Probabilités et introduction à la Statistique
M1 premier semestre
Résumé de cours

F. Barthe
Institut de Mathématiques.
Université Paul Sabatier.
31062 Toulouse cedex 09.
Email : barthe@math.univ-toulouse.fr

2018

Les lecteurs intéressés par les preuves des énoncés pourront consulter entre autres les livres suivants :

Philippe BARBE et Michel LEDOUX,
Probabilité. De la Licence à l'Agrégation.
Editions Espaces 34, Belin (1998).
Nouvelle édition EDP Sciences (2007).

Dominique FOURDRINIER,
Statistique Inférentielle 2ème Cycle - Cours et Exercices Corrigés.
Sciences Sup, Dunod (2002).

Table des matières

Chapitre 1

Théorie de la mesure.

1.1 Algèbres et tribus

Définition 1. Un ensemble de parties de Ω, $\mathcal{A} \subset \mathcal{P}(\Omega)$ est une *algèbre (de Boole)* si

1. $\Omega \in \mathcal{A}$,
2. $\mathcal{A}$ est stable par complémentation : $A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$,
3. $\mathcal{A}$ est stable par réunion finie : $A, B \in \mathcal{A} \;\Rightarrow\; A \cup B \in \mathcal{A}$.

Définition 2. Un ensemble de parties de Ω, $\mathcal{A} \subset \mathcal{P}(\Omega)$ est une *tribu* (ou σ-*algèbre*) si

1. $\Omega \in \mathcal{A}$,
2. $\mathcal{A}$ est stable par complémentation : $A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$,
3. $\mathcal{A}$ est stable par réunion dénombrable : $(\forall n \in \mathbb{N}, A_n \in \mathcal{A}) \Rightarrow \bigcup_{n\in\mathbb{N}} A_n \in \mathcal{A}$.

Remarque 1. $\cap A_n = \big(\cup\big(A_n^c\big)\big)^c$. Une tribu est donc aussi stable par intersection dénombrable.

Vocabulaire. Si $\mathcal{A} \subset \mathcal{P}(\Omega)$ est une tribu, $(\Omega, \mathcal{A})$ est un *espace mesurable.* Tout $A \in \mathcal{A}$ est un *ensemble mesurable.*

Proposition 1. *Une intersection d'algèbres sur Ω est une algèbre sur Ω. Une intersection de tribus sur Ω est une tribu sur Ω.*

Définition 3. Soit $\mathcal{E} \subset \mathcal{P}(\Omega)$.

1. L'algèbre engendrée par $\mathcal{E}$ est par définition l'intersection des algèbres contenant $\mathcal{E}$. On la note $a(\mathcal{E})$.
2. La tribu engendrée par $\mathcal{E}$ est par définition l'intersection des algèbres contenant $\mathcal{E}$. On la note $\sigma(\mathcal{E})$. C'est la plus petite tribu sur Ω contenant $\mathcal{E}$.
   Si $\mathcal{A}$ est une tribu alors $\mathcal{E} \subset \mathcal{A} \Leftrightarrow \sigma(\mathcal{E}) \subset \mathcal{A}$.

Définition 4. Soit Ω un espace topologique. On appelle tribu *Borélienne* sur Ω, notée $\mathcal{B}(\Omega)$, la tribu engendrée par les ouverts de Ω. Tout $\mathcal{A} \in \mathcal{B}(\Omega)$ est appelé *ensemble borélien.*

Exemple 2. On munit $\mathbb{R}$ et $\overline{\mathbb{R}} := \mathbb{R} \cup \{-\infty, +\infty\}$ de leurs topologies usuelles, induites par l'ordre. On a alors

$$\mathcal{B}(\mathbb{R}) = \sigma\Big(\big\{]-\infty, a];\ a \in \mathbb{R}\big\}\Big) \quad \text{et} \quad \mathcal{B}(\overline{\mathbb{R}}) = \sigma\Big(\big\{[-\infty, a];\ a \in \mathbb{R}\big\}\Big).$$

Définition 5. Soient (Ω_1, M_1), (Ω_2, M_2) deux espaces mesurables. La tribu produit $M_1 \otimes M_2$ sur $\Omega_1 \times \Omega_2$ est la tribu engendrée par les pavés

$$A_1 \times A_2,\ A_1 \in M_1,\ A_2 \in M_2.$$

1.2 Fonctions mesurables

Définition 6. Soient $(\Omega, \mathcal{A})$, $(E, \mathcal{B})$ des espaces mesurables.

— Une fonction $f : \Omega \to E$ est dite *mesurable* (pour $\mathcal{A}$ et $\mathcal{B}$) si

$$\forall B \in \mathcal{B}, f^{-1}(B) \in \mathcal{A},$$

c'est à dire si $f^{-1}(\mathcal{B}) \subset \mathcal{A}$ où l'on a posé $f^{-1}(\mathcal{B}) := \left\{f^{-1}(B)\ ,\ B \in \mathcal{B}\right\}$.

— On appelle *tribu engendrée par* f la tribu

$$\sigma(f) := \sigma\left(f^{-1}(\mathcal{B})\right) = f^{-1}(\mathcal{B}) = \left\{f^{-1}(B), B \in \mathcal{B}\right\}.$$

C'est la plus petite tribu sur Ω qui rend f mesurable.

— Plus généralement, si $\mathcal{F}$ est une famille de fonction de $\Omega \longrightarrow (E, \mathcal{B})$, la tribu engendrée par $\mathcal{F}$, noté $\sigma(\mathcal{F})$ est la plus petite tribu sur Ω qui rend mesurable toute fonction de $\mathcal{F}$

$$\sigma(\mathcal{F}) := \sigma\big(\left\{f^{-1}(B),\ B \in \mathcal{B},\ f \in \mathcal{F}\right\}\big).$$

Vocabulaire. Une application mesurable $f : (\Omega,\ \mathcal{A}) \longrightarrow (E, \mathcal{B}(E))$ est dite *borélienne.*

Exemple 3. Une fonction f définie sur $(\Omega, \mathcal{A})$ et à valeurs dans $\mathbb{R}$ ou $\overline{\mathbb{R}}$ est borélienne si et seulement si

$$\forall a \in \mathbb{R}, \quad \{\omega \in \Omega;\ f(\omega) \leq a\} \in \mathcal{A}.$$

Proposition 2. *Vérification de la mesurabilité des fonctions sur une partie génératrice.*

— *Soit* $f : (\Omega,\ \mathcal{A}) \to (E,\ \mathcal{B})$ *et soit* $\mathcal{E} \subset \mathcal{P}(E)$, *avec* $\sigma(\mathcal{E}) = \mathcal{B}$, *alors*

$$\sigma(f) = \sigma\left(f^{-1}(\mathcal{E})\right) = \sigma\big(\{f^{-1}(C),\ C \in \mathcal{E}\}\big).$$

Donc f est mesurable si et seulement si : $\forall C \in \mathcal{E},\ f^{-1}(C) \in \mathcal{A}$.

— *Plus généralement, si* $\mathcal{F}$ *est une famille de fonctions* $\Omega \to E$ *alors*

$$\sigma(\mathcal{F}) = \sigma\left(\left\{f^{-1}(C),\ C \in \mathcal{E},\ f \in \mathcal{F}\right\}\right).$$

Proposition 3. *Stabilité des fonctions mesurables.*

— *Composition : on considère* $(\Omega_1, \mathcal{A}_1) \xrightarrow{f_1} (\Omega_2, \mathcal{A}_2) \xrightarrow{f_2} (\Omega_3, \mathcal{A}_3)$. *Dans ce cas :* f_1, f_2 *mesurables* $\Rightarrow f_2 \circ f_1$ *mesurables.*

— *Soient pour* $i = 1$ *ou 2,* $f_i : (\Omega, \mathcal{A}) \to (E_i, \mathcal{B}_i)$ *et soit* $f : (\Omega, \mathcal{A}) \to (E_1 \times E_2, \mathcal{B}_1 \otimes B_2)$ *définie par* $f(\omega) = (f_1(\omega), f_2(\omega))$. *Alors* f *est mesurable si et seulement si* f_1 *et* f_2 *sont mesurables.*

— *Soient* $(f_n)_{n \in \mathbb{N}}$ *une suite de fonctions mesurables de* $(\Omega, \mathcal{A}) \to \left(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d)\right)$.

  — *L'ensemble* $C := \{\omega \in \Omega, \exists \lim_n f_n(\omega)\}$ *est mesurable et la fonction* f *qui vaut* $\lim_n f_n$ *quand la limite existe et* 0 *sinon est mesurable.*

  — $\forall \alpha \in \mathbb{R},\ \alpha f_1$ *est mesurable.*

  — $f_1 + f_2$ *est mesurable.*

— *Si* $f, g : (\Omega, \mathcal{A}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ *sont mesurables, alors* $f + g$, $f \times g$, $\max(f, g)$ *et* $\min(f, g)$ *sont mesurables.*

— *Si pour* $n \geq 1$, $f_n : (\Omega, \mathcal{A}) \to \left(\overline{\mathbb{R}}, \mathcal{B}\big(\overline{\mathbb{R}}\big)\right)$ *sont mesurables, alors* $\max(f_1, f_2)$, $\min(f_1, f_2)$, $\limsup_n f_n$ *et* $\liminf_n f_n$ *sont mesurables. Si* f_1 *et* f_2 *sont positives, alors* $f_1 + f_2$ *est bien définie et mesurable.*

Proposition 4. *Si* $f : (\Omega_1, \mathcal{B}(\Omega_1)) \to (\Omega_2, \mathcal{B}(\Omega_2))$ *est continue, alors elle est mesurable.*

Définition 7. Une fonction $f : (\Omega, \mathcal{A}) \to \overline{\mathbb{R}}$ est dite *étagée* s'il existe $k \in \mathbb{N},\ A_1, \dots, A_k \in \mathcal{A}$ disjoints et $a_1, \dots, a_n \in \mathbb{R}^+$ avec $f(\omega) = \sum_{i=1}^k a_i \mathbb{1}_{A_i}(\omega)$.

Une telle fonction est mesurable quelle que soit la tribu à l'arrivée.

Proposition 5. *Toute fonction $f : (\Omega, \mathcal{A}) \to \big(\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}})\big)$, à valeurs positives, est limite simple croissante de fonctions étagées.*

1.3 Classes monotones.

Cette notion, plus maniable que celle de tribu, sera utile pour définir les mesures.

Définition 8. $\mathcal{M} \subset \mathcal{P}(\Omega)$ est une *classe monotone* (c.m. en abrégé) si :

1. $\Omega \in \mathcal{M}$,
2. $\big[A, B \in \mathcal{M}$ et $B \subset A\big] \Rightarrow A \setminus B = A \cap B^c \in \mathcal{M}$,
3. $\big[\forall i \in \mathbb{N},\ A_i \in \mathcal{M}$ et $A_i \subset A_{i+1}\big] \Rightarrow \bigcup_{i\in\mathbb{N}} A_i \in \mathcal{M}$.

Définition-Proposition 9. Une intersection quelconque de c.m. est une c.m.

Soit $\mathcal{E} \subset \mathcal{P}(\Omega)$. La classe monotone engendrée par $\mathcal{E}$, notée $\mathcal{M}(\mathcal{E})$, est par définition l'intersection de toutes les c.m. contenant $\mathcal{E}$. C'est la plus petite c.m. sur Ω qui contient $\mathcal{E}$.

Proposition 6. *Soit $\mathcal{M}$ une classe monotone. Si elle est stable par intersection finie, alors c'est une tribu.*

Théorème 7 (Classes monotones)**.** *Soit $\mathcal{E} \subset \mathcal{P}(\Omega)$ stable par intersection finie. Alors $\mathcal{M}(\mathcal{E}) = \sigma(\mathcal{E})$.*

Définition 10. Un ensemble $\mathcal{H}$ de fonctions de $\Omega \to \mathbb{R}$ est dit stable par convergence monotone bornée si pour toute suite $(f_n)_{n\in\mathbb{N}}$, croissante et bornée ($\exists\ C,\ \forall n,\ \forall \omega,\ |f_n(\omega)| \leq C$) de fonctions de $\mathcal{H}$, la limite $f = \lim f_n$ est dans $\mathcal{H}$.

Théorème 8. (Version fonctionnelle du théorème de classe monotone)

Soit $\mathcal{H}$ un $\mathbb{R}$-espace vectoriel de fonctions bornées de $\Omega \to \mathbb{R}$ stable par convergence monotone bornée. Soit $\mathcal{C} \subset \mathcal{H}$ un ensemble de fonctions

— *contenant les constantes,*

— *stable par multiplication et addition d'une constante.*

Alors $\mathcal{H}$ contient l'ensemble $b\,(\sigma(\mathcal{C}))$ des fonctions bornées mesurables pour

$$\sigma(\mathcal{C}) = \sigma\left(\left\{f^{-1}(B),\ f \in \mathcal{C},\ B \in \mathcal{B}(\mathbb{R})\right\}\right).$$

Remarque 4. Ce résultat peut être vu comme un analogue borélien du théorème de Stone-Weierstrass.

1.4 Mesures

Définition 11. Soit $(\Omega, \mathcal{A})$ un espace mesurable. On appelle *mesure* sur $(\Omega, \mathcal{A})$ toute application $\mu : \mathcal{A} \to \mathbb{R}^+ \cup \{\infty\}$ telle que

— $\mu(\emptyset) = 0$,

— si $(A_i)_{i\in I}$ est une famille au plus dénombrable d'ensembles disjoints de $\mathcal{A}$ alors

$$\mu\Big(\bigcup_{i\in I} A_i\Big) = \sum_{i\in I} \mu(A_i).$$

Vocabulaire. On dit que $(\Omega, \mathcal{A}, \mu)$ est un *espace mesuré.*

Définition 12. On dit que μ est une *mesure de probabilité* si $\mu(\Omega) = 1$, une mesure *finie* si $\mu(\Omega) < +\infty$. On dit que μ est σ-finie s'il existe une suite $(A_n)_{n\in\mathbb{N}}$ dans $\mathcal{A}$ avec $\forall n,\ \mu(A_n) < \infty$ et $\Omega = \bigcup_n A_n$.

Proposition 9. *Soit $(\Omega, \mathcal{A}, \mu)$ un espace mesuré. Soient A, B* et *$(A_i)_{i\in I}$ des éléments de $\mathcal{A}$:*

1. *$A \subset B \Rightarrow \mu(A) \leq \mu(B)$.*
2. *$\mu(A \cup B) + \mu(A \cap B) = \mu(A) + \mu(B)$.*
3. *Si I est au plus dénombrable, alors $\mu(\bigcup_{i\in I} A_i) \leq \sum_{i\in I} \mu(A_i)$.*
4. *Si $\forall n \in \mathbb{N},\ A_n \subset A_{n+1}$ alors $\mu(\bigcup_{n\in\mathbb{N}} A_n) = \lim_{n\to\infty} \uparrow \mu(A_n)$.*
5. *Si $\forall n \in \mathbb{N},\ A_n \supset A_{n+1}$* et *$\exists n_o,\ \mu(A_{n_o}) < \infty$ alors $\mu(\bigcap_{n\in\mathbb{N}} A_n) = \lim_{n\to\infty} \downarrow \mu(A_n)$*

Proposition 10. (Constructions de nouvelles mesures, à partir de mesures)

1. *Si μ, ν sont des mesures sur $(\Omega, \mathcal{A})$ et si $\alpha \in \mathbb{R}^+$*
   alors $\alpha\mu + \nu :\ A \mapsto \alpha\mu(A) + \nu(A)$ est aussi une mesure.
2. *Si $A \in \mathcal{A}$ on peut définir la restriction de μ à A en posant pour $B \in \mathcal{A}$, $\mu_A(B) = \mu(A\cap B)$. C'est aussi une mesure.*
3. *Mesure image par une fonction.*
   Soit $f : (\Omega, \mathcal{A}) \mapsto (E, \mathcal{B})$ mesurable, soit μ mesure sur $(\Omega, \mathcal{A})$.
   L'application : $\mu_f : \mathcal{B} \to \mathbb{R}^+ \cup \{\infty\}$ définie par $\mu_f(B) = \mu\big(f^{-1}(A)\big)$ est une mesure sur $(E, \mathcal{B})$ appelée mesure image de μ par f .

