d'où

$$\partial A_i \subset \{x; f(x) = \alpha_{i-1}\} \cup \{x; f(x) = \alpha_i\}$$

Donc

$$Q(\partial A_i) = 0, \quad i = 1, \dots, \ell$$

Or

$$\left| \int f dQ_n - \sum_1^\ell \alpha_i Q_n(A_i) \right| \leq \varepsilon$$

$$\left| \int f dQ - \sum_1^\ell \alpha_i Q(A_i) \right| \leq \varepsilon$$

$$\sum_1^\ell \alpha_i Q_n(A_i) \rightarrow \sum_1^\ell \alpha_i Q(A_i)$$

donc

$$\begin{aligned} \int f dQ - 2\varepsilon &\leq \sum_1^\ell \alpha_i Q(A_i) - \varepsilon \\ &= \lim \sum_1^\ell \alpha_i Q_n(A_i) - \varepsilon \\ &\leq \liminf \int f dQ_n \\ &\leq \limsup \int f dQ_n \\ &\leq \lim \sum_1^\ell \alpha_i Q_n(A_i) + \varepsilon \\ &= \sum_1^\ell \alpha_i Q(A_i) + \varepsilon \\ &\leq \int f dQ + 2\varepsilon \end{aligned}$$

ceci $\forall \varepsilon > 0$.

□

Lemme 7.1.9. $E = \{\alpha \in \mathbb{R}; Qf^{-1}(\alpha) > 0\}$ est au plus dénombrable.

PREUVE Soit F la fonction de répartition de la loi de probabilité Qf^{-1} sur $(\mathbb{R}, \mathcal{B})$. E est l'ensemble des points de discontinuité de F . Posons

$$E_n = \left\{x; F(x^+) - F(x) \geq \frac{1}{n}\right\}$$

$$\text{card}E_n \leq n, \text{ et } E_n \uparrow E,$$

donc E est au plus dénombrable. \square

Théorème 7.1.10. Cas $d = 1$ Soit $\{X_n, n \in \mathbb{N}\}$ 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) $F_n(x) \rightarrow F(x), \forall x \in C(F)$, ensemble des points de continuité de F .

Remarque 7.1.11. D'après le Lemme 7.1.9, le complémentaire de $C(F)$ est au plus dénombrable.

PREUVE (i) \Rightarrow (ii) résulte du Théorème 7.1.6 (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

$$\mathbb{P}(X_n \in]a, b]) \rightarrow \mathbb{P}(X \in]a, b])$$

Soit G un ouvert de \mathbb{R} . G est une réunion dénombrable d'éléments disjoints de \mathcal{A} , $G = \bigcup_1^\infty A_m$.

$$\mathbb{P}\left(X \in \bigcup_{n \leq M} A_m\right) = \lim_n \mathbb{P}\left(X_n \in \bigcup_{m \leq M} A_m\right)$$

$$\leq \liminf \mathbb{P}(X_n \in G)$$

D'où en faisant tendre $M \rightarrow \infty$,

$$\mathbb{P}(X \in G) \leq \liminf \mathbb{P}(X_n \in G)$$

Il suffit alors d'utiliser le Théorème 7.1.7 \square

Un autre résultat en dimension 1 est le

Lemme 7.1.12. (Fatou) Soit $\{X_n, n \geq 0\}$ et X des v. a. r., telles que $X_n \xrightarrow{\mathcal{L}} X$. Supposons que $X_n \geq 0$ p. s. pour tout $n \geq 0$. Alors $X \geq 0$ p. s. et

$$\mathbb{E}(X) \leq \liminf_n \mathbb{E}(X_n).$$

PREUVE D'après le Théorème 7.1.7 (iii) avec $G =]-\infty, 0[$, $\mathbb{P}(X < 0) \leq \liminf_n \mathbb{P}(X_n < 0) = 0$. En outre, pour tout $K > 0$,

$$\begin{aligned} \mathbb{E}(X \wedge K) &= \lim_n \mathbb{E}(X_n \wedge K) \\ &\leq \liminf_n \mathbb{E}(X_n). \end{aligned}$$

Il reste à faire tendre $K \rightarrow \infty$ pour conclure. \square

Définition 7.1.13. Une famille $\{Q_\lambda, \lambda \in \Lambda\}$ de probabilités sur $(\mathbb{R}^d, \mathcal{B}_d)$ est dite **tendue** si $\forall \varepsilon > 0, \exists K$ compact $\subset \mathbb{R}^d$ t.q.

$$Q_\lambda(K) \geq 1 - \varepsilon, \forall \lambda \in \Lambda$$

Une famille $\{X_\lambda, \lambda \in \Lambda\}$ de v.a. de dimension d est dite **tendue** si la famille $\{\mathbb{P}_{X_\lambda}, \lambda \in \Lambda\}$ est tendue, i.e. si $\forall \varepsilon > 0, \exists K$ compact $\subset \mathbb{R}^d$ t.q.

$$\mathbb{P}(X_\lambda \in K) \geq 1 - \varepsilon, \forall \lambda \in \Lambda$$

\square

Si Q est une probabilité sur $(\mathbb{R}^d, \mathcal{B}_d)$, $\forall \varepsilon > 0, \exists$ 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 la famille finie de lois de probabilité est tendue.

Théorème 7.1.14. Si $X_n \xrightarrow{\mathcal{L}} X$, alors la famille $\{X_n, n \in \mathbb{N}\}$ de vecteurs aléatoires de dimension d est tendue.

PREUVE Soit $\varepsilon > 0, k > 0$ t.q. $\mathbb{P}(|X| \leq k) \geq 1 - \varepsilon/2; k' = k+1, \varphi \in C_b(\mathbb{R}^d)$ t.q. $\varphi(x) \leq 1, \varphi(x) = 1$ si $|x| \leq k, \varphi(x) = 0$ si $|x| \geq k'$.

$$\mathbb{E}[\varphi(X_n)] \rightarrow \mathbb{E}[\varphi(X)] \geq 1 - \varepsilon/2.$$

Donc $\exists n_0$ t.q. $\forall n \geq n_0 \mathbb{E}[\varphi(X_n)] \geq 1 - \varepsilon$ et $\forall n \geq n_0 \mathbb{P}(|X_n| \leq k') \geq 1 - \varepsilon$. D'après la remarque ci-dessus, $\exists K$ compact de \mathbb{R}^d , tel que

$$\mathbb{P}(X_n \in K) \geq 1 - \varepsilon, \forall n < n_0.$$

7.2. RELATION AVEC LES AUTRES TYPES DE CONVERGENCE, ET PROPRIÉTÉS SUPPLÉMENTAIRES

Posons $\bar{K} = K \cup \{x; |x| \leq k'\}$. \bar{K} est un compact, et

$$\mathbb{P}(X_n \in \bar{K}) \geq 1 - \varepsilon; \forall n \in \mathbb{N}$$

□

On a la réciproque partielle du Théorème 7.1.14, que nous admettrons :

Théorème 7.1.15. *Soit $\{X_n, n \in \mathbb{N}\}$ une suite **tendue** de vecteurs aléatoires de dimension d . Alors il existe une sous-suite $\{X_{n_k}, k \in \mathbb{N}\}$ et un vecteur aléatoire X de dimension d , t.q.*

$$X_{n_k} \xrightarrow{\mathcal{L}} X.$$

□

Corollaire 7.1.16. *Soit $\mathcal{K} \subset C_b(\mathbb{R}^d; \mathbb{C})$ une classe séparante, et $\{X_n, n \in \mathbb{N}\}$ 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 $\forall f \in \mathcal{K}$, la suite $\{\mathbb{E}[f(X_n)], n \in \mathbb{N}\}$ est convergente.

