



UFR MITSIC

NOTES DE COURS

Intégration et probabilités

Bastien MALLEIN

Version du 20 septembre 2021

`mallein@math.univ-paris13.fr`
Laboratoire Analyse, Géométrie et Applications
99, Avenue Jean-Baptiste Clément
93430 Villetaneuse FRANCE

Avant-propos

L'objectif de ce cours est de fournir l'essentiel des résultats et méthodes de théorie de la mesure nécessaires à la fondation de la théorie moderne des probabilités. Il s'adresse à tout étudiant de L3 ayant déjà suivi un cours de probabilités élémentaire. Ce manuscrit a été écrit pour être lisible de façon autonome, avec un minimum de références à des ouvrages extérieurs. La plupart des résultats énoncés ici sont fournis avec une preuve. Les chapitres sont construits pour s'enchaîner logiquement les uns après les autres, et se finissent par des exercices d'application directe pour tester sa compréhension.

L'exposition faite ici ne prétend être ni originale ni optimale. Les résultats mentionnés ici sont en grande partie empruntés à d'autres excellents supports de cours de probabilités comme *Intégration, Probabilités et Processus Aléatoires* de Jean-François Le Gall. Si vous trouvez des fautes et/ou erreurs dans le présent manuscrit, ou si vous avez des suggestions pour l'améliorer, n'hésitez pas à me contacter à l'adresse mallein@math.univ-paris13.fr.

Table des matières

Introduction	7
1 Espace de probabilité	9
1.1 Univers de probabilités	9
1.2 Tribu sur un univers de probabilités	11
1.3 Loi de probabilité	14
2 Variables aléatoires et leurs lois	21
2.1 Fonctions mesurables	21
2.2 Variables aléatoires	23
2.3 Vecteurs aléatoires	25
2.4 Suites infinies de variables aléatoires	26
2.5 Lois usuelles	26
3 Espérance d'une variable aléatoire	31
4 Construction de mesures	33
5 Indépendance	35
6 Convergence d'une suite de variables aléatoires	37

Introduction

*“Un voyage de milliers de kilomètres commence
toujours par un premier pas.”*

Lao-Tseu – Tao Te King.

La théorie des probabilités a pour objet la modélisation et l’analyse d’expériences aléatoires. Une expérience aléatoire est construite comme un phénomène dont on ne connaît pas l’issue, mais dont on connaît l’ensemble des éventualités possibles pour cette réalisation ainsi que la probabilité que ces éventualités se produisent. Un exemple classique d’expérience aléatoire est le lancer d’un dé à 6 faces, un exemple classique d’expérience aléatoire. Dans cette expérience, on ne sait pas sur quelle face le dé va tomber, mais on sait que quelle que soit cette face, ce sera un nombre entre 1 et 6. La probabilité d’observer une face est de $1/6$.

Ce cours est divisé en 5 parties de tailles inégales. Dans un premier temps on introduira la notion d’espace de probabilité, qui sert à définir une expérience aléatoire. Cet espace est la donnée par l’ensemble des éventuelles réalisations possibles de l’expérience aléatoire, une famille de sous-ensembles définissant les événements observables, et une mesure de probabilité donnant la probabilité de ces événements. Ce triplet est la structure essentielle sur laquelle le reste de la théorie des probabilités est construite.

On introduira ensuite les variables aléatoires, qui peuvent être pensées comme des observables associées à l’expérience aléatoire. Si le lancer d’un dé implique de nombreuses considérations (force et angle du lancer, matière sur laquelle le dé roule, humidité de l’air, etc.), le résultat observé par l’expérimentateur est simplement la face du dé pointant vers le haut, et peut donc être construit comme une fonction de la réalisation. Les variables aléatoires sont les briques élémentaires avec lesquelles des expériences aléatoires peuvent être combinées pour obtenir les résultats fondamentaux de la théorie des probabilités.

Dans le chapitre 3, on construira l’espérance d’une variable aléatoire réelle, comme la valeur moyenne attendue par la mesure associée à cette variable. L’espérance est un outil essentiel, puisque l’espérance de fonctions d’une variable aléatoires suffit à déterminer sa loi. Cette observation est à la base de la définition d’extensions des probabilités telle que la théorie des probabilités libres. Elle permet aussi de supprimer les références à l’espace de probabilité sous-jacent, puisque n’importe quel espace de probabilité suffisamment grand pour définir la variable aléatoire considérée peut être substitué pour simuler l’expérience d’intérêt.

On définira ensuite la notion d’indépendance de variables aléatoires, permettant de donner une définition simple pour des suites infinies de variables aléatoires. De très nombreux modèles aléatoires utilisés en probabilités sont construits comme image d’une famille plus ou moins grande de variables aléatoires indépendantes.

Enfin, on exposera dans le chapitre 6 les résultats principaux de ce cours : la loi des

grands nombres et le théorème central limite. La loi des grands nombres est un théorème fondamental de la théorie des probabilités, puisqu'il justifie a posteriori la notion de probabilité et d'espérance, en observant que l'espérance d'une variable aléatoire est la moyenne des résultats obtenus sur un grand nombre d'expériences indépendantes. On explorera également dans ce chapitre les différents modes de convergences ainsi que leurs relations.

Espace de probabilité

“The theory of probability combines commonsense reasoning with calculation. It domesticates luck, making it subservient to reason.”

La théorie des probabilités combine des raisonnements de bon sens avec des formules. Elle dompte la chance, l’asservissant à la raison.

Ivars Peterson – The Jungles of Randomness.

L’objectif de la théorie des probabilités est la modélisation, l’étude et la prédiction des résultats d’une expérience aléatoires (lancers de pièces, jets de dés, résultats d’une élection, temps de désintégration d’un atome radioactif, etc.). Afin de représenter ces expériences aléatoires, on utilisera un *espace de probabilité*, qui peut être défini comme un triplet $(\Omega, \mathcal{F}, \mathbf{P})$.

Dans ce triplet, Ω est l’*univers de probabilité*, l’ensemble des réalisations possibles de l’expérience aléatoire. L’ensemble \mathcal{F} est une famille de parties de Ω appelée *tribu*, dont les éléments sont les familles de réalisations qu’il est possible d’observer, les *événements*. Enfin, la fonction \mathbf{P} est une application de \mathcal{F} dans $[0, 1]$ associant à chaque événement la *probabilité* que cet événement se réalise.

Le reste de ce chapitre est dédié à la définition détaillée des trois éléments Ω , \mathcal{F} et \mathbf{P} formant l’espace de probabilité décrit ici. On introduira dans un premier temps la notion d’univers de probabilité associé à une expérience aléatoire. On explicitera ensuite la notion de tribu ainsi que ses principales propriétés. On introduira enfin les mesures de probabilité, qui associent à chaque événement sa probabilité de réalisation.

