

UNIVERSITÉ DE RENNES 1

FPR

FONDEMENTS
DES
PROBABILITÉS

AUTEUR
JÜRGEN ANGST

NOTES DE COURS
VICTOR LECERF



2020–2021

Table des matières

1	Espaces de probabilité, vocabulaire probabiliste	5
1.1	Espaces probabilisés	5
1.2	Propriétés élémentaires des mesures de probabilités	6
1.3	Limites supérieures et inférieures d'ensembles	7
1.4	Complétion et prolongement de mesure	8
1.4.1	Ensembles négligeables, tribu complété	8
1.4.2	Classes monotones	8
2	Variables aléatoires	11
2.1	Définition et premières propriétés	11
2.1.1	Variables et vecteurs aléatoires	11
2.1.2	Loi d'une variable aléatoire	12
2.1.3	Variables aléatoires indépendantes	12
2.2	Fonction de répartition et densité	14
2.2.1	Fonction de répartition	14
2.2.2	Densité	15
2.3	Variables aléatoires usuelles	18
2.3.1	Variables aléatoires discrètes	18
2.3.2	Variables aléatoires à densité	19
3	Espérance et moments	21
3.1	Espérance et moments	21
3.1.1	Espérance : définitions et premières propriétés	21
3.1.2	Moments d'ordre supérieurs	24
3.1.3	Moments de variables usuelles	26
3.2	Espérance et identification de loi	27
3.2.1	Identification et fonctions tests	27
3.2.2	Espérance et indépendance	27
3.2.3	Le problème de moments	28
3.3	Transformées exponentielles	30
3.3.1	Fonction caractéristique	30
3.3.2	Autres transformées exponentielles	32
3.4	Probabilités, lois et espérances conditionnelles	34
3.4.1	Probabilités conditionnelles	34
3.4.2	Loi et espérance conditionnelle	35

4	Convergence de variables aléatoires	37
4.1	Différents modes de convergence	37
4.1.1	Convergence presque sûre	38
4.1.2	Convergence en probabilité	39
4.1.3	Convergence \mathbb{L}^p	40
4.1.4	Convergence en loi	40
4.2	Articulation des modes de convergence	42
4.2.1	Convergence presque sûre en probabilité	42
4.2.2	Convergences \mathbb{L}^p , p.s, \mathbb{P}	42
4.2.3	Convergence en loi et autres modes	43
5	Théorèmes limites	45
5.1	Loi des grands nombres (LGN)	45
5.1.1	Loi faible des grands nombres	45
5.1.2	Loi forte des grands nombres	46
5.2	Théorème limite central (TLC)	46
5.2.1	Théorème limite central	46
5.2.2	Retour sur les applications de la loi des grands nombres	48
6	Vecteurs gaussiens	49
6.1	Vecteurs gaussiens	49
6.1.1	Définitions et propriétés élémentaires	49
6.1.2	Théorème limite central multidimensionnel	53
6.2	Projections orthogonales	55
6.2.1	Théorème de COCHRAN	55
6.2.2	Test d'adéquation du χ^2	56
6.2.3	Espérances conditionnelles gaussiennes	57

Chapitre 1

Espaces de probabilité, vocabulaire probabiliste

La théorie des probabilités remonte au XVIII^{ème} siècle et a été au début du XXI^{ème}. Il s'agit à l'origine de modéliser mathématiquement des phénomènes complexes dont le résultat ne peut être prédit, ou dont la modélisation déterministe est trop complexe pour être mise en œuvre effectivement, comme par exemple un lancer de dé, d'une pièce, ou la trajectoire d'une particule dans un fluide.

Au lieu de se focaliser sur une issue précise de l'expérience, on considère l'ensemble des résultats possibles et on leur alloue un "poids" selon qu'ils sont plus ou moins probables.

1.1 Espaces probabilisés

Soit Ω un ensemble, et $\mathcal{P}(\Omega)$ l'ensemble de ses parties. On rappelle la définition d'une tribu ainsi que de quelques concepts.

Définition 1.1.1 (tribu). Une tribu \mathcal{F} sur Ω est une classe de parties de Ω ($\mathcal{F} \subset \mathcal{P}(\Omega)$) telle que

1. $\Omega \in \mathcal{F}$.
2. Si $A \in \mathcal{F}$, alors $A^c \in \mathcal{F}$.
3. Si $(A_n)_n \in \mathcal{F}^{\mathbb{N}}$, alors $\bigcup_{n \in \mathbb{N}} A_n \in \mathcal{F}$.

Exemples. $\{\emptyset, \Omega\}$ (tribu grossière), $\mathcal{P}(\Omega)$ (tribu pleine), $\{\emptyset, A, A^c, \Omega\}$ pour $A \subset \Omega$ quelconque.

Remarques.

- (Ω, \mathcal{F}) est dit être un **espace mesurable** (ou **espace probabilisable**), et les éléments de \mathcal{F} sont appelés les *événements*.
 - Une tribu contient toujours l'ensemble vide (on peut remplacer d'ailleurs le premier axiome par " $\emptyset \in \mathcal{F}$ ").
 - Une tribu est stable par intersection au plus dénombrable.
 - Une intersection de tribus est encore une tribu. On peut alors construire pour tout $\mathcal{A} \subset \mathcal{P}(\Omega)$ la plus petite tribu contenant \mathcal{A} (comme l'intersection de toutes les tribus contenant \mathcal{A}), généralement notée $\sigma(\mathcal{A})$.
-

1.2. Propriétés élémentaires des mesures de probabilités

- Lorsque Ω est dénombrable, on le munit généralement de la tribu pleine $\mathcal{P}(\Omega)$.
- Lorsque Ω est muni d'une topologie τ , on le munit de la tribu borélienne $\mathcal{B}(\Omega) = \sigma(\tau)$.

Définition 1.1.2. Soit (Ω, \mathcal{F}) un espace mesurable. Une mesure de probabilité (ou probabilité) est une mesure sur \mathcal{F} à valeurs dans $[0, 1]$ vérifiant $\mathbb{P}(\Omega) = 1$. On dit que $(\Omega, \mathcal{F}, \mathbb{P})$ est un **espace probabilisé** (ou **espace de probabilité**).

Exemples.

- $([0, 1], \mathcal{B}([0, 1]), \lambda)$ où λ est la mesure de LEBESGUE.
- (Mesure de DIRAC) Soit $(\Omega, \mathcal{P}(\Omega))$ avec Ω un ensemble, et soit $a \in \Omega$. On note $\delta_a := \mathbb{1}_{\{a\}}$ la mesure de DIRAC en a . $(\Omega, \mathcal{P}(\Omega), \delta_a)$ est un espace probabilisé.
- Soit $(p_n)_n \in \mathbb{R}_+$ une famille de réels de somme 1. Alors,

$$\mathbb{P} = \sum_{n \in \mathbb{N}} p_n \delta_n$$

est bien définie, et est une probabilité sur $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$. On a alors

$$\mathbb{P}(A) = \sum_{n \in \mathbb{N}} \mathbb{1}_A(n).$$

1.2 Propriétés élémentaires des mesures de probabilités

Proposition 1.2.1. Une mesure de probabilité possède les mêmes propriétés élémentaires qu'une mesure quelconque. En outre, elle vérifie l'additivité sur deux événements disjoints, la croissance, la formule de la mesure de la différence symétrique, la σ -sous-additivité, la continuité à gauche et à droite (en termes d'ensembles).

Remarque. Sur $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé, et pour tout $A, B \in \mathcal{F}$, on $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$. La proposition suivante généralise cette formule.

Proposition 1.2.2. Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé, et A_1, \dots, A_n des événements de \mathcal{F} . Alors,

$$\mathbb{P}\left(\bigcup_{i=1}^n A_i\right) = \sum_{k=1}^n \left((-1)^{k+1} \sum_{1 \leq i_1 < i_2 < \dots < i_k \leq n} \mathbb{P}\left(\bigcap_{j=1}^k A_{i_j}\right) \right)$$

Exemple. Pour $n = 3$,

$$\mathbb{P}(A \cup B \cup C) = \mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(C) - \mathbb{P}(A \cap B) - \mathbb{P}(B \cap C) - \mathbb{P}(C \cap A) + \mathbb{P}(A \cap B \cap C).$$

1.3 Limites supérieures et inférieures d'ensembles

Soient $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité et $(A_n)_{n \in \mathbb{N}}$ une suite d'évènements. Deux évènements complexes, mais très utiles, sont la limite inférieure et la limite supérieure.

Définition 1.3.1. On appelle limite inférieure (resp. supérieure) de la famille d'évènements $(A_n)_n \in \mathcal{F}^{\mathbb{N}}$ l'ensemble

$$\liminf_{n \rightarrow \infty} \bigcup_{n \geq 1} \bigcap_{k > n} A_k$$

$$\left(\text{resp. } \limsup_{n \rightarrow \infty} \bigcap_{n \geq 1} \bigcup_{k > n} A_k \right).$$

On retrouve sur ces ensembles de propriétés analogues aux limites inférieures et supérieures sur les suites numériques.

Proposition 1.3.1. Soit $(A_n)_n \in \mathcal{F}^{\mathbb{N}}$. On a

- (i) $\liminf_{n \rightarrow \infty} A_n \subset \limsup_{n \rightarrow \infty} A_n$.
- (ii) $(\limsup_{n \rightarrow \infty} A_n)^c = \liminf_{n \rightarrow \infty} A_n^c$, et $(\liminf_{n \rightarrow \infty} A_n)^c = \limsup_{n \rightarrow \infty} A_n^c$.

Démonstration. (i) Soit $n \in \mathbb{N}$. Pour tout $p > n$, on a $\bigcap_{k > n} A_k \subset A_p$. Ainsi, $\bigcap_{k > n} A_k \subset \bigcup_{p > m} A_p$ pour tout $m \in \mathbb{N}$, et alors $\bigcap_{k > n} A_k \subset \bigcap_{m \geq 1} \bigcup_{p > m} A_p = \limsup_{n \rightarrow \infty} A_n$. Cette inclusion étant vraie pour tout $n \in \mathbb{N}$, elle reste vraie après passage à l'union sur $n \in \mathbb{N}$ à gauche. □

Proposition 1.3.2. Soit $(A_n)_n \in \mathcal{F}^{\mathbb{N}}$ une suite d'évènements. Alors,

$$\liminf_{n \rightarrow \infty} \mathbb{1}_{A_n} = \mathbb{1}_{\liminf_{n \rightarrow \infty} A_n} \quad \text{et} \quad \limsup_{n \rightarrow \infty} \mathbb{1}_{A_n} = \mathbb{1}_{\limsup_{n \rightarrow \infty} A_n}$$

Définition 1.3.2. Une suite d'évènements $(A_n)_n \in \mathcal{F}^{\mathbb{N}}$ est dite convergente si les limites inférieures et supérieures coïncident. Si elle existe, on la note $\lim_{n \rightarrow \infty} A_n$.

Proposition 1.3.3. Soit $(A_n)_n \in \mathcal{F}^{\mathbb{N}}$ une suite d'évènements. Alors,

$$\mathbb{P} \left(\liminf_{n \rightarrow \infty} A_n \right) \leq \liminf_{n \rightarrow \infty} \mathbb{P}(A_n) \leq \limsup_{n \rightarrow \infty} \mathbb{P}(A_n) \leq \mathbb{P} \left(\limsup_{n \rightarrow \infty} A_n \right).$$

En particulier si $(A_n)_n$ converge,

$$\mathbb{P} \left(\lim_{n \rightarrow \infty} A_n \right) = \lim_{n \rightarrow \infty} \mathbb{P}(A_n).$$

1.4 Complétion et prolongement de mesure

1.4.1 Ensembles négligeables, tribu complété

On fixe dans cette sous-section un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$.

Définition 1.4.1 (négligeabilité). Soit $N \in \mathcal{P}(\Omega)$. L'ensemble N est dit négligeable s'il est inclus dans un ensemble de mesure nulle. Deux $A, B \in \mathcal{P}(E)$ ensembles sont dits être égaux presque sûrement si $A \Delta B = A \setminus B \cup B \setminus A$ est négligeable.

Proposition 1.4.1. Soit \mathcal{N} l'ensemble des parties négligeables de Ω . Alors,

- (i) $\emptyset \in \mathcal{N}$,
- (ii) Si $A \in \mathcal{N}$ et $B \in \mathcal{P}(\Omega)$ sont tels que $B \subset A$, alors $B \in \mathcal{N}$,
- (iii) \mathcal{N} est stable par union au plus dénombrable,
- (iv) \mathcal{N} est stable par intersection quelconque.

Définition 1.4.2 (tribu complétée). On appelle la tribu complétée de \mathcal{F} par rapport à \mathbb{P} la tribu

$$\overline{\mathcal{F}} = \sigma(\mathcal{F} \cup \mathcal{N})$$

où \mathcal{N} désigne l'ensemble des parties négligeables de Ω .

Proposition 1.4.2. Soit $A \in \mathcal{F}$. Sont équivalents :

- (i) $A \in \overline{\mathcal{F}}$.
- (ii) Il existe $B, C \in \mathcal{F}$, tels que $B \subset A \subset C$ et $\mathbb{P}(C \setminus B) = 0$.
- (iii) Il existe $B \in \mathcal{F}$ et $N \in \mathcal{N}$ tels que $A = B \cup N$.
- (iv) Il existe $B \in \mathcal{F}$ tel que $A = B$ presque sûrement, i.e que $A \Delta B \in \mathcal{N}$.

Proposition 1.4.3. Soient $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité et $\overline{\mathcal{F}}$ la tribu complétée par rapport à \mathcal{F} . Alors, l'application

$$\overline{\mathbb{P}} : \begin{cases} \overline{\mathcal{F}} & \longrightarrow [0, 1] \\ A = B \cup N & \longmapsto \mathbb{P}(B) \end{cases}$$

est bien définie, est une probabilité et l'unique prolongement de \mathbb{P} à $\overline{\mathcal{F}}$ tel que $\overline{\mathbb{P}}|_{\mathcal{F}} = \mathbb{P}$.

Proposition 1.4.4. L'ensemble des négligeables pour $\overline{\mathbb{P}}$ coïncident avec ceux de \mathbb{P} .

1.4.2 Classes monotones

On rappelle ici un procédé d'extension des définitions de certains objets sur les tribus après les avoir définis sur une classe restreinte d'ensemble.

Définition 1.4.3. Une famille \mathcal{M} de parties de Ω est appelée **classe monotone** (ou λ -système) si

(i) $\Omega \in \mathcal{M}$,

(ii) \mathcal{M} est stable par différence :

$$\forall A, B \in \mathcal{M}, (B \subset A) \implies (A \setminus B) \in \mathcal{M},$$

(iii) \mathcal{M} est stable par union croissante :

$$\forall (A_n)_n \in \mathcal{M}^{\mathbb{N}}, (\forall n \in \mathbb{N}, A_n \subset A_{n+1}) \implies \left(\bigcup_{n \in \mathbb{N}} A_n \in \mathcal{M} \right).$$

Remarques.

- Une intersection quelconque de classes monotones est encore une classe monotone.
- Une tribu est une classe monotone, car $A \setminus B = A \cap B^c$.
- Une classe monotone stable par intersection finie est une tribu.
- Comme pour les tribus, on peut définir pour toute famille de partie $\mathcal{E} \subset \mathcal{P}(\Omega)$ la plus petite classe monotone contenant \mathcal{E} comme l'intersection de toutes les classes monotones contenant \mathcal{E} , notée $\mathcal{M}(\mathcal{E})$.

Théorème 1.4.1 (des classes monotone (DYNKIN)). Soit \mathcal{A} une famille de parties de Ω stable par intersection finie (appelée π -système). Alors, $\mathcal{M}(\mathcal{A}) = \sigma(\mathcal{A})$.

Démonstration. Probablement quelque part dans le cours d'intégration. □

Théorème 1.4.2. Soit (Ω, \mathcal{F}) un espace mesurable Si deux probabilités \mathbb{P}_1 et \mathbb{P}_2 sur \mathcal{F} coïncident sur une partie $\mathcal{A} \subset \Omega$ stable par intersections finies, alors elles coïncident sur $\sigma(\mathcal{A})$.