PREUVE La C.N. est évidente. Démontrons la C.S.

D'après le Théorème 7.1.15, \exists 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 $\exists g \in C_b(\mathbb{R}^d)$, $\varepsilon > 0$, et une sous-suite $\{X_{n_\ell}; \ell \in \mathbb{N}\}$ t.q. :

$$(*) \quad \mathbb{E}|g(X_{n_\ell}) - g(X)| \geq \varepsilon, \quad \forall \ell \in \mathbb{N}.$$

A nouveau d'après le Théorème 7.1.15, 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 $\mathbb{E}[f(X)] = \mathbb{E}[f(X')]$, $\forall f \in \mathcal{K}$, ce qui entraîne que $\mathbb{P}_X = \mathbb{P}'_{X'}$, donc $\mathbb{E}[g(X)] = \mathbb{E}[g(X')]$, $\forall g \in C_b(\mathbb{R}^d)$, ce qui contredit (*). □

7.2 Relation avec les autres types de convergence, et propriétés supplémentaires.

Théorème 7.2.1. *Soit $\{X_n, n \in \mathbb{N}\}$ et X des v.a. de dimension d , définis sur un même espace de probabilité (Ω, \mathcal{F}, P) . Alors*

$$(i) \quad X_n \xrightarrow{p} X \Rightarrow X_n \xrightarrow{\mathcal{L}} X$$

$$(ii) \quad \left. \begin{array}{l} X_n \xrightarrow{\mathcal{L}} X \\ P_X = \delta_{x_0} \end{array} \right\} \Rightarrow X_n \xrightarrow{p} X.$$

Commentaires

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. $\exists Y$ 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 $\mathbb{P}_X = \mathbb{P}_Y = \delta_{x_0}$, $X = Y$ 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, $\mathbb{E}[f(X_n)] \rightarrow \mathbb{E}[f(X)]$ par le Corollaire 6.2.7.

(ii)

$$\mathbb{P}(|X_n - X| > \varepsilon) \leq \mathbb{P}(|X_n - x_0| \geq \varepsilon)$$

$$\limsup \mathbb{P}(|X_n - X| > \varepsilon) \leq \limsup \mathbb{P}(X_n \in B^c(x_0; \varepsilon)) \leq \mathbb{P}(X \in B^c(x_0; \varepsilon)) = 0,$$

$$\text{où } B(x_0; \varepsilon) = \{x \in \mathbb{R}^d; |x - x_0| < \varepsilon\}.$$

□

Contrairement aux autres types de convergence, la convergence en loi n'est pas "additive", au sens où

$$\left. \begin{array}{l} X_n \xrightarrow{\mathcal{L}} X \\ Y_n \xrightarrow{\mathcal{L}} Y \end{array} \right\} \not\Rightarrow X_n + Y_n \xrightarrow{\mathcal{L}} X + Y$$

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} .

7.2. RELATION AVEC LES AUTRES TYPES DE CONVERGENCE, ET PROPRIÉTÉS SUPPLÉMENTAIRES

Théorème 7.2.2. Soit $\{X_n, Y_n; n \in \mathbb{N}\}$ et X des v.a. de dimension d , $x_0 \in \mathbb{R}^d$. On suppose :

$$\begin{aligned} X_n &\xrightarrow{\mathcal{L}} X \\ Y_n &\xrightarrow{p} x_0 \end{aligned}$$

Alors

$$(i) X_n + Y_n \xrightarrow{\mathcal{L}} X + x_0$$

$$(ii) \text{ si } d = 1; X_n Y_n \xrightarrow{\mathcal{L}} x_0 X$$

PREUVE (i) Tout fermé de \mathbb{R}^d peut s'écrire sous la forme $x_0 + F \triangleq \{z; z - x_0 \in F\}$, avec F fermé de \mathbb{R}^d .

$$\begin{aligned} \{X_n + Y_n \in x_0 + F\} &= \{X_n + Y_n \in x_0 + F\} \cap \{|Y_n - x_0| \leq \varepsilon\} \\ &\quad \cup \{X_n + Y_n \in x_0 + F\} \cap \{|Y_n - x_0| > \varepsilon\} \\ &\subset \{X_n \in F_\varepsilon\} \cup \{|Y_n - x_0| > \varepsilon\} \end{aligned}$$

$$\begin{aligned} \limsup \mathbb{P}(X_n + Y_n \in x_0 + F) &\leq \mathbb{P}(X \in F_\varepsilon) \quad \forall \varepsilon > 0 \\ \limsup \mathbb{P}(X_n + Y_n \in x_0 + F) &\leq \mathbb{P}(X \in F) \\ &= \mathbb{P}(X + x_0 \in x_0 + F) \end{aligned}$$

(ii)

a) **Le cas** $x_0 = 0$ Il s'agit de montrer que $X_n Y_n \xrightarrow{p} 0$;

$$\begin{aligned} \{|X_n Y_n| > \varepsilon\} &\subset \{|X_n Y_n| > \varepsilon\} \cap \{|X_n| \leq M\} \cup \{|X_n Y_n| > \varepsilon\} \cap \{|X_n| > M\} \\ &\subset \{|Y_n| > \frac{\varepsilon}{M}\} \cup \{|X_n| > M\} \end{aligned}$$

D'après le Théorème 7.1.14 $\forall \eta > 0, \exists M_0$ t.q. $\mathbb{P}(|X_n| > M_0) \leq \eta, \forall n$.

$$\begin{aligned} \mathbb{P}(|X_n Y_n| > \varepsilon) &\leq \mathbb{P}\left(|Y_n| > \frac{\varepsilon}{M_0}\right) + \eta \\ \limsup \mathbb{P}(|X_n Y_n| > \varepsilon) &\leq \eta, \quad \forall \eta > 0 \\ &= 0, \quad \forall \varepsilon > 0. \end{aligned}$$

b) **Le cas** $x_0 \neq 0$ On pose :

$$\begin{aligned} X_n Y_n &= X_n(Y_n - x_0) + x_0 X_n \\ &\xrightarrow{\mathcal{L}} 0 + x_0 X \end{aligned}$$

d'après a) et (i).

□

Notons que l'argument de (ii) a) montre aussi le

Corollaire 7.2.3. *Si $X_n \rightarrow 0$ en probabilité, et la suite $\{Y_n\}$ est tendue, alors $X_n Y_n \rightarrow 0$ en probabilité.*

Théorème 7.2.4. *Soit $\{X_n, Y_n \in \mathbb{N}\}$, 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 :

Proposition 7.2.5. *Soit $\{X_n, n \in \mathbb{N}\}$ et X des v.a. de dimension d , et $\varphi : \mathbb{R}^d \rightarrow \mathbb{R}^k$ une application continue. On suppose que $X_n \xrightarrow{\mathcal{L}} X$. Alors $\varphi(X_n) \xrightarrow{\mathcal{L}} \varphi(X)$.*

La proposition résulte immédiatement des définitions.