1.1 Univers de probabilités

L’univers de probabilité est l’espace sur lequel l’expérience aléatoire d’intérêt est modélisée. Cet espace est l’ensemble des éventualités possibles impliquant cette expérience aléatoire. Un univers de probabilité est construit pour modéliser une expérience aléatoire associée, considérer différentes expériences aléatoires peut mener à l’utilisation de différents univers de probabilité.

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

On dit qu'un univers de probabilité est *discret* si Ω est fini ou dénombrable. Dans le cas contraire, on parlera d'univers de probabilité continu. La plupart des univers de probabilité discrets considérés ici peuvent se représenter comme des sous-ensembles de \mathbb{Z}^n pour un certain $n \in \mathbb{N}$, tandis que les univers continus sont régulièrement construits comme des intervalles de \mathbb{R}^n , ou comme $E^{\mathbb{N}}$ l'ensemble des suites à valeurs dans un espace E contenant au moins deux points.

Exemple 1.1.2 (Quelques univers de probabilité.).

1. On considère l'expérience aléatoire consistant au jet d'un dé à 6 faces. Un univers de probabilité associé est $\Omega^{(1)} = \{1, 2, 3, 4, 5, 6\}$. Chaque entier $j \in \Omega^{(1)}$ correspond à l'éventualité « le dé tombe sur la face j ».
2. $\Omega^{(2)} = \{P, F\}^3$ est un univers de probabilités pour l'expérience aléatoire consistant au jet de 3 pièces de monnaie. L'éventualité $(P, F, P) \in \Omega$ correspond à l'événement « la première pièce tombe sur pile, la seconde sur face et la troisième sur pile ».
3. $\Omega^{(3)} = \{(1, 2, 3), (1, 3, 2), (2, 1, 3), (2, 3, 1), (3, 1, 2), (3, 2, 1)\}$ 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. $\Omega^{(4)} = \mathbb{R}_+$ est un univers de probabilité pour l'expérience consistant à observer la durée de vie d'une ampoule. L'éventualité $t \in \Omega^{(4)}$ correspond à une ampoule grillant au bout d'un temps t .
5. $\Omega^{(5)} = \{P, F\}^{\mathbb{N}}$ est un univers de probabilité pour l'expérience consistant au jet d'une infinité de pièces de monnaie, chaque suite de P et F représentant la suite de résultats obtenus lors des jets successifs de la pièce.

Les univers $\Omega^{(1)}$, $\Omega^{(2)}$ et $\Omega^{(3)}$ sont des univers de probabilité discrets, tandis que $\Omega^{(4)}$ et $\Omega^{(5)}$ sont des univers de probabilité continus.

Il n'existe pas un unique univers de probabilités permettant de représenter une expérience aléatoire, et différentes constructions sont possibles. Lorsqu'un univers de probabilité est l'univers le plus simple possible pour représenter cette expérience aléatoire, on parle d'*univers canonique*. Les univers décrits ci-dessus sont tous des univers canoniques. 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 1.1.3. 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'orientation de la pièce au moment où 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.

Si Ω_1 est un univers de probabilité permettant de modéliser une première expérience et Ω_2 un univers de probabilité permettant d'en modéliser une seconde, alors l'univers produit $\Omega_1 \times \Omega_2$ forme un univers de probabilité permettant la modélisation de deux expériences successives. Par exemple, l'univers $\Omega^{(1)} \times \Omega^{(2)}$ permet modéliser le lancer d'un dé, suivi du lancer de trois pièces.

Si Ω est un univers de probabilité permettant de modéliser une première expérience et si pour tout $\omega \in \Omega$, A_ω est un autre univers de probabilité, alors $\cup_{\omega \in \Omega} \{\omega\} \times A_\omega$ permet de modéliser l'expérience aléatoire consistant à réaliser la première expérience, puis en fonction du résultat de cette expérience de réaliser l'expérience associée à A_ω . Ainsi, l'univers $\cup_{j=1}^6 \{j\} \times \{P, F\}^j$ permet de modéliser l'expérience suivante : on jette un dé, puis un nombre de pièces égal au résultat du dé.

1.2 Tribu sur un univers de probabilités

Une tribu sur un univers de probabilités Ω est une collection de sous-ensembles de Ω . La tribu regroupe l'ensemble des événements d'une expérience aléatoire qu'il est possible d'observer, et auxquels on peut associer une probabilité d'occurrence.

Définition 1.2.1. Une *tribu* \mathcal{F} sur Ω est un sous-ensemble de $\mathcal{P}(\Omega)$ (l'ensemble des parties de Ω) qui satisfait les trois axiomes suivants :

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

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

On rappelle qu'un ensemble I est appelé dénombrable si il existe une injection de I dans \mathbb{N} . Dans le cas contraire on dit que I est indénombrable. Un ensemble I est infini dénombrable si et seulement si il est possible d'énumérer ses éléments par \mathbb{N} $I = \{e_j, j \in \mathbb{N}\}$.

Exemple 1.2.2. Les ensembles \mathbb{N}^2 , \mathbb{Z} et \mathbb{Q} sont dénombrables. Les ensembles $\{0, 1\}^{\mathbb{N}}$ et \mathbb{R} sont indénombrables.

Un ensemble $A \in \mathcal{F}$ est appelé un *événement* de la 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*.

Propriété 1.2.3. Si \mathcal{F} est une tribu, alors

1. on a $\emptyset \in \mathcal{F}$;
2. \mathcal{F} est stable par union finie ;
3. \mathcal{F} est stable par intersections finies ou dénombrables.

Démonstration. Soit \mathcal{F} une tribu. Par définition, $\Omega \in \mathcal{F}$, et donc $\Omega^c = \emptyset \in \mathcal{F}$.

Soit $n \in \mathbb{N}$ et B_1, \dots, B_n des événements de \mathcal{F} . On pose $B_{n+1} = B_{n+2} = \dots = \emptyset$. Dans ce cas, par définition, on a également $\bigcup_{k \geq 1} B_k = \bigcup_{k=1}^n B_k \in \mathcal{F}$, ce qui montre que \mathcal{F} est stable par union finie.

Enfin, si $(A_n, n \geq 0)$ est une suite d'événements de \mathcal{F} , alors pour tout $n \in \mathbb{N}$, on a aussi $A_n^c \in \mathcal{F}$. Par conséquent $\bigcup_{n \in \mathbb{N}} A_n^c$ appartient à \mathcal{F} , donc c'est également le cas de $(\bigcup_{n \in \mathbb{N}} A_n^c)^c = \bigcap_{n \in \mathbb{N}} A_n$. Cela montre que \mathcal{F} est stable par intersection dénombrable. En posant $A_{n+1} = A_{n+2} = \dots = \Omega$, on montre également la stabilité de \mathcal{F} par intersections finies. \square

Une tribu sur Ω est utilisée pour modéliser une connaissance accessible sur l'expérience aléatoire. Les ensembles $A \in \mathcal{F}$ représentent alors les événements pour lesquels, avec les observations considérées sur l'expérience, on peut répondre à la question suivante : « Est-ce que l'éventualité $\omega \in \Omega$ appartient à A ou à A^c ? ». On peut ainsi utiliser plusieurs tribus pour modéliser différents niveaux de connaissances sur une même expérience aléatoire.