Démonstration. Soit $\mathcal{M} = \{B \in \mathcal{F} \mid \mathbb{P}_1(B) = \mathbb{P}_2(B)\}$. Alors $\Omega \in \mathcal{M}$, et si $B, C \in \mathcal{M}$ avec $B \subset C$, alors $C \setminus B \in \mathcal{M}$ (calcul simple). De plus, si $(B_n)_n \in \mathcal{M}^{\mathbb{N}}$ est une suite croissante, alors $\bigcup_{n \in \mathbb{N}} B_n \in \mathcal{M}$ car

$$\mathbb{P}_1 \left(\bigcup_{n \in \mathbb{N}} B_n \right) = \lim_{n \rightarrow \infty} \mathbb{P}_1(B_n) = \lim_{n \rightarrow \infty} \mathbb{P}_2(B_n) = \mathbb{P}_2 \left(\bigcup_{n \in \mathbb{N}} B_n \right).$$

On en déduit que \mathcal{M} est une classe monotone telle que $\mathcal{A} \subset \mathcal{M}$. Ainsi, $\mathcal{M}(\mathcal{A}) \subset \mathcal{M}$. Mais \mathcal{A} étant un π -système, le théorème des classes monotones affirme que $\mathcal{M}(\mathcal{A}) = \sigma(\mathcal{A})$, d'où $\sigma(\mathcal{A}) \subset \mathcal{M}$. □

Chapitre 2

Variables aléatoires

2.1 Définition et premières propriétés

2.1.1 Variables et vecteurs aléatoires

Définition 2.1.1. Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité, et (E, \mathcal{E}) un espace mesurable. On appelle variable aléatoire à valeurs dans E toute application mesurable de (Ω, \mathcal{F}) dans (E, \mathcal{E}) .

Remarques.

- On notera souvent “v.a” pour variable aléatoire.
- Si $(E, \mathcal{E}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, on parle de *variable aléatoire réelle* (v.a.r).
- Si $(E, \mathcal{E}) = (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$, on parle de *vecteur aléatoire*. Si $X = (X_1, \dots, X_n)$ est un vecteur aléatoire, la composante $X_i : (\Omega, \mathcal{F}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ est appelé la i -ième marginale du vecteur X (pour $i \in \llbracket 1, n \rrbracket$).

Proposition 2.1.1. Soit $X : (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow (E, \mathcal{E})$ une application et $\mathcal{C} \subset \mathcal{P}(E)$. Alors,

$$X^{-1}(\sigma(\mathcal{C})) = \sigma(X^{-1}(\mathcal{C})).$$

En particulier, si $\sigma(\mathcal{C}) = \mathcal{E}$, pour vérifier que X est mesurable, il suffit de vérifier que $X^{-1}(\mathcal{C}) \subset \mathcal{F}$.

Remarque. La somme, le produit, le quotient, le minimum, le maximum, la limite supérieure, la limite inférieure, de variables aléatoires est une variable aléatoire (résultat général sur les fonctions mesurables).

Proposition 2.1.2. Soit (E, d) un espace métrique, et soit $(X_n)_n$ une suite de variables aléatoires de (Ω, \mathcal{F}) dans $(E, \mathcal{B}(E))$. Si $(X_n)_n$ converge simplement vers une fonction X , alors X est une variable aléatoire.

En bref, la limite simple de fonctions mesurables est mesurable.

Proposition 2.1.3. *Toute variable aléatoire réelle est **limite simple** de variables aléatoires étagées. De plus, si X est à valeurs positives, on peut choisir la suite **croissante**.*

2.1.2 Loi d'une variable aléatoire

Définition 2.1.2. *Soit $X : (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow (E, \mathcal{E})$ une variable aléatoire. On appelle loi de X la mesure de probabilité sur (E, \mathcal{E}) la mesure image \mathbb{P}_X de \mathbb{P} par X , définie pour tout $A \in \mathcal{E}$ par*

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

Remarque. Si $X = (X_1, \dots, X_n)$ un vecteur aléatoire à valeurs dans $E = (\prod_{i=1}^n E_i, \otimes_{i=1}^n \mathcal{E}_i)$, alors la loi \mathbb{P}_X est appelée *loi jointe* des X_i . Les lois \mathbb{P}_{X_i} sont appelées les *lois marginales*. Pour $i \in \llbracket 1, n \rrbracket$, si $A \subset E_i$, on a

$$\mathbb{P}_{X_i}(A) = \mathbb{P}_X(E_1 \times \dots \times E_{i-1} \times A \times E_{i+1} \times \dots \times E_n).$$

Définition 2.1.3. *On dit que deux variables aléatoires $X, Y : (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow (E, \mathcal{E})$ ont même loi si $\mathbb{P}_X = \mathbb{P}_Y$. On note alors $\mathcal{L}(X) = \mathcal{L}(Y)$.*

Remarque. Cette définition n'interdit pas particulièrement que X et Y soient définies sur des espaces de probabilités différents.

2.1.3 Variables aléatoires indépendantes

Définition 2.1.4. *Sur un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$, deux évènements A et $B \in \mathcal{F}$ sont dits indépendants si*

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

Plus généralement, les évènements d'une famille $(A_i)_{i \in I}$ sont dits :

— *mutuellement indépendants si pour tout $J \subset I$ fini,*

$$\mathbb{P}\left(\prod_{i \in J} A_i\right) = \prod_{i \in J} \mathbb{P}(A_i).$$

— *deux à deux indépendants si,*

$$\forall i, j \in I, (i \neq j) \implies (\mathbb{P}(A_i \cap A_j) = \mathbb{P}(A_i) \cdot \mathbb{P}(A_j)).$$

Notation. On note parfois $A \perp B$.

Définition 2.1.5. Deux familles d'évènements $\mathcal{G}, \mathcal{H} \subset \mathcal{F}$ sont dites **indépendantes** si pour tout $(A, B) \in \mathcal{G} \times \mathcal{H}$, A et B sont indépendants. Plus généralement, Les éléments d'une famille $(\mathcal{G}_i)_{i \in I}$ de familles d'évènements sont dits :

— **mutuellement indépendants**, si pour $J \subset I$ fini :

$$\forall (A_j)_{j \in J} \in \prod_{j \in J} \mathcal{G}_j, \mathbb{P} \left(\bigcap_{j \in J} A_j \right) = \prod_{j \in J} \mathbb{P}(A_j).$$

— **deux à deux indépendants**, si \mathcal{G}_i et \mathcal{G}_j sont indépendants pour tout $i \neq j$.

Proposition 2.1.4. Si $(\mathcal{A}_i)_{i \in I}$ est une famille de π -systèmes. Alors, les $(\mathcal{A}_i)_i$ sont mutuellement indépendants (resp. deux à deux indépendants) si, et seulement si, les $(\sigma(\mathcal{A}_i))_i$ sont mutuellement indépendants (resp. deux à deux indépendants).

Démonstration. Application du théorème des classes monotones. □

Définition 2.1.6. Soit $X : (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow (E, \mathcal{E})$. On note

$$\sigma(X) = X^{-1}(\mathcal{E}).$$

C'est la plus petite tribu sur Ω rendant X mesurable.

Définition 2.1.7. Deux variables aléatoires $X : (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow (E_1, \mathcal{E}_1)$ et $Y : (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow (E_2, \mathcal{E}_2)$ sont dites être deux à deux indépendantes si les tribus $\sigma(X)$ et $\sigma(Y)$ sont deux à deux indépendantes, i.e :

$$\forall A \in \mathcal{E}_1, \forall B \in \mathcal{E}_2, \mathbb{P}((X \in A) \wedge (Y \in B)) = \mathbb{P}(X \in A) \cdot \mathbb{P}(Y \in B).$$

De même, elles sont dites mutuellement indépendantes

Remarques.

- Il n'est pas nécessaire que X et Y soient à valeurs dans le même espace mesurable pour que la définition fasse sens.
- On définit de même l'indépendance d'une famille de variables aléatoires $(X_i)_{i \in I}$ comme l'indépendance deux à deux des variables aléatoires de la famille.

Proposition 2.1.5. Si $X = (X_1, \dots, X_n)$ est un vecteur aléatoire à valeurs dans $(\prod_{i=1}^n E_i, \otimes_{i=1}^n \mathcal{E}_i)$, alors les variables indépendantes X_i sont mutuellement indépendantes si, et seulement si, pour tout $(A_1, \dots, A_n) \in \prod_{i=1}^n \mathcal{E}_i$,

$$\mathbb{P}_X(A_1 \times \dots \times A_n) = \mathbb{P}_{X_1}(A_1) \cdots \mathbb{P}_{X_n}(A_n).$$

Ce qui est équivalent à dire que

$$\mathbb{P}_X = \mathbb{P}_{X_1} \otimes \cdots \otimes \mathbb{P}_{X_n}.$$

Démonstration. Par définition,

$$\begin{aligned} \mathbb{P}_X(A_1 \times \cdots \times A_n) &= \mathbb{P}(X \in (A_1, \dots, A_n)) \\ &= \mathbb{P}(X_1 \in A_1, X_2 \in A_2, \dots, X_n \in A_n) \\ &= \mathbb{P}(X_1 \in A_1) \cdots \mathbb{P}(X_n \in A_n) \\ &= \mathbb{P}_{X_1}(A_1) \cdots \mathbb{P}_{X_n}(A_n). \end{aligned}$$

□

2.2 Fonction de répartition et densité

2.2.1 Fonction de répartition

On étudie dans cette sous-section les fonctions de répartition¹ qui sont un ensemble de fonctions qui nous permettront de caractériser les lois des variables aléatoires réelles.

Définition 2.2.1. On appelle fonction de répartition d'une variable aléatoire réelle la fonction $F_X : [0, 1] \rightarrow \mathbb{R}$ vérifiant pour tout $x \in \mathbb{R}$,

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

Proposition 2.2.1. La fonction de répartition caractérise la loi d'une variable aléatoire réelle, au sens où $(F_X = F_Y) \iff (\mathbb{P}_X = \mathbb{P}_Y)$.

Démonstration. La famille $\{]-\infty, x] \mid x \in \mathbb{R}\}$ engendre les boréliens de \mathbb{R} , et est stable par intersection finie. On conclue par le théorème des classes monotones.

□

Proposition 2.2.2. Soit F_X la fonction de répartition d'une variable aléatoire réelle X . On a les résultats suivants.

- (i) F_X est croissante.
- (ii) La limite de F_X en $-\infty$ est 0.
- (iii) La limite de F_X en ∞ est 1.
- (iv) La fonction F_X est CÀDLÀG (continue à droite avec limite à gauche), i.e

$$F_X(x^-) = \lim_{y \rightarrow x^-} F_X(y) = \mathbb{P}(X < x), \text{ et est continue à droite.}$$

- (v) Pour tout $a, b \in \mathbb{R}$ avec $a < b$,

$$\mathbb{P}(X \in [a, b]) = F_X(b) - F_X(a^-).$$

1. Distribution function en anglais.

(vi) Si $x_0 \in \mathbb{R}$ n'est pas un atome de X , i.e $\mathbb{P}(X = x_0) \neq 0$, alors F_X est continue en x_0 .

Démonstration. (i) .

□

Corollaire 2.2.1. Une variable aléatoire réelle possède un nombre au plus dénombrable d'atomes.

Démonstration. Une fonction monotone possède un nombre au plus dénombrable de points de discontinuité.

□

Proposition 2.2.3. Soit $F : \mathbb{R} \rightarrow [0, 1]$ une fonction CÀDLÀG croissante, avec

$$\lim_{x \rightarrow -\infty} F(x) = 0 \quad \text{et} \quad \lim_{x \rightarrow \infty} F(x) = 1.$$

Alors, F est la fonction de répartition d'une certaine variable aléatoire réelle.

2.2.2 Densité

Définition 2.2.2. Soient μ et ν deux mesures σ -finies sur un espace mesurable (Ω, \mathcal{F}) . La mesure μ est dite **absolument continue** par rapport à ν , si :

$$\forall A \in \mathcal{F}, (\nu(A) = 0) \implies (\mu(A) = 0).$$

On note dans ce cas, $\mu \ll \nu$.

Théorème 2.2.1 (RADON-NIKODYM). En conservant la même quantification, si $\mu \ll \nu$, alors il existe une fonction mesurable f telle que pour tout $A \in \mathcal{F}$,

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

Dans ce cas, et si μ et ν sont positives, et que μ est finie, alors f est positive et élément $L^1(\Omega, \nu)$.

Remarque. Cette fonction f est appelée *densité* (ou *dérivée de RADON-NIKODYM*) de μ par rapport à ν .

Définition 2.2.3 (v.a.r à densité). Une variable aléatoire réelle X est dite être à **densité** si P_X est absolument continue par rapport à la mesure de LEBESGUE sur \mathbb{R} . Dans ce cas la dérivée de RADON-NIKODYM f_X est appelée **densité de la loi de X** (ou de X). Cette fonction est positive et intégrable pour la mesure de LEBESGUE.

Remarques.

— On a immédiatement pour tout $A \in \mathcal{F}$,

$$\mathbb{P}_X(A) = \mathbb{P}(X \in A) = \int_A f_X(x) dx.$$

En particulier, pour tout $a, b \in \mathbb{R}$ avec $a < b$,

$$\mathbb{P}_X([a, b]) = \int_a^b f_X(x) dx.$$

— Réciproquement, si f est une fonction mesurable positive d'intégrale 1 par rapport à la mesure de LEBESGUE, la formule

$$\mathbb{P}(A) = \int_A f d\lambda$$

définit une mesure de probabilité.

Exemple. Si $(\Omega, \mathcal{F}, \mathbb{P}) = ([0, 1], \mathcal{B}([0, 1]), \lambda_1)$, et si l'on pose

$$\begin{aligned} Y : ([0, 1], \mathcal{B}([0, 1])) &\longrightarrow ([-1, 0], \mathcal{B}([-1, 0])) \\ \omega &\longmapsto \omega^2 - 1, \end{aligned}$$

alors, pour tout $a, b \in \mathbb{R}$, avec $a < b$,

$$\begin{aligned} \mathbb{P}_Y([a, b]) &= \mathbb{P}(Y \in [a, b]) \\ &= \mathbb{P}(\{\omega \in [0, 1] \mid \omega^2 - 1 \in [a, b]\}) \\ &= \mathbb{P}(\{\omega \in [0, 1] \mid \omega \in [\sqrt{a+1}, \sqrt{b+1}]\}) \\ &= \sqrt{b+1} - \sqrt{a+1} \\ &= \int_a^b \frac{1}{2\sqrt{x+1}} dx. \end{aligned}$$

On trouve donc la densité $f_Y : x \mapsto \frac{1}{2\sqrt{x+1}}$.

Proposition 2.2.4. Soit X une variable aléatoire réelle de densité f_X , et de fonction de répartition F_X . Alors, on a les résultats suivants.

(i) Pour tout $x \in \mathbb{R}$,

$$F_X(x) = \int_{-\infty}^x f_X(t) dt,$$

(ii) F_X est continue sur \mathbb{R} .

(iii) Si f_X est continue en un point $x_0 \in \mathbb{R}$, alors F_X y est dérivable, et $F'_X(x_0) = f_X(x_0)$.

(iv) Inversement, si X a pour fonction de répartition F_X définie pour tout $x \in \mathbb{R}$ par

$$F_X(x) = \int_{-\infty}^x f(t) dt$$

avec f une fonction mesurable positive, alors X a pour densité f .

Remarque. Si X est une variable aléatoire réelle telle que F_X soit continue, alors X est nécessairement à densité. C'est le si F_X est absolument continue, i.e si il une fonction f intégrable telle que $F_X(b) - F_X(a) = \int_a^b f(t) dt$. Une fonction continue, de dérivée L^1 n'est pas forcément l'intégrale de sa dérivée (comme l'escalier de CANTOR).

Définition 2.2.4. Une fonction $f : \mathbb{R}^d \rightarrow \mathbb{R}$ mesurable est appelée **densité** si elle est positive et si

$$\int_{\mathbb{R}^d} f d\lambda_d = 1.$$

Un vecteur aléatoire $X = (X_1, \dots, X_d)$ a pour loi la loi densité f si pour tout $(a_i)_{1 \leq i \leq d}$ et $(b_i)_{1 \leq i \leq d} \in \mathbb{R}^d$ tels que $a_i \leq b_i$ pour tout $i \in \llbracket 1, d \rrbracket$,

$$\mathbb{P} \left(X \in \prod_{i=1}^d [a_i, b_i] \right) = \int_{\prod_{i=1}^d [a_i, b_i]} f d\lambda_d,$$

où λ_d désigne la mesure de LEBESGUE sur \mathbb{R}^d .

Proposition 2.2.5. Soit X un vecteur aléatoire à valeurs dans \mathbb{R}^d , de densité f_X . Soit φ un difféomorphisme de \mathbb{R}^d . Alors, $Y = \varphi \circ X$ est un vecteur aléatoire de densité f_Y définie par

$$f_Y(u) = f_X(\varphi^{-1}(u)) J\varphi^{-1}(u),$$

pour tout $u \in \mathbb{R}^d$, où $J\varphi^{-1}(u)$ est la valeur absolue du déterminant de la jacobienne de φ^{-1} en u .