□

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

$$X_n \xrightarrow{\mathcal{L}} X \Rightarrow \varphi_n(u) \rightarrow \varphi(u), \quad \forall u \in \mathbb{R}^d,$$

et d'autre part

$$\left. \begin{array}{l} \varphi_n(u) \rightarrow \varphi(u), \forall u \in \mathbb{R} \\ \{X_n\} \text{ est tendue} \end{array} \right\} \Rightarrow X_n \xrightarrow{\mathcal{L}} X$$

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

$$T\varphi(u) = 1 - \frac{1}{2}\varphi(u) - \frac{1}{2}\varphi(-u).$$

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 $\theta > 0$,*

$$\mathbb{P}(|X| > 1/\theta) \leq \frac{\alpha}{\theta} \int_0^\theta T\varphi(u) du.$$

PREUVE Notons que $T\varphi(u) = \mathbb{E}[1 - \cos(uX)]$ donc 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. *Soit $\{X_n, n \in \mathbb{N}\}$ une suite de v.a. de dimension d . Notons φ_n la fonction caractéristique de X_n . On suppose :*

(i) $\lim_{n \rightarrow \infty} \varphi_n(u)$ existe, $\forall u \in \mathbb{R}^d$.

(ii) $\varphi(u) \stackrel{\Delta}{=} \lim_n \varphi_n(u)$ est continue en $u = 0$.

Alors \exists un v.a. X de dimension d tel que :

$$X_n \xrightarrow{\mathcal{L}} X,$$

et φ est la fonction caractéristique de X .

PREUVE Il suffit, pour pouvoir appliquer le Corollaire 7.1.16, de montrer que la suite $\{X_n, n \in \mathbb{N}\}$ est tendue. Mais la tension de cette suite est équivalente à celle des d suites $\{X_n^i, n \in \mathbb{N}\}$, $1 \leq i \leq d$, 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 $\{X_n, n \in \mathbb{N}\}$.

Nous supposons donc dorénavant que $d = 1$. Posons $Q_n = \mathbb{P}_{X_n}$. D'après le Lemme 7.3.1, pour tout $\theta > 0$, si l'on pose

$$S_\theta = \{x; |x| \leq 1/\theta\},$$

on a

$$Q_n(S_\theta^c) \leq \frac{\alpha}{\theta} \int_0^\theta T\varphi_n(u) du.$$

D'après (ii) et le Théorème de convergence dominée,

$$\limsup Q_n(S_\theta^c) \leq \frac{\alpha}{\theta} \int_0^\theta T\varphi(u) du$$

Mais grâce à (ii), $T\varphi(u)$ est continue en $u = 0$, et est nulle pour $u = 0$ par définition de T .

Donc

$$\frac{1}{\theta} \int_0^\theta T\varphi(u) du \rightarrow 0, \text{ quand } \theta \rightarrow 0.$$

Fixons $\varepsilon > 0$. Soit alors $a > 0$ tel que

$$\limsup_n Q_n(S_a^c) \leq \varepsilon/2$$

Donc $\exists n_0$ t.q. $\forall n \geq n_0$, $Q_n(S_a^c) \leq \varepsilon$. Soit K un compact tel que $K \supset S_a$, et $Q_n(K) \geq 1 - \varepsilon$, $\forall n < n_0$. Alors $Q_n(K) \geq 1 - \varepsilon$, $\forall n$.

$\{Q_n; n \in \mathbb{N}\}$ est donc tendue. □

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.

$$\varphi_n(u) \rightarrow \varphi(u), \quad \forall u \in \mathbb{R}^d$$

Alors

$$X_n \xrightarrow{\mathcal{L}} X.$$

□

Théorème 7.3.4. Soit $\{X_n, n \in \mathbb{N}\}$ et X des vecteurs aléatoires de dimension d , et $\{\varphi_n, n \in \mathbb{N}\}$ et φ les fonctions caractéristiques correspondantes.

Supposons que $X_n \xrightarrow{\mathcal{L}} X$.

Alors $\varphi_n(u) \rightarrow \varphi(u)$ uniformément sur tout compact.

PREUVE

- a) Montrons tout d'abord que $X_n \xrightarrow{\mathcal{L}} X$ entraîne que les $\{\varphi_n, n \in \mathbb{N}\}$ sont équicontinues (on pose $\varphi = \varphi_0$) i.e. $\sup_{u \in \mathbb{R}^d} \sup_n |\varphi_n(u+h) - \varphi_n(u)| \rightarrow 0$, quand $h \rightarrow 0$.

Soit K un compact de \mathbb{R}^d . On pose $Q_n = \mathbb{P}_{X_n}$.

$$\begin{aligned} |\varphi_n(u+h) - \varphi_n(u)| &\leq \int_K |e^{i(u+h,x)} - e^{i(u,x)}| dQ_n(x) + 2Q_n(K^c) \\ &\leq \sup_{x \in K} |e^{i(h,x)} - 1| + 2Q_n(K^c) \end{aligned}$$

Fixons $\varepsilon > 0$. D'après le Théorème 7.1.14, $\exists K_\varepsilon$ compact tel que $Q_n(K_\varepsilon^c) \leq \varepsilon/4$. Alors $\exists \rho_\varepsilon$ t.q. $\forall |h| \leq \rho_\varepsilon$, $\sup_{x \in K_\varepsilon} |e^{i(h,x)} - 1| \leq \varepsilon/2$.

d'où

$$\sup_{u \in \mathbb{R}^d} \sup_n |\varphi_n(u+h) - \varphi_n(u)| \leq \varepsilon, \quad \forall |h| \leq \rho_\varepsilon,$$

et la même chose avec φ_n remplacée par φ .

- b) Soit maintenant K un compact quelconque de \mathbb{R}^d . Soit $u_1, u_2, \dots, u_m \in K$ 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 $\sup_{k \leq m} |\varphi_n(u_k) - \varphi(u_k)| \rightarrow 0$, quand $n \rightarrow \infty$.

7.4 Le Théorème Central Limite

Théorème 7.4.1. Soit $\{X_n, n \in \mathbb{N}\}$ une suite de v.a. de dimension d , i.i.d., la loi commune ayant un moment d'ordre 2 fini. On note $\mu = E(X_1)$, $\Gamma = \text{Cov}(X_1)$.

Alors

$$\frac{X_1 + \cdots + X_n - n\mu}{\sqrt{n}} \xrightarrow{\mathcal{L}} N(0, \Gamma).$$

Etablissons tout d'abord le :

Lemme 7.4.2. Soit z_1, \dots, z_m et z'_1, \dots, 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

$$\begin{aligned} S_n &= X_1 + \cdots + X_n \\ \varphi_{S_n/\sqrt{n}}(u) &= \varphi_{S_n} \left(\frac{u}{\sqrt{n}} \right) = \varphi_{X_1}^n \left(\frac{u}{\sqrt{n}} \right). \end{aligned}$$

Par hypothèse, $\mathbb{E}(|X_1|^2) < \infty$. Donc, d'après la Proposition 5.2.3

$$\varphi_{X_1}(u) = 1 - \frac{1}{2} \mathbb{E}[(u, X_1)^2] + \delta(u)|u|^2$$

avec $\delta(u) \rightarrow 0$, si $u \rightarrow 0$, et $|\delta(u)| \leq 2\mathbb{E}(|X_1|^2)$.

