

Probabilités

Notes de Cours

Licence L3 — 2025–2026

Yaë Ulrich Gaba

*« La théorie des probabilités n'est au
fond que le bon sens réduit en calcul. »*

— Pierre-Simon Laplace

Table des matières

Préface	1
1 Espaces Probabilisés	5
1.1 Expérience aléatoire et univers	5
1.2 Tribus et événements	6
1.3 Axiomes de Kolmogorov	7
1.4 Propriétés de la mesure de probabilité	7
1.5 Continuité de la mesure de probabilité	8
1.6 Construction de probabilités	8
1.6.1 Probabilité uniforme sur un ensemble fini	8
1.6.2 Probabilité définie par une densité discrète	9
1.6.3 Construction sur $\mathbb{R}$: densité continue	9
1.7 Lemme de Borel–Cantelli (première partie)	9
1.8 Exercices	10
2 Probabilité Conditionnelle et Indépendance	11
2.1 Probabilité conditionnelle	11
2.2 Règle de la chaîne et formule de Bayes	12
2.3 Indépendance d’événements	13
2.4 Second lemme de Borel–Cantelli	14
2.5 Indépendance de tribus et de variables aléatoires	14
2.6 Applications classiques	14
2.7 Exercices	15
3 Variables Aléatoires Discrètes	17
3.1 Définition et loi d’une variable aléatoire	17
3.2 Loïs discrètes classiques	18
3.2.1 Loi de Bernoulli	18
3.2.2 Loi binomiale	18
3.2.3 Loi géométrique	18
3.2.4 Loi de Poisson	19
3.2.5 Loi hypergéométrique	19
3.3 Tableau récapitulatif	19
3.4 Fonction génératrice	20
3.5 Exercices	20

4	Variables Aléatoires Continues	23
4.1	Densité de probabilité	23
4.2	Fonction de répartition	24
4.3	Lois continues classiques	24
4.3.1	Loi uniforme continue	24
4.3.2	Loi exponentielle	24
4.3.3	Loi normale (gaussienne)	25
4.3.4	Loi gamma	25
4.3.5	Loi bêta	25
4.4	Quantiles et médiane	26
4.5	Simulation par inversion	26
4.6	Exercices	26
5	Espérance, Variance, Moments	27
5.1	Espérance	27
5.2	Variance	28
5.3	Moments d'ordre supérieur	29
5.4	Inégalités fondamentales	29
5.5	Espérance conditionnelle (introduction)	30
5.6	Exercices	30
6	Lois Usuelles et leurs Propriétés	31
6.1	Panorama des lois discrètes	31
6.1.1	Loi binomiale négative	31
6.1.2	Loi uniforme discrète	31
6.1.3	Loi de Poisson — compléments	32
6.2	Panorama des lois continues	32
6.2.1	Loi log-normale	32
6.2.2	Loi de Cauchy	32
6.2.3	Loi de Weibull	32
6.2.4	Loi chi-deux	32
6.2.5	Loi de Student	33
6.2.6	Loi de Fisher	33
6.3	Relations entre les lois	33
6.4	Tableau récapitulatif général	33
6.5	Méthodes de reconnaissance	34
6.6	Exercices	34
7	Fonctions de Variables Aléatoires	35
7.1	Cas discret	35
7.2	Méthode de la fonction de répartition	35
7.3	Méthode du changement de variable	36
7.4	Cas non monotone	36
7.5	Fonction génératrice de moments	37
7.6	Transformation de probabilité intégrale	37
7.7	Exercices	37

8	Couples de Variables Aléatoires	39
8.1	Loi conjointe	39
8.1.1	Cas discret	39
8.1.2	Cas continu	39
8.2	Lois marginales	40
8.3	Loi conditionnelle	40
8.4	Indépendance	40
8.5	Covariance et corrélation	41
8.6	Loi normale bivariée	41
8.7	Somme de variables aléatoires	42
8.8	Exercices	42
9	Lois des Grands Nombres	43
9.1	Modes de convergence	43
9.2	Loi faible des grands nombres	44
9.3	Loi forte des grands nombres	44
9.4	Applications	45
9.4.1	Interprétation fréquentiste	45
9.4.2	Méthode de Monte Carlo	45
9.4.3	Théorème de Glivenko–Cantelli	45
9.5	Convergence en loi (introduction)	45
9.6	Exercices	46
10	Théorème Central Limite	47
10.1	Énoncé du théorème	47
10.2	Démonstration par les fonctions caractéristiques	48
10.3	Applications pratiques	48
10.3.1	Approximation de la loi binomiale	48
10.3.2	Intervalles de confiance	49
10.3.3	Approximation de la loi de Poisson	49
10.4	TCL multidimensionnel	49
10.5	Méthode delta	49
10.6	Vitesse de convergence : théorème de Berry–Esseen	49
10.7	Exercices	50
11	Fonctions Caractéristiques	51
11.1	Définition et premières propriétés	51
11.2	Calcul des moments	52
11.3	Fonctions caractéristiques des lois classiques	52
11.4	Théorèmes d’unicité et d’inversion	52
11.5	Théorème de continuité de Lévy	53
11.6	Application : preuve de stabilité de la loi normale	53
11.7	Application : stabilité de la loi de Poisson	53
11.8	Fonction caractéristique et convergence en loi	54
11.9	Exercices	54

12 Introduction aux Chaînes de Markov	55
12.1 Introduction	55
12.2 Définitions	55
12.3 Probabilités de transition en n pas	56
12.4 Classification des états	56
12.5 Distributions stationnaires	57
12.6 Théorème ergodique	57
12.7 Réversibilité et équations de bilan détaillé	58
12.8 Exercices	58

Préface

Objectifs du cours

Ce cours de **Théorie des Probabilités** s'adresse aux étudiants de deuxième année de licence (L2) en mathématiques, informatique ou sciences physiques. Il constitue une introduction rigoureuse à la théorie mathématique des probabilités, fondée sur l'axiomatique de Kolmogorov, tout en maintenant une forte intuition probabiliste.

Les probabilités occupent une place centrale dans les mathématiques modernes et leurs applications. De la physique statistique à l'apprentissage automatique, de la finance quantitative à la biologie moléculaire, la modélisation probabiliste est devenue un outil indispensable pour comprendre et quantifier l'incertitude.

Prérequis

Pour aborder ce cours dans de bonnes conditions, l'étudiant devra maîtriser :

- **Analyse réelle** : suites, séries, intégrales (Riemann), convergence uniforme, théorèmes de convergence dominée (notions de base).
- **Algèbre linéaire** : espaces vectoriels, matrices, valeurs propres (pour le chapitre sur les chaînes de Markov).
- **Théorie des ensembles** : opérations ensemblistes, dénombrabilité, notions de cardinalité.
- **Combinatoire** : arrangements, combinaisons, principe de multiplication.

Organisation du cours

Le cours est structuré en douze chapitres, organisés selon une progression logique :

Chapitres 1–2 : Fondements. On y définit l'espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$, les axiomes de Kolmogorov, la probabilité conditionnelle et la notion d'indépendance. Ces deux chapitres posent le cadre axiomatique sur lequel repose tout l'édifice.

Chapitres 3–5 : Variables aléatoires. On introduit les variables aléatoires discrètes puis continues, avec leurs lois, fonctions de répartition, densités. Le chapitre 5 développe les notions d'espérance, variance et moments, outils fondamentaux du calcul probabiliste.

Chapitres 6–8 : Lois et transformations. Le chapitre 6 présente un catalogue des lois classiques (binomiale, Poisson, normale, exponentielle, gamma...). Le chapitre 7 étudie les fonctions de variables aléatoires, et le chapitre 8 introduit les lois jointes, la covariance et la corrélation.

Chapitres 9–11 : Théorèmes limites. C'est le cœur de la théorie : la loi des grands nombres (faible et forte) au chapitre 9, le théorème central limite au chapitre 10, et les fonctions caractéristiques au chapitre 11, outil puissant qui permet de démontrer élégamment le TCL.

Chapitre 12 : Introduction aux chaînes de Markov. Ce dernier chapitre ouvre la porte aux processus stochastiques en étudiant les chaînes de Markov à espace d'états fini ou dénombrable.

Méthodologie et notation

Tout au long de ce cours, nous adoptons les conventions suivantes :

- Les **définitions** sont encadrées et numérotées. Chaque concept nouveau est d'abord défini rigoureusement avant d'être illustré par des exemples.
- Les **théorèmes** sont énoncés avec précision et, dans la mesure du possible, démontrés complètement. Lorsqu'une démonstration dépasse le cadre du cours, nous la remplaçons par une esquisse de preuve ou une référence.
- Les **exemples** sont nombreux et détaillés. Ils constituent un complément essentiel à la théorie abstraite.
- Les **exercices** en fin de section permettent de vérifier la compréhension et de développer les techniques de calcul.

Notations principales.

Symbole	Signification
Ω	Univers (ensemble des résultats possibles)
$\mathcal{F}$	Tribu (ensemble des événements)
$\mathbb{P}$	Mesure de probabilité
$\mathbb{E}[X]$	Espérance de la variable aléatoire X
$\text{Var}(X)$	Variance de X
$\text{Cov}(X, Y)$	Covariance de X et Y
F_X	Fonction de répartition de X
f_X	Densité de X (cas continu)
$\varphi_X(t)$	Fonction caractéristique de X
$X \sim \mathcal{L}$	X suit la loi $\mathcal{L}$
$\xrightarrow{\text{p.s.}}$	Convergence presque sûre
$\xrightarrow{\mathbb{P}}$	Convergence en probabilité
$\xrightarrow{\mathcal{L}}$	Convergence en loi

Perspective historique

La théorie des probabilités a une histoire riche, qui mérite d'être brièvement évoquée. Les premiers travaux remontent à la correspondance entre **Blaise Pascal** et **Pierre de Fermat** en 1654, à propos du problème des partis. Au XVIII^e siècle, **Jakob Bernoulli** démontre la première version de la loi des grands nombres dans son *Ars Conjectandi* (1713), tandis que **Abraham de Moivre** obtient une forme primitive du théorème central limite.

Au XIX^e siècle, **Pierre-Simon de Laplace** synthétise l'ensemble des connaissances probabilistes de son époque dans sa *Théorie analytique des probabilités* (1812). **Pafnouti Tchebychev**, **Andreï Markov** et **Aleksandr Liapounov** développent les outils analytiques qui mèneront aux théorèmes limites modernes.

La révolution décisive survient en 1933, quand **Andreï Kolmogorov** publie ses *Grundbegriffe der Wahrscheinlichkeitsrechnung*, où il fonde la théorie des probabilités sur la théorie de la mesure. Cette axiomatisation, que nous adoptons dans ce cours, a permis de résoudre de nombreux paradoxes et a ouvert la voie à la théorie moderne des processus stochastiques.

Conseils aux étudiants

1. **Travaillez régulièrement.** Les probabilités ne s'apprennent pas la veille de l'examen. Chaque chapitre s'appuie sur les précédents.
2. **Faites les exercices.** La compréhension passive de la théorie ne suffit pas. C'est en résolvant des problèmes que l'on développe l'intuition probabiliste.
3. **Dessinez.** Représentez les ensembles, tracez les densités, esquissez les fonctions de répartition. La visualisation est un allié précieux.

4. **Vérifiez les cas particuliers.** Lorsque vous obtenez une formule générale, testez-la sur des exemples simples. Cela permet de détecter les erreurs et de renforcer la compréhension.
5. **Distinguez le discret du continu.** De nombreuses erreurs proviennent de la confusion entre sommes et intégrales, entre loi de masse et densité. Soyez toujours conscient du cadre dans lequel vous travaillez.

Références principales

Ce cours s'inspire de nombreux ouvrages classiques, parmi lesquels :

- J. Jacod et P. Protter, *Probability Essentials*, Springer.
- P. Billingsley, *Probability and Measure*, Wiley.
- W. Feller, *An Introduction to Probability Theory and Its Applications*, volumes I et II, Wiley.
- J.-F. Le Gall, *Intégration, Probabilités et Processus Aléatoires*, École Normale Supérieure.
- D. Foata et A. Fuchs, *Calcul des Probabilités*, Dunod.
- A. Raugi, *Cours de probabilités*, Université de Rennes.
- G. Grimmett et D. Stirzaker, *Probability and Random Processes*, Oxford.

Bonne lecture et bon travail à tous.

Chapitre 1

Espaces Probabilisés

Le calcul des probabilités est né dans les tripots du XVII^e siècle. Blaise Pascal et Pierre de Fermat, dans leur célèbre correspondance de 1654, cherchaient à résoudre le « problème des partis » : comment répartir équitablement la mise d'un jeu interrompu ? Mais il a fallu près de trois siècles pour que la théorie des probabilités acquière des fondations mathématiques rigoureuses. C'est Andreï Kolmogorov qui, en 1933, dans ses *Grundbegriffe der Wahrscheinlichkeitsrechnung*, posa les axiomes définitifs : une probabilité n'est rien d'autre qu'une mesure de masse totale 1 sur un espace mesurable.

Ce premier chapitre construit ces fondations pierre par pierre : l'univers Ω des possibles, la tribu $\mathcal{F}$ des événements observables, et la mesure de probabilité $\mathbb{P}$ qui attribue à chaque événement un degré de vraisemblance.

Intuition

Ce premier chapitre pose les fondations mathématiques de la théorie des probabilités. Nous y définissons la notion d'espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$, présentons les axiomes de Kolmogorov et établissons les premières propriétés de la mesure de probabilité.

1.1 Expérience aléatoire et univers

Avant toute formalisation, posons la question naïve : qu'est-ce qu'une expérience aléatoire ? C'est une expérience dont nous ne pouvons pas prédire le résultat, mais dont nous pouvons décrire tous les résultats possibles. Un lancer de dé, le cours d'une action boursière demain, la durée de vie d'un transistor — autant de situations où l'incertitude est irréductible, mais où la structure de l'incertitude est accessible.

Définition 1.1 (Expérience aléatoire). Une **expérience aléatoire** est une expérience dont le résultat ne peut être prédit avec certitude avant sa réalisation, mais dont on peut décrire l'ensemble de tous les résultats possibles.

Définition 1.2 (Univers). L'**univers** (ou **espace fondamental**) est l'ensemble Ω de tous les résultats possibles d'une expérience aléatoire. Chaque élément $\omega \in \Omega$ est appelé un **résultat élémentaire** ou **épreuve**.

Exemple 1.3. 1. **Lancer d'un dé** : $\Omega = \{1, 2, 3, 4, 5, 6\}$.

2. **Lancer d'une pièce** : $\Omega = \{\text{Pile}, \text{Face}\}$.

3. **Durée de vie d'un composant** : $\Omega = [0, +\infty)$.

4. **Suite infinie de lancers de pièce** : $\Omega = \{0, 1\}^{\mathbb{N}} = \{(\omega_1, \omega_2, \dots) : \omega_i \in \{0, 1\}\}$.

Remarque 1.4. L'univers Ω peut être fini, dénombrable ou indénombrable. Le choix de Ω dépend de la modélisation : il n'est pas unique pour une expérience donnée.

1.2 Tribus et événements

Lorsque Ω est fini ou dénombrable, on peut souvent considérer tous les sous-ensembles de Ω comme des événements. Pour des espaces indénombrables (comme $\mathbb{R}$), il est nécessaire de restreindre la classe des ensembles « mesurables ».

Définition 1.5 (Tribu / σ -algèbre). Une **tribu** (ou σ -algèbre) sur Ω est une collection $\mathcal{F} \subseteq \mathcal{P}(\Omega)$ vérifiant :

- (i) $\Omega \in \mathcal{F}$;
- (ii) si $A \in \mathcal{F}$, alors $A^c \in \mathcal{F}$ (stabilité par complémentation) ;
- (iii) si $(A_n)_{n \geq 1}$ est une suite d'éléments de $\mathcal{F}$, alors $\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$ (stabilité par union dénombrable).

Proposition 1.6 (Propriétés immédiates). Soit $\mathcal{F}$ une tribu sur Ω . Alors :

- (a) $\emptyset \in \mathcal{F}$;
- (b) $\mathcal{F}$ est stable par intersection dénombrable : si $A_n \in \mathcal{F}$ pour tout n , alors $\bigcap_{n=1}^{\infty} A_n \in \mathcal{F}$;
- (c) $\mathcal{F}$ est stable par différence : si $A, B \in \mathcal{F}$, alors $A \setminus B \in \mathcal{F}$.

Démonstration. (a) $\emptyset = \Omega^c \in \mathcal{F}$ par (i) et (ii). (b) Par les lois de De Morgan, $\bigcap A_n = (\bigcup A_n^c)^c$, ce qui est dans $\mathcal{F}$ par (ii) et (iii). (c) $A \setminus B = A \cap B^c \in \mathcal{F}$ par (ii) et (b). $\square$

Définition 1.7 (Événement). Un **événement** est un élément de la tribu $\mathcal{F}$, c'est-à-dire un sous-ensemble de Ω auquel on peut attribuer une probabilité. On dit que l'événement A **se réalise** si le résultat ω de l'expérience appartient à A .

Exemple 1.8 (Tribus fondamentales). 1. **Tribu grossière** : $\mathcal{F} = \{\emptyset, \Omega\}$. C'est la plus petite tribu sur Ω .

2. **Tribu discrète** : $\mathcal{F} = \mathcal{P}(\Omega)$. C'est la plus grande tribu sur Ω .

3. **Tribu engendrée** : si $\mathcal{C}$ est une collection de sous-ensembles de Ω , la tribu engendrée $\sigma(\mathcal{C})$ est la plus petite tribu contenant $\mathcal{C}$.

Définition 1.9 (Tribu borélienne). La **tribu borélienne** de $\mathbb{R}$, notée $\mathcal{B}(\mathbb{R})$, est la tribu engendrée par les ouverts de $\mathbb{R}$:

$$\mathcal{B}(\mathbb{R}) = \sigma(\{O \subseteq \mathbb{R} : O \text{ ouvert}\}).$$

On montre que $\mathcal{B}(\mathbb{R}) = \sigma(\{]-\infty, a] : a \in \mathbb{R}\}) = \sigma(\{]a, b[: a < b\})$.

1.3 Axiomes de Kolmogorov

Définition 1.10 (Mesure de probabilité). Soit $(\Omega, \mathcal{F})$ un espace mesurable. Une **mesure de probabilité** est une application $\mathbb{P} : \mathcal{F} \rightarrow [0, 1]$ vérifiant les **axiomes de Kolmogorov** :

(K1) **Non-négativité** : pour tout $A \in \mathcal{F}$, $\mathbb{P}(A) \geq 0$.