Démonstration. Soit B un évènement. Alors,

$$\begin{aligned} \mathbb{P}_Y(B) &= \mathbb{P}(Y \in B) \\ &= \mathbb{P}(\varphi(X) \in B) \\ &= \mathbb{P}(X \in \varphi^{-1}(B)) \\ &= \int_{\varphi^{-1}(B)} f_X(x) dx \\ &= \int_B f_X(\varphi^{-1}(u)) J\varphi^{-1}(u) du, \end{aligned}$$

et ce par le changement de variable $x = \varphi^{-1}(u)$. On peut alors trouver la densité f_Y dans la dernière intégrale. \square

Proposition 2.2.6. Soit (X, Y) un couple de variables aléatoires réelles, de densité $f_{(X,Y)}$ sur \mathbb{R}^2 . Alors, X et Y sont aussi à densité, données par

$$f_X(x) = \int_{\mathbb{R}} f_{(X,Y)}(x, y) dy \quad \text{et} \quad f_Y(y) = \int_{\mathbb{R}} f_{(X,Y)}(x, y) dx,$$

pour tout x et $y \in \mathbb{R}$. Ce sont les **densités marginales**.

Démonstration. Soit $A \in \mathcal{B}(\mathbb{R})$. Alors,

$$\begin{aligned} \mathbb{P}_X(A) &= \mathbb{P}(X \in A) \\ &= \mathbb{P}(X \in A, Y \in \mathbb{R}) \\ &= \int_{A \times \mathbb{R}} f_{(X,Y)} d\lambda_2 \\ &= \int_A \left(\int_{\mathbb{R}} f_{(X,Y)}(x, y) dy \right) dx, \end{aligned}$$

tous ces calculs et inversions étant autorisées par le théorème d'inversion de FUBINI-LEBESGUE. La démonstration pour Y est identique. □

Proposition 2.2.7. Soit (X, Y) un couple de variables aléatoires réelles de densité $f_{(X,Y)}$. Alors, X et Y sont indépendantes si, et seulement si, , pour tout $(x, y) \in \mathbb{R}^2$,

$$f_{(X,Y)}(x, y) = f_X(x)f_Y(y).$$

Remarque. Cette propriété se généralise aux vecteurs aléatoires avec une indépendance mutuelle.

2.3 Variables aléatoires usuelles

Soit X une variable aléatoire réelle.

2.3.1 Variables aléatoires discrètes

Loi de BERNOULLI

On dit que X suit une loi de BERNOULLI de paramètre $p \in [0, 1]$ si $X(\Omega) = \{0, 1\}$, et

$$\mathbb{P}(X = 1) = p.$$

On note $X \sim \mathcal{B}(p)$.

Loi binomiale

On dit que X suit la loi binomiale de paramètre $p \in [0, 1]$ et $n \in \mathbb{N}^*$ si $X(\Omega) = \llbracket 0, n \rrbracket$, et

$$\mathbb{P}(X = k) = \binom{n}{k} p^k (1 - p)^{n-k},$$

pour tout $k \in \llbracket 0, n \rrbracket$. On note $X \sim \mathcal{B}(n, p)$.

Loi géométrique

On dit que X suit une loi géométrique de paramètre $p \in [0, 1]$ si $X(\Omega) = \mathbb{N}^*$ et

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

pour tout $k \in \mathbb{N}^*$. On note $X \sim \mathcal{G}(p)$.

Loi de POISSON

On dit que X suit une loi de POISSON de paramètre $\lambda > 0$ si $X(\Omega) = \mathbb{N}$ et

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

pour tout $k \in \mathbb{N}$. On note $X \sim \mathcal{P}(\lambda)$.

Loi uniforme sur un univers fini

On dit que X suit une loi uniforme sur $X(\Omega) = \{x_1, \dots, x_n\}$ si

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

pour tout $i \in \llbracket 1, n \rrbracket$. On a alors

$$\mathbb{P}(X \in A) = \frac{\text{card } A}{\text{card } \Omega},$$

2.3.2 Variables aléatoires à densité

Loi uniforme sur un segment de \mathbb{R}

Soit $a, b \in \mathbb{R}$ avec $a < b$. On dit que X est uniforme sur le segment $[a, b]$ si \mathbb{P}_X admet comme densité

$$x \mapsto f_X(x) = \frac{1}{b - a} \mathbb{1}_{[a, b]}(x).$$

On note $X \sim \mathcal{U}_{[a, b]}$.

Loi exponentielle

On dit que X suit une loi exponentielle de paramètre $\lambda > 0$ si $X(\Omega) = \mathbb{R}_+$ et X admet pour densité

$$x \mapsto f_X(x) = \lambda e^{-\lambda x} \mathbb{1}_{\mathbb{R}_+}.$$

On note $X \sim \mathcal{E}(\lambda)$.

Loi normale

On dit que X suit la loi normale (ou gaussienne) de paramètre $m \in \mathbb{R}$ et $\sigma \geq 0$, si $X(\Omega) \subset \mathbb{R}$ et X admet pour densité

$$x \mapsto f_X(x) = \frac{e^{-\frac{(x-m)^2}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}}.$$

On note $X \sim \mathcal{N}(m, \sigma^2)$.

Loi Gamma

Soit $n \in \mathbb{N}^*$, et $\lambda > 0$. On dit que X suit une loi $\Gamma(n, \lambda)$ si $X(\Omega) \subset \mathbb{R}_+$, et si X admet pour densité

$$x \mapsto f_X(x) = \frac{\lambda^n}{\Gamma(n)} e^{-\lambda x} x^{n-1} \mathbf{1}_{\mathbb{R}_+}.$$

Loi de CAUCHY

On dit que X suit la loi de CAUCHY de paramètre $\lambda > 0$ si $X(\Omega) = \mathbb{R}$ et admet pour densité

$$x \mapsto f_X(x) = \frac{1}{\pi} \frac{\lambda}{\lambda^2 + x^2}.$$

On note $X \sim \mathcal{C}(\lambda)$.

Chapitre 3

Espérance et moments

3.1 Espérance et moments

Commençons par rappeler la formule de transfert.

Définition 3.1.1 (mesure image). Soient (X, \mathcal{F}) et (Y, \mathcal{G}) des espaces mesurables, $\varphi : (X, \mathcal{F}) \rightarrow (Y, \mathcal{G})$ une application mesurable, et μ une mesure sur (X, \mathcal{F}) . Alors, la mesure image par φ est définie par

$$\nu(B) = \mu(\varphi^{-1}(B)),$$

pour tout $B \in \mathcal{G}$.

Proposition 3.1.1 (formule de transfert). Avec la même quantification, et si $h : (Y, \mathcal{G}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ est une fonction mesurable, alors, h est ν -intégrable si, et seulement si, $h \circ \varphi$ est μ -intégrable. Dans ce cas,

$$\int_X h \circ \varphi \, d\mu = \int_Y h \, d\nu.$$

3.1.1 Espérance : définitions et premières propriétés

Définition 3.1.2. Soit $X : (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ une variable aléatoire réelle. On dit que X est **intégrable** (ou \mathbb{P} -intégrable) si

$$\int_{\Omega} |X| \, d\mathbb{P} = \int_{\mathbb{R}} |x| \, d\mathbb{P}_X(x) < \infty.$$

Définition 3.1.3 (espérance). Soit $X : (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ une variable aléatoire réelle positive ou \mathcal{P} -intégrable. On appelle espérance de X , notée $\mathbb{E}(X)$, la quantité

$$\mathbb{E}(X) = \int_{\Omega} X \, d\mathbb{P} = \int_{\mathbb{R}} x \, d\mathbb{P}_X(x).$$

Plus généralement, si h est une fonction mesurable, telle que $h(X)$ est positive ou \mathbb{P} -intégrable, on a

$$\mathbb{E}(h(X)) = \int_{\Omega} h(X(\omega)) d\mathbb{P}(\omega) = \int_{\mathbb{R}} h(x) d\mathbb{P}_X(x).$$

Enfin, si $X = (X_1, \dots, X_n)$ est un vecteur aléatoire, et $h : \mathbb{R}^n \rightarrow \mathbb{R}$ est une fonction mesurable, on note lorsque cela est bien défini,

$$\mathbb{E}(h(X_1, \dots, X_n)) = \int_{\mathbb{R}} h(x) d\mathbb{P}_X(x).$$

Remarque. Les intégrales ci-dessus sont à considérer au sens de LEBESGUE¹. Par exemples,

- Si $X = \mathbb{1}_A$, $\mathbb{E}(X) = \mathbb{P}(A)$.
- Si $X = \sum_{i=1}^n a_i \mathbb{1}_{A_i}$, alors $\mathbb{E}[X] = \sum_{i=1}^n a_i \mathbb{P}(A_i)$, où les A_i sont des ensembles mesurables, et les a_i des éléments de \mathbb{R} .
- Si $X \geq 0$,

$$\mathbb{E}[X] = \sup\{\mathbb{E}(Y) \mid Y \text{ une fonction étagée bornée.}\}.$$

- Si $\mathbb{E}[|X|] < \infty$, alors

$$\mathbb{E}[X] = \mathbb{E}[X^+] - \mathbb{E}[X^-].$$

- Concrètement, si X est discrète, et en notant $X(\Omega) = \{x_1, \dots, x_n, \dots\}$ (au plus dénombrable), alors

$$\mathbb{E}[X] = \sum_{i=1}^{\infty} x_i \mathbb{P}(X = x_i).$$

De même, pour une fonction $h : \mathbb{R} \rightarrow \mathbb{R}$ mesurable,

$$\mathbb{E}[h(X)] = \sum_{i=1}^{\infty} h(x_i) \mathbb{P}(X = x_i).$$

- Si X est à densité f_X , alors,

$$\mathbb{E}[X] = \int_{\mathbb{R}} x f_X(x) dx \quad \text{et} \quad \mathbb{E}[h(X)] = \int_{\mathbb{R}} h(x) f_X(x) dx.$$

Définition 3.1.4 (espérance de vecteur aléatoire et variable aléatoire centrée). Si $X = (X_1, \dots, X_n)$ est un vecteur aléatoire, avec $\mathbb{E}[|X_i|] < \infty$ pour tout $i \in \llbracket 1, n \rrbracket$, alors, on définit

$$\mathbb{E}[X] = (\mathbb{E}[X_1], \dots, \mathbb{E}[X_n]).$$

On dit que X est centré si $\mathbb{E}[X] = 0_{\mathbb{R}^n}$.

1. Sinon c'est complexe.

Proposition 3.1.2. Soient X et Y sont deux variables aléatoires réelles, positives ou \mathcal{P} -intégrables. On a les résultats suivants.

- (i) Si $X \leq Y$ presque sûrement, alors $\mathbb{E}[X] \leq \mathbb{E}[Y]$.
- (ii) Pour tout $\lambda \in \mathbb{R}$, $\mathbb{E}[X + \lambda Y] = \mathbb{E}[X] + \lambda \mathbb{E}[Y]$.
- (iii) $|\mathbb{E}[X]| \leq \mathbb{E}[|X|]$.
- (iv) Si X est positive presque sûrement, et est d'espérance nulle, alors $X = 0$ presque partout.

Démonstration. Ces propriétés sont des redites de cours d'intégration de LEBESGUE. Voir le cours d'INTL. □

Remarque. L'espérance n'est autre qu'une intégrale par rapport à la mesure \mathbb{P} . Il va donc de soit que les théorèmes classiques de la théorie de l'intégrale de LEBESGUE. En voici quelques exemples.

- *Convergence monotone.* Si $(X_n)_n$ est une suite croissante de variables aléatoires positives, convergeant vers une variable aléatoire X , alors

$$\mathbb{E} \left[\lim_{n \rightarrow \infty} X_n \right] = \lim_{n \rightarrow \infty} \mathbb{E}[X_n].$$

- *Lemme de FATOU.* Si

Si $(X_n)_n$ est une suite de variables aléatoires positives, alors,

$$\mathbb{E} \left[\liminf_{n \rightarrow \infty} X_n \right] \leq \liminf_{n \rightarrow \infty} \mathbb{E}[X_n].$$

- *Théorème de convergence dominée.* Si $(X_n)_n$ est une suite de variables aléatoires convergeant vers une variable aléatoire X , et qu'il existe une fonction Y mesurable et intégrable telle que

$$|X_n(\omega)| \leq |Y(\omega)|$$

pour tout $n \in \mathbb{N}$ et tout $\omega \in \Omega$, alors,

$$\lim_{n \rightarrow \infty} \mathbb{E}[X_n] = \mathbb{E}[X].$$

Proposition 3.1.3 (inégalité de MARKOV). Soit X une variable aléatoire. Alors, pour tout $t > 0$,

$$\mathbb{P}(|X| > t) \leq \frac{\mathbb{E}[|X|]}{t}.$$

Démonstration. On a

$$\mathbb{E}[|X|] = \underbrace{\mathbb{E} \left[|X| \mathbf{1}_{|X| > t} \right]}_{\substack{t \mathbb{E}[\mathbf{1}_{|X| > t}] \\ = t \mathbb{P}(|X| > t)}} + \underbrace{\mathbb{E} \left[|X| \mathbf{1}_{|X| \leq t} \right]}_{\geq 0}.$$

□

Variante. Il a en fait plusieurs inégalités de MARKOV, mais dont l'idée reste semblable. Soit $\lambda > 0$ tel que $\mathbb{E}[e^{\lambda|X|}] < \infty$. Alors,

$$\mathbb{P}(|X| > t) = \mathbb{P}(e^{\lambda|X|} > e^{\lambda t}) \leq \frac{\mathbb{E}[e^{\lambda|X|}]}{e^{\lambda t}},$$

pour tout $t \in \mathbb{R}$.

Proposition 3.1.4 (inégalité de JENSEN). *Soit X une variable aléatoire réelle, et φ une fonction convexe, et tels que X et $\varphi(X)$ soient intégrables. Alors,*

$$\mathbb{E}[\varphi(X)] \geq \varphi(\mathbb{E}[X]).$$

Démonstration. On observe que $\varphi(x) = \sup\{f(x) \mid f \text{ affine}, f \leq \varphi\}$. On conclut alors par positivité et linéarité de l'espérance. □

Exemple. $\mathbb{E}[X^2] \leq \mathbb{E}[X]^2$.

Proposition 3.1.5. *Soit X une variable aléatoire **positive**. Alors,*

$$\mathbb{E}[X] = \int_0^\infty \mathbb{P}(X > t) dt = \int_0^\infty (1 - F_X(t)) dt.$$

En particulier, si $X(\Omega) \subset \mathbb{N}$, alors

$$\mathbb{E}[X] = \sum_{k=0}^\infty \mathbb{P}(X > k).$$

Démonstration. D'après FUBINI-TONELLI,

$$\begin{aligned} \int_0^\infty \mathbb{P}(X > t) dt &= \int_0^\infty \mathbb{E}[\mathbf{1}_{X>t}] dt \\ &= \mathbb{E}\left[\int_0^\infty \mathbf{1}_{X>t} dt\right] \\ &= \mathbb{E}\left[\int_0^X dt\right] = \mathbb{E}[X]. \end{aligned}$$

□

3.1.2 Moments d'ordre supérieurs

Définition 3.1.5 (moment). *Soit $p \in \mathbb{N}^*$. On dit qu'une variable aléatoire réelle admet un moment d'ordre p si*

$$\mathbb{E}[|X|^p] = \int_\Omega |X|^p d\mathbb{P} = \int_{\mathbb{R}} |x|^p d\mathbb{P}_X(x) < \infty.$$

Dans ce cas, on note $\|X\|_p = \mathbb{E}[|X|^p]^{1/p}$.

Notation. On note $\mathbb{L}^p(\Omega, \mathcal{F}, \mathbb{P})$ l'espace des variables aléatoires réelles admettant un moment d'ordre p , quotienté par la relation d'équivalence \sim "être égales presque partout". Enfin, on note

$$\mathbb{L}^\infty(\Omega, \mathcal{F}, \mathbb{P}) = \{X \text{ v.a.r} \mid \exists c > 0, \mathbb{P}(|X| > c) = 0\} / \sim .$$

Les inégalités classiques de la théorie de l'intégration s'étendent dans ce cadre probabiliste.

