

Rappels de probabilités
Master 1 Mathématiques Fondamentales et Applications

Bastien Mallein

Année 2020–2021

Avant-propos

L'objectif de ce support de cours est de fournir rapidement l'essentiel des informations nécessaires à la poursuite d'étude au niveau master d'un étudiant ayant déjà suivi un cours de licence de théorie de la mesure et/ou de probabilités. Pour cette raison, la plupart des résultats mentionnés sont rédigés sans donner trop de détails, et sous l'hypothèse d'une connaissance suffisante des bases de la théorie de la mesure et de l'intégration. Les preuves ne sont esquissées que dans l'hypothèses où elles apportent une information importante, ou forment un exemple de manipulations typiques nécessaire en théorie des probabilités.

Ce polycopié n'a pas de prétentions à être original, et emprunte en grande partie à d'autres manuscrits d'introductions aux probabilités. Toute personne intéressée peut trouver des références supplémentaires dans l'excellent document *Intégration, Probabilités et Processus Aléatoires* de Jean-François Le Gall. Des compléments intéressants peuvent également être trouvés dans les livres *Foundation of Modern Probability* d'Olav Kallenberg ou *Probability: Theory and Example* de Rick Durrett.

Table des matières

1	Espace de probabilité	2
1.1	Univers de probabilités	2
1.2	Tribu sur un univers de probabilités	3
1.3	Loi de probabilité	5
1.4	Exercices	7
2	Variables aléatoires	9
2.1	Fonctions mesurables	9
2.2	Variables aléatoires et leurs lois	10
2.3	Vecteurs aléatoires	11
2.4	Espérance, variance	12
2.5	Quelques inégalités sur les espérances	14
2.6	Lois usuelles	14
2.6.1	Lois discrètes	14
2.6.2	Lois à densité	15
2.7	Exercices	16
3	Indépendance	18
3.1	Indépendance d'événements	18
3.2	Indépendances de variables et tribus	19
3.3	Théorèmes de Fubini	21
3.4	Lemme de Borel-Cantelli	21
3.5	Exercices	21
4	Convergence	23
4.1	Préliminaires sur la convergence	23
4.2	Différents modes de convergence	23
4.3	Relations entre les modes de convergence	24
4.4	Convergence de sommes de variables aléatoires	25
4.5	Exercices	25

Chapitre 1

Espace de probabilité

La théorie des probabilités a pour but de modéliser, d'étudier et de prédire des observations liées au résultat d'une expérience aléatoire (lancer d'une pièce, jet d'un dé, résultat d'une élection, désintégration d'un atome radioactif, etc.). L'expérience aléatoire est modélisée grâce à un espace de probabilités, dont on rappelle ici la définition.

Définition 1.0.1. Un espace de probabilités est un triplet $(\Omega, \mathcal{F}, \mathbf{P})$, où Ω est un univers de probabilité, \mathcal{F} est une tribu sur Ω et \mathbf{P} est une mesure de probabilité sur l'espace mesurable (Ω, \mathcal{F}) .

Le reste de ce chapitre est dédié à la définitions des trois éléments formant l'espace de probabilité. On introduira ainsi d'abord l'univers de probabilité, qui permet de décrire l'ensemble des réalisations possibles de l'expérience aléatoire. On introduira ensuite la tribu, qui décrit l'ensemble des événements observables de l'expérience aléatoire. Enfin, la mesure de probabilités est introduite, qui permet de mesurer la vraisemblance avec laquelle un événement peut se produire.

1.1 Univers de probabilités

L'univers de probabilité est l'espace sur lequel l'expérience aléatoire d'intérêt est construite. Cet espace est construit comme regroupant l'ensemble des résultats possibles impliquant cette expérience aléatoire.

Définition 1.1.1. Un *univers de probabilité* Ω est défini comme l'ensemble des éventualités d'une expérience aléatoire. Un élément $\omega \in \Omega$ est appelé une *éventualité*, et représente une issue possible pour l'expérience aléatoire.

Un univers de probabilité est dit discret si Ω est fini ou dénombrable. Dans le cas contraire, on parle d'univers de probabilité continu. La plupart des univers de probabilités discrets considérés ici sont choisis comme des sous-ensembles de \mathbb{Z}^n pour un certain $n \in \mathbb{N}$, tandis que les univers continus seront choisis comme des sous-ensembles de \mathbb{R}^n pour $n \in \mathbb{N} \cup \{\infty\}$.

Exemple 1.1.2 (Quelques univers de probabilité.).

1. $\Omega = \{1, 2, 3, 4, 5, 6\}$ est un univers de probabilité pour l'expérience aléatoire consistant au jet d'un dé à 6 faces. Chaque entier $j \in \Omega$ correspond à l'éventualité "le dé tombe sur la face j ".
2. $\Omega = \{P, F\}^3$ est un univers de probabilités pour l'expérience aléatoire consistant au jet de 3 pièces de monnaie. L'éventualité $(P, P, P) \in \Omega$ correspond à "les trois pièces tombent sur pile".
3. $\Omega = \{(1, 2, 3), (1, 3, 2), (2, 1, 3), (2, 3, 1), (3, 1, 2), (3, 2, 1)\}$ est un univers de probabilité pour l'expérience aléatoire consistant à observer l'ordre d'arrivée d'une course de trois coureurs. Chaque éventualité correspond à un ordre d'arrivée.
4. $\Omega = \mathbb{R}_+$ est un univers de probabilité pour l'expérience consistant à observer la durée de vie d'une ampoule.

5. $\Omega = \{P, F\}^{\mathbb{N}}$ est un univers de probabilité pour l'expérience consistant au jet d'une infinité de pièces de monnaie.

Remarque 1.1.3. Lorsqu'un univers de probabilité est le plus petit univers possible pour représenter cette expérience aléatoire, on parle d'*univers canonique*. Les univers décrits ci-dessus sont tous des univers canoniques, mais on aurait pu décrire ces expériences aléatoires comme ayant un univers de probabilité \mathbb{R} .

Il n'existe pas un unique univers de probabilités permettant de représenter une expérience aléatoire, et différentes constructions sont possibles. On fera souvent le choix de poser l'univers de probabilités sans préciser en détails les éventualités qu'il contient (sauf si on parle de l'univers canonique).

Exemple 1.1.4. L'univers de probabilité $\Omega = [0, 2\pi)$ est un univers de probabilité qui peut être utilisé pour modéliser l'expérience aléatoire correspondant au jet d'une pièce. Chaque éventualité $\theta \in \Omega$ correspond à l'angle que la pièce fait avec la main lorsqu'elle est rattrapée. La pièce tombe donc sur pile si $\theta \in [0, \pi)$ et sur face si $\theta \in [\pi, 2\pi)$. Cet univers n'est pas l'univers canonique $\{P, F\}$ utilisé pour représenter cette expérience.

1.2 Tribu sur un univers de probabilités

Une tribu sur un univers de probabilités Ω est une collection de sous-ensembles de Ω qui correspond aux événements observables de l'expérience aléatoire. La tribu regroupe l'ensemble des événements d'une expérience aléatoire auxquels on peut associer une probabilité d'occurrence. Elle satisfait les conditions suivantes.

Définition 1.2.1. Une *tribu* \mathcal{F} sur Ω est un sous-ensemble de $\mathcal{P}(\Omega)$ satisfaisant les propriétés suivantes :

1. $\Omega \in \mathcal{F}$,
2. pour tout $A \in \mathcal{F}$, $A^c \in \mathcal{F}$,
3. pour toute suite $(A_n, n \geq 0) \in \mathcal{F}^{\mathbb{N}}$, $\cup_{n \in \mathbb{N}} A_n \in \mathcal{F}$.

En d'autres termes, \mathcal{F} contient l'ensemble total, et est stable par passage au complémentaire et par union dénombrable.

Un ensemble $A \in \mathcal{F}$ est appelé un *événement* de la tribu. Les événements d'une tribu correspondent à l'ensemble des observations possibles grâce à cette tribu. Pour tout $\omega \in \Omega$, on appelle un événement de la forme $\{\omega\}$ un événement élémentaire. Le couple (Ω, \mathcal{F}) est appelé un *espace mesuré*.

Remarque 1.2.2. Si \mathcal{F} est une tribu, alors

1. on a $\emptyset \in \mathcal{F}$, grâce aux points 1 et 2 de la définition ;
2. en posant $A_n = \emptyset$ pour tout $n \geq n_0$, on observe que \mathcal{F} est également stable par union finie ;
3. en passant au complémentaire, on obtient que \mathcal{F} est stable par intersections finies ou dénombrables.

On utilise une tribu sur Ω pour modéliser une connaissance accessible sur l'expérience aléatoire. Les ensembles $A \in \mathcal{F}$ représentent alors les événements pour lesquels, avec le niveau de connaissance considéré, on peut répondre à la question : est-ce que la réalisation $\omega \in \Omega$ appartient à A ou à A^c . On utilise différentes tribus pour modéliser différents niveaux de connaissances sur une même expérience aléatoire.

Exemple 1.2.3 (Quelques tribus classiques.).

1. La tribu $\mathcal{F} = \{\emptyset, \Omega\}$ est appelée la tribu triviale, et correspond à l'absence totale de connaissance sur un événement.
2. Pour $A \subset \Omega$, la tribu $\mathcal{F} = \{\emptyset, A, A^c, \Omega\}$ est une tribu correspondant à l'observation de l'événement A .

3. La tribu $\mathcal{F} = \mathcal{P}(\Omega)$ est la tribu la plus fine définissable, et correspond à l'observation de l'intégralité de l'expérience aléatoire.
4. Dans l'exemple 1.1.4, la tribu associée à l'observation du côté duquel la pièce est tombée est donnée par $\mathcal{F} = \{\emptyset, [0, \pi), [\pi, 2\pi), \Omega\}$.

Une tribu est une collection de sous-ensembles de Ω . On peut ainsi lui appliquer les opérations ensemblistes d'union et d'intersection. Il est intéressant de remarquer que si une tribu n'est pas stable par union, elle est stable par intersections.

Proposition 1.2.4. Soit $(\mathcal{F}_i)_{i \in I}$ une famille de tribus sur Ω , alors $\mathcal{C} = \bigcap_{i \in I} \mathcal{F}_i$ est une tribu.

Cette proposition permet alors de définir la plus petite tribu satisfaisant une propriété, qui est définie par l'intersection de toutes les tribus satisfaisant cette propriété. On peut ainsi définir la notion de tribu engendrée.

Définition 1.2.5. Pour tout sous-ensemble $\mathcal{A} \subset \mathcal{P}(\Omega)$, la *tribu engendrée* par \mathcal{A} est définie par

$$\sigma(\mathcal{A}) := \bigcap_{\substack{\mathcal{F} \text{ tribu,} \\ \mathcal{A} \subset \mathcal{F}}} \mathcal{F}.$$

C'est la plus petite tribu qui contient tous les événements de \mathcal{A} .

En particulier, pour $\Omega = \mathbb{R}$, on appelle *tribu borélienne* la tribu engendrée par l'ensemble \mathcal{O} des ouverts de \mathbb{R} . C'est également la tribu engendrée par l'ensemble des intervalles ouverts, ou l'ensemble des intervalles fermés.

Si $\mathcal{E} \subset \mathcal{F}$ sont deux tribus sur Ω , on dit que \mathcal{E} est une sous-tribu de \mathcal{F} . La tribu \mathcal{E} est alors plus grossière (correspond à une connaissance moins détaillée), et la tribu \mathcal{F} plus fine (correspond à une connaissance plus précise).