La construction de mesures à partir de rien est délicate car les tribus sont souvent décrites par parties génératrices. Il est ainsi difficile de définir $\mu(A)$ pour $A \in \mathcal{A}$ quelconque. On souhaiterait définir la mesure sur une partie génératrice l'étendre de façon unique.

Théorème 11. (Caractérisation des mesures finies)
Soit $(\Omega, \mathcal{A})$ un espace mesuré. Soit $\mathcal{E} \subset \mathcal{A} \subset \mathcal{P}(\Omega)$ telle que
— $\Omega \in \mathcal{E}$,
— $A, B \in \mathcal{E} \Rightarrow A \cap B \in \mathcal{E}$,
— $\sigma(\mathcal{E}) = \mathcal{A}$.
Si μ et *ν sont deux mesures finies sur $(\Omega, \mathcal{A})$ qui coincident sur $\mathcal{E}$ alors $\mu = \nu$.*

Théorème 12. (Prolongement)
Soit $\mathcal{C}$ une algèbre de Boole de parties de Ω. Soit $\mu : \mathcal{C} \to \mathbb{R}^+ \cup \{\infty\}$ telle que
— $\mu(\emptyset) = 0$,
— $A_1, \ldots, A_n \in \mathcal{C} \Rightarrow \mu(A_1 \cup \ldots \cup A_n) = \sum_{i\leq n} \mu(A_i)$,
— *si $\forall n \in \mathbb{N},\ A_n \subset A_{n+1} \in \mathcal{C}$* et *$\bigcup_{n\in\mathbb{N}} A_n \in \mathcal{C}$ alors $\mu(A) = \lim_{n\to\infty} \mu(A_n)$.*
Alors μ se prolonge en une mesure sur $(\Omega, \sigma(\mathcal{C}))$. Si μ est σ-finie sur $\mathcal{C}$, le prolongement est unique et σ-fini.

Exemple 5. Applications : construction de la mesure de Lebesgue sur $\mathbb{R}$, construction de mesures produits.

Chapitre 2

Rappels d'intégration

2.1 Construction de l'intégrale de Lebesgue.

Soient $(\Omega, \mathcal{A}, \mu)$ un espace mesuré, $f : (\Omega, \mathcal{A}) \mapsto (\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}}))$ une application borélienne et $B \in \mathcal{A}$. L'intégrale de f sur B par rapport à μ sera notée

$$\int_B f\, d\mu = \int_B f(\omega)\, d\mu(\omega) = \int_B f(\omega)\, \mu(d\omega).$$

Si $B = \Omega$ on note simplement $\int f\, d\mu$, $\int f(\omega)\, d\mu(\omega)$ ou $\int f(\omega)\, \mu(d\omega)$.

Définition 13. (Définition de l'intégrale par étapes)

1. Si $f = \mathbb{1}_A$ avec $A \in \mathcal{A}$ alors $\int_B f\, d\mu := \mu(A \cap B)$.
2. Si f est étagée : $f = \sum_{i=1}^n a_i \mathbb{1}_{A_i}$ avec $a_i \in \mathbb{R}^+$, et $A_i \in \mathcal{A}$ disjoints alors, avec la convention $0 \cdot \infty = 0$,
$$\int_B f\, d\mu := \sum_{i=1}^n a_i\, \mu(A_i \cap B).$$
3. Si $f \geq 0$, on pose
$$\int_B f\, d\mu := \sup \left\{ \int_B g\, d\mu,\ g \quad \text{étagée } \leq f \right\} \in [0, +\infty].$$
On dit que f est intégrable sur B si $\int_B f\, d\mu < +\infty$.
4. Si f est quelconque elle s'écrit $f = f_+ - f_-$ et $|f| = f_+ + f_-$.
On dit que f est μ-intégrable sur B si $\int_B |f|\, d\mu < \infty$, ce qui équivaut à
$$\int_B f_+\, d\mu < \infty \text{ et } \int_B f_-\, d\mu < \infty.$$
Dans ce cas $\int_B f\, d\mu := \int_B f_+\, d\mu - \int_B f_-\, d\mu \in \mathbb{R}$.

Remarque 6. Si f est intégrable par rapport à μ alors $\mu\Big(\{\omega;\ f(\omega) \in \{-\infty, +\infty\}\}\Big) = 0$.

Proposition 13. (Opérations et propriétés élémentaires)

— *Linéarité : si f, g sont des intégrables et α, $\beta \in \mathbb{R}$.*
$$\int_B (\alpha f + \beta g)\, d\mu = \alpha \int_B f\, d\mu + \beta \int_B g\, d\mu.$$
(vrai aussi si $f, g \geq 0$ et $\alpha, \beta \geq 0$.)

- *Ordre :*
  - *Si* $f \geq g$ *sont intégrables alors* $\int_B f\,d\mu \geq \int_B g\,d\mu$.
    Si $f \geq g \geq 0$ *alors* $\int_B f\,d\mu \geq \int_B g\,d\mu \geq 0$.
  - *Si* $A \subset B$ *et* $f \geq 0$ *alors* $\int_B f\,d\mu \geq \int_A f\,d\mu$.
  - *Si* $f = 0$ *alors* $\int f\,d\mu = 0$.
  - *Si* $f \geq 0$ et $\int f\,d\mu = 0$ *alors* $f = 0$ μ*-p.p.*
  - *Si* $\mu(B) = 0$ *alors* $\int_B f\,d\mu = 0$.
- *Si* $f \geq 0$ *(ou intégrable sur* B*) alors* $f\,1_B$ *l'est aussi et* $\int_B f\,d\mu = \int 1_B f\,d\mu$.

2.2 Théorèmes de convergence

Théorème 14 (Convergence monotone)**.** *Soit* $(f_n)_{n\in\mathbb{N}}$ *une suite croissante de fonctions mesurables à valeurs dans* $[0,+\infty]$*. Alors*

$$\lim_{n\to\infty} \int f_n\,d\mu = \int \lim_{n\to\infty} f_n\,d\mu.$$

Théorème 15 (Fatou)**.** *Soient* $(f_n)_{n\in\mathbb{N}}$ *des fonctions mesurables à valeurs dans* $[0,+\infty]$*, on a alors :*

$$\int \liminf_{n\to\infty} f_n\,d\mu \leq \liminf_{n\to\infty} \int f_n\,d\mu.$$

Théorème 16 (Convergence dominée)**.** *Si* $(f_n)_{n\in\mathbb{N}}$ *est une suite de fonctions mesurables, et s'il existe une fonction* g *intégrable (* $\int g\,d\mu < \infty$*) telle que* $|f_n| \leq g$ *et* μ*-presque partout* $f_n \to f$ *alors* f *est intégrable et*

$$\lim_{n\to\infty} \int f_n\,d\mu = \int f\,d\mu.$$

2.3 Mesures à densité

Proposition 17. *Soit* $f : (\Omega,\ \mathcal{A}) \to \left(\overline{\mathbb{R}},\ \mathcal{B}(\overline{\mathbb{R}})\right)$ *mesurable* ≥ 0 *et soit* μ *une mesure sur* $(\Omega,\ \mathcal{A})$*. On peut définir l'application* $\nu : \mathcal{A} \to \mathbb{R}^+ \cup \{\infty\}$ *par*

$$\nu(A) := \int_A f\,d\mu.$$

Alors ν *est une mesure et*
- *si* $\mu(A) = 0$ *alors* $\nu(A) = 0$.
- *si* g *est mesurable et positive alors* $\int g\,d\nu = \int fg\,d\mu$.
- *gest* ν*-intégrable si et seulement si* fg *est* μ*-intégrable et dans ce cas* $\int g\,d\nu = \int fg\,d\mu$.

Vocabulaire. On dit que ν est une *mesure à densité* f *par rapport à* μ et on note $d\nu = f\,d\mu$.

Définition 14. Soit μ,ν deux mesures sur $(\Omega,\mathcal{A})$. On dit que ν est *absolument continue par rapport à* μ si :

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

On note $\nu \ll \mu$. Si $\nu \ll \mu$ et $\mu \ll \nu$ on dit qu'elles sont *équivalentes*.

Théorème 18 (Radon-Nikodym)**.** *Soient* μ,ν *deux mesures* σ*-finies sur* $(\Omega,\mathcal{A})$ *telles que* $\nu \ll \mu$*. Alors il existe une fonction positive mesurable* f *telle que* $\forall A \in \mathcal{A}$ *on a* $\nu(A) = \int_A f\,d\mu$ *i.e.* $d\nu = f\,d\mu$.

Théorème 19. *Soient μ, ν des mesures σ-finies sur $(\Omega, \mathcal{A})$. Alors il existe une unique décomposition $\nu = \nu_{ac} + \nu_\perp$ avec*

- *— $\nu_{ac} \ll \mu$*
- *— $\nu_\perp$ étrangère à μ (i.e. $\exists A \in \mathcal{A}$ avec $\mu(A) = 0$ et $\nu_\perp(A^c) = 0$).*

2.4 Changements de variables.

On commence par le cas général.

Théorème 20 (Transport)**.** *Soit $f : (\Omega, \mathcal{A}, \mu) \to (E, \mathcal{E})$ mesurable et soit $\varphi : E \to \mathbb{R}$ borélienne.*

- *— Si $\varphi \geq 0$ alors $\int_\Omega \varphi \circ f \, d\mu = \int_E \varphi \, d\mu_f$ où μ_f est la mesure image définie par $\mu_f(B) = \mu(f^{-1}(B))$.*
- *— Si φ est quelconque : $\varphi \circ f$ est μ-intégrable si et seulement si φ est μ_f-intégrable et dans ce cas la formule est vraie.*

Ensuite un cas où le calcul différentiel aide à calculer μ_f.

Théorème 21 (Changement de variables)**.** *Soient A, B ouverts de $\mathbb{R}^n, f : A \to B$ un $\mathcal{C}^1$ difféomorphisme et $g = f^{-1}$. Alors*

$$\int_A \varphi(f(x)) \, dx = \int_B \varphi(y) \, |\det(Dg(y))| \, dy$$

2.5 Interversions

Théorème 22 (Tonelli et Fubini)**.** *Soient $(\Omega_1, \mathcal{A}_1, \mu_1)$ et $(\Omega_2, \mathcal{A}_2, \mu_2)$ deux espaces mesurés σ-finis. On considère l'espace produit $(\Omega := \Omega_1 \times \Omega_2, \mathcal{A}_1 \otimes \mathcal{A}_2, \mu := \mu_1 \otimes \mu_2)$. Soit $f : \Omega \to \overline{\mathbb{R}}$ mesurable. Alors la formule d'interversion suivante*

$$\begin{aligned} \int_\Omega f \, d\mu &= \int_{\Omega_1} \left(\int_{\Omega_2} f(\omega_1, \omega_2) \, d\mu_2(\omega_2) \right) d\mu_1(\omega_1) \\ &= \int_{\Omega_2} \left(\int_{\Omega_1} f(\omega_1, \omega_2) \, d\mu_1(\omega_1) \right) d\mu_2(\omega_2) \end{aligned}$$

est valable

- *— si f est positive (Tonelli)*
- *— ou si f est μ-intégrable, i.e. $\int_\Omega |f| \, d\mu < +\infty$ (Fubini). Pour le vérifier on peut utiliser le théorème de Tonelli pour $|f| \geq 0$.*

Corollaire 23. *On considére une suite de fonctions mesurables $f_n : (\Omega, \mathcal{A}, \mu) \to \overline{\mathbb{R}}$, $n \geq 0$. Alors l'égalité*

$$\int_\Omega \Big(\sum_{n \in \mathbb{N}} f_n \Big) \, d\mu = \sum_{n \in \mathbb{N}} \Big(\int_\Omega f_n \, d\mu \Big)$$

est valable si

- *— $f_n \geq 0$ pour tout n,*
- *— ou bien si $\int \big(\sum |f_n| \big) \, d\mu = \sum \big(\int |f_n| \, d\mu \big) < +\infty$.*

Théorème 24. (Dérivation sous le signe d'intégrale)

Soit $(\Omega, \mathcal{A}, \mu)$ un espace mesuré. Soit I un intervalle de $\mathbb{R}$, t_0 un point de l'intérieur de I et $\varepsilon > 0$ tel que $[t_0 - \varepsilon, t_0 + \varepsilon] \subset I$. Soit $f : \Omega \times I \to \mathbb{R}$ telle que

- *— pour tout $t \in I$, l'application $\omega \mapsto f(\omega, t)$ est μ-intégrable,*

— *le complémentaire de* $\Omega' := \{\omega \in \Omega;\ t \mapsto f(\omega,t)$ *est dérivable sur* $[t_0-\varepsilon, t_0+\varepsilon]\}$ *est négligeable pour* μ,

— *il existe une fonction* g *sur* Ω, μ*-intégrable telle que*

$$\forall t \in [t_0-\varepsilon, t_0+\varepsilon], \quad \forall \omega \in \Omega', \quad |\partial_2 f(\omega,t)| \leq g(\omega).$$

Alors l'application $t \mapsto F(t) := \int_\Omega f(\omega,t)\, d\mu(\omega)$ *est dérivable en* t_0 *et*

$$F'(t_0) = \int_\Omega \partial_2 f(\omega,t_0)\, d\mu(\omega).$$

2.6 Inégalités, espaces $L^p(\Omega, \mathcal{A}, \mu)$.

Définition 15. Sur $\mathcal{L}^\circ(\Omega,\mathcal{A},\mu) = \{f : (\Omega,\mathcal{A},\mu) \to (\mathbb{R},\mathcal{B}(\mathbb{R}))$ mesurables$\}$ on peut définir

$$\begin{aligned} \|f\|_p &:= \left(\int |f|^p\, d\mu\right)^{\frac{1}{p}} \quad \text{pour} \quad p \in]0,+\infty[\\ \|f\|_\infty &:= \inf\Big\{a \geq 0,\ \mu\big(\omega, |f(\omega)| > a\big) = 0\Big\}. \end{aligned}$$

Pour $p \in]0,+\infty]$, l'espace $L^p(\Omega,\mathcal{A},\mu)$ est obtenu comme quotient de $\{f \in \mathcal{L}^\circ(\Omega,\mathcal{A},\mu), \|f\|_p < \infty\}$ par la relation d'équivalence $f \overset{\mu-\text{p.p.}}{=} g$.

Proposition 25. (Structure et dualité des espaces L^p)

— *Pour* $p \in [1,\infty]$, $(L^p(\Omega,\mathcal{A},\mu), \|f\|_p)$ *est un espace de Banach. En particulier on a l'inégalité triangulaire*

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

— *Si* $p \in [1,\infty)$, *le dual de* L_p *est* $L_{p'}$ *où* p' *est l'exposant conjugué tel que* $\frac{1}{p} + \frac{1}{p'} = 1$. *Les inégalités de Hölder assurent que*

$$\begin{aligned} \int |fg|\, d\mu &\leq \left(\int |f|^p\, d\mu\right)^{\frac{1}{p}} \left(\int |g|^{p'}\, d\mu\right)^{\frac{1}{p'}} \\ \|fg\|_1 &\leq \|f\|_p\, \|g\|_{p'}. \end{aligned}$$

— *L'espace* $L^2(\Omega,\mathcal{A},\mu)$ *muni du produit scalaire* $(f,g) := \int fg\, d\mu$ *est un espace de Hilbert.*

Chapitre 3

Mesures de probabilité, variables aléatoires.

