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

Basées sur des notes de F. Barthe et complétées par G. Cébron et C. Pellegrini
Ces notes sont en cours de correction et donc peuvent présenter des erreurs.
Institut de Mathématiques.
Université Paul Sabatier.
31062 Toulouse cedex 09.
Email : barthe@math.univ-toulouse.fr

4 février 2021

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. $\cap A_n = \left(\cup\left(A_n^c\right)\right)^c$. 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.*

Vocabulaire. La tribu $\{\emptyset, \Omega\}$ est appelée tribu triviale et $\mathcal{P}(\Omega)$ est appelée tribu grossière.

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

Remarque 2. Une union d'algèbres ou de tribus n'est pas en générale une tribu ou une algèbre.

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 $\mathcal{E} \subset \mathcal{A} \Leftrightarrow \sigma(\mathcal{E}) \subset \mathcal{A}$.

Remarque 3. La définition ci dessus n'est pas vide car il existe toujours une tribu ou une algèbre qui contient $\mathcal{E}$. Laquelle ?

Remarque 4. Si $\mathcal{E} \subset \mathcal{F}$ alors $\sigma(\mathcal{E}) \subset \sigma(\mathcal{F})$. De même pour les algèbres.

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 5. On munit $\mathbb{R}$ sa topologie usuelle. On a alors

$$\mathcal{B}(\mathbb{R}) = \sigma\Big(\big\{]a,b[;\ a,b \in \mathbb{R}\big\}\Big).$$

Pour montrer ce résultat il faut utiliser la caractérisation des tribus engendrée. On veut montrer que $\sigma\Big(\big\{]a,b[;\ a,b \in \mathbb{R}\big\}\Big) = \sigma\left(\{O \text{ ouverts de } \mathbb{R}\}\right)$. Il est clair que

$$\big\{]a,b[;\ a,b \in \mathbb{R}\big\} \subset \{O \text{ ouverts de } \mathbb{R}\}$$

donc $\sigma\left(\big\{]a,b[;\ a,b \in \mathbb{R}\big\}\right) \subset \sigma\left(\{O \text{ ouverts de } \mathbb{R}\}\right)$. Réciproquement tout ouvert de $\mathbb{R}$ peut s'écrire comme union dénombrable d'intervalles ouverts. En effet soit O un ouvert alors si $x \in O$ il existe $\epsilon > 0$ tel que $]x-\epsilon, x+\epsilon[\subset O$, on peut alors choisir r et r' rationnels tel que $x - \epsilon < r < x < r' < x + \epsilon$, d'où $]r, r'[\subset O$. Pour chaque $x \in O$ on associe donc deux tels rationnels. L'ensemble O peut alors s'écrire comme l'union des intervalles ainsi construit. Comme l'ensemble des rationnels est dénombrable , on a écrit O comme une union dénombrable d'intervalle. Au passage on a montré également que

$$\mathcal{B}(\mathbb{R}) = \sigma\Big(\big\{]a,b[;\ a,b \in \mathbb{Q}\big\}\Big).$$

On aurait pu également considérer les composantes connexes. On a aussi

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

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

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

Exemple 6. On a le résultat suivant

$$\mathcal{B}(\mathbb{R}^2) = \mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R})$$

On peut montrer ce résultat de plusieurs façons. Tout d'abord pour montrer que

$$\mathcal{B}(\mathbb{R}^2) \subset \mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R}).$$

on peut considérer un ouvert de$\mathbb{R}^2$ et l'écrire comme une union dénombrable d'intervalles (c'est la même idée que celle utilisée dans la caractérisation de $\mathcal{B}(\mathbb{R})$). Pour la réciproque : soient B_1 et B_2 deux ouverts de $\mathbb{R}$ on veut montrer que $B_1 \times B_2 \in \mathcal{B}(\mathbb{R}^2)$. Tout d'abord considérons O_1 un ouvert de $\mathbb{R}$ et l'ensemble

$$\mathcal{E}_2 = \{B \in \mathcal{B}(\mathbb{R}) | O_1 \times B \in \mathcal{B}(\mathbb{R}^2)\}$$

Il convient de voir que $\mathcal{E}_2$ est une tribu. En particulier on peut voir que

$$O_1 \times (B)^c = (O_1 \times \mathbb{R}) \cap (O_1 \times B^c).$$

De plus $\mathcal{E}_2$ contient les ouverts donc $\mathcal{B}(\mathbb{R}) \subset \mathcal{E}_2$. Ainsi pour tout borélien B on a montré que $O_1 \times B \in \mathcal{B}(\mathbb{R}^2)$ et ceci est valable pour tout ouvert O_1. Nous on veut le résultat pour $B_1 \times B_2$ considérons donc l'ensemble

$$\mathcal{E}_1 = \{C \in \mathcal{B}(\mathbb{R}) | C \times B_2 \in \mathcal{B}(\mathbb{R}^2)\}.$$

On montre que $\mathcal{E}_1$ est une tribu qui contient les ouverts d'après ce qui précéde donc $\mathcal{B}(\mathbb{R}) \subset \mathcal{E}_1$. Ainsi pour tout borélien $B \in \mathcal{B}(\mathbb{R})$ on a que $B \times B_2 \in \mathcal{B}(\mathbb{R}^2)$ et donc pour $B = B_1$ on a le résultat attendu.

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 7. 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* $n \geq 1$*,* $f_n : (\Omega, \mathcal{A}) \to \left(\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}})\right)$ *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.*

Proposition 6. *On note* π_i *les application projections i.e*

$$\begin{array}{rcl} \pi_i : \Omega_1 \times \Omega_2 & \to & \Omega_i \\ (\omega_1, \omega_2) & \mapsto & \omega_i. \end{array}$$

La tribu produit est la plus petite tribu qui rend mesurable les applications projections mesurables

Démonstration. Soient (Ω_1, M_1), (Ω_2, M_2). Si on note $\mathcal{T} = \sigma(\pi_i^{-1}(A_i), A_i \in M_i, i = 1, 2)$ la plus petite tribu qui rend les applications mesurables. Soient $A_1 \in M_1$ et $A_2 \in M_2$ on a que $\pi_1^{-1}(A_1) = A_1 \times \nleq_2$ et $\pi_2^{-1}(A_2) = \Omega_1 \times A_2$ donc $\pi_1^{-1}(A_1) \in M_1 \otimes M_2$ et $\pi_2^{-1}(A_2) \in M_1 \otimes M_2$ d'où $\mathcal{T} \subset M_1 \otimes M_2$. On observe également que

$$A_1 \times A_2 = \pi_1^{-1}(A_1) \cap \pi_1^{-2}(A_2) \in \mathcal{T}$$

donc $M_1 \otimes M_2 \subset \mathcal{T}$. □

Exemple 8. En appliquant cela avec $\Omega_1 = \Omega_2 = \mathbb{R}$. On a que les applications π_i sont continues de $\mathbb{R}^2$ dans $\mathbb{R}$ donc sont mesurables par rapport à la tribu $\mathcal{B}(\mathbb{R}^2)$ et donc $\mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R}) \subset \mathcal{B}(\mathbb{R}^2)$. On retrouve une partie du résultat de la partie précédente.

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 $A_i \subset A_{i+1}\big] \Rightarrow \bigcup_{i \in \mathbb{N}} A_i \in \mathcal{M}$.

Exemple 9. Toute tribu est une classe monotone

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 7. *Soit* $\mathcal{M}$ *une classe monotone. Si elle est stable par intersection (ou union) finie, alors c'est une tribu.*

Démonstration. En effet il suffit de voir que si $(A_n)_n$ est une suite d'éléments de $\mathcal{M}$ alors

$$\bigcup_{n \in \mathbb{N}} A_n = \bigcup_{n \in \mathbb{N}} (A_1 \cup \ldots \cup A_n)$$

et si on a stabilité par union finie alors la propriété 3 de la définition ci-dessus permet de conclure. □

Théorème 8 (Classes monotones)**.** *Soit $\mathcal{E} \subset \mathcal{P}(\Omega)$ stable par intersection finie. Alors $\mathcal{M}(\mathcal{E}) = \sigma(\mathcal{E})$.*

Démonstration. L'ensemble $\sigma(\mathcal{E})$ est une tribu qui contient $\mathcal{E}$ donc elle contient la c.m engendrée par $\mathcal{E}$. Ainsi on a $\mathcal{M}(\mathcal{E}) \subset \sigma(\mathcal{E})$. Pour montrer la réciproque, il suffit de montrer que $\mathcal{M}(\mathcal{E})$ est stable par intersection finie. Pour cela considérons deux événements A et B dans $\mathcal{M}(\mathcal{E})$ et montrons que $A \cap B \in \mathcal{M}(\mathcal{E})$. Pour cela mimons ce qu'on a fait pour $\mathcal{B}(\mathbb{R}^2)$. Soit $E \in \mathcal{E}$, considérons

$$\mathcal{U} = \{C \in \mathcal{M}(\mathcal{E}) | E \cap C \in \mathcal{M}(\mathcal{E})\}.$$

Nous allons voir que $\mathcal{U}$ est une classe monotone qui contient $\mathcal{E}$. Comme nous avons supposé que $\mathcal{E}$ était stable par intersection finie, il est donc clair que $\mathcal{E} \subset \mathcal{U}$. La propriété 1 et 3 de la définition sont claires car $\mathcal{M}(\mathcal{E})$ est une c.m. La propriété 2 est plus délicate. Soient V et W deux éléments de $\mathcal{U}$ tels que $V \supset W$ alors

$$\begin{aligned} C \cap (V \setminus W) &= C \cap (V \cap W^c) \\ &= C \cap V \cap (C^c \cup W^c) \\ &= (C \cap V) \cap (C \cap W)^c \end{aligned}$$

Or $C \cap V$ et $C \cap W$ sont des éléments de $\mathcal{M}(\mathcal{E})$ telle que $C \cap V \supset C \cap W$ et donc $(C \cap V) \cap (C \cap W)^c \in \mathcal{M}(\mathcal{E})$. On a donc montré que pour tout $E \in \mathcal{E}$ et pour tout $C \in \mathcal{M}(\mathcal{E})$ on aque $E \cap C \in \mathcal{E}$. Introduisons maintenant

$$\mathcal{U}' = \{C \in \mathcal{M}(\mathcal{E}) | A \cap C \in \mathcal{M}(\mathcal{E})\}.$$

Grâce au travail fait sur $\mathcal{U}$ on sait que $\mathcal{U}'$ est une c.m qui contient $\mathcal{E}$ donc c'est une c.m et donc pour tout $C \in \mathcal{M}(\mathcal{E})$ on a que $A \cap C \in \mathcal{M}(\mathcal{E})$ et donc pour B on a que $a \cap B \in \mathcal{M}(\mathcal{E})$ cqfd. On utilsie ensuite la proposition précédente pour conclure. □

Remarque 10. Ce théorème est précieux car il permet de vérifier qu'une propriété est vraie sur une classe moins grosse que la tribu entière notamment sur des classes de générateurs (classe monotone). On peut comparer cela aux espaces vectoriels où une propriété est vraie pour des applications linéaires dès quelle est vraie sur une base.

Dans la suite nous aurons besoin de manipuler des fonctions et de vérifier des propriétés sur des classes de fonctions d'où le résultat suivant qui généralise la notion de classe monotone à celle des fonctions.

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 $\mathcal{H}$.

Remarque 11. Ici la monotonie se compare à la propriété 3 de la définition 8. En effet si $(A_n)_n$ est une suite croissante déléments alors la famille de fonctions $(\mathbf{1}_{A_n})$ est une suite croissante de fonctions bornées. Sa limite est $\mathbf{1}_{\cup A_n}$ et donc on peut directement comparer la stabilité par limite et par union.

Théorème 9. (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 et contenant les constantes. Soit $\mathcal{E} \subset \mathcal{H}$ un ensemble de fonctions stable par multiplication.

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

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

Remarque 12. Ce résultat peut être vu comme un analogue borélien du théorème de Stone-Weierstrass.

Démonstration. Supposons dans un premier temps que

$$\mathcal{E} = \{\mathbf{1}_A, A \in \mathcal{E}'\},$$

pour une certaine famille $\mathcal{E}'$. Considérons

$$\{A, \mathbf{1}_A \in \mathcal{H}\}$$

c'est une c.m quui contient $\mathcal{E}'$ donc par le lemme des c.m elle contient aussi $\sigma(\mathcal{E}') = \sigma(\mathcal{E})$. Il s'en suit que $\mathcal{H}$ contient les fonctions étagées, puis positive par monotonie et enfin quelconque par linéarité qui sont $\sigma(\mathcal{E}))$ mesurables et donc on a le résultat.

Maintenant supposons $\mathcal{E}$ quelconque, on va montrer que $\mathcal{H}$ contient

$$\{\mathbf{1}_{f_1^{-1}(]a_1,+\infty[)\cap\ldots\cap f_n^{-1}(]a_n,+\infty[)}, n \in \mathbb{N}^*, a_1, \ldots, a_n \in \mathbb{R}, f_1, \ldots, f_n \in \mathcal{E}\}.$$

Cet ensemble engendre $\sigma(\mathcal{E})$. De plus cet ensemble est stable par multiplication. Si on prouve que $\mathcal{H}$ contient un tel ensemble alors on peut répéter le raisonnement que l'on vient juste de faire avec $\mathcal{E}'$.

Pour montrer cela, considérons $f_1, \ldots, f_n \in \mathcal{E}$. Tout polynôme $P(f_1, \ldots, f_n) \in \mathcal{H}$ par linéarité. En utilisant Stone Weierstrass, par approximation uniforme sur les compacts on a que $\phi(f_1, \ldots, f_n) \in \mathcal{H}$, pour toute fonction continue. Attention cela suppose de savoir montrer que $\mathcal{H}$ est stable par limite uniforme. Admettons que cela est vraie pour l'instant alors en utilisant une limite croissante on a que

$$\mathbf{1}_{]a_1,+\infty[\cap\ldots\cap]a_n,+\infty[}(f_1, \ldots, f_n) \in \mathcal{H}$$

Stabilité par limite uniforme : Soit (f_n) une suite d'éléments de $\mathcal{H}$ telle que $\|f_n - f\|_\infty \to 0$ on veut montrer que $f \in \mathcal{M}$. Quitte à extraire une sous suite telle que $\|f_n - f\|_\infty < 2^{-n}$. On a alors

$$f_{n+1} - f_n + 2^{1-n} = f_{n+1} - f + 2^{-n} + f - f_n + 2^n > 0$$

et donc

$$\sum_{k=1}^{n} f_{k+1} - f_k + 2^{1-k}$$

est donc une suite croissante qui converge vers $f - f_1 + 2$ en croissant. Donc $f - f_1 + 2 \in \mathcal{H}$ et par suite $f \in \mathcal{H}$ □

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

1. $A \subset B \Rightarrow\ \mu(A)\ \le\ \mu(B)$.
2. $\mu(A \cup B)\ +\ \mu(A \cap B) = \mu\ (A)\ +\ \mu\ (B)$.
3. *Si* I *est au plus dénombrable, alors* $\mu(\bigcup_{i\in I} A_i) \le \sum_{i\in I} \mu(A_i)$.
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* $\mu(\bigcap_{n\in\mathbb{N}} A_n) = \lim_{n\to\infty} \downarrow \mu(A_n)$

Démonstration. Pour 1) on remarque que

$$B = A \cup (B \setminus A)$$

et que l'intersection est disjointe d'où $\mu(B) = \mu(A) + \mu(B \setminus A)$ et le résultat suit.

Pour 2) on a $A \cup B = A \cup (B \setminus (A \cap B))$ qui est une union disjointe. Avec ce que l'on vient juste d'écrire on a $\mu(B \setminus (A \cap B)) + \mu(A \cap B) = \mu(B)$ et donc

$$\mu(A \cup B) + \mu(A \cap B) = \mu(A) + \mu(B \setminus (A \cap B)) + \mu(A \cap B) = \mu(A) + \mu(B)$$

Pour 3), pour tout n on a

$$\mu(A_1 \cup \ldots \cup A_n) \le \mu(A_1) + \ldots + \mu(A_n) \le \sum_n \mu(A_n)$$

et on utilise ensuite le point 4 avec

$$D_n = A_1 \cup \ldots A_n.$$

On écrit $B_n = A_{n+1} \setminus A_n$ qui sont des ensembles disjoints et $A_0 = \emptyset$ on a

$$\bigcup_{n\in\mathbb{N}} A_n = \bigcup_{n\in\mathbb{N}} B_n$$

et on a

$$A_n = \bigcup_{k=0}^{n-1} B_k$$

et donc

$$\mu(A_n) = \sum_{k=0}^{n} \mu(B_k)$$

d'où le résultat 4.

