

UNIVERSITÉ PARIS-DAUPHINE PSL

# Introduction aux Probabilités

LICENCE 2, Mathématiques - Informatique

BÉATRICE DE TILIÈRE

*Inspiré de*

BÉATRICE DE TILIÈRE ET FRÉDÉRIQUE PETIT : “Probabilités et statistique pour le CAPES”

FRANÇOIS SIMENHAUS, JUSTIN SALEZ : “Introduction aux probabilités”

JULIEN POISAT, JOSÉ TRASHORRAS : “Probabilités 2”

*et certaines parties inspirées du livre de*

SYLVIE MÉLÉARD : “Aléatoire”

---



# TABLE DES MATIÈRES

|          |                                                                       |           |
|----------|-----------------------------------------------------------------------|-----------|
| <b>1</b> | <b>MODÉLISATION DES PHÉNOMÈNES ALÉATOIRES</b>                         | <b>3</b>  |
| 1.1      | Introduction . . . . .                                                | 3         |
| 1.2      | L'espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$ . . . . .    | 4         |
| 1.2.1    | Espace des états/univers . . . . .                                    | 4         |
| 1.2.2    | Évènements - tribu . . . . .                                          | 5         |
| 1.2.3    | Probabilité . . . . .                                                 | 7         |
| 1.3      | Le cas dénombrable et le cas fini . . . . .                           | 11        |
| 1.3.1    | Univers finis . . . . .                                               | 12        |
| 1.3.2    | Univers dénombrables . . . . .                                        | 16        |
| 1.4      | Le cas général . . . . .                                              | 17        |
| <b>2</b> | <b>CONDITIONNEMENT ET INDÉPENDANCE</b>                                | <b>19</b> |
| 2.1      | Probabilité conditionnelle . . . . .                                  | 19        |
| 2.1.1    | Définition . . . . .                                                  | 19        |
| 2.1.2    | Formule des probabilités totales et formule de Bayes . . . . .        | 21        |
| 2.2      | Indépendance des évènements . . . . .                                 | 22        |
| 2.3      | Lemme de Borel-Cantelli . . . . .                                     | 24        |
| 2.4      | Expériences aléatoires indépendantes et probabilité produit . . . . . | 25        |
| <b>3</b> | <b>VARIABLES ALÉATOIRES</b>                                           | <b>27</b> |
| 3.1      | Définitions . . . . .                                                 | 27        |
| 3.2      | Variables aléatoires discrètes . . . . .                              | 28        |
| 3.2.1    | Définition . . . . .                                                  | 28        |
| 3.2.2    | Variables aléatoires discrètes classiques . . . . .                   | 29        |
| 3.3      | Variables aléatoires réelles . . . . .                                | 32        |
| 3.3.1    | Définition . . . . .                                                  | 32        |
| 3.3.2    | Fonction de répartition . . . . .                                     | 33        |
| 3.3.3    | Variables aléatoires discrètes réelles . . . . .                      | 34        |

---

|          |                                                                           |           |
|----------|---------------------------------------------------------------------------|-----------|
| 3.3.4    | Variables aléatoires à densité . . . . .                                  | 34        |
| <b>4</b> | <b>ESPÉRANCE, VARIANCE &amp; INÉGALITÉS</b>                               | <b>39</b> |
| 4.1      | Espérance . . . . .                                                       | 39        |
| 4.1.1    | Cas d'un espace des états fini ou dénombrable . . . . .                   | 39        |
| 4.1.2    | Cas des variables aléatoires réelles . . . . .                            | 40        |
| 4.1.3    | Propriétés fondamentales . . . . .                                        | 41        |
| 4.1.4    | Cas des variables aléatoires discrètes et à densité . . . . .             | 43        |
| 4.2      | Variance, covariance, moments d'ordres supérieurs . . . . .               | 44        |
| 4.2.1    | Variance, covariance, corrélation, écart-type . . . . .                   | 44        |
| 4.3      | Inégalité de Markov, Bienaymé-Tchebychev et autres . . . . .              | 47        |
| <b>5</b> | <b>VECTEURS ALÉATOIRES DISCRETS, INDÉPENDANCE, LOI DES GRANDS NOMBRES</b> | <b>51</b> |
| 5.1      | Vecteurs aléatoires discrets . . . . .                                    | 51        |
| 5.1.1    | Définitions et premier résultat . . . . .                                 | 51        |
| 5.1.2    | Lois et espérance conditionnelles . . . . .                               | 53        |
| 5.2      | Indépendance . . . . .                                                    | 56        |
| 5.3      | Conséquences . . . . .                                                    | 58        |
| 5.4      | Loi faible des grands nombres . . . . .                                   | 60        |
| <b>6</b> | <b>INTRODUCTION À LA MARCHE ALÉATOIRE</b>                                 | <b>63</b> |
| 6.1      | Définition . . . . .                                                      | 63        |
| 6.2      | Retours en zéro . . . . .                                                 | 64        |
| 6.3      | Ruine de la joueuse . . . . .                                             | 66        |

---

# CHAPITRE 1

## MODÉLISATION DES PHÉNOMÈNES ALÉATOIRES

### 1.1 INTRODUCTION

Une *expérience (ou phénomène) aléatoire* consiste en une expérience pour laquelle toutes les issues possibles sont connues, mais où interviennent de nombreux facteurs, dont nous ne connaissons ou maîtrisons qu'une petite partie. Dans ce cas, l'issue n'est pas prévisible avec certitude. La *théorie des probabilités* consiste en l'étude de ces expériences aléatoires.

Citons quelques exemples : le résultat d'un jeu de hasard (pile ou face, jet de dé, roulette etc.) ; durée de vie d'un atome radioactif, d'un individu, d'une ampoule ; les instants de passage d'un bus à un arrêt donné ; la promenade d'un ivrogne dans la rue ; la trajectoire d'une poussière à la surface de l'eau etc.

Les applications de la théorie des probabilités sont nombreuses : base de la statistique, outil puissant en finance, dans les assurances, théorie des jeux. Elle permet également de modéliser de nombreux phénomènes complexes en physique, biologie, médecine, sciences humaines, climatologie. Elle s'est aussi révélée utile dans de nombreux domaines des mathématiques pures. Mais surtout, elle a acquis une place importante au sein des mathématiques en tant que discipline à part entière, de part son intérêt intrinsèque.

Historiquement, les jeux des hasards sont présents en Égypte, en Grèce et à Rome dès l'Antiquité. Il est cependant intéressant de constater qu'un traitement systématique n'est apparu qu'au XVI<sup>e</sup> siècle dans le livre *Liber de Ludo Alea* de Gerolamo Cardano (1501-1576). La véritable étincelle se trouve dans la correspondance entre Blaise Pascal (1623-1662) et Pierre de Fermat (~1605-1665), au sujet de problèmes posés par le chevalier de Méré. Encouragé par Pascal, Christian Huygens (1629-1695) publie *De ratiocinis in ludo aleae* (raisonnements sur les jeux de dés) en 1657. Ce livre est le premier ouvrage important sur les probabilités. Il y définit la notion d'espérance et y développe plusieurs problèmes de partages de gains lors de jeux ou de tirages dans des urnes. Deux ouvrages fondateurs sont également à noter : *Ars Conjectandi* de Jacques Bernoulli (1654-1705) qui définit la notion de variable aléatoire et donne la première version de la loi des grands nombres, et *The Doctrine of Chance* d'Abraham de Moivre (1668-1754) qui généralise l'usage de la combinatoire. On mentionnera également Pierre-Simon de Laplace (1749-1827), Leonhard Euler (1707-1783) et Johann Carl Friedrich Gauss (1777-1855).

La théorie des probabilités classique ne prend réellement son essor qu'avec les notions de mesure et d'ensembles mesurables qu'Émile Borel (1871-1956) introduit en 1897. Cette notion de mesure

est complétée par Henri Léon Lebesgue (1875-1941) et sa théorie de l'intégration. La première version moderne du théorème central limite est donnée par Alexandre Liapounov en 1901 et la première preuve du théorème moderne est due à Paul Lévy en 1910. Il faudra attendre 1933 pour que la théorie des probabilités sorte d'un ensemble de méthodes et d'exemples divers et devienne une véritable théorie, axiomatisée par Andreï Nikolaïevitch Kolmogorov (1903-1987).

## 1.2 L'ESPACE PROBABILISÉ $(\Omega, \mathcal{F}, \mathbb{P})$

Le but de la théorie des probabilités est de fournir un modèle mathématique pour décrire les expériences aléatoires. Sous sa forme moderne, la formulation de cette théorie contient trois ingrédients : l'*espace des états* ou *univers*, les *événements*, et la *loi de probabilité* ou simplement la *probabilité*. Dans toute la suite, nous considérons une expérience aléatoire que nous cherchons à modéliser.

### 1.2.1 ESPACE DES ÉTATS/UNIVERS

**Définition 1.1.** L'*espace des états*, appelé aussi *univers*, noté  $\Omega$ , est l'ensemble des résultats possibles de l'expérience. Un résultat possible de l'expérience est noté  $\omega$ , donc  $\omega \in \Omega$ .

*Remarque.* A noter que le choix d'espace des états n'est pas forcément unique. Il peut également être de natures très différentes : fini ou dénombrable, infini non dénombrable, espace de fonctions, etc.

EXEMPLE 1.2.

1. Lancer d'une pièce de monnaie :  $\Omega = \{P, F\}$
2. Deux lancers successifs d'une même pièce de monnaie :  $\Omega = \{(P, P), (P, F), (F, P), (F, F)\} = \{P, F\}^2$ .
3. Lancer d'un dé :  $\Omega = \{1, 2, 3, 4, 5, 6\}$ .
4. Deux lancers successifs d'un même dé, et on s'intéresse à la somme des nombres obtenus. Il y a deux choix raisonnables :

$$\begin{aligned}\Omega_1 &= \{(i, j) : i, j \in \{1, \dots, 6\}\} = \{1, \dots, 6\}^2, \\ \Omega_2 &= \{2, 3, 4, \dots, 12\}.\end{aligned}$$

5. Lancer d'un même dé indéfiniment :  $\Omega = \{(u_n)_{n \geq 1} : \forall n \in \mathbb{N}^*, u_n \in \{1, \dots, 6\}\} = \{1, \dots, 6\}^{\mathbb{N}^*}$ .
6. Durée de vie d'un individu :  $\Omega = \{x \in \mathbb{R}^+ : 0 \leq x \leq 120\}$ .
7. Promenade d'un ivrogne dans une rue (un pas en avant, un pas en arrière) :

$$\Omega = \{(u_n)_{n \geq 1} : \forall n \in \mathbb{N}^*, u_n \in \{-1, 1\}\} = \{-1, 1\}^{\mathbb{N}^*}.$$

8. Trajectoire d'une poussière à la surface de l'eau pendant un intervalle de temps  $[0, T]$  :  $\Omega = \mathcal{C}([0, T]; \mathbb{R}^2)$ .

### 1.2.2 ÉVÈNEMENTS - TRIBU

**Définition heuristique.** Un *évènement* est un sous-ensemble de  $\Omega$  dont on peut dire, une fois l'issue de l'expérience connue, s'il est réalisé ou non.

Un singleton, c'est-à-dire un évènement réduit à un seul élément de  $\Omega$ , est appelé un *évènement élémentaire*, sinon on parle d'*évènement composite*. On note un évènement par une lettre majuscule  $A, B, C, \dots$  et l'ensemble de tous les évènements de  $\Omega$  par  $\mathcal{F}$ .

L'ensemble des évènements  $\mathcal{F}$  associés à une expérience aléatoire est donc un sous-ensemble des parties de  $\Omega$ ,  $\mathcal{F} \subset \mathcal{P}(\Omega)$ . Il semblerait naturel de prendre  $\mathcal{F} = \mathcal{P}(\Omega)$ , mais il y a alors des exemples où il est impossible d'associer à chaque évènement une probabilité de façon cohérente. Dans ces cas-là, il est donc nécessaire de se restreindre à un sous-ensemble strict de  $\mathcal{P}(\Omega)$ . L'ensemble des évènements considéré doit satisfaire à quelques propriétés naturelles, il doit former une tribu, dont voici la définition.

**Définition 1.3.** Un ensemble  $\mathcal{F}$  de parties de  $\Omega$  est une *tribu*, ou  *$\sigma$ -algèbre*, s'il satisfait aux conditions suivantes :

1.  $\Omega \in \mathcal{F}$ .
2. Si  $A \in \mathcal{F}$ , alors  $A^c \in \mathcal{F}$ .
3. Si  $(A_n)_{n \geq 1}$  est une suite d'éléments de  $\mathcal{F}$ , alors  $\bigcup_{n \geq 1} A_n \in \mathcal{F}$ .

Ainsi, l'*ensemble des évènements* associé à une expérience est la tribu  $\mathcal{F}$  choisie sur  $\Omega$ .

EXEMPLE 1.4. (Suite de l'Exemple 1.2 ; on reprend la numérotation déjà utilisée.) Décrivons les évènements suivants comme des sous-ensembles de l'espace des états  $\Omega$ .

2. "Le premier jet donne pile" est le sous-ensemble  $\{PP, PF\}$  de  $\Omega$ .
4. "La somme des résultats obtenus est égale à 4" est le sous-ensemble  $\{(1, 3), (2, 2), (3, 1)\}$  de  $\Omega_1$  et le sous-ensemble  $\{4\}$  de  $\Omega_2$ . Remarquons que si l'on choisit  $\Omega_2$ , "le premier dé a une valeur inférieure ou égale à 3" n'est pas un évènement de  $\Omega$ .
5. "Le premier 1 est obtenu au  $N$ -ième lancer" est le sous-ensemble suivant de  $\Omega$

$$\{(u_n)_{n \geq 1} \in \Omega : u_1 \geq 2, \dots, u_{N-1} \geq 2, u_N = 1\}.$$

6. "L'individu atteint au moins 50 ans" est le sous-ensemble suivant de  $\Omega$

$$\{x \in \mathbb{R}^+ : 50 \leq x \leq 120\}.$$

7. "L'ivrogne avance au  $N$ -ième pas" est le sous-ensemble suivant de  $\Omega$

$$\{(u_n)_{n \geq 1} \in \Omega : u_N = 1\}.$$

*Remarque.* Les évènements, qui sont par définition des sous-ensembles de l'univers  $\Omega$ , sont en général décrits à l'aide de phrases dans un premier temps. En effet, on commence par se poser une question liée à une expérience aléatoire, puis on introduit un modèle probabiliste pour y répondre. Par exemple, on cherche la probabilité que la somme de deux dés lancés au hasard soit égale à 4 ; l'évènement considéré est alors "la somme des dés est égale à 4".

Une fois fixé le choix de l'univers, un évènement correspond à un *unique* sous-ensemble de ce dernier. Comme il n'y a pas forcément unicité du modèle et qu'alors les évènements peuvent



s'écrire en termes de sous-ensembles sous des formes différentes, la phrase qui décrit un évènement permet de se comprendre, quel que soit le modèle choisi. Remarquons aussi que, étant donné un sous-ensemble d'un univers, il est souvent possible de le décrire par différentes phrases, qui représentent toutes le même évènement. Par exemple l'évènement  $\{PF, FF\}$  de l'Exemple 1.4 numéro 2 peut se traduire par "le premier jet donne pile" ou "le premier jet ne donne pas face".

Puisque les évènements sont des ensembles, on peut effectuer les opérations habituelles, avec la correspondance suivante entre les terminologies ensemblistes et probabilistes.

| Notation               | Terminologie ensembliste   | Terminologie probabiliste                                   |
|------------------------|----------------------------|-------------------------------------------------------------|
| $\Omega$               | ensemble entier            | espace des états, évènement certain                         |
| $\emptyset$            | ensemble vide              | évènement impossible                                        |
| $\omega$               | élément de $\Omega$        | évènement élémentaire/singleton                             |
| $A$                    | sous-ensemble de $\Omega$  | évènement                                                   |
| $\omega \in A$         | $\omega$ appartient à $A$  | $A$ est réalisé si $\omega$ est le résultat de l'expérience |
| $A \subset B$          | $A$ est inclu dans $B$     | si $A$ est réalisé alors $B$ aussi                          |
| $A \cup B$             | réunion de $A$ et $B$      | l'évènement " $A$ ou $B$ " (ou non exclusif!)               |
| $A \cap B$             | intersection de $A$ et $B$ | l'évènement " $A$ et $B$ "                                  |
| $A^c$                  | complémentaire de $A$      | l'évènement contraire de $A$                                |
| $A \cap B = \emptyset$ | $A$ et $B$ sont disjoints  | $A$ et $B$ sont incompatibles                               |

EXEMPLE 1.5. Deux lancers successifs d'une même pièce de monnaie,  $\Omega = \{P, F\}^2$ . Soient  $A = \{PP\}$ ,  $B = \{(P, F)\}$ ,  $C = \{(F, P), (F, F)\}$ . Alors,

- $A \cup B = \{(P, P), (P, F)\} = C^c$ , est l'évènement "le premier jet donne pile",
- $A \cap B = \emptyset$ , est l'évènement impossible,  $A$  et  $B$  sont incompatibles.

**Propriété 1.6.** Les opérations sur les évènements satisfont aux règles suivantes. Pour tout évènement  $A, B, C$ , on a

- *commutativité* :  $A \cup B = B \cup A$  ;
- *associativité* :  $(A \cup B) \cup C = A \cup (B \cup C)$  ;
- *distributivité* :  $(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$  ;
- *lois de De Morgan* :  $(A \cup B)^c = A^c \cap B^c$ , et  $(A \cap B)^c = A^c \cup B^c$ .

Maintenant que nous avons bien étudié les évènements, revenons sur la notion de tribu et commençons par deux exemples.

EXEMPLE.

1.  $\{\emptyset, \Omega\}$  est une tribu et c'est la plus petite (au sens de l'inclusion). Il s'agit de la *tribu grossière* ou *triviale*.
2.  $\mathcal{P}(\Omega)$  est une tribu et c'est la plus grande. Elle est choisie lorsque  $\Omega$  est fini ou dénombrable. Cette tribu est typiquement trop grande lorsque  $\Omega$  est infini non dénombrable.

**Définition 1.7.**

- Si  $\mathcal{C} \subset \mathcal{P}(\Omega)$ , on appelle *tribu engendrée par  $\mathcal{C}$* , notée  $\sigma(\mathcal{C})$ , la plus petite tribu contenant  $\mathcal{C}$ .
- On appelle *tribu borélienne* sur  $\mathbb{R}$ , notée  $\mathcal{B}(\mathbb{R})$ , la tribu engendrée par l'ensemble des intervalles ouverts, *i.e.*, de la forme  $]a, b[$  avec  $a, b \in \mathbb{R}$ ,  $a < b$ .

*Remarque 1.8.*

- La tribu engendrée par  $\mathcal{C}$  existe toujours. En effet, on peut définir

$$\sigma(\mathcal{C}) := \bigcap_{\{\mathcal{F} : \mathcal{F} \text{ tribu contenant } \mathcal{C}\}} \mathcal{F}.$$

Cette intersection est non-vide car  $\mathcal{P}(\Omega)$  est une telle tribu ; c'est bien une tribu, car une intersection de tribus est une tribu (voir exercices) ; elle contient  $\mathcal{C}$  par définition, et c'est la plus petite car on prend l'intersection.

- En exercices, vous montrerez que la tribu borélienne est aussi la tribu engendrée par l'ensemble des intervalles de la forme  $] - \infty, a]$ , avec  $a \in \mathbb{R}$ .

EXEMPLE 1.9.