3.1 Mesures de probabilités, lois

Définition 16. Soit $(\Omega, \mathcal{A})$ un espace mesurable. Une *mesure de probabilité* sur $(\Omega, \mathcal{A})$ est une mesure P sur $(\Omega, \mathcal{A})$ avec $P(\Omega) = 1$. On parle aussi de *loi de probabilité*, où de loi.

— Si $P = \sum_{i\in I} p_i \delta_{x_i}$ avec $\sum_{i\in I} p_i = 1$ et I au plus dénombrable, on parle de loi discrète.
— Si $P \ll \mu$ on parle de loi à densité par rapport à μ.
— Si $\lambda_{\mathbb{R}^n}$ est la mesure de Lebesgue de $\mathbb{R}^n$ et si P est une mesure de probabilité sur $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ avec $P \ll \lambda_{\mathbb{R}^n}$, on parle de mesure à densité.

Remarque 7. Il n'y a pas que des mesures discrètes ou à densité !

Définition 17. Soit $A \in \mathcal{A}$ un événement.

— il a lieu *presque sûrement* (P-p.s en abrégé) si $P(A) = 1$.
— il est *négligeable* si $P(A) = 0$.

Proposition 26. *(On spécifie les propriétés des mesures dans le cas de probabilités.)*

— $A \subset B \Rightarrow P(A) \leq P(B)$
— $P(A^c) = 1 - P(A)$
— $P(A \cup B) = P(B) + P(B) - P(A \cap B)$
— $P(\bigcup_{n\in\mathbb{N}} A_n) \leq \sum_{n\in\mathbb{N}} P(A_n) \bigcup_{n\in\mathbb{N}}$.
— *Si* $\forall n,\ A_n \subset A_{n+1}$ *alors* $P(\bigcup_n A_n) = \lim \uparrow P(A_n)$
— *Si* $\forall n,\ A_n \supset A_{n+1}$ *alors* $P(\bigcap_{n\in\mathbb{N}} A_n) = \lim \downarrow P(A_n)$

Définition 18. On appelle *variable aléatoire* (v.a. en abrégé) sur $(\Omega, \mathcal{A}, P)$ toute fonction mesurable $X : (\Omega, \mathcal{A}, P) \to (E, \mathcal{B})$.

On appelle *loi de* X sous la probabilité P la mesure image P_X de P par X.

Par définition $\forall B \in \mathcal{B}$,

$$P_X(B) = P(X^{-1}(B)) = P\left(\{\omega \in \Omega, X(\omega) \in B\}\right) = P(X \in B).$$

Dans la pratique : on oubliera souvent $(\Omega, \mathcal{A}, P)$ pour se concentrer sur certaines variables, dont le comportement aléatoire est défini par la loi.

Vocabulaire. Une v.a. $X : (\Omega, \mathcal{A}, P) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ (ou $(\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}}))$) est une *variable aléatoire réelle* (v.a.r. en abrégé).

3.2 Moyennes, calculs et inégalités

Soit $X : (\Omega, \mathcal{A}, P) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ une v.a.r.

— Si X est P-intégrable ou ≥ 0 alors on appelle *espérance* de X ou *moyenne* de X le nombre $E(X) := \int_\Omega X\, dP$.

— Si $X \in L^p$, $p > 0$ son *moment absolu d'ordre* p est $E(|X|^p)$ et si $X \in L^p$, $p \in \mathbb{N}$ son *moment d'ordre* p est $E(X^p)$.

— Si $X \in L^2$ sa *variance* est $\mathrm{Var}(X) := E[(X - EX)^2] = E(X^2) - (EX)^2$.

— Si $X, Y \in L^2$, leur *covariance* est

$$\mathrm{Cov}(X, Y) := E\big[(X - EX)(Y - EY)\big] = E(XY) - (EX)(EY).$$

Théorème 27. *Soient $X : (\Omega, \mathcal{A}, P) \to (E, \mathcal{B})$ et $\varphi : (E, B) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$*

— *si $\varphi \geq 0$ alors $E\varphi(X) = \int_E \varphi(x)\, dP_X(x)$*

— *si φ est générale, φ est P_X-intégrable si et seulement si $\varphi(X)$ est P-intégrable et*

$$E\varphi(X) = \int_E \varphi\, dP_X.$$

C'est la traduction du théorème de transport. P_X suffit à calculer $Ef(X)$.

Proposition 28. (Quelques inégalités utiles)

— *Jensen : Si X est une v.a.r. avec $E|X| < \infty$ et $f : \mathbb{R} \to \mathbb{R}^+$ est une fonction convexe, alors $f(EX) \leq Ef(X)$.*

— *Markov : Si $a \in]0, +\infty[$ et X est une v.a.r. ≥ 0 alors $P(X \geq a) \leq \frac{E(x)}{a}$.*
Conséquence : Si X est une v.a.r. dans L^2 alors pour $a > 0$,

$$P\left(|X - EX| \geq a\right) \leq \frac{\mathrm{Var} X}{a^2}.$$

— *Si $1 \leq p \leq q \leq +\infty$, alors $\|X\|_p \leq \|X\|_q$ et $L^q(\Omega, \mathcal{A}, P) \subset L^p(\Omega, \mathcal{A}, P)$.*

Définition 19. $X : (\Omega, \mathcal{A}, P) \to (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ est dit p-intégrable si $E(\|X\|^p) < \infty$. Si on note $X = (X_1, \ldots, X_d)$, son espérance est le vecteur $EX = (EX_1, \ldots, EX_d) \in \mathbb{R}^d$. Sa matrice de covariance est

$$\mathrm{Cov}(X) = \big(\mathrm{Cov}(X_i, X_j)\big)_{\leq i,j \leq d} = \big(E(X_i - EX_i)(X_j - EX_j)\big)_{i,j}$$

C'est une matrice symétrique positive car

$$\sum \alpha_i \alpha_j \,\mathrm{Cov}(X_i, X_j) = \mathrm{Var}\big(\sum \alpha_i X_i\big) \geq 0$$

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

Définition 20. Soit $X : (\Omega, \mathcal{A}, P) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ une v.a.r.. On définit sa *fonction de répartition* $F_X : \mathbb{R} \to [0, 1]$ par

$$F_X(t) = P_X(] - \infty, t]) = P(X \leq t).$$

Proposition 29. $F_X = F_Y \Rightarrow P_X = P_Y$.

Proposition 30. — *F_X est croissante, continue à droite.*

$$\lim_{+\infty} F_X = 1,\ \lim_{-\infty} F_X = 0$$

— F_X *a une limite à gauche en tout point*

$$F_X(t^-) := \lim_{u \overset{<}{\to} t} F_X(u) = P(X < t).$$

— F_X *est discontinue en* $t \Leftrightarrow P(X = t) > 0$ *i.e.* t *est un atome pour* P_X.
— F_X *admet un nombre au plus dénombrable de discontinuités.*

Définition 21. Soit $F : \mathbb{R} \to [0, 1]$ croissante telle que $\lim_{+\infty} F = 1$, $\lim_{-\infty} F = 0$. Pour $u \in]0, 1[$ on définit sa fonction *inverse (à gauche) généralisée* comme suit :

$$F^{\leftarrow}(u) = \inf\{x, F(x) \geq u\}.$$

Proposition 31. *Soit* $F : \mathbb{R} \to [0, 1]$ *croissante, continue à droite, avec* $\lim_{+\infty} F = 1$ *et* $\lim_{-\infty} F = 0$.

1. *Pour tout* $u \in]0, 1[$, *pour tout* t

$$F^{\leftarrow}(u) \leq t \Leftrightarrow u \leq F(t).$$

2. *Si* U *est une v.a. uniforme sur* $[0, 1]$, *(i.e.* $dP_U(x) = \mathbb{1}_{[0,1]}(x)\, dx$*) alors* $F^{\leftarrow}(U)$ *admet* F *pour fonction de répartition.*
3. *Si* X *est une v.a.r. alors* $F_X^{\leftarrow}(U)$ *a même loi que* X *(en abrégé* $F_X^{\leftarrow}(U) \sim X$*).*

Proposition 32. P_X *est la dérivée de* F_X *au sens des distributions.*
Si F_X *est de classe* $\mathcal{C}^1$ *alors* $P_X(dt) = F_X'(t)\, dt$.

3.4 Fonction caractéristique

Définition 22. Soit μ une mesure finie sur $\big(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d)\big)$. Sa *transformée de Fourier* est la fonction définie pour $\xi \in \mathbb{R}^d$ par

$$\widehat{\mu}(\xi) = \int_{\mathbb{R}^d} e^{i\langle \xi, x\rangle} d\mu(x) = \int_{\mathbb{R}^d} \cos\big(\langle \xi, x\rangle\big)\, d\mu(x) + i \int_{\mathbb{R}^d} \sin\big(\langle \xi, x\rangle\big)\, d\mu(x).$$

Définition 23. Soit $X = (X_1, \ldots, X_d)$ une v.a. à valeurs dans $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. Sa *fonction caractéristique* est définie pour $\xi = (\xi_1, \ldots, \xi_d) \in \mathbb{R}^d$ par

$$\Phi_X(\xi) := E\big(e^{i\langle \xi, X\rangle}\big) = E\big(e^{i\sum \xi_j X_j}\big) = \widehat{P}_X(\xi).$$

Théorème 33. *La transformée de Fourier est injective sur les mesures de probabilité boréliennes de* $\mathbb{R}^d$ *:*

$$\Phi_X = \Phi_Y \Rightarrow P_X = P_Y.$$

Théorème 34. *Soit* X *v.a. à valeurs dans* $\mathbb{R}^d$. *On note* $\lambda_{\mathbb{R}^d}$ *la mesure de Lebesgue sur* $\mathbb{R}^d$. *Si* $\Phi_X \in L^1(\mathbb{R}^d, \lambda_{\mathbb{R}^d})$ *alors* $P_X \ll \lambda_{\mathbb{R}^d}$ *et sa densité est donnée par*

$$f_X(x) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} e^{-i\langle \xi, x\rangle} \Phi_X(\xi) d\xi.$$

Proposition 35. *Pour toute v.a.r.* X,

— Φ_X *est continue,*
— *Si* $E|X|^n < \infty$ *alors* Φ_X *est* n *fois dérivable et pour* $k \leq n$, $t \in \mathbb{R}$

$$\Phi_X^{(k)}(t) = i^k\, E[X^k e^{itX}].$$

En particulier on retrouve les moments par la relation $\Phi_X^{(k)}(0) = i^k\, E(X^k)$.

— *Si* Φ_X *est* n *fois dérivable en* 0 *alors* X *admet des moments jusqu'à l'ordre* $2\left\lfloor \frac{n}{2} \right\rfloor$.

3.5 Transformée de Laplace

Définition 24. Soit X une v.a. à valeur dans $\mathbb{R}^d$. Sa *transformée de Laplace* est définie pour $s \in \mathbb{R}^d$ par

$$L_X(s) := Ee^{\langle s,X\rangle} \in [0, \infty].$$

Proposition 36. *L'ensemble* $\{s \in \mathbb{R}^d;\ L_X(s) < \infty\}$ *est convexe. Pour tout* $\lambda \in [0,1]$ *et tous* $s_1, s_2 \in \mathbb{R}^d$*, on a*

$$L_X\big(\lambda s_1 + (1-\lambda)s_2\big) \le L_X(s_1)^\lambda\, L_X(s_2)^{1-\lambda}.$$

Proposition 37. *Soit* X *une v.a.r. telle que* $\forall t \in [-\epsilon, \epsilon]$, $Ee^{tX} < \infty$*. Alors*

$$\forall t \in [-\epsilon, \epsilon], \quad L_X(t) = \sum_{n\ge 0} \frac{t^n}{n!} E\big(X^n\big).$$

Théorème 38. *Soient* X, Y *deux v.a. à valeurs dans* $\mathbb{R}^d$*. Soit* $\epsilon > 0$*. Si pour tout* s *tel que* $\|s\| \le \epsilon$ *on a* $L_X(s) = L_Y(s) < +\infty$*, alors* $P_X = P_Y$*.*

Plus généralement deux mesures sont égales des que leurs tranformées de Laplace sont finies et égales sur une boule :

Théorème 39. *Soient* μ, ν *des mesures boréliennes sur* $\mathbb{R}^d$*. S'il existe* $x_0 \in \mathbb{R}^d$ *et* $\epsilon > 0$ *tels que*

$$\forall x \in B(x_0, \epsilon),\ L_\mu(x) = L_\nu(x) < \infty$$

alors $\mu = \nu$*.*

Définition 25. Soit X une v.a.r. qui admet des moments de tous ordres. On dit que sa loi est *caractérisée par ses moments* si pour toute v.a.r. Y on a

$$[\forall n \ge 0,\ EX^n = EY^n] \Rightarrow P_X = P_Y.$$

Proposition 40. *Soit* X *une v.a.r.. Sa loi est caractérisée par les moments si l'une des hypothèses suivantes (qui sont équivalentes) est vérifiée :*

— *il existe* $\varepsilon > 0$ *tel que* $Ee^{\varepsilon|X|} < +\infty$*.*

— *il existe* $C \in \mathbb{R}^+$ *tel que pour tout* $n \ge 1$, $E\big(X^{2n}\big)^{\frac{1}{2n}} \le Cn$*.*

Chapitre 4

Rappels sur la notion d'indépendance

4.1 Indépendance

Soit $(\Omega, \mathcal{A}, P)$ un espace de probabilité.

Définition 26. (Indépendance d'événements)

— Deux événements $A, B \in \mathcal{A}$ sont indépendants si $P(A \cap B) = P(A)P(B)$.

— Des événements $A_1, \dots, A_n \in \mathcal{A}$ sont indépendants si $\forall \{j_1, \dots, j_p\} \subset \{1, \dots, N\}$ on a

$$P(A_{j_1} \cap \cdots \cap A_{j_p}) = P(A_{j_1}) \times \cdots \times P(A_{j_p}).$$

Définition 27. (Indépendance de sous-tribus)

— Des sous-tribus $\mathcal{B}_1, \dots, \mathcal{B}_n$ de $\mathcal{A}$ sont mutuellement indépendantes si

$$\forall C_i \in \mathcal{B}_i, \quad P(C_1 \cap \cdots \cap C_n) = P(C_1) \times \cdots \times P(C_n).$$

— Des sous-tribus de $\mathcal{A}$, $(\mathcal{B}_i)_{i \in I}$ sont mutuellement indépendantes si toute sous-famille finie l'est.

L'indépendance de tribus se vérifie sur des parties génératrices stables par intersection :

Proposition 41. *Soient $\mathcal{B}_1, \dots, \mathcal{B}_n$ des sous-tribus de $\mathcal{A}$. Pour chaque i soit $\mathcal{F}_i \subset \mathcal{B}_i$ avec*

— *$\Omega \in \mathcal{F}_i$, $\sigma(\mathcal{F}_i) = \mathcal{B}_i$*

— *$\mathcal{F}_i$ stable par intersection finie.*

Si $\forall F_i \in \mathcal{F}_i$ on a $P(F_1 \cap \cdots \cap F_n) = \prod_{i=1}^n P(F_i)$ alors les $(\mathcal{B}_i)_{i=1}^n$ sont indépendantes.

