

# Intégration et Probabilités

Anne-Laure Dalibard

Année 2023-2024



# Table des matières

<b>I</b>	<b>Théorie de la Mesure</b>	<b>3</b>
<b>1</b>	<b>Espaces Mesurés</b>	<b>3</b>
1.1	Ensembles Mesurables . . . . .	3
1.2	Mesures Positives . . . . .	3
1.3	Fonctions Mesurables . . . . .	4
1.4	Classe Monotone . . . . .	4
<b>2</b>	<b>Intégration par rapport à une mesure</b>	<b>5</b>
2.1	Intégration Positive . . . . .	5
2.2	Fonctions Intégrables . . . . .	6
2.3	Intégrales dépendant d'un Paramètre . . . . .	7
<b>3</b>	<b>Mesures Produits</b>	<b>8</b>
3.1	Préliminaires . . . . .	8
3.2	Construction de la mesure-produit. . . . .	9
3.3	Théorème de Fubini . . . . .	9
3.4	Applications . . . . .	9
3.4.1	Intégration par Parties . . . . .	9
3.4.2	Convolution . . . . .	10
3.4.3	Calcul du Volume de la Boule Unité . . . . .	10
<b>4</b>	<b>Espaces <math>L^p</math></b>	<b>10</b>
4.1	Définition et Inégalités . . . . .	10
4.2	L'espace de Banach $L^p$ . . . . .	12
4.3	Théorie de Densité dans les Espaces $L^p$ . . . . .	12
4.4	Radon-Nikodym . . . . .	12
<b>II</b>	<b>Probabilités</b>	<b>12</b>
<b>5</b>	<b>Théorie des Probabilités</b>	<b>12</b>
5.1	Loi de Probabilité . . . . .	12
5.1.1	Lois de Probabilité . . . . .	13
5.1.2	Des Propriétés . . . . .	14
5.2	Moments de Variables Aléatoires . . . . .	14
<b>6</b>	<b>Indépendance</b>	<b>15</b>
6.1	Définitions . . . . .	15
6.2	Hahn-Kolmogorov . . . . .	16
6.3	Convergence de sommes de variables aléatoires indépendantes . . . . .	17
6.4	Semigroupes de Convolution . . . . .	18
<b>7</b>	<b>Convergence de Variables Aléatoires</b>	<b>18</b>
7.1	Loi Forte des Grands Nombres . . . . .	19
7.2	Convergence en Loi . . . . .	19
7.3	Conséquences . . . . .	20
7.3.1	Convergence des Mesures Empiriques . . . . .	20
7.3.2	Théorème Central Limite . . . . .	20
7.4	Intervalle de Confiance . . . . .	20

<b>8</b>	<b>Espérance Conditionnelle</b>	<b>22</b>
8.1	Variables Aléatoires Intégrables . . . . .	22
8.2	Cas des Variables Positives . . . . .	23

## Première partie

# Théorie de la Mesure

## 1 Espaces Mesurés

### 1.1 Ensembles Mesurables

**Définition 1.1.1.** Une tribu sur un ensemble  $E$  est un ensemble  $\mathcal{A} \subset \mathcal{P}(E)$  telle que :

- $E \in \mathcal{A}$
- $\mathcal{A}^c = \mathcal{A}$
- Si  $(A_i)_{i \in \mathbb{N}} \in \mathcal{A}$ ,  $\bigcup_{i \in \mathbb{N}} A_i \in \mathcal{A}$ .

Les éléments de  $\mathcal{A}$  sont appelés parties mesurables.

**Définition 1.1.2.** Soit  $\mathcal{C} \subset \mathcal{P}(E)$ . On appelle tribu engendrée par  $\mathcal{C}$  l'ensemble  $\sigma(\mathcal{C}) = \bigcap_{\mathcal{A} \text{ tribu}, \mathcal{C} \subset \mathcal{A}} \mathcal{A}$ . C'est la plus petite tribu contenant  $\mathcal{C}$ .

**Définition 1.1.3.** Soit  $E$  un espace topologie,  $\mathcal{O}$  la classe des ouverts. On appelle tribu borélienne sur  $E$  la tribu engendrée par  $\mathcal{O}$  notée  $\mathcal{B}(E)$ .

**Proposition 1.1.1.**  $\mathcal{B}(\mathbb{R})$  est engendrée aussi par les intervalles ouverts  $]a, b[$   $_{a, b \in \mathbb{R}}$ ,  $]a, +\infty[$   $_{a \in \mathbb{R}}$ ,  $]a, +\infty[$   $_{a \in \mathbb{Q}}$ .

**Définition 1.1.4.** La tribu produit sur  $(E_1, \mathcal{A}_1), (E_2, \mathcal{A}_2)$  est la tribu sur  $E_1 \times E_2$  définie  $\mathcal{A}_1 \otimes \mathcal{A}_2 = \sigma(\{A_1 \times A_2 \mid (A_1, A_2) \in \mathcal{A}_1 \times \mathcal{A}_2\})$

**Proposition 1.1.2.** On a  $\mathcal{B}(\mathbb{R}^2) = \mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R})$ .

### 1.2 Mesures Positives

**Définition 1.2.1.** Une mesure positive sur  $(E, \mathcal{A})$  est une application  $\mu : \mathcal{A} \rightarrow [0, +\infty]$  qui vérifie :

- $\mu(\emptyset) = 0$ .
- Pour toute famille  $(A_n)_{n \in \mathbb{N}}$  de parties mesurables disjointes :

$$\mu \left( \bigcup_{n \in \mathbb{N}} A_n \right) = \sum_{n \in \mathbb{N}} \mu(A_n)$$

**Proposition 1.2.1.** On a :

- Si  $(A_n)_{n \in \mathbb{N}} \in \mathcal{A}^{\mathbb{N}}$  est une suite croissante :

$$\mu \left( \bigcup_{n \in \mathbb{N}} A_n \right) = \lim_{n \rightarrow \infty} \uparrow \mu(A_n)$$

- Si  $(B_n)_{n \in \mathbb{N}} \in \mathcal{A}^{\mathbb{N}}$  est une suite décroissante et si  $\mu(B_0) < \infty$  :

$$\mu \left( \bigcap_{n \in \mathbb{N}} B_n \right) = \lim_{n \rightarrow \infty} \downarrow \mu(B_n)$$

- Si  $A_n \in \mathcal{A}$  :

$$\mu \left( \bigcup_{n \in \mathbb{N}} A_n \right) \leq \sum_{n \in \mathbb{N}} \mu(A_n)$$

**Définition 1.2.2.** Il existe une unique mesure positive sur  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$  telle que  $\lambda([a, b]) = b - a$ , pour tous  $a, b \in \mathbb{R}$ .

**Définition 1.2.3.** •  $\mu$  est finie si  $\mu(E) < \infty$

- $\mu$  est une mesure de probabilité si  $\mu(E) = 1$
- $\mu$  est  $\sigma$ -finie s'il existe une suite croissante de parties mesurables  $E_n$  d'union  $E$  et de mesure toujours finie.
- $x \in E$  est un atome de  $\mu$  si  $\mu(\{x\}) > 0$
- $\mu$  est dite diffuse si elle n'a pas d'atomes.

### 1.3 Fonctions Mesurables

**Définition 1.3.1.** Une application  $f : (E, \mathcal{A}) \rightarrow (F, \mathcal{B})$  est dite mesurable si  $\forall B \in \mathcal{B}, f^{-1}(B) \in \mathcal{A}$ .

**Théorème 1.3.1.** La composition de deux applications mesurables est mesurable.

**Remarque 1.3.1.1** (Composition Mesurable). Il faut bien que les applications  $f$  et  $g$  partagent un espace, avec la même tribu (comme la chanson). On définit fréquemment deux tribus différentes sur  $\mathbb{R}^d$  : la tribu borélienne et la tribu de Lebesgue, tribu complétée de la tribu borélienne pour la mesure de Lebesgue  $\mathcal{M}(\lambda) = \{A \subset \mathbb{R}^d, \exists B_1, B_2 \in \mathcal{B}(\mathbb{R}^d), B_1 \subset A \subset B_2 \text{ et } \lambda(B_2 \setminus B_1) = 0\}$  et on a  $\mathcal{B}(\mathbb{R}^d) \subsetneq \mathcal{M}(\lambda)$ . Dans certains livres :  $f$  est mesurable si  $f : (\mathbb{R}, \mathcal{M}(\lambda)) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$  est mesurable.

**Proposition 1.3.1.** Pour que  $f$  soit mesurable, il suffit qu'il existe une sous-classe engendrant  $\mathcal{B}$  pour laquelle la propriété est vraie.

**Corollaire 1.3.1.1.** Si  $f : \mathbb{R}^{d_1} \rightarrow \mathbb{R}^{d_2}$  est continue, elle est mesurable pour les boréliens.

**Corollaire 1.3.1.2.** Une application produit est mesurable.

*Démonstration.* On a :  $A_1 \otimes A_2 = \sigma(A_1 \times A_2)$  ■

**Lemme 1.3.2.** Les applications  $(+)(\times)(\max)(\min)$  de deux fonctions réelles sont mesurables

**Corollaire 1.3.2.1.** Les parties positives et négatives d'une fonction sont mesurables

**Proposition 1.3.2.** Si les  $f_n$  sont mesurables de  $E$  dans  $\overline{\mathbb{R}}$  alors :  $\sup_n f_n, \inf_n f_n, \liminf f_n, \limsup f_n$  sont mesurables. En particulier :  $\lim_n f_n$  est mesurable si la suite CS.

*Démonstration.* 1. Si  $f(x) = \inf f_n(x) : f^{-1}[-\infty, a[ = \bigcup_n \{x \mid f_n(x) < a\}$ . De même pour sup.

On en déduit immédiatement  $\liminf f_n = \sup_{n \geq 0} \inf_{k \geq n} f_k$ .