1. Soit  $A \in \mathcal{P}(\Omega)$ ,  $A \neq \emptyset$ ,  $A \neq \Omega$ , alors la tribu engendrée par  $A$  est  $\sigma(A) = \{\emptyset, A, A^c, \Omega\}$ .
2. Soit  $\Omega = \{1, 2, 3\}$  et  $\mathcal{C} = \{\{1\}, \{2\}\}$ . Alors  $\sigma(\mathcal{C}) = \mathcal{P}(\Omega)$ . En effet, la tribu engendrée par  $\mathcal{C}$  est stable par réunion et complémentation, elle doit donc contenir  $\{1, 2\}$  et  $\{3\}$  ; en particulier elle contient  $\{1\}, \{2\}, \{3\}$ . Par réunion et complémentation à nouveau,  $\sigma(\mathcal{C})$  doit contenir  $\mathcal{P}(\Omega)$  et est donc égale.

### 1.2.3 PROBABILITÉ

Nous souhaitons maintenant associer à chacun des événements une *probabilité*, qui mesure la vraisemblance que l'on accorde a priori à l'évènement avant la réalisation de l'expérience. C'est une des données du modèle, que l'on peut comprendre intuitivement de différentes manières, en voici deux.

*Approche utilisant les symétries.* On considère un dé non-pipé. Il est alors naturel de supposer que chacune des issues possibles ait la même probabilité égale à  $1/6$ . Il faut cependant être prudent avec cette approche. En effet, supposons que nous souhaitions déterminer la probabilité du sexe d'un nouveau né. Il n'y a aucune raison de penser qu'il y a plus de chances d'avoir un garçon ou une fille, de sorte qu'il est naturel d'associer une probabilité  $1/2$  à chacun des événements élémentaires. Cependant, les statistiques montrent que la proportion de garçons nouvellement né est de 51,17% (INED, France métropolitaine).

*Approche fréquentiste.* On suppose qu'une expérience d'univers  $\Omega$  est exécutée plusieurs fois sous les mêmes conditions. Pour chaque évènement  $A$  de  $\Omega$ , on définit  $n_N(A)$  comme le nombre de fois où l'évènement  $A$  survient lors des  $N$  premières répétitions de l'expérience. Alors la *probabilité de l'évènement*  $A$ , notée  $\mathbb{P}(A)$ , est définie comme la limite, dans un sens à préciser, du quotient  $n_N(A)/N$ .

Cela veut dire que  $\mathbb{P}(A)$  est définie comme la limite du pourcentage du nombre de fois où  $A$  survient par rapport au nombre total des répétitions. C'est donc la fréquence limite de  $A$ . Bien que cette définition soit intuitivement commode, elle présente un sérieux inconvénient. En effet, il faut justifier de l'existence de la limite, ce qui est difficile a priori.

Il est plus raisonnable d'admettre que les probabilités satisfont à un ensemble d'axiomes simples et intuitivement acceptables, pour ensuite démontrer qu'une telle fréquence limite existe dans un certain sens (voir plus loin la *loi des grands nombres*).

**Définition 1.10.** Étant donné un espace d'états  $\Omega$  et une tribu d'évènements  $\mathcal{F}$ , une *mesure de probabilité*, ou simplement *probabilité*,  $\mathbb{P}$  sur  $(\Omega, \mathcal{F})$ , est une application de  $\mathcal{F}$  dans  $[0, 1]$ , possédant les propriétés suivantes.

1. L'évènement certain est de probabilité 1 :  $\mathbb{P}(\Omega) = 1$ .
2. *Axiome de  $\sigma$ -additivité* : pour toute famille dénombrable  $(A_n)_{n \geq 1}$  d'évènements de  $\mathcal{F}$ , deux-à-deux disjoints, on a

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

*Remarque.* Le point 2. de la définition implique en particulier que la série de terme général  $\mathbb{P}(A_n)_{n \geq 1}$  est convergente (et de somme  $\mathbb{P}(\bigcup_{n \geq 1} A_n)$ ).

**Définition 1.11.**

- Le triplet  $(\Omega, \mathcal{F}, \mathbb{P})$  est alors appelé un *espace probabilisé*.
- Un évènement  $A \in \mathcal{F}$  tel que  $\mathbb{P}(A) = 0$  est appelé *négligeable*.
- Un évènement  $A \in \mathcal{F}$  tel que  $\mathbb{P}(A) = 1$  est appelé *presque-sûr*, il se produit  *$\mathbb{P}$ -presque sûrement*.

On a les conséquences suivantes de la définition d'une probabilité.

**Corollaire 1.12.** Soit  $(\Omega, \mathcal{F}, \mathbb{P})$  un espace probabilisé, soient deux évènements  $A, B \in \mathcal{F}$ , et  $(A_i)_{i=1}^n$  une famille finie d'évènements de  $\mathcal{F}$ .

1.  $\mathbb{P}(\emptyset) = 0$ .
2. Additivité. Si les  $(A_i)_{i=1}^n$  sont deux-à-deux disjoints, alors

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

3. Complémentaire.  $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$ .
4. Différence.  $\mathbb{P}(B \setminus A) = \mathbb{P}(B) - \mathbb{P}(A \cap B)$ .
5. Monotonie. Si  $A \subset B$ , alors :  $\mathbb{P}(A) \leq \mathbb{P}(B)$ .
6. Formule du crible.  $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$ . Plus généralement, on a la formule de Poincaré :

$$\mathbb{P}\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n (-1)^{i-1} \sum_{\substack{J \subset \{1, \dots, n\} \\ \text{card}(J) = i}} \mathbb{P}\left(\bigcap_{k \in J} A_k\right).$$

*Démonstration.*

1. On applique la  $\sigma$ -additivité avec la famille  $(A_n)_{n \geq 1}$  où, pour tout  $n \geq 1$ ,  $A_n = \emptyset$ . Ces évènements sont bien deux-à-deux disjoints et on a donc,  $\mathbb{P}(\emptyset) = \sum_{n \geq 1} \mathbb{P}(\emptyset)$  où le terme dans la somme est indépendant de  $n$ . On déduit que  $\mathbb{P}(\emptyset) = 0$ .
2. On considère la famille  $(A_i)_{i \geq 1}$  où, pour tout  $i \geq n+1$ ,  $A_i := \emptyset$ . Ces évènements sont deux-à-deux disjoints, ainsi,

$$\begin{aligned} \mathbb{P}(\cup_{i=1}^n A_i) &= \mathbb{P}(\cup_{i \geq 1} A_i) = \sum_{i \geq 1} \mathbb{P}(A_i), \text{ par la } \sigma\text{-additivité} \\ &= \sum_{i=1}^n \mathbb{P}(A_i) + \sum_{i \geq n+1} \mathbb{P}(\emptyset) = \sum_{i=1}^n \mathbb{P}(A_i), \text{ en utilisant le Point 1.} \end{aligned}$$

3. On utilise le Point 2. avec  $n = 2$ ,  $A_1 = A$ ,  $A_2 = A^c$ , qui satisfont  $A_1 \cap A_2 = \emptyset$  et  $A_1 \cup A_2 = \Omega$ . En utilisant de plus le Point 1. de la définition, on obtient :  $1 = \mathbb{P}(\Omega) = \mathbb{P}(A) + \mathbb{P}(A^c)$ .
4. On écrit  $B = A_1 \cup A_2$  avec  $A_1 = B \cap A$  et  $A_2 = B \cap A^c := B \setminus A$ . Les évènements  $A_1$  et  $A_2$  sont deux-à-deux disjoints car  $A_1 \cap A_2 = B \cap (A \cap A^c) = \emptyset$ . Ainsi,

$$\mathbb{P}(B) = \mathbb{P}(B \cap A) + \mathbb{P}(B \cap A^c), \text{ d'après le Point 2.}$$

$$\text{d'où } \mathbb{P}(B \setminus A) = \mathbb{P}(B \cap A^c) = \mathbb{P}(B) - \mathbb{P}(B \cap A).$$

5. D'après le point précédent,

$$\begin{aligned} \mathbb{P}(B) &= \mathbb{P}(A \cap B) + \mathbb{P}(B \setminus A) \\ &\geq \mathbb{P}(A \cap B), \text{ car une probabilité est positive} \\ &= \mathbb{P}(A), \text{ car } A \subset B. \end{aligned}$$

6. Démontrons le cas de deux évènements; le cas général sera démontré en exercices. On écrit  $A \cup B = (A \setminus B) \cup (B \setminus A) \cup (A \cap B)$ . Ces évènements sont deux-à-deux disjoints, ainsi

$$\begin{aligned} \mathbb{P}(A \cup B) &= \mathbb{P}(A \setminus B) + \mathbb{P}(B \setminus A) + \mathbb{P}(A \cap B), \text{ d'après le Point 2.} \\ &= \mathbb{P}(A) - \mathbb{P}(A \cap B) + \mathbb{P}(B) - \mathbb{P}(A \cap B) + \mathbb{P}(A \cap B), \text{ d'après le Point 4.} \\ &= \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B). \end{aligned} \quad \square$$

*Remarque* (Cas fini). Lorsque le cardinal de  $\Omega$  est fini, on munit  $\Omega$  de la tribu  $\mathcal{P}(\Omega)$ . Notons que toute famille dénombrable d'évènements disjoints deux-à-deux contient un nombre fini d'évènements non vides, et la définition de probabilité est équivalente à la définition suivante.

Une *probabilité*  $\mathbb{P}$  sur  $(\Omega, \mathcal{P}(\Omega))$  est une application de  $\mathcal{P}(\Omega)$  dans  $[0, 1]$ , possédant les propriétés suivantes.

1. L'évènement certain est de probabilité 1 :  $\mathbb{P}(\Omega) = 1$ .
2. *Axiome d'additivité* : pour tous  $A \in \mathcal{P}(\Omega)$ ,  $B \in \mathcal{P}(\Omega)$ , tels que  $A$  et  $B$  sont disjoints, on a

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

Voici quelques autres conséquences fréquemment utilisées.

**Proposition 1.13.** *Soit  $(\Omega, \mathcal{F}, \mathbb{P})$  un espace probabilisé.*

1. *Soit  $(A_n)_{n \geq 1}$  une suite d'évènements, alors :*

$$\mathbb{P}(\cup_{n \geq 1} A_n) \leq \sum_{n \geq 1} \mathbb{P}(A_n).$$

2. *Soit  $(A_n)_{n \geq 1}$  une suite croissante d'évènements de  $\mathcal{F}$ , c'est-à-dire, pour tout  $n \geq 1$ ,  $A_n \subset A_{n+1}$ . Soit  $A = \bigcup_{n \geq 1} A_n$ , alors :*

$$\mathbb{P}(A) = \lim_{n \rightarrow +\infty} \mathbb{P}(A_n).$$

3. *Soit  $(B_n)_{n \geq 1}$  une suite décroissante d'évènements de  $\mathcal{F}$ , c'est-à-dire, pour tout  $n \geq 1$ ,  $B_n \supset B_{n+1}$ . Soit  $B = \bigcap_{n \geq 1} B_n$ , alors :*

$$\mathbb{P}(B) = \lim_{n \rightarrow +\infty} \mathbb{P}(B_n).$$

*Remarque* (Points 2. et 3.). Les suites réelles  $(\mathbb{P}(A_n))_{n \geq 1}$  et  $(\mathbb{P}(B_n))_{n \geq 1}$  sont respectivement croissante et décroissante, bornées, donc convergentes, justifiant ainsi l'existence des limites. Par ailleurs, la suite  $(\cup_{n=1}^N A_n)_{N \geq 1}$ , égale à  $(A_N)_{N \geq 1}$ , est une suite croissante au sens de l'inclusion, et l'ensemble  $\cup_{n \geq 1} A_n$  est habituellement noté  $A = \lim_{n \rightarrow \infty} A_n$ , de sorte que le résultat de la proposition précédente peut encore s'écrire :

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

De même, la suite  $(\cap_{n=1}^N B_n)_{N \geq 1}$ , égale à  $(B_N)_{N \geq 1}$ , est une suite décroissante au sens de l'inclusion, et l'ensemble  $\cap_{n \geq 1} B_n$  est habituellement noté  $B = \lim_{n \rightarrow +\infty} B_n$ , de sorte que le résultat de la proposition précédente s'écrit :

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

*Démonstration.* Posons  $C_1 = A_1$  et pour tout  $n \geq 2$ ,  $C_n = A_n \setminus (\cup_{k=1}^{n-1} A_k)$ . Alors, les événements  $(C_n)_{n \geq 1}$  sont deux-à-deux disjoints et  $\cup_{n \geq 1} C_n = \cup_{n \geq 1} A_n$ . Montrons le Point 1. On a

$$\begin{aligned} \mathbb{P}(\cup_{n \geq 1} A_n) &= \mathbb{P}(\cup_{n \geq 1} C_n) \\ &= \sum_{n \geq 1} \mathbb{P}(C_n), \text{ d'après la } \sigma\text{-additivité} \\ &\leq \sum_{n \geq 1} \mathbb{P}(A_n), \text{ car, pour tout } n \geq 1 \text{ } C_n \subset A_n. \end{aligned}$$

Montrons maintenant le Point 2. Dans ce cas, la suite d'événements  $(A_n)_{n \geq 1}$  est de plus croissante ce qui implique que, pour tout  $n \geq 2$ ,  $\cup_{k=1}^{n-1} A_k = A_{n-1}$  et  $C_n = A_n \setminus A_{n-1}$ .

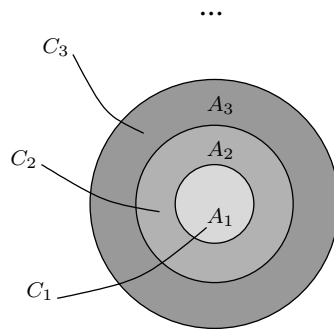


FIGURE 1.1 – Une vision schématique d'une suite croissante d'événements  $(A_n)_{n \geq 1}$ . Les événements  $(C_n)_{n \geq 1}$  sont les anneaux successifs.

Ainsi, en reprenant le début du calcul précédent, on a

$$\begin{aligned}
 \mathbb{P}(A) &= \mathbb{P}(\cup_{n \geq 1} A_n) = \sum_{n \geq 1} \mathbb{P}(C_n) = \lim_{N \rightarrow \infty} \left( \mathbb{P}(A_1) + \sum_{n=2}^N \mathbb{P}(A_n \setminus A_{n-1}) \right) \\
 &= \lim_{N \rightarrow +\infty} \left( \mathbb{P}(A_1) + \sum_{n=2}^N [\mathbb{P}(A_n) - \mathbb{P}(A_n \cap A_{n-1})] \right), \text{ d'après le Point 4. du Corollaire 1.12} \\
 &= \lim_{N \rightarrow +\infty} \left( \mathbb{P}(A_1) + \sum_{n=2}^N [\mathbb{P}(A_n) - \mathbb{P}(A_{n-1})] \right), \text{ car } A_{n-1} \subset A_n \\
 &= \lim_{N \rightarrow +\infty} \mathbb{P}(A_N).
 \end{aligned}$$

Pour le cas 3. d'une suite décroissante  $(B_n)_{n \geq 1}$  d'évènements, on écrit

$$\begin{aligned}
 \mathbb{P}(B) &= \mathbb{P}(\cap_{n \geq 1} B_n) = 1 - \mathbb{P}((\cap_{n \geq 1} B_n)^c), \text{ d'après le Point 3. du Corollaire 1.12} \\
 &= 1 - \mathbb{P}(\cup_{n \geq 1} B_n^c), \text{ complémentaire d'une intersection} \\
 &= 1 - \lim_{n \rightarrow \infty} \mathbb{P}(B_n^c), \text{ d'après le point précédent car la suite } (B_n^c)_{n \geq 1} \text{ est croissante} \\
 &= \lim_{n \rightarrow \infty} \mathbb{P}(B_n), \text{ d'après le Point 3. du Corollaire 1.12.} \quad \square
 \end{aligned}$$

### 1.3 LE CAS DÉNOMBRABLE ET LE CAS FINI

Dans cette section, nous allons voir que dans le cas où l'univers  $\Omega$  est dénombrable ou fini, on peut définir une probabilité de manière simple. Plus précisément, nous allons montrer qu'une probabilité est caractérisée par sa valeur sur les singletons (évènements élémentaires). Dans ces deux cas, on choisit toujours  $\mathcal{F} = \mathcal{P}(\Omega)$ . Nous avons alors la caractérisation suivante.

**Proposition 1.14.** *Soit une suite  $(p_\omega)_{\omega \in \Omega}$  de nombres réels. Il lui correspond une unique probabilité  $\mathbb{P}$  sur  $(\Omega, \mathcal{P}(\Omega))$  telle que pour tout  $\omega \in \Omega$ ,  $\mathbb{P}(\{\omega\}) = p_\omega$ , si et seulement si*

$$\forall \omega \in \Omega \quad 0 \leq p_\omega \leq 1, \quad \text{et} \quad \sum_{\omega \in \Omega} p_\omega = 1.$$

Nous avons alors,

$$\forall A \in \mathcal{P}(\Omega), \quad \mathbb{P}(A) = \sum_{\omega \in A} p_\omega. \quad (1.1)$$

*Remarque.* Lorsque l'univers est fini, il n'y a pas de commentaire particulier à faire. Lorsque l'univers est dénombrable, on peut numéroter les singletons  $\Omega = \{\omega_1, \omega_2, \dots\}$ . On pose ensuite, pour tout  $A \in \mathcal{F}$ ,

$$\mathbb{P}(A) = \sum_{\omega \in A} p_\omega = \sum_{n=1}^{+\infty} p_{\omega_n} \mathbb{I}_A(\omega_n).$$

Cette définition ne dépend pas de l'ordre choisi pour les éléments de  $\Omega$  car les termes intervenant dans la série sont tous positifs.

*Démonstration.* Soit  $\mathbb{P}$  une probabilité sur  $(\Omega, \mathcal{P}(\Omega))$  et pour tout  $\omega \in \Omega$ , posons  $p_\omega := \mathbb{P}(\{\omega\})$ . Alors, par définition d'une probabilité, nous avons bien que  $0 \leq p_\omega \leq 1$ . De plus, tout évènement  $A$  de  $\mathcal{P}(\Omega)$  est réunion dénombrable, resp. finie, d'éléments de  $\Omega$ , si bien que par définition d'une probabilité (cas dénombrable) et par le Corollaire 1.12 (cas fini), on a :

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

On a aussi que  $\Omega$  est réunion dénombrable, resp. finie, de ses singletons. Ainsi, en appliquant l'identité ci-dessus à  $\Omega$  et en utilisant de plus la définition d'une probabilité, nous avons

$$1 = \mathbb{P}(\Omega) = \mathbb{P}(\cup_{\omega \in \Omega} \{\omega\}) = \sum_{\omega \in \Omega} \mathbb{P}(\{\omega\}) = \sum_{\omega \in \Omega} p_\omega.$$

Nous avons donc démontré un sens. Réciproquement, étant donné une suite de nombre  $(p_\omega)_{\omega \in \Omega}$ , pour tout  $\omega \in \Omega$ , posons  $\mathbb{P}(\{\omega\}) = p_\omega$  et étendons  $\mathbb{P}$  à  $\mathcal{P}(\Omega)$  en utilisant la Condition (1.1). Alors, on vérifie aisément que  $\mathbb{P}$  ainsi définie vérifie les conditions de la Définition 1.10.  $\square$

### 1.3.1 UNIVERS FINIS

Dans toute cette section, on suppose que  $\Omega$  est de cardinal fini. Un cas important de probabilité sur  $\Omega$  est celui de la probabilité uniforme, pour laquelle chaque singleton de  $\Omega$  a la même chance de réalisation. La définition suivante découle de la Proposition 1.14.