Donc

$$(*) \quad \varphi_{X_1} \left(\frac{u}{\sqrt{n}} \right) = 1 - \frac{1}{2n} \mathbb{E}[(u, X_1)^2] + \delta \left(\frac{u}{\sqrt{n}} \right) \frac{|u|^2}{n}.$$

Il résulte du Lemme 7.4.2 que pour n assez grand,

$$\left| \varphi_{X_1}^n \left(\frac{u}{\sqrt{n}} \right) - \left(1 - \frac{\mathbb{E}[(u, X_1)^2]}{2n} \right)^n \right| \leq n \left| \varphi_{X_{n_1}} \left(\frac{u}{\sqrt{n}} \right) - \left(1 - \frac{\mathbb{E}[(u, X_1)^2]}{2n} \right) \right|$$

Et d'après (*) le membre de droite de cette inégalité tend vers 0, quand $n \rightarrow +\infty$. De plus :

$$\left(1 - \frac{t^2}{2n} \right)^n \rightarrow e^{-t^2/2}, \quad \forall t \in \mathbb{R}$$

Donc

$$\varphi_{S_{n/\sqrt{n}}}(u) \rightarrow \exp \left[-\frac{1}{2}(\Gamma u, u) \right], \quad \forall u \in \mathbb{R}^d.$$

Le Théorème découle alors du Théorème 7.3.3

7.5 Uniforme intégrabilité

Soit X une v. a. r. p. s. finie. Alors

$$|X| \mathbf{1}_{|X| \geq K} \rightarrow 0 \quad \text{p. s. quand } K \rightarrow \infty.$$

Si en outre X est intégrable, alors par convergence dominée,

$$\mathbb{E}(|X| \mathbf{1}_{|X| \geq K}) \rightarrow 0, \quad \text{quand } K \rightarrow \infty.$$

On pose la

Définition 7.5.1. *La suite de v. a. r. $\{X_n, n \geq 1\}$ est dite uniformément intégrable si*

$$\lim_{K \rightarrow \infty} \sup_{n \geq 1} \mathbb{E}(|X_n| \mathbf{1}_{|X_n| \geq K}) = 0.$$

Bien sûr, une suite uniformément intégrable est formée de v. a. intégrables, puisque il existe K' t. q.

$$\sup_{n \geq 1} \mathbb{E}(|X_n| \mathbf{1}_{|X_n| \geq K'}) \leq 1, \quad \text{donc } \mathbb{E}|X_n| \leq 1 + K'.$$

On a le

Théorème 7.5.2. 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

$$(7.1) \quad \mathbb{E}(X_n) \rightarrow \mathbb{E}(X), \text{ quand } n \rightarrow \infty.$$

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 7.1.12) 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

$$|x - \varphi_K(x)| \leq |x| \mathbf{1}_{\{|x| > K\}}, \quad \forall x \in \mathbb{R}.$$

On a alors

$$\begin{aligned} \mathbb{E}\varphi_K(X_n) &\rightarrow \mathbb{E}\varphi_K(X), \\ |\mathbb{E}X_n - \mathbb{E}\varphi_K(X_n)| &\leq \mathbb{E}(|X_n| \mathbf{1}_{\{|X_n| > K\}}), \\ |\mathbb{E}X - \mathbb{E}\varphi_K(X)| &\leq \mathbb{E}(|X| \mathbf{1}_{\{|X| > K\}}), \end{aligned}$$

Donc

$$\limsup_n |\mathbb{E}X - \mathbb{E}X_n| \leq \sup_n \mathbb{E}(|X_n| \mathbf{1}_{\{|X_n| > K\}}) + \mathbb{E}(|X| \mathbf{1}_{\{|X| > K\}}),$$

et le membre de droite tend vers 0 quand $K \rightarrow \infty$. 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 $n \rightarrow \infty$,

$$\mathbb{E}(X_n \mathbf{1}_{\{X_n \leq K\}}) \rightarrow \mathbb{E}(X \mathbf{1}_{\{X \leq K\}}).$$

Donc si $\mathbb{E}X_n \rightarrow \mathbb{E}X$, par différence,

$$\mathbb{E}(X_n \mathbf{1}_{\{X_n > K\}}) \rightarrow \mathbb{E}(X \mathbf{1}_{\{X > K\}}).$$

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

$$\mathbb{E}(X_n \mathbf{1}_{\{X_n > K\}}) \leq \varepsilon, \quad \forall n \geq n_0.$$

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. \square

Corollaire 7.5.3. *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. *$\|X_n\|_1 \rightarrow \|X\|_1$, quand $n \rightarrow \infty$.*

Si en outre $X_n \rightarrow X$ en probabilité, chacune des deux conditions ci-dessus est équivalente à

$$X_n \rightarrow X \quad \text{dans } L^1(\Omega).$$

Remarque 7.5.4. *Rappelons que $\|X\|_1$ désigne la norme de la v. a. X dans $L^1(\Omega)$, i. e. $\|X\|_1 = \mathbb{E}(|X|)$.*

PREUVE Le Théorème, appliqué à la suite de v. a. $\{|X_n|, n \geq 1\}$, implique la première partie du Corollaire. Si $X_n \rightarrow X$ en probabilité, alors $|X - X_n| \rightarrow 0$ 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 $\mathbb{E}|X - X_n| \rightarrow 0$. Réciproquement, si $X_n \rightarrow X$ dans $L^1(\Omega)$, $\|X_n\|_1 \rightarrow \|X\|_1$. \square

Concluons ce chapitre en donnant des critères de tension et d'uniforme intégrabilité.

Proposition 7.5.5. *Une suite $\{X_n, n \geq 1\}$ de v. a. r. est **tendue** si et seulement si il existe une application $G : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ telle que $G(x) \rightarrow \infty$, quand $x \rightarrow \infty$, 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 $H : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ telle que $H(x)/x \rightarrow \infty$, quand $x \rightarrow \infty$, 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 \square

Chapitre 8

Espérance et Probabilité 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 :

$$\mathbb{P}_X(A|Y = y) \triangleq \mathbb{P}(X \in A|Y = y), \quad A \in \mathcal{B}.$$

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,

$$\mathbb{P}_X(A|Y = y) = \frac{\mathbb{P}(\{X \in A\} \cap \{Y = y\})}{\mathbb{P}(Y = y)}$$

$A \rightarrow P_X(A|Y = y)$ est une probabilité, et on peut aussi définir, par exemple pour φ mesurable bornée,

$$\mathbb{E}[\varphi(X)|Y = y] \triangleq \int \varphi(x) \mathbb{P}_X(dx|Y = y),$$

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 apparaîtront 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, \dots, y_n , avec $\mathbb{P}(Y = y_i) > 0$, $i = 1 \dots n$. 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 \exists une fonction $g : \mathbb{R} \rightarrow \mathbb{R}$ 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, \dots, z_n , et t.q. $\{Z = z_i\} = \{Y = y_i\}$, $i = 1, \dots, n$, alors si h est telle que $\mathbb{E}(X|Z = z_i) = h(z_i)$, $i = 1, \dots, n$, on a :

$$g(y_i) = h(z_i), \quad i = 1; \dots, n, \text{ soit}$$

$$g(Y(\omega)) = h(Z(\omega)), \quad \forall \omega \in \Omega$$

$$\text{donc } E(X|Y) = E(X|Z)$$

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 Espérance conditionnelle 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

a) X 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 $X \in L^2(\Omega, \mathcal{F}, \mathbb{P})$.