2. On a :  $\{x \in E \mid \lim f_n(x) \text{ existe}\} = \{x \in E \mid \liminf f_n(x) = \limsup f_n(x)\} = \mathcal{G}^{-1}(\Delta)$  où  $\mathcal{G} = (\liminf f_n, \limsup f_n)$  et  $\Delta$  est la diagonale de  $\overline{\mathbb{R}}^2$ . ■

**Définition 1.3.2** (Mesure-Image). On appelle mesure image de  $\mu$  par  $f$ , notée  $f_{\#}\mu$  la mesure  $f_{\#}\mu(B) = \mu(f^{-1}(B))$

### 1.4 Classe Monotone

**Définition 1.4.1** (Classe Monotone).  $\mathcal{M} \in \mathcal{P}(E)$  est une classe monotone si :

1.  $E \in \mathcal{M}$
2. Si  $A, B \in \mathcal{M}$  avec  $A \subset B$ ,  $B \setminus A \in \mathcal{M}$
3. Si  $(A_n) \in \mathcal{M}^{\mathbb{N}}$  croissante,  $\bigcup_{n \in \mathbb{N}} A_n \in \mathcal{M}$

**Remarque 1.4.0.1.** Toute tribu est une classe monotone

**Lemme 1.4.1.** Si  $\mathcal{M}$  est une classe monotone stable par intersections finies, c'est une tribu.

**Définition 1.4.2.** Si  $\mathcal{C} \subset \mathcal{P}(E) : \mathcal{M}(\mathcal{C}) = \bigcap_{\mathcal{M} \text{ classe monotone, } \mathcal{C} \subset \mathcal{M}}$

**Théorème 1.4.2** (Lemme de Classe Monotone). Si  $\mathcal{C} \subset \mathcal{P}(E)$  est stable par intersections finies :  $\mathcal{M}(\mathcal{C}) = \sigma(\mathcal{C})$

**Remarque 1.4.2.1.** Les classes monotones sont des outils plus maniables que les tribus et se marient mieux avec les propriétés des mesures. Le théorème fait le lien entre tribus et classes monotones, ce qui facilite la vie avec les mesures.

*Démonstration.* Point Méthodologique : ne pas essayer d'exprimer des éléments de  $\mathcal{C}$ . T'façon les preuves constructives, c'est pour les salopes. ■

**Remarque 1.4.2.2.** On peut en déduire l'unicité de la mesure de Lebesgue. C'est une conséquence du théorème suivant.

**Théorème 1.4.3.** Soit  $\mathcal{C}$  stable par intersections telle que  $\sigma\mathcal{C} = \mathcal{A}$ . On suppose  $\mu_1(A) = \mu_2(A), \forall A \in \mathcal{C}$  Alors :

1.  $\mu$ p.p., l'application  $u \mapsto f(u, x)$  est continue en  $u_0$
2. Si  $\mu_1(E) = \mu_2(E) < +\infty$  alors  $\mu_1 = \mu_2$
3. S'il existe  $(E_n) \in \mathcal{A}^{\mathbb{N}}$  croissante d'union  $E$  et de mesures égales et finies par  $\mu_1$  et  $\mu_2$  alors  $\mu_1 = \mu_2$

*Démonstration.* 1. Cas fini :  $\mathcal{M} = \{A \in \mathcal{A} \mid \mu_1(A) = \mu_2(A)\}$  est une classe monotone. Donc  $\mathcal{M} = \mathcal{A}$  par Lemme de Classe Monotone

2. Cas Infini : On applique le cas fini à  $E_n$  en prenant la restriction. Par continuité croissante, on obtient bien le résultat. ■

## 2 Intégration par rapport à une mesure

### 2.1 Intégration Positive

**Définition 2.1.1.**  $f$  mesurable à valeurs réelles est étagée si elle prend un nombre fini de valeurs.

**Remarque 2.1.0.1.** 1. Les Fonctions en escalier sur un intervalle sont étagées  
2. L'indicatrice d'un ensemble est étagée, en particulier :  $1_{\mathbb{Q}}$  est étagée.

**Définition 2.1.2.** Si  $f$  est étagée, et prend les valeurs :  $\alpha_1 < \dots < \alpha_n$ , l'écriture canonique de  $f$  est, avec  $A_i = f^{-1}(\{\alpha_i\}) : f = \sum_i^n \alpha_i 1_{A_i}$

Pour  $f$  mesurable, on pose alors :  $\int f d\mu = \sup_{h \in \mathcal{E}, h \leq f} \int h d\mu$

**Proposition 2.1.1.** L'intégrale est une forme linéaire monotone i.e. l'intégrale d'une fonction positive est positive. Ceci s'étend aux fonctions intégrables.

**Théorème 2.1.1** (De Convergence Monotone). Si  $f_n$  est croissante positive et tend vers  $f$  :

$$\int f d\mu = \lim_{n \rightarrow \infty} \int f_n d\mu$$

*Démonstration.* Par croissance :

$$\int f d\mu \geq \lim_{n \rightarrow \infty} \int f_n d\mu.$$

Il suffit donc d'établir l'autre inégalité : soit  $h = \sum \alpha_i 1_{A_i}$  étagée positive inférieure à  $f$ . Soit  $a \in [0, 1[$ . On pose :

$$E_n = \{x \in E \mid ah(x) \leq f_n(x)\}$$

Les  $E_n$  sont mesurables et, puisque  $a < 1$  et  $f = \lim \uparrow f_n$ ,  $E = \bigcup \uparrow E_n$ .  
Comme  $f_n \geq a \mathbb{1}_{E_n} h$ ,

$$\int f_n d\mu \geq \int \mathbb{1}_{E_n} h = a \sum \alpha_i \mu(A_i \cap E_n)$$

En passant à la limite croissante, comme  $A_i \cap E_n \uparrow A_i$ , puis en faisant tendre  $a$  vers 1 on trouve le résultat. ■

**Théorème 2.1.2.** *Soit  $f$  mesurable positive. Il existe une suite croissante de fonctions étagées positives de limite  $f$ .*

**Corollaire 2.1.2.1.** *Si les  $f_n$  sont positives :*

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

**Théorème 2.1.3** (Lemme de Fatou). *Si les  $f_n$  sont mesurables positives :*

$$\int \liminf f_n d\mu \leq \liminf \int f_n d\mu$$

*Démonstration.* On a :

$$\liminf f_n = \lim_{k \rightarrow \infty} \left( \inf_{n \geq k} f_n \right)$$

Donc, par théorème de convergence monotone 2.1.1 :

$$\int \liminf f_n d\mu = \lim_{k \rightarrow \infty} \int \left( \inf_{n \geq k} f_n \right) d\mu$$

Par ailleurs, si  $p \geq k$ , on a :  $\inf_{n \geq k} f_n \leq f_p$ , d'où :

$$\int \left( \inf_{n \geq k} f_n \right) d\mu \leq \inf_{p \geq k} \int f_p d\mu$$

En passant à la limite croissante quand  $k \rightarrow \infty$ , on a :

$$\lim_{k \rightarrow \infty} \int \left( \inf_{n \geq k} f_n \right) d\mu \leq \lim_{k \rightarrow \infty} \inf_{p \geq k} \int f_p d\mu = \liminf \int f_n d\mu$$

ce qui conclut. ■

**Proposition 2.1.2.** *Soit  $f$  mesurable positive :*

- $\forall a > 0, \mu(\{x \in E \mid f(x) \geq a\}) \leq \frac{1}{a} \int f d\mu$  [Inégalité de Markov]
- $\int f d\mu < \infty \Rightarrow f < \infty p.p.$
- $\int f d\mu \Leftrightarrow f = 0 p.p.$
- $f = gp.p. \Rightarrow \int f d\mu = \int g d\mu$

*On peut généraliser cette dernière proposition aux fonctions intégrables.*

## 2.2 Fonctions Intégrables

**Définition 2.2.1.** *On dit qu'une fonction est intégrable si l'intégrale de sa norme est finie. En ce cas, l'intégrale de la fonction est la somme des intégrales de ses fonctions composantes.*

**Théorème 2.2.1.** *L'intégrale vérifie l'inégalité triangulaire : Soit  $f$  intégrable*

$$\int |f| d\mu \geq \left| \int f d\mu \right|$$

*Démonstration.* Ecrire  $|z|^2 = z\bar{z}$  pour  $z$  l'intégrale de  $f$ . En notant  $\bar{a}$  le conjugué de  $z$ , on a le résultat. ■

**Théorème 2.2.2** (De Convergence Dominée). Soit  $f_n$  une suite de fonctions mesurables à valeurs dans  $\mathbb{R}$  ou  $\mathbb{C}$ . On suppose :

1. Il existe  $f$  mesurable telle que  $f_n(x) \rightarrow f(x)$   $\mu$ p.p.
2. Il existe  $g$  telle que  $|f_n(x)| \leq g(x)$   $\mu$ p.p.

Alors :

$$\lim_{n \rightarrow \infty} \int |f_n(x) - f(x)| \mu(dx) = 0$$

et

$$\lim_{n \rightarrow \infty} \int f_n d\mu = \int f d\mu$$

*Démonstration.* 1. On suppose dans un premier temps que les hypothèses (1) et (2) sont vérifiées sur tout l'ensemble. On remarque  $|f| \leq g$  donc  $f$  est intégrable.

Ensuite, puisque par inégalité triangulaire,  $|f - f_n| \leq 2g$  et  $|f - f_n| \rightarrow 0$ , par lemme de Fatou :

$$\liminf \int (2g - |f - f_n|) d\mu \geq \int \liminf (2g - |f - f_n|) d\mu = 2 \int g d\mu.$$

Par linéarité, comme  $\liminf -u_n = -\limsup u_n$  :

$$2 \int g d\mu - \limsup \int |f - f_n| d\mu \geq 2 \int g d\mu$$

D'où,  $\int |f - f_n| d\mu \rightarrow 0$ . Par inégalité triangulaire :

$$\left| \int f d\mu - \int f_n d\mu \right| \leq \int |f - f_n| d\mu$$

2. Dans le cas général, on suppose cette fois ci (1) et (2). On pose alors :

$$A = \{x \in E \mid f_n(x) \rightarrow f(x) \text{ et pour tout } n \mid |f_n(x)| \leq g(x)\}.$$

Par hypothèses,  $\mu(A^c) = 0$  et par la première partie de la preuve appliquée à :

$$\tilde{f}_n(x) = \mathbb{1}_A(x) f_n(x), \tilde{f}(x) = \mathbb{1}_A(x) f(x)$$

Comme  $f = \tilde{f}$  p.p. et  $f_n = \tilde{f}_n$  p.p., on a bien le résultat par la première partie de la preuve. ■

## 2.3 Intégrales dépendant d'un Paramètre

Principe : Utiliser le TCD pour montrer des propriétés de régularité.