(K2) **Normalisation** : $\mathbb{P}(\Omega) = 1$.

(K3) **σ -additivité** : pour toute suite $(A_n)_{n \geq 1}$ d'événements *mutuellement disjoints* (i.e. $A_i \cap A_j = \emptyset$ pour $i \neq j$) :

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mathbb{P}(A_n).$$

Définition 1.11 (Espace probabilisé). Le triplet $(\Omega, \mathcal{F}, \mathbb{P})$ est appelé **espace probabilisé**.

Additivité finie vs. σ -additivité

L'additivité finie ($\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$ pour $A \cap B = \emptyset$) découle de (K3). L'axiome de σ -additivité est plus fort : il garantit la cohérence du passage à la limite et est essentiel pour les théorèmes limites (chapitres 9–11).

1.4 Propriétés de la mesure de probabilité

Théorème 1.12 (Propriétés fondamentales). Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé. Pour tous $A, B \in \mathcal{F}$:

(i) $\mathbb{P}(\emptyset) = 0$.

(ii) $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$.

(iii) Si $A \subseteq B$, alors $\mathbb{P}(A) \leq \mathbb{P}(B)$ (*monotonie*).

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

(v) $\mathbb{P}(A) \leq 1$.

Démonstration. (i) Posons $A_1 = \Omega$ et $A_n = \emptyset$ pour $n \geq 2$. La suite est disjointe, et $\bigcup A_n = \Omega$. Par (K3) : $1 = \mathbb{P}(\Omega) = \mathbb{P}(\Omega) + \sum_{n \geq 2} \mathbb{P}(\emptyset)$. Donc $\mathbb{P}(\emptyset) = 0$.

(ii) $\Omega = A \cup A^c$ (union disjointe), donc $1 = \mathbb{P}(A) + \mathbb{P}(A^c)$.

(iii) $B = A \cup (B \setminus A)$ (union disjointe), donc $\mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(B \setminus A) \geq \mathbb{P}(A)$.

(iv) $A \cup B = A \cup (B \setminus A)$ et $B = (A \cap B) \cup (B \setminus A)$, d'où $\mathbb{P}(B \setminus A) = \mathbb{P}(B) - \mathbb{P}(A \cap B)$ et le résultat suit.

(v) Conséquence de (ii) : $\mathbb{P}(A) = 1 - \mathbb{P}(A^c) \leq 1$.

□

Théorème 1.13 (Formule d'inclusion-exclusion). *Pour tous événements $A_1, \dots, A_n \in \mathcal{F}$:*

$$\mathbb{P}\left(\bigcup_{i=1}^n A_i\right) = \sum_i \mathbb{P}(A_i) - \sum_{i < j} \mathbb{P}(A_i \cap A_j) + \sum_{i < j < k} \mathbb{P}(A_i \cap A_j \cap A_k) - \dots + (-1)^{n+1} \mathbb{P}(A_1 \cap \dots \cap A_n).$$

Corollaire 1.14 (Inégalité de Boole / sous-additivité). *Pour toute suite $(A_n)_{n \geq 1}$ d'événements :*

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) \leq \sum_{n=1}^{\infty} \mathbb{P}(A_n).$$

1.5 Continuité de la mesure de probabilité

Théorème 1.15 (Continuité monotone). *Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé.*

(i) **Continuité croissante** : si $A_1 \subseteq A_2 \subseteq \dots$ (suite croissante), alors

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

(ii) **Continuité décroissante** : si $A_1 \supseteq A_2 \supseteq \dots$ (suite décroissante), alors

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

Démonstration. (i) Posons $B_1 = A_1$ et $B_n = A_n \setminus A_{n-1}$ pour $n \geq 2$. Les B_n sont mutuellement disjoints et $\bigcup_{n=1}^N B_n = A_N$. Par σ -additivité :

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \mathbb{P}\left(\bigcup_{n=1}^{\infty} B_n\right) = \sum_{n=1}^{\infty} \mathbb{P}(B_n) = \lim_{N \rightarrow \infty} \sum_{n=1}^N \mathbb{P}(B_n) = \lim_{N \rightarrow \infty} \mathbb{P}(A_N).$$

(ii) Appliquer (i) à la suite croissante $A_1^c \subseteq A_2^c \subseteq \dots$ et utiliser $\mathbb{P}(A_n^c) = 1 - \mathbb{P}(A_n)$. $\square$

1.6 Construction de probabilités

1.6.1 Probabilité uniforme sur un ensemble fini

Définition 1.16 (Probabilité uniforme). Si Ω est un ensemble fini, la **probabilité uniforme** (ou équiprobabilité) est définie par :

$$\mathbb{P}(A) = \frac{|A|}{|\Omega|} = \frac{\text{nombre de cas favorables}}{\text{nombre de cas possibles}}.$$

Quand utiliser l'équiprobabilité ?

L'hypothèse d'équiprobabilité n'est légitime que si la symétrie du problème la justifie (dé équilibré, tirage au sort...). Ne l'invoquez jamais sans vérification !

Exemple 1.17 (Problème des anniversaires). Dans un groupe de n personnes, quelle est la probabilité qu'au moins deux personnes aient le même anniversaire ? On suppose 365 jours équiprobables.

L'univers est $\Omega = \{1, \dots, 365\}^n$ avec $|\Omega| = 365^n$. L'événement complémentaire « tous les anniversaires sont distincts » a pour cardinal $365 \times 364 \times \dots \times (365 - n + 1)$. Donc :

$$\mathbb{P}(\text{au moins une coïncidence}) = 1 - \frac{365!}{(365 - n)! 365^n}.$$

Pour $n = 23$, on obtient $\mathbb{P} \approx 0,507$: il y a plus d'une chance sur deux !

1.6.2 Probabilité définie par une densité discrète

Proposition 1.18. Si Ω est dénombrable et si $(p_\omega)_{\omega \in \Omega}$ est une famille de réels tels que $p_\omega \geq 0$ pour tout ω et $\sum_{\omega \in \Omega} p_\omega = 1$, alors :

$$\mathbb{P}(A) = \sum_{\omega \in A} p_\omega, \quad A \subseteq \Omega,$$

définit une mesure de probabilité sur $(\Omega, \mathcal{P}(\Omega))$.

1.6.3 Construction sur $\mathbb{R}$: densité continue

Proposition 1.19. Soit $f : \mathbb{R} \rightarrow [0, +\infty)$ une fonction intégrable telle que $\int_{-\infty}^{+\infty} f(x) dx = 1$. Alors :

$$\mathbb{P}(A) = \int_A f(x) dx, \quad A \in \mathcal{B}(\mathbb{R}),$$

définit une mesure de probabilité sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. La fonction f est appelée **densité de probabilité**.

1.7 Lemme de Borel–Cantelli (première partie)

Théorème 1.20 (Premier lemme de Borel–Cantelli). Soit $(A_n)_{n \geq 1}$ une suite d'événements. Si $\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty$, alors :

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

où $\limsup_{n \rightarrow \infty} A_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k$ est l'événement « une infinité de A_n se réalisent ».

Démonstration. Posons $B_n = \bigcup_{k=n}^{\infty} A_k$. La suite (B_n) est décroissante et $\limsup A_n = \bigcap B_n$. Par continuité décroissante :

$$\mathbb{P}(\limsup A_n) = \lim_{n \rightarrow \infty} \mathbb{P}(B_n) \leq \lim_{n \rightarrow \infty} \sum_{k=n}^{\infty} \mathbb{P}(A_k) = 0,$$

car le reste d'une série convergente tend vers zéro. □

Remarque 1.21. Le **second** lemme de Borel–Cantelli (réciproque partielle) nécessite l'indépendance des événements et sera énoncé au chapitre 2.

1.8 Exercices

Exercice 1.1. Montrer que l'intersection de deux tribus est une tribu. L'union de deux tribus est-elle toujours une tribu ? Donner un contre-exemple.

Exercice 1.2. Soit $\Omega = \{1, 2, 3, 4, 5, 6\}$ (dé équilibré). Écrire la tribu engendrée par la partition $\{\{1, 2\}, \{3, 4\}, \{5, 6\}\}$ et dénombrer ses éléments.

Exercice 1.3. On lance un dé équilibré deux fois. Calculer la probabilité que :

- (a) la somme soit égale à 7 ;
- (b) la somme soit supérieure ou égale à 10 ;
- (c) les deux faces soient identiques.

Exercice 1.4. On tire trois cartes successivement et sans remise dans un jeu de 52 cartes. Calculer la probabilité d'obtenir exactement deux as.

Exercice 1.5. Soit (A_n) une suite décroissante d'événements avec $\mathbb{P}(A_n) = 1/n$. Que vaut $\mathbb{P}(\bigcap_{n=1}^{\infty} A_n)$? Justifier.

Exercice 1.6 (Lemme de Borel–Cantelli). Soit (A_n) une suite d'événements avec $\mathbb{P}(A_n) = 1/n^2$. Montrer que $\mathbb{P}(\limsup A_n) = 0$. Même question avec $\mathbb{P}(A_n) = 1/n$: peut-on conclure ?

Exercice 1.7. Montrer la formule d'inclusion-exclusion (théorème 1.13) par récurrence sur n .

Exercice 1.8. On dispose n lettres dans n enveloppes au hasard (une lettre par enveloppe). Soit D_n le nombre de lettres « bien placées » (dans la bonne enveloppe). En utilisant la formule d'inclusion-exclusion, montrer que :

$$\mathbb{P}(D_n = 0) = \sum_{k=0}^n \frac{(-1)^k}{k!} \xrightarrow{n \rightarrow \infty} e^{-1} \approx 0,368.$$

Chapitre 2

Probabilité Conditionnelle et Indépendance

Un médecin apprend qu'un patient a un test positif pour une maladie rare. Quelle est la probabilité que le patient soit réellement malade ? La réponse — souvent contre-intuitive — dépend crucialement de la *prévalence* de la maladie et de la qualité du test. C'est le célèbre « paradoxe » de Bayes, et pour le résoudre, il faut maîtriser la notion de *probabilité conditionnelle* : comment mettre à jour nos croyances lorsqu'une information nouvelle est disponible.

Intuition

La notion de probabilité conditionnelle permet de mettre à jour l'information disponible lorsqu'un événement est observé. L'indépendance, quant à elle, formalise l'idée que deux événements n'ont aucune influence l'un sur l'autre. Ces deux concepts sont au cœur de toute la théorie des probabilités.

2.1 Probabilité conditionnelle

Définition 2.1 (Probabilité conditionnelle). Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé et $B \in \mathcal{F}$ un événement tel que $\mathbb{P}(B) > 0$. La **probabilité conditionnelle** de A sachant B est :

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

Proposition 2.2. Pour B fixé avec $\mathbb{P}(B) > 0$, l'application $A \mapsto \mathbb{P}(A \mid B)$ est une mesure de probabilité sur $(\Omega, \mathcal{F})$.

Démonstration. Vérifions les axiomes de Kolmogorov :

- $\mathbb{P}(A \mid B) = \mathbb{P}(A \cap B)/\mathbb{P}(B) \geq 0$ pour tout A .
- $\mathbb{P}(\Omega \mid B) = \mathbb{P}(B)/\mathbb{P}(B) = 1$.
- Si (A_n) sont mutuellement disjoints, alors les $(A_n \cap B)$ le sont aussi, et par σ -additivité de $\mathbb{P}$:

$$\mathbb{P}\left(\bigcup_n A_n \mid B\right) = \frac{\mathbb{P}(\bigcup_n (A_n \cap B))}{\mathbb{P}(B)} = \frac{\sum_n \mathbb{P}(A_n \cap B)}{\mathbb{P}(B)} = \sum_n \mathbb{P}(A_n \mid B). \quad \square$$

Exemple 2.3. On lance deux dés équilibrés. Sachant que la somme est 8, quelle est la probabilité que le premier dé affiche 3 ?

Soit $A = \{\text{premier dé} = 3\}$ et $B = \{\text{somme} = 8\}$. Les résultats donnant somme 8 sont : $(2, 6), (3, 5), (4, 4), (5, 3), (6, 2)$, donc $\mathbb{P}(B) = 5/36$. Comme $A \cap B = \{(3, 5)\}$, on a $\mathbb{P}(A \cap B) = 1/36$. Donc $\mathbb{P}(A | B) = (1/36)/(5/36) = 1/5$.

2.2 Règle de la chaîne et formule de Bayes

Théorème 2.4 (Règle de multiplication / règle de la chaîne). *Pour tous événements $A_1, \dots, A_n$ avec $\mathbb{P}(A_1 \cap \dots \cap A_{n-1}) > 0$:*

$$\mathbb{P}(A_1 \cap \dots \cap A_n) = \mathbb{P}(A_1) \mathbb{P}(A_2 | A_1) \mathbb{P}(A_3 | A_1 \cap A_2) \cdots \mathbb{P}(A_n | A_1 \cap \dots \cap A_{n-1}).$$

Définition 2.5 (Système complet d'événements). Une famille $(B_i)_{i \in I}$ d'événements forme un **système complet d'événements** (ou partition de Ω) si :

- (i) $B_i \cap B_j = \emptyset$ pour $i \neq j$ (disjoints) ;
- (ii) $\bigcup_{i \in I} B_i = \Omega$ (exhaustifs) ;
- (iii) $\mathbb{P}(B_i) > 0$ pour tout i .

Théorème 2.6 (Formule des probabilités totales). *Soit $(B_i)_{i \in I}$ un système complet d'événements (avec I fini ou dénombrable). Pour tout $A \in \mathcal{F}$:*

$$\mathbb{P}(A) = \sum_{i \in I} \mathbb{P}(A | B_i) \mathbb{P}(B_i).$$

Démonstration. Comme $A = A \cap \Omega = \bigcup_i (A \cap B_i)$ (union disjointe) :

$$\mathbb{P}(A) = \sum_i \mathbb{P}(A \cap B_i) = \sum_i \mathbb{P}(A | B_i) \mathbb{P}(B_i). \quad \square$$

Théorème 2.7 (Formule de Bayes). *Soit $(B_i)_{i \in I}$ un système complet d'événements et A un événement avec $\mathbb{P}(A) > 0$. Alors pour tout $j \in I$:*

$$\mathbb{P}(B_j | A) = \frac{\mathbb{P}(A | B_j) \mathbb{P}(B_j)}{\sum_{i \in I} \mathbb{P}(A | B_i) \mathbb{P}(B_i)}.$$

Interprétation bayésienne

Les $\mathbb{P}(B_i)$ sont les probabilités *a priori* des hypothèses B_i . Après observation de A , la formule de Bayes fournit les probabilités *a posteriori* $\mathbb{P}(B_i | A)$, mises à jour par la vraisemblance $\mathbb{P}(A | B_i)$.

Exemple 2.8 (Test de dépistage). Un test médical a une sensibilité de 99% (vrai positif) et une spécificité de 95% (vrai négatif). La prévalence de la maladie est de 0,1%. Quelle est la probabilité qu'un patient soit malade sachant que le test est positif ?

Soit M = « malade » et T^+ = « test positif ». On a : $\mathbb{P}(M) = 0,001$, $\mathbb{P}(T^+ | M) = 0,99$, $\mathbb{P}(T^+ | M^c) = 0,05$.

$$\begin{aligned}\mathbb{P}(M | T^+) &= \frac{\mathbb{P}(T^+ | M) \mathbb{P}(M)}{\mathbb{P}(T^+ | M) \mathbb{P}(M) + \mathbb{P}(T^+ | M^c) \mathbb{P}(M^c)} \\ &= \frac{0,99 \times 0,001}{0,99 \times 0,001 + 0,05 \times 0,999} = \frac{0,00099}{0,05094} \approx 0,0194.\end{aligned}$$

Même avec un test très performant, la probabilité d'être malade sachant un test positif n'est que d'environ 2%, à cause de la faible prévalence.

2.3 Indépendance d'événements

Définition 2.9 (Indépendance de deux événements). Deux événements A et B sont **indépendants**, noté $A \perp\!\!\!\perp B$, si :

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

Remarque 2.10. Si $\mathbb{P}(B) > 0$, l'indépendance équivaut à $\mathbb{P}(A | B) = \mathbb{P}(A)$: savoir que B s'est réalisé ne modifie pas la probabilité de A .

Indépendance $\neq$ incompatibilité

Deux événements disjoints ($A \cap B = \emptyset$) de probabilité non nulle ne sont *jamais* indépendants, car $\mathbb{P}(A \cap B) = 0 \neq \mathbb{P}(A)\mathbb{P}(B)$. Ne confondez pas ces deux notions !

Définition 2.11 (Indépendance mutuelle). Les événements $A_1, \dots, A_n$ sont **mutuellement indépendants** si, pour tout sous-ensemble $J \subseteq \{1, \dots, n\}$ avec $|J| \geq 2$:

$$\mathbb{P}\left(\bigcap_{j \in J} A_j\right) = \prod_{j \in J} \mathbb{P}(A_j).$$

Indépendance deux à deux $\neq$ mutuelle

L'indépendance deux à deux ne suffit pas pour l'indépendance mutuelle. Il faut vérifier *toutes* les sous-familles, pas seulement les paires.

Exemple 2.12 (Contre-exemple de Bernstein). On lance deux dés équilibrés. Posons :

- A = « le premier dé est pair »,
- B = « le second dé est pair »,
- C = « la somme est paire ».

On vérifie que $\mathbb{P}(A) = \mathbb{P}(B) = \mathbb{P}(C) = 1/2$ et que toute paire est indépendante : $\mathbb{P}(A \cap B) = 1/4 = \mathbb{P}(A)\mathbb{P}(B)$, etc. Cependant, $\mathbb{P}(A \cap B \cap C) = 1/4 \neq 1/8 = \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C)$. Les trois événements sont indépendants deux à deux mais pas mutuellement.

Proposition 2.13. Si A et B sont indépendants, alors les couples suivants le sont aussi : (A, B^c) , (A^c, B) , (A^c, B^c) .

Démonstration. $\mathbb{P}(A \cap B^c) = \mathbb{P}(A) - \mathbb{P}(A \cap B) = \mathbb{P}(A) - \mathbb{P}(A)\mathbb{P}(B) = \mathbb{P}(A)(1 - \mathbb{P}(B)) = \mathbb{P}(A)\mathbb{P}(B^c)$. Les autres cas sont analogues. $\square$

2.4 Second lemme de Borel–Cantelli

Théorème 2.14 (Second lemme de Borel–Cantelli). *Soit $(A_n)_{n \geq 1}$ une suite d'événements mutuellement indépendants. Si $\sum_{n=1}^{\infty} \mathbb{P}(A_n) = +\infty$, alors :*

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

Autrement dit, presque sûrement, une infinité de A_n se réalisent.