Proposition 42. (Regroupement par paquets) *Soit $(\mathcal{B}_i)_{i \in I}$ une famille de sous-tribus de $\mathcal{A}$, indépendantes. Soit une partition de $I, I = \cup_{j \in J} I_j$ et pour $j \in J$, $\widetilde{\mathcal{B}}_j = \sigma(\mathcal{B}_i, i \in I_j)$. Alors les tribus $(\widetilde{\mathcal{B}}_j)_{j \in J}$ sont indépendantes.*

Définition 28. (Indépendance de variables aléatoires)

Pour $i \in I$, soit $X_i : (\Omega, \mathcal{A}, P) \to (E_i, \mathcal{E}_i)$ une v.a.. Les $(X_i)_{i \in I}$ sont indépendantes si les tribus $\big(\sigma(X_i)\big)_{i \in I}$ le sont. En particulier $X_1, \dots, X_n$ sont indépendants si et seulement si $\forall F_i \in \mathcal{E}_i$,

$$P\Big((X_1 \in F_1) \cap \cdots \cap (X_n \in F_n)\Big) = \prod_{i=1}^n P(X_i \in F_i).$$

Théorème 43. *Des v.a. $X_1, \dots, X_n$ sont indépendants si et seulement si*

$$P_{(X_1,\dots,X_n)} = P_{X_1} \otimes \cdots \otimes P_{X_n}.$$

Dans ce cas :

— $\forall f_i$ *mesurables positives :* $E\big(\prod_{i=1}^n f_i(X_i)\big) = \prod_{i=1}^n Ef_i(X_i)$.

— *Si f_i mesurable de signe quelconque, l'égalité précédente est vérifiée si*

$$E\Big(\prod_{i=1}^n |f_i(X_i)|\Big) = \prod_{i=1}^n E|f_i(X_i)| < \infty.$$

Proposition 44. *Soient $X_1, \dots, X_n$ des variables aléatoires réelles. Les assertions suivantes sont équivalentes :*

— *$X_1, \dots, X_n$ sont indépendantes.*

— $\forall \xi \in \mathbb{R}^n,\ \Phi_{(X_1,\dots,X_n)}(\xi_1, \dots, \xi_n) = \prod_{i=1}^n \Phi_{X_i}(\xi_i)$.

— $\forall a_1, \dots, a_n \in \mathbb{R},\ P(X_1 \leq a_1, \dots, X_n \leq a_n) = \prod_{j=1}^n P(X_j \leq a_j)$.

— $\forall f_i : \mathbb{R} \to \mathbb{R}^+,\ E\Big(\prod_{i=1}^n f_i(X_i)\Big) = \prod_{i=1}^n Ef_i(X_i)$.

4.2 Sommes de variables aléatoire indépendantes

Définition 29. Des v.a.r. $X, Y \in L^2(\Omega, \mathcal{A}, P)$ sont non corrélées si

$$\mathrm{Cov}(X,Y) = E(XY) - E(X)E(Y) = E[(X - E(X))(Y - E(Y))] = 0$$

Remarque 8. Des variables indépendantes sont non corrélées.

Proposition 45. *Soit $X_1, \dots, X_n$ des v.a.r. dans $L^2(\Omega, \mathcal{A}, P)$ indépendantes (non corrélées deux à deux suffit), alors*

$$\mathrm{Var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \mathrm{Var}(X_i).$$

Loi de la somme de v.a. indépendantes :

Définition 30. Soient μ, ν deux probabilités sur $\mathbb{R}^d$. Leur produit de convolution $\mu * \nu$ est défini comme la mesure image de $\mu * \nu$ par $(y,z) \mapsto y + z$. En d'autres termes

$$\forall A \in \mathcal{B}(\mathbb{R}^d),\ \mu * \nu(A) = \mu \otimes \nu\Big(\big\{(y,z), y + z \in A\big\}\Big)$$

$$\forall \varphi \geq 0,\ \int \varphi(x)\, d(\mu * \nu)(x) = \int\int \varphi(y+z)\, d\mu(y)\, d\nu(z)$$

Exemple 9. Si $d\mu(x) = p(x)\,dx$ et $d\nu(y) = q(x)\,dx$ alors $\mu * \nu$ admet pour densité par rapport à la mesure de Lebesgue la fonction $p * q(x) = \int p(x - z)\, q(z)\, dz$.

Proposition 46. *Soient X, Y des v.a. indépendantes à valeurs $\mathbb{R}^d$ alors*

$$\begin{aligned} P_{X+Y} &= P_X * P_Y, \\ \Phi_{X+Y} &= \Phi_X\, \Phi_Y, \\ L_{X+Y} &= L_X\, L_Y. \end{aligned}$$

Si X, Y sont à valeurs dans $\mathbb{N}$, alors $G_{X+Y} = G_X\, G_Y$.

Exemple 10. **(Lois classiques dont les convolées sont très simples)**

— Lois binomiales : pour $n \in \mathbb{N}^*$ et $p \in (0,1)$,

$$\mathcal{B}(n,p) := \sum_{k=0}^{n} C_n^k p^k (1-p)^{n-k} \delta_k$$

On a $\mathcal{B}(n,p) * \mathcal{B}(n',p') = \mathcal{B}(n+n',p)$. En d'autres termes, si $X \sim \mathcal{B}(n,p)$ et $Y \sim \mathcal{B}(n',p')$ sont indépendantes alors :

$$X + Y \sim \mathcal{B}(n+n',p).$$

On appelle loi de Bernoulli avec probabilité de succès $p \in [0,1]$, la loi

$$b(p) := \mathcal{B}(1,p) = (1-p)\delta_0 + p\delta_1.$$

— Lois binomiales négatives : $\mathcal{B}(-m,p)$ avec $p \in (0,1)$ et $m \in \mathbb{N}^*$, définies par

$$\mathcal{B}(-m,p) := \sum_{k \geq 0} C_{m+k-1}^{m-1} \, p^m (1-p)^k \, \delta_k.$$

Pour $m, n \in \mathbb{N}^*$, on a $\mathcal{B}(-m,p) * \mathcal{B}(-n,p) = \mathcal{B}\big(-(m+n),p\big)$. On appelle loi géométrique $\mathcal{G}(p) := \mathcal{B}(-1,p) = \sum_{k \geq 0} p(1-p)^k \, \delta_k$.

— Loi de Poisson : $\mathcal{P}(\lambda) := \sum_{k \geq 0} e^{-\lambda} \frac{\lambda^k}{k!} \delta_k$ avec $\lambda > 0$. Elle vérifie

$$\mathcal{P}(\lambda) * \mathcal{P}(\lambda') = \mathcal{P}(\lambda + \lambda').$$

— Lois $\gamma(t,a)$ avec $t > 0$ et $a > 0$ de densité $\frac{a^t}{\Gamma(t)} x^{t-1} e^{-ax} 1_{x>0}$ par rapport à la mesure de Lebesgue sur $\mathbb{R}$. On rappelle que pour $t > 0$, $\Gamma(t) = \int_0^{+\infty} e^{-x} x^{t-1} dx$. Si $s, t > 0$,

$$\gamma(t,a) * \gamma(s,a) = \gamma(t+s,a).$$

On appelle loi exponentielle de paramètre $a > 0$ la mesure de probabilité $\mathcal{E}xp(a) := \gamma(1,a)$.

— Lois gaussiennes : $\mathcal{N}(m,\sigma^2), m \in \mathbb{R}, \sigma \geq 0$. Elles sont définies par $\mathcal{N}(m,0) := \delta_m$ et pour $\sigma > 0$

$$\mathcal{N}(m,\sigma^2)(dx) := e^{\frac{-(x-m)^2}{2\sigma^2}} \frac{dx}{\sigma\sqrt{2\pi}}, \; x \in \mathbb{R}.$$

On a $\mathcal{N}(m,\sigma^2) * \mathcal{N}(m',\sigma'^2) = \mathcal{N}(m+m',\sigma^2+\sigma'^2)$.

4.3 Suites de tribus ou de v.a. indépendantes

Théorème 47. *Soit $(\mathcal{T}_n)_{n \in \mathbb{N}}$ une suite de sous-tribus indépendantes sur $(\Omega, \mathcal{A}, P)$. On pose $\mathcal{B}_k := \sigma(\mathcal{T}_n, n \geq k)$ et l'on définit la tribu asymptotique ou terminale $\mathcal{B}_\infty := \cap_{k \geq 1} \mathcal{B}_k$. Alors $\forall B \in \mathcal{B}_\infty$, $P(B) \in \{0,1\}$.*

Définition 31. Soit $(A_n)_{n \in \mathbb{N}}$ une suite d'événements de $(\Omega, \mathcal{A})$,

$$\limsup_n A_n := \bigcap_{n \in \mathbb{N}} \Big(\bigcup_{k \geq n} A_k \Big) = \{\omega \in \Omega; \text{ pour une infinité de valeurs de } n, \, \omega \in A_n\},$$

$$\liminf_n A_n := \bigcup_{n \in \mathbb{N}} \Big(\bigcap_{k \geq n} A_k \Big) = \{\omega \in \Omega; \, \exists n_0, \, \forall n \geq n_0, \, \omega \in A_n\}.$$

Exemple 11. Soit $(A_n)_{n\in\mathbb{N}}$ une suite d'événements indépendants de $(\Omega, \mathcal{A}, P)$. Les tribus $\mathcal{T}_n = \{\emptyset, A_n, A_n^c, \Omega\}$ sont indépendantes. Donc

$$\begin{aligned} P\left(\liminf A_n\right) &\in \{0,1\}. \\ P\left(\limsup A_n\right) &\in \{0,1\}. \end{aligned}$$

Lemme 48 (Borel-Cantelli)**.** *Soient* $(A_n)_{n\in\mathbb{N}}$ *avec* $A_i \in \mathcal{A}$.

1. *Si* $\sum_{n\geq 0} P(A_n) < \infty$ *alors* $P(\limsup A_n) = 0$ *i.e. p.s. card* $\{n, \omega \in A_n\} < \infty$.
2. *Si les* A_n *sont indépendants et si* $\sum P(A_n) = +\infty$ *alors* $P(\limsup A_n) = 1$ *i.e. p.s.* A_n *a lieu une infinité de fois.*

Chapitre 5

Calcul conditionnel

5.1 Conditionnement discret

5.1.1 Par rapport à un événement

L'idée de conditionnement est réévaluer les probabilités d'événements en tenant compte d'une information supplémentaire.

Définition-Proposition 32. Soit $(\Omega, \mathcal{A}, P)$ un espace de probabilités et $B \in \mathcal{A}$ avec $P(B) > 0$.

1. La *probabilité conditionnelle de* $A \in \mathcal{A}$ *sachant* B est
$$P(A|B) := \frac{P(A \cap B)}{P(B)}\cdot$$
2. La *mesure de probabilité conditionnelle sachant* B est l'application
$$\begin{array}{ccc} \mathcal{A} & \to & [0,1] \\ A & \mapsto & \dfrac{P(A \cap B)}{P(B)} = P(A|B). \end{array}$$
On note la $P(\cdot|B)$.
3. Si Y est une v.a.r. sur $(\Omega, \mathcal{A}, P)$ positive ou intégrable, on définit *son espérance conditionnelle sachant* B par
$$E(Y|B) := \frac{E(Y 1_B)}{P(B)} = \frac{1}{P(B)} \int_B Y\, dP = \int_\Omega Y\, dP(\cdot|B).$$
4. La loi de Y sachant B, notée $P_{Y|B}$ est définie comme l'image par Y de $P(\cdot|B)$, $P_{Y|B}(A) := P(Y \in A|B)$. Si φ est mesurable ≥ 0 ou si $\varphi(Y)$ est P-intégrable, on a
$$E\big(\varphi(Y)|B\big) = \int \varphi(Y)\, dP(\cdot|B) = \int \varphi\, dP_{Y|B}.$$

5.1.2 Par rapport à une v.a. discrète

On suppose que l'on a accès aux valeurs d'une v.a. et on réactualise suivant cet aléa.

Définition 33. Soit X une v.a. sur $(\Omega, \mathcal{A}, P)$ à valeur dans E dénombrable.

Soit $E' = \{x \in E, P(X = x) > 0\}$, on a $P(X \in E \setminus E') = 0$. Pour $A \in \mathcal{A}$, on pose
$$\varphi(x) := \begin{cases} P(A|X = x) = \frac{P\left(A \cap \{X = x\}\right)}{P(X = x)} & \text{si} \quad x \in E' \\ 0 & \text{si} \quad x \in E \setminus E' \end{cases}$$
et on définit $P(A|X) := \varphi(X)$. C'est une v.a. qui dépend de X.

Définition 34. Si Y est une v.a.r. avec $Y \in L^1(\Omega, \mathcal{A}, P)$ ou $Y \geq 0$, on définit

$$\Psi(x) = \begin{cases} E(Y|X=x) & \text{si} \quad P(X=x) > 0 \\ 0 & \text{si} \quad P(X=x) = 0 \end{cases}$$

Par définition l'*espérance conditionnelle de Y sachant X* est la v.a. $\Psi(X)$.

Lemme 49 (Doob)**.** *Soit $X : (\Omega, \mathcal{A}, P) \to (E, \mathcal{E})$ et $Y : \Omega \to \mathbb{R}^d$ deux v.a.. Alors Y est $\sigma(X)$-mesurable si et seulement s'il existe une fonction $h : E \to \mathbb{R}^d$ borélienne avec $Y = h(X)$.*

Proposition 50. *Soit X une v.a. discrète et $Y \in L^1$, alors*

1. $E\big(|E(Y|X)|\big) \leq E\big(|Y|\big)$
2. *Pour toute v.a. Z, $\sigma(X)-$mesurable et bornée*

$$E\big(ZE(Y|X)\big) = E(ZY).$$

Cette dernière condition est semblable à celle qui caractérise la projection orthogonale sur un sous-espace d'un espace hilbertien.

Théorème 51 (Projection dans un Hilbert)**.** *Soit $\big(H, \langle\cdot,\cdot\rangle, |\cdot|\big)$ un espace de Hilbert et S un sous-espace vectoriel fermé de H. Alors pour tout $y \in H$, il existe un unique $p(y) \in S$ tel que*

$$|y - p(y)| = \inf_{u \in S} |y - u|.$$

De plus il vérifie $\forall u \in S, \langle y - p(y), u\rangle = 0$ et cette propriété le caractérise. On dit que $p(y)$ est la projection orthogonale de y sur S.

5.2 Espérance conditionnelle d'une v.a. par rapport à une tribu

5.2.1 Définition pour des variables aléatoires dans L^2

Soient $(\Omega, \mathcal{A}, P)$ un espace de probabilité et $\mathcal{B}$ une sous tribu de $\mathcal{A}$, alors $L^2(\Omega, \mathcal{B}, P)$ est un sous espace fermé de $L^2(\Omega, \mathcal{A}, P)$.

Définition 35. Soit $Y \in L^2(\Omega, \mathcal{A}, P)$. L'*espérance conditionnelle de Y sachant $\mathcal{B}$*, notée $E(Y|\mathcal{B})$ est la projection orthogonale de Y sur $L^2(\Omega, \mathcal{B}, P)$. Elle est caractérisée par :

1. $E(Y|\mathcal{B}) \in L^2(\Omega, \mathcal{B}, P)$
2. $\forall Z \in L^2(\Omega, \mathcal{B}, P), E(ZY) = E\big(ZE(Y|\mathcal{B})\big)$.

Si $\mathcal{B} = \sigma(X)$ on écrit $E(Y|\mathcal{B}) = E(Y|X)$.

Remarque 12. $E[Y|\mathcal{B}]$ est un élément de $L^2(\Omega, \mathcal{A}, P)$ donc, en toute rigueur, défini modulo égalité p.s. Il faut s'en souvenir lorsque l'on écrit des propriétés ponctuelles.

