

Sorbonne Université — Faculté des sciences
Année universitaire 2019–2020

Introduction aux probabilités (2MA241)

Quentin BERGER et Shen LIN

Table des matières

Préambule	5
1 Préliminaires : ensembles, dénombrement et dénombrabilité	6
1.1 Opérations sur les ensembles	6
1.2 Quelques notions importantes de dénombrement	7
1.3 Dénombrabilité	8
2 Espaces probabilisés $(\Omega, \mathcal{F}, \mathbb{P})$	10
2.1 Axiomes des probabilités et premières propriétés	10
2.1.1 Espaces probabilisés	10
2.1.2 Propriétés fondamentales	12
2.2 Espaces d'états dénombrables	16
2.2.1 Probabilité et densité discrète	16
2.2.2 Équiprobabilité : quelques exemples	17
2.3 Probabilité conditionnelle et indépendance	20
2.3.1 Probabilité conditionnelle	20
2.3.2 Indépendance d'événements	22
3 Variables aléatoires discrètes	25
3.1 Variables aléatoires discrètes : premières définitions et propriétés	25
3.2 Indépendance de variables aléatoires	29
3.2.1 Loi jointe et lois marginales	29
3.2.2 Indépendance de variables aléatoires	30
3.3 Exemples essentiels de variables aléatoires discrètes	33
3.3.1 Loi uniforme discrète	33
3.3.2 Loi de Bernoulli	34
3.3.3 Loi binomiale	34
3.3.4 Loi de Poisson	34
3.3.5 Loi géométrique	35
3.4 Espérance, variance et moments	39
3.4.1 Définition de l'espérance	39
3.4.2 Propriétés de l'espérance	40
3.4.3 Moments, variance et covariance	43
3.4.4 Espérance et indépendance	45
3.4.5 Espérance et variance des lois classiques	46
3.5 Inégalités probabilistes et applications	47
3.6 Fonction génératrice et applications	50

4	Intermédiaire : variables aléatoires réelles générales	55
4.1	Définition et propriétés	55
4.2	Fonction de répartition	57
4.3	Fonction génératrice des moments	59
5	Variables aléatoires à densité	61
5.1	Variables aléatoires absolument continues	61
5.1.1	Définition et premières propriétés	61
5.1.2	Calculer une densité	62
5.1.3	Espérance et formule de transfert	63
5.2	Exemples essentiels de variables aléatoires absolument continues	66
5.2.1	Loi uniforme continue	66
5.2.2	Loi exponentielle	67
5.2.3	Loi normale (gaussienne)	68
5.3	Vecteurs aléatoires absolument continus	70
5.3.1	Définition et premières propriétés	70
5.3.2	Densité jointe, densités marginales, indépendance	71
5.3.3	Formule de transfert et calculs avec les densités	73
5.4	Quelques exemples supplémentaires...	75
6	Simulation informatique de variables aléatoires	78
6.1	Simulation de lois discrètes	78
6.2	Méthode d'inversion et simulation de lois absolument continues	81
6.3	Le cas de la loi normale	83
6.4	Méthode de rejet : cas d'une densité bornée et à support borné	83
7	Loi des grands nombres et convergences de suites de variables aléatoires	88
7.1	La loi (faible) des grands nombres et applications	88
7.2	Une application : la méthode de Monte-Carlo	92
7.3	Convergence en probabilité pour une suite de variables aléatoires	94
8	Introduction à la statistique	97
8.1	Modèles statistiques paramétriques	97
8.2	Inégalités de concentration et construction d'intervalles de confiance	99
8.2.1	Inégalités de concentration et intervalles de confiance	99
8.2.2	Intervalles de confiance pour la méthode de Monte-Carlo	101
	Variables aléatoires importantes : lois, espérance, variance	103
	Index	104

Préambule

Une expérience est dite *aléatoire* si, “reproduite plusieurs fois dans des conditions identiques”, le résultat de l’expérience change d’une fois sur l’autre, de sorte qu’il ne peut être prédit. Des exemples classiques que l’on souhaiterait étudier sont : le résultat du lancer d’une pièce ou d’un dé, le tirage du loto, le fait de tirer un nombre au hasard. Mais on peut aussi penser à des expériences plus complexes, comme un marcheur qui se déplacerait de manière “aléatoire” dans \mathbb{R}^d , comme dans la figure ci-dessous.

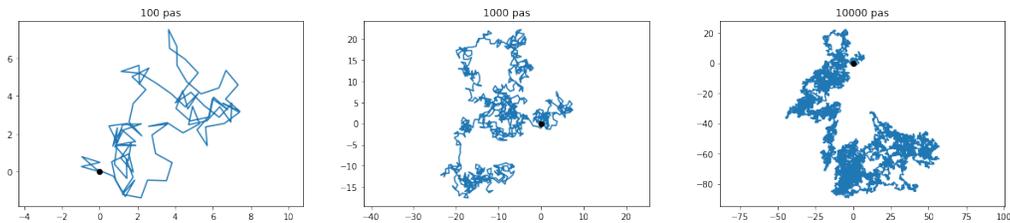


FIGURE 1 – Trajectoire d’un “marcheur aléatoire” dans \mathbb{R}^2 , après $n = 100$, $n = 1000$ et $n = 10000$ pas.

Le but de ce cours est de donner une introduction à la notion mathématique d’expérience aléatoire et de probabilité, et de présenter des concepts et outils mathématiques utilisés dans la modélisation de phénomènes aléatoires, sans avoir recours à la théorie de la mesure. On étudiera plus en détails certains modèles aléatoires, comme par exemple la marche aléatoire...

Les exemples précédés de ♠ et les sections précédées de ★ sont des exemples et problèmes importants en théorie des probabilités : ils sont considérés comme *optionnels* et pourront ne pas être étudiés, mais ils présentent un aperçu de certains modèles probabilistes, et permettent d’illustrer la puissance des différents outils introduits dans ce cours.

Quelques références

Donnons ici quelques références de livres traitant la théorie des probabilités à un niveau proche (mais souvent plus élevé) de ce cours :

- Sheldon M. Ross, *Initiation aux probabilités*, PPUR presses polytechniques.
- Pierre Brémaud, *Initiation aux Probabilités – et aux chaînes de Markov*, Springer.
- Walter Appel, *Probabilités pour les non-probabilistes*, H&K.
- Jean-Yves Oувrard, *Probabilités, Tome 1*, Cassini.

Chapitre 1

Préliminaires : ensembles, dénombrement et dénombrabilité

Nous rappelons dans ce chapitre quelques notions et notations élémentaires de théorie des ensembles. On notera $\mathbb{N} = \{0, 1, 2, \dots\}$ l'ensemble des entiers naturels, et $\mathbb{N}_* = \{1, 2, \dots\}$ l'ensemble des entiers naturels non nuls. Si Ω est un ensemble, on notera $\mathcal{P}(\Omega) = \{A, A \subset \Omega\}$ l'ensemble des parties (c'est-à-dire des sous-ensembles) de Ω .

1.1 Opérations sur les ensembles

Soit Ω un ensemble et soient A, B deux sous-ensembles de Ω . Rappelons les notations standard

$$A \cup B := \{\omega \in \Omega : \omega \in A \text{ ou } \omega \in B\}, \quad A \cap B := \{\omega \in \Omega : \omega \in A \text{ et } \omega \in B\},$$

$$A^c := \{\omega \in \Omega : \omega \notin A\}, \quad A \setminus B := A \cap B^c$$

$$A \Delta B := (A \setminus B) \cup (B \setminus A) = (A \cup B) \setminus (A \cap B).$$

Les notions d'union et d'intersection s'étendent de manière naturelle à une famille arbitraire $(A_i)_{i \in I}$ de sous-ensembles de Ω :

$$\bigcup_{i \in I} A_i := \{\omega \in \Omega : \exists i \in I \text{ tel que } \omega \in A_i\}, \quad \bigcap_{i \in I} A_i := \{\omega \in \Omega : \forall i \in I \text{ on a } \omega \in A_i\}.$$

On dira qu'une union $\bigcup_{i \in I} A_i$ est *disjointe* si les ensembles $(A_i)_{i \in I}$ sont *deux à deux disjoints*, c'est-à-dire si $A_i \cap A_j = \emptyset$ pour tous $i \neq j$.

Rappelons les relations de distributivité et de passage au complémentaire, qui sont valables de manière générale pour des familles arbitraires d'ensembles :

$$\begin{aligned} \left(\bigcup_{i \in I} A_i \right) \cap C &= \bigcup_{i \in I} (A_i \cap C), & \left(\bigcap_{i \in I} A_i \right) \cup C &= \bigcap_{i \in I} (A_i \cup C), \\ \left(\bigcup_{i \in I} A_i \right)^c &= \bigcap_{i \in I} A_i^c, & \left(\bigcap_{i \in I} A_i \right)^c &= \bigcup_{i \in I} A_i^c. \end{aligned}$$

Le produit cartésien $A \times B$ de deux ensembles A et B est l'ensemble des couples ordonnés (a, b) avec $a \in A$ et $b \in B$. On définit de manière analogue le produit de plus de deux ensembles. On notera A^n le produit $A \times \cdots \times A$ (n fois).

Pour tout sous-ensemble $A \subseteq \Omega$ on définit la fonction $\mathbb{1}_A : \Omega \rightarrow \mathbb{R}$, appelée *indicatrice de A* , qui vaut 1 sur A et 0 sur A^c :

$$\mathbb{1}_A(x) := 1 \text{ si } x \in A, \quad \mathbb{1}_A(x) := 0 \text{ si } x \notin A. \quad (1.1)$$

Dans d'autres contextes, cette fonction est appelée caractéristique, et notée χ_A (mais en probabilités, la fonction caractéristique est un autre objet...).

1.2 Quelques notions importantes de dénombrement

Pour un ensemble *fini* A , on notera $|A|$ sa cardinalité, c'est-à-dire le nombre d'éléments qu'il contient. On écrira $|A| < \infty$ pour indiquer que l'ensemble est fini, et $|A| = \infty$ pour indiquer qu'il est infini.

Rappelons quelques principes de base du dénombrement :

- si A_1, \dots, A_n sont des ensembles finis *deux à deux disjoints*, alors $A_1 \cup \cdots \cup A_n$ est fini et de cardinal $\sum_{i=1}^n |A_i|$;
- Si A_1, \dots, A_n sont des ensembles finis, alors $A_1 \times \cdots \times A_n$ est fini, et de cardinal $|A_1| \times \cdots \times |A_n|$.

Pour A, B deux ensembles, on rappelle qu'une application $f : A \rightarrow B$ est dite :

- *injective* si tout élément de B est l'image d'au plus un élément de A ;
- *surjective* si tout élément de B est l'image d'au moins un élément de A ;
- *bijective* si elle est à la fois injective et surjective.

S'il existe une injection de A dans B alors on a $|A| \leq |B|$. S'il existe une surjection de A sur B alors on a $|A| \geq |B|$. S'il existe une bijection entre A et B alors on a $|A| = |B|$.

Exercice 1.1. Montrer que si Ω est un ensemble, la fonction $\varphi : \mathcal{P}(\Omega) \rightarrow \{0, 1\}^\Omega$ qui à un ensemble $A \subset \Omega$ associe la fonction indicatrice $\mathbb{1}_A$ est une bijection. En déduire que si Ω est fini, on a $|\mathcal{P}(\Omega)| = 2^{|\Omega|}$.

Nombre d'arrangements. Si A est un ensemble fini de cardinal n , on appelle *arrangement (sans répétition)* de k éléments de A , un k -uplet (ordonné) (a_1, \dots, a_k) d'éléments distincts de A . Le nombre d'arrangements de k éléments de A est égal à $n(n-1) \cdots (n-k+1)$ si $0 \leq k \leq n$ et à 0 si $k > n$. Il s'agit aussi du nombre d'applications injectives d'un ensemble à k éléments dans un ensemble à n éléments.

Nombre de permutations d'un ensemble. Si A est un ensemble de cardinal n , on appelle *permutation* de A une application bijective de A dans A . Le nombre de *permutations* de A est égal à $n! = n(n-1) \cdots 2 \cdot 1$, où par convention $0! = 1$. On notera \mathfrak{S}_n l'ensemble des permutations de $\{1, \dots, n\}$.

Nombre de combinaisons. Si A est un ensemble fini de cardinal n , on appelle *combinaison* de k éléments de A un sous-ensemble $\{a_1, \dots, a_k\}$ d'exactly k éléments de A . Le nombre de combinaisons de k éléments de A est égal à

$$\binom{n}{k} := \frac{n(n-1) \cdots (n-k+1)}{k!} = \frac{n!}{k!(n-k)!},$$

où par convention $\binom{n}{k} = 0$ si $k > n$ ou $n < 0$, et $\binom{n}{0} = 1$. Notons que le nombre d'arrangements de k éléments est égal à $k!$ fois le nombre de combinaisons : en effet, pour chaque sous-ensemble $\{a_1, \dots, a_k\}$, il y a $k!$ manière d'ordonner ses éléments, et donc $k!$ arrangements correspondants.

Le nombre $\binom{n}{k}$ est appelé coefficient binomial. Rappelons quelques-unes de ses propriétés :

- $\binom{n}{k} = \binom{n}{n-k}$ pour tout $0 \leq k \leq n$;
- $\binom{n}{k} + \binom{n}{k+1} = \binom{n+1}{k+1}$ pour tout $0 \leq k \leq n$ (triangle de Pascal) ;
- $(x + y)^n = \sum_{k=0}^n \binom{n}{k} x^k y^{n-k}$ pour tout $x, y \in \mathbb{R}$ (binôme de Newton).

1.3 Dénombrabilité

Définition 1.2. *Un ensemble est dit dénombrable s'il est fini, ou en bijection avec l'ensemble \mathbb{N} des entiers naturels.*

Concrètement, un ensemble est dénombrable si on peut énumérer ses éléments, c'est-à-dire écrire ses éléments comme une suite finie x_1, \dots, x_N ou infinie $(x_i)_{i \geq 1}$. Tout sous-ensemble $A \subset \mathbb{N}$ est dénombrable, et un ensemble en bijection avec une partie de \mathbb{N} est dénombrable*. Le résultat suivant permet de déterminer si des ensembles sont dénombrables.

Théorème 1.3. *i) Si A_1, \dots, A_n sont des ensembles dénombrables, le produit $A_1 \times \dots \times A_n$ est également dénombrable.*
ii) Si $(A_i)_{i \in I}$ est une famille dénombrable (c'est-à-dire si I est dénombrable) d'ensembles finis ou dénombrables (c'est-à-dire A_i est dénombrable pour tout $i \in I$) alors l'union $\bigcup_{i \in I} A_i$ est également dénombrable.

Démonstration. La démonstration est basée sur le lemme suivant.

Lemme 1.4. *Soit A et B deux ensembles. S'il existe une injection de A dans B et si B est dénombrable alors A est dénombrable. S'il existe une surjection de A dans B et si A est dénombrable, alors B est dénombrable.*

La première partie de ce lemme repose sur le fait qu'une injection $f : A \rightarrow B$ induit une bijection entre A et $f(A) = \{f(a), a \in A\}$: si B est dénombrable, alors $f(A) \subset B$ est dénombrable, et donc A aussi. De même, à partir d'une surjection $f : A \rightarrow B$ on peut construire une bijection entre un sous-ensemble A' de A et B (pour tout $b \in B$, on choisit un antécédent a_b de b par f , et on considère $A' = \{a_b, b \in B\}$) : si A est dénombrable, alors $A' \subset A$ est dénombrable, et donc B aussi.

i) On peut supposer que $A_1 = \dots = A_n = \mathbb{N}_*$. Notons $p_1 = 2, p_2 = 3, \dots, p_n$ les n premiers nombres premiers. L'application qui à $(i_1, \dots, i_n) \in \mathbb{N}_*^n$ associe le nombre $p_1^{i_1} \cdot p_2^{i_2} \cdot \dots \cdot p_n^{i_n}$ est une injection de \mathbb{N}_*^n dans \mathbb{N}_* , par unicité de la décomposition en facteurs premiers d'un nombre donné. Le Lemme 1.4 permet de conclure.

ii) On peut supposer que $I = \mathbb{N}_*$. L'application de $\mathbb{N}_* \times \mathbb{N}_*$ dans $\bigcup_{i \in I} A_i$ qui au couple (n, m) associe le m -ième élément de l'ensemble A_n est une surjection. Le Lemme 1.4 permet de conclure. \square

*. Attention, on trouve parfois dans la littérature des définitions où dénombrable signifie *infini dénombrable*.

Corollaire 1.5. *L'ensemble \mathbb{Z} des entiers relatifs et l'ensemble \mathbb{Q} des nombres rationnels sont (infinis) dénombrables.*

Démonstration. L'ensemble \mathbb{Z} est dénombrable car l'application de l'ensemble dénombrable $\{1, -1\} \times \mathbb{N}$ dans \mathbb{Z} qui au couple (ε, n) associe le produit εn est une surjection. De même, \mathbb{Q} est dénombrable car l'application de l'ensemble dénombrable $\mathbb{Z} \times \mathbb{N}_*$ dans \mathbb{Q} qui au couple (p, q) associe le rationnel p/q est une surjection. \square

Exercice 1.6. Montrer que l'ensemble $\mathbb{Z}[X]$ des polynômes à coefficients entiers est dénombrable. En déduire que l'ensemble \mathcal{A} des nombres algébriques est dénombrable. (Un nombre est algébrique s'il est racine d'un polynôme à coefficients entiers.)

Théorème 1.7. *L'ensemble des nombres réels \mathbb{R} n'est pas dénombrable.*

Démonstration. Il suffit de montrer qu'un sous-ensemble de \mathbb{R} n'est pas dénombrable (pourquoi?). Montrons que l'ensemble $\mathcal{E} := \{x \in [0, 1] : \text{l'écriture décimale de } x \text{ ne comporte que des 1 et des 2}\}$ n'est pas dénombrable (notons que l'écriture décimale d'un élément de \mathcal{E} est unique).

Raisonnons par l'absurde : supposons que \mathcal{E} soit dénombrable. Il existe alors une bijection $\varphi : \mathbb{N}_* \rightarrow \mathcal{E}$. On va construire un nombre qui n'est pas dans l'image de φ , grâce au procédé diagonal de Cantor. Posons $x_i = \varphi(i)$, et notons $x_i^{(k)}$ le k -ème chiffre dans son écriture décimale $x_i = 0, x_i^{(1)} x_i^{(2)} x_i^{(3)} \dots$. Alors, on construit un nombre y à travers son écriture décimale : on pose $y = 0, y_1 y_2 \dots$, où $y_i = 2$ si $x_i^{(i)} = 1$ et $y_i = 1$ si $x_i^{(i)} = 2$. Ainsi, l'écriture décimale de y ne contient que des 1 et des 2 : on a bien $y \in \mathcal{E}$. Plus important, on a $y_i \neq x_i^{(i)}$ pour tout i . Cela signifie que y ne peut pas être dans l'image de φ : s'il existe un $j \in \mathbb{N}_*$ tel que $y = \varphi(j)$, on aurait $y = x_j$ et donc $y_j = x_j^{(j)}$ ce qui contredit la construction de y . \square

Notons que $\mathcal{P}(\mathbb{N})$ est en bijection avec $\{0, 1\}^{\mathbb{N}}$ (voir l'Exercice 1.1). De plus, l'application qui à un réel $x \in [0, 1[$ associe son développement en base 2 est une injection de $[0, 1[$ dans $\{0, 1\}^{\mathbb{N}}$: on en déduit donc que $\{0, 1\}^{\mathbb{N}}$ et $\mathcal{P}(\mathbb{N})$ ne sont pas dénombrables.

Somme indexée par un ensemble dénombrable

Si A est un ensemble dénombrable et $\{x_a\}_{a \in A}$ est une famille de réels indexée par A , alors on peut énumérer ses éléments $\{x_1, x_2, \dots\}$, et considérer la somme $\sum_{n \geq 1} x_n$. Cependant, pour que la somme $\sum_{a \in A} x_a$ soit bien définie, il faut qu'elle ne dépende pas de l'ordre dans lequel on a énuméré ses éléments.

- Dans le cas où la famille $\{x_a\}_{a \in A}$ est à termes positifs, on peut montrer que la somme ne dépend pas de l'ordre de sommation, avec la définition

$$\sum_{a \in A} x_a = \sup \left\{ \sum_{a \in A'} x_a : A' \text{ sous-ensemble fini de } A \right\} \in [0, +\infty].$$

- De manière générale, la somme $\sum_{a \in A} x_a$ est bien définie si au moins une des deux sommes $\sum_{a \in A} x_a^+$ et $\sum_{a \in A} x_a^-$ est finie (où $x^+ = \max\{x, 0\}$ et $x^- = \max\{-x, 0\}$ désignent les parties positives et négatives de x) et on pose alors

$$\sum_{a \in A} x_a = \sum_{a \in A} x_a^+ - \sum_{a \in A} x_a^- \in [-\infty, +\infty].$$

Chapitre 2

Espaces probabilisés $(\Omega, \mathcal{F}, \mathbb{P})$

2.1 Axiomes des probabilités et premières propriétés

2.1.1 Espaces probabilisés

Notre objectif est de fournir une description mathématique d'une expérience aléatoire.

Exemple 2.1. Voici trois exemples d'expériences aléatoires :

1. on lance un dé ordinaire à 6 faces ;
2. on lance une pièce équilibrée jusqu'à ce que l'on obtienne Pile ;
3. on effectue une suite *infinie* de Pile ou Face.

L'espace d'état, Ω . La première étape consiste à identifier un ensemble Ω , appelé *espace d'état* ou *univers*, qui contient tous les résultats possibles de l'expérience :

- pour le lancer d'un dé à 6 faces, on considérera l'espace d'état $\Omega = \{1, 2, 3, 4, 5, 6\}$;
- pour le lancer d'une pièce équilibrée jusqu'à l'obtention de Pile, on pourra identifier le résultat de l'expérience au nombre de lancers effectués : on considérera l'espace d'état $\Omega = \mathbb{N}_*$;
- pour une suite *infinie* de Pile ou Face, on considérera l'espace d'état $\Omega = \{0, 1\}^{\mathbb{N}^*}$: une suite $\omega = (\omega_1, \omega_2, \dots)$ correspondra à une suite de Pile ou Face, où $\omega_n = 1$ si le n -ème lancer est Pile et $\omega_n = 0$ si le n -ème lancer est Face.

L'ensemble des événements, \mathcal{F} . La deuxième étape consiste à identifier un ensemble \mathcal{F} d'*événements*, qui sont de manière informelle des affirmations sur le résultat de l'expérience aléatoire. Les événements sont décrits par des *sous-ensembles de l'espace d'état* Ω ; pour une affirmation donnée on peut considérer l'ensemble des résultats de l'expérience pour lesquels cette affirmation est vérifiée. Dans les exemples précédents, on peut considérer les événements suivants :

- $A =$ "le résultat du dé est un nombre pair" $= \{2, 4, 6\}$
- $B =$ "on n'obtient pas Pile lors des 3 premiers lancers" $= \{4, 5, \dots\}$
- $C =$ "les 3 premiers lancers sont Face" $= \{\omega \in \{0, 1\}^{\mathbb{N}^*} : \omega_1 = \omega_2 = \omega_3 = 0\}$.

Quand aucun des résultats possibles de l'expérience ne vérifie l'affirmation, l'événement est dit impossible, et identifié à l'ensemble vide \emptyset . À l'opposé, une affirmation qui est vérifiée quel que soit le résultat de l'expérience correspond à l'événement certain, donné par l'ensemble Ω tout entier.

Notons que les opérations logiques de *conjonction*, de *disjonction* et de *négation* des affirmations correspondent à l'*intersection*, à l'*union* et au *complémentaire* des ensembles :

$$\begin{aligned} \text{“}A \text{ et } B \text{ sont tous les deux vérifiés”} &\longrightarrow A \cap B, \\ \text{“au moins un parmi } A \text{ ou } B \text{ est vérifié”} &\longrightarrow A \cup B, \\ \text{“}A \text{ n'est pas vérifié”} &\longrightarrow A^c. \end{aligned}$$

Une question délicate est de savoir s'il faut considérer que *tous* les sous-ensembles d'un espace d'état Ω sont des événements, c'est-à-dire si on doit prendre $\mathcal{F} = \mathcal{P}(\Omega)$. On peut le faire quand Ω est dénombrable, mais si Ω est infini non dénombrable (comme dans le troisième exemple), il peut s'avérer nécessaire de ne considérer qu'une classe restreinte de sous-ensembles de Ω , que l'on peut interpréter comme les événements “descriptibles”. De manière générale, on considérera un ensemble d'événements $\mathcal{F} \subset \mathcal{P}(\Omega)$ *non vide*, qui vérifie les propriétés suivantes.

Définition 2.2 (Tribu). *On dit que \mathcal{F} est une tribu sur Ω (ou une σ -algèbre), si \mathcal{F} est une partie non vide de $\mathcal{P}(\Omega)$ vérifiant :*

- (i) $\Omega \in \mathcal{F}$;
- (ii) si $A \in \mathcal{F}$ alors $A^c \in \mathcal{F}$;
- (iii) si $(A_i)_{i \in \mathbb{N}}$ est une suite d'événements de \mathcal{F} , alors $\bigcup_{i \in \mathbb{N}} A_i \in \mathcal{F}$.

Notons que l'on a alors $\emptyset = \Omega^c \in \mathcal{F}$ et aussi $\bigcap_{i \in \mathbb{N}} A_i = (\bigcup_{i \in \mathbb{N}} A_i^c)^c \in \mathcal{F}$: l'ensemble \mathcal{F} est donc stable par passage au complémentaire, et par union ou intersection dénombrable.

En pratique, si Ω est dénombrable, on choisira $\mathcal{F} = \mathcal{P}(\Omega)$. Si Ω est infini non dénombrable (comme dans le troisième exemple un peu plus haut), alors on pourra considérer des événements “de base”, à partir desquels on construit l'ensemble \mathcal{F} (on parle de tribu engendrée). Par exemple, si $\Omega = \{0, 1\}^{\mathbb{N}^*}$ correspond à une suite *infinie* de lancers de Pile ou Face, alors on a envie de considérer les ensembles de base $B_i = \{\omega, \omega_i = 1\}$, correspondant à l'affirmation “le i -ème lancer est Pile”. L'ensemble \mathcal{F} doit alors contenir tous les événements que l'on peut construire avec les B_i par passage au complémentaire, union et intersection dénombrables (et il existe un moyen canonique de le faire) : on doit ainsi pouvoir construire des événements plus “compliqués”, comme l'événement $B = \bigcup_{n \geq 1} \bigcap_{i \geq n} B_i$ = “il n'y a que des Pile à partir d'un certain rang”, ou B^c = “il y a une infinité de Face”.

La probabilité, \mathbb{P} . Le dernier ingrédient est la donnée d'une (*mesure de*) *probabilité* \mathbb{P} . D'un point de vue mathématique, une probabilité est décrite par une application \mathbb{P} qui à tout événement $A \in \mathcal{F}$ associe un nombre $\mathbb{P}(A) \in [0, 1]$.

Si l'expérience aléatoire pouvait être répétée un très grand nombre N de fois, on pourrait en principe compter le nombre de fois $N_A \in \{0, \dots, N\}$ où l'événement A est vérifié. Dans ce cas, il est naturel d'interpréter $\mathbb{P}(A)$ comme la proportion de fois où l'événement A est vérifié, à savoir $\mathbb{P}(A) \approx N_A/N$. Il s'agit de l'*interprétation fréquentiste* des probabilités (qui est justifiée a posteriori par la *loi des grands nombres*, un théorème que nous verrons dans le Chapitre 7).

Dans les faits, le choix de la probabilité est un point délicat. Dans certains cas, il existe un choix “naturel”, issu de considérations sur la nature de l’expérience aléatoire considérée, notamment de symétries. Mais dans tous les cas, quelle que soit la manière dont elle est choisie, une probabilité \mathbb{P} devra satisfaire certaines propriétés, par exemple :

$$\mathbb{P}(\Omega) = 1, \quad \mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) \quad \text{si } A \cap B = \emptyset. \quad (2.1)$$

Ces relations sont tout à fait naturelles si l’on pense à l’interprétation fréquentiste $\mathbb{P}(A) \simeq N_A/N$, car d’une part $N_\Omega = N$ (Ω contient *tous* les résultats possibles de l’expérience), et d’autre part $N_{A \cup B} = N_A + N_B$ si $A \cap B = \emptyset$. Le point de vue moderne consiste à appeler probabilité toute fonction \mathbb{P} qui satisfait une version renforcée des propriétés (2.1).

Définition 2.3 (Axiomes des probabilités). *Soit Ω un ensemble (non vide) muni d’un ensemble d’événements (une tribu) \mathcal{F} vérifiant la Définition 2.2. Une fonction $\mathbb{P} : \mathcal{F} \rightarrow [0, 1]$ est appelée probabilité si elle vérifie les propriétés suivantes :*

A1. $\mathbb{P}(\Omega) = 1$.

A2. (σ -additivité) *Pour toute suite $(A_n)_{n \geq 1}$ d’événements de \mathcal{F} disjoints (c’est-à-dire $A_n \cap A_m = \emptyset$ pour $n \neq m$) on a*

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

Conclusion. Le triplet $(\Omega, \mathcal{F}, \mathbb{P})$ est appelé *espace probabilisé*. L’interprétation d’un espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$ est donc la suivante : l’ensemble Ω contient tous les résultats possibles de l’expérience aléatoire, et pour tout événement $A \in \mathcal{F}$, le réel $\mathbb{P}(A) \in [0, 1]$ exprime le “degré de confiance” que l’on attribue à la possibilité que le résultat de l’expérience soit un élément de A . La propriété A1 exprime le fait que l’espace d’état entier est un événement *certain*, alors que la propriété A2 est une extension de (2.1).

Notons dans un premier temps que la Définition 2.3 implique l’*additivité finie* : si A_1, A_2, \dots, A_k est une famille finie d’événements disjoints (c’est-à-dire $A_i \cap A_j = \emptyset$ pour $i \neq j$), alors

$$\mathbb{P}(A_1 \cup A_2 \cup \dots \cup A_k) = \sum_{j=1}^k \mathbb{P}(A_j). \quad (2.2)$$

Il suffit en effet de prolonger la famille d’événements disjoints A_1, A_2, \dots, A_k à une suite infinie d’événements disjoints, en posant $A_n := \emptyset$ pour $n > k$. Alors, on a $\bigcup_{j=1}^k A_j = \bigcup_{j=1}^{+\infty} A_j$, et on obtient la conclusion grâce à l’axiome A2.

Remarque 2.4. Si Ω est infini, alors les axiomes $\{A1, A2\}$ sont strictement plus forts que $\{A1, (2.2)\}$, dans le sens où il existe des fonctions $\mathbb{P} : \Omega \rightarrow [0, 1]$ qui satisfont (2.2) mais pas A2. Ainsi, la σ -additivité *n’est pas* une conséquence de l’additivité finie.

2.1.2 Propriétés fondamentales

Commençons par l’énoncé de quelques conséquences simples, mais importantes, des axiomes A1 et A2.

Proposition 2.5. Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé.

1. Pour tous événements $A, B \in \mathcal{F}$ tels que $A \subseteq B$

$$\mathbb{P}(B \setminus A) = \mathbb{P}(B) - \mathbb{P}(A),$$

et par conséquent $\mathbb{P}(A) \leq \mathbb{P}(B)$. Pour tout $A \in \mathcal{F}$, $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$.

2. Pour tous événements $A, B \in \mathcal{F}$

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

et par conséquent $\mathbb{P}(A \cup B) \leq \mathbb{P}(A) + \mathbb{P}(B)$.

Démonstration. Commençons par le point 1. Si $A \subseteq B$, on peut écrire $B = A \cup (B \setminus A)$ et l'union est disjointe : par additivité de \mathbb{P} , on a

$$\mathbb{P}(B) = \mathbb{P}(A \cup (B \setminus A)) = \mathbb{P}(A) + \mathbb{P}(B \setminus A),$$

c'est-à-dire $\mathbb{P}(B \setminus A) = \mathbb{P}(B) - \mathbb{P}(A)$. Étant donné que la probabilité d'un événement quelconque est positive, il s'ensuit que $\mathbb{P}(A) \leq \mathbb{P}(B)$.

Venons-en au point 2. Pour tout choix de $A, B \in \mathcal{F}$, on a $B \setminus A = B \setminus (A \cap B)$. Étant donné que $(A \cap B) \subseteq B$, on obtient $\mathbb{P}(B \setminus A) = \mathbb{P}(B) - \mathbb{P}(A \cap B)$ grâce au point précédent. Enfin, en écrivant $A \cup B = A \cup (B \setminus A)$, l'union étant disjointe on obtient par additivité de \mathbb{P}

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

ce qui conclut la démonstration. \square

Exemple 2.6. La formule $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$ s'avère souvent très utile. À titre d'exemple, calculons la probabilité d'obtenir au moins un 6 en lançant 8 dés. Un choix naturel pour l'espace d'état est $\Omega = \{1, 2, 3, 4, 5, 6\}^8$, où $\omega = (\omega_1, \dots, \omega_8)$ avec $\omega_i \in \{1, 2, 3, 4, 5, 6\}$ représente les 8 tirages. L'événement qui nous intéresse est donné par le sous-ensemble

$$A := \{\omega = (\omega_1, \dots, \omega_8) \in \Omega : \omega_i = 6 \text{ pour au moins un } 1 \leq i \leq 8\}.$$

Si les dés sont réguliers, il est naturel de munir Ω de la probabilité uniforme \mathbb{P} , à savoir $\mathbb{P}(A) = |A|/|\Omega|$. Il est clair que $|\Omega| = 6^8$, mais il n'est pas évident de savoir comment calculer la cardinalité de A . Le problème se résout facilement en remarquant que

$$A^c = \{\omega = (\omega_1, \dots, \omega_8) \in \Omega : \omega_i \neq 6 \text{ pour tout } 1 \leq i \leq 8\} = \{1, 2, 3, 4, 5\}^8,$$

qui nous donne $|A^c| = 5^8$. Ainsi, on obtient $\mathbb{P}(A) = 1 - \mathbb{P}(A^c) = 1 - \frac{|A^c|}{|\Omega|} = 1 - \left(\frac{5}{6}\right)^8 \approx 0,77$.

L'égalité de la partie 2 de la Proposition 2.5 peut être généralisée à l'union de plus de deux événements. Par exemple, on peut chercher à calculer $\mathbb{P}(A \cup B \cup C)$ pour trois événements A, B, C . En utilisant deux fois l'égalité, on obtient

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

À partir de là, il n'est pas difficile, de "deviner" la formule générale pour l'union d'un nombre fini (arbitraire) d'événements non nécessairement disjoints. Le résultat suivant est appelé *formule d'inclusion-exclusion*, ou *formule du crible*.

Proposition 2.7 (Formule d'inclusion-exclusion). *Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé. Quel que soit le choix d'événements $A_1, A_2, \dots, A_n \in \mathcal{F}$, on a*

$$\mathbb{P}(A_1 \cup A_2 \cup \dots \cup A_n) = \sum_{k=1}^n \sum_{\substack{J \subseteq \{1,2,\dots,n\} \\ \text{tel que } |J|=k}} (-1)^{k+1} \mathbb{P}\left(\bigcap_{i \in J} A_i\right). \quad (2.3)$$

Démonstration. Une démonstration plus simple sera donnée plus tard, voir l'Exemple 3.36, mais nous en donnons une autre maintenant. Montrons par récurrence sur n que la relation (2.3) est vraie pour tout n -uplet d'événements A_1, A_2, \dots, A_n . Pour $n = 1$ la formule (2.3) est réduite à $\mathbb{P}(A_1) = \mathbb{P}(A_1)$ et il n'y a donc rien à démontrer, et pour $n = 2$ la formule est $\mathbb{P}(A_1 \cup A_2) = \mathbb{P}(A_1) + \mathbb{P}(A_2) - \mathbb{P}(A_1 \cap A_2)$, ce qui a déjà été vu dans la Proposition 2.5. Soit $n \geq 2$, supposons que l'assertion soit vraie pour tout $k \leq n$, et montrons qu'elle est vraie pour $n + 1$. Soient $A_1, A_2, \dots, A_n, A_{n+1}$ des événements. Par hypothèse de récurrence, (2.3) est vraie pour $n = 2$, donc

$$\begin{aligned} \mathbb{P}(A_1 \cup A_2 \cup \dots \cup A_n \cup A_{n+1}) &= \mathbb{P}(A_1 \cup A_2 \cup \dots \cup A_n) + \mathbb{P}(A_{n+1}) - \mathbb{P}((A_1 \cup A_2 \cup \dots \cup A_n) \cap A_{n+1}) \\ &= \mathbb{P}(A_1 \cup A_2 \cup \dots \cup A_n) + \mathbb{P}(A_{n+1}) - \mathbb{P}(B_1 \cup B_2 \cup \dots \cup B_n), \end{aligned} \quad (2.4)$$

où nous avons posé $B_i = A_i \cap A_{n+1}$ pour $i = 1, 2, \dots, n$ pour alléger les notations. En utilisant de nouveau l'hypothèse de récurrence, cette fois pour n événements, on obtient

$$\begin{aligned} \mathbb{P}(A_1 \cup A_2 \cup \dots \cup A_n) &= \sum_{k=1}^n \sum_{\substack{J \subseteq \{1,2,\dots,n\} \\ \text{tel que } |J|=k}} (-1)^{k+1} \mathbb{P}\left(\bigcap_{i \in J} A_i\right) \\ &= \sum_{k=1}^n \sum_{\substack{J \subseteq \{1,2,\dots,n+1\} \\ \text{tel que } |J|=k \text{ et } n+1 \notin J}} (-1)^{k+1} \mathbb{P}\left(\bigcap_{i \in J} A_i\right), \end{aligned} \quad (2.5)$$

et de manière analogue

$$\begin{aligned} \mathbb{P}(B_1 \cup B_2 \cup \dots \cup B_n) &= \sum_{k=1}^n \sum_{\substack{J \subseteq \{1,2,\dots,n\} \\ \text{tel que } |J|=k}} (-1)^{k+1} \mathbb{P}\left(\bigcap_{i \in J} B_i\right) \\ &= \sum_{k=1}^n \sum_{\substack{J \subseteq \{1,2,\dots,n\} \\ \text{tel que } |J|=k}} (-1)^{k+1} \mathbb{P}\left(A_{n+1} \cap \left(\bigcap_{i \in J} A_i\right)\right) \\ &= - \sum_{k=2}^{n+1} \sum_{\substack{J \subseteq \{1,2,\dots,n+1\} \\ \text{tel que } |J|=k \text{ et } n+1 \in J}} (-1)^{k+1} \mathbb{P}\left(\bigcap_{i \in J} A_i\right). \end{aligned} \quad (2.6)$$

En substituant (2.5) et (2.6) dans la dernière ligne de (2.4), on obtient

$$\mathbb{P}(A_1 \cup A_2 \cup \dots \cup A_n \cup A_{n+1}) = \sum_{k=1}^{n+1} \sum_{\substack{J \subseteq \{1,2,\dots,n+1\} \\ \text{tel que } |J|=k}} (-1)^{k+1} \mathbb{P}\left(\bigcap_{i \in J} A_i\right),$$

ce qui conclut la récurrence □

Une propriété fondamentale, qui se déduit de la Définition 2.3 est la suivante.

Proposition 2.8 (Continuité par le bas et par le haut des probabilités). *Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé. Alors on a*

1. Continuité par le bas : si $(A_n)_{n \geq 1}$ est une suite croissante d'événements, c'est-à-dire $A_{n+1} \supseteq A_n$ pour tout $n \in \mathbb{N}$, alors

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

2. Continuité par le haut : si $(A_n)_{n \geq 1}$ est une suite décroissante d'événements, c'est-à-dire $A_{n+1} \subseteq A_n$ pour tout $n \in \mathbb{N}$, alors

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

Démonstration. Pour une suite croissante $(A_n)_{n \in \mathbb{N}}$ d'événements, on définit une autre suite $(B_n)_{n \in \mathbb{N}}$ de la façon suivante : on pose $B_1 := A_1$, et $B_n := A_n \setminus A_{n-1}$ pour $n \geq 2$. Par construction, les événements B_n sont disjoints, et pour tout $n \in \mathbb{N}$, $\bigcup_{k=1}^n B_k = A_n$. En outre, $\bigcup_{n=1}^{\infty} B_n = \bigcup_{n=1}^{\infty} A_n$. Ainsi, par σ -additivité, on a

$$\begin{aligned} \mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) &= \mathbb{P}\left(\bigcup_{n=1}^{\infty} B_n\right) = \sum_{n=1}^{\infty} \mathbb{P}(B_n) = \lim_{n \rightarrow +\infty} \sum_{k=1}^n \mathbb{P}(B_k) \\ &= \lim_{n \rightarrow +\infty} \mathbb{P}\left(\bigcup_{k=1}^n B_k\right) = \lim_{n \rightarrow +\infty} \mathbb{P}(A_n). \end{aligned}$$

Soit $(A_n)_{n \geq 1}$ une suite décroissante d'événements décroissants. Posons $B_n := A_n^c$, de sorte que $(B_n)_{n \geq 1}$ est une suite croissante d'événements. Ainsi, en utilisant le point 1, on obtient

$$\mathbb{P}\left(\bigcap_{n=1}^{\infty} A_n\right) = \mathbb{P}\left(\left(\bigcup_{n=1}^{\infty} B_n\right)^c\right) = 1 - \mathbb{P}\left(\bigcup_{n=1}^{\infty} B_n\right) = 1 - \lim_{n \rightarrow +\infty} \mathbb{P}(B_n) = \lim_{n \rightarrow +\infty} \mathbb{P}(A_n),$$

ce qui conclut la démonstration du deuxième point. \square

Un corollaire utile est le suivant.

Corollaire 2.9 (Sous-additivité). *Soit $(A_n)_{n \geq 1}$ une suite d'événements. Alors*

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

Démonstration. Soit $B_n := \bigcup_{k=1}^n A_k$. Il est clair que $(B_n)_{n \geq 1}$ est une suite croissante d'événements. En outre $\bigcup_{n \geq 1} B_n = \bigcup_{n \geq 1} A_n$. Grâce à la partie 2 de la Proposition 2.5, on sait que $\mathbb{P}(A_1 \cup A_2) \leq \mathbb{P}(A_1) + \mathbb{P}(A_2)$: par récurrence, on peut facilement étendre cette inégalité,

et obtenir $\mathbb{P}(\bigcup_{k=1}^n A_k) \leq \sum_{k=1}^n \mathbb{P}(A_k)$ pour tout $n \geq 1$. Mais alors, en utilisant aussi la Proposition 2.8, on obtient

$$\begin{aligned} \mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) &= \mathbb{P}\left(\bigcup_{n=1}^{\infty} B_n\right) = \lim_{n \rightarrow +\infty} \mathbb{P}(B_n) = \lim_{n \rightarrow +\infty} \mathbb{P}\left(\bigcup_{k=1}^n A_k\right) \\ &\leq \lim_{n \rightarrow +\infty} \sum_{k=1}^n \mathbb{P}(A_k) = \sum_{n=1}^{+\infty} \mathbb{P}(A_n), \end{aligned}$$

qui est l'inégalité recherchée. \square

♠ **Exemple 2.10** (Des événements de probabilité 0 ? 1 ?). Reprenons l'exemple d'une suite *infinie* de Pile ou Face équilibrés : prenons comme espace d'état $\Omega = \{0, 1\}^{\mathbb{N}^*}$, et laissons la construction de l'ensemble des événements \mathcal{F} et de la probabilité \mathbb{P} sur \mathcal{F} de côté (on a pour cela besoin d'outils de théorie de la mesure, rappelons simplement que \mathcal{F} est construit à partir d'événements de base $B_i = \text{"le } i\text{-ème lancer est Pile"}$).

Commençons par calculer la probabilité de l'événement $A = \bigcap_{i=1}^{\infty} B_i = \text{"on n'obtient que des Pile"}$. Notons que $A = \bigcap_{n \geq 1} C_n$, où $C_n = \bigcap_{i=1}^n B_i$ est l'événement qu'il n'y a que des Pile lors des n premiers lancers. Il est naturel de demander à ce que la probabilité \mathbb{P} soit telle que $\mathbb{P}(C_n) = (\frac{1}{2})^n$, car à chaque lancer on a une chance sur deux d'obtenir Pile. Ainsi, la suite $(C_n)_{n \geq 1}$ étant décroissante (exercice), on a d'après la Proposition 2.8 que $\mathbb{P}(A) = \lim_{n \rightarrow \infty} \mathbb{P}(C_n) = 0$.

Considérons maintenant l'événement $B = \bigcup_{n \geq 1} \bigcap_{i=n}^{\infty} B_i = \text{"il n'y a que des Pile à partir d'un certain rang"}$. Notons $A_n := \bigcap_{i=n}^{\infty} B_i$, de sorte que $B = \bigcup_{n \geq 1} A_n$. Avec le même raisonnement que précédemment, on a $\mathbb{P}(A_n) = \lim_{m \rightarrow \infty} \mathbb{P}(\bigcap_{i=n}^m B_i)$, de sorte que $\mathbb{P}(A_n) = \lim_{m \rightarrow \infty} (\frac{1}{2})^{m-n} = 0$. Par sous-additivité, on a donc $\mathbb{P}(B) \leq \sum_{n=1}^{\infty} \mathbb{P}(A_n) = 0$. On en conclut que $\mathbb{P}(B) = 0$, et donc que $\mathbb{P}(B^c) = 1$, où B^c est l'événement "il y a une infinité de Face". On a donc montré qu'*avec probabilité 1* (on dit aussi *presque sûrement*), dans une suite infinie de Pile ou Face, il y a une infinité de Face. (Notons que l'une des principales difficultés est de garantir qu'il existe une probabilité \mathbb{P} définie sur \mathcal{F} qui vérifie les propriétés voulues.)

Un événement $A \in \mathcal{F}$ de probabilité $\mathbb{P}(A) = 1$ est appelé *presque sûr*.

2.2 Espaces d'états dénombrables

Dans le cas où Ω est dénombrable, on considérera l'ensemble d'événements $\mathcal{F} = \mathcal{P}(\Omega)$.

2.2.1 Probabilité et densité discrète

On montre dans ce paragraphe que dans le cas où Ω est dénombrable, pour définir une probabilité $\mathbb{P} : \mathcal{P}(\Omega) \rightarrow \mathbb{R}$, il suffit de définir un objet plus simple, appelé *densité discrète*.

Définition 2.11 (Densité discrète). *Soit Ω un ensemble non vide, dénombrable. On appelle densité discrète sur Ω toute fonction $p : \Omega \rightarrow \mathbb{R}$ qui satisfait*

$$p(\omega) \geq 0 \quad \forall \omega \in \Omega, \quad \sum_{\omega \in \Omega} p(\omega) = 1. \quad (2.7)$$

Proposition 2.12. Soit p une densité discrète sur un ensemble dénombrable Ω . La fonction $\mathbb{P} : \mathcal{P}(\Omega) \rightarrow \mathbb{R}$ définie par

$$\mathbb{P}(A) := \sum_{\omega \in A} p(\omega), \quad \forall A \subseteq \Omega, \quad (2.8)$$

satisfait les axiomes A1 et A2 de la Définition 2.3.

Démonstration. Clairement, $\mathbb{P}(\Omega) = 1$, grâce à la deuxième relation dans (2.7), donc l'axiome A1 est vérifié. En ce qui concerne l'axiome A2, étant donné que la famille de réels $(p(\omega))_{\omega \in \Omega}$ est positive, on peut *sommer par paquets* : si $(A_k)_{k \in \mathbb{N}}$ est une suite de sous-ensembles de Ω disjoints (c'est-à-dire $A_i \cap A_j = \emptyset$ pour $i \neq j$), en posant $A := \bigcup_{k \in \mathbb{N}} A_k$, on a

$$\mathbb{P}(A) = \sum_{\omega \in A} p(\omega) = \sum_{k \in \mathbb{N}} \left(\sum_{\omega \in A_k} p(\omega) \right) = \sum_{k \in \mathbb{N}} \mathbb{P}(A_k),$$

ce qui complète la démonstration. \square

Ainsi, si Ω est dénombrable, la Proposition 2.12 montre que toute densité discrète p sur Ω détermine une *probabilité* \mathbb{P} sur Ω , via la formule (2.8). Le point intéressant est que la réciproque est aussi vraie. En effet, si \mathbb{P} est une probabilité sur Ω , on peut définir la fonction $p : \Omega \rightarrow \mathbb{R}$ par $p(\omega) = \mathbb{P}(\{\omega\})$. Clairement, $p(\omega) = \mathbb{P}(\{\omega\}) \geq 0$ pour tout $\omega \in \Omega$. Pour tout $A \in \mathcal{P}(\Omega)$ on peut écrire $A = \bigcup_{\omega \in A} \{\omega\}$, où l'union est disjointe et *dénombrable*, étant donné que Ω est dénombrable : en utilisant l'axiome A2, on obtient donc que $\mathbb{P}(A) = \sum_{\omega \in A} \mathbb{P}(\{\omega\}) = \sum_{\omega \in A} p(\omega)$. En prenant $A = \Omega$, on voit que p satisfait la deuxième relation dans (2.7) et est bien une densité discrète.