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).

Exemple 1.2.4 (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 1.1.3, la tribu associée à l'observation du côté duquel la pièce est tombée est donnée par $\mathcal{F} = \{\emptyset, [0, \pi), [\pi, 2\pi), \Omega\}$.

Exemple 1.2.5 (Différentes tribus pour une même expérience aléatoire). . On considère l'expérience suivante : on lance un dé à 6 faces. On communique à un premier observateur la parité de la face du dé, et à un second si ce résultat est plus grand que 4 ou plus petit que 3. L'univers de probabilité associé à cette expérience est $\Omega = \{1, 2, 3, 4, 5, 6\}$. Les tribus correspondant aux mesures des différents observateurs sont

$$\begin{aligned}\mathcal{F}_1 &= \{\emptyset, \{1, 3, 5\}, \{2, 4, 6\}, \{1, 2, 3, 4, 5, 6\}\} \\ \mathcal{F}_2 &= \{\emptyset, \{1, 2, 3\}, \{4, 5, 6\}, \{1, 2, 3, 4, 5, 6\}\}.\end{aligned}$$

Une tribu étant un ensemble de sous-ensembles de Ω , on peut lui appliquer les opérations ensemblistes d'union et d'intersection. Il est notable que si une union de tribus n'est pas toujours une tribu, la notion de tribu est stable par intersections.

Proposition 1.2.6. *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.*

Démonstration. On montre que \mathcal{C} satisfait les trois axiomes des tribus. Observons tout d'abord que pour tout $i \in I$, $\Omega \in \mathcal{F}_i$, par conséquent $\Omega \in \mathcal{C}$.

Considérons maintenant un élément $A \in \mathcal{C}$. Pour tout $i \in I$, on a $A \in \mathcal{F}_i$, donc $A^c \in \mathcal{F}_i$. Par conséquent, on a également $A^c \in \bigcap_{i \in I} \mathcal{F}_i = \mathcal{C}$.

Enfin, soit $(A_n) \in \mathcal{C}^{\mathbb{N}}$. Pour tout $i \in I$, on a également $\bigcup_{n \in \mathbb{N}} A_n \in \mathcal{F}_i$. Par conséquent $\bigcup_{n \in \mathbb{N}} A_n \in \mathcal{C}$. On a donc bien montré que \mathcal{C} est une tribu. \square

La proposition 1.2.6 permet de définir la plus petite tribu vérifiant une propriété, qui est définie comme étant l'intersection de toutes les tribus satisfaisant cette propriété. On définit en particulier la notion de tribu engendrée.

Définition 1.2.7 (Tribu engendrée). Pour toute famille $\mathcal{A} \subset \mathcal{P}(\Omega)$ de sous-ensembles de Ω , 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} .

Une *partie génératrice* de \mathcal{F} est un ensemble $\mathcal{A} \subset \mathcal{P}(\Omega)$ tel que $\sigma(\mathcal{A}) = \mathcal{F}$.

Soient \mathcal{E} et \mathcal{F} deux tribus sur Ω . On note $\mathcal{E} \vee \mathcal{F} := \sigma(\mathcal{E} \cup \mathcal{F})$ la *tribu engendrée par \mathcal{E} et \mathcal{F}* . Cette tribu est généralement plus grande 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 1.2.8. En reprenant les notations de l'exemple 1.2.5, la tribu $\mathcal{F}_1 \vee \mathcal{F}_2$ représente la connaissance que les observateurs obtiennent lorsqu'ils mettent en commun leurs informations. On a

$$\mathcal{F}_1 \vee \mathcal{F}_2 = \sigma(\{\{1, 3\}, \{2\}, \{4, 6\}, \{5\}\}),$$

qui contient 16 événements. C'est un ensemble strictement plus grand que $\mathcal{F}_1 \cup \mathcal{F}_2$, qui contient 6 éléments.

Un cas particulier de tribu engendrée est la tribu produit, définie sur l'espace produit de deux espaces mesurés.

Définition 1.2.9 (Tribu produit). Soient (E, \mathcal{E}) et (F, \mathcal{F}) deux espaces mesurables. On peut munir le produit $E \times F$ de la *tribu produit*

$$\mathcal{E} \otimes \mathcal{F} = \sigma(A \times B; A \in \mathcal{E}, B \in \mathcal{F}).$$

Les ensembles de la forme $A \times B$ sont appelés les *pavés mesurables*.

La tribu engendrée par l'ensemble des ouverts d'un espace topologique est la tribu borélienne de cette ensemble.

Définition 1.2.10 (Tribu borélienne). Si E est un espace topologique, on appelle *tribu borélienne sur E* la tribu engendrée par l'ensemble \mathcal{O} des ouverts de E . C'est également la tribu engendrée par l'ensemble des boules ouvertes, ou l'ensemble des boules fermées.

Intermède : les classes monotones

Une tribu, en temps que famille de sous-ensembles de Ω , satisfait trois propriétés dont la troisième (stabilité par union dénombrable) est relativement restrictive. Il peut être complexe de démontrer cette troisième propriété en pratique. On peut alors se tourner sur la notion plus générale de classe monotone, avec laquelle il peut être plus simple de travailler.

Définition 1.2.11. 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 $\bigcup_{n \in \mathbb{N}} A_n \in \mathcal{C}$.

Notons que l'on peut écrire

$$B \setminus A = B \cap A^c = (B^c \cup A)^c,$$

donc si A et B appartiennent à la tribu \mathcal{E} , alors $B \setminus A \in \mathcal{E}$. Ainsi la différence entre une tribu et une classe monotones est que, contrairement à une tribu, une classe monotone est uniquement stable par union croissante.

Remarque 1.2.12.

1. Toute tribu est une classe monotone.
2. Une intersection de classes monotones est une classe monotone. On peut donc introduire la notion de classe monotone engendrée par $\mathcal{A} \subset \mathcal{P}(\Omega)$, définie par

$$\mathcal{M}(\mathcal{A}) := \bigcap_{\substack{\mathcal{C} \text{ classe monotone,} \\ \mathcal{A} \subset \mathcal{C}}} \mathcal{C}.$$

3. On a $\mathcal{M}(\mathcal{A}) \subset \sigma(\mathcal{A})$.

Le résultat suivant justifie l'introduction de la notion de classe monotone, en la reliant à celle de tribu. Il montre que sous certaines hypothèses satisfaites par 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 1.2.13 (Lemme des classes monotones). *Pour toute famille \mathcal{A} de sous-ensembles de Ω qui vérifie $\forall n \in \mathbb{N}, \forall A_1, \dots, A_n \in \mathcal{A}, \bigcap_{i=1}^n A_i \in \mathcal{A}$, on a $\mathcal{M}(\mathcal{A}) = \sigma(\mathcal{A})$.*

Démonstration. Notons pour commencer qu'il est immédiat que $\mathcal{M}(\mathcal{A}) \subset \sigma(\mathcal{A})$, puisque toute tribu est également une classe monotone. On montre le lemme des classes monotones en prouvant que $\mathcal{M}(\mathcal{A})$ est une tribu. Pour ce faire, on montre dans un premier temps que $\mathcal{M}(\mathcal{A})$ est stable par intersections finies.