Démonstration. Il suffit de montrer que $\mathbb{P}(\bigcup_{k=n}^{\infty} A_k) = 1$ pour tout n . Passant au complémentaire :

$$\mathbb{P}\left(\bigcap_{k=n}^N A_k^c\right) = \prod_{k=n}^N \mathbb{P}(A_k^c) = \prod_{k=n}^N (1 - \mathbb{P}(A_k)) \leq \prod_{k=n}^N e^{-\mathbb{P}(A_k)} = \exp\left(-\sum_{k=n}^N \mathbb{P}(A_k)\right).$$

Quand $N \rightarrow \infty$, la somme diverge, donc le produit tend vers 0. Par continuité décroissante : $\mathbb{P}(\bigcap_{k=n}^{\infty} A_k^c) = 0$, d'où $\mathbb{P}(\bigcup_{k=n}^{\infty} A_k) = 1$. $\square$

Résumé : Borel–Cantelli

- $\sum \mathbb{P}(A_n) < \infty \implies \mathbb{P}(\limsup A_n) = 0$ (sans hypothèse).
- $\sum \mathbb{P}(A_n) = \infty$ et indépendance $\implies \mathbb{P}(\limsup A_n) = 1$.

2.5 Indépendance de tribus et de variables aléatoires

Définition 2.15 (Tribus indépendantes). Deux sous-tribus $\mathcal{G}_1, \mathcal{G}_2 \subseteq \mathcal{F}$ sont **indépendantes** si, pour tous $A \in \mathcal{G}_1$ et $B \in \mathcal{G}_2$:

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

Définition 2.16 (Variables aléatoires indépendantes — préambule). Deux variables aléatoires X et Y sont **indépendantes** si les tribus qu'elles engendrent, $\sigma(X)$ et $\sigma(Y)$, sont indépendantes. En termes de lois (voir chapitre 3 pour les définitions complètes), cela revient à :

$$\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A) \mathbb{P}(Y \in B)$$

pour tous boréliens $A, B \in \mathcal{B}(\mathbb{R})$.

2.6 Applications classiques

Exemple 2.17 (Problème de Monty Hall). Un candidat choisit une porte parmi trois. Derrière l'une se trouve une voiture, derrière les deux autres des chèvres. Le présentateur, qui sait où est la voiture, ouvre une porte révélant une chèvre, puis propose au candidat de changer de porte.

Soit V_i = « la voiture est derrière la porte i » ($i = 1, 2, 3$), et supposons que le candidat choisisse la porte 1. Par la formule de Bayes, on montre que la probabilité de gagner en changeant est $2/3$, tandis que rester donne $1/3$. Il est donc avantageux de changer.

Exemple 2.18 (Marche aléatoire et ruine du joueur). Un joueur commence avec un capital $k \in \{1, \dots, N-1\}$. À chaque tour, il gagne 1 € avec probabilité p ou perd 1 € avec probabilité $q = 1 - p$, indépendamment. Le jeu s'arrête quand le capital atteint 0 (ruine) ou N (objectif atteint).

En utilisant la probabilité conditionnelle et un argument de récurrence, on montre que la probabilité de ruine est :

$$\mathbb{P}(\text{ruine} \mid X_0 = k) = \begin{cases} \frac{(q/p)^k - (q/p)^N}{1 - (q/p)^N} & \text{si } p \neq q, \\ 1 - \frac{k}{N} & \text{si } p = q = \frac{1}{2}. \end{cases}$$

2.7 Exercices

Exercice 2.1. Un lot contient 10 pièces dont 3 sont défectueuses. On tire 2 pièces successivement sans remise. Calculer la probabilité que la seconde soit défectueuse sachant que la première l'est.

Exercice 2.2. Trois machines M_1, M_2, M_3 produisent respectivement 50%, 30% et 20% de la production. Leurs taux de défaut sont 2%, 3% et 5%. Un article pris au hasard est défectueux. Quelle est la probabilité qu'il provienne de M_2 ?

Exercice 2.3. Montrer que si A, B, C sont mutuellement indépendants, alors $A \cup B$ et C sont indépendants.

Exercice 2.4. On lance une pièce équilibrée n fois. Soit A_k l'événement « le k -ième lancer donne Pile ». Montrer que les A_k sont mutuellement indépendants.

Exercice 2.5 (Formule de Bayes itérée). On dispose d'une urne contenant 1 boule rouge et 1 boule bleue. On tire une boule au hasard, on la remet et on ajoute une boule de la même couleur. Après n tirages, calculer la probabilité que toutes les boules tirées soient rouges.

Exercice 2.6. Soit $(A_n)_{n \geq 1}$ une suite d'événements indépendants avec $\mathbb{P}(A_n) = 1/(n+1)$. Montrer que $\mathbb{P}(\limsup A_n) = 1$. Interpréter.

Exercice 2.7 (Paradoxe de Simpson). Donner un exemple numérique où $\mathbb{P}(A \mid B) > \mathbb{P}(A \mid B^c)$ mais $\mathbb{P}(A \mid B \cap C) < \mathbb{P}(A \mid B^c \cap C)$ et $\mathbb{P}(A \mid B \cap C^c) < \mathbb{P}(A \mid B^c \cap C^c)$. Expliquer le phénomène.

Exercice 2.8. Soit $p \in (0, 1)$ et $(X_n)_{n \geq 1}$ une suite de lancers de Bernoulli indépendants de paramètre p . Montrer que $\mathbb{P}(\exists n : X_n = 1) = 1$.

Chapitre 3

Variables Aléatoires Discrètes

Lancer un dé, compter le nombre de clients dans une file d'attente, observer le nombre de désintégrations radioactives par seconde : dans chacune de ces situations, le hasard produit un résultat qu'on peut énumérer. Les variables aléatoires discrètes — celles qui ne prennent qu'un nombre fini ou dénombrable de valeurs — sont les premières que l'on rencontre, et les plus intuitives. Pourtant, elles réservent des surprises : la loi de Poisson, découverte par Siméon Denis Poisson en 1837 pour modéliser les erreurs judiciaires, réapparaît partout, des centres d'appels aux impacts de bombes V-2 sur Londres. La loi géométrique capture la patience du joueur obstiné, tandis que la loi binomiale, héritière directe des travaux de Jacob Bernoulli, reste le modèle de référence pour les expériences répétées.

Ce chapitre construit le cadre formel des variables aléatoires discrètes et développe les lois classiques qui peuplent toute la théorie des probabilités.

Intuition

Une variable aléatoire est une fonction qui associe une valeur numérique au résultat d'une expérience aléatoire. Ce chapitre traite le cas discret : l'ensemble des valeurs prises est fini ou dénombrable.

3.1 Définition et loi d'une variable aléatoire

Définition 3.1 (Variable aléatoire). Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé. Une **variable aléatoire réelle** est une application $X : \Omega \rightarrow \mathbb{R}$ mesurable, c'est-à-dire telle que pour tout $B \in \mathcal{B}(\mathbb{R})$:

$$X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\} \in \mathcal{F}.$$

Définition 3.2 (Variable aléatoire discrète). Une variable aléatoire X est **discrète** si elle prend ses valeurs dans un ensemble fini ou dénombrable $\{x_1, x_2, \dots\} \subseteq \mathbb{R}$.

Définition 3.3 (Loi de probabilité / fonction de masse). La **loi** (ou **distribution**) de X est la mesure de probabilité P_X sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ définie par $P_X(B) = \mathbb{P}(X \in B)$. Pour une variable discrète, la loi est entièrement déterminée par la **fonction de masse** :

$$p_X(x_k) = \mathbb{P}(X = x_k), \quad k = 1, 2, \dots$$

avec $p_X(x_k) \geq 0$ et $\sum_k p_X(x_k) = 1$.

Définition 3.4 (Fonction de répartition). La **fonction de répartition** de X est :

$$F_X(x) = \mathbb{P}(X \leq x) = \sum_{x_k \leq x} p_X(x_k), \quad x \in \mathbb{R}.$$

Proposition 3.5 (Propriétés de F_X). (i) F_X est croissante.

(ii) $\lim_{x \rightarrow -\infty} F_X(x) = 0$ et $\lim_{x \rightarrow +\infty} F_X(x) = 1$.

(iii) F_X est continue à droite.

(iv) $\mathbb{P}(X = a) = F_X(a) - F_X(a^-)$ (saut en a).

(v) Pour une variable discrète, F_X est une fonction en escalier.

3.2 Lois discrètes classiques

3.2.1 Loi de Bernoulli

Définition 3.6 (Bernoulli). $X \sim \text{Bernoulli}(p)$ avec $p \in [0, 1]$:

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

$$\mathbb{E}[X] = p, \quad \text{Var}(X) = pq.$$

3.2.2 Loi binomiale

Définition 3.7 (Binomiale). $X \sim \text{Bin}(n, p)$: nombre de succès en n essais de Bernoulli indépendants.

$$\mathbb{P}(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}, \quad k = 0, 1, \dots, n.$$

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

Exemple 3.8. On lance 10 fois une pièce équilibrée. La probabilité d'obtenir exactement 6 Pile est :

$$\mathbb{P}(X = 6) = \binom{10}{6} \left(\frac{1}{2}\right)^{10} = \frac{210}{1024} \approx 0,205.$$

3.2.3 Loi géométrique

Définition 3.9 (Géométrique). $X \sim \text{Geom}(p)$: nombre d'essais nécessaires pour obtenir le premier succès.

$$\mathbb{P}(X = k) = (1 - p)^{k-1} p, \quad k = 1, 2, 3, \dots$$

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

Proposition 3.10 (Perte de mémoire). La loi géométrique est la seule loi discrète à support dans $\mathbb{N}^*$ vérifiant la propriété de **perte de mémoire** :

$$\mathbb{P}(X > m + n \mid X > m) = \mathbb{P}(X > n) \quad \forall m, n \geq 0.$$

Démonstration. $\mathbb{P}(X > k) = (1 - p)^k$. Donc :

$$\mathbb{P}(X > m + n \mid X > m) = \frac{(1 - p)^{m+n}}{(1 - p)^m} = (1 - p)^n = \mathbb{P}(X > n). \quad \square$$

3.2.4 Loi de Poisson

Définition 3.11 (Poisson). $X \sim \text{Poisson}(\lambda)$ avec $\lambda > 0$:

$$\mathbb{P}(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}, \quad k = 0, 1, 2, \dots$$

$$\mathbb{E}[X] = \lambda, \quad \text{Var}(X) = \lambda.$$

Théorème 3.12 (Approximation de Poisson). Si $X_n \sim \text{Bin}(n, p_n)$ avec $np_n \rightarrow \lambda$ quand $n \rightarrow \infty$, alors pour tout $k \in \mathbb{N}$:

$$\mathbb{P}(X_n = k) \xrightarrow{n \rightarrow \infty} e^{-\lambda} \frac{\lambda^k}{k!}.$$

Démonstration.

$$\begin{aligned} \mathbb{P}(X_n = k) &= \binom{n}{k} p_n^k (1 - p_n)^{n-k} \\ &= \frac{n!}{k!(n-k)!} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} \\ &= \frac{\lambda^k}{k!} \cdot \underbrace{\frac{n(n-1) \cdots (n-k+1)}{n^k}}_{\rightarrow 1} \cdot \underbrace{\left(1 - \frac{\lambda}{n}\right)^n}_{\rightarrow e^{-\lambda}} \cdot \underbrace{\left(1 - \frac{\lambda}{n}\right)^{-k}}_{\rightarrow 1} \xrightarrow{n \rightarrow \infty} e^{-\lambda} \frac{\lambda^k}{k!}. \quad \square \end{aligned}$$

Quand utiliser Poisson ?

La loi de Poisson modélise le nombre d'événements rares dans un intervalle donné (appels téléphoniques, particules radioactives, pannes...). Règle pratique : si $n \geq 30$ et $p \leq 0,1$, on approxime $\text{Bin}(n, p)$ par $\text{Poisson}(np)$.

3.2.5 Loi hypergéométrique

Définition 3.13 (Hypergéométrique). On tire n objets sans remise d'une population de N objets dont K sont marqués. X = nombre d'objets marqués parmi les n tirés :

$$\mathbb{P}(X = k) = \frac{\binom{K}{k} \binom{N-K}{n-k}}{\binom{N}{n}}, \quad k = \max(0, n + K - N), \dots, \min(n, K).$$

$$\mathbb{E}[X] = nK/N.$$

3.3 Tableau récapitulatif

Lois discrètes fondamentales

Loi	Paramètres	$\mathbb{E}[X]$	$\text{Var}(X)$
Bernoulli(p)	$p \in [0, 1]$	p	$p(1 - p)$
Binomiale(n, p)	$n \in \mathbb{N}^*, p \in [0, 1]$	np	$np(1 - p)$
Géométrique(p)	$p \in (0, 1]$	$1/p$	$(1 - p)/p^2$
Poisson(λ)	$\lambda > 0$	λ	λ
Uniforme $\{1, \dots, n\}$	$n \in \mathbb{N}^*$	$(n + 1)/2$	$(n^2 - 1)/12$

3.4 Fonction génératrice

Définition 3.14 (Fonction génératrice). Soit X une variable aléatoire à valeurs dans $\mathbb{N}$. La **fonction génératrice** de X est :

$$G_X(s) = \mathbb{E}[s^X] = \sum_{k=0}^{\infty} \mathbb{P}(X = k) s^k, \quad |s| \leq 1.$$

Proposition 3.15 (Propriétés de G_X). (i) $G_X(1) = 1$.

(ii) $\mathbb{E}[X] = G'_X(1)$.

(iii) $\mathbb{E}[X(X - 1)] = G''_X(1)$, d'où $\text{Var}(X) = G''_X(1) + G'_X(1) - [G'_X(1)]^2$.

(iv) $\mathbb{P}(X = k) = G_X^{(k)}(0)/k!$.

(v) Si X et Y sont indépendantes, $G_{X+Y}(s) = G_X(s) \cdot G_Y(s)$.

Exemple 3.16 (Fonction génératrice de la loi de Poisson). Si $X \sim \text{Poisson}(\lambda)$:

$$G_X(s) = \sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!} s^k = e^{-\lambda} \sum_{k=0}^{\infty} \frac{(\lambda s)^k}{k!} = e^{-\lambda} \cdot e^{\lambda s} = e^{\lambda(s-1)}.$$

On retrouve $G'_X(1) = \lambda = \mathbb{E}[X]$ et $G''_X(1) = \lambda^2$, d'où $\text{Var}(X) = \lambda^2 + \lambda - \lambda^2 = \lambda$.

3.5 Exercices

Exercice 3.1. Soit $X \sim \text{Bin}(n, p)$. Montrer directement (par calcul) que $\mathbb{E}[X] = np$ et $\text{Var}(X) = np(1 - p)$.

Exercice 3.2. Soit $X \sim \text{Geom}(p)$. Calculer $\mathbb{E}[X]$ en utilisant la fonction génératrice.

Exercice 3.3. Soit $X \sim \text{Poisson}(\lambda)$ et $Y \sim \text{Poisson}(\mu)$, indépendantes. Montrer que $X + Y \sim \text{Poisson}(\lambda + \mu)$ en utilisant les fonctions génératrices.

Exercice 3.4. Un standard téléphonique reçoit en moyenne 4 appels par minute, modélisés par une loi de Poisson. Calculer la probabilité de recevoir :

- (a) exactement 3 appels en une minute ;
- (b) au moins 1 appel en 30 secondes ;
- (c) plus de 10 appels en 2 minutes.

Exercice 3.5. Un sac contient 5 boules rouges et 15 boules bleues. On tire 4 boules sans remise. Calculer la probabilité d'obtenir exactement 2 boules rouges.

Exercice 3.6. On lance un dé équilibré jusqu'à obtenir un 6. Soit X le nombre de lancers.

- (a) Quelle est la loi de X ?
- (b) Calculer $\mathbb{P}(X > 10)$.
- (c) Calculer $\mathbb{P}(X > 15 \mid X > 5)$. Retrouver la perte de mémoire.

Exercice 3.7. Montrer que la loi géométrique est la seule loi discrète sur $\mathbb{N}^*$ vérifiant la propriété de perte de mémoire. *Indication* : poser $g(n) = \mathbb{P}(X > n)$ et utiliser l'équation fonctionnelle $g(m+n) = g(m)g(n)$.

Exercice 3.8. Soit $N \sim \text{Poisson}(\lambda)$ et $(X_i)_{i \geq 1}$ des variables i.i.d. de Bernoulli(p), indépendantes de N . Poser $S = \sum_{i=1}^N X_i$. Montrer que $S \sim \text{Poisson}(\lambda p)$.

Chapitre 4

Variables Aléatoires Continues

Avec les variables discrètes, on comptait. Avec les variables continues, on mesure. La transition est plus qu'un changement de vocabulaire : elle exige de remplacer les sommes par des intégrales, les masses ponctuelles par des densités, et d'apprivoiser l'idée qu'en continu, la probabilité d'obtenir une valeur *exacte* est toujours nulle. C'est Abraham de Moivre qui, vers 1733, a ouvert cette voie en approchant la loi binomiale par la courbe en cloche — ce que nous appelons aujourd'hui la loi normale, et que Gauss immortalisera dans ses travaux d'astronomie.

Intuition

Une variable aléatoire continue prend ses valeurs dans un intervalle (ou tout $\mathbb{R}$) et sa loi est décrite par une **densité de probabilité**. Ce chapitre développe les outils spécifiques au cas continu : densité, fonction de répartition, quantiles.

4.1 Densité de probabilité

Définition 4.1 (Variable aléatoire à densité). Une variable aléatoire X est dite **continue** (ou à densité) s'il existe une fonction $f_X : \mathbb{R} \rightarrow [0, +\infty)$, intégrable, telle que pour tout $a \leq b$:

$$\mathbb{P}(a \leq X \leq b) = \int_a^b f_X(x) dx.$$

La fonction f_X est appelée **densité de probabilité** de X .

Proposition 4.2 (Propriétés de la densité). (i) $f_X(x) \geq 0$ pour tout $x \in \mathbb{R}$.

(ii) $\int_{-\infty}^{+\infty} f_X(x) dx = 1$.

(iii) $\mathbb{P}(X = a) = 0$ pour tout $a \in \mathbb{R}$.