Pour la propriété suivante passons àu complémentaire. Posons pour $n > n_o$, $C_n = A_{n_o} \setminus A_n = A_{n_o} \cap A_n^c$. On a que

$$\mu(C_n) = \mu(A_{n_o}) - \mu(A_n)$$

car $\mu(A_{n_o}) < +\infty$. On a que $C_n \subset C_{n+1}$ ainsi

$$\lim \mu(C_n) = \mu(\bigcup C_n).$$

Remarquons que

$$\bigcap_{n\in\mathbb{N}} A_n = \bigcap_{n=n_o}^{+\infty} A_n = A_{n_o} \cap \bigcap_{n>n_o} A_n$$

et

$$\bigcup_{n=n_o}^{+\infty} C_n = A_{n_0} \cap \bigcup_{n>n_o} A_n^c = A_{n_o} \cap \left(\bigcap_{n>n_o} A_n\right)^c = A_{n_o} \setminus \left(\bigcap_{n>n_o} A_n\right)$$

et donc à nouveau $\mu(\bigcup_{n=n_o}^{+\infty} C_n) = \mu(A_{n_o}) - \mu(\bigcap_{n>n_o} A_n)$. On a alors

$$\lim \mu(A_n) = \mu(\bigcap_{n\in\mathbb{N}} A_n)$$

□

Proposition 11. (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 faÃ§on unique.

Théorème 12. (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 $\mu = \nu$.

Démonstration. $\mathcal{E}$ est une classe stable par intersection et on a que

$$\mathcal{E} \subset \{A \in \mathcal{A} | \mu(A) = \nu(A)\} \subset \mathcal{A}$$

Or $\{A \in \mathcal{A} | \mu(A) = \nu(A)\}$ est une c.m donc contient $\mathcal{M}(\mathcal{E}) = \sigma(\mathcal{E})$ par le lemme des classes monotones. □

Remarque 13. En particulier $\lambda(]a, b]) = b - a$ définit de manière unique la mesure de Lebesgue sur tout compact.

Remarque 14. La mesure produit est définie de manière unique.

Théorème 13. (Prolongement de Carathéodory)
Soit $\mathcal{C}$ une algèbre de Boole de parties de Ω. Soit $\mu : \mathcal{C} \to \mathbb{R}^+ \cup \{\infty\}$ telle que
— $\mu(\emptyset) = 0$,
— *$A_1, \ldots, A_n \in \mathcal{C}$ disjoints $\Rightarrow \mu(A_1 \cup \ldots \cup A_n) = \sum_{i\le n} \mu(A_i)$,*

— *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 15. Applications : construction de la mesure de Lebesgue sur $\mathbb{R}$, construction de mesures produits.

Chapitre 2

Mesures de probabilité, variables aléatoires.

2.1 Mesures de probabilités, lois

Définition 13. 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 16. Il n'y a pas que des mesures discrètes ou à densité !

Définition 14. 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 14. *(On spécifie les propriétés des mesures dans le cas de probabilités.)*

— $A \subset B \Rightarrow P(A) \leq P(B)$

— $P(A^c) = 1 - P(A)$

— $P(A \cup B) = P(B) + P(B) - P(A \cap B)$

— $P(\bigcup_{n \in \mathbb{N}} A_n) \leq \sum_{n \in \mathbb{N}} P(A_n)$.

— *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 15. 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).

Proposition 15. *Soit X une v.a à valeurs dans $(E, \mathcal{A})$ et Y une v.a à valeurs dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Alors Y est mesurable par rapport à $\sigma(X)$ ssi il existe une fonction $f : (E, \mathcal{A}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ telle que $Y = f(X)$.*

Démonstration. Commençons par les fonctions étagées. Si Y est étagée i.e

$$Y = \sum_{i=1}^{k} \alpha_i \mathbf{1}_{A_i}$$

avec $A_i \in \sigma(X)$ donc il existe B_i tel que $A_i = X^{-1}(B_i)$ d'où

$$Y = \sum_{i=1}^{k} \alpha_i \mathbf{1}_{X^{-1}(B_i)} = \sum_{i=1}^{k} \alpha_i \mathbf{1}_{B_i} \circ (X).$$

Supposons maintenant que Y est positive. Il existe une suite croissante (Y_n) de fonction étagées telle que

$$Y = \lim_n Y_n$$

Donc

$$\begin{aligned} Y &= \lim_n Y_n \\ &= \limsup_n (Y_n) \\ &= \limsup_n (f_n(X)) \\ &= (\limsup_n f_n)(X) \end{aligned}$$

On pose $\tilde{f} = (\limsup_n f_n) : E \to [0, +\infty]$ et $f = \tilde{f} \times (\mathbf{1}_{\mathbb{R}^+} \circ \tilde{f}) : E \to \mathbb{R}^+$ et on vérifie que $Y = f(X)$ $(Y = Y \times \mathbf{1}_{\mathbb{R}^+} \circ (Y))$.

Dans le cas général on écrit $Y = Y^+ - Y^-$ avec $Y^+ = \max(Y, 0)$ et $Y^- = -\min(Y, 0)$ □

Remarque 17. Les variables $\sigma(X)$-mesurables sont celles auxquelles on a accès connaissant la v.a X

2.2 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 16. *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 $Ef(X)$.

Démonstration. On suppose d'abord $\varphi \geq 0$. Dans le cas ou φ est étagée on a que

$$\varphi(X) = \sum_i \alpha_i \mathbf{1}_{B_i}(X)$$

Ainsi

$$\begin{aligned} E\varphi(X) &= \sum_i \alpha_i E\mathbf{1}_{B_i}(X) \\ &= \sum_i \alpha_i E\mathbf{1}_{B_i}(X) \\ &= \sum_i \alpha_i P(X \in B_i) \\ &= \sum_i \alpha_i P_X(B_i) \\ &= \int \sum_i \alpha_i \mathbf{1}_{B_i} dP_X \end{aligned}$$

Sinon φ est limite croissante de fonctions étagées i.e

$$\varphi = \lim_n \varphi_n$$

Donc par convergence monotone 2 fois

$$\begin{aligned} E\varphi(X) &= E \lim_n \varphi_n(X) \\ &= \lim_n E\varphi_n(X) \\ &= \lim_n \int \varphi_n dP_X \\ &= \int \lim_n \varphi_n dP_X \\ &= \int \varphi dP_X \end{aligned}$$

Dans le cas général on décompose $\varphi = \varphi^+ - \varphi^-$. Puis φ est intégrable ssi $\varphi+$ et φ^- le sont. □

Exemple 18. Soit X une v.a discrète i.e $X : (\Omega, \mathcal{A}) \to S$ avec S finie ou dénombrable alors

$$Ef(X) = \sum_{i \in S} f(i)P(X = i)$$

Espérance d'une Bernoulli, d'une binomiale, d'une Poisson, d'une géometrique

Exemple 19. Soit X une v.a uniforme sur $[0, 1]$. Soit X une v.a exponentielle de paramètre λ. Calculer l'espérance...

Remarque 20. La remarque qui suit est un outil majeur permettant de déterminer P_X lorsque celui-ci n'est pas donné.

Proposition 17. *Soit X une v.a alors X a pour loi μ ssi pour toute fonction g continue bornée*

$$\mathbb{E}(g(X)) = \int g(x) d\mu(x)$$

Démonstration. On note $\mathcal{H} = \{\phi$ borélienne bornée $\mid \int \phi(x)dP_X(x) = \int \phi(x)d\mu(x)\}$ l'espace vectoriel monotone contenant $\mathcal{E} = \{f$ fonctions continues bornées$\}$ qui est stable par produit. Donc $\mathcal{H}$ contient $\sigma(\mathcal{E})$. Ainsi les fonctions $\phi_i^{(n)} : (x_i) \to -n\mathbf{1}_{x_i<-n} + x_i\mathbf{1}_{-n\leq x_i\leq n} + n\mathbf{1}_{x_i>n}$ qui sont continues convergent simplement vers $\pi_i = (x_i) \to x_i$ qui est donc $\sigma(\mathcal{E})$ mesurable. Donc pour tout B_i borélien $\pi_i^{-1}(B_i) = \mathbb{R} \times \ldots \times B_i \times \ldots \times \mathbb{R} \in \sigma(\mathcal{E})$. La collection de tels événements lorsque $i = 1, \ldots, d$ engendre $\mathcal{B}(\mathbb{R}^d)$ donc $\mathcal{B}(\mathbb{R}^d) \subset \sigma(\mathcal{E})$ et donc $P_X = \mu$. $\square$

Exemple 21. Soit $X \sim \mathcal{N}(0,1)$. Déterminer la loi de X^2, e^X.

Exemple 22. Soit $X \sim \mathcal{U}([0,1])$. Déterminer la loi de $-\ln(U)/\lambda$ avec $\lambda > 0$.

Remarque 23. On peut restreindre encore la classe des fonctions avec par exemple les fonctions lipschitziennes bornées, fonctions $\mathcal{C}^k$ bornées, fonctions C^k à support compact.....

Proposition 18. (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 $L^q(\Omega, \mathcal{A}, P) \subset L^p(\Omega, \mathcal{A}, P)$.*

Remarque 24. Attention ceci n'est vraie que parce que la mesure est finie. Trouver des contre-exemples dans le cas de la mesure de Lebesgue.

Définition 16. $X : (\Omega, \mathcal{A}, P) \to (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ 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

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

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

Définition 17. 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 19. $F_X = F_Y \Rightarrow P_X = P_Y$.

Démonstration. Considérons

$$\mathcal{C} = \{\mathbb{R},]-\infty, x], x \in \mathbb{R}\}$$

c'est une famille stable par intersection et donc par le résultat sur les c.m comme P_X et P_Y coincident sur $\mathcal{C}$ alors elles coincident sur $\sigma(\mathcal{C})$. or $\mathcal{B}(\mathbb{R}) = \sigma(\mathcal{C})$. $\square$

Proposition 20. — *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.*

Démonstration. Soient $t \leq t'$ alors $]-\infty, t] \subset]-\infty, t']$ et donc $F_X(t) \leq F_X(t')$ d'où F_X est croissante. Soit (t_n) une suite croissante telle que $t_n \to +\infty$ alors $]-\infty, t_n] \subset]-\infty, t_{n+1}]$ et donc

$$\lim_n F_x(t_n) = P_X(\bigcup]-\infty, t_n]) = P_X(\mathbb{R}) = P(X \in \mathbb{R}) = 1$$

Pour la limite en 0 on utilise une suite décroissante qui tend vers $-\infty$.

Pour la continuité à droite en t_0 on considère (t_n) qui décroit vers t_0 on alors

$$\bigcup]-\infty, t_n] =]-\infty, t_0]$$

et on utilise la limite croissante i.e

$$\lim F_X(t_n) = P_X(\bigcup]-\infty, t_n]) = P_X(]-\infty, t_0]) = F_X(t_0).$$

D'où $F_X(t_0-) = F_X(t_0)$. Pour la limite à gauche soit (t_n) une suite croissante vers t_0 avec $t_n > t_0$ on a donc

$$\bigcup]-\infty, t_n] =]-\infty, t_0[.$$

Par limite décroissante $F_X(t_0+) = P_X(]-\infty, t_0[)$. Ainsi F_X est continue en t_0 ssi $P_X(]-\infty, t_0[) = P_X(]-\infty, t_0])$ i.e $P_X(\{t_0\}) = 0$.

Notons $\mathcal{D}$ l'ensemble des points de discontinuité de F_X. Soient $a > b$ deux points de $\mathcal{D}$. Considérons les intervalles non vides $]F_X(a-), F_X(a+)[$ et $]F_X(b-), F_X(b+)[$. La croissance de F_X implique que ces intervalles sont disjoints. On peut donc sélectionner un rationnel dans et injecter $\mathcal{D}$ dans $\mathbb{Q}$ d'où la dénombrabilité. $\square$

Exemple 25. Uniforme, Bernoulli, $\mathcal{E}(\lambda), \lambda > 0$.

Définition 18. 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 21. *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 $F_X^{\leftarrow}(U) \sim X$).*

Démonstration. Soit $u \in]0,1[$ et soit $t \in \mathbb{R}$ si $u \leq F(t)$ par définition de $F^{\leftarrow}(u)$ on a que $t \geq F^{\leftarrow}(u)$. Réciproquement si $t \geq F^{\leftarrow}(u)$ on a deux possibilités $t > F^{\leftarrow}(u)$ par définition de l'inf il existe (t_n) qui appartient à $\inf\{x, F(x) \geq u\}$ et qui décroit vers $F^{\leftarrow}(u)$. A partir d'un certain rang on a que $t > t_n$ et donc $F_X(t) \geq F_X(t_n) \geq u$ d'où $F_X(t) \geq u$. Si $t = F^{\leftarrow}(u)$, on considère à nouveau une suite (t_n) qui tend vers t telle que $F_X(t_n) \geq u$ par continuité à droite on a alors que $F_x(t_n) \to F_X(t)$ et donc $F_X(t) \geq u$

Soit $t \in \mathbb{R}$ alors

$$P(F^{\leftarrow}(U) \leq t) = P(U \leq F(t)) = F(t)$$

et donc on a le résultat. Pour le dernier point on utilise la caractérisation de la loi par la fonction de répartition. □

Proposition 22. *P_X est la dérivée de F_X au sens des distributions.*

Si F_X est de classe $\mathcal{C}^1$ alors $P_X(dt) = F'_X(t)\,dt$.

2.4 Fonction caractéristique

Définition 19. 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 20. Soit $X = (X_1, \dots, X_d)$ une v.a. à valeurs dans $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. Sa *fonction caractéristique* est définie pour $\xi = (\xi_1, \dots, \xi_d) \in \mathbb{R}^d$ 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 23. *La transformée de Fourier est injective sur les mesures de probabilité boréliennes de $\mathbb{R}^d$:*

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

Démonstration. Pour cela on va utiliser la version fonctionnelle des classes monotones. On considère

$$\mathcal{H} = \{\phi \text{ borélienne bornée} | E(\phi(X)) = E(\phi(Y))\}$$

C'est un e.v monotone par convergence dominé. On a les égalités suivante pour tout x et a

$$\begin{aligned} \cos(\langle a, x\rangle) &= \frac{e^{i\langle a,x\rangle} + e^{-i\langle a,x\rangle}}{2} \\ \sin(\langle a, x\rangle) &= \frac{e^{i\langle a,x\rangle} - e^{-i\langle a,x\rangle}}{2i}. \end{aligned}$$

Les fonctions qui en découlent sont donc des éléments de $\mathcal{H}$. Donc $\mathcal{H}$ contient l'espace vectoriel $\mathcal{C}$ engendré par les fonctions

$$\{\cos(\langle a, x\rangle), \sin(\langle a, x\rangle), a \in \mathbb{R}\}.$$

L'e.v $\mathcal{C}$ est stable par multiplication (formule de trigo !), donc $\mathcal{H}$ contient les fonctiosn mesurables pour $\sigma(\mathcal{C})$. pour conclure, il suffit de montrer que

$$\mathcal{B}(\mathbb{R}^d) \subset \sigma(\mathcal{C}).$$

Considérons l'application

$$\begin{array}{ccc} \mathbb{R}^d & \to & \mathbb{R} \\ (x_i) & \mapsto & n\sin(\frac{x_i}{n}) \end{array}$$

qui est $\sigma(\mathcal{C}))$ mesurable. Par limite la fonction $\pi_i : (x_i) \leftarrow x_i$ est alors $\sigma(\mathcal{C})$ mesurable. Ainsi soit B^i un borélien on a que

$$\pi_i^{-1}(B_i) = \mathbb{R} \times \ldots \times B_i \times \ldots \times \mathbb{R} \in \sigma(\mathcal{C}).$$

Il s'agit d'un système de générateur de $\sigma(\mathbb{R}^d)$ et donc on a le résultat. □

Démonstration. Voilà une preuve alternative en dimension 1. Soit X une variable aléatoire réelle. On note P_X sa loi, Φ_X sa fonction caractéristique et F_X sa fonction de répartition.