Si \mathcal{E} et \mathcal{F} sont deux tribus sur Ω , on notera $\mathcal{E} \vee \mathcal{F}$ la tribu $\sigma(\mathcal{E} \cup \mathcal{F})$. C'est un ensemble généralement plus grand que $\mathcal{E} \cup \mathcal{F}$. Elle correspond à la connaissance obtenue en croisant celle associée aux deux tribus \mathcal{E} et \mathcal{F} .

Exemple 1.2.6. On note $\Omega = \{1, \dots, 6\}^2$ l'univers de probabilité correspondant au lancer de deux dés à 6 faces.

1. La tribu $\mathcal{P}(\Omega)$ correspond à la connaissance obtenue en observant le résultat des deux dés.
2. La tribu $\mathcal{A} = \{A \times \{1, \dots, 6\}, A \subset \{1, \dots, 6\}\}$ correspond à la connaissance obtenue en observant le résultat du premier dé.
3. La tribu $\mathcal{B} = \{\emptyset, \{1, \dots, 6\} \times \{1, 3, 5\}, \{1, \dots, 6\} \times \{2, 4, 6\}, \Omega\}$ correspond à la connaissance obtenue en observant la parité du second dé.

Une tribu est une famille de sous-ensemble de Ω satisfaisant trois propriétés relativement restrictives. On introduit maintenant la notion de classe monotone, plus générale que celle de tribu, et avec laquelle il peut être plus simple de travailler.

Définition 1.2.7. Une classe monotone \mathcal{C} sur Ω est une partie de $\mathcal{P}(\Omega)$ satisfaisant les propriétés suivantes :

1. $\Omega \in \mathcal{C}$,
2. pour tout $A, B \in \mathcal{C}$ tel que $A \subset B$, $B \setminus A \in \mathcal{C}$,
3. pour toute suite croissante d'événements $(A_n, n \geq 0) \in \mathcal{C}^{\mathbb{N}}$ (i.e. telle que $A_i \subset A_j$ pour tout $i < j$), on a $\bigcup_{n \in \mathbb{N}} A_n \in \mathcal{C}$.

Remarque 1.2.8.

1. Toute tribu est une classe monotone.
2. Une intersection de classes monotones est une classe monotone, d'où la notion de classe monotone engendrée par $\mathcal{A} \subset \mathcal{P}(\Omega)$, définie par $\mathcal{M}(\mathcal{A}) := \bigcap_{\substack{\mathcal{C} \text{ classe monotone,} \\ \mathcal{A} \subset \mathcal{C}}} \mathcal{C}$.

Le résultat suivant justifie l'introduction de la notion de classe monotone. Il montre que sous certaines hypothèses sur l'ensemble générateur \mathcal{A} , la classe monotone engendrée par \mathcal{A} est une tribu. C'est alors la plus petite tribu contenant \mathcal{A} .

Théorème 1.2.9 (Lemme des classes monotones). *Pour toute famille \mathcal{A} de sous-ensembles de Ω satisfaisant*

$$\forall n \in \mathbb{N}, \forall A_1, \dots, A_n \in \mathcal{A}, \bigcap_{i=1}^n A_i \in \mathcal{A},$$

on a $\mathcal{M}(\mathcal{A}) = \sigma(\mathcal{A})$.

On résume souvent le lemme des classes monotones de la façon suivante : toute famille d'événements stable par passage au complémentaire, unions croissantes et intersections finies est une tribu. Le lemme des classes monotones est souvent employé comme suit : pour montrer qu'une propriété est vérifiée par tous les événements d'une tribu, on montre qu'elle est vérifiée par une famille génératrice stable par intersections finies, puis que l'ensemble des événements satisfaisant cette propriété forme une classe monotone.

1.3 Loi de probabilité

La loi de probabilité est le dernier élément constitutif d'un espace de probabilité. Étant donné un espace mesuré (Ω, \mathcal{F}) , elle permet de mesurer la probabilité d'occurrence de chaque événement de \mathcal{F} . Elle peut représenter le degré de confiance que l'on a dans la réalisation d'un événement, ou bien, dans une vision fréquentiste, la proportion d'expériences vérifiant cette événement si on répétait l'expérience aléatoire un grand nombre de fois. On verra en fin de cours le lien entre cette vision historique "fréquentiste" de la théorie des probabilités et la théorie moderne, issue des travaux de Kolmogorov, qu'on développe ici.

Définition 1.3.1. Une *mesure* (ou *loi*) de probabilité sur l'espace mesuré (Ω, \mathcal{F}) est une application $\mathbf{P} : \mathcal{F} \rightarrow \mathbb{R}$ satisfaisant les probabilités suivantes

1. $\mathbf{P}(\Omega) = 1$,
2. pour tout $A \in \mathcal{F}$, $\mathbf{P}(A) \in [0, 1]$,
3. pour toute suite $(A_n, n \geq 0) \in \mathcal{F}^{\mathbb{N}}$ d'événements deux à deux disjoints (i.e. tels que $A_i \cap A_j = \emptyset$ pour tout $i \neq j$), $\mathbf{P}(\bigcup_{n \in \mathbb{N}} A_n) = \sum_{n \in \mathbb{N}} \mathbf{P}(A_n)$.

Une mesure de probabilité peut être associée à une notion de "masse". Pour chaque événement $A \in \mathcal{F}$, la fonction \mathbf{P} donne la masse associée à cet événement. La masse totale est normée pour être égale à 1, et on a bien que la masse de la réunion d'événements indépendants est égale à la somme des masses.

Remarque 1.3.2. Grâce à la 3e propriété, on observe que si Ω est discret et $\mathcal{F} = \mathcal{P}(\Omega)$, alors pour tout $A \subset \Omega$, on a

$$\mathbf{P}(A) = \mathbf{P}(\bigcup_{\omega \in A} \{\omega\}) = \sum_{\omega \in A} \mathbf{P}(\{\omega\}),$$

par conséquent, une mesure de probabilité sur un espace discret est caractérisée par sa valeur sur les événements élémentaires.

Exemple 1.3.3 (Quelques exemples de mesures de probabilités.).

1. Pour $\Omega = \{1, \dots, n\}$ et $\mathcal{F} = \mathcal{P}(\Omega)$, les mesures définies par

$$\forall k \in \Omega, \mathbf{P}(k) = \frac{1}{n} \quad \text{et} \quad \forall k \in \Omega, \mathbf{Q}(k) = \frac{1}{2^n} \binom{n}{k}$$

sont des mesures de probabilités.

2. Pour tout $a \in \Omega$, la mesure définie par

$$\mathbf{P}(A) = \begin{cases} 1 & \text{si } a \in A \\ 0 & \text{sinon} \end{cases}$$

est une mesure de probabilité qu'on appelle la mesure de Dirac en a et qu'on note δ_a .

Propriété 1.3.4. Une mesure de probabilité \mathbf{P} sur (Ω, \mathcal{F}) satisfait les égalités suivantes :

1. $\mathbf{P}(\emptyset) = 0$;
2. pour tout $A \in \mathcal{F}$, $\mathbf{P}(A^c) = 1 - \mathbf{P}(A)$;
3. pour tout $A, B \in \mathcal{F}$ tels que $A \subset B$, $\mathbf{P}(A) \leq \mathbf{P}(B)$.
4. pour tout $A, B \in \mathcal{F}$, $\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(A \cap B)$.

Démonstration. On observe pour commencer $\emptyset = \emptyset \cup \emptyset$, et par conséquent

$$\mathbf{P}(\emptyset) = \mathbf{P}(\emptyset) + \mathbf{P}(\emptyset).$$

On en déduit immédiatement que $\mathbf{P}(\emptyset) = 0$.

De même pour tout $A \in \mathcal{F}$, on a

$$1 = \mathbf{P}(\Omega) = \mathbf{P}(A \cup A^c) = \mathbf{P}(A) + \mathbf{P}(A^c),$$

d'où $\mathbf{P}(A^c) = 1 - \mathbf{P}(A)$.

Soit $A, B \in \mathcal{F}$ tel que $A \subset B$. On a alors

$$\mathbf{P}(B) = \mathbf{P}((B \cap A) \cup (B \cap A^c)) = \mathbf{P}(B \cap A) + \mathbf{P}(B \cap A^c) = \mathbf{P}(A) + \mathbf{P}(B \cap A^c) \geq \mathbf{P}(A).$$

Enfin, pour tout $A, B \in \mathcal{F}$, on peut écrire

$$\begin{aligned} \mathbf{P}(A \cup B) &= \mathbf{P}((A \cap B^c) \cup (A \cap B) \cup (A^c \cap B)) \\ &= \mathbf{P}(A \cap B^c) + \mathbf{P}(A \cap B) + \mathbf{P}(A^c \cap B) \\ &= (\mathbf{P}(A \cap B^c) + \mathbf{P}(A \cap B)) + (\mathbf{P}(A \cap B) + \mathbf{P}(A^c \cap B)) - \mathbf{P}(A \cap B) \\ &= \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(A \cap B), \end{aligned}$$

ce qui complète la preuve. □

Définition 1.3.5. Un événement $A \in \mathcal{F}$ tel que $\mathbf{P}(A) = 0$ est dit *négligeable*. Une propriété vraie pour tout $\omega \in B$ tel que B^c est négligeable est dite *vraie presque sûrement* (ou p.s.).

Proposition 1.3.6 (Convergence monotone). Soit \mathbf{P} une mesure de probabilité de (Ω, \mathcal{F}) et $(A_n, n \geq 0)$ une suite croissante d'événements de \mathcal{F} . On a

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

De la même façon, si $(B_n, n \geq 0)$ est une suite décroissante d'événements de \mathcal{F} , on a

$$\mathbf{P}\left(\bigcap_{n \geq 1} B_n\right) = \lim_{n \rightarrow \infty} \mathbf{P}(B_n).$$

Pour déterminer la valeur de \mathbf{P} , il est souvent suffisant de donner la valeur de $\mathbf{P}(A)$ pour $A \in \mathcal{A}$ une partie relativement petite de la tribu \mathcal{F} . C'est le cas dans la remarque 1.3.2 : dans un univers discret, il suffit de donner la valeur de \mathbf{P} sur les singletons pour caractériser (définir de façon unique) la mesure de probabilité \mathbf{P} . Dans le cas d'un univers continu, on utilise le résultat suivant.

Lemme 1.3.7. Soit \mathbf{P} et \mathbf{Q} deux mesures de probabilités sur $(\Omega, \sigma(\mathcal{A}))$. Si la famille d'événements \mathcal{A} est stable par intersections finies, et si $\mathbf{P}(A) = \mathbf{Q}(A)$ pour tout $A \in \mathcal{A}$, alors $\mathbf{P} = \mathbf{Q}$.

Ainsi, pour définir \mathbf{P} , il suffit de donner la valeur de $\mathbf{P}(A)$ pour tout $A \in \mathcal{A}$ une famille stable par intersection finie. Notons que si cette donnée garantit l'unicité de la loi, elle ne garantit pas son existence.

Démonstration. On note $\mathcal{G} = \{A \in \sigma(\mathcal{A}) : \mathbf{P}(A) = \mathbf{Q}(A)\}$. L'objectif de cette preuve est de montrer que $\mathcal{G} = \sigma(\mathcal{A})$. Pour ce faire, on montre que \mathcal{G} est une classe monotone qui contient \mathcal{A} . Par définition, on aura alors $\mathcal{M}(\mathcal{A}) \subset \mathcal{G}$, et grâce au théorème 1.2.9, on en déduira que $\sigma(\mathcal{A}) \subset \mathcal{G}$, ce qui complètera la preuve.