**Définition 1.15.** Une probabilité  $\mathbb{P}$  sur  $(\Omega, \mathcal{P}(\Omega))$  est dite *uniforme*, si  $\mathbb{P}(\{\omega\})$  ne dépend pas de  $\omega \in \Omega$ . On dit alors que l'on est en situation d'*équiprobabilité* et on a,

$$\forall \omega \in \Omega, \quad \mathbb{P}(\{\omega\}) = \frac{1}{\text{card}(\Omega)}, \quad \forall A \in \mathcal{P}(\Omega), \quad \mathbb{P}(A) = \frac{\text{card}(A)}{\text{card}(\Omega)}.$$

EXEMPLE 1.16. Reprenons la question 4 des Exemples 1.2 et 1.4 et calculons la probabilité de l'évènement  $A$  "la somme des dés est égale à 4".

Supposons que l'on ait choisi l'espace  $\Omega_1$ . Alors, on est en situation d'équiprobabilité et la probabilité  $\mathbb{P}_1$  sur  $\Omega_1$  est uniforme, de sorte que pour tout  $(i, j) \in \{1, \dots, 6\}^2$  :

$$\mathbb{P}_1[\{(i, j)\}] = \frac{1}{36}.$$

$$\text{Ainsi, } \mathbb{P}_1(A) = \mathbb{P}_1[\{(1, 3), (2, 2), (3, 1)\}] = \frac{\text{card}(A)}{\text{card}(\Omega_1)} = \frac{3}{36}.$$

Supposons maintenant que l'on ait choisi l'espace  $\Omega_2$ . Alors, on n'est plus en situation d'équiprobabilité. Au vu des conditions de l'expérience, on définit  $\mathbb{P}_2$  ainsi :

$$\begin{aligned} \mathbb{P}_2(\{2\}) = \mathbb{P}_2(\{12\}) &= \frac{1}{36}, & \mathbb{P}_2(\{3\}) = \mathbb{P}_2(\{11\}) &= \frac{1}{18}, & \mathbb{P}_2(\{4\}) = \mathbb{P}_2(\{10\}) &= \frac{1}{12}, \\ \mathbb{P}_2(\{5\}) = \mathbb{P}_2(\{9\}) &= \frac{1}{9}, & \mathbb{P}_2(\{6\}) = \mathbb{P}_2(\{8\}) &= \frac{5}{36}, & \mathbb{P}_2(\{7\}) &= \frac{1}{6}. \end{aligned}$$

$$\text{Ainsi, } \mathbb{P}_2(A) = \mathbb{P}_2(\{4\}) = \frac{1}{12} \text{ mais } \frac{\text{card}(A)}{\text{card}(\Omega_2)} = \frac{1}{11}, \text{ d'où } \mathbb{P}_2(A) \neq \frac{\text{card}(A)}{\text{card}(\Omega_2)}.$$

Cet exemple montre qu'il est très important de spécifier le choix d'univers et de probabilité. Bien que les résultats finaux ne changent pas, les raisonnements pour y arriver sont différents et doivent être explicités.

Dans le cas de la probabilité uniforme, les calculs de probabilités se ramènent à des questions de **dénombrements**, nous sommes dans le cadre du calcul combinatoire. Un cas très classique de dénombrement est celui des modèles d'urne. Il y a différentes manières d'effectuer les tirages dans une urne et il est important de bien identifier dans quel cas on se trouve afin de ne pas commettre d'erreur d'énumération.

**Définition 1.17.** Une *population* de taille  $N$  est un ensemble  $\mathcal{S} = \{s_1, \dots, s_N\}$  de  $N$  éléments. Les éléments de  $\mathcal{S}$  sont les *individus* de la population  $\mathcal{S}$ . La *taille*  $N$  est le nombre d'éléments de  $\mathcal{S}$ .

**TIRAGES ORDONNÉS.** Un *échantillon* de taille  $r$  est un  $r$ -uplet  $(s_{i_1}, \dots, s_{i_r})$  d'éléments de  $\mathcal{S}$ . Deux procédures sont possibles.

- *Le tirage avec remise.* Dans ce cas, chaque élément de l'ensemble peut être choisi à plusieurs reprises. On parle alors d'*échantillon de taille  $r$  avec répétitions*. Soit  $\Omega_1$  l'ensemble de ces échantillons, alors  $\Omega_1 = \{s_1, \dots, s_N\}^r$ , et :

$$\text{card}(\Omega_1) = N^r.$$

EXEMPLE 1.18. On jette un dé 5 fois de suite. Alors  $\Omega_1$  est l'ensemble des tirages possibles et on a  $\text{card}(\Omega_1) = 6^5$ .

- *Le tirage sans remise.* Dans ce cas, chaque élément de l'ensemble peut être choisi au plus une fois. On parle alors d'*échantillon de taille  $r$  sans répétition*, ou d'*arrangement des éléments de  $\mathcal{S}$  pris  $r$  à  $r$* . Naturellement, on impose les conditions supplémentaires  $r \leq N$ , et  $\forall j \neq k, i_j \neq i_k$ . Soit  $\Omega_2$  l'ensemble de ces échantillons, on a alors :

$$\text{card}(\Omega_2) = N(N-1) \dots (N-r+1) = \frac{N!}{(N-r)!} := A_N^r.$$

EXEMPLE 1.19. On considère une population de taille  $N$  et un échantillon aléatoire de taille  $r$  avec répétition. On choisit alors comme univers  $\Omega_1$  que l'on munit de la probabilité uniforme, notée  $\mathbb{P}$ . On s'intéresse à l'évènement  $A$  "aucun individu n'a été choisi plus d'une fois" qui est la même chose que "tous les individus sont distincts". Alors on a,  $A = \Omega_2$  et :

$$\mathbb{P}(A) = \frac{\text{card}(\Omega_2)}{\text{card}(\Omega_1)} = \frac{A_N^r}{N^r}.$$

Donnons quelques applications de ce résultat.

1. On jette un dé cinq fois de suite. Alors la probabilité d'obtenir cinq nombres distincts est  $\frac{6 \cdot 5 \cdot 4 \cdot 3 \cdot 2}{6^5} \sim 0,09$ .
2. Supposons que dans une ville, il y ait sept accidents par semaine. Alors la probabilité d'avoir exactement un accident chaque jour de la semaine est  $\frac{7!}{7^7} \sim 0,00612$ .

**TIRAGES NON ORDONNÉS.** Une *sous-population* de taille  $r$  est un sous-ensemble  $\{s_{i_1}, \dots, s_{i_r}\}$  d'éléments de  $\mathcal{S}$ . De manière similaire aux tirages ordonnés, deux procédures sont possibles.

- *Le tirage sans remise.* On parle alors de *sous-population de taille  $r$  sans répétition*, ou de *combinaison de  $r$  éléments*. On impose à nouveau les conditions supplémentaires,  $r \leq N$ , et  $\forall j \neq k, i_j \neq i_k$ . Soit  $\Omega_3$  l'ensemble de ces populations, on a alors :

$$\text{card}(\Omega_3) = \frac{N!}{(N-r)!r!} := \binom{N}{r}.$$



Ce nombre est appelé *coefficient binomial*.

*Démonstration.* Chacun des sous-ensembles à  $r$  éléments fournit  $r!$  échantillons de taille  $r$  sans répétition, de sorte que  $\text{card}(\Omega_2) = r! \text{card}(\Omega_3)$ .  $\square$

EXEMPLE 1.20.

1. On appelle main de poker l'ensemble des 5 cartes que chacun des quatre joueurs reçoit lors de la distribution d'un jeu qui en contient 32. Alors il existe  $\binom{32}{5}$  mains différentes. Soit  $A$  l'évènement "les hauteurs des 5 cartes sont différentes", calculons  $\text{card}(A)$ . On peut choisir ces hauteurs de  $\binom{8}{5}$  manières différentes. Il faut ensuite choisir la couleur (trèfle, carreau, cœur, pique) de chacune des hauteurs. Ainsi :

$$\text{card}(A) = \binom{8}{5} 4^5.$$

Étant donné que toutes les mains sont supposées équiprobables, la probabilité d'obtenir une main dont les 5 cartes ont une hauteur différente est :

$$\mathbb{P}(A) = \frac{\binom{8}{5} 4^5}{\binom{32}{5}}.$$

2. Une urne contient  $N_b$  boules blanches et  $N_n$  boules noires. Posons  $N = N_b + N_n$ . On tire  $r$  boules avec remise dans l'urne, il y a alors  $N^r$  tirages possibles. Soit  $A_k$  l'évènement "on a tiré exactement  $k$  boules blanches", calculons  $\text{card}(A_k)$ . L'évènement  $A_k$  est réalisé lorsque l'issue est constituée de  $k$  boules blanches et  $r - k$  boules noires. Il y a  $\binom{r}{k}$  façons de choisir la position des boules blanches, la position des boules noires est ensuite fixée. Pour chacune des positions de boule blanche, il y a ensuite  $N_b$  choix de boules blanches possibles, et pour chacune des positions de boule noire, il y a  $N_n$  choix possibles, ainsi :

$$\text{card}(A_k) = \binom{r}{k} N_b^k N_n^{r-k}.$$

Étant donné que tous les tirages sont supposés équiprobables, la probabilité d'obtenir exactement  $k$  boules blanches lors d'un tirage de  $r$  boules avec remise est :

$$\mathbb{P}(A_k) = \frac{\binom{r}{k} N_b^k N_n^{r-k}}{N^r} = \binom{r}{k} \left(\frac{N_b}{N}\right)^k \left(\frac{N_n}{N}\right)^{r-k}.$$

Ceci est un exemple de la *loi binomiale*, que nous reverrons plus tard.

3. Soit  $\mathcal{S}$  une population de taille  $N$  (*ex.* des étudiants), que l'on range en deux catégories  $a$  et  $b$  incompatibles (*ex.* filles et garçons), de tailles respectives  $N_a$  et  $N_b = N - N_a$ . On choisit "au hasard" une sous-population de taille  $r$  sans répétition, il y a alors  $\binom{N}{r}$  choix possibles. Soit  $A_k$  l'évènement "on a choisi exactement  $k$  individus de la catégorie  $a$ ", calculons  $\text{card}(A_k)$ . L'évènement est réalisé lorsque l'issue est constituée de  $k$  individus de la catégorie  $a$  et  $r - k$  de la catégorie  $b$ . Il y a  $\binom{N_a}{k}$  façons de choisir les  $k$  individus de la catégorie  $a$  et pour chacune il y a  $\binom{N - N_a}{r - k}$  façons de choisir les individus restants dans la catégorie  $b$ , ainsi :

$$\text{card}(A_k) = \binom{N_a}{k} \binom{N - N_a}{r - k}.$$

Remarquer que pour que ceci ait un sens, il faut que  $0 \leq k \leq \min\{r, N_a\}$ . Étant donné que tous les tirages sont supposés équiprobables, la probabilité d'obtenir  $k$  individus de la catégorie  $a$  lors de ce tirage est :

$$\mathbb{P}(A_k) = \frac{\binom{N_a}{k} \binom{N_b}{r-k}}{\binom{N}{r}}.$$

Ceci est un exemple de la *loi hypergéométrique*.

*Remarque.* Supposons que  $N_a = N_a(N)$  soit une fonction de  $N$  et que le nombre total de boules tende vers l'infini, de sorte que la proportion  $\frac{N_a}{N}$  tende vers  $p$  (et donc que  $\frac{N_b}{N}$  tende vers  $1-p$ ), avec  $0 < p < 1$ . Ainsi,  $N_a$  et  $N_b$  tendent vers  $+\infty$  avec  $N$ . Fixons  $r \geq 0$  et  $k$  compris entre 0 et  $r$ . Alors, pour  $N$  assez grand, on a  $N_a \geq k$ ,  $N_b \geq r-k$  et  $\mathbb{P}(A_k)$  peut s'écrire :

$$\begin{aligned} \mathbb{P}(A_k) &= \frac{N_a(N_a-1)\dots(N_a-k+1)}{k!} \frac{N_b(N_b-1)\dots(N_b-r+k+1)}{(r-k)!} \frac{r!}{N(N-1)\dots(N-r+1)} \\ &= \frac{r!}{k!(r-k)!} \frac{N_a(N_a-1)\dots(N_a-k+1) N_b(N_b-1)\dots(N_b-r+k+1)}{N(N-1)\dots(N-r+1)} \\ &= \frac{r!}{k!(r-k)!} \frac{N^k \frac{N_a}{N} (\frac{N_a}{N} - \frac{1}{N}) \dots (\frac{N_a}{N} - \frac{k-1}{N}) N^{r-k} \frac{N_b}{N} (\frac{N_b}{N} - \frac{1}{N}) \dots (\frac{N_b}{N} - \frac{r-k-1}{N})}{N^r 1(1 - \frac{1}{N}) \dots (1 - \frac{r-1}{N})} \\ &= \binom{r}{k} \frac{\frac{N_a}{N} (\frac{N_a}{N} - \frac{1}{N}) \dots (\frac{N_a}{N} - \frac{k-1}{N}) \frac{N_b}{N} (\frac{N_b}{N} - \frac{1}{N}) \dots (\frac{N_b}{N} - \frac{r-k-1}{N})}{1(1 - \frac{1}{N}) \dots (1 - \frac{r-1}{N})} \\ &\xrightarrow{N \rightarrow +\infty} \frac{\binom{r}{k} p^k (1-p)^{r-k}}{1}. \end{aligned}$$

Ainsi,  $\mathbb{P}(A_k)$  tend vers  $\binom{r}{k} p^k (1-p)^{r-k}$ . On a donc obtenu la loi binomiale comme limite de lois hypergéométriques. Ce résultat est intuitif, car lorsque le nombre de boules est très grand, que le tirage s'effectue avec ou sans remise ne change pas grand chose : on a peu de chance de tirer deux fois la même boule.

• *Partitionnement.* Soient  $r_1, \dots, r_k$  des entiers positifs (éventuellement nuls) tels que,  $r_1 + \dots + r_k = N$ . Le nombre de façons de répartir  $N$  objets dans  $k$  familles de sorte que la  $i$ -ième famille contienne  $r_i$  éléments est égal à :

$$\frac{N!}{r_1! \dots r_k!}.$$

Ce nombre se note  $\binom{N}{r_1 \dots r_k}$  et s'appelle *coefficient multinomial*.

*Démonstration.* Pour remplir la première famille, il faut choisir  $r_1$  objets parmi  $N$ , ce qui peut se faire de  $\binom{N}{r_1}$  façons. Pour remplir la seconde famille, il faut choisir  $r_2$  objets parmi  $N - r_1$ , soit  $\binom{N-r_1}{r_2}$ . En continuant ainsi, on obtient que le nombre de telles répartitions est de :

$$\binom{N}{r_1} \binom{N-r_1}{r_2} \dots \binom{N-r_1-\dots-r_{k-1}}{r_k} = \frac{N!}{r_1! \dots r_k!}.$$

□

EXEMPLE. Le nombre d'anagrammes du mot CHERCHER est  $\frac{8!}{2!2!2!2!}$ .

• *Le tirage avec remise.* On parle alors de *sous-population de taille  $r$  avec répétitions*. Soit  $\Omega_4$  l'ensemble de ces populations, on a alors :

$$\text{card}(\Omega_4) = \binom{N+r-1}{N-1} = \binom{N+r-1}{r}.$$

*Démonstration.* Ce problème revient à placer  $r$  boules indistinguables dans  $N$  urnes. En effet, le nombre de boules dans la  $i$ -ième urne compte le nombre de répétitions de l'individu  $i$  lors du tirage. Représentons les  $r$  boules par  $r$  étoiles alignées, avec une cloison à chacune des extrémités. Par exemple, lorsque  $r = 7$ ,

$$|*****|$$

Répartir les  $r$  boules dans  $N$  urnes revient à rajouter  $N - 1$  cloisons formant les  $N$  urnes. Par exemple, lorsque  $r = 7$ ,  $N = 3$ ,

$$|**||*****|,$$

représente le tirage :  $s_1, s_1, s_3, s_3, s_3, s_3, s_3$ . Ainsi, ce problème revient à placer  $N - 1$  cloisons sur  $N + r - 1$  positions, les positions restantes étant occupées par des  $*$ . □

EXEMPLE. Soient  $r \in \mathbb{N}^*$  et  $n \in \mathbb{N}^*$ . On cherche à compter le nombre de suites d'entiers naturels  $r_1, \dots, r_n$ , telles que :

$$r_1 + \dots + r_n = r.$$

Ce problème revient à placer  $r$  boules indistinguables dans  $n$  urnes, où le nombre de boules dans la  $i$ -ième urne représente  $r_i$ . Ainsi, le nombre de ces suites est  $\binom{n+r-1}{n-1}$ . Par exemple, si  $r = 10$ ,  $n = 3$ ,

$$|**|*****|***|$$

représente la partition  $(2, 5, 3)$  de 10. Remarquer que ces suites sont naturellement ordonnées de sorte que l'on distingue  $(2, 5, 3)$  de  $(5, 3, 2)$ .

### 1.3.2 UNIVERS DÉNOMBRABLES

Souvenez-vous que, dans ce cas, on prend  $\mathcal{F} = \mathcal{P}(\Omega)$  et qu'une probabilité sur  $(\Omega, \mathcal{P}(\Omega))$  est entièrement déterminée par la donnée des nombres  $(p_\omega)_{\omega \in \Omega}$  tels que, pour tout  $\omega \in \Omega$ ,  $0 \leq p_\omega \leq 1$  et  $\sum_{\omega \in \Omega} p_\omega = 1$  ; on définit alors, pour tout  $\omega \in \Omega$ ,  $\mathbb{P}(\{\omega\}) = p_\omega$ .

EXEMPLE 1.21. On jette une pièce de monnaie jusqu'à l'obtention du premier pile. On peut choisir  $\Omega = \mathbb{N}^* \cup \{\infty\}$  où, pour tout  $k \in \mathbb{N}^*$ ,  $\{k\}$  représente l'évènement "le premier pile est obtenu au  $k$ -ième jet", et  $\{\infty\}$  représente l'évènement "pile ne sort jamais". Si la pièce est équilibrée, on aura, pour tout  $k \in \mathbb{N}^*$  :

$$\mathbb{P}(\{k\}) = \frac{1}{2^k}.$$

Comme  $\mathbb{N}^*$  et  $\{\infty\}$  sont disjoints,  $1 = \mathbb{P}(\Omega) = \mathbb{P}(\{\infty\}) + \mathbb{P}(\bigcup_{k \in \mathbb{N}^*} \{k\})$ . Ainsi, la probabilité que

pile ne sorte jamais est donnée par :

$$\begin{aligned}\mathbb{P}(\{\infty\}) &= 1 - \mathbb{P}\left(\bigcup_{k \in \mathbb{N}^*} \{k\}\right) \\ &= 1 - \sum_{k=1}^{+\infty} \mathbb{P}(\{k\}), \text{ car les évènements sont disjoints} \\ &= 1 - \sum_{k=1}^{+\infty} \frac{1}{2^k} = 0.\end{aligned}$$

La probabilité que le premier pile sorte après un nombre pair de lancers est :

$$\mathbb{P}(\{2, 4, 6, \dots\}) = \sum_{k=1}^{+\infty} \mathbb{P}(\{2k\}) = \sum_{k=1}^{+\infty} \frac{1}{2^{2k}} = \frac{1}{3}.$$

## 1.4 LE CAS GÉNÉRAL

Lors que l'univers  $\Omega$  est infini non-dénombrable, la situation est en général beaucoup plus subtile que dans les cas précédents. Il est rarement imaginable de donner la probabilité de tous les évènements, et la probabilité des singletons est souvent nulle.