Proposition 3.1.6 (HÖLDER, CAUCHY-SCHWARTZ). Soient $p, q \geq 1$ des réels conjugués. Si $X \in \mathbb{L}^p(\Omega, \mathcal{F}, \mathbb{P})$ et $Y \in \mathbb{L}^q(\Omega, \mathcal{F}, \mathbb{P})$, alors

$$\|XY\|_1 = \mathbb{E}[|XY|] \leq \mathbb{E}[|X|^p]^{1/p} \cdot \mathbb{E}[|Y|^q]^{1/q}.$$

En particulier, pour $p = q = 2$,

$$\mathbb{E}[|XY|]^2 \leq \mathbb{E}[|X|^2] \cdot \mathbb{E}[|Y|^2].$$

Proposition 3.1.7 (MINKOWSKI). Soit $p \geq 1$, et $X, Y \in \mathbb{L}^p(\Omega, \mathcal{F}, \mathbb{P})$. Alors,

$$\|X + Y\|_p \leq \|X\|_p + \|Y\|_p.$$

Définition 3.1.6 (variance, écart-type). Soit $X \in \mathbb{L}^2(\Omega, \mathcal{F}, \mathbb{P})$. On appelle **variance** de X la quantité

$$\text{var}(X) = \mathbb{E}[|X - \mathbb{E}[X]|^2] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 \geq 0.$$

C'est l'écart quadratique à la moyenne. On appelle **écart-type** la quantité

$$\sqrt{\text{var}(X)} = \|X - \mathbb{E}[X]\|_2.$$

Enfin, si X et $Y \in \mathbb{L}^2(\Omega, \mathcal{F}, \mathbb{P})$, on définit la **covariance** de X et Y par

$$\text{cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] \in \mathbb{R}.$$

C'est la version polarisée de la variance.

Proposition 3.1.8. Pour tout $X \in \mathbb{L}^2(\Omega, \mathcal{F}, \mathbb{P})$, on a les propriétés suivantes.

- (i) $\text{var}(X) \geq 0$.
- (ii) Pour tout $\lambda \in \mathbb{R}$, $\text{var}(\lambda X) = \lambda^2 \text{var}(X)$.
- (iii) Pour tout $\alpha \in \mathbb{R}$, $\text{var}(X + \alpha) = \text{var}(X)$.
- (iv) Si $\text{var}(X) = 0$, alors X est constante presque partout. En particulier, X est constante à $\mathbb{E}[X]$ presque partout.
- (v) Si $Y \in \mathbb{L}^2(\Omega, \mathcal{F}, \mathbb{P})$, alors

$$\text{var}(X + Y) = \text{var}(X) + \text{var}(Y) + 2 \text{cov}(X, Y).$$

Proposition 3.1.9 (inégalité de BIENAYMÉ-TCHEBYTCHEV). Soit $X \in \mathbb{L}^2(\Omega, \mathcal{F}, \mathbb{P})$, et $t > 0$. Alors,

$$\mathbb{P}(|X - \mathbb{E}[X]| > t) \leq \frac{\text{var}(X)}{t^2}.$$

Démonstration. On se ramène à l'inégalité de MARKOV :

$$\mathbb{P}(|X - \mathbb{E}[X]| > t) = \mathbb{P}(|X - \mathbb{E}[X]|^2 > t^2) \leq \frac{\mathbb{E}[|X - \mathbb{E}[X]|^2]}{t^2} = \frac{\text{var}(X)}{t^2}.$$

□

3.1.3 Moments de variables usuelles

Loi de BERNOULLI

Si $X \sim \mathcal{B}(p)$, alors,

- $\mathbb{E}[X] = p$,
- $\mathbb{E}[X^2] = p$,
- $\text{var}(X) = p(1 - p)$.

Loi uniforme sur un univers fini

Si $X \sim \mathcal{U}_{\{x_1, \dots, x_n\}}$, alors,

$$\mathbb{E}(X) = \frac{1}{n} \sum_{i=1}^n x_i.$$

Loi binomiale

Si $X \sim \mathcal{B}(n, p)$, alors,

$$\mathbb{E}[X] = np \quad \text{et} \quad \text{var}(X) = np(1 - p).$$

Loi géométrique

Si $X \sim \mathcal{G}(p)$, alors,

$$\mathbb{E}[X] = \frac{1}{p} \quad \text{et} \quad \text{var}(X) = \frac{1 - p}{p^2}.$$

Loi de POISSON

Si $X \sim \mathcal{P}(\lambda)$, alors,

$$\mathbb{E}[X] = \text{var}(X) = \lambda.$$

Loi uniforme sur un segment de \mathbb{R}

Si $X \sim \mathcal{E}(\lambda)$.

Loi exponentielle

Loi normale

Loi de CAUCHY

3.2 Espérance et identification de loi

3.2.1 Identification et fonctions tests

Théorème 3.2.1. Soit X un vecteur aléatoire à valeurs dans \mathbb{R}^d et μ une mesure de probabilité sur \mathbb{R}^d . Alors, X suit la loi μ ($X \sim \mu$) si, et seulement si, pour toute fonction $h : \mathbb{R}^d \rightarrow \mathbb{R}$ continue à support compact (ou de classe \mathcal{C}^∞),

$$\mathbb{E}[h(X)] = \int_{\mathbb{R}^d} h(x) d\mu(x).$$

Exemple. Soit X une v.a.r suivant une loi de CAUCHY de paramètre 1. On a

$$f_X(x) = \frac{1}{\pi(1+x^2)},$$

pour tout $x \in \mathbb{R}$. Soit $Y = X^+ = X\mathbf{1}_{X \geq 0}$. On pose de même $x^+ = x\mathbf{1}_{\mathbb{R}_+}(x)$ pour tout $x \in \mathbb{R}$. Pour toute fonction h continue à support compact,

$$\begin{aligned} \mathbb{E}[h(Y)] &= \int_{\mathbb{R}_+} h(x^+) d\mathbb{P}_X(x) \\ &= \int_{\mathbb{R}_+} h(x^+) f_X(x) dx \\ &= h(0)\mathbb{P}(X \leq 0) + \int_{\mathbb{R}_+} \frac{h(x)}{\pi(1+x^2)} dx. \end{aligned}$$

On en déduit que $d\mathbb{P}_Y(x) = \frac{1}{2}\delta_0(x) + \frac{1}{\pi(1+x^2)}\mathbf{1}_{x>0} dx$.

3.2.2 Espérance et indépendance

Proposition 3.2.1. Soit $X = (X_1, \dots, X_n)$ un vecteur aléatoire à valeurs dans \mathbb{R}^n . Sont équivalents.

- (i) Les variables aléatoires sont mutuellement indépendantes.
- (ii) Pour toutes fonctions mesurables bornées (ou positives) $h_i : \mathbb{R} \rightarrow \mathbb{R}$ ($i \in \llbracket 1, n \rrbracket$),

$$\mathbb{E}[h_1(X_1), \dots, h_n(X_n)] = \prod_{i=1}^n \mathbb{E}[h_i(X_i)]$$

Corollaire 3.2.1. Soient X, Y des variables aléatoires réelles indépendantes admettant des moments d'ordre 2. Alors, $\text{cov}(X, Y) = 0$ et $\text{var}(X + Y) = \text{var}(X) + \text{var}(Y)$.

Attention. La réciproque est *fausse*. Si par exemple U est une variable aléatoire de loi uniforme sur $] -1, 1[$ et $V = U^2$, alors un calcul montre que $\text{cov}(U, V) = 0$. Pourtant, U et V ne sont pas indépendantes. En effet,

$$\mathbb{P}(|U| < 1/2, V > 1/2) = 0 \quad \text{mais} \quad \mathbb{P}(|U| > 1/2)\mathbb{P}(V > 1/2) > 0.$$

Démonstration. Si $X \perp Y$, alors

$$\text{cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \mathbb{E}[X - \mathbb{E}[X]] \cdot \mathbb{E}[Y - \mathbb{E}[Y]] = 0.$$

□

3.2.3 Le problème de moments

Le problème de moments consiste à se demander si la donnée des $\mathbb{E}[X^p]$ pour p parcourant \mathbb{N} permet d'identifier \mathbb{P}_X . En fait, on sous-entend par là deux questions :

- *Existence.* Étant donné une suite $(m_p)_{p \in \mathbb{N}}$, existe-t-il une variable aléatoire réelle X telle que $\mathbb{E}[X^p] = m_p$ pour tout $p \geq 0$?
- Si X et Y sont deux variables aléatoires telles que $\mathbb{E}[X^p] = \mathbb{E}[Y^p]$, a-t-on $\mathbb{P}_X = \mathbb{P}_Y$.

Nous verrons que la réponse à ces questions dépend de du support de la loi cible.

Définition 3.2.1. Soit $(m_p)_p \in \mathbb{R}^{\mathbb{N}}$. On dit que $(m_p)_p$ est complètement monotone si :

$$\forall k \geq 0, \forall p \geq 0, (-1)^k (\Delta^k m)_p \geq 0.$$

Ici, $(\Delta m)_p = m_{p+1} - m_p$, $(\Delta^2 m)_p = (m_{p+2} - m_{p+1}) - (m_{p+1} - m_p) = m_{p+2} - 2m_{p+1} + m_p$.

Théorème 3.2.2 (des moments de HAUSDORFF). Soit $(m_p)_p \in \mathbb{R}^{\mathbb{N}}$. Cette suite est une suite des moments d'une mesure μ à support dans $[0, 1]$, i.e

$$m_p = \int_0^1 x^p d\mu(x)$$

pour tout $p \in \mathbb{N}$ si, et seulement si, la suite $(m_p)_p$ est complètement monotone.

Démonstration. Le caractère nécessaire de la complète monotonie est clair si l'on remarque que

$$(-1)^k (\Delta^k m)_p = \int_0^1 x^p (1-x)^k d\mu(x) \geq 0.$$

On admettra l'autre implication.

□

Théorème 3.2.3. Une suite $(m_p)_p \in \mathbb{R}^{\mathbb{N}}$ est la suite des moments d'une mesure à support non borné dans \mathbb{R} si, et seulement si, pour toute suite non nulle $(a_j)_j \in \mathbb{C}^{\mathbb{N}}$ à support fini,

$$\sum_{j,k \in \mathbb{N}} m_{j+k} a_j \bar{a}_k > 0$$

Démonstration. On admet que la condition est suffisante. Pour la réciproque, on a

$$\sum_{j,k \in \mathbb{N}} m_{j+k} a_j \bar{a}_k = \int_{\mathbb{R}} \left| \sum_{j=0}^{\infty} a_j x^j \right|^2 d\mu(x) \geq 0.$$

□

Remarque. Le support d'une mesure borélienne est l'intersection des fermés dont le complémentaire est de mesure nulle.

Proposition 3.2.2. Si μ est une mesure borélienne sur $[0, 1]$ à support compact, alors μ est caractérisé par la suite de ses moments,

$$\left(\int_0^1 x^p d\mu(x) \right)_{p \in \mathbb{N}}.$$

Remarque. Sans l'hypothèse de compacité du support, on perd l'unicité. Voir par exemple l'exercice 7 de la feuille du TD 5.

Démonstration. On considère deux mesures μ et ν telles que pour tout $p \in \mathbb{N}$,

$$\int_0^1 x^p d\mu(x) = \int_0^1 x^p d\nu(x).$$

On a donc par linéarité que pour tout $Q \in \mathbb{R}[X]$,

$$\int_0^1 Q(x) d\mu(x) = \int_0^1 Q(x) d\nu(x).$$

Maintenant, soit $f : [0, 1] \rightarrow \mathbb{R}$ continue. Par le théorème de STONE-WEIERSTRASS, pour tout $\varepsilon > 0$, il existe $Q_\varepsilon \in \mathbb{R}[X]$ tel que $\|f - Q_\varepsilon\|_{\infty, [0,1]} \leq \varepsilon$. On a alors

$$\begin{aligned} \int_0^1 f(x) d\mu(x) - \int_0^1 f(x) d\nu(x) &= \int_0^1 (f(x) - Q_\varepsilon(x)) d\mu(x) \\ &\quad - \int_0^1 (f(x) - Q_\varepsilon(x)) d\nu(x) \\ &\quad + \underbrace{\int_0^1 Q_\varepsilon(x) d\mu(x) - \int_0^1 Q_\varepsilon(x) d\nu(x)}_{=0}. \end{aligned}$$

Enfin,

$$\begin{aligned}
 \left| \int_0^1 f(x) d\mu(x) - \int_0^1 f(x) d\nu(x) \right| &\leq \left| \int_0^1 (f(x) - Q_\varepsilon(x)) d\mu(x) \right| \\
 &\quad + \left| \int_0^1 (f(x) - Q_\varepsilon(x)) d\nu(x) \right| \\
 &\leq \int_0^1 \|f - Q_\varepsilon\|_{\infty, [0,1]} d\mu(x) + \int_0^1 \|f - Q_\varepsilon\|_{\infty, [0,1]} d\nu(x). \\
 &\leq \varepsilon(\mu([0, 1]) + \nu([0, 1])) \leq 2\varepsilon.
 \end{aligned}$$

Ainsi, $\int_0^1 f(x) d\mu(x) = \int_0^1 f(x) d\nu(x)$ pour tout $f \in \mathcal{C}([0, 1], \mathbb{R})$. On en déduit que $\mu = \nu$. □

3.3 Transformées exponentielles

3.3.1 Fonction caractéristique

Définition 3.3.1. Soit $X = (X_1, \dots, X_d)$ un vecteur aléatoire. On définit sa fonction caractéristique $\varphi_X : \mathbb{R}^d \rightarrow \mathbb{C}$ par

$$\varphi_X(t) = \mathbb{E} [e^{itX}] = \mathbb{E} \left[e^{i \sum_{k=1}^d t_k X_k} \right],$$

pour tout $t = (t_1, \dots, t_k) \in \mathbb{R}^d$.

Remarques.

- φ_X est toujours bien définie car $|e^{itX}| \leq 1$ et $e^{itX} \in \mathbb{L}^1(\Omega, \mathcal{F}, \mathbb{P})$.
- Pour tout $t \in \mathbb{R}^d$, on a

$$\varphi_X(t) = \int_{\mathbb{R}} e^{itx} d\mathbb{P}_X(x).$$

C'est la transformée de FOURIER de \mathbb{P}_X .

Exemple. Soit $\lambda > 0$ et U une variable aléatoire suivant une loi de LAPLACE de paramètre λ , i.e que U a pour densité $f_U(x) = \frac{\lambda}{2} e^{-\lambda|x|}$. Pour tout $t \in \mathbb{R}$, on a

$$\mathbb{E}[e^{itU}] = \int_{\mathbb{R}} e^{itx} \frac{\lambda}{2} e^{-\lambda|x|} dx = \frac{\lambda^2}{\lambda^2 + t^2}.$$

On a alors $\varphi_U(\lambda t) = \frac{1}{1+t^2}$.

Théorème 3.3.1. La fonction caractérise la loi : si X et Y sont deux variables aléatoires réelles telles que $\varphi_X = \varphi_Y$, alors $\mathbb{P}_X = \mathbb{P}_Y$.

Proposition 3.3.1. Soit X un vecteur aléatoire et φ_X sa fonction caractéristique. On a les propriétés suivantes.

- (i) Pour tout $t \in \mathbb{R}^d$, $|\varphi_X(t)| \leq 1$.
- (ii) Pour tout $t \in \mathbb{R}^d$, $\varphi_X(-t) = \overline{\varphi_X(t)}$.
- (iii) $\varphi_X(0) = 1$.
- (iv) φ_X est uniformément continue.
- (v) φ_X est de type positif.

Théorème 3.3.2 (BOCHNER-HERGLOTZ). Soit $\varphi : \mathbb{R}^d \rightarrow \mathbb{C}$ continue en 0 avec $\varphi(0) = 1$ et de type positif. Alors, φ est la fonction caractéristique d'un vecteur aléatoire.

Théorème 3.3.3 (inversion de FOURIER). Soit X un vecteur aléatoire à valeurs dans \mathbb{R}^d de fonction caractéristique φ_X . On suppose que φ_X est intégrable sur \mathbb{R}^d . Alors, X admet une densité f_X continue et bornée définie par

$$f_X(x) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} e^{-itx} \varphi_X(t) dt,$$

pour tout $x \in \mathbb{R}^d$.

Théorème 3.3.4. Soit $X = (X_1, \dots, X_n)$ un vecteur aléatoire. Alors, les X_i sont mutuellement indépendantes si, et seulement si, pour tout $t = (t_1, \dots, t_n) \in \mathbb{R}^n$,

$$\varphi_X(t) = \prod_{k=1}^n \varphi_{X_k}(t_k).$$

Proposition 3.3.2. Si X et Y sont des variables aléatoires indépendantes réelles, alors pour tout $t \in \mathbb{R}$,

$$\varphi_{X+Y}(t) = \varphi_X(t)\varphi_Y(t).$$

Remarque. On généralise bien évidemment le résultat à une somme de n variables aléatoires indépendantes de même loi.

Proposition 3.3.3. Soit X est une variable aléatoire réelle de fonction caractéristique φ_X et $p \in \mathbb{N}^*$. On a les résultats suivants.

- (i) Si $\mathbb{E}[|X|^p] < \infty$, alors φ_X est de classe \mathcal{C}^p , et

$$\varphi_X^{(k)}(0) = i^k \mathbb{E}[X^k e^{itX}]$$

pour tout $k \in \llbracket 1, p \rrbracket$.