1. Pour tous $a \neq b \in \mathbb{R}$ et $T > 0$, on a

$$\frac{1}{2\pi}\int_{-T}^{T} \frac{e^{-ita} - e^{-itb}}{it}\Phi_X(t)\,dt$$
$$= \int_{\mathbb{R}} \left(\frac{1}{\pi}\int_0^T \frac{\sin\big(t(x-a)\big)}{t} dt - \frac{1}{\pi}\int_0^T \frac{\sin\big(t(x-b)\big)}{t} dt \right) P_X(dx).$$

2. Par une IPP on montre que $\lim_{T\to\infty} \int_0^T \frac{\sin t}{t} dt = \int_0^{+\infty} \frac{1-\cos t}{t^2}\, dt \in]0, +\infty[$. On admettra pour la suite que cette limite vaut $\frac{\pi}{2}$. On en déduit que $x \in \mathbb{R}$,

$$\lim_{T\to\infty} \int_0^T \frac{\sin(tx)}{t}\, dt = \frac{\pi}{2}\text{signe}(x),$$

où $\text{signe}(x) = 1$ si $x > 0$, $\text{signe}(x) = -1$ si $x < 0$ et $\text{signe}(0) = 0$.

3. Pour $a < b$ on a donc

$$\lim_{T\to+\infty} \frac{1}{2\pi}\int_{-T}^{T} \frac{e^{-ita} - e^{-itb}}{it}\Phi_X(t)\,dt = P_X\big(]a, b[\big) + \frac{P_X(\{a\}) + P_X(\{b\})}{2}.$$

4. Ainsi si F_X est continue en a et b, on a $P_X(\{a\}) = P_X(\{b\}) = 0$ et par suite

$$F_X(b) - F_X(a) = \lim_{T\to+\infty} \frac{1}{2\pi}\int_{-T}^{T} \frac{e^{-ita} - e^{-itb}}{it}\Phi_X(t)\,dt.$$

Ainsi en tout point de continuité de F_X et F_Y on a que

$$F_X(b) - F_X(a) = F_Y(b) - F_Y(a)$$

Ainsi $F_X = F_Y$ en tout point de continuité de F_X et F_Y. Les points de discontinuité de F_X ou F_Y sont en nombre dénombrable donc si t_0 est l'un de ces points on peut trouver une suite (t_n) avec $t_n > t_0$ pour tout n et (t_n) sont des points de continuitÃ© de F_X et F_Y. On a donc par continuité à droite

$$F_X(t_0) = \lim F_X(t_n) = \lim F_Y(t_n) = F_Y(t_0).$$

Donc $F_X = F_Y$ sur tout $\mathbb{R}$ et donc $P_X = P_Y$.

□

Exemple 26. Uniforme, Poisson, Binomiale.

Exemple 27. Soit $X \sim \mathcal{N}(0,1)$, on a que pour tout $t \in \mathbb{R}$

$$\Phi_X(t) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{itx} e^{-x^2/2} dx$$

On vérifie que Φ_X est $\mathcal{C}^1$ et $\Phi'_X(t) = -t^2 \Phi_X(t)$ (on dérive sous le signe intégrale et on fait une IPP) et donc $\Phi_X(t) = e^{-t^2/2}$.

Si $X \sim \mathcal{N}(m, \sigma^2)$ alors $Z = (X-m)/\sigma \sim \mathcal{N}(0,1)$

$$\Phi_X(t) = E(e^{itX}) = E(e^{it(\sigma Z + m)}) = e^{imt - \frac{\sigma^2 t^2}{2}}$$

Théorème 24. *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*

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

Démonstration. On va considérer l'ensemble des fonctions continues à support compact et montrer que pour tout ϕ dans cet ensemble

$$\int \phi(t) dP_X(t) = \int \phi(t) f_X(t) dt$$

ce qui par la version fonctionnelle des c.m suffit à conclure. Introduisons $\gamma_{n,t}$ la densité d'une gaussienne de variance $\frac{1}{n^2}$ centrée en t i.e

$$\gamma_{n,t}(y) = \frac{1}{(\sqrt{2\pi}/n)^d} e^{\frac{-|y-t|^2}{2/n^2}}$$

On a que (changement de varaiable plus convergence dominée)

$$\phi(t) = \lim_n \int \phi(y) \gamma_{n,t}(y) dy$$

Ainsi

$$\begin{aligned}
\int \phi(t) dP_X(t) &= \int \lim_n \int \phi(y)\gamma_{n,t}(y) dy dP_X(t) \\
&= \lim_n \int \int \phi(y)\gamma_{n,t}(y) dy dP_X(t) \quad \left(\left| \int \phi(y)\gamma_{n,t}(y) dy \right| \leq \|\phi\|_\infty \right) \\
&= \lim_n \int \phi(y) \int \gamma_{n,t}(y) dP_X(t) dy \quad \text{(Fubini avec la même raison que ci-dessus)}
\end{aligned}$$

A ce stade on veut à nouveau intervertir limite et intégrale. Pour cela on utilise l'inversion de la transformée de Fourier d'une gaussienne c'est à dire

$$\gamma_{n,t}(y) = \frac{1}{(2\pi)^d} \int e^{i\langle \zeta, t-y\rangle} e^{-|\zeta|^2/2n^2} d\zeta$$

(ce résultat est une généralisation multi d de la fonction caractéristique de la gaussienne $\mathcal{N}(0,1)$).

Maintenant on peut observer que

$$\begin{aligned}
\int \gamma_{n,t}(y) dP_X(t) &= \int \frac{1}{(2\pi)^d} \int e^{i\langle \zeta, t-y\rangle} e^{-|\zeta|^2/2n^2} d\zeta dP_X(t) \\
&= \int \frac{1}{(2\pi)^d} \int e^{i\langle \zeta, t-y\rangle} e^{-|\zeta|^2/2n^2} dP_X(t) d\zeta \quad (Fubini) \\
&= \frac{1}{(2\pi)^d} \int e^{-i\langle \zeta, y\rangle} e^{-|\zeta|^2/2n^2} \Phi_X(\zeta) d\zeta
\end{aligned}$$

Par suite on a que

$$\left|\phi(y) \int \gamma_{n,t}(y) dP_X(t)\right| \leq \frac{1}{(2\pi)^d} |\phi(y)| \int |\Phi_X(\zeta)| d\zeta$$

Cette borne est indépendante de n et on a que Φ_X est supposé L^1 et ϕ est à support compact donc on peut appliquer le théorème de convergence dominée. On a

$$\begin{aligned}
\int \phi(t) dP_X(t) &= \int \phi(y) \lim_n \int \gamma_{n,t} dP_X(t) dy \\
&= \int \phi(y) \lim_n \frac{1}{(2\pi)^d} \int e^{-i\langle \zeta, y\rangle} e^{-|\zeta|^2/2n^2} \Phi_X(\zeta) d\zeta dy \\
&= \int \phi(y) \frac{1}{(2\pi)^d} \int e^{-i\langle \zeta, y\rangle} \Phi_X(\zeta) d\zeta d
\end{aligned}$$

d'où le résultat □

Exemple 28. Une loi de Cauchy de paramètre a à pour fonction caractéristique $t \to e^{-a|t|}$

Proposition 25. *Pour toute v.a.r. X,*

- *Φ_X est continue,*
- *Si $E|X|^n < \infty$ alors Φ_X est de classe $\mathcal{C}^n$ 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 $2\left\lfloor \frac{n}{2} \right\rfloor$.*

Démonstration. Pour montrer cela on utilise le théorème de continuité et de dérivation sous le signe intégrale.

Pour la réciproque, faisons le cas $n = 2$ on suppose donc que $\Phi_X(t)$ est 2 fois dérivable en 0. On peut donc écrire

$$\Phi_X(t) = 1 + t\Phi'_X(0) + \frac{t^2}{2}\Phi''_X(0) + \circ(t^2)$$
$$\Phi_X(-t) = 1 - t\Phi'_X(0) + \frac{t^2}{2}\Phi''_X(0) + \circ(t^2)$$

Donc on a

$$\Phi_X(t) + \Phi_X(-t) = E(2\cos(Xt))) = 2 + t^2\Phi''_X(0) + \circ(t^2)$$

Ainsi on a que pour t suffisamment petit

$$E(2 - \frac{2\cos(Xt)}{t^2}) = -\Phi''_X(0) + \circ(1)$$

On peut alors appliquer le Lemme de Fatou avec une suite (t_n) qui tend vers 0.

Lemme 26. *Soit $(E, \mathcal{A}, \mu)$ un espace mesuré et soit (f_n) une suite de fonctions positives alors*

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

On a alors

$$E\left[\liminf \frac{2 - 2\cos(Xt_n)}{t_n^2}\right] \leq \liminf E\left[\frac{2\cos(Xt_n) - 2}{t_n^2}\right] = -\Phi_X''(0)$$

Or on a que

$$\liminf \frac{2 - 2\cos(Xt_n)}{t_n^2} = X^2$$

et donc

$$EX^2 \leq -\Phi_X''(0)$$

□

2.5 Transformée de Laplace

Définition 21. 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 27. *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*

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

Proposition 28. *Soit X une v.a.r. telle que $\forall t \in [-\epsilon, \epsilon]$, $Ee^{tX} < \infty$. Alors*

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

Théorème 29. *Soient X, Y deux v.a. à valeurs dans $\mathbb{R}^d$. Soit $\epsilon > 0$. Si pour tout s tel que $\|s\| \leq \epsilon$ on a $L_X(s) = L_Y(s) < +\infty$, alors $P_X = P_Y$.*

Démonstration. Traitons le cas de la dimension $d = 1$. On peut remarquer que les fonctions

$$s \mapsto E(e^{sX}), \quad s \mapsto E(e^{sY})$$

sont définies pour tout complexe s tels que $|Re(s)| < \epsilon$.

Considérons s_0 tel que $Re(s_0) = 0$ et s tel que $|s| < \epsilon$. En remarquant que $|e^{s_0X}| = 0$ et $|e^{sX}| = |e^{|\Re(s)X|}|$ est inégrable, on a en utilisant Fubini

$$E[e^{(s_0+s)X}] = \sum_n \frac{s^n}{n!} E(e^{s_0X} X^n).$$

Il s'agit donc d'un développement en série entière valable sur $\mathring{B}(0, \epsilon)$. On a donc que $s \mapsto E(e^{sX})$ est analytique sur $\bigcup_{s_0 \in i\mathbb{R}} \mathring{B}(s_0, \epsilon) = \{s \in \mathbb{C} : |Re(s)| < \epsilon\}$. On a également que $s \mapsto E(e^{sY})$ est analytique. On en déduit que $s \mapsto E(e^{sY})$ et $s \mapsto E(e^{sX})$ coincident sur $\{s \in \mathbb{C} : |Re(s)| < \epsilon\}$ et donc pour $s = it$ avec $t \in \mathbb{R}$ on a que $\Phi_X = \Phi_Y$. □

Exemple 29. Soit $X \sim \mathcal{N}(0,1)$, soit $z \in \mathbb{R}$ on a que

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

en remarquant qu'on a fait apparaître la densité de $\mathcal{N}(z,1)$.

En utilisant le raisonnement sur l'analycité on retrouve la valeur de la fonction caractéristique de $\mathcal{N}(0,1)$.

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

Théorème 30. *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 $\mu = \nu$.

2.6 Moments

Définition 22. 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

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

Proposition 31. *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 $n \geq 1$, $E\big(X^{2n}\big)^{\frac{1}{2n}} \leq Cn$.*

Exemple 30. Toute v.a à support compact

Exemple 31. La gaussienne $\mathcal{N}(0,1)$.

$$E(X^{2k}) = \frac{(2k)!}{2^k k!}$$

or avec l'équivalent de Stirling $n! \sim \sqrt{2\pi n}\left(\frac{n}{e}\right)^n$ on a

$$\begin{aligned} E(X^{2k}) &\sim \frac{\sqrt{2\pi 2k}\left(\frac{2k}{e}\right)^{2k}}{2^k\sqrt{2\pi k}\left(\frac{k}{e}\right)^k} \\ &\sim \sqrt{2}2^k\left(\frac{k}{e}\right)^k \end{aligned}$$

Ainsi

$$\lim_k \frac{(E(X^{2k}))^{1/2k}}{k} = 0$$

La première propriété était également valable.

2.7 Fonctions génératrices

Définition 23. Si X est à valeurs dans $\mathbb{N}$ on définit sa fonction génératrice

$$\begin{array}{cccc} G_X : & [0,1] & \to & [0,1] \\ & s & \mapsto & E(s^X) = \sum_n s^n P(X=n) \end{array}$$

la proposition suivante est basée sur l'unicité du développement en série entière

Proposition 32. *Si X et Y sont à valeurs dans $\mathbb{N}$ alors $G_X = G_Y$ implique que $P_X = P_Y$.*

Chapitre 3

Rappels sur la notion d'indépendance

3.1 Indépendance

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

Définition 24. (Indépendance d'événements)

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

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

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

Définition 25. (Indépendance de sous-tribus)

— Des sous-tribus $\mathcal{B}_1, \ldots, \mathcal{B}_n$ 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 33. *Soient $\mathcal{B}_1, \ldots, \mathcal{B}_n$ 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.

Démonstration. On pose

$$\mathcal{E}_1 = \{G_1 \in \mathcal{B}_1 \ tq \ \forall j \geq 2, F_j \in \mathcal{F}_j \ P(G_1 \cap F_2 \cap \ldots \cap F_n) = P(B_1) \times \prod_{i=2}^n P(F_i)\}$$

On a que

$$\mathcal{F}_1 \subset \mathcal{E}_1 \subset \mathcal{B}_1$$

Or $\mathcal{E}_1$ est une c.m. Donc

$$\mathcal{M}(\mathcal{F}_1) \subset \mathcal{E}_1$$

et comme $\mathcal{F}_1$ est stable par intersection finie on a $\mathcal{M}(\mathcal{F}_1) = \sigma(\mathcal{F}_1)$. Or $\sigma(\mathcal{F}_1) = \mathcal{B}_1$ et donc $\forall G_1 \in \mathcal{B}_1, \forall F_j \in \mathcal{F}_j, j \geq 2$ $P(G_1 \cap F_2 \cap \ldots \cap F_n) = P(B_1) \times \prod_{i=2}^n P(F_i)$. On raisonne de même avec $\mathcal{F}_2$ en posant

$$\mathcal{E}_2 = \{G_2 \in \mathcal{B}_2 \ tq \ \forall G_1 \in \mathcal{B}_1 \forall j \geq 3, F_j \in \mathcal{F}_j \ P(G_1 \cap G_2 \cap \ldots \cap F_n) = P(G_1)P(G_2) \times \prod_{i=3}^n P(F_i)\}$$

et par récurrence on obtiendra le résultat. □

Proposition 34. (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.*

Démonstration. On fixe d'abord une sous famille finie d'indice dans J, notons J_f ce sous ensemble. Il faut montrer que les $\widetilde{\mathcal{B}}_j, j \in J_f$ sont indépendantes. On introduit pour $j \in J_f$

$$\mathcal{E}_j = \{A, A \text{ intersection finie d'ensembles } A_i \in \mathcal{B}_i \text{ pour } i \in I_j\}$$

Remarquons que $\widetilde{\mathcal{B}}_j = \sigma(\mathcal{E}_j)$, ensuite que les ensembles $\mathcal{E}_j$ satisfont la proposition précédente. □

Définition 26. (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 $X_1, \ldots, X_n$ 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 35. *Des v.a. $X_1, \ldots, X_n$ sont indépendants si et seulement si*

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

Dans ce cas :
- *$\forall f_i$ mesurables positives : $E\big(\prod_{i=1}^n f_i(X_i)\big) = \prod_{i=1}^n E f_i(X_i)$.*
- *Si f_i mesurable de signe quelconque, l'égalité précédente est vérifiée si*

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

Démonstration. On vérifie l'égalité des lois $P_{(X_1,\ldots,X_n)} = P_{X_1} \otimes \cdots \otimes P_{X_n}$ sur les pavés et ensuite on utilise le théorème de caractérisation des mesures finies. □

Proposition 36. *Soient $X_1, \ldots, X_n$ des variables aléatoires réelles. Les assertions suivantes sont équivalentes :*
- *$X_1, \ldots, X_n$ sont indépendantes.*
- *$\forall \xi \in \mathbb{R}^n$, $\Phi_{(X_1,\ldots,X_n)}(\xi_1, \ldots, \xi_n) = \prod_{i=1}^n \Phi_{X_i}(\xi_i)$.*
- *$\forall a_1, \ldots, a_n \in \mathbb{R}$, $P(X_1 \leq a_1, \ldots, X_n \leq a_n) = \prod_{j=1}^n P(X_j \leq a_j)$.*
- *$\forall f_i : \mathbb{R} \to \mathbb{R}^+$, continues bornées, $E\Big(\prod_{i=1}^n f_i(X_i)\Big) = \prod_{i=1}^n E f_i(X_i)$.*

Démonstration. Pour la fonction caractéristique c'est évident car la fonction caractéristique caractérise la loi. Pour le second point on raisonne comme pour la fonction de répartition. Pour le troisième point il suffit d'appliquer la version fonctionnelle du lemme des classes monotones. $\square$

Exemple 32. Soit X et Y deux v.a $\mathcal{N}(0,1)$ alors $U = X+Y$ et $V = X-Y$ sont indépendantes. Pour cela utilisons deux méthodes.