Un résultat très utile du cours de “théorie de la mesure” dit la chose suivante. Considérons une classe d'évènements  $\mathcal{C}$  stable par intersections finies (un  $\pi$ -système), qui est telle que  $\sigma(\mathcal{C}) = \mathcal{F}$ . Alors comme conséquence du lemme de classe monotone, une mesure de probabilité est entièrement déterminée par ses valeurs sur  $\mathcal{C}$ . Par exemple, une mesure de probabilité sur  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$  est entièrement déterminée par sa valeur sur les intervalles de la forme  $] - \infty, x]$ ,  $x \in \mathbb{R}$ . Ainsi, sachant qu'une mesure existe, on n'a pas besoin de la connaître sur tous les évènements pour la caractériser. Par ailleurs, la question de l'existence de mesures de probabilité est très délicate. Elle utilise par exemple le théorème d'extension de Carathéodory, qui dépasse largement le contenu de ce cours.

Il existe cependant des exemples de mesures de probabilité très facile à décrire, même dans le cadre d'espace d'états généraux.

EXEMPLE 1.22. Soit  $(\Omega, \mathcal{F})$  un univers et une tribu, et soit  $\omega_0 \in \Omega$  fixé. La *mesure de Dirac* en  $\omega_0$ , notée  $\delta_{\omega_0}$ , est la probabilité définie par :

$$\forall A \in \mathcal{F}, \quad \delta_{\omega_0}(A) = \begin{cases} 1 & \text{si } \omega_0 \in A, \\ 0 & \text{sinon.} \end{cases}$$

Ainsi, une propriété/un évènement est presque sûre pour  $\delta_{\omega_0}$  si  $\omega_0$  appartient à l'évènement, autrement dit, la propriété est satisfaite par  $\omega_0$ . L'ensemble  $\Omega \setminus \{\omega_0\}$  est négligeable pour  $\delta_{\omega_0}$ .

Si de plus  $\Omega$  est dénombrable,  $\Omega = \{\omega_1, \omega_2, \dots\}$  et si on a  $\mathbb{P}$  une probabilité sur  $(\Omega, \mathcal{P}(\Omega))$  :  $p_\omega = \mathbb{P}(\{\omega\})$ . Alors, pour tout  $A \in \mathcal{F}$ ,

$$\mathbb{P}(A) = \sum_{n \geq 1} p_{\omega_n} \delta_{\omega_n}(A); \text{ ainsi } \mathbb{P} = \sum_{n \geq 1} p_{\omega_n} \delta_{\omega_n}.$$

Autrement dit, une mesure de probabilité sur un espace dénombrable s'écrit comme une somme pondérée de mesures de Dirac sur les singletons de  $\Omega$ .



## CHAPITRE 2

# CONDITIONNEMENT ET INDÉPENDANCE

### 2.1 PROBABILITÉ CONDITIONNELLE

#### 2.1.1 DÉFINITION

Motivons la définition de probabilité conditionnelle sur un exemple. Soit  $\Omega$  un groupe d'étudiant.e.s partitionné en  $\Omega = R \cup R^c$ , où  $R$  représente l'ensemble des étudiant.e.s qui ont révisé l'examen. Soit  $F$  l'ensemble des femmes, de sorte que  $F \cup F^c$  représente une autre partition de  $\Omega$ . On suppose que l'on choisit un.e étudiant.e "au hasard", de sorte que l'on munit  $\Omega$  de la probabilité uniforme, notée  $\mathbb{P}$ . Ainsi, la probabilité que l'étudiant.e. choisi ait révisé l'examen est :

$$\mathbb{P}(R) = \frac{\text{card}(R)}{\text{card}(\Omega)}.$$

Si maintenant on choisit un.e étudiant.e avec l'information supplémentaire qu'il s'agit d'une femme, tout se passe comme si l'univers considéré est  $F$ , et que l'on a partitionné  $F$  (et non plus  $\Omega$ ) en  $R \cap F$  et  $R^c \cap F$ . Ainsi la probabilité que l'étudiant.e choisi ait révisé, étant donné l'information que c'est une femme, est égale à :

$$\frac{\text{card}(R \cap F)}{\text{card}(F)} = \frac{\text{card}(R \cap F) / \text{card}(\Omega)}{\text{card}(F) / \text{card}(\Omega)} = \frac{\mathbb{P}(R \cap F)}{\mathbb{P}(F)}.$$

L'idée est qu'une information supplémentaire concernant l'expérience modifie la vraisemblance que l'on accorde à l'évènement étudié.

**Définition 2.1.** Soit  $(\Omega, \mathcal{F}, \mathbb{P})$  un espace probabilisé, et  $B$  un évènement tel que  $\mathbb{P}(B) > 0$ . Pour tout  $A \in \mathcal{F}$ , on définit la *probabilité conditionnelle de  $A$  sachant  $B$* , notée  $\mathbb{P}(A|B)$ , par :

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}. \quad (2.1)$$

**Lemme 2.2.** *Sous les hypothèses de la définition ci-dessus, l'application  $\mathbb{P}(\cdot|B) : \mathcal{F} \rightarrow \mathbb{R}^+$ , définie par  $A \mapsto \mathbb{P}(A|B)$ , est une probabilité sur  $(\Omega, \mathcal{F})$ , ainsi  $(\Omega, \mathcal{F}, \mathbb{P}(\cdot|B))$  est un espace probabilisé.*

*Démonstration.* Il faut montrer que  $\mathbb{P}(\cdot|B)$  satisfait aux trois axiomes caractérisant une probabilité.

- Pour tout  $A \in \mathcal{F}$ , on a  $A \cap B \subset B$ , d'où  $0 \leq \mathbb{P}(A \cap B) \leq \mathbb{P}(B)$ , et donc :

$$0 \leq \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \leq 1.$$

- Comme  $\Omega \cap B = B$ , on a  $\mathbb{P}(\Omega|B) = 1$ .
- Soit  $(A_n)_{n \geq 1}$  une famille dénombrable d'évènements de  $\mathcal{F}$  deux-à-deux disjoints. Alors on a l'égalité  $(\cup_{n \geq 1} A_n) \cap B = \cup_{n \geq 1} (A_n \cap B)$ . Comme les évènements  $(A_n)_{n \geq 1}$  sont deux-à-deux disjoints, il en est de même pour les évènements  $(A_n \cap B)_{n \geq 1}$ . Ainsi la  $\sigma$ -additivité de  $\mathbb{P}$  implique  $\mathbb{P}(\cup_{n \geq 1} (A_n \cap B)) = \sum_{n \geq 1} \mathbb{P}(A_n \cap B)$ , et on conclut :

$$\mathbb{P}\left(\bigcup_{n \geq 1} A_n \mid B\right) = \frac{\mathbb{P}\left(\left(\bigcup_{n \geq 1} A_n\right) \cap B\right)}{\mathbb{P}(B)} = \sum_{n \geq 1} \frac{\mathbb{P}(A_n \cap B)}{\mathbb{P}(B)} = \sum_{n \geq 1} \mathbb{P}(A_n|B). \quad \square$$

*Remarque 2.3.* Comme  $(\Omega, \mathcal{F}, \mathbb{P}(\cdot|B))$  est un espace probabilisé,  $\mathbb{P}(\cdot|B)$  satisfait à toutes les propriétés du Corollaire 1.12 et de la Proposition 1.13.

**Proposition 2.4** (Formule des probabilités composées).

1. Si  $\mathbb{P}(A) > 0$  et  $\mathbb{P}(B) > 0$ , on peut réécrire l'équation (2.1) sous la forme :

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

2. Si  $n \geq 2$  et  $A_1, \dots, A_n$  sont des évènements de  $\mathcal{F}$  tels que  $\mathbb{P}(\cap_{i=1}^{n-1} A_i) > 0$ , alors :

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

*Démonstration.* Le Point 1. est évident. Le Point 2. sera démontré en exercices.  $\square$

**EXEMPLE 2.5.** On considère deux lancers successifs d'un même dé. Sachant que le premier jet donne 3, on souhaite calculer la probabilité que la somme soit strictement supérieure à 6.

Supposons que l'on ait choisi l'univers  $\Omega = \{1, \dots, 6\}^2$ , que l'on munit de la tribu  $\mathcal{F} = \mathcal{P}(\Omega)$ . On est alors en situation d'équiprobabilité, et on munit  $\Omega$  de la probabilité uniforme, notée  $\mathbb{P}$ . Ainsi, pour tout évènement  $A$  de  $\mathcal{F}$ ,  $\mathbb{P}(A) = \frac{\text{card}(A)}{\text{card}(\Omega)} = \frac{\text{card}(A)}{36}$ .

Soit  $B$  l'évènement "le premier jet donne 3";  $B$  est le sous-ensemble  $\{(3, j) : j \in \{1, \dots, 6\}\}$  de  $\mathcal{F}$ , son cardinal est égal à 6, d'où  $\mathbb{P}(B) = \frac{6}{36} = \frac{1}{6} > 0$ . La probabilité  $\mathbb{P}(B)$  étant strictement positive, on peut conditionner par l'évènement  $B$ .

Soit  $A$  l'évènement "la somme des dés est strictement supérieure à 6";  $A$  est le sous-ensemble  $\{(i, j) \in \Omega : i + j > 6\}$  de  $\mathcal{F}$ , et  $A \cap B = \{(3, 4), (3, 5), (3, 6)\}$ , d'où :  $\mathbb{P}(A \cap B) = \frac{3}{36} = \frac{1}{12}$ . On conclut que :

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{1}{2}.$$

### 2.1.2 FORMULE DES PROBABILITÉS TOTALES ET FORMULE DE BAYES

**Définition 2.6.** Soit  $I$  une partie finie ou dénombrable de  $\mathbb{N}$ . Une famille  $(B_i)_{i \in I}$  d'évènements de  $\Omega$  forme un *système complet d'évènements* de  $\Omega$ , si

$$\forall i \neq j, B_i \cap B_j = \emptyset, \text{ et } \bigcup_{i \in I} B_i = \Omega.$$

Autrement dit, la famille  $(B_i)_{i \in I}$  est une *partition* de  $\Omega$ .

**Théorème 2.7.** Soit  $(\Omega, \mathcal{F}, \mathbb{P})$  un espace probabilisé.

- **Formule des probabilités totales.** Soit  $(B_i)_{i \in I}$  un système complet d'évènements de  $\mathcal{F}$ , tel que pour tout  $i \in I$ ,  $\mathbb{P}(B_i) > 0$ , et soit  $A \in \mathcal{F}$ . Alors :

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

- **Formule de Bayes.** Sous les mêmes hypothèses que précédemment et si  $\mathbb{P}(A) > 0$ ,

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

*Démonstration.* Comme  $(B_i)_{i \in I}$  est une partition de  $\Omega$ , on a :

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

où les évènements  $(A \cap B_i)_{i \in I}$  sont deux-à-deux disjoints. Ainsi, en utilisant la  $\sigma$ -additivité de  $\mathbb{P}$  et la Proposition 2.4 :

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

ce qui démontre le premier point. De plus, d'après la définition de la probabilité conditionnelle et la Proposition 2.4, on a :

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

La formule de Bayes est obtenue en appliquant la formule des probabilités totales au dénominateur avec la partition  $(B_j)_{j \in I}$ .  $\square$

**EXEMPLE 2.8.** (Paradoxe de Simpson). Cet exemple réel<sup>1</sup> montre un paradoxe surprenant qui s'explique grâce aux probabilités conditionnelles et à la formule des probabilités totales. Il vous convaincra de l'importance de bien comprendre ce concept pour interpréter correctement les résultats d'études statistiques. Il provient d'une étude médicale sur le succès de deux traitements contre les calculs rénaux. Le traitement A a été effectué dans les années 1972-1980, et le traitement B dans les années 1980-1985.

La première table montre le succès global et le nombre de traitements pour chaque méthode.

---

1. Charig CR, ; Webb DR, ; Payne SR, ; Wickham OE . Comparison of treatment of renal calculi by operative surgery, percutaneous nephrolithotomy, and extracorporeal shock wave lithotripsy. BMJ 1986 ;292 : 879–82. 3



| Succès (taux de succès) |               |
|-------------------------|---------------|
| Traitement A            | Traitement B  |
| 273/350 (78%)           | 289/350 (83%) |

Cela semble révéler que traitement B, qui est nouveau, est plus efficace. Maintenant, en ajoutant des données concernant la taille des calculs, la comparaison prend une autre tournure :

| Résultats en fonction de la taille des calculs |               |                |              |
|------------------------------------------------|---------------|----------------|--------------|
| petits calculs                                 |               | grands calculs |              |
| Traitement A                                   | Traitement B  | Traitement A   | Traitement B |
| (81/87) 93%                                    | (234/270) 87% | (192/263) 73%  | (55/80) 69%  |

L'information au sujet de la taille des calculs a inversé les conclusions concernant l'efficacité de chaque traitement. Le traitement A est maintenant considéré comme plus efficace dans les deux cas. Le rebroussement de cette inégalité, qui conduit au paradoxe, se produit à cause de deux effets concurrents :

1. la variable supplémentaire (ici la taille) a un impact significatif sur les rapports ;
2. les tailles des groupes qui sont combinés quand la variable supplémentaire est ignorée sont très différentes. (Les groupes utilisés pour le traitement A et B ont la même taille, mais n'ont pas la même répartition de petits et grands calculs).

Vérifions les calculs pour le traitement A. On choisit comme univers  $\Omega$  les 350 patients de l'échantillon, que l'on munit de la tribu  $\mathcal{P}(\Omega)$ . Appelons  $S$  l'évènement "le traitement est un succès",  $P$  l'évènement "les calculs sont petits". Alors d'après le premier tableau,  $\mathbb{P}(S) = \frac{273}{350}$ , et d'après le deuxième,  $\mathbb{P}(P) = \frac{87}{350}$ ,  $\mathbb{P}(P^c) = \frac{263}{350}$ . Ces deux probabilités étant strictement positives, on peut conditionner par les évènements correspondants. D'après le deuxième tableau toujours, on a  $\mathbb{P}(S|P) = \frac{81}{87}$ ,  $\mathbb{P}(S|P^c) = \frac{192}{263}$ . Utilisons la formule des probabilités totales pour calculer  $\mathbb{P}(S)$ .

$$\begin{aligned}\mathbb{P}(S) &= \mathbb{P}(S|P)\mathbb{P}(P) + \mathbb{P}(S|P^c)\mathbb{P}(P^c) \\ &= \frac{81}{87} \frac{87}{350} + \frac{192}{263} \frac{263}{350} = \frac{273}{350}.\end{aligned}$$

On retrouve bien le résultat du premier tableau. Des calculs similaires permettent de vérifier les résultats pour le traitement B. Ainsi, ces résultats apparemment contradictoires s'expliquent aisément grâce aux probabilités conditionnelles.

## 2.2 INDÉPENDANCE DES ÉVÈNEMENTS

Dans toute cette section,  $(\Omega, \mathcal{F}, \mathbb{P})$  est un espace probabilisé.

**Définition 2.9.** Deux évènements  $A$  et  $B$  de  $\mathcal{F}$  sont dits *indépendants*, si

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

EXEMPLE 2.10.

1. Les évènements  $\emptyset$  et  $\Omega$  sont indépendants. En effet,

$$\mathbb{P}(\emptyset \cap \Omega) = \mathbb{P}(\emptyset) = 0, \text{ et } \mathbb{P}(\emptyset)\mathbb{P}(\Omega) = 0.1 = 0,$$

d'où,  $\mathbb{P}(\emptyset \cap \Omega) = \mathbb{P}(\emptyset)\mathbb{P}(\Omega)$ .

2. On jette un dé parfaitement équilibré. Soit  $A$  l'évènement "obtenir 1,2 ou 3", et  $B$  l'évènement "obtenir 1,2,4 ou 5". On choisit comme espace des états,  $\Omega = \{1, \dots, 6\}$ , que l'on munit de la probabilité uniforme. On a alors,

$$\mathbb{P}(A) = \frac{1}{2}, \mathbb{P}(B) = \frac{2}{3}, \mathbb{P}(A \cap B) = \mathbb{P}(\{1, 2\}) = \frac{1}{3}.$$

Ainsi, comme  $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$ , on déduit que les évènements  $A$  et  $B$  sont indépendants.

*Remarque 2.11.*

1. L'indépendance est liée au choix de probabilité  $\mathbb{P}$  et n'est donc pas une notion ensembliste. En particulier, cela n'a rien à voir avec le fait que  $A$  et  $B$  soient disjoints ou non. Dans l'Exemple 2.10 ci-dessus, les évènements sont indépendants, mais non disjoints ( $A \cap B \neq \emptyset$ ).
2. Si  $A$  et  $B$  sont deux évènements de probabilité strictement positive, alors :

$$A \text{ et } B \text{ sont indépendants} \Leftrightarrow \mathbb{P}(A) = \mathbb{P}(A|B) \Leftrightarrow \mathbb{P}(B) = \mathbb{P}(B|A).$$

Le fait d'avoir une information supplémentaire, à savoir que  $B$  est réalisé, ne modifie pas la probabilité de  $A$  (de même pour  $B$  lorsqu'on sait que  $A$  est réalisé) ce qui justifie la terminologie d'*indépendance*. Ces critères ne sont cependant pas utilisés comme définition car ils nécessitent l'hypothèse supplémentaire,  $\mathbb{P}(A) > 0$  et  $\mathbb{P}(B) > 0$ .

**Proposition 2.12.** *Si les évènements  $A$  et  $B$  sont indépendants, il en est de même des évènements  $A^c$  et  $B$ ,  $A$  et  $B^c$ ,  $A^c$  et  $B^c$ .*

*Démonstration.* Démontrons que  $A^c$  et  $B$  sont indépendants si  $A$  et  $B$  le sont.

$$\begin{aligned} \mathbb{P}(A^c \cap B) &= \mathbb{P}(B) - \mathbb{P}(A \cap B), \text{ d'après la formule des probabilités totales} \\ &= \mathbb{P}(B)(1 - \mathbb{P}(A)), \text{ car } A \text{ et } B \text{ sont indépendants} \\ &= \mathbb{P}(B)\mathbb{P}(A^c), \text{ donc } A^c \text{ et } B \text{ sont indépendants.} \end{aligned}$$

Les autres implications s'obtiennent de manière similaire. □

**Définition 2.13.** Soit  $(A_i)_{i \in I}$  une famille d'évènements de  $\mathcal{F}$ , avec  $I$  dénombrable ou fini.

- Les évènements  $(A_i)_{i \in I}$  sont dits *mutuellement indépendants*, ou simplement *indépendants* si, pour toute partie finie  $K$  de  $I$ ,

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

- Les évènements  $(A_i)_{i \in I}$  sont dits *deux-à-deux indépendants*, si pour tout  $i, j \in I$ ,  $i \neq j$ , on a :

$$\mathbb{P}(A_i \cap A_j) = \mathbb{P}(A_i)\mathbb{P}(A_j).$$

EXEMPLE 2.14. L'exemple suivant a pour but de montrer que la notion d'indépendance deux-à-deux est plus faible que celle d'indépendance.

Soit  $\Omega = \{1, 2, 3, 4\}$  muni de la tribu  $\mathcal{F} = \mathcal{P}(\Omega)$  et de la probabilité uniforme, notée  $\mathbb{P}$ . On considère les événements  $A = \{1, 2\}$ ,  $B = \{2, 3\}$ ,  $C = \{1, 3\}$ . Montrons que les événements  $A, B, C$  sont deux-à-deux indépendants, mais pas indépendants.