Remarque 2.13. Pour résumer, dans un espace probabilisé discret, la probabilité est déterminée par ses valeurs sur les *singletons* $\{\omega\}$, c'est-à-dire sur les événements constitués d'un seul élément de Ω (parfois appelés événements élémentaires). Ce n'est plus le cas pour les espaces probabilisés généraux, comme on le verra plus loin.

2.2.2 Équiprobabilité : quelques exemples

Soit Ω un ensemble *fini*. On définit, pour tout sous-ensemble $A \subseteq \Omega$

$$\mathbb{P}(A) := \frac{|A|}{|\Omega|},$$

où on rappelle que $|\cdot|$ désigne la cardinalité d'un ensemble. On voit facilement que \mathbb{P} vérifie les axiomes A1 et A2 de la Définition 2.3 (Ω étant fini, il suffit de vérifier (2.1) au lieu de A2). Ainsi, \mathbb{P} est une probabilité, appelée *probabilité uniforme sur Ω* . La densité discrète associée est donnée par

$$p(\omega) = \mathbb{P}(\{\omega\}) = \frac{1}{|\Omega|},$$

et donc *ne dépend pas de ω* .

Exercice 2.14. Soit Ω un ensemble fini, et \mathbb{P} une probabilité sur $\mathcal{P}(\Omega)$ telle que $p(\omega) = \mathbb{P}(\{\omega\}) = c$ ne dépend pas de $\omega \in \Omega$. Montrer que \mathbb{P} est la probabilité uniforme sur Ω .

- ♠ **Exemple 2.15** (Paradoxe des anniversaires). Déterminons la probabilité p_n que, dans un groupe de n personnes sélectionnées au hasard (nées une année non bissextile pour simplifier), au moins deux d'entre elles aient leurs anniversaires le même jour. À quel point n doit-il être grand pour que $p_n > 1/2$?

L'espace d'état naturel est $\Omega = \{1, \dots, 365\}^n$, où pour un élément $\omega = (\omega_1, \dots, \omega_n)$ ω_i représente le jour de l'anniversaire de la i -ème personne. On munit Ω de la probabilité uniforme \mathbb{P} : pour tout événement A , $\mathbb{P}(A) = |A|/|\Omega|$. Il est clair que $|\Omega| = 365^n$. L'événement qui nous intéresse est $A = \{\omega \in \Omega : \exists i \neq j, \omega_i = \omega_j\}$. La cardinalité de A est difficile à appréhender, et on va plutôt calculer la cardinalité de $A^c = \{\omega \in \Omega : \forall i \neq j, \omega_i \neq \omega_j\}$: il s'agit d'*arrangements* de n éléments dans un ensemble à 365 éléments, et on a donc $|A^c| = 365 \cdot 364 \cdots (365 - n + 1)$. On obtient donc que $p_n = \mathbb{P}(A) = 1 - \mathbb{P}(A^c)$, avec $\mathbb{P}(A^c) = \frac{|A^c|}{|\Omega|} = \frac{\prod_{i=0}^{n-1} (365-i)}{365^n} = \prod_{i=0}^{n-1} \frac{365-i}{365}$, c'est-à-dire

$$p_n = 1 - \prod_{i=0}^{n-1} \left(1 - \frac{i}{365}\right).$$

On a obtenu une expression exacte, mais peu "explicite". Avec l'aide d'une calculatrice, on vérifie facilement que $p_n > \frac{1}{2}$ si et seulement si $n \geq 23$. Ainsi, dans un groupe d'au moins 23 personnes, il y a une probabilité supérieure à 50% qu'au moins deux d'entre elles aient leurs anniversaires le même jour, un résultat surprenant à première vue. Pour un groupe de 50 personnes, la probabilité est supérieure à 97%.

- ♠ **Exemple 2.16** (Père Noël secret). Un groupe de n personnes décide d'organiser un Père Noël secret : on inscrit les noms des n personnes sur des étiquettes, et chaque personne tire une étiquette au hasard : elle devra faire un cadeau à la personne dont le nom est inscrit sur l'étiquette. Quelle est la probabilité p_n qu'une personne tire son propre nom ?

L'espace d'état naturel associé à cette expérience aléatoire est $\Omega = \mathfrak{S}_n$ l'ensemble des permutations, où pour un élément $\sigma \in \mathfrak{S}_n$, $\sigma(i) = j$ signifie que la i -ème personne tire le nom de la j -ème personne. On munit \mathfrak{S}_n de la probabilité uniforme \mathbb{P} , et on rappelle que $|\mathfrak{S}_n| = n!$. L'événement qui nous intéresse est $A = \{\sigma \in \mathfrak{S}_n : \exists i, \sigma(i) = i\}$, que la permutation possède un point fixe. On peut écrire A sous la forme $\bigcup_{i=1}^n A_i$, où $A_i = \{\sigma \in \mathfrak{S}_n : \sigma(i) = i\}$ est l'événement que i est un point fixe de la permutation. D'après la formule d'inclusion-exclusion (Proposition 2.7), on peut écrire

$$p_n = \mathbb{P}(A) = \mathbb{P}\left(\bigcup_{i=1}^n A_i\right) = \sum_{k=1}^n \sum_{J \subset \{1, \dots, n\}, |J|=k} (-1)^{k+1} \mathbb{P}\left(\bigcap_{i \in J} A_i\right).$$

Maintenant, on est capable de calculer la probabilité de l'événement $A_J = \bigcap_{i \in J} A_i$ quel que soit l'ensemble $J \subset \{1, \dots, n\}$. En effet, on peut calculer la cardinalité de $A_J = \{\sigma \in \mathfrak{S}_n : \sigma(i) = i \text{ pour tout } i \in J\}$: si $|J| = k$, on a $|A_J| = (n-k)!$, car la valeur de σ sur J est déterminée, et σ restreinte à $\{1, \dots, n\} \setminus J$ est une permutation de $\{1, \dots, n\} \setminus J$. On a donc que $\mathbb{P}(A_J) = \frac{(n-k)!}{n!}$, et ainsi

$$p_n = \sum_{k=1}^n (-1)^{k+1} \binom{n}{k} \times \frac{(n-k)!}{n!} = \sum_{k=1}^n \frac{(-1)^{k+1}}{k!},$$

où on a utilisé qu'il y a exactement $\binom{n}{k}$ ensembles $J \subset \{1, \dots, n\}$ à k éléments, puis la formule $\binom{n}{k} = \frac{n!}{k!(n-k)!}$. Au final, après quelques manipulations, on trouve que $p_n = 1 - \sum_{k=0}^n \frac{(-1)^k}{k!}$: ainsi, si n est grand, on a $p_n \approx 1 - e^{-1} \approx 0,6321$.

★ Probabilité que deux entiers soient premiers entre eux

On souhaite répondre à la question suivante : pour $n \geq 1$, quelle est la probabilité p_n que deux entiers pris au hasard dans $\{1, \dots, n\}$ soient premiers entre eux ? Que vaut la limite de p_n quand $n \rightarrow \infty$, si elle existe ? Pour formaliser ce problème, on prend comme espace d'état $\Omega = \{1, \dots, n\}^2$ (les couples d'entiers possibles), que l'on munit de la probabilité uniforme \mathbb{P} . L'événement dont l'on souhaite estimer la probabilité est

$$A = \text{“les deux entiers sont premiers entre eux”} = \{(a, b) \in \Omega_n : \text{pgcd}(a, b) = 1\}.$$

Pour $q \in \mathbb{N}$, on pose $D_q = \text{“}q \text{ divise les deux entiers”} = \{(a, b) \in \Omega_n : q|a \text{ et } q|b\}$. On remarque alors que $D_q = M_q \times M_q$, où $M_q = \{mq : 1 \leq m \leq \lfloor n/q \rfloor\}$: on obtient donc $\mathbb{P}(D_q) = \frac{|D_q|}{|\Omega_n|} = \frac{\lfloor n/q \rfloor^2}{n^2}$.

Pour calculer la probabilité de A , on écrit $A^c = D_{p_1} \cup \dots \cup D_{p_\ell}$ où p_1, \dots, p_ℓ désignent les nombres premiers inférieurs à n , ce qui permet d'obtenir, en utilisant la formule d'inclusion-exclusion (Proposition 2.7),

$$\mathbb{P}(A^c) = \sum_{k=1}^{\ell} \sum_{J \subset \{1, \dots, \ell\}, |J|=k} (-1)^{k+1} \mathbb{P}\left(\bigcap_{i \in J} D_{p_i}\right).$$

Comme les p_i sont des nombres premiers distincts, on a $D_{p_{i_1}} \cap \dots \cap D_{p_{i_k}} = D_{p_{i_1} \dots p_{i_k}}$. Au vu de la probabilité $\mathbb{P}(D_q)$ calculée plus haut, on obtient

$$\begin{aligned} \mathbb{P}(A) = 1 - \mathbb{P}(A^c) &= 1 + \sum_{k=1}^{\ell} \sum_{J \subset \{1, \dots, \ell\}, |J|=k} (-1)^k \frac{1}{n^2} \left[\frac{n}{\prod_{i \in J} p_i} \right]^2 \\ p_n &= \sum_{d=1}^n \mu(d) \frac{1}{n^2} \lfloor n/d \rfloor^2, \end{aligned} \quad (2.9)$$

où on a réécrit la somme comme portant sur tous les entiers $1 \leq d \leq n$: $\mu(d)$ vaut 0 si d possède un facteur premier avec une multiplicité supérieure à 2, et sinon $\mu(d) = (-1)^k$ où k est le nombre de facteurs premiers de d (pour le terme $d = 1$, on a par convention $\mu(d) = 1$). La fonction $\mu(d)$ est connue en arithmétique sous le nom de *fonction de Möbius*.

Étudions maintenant la limite de p_n quand $n \rightarrow +\infty$. Tout d'abord, montrons que

$$\lim_{n \rightarrow +\infty} p_n = \sum_{d=1}^{+\infty} \mu(d) \frac{1}{d^2}, \quad (2.10)$$

cette dernière somme étant absolument convergente (car $|\mu(d)/d^2| \leq 1/d^2$). En effet, en partant de (2.9), on obtient par inégalité triangulaire et en utilisant que $|\mu(d)| \leq 1$

$$\left| p_n - \sum_{d=1}^n \mu(d) \frac{1}{d^2} \right| \leq \sum_{d=1}^n \left| \frac{\lfloor n/d \rfloor^2}{n^2} - \frac{1}{d^2} \right| \leq \frac{2}{n} \sum_{d=1}^n \frac{1}{d},$$

où on a utilisé pour la dernière inégalité que $n/d - 1 \leq \lfloor n/d \rfloor \leq n/d$, de sorte que $n^2/d^2 - 2n/d \leq \lfloor n/d \rfloor^2 \leq n^2/d^2$. En utilisant le fait que pour la série harmonique on a $\sum_{d=1}^n \frac{1}{d} \sim \ln n$ quand $n \rightarrow \infty$, on obtient donc que $\left| p_n - \sum_{d=1}^n \mu(d)/d^2 \right| \rightarrow 0$, ce qui démontre (2.10).

Maintenant, on peut calculer explicitement la série dans le membre de droite de (2.10). Écrivons le produit des deux séries convergentes $\sum_{d=1}^{\infty} \mu(d) \frac{1}{d^2}$ et $\sum_{m=1}^{\infty} \frac{1}{m^2}$ (cette dernière série vaut $\pi^2/6$, d'après un résultat dû à Euler) :

$$\left(\sum_{m=1}^{+\infty} \frac{1}{m^2} \right) \left(\sum_{d=1}^{+\infty} \frac{\mu(d)}{d^2} \right) = \sum_{(m,d) \in \mathbb{N}_* \times \mathbb{N}_*} \frac{\mu(d)}{(md)^2} = \sum_{(k,d) \in \mathbb{N}_* \times \mathbb{N}_* : d|k} \frac{\mu(d)}{k^2} = \sum_{k=1}^{+\infty} \frac{1}{k^2} \sum_{d|k} \mu(d).$$

Maintenant, on utilise une propriété importante de la fonction de Möbius, dont nous laissons la démonstration au lecteur : la somme $\sum_{d|k} \mu(d)$ vaut 1 pour $k = 1$, et vaut 0 pour $k \geq 2$.

Le seul terme non nul dans la dernière somme est donc le terme $k = 1$, ce qui montre que $(\sum_{d=1}^{\infty} \frac{\mu(d)}{d^2})(\sum_{m=1}^{\infty} \frac{1}{m^2}) = 1$. On a donc démontré le résultat suivant, dû à Cesàro,

$$\lim_{n \rightarrow +\infty} p_n = \left(\sum_{m=1}^{+\infty} \frac{1}{m^2} \right)^{-1} = \frac{6}{\pi^2}.$$

Ainsi, si l'on prend deux entiers au hasard dans $\{1, \dots, n\}$, la probabilité qu'ils soient premiers entre eux converge vers $6/\pi^2$ quand $n \rightarrow +\infty$.

2.3 Probabilité conditionnelle et indépendance

2.3.1 Probabilité conditionnelle

Considérons un événement A d'une expérience aléatoire, qui possède une probabilité $\mathbb{P}(A)$. Si l'on prend connaissance du fait qu'un autre événement B est vérifié, comment doit-on ajuster la valeur de $\mathbb{P}(A)$ pour tenir compte de cette information ? La réponse est la probabilité conditionnelle $\mathbb{P}(A|B)$ de A sachant B .

Définition 2.17 (Probabilité conditionnelle). Soient A et B deux événements d'un espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$, avec $\mathbb{P}(B) > 0$. On appelle probabilité conditionnelle de A sachant B la quantité

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

Pour motiver et "deviner" cette définition, revenons un instant sur l'interprétation fréquentiste des probabilités. En répétant un grand nombre N de fois l'expérience aléatoire, et en comptant le nombre de fois où l'événement A est vérifié, on a $\mathbb{P}(A) \approx N_A/N$. Pour tenir compte de l'information supplémentaire que l'événement B a été vérifié, il faut se restreindre aux fois où l'expérience a donné un résultat dans B (au nombre de N_B), et de compter les fois où A a aussi été vérifié (au nombre de $N_{A \cap B}$). En définissant la probabilité conditionnelle comme la proportion de fois où A est réalisé parmi les fois où B est réalisé, on obtient pour N grand $\mathbb{P}(A|B) \approx \frac{N_{A \cap B}}{N_B} = \frac{N_{A \cap B}/N}{N_B/N} \approx \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$.

Certaines propriétés importantes des probabilités conditionnelles sont résumées dans le résultat suivant, laissé en exercice.

Exercice 2.18. Soit B un événement d'un espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$, tel que $\mathbb{P}(B) > 0$. Montrer que la fonction $\mathbb{P}(\cdot | B) : A \in \mathcal{F} \mapsto \mathbb{P}(A|B)$, est une probabilité sur (Ω, \mathcal{F}) .

Remarquons que la probabilité conditionnelle permet d'exprimer la probabilité de l'intersection de n événements de manière récursive.

Proposition 2.19 (Probabilités conditionnelles en chaîne). Soient $n \geq 2$ et A_1, \dots, A_n des événements tels que $\mathbb{P}(A_1 \cap A_2 \cap \dots \cap A_{n-1}) > 0$. Alors

$$\begin{aligned} \mathbb{P}\left(\bigcap_{i=1}^n A_i\right) &= \mathbb{P}(A_1)\mathbb{P}(A_2|A_1)\mathbb{P}(A_3|A_1 \cap A_2) \cdots \mathbb{P}(A_n|A_1 \cap \dots \cap A_{n-1}) \\ &= \mathbb{P}(A_1) \cdot \prod_{i=2}^n \mathbb{P}(A_i|A_1 \cap \dots \cap A_{i-1}). \end{aligned} \tag{2.11}$$

Démonstration. Remarquons que $(A_1 \cap A_2 \cap \dots \cap A_j) \subseteq (A_1 \cap A_2 \cap \dots \cap A_i)$ pour $i \leq j$: en particulier, si $\mathbb{P}(A_1 \cap A_2 \cap \dots \cap A_{n-1}) > 0$ alors $\mathbb{P}(A_1 \cap A_2 \cap \dots \cap A_i) > 0$ pour tout $1 \leq i \leq n-1$, donc les probabilités conditionnelles dans (2.11) sont bien définies.

La démonstration se fait par récurrence. Pour deux événements B, C avec $\mathbb{P}(C) > 0$, on a, par définition de la probabilité conditionnelle,

$$\mathbb{P}(B \cap C) = \mathbb{P}(B)\mathbb{P}(C|B), \quad (2.12)$$

ce qui implique que la relation (2.11) est vérifiée pour $n = 2$. Pour l'étape d'hérédité, on remarque que grâce à la relation (2.12) on a

$$\mathbb{P}\left(\bigcap_{i=1}^n A_i\right) = \mathbb{P}\left(\left(\bigcap_{i=1}^{n-1} A_i\right) \cap A_n\right) = \mathbb{P}\left(\bigcap_{i=1}^{n-1} A_i\right) \mathbb{P}\left(A_n \mid \bigcap_{i=1}^{n-1} A_i\right).$$

En appliquant l'hypothèse de récurrence, on obtient la relation (2.11). \square

On donne maintenant des formules essentielles, qui permettent de “décomposer” la probabilité d'un événement en probabilités d'événements plus élémentaires.

Proposition 2.20. *Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé, et soit $(B_i)_{i \in I}$ une partition dénombrable de Ω , c'est-à-dire $B_i \cap B_j = \emptyset$ pour tout $i \neq j$ et $\bigcup_{i \in I} B_i = \Omega$. Pour un événement A , on a la formule de décomposition suivante*

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

Si de plus $\mathbb{P}(B_i) > 0$ pour tout $i \in I$, on a la formule des probabilités totales :

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

Démonstration. Par hypothèse $\Omega = \bigcup_{i \in I} B_i$, donc

$$A = A \cap \Omega = A \cap \left(\bigcup_{i \in I} B_i\right) = \bigcup_{i \in I} (A \cap B_i).$$

Étant donné que les événements $(B_i)_{i \in I}$ sont disjoints, les événements $(A \cap B_i)_{i \in I}$ le sont aussi. Comme I est dénombrable on obtient par σ -additivité $\mathbb{P}(A) = \sum_{i \in I} \mathbb{P}(A \cap B_i)$. Enfin, si $\mathbb{P}(B_i) > 0$, on peut écrire $\mathbb{P}(A \cap B_i) = \mathbb{P}(A|B_i)\mathbb{P}(B_i)$ par définition des probabilités conditionnelles. \square

Exemple 2.21. Deux urnes contiennent respectivement 3 boules rouges et 1 verte (l'urne α) et 1 boule rouge et 1 verte (l'urne β). On choisit, avec la même probabilité, l'une des deux urnes, et de l'urne choisie on tire une boule au hasard. Quelle est la probabilité de tirer une boule rouge? Notons R l'événement “tirer une boule rouge”, et A l'événement “choisir l'urne α ” : les renseignements de l'énoncé donnent clairement $\mathbb{P}(A) = \frac{1}{2}$, $\mathbb{P}(R|A) = \frac{3}{4}$ et $\mathbb{P}(R|A^c) = \frac{1}{2}$. En appliquant la formule des probabilités totales, on obtient donc

$$\mathbb{P}(R) = \mathbb{P}(R|A)\mathbb{P}(A) + \mathbb{P}(R|A^c)(1 - \mathbb{P}(A)) = \frac{3}{4} \frac{1}{2} + \frac{1}{2} \frac{1}{2} = \frac{5}{8}.$$

Soient maintenant A et B deux événements tels que $\mathbb{P}(A) > 0$, $\mathbb{P}(B) > 0$, de sorte que les deux probabilités conditionnelles $\mathbb{P}(A|B)$ et $\mathbb{P}(B|A)$ sont bien définies. Il est presque immédiat d'établir la relation suivante.

Théorème 2.22 (Formule de Bayes). *Si A, B sont des événements d'un espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$, avec $\mathbb{P}(A) > 0$ et $\mathbb{P}(B) > 0$, alors on a la formule*

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(A|B)\mathbb{P}(B)}{\mathbb{P}(A)}. \quad (2.13)$$

Démonstration. La formule de Bayes (2.13) est équivalente à $\mathbb{P}(B|A)\mathbb{P}(A) = \mathbb{P}(A|B)\mathbb{P}(B)$, ce qui est vérifié parce que, par définition des probabilités conditionnelles, les deux membres sont égaux à $\mathbb{P}(A \cap B)$. \square

La formule de Bayes, bien que d'apparence simple, revêt une importance fondamentale, et est à l'origine d'un pan entier de la statistique, appelée *statistique bayésienne*.

Exemple 2.23. Pour déterminer la présence d'un virus, on dispose d'un test clinique : si le virus est présent, le test est positif dans 100% des cas ; si le virus est absent, le test est positif dans 2% des cas (un *faux positif*). On sait que 5 personnes sur 10 000 ont le virus. On fait un test sur un individu choisi au hasard, et le résultat est positif. Avec quelle confiance peut-on affirmer qu'il ou elle soit malade ?

On considère les deux événements suivants : A = "le test est positif" et B = "l'individu est malade". Les données de l'énoncé permettent de déterminer $\mathbb{P}(B) = 0,0005$, $\mathbb{P}(A|B) = 1$, et $\mathbb{P}(A|B^c) = 0,02$. La réponse à la question posée est donnée par $\mathbb{P}(B|A)$. En utilisant la formule de Bayes et la formule des probabilités totales, on a

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(A|B)\mathbb{P}(B)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A|B)\mathbb{P}(B)}{\mathbb{P}(A|B)\mathbb{P}(B) + \mathbb{P}(A|B^c)\mathbb{P}(B^c)} = \frac{0,0005}{0,02049} \approx 0,025$$

qui est extrêmement faible. Ainsi, même si le résultat du test est positif, il est très peu probable que l'individu soit malade ! Ce test donnera en réalité une très grande proportion de faux positifs.

2.3.2 Indépendance d'événements

On a vu plus haut comment la probabilité conditionnelle $\mathbb{P}(A|B)$ représente la probabilité de l'événement A sous la condition que l'événement B soit vérifié. Il est possible qu'une telle condition ne modifie pas la probabilité de A , c'est-à-dire $\mathbb{P}(A|B) = \mathbb{P}(A)$. En utilisant la définition de probabilité conditionnelle, on voit que c'est équivalent à $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$: cette identité possède l'avantage d'être explicitement symétrique en A et B , et d'être bien définie même quand $\mathbb{P}(B) = 0$.

Définition 2.24 (Indépendance de deux événements). *Deux événements A et B d'un espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$ sont dit indépendants si*

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

Exemple 2.25. Si Aïssata, Bérénice et Clémentine lancent chacune un dé régulier à 6 faces, les événements $A :=$ “Aïssata et Bérénice obtiennent le même résultat” et $B :=$ “Bérénice et Clémentine obtiennent le même résultat” sont indépendants. On laisse en exercice, le soin de montrer que $\mathbb{P}(A) = \frac{1}{6}$, $\mathbb{P}(B) = \frac{1}{6}$ et $\mathbb{P}(A \cap B) = \frac{1}{36}$, ce qui permet de vérifier que $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.

Dans cet exemple, bien qu’il puisse y avoir intuitivement un type “d’influence” entre les événements A et B , ils sont indépendants au sens de la Définition 2.24 : *le fait que l’un soit vérifié ne modifie pas la probabilité de l’autre.*

Même si c’est évident à partir de la définition, soulignons que deux événements indépendants *ne sont pas disjoints* (sauf dans le cas trivial où un des événements est de probabilité 0) : en effet, si A, B sont indépendants et $\mathbb{P}(A) > 0$, $\mathbb{P}(B) > 0$, il s’ensuit que $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B) > 0$, et donc $A \cap B \neq \emptyset$.

Remarque 2.26. Si A et B sont des événements indépendants, alors A^c et B le sont aussi :

$$\begin{aligned} \mathbb{P}(A^c \cap B) &= \mathbb{P}(B \setminus (A \cap B)) = \mathbb{P}(B) - \mathbb{P}(A \cap B) = \mathbb{P}(B) - \mathbb{P}(A)\mathbb{P}(B) \\ &= (1 - \mathbb{P}(A))\mathbb{P}(B) = \mathbb{P}(A^c)\mathbb{P}(B). \end{aligned}$$

De même, A et B^c sont aussi indépendants, ainsi que A^c et B^c .

Cherchons maintenant à comprendre comment définir l’indépendance de plus de deux événements. Revenons à l’Exemple 2.25, et considérons les trois événements $A :=$ “Aïssata et Bérénice obtiennent le même résultat” $B :=$ “Bérénice et Clémentine obtiennent le même résultat” et $C :=$ “Aïssata et Clémentine obtiennent le même résultat”. Les événements $\{A, B\}$ sont indépendants, et il en va de même pour les événements $\{A, C\}$ et $\{B, C\}$. Cependant, on remarque que $A \cap B \cap C = A \cap B$, de sorte que

$$\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A \cap B) = \frac{1}{36} \neq \frac{1}{216} = \frac{1}{6^3} = \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C).$$

Cela suggère que l’indépendance de trois événements A, B, C n’est pas la même chose que l’indépendance deux à deux, de $\{A, B\}$, de $\{A, C\}$ et de $\{B, C\}$.

Définition 2.27 (Indépendance d’événements). *Soit $(A_i)_{i \in I}$ une famille d’événements d’un espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$, où l’ensemble d’indices I est arbitraire. Cette famille d’événements est dite indépendante si pour tout sous-ensemble fini $J \subseteq I$ ($|J| \geq 2$) on a*

$$\mathbb{P}\left(\bigcap_{j \in J} A_j\right) = \prod_{j \in J} \mathbb{P}(A_j). \quad (2.14)$$

Pour l’indépendance de n événements A_1, \dots, A_n , il ne suffit donc pas de vérifier que

$$\mathbb{P}(A_1 \cap A_2 \cap \dots \cap A_n) = \mathbb{P}(A_1) \cdot \mathbb{P}(A_2) \cdots \mathbb{P}(A_n), \quad (2.15)$$

mais il faut montrer que cette propriété de factorisation est vraie *pour toute sous-famille*, comme demandé dans la relation (2.14).

Comme précédemment, si dans une famille arbitraire d’événements indépendants, on remplace certains événements par leurs complémentaires, on obtient à nouveau une famille d’événements indépendants. On laisse ce résultat en exercice.

Exemple 2.28. Soient n événements A_1, \dots, A_n , indépendants, tous de probabilité $\mathbb{P}(A_i) = p$, où $p \in]0, 1[$ est un paramètre : on interprète A_i comme l'événement d'avoir un succès lors la i -ème épreuve de la répétition de n épreuves indépendantes de probabilité de succès p . Pour $k \in \{0, \dots, n\}$, on considère l'événement $B_k =$ "il y a exactement k succès parmi les n tentatives". En décomposant suivant les indices des épreuves qui sont des succès, on peut écrire

$$B_k = \bigcup_{I \subset \{1, \dots, n\}, |I|=k} \left(\bigcap_{i \in I} A_i \right) \left(\bigcap_{i \notin I} A_i^c \right).$$

L'union est disjointe (pourquoi ?), et par σ -additivité, on a donc

$$\mathbb{P}(B_k) = \sum_{I \subset \{1, \dots, n\}, |I|=k} \mathbb{P} \left(\bigcap_{i \in I} A_i \bigcap_{i \notin I} A_i^c \right) = \sum_{I \subset \{1, \dots, n\}, |I|=k} \prod_{i \in I} \mathbb{P}(A_i) \prod_{i \notin I} \mathbb{P}(A_i^c),$$

où on a utilisé l'indépendance des $(A_i)_{1 \leq i \leq n}$ pour la dernière égalité. Comme $\mathbb{P}(A_i) = p$, $\mathbb{P}(A_i^c) = 1 - p$, et qu'il y a $\binom{n}{k}$ sous-ensembles $I \subset \{1, \dots, n\}$ de cardinal k , on obtient finalement

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

♠ **Exemple 2.29** (Paradoxe du singe de Borel). Un singe tape au hasard sur les touches d'un clavier. Arrivera-t-il, un jour ou l'autre, à écrire *Les Misérables* de Victor Hugo ?

Disons que le clavier possède $K := 100$ touches, et supposons que chaque touche ait la même probabilité $q = \frac{1}{K}$ d'être tapée. On fait aussi l'hypothèse que les touches tapées par le singe sont "choisies" indépendamment. Les Misérables compte $N := 3\,000\,000$ (pour simplifier) caractères, espaces inclus. Nous appellerons "première tentative" la suite des N premières touches tapées par le singe, "deuxième tentative" la suite des N touches suivantes tapées, etc. et pour chacune des tentatives nous appellerons "succès" l'événement que la suite des touches tapées coïncide *exactement* avec le texte des Misérables : on notera A_i l'événement que la i -ème tentative soit un succès. La probabilité p d'un seul succès (c'est-à-dire que N touches tapées au hasard soient précisément les "bonnes") est ridiculement faible, mais elle est néanmoins strictement positive : par indépendance des touches tapées par le singe lors d'une tentative, on aura

$$p = \mathbb{P}(A_i) = q^N = \frac{1}{K^N} = 10^{-3\,000\,000} > 0.$$

Notons que l'événement que le singe réussisse un jour à taper le texte des Misérables s'écrit $A = \bigcup_{i=1}^{\infty} A_i$. On peut calculer la probabilité de A^c en utilisant la continuité des probabilités par le haut :

$$\mathbb{P}(A^c) = \mathbb{P} \left(\bigcap_{i=1}^{\infty} A_i^c \right) = \lim_{n \rightarrow \infty} \mathbb{P} \left(\bigcap_{i=1}^n A_i^c \right) = \lim_{n \rightarrow \infty} (1-p)^n,$$

où on a utilisé l'indépendance des tentatives, c'est-à-dire de la famille d'événements $(A_i)_{i \in \mathbb{N}}$. Ainsi, comme $1 - p < 1$, on a $\mathbb{P}(A^c) = 0$, c'est-à-dire $\mathbb{P}(A) = 1$: *presque sûrement*, le singe réussira à écrire le texte complet des Misérables (on peut même montrer que presque sûrement il l'écrira même une infinité de fois). La solution du "paradoxe" tient dans le fait que le nombre de tentatives nécessaires est gigantesque !

Chapitre 3

Variables aléatoires discrètes

La notion mathématique de *variable aléatoire* formalise l'idée d'une quantité qui dépend du résultat d'une expérience aléatoire. Deux exemples de variables aléatoires sont

- le nombre X obtenu lors du lancer d'une dé équilibré ;
- le nombre Y de Pile obtenus lors de n Pile ou Face.

De nombreux événements peuvent s'exprimer en termes de variables aléatoires, par exemple

- $A :=$ "Le résultat du dé est un nombre pair" = " $X \in \{2, 4, 6\}$ ";
- $B :=$ "on n'obtient aucun Pile lors de n Pile ou Face" = " $Y = 0$ ".

Soulignons la différence entre *événements* et *variables aléatoires* : les premiers correspondent à des *affirmations* sur le résultat d'une expérience aléatoire, alors que les secondes correspondent à des *quantités* qui sont fonctions de ce résultat. Pour un événement A , la question que l'on se pose est " A est-il vérifié?", et la réponse est "oui" ou bien "non", alors que pour une variable aléatoire X , la question que l'on se pose est "quelle valeur prend X ?".

Dans tout ce chapitre, on se donne $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité, et E un ensemble quelconque.

3.1 Variables aléatoires discrètes : premières définitions et propriétés

Définition 3.1 (Variable aléatoire discrète). *On appelle variable aléatoire discrète toute fonction $X : \Omega \rightarrow E$ telle que $\mathcal{X} := X(\Omega) = \{X(\omega), \omega \in \Omega\}$ est dénombrable, et telle que pour tout $x \in \mathcal{X}$ les ensembles $\{\omega \in \Omega, X(\omega) = x\}$ soient des événements de \mathcal{F} . (Dans le cas où Ω est dénombrable, $\mathcal{F} = \mathcal{P}(\Omega)$ et c'est automatiquement le cas.)*

Si $E = \mathbb{R}$, on dit que X est une variable aléatoire *réelle*. Si $E = \mathbb{R}^n$ on dit que X est un *vecteur aléatoire* de dimension n .

Notons que pour $A \subset E$, on a $\{\omega \in \Omega, X(\omega) \in A\} = \{\omega \in \Omega, X(\omega) \in A \cap \mathcal{X}\}$. Ainsi, on peut écrire $\{\omega \in \Omega, X(\omega) \in A\} = \bigcup_{x \in A \cap \mathcal{X}} \{\omega \in \Omega, X(\omega) = x\}$, et comme l'union est dénombrable (car $\mathcal{X} := X(\Omega)$ est dénombrable) et que tous les ensembles sont des événements de \mathcal{F} (d'après la définition), l'ensemble $\{\omega \in \Omega, X(\omega) \in A\}$ est aussi un événement de \mathcal{F} .

Exemple 3.2. Un cas “banal” de variable aléatoire est donné par une fonction X constante : $X(\omega) = c$ pour tout $\omega \in \Omega$, où $c \in E$ est un élément fixé. Une telle variable aléatoire est définie sur n’importe quel espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$ et sera identifiée à l’élément $c \in E$.

Exemple 3.3. Pour tout événement fixé $A \in \mathcal{F}$, la fonction indicatrice $X = \mathbb{1}_A : \Omega \rightarrow \mathbb{R}$ définit une variable aléatoire réelle, appelée *indicatrice de l’événement A* . La variable aléatoire X prend seulement deux valeurs : $X(\omega) = 1$ si $\omega \in A$ et $X(\omega) = 0$ si $\omega \notin A$. Réciproquement, toute variable aléatoire réelle qui ne prend que les valeurs 0 et 1 est nécessairement une indicatrice : on a en effet $X = \mathbb{1}_A$, où $A := \{\omega \in \Omega : X(\omega) = 1\}$. Malgré leur apparente simplicité, les variables aléatoires indicatrices, dites *de Bernoulli*, jouent un rôle important en probabilité. Nous les rencontrerons souvent par la suite...

Un point sur les notations. Soit $X : \Omega \rightarrow E$ une variable aléatoire. Pour tout $A \subset E$, on notera $\{X \in A\}$ la préimage de A par X , à savoir

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

En d’autres termes, $\{X \in A\}$ est le sous-ensemble de Ω (l’événement) constitué des résultats ω de l’expérience aléatoire pour lesquels $X(\omega) \in A$.

Les événements du type $\{X \in A\}$, pour $A \subseteq E$, sont appelés *événements générés par X* , et sont décrits informellement par “ X prend une valeur dans A ”. Par exemple, si $X : \Omega \rightarrow \mathbb{R}$ est une variable aléatoire qui décrit le nombre obtenu en lançant un dé à six faces, l’affirmation “le nombre obtenu est pair” se traduit par l’écriture $\{X \in \{2, 4, 6\}\}$, qui définit un sous-ensemble de Ω , c’est-à-dire un événement.

On écrira aussi pour simplifier $\{X = x\}$ au lieu de $\{X \in \{x\}\}$, et si X est une variable aléatoire réelle, $\{X > a\}$ au lieu de $\{X \in]a, \infty[\}$, etc. De plus, si X, Y sont des variables aléatoires réelles définies sur le même espace de probabilité, on écrira $\{X = Y\}$ au lieu de $\{X - Y = 0\}$, etc. On écrira aussi $\mathbb{P}(X \in A)$ et $\mathbb{P}(X = x)$ au lieu de $\mathbb{P}(\{X \in A\})$ et $\mathbb{P}(\{X = x\})$, pour alléger les notations (soulignons que les probabilités sont bien définies car les ensembles $\{X \in A\}$ et $\{X = x\}$ sont des événements de \mathcal{F} , comme vu plus haut).

Opérations sur les variables aléatoires. On *admettra* les résultats suivants, qui bien qu’ils découlent de la Définition 3.1, requièrent une manipulation plus avancée du concept de tribu (certains résultats sont accessibles, mais on y reviendra en L3).

\rightsquigarrow Si X et Y sont des variables aléatoires discrètes *réelles* définies sur le même espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$, alors les fonctions suivantes sont des variables aléatoires :

- $f(X) : \omega \mapsto f(X(\omega))$, où f est une fonction arbitraire ;
- $aX + bY : \omega \mapsto aX(\omega) + bY(\omega)$ pour $a, b \in \mathbb{R}$;
- $\max(X, Y) : \omega \mapsto \max\{X(\omega), Y(\omega)\}$ et $\min(X, Y) : \omega \mapsto \min\{X(\omega), Y(\omega)\}$.

\rightsquigarrow Si $(X_n)_{n \in \mathbb{N}}$ est une suite de variables aléatoires discrètes à valeurs dans $\mathbb{R} \cup \{\pm\infty\}$ définies sur le même espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$, les fonctions suivantes sont des variables aléatoires :

- $\sup_{n \in \mathbb{N}} X_n : \omega \mapsto \sup\{X_n(\omega), n \in \mathbb{N}\}$ et $\inf_{n \in \mathbb{N}} X_n : \omega \mapsto \inf\{X_n(\omega), n \in \mathbb{N}\}$;
- $\lim_{n \rightarrow \infty} X_n : \omega \mapsto \lim_{n \rightarrow \infty} X_n(\omega)$, si la limite existe pour tout $\omega \in \Omega$ (par exemple si pour tout $\omega \in \Omega$ la suite $(X_n(\omega))_{n \in \mathbb{N}}$ est monotone) ;
- $\sum_{n \in \mathbb{N}} X_n : \omega \mapsto \sum_{n \in \mathbb{N}} X_n(\omega)$, si la somme est bien définie pour tout $\omega \in \Omega$ (par exemple si $X_n(\omega) \geq 0$ pour tout ω).

Loi et densité discrète

À toute variable aléatoire discrète $X : \Omega \rightarrow E$ on associe la *loi* μ_X de X , qui est une probabilité sur l'espace d'arrivée E .

Définition 3.4 (Loi et densité discrète). Soit $X : \Omega \rightarrow E$ une variable aléatoire discrète.

- On appelle loi de X l'application $\mu_X : \mathcal{P}(E) \rightarrow [0, 1]$ définie par

$$\mu_X(A) := \mathbb{P}(X \in A), \quad \forall A \subseteq E. \quad (3.1)$$

- On appelle densité discrète de X l'application $p_X : E \rightarrow [0, 1]$ définie par

$$p_X(x) := \mathbb{P}(X = x) = \mu_X(\{x\}), \quad \forall x \in E. \quad (3.2)$$

Le fait de vérifier que μ_X vérifie la Définition 2.3 d'une probabilité est laissée en exercice. L'importance de la loi μ_X d'une variable aléatoire X réside dans le fait qu'elle décrit *quelles valeurs prend X , et avec quelle probabilité*. Si l'on souhaite calculer la probabilité "que X fasse quelque chose" (c'est-à-dire d'un événement généré par X), il n'est pas nécessaire de connaître en détail comment est définie la fonction X , mais il suffit de connaître sa loi.

Exemple 3.5. Soit $A \subseteq \Omega$ un événement, et $X := \mathbf{1}_A$ la variable aléatoire indicatrice de l'événement A . Clairement $X(\Omega) = \{0, 1\}$ et en outre $\mathbb{P}(X = 1) = \mathbb{P}(A)$, donc $\mathbb{P}(X = 0) = 1 - \mathbb{P}(A)$. La densité discrète de X est donc donnée par $p_X(1) = \mathbb{P}(A)$, $p_X(0) = 1 - \mathbb{P}(A)$ et $p_X(x) = 0$ si $x \notin \{0, 1\}$. Une telle variable aléatoire est dite de *Bernoulli*.

Proposition 3.6. Soit $X : \Omega \rightarrow E$ une variable aléatoire discrète. La fonction p_X est une densité discrète sur \mathcal{X} au sens de la Définition 2.11, et est nulle en dehors de \mathcal{X} . La loi μ_X est la probabilité sur $(\mathcal{X}, \mathcal{P}(\mathcal{X}))$ associée à la densité discrète p_X (Proposition 2.12), étendue à une probabilité sur $(E, \mathcal{P}(E))$ par

$$\mu_X(A) = \sum_{x \in A} p_X(x) = \sum_{x \in A \cap \mathcal{X}} p_X(x), \quad A \in \mathcal{P}(E).$$

Démonstration. Il est clair d'après la définition (3.2) que $p_X(x) \geq 0$, que $\sum_{x \in \mathcal{X}} p_X(x) = 1$, et que $p_X(x) = 0$ pour $x \notin \mathcal{X}$ (car $\{X = x\} = \emptyset$ si $x \notin \mathcal{X}$). On a déjà remarqué plus haut que pour tout $A \subseteq E$ on peut écrire

$$\{X \in A\} = \bigcup_{x \in A \cap \mathcal{X}} \{X = x\}. \quad (3.3)$$

L'union dans le membre de droite est dénombrable (\mathcal{X} est dénombrable) et *disjointe* : en effet, s'il existe $\omega \in \{X = x\} \cap \{X = y\}$, cela signifie que $X(\omega) = x$ et $X(\omega) = y$, donc que $x = y$. Par conséquent, par σ -additivité de \mathbb{P} , et avec les définitions (3.1) et (3.2), on a

$$\mu_X(A) = \mathbb{P}(X \in A) = \sum_{x \in A \cap \mathcal{X}} \mathbb{P}(X = x) = \sum_{x \in A \cap \mathcal{X}} p_X(x) = \sum_{x \in A} p_X(x),$$

où la dernière égalité est due au fait que $p_X(x) = 0$ si $x \notin \mathcal{X}$. \square

Exemple 3.7. On considère le lancer d'un dé régulier à six faces. L'espace probabilisé naturel pour cette expérience est $\Omega = \{1, 2, 3, 4, 5, 6\}$, muni de la probabilité uniforme \mathbb{P} . Le nombre obtenu en lançant le dé est représenté par la variable aléatoire $X : \Omega \rightarrow \mathbb{R}$, définie simplement par $X(\omega) = \omega$ pour tout $\omega \in \Omega$.

Si on suppose maintenant que le lancer de dé fait partie d'une expérience plus grande, qui comprend aussi un Pile ou Face, on modifie l'espace probabilisé : on prend $\Omega' = \{1, 2, 3, 4, 5, 6\} \times \{P, F\}$ muni de la probabilité \mathbb{P}' uniforme sur Ω' . La variable aléatoire qui correspond au nombre obtenu par le lancer du dé est maintenant représenté par une *fonction différente* $X' : \Omega' \rightarrow \mathbb{R}$, le domaine de définition étant différent : $X'(\omega') := i$ pour tout $\omega' = (i, a) \in \Omega'$ (avec $1 \leq i \leq 6$ et $a \in \{P, F\}$).

On remarque que les variables aléatoires X et X' , définies sur des espaces différents Ω et Ω' , sont toutes les deux à valeurs dans le même espace \mathbb{R} et ont la même loi, c'est-à-dire $\mu_X = \mu_{X'}$ (exercice). Ainsi, afin de calculer des probabilités qui ne concernent que la seule variable aléatoire X ou X' , travailler avec l'une ou avec l'autre n'a pas d'importance : comme elles ont la même loi, les probabilités correspondantes sont égales.

Deux variables aléatoires X, X' , définies sur le même espace Ω et à valeurs dans le même ensemble E sont dites *presque sûrement égales* si $\mathbb{P}(X = X') = \mathbb{P}(\{\omega \in \Omega : X(\omega) = X'(\omega)\}) = 1$.

En d'autres mots, deux variables aléatoires X et X' sont presque sûrement égales si elles diffèrent sur un ensemble de probabilité nulle : $\mathbb{P}(X \neq X') = 0$. On laisse le résultat suivant en exercice.

Exercice 3.8. Montrer que si deux variables aléatoires discrètes X et X' sont presque sûrement égales alors elles ont la même loi (et la même densité discrète).

Indication : on pourra démontrer que si C est un événement de probabilité 1, alors $\mathbb{P}(B \cap C) = \mathbb{P}(B)$ pour tout événement B ; et ensuite l'appliquer à $C = \{X = X'\}$ et $B = \{X \in A\}$, $B' = \{X' \in A\}$.

Dans la plupart des cas, pour étudier la loi d'une variable aléatoire, il s'avère plus pratique de travailler avec la densité discrète, qui caractérise la loi.

Exemple 3.9. Comme on l'a vu plus haut, l'exemple le plus simple de variable aléatoire est donné par une constante : $X(\omega) = c$ pour tout $\omega \in \Omega$, où $c \in E$ est un élément fixé. Dans ce cas $X(\Omega) = \{c\}$ et $p_X(c) = \mathbb{P}(X = c) = 1$. La densité discrète de X est donc donnée par $p_X(c) = 1$ et $p_X(x) = 0$ si $x \neq c$. Une telle variable aléatoire est dite *presque sûrement constante*.

Exemple 3.10. 120 étudiants sont divisés en 3 groupes, appelés A, B et C, de respectivement 36, 40 et 44 étudiants. On considère les expériences aléatoires suivantes, où "au hasard" signifie "uniformément parmi toutes les possibilités".

- On choisit un groupe au hasard, et on note X le nombre d'étudiant du groupe choisi.
- On choisit un étudiant au hasard, et on note Y le nombre d'étudiants dans le groupe de l'étudiant choisi.

Déterminons les lois de X et Y , ou plutôt leurs densités discrètes. Tout d'abord, notons que $X(\Omega) = Y(\Omega) = \{36, 40, 44\}$.

Comme la variable aléatoire X est déterminée en choisissant un groupe au hasard parmi les 3 possibles, il est clair que $p_X(36) = p_X(40) = p_X(44) = \frac{1}{3}$.

La densité discrète de Y est différente. En effet, l'événement $\{Y = 36\}$ correspond à l'événement "on choisit un étudiant du groupe A" et dans ce cas l'étudiant (pas le groupe) est choisi uniformément parmi les 120 possibles. On a donc $p_Y(36) = \frac{36}{120} = \frac{3}{10}$, $p_Y(40) = \frac{40}{120} = \frac{1}{3}$, $p_Y(44) = \frac{44}{120} = \frac{11}{30}$.

3.2 Indépendance de variables aléatoires

3.2.1 Loi jointe et lois marginales

Dans de nombreux cas, on s'intéresse à deux variables aléatoires (ou plus), qui sont issues d'une même expérience aléatoire. Par exemple, en lançant deux dés, on peut s'intéresser au nombre X obtenu avec le premier dé, mais aussi à la somme Y des nombres obtenus avec les deux dés. Cela signifie que, sur le même espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$ on définit deux variables aléatoires $X : \Omega \rightarrow E$ et $Y : \Omega \rightarrow F$ (pas nécessairement à valeurs dans le même espace) : on peut alors considérer le couple $Z := (X, Y)$, qui est une variable aléatoire définie sur Ω , à valeurs dans l'espace produit $E \times F$. Aux variables aléatoires $X : \Omega \rightarrow E$ et $Y : \Omega \rightarrow F$ sont associées les lois respectives μ_X et μ_Y , qui sont des probabilités sur les ensembles E et F . À la variable aléatoire $Z = (X, Y) : \Omega \rightarrow E \times F$ on associe aussi sa loi μ_Z , que nous noterons $\mu_{X,Y}$, qui est une probabilité sur $E \times F$.

La loi $\mu_{X,Y}$ est appelée *loi jointe* des variables aléatoires X et Y , et les lois μ_X et μ_Y *lois marginales*. Les densités discrètes p_X et p_Y des variables aléatoires X et Y , dites *densités (discrètes) marginales*, peuvent se déduire de la densité discrète $p_{X,Y}$ de la variable aléatoire (X, Y) , dite *densité (discrète) jointe* (voir la Proposition 3.11 ci-dessous).

Pour $x \in E$ et $y \in F$ on rappelle que

$$\begin{aligned} p_X(x) &= \mathbb{P}(X = x), & p_Y(y) &= \mathbb{P}(Y = y), \\ p_{X,Y}(x, y) &= \mathbb{P}((X, Y) = (x, y)) = \mathbb{P}(X = x, Y = y), \end{aligned}$$

où dans la dernière probabilité, la virgule désigne l'intersection d'événements, c'est-à-dire que l'on utilise la notation $\{X = x, Y = y\} := \{X = x\} \cap \{Y = y\}$.