(ii) Réciproquement, si φ_X est p -fois dérivable en zéro, alors $\mathbb{E}[|X|^k] < \infty$ pour tout $k \in \llbracket 1, 2\lfloor p/2 \rfloor \rrbracket$.

Exemple. Si $X \sim \mathcal{E}(\lambda)$, $f_X(x) = \lambda e^{-\lambda x}$ pour tout $x \in \mathbb{R}_+$. On a pour tout $t \in \mathbb{R}$,

$$\varphi_X(t) = \int_0^\infty e^{itx} \lambda e^{-\lambda x} dx = \frac{\lambda}{\lambda - it}.$$

Ainsi, $\varphi'_X(t) = \frac{\lambda i}{(\lambda - it)^2}$, donc $\mathbb{E}[X] = \frac{1}{\lambda}$. De même, $\varphi''_X(t) = \frac{-2\lambda}{(\lambda - it)^3}$, donc $\mathbb{E}[X^2] = \frac{1}{\lambda^2}$. On retrouve que $\text{var}(X) = \frac{1}{\lambda^2}$.

3.3.2 Autres transformées exponentielles

Définition 3.3.2 (fonction génératrice). Soit X une variable aléatoire à valeurs dans \mathbb{N} . On appelle fonction génératrice de X la fonction $G_X : [0, 1] \rightarrow \mathbb{R}$, $s \mapsto \mathbb{E}[s^X]$. Elle est telle que pour tout $s \in [0, 1]$,

$$G_X(s) = \sum_{k=0}^{\infty} \mathbb{P}(X = k) s^k.$$

Plus généralement, si $X = (X_1, \dots, X_d)$ est un vecteur aléatoire à valeurs dans \mathbb{N}^d . On pose

$$G_X(s_1, \dots, s_d) = \mathbb{E} \left[s_1^{X_1} \cdots s_d^{X_d} \right].$$

Remarques.

- On a le lien “formel” $G_X(e^{it}) = \varphi_X(t)$.
- Le rayon de convergence de la série dont G_X est la somme a un rayon de convergence d’au moins 1 puisque $G_X(1) = 1$. On en déduit alors que G_X est de classe \mathcal{C}^∞ sur $[0, 1[$, et pour tout $n \in \mathbb{N}$,

$$G_X^{(n)}(0) = n! \mathbb{P}(X = n).$$

La fonction génératrice caractérise la loi.

- Soit $n \in \mathbb{N}^*$. Par le théorème de dérivation sous le signe somme, G_X est n -fois dérivable en 1 si, et seulement si, $\mathbb{E}[X^n] < \infty$ et

$$G_X^{(n)}(1) = \mathbb{E}[X(X-1) \cdots (X-n+1)].$$

C’est le moment factoriel.

Exemples.

- Si $X \sim \mathcal{B}(p)$, alors $G_X(s) = 1 - p + ps$ pour tout $s \in [0, 1]$.
- Si $X \sim \mathcal{G}(p)$, $G_X(s) = \frac{ps}{1 - (1-p)s}$ pour tout $s \in [0, 1]$.
- Si $X \sim \mathcal{P}(\lambda)$, $G_X(s) = e^{\lambda(s-1)}$ pour tout $s \in [0, 1]$.

Proposition 3.3.4. Soient X et Y des variables aléatoires à valeurs dans \mathbb{N} . Alors,

- (i) X et Y sont indépendantes si, et seulement si, pour tout $s, t \in [0, 1]$, $G_{(X,Y)}(s, t) = G_X(s)G_Y(t)$.
- (ii) Si X et Y sont indépendantes, alors $G_{X+Y} = G_X G_Y$.

Exemples.

- Si $X \sim \mathcal{B}(n, p)$, alors $G_X(s) = (1 - p + ps)^n$ pour tout $s \in [0, 1]$.
- Si $X \sim \mathcal{P}(\lambda)$ et $Y \sim \mathcal{P}(\mu)$ sont des v.a.r indépendantes, alors on voit que $G_{X+Y}(s) = e^{(\lambda+\mu)(s-1)}$ pour tout $s \in [0, 1]$. Puisque la fonction génératrice de $X + Y$ caractérise sa loi, on voit que $X + Y \sim \mathcal{P}(\lambda + \mu)$

Définition 3.3.3. Soit X un vecteur aléatoire à valeurs dans \mathbb{R}^d . On définit la transformée de LAPLACE de X par

$$L_X : \mathbb{R}^d \longrightarrow [0, \infty]$$

$$t \longmapsto \mathbb{E} [e^{(t,X)}]$$

Remarque.

- On a “ $\varphi_X(t) = L_X(it)$ ”.
- Contrairement à la fonction caractéristique ou la fonction génératrice, L_X peut prendre ∞ comme valeur.

Proposition 3.3.5. Soit X une variable aléatoire.

- (i) $L_X(0) = 1$.
- (ii) Si $L_X(t) < \infty$ pour tout t dans un voisinage de zéro, alors L_X y est analytique :

$$L_X(t) = \sum_{n=0}^{\infty} \frac{t^n}{n!} \mathbb{E}[X^n] \quad \text{et} \quad \mathbb{E}[X^n] = L_X^{(n)}(0) \text{ pour tout } n \in \mathbb{N}.$$

En particulier, $X \in \mathbb{L}^p(\Omega, \mathcal{F}, \mathbb{P})$ pour tout $p \geq 1$.

Remarque. Si X est une variable positive, on pourra considérer alternativement

$$\tilde{L}_X(t) = \mathbb{E} [e^{-tX}] \quad \text{pour tout } t \geq 0.$$

Proposition 3.3.6. Soient X et Y des variables aléatoires réelles.

- (i) X et Y sont indépendantes si, et seulement si, $L_{(X,Y)}(s, t) = L_X(s)L_Y(t)$ pour tout

$s, t \in \mathbb{R}$.

(ii) Si X et Y sont indépendantes, alors $L_{X+Y} = L_X L_Y$.

Exemples.

— Si $X \sim \mathcal{E}(\lambda)$, alors $f_X(x) = \lambda e^{-\lambda x} \mathbf{1}_{x \geq 0}$ avec $\lambda > 0$. Alors,

$$\tilde{L}_X(t) = \frac{\lambda}{\lambda + t},$$

pour tout $t \in \mathbb{R} \setminus \{-\lambda\}$.

— Si $Y \sim \Gamma(n, \lambda)$, alors $f_Y(x) = \frac{\lambda^n}{\Gamma(n)} x^{n-1} e^{-\lambda x} \mathbf{1}_{x \geq 0}$ avec $\lambda > 0$. On a alors,

$$\tilde{L}_Y(t) = \left(\frac{\lambda}{\lambda + t} \right)^n,$$

pour tout $t \in \mathbb{R} \setminus \{-\lambda\}$. Ainsi, si $X_1, \dots, X_n \sim \mathcal{E}(\lambda)$ sont des variables mutuellement indépendantes, alors $X_1 + \dots + X_n \sim \Gamma(n, \lambda)$.

3.4 Probabilités, lois et espérances conditionnelles

3.4.1 Probabilités conditionnelles

Définition 3.4.1. Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé. Soit $A \in \mathcal{F}$, tel que $\mathbb{P}(A) > 0$. On définit la probabilité conditionnelle “sachant A ” comme la probabilité définie pour $B \in \mathcal{F}$ par

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

Remarques.

- On doit bien évidemment vérifier que $B \mapsto \mathbb{P}(B|A)$ définit une probabilité.
- Si $B \in \mathcal{F}$ est un évènement indépendant de A , alors $\mathbb{P}(B|A) = \mathbb{P}(B)$. Cela est cohérent avec le sens que l’on veut donner à des probabilités conditionnelles.

Proposition 3.4.1 (formule des probabilités totales). Si $\Omega = \bigsqcup_{i \in \mathbb{N}} A_i$ est une partition au plus dénombrable d’évènements de Ω telle que $\mathbb{P}(A_i) > 0$ pour tout $i \in I$. Alors, pour tout $B \in \mathcal{F}$,

$$\mathbb{P}(B) = \sum_{i \in \mathbb{N}} \mathbb{P}(B|A_i) \mathbb{P}(A_i).$$

Démonstration. Écrire $B = B \cap \Omega = \bigsqcup_{i \in \mathbb{N}} (B \cap A_i)$ et utiliser la σ -additivité de la probabilité \mathbb{P} . □

Exemple. Si l'on écrit $\Omega = A \sqcup A^c$, alors pour tout $B \in \mathcal{F}$,

$$\mathbb{P}(B) = \mathbb{P}(B|A)\mathbb{P}(A) + \mathbb{P}(B|A^c)\mathbb{P}(A^c).$$

Proposition 3.4.2 (formule de BAYES). Si $\Omega = \bigsqcup_{i \in \mathbb{N}} A_i$ est une partition dénombrable d'évènements avec $\mathbb{P}(A_i) > 0$ pour tout $i \in \mathbb{N}$. Soit $B \in \mathcal{F}$ tel que $\mathbb{P}(B) > 0$. Alors, pour tout $i \in \mathbb{N}$,

$$\mathbb{P}(A_i|B) = \frac{\mathbb{P}(B|A_i)\mathbb{P}(A_i)}{\sum_{j \in \mathbb{N}} \mathbb{P}(B|A_j)\mathbb{P}(A_j)}.$$

C'est une inversion de conditionnement.

Démonstration. On écrit

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

formule à laquelle on applique la formule des probabilités totales (au dénominateur). □

3.4.2 Loi et espérance conditionnelle

Définition 3.4.2. Soit X une variable aléatoire discrète et $x_0 \in X(\Omega)$ tel que $\mathbb{P}(X = x_0) > 0$. On appelle loi conditionnelle sachant $\{X = x_0\}$ la loi définie par

$$\mathbb{P}^{X=x_0} = \mathbb{P}(B|X = x_0) \quad \text{pour tout } B \in \mathcal{F}.$$

On appelle espérance conditionnelle sachant $\{X = x_0\}$, noté $\mathbb{E}[\cdot|X = x_0]$, l'espérance sous la loi $\mathbb{P}^{X=x_0}$.

Exemple. On lance deux dés équilibrés à six faces. On note X et Y les résultats. Alors,

$$\begin{aligned} \mathbb{E}[X | X + Y = 8] &= \sum_{k=1}^6 k \mathbb{P}(X = k | X + Y = 8) \\ &= \sum_{k=1}^6 k \frac{\mathbb{P}(X = k)\mathbb{P}(Y = 8 - k)}{\mathbb{P}(X + Y = 8)} \\ &= \sum_{k=1}^6 k \frac{\mathbb{P}(X = k, X + Y = 8)}{\mathbb{P}(X + Y = 8)} \\ &= \sum_{k=2}^6 k \frac{\mathbb{P}(X = k, X + Y = 8)}{\mathbb{P}(X + Y = 8)} \\ &= \sum_{k=2}^6 k \frac{1/36}{5/36} = 4 \end{aligned}$$

En moyenne, un dé vaut 4. On aurait pu faire le calculer en écrivant que $\mathbb{E}[X | X + Y = 8] = \mathbb{E}[Y | X + Y = 8]$ et $\mathbb{E}[X + Y | X + Y = 8] = 8$.

Définition 3.4.3. Soit (X, Y) un couple de variables aléatoires, de densité $f_{(X,Y)}$. On définit la densité conditionnelle de Y sachant $\{X = x_0\}$ par

$$f_{Y|X=x_0}(y) = \frac{f_{(X,Y)}(x, y)}{f_X(x)} = \frac{f_{(X,Y)}(x, y)}{\int_{\mathbb{R}} f_{(X,Y)}(x, y) dy},$$

pour tout $y \in \mathbb{R}$. C'est une densité de probabilité. L'espérance associée est l'espérance conditionnelle sachant $\{X = x_0\}$, i.e que si $h : \mathbb{R} \rightarrow \mathbb{R}$ est une fonction mesurable bornée, alors,

$$\mathbb{E}[h(Y) | X = x_0] = \int_{\mathbb{R}} h(y) f_{Y|X=x_0}(y) dy.$$

Chapitre 4

Convergence de variables aléatoires

4.1 Différents modes de convergence

On commence par rappeler l'énoncé du lemme de BOREL-CANTELLI.

Lemme 4.1.1. *Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilités, et $(A_n)_{n \in \mathbb{N}}$ une suite d'évènements. Alors,*

(i) *Si $\sum_{n=0}^{\infty} \mathbb{P}(A_n) < \infty$, alors,*

$$\mathbb{P} \left(\limsup_{n \rightarrow \infty} A_n \right) = 0.$$

Les évènements se produisent un nombre fini de fois.

(ii) *Si les éléments de $(A_n)_n$ sont mutuellement indépendants, et si $\sum_{n=0}^{\infty} \mathbb{P}(A_n) = \infty$, alors,*

$$\mathbb{P} \left(\limsup_{n \rightarrow \infty} A_n \right) = 1.$$

Les évènements se produisent une infinité de fois.

Démonstration. On rappelle que $\limsup_{n \rightarrow \infty} A_n = \bigcap_{n \in \mathbb{N}} \bigcup_{k \geq n} A_k$ (intersection d'union décroissante). Ainsi,

$$\mathbb{P} \left(\limsup_{n \rightarrow \infty} A_n \right) = \lim_{n \rightarrow \infty} \mathbb{P} \left(\bigcup_{k \geq n} A_k \right).$$

(i) Si $\sum_{n=0}^{\infty} \mathbb{P}(A_n) < \infty$, alors par sous-additivité,

$$\mathbb{P} \left(\bigcup_{k \geq n} A_k \right) \leq \sum_{k=n}^{\infty} \mathbb{P}(A_k) \xrightarrow{n \rightarrow \infty} 0,$$

d'où le résultat.

(ii) Si les éléments de $(A_n)_n$ sont mutuellement indépendants, et si $\sum_{n=0}^{\infty} \mathbb{P}(A_n) = \infty$, on a

$$\begin{aligned} \mathbb{P} \left(\left(\limsup_{n \rightarrow \infty} A_n \right)^c \right) &= \mathbb{P} \left(\liminf_{n \rightarrow \infty} A_n^c \right) \\ &= \mathbb{P} \left(\bigcup_{n \rightarrow \infty} \bigcap_{k \geq n} A_k^c \right) \\ &= \lim_{n \rightarrow \infty} \mathbb{P} \left(\bigcap_{k \geq n} A_k^c \right). \end{aligned}$$

Par indépendance mutuelle, on a pour tout $n \in \mathbb{N}$ que

$$\mathbb{P} \left(\bigcap_{k \geq n} A_k^c \right) = \prod_{k \geq n} \mathbb{P}(A_k^c) = \prod_{k \geq n} (1 - \mathbb{P}(A_k)).$$

En passant au logarithme et en utilisant l'inégalité de concavité " $\ln(1 - x) \leq -x$ ", on a

$$\ln \left(\mathbb{P} \left(\bigcap_{k \geq n} A_k^c \right) \right) = \sum_{k \geq n} \ln(1 - \mathbb{P}(A_k)) \leq$$

□

4.1.1 Convergence presque sûre

Définition 4.1.1. On dit qu'une suite $(X_n)_{n \in \mathbb{N}}$ de variables aléatoires converge presque sûrement vers une variable aléatoire X si

$$\mathbb{P} \left(\left\{ \omega \in \Omega \mid \lim_{n \rightarrow \infty} X_n(\omega) = X(\omega) \right\} \right) = 1.$$

Notation. $X_n \xrightarrow[n \rightarrow \infty]{p.s.} X$.

Remarques.

- En d'autres termes, $(X_n)_n$ converge simplement vers X sur un ensemble de mesure pleine.
- Cet ensemble est mesurable comme ensemble de points d'égalités des \liminf et des \limsup des X_n , qui sont des fonctions mesurables.
- On a

$$\mathbb{P} \left(\left\{ \omega \in \Omega \mid \lim_{n \rightarrow \infty} X_n(\omega) = X(\omega) \right\} \right) = \underbrace{\bigcap_{p \geq 1} \bigcup_{n \geq 1} \bigcap_{k \geq n} \left\{ |X_k(\omega) - X(\omega)| < \frac{1}{p} \right\}}_{\text{décroissante en } p}.$$

Ce qui justifie aussi la mesurabilité.

— Par monotonie,

$$\begin{aligned} X_n \xrightarrow[n \rightarrow \infty]{p.s} X &\iff \forall p \geq 1, \mathbb{P} \left(\bigcup_{n \geq 1} \bigcap_{k \geq n} \left\{ |X_k(\omega) - X(\omega)| < \frac{1}{p} \right\} \right) = 1 \\ &\iff \forall \varepsilon > 0, \mathbb{P} \left(\bigcup_{n \geq 1} \bigcap_{k \geq n} \left\{ |X_k(\omega) - X(\omega)| < \varepsilon \right\} \right) = 1 \\ &\iff \forall \varepsilon > 0, \mathbb{P} \left(\liminf_{n \rightarrow \infty} \left\{ |X_n - X| < \varepsilon \right\} \right) = 1 \\ &\iff \forall \varepsilon > 0, \mathbb{P} \left(\limsup_{n \rightarrow \infty} \left\{ |X_n - X| > \varepsilon \right\} \right) = 0. \end{aligned}$$

Corollaire 4.1.1. Soit $(X_n)_n$ une suite de variables aléatoires, et X une variable aléatoire.