Proposition 52. *Soit $Y \in L^2(\Omega, \mathcal{A}, P)$.*

1. *Si Y est $\mathcal{B}$-mesurable alors $E(Y|\mathcal{B}) = Y$ (p.s.)*
2. $Y \mapsto E(Y|\mathcal{B})$ *est linéaire*
3. *Si $Y \geq 0$ alors $E(Y|\mathcal{B}) \geq 0$ p.s.*
   Si $Y' \geq Y$ alors $E(Y|\mathcal{B}) \geq E(Y'|\mathcal{B})$ p.s.

5.2.2 Extension aux variables aléatoires positives

Soit $\mathcal{B}$ une sous-tribu de $\mathcal{A}$.

Théorème 53 (définition)**.** *Soit $Y : (\Omega, \mathcal{A}, P) \to [0, +\infty]$ une v.a. positive. Il existe une v.a. notée $E(Y|\mathcal{B}) \geq 0$ p.s., unique p.s. telle que :*

1. *$E(Y|\mathcal{B})$ est $\mathcal{B}$-mesurable,*
2. *Pour toute variable Z, $\mathcal{B}$-mesurable et positive, $E(ZY) = E\big(ZE(Y|\mathcal{B})\big)$.*

Proposition 54. *Soit Y une v.a. ≥ 0*

— *Si Y est $\mathcal{B}$-mesurable alors $E(Y|\mathcal{B}) = Y$ p.s.*
— *Si $Y \leq Y'$ alors $E(Y|\mathcal{B}) \leq E(Y'|\mathcal{B})$ p.s.*

5.2.3 Définition pour des variables aléatoires dans L^1

Définition-Proposition 36. Soit Y une v.a.r. de $L^1(\Omega, \mathcal{A}, P)$. On pose $E(Y|\mathcal{B}) := E(Y_+|\mathcal{B}) - E(Y_-|\mathcal{B})$. Cette définition a un sens dans L^1 car

$$E\big|E(Y|\mathcal{B})\big| \leq E\big(|Y|\big).$$

De plus $E(Y|\mathcal{B})$ est caractérisée par les propriétées suivantes :

1. $E(Y|\mathcal{B})$ est $\mathcal{B}$-mesurable,
2. Pour toute variable Z, $\mathcal{B}$-mesurable et bornée, $E\big(Z\ E(Y|\mathcal{B})\big) = E(ZY)$.

5.2.4 Propriétés de l'espérance conditionnelle

Proposition 55. *(Sommes et limites)*

1. *Si $X, Y \geq 0$ et $a, b \geq 0$ (resp. $X, Y \in L^1$ et $a, b \in \mathbb{R}$) alors*
$$E(aX + bY|\mathcal{B}) = aE(X|\mathcal{B}) + bE(Y|\mathcal{B}) \quad p.s.$$
2. *Si $(X_n)_{n\geq 0}$ est une suite croissante de v.a.r. positives telle que $X_n \underset{n\to\infty}{\nearrow} X$ p.s. alors*
$$E(X_n|\mathcal{B}) \underset{n\to\infty}{\nearrow} E(X|\mathcal{B}) \quad p.s.$$
3. *Si $(X_n)_{n\geq 0}$ est une suite de variables positives, alors*
$$E(\liminf X_n|\mathcal{B}) \leq \liminf E(X_n|\mathcal{B}) \quad p.s.$$
4. *Soit $(X_n)_{n\geq 0}$ une suite de v.a.r. dominées (il existe Y avec $|X_n| \leq Y$ et $EY < +\infty$) et telle que $X_n \xrightarrow[n\to\infty]{p.s.} X$ alors*
$$E(X_n|\mathcal{B}) \xrightarrow[n\to\infty]{p.s., L^1} E(X|\mathcal{B}).$$
5. *Si $X \in L^1$ et $f : \mathbb{R} \to \mathbb{R}^+$ est convexe alors $E\big(f(X)|\mathcal{B}\big) \geq f\big(E(X|\mathcal{B})\big)$ p.s.*

Proposition 56. *Soient X, Y des v.a.r. définies sur $(\Omega, \mathcal{A}, P)$ et $\mathcal{C} \subset \mathcal{B}$ des sous-tribus de $\mathcal{A}$.*

1. *Si Y est $\mathcal{B}$-mesurable alors l'égalité*
$$E(YX|\mathcal{B}) = YE(X|\mathcal{B}) \quad p.s.$$
est valable dès que $X \geq 0,\ Y \geq 0$ ou $X,\ XY \in L^1$.

2. Si $X \geq 0$ ou $X \in L^1$, alors p.s.

$$\begin{aligned} E\big(E(X|\mathcal{B})|\mathcal{C}\big) &= E(X|\mathcal{C}), \\ E\big(E(X|\mathcal{C})|\mathcal{B}\big) &= E(X|\mathcal{C}). \end{aligned}$$

Proposition 57. *Soient $\mathcal{B}, \mathcal{C}$ deux sous-tribus de $\mathcal{A}$. Les assertions suivantes sont équivalentes :*

1. *$\mathcal{B}, \mathcal{C}$ sont indépendantes*
2. *$\forall Z \geq 0$, $\mathcal{C}$-mesurable, $E(Z|\mathcal{B}) = E(Z)$.*

5.3 Loi conditionnelle sachant une variable aléatoire

Définition 37. Soient $(E, \mathcal{E})$ et $(F, \mathcal{F})$ deux espaces mesurables. Un *noyau de transition* (ou *probabilité de transition*) de $(E, \mathcal{E})$ dans $(F, \mathcal{F})$ est une application $K : E \times \mathcal{F} \to [0,1]$ telle que

1. $\forall x \in E, A \mapsto K(x, A)$ est une probabilité sur $(F, \mathcal{F})$.
2. $\forall A \in \mathcal{F}, x \mapsto K(x, A)$ est $\mathcal{E}$-mesurable.

Proposition 58. *Si $h \geq 0$ est mesurable sur $(F, \mathcal{F})$ alors $x \mapsto \int h(y)K(x, dy)$ est mesurable.*

Définition 38. Soient $X : (\Omega, \mathcal{A}, P) \to (E, \mathcal{E})$ et $Y : (\Omega, \mathcal{A}, P) \to (F, \mathcal{F})$ deux v.a. On appelle *(version de) la loi conditionnelle de Y sachant X* toute probabilité de transition $K(x, dy)$ de $(E, \mathcal{E})$ dans $(F, \mathcal{F})$ telle que pour toute fonction $h : (F, \mathcal{F}) \to \mathbb{R}^+$ mesurable

$$E\big(h(Y)|X\big) = \int h(y)K(X, dy) \qquad p.s.$$

Pour $h = 1_A$,

$$P(Y \in A|X) = E\big(\mathbb{1}_A(Y)|X\big) = K(X, A) \qquad p.s.$$

Remarque 13. On dira souvent que $K(x, \cdot)$ est la loi conditionnelle de Y sachant que $X = x$. Cette appellation très intuitive est cependant un peu trompeuse. En effet $K(x, \cdot)$ est uniquement définie P_X-presque sûrement en x, donc cette notion n'a de sens que globalement et pas pour un seul x. De plus l'événement $\{X = x\}$ est souvent négligeable et il est délicat de définir le conditionnement qui lui est associé.

Théorème 59. *Soient E et F deux espaces métriques complets et séparables (i.e. contenant une suite dense). Soient $X : (\Omega, \mathcal{A}, P) \to \big(E, \mathcal{B}(E)\big)$ et $Y : (\Omega, \mathcal{A}, P) \to \big(F, \mathcal{B}(F)\big)$ deux v.a., alors il existe une loi conditionnelle de Y sachant X. C'est le cas si X et Y sont à valeur dans $\big(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d)\big)$.*

Proposition 60. *Soient $X : (\Omega, \mathcal{A}, P) \to (E, \mathcal{E})$* et *$Y : (\Omega, \mathcal{A}, P) \to (F, \mathcal{F})$ deux v.a. Si K est la loi conditionnelle de Y sachant X, alors pour toute fonction mesurable positive sur $(E \times F, \mathcal{E} \otimes \mathcal{F})$ on a*

$$E\big(h(X,Y)|X\big) = \int_F h(X, y)\, K(X, dy) \quad p.s.,$$

ainsi que la formule de désintégration suivante :

$$E\big(h(X,Y)\big) = \int_E \left(\int_F h(x,y)K(x,dy)\right) dP_X(x).$$

Chapitre 6

Vecteurs gaussiens

6.1 Variables gaussiennes

Définition 39. Pour $m \in \mathbb{R}$, $\sigma \geq 0$ on définit la loi *gaussienne* (ou *normale*) $\mathcal{N}(m, \sigma^2)$ qui est une probabilité sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ par $\mathcal{N}(m, 0) = \delta_m$ et pour $\sigma > 0$,

$$\mathcal{N}(m, \sigma^2)(dx) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}} dx.$$

$\mathcal{N}(0, 1)$ est appelée loi gaussienne *standard* ou *centrée réduite*. Une v.a.r. X est *gaussienne* s'il existe $(m, \sigma) \in \mathbb{R} \times \mathbb{R}^+$ telle que $P_X = \mathcal{N}(m, \sigma^2)$, elle est centrée si $m = 0$.

Proposition 61. *Soit $X \sim \mathcal{N}(m, \sigma^2)$ alors*
- $EX = m, \qquad \mathrm{Var}(X) = \sigma^2$,
- *pour tout* $\xi \in \mathbb{R}$, $\Phi_X(\xi) = \exp\left(im\xi - \frac{\sigma^2\xi^2}{2}\right)$.

Si $Y \sim \mathcal{N}(m', \sigma'^2)$ est indépendante de X alors $X + Y \sim \mathcal{N}(m + m', \sigma^2 + \sigma'^2)$.

6.2 Vecteurs gaussiens de $\mathbb{R}^d$

Définition 40. Un vecteur aléatoire $X = (X_1, \ldots, X_d)$ est dit gaussien si pour tout $\alpha \in \mathbb{R}^d$, la v.a.r. $\langle \alpha, X \rangle$ est gaussienne. On note son espérance $m = EX = (EX_1, \ldots, EX_d)$ et sa matrice de covariance

$$\Gamma = \mathrm{Cov}(X) = \left(\mathrm{Cov}(X_i, X_j)\right)_{1 \leq i,j \leq d},$$

c'est une matrice symétrique positive.

Remarque 14. On pourrait définir de manière analogue la notion de vecteur gaussien à valeurs dans un espace euclidien. Comme tout espace euclidien de dimension d s'identifie à $\mathbb{R}^d$ muni de sa structure canonique, ce point de vue ne donne pas de nouveaux objet mais seulement des notations plus intrinsèques.

Proposition 62 (Unicité)**.** *La loi d'un vecteur gaussien X dans $\mathbb{R}^d$ est caractérisée par sa moyenne m et sa matrice de covariance Γ. Plus précisément on a*

$$\Phi_X(\alpha) = \exp\left(i\langle \alpha, m \rangle - \frac{\langle \alpha, \Gamma\alpha \rangle}{2}\right), \quad \forall \alpha \in \mathbb{R}^d.$$

On note $P_X = \mathcal{N}_d(m, \Gamma)$.

Proposition 63. *Si $X \sim \mathcal{N}_d(m, \Sigma)$, $b \in \mathbb{R}^k$ et $A \in M_{k,d}(\mathbb{R})$ alors $AX + b$ est un vecteur gaussien à valeurs dans $\mathbb{R}^k$, de loi $\mathcal{N}_k(Am + b, A\Sigma\, {}^tA)$.*

Proposition 64 (Existence)**.** *Soient $G_1, \dots, G_d$ des v.a.r. indépendantes de loi $\mathcal{N}(0,1)$.*

- *Le vecteur aléatoire $G = (G_1, \dots, G_d)$ est gaussien de loi $\mathcal{N}_d(0, I_d)$, où I_d est la matrice identité de taille d.*
- *Si $m \in \mathbb{R}^d$ et $\Gamma \in M_d(\mathbb{R})$ est une matrice symétrique positive, alors $m + \sqrt{\Gamma}\, G \sim \mathcal{N}_d(m, \Gamma)$.*

Proposition 65. *Si $m \in \mathbb{R}^d, \Gamma$ est une matrice $d \times d$ et symétrique positive et inversible (i.e. définie positive) alors*

$$\mathcal{N}_d(m,\Gamma)(dx) = \exp\Big(-\frac{1}{2}\big\langle x-m, \Gamma^{-1}(x-m)\big\rangle\Big)\frac{dx}{(2\pi)^{\frac{d}{2}}\sqrt{\det\Gamma}}.$$

Théorème 66. *Soit $X \sim \mathcal{N}_d(m,\Gamma)$. Soient $\alpha_1 \geq \dots \geq \alpha_d$ les valeurs propres (avec répétition) de Γ et soit $(V_1, \dots, V_d)$ une base orthonormée formée de vecteurs propres associés. Soient $G_1, \dots, G_d$ des variables indépendantes de loi $\mathcal{N}(0,1)$. Alors X a même loi que*

$$m + \sum_{i=1}^d \sqrt{\alpha_i}\, G_i\, V_i.$$

6.3 Indépendance, conditionnement

L'indépendance des composantes d'un vecteur gaussien se lit sur les covariances.

Proposition 67. *Soit $(X_1, \dots, X_n, Y)$ un vecteur gaussien de $\mathbb{R}^{n+1}$. Les assertions suivantes sont équivalentes :*

1. *Y est indépendante de $(X_1, \dots, X_n)$.*
2. *$\forall j \in \{1, \dots, n\}$, $\mathrm{Cov}(Y, X_i) = 0$.*

Plus généralement si $(X_1, \dots, X_n)$ vecteur gaussien dont la matrice $\mathrm{Cov}(X)$ est diagonale par blocs de tailles $n_1, \dots, n_k$ avec $n_1 + \ldots + n_k = n$, alors les vecteurs correspondant aux blocs

$$(X_1, \dots, X_{n_1}),\ (X_{n_1+1}, \dots, X_{n_1+n_2}), \dots, (X_{n_1+\ldots+n_{k-1}+1}, \dots, X_n)$$

sont mutuellement indépendants.

6.3.1 Calculs d'espérance conditionnelle

Proposition 68. *Soit $(X_1, \dots, X_n, Y)$ un vecteur gaussien centré de $\mathbb{R}^{n+1}$, défini sur un espace de probabilité $(\Omega, \mathcal{A}, P)$.*

1. *Alors $E(Y|X_1, \dots, X_n)$ est la projection orthogonale de Y sur $\mathrm{vect}(X_1, \dots, X_n)$ au sens de $L^2(\Omega, \mathcal{A}, P)$. Donc, il existe des nombres réels $\lambda_1, \dots, \lambda_n$ tels que*

   $$E(Y|X_1, \dots, X_n) = \sum_{j=1}^n \lambda_j X_j.$$

   Ils sont caractérisés par les équations :

   $$E(YY_i) = \sum_{j=1}^n \lambda_j E(Y_j Y_i), \quad i = 1, \dots, n.$$

2. *Soit $\sigma^2 = E\Big(Y - \sum_{j=1}^n \lambda_j X_j\Big)^2 \in \mathbb{R}^+$. La loi conditionnelle de Y sachant $X_1, \dots, X_n$ est donnée par*

   $$K\big((x_1, \dots, x_n), \cdot\big) = \mathcal{N}\Big(\sum_{j=1}^n \lambda_j x_j, \sigma^2\Big).$$

6.3.2 Lois associées

Définition 41. La loi du *Chi-deux* à n degrés de liberté est $\chi^2(n) := \gamma\left(\frac{n}{2}, \frac{1}{2}\right)$. C'est la loi de $|G|^2 = \sum_{i=1}^n G_i^2$ si $G \sim \mathcal{N}_n(0, I_n)$.