On observe pour commencer que $\mathbf{P}(\Omega) = 1 = \mathbf{Q}(\Omega)$, par conséquent $\Omega \in \mathcal{G}$. Soit $A, B \in \mathcal{G}$ tel que $A \subset B$, on a $\mathbf{P}(B \setminus A) = \mathbf{P}(B) - \mathbf{P}(A) = \mathbf{Q}(B) - \mathbf{Q}(A) = \mathbf{Q}(B \setminus A)$, donc $B \setminus A \in \mathcal{G}$.

Enfin, si (A_n) est une suite croissante d'événements de \mathcal{G} on a grâce à la proposition 1.3.6 :

$$\mathbf{P}(\cup_{n \in \mathbb{N}} A_n) = \lim_{n \rightarrow \infty} \mathbf{P}(A_n) = \lim_{n \rightarrow \infty} \mathbf{Q}(A_n) = \mathbf{Q}(\cup_{n \in \mathbb{N}} A_n).$$

On a donc $\cup_{n \in \mathbb{N}} A_n \in \mathcal{G}$, ce qui montre bien que \mathcal{G} est une classe monotone, ce qui complète cette démonstration. \square

Exemple 1.3.8. On pose $\Omega = \{1, \dots, 6\}^{\mathbb{N}}$ l'ensemble des suites d'entiers entre 1 et 6. Cela correspond à l'expérience du lancer d'une infinité de dés à 6 faces. On lui associe la tribu \mathcal{F} sur Ω engendrée par les ensembles $\Pi_{i_1, \dots, i_n} = \{(i_1, \dots, i_n, j_1, j_2, \dots), (j_k) \in \{1, \dots, 6\}^{\mathbb{N}}\}$ pour $(i_1, \dots, i_n) \in \{1, \dots, 6\}^n$, qui correspond à l'observation des n premiers lancers pour tout n . Alors, une mesure de probabilités \mathbf{P} sur (Ω, \mathcal{F}) est caractérisée $\mathbf{P}(\Pi_{i_1, \dots, i_n})$ pour $n \in \mathbb{N}$ et $(i_1, \dots, i_n) \in \{1, \dots, 6\}^n$.

Exemple 1.3.9 (Mesure de Lebesgue). La mesure de Lebesgue sur $[0, 1]$ est la seule mesure de probabilité tel que $\mathbf{P}([a, b]) = b - a$ pour tout $0 \leq a \leq b \leq 1$. L'unicité de cette mesure est une conséquence du lemme 1.3.7. Montrer l'existence d'une telle mesure de probabilité dépasse le cadre de ces rappels, on supposera donc la mesure de Lebesgue construite par la suite.

1.4 Exercices

Exercice 1.1. Construire un espace de probabilité associé aux expériences aléatoires suivantes :

1. On considère une urne contenant 5 boules numérotées de 1 à 5, et on tire 2 boules dans cette urne au hasard.
2. On considère deux urnes, l'une contenant 3 boules numérotées de A à C , et l'autre contenant 5 boules numérotées de A à E . On choisit au hasard une urne, dans laquelle on tire au hasard une boule.
3. On lance un dé à 6 faces le nombre de fois nécessaires pour obtenir un 6, et on s'intéresse au nombre de lancers qui a été nécessaire.

Exercice 1.2. On reprend les notations de l'exemple 1.2.6.

1. Donner une partie génératrice des tribus \mathcal{A} et \mathcal{B} .
2. Décrire la tribu $\mathcal{A} \vee \mathcal{B}$. À quelle type de connaissance sur l'expérience aléatoire correspond-elle ? Quelle tribu \mathcal{C} peut-on ajouter pour obtenir une connaissance complète sur l'expérience aléatoire.

Exercice 1.3. On jette successivement trois pièces de monnaie, et on s'intéresse aux côtés qu'elles montrent.

1. Construire un espace de probabilité associé à cette expérience aléatoire.
2. On considère les trois événements A : "la première pièce est tombée sur face", B : "la deuxième pièce est tombée sur face" et C : la troisième pièce est tombée sur face.
 - (a) Donner une description de l'événement $A \cup B^c$.
 - (b) Écrire, grâce aux événements A, B et C et des opérations ensemblistes l'événement D : "la première et la troisième pièce montrent des côtés différents".
 - (c) Calculer la probabilité des événements $A \cup B^c$ et D .
3. Une deuxième personne arrive et observe le résultat des trois pièces de monnaie sur la table sans connaître l'ordre dans lequel elles ont été lancées.

- (a) Donner la tribu \mathcal{G} associée à la connaissance de cette seconde personne.
- (b) Calculer la probabilité des différents événements de cette tribu.
- (c) L'événement B appartient-il à \mathcal{G} ? Et $A \cup B \cup C$?

Exercice 1.4. Un tiroir contient n paires de chaussettes de couleurs différentes.

1. On tire deux chaussettes au hasard. Quelle est la probabilité qu'elles viennent de la même paire?
2. On tire six chaussettes au hasard. Calculer la probabilité des événements suivants :
 - on a tiré 3 paires de chaussettes ;
 - on a tiré au moins une paire de chaussettes ;
 - on a tiré exactement une paire de chaussettes.

Exercice 1.5. On teste les capacités cognitives d'un animal de laboratoire de la façon suivante : on le place devant quatre boutons sur lesquels il peut appuyer. Trois de ces boutons provoquent un bruit désagréable, tandis que le quatrième provoque la distribution d'une friandise. Déterminer l'espace de probabilité associé à cette expérience dans les trois cas suivants :

- l'animal a une mémoire parfaite, et se souvient du résultat de toutes ses actions passées ;
- l'animal n'a aucune mémoire, et ne se souvient pas du résultat de ses actions ;
- l'animal a une mémoire partielle, et se souvient seulement du résultat de sa dernière action.

Chapitre 2

Variables aléatoires

Dans tout le reste de ce manuscrit, on suppose donné un univers de probabilité $(\Omega, \mathcal{F}, \mathbf{P})$. L'objectif de ce chapitre est d'introduire la notion de variable aléatoire, qui est un objet central de la théorie des probabilités. Une variable aléatoire est construite comme une observable sur une expérience aléatoire, c'est-à-dire une valeur numérique associée au résultat de cette expérience.

On introduira dans un premier temps les notions de variable aléatoire et de loi de la variable aléatoire. On introduira ensuite l'espérance et la variance de ces variables aléatoires, qui permettent de mesurer respectivement la valeur moyenne et la dispersion attendues pour l'observable lorsque l'expérience aléatoire est réalisée. On introduira enfin quelques lois de probabilité usuelles.

2.1 Fonctions mesurables

On introduit dans un premier temps la notion de fonction mesurable.

Définition 2.1.1. Soit (E, \mathcal{E}) un espace mesuré. On dit que l'application $f : \Omega \rightarrow E$ est mesurable si $f^{-1}(A) \in \mathcal{F}$ pour tout $A \in \mathcal{E}$.

Soit \mathcal{G} une sous-tribu de \mathcal{F} , on dit que f est mesurable par rapport à \mathcal{G} si $f^{-1}(A) \in \mathcal{G}$ pour tout $A \in \mathcal{E}$.

La mesurabilité est une propriété issue de la théorie de la mesure. Elle peut être interprétée comme étant la marque que la connaissance fournie par \mathcal{F} (ou \mathcal{G}) est suffisante pour donner la valeur de f . Mais cette interprétation ne fait qu'explorer la surface de cette notion, et on se référera aux ouvrages cités plus haut pour une définition complète de la notion de fonction mesurable. Citons toutefois quelques propriétés permettant de montrer qu'une fonction est mesurable

Propriété 2.1.2.

1. Si $A \in \mathcal{F}$, la fonction $\mathbf{1}_A$ est mesurable.
2. Si f et g sont des fonctions mesurables, alors $\lambda f + \mu g$ est une fonction mesurable pour tout $\lambda, \mu \in \mathbb{R}$.
3. Si $(f_n, n \geq 0)$ est une suite de fonctions mesurables, alors $\sup_{n \in \mathbb{N}} f_n$ et $\limsup_{n \rightarrow \infty} f_n$ sont des fonctions mesurables.

La tribu borélienne joue un rôle fondamental pour montrer la mesurabilité de fonctions.

Proposition 2.1.3. La tribu borélienne est la plus petite tribu qui rend toutes les fonctions continues de \mathbb{R} dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ mesurables.

Sauf mention explicite du contraire, \mathbb{R} est toujours muni de la tribu borélienne. Grâce aux propriétés précédentes, on obtient que toute fonction continue, toute limite de fonction continue, tout maximum de fonction continue sont des fonctions mesurables. Au contraire, la construction de fonctions non-mesurables est complexe et dépasse le cadre de ces notes.

2.2 Variables aléatoires et leurs lois

Définition 2.2.1. Une variable aléatoire est une fonction mesurable $X : (\Omega, \mathcal{F}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

Une variable aléatoire est souvent définie à un événement de probabilité négligeable près (en particulier si Ω est un univers continu). Par conséquent, on dira par exemple qu'une variable aléatoire est positive si $X \geq 0$ p.s., i.e.

$$\mathbf{P}(\{\omega \in \Omega : X(\omega) < 0\}) = 0.$$

Pour simplifier les notations, on écrira souvent pour $A \in \mathcal{B}(\mathbb{R})$:

$$\mathbf{P}(\{\omega \in \Omega : X(\omega) \in A\}) = \mathbf{P}(X^{-1}(A)) =: \mathbf{P}(X \in A).$$

Exemple 2.2.2.

1. On considère l'univers de probabilité $(\Omega, \mathcal{F}, \mathbf{P})$, avec $\Omega = \{1, \dots, 6\}^2$, la tribu $\mathcal{F} = \mathcal{P}(\Omega)$ et $\mathbf{P}(\{(i, j)\}) = 1/36$, la fonction $X : (i, j) \mapsto i + j$ est une variable aléatoire modélisant le résultat obtenu en sommant le lancer de deux dés à 6 faces.
2. On considère l'univers de probabilité $([0, 1], \mathcal{B}([0, 1]), \lambda)$, où λ est la mesure de Lebesgue sur $[0, 1]$. La fonction $X : \omega \mapsto \omega$ est une variable aléatoire modélisant le résultat obtenu en tirant un nombre au hasard entre 0 et 1. C'est un simple exercice de montrer que $X \notin \mathbb{Q}$ p.s.

Une variable aléatoire X est souvent décrite à partir de sa loi de probabilité, qui est définie comme la mesure image de la mesure \mathbf{P} par l'application X .

Définition 2.2.3. Soit X une variable aléatoire, la loi de X est la mesure de probabilité sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ définie par

$$\mu_X(A) := \mathbf{P}(X^{-1}(A)) = \mathbf{P}(X \in A).$$

Le support de X est le plus petit fermé tel que $\mu_X(F) = 1$, qu'on peut définir comme

$$\text{Supp}(X) = \text{Supp}(\mu_X) := \bigcap_{\substack{F \text{ fermé,} \\ \mu_X(F)=1}} F.$$

La loi d'une variable aléatoire est donc une mesure de probabilités sur l'espace mesuré $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Son support décrit l'ensemble des valeurs prises par la variable aléatoire X , à un ensemble négligeable près.

De la même façon qu'au chapitre précédent, on peut caractériser cette loi en donnant sa valeur pour une famille de sous-ensembles de \mathbb{R} bien choisis.

Propriété 2.2.4. La loi de X est caractérisée par $\mathbf{P}(X \in A)$ pour tout intervalle ouvert (ou fermé) $A \subset \mathbb{R}$.

La loi de X est caractérisée par $\mathbf{P}(X \leq x)$ pour tout $x \in \mathbb{R}$.