Proposition 3.11. Soient $X : \Omega \rightarrow E$ et $Y : \Omega \rightarrow F$ deux variables aléatoires définies sur le même espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$. Les densités marginales p_X, p_Y vérifient

$$\begin{aligned} \forall x \in E : \quad p_X(x) &= \sum_{y \in F} p_{X,Y}(x, y); \\ \forall y \in F : \quad p_Y(y) &= \sum_{x \in E} p_{X,Y}(x, y). \end{aligned} \tag{3.4}$$

Démonstration. On observe que l'espace Ω peut s'écrire comme une union dénombrable d'événements disjoints, sous la forme $\Omega = \bigcup_{y \in Y(\Omega)} \{Y = y\}$ (pourquoi?). Comme on a $\{X = x\} = \{X = x\} \cap \Omega$, on obtient que

$$\{X = x\} = \bigcup_{y \in Y(\Omega)} \{X = x, Y = y\}, \quad \forall x \in E,$$

où l'union est disjointe et dénombrable. Ainsi, par σ -additivité,

$$\begin{aligned} p_X(x) &= \mathbb{P}(X = x) = \sum_{y \in Y(\Omega)} \mathbb{P}(X = x, Y = y) = \sum_{y \in Y(\Omega)} \mathbb{P}((X, Y) = (x, y)) \\ &= \sum_{y \in Y(\Omega)} p_{X,Y}(x, y) = \sum_{y \in F} p_{X,Y}(x, y), \end{aligned}$$

où, pour la dernière égalité, on a utilisé le fait que si $y \notin Y(\Omega)$, alors l'événement $\{(X, Y) = (x, y)\} \subseteq \{Y = y\} = \emptyset$ est vide et a donc une probabilité nulle. \square

Ce résultat montre que les lois marginales des deux variables aléatoires peuvent être déterminées à partir de la loi jointe. La réciproque est fautive, comme le montre l'exemple suivant. Intuitivement, la loi jointe décrit le comportement collectif des variables aléatoires, et contient *plus d'information* que les lois marginales, qui ne décrivent que les comportements individuels.

Exemple 3.12. Une urne contient n boules, numérotées de 1 à n .

I. On tire deux boules *avec* remise, et on note X_1 et X_2 les numéros des boules tirées : le vecteur aléatoire (X_1, X_2) est à valeurs dans $\{1, \dots, n\}^2$. On laisse en exercice le soin de montrer que la densité discrète jointe est $p_{X_1, X_2}(x_1, x_2) = \frac{1}{n^2}$ pour tout couple $(x_1, x_2) \in \{1, \dots, n\}^2$; et d'en déduire que les densités marginales vérifient $p_{X_1}(x) = p_{X_2}(x) = \frac{1}{n}$ pour tout $x \in \{1, \dots, n\}$.

II. On tire deux boules *sans* remise, et on note X'_1 et X'_2 les numéros des boules tirées : le vecteur aléatoire (X_1, X_2) est aussi à valeurs dans $\{1, \dots, n\}^2$. On laisse en exercice le soin de montrer que la densité discrète jointe est $p_{X'_1, X'_2}(x_1, x_2) = \frac{1}{n(n-1)}$ pour tout couple $(x_1, x_2) \in \{1, \dots, n\}^2$ tel que $x_1 \neq x_2$. On peut en déduire que les densités marginales vérifient $p_{X'_1}(x) = p_{X'_2}(x) = \frac{1}{n}$ pour tout $x \in \{1, \dots, n\}$.

En conclusion, X_1 a la même loi que X'_1 , X_2 a la même loi que X'_2 , mais (X_1, X_2) n'a pas la même loi que (X'_1, X'_2) .

La Proposition 3.11 s'étend sans difficulté au cas de plus de deux variables aléatoires. Plus précisément, soient X_1, \dots, X_n des variables aléatoires définies sur le même espace probabilisé, à valeurs respectivement dans les ensembles E_1, \dots, E_n . Le n -uplet (X_1, \dots, X_n) est alors une variable aléatoire à valeurs dans l'espace produit $E_1 \times \dots \times E_n$. La densité (discrète) jointe p_{X_1, \dots, X_n} (c'est-à-dire la densité discrète de la variable aléatoire (X_1, \dots, X_n)) est reliée aux densités discrètes marginales p_{X_i} par la relation suivante :

$$p_{X_i}(x_i) = \sum_{j \neq i} \sum_{x_j \in E_j} p_{X_1, \dots, X_n}(x_1, \dots, x_n), \quad \text{pour tout } x_i \in E_i,$$

en complète analogie avec la relation (3.4).

3.2.2 Indépendance de variables aléatoires

Dans le chapitre précédent on a défini l'indépendance d'événements. Commençons par définir la notion d'indépendance pour une famille finie de variables aléatoires.

Définition 3.13 (Indépendance de n variables aléatoires). Soient X_1, X_2, \dots, X_n des variables aléatoires, définies sur le même espace probabilisé, et à valeurs respectivement dans les ensembles E_1, E_2, \dots, E_n . Elles sont dites indépendantes si pour tout choix de sous-ensembles $A_1 \subseteq E_1, A_2 \subseteq E_2, \dots, A_n \subseteq E_n$ on a

$$\mathbb{P}(X_1 \in A_1, X_2 \in A_2, \dots, X_n \in A_n) = \prod_{i=1}^n \mathbb{P}(X_i \in A_i). \quad (3.5)$$

Dans le cas de deux variables aléatoires $X : \Omega \rightarrow E$ et $Y : \Omega \rightarrow F$, la relation (3.5) s'écrit

$$\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B), \quad \forall A \subseteq E, \forall B \subseteq F. \quad (3.6)$$

Au vu de la Définition 2.27 de l'indépendance d'une famille d'événements, il peut sembler étrange que la relation (3.5) ne soit pas imposée aussi à toute sous-famille d'indices $\{i_1, i_2, \dots, i_k\} \subset \{1, 2, \dots, n\}$. La raison est que cela n'est pas nécessaire : en effet, en choisissant $A_j = E_j$ pour $j \notin \{i_1, i_2, \dots, i_k\}$, on a que $\{X_j \in A_j\} = \{X_j \in E_j\} = \Omega$ pour ces j , et ainsi on obtient de (3.5)

$$\begin{aligned} \mathbb{P}(X_{i_1} \in A_{i_1}, X_{i_2} \in A_{i_2}, \dots, X_{i_k} \in A_{i_k}) &= \mathbb{P}(X_1 \in A_1, X_2 \in A_2, \dots, X_n \in A_n) \\ &= \prod_{i=1}^n \mathbb{P}(X_i \in A_i) = \prod_{j=1}^k \mathbb{P}(X_{i_j} \in A_{i_j}). \end{aligned}$$

En particulier, si X_1, X_2, \dots, X_n sont des variables aléatoires indépendantes, alors $X_{i_1}, X_{i_2}, \dots, X_{i_k}$ le sont aussi, pour tout choix de $\{i_1, i_2, \dots, i_k\} \subseteq \{1, 2, \dots, n\}$.

Définition 3.14 (Indépendance de variables aléatoires). Soient $(X_i)_{i \in I}$ des variables aléatoires définies sur le même espace probabilisé, où l'ensemble des indices I est arbitraire. Ces variables aléatoires sont dites indépendantes si pour tout sous-ensemble $J \subseteq I$ fini ($|J| < \infty$), les variables aléatoires $(X_j)_{j \in J}$ sont indépendantes.

Montrons maintenant comment l'indépendance de variables aléatoires peut être caractérisée en terme de leur loi jointe. Pour simplifier les notations, nous ne démontrons la proposition suivante que dans le cas de deux variables aléatoires, mais le résultat s'étend facilement au cas de n variables aléatoires.

Proposition 3.15. Soient $X : \Omega \rightarrow E$ et $Y : \Omega \rightarrow F$ des variables aléatoires définies sur le même espace probabilisé, dont les densités marginales sont notées p_X, p_Y et la densité jointe $p_{X,Y}$. Ces variables aléatoires sont indépendantes si et seulement si on a la relation suivante :

$$p_{X,Y}(x, y) = p_X(x)p_Y(y), \quad \forall x \in E, y \in F. \quad (3.7)$$

Démonstration. Si X et Y sont indépendantes, alors pour tout $x \in E$ et $y \in F$, en appliquant la définition de l'indépendance avec $A = \{x\}$ et $B = \{y\}$, on obtient

$$p_{X,Y}(x, y) = \mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x)\mathbb{P}(Y = y) = p_X(x)p_Y(y).$$

Réciproquement, supposons que l'on ait la relation (3.7). Pour tout $A \subset E$, $B \subset F$, on a $\mathbb{P}(X \in A, Y \in B) = \mathbb{P}((X, Y) \in A \times B)$, de sorte qu'en utilisant la densité discrète du vecteur aléatoire (X, Y) on a

$$\begin{aligned} \mathbb{P}(X \in A, Y \in B) &= \sum_{x \in A, y \in B} p_{X,Y}(x, y) = \sum_{x \in A, y \in B} p_X(x) p_Y(y) \\ &= \left(\sum_{x \in A} p_X(x) \right) \left(\sum_{y \in B} p_Y(y) \right) = \mathbb{P}(X \in A) \mathbb{P}(Y \in B). \end{aligned}$$

On a donc obtenu la relation (3.6), qui caractérise l'indépendance de X et Y . \square

La Proposition 3.15 s'étend au cas de n variables aléatoires X_1, X_2, \dots, X_n , définies dans le même espace probabilisé, à valeurs respectivement dans les ensembles E_1, E_2, \dots, E_n . En notant p_{X_1, \dots, X_n} leur densité jointe, et p_{X_i} les densités marginales, ces variables aléatoires sont indépendantes si et seulement si

$$p_{X_1, \dots, X_n}(x_1, \dots, x_n) = \prod_{i=1}^n p_{X_i}(x_i), \quad \forall (x_1, \dots, x_n) \in E_1 \times \dots \times E_n.$$

Remarque 3.16. Si X, Y sont deux variables aléatoires définies sur le même espace probabilisé, à valeurs respectivement dans les ensembles E et F , dont la densité jointe se factorise sous la forme $p_{X,Y}(x, y) = \alpha(x)\beta(y)$ pour tous $x \in E$, $y \in F$ (où $\alpha : E \rightarrow \mathbb{R}^+$ et $\beta : F \rightarrow \mathbb{R}^+$ sont des fonction positives arbitraires), alors X et Y sont indépendantes.

En effet, grâce à la Proposition 3.11, on a $p_X(x) = \sum_{y \in F} p_{X,Y}(x, y) = B\alpha(x)$, où $B := \sum_{y \in F} \beta(y) \in [0, +\infty]$. De même, $p_Y(y) = A\beta(y)$, avec $A = \sum_{x \in E} \alpha(x) \in [0, +\infty]$. De plus, en sommant à la fois sur x et y , on obtient $1 = \sum_{x \in E, y \in F} p_{X,Y}(x, y) = AB$: cela montre que $0 < A < \infty$, $0 < B < \infty$, et ainsi $p_{X,Y}(x, y) = \alpha(x)\beta(y) = p_X(x)p_Y(y)$ pour tous $x \in E$, $y \in F$. On en conclut que X et Y sont indépendantes d'après la Proposition 3.15.

Montrons maintenant une propriété, plutôt intuitive, selon laquelle, si on divise une famille de variables aléatoires indépendantes en sous-familles disjointes (des "paquets"), on obtient des variables aléatoires indépendantes. Pour simplifier les notations, nous n'énonçons le résultat que dans le cas de deux paquets.

Proposition 3.17 (Indépendance par paquets). *Soient X_1, X_2, \dots, X_n des variables aléatoires indépendantes et soient $I = \{i_1, \dots, i_\ell\}$, $J = \{j_1, \dots, j_k\}$ des sous-ensembles non vides et disjoints de $\{1, 2, \dots, n\}$. Alors les variables aléatoires*

$$X_I := (X_{i_1}, \dots, X_{i_\ell}) \quad \text{et} \quad X_J := (X_{j_1}, \dots, X_{j_k})$$

sont indépendantes.

Démonstration. Observons que les variables aléatoires $(X_h)_{h \in I \cup J}$ sont indépendantes, ainsi que les variables aléatoires $(X_i)_{i \in I}$ et les variables aléatoires $(X_j)_{j \in J}$. Par conséquent, en notant $x_I := (x_{i_1}, \dots, x_{i_\ell})$ et $x_J := (x_{j_1}, \dots, x_{j_k})$, on a grâce à la Proposition 3.11 :

$$\begin{aligned} p_{X_I, X_J}(x_I, x_J) &= p_{X_{i_1}, \dots, X_{i_\ell}, X_{j_1}, \dots, X_{j_k}}(x_{i_1}, \dots, x_{i_\ell}, x_{j_1}, \dots, x_{j_k}) \\ &= \prod_{r=1}^{\ell} \prod_{s=1}^k p_{X_{i_r}}(x_{i_r}) p_{X_{j_s}}(x_{j_s}) = \left(\prod_{r=1}^{\ell} p_{X_{i_r}}(x_{i_r}) \right) \left(\prod_{s=1}^k p_{X_{j_s}}(x_{j_s}) \right) \\ &= p_{X_I}(x_I) p_{X_J}(x_J), \end{aligned}$$

pour tous x_I et x_J . L'indépendance de X_I et X_J en découle, grâce à la Proposition 3.15. \square

La proposition qui suit établit que des fonctions de variables aléatoires indépendantes sont indépendantes (le résultat s'étend facilement à plus de deux variables aléatoires). Si $X : \Omega \rightarrow E$ est une variable aléatoire et $f : E \rightarrow H$ une fonction, on notera $f(X)$ leur composition $f \circ X : \Omega \rightarrow H$, qui est aussi une variable aléatoire (exercice).

Proposition 3.18 (Indépendance par transformation). *Soient X et Y des variables aléatoires indépendantes, à valeurs respectivement dans les ensembles E et F . De plus, soient H, K deux ensembles et $f : E \rightarrow H$, $g : F \rightarrow K$ des fonctions arbitraires. Alors les variables aléatoires $f(X)$ et $g(Y)$ sont aussi indépendantes.*

Démonstration. Il suffit d'observer que, si $A \subseteq H$, $B \subseteq K$,

$$\begin{aligned} \mathbb{P}(f(X) \in A, g(Y) \in B) &= \mathbb{P}(X \in f^{-1}(A), Y \in g^{-1}(B)) \\ &= \mathbb{P}(X \in f^{-1}(A))\mathbb{P}(Y \in g^{-1}(B)) = \mathbb{P}(f(X) \in A)\mathbb{P}(g(Y) \in B), \end{aligned}$$

donc $f(X)$ et $g(Y)$ sont indépendantes. \square

Voyons maintenant un exemple d'application combinée des Propositions 3.17 et 3.18.

Corollaire 3.19. *Soient $X_1, \dots, X_n, X_{n+1}, \dots, X_{n+m}$ des variables aléatoires réelles indépendantes, avec $n, m \geq 1$. Alors $X_1 + \dots + X_n$ et $X_{n+1} + \dots + X_{n+m}$ sont des variables aléatoires indépendantes.*

Démonstration. En utilisant la Proposition 3.17 avec $I = \{1, \dots, n\}$, $J = \{n+1, \dots, n+m\}$, on obtient que $X_I = (X_1, \dots, X_n)$ et $X_J = (X_{n+1}, \dots, X_{n+m})$ sont des variables aléatoires indépendantes. En notant $f_k : \mathbb{R}^k \rightarrow \mathbb{R}$ la fonction somme, définie par $f_k(x_1, \dots, x_k) := x_1 + \dots + x_k$, on peut alors écrire $X_1 + \dots + X_n = f_n(X_I)$ et $X_{n+1} + \dots + X_{n+m} = f_m(X_J)$. Le conclusion découle donc de la Proposition 3.18. \square

3.3 Exemples essentiels de variables aléatoires discrètes

3.3.1 Loi uniforme discrète

Soit E un ensemble arbitraire fini (un cas typique est celui où $E = \{1, 2, \dots, n\}$, avec $n \in \mathbb{N}_*$). Une variable aléatoire qui prend avec probabilité égale toutes les valeurs $x \in E$ est appelée *variable aléatoire uniforme discrète à valeurs dans E* , et on écrira $X \sim \text{Unif}(E)$. Une telle variable aléatoire a pour densité discrète

$$p_X(x) = \frac{1}{|E|}, \quad \forall x \in E,$$

et sa loi est donc donnée par $\mu_X(A) = \sum_{x \in A} p_X(x) = \frac{|A|}{|E|}$ pour tout $A \subseteq E$. Dit autrement, $X \sim \text{Unif}(E)$ si et seulement si la loi μ_X de X est la probabilité uniforme sur E .

3.3.2 Loi de Bernoulli

Une variable aléatoire X est dite *de Bernoulli* si $X(\Omega) = \{0, 1\}$. La loi de X est complètement déterminée par le paramètre $p := p_X(1) \in [0, 1]$, car $p_X(0) = 1 - p$, et $p_X(x) = 0$ si $x \notin \{0, 1\}$. Dans ce cas X est appelée variable aléatoire de Bernoulli de paramètre p , et on écrira $X \sim \text{Bern}(p)$.

3.3.3 Loi binomiale

Pour $n \in \mathbb{N}_*$ et $p \in [0, 1]$ donnés, on dira que la variable aléatoire X suit une loi binomiale de paramètres n et p , et on écrira $X \sim \text{Bin}(n, p)$, si X est à valeurs dans $\{1, \dots, n\}$ et a pour densité discrète

$$p_X(k) := \binom{n}{k} p^k (1-p)^{n-k} \quad \text{pour tout } k \in \{0, \dots, n\}. \quad (3.8)$$

Si on considère n épreuves indépendantes de probabilité de succès p , on a vu dans l'Exemple 2.28 que la probabilité qu'on ait exactement k succès est donnée par l'expression (3.8) : on peut donc interpréter X comme le nombre de succès parmi ces n épreuves. On a ainsi le résultat suivant.

Soient X_1, \dots, X_n des variables aléatoires $\text{Bern}(p)$ *indépendantes*. Alors $X_1 + \dots + X_n$ suit la loi $\text{Bin}(n, p)$.

Notons que si l'on pose $A_i = \{X_i = 1\}$, on peut interpréter A_i comme l'événement que la i -ème épreuve soit un succès (X_i vaut 1 si on a un succès et 0 si on a un échec) : les événements $(A_i)_{1 \leq i \leq n}$ sont indépendants et de probabilité $\mathbb{P}(A_i) = p$. Ainsi $X_1 + \dots + X_n$ compte le nombre de succès parmi n épreuves indépendantes de probabilité de succès p .

Exercice 3.20. Soient $X \sim \text{Bin}(n, p)$ et $Y \sim \text{Bin}(m, p)$ des variables aléatoires *indépendantes*. Montrer que $X + Y \sim \text{Bin}(n + m, p)$.

Indication : prendre $(X_i)_{1 \leq i \leq n+m}$ des variables aléatoires $\text{Bern}(p)$ indépendantes, et considérer $X' = \sum_{i=1}^n X_i, Y' = \sum_{i=n+1}^{n+m} X_i$; noter que le vecteur aléatoire (X', Y') a la même loi que (X, Y) ...

3.3.4 Loi de Poisson

Pour $\lambda \in]0, \infty[$ donné, on dira qu'une variable aléatoire X est de *Poisson* de paramètre λ et on notera $X \sim \text{Poi}(\lambda)$, si X est à valeurs dans \mathbb{N} , et est de densité discrète donnée par

$$p_X(k) = e^{-\lambda} \frac{\lambda^k}{k!} \quad \text{pour tout } k \geq 0. \quad (3.9)$$

Ces variables aléatoires apparaissent naturellement lorsque l'on considère des variables aléatoires de loi $\text{Bin}(n, p)$ où le nombre de tentatives n est très grand, mais la probabilité de succès est très petite, proportionnelle à $1/n$.

Proposition 3.21. Soit $(p_n)_{n \geq 1}$ une suite à valeurs dans $]0, 1[$. On suppose que $p_n \sim \frac{\lambda}{n}$ quand $n \rightarrow \infty$, où $\lambda \in]0, \infty[$. Alors, si $X_n \sim \text{Bin}(n, p_n)$, on a

$$\lim_{n \rightarrow \infty} \mathbb{P}(X_n = k) = \mathbb{P}(X = k), \quad \text{pour tout } k \geq 0,$$

où $X \sim \text{Poi}(\lambda)$.

Démonstration. Pour simplifier les notations, supposons que $p_n = \frac{\lambda}{n}$, avec $\lambda > 0$ (on laisse le lecteur adapter la démonstration). Soit $k \geq 0$ fixé. Pour $n \geq k$, on a

$$\begin{aligned} \mathbb{P}(X_n = k) &= \binom{n}{k} p_n^k (1 - p_n)^{n-k} = \frac{n!}{k!(n-k)!} \frac{\lambda^k}{n^k} \left(1 - \frac{\lambda}{n}\right)^{n-k} \\ &= \frac{\lambda^k}{k!} \frac{n(n-1)\cdots(n-k+1)}{n^k} \frac{1}{\left(1 - \frac{\lambda}{n}\right)^k} \left(1 - \frac{\lambda}{n}\right)^n. \end{aligned}$$

On observe que, k étant fixé, $n(n-1)\cdots(n-k+1)$ est un polynôme de degré k en n , dont le terme dominant quand $n \rightarrow \infty$ est n^k , et donc $\lim_{n \rightarrow +\infty} \frac{n(n-1)\cdots(n-k+1)}{n^k} = 1$. On a aussi, toujours pour k fixé, $\lim_{n \rightarrow +\infty} \left(1 - \frac{\lambda}{n}\right)^k = 1$ et $\lim_{n \rightarrow +\infty} \left(1 - \frac{\lambda}{n}\right)^n = e^{-\lambda}$. On en déduit que $\lim_{n \rightarrow +\infty} \mathbb{P}(X_n = k) = e^{-\lambda} \frac{\lambda^k}{k!}$, qui est ce que l'on voulait. \square

De la même manière que pour les variables aléatoires binomiales, la somme de deux variables aléatoires de Poisson *indépendantes* est encore une variable de Poisson.

Proposition 3.22. Soient $X \sim \text{Poi}(\lambda)$ et $Y \sim \text{Poi}(\mu)$ des variables aléatoires indépendantes. Alors $X + Y \sim \text{Poi}(\lambda + \mu)$.

Démonstration. Soit $n \geq 0$, et calculons $\mathbb{P}(X+Y = n)$. L'idée est de décomposer l'événement $\{X + Y = n\}$ suivant la valeur de X : on écrit $\{X + Y = n\} = \bigcup_{k=0}^n \{X = k, Y = n - k\}$. L'union étant disjointe et les variables aléatoires X et Y indépendantes, on a

$$\begin{aligned} \mathbb{P}(X + Y = n) &= \sum_{k=0}^n \mathbb{P}(X = k, Y = n - k) = \sum_{k=0}^n \mathbb{P}(X = k) \mathbb{P}(Y = n - k) \\ &= \sum_{k=0}^n e^{-\lambda} \frac{\lambda^k}{k!} e^{-\mu} \frac{\mu^{n-k}}{(n-k)!} \\ &= e^{-(\lambda+\mu)} \frac{1}{n!} \sum_{k=0}^n \binom{n}{k} \lambda^k \mu^{n-k} = e^{-(\lambda+\mu)} \frac{(\lambda + \mu)^n}{n!}, \end{aligned}$$

où dans la dernière égalité on a appliqué la formule du binôme de Newton. \square

3.3.5 Loi géométrique

Pour $p \in]0, 1[$, on dira que la variable aléatoire réelle X suit la loi *géométrique de paramètre* p , et on écrira $X \sim \text{Géom}(p)$, si X est à valeurs dans \mathbb{N}_* et a pour densité discrète

$$p_X(k) := p(1-p)^{k-1} \quad \text{pour tout } k \geq 1. \quad (3.10)$$

Si $(X_n)_{n \geq 1}$ est une suite de variables aléatoires indépendantes $\text{Bern}(p)$, représentant une suite infinie d'épreuves indépendantes ($X_n = 1$ représente un succès à la n -ème épreuve, et $X_n = 0$ un échec), alors l'instant de premier succès suit une loi géométrique. En effet, l'instant de premier succès est donné par la variable aléatoire $T : \Omega \rightarrow \mathbb{N}_* \cup \{+\infty\}$, définie par $T(\omega) := \min\{n \in \mathbb{N} : X_n(\omega) = 1\}$, avec la convention que $\min \emptyset := +\infty$. On remarque alors que, pour $k \geq 1$, on a l'égalité d'événements $\{T = k\} = \{X_1 = 0, \dots, X_{k-1} = 0, X_k = 1\}$: par indépendance des variables $(X_n)_{n \in \mathbb{N}}$, on a donc bien $\mathbb{P}(T = k) = (1-p)^{k-1}p$. Soulignons que $\mathbb{P}(T < +\infty) = \lim_{n \rightarrow \infty} \mathbb{P}(T \leq n) = \sum_{k=1}^{\infty} \mathbb{P}(T = k) = 1$ et donc que $\mathbb{P}(T = +\infty) = 0$.

Remarque 3.23. Dans d'autres contextes il peut être utile de considérer des variables aléatoires à valeurs dans \mathbb{N} (plutôt que \mathbb{N}_*), dont la densité discrète est donnée par $p_X(k) = (1-p)^k p$ pour tout $k \geq 0$. Cela correspond par exemple à la situation où plutôt que de considérer l'instant de premier succès, on considère le nombre d'échecs avant le premier succès. Il s'agit aussi de variables aléatoires appelées *géométriques*, dont on notera la loi $\text{Géom}_0(p)$. Il faudra donc faire attention de savoir quelle variable géométrique on manipule !

Calculons la probabilité qu'une variable géométrique prenne une valeur strictement plus grande qu'un entier n donné :

$$\mathbb{P}(X > n) = \sum_{k=n+1}^{+\infty} \mathbb{P}(X = k) = \sum_{k=n+1}^{+\infty} p(1-p)^{k-1} = (1-p)^n \sum_{\ell=0}^{+\infty} p(1-p)^\ell = (1-p)^n$$

Cela va nous permettre de montrer que les variables aléatoires géométriques possèdent une propriété importante, dite d'*absence de mémoire*.

Proposition 3.24. Soit $X \sim \text{Géom}(p)$. Alors pour tout $n, m \geq 0$

$$\mathbb{P}(X > n + m \mid X > n) = \mathbb{P}(X > m). \quad (3.11)$$

Démonstration. Tout d'abord, comme $\{X > n + m\} \subseteq \{X > n\}$, on a

$$\mathbb{P}(X > n + m \mid X > n) = \frac{\mathbb{P}(X > n + m, X > n)}{\mathbb{P}(X > n)} = \frac{\mathbb{P}(X > n + m)}{\mathbb{P}(X > n)}.$$

On obtient $\mathbb{P}(X > n + m \mid X > n) = \frac{(1-p)^{n+m}}{(1-p)^n} = (1-p)^m = \mathbb{P}(X > m)$, comme voulu. \square

Exercice 3.25. Soit X une variable aléatoire à valeurs dans \mathbb{N}_* , telle que la propriété (3.11) est vérifiée. Alors $X \sim \text{Géom}(p)$, avec $p := \mathbb{P}(X = 1)$.

Deux exemples liés à la marche aléatoire

★ La ruine du joueur

On a ici les outils pour étudier un modèle aléatoire important, connu sous le nom de *ruine du joueur*. Un joueur arrive à un casino avec x euros en poche, et espère gagner n euros pour s'acheter un téléphone hors de prix ($n \geq x$). Il parie successivement 1 euro : à chaque tour, soit il gagne 1 euro, avec probabilité $\frac{1}{2}$, soit il perd 1 euro, avec probabilité $\frac{1}{2}$. Le jeu continue jusqu'à ce que le joueur atteigne son objectif de n euros ou bien se retrouve avec 0 euro et ne puisse plus jouer. Quel est la probabilité $p_n(x)$ que, partant de x euros, le joueur réussisse à gagner n euros avant d'être ruiné ?

Formalisons un peu ce problème. On note $(X_i)_{i \geq 1}$ les gains successifs du joueur : les variables aléatoires X_i sont indépendantes, à valeur dans $\{-1, 1\}$, et de même loi donnée par $\mathbb{P}(X_i = 1) = \mathbb{P}(X_i = -1) = \frac{1}{2}$. Si on note $S_k := \sum_{i=1}^k X_i$ (par convention $S_0 = 0$), alors le gain (algébrique) du joueur après k tours est $x + S_k$, si le jeu n'était jamais interrompu. La probabilité que l'on cherche à calculer est donc, pour $0 \leq x \leq n$,

$$p_n(x) = \mathbb{P}(\exists k \geq 0 \text{ tel que } x + S_k = n ; x + S_j > 0 \text{ pour tout } 0 \leq j \leq k). \quad (3.12)$$

Notons tout d'abord que l'on a clairement $p_n(0) = 0$ et $p_n(n) = 1$. Si $1 \leq x \leq n-1$, alors on peut prendre $k \geq 1$ et $j \geq 1$ dans la probabilité ci-dessus :

$$\begin{aligned} p_n(x) &= \mathbb{P}(\exists k \geq 1, x + S_k = n \text{ et } x + S_j > 0 \forall 1 \leq j \leq k) \\ &= \mathbb{P}(\exists k \geq 1, x + X_1 + \tilde{S}_{k-1} = n \text{ et } x + X_1 + \tilde{S}_{j-1} > 0 \forall 1 \leq j \leq k) \end{aligned}$$

où on a noté $\tilde{S}_{j-1} := \sum_{i=2}^j X_i = X_2 + X_3 + \dots + X_j$ pour $j \geq 1$. Ainsi, en décomposant suivant la valeur de X_1 , on obtient

$$\begin{aligned} p_n(x) &= \mathbb{P}(X_1 = 1) \mathbb{P}(\exists k \geq 1, x + 1 + \tilde{S}_{k-1} = n \text{ et } x + 1 + \tilde{S}_{j-1} > 0 \forall 1 \leq j \leq k \mid X_1 = 1) \\ &\quad + \mathbb{P}(X_1 = -1) \mathbb{P}(\exists k \geq 1, x - 1 + \tilde{S}_{k-1} = n \text{ et } x - 1 + \tilde{S}_{j-1} > 0 \forall 1 \leq j \leq k \mid X_1 = -1) \end{aligned}$$

Comme X_1 est indépendant de $(\tilde{S}_j)_{j \geq 1}$ (par indépendance par paquets), on a que les événements $\{\exists k \geq 1, x + 1 + \tilde{S}_{k-1} = n \text{ et } x + 1 + \tilde{S}_{j-1} > 0 \forall 1 \leq j \leq k\}$ et $\{X_1 = 1\}$ sont indépendants, et on peut enlever le conditionnement par $X_1 = 1$ (on rappelle que $\mathbb{P}(A|B) = \mathbb{P}(A)$ si A et B sont indépendants), et de même pour le conditionnement par $X_1 = -1$. On en déduit que

$$\begin{aligned} p_n(x) &= \frac{1}{2} \mathbb{P}(\exists k \geq 1, x + 1 + \tilde{S}_{k-1} = n \text{ et } x + 1 + \tilde{S}_{j-1} > 0 \forall 1 \leq j \leq k) \\ &\quad + \frac{1}{2} \mathbb{P}(\exists k \geq 1, x - 1 + \tilde{S}_{k-1} = n \text{ et } x - 1 + \tilde{S}_{j-1} > 0 \forall 1 \leq j \leq k) \\ &= \frac{1}{2} \mathbb{P}(\exists k \geq 0, x + 1 + S_k = n \text{ et } x + 1 + S_j > 0 \forall 0 \leq j \leq k-1) \\ &\quad + \frac{1}{2} \mathbb{P}(\exists k \geq 0, x - 1 + S_k = n \text{ et } x - 1 + S_j > 0 \forall 0 \leq j \leq k-1), \end{aligned}$$

où on a utilisé le fait que $(\tilde{S}_{n-1})_{n \geq 1}$ a la même loi que $(S_n)_{n \geq 0}$. Au final, en utilisant la définition (3.12) (noter que la condition $x + 1 + S_j > 0$, resp. $x - 1 + S_j > 0$, est automatiquement vérifiée pour $j = k$), on obtient la relation

$$p_n(x) = \frac{1}{2} p_n(x+1) + \frac{1}{2} p_n(x-1) \quad \text{pour tout } 1 \leq x \leq n-1. \quad (3.13)$$

Cela va nous permettre calculer $p_n(x)$. En effet, la relation (3.13) implique que $p_n(x+1) - p_n(x) = p_n(x) - p_n(x-1)$ * pour tout $1 \leq x \leq n-1$: en itérant cette relation, on obtient $p_n(x+1) - p_n(x) = p_n(1) - p_n(0) = p_n(1)$. On en conclut, en utilisant de nouveau que $p_n(0) = 0$, que

$$p_n(x) = \sum_{y=0}^{x-1} (p_n(y+1) - p_n(y)) = x \times p_n(1).$$

En prenant $x = n$, on trouve donc $1 = p_n(n) = n p_n(1)$, c'est-à-dire $p_n(1) = \frac{1}{n}$. On en conclut que $p_n(x) = x p_n(1) = \frac{x}{n}$ pour tout $0 \leq x \leq n$.

★ Marche aléatoire simple sur \mathbb{Z}^d : transience vs. récurrence

Soit $(X_i)_{i \geq 1}$ une suite de variables aléatoires indépendantes, à valeurs dans $\{-1, 1\}^d$: on suppose que $X_i = (X_i^{(1)}, \dots, X_i^{(d)})$ où les $(X_i^{(j)})_{1 \leq k \leq d}$ sont indépendantes uniformes dans $\{-1, 1\}$, c'est-à-dire $\mathbb{P}(X_i^{(j)} = +1) = \mathbb{P}(X_i^{(j)} = -1) = \frac{1}{2}$. Les X_i sont interprétés comme les pas (aléatoires) d'un marcheur dans \mathbb{Z}^d , et on pose $S_0 = 0_d := (0, \dots, 0)$, $S_n := \sum_{i=1}^n X_i \in \mathbb{Z}^d$, qui représente la position du marcheur après n pas : on appelle $(S_n)_{n \in \mathbb{N}}$ la *marche aléatoire sur \mathbb{Z}^d* . On note

*. Indication : écrire $p_n(x) = \frac{1}{2} p_n(x) + \frac{1}{2} p_n(x)$.

$T = \inf\{k \geq 1, S_{2k} = 0_d\}$ le temps de premier retour en 0_d du marcheur (soulignons que le marcheur ne peut pas revenir en 0_d après un nombre impair de pas).

Théorème 3.26. *Si $d = 1, 2$, alors $\mathbb{P}(T < +\infty) = 1$; si $d \geq 3$, alors $\mathbb{P}(T < +\infty) < 1$.*

Ce théorème signifie qu'en dimension $d = 1, 2$, *presque sûrement* le marcheur finit par repasser en l'origine (on dit que la marche est *récurrente*), alors qu'en dimension $d \geq 3$, le marcheur a une probabilité strictement positive de ne jamais revenir en l'origine (on dit que la marche est *transiente*).

Démonstration. On pose $f_k = \mathbb{P}(T = k)$ de sorte que l'on a $\mathbb{P}(T < +\infty) = \sum_{k=1}^{\infty} f_k$, et on note $u_n := \mathbb{P}(S_{2n} = 0_d)$. On a $u_0 = 1$, et pour tout $n \geq 1$, en décomposant suivant la valeur de T , on a la relation

$$\begin{aligned} u_n &= \sum_{k=1}^n \mathbb{P}(T = k, S_{2n} = 0_d) = \sum_{k=1}^n \mathbb{P}(T = k, S_{2n} - S_{2k} = 0_d) \\ &= \sum_{k=1}^n \mathbb{P}(T = k) \mathbb{P}(S_{2n} - S_{2k} = 0_d) = \sum_{k=1}^n f_k u_{n-k}, \end{aligned}$$

où on a utilisé que les événements $\{T = k\}$ et $\{S_{2n} - S_{2k}\}$ sont indépendants (pourquoi?). Si $\sum_{n=1}^{\infty} u_n < +\infty$, alors on peut écrire

$$\sum_{n=1}^{\infty} u_n = \sum_{n=1}^{\infty} \sum_{k=1}^n f_k u_{n-k} = \sum_{k=1}^{\infty} f_k \sum_{n=k}^{\infty} u_{n-k} = \sum_{k=1}^{\infty} f_k \left(1 + \sum_{n=1}^{\infty} u_n\right).$$

D'autre part, si $\sum_{n=1}^{\infty} u_n = +\infty$, on écrit, pour tout $N \geq 1$

$$\sum_{n=1}^N u_n = \sum_{n=1}^N \sum_{k=1}^n f_k u_{n-k} = \sum_{k=1}^N f_k \sum_{n=k}^N u_{n-k} \leq \sum_{k=1}^N f_k \left(1 + \sum_{n=1}^N u_n\right);$$

on obtient $\sum_{k=1}^N f_k \geq \frac{\sum_{n=1}^N u_n}{1 + \sum_{n=1}^N u_n}$, ce qui montre que $\sum_{k=1}^{\infty} f_k \geq 1$ en prenant $N \rightarrow \infty$. On a ainsi montré la relation suivante :

$$\sum_{k=1}^{\infty} f_k = \frac{\sum_{n=1}^{\infty} u_n}{1 + \sum_{n=1}^{\infty} u_n} \begin{cases} < 1 & \text{si } \sum_{n=1}^{\infty} u_n < +\infty, \\ = 1 & \text{si } \sum_{n=1}^{\infty} u_n = +\infty, \end{cases}$$

et il reste à étudier la nature de la série $\sum_{n=1}^{\infty} u_n$. Notons que l'on peut écrire $S_{2n} = (S_{2n}^{(1)}, \dots, S_{2n}^{(d)})$ avec $S_{2n}^{(j)} = \sum_{i=1}^{2n} X_i^{(j)}$: on peut donc calculer explicitement

$$u_n := \mathbb{P}(S_{2n}^{(j)} = 0, \forall 1 \leq j \leq d) = \mathbb{P}(S_{2n}^{(1)} = 0)^d = \left(\frac{1}{2^{2n}} \binom{2n}{n}\right)^d,$$

où l'on a utilisé pour le calcul de la dernière probabilité que $S_{2n}^{(1)} = 0$ si et seulement si exactement n des $X_i^{(1)}$ valent $+1$ et n des $X_i^{(1)}$ valent -1 . En utilisant la formule de Stirling $k! \sim \sqrt{2\pi k} k^k e^{-k}$, on trouve que $u_n \sim (\pi n)^{-d/2}$ quand $n \rightarrow +\infty$: d'après le critère de Riemann, on trouve que $\sum_{n=1}^{\infty} u_n = +\infty$ si $d/2 \leq 1$ (c'est-à-dire $d = 1, 2$) et $\sum_{n=1}^{\infty} u_n < +\infty$ si $d/2 > 1$ (c'est-à-dire $d \geq 3$), ce qui conclut la démonstration. \square

3.4 Espérance, variance et moments

3.4.1 Définition de l'espérance

La notion mathématique d'espérance d'une variable aléatoire X correspond à la notion de barycentre (ou de moyenne pondérée) des valeurs x_i que peut prendre X , pondérées par les poids $p_i = \mathbb{P}(X = x_i)$. Si on répète un grand nombre N de fois une expérience aléatoire, et que l'on note X_1, \dots, X_N les valeurs de la variable X prises lors de ces différentes expériences, on peut réécrire la moyenne $\frac{1}{N}(X_1 + \dots + X_N)$ sous la forme $\frac{1}{N} \sum_x x N_x$, où N_x est le nombre de fois où la variable X_i vaut x (la somme porte sur les valeurs x que peut prendre la variable aléatoire). Comme on a $\frac{N_x}{N} \approx \mathbb{P}(X = x)$, cela suggère que la moyenne des valeurs observées de la variable aléatoire vérifie $\frac{1}{N}(X_1 + \dots + X_N) \approx \sum_x x \mathbb{P}(X = x)$.

Définition 3.27 (Espérance). Soit X une variable aléatoire discrète réelle, à valeurs dans l'ensemble $\mathcal{X} := X(\Omega)$ dénombrable, de densité discrète p_X . On dit que X admet une espérance si la somme $\sum_{x \in \mathcal{X}} x p_X(x)$ est bien définie, c'est-à-dire si au moins une des sommes suivantes est finie $\sum_{x \in \mathcal{X}} x^+ p_X(x) < +\infty$ ou $\sum_{x \in \mathcal{X}} x^- p_X(x) < +\infty$ ^a. Si X admet une espérance, alors on définit l'espérance de X

$$\mathbb{E}(X) := \sum_{x \in \mathcal{X}} x p_X(x) \in [-\infty, +\infty]. \quad (3.14)$$

a. On rappelle que $x^+ = \max(x, 0)$ et $x^- = \max(-x, 0)$ sont les parties positives et négatives de x .

Notons que si X ne prend qu'un nombre fini de valeurs, alors X admet une espérance finie.

Exemple 3.28. Soit X une variable aléatoire réelle presque sûrement constante : il existe un $c \in \mathbb{R}$ tel que l'on ait $p_X(c) = 1$ (X est à valeurs dans $\mathcal{X} = \{c\}$). Ainsi, X admet une espérance, donnée par $\mathbb{E}(X) = c p_X(c) = c$.

Exemple 3.29. Soit $X := \mathbb{1}_A$ la variable aléatoire indicatrice d'un événement $A \in \mathcal{F}$: X est à valeur dans $\mathcal{X} = \{0, 1\}$, et $p_X(1) = \mathbb{P}(A)$ et $p_X(0) = 1 - \mathbb{P}(A)$. Ainsi, X admet une espérance, donnée par $\mathbb{E}(X) = 0 \cdot (1 - \mathbb{P}(A)) + 1 \cdot \mathbb{P}(A) = \mathbb{P}(A)$.

Nous avons vu que $\mathbb{E}(X)$ représente une sorte de "barycentre" des valeurs prises par X , pondérées par leur probabilités respectives. Cependant, $\mathbb{E}(X)$ n'est (souvent) pas "la valeur la plus probable" de X , ni même une valeur que l'on s'attend typiquement à observer.

Exemple 3.30. Lançons un dé régulier à six faces, et notons X le nombre obtenu, à valeurs dans $\mathcal{X} = \{1, 2, 3, 4, 5, 6\}$: $\mathbb{E}(X)$ est bien défini et fini. Comme $p_X(x) = \frac{1}{6}$ pour tout $x \in \mathcal{X}$, on obtient $\mathbb{E}(X) = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + \dots + 6 \cdot \frac{1}{6} = \frac{7}{2}$. Cela donne un exemple où $\mathbb{E}(X) \notin X(\Omega)$.

Exercice 3.31. Reprendre l'Exemple 3.10, et montrer que $\mathbb{E}(X) = 40$, $\mathbb{E}(Y) = \frac{604}{15} \approx 40,3$. Intuitivement, il est normal que $\mathbb{E}(Y) > \mathbb{E}(X)$: en choisissant un étudiant au hasard, il est plus probable d'en choisir un provenant d'un groupe nombreux.

Remarque 3.32. Une observation fondamentale est que toute variable aléatoire positive X admet une espérance (qui peut éventuellement valoir $+\infty$). Maintenant, pour toute variable aléatoire réelle X , on peut considérer les variables aléatoires positives $X^+ := \max\{X, 0\}$, $X^- := \max\{-X, 0\}$ et $|X| = X^+ + X^-$. Les espérances $\mathbb{E}(X^+)$, $\mathbb{E}(X^-)$ sont toujours bien définies, et X admet une espérance si et seulement si au moins l'une des deux espérances

$\mathbb{E}(X^+)$, $\mathbb{E}(X^-)$ est finie : on a alors $\mathbb{E}(X) = \mathbb{E}(X^+) - \mathbb{E}(X^-)$. On vérifie aussi facilement que $\mathbb{E}(|X|) = \mathbb{E}(X^+) + \mathbb{E}(X^-)$, de sorte que X admet une espérance *finie* si et seulement si $\mathbb{E}(|X|) < \infty$ (si et seulement si les deux espérances $\mathbb{E}(X^+)$ et $\mathbb{E}(X^-)$ sont finies).

3.4.2 Propriétés de l'espérance

Proposition 3.33 (Formule de transfert). *Soit X une variable aléatoire discrète à valeurs dans $\mathcal{X} := X(\Omega) \subset E$, où E est un ensemble quelconque, et soit $g : E \rightarrow \mathbb{R}$ une fonction. La variable aléatoire $g(X)$ admet une espérance si et seulement si $\sum_{x \in \mathcal{X}} g(x) p_X(x)$ est bien définie, et si c'est le cas, $\mathbb{E}(g(X))$ est donnée par*

$$\mathbb{E}(g(X)) = \sum_{x \in \mathcal{X}} g(x) p_X(x) \in [-\infty, +\infty]. \quad (3.15)$$

La formule (3.15) (dite *de transfert*) s'applique toujours aux variables aléatoires $g(X)^+$, $g(X)^-$ et $|g(X)|$, qui sont positives et donc admettent une espérance. On en déduit par exemple que $g(X)$ admet une espérance finie si et seulement si $\sum_{x \in \mathcal{X}} |g(x)| p_X(x) < +\infty$.

Démonstration. Notons $Y = g(X)$, qui est à valeurs dans $\mathcal{Y} := g(\mathcal{X}) = \{g(x), x \in \mathcal{X}\}$. Supposons pour l'instant que la fonction g est à valeurs dans \mathbb{R}^+ . Dans ce cas, $g(X)$ admet une espérance (parce qu'elle est positive), et la somme $\sum_{x \in \mathcal{X}} g(x) p_X(x)$ est bien définie car tous les termes sont positifs. Maintenant, on remarque que pour $y \in \mathcal{Y}$

$$p_Y(y) = \mathbb{P}(g(X) = y) = \mathbb{P}(X \in g^{-1}(\{y\})) = \sum_{x \in g^{-1}(\{y\})} \mathbb{P}(X = x) = \sum_{x \in \mathcal{X}, g(x)=y} p_X(x).$$

En appliquant la formule (3.14), on obtient

$$\mathbb{E}(Y) = \sum_{y \in \mathcal{Y}} y p_Y(y) = \sum_{y \in \mathcal{Y}} \sum_{x \in \mathcal{X}: g(x)=y} y p_X(x) = \sum_{y \in \mathcal{Y}} \sum_{x \in \mathcal{X}: g(x)=y} g(x) p_X(x).$$

Maintenant, on remarque que l'on a la décomposition $\mathcal{X} = \bigcup_{y \in \mathcal{Y}} \{x \in \mathcal{X}, g(x) = y\}$: l'union étant disjointe, on obtient donc $\mathbb{E}(Y) = \sum_{x \in \mathcal{X}} g(x) p_X(x)$.

Enlevons maintenant l'hypothèse que g est positive, et posons $a_x := g(x) p_X(x)$ pour simplifier. La famille de réels $(a_x)_{x \in \mathcal{X}}$ admet une somme si et seulement si une des deux sommes $\sum_{x \in \mathcal{X}} a_x^+$ et $\sum_{x \in \mathcal{X}} a_x^-$ est finie, auquel cas $\sum_{x \in \mathcal{X}} a_x := \sum_{x \in \mathcal{X}} a_x^+ - \sum_{x \in \mathcal{X}} a_x^-$. Vu que $a_x^\pm = g(x)^\pm p_X(x)$, on peut appliquer le résultat que l'on vient de démontrer aux fonctions positives g^+ et g^- : on obtient que $\mathbb{E}(g(X)^+) = \sum_{x \in \mathcal{X}} a_x^+$ et $\mathbb{E}(g(X)^-) = \sum_{x \in \mathcal{X}} a_x^-$. Ainsi, la famille $(a_x)_{x \in E}$ admet une somme si et seulement si au moins une des deux espérances $\mathbb{E}(g(X)^+)$ et $\mathbb{E}(g(X)^-)$ est finie, auquel cas $\sum_{x \in \mathcal{X}} a_x = \mathbb{E}(g(X)^+) - \mathbb{E}(g(X)^-)$, ce qui conclut la démonstration. \square

Donnons maintenant une formule alternative pour l'espérance, qui va nous permettre de démontrer des propriétés importantes.

Proposition 3.34. *Soit X une variable aléatoire discrète réelle, définie sur un espace probabilisé $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$ avec Ω dénombrable. Alors X admet une espérance si et seulement si la somme $\sum_{\omega \in \Omega} X(\omega) \mathbb{P}(\{\omega\})$ est bien définie. Si c'est le cas, on a $\mathbb{E}(X) = \sum_{\omega \in \Omega} X(\omega) \mathbb{P}(\{\omega\})$.*

Comme on l'a vu dans le chapitre précédent, si un espace d'état Ω est dénombrable, alors toute probabilité \mathbb{P} sur Ω est déterminée par sa valeur sur les singletons $\{\omega\}$ (ce qui n'est pas le cas pour les espaces probabilisés généraux). Notons que si X est une variable aléatoire discrète définie sur un espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$, alors il est possible de définir une variable aléatoire X' sur un espace probabilisé $(\Omega', \mathcal{P}(\Omega'), \mathbb{P}')$ avec Ω' dénombrable, telle que X' possède la même loi (et donc la même espérance) que X : il suffit de choisir $\Omega' = X(\Omega)$ et $\mathbb{P}' = \mu_X$ la loi de X . En d'autres termes, si on considère une seule variable aléatoire discrète, alors on peut toujours se ramener à un espace probabilisé dénombrable.