On voudrait considérer $L^2(\Omega, \mathcal{G}, \mathbb{P})$ comme un sous espace vectoriel de $L^2(\Omega, \mathcal{F}, \mathbb{P})$. Ceci n'est pas tout à fait correct, car si \mathcal{G} ne contient pas tous les ensembles de \mathbb{P} mesure nulle de \mathcal{F} , les classes d'équivalence ne sont pas mes mêmes, donc les deux espaces ne contiennent pas les mêmes objets.

Notation : On notera $L^2(\mathcal{G}, \mathbb{P}_{\mathcal{F}})$ 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.

$L^2(\mathcal{G}, \mathbb{P}_{\mathcal{F}})$ 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 $L^2(\Omega, \mathcal{F}, \mathbb{P})$.

Définition 8.2.1. On appelle **espérance conditionnelle par rapport à \mathcal{G}** l'opérateur de projection orthogonale dans $L^2(\Omega, \mathcal{F}, \mathbb{P})$, sur $L^2(\mathcal{G}, \mathbb{P}_{\mathcal{F}})$.

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 un élément de $L^2(\mathcal{G}, \mathbb{P}_{\mathcal{F}})$, 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,

(*) $\mathbb{E}(YX) = \mathbb{E}(YE^{\mathcal{G}}(X)), \forall Y$ v.a.r. \mathcal{G} -mesurable et de carré intégrable.

En particulier, en choisissant $Y = 1$, on obtient :

$$\mathbb{E}(E^{\mathcal{G}}(X)) = \mathbb{E}(X).$$

Lemme 8.2.2. 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 :

$$0 \leq \mathbb{E} [\mathbf{1}_{\{\mathbb{E}^{\mathcal{G}}(X) < 0\}} X] = \mathbb{E} [\mathbf{1}_{\{\mathbb{E}^{\mathcal{G}}(X) < 0\}} \mathbb{E}^{\mathcal{G}}(X)] \leq 0$$

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'après la linéarité de la projection orthogonale,

$$\mathbb{E}^{\mathcal{G}}(X) = \mathbb{E}^{\mathcal{G}}(X^+) - \mathbb{E}^{\mathcal{G}}(X^-)$$

Le Lemme 8.2.2 entraîne que $\mathbb{E}^{\mathcal{G}}(X^+) \geq 0$, $\mathbb{E}^{\mathcal{G}}(X^-) \geq 0$ p.s. Donc :

$$|\mathbb{E}^{\mathcal{G}}(X)| \leq \mathbb{E}^{\mathcal{G}}(X^+) + \mathbb{E}^{\mathcal{G}}(X^-) = \mathbb{E}^{\mathcal{G}}(|X|) \text{ p.s.}$$

Donc

$$(**) \quad \mathbb{E}[|\mathbb{E}^{\mathcal{G}}(X)|] \leq \mathbb{E}(|X|).$$

b) X intégrable

Rappel d'analyse Soit H et B deux espaces de Banach, avec $H \subset B$ et H dense dans B . Si $A \in \mathcal{L}(H)$ vérifie : $\exists c > 0$ t.q.

$$(\clubsuit) \quad \|Ah\|_B \leq c\|h\|_B, \quad \forall h \in H,$$

alors A s'étend en un opérateur linéaire continu de B dans B , qui vérifie $(\clubsuit) \forall h \in B$. On applique ce résultat avec $H = L^2(\Omega, \mathcal{F}, \mathbb{P})$, $B = L^1(\Omega, \mathcal{F}, \mathbb{P})$, $A = E^{\mathcal{G}}(\cdot)$. L'inégalité (\clubsuit) est satisfaite grâce à (**). $\mathbb{E}^{\mathcal{G}}(\cdot)$ s'étend donc en un opérateur linéaire continu de $L^1(\Omega, \mathcal{F}, \mathbb{P})$ dans lui-même, qui vérifie (**).

En particulier, $\forall X$ v.a.r. intégrable, $\{X_n, n \in \mathbb{N}\}$ suite de v.a.r. de carré intégrable, t.q. $X_n \rightarrow X$ dans $L^1(\Omega, \mathcal{F}, P)$, alors $E^{\mathcal{G}}(X) = L^1 - \lim_n \mathbb{E}^{\mathcal{G}}(X_n)$.

On déduit alors aisément de (*) que si X est intégrable :

$$(*) \quad \mathbb{E}(YX) = \mathbb{E}(Y\mathbb{E}^{\mathcal{G}}(X)), \quad \forall Y \text{ v.a.r. } \mathcal{G}\text{-mesurable et bornée.}$$

c) Complément : $X \geq 0$ (non nécessairement intégrable)

Soit X une v.a.r. $X \geq 0$, p.s. Alors si $X_n \triangleq \inf(X, n)$, X_n est de carré intégrable $\forall n$, et $X_n \uparrow X$ p.s. Il résulte du Lemme 8.2.2 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 $\mathbb{E}^{\mathcal{G}}(X_n)$.

8.2.2 Propriétés de l'espérance conditionnelle

Théorème 8.2.3. *Si X est une v.a.r. intégrable, $\mathbb{E}^{\mathcal{G}}(X)$ est l'unique classe d'équivalence de v.a.r. qui vérifie :*

- (i) $\mathbb{E}^{\mathcal{G}}(X)$ est \mathcal{G} -mesurable et
- (ii) $\mathbb{E}(ZX) = \mathbb{E}(Z\mathbb{E}^{\mathcal{G}}(X))$, $\forall Z \mathcal{G}$ mesurable et bornée ou
- (ii)' $\int_G X dP = \int_G \mathbb{E}^{\mathcal{G}}(X) dP$, $\forall G \in \mathcal{G}$

PREUVE

- a) (ii) \Rightarrow (ii)' (trivial)
- b) (i) + (ii)' $\Rightarrow \mathbb{E}^{\mathcal{G}}(X)$ est intégrable. En effet, si l'on pose $G_1 = \{\mathbb{E}^{\mathcal{G}}(X) \geq 0\}$, $G_2 = \{\mathbb{E}^{\mathcal{G}}(X) < 0\}$,

$$\begin{aligned} \mathbb{E}(|\mathbb{E}^{\mathcal{G}}(X)|) &= \int_{G_1} \mathbb{E}^{\mathcal{G}}(X) dP - \int_{G_2} \mathbb{E}^{\mathcal{G}}(X) dP \\ &= \int_{G_1} X dP - \int_{G_2} X dP \\ &\leq \mathbb{E}(|X|) < \infty \end{aligned}$$