La propriété précédente motive l'introduction de la fonction de répartition d'une loi, comme une fonction permettant de caractériser une loi de probabilité.

Définition 2.2.5. La fonction de répartition de X est la fonction

$$F : \begin{array}{ccc} \mathbb{R} & \longrightarrow & [0, 1] \\ x & \longmapsto & \mathbf{P}(X \leq x). \end{array}$$

Propriété 2.2.6. Une fonction de répartition est croissante, continue à droite et avec des limites à gauche en tout point, i.e.

$$\lim_{\substack{y \rightarrow x \\ y < x}} F(y) \leq F(x) = \lim_{\substack{y \rightarrow x \\ y > x}} F(y).$$

La réciproque de la propriété précédente est également vraie : toute fonction f croissante, continue à droite avec des limites à gauche en tout point convergeant vers 1 en $+\infty$ et vers 0 en $-\infty$ est la fonction de répartition d'une loi de probabilité sur \mathbb{R} . Deux familles de lois de probabilités sont particulièrement communes.

On notera souvent $F(x-) = \lim_{y \rightarrow x} F(y) = \mu((-\infty, x))$. Pour tout $a < b$, on notera que

$$F(b) - F(a) = \mu((a, b]), \quad F(b) - F(a-) = \mu([a, b]) \quad \text{et} \quad F(b-) - F(a) = \mu((a, b)).$$

Propriété 2.2.7. On dit qu'un point $x \in \mathbb{R}$ est un atome de la mesure μ si $\mu(\{x\}) > 0$. Une loi μ a un atome en x si et seulement si sa fonction de répartition est discontinue en x . On a alors $\mu(\{x\}) = F(x) - F(x-)$.

Lois discrètes

Une loi de probabilité est dite discrète si son support est fini ou dénombrable. Une loi de probabilité sur \mathbb{R} est discrète si et seulement si sa fonction de répartition est une fonction étagée.

Lois à densité

Une loi μ admet pour densité (par rapport à la mesure de Lebesgue) la fonction mesurable φ si elle est définie par

$$\mu(A) = \int_A \varphi(x) dx = \int_{\mathbb{R}} \mathbf{1}_A(x) \varphi(x) dx.$$

Une fonction φ est la densité d'une loi de probabilité si et seulement si

- $\varphi \geq 0$ presque partout ;
- $\int_{-\infty}^{\infty} \varphi(x) dx = 1$.

Si μ est une loi à densité, alors sa fonction de répartition est continue, dérivable presque partout, et on a $F'(x) = \varphi(x)$ presque partout. Réciproquement, si F est continue, dérivable presque partout et si $\int F'(x) dx = 1$ alors F est la fonction de répartition d'une loi à densité.

Notons qu'il existe des lois qui ne sont ni à densité, ni discrètes.

2.3 Vecteurs aléatoires

La notion de variable aléatoire peut être étendue à des espaces plus généraux que \mathbb{R} . On parle de variable aléatoire dans un espace E comme une fonction mesurable à valeurs dans (E, \mathcal{E}) . Un cas particulier de cette notion générale est celle de vecteur aléatoire, défini comme une variable aléatoire à valeurs dans \mathbb{R}^n .

Définition 2.3.1. Un vecteur aléatoire est une famille $X = (X_1, \dots, X_n)$ de variables aléatoires. La loi du vecteur aléatoire est la mesure de probabilité sur $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ définie par

$$\mu_X(A) = \mathbf{P}((X_1, \dots, X_n) \in A).$$

Les lois marginales du vecteur aléatoire (X_1, \dots, X_n) sont les lois des variables X_1, \dots, X_n .

Remarque 2.3.2. Les lois marginales d'un vecteur aléatoire ne déterminent pas la loi de ce vecteur. En effet, considérons le vecteur aléatoire (X, Y) , dont la loi est donnée par

$$\mathbf{P}(X = 1, Y = 1) = \mathbf{P}(X = 1, Y = 0) = \mathbf{P}(X = 0, Y = 1) = \mathbf{P}(X = 0, Y = 0) = \frac{1}{4},$$

alors les vecteurs (X, X) et (X, Y) ont les mêmes lois marginales, mais des lois différentes.

Propriété 2.3.3. La loi d'un vecteur aléatoire est caractérisée par $\mathbf{P}(X_1 \in A_1, \dots, X_n \in A_n)$ pour toute famille d'intervalles ouverts (A_1, \dots, A_n) . On appelle l'ensemble $A_1 \times A_2 \times \dots \times A_n$ un pavé de \mathbb{R}^n .

Pour calculer les lois marginales d'une variable aléatoire, on peut utiliser la formule suivante.

Propriété 2.3.4. Soit $X = (X_1, \dots, X_n)$ un vecteur aléatoire, la loi marginale de X_1 est donnée par $\mu_{X_1}(A) = \mu(A \times \Omega^{n-1})$.

Corollaire 2.3.5. 1. Si X_1, \dots, X_n sont des variables aléatoires à densité par rapport à la mesure de Lebesgue, pour tout $x_1 \in \mathbb{R}$, on a

$$\varphi_{X_1}(x) = \int_{\mathbb{R}^{n-1}} \varphi_X(x, x_2, \dots, x_n) dx_2 \dots dx_n.$$

2. Si X_1, \dots, X_n sont des variables aléatoires discrètes, pour tout $x \in X_1(\Omega)$, on a

$$\mathbf{P}(X_1 = x) = \sum_{x_2 \in X_2(\Omega), \dots, x_n \in X_n(\Omega)} \mathbf{P}(X_1 = x, X_2 = x_2, \dots, X_n = x_n).$$

2.4 Espérance, variance

Définition 2.4.1. L'espérance d'une variable aléatoire est la quantité définie par

$$\mathbf{E}(X) = \int_{\Omega} X(\omega) \mathbf{P}(d\omega),$$

qui est bien définie dans les deux cas suivants :

- $X \geq 0$ (auquel cas $\mathbf{E}(X) \in [0, \infty]$);
- $\mathbf{E}(|X|) < \infty$ (auquel cas $\mathbf{E}(X) \in \mathbb{R}$).

Si X est une variable aléatoire telle que $\mathbf{E}(|X|) < \infty$, on dit que $X \in \mathbb{L}^1$, ou que X est intégrable. On étend ces définitions aux vecteurs aléatoires en posant $\mathbf{E}(X) = (\mathbf{E}(X_1), \dots, \mathbf{E}(X_n)) \in \mathbb{R}^n$ pour peu que chacune des variables aléatoires X_i soit intégrable.

Propriété 2.4.2. Si X et Y sont deux variables aléatoires intégrables, alors pour tout $\lambda, \mu \in \mathbb{R}$, on a $\mathbf{E}(\lambda X + \mu Y) = \lambda \mathbf{E}(X) + \mu \mathbf{E}(Y)$ (en particulier, $\lambda X + \mu Y$ est intégrable).

Si X et Y sont deux variables aléatoires intégrables et $X \leq Y$ p.s., alors $\mathbf{E}(X) \leq \mathbf{E}(Y)$.

On se réfère à la première propriété comme la linéarité de l'espérance. La seconde propriété est appelée la positivité de l'espérance (en effet, $Y - X$ est une variable positive, et $\mathbf{E}(Y - X) \geq 0$). Notons que la contraposée de cette propriété est généralement fautive. Même si $\mathbf{E}(Y) \geq \mathbf{E}(X)$, on ne peut pas déduire que Y est plus grande, ou plus petite que X .

Une conséquence de la positivité de l'espérance est l'inégalité de Markov, qui permet de majorer la probabilité pour une variable aléatoire d'être grande par son espérance.

Proposition 2.4.3 (Inégalité de Markov). Soit X une variable aléatoire positive p.s. Pour tout $a > 0$, on a

$$\mathbf{P}(X \geq a) \leq \mathbf{E}(X)/a.$$

Corollaire 2.4.4. Soit X une variable aléatoire à valeurs dans $[0, \infty]$.

1. Si $\mathbf{E}(X) < \infty$ alors $X < \infty$ p.s.
2. Si $\mathbf{E}(X) = 0$, alors $X = 0$ p.s.

La proposition suivante permet le calcul explicite de l'espérance d'une variable aléatoire.

Proposition 2.4.5 (Formule de transfert). Soit $f : \mathbb{R} \rightarrow \mathbb{R}$ une fonction mesurable (par rapport à la tribu borélienne). L'espérance de la variable aléatoire $f(X)$, lorsqu'elle est bien définie, vaut

$$\mathbf{E}(f(X)) = \int_{\mathbb{R}} f(x) \mu_X(dx),$$

où μ_X est la loi de X .

La formule de transfert devient particulièrement simple à utiliser lorsque la variable aléatoire X est discrète, c'est-à-dire lorsqu'elle prend un nombre fini ou dénombrable de valeurs différentes.

Propriété 2.4.6 (Formule de transfert, cas discret). Si X est une variable aléatoire intégrable discrète, alors

$$\mathbf{E}(X) = \sum_{k \in X(\Omega)} k \mathbf{P}(X = k),$$

et de plus, pour toute fonction mesurable f telle que $f(X)$ est intégrable, on a

$$\mathbf{E}(f(X)) = \sum_{k \in X(\Omega)} f(k) \mathbf{P}(X = k)$$

Dans le cas d'une variable continue, on peut également disposer d'une formule plus simple dans le cas des lois ayant une densité par rapport à la mesure de Lebesgue.

Définition 2.4.7. La loi μ d'une variable aléatoire est dite à densité par rapport à la mesure de Lebesgue s'il existe une fonction mesurable φ telle que

$$\forall A \in \mathcal{B}(\mathbb{R}), \mu(A) = \int_A \varphi(x) dx.$$

Notons que φ est une fonction positive telle que $\int_{\mathbb{R}} \varphi(x) dx = 1$.

Remarque 2.4.8. Si la fonction de répartition F de μ est continue sur \mathbb{R} , et dérivable sauf en un nombre au plus fini de points, alors μ admet une densité par rapport à la mesure de Lebesgue, et $\varphi(x) = F'(x)$ sauf en un nombre au plus fini de points.

Propriété 2.4.9 (Formule de transfert, cas à densité). *Si X est une variable aléatoire de loi ayant pour densité φ par rapport à la mesure de Lebesgue, alors*

$$\mathbf{E}(X) = \int x\varphi(x) dx,$$

et de plus, pour toute fonction mesurable f telle que $f(X)$ est intégrable, on a

$$\mathbf{E}(f(X)) = \int f(x)\varphi(x) dx.$$

L'espérance de fonctions de variables aléatoires peut être utilisé pour caractériser la loi de ces variables aléatoires. Ainsi, les résultats suivants montrent qu'il suffit de déterminer $\mathbf{E}(f(X))$ pour un nombre réduit de fonctions f afin de déterminer la loi de X .

Lemme 2.4.10. *La donnée de $\mathbf{E}(f(X))$ pour toute fonction f mesurable bornée $\mathbb{R} \rightarrow \mathbb{R}$ caractérise la loi de X .*

Proposition 2.4.11. *La donnée de $\mathbf{E}(f(X))$ pour toute fonction f de classe C^∞ , positive et bornée $\mathbb{R} \rightarrow \mathbb{R}$ caractérise la loi de X .*

En particulier, la loi de X est caractérisée par $\mathbf{E}(f(X))$ pour l'ensemble des fonctions continues bornées $\mathbb{R} \rightarrow \mathbb{R}$.

Théorème 2.4.12. *On appelle transformée de Fourier de X la fonction*

$$\varphi_X : \begin{array}{l} \mathbb{R} \longrightarrow \mathbb{C} \\ \xi \longmapsto \mathbf{E}(e^{i\xi X}). \end{array}$$

La transformée de Fourier de X caractérise la loi de X .