**Remarque 2.3.0.1** (Exemples d'utilisation). • Pour  $f$  intégrable à variables dans  $\mathbb{R}^d$ , on définit la transformée de Fourier  $\hat{f}$  par :

$$\hat{f}(\xi) = \int_{\mathbb{R}^d} \exp(-\xi \cdot x) f(x) dx$$

- Soit  $f$  intégrable à variables dans  $\mathbb{R}^d$ ,  $g$  continue bornée à variables dans  $\mathbb{R}^d$ . On définit la convolée de  $f$  et  $g$  par :

$$f \star g : x \mapsto \int_{\mathbb{R}^d} f(y) g(x - y) dy$$

**Théorème 2.3.1** (De Continuité sous l'intégrale). Soit  $U$  un ouvert de  $\mathbb{R}^d$ ,  $u_0 \in U$ . Soit  $f : U \times E \rightarrow \mathbb{R}$  vérifiant :

1.  $\forall u \in U$ , l'application  $x \in E \mapsto f(u, x)$  est mesurable.
- 2.

3.  $\mu p.p.$ , l'application  $u \mapsto f(u, x)$  est continue en  $u_0$
4. Il existe une application  $g : E \rightarrow [0, +\infty[$ , intégrable telle

$$\forall u \in U, \mu p.p., |f(u, x)| \leq g(x)$$

Alors,  $F(u) = \int f(u, x) \mu(dx)$  est bien définie et est continue en  $u_0$ .

*Démonstration.* Par l'hypothèse (iii.),  $F$  est bien définie.

Soit  $(u_n)$  une suite de limite  $u_0$ . Par (ii.),  $f(u_n, x) \rightarrow_{n \rightarrow \infty} f(u_0, x), \mu p.p.$  Par (iii.), on peut appliquer le théorème de convergence dominée 2.2.2, ce qui donne le résultat par caractérisation séquentielle de la limite. ■

**Corollaire 2.3.1.1.** *Les fonctions définies en exemple sont continues.*

**Théorème 2.3.2** (De Dérivation sous l'intégrale). *Soit  $U$  un ouvert de  $\mathbb{R}$ ,  $u_0 \in U$ . Soit  $f : U \times E \rightarrow \mathbb{R}$  vérifiant :*

1.  $\forall u \in U, x \mapsto f(u, x)$  est mesurable
2.  $\mu p.p.$ ,  $u \mapsto f(u, x)$  est dérivable en  $u_0$
3. il existe  $g$  positive intégrable, telle que  $\mu p.p.$  :

$$\forall u \in U, |f(u, x) - f(u_0, x)| \leq g(x) |u - u_0|$$

Alors :

$$F(u) = \int f(u, x) \mu(dx) \text{ est dérivable et } F'(u_0) = \int \frac{\partial f}{\partial u}(u_0, x) \mu(dx)$$

*Démonstration.* On applique le TCD 2.2.2 à  $\varphi_n(x) = \frac{f(u_n, x) - f(u_0, x)}{u_n - u_0}$  où  $u_n \rightarrow u_0$ . ■

**Corollaire 2.3.2.1.** *En choisissant pour la transformée de fourier des fonctions dont les premiers moments sont intégrables, on gagne en régularité. En particulier, si une fonction est à support compact, sa transformée de fourier est indéfiniment dérivable.*

### 3 Mesures Produits

On se donne  $(E_1, \mathcal{A}_1, \mu_1)$  et  $(E_2, \mathcal{A}_2, \mu_2)$ , et on veut :

1. Définir une mesure produit sur  $(E_1 \times E_2, \mathcal{A}_1 \otimes \mathcal{A}_2)$
2. Démontrer les théorèmes de Fubini sur la mesure ainsi définie.

#### 3.1 Préliminaires

**Définition 3.1.1.** *Soit  $C \in \mathcal{A}_1 \otimes \mathcal{A}_2, x_1 \in E_1, x_2 \in E_2$ . On note :*

$$C_{x_1} = \{y \in E_2 \mid (x_1, y) \in C\}$$

$$C_{x_2} = \{y \in E_1 \mid (y, x_2) \in C\}$$

*On définit par ailleurs, si  $f : (E_1 \times E_2, \mathcal{A}_1 \otimes \mathcal{A}_2) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ , les applications partielles  $f_{x_1}$  et  $f^{x_2}$ .*

**Lemme 3.1.1.** •  $\forall C \in \mathcal{A}_1 \otimes \mathcal{A}_2, x_1 \in E_1, x_2 \in E_2, C_{x_1} \in \mathcal{A}_2$  et  $C^{x_2} \in \mathcal{A}_1$

- $f_{x_1}$  et  $f^{x_2}$  sont mesurables.

*Démonstration.* • On introduit la classe des  $C_{x_1}$  pour  $x_1 \in E_1$ . Cette classe contient les pavés et est une tribu, i.e.  $\mathcal{C} = \mathcal{A} \otimes \mathcal{B}$ .

- Pour toute partie mesurable  $D$  de  $\mathbb{R}$ ,

$$f_{x_1}^{-1}(D) = \{x_2 \in E_2 \mid (x_1, x_2) \in f^{-1}(D)\} = (f^{-1}(D))_{x_1}$$

■



### 3.2 Construction de la mesure-produit.

**Théorème 3.2.1.** 1. Il existe une unique mesure  $m$  sur  $(E_1 \times E_2, \mathcal{A}_1 \otimes \mathcal{A}_2)$  telle que :

$$\forall A \in \mathcal{A}_1, \forall A_2 \in \mathcal{A}_2, m(A_1 \times A_2) = \mu_1(A_1) \mu_2(A_2)$$

Cette mesure est  $\sigma$ -finie et notée  $m = \mu_1 \otimes \mu_2$

2. Pour tout  $C \in \mathcal{A}_1 \otimes \mathcal{A}_2$  :

$$\mu_1 \otimes \mu_2(C) = \int_E \mu_2(C_{x_1}) \mu_1(dx_1) = \int_F \mu_1(C^{x_2}) \mu_2(dx_2)$$

*Démonstration.* • Unicité : Lemme de Classe Monotone sur la classe des pavés mesurables.

- Existence : On pose  $m(C)$  comme dans le second point et on vérifie le résultat en supposant d'abord  $\mu_2$  finie puis  $\sigma$ -finie. ■

### 3.3 Théorème de Fubini

**Théorème 3.3.1** (Fubini-Tonelli). On suppose  $\mu_1, \mu_2$   $\sigma$ -finies. Soit  $f : E_1 \times E_2 \rightarrow [0, \infty]$  une fonction mesurable :

- Les fonctions :

$$\begin{aligned} x &\mapsto \int f(x, y) \mu_2(dy) \\ y &\mapsto \int f(x, y) \mu_1(dx) \end{aligned}$$

sont  $E_1$  et  $E_2$  mesurables respectivement.

- On a :

$$\int_{E_1 \times E_2} f d(\mu_1 \otimes \mu_2) = \int_{E_1} \left( \int_{E_2} f(x, y) \mu_2(dy) \right) \mu_1(dx) = \int_{E_2} \left( \int_{E_1} f(x, y) \mu_1(dx) \right) \mu_2(dy)$$

Autrement dit, pour une fonction positive, ça marche.

**Théorème 3.3.2** (Fubini-Lebesgue). Soit  $f \in \mathcal{L}^1(E_1 \times E_2, \mathcal{A}_1 \otimes \mathcal{A}_2, \mu_1 \otimes \mu_2)$ . Alors :

1.  $\mu_1(dx_1)$  p.p.,  $y \mapsto f(x, y)$  est dans  $\mathcal{L}^1(E_2, \mathcal{A}_2, \mu_2)$ ,  
 $\mu_2(dx_2)$  p.p.,  $x \mapsto f(x, y)$  est dans  $\mathcal{L}^1(E_1, \mathcal{A}_1, \mu_1)$
2. Les fonctions  $x \mapsto \int f(x, y) \mu_2(dy)$  et  $y \mapsto \int f(x, y) \mu_1(dx)$ , bien définies sauf sur un ensemble mesurable de mesure nulle sont respectivement dans  $\mathcal{L}^1(E_1, \mathcal{A}_1, \mu_1)$  et  $\mathcal{L}^1(E_2, \mathcal{A}_2, \mu_2)$ .
3. On a :

$$\int_{E_1 \times E_2} f d(\mu_1 \otimes \mu_2) = \int_{E_1} \left( \int_{E_2} f(x_1, x_2) \mu_2(dx_2) \right) \mu_1(dx_1) = \int_{E_2} \left( \int_{E_1} f(x_1, x_2) \mu_1(dx_1) \right) \mu_2(dx_2)$$

Autrement dit, pour une fonction intégrable (même à valeurs complexes), ça se passe bien.

### 3.4 Applications

#### 3.4.1 Intégration par Parties

**Définition 3.4.1.** Pour  $f, g$  deux fonctions mesurables de  $\mathbb{R}$  dans  $\mathbb{R}$  localement intégrables (i.e. intégrables sur tout compact pour la mesure de Lebesgue), on pose pour  $x \in \mathbb{R}$  :

$$\begin{aligned} F(x) &= \int_0^x f(t) dt \\ G(x) &= \int_0^x g(t) dt \end{aligned}$$

**Théorème 3.4.1** (IPP). Pour tous  $a < b$ , on a :

$$F(b)G(b) = F(a)G(a) + \int_a^b f(t)G(t)dt + \int_a^b F(t)g(t)dt$$

### 3.4.2 Convolution

**Définition 3.4.2.** Si  $f$  et  $g$  sont deux fonctions mesurables sur  $\mathbb{R}^d$ , la convolution :

$$f \star g(x) = \int_{\mathbb{R}^d} f(x-y)g(y)dy$$

est bien définie si :

$$\int_{\mathbb{R}^d} |f(x-y)g(y)| dy < \infty$$

Alors, on a  $g \star f(x) = f \star g(x)$

**Proposition 3.4.1.** Soient  $f, g \in \mathcal{L}^1(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \lambda)$ . Alors, pour  $\lambda$  presque tout  $x \in \mathbb{R}^d$ , la convolution  $f \star g(x)$  est bien définie. De plus  $f \star g \in \mathcal{L}^1(\lambda)$  et  $\|f \star g\|_1 \leq \|f\|_1 \|g\|_1$ .