(iv) $\mathbb{P}(a \leq X \leq b) = \mathbb{P}(a < X < b) = \mathbb{P}(a \leq X < b) = \mathbb{P}(a < X \leq b)$.

La densité n'est pas une probabilité

On peut avoir $f_X(x) > 1$ pour certaines valeurs de x . La densité est une « concentration de probabilité », pas une probabilité. Seule son intégrale sur un intervalle donne une probabilité.

4.2 Fonction de répartition

Définition 4.3 (Fonction de répartition).

$$F_X(x) = \mathbb{P}(X \leq x) = \int_{-\infty}^x f_X(t) dt, \quad x \in \mathbb{R}.$$

Théorème 4.4 (Lien densité-répartition). *Si F_X est dérivable en x , alors $f_X(x) = F'_X(x)$. Plus généralement, f_X est une dérivée de F_X presque partout.*

Proposition 4.5 (Propriétés de F_X). (i) F_X est croissante et continue (dans le cas à densité).

$$(ii) \lim_{x \rightarrow -\infty} F_X(x) = 0, \lim_{x \rightarrow +\infty} F_X(x) = 1.$$

$$(iii) \mathbb{P}(a < X \leq b) = F_X(b) - F_X(a).$$

4.3 Lois continues classiques

4.3.1 Loi uniforme continue

Définition 4.6 (Uniforme sur $[a, b]$). $X \sim \mathcal{U}([a, b])$ avec $a < b$:

$$f_X(x) = \frac{1}{b-a} \mathbf{1}_{[a,b]}(x), \quad F_X(x) = \begin{cases} 0 & \text{si } x < a, \\ \frac{x-a}{b-a} & \text{si } a \leq x \leq b, \\ 1 & \text{si } x > b. \end{cases}$$

$$\mathbb{E}[X] = (a+b)/2, \text{Var}(X) = (b-a)^2/12.$$

4.3.2 Loi exponentielle

Définition 4.7 (Exponentielle). $X \sim \text{Exp}(\lambda)$ avec $\lambda > 0$:

$$f_X(x) = \lambda e^{-\lambda x} \mathbf{1}_{[0,+\infty)}(x), \quad F_X(x) = (1 - e^{-\lambda x}) \mathbf{1}_{[0,+\infty)}(x).$$

$$\mathbb{E}[X] = 1/\lambda, \text{Var}(X) = 1/\lambda^2.$$

Théorème 4.8 (Perte de mémoire). *La loi exponentielle est la seule loi continue sur $[0, +\infty)$ vérifiant :*

$$\mathbb{P}(X > s+t \mid X > s) = \mathbb{P}(X > t) \quad \forall s, t \geq 0.$$

Démonstration. Soit $g(t) = \mathbb{P}(X > t) = e^{-\lambda t}$. Alors :

$$\mathbb{P}(X > s+t \mid X > s) = \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} = e^{-\lambda t} = g(t).$$

Réciproquement, si $g(s+t) = g(s)g(t)$ avec g continue décroissante et $g(0) = 1$, alors $g(t) = e^{-\lambda t}$ pour un certain $\lambda > 0$, ce qui caractérise la loi exponentielle. $\square$

Exponentielle et Poisson

Si les arrivées suivent un processus de Poisson de taux λ , alors le temps entre deux arrivées consécutives suit une loi $\text{Exp}(\lambda)$.

4.3.3 Loi normale (gaussienne)

Définition 4.9 (Loi normale). $X \sim \mathcal{N}(\mu, \sigma^2)$ avec $\mu \in \mathbb{R}$ et $\sigma > 0$:

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \quad x \in \mathbb{R}.$$

$$\mathbb{E}[X] = \mu, \quad \text{Var}(X) = \sigma^2.$$

Définition 4.10 (Loi normale centrée réduite). Si $Z \sim \mathcal{N}(0, 1)$, on note $\Phi(x) = \mathbb{P}(Z \leq x)$ sa fonction de répartition. On a la relation :

$$\text{si } X \sim \mathcal{N}(\mu, \sigma^2), \quad \mathbb{P}(X \leq x) = \Phi\left(\frac{x-\mu}{\sigma}\right).$$

Valeurs clés de Φ

z	1,645	1,96	2,326	2,576	3,291
$\Phi(z)$	0,95	0,975	0,99	0,995	0,9995

Par symétrie : $\Phi(-z) = 1 - \Phi(z)$.

Proposition 4.11 (Règle des $k\sigma$). Si $X \sim \mathcal{N}(\mu, \sigma^2)$:

- $\mathbb{P}(|X - \mu| \leq \sigma) \approx 0,6827$ (règle du 1σ),
- $\mathbb{P}(|X - \mu| \leq 2\sigma) \approx 0,9545$ (règle du 2σ),
- $\mathbb{P}(|X - \mu| \leq 3\sigma) \approx 0,9973$ (règle du 3σ).

4.3.4 Loi gamma

Définition 4.12 (Gamma). $X \sim \Gamma(\alpha, \lambda)$ avec $\alpha, \lambda > 0$:

$$f_X(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} \mathbf{1}_{(0,+\infty)}(x),$$

où $\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} dt$. On a $\mathbb{E}[X] = \alpha/\lambda$ et $\text{Var}(X) = \alpha/\lambda^2$.

Remarque 4.13. $\Gamma(1, \lambda) = \text{Exp}(\lambda)$ et $\Gamma(n/2, 1/2) = \chi^2(n)$ (loi du chi-deux à n degrés de liberté).

4.3.5 Loi bêta

Définition 4.14 (Bêta). $X \sim \text{Beta}(\alpha, \beta)$ avec $\alpha, \beta > 0$:

$$f_X(x) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha, \beta)} \mathbf{1}_{(0,1)}(x),$$

où $B(\alpha, \beta) = \Gamma(\alpha)\Gamma(\beta)/\Gamma(\alpha + \beta)$. $\mathbb{E}[X] = \alpha/(\alpha + \beta)$.

4.4 Quantiles et médiane

Définition 4.15 (Quantile). Le **quantile d'ordre** p ($0 < p < 1$) de X est :

$$q_p = \inf\{x \in \mathbb{R} : F_X(x) \geq p\} = F_X^{-1}(p).$$

Définition 4.16 (Médiane). La **médiane** de X est le quantile d'ordre $1/2$: $\text{Med}(X) = q_{1/2}$.

Exemple 4.17. Si $X \sim \text{Exp}(\lambda)$, alors $F_X(x) = 1 - e^{-\lambda x}$ pour $x \geq 0$. Le quantile d'ordre p est :

$$q_p = -\frac{\ln(1-p)}{\lambda}.$$

En particulier, $\text{Med}(X) = \ln 2/\lambda$.

4.5 Simulation par inversion

Théorème 4.18 (Méthode d'inversion). Si $U \sim \mathcal{U}([0, 1])$ et F est une fonction de répartition continue strictement croissante, alors $X = F^{-1}(U)$ a pour fonction de répartition F .

Démonstration. $\mathbb{P}(X \leq x) = \mathbb{P}(F^{-1}(U) \leq x) = \mathbb{P}(U \leq F(x)) = F(x)$. □

Exemple 4.19. Pour simuler $X \sim \text{Exp}(\lambda)$, on génère $U \sim \mathcal{U}([0, 1])$ et on pose $X = -\ln(1-U)/\lambda = -\ln(U)/\lambda$ (car $1-U$ a la même loi que U).

4.6 Exercices

Exercice 4.1. Soit X de densité $f(x) = cx^2 \mathbf{1}_{[0,1]}(x)$. Déterminer c , $\mathbb{E}[X]$, $\text{Var}(X)$ et F_X .

Exercice 4.2. Soit $X \sim \mathcal{N}(100, 25)$. Calculer :

- (a) $\mathbb{P}(X > 105)$;
- (b) $\mathbb{P}(90 < X < 110)$;
- (c) le quantile $q_{0,9}$.

Exercice 4.3. Montrer que si $X \sim \mathcal{N}(\mu, \sigma^2)$, alors $Z = (X - \mu)/\sigma \sim \mathcal{N}(0, 1)$.

Exercice 4.4. Soit $X \sim \text{Exp}(\lambda)$. Calculer la densité et l'espérance de $Y = X^2$.

Exercice 4.5. Montrer que l'intégrale de Gauss vaut : $\int_{-\infty}^{+\infty} e^{-x^2/2} dx = \sqrt{2\pi}$. *Indication :* calculer I^2 en coordonnées polaires.

Exercice 4.6. Soit $X \sim \Gamma(\alpha, \lambda)$. Montrer que $Y = 2\lambda X \sim \chi^2(2\alpha)$ lorsque $\alpha \in \mathbb{N}^*/2$.

Exercice 4.7 (Simulation). Décrire une méthode pour simuler une variable aléatoire de densité $f(x) = 2x \mathbf{1}_{[0,1]}(x)$ par la méthode d'inversion.

Exercice 4.8. Soit X de densité $f(x) = \frac{2}{\pi(1+x^2)} \mathbf{1}_{(0,+\infty)}(x)$. Vérifier que f est bien une densité et calculer F_X . L'espérance de X existe-t-elle ?

Chapitre 5

Espérance, Variance, Moments

Si vous jouez à un jeu de hasard un grand nombre de fois, votre gain moyen par partie convergera vers un nombre précis : l'*espérance*. Ce résultat, pressenti par Pascal et Fermat dès 1654, est la loi des grands nombres. L'espérance résume la « position » d'une variable aléatoire ; la *variance*, introduite par Ronald Fisher en 1918, mesure sa « dispersion ». Ensemble, ces deux paramètres capturent l'essentiel du comportement d'une distribution, et les moments d'ordre supérieur — asymétrie, aplatissement — en raffinent le portrait.

Intuition

L'espérance et la variance sont les deux paramètres fondamentaux résumant le comportement d'une variable aléatoire : position et dispersion. Ce chapitre les définit rigoureusement, établit leurs propriétés et introduit les moments d'ordre supérieur.

5.1 Espérance

Définition 5.1 (Espérance — cas discret). Soit X une variable aléatoire discrète prenant les valeurs $x_1, x_2, \dots$. L'**espérance** de X est :

$$\mathbb{E}[X] = \sum_k x_k \mathbb{P}(X = x_k),$$

pourvu que $\sum_k |x_k| \mathbb{P}(X = x_k) < \infty$ (convergence absolue).

Définition 5.2 (Espérance — cas continu). Si X admet une densité f_X , l'**espérance** est :

$$\mathbb{E}[X] = \int_{-\infty}^{+\infty} x f_X(x) dx,$$

pourvu que $\int |x| f_X(x) dx < \infty$.

Théorème 5.3 (Formule de transfert (LOTUS)). Si $g : \mathbb{R} \rightarrow \mathbb{R}$ est mesurable :

- **Cas discret** : $\mathbb{E}[g(X)] = \sum_k g(x_k) \mathbb{P}(X = x_k)$.
- **Cas continu** : $\mathbb{E}[g(X)] = \int_{-\infty}^{+\infty} g(x) f_X(x) dx$.

(Pourvu que les sommes/intégrales convergent absolument.)

Pourquoi LOTUS est si utile

La formule de transfert permet de calculer $\mathbb{E}[g(X)]$ directement à partir de la loi de X , sans avoir à déterminer la loi de $Y = g(X)$.

Théorème 5.4 (Propriétés de l'espérance). (i) **Linéarité** : $\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]$ pour tous $a, b \in \mathbb{R}$.

(ii) **Positivité** : si $X \geq 0$ p.s., alors $\mathbb{E}[X] \geq 0$.

(iii) **Monotonie** : si $X \leq Y$ p.s., alors $\mathbb{E}[X] \leq \mathbb{E}[Y]$.

(iv) $|\mathbb{E}[X]| \leq \mathbb{E}[|X|]$.

(v) **Produit pour variables indépendantes** : si $X \perp\!\!\!\perp Y$ et $\mathbb{E}[|X|], \mathbb{E}[|Y|] < \infty$, alors $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$.

Exemple 5.5. Soit $X \sim \text{Exp}(\lambda)$. On calcule :

$$\mathbb{E}[X] = \int_0^\infty x \lambda e^{-\lambda x} dx = [-x e^{-\lambda x}]_0^\infty + \int_0^\infty e^{-\lambda x} dx = 0 + \frac{1}{\lambda} = \frac{1}{\lambda}.$$

5.2 Variance

Définition 5.6 (Variance et écart-type). La **variance** de X (si $\mathbb{E}[X^2] < \infty$) est :

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

L'**écart-type** est $\sigma(X) = \sqrt{\text{Var}(X)}$.

Théorème 5.7 (Formule de König-Huygens).

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

Démonstration.

$$\begin{aligned} \text{Var}(X) &= \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2 - 2X\mathbb{E}[X] + (\mathbb{E}[X])^2] \\ &= \mathbb{E}[X^2] - 2\mathbb{E}[X]\mathbb{E}[X] + (\mathbb{E}[X])^2 = \mathbb{E}[X^2] - (\mathbb{E}[X])^2. \end{aligned} \quad \square$$

Proposition 5.8 (Propriétés de la variance). (i) $\text{Var}(X) \geq 0$, et $\text{Var}(X) = 0$ si et seulement si X est constante p.s.

(ii) $\text{Var}(aX + b) = a^2\text{Var}(X)$.

(iii) Si $X \perp\!\!\!\perp Y$, alors $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$.

(iv) Plus généralement : $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y)$.

Résumé : espérance et variance

$$\mathbb{E}[aX + b] = a\mathbb{E}[X] + b, \quad \text{Var}(aX + b) = a^2\text{Var}(X), \quad \text{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2.$$

5.3 Moments d'ordre supérieur

Définition 5.9 (Moment d'ordre r). Le **moment d'ordre** r de X ($r \in \mathbb{N}^*$) est :

$$m_r = \mathbb{E}[X^r],$$

lorsque cette espérance existe. Le **moment centré d'ordre** r est : $\mu_r = \mathbb{E}[(X - \mathbb{E}[X])^r]$.

Définition 5.10 (Coefficient d'asymétrie et kurtosis). • Le **coefficient d'asymétrie** (skewness) est : $\gamma_1 = \mu_3/\sigma^3$.

- Le **kurtosis** est : $\gamma_2 = \mu_4/\sigma^4 - 3$.
Pour la loi normale, $\gamma_1 = 0$ et $\gamma_2 = 0$.

Remarque 5.11. $\gamma_1 > 0$ indique une queue droite plus lourde, $\gamma_1 < 0$ une queue gauche plus lourde. $\gamma_2 > 0$ indique des queues plus lourdes que la normale (distribution leptokurtique).

5.4 Inégalités fondamentales

Théorème 5.12 (Inégalité de Markov). Si $X \geq 0$ et $a > 0$:

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}[X]}{a}.$$

Démonstration. $\mathbb{E}[X] = \mathbb{E}[X \mathbf{1}_{X \geq a}] + \mathbb{E}[X \mathbf{1}_{X < a}] \geq \mathbb{E}[X \mathbf{1}_{X \geq a}] \geq a \mathbb{P}(X \geq a)$. □

Théorème 5.13 (Inégalité de Bienaymé–Tchebychev). Si $\text{Var}(X)$ existe :

$$\mathbb{P}(|X - \mathbb{E}[X]| \geq \varepsilon) \leq \frac{\text{Var}(X)}{\varepsilon^2}, \quad \forall \varepsilon > 0.$$

Démonstration. Appliquer l'inégalité de Markov à $Y = (X - \mathbb{E}[X])^2$ avec $a = \varepsilon^2$:

$$\mathbb{P}(|X - \mathbb{E}[X]| \geq \varepsilon) = \mathbb{P}(Y \geq \varepsilon^2) \leq \frac{\mathbb{E}[Y]}{\varepsilon^2} = \frac{\text{Var}(X)}{\varepsilon^2}. \quad \square$$

Théorème 5.14 (Inégalité de Jensen). Si φ est convexe et $\mathbb{E}[|X|] < \infty$:

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

Si φ est concave, l'inégalité est renversée.

Exemple 5.15. En prenant $\varphi(x) = x^2$ (convexe), on retrouve $(\mathbb{E}[X])^2 \leq \mathbb{E}[X^2]$, c'est-à-dire $\text{Var}(X) \geq 0$.

Théorème 5.16 (Inégalité de Cauchy–Schwarz).

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

Égalité si et seulement si $Y = aX + b$ p.s. pour certains $a, b \in \mathbb{R}$.

5.5 Espérance conditionnelle (introduction)

Définition 5.17 (Espérance conditionnelle discrète). Si X et Y sont discrètes :

$$\mathbb{E}[X \mid Y = y] = \sum_x x \mathbb{P}(X = x \mid Y = y).$$

Théorème 5.18 (Formule de l'espérance totale).

$$\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X \mid Y]] = \sum_y \mathbb{E}[X \mid Y = y] \mathbb{P}(Y = y).$$

Théorème 5.19 (Formule de la variance totale).

$$\text{Var}(X) = \mathbb{E}[\text{Var}(X \mid Y)] + \text{Var}(\mathbb{E}[X \mid Y]).$$

Exemple 5.20. Soit $N \sim \text{Poisson}(\lambda)$ et, conditionnellement à $N = n$, $S = X_1 + \dots + X_n$ où les X_i sont i.i.d. avec $\mathbb{E}[X_i] = \mu$ et $\text{Var}(X_i) = \sigma^2$. Alors :

$$\begin{aligned} \mathbb{E}[S] &= \mathbb{E}[\mathbb{E}[S \mid N]] = \mathbb{E}[N\mu] = \lambda\mu, \\ \text{Var}(S) &= \mathbb{E}[\text{Var}(S \mid N)] + \text{Var}(\mathbb{E}[S \mid N]) = \mathbb{E}[N\sigma^2] + \text{Var}(N\mu) = \lambda\sigma^2 + \lambda\mu^2. \end{aligned}$$

5.6 Exercices

Exercice 5.1. Calculer $\mathbb{E}[X]$ et $\text{Var}(X)$ pour $X \sim \mathcal{N}(\mu, \sigma^2)$ directement par intégration.

Exercice 5.2. Soit X de densité $f(x) = 6x(1-x) \mathbf{1}_{[0,1]}(x)$. Calculer $\mathbb{E}[X]$, $\mathbb{E}[X^2]$, $\text{Var}(X)$, γ_1 et γ_2 .

Exercice 5.3. Montrer l'inégalité de Markov. En déduire que si $\mathbb{E}[e^{tX}]$ existe pour $t > 0$, alors $\mathbb{P}(X \geq a) \leq e^{-ta} \mathbb{E}[e^{tX}]$ pour tout a (borne de Chernoff).