De la même façon que pour une variable aléatoire, la loi d'un vecteur aléatoire est caractérisé en outre par sa transformée de Fourier.

Théorème 2.4.13. *Soit $X = (X_1, \dots, X_n)$ un vecteur aléatoire. On appelle transformée de Fourier de X la fonction*

$$\varphi_X : \begin{array}{l} \mathbb{R}^n \longrightarrow \mathbb{C} \\ (\xi_1, \dots, \xi_n) \longmapsto \mathbf{E}(e^{i(\xi_1 X_1 + \dots + \xi_n X_n)}). \end{array}$$

La transformée de Fourier de X caractérise la loi de X .

La variance d'une variable aléatoire mesure son étalement. Elle forme une mesure d'à quelle point la variable aléatoire peut s'éloigner de son espérance.

Définition 2.4.14. La variance d'une variable aléatoire intégrable est la quantité

$$\text{Var}(X) = \mathbf{E}((X - \mathbf{E}(X))^2) = \mathbf{E}(X^2) - \mathbf{E}(X)^2.$$

Cette variance est finie si et seulement si $\mathbf{E}(X^2) < \infty$, que l'on note $X \in L^2$.

Proposition 2.4.15 (Inégalité de Bienaymé-Tchebychev). *Pour tout $a > 0$, on a*

$$\mathbf{P}(|X - \mathbf{E}(X)| > a) \leq \frac{\text{Var}(X)}{a^2}.$$

On finit cette section en rappelant la définition de la covariance.

Définition 2.4.16. Soit X, Y des variables aléatoires telles que $\mathbf{E}(X^2) < \infty$, $\mathbf{E}(Y^2) < \infty$. La covariance de X et Y est la quantité définie par

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

Propriété 2.4.17. On a $\mathbf{Var}(X + Y) = \mathbf{Var}(X) + \mathbf{Var}(Y) + 2\text{Cov}(X, Y)$

2.5 Quelques inégalités sur les espérances

On évoque d'abord l'inégalité de Cauchy-Schwartz.

Proposition 2.5.1 (Inégalité de Cauchy-Schwartz). Soit X et Y deux variables aléatoires dans L^2 , on a

$$|\mathbf{E}(XY)| \leq \sqrt{\mathbf{E}(X^2)\mathbf{E}(Y^2)}.$$

Cette inégalité se généralise sous la forme de l'inégalité de Hölder.

Proposition 2.5.2 (Inégalité de Hölder). Soit $p, q \in [1, +\infty]$ tel que $\frac{1}{p} + \frac{1}{q} = 1$, soit X et Y deux variables aléatoires telles que $\mathbf{E}(|X|^p) < \infty$ et $\mathbf{E}(|Y|^q) < \infty$, on a

$$\mathbf{E}(|XY|) \leq \mathbf{E}(|X|^p)^{1/p} \mathbf{E}(|Y|^q)^{1/q}$$

On finit en rappelant l'inégalité de Jensen, qui peut être vue comme une généralisation de la formule caractérisant la convexité d'une fonction.

Proposition 2.5.3 (Inégalité de Jensen). Soit X une variable aléatoire dans L^1 et f une fonction convexe, on a

$$\mathbf{E}(f(X)) \geq f(\mathbf{E}(X)).$$

2.6 Lois usuelles

2.6.1 Lois discrètes

Loi de Dirac. Pour $a \in \mathbb{R}$, la loi de Dirac notée δ_a représente la loi d'une variable aléatoire constante égale à a . Si X est de loi δ_a , on a $\mathbf{E}(X) = a$ et $\mathbf{Var}(X) = 0$. D'ailleurs, toute variable aléatoire ayant une variance nulle suit une loi de Dirac.

Loi uniforme discrète. Pour $n \in \mathbb{N}$, la loi uniforme sur $\{1, \dots, n\}$ modélise le choix, uniformément au hasard, d'un nombre compris entre 1 et n . Soit X une variable aléatoire de loi uniforme sur $\{1, \dots, n\}$, on a

$$\mathbf{P}(X = k) = \frac{\mathbf{1}_{\{1 \leq k \leq n\}}}{n}, \quad \mathbf{E}(X) = \frac{n+1}{2}, \quad \text{et} \quad \mathbf{Var}(X) = \frac{n^2-1}{12}.$$

On peut généraliser la notion de loi uniforme discrète au cas d'un espace E de cardinal fini différent de $\{1, \dots, n\}$. Dans ce cas, la probabilité de choisir un élément de l'ensemble est toujours donnée par $1/\text{Card}(E)$.

Loi de Bernoulli. Pour $p \in [0, 1]$, la loi de Bernoulli de paramètre p , que l'on note $\mathcal{B}(p)$ représente la loi d'une variable aléatoire valant 1 avec probabilité p , et 0 avec probabilité $1 - p$. Soit X une variable aléatoire de loi $\mathcal{B}(p)$, on a

$$\mathbf{P}(X = i) = \begin{cases} p & \text{si } i = 1 \\ 1 - p & \text{si } i = 0 \end{cases}, \quad \mathbf{E}(X) = p \quad \text{et} \quad \mathbf{Var}(X) = p(1 - p).$$

On utilise souvent la loi de Bernoulli pour représenter le résultat d'une expérience aléatoire ayant probabilité p de réussir.

Loi binomiale. Pour $n \in \mathbb{N}$ et $p \in [0, 1]$, la loi binomiale de paramètres n et p , notée $\mathcal{B}(n, p)$ est une loi à support sur $\{0, \dots, n\}$ définie comme suit. Étant donné X une variable aléatoire de loi $\mathcal{B}(n, p)$, on a

$$\mathbf{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad \mathbf{E}(X) = np, \quad \text{et} \quad \mathbf{Var}(X) = np(1-p).$$

Cette loi représente le nombre d'expériences réussies lorsqu'on répète n fois une expérience ayant probabilité p de réussir.

Loi de Poisson. Pour $\lambda \geq 0$, la loi de Poisson de paramètre λ , que l'on note $\mathcal{P}(\lambda)$ représente la loi d'une variable aléatoire à support dans \mathbb{N} définie comme suit. Soit X une variable aléatoire de loi $\mathcal{P}(\lambda)$, on a

$$\mathbf{P}(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}, \quad \mathbf{E}(X) = \lambda \quad \text{et} \quad \mathbf{Var}(X) = \lambda.$$

On se réfère souvent à cette loi comme la "loi des événements rares". En effet, elle apparaît comme la limite, lorsque $n \rightarrow \infty$, des lois $\mathcal{B}(n, \lambda/n)$ (c.f. Chapitre 4. pour la notion de limite de lois). En d'autres termes, elle représente le nombre d'occurrences positives d'une expérience répétée à de très nombreuses reprises mais ayant une probabilité infinitésimale de se produire. Elle apparaît ainsi naturellement dans de nombreux domaines de la physique (nombre d'atomes se désintégrant sur une période donnée), ou de la biologie (nombre d'individus contaminés en une génération par un infecté).

Loi géométrique. Pour $p \in (0, 1]$, la loi Géométrique de paramètre p , notée $\mathcal{G}(p)$ est la loi d'une variable aléatoire à support dans \mathbb{N}^* définie comme suit. Soit X une variable de loi $\mathcal{G}(p)$, on a

$$\mathbf{P}(X = k) = p(1-p)^{k-1}, \quad \mathbf{E}(X) = 1/p \quad \text{et} \quad \mathbf{Var}(X) = \frac{1-p}{p^2}.$$

Cette loi représente, pour une expérience ayant probabilité p de se produire, le nombre de fois qu'il faut reproduire cette expérience pour observer une issue positive.

2.6.2 Lois à densité

Loi uniforme. Pour $a < b$, la loi uniforme sur $[a, b]$ est définie comme suit. Soit X une variable aléatoire de loi uniforme sur $[a, b]$, on a

$$\varphi_X(x) = \frac{1}{b-a} \mathbf{1}_{\{x \in [a, b]\}}, \quad \mathbf{E}(X) = \frac{a+b}{2} \quad \text{et} \quad \mathbf{Var}(X) = \frac{(b-a)^2}{12}.$$

Loi exponentielle. Pour $\lambda \in \mathbb{R}_+$, la loi exponentielle de paramètre λ , notée $\mathcal{E}(\lambda)$ est définie comme suit. Étant donné X une variable aléatoire de loi $\mathcal{E}(\lambda)$, on a

$$\varphi_X(x) = \lambda e^{-\lambda x} \mathbf{1}_{\{x \geq 0\}}, \quad \mathbf{E}(X) = \frac{1}{\lambda} \quad \text{et} \quad \mathbf{Var}(X) = \frac{1}{\lambda^2}.$$

La loi exponentielle est utilisée pour modéliser la durée de vie d'un phénomène sans mémoire ou sans vieillissement. En effet, c'est l'unique famille de lois satisfaisant, $\mathbf{P}(X \geq x+y | X \geq x) = \mathbf{P}(X \geq y)$ pour tout $x, y \in \mathbb{R}_+$ (c.f. Chapitre 3 pour la définition de probabilité conditionnelle).

Loi gaussienne. Pour $\mu \in \mathbb{R}$ et $\sigma^2 \in \mathbb{R}_+$, la loi normale de moyenne μ et de variance σ^2 est notée $\mathcal{N}(\mu, \sigma^2)$. Soit X une variable aléatoire de loi $\mathcal{N}(\mu, \sigma^2)$, on a

$$\varphi_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2}, \quad \mathbf{E}(X) = \mu, \quad \mathbf{Var}(X) = \sigma^2.$$

La loi normale (ou gaussienne) est souvent appelée "loi des erreurs" de par son apparition dans le théorème central limite. C'est une loi utilisée à de très nombreuses reprises pour modéliser notamment une erreur de mesure, ou la variabilité d'un paramètre dépendant d'un grand nombre de facteurs extérieurs (taille d'une personne, rendement d'une récolte, etc.). On remarquera que si Y est de loi $\mathcal{N}(0, 1)$, alors $\mu + \sigma Y \sim \mathcal{N}(\mu, \sigma^2)$.

Loi Gamma. Pour $a, \theta \geq 0$, la loi Gamma de paramètres a et θ est notée $\Gamma(a, \theta)$ est une loi à support dans \mathbb{R}_+ . Étant donné X une variable aléatoire de loi $\Gamma(a)$, on a

$$\varphi_X(x) = \frac{1}{\Gamma(a)\theta^a} x^{a-1} e^{-x/\theta} \mathbf{1}_{\{x>0\}}, \quad \mathbf{E}(X) = a\theta, \quad \mathbf{Var}(X) = a\theta^2.$$

On notera que si Y est de loi $\Gamma(a, 1)$, alors $\theta Y \sim \Gamma(a, \theta)$. La fonction Γ éponyme introduite dans la densité de la loi est la fonction Gamma d'Euler, définie par

$$\Gamma(a) = \int_0^\infty x^{a-1} e^{-x} dx.$$

On notera que $\Gamma(a+1) = a\Gamma(a)$ par intégration par partie, et en particulier $\Gamma(n+1) = n!$.

Loi Beta. Pour $a, b \geq 0$, la loi Beta de paramètres a et b est notée $B(a, b)$, et est définie comme suit. Soit X une variable aléatoire de loi $B(a, b)$, on a

$$\varphi_X(x) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1} (1-x)^{b-1} \mathbf{1}_{\{x \in (0,1)\}}, \quad \mathbf{E}(X) = \frac{a}{a+b} \quad \text{et} \quad \mathbf{Var}(X) = \frac{ab}{(a+b)^2(a+b+1)}.$$

Notons que si $a = b = 1$, on retrouve la loi uniforme sur $[0, 1]$.