Théorème 69 (Cochran)**.** *Soit $Y \sim \mathcal{N}_n(0, I_n)$ et H un sous-espace vectoriel de $\mathbb{R}^n$ de dimension $r \leq n$, alors le vecteur gaussien $X = P_H^{\perp}(Y)$,*

- *vu comme vecteur aléatoire à valeurs dans $\mathbb{R}^n$, suit à la loi $\mathcal{N}_n\big(0, P_H^{\perp}\big)$,*
- *vu comme vecteur aléatoire à valeurs dans $H \approx \mathbb{R}^r$, suit $\mathcal{N}_r(0, I_r)$.*

De plus $|X|^2$ suit la loi $\chi^2(r)$.

Définition 42. Pour $\nu > 0$ la loi de Student $s(\nu)$ de paramètre ν est la loi à densité sur $\mathbb{R}^+$ donnée par

$$s(\nu)(dt) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\,\Gamma(\frac{\nu}{2})} \left(1 + \frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}} dt.$$

Théorème 70. *Soit $X = (X_1, \ldots, X_n)$ où les X_i sont des v.a.r. indépendantes et de loi $\mathcal{N}(m, \sigma^2)$. On considére les variables*

$$\overline{X}_n := \frac{X_1 + \cdots + X_n}{n} \quad \text{et} \quad S_n := \left(\frac{1}{n-1} \sum_{k=1}^{n} \left(X_n - \overline{X}_n\right)^2\right)^{\frac{1}{2}},$$

appelées respectivement moyenne empirique et variance empirique. Alors

$$\left(\sqrt{n}\,\frac{\overline{X_n} - m}{\sigma},\ \frac{n-1}{\sigma^2}\,S_n^2\right) \quad \text{suit la loi} \quad \mathcal{N}(0,1) \otimes \chi^2(n-1).$$

En particulier moyenne empirique et variance empirique sont indépendantes. De plus le quotient

$$\sqrt{n}\,\frac{\overline{X_n} - m}{S_n}$$

suit la loi de Student $s(n-1)$.

Chapitre 7

Convergence des suites de v.a. et de lois

7.1 Convergence presque sure

Définition 43. Soient $X, (X_n)_{n\geq 0}$ des variables aléatoires définies sur $(\Omega, \mathcal{A}, P)$ et à valeurs dans $\big(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d)\big)$. On dit que la suite $(X_n)_{n\geq 0}$ tend vers X *presque sûrement* et on note $X_n \xrightarrow{p.s.} X$ si

$$P\Big(\big\{\omega \in \Omega;\ X(\omega) = \lim_{n\to\infty} X_n(\omega)\big\}\Big) = 1.$$

Proposition 71. *(Critère de Borel-Cantelli).*

1. *Si* $\forall \epsilon > 0,\ \sum_{n\in\mathbb{N}} P\big(|X_n - X| \geq \varepsilon\big) < \infty$ *alors* $X_n \xrightarrow{p.s.} X$.
2. *Si les* X_n *sont indépendantes et* α *est un nombre réel fixé,*

$$X_n \xrightarrow{p.s.} \alpha \quad \Leftrightarrow \quad \forall \epsilon > 0, \sum_{n\in\mathbb{N}} P\big(|X_n - \alpha| \geq \varepsilon\big) < \infty.$$

Proposition 72. *Soient* $(X_n)_{n\geq 0}, (Y_n)_{n\geq 0}, X, Y$ *des v.a.r. définies sur* $(\Omega, \mathcal{A}, P)$. *Si* $X_n \xrightarrow{p.s.} X$ *et* $Y_n \xrightarrow{p.s.} Y$ *alors* $X_n + Y_n \xrightarrow{p.s.} X + Y$ *et* $X_n Y_n \xrightarrow{p.s.} XY$.

7.2 Convergence en probabilité

Définition 44. Soient $X, (X_n)_{n\geq 0}$ des variables aléatoires définies sur $(\Omega, \mathcal{A}, P)$ et à valeurs dans $\big(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d)\big)$. On dit que la suite $(X_n)_{n\geq 0}$ tend vers X *en probabilité* et on note $X_n \xrightarrow{P} X$ si

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

Proposition 73. *(Lien avec la convergence p.s.)*

1. $X_n \xrightarrow{p.s.} X \quad \Rightarrow \quad X_n \xrightarrow{P} X$.
2. $X_n \xrightarrow{P} X \quad \Rightarrow \quad$ *il existe une sous-suite telle que* $X_{n_k} \xrightarrow{p.s.} X$.

Définition-Proposition 45. Soit $L^0_{\mathbb{R}^d}(\Omega, \mathcal{A}, P)$ l'ensemble des v.a de $(\Omega, \mathcal{A}, P)$ dans $\big(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d)\big)$ quotienté par la relation d'égalité p.s.. L'application $d(X, Y) := E \min\big(1, |X - Y|\,\big)$ est une distance sur cet espace et

$$X_n \xrightarrow{P} X \quad \Leftrightarrow \quad \lim_{n\to\infty} d(X_n, X) = 0.$$

On dit que la convergence en probabilité est *métrisable.*

Proposition 74. *Les assertions suivantes sont équivalentes*
- $X_n \xrightarrow{P} X$,
- *de toute sous-suite* $(X_{n_k})_{k\geq 0}$ *on peut extraire une sous-suite* $\big(X_{n_{k_j}}\big)_{j\geq 0}$ *telle que* $X_{n_{k_j}} \xrightarrow[j\to\infty]{p.s.} X$.

Corollaire 75. *Soient* (X_n) *et* (Y_n) *deux suites de v.a.r. sur* $(\Omega, \mathcal{A}, P)$ *telles que* $X_n \xrightarrow{P} X$ *et* $Y_n \xrightarrow{P} Y$.
- *Si* ϕ *est continue alors* $\phi(X_n) \xrightarrow{P} \phi(X)$,
- *Si* $\alpha, \beta \in \mathbb{R}$ *alors* $\alpha X_n + \beta Y_n \xrightarrow{P} \alpha X + \beta Y$,
- $X_n Y_n \xrightarrow{P} XY$.

Théorème 76. *L'espace* $\big(L^0_{\mathbb{R}^d}(\Omega, \mathcal{A}, P), d\big)$ *est complet.*

7.3 Convergence dans $L^p, p \geq 1$

Définition 46. Soient X et $(X_n)_{n\geq 0}$ des v.a.r. dans $L^p(\Omega, \mathcal{A}, P)$. On dit que la suite (X_n) tend vers X au sens L^p et on note $X_n \xrightarrow{L^p} X$ si $\lim_{n\to\infty} \|X - X_n\|_p = 0$, c'est-à-dire si

$$\lim_{n\to\infty} E\big(|X_n - X|^p\big) = 0.$$

Proposition 77. *Si* $q \geq p \geq 1$,

$$X_n \xrightarrow{L^q} X \quad \Rightarrow \quad X_n \xrightarrow{L^p} X \quad \Rightarrow \quad X_n \xrightarrow{L^1} X \quad \Rightarrow \quad X_n \xrightarrow{P} X.$$

Définition 47. Une famille $(X_i)_{i\in I}$ de v.a.r. sur $(\Omega, \mathcal{A}, P)$ est *uniformément intégrable*, (UI en abrégé) ou *équiintégrable* si

$$\lim_{c\to+\infty} \sup_{i\in I} E\left(|X_i| \, 1_{|X_i|>c}\right) = 0.$$

Remarque 15. Si $(X_i)_{i\in I}$ est UI, chacune des variables X_i est intégrable.

Proposition 78. *(Conditions suffisantes d'uniforme intégrabilité).*
- *Une famille finie de variables intégrables est UI.*
- *Une famille* $(X_i)_{i\in I}$ *telle qu'il existe une v.a.r.* Y *intégrable avec* $|X_i| \leq Y$, *pour tout* $i \in I$ *(on parle de famille dominée), est UI.*
- *S'il existe* $\Phi : \mathbb{R}^+ \to \mathbb{R}^+$ *telle que* $\lim_{x\to+\infty} \Phi(x)/x = +\infty$ *et* $\sup_{i\in I} E\Phi\big(|X_i|\big) < +\infty$ *alors la famille* $(X_i)_{i\in I}$ *est UI.*

L'uniforme intégrabilité permet de remonter de la convergence en probabilité à celle dans L^1. L'énoncé suivant peut être vu comme une généralisation du théorème de convergence dominée.

Théorème 79. *Si la suite* $(X_n)_{n\in\mathbb{N}}$ *est* UI *et* $X_n \xrightarrow{P} X$ *alors* X *est intégrable et* $X_n \xrightarrow{L^1} X$.

Proposition 80. *Si* $X_n \xrightarrow{P} X$ *et s'il existe* $q > 1$ *tel que* $\sup_n \|X_n\|_q < +\infty$ *alors*

$$X_n \xrightarrow{L^p} X, \quad \forall p \in [1, q[\,.$$

7.4 Convergence en loi

On note $C_b(\mathbb{R}^d)$ l'ensemble des fonctions continues bornées de $\mathbb{R}^d$ dans $\mathbb{R}$ et $C_0(\mathbb{R}^d)$ l'ensemble des fonctions continues $f : \mathbb{R}^d \to \mathbb{R}$ avec $\lim\limits_{|x|\to+\infty} f(x) = 0$.

Proposition 81. *Soient μ et ν deux mesures de probabilité boréliennes sur $\mathbb{R}^d$. Alors $\mu = \nu$ si et seulement si*

$$\forall \varphi \in C_b(\mathbb{R}^d), \quad \int \varphi\, d\mu = \int \varphi\, d\nu.$$

Définition 48. Soient μ et $(\mu_n)_{n\geq 0}$ des mesures finies sur $\big(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d)\big)$. On dit que la suite de mesures $(\mu_n)_{n\geq 0}$ converge *étroitement* vers μ et on note $\mu_n \xrightarrow{\text{étroit}} \mu$ si

$$\forall \varphi \in C_b(\mathbb{R}^d), \quad \lim_{n\to\infty} \int \varphi\, d\mu_n = \int \varphi\, d\mu.$$

Proposition 82. *Soit H un sous-ensemble dense de $\big(C_0(\mathbb{R}^d), \|\cdot\|_\infty\big)$. Soient μ et $(\mu_n)_{n\geq 0}$ des mesures de* **probabilité** *sur $\big(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d)\big)$. Alors*

$$\mu_n \xrightarrow{\textit{étroit}} \mu \quad \Leftrightarrow \quad \forall \varphi \in H, \lim_{n\to\infty} \int \varphi\, d\mu_n = \int \varphi\, d\mu.$$

Définition 49. Soient X et $(X_n)_{n\geq 0}$ v.a. à valeurs dans $\big(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d)\big)$, mais pas forcément définies sur le même espace de probabilité. On dit que la suite $(X_n)_{n\geq 0}$ converge *en loi* vers X et on note $X_n \xrightarrow{\mathcal{L}} X$ si $P_{X_n} \xrightarrow{\text{étroit}} P_X$, c'est-à-dire si

$$\forall \varphi \in C_b(\mathbb{R}^d), \quad \lim_{n\to\infty} E\varphi(X_n) = E\varphi(X).$$

On note aussi parfois $X_n \xrightarrow{\mathcal{L}} P_X$.

Proposition 83. $X_n \xrightarrow{P} X \quad \Rightarrow \quad X_n \xrightarrow{\text{loi}} X.$

Proposition 84. *Soient μ_n, μ des mesures de probabilité sur $\big(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d)\big)$. Les assertions suivantes sont équivalentes :*

1. $\mu_n \xrightarrow{\textit{étroit}} \mu$.
2. *Pour tout ouvert $G \subset \mathbb{R}^d$, $\liminf\limits_{n\to\infty} \mu_n(G) \geq \mu(G)$.*
3. *Pour tout fermé $F \subset \mathbb{R}^d$, $\limsup\limits_{n\to\infty} \mu_n(F) \leq \mu(F)$.*
4. *Pour tout borélien $B \subset \mathbb{R}^d$ tel que $\mu(\partial B) = 0$ on a $\lim\limits_{n\to+\infty} \mu_n(B) = \mu(B)$.*

La convergence en loi des v.a. réelles peut se vérifier sur les fonctions de répartition :

Proposition 85. *Soient X et $(X_n)_{n\geq 0}$ des v.a.réelles. Alors $X_n \xrightarrow{\mathcal{L}} X$ si et seulement si en tout $t \in \mathbb{R}$ où F_X est continue on a*

$$\lim_{n\to\infty} F_{X_n}(t) = F_X(t).$$

Définition 50. Une famille $(\mu_i)_{i\in I}$ de mesures de probabilité sur $\big(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d)\big)$ est dite *tendue* ou *équi-tendue* si pour tout $\varepsilon > 0$ il existe un ensemble compact $K_\varepsilon \subset \mathbb{R}^d$ tel que

$$\forall i \in I, \quad \mu_i(K_\varepsilon) \geq 1 - \varepsilon.$$

Théorème 86 (Prokhorov)**.** *Soit* $(\mu_n)_{n\in\mathbb{N}}$ *une suite tendue de mesures probabilités sur* $\big(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d)\big)$*. Alors on peut en extraire une sous-suite qui converge étroitement vers une mesure de probabilité.*

Théorème 87 (Lévy)**.** *1. Si* $X_n \xrightarrow{\mathcal{L}} X$ *alors* $\forall \xi \in \mathbb{R}^d$, $\lim_{n\to\infty} \Phi_{X_n}(\xi) = \Phi_X(\xi)$.

2. S'il existe une fonction $\Phi : \mathbb{R}^d \to \mathbb{C}$ *continue en 0 telle que pour tout* $\xi \in \mathbb{R}^d$,

$$\lim_{n\to\infty} \Phi_{X_n}(\xi) = \Phi(\xi),$$

alors il existe une mesure de probabilité μ *sur* $\big(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d)\big)$ *telle que* $\Phi = \hat{\mu}$*. De plus* $\mu_n \xrightarrow{\text{étroit}} \mu$ *et si* X *est une v.a. de loi* μ *on a* $X_n \xrightarrow{\mathcal{L}} X$*.*

Théorème 88 (Skorokhod)**.** *On considère une suite de v.a.* $X_n : (\Omega_n, \mathcal{A}_n, P_n) \to \big(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d)\big)$, $n \geq 0$*. Si la suite* $(X_n)_{n\geq 0}$ *converge en loi, alors il existe un espace de probabilité* $(\Omega, \mathcal{A}, P)$ *et une suite de variables aléatoires* $Y_n : (\Omega, \mathcal{A}, P) \to \big(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d)\big)$ *telle que*

— *pour tout* $n \geq 0$, $P_{X_n} = P_{Y_n}$,

— *la suite* $(Y_n)_{n\geq 0}$ *converge presque sûrement.*

Chapitre 8

Théorèmes limites

8.1 Loi des grands nombres

Définition 51. Soit $(X_n)_{n\geq 1}$ une suite de v.a.r. définies sur un même espace de probabilité $(\Omega, \mathcal{A}, P)$. A un échantillon de taille n, $\big(X_1(\omega), \dots, X_n(\omega)\big)$, on associe

— la *moyenne empirique* $\overline{X}_n(\omega) := \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} X_i(\omega)$,

— la *mesure empirique* $\mu_n^\omega := \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} \delta_{X_i(\omega)}$.

Théorème 89 (Loi faible des grands nombres)**.** *Soit $(X_n)_{n\in\mathbb{N}^*}$ une suite de v.a.r indépendantes et de même loi avec $E(X_1^2) < +\infty$, alors*

$$\overline{X}_n = \frac{1}{n}\sum_{j=1}^{n} X_j \xrightarrow[n\to\infty]{L^2} EX_1.$$

Théorème 90 (Loi forte des grands nombres)**.** *Soit $(X_n)_{n\in\mathbb{N}^*}$ une suite de v.a.r indépendantes et de même loi avec $E|X_1| < +\infty$, alors*

$$\overline{X}_n \xrightarrow[n\to\infty]{p.s.} EX_1.$$

Théorème 91 (Fondement de la statistique)**.** *Soit $(X_n)_{n\in\mathbb{N}^*}$ une suite de v.a.r indépendantes et de même loi sur $(\Omega, \mathcal{A}, P)$, alors*

$$P\left(\left\{\omega \in \Omega,\ \mu_n^\omega \xrightarrow[n\to\infty]{\textit{étroit}} P_{X_1}\right\}\right) = 1.$$

i.e. presque sûrement, la mesure empirique converge étroitement vers la loi de X_1.