Démonstration. La démonstration est analogue à celle de la Proposition 3.33. Si X est positive, son espérance $\mathbb{E}(X)$ est bien définie, et on a

$$\begin{aligned} \mathbb{E}(X) &= \sum_{x \in \mathcal{X}} x p_X(x) = \sum_{x \in \mathcal{X}} x \mathbb{P}(X = x) = \sum_{x \in \mathcal{X}} x \sum_{\omega \in \{X=x\}} \mathbb{P}(\{\omega\}) \\ &= \sum_{x \in \mathcal{X}} \sum_{\omega \in \{X=x\}} x \mathbb{P}(\{\omega\}) = \sum_{x \in \mathcal{X}} \sum_{\omega \in \{X=x\}} X(\omega) \mathbb{P}(\{\omega\}), \end{aligned}$$

où pour la dernière égalité on a utilisé que $X(\omega) = x$ pour tout $\omega \in \{X = x\}$. En remarquant que $\Omega = \bigcup_{x \in \mathcal{X}} \{X = x\}$, l'union étant disjointe, on obtient que $\mathbb{E}(X) = \sum_{\omega \in \Omega} X(\omega) \mathbb{P}(\{\omega\})$.

Dans le cas où X n'est pas positive, on applique la formule à X^+ et X^- . Il s'ensuit que la somme $\sum_{\omega \in \Omega} X(\omega) \mathbb{P}(\{\omega\})$ est bien définie si et seulement si au moins une des deux espérances $\mathbb{E}(X^+)$ et $\mathbb{E}(X^-)$ est finie, auquel cas $\sum_{\omega \in \Omega} X(\omega) \mathbb{P}(\{\omega\}) = \mathbb{E}(X^+) - \mathbb{E}(X^-)$. Cela conclut la démonstration. \square

Proposition 3.35 (Propriétés de l'espérance). *Soient X, Y des variables aléatoires discrètes réelles, définies sur le même espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$.*

1. **(Monotonie)** *Si X, Y admettent toutes deux une espérance, et $X(\omega) \leq Y(\omega)$ pour tout $\omega \in \Omega$, alors $\mathbb{E}(X) \leq \mathbb{E}(Y)$.*
2. *Si X admet une espérance, alors $|\mathbb{E}(X)| \leq \mathbb{E}(|X|)$.
Si $\mathbb{E}(|X|) < +\infty$, alors $\mathbb{E}(X)$ existe et est finie.*
3. **(Linéarité)** *Si X et Y admettent une espérance (resp. si X et Y sont positives) et $a, b \in \mathbb{R}$ (resp. $a, b \in \mathbb{R}^+$), alors la variable aléatoire $aX + bY$ admet une espérance (resp. est positive) et*

$$\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y).$$

Les démonstrations sont de simples conséquences de la Proposition 3.34 et des propriétés des sommes infinies, et sont laissées en exercice. La propriété de linéarité de l'espérance, en particulier, est utilisée très souvent (voir les deux exemples qui suivent), et peut avoir des conséquences moins évidentes que ce que l'on pourrait supposer.

Exemple 3.36. Redémontrons ici la formule d'inclusion-exclusion de la Proposition 2.7. L'idée est d'écrire $\mathbb{P}(\bigcup_{i=1}^n A_i) = \mathbb{E}(\mathbb{1}_{\bigcup_{i=1}^n A_i})$, et d'utiliser les propriétés de la fonction indicatrice. On peut en effet écrire

$$\mathbb{1}_{\bigcup_{i=1}^n A_i} = 1 - \mathbb{1}_{\bigcap_{i=1}^n A_i^c} = 1 - \prod_{i=1}^n \mathbb{1}_{A_i^c} = 1 - \prod_{i=1}^n (1 - \mathbb{1}_{A_i}).$$

Ensuite, en développant le produit, on obtient

$$\begin{aligned} \prod_{i=1}^n (1 - \mathbb{1}_{A_i}) &= 1 - \sum_{i=1}^n \mathbb{1}_{A_i} + \sum_{1 \leq i_1 < i_2 \leq n} \mathbb{1}_{A_{i_1}} \mathbb{1}_{A_{i_2}} - \sum_{1 \leq i_1 < i_2 < i_3 \leq n} \mathbb{1}_{A_{i_1}} \mathbb{1}_{A_{i_2}} \mathbb{1}_{A_{i_3}} + \cdots \\ &= 1 + \sum_{k=1}^n (-1)^k \sum_{\{i_1, \dots, i_k\} \subset \{1, \dots, n\}} \mathbb{1}_{A_{i_1} \cap \dots \cap A_{i_k}}, \end{aligned}$$

de sorte que

$$\mathbb{1}_{\bigcup_{i=1}^n A_i} = \sum_{k=1}^n (-1)^{k+1} \sum_{\{i_1, \dots, i_k\} \subset \{1, \dots, n\}} \mathbb{1}_{A_{i_1} \cap \dots \cap A_{i_k}}.$$

Par linéarité de l'espérance, et en utilisant que $\mathbb{E}(\mathbb{1}_A) = \mathbb{P}(A)$, on obtient donc

$$\mathbb{P}\left(\bigcup_{i=1}^n A_i\right) = \mathbb{E}\left(\mathbb{1}_{\bigcup_{i=1}^n A_i}\right) = \sum_{k=1}^n (-1)^{k+1} \sum_{\{i_1, \dots, i_k\} \subset \{1, \dots, n\}} \mathbb{P}(A_{i_1} \cap \dots \cap A_{i_k}),$$

qui est la formule d'inclusion-exclusion.

Exemple 3.37. Soient A_1, \dots, A_m des événements quelconques (en particulier, ils ne sont pas nécessairement indépendants). On considère la variable aléatoire réelle

$$X := \mathbb{1}_{A_1} + \cdots + \mathbb{1}_{A_m},$$

qui compte le nombre d'événements A_i qui sont vérifiés. La loi de X dépend des relations qu'il y a entre les événements A_1, \dots, A_m et, en général, peut être compliquée. Cependant, par linéarité de l'espérance, on a toujours

$$\mathbb{E}(X) = \mathbb{E}(\mathbb{1}_{A_1}) + \cdots + \mathbb{E}(\mathbb{1}_{A_m}) = \mathbb{P}(A_1) + \cdots + \mathbb{P}(A_m).$$

En particulier, $\mathbb{E}(X) = mp$ si tous les événements ont la même probabilité $p = \mathbb{P}(A_i)$.

Revenons alors à l'Exemple 2.15. On sélectionne au hasard n personnes nées une année non bissextile, on les numérote de 1 à n , et on introduit, pour toute paire de personnes $\{i, j\} \subseteq \{1, \dots, n\}$ (avec $i \neq j$), l'événement $A_{\{i, j\}} =$ "les personnes i et j ont leur anniversaire le même jour". On a alors $m = \binom{n}{2} = \frac{1}{2}n(n-1)$ événements $A_{\{i, j\}}$ (non indépendants!), tous de même probabilité $p = \mathbb{P}(A_{\{i, j\}}) = \frac{1}{365}$. Dans ce cas, la variable aléatoire

$$X := \sum_{\{i, j\} \subseteq \{1, \dots, n\}} \mathbb{1}_{A_{\{i, j\}}}$$

compte le nombre de paires de personnes qui ont leurs anniversaires le même jour, et a pour espérance

$$\mathbb{E}(X) = mp = \frac{1}{2}n(n-1) \cdot \frac{1}{365} = \frac{(n-1)n}{730}. \quad (3.16)$$

On sait déjà que pour $n = 23$, donc pour un groupe relativement petit, la probabilité qu'au moins deux personnes aient leurs anniversaires le même jour, c'est-à-dire $\mathbb{P}(X \geq 1)$, est strictement plus grand que $\frac{1}{2}$. Ce résultat est "confirmé" par la relation (3.16) : en effet, pour $n = 23$, on a $\mathbb{E}(X) \approx 0,7$, une valeur supérieure à $\frac{1}{2}$. Cela fournit une "explication" possible au paradoxe des anniversaires : bien que le nombre de personnes $n = 23$ soit relativement faible, la quantité pertinente est le nombre de *paires de personnes*, $\frac{1}{2}n(n-1) = 253$, qui est comparable au nombre 365 de jours de l'année.

On peut aussi exploiter la propriété de monotonie de l'espérance, pour montrer qu'une variable aléatoire réelle X admet une espérance finie. Par exemple, si Z, X sont deux variables aléatoires réelles définies sur le même espace probabilisé telles que $|X| \leq Z$, alors, si Z admet une espérance finie, X admet aussi une espérance finie.

On a aussi le résultat suivant.

Proposition 3.38. *Soit X une variable aléatoire discrète réelle positive (définie sur un espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$ avec Ω dénombrable) : $X(\omega) \geq 0$ pour tout $\omega \in \Omega$. Si $\mathbb{E}(X) = 0$, alors X est presque sûrement égale à 0, c'est-à-dire $\mathbb{P}(X = 0) = 1$.*

Démonstration. La somme $\sum_{\omega \in \Omega} X(\omega)\mathbb{P}(\{\omega\})$ est à termes positifs, et est donc bien définie, égale à $\mathbb{E}(X) = 0$. Mais une somme de termes positifs vaut zéro si et seulement si tous les termes sont nuls : $X(\omega)\mathbb{P}(\{\omega\}) = 0$ pour tout $\omega \in \Omega$, c'est-à-dire $\mathbb{P}(\{\omega\}) = 0$ pour tous les ω tels que $X(\omega) > 0$. Cela implique que $\mathbb{P}(X > 0) = 0$ (pourquoi?) et donc $\mathbb{P}(X = 0) = \mathbb{P}(X \geq 0) - \mathbb{P}(X > 0) = 1$, comme voulu. \square

On souligne ici que la propriété de linéarité de l'espérance s'étend facilement à des sommes finies de variables aléatoires. On énonce ici un résultat que l'on admettra (correspondant au théorème de convergence monotone en théorie de la mesure) pour traiter le cas de sommes infinies — on ne s'en servira d'ailleurs pas souvent dans ce cours.

Proposition 3.39 (admise). *Soit $(X_i)_{i \geq 1}$ une suite de variables aléatoires positives définies sur le même espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$. Alors $\mathbb{E}\left(\sum_{i=1}^{\infty} X_i\right) = \sum_{i=1}^{\infty} \mathbb{E}(X_i)$.*

Rappelons que $\sum_{i=1}^{\infty} X_i$ est bien une variable aléatoire (les termes étant tous positifs, la somme est bien définie), qui peut potentiellement valoir $+\infty$.

3.4.3 Moments, variance et covariance

Définition 3.40 (Moments). *Pour un $k \geq 1$ donné, on dit que la variable aléatoire X admet un moment d'ordre k fini si $\mathbb{E}(|X|^k) < \infty$. Si c'est le cas, la quantité $\mathbb{E}(X^k)$ est finie, et appelée moment d'ordre k de X .*

Soulignons quelques propriétés importantes :

- si X et Y sont deux variables aléatoires qui admettent un moment d'ordre k fini, alors il en va de même pour $X + Y$.
- si X admet un moment d'ordre k fini, alors X admet un moment d'ordre j fini, pour tout $1 \leq j \leq k$;

Pour le premier point, on utilise que pour tous $x, y \in \mathbb{R}$, $|x+y| \leq 2 \max\{|x|, |y|\}$ de sorte que $|x+y|^k \leq 2^k \max\{|x|^k, |y|^k\} \leq 2^k|x|^k + 2^k|y|^k$: par monotonie et linéarité de l'espérance, on a donc $\mathbb{E}(|X+Y|^k) \leq 2^k\mathbb{E}(|X|^k) + 2^k\mathbb{E}(|Y|^k)$. Pour le deuxième point, on utilise que pour $1 \leq j \leq k$ on a pour tout $x \in \mathbb{R}$, $|x|^j \leq 1 + |x|^k$ (distinguer les cas $|x| \leq 1$ et $|x| > 1$) : par monotonie et linéarité de l'espérance, on a donc $\mathbb{E}(|X|^j) \leq 1 + \mathbb{E}(|X|^k)$.

Les moments les plus utilisés sont ceux d'ordre 1, qui n'est autre que l'espérance $\mathbb{E}(X)$, et celui d'ordre 2, c'est-à-dire $\mathbb{E}(X^2)$.

Définition 3.41 (Variance et covariance). Soient X et Y des variables aléatoires réelles définies sur le même espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$.

- Si X admet un moment d'ordre 2 fini, on appelle variance de X la quantité

$$\text{Var}(X) = \mathbb{E}((X - \mathbb{E}(X))^2) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 \geq 0.$$

- Si X, Y et XY admettent une espérance finie, on appelle covariance de X et Y la quantité

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

Soulignons que les deux expressions pour la variance découlent simplement de la linéarité de l'espérance : si on note $\mu := \mathbb{E}(X)$, en développant $(X - \mu)^2 = X^2 - 2\mu X + \mu^2$ et en utilisant la linéarité de l'espérance, on obtient que $\mathbb{E}((X - \mu)^2) = \mathbb{E}(X^2) - 2\mu\mathbb{E}(X) + \mu^2 = \mathbb{E}(X^2) - \mu^2$. Il en va de même pour les deux expressions pour la covariance de X et Y .

Remarque 3.42. Si X et Y admettent un moment d'ordre 2 fini, alors XY admet une espérance finie : en effet, on utilise que $|xy| \leq \frac{1}{2}x^2 + \frac{1}{2}y^2$ pour tous $x, y \in \mathbb{R}$, de sorte que par monotonie et linéarité de l'espérance, on a $\mathbb{E}(|XY|) \leq \frac{1}{2}\mathbb{E}(X^2) + \frac{1}{2}\mathbb{E}(Y^2)$. Ainsi, si deux variables aléatoires X et Y admettent un moment d'ordre 2 fini, leur covariance $\text{Cov}(X, Y)$ est bien définie.

La variance $\text{Var}(X)$ est l'espérance de la distance, élevée au carré, entre X et $\mathbb{E}(X)$: elle est intuitivement liée à la “dispersion” des valeurs de X autour son espérance.

La racine carrée de la variance $\sqrt{\text{Var}(X)}$ est appelée *écart type*.

Proposition 3.43 (Propriétés de la covariance). La covariance $\text{Cov}(\cdot, \cdot)$ est une forme symétrique et bilinéaire : pour toutes variables aléatoires X, Y, Z admettant un moment d'ordre 2 fini, pour $\alpha, \beta \in \mathbb{R}$ on a

$$\begin{aligned} \text{Cov}(X, Y) &= \text{Cov}(Y, X), \\ \text{Cov}(\alpha X + \beta Y, Z) &= \alpha \text{Cov}(X, Z) + \beta \text{Cov}(Y, Z). \end{aligned}$$

De plus, si X ou Y est presque sûrement constante, on a $\text{Cov}(X, Y) = 0$.

Démonstration. Les propriétés de symétrie et de bilinéarité découlent facilement de la définition $\text{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$, et de la linéarité de l'espérance.

Si X est presque sûrement constante, égale à la constante $c \in \mathbb{R}$, alors $\mathbb{E}(X) = c$, donc la variable aléatoire $X - \mathbb{E}(X)$ est presque sûrement égale à 0. Par conséquent, la variable aléatoire $(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))$ est elle aussi presque sûrement égale à 0, et donc $\text{Cov}(X, Y) = \mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))) = 0$. \square

Proposition 3.44 (Propriétés de la variance). Soit X une variable aléatoire admettant un moment d'ordre 2 fini, alors :

- $\text{Var}(aX + b) = a^2 \text{Var}(X)$ pour tous $a, b \in \mathbb{R}$;
- $\text{Var}(X) = 0$ si et seulement si X est presque sûrement constante ;
- $\text{Var}(X) \leq \mathbb{E}((X - c)^2)$ pour tout $c \in \mathbb{R}$, avec égalité si et seulement si $c = \mathbb{E}(X)$.

Si X_1, \dots, X_n sont des variables aléatoires admettant un moment d'ordre 2 fini, alors

$$\text{Var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \text{Var}(X_i) + \sum_{1 \leq i, j \leq n, i \neq j} \text{Cov}(X_i, X_j). \quad (3.17)$$

Démonstration. Grâce à la Proposition 3.43, la covariance est une forme bilinéaire : comme $\text{Var}(Z) = \text{Cov}(Z, Z)$, cela démontre donc (3.17). Le fait que $\text{Var}(aX + b) = a^2 \text{Var}(X)$ découle aussi de la bilinéarité de la covariance, et du fait qu'elle s'annule si l'un de ses deux argument est constant.

Pour le deuxième point, si X est presque sûrement constante, alors d'après la Proposition 3.43, $\text{Var}(X) = \text{Cov}(X, X) = 0$. Réciproquement, si $\text{Var}(X) = 0$, en posant $Y := (X - \mathbb{E}(X))^2$, cela signifie que $\mathbb{E}(Y) = 0$. Comme Y est une variable aléatoire positive, la Proposition 3.38 montre que $\mathbb{P}(Y = 0) = 1$: cela montre que presque sûrement $X - \mathbb{E}(X) = 0$, c'est-à-dire que X est presque sûrement égale à la constante $\mathbb{E}(X)$.

Pour le troisième point, il suffit d'observer que pour tout $c \in \mathbb{R}$ on a l'identité $0 \leq \text{Var}(X) = \text{Var}(X - c) = \mathbb{E}((X - c)^2) - (\mathbb{E}(X) - c)^2$. \square

3.4.4 Espérance et indépendance

Proposition 3.45. Soient X, Y des variables aléatoires réelles indépendantes.

* Si X et Y admettent toutes les deux une espérance finie, alors XY admet une espérance finie, et $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$.

* Si X et Y sont des variables positives, alors $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$ même sans l'hypothèse que l'espérance de X et Y sont finies.

Démonstration. Comme X et Y sont indépendantes, on a que $p_{X,Y}(x, y) = p_X(x)p_Y(y)$ pour tout $(x, y) \in \mathbb{R}^2$, grâce à la Proposition 3.15. Par conséquent, en appliquant la formule de transfert (3.15) à la variable aléatoire (X, Y) (à valeurs dans $\mathcal{X} \times \mathcal{Y} \subset \mathbb{R}^2$) pour la fonction positive $g(z) = g(x, y) := |xy|$, on obtient

$$\begin{aligned} \mathbb{E}(|XY|) &= \sum_{(x,y) \in \mathcal{X} \times \mathcal{Y}} |x||y| p_{X,Y}(x, y) = \left(\sum_{x \in \mathcal{X}} |x| p_X(x) \right) \left(\sum_{y \in \mathcal{Y}} |y| p_Y(y) \right) \\ &= \mathbb{E}(|X|) \mathbb{E}(|Y|). \end{aligned}$$

Cela montre que si X et Y sont positives, on a toujours $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$. De plus, si $\mathbb{E}(|X|) < \infty$ et $\mathbb{E}(|Y|) < \infty$ alors on a aussi $\mathbb{E}(|XY|) < \infty$. En répétant les mêmes étapes en enlevant les valeurs absolues, on montre que $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$. \square

En se rappelant la définition de covariance et la formule (3.17), on obtient immédiatement les corollaires suivants de la Proposition 3.45.

Corollaire 3.46. Soient X, Y deux variables aléatoires indépendantes, admettant une espérance finie, alors $\text{Cov}(X, Y) = 0$.

Attention, l'exemple suivant montre que la réciproque est fautive !

Exercice 3.47. Soit X une variable aléatoire à valeurs dans $\{-1, 0, 1\}$ de loi uniforme sur cet ensemble. On pose $Y = \mathbb{1}_{\{X=0\}}$. Montrer que $\text{Cov}(X, Y) = 0$, mais que X et Y ne sont pas indépendantes.

La formule (3.17), avec le fait que $\text{Cov}(X, Y) = 0$ si X et Y sont indépendantes, donne le résultat suivant, qui nous sera extrêmement utile (par exemple dans le Chapitre 7).

Corollaire 3.48. Si X_1, \dots, X_n sont des variables aléatoires indépendantes admettant in moment d'ordre 2 fini,

$$\text{Var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \text{Var}(X_i).$$

3.4.5 Espérance et variance des lois classiques

Calculons l'espérance et la variance des variables aléatoires introduites dans la Section 3.3.

Loi uniforme. Soit X une variable aléatoire de loi uniforme sur $\{1, \dots, n\}$. En appliquant la formule de transfert (3.15) on a

$$\begin{aligned} \mathbb{E}(X) &= \sum_{k=1}^n k p_X(k) = \frac{1}{n} \sum_{k=1}^n k = \frac{n+1}{2}, \\ \mathbb{E}(X^2) &= \sum_{k=1}^n k^2 p_X(k) = \frac{1}{n} \sum_{k=1}^n k^2 = \frac{(n+1)(2n+1)}{6}. \end{aligned}$$

Loi de Bernoulli. Soit X une variable aléatoire de loi de Bernoulli de paramètre p . On a déjà vu que $\mathbb{E}(X) = 0 \cdot \mathbb{P}(X=0) + 1 \cdot \mathbb{P}(X=1) = p$. Comme on a $X^2 = X$ car $X \in \{0, 1\}$, il est immédiat de voir que $\mathbb{E}(X^2) = \mathbb{E}(X) = p$ et donc $\text{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2 = p(1-p)$.

Loi binomiale. Pour calculer l'espérance et la variance d'une variables aléatoire X de loi binomiale, on utilise l'interprétation d'une loi binomiale. Soient X_1, \dots, X_n des variables aléatoires indépendantes, de loi $X_i \sim \text{Bern}(p)$. Alors $X := X_1 + \dots + X_n \sim \text{Bin}(n, p)$, et par linéarité de l'espérance on obtient donc

$$\mathbb{E}(X) = \mathbb{E}(X_1 + \dots + X_n) = np.$$

De plus, étant donné que les X_i sont indépendantes, en utilisant le Corollaire 3.48 on obtient

$$\text{Var}(X) = \sum_{i=1}^n \text{Var}(X_i) = np(1-p).$$

De manière alternative, des calculs directs sont possibles (mais plus techniques), en appliquant la formule de transfert (3.15). On les laisse en exercice.

Loi de Poisson. Soit X une variable aléatoire de loi de Poisson de paramètre $\lambda > 0$. D'après la définition (3.14) de l'espérance on a

$$\mathbb{E}(X) = \sum_{k=0}^{+\infty} k p_X(k) = \sum_{k=1}^{+\infty} k \frac{\lambda^k}{k!} e^{-\lambda} = \lambda e^{-\lambda} \sum_{k=1}^{+\infty} \frac{\lambda^{k-1}}{(k-1)!},$$

où on a utilisé que le terme $k = 0$ dans la somme vaut 0. Comme la dernière somme est le développement en série entière de e^λ (en faisant un changement d'indice $k' = k - 1$), on obtient $\mathbb{E}(X) = \lambda$. En appliquant la formule de transfert (3.15), on a, de la même manière

$$\mathbb{E}(X(X-1)) = \sum_{k=0}^{+\infty} k(k-1) p_X(k) = \sum_{k=2}^{+\infty} k(k-1) \frac{\lambda^k}{k!} e^{-\lambda} = \lambda^2 \sum_{k=2}^{+\infty} \frac{\lambda^{k-2}}{(k-2)!} e^{-\lambda} = \lambda^2.$$

Comme $X^2 = X(X-1) + X$, on en déduit par linéarité de l'espérance que $\mathbb{E}(X^2) = \mathbb{E}(X(X-1)) + \mathbb{E}(X) = \lambda^2 + \lambda$. On en conclut que $\text{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \lambda^2 + \lambda - \lambda^2 = \lambda$.

Loi géométrique. Soit X une variable aléatoire géométrique de paramètre $p \in]0, 1[$. D'après la définition (3.14) de l'espérance on a

$$\mathbb{E}(X) = \sum_{k=1}^{+\infty} k p_X(k) = \sum_{k=1}^{+\infty} k(1-p)^{k-1} p.$$

Si on considère la série entière $f(x) = \sum_{k=0}^{\infty} x^k$ de rayon de convergence 1, donc définie et dérivable sur $] -1, 1[$, le théorème de dérivation des séries entières nous donne que $f'(1-p) = \sum_{k=0}^{\infty} k(1-p)^{k-1}$: on a donc $\mathbb{E}(X) = p f'(1-p) = \frac{1}{p}$ (car $f(x) = \frac{1}{1-x}$, et $f'(x) = \frac{1}{(1-x)^2}$). En appliquant la formule de transfert (3.15), on a, de la même manière

$$\mathbb{E}(X(X-1)) = \sum_{k=1}^{+\infty} k(k-1) p_X(k) = \sum_{k=2}^{+\infty} k(k-1) (1-p)^{k-1} p = p(1-p) f''(1-p),$$

ce qui donne $\mathbb{E}(X(X-1)) = \frac{2(1-p)}{p^2}$. On en déduit par linéarité de l'espérance que $\mathbb{E}(X^2) = \mathbb{E}(X(X-1)) + \mathbb{E}(X) = \frac{2(1-p)}{p^2} + \frac{1}{p} = \frac{2-p}{p^2}$, et donc $\text{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \frac{1-p}{p^2}$.

3.5 Inégalités probabilistes et applications

Les inégalités que l'on va énoncer dans cette section peuvent paraître simples, mais leurs applications sont extrêmement importantes, et à la base de nombreux raisonnements probabilistes.

Théorème 3.49 (Inégalité de Markov). Soit X une variable aléatoire réelle, à valeurs positives. Alors, pour tout $\varepsilon > 0$,

$$\mathbb{P}(X \geq \varepsilon) \leq \frac{\mathbb{E}(X)}{\varepsilon}.$$

Démonstration. Par hypothèse $X(\omega) \geq 0$ pour tout $\omega \in \Omega$. Par conséquent, on a l'inégalité entre variables aléatoires

$$X = X\mathbf{1}_{\{X \geq \varepsilon\}} + X\mathbf{1}_{\{X < \varepsilon\}} \geq \varepsilon\mathbf{1}_{\{X \geq \varepsilon\}} + 0.$$

Par monotonie et linéarité de l'espérance, on obtient

$$\mathbb{E}(X) \geq \mathbb{E}(\varepsilon\mathbf{1}_{\{X \geq \varepsilon\}}) = \varepsilon\mathbb{E}(\mathbf{1}_{\{X \geq \varepsilon\}}) = \varepsilon\mathbb{P}(X \geq \varepsilon),$$

ce qui donne l'inégalité voulue. \square

Théorème 3.50 (Inégalité de Bienaymé-Tchebychev). *Soit X une variable aléatoire réelle, admettant un moment d'ordre 2 fini. Alors pour tout $\varepsilon > 0$*

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq \varepsilon) \leq \frac{\text{Var}(X)}{\varepsilon^2}.$$

Démonstration. On remarque que l'on a l'égalité d'événements suivante :

$$\{|X - \mathbb{E}(X)| \geq \varepsilon\} = \{|X - \mathbb{E}(X)|^2 \geq \varepsilon^2\}.$$

Ainsi, en appliquant l'inégalité de Markov à la variable aléatoire positive $(X - \mathbb{E}(X))^2$, on obtient

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq \varepsilon) = \mathbb{P}((X - \mathbb{E}(X))^2 \geq \varepsilon^2) \leq \frac{\mathbb{E}((X - \mathbb{E}(X))^2)}{\varepsilon^2},$$

et comme $\mathbb{E}((X - \mathbb{E}(X))^2) = \text{Var}(X)$, cela donne l'inégalité voulue. \square

Cette inégalité justifie l'appellation "d'écart-type" : en l'appliquant avec $\varepsilon = c\sigma$, où $\sigma = \sqrt{\text{Var}(X)}$ est l'écart-type de X et $c > 0$, on obtient $\mathbb{P}(|X - \mathbb{E}(X)| > c\sigma) \leq 1/c^2$. Par exemple, la probabilité que la variable aléatoire X s'écarte de son espérance $\mathbb{E}(X)$ de plus de 2σ est plus petite qu'un quart. Autrement dit, typiquement, X s'écarte de son espérance $\mathbb{E}(X)$ d'au plus une constante fois σ .

Théorème 3.51 (Inégalité de Cauchy-Schwarz). *Soient X et Y des variables aléatoires réelles admettant un moment d'ordre 2 fini. Alors le produit XY admet une espérance finie, et on a l'inégalité*

$$|\mathbb{E}(XY)| \leq \sqrt{\mathbb{E}(X^2)\mathbb{E}(Y^2)}. \quad (3.18)$$

Démonstration. On a déjà montré que XY admet une espérance finie si X et Y admettent un moment d'ordre 2 fini. Supposons que $\mathbb{E}(Y^2) > 0$: cela implique que Y n'est pas nulle presque sûrement — si $Y = 0$ presque sûrement l'inégalité (3.18) est triviale. Maintenant pour tout $\lambda \in \mathbb{R}$, on a $(\lambda Y - X)^2 = \lambda^2 Y^2 - 2\lambda XY + X^2 \geq 0$, de sorte que par monotonie et linéarité de l'espérance, on a

$$\lambda^2 \mathbb{E}(Y^2) - 2\lambda \mathbb{E}(XY) + \mathbb{E}(X^2) \geq 0$$

pour tout $\lambda \in \mathbb{R}$. Il s'agit d'un polynôme de degré 2 en λ qui reste positif, et donc son discriminant $\Delta = 4\mathbb{E}(XY)^2 - 4\mathbb{E}(X^2)\mathbb{E}(Y^2)$ est négatif. On en déduit donc que $\mathbb{E}(XY)^2 \leq \mathbb{E}(X^2)\mathbb{E}(Y^2)$, ce qui donne (3.18).

Notons que si $\mathbb{E}(Y^2) > 0$ on a égalité dans (3.18) si et seulement si il existe un $\lambda > 0$ tel que $\mathbb{E}((\lambda Y - X)^2) = 0$, ce qui implique que $X = \lambda Y$ presque sûrement d'après la Proposition 3.38. \square

★ Une application : la méthode des moments pour le problème du collectionneur

Lorsque l'on est confronté à une variable aléatoire X "compliquée", il n'est pas toujours possible d'en calculer la loi explicitement. On cherche alors à obtenir des estimées sur les probabilités d'événements liés à cette variable : les inégalités de Markov et de Tchebychev montrent que certaines de ces probabilités s'estiment à partir des moments de X , dont le calcul se révèle souvent plus simple. La *méthode des moments* consiste à estimer les moments d'une variable aléatoire, dans le but d'en déduire des informations sur cette variable.

Considérons le problème suivant. Une marque de céréales offre une pièce d'un puzzle en cadeau dans chacun ses paquets. On suppose que le puzzle complet comporte n pièces, et que la pièce offerte dans chaque paquet est indépendante des autres paquets, et est choisie uniformément parmi les n pièces possibles. Combien de paquets faut-il acheter pour avoir de bonnes chances d'avoir toutes les pièces du paquet ?

Plus formellement, on considère $(X_k)_{k \geq 1}$ des variables aléatoires indépendantes, de loi uniforme sur $\{1, \dots, n\}$: la variable X_k représente le numéro de la pièce contenue dans le k -ème paquet acheté. Considérons Y_k la variable aléatoire égale au nombre de pièces qu'il reste encore à récupérer après avoir acheté k paquets : en particulier, si $Y_k \geq 1$ cela veut dire que l'on n'a pas toutes les pièces, alors que si $Y_k = 0$ on a toutes les pièces. Remarquons que l'on peut écrire

$$Y_k = \sum_{i=1}^n \mathbb{1}_{A_i^{(k)}},$$

où $A_i^{(k)}$ désigne l'événement que la pièce n° i n'a pas encore été trouvée après avoir acheté k paquets, c'est-à-dire $A_i^{(k)} = \bigcap_{j=1}^k \{X_j \neq i\}$.

On peut déjà utiliser la linéarité de l'espérance et le fait que $\mathbb{E}(\mathbb{1}_A) = \mathbb{P}(A)$ pour avoir

$$\mathbb{E}(Y_k) = \sum_{i=1}^n \mathbb{P}(A_i^{(k)}) = n \times \left(1 - \frac{1}{n}\right)^k,$$

où on a utilisé le fait que $\mathbb{P}(A_i^{(k)}) = \prod_{j=1}^k \mathbb{P}(X_j \neq i) = \left(1 - \frac{1}{n}\right)^k$, par indépendance des X_j . D'après l'inégalité de Markov, on obtient

$$\mathbb{P}(Y_k \geq 1) \leq \mathbb{E}(Y_k) = n \left(1 - \frac{1}{n}\right)^k \leq n e^{-\frac{k}{n}},$$

où on a aussi utilisé que $1 - x \leq e^{-x}$ pour tout $x \in \mathbb{R}$. Ainsi, on voit clairement que pour tout $\varepsilon > 0$, si $k \geq (1 + \varepsilon)n \ln n$, alors $\mathbb{P}(Y_k \geq 1) \leq n^{-\varepsilon} \rightarrow 0$ quand $n \rightarrow \infty$. On a donc montré que pour $\varepsilon > 0$ arbitraire fixé, si le puzzle contient un grand nombre n de pièces, il sera très peu probable qu'il reste encore des pièces à récupérer après avoir acheté $(1 + \varepsilon)n \ln n$ paquets.

On peut aussi calculer la variance de Y_k : d'après la Proposition 3.44, on a

$$\text{Var}(Y_k) = \sum_{i=1}^n \text{Var}(\mathbb{1}_{A_i^{(k)}}) + \sum_{1 \leq i, j \leq n, i \neq j} \text{Cov}(\mathbb{1}_{A_i^{(k)}}, \mathbb{1}_{A_j^{(k)}}) \leq n \times \left(1 - \frac{1}{n}\right)^k,$$

où on a utilisé que $\text{Var}(\mathbb{1}_A) = \mathbb{P}(A)(1 - \mathbb{P}(A)) \leq \mathbb{P}(A)$ et le fait que pour $i \neq j$ on ait $\text{Cov}(\mathbb{1}_{A_i^{(k)}}, \mathbb{1}_{A_j^{(k)}}) = \left(1 - \frac{2}{n}\right)^k - \left(1 - \frac{1}{n}\right)^{2k} \leq 0$ (par un calcul laissé en exercice). En remarquant que l'événement $\{Y_k = 0\}$

est égal à $\{Y_k \leq 0\} = \{Y_k - \mathbb{E}(Y_k) \leq -\mathbb{E}(Y_k)\} \subset \{|Y_k - \mathbb{E}(Y_k)| \geq \mathbb{E}(Y_k)\}$, on peut utiliser l'inégalité de Bienaymé-Tchebychev, pour obtenir

$$\mathbb{P}(Y_k = 0) \leq \mathbb{P}\left(|Y_k - \mathbb{E}(Y_k)| \geq \mathbb{E}(Y_k)\right) \leq \frac{\text{Var}(Y_k)}{\mathbb{E}(Y_k)^2} \leq \frac{1}{n(1 - \frac{1}{n})^k}.$$

On remarque ici que pour tout $\varepsilon > 0$, si $k \leq (1 - \varepsilon)n \ln n$, alors $n(1 - \frac{1}{n})^k \rightarrow +\infty$ quand $n \rightarrow +\infty$ (exercice : faire un développement limité), et donc $\mathbb{P}(Y_k = 0) \rightarrow 0$. On a donc montré que si le puzzle contient un grand nombre n de pièces, il sera très peu probable d'avoir récupéré toutes les pièces après avoir acheté seulement $(1 - \varepsilon)n \ln n$ paquets.

3.6 Fonction génératrice et applications

On introduit dans cette section un outil pour étudier des variables aléatoires à valeurs dans \mathbb{N} , appelé *fonction génératrice*. Pour X une variable aléatoire à valeurs dans \mathbb{N} , de densité discrète p_X , la fonction génératrice de X est la fonction

$$G_X(s) := \mathbb{E}(s^X) = \sum_{n=0}^{+\infty} p_X(n) s^n, \quad (3.19)$$

définie au moins pour $s \in [-1, 1]$ (on utilise la convention que $0^0 = 1$).

Exemple 3.52. Si X est une variable aléatoire de loi de Poisson de paramètre $\lambda > 0$, alors la fonction génératrice de X est

$$G_X(s) = \sum_{k=0}^{+\infty} \frac{\lambda^k}{k!} e^{-\lambda} s^k = e^{\lambda(s-1)} \quad \text{pour tout } s \in \mathbb{R},$$

où on a utilisé que $\sum_{k=0}^{+\infty} \frac{(\lambda s)^k}{k!} = e^{\lambda s}$.

Exercice 3.53. Soit $X \sim \text{Bern}(p)$. Montrer que $G_X(s) = 1 + p(s - 1)$ pour tout $s \in \mathbb{R}$.

Le premier résultat essentiel est que la fonction génératrice G_X d'une variable aléatoire X à valeurs entières *caractérise* sa loi, dans le sens où si l'on connaît la fonction G_X alors on peut récupérer la densité discrète $(p_X(n))_{n \geq 0}$: deux variables aléatoires à valeurs dans \mathbb{N} ayant la même fonction génératrice possèdent donc la même loi.

Proposition 3.54. Soit X une variable aléatoire à valeurs dans \mathbb{N} , et soit G_X sa fonction génératrice. Alors G_X est bien définie (au moins) sur $[-1, 1]$, et est infiniment dérivable (au moins) sur $] -1, 1[$. De plus, la fonction génératrice caractérise la loi : on a $p_X(k) = \frac{1}{k!} G_X^{(k)}(0)$ pour tout $k \in \mathbb{N}$, où $G_X^{(k)}$ est la dérivée k -ème de G_X .

Démonstration. Tout d'abord, le fait que G_X soit définie pour tout $s \in [-1, 1]$ provient de la définition (3.19) : pour tout $s \in [-1, 1]$, la série converge absolument, car $|p_X(n)s^n| \leq p_X(n)$, dont la somme vaut 1. La série entière définie dans (3.19) possède donc un *rayon de convergence* au moins égal à 1. Le théorème de dérivation des séries entières assure le fait

que G_X soit infiniment dérivable sur l'intervalle $] -1, 1[$, et donne aussi que pour $k \geq 0$, pour tout $s \in] -1, 1[$ on a,

$$G_X^{(k)}(s) = \sum_{n=k}^{+\infty} n(n-1)\cdots(n-k+1)p_X(n)s^{n-k} = \sum_{n=k}^{+\infty} \frac{n!}{(n-k)!} p_X(n)s^{n-k}. \quad (3.20)$$

En prenant $s = 0$ dans cette expression, tous les termes de la somme sont nuls, sauf le terme $n = k$, qui vaut $k!p_X(k)$. On a donc montré que $G_X^{(k)}(0) = k!p_X(k)$. (Il s'agit en fait de l'unicité de la décomposition en série entière.) \square

Exercice 3.55. Soit X une variable aléatoire à valeurs dans \mathbb{N} . Montrer que G_X est croissante et convexe sur $[0, 1]$.

Proposition 3.56. Si X et Y sont deux variables aléatoires indépendantes à valeurs dans \mathbb{N} , alors pour $s \in \mathbb{R}$ tel que $G_X(s)$ et $G_Y(s)$ sont bien définies, on a $G_{X+Y}(s) = G_X(s)G_Y(s)$.

Démonstration. Il suffit d'écrire $G_{X+Y}(s) := \mathbb{E}(s^{X+Y}) = \mathbb{E}(s^X s^Y) = \mathbb{E}(s^X)\mathbb{E}(s^Y) = G_X(s) \times G_Y(s)$, où on a utilisé que pour tout $s \in \mathbb{R}$ les variables aléatoires s^X et s^Y sont indépendantes. \square

Cette propriété s'avère extrêmement utile : elle permet de calculer la fonction génératrice d'une somme de variables aléatoires indépendantes.

Exemple 3.57. Soit $X \sim \text{Poi}(\lambda)$ et $Y \sim \text{Poi}(\mu)$ deux variables aléatoires indépendantes, où $\lambda, \mu > 0$. Alors en utilisant le calcul effectué dans l'Exemple 3.52, la fonction génératrice de $X + Y$ est égale à $G_{X+Y}(s) = G_X(s)G_Y(s) = e^{\lambda(s-1)}e^{\mu(s-1)} = e^{(\lambda+\mu)(s-1)}$. On reconnaît la fonction génératrice d'une loi de Poisson de paramètre $\lambda + \mu$: comme la fonction génératrice caractérise la loi des variables à valeurs dans \mathbb{N} , on a $X + Y \sim \text{Poi}(\lambda + \mu)$.

Exercice 3.58. Soit $X \sim \text{Bin}(n, p)$. Montrer que $G_X(s) = (1 + p(s-1))^n$ pour tout $s \in \mathbb{R}$.

Proposition 3.59. Soit X une variable aléatoire à valeurs dans \mathbb{N} , et soit G_X sa fonction génératrice. Alors X admet un moment d'ordre k fini si et seulement si G_X est k fois dérivable à gauche en 1, et si c'est le cas on a

$$\lim_{s \uparrow 1} G_X^{(k)}(s) = \mathbb{E}(X(X-1)\cdots(X-k+1)).$$

Avant de montrer ce résultat, soulignons que X admet un moment d'ordre k fini si et seulement si $X(X-1)\cdots(X-k)$ admet une espérance finie (exercice \dagger).

Démonstration. Tout d'abord, au vu de la dérivée (3.20), le théorème de convergence radiale d'Abel \ddagger assure que si $\mathbb{E}(X(X-1)\cdots(X-k+1)) = \sum_{n=k}^{+\infty} n(n-1)\cdots(n-k+1)p_X(n) < +\infty$,

\dagger . Procéder par récurrence pour montrer l'énoncé : " X admet un moment d'ordre n si et seulement si $P(X)$ admet une espérance finie pour tout polynôme P de degré n ".

\ddagger . **Théorème** (de convergence radiale d'Abel). Si $\sum_{n \in \mathbb{N}} a_n x^n$ converge en un point x_0 , alors la convergence est uniforme sur $[0, x_0]$; en particulier, $\lim_{x \uparrow x_0} \sum_{n \in \mathbb{N}} a_n x^n = \sum_{n \in \mathbb{N}} a_n x_0^n$.

alors

$$\lim_{s \uparrow 1} G_X^{(k)}(s) = \sum_{n=k}^{+\infty} n(n-1) \cdots (n-k+1) p_X(n) = \mathbb{E}(X(X-1) \cdots (X-k+1)),$$

ce qui démontre une première implication.

Pour l'autre implication, on utilise le lemme suivant, qui donne une réciproque (partielle) au théorème d'Abel, dans le cas d'une série entière à coefficients positifs.

Lemme 3.60. Soit $(a_n)_{n \geq 0}$ une suite de réels positifs, et soit $f(x) = \sum_{n=0}^{+\infty} a_n x^n$, de rayon de convergence $R > 0$. Alors f est croissante sur $[0, R[$, et la limite $\ell = \lim_{x \uparrow R} f(x)$ existe (et peut valoir $+\infty$). On a $\sum_n a_n R^n = \ell \in \mathbb{R}_+ \cup \{+\infty\}$.

Démonstration. Comme tous les termes sont positifs, pour tout $m \in \mathbb{N}$ et $0 \leq x \leq y \leq R$ on a $\sum_{n=0}^m a_n x^n \leq \sum_{n=0}^m a_n y^n$, de sorte qu'en passant à la limite $m \rightarrow \infty$, on a $f(x) \leq f(y)$; notons que $f(R) = \sum_n a_n R^n$ est bien défini (mais peut valoir $+\infty$) car tous les termes sont positifs. Cela montre que f est croissante sur $[0, R[$ et que on a $f(x) \leq f(R)$ pour tout $x \in [0, R[$: on obtient donc que $\ell = \lim_{x \uparrow R} f(x) \leq \sum_n a_n R^n$.

Réciproquement, pour tout $m \in \mathbb{N}$ et $x \in [0, R[$, on a $f(x) \geq \sum_{n=0}^m a_n x^n$: en prenant la limite quand $x \uparrow R$, on obtient $\ell = \lim_{x \uparrow R} f(x) \geq \sum_{n=0}^m a_n R^n$, quel que soit $m \in \mathbb{N}$. En prenant la limite quand $m \rightarrow \infty$ dans le terme de droite, on obtient $\ell \geq \sum_{n=0}^{\infty} a_n R^n$, ce qui conclut la démonstration. \square

D'après ce lemme, si la série dérivée $G_X^{(k)}(s)$ admet une limite finie à gauche, alors $\sum_{n=k}^{\infty} n(n-1) \cdots (n-k+1) p_X(n)$ converge, c'est-à-dire X admet un moment d'ordre k fini. \square

Exemple 3.61. Soit X une variable aléatoire de loi géométrique de paramètre $p \in]0, 1[$: $p_X(n) = (1-p)^{n-1} p$ pour tout $n \geq 1$. La fonction génératrice de X est

$$G_X(s) = \sum_{n=1}^{+\infty} p(1-p)^{n-1} s^n = \frac{ps}{1 - (1-p)s}$$

pour tout $|s| < \frac{1}{1-p}$. On a alors

$$G'_X(s) = \frac{p}{(1 - (1-p)s)^2}, \quad G''_X(s) = \frac{2p(1-p)}{(1 - (1-p)s)^3},$$

et donc $\mathbb{E}(X) = G'_X(1) = \frac{1}{p}$, $\mathbb{E}(X(X-1)) = G''_X(1) = \frac{2(1-p)}{p^2}$, qui donne $\text{Var}(X) = \frac{1-p}{p^2}$.

★ Une application : la marche aléatoire simple sur \mathbb{Z}

Considérons la marche aléatoire simple symétrique sur \mathbb{Z} : soit $(X_i)_{i \geq 1}$ une suite de variables aléatoires indépendantes, à valeurs dans $\{-1, 1\}$, de loi $\mathbb{P}(X_i = +1) = \mathbb{P}(X_i = -1) = \frac{1}{2}$, et posons $S_0 = 0$, et $S_n = \sum_{i=1}^n X_i$. On peut interpréter S_n comme la position d'un marcheur sur \mathbb{Z} après n pas, le i -ème pas étant donné par la variable aléatoire X_i . On introduit, pour $m \geq 0$, la variable aléatoire $T_m := \inf\{n \geq 0, S_n = m\}$, qui est le premier temps d'atteinte de m , avec la convention $\inf \emptyset = +\infty$. Cette variable aléatoire est compliquée, mais il s'avère que l'on peut calculer sa fonction génératrice (et donc en déduire sa densité discrète). Notons qu'a priori, T_m est à valeurs dans $\mathbb{N} \cup \{+\infty\}$: on va considérer la fonction $G_m : [0, 1] \rightarrow \mathbb{R}$ définie par

$$G_m(s) = \sum_{n=0}^{+\infty} \mathbb{P}(T_m = n) s^n. \quad (3.21)$$

Si $\mathbb{P}(T_m < +\infty) = 1$, il s'agit de la fonction génératrice de T_m . Dans tous les cas, on a $G_m(1) = \sum_{n=0}^{+\infty} \mathbb{P}(T_m = n) = \mathbb{P}(T_m < +\infty)$ (pourquoi?).

Théorème 3.62. On a $G_m(s) = G_1(s)^m$, avec $G_1(s) = \frac{1}{s}(1 - \sqrt{1 - s^2})$, pour tout $s \in]0, 1[$.

Ce théorème montre en particulier que $G_m(1) = 1$, et donc que $\mathbb{P}(T_m < +\infty) = 1$, pour tout $m \in \mathbb{N}$. Ainsi, presque sûrement, le marcheur finira par atteindre la position m , quel que soit $m \in \mathbb{N}$ (et même $m \in \mathbb{Z}$, par symétrie). On en déduit par exemple que $\mathbb{P}(\sup_{n \in \mathbb{N}} S_n \geq m) = 1$ pour tout $m \in \mathbb{N}$, ce qui par continuité des probabilités donne $\mathbb{P}(\sup_n S_n = +\infty) = 1$; et par symétrie $\mathbb{P}(\inf_n S_n = -\infty) = 1$. On en conclut que, presque sûrement, $\inf_{n \in \mathbb{N}} S_n = -\infty$ et $\sup_{n \in \mathbb{N}} S_n = +\infty$; et donc le marcheur visitera tous les sites de \mathbb{Z} , une infinité de fois §...