#### 3.4.2.1 Approximations de Dirac

**Définition 3.4.3.** On dit qu'une suite  $\varphi_n \in C_c(\mathbb{R}^d)$  est une approximation de  $\delta_0$  si :

- Il existe un compact  $K$  tel que  $\text{supp}(\varphi_n) \subset K$  pour tout  $n$ .
- Pour tout  $n$ ,  $\varphi_n \geq 0$  et

$$\int_{\mathbb{R}^d} \varphi_n(x) dx = 1$$

- Pour tout  $\delta > 0$ ,

$$\lim_{n \rightarrow \infty} \int_{\{|x| > \delta\}} \varphi_n(x) dx = 0$$

**Proposition 3.4.2.** En prenant  $\varphi : \mathbb{R}^d \rightarrow \mathbb{R}_+$  est continue à support compact de mesure 1, avec  $\varphi_n(x) = n^d \varphi(nx)$ , on a construit des approximations de Dirac.

En prenant par exemple  $\varphi(x) = c \exp\left(-\frac{1}{1-|x|^2}\right) \mathbf{1}_{\{|x| < 1\}}$ , et en choisissant  $c > 0$  correctement, on a même une approximation  $C^\infty$ .

**Proposition 3.4.3.** 1. Si  $f : \mathbb{R}^d \rightarrow \mathbb{R}$  est continue, on a  $\varphi_n \star f \rightarrow f$  quand  $n \rightarrow \infty$ , uniformément sur tout compact.

2. Si  $f \in \mathcal{L}^p(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \lambda)$ , avec  $p \in [1, \infty[$ , on a  $\varphi_n \star f \rightarrow f$  dans  $\mathcal{L}^p$ .

### 3.4.3 Calcul du Volume de la Boule Unité

On note ici  $B_d$  la boule unité fermée de  $\mathbb{R}^d$  et  $\lambda_d$  la mesure de Lebesgue sur  $\mathbb{R}^d$ . En vue de calculer  $\gamma_d = \lambda_d(B_d)$ , on observe que l'image de  $\lambda_d$  par  $x \mapsto ax$  est  $a^{-d} \lambda_d$  pour tout  $a > 0$ . Par théorème de Fubini, si  $d \geq 2$ , on montre que  $\gamma_d = \gamma_{d-1} I_{d-1}$  en posant pour  $n \geq 0$ ,  $I_n = \int_{-1}^1 (1-x^2)^{n/2} dx$ . Or, par IPP,  $I_n = \frac{n}{n+1} I_{n-2}$ . D'où  $\gamma_d = I_{d-1} I_{d-2} \gamma_{d-2} = \frac{2\pi}{d} \gamma_{d-2}$  et donc  $\gamma_{2k} = \frac{\pi^k}{k!}$  et  $\gamma_{2k+1} = \frac{\pi^{k+1/2}}{(k+\frac{1}{2})(k-\frac{1}{2}) \dots \frac{3}{2} \cdot \frac{1}{2}}$ . On peut regrouper ces résultats sous :

$$\gamma_d = \frac{\pi^{d/2}}{\Gamma\left(\frac{d}{2} + 1\right)}$$

## 4 Espaces $L^p$

### 4.1 Définition et Inégalités

**Définition 4.1.1.** Soit  $f$  et  $g$  des fonctions mesurables à valeurs réelles, on dit que  $f \sim g$  si et seulement si  $f = g$ ,  $\mu$  p.p. C'est une relation d'équivalence.

**Définition 4.1.2.** Soit alors :  $p \in [1, +\infty)$ . On note :

$$\mathcal{L}^p(E, \mathcal{A}, \mu) = \left\{ f : E \rightarrow \mathbb{R}, \text{ mesurable } \mid \int_E |f|^p d\mu < \infty \right\}$$

On définit alors :  $L^p(E, \mathcal{A}, \mu) = \mathcal{L}^p(E, \mathcal{A}, \mu) / \sim$ .

On définit de plus :

$$\mathcal{L}^\infty(E, \mathcal{A}, \mu) = \{ f : E \rightarrow \mathbb{R}, \text{ mesurable } \mid \exists C > 0, |f| \leq C, \mu \text{ p.p.} \}$$

On définit alors :  $L^\infty(E, \mathcal{A}, \mu) = \mathcal{L}^\infty(E, \mathcal{A}, \mu) / \sim$ .

On fera systématiquement l'abus d'utiliser un système de représentants.

**Définition 4.1.3.** On note, pour  $f : E \rightarrow \mathbb{R}$  mesurable, avec  $\infty^{\frac{1}{p}}$ , si  $p \in [1, \infty[$  :

$$\|f\|_p = \left( \int |f|^p d\mu \right)^{\frac{1}{p}}$$

et :

$$\|f\|_\infty = \inf \{ C \in [0, \infty] \mid |f| \leq C, \mu \text{ p.p.} \}$$

**Proposition 4.1.1.** • Si  $f \in L^\infty$ , alors  $|f| \leq \|f\|_\infty, \mu \text{ p.p.}$

• Si  $f \in L^p, p \in [1, \infty], \|f\|_p = 0 \Leftrightarrow f = 0$ .

• Les  $\|\cdot\|_p$  sont des normes.

**Définition 4.1.4.** Soient  $p, q \in [1, \infty]^2$ , on dit que  $p$  et  $q$  sont conjugués si

$$\frac{1}{p} + \frac{1}{q} = 1$$

En particulier,  $p = 1$  et  $q = \infty$  tout comme 2 et 2 sont conjugués

**Lemme 4.1.1** (Inégalité d'Young). Soit  $p, q \in ]1, \infty[$  conjugués. Pour tous  $a, b \in \mathbb{R}$  :

$$|ab| \leq \frac{1}{p} |a|^p + \frac{1}{q} |b|^q$$

**Proposition 4.1.2** (Inégalité de Hölder). Soient  $p, q$  conjugués. Soit  $f \in L^p, g \in L^q$ . Alors,  $fg$  est intégrable et :

$$\int |fg| d\mu \leq \|f\|_p \|g\|_q$$

**Remarque 4.1.1.1.** Cas d'égalité :

- Si  $p = 1, q = \infty$ , et que  $f = 0$  ou  $g = \|g\|_\infty$  presque partout.
- Si  $p, q < \infty$  : Il y a égalité si  $f$  ou  $g$  est nulle. Sinon si et seulement si il y a égalité dans l'inégalité de Young  $\mu$  presque partout.

**Remarque 4.1.1.2.** Si  $g \in L^q$  est fixée : L'application

$$f \in L^p \mapsto \int_E fg d\mu$$

est une forme linéaire continue.

**Remarque 4.1.1.3.** Dans le cas  $E = \mathbb{N}, \mathcal{A} = \mathcal{P}(\mathbb{N})$ , on note  $\uparrow^p(\mathbb{N})$  l'ensemble des suites dont la série des puissances  $p$ -ième est sommable.

**Remarque 4.1.1.4.** Dans le cas où  $\mu(E) < \infty$ , en prenant  $g = 1$  dans l'inégalité de Hölder :

$$\int |f| d\mu \leq \mu(E)^{1-\frac{1}{p}} \left( \int |f|^p d\mu \right)^{\frac{1}{p}} < +\infty$$

**Proposition 4.1.3** (Inégalité de Jensen). Soit  $\varphi : \mathbb{R} \rightarrow \mathbb{R}_+$  convexe. Pour  $f \in L^1$  :

$$\int \varphi \circ f d\mu \geq \varphi \left( \int f d\mu \right)$$

## 4.2 L'espace de Banach $L^p$

**Proposition 4.2.1** (Inégalité de Minkowski). Soit  $p \in [1, \infty]$  :

$$\|f + g\|_p \leq \|f\|_p + \|g\|_p$$

**Théorème 4.2.1** (Riesz/de Complétude des  $L^p$ ). L'espace  $L^p$  est un Banach.

**Proposition 4.2.2.** Soit  $p \in [1, \infty[$  et soit  $(f_n)$  qui converge vers  $f$  dans  $L^p$ . Il existe alors une sous-suite  $(f_{k_n})$  qui converge  $\mu$  p.p. vers  $f$ .

**Proposition 4.2.3.** L'espace  $L^2$  muni du produit scalaire :

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

est un Hilbert réel.

## 4.3 Théorie de Densité dans les Espaces $L^p$

**Définition 4.3.1.** Si  $E$  est un espace métrique, une mesure  $\mu$  sur  $(E, \mathcal{B}(E))$  est dite extérieurement régulière si

$$\forall A \in \mathcal{B}(E), \mu(A) = \inf \{ \mu(U) \mid U \text{ ouvert}, A \subseteq U \}$$

**Théorème 4.3.1.** L'espace des fonctions étagées appartenant à  $L^p$  est dense dans  $L^p$ .

**Théorème 4.3.2.** Les fonctions  $C^\infty(\mathbb{R}^d)$  à support compact sont denses dans  $L^p(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \lambda_d)$ .

**Lemme 4.3.3.** Les fonctions Lipschitziennes qui sont dans  $L^p$  sont denses dans  $L^p$ .

## 4.4 Radon-Nikodym

**Définition 4.4.1.** Soient  $\mu$  et  $\nu$  deux mesures sur  $(E, \mathcal{A})$ . On dit que :

1.  $\nu$  est absolument continue par rapport à  $\mu$ , noté  $\nu \ll \mu$  si :

$$\forall A \in \mathcal{A}, \mu(A) = 0 \Rightarrow \nu(A) = 0$$

2.  $\nu$  est étrangère à  $\mu$ , noté  $\nu \perp \mu$  s'il existe  $N \in \mathcal{A}$  tel que  $\mu(N) = 0$  et  $\nu(N^c) = 0$

**Théorème 4.4.1** (Radon-Nikodym). Soit  $\mu, \nu$  deux mesures  $\sigma$ -finies sur  $(E, \mathcal{A})$ . Il existe alors un unique couple  $(\nu_a, \nu_s)$  de mesures  $\sigma$ -finies sur  $(E, \mathcal{A})$  telle que :

- $\nu = \nu_a + \nu_s$
- $\nu_a \ll \mu$  et  $\nu_s \perp \mu$ .

De plus il existe une fonction mesurable  $h : E \rightarrow \mathbb{R}_*^+$  telle que :

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

## Deuxième partie

# Probabilités

## 5 Théorie des Probabilités

### 5.1 Loi de Probabilité

**Définition 5.1.1.** Un espace probabilisé est un espace mesurable  $(\Omega, \mathcal{A})$  muni d'une mesure  $P$  de masse totale 1. Pour  $A \in \mathcal{A}$ ,  $P(A)$  représente la probabilité d'occurrence de l'évènement  $A$ .

**Définition 5.1.2.** Une application mesurable  $X : \Omega \rightarrow E$  est appelée variable aléatoire à valeurs dans  $E$ .