Exercice 5.4. Soit X uniformément distribuée sur $\{1, 2, \dots, n\}$. Calculer $\mathbb{E}[X]$, $\mathbb{E}[X^2]$ et $\text{Var}(X)$.

Exercice 5.5. Montrer l'inégalité de Cauchy-Schwarz en considérant le polynôme $p(t) = \mathbb{E}[(X + tY)^2] \geq 0$ et en étudiant son discriminant.

Exercice 5.6. Un dé équilibré est lancé N fois, où $N \sim \text{Poisson}(10)$. Soit S la somme des résultats. Calculer $\mathbb{E}[S]$ et $\text{Var}(S)$.

Exercice 5.7. Soit X une variable aléatoire positive. Montrer que :

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

Indication : intervertir l'intégrale et l'espérance.

Exercice 5.8. Montrer que $\text{Var}(X) = \min_{c \in \mathbb{R}} \mathbb{E}[(X - c)^2]$, le minimum étant atteint en $c = \mathbb{E}[X]$.

Chapitre 6

Lois Usuelles et leurs Propriétés

La théorie des probabilités a produit, au fil des siècles, une galerie de distributions remarquables — chacune née d'un problème concret, et chacune dotée de propriétés spécifiques qui la rendent irremplaçable dans son domaine. La loi normale de Gauss, la loi exponentielle des temps d'attente, la loi de Poisson des événements rares, la loi beta de la statistique bayésienne, la loi gamma de la théorie des files d'attente : ce chapitre les rassemble dans un panorama organisé, mettant en lumière leurs relations mutuelles (somme, conditionnement, passage à la limite) et les propriétés qui les caractérisent.

Intuition

Ce chapitre présente un catalogue organisé des lois de probabilité les plus courantes, tant discrètes que continues, avec leurs propriétés, relations mutuelles et domaines d'application.

6.1 Panorama des lois discrètes

6.1.1 Loi binomiale négative

Définition 6.1 (Binomiale négative). $X \sim \text{BN}(r, p)$: nombre d'essais pour obtenir le r -ième succès.

$$\mathbb{P}(X = k) = \binom{k-1}{r-1} p^r (1-p)^{k-r}, \quad k = r, r+1, \dots$$

$\mathbb{E}[X] = r/p$, $\text{Var}(X) = r(1-p)/p^2$. Pour $r = 1$, on retrouve la loi géométrique.

6.1.2 Loi uniforme discrète

Définition 6.2 (Uniforme discrète). $X \sim \mathcal{U}\{a, a+1, \dots, b\}$:

$$\mathbb{P}(X = k) = \frac{1}{b-a+1}, \quad k = a, a+1, \dots, b.$$

$\mathbb{E}[X] = (a+b)/2$, $\text{Var}(X) = ((b-a+1)^2 - 1)/12$.

6.1.3 Loi de Poisson — compléments

Proposition 6.3 (Stabilité par somme). Si $X_1 \sim \text{Poisson}(\lambda_1)$ et $X_2 \sim \text{Poisson}(\lambda_2)$ sont indépendantes, alors $X_1 + X_2 \sim \text{Poisson}(\lambda_1 + \lambda_2)$.

Proposition 6.4 (Décomposition de Poisson). Si $X \sim \text{Poisson}(\lambda)$ et chaque événement est classé de type 1 avec probabilité p et de type 2 avec probabilité $1 - p$, indépendamment, alors les nombres N_1, N_2 de chaque type sont indépendants avec $N_1 \sim \text{Poisson}(\lambda p)$ et $N_2 \sim \text{Poisson}(\lambda(1 - p))$.

6.2 Panorama des lois continues

6.2.1 Loi log-normale

Définition 6.5 (Log-normale). $X \sim \text{LogN}(\mu, \sigma^2)$ si $\ln X \sim \mathcal{N}(\mu, \sigma^2)$.

$$f_X(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{(\ln x - \mu)^2}{2\sigma^2}\right), \quad x > 0.$$

$$\mathbb{E}[X] = e^{\mu+\sigma^2/2}, \quad \text{Var}(X) = e^{2\mu+2\sigma^2}(e^{\sigma^2} - 1).$$

6.2.2 Loi de Cauchy

Définition 6.6 (Cauchy). $X \sim \text{Cauchy}(x_0, \gamma)$:

$$f_X(x) = \frac{1}{\pi\gamma \left(1 + \left(\frac{x-x_0}{\gamma}\right)^2\right)}.$$

L'espérance et la variance n'existent pas.

Cauchy n'a pas de moments

La loi de Cauchy montre que toute variable aléatoire n'a pas forcément d'espérance. La moyenne empirique de variables de Cauchy i.i.d. ne converge pas : la loi des grands nombres ne s'applique pas.

6.2.3 Loi de Weibull

Définition 6.7 (Weibull). $X \sim \text{Weibull}(k, \lambda)$ avec $k, \lambda > 0$:

$$f_X(x) = \frac{k}{\lambda} \left(\frac{x}{\lambda}\right)^{k-1} e^{-(x/\lambda)^k} \mathbf{1}_{(0,+\infty)}(x).$$

Pour $k = 1$, on retrouve la loi exponentielle de paramètre $1/\lambda$.

6.2.4 Loi chi-deux

Définition 6.8 (Chi-deux). Si $Z_1, \dots, Z_n$ sont i.i.d. $\mathcal{N}(0, 1)$, alors $\chi_n^2 = Z_1^2 + \dots + Z_n^2 \sim \Gamma(n/2, 1/2)$ est la loi du **chi-deux** à n **degrés de liberté**. $\mathbb{E}[\chi_n^2] = n$, $\text{Var}(\chi_n^2) = 2n$.

6.2.5 Loi de Student

Définition 6.9 (Student). Si $Z \sim \mathcal{N}(0, 1)$ et $V \sim \chi_n^2$ sont indépendants, alors $T = Z/\sqrt{V/n}$ suit la **loi de Student à n degrés de liberté**. $\mathbb{E}[T] = 0$ (pour $n > 1$), $\text{Var}(T) = n/(n-2)$ (pour $n > 2$). Pour $n = 1$, on retrouve la loi de Cauchy standard.

6.2.6 Loi de Fisher

Définition 6.10 (Fisher). Si $U \sim \chi_m^2$ et $V \sim \chi_n^2$ sont indépendants, alors $F = (U/m)/(V/n)$ suit la **loi de Fisher à (m, n) degrés de liberté**.

6.3 Relations entre les lois

Carte des relations entre lois

- $\text{Bernoulli}(p) = \text{Bin}(1, p)$.
- $\text{Geom}(p) = \text{BN}(1, p)$.
- $\text{Exp}(\lambda) = \Gamma(1, \lambda)$.
- $\chi^2(n) = \Gamma(n/2, 1/2)$.
- $\text{Bin}(n, p) \xrightarrow{n \rightarrow \infty, np \rightarrow \lambda} \text{Poisson}(\lambda)$.
- $\text{Bin}(n, p) \xrightarrow{n \rightarrow \infty} \mathcal{N}(np, np(1-p))$ (TCL).
- $\text{Poisson}(\lambda) \xrightarrow{\lambda \rightarrow \infty} \mathcal{N}(\lambda, \lambda)$.
- $t_n \xrightarrow{n \rightarrow \infty} \mathcal{N}(0, 1)$.
- $\Gamma(n, \lambda) =$ loi de la somme de n v.a. $\text{Exp}(\lambda)$ indépendantes.

6.4 Tableau récapitulatif général

Récapitulatif des lois continues

Loi	Densité	Support	$\mathbb{E}[X]$	$\text{Var}(X)$
$\mathcal{U}([a, b])$	$\frac{1}{b-a}$	$[a, b]$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
$\text{Exp}(\lambda)$	$\lambda e^{-\lambda x}$	$[0, \infty)$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
$\mathcal{N}(\mu, \sigma^2)$	$\frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/(2\sigma^2)}$	$\mathbb{R}$	μ	σ^2
$\Gamma(\alpha, \lambda)$	$\frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}$	$(0, \infty)$	$\frac{\alpha}{\lambda}$	$\frac{\alpha}{\lambda^2}$
$\text{Beta}(\alpha, \beta)$	$\frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha, \beta)}$	$(0, 1)$	$\frac{\alpha}{\alpha+\beta}$	$\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$

6.5 Méthodes de reconnaissance

Remarque 6.11. Pour identifier une loi, on peut procéder ainsi :

1. Identifier le **support** : $\mathbb{N}$, $\{0, 1, \dots, n\}$, $(0, \infty)$, $\mathbb{R}$, $(0, 1)$, etc.
2. Repérer la **forme fonctionnelle** : exponentielle décroissante, cloche symétrique, puissance...
3. Vérifier les **paramètres** en calculant espérance et variance.

6.6 Exercices

Exercice 6.1. Soit $X \sim \Gamma(n, \lambda)$ avec $n \in \mathbb{N}^*$. Montrer que X a même loi que la somme de n variables i.i.d. $\text{Exp}(\lambda)$.

Exercice 6.2. Montrer que si $X \sim \mathcal{N}(0, 1)$, alors $X^2 \sim \chi^2(1)$.

Exercice 6.3. Soit $X \sim \text{Poisson}(\lambda)$ avec $\lambda = 100$. En utilisant l'approximation normale, estimer $\mathbb{P}(90 \leq X \leq 110)$.

Exercice 6.4. Soit $U \sim \mathcal{U}([0, 1])$. Calculer la loi de $X = -\ln U$.

Exercice 6.5. Soit $X \sim \text{Beta}(1, 1)$. Montrer que $X \sim \mathcal{U}([0, 1])$.

Exercice 6.6. On observe le nombre de clients arrivant dans un magasin chaque heure, supposé suivre une loi de Poisson de paramètre $\lambda = 12$. Chaque client achète indépendamment avec probabilité $p = 0,3$. Quelle est la loi du nombre de clients acheteurs par heure ?

Exercice 6.7. Montrer que si $X \sim t_n$ (Student à n ddl), alors $X^2 \sim F(1, n)$ (Fisher).

Exercice 6.8. Soit X de loi log-normale $\text{LogN}(0, 1)$. Calculer $\mathbb{E}[X]$, $\text{Var}(X)$ et la médiane de X .

Chapitre 7

Fonctions de Variables Aléatoires

En sciences et en ingénierie, on mesure rarement la quantité qui nous intéresse directement. Un physicien mesure un angle et en déduit une vitesse ; un financier observe un rendement logarithmique et en déduit un prix ; un ingénieur mesure une tension et en déduit une puissance. À chaque fois, on applique une transformation g à une variable aléatoire X pour obtenir une nouvelle variable aléatoire $Y = g(X)$. Mais quelle est la loi de Y ? La question, d'apparence innocente, conduit à des techniques riches et variées : formule du changement de variable, méthode de la fonction de répartition, jacobiens pour les transformations multidimensionnelles. Maîtriser ces outils, c'est passer du statut d'observateur passif de l'aléa à celui d'architecte capable de manipuler les distributions à volonté.

Intuition

Lorsqu'on applique une transformation à une variable aléatoire, on obtient une nouvelle variable aléatoire dont la loi peut être très différente. Ce chapitre développe les méthodes permettant de déterminer la loi de $Y = g(X)$ à partir de celle de X .

7.1 Cas discret

Proposition 7.1 (Loi de $g(X)$ — cas discret). Si X est une variable discrète prenant les valeurs $x_1, x_2, \dots$ et $g : \mathbb{R} \rightarrow \mathbb{R}$, alors $Y = g(X)$ est discrète et :

$$\mathbb{P}(Y = y) = \sum_{k: g(x_k)=y} \mathbb{P}(X = x_k).$$

Exemple 7.2. Si $X \sim \text{Bin}(n, 1/2)$ et $Y = 2X - n$, alors Y prend les valeurs $-n, -n + 2, \dots, n - 2, n$ et :

$$\mathbb{P}(Y = 2k - n) = \binom{n}{k} \left(\frac{1}{2}\right)^n, \quad k = 0, 1, \dots, n.$$

7.2 Méthode de la fonction de répartition

Théorème 7.3 (Méthode de la CDF). Pour trouver la loi de $Y = g(X)$:

1. Calculer $F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(g(X) \leq y)$.

2. Dériver : $f_Y(y) = F'_Y(y)$.

Exemple 7.4 ($Y = X^2$ avec $X \sim \mathcal{N}(0, 1)$). Pour $y > 0$:

$$F_Y(y) = \mathbb{P}(X^2 \leq y) = \mathbb{P}(-\sqrt{y} \leq X \leq \sqrt{y}) = \Phi(\sqrt{y}) - \Phi(-\sqrt{y}) = 2\Phi(\sqrt{y}) - 1.$$

En dérivant :

$$f_Y(y) = 2 \cdot \frac{1}{\sqrt{2\pi}} e^{-y/2} \cdot \frac{1}{2\sqrt{y}} = \frac{1}{\sqrt{2\pi}} y^{-1/2} e^{-y/2}, \quad y > 0.$$

C'est la densité d'une loi $\Gamma(1/2, 1/2) = \chi^2(1)$.

Exemple 7.5 ($Y = e^X$ avec $X \sim \mathcal{N}(\mu, \sigma^2)$). Pour $y > 0$:

$$F_Y(y) = \mathbb{P}(e^X \leq y) = \mathbb{P}(X \leq \ln y) = \Phi\left(\frac{\ln y - \mu}{\sigma}\right).$$

En dérivant :

$$f_Y(y) = \frac{1}{y\sigma\sqrt{2\pi}} \exp\left(-\frac{(\ln y - \mu)^2}{2\sigma^2}\right), \quad y > 0.$$

On retrouve la densité de la loi log-normale.

7.3 Méthode du changement de variable

Théorème 7.6 (Changement de variable — cas monotone). *Soit X une variable continue de densité f_X et g une fonction **strictement monotone** et de classe C^1 sur le support de X , avec g^{-1} de classe C^1 . Alors $Y = g(X)$ a pour densité :*

$$f_Y(y) = f_X(g^{-1}(y)) \cdot |(g^{-1})'(y)|.$$

Démonstration. Supposons g strictement croissante (le cas décroissant est analogue).

$$F_Y(y) = \mathbb{P}(g(X) \leq y) = \mathbb{P}(X \leq g^{-1}(y)) = F_X(g^{-1}(y)).$$

En dérivant par la règle de la chaîne : $f_Y(y) = f_X(g^{-1}(y)) \cdot (g^{-1})'(y)$. Comme g est croissante, $(g^{-1})' > 0$, d'où la valeur absolue est superflue. Dans le cas décroissant, on obtient un signe moins qui est absorbé par $|\cdot|$. $\square$

Exemple 7.7 (Transformation affine). Si $Y = aX + b$ avec $a \neq 0$: $g^{-1}(y) = (y - b)/a$, $(g^{-1})'(y) = 1/a$. Donc :

$$f_Y(y) = \frac{1}{|a|} f_X\left(\frac{y - b}{a}\right).$$

En particulier, si $X \sim \mathcal{N}(\mu, \sigma^2)$, alors $Z = (X - \mu)/\sigma \sim \mathcal{N}(0, 1)$.

7.4 Cas non monotone

Théorème 7.8 (Changement de variable — cas général). *Si g est de classe C^1 et le support de X peut être décomposé en un nombre fini d'intervalles $I_1, \dots, I_m$ sur chacun desquels g est strictement monotone, avec inverses $h_1, \dots, h_m$, alors :*

$$f_Y(y) = \sum_{j=1}^m f_X(h_j(y)) \cdot |h'_j(y)| \mathbf{1}_{g(I_j)}(y).$$

Exemple 7.9 ($Y = X^2$ avec X de densité f_X symétrique). $g(x) = x^2$ est décroissante sur $(-\infty, 0)$ et croissante sur $(0, +\infty)$. Les deux inverses sont $h_1(y) = -\sqrt{y}$ et $h_2(y) = \sqrt{y}$, avec $|h'_j(y)| = 1/(2\sqrt{y})$. Donc :

$$f_Y(y) = \frac{f_X(\sqrt{y}) + f_X(-\sqrt{y})}{2\sqrt{y}}, \quad y > 0.$$

7.5 Fonction génératrice de moments

Définition 7.10 (Fonction génératrice de moments (MGF)). La **fonction génératrice de moments** de X est :

$$M_X(t) = \mathbb{E}[e^{tX}],$$

définie pour les t dans un voisinage de 0 où cette espérance existe.

Proposition 7.11 (Propriétés de la MGF). (i) $M_X(0) = 1$.

(ii) $M_X^{(n)}(0) = \mathbb{E}[X^n]$ (les moments s'obtiennent par dérivation).

(iii) Si $M_X(t) = M_Y(t)$ pour tout t dans un voisinage de 0, alors X et Y ont même loi.

(iv) Si $X \perp\!\!\!\perp Y$, alors $M_{X+Y}(t) = M_X(t) \cdot M_Y(t)$.

Exemple 7.12 (MGF de la loi normale). Si $X \sim \mathcal{N}(\mu, \sigma^2)$:

$$M_X(t) = \exp\left(\mu t + \frac{\sigma^2 t^2}{2}\right).$$

On retrouve $M'_X(0) = \mu = \mathbb{E}[X]$ et $M''_X(0) = \sigma^2 + \mu^2 = \mathbb{E}[X^2]$.

7.6 Transformation de probabilité intégrale

Théorème 7.13 (Transformation de probabilité intégrale). Si X est une variable continue de CDF F_X strictement croissante, alors $U = F_X(X) \sim \mathcal{U}([0, 1])$.

Démonstration. Pour $u \in (0, 1)$: $\mathbb{P}(U \leq u) = \mathbb{P}(F_X(X) \leq u) = \mathbb{P}(X \leq F_X^{-1}(u)) = F_X(F_X^{-1}(u)) = u$. $\square$

Remarque 7.14. Ce résultat est la base théorique de la méthode d'inversion pour la simulation (chapitre 4) et des tests d'adéquation (p-values).

7.7 Exercices

Exercice 7.1. Soit $X \sim \mathcal{U}([0, 1])$ et $Y = -\ln X$. Déterminer la loi de Y .

Exercice 7.2. Soit $X \sim \text{Exp}(1)$. Calculer la densité de $Y = \sqrt{X}$.

Exercice 7.3. Soit $X \sim \mathcal{N}(0, 1)$. Calculer la densité de $Y = |X|$ (loi normale repliée).

Exercice 7.4. Soit X de densité $f_X(x) = 2x \mathbf{1}_{[0,1]}(x)$ et $Y = 1 - X^2$. Déterminer la densité de Y .