Démonstration. L'ingrédient principal est la relation de récurrence suivante entre les variables T_m et T_{m-1}, T_{m+1} : si $m \geq 1$, pour tout $n \geq 1$ on a

$$\mathbb{P}(T_m = n) = \frac{1}{2} \mathbb{P}(T_{m-1} = n-1) + \frac{1}{2} \mathbb{P}(T_{m+1} = n-1). \quad (3.22)$$

L'idée derrière cette identité est simple : on décompose l'événement $\{T_m = n\}$ suivant le premier pas du marcheur, ce qui donne

$$\mathbb{P}(T_m = n) = \mathbb{P}(X_1 = 1) \mathbb{P}(T_m = n \mid X_1 = 1) + \mathbb{P}(X_1 = -1) \mathbb{P}(T_m = n \mid X_1 = -1).$$

Maintenant, on peut écrire

$$\begin{aligned} \mathbb{P}(T_m = n \mid X_1 = 1) &= \mathbb{P}(X_1 + \dots + X_n = m, X_1 + \dots + X_j < m \text{ pour } j < n \mid X_1 = 1) \\ &= \mathbb{P}(X_2 + \dots + X_n = m-1, X_2 + \dots + X_j < m-1 \text{ pour } j < n \mid X_1 = 1) \\ &= \mathbb{P}(X_2 + \dots + X_n = m-1, X_2 + \dots + X_j < m-1 \text{ pour } j < n), \end{aligned}$$

où on a utilisé que X_1 et X_2, \dots, X_n sont indépendants pour enlever le conditionnement dans la deuxième égalité (on rappelle que $\mathbb{P}(A \mid B) = \mathbb{P}(A)$ si A et B sont indépendants). Maintenant, comme (X_2, \dots, X_n) a la même loi que (X_1, \dots, X_{n-1}) , on en déduit que

$$\begin{aligned} \mathbb{P}(T_m = n \mid X_1 = 1) &= \mathbb{P}(X_1 + \dots + X_{n-1} = m-1, X_1 + \dots + X_j < m-1 \text{ pour } j < n-1) \\ &= \mathbb{P}(T_{m-1} = n-1). \end{aligned}$$

De la même manière, on a $\mathbb{P}(T_m = n \mid X_1 = -1) = \mathbb{P}(T_{m+1} = n-1)$, ce qui permet d'obtenir (3.22).

Pour $m \geq 1$, on a $\mathbb{P}(T_m = 0) = 0$, de sorte le terme $n = 0$ est nul dans (3.21) : en appliquant la relation (3.22), on obtient pour tout $s \in]0, 1[$

$$\begin{aligned} G_m(s) &= \frac{1}{2} \sum_{n=1}^{+\infty} \mathbb{P}(T_{m-1} = n-1) s^n + \frac{1}{2} \sum_{n=1}^{+\infty} \mathbb{P}(T_{m+1} = n-1) s^n \\ &= \frac{s}{2} (G_{m-1}(s) + G_{m+1}(s)). \end{aligned}$$

Si on fixe $s \in]0, 1[$ et que l'on note $G_m := G_m(s)$, on a que la suite $(G_m)_{m \geq 0}$ satisfait la relation de récurrence $G_m = \frac{s}{2} G_{m-1} - G_{m-2}$, pour $m \geq 2$. On peut résoudre cette récursion : les solutions sont de la forme $(a\lambda_s^m + b\mu_s^m)_{m \geq 0}$, où $\lambda_s := \frac{1}{s}(1 - \sqrt{1 - s^2})$ et $\mu_s := \frac{1}{s}(1 + \sqrt{1 - s^2})$ sont les deux solutions de l'équation caractéristique $x^2 - \frac{s}{2}x + 1 = 0$. Il nous reste alors à déterminer a et b . Notons que $\mu_s > 1$ et que $\lambda_s \in]0, 1[$ quel que soit $s \in]0, 1[$ (exercice) : si on avait $b \neq 0$, on aurait $\lim_{m \rightarrow +\infty} |G_m| = +\infty$. Or, d'après la définition (3.21) on doit avoir $G_m \leq 1$ pour tout $m \in \mathbb{N}$: on en déduit donc que $b = 0$, et que G_m est de la forme $a\lambda_s^m$. Pour $m = 0$, on a $G_0 = 1$ (car

§. Si $x \in \mathbb{Z}$ n'est visité qu'un nombre fini de fois, alors S_n reste forcément d'un côté ou de l'autre de x à partir d'un certain rang.

$\mathbb{P}(T_0 = 0) = 1)$, et on en conclut donc que $a = 1$. On a donc montré que $G_m(s) = \lambda_s^m$ pour tout $s \in]0, 1[$, avec $\lambda_s = \frac{1}{s}(1 - \sqrt{1 - s^2})$.

Le fait que G_m soit continue en 0 vient de sa définition sous la forme d'une série entière ; notons qu'en $s = 0$, on peut prolonger λ_s en posant $\lambda_s = 0$. Le fait que G_m soit continue en 1 vient du théorème d'Abel, car la somme $\sum_{n=0}^{+\infty} \mathbb{P}(T_m = n)$ est absolument convergente (elle est majorée par 1). \square

Chapitre 4

Intermède : variables aléatoires réelles générales

4.1 Définition et propriétés

On peut étendre la Définition 3.1 de variable aléatoire discrète à une variable aléatoire à valeurs dans \mathbb{R} , de la façon suivante.

Définition 4.1 (Variable aléatoire réelle). *On appelle variable aléatoire réelle sur un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$ toute fonction $X : \Omega \rightarrow \mathbb{R}$ telle que pour tout intervalle I les ensembles $\{\omega, X(\omega) \in I\}$ soient des événements de \mathcal{F} .*

Cette définition est naturelle : on veut pouvoir calculer les probabilités d'événements du type “ X est positif” (en prenant $I = [0, +\infty[$) ou “ $|X|$ est strictement inférieur à 1” (en prenant $I =]-\infty, 1[$). Notons déjà qu’une variable aléatoire discrète, au sens de la Définition 3.1, et à valeurs dans $\mathcal{X} \subset \mathbb{R}$ dénombrable, est une variable aléatoire réelle au sens de la Définition 4.1. En effet, pour tout intervalle I , l’ensemble $\{X \in I\}$ s’écrit comme l’union dénombrable $\bigcup_{x \in I \cap \mathcal{X}} \{X = x\}$: d’après la définition d’une tribu, comme $\{X = x\}$ est un événement de \mathcal{F} pour tout $x \in \mathcal{X}$, on obtient que $\{X \in I\}$ est un événement de \mathcal{F} .

Remarque 4.2. Pour des variables aléatoires réelles $X : \Omega \rightarrow \mathbb{R}$ il suffit en fait de considérer les demi-droites $]-\infty, t]$: plus précisément, X est une variable aléatoire si (et seulement si) $\{X \leq t\} \in \mathcal{F}$ pour tout $t \in \mathbb{R}$. On ne le démontrera pas ici, mais l’idée est que l’on peut toujours exprimer un intervalle I à l’aide d’intervalles du type $]-\infty, t]$ (avec des passages au complémentaire, unions et/ou intersections dénombrables) : par exemple $]-\infty, t[= \bigcup_{r \in \mathbb{Q}, r < t}]-\infty, r]$, $[s, +\infty[=]-\infty, s]^c$, $[s, t] = [s, +\infty[\cap]-\infty, t]$, etc...

Remarque 4.3. On peut étendre cette définition à des variables aléatoires à valeurs dans \mathbb{R}^n (des vecteurs aléatoires), en requérant que pour tout n -uplet d’intervalles I_1, \dots, I_n les ensembles $\{X \in I_1 \times \dots \times I_n\}$ soient des événements de \mathcal{F} .

La Définition 4.1 assure que pour une variable aléatoire réelle X , les ensembles $\{X \in A\}$ sont des événements pour une large classe d’ensembles $A \subset \mathbb{R}$, appelés ensembles boréliens : l’ensemble $\mathcal{B}(\mathbb{R})$ des boréliens de \mathbb{R} est la “tribu engendrée” par les intervalles de \mathbb{R} (on rappelle la Définition 2.2 et la remarque qui la suit). On souligne cependant ici que tous les

ensembles ne sont pas boréliens ! Il s'agit d'une question subtile de théorie de la mesure, qui sera vue en L3. Une conséquence de cette subtilité est que, si X est une variable aléatoire réelle, et $g : \mathbb{R} \rightarrow \mathbb{R}$ est une fonction, $g(X)$ n'est pas forcément une variable aléatoire... Par contre, c'est le cas pour toutes les fonctions g avec lesquelles on travaille habituellement (par exemple les fonctions continues par morceaux), et on ne se posera donc ici pas trop de questions (de nouveau, on repousse cette subtilité au cours de théorie de la mesure en L3).

Loi d'une variable aléatoire réelle

Redonnons maintenant la définition de la loi d'une variable aléatoire réelle, qui étend celle d'une variable aléatoire discrète. On rappelle que $\mathcal{B}(\mathbb{R})$ est la tribu engendrée par les intervalles, appelée tribu borélienne.

Définition 4.4 (Loi d'une variable aléatoire). *Soit X une variable aléatoire réelle. On appelle loi de X la probabilité $\mu_X : \mathcal{B}(\mathbb{R}) \rightarrow [0, 1]$ définie par*

$$\mu_X(B) := \mathbb{P}(X \in B).$$

(La probabilité $\mathbb{P}(X \in B)$ est bien définie parce que $\{X \in B\} \in \mathcal{F}$ est un événement pour tout $B \in \mathcal{B}(\mathbb{R})$, par définition d'une variable aléatoire.)

Le fait que la loi μ_X est une probabilité sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ est laissée en exercice.

Remarque 4.5. De façon générale, tous les résultats du Chapitre 3 pour lesquels l'énoncé ne fait pas référence explicitement à la densité discrète de la variable aléatoire peuvent s'étendre à des variables aléatoires réelles.

Indépendance

La notion d'indépendance de variables aléatoires réelles est complètement analogue à la Définition 3.13.

Définition 4.6. *Soient X_1, X_2, \dots, X_n des variables aléatoires réelles, définies sur le même espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$. Elles sont dites indépendantes si pour tous intervalles I_1, \dots, I_n on a*

$$\mathbb{P}(X_1 \in I_1, X_2 \in I_2, \dots, X_n \in I_n) = \prod_{i=1}^n \mathbb{P}(X_i \in I_i).$$

Pour une famille arbitraire $(X_i)_{i \in I}$ de variables aléatoires, toutes définies sur le même espace probabilisé, celles-ci sont dites indépendantes si toute sous-famille finie est composée de variables aléatoires indépendantes.

Les propriétés de l'indépendance, comme l'indépendance par paquets ou le fait que des fonctions de variables aléatoires indépendantes sont indépendantes, continuent à être valables.

Espérance

La Définition 3.27 de l'espérance d'une variable aléatoire réelle discrète X reposait fortement sur la densité discrète de X , et ne s'étend donc pas directement à des variables

aléatoires réelles générales. On va commencer par définir l'espérance d'une variable aléatoire positive en l'*approchant* par une variable aléatoire discrète. On définit ensuite l'espérance pour une variable aléatoire générale en remarquant que l'on peut écrire $X = X^+ - X^-$, où $X^+ = \max\{X, 0\}$ et $X^- = \min\{-X, 0\}$ sont les parties positives et négatives de X .

Définition 4.7. Soit X une variable aléatoire réelle sur un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$.

- Si $X \geq 0$, on définit l'espérance de X comme la quantité

$$\mathbb{E}(X) := \sup \left\{ \mathbb{E}(Y) : Y : \Omega \rightarrow \mathbb{R} \text{ est une variable} \right. \\ \left. \text{aléatoire discrète telle que } 0 \leq Y \leq X \right\} \in [0, +\infty].$$

(En particulier, $\mathbb{E}(X)$ coïncide avec la formule (3.14) si X est discrète.)

- En général, on dit que X admet une espérance si au moins une des espérances $\mathbb{E}(X^+)$ et $\mathbb{E}(X^-)$ est finie, et si c'est le cas on définit

$$\mathbb{E}(X) := \mathbb{E}(X^+) - \mathbb{E}(X^-) \in [-\infty, +\infty].$$

Par conséquent, X admet une espérance finie si et seulement si les deux quantités $\mathbb{E}(X^+)$ et $\mathbb{E}(X^-)$ sont finies, ou de manière équivalente (par linéarité, voir plus bas) si $\mathbb{E}(|X|) = \mathbb{E}(X^+) + \mathbb{E}(X^-)$ est finie.

Toutes les propriétés de l'espérance (notamment la Proposition 3.35) et inégalités probabilistes énoncées dans le Chapitre 3, continuent à être vérifiées pour des variables aléatoires réelles générales. On rappelle en particulier la *monotonie*, la *positivité*, la *linéarité* de l'espérance :

- si $X \leq Y$ alors $\mathbb{E}(X) \leq \mathbb{E}(Y)$, pourvu que X et Y admettent une espérance ;
- si X admet une espérance, alors $|\mathbb{E}(X)| \leq \mathbb{E}(|X|)$; et X admet une espérance finie si et seulement si $\mathbb{E}(X) < \infty$;
- $\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y)$ si X et Y admettent une espérance finie, pour tout $a, b \in \mathbb{R}$; ou bien si X et Y sont positives, pour tout $a, b \geq 0$.

4.2 Fonction de répartition

Un objet très important pour l'étude des variables aléatoires réelles est la *fonction de répartition*.

Définition 4.8 (Fonction de répartition). Soit X une variable aléatoire réelle. On appelle fonction de répartition de X la fonction $F_X : \mathbb{R} \rightarrow [0, 1]$ définie par

$$F_X(x) := \mathbb{P}(X \leq x), \quad \forall x \in \mathbb{R}.$$

Exercice 4.9. Soient X et Y deux variables aléatoires réelles indépendantes. On définit $Z := \max\{X, Y\}$. Montrer que $F_Z(t) = F_X(t)F_Y(t)$ pour tout $t \in \mathbb{R}$.

On remarque que $F_X(x) = \mu_X(\cdot - \infty, x]$: la fonction de répartition F_X ne dépend donc que de la loi μ_X de la variable aléatoire X . Mais la réciproque est aussi vraie : la fonction de répartition d'une variable aléatoire réelle en caractérise la loi. La démonstration de ce résultat requiert des outils plus avancés de théorie de la mesure, qui seront vus en L3.

Proposition 4.10. *Deux variables aléatoires réelles X et Y qui ont la même fonction de répartition $F_X = F_Y$ ont la même loi : $\mu_X = \mu_Y$.*

Démontrons maintenant quelques propriétés de la fonction de répartition.

Proposition 4.11 (Propriétés de la fonction de répartition). *Soit F_X la fonction de répartition d'une variable aléatoire réelle X . Alors, on a les propriétés suivantes :*

1. F_X est croissante ;
2. F_X est continue à droite, c'est-à-dire $F_X(x) := \lim_{y \downarrow x} F_X(y)$ pour tout $x \in \mathbb{R}$;
3. $\lim_{x \rightarrow -\infty} F_X(x) = 0$;
4. $\lim_{x \rightarrow +\infty} F_X(x) = 1$.

On verra plus loin que toute fonction $F : \mathbb{R} \rightarrow \mathbb{R}$ qui possède ces propriétés est la fonction de répartition d'une variable aléatoire, cf. Théorème 5.16.

Démonstration. 1. Si $x \leq y$ alors $\{X \leq x\} \subseteq \{X \leq y\}$, et donc $F_X(x) = \mathbb{P}(X \leq x) \leq \mathbb{P}(X \leq y) = F_X(y)$.

2. Soit $x \in \mathbb{R}$. La fonction F_X étant croissante, on sait que la limite $\lim_{h \downarrow 0} F_X(x + h)$ existe : par unicité de la limite, on a $\lim_{h \downarrow 0} F_X(x + h) = \lim_{n \rightarrow \infty} F_X(x + \frac{1}{n})$. On observe alors que $\{X \leq x\} = \bigcap_{n \in \mathbb{N}} \{X \leq x + \frac{1}{n}\}$, et que $\{X \leq x + \frac{1}{n+1}\} \subseteq \{X \leq x + \frac{1}{n}\}$ pour tout $n \geq 1$. D'après la continuité par le haut des probabilités (Proposition 2.8), on a donc

$$F_X(x) = \mathbb{P}(X \leq x) = \lim_{n \rightarrow +\infty} \mathbb{P}(X \leq x + \frac{1}{n}) = \lim_{n \rightarrow +\infty} F_X(x + \frac{1}{n}) = \lim_{h \downarrow 0} F_X(x + h).$$

3. De même que précédemment, la fonction F_X étant croissante, on sait que la limite $\lim_{x \rightarrow -\infty} F_X(x)$ existe, et est égale à $\lim_{n \rightarrow \infty} F_X(-n)$, par unicité de la limite. Considérons la famille décroissante d'événements $\{X \leq -n\}$, et observons que $\bigcap_{n \in \mathbb{N}} \{X \leq -n\} = \emptyset$, parce qu'il n'existe aucun réel $x = X(\omega)$ tel que $x \leq -n$ pour tout $n \in \mathbb{N}$. D'après la continuité par le haut des probabilités (Proposition 2.8), on obtient que

$$0 = \mathbb{P}(\emptyset) = \lim_{n \rightarrow \infty} \mathbb{P}(X \leq -n) = \lim_{n \rightarrow \infty} F_X(-n) = \lim_{x \rightarrow -\infty} F_X(x).$$

4. La famille d'événements $\{X \leq n\}$ est croissante et $\bigcup_{n \in \mathbb{N}} \{X \leq n\} = \Omega$, donc d'après la continuité par le bas des probabilités, on a de la même manière que précédemment

$$1 = \lim_{n \rightarrow \infty} \mathbb{P}(X \leq n) = \lim_{n \rightarrow \infty} F_X(n) = \lim_{x \rightarrow \infty} F_X(x).$$

□

Proposition 4.12. *Pour tout $x \in \mathbb{R}$, on a*

$$\mathbb{P}(X = x) = F_X(x) - F_X(x^-), \quad \forall x \in \mathbb{R},$$

où $F_X(x^-) := \lim_{y \uparrow x} F_X(y)$ désigne la limite à gauche de la fonction F_X au point x , qui existe par monotonie de F_X .

En particulier, si X est une variable aléatoire discrète, la fonction de répartition détermine complètement sa densité discrète ! On a que F_X est discontinue en x si et seulement si $\mathbb{P}(X = x) > 0$, et la valeur de $\mathbb{P}(X = x)$ représente la hauteur du “saut” de F_X en x .

Démonstration. Si $y < x$, on a $\{X \in]y, x]\} = \{X \leq x\} \setminus \{X \leq y\}$, et ainsi

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

La suite d'événements $\{X \in]x - \frac{1}{n}, x]\}$ est décroissante et est telle que $\{X = x\} = \bigcap_{n \in \mathbb{N}} \{X \in]x - \frac{1}{n}, x]\}$. Ainsi, d'après la continuité par le haut des probabilités (Proposition 2.8), et en utilisant le fait que la limite $F_X(x^-) := \lim_{h \downarrow 0} F_X(x - h)$ existe et est égale à $\lim_{n \rightarrow \infty} F_X(x - \frac{1}{n})$, on obtient

$$\begin{aligned} \mathbb{P}(X = x) &= \lim_{n \rightarrow +\infty} \mathbb{P}(X \in]x - \frac{1}{n}, x]) = \lim_{n \rightarrow +\infty} \left(F_X(x) - F_X\left(x - \frac{1}{n}\right) \right) \\ &= F_X(x) - \lim_{n \rightarrow +\infty} F_X\left(x - \frac{1}{n}\right) = F_X(x) - F_X(x^-), \end{aligned}$$

ce qui donne le résultat. □

4.3 Fonction génératrice des moments

Un outil important pour l'étude de variables aléatoires réelles, similaire à la fonction génératrice pour les variables à valeurs dans \mathbb{N} définie dans la Section 3.6, est la fonction génératrice des moments.

Définition 4.13 (Fonction génératrice des moments). Soit X une variable aléatoire réelle. La fonction $M_X : \mathbb{R} \rightarrow \mathbb{R}_+ \cup \{+\infty\}$ définie par $M_X(t) = \mathbb{E}(e^{tX})$ est appelée fonction génératrice des moments de la variable aléatoire X .

Soulignons que la fonction génératrice des moments d'une variable aléatoire est toujours bien définie (car e^{tX} est une variable aléatoire positive), mais peut prendre la valeur $+\infty$. On a le lien suivant entre la fonction génératrice des moments et la fonction génératrice : $M_X(t) = G_X(e^t)$. En particulier, la fonction génératrice des moments caractérise la loi d'une variable aléatoire à valeurs dans \mathbb{N} , mais on souligne que ce n'est pas le cas de manière générale ! On peut cependant l'utiliser pour calculer les moments d'une variable aléatoire, pourvu qu'on puisse appliquer le résultat suivant.

Théorème 4.14. Soit X une variable aléatoire réelle, et M_X sa fonction génératrice des moments. S'il existe $a > 0$ tel que $M_X(t) < +\infty$ pour tout $t \in]-a, a[$, alors :

1. la variable aléatoire X admet un moment d'ordre k fini pour tout $k \in \mathbb{N}$, et pour tout $t \in]-a, a[$, on a

$$M_X(t) = \sum_{k=0}^{\infty} \mathbb{E}(X^k) \frac{t^k}{k!};$$

2. la fonction M_X est infiniment dérivable sur $]-a, a[$ et, en notant $M_X^{(k)}$ la dérivée k -ème de M_X , on a

$$M_X^{(k)}(0) = \mathbb{E}(X^k), \quad \forall k \in \mathbb{N}.$$

Démonstration. Commençons par quelques considérations préliminaires. Tout d'abord, en utilisant que pour tout $x \in \mathbb{R}$ on a $e^{|x|} \leq e^{tx} + e^{-tx}$, on obtient que pour tout $t \in]-a, a[$

$$M_{|X|}(t) = \mathbb{E}(e^{t|X|}) \leq \mathbb{E}(e^{tX}) + \mathbb{E}(e^{-tX}) = M_X(t) + M_X(-t) < +\infty.$$

De plus, comme $e^u \geq \frac{u^k}{k!}$ pour tout $u \geq 0$ et $k \geq 0$, on a $|X|^k \leq k! h^{-k} e^{h|X|}$ pour tout $h > 0$: pour $t \in]-a, a[$, on peut choisir $h > 0$ tel que $|t| + h \in]-a, a[$, et on obtient que pour tout $k \geq 0$

$$\mathbb{E}\left(|X|^k e^{tX}\right) \leq k! h^{-k} \mathbb{E}\left(e^{(|t|+h)|X|}\right) = k! h^{-k} M_{|X|}(|t| + h) < +\infty. \quad (4.1)$$

Point 1. Le fait que $\mathbb{E}(|X|^k) < +\infty$ pour tout $k \in \mathbb{N}$ découle directement de l'inégalité précédente (en prenant $t = 0$). Maintenant, on peut écrire, pour tout $t \in]-a, a[$, en utilisant la linéarité de l'espérance

$$\begin{aligned} \left| M_X(t) - \sum_{k=0}^n \mathbb{E}(X^k) \frac{t^k}{k!} \right| &= \left| \mathbb{E}\left(e^{tX} - \sum_{k=0}^n X^k \frac{t^k}{k!}\right) \right| \\ &\leq \mathbb{E}\left(\left| e^{tX} - \sum_{k=0}^n X^k \frac{t^k}{k!} \right|\right) = \mathbb{E}\left(\left| \sum_{k=n+1}^{\infty} X^k \frac{t^k}{k!} \right|\right) \end{aligned}$$

où on a aussi utilisé le développement de e^{tX} . Maintenant, pour tout $x \in \mathbb{R}$ on a

$$\left| \sum_{k=n+1}^{\infty} \frac{x^k}{k!} \right| \leq \sum_{k=n+1}^{\infty} \frac{|x|^k}{k!} \leq \frac{|x|}{n+1} \sum_{k=n+1}^{\infty} \frac{|x|^{k-1}}{(k-1)!} \leq \frac{|x|}{n+1} e^{|x|},$$

où on a utilisé que $\frac{1}{k!} \leq \frac{1}{n+1} \frac{1}{(k-1)!}$ pour tout $k \geq n+1$. On a donc, en appliquant cette relation à $x = tX$,

$$\left| M_X(t) - \sum_{k=0}^n \mathbb{E}(X^k) \frac{t^k}{k!} \right| \leq \mathbb{E}\left(\left| \sum_{k=n+1}^{\infty} X^k \frac{t^k}{k!} \right|\right) \leq \frac{1}{n+1} \mathbb{E}\left(|t||X| e^{|tX|}\right).$$

Comme la dernière espérance est finie pour $t \in]-a, a[$ d'après (4.1), le membre de gauche tend vers 0 quand $n \rightarrow \infty$, ce qui conclut la démonstration du premier point.

Point 2. Maintenant, comme on a démontré que M_X est donnée par la *série entière* (??), le théorème de dérivation des séries entières nous donne directement que M_X est k fois dérivable sur $]-a, a[$, et que

$$M_X^{(n)}(t) = \mathbb{E}(X^n e^{tX}).$$

Le résultat voulu s'ensuit en prenant $t = 0$. □

Résumé du chapitre

Une variable aléatoire réelle X , définie sur un espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$, peut prendre une quantité plus que dénombrable de valeurs de \mathbb{R} , et donc ne pas être discrète. La loi de X est toujours déterminée par la fonction de répartition $F_X(x) := \mathbb{P}(X \leq x)$, dont la connaissance permet en principe de calculer toutes les quantités d'intérêt relatives à X . Les notions d'indépendance et d'espérance peuvent être étendues à des variables aléatoires générales, et des propriétés analogues à celles vues dans le cas discret sont vérifiées.

Chapitre 5

Variables aléatoires à densité

On introduit ici une famille importante de variables aléatoires, qui peuvent prendre une infinité continue de valeurs, typiquement dans un intervalle $]a, b[\subseteq \mathbb{R}$, et qui possèdent la propriété que leur loi est déterminée par une fonction $f_X : \mathbb{R} \rightarrow \mathbb{R}$, appelée *densité*, qui permet le calcul de toutes les quantités d'intérêt.

Beaucoup de propriétés des *variables aléatoires à densité* (on dit aussi *absolument continues*) sont analogues à celles des variables aléatoires discrètes, la densité f_X jouant le même rôle que la densité discrète p_X , les intégrales remplaçant les sommes. Il faut cependant faire attention à ne pas tirer de généralisations hasardeuses (voir notamment (5.3)).

Dans tout le reste du chapitre, on considérera $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé, sur lequel seront définies les variables aléatoires considérées.

5.1 Variables aléatoires absolument continues

5.1.1 Définition et premières propriétés

Définition 5.1 (Variables aléatoires absolument continues). *Une variable aléatoire réelle X est dite absolument continue ou à densité s'il existe une fonction positive $f_X : \mathbb{R} \rightarrow [0, +\infty[$, intégrable sur \mathbb{R} , telle que la fonction de répartition de X peut s'écrire sous la forme*

$$F_X(t) = \mathbb{P}(X \leq t) = \int_{-\infty}^t f_X(x) dx, \quad \forall t \in \mathbb{R}. \quad (5.1)$$

Une telle fonction f_X est appelé densité de X .

En prenant la limite $t \rightarrow +\infty$ dans la relation (5.1), la densité f_X doit donc vérifier

$$\int_{-\infty}^{+\infty} f_X(x) dx = 1. \quad (5.2)$$

On souligne que toute fonction positive $f : \mathbb{R} \rightarrow [0, +\infty[$ intégrable sur \mathbb{R} et telle que $\int_{-\infty}^{+\infty} f(x) dx = 1$ est la densité d'une variable aléatoire absolument continue. En effet, la

fonction $F(x) := \int_{-\infty}^x f(t) dt$ possède les propriétés d'une fonction de répartition (cf. Proposition 4.11) : le Théorème 5.16 plus loin garantit qu'il existe une variable aléatoire X dont la fonction de répartition est $F_X = F$, donc qui possède pour densité f_X .

Exemple 5.2. On considère l'espace d'état $\Omega = [0, 1]$, et une probabilité \mathbb{P} telle que pour tout $0 \leq a \leq b \leq 1$ on ait $\mathbb{P}([a, b]) = b - a$ * : cet espace probabilisé correspond à l'expérience qui consiste à choisir "au hasard" un réel x dans l'intervalle $[0, 1]$... Soit X la variable aléatoire définie par $X(\omega) = \omega$ pour tout $\omega \in [0, 1]$. On peut calculer explicitement sa fonction de répartition : $F_X(t) = 0$ si $t < 0$ (car alors $\{X \leq t\} = \emptyset$); $F_X(t) = \mathbb{P}(X \in [0, t]) = t$ si $t \in [0, 1]$, et $F_X(t) = 1$ si $t > 1$ (car alors $\{X \leq t\} = \Omega$). On vérifie alors facilement que $F_X(t)$ s'écrit sous la forme $\int_{-\infty}^t f_X(x) dx$ où la fonction $f : \mathbb{R} \rightarrow \mathbb{R}$ est définie par

$$f(x) := \mathbb{1}_{[0,1]}(x) = \begin{cases} 1 & \text{si } 0 \leq x \leq 1, \\ 0 & \text{sinon.} \end{cases}$$

Soulignons que si X est une variable aléatoire absolument continue, sa densité f_X n'est pas déterminée de manière unique. En effet, si f_X est une densité de X , alors pour toute fonction g obtenue à partir de f_X en modifiant sa valeur sur un ensemble fini de points, on a que l'intégrale dans (5.1) n'est pas modifiée, et donc g est aussi une densité de X . Ainsi, dans l'Exemple 5.2, la fonction $x \mapsto \mathbb{1}_{]0,1[}(x)$ est aussi une densité de la variable aléatoire X .

Si X est une variable aléatoire à densité, sa fonction de répartition $F_X(x) = \int_{-\infty}^x f_X(t) dt$ est une fonction continue (pourquoi?). On a donc $F_X(x) = F_X(x^-)$ pour tout $x \in \mathbb{R}$, et comme $F_X(x) - F_X(x^-) = \mathbb{P}(X = x)$ (cf. Proposition 4.12), on obtient que

$$\mathbb{P}(X = x) = 0, \quad \forall x \in \mathbb{R}. \quad (5.3)$$

Pour tout $a, b \in [-\infty, +\infty]$ avec $a \leq b$, on a $\mathbb{P}(X \in]a, b]) = \mathbb{P}(X \leq b) - \mathbb{P}(X \leq a) = F_X(b) - F_X(a) = \int_a^b f_X(t) dt$. Ainsi, grâce à (5.3), on a

$$\begin{aligned} \mathbb{P}(X \in [a, b]) &= \mathbb{P}(X \in]a, b]) = \mathbb{P}(X \in [a, b[) = \mathbb{P}(X \in]a, b[) \\ &= F_X(b) - F_X(a) = \int_a^b f_X(x) dx. \end{aligned} \quad (5.4)$$

Si $f_X(x) = 0$ pour tout $x \in]a, b[$, alors $\mathbb{P}(X \in]a, b[) = 0$: les intervalles où la densité est nulle représentent des valeurs qui ne sont (presque sûrement) pas prises par X .

Remarque 5.3. L'identité (5.3) montre que la densité $f_X(x)$ d'une variable aléatoire X absolument continue n'est pas la probabilité que la variable aléatoire prenne la valeur x , c'est-à-dire $f_X(x) \neq \mathbb{P}(X = x)$.

5.1.2 Calculer une densité

Une question typique consiste à déterminer si une variable aléatoire X est absolument continue, et si c'est le cas, à en calculer une densité.

*. En réalité \mathbb{P} est définie sur une tribu \mathcal{F} qui contient les intervalles (la tribu borélienne...), et le fait qu'une telle probabilité existe est non trivial, et sera vu en L3!

Proposition 5.4. *Soit X une variable aléatoire réelle. Si la fonction de répartition F_X est continue et \mathcal{C}^1 par morceaux^a, alors X est absolument continue, de densité $f_X(x) := F'_X(x)$ (définie arbitrairement aux points où F_X n'est pas dérivable).*

a. Une fonction $f : \mathbb{R} \rightarrow \mathbb{R}$ est \mathcal{C}^1 par morceaux si la restriction de f à n'importe quel intervalle fermé borné $[a, b]$ est \mathcal{C}^1 par morceaux.

Démonstration. Il s'agit d'un résultat classique d'analyse, qui énonce que si F est \mathcal{C}^1 par morceaux, alors $F(b) - F(a) = \int_a^b f(x)dx$, où $f(x) = F'(x)$. Comme $\lim_{a \rightarrow -\infty} F_X(a) = 0$ d'après la Proposition 4.11, on trouve que $F(b) = \int_{-\infty}^b f(x)dx$ pour tout $b \in \mathbb{R}$. \square

Exemple 5.5. Soit X une variable aléatoire *uniforme* sur $[0, 1]$, comme introduite dans l'Exemple 5.2. On pose $Y = X^2$: montrons que Y est une variable aléatoire à densité, et déterminons cette densité.

On rappelle que la fonction de répartition de X vaut $F_X(t) = 0$ pour $t < 0$, $F_X(t) = t$ pour $t \in [0, 1]$ et $F_X(t) = 1$ pour $t > 1$. Déterminons maintenant la fonction de répartition de Y . Tout d'abord, on remarque que l'on a encore $Y \in [0, 1]$: on a donc $F_Y(y) = 0$ pour $y < 0$ (car alors $\{Y \leq y\} = \emptyset$) et $F_Y(y) = 1$ pour $y > 1$ (car alors $\{Y \leq y\} = \Omega$). Pour $y \in [0, 1]$, on a l'égalité d'événements $\{Y \leq y\} = \{X^2 \leq y\} = \{-\sqrt{y} \leq X \leq \sqrt{y}\}$: on obtient donc que $F_Y(y) = F_X(\sqrt{y}) - F_X(-\sqrt{y}) = F_X(\sqrt{y})$, car $F_X(-\sqrt{y}) = 0$. On obtient finalement que $F_Y(y) = 0$ si $y < 0$, $F_Y(y) = \sqrt{y}$ si $y \in [0, 1]$ et $F_Y(y) = 1$ si $y > 1$. La fonction F_Y est continue sur tout \mathbb{R} . De plus, elle est dérivable, de dérivée continue, sur les intervalles $]-\infty, 0[$, $]0, 1[$ et $]1, +\infty[$. Ainsi, F_Y est continue et \mathcal{C}^1 par morceaux, et on en conclut que la variable aléatoire Y est absolument continue, de densité $f_Y(y) = F'_Y(y) = \frac{1}{2\sqrt{y}} \mathbb{1}_{]0, 1[}(y)$.

Exercice 5.6. Soit X une variable aléatoire réelle absolument continue, de densité f_X . Pour tout $a, b \in \mathbb{R}$ avec $a \neq 0$, montrer que la variable aléatoire $Y := aX + b$ est absolument continue, et a pour densité $f_Y(y) = \frac{1}{|a|} f_X(\frac{y-b}{a})$.

Exercice 5.7. Soit X une variable aléatoire uniforme sur $[0, 1]$, définie dans l'Exemple 5.2. On pose $Y = X \mathbb{1}_{\{X \leq 1/2\}}$. Calculer la fonction de répartition de Y , et montrer que Y n'est ni absolument continue, ni discrète.

5.1.3 Espérance et formule de transfert

Si X est une variable aléatoire absolument continue de densité f_X , on va énoncer un résultat qui va nous permettre de calculer l'espérance $\mathbb{E}(g(X))$, où $g : \mathbb{R} \rightarrow \mathbb{R}$ est une fonction continue par morceaux (de sorte que $g(X)$ est bien une variable aléatoire). Notons que si g est positive, l'intégrale sur \mathbb{R} de la fonction $g(x)f_X(x)$ est toujours bien définie (et vaut éventuellement $+\infty$).

Proposition 5.8 (Formule de transfert, cas positif). *Soit X une variable aléatoire réelle absolument continue, de densité f_X , et soit $g : \mathbb{R} \rightarrow [0, +\infty[$ une fonction positive et continue par morceaux. L'espérance de la variable aléatoire positive $g(X)$ est toujours un élément bien défini de $[0, +\infty]$ et est donnée par*

$$\mathbb{E}(g(X)) = \int_{-\infty}^{+\infty} g(x)f_X(x) dx. \quad (5.5)$$

Exemple 5.9. Si X est une variable aléatoire réelle absolument continue, de densité f_X , la fonction génératrice des moments de X est donnée par l'intégrale

$$M_X(t) = \mathbb{E}(e^{tX}) = \int_{-\infty}^{+\infty} e^{tx} f_X(x) dx.$$

Dans le cas d'une fonction $g : \mathbb{R} \rightarrow \mathbb{R}$ non nécessairement positive, on peut appliquer la Proposition 5.8 aux fonctions positives g^+ , g^- et $|g|$: on obtient les résultats suivants.

Proposition 5.10 (Formule de transfert, cas intégrable). *Soit X une variable aléatoire réelle absolument continue, de densité f_X , et soit $g : \mathbb{R} \rightarrow \mathbb{R}$ une fonction continue par morceaux. La variable aléatoire $g(X)$ admet une espérance finie si et seulement si la fonction $|g(x)| f_X(x)$ est intégrable sur \mathbb{R} :*

$$\mathbb{E}(|g(X)|) < +\infty \quad \Longleftrightarrow \quad \int_{-\infty}^{+\infty} |g(x)| f_X(x) dx < +\infty,$$

et dans ce cas $\mathbb{E}(g(X))$ est donné par la formule (5.5).

Corollaire 5.11. *Une variable aléatoire réelle X absolument continue, de densité f_X , admet un moment d'ordre k fini si et seulement si $\mathbb{E}(|X|^k) = \int_{-\infty}^{+\infty} |x|^k f_X(x) dx < +\infty$, et dans ce cas, le moment d'ordre k de X est donné par l'intégrale*

$$\mathbb{E}(X^k) = \int_{-\infty}^{+\infty} x^k f_X(x) dx.$$

En particulier, si X est une variable aléatoire à densité qui admet un moment d'ordre 2, cette formule permet de calculer $\text{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$.

Exercice 5.12. Soit X une variable aléatoire de densité donnée par $f_X(x) = \mathbb{1}_{]0,1[}(x)$ pour $x \in \mathbb{R}$. Montrer que $\mathbb{E}(X^k) = \frac{1}{k+1}$ pour tout $k \geq 1$.

La formule de transfert (5.5) fournit une autre méthode pour calculer les densités de variables aléatoires absolument continues.

Proposition 5.13. *Une variable aléatoire X est absolument continue et admet pour densité f_X si et seulement si pour toute fonction $h : \mathbb{R} \rightarrow \mathbb{R}$ continue et bornée, on a la formule de transfert (5.5)*

$$\mathbb{E}(h(X)) = \int_{-\infty}^{+\infty} h(x) f_X(x) dx. \quad (5.6)$$

Démonstration. Si X est absolument continue et a pour densité f_X , alors si $h : \mathbb{R} \rightarrow \mathbb{R}$ est une fonction continue par morceaux et bornée quelconque, la fonction $|h(x)| f_X(x)$ est intégrable car $|h(x)| f_X(x) \leq M f_X(x)$ pour une certaine constante M , et la formule (5.6) découle directement de la Proposition 5.10.

Réciproquement, si la formule (5.6) est vraie pour toutes les fonctions continues et bornées, alors montrons que pour tout $t \in \mathbb{R}$

$$F_X(t) = \mathbb{P}(X \leq t) = \int_{-\infty}^t f_X(x) dx. \quad (5.7)$$

Cela serait facile si on pouvait prendre la fonction $h(x) = \mathbb{1}_{]-\infty, t]}(x)$ dans (5.6), mais elle n'est pas continue ! Pour $t \in \mathbb{R}$ fixé, et pour tout $\varepsilon > 0$, on définit donc la fonction *continue* h_ε comme une fonction affine par morceaux qui vaut 1 sur $]-\infty, t]$ et 0 sur $[t+\varepsilon, +\infty[$: $h_\varepsilon(x) = 1$ pour $x \in]-\infty, t]$, $h_\varepsilon(x) = 0$ pour $x \in [t+\varepsilon, +\infty[$, et $h_\varepsilon(x) = 1 - \frac{1}{\varepsilon}(x-t)$ pour $x \in]t, t+\varepsilon[$. La fonction h_ε est une "approximation continue" de la fonction indicatrice $\mathbb{1}_{]-\infty, t]}$. Notons que l'on a $\mathbb{1}_{]-\infty, t]}(x) \leq h_\varepsilon(x) \leq \mathbb{1}_{]-\infty, t+\varepsilon]}(x)$ pour tout $x \in \mathbb{R}$: par monotonie de l'espérance, on a

$$F_X(t) = \mathbb{P}(X \leq t) \leq \mathbb{E}(h_\varepsilon(X)) \leq \mathbb{P}(X \leq t + \varepsilon) = F_X(t + \varepsilon),$$

où on a utilisé que $\mathbb{E}(\mathbb{1}_{]-\infty, a]}(X)) = \mathbb{P}(X \leq a)$ pour tout $a \in \mathbb{R}$. Lorsque $\varepsilon \downarrow 0$, comme la fonction de répartition est continue à droite on a $\lim_{\varepsilon \downarrow 0} F_X(t + \varepsilon) = F_X(t)$, de sorte que l'on obtient

$$\mathbb{P}(X \leq t) = \lim_{\varepsilon \downarrow 0} \mathbb{E}(h_\varepsilon(X)).$$

Comme h_ε est continue et bornée, elle vérifie (5.6), et on a donc

$$\mathbb{E}(h_\varepsilon(X)) = \int_{-\infty}^{+\infty} h_\varepsilon(x) f_X(x) dx = \int_{-\infty}^t f_X(x) dx + \int_t^{t+\varepsilon} (1 - \frac{1}{\varepsilon}(x-t)) f_X(x) dx,$$

avec

$$0 \leq \int_t^{t+\varepsilon} (1 - (x-t)/\varepsilon) f_X(x) dx \leq \int_t^{t+\varepsilon} f_X(x) dx \rightarrow 0 \quad \text{quand } \varepsilon \downarrow 0,$$

où pour la dernière limite on a utilisé le fait que $f_X(x)$ est intégrable. Cela démontre donc l'identité (5.7), ce qui, d'après la Définition 5.1, montre que X est absolument continue et a pour densité f_X . \square

Exemple 5.14. Reprenons l'Exemple 5.5 : soit X une variable aléatoire de densité $f_X(x) = \mathbb{1}_{]0,1]}(x)$, et déterminons la densité de $Y = X^2$. Pour cela, on se fixe une fonction $h : \mathbb{R} \rightarrow \mathbb{R}$ continue bornée quelconque (une *fonction test* pour la formule de transfert), et on essaie de calculer $\mathbb{E}(h(Y))$.

Comme $Y = X^2$, on peut appliquer la formule de transfert (5.5) à X , pour la fonction $g(x) = h(x^2)$: on obtient donc

$$\mathbb{E}(h(Y)) = \mathbb{E}(h(X^2)) = \int_{-\infty}^{\infty} h(x^2) \cdot \mathbb{1}_{]0,1]}(x) dx = \int_0^1 h(x^2) dx,$$

où on a utilisé que l'indicatrice vaut 1 sur l'intervalle $[0, 1]$ et 0 en dehors. Comme on cherche à obtenir une formule du type (5.6), on va effectuer le changement de variable $y = x^2$ ($x = \sqrt{y}$) : on obtient

$$\mathbb{E}(h(Y)) = \int_0^1 h(y) \frac{1}{2\sqrt{y}} dy = \int_{-\infty}^{+\infty} h(y) \underbrace{\frac{1}{\sqrt{y}} \mathbb{1}_{]0,1]}(y)}_{\text{densité}} dy,$$

où on a rajouté l'indicatrice de l'intervalle $]0, 1[$ de façon à faire porter l'intégrale sur \mathbb{R} tout entier. D'après la Proposition 5.13, cette relation étant valable pour une fonction arbitraire $h : \mathbb{R} \rightarrow \mathbb{R}$ continue et bornée, on obtient que Y est absolument continue, et que la fonction $\frac{1}{2\sqrt{y}} \mathbb{1}_{]0,1]}(y)$ en est sa densité.

5.2 Exemples essentiels de variables aléatoires absolument continues

5.2.1 Loi uniforme continue

Soient $a, b \in \mathbb{R}$ avec $a < b$. Une variable aléatoire réelle absolument continue X est dite *uniforme sur $]a, b[$* , et on écrira $X \sim \mathcal{U}(a, b)$, si elle a pour densité

$$f_X(x) = \frac{1}{b-a} \mathbb{1}_{]a, b[}(x). \quad (5.8)$$

On vérifie facilement que cette fonction est effectivement une densité, c'est-à-dire qu'on a $\int_{-\infty}^{+\infty} f_X(x) dx = 1$. Notons que l'on peut remplacer l'intervalle ouvert $]a, b[$ par l'intervalle fermé $[a, b]$ dans la densité : cela ne change pas la fonction de répartition ni donc la loi.

Une variable aléatoire $X \sim \mathcal{U}(a, b)$ s'interprète comme "un point choisi uniformément dans l'intervalle $]a, b[$ ", parce que la valeur de la "densité de probabilité" $f_X(x)$ est la même, strictement positive, pour tout $x \in]a, b[$, et nulle en dehors de cet intervalle.

On peut calculer l'espérance et la variance de $X \sim \mathcal{U}(a, b)$. Comme la densité f_X est nulle en dehors de l'intervalle $]a, b[$, on a $\mathbb{P}(X \in]a, b[) = 1$, ce qui signifie que X est presque sûrement bornée ($|X| \leq \max\{|a|, |b|\}$), et donc admet des moments finis de tout ordre. En appliquant la formule de transfert, on obtient

$$\mathbb{E}(X) = \frac{1}{b-a} \int_a^b x dx = \frac{a+b}{2}, \quad \mathbb{E}(X^2) = \frac{1}{b-a} \int_a^b x^2 dx = \frac{a^2 + ab + b^2}{3},$$

d'où $\text{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \frac{(b-a)^2}{12}$. La fonction de répartition est elle aussi explicite :

$$F_X(t) = \int_{-\infty}^t f_X(x) dx = \begin{cases} 0 & \text{si } t \leq a, \\ \frac{t-a}{b-a} & \text{si } t \in]a, b[, \\ 1 & \text{si } t \geq b. \end{cases} \quad (5.9)$$

Exercice 5.15. Soit $U \sim \mathcal{U}(a, b)$, et soit $\alpha > 0, \beta \in \mathbb{R}$. On pose $V = \alpha U + \beta$. Montrer que V suit une loi uniforme, sur un intervalle dont on déterminera les bornes.

On énonce maintenant un théorème très important, qui permet de construire une variable aléatoire de fonction de répartition donnée, à partir d'une variable aléatoire $X \sim \mathcal{U}(0, 1)$. Cela montre en particulier que toute fonction qui vérifie les propriétés énoncées dans la Proposition 4.11 est la fonction de répartition d'une variable aléatoire.

Théorème 5.16. Soit $F : \mathbb{R} \rightarrow \mathbb{R}$ une fonction croissante, continue à droite et telle que $\lim_{x \rightarrow -\infty} F_X(x) = 0$ et $\lim_{x \rightarrow +\infty} F_X(x) = 1$. Alors on définit le pseudo-inverse de F comme la fonction $\varphi :]0, 1[\rightarrow \mathbb{R}$ définie par :

$$\varphi(x) := \inf\{y \in \mathbb{R} : F(y) \geq x\}.$$

Soit $X \sim \mathcal{U}(0, 1)$, alors $Y := \varphi(X)$ est une variable aléatoire dont la fonction de répartition est $F_Y = F$.

Démonstration. Calculons la fonction de répartition $F_Y(t) = \mathbb{P}(Y \leq t) = \mathbb{P}(\varphi(X) \leq t)$ pour tout $t \in \mathbb{R}$. La clé de ce théorème est que l'on a l'égalité d'événements

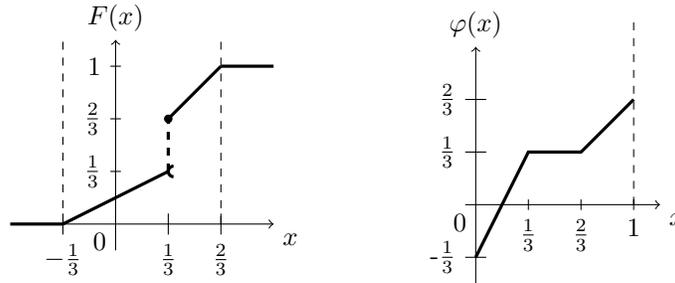
$$\{\varphi(X) \leq t\} = \{F(t) \geq X\},$$

qui vient du fait que $\varphi(x) \leq t$ si et seulement si $F(x) \geq t$. En effet, si $F(t) \geq x$, il découle immédiatement de la définition de φ que $\varphi(x) \leq t$. Réciproquement, si $F(t) < x$, comme F est une fonction continue à droite, il existe $\varepsilon > 0$ tel que $F(t + \varepsilon) < x$; par monotonie, $F(y) < x$ pour tout $y \leq t + \varepsilon$ et donc $\varphi(x) \geq t + \varepsilon$, en particulier $\varphi(x) > t$.