2.7 Exercices

Exercice 2.1 (Fonction de répartition inverse). Soit X une variable aléatoire réelle, on note F sa fonction de répartition. Pour tout $u \in (0, 1)$, on pose

$$F^{-1}(u) = \inf\{t \in \mathbb{R} : F(t) \geq u\},$$

qu'on appelle l'inverse généralisée continue à gauche de F .

1. Montrer que pour tout $t \in \mathbb{R}$ et $u \in (0, 1)$, on a $F^{-1}(u) \leq t \iff u \leq F(t)$.
2. En déduire que si U est de loi uniforme sur $[0, 1]$, alors $\bar{X} = F^{-1}(U)$ suit la même loi que X .
3. On suppose que X est de loi exponentielle de paramètre 1, déterminer F et F^{-1} et en déduire la loi de $-\log U$.
4. On appelle loi de Cauchy une loi ayant pour densité $\frac{1}{\pi(1+x^2)}$ par rapport à la mesure de Lebesgue. Déterminer G tel que $G(U)$ suit une loi de Cauchy.
5. Montrer que si F est une fonction croissante continue à droite telle que $\lim_{x \rightarrow -\infty} F(x) = 0$ et $\lim_{x \rightarrow \infty} F(x) = 1$, alors F est la fonction de répartition d'une variable aléatoire.

Exercice 2.2 (Lois images).

1. Soit X une variable aléatoire de loi $\mathcal{E}(\lambda)$, déterminer la loi de $\lfloor X \rfloor + 1$.
2. Soit U une variable aléatoire de loi uniforme sur $[0, 1]$, déterminer la loi de U^2 .
3. Soit U une variable aléatoire de loi uniforme sur $[0, 1]$, déterminer la loi de $\lfloor 1/U \rfloor$.
4. Soit X une variable aléatoire de loi $\mathcal{N}(0, 1)$, déterminer la loi de X^2 .

Exercice 2.3 (Calculs d'espérances). Calculer les espérances et variances des lois données dans la section 2.6. Tracer leur fonction de répartition.

Exercice 2.4. Calculer l'espérance des variables aléatoires dont les lois sont données ci-dessous.

1. X telle que $\mathbf{P}(X = n) = \frac{1}{n(n+1)}$ pour tout $n \geq 1$ (justifier que c'est une mesure de probabilité).
2. $Y = e^Z$, où Z suit une loi exponentielle de paramètre λ .
3. W ayant pour densité $\frac{1}{\pi(1+x^2)}$ par rapport à la mesure de Lebesgue.

Exercice 2.5 (Une formule utile).

1. Soit X une variable aléatoire à valeurs dans \mathbb{N} . Montrer que

$$\mathbf{E}(X) = \sum_{n \geq 1} \mathbf{P}(X \geq n).$$

2. Soit X une variable aléatoire positive p.s. Montrer que pour tout $\alpha > 0$, on a

$$\mathbf{E}(X^\alpha) = \alpha \int t^{\alpha-1} \mathbf{P}(X > t) dt$$

Exercice 2.6. Soit (X, Y, Z) un point tiré au hasard sur la sphère \mathbb{R}^3 . Déterminer la loi de X .

Chapitre 3

Indépendance

La notion d'indépendance est la première notion importante apparaissant dans ces notes où la théorie des probabilités se différencie nettement de l'intégration. Intuitivement, l'indépendance de deux événements A et B se caractérise par le fait qu'on ne peut obtenir de connaissance sur B en sachant que l'événement A est réalisé ou non. De la même façon, l'indépendance de deux variables aléatoires X et Y signifie que l'observation de X ne permet pas d'obtenir de connaissance sur la variable Y .

Toutefois, cette explication intuitive doit être formalisée pour entrer dans le cadre de la théorie des probabilités telle que construite plus haut. Nous commencerons par définir l'indépendance mathématique entre deux ou plusieurs événements. On définira ensuite l'indépendance entre deux ou plusieurs variables aléatoires, qui sera reliée à la notion d'indépendance des tribus engendrées par ces variables aléatoires. Cette notion permettra par la suite de définir aisément des familles infinies de variables aléatoires, en les construisant à partir d'une famille de variables indépendantes et identiquement distribuées.

Un résultat fondamental de cette section est que si (X, Y) est un couple de variables aléatoires indépendantes, alors la loi de (X, Y) s'exprime sous la forme d'une mesure produit. On rappellera à cette occasion les théorèmes de Fubini-Tonelli et Fubini-Lebesgue, qui permettent d'échanger l'ordre d'intégration sous certaines hypothèses.

3.1 Indépendance d'événements

Nous définissons dans un premier temps la notion d'indépendance de deux événements.

Définition 3.1.1. Deux événements A et B de \mathcal{F} sont dits *indépendants* si $\mathbf{P}(A \cap B) = \mathbf{P}(A)\mathbf{P}(B)$.

La notion d'indépendance de deux événements marque le fait que le fait de savoir si A se réalise ou non n'influence pas la probabilité qu'on attribue à l'événement B .

Exemple 3.1.2.

1. Considérons le lancer d'un dé équilibré à 6 faces, les événements A "la face sur laquelle le dé tombe est paire" et B "la face sur laquelle le dé tombe est un multiple de 3" sont indépendants. Sachant que A est réalisé, on sait que le dé est tombé sur la face 2, 4 ou 6, et ces trois événements sont équiprobables. Alors, B arrive toujours avec une chance sur trois, qu'on sache que l'événement A s'est produit ou non.
2. Lorsqu'on considère le lancer de plusieurs dés, on fait justement l'hypothèse que ces lancers sont indépendants, c'est-à-dire que tout événement dépendant du résultat d'un dé est indépendant de tout événement dépendant du résultat de l'autre. C'est le cas en posant $\Omega = \{1, \dots, 6\}^2$ avec $\mathbf{P}(\{\omega\}) = 1/36$, dans ce cas les événements $\{6\} \times \{1, \dots, 6\}$ et $\{1, \dots, 6\} \times \{6\}$ sont indépendants.

Propriété 3.1.3. Si A et B sont indépendants, alors A et B^c , A^c et B , et A^c et B^c sont également indépendants.

Remarque 3.1.4. Un événement A est indépendant de lui-même si et seulement si A ou A^c est négligeable.

La notion d'indépendance d'événements se généralise à toute suite finie ou infinie d'événements.

Définition 3.1.5. Une famille A_1, \dots, A_n d'événements de \mathcal{F} sont dits *indépendants* si pour tout $k \leq n$ et i_1, \dots, i_k on a

$$\mathbf{P} \left(\bigcap_{j=1}^k A_{i_j} \right) = \prod_{j=1}^k \mathbf{P}(A_{i_j}).$$

Observons que la condition $\mathbf{P}(\bigcap_{i=1}^n A_i) = \prod_{i=1}^n \mathbf{P}(A_i)$ n'est pas une condition suffisante pour l'indépendance d'événements. Il est nécessaire de vérifier que la formule est vraie pour les intersections de toutes les sous-familles de A_1, \dots, A_n .

Proposition 3.1.6. Une famille A_1, \dots, A_n d'événements de \mathcal{F} sont indépendants si et seulement si on a

$$\mathbf{P}(\bigcap_{j=1}^n B_j) = \prod_{j=1}^n \mathbf{P}(B_j)$$

pour tous B_j tels que $B_j \in \sigma(A_j) = \{\emptyset, A_j, A_j^c, \Omega\}$, $j \leq n$.

Définition 3.1.7 (Probabilité conditionnelle). Soit A et B deux événements de \mathcal{F} tel que $\mathbf{P}(A) > 0$, la *probabilité conditionnelle* de B sachant A est la quantité définie par

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

La probabilité conditionnelle de B sachant A modélise la probabilité qu'on attribue à l'événement B sachant que l'événement A s'est réalisé. Notons que cette interprétation est compatible avec la notion d'indépendance, lorsque $\mathbf{P}(A) > 0$, les événements A et B sont indépendants si et seulement si $\mathbf{P}(B|A) = \mathbf{P}(B)$, c'est-à-dire si la probabilité que B se réalise sachant que A se réalise est inchangée.

On termine cette section avec la formule de Bayes, qui relie les probabilités conditionnelles de B sachant A et de A sachant B .

Théorème 3.1.8 (Formule de Bayes). Soit A et B deux événements vérifiant $\mathbf{P}(A), \mathbf{P}(B) > 0$. On a

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

3.2 Indépendances de variables et tribus

Avant de définir l'indépendance de variables aléatoires, une notion plus simple à définir est celle de l'indépendance de tribus.

Définition 3.2.1. Deux tribus \mathcal{F} et \mathcal{G} sont dites *indépendantes* si

$$\forall A \in \mathcal{F}, \forall B \in \mathcal{G}, \mathbf{P}(A \cap B) = \mathbf{P}(A)\mathbf{P}(B).$$

Les tribus d'une famille $(\mathcal{F}_1, \dots, \mathcal{F}_n)$ sont dites *indépendantes* si

$$\forall A_i \in \mathcal{F}_i, i \leq n, \mathbf{P} \left(\bigcap_{j=1}^n A_j \right) = \prod_{j=1}^n \mathbf{P}(A_j).$$

En d'autres termes, des tribus $\mathcal{F}_1, \dots, \mathcal{F}_n$ sont indépendantes si toute collection d'événements $A_i \in \mathcal{F}_i$ forment une suite d'événements indépendants. On remarque que cette notion d'indépendance coïncide avec celle définie pour les événements : A et B sont deux événements indépendants si et seulement si $\sigma(A)$ et $\sigma(B)$ sont indépendantes.

Étant donné l'interprétation de la notion de tribu en terme de connaissance, on retrouve ici la notion empirique d'indépendance, c'est-à-dire que deux tribus \mathcal{F} et \mathcal{G} sont indépendantes si et seulement si la connaissance représentée par la tribu \mathcal{F} ne contribue pas à obtenir des connaissances sur les événements de la tribu \mathcal{G} et réciproquement.

Exemple 3.2.2. Lorsqu'on considère le lancer de deux dés, la tribu \mathcal{F} associée au résultat du premier dé est indépendante de la tribu associée au résultat du second dé \mathcal{G} . En revanche, la tribu $\tilde{\mathcal{F}}$ associée au plus petit résultat des deux dés n'est pas indépendante de $\tilde{\mathcal{G}}$, associée au résultat du plus grand des deux dés.

Avant de définir la notion d'indépendance de variables aléatoires, on définit la notion de connaissance apportée par leur observation.

Définition 3.2.3. Soit X une variable aléatoire réelle, la *tribu engendrée par X* est la tribu définie par $\sigma(X) = \{X^{-1}(A), A \in \mathcal{B}(\mathbb{R})\}$. C'est la plus petite tribu qui rend l'application X à valeurs dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ mesurable.

Grâce à cette notion, on peut maintenant définir l'indépendance entre deux variables aléatoires.

Définition 3.2.4. Deux variables aléatoires sont dites indépendantes si les tribus $\sigma(X)$ et $\sigma(Y)$ sont indépendantes.

Les variables aléatoires d'une famille $(X_j, j \leq n)$ sont dites indépendantes si $(\sigma(X_i), i \leq n)$ est une famille de tribus indépendantes.

On liste ici quelques façons de caractériser des familles de variables aléatoires indépendantes.

Proposition 3.2.5. Soit X_1, \dots, X_n une famille de variables aléatoires.