- (i) Si l'on suppose que pour tout $\varepsilon > 0$, $\sum_{n=0}^{\infty} \mathbb{P}(|X_n - X| > \varepsilon) < \infty$. Alors, $(X_n)_n$ converge presque sûrement vers X .
- (ii) Si l'on suppose que les X_n sont mutuellement indépendantes, alors :

$$X_n \xrightarrow[n \rightarrow \infty]{p.s} 0 \iff \forall \varepsilon > 0, \sum_{n=0}^{\infty} \mathbb{P}(|X_n| > \varepsilon) < \infty.$$

Démonstration. Appliquer le lemme de BOREL-CANTELLI à la définition de convergence presque sûre. □

Proposition 4.1.1. Si $X_n \xrightarrow[n \rightarrow \infty]{p.s} X$ et si f est une fonction continue, alors

$$f(X_n) \xrightarrow[n \rightarrow \infty]{p.s} f(X).$$

Démonstration. Par continuité de f , $f(X_n(\omega)) \xrightarrow[n \rightarrow \infty]{} f(X(\omega))$. On en déduit que

$$\mathbb{P} \left(\left\{ \omega \in \Omega \mid f(X_n(\omega)) \xrightarrow[n \rightarrow \infty]{} f(X(\omega)) \right\} \right) \geq \mathbb{P} \left(\left\{ \omega \in \Omega \mid X_n(\omega) \xrightarrow[n \rightarrow \infty]{} X(\omega) \right\} \right) = 1. \quad \square$$

4.1.2 Convergence en probabilité

Définition 4.1.2. On dit qu'une suite $(X_n)_n$ de variables aléatoires convergence en probabilité vers X si :

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

Notation. $X_n \xrightarrow[n \rightarrow \infty]{\mathbb{P}} X$.

Proposition 4.1.2 (unicité de la limite). Si $X_n \xrightarrow[n \rightarrow \infty]{\mathbb{P}} X$ et $X_n \xrightarrow[n \rightarrow \infty]{\mathbb{P}} Y$, alors $X = Y$ \mathbb{P} -p.p.

Proposition 4.1.3. Si $X_n \xrightarrow[n \rightarrow \infty]{\mathbb{P}} X$ et si f est une fonction continue, alors

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

Remarque. On peut métriser la convergence en probabilité de la manière suivante : pour tout $X, Y \in \mathbb{L}^0(\Omega, \mathcal{F}, \mathbb{P})$, on pose

$$d_{\mathbb{P}}(X, Y) = \mathbb{E}[\min\{|X - Y|, 1\}].$$

L'espace métrique $(\mathbb{L}^0(\Omega, \mathcal{F}, \mathbb{P}), d_{\mathbb{P}})$ est un espace complet.

4.1.3 Convergence \mathbb{L}^p

Définition 4.1.3. Soit $p \geq 1$. On dit que qu'une suite $(X_n)_n$ de variables aléatoires converge vers la variable aléatoire X dans \mathbb{L}^p si

$$\|X_n - X\|_p \xrightarrow[n \rightarrow \infty]{} 0,$$

$$\text{i.e. } \mathbb{E}[|X_n - X|^p] \xrightarrow[n \rightarrow \infty]{} 0.$$

Notation. $X_n \xrightarrow[n \rightarrow \infty]{\mathbb{L}^p} X$.

Remarque. En vertu de l'inégalité de HÖLDER, si $q \leq p$, alors une convergence \mathbb{L}^p implique une convergence \mathbb{L}^q .

4.1.4 Convergence en loi

Définition 4.1.4. On dit que qu'une suite $(X_n)_n$ de variables aléatoires converge en loi vers la variable aléatoire X si pour toute fonction f continue et bornée,

$$\mathbb{E}[f(X_n)] \xrightarrow[n \rightarrow \infty]{} \mathbb{E}[f(X)].$$

Notation. $X_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} X$.

Remarque. Une convergence en loi se réécrit de la manière suivante : pour toute fonction f continue et bornée,

$$\int_{\mathbb{R}} f(x) d\mathbb{P}_{X_n}(x) \xrightarrow[n \rightarrow \infty]{} \int_{\mathbb{R}} f(x) d\mathbb{P}_X(x).$$

C'est la convergence "étroite" de \mathbb{P}_{X_n} vers \mathbb{P}_X . La particularité de cette loi est qu'elle ne dépend que des lois des X_n et de X .

Lemme 4.1.2 (portemanteau). Soit $(X_n)_n$ une suite de variable aléatoire, et X une variable aléatoire. Sont équivalents.

- (i) $(X_n)_n$ converge en loi vers X .
- (ii) Pour tout fermé F ,

$$\limsup_{n \rightarrow \infty} \mathbb{P}(X_n \in F) \leq \mathbb{P}(X \in F).$$

- (iii) Pour tout ouvert O ,

$$\liminf_{n \rightarrow \infty} \mathbb{P}(X_n \in O) \geq \mathbb{P}(X \in O).$$

- (iv) Pour tout borélien B tel que $\mathbb{P}(X \in \partial B) = 0$, on a

$$\lim_{n \rightarrow \infty} \mathbb{P}(X_n \in B) = \mathbb{P}(X \in B).$$

Proposition 4.1.4. Soit $(X_n)_n$ une suite de variables aléatoires réelles. Alors, $(X_n)_n$ converge en loi vers la variable aléatoire X si, et seulement si, $F_{X_n}(t) \xrightarrow[n \rightarrow \infty]{} F_X(t)$ pour tout point de continuité $t \in \mathbb{R}$ de F_X .

Théorème 4.1.1. La suite $(X_n)_{n \in \mathbb{N}}$ converge en loi vers X si, et seulement si, pour tout $t \in \mathbb{R}$, $\varphi_{X_n}(t) \xrightarrow[n \rightarrow \infty]{} \varphi_X(t)$.

Théorème 4.1.2 (LEVY). Soit $(\varphi_{X_n})_n$ une suite de fonctions caractéristiques qui converge ponctuellement vers une fonction φ . Alors, si φ est **continue en zéro**, elle est une fonction caractéristique d'une variable aléatoire X telle que

$$X_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} X.$$

Théorème 4.1.3 (continuous mapping theorem). Soit $(X_n)_n$ une suite de variables aléatoires, et X une variable aléatoire telle que

$$X_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} X.$$

Si f est continue \mathbb{P}_X -p.p (presque sûrement), alors

$$f(X_n) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} f(X).$$

4.2 Articulation des modes de convergence

4.2.1 Convergence presque sûre en probabilité

Lemme 4.2.1. — Si $X_n \xrightarrow[n \rightarrow \infty]{p.s.} X$, alors $X_n \xrightarrow[n \rightarrow \infty]{\mathbb{P}} X$.

— Si $X_n \xrightarrow[n \rightarrow \infty]{\mathbb{P}} X$, alors il existe une sous-suite $(X_{\varphi(n)})_n$ telle que $X_{\varphi(n)} \xrightarrow[n \rightarrow \infty]{p.s.} X$.

4.2.2 Convergences \mathbb{L}^p , p.s, \mathbb{P}

Lemme 4.2.2. Si $X_n \xrightarrow[n \rightarrow \infty]{p.s.} X$, et qu'il existe une variable aléatoire $Y \in \mathbb{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ telle que $|X_n| \leq Y$ pour tout $n \in \mathbb{N}$, alors

$$X_n \xrightarrow[n \rightarrow \infty]{\mathbb{L}^1} X.$$

Démonstration. Appliquer le théorème de convergence dominée. □

Lemme 4.2.3. Si $X_n \xrightarrow[n \rightarrow \infty]{\mathbb{L}^p} X$, alors $X_n \xrightarrow[n \rightarrow \infty]{\mathbb{P}} X$.

Définition 4.2.1. Une famille de $(X_i)_{i \in I}$ de variables aléatoires est dite **uniformément intégrable** (ou *équii-intégrable*) si

$$\lim_{n \rightarrow \infty} \sup_{i \in I} \mathbb{E}[|X_i| \mathbb{1}_{|X_i| > n}] = 0.$$

Proposition 4.2.1. La famille $(X_i)_{i \in I}$ est uniformément intégrable si, et seulement si,

— $(X_i)_i$ est bornée dans \mathbb{L}^1 , i.e

$$\sup_{i \in I} \mathbb{E}[|X_i|] < \infty,$$

— pour tout $\varepsilon > 0$, il existe $\delta > 0$, tel que pour tout $A \in \mathcal{F}$,

$$(\mathbb{P}(A) < \delta) \implies \left(\sup_{i \in I} \mathbb{E}[|X_i| \mathbb{1}_A] < \varepsilon \right).$$

Proposition 4.2.2. Soit $(X_n)_n$ une suite de variables aléatoires intégrables. Sont équivalents.

(i) $X_n \xrightarrow[n \rightarrow \infty]{\mathbb{L}^1} X$.

(ii) $(X_n)_{n \in \mathbb{N}}$ est uniformément intégrable et $X_n \xrightarrow[n \rightarrow \infty]{\mathbb{P}} X$.

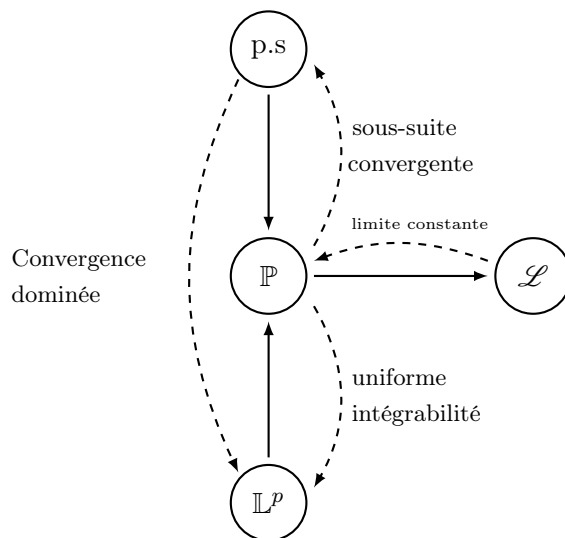


FIGURE 4.1 – En résumé.

4.2.3 Convergence en loi et autres modes

Lemme 4.2.4. Si $X_n \xrightarrow[n \rightarrow \infty]{\mathbb{P}} X_\infty$, alors $X_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} X_\infty$.

Démonstration. Soit $t \in \mathbb{R}$, $\varepsilon > 0$, et $n \in \mathbb{N}$. On a

$$\begin{aligned}
 |\varphi_{X_n}(t) - \varphi_{X_\infty}(t)| &= |\mathbb{E}[e^{itX_n} - e^{itX_\infty}]| \\
 &\leq |t| \cdot \mathbb{E}[\min(|X_n - X_\infty|, 2)] \\
 &\leq |t| \mathbb{E}[|X_n - X_\infty| \mathbf{1}_{|X_n - X_\infty| < \varepsilon}] + |t| \mathbb{E}[\min(|X_n - X_\infty|, 2) \mathbf{1}_{|X_n - X_\infty| \geq \varepsilon}] \\
 &\leq |t| \varepsilon + 2|t| \mathbb{P}(|X_n - X_\infty| \geq \varepsilon).
 \end{aligned}$$

On en déduit que $\limsup |\varphi_{X_n}(t) - \varphi_{X_\infty}(t)| \dots$ □

Corollaire 4.2.1. Si $X_n \xrightarrow[n \rightarrow \infty]{p.s, L^1} X_\infty$, alors $X_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} X_\infty$.

Lemme 4.2.5. Si $X_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} c \in \mathbb{R}$, alors $X_n \xrightarrow[n \rightarrow \infty]{\mathbb{P}} c$.

Démonstration. Soit $\varepsilon > 0$. On a $\mathbb{P}(|X_n - c| > \varepsilon) = \mathbb{P}(X_n < c - \varepsilon) + \mathbb{P}(X_n > c + \varepsilon)$. La variable aléatoire constante à c a pour fonction de répartition la fonction $F_c = \mathbf{1}_{[c, \infty[}$. On a alors

$$\mathbb{P}(|X_n - c| > \varepsilon) = \underbrace{F_c((c - \varepsilon)^-)}_{=0} + 1 - \underbrace{F_c(c + \varepsilon)}_{=0} = 0 + 1 - 1 = 0.$$

□

Chapitre 5

Théorèmes limites

Dans tout le chapitre, on se place dans un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$. On suppose que dans cette espace est définie une suite de variables $(X_n)_{n \in \mathbb{N}}$ *indépendantes (mutuellement) identiquement distribuées* (i.i.d).

5.1 Loi des grands nombres (LGN)

5.1.1 Loi faible des grands nombres

Lemme 5.1.1. Soit $a, b \in \mathbb{C}^{\mathbb{N}^*}$, des suites telles que $|a_i| \leq 1$ et $|b_i| \leq 1$ pour tout $i \in \mathbb{N}^*$. Alors pour tout $n \in \mathbb{N}^*$,

$$\left| \prod_{i=1}^n a_i - \prod_{i=1}^n b_i \right| \leq \sum_{i=1}^n |a_i b_i|.$$

Démonstration. Par récurrence. □

Théorème 5.1.1. Soit $(X_n)_n$ une suite de variables aléatoire i.i.d intégrables (i.e $\mathbb{E}[|X_1|] < \infty$), alors

$$\frac{S_n}{n} = \frac{1}{n} \sum_{i=1}^n X_i \xrightarrow[n \rightarrow \infty]{\mathbb{P}} \mathbb{E}[X_1].$$

Remarque. Alors que la moyenne des n premiers X_i est une variable aléatoire, la limite elle est purement déterministe.

Démonstration. — On suppose dans un premier temps les X_i sont éléments de $\mathbb{L}^2(\Omega, \mathcal{F}, \mathbb{P})$. Soit $\varepsilon > 0$. On remarque que $\mathbb{E}[\frac{S_n}{n}] = \mathbb{E}[X_1]$. Par l'inégalité de TCHEBYTCHEV, on a

$$\mathbb{P} \left(\frac{S_n}{n} - \mathbb{E}[X_1] > \varepsilon \right) \leq \frac{\text{var}(\frac{S_n}{n})}{\varepsilon^2} = \frac{1}{n^2 \varepsilon^2} \text{var}(S_n) = \frac{1}{n^2 \varepsilon^2} n \text{var}(X_1) = \frac{1}{n \varepsilon^2} \text{var}(X_1) \xrightarrow[n \rightarrow \infty]{} 0.$$

— Quitte à considérer les $X_i - \mathbb{E}[X_i]$, on peut supposer $\mathbb{E}[X_i] = 0$. Selon le lemme 4.2.5, il suffit ici de montrer que $\frac{S_n}{n} \xrightarrow[n \rightarrow \infty]{\mathcal{L}} 0$. Soit $t \in \mathbb{R}$, par indépendance,

$$\varphi_{\frac{S_n}{n}}(t) = \mathbb{E} \left[e^{i \frac{t}{n} \sum_{i=1}^n X_i} \right] = \varphi_{X_1} \left(\frac{t}{n} \right)^n.$$

D'après le lemme précédent,

$$\left| \varphi_{\frac{S_n}{n}}(t) - 1 \right| \leq n \left| \varphi_{X_1} \left(\frac{t}{n} \right) - 1 \right|.$$

Puisque $\mathbb{E}[|X_i|] < \infty$, φ_{X_1} est dérivable en zéro et $\varphi'_{X_1}(0) = i\mathbb{E}[X_1] = 0$. On écrit alors

$$t \cdot \frac{n}{t} \left| \varphi_{X_1} \left(\frac{t}{n} \right) - 1 \right| \xrightarrow[n \rightarrow \infty]{} t \cdot |\varphi'_{X_1}(0)| = 0$$

en remarquant que le membre de gauche est un taux d'accroissement en zéro. En conclusion, \square

5.1.2 Loi forte des grands nombres

Théorème 5.1.2. Soit $(X_n)_n$ une suite de variables aléatoire i.i.d intégrables (i.e $\mathbb{E}[|X_1|] < \infty$), alors

$$\frac{S_n}{n} = \frac{1}{n} \sum_{i=1}^n X_i \xrightarrow[n \rightarrow \infty]{p.s., L^1} \mathbb{E}[X_1].$$

Remarque. Si une suite $(X_n)_n$ de variables aléatoires i.i.d vérifie $\frac{S_n}{n} \xrightarrow[n \rightarrow \infty]{p.s.} c \in \mathbb{R}$, alors $\mathbb{E}[|X_1|] < \infty$ et $c = \mathbb{E}[X_1]$.