Ainsi, on a $F_Y(t) = \mathbb{P}(\varphi(X) \leq t) = \mathbb{P}(X \leq F(t))$, pour tout $t \in \mathbb{R}$. D'après la formule (5.9) pour la fonction de répartition de X , on a $\mathbb{P}(X \leq F(t)) = 0$ si $F(t) = 0$, et $\mathbb{P}(X \leq F(t)) = \mathbb{P}(X \in]0, F(t]) = F(t)$ si $F(t) > 0$. Cela montre que $F_Y(t) = F(t)$ pour tout $t \in \mathbb{R}$, ce qui conclut la démonstration. \square

Soulignons que si F est bijective de $I = F^{-1}(]0, 1[)$ dans $]0, 1[$, alors on a $\varphi = F^{-1}$. De manière générale, on peut construire le graphe du pseudo-inverse φ de la manière suivante : i) on restreint le graphe de F à l'intervalle $I = F^{-1}(]0, 1[)$; ii) on complète le graphe en ajoutant un segment vertical à chaque point de discontinuité de F ; iii) on effectue la symétrie par rapport à la diagonale $y = x$ (les "sauts" de F deviennent des "plateaux" pour φ). Cela justifie la terminologie de pseudo-inverse, et on renvoie à l'Exemple 5.17 qui suit pour une illustration.

Exemple 5.17. Soit F la fonction définie par $F(x) = 0$ pour $x < -\frac{1}{3}$, $F(x) = \frac{1}{2}(x + \frac{1}{3})$ pour $x \in [-\frac{1}{3}, \frac{1}{3}[$, $F(x) = x + \frac{1}{3}$ pour $x \in [\frac{1}{3}, \frac{2}{3}[$, $F(x) = 1$ pour $x \geq \frac{2}{3}$, dont le graphe est donné ci-dessous (à gauche). Le graphe du pseudo-inverse φ de F est donné dans le graphe ci-dessous (à droite) : on a complété le graphe de F sur $]-\frac{1}{3}, \frac{2}{3}[$ avec un segment vertical en $x = \frac{1}{2}$, et effectué la symétrie par rapport à la droite $y = x$.



5.2.2 Loi exponentielle

Soit $\lambda > 0$. Une variable aléatoire réelle absolument continue X est dite *exponentielle de paramètre λ* , et on écrira $X \sim \mathcal{E}(\lambda)$, si elle a pour densité

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

(En particulier, $X > 0$ presque sûrement.) La fonction de répartition est donnée par : $F_X(t) = 0$ pour $t < 0$, et pour $t \geq 0$

$$F_X(x) = \int_{-\infty}^x f_X(x) dx = \int_0^x \lambda e^{-\lambda x} dx = [-e^{-\lambda x}]_0^x = 1 - e^{-\lambda x}. \quad (5.10)$$

En utilisant la formule de transfert, on obtient, avec une intégration par parties.

$$\begin{aligned}\mathbb{E}(X) &= \int_0^{\infty} x \cdot \lambda e^{-\lambda x} dx = \left[-x e^{-\lambda x} \right]_0^{\infty} + \int_0^{\infty} e^{-\lambda x} dx \\ &= 0 + \left[-\frac{1}{\lambda} e^{-\lambda x} \right]_0^{\infty} = \frac{1}{\lambda}.\end{aligned}$$

De même, on trouve que

$$\mathbb{E}(X^2) = \int_0^{\infty} x^2 \lambda e^{-\lambda x} dx = \left[-x^2 e^{-\lambda x} \right]_0^{\infty} + \frac{2}{\lambda} \int_0^{\infty} x \lambda e^{-\lambda x} dx = \frac{2}{\lambda} \mathbb{E}(X) = \frac{2}{\lambda^2},$$

de sorte que $\text{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \frac{1}{\lambda^2}$.

Proposition 5.18 (Absence de mémoire). *Si $X \sim \mathcal{E}(\lambda)$, alors*

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

Démonstration. On a $\mathbb{P}(X > s + t | X > s) = \frac{\mathbb{P}(X > s + t)}{\mathbb{P}(X > s)}$, et comme $\mathbb{P}(X > s) = 1 - F_X(s)$, on a d'après (5.10) que $\mathbb{P}(X > s) = e^{-\lambda s}$, $\mathbb{P}(X > s + t) = e^{-\lambda(s+t)}$. On obtient donc que $\mathbb{P}(X > s + t | X > s) = e^{-\lambda t} = \mathbb{P}(X > t)$. \square

Exercice 5.19. Soient X, Y deux variables aléatoires indépendantes, avec $X \sim \mathcal{E}(\lambda)$, $Y \sim \mathcal{E}(\mu)$, et soit $Z = \min\{X, Y\}$. Montrer que $\mathbb{P}(Z > t) = \mathbb{P}(X > t)\mathbb{P}(Y > t)$ pour tout $t \in \mathbb{R}$, et en déduire la fonction de répartition de Z . Conclure que $Z \sim \mathcal{E}(\lambda + \mu)$.

5.2.3 Loi normale (gaussienne)

Une variable aléatoire réelle absolument continue Z est dite *normale (ou gaussienne) standard*, et on écrira $Z \sim \mathcal{N}(0, 1)$, si elle a pour densité

$$f_Z(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}. \quad (5.11)$$

Notons que l'on a $\int_{-\infty}^{+\infty} e^{-\frac{x^2}{2}} dx = \sqrt{2\pi}$ (par un simple changement de variable dans l'intégrale de Gauss $\int_0^{\infty} e^{-x^2} dx = \frac{\sqrt{\pi}}{2}$, voir le cours d'intégrale de Lebesgue), ce qui assure que l'on a bien $\int_{-\infty}^{\infty} f_Z(x) dx = 1$. D'après la formule de l'espérance, on a

$$\mathbb{E}(Z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x e^{-\frac{x^2}{2}} dx = 0,$$

car la fonction $x e^{-x^2/2}$ est intégrable sur \mathbb{R} et impaire. De plus, en intégrant par parties,

$$\begin{aligned}\text{Var}(Z) &= \mathbb{E}(Z^2) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x^2 e^{-\frac{x^2}{2}} dx \\ &= \frac{1}{\sqrt{2\pi}} \left[\left[-x e^{-\frac{x^2}{2}} \right]_{-\infty}^{+\infty} + \int_{-\infty}^{+\infty} e^{-\frac{x^2}{2}} dx \right] = 1.\end{aligned}$$

Pour $Z \sim \mathcal{N}(0, 1)$ et des réels $\mu \in \mathbb{R}$, $\sigma \geq 0$ donnés, on définit la variable aléatoire $X := \sigma Z + \mu$. Des propriétés élémentaires de l'espérance et de la variance (Proposition 3.35

et Proposition 3.44), il s'ensuit que $\mathbb{E}(X) = \mu$, et $\text{Var}(X) = \sigma^2$. Si $\sigma = 0$, X n'est autre que la constante μ . Si en revanche $\sigma > 0$, d'après l'Exercice 5.6, la variable aléatoire X est absolument continue, de densité

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}. \quad (5.12)$$

Une variable aléatoire X absolument continue de densité donnée par (5.12) est appelée *normale (ou gaussienne) d'espérance $\mu \in \mathbb{R}$ et de variance $\sigma^2 \geq 0$* , et on écrit $X \sim \mathcal{N}(\mu, \sigma^2)$. (Par convention, X est la variable aléatoire constante égale à μ si $\sigma = 0$.)

En inversant la construction précédente, il s'ensuit que

$$\text{pour } \sigma > 0 : \quad X \sim \mathcal{N}(\mu, \sigma^2) \iff Z := \frac{X - \mu}{\sigma} \sim \mathcal{N}(0, 1).$$

Comme simple corollaire (exercice), on obtient le résultat suivant.

Proposition 5.20. *Une transformation affine d'une variable aléatoire normale est une variable aléatoire normale : $\forall a, b \in \mathbb{R}, X \sim \mathcal{N}(\mu, \sigma^2) \implies aX + b \sim \mathcal{N}(\tilde{\mu}, \tilde{\sigma}^2)$, avec $\tilde{\mu} = a\mu + b$ et $\tilde{\sigma}^2 = a^2\sigma^2$.*

Il est aussi possible de calculer la fonction génératrice des moments pour les variables aléatoires normales. Si $Z \sim \mathcal{N}(0, 1)$, on a

$$M_Z(t) = \int_{-\infty}^{+\infty} e^{tz} f_Z(z) dz = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{tx - \frac{1}{2}x^2} dx.$$

Cette intégrale peut se calculer en utilisant la méthode de complétion des carrés : on écrit $tx - \frac{1}{2}x^2 = -\frac{1}{2}(x - t)^2 + \frac{t^2}{2}$, ce qui donne $M_Z(t) = e^{\frac{t^2}{2}} \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-t)^2} dx$. Comme la fonction intégrée est la densité d'une variable aléatoire normale d'espérance t et de variance 1, l'intégrale vaut 1 : on en déduit que

$$M_Z(t) = e^{\frac{t^2}{2}} \quad \forall t \in \mathbb{R}. \quad (5.13)$$

Cela permet en particulier de calculer tous les moments de $Z \sim \mathcal{N}(0, 1)$, en écrivant $M_Z(t) = \sum_{k=0}^{\infty} \mathbb{E}(Z^k) \frac{t^k}{k!}$ et $e^{t^2/2} = \sum_{n=0}^{\infty} \frac{t^{2n}}{2^n n!}$: on trouve que $\mathbb{E}(Z^k) = 0$ si k est impair, et $\mathbb{E}(Z^{2k}) = \frac{(2k)!}{2^k k!}$ dans le cas pair.

Si X est une variable aléatoire de loi $\mathcal{N}(\mu, \sigma^2)$, elle peut s'écrire sous la forme $X = \sigma Z + \mu$, avec $Z := \frac{1}{\sigma}(X - \mu) \sim \mathcal{N}(0, 1)$: on a ainsi

$$M_X(t) = \mathbb{E}(e^{tX}) = \mathbb{E}(e^{t(\sigma Z + \mu)}) = e^{\mu t} \mathbb{E}(e^{t\sigma Z}) = e^{\mu t + \frac{\sigma^2}{2} t^2}.$$

Concluons avec une propriété fondamentale que l'on démontrera plus bas (Exemple 5.31) : la somme de deux variables aléatoires normales *indépendantes* est normale.

Proposition 5.21 (Somme de normales indépendantes). *Soient X et Y des variables aléatoires normales indépendantes : $X \sim \mathcal{N}(\mu_1, \sigma_1^2)$ et $Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$. Alors leur somme est aussi normale : $X + Y \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.*

Corollaire 5.22. Soient X_1, \dots, X_n des variables aléatoires normales indépendantes. Toute combinaison linéaire $X := a_1X_1 + \dots + a_nX_n + b$, avec $a_1, \dots, a_n, b \in \mathbb{R}$, est une variable aléatoire normale.

Démonstration. Le cas $n = 1$ n'est autre que la Proposition 5.20. En procédant par récurrence, on définit

$$Y := a_1X_1 + \dots + a_{n-1}X_{n-1}, \quad W := a_nX_n + b,$$

de sorte que, par hypothèse de récurrence, aussi bien Y que W sont des variables aléatoires normales. Celles-ci sont de plus indépendantes, parce qu'elles sont construites à partir d'ensembles disjoints des variables aléatoires indépendantes (voir les Propositions 3.17-3.18). Par conséquent $X = Y + W$ est normale, d'après la Proposition 5.21. \square

5.3 Vecteurs aléatoires absolument continus

5.3.1 Définition et premières propriétés

Définition 5.23. Un vecteur aléatoire n -dimensionnel X (c'est-à-dire une variable aléatoire X à valeurs dans \mathbb{R}^n), est dit absolument continu ou à densité s'il existe une fonction positive $f_X : \mathbb{R}^n \rightarrow [0, +\infty[$ intégrable sur \mathbb{R}^n telle que, pour tout choix d'intervalles I_1, \dots, I_n de \mathbb{R} , on ait

$$\mathbb{P}(X \in I_1 \times \dots \times I_n) = \int_{I_1 \times \dots \times I_n} f_X(x) dx. \quad (5.14)$$

La fonction f_X est appelée densité du vecteur aléatoire X .

En analogie avec le cas unidimensionnel, en prenant $I_1 = \dots = I_n = \mathbb{R}$ dans (5.14), on a que $\int_{\mathbb{R}^n} f_X(x) dx = 1$. Réciproquement, toute fonction positive $f : \mathbb{R}^n \rightarrow [0, +\infty[$, intégrable sur \mathbb{R}^n et d'intégrale 1 est la densité d'un vecteur aléatoire dans \mathbb{R}^n .

La densité f_X d'un vecteur aléatoire absolument continu X n'est pas déterminée de manière unique : si $g : \mathbb{R}^n \rightarrow [0, +\infty[$ coïncide avec f_X en dehors d'un ensemble de mesure n -dimensionnelle nulle (voir le cours d'intégrale de Lebesgue), la relation (5.14) est toujours valable avec g à la place de f_X .

Remarque 5.24. On peut démontrer que de (5.14) il découle que

$$\mathbb{P}(X \in C) = \int_C f_X(x) dx = \int_{\mathbb{R}^n} f_X(x) \mathbb{1}_C(x) dx, \quad (5.15)$$

pour une grande classe d'ensembles $C \subseteq \mathbb{R}^n$ [†]. Cela permet donc de calculer les probabilités de beaucoup d'événements d'intérêt impliquant X . Soulignons que la densité d'un vecteur aléatoire absolument continu en caractérise la loi (pour montrer cela, il faut des notions de théorie de la mesure).

[†]. Les boréliens de \mathbb{R}^n .

Exemple 5.25 (Vecteurs aléatoires uniformes continus). Soit $C \subseteq \mathbb{R}^n$ un sous-ensemble avec $0 < \text{mes}(C) := \int_{\mathbb{R}^n} \mathbb{1}_C(x) dx < +\infty$. Un vecteur aléatoire n -dimensionnel X est dit *uniforme sur C* , et on écrit $X \sim \mathcal{U}(C)$, si X est absolument continu, de densité

$$f_X(x) = \frac{1}{\text{mes}(C)} \mathbb{1}_C(x).$$

Un vecteur aléatoire $X \sim \mathcal{U}(C)$ traduit en termes mathématiques l'idée de "choisir au hasard uniformément" un point de l'ensemble C .

5.3.2 Densité jointe, densités marginales, indépendance

Pour un vecteur aléatoire absolument continu $X = (X_1, \dots, X_n)$ à valeurs dans \mathbb{R}^n , on notera sa densité f_{X_1, \dots, X_n} .

Proposition 5.26 (Densités jointes et marginales). *Soit $X = (X_1, \dots, X_n)$ un vecteur aléatoire absolument continu, à valeurs dans \mathbb{R}^n , de densité f_{X_1, \dots, X_n} . Pour tout $k < n$, le vecteur aléatoire (X_1, \dots, X_k) , à valeurs dans \mathbb{R}^k , est absolument continu, de densité*

$$f_{X_1, \dots, X_k}(x_1, \dots, x_k) = \int_{\mathbb{R}^{n-k}} f_X(x_1, \dots, x_k, x_{k+1}, \dots, x_n) dx_{k+1} \cdots dx_n.$$

Un résultat analogue est valable pour n'importe quel sous-ensemble de composantes $X_J := (X_i)_{i \in J}$, avec $J \subseteq \{1, \dots, n\}$.

La densité f_{X_1, \dots, X_n} d'un vecteur aléatoire (X_1, \dots, X_n) est aussi appelée *densité jointe* des variables aléatoires X_1, \dots, X_n , et les densités f_{X_i} des différentes composantes sont appelées *densités marginales*.

Démonstration. Soient I_1, \dots, I_k des intervalles de \mathbb{R} , et soit $I_m = \mathbb{R}$ pour $k < m \leq n$. En appliquant le théorème de Fubini-Tonelli à la relation (5.14), on a

$$\begin{aligned} \mathbb{P}(X_1 \in I_1, \dots, X_k \in I_k) &= \mathbb{P}(X_1 \in I_1, \dots, X_k \in I_k, X_{k+1} \in \mathbb{R}, \dots, X_n \in \mathbb{R}) \\ &= \int_{I_1 \times \dots \times I_k} \left[\int_{\mathbb{R}^{n-k}} f_X(x_1, \dots, x_k, x_{k+1}, \dots, x_n) dx_{k+1} \cdots dx_n \right] dx_1 \cdots dx_k, \end{aligned}$$

qui donne le résultat. □

En particulier, si (X, Y) est un vecteur aléatoire absolument continu à valeurs dans \mathbb{R}^2 , ses composantes X et Y sont des variables aléatoires absolument continues, de densité

$$f_X(x) = \int_{-\infty}^{+\infty} f_{X,Y}(x, y) dy, \quad f_Y(y) = \int_{-\infty}^{+\infty} f_{X,Y}(x, y) dx. \quad (5.16)$$

Donnons un exemple du type de calculs que l'on peut faire : déterminons la loi de la somme des composantes d'un vecteur aléatoire à densité.

Proposition 5.27. Soit (X, Y) un vecteur aléatoire bidimensionnel absolument continu, de densité $f_{X,Y}$. Alors la somme des composantes $X + Y$ est une variable aléatoire réelle absolument continue, de densité

$$f_{X+Y}(z) = \int_{-\infty}^{+\infty} f_{X,Y}(x, z-x) dx = \int_{-\infty}^{+\infty} f_{X,Y}(z-x, x) dx, \quad (5.17)$$

à condition que $x \mapsto f_{X,Y}(x, z-x)$ et $x \mapsto f_{X,Y}(z-x, x)$ soient intégrables sur \mathbb{R} .

Démonstration. Calculons la fonction de répartition de $X + Y$: en introduisant pour $t \in \mathbb{R}$ le sous-ensemble $D_z := \{(x, y) \in \mathbb{R}^2 : x + y \leq t\}$, on peut écrire

$$\begin{aligned} F_{X+Y}(z) &= \mathbb{P}(X + Y \leq z) = \mathbb{P}((X, Y) \in D_z) = \int_{D_z} f_{X,Y}(x, y) dx dy \\ &= \int_{-\infty}^{+\infty} \left(\int_{-\infty}^{+\infty} f_{X,Y}(x, y) \mathbb{1}_{D_z}(x, y) dy \right) dx = \int_{-\infty}^{+\infty} \left(\int_{-\infty}^{z-x} f_{X,Y}(x, y) dy \right) dx, \end{aligned}$$

où on a appliqué la formule (5.15) et le théorème de Fubini-Tonelli. En opérant le changement de variable $y = \psi(t) := t - x$ dans l'intégrale interne et en appliquant de nouveau le théorème de Fubini-Tonelli, on obtient

$$F_{X+Y}(z) = \int_{-\infty}^{+\infty} \left(\int_{-\infty}^z f_{X,Y}(x, t-x) dt \right) dx = \int_{-\infty}^z \left(\int_{-\infty}^{+\infty} f_{X,Y}(x, t-x) dx \right) dt.$$

Nous avons donc montré que $F_{X+Y}(z) = \int_{-\infty}^z g(t) dt$, où $g(t)$ désigne l'intégrale interne dans la dernière expression, ce qui démontre que $X + Y$ est une variable aléatoire absolument continue. Avec des arguments similaires, on obtient la deuxième expression dans (5.17). \square

De manière analogue à la Proposition 3.15, l'indépendance de variables aléatoires peut être caractérisée en termes de leur loi jointe.

Proposition 5.28 (Indépendance et densité). Soit $X = (X_1, \dots, X_n)$ un vecteur aléatoire absolument continu n -dimensionnel, pour lequel on a

$$f_{X_1, \dots, X_n}(x_1, \dots, x_n) = f_{X_1}(x_1) \cdots f_{X_n}(x_n) \quad (5.18)$$

pour tout $x = (x_1, \dots, x_n) \in \mathbb{R}^n \setminus C$, où C est un ensemble de mesure n -dimensionnelle nulle (qui peut être vide). Alors les composantes X_1, \dots, X_n sont des variables aléatoires réelles indépendantes.

Réciproquement, si X_1, \dots, X_n sont des variables aléatoires réelles indépendantes et absolument continues, de densité f_{X_1}, \dots, f_{X_n} , le vecteur aléatoire $X = (X_1, \dots, X_n)$ est absolument continu, de densité jointe donnée par (5.18).

Dans le cas de deux variables aléatoires réelles X et Y indépendantes, la relation (5.18) devient

$$f_{X,Y}(x, y) = f_X(x) f_Y(y). \quad (5.19)$$

Dans les faits, pour démontrer l'indépendance de deux variables aléatoires X et Y , il suffit de montrer que la densité jointe $f_{X,Y}(x, y)$ se factorise sous la forme d'un produit d'une fonction de x et d'une fonction de y (exercice : voir la Remarque 3.16).

Remarque 5.29. D'après la Proposition 5.26, si un vecteur aléatoire est absolument continu, alors ses composantes sont des variables aléatoires réelles à densité. La réciproque est vraie si les composantes sont indépendantes, d'après la Proposition 5.28, mais *elle est fautive* en général. Par exemple, si X est une variable aléatoire réelle, le vecteur (X, X) ne peut pas être absolument continu (exercice †).

En combinant la Proposition 5.28 avec la Proposition 5.27, on obtient le corollaire suivant.

Proposition 5.30. *Soit X et Y deux variables aléatoires réelles indépendantes, de densités respectives f_X et f_Y . Alors $X + Y$ est une variable aléatoire absolument continue, de densité donnée par*

$$f_{X+Y}(z) = f_X * f_Y(z) := \int_{-\infty}^{+\infty} f_X(x)f_Y(z-x)dx = \int_{-\infty}^{+\infty} f_X(z-x)f_Y(x)dx.$$

*La fonction $f_X * f_Y = f_Y * f_X$ est appelé produit de convolution de f_X et f_Y .*

Exemple 5.31. Soient $X \sim \mathcal{N}(0, 1)$ et $Y \sim \mathcal{N}(0, 1)$ indépendantes. Déterminons la densité de $X + Y$. D'après la Proposition 5.30, on a

$$f_{X+Y}(z) = \int_{-\infty}^{+\infty} f_X(x)f_Y(z-x) dx = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \exp\left(-\frac{x^2}{2} - \frac{(z-x)^2}{2}\right) dx.$$

On vérifie par un calcul qu'on a $\frac{x^2}{2} + \frac{(z-x)^2}{2} = \frac{z^2}{4} + (x - \frac{z}{2})^2$: on obtient donc

$$f_{X+Y}(z) = \frac{1}{2\pi} e^{-\frac{z^2}{4}} \int_{-\infty}^{+\infty} e^{-(x-\frac{z}{2})^2} dx = \frac{1}{\sqrt{4\pi}} e^{-\frac{z^2}{4}},$$

où on a utilisé que $\int_{-\infty}^{+\infty} e^{-(x-z)^2} dx = \sqrt{\pi}$ (en reconnaissant par exemple que la fonction qui apparaît dans l'intégrale est égale à la densité d'une $\mathcal{N}(z, 1/2)$ à un facteur multiplicatif $\frac{1}{\sqrt{\pi}}$ près, cf. (5.12)). On a donc montré que $X + Y \sim \mathcal{N}(0, 2)$.

Exercice : traiter le cas général $X \sim \mathcal{N}(\mu_1, \sigma_1^2)$ et $Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$ indépendantes.

5.3.3 Formule de transfert et calculs avec les densités

Proposition 5.32 (Formule de transfert). *Soit X un vecteur aléatoire absolument continu n -dimensionnel, de densité f_X , et soit $g : \mathbb{R}^n \rightarrow \mathbb{R}$ une fonction continue par morceaux. Alors $g(X)$ admet une espérance finie si et seulement si la fonction $|g(x)|f_X(x)$ est intégrable sur \mathbb{R}^n , et dans ce cas*

$$\mathbb{E}(g(X)) = \int_{\mathbb{R}^n} g(x) f_X(x) dx. \tag{5.20}$$

La formule (5.20), dite *de transfert* est à la base de nombreux calculs avec des vecteurs aléatoires. Par exemple, de la même manière que dans la Proposition 5.13, la formule de

†. Remarquer que $\mathbb{P}((X, X) \in A) = 1$ où $A = \{(x, y) \in \mathbb{R}^2 : y = x\}$, mais que A est de dimension 1, donc de mesure 2-dimensionnelle nulle.

transfert (5.20) permet de déterminer la densité d'un vecteur aléatoire. (La démonstration est similaire à celle de la Proposition 5.13.)

Proposition 5.33. *Un vecteur aléatoire X de dimension n est absolument continu et a pour densité f_X si et seulement si pour toute fonction $h : \mathbb{R}^n \rightarrow \mathbb{R}$ continue et bornée on a la formule de transfert*

$$\mathbb{E}(h(X)) = \int_{\mathbb{R}^n} h(x) f_X(x) dx. \quad (5.21)$$

Avant de donner un exemple, rappelons la formule de changement de variable. Pour $V, U \subseteq \mathbb{R}^n$ deux ouverts, une fonction $\varphi : V \rightarrow U$ est appelée \mathcal{C}^1 -difféomorphisme si :

- φ est bijective ;
- les dérivées partielles de φ existent et sont continues sur V ;
- la matrice jacobienne de $\varphi = (\varphi_1, \dots, \varphi_n)$ en z , $J_\varphi(z) = \begin{pmatrix} \frac{\partial \varphi_1}{\partial x_1} & \dots & \frac{\partial \varphi_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial \varphi_n}{\partial x_1} & \dots & \frac{\partial \varphi_n}{\partial x_n} \end{pmatrix}$ est inversible (autrement dit, $\det J_\varphi(z)$, appelé jacobien, est non nul) pour tout $z \in V$.

La formule de changement de variable s'énonce ainsi : pour toute fonction $g : U \rightarrow \mathbb{R}$ intégrable, la fonction $g(\varphi(z)) |\det J_\varphi(z)| : V \rightarrow \mathbb{R}$ est intégrable, et on a

$$\int_U g(x) dx = \int_V g(\varphi(z)) |\det J_\varphi(z)| dz. \quad (5.22)$$

Exemple 5.34 (Box-Müller). Soit U_1, U_2 deux variables aléatoires indépendantes, de loi $\mathcal{U}(0, 1)$. On pose $R = \sqrt{-2 \ln U_1}$ et $\Theta = 2\pi U_2$, qui sont indépendantes, et de densités respectives $f_R(r) = r e^{-\frac{r^2}{2}} \mathbb{1}_{]0, \infty[}(r)$ et $f_\Theta(\theta) = \frac{1}{2\pi} \mathbb{1}_{]0, 2\pi[}(\theta)$ (exercice). On pose enfin

$$X = R \cos \Theta, \quad Y = R \sin \Theta.$$

Déterminons la densité jointe de X et Y . Pour cela, on se fixe une fonction $h : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ continue et bornée, et on va essayer d'exprimer $\mathbb{E}(h(X, Y))$ sous la forme (5.21).

Une première étape consiste à écrire que $\mathbb{E}(h(X, Y)) = \mathbb{E}(h(R \cos \Theta, R \sin \Theta))$, qui est l'espérance d'une fonction g de R et Θ continue et bornée[§], donc intégrable, de sorte que l'on peut appliquer la formule de transfert à cette fonction-là. Comme on connaît la densité de (R, Θ) , qui vaut $f_{R, \Theta}(r, \theta) = f_R(r) f_\Theta(\theta)$ car R et Θ sont indépendantes, on obtient

$$\mathbb{E}(h(X, Y)) = \int_{]0, +\infty[\times]0, 2\pi[} h(r \cos \theta, r \sin \theta) \frac{1}{2\pi} r e^{-\frac{r^2}{2}} dr d\theta.$$

La deuxième étape consiste à effectuer un changement de variable, afin de passer en coordonnées cartésiennes : on pose $L = \{(x, y) \in \mathbb{R}^2 : x \geq 0, y = 0\}$,

$$V :=]0, +\infty[\times]0, 2\pi[, \quad U := \mathbb{R}^2 \setminus L,$$

[§]. La fonction $g :]0, +\infty[\times]0, 2\pi[\rightarrow \mathbb{R}^2$ définie par $g(r, \theta) = h(r \cos \theta, r \sin \theta)$.

et on considère la fonction $\varphi : V \rightarrow U$ définie par $\varphi(r, \theta) := (r \cos \theta, r \sin \theta)$, dont on vérifie facilement qu'il s'agit d'un difféomorphisme, de matrice jacobienne

$$J_\varphi(r, \theta) = \begin{pmatrix} \cos \theta & r \sin \theta \\ \sin \theta & -r \cos \theta \end{pmatrix},$$

de jacobien $|\det J_\varphi(r, \theta)| = r$. La formule du changement de variable (5.22) nous donne, en écrivant $\exp(-\frac{r^2}{2})$ sous la forme $\exp(-\frac{1}{2}((r \cos \theta)^2 + (r \sin \theta)^2))$,

$$\begin{aligned} \mathbb{E}(h(X, Y)) &= \int_V h(r \cos \theta, r \sin \theta) \frac{1}{2\pi} e^{-\frac{1}{2}((r \cos \theta)^2 + (r \sin \theta)^2)} |\det \text{Jac}_\varphi(r, \theta)| dr d\theta \\ &= \int_U h(x, y) \frac{1}{2\pi} e^{-\frac{x^2+y^2}{2}} dx dy. \end{aligned}$$

Comme l'ensemble L est une demi droite, il possède une mesure 2-dimensionnelle nulle, et on peut remplacer le domaine d'intégration $U = \mathbb{R}^2 \setminus L$ par \mathbb{R}^2 . Cette formule étant valable quelle que soit la fonction $h : \mathbb{R}^2 \rightarrow \mathbb{R}$ continue bornée, la Proposition 5.33 nous dit que le vecteur (X, Y) possède pour densité

$$f_{X,Y}(x, y) = \frac{1}{2\pi} e^{-\frac{x^2+y^2}{2}} = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \times \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}.$$

Autrement dit, $X \sim \mathcal{N}(0, 1)$, $Y \sim \mathcal{N}(0, 1)$ avec X et Y indépendants.

5.4 Quelques exemples supplémentaires...

★ Loi Gamma

Une variable aléatoire réelle X absolument continue est dite de *Gamma de paramètre* $\alpha, \lambda > 0$, et on écrira $X \sim \text{Gamma}(\alpha, \lambda)$, si elle a pour densité

$$f_X(x) = \frac{1}{\Gamma(\alpha)} \lambda^\alpha x^{\alpha-1} e^{-\lambda x} \mathbf{1}_{]0, +\infty[}(x), \quad (5.23)$$

où $\Gamma :]0, \infty[\rightarrow \mathbb{R}$ est la fonction "Gamma" d'Euler, définie par $\Gamma(\alpha) = \int_0^{+\infty} x^{\alpha-1} e^{-x} dx$, qui est finie si $\alpha > 0$ (exercice). Pour $\alpha = 1$, on retrouve la loi exponentielle de paramètre λ .

Soulignons que le deuxième paramètre d'une loi $\text{Gamma}(\alpha, \lambda)$ est simplement un facteur d'échelle : plus précisément $X \sim \text{Gamma}(\alpha, \lambda)$ si et seulement si $Y = \lambda X \sim \text{Gamma}(\alpha, 1)$ (voir l'Exercice 5.6).

Espérance et variance se calculent facilement : si $Y \sim \text{Gamma}(\alpha, 1)$, on obtient

$$\mathbb{E}(Y) = \int_{-\infty}^{+\infty} y f_Y(y) dy = \frac{1}{\Gamma(\alpha)} \int_0^{+\infty} y^\alpha e^{-y} dy = \frac{\Gamma(\alpha+1)}{\Gamma(\alpha)} = \alpha,$$

où on a utilisé que $\Gamma(\alpha+1) = \alpha\Gamma(\alpha)$ pour tout $\alpha > 0$ (exercice). De même, on a

$$\mathbb{E}(Y^2) = \int_{-\infty}^{+\infty} y^2 f_Y(y) dy = \frac{1}{\Gamma(\alpha)} \int_0^{+\infty} y^{\alpha+1} e^{-y} dy = \frac{\Gamma(\alpha+2)}{\Gamma(\alpha)} = \alpha(\alpha+1),$$

dont on déduit que $\text{Var}(Y) = \mathbb{E}(Y^2) - \mathbb{E}(Y)^2 = \alpha$. Comme on a $X := \frac{1}{\lambda}Y \sim \text{Gamma}(\alpha, \lambda)$, on obtient donc l'espérance et la variance d'une variable aléatoire $X \sim \text{Gamma}(\alpha, \lambda)$:

$$\mathbb{E}(X) = \frac{\alpha}{\lambda}, \quad \text{Var}(X) = \frac{\alpha}{\lambda^2}. \quad (5.24)$$

Proposition 5.35 (Somme de Gamma indépendantes). *Soient X et Y des variables aléatoires réelles indépendantes, avec $X \sim \text{Gamma}(\alpha, \lambda)$ et $Y \sim \text{Gamma}(\beta, \lambda)$. Alors $X + Y \sim \text{Gamma}(\alpha + \beta, \lambda)$.*

Démonstration. On a la formule suivante pour la densité de la somme de deux variables aléatoires indépendantes à densité (Proposition 5.30)

$$\begin{aligned} f_{X+Y}(z) &= \int_{-\infty}^{+\infty} f_X(x)f_Y(z-x) dx \\ &= \frac{\lambda^\alpha \lambda^\beta}{\Gamma(\alpha)\Gamma(\beta)} \int_{-\infty}^{+\infty} \mathbb{1}_{]0,+\infty[}(x) x^{\alpha-1} e^{-\lambda x} \mathbb{1}_{]0,+\infty[}(z-x) (z-x)^{\beta-1} e^{-\lambda(z-x)} dx \\ &= \frac{\lambda^{\alpha+\beta}}{\Gamma(\alpha)\Gamma(\beta)} e^{-\lambda z} \int_0^z x^{\alpha-1} (z-x)^{\beta-1} dx. \end{aligned}$$

Avec le changement de variable $x = zt$ on obtient donc

$$f_{X+Y}(z) = C \lambda^{\alpha+\beta} z^{\alpha+\beta-1} e^{-\lambda z}, \quad \text{avec } C := \frac{1}{\Gamma(\alpha)\Gamma(\beta)} \int_0^1 t^{\alpha-1} (1-t)^{\beta-1} dt.$$

La densité $f_{X+Y}(z)$ étant proportionnelle à $z^{\alpha+\beta-1} e^{-\lambda z}$, elle est nécessairement la densité d'une $\text{Gamma}(\alpha + \beta, \lambda)$ (et la constante C est égale à $1/\Gamma(\alpha + \beta)$). \square

Remarque 5.36. D'après la démonstration précédente, on a l'identité suivante, pour tous $a, b \in]0, \infty[$

$$\beta(a, b) := \int_0^1 t^{a-1} (1-t)^{b-1} dt = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}, \quad \forall a, b > 0, \quad (5.25)$$

qui n'est pas du tout évidente! (La fonction $\beta :]0, \infty[^2 \rightarrow]0, \infty[$ est appelée *fonction Beta*.)

Exercice 5.37. Montrer que si $X \sim \mathcal{N}(0, 1)$, alors $X^2 \sim \Gamma(\frac{1}{2}, \frac{1}{2})$. En déduire que si X_1, \dots, X_n sont des variables aléatoires $\mathcal{N}(0, 1)$ indépendantes, alors $X_1^2 + \dots + X_n^2 \sim \Gamma(\frac{n}{2}, \frac{1}{2})$. Une telle variable aléatoire est dite de χ^2 ("khi-deux") à n degrés de liberté.

★ Loi de Cauchy

Une variable aléatoire réelle X absolument continue est dite *de Cauchy* si elle a pour densité

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

Comme $\arctan'(x) = \frac{1}{1+x^2}$ et que $\arctan(\pm\infty) = \pm\frac{\pi}{2}$, on a

$$F_X(t) = \int_{-\infty}^t f_X(x) dx = \frac{1}{\pi} (\arctan(t) - \arctan(-\infty)) = \frac{1}{\pi} \arctan(t) + \frac{1}{2}.$$

La première remarque est que X n'admet pas d'espérance. On en effet

$$\mathbb{E}(X^+) = \int_{-\infty}^{+\infty} x^+ f_X(x) dx = \frac{1}{\pi} \int_0^{+\infty} \frac{x}{1+x^2} dx = +\infty,$$

où la dernière égalité vient du critère de Riemann, car $\frac{x}{1+x^2} \sim \frac{1}{x}$ quand $x \rightarrow +\infty$. De même, par symétrie, $\mathbb{E}(X^-) = +\infty$. Donc X n'admet pas d'espérance, même si $\int_{-a}^{+a} x f_X(x) dx = 0$ pour tout $a > 0$!

Somme de variables de Cauchy indépendantes (difficile!). Soient maintenant X, Y deux variables aléatoires *indépendantes*, de même loi de Cauchy. D'après la Proposition 5.30, la densité de $X + Y$ vaut

$$f_{X+Y}(z) = \frac{1}{\pi^2} \int_{-\infty}^{+\infty} g_z(x) dx, \quad \text{où } g_z(x) := \frac{1}{1+x^2} \cdot \frac{1}{1+(z-x)^2}.$$

Notons que g_z est positive, et vérifie $g_z(x) \sim 1/x^4$ quand $x \rightarrow \pm\infty$: l'intégrale (doublement) généralisée de la fonction g_z est donc bien définie, et on a $\int_{-\infty}^{+\infty} g_z(x) dx = \lim_{a \rightarrow \infty} \int_{-a}^a g_z(x) dx$.

Pour calculer cette intégrale, on effectue une décomposition en éléments simples de $g_z(x)$: après calcul et quelques manipulations algébriques, on trouve, pour $z \neq 0$

$$g_z(x) = \frac{\alpha_z}{1+x^2} + \frac{2\beta_z x}{1+x^2} + \frac{\alpha_z}{1+(x-z)^2} - \frac{2\beta_z(x-z)}{1+(x-z)^2} \quad \text{avec } \alpha_z = \frac{1}{4+z^2}, \beta_z = z^{-1}\alpha_z.$$

Des calculs élémentaires donnent :

$$\begin{aligned} \int_{-a}^a \frac{1}{1+x^2} dx &= \arctan(a) - \arctan(-a), & \int_{-a}^a \frac{2x}{1+x^2} dx &= 0, \\ \int_{-a}^a \frac{1}{1+(x-z)^2} dx &= \arctan(a+z) - \arctan(-a+z), \\ \int_{-a}^a \frac{2(x-z)}{1+(x-z)^2} dx &= \log\left(\frac{1+(a+z)^2}{1+(-a+z)^2}\right). \end{aligned}$$

Quand $a \rightarrow +\infty$, la première et la troisième intégrale convergent vers π , alors que la deuxième et la quatrième convergent vers 0*. On obtient donc, pour tout $z \neq 0$,

$$\int_{-\infty}^{+\infty} g_z(x) dx = \lim_{a \rightarrow \infty} \int_{-a}^a g_z(x) dx = 2\pi\alpha_z = \frac{2\pi}{4+z^2},$$

de sorte que

$$f_{X+Y}(z) = \frac{1}{\pi^2} \int_{-\infty}^{+\infty} g_z(x) dx = \frac{2}{\pi(4+z^2)}.$$

Comme on peut changer la valeur en 0 d'une densité arbitrairement, on peut choisir cette fonction (définie aussi en 0) pour densité de $X + Y$.

D'après l'Exercice 5.6, si W est une variable aléatoire de Cauchy, alors $2W$ possède pour densité $f_{2W}(x) = \frac{1}{2} f_W\left(\frac{x}{2}\right) = \frac{2}{\pi(4+w^2)}$. On a ainsi que $X + Y$ et $2W$ ont la densité, et donc la même loi. *La conclusion est donc que si X, Y sont deux variables aléatoires indépendantes de loi de Cauchy, alors $\frac{1}{2}(X + Y)$ suit une loi de Cauchy.*

*. Notons qu'il y a ici une subtilité : g_z est intégrable sur \mathbb{R} , mais ce n'est pas le cas des différents termes de sa décomposition en éléments simples... Le fait d'écrire $\int_{-\infty}^{+\infty} g_z(x) dx = \lim_{a \rightarrow \infty} \int_{-a}^a g_z(x) dx$ permet de voir que certaines intégrales des éléments simples "se compensent".

Chapitre 6

Simulation informatique de variables aléatoires

Le principe des simulations est d'utiliser un ordinateur pour effectuer des expériences aléatoires à notre place : on va les *simuler* plutôt que de les effectuer en pratique...

Exemple 6.1. Soient X_1, \dots, X_n des variables aléatoires indépendantes, de loi uniforme sur $\{1, \dots, 365\}$, représentant les jours d'anniversaire de n personnes. On se demande quelle est la probabilité p_n que "au moins trois personnes aient leurs anniversaires le même jour". Contrairement au problème habituel des anniversaires, cette probabilité est difficile à calculer... Une approche numérique consiste à utiliser la méthode de Monte-Carlo (détaillée dans le chapitre 7) : si on arrive à simuler un très grand nombre $N \gg 1$ de n -uplets (X_1, \dots, X_n) , la proportion de n -uplets qui vérifient la propriété qu'au moins trois de ses composantes sont égales est une bonne approximation de la probabilité recherchée...

Pour construire des fonctions qui génèrent différentes variables aléatoires, on utilisera essentiellement la fonction `random` (présente dans la librairie `random` de Python) : un appel à la fonction `random` renvoie un nombre *aléatoire*, choisi uniformément dans l'intervalle $]0, 1[$; de plus, des appels successifs à la fonction `random` renvoient des résultats indépendants. On dit que `random` *génère* des variables aléatoires indépendantes, de loi uniforme sur $]0, 1[$.

6.1 Simulation de lois discrètes

Loi uniforme discrète

Pour simuler une variable aléatoire de loi uniforme sur $\{1, \dots, n\}$, on utilise le fait suivant.

Si $U \sim \mathcal{U}(0, 1)$ et si $n \geq 1$, alors $\lfloor nU \rfloor + 1$ suit une loi uniforme sur $\{1, \dots, n\}$.

Démonstration. Pour montrer cela, il suffit de calculer la densité discrète de $Y := \lfloor nU \rfloor + 1$, qui est clairement à valeurs dans $\{1, \dots, n\}$: pour tout $k \in \{1, \dots, n\}$, on a

$$\mathbb{P}(Y = k) = \mathbb{P}(\lfloor nU \rfloor = k - 1) = \mathbb{P}\left(U \in \left[\frac{k-1}{n}, \frac{k}{n}\right)\right) = \frac{k}{n} - \frac{k-1}{n} = \frac{1}{n},$$

où on a utilisé le fait que $\lfloor nu \rfloor = k - 1$ si et seulement si $nu \in [k - 1, k[$, puis le fait que $\mathbb{P}(U \in [a, b]) = b - a$ pour $0 \leq a < b \leq 1$. \square

Ainsi, une manière d'écrire une fonction `Uniforme` qui prend en argument un entier $n \geq 1$ et qui génère une variable aléatoire de loi uniforme sur $\{1, \dots, n\}$ est la suivante :

```
1 def Uniforme(n):
2     U=random()
3     return int(n*U)+1
```

Loi de Bernoulli

Pour $p \in]0, 1[$, si $U \sim \mathcal{U}(0, 1)$, alors l'événement $\{U < p\}$ est un événement de probabilité p . On a donc clairement que :

Si $U \sim \mathcal{U}(0, 1)$, alors $1_{\{U < p\}}$ est une variable aléatoire de Bernoulli de paramètre p .

Pour simuler une variable aléatoire de Bernoulli, on va donc générer une variable aléatoire U de loi $\mathcal{U}(0, 1)$ (grâce à la fonction `random`), et renvoyer 1 si $U < p$, et 0 si $U \geq p$. Une manière d'écrire une fonction `Bern` qui prend en argument un réel $p \in]0, 1[$ et génère une variable aléatoire de loi `Bern(p)` est la suivante :

```
1 def Bern(p):
2     U=random()
3     if U<p:
4         return 1
5     else:
6         return 0
```

Loi binomiale

Pour simuler une variable aléatoire de loi `Bin(n, p)`, il suffit d'utiliser le fait que la somme de n variables de Bernoulli de paramètre p indépendantes suit la loi `Bin(n, p)`. On va donc faire n appels successifs à la fonction `Bern` que l'on vient de définir, et sommer au fur à mesure les résultats que l'on obtient. On rappelle que des appels successifs à la fonction `random` renvoient des simulations *indépendantes* de variables aléatoires $\mathcal{U}(0, 1)$; des appels successifs à la fonction `Bern` renverront donc des simulations indépendantes de variables aléatoires de loi de Bernoulli. Deux manières d'écrire une fonction `Binom` qui prend en argument un entier n et un réel $p \in]0, 1[$ et qui génère une variable aléatoire de loi `Bin(n, p)` sont donc :

```
1 def Binom(n,p):
2     X=0
3     for i in range(n):
4         X=X+Bern(p)
5     return X
```

```
1 def Binom(n,p):
2     X=0
3     for i in range(n):
4         if random()<p:
5             X+= 1
6     return X
```

Loi géométrique

Pour simuler une variable aléatoire de loi `Géom(p)`, on va utiliser le fait qu'une telle variable aléatoire correspond au premier instant de succès lors de la répétition d'épreuves indépendantes, de probabilité de succès p .

Si $(U_i)_{i \geq 1}$ est une suite de variables aléatoires indépendantes de loi $\mathcal{U}(0, 1)$, alors $T = \min\{i \geq 1, U_i < p\}$ suit une loi géométrique $\text{Geom}(p)$.

On va ainsi faire des appels successifs à la fonction `random`, en continuant tant que l'on obtient une valeur plus grande que p , et on va compter le nombre d'appels à la fonction `random` que l'on a effectué : on va utiliser une boucle `while` qui s'arrêtera la première fois que l'on aura obtenu une valeur plus petite que p ; avec un compteur qui compte le nombre de répétitions de la boucle. Une manière d'écrire une fonction `Geom` qui prend en argument un $p \in]0, 1[$ et génère une variable aléatoire de loi $\text{Geom}(p)$ est donc :

```

1 def Geom(p):
2     X=1
3     while random() >= p:
4         X=X+1
5     return X

```

Remarque 6.2. Soulignons ici que dans la boucle `while`, on fait bien des appels *successifs* à la fonction `random`, ce qui fait que l'on simule bien une suite de variables $\mathcal{U}(0, 1)$ indépendantes (et la boucle s'arrête dès que l'une des variables est plus grande que p).

Loi de Poisson

Pour simuler une loi de Poisson de paramètre $\lambda > 0$, on va utiliser le fait suivant.

Si $(Y_i)_{i \geq 1}$ est une suite de variables aléatoires indépendantes de loi $\mathcal{E}(1)$, et si $S_k = \sum_{i=1}^k Y_i$ pour $k \geq 0$ ($S_0 = 0$), alors $N = \sup\{k \geq 0, S_k \leq \lambda\}$ suit une loi de Poisson de paramètre λ .

Démonstration. Pour montrer cela, notons que l'on a l'égalité d'événements $\{N \geq n\} = \{S_n \leq \lambda\}$, et que S_n suit une loi Gamma($n, 1$), de densité $\frac{1}{(n-1)!} x^{n-1} e^{-x}$ sur \mathbb{R}_+ (voir la Section 5.4, ou le démontrer par récurrence). On a donc

$$\begin{aligned} \mathbb{P}(N \geq n) &= \mathbb{P}(S_n \leq \lambda) = \int_0^\lambda \frac{x^{n-1}}{(n-1)!} e^{-x} dx \\ &\stackrel{(\text{IPP})}{=} \frac{\lambda^n}{n!} e^{-\lambda} + \int_0^\lambda \frac{x^n}{n!} e^{-x} dx = \frac{\lambda^n}{n!} e^{-\lambda} + \mathbb{P}(N \geq n+1). \end{aligned}$$

Pour conclure, on écrit que $\mathbb{P}(N = n) = \mathbb{P}(N \geq n+1) - \mathbb{P}(N \geq n) = \frac{\lambda^n}{n!} e^{-\lambda}$. □

Pour générer une variable aléatoire de Poisson, il suffit donc de calculer S_1, S_2, \dots , en vérifiant à chaque étape si $S_i \leq \lambda$: au premier instant k où $S_k > \lambda$, on renvoie $k-1$. Il faut donc utiliser une boucle `while`, où à chaque étape on augmente un compteur de 1 et on rajoute à S_i une variable aléatoire $\mathcal{E}(1)$ — on va utiliser le fait que si $U \sim \mathcal{U}(0, 1)$, alors $-\ln U$ suit une loi $\mathcal{E}(1)$ (voir plus bas). Une fonction `Poisson` qui prend en argument un réel $a > 0$ et génère une variable aléatoire de loi $\text{Poi}(a)$ peut donc s'écrire :

```

1 def Poisson(a):
2     k=0
3     S=0
4     while S <= a:
5         k=k+1
6         S=S-log(random())
7     return k-1

```

Le cas de variables à valeurs dans un ensemble fini

Considérons une variable aléatoire X , à valeurs dans un ensemble $E = \{x_1, \dots, x_n\}$ fini. On notera $p_k := \mathbb{P}(X = x_k)$ pour $k \in \{1, \dots, n\}$ la densité discrète de X . Pour simuler la variable aléatoire X , on peut procéder de la manière suivante :

- on découpe l'intervalle $[0, 1[$ en intervalles I_k (disjoints) de longueur p_k — on prend de manière canonique $I_k = [P_{k-1}, P_k[$ où $P_k = \sum_{i=1}^k p_i$ (et $P_0 = 0$) ;
- on génère une variable aléatoire $U \sim \mathcal{U}(0, 1)$, qui tombera dans un des intervalles I_k ;
- si U tombe dans l'intervalle I_k alors on renvoie la valeur x_k .