Comme  $(\Omega, \mathcal{F})$  est munit de la probabilité uniforme, on a pour tout événement  $A$  de  $\mathcal{F}$ ,  $\mathbb{P}(A) = \frac{\text{card}(A)}{\text{card}(\Omega)} = \frac{\text{card}(A)}{4}$ . Ainsi,

$$\mathbb{P}(A) = \mathbb{P}(B) = \mathbb{P}(C) = \frac{2}{4} = \frac{1}{2}.$$

D'autre part,  $A \cap B = \{2\}$ ,  $B \cap C = \{3\}$ ,  $A \cap C = \{1\}$ ,  $A \cap B \cap C = \{\emptyset\}$ . D'où :

$$\mathbb{P}(A \cap B) = \mathbb{P}(B \cap C) = \mathbb{P}(A \cap C) = \frac{1}{4}, \quad \text{et } \mathbb{P}(A \cap B \cap C) = 0.$$

D'après les calculs ci-dessus, on a :

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B) = \frac{1}{4}, \quad \mathbb{P}(B \cap C) = \mathbb{P}(B)\mathbb{P}(C) = \frac{1}{4}, \quad \mathbb{P}(A \cap C) = \mathbb{P}(A)\mathbb{P}(C) = \frac{1}{4},$$

d'où les événements  $A, B, C$  sont deux-à-deux indépendants.

Cependant,  $\mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C) = \frac{1}{8}$  et  $\mathbb{P}(A \cap B \cap C) = 0$ , d'où les événements  $A, B, C$  ne sont pas mutuellement indépendants.

## 2.3 LEMME DE BOREL-CANTELLI

Voici un résultat fameux et très utile de probabilités, qui sert en particulier pour démontrer la loi des grands nombres. Au préalable, nous avons besoin de la définition suivante.

**Définition 2.15.** Soit  $(A_n)_{n \geq 1}$  une suite d'événements de  $\mathcal{F}$ .

- L'ensemble  $\limsup_n A_n$  est défini par :

$$\limsup_n A_n = \bigcap_{k \geq 1} \bigcup_{n \geq k} A_n = \{\omega \in \Omega : \forall k \geq 1, \exists n \geq k, \omega \in A_n\}.$$

Il s'agit de l'ensemble des  $\omega$  qui appartiennent à une infinité de  $A_n$ .

- L'ensemble  $\liminf_n A_n$  est défini par :

$$\liminf_n A_n = \bigcup_{k \geq 1} \bigcap_{n \geq k} A_n = \{\omega \in \Omega : \exists k \geq 1, \forall n \geq k, \omega \in A_n\}.$$

Il s'agit de l'ensemble des  $\omega$  qui appartiennent à tous les  $A_n$  à partir d'un certain rang.

*Remarque 2.16.* Les événements  $\limsup_n A_n$  et  $\liminf_n A_n$  appartiennent bien à  $\mathcal{F}$  car ils sont intersection et réunion dénombrable d'événements de  $\mathcal{F}$ .

**Théorème 2.17** (Lemme de Borel-Cantelli). *Soit  $(A_n)_{n \geq 1}$  une suite d'événements de  $\mathcal{F}$ .*

1. On a :

$$\sum_{n \geq 1} \mathbb{P}(A_n) < \infty \quad \Rightarrow \quad \mathbb{P}(\limsup_n A_n) = 0.$$

*Autrement dit,  $\mathbb{P}$ -presque sûrement, il y a au plus un nombre fini de  $A_n$  qui sont réalisés.*

2. Si de plus la suite  $(A_n)_{n \geq 1}$  est indépendante, alors

$$\sum_{n \geq 1} \mathbb{P}(A_n) = \infty \quad \Rightarrow \quad \mathbb{P}(\limsup_n A_n) = 1.$$

Autrement dit,  $\mathbb{P}$ -presque sûrement, il y a une infinité de  $A_n$  qui sont réalisés.

*Démonstration.*

1. Pour tout  $k \geq 1$ , soit  $B_k = \cup_{n \geq k} A_n$ . Alors  $(B_k)_{k \geq 1}$  est une suite décroissante d'évènements et

$$\begin{aligned} \mathbb{P}(\limsup_n A_n) &= \mathbb{P}(\cap_{k \geq 1} B_k) = \lim_{k \rightarrow \infty} \mathbb{P}(B_k), \text{ d'après la Proposition 1.13, Point 3.} \\ &\leq \lim_{k \rightarrow \infty} \sum_{n \geq k} \mathbb{P}(A_n), \text{ d'après la Proposition 1.13, Point 1.} \end{aligned}$$

Si la série  $\sum_{n \geq 1} \mathbb{P}(A_n)$  est convergente, alors son reste tend vers 0 et, par comparaison, on en déduit que  $\mathbb{P}(\limsup_n A_n) = 0$ , ce qui démontre le Point 1.

2. Il est équivalent de démontrer que,

$$\mathbb{P}([\limsup_n A_n]^c) = \mathbb{P}(\cup_{k \geq 1} \cap_{n \geq k} A_n^c) = 0.$$

Pour tout  $k \geq 1$ , soit  $C_k = \cap_{n \geq k} A_n^c$ . Alors  $(C_k)_{k \geq 1}$  est une suite croissante d'évènements et d'après la Proposition 1.13, Point 2.,

$$\mathbb{P}(\cup_{k \geq 1} \cap_{n \geq k} A_n^c) = \lim_{k \rightarrow \infty} \mathbb{P}(C_k). \quad (2.2)$$

Or, pour tout  $\ell \geq k$ ,

$$\begin{aligned} \mathbb{P}(C_k) &\leq \mathbb{P}(\cap_{n=k}^{\ell} A_n^c), \text{ d'après le Corollaire 1.12, Point 5.} \\ &= \prod_{n=k}^{\ell} \mathbb{P}(A_n^c) = \prod_{n=k}^{\ell} (1 - \mathbb{P}(A_n)), \text{ par indépendance, et passage au complémentaire} \\ &\leq \prod_{n=k}^{\ell} e^{-\mathbb{P}(A_n)} = e^{-\sum_{n=k}^{\ell} \mathbb{P}(A_n)}, \text{ car } 1 - x \leq e^{-x} \text{ pour } x \geq 0. \end{aligned}$$

Le membre de gauche est indépendant de  $\ell$  ainsi, pour tout  $k \geq 1$ ,  $\mathbb{P}(C_k) \leq e^{-\sum_{n=k}^{\infty} \mathbb{P}(A_n)}$ . Or, si  $\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty$ , pour tout  $k \geq 1$ ,  $e^{-\sum_{n=k}^{\infty} \mathbb{P}(A_n)} = 0$ , ce qui implique que  $\mathbb{P}(C_k) = 0$ . On conclut que  $\mathbb{P}(\cup_{k \geq 1} \cap_{n \geq k} A_n^c) = 0$  en utilisant l'équation (2.2).  $\square$

## 2.4 EXPÉRIENCES ALÉATOIRES INDÉPENDANTES ET PROBABILITÉ PRODUIT

Le but de cette section est d'étudier la description mathématique d'une expérience aléatoire répétée de manière indépendante, un nombre fini ou infini de fois. Regardons d'abord le cadre général, puis nous donnerons un exemple.

Soit  $((\Omega_n, \mathcal{F}_n, \mathbb{P}_n))_{n \geq 1}$  une suite d'espaces probabilisés, où chacun rend compte d'une expérience faite au temps  $n$  et les expériences sont indépendantes les unes des autres.

**CAS DE  $k$  ESPACES.** On prend  $\Omega = \Omega_1 \times \cdots \times \Omega_k$ , muni de la *tribu produit*  $\mathcal{F} = \mathcal{F}_1 \otimes \cdots \otimes \mathcal{F}_k$ , qui est la tribu engendrée par les pavés  $A_1 \times \cdots \times A_k$ , avec  $A_n \in \mathcal{F}_n$ ,  $n = 1, \dots, k$ . On définit  $\mathbb{P}$  sur les pavés par

$$\mathbb{P}(A_1 \times \cdots \times A_k) = \mathbb{P}_1(A_1) \cdots \mathbb{P}_k(A_k),$$

et en se référant à la Section 1.4, on montre que cela suffit pour caractériser une probabilité  $\mathbb{P}$  sur  $(\Omega, \mathcal{F})$ , appelée *probabilité produit*. En utilisant le cours de théorie de la mesure, on peut aussi montrer qu'une telle mesure existe.

**NOMBRE DÉNOMBRABLE D'ESPACES.** On prend  $\Omega = \prod_{n \geq 1} \Omega_n$ , muni de la *tribu produit*  $\mathcal{F}$ , aussi appelée *tribu des cylindres*, définie comme la tribu engendrée par les produits cartésiens finis d'événements des tribus  $(\mathcal{F}_n)_{n \geq 1}$ . Elle contient tous les événements de la forme

$$A_1 \times \cdots \times A_k \times \Omega_{k+1} \times \Omega_{k+2} \times \cdots, \quad A_n \in \mathcal{F}_n, \quad n \in \{1, \dots, k\}.$$

On peut montrer, voir le cours à venir de "théorie de la mesure", qu'il existe une unique probabilité  $\mathbb{P}$  sur  $(\Omega, \mathcal{F})$ , appelée *probabilité produit*, qui vérifie

$$\mathbb{P}(A_1 \times \cdots \times A_k \times \Omega_{k+1} \times \Omega_{k+2} \times \cdots) = \prod_{n=1}^k \mathbb{P}_n(A_n).$$

Un point important est que cette probabilité rend indépendantes les expériences aléatoires correspondants à chacun des espaces  $(\Omega_n, \mathcal{F}_n, \mathbb{P}_n)$ .

EXEMPLE 2.18. On lance un dé à 6 faces non-truqué, indéfiniment. Pour tout  $n \geq 1$ , on a  $(\Omega_n, \mathcal{F}_n, \mathbb{P}_n)$ , avec  $\Omega_n = \{1, \dots, 6\}$ ,  $\mathcal{F}_n = \mathcal{P}(\Omega_n)$ ,  $\mathbb{P}_n$  est la probabilité uniforme. Ainsi,

$$\Omega = \{1, \dots, 6\}^{\mathbb{N}^*} = \{(i_n)_{n \geq 1} : \forall n \geq 1, i_n \in \{1, \dots, 6\}\},$$

que l'on munit de la tribu produit  $\mathcal{F}$  et de la probabilité produit  $\mathbb{P}$ .

Pour tout  $k \geq 1$ , on définit  $A_k$  l'événement "on obtient 1 pour la première fois au  $k$ -ième jet". Alors,  $A_k$  est le sous-ensemble suivant de  $\Omega$  :

$$\{(i_n)_{n \geq 1} \in \Omega : i_1 \neq 1, \dots, i_{k-1} \neq 1, i_k = 1\},$$

et par définition de la probabilité produit, on a

$$\mathbb{P}(A_k) = \left[ \prod_{n=1}^{k-1} \mathbb{P}_n(i_n > 1) \right] \mathbb{P}_k(i_k = 1) = \left( \frac{5}{6} \right)^{k-1} \cdot \frac{1}{6}.$$

Ceci est un exemple de *loi géométrique* que nous reverrons plus tard.

# CHAPITRE 3

## VARIABLES ALÉATOIRES

Soit  $(\Omega, \mathcal{F}, \mathbb{P})$  un espace probabilisé. Plutôt que de travailler avec des événements de  $\mathcal{F}$ , il est souvent plus commode d'associer une valeur numérique aux résultats d'une expérience aléatoire. Par exemple, lors de  $n$  jets de pile ou face, il sera intéressant d'étudier le nombre de piles obtenus. Cela motive l'introduction de la notion de *variable aléatoire*, qui est une application  $X$  de  $\Omega$  dans un ensemble  $E$  qui sera typiquement,  $\mathbb{N}^d$ ,  $\mathbb{Z}^d$  ou  $\mathbb{R}^d$  ( $d \geq 1$ ), mais qui peut être plus général. On va alors transférer la structure probabiliste de l'espace de départ sur l'espace d'arrivée.

### 3.1 DÉFINITIONS

Afin de pouvoir étudier l'espace d'arrivée de manière probabiliste, nous devons également le munir d'une structure adéquate. Ainsi, soit  $E$  l'espace d'arrivée, que l'on munit d'une tribu  $\mathcal{E}$ .

**Définition 3.1.** Une *variable aléatoire*  $X$  est une application de  $(\Omega, \mathcal{F})$  dans  $(E, \mathcal{E})$  qui préserve les structures respectives de tribus, c'est-à-dire qu'elle vérifie

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

*Remarque 3.2.*

- Attention à la notation  $X^{-1}$  qui n'a rien à voir avec le fait que la fonction  $X$  soit bijective ;  $X^{-1}(B)$  désigne l'image réciproque de la partie  $B$  par l'application  $X$ .
- La terminologie de variable aléatoire est trompeuse, car il s'agit en fait d'une fonction de  $\Omega$  dans  $E$ . Une variable aléatoire  $X$  est une *application mesurable* de  $(\Omega, \mathcal{F})$  dans  $(E, \mathcal{E})$  (voir cours de théorie de la mesure de l'année prochaine). Lorsque le choix de tribus est clair, on écrira simplement que  $X$  est une variable aléatoire de  $\Omega$  dans  $E$ .

**Définition 3.3** (Proposition). Soit  $X$  une variable aléatoire. L'application  $\mathbb{P}^X : \mathcal{E} \rightarrow [0, 1]$  définie par :

$$\forall B \in \mathcal{E}, \quad \mathbb{P}^X(B) := \mathbb{P}(X^{-1}(B)) = \mathbb{P}(\{X \in B\}),$$

est une probabilité sur l'espace d'arrivée  $(E, \mathcal{E})$ , appelée *loi de la variable aléatoire*  $X$ .

*Remarque.*

- Il s'agit de “la probabilité image” de la probabilité  $\mathbb{P}$  sur l'espace d'arrivée. L'espace d'arrivée  $E$ , muni de la tribu  $\mathcal{E}$  et de la probabilité  $\mathbb{P}^X$  est un espace probabilisé  $(E, \mathcal{E}, \mathbb{P}^X)$ .

- On utilisera la notation  $X \sim \mu$  pour dire que  $X$  a loi  $\mu$ , autrement dit  $\mathbb{P}^X = \mu$ .

*Démonstration.* Notons d'abord que cette définition fait sens car, comme  $X$  est une variable aléatoire, l'ensemble  $\{X \in B\} \in \mathcal{F}$ . Vérifions ensuite les axiomes d'une probabilité. Comme  $0 \leq \mathbb{P} \leq 1$ , on a aussi  $0 \leq \mathbb{P}^X \leq 1$ . Ensuite,

$$\mathbb{P}^X(E) = \mathbb{P}(\{\omega \in \Omega : X(\omega) \in E\}) = \mathbb{P}(\Omega) = 1,$$

où dans la dernière égalité, on a utilisé que  $\mathbb{P}$  est une probabilité. Soit maintenant une famille dénombrable  $(B_n)_{n \geq 1}$  d'événements de  $\mathcal{E}$ , deux-à-deux disjoints. Alors, on a également que les événements  $(\{X \in B_n\})_{n \geq 1}$  de  $\mathcal{F}$  sont deux-à-deux disjoints, et que  $\{X \in \cup_{n \geq 1} B_n\} = \cup_{n \geq 1} \{X \in B_n\}$ . Ainsi,

$$\mathbb{P}^X(\cup_{n \geq 1} B_n) = \mathbb{P}(\{X \in \cup_{n \geq 1} B_n\}) = \mathbb{P}(\cup_{n \geq 1} \{X \in B_n\}) = \sum_{n \geq 1} \mathbb{P}(X \in B_n) = \sum_{n \geq 1} \mathbb{P}^X(B_n),$$

où dans l'avant dernière égalité, on a utilisé que  $\mathbb{P}$  est une probabilité. On a bien démontré tous les axiomes d'une probabilité.  $\square$

## 3.2 VARIABLES ALÉATOIRES DISCRÈTES

Dans toute la suite  $(\Omega, \mathcal{F}, \mathbb{P})$  est un espace probabilisé.

### 3.2.1 DÉFINITION

**Définition 3.4.** Soit  $X$  une application de  $\Omega$  dans  $E$ , où  $E = \{x_i\}_{i \in I}$  avec  $I$  dénombrable ou fini. On munit  $E$  de la tribu  $\mathcal{P}(E)$ . Alors, si  $X$  est une variable aléatoire de  $(\Omega, \mathcal{F})$  dans  $(E, \mathcal{P}(E))$ , elle est appelée *variable aléatoire discrète*.

D'après la Proposition 1.14 appliquée à l'espace d'arrivée, on obtient la caractérisation suivante.

**Proposition 3.5.** La loi d'une variable aléatoire discrète  $X$  à valeurs dans  $E = \{x_i\}_{i \in I}$  est caractérisée par la donnée de :

$$\forall i \in I, \quad \mathbb{P}^X(\{x_i\}) = \mathbb{P}(\{X = x_i\}) = \mathbb{P}(\{\omega \in \Omega : X(\omega) = x_i\}).$$

Pour tout  $B \in \mathcal{P}(E)$ , on a alors  $\mathbb{P}^X(B) = \sum_{x_i \in B} \mathbb{P}^X(x_i) = \mathbb{P}(X \in B)$ .

Avant de présenter les variables aléatoires discrètes classiques, voici deux premiers exemples.

EXEMPLE 3.6.

1. On considère deux lancers successifs d'un même dé, et on note  $S$  la variable aléatoire correspondant à la somme des valeurs obtenues, ainsi  $S$  prend ses valeurs dans  $\{2, 3, \dots, 12\}$ . Choisissons comme univers  $\Omega = \{1, \dots, 6\}^2$ , que l'on munit de la tribu  $\mathcal{F} = \mathcal{P}(\Omega)$ , et de la probabilité uniforme, notée  $\mathbb{P}$ . Ainsi pour tout sous-ensemble  $A$  de  $\mathcal{P}(\Omega)$ ,  $\mathbb{P}(A) = \frac{\text{card}(A)}{\text{card}(\Omega)}$ .

Étant donné que  $S$  prend ses valeurs dans  $\{2, \dots, 12\}$ , pour déterminer la loi de probabilité  $\mathbb{P}^S$  de  $S$ , il suffit de calculer, pour tout  $i \in \{2, \dots, 12\}$ ,  $\mathbb{P}^S(i)$  :

$$\mathbb{P}^S(2) = \mathbb{P}(S = 2) = \mathbb{P}(\{\omega \in \Omega : S(\omega) = 2\}) = \mathbb{P}((1, 1)) = \frac{1}{36},$$

$$\mathbb{P}^S(3) = \mathbb{P}(\{\omega \in \Omega : S(\omega) = 3\}) = \mathbb{P}(\{(1, 2), (2, 1)\}) = \frac{1}{18},$$

etc.