1. Les variables aléatoires X_1, \dots, X_n sont indépendantes si et seulement si, pour tous boréliens $A_1, \dots, A_n \in \mathcal{B}(\mathbb{R})$, on a

$$\mathbf{P}(X_1 \in A_1, \dots, X_n \in A_n) = \prod_{j=1}^n \mathbf{P}(X_j \in A_j).$$

2. Les variables aléatoires X_1, \dots, X_n sont indépendantes si et seulement si, pour toutes fonctions f_1, \dots, f_n mesurables (ou continues, ou \mathcal{C}^∞ , etc.), on a

$$\mathbf{E} \left(\prod_{j=1}^n f_j(X_j) \right) = \prod_{j=1}^n \mathbf{E}(f_j(X_j)).$$

3. Les variables aléatoires X_1, \dots, X_n sont indépendantes si et seulement si pour tout vecteur $(\xi_1, \dots, \xi_n) \in \mathbb{R}^n$, on a

$$\mathbf{E} \left(\exp \left(i \sum_{j=1}^n \xi_j X_j \right) \right) = \prod_{j=1}^n \mathbf{E}(e^{i\xi_j X_j}).$$

Théorème 3.2.6. Une famille X_1, \dots, X_n de variables aléatoires est indépendante si et seulement si la loi du vecteur aléatoire $X = (X_1, \dots, X_n)$ est égale au produit des lois de ses marginales, i.e. $\mu_X = \prod_{j=1}^n \mu_{X_j}$.

En particulier, si le vecteur X possède des coordonnées indépendantes, la loi de X est caractérisée par la loi de ses marginales uni-dimensionnelles.

Corollaire 3.2.7. Si X_1, \dots, X_n sont des variables à densités, X_1, \dots, X_n sont indépendantes si et seulement si

$$\varphi_X(x_1, \dots, x_n) = \prod_{j=1}^n \varphi_{X_j}(x_j).$$

On finit cette section par l'extension de la notion d'indépendance à des suites infinies de variables aléatoires.

Définition 3.2.8. Une famille $(\mathcal{F}_i, i \in I)$ de tribus est indépendante si et seulement si pour tout $i_1, \dots, i_n \in I$, la famille $\mathcal{F}_{i_1}, \dots, \mathcal{F}_{i_n}$ est indépendante.

Une famille $(X_i, i \in I)$ de variables aléatoires est indépendante si et seulement si pour tout $i_1, \dots, i_n \in I$, la famille X_{i_1}, \dots, X_{i_n} est indépendante.

On dit que qu'une suite $(X_n, n \in \mathbb{N})$ de variables aléatoires est une suite de variables aléatoires indépendantes et identiquement distribuées (i.i.d.) si les variables aléatoires sont indépendantes et suivent toutes la même loi.

3.3 Théorèmes de Fubini

Les théorèmes de Fubini sont des formules générales de la théorie de l'intégration. On les rappelle ici puisque ce sont des résultats impliquant des mesures produits, c'est-à-dire les lois de vecteurs de coordonnées indépendantes.

Théorème 3.3.1 (Théorème de Fubini-Tonelli). *Si $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ est une fonction mesurable positive, et (X, Y) est une paire de variables aléatoires indépendantes, on a*

$$\mathbf{E}(f(X, Y)) = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} f(x, y) \mu_X(dx) \right) \mu_Y(dy) = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} f(x, y) \mu_Y(dy) \right) \mu_X(dx).$$

Théorème 3.3.2 (Théorème de Fubini-Lebesgue). *Si $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ est une fonction mesurable, (X, Y) est une paire de variables aléatoires indépendantes, et $f(X, Y)$ est intégrable, on a*

$$\mathbf{E}(f(X, Y)) = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} f(x, y) \mu_X(dx) \right) \mu_Y(dy) = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} f(x, y) \mu_Y(dy) \right) \mu_X(dx).$$

3.4 Lemme de Borel-Cantelli

On finit cette section par le rappel du théorème de Borel-Cantelli, qui permet de déterminer si une suite d'événements peut arriver ou non une infinité de fois. Si $(A_n, n \geq 1)$ est une suite d'événements, on note

$$\limsup_{n \rightarrow \infty} A_n = \bigcap_{n \in \mathbb{N}} \bigcup_{j \geq n} A_j$$

$$\liminf_{n \rightarrow \infty} A_n = \bigcup_{n \in \mathbb{N}} \bigcap_{j \geq n} A_j.$$

Ils représentent respectivement l'événement "une infinité d'événements A_j se produisent" et "à partir d'un certain rang, uniquement les événements A_j se produisent". Remarquons en passant que par de simples calculs d'ensembles, on a

$$\left(\limsup_{n \rightarrow \infty} A_n \right)^c = \liminf_{n \rightarrow \infty} A_n^c.$$

Théorème 3.4.1 (Lemme de Borel-Cantelli). *Soit $(A_n, n \geq 1)$ une suite d'événements.*

1. *Si $\sum_{n=1}^{\infty} \mathbf{P}(A_n) < \infty$, alors $\mathbf{P}(\limsup_{n \rightarrow \infty} A_n) = 0$.*
2. *Si $\sum_{n=1}^{\infty} \mathbf{P}(A_n) = \infty$ et les (A_n) sont indépendants, alors $\mathbf{P}(\limsup_{n \rightarrow \infty} A_n) = 1$.*

Les théorèmes de Borel-Cantelli sont l'un des rares outils de la théorie des probabilités permettant d'obtenir un résultat presque sûr. La loi des grands nombres, qu'on verra au chapitre suivant est un autre exemple, et le théorème de convergence des martingales, vu dans le cours de probabilités de M1 en est un troisième. Il n'y a que très peu d'autres résultats permettant de montrer cette convergence presque sûre.

3.5 Exercices

Exercice 3.1. Rappelons qu'en probabilités, on précise rarement l'espace $(\Omega, \mathcal{F}, \mathbf{P})$ utilisé, car ce choix n'a pas d'importance, tant que les variables aléatoires considérées ont la loi voulue. Cependant, il faut tout de même savoir justifier l'existence d'un tel espace de probabilité. De plus ce choix est parfois important pour coupler des variables aléatoires. On va voir ici quelques méthodes classiques de construction d'un espace de probabilité.

1. Soit μ une probabilité sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. Donner un espace de probabilité et une variable aléatoire aléatoire X sur celui-ci tels que X suit la loi μ .

2. Soit μ une probabilité sur $(\mathbb{R}^m, \mathcal{B}(\mathbb{R}^m))$ et ν une probabilité sur $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$. Donner un espace de probabilité et des variables aléatoires X et Y sur celui-ci tels que X et Y sont indépendantes et de lois respectives μ et ν .
3. Soit $0 \leq p_X \leq p_Y \leq 1$. Donner un espace de probabilité et des variables aléatoires X et Y sur celui-ci tels que X et Y suivent les lois de Bernoulli de paramètre p_X et p_Y respectivement, et tels que $X \leq Y$ p.s.
4. Soit μ et ν deux mesures de probabilité sur \mathbb{R} telles que $\mu((-\infty, x]) \leq \nu((-\infty, x])$ pour tout $x \in \mathbb{R}$. Construire deux variables aléatoires X et Y telle que X est de loi μ , Y est de loi ν et $X \geq Y$ p.s.
5. Donner un espace de probabilité et une famille croissante $(X_p)_{p \in [0,1]}$ de variables aléatoires tels que, pour tout $p \in [0, 1]$, X_p suit la loi $\mathcal{B}(p)$.

Exercice 3.2. Soit (X, Y) un vecteur aléatoire dont la loi admet pour densité

$$\varphi_{X,Y}(x, y) = \frac{1}{y} e^{-y} \mathbf{1}_{\{0 < x < y\}}.$$

1. Montrer que $\varphi_{X,Y}$ est bien une densité de probabilité.
2. Déterminer les lois marginales de X et Y .
3. Montrer que X et Y/X sont des variables aléatoires indépendantes.

Exercice 3.3. Une maladie M affecte une personne sur 1000 dans une population donnée. On dispose d'un test sanguin qui détecte M avec une fiabilité de 99% lorsque cette maladie est effectivement présente. Cependant, on obtient aussi un résultat faussement positif pour 0,2% des personnes saines testées. Quelle est la probabilité qu'une personne soit réellement malade lorsque son test est positif?

Exercice 3.4. Soit N une variable aléatoire de loi de Poisson de paramètre $\lambda > 0$. Conditionnellement à l'événement $\{N = n\}$, Z suit la loi $\text{Bin}(n, p)$ pour $p \in [0, 1]$.

1. Déterminer la loi de Z .
2. Montrer que Z et $N - Z$ sont deux variables aléatoires indépendantes.
3. Soit M une variable aléatoire de loi de Poisson de paramètre μ indépendante de N . Montrer que $M + N$ suit une loi de Poisson de paramètre $\lambda + \mu$.

Exercice 3.5 (Simulation d'une loi de Poisson). Soit $(U_n, n \geq 0)$ une suite de variables aléatoires indépendantes de loi uniforme sur $[0, 1]$ et $\lambda > 0$. On pose

$$T = \inf\{n \geq 0 : (-\log(U_0)) + \dots + (-\log U_n) \geq \lambda\}.$$

Déterminer la loi de T .

Exercice 3.6 (Simulation d'une loi conditionnelle). Soit μ une mesure de probabilité sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ et B un ensemble satisfaisant $\mu(B) > 0$. Le but de cet exercice est de déterminer une façon de simuler la loi μ_B de μ conditionnée à B , définie par

$$\forall A \in \mathcal{B}(\mathbb{R}), \mu_B(A) = \mu(B \cap A) / \mu(B).$$

1. Montrer que μ_B est bien une mesure de probabilité.
2. Soit (X_n) une suite de variables i.i.d. de loi μ , et $N = \inf\{n \in \mathbb{N} : X_n \in B\}$.
 - (a) Quelle est la loi de N ? Calculer son espérance.
 - (b) On pose $Z = X_N$, montrer que Z et N sont indépendantes. Déterminer la loi de Z .

Chapitre 4

Convergence

On considère dans ce chapitre les différents modes de convergence observables pour une suite $(X_n, n \geq 0)$ de variables aléatoires. On considère dans ce chapitre que $(X_n, n \geq 1)$ est une suite de variables aléatoires construites sur le même espace de probabilité $(\Omega, \mathcal{F}, \mathbf{P})$. Par lemme des classes monotones, on notera que la loi de la suite $(X_n, n \geq 1)$ est caractérisée par la loi des vecteurs (X_1, \dots, X_k) pour tout $k \in \mathbb{N}$.

4.1 Préliminaires sur la convergence

On rappelle les trois résultats suivants, issus de la théorie de la mesure.

Théorème 4.1.1 (Convergence monotone). *Soit (X_n) une suite croissante de variables aléatoires positives, on a*

$$\lim_{n \rightarrow \infty} \mathbf{E}(X_n) = \mathbf{E} \left(\lim_{n \rightarrow \infty} X_n \right).$$

Théorème 4.1.2 (Lemme de Fatou). *Soit (X_n) une suite de variables aléatoires positives, on a*

$$\liminf_{n \rightarrow \infty} \mathbf{E}(X_n) \geq \mathbf{E} \left(\liminf_{n \rightarrow \infty} X_n \right).$$

Théorème 4.1.3 (Convergence dominée). *Soit (X_n) une suite de variables aléatoires convergeant vers X , si il existe une variable aléatoire $Y \in L^1$ telle que $|X_n| \leq Y$ pour tout $n \in \mathbb{N}$, on a*

$$\lim_{n \rightarrow \infty} \mathbf{E}(|X_n - X|) = 0.$$

En particulier, $\lim_{n \rightarrow \infty} \mathbf{E}(X_n) = \mathbf{E}(\lim_{n \rightarrow \infty} X_n)$.