La première utilise la fonction caractéristique. Soit $(u,v) \in \mathbb{R}^2$ on a

$$\begin{aligned}
\phi_{(U,V)}(u,v) &= E(e^{i(uU+vV)}) \\
&= E(e^{i((u+v)X+(u-v)Y)}) \\
&= E(e^{i((u+v)X)}E(e^{i((u-v)Y)} \quad \text{par indépendance} \\
&= e^{-(u+v)^2/2}e^{-(u-v)^2/2} \\
&= e^{-u^2}e^{-v^2}
\end{aligned}$$

La fonction caractéristique est à variables séparées donc les deux v.a sont indépendantes. Or la fonction $u \mapsto e^{-u^2}$ est la fonction caractéristique de $\mathcal{N}(0,2)$. On a donc que U et V ont même loi et que $U \sim \mathcal{N}(0,2)$.

La deuxième utilise le théorème du transfert. Soient f et g deux fonctions continues bornées. on a

$$\begin{aligned}
E(f(U)g(V)) &= E(f(X+Y)g(X-Y)) \\
&= \int_{\mathbb{R}^2} f(x+y)g(x-y)\frac{1}{2\pi}e^{-(x^2+y^2)/2}dxdy \quad \text{par independance de X et Y} \\
&= \int_{\mathbb{R}^2} f(u)g(v)\frac{1}{2\pi}e^{-(((u+v)/2)^2+((u-v)/2)^2)/2}\frac{1}{2}dudv \quad \text{changment de variables u=x+y, v=x-y} \\
&= \int_{\mathbb{R}^2} f(u)g(v)\frac{1}{2\pi}e^{-(u^2+v^2)/4}\frac{1}{2}dudv \\
&= \int_{\mathbb{R}^2} f(u)g(v)\frac{1}{2\sqrt{\pi}}e^{-(u^2)/4}\frac{1}{2\sqrt{\pi}}e^{-(v^2)/4}dudv
\end{aligned}$$

On identifie alors les densités de U et de V qui sont les mêmes et on retrouve le résultat sur le fait que U et V sont indépendantes et de distribution $\mathcal{N}(0,2)$.

Exemple 33. Soit X et Y deux v.a $\mathcal{N}(0,1)$ alors il existe des unique v.a $R>0$ et $\Theta \in]0,2\pi[$, telles que

$$X = R\cos(\Theta), \quad Y = R\sin(\Theta)$$

Alors R et Θ sont indépendantes. On note $\phi(r,\theta) = (r\cos(\theta); r\sin(\theta))$ le changement de coordonnée polaire qui est un isomorphisme de $\mathbb{R}_+^* \times]0,2\pi]$ dans $\mathbb{R}^2 \setminus (0,0)$. Soit $f(r,\theta) = h(r)g(\theta)$ avec h et g deux fonctions continues bornées, on a

$$\begin{aligned}
E(f(R,\Theta)) &= E(f \circ \phi^{(-1)}(X,Y)) \\
&= \int_{\mathbb{R}^2} f \circ \phi^{(-1)}(x,y)\frac{1}{2\pi}e^{-(x^2+y^2)/2}dxdy \\
&= \int_0^\infty \int_0^{2\pi} f(r,\theta)\frac{1}{2\pi}e^{-r^2/2}rdrd\theta \quad \text{changement de v.a } (r,\theta) = \phi^{(-1)}(x,y) \\
&= \int_0^\infty f(r)e^{-r^2/2}rdr \int_0^{2\pi} \frac{1}{2\pi}g(\theta)d\theta
\end{aligned}$$

On voit donc que R et Θ sont indépendantes et que la densité de R est $r \mapsto e^{-r^2/2}r\mathbf{1}_{\mathbb{R}^+}$ et que Θ suit une loi uniforme sur $[0,2\pi]$.

3.2 Sommes de variables aléatoire indépendantes

Définition 27. 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 34. Des variables indépendantes sont non corrélées. La réciproque est fausse. Soit ε telle que $P[\varepsilon = 1] = P[\varepsilon = -1] = 1/2$. Soit X une v.a indépendante de ε alors X et εX ne sont pas indépendante (?) mais $\mathrm{Cov}(X, \varepsilon X) = 0$.

Proposition 37. *Soit $X_1, \ldots, X_n$ 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 28. 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 35. 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 38. *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 36. **(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 $\mathcal{N}(m, \sigma^2) * \mathcal{N}(m', \sigma'^2) = \mathcal{N}(m + m', \sigma^2 + \sigma'^2)$.

3.3 Suites de tribus ou de v.a. indépendantes

Théorème 39. *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 $\forall B \in \mathcal{B}_\infty$, $P(B) \in \{0, 1\}$.*

Démonstration. On pose pour $k \geq 1$

$$\mathcal{D}_k = \sigma(\mathcal{T}_1, \ldots, \mathcal{T}_k)$$

On a que $\mathcal{D}_k$ est indépendante de $\mathcal{B}_{k+1}$ et donc de $\mathcal{B}_\infty$. Remarquons que $\bigcup \mathcal{D}_k$ forme une classe d'événements stable par intersection donc $\mathcal{B}_\infty$ est indépendante de $\sigma(\bigcup \mathcal{D}_k) \supset \mathcal{B}_\infty$ donc

$$\forall B \in \mathcal{B}_\infty,\ P(B) = P(B \cap B) = P(B)^2.$$

□

Exemple 37. Les événements suivants sont des événements de $\mathcal{B}_\infty$

— Une suite (X_n) de v.a.r converge

— La série $\sum X_k$ converge

Définition 29. Soit $(A_n)_{n\in\mathbb{N}}$ une suite d'événements de $(\Omega, \mathcal{A})$,

$$\begin{aligned}\limsup_n A_n &:= \bigcap_{n\in\mathbb{N}}\Big(\bigcup_{k\geq n} A_k\Big) = \{\omega \in \Omega;\ \text{pour une infinité de valeurs de } n,\ \omega \in A_n\},\\ \liminf_n A_n &:= \bigcup_{n\in\mathbb{N}}\Big(\bigcap_{k\geq n} A_k\Big) = \{\omega \in \Omega;\ \exists n_0,\ \forall n \geq n_0,\ \omega \in A_n\}.\end{aligned}$$

Exemple 38. 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

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

Lemme 40 (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.*

Démonstration. Pour le point 1) on a que pour tout $n \in \mathbb{N}$ $\bigcup_{k>n} A_k \supset \bigcup_{k>n+1} A_k$ donc

$$P(\bigcap_n \bigcup_{k>n} A_k) = \lim_n P(\bigcup_{k>n} A_k)$$

or

$$P(\bigcup_{k>n} A_k) \leq \sum_{k>n} P(A_k) \to 0$$

comme reste d'une série convergente.

Pour le point 2) on a

$$P(\limsup A_n) = \lim_n \lim_N P(\bigcup_{k=n}^{N} A_k)$$

Or on a

$$P(\bigcup_{k=n}^{N} A_k) = 1 - \prod_{k=n}^{N} (1 - P(A_k))$$

et

$$\log(\prod_{k=n}^{N} (1 - P(A_k))) = \sum_{k=n}^{N} \log(1 - P(A_k)) \leq -\sum_{k=n}^{N} P(A_k) \underset{N}{\to} -\infty$$

Ainsi

$$\prod_{k=n}^{N} (1 - P(A_k)) \underset{N}{\to} 0$$

et le résultat suit. □

Exemple 39. Une suite de v.a.r (X_n) indépendante converge vers 0 ssi

$$\sum \mathbb{P}(|X_n| > \varepsilon) < \infty$$

Dans ce cas que dire si $X_n \sim \mathcal{B}(p_n)$.

Chapitre 4

Calcul conditionnel

4.1 Conditionnement discret

4.1.1 Par rapport à un événement

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

Définition-Proposition 30. 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

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

2. La *mesure de probabilité conditionnelle sachant B* est l'application

$$\begin{array}{rcl} \mathcal{A} & \to & [0,1] \\ A & \mapsto & \dfrac{P(A \cap B)}{P(B)} = P(A|B). \end{array}$$

   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

$$E(Y|B) := \int_\Omega Y \, dP(\cdot|B) = \frac{E(Y 1_B)}{P(B)} = \frac{1}{P(B)} \int_B Y \, dP.$$

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

$$E\big(\varphi(Y)|B\big) = \int \varphi(Y) \, dP(\cdot|B) = \int \varphi \, dP_{Y|B}.$$

Démonstration. Pour les points 3) et 4) il s'agit de la construction usuelle des mesures. On considère d'abord $Y = \mathbf{1}_A$, puis ensuite les fonctions étagées, puis les fonctions positives et enfin $Y = Y^+ - Y^-$. $\square$

4.1.2 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 31. 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.

Remarque 40. Comme X est une v.a discrète on peut écrire

$$\varphi(X) = \sum_{x \in E} \varphi(x) \mathbf{1}_{X=x} = \sum_{x \in E} P\big(A|X = x\big) \mathbf{1}_{X=x}$$

Définition 32. 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* notée $E(Y|X)$ est la v.a. $\Psi(X)$.

Remarque 41. De même

$$\Psi(X) = \sum_{x \in E} \Psi(x) \mathbf{1}_{X=x} = \sum_{x \in E} E(Y|X = x) \mathbf{1}_{X=x}$$

Exemple 42. Soit X une v.a uniforme sur $\{-1, 1, 2\}$.

$$\begin{aligned} E(X||X|) &= E(X||X| = 1)\mathbf{1}_{|X|=1} + E(X||X| = 2)\mathbf{1}_{|X|=2} \\ &= \frac{E(X\mathbf{1}_{|X|=1})}{P(|X| = 1)}\mathbf{1}_{|X|=1} + \frac{E(X\mathbf{1}_{|X|=2})}{P(|X| = 2)}\mathbf{1}_{|X|=2} \\ &= 2\mathbf{1}_{|X|=2} \end{aligned}$$

Exemple 43. Soient X_1 et X_2 deux v.a indépendantes de loi de Poisson de paramètre $\lambda > 0$. On sait que $X_1 + X_2$ est une v.a de Poisson de paramètre 2λ.

$$
\begin{aligned}
E(X_1|X_1+X_2) &= \sum_{k=0}^{\infty} E(X_1|X_1+X_2=k)\mathbf{1}_{X_1+X_2} = k \\
&= \sum_{k=0}^{\infty} \frac{E(X_1\mathbf{1}_{X_1+X_2=k})}{P(X_1+X_2=k)}\mathbf{1}_{X_1+X_2} = k \\
&= \sum_{k=0}^{\infty}\sum_{j=0}^{k} \frac{E(j\mathbf{1}_{X_1=j}\mathbf{1}_{X_2=k-j})}{P(X_1+X_2=k)}\mathbf{1}_{X_1+X_2} = k \\
&= \sum_{k=0}^{\infty}\sum_{j=1}^{k} j\frac{k!}{j!(k-j)!}\mathbf{1}_{X_1+X_2} = k \\
&= \sum_{k=0}^{\infty}\sum_{j=1}^{k} \frac{k!}{(j-1)!(k-j)!2^k}\mathbf{1}_{X_1+X_2} = k \\
&= \sum_{k=0}^{\infty}\sum_{j=1}^{k} \frac{k!}{(j-1)!((k-1)-(j-1))!2^k}\mathbf{1}_{X_1+X_2} = k \\
&= \sum_{k=0}^{\infty} k\sum_{j=1}^{k} \frac{(k-1)!}{(j-1)!((k-1)-(j-1))!2^k}\mathbf{1}_{X_1+X_2} = k \\
&= \sum_{k=0}^{\infty} \frac{k}{2^k}\sum_{j=0}^{k-1} \frac{(k-1)!}{(j)!((k-1)-(j))!}\mathbf{1}_{X_1+X_2} = k \\
&= \sum_{k=0} \frac{k}{2^k}2^{k-1}\mathbf{1}_{X_1+X_2} = k \\
&= \sum_{k=0} \frac{k}{2}\mathbf{1}_{X_1+X_2} = k \\
&= \frac{X_1+X_2}{2}
\end{aligned}
$$

Lemme 41 (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 $Y = h(X)$.*

Remarque 44. Il est alors essentiel de remarquer que les espérances conditionnelles sachant une v.a X sont $\sigma(X)$ mesurables

Proposition 42. *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*

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

Réciproquement, si W est une v.a $\sigma(X)$ mesurable et que pour toute v.a. Z, $\sigma(X)-$mesurable et bornée

$$E\big(ZW\big) = E(ZY)$$

alors $E(Y|X) = W$ presque sûrement. De plus une telle v.a est unique.

Démonstration. Montrons la réciproque du point 2). Comme W est $\sigma(X)$ mesurable il existe h mesurable telle que $W = h(X) = \sum_{x\in E} h(x)\mathbf{1}_{X=x}$. On applique pour $Z = \mathbf{1}_{X=x}$ on a donc

$$\begin{aligned} E(\mathbf{1}_{X=x}Y) &= E\big(\mathbf{1}_{X=x}W\big) \\ &= E\big(\mathbf{1}_{X=x}\sum_{z\in E} h(z)\mathbf{1}_{X=z}\big) \\ &= h(x)E(\mathbf{1}_{X=x}) \\ &= h(x)P(X=x) \end{aligned}$$

Ainsi pour tout $x \in E$ on a

$$h(x) = E(\mathbf{1}_{X=x}Y)/P(X=x) = E(Y|\mathbf{1}_{X=x})$$

et le résultat en découle. Pour l'unicité, on considère deux v.a W et W', $\sigma(X)$ mesurables satisfaisant l'égalité, on a donc

$$E((W-W')Z) = 0$$

pour tout fonction Z $\sigma(X)$ mesurable bornée. On choisit $Z = \mathbf{1}_{W-W'>0}$ qui est $\sigma(X)$ mesurabme, on a donc que

$$E((W-W')\mathbf{1}_{W-W'\geq 0}) = 0.$$

On en déduit que $W - W' \| eq0$ p.s. Avec $Z' = \mathbf{1}_{W-W'<0}$ on en déduit que $W - W' \geq 0$ et donc que $W = W'$. □

Remarque 45. Lorsqu'on applique le point 2) avec $Z = \mathbf{1}_\Omega = 1$ on obtient

$$E(E(Y|X)) = E(Y).$$

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

Théorème 43 (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.