2. Soit  $(\Omega, \mathcal{F}, \mathbb{P})$  un espace probabilisé, et  $A \in \mathcal{F}$  un évènement. Alors l'indicateur de  $A$ , notée  $\mathbb{I}_A$ , est la variable aléatoire définie sur  $\Omega$  par :

$$\forall \omega \in \Omega, \mathbb{I}_A(\omega) = \begin{cases} 1 & \text{si } \omega \in A, \\ 0 & \text{sinon.} \end{cases}$$

Ainsi l'indicateur de  $A$  prend ses valeurs dans  $\{0, 1\}$ . Rappelons quelques propriétés de l'indicateur. Si  $A$  et  $B$  sont deux évènements de  $\mathcal{F}$ , alors :

$$\mathbb{I}_{A^c} = 1 - \mathbb{I}_A, \quad \mathbb{I}_{A \cap B} = \mathbb{I}_A \mathbb{I}_B, \quad \mathbb{I}_{A \cup B} = \mathbb{I}_A + \mathbb{I}_B - \mathbb{I}_{A \cap B}.$$

Étant donné que l'indicateur de  $A$  prend ses valeurs dans  $\{0, 1\}$ , il suffit de déterminer, pour  $i \in \{0, 1\}$ ,  $\mathbb{P}^{\mathbb{I}_A}(i)$  :

$$\mathbb{P}^{\mathbb{I}_A}(1) = \mathbb{P}(\mathbb{I}_A = 1) = \mathbb{P}(\{\omega \in \Omega : \mathbb{I}_A(\omega) = 1\}) = \mathbb{P}(\{\omega \in \Omega : \omega \in A\}) = \mathbb{P}(A),$$

$$\mathbb{P}^{\mathbb{I}_A}(0) = \mathbb{P}(\mathbb{I}_A = 0) = \mathbb{P}(\{\omega \in \Omega : \mathbb{I}_A(\omega) = 0\}) = \mathbb{P}(\{\omega \in \Omega : \omega \notin A\}) = \mathbb{P}(A^c) = 1 - \mathbb{P}(A).$$

### 3.2.2 VARIABLES ALÉATOIRES DISCRÈTES CLASSIQUES

Rappelons que la loi d'une variable aléatoire  $X$  discrète définie sur  $\Omega$  et à valeurs dans  $E = \{x_i\}_{i \in I}$  est caractérisée par la donnée des  $(\{x_i\}, \mathbb{P}^X(x_i))_{i \in I}$ . Souvent, on donne directement la loi  $\mathbb{P}^X$  d'une variable aléatoire sans passer par la probabilité  $\mathbb{P}$ . En utilisant à nouveau la Proposition 1.14 sur l'espace d'arrivée, pour s'assurer que la loi  $\mathbb{P}^X$  de  $X$  définit bien une probabilité, il suffit de s'assurer que

$$\forall i \in I, \quad 0 \leq \mathbb{P}^X(x_i) \leq 1, \quad \text{et} \quad \sum_{i \in I} \mathbb{P}^X(x_i) = 1.$$

Voici les variables aléatoires discrètes classiques, c'est-à-dire celles qui apparaissent naturellement dans de nombreuses expériences. Nous disons qu'une variable aléatoire  $X$  définie sur  $\Omega$  suit la :

**Loi uniforme discrète** sur  $\mathcal{X} = \{x_1, \dots, x_n\}$ , avec  $n \geq 1$ ,  $x_i \neq x_j$ ,  $i \neq j$ , notée  $\mathcal{U}(\mathcal{X})$ , si  $X(\Omega) = \mathcal{X}$  et

$$\forall i \in \{1, \dots, n\}, \quad \mathbb{P}^X(x_i) = \frac{1}{n}.$$

EXEMPLE 3.7. On lance une pièce équilibrée. Soit  $\Omega = \{P, F\}$  que l'on munit de la probabilité uniforme  $\mathbb{P}$ . On définit la variable aléatoire  $X : \Omega \rightarrow \{-1, 1\}$ , par  $X(P) = 1$ ,  $X(F) = -1$ . Ainsi, la loi de probabilité de  $X$  est :

$$\mathbb{P}^X(1) = \mathbb{P}(X = 1) = \mathbb{P}(P) = \frac{1}{2}, \quad \mathbb{P}^X(-1) = \mathbb{P}(X = -1) = \mathbb{P}(F) = \frac{1}{2},$$

et la variable aléatoire  $X$  suit une loi uniforme sur  $\{-1, 1\}$ .



**Loi de Bernoulli de paramètre  $p$**  avec  $p \in [0, 1]$ , notée  $\text{Ber}(p)$ , si  $X(\Omega) = \{0, 1\}$  et

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

EXEMPLE 3.8. Une *épreuve de Bernoulli de paramètre  $p$*  est une expérience avec une notion de succès/échec, et probabilité  $p$  de succès. Soit  $\Omega$  l'ensemble des issues de l'expérience,  $\mathbb{P}$  une probabilité sur  $(\Omega, \mathcal{F})$  et  $A$  l'évènement "obtenir un succès". Alors, on a  $\mathbb{P}(A) = p$ , et l'indicatrice de  $A$ ,  $\mathbb{I}_A$ , suit une loi de Bernoulli de paramètre  $p$ . En effet,  $\mathbb{I}_A$  est à valeurs dans  $\{0, 1\}$  et nous avons déjà vu que :

$$\mathbb{P}^{\mathbb{I}_A}(1) = \mathbb{P}(A) = p, \quad \mathbb{P}^{\mathbb{I}_A}(0) = 1 - \mathbb{P}(A).$$

Par exemple, on jette un dé équilibré une fois. On appelle succès l'évènement  $A$  "obtenir un nombre plus grand ou égal à 2". On choisit comme univers  $\Omega = \{1, \dots, 6\}$ , alors  $\mathbb{P}(A) = \mathbb{P}(\{2, 3, 4, 5, 6\}) = \frac{5}{6}$  et  $\mathbb{I}_A$  suit une loi de Bernoulli de paramètre  $5/6$ .

**Loi binomiale de paramètre  $(n, p)$** , avec  $n \in \mathbb{N}^*$  et  $p \in [0, 1]$ , notée  $\text{Bin}(n, p)$ , si  $X(\Omega) = \{0, \dots, n\}$  et

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

EXEMPLE 3.9. On appelle *schéma de Bernoulli de paramètres  $n$  et  $p$*  l'expérience qui consiste à répéter  $n$  fois de manière indépendante une épreuve de Bernoulli de paramètre  $p$ . On choisit comme univers  $\Omega^n$ , que l'on munit de tribu produit et de la probabilité produit  $\mathbb{P}_n$ . Pour  $i \in \{1, \dots, n\}$ , on note  $A_i$  l'évènement "obtenir un succès lors de la  $i$ -ième expérience". Soit  $k \in \{0, \dots, n\}$ , comme les expériences sont indépendantes, la probabilité d'obtenir un succès lors des  $k$  premières expériences et un échec lors des  $n - k$  dernières est :

$$\begin{aligned} \mathbb{P}_n(A_1 \cap \dots \cap A_k \cap A_{k+1}^c \cap \dots \cap A_n^c) &= \mathbb{P}(A_1) \dots \mathbb{P}(A_k) \mathbb{P}(A_{k+1}^c) \dots \mathbb{P}(A_n^c) \\ &= p^k (1 - p)^{n-k}. \end{aligned}$$

Soit  $X$  la variable aléatoire qui compte le nombre de succès dans un schéma de Bernoulli, alors  $X$  suit une loi binomiale de paramètres  $n$  et  $p$ . En effet,

$$\mathbb{P}^X(k) = \mathbb{P}_n(X = k) = \mathbb{P}_n(\{\text{obtenir } k \text{ succès sur les } n \text{ expériences}\}).$$

Il y a  $\binom{n}{k}$  façons d'obtenir ces succès, et pour chacune des façons la probabilité est  $p^k (1 - p)^{n-k}$ . Ainsi,

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

Par exemple, si on jette  $n$  fois un dé équilibré, et on appelle  $X$  la variable aléatoire qui compte le nombre de fois où l'on obtient un nombre plus grand ou égal à 2, alors  $X$  suit une loi binomiale de paramètres  $n$  et  $5/6$ .

*Remarque.* Nous avons vu que pour tout  $i \in \{1, \dots, n\}$ , la variable aléatoire  $\mathbb{I}_{A_i}$  suit une loi de Bernoulli de paramètre  $p$ . De plus,

$$X = \sum_{i=1}^n \mathbb{I}_{A_i}.$$

La variable aléatoire  $X$  est la somme de  $n$  variables aléatoires de Bernoulli indépendantes de paramètre  $p$  (nous verrons plus tard la définition de variables aléatoires indépendantes). En particulier, la somme de deux variables aléatoires indépendantes, de loi binomiale de paramètres  $n, p$  et  $m, p$  respectivement, est une variable aléatoire binomiale de paramètres  $n + m$  et  $p$ .

**Loi géométrique de paramètre  $p$** , avec  $p \in ]0, 1]$ , notée  $\mathcal{G}(p)$ , si  $X(\Omega) = \mathbb{N}^*$  et

$$\forall k \in \mathbb{N}^*, \quad \mathbb{P}^X(k) = (1-p)^{k-1}p.$$

En translatant de 1 les valeurs de la variable aléatoire, on obtient la loi géométrique sur  $\mathbb{N}$ .

**EXEMPLE 3.10.** Considérons un schéma de Bernoulli où l'expérience est répétée indéfiniment. On choisit comme univers  $\Omega^{\mathbb{N}^*}$ , que l'on munit de la tribu des cylindres et de la probabilité produit notée  $\mathbb{P}_\infty$ . Si  $X$  est la variable aléatoire égale au temps passé jusqu'au premier succès, alors  $X$  suit une loi géométrique de paramètre  $p$ . En effet, pour tout  $k \in \mathbb{N}^*$ ,

$$\begin{aligned} \mathbb{P}^X(k) &= \mathbb{P}_\infty(X = k) \\ &= \mathbb{P}_k(A_1^c \cap \dots \cap A_{k-1}^c \cap A_k) \\ &= \mathbb{P}(A_1^c) \dots \mathbb{P}(A_{k-1}^c) \mathbb{P}(A_k), \text{ par définition de la probabilité produit} \\ &= (1-p)^{k-1}p. \end{aligned}$$

**Loi de Poisson de paramètre  $\lambda$** , avec  $\lambda > 0$ , notée  $\mathcal{P}(\lambda)$ , si  $X(\Omega) = \mathbb{N}$  et

$$\forall k \in \mathbb{N}, \quad \mathbb{P}^X(k) = e^{-\lambda} \frac{\lambda^k}{k!}.$$

**EXEMPLE 3.11.** La loi de Poisson modélise le nombre d'autobus passés à un arrêt avant un instant  $T$  donné, et  $\lambda$  représente le nombre moyen d'arrivées dans cet intervalle.

*Remarque.* Lorsque  $p$ ,  $n$  et  $np_n$  tendent vers 0,  $+\infty$  et  $\lambda$  respectivement, la loi binomiale de paramètres  $n$  et  $p$  tend vers une loi de Poisson de paramètre  $\lambda$ . Plus précisément, on a la proposition suivante.

**Proposition 3.12.** Soit  $(X_n)_{n \geq 1}$  une suite de variables aléatoires. On suppose que, pour tout entier  $n \geq 1$ , la variable aléatoire  $X_n$  suit une loi binomiale de paramètres  $n$  et  $p_n$ , et que  $np_n$  tende vers un nombre réel strictement positif  $\lambda$  lorsque  $n$  tend vers  $+\infty$ . Alors, pour tout entier naturel  $k$  :

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

*Démonstration.* Soit  $k$  un entier naturel. Alors, pour tout entier  $n \geq k$ , on a :

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

De  $np_n = \lambda + o(1)$ , on déduit que  $p_n = \frac{\lambda}{n} + o\left(\frac{1}{n}\right)$ , puis :

- i)  $\lim_{n \rightarrow +\infty} p_n = 0$  et  $\lim_{n \rightarrow +\infty} (1-p_n)^{-k} = 1$  ;
- ii)  $\lim_{n \rightarrow +\infty} (np_n)^k = \lambda^k$  ;
- iii)  $n \ln(1-p_n) = n \ln\left(1 - \frac{\lambda}{n} + o\left(\frac{1}{n}\right)\right) = n \left(-\frac{\lambda}{n} + o\left(\frac{1}{n}\right)\right) = -\lambda + o(1)$  et  $\lim_{n \rightarrow +\infty} n \ln(1-p_n) = -\lambda$  ;
- iv) par continuité de l'application exponentielle,  $\lim_{n \rightarrow +\infty} e^{n \ln(1-p_n)} = e^{-\lambda}$ .

Il s'ensuit que

$$\mathbb{P}^{X_n}(\{k\}) \xrightarrow{n \rightarrow +\infty} \frac{1}{k!} \cdot \lambda^k \cdot 1 \cdot e^{-\lambda},$$

d'où le résultat. □

### 3.3 VARIABLES ALÉATOIRES RÉELLES

#### 3.3.1 DÉFINITION

**Définition 3.13.** Soit  $X$  une variable aléatoire de  $(\Omega, \mathcal{F})$  dans  $(E, \mathcal{E})$ . Si  $E = \mathbb{R}$  et  $\mathcal{E} = \mathcal{B}(\mathbb{R})$  est la tribu borélienne, alors  $X$  est appelée une *variable aléatoire réelle*.

*Remarque 3.14.*

- On obtient la définition de la loi de la variable aléatoire réelle  $X$  comme cas particulier de la Définition 3.3.
- Pour montrer qu'une application  $X : \Omega \rightarrow \mathbb{R}$  est une variable aléatoire réelle, il suffit de montrer que, pour tout  $x \in \mathbb{R}$ ,  $X^{-1}(]-\infty, x]) \in \mathcal{F}$ . En effet, supposons que ce soit le cas, notons  $\mathcal{G} = \{B \in \mathcal{B}(\mathbb{R}) : X^{-1}(B) \in \mathcal{F}\}$  et montrons que  $\mathcal{G} = \mathcal{B}(\mathbb{R})$ . Par définition,  $\mathcal{G} \subset \mathcal{B}(\mathbb{R})$ . Comme l'image réciproque commute avec la complémentation et la réunion, et que  $\mathcal{G}$  contient  $\mathbb{R}$ , on déduit de plus que  $\mathcal{G}$  est une tribu. Par hypothèse,  $\mathcal{G}$  contient  $\{]-\infty, x] : x \in \mathbb{R}\}$ , ainsi  $\mathcal{G}$  est une tribu contenant les intervalles  $\{]-\infty, x] : x \in \mathbb{R}\}$ . D'après la Remarque 1.8,  $\mathcal{B}(\mathbb{R})$  est la plus petite tribu contenant ces intervalles, d'où  $\mathcal{B}(\mathbb{R}) \subset \mathcal{G}$  et par suite on a  $\mathcal{G} = \mathcal{B}(\mathbb{R})$ .

**Proposition 3.15.** Si  $X_1, \dots, X_n$  sont des variables aléatoires réelles, et  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  est une fonction continue, alors  $Y = f(X_1, \dots, X_n)$  est une variable aléatoire réelle.

*Démonstration.* D'après la Remarque 3.14, il suffit de montrer que, pour tout  $x \in \mathbb{R}$ ,  $Y^{-1}(]-\infty, x]) \in \mathcal{F}$  ou de manière équivalente  $Y^{-1}(]x, \infty[) \in \mathcal{F}$ . On a,

$$Y^{-1}(]x, \infty[) = \{\omega \in \Omega : f(X_1, \dots, X_n)(\omega) > x\}.$$

Or, comme l'application  $f$  est continue, on sait d'après le cours de topologie, que l'image réciproque de l'intervalle ouvert  $]x, \infty[$  s'écrit comme réunion dénombrable de rectangles ouverts de  $\mathbb{R}^n$ ; ainsi,  $Y^{-1}(]x, \infty[) = \cup_{i \in \mathbb{N}} A_i$ , avec  $A_i = \cap_{j=1}^n \{\omega \in \Omega : x_{i,j} < X_j(\omega) < y_{i,j}\}$ . Or, comme pour tout  $j$ ,  $X_j$  est une variable aléatoire, on sait que l'image réciproque de chacun des intervalles ouverts  $]x_{i,j}, y_{i,j}[$  est un élément de  $\mathcal{F}$ . En conclusion,  $Y^{-1}(]x, \infty[)$  s'écrit comme réunion dénombrable d'intersections finies d'éléments de  $\mathcal{F}$  et appartient donc à  $\mathcal{F}$ .  $\square$

En conséquence de la Proposition 3.15 (Point 1.) ou de la Remarque 3.14 (Point 2. et 3.) nous avons les résultats suivants.

**Corollaire 3.16.** Soient  $X, Y$  et  $(X_n)_{n \geq 1}$  des variables aléatoires réelles. Alors, les applications suivantes sont des variables aléatoires réelles :

1.  $X + Y, XY, \frac{X}{Y}$  si  $Y \neq 0$ ,
2.  $\sup_{1 \leq n \leq k} X_n, \inf_{1 \leq n \leq k} X_n, \sup_{n \geq 1} X_n, \inf_{n \geq 1} X_n$ ,
3.  $\limsup_{n \geq 1} X_n, \liminf_{n \geq 1} X_n$ , ainsi que  $\lim_{n \rightarrow \infty} X_n$  lorsque cette variable aléatoire est bien définie.

*Démonstration.* La preuve des Points 2. et 3. sera faite en exercices.  $\square$

**3.3.2 FONCTION DE RÉPARTITION**

Spécifier  $\mathbb{P}^X$  pour tous les boréliens de  $\mathcal{B}(\mathbb{R})$  serait en pratique fastidieux. Comme nous allons le voir, il suffit en fait de connaître la loi  $\mathbb{P}^X$  sur tous les intervalles de la forme  $] - \infty, x]$  avec  $x \in \mathbb{R}$ . Ceci motive la définition suivante.

**Définition 3.17.** Soit  $X$  une variable aléatoire réelle. On appelle *fonction de répartition de  $X$* , la fonction  $F^X : \mathbb{R} \rightarrow [0, 1]$  définie par,

$$\forall x \in \mathbb{R}, \quad F^X(x) = \mathbb{P}^X(]-\infty, x]) = \mathbb{P}(X \leq x).$$

Voici des propriétés importantes de la fonction de répartition.

**Proposition 3.18** (Caractérisation de la loi). *La fonction de répartition  $F^X$  caractérise la loi de probabilité  $\mathbb{P}^X$  et elle vérifie les trois conditions suivantes :*

1. *elle est croissante,*
2. *elle est continue à droite,*
3.  $\lim_{x \rightarrow -\infty} F^X(x) = 0, \quad \lim_{x \rightarrow \infty} F^X(x) = 1.$

*Démonstration.* La première partie du résultat signifie que deux variables aléatoires ont même loi si et seulement si elles ont même fonction de répartition. Un sens est facile. L'autre sens dépasse le cadre de ce cours, mais a été mentionné à la Section 1.4 : le fait qu'une probabilité définie sur  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$  soit caractérisée par sa valeur sur les intervalles de la forme  $] - \infty, x]$ ,  $x \in \mathbb{R}$ , est une conséquence du lemme de classe monotone (cf cours à venir de "théorie de la mesure").

Notons  $F^X := F$  et démontrons maintenant le Point 1. Si  $x \leq y$ , alors  $] - \infty, x] \subset ] - \infty, y]$  et d'après le Corollaire 1.12, on a  $F(x) = \mathbb{P}^X(]-\infty, x]) \leq \mathbb{P}^X(]-\infty, y]) = F(y)$ . Pour le Point 2. on considère une suite  $(x_n)_{n \geq 1}$  qui décroît vers  $x$ . Alors la suite d'événements  $(]-\infty, x_n])_{n \geq 1}$  est décroissante et  $\bigcap_{n \geq 1} ] - \infty, x_n] = ] - \infty, x]$ . Ainsi, par le Point 3. de la Proposition 1.13, on a  $F(x) = \mathbb{P}^X(]-\infty, x]) = \lim_{n \rightarrow \infty} \mathbb{P}^X(]-\infty, x_n]) = \lim_{n \rightarrow \infty} F(x_n)$ . Pour le Point 3., montrons la deuxième égalité. Soit  $(x_n)_{n \geq 1}$  une suite croissante qui tend vers  $\infty$ . la suite d'événements  $(]-\infty, x_n])_{n \geq 1}$  est croissante et  $\bigcup_{n \geq 1} ] - \infty, x_n] = ] - \infty, \infty[$ . Ainsi, par le Point 2. de la Proposition 1.13, on a  $1 = \mathbb{P}^X(]-\infty, \infty[) = \lim_{n \rightarrow \infty} \mathbb{P}^X(]-\infty, x_n]) = \lim_{n \rightarrow \infty} F(x_n)$ . L'autre égalité se montre de manière analogue avec une suite décroissante qui tend vers  $-\infty$ . A noter que pour les Points 2. et 3., il suffit de considérer des suites croissantes/décroissantes car la fonction  $F$  est croissante.  $\square$

Voici d'autres propriétés de la fonction de répartition, qui seront démontrées en exercice.