Exercice 7.5. Calculer la MGF de la loi $\text{Exp}(\lambda)$ et en déduire $\mathbb{E}[X]$, $\mathbb{E}[X^2]$, $\text{Var}(X)$.

Exercice 7.6. Soit $X \sim \text{Cauchy}(0, 1)$ de densité $f(x) = 1/(\pi(1 + x^2))$. Montrer que la MGF de X n'existe pas (i.e. $\mathbb{E}[e^{tX}] = \infty$ pour tout $t \neq 0$).

Exercice 7.7. Soit U_1, U_2 i.i.d. $\mathcal{U}([0, 1])$. Par la méthode de Box–Muller, on pose : $Z_1 = \sqrt{-2 \ln U_1} \cos(2\pi U_2)$ et $Z_2 = \sqrt{-2 \ln U_1} \sin(2\pi U_2)$. Montrer que Z_1, Z_2 sont i.i.d. $\mathcal{N}(0, 1)$.

Exercice 7.8. En utilisant la MGF, montrer que si $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$ et $X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$ sont indépendantes, alors $X_1 + X_2 \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.

Chapitre 8

Couples de Variables Aléatoires

Jusqu'ici, nous avons étudié des variables aléatoires isolées. Mais dans le monde réel, les grandeurs sont rarement indépendantes : la taille et le poids d'un individu sont corrélés, le rendement d'une action dépend du marché global, la température et la pression atmosphérique évoluent conjointement. Pour comprendre ces dépendances, il faut passer de la loi d'une variable à la *loi conjointe* de plusieurs variables. Francis Galton, dans les années 1880, a été le premier à étudier systématiquement la corrélation — en mesurant la taille des parents et de leurs enfants. Karl Pearson a ensuite formalisé le coefficient de corrélation qui porte son nom. Ce chapitre développe les outils fondamentaux : lois conjointes, marginales, conditionnelles, covariance et corrélation.

Intuition

Ce chapitre étudie les lois conjointes de deux (ou plusieurs) variables aléatoires, les lois marginales et conditionnelles, la covariance et la corrélation. Ces outils sont essentiels pour modéliser les dépendances entre grandeurs aléatoires.

8.1 Loi conjointe

8.1.1 Cas discret

Définition 8.1 (Loi conjointe discrète). Soit (X, Y) un couple de variables aléatoires discrètes. La **loi conjointe** est la fonction :

$$p_{X,Y}(x_i, y_j) = \mathbb{P}(X = x_i, Y = y_j), \quad \text{avec} \quad \sum_{i,j} p_{X,Y}(x_i, y_j) = 1.$$

8.1.2 Cas continu

Définition 8.2 (Densité conjointe). Le couple (X, Y) admet une **densité conjointe** $f_{X,Y}$ si pour tout borélien $B \subseteq \mathbb{R}^2$:

$$\mathbb{P}((X, Y) \in B) = \iint_B f_{X,Y}(x, y) \, dx \, dy,$$

avec $f_{X,Y}(x, y) \geq 0$ et $\iint_{\mathbb{R}^2} f_{X,Y}(x, y) \, dx \, dy = 1$.

Définition 8.3 (Fonction de répartition conjointe).

$$F_{X,Y}(x, y) = \mathbb{P}(X \leq x, Y \leq y).$$

Dans le cas à densité : $f_{X,Y}(x, y) = \frac{\partial^2 F_{X,Y}}{\partial x \partial y}(x, y)$.

8.2 Lois marginales

Définition 8.4 (Marginales). Les **lois marginales** s'obtiennent par sommation (discret) ou intégration (continu) sur l'autre variable :

$$\begin{aligned} \text{Discret :} \quad p_X(x_i) &= \sum_j p_{X,Y}(x_i, y_j), & p_Y(y_j) &= \sum_i p_{X,Y}(x_i, y_j). \\ \text{Continu :} \quad f_X(x) &= \int_{-\infty}^{+\infty} f_{X,Y}(x, y) dy, & f_Y(y) &= \int_{-\infty}^{+\infty} f_{X,Y}(x, y) dx. \end{aligned}$$

La loi conjointe détermine les marginales, mais pas l'inverse

Connaissant $f_{X,Y}$, on calcule f_X et f_Y . Mais connaissant seulement f_X et f_Y , on ne peut pas reconstituer $f_{X,Y}$ sans information sur la dépendance entre X et Y .

Exemple 8.5. Soit (X, Y) de densité conjointe $f_{X,Y}(x, y) = 6(1 - y)$ pour $0 < x < y < 1$ et 0 sinon. Alors :

$$\begin{aligned} f_X(x) &= \int_x^1 6(1 - y) dy = 3(1 - x)^2, \quad 0 < x < 1. \\ f_Y(y) &= \int_0^y 6(1 - y) dx = 6y(1 - y), \quad 0 < y < 1. \end{aligned}$$

8.3 Loi conditionnelle

Définition 8.6 (Densité conditionnelle). La **densité conditionnelle** de Y sachant $X = x$ est :

$$f_{Y|X}(y | x) = \frac{f_{X,Y}(x, y)}{f_X(x)}, \quad f_X(x) > 0.$$

De même dans le cas discret : $\mathbb{P}(Y = y_j | X = x_i) = p_{X,Y}(x_i, y_j)/p_X(x_i)$.

8.4 Indépendance

Théorème 8.7 (Caractérisation de l'indépendance). X et Y sont indépendantes si et seulement si :

- **Cas discret** : $p_{X,Y}(x, y) = p_X(x) p_Y(y)$ pour tout (x, y) .
- **Cas continu** : $f_{X,Y}(x, y) = f_X(x) f_Y(y)$ pour tout (x, y) .
- **Via la CDF** : $F_{X,Y}(x, y) = F_X(x) F_Y(y)$ pour tout (x, y) .

Proposition 8.8 (Critère de factorisation). Le couple (X, Y) continu est indépendant si et seulement si il existe des fonctions g, h telles que $f_{X,Y}(x, y) = g(x) h(y)$ sur le support (qui doit être un produit cartésien).

8.5 Covariance et corrélation

Définition 8.9 (Covariance).

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

Proposition 8.10 (Propriétés de la covariance). (i) $\text{Cov}(X, X) = \text{Var}(X)$.

(ii) $\text{Cov}(X, Y) = \text{Cov}(Y, X)$ (symétrie).

(iii) $\text{Cov}(aX + b, cY + d) = ac \text{Cov}(X, Y)$ (bilinéarité).

(iv) $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y)$.

(v) Si $X \perp\!\!\!\perp Y$, alors $\text{Cov}(X, Y) = 0$.

(vi) $\text{Var}(\sum_{i=1}^n X_i) = \sum_{i=1}^n \text{Var}(X_i) + 2 \sum_{i < j} \text{Cov}(X_i, X_j)$.

Définition 8.11 (Coefficient de corrélation).

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X) \text{Var}(Y)}}.$$

Théorème 8.12 (Propriétés de ρ). (i) $-1 \leq \rho(X, Y) \leq 1$ (conséquence de Cauchy-Schwarz).

(ii) $|\rho(X, Y)| = 1$ si et seulement si $Y = aX + b$ p.s. avec $a \neq 0$.

(iii) $\rho(X, Y) = 0$ signifie que X et Y sont **non corrélées**.

Non corrélé $\neq$ indépendant

$\text{Cov}(X, Y) = 0$ n'implique pas l'indépendance. Exemple : si $X \sim \mathcal{N}(0, 1)$ et $Y = X^2$, alors $\text{Cov}(X, Y) = \mathbb{E}[X^3] = 0$ (par symétrie), mais X et Y ne sont évidemment pas indépendants.

8.6 Loi normale bivariée

Définition 8.13 (Normale bivariée). Le vecteur (X, Y) suit une **loi normale bivariée**

$\mathcal{N}_2(\boldsymbol{\mu}, \Sigma)$ avec $\boldsymbol{\mu} = (\mu_X, \mu_Y)^T$ et $\Sigma = \begin{pmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{pmatrix}$ si :

$$f_{X,Y}(x, y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)} \left[\frac{(x-\mu_X)^2}{\sigma_X^2} - 2\rho\frac{(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y} + \frac{(y-\mu_Y)^2}{\sigma_Y^2} \right]\right).$$

Théorème 8.14 (Propriétés de la normale bivariée). (i) Les marginales sont $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$ et $Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$.

(ii) $\rho(X, Y) = \rho$ (le paramètre de la loi).

(iii) X et Y sont indépendantes si et seulement si $\rho = 0$.

(iv) Toute combinaison linéaire $aX + bY$ est normale.

Remarque 8.15. Le point (iii) est spécifique à la loi normale : en général, non corrélé n'implique pas indépendant, mais pour un vecteur gaussien, si.

8.7 Somme de variables aléatoires

Théorème 8.16 (Convolution). *Si X et Y sont indépendantes avec densités f_X et f_Y , alors $Z = X + Y$ a pour densité la **convolution** :*

$$f_Z(z) = (f_X * f_Y)(z) = \int_{-\infty}^{+\infty} f_X(t) f_Y(z - t) dt.$$

Exemple 8.17. Si $X \sim \text{Exp}(\lambda)$ et $Y \sim \text{Exp}(\lambda)$ sont indépendantes, alors pour $z > 0$:

$$f_Z(z) = \int_0^z \lambda e^{-\lambda t} \lambda e^{-\lambda(z-t)} dt = \lambda^2 e^{-\lambda z} \int_0^z dt = \lambda^2 z e^{-\lambda z}.$$

C'est la densité de $\Gamma(2, \lambda)$.

8.8 Exercices

Exercice 8.1. Soit (X, Y) de densité $f(x, y) = cxy$ pour $0 < x < 1$, $0 < y < 1$ et 0 sinon. Déterminer c , les marginales, et vérifier l'indépendance.

Exercice 8.2. Soit (X, Y) de densité $f(x, y) = 2$ pour $0 < x < y < 1$, 0 sinon. Calculer f_X , f_Y , $\mathbb{E}[X]$, $\mathbb{E}[Y]$, $\text{Cov}(X, Y)$, $\rho(X, Y)$.

Exercice 8.3. Montrer que si $(X, Y) \sim \mathcal{N}_2$ avec $\rho = 0$, alors X et Y sont indépendantes (factoriser la densité conjointe).

Exercice 8.4. Soit $X \sim \mathcal{N}(0, 1)$ et $Y = X^2$. Vérifier que $\text{Cov}(X, Y) = 0$ mais que X et Y ne sont pas indépendantes.

Exercice 8.5. Calculer par convolution la densité de $Z = X + Y$ où $X \sim \mathcal{N}(0, 1)$ et $Y \sim \mathcal{N}(0, 1)$ sont indépendantes.

Exercice 8.6. Soit (X, Y) uniforme sur le disque $\{(x, y) : x^2 + y^2 \leq 1\}$. Calculer les marginales, $\mathbb{E}[X]$, $\text{Var}(X)$, $\text{Cov}(X, Y)$.

Exercice 8.7. On tire au hasard un point (X, Y) uniformément dans le triangle de sommets $(0, 0)$, $(1, 0)$, $(0, 1)$. Déterminer la densité conjointe, les marginales, et calculer $\rho(X, Y)$.

Exercice 8.8. Soit X_1, X_2, X_3 i.i.d. $\text{Exp}(1)$. Calculer $\text{Cov}(X_1 + X_2, X_2 + X_3)$.

Chapitre 9

Lois des Grands Nombres

Pourquoi la fréquence observée d'un événement se stabilise-t-elle autour de sa probabilité théorique quand on répète l'expérience un grand nombre de fois ? Cette question, au cœur de la pensée probabiliste, a reçu sa première réponse de Jacob Bernoulli en 1713, dans son *Ars Conjectandi* publié à titre posthume. Bernoulli montre que la fréquence empirique converge vers la probabilité — mais sa démonstration est longue et laborieuse. Tchebychev, en 1867, simplifie considérablement la preuve grâce à son inégalité. Kolmogorov, en 1933, couronne l'édifice avec la *loi forte des grands nombres*, qui garantit la convergence presque sûre. Ce chapitre retrace cette progression, des inégalités de concentration aux différents modes de convergence stochastique.

Intuition

La loi des grands nombres est l'un des résultats les plus fondamentaux de la théorie des probabilités. Elle justifie l'interprétation fréquentiste de la probabilité : la fréquence empirique d'un événement converge vers sa probabilité théorique lorsque le nombre d'expériences augmente.

9.1 Modes de convergence

Définition 9.1 (Convergence en probabilité). Une suite (X_n) **converge en probabilité** vers X , noté $X_n \xrightarrow{\mathbb{P}} X$, si pour tout $\varepsilon > 0$:

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

Définition 9.2 (Convergence presque sûre). (X_n) **converge presque sûrement** vers X , noté $X_n \xrightarrow{p.s.} X$, si :

$$\mathbb{P}\left(\lim_{n \rightarrow \infty} X_n = X\right) = 1,$$

ou de manière équivalente, $\mathbb{P}(\bigcap_{m=1}^{\infty} \bigcup_{n=m}^{\infty} \{|X_n - X| > \varepsilon\}) = 0$ pour tout $\varepsilon > 0$.

Définition 9.3 (Convergence dans L^p). (X_n) **converge dans L^p** ($p \geq 1$) vers X , noté $X_n \xrightarrow{L^p} X$, si :

$$\lim_{n \rightarrow \infty} \mathbb{E}[|X_n - X|^p] = 0.$$

Théorème 9.4 (Hiérarchie des convergences).

$$X_n \xrightarrow{L^p} X \implies X_n \xrightarrow{\mathbb{P}} X, \quad X_n \xrightarrow{p.s.} X \implies X_n \xrightarrow{\mathbb{P}} X.$$

Les réciproques sont fausses en général. En revanche : $X_n \xrightarrow{\mathbb{P}} X$ implique l'existence d'une sous-suite $X_{n_k} \xrightarrow{p.s.} X$.

Convergence p.s. et convergence L^p

La convergence p.s. n'implique pas la convergence L^p , ni réciproquement. Chacune a ses propres conditions.

9.2 Loi faible des grands nombres

Théorème 9.5 (Loi faible des grands nombres (LFGN)). Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires **indépendantes et identiquement distribuées** (i.i.d.) avec $\mathbb{E}[X_1] = \mu$ et $\text{Var}(X_1) = \sigma^2 < \infty$. Posons $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$. Alors :

$$\bar{X}_n \xrightarrow{\mathbb{P}} \mu, \quad \text{i.e.} \quad \forall \varepsilon > 0, \quad \lim_{n \rightarrow \infty} \mathbb{P}(|\bar{X}_n - \mu| > \varepsilon) = 0.$$

Démonstration. C'est une application directe de l'inégalité de Bienaymé–Tchebychev.

$$\mathbb{P}(|\bar{X}_n - \mu| > \varepsilon) \leq \frac{\text{Var}(\bar{X}_n)}{\varepsilon^2} = \frac{\text{Var}(\frac{1}{n} \sum X_i)}{\varepsilon^2} = \frac{1}{n^2} \cdot \frac{n\sigma^2}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2} \xrightarrow{n \rightarrow \infty} 0. \quad \square$$

Remarque 9.6. La LFGN reste vraie sous l'hypothèse plus faible que les X_i sont non corrélées (pas nécessairement indépendantes) avec les mêmes deux premiers moments.

Signification de la LFGN

La moyenne empirique $\bar{X}_n$ se concentre autour de μ quand n est grand. C'est le fondement théorique des méthodes de Monte Carlo et de l'estimation statistique.

9.3 Loi forte des grands nombres

Théorème 9.7 (Loi forte des grands nombres (LFGN forte)). Soit $(X_n)_{n \geq 1}$ une suite i.i.d. avec $\mathbb{E}[|X_1|] < \infty$ et $\mathbb{E}[X_1] = \mu$. Alors :

$$\bar{X}_n \xrightarrow{p.s.} \mu, \quad \text{i.e.} \quad \mathbb{P}\left(\lim_{n \rightarrow \infty} \bar{X}_n = \mu\right) = 1.$$

Remarque 9.8. La loi forte ne nécessite que l'existence du moment d'ordre 1 (pas de variance finie). Sa démonstration complète utilise des outils plus avancés (troncation, lemme de Borel–Cantelli).

Preuve dans le cas $\mathbb{E}[X_1^4] < \infty$. Nous donnons la preuve sous l'hypothèse plus forte $\mathbb{E}[X_1^4] < \infty$. Sans perte de généralité, supposons $\mu = 0$. Posons $S_n = \sum_{i=1}^n X_i$.

$$\mathbb{E}[S_n^4] = \mathbb{E}\left[\left(\sum_{i=1}^n X_i\right)^4\right] = \sum_i \mathbb{E}[X_i^4] + \binom{4}{2} \sum_{i \neq j} \mathbb{E}[X_i^2] \mathbb{E}[X_j^2]$$

(les termes croisés d'ordre impair s'annulent par indépendance et $\mathbb{E}[X_i] = 0$). Donc $\mathbb{E}[S_n^4] = n\mathbb{E}[X_1^4] + 3n(n-1)(\mathbb{E}[X_1^2])^2 \leq Cn^2$ pour une constante C . Par l'inégalité de Markov :

$$\mathbb{P}(|\bar{X}_n| > \varepsilon) = \mathbb{P}(|S_n| > n\varepsilon) \leq \frac{\mathbb{E}[S_n^4]}{n^4\varepsilon^4} \leq \frac{C}{n^2\varepsilon^4}.$$

Comme $\sum_n 1/n^2 < \infty$, le premier lemme de Borel–Cantelli donne $\mathbb{P}(\limsup\{|\bar{X}_n| > \varepsilon\}) = 0$ pour tout $\varepsilon > 0$, d'où $\bar{X}_n \rightarrow 0$ p.s. $\square$

9.4 Applications

9.4.1 Interprétation fréquentiste

Corollaire 9.9. *Soit A un événement de probabilité p . Si l'on répète l'expérience n fois indépendamment, la fréquence empirique $\hat{p}_n = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{A_i}$ converge presque sûrement vers p .*

9.4.2 Méthode de Monte Carlo

Exemple 9.10 (Estimation de π). Soit $(X_i, Y_i)_{i \geq 1}$ i.i.d. uniformes sur $[0, 1]^2$. Posons $Z_i = \mathbf{1}_{X_i^2 + Y_i^2 \leq 1}$. Alors $\mathbb{E}[Z_i] = \pi/4$ et par la loi des grands nombres :

$$\frac{4}{n} \sum_{i=1}^n Z_i \xrightarrow{\text{p.s.}} \pi.$$

9.4.3 Théorème de Glivenko–Cantelli

Théorème 9.11 (Glivenko–Cantelli). *Soit (X_i) i.i.d. de CDF F et $\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{X_i \leq x}$ la fonction de répartition empirique. Alors :*

$$\sup_{x \in \mathbb{R}} |\hat{F}_n(x) - F(x)| \xrightarrow{\text{p.s.}} 0.$$

9.5 Convergence en loi (introduction)

Définition 9.12 (Convergence en loi). (X_n) **converge en loi** vers X , noté $X_n \xrightarrow{\mathcal{L}} X$, si pour toute fonction $f : \mathbb{R} \rightarrow \mathbb{R}$ continue et bornée :

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

De manière équivalente : $F_{X_n}(x) \rightarrow F_X(x)$ en tout point x de continuité de F_X .