La loi de la variable aléatoire  $X$  est la mesure image de  $P$  par  $X$  notée

$$\mathbb{P}(X \in B) = \mathbb{P}_X(B) = \mathbb{P}(X^{-1}(B))$$

**Définition 5.1.3.** • *Variables Aléatoires Discrètes* : C'est le cas où  $E$  est dénombrable et  $\mathcal{E} = \mathcal{P}(E)$ . La loi de  $X$  est alors

$$P_X = \sum_{x \in E} P(X = x) \delta_x$$

- *Variables Aléatoires à Densité* : Une variable aléatoire  $X$  à valeurs dans  $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$  est dite à densité si  $P_X$  est absolument continue par rapport à la mesure de Lebesgue  $\lambda$ . Il existe alors  $p : \mathbb{R}^d \rightarrow \mathbb{R}_+$  telle que

$$P_X(B) = \int_B p(x) dx$$

La fonction  $p$ , unique à un ensemble de mesure nulle près, est appelée la densité de  $X$ .

**Définition 5.1.4.** Soit  $X$  une variable aléatoire réelle. On note

$$E[X] = \int_{\Omega} X(\omega) P(d\omega)$$

qui est bien définie si  $X \geq 0$  et alors  $E[X] \in [0, \infty]$  ou si  $E[|X|] < \infty$ . On étend cette définition dans  $\mathbb{R}^d$  avec  $E[(X_1, \dots, X_d)] = (E[X_1], \dots, E[X_d])$ .

**Proposition 5.1.1.** Si  $X = \mathbb{1}_B$ ,  $E[X] = \mathbb{P}(B)$

**Proposition 5.1.2.** Soit  $X$  v.a. à valeurs dans  $E$ . Si  $f : E \rightarrow [0, \infty]$  est mesurable :

$$E[f(X)] = \int_E f(x) \mathbb{P}_X(dx)$$

*Démonstration.* Découle des propositions sur la composition de fonctions mesurables. ■

**Proposition 5.1.3.** Soit  $X = (X_1, \dots, X_d)$  une v.a. à valeurs dans  $\mathbb{R}^d$ . Supposons que la loi de  $X$  a densité  $p(x_1, \dots, x_d)$ . Alors, pour tout  $j$ , la loi de  $X_j$  a une densité donnée par

$$p_j(x) = \int_{\mathbb{R}^{d-1}} p(x_1, \dots, x_{j-1}, x, x_{j+1}, \dots, x_d) dx_1 \dots dx_{j-1} dx_{j+1} \dots dx_d$$

La loi de  $X$  détermine entièrement les lois marginales de  $X$ , mais la réciproque est fausse.

*Démonstration.* Direct. ■

### 5.1.1 Lois de Probabilité

**Définition 5.1.5.** • *Lois Discrètes* :

- *Loi Uniforme* : Si  $E$  est un ensemble fini,  $|E| = n$ , une v.a.  $X$  est de loi uniforme sur  $E$ ,  $X \sim \mathcal{U}(E)$ , si

$$\mathbb{P}(X = x) = \frac{1}{n}$$

- *Loi de Bernoulli de paramètre  $p \in [0, 1]$*  :  $X$  à valeurs dans  $\{0, 1\}$  suit une telle loi,  $X \sim \mathcal{B}(p)$  si  $\mathbb{P}(X = 1) = p$ ,  $\mathbb{P}(X = 0) = 1 - p$ .
- *Loi Binômiale  $\mathcal{B}(n, p)$ ,  $n \in \mathbb{N}^*$ ,  $p \in [0, 1]$*  :  $\mathbb{P}(X = k) = C_n^k (1 - p)^{n-k} p^k$ .
- *Loi Géométrique de paramètre  $p \in ]0, 1[$*  :  $X \sim \mathcal{G}(p)$  si  $\mathbb{P}(X = k) = (1 - p) p^k$
- *Loi de poisson de paramètre  $\lambda > 0$*  :  $X \sim \mathcal{P}(\lambda)$  si  $\mathbb{P}(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}$ .

- *Lois Continues/à Densité* :

- *Loi Uniforme sur  $[a, b]$*  :  $p(x) = \frac{1}{b-a} \mathbb{1}_{[a,b]}(x)$
- *Loi Exponentielle de paramètre  $\lambda > 0$*  :  $p(x) = \lambda e^{-\lambda x} \mathbb{1}_{\mathbb{R}_+}(x)$
- *Loi Gaussienne ou Normale  $\mathcal{N}(m, \sigma^2)$ ,  $m \in \mathbb{R}$ ,  $\sigma > 0$*  :  $p(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right)$

### 5.1.2 Des Propriétés

**Définition 5.1.6.** La fonction de répartition d'une variable aléatoire réelle  $X$  est la fonction  $F_X : \mathbb{R} \rightarrow [0, 1]$  définie par :

$$F_X(t) = \mathbb{P}(X \leq t) = \mathbb{P}_X([-\infty, t])$$

La fonction  $F_X$  est croissante, continue à droite, et a pour limites 0 en  $-\infty$  et 1 en  $+\infty$ .

**Proposition 5.1.4.** La fonction de répartition caractérise la variable aléatoire et les sauts de  $F_X$  correspondent aux atomes de  $\mathbb{P}_X$ . De plus :

$$\mathbb{P}(a \leq X \leq b) = F_X(b) - F_X(a^-)$$

$$\mathbb{P}(a < X < b) = F_X(b^-) - F_X(a)$$

*Démonstration.* Découle des résultats d'intégration. ■

**Définition 5.1.7.** On définit la tribu engendrée par une variable aléatoire dans un espace mesurable comme la plus petite tribu sur  $\Omega$  qui rende  $X$  mesurable :

$$\sigma(X) = \{A = X^{-1}(B) \mid B \in \mathcal{E}\}$$

**Proposition 5.1.5.** Soit  $X$  une variable aléatoire à valeurs dans  $E, \mathcal{E}$ . Soit  $Y$  une v.a. réelle. Il y a équivalence entre :

1.  $Y$  est  $\sigma(X)$ -mesurable.
2. Il existe une fonction mesurable de  $E, \mathcal{E}$  dans  $\mathbb{R}, \mathcal{B}(\mathbb{R})$  telle que  $Y = f(X)$ .

*Démonstration.* ii) implique i) est directe. Si  $Y$  est  $\sigma(X)$ -mesurable. On traite le cas où  $Y$  est étagée, puis on conclut car  $Y$  est limite simple de v.a.  $Y_n$  étagées et  $\sigma(X)$ -mesurables. ■

## 5.2 Moments de Variables Aléatoires

**Définition 5.2.1.** Le moment d'ordre  $p$  de  $X$  est  $E[X^p]$ . On appelle moment absolu d'ordre  $p$  de  $X$  la quantité  $E[|X|^p]$ .

**Proposition 5.2.1.** • Par Convergence Monotone :  $X_n \geq 0, X_n \uparrow X \Rightarrow E[X_n] \uparrow E[X]$

- Lemme de Fatou :  $X_n \geq 0 \Rightarrow E[\liminf X_n] \leq \liminf E[X_n]$
- Convergence Dominée :  $|X_n| \leq Z, E[Z] < \infty, X_n \rightarrow X \text{ p.p.} \Rightarrow E[X_n] \rightarrow E[X]$ .

**Proposition 5.2.2.** Inégalité de Cauchy-Schwarz :  $E[|XY|] \leq \sqrt{E[X^2]E[Y^2]}$ . Dans le cas où  $Y = 1$  on a :  $E[|X|^2] \leq E[X^2]$

Les espaces  $\mathcal{L}^p(\Omega, \mathcal{A}, \mathbb{P})$  sont définis pour tout  $p \in [1, \infty]$  par le quotient pour la relation "P presque partout" des fonctions dont le moment d'ordre  $p$  est défini.

**Définition 5.2.2.** Soit  $X \in \mathcal{L}^2$ , la variance de  $X$  est :

$$\text{var}(X) = E[(X - E[X])^2]$$

et l'écart-type de  $X$  est :

$$\sigma_X = \sqrt{\text{var}(X)}$$

**Proposition 5.2.3.** On a :  $\text{var}(X) = E[X^2] - (E[X])^2$  et si  $a \in \mathbb{R}$  :

$$E[(X - a)^2] = \text{var}(X) + (E[X] - a)^2$$

Ainsi :

$$\text{var}(X) = \inf_{a \in \mathbb{R}} E[(X - a)^2]$$

*Démonstration.* Calculer. ■

**Proposition 5.2.4.** Si  $X \in \mathcal{L}^2$  et  $a > 0$  :

- Inégalité de Markov :  $\mathbb{P}(X \geq a) \leq \frac{1}{a} E[X]$
- Inégalité de Bienaymé-Tchébychev :  $\mathbb{P}(|X - E[X]| \geq a) \leq \frac{1}{a^2} \text{var}(X)$

*Démonstration.* Utiliser l'Inégalité de Markov pour  $(X - E[X])^2$ . ■

**Définition 5.2.3.** Si  $X, Y \in \mathcal{L}^2$ , la covariance de  $X$  et  $Y$  est :

$$\text{cov}(X, Y) = E[(X - E[X])(Y - E[Y])] = E[XY] - E[X]E[Y]$$

On définit pour un vecteur aléatoire à valeurs dans  $\mathbb{R}^d$  :

$$K_X = (\text{cov}(X_i, X_j))_{i,j \in \llbracket 1, d \rrbracket}$$

$X, Y \mapsto \text{cov}(X, Y)$  est bilinéaire et  $K_X$  est symétrique positive.