- c) (i) + (ii)' \Rightarrow (i) + (ii) résulte du Théorème des classes monotones 1.7.8, en utilisant le Théorème de convergence dominée de Lebesgue, puisque X et $\mathbb{E}^{\mathcal{G}}(X)$ sont intégrables (grâce à b).
- d) Par définition, $\mathbb{E}^{\mathcal{G}}(X)$ vérifie (i). Etant donné X intégrable, posons $X^n(\omega) = X(\omega)\mathbf{1}_{\{|X(\omega)| \leq n\}}$. X^n est de carré intégrable, donc par la Définition 8.2.1,

$$\mathbb{E}(ZX^n) = \mathbb{E}(Z\mathbb{E}^{\mathcal{G}}[X^n]), \quad \forall Z \mathcal{G}\text{-mesurable et bornée.}$$

(ii) s'obtient alors par passage à la limite dans $L^1(\Omega)$.

- e) L'unicité s'obtient comme suit. Si Y est \mathcal{G} -mesurable et intégrable,

$$\int_G Y dP = 0, \forall G \in \mathcal{G} \Rightarrow \left\{ \begin{array}{l} \int_{\{Y>0\}} Y dP = 0 \\ \int_{\{Y<0\}} Y dP = 0 \end{array} \right\} \Rightarrow Y = 0 \text{ p. s.}$$

Corollaire 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. qui vérifie :*

- a) $\mathbb{E}^{\mathcal{G}}[X]$ est \mathcal{G} -mesurable et intégrable.
- b) $\forall C \in \mathcal{C}, \int_C X d\mathbb{P} = \int_C \mathbb{E}^{\mathcal{G}}[X] d\mathbb{P}$, où \mathcal{C} est un π -système qui contient Ω , avec $\sigma(\mathcal{C}) = \mathcal{G}$.

PREUVE Il suffit de montrer que (a) + (b) \Rightarrow (ii)' du Théorème. Ceci résulte du Théorème $\pi - \lambda$ (1.3.9), en remarquant que :

$$\mathcal{L} = \left\{ B; \int_B X d\mathbb{P} = \int_B \mathbb{E}^{\mathcal{G}}[X] d\mathbb{P} \right\}$$

est un λ -système. □

Exemple 8.2.5. *Supposons que \mathcal{G} est engendrée par une partition finie (B_1, \dots, B_n) de Ω . Alors $\mathbb{E}^{\mathcal{G}}[X]$ est la v.a.r. constante sur chaque B_i t.q. :*

$$\mathbb{E}^{\mathcal{G}}[X](\omega_i) = \frac{1}{\mathbb{P}(B_i)} \int_{B_i} X(\omega) \mathbb{P}(d\omega), \quad \omega_i \in B_i.$$

En particulier, si $\mathcal{G} = \{\emptyset, \Omega\}$, $\mathbb{E}^{\mathcal{G}}[X] = \mathbb{E}[X]$.

Proposition 8.2.6. *Si X, X_1, X_2 sont des v.a.r. intégrables,*

- a) $\mathbb{E}^{\mathcal{G}}[\alpha X_1 + \beta X_2] = \alpha \mathbb{E}^{\mathcal{G}}[X_1] + \beta \mathbb{E}^{\mathcal{G}}[X_2]$
- 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 Y 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, et XY est intégrable,

$$\mathbb{E}^{\mathcal{G}}[YX] = X \mathbb{E}^{\mathcal{G}}[Y]$$

PREUVE a) découle de la définition.

b) la 1ère partie se démontre comme au Lemme 8.2.2. 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 8.2.4, 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) $\forall H \in \mathcal{H}, \int_H X d\mathbb{P} = \int_H \mathbb{E}^{\mathcal{G}}(X) d\mathbb{P} = \int_H \mathbb{E}^{\mathcal{H}}[\mathbb{E}^{\mathcal{G}}(X)] d\mathbb{P}$. La deuxième partie s'obtient alors en choisissant $\mathcal{H} = \{\emptyset, \Omega\}$.

e) Supposons tout d'abord X \mathcal{G} -mesurable et bornée. $\forall Z$ \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 $X_n \rightarrow X$ p.s. et $X_n Y \rightarrow XY$ dans $L^1(\Omega)$. \square

On a un Théorème de convergence dominée pour les espérances conditionnelles :

Proposition 8.2.7. *Si $X_n \rightarrow X$ p.s. et si $\exists Z$ intégrable t.q. $|X_n| \leq Z$ p.s., alors X_n et X sont intégrables et*

$$\mathbb{E}^{\mathcal{G}}[X_n] \rightarrow \mathbb{E}^{\mathcal{G}}[X] \text{ p.s. et dans } L^1(\Omega)$$

Montrons d'abord le :

Lemme 8.2.8. *Soit $\{X_n\}_{n \in \mathbb{N}}$ 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) $X_n \rightarrow X$ p.s.

(ii) $X_n \rightarrow X$ dans $L^1(\Omega)$.

De plus, si l'une des deux assertions est vraie, $\mathbb{E}^{\mathcal{G}}[X_n] \rightarrow \mathbb{E}^{\mathcal{G}}[X]$ 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 $\mathbb{E}(|X - X_n|) \rightarrow 0$. Réciproquement, $X_n \rightarrow X$ dans $L^1(\Omega)$ entraîne $X_n \leq X$ p.s. [resp. $X_n \geq X$ p.s.]. Donc $\exists Z$

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) $\Rightarrow \mathbb{E}^{\mathcal{G}}(X_n) \rightarrow \mathbb{E}^{\mathcal{G}}(X)$ 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 8.2.7

$$\mathbb{E}^{\mathcal{G}} \left(\inf_{m \geq n} X_m \right) \leq \inf_{m \geq n} \mathbb{E}^{\mathcal{G}}(X_m) \leq \sup_{m \geq n} \mathbb{E}^{\mathcal{G}}(X_m) \leq \mathbb{E}^{\mathcal{G}} \left(\sup_{m \geq n} X_m \right)$$

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

$$\mathbb{E}^{\mathcal{G}}(X_n) \rightarrow \mathbb{E}^{\mathcal{G}}(X) \text{ p.s.}$$

□

Proposition 8.2.9. *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 $X_n \rightarrow X$ dans $L^p(\Omega)$,*

$$\mathbb{E}^{\mathcal{G}}[X_n] \rightarrow \mathbb{E}^{\mathcal{G}}[X] \text{ dans } L^p(\Omega).$$

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 8.2.10. *Soit $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ une fonction convexe. Alors $\forall X$ v.a.r. intégrable, t.q. $\varphi(X)$ soit 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) : \exists 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

$$\begin{aligned} \varphi(X) &\geq a_n X + b_n \text{ p.s.}, \forall n \in \mathbb{N}, \\ \text{et } \mathbb{E}^{\mathcal{G}}[\varphi(X)] \text{ p.s.} &\geq \sup_n (a_n \mathbb{E}^{\mathcal{G}}(X) + b_n) = \varphi(\mathbb{E}^{\mathcal{G}}[X]) \end{aligned}$$