Soit $A \in \mathcal{A}$, on pose $\mathcal{M}_A = \{B \in \mathcal{M}(\mathcal{A}) : A \cap B \in \mathcal{M}(\mathcal{A})\}$. On observe alors que \mathcal{M}_A est une classe monotone, en effet :

1. on a $\Omega \cap A = A \in \mathcal{A} \subset \mathcal{M}(\mathcal{A})$, donc $\Omega \in \mathcal{M}_A$;
2. si $B, C \in \mathcal{M}_A$ avec $B \subset C$, on a $A \cap B$ et $A \cap C \in \mathcal{M}(\mathcal{A})$, par conséquent on a également $(B \setminus C) \cap A = (B \cap A) \setminus (C \cap A) \in \mathcal{M}(\mathcal{A})$, d'où $B \setminus C \in \mathcal{M}_A$;
3. enfin, si (B_n) est une suite croissante d'événements de \mathcal{M}_A , on a $A \cap B_n \in \mathcal{M}(\mathcal{A})$ pour tout $n \in \mathbb{N}$, donc $A \cap \bigcup_{n \in \mathbb{N}} B_n = \bigcup_{n \in \mathbb{N}} A \cap B_n \in \mathcal{M}(\mathcal{A})$, donc $\bigcup_{n \in \mathbb{N}} B_n \in \mathcal{M}_A$.

Comme \mathcal{A} est stable par intersections finies, on a $\mathcal{A} \subset \mathcal{M}_A$. Puisque $\mathcal{M}(\mathcal{A})$ est la plus petite classe monotone contenant \mathcal{A} , on obtient $\mathcal{M}(\mathcal{A}) \subset \mathcal{M}_A$. On a ainsi montré que pour tout $A \in \mathcal{A}$ et $B \in \mathcal{M}(\mathcal{A})$, $A \cap B \in \mathcal{M}(\mathcal{A})$.

On se fixe ensuite $A \in \mathcal{M}(\mathcal{A})$, et on pose $\overline{\mathcal{M}}_A = \{B \in \mathcal{M}(\mathcal{A}) : A \cap B \in \mathcal{M}(\mathcal{A})\}$. Par le même raisonnement que précédemment, on observe que $\overline{\mathcal{M}}_A$ est une classe monotone, et donc $\mathcal{M}(\mathcal{A}) = \overline{\mathcal{M}}_A$. Cela montre que pour tout $A, B \in \mathcal{M}(\mathcal{A})$, $A \cap B \in \mathcal{M}(\mathcal{A})$.

Plus généralement, si $A_1, \dots, A_n \in \mathcal{M}(\mathcal{A})$, on a

$$A_1 \cap A_2 \cap \dots \cap A_n = (A_1 \cap A_2) \cap A_3 \cap \dots \cap A_n,$$

et $A_1 \cap A_2 \in \mathcal{M}(\mathcal{A})$. Donc par une récurrence immédiate, on obtient que $\mathcal{M}(\mathcal{A})$ est stable par toute intersection finies.

Pour compléter cette preuve, on montre qu'une classe monotone stable par intersections finies est une tribu. En effet, par passage au complémentaire, pour tout $B_1, \dots, B_n \in \mathcal{M}(\mathcal{A})$, on a $B_1 \cup \dots \cup B_n \in \mathcal{M}(\mathcal{A})$. Soit (B_n) une suite (quelconque) d'événements de $\mathcal{M}(\mathcal{A})$, on pose $C_n = B_1 \cup \dots \cup B_n$. Donc (C_n) est une suite croissante d'événements de $\mathcal{M}(\mathcal{A})$. Par propriété de classe monotone, on obtient $\bigcup_{n \in \mathbb{N}} C_n = \bigcup_{n \in \mathbb{N}} B_n \in \mathcal{M}(\mathcal{A})$, ce qui montre la stabilité de $\mathcal{M}(\mathcal{A})$ par union dénombrable.

Finalement, puisque $\mathcal{M}(\mathcal{A})$ est une tribu contenant \mathcal{A} , on en déduit que $\sigma(\mathcal{A}) \subset \mathcal{M}(\mathcal{A})$, ce qui, au vu de l'inclusion observée en début de preuve, permet de conclure à l'égalité entre $\sigma(\mathcal{A})$ et $\mathcal{M}(\mathcal{A})$. \square

Le lemme des classes monotones est également souvent employé sous la forme suivante : « toute famille d'événements –contenant l'espace total– stable par passage au complémentaire, union croissante et intersection finie est une tribu ».

On peut utiliser ce théorème pour démontrer que tous les événements d'une tribu satisfont une propriété : on montre que la propriété est satisfaite par une famille génératrice stable par intersections finies, puis que l'ensemble des événements satisfaisant cette propriété forme une classe monotone, ce qui permet de conclure.

1.3 Loi de probabilité

Une loi de probabilité forme le dernier élément constitutif d'un espace de probabilité. Étant donné un espace mesuré (Ω, \mathcal{F}) , une loi de probabilité mesure la *probabilité d'occurrence* de chaque événement de \mathcal{F} . Dans une vision fréquentiste de la théorie des probabilités, elle

représente la proportion asymptotique d'expériences pour lesquelles cet événement est réalisé lorsqu'on répète un grand nombre de fois cette expérience aléatoire. Lorsque l'expérience ne peut pas être répétée (météo du lendemain au vu des conditions actuelles, résultats d'une élection au vu des sondages actuels, etc.), cette probabilité peut représenter le degré de croyance de réalisation de cet événement.

On verra au chapitre 6 que la théorie moderne des probabilités telle qu'exposée ici, et issue des travaux de Kolmogorov, est bien en accord avec l'interprétation fréquentiste de la probabilité. On verra également au chapitre 5 la formule de Bayes, qui est à la base de l'interprétation en terme de degré de croyance de la probabilité d'un événement.

Définition 1.3.1. Une *mesure* (ou *loi*) de probabilité sur l'espace mesuré (Ω, \mathcal{F}) est une application $\mathcal{F} \rightarrow \mathbb{R}$ satisfaisant les propriétés 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 $i \neq j$), $\mathbf{P}(\cup_{n \in \mathbb{N}} A_n) = \sum_{n \in \mathbb{N}} \mathbf{P}(A_n)$.