**Proposition 3.19.** *Soit  $x \in \mathbb{R}$  et  $y \in \mathbb{R}$  tel que  $x < y$ . Alors*

1. *La fonction de répartition  $F^X$  admet une limite à gauche en tout point  $x$ . Notons  $F^X(x_-)$  cette limite.*
2.  $\mathbb{P}(X > x) = 1 - F^X(x),$
3.  $\mathbb{P}(x < X \leq y) = F^X(y) - F^X(x),$
4.  $\mathbb{P}(x < X < y) = F^X(y_-) - F^X(x),$
5.  $\mathbb{P}(x \leq X \leq y) = F^X(y) - F^X(x_-),$
6.  $\mathbb{P}(x \leq X < y) = F^X(y_-) - F^X(x_-)$

$$7. \mathbb{P}(X = x) = F^X(x) - F^X(x_-).$$

Comme conséquence du Point 7., nous obtenons la caractérisation suivante de la continuité de  $F^X$ .

**Corollaire 3.20.**

$$\mathbb{P}^X(x) = \mathbb{P}(X = x) = 0 \Leftrightarrow F^X \text{ est continue en } x.$$

Le théorème suivant va plus loin et caractérise les fonctions qui sont effectivement fonction de répartition d'une variable aléatoire. Sa démonstration qui nécessite la preuve de l'existence d'une mesure de probabilité est un résultat subtil de théorie de la mesure. Nous énonçons ce résultat sans donner de preuve.

**Théorème 3.21.** *Soit  $F$  une fonction définie sur  $\mathbb{R}$  qui satisfait les propriétés 1. à 3. de la Proposition 3.18. Alors, c'est la fonction de répartition d'une loi  $\mu$  sur  $\mathbb{R}$  muni de sa tribu borélienne  $\mathcal{B}(\mathbb{R})$ . On ne peut pas en général définir  $\mu$  sur la tribu  $\mathcal{P}(\mathbb{R})$  de toutes les parties de  $\mathbb{R}$ .*

### 3.3.3 VARIABLES ALÉATOIRES DISCRÈTES RÉELLES

Considérons le cas d'une variable aléatoire  $X$  discrète à valeurs réelles. La variable aléatoire  $X$  étant discrète, dans la Section 3.2, nous avons considéré l'espace d'arrivée  $E = X(\Omega)$  muni de la tribu  $\mathcal{P}(E)$ . On peut montrer que  $X$  est aussi une variable aléatoire à valeurs dans  $\mathbb{R}$  muni de la tribu  $\mathcal{B}(\mathbb{R})$ . Ainsi, une variable aléatoire discrète réelle est bien une variable aléatoire réelle au sens où nous l'avons défini.

Comme conséquence de sa définition, la fonction de répartition satisfait aux propriétés suivantes.

**Proposition 3.22.** *la fonction de répartition  $F^X$  d'une variable aléatoire  $X$  discrète réelle vérifie :*

1.  $F^X(x) = \sum_{\{y \in X(\Omega) : y \leq x\}} \mathbb{P}^X(y),$
2. si  $x$  et  $y$  sont deux points consécutifs de  $X(\Omega)$ , alors  $F^X$  est constante sur  $[x, y[$ ,
3. La hauteur du saut en  $x \in X(\Omega)$  est donnée par  $\mathbb{P}^X(x)$ .

**EXEMPLE 3.23.** En exercices, vous calculerez la fonction de répartition des lois discrètes classiques : uniforme sur  $\{1, \dots, n\}$ , Bernoulli de paramètre  $p$ , binomiale de paramètres  $(n, p)$ , géométrique de paramètre  $p$ .

### 3.3.4 VARIABLES ALÉATOIRES À DENSITÉ

**PRÉAMBULE.** Dans la suite, nous allons considérer des intégrales de fonctions sur  $\mathbb{R}$  ou des intervalles de  $\mathbb{R}$ . Nous parlerons aussi de *fonctions intégrables*, c'est-à-dire de fonctions intégrables au sens de Riemann ou de Lebesgue (que vous verrez dans le cours d'intégration — une fonction intégrable au sens de Riemann est intégrable au sens de Lebesgue, la réciproque est fautive en général). Une *fonction intégrable*  $f$  est une fonction telle que l'intégrale généralisée  $\int_{-\infty}^{\infty} |f(x)| dx$  est finie, ce qui implique que l'intégrale généralisée  $\int_{-\infty}^{\infty} f(x) dx$  existe.

**Définition 3.24.**

- Une fonction réelle  $f$  définie sur  $\mathbb{R}$  est une *densité de probabilités*, ou simplement *densité*, si elle est positive, intégrable et vérifie

$$\int_{-\infty}^{\infty} f(x) dx = 1.$$

- Si  $f$  est comme ci-dessus, la fonction  $F$  définie par,

$$\forall x \in \mathbb{R}, \quad F(x) = \int_{-\infty}^x f(y) dy, \quad (3.1)$$

satisfait aux Point 1. à 3. de la Proposition 3.18. D'après le Théorème 3.21, il s'agit donc de la fonction de répartition d'une variable aléatoire  $X$ .

- Une variable aléatoire  $X$  dont la fonction de répartition  $F^X$  est de la forme (3.1) est dite *variable aléatoire à densité*, de densité  $f$ .

**Proposition 3.25.** *Soit  $X$  une variable aléatoire réelle. Pour les Points 1., 2. et 4., nous supposons que  $X$  est à densité  $f$ . Nous avons alors les propriétés suivantes :*

1. la fonction de répartition  $F^X$  est continue sur  $\mathbb{R}$ , de sorte que

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

2. La fonction de répartition  $F^X$  est dérivable partout où  $f$  est continue et, en ces points,  $(F^X)'(x) = f(x)$ .
3. Réciproquement, si la fonction de répartition  $F^X$  est continue et dérivable par morceaux, alors  $X$  est une variable aléatoire à densité, de densité égale à  $(F^X)'$  partout où elle est dérivable.
4. Pour tout intervalle  $J$  de  $\mathbb{R}$ , non réduit à un point :

$$\mathbb{P}(X \in J) = \int_{-\infty}^{+\infty} \mathbb{I}_J(x) f(x) dx = \int_J f(x) dx.$$

*Démonstration.* Les trois premiers points découlent de résultats généraux d'analyse et de la définition d'une densité. Pour le quatrième, il suffit de retourner à la définition de fonction de répartition.  $\square$

*Remarque 3.26.*

1. On parle de la densité de la loi d'une variable aléatoire, alors qu'on devrait parler d'une densité. En effet, si par exemple on modifie  $f$  en un nombre fini de points en lui donnant d'autres valeurs positives, on obtient une autre densité définissant la même variable aléatoire.
2. Il existe des variables aléatoires qui ne sont pas à densité, telles que les variables aléatoires discrètes par exemple.

Donnons maintenant les exemples classiques de variables aléatoires à densité. Nous disons qu'une variable aléatoire réelle  $X$  définie sur  $\Omega$  suit la :

**Loi uniforme** sur  $I = [a, b]$ , avec  $a, b \in \mathbb{R}$ ,  $a < b$ , notée  $\mathcal{U}([a, b])$ , si elle admet pour densité

$$\forall x \in \mathbb{R}, \quad f(x) = \frac{1}{b-a} \mathbb{I}_{[a,b]}(x) = \begin{cases} \frac{1}{b-a} & \text{si } a \leq x \leq b \\ 0 & \text{sinon.} \end{cases}$$

Notons qu'en vertu de la Remarque 3.26, on peut définir  $f(a) = 0$  et/ou  $f(b) = 0$  et obtenir une autre densité définissant la même variable aléatoire.

**Loi gaussienne ou loi normale de moyenne  $\mu$  et de variance  $\sigma^2$** , notée  $\mathcal{N}(\mu, \sigma^2)$ , si elle admet pour densité

$$\forall x \in \mathbb{R}, \quad f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

Si  $\mu = 0$  et  $\sigma^2 = 1$ , on parle de loi gaussienne *centrée-réduite*. Cette terminologie deviendra claire lorsqu'on aura calculé sa moyenne et sa variance.

*Remarque.*

1. Vous montrerez en exercices que si  $X$  suit une loi normale  $\mathcal{N}(0, 1)$ , alors  $Y = \sigma X + \mu$  suit une loi  $\mathcal{N}(\mu, \sigma^2)$ . De manière analogue, si  $Y$  suit une loi  $\mathcal{N}(\mu, \sigma^2)$ , alors  $X = \frac{Y-\mu}{\sigma}$  suit une loi  $\mathcal{N}(0, 1)$ .
2. La fonction de répartition de  $X$  ne se calcule pas à l'aide des fonctions usuelles. Si on note  $\Pi$  la fonction de répartition de la loi gaussienne centrée réduite, des tables donnent des valeurs approchées de  $\Pi(x)$  pour  $x > 0$ . En remarquant que :  $\Pi(x) + \Pi(-x) = 1$ , on peut en déduire des valeurs approchées de  $\Pi(x)$  pour les valeurs négatives de  $x$ .

**Loi exponentielle de paramètre  $\lambda$ ,  $\lambda > 0$** , notée  $\mathcal{E}(\lambda)$ , si elle admet pour densité

$$\forall x \in \mathbb{R}, \quad f(x) = \mathbb{I}_{[0, +\infty[}(x) \lambda e^{-\lambda x} = \begin{cases} \lambda e^{-\lambda x} & \text{si } x \geq 0, \\ 0 & \text{si } x < 0. \end{cases}$$

Dans la pratique, la loi exponentielle modélise souvent une durée de vie ou le temps d'attente avant l'arrivée d'un événement spécifique. Elle a la propriété d'être sans mémoire.

**Loi de Cauchy de paramètre  $\lambda$ ,  $\lambda > 0$** , notée  $\mathcal{C}(\lambda)$ , si elle admet pour densité

$$\forall x \in \mathbb{R}, \quad f(x) = \frac{\lambda}{\pi(\lambda^2 + x^2)}.$$

La loi de Cauchy apparaît comme la loi d'une variable aléatoire qui est le quotient de variables aléatoires gaussiennes centrées indépendantes, de même variance. L'inverse d'une variable aléatoire qui suit une loi de Cauchy suit également une loi de Cauchy.

**Loi du  $\chi^2$  à  $n$  degrés de libertés**, avec  $n \in \mathbb{N}^*$ , notée  $\chi^2(n)$ , si elle admet pour densité l'application  $f$  définie par :

$$\forall x \in \mathbb{R}, \quad f(x) = \frac{1}{2^{\frac{n}{2}} \Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} e^{-\frac{x}{2}} \mathbb{I}_{[0, \infty[}(x).$$

La loi du  $\chi^2$  est la loi de la somme des carrés de  $n$  variables aléatoires gaussiennes centrées réduites indépendantes.

**Loi gamma de paramètres  $(\alpha, \lambda)$** , avec  $\alpha, \lambda > 0$ , notée  $\Gamma(\alpha, \lambda)$ , si elle admet pour densité l'application  $f$  définie par :

$$\forall x \in \mathbb{R}, \quad f(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} \mathbb{I}_{[0, \infty[}(x),$$

où la fonction  $\Gamma : ]0, \infty[ \rightarrow ]0, \infty[$  est définie par :

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

Notons que lorsque  $\alpha = 1$ , on retrouve la loi exponentielle, et lorsque  $\alpha = \frac{n}{2}$  et  $\lambda = \frac{1}{2}$ , on retrouve la loi du  $\chi^2$  à  $n$  degrés de liberté.





# CHAPITRE 4

## ESPÉRANCE, VARIANCE & INÉGALITÉS

### 4.1 ESPÉRANCE

Dans cette section, nous allons voir une définition très importante qui est celle de l'*espérance* d'une variable aléatoire réelle. Nous allons nous échauffer avec le cas où l'espace des états est dénombrable ou fini. Sans le préciser à chaque fois, toutes les variables aléatoires considérées dans ce chapitre seront *réelles*.

#### 4.1.1 CAS D'UN ESPACE DES ÉTATS FINI OU DÉNOMBRABLE

Intuitivement, l'espérance d'une variable aléatoire est la moyenne de  $X$ . Pour calculer cette moyenne, on pondère les valeurs prises par  $X$  par leur chance d'arriver, c'est-à-dire par leur probabilité.

Soit  $(\Omega, \mathcal{F}, \mathbb{P})$  un espace probabilisé. On suppose que  $\Omega$  est fini ou dénombrable, et pour tout  $\omega \in \Omega$ , on note  $p_\omega = \mathbb{P}(\{\omega\})$ . Soit  $X$  une variable aléatoire définie sur  $\Omega$ ; on note  $X(\Omega) = \{x_i\}_{i \in I}$  avec  $I$  fini ou dénombrable.

**Définition 4.1.** Si la somme  $\sum_{\omega \in \Omega} p_\omega |X(\omega)|$  est finie, alors l'*espérance* de  $X$ , ou *moyenne* de  $X$ , est définie par

$$\mathbb{E}(X) = \sum_{\omega \in \Omega} p_\omega X(\omega).$$

Le théorème suivant, bien que simple à démontrer, est important car il permet de passer de l'espace de départ  $(\Omega, \mathcal{F}, \mathbb{P})$ , qui est souvent abstrait et inconnu, à l'espace d'arrivée, qui dans ce cas est un sous-ensemble fini ou dénombrable de  $\mathbb{R}$ . Ce résultat montre qu'en fait l'espérance ne dépend que de la loi de la variable aléatoire  $X$ .

**Théorème 4.2.** Si  $\sum_{\omega \in \Omega} p_\omega |X(\omega)| < \infty$ . Alors,

$$\mathbb{E}(X) = \sum_{\omega \in \Omega} p_\omega X(\omega) = \sum_{i \in I} x_i \mathbb{P}^X(x_i) = \sum_{i \in I} x_i \mathbb{P}(X = x_i).$$

*Démonstration.* La preuve consiste à décomposer la somme et à sommer par paquets.

$$\begin{aligned}\mathbb{E}(X) &= \sum_{\omega \in \Omega} p_{\omega} X(\omega) = \sum_{i \in I} \sum_{\{\omega \in \Omega : X(\omega) = x_i\}} p_{\omega} X(\omega) \\ &= \sum_{i \in I} x_i \left( \sum_{\{\omega \in \Omega : X(\omega) = x_i\}} p_{\omega} \right) = \sum_{i \in I} x_i \mathbb{P}(X = x_i). \quad \square\end{aligned}$$

#### 4.1.2 CAS DES VARIABLES ALÉATOIRES RÉELLES

Nous allons maintenant voir le cas général des variables aléatoires réelles. A noter que certains points dépassent le cadre du programme et seront vus de manière complète dans le cours de théorie de la mesure.

Un premier point à noter est que la partie de droite du Théorème 4.2 a un sens si l'espace d'arrivée est fini, indépendamment de ce qui se passe pour l'espace de départ. L'idée est donc de définir l'espérance pour les variables aléatoires étagées en utilisant le Théorème 4.2, puis d'utiliser ceci pour définir l'espérance de variables aléatoires *positives* et finalement passer aux variables aléatoires *intégrables*.

**Définition 4.3.** Une variable aléatoire  $X$  est dite *étagée*, si elle ne prend qu'un nombre fini de valeurs  $x_1, \dots, x_m$ . Pour tout  $i \in \{1, \dots, m\}$ , notons  $A_i = X^{-1}(\{x_i\}) = \{X = x_i\}$ , alors la variable aléatoire  $X$  s'écrit :

$$X = \sum_{i=1}^m x_i \mathbb{I}_{A_i}.$$

**Définition 4.4.**

- **Variables aléatoires étagées.** Si  $X$  est une variable aléatoire étagée prenant les valeurs  $\{x_1, \dots, x_m\}$ , alors son *espérance* est définie par :

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

- **Variables aléatoires positives.** Si  $X$  est une variable aléatoire positive, alors son *espérance* est définie par :

$$\mathbb{E}(X) = \sup\{\mathbb{E}(Z) : Z \text{ étagée positive, } Z \leq X\}.$$

Cette quantité est toujours positive, mais peut être égale à l'infini !

- **Variables aléatoires intégrables.** Soit  $X$  une variable aléatoire. On écrit alors

$$X = X^+ - X^-,$$

où  $X^+ = \max\{0, X\}$ ,  $X^- = \max\{0, -X\}$ . Comme  $X^+$  et  $X^-$  sont des variables aléatoires positives, leur espérance est bien définie. On dit que  $X$  est *intégrable* si  $\mathbb{E}(X^+) < \infty$  et  $\mathbb{E}(X^-) < \infty$ . A noter que ceci est équivalent à  $\mathbb{E}(|X|) < \infty$  car  $|X| = X^+ + X^-$ .

Si  $X$  est une variable aléatoire *intégrable*, alors son *espérance* est définie par :

$$\mathbb{E}(X) = \mathbb{E}(X^+) - \mathbb{E}(X^-).$$

L'ensemble des variables aléatoires intégrables est noté  $L^1$  ou  $L^1(\Omega, \mathcal{F}, \mathbb{P})$ .

**Définition 4.5.** Soit  $X$  une variable aléatoire positive ou intégrable. Alors  $X$  est dite *centrée* si  $\mathbb{E}(X) = 0$ .

### 4.1.3 PROPRIÉTÉS FONDAMENTALES

Voici les propriétés fondamentales de l'espérance. Sauf la première propriété qui est une conséquence directe de la définition de l'espérance pour les variables aléatoires étagées, nous ne démontrons pas cette proposition dans ce cours, mais vous verrez la preuve dans le cours de mesure et intégration.

**Proposition 4.6.**

1. **Espérance d'une indicatrice.** Pour tout événement  $A \in \mathcal{F}$ , on a

$$\mathbb{E}(\mathbb{I}_A) = \mathbb{P}(A).$$

2. **Linéarité.** Soit  $X, Y$  deux variables aléatoires intégrables, et soit  $a, b \in \mathbb{R}$ . Alors  $aX + bY$  est une variable aléatoire intégrable, et on a

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

Autrement dit, l'ensemble  $L^1(\Omega, \mathcal{F}, \mathbb{P})$  est un espace vectoriel. La propriété ci-dessus reste vraie si  $a, b \geq 0$  et  $X$  et  $Y$  sont positives.

3. **Monotonie.** Soient  $X, Y$  deux variables aléatoires positives ou intégrables, alors

$$X \leq Y \quad p.s. \implies \mathbb{E}(X) \leq \mathbb{E}(Y),$$

ou, de manière équivalente,

$$X \geq 0 \quad p.s. \implies \mathbb{E}(X) \geq 0.$$

4. **Valeur absolue.** Soit  $X$  une variable aléatoire intégrable, alors

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

Les deux propriétés suivantes sont également fondamentales.