**Définition 5.2.4.** Si  $X$  est une v.a. à valeurs dans  $\mathbb{R}^d$  la fonction caractéristique de  $X$  est  $\Phi_X : \mathbb{R}^d \rightarrow \mathbb{C}$  définie par :

$$\Phi_X(\xi) = E[\exp(i\xi \cdot X)] = \int e^{i\xi \cdot x} \mathbb{P}_X(dx)$$

**Lemme 5.2.1.** Soit  $X \sim \mathcal{N}(0, \sigma^2)$  :

$$\Phi_X(\xi) = \exp\left(-\frac{\sigma^2 \xi^2}{2}\right)$$

*Démonstration.* On calcule, en dérivant sous le signe intégrale. ■

**Théorème 5.2.2.** La fonction caractéristique caractérise la loi.

*Démonstration.* On traite le cas  $d = 1$  puis on applique le théorème de Fubini-Lebesgue. ■

**Définition 5.2.5.** On définit la fonction génératrice  $g_X$  d'une variable aléatoire à valeurs dans  $\mathbb{N}$  sur l'intervalle  $[0, 1]$  par :

$$g_X(r) = E[r^X] = \sum_{n=0}^{\infty} \mathbb{P}(X = n) r^n$$

**Proposition 5.2.5.** •  $g_X$  est continue sur  $[0, 1]$

- $g_X(0) = \mathbb{P}(X = 0), g_X(1) = 1$ .
- La fonction caractéristique caractérise la loi puisque les  $\mathbb{P}(X = n)$  sont les coefficients (uniques) du développement de Taylor de  $g_X$  en 0.
- $\lim_{r \uparrow 1} g_X^{(p)}(r) = E[X(X-1) \cdots (X-p+1)]$

## 6 Indépendance

### 6.1 Définitions

**Définition 6.1.1.** Deux événements sont indépendants si  $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$ .  $n$  événements sont indépendants si pour tout sous-ensemble non vide de  $\{1, \dots, n\}$  on a :

$$\mathbb{P}\left(\bigcap_{j \in J \subset \llbracket 1, n \rrbracket} A_j\right) = \prod_{j \in J} \mathbb{P}(A_j)$$

**Proposition 6.1.1.** Les  $n$  événements  $A_1, \dots, A_n$  sont indépendants si et seulement si, quand  $B_i \in \sigma(A_i) = \{\emptyset, A_i, A_i^c, \Omega\}$  :

$$\mathbb{P}\left(\bigcap_{j \in \llbracket 1, n \rrbracket} B_j\right) = \prod_{j \in \llbracket 1, n \rrbracket} \mathbb{P}(B_j)$$

**Définition 6.1.2.** • Soient  $\mathcal{B}_1, \dots, \mathcal{B}_n$   $n$  sous-tribus de  $\mathcal{A}$ . Ces sous-tribus sont indépendantes si et seulement si toute familles d'évènements de ces tribus l'est.

- Soient  $X_1, \dots, X_n$   $n$  v.a. à valeurs dans  $E_1, \dots, E_n$ . Ces variables sont indépendantes si les tribus qu'elles engendrent le sont, i.e. si

$$\forall F_1 \in \mathcal{E}_1, \dots, \forall F_n \in \mathcal{E}_n, \mathbb{P} \left( \bigcap_{i \in \llbracket 1, n \rrbracket} \{X_i \in F_i\} \right) = \prod_{i \in \llbracket 1, n \rrbracket} \mathbb{P}(X_i \in F_i)$$

**Théorème 6.1.1.** Les  $n$  variables aléatoires sont indépendantes si et seulement si la loi du  $n$ -uplet  $(X_1, \dots, X_n)$  est le produit des lois de ses composantes :

$$\mathbb{P}_{(X_1, \dots, X_n)} = \mathbb{P}_{X_1} \otimes \dots \otimes \mathbb{P}_{X_n}$$

On a alors, si les  $f_i$  sont mesurables positives sur  $(E_i, \mathcal{E}_i)$  :

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

**Corollaire 6.1.1.1.** Si  $X_1, X_2 \in \mathcal{L}^2$  sont indépendantes :  $\text{cov}(X_1, X_2) = 0$ .

**Corollaire 6.1.1.2.** Soient  $X_1, X_n$   $n$  v.a. réelles.

1. Si  $P_{X_i}$  est de densité  $p_i$  et que les variables sont indépendantes. Alors :

$$p(x_1, \dots, x_n) = \prod_{i \in \llbracket 1, n \rrbracket} p_i(x_i)$$

2. Inversement, si la loi de  $(X_1, \dots, X_n)$  a une densité de la forme

$$p(x_1, \dots, x_n) = \prod_{i \in \llbracket 1, n \rrbracket} q_i(x_i)$$

où les  $q_i$  sont boréliennes positives sur  $\mathbb{R}$ . Alors les variables sont indépendantes et leurs densités sont proportionnelles à  $q_i$ .

**Proposition 6.1.2.** Soient  $\mathcal{B}_1, \dots, \mathcal{B}_n$  des sous-tribus de  $\mathcal{A}$ . Pour tout  $i \in \llbracket 1, n \rrbracket$ , soit  $\mathcal{C}_i \subset \mathcal{B}_i$  une classe stable par intersections finies, contenant  $\Omega$  et engendrant  $\mathcal{B}_i$ . Si

$$\forall C_1, \dots, C_n \in \mathcal{C}_1 \times \dots \times \mathcal{C}_n, P \left( \bigcap_{i \in \llbracket 1, n \rrbracket} C_i \right) = \prod_{i \in \llbracket 1, n \rrbracket} P(C_i)$$

alors les tribus  $\mathcal{B}_i$  sont indépendantes.

**Corollaire 6.1.1.3.** On en déduit le lemme de regroupement par paquets. Si on regroupe des tribus en partition, les tribus que chacune des classes engendrent sont indépendantes, et de même pour des variables aléatoires.

## 6.2 Hahn-Kolmogorov

**Théorème 6.2.1** (de Hahn-Kolmogorov). Soit  $\Omega$  un ensemble,  $\mathcal{A}$  une algèbre sur  $\Omega$ . Soit  $m : \mathcal{A} \rightarrow \mathbb{R}$  telle que :

- $m(\emptyset) = 0$  et  $m(\Omega) = 1$
- Si  $A_n$  est une suite d'éléments disjoints de  $\mathcal{A}$  telle que  $\cup_n A_n \in \mathcal{A}$ , alors  $m(\cup_n A_n) = \sum_n m(A_n)$ .

Alors, il existe une tribu  $\mathcal{T}$  contenant  $\mathcal{A}$  et une mesure de probabilité  $\mathbb{P}$  telle que restreinte à  $\mathcal{A}$  ce soit  $m$ .



**Remarque 6.2.1.1.** On appliquera ce résultat avec  $\Omega = \prod E_n$ ,  $\mathcal{A} = B_\infty$ .

**Lemme 6.2.2.** On définit pour  $A \in P(\Omega)$ ,

$$m^*(A) = \inf_{A \subset \bigcup_n A_n} \sum_{n \in \mathbb{N}} m(A_n)$$

Alors  $m^*$  est une mesure extérieure :

- $m^*(\emptyset) = 0$
- Si  $A \subset B$ ,  $m^*(A) \leq m^*(B)$ .
- $\forall A_n \in P(\Omega)^\mathbb{N}$

$$m^*\left(\bigcup_n A_n\right) \leq \sum_{n \in \mathbb{N}} m^*(A_n)$$

De plus  $m^*$  est égale à  $m$  sur  $\mathcal{A}$ .

### 6.3 Convergence de sommes de variables aléatoires indépendantes

**Définition 6.3.1.** Soit  $\mu$  et  $\nu$  deux lois de probabilités sur  $\mathbb{R}^d$ . On note  $\mu \star \nu$  la loi de probabilité définie par :

$$(\mu \star \nu)(A) = \int_{\mathbb{R}^d \times \mathbb{R}^d} \mathbb{1}_{(x+y) \in A} \mu(dx) \nu(dy)$$

Autrement dit, pour  $f \geq 0$  mesurable :

$$\int f d(\mu \star \nu) = \int_{\mathbb{R}^d \times \mathbb{R}^d} f(x+y) \mu(dx) \nu(dy)$$

**Proposition 6.3.1.** Soient  $X$  et  $Y$  deux v.a. indépendantes à valeurs dans  $\mathbb{R}^d$  :

1. La loi de  $X+Y$  est  $P_X \star P_Y$ . En particulier, si  $X$  et  $Y$  sont à densité :  $X+Y$  a pour densité  $p_X \star p_Y$
2. La fonction caractéristique de  $X+Y$  est :  $\Phi_{X+Y}(\xi) = \Phi_X(\xi) \Phi_Y(\xi)$
3. Si  $X$  et  $Y$  sont de carré intégrable :  $K_{X+Y} = K_X + K_Y$ .

**Définition 6.3.2.** Soit  $X_n$  une suite de v.a. à valeurs dans  $\mathbb{R}^d$ ,  $X$  une v.a. à valeurs dans  $\mathbb{R}^d$ . On dit que  $X_n$  converge en probabilité vers  $X$  si :

$$\forall \varepsilon > 0, \lim_{n \rightarrow \infty} \mathbb{P}(|X_n - X| \geq \varepsilon) = 0$$

**Définition 6.3.3.** Pour  $X, Y$  dans  $L^0_{\mathbb{R}^d}$ , on pose :  $d(X, Y) = E[\text{Inf}(|X - Y|, 1)]$ . C'est une distance.

**Proposition 6.3.2.** Si  $(X_n)$  est une suite de  $L^0$ , alors  $X_n \rightarrow X$  si et seulement si  $d(X_n, X) \xrightarrow[n \rightarrow \infty]{} 0$ . De plus,  $L^0$  est complet.

**Corollaire 6.3.0.1.** On a de plus que si  $d(X_n, X) \rightarrow 0$ , il existe une sous-suite qui converge p.s.

**Corollaire 6.3.0.2.** Soit  $X_n$  une suite de v.a. qui converge en proba vers  $X$ . S'il existe  $r > 1$  tel que :

$$\sup_{n \in \mathbb{N}} E[|X_n|^r] < \infty \text{ et } E[|X|^r] < \infty$$

Il y a convergence dans  $L^1$ .