Remarque 46. L'égalité $E\big(ZE(Y|X)\big) = E(ZY)$ peut en effet se lire

$$E\big(Z(E(Y|X) - Y)\big) = 0$$

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

4.2.1 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 33. 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 47. $E[Y|\mathcal{B}]$ 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 44. *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.*

Démonstration. Les deux premiers points viennent de la définition en termes de projection. Pour le 3ème point, on considère $Z = \mathbf{1}_{E(Y|\mathcal{B})<0}$ on a

$$0 \leq E(YZ) = E(E(Y|\mathcal{B})Z) \leq 0$$

Donc $E(E(Y|\mathcal{B})Z) = 0$. Or $E(Y|\mathcal{B})Z \leq 0$ p.s. Ainsi $E(Y|\mathcal{B})Z = 0$ p.s et donc $E(Y|\mathcal{B}) = 0$ sur $\{Z = 1\}$ or $E(Y|\mathcal{B}) < 0$ sur $\{Z = 1\}$. Donc $P(Z = 1) = 0$. Le résultat en découle. Si $Y \geq Y'$ on applique le point précédent à $Y - Y' \geq 0$. $\square$

4.2.2 Extension aux variables aléatoires positives

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

Théorème 45 (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,* $E(ZY) = E\big(ZE(Y|\mathcal{B})\big)$*.*

Démonstration. On pose

$$E(Y|\mathcal{B}) = \lim E(\inf(Y, n)|\mathcal{B}).$$

Cette limite existe car la suite $(E(\inf(Y, n)|\mathcal{B}))$ est croissante. Soit Z une v.a $\mathcal{B}$ mesurable positive, alors pour tout $n \in \mathbb{N}$ les v.a $\inf(Z, n), \inf(Y, n)$ sont L^2 on a

$$E(\inf(Z, n) \inf(Y, n)) \quad = \quad E(\inf(Z, n)E(\inf(Y, n)|\mathcal{B}))$$

On peut alors utiliser le théorème de convergence monotone sur chaque membre et on obtient le résultat. Il reste à montrer l'unicité. Attention on ne peut pas considérer la v.a $Y - E(Y|\mathcal{B})$. En effet $E(Y|\mathcal{B})$ peut éventuellement être infini. Considérons $\tilde{Y}$ une v.a satisfaisant les deux points et posons

$$Z = \mathbf{1}_{\tilde{Y} \leq a < b \leq E(Y|\mathcal{B})}.$$

On a

$$\begin{aligned} aP(\tilde{Y} \leq a < b \leq E(Y|\mathcal{B})) = E(Z) &\geq E(\tilde{Y}Z) \\ &= E(E(Y|\mathcal{B})Z) \\ &\geq bP(\tilde{Y} \leq a < b \leq E(Y|\mathcal{B})) \end{aligned}$$

Comme $a < b$ on a $P(\tilde{Y} \leq a < b \leq E(Y|\mathcal{B})) = 0$ et ceci pour tout $a < b$. En considérant l'union sur a et b rationnels on a que $P(\bigcup_{a>b,(a,b)\in\mathbb{Q}}\{\tilde{Y} \leq a < b \leq E(Y|\mathcal{B})\}) = P(\tilde{Y} < E(Y|\mathcal{B})) = 0$. On fait de même pour montrer que $P(\tilde{Y} > E(Y|\mathcal{B})) = 0$. Ainsi $\tilde{Y} = E(Y|\mathcal{B})$ presque sûrement. $\square$

Proposition 46. *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.*

Démonstration. Si Y est $\mathcal{B}$-mesurable, il en est de même pour $\inf(Y, n)$ et donc par le cas L^2 on a $E(\inf(Y, n)|\mathcal{B}) = \inf(Y, n)$ puis on passe à la limite quand n tend vers l'infini. On a que $\inf(Y, n) \leq \inf(Y', n)$ puis passage à la limite. □

4.2.3 Définition pour des variables aléatoires dans L^1

Définition-Proposition 34. 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, $E\big(Z\ E(Y|\mathcal{B})\big) = E(ZY)$.

Démonstration. Si $Y \geq 0$ et L^1 on a $E(Y|\mathcal{B}) \geq 0$ et pour $Z = 1$ on a

$$E(E(Y|\mathcal{B})) = E(Y) < +\infty.$$

Donc $E(Y|\mathcal{B})$ est dans L^1. Ainsi $E(Y|\mathcal{B}) := E(Y_+|\mathcal{B}) - E(Y_-|\mathcal{B})$ a bien un sens dans L^1. Il est alors clair que

$$|E(Y|\mathcal{B})| \leq E(Y_+|\mathcal{B}) + E(Y_-|\mathcal{B})$$

Ainsi $E(|E(Y|\mathcal{B})|) \leq E(Y_+ + Y_-) = E(|Y|)$. Pour l'unicité comme on peut considérer $Y - E(Y|\mathcal{B})$, on peut adapter l'unicité dans le cas discret. Pour Z mesurable bornée on écrit $Z = Z_+ - Z_-$ et on applique le cas positif. □

4.2.4 Propriétés de l'espérance conditionnelle

Proposition 47. *(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 48. *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.*
$$\begin{aligned} E\big(E(X|\mathcal{B})|\mathcal{C}\big) &= E(X|\mathcal{C}), \\ E\big(E(X|\mathcal{C})|\mathcal{B}\big) &= E(X|\mathcal{C}). \end{aligned}$$

Démonstration. Soit Z une v.a $\mathcal{B}$ mesurable bornée, on a pour tout $n \in \mathbb{N}$
$$E(ZE(\inf(Y,n)X|\mathcal{B})) = E(Z\inf(Y,n)X) = E(Z\inf(Y,n)E(X|\mathcal{B}))$$
Ainsi presque surement
$$E(\inf(Y,n)X|\mathcal{B}) = \inf(Y,n)E(X|\mathcal{B})$$
on peut alors appliquer le point 4) de la proposition précédente. □

Proposition 49. *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, $E(Z|\mathcal{B}) = E(Z)$.*

4.3 Loi conditionnelle sachant une variable aléatoire

Définition 35. 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 50. *Si $h \geq 0$ est mesurable sur $(F, \mathcal{F})$ alors $x \mapsto \int h(y)K(x, dy)$ est mesurable.*

Définition 36. 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 48. On peut remplacer positif par fonction continue bornée dans la définition.

Remarque 49. 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.

Exemple 50. Si $X : \Omega \to E$ est une v.a discrète avec $P(X = x) > 0$ pour tout $x \in E$, alors $P(Y \in A|X = x) = K(x, A)$ a un sens pour tout $x \in E$.

Exemple 51. Soit (X,Y) un couple de v.a de loi jointe $p(x,y)dxdy$. Déterminons $E(h(Y)|X)$ pour identifier la loi condtionnelle de Y sachant $X=x$. Pour cela on sait qu'il existe g telle que $E(h(Y)|X)=g(X)$ et pour toute fonction f continue bornée

$$E(f(X)g(X)) = E(f(X)E(h(Y)|X)) = E(f(X)h(Y))$$

De manière générale, on va donc cherche à écrire $E(f(X)h(Y)) = \int f(x)g(x)dP_X(x)$. Ici on sait que la loi de X est à densité donnée par

$$f_X(x) = \int p(x,y)dy,$$

on a va donc chercher une formule du genre $E(f(X)h(Y)) = \int f(x)g(x)f_X(x)dx$ On a donc

$$\begin{aligned}
E(f(X)h(Y)) &= \int\int f(x)h(y)p(x,y)dxdy \\
&= \int f(x)\left(\int h(y)p(x,y)dy\right)dx \quad (Fubini) \\
&= \int f(x)\left(\int h(y)p(x,y)dy\right)dx \\
&= \int f(x)\left(\int h(y)\frac{p(x,y)}{f_X(x)}\mathbf{1}_{f_X(x)>0}dy\right)f_X(x)dx
\end{aligned}$$

Finalement on a donc

$$g(x) = \left(\int h(y)\frac{p(x,y)}{f_X(x)}\mathbf{1}_{f_X(x)>0}dy\right)$$

et par suite

$$E(h(Y)|X) = g(X) = \left(\int h(y)\frac{p(X,y)}{f_X(X)}\mathbf{1}_{f_X(x)>0}dy\right).$$

Ainsi la loi de Y sachant $X=x$ est donnée par

$$f_{Y|X=x}(y) = \frac{p(x,y)}{f_X(x)}\mathbf{1}_{f_X(x)>0}$$

Exemple 52. Si X et Y sont deux v.a indépendantes alors

$$K(x,dy) = P_Y(dy)$$

Théorème 51. *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 52. *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 :

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

Démonstration. Pour $A \in \mathcal{E}$ et $B \in \mathcal{F}$ et $h = \mathbf{1}_{A\times B}$. On a

$$\begin{aligned} E\big(h(X,Y)|X\big) &= E\big(\mathbf{1}_A(X)\mathbf{1}_B(Y)|X\big) \\ &= \mathbf{1}_A(X)E\big(\mathbf{1}_B(Y)|X\big) \\ &= \mathbf{1}_A(X)\int_F \mathbf{1}_B(y)K(X,dy) \\ &= \int_F \mathbf{1}_A(X)\mathbf{1}_B(y)K(X,dy) \end{aligned}$$

Ainsi l'ensemble des fonctionc vérifiant l'égalité $E\big(h(X,Y)|X\big) = \int_F h(X,y)\ K(X,dy) \quad p.s.$, est une classe monotone contenant les pavés (stable par multiplication) donc pour tout ensemble $M \in \mathcal{E}\otimes\mathcal{F}$

$$E\big(\mathbf{1}_M(X,Y)|X\big) = \int_F \mathbf{1}_M(X,y)\ K(X,dy) \quad p.s.$$

On obtient légalité pour toute fonction h en passant par les fonctions étagées puis positive et puis $h = h_+ - h_-$. ☐

Exemple 53. Soient X et Y deux variables aléatoire indépendantes $\mathcal{N}(0,1)$. Déterminer la loi conditionnelle de X sachant $X - Y$. On va raisonner de 2 manières différentes. On pose $Z = X - Y$, le but est de déterminer le noyau K tel que pour tout fonction h continue bornée

$$E(h(X)|Z) = \int h(x)K(Y,dx)$$

1ere méthode (déterminons la densité du couple). Pour cela on considère θ une fonction continue bornée et on calcul

$$\begin{aligned} E(h(X,Z)) &= E(h(X,X-Y)) && (4.1)\\ &= \int\int \theta(x,x-y))\frac{e^{-x^2/2}}{\sqrt{2\pi}}\frac{e^{-y^2/2}}{\sqrt{2\pi}}dxdy \quad (u=x, v=x-y) && (4.2)\\ &= \int\int \theta(u,v)\frac{e^{-u^2/2}}{\sqrt{2\pi}}\frac{e^{-(u-v)^2/2}}{\sqrt{2\pi}}dudv && (4.3) \end{aligned}$$

La densité de (X,Z) est donc

$$f_{(X,Z)}(x,z) = \frac{e^{-x^2/2}}{\sqrt{2\pi}}\frac{e^{-(x-z)^2/2}}{\sqrt{2\pi}}\mathbf{1}_{\mathbb{R}^2}(x,z)$$

Or la densité de $Z = X - Y$ est celle d'une $\mathcal{N}(0,2)$

$$f_Z(z) = \frac{e^{-z^2/4}}{2\sqrt{\pi}}$$

Ainsi

$$K(z,dx) = \frac{f_{(X,Z)}(x,z)}{f_Z(z)}dx = \frac{e^{-x^2/2}}{\sqrt{2\pi}}\frac{e^{-(x-z)^2/2}}{\sqrt{2\pi}}\frac{2\sqrt{\pi}}{e^{-z^2/4}}dx = \frac{e^{-(x-z/2)^2}}{\sqrt{\pi}}dx \qquad (4.4)$$

La loi de X sachant $X - Y = z$ est donc une loi $\mathcal{N}(z/2, 1/2)$.

2eme méthode. On sait d'après Doob qu'il existe une fonction p telle que

$$E(h(X)|Z) = g(Z)$$

et que pour toute fonction θ continue bornée

$$E(\theta(Z)h(X)) = E(\theta(Z)E(h(X)|Z)) = E(\theta(Z)p(Z)) = \int \theta(z)p(z)f_Z(z)dz$$

On sait également $f_Z(z) = \frac{e^{-z^2/4}}{2\sqrt{\pi}}$. On va donc transformée $E(\theta(Z)h(X))$ comme une intégrale en fonction de la densitée de Z. Ainsi on a

$$\begin{aligned}
E(\theta(Z)h(X)) &= E(\theta(X-Y)h(X)) && (4.5)\\
&= \int \theta(x-y)h(x)\frac{e^{-x^2/2}}{\sqrt{2\pi}}\frac{e^{-y^2/2}}{\sqrt{2\pi}}dxdy \quad (u=x, v=x-y) && (4.6)\\
&= \int \theta(v)h(u)\frac{e^{-u^2/2}}{\sqrt{2\pi}}\frac{e^{-(u-v)^2/2}}{\sqrt{2\pi}}dudv && (4.7)\\
&= \int \theta(v)\left(\int h(u)\frac{e^{-u^2/2}}{\sqrt{2\pi}}\frac{e^{-(u-v)^2/2}}{\sqrt{2\pi}}du\right)dv && (4.8)\\
&= \int \theta(v)\left(\int h(u)\frac{e^{-u^2/2}}{\sqrt{2\pi}}\frac{e^{-(u-v)^2/2}}{\sqrt{2\pi}}\frac{1}{f_Z(v)}du\right)f_Z(v)dv && (4.9)\\
&= \int \left(\int h(u)\frac{e^{-(u-z/2)^2}}{\sqrt{\pi}}du\right)f_Z(v) && (4.10)
\end{aligned}$$

On a alors

$$E(h(X)|Z) = \int h(u)\frac{e^{-(u-Z/2)^2}}{\sqrt{\pi}}du$$

et on retrouve le résultat précédent.

Exemple 54. Même cadre que l'exercice précédent. Donner la loi conditionnelle de X^2+Y^2 sachant $X-Y$. On rappelle que $X+Y$ et $X-Y$ sont indépendantes. On observe que l'on peut écrire

$$X^2+Y^2 = \frac{(X+Y)^2}{2} + \frac{(X-Y)^2}{2}$$

Ici on peut décrire la loi conditionnelle de X^2+Y^2 sachant $X-Y=z$ comme la loi du carré d'une $\mathcal{N}(0,2)$ divisé par 2 plus z^2. Trouver la densité est un peu pénible.

Chapitre 5

Convergence des suites de v.a. et de lois

5.1 Convergence presque sure

Définition 37. 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 53. *(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,*

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

Démonstration. L'événement $\{X = \lim X_n\}$ peut sécrire

$$\{X = \lim X_n\} = \bigcap_{k\geq 1} \bigcup_{N\in\mathbb{N}} \bigcap_{n\geq N} \{|X - X_n| \leq 1/k\}.$$

Par Borel Cantelli pour tout $\epsilon > 0$ si $\sum_{n\in\mathbb{N}} P\big(|X_n - X| \geq \varepsilon\big) < \infty$ alors on a que

$$P(\limsup\{|X_n - X| \geq \varepsilon\}) = P(\bigcap_{N\in\mathbb{N}} \bigcup_{n\geq N} \{|X_n - X| \geq \varepsilon\}) = 0$$

Donc pour tout $\epsilon > 0$ on a que $P(\bigcup_{N\in\mathbb{N}} \bigcap_{n\geq N} \{|X_n - X| < \varepsilon\}) = 1$ Donc pour tout k avec $\epsilon = 1/k$ on a que $P(\bigcup_{N\in\mathbb{N}} \bigcap_{n\geq N} \{|X_n - X| < 1/k\}) = 1$ était par suite

$$P(\bigcap_{k\geq 1} \bigcup_{N\in\mathbb{N}} \bigcap_{n\geq N} \{|X_n - X| < 1/k\}) = 1$$

et finalement

$$P(X = \lim X_n) = 1$$

Pour le deuxième point il faut montrer le sens de gauche à droite. Supposons par l'absurde qu'il existe ε tel que $\sum_{n\in\mathbb{N}} P\big(|X_n-\alpha|\geq\varepsilon\big)=\infty$. Comme les (X_n) sont indépendantes alors les ensembles $\{|X_n-\alpha|\geq\varepsilon\}$ sont indépendant ainsi par Borel cantelli

$$P(\limsup\{|X_n-X|\geq\varepsilon\})=P\left(\bigcap_{N\in\mathbb{N}}\bigcup_{n\geq N}\{|X_n-X|\geq\varepsilon\}\right)=1.$$

Il existe donc un ensemble $\tilde{\Omega}$ de mesure 1 tel que pour tout $\omega\in\tilde{\Omega}$, pour tout $N\in\mathbb{N}$ il existe $n\geq N$ tel que $|X_n-X|\geq\varepsilon$. C'est absurde car cela contredit la convergence de la suite (X_n).