8.2 Théorème de la limite centrale

Théorème 92. *Soit $(X_n)_{n\geq 1}$ une suite de v.a.r. indépendantes et de même loi avec $E(X_1^2) < \infty$. Soit $m = EX_1$ et $\sigma^2 = \mathrm{Var}(X_1)$, alors*

$$\sqrt{n}\,\frac{\overline{X}_n - EX_1}{\sigma} = \frac{\sum_{i=1}^{n} X_i - nEX_1}{\sigma\sqrt{n}} \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0,1).$$

Soit $(X_n)_{n\geq 1}$ une suite de v.a. à valeur dans $\mathbb{R}^d$, indépendantes et de même loi. On suppose que $E\big(|X_1|^2\big) < \infty$ et on pose $\Gamma = \mathrm{Cov}(X_1)$. Dans ce cas

$$\frac{X_1 + \cdots + X_n - nEX_1}{\sqrt{n}} \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}_d(0,\Gamma).$$

Corollaire 93. *Soit $(X_n)_{n\geq 1}$ une suite de v.a.r. indépendantes et de même loi avec $E(X_1^2) < \infty$. On note $\sigma^2 = \mathrm{Var}(X_1)$. Alors pour tout $a \geq 0$,*

$$\lim_{n\to\infty} P\left(\left|\overline{X}_n - EX_1\right| \geq \frac{\sigma a}{\sqrt{n}}\right) = 2\int_a^{+\infty} e^{-t^2/2}\frac{dt}{\sqrt{2\pi}}.$$

8.3 Valeurs extrêmes (hors programme)

Définition 52. La *borne supérieure essentielle* d'une v.a.r. X est donnée par

$$s_X := \sup\big\{x,\ P(X \leq x) < 1\big\} \in]-\infty, +\infty].$$

Théorème 94. *Soient $(X_n)_{n\geq 1}$ une suite de v.a.r. indépendantes de même loi, alors*

$$M_n := \max(X_1, \ldots, X_n) \xrightarrow[n\to\infty]{p.s.} s_{X_1}.$$

Théorème 95. *Soit $(X_n)_{n\geq 1}$ une suite de v.a.r. indépendantes et de même loi. Soit s le sup essentiel de leur loi commune et F sa fonction de répartition.*

1. *S'il existe une fonction $g > 0$ telle que pour tout $x \in \mathbb{R}$,*

$$\lim_{t \overset{<}{\to} s} \frac{1 - F(t + xg(t))}{1 - F(t)} = e^{-x}$$

alors il existe des suites $(a_n)_{n\in\mathbb{N}}$ et $(b_n)_{n\in\mathbb{N}}$ avec $a_n > 0$ et

$$P\big(a_n(M_n - b_n) \leq t\big) \underset{n\to\infty}{\longrightarrow} G_I(t) := e^{-e^{-t}}.$$

2. *Si $s = +\infty$ et s'il existe $\alpha > 0$ tel que pour tout $x > 0$,*

$$\lim_{t\to\infty} \frac{1 - F(tx)}{1 - F(t)} = x^{-\alpha},$$

alors il existe des suites $(a_n)_{n\in\mathbb{N}}$ et $(b_n)_{n\in\mathbb{N}}$ avec $a_n > 0$ et

$$P\big(a_n(M_n - b_n) \leq t\big) \underset{n\to\infty}{\longrightarrow} G_{II}(t) := \begin{cases} 0 & \text{si} \quad t < 0, \\ e^{-(x^{-\alpha})} & \text{si} \quad t \geq 0. \end{cases}$$

3. *Si $s < +\infty$ et s'il existe $\alpha > 0$ tel que pour tout $x > 0$,*

$$\lim_{t \overset{>}{\to} 0} \frac{1 - F(s - tx)}{1 - F(s - t)} = x^{\alpha},$$

alors il existe une suite $(a_n)_{n\in\mathbb{N}}$ avec $a_n > 0$ et

$$P\big(a_n(M_n - s) \leq t\big) \underset{n\to\infty}{\longrightarrow} G_{III}(t) := \begin{cases} e^{-(-t)^{\alpha}} & \text{si} \quad t < 0, \\ 1 & \text{si} \quad t \geq 0. \end{cases}$$

Remarque 16. On peut montrer que si la suite des $a_n(M_n - b_n)$ converge en loi vers une variable non constante X, alors il existe a, b tels que la fonction de répartition de $aX + b$ soit de la forme G_I, G_{II} ou G_{III}. Cependant il ne peut pas toujours exister une limite non constante :

Proposition 96. *Soit $(X_n)_{n\geq 1}$ une suite de v.a.r. indépendantes et de même loi, avec*

$$s_{X_1} < \infty \quad et \quad P_{X_1}\big(\{s_{X_1}\}\big) > 0.$$

Si une suite $(\mu_n)_{n\geq 1}$ vérifie $\lim_{n\to\infty} P\left(M_n \leq \mu_n\right) = \rho$ alors $\rho \in \{0, 1\}$. En d'autres termes, si la suite $\big(a_n(M_n - b_n)\big)_{n\geq 1}$ converge en loi, sa limite est une constante.

8.4 Méthode Delta

C'est un principe qui permet d'enrichir les résultats précédents de convergence en loi :

Théorème 97. *Soit $(T_n)_{n\in\mathbb{N}}$ et V des vecteurs aléatoires à valeurs dans $\mathbb{R}^d$. Soit (r_n) une suite de nombres réels vérifiant $\lim_n r_n = +\infty$, et soit $\theta \in \mathbb{R}^d$. On suppose que*

$$r_n(T_n - \theta) \xrightarrow[n\to\infty]{\mathcal{L}} V.$$

Soit alors $\phi : \mathbb{R}^d \to \mathbb{R}^p$ une application différentiable. Alors

$$r_n(\phi(T_n) - \phi(\theta)) \xrightarrow[n\to\infty]{\mathcal{L}} d\phi(\theta) \cdot V,$$

où $d\phi(\theta)$ est la différentielle de ϕ au point θ.

Chapitre 9

Introduction à la statistique

9.1 Structure statistique

Définition 53. Une *structure statististique*, ou *modèle statistique* est un triplet $\big(\Omega, \mathcal{A}, (P_\theta)_{\theta\in\Theta}\big)$ formé d'un espace mesurable $(\Omega, \mathcal{A})$ muni d'une famille de mesures de probabilités. Le modèle est dit *paramétrique* si Θ est un sous-ensemble de $\mathbb{R}^d$ pour un certain d. Un élément ω de Ω est appelé une *observation*.

Les probabilités P_θ correspondent à divers modèles probabilistes susceptibles de représenter un phénomène. L'objectif principal de la statistique est de proposer parmi ces modèles celui qui est le plus fidèle aux résultats d'expériences. En d'autre termes, à partir d'une observation ω on cherche à déterminer le paramètre (inconnu) θ qui est le plus proche de la réalité.

Définition 54. Une *structure d'échantillonnage* est une structure statistique produit

$$\big(\Omega, \mathcal{A}, (P_\theta)_{\theta\in\Theta}\big)^{\otimes n} := \big(\Omega^n, \mathcal{A}^{\otimes n}, (P_\theta^{\otimes n})_{\theta\in\Theta}\big).$$

Une telle structure sert à modéliser une expérience répétée n fois. On définit souvent les variables aléatoires $X_i : (\Omega^n, \mathcal{A}^{\otimes n}) \to (\Omega, \mathcal{A})$ par $X_i(\omega_1, \ldots, \omega_n) := \omega_i$. Si l'on munit $(\Omega^n, \mathcal{A}^{\otimes n})$ de la probabilité $P_\theta^{\otimes n}$ alors $X_1, \ldots, X_n$ sont indépendantes et de loi P_θ. On dit souvent que $(X_1(\omega), \ldots, X_n(\omega))$ est un *n-échantillon*.

Remarque 17. On peut aussi considérer des structures produit dénombrable $\big(\Omega^{\mathbb{N}}, \mathcal{A}^{\otimes\mathbb{N}}, (P_\theta^{\otimes\mathbb{N}})_{\theta\in\Theta}\big)$.

Définition 55. Le modèle $\big(\Omega, \mathcal{A}, (P_\theta)_{\theta\in\Theta}\big)$ est *dominé* s'il existe une mesure σ-finie μ sur $(\Omega, \mathcal{A})$ telle que

$$\forall\theta \in \Theta,\ P_\theta \ll \mu.$$

Dans ce cas, toute fonction $L : \Omega \times \Theta \to \mathbb{R}$ telle que pour tout θ, $dP_\theta(\omega) = L(\omega, \theta)\, d\mu(\omega)$ est appelée *vraisemblance* du modèle. Autrement dit, $L(\cdot, \theta)$ est une densité de P_θ par rapport à μ.

Définition 56. Soit $\big(\Omega, \mathcal{A}, (P_\theta)_{\theta\in\Theta}\big)$ un modèle statistique et $(E, \mathcal{B})$ un espace mesurable. On appelle *variable statitistique* ou simplement *statistique* toute application $T : (\Omega, \mathcal{A}) \to (E, \mathcal{B})$, c'est-à-dire toute variable aléatoire. Il est à noter que T ne dépend pas du paramètre θ.

— On dit que T, à valeurs dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, est *sommable* si

$$\forall\theta \in \Theta, \quad E_\theta|T| := \int_\Omega |T(\omega)|\, dP_\theta(\omega) < +\infty.$$

— Si T^2 est sommable on note $\mathrm{Var}_\theta(T) := E_\theta(T^2) - (E_\theta T)^2$.

— La loi de T dépend de θ puisque c'est par définition la mesure image de P_θ par T :

$$P_{\theta,T}(B) = P_\theta\big(T^{-1}(B)\big), \quad \forall B \in \mathcal{B}.$$

— La structure statistique *induite* par T est $\big(E, \mathcal{B}, (P_{\theta,T})_{\theta\in\Theta}\big)$.

Dans la suite l'ensemble Θ des paramètres sera un sous-ensemble de $\mathbb{R}^d$ muni d'une tribu. Le but de l'estimation statistique sera de retrouver la vraie valeur de θ étant donnée une observation ω. Parfois on cherche seulement à déterminer une partie de l'information contenue dans θ (par exemple, une de ses coordonnées). Ce nouveau *paramètre d'intéret* est de la forme $\lambda := g(\theta)$.

Définition 57. Soit $g : (\Theta, \mathcal{T}) \to (\Lambda, \mathcal{L})$ une application mesurable à valeurs dans un sous-ensemble de $\mathbb{R}^k$. Un *estimateur* de $g(\theta)$ est une statistique $T : (\Omega, \mathcal{A}) \to (\Lambda, \mathcal{L})$.

Pour une observation ω, la valeur $T(\omega)$ est une proposition de valeur de $g(\theta)$. Il convient de quantifier la précision d'une telle estimation.

Définition 58. Le *biais* d'un estimateur T (sommable) de $g(\theta)$ est l'application $b_T : \Theta \to \mathbb{R}^k$ définie par

$$b_T(\theta) := E_\theta(T) - g(\theta).$$

Le *risque quadratique* de T est l'application $r_T : \Theta \to [0, +\infty]$ définie par

$$r_T(\theta) := E_\theta\big(|T - g(\theta)|^2\big).$$

Ici on a noté $|x|$ la norme euclidienne d'un vecteur $x \in \mathbb{R}^k$.

Remarque 18. On peut considérer des notions plus générales : si $L : \mathbb{R}^k \times \mathbb{R}^k \to \mathbb{R}^+$ vérifie $L(x, x) = 0$ pour tout x, la fonction de risque de T au sens de L est $\theta \mapsto R_T^L(\theta) := E_\theta L(T, g(\theta))$.

Définition 59. Soient S, T deux estimateurs de $g(\theta) \in \mathbb{R}^k$. On dit que T est *meilleur* que S en terme de risque quadratique si

$$\forall \theta \in \Theta, \quad E_\theta\big(|T - g(\theta)|^2\big) \leq E_\theta\big(|S - g(\theta)|^2\big).$$

On dit que T est strictement meilleur si en plus il existe une valeur de θ pour laquelle l'inégalité est stricte.

Définition 60. On dit qu'un estimateur est *admissible* s'il n'existe pas d'estimateur strictement meilleur.

Dans le cadre d'un modèle d'échantillonnage $\big(\Omega^{\mathbb{N}}, \mathcal{A}^{\otimes\mathbb{N}}, (P_\theta^{\otimes\mathbb{N}})_{\theta\in\Theta}\big)$, on considère généralement des suites d'estimateurs $(T_n)_{n\geq 1}$ de $g(\theta)$ avec la propriété que T_n est une fonction de $X_1, \dots, X_n$ seulement. On peut alors considérer des propriétés asymptotiques de la suite d'estimateurs : on dit que

— (T_n) est *asymptotiquement sans biais* si $\forall \theta \in \Theta, \quad \lim_{n\to\infty} E_\theta T_n = g(\theta)$.

— (T_n) converge vers λ au sens L^1 (ou L^2, p.s., en proba) si $\forall \theta \in \Theta,\ T_n$ converge vers $g(\theta)$ au sens L^1 (ou L^2, p.s., en proba).

Remarque 19. Il faut noter que toutes les notions de convergence dépendent de θ puisque leur définition fait intervenir la probabilité sous-jacente.

9.2 Quelques techniques d'estimation

9.2.1 Moments empiriques

On considère un modèle d'échantillonnage $\big(\mathbb{R}^{\mathbb{N}}, \mathcal{B}(\mathbb{R})^{\otimes\mathbb{N}}, (P_\theta^{\otimes\mathbb{N}})_{\theta\in\Theta}\big)$. On suppose que X_1 est sommable. Si le paramètre d'intérêt est $m(\theta) = E_\theta X_1$, un estimateur naturel est la moyenne empirique

$$\overline{X}_n := \frac{X_1 + \cdots + X_n}{n}.$$

C'est un estimateur sans biais qui converge vers $m(\theta)$ au sens presque sûr.

Plus généralement, supposons que X_1^k est sommable et que le paramètre d'intérêt est $g(\theta) = \psi(m_1(\theta), \ldots, m_k(\theta))$, où l'on a posé $m_\ell(\theta) := E_\theta(X_1^\ell)$. On définit les moments empiriques par la formule

$$\widehat{m}_\ell := \frac{X_1^\ell + \cdots + X_n^\ell}{n}$$

et on propose l'estimateur

$$\widehat{g} := \psi\big(\widehat{m}_1, \ldots, \widehat{m}_k\big).$$

9.2.2 Maximum de vraisemblance

Soit $(\Omega, \mathcal{A}, (P_\theta)_{\theta\in\Theta})$ un modèle dominé de fonction de vraisemblance $L(\omega, \theta)$. On appelle estimateur du *maximum de vraisemblance* de θ tout estimateur $\widehat{\theta}$ tel que

$$\forall \omega \in \Omega,\ L\big(\omega, \widehat{\theta}(\omega)\big) = \max_{\theta\in\Theta} L(\omega, \theta).$$

En d'autres termes, un tel estimateur propose la valeur de θ qui rend l'événement observé le plus probable.

Remarque 20. Un tel estimateur n'existe pas forcément. S'il existe, il n'est pas forcément unique.