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

$$\mathbf{P}(A) = \mathbf{P}(\cup_{\omega \in A} \{\omega\}) = \sum_{\omega \in A} \mathbf{P}(\{\omega\}).$$

Par conséquent, une mesure de probabilité sur un espace discret est caractérisée par sa *fonction de masse*, définie par $\omega \in \Omega \mapsto \mathbf{P}(\{\omega\})$.

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 sur (Ω, \mathcal{F}) . On considère une suite croissante $(A_n, n \geq 0)$ 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

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

Démonstration. Pour tout $n \in \mathbb{N}$, on pose $R_n = A_{n+1} \setminus A_n$. On a alors, pour tout $n \in \mathbb{N}$

$$A_n = \bigcup_{j=1}^n A_j = A_1 \cup \bigcup_{j=1}^{n-1} R_j,$$

avec (A_1, R_1, \dots) deux à deux disjoints. Grâce à la σ -additivité de \mathbf{P} , on a donc

$$\mathbf{P}(A_n) = \mathbf{P}(A_1) + \sum_{j=1}^n \mathbf{P}(R_j) \quad \text{et} \quad \mathbf{P}\left(\bigcup_{j \geq 1} A_j\right) = \mathbf{P}(A_1) + \sum_{j=1}^{\infty} \mathbf{P}(R_j) \leq 1.$$

La série de terme positifs ayant un plus grand élément borné par 1, elle converge donc. En utilisant que $\mathbf{P}(R_j) = \mathbf{P}(A_{j+1}) - \mathbf{P}(A_j)$, on obtient

$$\mathbf{P}\left(\bigcup_{j \geq 1} A_j\right) = \lim_{n \rightarrow \infty} \mathbf{P}(A_1) + \sum_{j=1}^n \mathbf{P}(A_{j+1}) - \mathbf{P}(A_j) = \lim_{n \rightarrow \infty} \mathbf{P}(A_n),$$

par télescopage. □

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 la loi \mathbf{P} sur (Ω, \mathcal{F}) , il suffit de donner la valeur de $\mathbf{P}(A)$ pour tout $A \in \mathcal{A}$, où \mathcal{A} est une famille stable par intersection finie génératrice de \mathcal{F} . Notons que définir \mathbf{P} sur \mathcal{A} garantit bien l'unicité de la loi décrite, 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 1.2.13, on en déduira que $\sigma(\mathcal{A}) \subset \mathcal{G}$, ce qui complétera 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} , par la proposition 1.3.6 on a :

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

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. On montera l'existence de la mesure de Lebesgue au chapitre 4. Dans l'intervalle, on supposera l'existence de cette mesure de Lebesgue.

Une mesure de probabilité peut être associée à une notion de *mesure*, ou « masse », d'un ensemble. Pour chaque événement $A \in \mathcal{F}$, la fonction \mathbf{P} donne la masse associée à cet événement. La masse totale est normalisée à 1, et la masse de la réunion d'événements disjoints est égale à la somme des masses. Une notion plus générale de mesure σ -finie est définie comme suit.

Définition 1.3.10. Une *mesure* sur l'espace mesuré (E, \mathcal{E}) est une application $\mu : \mathcal{E} \rightarrow \mathbb{R}$ satisfaisant les propriétés suivantes :

- pour tout $A \in \mathcal{F}$, $\mu(A) \geq 0$,
- pour toute suite $(A_n, n \geq 0) \in \mathcal{F}^{\mathbb{N}}$ d'événements deux à deux disjoints, $\mu(\cup_{n \in \mathbb{N}} A_n) = \sum_{n \in \mathbb{N}} \mu(A_n)$.

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

1. On note \mathcal{A} la tribu correspondant à l'observation de la parité du résultat du premier dé. Décrire la tribu \mathcal{A} .
2. On note \mathcal{B} la tribu correspondant à l'observation du résultat du second dé. Donner une partie génératrice de \mathcal{B} .
3. Décrire la tribu $\mathcal{A} \vee \mathcal{B}$. À quelle type de connaissance sur l'expérience aléatoire correspond-t-elle? Quelle tribu \mathcal{C} peut-on ajouter pour obtenir la connaissance complète du résultat des deux dés.

Exercice 1.3. Soit Ω un univers de probabilité, on note (A_1, \dots, A_n) une partition de Ω , c'est-à-dire une famille d'événements deux à deux disjoints tels que $\bigcup_{j=1}^n A_j = \Omega$. Quel est le cardinal de $\sigma(A_1, \dots, A_n)$?

Exercice 1.4. 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-t-il à \mathcal{G} ? Et $A \cup B \cup C$?

Exercice 1.5 (Union et intersection de tribus). 1. Montrer qu'une intersection de tribus est une tribu, mais qu'une union de tribus n'est pas forcément une tribu.

2. Pour chaque entier n soit \mathcal{F}_n la tribu de \mathbb{N} engendrée par l'ensemble $\{0\}, \{1\}, \dots, \{n\}$. Montrer que (\mathcal{F}_n) est une suite croissante de tribus mais que $\bigcup \mathcal{F}_n$ n'est pas une tribu.

Exercice 1.6 (Restriction d'une tribu). Soient \mathcal{F} une tribu sur E et B un élément de \mathcal{F} . Montrer que $\mathcal{F}_B := \{A \cap B, A \in \mathcal{F}\}$ est une tribu de B .

Exercice 1.7 (Tribu image réciproque). Soient (E, \mathcal{E}) et (F, \mathcal{F}) des espace mesurable. Soit $f : E \rightarrow F$ une application.

1. On définit

$$\mathcal{A} := \{f^{-1}(B) : B \in \mathcal{F}\}.$$

Montrer que \mathcal{E} est une tribu sur E . On l'appelle la tribu image réciproque de \mathcal{F} par f .

2. On définit

$$\mathcal{B} := \{B \subset F : f^{-1}(B) \in \mathcal{E}\},$$

montrer que \mathcal{B} est une tribu.

Exercice 1.8 (Tribu dyadique). Pour tout $n \in \mathbb{N}$, on note

$$\mathcal{B}_n = \sigma(\{(k/2^n, (k+1)/2^n], 0 \leq k \leq 2^n - 1\}).$$

1. Décrire la tribu \mathcal{B}_n .
2. Montrer que la tribu engendrée par $\cup_{n \in \mathbb{N}} \mathcal{B}_n$ est la tribu des boréliens de l'intervalle $[0, 1]$.