□

Proposition 54. *Soient* $(X_n)_{n\geq0},(Y_n)_{n\geq0},X,Y$ *des v.a.r. définies sur* $(\Omega,\mathcal{A},P)$. *Si* $X_n\xrightarrow{p.s.}X$ *et* $Y_n\overset{p.s.}{\to}Y$ *alors* $X_n+Y_n\xrightarrow{p.s.}X+Y$ *et* $X_nY_n\xrightarrow{p.s.}XY$.

5.2 Convergence en probabilité

Définition 38. Soient $X,(X_n)_{n\geq0}$ 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 $(X_n)_{n\geq0}$ 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 55. *(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* $X_{n_k}\xrightarrow{p.s.}X$.

Démonstration. Pour le point 1, on a que $P\big(\big(|X_n-X|\geq\varepsilon\big)\big)=E(\mathbf{1}_{|X_n-X|\geq\varepsilon})$ or $\mathbf{1}_{|X_n-X|\geq\varepsilon}\xrightarrow{p.s.}0$ puis on applique le théorème de convergence dominée.

Pour le point 2 on sait que pour tout $\varepsilon>0$, $\lim_{n\to\infty}P\big(|X_n-X|\geq\varepsilon\big)=0$. Donc pour tout $k\in\mathbb{N}$ il existe $n_k\in\mathbb{N}$ telle que pour tout $n\geq n_k$ $P\big(|X_n-X|\geq1/k\big)\leq1/2^k$. on peut toujours choisir (n_k) une suite strictement croissante. Ainsi on a que

$$\sum_k P\big(|X_{n_k}-X|\geq1/k\big)\leq\sum_k 1/2^k<\infty.$$

On applique alors Borel Cantelli pour conclure. En effet on a que

$$P\left(\bigcap_{k\in\mathbb{N}}\bigcup_{j\geq k}|X_{n_j}-X|\geq1/j\right)=0.$$

Avec probabilité 1 il existe $k\in\mathbb{N}$ tel que pour tout $j\geq k$ on a

$$|X_{n_j}-X|\leq1/j$$

et donc (X_{n_j}) converge vers X.

□

Remarque 55. Il y existe des suites de v.a qui converge en proba mais pas p.s. Soit $[0,1]$ munit de la loi uniforme et $X_1=\mathbf{1}_{[0,1/2]},X_2=\mathbf{1}_{[1/2,1]},X_3=\mathbf{1}_{[0,1/4]},X_4=\mathbf{1}_{[1/4,1/2]}$ etc. On construit ainsi une vague avec les nombres dyadique. Alors on a que X_n converge en proba vers 0 mais pas presque surement. Pour $\omega\in]0,1]$, on peut exhiber une sous suite telle que $X_n(\omega)=1$. Essayer de formaliser proprement ce contre-exemple est un bon exercice.

Remarque 56. Soit (X_n) une suite de v.a indépendante telle que $X_n\mathcal{B}(1/n)$ alors (X_n) converge en proba vers 0 mais pas presque surement. Ceci résulte de l'application du Lemme de Borel Cantelli.

Définition-Proposition 39. 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.*

Démonstration. Il est facile de vérifier que d définie une distance. Montrons l'équivalence. De droite à gauche, on a que pour tout $1>\varepsilon>0$

$$E(\min(1,|X-Y|)) \geq \varepsilon P(\min(1,|X-Y|)\geq\varepsilon) = \varepsilon P(|X-Y|\geq\varepsilon)$$

Ainsi si $d(X,X_n)\longrightarrow 0$ alors $P(|X-X_n|\geq\varepsilon)\longrightarrow 0$ pour tout ε.

Pour la réciproque on observe que

$$E(\min(1,|X-X_n|)) \leq \varepsilon + P(\min(1,|X-X_n|)\geq\varepsilon)$$

et donc on a que

$$\limsup E(\min(1,|X-X_n|)) \leq \varepsilon$$

pour tout ε. Ainsi

$$\limsup E(\min(1,|X-X_n|)) = 0$$

et le résultat en découle. $\square$

Proposition 56. *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* $X_{n_{k_j}} \xrightarrow[j\to\infty]{p.s.} X$.

Démonstration. De la première assertion à la deuxième on que si $X_n \xrightarrow{P} X$ alors $d(X_n,X)\to 0$ ainsi pour toute sous suite $d(X_{n_k},X)\to 0$ et on peut alors extraire de cette sous suite une sous suite convergente presque sûrement par la proposition ci-dessus.

Dans l'autre sens, supposons par l'absurde que $d(X,X_n)$ ne converge pas vers 0, il existe donc $\varepsilon>0$ et une sous suite telle que $d(X_{n_k}X)\geq\varepsilon$. De cette sous suite on peut extraire une sous suite qui converge p.s donc en proba et donc $d(X_{n_{k_j}},X)$ converge vers 0 mais on a aussi $d(X_{n_{k_j}}X)\geq\varepsilon$ ce qui est impossible. $\square$

Corollaire 57. *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* $\alpha X_n+\beta Y_n \xrightarrow{P} \alpha X+\beta Y$,

— $X_nY_n \xrightarrow{P} XY$.

Démonstration. Soit (X_{n_k}) une sous suite extraite. De cette sous suite, on peut extraire une sous suite $(X_{n_{k_j}})$ qui converge p.s. Alors comme ϕ est continue on a que $(\phi(X_{n_{k_j}}))$ converge p.s. Donc de toute suite extraite $(\phi(X_{n_k}))$ on peut extraire une sous suite qui converge p.s. Donc $(\phi(X_n))$ converge en proba. $\square$

Théorème 58. *L'espace* $\left(L^0_{\mathbb{R}^d}(\Omega, \mathcal{A}, P), d\right)$ *est complet.*

Démonstration. Soit (X_n) une suite de Cauchy pour d montrons qu'elle converge. Pour cela il suffit de montrer qu'il n'y a qu'une seule valeur d'adhérence. Pour cela on peut extraire une suite telle que

$$d(X_{n_k}, X_{n_{k+1}}) \leq 1/2^k$$

Donc

$$\sum E(\min(1, |X_{n_k} - X_{n_{k+1}}|)) < \infty.$$

Donc

$$\sum \min(1, |X_{n_k} - X_{n_{k+1}}|) < \infty \quad p.s.$$

Donc $\sum X_{n_k} - X_{n_{k+1}}$ converge p.s. Enfin $X_{n_{k+1}} = X_{n_0} + \sum_{j=0}^{k} X_{n_{j+1}} - X_{n_j}$ et donc (X_{n_k}) converge p.s et donc en proba. On a donc montré que (X_n) a une valeur d'adhérence et comme toute suite de Cauchy n'en a qu'une on a le résultat. $\square$

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

Définition 40. 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

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

Proposition 59. *Si* $q \geq p \geq 1$,

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

Démonstration. Pour les convergences L^p il suffit de remarquer que pour $1 \leq p \leq q$

$$\|X_n - X\|_1 \leq \|X_n - X\|_p \leq \|X_n - X\|_q.$$

Ensuite pour montrer que la convergence L^1 entraine la convergence en probabilité, il suffit de remarquer que toute suite extraite converge en norme L^1 et de cette soute suite on peut extraire une sous suite qui converge p.s. Alors on obtient la convergence en probabilité. $\square$

Exemple 57. Soit X_n une suite de v.a telle que

$$P(X_n = n^{1/r}) = \frac{1}{n} \quad P(X_n = 0) = 1 - \frac{1}{n}$$

alors pour tout $\varepsilon >$ on a que

$$P(|X_n| > \varepsilon) = P(X_n = n^{1/r}) = \frac{1}{n} \to 0$$

donc (X_n) converge en proba vers 0. Mais

$$E(X_n^r) = 1$$

donc ne converge pas en norme L^r.

Définition 41. 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 58. Si $(X_i)_{i\in I}$ est UI, chacune des variables X_i est intégrable. En effet

$$E(|X_i|) \leq c + E\left(|X_i|\, 1_{|X_i|>c}\right)$$

et donc on conclut pour c suffisamment grand. De plus

$$\sup_i E(|X_i|) < \infty$$

Remarque 59. Une v.a L^1 est toujours UI

Proposition 60. *(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.*

Démonstration. Seul le dernier point mérite une preuve. On a que $\forall M > 0$, $\exists k \geq 0$ tel que $\forall x \geq K$

$$\frac{\phi(x)}{x} \geq M$$

Donc $\forall M > 0$, $\exists k \geq 0$ et pour tout $i \in I$

$$E\left(|X_i|\, 1_{|X_i|>k}\right) \leq \frac{E(\phi(|X_i|))}{M}$$

on peut donc prendre la limite quand k tend vers l'infini et on a le résultat. □

Remarque 60. On sent ici que la seule intégrabilité i.e $\sup_i E(|X_i|) < \infty$ ne suffit pas (i.e $\phi(x) = x$)

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

Démonstration. Montrons l'intégrabilité. On sait qu'il existe une suite (X_{n_k}) qui converge p.s vers X. Ainsi

$$E(|X|) = E(\lim |X_{n_k}|) \leq \liminf E(|X_{n_k}|) \leq \sup E(|X_n|).$$

En remarquant que

$$|X_n - X|\mathbf{1}_{|X_n-X|\geq k} \leq 2|X_n|\mathbf{1}_{|X_n|\geq k/2} + 2|X|\mathbf{1}_{|X|\geq k/2}$$

on voit que $(X_n - X)$ est UI. En posant $Y_n = |X_n - X|$ on a que pour tout $\varepsilon > 0$

$$EY_n \leq \varepsilon + KP(\varepsilon \leq Y_n \leq K) + E(Y_n\mathbf{1}_{Y_n\geq K}).$$

A ce stade on fait tendre n vers l'infini on a donc

$$\limsup E(|Y_n|) \leq \varepsilon + \sup_n E(Y_n\mathbf{1}_{Y_n\geq K})$$

Pour K tendant vers l'infini on a donc

$$\limsup E(|Y_n|) \leq \varepsilon$$

pour tout ε. Donc (Y_n) converge vers 0 dans L^1 et enfin (X_n) converge vers X dans L^1. □

Proposition 62. *Si* $X_n \xrightarrow{P} X$ *et s'il existe* $q > 1$ *tel que* $\sup_n \|X_n\|_q < +\infty$ *alors*

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

Démonstration. La suite (X_n) est UI avec $\phi(x) = x^q$. Donc (X_n) converge vers X dans L^1. Avec un argument similaire à la démonstration du théorème précédent on a aussi que $X \in L^q$. Pour $p \in [1, q[$ on peut écrire $p = \lambda + (1-\lambda)q$ alors

$$E|X_n - X|p = E|X_n - X|^{\lambda + (1-\lambda)q} = E|X_n - X|^{\lambda}|X_n - X|^{(1-\lambda)q} \leq (E|X_n - X|)^{\lambda}(E|X_n - X|^q)^{(1-\lambda)}$$

On a donc le résultat attendu. □

5.4 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 $\lim_{|x| \to +\infty} f(x) = 0$.

Proposition 63. *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 42. 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

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

On rappelle que le dual de $C_0(\mathbb{R}^d)$

$$C_0(\mathbb{R}^d)^* = \{\text{mesures boréliennes signées finies}\}$$

Proposition 64. *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*

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

Démonstration. Seul $\Leftarrow$ nécessite une preuve car l'autre sens est immédiat puisque $C_0(\mathbb{R}^d) \subset C_b(\mathbb{R}^d)$. On introduit la suite de fonction (f_k) continue telle que

$$f_k(x) = \begin{cases} 1 & si \quad |x| \leq k \\ 0 & si \quad |x| \geq k+1 \end{cases}$$

Pour toutes fonctions $\phi \in C_b(\mathbb{R}^d)$ on note $\phi_k = \phi f_k \in C_0(\mathbb{R}^d)$. Si

$$\lim_{n \to \infty} \int \phi \, d\mu_n = \int \phi \, d\mu$$

pour tout fonction $\phi \in H$ alors μ_n converge vers μ dans $C_0(\mathbb{R}^d)^*$ pour la topologie faible $*$. Par densité si c'est vrai pour H cela reste vrai pour $C_0(\mathbb{R}^d)$ et par suite

$$\lim_{n \to \infty} \int \phi \, d\mu_n = \int \phi \, d\mu$$

pour toute fonction $\phi \in C_0(\mathbb{R}^d)$. Maintenant pour $\phi \in C_b(\mathbb{R}^d)$ on a que

$$\phi = \phi(1 - f_k) + \phi f_k$$

Ainsi

$$\begin{aligned}\left|\int \phi d\mu_n - \int \phi d\mu\right| &\leq \int |\phi(1-f_k)|d\mu_n + \int |\phi(1-f_k)|d\mu + \left|\int \phi_k d\mu_n - \int \phi_k d\mu\right| \\ &\leq \|\phi\|_\infty \int |1-f_k|d\mu_n + \|\phi\|_\infty \int |1-f_k|d\mu + \left|\int \phi_k d\mu_n - \int \phi_k d\mu\right|\end{aligned}$$

Ainsi

$$\limsup_n \left|\int \phi d\mu_n - \int \phi d\mu\right| \leq 2\|\phi\|_\infty \int |1-f_k|d\mu$$

On a alors le résultat en prenant k qui tend vers l'infini car les mesures sont finies. □

Définition 43. Soient X et $(X_n)_{n\geq 0}$ v.a. à valeurs dans $\left(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d)\right)$, 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 $X_n \xrightarrow{\mathcal{L}} P_X$.

Remarque 61. Dans le cas où (X_n) est une suite de v.a discrète à valeurs dans le même espace E, la convergence en loi vers X est équivalente à

$$P(X_n = x) \to P(X = x),$$

pour tout $x \in E$.

Remarque 62. Notons que des v.a discrète peuvent converger en loi vers des v.a continues. Par exemple si pour $n \in \mathbb{N}^*$ $X_n \sim \mathcal{U}(\{\frac{1}{n}, \dots, \frac{k}{n}, \dots 1\})$ alors (X_n) converge vers $U \sim \mathcal{U}([0,1])$. En effet soit ϕ une fonction continue bornée alors par les intégrales de Riemann

$$E(\phi(X_n)) = \sum_{k=1}^n \frac{1}{n}\phi\left(\frac{k}{n})\right) \to \int_0^1 \phi(x)dx.$$

L'inverse est également vrai. Si (x_n) est une suite strictement positive qui converge vers 0 alors $X_n \sim \mathcal{N}(0, x_n^2)$ converge en loi vers $X \sim \delta_0$.

Proposition 65. $X_n \xrightarrow{P} X \quad \Rightarrow \quad X_n \xrightarrow{\text{loi}} X.$

Démonstration. Soit ϕ une fonction continue bornée alors la suite $(\mathbb{E}(\phi(X_n))$ est une suite bornée. Montrons que sa seule valeur d'adhérence est $E(\phi(X))$. Comme X_n converge en probabilité. De toute suite extraite (X_{n_k}) on peut extraire une sous suite $(X_{n_{k_j}})$ qui converge p.s vers X. Alors par convergence dominée, on a

$$\lim_j E[\phi(X_{n_{k_j}})] = E[\phi(X)].$$

Ainsi si (X_{n_k}) est une suite extraite telle que $(\mathbb{E}(\phi(X_{n_k}))$ converge alors en extrayant uen sous suite qui converge p ;s la limite est forcément $E[\phi(X)]$.