**Définition 6.3.4.** Une suite  $X_n$  de variable aléatoires à valeurs dans  $\mathbb{R}^d$  converge en loi vers une variable aléatoire  $X$  à valeurs dans  $\mathbb{R}^d$  si

$$\mathbb{P}_{X_n} \xrightarrow[n \rightarrow \infty]{(e)} \mathbb{P}_X$$

Cela équivaut à :

$$\forall \varphi \in \mathcal{C}_b(\mathbb{R}^d, \mathbb{R}), E[\varphi(X_n)] \xrightarrow[n \rightarrow \infty]{} E[\varphi(X)]$$

**Proposition 6.3.3.** *Si la suite  $(X_n)$  converge en probabilité vers  $X$  alors la suite  $X_n$  converge en loi vers  $X$ .*

**Théorème 6.3.1** (Loi Faible des Grands Nombres). *Soit  $X_n$  une suite de variables aléatoires réelles indépendantes et de même loi. Si  $E[X_1^2] < \infty$  :*

$$\frac{1}{n} (X_1 + \dots + X_n) \xrightarrow[n \rightarrow \infty]{L^2} E[X_1]$$

**Proposition 6.3.4.** *Sous les hypothèses du théorème précédent, si  $E[X_1^4] < \infty$ , on a presque sûrement*

$$\frac{1}{n} (X_1 + \dots + X_n) \xrightarrow[n \rightarrow \infty]{} E[X_1]$$

**Corollaire 6.3.1.1.** *Si  $(A_n)_{n \geq 1}$  est une suite d'événements indépendants de même probabilité, on a :*

$$\frac{1}{n} \sum_{i=1}^n \mathbb{1}_{A_i} \xrightarrow[n \rightarrow \infty]{p.s.} \mathbb{P}(A_1)$$

## 6.4 Semigroupes de Convolution

Soit  $I \in \{\mathbb{N}, \mathbb{R}_+\}$

**Définition 6.4.1.** *Soit  $(\mu_t)_{t \in I}$  une famille de mesures de probabilités sur  $\mathbb{R}^d$ . On dit que  $(\mu_t)$  est un semigroupe de convolution si  $\mu_0 = \delta_0$  et si :*

$$\mu_t \star \mu_{t'} = \mu_{t+t'}, \forall t, t' \in I$$

**Lemme 6.4.1.** *Pour que  $\mu_t$  soit un semigroupe de convolution, il suffit qu'il existe  $\varphi : \mathbb{R} \rightarrow \mathbb{C}$  telle que :*

- si  $I = \mathbb{N}$ ,  $\hat{\mu}_t(\xi) = \varphi(\xi)^t$ , pour tout  $t \in \mathbb{N}$
- si  $I = \mathbb{R}$ ,  $\hat{\mu}_t(\xi) = \exp(-t\varphi(\xi))$ , pour tout  $t$

**Proposition 6.4.1.** *Si  $X, Y$  sont deux v.a. réelles indépendantes et :*

- Si  $X \sim \mathcal{P}(\lambda), Y \sim \mathcal{P}(\lambda')$  alors  $X + Y \sim \mathcal{P}(\lambda + \lambda')$
- Si  $X \sim \mathcal{N}(m, \sigma^2), Y \sim \mathcal{N}(m', \sigma'^2)$  alors  $X + Y \sim \mathcal{N}(m + m', \sigma^2 + \sigma'^2)$

## 7 Convergence de Variables Aléatoires

**Définition 7.0.1.** *On dit que la suite  $(X_n)$  converge en probabilité vers  $X$  noté*

$$X_n \xrightarrow[n \rightarrow \infty]{(\mathbb{P})} X$$

si pour tout  $\varepsilon > 0$

$$\lim_{n \rightarrow \infty} \mathbb{P}(|X_n - X| > \varepsilon) = 0$$

**Proposition 7.0.1.** *Soit  $L_{\mathbb{R}^d}^0(\Omega, \mathcal{A}, P)$  le quotient de l'espace de toutes les variables aléatoires par la relation  $X = Y$  presque sûrement. Alors la formule :*

$$d(X, Y) = E[|X - Y| \wedge 1]$$

définit une distance compatible avec la convergence en probabilité.

**Proposition 7.0.2.** *Soit  $X_n$  une suite convergeant en probabilité vers  $X$ . Supposons qu'il existe  $r \in ]1, \infty[$  tel que la suite  $(X_n)$  soit bornée dans  $L^r$ . Alors, pour  $p \in [1, r[$*

## 7.1 Loi Forte des Grands Nombres

**Théorème 7.1.1** (Loi du Tout ou Rien). *Soit  $X_n$  une suite de v.a. indépendantes, à valeurs dans des espaces mesurables quelconques. Pour  $n \geq 1$  soit  $\mathfrak{B}_n$  la tribu  $\sigma(X_k, k \geq n)$ . Alors la tribu asymptotique*

$$\mathfrak{B}_\infty = \bigcap_{n=0}^{\infty} \mathfrak{B}_n$$

*est grossière, au sens où  $P(B) \in \{0, 1\}$  pour tout  $B \in \mathfrak{B}_\infty$*

**Proposition 7.1.1.** *Soit  $X_n$  une suite de variables aléatoires i.i.d de Rademacher. On note  $S_n = X_1 + \dots + X_n$ . Alors, presque sûrement :*

$$\sup S_n = +\infty \text{ et } \inf S_n = -\infty$$

*En particulier, il existe presque sûrement des entiers arbitrairement grands pour lesquels  $S_n = 0$ .*

**Théorème 7.1.2** (Loi Forte des Grands Nombres). *Soit  $X_n$  une suite de v.a. i.i.d. dans  $L^1$ . Alors,*

$$\frac{1}{n} (X_1 + \dots + X_n) \xrightarrow[n \rightarrow \infty]{p.s.} E[X_1]$$

## 7.2 Convergence en Loi

**Définition 7.2.1.** *Une suite  $\mu_n$  de mesures de probabilité sur  $\mathbb{R}^d$  converge étroitement vers une mesure de probabilité  $\mu$  sur  $\mathbb{R}^d$  si*

$$\forall \varphi \in \mathcal{C}_b(\mathbb{R}^d, \mathbb{R}), \int \varphi d\mu_n \xrightarrow[n \rightarrow \infty]{} \int \varphi d\mu$$

**Théorème 7.2.1** (du Porte Manteau). *Soient  $\mu_n, \mu$  des mesures de probabilité sur  $\mathbb{R}^d$ . Les quatre assertions suivantes sont équivalentes :*

$$1. \mu_n \xrightarrow{(e)} \mu$$

2. *Pour tout ouvert  $G$  de  $\mathbb{R}^d$*

$$\liminf \mu_n(G) \geq \mu(G)$$

3. *Pour tout fermé  $F$  de  $\mathbb{R}^d$*

$$\limsup \mu_n(F) \leq \mu(F)$$

4. *Pour tout  $B \in \mathcal{B}(\mathbb{R}^d)$  tel que  $\mu(\partial B) = 0$  :*

$$\lim \mu_n(B) = \mu(B)$$

**Proposition 7.2.1.** *Une suite de v.a. réelles converge en loi vers une v.a. si et seulement si les fonctions de répartition convergent en tout point de continuité.*

On note  $\mathcal{C}_c(\mathbb{R}^d)$  l'espace des fonctions continues à support compact.

**Proposition 7.2.2.** *Soient  $\mu_n, \mu$  des mesures de probabilité sur  $\mathbb{R}^d$ . Soit  $H$  un sous-ensemble de  $\mathcal{C}_b(\mathbb{R}^d)$  dont l'adhérence pour la norme infinie contient  $\mathcal{C}_c(\mathbb{R}^d)$ . Alors les propositions suivantes sont équivalentes :*

$$1. \mu_n \xrightarrow{(e)} \mu$$

2. *On a :*

$$\forall \varphi \in \mathcal{C}_c(\mathbb{R}^d), \int \varphi d\mu_n \xrightarrow[n \rightarrow \infty]{} \int \varphi d\mu$$

3. *On a :*

$$\forall \varphi \in H, \int \varphi d\mu_n \xrightarrow[n \rightarrow \infty]{} \int \varphi d\mu$$

**Théorème 7.2.2** (Lévy). Une suite  $\mu_n$  de mesures de probabilité sur  $\mathbb{R}^d$  converge étroitement vers une mesure de probabilité  $\mu$  sur  $\mathbb{R}^d$  si et seulement si

$$\forall \xi \in \mathbb{R}^d, \hat{\mu}_n(\xi) \xrightarrow[n \rightarrow \infty]{} \hat{\mu}(\xi)$$

De manière équivalente, une suite  $X_n$  de variables aléatoires à valeurs dans  $\mathbb{R}^d$  converge en loi vers  $X$  si et seulement si

$$\forall \xi \in \mathbb{R}^d, \Phi_{X_n}(\xi) \xrightarrow[n \rightarrow \infty]{} \Phi_X(\xi)$$

## 7.3 Conséquences

### 7.3.1 Convergence des Mesures Empiriques

**Théorème 7.3.1.** Soit  $X_n$  une suite de v.a. i.i.d à valeurs dans  $\mathbb{R}^d$ . Pour tout  $\omega \in \Omega$  et tout  $n \geq 1$ , soit  $\mu_{n,\omega}$  définie par :

$$\mu_{n,\omega} = \frac{1}{n} \sum_{i=1}^n \delta_{X_i(\omega)}$$

Alors presque sûrement :

$$\mu_{n,\omega} \xrightarrow[n \rightarrow \infty]{(e)} \mathbb{P}_{X_1}$$

### 7.3.2 Théorème Central Limite

**Théorème 7.3.2.** Soit  $X_n$  une suite de variables aléatoires réelles indépendantes et de même loi dans  $L^2$ . Soit  $\sigma^2 = \text{var } X_1$ . Alors :

$$\frac{1}{\sqrt{n}} (X_1 + \dots + X_n - nE[X_1]) \xrightarrow[n \rightarrow \infty]{(loi)} \mathcal{N}(0, \sigma^2)$$

De manière équivalente, pour tous  $a, b \in \bar{\mathbb{R}}$  :

$$\lim_{n \rightarrow \infty} \mathbb{P} (X_1 + \dots + X_n \in [nE[X_1] + a\sqrt{n}, nE[X_1] + b\sqrt{n}]) = \frac{1}{\sigma\sqrt{2\pi}} \int_a^b \exp\left(-\frac{x^2}{2\sigma^2}\right) dx$$

**Définition 7.3.1.** Soit  $C$  une matrice  $d \times d$  réelle symétrique positive. Une v.a.  $X$  à valeurs dans  $\mathbb{R}^d$  de carré intégrable, est appelée vecteur gaussien centré de covariance  $C$  si :

$$\forall \xi \in \mathbb{R}^d, \Phi_X(\xi) = E[e^{i\xi \cdot X}] = \exp\left(-\frac{1}{2} \xi^t C \xi\right)$$

On dit aussi que  $X$  suit la loi  $\mathcal{N}(0, C)$ .

**Proposition 7.3.1.** Soit  $C$  une matrice symétrique positive. Il existe un vecteur gaussien centré de covariance  $C$ .