9.2.3 Intervalle de confiance

On considère ici un paramètre d'intérêt $g(\theta) \in \mathbb{R}$. Soit $\alpha \in [0, 1]$. Si deux estimateurs $\widehat{a}, \widehat{b}$ vérifient

$$\forall \theta \in \Theta, \quad P_\theta\big(\widehat{a} \le g(\theta) \le \widehat{b}\big) \ge 1 - \alpha,$$

on dit que $[\widehat{a}, \widehat{b}]$ est un *intervalle de confiance*, de sécurité $1 - \alpha$.

Dans le cas d'un modèle d'échantillonnage, on parle d'*intervalle de confiance asymptotique* si l'on a des suites d'estimateurs $(\widehat{a}_n), (\widehat{b}_n)$ telles que

$$\forall \theta \in \Theta, \quad \lim_{n\to+\infty} P_\theta\big(\widehat{a}_n \le g(\theta) \le \widehat{b}_n\big) \ge 1 - \alpha.$$

Les théorèmes limite du chapitre précédent permettent d'en construire.

9.3 Exhaustivité

Définition 61. Soit $(\Omega, \mathcal{A}, (P_\theta)_{\theta\in\Theta})$ une structure statistique. On dit qu'une statistique $T : (\Omega, \mathcal{A}) \to (\Omega', \mathcal{A}')$ est *exhaustive* si pour toute statistique X à valeurs réelles, positive ou sommable, il existe une fonction mesurable $f : \Omega' \to \mathbb{R}$ telle que

$$\forall \theta \in \Theta, \quad E_\theta\big(X|T\big) = f(T) \quad P_\theta - \text{p.s.}$$

En d'autres termes la loi conditionnelle sachant T ne dépend plus du paramètre θ. Ainsi la statistique T contient toute l'information relative à θ. Comme $E_\theta\big(X|T\big)$ ne dépend plus de θ, on le note simplement $E\big(X|T\big)$.

Proposition 98. *Soit S un estimateur de $g(\theta)$ et soit T une statistique exhaustive. Alors $E(S|T) = E_\theta(S|T)$ est un estimateur de $g(\theta)$, de même fonction de biais que S. De plus $E(S|T)$ est meilleur que S en termes de risque quadratique.*

Le théorème suivant donne un critère d'exhaustivité pour les modèles dominés :

Théorème 99 (Neyman-Fisher)**.** *Soit $(\Omega, \mathcal{A}, (P_\theta)_{\theta\in\Theta})$ un modèle dominé par μ, et de fonction de vraisemblance L. S'il existe*

— *une application borélienne $g : (\Omega, \mathcal{A}) \to \mathbb{R}^+$,*

— *pour chaque $\theta \in \Theta$, une application $g_\theta : (E, \mathcal{B}) \to \mathbb{R}^+$ borélienne,*

— *une statistique $T : (\Omega, \mathcal{A}) \to (E, \mathcal{B})$*

telles que

$$\forall\theta \in \Theta,\ \forall\omega \in \Omega, \quad L(\omega, \theta) = g(\omega)g_\theta\big(T(\omega)\big)$$

alors T est une statistique exhaustive. La réciproque est aussi vraie.

Remarque 21. La statistique $T(\omega) := \omega$ est donc exhaustive. Cependant une statistique exhaustive est d'autant plus intéressante que son espace d'arrivée E est de petite dimension.

9.4 Complétude

Définition 62. Une statistique $T : (\Omega, \mathcal{A}) \to (E, \mathcal{B})$ est dite *complète* si pour toute application borélienne $h : (E, \mathcal{B}) \to \mathbb{R}$ telle que $h(T)$ est sommable, on a

$$\big(\forall\theta \in \Theta,\ E_\theta h(T) = 0\big) \implies \big(\forall\theta \in \Theta,\ h(T) = 0\ \ P_\theta - \text{p.s.}\big).$$

Théorème 100 (Lehmann-Scheffé)**.** *On suppose que $g(\theta) \in \mathbb{R}$ est le paramètre d'intérêt. Soit S une statistique réelle et T une statistique exhaustive et complète.*

Alors $E(S|T)$ est un estimateur de $g(\theta)$ de même biais que S. Parmi les estimateurs de même biais que S, c'est "l'unique" estimateur ayant un risque quadratique minimum.

Remarque 22. L'unicité est à comprendre au sens suivant : si S' est un autre estimateur de même biais que S et de risque quadratique minimum, alors pour tout $\theta \in \Theta$, $P_\theta\big(S' = E(S|T)\big) = 1$.

Corollaire 101. *Soit T une statistique exhaustive et complète. Si $U = \varphi(T)$ estime $g(\theta)$ sans biais, alors U est "le" meilleur estimateur sans biais en terme de risque quadratique.*

9.5 Liberté

Définition 63. Une statistique $T : \big(\Omega, \mathcal{A}, (P_\theta)_{\theta\in\Theta}\big) \to (\Omega', \mathcal{A}')$ est *libre* si sa loi sous P_θ (notée $P_{\theta,T}$) ne dépend pas de θ.

Proposition 102. *Si T_1 est exhaustive et complète et si T_2 est libre alors pour tout $\theta \in \Theta$, T_1 et T_2 sont indépendantes pour P_θ.*

9.6 Modèles exponentiels

Définition 64. Un modèle statistique dominé $\big(\Omega, \mathcal{A}, (P_\theta)_{\theta\in\Theta}\big)$ est dit *exponentiel* si pour un entier p il existe

— des applications $Q : \Theta \to \mathbb{R}^p$, $c : \Theta \to \mathbb{R}^+$,

— une application mesurable positive d sur Ω,

— une statistique $T : \Omega \to \mathbb{R}^p$

telles qu'une fonction de vraisemblance se factorise sous la forme

$$L(\omega, \theta) = c(\theta)\, d(\omega)\, \exp\big(\langle Q(\Theta), T(\omega)\rangle\big).$$

On dit alors que T est une statistique *naturelle.*

Remarque 23. Le nombre $C(\theta)$ s'exprime en fonction de $\eta := Q(\theta)$. On peut donc choisir η comme nouveau paramètre du modèle.

Théorème 103. *La statistique naturelle d'une structure exponentielle est*

— *exhaustive*

— *complète dès que $Q(\Theta) := \{Q(\theta);\ \theta \in \Theta\}$ contient un ouvert non-vide (de $\mathbb{R}^p$).*

9.7 Information et efficacité

Dans cette section, on considère un modèle dominé par une mesure μ. On suppose pour simplifier que l'ensemble des paramètres est un intervalle de $\mathbb{R}$. Des hypothèses de régularité sont requises : on suppose que "la" vraisemblance L du modèle est strictement positive et qu'elle est dérivable en θ. On note X la variable aléatoire définie pour $\omega \in \Omega$ par $X(\omega) = \omega$.

Définition 65. L'information de Fisher du modèle est la fonction

$$\theta \mapsto I(\theta) := \int_\Omega \left(\frac{\partial \log L}{\partial \theta}(\omega, \theta)\right)^2 P_\theta(d\omega) = E_\theta\left(\frac{\partial \log L}{\partial \theta}(X, \theta)^2\right).$$

Théorème 104 (Fréchet-Darmois, Cramer-Rao)**.** *Soit T un estimateur sans biais de $g(\theta) \in \mathbb{R}$. On suppose en plus de ce qui précède que g est dérivable et que les dérivées en θ des expressions $\int L(x, \theta)\, d\mu(x)$ et $\int T(x) L(x, \theta)\, d\mu(x)$ s'obtiennent en dérivant sous le signe somme. Alors pour tout $\theta \in \Theta$,*

$$\mathrm{Var}_\theta(T) = r_T(\theta) \geq \frac{g'(\theta)^2}{I(\theta)}.$$

Ce résultat peut être vu comme un principe d'incertitude : si T estime correctement en moyenne pour tout θ, il doit fluctuer autour de cette moyenne.

Définition 66. On dit qu'un estimateur sans biais de $g(\theta)$ est *efficace* s'il réalise l'égalité dans la précédente inégalité.

9.8 Initiation aux tests

Le cadre de travail sera une structure statistique $\big(\Omega, \mathcal{A}, (P_\theta)_{\theta\in\Theta}\big)$. Les tests ont pour but d'infirmer ou de confirmer une hypothèse sur le paramètre d'intérêt. Plus précisément, on appelle *hypothèse* sur θ un sous-ensemble de Θ. On note H : "$\theta \in A$" pour signifier que l'on considère une hypothèse nommée H et qui correspond au sous-ensemble A.

9.8.1 Definitions

Définition 67. Etant données
- une hypothèse dite *nulle* H_0 : "$\theta \in \Theta_0$",
- une hypothèse dite *alternative* H_1 : "$\theta \in \Theta_1$", avec $\Theta_0 \cap \Theta_1 = \emptyset$,

un *test déterministe de* H_0 *contre* H_1 est une statistique $\Phi : (\Omega, \mathcal{A}) \to \big(\{0,1\}, \mathcal{P}(\{0,1\})\big)$. Pour une observation $\omega \in \Omega$
- si $\Phi(\omega) = 0$ on conclut que $\theta \in \Theta_0$. On dit que l'on *accepte* H_0.
- si $\Phi(\omega) = 1$ on conclut que $\theta \in \Theta_1$. On dit que l'on *rejette* H_0.

L'ensemble $R := \{\omega \in \Omega;\ \Phi(\omega) = 1\}$ est appelé *région critique* ou *zone de rejet.*

Ce faisant, on peut se tromper de deux manières :
- $\theta \in \Theta_0$ mais $\Phi(\omega) = 1$; on parle d'erreur de *première espèce*,
- $\theta \in \Theta_1$ mais $\Phi(\omega) = 0$; on parle d'erreur de *deuxième espèce.*

Pour des raisons théoriques, on considère une définition plus générale :

Définition 68. Un *test de* H_0 *contre* H_1 est une statistique

$$\Phi : (\Omega, \mathcal{A}) \to \big([0,1], \mathcal{B}([0,1])\big).$$

On interprète $\Phi(\omega)$ comme la probabilité de rejeter H_0 :
- si $\Phi(\omega) = 0$ on accepte H_0,
- si $\Phi(\omega) = 1$ on rejette H_0,
- si $\Phi(\omega) = \alpha \in]0,1[$ on est dans la région d'*hésitation* ; on tire le résultat au hasard en rejetant avec probabilité α et en acceptant avec probabilité $1 - \alpha$.

9.8.2 Mesures de qualité des tests

Définition 69. L'*image* du test Φ est l'application $\beta_\Phi : \Theta \to [0,1]$ définie par

$$\beta_\Phi(\theta) := E_\theta \Phi = P_\theta(\text{Rejet}).$$

Cette fonction mesure la probabilité de rejet si la vraie valeur est θ.

Définition 70. La *fonction niveau* de Φ est la restriction de β_Φ à $H_0 : \theta \in \Theta_0$. Le *niveau* de Φ est le nombre

$$\sup_{\theta \in \Theta_0} \beta_\Phi(\theta).$$

On dit que Φ est de *seuil* α si $\sup_{\theta \in \Theta_0} \beta_\Phi(\theta) \leq \alpha$ c'est-à-dire si la probabilité d'erreur de première espèce reste toujours inférieure à α.

Définition 71. La *fonction puissance* de Φ est la restriction de β_Φ à $H_1 : \theta \in \Theta_1$,

$$\forall \theta \in \Theta_1, \quad \beta_\Phi(\theta) = P_\theta(\text{Rejet}) = 1 - P_\theta(\text{Acceptation}).$$

Donc cette fonction vaut un moins la probabilité d'erreur de deuxième espèce.

Définition 72. Soient Φ et Φ' deux tests de H_0 contre H_1. On dit que
- Φ est *meilleur* que Φ' et on note $\Phi \leq \Phi'$, si $\begin{cases} \forall \theta \in \Theta_0, & \beta_\Phi(\theta) \leq \beta_{\Phi'}(\theta), \\ \forall \theta \in \Theta_1, & \beta_\Phi(\theta) \geq \beta_{\Phi'}(\theta), \end{cases}$
- Φ et Φ' sont *equivalents* et on note $\Phi \sim \Phi'$ si $\Phi \leq \Phi'$ et $\Phi' \leq \Phi$ ce qui revient à dire que $\beta_\Phi = \beta_{\Phi'}$,
- Φ est *strictement meilleur* que Φ' si $\Phi \leq \Phi'$ et $\Phi \not\sim \Phi'$,

— Φ est *admissible* s'il n'existe pas de test strictement meilleur.

Comme il existe deux types d'erreurs contradictoires, il n'est pas possible de minimiser simultanément les deux. On introduit donc une notion dissymétrique de test optimal :

Définition 73. Un test Φ de seuil α est *uniformément plus puissant au niveau* α (UPP_α) si sa fonction puissance est plus grande que celle de tout test de seuil α.

En d'autres termes, parmi tous les tests dont la probabilité d'erreur de première espèce est uniformément majorée par α, un test UPP_α a toujours une probabilité d'erreur de deuxième espèce minimale.

9.8.3 Construction de tests

On considère un modèle dominé, de fonction de vraisemblance L. Dans ce cas on peut construire des tests de $H_0 : \theta \in \Theta_0$ contre $H_1 : \theta \in \Theta_1$ en utilisant le rapport de vraisemblance, défini pour $\omega \in \Omega$ par

$$\ell(\omega) = \frac{\sup_{\theta \in \Theta_1} L(\omega, \theta)}{\sup_{\theta \in \Theta_0} L(\omega, \theta)}.$$

Pour toute fonction f croissante et à valeurs dans $[0, 1]$, le test défini par

$$\Phi(\omega) := f(\ell(\omega)), \quad \forall \omega \in \Omega$$

est un *test du rapport de vraisemblance.*

On peut par exemple considérer $\Phi(\omega) = \mathbb{1}_{\ell(\omega) \geq k}$ où k est un paramètre à déterminer afin de contrôler le niveau du test. Cependant on obtient souvent des résultats plus précis avec un test non-déterministe. C'est le cas pour des hypothèses simples (c'est-à-dire telles que Θ_0 et Θ_1 sont des singletons) :

Théorème 105 (Neyman-Pearson)**.** *Soit $\alpha \in]0, 1[$ et $\theta_0, \theta_1 \in \Theta$. Il existe $(k, p) \in \mathbb{R}^+ \times [0, 1]$ pour lesquels le test Φ défini pour $\omega \in \Omega$ par*

$$\Phi(\omega) := \begin{cases} 1 & si \quad L(\omega, \theta_1) > k\, L(\omega, \theta_0), \\ p & si \quad L(\omega, \theta_1) = k\, L(\omega, \theta_0), \\ 0 & si \quad L(\omega, \theta_1) < k\, L(\omega, \theta_0), \end{cases}$$

est UPP_α pour tester H_0 : "$\theta = \theta_0$" contre H_1 : "$\theta = \theta_1$".

Ces tests UPP_α pour hypothèses simples permettent parfois de construire des tests UPP_α impliquant des hypothèses composites :

Proposition 106. *Soit $\alpha \in]0, 1[$. Soit $\{\theta_i,\ i \in I\}$ un sous-ensemble de Θ ne contenant pas θ_0. On considère les hypothèses H_0 : "$\theta = \theta_0$", H_1 : "$\theta \in \{\theta_i,\ i \in I\}$" et pour $i \in I$, H'_i : "$\theta = \theta_i$".*

1. *Si Φ est un test qui ne dépend pas de i et qui pour tout $i \in I$ est UPP_α pour tester H_0 contre H'_i alors Φ est un test UPP_α de H_0 contre H_1.*
2. *Réproquement si Φ est un test UPP_α de H_0 contre H_1 alors pour tout $i \in I$, Φ est un test UPP_α de H_0 contre H'_i.*