Sinon on aurait pu considérer les fonctions continues à support compact qui sont dense dans $C_0(\mathbb{R}^d)$. On a alors pour tout $\epsilon > 0$

$$\begin{aligned}|E\phi(X_n) - E\phi(X)| &\leq E|\phi(X_n) - \phi(X)|\mathbf{1}_{|X_n - X| > \epsilon} + E|\phi(X_n) - \phi(X)|\mathbf{1}_{|X_n - X| \leq \epsilon} \\ &\leq 2\|\phi\|_\infty P(|X_n - X| > \epsilon) + E|\phi(X_n) - \phi(X)|\mathbf{1}_{|X_n - X| \leq \epsilon}\end{aligned}$$

On utilise alors la convergence en proba pour le premier terme et comme ϕ est continue à support compact, elle est uniformément continue ce qui permet de contrôler le second terme. □

Proposition 66. *Soient μ_n, μ des mesures de probabilité sur $\left(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d)\right)$. Les assertions suivantes sont équivalentes :*

1. *$\mu_n \stackrel{\text{étroit}}{\longrightarrow} \mu$.*
2. *Pour tout ouvert $G \subset \mathbb{R}^d$, $\liminf_{n\to\infty} \mu_n(G) \geq \mu(G)$.*
3. *Pour tout fermé $F \subset \mathbb{R}^d$, $\limsup_{n\to\infty} \mu_n(F) \leq \mu(F)$.*
4. *Pour tout borélien $B \subset \mathbb{R}^d$ tel que $\mu(\partial B) = 0$ on a $\lim_{n\to+\infty} \mu_n(B) = \mu(B)$.*

Démonstration. 2) ⇔ 3) vient du fait que le complémentaire d'un fermé est un ouvert et vice versa.

Montrons que 1) ⇒ 2). Soit O un ouvert. On considère la fonction

$$\phi_k(x) = \min(1, kd(x, O^c))$$

qui est une fonction continue bornée. Ainsi $\phi_k \leq \mathbf{1}_O$ et $\phi_k \to_k \mathbf{1}_O$. On a donc pour tout n

$$\mu_n(O) \geq \int \phi_k d\mu_n$$

et donc

$$\liminf \mu_n(O) \geq \liminf \int \phi_k d\mu_n = \int \phi_k d\mu$$

Ensuite par convergence dominée en faisant tendre k vers l'infini on a

$$\liminf \mu_n(O) \geq \liminf \int \phi_k d\mu_n = \int \mathbf{1}_O d\mu = \mu(O)$$

Montrons que 2) et 3) implique 4). Si $\mu(\partial B) = 0$ alors

$$\mu(B) = \mu(\mathring{B}) = \mu(\bar{B})$$

donc

$$\liminf \mu_n(B) \geq \liminf \mu_n(\mathring{B}) \geq \mu(\mathring{B}) = \mu(B) \tag{5.1}$$
$$\limsup \mu_n(B) \leq \limsup \mu_n(\bar{B}) \leq \mu(\bar{B}) = \mu(B) \tag{5.2}$$

Donc

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

Montrons enfin que 4) ⇒ 1). Soit $\phi \in C_b(\mathbb{R}^d)$. On peut supposer quitte à considérer la partie positive puis ensuite négative que $\phi \geq 0$. On a que

$$\begin{aligned}\int_{\mathbb{R}^d} \phi d\mu_n &= \int_{\mathbb{R}^d} \left(\int_0^{\|\phi\|_\infty} \mathbf{1}_{y \leq \phi(x)} dy \right) d\mu_n(x) \\ &= \int_0^{\|\phi\|_\infty} \mu_n(\{x : \phi(x) \geq y\}) dy \quad (Fubini)\end{aligned}$$

A ce stade considérons le fermé $F_y = \{x : \phi(x) \geq y\}$ qui contient l'ouvert $\{x : \phi(x) > y\}$. On a donc

$$\partial F_y \subset F_y \setminus \{x : \phi(x) > y\} = \phi^{-1}(\{y\}).$$

On a donc que $\mu(\partial F_y) = 0$ pour presque tout y. En effet

$$\{y : \mu(\partial F_y) > 0\} \subset \{y : \mu(\phi^{-1}(\{y\})) > 0\}$$

qui sont les atomes de la mesure image de μ par ϕ. Donc $\{y : \mu(\partial F_y) > 0\}$ est au plus dénombrable. Ainsi pour presque tout y

$$\mu_n(F_y) \to \mu(F_y)$$

et donc par convergence dominée

$$\int_0^{\|\phi\|_\infty} \mu_n(\{x : \phi(x) \geq y\})dy \to \int_0^{\|\phi\|_\infty} \mu(\{x : \phi(x) \geq y\})dy = \int_{\mathbb{R}^d} \phi d\mu$$

et le résultat en découle. □

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

Proposition 67. *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*

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

Démonstration. L'implication $\Rightarrow$ est une conséquence du théorème du portemanteau. Pour la réciproque on va montrer le point 2) du théorème du portemanteau. Ainsi considérons un ouvert $O =]a, b[$ (le cas général s'en déduit par union dénombrable). Comme le nombre de points de discontinuité de F_X est dénombrable on peut trouver une suite (a_k) décroissante et convergeante vers a telle que F_X est continue en a_k pour tout k et une suite (b_k) croissante qui converge vers b telle que F_X est continue en b_k. Par continuité à droite et à gauche, on a que

$$\begin{aligned}
\limsup F_{X_n}(a) &\leq \limsup F_{X_n}(a_k) = F_X(a_k) \to F_X(a) \\
\liminf F_{X_n}(b^-) &\geq \liminf F_{X_n}(b_k) = F_X(b_k) \to F_X(b-)
\end{aligned}$$

Ainsi

$$\begin{aligned}
\liminf P_{X_n}(]a, b[) &= \liminf F_{X_n}(b^-) - F_{X_n}(a) \\
&\geq F_X(b^-) - F_X(a) = P_X(]a, b[)
\end{aligned}$$

et donc on a le résultat attendu. □

La convergence en loi concerne donc la convergence d'une suite de mesure. Une suite de mesure de probabilité (μ_n) est bornée pour la topologie faible $*$. Il est naturel de savoir si on peut en extraire une sous suite convergente. Dans le cas où le support des μ_n est discret $\{x_1, \ldots, x_K\}$, alors

$$\{\mu_n(x_i), i = 1, \ldots, K, n \in \mathbb{N}\} \subset \left\{(y_1, \ldots, y_K) \in \mathbb{R}^K, y_i \geq 0, \sum_{i=1}^{K} y_i = 1\right\}.$$

On a donc une suite $(\mu_n(x_i)_{i=1,\dots,K})$ est une suite dans un compact donc on peut extraire une sous suite qui converge et la limite définie une loi de probabilité sur $\{x_1, \dots, x_K\}$. La sous suite converge donc en loi. Dans le cas général c'est plus délicat.

D'après le théorème de Banach Alaoglu, la boule unité du dual $C_0(\mathbb{R}^d)^*$ est $*$ faiblement compact donc une suite de mesure (μ_n) admet une sous suite qui converge vers une mesure positive mais qui n'est pas forcément une loi de probabilité. Par exemple (δ_n).

Pour garantir le fait que la mesure limite est une loi de probabilité, on utilise la notion de tension.

Définition 44. 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 68 (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.*

Démonstration. D'après Banach Alaoglu, on peut extraire uen sous suite (μ_{n_k}) qui converge $*$ faiblement vers une mesure positive μ. Pour toute fonction $f \in C_0(\mathbb{R}^d)$ on a que

$$\int f d\mu_{n_k} \to \int f d\mu.$$

On a que $\mu(\mathbb{R}^d) \leq 1$. Pour conclure il suffit de montrer que $\mu(\mathbb{R}^d) = 1$. Soit $\epsilon > 0$, on peut donc trouver un compact K tel que

$$\sup_n \mu_{n_k}(K) \geq 1 - \epsilon.$$

On peut trouver $f \in C_0(\mathbb{R}^d)$ telle que $\mathbf{1} \geq f \geq \mathbf{1}_K$. On a alors

$$\mu(\mathbb{R}^d) \geq \int f d\mu = \lim \int f d\mu_n \geq \lim \int \mathbf{1}_K d\mu_{n_k} \geq 1 - \epsilon$$

Maintenant en faisant tendre ϵ vers 0, on a que $\mu(\mathbb{R}^d) \geq 1$ et le résultat est prouvé. $\square$

Théorème 69. (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 $X_n \xrightarrow{\mathcal{L}} X$.

Démonstration. Le point 1) est évident en considérant la partie réelle et imaginaire de $x \mapsto e^{ix}$.

Pour le point 2) il suffit de montrer que μ_n est tendue. En effet le théorème de Prokhorov nous donnera l'existence d'une sous suite convergente vers une mesure μ. Ainsi Φ est n :'ecessairement la fonction caractéristique de μ. De cette manière toutes les valeurs d'adhérence seront égale à μ (car elles auront même fonction caractéristique).

Pour montrer la tension, on va montrer que $P_{X_n}([-M, M])$ converge uniformément vers 1 lorsque M tend vers l'infini. Sur l'ensemble $[-M, M]^c$ la quantité

$$1 - \frac{\sin(X/M)}{X/M}$$

est minorée par une certaine constante α indépendante de M. On a donc

$$\begin{array}{rcl} P_{X_n}([-M,M]^c) & = & E\left(\mathbf{1}_{[-M,M]^c}(X_n)\right) \\ & \leq & \frac{1}{\alpha}E\left(1-\frac{\sin(X_n/M)}{X_n/M}\right) \\ & \leq & \frac{1}{\alpha}E\left(\frac{M}{2}\int_{-1/M}^{1/M}(1-e^{iX_nt})dt\right) \\ & \leq & \frac{1}{\alpha}\frac{M}{2}\int_{-1/M}^{1/M}(1-\Phi_{X_n}(t))dt \end{array}$$

Lorsque n tend vers l'infini, par convergence dominée, on a

$$\frac{M}{2}\int_{-1/M}^{1/M}(1-\Phi_{X_n}(t)) \to \frac{M}{2}\int_{-1/M}^{1/M}(1-\Phi(t))$$

Donc

$$\limsup_n P_{X_n}([-M,M]^c) \leq \frac{1}{\alpha}\frac{M}{2}\int_{-1/M}^{1/M}(1-\Phi(t))$$

Ainsi il existe N telle que

$$\sup_{n\geq N} P_{X_n}([-M,M]^c) \leq \frac{1}{\alpha}\frac{M}{2}\int_{-1/M}^{1/M}(1-\Phi(t))$$

Or

$$\lim_M \frac{M}{2}\int_{-1/M}^{1/M}(1-\Phi(t)) = 1-\Phi(0) = 0$$

donc $P_{X_n}([-M,M]^c)$ converge uniformmñent vers 0 et le résultat est obtenu. □

Théorème 70. (Lévy)(Version faible). $X_n \xrightarrow{\mathcal{L}} X$ *ssi* $\forall\xi \in \mathbb{R}^d,\ \lim_{n\to\infty}\Phi_{X_n}(\xi) = \Phi_X(\xi)$.

Démonstration. Naturellement c'est une conséquence du théorème précédent mais on peut directement le résultat sans faire apppel à la tension. Le sens $\Rightarrow$ est immédiat car $x \mapsto e^{itx}$ est continue et bornée (au pire on prend la partie imaginaire et réelle). Pour le sens $\Leftarrow$, on considère une fonction $\phi \in L^1$ on considère

$$\xi \mapsto f(\xi) = \int e^{it\xi}\phi(t)dt$$

qui est une fonction C_0. Alors

$$\begin{array}{rcl} E(f(X_n)) & = & E(\int e^{itX_n}\phi(t)dt) \\ & = & \int E(e^{itX_n}\phi(t)) \quad (Fubini) \\ & = & \int \phi(t)\Phi_{X_n}(t)dt \\ & \to_n & \int \phi(t)\Phi_X(t) = E(f(X)) \quad \text{(convergence dominée)} \end{array}$$

En notant C_c^∞ les fonctions C^∞ à support compact on sait que C_c^∞ est dense dans C_0. Or $C_c^\infty \subset \mathcal{S}$ où $\mathcal{S} = \{f \in \mathcal{C}^\infty(\mathbb{R}^d) \mid \forall(\alpha,\beta)\ \|f\|_{\alpha,\beta} < +\infty\}$ désigne l'espace de Schwarz avec

$\|f\|_{\alpha,\beta} = \|x^\alpha D^\beta f\|_\infty$. Or la transformée de Fourier est un automorphisme sur l'espace de Schwarz donc pour tout $f \in C_c^\infty$ il existe $\phi \in \mathcal{S}$ telle que $\xi \mapsto f(\xi) = \int e^{it\xi}\phi(t)dt$. Ainsi pour tout $f \in C_c^\infty$

$$\lim_n Ef(X_n) = Ef(X).$$

En utilisant une proposition ci dessus on obtient le résultat car C_c^∞ est dense dans C_0 pour la norme $\|.\|_\infty$. □

Théorème 71 (Skorokhod)**.** *On considère une suite de v.a.* $X_n : (\Omega_n, \mathcal{A}_n, P_n) \to \left(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d)\right)$, $n \geq 0$. *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* $Y_n : (\Omega, \mathcal{A}, P) \to \left(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d)\right)$ *telle que*

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

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

Lemme 72 (Lemme de Slutsky)**.** $X_n \xrightarrow{\mathcal{L}} X$ *et* $Y_n \xrightarrow{P} c$ *alors*

$$(X_n, Y_n) \xrightarrow{\mathcal{L}} (X, c)$$

Démonstration. Soient f et $g$deux fonctions continues bornées, on va montrer que

$$E(f(X_n)g(Y_n) \to_n E(f(X)g(c))$$

Soit $\epsilon > 0$, il existe N tel que pour tout $n \geq N$

$$E|f(X_n) - f(X)| \leq \epsilon \quad P(|g(X_n) - g(c)| \geq \epsilon) \leq \epsilon$$

Ainsi

$$\begin{aligned} E(f(X_n)g(X_n)) - E(f(X)g(c)) &= E(f(X_n)(g(X_n)) - g(c)) + E(f(X_n) - f(X))g(c) \\ &\leq \|f\|_\infty \|2g\|_\infty P(|g(Y_n) - g(c)| \geq \epsilon) + \|f\|_\infty \epsilon P(|g(Y_n) - g(c)| \leq \epsilon) \\ &\quad + \|g\|_\infty \epsilon \\ &\leq (2\|f\|_\infty \|g\|_\infty + \|f\|_\infty + \|g\|_\infty)\epsilon \end{aligned}$$

et le résultat est obtenu. □

Chapitre 6

Théorèmes limites

6.1 Loi des grands nombres

Définition 45. 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, $\big(X_1(\omega), \ldots, X_n(\omega)\big)$, 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 73 (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*

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

Démonstration. Pour tout $n \in \mathbb{N}^*$ on a $E(\bar{X}_n) = E(X_1)$. Ainsi

$$\begin{aligned} E\left(\left|\bar{X}_n - E(X_1)\right|^2\right) &= Var(\bar{X}_n) \\ &= \frac{1}{n}Var(X_1) \\ &\to_n 0 \end{aligned}$$

□

On a une autre version de ce théorème en termes de convergence en probabilité

Théorème 74 (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*

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

Démonstration. Naturellement il s'agit d'une conséquence de la convergence L^2 mais la preuve directe peut être utilisé pour établir des intervalles de confiance non asymptotiques. Soit $\epsilon > 0$ on a

$$P\left(|\bar{X}_n - EX_1| \geq \epsilon\right) \leq \frac{Var(\bar{X}_n)}{\epsilon^2} = \frac{VarX_1}{n\epsilon^2} \to 0$$

□

Exemple 63. Lorsque $X_i \sim \mathcal{B}(p)$ on a

$$P\left(|\bar{X}_n - p| \geq \epsilon\right) \leq \frac{Var(\bar{X}_n)}{\epsilon^2} = \frac{Var X_1}{n\epsilon^2} = \frac{p(1-p)}{n\epsilon^2} \leq \frac{1}{4n\epsilon^2}$$

Si on pose $\alpha = \frac{1}{4n\epsilon^2}$ ce qui implique $\epsilon = \frac{1}{\sqrt{4n\alpha}}$. Ainsi

$$P(\bar{X}_n - \frac{1}{\sqrt{4n\alpha}} < p < \bar{X}_n - \frac{1}{\sqrt{4n\alpha}}) = P(|\bar{X}_n - p| < \frac{1}{\sqrt{4n\alpha}}) \geq 1 - \alpha.$$

L'intervalle aléatoire

$$]\bar{X}_n - \frac{1}{\sqrt{4n\alpha}}, \bar{X}_n - \frac{1}{\sqrt{4n\alpha}}[$$

est un intervalle dit de confiance au niveau de confiance α.