□

8.3 Espérance conditionnelle par rapport à une variable aléatoire

A tout v.a. X , on associe sa “tribu naturelle” $\sigma(X)$ -cf. la Définition 4.4.3- qui est la plus petite sous-tribu de \mathcal{F} qui rende X mesurable. On pose :

Définition 8.3.1. *Si Y est une v.a.r. intégrable, on définit l’espérance conditionnelle de Y , sachant X , par :*

$$\mathbb{E}(Y|X) \triangleq \mathbb{E}(Y|\sigma(X))$$

□

Dans cette définition, X peut être un vecteur aléatoire de dimension quelconque. Remarquons que si $\sigma(X) = \sigma(X')$

$$\mathbb{E}(Y|X) = \mathbb{E}(Y|X') \text{ p.s.}$$

Dans la pratique, on observe une réalisation x de X , et on voudrait définir $\mathbb{E}(Y|Y = x)$. Pour cela, on va utiliser la :

Proposition 8.3.2. *Soit X un v.a. de dimension d , et Z une v.a.r., tous deux définis sur (Ω, \mathcal{F}) . Si Z est $\sigma(X)$ -mesurable, alors \exists une application φ borélienne de \mathbb{R}^d dans \mathbb{R} , telle que :*

$$Z = \varphi(X) \text{ p.s.}$$

PREUVE Considérons la classe des v.a.r. sur (Ω, \mathcal{F}) : $\mathcal{C} = \{\varphi(X); \varphi \text{ mesurable } : \mathbb{R}^d \rightarrow \mathbb{R}\}$.

Il suffit de montrer que \mathcal{C} contient toutes les v.a.r. $\sigma(X)$ -mesurables. Pour cela, on va appliquer le Théorème des classes monotones 1.7.8 ; \mathcal{C} est un espace vectoriel, et $\forall A \in \sigma(X)$, $\exists B \in \mathcal{B}_d$ t.q. $A = \{X \in B\}$, donc \mathcal{C} contient $\mathbf{1}_A = \mathbf{1}_{\{X \in B\}}$.

Il nous reste à vérifier que si $Y_n \in \mathcal{C}$, $Y_n(\omega) \rightarrow Y(\omega)$, $\forall \omega \in \Omega$, alors $Y \in \mathcal{C}$.

Soit donc φ_n une suite d'applications mesurables de \mathbb{R}^d dans \mathbb{R} , t.q. $\varphi_n(X(\omega)) \rightarrow Y(\omega)$, $\forall \omega \in \Omega$. Posons $B \in \{x \in \mathbb{R}^d; \lim \varphi_n(x) \text{ existe}\}$, $B \in \mathcal{B}_d$, car

$$B = \bigcap_{k=1}^{\infty} \bigcup_{n=1}^{\infty} \bigcap_{i=1}^{\infty} \left\{ |\varphi_{n+i}(x) - \varphi_n(x)| < \frac{1}{k} \right\}.$$

On pose

$$\varphi(x) = \begin{cases} \lim \varphi_n(x) & \text{si } x \in B \\ 0 & \text{si } x \in B^c \end{cases}$$

Alors φ est mesurable, et $Y = \varphi(X)$. □

D'après la Proposition 8.3.2, \exists une application mesurable $\varphi : \mathbb{R}^d \rightarrow \mathbb{R}$, t.q. :

$$\mathbb{E}(Y|X) = \varphi(X) \text{ p.s.}$$

φ peut être considérée comme une v.a.r. définie sur $(\mathbb{R}^d, \mathcal{B}_d, \mathbb{P}_X)$. On vérifie qu'elle est unique à classe de \mathbb{P}_X -équivalence près. On pose alors, si Y est une v.a.r. intégrable :

Définition 8.3.3. $\mathbb{E}(Y|X = x)$ est l'unique classe d'équivalence $\varphi(x)$ de v.a.r. définies sur $(\mathbb{R}^d, \mathcal{B}_d, \mathbb{P}_X)$, t.q. $\mathbb{E}(Y|X)$ et $\varphi(X)$ soient p.s. égales. □

Remarque 8.3.4. Fixons x_0 t.q. $\mathbb{P}_X(\{x_0\}) = 0$. On ne peut pas définir $\mathbb{E}(Y|X = x_0)$, car cette quantité dépend du choix arbitraire du représentant dans la classe de fonctions $x \rightarrow \mathbb{E}(Y|X = x)$.

Proposition 8.3.5. Soit X un v.a. de dimension d , Y un v.a. de dimension k , f une application mesurable de \mathbb{R}^d dans \mathbb{R} , t.q.

$$\mathbb{E}(|f(X)Y|) < \infty \text{ et } \mathbb{E}(|Y|) < \infty.$$

$$\text{Alors } \mathbb{E}[f(X)Y|X = x] = f(x)\mathbb{E}[Y|X = x] \text{ p.s.}$$

Si de plus X et Y sont indépendants, et si g est une application mesurable de \mathbb{R}^{d+k} dans \mathbb{R} , t.q. $\mathbb{E}(|g(X, Y)|) < \infty$, alors

$$(*) \quad \mathbb{E}[g(X, Y)|X = x] = \mathbb{E}[g(x, Y)]$$

PREUVE Il résulte de 8.2.6 e) :

$$\mathbb{E}[f(X)Y|X] = f(X)\mathbb{E}[Y|X]$$

La première égalité découle alors de la Définition 8.3.3.

Il suffit de démontrer la seconde égalité pour g bornée. Soit d'abord $C \in \mathcal{B}_d$, $D \in \mathcal{B}_k$.

$$\mathbb{E}[\mathbf{1}_C(X)\mathbf{1}_D(Y)|X] = \mathbf{1}_C(X)\mathbb{E}[\mathbf{1}_D(Y)] \text{ p.s.}$$

$$\text{d'où } \mathbb{E}[\mathbf{1}_C(X)\mathbf{1}_D(Y)|X = x] = \mathbf{1}_C(x)\mathbb{E}[\mathbf{1}_D(Y)] \mathbb{P}_X \text{ p.s.}$$

Posons $\Lambda = \{C \times D, C \in \mathcal{B}_d, D \in \mathcal{B}_k\}$. Λ est un π -système qui contient \mathbb{R}^{d+k} . Si $\mathcal{C} = \{g \text{ mesurables bornées qui vérifient } (*)\}$, on vient de voir que \mathcal{C} contient $\mathbf{1}_A, \forall A \in \Lambda$. Or \mathcal{C} est un espace vectoriel stable par limite croissante, d'où le résultat par application du Théorème des classes monotones. \square

Il résulte de la Proposition que si X et Y sont indépendantes et $\varphi(X, Y)$ intégrable,

$$\mathbb{E}[\varphi(X, Y)|X] = \int_{\mathbb{R}^k} \varphi(X, y) d\mathbb{P}_Y(y)$$

On montre de plus, par la méthode utilisée dans la Proposition :