Exercice 1.9 (Limsup et liminf de suites). Soit $(a_n)_{n \geq 0}$ une suite de réels, on pose

$$\limsup_{n \rightarrow \infty} a_n = \lim_{n \rightarrow \infty} \sup_{k \geq n} a_k \quad \text{et} \quad \liminf_{n \rightarrow \infty} a_n = \lim_{n \rightarrow \infty} \inf_{k \geq n} a_k.$$

1. Montrer que $\limsup_{n \rightarrow \infty} a_n$ et $\liminf_{n \rightarrow \infty} a_n$ sont respectivement la plus grande et la plus petite valeur d'adhérence de la suite $(a_n)_{n \geq 0}$, en autorisant les valeurs d'adhérence infinies.
2. Vérifier que a_n converge vers $\ell \in \mathbb{R} \cup \{-\infty, +\infty\}$ si et seulement si

$$\limsup_{n \rightarrow \infty} a_n = \liminf_{n \rightarrow \infty} a_n = \ell.$$

3. Soit $(b_n)_{n \geq 0}$ une autre suite de réels. Montrer que

$$\limsup_{n \rightarrow \infty} (a_n + b_n) \leq \limsup_{n \rightarrow \infty} a_n + \limsup_{n \rightarrow \infty} b_n.$$

A-t-on toujours $\limsup_{n \rightarrow \infty} (a_n + b_n) = \limsup_{n \rightarrow \infty} a_n + \limsup_{n \rightarrow \infty} b_n$?

Variables aléatoires et leurs lois

“Le nom seul de calcul des probabilités est un paradoxe : la probabilité, opposée à la certitude, c’est ce qu’on ne sait pas, et comment peut-on calculer ce que l’on ne connaît pas ?”

Henri Poincaré – La science et l’hypothèse.

On supposera fixé, dans tout le reste de ce cours, un univers de probabilités $(\Omega, \mathcal{F}, \mathbf{P})$. 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, révélant une information partielle sur cette expérience.

La notion de variable aléatoire permet de construire de nombreux raisonnements sans nécessité de définir explicitement, ou de construire, l’espace de probabilité associé à l’expérience aléatoire, ce qui peut être fastidieux ou compliqué. Si la manipulation des variables aléatoires demande quelques précautions, elle permet aussi de simplifier grandement l’écriture des raisonnements.

On introduira tout d’abord la notion de fonction mesurable et de variable aléatoire, ainsi que la loi d’une variable aléatoire. On s’intéressera ensuite à des versions multidimensionnelles, les vecteurs aléatoires et les suites de variables aléatoires. Enfin, on introduira quelques lois de probabilité usuelles ainsi que leurs principales propriétés.

2.1 Fonctions mesurables

Une variable aléatoire est une fonction de Ω dans \mathbb{R} satisfaisant une propriété permettant de calculer la probabilité que cette variable vérifie une propriété : la *mesurabilité*. On introduit dans un premier temps la notion de fonction mesurable.

Définition 2.1.1. Soit (E, \mathcal{E}) et (F, \mathcal{F}) deux espaces mesurés. On dit que l’application $f : E \rightarrow F$ est mesurable si $f^{-1}(A) \in \mathcal{E}$ pour tout $A \in \mathcal{F}$.

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

La mesurabilité est une propriété issue de la théorie de la mesure. Elle peut être interprétée comme la marque que la connaissance fournie par \mathcal{E} (ou $\bar{\mathcal{E}}$) est suffisante pour donner la valeur de f . Nous allons toutefois citer quelques propriétés permettant de montrer qu’une fonction est mesurable.

Proposition 2.1.2. Soient (E, \mathcal{E}) et (F, \mathcal{F}) deux espaces mesurés. Une fonction $f : E \rightarrow F$ est mesurable si et seulement si il existe $\mathcal{C} \subset \mathcal{F}$ tel que $\sigma(\mathcal{C}) = \mathcal{F}$ et $f^{-1}(\mathcal{C}) \subset \mathcal{A}$.

Démonstration. Soit $\mathcal{G} = \{B \in \mathcal{B} : f^{-1}(B) \in \mathcal{A}\}$. On observe que \mathcal{G} est une tribu, et de plus $\mathcal{C} \subset \mathcal{G}$. On en déduit $\mathcal{B} = \sigma(\mathcal{C}) \subset \mathcal{G}$, et donc f est mesurable. \square

Exemple 2.1.3. 1. Lorsque (F, \mathcal{F}) est $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, il suffit de montrer que $f^{-1}(]-\infty, a]) \in \mathcal{E}$ pour tout $a \in \mathbb{R}$.

2. Lorsque \mathcal{F} est la tribu borélienne de \mathcal{F} , il suffit de montrer que $f^{-1}(O) \in \mathcal{E}$ pour tout ouvert O de F .

Proposition 2.1.4. Soit (E, \mathcal{E}) , (F, \mathcal{F}) et (G, \mathcal{G}) trois espaces mesurés.

1. Si $f : E \rightarrow F$ et $g : F \rightarrow G$ sont mesurables, alors $g \circ f$ est mesurable de E dans G .
2. Si $f : E \rightarrow F$ et $g : E \rightarrow G$ sont mesurables, alors $(f, g) : E \rightarrow F \times G$ est mesurable par rapport à la tribu produit $\mathcal{F} \otimes \mathcal{G}$.

Démonstration. 1. Soit $A \in \mathcal{G}$, on observe que $g^{-1}(A) \in \mathcal{F}$ par mesurabilité de g . On en déduit donc bien que

$$(g \circ f)^{-1}(A) = f^{-1}(g^{-1}(A)) \in \mathcal{E},$$

par mesurabilité de F .

2. Par proposition 2.1.2, il suffit de montrer que pour tout pavé mesurable $A \times B \in \mathcal{F} \otimes \mathcal{G}$, on a

$$(f, g)^{-1}(A, B) = f^{-1}(A) \cap g^{-1}(B) \in \mathcal{E},$$

ce qui est immédiat par stabilité de \mathcal{E} par intersections finies. \square

Le cas d'espaces topologiques munis de la tribu borélienne permet de vérifier aisément la mesurabilité d'un grand nombre de fonctions.

Proposition 2.1.5. Soient E et F deux espaces topologiques. On note \mathcal{E} et \mathcal{F} les tribus boréliennes associées à E et F . Toute fonction continue de E dans F est alors mesurable.

Démonstration. Soit f une fonction continue de E dans F . Par définition, pour tout O ouvert de F , $f^{-1}(O)$ est un ouvert de E . Par proposition 2.1.2, on en conclut que f est mesurable. \square

Dans le cas des fonctions à valeurs réelles, de nombreuses opérations sur les fonctions préservent la mesurabilité.

Propriété 2.1.6. Soit (E, \mathcal{E}) un espace mesuré, et \mathbb{R} est équipé de la tribu borélienne.

1. Si $A \in \mathcal{E}$, la fonction $\mathbb{1}_A : E \rightarrow \mathbb{R}$ est mesurable.
2. Si f et g sont des fonctions mesurables à valeurs dans \mathbb{R} , alors fg et $\lambda f + \mu g$ sont des 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.