Théorème 75 (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*

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

Démonstration. **Etape 1) :** On peut supposer que $X_i \geq 0$ en décomposant en partie positive et négative.

Etape 2) : On peut remplacer X_i par $Y_i = X_i \mathbf{1}_{X_i \leq i}$. En effet les suites (X_n) et (Y_n) sont égales à partir d'un rang (aléatoire). En effet on a que

$$\begin{aligned}
\sum_{i=1}^{\infty} P[X_i \neq Y_i] = \; &= \sum_{i=1}^{\infty} P(X_i > i) \\
&= \sum_{i=1}^{\infty} P(X_1 > i) \\
&\leq \sum_{i=1}^{\infty} \int_{i-1}^{i} P(X_1 > x)dx \\
&= \int_0^{\infty} P(X_1 > x)dx \\
&= E\left(\int_0^{\infty} \mathbf{1}_{X_1 > x} dx\right) = E(X_1) < +\infty
\end{aligned}$$

Ainsi par Borel Cantelli

$$P\left(\bigcup_{n\in\mathbb{N}} \bigcap_{k\geq n} \{X_k = Y_k\}\right) = 1$$

Ainsi pour ω dans un ensemble de probabilité il existe $n \in \mathbb{N}$ telle que pour tout $k \geq n$ on a $X_k(\omega) = Y_k(\omega)$. On va donc démontrer que p.s

$$\bar{Y}_n = \frac{1}{n}\sum_{i=1}^{n} Y_i \to E[X_1].$$

Etape 3) : On fixe $\alpha > 1$ et on définit $k_n = \lfloor \alpha^n \rfloor$. On va montrer que

$$\bar{Y}_{k_n} \to E(X_1)$$

on comblera les trous ensuite. Par convergence monotone on a que

$$E(Y_i) = E(X_1 \mathbf{1}_{X_1 \leq i}) \to E(X_1).$$

On a donc que

$$E(\bar{Y}_{k_n}) = \frac{1}{k_n} \sum_{i=1}^{k_n} E(X_1 \mathbf{1}_{X_1 \leq i}) \to E(X_1)$$

par Césaro. On est donc ramener à montrer que p.s

$$\bar{Y}_{k_n} - E(\bar{Y}_{k_n}) \to 0$$

Utilisons pour cela le critère de Borel Cantelli. Soit $\epsilon > 0$ on a

$$\begin{aligned} \sum_{n=1}^{\infty} P(|\bar{Y}_{k_n} - E(\bar{Y}_{k_n}| \geq \epsilon) &\leq \sum_{n=1}^{\infty} \frac{Var(\bar{Y}_{k_n})}{\epsilon^2} \\ &\leq \sum_{n=1}^{\infty} \frac{1}{\epsilon^2} \sum_{i=1}^{k_n} \frac{Var(Y_i)}{k_n^2}. \end{aligned}$$

Par ailleurs on a que

$$Var(Y_i) \leq E(|X_i|^2 \mathbf{1}_{X_i \leq i}) \leq E(|X_1|^2 \mathbf{1}_{X_1 \leq k_n})$$

ainsi

$$\begin{aligned} \sum_{n=1}^{\infty} P(|\bar{Y}_{k_n} - E(\bar{Y}_{k_n}| \geq \epsilon) &\leq \sum_{n=1}^{\infty} \frac{1}{\epsilon^2} \sum_{i=1}^{k_n} \frac{E(|X_1|^2 \mathbf{1}_{X_1 \leq k_n})}{k_n^2} \\ &\leq \frac{1}{\epsilon^2} E\left(\sum_{n=1}^{\infty} \frac{\mathbf{1}_{X_1 \leq k_n}}{k_n} |X_1|^2 \right) \end{aligned}$$

Comme $k_n \geq \alpha^n - 1 \geq (1 - 1/\alpha)\alpha^n$ on a que

$$\sum_{n=1}^{\infty} \frac{\mathbf{1}_{X_1 \leq k_n}}{k_n} \leq \sum_{n: X_1 \leq k_n} \frac{1}{(1 - 1/\alpha)\alpha^n}.$$

De plus si $X_1 \leq k_n$ alors $n \geq \frac{\log(X_1}{\log \alpha}$. Donc

$$\begin{aligned} \sum_{n=1}^{\infty} \frac{\mathbf{1}_{X_1 \leq k_n}}{k_n} &\leq \frac{1}{1 - \frac{1}{\alpha}} \sum_{n: n \geq \frac{\log(X_1}{\log \alpha}} \frac{1}{\alpha^n} \\ &\leq \frac{1}{1 - \frac{1}{\alpha}} \frac{\alpha^{-\log(X_1)/\log(\alpha)}}{1 - \frac{1}{\alpha}} \\ &\leq \frac{X_1^{-1}}{(1 - 1/\alpha)^2} \end{aligned}$$

Ainsi

$$E\left(\sum_{n=1}^{\infty} \frac{\mathbf{1}_{X_1 \leq k_n}}{k_n} |X_1|^2 \right) \leq \frac{EX_1}{(1 - 1/\alpha)^2}$$

On peut donc appliquer le critère de Borel Cantelli et conclure que

$$\bar{Y}_{k_n} - E(\bar{Y}_{k_n}) \to 0.$$

Etape 4) : Pour $n \geq 1$, on note $m(n)$ l'entier tel que

$$\lfloor \alpha^{m(n)} \rfloor = k_{m(n)} \leq n \leq \lfloor \alpha^{m(n+1)} \rfloor.$$

On a que

$$\begin{aligned} \liminf \frac{\sum_{i=1}^{n} Y_i}{n} &\geq \liminf \frac{\sum_{i=1}^{k_{m(n)}} Y_i}{k_{m(n+1)}} \\ &= \liminf \bar{Y}_{k_{m(n)}} \frac{k_{m(n)}}{k_{m(n+1)}} \\ &= \frac{EX_1}{\alpha} \end{aligned}$$

Ici il faut voir que

$$\frac{\lfloor \alpha^n \rfloor}{\lfloor \alpha^{n+1} \rfloor} \to \frac{1}{\alpha}$$

De manière similaire on montre que

$$\limsup \frac{\sum_{i=1}^{n} Y_i}{n} \leq \alpha E(X_1)$$

Comme α est arbitrairement proche de 1 on conclut que

$$E(X_1) \leq \liminf \frac{\sum_{i=1}^{n} Y_i}{n} \leq \limsup \frac{\sum_{i=1}^{n} Y_i}{n} \leq E(X_1)$$

et on obtient le résultat attendu. $\square$

Théorème 76 (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 si on définit*

$$\mu_n^\omega = \frac{\delta_{X_1(\omega)} + \dots \delta_{X_n(\omega)}}{n}$$

$$P\left(\left\{\omega \in \Omega,\ \mu_n^\omega \underset{n\to\infty}{\overset{\text{étroit}}{\longrightarrow}} P_{X_1}\right\}\right) = 1.$$

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

Démonstration. Comme $C_0(\mathbb{R}^d)$ est séparable, on peut trouver un ensemble $\mathcal{H}$ de fonctions dense et dénombrable. Par la loi forte des grands nombre pour tout $\phi \in \mathcal{H}$

$$\frac{\phi(X_1) + \dots \phi(X_n)}{n} = \int \phi d\mu_n^\omega \to E(\phi(X_1)) \ p.s$$

Ainsi

$$P(\int \phi d\mu_n \to \int \phi dP_{X_1}) = 1$$

Comme $\mathcal{H}$ est dénombrable on a

$$P(\forall \phi \in \mathcal{H} \int \phi d\mu_n \to \int \phi dP_{X_1}) = 1$$

ce qui revient à dire

$$P\left(\left\{\omega \in \Omega,\ \mu_n^\omega \underset{n\to\infty}{\overset{\text{étroit}}{\longrightarrow}} P_{X_1}\right\}\right) = 1$$

$\square$

6.2 Théorème central limite

Théorème 77. *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*

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

Démonstration. Quitte à centrer et à réduire les v.a on peut supposer que $E[X_1] = 0$ et $\sigma = 1$. Soit $T \in \mathbb{R}$, on a

$$\begin{aligned}
\Phi_{\sqrt{n}\bar{X}_n}(t) &= E(e^{it\sqrt{n}\bar{X}_n}) \\
&= E(\prod_{i=1}^n e^{it/\sqrt{n}X_i}) \\
&= \Big(E(e^{it\sqrt{n}X_1})\Big)^n \\
&= \big(\Phi_{X_1}(t\sqrt{n})\big)^n
\end{aligned}$$

Or $E(X^2) \leq \infty$ donc $\Phi_{X_1} \in C^2$ et on a

$$\Phi_{X_1}(t) = 1 - \frac{t^2}{2} + \circ(t^2)$$

Donc

$$\Phi_{\sqrt{n}\bar{X}_n}(t) = \left(1 - \frac{t^2}{2n} + \circ(\frac{t^2}{n})\right)^n.$$

Pour $z \in \mathbb{C}$ on a

$$\begin{aligned}
\left|e^z - \Big(1 + \frac{z}{n}\Big)^n\right| &= \left|\sum_{k=0}^{\infty} \frac{z^k}{k!} - \sum_{k=0}^{n} \frac{n!}{k!(n-k)!}\frac{z^k}{n^k}\right| \\
&\leq \sum_{k=0}^{n} \left|\frac{1}{k!} - \frac{n!}{k!(n-k)!n^k}\right| |z|^k + \sum_{k\geq n+1} \frac{|z|^k}{k!}
\end{aligned}$$

On remarque que la quantité

$$\frac{1}{k!} - \frac{n!}{k!(n-k)!n^k} = \frac{1}{k!}\left(1 - \frac{n(n-1)\ldots(n-(k-1))}{n^k}\right) \geq 0$$

$$\begin{aligned}
\left|e^z - \Big(1 + \frac{z}{n}\Big)^n\right| &\leq \sum_{k=0}^{n} \left(\frac{1}{k!} - \frac{n!}{k!(n-k)!n^k}\right) |z|^k + \sum_{k\geq n+1} \frac{|z|^k}{k!} \\
&= e^{|z|} - \left(1 + \frac{|z|}{n}\right)^n \to_n 0
\end{aligned}$$

Ainsi

$$\Phi_{\sqrt{n}\bar{X}_n}(t) \to e^{-t^2/2}$$

qui est la fonction caractéristique de la $\mathcal{N}(0,1)$ et on applique le théorème de Levy. □

Corollaire 78. *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 $a \geq 0$,*

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

6.3 Méthode Delta

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

Théorème 79. *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.

Remarque 64. Un cadre d'application est celui du théorème central limit. En effet sous les bonnes hypothèse, i. e (X_i) i.i.d L^2 avec $E(X_1) = \theta$, $VarX_1 = 1$ on a que

$$\sqrt{n}(\bar{X}_n - \theta) \xrightarrow[n\to\infty]{\mathcal{L}} Z$$

avec $Z \sim \mathcal{N}(0,1)$. Ona. donc pour tout ϕ différentiable

$$\sqrt{n}(\phi(\bar{X}_n) - \phi(\theta)) \xrightarrow[n\to\infty]{\mathcal{L}} \phi'(\theta)Z$$

Dans ce cadre la preuve n'est pas difficile. On a que

$$g(\bar{X}_n) = g(\theta) + g'(\theta)(\bar{X}_n - \theta) + \circ(|\bar{X}_n - \theta|)$$

et donc

$$\sqrt{n}(g(\bar{X}_n) - g(\theta)) = g'(\theta)\sqrt{n}(\bar{X}_n - \theta) + \circ(\sqrt{n}|\bar{X}_n - \theta|)$$

Or $\sqrt{n}|\bar{X}_n - \theta|$ tend en loi vers 0 donc en proba vers 0. Donc $\circ(\sqrt{n}|\bar{X}_n - \theta|)$ tend en proba vers 0. En effet si $Z_n = \circ(\sqrt{n}|\bar{X}_n - \theta|)$ alors

$$Z_n = \frac{Z_n}{\sqrt{n}|\bar{X}_n - \theta|}\sqrt{n}|\bar{X}_n - \theta|$$

et

$$\frac{Z_n}{\sqrt{n}|\bar{X}_n - \theta|}$$

tend vers 0 p.s. On peut alors appliquer le Lemme de Slutsky et conclure que

$$\sqrt{n}(g(\bar{X}_n) - g(\theta)) \xrightarrow[n\to\infty]{\mathcal{L}} g'(\theta)Z$$

Chapitre 7

Introduction à la statistique

7.1 Structure statistique

Définition 46. 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 47. 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 $X_i(\omega_1, \dots, \omega_n) := X_i(\omega_i)$. Si l'on munit $(\Omega^n, \mathcal{A}^{\otimes n})$ de la probabilité $P_\theta^{\otimes n}$ alors $X_1, \dots, X_n$ sont indépendantes et de loi P_θ. On dit souvent que $(X_1(\omega), \dots, X_n(\omega))$ est un *n-échantillon*.

Remarque 65. 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 48. 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 49. Soit $\big(\Omega, \mathcal{A}, (P_\theta)_{\theta\in\Theta}\big)$ un modèle statistique et $(E, \mathcal{B})$ un espace mesurable. On appelle *variable statistique* 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 50. 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 51. 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 66. 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 52. 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 53. 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 $X_1, \ldots, X_n$ 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 67. Il faut noter que toutes les notions de convergence dépendent de θ puisque leur définition fait intervenir la probabilité sous-jacente.

7.2 Quelques techniques d'estimation

7.2.1 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

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

7.2.2 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 68. Un tel estimateur n'existe pas forcément. S'il existe, il n'est pas forcément unique.

7.2.3 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

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

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

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

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

7.3 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.

7.3.1 Definitions

Définition 54. 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 55. 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 $1 - \alpha$.

7.3.2 Mesures de qualité des tests

Définition 56. 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 57. 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 58. 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 59. 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 60. 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.

7.3.3 Construction de tests

Les tests déterministes les plus simples considèrent une statistique $T : \Omega \to \mathbb{R}$ et décident d'accepter ou de refuser H_0 grâce à une *valeur critique* $c \in \mathbb{R}$: on accepte H_0 si T est observée dans la *région d'acceptation* $]-\infty; c[$, tandis qu'on rejette H_0 si T est observée dans la *région de rejet* $[c\ ; +\infty[$. Remarquons que le niveau $\sup_{\theta \in \Theta_0} P_\theta(T \geq c)$ de ces tests $\Phi = \mathbb{1}_{T \geq c}$ est décroissant en fonction de c.

En pratique, on fixe un seuil α a priori, et on choisit une valeur critique c assez grande pour que le test $\Phi = \mathbb{1}_{T \geq c}$ soit de seuil α, c'est-à-dire tel que la probabilité d'erreur de première espèce soit inférieure à α. Plus α est choisi petit, plus la règle de décision est conservative, au sens où elle ne rejette H_0 que très rarement.

Définition 61. La *p-valeur* (ou *seuil de signification observé*) de la réalisation $t \in \mathbb{R}$ d'une statistique $T : \Omega \to \mathbb{R}$ est $\sup_{\theta \in \Theta_0} P_\theta(T \geq t)$: c'est le plus petit seuil α pour lequel les tests de seuil α de la forme $\Phi = \mathbb{1}_{T \geq c}$ conduisent à rejeter H_0.

On considère maintenant 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 $\Phi(\omega) = \mathbb{1}_{\ell(\omega) \geq c}$ où c est la valeur critique à déterminer afin de contrÃ´ler 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 80 (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 81. *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'.*