Théorème 9.13 (Hiérarchie complète).

$$X_n \xrightarrow{\text{p.s.}} X \implies X_n \xrightarrow{\mathbb{P}} X \implies X_n \xrightarrow{\mathcal{L}} X.$$

La convergence en loi est la plus faible. Elle ne concerne que les lois, pas les variables elles-mêmes.

Résumé des convergences

Convergence	Définition
Presque sûre	$\mathbb{P}(\lim X_n = X) = 1$
En probabilité	$\mathbb{P}(X_n - X > \varepsilon) \rightarrow 0$
Dans L^p	$\mathbb{E}[X_n - X ^p] \rightarrow 0$
En loi	$\mathbb{E}[f(X_n)] \rightarrow \mathbb{E}[f(X)]$ pour f continue bornée

9.6 Exercices

Exercice 9.1. Soit $X_n \sim \mathcal{U}([0, 1/n])$. Montrer que $X_n \xrightarrow{\mathbb{P}} 0$ et que $X_n \xrightarrow{L^2} 0$. La convergence est-elle presque sûre ?

Exercice 9.2. Soit (X_n) i.i.d. avec $\mathbb{P}(X_n = 1) = \mathbb{P}(X_n = -1) = 1/2$. Montrer que $\bar{X}_n \rightarrow 0$ p.s. et que $\mathbb{P}(\bar{X}_n = 0) \rightarrow 0$ quand n est impair. Commenter.

Exercice 9.3. Donner un exemple de suite (X_n) convergeant en probabilité vers 0 mais pas presque sûrement. *Indication* : considérer des indicatrices d'intervalles se déplaçant sur $[0, 1]$.

Exercice 9.4. On tire n nombres indépendamment et uniformément dans $[0, 1]$. Soit $M_n = \max(X_1, \dots, X_n)$. Montrer que $M_n \rightarrow 1$ p.s.

Exercice 9.5. Soit (X_n) i.i.d. de loi exponentielle $\text{Exp}(1)$. Montrer que $\frac{1}{n} \max(X_1, \dots, X_n) \rightarrow 0$ p.s. et même que $\max(X_1, \dots, X_n)/\ln n \rightarrow 1$ p.s.

Exercice 9.6. Soit (X_n) i.i.d. de loi de Cauchy standard. Montrer que $\bar{X}_n$ ne converge vers aucune constante en probabilité. *Indication* : utiliser les fonctions caractéristiques.

Exercice 9.7 (Monte Carlo). On veut estimer $I = \int_0^1 e^{-x^2} dx$ par Monte Carlo. Écrire l'estimateur de Monte Carlo et, en utilisant l'inégalité de Tchebychev, déterminer le nombre de tirages nécessaires pour que l'erreur soit inférieure à 0,01 avec probabilité 0,95.

Exercice 9.8. Soit (X_n) une suite i.i.d. avec $\mathbb{E}[X_1] = 0$ et $\text{Var}(X_1) = 1$. Montrer par la LFGN que $S_n/n \rightarrow 0$ en probabilité, puis que $S_n^2/n^2 \rightarrow 0$ en probabilité. En déduire un résultat sur $S_n/n^{3/4}$.

Chapitre 10

Théorème Central Limite

Pourquoi la loi normale apparaît-elle partout — dans la taille des individus, les erreurs de mesure, les rendements financiers, le bruit thermique ? La réponse tient en un théorème : le *théorème central limite* (TCL). Abraham de Moivre en donne la première version en 1733 pour les lancers de pièces ; Laplace le généralise vers 1810 ; Lindeberg (1922) et Lévy (1925) établissent la version définitive sous des hypothèses minimales. Le TCL affirme que la somme normalisée de variables aléatoires indépendantes et identiquement distribuées, quelle que soit leur loi commune (pourvu que la variance soit finie), converge en distribution vers la loi normale. C'est le résultat le plus célèbre — et le plus utilisé — de toute la théorie des probabilités.

Intuition

Le théorème central limite (TCL) est sans doute le résultat le plus célèbre de la théorie des probabilités. Il affirme que la somme normalisée de variables aléatoires i.i.d. converge en loi vers la loi normale, quelle que soit la loi commune des variables, pourvu que la variance soit finie.

10.1 Énoncé du théorème

Théorème 10.1 (Théorème central limite). *Soit $(X_n)_{n \geq 1}$ une suite i.i.d. avec $\mathbb{E}[X_1] = \mu$ et $\text{Var}(X_1) = \sigma^2 \in (0, \infty)$. Posons $S_n = \sum_{i=1}^n X_i$. Alors :*

$$Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}} = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1).$$

Autrement dit, pour tout $x \in \mathbb{R}$:

$$\lim_{n \rightarrow \infty} \mathbb{P}(Z_n \leq x) = \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2} dt.$$

Pourquoi la loi normale est-elle omniprésente ?

Le TCL explique pourquoi tant de phénomènes naturels suivent approximativement une loi normale : toute grandeur résultant de la somme de nombreuses petites contributions indépendantes sera approximativement gaussienne.

10.2 Démonstration par les fonctions caractéristiques

Nous renvoyons au chapitre 11 pour la théorie complète des fonctions caractéristiques. Ici, nous esquissons la preuve en admettant le théorème de Lévy.

Preuve du TCL. Sans perte de généralité, posons $\mu = 0$ (sinon remplacer X_i par $X_i - \mu$). Soit $\varphi(t) = \mathbb{E}[e^{itX_1}]$ la fonction caractéristique commune des X_i .

Par hypothèse, $\varphi(0) = 1$, $\varphi'(0) = i\mathbb{E}[X_1] = 0$, et $\varphi''(0) = i^2\mathbb{E}[X_1^2] = -\sigma^2$. Le développement de Taylor donne :

$$\varphi(t) = 1 - \frac{\sigma^2 t^2}{2} + o(t^2) \quad (t \rightarrow 0).$$

La fonction caractéristique de $Z_n = S_n/(\sigma\sqrt{n})$ est :

$$\begin{aligned} \varphi_{Z_n}(t) &= \mathbb{E}\left[\exp\left(it \cdot \frac{S_n}{\sigma\sqrt{n}}\right)\right] = \left[\varphi\left(\frac{t}{\sigma\sqrt{n}}\right)\right]^n \\ &= \left[1 - \frac{\sigma^2}{2} \cdot \frac{t^2}{\sigma^2 n} + o\left(\frac{1}{n}\right)\right]^n = \left[1 - \frac{t^2}{2n} + o\left(\frac{1}{n}\right)\right]^n. \end{aligned}$$

En utilisant $\lim_{n \rightarrow \infty} (1 + a/n)^n = e^a$:

$$\varphi_{Z_n}(t) \xrightarrow[n \rightarrow \infty]{} e^{-t^2/2}.$$

Or $e^{-t^2/2}$ est la fonction caractéristique de $\mathcal{N}(0, 1)$. Par le théorème de continuité de Lévy, $Z_n \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1)$. □

10.3 Applications pratiques

10.3.1 Approximation de la loi binomiale

Corollaire 10.2 (Théorème de Moivre–Laplace). Si $S_n \sim \text{Bin}(n, p)$ avec $0 < p < 1$, alors :

$$\frac{S_n - np}{\sqrt{np(1-p)}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1).$$

Exemple 10.3. On lance une pièce équilibrée 100 fois. Quelle est la probabilité d'obtenir entre 45 et 55 Pile ?

Avec $S_{100} \sim \text{Bin}(100, 0,5)$, $\mu = 50$, $\sigma = 5$:

$$\begin{aligned} \mathbb{P}(45 \leq S_{100} \leq 55) &= \mathbb{P}\left(\frac{45 - 50}{5} \leq Z \leq \frac{55 - 50}{5}\right) \approx \mathbb{P}(-1 \leq Z \leq 1) \\ &= 2\Phi(1) - 1 \approx 0,6827. \end{aligned}$$

Correction de continuité

Pour améliorer l'approximation, on utilise la **correction de continuité** :

$$\mathbb{P}(a \leq S_n \leq b) \approx \Phi\left(\frac{b + 0,5 - np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{a - 0,5 - np}{\sqrt{np(1-p)}}\right).$$

10.3.2 Intervalles de confiance

Corollaire 10.4 (Intervalle de confiance pour la moyenne). *Si (X_i) sont i.i.d. avec $\mathbb{E}[X_1] = \mu$ et $\text{Var}(X_1) = \sigma^2$ connus, un **intervalle de confiance à 95%** pour μ est :*

$$\left[\bar{X}_n - 1,96 \frac{\sigma}{\sqrt{n}}, \bar{X}_n + 1,96 \frac{\sigma}{\sqrt{n}} \right].$$

Quantiles courants pour les IC

Niveau de confiance	α	$z_{\alpha/2}$
90%	0,10	1,645
95%	0,05	1,960
99%	0,01	2,576

10.3.3 Approximation de la loi de Poisson

Corollaire 10.5. *Si $X \sim \text{Poisson}(\lambda)$ avec λ grand, alors $(X - \lambda)/\sqrt{\lambda} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1)$.*

10.4 TCL multidimensionnel

Théorème 10.6 (TCL multidimensionnel). *Soit (X_n) une suite i.i.d. de vecteurs aléatoires dans $\mathbb{R}^d$, avec $\mathbb{E}[X_1] = \boldsymbol{\mu}$ et $\text{Cov}(X_1) = \Sigma$. Alors :*

$$\sqrt{n}(\bar{X}_n - \boldsymbol{\mu}) \xrightarrow{\mathcal{L}} \mathcal{N}_d(0, \Sigma).$$

10.5 Méthode delta

Théorème 10.7 (Méthode delta). *Si $\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2)$ et si g est dérivable en μ avec $g'(\mu) \neq 0$, alors :*

$$\sqrt{n}(g(\bar{X}_n) - g(\mu)) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2[g'(\mu)]^2).$$

Exemple 10.8. Si $\bar{X}_n \rightarrow p$ et $g(x) = x(1 - x)$, alors $g'(p) = 1 - 2p$ et :

$$\sqrt{n}(\bar{X}_n(1 - \bar{X}_n) - p(1 - p)) \xrightarrow{\mathcal{L}} \mathcal{N}(0, p(1 - p)(1 - 2p)^2).$$

10.6 Vitesse de convergence : théorème de Berry–Esseen

Théorème 10.9 (Berry–Esseen). *Sous les hypothèses du TCL, si $\mathbb{E}[|X_1|^3] < \infty$:*

$$\sup_{x \in \mathbb{R}} |\mathbb{P}(Z_n \leq x) - \Phi(x)| \leq \frac{C \mathbb{E}[|X_1 - \mu|^3]}{\sigma^3 \sqrt{n}},$$

où C est une constante universelle ($C \leq 0,4748$).

Remarque 10.10. Berry–Esseen montre que la convergence dans le TCL est de l'ordre de $1/\sqrt{n}$. Cela justifie l'utilisation de l'approximation normale dès que n est modérément grand (typiquement $n \geq 30$).

10.7 Exercices

Exercice 10.1. Soit (X_i) i.i.d. avec $\mathbb{E}[X_1] = 2$ et $\text{Var}(X_1) = 9$. Donner une approximation de $\mathbb{P}(S_{100} > 220)$.

Exercice 10.2. Un dé équilibré est lancé 360 fois. Approximer la probabilité que la somme dépasse 1300.

Exercice 10.3. Soit $S_n \sim \text{Bin}(400, 0,4)$. Calculer approximativement $\mathbb{P}(S_n \geq 180)$ avec et sans correction de continuité.

Exercice 10.4. On mesure une grandeur physique $n = 50$ fois. Les mesures sont i.i.d. de variance $\sigma^2 = 4$. Quel est l'intervalle de confiance à 99% pour la moyenne ?

Exercice 10.5. Combien de sondages faut-il réaliser pour estimer une proportion p à 2% près avec une confiance de 95% ? (On ne connaît pas p .)

Exercice 10.6. En utilisant la méthode delta, donner la loi limite de $\sqrt{n}(\ln \bar{X}_n - \ln \mu)$ lorsque les X_i sont i.i.d. positives avec $\mathbb{E}[X_1] = \mu > 0$ et $\text{Var}(X_1) = \sigma^2$.

Exercice 10.7. Vérifier le TCL pour $X_i \sim \text{Exp}(\lambda)$ en calculant directement la fonction caractéristique de Z_n et en montrant qu'elle converge vers $e^{-t^2/2}$.

Exercice 10.8. Soit (X_i) i.i.d. Poisson(4). Approximer $\mathbb{P}(\sum_{i=1}^{100} X_i \leq 380)$.

Chapitre 11

Fonctions Caractéristiques

Intuition

La fonction caractéristique est un outil analytique puissant qui transforme les problèmes probabilistes en problèmes d'analyse complexe. Elle existe toujours (contrairement à la MGF), caractérise entièrement la loi, et permet de démontrer élégamment le théorème central limite.

11.1 Définition et premières propriétés

Définition 11.1 (Fonction caractéristique). La **fonction caractéristique** d'une variable aléatoire réelle X est :

$$\varphi_X(t) = \mathbb{E}[e^{itX}] = \mathbb{E}[\cos(tX)] + i \mathbb{E}[\sin(tX)], \quad t \in \mathbb{R}.$$

Théorème 11.2 (Propriétés fondamentales). Soit φ_X la fonction caractéristique de X . Alors :

- (i) $\varphi_X(0) = 1$.
- (ii) $|\varphi_X(t)| \leq 1$ pour tout $t \in \mathbb{R}$.
- (iii) φ_X est **uniformément continue** sur $\mathbb{R}$.
- (iv) $\varphi_X(-t) = \overline{\varphi_X(t)}$ (conjugué complexe).
- (v) Si $Y = aX + b$, alors $\varphi_Y(t) = e^{ibt} \varphi_X(at)$.
- (vi) Si $X \perp\!\!\!\perp Y$, alors $\varphi_{X+Y}(t) = \varphi_X(t) \varphi_Y(t)$.

Preuve de (iii).

$$|\varphi_X(t+h) - \varphi_X(t)| = |\mathbb{E}[e^{itX}(e^{ihX} - 1)]| \leq \mathbb{E}[|e^{ihX} - 1|].$$

Comme $|e^{ihX} - 1| \leq 2$ et $e^{ihX} - 1 \rightarrow 0$ quand $h \rightarrow 0$, la convergence dominée donne $\mathbb{E}[|e^{ihX} - 1|] \rightarrow 0$ uniformément en t . $\square$

11.2 Calcul des moments

Théorème 11.3 (Moments et dérivées). *Si $\mathbb{E}[|X|^n] < \infty$, alors φ_X est n fois dérivable et :*

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

En particulier : $\mathbb{E}[X] = \varphi_X'(0)/i$ et $\mathbb{E}[X^2] = -\varphi_X''(0)$, d'où $\text{Var}(X) = -\varphi_X''(0) - [\varphi_X'(0)/i]^2$.

11.3 Fonctions caractéristiques des lois classiques

Table des fonctions caractéristiques

Loi de X	$\varphi_X(t)$
Bernoulli(p)	$1 - p + pe^{it}$
Bin(n, p)	$(1 - p + pe^{it})^n$
Poisson(λ)	$\exp(\lambda(e^{it} - 1))$
Geom(p)	$\frac{pe^{it}}{1 - (1 - p)e^{it}}$
$\mathcal{U}([a, b])$	$\frac{e^{itb} - e^{ita}}{it(b - a)}$
Exp(λ)	$\frac{\lambda}{\lambda - it}$
$\mathcal{N}(\mu, \sigma^2)$	$\exp\left(i\mu t - \frac{\sigma^2 t^2}{2}\right)$
$\Gamma(\alpha, \lambda)$	$\left(\frac{\lambda}{\lambda - it}\right)^\alpha$
Cauchy($0, 1$)	$e^{- t }$

Exemple 11.4 (Calcul pour la loi normale). Si $X \sim \mathcal{N}(0, 1)$:

$$\begin{aligned} \varphi_X(t) &= \int_{-\infty}^{+\infty} e^{itx} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-(x^2 - 2itx)/2} dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-(x - it)^2/2} e^{-t^2/2} dx = e^{-t^2/2}. \end{aligned}$$

Pour $X \sim \mathcal{N}(\mu, \sigma^2)$, $X = \sigma Z + \mu$, donc $\varphi_X(t) = e^{i\mu t} e^{-\sigma^2 t^2/2} = e^{i\mu t - \sigma^2 t^2/2}$.

11.4 Théorèmes d'unicité et d'inversion

Théorème 11.5 (Théorème d'unicité). *Deux variables aléatoires X et Y ont même loi si et seulement si $\varphi_X(t) = \varphi_Y(t)$ pour tout $t \in \mathbb{R}$.*

Théorème 11.6 (Formule d'inversion). *Si $\varphi_X \in L^1(\mathbb{R})$ (i.e. $\int |\varphi_X(t)| dt < \infty$), alors X admet une densité continue donnée par :*

$$f_X(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-itx} \varphi_X(t) dt.$$

Remarque 11.7. La formule d'inversion est l'analogue probabiliste de la transformée de Fourier inverse. La fonction caractéristique n'est autre que la transformée de Fourier de la loi de X .

11.5 Théorème de continuité de Lévy

Théorème 11.8 (Théorème de continuité de Lévy). *Soit (X_n) une suite de variables aléatoires et $\varphi_n = \varphi_{X_n}$.*

- (i) *Si $X_n \xrightarrow{\mathcal{L}} X$, alors $\varphi_n(t) \rightarrow \varphi_X(t)$ pour tout t .*
- (ii) *Réciproquement, si $\varphi_n(t) \rightarrow \psi(t)$ pour tout t et si ψ est **continue en 0**, alors ψ est la fonction caractéristique d'une variable X et $X_n \xrightarrow{\mathcal{L}} X$.*

Remarque 11.9. C'est ce théorème qui est utilisé dans la preuve du TCL (chapitre 10) : on montre la convergence ponctuelle de $\varphi_{Z_n}(t) \rightarrow e^{-t^2/2}$, et la continuité en 0 de la limite est évidente.