Démonstration. 1. On observe que pour tout ouvert O de \mathbb{R} , on a

$$\mathbb{1}_A^{-1}(O) \in \{ \emptyset, A, A^c, E \} \subset \mathcal{E},$$

donc $\mathbb{1}_A$ est mesurable.

2. Ce sont des conséquences immédiates des propositions 2.1.4 et 2.1.5. Ainsi par exemple fg est la composition de l'applications (f, g) par $(x, y) \mapsto xy$, qui sont toutes deux mesurables (la première comme couple de fonctions mesurables, la seconde comme fonction continue). \square

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. En pratique, la plupart des fonctions numériques sont mesurables par rapport à la tribu borélienne. Au contraire, la construction de fonctions non-mesurables est complexe et dépasse le cadre de ces notes.

Remarque 2.1.7. Grâce aux propriétés précédentes, on observe que si $(A_n, n \in \mathbb{N})$ est une suite d'événements deux à deux disjoints et $(\lambda_n, n \in \mathbb{N})$ est une suite de réels, alors la fonction $f = \sum_{n \in \mathbb{N}} \lambda_n \mathbb{1}_{A_n}$ est une fonction mesurable de E dans \mathbb{R} qu'on appelle *fonction étagée*.

2.2 Variables aléatoires

On introduit maintenant la notion de variable aléatoire, un objet central de la théorie des probabilités.

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

En d'autres termes, une variable aléatoire est une fonction mesurable d'un espace de probabilité dans \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.

La loi d'une variable aléatoire est définie comme la mesure image de la mesure \mathbf{P} par l'application X . C'est une mesure de probabilité sur \mathbb{R} , munie de la tribu borélienne.

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

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

Le support de la loi de la variable aléatoire X décrit l'ensemble des valeurs prises par la variable aléatoire X , à un ensemble négligeable près. Pour éviter toute référence à l'univers de probabilités, une variable aléatoire X est souvent décrite à partir de sa loi de probabilité.

Grâce au lemme des classes monotones, 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 $x \in \mathbb{R}$.

Démonstration. Ces caractérisations de la loi de X sont des conséquences immédiates du théorème des classes monotones, en particulier du Lemme 1.3.7. Les familles d'intervalles considérées sont stables par intersection finies, et engendrent la tribu borélienne entière.

Par exemple en considérant $\mathcal{A} = \{] - \infty, x], x \in \mathbb{R} \}$, on observe que pour tout $a < b$, $] - \infty, b[\setminus] - \infty, a[=]a, b[$. Puis, en choisissant une suite $(b_n, n \in \mathbb{N})$ croissant vers b , on a $\cup_{n \in \mathbb{N}}]a, b_n[=]a, b[$. Par conséquent, la tribu engendrée par \mathcal{A} contient tous les intervalles ouverts. Tout ouvert de \mathbb{R} pouvant être construit comme union dénombrable d'intervalles ouverts, on en déduit que $\sigma(\mathcal{A})$ contient tous les ouverts de \mathbb{R} , et donc est une tribu plus fine que la tribu borélienne sur \mathbb{R} .

On en déduit que si deux lois de probabilités sur \mathbb{R} prennent les mêmes valeurs sur \mathcal{A} , alors elles sont égales. Cela montre bien que la loi de X est caractérisée par $\mathbf{P}(X \in A)$ pour tout $A \in \mathcal{A}$. \square

Cette propriété permet d'introduire une fonction caractérisant la loi d'une variable aléatoire réelle : la fonction de répartition. Avec un léger abus de notation, on parlera également de fonction de répartition d'une variable aléatoire pour parler de la fonction de répartition de sa loi.

Définition 2.2.5. La fonction de répartition de la loi de probabilité μ sur \mathbb{R} est la fonction

$$F : \begin{array}{ll} \mathbb{R} & \longrightarrow [0, 1] \\ x & \longmapsto \mu((-\infty, x]). \end{array}$$

La fonction de répartition de la variable aléatoire X est la fonction

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

Propriété 2.2.6. Une fonction F est la fonction de répartition d'une variable aléatoire réelle si et seulement si

1. F est croissante ;
2. F est continue à droite ;
3. $\lim_{x \rightarrow -\infty} F(x) = 0$ et $\lim_{x \rightarrow \infty} F(x) = 1$.

Démonstration. Soit X une variable aléatoire réelle, on pose $F : x \mapsto \mathbf{P}(X \leq x)$. On observe immédiatement que pour tout $a < b$, on a $\{X \leq a\} \subset \{X \leq b\}$, donc $F(a) \leq F(b)$, ce qui montre que F est croissante. Soit (b_n) une suite décroissante de réels convergeant vers b . Par convergence monotone (proposition 1.3.6), on a

$$\lim_{n \rightarrow \infty} F(b_n) = \lim_{n \rightarrow \infty} \mathbf{P}(X \leq b_n) = \mathbf{P}(\cap_{n \in \mathbb{N}} \{X \leq b_n\}) = \mathbf{P}(X \leq b) = F(b),$$

donc F est continue à droite. En utilisant à nouveau la convergence monotone, on obtient également

$$\lim_{x \rightarrow \infty} F(x) = \mathbf{P}(X \in \mathbb{R}) = 1 \text{ et } \lim_{x \rightarrow -\infty} F(x) = \mathbf{P}(X \in \emptyset) = 0.$$

Afin de montrer la réciproque, c'est-à-dire qu'une fonction croissante continue à droite telle que $\lim_{x \rightarrow -\infty} F(x) = 0$ et $\lim_{x \rightarrow \infty} F(x) = 1$ est une fonction de répartition, on construit une variable aléatoire dont la fonction de répartition est F . Une telle construction est proposée dans l'exercice 2.2. \square

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 $\mu(\{x\}) = F(x) - F(x-)$.*

Deux familles de lois de probabilités sont particulièrement communes.

Définition 2.2.8 (Loi discrète). Une loi de probabilité est dite *discrète* si son support est fini ou dénombrable.

On observera que si la fonction de répartition d'une variable aléatoire est étagée, alors cette variable aléatoire est discrète.

Définition 2.2.9. Une loi de probabilité μ sur \mathbb{R} est dite à *densité* par rapport à la mesure de Lebesgue s'il existe une fonction mesurable φ (appelée la *densité* de μ) telle que

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

pour tout borélien A de \mathbb{R} .

On observe qu'une loi de probabilité sur \mathbb{R} à densité si et seulement si sa fonction de répartition est dérivable presque partout (autrement dit l'ensemble des points où F n'est pas dérivable est un ensemble de mesure de Lebesgue 0) et $\int_{\mathbb{R}} F'(x) dx = 1$. Dans ce cas, la densité de la loi est la fonction $x \mapsto F'(x)$ définie presque partout.

Propriété 2.2.10. *Une fonction mesurable φ est la densité d'une loi de probabilité si elle satisfait les deux propriétés suivantes*

1. $\varphi(x) \geq 0$ pour presque tout x ,
2. $\int_{\mathbb{R}} \varphi(x) dx = 1$.