5.2 Théorème limite central (TLC)

Il faut bien comprendre dans le théorème “limite centrale” que les mots “limite” et “central” sont des adjectifs. Ce théorème est en fait un raffinement de la loi des grands nombres.

5.2.1 Théorème limite central

Théorème 5.2.1. Soit $(X_n)_n$ une suite de variables aléatoires i.i.d avec $\mathbb{E}[|X_i|^2] < \infty$. On pose $\sigma^2 = \text{var}(X_1)$. Alors,

$$\sqrt{n} \left(\frac{S_n}{n} - \mathbb{E}[X_1] \right) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, \sigma^2).$$

De manière équivalente,

$$\frac{\sqrt{n}}{\sigma} \left(\frac{S_n}{n} - \mathbb{E}[X_1] \right) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, 1).$$

Remarques.

- Le théorème limite central assure que la vitesse de convergence dans la loi des grands nombres est $\mathcal{O}\left(\frac{1}{\sqrt{n}}\right)$.
- La loi gaussienne $\mathcal{N}(0, 1)$ apparaît comme une loi limite universelle.

Avant de démontrer ce théorème, nous avons besoin d'une estimation sur le reste d'ordre quelconque du développement de TAYLOR de l'exponentielle complexe.

Lemme 5.2.1. *Pour tout $x \in \mathbb{R}$, et $p \in \mathbb{N}^*$. Alors,*

$$\left| e^{ix} - \sum_{k=0}^p \frac{(ix)^k}{k!} \right| \leq \min \left(\frac{|x|^{p+1}}{(p+1)!}, \frac{2|x|^p}{p!} \right).$$

Démonstration. Utiliser le théorème de TAYLOR reste intégral. On majore sans soucis pour faire apparaître le premier terme du minimum. Pour le second, on intègre par parties le reste. □

On montre désormais le théorème limite central.

Démonstration. Quitte à considérer $\frac{X_i - \mathbb{E}[X_i]}{\sigma}$, on peut supposer que $\mathbb{E}[X_i] = 0$ et $\text{var}(X_i) = 1$. On calcule la fonction caractéristique de $\frac{S_n}{\sqrt{n}}$: par indépendance, pour tout $t \in \mathbb{R}$,

$$\varphi_{\frac{S_n}{\sqrt{n}}}(t) = \varphi_{X_1} \left(\frac{t}{\sqrt{n}} \right)^n.$$

Puisque X_1 admet un moment d'ordre 2, φ_{X_1} est deux fois dérivable. Alors,

$$\varphi_{X_1}(t) = 1 + t\varphi'_{X_1}(0) + \frac{t^2}{2}\varphi''_{X_1}(0) + o(t^2) = 1 - \frac{t^2}{2} + o(t^2).$$

D'après le lemme 5.1.1 avec $a_i = \varphi_{X_1}\left(\frac{t}{\sqrt{n}}\right)$ et $b_i = 1 - \frac{t^2}{2n}$, on a

$$\left| \varphi_{\frac{S_n}{\sqrt{n}}}(t) - \left(1 - \frac{t^2}{2n}\right)^n \right| \leq n \left| \varphi_{X_1} \left(\frac{t}{\sqrt{n}} \right) - \left(1 - \frac{t^2}{2n}\right) \right|$$

On applique ensuite le lemme précédent, et alors,

$$\left| \varphi_{X_1} \left(\frac{t}{\sqrt{n}} \right) - \left(1 - \frac{t^2}{2n}\right) \right| \leq \min \left(\frac{|t|^3|X_1|^3}{6n^{3/2}}, \frac{|t|^2|X_1|^2}{n} \right).$$

On en déduit que

$$\left| \varphi_{\frac{S_n}{\sqrt{n}}}(t) - \left(1 - \frac{t^2}{2n}\right)^n \right| \leq \min \left(\frac{|t|^3|X_1|^3}{6\sqrt{n}}, |t|^2|X_1|^2 \right).$$

□

Remarque.

Théorème 5.2.2. Soit $(X_n)_n$ une suite de variables indépendantes avec $\text{var}(X_i) < \infty$. On pose $\sigma_n = \sqrt{\sum_{k=1}^n \text{var}(X_k)}$ pour tout $n \in \mathbb{N}^*$. Alors,

$$Z_n = \frac{1}{\sigma_n} \sum_{k=1}^n (X_k - \mathbb{E}[X_k]) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, 1).$$

5.2.2 Retour sur les applications de la loi des grands nombres

Méthode de Monte-Carlo

Chapitre 6

Vecteurs gaussiens

On rappelle qu'une variable gaussienne est une variable aléatoire réelle $X \sim \mathcal{N}(m, \sigma^2)$ si elle admet la densité

$$x \mapsto f_X(x) = \frac{e^{-\frac{(x-m)^2}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}}.$$

De plus, on a pour tout $t \in \mathbb{R}$,

$$\varphi_X(t) = e^{itm} e^{-\frac{\sigma^2 t^2}{2}}.$$

On remarque de plus que $\frac{X-m}{\sigma} \sim \mathcal{N}(0, 1)$. Inversement, si $Y \sim \mathcal{N}(0, 1)$, alors $\sigma Y + m \sim \mathcal{N}(m, \sigma^2)$. Dans le cas où $\sigma^2 = 0$, on dit que X est dégénérée : $X = m$ presque sûrement. On rappelle enfin des notations usuelles : $\langle \cdot, \cdot \rangle$ dénotera le produit scalaire euclidien sur \mathbb{R}^d et $\|\cdot\|$ la norme euclidienne usuelle.

6.1 Vecteurs gaussiens

6.1.1 Définitions et propriétés élémentaires

Définition 6.1.1 (vecteurs gaussiens). *Un vecteur aléatoire $X = (X_1, \dots, X_d)^T$ est dit être **gaussien** si toute combinaison linéaire de ses composantes est gaussienne, i.e que pour tout $a \in \mathbb{R}^d$, $\langle a, X \rangle$ est gaussienne.*

Remarques.

- Si $X = (X_1, \dots, X_d)^T$ est gaussien, alors en particulier pour tout $i \in \llbracket 1, d \rrbracket$, X_i est une variable gaussienne. La réciproque est fautive. Par exemple, si $X \sim \mathcal{N}(0, 1)$, et $\varepsilon \sim \mathcal{B}(\pm 1, 1/2)$, alors les composantes de $(X, \varepsilon X)$ sont des variables gaussiennes mais $(X, \varepsilon X)$ n'est pas un vecteur gaussien car $X + \varepsilon X$ car elle n'est pas constante et

$$\mathbb{P}(X + \varepsilon X = 0) \geq \mathbb{P}(\varepsilon = -1) = \frac{1}{2}.$$

On en déduit que cette variable a un atome donc elle ne peut pas être à densité (donc *a fortiori* gaussienne). Un autre contre-exemple classique est de prendre $X \sim \mathcal{N}(0, 1)$ et de poser $Y = X$ si $|X| \geq 1$ et $Y = -X$ sinon. Alors $Y \sim \mathcal{N}(0, 1)$ mais (X, Y) n'est pas gaussien.

— Si X est un vecteur gaussien, alors X est \mathbb{L}^2 . En effet, par CAUCHY-SCHWARZ, on a

$$\mathbb{E}[|X_i X_j|]^2 \leq \mathbb{E}[X_i^2] \mathbb{E}[X_j^2] < \infty.$$

— L'intérêt de l'étude des vecteurs gaussiens sera en fait de remarquer que tout se passe "bien" dans l'espace des vecteurs gaussiens car cet espace est stable pour de nombreuses transformations linéaires.

Définition 6.1.2. Si $X = (X_1, \dots, X_d)^T$ est un vecteur gaussien, on définit sa moyenne comme la quantité

$$m = (\mathbb{E}[X_1], \dots, \mathbb{E}[X_d])^T.$$

De plus, on appelle **matrice de covariance** de X la matrice $K = (\text{cov}(X_i, X_j))_{i,j \in \llbracket 1, d \rrbracket}$. Cette matrice est bien définie selon la remarque précédente.

Remarques.

— On rappelle que

$$\text{cov}(X_i, X_j) = \mathbb{E}[X_i X_j - \mathbb{E}[X_i] \mathbb{E}[X_j]].$$

— La matrice de covariance K est symétrique positive. En effet, si $x \in \mathbb{R}^d$,

$$x^T K x = \sum_{i,j \in \llbracket 1, d \rrbracket} x_i x_j \text{cov}(X_i, X_j) = \text{var} \left(\sum_{i=1}^d x_i X_i \right) \geq 0.$$

Proposition 6.1.1. Un vecteur gaussien $X = (X_1, \dots, X_d)^T$ de moyenne m et de matrice de covariance K a pour fonction caractéristique la fonction définie pour tout $x \in \mathbb{R}^d$,

$$\varphi_X(x) = \mathbb{E} \left[e^{i \langle x, X \rangle} \right] = e^{i \langle x, m \rangle} e^{-\frac{1}{2} x^T K x}.$$

Démonstration. Par définition, si X est gaussien et $x \in \mathbb{R}^d$, alors $\langle x, X \rangle$ est un variable gaussienne. On a

$$\mathbb{E}[\langle x, X \rangle] = \mathbb{E} \left[\sum_{k=1}^d x_k X_k \right] = \sum_{k=1}^d x_k \mathbb{E}[X_k] = \langle x, m \rangle.$$

De plus, selon la remarque précédente,

$$\text{var}(\langle x, X \rangle) = \text{var} \left(\sum_{i=1}^d x_i X_i \right) = x^T K x.$$

Ainsi, la fonction caractéristique de $\langle x, X \rangle$ prise en $t = 1$ vaut

$$\mathbb{E} \left[e^{i \langle x, X \rangle} \right] = e^{i \langle x, m \rangle} e^{-\frac{1}{2} x^T K x}.$$

On remarque alors que la loi de X est entièrement caractérisée par $m = \mathbb{E}[X]$ et la matrice de covariance K . □

Notation. On note $X \sim \mathcal{N}_d(m, K)$. La proposition précédente assure que cette notation est bien fondée.

Remarque. Les caractéristiques d'un vecteur gaussien se lisent sur la transformée de FOURIER (*i.e* la fonction caractéristique). Par exemple, si

$$\varphi_X(s, t) = e^{2is+3it} e^{-\frac{1}{2}(s^2-2st+2t^2)},$$

alors $m = (2, 3)$ et

$$K = \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix}.$$

En fait, on trouve dans la partie imaginaire de l'exponentielle les composantes de m , et dans les coefficients de la forme quadratique de l'exponentielle réelle les coefficients de K (qui sont les coefficients de la forme quadratique en question).

Proposition 6.1.2. Soient $X \sim \mathcal{N}_d(m_X, K_X)$ et $Y \sim \mathcal{N}_d(m_Y, K_Y)$. Si X et Y sont indépendants, alors $X + Y$ est gaussien et $X + Y \sim \mathcal{N}_d(m_X + m_Y, K_X + K_Y)$.

Démonstration. Soit $x \in \mathbb{R}^d$. On a

$$\begin{aligned} \varphi_{X+Y}(x) &= \mathbb{E} \left[e^{i\langle x, X+Y \rangle} \right] \\ &= \mathbb{E} \left[e^{i\langle x, X \rangle + i\langle x, Y \rangle} \right] \\ &= \mathbb{E} \left[e^{i\langle x, X \rangle} \right] + \mathbb{E} \left[e^{i\langle x, Y \rangle} \right] \\ &= \left(e^{i\langle x, m_X \rangle} e^{-\frac{1}{2}x^T K_X x} \right) \left(e^{i\langle x, m_Y \rangle} e^{-\frac{1}{2}x^T K_Y x} \right) \\ &= e^{i\langle x, m_X + m_Y \rangle} e^{-\frac{1}{2}x^T (K_X + K_Y) x}. \end{aligned}$$

□

Soit $(X, Y) = (X_1, \dots, X_d, Y_1, \dots, Y_p)^T \in \mathbb{R}^{d+p}$ un vecteur gaussien de moyenne $(m_X, m_Y) \in \mathbb{R}^{d+p}$ et de matrice de covariance $K_{(X,Y)} \in \mathcal{M}_{d+p}(\mathbb{R})$.

Proposition 6.1.3. Les vecteurs gaussiens X et Y sont indépendants si, et seulement si, $K_{(X,Y)}$ est diagonable par bloc, *i.e*

$$K_{(X,Y)} = \begin{pmatrix} K_X & (0) \\ (0) & K_Y \end{pmatrix}$$

avec $K_X \in \mathcal{M}_d(\mathbb{R})$ et $K_Y \in \mathcal{M}_p(\mathbb{R})$.

Démonstration. Le sens direct est clair. Pour la réciproque, on calcule la fonction caractéristique. Soit $(x, y)^T \in \mathbb{R}^{d+p}$, on a

□

Proposition 6.1.4. Soient $(X_n)_{n \in \mathbb{N}}$ une suite de vecteurs gaussiens dans \mathbb{R}^d . On les note tels que pour tout $n \in \mathbb{N}$, $X_n \sim \mathcal{N}_d(m_n, K_n)$. Alors, si $(X_n)_n$ converge en loi vers un certain X , alors X est nécessairement gaussien. Plus particulièrement, $(X_n)_n$ converge en loi vers $X \sim \mathcal{N}_d(m, K)$ si, et seulement si, $m_n \xrightarrow{n \rightarrow \infty} m$ et $K_n \xrightarrow{n \rightarrow \infty} K$.

Démonstration. Voir TD.

□

Proposition 6.1.5. Soit $X \sim \mathcal{N}_d(m, K)$ et $A \in \mathcal{M}_{p,d}(\mathbb{R})$. Alors, le vecteur AX est gaussien et suit une loi $\mathcal{N}_d(Am, AK A^T)$.

Démonstration. Pour commencer, AX est gaussien car ses composantes sont des combinaisons linéaires du vecteur gaussien X . Ainsi, AX est aussi gaussien. En fait, pour tout $x \in \mathbb{R}^d$,

$$\langle AX, x \rangle = \langle X, A^T x \rangle.$$

Déterminons maintenant sa moyenne et sa matrice de covariance. Par définition, si $x \in \mathbb{R}^d$, alors $\langle AX, x \rangle = \langle X, A^T x \rangle$ de sorte que

$$\mathbb{E}[\langle AX, x \rangle] = \langle \mathbb{E}[X], A^T x \rangle = \langle m, A^T x \rangle = \langle Am, x \rangle.$$

De plus,

$$\text{var}(\langle AX, x \rangle) = \text{var}(\langle X, A^T x \rangle) = (A^T x)^T K (A^T x) = x^T (AK A^T) x.$$

On conclue par la caractérisation des vecteurs gaussiens.

□

Définition 6.1.3 (vecteur non dégénéré). Un vecteur gaussien $X \sim \mathcal{N}_d(m, K)$ est dit être **non dégénéré** si sa matrice de covariance est inversible.

Remarques.

- On dit alors que X est *dégénéré* si $\det K = 0$.
- Le X est dégénéré s'il existe $a \in \mathbb{R}^d$ tel que $Ka = 0$. Alors, $\text{var}(\langle a, X \rangle) = a^T K a = 0$ de sorte que aX est constante presque sûrement. On peut voir un vecteur dégénéré comme un vecteur vivant dans un hyperplan de \mathbb{R}^d . Par exemple, si $Y_1, \dots, Y_{d-1} \sim \mathcal{N}(0, 1)$ sont des variables indépendantes, alors le vecteur gaussien

$$X = \left(Y_1, \dots, Y_{d-1}, -\sum_{i=1}^{d-1} Y_i \right)$$

est dégénéré. On remarquera que X est gaussien parce que les Y_i sont indépendantes.

- En fait, X est non dégénéré si la forme quadratique associée à sa matrice de covariance est non dégénérée (?).
- Si $X \sim \mathcal{N}_d(m, K)$ est non dégénéré, K est symétrique définie positive. Si l'on diagonalise orthogonalement K sous la forme $K = PDP^T$ avec P une matrice orthonormée et $D = \text{diag}(\lambda_1, \dots, \lambda_d)$ avec $\lambda_i > 0$. Alors K admet une racine carrée matricielle qu'on note ici \sqrt{K} . On a alors

$$\sqrt{K} = P \text{diag} \left(\sqrt{\lambda_1}, \dots, \sqrt{\lambda_d} \right) P^T,$$

et on a

$$(\sqrt{K})^{-1} = P \text{diag} \left(\frac{1}{\sqrt{\lambda_1}}, \dots, \frac{1}{\sqrt{\lambda_d}} \right) P^T.$$

Avec une telle notation, on va pouvoir transformer X en l'équivalent d'une variable centrée réduite.