11.6 Application : preuve de stabilité de la loi normale

Théorème 11.10 (Stabilité de la loi normale). *Si $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$ et $X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$ sont indépendantes, alors :*

$$X_1 + X_2 \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2).$$

Démonstration.

$$\begin{aligned} \varphi_{X_1+X_2}(t) &= \varphi_{X_1}(t) \varphi_{X_2}(t) \\ &= e^{i\mu_1 t - \sigma_1^2 t^2/2} e^{i\mu_2 t - \sigma_2^2 t^2/2} \\ &= e^{i(\mu_1 + \mu_2)t - (\sigma_1^2 + \sigma_2^2)t^2/2}, \end{aligned}$$

qui est la FC de $\mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$. Le résultat suit par unicité. □

11.7 Application : stabilité de la loi de Poisson

Proposition 11.11. Si $X_1 \sim \text{Poisson}(\lambda_1)$ et $X_2 \sim \text{Poisson}(\lambda_2)$ sont indépendantes, alors $X_1 + X_2 \sim \text{Poisson}(\lambda_1 + \lambda_2)$.

Démonstration. $\varphi_{X_1+X_2}(t) = e^{\lambda_1(e^{it}-1)} \cdot e^{\lambda_2(e^{it}-1)} = e^{(\lambda_1+\lambda_2)(e^{it}-1)}$, qui est la FC de $\text{Poisson}(\lambda_1 + \lambda_2)$. □

11.8 Fonction caractéristique et convergence en loi

Exemple 11.12 (Convergence de Poisson vers la normale). Soit $X_n \sim \text{Poisson}(n)$. Posons $Z_n = (X_n - n)/\sqrt{n}$.

$$\begin{aligned}\varphi_{Z_n}(t) &= e^{-itn/\sqrt{n}} \varphi_{X_n}(t/\sqrt{n}) = e^{-it\sqrt{n}} e^{n(e^{it/\sqrt{n}}-1)} \\ &= \exp\left(n\left(e^{it/\sqrt{n}} - 1 - \frac{it}{\sqrt{n}}\right)\right).\end{aligned}$$

Par développement : $e^{it/\sqrt{n}} = 1 + it/\sqrt{n} - t^2/(2n) + o(1/n)$, d'où :

$$\varphi_{Z_n}(t) = \exp\left(n\left(-\frac{t^2}{2n} + o(1/n)\right)\right) \rightarrow e^{-t^2/2}.$$

Donc $Z_n \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1)$ par le théorème de Lévy.

11.9 Exercices

Exercice 11.1. Calculer la fonction caractéristique de la loi $\mathcal{U}(\{-1, 0, 1\})$.

Exercice 11.2. Soit $X \sim \text{Exp}(\lambda)$. Calculer $\varphi_X(t)$ et en déduire $\mathbb{E}[X]$, $\mathbb{E}[X^2]$, $\text{Var}(X)$.

Exercice 11.3. Montrer que $\varphi_X(t) = e^{-|t|}$ est la fonction caractéristique de la loi de Cauchy standard. En déduire que si X_1, X_2 sont i.i.d. Cauchy standard, alors $(X_1 + X_2)/2$ est encore Cauchy standard.

Exercice 11.4. En utilisant la formule d'inversion, retrouver la densité de la loi $\mathcal{N}(0, 1)$ à partir de sa FC $\varphi(t) = e^{-t^2/2}$.

Exercice 11.5. Montrer que la FC de la loi $\Gamma(\alpha, \lambda)$ est $(\lambda/(\lambda - it))^\alpha$. En déduire que la somme de n exponentielles i.i.d. de paramètre λ suit une loi $\Gamma(n, \lambda)$.

Exercice 11.6. Soit $X_n \sim \text{Bin}(n, \lambda/n)$ avec $\lambda > 0$ fixé. En calculant $\varphi_{X_n}(t)$, retrouver l'approximation de Poisson : $X_n \xrightarrow{\mathcal{L}} \text{Poisson}(\lambda)$.

Exercice 11.7. Montrer que si φ_X est à valeurs réelles, alors X et $-X$ ont même loi (la loi de X est symétrique autour de 0).

Exercice 11.8. Soit (X_n) i.i.d. avec $\mathbb{E}[X_1] = 0$, $\text{Var}(X_1) = \sigma^2$ et $\varphi_{X_1}(t) = 1/(1 + t^2)$ (Cauchy). Montrer que $\bar{X}_n$ a même loi que X_1 et que le TCL ne s'applique pas (pourquoi?).

Chapitre 12

Introduction aux Chaînes de Markov

12.1 Introduction

Les chaînes de Markov constituent l'un des modèles stochastiques les plus fondamentaux. Elles modélisent l'évolution aléatoire d'un système dont le futur ne dépend du passé qu'à travers le présent (propriété de Markov). Leurs applications couvrent la physique (marches aléatoires), la biologie (évolution de populations), l'informatique (algorithmes de Monte-Carlo, PageRank) et la finance.

La propriété de Markov : absence de mémoire

Imaginez un pion sur un échiquier qui se déplace aléatoirement. À chaque étape, la probabilité de se déplacer vers une case voisine ne dépend que de la position actuelle du pion, pas de la trajectoire passée. C'est la propriété de Markov : « le futur est indépendant du passé, conditionné au présent ».

12.2 Définitions

Définition 12.1 (Chaîne de Markov). Soit E un ensemble dénombrable (l'*espace d'états*). Un processus stochastique $(X_n)_{n \geq 0}$ à valeurs dans E est une *chaîne de Markov* si pour tous $n \geq 0$ et tous états $i_0, \dots, i_{n-1}, i, j \in E$:

$$\mathbb{P}(X_{n+1} = j \mid X_n = i, X_{n-1} = i_{n-1}, \dots, X_0 = i_0) = \mathbb{P}(X_{n+1} = j \mid X_n = i),$$

dès que le conditionnement est bien défini ($\mathbb{P}(X_n = i, \dots) > 0$).

Définition 12.2 (Chaîne homogène et matrice de transition). La chaîne est dite *homogène* (ou *stationnaire*) si les probabilités de transition ne dépendent pas de n :

$$p_{ij} = \mathbb{P}(X_{n+1} = j \mid X_n = i), \quad \forall n \geq 0.$$

La matrice $P = (p_{ij})_{i,j \in E}$ est appelée *matrice de transition*. Elle vérifie :

- (i) $p_{ij} \geq 0$ pour tous i, j ;
- (ii) $\sum_{j \in E} p_{ij} = 1$ pour tout i (matrice stochastique).

Définition 12.3 (Loi initiale). La *loi initiale* est la loi de X_0 , c'est-à-dire le vecteur $\mu_0 = (\mathbb{P}(X_0 = i))_{i \in E}$. La loi de la chaîne est entièrement déterminée par μ_0 et P .

Proposition 12.4 (Loi jointe). Pour tous $i_0, i_1, \dots, i_n \in E$:

$$\mathbb{P}(X_0 = i_0, X_1 = i_1, \dots, X_n = i_n) = \mu_0(i_0) p_{i_0 i_1} p_{i_1 i_2} \cdots p_{i_{n-1} i_n}.$$

Exemple 12.5 (Marche aléatoire sur $\mathbb{Z}$). Soit $X_0 = 0$ et $X_{n+1} = X_n + \xi_{n+1}$ où les (ξ_n) sont i.i.d. avec $\mathbb{P}(\xi = +1) = p$ et $\mathbb{P}(\xi = -1) = 1 - p$. C'est une chaîne de Markov sur $E = \mathbb{Z}$ avec $p_{i, i+1} = p$ et $p_{i, i-1} = 1 - p$.

Exemple 12.6 (Modèle d'Ehrenfest). $2N$ particules sont réparties entre deux boîtes. À chaque étape, on choisit une particule au hasard et on la déplace dans l'autre boîte. Si X_n est le nombre de particules dans la première boîte, (X_n) est une chaîne de Markov sur $\{0, 1, \dots, 2N\}$ avec $p_{k, k-1} = k/(2N)$ et $p_{k, k+1} = 1 - k/(2N)$.

12.3 Probabilités de transition en n pas

Définition 12.7. Les *probabilités de transition en n pas* sont :

$$p_{ij}^{(n)} = \mathbb{P}(X_n = j \mid X_0 = i).$$

Théorème 12.8 (Équation de Chapman-Kolmogorov). Pour tous $m, n \geq 0$ et tous $i, j \in E$:

$$p_{ij}^{(m+n)} = \sum_{k \in E} p_{ik}^{(m)} p_{kj}^{(n)}.$$

En notation matricielle : $P^{(m+n)} = P^m \cdot P^n$, soit $P^{(n)} = P^n$.

Démonstration. Par la formule des probabilités totales :

$$\begin{aligned} p_{ij}^{(m+n)} &= \mathbb{P}(X_{m+n} = j \mid X_0 = i) = \sum_{k \in E} \mathbb{P}(X_{m+n} = j \mid X_m = k) \mathbb{P}(X_m = k \mid X_0 = i) \\ &= \sum_{k \in E} p_{kj}^{(n)} p_{ik}^{(m)}. \end{aligned}$$

□

12.4 Classification des états

Définition 12.9 (Accessibilité et communication). • L'état j est *accessible* depuis i , noté $i \rightarrow j$, s'il existe $n \geq 0$ tel que $p_{ij}^{(n)} > 0$.

- Deux états i et j *communiquent*, noté $i \leftrightarrow j$, si $i \rightarrow j$ et $j \rightarrow i$.

La relation $\leftrightarrow$ est une relation d'équivalence. Ses classes sont les *classes de communication*.

Définition 12.10 (Chaîne irréductible). La chaîne est *irréductible* si tous les états communiquent, c'est-à-dire s'il n'y a qu'une seule classe de communication.

Définition 12.11 (Temps de retour et récurrence). Le *temps de premier retour* en i est : $T_i = \inf\{n \geq 1 : X_n = i\}$.

- L'état i est *récurent* si $\mathbb{P}_i(T_i < \infty) = 1$.

- L'état i est *transitoire* si $\mathbb{P}_i(T_i < \infty) < 1$.

Théorème 12.12 (Caractérisation de la récurrence). *Un état i est récurrent si et seulement si $\sum_{n=0}^{\infty} p_{ii}^{(n)} = +\infty$. Un état i est transitoire si et seulement si $\sum_{n=0}^{\infty} p_{ii}^{(n)} < +\infty$.*

Démonstration. Soit $N_i = \sum_{n=1}^{\infty} \mathbb{1}_{\{X_n=i\}}$ le nombre de retours en i . On a $\mathbb{E}_i[N_i] = \sum_{n=1}^{\infty} p_{ii}^{(n)}$. De plus, $\mathbb{P}_i(N_i \geq k) = \mathbb{P}_i(T_i < \infty)^k$, donc N_i suit une loi géométrique. Si i est récurrent, $\mathbb{P}_i(T_i < \infty) = 1$, donc $\mathbb{E}_i[N_i] = +\infty$. Si i est transitoire, $\mathbb{P}_i(T_i < \infty) < 1$, donc $\mathbb{E}_i[N_i] = \mathbb{P}_i(T_i < \infty)/(1 - \mathbb{P}_i(T_i < \infty)) < +\infty$. $\square$

Proposition 12.13 (Propriété de classe). La récurrence et la transience sont des propriétés de classe : si i communique avec j , alors i est récurrent si et seulement si j l'est.

Définition 12.14 (État absorbant). Un état i est *absorbant* si $p_{ii} = 1$. Un état absorbant est récurrent.

Définition 12.15 (Période). La *période* d'un état i est $d(i) = \gcd\{n \geq 1 : p_{ii}^{(n)} > 0\}$. L'état i est *apériodique* si $d(i) = 1$. La période est une propriété de classe.

12.5 Distributions stationnaires

Définition 12.16 (Distribution stationnaire). Un vecteur de probabilité $\pi = (\pi_i)_{i \in E}$ (avec $\pi_i \geq 0$ et $\sum_i \pi_i = 1$) est une *distribution stationnaire* (ou *invariante*) pour la matrice P si :

$$\pi P = \pi, \quad \text{c'est-à-dire} \quad \pi_j = \sum_{i \in E} \pi_i p_{ij}, \quad \forall j \in E.$$

Théorème 12.17 (Existence pour les chaînes irréductibles récurrentes positives). *Soit (X_n) une chaîne de Markov irréductible et récurrente. Alors :*

1. *Il existe une mesure invariante π (unique à un facteur multiplicatif près) donnée par $\pi_j = 1/\mathbb{E}_j[T_j]$.*
2. *π est une distribution de probabilité (i.e. $\sum_j \pi_j = 1$) si et seulement si la chaîne est récurrente positive, c'est-à-dire $\mathbb{E}_i[T_i] < \infty$ pour tout (ou de façon équivalente, pour un) état i .*

Remarque 12.18. Si l'espace d'états E est fini et la chaîne est irréductible, alors elle est automatiquement récurrente positive et admet une unique distribution stationnaire.

12.6 Théorème ergodique

Théorème 12.19 (Théorème ergodique pour les chaînes de Markov). *Soit (X_n) une chaîne de Markov irréductible, récurrente positive et apériodique, de distribution stationnaire π . Alors pour tout état initial i et tout état j :*

$$\lim_{n \rightarrow \infty} p_{ij}^{(n)} = \pi_j.$$

De plus, pour toute fonction bornée $f : E \rightarrow \mathbb{R}$:

$$\frac{1}{n} \sum_{k=0}^{n-1} f(X_k) \xrightarrow[n \rightarrow \infty]{p.s.} \sum_{j \in E} f(j) \pi_j.$$

Formules clés des chaînes de Markov

- **Chapman-Kolmogorov** : $P^{(m+n)} = P^m \cdot P^n$
- **Distribution au temps n** : $\mu_n = \mu_0 P^n$
- **Stationnarité** : $\pi P = \pi$
- **Lien avec le temps de retour** : $\pi_j = 1/\mathbb{E}_j[T_j]$
- **Convergence** : $\lim_{n \rightarrow \infty} P^n = \mathbf{1}\pi^T$ (si irréductible, récurrente positive, apériodique)

12.7 Réversibilité et équations de bilan détaillé

Définition 12.20 (Réversibilité). Une chaîne de Markov de matrice de transition P et de distribution stationnaire π est dite *réversible* si les *équations de bilan détaillé* sont satisfaites :

$$\pi_i p_{ij} = \pi_j p_{ji}, \quad \forall i, j \in E.$$

Proposition 12.21. Si π vérifie les équations de bilan détaillé, alors π est une distribution stationnaire.

Démonstration. En sommant sur i : $\sum_i \pi_i p_{ij} = \sum_i \pi_j p_{ji} = \pi_j \sum_i p_{ji} = \pi_j$, car $\sum_i p_{ji} = 1$. $\square$

Exemple 12.22 (Marche aléatoire sur un graphe). Soit $G = (V, E)$ un graphe connexe fini non orienté. La marche aléatoire simple sur G a pour matrice de transition $p_{ij} = 1/\deg(i)$ si (i, j) est une arête. La distribution stationnaire est $\pi_i = \deg(i)/(2|E|)$, et les équations de bilan détaillé sont satisfaites.

Bilan détaillé $\neq$ stationnarité

Le bilan détaillé est une condition *suffisante* mais pas nécessaire pour la stationnarité. Il existe des chaînes de Markov avec une distribution stationnaire qui ne vérifie pas le bilan détaillé (par exemple, les chaînes avec un cycle orienté).

12.8 Exercices

Exercice 12.1. Soit (X_n) la marche aléatoire symétrique sur $\mathbb{Z}$ ($p = 1/2$). Montrer que chaque état est récurrent. *Indication* : utiliser $p_{00}^{(2n)} = \binom{2n}{n} 2^{-2n} \sim (\pi n)^{-1/2}$.

Exercice 12.2. Montrer que la marche aléatoire symétrique sur $\mathbb{Z}^3$ est transitoire. *Indication* : montrer que $\sum_n p_{00}^{(n)} < \infty$.

Exercice 12.3. Soit $E = \{1, 2, 3\}$ et $P = \begin{pmatrix} 0 & 1/2 & 1/2 \\ 1/3 & 1/3 & 1/3 \\ 1/2 & 0 & 1/2 \end{pmatrix}$.

1. Vérifier que la chaîne est irréductible et apériodique.

2. Calculer la distribution stationnaire π .

3. La chaîne est-elle réversible ?

Exercice 12.4 (Modèle de file d'attente). Soit $N \geq 1$ la capacité d'une file d'attente. À chaque étape, un client arrive avec probabilité λ et un client est servi avec probabilité μ (si la file n'est pas vide), indépendamment. Soit X_n le nombre de clients au temps n . Montrer que (X_n) est une chaîne de Markov, écrire sa matrice de transition, et trouver sa distribution stationnaire lorsque $\lambda < \mu$.

Exercice 12.5 (Algorithme de Metropolis-Hastings). Soit π une distribution de probabilité cible sur un ensemble fini E et soit Q une matrice de transition irréductible (la *proposition*). On définit la matrice de transition :

$$p_{ij} = q_{ij} \min\left(1, \frac{\pi_j q_{ji}}{\pi_i q_{ij}}\right) \quad (i \neq j), \quad p_{ii} = 1 - \sum_{j \neq i} p_{ij}.$$

Montrer que π est une distribution stationnaire pour P et que le bilan détaillé est satisfait.

Exercice 12.6. Soit (X_n) une chaîne de Markov irréductible sur un espace d'états fini E avec $|E| = N$. Montrer qu'il existe un unique vecteur stationnaire π et que $\pi_i > 0$ pour tout i .

Bibliographie

- [1] Billingsley, P., *Probability and Measure*, 3rd ed., Wiley, 1995.
- [2] Feller, W., *An Introduction to Probability Theory and Its Applications, Vol. I*, 3rd ed., Wiley, 1968.
- [3] Feller, W., *An Introduction to Probability Theory and Its Applications, Vol. II*, 2nd ed., Wiley, 1971.
- [4] Durrett, R., *Probability : Theory and Examples*, 5th ed., Cambridge, 2019.
- [5] Dacunha-Castelle, D. and Duflo, M., *Probabilités et Statistiques*, 2 vols., Masson, 1982–1983.