Comme $U \sim \mathcal{U}(0, 1)$, on a en effet $\mathbb{P}(U \in I_k) = p_k$: la probabilité que cet algorithme renvoie la valeur x_k est donc p_k ($= \mathbb{P}(X = x_k)$). En pratique, on va générer une variable aléatoire $U \sim \mathcal{U}(0, 1)$ *une fois pour toute*, et passer en revue les intervalles $I_1 = [0, P_1[$, $I_2 = [P_1, P_2[$, ..., $I_n = [P_{n-1}, 1[$, jusqu'à ce que l'on trouve l'intervalle qui contient la valeur U (c'est-à-dire que l'on s'arrête au premier moment où $U < P_k$) ; on renverra alors la valeur x_k correspondant à l'intervalle dans lequel est tombé U .

Exemple 6.3. Simulons une variable aléatoire X à valeurs dans $\{-1, 0, 2\}$, de loi donnée par $\mathbb{P}(X = -1) = \frac{1}{3}$, $\mathbb{P}(X = 0) = \frac{1}{2}$, $\mathbb{P}(X = 2) = \frac{1}{6}$. On découpe $[0, 1[$ en trois intervalles $[0, \frac{1}{3}[$, $[\frac{1}{3}, \frac{5}{6}[$ et $[\frac{5}{6}, 1[$, et on peut donc écrire le programme suivant pour générer une telle variable aléatoire :

```
1 def X():
2     U=random()
3     if U<1/3:
4         return -1
5     elif U<5/6:
6         return 0
7     else:
8         return 2
```

De manière générale, si on suppose données deux fonctions p et x telles que $p(k)$ renvoie p_k et $x(k)$ renvoie x_k , une fonction `Variable` qui permet de simuler une variable aléatoire de loi donnée par la densité discrète $p_k = \mathbb{P}(X = x_k)$ peut s'écrire comme suit :

```
1 def Variable():
2     U=random()
3     k,P=0,0
4     while P<=U:
5         k=k+1
6         P=P+p(k)
7     return x(k)
```

6.2 Méthode d'inversion et simulation de lois absolument continues

On a vu dans le chapitre précédent (Théorème 5.16), que l'on peut utiliser une variable aléatoire $U \sim \mathcal{U}(0, 1)$ pour construire n'importe quelle variable aléatoire réelle. Rappelons le résultat du Théorème 5.16.

Soit X une variable aléatoire réelle, de fonction de répartition F , et soit $\varphi :]0, 1[\rightarrow \mathbb{R}$ le *pseudo-inverse* de F , c'est-à-dire

$$\varphi(y) := \inf\{z \in \mathbb{R} : F(z) \geq y\}. \quad (6.1)$$

Alors si $U \sim \mathcal{U}(0, 1)$, la variable aléatoire $\varphi(U)$ a pour fonction de répartition F , et a donc la même loi que X .

En pratique, la *méthode d'inversion* pour simuler une variable aléatoire X consiste à :

- calculer (explicitement) la fonction de répartition $F = F_X$ de X ;
- calculer le pseudo-inverse φ de F (si F est bijective d'un intervalle $]a, b[$ dans $]0, 1[$, alors φ est simplement l'inverse de F) ;
- simuler $\varphi(U)$, où $U \sim \mathcal{U}(0, 1)$, en utilisant la fonction `random`.

Cette méthode est extrêmement utile, mais il faut cependant être capable de *calculer* et *d'inverser* (ou de pseudo-inverser) la fonction de répartition F ...

Loi uniforme continue sur un intervalle $]a, b[$

Soient $a, b \in \mathbb{R}$ avec $a < b$. Appliquons la méthode d'inversion pour simuler une variable aléatoire X de loi $\mathcal{U}(a, b)$. La fonction de répartition de X est explicite, et donnée par

$$F_X(x) = 0 \text{ si } x \leq a ; \quad F_X(x) = \frac{x - a}{b - a} \text{ si } x \in]a, b[; \quad F_X(x) = 1 \text{ si } x \geq b.$$

Ainsi, F_X est une bijection de $]a, b[$ dans $]0, 1[$, et on peut calculer son inverse φ en inversant l'équation $F_X(\varphi(x)) = \frac{\varphi(x) - a}{b - a} = x$: on trouve donc que $\varphi(x) = (b - a)x + a$. Donc si $U \sim \mathcal{U}(0, 1)$, alors $\varphi(U) = (b - a)U + a$ suit une loi uniforme sur $]a, b[$. Voici donc une fonction `Unif` qui prend en argument deux réels a, b avec $a < b$ et génère une variable aléatoire de loi $\mathcal{U}(a, b)$:

```
1 def Unif(a, b):
2     return (b-a)*random()+a
```

Loi exponentielle

Soit $\lambda > 0$ et X une variable aléatoire $\mathcal{E}(\lambda)$. Alors la fonction de répartition de X est explicite, et donnée par

$$F_X(x) = 0 \text{ si } x \leq 0 ; \quad F_X(x) = 1 - e^{-\lambda x} \text{ si } x > 0.$$

Ainsi, F_X est une bijection de $]0, +\infty[$ dans $]0, 1[$, et on peut calculer son inverse φ en inversant l'équation $F_X(\varphi(x)) = 1 - e^{-\lambda\varphi(x)} = x$: on trouve donc que $\varphi(x) = -\frac{1}{\lambda} \ln(1 - x)$. Donc si $U \sim \mathcal{U}(0, 1)$, alors $\varphi(U) = -\frac{1}{\lambda} \ln(1 - U)$ suit une loi exponentielle de paramètre λ (notons que l'on a aussi $1 - U \sim \mathcal{U}(0, 1)$, et on a donc aussi $-\frac{1}{\lambda} \ln U \sim \mathcal{E}(\lambda)$). Voici une fonction `Expo` qui prend en argument un réel $a > 0$ et qui permet de simuler une variable aléatoire de loi $\mathcal{E}(a)$:

```
1 def Expo(a):
2     return -log(random())/a
```

Remarque 6.4. Notons que si $X \sim \mathcal{E}(\lambda)$, alors $Y = \lfloor X \rfloor + 1$ suit une loi géométrique de paramètre $p = 1 - e^{-\lambda}$ (exercice). Ainsi, pour simuler une loi géométrique de paramètre p , il suffit de simuler une loi exponentielle de paramètre $\lambda = -\ln(1 - p)$, d'en prendre la partie entière et de rajouter 1.

Loi de Cauchy

Pour simuler une variable aléatoire de Cauchy dont la densité est $f(x) = \frac{1}{\pi(1+x^2)}$ sur \mathbb{R} , on peut aussi utiliser la méthode d'inversion. Sa fonction de répartition est $F_X(x) = \frac{1}{\pi} \arctan(x) + \frac{1}{2}$, qui est une fonction bijective de \mathbb{R} dans \mathbb{R} , d'inverse $\varphi(x) = F_X^{-1}(x) = \tan(\pi(x - \frac{1}{2}))$. Ainsi, si $U \sim \mathcal{U}(0, 1)$, $\varphi(U) = \tan(\pi(U - \frac{1}{2}))$ suit la loi de Cauchy. Voici une fonction `Cauchy` qui permet de générer une variable de Cauchy :

```
1 def Cauchy():
2     return tan(pi*(random()-0.5))
```

6.3 Le cas de la loi normale

Pour simuler une variable aléatoire de loi normale $\mathcal{N}(0, 1)$, la méthode d'inversion ne nous aide pas vraiment : en effet, la fonction de répartition n'admet pas de formule explicite, et s'écrit simplement $\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$. Calculer explicitement l'inverse de Φ n'est pas non plus possible... Cependant, on a vu dans l'Exemple 5.34 le résultat suivant :

Si U_1, U_2 sont deux variables aléatoires indépendantes, de loi $\mathcal{U}(0, 1)$, alors

$$X = \sqrt{-2 \log U_1} \times \cos(2\pi U_2), \quad Y = \sqrt{-2 \log U_1} \times \sin(2\pi U_2),$$

sont deux variables aléatoires indépendantes, de loi $\mathcal{N}(0, 1)$.

Il s'agit de la méthode dite de *Box-Muller* pour simuler une loi normale. Voici donc une fonction `Normale` qui permet de simuler une variable aléatoire de loi normale $\mathcal{N}(0, 1)$:

```
1 def Normale():
2     return sqrt(-2*log(random()))*cos(2*pi*random())
```

On souligne que dans ce programme, les deux appels à la fonction `random` renvoient deux réalisations de variables aléatoires $\mathcal{U}(0, 1)$ indépendantes.

Pour simuler une variable aléatoire de loi $\mathcal{N}(\mu, \sigma^2)$, où $\mu \in \mathbb{R}$ et $\sigma^2 \geq 0$ sont des paramètres donnés, il suffit d'utiliser le fait que si $X \sim \mathcal{N}(0, 1)$, alors $\sigma X + \mu \sim \mathcal{N}(\mu, \sigma^2)$.

6.4 Méthode de rejet : cas d'une densité bornée et à support borné

Si X est une variable aléatoire de densité f dont on n'est pas capable de calculer la fonction de répartition ou sa pseudo-inverse explicitement, on ne peut pas appliquer la méthode d'inversion décrite plus haut. On considère ici la méthode dite *de rejet*, que l'on présentera uniquement dans le cas où la densité f est : i) bornée ; ii) à support $\text{supp}(f) := \{x, f(x) > 0\}$ borné.

Proposition 6.5. Soit g une fonction positive, nulle en dehors d'un intervalle $[a, b]$, et bornée par une constante M . Soient $(U_i)_{i \in \mathbb{N}}$ et $(V_i)_{i \in \mathbb{N}}$ deux suites de variables aléatoires indépendantes, de loi $\mathcal{U}(a, b)$ et $\mathcal{U}(0, M)$ respectivement. On pose

$$T = \min\{i, V_i < g(U_i)\}, \quad X := U_T \quad (X(\omega) = U_{T(\omega)}(\omega) \text{ pour } \omega \in \Omega).$$

Alors X est une variable aléatoire de densité $f(x) := \frac{1}{A}g(x)$, où $A := \int_a^b g(x)dx < +\infty$.

En utilisant cette proposition, on peut simuler une variable aléatoire dont la densité est bornée et à support borné, proportionnelle à une fonction g comme dans l'énoncé de la proposition : il suffit de générer successivement des variables aléatoires $U_1, V_1, U_2, V_2, \dots$ telles que $U_i \sim \mathcal{U}(a, b)$, $V_i \sim \mathcal{U}(0, M)$, et on les rejette jusqu'au premier instant T pour lequel $V_i < g(U_i)$. En d'autres termes (U_i, V_i) sont des points pris uniformément dans le rectangle $[a, b] \times [0, M]$, et on les rejette jusqu'à ce que (U_i, V_i) tombe sous la courbe de g . On renverra alors la valeur de U_T , qui est une variable aléatoire de densité f proportionnelle à g , le coefficient de proportionnalité étant égal à $1/A$, où A est l'aire sous la courbe de g .

Exemple 6.6. Soit X une variable aléatoire de densité $f(x) = \frac{2}{\pi}\sqrt{1-x^2}\mathbb{1}_{]-1,1[}(x)$, qui suit la loi dite du *demi-cercle**. La densité est nulle en dehors de $[-1, 1]$, et notons de plus que $g(x) := \sqrt{1-x^2} \leq 1$ pour tout $x \in [-1, 1]$. D'après la méthode de rejet, on peut donc simuler une variable aléatoire de densité $f(x) = \frac{2}{\pi}g(x)$ (notons que $A = \int_{-1}^1 g(x)dx = \frac{\pi}{2}$) en générant successivement des variables aléatoires $U_1, V_1, U_2, V_2, \dots$ telles que $U_i \sim \mathcal{U}(-1, 1)$, $V_i \sim \mathcal{U}(0, 1)$, jusqu'à ce que $V_i \leq g(U_i)$; à cet instant T , U_T est une variable aléatoire dont la densité est donnée par $f(x) = \frac{2}{\pi}g(x)$.

Ainsi, une fonction `DemiCercle` qui permet de simuler une variable aléatoire de densité $f(x) = \frac{2}{\pi}\sqrt{1-x^2}\mathbb{1}_{]-1,1[}(x)$ peut donc s'écrire comme suit :

```

1 def DemiCercle():
2     U=2*random()-1
3     V=random()
4     while V>=sqrt(1-U**2):
5         U=2*random()-1
6         V=random()
7     return U

```

Démonstration de la Proposition 6.5. Posons $X = U_T$, et déterminons la fonction de répartition $F_X(t)$ de X . Comme $U_i \in [a, b]$ pour tout $i \in \mathbb{N}$, on sait déjà que $X \in [a, b]$, de sorte que l'on ne considère que le cas $t \in [a, b]$. L'idée essentielle est de décomposer la probabilité $F_X(t) = \mathbb{P}(X \leq t) = \mathbb{P}(U_T \leq t)$ suivant la valeur de T : on écrit, pour tout $t \in [a, b]$,

$$F_X(t) = \mathbb{P}(U_T \leq t) = \sum_{n=1}^{+\infty} \mathbb{P}(T = n, U_n \leq t).$$

Ici, les événements $\{T = n\}$ et $U_n \leq t$ ne sont pas indépendants, mais on peut décomposer l'événement $\{T = n\}$, pour écrire

$$\begin{aligned} \mathbb{P}(T = n, U_n \leq t) &= \mathbb{P}(V_1 \geq g(U_1), \dots, V_{n-1} \geq g(U_{n-1}), V_n < g(U_n), U_n \leq t) \\ &= \mathbb{P}(V_1 \geq g(U_1)) \cdots \mathbb{P}(V_{n-1} \geq g(U_{n-1})) \mathbb{P}(V_n < g(U_n), U_n \leq t) \end{aligned} \quad (6.2)$$

*. Notons qu'il est possible de calculer la fonction de répartition de X , mais pas de l'inverser.

où on a utilisé le fait que les vecteurs aléatoires $((U_i, V_i))_{i \in \mathbb{N}}$ sont indépendants, de sorte que les événements $\{V_i \geq g(U_i)\}$ sont indépendants.

Maintenant, on peut calculer les probabilités $\mathbb{P}(V_i \geq g(U_i))$ et $\mathbb{P}(V_n < g(U_n), U_n \leq t)$. Si $U \sim \mathcal{U}(a, b)$ et $V \sim \mathcal{U}(0, M)$ sont indépendantes, la densité jointe du vecteur (U, V) est le produit des densités de U et V , c'est-à-dire

$$f_{(U,V)}(u, v) = \frac{1}{b-a} \mathbb{1}_{[a,b]}(u) \times \frac{1}{M} \mathbb{1}_{[0,M]}(v) = \frac{1}{(b-a)M} \mathbb{1}_{[a,b] \times [0,M]}(u, v)$$

(on reconnaît la densité d'un vecteur aléatoire de loi uniforme sur $[a, b] \times [0, M]$). Ainsi, en utilisant la formule de transfert, et en notant $D_g := \{(x, y), a \leq x \leq b, 0 \leq y < g(x)\}$ le domaine situé sous la courbe de g , on peut calculer la probabilité $\mathbb{P}(V < g(U))$, égale à

$$\mathbb{P}((U, V) \in D_g) = \int_{D_g} f_{(U,V)}(u, v) du dv = \frac{1}{(b-a)M} \int_{D_g} du dv = \frac{A}{(b-a)M}.$$

où on a utilisé que $D_g \subset [a, b] \times [0, M]$, puis que la dernière intégrale est égale à l'aire du domaine D_g , c'est-à-dire $A = \int_a^b g(x) dx$. Soulignons au passage que comme les vecteurs aléatoires $((U_i, V_i))_{i \in \mathbb{N}}$ sont indépendants, la variable aléatoire T est donc une variable géométrique, de paramètre $\rho := \frac{A}{(b-a)M}$. En passant au complémentaire, on obtient donc que $\mathbb{P}(V \geq g(U)) = 1 - \rho$.

De la même manière, pour $t \in [a, b]$, en notant $D_g(t) := \{(x, y), a \leq x \leq t, 0 \leq y < g(x)\}$ le domaine situé sous la courbe de g à gauche de t , on obtient

$$\mathbb{P}(V < g(U), U \leq t) = \mathbb{P}((U, V) \in D_g(t)) = \frac{1}{(b-a)M} \int_a^t g(x) dx = \rho \times \frac{1}{A} \int_a^t g(x) dx,$$

En revenant à (6.2), on obtient donc que pour $t \in [a, b]$

$$\mathbb{P}(T = n, U_n \leq t) = (1 - \rho)^{n-1} \rho \times \frac{1}{A} \int_a^t g(x) dx,$$

et en sommant sur n (en utilisant que $\sum_{n=1}^{+\infty} (1 - \rho)^{n-1} \rho = 1$), on trouve donc

$$F_X(t) = \sum_{n=1}^{+\infty} \mathbb{P}(T = n, U_n \leq t) = \frac{1}{A} \int_a^t g(x) dx.$$

Cela conclut donc la démonstration, au vu de la Définition 5.1 (on rappelle que g est nulle en dehors de $[a, b]$). \square

★ Simulation d'un modèle aléatoire : la série harmonique de signe aléatoire

On sait que la série alternée $\sum_{i=1}^n \frac{(-1)^i}{i}$ converge (vers $\ln 2$), mais ne converge pas absolument... Le *théorème de réarrangement de Riemann*[†] assure alors que quel que soit $\ell \in \mathbb{R} \cup \{-\infty, +\infty\}$ que l'on se fixe, on peut "réarranger" les termes de la somme pour qu'elle converge vers ℓ : plus précisément il existe une bijection $\sigma : \mathbb{N}_* \rightarrow \mathbb{N}_*$ telle que $\lim_{n \rightarrow +\infty} \sum_{i=1}^n \frac{(-1)^{\sigma(i)}}{i} = \ell$. En tant que

[†]. Voir par exemple https://fr.wikipedia.org/wiki/Théorème_de_réarrangement_de_Riemann.

probabiliste, on se pose la question suivante : que se passe-t-il si on met un “signe aléatoire” à la place du signe alterné $(-1)^i$?

Soit $(X_i)_{i \in \mathbb{N}^*}$ une suite de variables aléatoires indépendantes, de même loi $\mathbb{P}(X_i = +1) = \mathbb{P}(X_i = -1) = \frac{1}{2}$, et posons

$$W_0 = 0, \quad W_n := \sum_{i=1}^n \frac{X_i}{i}.$$

La question est donc de savoir si la suite aléatoire $(W_n)_{n \in \mathbb{N}}$ converge. Cette question est a priori difficile : effectuons des simulations informatiques de la suite $(W_n)_{n \in \mathbb{N}}$, pour se faire une idée. La fonction suivante génère une variable aléatoire X_i définie plus haut.

```

1 def Signe():
2     if random() < 0.5:
3         return 1
4     else:
5         return -1

```

La fonction suivante prend en argument un entier n , et génère une suite $(W_i)_{0 \leq i \leq n}$:

```

1 def Serie(n):
2     W = [0]
3     for i in range(n):
4         W.append(W[i] + Signe() / (i+1))
5     return W

```

La liste $W = [W_0, W_1, \dots, W_n]$ contient donc les valeurs successives de la suite $(W_i)_{i \in \mathbb{N}}$, et on peut donc dessiner le graphe de la fonction $i \mapsto W_i$ pour avoir une idée de si la suite $(W_i)_{i \in \mathbb{N}}$ se “stabilise”. La Figure 6.1 suggère que : i) dans tous les cas, la suite W_n converge ; ii) la limite de la suite W_n est différente à chaque fois.

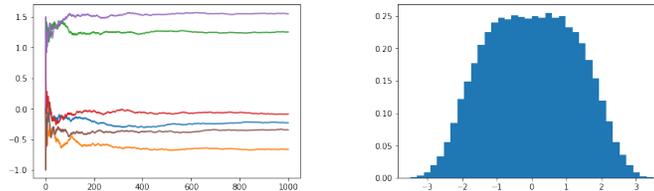


FIGURE 6.1 – À gauche, une représentation graphique de 6 simulations informatiques de la suite $(W_n)_{0 \leq n \leq 1000}$. À droite, un histogramme de 100 000 répétitions indépendantes de W_{1000} , qui donne une idée de la loi de la “variable aléatoire limite” $W = \lim_{n \rightarrow +\infty} W_n$.

Ainsi, on peut s’attendre à un résultat du type : “la suite $(W_n)_{n \in \mathbb{N}}$ converge, et la limite $W = \lim_{n \rightarrow +\infty} W_n$ est aléatoire”. On peut en fait démontrer cette intuition (il s’agit d’un résultat difficile).

Proposition 6.7. Soit A l’événement “la suite $(W_n)_{n \in \mathbb{N}}$ est de Cauchy”, que l’on peut écrire sous la forme $A = \bigcap_{k \in \mathbb{N}} \bigcup_{N \in \mathbb{N}} \bigcap_{m, n \geq N} \{|W_m - W_n| \leq \frac{1}{k}\}$. Alors $\mathbb{P}(A) = 1$.

Autrement dit, la probabilité que la suite (aléatoire) $(W_n)_{n \geq 0}$ converge vaut 1 (car une suite réelle est convergente si et seulement si elle est de Cauchy).

Démonstration (difficile !) On va chercher à montrer que $\mathbb{P}(A^c) = 0$: on a par sous-additivité, et par continuité par le haut des probabilités

$$\begin{aligned} \mathbb{P}(A^c) &= \mathbb{P}\left(\bigcup_{k \in \mathbb{N}} \bigcap_{N \in \mathbb{N}} \bigcup_{m, n \geq N} \left\{|W_m - W_n| > \frac{1}{k}\right\}\right) \\ &\leq \sum_{k \in \mathbb{N}} \mathbb{P}\left(\bigcap_{N \in \mathbb{N}} \bigcup_{m, n \geq N} \left\{|W_m - W_n| > \frac{1}{k}\right\}\right) = \sum_{k \in \mathbb{N}} \lim_{N \rightarrow \infty} \mathbb{P}\left(\bigcup_{m, n \geq N} \left\{|W_m - W_n| > \frac{1}{k}\right\}\right). \end{aligned}$$

Notons que si $|W_m - W_n| > \frac{1}{k}$, alors au moins une des inégalités $|W_m - W_N| > \frac{1}{2k}$ ou $|W_n - W_N| > \frac{1}{2k}$ est vérifiée, de sorte que

$$\mathbb{P}(A^c) \leq \sum_{k \in \mathbb{N}} \lim_{N \rightarrow \infty} \mathbb{P} \left(\bigcup_{m \geq N} \left\{ |W_m - W_N| > \frac{1}{2k} \right\} \right). \quad (6.3)$$

Pour estimer cette dernière probabilité, on utilise la proposition suivante :

Proposition 6.8 (Inégalité maximale de Kolmogorov). *Soit $(Y_i)_{i \in \mathbb{N}}$ une suite de variables aléatoires indépendantes (pas nécessairement de même loi), qui admettent un moment d'ordre 2 fini et telles que $\mathbb{E}(Y_i) = 0$ pour tout $i \in \mathbb{N}$. Si on pose $S_0 = 0$ et $S_n := \sum_{i=1}^n Y_i$, alors pour tout $x > 0$, on a l'inégalité $\mathbb{P}(\max_{1 \leq i \leq n} |S_i| \geq x) \leq \frac{1}{x^2} \text{Var}(S_n) = \frac{1}{x^2} \sum_{i=1}^n \text{Var}(Y_n)$.*

Comme $\mathbb{E}(X_i) = 0$ et $\text{Var}(X_i) = 1$, en appliquant ce résultat à $W_m - W_N = \sum_{i=N+1}^m \frac{X_i}{i}$, on obtient pour tout $M > N$

$$\mathbb{P} \left(\max_{M \geq m \geq N} |W_m - W_N| \geq \frac{1}{2k} \right) \leq 4k^2 \sum_{i=N+1}^M \text{Var} \left(\frac{X_i}{i} \right) = 4k^2 \sum_{i=N+1}^M \frac{1}{i^2}.$$

Ainsi, en prenant $M \rightarrow +\infty$, par continuité par le bas des probabilités, on a

$$\mathbb{P} \left(\bigcup_{m \geq N} \left\{ |W_m - W_N| > \frac{1}{2k} \right\} \right) = \lim_{M \rightarrow \infty} \mathbb{P} \left(\bigcup_{M \geq m \geq N} \left\{ |W_m - W_N| > \frac{1}{2k} \right\} \right) \leq 4k^2 \sum_{i=N+1}^{+\infty} \frac{1}{i^2} \xrightarrow{N \rightarrow \infty} 0.$$

On en conclut que tous les termes dans la somme dans (6.3) sont nuls, de sorte que $\mathbb{P}(A^c) = 0$. \square

Démonstration de la Proposition 6.8. On note $T = \min\{i \geq 1, |S_i| \geq x\}$ le premier instant où la suite $(S_i)_{i \geq 0}$ dépasse, en valeur absolue, x : on a $\{\max_{1 \leq i \leq n} |S_i| \geq x\} = \{T \leq n\}$, et on doit donc majorer la probabilité $\mathbb{P}(T \leq n)$. Par monotonie de l'espérance, on a

$$\text{Var}(S_n) = \mathbb{E}(S_n^2) \geq \mathbb{E}(S_n^2 \mathbb{1}_{\{T \leq n\}}) = \sum_{k=1}^n \mathbb{E}(S_n^2 \mathbb{1}_{\{T=k\}}),$$

où on a décomposé l'indicatrice $\mathbb{1}_{\{T \leq n\}} = \sum_{k=1}^n \mathbb{1}_{\{T=k\}}$, et utilisé la linéarité de l'espérance. Maintenant, on peut développer le carré

$$S_n^2 = (S_k + S_n - S_k)^2 = S_k^2 + 2S_k(S_n - S_k) + (S_n - S_k)^2 \geq S_k^2 + 2S_k(S_n - S_k).$$

En multipliant par $\mathbb{1}_{\{T=k\}}$ et en prenant l'espérance, on a donc

$$\mathbb{E}(S_n^2 \mathbb{1}_{\{T=k\}}) \geq \mathbb{E}(S_k^2 \mathbb{1}_{\{T=k\}}) + 2\mathbb{E}(S_k(S_n - S_k) \mathbb{1}_{\{T=k\}}).$$

Notons que si $T = k$ on a nécessairement $|S_k| \geq x$, et donc $\mathbb{E}(S_k^2 \mathbb{1}_{\{T=k\}}) \geq x^2 \mathbb{E}(\mathbb{1}_{\{T=k\}}) = x^2 \mathbb{P}(T = k)$. De plus, $S_k \mathbb{1}_{\{T=k\}}$ est indépendant de $S_n - S_k$: en effet, S_k et $\mathbb{1}_{\{T=k\}}$ ne dépendent que de (Y_1, \dots, Y_k) , et $S_n - S_k$ de (Y_{k+1}, \dots, Y_n) . On obtient donc que $\mathbb{E}(S_k(S_n - S_k) \mathbb{1}_{\{T=k\}}) = \mathbb{E}(S_k \mathbb{1}_{\{T=k\}}) \mathbb{E}(S_n - S_k) = 0$. On a donc finalement obtenu que $\mathbb{E}(S_n^2 \mathbb{1}_{\{T=k\}}) \geq x^2 \mathbb{P}(T = k)$, et ainsi

$$\text{Var}(S_n) \geq \sum_{k=1}^n x^2 \mathbb{P}(T = k) = x^2 \mathbb{P}(T \leq n),$$

ce qui conclut la démonstration. \square

Chapitre 7

Loi des grands nombres et convergences de suites de variables aléatoires

Comme pour les suites de fonctions, on peut définir différentes notions de convergence pour les suites de variables aléatoires, qui ne sont pas toutes équivalentes. On se concentrera dans ce cours sur un type de convergence particulier, appelé *convergence en probabilité*. Les autres types de convergence seront étudiés plus en détail en L3.

7.1 La loi (faible) des grands nombres et applications

La loi des grands nombres rend rigoureuse l'idée de l'approche fréquentiste des probabilités. Si on répète une expérience un très grand nombre N de fois, et que l'on note les valeurs X_1, \dots, X_N d'une variable aléatoire X associée à l'expérience, alors la moyenne $\frac{X_1 + \dots + X_N}{N}$ des valeurs observées devrait être "proche" de l'espérance $\mathbb{E}(X)$. Dans le cas où $X = \mathbb{1}_A$, alors $X_1 + \dots + X_N$ compte le nombre N_A de fois où on a observé l'événement A , et $\frac{X_1 + \dots + X_N}{N} = \frac{N_A}{N}$ devrait être proche de $\mathbb{E}(X) = \mathbb{E}(\mathbb{1}_A) = \mathbb{P}(A)$.

Si $(X_i)_{i \geq 1}$ est une suite de variables aléatoires, on définit pour $n \geq 1$

$$\bar{X}_n := \frac{1}{n} \sum_{i=1}^n X_i, \quad (7.1)$$

que l'on appelle *moyenne empirique*. Soulignons que \bar{X}_n est elle-même une variable aléatoire... La loi des grands nombres montre que, sous certaines hypothèses, la suite de variables aléatoires $(\bar{X}_n)_{n \geq 1}$ "se concentre" autour de la valeur $\mu = \mathbb{E}(X)$.

Théorème 7.1 (Loi des grands nombres). Soit $(X_i)_{i \geq 1}$ une suite de variables aléatoires réelles i.i.d. (indépendantes et identiquement distribuées), qui admettent une espérance finie. On note $\mu = \mathbb{E}(X_i)$. Alors pour tout $\varepsilon > 0$, on a

$$\lim_{n \rightarrow \infty} \mathbb{P}\left(|\bar{X}_n - \mu| > \varepsilon\right) = 0.$$

On va commencer par démontrer une version plus faible de ce résultat.

Proposition 7.2. Soit $(X_i)_{i \geq 1}$ une suite de variables aléatoires réelles i.i.d., qui admettent un moment d'ordre 2 fini — les X_i admettent aussi une espérance finie (voir Section 3.4.3). On note $\mu = \mathbb{E}(X_i)$, et $\sigma^2 = \text{Var}(X_i) < +\infty$. Alors pour tout $\varepsilon > 0$, on a $\lim_{n \rightarrow \infty} \mathbb{P}(|\bar{X}_n - \mu| > \varepsilon) = 0$.

Démonstration. Notons que la variable aléatoire \bar{X}_n admet un moment d'ordre 2 fini. Ainsi, \bar{X}_n admet une espérance et une variance finie, que l'on peut calculer explicitement.

Par linéarité de l'espérance, on obtient directement que $\mathbb{E}(\bar{X}_n) = \mu$, par définition de \bar{X}_n . De plus, comme les variables aléatoires $(X_i)_{i \in \mathbb{N}}$ sont indépendantes, on a

$$\text{Var}(\bar{X}_n) = \frac{1}{n^2} \sum_{i=1}^n \text{Var}(X_i) = \frac{\sigma^2}{n}. \quad (7.2)$$

En appliquant l'inégalité de Bienaymé-Tchebychev à la variable aléatoire \bar{X}_n , on obtient donc, pour tout $\varepsilon > 0$

$$\mathbb{P}(|\bar{X}_n - \mu| \geq \varepsilon) \leq \frac{\text{Var}(\bar{X}_n)}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2}, \quad (7.3)$$

qui tend bien vers 0 quand $n \rightarrow \infty$. □

Exercice 7.3. En reprenant la démonstration, montrer que l'on peut affaiblir les hypothèses de la Proposition 7.2 : la conclusion reste valable si les $(X_i)_{i \in \mathbb{N}}$ admettent un moment d'ordre 2 fini et que

- a) les X_i ont la même espérance, $\mathbb{E}(X_i) = \mu$;
- b) $|\text{Cov}(X_i, X_j)| \leq \rho(|j - i|)$ pour tous $i, j \geq 1$, où $\rho : \mathbb{N} \rightarrow \mathbb{R}_+$ est telle que $\rho(k) \rightarrow 0$ quand $k \rightarrow \infty$.

Indication : montrer que $\text{Var}(\bar{X}_n) \leq \frac{1}{n} \sum_{k=1}^n \rho(k)$.

Démonstration du Théorème 7.1 (difficile). La démonstration est un peu technique, mais pas insurmontable : elle utilise les mêmes principes que dans la démonstration de la Proposition 7.2, combinée avec une technique dite de *troncation*.

On va accepter ici le résultat suivant, naturel au vu de la continuité par le bas des probabilités : si $(A_m)_{m \in \mathbb{N}}$ est une suite *croissante* d'événements telle que $\mathbb{P}(\bigcup_{m \in \mathbb{N}} A_m) = 1$, et si X est une variable aléatoire d'espérance finie, alors on a $\lim_{m \rightarrow \infty} \mathbb{E}(X \mathbb{1}_{A_m}) = \mathbb{E}(X)$ et aussi $\lim_{m \rightarrow \infty} \mathbb{E}(|X| \mathbb{1}_{A_m^c}) = 0$ (il s'agit du théorème de convergence monotone en théorie de la mesure, qui sera vu en L3).

Fixons $\varepsilon > 0$. Prenons $\delta > 0$ arbitraire. Alors, on peut choisir $M > 0$ suffisamment grand de sorte que $|\mathbb{E}(X \mathbb{1}_{\{|X| \leq M\}}) - \mu| \leq \varepsilon/3$ et $\mathbb{E}(|X| \mathbb{1}_{\{|X| \leq M\}}) \leq \delta\varepsilon/3$ (les choix de $\varepsilon/3$ et $\delta\varepsilon/3$ au lieu de ε et δ sont faits pour simplifier certains calculs par la suite). Pour tout $i \in \mathbb{N}$, on note alors

$$\tilde{X}_i = X_i \mathbb{1}_{\{X_i \leq M\}} \quad \text{et} \quad \hat{X}_i = X_i \mathbb{1}_{\{X_i > M\}}, \quad (7.4)$$

de sorte que $X_i = \tilde{X}_i + \hat{X}_i$, et $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n \tilde{X}_i + \frac{1}{n} \sum_{i=1}^n \hat{X}_i$. Par inégalité triangulaire, on obtient alors l'inclusion d'événements suivante

$$\left\{ |\bar{X}_n - \mu| > \varepsilon \right\} \subset \left\{ \left| \frac{1}{n} \sum_{i=1}^n \tilde{X}_i - \mathbb{E}(\tilde{X}_1) \right| > \frac{\varepsilon}{3} \right\} \cup \left\{ |\mathbb{E}(\tilde{X}_1) - \mu| > \frac{\varepsilon}{3} \right\} \cup \left\{ \left| \frac{1}{n} \sum_{i=1}^n \hat{X}_i \right| > \frac{\varepsilon}{3} \right\},$$

en ayant aussi noté que si la somme de trois termes est plus grande que ε , alors nécessairement au moins l'un des trois termes est plus grand que $\varepsilon/3$. Notons que l'on a choisi $M > 0$ de sorte que $|\mathbb{E}(\tilde{X}_1) - \mu| \leq \varepsilon/3$, et donc $\{|\mathbb{E}(\tilde{X}_1) - \mu| > \frac{\varepsilon}{3}\} = \emptyset$. En utilisant la sous-additivité des probabilités, on a donc

$$\mathbb{P}(|\bar{X}_n - \mu| > \varepsilon) \leq \mathbb{P}\left(\left|\frac{1}{n} \sum_{i=1}^n \tilde{X}_i - \mathbb{E}(\tilde{X}_1)\right| > \frac{\varepsilon}{3}\right) + \mathbb{P}\left(\left|\frac{1}{n} \sum_{i=1}^n \hat{X}_i\right| > \frac{\varepsilon}{3}\right). \quad (7.5)$$

Maintenant, les variables aléatoires \tilde{X}_i sont indépendantes, de même loi, et bornées par M : elles admettent un deuxième moment fini, et on peut donc leur appliquer la Proposition 7.2, et obtenir que

$$\lim_{n \rightarrow +\infty} \mathbb{P}\left(\left|\frac{1}{n} \sum_{i=1}^n \tilde{X}_i - \mathbb{E}(\tilde{X}_1)\right| > \frac{\varepsilon}{3}\right) = 0.$$

Ainsi, il existe n_0 tel que pour tout $n \geq n_0$, le premier terme dans (7.5) est plus petit que δ . Pour le deuxième terme, on peut appliquer l'inégalité de Markov (Théorème 3.49), combinée à l'inégalité triangulaire, pour obtenir que

$$\mathbb{P}\left(\left|\frac{1}{n} \sum_{i=1}^n \hat{X}_i\right| > \frac{\varepsilon}{3}\right) \leq \frac{3}{\varepsilon} \mathbb{E}\left(\frac{1}{n} \sum_{i=1}^n |\hat{X}_i|\right) = \frac{3}{\varepsilon} \mathbb{E}(\hat{X}_1) \leq \delta,$$

où on a utilisé la linéarité de l'espérance et le fait que l'on ait choisi $M > 0$ précisément pour que $\mathbb{E}(\hat{X}_1) \leq \delta\varepsilon/3$. Ainsi, pour tout $n \geq n_0$, on a $\mathbb{P}(|\bar{X}_n - \mu| > \varepsilon) \leq 2\delta$: la valeur de δ étant arbitraire, on obtient finalement que $\lim_{n \rightarrow \infty} \mathbb{P}(|\bar{X}_n - \mu| > \varepsilon) = 0$. \square

Remarque 7.4. Dans le Théorème 7.1, la condition que les variables aléatoires X_i admettent une espérance finie est cruciale. Par exemple, si l'on considère une suite $(X_i)_{i \geq 1}$ de variables aléatoires i.i.d. de loi de Cauchy, de densité donnée par $f(x) = \frac{1}{\pi} \frac{1}{1+x^2}$, on a vu dans le chapitre précédent que les X_i n'admettent pas d'espérance. Maintenant, on a aussi vu que si X, Y sont deux variables aléatoires indépendantes de loi de Cauchy, alors $\frac{1}{2}(X + Y)$ suit une loi de Cauchy. On peut généraliser cela au cas de n variables aléatoires indépendantes : si $(X_i)_{i \geq 1}$ est une suite de variables aléatoires i.i.d. de loi de Cauchy, alors $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ suit une loi de Cauchy. Ainsi, quelle que soit $\mu \in \mathbb{R}$ (la valeur que l'on souhaiterait assigner à l'espérance de X_i), on a pour tout $\varepsilon > 0$

$$\begin{aligned} \mathbb{P}(|\bar{X}_n - \mu| > \varepsilon) &= \int_{-\infty}^{\mu-\varepsilon} \frac{dx}{\pi(1+x^2)} + \int_{\mu+\varepsilon}^{+\infty} \frac{dx}{\pi(1+x^2)} \\ &= 1 - \frac{1}{\pi} (\arctan(\mu + \varepsilon) - \arctan(\mu - \varepsilon)). \end{aligned}$$

Cette expression ne dépend pas de n , et en particulier elle ne tend pas vers 0 quand $n \rightarrow +\infty$! La suite $(X_i)_{i \geq 1}$ ne vérifie donc pas la loi faible des grands nombres, voir la Figure 7.1.

Dans la Proposition 7.10, on verra que si $(X_i)_{i \geq 1}$ est une suite de variables aléatoires i.i.d. positives dont l'espérance est égale à $+\infty$, alors $\bar{X}_n \rightarrow +\infty$, en un certain sens.

★ Une démonstration d'un théorème d'approximation de Weierstrass

À toute fonction $f : [0, 1] \rightarrow \mathbb{R}$ on associe la suite suivante de polynômes, appelés *polynômes de Bernstein de f* :

$$B_n^f(x) := \sum_{k=0}^n \binom{n}{k} f\left(\frac{k}{n}\right) x^k (1-x)^{n-k}.$$

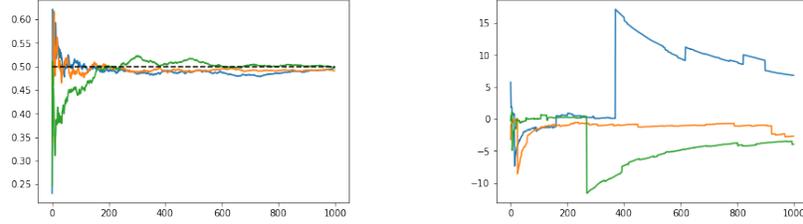


FIGURE 7.1 – Représentation graphique de trois suites $(\bar{X}_n)_{1 \leq n \leq 1000}$, où $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ avec $(X_i)_{i \in \mathbb{N}}$ une suite de variables i.i.d. À gauche, les X_i suivent la loi uniforme $\mathcal{U}(0, 1)$, et on observe que les trois suites “convergent” vers l’espérance de X_i , qui vaut $1/2$. À droite, les X_i suivent la loi de Cauchy, et les trois suites semblent “ne pas converger”.

Théorème 7.5 (Weierstrass-Bernstein). *Pour toute fonction $f : [0, 1] \rightarrow \mathbb{R}$ continue, la suite des polynômes de Bernstein de f converge uniformément vers f :*

$$\lim_{n \rightarrow +\infty} \sup_{x \in [0, 1]} |f(x) - B_n^f(x)| = 0.$$

Avec quelques transformations simples, cela démontre le théorème d’approximation suivant : « toute fonction continue définie sur un intervalle fermé et borné de \mathbb{R} peut être approchée uniformément par une suite de polynômes ».

Démonstration. L’idée de la démonstration est que grâce à la formule de transfert on peut écrire $B_n^f(x) = \mathbb{E}(f(\frac{X}{n}))$, où $X \sim \text{Bin}(n, x)$. Pour $x \in [0, 1]$, soient X_1, \dots, X_n des variables aléatoires i.i.d. de loi $X_i \sim \text{Bern}(x)$: la variable aléatoire $X_1 + \dots + X_n$ suit une loi $\text{Bin}(n, x)$: on a donc

$$B_n^f(x) = \mathbb{E}\left(f\left(\frac{1}{n} \sum_{i=1}^n X_i\right)\right) = \mathbb{E}(f(\bar{X}_n)).$$

D’après la loi des grands nombres, on a $\frac{1}{n} \sum_{i=1}^n X_i \approx x$ quand $n \rightarrow \infty$, ce qui suggère que $B_n^f(x) \approx f(x)$, car f est continue. Il reste à rendre cet argument rigoureux.

Fixons un $\varepsilon > 0$. Comme une fonction continue sur un intervalle fermé et borné de \mathbb{R} est bornée et uniformément continue : i) il existe un $M > 0$ tel que $|f(x)| \leq M$ pour tout $x \in [0, 1]$; ii) il existe un $\delta_\varepsilon > 0$ tel que si $|y - x| \leq \delta_\varepsilon$ alors $|f(y) - f(x)| \leq \frac{\varepsilon}{2}$. En écrivant $f(x) = \mathbb{E}(f(x))$, on a par linéarité de l’espérance

$$|f(x) - B_n^f(x)| = |\mathbb{E}(f(x) - f(\bar{X}_n))| \leq \mathbb{E}(|f(x) - f(\bar{X}_n)|).$$

On peut maintenant diviser l’espérance en deux parties,

$$\mathbb{E}(|f(x) - f(\bar{X}_n)|) = \mathbb{E}(|f(x) - f(\bar{X}_n)| \mathbb{1}_{\{|\bar{X}_n - x| \leq \delta_\varepsilon\}}) + \mathbb{E}(|f(x) - f(\bar{X}_n)| \mathbb{1}_{\{|\bar{X}_n - x| > \delta_\varepsilon\}}).$$

En utilisant l’uniforme continuité de f , le premier terme est majoré par $\frac{\varepsilon}{2}$ (on majore aussi l’indicatrice par 1). Pour le deuxième terme, on utilise la majoration $|f(x) - f(y)| \leq |f(x)| + |f(y)| \leq 2M$. On obtient donc

$$|f(x) - B_n^f(x)| \leq \frac{\varepsilon}{2} + 2M \mathbb{P}\left(|\bar{X}_n - x| > \delta_\varepsilon\right) \leq \frac{\varepsilon}{2} + 2M \frac{x(1-x)}{n\delta_\varepsilon^2},$$

où on a utilisé l’inégalité (7.3) pour le deuxième terme (rappelons que l’on a $\mu := \mathbb{E}(X_i) = x$ et $\sigma^2 := \text{Var}(X_i) = x(1-x)$). En prenant le supremum sur $x \in [0, 1]$ (atteint en $x = \frac{1}{2}$), on aboutit à

$$\sup_{x \in [0, 1]} |f(x) - B_n^f(x)| \leq \frac{\varepsilon}{2} + \frac{M}{2n\delta_\varepsilon^2},$$

qui est plus petit que ε pour tout $n \geq n_\varepsilon := M/(\delta_\varepsilon^2 \varepsilon)$. Cela conclut la démonstration. \square

7.2 Une application : la méthode de Monte-Carlo

La loi des grands nombres affirme que si X est une variable aléatoire qui admet une espérance finie et si $(X_i)_{i \geq 1}$ est une suite de variables aléatoires i.i.d. de même loi que X , alors pour n grand, la moyenne empirique $\bar{X}_n := \frac{1}{n} \sum_{i=1}^n X_i$ est “proche” de $\mathbb{E}(X)$. Ainsi, si on sait simuler informatiquement la variable aléatoire X , on va être capable de générer un très grand nombre de variables aléatoires X_i , calculer \bar{X}_n , ce qui nous donnera une valeur approchée de $\mathbb{E}(X)$.

Estimer numériquement une intégrale.

Supposons que l’on ait une fonction $f : [a, b] \rightarrow \mathbb{R}$ intégrable, dont on cherche à estimer l’intégrale $\int_a^b f(x) dx$. Il existe des méthodes exactes d’approximation (approximation de Riemann, méthode des trapèzes, etc...) : cependant, l’efficacité de ces méthodes dépend des propriétés de “régularité” de la fonction f en question (c’est-à-dire de des propriétés de continuité, dérivabilité, etc.).

Si $(X_i)_{i \geq 1}$ sont des variables aléatoires indépendantes, avec $X_i \sim \mathcal{U}(a, b)$, alors d’après la formule de transfert on a $\mathbb{E}(f(X_i)) = \frac{1}{b-a} \int_a^b f(x) dx$. En appliquant la loi des grands nombres aux variables aléatoires $f(X_1), f(X_2), \dots$ (qui sont i.i.d. et d’espérance finie) on obtient alors, pour tout $\varepsilon > 0$,

$$\lim_{n \rightarrow +\infty} \mathbb{P}\left(\left|\frac{1}{n} \sum_{i=1}^n f(X_i) - \frac{1}{b-a} \int_a^b f(x) dx\right| > \varepsilon\right) = 0, \quad \forall \varepsilon > 0,$$

de sorte que $\frac{1}{n} \sum_{i=1}^n f(X_i)$ est une bonne approximation de $\frac{1}{b-a} \int_a^b f(x) dx$. En remarquant que $\sigma^2 := \text{Var}(f(X_1)) \leq \mathbb{E}((f(X_1))^2)$, on peut quantifier cette approximation : grâce à l’inégalité (7.3) (de Bienaymé-Tchebychev), on a

$$\mathbb{P}\left(\left|\frac{1}{n} \sum_{i=1}^n f(X_i) - \frac{1}{b-a} \int_a^b f(x) dx\right| > \varepsilon\right) \leq \frac{\mathbb{E}(f(X_1)^2)}{\varepsilon^2 n}. \quad (7.6)$$

Ainsi, si on souhaite approcher $\frac{1}{b-a} \int_a^b f(x) dx$ avec un seuil d’erreur maximal de ε , et si δ est la probabilité avec laquelle on accepte de faire une erreur plus grande que ε , alors l’inégalité (7.6) indique qu’il est suffisant de prendre $n \geq \frac{\mathbb{E}(f(X_1)^2)}{\delta \varepsilon^2}$ (pour que $\frac{\mathbb{E}(f(X_1)^2)}{\varepsilon^2 n} \leq \delta$) — si la fonction f est bornée par une constante M , on peut majorer $\mathbb{E}(f(X_1)^2)$ par M^2 .