**Théorème 7.3.3** (Théorème Central Limite Vectoriel). Soit  $X_n$  une suite de v.a. i.i.d. dans  $\mathbb{R}^d$  de carré intégrable. Alors,

$$\frac{1}{\sqrt{n}} (X_1 + \dots + X_n - nE[X_1]) \xrightarrow[n \rightarrow \infty]{(loi)} \mathcal{N}(0, K_{X_1})$$

## 7.4 Intervalles de Confiance

Pour un institut de sondage qui cherche à évaluer les intentions de vote entre deux candidats  $A$  et  $B$ , on peut modéliser chaque intention de vote par une variable de Bernoulli de paramètre  $p \in ]0, 1[$  inconnu. On suppose que les électeurs sont indépendants.

On observe la moyenne empirique

$$\hat{m}_n = \frac{1}{n} (X_1 + \dots + X_n)$$

Par la loi forte des grands nombre,  $\hat{m}_n \xrightarrow[n \rightarrow \infty]{p.s.} p$ .

On cherche à quantifier

$$\mathbb{P}(|\hat{m}_n - p| > \varepsilon), \text{ pour } \varepsilon > 0$$

- Par l'inégalité de Bienaymé-Tchebychev :

$$P\mathbb{P}|\hat{m}_n - p| > \varepsilon \leq \frac{1}{\varepsilon^2} \mathbb{E} \left[ |\hat{m}_n - p|^2 \right] = \frac{1}{n\varepsilon^2} \text{var } X_1 \leq \frac{p(1-p)}{n\varepsilon^2} \leq \frac{1}{4n\varepsilon^2}$$

Ici, par exemple, avec une probabilité de 95%, on a

$$p \in [\hat{m}_n - \varepsilon, \hat{m}_n + \varepsilon] \text{ où } \varepsilon = \frac{1}{\sqrt{4n \times 5.10^{-2}}}$$

**Remarque 7.4.0.1.** *L'intervalle de confiance est toujours en taille  $\frac{1}{\sqrt{n}}$*

**Lemme 7.4.1.** *Soit  $(Y_n)$  une suite de v.a. i.i.d. centrées bornées tel que  $|Y_1| \leq c$  presque sûrement. Alors,*

$$\mathbb{P} \left( \left| \frac{1}{n} (Y_1 + \dots + Y_n) \right| > \varepsilon \right) \leq 2 \exp \left( -n \frac{\varepsilon^2}{2c^2} \right)$$

Dans le lemme précédent, on obtient, avec une probabilité de 95% l'intervalle précédent pour

$$\varepsilon = \sqrt{\frac{2c(\ln(2) - \ln(5.10^{-2}))}{n}}$$

- Intervalles de Confiance Asymptotiques Par le Théorème Central Limite :

$$\frac{\sqrt{n}}{\sigma} (\hat{m}_n - p) \xrightarrow[n \rightarrow \infty]{(loi)} \mathcal{N}(0, 1) \text{ où } \sigma = p(1-p)$$

En particulier, si  $a > 0$  :

$$\mathbb{P} \left( \left| \frac{\sqrt{n}}{\sigma} (\hat{m}_n - p) \right| \geq a \right) \simeq_{n \gg 1} \frac{1}{\sqrt{2\pi}} \int_{|x| \geq a} \exp \left( -\frac{x^2}{2} \right) dx$$

L'intervalle de confiance asymptotique à 95% est donc (les valeurs de  $a$  sont tabulées, ici  $a = 1,96$ ) :

$$\left[ \hat{m}_n - \frac{a\sigma}{\sqrt{n}}, \hat{m}_n + \frac{a\sigma}{\sqrt{n}} \right]$$

On peut ici remplacer la convergence en loi par la convergence en probabilité sur les intervalles par le théorème de Porte-Manteau.

On aimerait remplacer  $\sigma$  par un écart type empirique, par exemple

$$\begin{aligned} \sigma_n^2 &= \frac{1}{n} \sum_{k=1}^n (X_k - \hat{m}_n)^2 \\ &= \frac{1}{n} \sum_{k=1}^n X_k^2 - 2\hat{m}_n^2 + \hat{m}_n^2 \\ &= \frac{1}{n} \sum_{k=1}^n X_k^2 - \hat{m}_n^2 \\ &= \hat{m}_n - \hat{m}_n^2 \end{aligned}$$

Par la loi forte des grands nombres,

$$\sigma_n^2 \xrightarrow[n \rightarrow \infty]{(p.s.)} p(1-p)$$

Toutefois,  $E[\sigma_n^2] \neq p(1-p)$ . En effet,

$$\begin{aligned} E[\sigma_n^2] &= E[\hat{m}_n - \hat{m}_n^2] \\ &= p - \frac{1}{n^2} \sum_{1 \leq i, j \leq n} E[X_i X_j] \\ &= p - \frac{1}{n} p - \frac{1}{n^2} n(n-1)p^2 \\ &= \frac{n-1}{n} p(1-p) \end{aligned}$$

On dit que  $\sigma_n$  est un estimateur avec biais. On le remplace fréquemment par l'estimateur sans biais

$$\tilde{\sigma}_n^2 = \frac{n}{n-1} \sigma_n^2 = \frac{1}{n-1} \sum_{k=1}^n (X_k - \hat{m}_n)^2$$

**Lemme 7.4.2** (Lemme de Slutsky). *Soit  $Z_n, \alpha_n$  deux suites de v.a. à valeurs dans  $\mathbb{R}^d$ . On suppose*

$$Z_n \xrightarrow[n \rightarrow \infty]{(loi)} Z$$

et

$$\alpha_n \xrightarrow[n \rightarrow \infty]{(loi)} \alpha$$

avec  $\alpha$  une constante i.e.  $\alpha_n \xrightarrow[n \rightarrow \infty]{proba} \alpha$ .

Alors,

$$(Z_n, \alpha_n) \xrightarrow[n \rightarrow \infty]{(loi)} (Z, \alpha)$$

*Démonstration.* On utilise le théorème de Lévy. ■

Ici,  $\sigma_n^2, \tilde{\sigma}_n^2 \xrightarrow[n \rightarrow \infty]{(p.s.)} p(1-p)$  donc en probabilité et la limite est constante.

Donc,

$$(\sqrt{n}(\hat{m}_n - p), \sigma_n) \xrightarrow[n \rightarrow \infty]{(loi)} (Z, \sigma) \text{ où } \mathcal{N}(0, 1)$$

et par conséquent,

$$\mathbb{1}_{\tilde{\sigma}_n \neq 0} \frac{\sqrt{n}}{\tilde{\sigma}_n} (\hat{m}_n - p) \xrightarrow[n \rightarrow \infty]{(loi)} \mathcal{N}(0, 1)$$

## 8 Espérance Conditionnelle

**Définition 8.0.1.** *Soit  $(\Omega, \mathcal{A}, \mathbb{P})$  un espace de probabilité,  $B \in \mathcal{A}$  de probabilité non nulle. On définit*

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

*De même, pour toute v.a.  $X \geq 0$ , ou pour  $X \in L^1(\Omega, \mathcal{A}, \mathbb{P})$ , on définit l'espérance conditionnelle de  $X$  sachant  $B$  comme*

$$E[X | B] = \frac{E[X \mathbb{1}_B]}{P(B)}$$

### 8.1 Variables Aléatoires Intégrables

**Définition 8.1.1.** *Soit  $\mathcal{B}$  une sous-tribu de  $\mathcal{A}$  et  $X$  une v.a. intégrable. Alors, il existe une unique v.a.  $\tilde{X} \in L^1$  telle que*

$$\forall B \in \mathcal{B}, E[\tilde{X} \mathbb{1}_B] = E[X \mathbb{1}_B]$$

*On appelle  $\tilde{X}$  l'espérance de conditionnelle de  $X$  sachant  $\mathcal{B}$  et on note  $\tilde{X} = E[X | \mathcal{B}]$ .*

**Lemme 8.1.1.** *La définition précédente est équivalente à  $\forall Z$  v.a.  $\mathcal{B}$ -mesurable bornée,  $E[XZ] = E[E[X | \mathcal{B}]Z]$ .*

**Proposition 8.1.1.** *L'espérance conditionnelle est*

- *Linéaire* :  $\forall X, Y \in L^1, \lambda \in \mathbb{R}, E[X + \lambda Y | \mathcal{B}] = E[X | \mathcal{B}] + \lambda E[Y | \mathcal{B}]$ .
- *Positive* :  $X \geq 0 p.s. \Rightarrow E[X | \mathcal{B}] \geq 0 p.s.$
- *Continue* :  $E[|E[X | \mathcal{B}]|] \leq E[|X|]$ .

**Remarque 8.1.1.1.**  $E[X | \mathcal{B}]$  est la meilleure approximation de  $X$  compte tenu de l'information donnée par la tribu  $\mathcal{B}$

## 8.2 Cas des Variables Positives

**Théorème 8.2.1.** *Soit  $X$  une v.a.  $\mathcal{A}$  mesurable et positive, il existe une unique application (à un ensemble de mesure nulle près)  $\mathcal{B}$ -mesurable et positive, notée  $E[X | \mathcal{B}]$  telle que*

$$\forall Z \in L^0(\mathcal{B}), Z \geq 0, E[E[X | \mathcal{B}]Z] = E[XZ]$$

*Elle se définit par la formule*

$$E[X | \mathcal{B}] = \lim_{n \rightarrow \infty} \uparrow E[X \wedge n | \mathcal{B}] p.s$$

**Proposition 8.2.1.** *L'espérance conditionnelle est linéaire (pour des coefficients positifs), monotone, vérifie les théorèmes de convergence.*

**Lemme 8.2.2.** *Soit  $\varphi$  une fonction convexe positive,  $X \in L^1$ ,  $\mathcal{B}$  une sous-tribu de  $\mathcal{A}$ . Alors, presque sûrement,*

$$\varphi(E[X | \mathcal{B}]) \leq E[\varphi(X) | \mathcal{B}]$$

**Proposition 8.2.2.** • *Soit  $X$   $\mathcal{B}$ -mesurable, intégrable. Alors  $E[X | \mathcal{B}] = X p.s$*

- *Si  $X$  est indépendante de la tribu  $\mathcal{B}$ , alors  $E[X | \mathcal{B}] = E[X]$*
- *Soit  $\mathcal{B}_1 \subset \mathcal{B}_2$  deux sous-tribus de  $\mathcal{A}$ . Alors,*

$$E[E[X | \mathcal{B}_2] | \mathcal{B}_1] = E[X | \mathcal{B}_1]$$

**Théorème 8.2.3.** *Si  $X \in L^2$ , alors  $E[X | \mathcal{B}]$  est la projection orthogonale de  $X$  sur  $L^2(\Omega, \mathcal{B}, \mathbb{P})$ .*