4.2 Différents modes de convergence

Définition 4.2.1. On dit que la suite $(X_n, n \geq 1)$ converge presque sûrement vers X si l'événement

$$\{\omega \in \Omega : \liminf_{n \rightarrow \infty} X_n(\omega) = \limsup_{n \rightarrow \infty} X_n(\omega) = X(\omega)\}^c$$

est négligeable.

La convergence presque sûre est la notion de convergence la plus intuitive. Elle correspond à la convergence de X_n sur un ensemble de mesure pleine. C'est également une notion de convergence très forte. Toutefois, elle n'est pas métrisable, ce qui signifie que certaines méthodes que l'on a l'habitude d'utiliser pour montrer la convergence de suites ne s'appliquent pas ici. Par exemple, le fait que de toute sous-suite de (X_n) on peut extraire une sous-suite convergeant p.s. vers X n'est pas suffisant à montrer la convergence de (X_n) vers X .

Définition 4.2.2. Pour tout $p \geq 1$, on dit que la suite $(X_n, n \geq 1)$ converge dans L^p vers X si on a

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

La convergence dans L^p est une notion de convergence très robuste, car elle est associée à la norme $\|\cdot\|_p$, tel que l'espace des variables aléatoires devient un espace de Banach. Deux notions de convergence sont particulièrement intéressantes : la convergence dans L^1 , qu'on peut obtenir grâce au théorème de convergence dominée, et la convergence dans L^2 . En effet, dans ce dernier cas, l'espace des variables aléatoires telles que $\|X\|_2 < +\infty$ forme un espace de Hilbert.

Propriété 4.2.3. Soit $1 \leq p < q$, si $X_n \rightarrow X$ dans L^q alors $X_n \rightarrow X$ dans L^p .

Définition 4.2.4. On dit que la suite $(X_n, n \geq 1)$ converge en probabilité vers X si pour tout $\varepsilon > 0$,

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

La notion de convergence en probabilité est bien plus faible que les deux notions précédentes, cependant elle permet d'identifier et de construire de façon unique la variable X sur le même espace de probabilité que la suite (X_n) . On notera qu'on peut réécrire la convergence presque sûre de (X_n) comme

$$\forall \varepsilon > 0, \mathbf{P}(\exists n \in \mathbb{N} : \forall k \geq n, |X_k - X| > \varepsilon) < \varepsilon,$$

et la convergence en probabilité comme

$$\forall \varepsilon > 0, \exists n \in \mathbb{N} : \forall k \geq n, \mathbf{P}(|X_k - X| > \varepsilon) < \varepsilon.$$

Propriété 4.2.5. On a $X_n \rightarrow X$ en loi si et seulement si $\lim_{n \rightarrow \infty} \mathbf{E}(|X_n - X| \wedge 1) = 0$.

Définition 4.2.6. On dit que la suite $(X_n, n \geq 1)$ converge en loi vers X si pour toute fonction continue bornée f , on a

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

La convergence en loi est la notion de convergence la plus faible, puisqu'elle converge uniquement la suite des lois $(\mu_{X_n}, n \geq 1)$. En d'autres termes, la convergence en loi dépend des marginales unidimensionnelles, pas de la loi de la suite infinie. En particulier, la variable X vers laquelle (X_n) converge n'est pas nécessairement construite sur l'espace de probabilité $(\Omega, \mathcal{F}, \mathcal{P})$.

Proposition 4.2.7. On a $X_n \rightarrow X$ en loi si et seulement si $\lim_{n \rightarrow \infty} \varphi_{X_n} = \varphi_X$.

4.3 Relations entre les modes de convergence

On explore ici les relations entre les différents modes de convergence. On commence par les implications liées aux modes de convergence les plus forts vers les plus faibles.

Proposition 4.3.1. Si $X_n \rightarrow X$ p.s. ou $X_n \rightarrow X$ dans L^p , alors $X_n \rightarrow X$ en probabilité.

Proposition 4.3.2. Si $X_n \rightarrow X$ en probabilité alors $X_n \rightarrow X$ en loi.

On explore ensuite quelques conditions permettant de prouver les réciproques.

Proposition 4.3.3. Si $X_n \rightarrow c$ en loi (où c est une constante), alors $X_n \rightarrow c$ en probabilité.

Une généralisation de ce résultat est le théorème de Slutsky.

Proposition 4.3.4. Si $X_n \rightarrow c$ en loi et $Y_n \rightarrow Y$ en loi, alors $(X_n, Y_n) \rightarrow (c, Y)$ en loi.

Proposition 4.3.5. Si $X_n \rightarrow X$ en probabilité, alors il existe une suite extraite de (X_n) qui converge presque sûrement.

Proposition 4.3.6. Si $X_n \rightarrow X$ en probabilité et si $\lim_{A \rightarrow \infty} \sup_{n \in \mathbb{N}} \mathbf{E}(|X_n| \mathbf{1}_{\{|X_n| > A\}}) = 0$, alors $X_n \rightarrow X$ dans L^1 .

Proposition 4.3.7. Si $X_n \rightarrow X$ en probabilité et si il existe $r > 1$ tel que $\sup_{n \in \mathbb{N}} \mathbf{E}(|X_n|^r) < \infty$, alors $X_n \rightarrow X$ dans L^p pour tout $p < r$.

4.4 Convergence de sommes de variables aléatoires

On commence par rappeler la propriété suivante, qui permet de calculer la loi de la somme de variables aléatoires indépendantes.

Propriété 4.4.1. Soit (X, Y) deux variables aléatoires indépendantes, de loi respective μ et ν . La loi de $X + Y$ est notée $\mu * \nu$ et est appelée la convolée de μ et de ν . On a

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

On introduit maintenant la loi des grands nombres.

Théorème 4.4.2. Soit $(X_n, n \geq 1)$ une suite de variables aléatoires indépendantes et identiquement distribuées, on suppose que $X_1 \in L^1$. On a

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=1}^n X_j = \mathbf{E}(X) \quad p.s.$$

Ce théorème représente une justification a posteriori de la théorie des probabilités, et la notion d'espérance. Ainsi, étant donné A un événement ayant la probabilité p de se produire, si l'expérience est répétée un grand nombre de fois, une proportion environ p vérifiera l'événement A . De la même façon, si on fait la moyenne d'un grand nombre de copies i.i.d. d'une variable aléatoire X , on obtient une valeur proche de $\mathbf{E}(X)$.

Toutefois, il ne faut pas perdre de vue que cela reste une abstraction. En paraphrasant Persi Diaconis : "Un dé lancé un million de fois devient rond". La répétition à l'identique une infinité de fois d'une expérience aléatoire reste une vision de l'esprit.

On s'intéresse maintenant à la vitesse de convergence de la suite $\left(\frac{1}{n} \sum_{i=1}^n X_i, n \geq 0\right)$ vers $\mathbf{E}(X)$ prouvée dans la loi des grands nombres. On montre que les fluctuations qu'on observe à cet ordre sont universelles, c'est le théorème central limite.

Théorème 4.4.3 (Théorème central limit). Soit (X_n) une suite de variables aléatoires i.i.d. dans L^2 . On a

$$\lim_{n \rightarrow \infty} \sqrt{n} \left(\frac{1}{n} \sum_{j=1}^n X_j - \mathbf{E}(X) \right) = \mathcal{N}(0, \mathbf{Var}(X)) \quad \text{en loi.}$$

La convergence en loi obtenue dans le théorème précédent ne peut pas être améliorée. Ainsi, il n'y a pas de variable aléatoire N de loi gaussienne vers laquelle les fluctuations de la marche aléatoire convergent.

On finit ce manuscrit en évoquant le théorème des trois séries de Kolmogorov.

Théorème 4.4.4. Soit $(X_n, n \geq 1)$ une suite de variables aléatoires indépendantes. La série $\sum_{n \in \mathbb{N}} X_n$ converge presque sûrement si et seulement

- $\sum_{n \in \mathbb{N}} \mathbf{P}(|X_n| > 1) < \infty$;
- $\sum_{n \in \mathbb{N}} \mathbf{E}(X_n \mathbf{1}_{\{|X_n| \leq 1\}}) < \infty$;
- $\sum_{n \in \mathbb{N}} \mathbf{Var}(X_n \mathbf{1}_{\{|X_n| \leq 1\}}) < \infty$.

4.5 Exercices

Exercice 4.1. Soit (X_n) une suite de variables aléatoires indépendantes, de loi $\mathcal{E}(\lambda)$.

1. Montrer la convergence en probabilité suivante

$$\frac{1}{\ln n} \max_{1 \leq k \leq n} X_k \rightarrow \frac{1}{\lambda}.$$

2. Démontrer que la suite de terme général $\max_{1 \leq k \leq n} X_k - \frac{\ln n}{\lambda}$ converge en loi, et donner la loi de la limite.

3. Montrer que $(X_n/\ln n)$ converge en probabilités vers 0. La suite $(X_n/\ln n)$ converge-t-elle vers 0 presque sûrement ?

Exercice 4.2. Soit (X_n) une suite de variables aléatoires indépendantes et identiquement distribuées, montrer que $X_n/n \rightarrow 0$ presque sûrement si et seulement si $\mathbf{E}(|X_1|) < \infty$.

Exercice 4.3. On définit la notion de Césaro-convergence pour une suite. On dit que $(x_n, n \geq 1)$ Césaro-converge vers x si $\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=1}^n x_j = x$. On dit qu'une fonction mesurable f est Césaro-continue si pour toute suite (x_n) Césaro-convergeant vers x , on a $f(x_n)$ qui Césaro-converge vers $f(x)$.

1. Montrer que si (X_n) est une suite de variables aléatoires i.i.d. dans L^1 , alors (X_n) est une suite Césaro-convergente.
2. Soit (X_n) une suite de v.a. i.i.d. de Bernoulli de paramètre p , montrer que (X_n) et $(f(X_n))$ sont Césaro-convergentes, et déterminer leur limite.
3. En déduire que si f est Césaro-continue, alors $f(x) = f(0) + xf(1)$ pour tout $x \in [0, 1]$.
4. Conclure que les fonction Césaro-continues sont les fonctions affines.

Exercice 4.4. Soit $(X_n, n \geq 1)$ une suite de variables aléatoires i.i.d. de loi de Cauchy, c'est à dire ayant la densité $\frac{1}{\pi(1+x^2)}$ par rapport à la mesure de Lebesgue. On s'intéresse à la convergence de la suite $(\frac{1}{n} \sum_{j=1}^n X_j)$.

1. Soit Y une variable ayant pour densité $\frac{\lambda}{2} e^{-\lambda|x|}$. Calculer la transformée de Fourier de la loi de Y .
2. En déduire que la transformée de Fourier de la loi de Cauchy est $\xi \mapsto e^{-|\xi|}$.
3. Montrer alors que pour tout $n \in \mathbb{N}$, $\frac{1}{n} \sum_{j=1}^n X_j$ suit une loi de Cauchy.
4. Conclure sur la convergence de $(\frac{1}{n} \sum_{j=1}^n X_j)$.

Bibliographie

- [Dur19] Rick Durrett. *Probability : Theory and Examples*. Cambridge University Press, 2019.
- [Kal02] Olav Kallenberg. *Foundations of Modern Probability*. Springer New York, 2002.
- [LG12] Jean-François Le Gall. *Intégration, Probabilités et Processus Aléatoires*. École Normale Supérieure, 2012.
- [Wil91] David Williams. *Probability with Martingales*. Cambridge University Press, 1991.