Exemple 7.6. Donnons ici un exemple numérique : calculons une valeur approchée de $\int_0^1 f(x) dx$, avec $f(x) = \sqrt{\cos(x)} \ln(1/x)$ (notons que f n’est pas bornée, mais elle est intégrable). Le programme suivant met en pratique la méthode de Monte-Carlo, et permet d’obtenir une valeur approchée de cette intégrale :

```
1 n=100000
2 S=0
3 for i in range(n):
4     x=random()
5     S=S+ sqrt(cos(x))*log(1/x)
6 print(S/n)
```

On souligne qu’il *ne faut pas écrire* `sqrt(cos(random()))*log(1/random())`, car dans ce cas, les deux valeurs de `random` sont indépendantes, et en particulier ne sont pas les mêmes.

En prenant $n = 10^6$, l'ordinateur renvoie la valeur (aléatoire) $\frac{1}{n} \sum_{i=1}^n f(X_i) \approx 0,9713$: d'après l'inégalité (7.6), en utilisant que $\mathbb{E}(f(X_1)^2) = \int_0^1 f(x)^2 dx \leq \int_0^1 \ln(x)^2 dx = 2$ (exercice), on a donc, en prenant $\varepsilon = 0,01$,

$$\mathbb{P}\left(\left|0,9713 - \int_0^1 f(x)dx\right| > 0,01\right) \leq \frac{2}{(0,01)^2 \cdot 10^6} = 0,02.$$

Dit autrement, cela signifie qu'il y a 98% de chances que l'intégrale $\int_0^1 f(x)dx$ soit proche de la valeur 0,9713 à 0,01 pres.

Estimer la probabilité d'un événement.

Supposons que l'on ne sache pas calculer explicitement la probabilité d'un événement A , relatif à une expérience aléatoire donnée. Si on sait simuler informatiquement l'expérience en question, on va être capable d'en générer un grand nombre : en comptant le nombre de fois où l'événement A est vérifié, on va ainsi pouvoir estimer la probabilité de l'événement A .

Plus précisément, soit $(X_i)_{i \geq 1}$ une suite de variables aléatoires i.i.d. de loi de Bernoulli de paramètre $\mathbb{P}(A)$ — on pense à X_i comme à l'indicatrice que l'événement A soit réalisé lors de la i -ème expérience. On a $\sigma^2 = \mathbb{P}(A)(1 - \mathbb{P}(A))$, et on peut alors *quantifier* l'approximation $\frac{1}{n} \sum_{i=1}^n X_i \approx \mathbb{P}(A)$: au vu de l'inégalité de Bienaymé-Tchebychev (7.3), on a

$$\mathbb{P}\left(\left|\frac{1}{n} \sum_{i=1}^n X_i - \mathbb{P}(A)\right| > \varepsilon\right) \leq \frac{\sigma^2}{\varepsilon n} \leq \frac{1}{4\varepsilon^2 n}, \quad (7.7)$$

où on a utilisé que $p(1-p) \leq \frac{1}{4}$ pour tout $p \in [0, 1]$.

Exemple 7.7. Mettons en pratique cette méthode, dans le cadre de l'Exemple 6.1 (en préambule du Chapitre 6) : on cherche à estimer la probabilité que, parmi n personnes, trois personnes aient le même anniversaire. Soit $(X_i)_{i \geq 1}$ une suite de variables aléatoires indépendantes, de loi uniforme sur $\{1, \dots, 365\}$, représentant les dates d'anniversaire de différentes personnes. On cherche à estimer la probabilité de l'événement

$$A_n := \{\exists\{i, j, k\} \subset \{1, \dots, n\}, X_i = X_j = X_k\}.$$

Pour appliquer la méthode de Monte-Carlo, il faut réaliser l'expérience un très grand nombre de fois, et compter le nombre de fois où l'événement est vérifié. La fonction suivante permet de créer une liste de n variables aléatoires uniformes sur $\{1, \dots, 365\}$:

```
1 def anniv(n):
2     return [int(365*random()+1) for i in range(n)]
```

La fonction suivante teste si une liste donnée contient trois fois la même entrée, c'est-à-dire renvoie 1 si la liste contient une entrée triple, et 0 sinon* :

```
1 def triplet(liste):
2     triple=0
3     for i in range(len(liste)):
4         a=liste[i]
5         if liste.count(a)>2:
6             triple=1
7     return triple
```

*. On peut optimiser cette fonction pour ne pas avoir à vérifier le reste de la liste si on a déjà trouvé un triplet, mais on s'en tient à cette version pour simplifier.

(`liste.count(a)` compte le nombre de fois où l'élément `a` apparaît dans la liste).

Il reste à appliquer un grand nombre k de fois l'expérience, et à renvoyer le nombre de fois où l'événement A_n est vérifié :

```

1 k=10000
2 S=0
3 for i in range(k):
4     S=S+triplet(anniv(n))
5 print(S/n)

```

Pour $n = 80$, on trouve une valeur approchée de 0,42 ; pour $n = 100$ de 0,64 ; pour $n = 120$ de 0,83 ; pour $n = 140$ de 0,94... Par exemple, dans un amphi de 140 étudiants, il y a environ 94% de chances que trois personnes soient nées le même jour !

7.3 Convergence en probabilité pour une suite de variables aléatoires

L'énoncé que l'on a donné de la loi des grands nombres donne en réalité une notion de convergence pour des variables aléatoires, appelée *convergence en probabilité*.

Définition 7.8. Soient $(X_n)_{n \in \mathbb{N}}$ et X des variables aléatoires réelles, toutes définies sur le même espace probabilisé. On dit que la suite $(X_n)_{n \in \mathbb{N}}$ converge en (\mathbb{P}) -probabilité vers X , et on note $(X_n) \xrightarrow{\mathbb{P}} X$ si pour tout $\varepsilon > 0$ on a

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

On dira que $(X_n)_{n \in \mathbb{N}}$ converge en (\mathbb{P}) -probabilité vers $+\infty$ (resp. $-\infty$) si pour tout $A > 0$ on a $\lim_{n \rightarrow \infty} \mathbb{P}(X_n < A) = 0$ (resp. $\lim_{n \rightarrow \infty} \mathbb{P}(X_n > -A) = 0$).

Notons que sur un même espace d'état Ω muni d'une tribu \mathcal{F} , on peut considérer plusieurs probabilités $\mathbb{P}_1, \mathbb{P}_2, \dots$ † : on écrira donc que la suite $(X_n)_{n \in \mathbb{N}}$ converge en \mathbb{P} -probabilité pour préciser que c'est relativement à la probabilité \mathbb{P} que la convergence a lieu (on omettra cette précision lorsqu'il n'y a pas d'ambiguïté possible).

On peut donc reformuler la loi des grands nombres de la façon suivante : « si $(X_i)_{i \in \mathbb{N}}$ est une suite de variables aléatoires i.i.d. d'espérance finie $\mu := \mathbb{E}(X_i)$, alors $(\frac{1}{n} \sum_{i=1}^n X_i)_{n \geq 1}$ converge en probabilité vers μ ».

Exemple 7.9. Soit $(X_i)_{i \geq 1}$ une suite de variables aléatoires i.i.d., de même loi $\mathcal{E}(1)$ (voir la Section 5.2.2). On pose $M_n := \max\{X_1, \dots, X_n\}$. Pour tout $A > 0$, on a

$$\mathbb{P}(M_n < A) = \mathbb{P}(X_i < A \text{ pour tout } 1 \leq i \leq n) = \mathbb{P}(X_1 < A)^n = (1 - e^{-A})^n,$$

où on a utilisé l'indépendance des X_i et la forme explicite de la fonction de répartition d'une variable aléatoire $\mathcal{E}(1)$. Comme pour tout $A > 0$ on a $1 - e^{-A} < 1$, on a donc $\lim_{n \rightarrow \infty} \mathbb{P}(M_n < A) = 0$: la suite $(M_n)_{n \geq 1}$ converge en probabilité vers $+\infty$.

†. Par exemple, sur $\Omega = \{0, 1\}$ muni de $\mathcal{F} = \mathcal{P}(\Omega)$, on peut considérer les probabilités \mathbb{P}_θ définies par $\mathbb{P}_\theta(\{1\}) = \theta$, $\mathbb{P}_\theta(\{0\}) = 1 - \theta$, pour n'importe quel $\theta \in [0, 1]$.

On peut rendre ce résultat plus précis : pour tout $\varepsilon > 0$, en reprenant les calculs précédents, on trouve facilement que

$$\begin{aligned}\mathbb{P}(M_n < (1 - \varepsilon) \ln n) &= (1 - e^{-(1-\varepsilon) \ln n})^n = (1 - n^{-1+\varepsilon})^n, \\ \mathbb{P}(M_n > (1 + \varepsilon) \ln n) &= 1 - (1 - e^{-(1+\varepsilon) \ln n})^n = 1 - (1 - n^{-1-\varepsilon})^n.\end{aligned}$$

On obtient donc (exercice) que $\mathbb{P}(M_n < (1 - \varepsilon) \ln n) \rightarrow 0$ et $\mathbb{P}(M_n > (1 + \varepsilon) \ln n) \rightarrow 0$ quand $n \rightarrow \infty$: on en conclut que pour tout $\varepsilon > 0$ on a $\lim_{n \rightarrow \infty} \mathbb{P}(|\frac{M_n}{\ln n} - 1| > \varepsilon) = 0$, c'est-à-dire que $(\frac{M_n}{\ln n})_{n \geq 1}$ converge en probabilité vers 1.

Donnons un complément à la loi des grand nombres (Théorème 7.1).

Proposition 7.10. *Soit $(X_i)_{i \geq 1}$ une suite de variables aléatoires i.i.d. positives, d'espérance égale à $+\infty$. Alors la suite $(\frac{1}{n} \sum_{i=1}^n X_i)_{n \geq 1}$ converge en probabilité vers $+\infty$.*

Démonstration. En comparant X_i à la partie entière $\lfloor X_i \rfloor$ de X_i , on peut se ramener au cas d'une variable aléatoire à valeurs dans \mathbb{N} (exercice). Pour une variable aléatoire X à valeurs dans \mathbb{N} , il est clair que si $\mathbb{E}(X) = \sum_{n=0}^{\infty} n \mathbb{P}(X = n) = +\infty$, alors pour tout $A > 0$, il existe $K = K_A \in \mathbb{N}$ tel que $\mathbb{E}(X \mathbb{1}_{\{X \leq K\}}) = \sum_{n=0}^K n \mathbb{P}(X = n) \geq 2A$.

Fixons $A > 0$, et posons $\tilde{X}_i = X_i \mathbb{1}_{\{X_i \leq K\}}$, où $K = K_A$ est tel que $\tilde{\mu} := \mathbb{E}(\tilde{X}_i) \geq 2A$; on a aussi $\tilde{\mu} < +\infty$, car $0 \leq \tilde{X}_i \leq K$. Ainsi, les variables aléatoires $(\tilde{X}_i)_{i \geq 1}$ sont i.i.d. et d'espérance finie, et on peut leur appliquer la loi des grands nombres (Théorème 7.1) : en prenant $\varepsilon = A$, on obtient

$$\lim_{n \rightarrow \infty} \mathbb{P}\left(\left|\frac{1}{n} \sum_{i=1}^n \tilde{X}_i - \tilde{\mu}\right| \leq A\right) = 1.$$

Comme $\tilde{\mu} \geq 2A$, en raisonnant par inclusion d'événements, on a que

$$\mathbb{P}\left(\left|\frac{1}{n} \sum_{i=1}^n \tilde{X}_i - \tilde{\mu}\right| \leq A\right) \leq \mathbb{P}\left(\frac{1}{n} \sum_{i=1}^n \tilde{X}_i \geq A\right) \leq \mathbb{P}\left(\frac{1}{n} \sum_{i=1}^n X_i \geq A\right),$$

où on a aussi utilisé que $X_i \geq \tilde{X}_i$ pour tout i . On en conclut donc que pour $A > 0$ arbitraire on a $\lim_{n \rightarrow \infty} \mathbb{P}(\bar{X}_n \geq A) = 1$, c'est-à-dire $\lim_{n \rightarrow \infty} \mathbb{P}(\bar{X}_n < A) = 0$, en passant au complémentaire. \square

Propriétés de la convergence en probabilité

Proposition 7.11. *Si $(X_n)_{n \in \mathbb{N}} \xrightarrow{\mathbb{P}} X$ et $(Y_n)_{n \in \mathbb{N}} \xrightarrow{\mathbb{P}} Y$ (où toutes les variables aléatoires sont définies sur le même espace probabilisé), alors pour toute fonction $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ continue on a $(f(X_n, Y_n))_{n \in \mathbb{N}} \xrightarrow{\mathbb{P}} f(X, Y)$.*

En particulier, cette proposition implique que :

- pour toute fonction $g : \mathbb{R} \rightarrow \mathbb{R}$ continue on a $(g(X_n))_{n \in \mathbb{N}} \xrightarrow{\mathbb{P}} g(X)$;
- on a $(X_n + Y_n)_{n \in \mathbb{N}} \xrightarrow{\mathbb{P}} X + Y$; ou encore $(X_n Y_n)_{n \in \mathbb{N}} \xrightarrow{\mathbb{P}} XY$.

Démonstration. Fixons $\varepsilon > 0$. Commençons par un cas plus simple : supposons que les variables aléatoires X et Y soient bornées, c'est-à-dire telles que $|X|, |Y| \leq A$ (presque sûrement). Dans ce cas, on va utiliser le fait que f est uniformément continue sur l'intervalle fermé borné $[-A-1, A+1]^2$ (théorème de Heine) : il existe $\delta = \delta_{\varepsilon, A} < 1$ tel que si on a $|x - x'| < \delta$ et $|y - y'| < \delta$, alors $|f(x, y) - f(x', y')| \leq \varepsilon$. Ainsi, si $|f(x, y) - f(x', y')| > \varepsilon$ avec $(x, y) \in [-A, A]$, alors nécessairement $|x - x'| > \delta$ ou $|y - y'| > \delta$: on a donc l'inégalité suivante

$$\mathbb{P}(|f(X_n, Y_n) - f(X, Y)| > \varepsilon) \leq \mathbb{P}(|X_n - X| > \delta) + \mathbb{P}(|Y_n - Y| > \delta),$$

et par définition de la convergence en probabilité de $(X_n)_{n \in \mathbb{N}}$ vers X et de $(Y_n)_{n \in \mathbb{N}}$ vers Y , on en conclut que $\lim_{n \rightarrow \infty} \mathbb{P}(|f(X_n, Y_n) - f(X, Y)| > \varepsilon) = 0$.

Traitons maintenant le cas général. En utilisant que $\mathbb{P}(|X| = +\infty) = \mathbb{P}(|Y| = +\infty) = 0$, on a que pour tout $\eta > 0$ on peut trouver un $A = A_\eta$ tel que $\mathbb{P}(|X| > A) \leq \eta/2$ et $\mathbb{P}(|Y| > A) \leq \eta/2$. Ainsi, en décomposant suivant que $|X|, |Y| \leq A$, $|X| \geq A$ ou $|Y| \geq A$ (l'union n'est pas disjointe, on utilise la sous-additivité), on peut écrire

$$\begin{aligned} \mathbb{P}(|f(X_n, Y_n) - f(X, Y)| > \varepsilon) &\leq \mathbb{P}(|f(X_n, Y_n) - f(X, Y)| > \varepsilon, |X| \leq A, |Y| \leq A) \\ &\quad + \mathbb{P}(|f(X_n, Y_n) - f(X, Y)| > \varepsilon, |X| > A) \\ &\quad + \mathbb{P}(|f(X_n, Y_n) - f(X, Y)| > \varepsilon, |Y| > A) \\ &\leq \mathbb{P}(|f(X_n, Y_n) - f(X, Y)| > \varepsilon, |X|, |Y| \leq A) + \eta/2 + \eta/2. \end{aligned}$$

À ce moment-là, on peut reprendre la démonstration précédente : en prenant $\delta = \delta_{\varepsilon, A}$ défini comme plus haut, on a

$$\mathbb{P}(|f(X_n, Y_n) - f(X, Y)| > \varepsilon) \leq \mathbb{P}(|X_n - X| > \delta) + \mathbb{P}(|Y_n - Y| > \delta) + \eta.$$

Comme $\lim_{n \rightarrow \infty} \mathbb{P}(|X_n - X| > \delta) = 0 = \lim_{n \rightarrow \infty} \mathbb{P}(|Y_n - Y| > \delta) = 0$, et comme η est arbitraire, on a $\lim_{n \rightarrow \infty} \mathbb{P}(|f(X_n, Y_n) - f(X, Y)| > \varepsilon) = 0$. \square

Un exemple servant de “contre-exemple(s)”. Donnons un exemple qui montre que même si (X_n) converge en probabilité vers X , on n'a pas forcément $\lim_{n \rightarrow \infty} \mathbb{E}(X_n) = \mathbb{E}(X)$.

Exemple 7.12. Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires dont la loi est donnée par

$$\mathbb{P}(X_n = n^2) = \frac{1}{n}, \quad \mathbb{P}(X_n = 0) = 1 - \frac{1}{n}.$$

Alors on a facilement que, pour tout $\varepsilon > 0$, $\mathbb{P}(|X_n| > \varepsilon) \leq \mathbb{P}(X_n = n^2) = \frac{1}{n} \rightarrow 0$: la suite $(X_n)_{n \geq 1}$ converge donc en probabilité vers $X = 0$. D'autre part, on a $\mathbb{E}(X_n) = 0 \cdot (1 - \frac{1}{n}) + n^2 \cdot \frac{1}{n} = n$: ainsi, $\mathbb{E}(X_n) = n$ pour tout $n \geq 1$, de sorte que $\lim_{n \rightarrow \infty} \mathbb{E}(X_n) = +\infty$, qui est différent de $\mathbb{E}(X) = 0$...

Notons que le même exemple montre que même si (X_n) converge en probabilité vers X , on n'a pas forcément que la suite des moyennes de Césaro $(\frac{1}{n} \sum_{i=1}^n X_i)_{n \geq 1}$ converge en probabilité vers X . C'est l'objet de l'exercice suivant.

Exercice 7.13. Reprendre la suite de variables aléatoires de l'Exemple 7.12, en supposant de plus que les X_n sont indépendantes.

1. Montrer que la probabilité de l'événement $A_n = \bigcup_{i=\lfloor \frac{n}{2} \rfloor + 1}^n \{X_i = i^2\}$ tend vers $1 - e^{-2}$ quand $n \rightarrow \infty$. (*Indication : utiliser l'approximation de Poisson (Proposition 3.21) pour une loi $\text{Bin}(n - \lfloor \frac{n}{2} \rfloor - 1, \frac{1}{n})$.)*)
2. En déduire que $\mathbb{P}(\frac{1}{n} \sum_{i=1}^n X_i \geq \frac{n}{4}) \geq \mathbb{P}(A_n) \geq \frac{1}{2}$ pour n suffisamment grand. Conclure.

Chapitre 8

Introduction à la statistique

8.1 Modèles statistiques paramétriques

Lorsque l'on étudie un phénomène aléatoire, on peut se donner une classe de modèles qui dépendent d'un paramètre θ . L'approche statistique consiste à supposer qu'il existe une valeur de θ qui fournit la bonne modélisation du phénomène : on cherche alors à estimer la valeur de ce paramètre à partir de données expérimentales. (L'approche probabiliste consiste quant à elle à étudier les propriétés des modèles en fonction de θ , mais ne s'attache pas à déterminer quel est "le bon θ ".)

Définition 8.1. On appelle modèle statistique (paramétrique) une famille de probabilités $(\mathbb{P}_\theta)_{\theta \in \Theta}$ sur un même espace (Ω, \mathcal{F}) , où Θ est l'ensemble des paramètres.

Exemple 8.2. Soit $\Omega = \mathbb{N}$, $\mathcal{F} = \mathcal{P}(\Omega)$, et $(\mathbb{P}_\theta)_{\theta \in]0,1[}$ une famille de probabilités indexée par θ , définies par $\mathbb{P}_\theta(\{n\}) = \theta(1 - \theta)^n$ pour $n \in \mathbb{N}$ — autrement dit, \mathbb{P}_θ est la loi d'une variable aléatoire Géom₀(θ).

On appelle *échantillon* une suite $(X_i)_{i \geq 1}$ de variables aléatoires définies sur (Ω, \mathcal{F}) qui, pour toute probabilité \mathbb{P}_θ , $\theta \in \Theta$, sont indépendantes et de même loi (la loi peut dépendre de θ). Un n -uplet (X_1, \dots, X_n) est alors appelé échantillon de taille n .

Les variables X_i représentent les observations successives relatives au phénomène aléatoire. Souvent, pour définir un modèle statistique, on identifiera $(\mathbb{P}_\theta)_{\theta \in \Theta}$ aux lois possibles d'un échantillon $(X_i)_{i \geq 1}$.

Exemple 8.3. Supposons que l'on fasse tomber une tartine de confiture, qui atterrit du mauvais côté (celui de la confiture) avec une probabilité $p \in [0, 1]$ inconnue. Un échantillon sera une suite $(X_i)_{i \geq 1}$ de variables aléatoires de Bernoulli de paramètre p indépendantes ($X_i = 1$ si la i -ème tartine de confiture tombe du mauvais côté, et $X_i = 0$ sinon). Le modèle statistique que l'on considère est donc l'ensemble des probabilités $(\mathbb{P}_p)_{p \in [0,1]}$ (sur $\Omega = \{0, 1\}^{\mathbb{N}^*}$, muni d'une tribu \mathcal{F} que l'on ne précisera pas ici), où \mathbb{P}_p est la loi d'une suite infinie de variables aléatoires de Bernoulli indépendantes, de paramètre p .

Définition 8.4. Soit $(X_i)_{i \geq 1}$ un échantillon (pour un modèle statistique fixé). On appelle statistique d'échantillon une suite $(Y_n)_{n \geq 1}$ de variables aléatoires de la forme

$$Y_n = h_n(X_1, X_2, \dots, X_n),$$

où $h_n : \mathbb{R}^n \rightarrow \mathbb{R}$ sont des fonctions continues sur $\mathbb{R}^n \setminus C$ où C est de mesure n -dimensionnelle nulle (C peut être vide).

Pour certaines statistiques d'échantillons, qui permettent l'estimation d'un paramètre donné du modèle, on parlera d'*estimateur*.

Exemple 8.5. Revenons sur l'Exemple 8.3. Soit $(\mathbb{P}_p)_{p \in [0,1]}$ un modèle statistique et un échantillon $(X_i)_{i \geq 1}$ tel que, sous \mathbb{P}_p , les $(X_i)_{i \geq 1}$ sont des variables aléatoires indépendantes de loi de Bernoulli de paramètre $p \in [0, 1]$. Alors la suite $(Y_n)_{n \geq 1}$ donnée par $Y_n := \bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ définit une statistique d'échantillon, qui est un estimateur du paramètre p .

Définition 8.6 (Estimateur sans biais). Soit $(\mathbb{P}_\theta)_{\theta \in \Theta}$ un modèle statistique, et soit $h : \Theta \rightarrow \mathbb{R}$ une fonction. Une statistique d'échantillon $(Y_n)_{n \geq 1}$ est appelée estimateur sans biais pour $h(\theta)$ si $\mathbb{E}_\theta(Y_n) = h(\theta)$, pour tout $\theta \in \Theta$ et pour tout $n \geq 1$.

Exemple 8.7. On se donne un échantillon $(X_i)_{i \geq 1}$ de variables aléatoires indépendantes pour un modèle statistique $(\mathbb{P}_\theta)_{\theta \in \Theta}$ donné. On suppose que pour tout $\theta \in \Theta$, X_i admet un moment d'ordre deux fini par rapport à \mathbb{P}_θ , et on note $\mu(\theta) := \mathbb{E}_\theta(X_i)$ et $\sigma^2(\theta) := \text{Var}_\theta(X_i)$. Alors les statistiques d'échantillon

$$\bar{X}_n := \frac{1}{n} \sum_{i=1}^n X_i, \quad S_n^2 := \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2, \quad (8.1)$$

appelés moyenne et variance empiriques, sont des estimateurs sans biais de $\mu(\theta)$ et $\sigma^2(\theta)$.

En effet, il est clair que, par linéarité de l'espérance, on a $\mathbb{E}_\theta(\bar{X}_n) = \mathbb{E}_\theta(X_i) = \mu(\theta)$. Pour ce qui est de la variance empirique (définie pour $n \geq 2$), on a

$$\begin{aligned} (n-1)\mathbb{E}_\theta(S_n^2) &= \sum_{i=1}^n \mathbb{E}_\theta \left((X_i - \bar{X}_n)^2 \right) = \sum_{i=1}^n \left(\mathbb{E}_\theta(X_i^2) + \mathbb{E}_\theta(\bar{X}_n^2) - 2\mathbb{E}_\theta(X_i \bar{X}_n) \right) \\ &= n\mathbb{E}_\theta(X_i^2) + n\mathbb{E}_\theta(\bar{X}_n^2) - 2\mathbb{E}_\theta \left(\left(\sum_{i=1}^n X_i \right) \bar{X}_n \right) = n\mathbb{E}_\theta(X_i^2) - n\mathbb{E}_\theta(\bar{X}_n^2) \\ &= n(\mu(\theta)^2 + \sigma(\theta)^2) - n \left(\mu(\theta)^2 + \frac{\sigma(\theta)^2}{n} \right) = (n-1)\sigma(\theta)^2, \end{aligned}$$

où on a utilisé que $\mathbb{E}_\theta(\bar{X}_n^2) = \mathbb{E}_\theta(\bar{X}_n)^2 + \text{Var}_\theta(\bar{X}_n) = \mu(\theta)^2 + \frac{1}{n}\sigma(\theta)^2$.

Définition 8.8 (Estimateur consistant). Soit $(\mathbb{P}_\theta)_{\theta \in \Theta}$ un modèle statistique, et soit $h : \Theta \rightarrow \mathbb{R}$ une fonction. Une statistique d'échantillon $(Y_n)_{n \geq 1}$ est appelée estimateur consistant pour $h(\theta)$ si pour tout $\theta \in \Theta$ on a $\lim_{n \rightarrow \infty} \mathbb{P}_\theta(|Y_n - h(\theta)| > \varepsilon) = 0$ pour tout $\varepsilon > 0$, c'est-à-dire si la suite $(Y_n)_{n \geq 1}$ converge en \mathbb{P}_θ -probabilité vers $h(\theta)$.

Exemple 8.9. Dans l'Exemple 8.5, \bar{X}_n est un estimateur consistant pour p : en effet, d'après la loi des grands nombres, la suite $(Y_n)_{n \geq 1}$ converge en \mathbb{P}_p -probabilité vers $\mathbb{E}_p(X_i) = p$.

8.2 Inégalités de concentration et construction d'intervalles de confiance

Une suite de n données observées x_1, x_2, \dots, x_n peut être interprétée comme une réalisation $X_1(\omega), \dots, X_n(\omega)$ des variables aléatoires X_1, \dots, X_n . Si $Y_n = h_n(X_1, \dots, X_n)$ est un estimateur pour $h(\theta)$, la valeur de $h_n(x_1, x_2, \dots, x_n)$ est donc vue comme une estimation de $h(\theta)$ basée sur les données x_1, x_2, \dots, x_n .

On introduit maintenant une notion de “marge d’erreur” dans l’estimation d’un paramètre.

Définition 8.10 (Intervalle de confiance). Soit $(\mathbb{P}_\theta)_{\theta \in \Theta}$ un modèle statistique, $(X_i)_{i \geq 1}$ un échantillon, et soit $h : \Theta \rightarrow \mathbb{R}$ une fonction. Pour deux statistiques d’échantillon $(U_n)_{n \geq 1}$ et $(V_n)_{n \geq 1}$ telles que $U_n \leq V_n$ pour tout n , on dit que l’intervalle aléatoire $[U_n, V_n]$ est un intervalle de confiance pour $h(\theta)$ de niveau de confiance $\gamma \in]0, 1[$ si pour tout $\theta \in \Theta$ on a

$$\mathbb{P}_\theta(h(\theta) \in [U_n, V_n]) \geq \gamma. \quad (8.2)$$

Exemple 8.11. Soit $(\mathbb{P}_\theta)_{\theta \in [0, 2]}$ un modèle statistique, et $(X_i)_{i \geq 1}$ un échantillon tel que, sous \mathbb{P}_θ , les X_i soient des variables aléatoires indépendantes, de loi $\mathcal{U}(0, \theta)$. On pose $Y_n = \max\{X_1, \dots, X_n\}$, que l’on voit comme un estimateur de θ : on va utiliser Y_n pour construire un intervalle de confiance pour θ . Notons que pour tout $\varepsilon > 0$, en remarquant que $\{Y_n < \theta - \varepsilon\} = \{X_i < \theta - \varepsilon \text{ pour tout } 1 \leq i \leq n\}$, on a par indépendance des X_i que

$$\mathbb{P}_\theta(Y_n < \theta - \varepsilon) = \mathbb{P}_\theta(X_i < \theta - \varepsilon)^n = \begin{cases} 0 & \text{si } \varepsilon > \theta, \\ (1 - \frac{\varepsilon}{\theta})^n & \text{si } \varepsilon \leq \theta, \end{cases}$$

où on a utilisé que $\mathbb{P}_\theta(X_i < t) = \frac{t}{\theta}$ si $t \in [0, \theta]$ et 0 si $t < 0$. En utilisant que $1 - x \leq e^{-x}$ pour tout $x \in \mathbb{R}$ (exercice), on en déduit que $\mathbb{P}_\theta(Y_n < \theta - \varepsilon) \leq \exp(-\frac{n\varepsilon}{\theta}) \leq \exp(-\frac{n\varepsilon}{2})$, car $\theta \leq 2$. Ainsi, pour tout $\varepsilon > 0$, on obtient

$$\mathbb{P}_\theta(\theta \in [Y_n, Y_n + \varepsilon]) = \mathbb{P}_\theta(Y_n \in [\theta - \varepsilon, \theta]) = 1 - \mathbb{P}_\theta(Y_n < \theta - \varepsilon) \geq 1 - e^{-\frac{n\varepsilon}{2}}.$$

Pour $\alpha \in]0, 1[$, en prenant $\varepsilon = \frac{2}{n} \ln(\frac{1}{\alpha})$ on a donc

$$\mathbb{P}_\theta(\theta \in [Y_n, Y_n + \frac{2}{n} \ln(\frac{1}{\alpha})]) \geq 1 - \alpha,$$

ce qui signifie que $[Y_n, Y_n + \frac{2}{n} \ln(\frac{1}{\alpha})]$ est un intervalle de confiance pour θ de niveau $1 - \alpha$.
Remarque : $(Y_n)_{n \geq 1}$ est un estimateur consistant de θ , mais pas sans biais (exercice).

8.2.1 Inégalités de concentration et intervalles de confiance

Une méthode générale pour construire un intervalle de confiance consiste à utiliser des inégalités dites “de concentration” : on cherche à savoir à quel point l’estimateur $Y_n = h_n(X_1, \dots, X_n)$ est “concentré” autour de $h(\theta)$. Dans cette optique, l’inégalité de Bienaymé-Tchebychev s’avère utile.

Exemple 8.12. Reprenons les Exemples 8.3 et 8.5. Soit $(\mathbb{P}_p)_{p \in [0,1]}$ un modèle statistique, et un échantillon $(X_i)_{i \geq 1}$ tel que, sous \mathbb{P}_p , les $(X_i)_{i \geq 1}$ sont des variables aléatoires indépendantes de loi de Bernoulli de paramètre $p \in [0, 1]$. On considère $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$, qui est un estimateur sans biais et consistant de p (rappelons que $\mathbb{E}_p(X_i) = p$). D'après l'inégalité de Bienaymé-Tchebychev, voir (7.3), on a, pour tout $\varepsilon > 0$

$$\mathbb{P}_p\left(|\bar{X}_n - p| > \varepsilon\right) \leq \frac{p(1-p)}{n\varepsilon^2} \leq \frac{1}{4n\varepsilon^2},$$

où on a utilisé que $\text{Var}_p(X_i) = p(1-p) \leq \frac{1}{4}$ pour tout $p \in [0, 1]$. On peut voir cette inégalité de deux manières : comme une borne sur la probabilité que \bar{X}_n soit à une distance supérieure à ε de p (interprétation probabiliste) ; comme une borne sur la probabilité que p soit à distance supérieure à ε de \bar{X}_n (interprétation statisticienne). On peut donc réécrire l'inégalité précédente sous la forme $\mathbb{P}_\theta(p \notin [\bar{X}_n - \varepsilon, \bar{X}_n + \varepsilon]) \leq \frac{1}{4n\varepsilon^2}$. Ainsi, si on se fixe un niveau de confiance $1 - \alpha$ avec $\alpha \in]0, 1[$, en prenant $\varepsilon = \frac{1}{2\sqrt{n\alpha}}$ (de sorte que $\frac{1}{4n\varepsilon^2} = \alpha$), on obtient finalement

$$\mathbb{P}_\theta\left(p \in \left[\bar{X}_n - \frac{1}{2\sqrt{n\alpha}}, \bar{X}_n + \frac{1}{2\sqrt{n\alpha}}\right]\right) \geq 1 - \alpha,$$

et $[\bar{X}_n - \frac{1}{2\sqrt{n\alpha}}, \bar{X}_n + \frac{1}{2\sqrt{n\alpha}}]$ est un intervalle de confiance pour p de niveau de confiance $1 - \alpha$.

On remarque ainsi que si l'on obtient une inégalité "meilleure" que celle de Bienaymé-Tchebychev, on obtiendra un intervalle de confiance "meilleur". On énonce ici une inégalité qui s'avère utile en statistique (et aussi dans d'autres contextes).

Proposition 8.13 (Inégalité de Hoeffding). *Soit $(X_i)_{i \geq 1}$ une suite de variables aléatoires indépendantes, centrées $\mathbb{E}(X_i) = 0$, et telles que pour tout $i \geq 1$ il existe c_i tel que $\mathbb{P}(|X_i| \leq c_i) = 1$. Alors, pour tout $t > 0$, en posant $r_n := \sum_{i=1}^n c_i^2$, on a*

$$\mathbb{P}\left(\sum_{i=1}^n X_i > t\right) \leq e^{-\frac{t^2}{2r_n}}, \quad \mathbb{P}\left(\left|\sum_{i=1}^n X_i\right| > t\right) \leq 2e^{-\frac{t^2}{2r_n}}.$$

Notons que s'il existe un $c > 0$ tel que $\mathbb{P}(|X_i| \leq c) = 1$ pour tout $i \geq 1$, les inégalités deviennent $\mathbb{P}\left(\sum_{i=1}^n X_i > t\right) \leq e^{-\frac{t^2}{2nc^2}}$ et $\mathbb{P}\left(\left|\sum_{i=1}^n X_i\right| > t\right) \leq 2e^{-\frac{t^2}{2nc^2}}$.

(difficile). L'idée est d'appliquer l'inégalité de Markov (Théorème 3.49) à la variable aléatoire $\exp(a \sum_{i=1}^n X_i)$, pour un $a > 0$ bien choisi. Pour $t > 0$ et pour tout $a > 0$, comme $x > t$ si et seulement si $e^{ax} > e^{at}$, on obtient

$$\mathbb{P}\left(\sum_{i=1}^n X_i > t\right) = \mathbb{P}\left(\exp\left(a \sum_{i=1}^n X_i\right) > e^{at}\right) \leq e^{-at} \mathbb{E}\left(\exp\left(a \sum_{i=1}^n X_i\right)\right),$$

où on a appliqué l'inégalité de Markov. En utilisant le fait que les X_i sont indépendants, on obtient

$$\mathbb{P}\left(\sum_{i=1}^n X_i > t\right) = e^{-at} \prod_{i=1}^n \mathbb{E}(e^{aX_i}). \quad (8.3)$$

Maintenant, par convexité de la fonction $x \mapsto e^{ax}$, pour tout $-c \leq x \leq c$ on a l'inégalité, $e^{ax} \leq \frac{1}{2}(e^{ac} + e^{-ac}) + \frac{x}{2c}(e^{ac} - e^{-ac})$. Ainsi, comme $|X_i| \leq c_i$ avec probabilité 1, on a

$$\mathbb{E}(e^{aX_i}) \leq \frac{1}{2}(e^{ac_i} + e^{-ac_i}) + \frac{\mathbb{E}(X_i)}{2c_i}(e^{ac_i} - e^{-ac_i}) = \frac{1}{2}(e^{ac_i} + e^{-ac_i}),$$

où on a utilisé la linéarité de l'espérance et le fait que $\mathbb{E}(X_i) = 0$. En remarquant que pour tout $x \in \mathbb{R}$ on a $\frac{1}{2}(e^x + e^{-x}) \leq e^{x^2/2}$ *, on obtient finalement que $\mathbb{E}(e^{aX_i}) \leq \exp(\frac{a^2 c_i^2}{2})$. En revenant à (8.3), on obtient donc que pour $t > 0$, tout $a > 0$

$$\mathbb{P}\left(\sum_{i=1}^n X_i > t\right) \leq e^{-at} e^{\frac{1}{2}a^2 r_n},$$

où on rappelle que $r_n = \sum_{i=1}^n c_i^2$. Cette inégalité étant valable pour tout $a > 0$, on peut "optimiser" la borne supérieure, en prenant $a = t/r_n$, de sorte que $-at + \frac{1}{2}a^2 r_n = \frac{t^2}{2r_n}$, ce qui donne la première inégalité.

Pour la deuxième inégalité, on remarque simplement que

$$\mathbb{P}\left(\left|\sum_{i=1}^n X_i\right| > t\right) = \mathbb{P}\left(\sum_{i=1}^n X_i > t\right) + \mathbb{P}\left(\sum_{i=1}^n X_i < -t\right),$$

et on peut utiliser l'inégalité déjà démontrée, à la fois pour le premier terme, et pour le deuxième terme (appliquée à $-X_i$). \square

Exemple 8.14. Revenons à l'Exemple 8.12 : sous \mathbb{P}_p , les variables aléatoires $(X_i)_{i \geq 1}$ sont indépendantes, de loi de Bernoulli de paramètre $p \in [0, 1]$. On peut alors appliquer l'inégalité de Hoeffding aux variables aléatoires centrées $\tilde{X}_i = X_i - p$, qui vérifient $\mathbb{P}_\theta(|\tilde{X}_i| \leq 1) = 1$: on obtient, pour tout $\varepsilon > 0$

$$\mathbb{P}_p\left(p \notin [\bar{X}_n - \varepsilon, \bar{X}_n + \varepsilon]\right) = \mathbb{P}_p\left(|\bar{X}_n - p| > \varepsilon\right) = \mathbb{P}_p\left(\left|\sum_{i=1}^n \tilde{X}_i\right| > \varepsilon n\right) \leq 2e^{-\frac{\varepsilon^2 n}{2}}. \quad (8.4)$$

Ainsi, pour $\alpha \in]0, 1[$, en posant $\varepsilon_{n,\alpha} := \frac{1}{\sqrt{n}} \sqrt{2 \ln(\frac{2}{\alpha})}$, on obtient que

$$\mathbb{P}_p\left(p \in [\bar{X}_n - \varepsilon_{n,\alpha}, \bar{X}_n + \varepsilon_{n,\alpha}]\right) \geq 1 - \alpha.$$

Cela montre que $[\bar{X}_n - \varepsilon_{n,\alpha}, \bar{X}_n + \varepsilon_{n,\alpha}]$ est un intervalle de confiance pour p de niveau de confiance $1 - \alpha$.

Notons que si α est petit (ce qui est souhaitable), $\sqrt{2 \ln(\frac{2}{\alpha})}$ est bien plus petit que $\frac{1}{2\sqrt{\alpha}}$: on a donc obtenu un intervalle de confiance "meilleur" que dans l'Exemple 8.12.

8.2.2 Intervalles de confiance pour la méthode de Monte-Carlo

Revenons sur les exemples d'utilisation de la méthode de Monte-Carlo de la Section 7.2.

*. Ce qu'on peut voir en comparant les développements en série entière des fonctions $\cosh(x) = \frac{1}{2}(e^x + e^{-x})$ et $e^{x^2/2}$.

Estimation numérique d'une intégrale.

Supposons que l'on cherche à estimer l'intégrale $\int_a^b f(x)dx$ d'une fonction f bornée par une constante M . Si $(X_i)_{i \geq 1}$ est une suite de variables aléatoires i.i.d. de loi $\mathcal{U}(a, b)$, alors la suite $(Y_n)_{n \geq 1}$ définie par $Y_n := \frac{b-a}{n} \sum_{i=1}^n f(X_i)$ est un estimateur (sans biais et consistant) de $\mathbb{E}(f(X_1)) = \int_a^b f(x)dx$.

Alors, en appliquant l'inégalité de Hoeffding (Proposition 8.13) aux variables aléatoires centrées $\tilde{X}_i := f(X_i) - \mathbb{E}(f(X_i))$, qui vérifient que $|\tilde{X}_i| \leq 2M$, on obtient que pour tout $\varepsilon > 0$

$$\mathbb{P}\left(\left|Y_n - \int_a^b f(x)dx\right| > \varepsilon\right) = \mathbb{P}\left(\left|\sum_{i=1}^n \tilde{X}_i\right| > n\varepsilon(b-a)\right) \leq 2 \exp\left(-\frac{n\varepsilon^2(b-a)^2}{8M}\right)$$

où on rappelle que $\mathbb{E}(f(X_i)) = \frac{1}{b-a} \int_a^b f(x)dx$. En posant $\varepsilon = \varepsilon_{n,\alpha} := \frac{1}{\sqrt{n}} \frac{\sqrt{8M}}{b-a} \sqrt{\ln\left(\frac{2}{\alpha}\right)}$, on en déduit que $[Y_n - \varepsilon_{n,\alpha}, Y_n + \varepsilon_{n,\alpha}]$ est un intervalle de confiance pour $\int_a^b f(x)dx$ de niveau de confiance $1 - \alpha$.

Exercice 8.15. En utilisant les données de l'Exemple 7.6, donner un intervalle de confiance à 95% pour $\int_0^1 \sqrt{\cos x} \ln\left(\frac{1}{x}\right)dx$.

Estimation numérique d'une probabilité.

Supposons que l'on cherche à estimer la probabilité d'un événement A . Si $(X_i)_{i \geq 1}$ est une suite de variables aléatoires i.i.d. de loi de Bernoulli de paramètre $\mathbb{P}(A)$, alors la suite $(\bar{X}_n)_{n \geq 1}$ définie par $\bar{X}_n := \frac{1}{n} \sum_{i=1}^n X_i$ est un estimateur (sans biais et consistant) de $\mathbb{P}(A)$. En reprenant les calculs de l'Exemple 8.14, en posant $\varepsilon_{n,\alpha} := \frac{1}{\sqrt{n}} \sqrt{2 \ln\left(\frac{2}{\alpha}\right)}$, on obtient que $[\bar{X}_n - \varepsilon_{n,\alpha}, \bar{X}_n + \varepsilon_{n,\alpha}]$ est un intervalle de confiance pour $\mathbb{P}(A)$ de niveau de confiance $1 - \alpha$.

Exercice 8.16. Soient $(X_i)_{i \geq 1}, (Y_i)_{i \geq 1}$ deux suites de variables aléatoires indépendantes, toutes de loi uniforme sur $[-1, 1]$.

1. Montrer que $\mathbb{P}(X_i^2 + Y_i^2 < 1) = \frac{\pi}{4}$.
2. On pose $A_i = \{X_i^2 + Y_i^2 < 1\}$, et $W_n := \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{A_i}$. Donner un intervalle de confiance de niveau $1 - \alpha$ pour $\mathbb{P}(A_i)$, en utilisant l'estimateur W_n ; en déduire un intervalle de confiance pour π .
3. On considère le programme suivant :

```
1 n=10**8
2 W=0
3 for i in range(n):
4     if (2*random()-1)**2+(2*random()-1)**2<1:
5         W=W+1
6 print(4*W/n)
```

qui renvoie (après un certain temps de calcul) la valeur (aléatoire) 3,14163804. Donner un intervalle de confiance à 95% pour π .

Variables aléatoires importantes : lois, espérance, variance

Variables aléatoires discrètes				
	Densité discrète $p_X(k) = \mathbb{P}(X = k)$	Espérance $\mathbb{E}(X)$	Variance $\text{Var}(X)$	Fonction génératrice des moments $M_X(t) = \mathbb{E}(e^{tX})$
Bernoulli , $\text{Bern}(p)$ ($p \in [0, 1]$)	p si $k = 1$, $1 - p$ si $k = 0$.	p	$p(1 - p)$	$pe^t + (1 - p)$
Binomiale , $\text{Bin}(n, p)$ ($n \in \mathbb{N}_*$, $p \in [0, 1]$)	$\binom{n}{k} p^k (1 - p)^{n-k}$, $k \in \{0, 1, \dots, n\}$	np	$np(1 - p)$	$(pe^t + (1 - p))^n$
Poisson , $\text{Poi}(\lambda)$ ($\lambda \in]0, \infty[$)	$e^{-\lambda} \frac{\lambda^k}{k!}$, $k \in \mathbb{N}$	λ	λ	$e^{\lambda(e^t - 1)}$
Géométrique , $\text{Géom}(p)$ ($p \in]0, 1]$)	$p(1 - p)^{k-1}$, $k \in \mathbb{N}_*$	$\frac{1}{p}$	$\frac{1 - p}{p^2}$	$\frac{pe^t}{1 - (1 - p)e^t}$ si $t < \ln(\frac{1}{1 - p})$ (+ ∞ sinon)

Variables aléatoires à densité				
	Densité $f_X(x)$	Espérance $\mathbb{E}(X)$	Variance $\text{Var}(X)$	Fonction génératrice des moments $M_X(t) = \mathbb{E}(e^{tX})$
Uniforme continue , $\mathcal{U}(a, b)$ ($a, b \in \mathbb{R}$, $a < b$)	$\frac{1}{b - a} \mathbb{1}_{]a, b[}(x)$	$\frac{a + b}{2}$	$\frac{(b - a)^2}{12}$	$\frac{e^{tb} - e^{ta}}{t(b - a)}$
Exponentielle , $\mathcal{E}(\lambda)$ ($\lambda \in]0, \infty[$)	$\lambda e^{-\lambda x} \mathbb{1}_{]0, \infty[}(x)$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	$\frac{\lambda}{\lambda - t}$ si $t < \lambda$ (+ ∞ sinon)
Normale , $\mathcal{N}(\mu, \sigma^2)$ ($\mu \in \mathbb{R}$, $\sigma \in]0, \infty[$)	$\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x - \mu)^2}{2\sigma^2}}$	μ	σ^2	$e^{\mu t + \frac{\sigma^2}{2} t^2}$

Index

- ★ La ruine du joueur, 36
 - ★ Méthode des moments : le problème du collectionneur, 49
 - ★ Marche aléatoire simple sur \mathbb{Z} , 52
 - ★ Marche aléatoire simple sur \mathbb{Z}^d , 37
 - ★ Probabilité que deux entiers soient premiers entre eux, 19
 - ★ Série harmonique de signe aléatoire, 85
 - ★ Un théorème d'approximation de Weierstrass, 90
- Bayes (formule de), 22
- Bienaymé-Tchebychev (inégalité de), 48
- Cauchy-Schwarz (inégalité de), 48
- Continuité (par le haut et le bas) des probabilités, 15
- Convergence en probabilité, 94
- Espérance, 39, 57, 63
- Estimateur
 - consistant, 98
 - sans biais, 98
- Fonction
 - de répartition, 57
 - génératrice (d'une variable aléatoire à valeurs dans \mathbb{N}), 50
 - génératrice des moments, 59
- Formule
 - d'inclusion/exclusion, 14
 - de Bayes, 22
 - de transfert, 40, 63, 73
 - des probabilités totales, 21
- Indépendance
 - d'événements, 22
 - de variables aléatoires, 30, 72
 - par paquets, 32
- Inégalité
 - de Bienaymé-Tchebychev, 48
 - de Cauchy-Schwarz, 48
 - de Hoeffding, 100
 - de Markov, 47
- Intervalle de confiance, 99
- Loi
 - binomiale, 34, 79
 - de Bernoulli, 34, 79
 - de Cauchy (★), 76, 83, 90
 - de Poisson, 34, 80
 - exponentielle, 67, 82
 - géométrique, 35, 79
 - Gamma (★), 75
 - normale (gaussienne), 68, 83
 - uniforme continue, 66, 82
 - uniforme discrète, 33, 78
- Loi des grands nombres, 88
- Loi jointe et lois marginales, 29, 71
- Markov (inégalité de), 47
- Méthode
 - d'inversion, 81
 - de Monte-Carlo, 92, 101
 - de rejet, 83
- Moments d'une variable aléatoire, 43, 64
- Paradoxe des anniversaires, 18, 93
- Paradoxe du singe de Borel, 24
- Presque sûr (p.s.), 16, 28
- Probabilité conditionnelle, 20
- Simulation de lois à densité, 81
- Simulation de lois discrètes, 78
- Sous-additivité, 15
- Statistique d'échantillon, 98
- Variable aléatoire à densité, 61
- Variable aléatoire discrète, 25
- Variance et Covariance, 44
- Vecteurs aléatoires à densité, 70