Exercice 8.3.6. Si X est \mathcal{G} -mesurable, Y et \mathcal{G} indépendantes, et $\varphi(X, Y)$ intégrable,

$$\mathbb{E}[\varphi(X, Y)|\mathcal{G}] = \mathbb{E}[\varphi(X, Y)|X]$$

\square

8.4 Probabilité conditionnelle

Définition 8.4.1. On appelle probabilité conditionnelle de l'événement A , sachant \mathcal{G} , la (classe d'équivalence de) v.a.r. :

$$\mathbb{P}(A|\mathcal{G}) \triangleq \mathbb{E}[\mathbf{1}_A|\mathcal{G}]$$

On appelle probabilité conditionnelle de A , sachant X :

$$\mathbb{P}(A|X) \triangleq \mathbb{E}[\mathbf{1}_A|X]$$

On appelle probabilité conditionnelle de A sachant que $X = x$:

$$\mathbb{P}(A|X = x) \triangleq \mathbb{E}[\mathbf{1}_A|X = x]$$

□

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 $\forall A \in \mathcal{F}$ un représentant de $\mathbb{P}(A|\mathcal{G})$, en faisant tous ces choix de façon "cohérente", pour que à ω fixé, $A \rightarrow \mathbb{P}(A|\mathcal{G})(\omega)$ soit une probabilité.

Soit $\begin{pmatrix} X \\ Y \end{pmatrix}$ un vecteur aléatoire à valeurs dans \mathbb{R}^{d+k} ; Posons la :

Définition 8.4.2. a) $\mathbb{P}_Y(\cdot|X)$ est appelée loi de probabilité conditionnelle régulière de Y sachant X si :

$$(i) \forall B \in \mathcal{B}_k, \mathbb{P}_Y(B|X) = \mathbb{P}(Y \in B|X) \text{ p.s.}$$

$$(ii) \forall \omega \in \Omega, B \rightarrow \mathbb{P}_Y(B|X)(\omega) \text{ est une mesure de probabilité sur } (\mathbb{R}^k, \mathcal{B}_k).$$

b) $\mathbb{P}_Y(\cdot|X = x)$ est appelée loi de probabilité conditionnelle régulière de Y , sachant que $X = x$ si :

$$(i) \forall B \in \mathcal{B}_k, \mathbb{P}_Y(B|X = x) = \mathbb{P}(Y \in B|X = x) \mathbb{P}_X \text{ p.s.}$$

$$(ii) \forall x \in \mathbb{R}^d, B \rightarrow \mathbb{P}_Y(B|X = x) \text{ est une mesure de probabilité sur } (\mathbb{R}^k, \mathcal{B}_k).$$

□

On démontre aisément à l'aide du Théorème des classes monotones la :

Proposition 8.4.3. 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 $\forall \varphi$ mesurable de \mathbb{R}^k dans \mathbb{R} , t.q.

$$\begin{aligned} \mathbb{E}(|\varphi(Y)|) &< \infty \\ \mathbb{E}[\varphi(Y)|X] &= \int_{\mathbb{R}^k} \varphi(y) \mathbb{P}_Y(dy|X) \text{ p.s.} \\ \mathbb{E}[\varphi(Y)|X = x] &= \int_{\mathbb{R}^k} \varphi(y) \mathbb{P}_Y(dy|X = x) \mathbb{P}_X \text{ p.s.} \end{aligned}$$

□

On démontre qu'une loi de probabilité conditionnelle régulière existe toujours. Nous allons le préciser dans deux cas :

1. celui où la loi du couple $\begin{pmatrix} X \\ Y \end{pmatrix}$ admet une densité,

2. celui où le couple est gaussien.

Proposition 8.4.4. 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) $\forall x \in Q, f(y|x) \triangleq \frac{f(x, y)}{g(x)}$ 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 trivial.

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, $\forall \omega, \bar{f}(y|X)(\omega)$ est la densité d'une probabilité sur $(\mathbb{R}^k, \mathcal{B}_k)$.

Il reste à vérifier que $\forall B \in \mathcal{B}_k$,

$$\int_B \bar{f}(y|X)dy = P(Y \in B|X) \text{ p.s.}$$

D'après 8.3.2 et 8.2.3, ceci est vrai si $\forall \varphi$ 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 8.4.5. 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 $\hat{\Sigma} = \text{Cov}(X - \hat{X})$. 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

$$\mathbb{E}(Z) = \mathbb{E}(X), \quad \text{Cov}(Z, Y) = \text{Cov}(X, Y),$$

où par définition $\text{Cov}(X, Y) = \mathbb{E}(XY') - \mathbb{E}(X)\mathbb{E}(Y)' = \Sigma_{12}$. 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

$$\text{Cov}(\tilde{X}, Y) = \text{Cov}(X, Y) - \text{Cov}(Z, Y) = 0.$$

On utilise ici la proposition 8.4.6 ci-dessous, qui est une variante de la Proposition 5.3.4. 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

$$\text{Cov}(\tilde{X}) = \Sigma_{12}\Sigma_{22}^{-1}\text{Cov}(Y)\Sigma_{22}^{-1}\Sigma_{21} = \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21},$$

et par l'indépendance de \hat{X} et \tilde{X} ,

$$\text{Cov}(X) = \text{Cov}(\hat{X}) + \text{Cov}(\tilde{X}).$$

Or si $f \in C_K(\mathbb{R}^d)$, on déduit de la Proposition 8.3.5

$$\begin{aligned} \mathbb{E}[f(X)|Y] &= \mathbb{E}[f(\hat{X} + \tilde{X})|Y] \\ &= \int_{\mathbb{R}^d} f(\hat{X} + x) \mathbb{P}_{\tilde{X}}(dx) \\ &= \int_{\mathbb{R}^d} f(x) N_{(\hat{X}, \hat{\Sigma})}(dx), \end{aligned}$$

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 $N(\hat{X}, \hat{\Sigma})$. \square

On a utilisé la

Proposition 8.4.6. *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 $\text{Cov}(U, V) = 0$.*

PREUVE La preuve est la même que celle de la Proposition 5.3.4, c'est à dire que l'on montre que la fonction caractéristique du v. a. $\begin{pmatrix} U \\ V \end{pmatrix}$ est le produit de celle de U par celle de V . \square

Ce résultat se généralise bien sûr au cas d'un vecteur aléatoire gaussien se décomposant en sous-vecteur U_1, \dots, U_k tels que $\text{Cov}(U_i, U_j) = 0$, pour tout $1 \leq i \neq j \leq k$.

Finissons ce chapitre par la

Remarque 8.4.7. *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 $\varphi(Y)$ caractérisée par le fait que*

$$\mathbb{E}[|X - \varphi(Y)|^2] = \min_{\psi} \mathbb{E}[|X - \psi(Y)|^2],$$

où le minimum est pris sur toutes les applications boréliennes $\psi : \mathbb{R}^k \rightarrow \mathbb{R}^d$, 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 $y \rightarrow Ay + b$ où A et b sont choisis tels que

$$\mathbb{E}(AY + b) = \mathbb{E}(X), \quad \text{Cov}(AY + b, Y) = \text{Cov}(X, Y).$$

C'est exactement comme cela que nous avons caractérisé Z au début de la preuve du Théorème 8.4.5. 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).