Remarque 2.2.11. Notons qu'il existe des lois qui ne sont ni des lois discrètes ni des lois à densité.

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 Suites infinies de variables aléatoires

De la même façon qu'un vecteur aléatoire est une variable aléatoire à valeur dans \mathbb{R}^n , une suite de variables aléatoires est une fonction mesurable d'un espace de probabilité à valeurs dans l'ensemble des suites réelles.

Définition 2.4.1. Une suite aléatoire est une famille $X = (X_n, n \in \mathbb{N})$ de variables aléatoires. La loi de la suite aléatoire est la mesure de probabilité sur $(\mathbb{R}^{\mathbb{N}}, \mathcal{B}(\mathbb{R}^{\mathbb{N}}))$ définie par

$$\mu_X(A) = \mathbf{P}((X_n, n \in \mathbb{N}) \in A).$$

Les lois marginales de la suite $(X_n, n \geq 1)$ sont les lois des variables X_1, \dots, X_n .

On observe que la loi d'une suite aléatoire peut être caractérisée par la loi des premiers termes de la suite.

Proposition 2.4.2. *La loi d'une suite de variable aléatoire $(X_n, n \geq 1)$ est caractérisée par la donnée des lois des vecteurs (X_1, \dots, X_k) pour tout $k \in \mathbb{N}$.*

Démonstration. On observe que $\mathcal{C} = \cup_{k \in \mathbb{N}} \{A_1 \times A_2 \times A_k \times \mathbb{R}^{\mathbb{N}}, A_1, \dots, A_k \in \mathcal{B}(\mathbb{R})\}$ est une classe monotone de $\mathcal{B}(\mathbb{R}^{\mathbb{N}})$. Par conséquent, la loi de $(X_n, n \geq 1)$ est caractérisée par la donnée de

$$\mathbf{P}(X_1 \in A_1, \dots, X_k \in A_k) \text{ pour } k \in \mathbb{N} \text{ et } A_1, \dots, A_k \in \mathcal{B}(\mathbb{R}). \quad \square$$

2.5 Lois usuelles

On recense ici les familles de lois usuelles, discrètes et à densité, ainsi que leurs propriétés élémentaires.

2.5.1 Lois discrètes

Si X est une variable aléatoire discrète de loi μ , on appelle espérance de X et on note $\mathbf{E}(X)$ la quantité

$$\mathbf{E}(X) = \sum_{x \in \text{Supp}(\mu)} x\mu(x) = \sum_{x \in \text{Supp}(\mu)} x\mathbf{P}(X = x).$$

On appelle variance de X la quantité $\mathbf{V}\text{ar}(X) = \mathbf{E}(X^2) - \mathbf{E}(X)^2$.

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

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

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.

2.5.2 Lois à densité

Si X est une variable aléatoire à densité de densité φ , on appelle espérance de X et on note $\mathbf{E}(X)$ la quantité

$$\mathbf{E}(X) = \int_{\mathbb{R}} x\varphi(x)dx.$$

On appelle variance de X la quantité $\mathbf{Var}(X) = \mathbf{E}(X^2) - \mathbf{E}(X)^2$.

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

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

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

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

La loi exponentielle est utilisée pour modéliser la durée de vie d'un phénomène sans mémoire ou sans vieillissement. En effet, on observe que X est une variable aléatoire de loi exponentielle si et seulement si $P(X \geq x + y | X \geq x) = P(X \geq y)$ pour tout $x, y \in \mathbb{R}_+$ (c.f. Chapitre 5 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 $\Gamma(a, \theta)$, on a

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

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

$$\varphi_X(x) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1} (1-x)^{b-1} \mathbb{1}_{\{x \in (0,1)\}},$$

$$\mathbf{E}(X) = \frac{a}{a+b} \quad \text{et} \quad \mathbf{Var}(X) = \frac{ab}{(a+b)^2(a+b+1)}.$$

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

Exercices

Exercice 2.1 (Tribu produit et mesurabilité). 1. Soient (E, \mathcal{E}) et (F, \mathcal{F}) deux espaces de probabilités, montrer que $\mathcal{E} \otimes \mathcal{F}$ est la plus petite tribu sur $E \times F$ telle que les projections π_E et π_F de $E \times F$ sur E et F respectivement sont mesurables.

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

Exercice 2.2 (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

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

alors F est la fonction de répartition d'une variable aléatoire.

Exercice 2.3. Soit $\alpha > 0$ et X une variable aléatoire à valeurs dans $[1, \infty[$ telle que pour tout $x > 1$, on a $\mathbf{P}(X \geq x) = x^{-\alpha}$. Montrer que X est une variable aléatoire à densité. La variable aléatoire X admet-elle une espérance? une variance?

Exercice 2.4 (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 2.5 (Le paradoxe de Bertrand). On s'intéresse à la probabilité de l'événement \mathcal{A} : « une corde choisie au hasard sur le cercle unité ait une longueur plus grande que le côté du triangle équilatéral inscrit ».

1. On choisit deux extrémités de la corde au hasard sur le cercle. Montrer que $\mathbf{P}(\mathcal{A}) = 1/3$.
2. On choisit le centre de la corde au hasard sur le disque unité. Déterminer la valeur de $\mathbf{P}(\mathcal{A})$.
3. On choisit au hasard la direction du rayon orthogonal à la corde, puis le centre de la corde uniformément sur ce rayon. Déterminer $\mathbf{P}(\mathcal{A})$.

Exercice 2.6. 1. Soit (X, Y) un point tiré au hasard sur le disque unité \mathbb{R}^2 . Déterminer la loi marginale de X .

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

Espérance d'une variable aléatoire

“Imaginez que je doive payer une certaine somme ; je peux sortir les pièces de mon porte-monnaie comme elles viennent pour arriver à la somme indiquée, ou sortir toutes les pièces et les choisir selon leur valeur. La première méthode est l'intégrale de Riemann, la deuxième correspond à mon intégrale.”

Henri Lebesgue (d'après internet).

Construction de mesures

“Where our language suggests a body and there is none : there, we should like to say, is a spirit.”

Lorsque notre langage suggère un corps et qu’il n’y en a pas :
là, tenons-nous à dire, il y a un esprit.

Ludwig Wittgenstein – Tractatus
Logico-Philosophicus.

Indépendance

“They pretend to themselves they are in control of events where perhaps they are not.”

Ils nourrissent l'illusion de contrôler les événements quand ça n'est peut-être pas le cas.

Anton Chigurh – No Country for Old Men, de
Cormac McCarthy.

Convergence d'une suite de variables aléatoires

“Un homme tirait au sort toutes ses décisions. Il ne lui arriva pas plus de mal qu'aux autres qui réfléchissent.”

Paul Valéry – Tel quel.