Proposition 6.1.6. *Si $X \sim \mathcal{N}_d(m, K)$ est un vecteur gaussien non dégénéré, alors $(\sqrt{K})^{-1}(X - m) \sim \mathcal{N}_d(0, I_d)$.*

Corollaire 6.1.1. *Soit $X \sim \mathcal{N}_d(m, K)$ un vecteur gaussien non dégénéré, alors X admet comme densité (pour la mesure de LEBESGUE sur \mathbb{R}^d) la fonction*

$$f_X : \begin{cases} \mathbb{R}^d & \longrightarrow \mathbb{R} \\ (x_1, \dots, x_d) & \longmapsto \frac{1}{(2\pi)^{d/2} \sqrt{\det K}} e^{-\frac{1}{2}(x-m)^T K^{-1}(x-m)}. \end{cases}$$

Démonstration. Si $K = I_d$, i.e que $X = (X_1, \dots, X_d)^T$ et les X_i sont i.i.d et $X_1 \sim \mathcal{N}(0, 1)$ alors pour tout $x \in \mathbb{R}^n$,

$$f_X(x) = \prod_{i=1}^d \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x_i^2} = \frac{1}{(2\pi)^{d/2}} e^{-\frac{1}{2}\|x\|^2}.$$

Dans le cas général, on utilise le changement de variable $x \mapsto (\sqrt{K})^{-1}(x - m)$. □

6.1.2 Théorème limite central multidimensionnel

Théorème 6.1.1. *Soit $(X^n)_{n \in \mathbb{N}} = ((X_1^n, \dots, X_d^n)^T)_{n \in \mathbb{N}}$ une suite de vecteurs aléatoires indépendants, de même loi avec $m = (\mathbb{E}[X_1^1], \dots, \mathbb{E}[X_d^1])$, et soit K la matrice de covariance, qu'on suppose de déterminant strictement positif. Alors,*

$$\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n X^i - m \right) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}_d(0, K).$$

De manière équivalente,

$$Z_n = \left(\sqrt{K}\right)^{-1} \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n X^i - m \right) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}_d(0, I_d).$$

Démonstration. Il faut et il suffit de montrer que

$$\varphi_{Z_n}(x) = \mathbb{E} \left[e^{i \langle x, Z_n \rangle} \right] \xrightarrow[n \rightarrow \infty]{} e^{-\frac{1}{2} \|x\|^2}.$$

Or, $\langle x, Z_n \rangle = y \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n X^i - m \right)$ où $Y = \left\langle \left(\sqrt{K}^{-1} \right)^T, x \right\rangle$.

...

□

Remarque. Un urne contient un *très grand* nombre M de boules numérotées de 1 à d . On note p_i la proportion de boules numérotées “ i ”. On suppose que $p_i > 0$ pour tout $i \in \llbracket 1, d \rrbracket$. On a évidemment $\sum_{i=1}^d p_i = 1$. On tire avec remise n boules dans l’urne. On note N_i^n le nombre de boules numérotées “ i ” sur les n tirages, de sorte que

$$\sum_{i=1}^d N_i^n = n.$$

Alors, la loi du vecteur $N^n = (N_1^n, \dots, N_d^n)^T$ est la loi multinomiale : pour tout $(k_1, \dots, k_d) \in \mathbb{N}^d$,

$$\mathbb{P}((N_1^n, \dots, N_d^n) = (k_1, \dots, k_d)) = \frac{n!}{\prod_{i=1}^d k_i!} p_1^{k_1} \dots p_d^{k_d}.$$

On veut connaître l’asymptotique de N^n lors que n tend vers l’infini. On remarque que si X_k désigne le numéro de la k -ième boule tirée, alors pour tout $i \in \llbracket 1, d \rrbracket$ et $n \in \mathbb{N}^*$,

$$N_i^n = \sum_{k=1}^n \mathbb{1}_{X_k=i}.$$

Par hypothèse, les X_k sont indépendants et

$$\mathbb{E}[\mathbb{1}_{X_k=i}] = \mathbb{P}(X_k = i) = p_i,$$

et

$$\text{cov}(\mathbb{1}_{X_k=i}, \mathbb{1}_{X_k=j}) = \mathbb{E}[\mathbb{1}_{X_k=i} \mathbb{1}_{X_k=j}] - \mathbb{E}[\mathbb{1}_{X_k=i}] \mathbb{E}[\mathbb{1}_{X_k=j}].$$

Cette covariance vaut $p_i(1 - p_i)$ si $i = j$, et $-p_i p_j$ si $i \neq j$. D’après la loi des grands nombres (appliquée à chaque coordonnée),

$$\frac{N^n}{n} \xrightarrow[n \rightarrow \infty]{p.s.} p = (p_1, \dots, p_d)^T.$$

Et même, d’après le théorème limite central multidimensionnel,

$$\sqrt{n} \left(\frac{N^n}{n} - p \right) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}_d(0, K).$$

6.2 Projections orthogonales

6.2.1 Théorème de COCHRAN

Définition 6.2.1 (loi du χ^2). Soit $X \sim \mathcal{N}_d(0, I_d)$. On appelle loi χ^2 (“khi-deux”) à d degrés de liberté la loi de $\|X\|^2 = \sum_{i=1}^d X_i^2$. En fait, $X^2 \sim \Gamma(d/2, 1/2)$, i.e que la densité associée est

$$x \longmapsto \frac{1}{2^d \Gamma(d/2)} \dots$$

On considère une décomposition orthogonale

$$\mathbb{R}^d = \bigoplus_{i=1}^k V_i \quad \text{avec } \dim(V_i) = d_i \text{ pour tout } i \in \llbracket 1, k \rrbracket.$$

On a évidemment $d = \sum_{i=1}^k d_i$. Pour tout $i \in \llbracket 1, k \rrbracket$, on note β_{V_i} la projection orthogonale sur V_i .

Théorème 6.2.1 (COCHRAN). Soit $X \sim \mathcal{N}_d(0, I_d)$. Alors, les projections orthogonales $\pi_{V_i}(X)$ sont des vecteurs gaussiens indépendants et pour tout $i \in \llbracket 1, k \rrbracket$, $\|\pi_{V_i}(X)\|^2 \sim \chi^2(\dim(V_i))$.

Remarques.

— Si $k = 2$, alors $\mathbb{R}^d = V \oplus V^\perp$. Alors, $\pi_V(X)$ et $\pi_{V^\perp}(X)$ sont des variables indépendantes et

$$\|\pi_V(X)\|^2 \sim \chi^2(\dim(V)) \quad \text{et} \quad \|\pi_{V^\perp}(X)\|^2 \sim \chi^2(d - \dim(V)).$$

En fait, le théorème de COCHRAN s’apparente fortement à un théorème de PYTHAGORE en loi.

— Il existe un résultat analogue lorsque la loi n’est pas centré : si $X \sim \mathcal{N}_d(0, K)$, il faut considérer une décomposition orthogonale pour la forme quadratique définie pour tout $x, y \in \mathbb{R}^d$ par

$$\langle x, y \rangle_K = \langle x, Ky \rangle.$$

Démonstration. On considère $(e_{i,j})_{\substack{1 \leq i \leq k \\ 1 \leq j \leq d_i}}$ une base orthonormée associée à la décomposition

$$\mathbb{R}^d = \bigoplus_{i=1}^k V_i.$$

Alors, pour tout $i \in \llbracket 1, k \rrbracket$,

$$\pi_{V_i}(X) = \sum_{j=1}^{d_i} \langle X, e_{i,j} \rangle e_{i,j}.$$

On en déduit que les vecteurs $\pi_{V_i}(X)$ sont gaussiens car toute combinaison linéaire de leurs composantes est une combinaison linéaire des composantes de X . Par orthogonalité, si $(i, j) \neq (k, l)$, alors

$$\text{cov}(\langle X, e_{i,j} \rangle, \langle X, e_{k,l} \rangle) = e_{i,j}^T I_d e^{k,l} = 0.$$

Cela revient à dire que la matrice covariance du vecteur

$$(\pi_{V_1}(X), \pi_{V_2}(X), \dots, \pi_{V_k}(X))$$

est diagonale par blocs, *i.e* que les variables $\pi_{V_i}(X)$ sont indépendantes. Puisque $x \mapsto \pi_V(x)$ est une application linéaire, on en déduit que $\pi_V(X) \sim \mathcal{N}(0, \pi_V I_d \pi_V^T)$, donc $\pi_V(X) = \mathcal{N}(0, \pi_V)$ (??). Ainsi,

$$\|\pi_V(X)\|^2 = \chi^2(\dim(V)).$$

□

6.2.2 Test d'adéquation du χ^2

Soit X une variable aléatoire à valeurs dans un ensemble fini $E = \{a_1, \dots, a_d\}$. On note $p = (p_1, \dots, p_d)$ la loi de X , *i.e* que $\mathbb{P}(X = a_i) = p_i > 0$ pour tout $i \in \llbracket 1, d \rrbracket$. Par ailleurs, on se donne une loi cible $\Pi = (\pi_1, \dots, \pi_d)$ sur E . Le problème est le suivant : au vu de n réalisations (x_1, \dots, x_n) de variables (X_1, \dots, X_n) (*i.e* que $x_i \in X_i(\Omega)$) mutuellement indépendantes de loi X , peut-on décider si $p = \pi$ ou non ? On introduit la statistique du χ^2 : pour tout $n \in \mathbb{N}^*$,

$$T_n = n \sum_{i=1}^d \frac{\left(\frac{N_i^n}{n} - \pi\right)^2}{\pi}$$

où $N_i^n = \text{card}\{k \in \llbracket 1, n \rrbracket \mid x_k = a_i\}$ pour tout $i \in \llbracket 1, d \rrbracket$ et $n \in \mathbb{N}^*$. En fait, le terme $\frac{N_i^n}{n}$ est la proportion empirique, et π_i est la proportion théorique. L'idée est que T_n mesure l'écart entre les proportions empiriques et théoriques.

Théorème 6.2.2. *On conserve les notations du paragraphe précédent.*

(i) *Sous l'hypothèse que " $p = \pi$ ", alors*

$$T_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \chi^2(d-1).$$

(ii) *Sous l'hypothèse que " $p \neq \pi$ ", alors*

$$T_n \xrightarrow[n \rightarrow \infty]{p.s.} \infty.$$

Démonstration. (i) D'après le théorème limite central multidimensionnel (application multinomiale),

$$\sqrt{n} \left(\frac{N_i^n}{n} - p \right) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}_d(0, K)$$

avec $K_{i,j}$ valant $p_i(1-p_i)$ si $i = j$ et $-p_i p_j$ sinon. Si $D = \text{diag}\left(\frac{1}{\sqrt{\pi_1}}, \dots, \frac{1}{\sqrt{\pi_d}}\right)$. On a par le *continuous mapping theorem*,

$$\sqrt{n}D \left(\sqrt{n} \left(\frac{N^n}{n} - p \right) \right) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} D\mathcal{N}_d(0, K) = \mathcal{N}_d(0, DKD^{-1}).$$

On a $\Sigma := DKD^{-1} = I_d - \sqrt{\pi}\sqrt{\pi}^T$, i.e que Σ est la projection sur $(\sqrt{\pi})^\perp$, qui est un espace vectoriel de dimension $d - 1$. Sous l'hypothèse que $p = \pi$, on a

$$\sqrt{n}D \left(\frac{N^n}{n} - p \right) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}_d(0, \Sigma).$$

Ainsi,

$$T_n = \left\| \sqrt{n}D \left(\frac{N^n}{n} - p \right) \right\|^2 \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \|\mathcal{N}_d(0, \Sigma)\|^2 \sim \chi^2(d - 1).$$

(ii) D'après la loi des grands nombres, $\frac{N_i^n}{n} \rightarrow p_i$. Si " $p \neq \pi$ ", alors il existe $k \in \llbracket 1, d \rrbracket$ tel que $p_k \neq \pi_k$. On en déduit que

$$T_n \geq n \left(\frac{N_k^n}{n} - \pi_k \right)^2 \xrightarrow[n \rightarrow \infty]{p.s} \infty.$$

□

Mise en place du test.

- Soit α un seuil (par exemple 5%).
- Soit η_α un réel tel que

$$\mathbb{P}(\chi^2(d - 1) > \eta_\alpha) \leq \alpha.$$

- On calcule T_n sur les données x_i :

$$T_n = \frac{n}{\pi_i} \sum_{i=1}^d \left(\frac{N_i^n}{n} - \pi \right)^2.$$

- Si $T_n = T_n(x_i)$

6.2.3 Espérances conditionnelles gaussiennes

Rappel. Si X, Y est un couple de variable aléatoires, l'espérance conditionnelle $\mathbb{E}[X|Y]$ est l'espérance sous la loi conditionnelle de X sachant Y . Par exemple, si X, Y sont discrètes, i.e à valeurs dans des univers finis qu'on note $\{x_1, \dots, x_n\} \times \{y_1, \dots, y_m\}$. Alors, pour tout $k \in \llbracket 1, m \rrbracket$,

$$\mathbb{E}[X|Y = y_k] = \sum_{i=1}^n x_i \mathbb{P}(X = x_i | Y = y_k).$$

Autre exemple, si (X, Y) est a pour densité $f_{(X,Y)}$, alors la loi conditionnelle de X sachant $\{Y = y_0\}$ a pour densité

$$f_{X|Y=y_0}(x) = \frac{f_{(X,Y)}(x, y_0)}{\int_{\mathbb{R}} f_{(X,Y)}(z, y) dz}$$

et

$$\mathbb{E}[X|Y = y_0] = \frac{\int_{\mathbb{R}} x f_{(X,Y)}(x, y_0) dx}{\int_{\mathbb{R}} f_{(X,Y)}(x, y_0) dx}.$$

Ici, le calcul d'espérances conditionnelles relève purement de l'analyse et du calcul intégral. La proposition suivante va nous montrer que dans le cadre des vecteurs gaussiens, l'algèbre domine.

Proposition 6.2.1. *Si $(X_1, \dots, X_n, Y_1, \dots, Y_m)$ est un vecteur gaussien. Alors, pour tout $i \in \llbracket 1, n \rrbracket$, $\mathbb{E}[X_i|Y_1, \dots, Y_m]$ est la projection orthogonale (au sens L^2) de X_i sur l'espace vectoriel $V = \text{Vect}(1, Y_1, \dots, Y_m)$: il existe $\lambda_0, \dots, \lambda_m \in \mathbb{R}$, tel que*

$$\mathbb{E}[X_i|Y_1, \dots, Y_m] = \lambda_0 + \sum_{j=1}^m \lambda_j Y_j.$$

En particulier, on peut expliciter ces coefficients : si l'on note $Z = (X_i, Y_1, \dots, Y_m)^T$, alors ce vecteur gaussien a pour matrice de covariance

$$\text{cov}(Z) = \left(\begin{array}{c|c} \Gamma_{X_i} & \Gamma_{X_i, Y}^T \\ \hline \Gamma_{X_i, Y} & \Gamma_Y \end{array} \right)$$

où $\Gamma_{X_i} = \text{var}(X_i)$, $\Gamma_Y = \text{cov} Y$, et $\Gamma_{X_i, Y}^T = (\text{cov}(X_i, Y_1), \dots, \text{cov}(X_i, Y_m))$ (vecteur ligne de taille m). Alors,

$$\mathbb{E}[X_i|Y_1, \dots, Y_m] = \mathbb{E}[X_i] + \Gamma_{X_i, Y}^T \Gamma_Y^{-1} \begin{pmatrix} Y_1 - \mathbb{E}[Y_1] \\ \vdots \\ Y_m - \mathbb{E}[Y_m] \end{pmatrix}.$$

Exemples.

— *En dimension 2.* Si (X, Y) est un vecteur gaussien suivant une loi $\mathcal{N}_2((1, 5)^T, K)$ où

$$K = \begin{pmatrix} 3 & 1 \\ 1 & 2 \end{pmatrix},$$

alors, $\mathbb{E}[X|Y] = 1 + \frac{1}{2}(Y - 5)$.

— *En dimension 3.* Si $(X, Y_1, Y_2) \sim \mathcal{N}_3(m, K)$ avec $m = (3, 2, 1)^T$, et

$$K = \begin{pmatrix} 2 & -2 & 2 \\ -2 & 5 & 1 \\ 2 & 1 & 5 \end{pmatrix},$$

alors,

$$\mathbb{E}[X|Y_1, Y_2] = 3 + (-2, 2) \begin{pmatrix} 5 & 1 \\ 1 & 5 \end{pmatrix}^{-1} \begin{pmatrix} Y_1 - 2 \\ Y_2 - 1 \end{pmatrix} = \frac{7}{2} + \frac{1}{2}(Y_2 - Y_1).$$