5. **Théorème de convergence monotone.** Soit  $(X_n)_{n \geq 1}$  une suite de variables aléatoires, telles que :

- la suite  $(X_n)_{n \geq 1}$  est croissante et positive p.s.
- $(X_n)_{n \geq 1}$  converge vers  $X$  p.s.

Alors,

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

6. **Théorème de convergence dominée.** Soit  $(X_n)_{n \geq 1}$  une suite de variables aléatoires, telles que :

- $(X_n)_{n \geq 1}$  converge vers  $X$  p.s.
- il existe une variable aléatoire  $Y$ , intégrable, telle que, pour tout  $n \geq 1$ ,  $|X_n| \leq Y$  p.s.

Alors  $X$  est intégrable et

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

*Remarque 4.7.* Soit  $X$  une variable aléatoire positive. Alors, la suite de variables aléatoires  $(X_n)_{n \geq 1}$  suivante, positive et croissante converge simplement vers  $X$  :

$$\forall n \geq 1, \quad X_n(\omega) = \sum_{k=0}^{n2^n-1} \frac{k}{2^n} \mathbb{I}_{\{X \in [\frac{k}{2^n}, \frac{k+1}{2^n}[ \}}(\omega).$$

Ainsi, d'après le théorème de convergence monotone, on a

$$\begin{aligned} \mathbb{E}(X) &= \lim_{n \rightarrow \infty} \mathbb{E}(X_n) \\ &= \lim_{n \rightarrow \infty} \sum_{k=0}^{n2^n-1} \frac{k}{2^n} \mathbb{P}^X \left( \left[ \frac{k}{2^n}, \frac{k+1}{2^n} \right] \right), \end{aligned}$$

où dans la deuxième égalité, on a utilisé la définition de l'espérance pour les variables aléatoires étagées.

Si  $X$  est une variable aléatoire intégrable, on peut appliquer le raisonnement ci-dessus à  $X^+$  et  $X^-$  et conclure que *la loi d'une variable aléatoire caractérise son espérance* si elle existe. La réciproque est cependant fausse. Pour avoir une réciproque, il faut connaître l'espérance de toute fonction raisonnable de la variable aléatoire  $X$ , tel qu'énoncé dans le théorème suivant.

**Théorème 4.8.** *Soient  $X, Y$  deux variables aléatoires. Alors, les conditions suivantes sont équivalentes.*

1. *Les variables aléatoires  $X$  et  $Y$  ont même loi.*
2. *Pour tout fonction  $h : \mathbb{R} \rightarrow \mathbb{R}$ , continue et bornée*

$$\mathbb{E}(h(X)) = \mathbb{E}(h(Y)).$$

*Démonstration. (Idée).*

- Implication 1.  $\Rightarrow$  2. Pour tout borélien  $B$  de  $\mathbb{R}$ , on a

$$\mathbb{P}^{h(X)}(B) = \mathbb{P}(h(X) \in B) = \mathbb{P}((h \circ X)^{-1}(B)) = \mathbb{P}(X^{-1}(h^{-1}(B))) = \mathbb{P}^X(h^{-1}(B)).$$

Ainsi, si  $X$  et  $Y$  ont même loi, il en est de même pour  $h(X)$  et  $h(Y)$ . D'après la Remarque 4.7, ceci implique que  $\mathbb{E}(h(X)) = \mathbb{E}(h(Y))$ .

- L'idée est de montrer que si la Condition 2. est satisfaite, les variables aléatoires  $X$  et  $Y$  ont la même fonction de répartition et de conclure avec la Proposition 3.18, qui dit que la fonction de répartition caractérise la loi. Pour cela, on a envie d'utiliser la Condition 2. avec la fonction  $h := \mathbb{I}_{]-\infty, x]}$ , car alors,  $\mathbb{E}(h(X)) = \mathbb{E}(\mathbb{I}_{]-\infty, x]}(X)) = \mathbb{P}(X \in ]-\infty, x]) = \mathbb{P}(X \leq x) = F^X(x)$ . La difficulté vient du fait que la fonction  $h = \mathbb{I}_{]-\infty, x]}$  n'est pas continue (mais elle bien bornée). Pour palier à cela, on approche  $h$  par une suite de fonctions continues bornées et on utilise le théorème de convergence dominée pour passer à la limite dans l'espérance.  $\square$

*Remarque 4.9.*

1. L'implication 1.  $\Rightarrow$  2. reste vraie dès que la fonction  $h$  est telle que  $h(X)$  est intégrable.
2. Dans le Point 2. du Théorème 4.8, on peut remplacer l'ensemble des fonctions continues bornées par l'ensemble des fonctions boréliennes positives par exemple.

## 4.1.4 CAS DES VARIABLES ALÉATOIRES DISCRÈTES ET À DENSITÉ

Dans le cas des variables aléatoires discrètes et à densité, l'espérance se calcule de manière explicite et cette section est très utilisée en pratique.

**Théorème 4.10** (Théorème de transfert).

- **Variables aléatoires discrètes.** Soit  $X$  une variable aléatoire discrète réelle à valeurs dans  $E = X(\Omega) = \{x_i\}_{i \in I}$ , avec  $I$  fini ou dénombrable. Soit  $h : E = X(\Omega) \rightarrow \mathbb{R}$  une application telle que  $h(X)$  est positive ou intégrable, c'est-à-dire telle que  $\sum_{i \in I} |h(x_i)| \mathbb{P}^X(x_i) < \infty$ , alors :

$$\mathbb{E}(h(X)) = \sum_{i \in I} h(x_i) \mathbb{P}^X(x_i) = \sum_{i \in I} h(x_i) \mathbb{P}(X = x_i).$$

- **Variables aléatoires à densité.** Soit  $X$  une variable aléatoire réelle à densité, et soit  $h : E = X(\Omega) \rightarrow \mathbb{R}$  une application telle que  $h(X)$  est positive ou intégrable, c'est-à-dire telle que  $\int_{-\infty}^{\infty} |h(x)| f(x) dx < \infty$ , alors :

$$\mathbb{E}(h(X)) = \int_{-\infty}^{\infty} h(x) f(x) dx.$$

*Démonstration.* Démontrons ce résultat dans le cas d'une variable aléatoire  $X$  discrète ne prenant qu'un nombre fini de valeurs  $E := \{x_i\}_{i \in I}$ , avec  $I$  fini. La variable aléatoire  $Y = h(X)$  est également discrète et ne prend qu'un nombre fini de valeurs  $h(E) := F = \{y_j\}_{j \in J}$ , avec  $J$  fini. Alors, par définition de l'espérance pour les variables aléatoires étagées, on a :

$$\begin{aligned} \mathbb{E}(Y) &= \sum_{j \in J} y_j \mathbb{P}(Y = y_j) = \sum_{j \in J} y_j \mathbb{P}(h(X) = y_j) = \sum_{j \in J} y_j \mathbb{P}(X \in h^{-1}(y_j)) \\ &= \sum_{j \in J} y_j \sum_{\{i \in I : h(x_i) = y_j\}} \mathbb{P}(X = x_i) = \sum_{j \in J} \sum_{\{i \in I : h(x_i) = y_j\}} y_j \mathbb{P}(X = x_i) \\ &= \sum_{j \in J} \sum_{\{i \in I : h(x_i) = y_j\}} h(x_i) \mathbb{P}(X = x_i) = \sum_{i \in I} h(x_i) \mathbb{P}(X = x_i). \end{aligned}$$

A noter que cette preuve est analogue à la preuve du Théorème 4.2. □

**EXEMPLE 4.11.** En exercices vous calculerez l'espérance, quand elle existe, des variables aléatoires classiques : discrètes ou à densité. A noter que vous utiliserez le théorème de transfert.

*Remarque 4.12.* Supposons qu'une variable aléatoire  $X$  est telle que pour toute fonction continue bornée  $h$ , on ait  $\mathbb{E}(h(X)) = \int_{-\infty}^{\infty} h(x) f(x) dx$ . Alors, en utilisant le Théorème 4.8 qui dit que ces quantités caractérisent la loi, on déduit que  $X$  est une variable aléatoire à densité, avec pour densité  $f$ . En pratique, c'est une méthode efficace pour montrer qu'une variable aléatoire est à densité et pour déterminer cette densité.

EXEMPLE 4.13. Soit  $X$  une variable aléatoire à densité, de densité  $f$ . Soit  $Y = aX + b$  avec  $a \neq 0$ . Alors, d'après le théorème de transfert, pour tout fonction  $h$  continue bornée, on a

$$\begin{aligned}\mathbb{E}(h(Y)) &= \int_{-\infty}^{\infty} h(ax + b)f(x) dx \\ &= \int_{-\infty}^{\infty} h(y)f\left(\frac{y-b}{a}\right)\frac{1}{|a|} dy,\end{aligned}$$

avec le changement de variable  $y = ax + b$ . On déduit du Théorème 4.8 que  $Y$  est une variable aléatoire à densité, de densité  $f^Y$  donnée par

$$\forall y \in \mathbb{R}, \quad f^Y(y) = f\left(\frac{y-b}{a}\right)\frac{1}{|a|}.$$

## 4.2 VARIANCE, COVARIANCE, MOMENTS D'ORDRES SUPÉRIEURS

### 4.2.1 VARIANCE, COVARIANCE, CORRÉLATION, ÉCART-TYPE

Dans toute cette section, on considère une variable aléatoire réelle  $X$  définie sur un espace probabilisé  $(\Omega, \mathcal{F}, \mathbb{P})$ .

**Définition 4.14.** La variable aléatoire  $X$  est dite de *carré intégrable* si  $X^2$  est intégrable. La quantité  $\mathbb{E}(X^2)$  est alors bien définie et est appelée *moment d'ordre 2* de  $X$ . L'ensemble des variables aléatoires de carré intégrable est noté  $L^2$  ou  $L^2(\Omega, \mathcal{F}, \mathbb{P})$ .

*Remarque 4.15.*

- De l'inégalité  $|X| \leq 1 + X^2$  et de la linéarité de l'espérance, on déduit que si  $X$  est de carré intégrable, alors elle est intégrable. La réciproque est fautive en général.
- Soient  $X, Y$  des variables aléatoires de carré intégrables, et soit  $a, b \in \mathbb{R}$ , alors de l'inégalité  $(aX + bY)^2 \leq 2(a^2X^2 + b^2Y^2)$ , nous déduisons que  $aX + bY$  est de carré intégrable. Ainsi, l'ensemble  $L^2(\Omega, \mathcal{F}, \mathbb{P})$  forme un espace vectoriel.

**Proposition 4.16.** Si  $X$  est une variable aléatoire de carré intégrable, alors :

$$\mathbb{E}(|X|) \leq \sqrt{\mathbb{E}(X^2)}.$$

Autrement dit,  $L^2(\Omega, \mathcal{F}, \mathbb{P})$  est un sous-espace vectoriel de  $L^1(\Omega, \mathcal{F}, \mathbb{P})$ .

*Démonstration.* Soit  $\lambda \in \mathbb{R}$ . Si la variable aléatoire  $X$  est de carré intégrable, alors il en est de même pour  $|X| + \lambda$  et nous posons :  $f(\lambda) = \mathbb{E}[(|X| + \lambda)^2]$  En utilisant la linéarité de l'espérance,

$$f(\lambda) = \mathbb{E}(X^2) + 2\lambda\mathbb{E}(|X|) + \lambda^2.$$

Ainsi,  $f$  est un polynôme de degré 2, positif ou nul. Son discriminant est donc négatif ou nul, ce qui implique le résultat. A noter que l'on peut aussi utiliser la preuve à venir de la positivité de la variance avec la variable aléatoire  $|X|$ .  $\square$

**Définition 4.17.** Soit  $X$  une variable aléatoire de carré intégrable.

- La *variance* de  $X$ , notée  $\text{Var}(X)$ , est définie par :

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

- L'*écart-type* de  $X$ , noté  $\sigma(X)$ , est défini par  $\sigma(X) = \sqrt{\text{Var}(X)}$ .
- La variable aléatoire  $X$  est dite *réduite* si sa variance est égale à 1. Elle est dite *centrée-réduite* si son espérance est de plus nulle.

*Remarque.*

- La variance de  $X$  mesure la façon dont  $X$  s'écarte de sa moyenne  $\mathbb{E}(X)$ .
- L'écart-type s'utilise surtout en statistique. Remarquer que si la variable aléatoire  $X$  a une unité, alors l'écart-type a la même unité, tandis que la variance a l'unité au carré, d'où l'intérêt dans la pratique de travailler avec l'écart-type.

**Propriété 4.18.** *Soit  $X$  une variable aléatoire de carré intégrable. Alors,*

1.  $\text{Var}(X) \geq 0$ ,
2.  $\forall a \in \mathbb{R}, \text{Var}(X + a) = \text{Var}(X)$ ,
3.  $\forall a \in \mathbb{R}, \text{Var}(aX) = a^2 \text{Var}(X)$ ,
4.  $\text{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$ .

*Démonstration.*

1. Découle de la propriété de monotonie de l'espérance et du fait que  $(X - \mathbb{E}(X))^2 \geq 0$ .
2. En utilisant la définition de la variance et la linéarité de l'espérance, on a :

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

3. D'après les mêmes propriétés,

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

4. On a aussi

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

**EXEMPLE 4.19.** En exercices vous calculerez la variance, quand elle existe, des variables aléatoires classiques : discrètes ou à densité. A noter que vous utiliserez le théorème de transfert.

Notons que si  $X$  et  $Y$  sont deux variables aléatoires dans  $L^2$ , alors  $XY$  est dans  $L^1$ . En effet, ceci découle de l'inégalité  $|XY| \leq \frac{1}{2}(X^2 + Y^2)$ . Ainsi, la définition suivante fait sens.

**Définition 4.20.** Soient  $X, Y$  deux variables aléatoires de carré intégrable.

- La *covariance* de  $X$  et  $Y$ , notée  $\text{Cov}(X, Y)$ , est définie par :

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



- Si  $X$  et  $Y$  sont de plus de variances non-nulles, la *corrélation de  $X$  et  $Y$* , notée  $\text{Cor}(X, Y)$ , est défini par :

$$\text{Cor}(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma(X)\sigma(Y)}.$$

La corrélation est sans unité et est très utilisée en statistique.

De manière analogue à la variance, la co-variance satisfait à quelques propriétés.

**Proposition 4.21.** *Soient  $X, Y, Z$  des variables aléatoires de carré intégrable. Alors,*

1.  $\text{Cov}(X, X) = \text{Var}(X)$  et  $\text{Cov}(X, Y) = \text{Cov}(Y, X)$ ,
2.  $\forall a, b \in \mathbb{R}, \text{Cov}(aX + bY, Z) = a \text{Cov}(X, Z) + b \text{Cov}(Y, Z)$ .

*Les propriétés 1. et 2. implique que  $\text{Cov}$  est une forme bilinéaire symétrique sur l'espace  $L^2$ , de forme quadratique associée  $\text{Var}$ .*

3.  $\forall a \in \mathbb{R}, \text{Cov}(X + a, Y) = \text{Cov}(X, Y)$ ,
4.  $\text{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$ ,
5. *La variance et la covariance sont reliées par l'égalité :*

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

*Remarque 4.22.* Vous remarquerez que les propriétés de la variance (sauf la positivité) se déduisent du Point 1. et des propriétés de la covariance.

*Démonstration.*

1. Immédiat.
4. Par définition et en utilisant la linéarité de l'espérance,

$$\begin{aligned} \text{Cov}(X, Y) &= \mathbb{E}[(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))] \\ &= \mathbb{E}[XY - \mathbb{E}(X)Y - X\mathbb{E}(Y) + \mathbb{E}(X)\mathbb{E}(Y)] \\ &= \mathbb{E}(XY) - 2\mathbb{E}(X)\mathbb{E}(Y) + \mathbb{E}(X)\mathbb{E}(Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y). \end{aligned}$$

2. D'après la propriété 4. et en utilisant la linéarité de l'espérance,

$$\begin{aligned} \text{Cov}(aX + bY, Z) &= \mathbb{E}[(aX + bY)(Z)] - \mathbb{E}(aX + bY)\mathbb{E}(Z) \\ &= a\mathbb{E}(XZ) + b\mathbb{E}(YZ) - a\mathbb{E}(X)\mathbb{E}(Z) - b\mathbb{E}(Y)\mathbb{E}(Z) \\ &= a \text{Cov}(X, Z) + b \text{Cov}(Y, Z). \end{aligned}$$

3. D'après la propriété 4. et en utilisant la linéarité de l'espérance,

$$\begin{aligned} \text{Cov}(X + a, Y) &= \mathbb{E}[(X + a)Y] - \mathbb{E}(X + a)\mathbb{E}(Y) \\ &= \mathbb{E}(XY) + a\mathbb{E}(Y) - \mathbb{E}(X)\mathbb{E}(Y) - a\mathbb{E}(Y) \\ &= \text{Cov}(X, Y). \end{aligned}$$

5. Par définition de la variance, et en utilisant la linéarité de l'espérance,

$$\begin{aligned} \text{Var}(X + Y) &= \mathbb{E}[(X + Y)^2] - [\mathbb{E}(X + Y)]^2 \\ &= \mathbb{E}(X^2) + 2\mathbb{E}(XY) + \mathbb{E}(Y^2) - [\mathbb{E}(X)^2 + 2\mathbb{E}(X)\mathbb{E}(Y) + \mathbb{E}(Y)^2] \\ &= \mathbb{E}(X^2) - \mathbb{E}(X)^2 + \mathbb{E}(Y^2) - \mathbb{E}(Y)^2 + 2[\mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)] \\ &= \text{Var}(X) + \text{Var}(Y) + 2 \text{Cov}(X, Y). \end{aligned}$$

□

**Définition 4.23.** Si  $X_1, \dots, X_n$  sont des variables aléatoires de carré intégrables, alors il en est de même de leur somme et on définit la *matrice de covariance*  $\text{Cov}$ , par :

$$\forall i, j \in \{1, \dots, n\}, \quad (\text{Cov})_{i,j} = \text{Cov}(X_i, X_j).$$

**Propriété 4.24.** Soient  $X_1, \dots, X_n$  des variables aléatoires de carré intégrables. Alors,

1. La matrice de covariance est une matrice réelle symétrique, dont la diagonale est formée des variances des variables aléatoires  $X_1, \dots, X_n$ .
2.  $\text{Var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \text{Var}(X_i) + 2 \sum_{1 \leq i < j \leq n} \text{Cov}(X_i, X_j).$

*Démonstration.* Le Point 1. est une conséquence directe de la Propriété 4.21. Le Point 2. se démontre par récurrence sur le nombre de variables aléatoires considérées.  $\square$

*Remarque 4.25.* Afin de faire des calculs explicites de covariances, il est utile d'avoir le théorème de transfert pour les vecteurs. Nous verrons ceci dans le chapitre suivant, dans le cas des vecteurs discrets uniquement.

Passons maintenant aux moments d'ordres supérieurs.

**Définition 4.26.** Soit  $X$  une variable aléatoires, et soit  $n \in \mathbb{N}^*$ . Si  $X^n$  est intégrable, la quantité  $\mathbb{E}(X^n)$  est bien définie et on l'appelle *moment d'ordre  $n$*  de la variable aléatoire  $X$ .

Comme conséquence du Théorème de transfert 4.10, on obtient le corollaire suivant.

**Corollaire 4.27.**

- Soit  $X$  une variable aléatoire discrète telle que  $X^n$  est intégrable, c'est-à-dire telle que,  $\sum_{i \in I} |x_i|^n \mathbb{P}^X(x_i) < \infty$ , alors son moment d'ordre  $n$  est égal à

$$\mathbb{E}(X^n) = \sum_{i \in I} x_i^n \mathbb{P}(X = x_i).$$

- Soit  $X$  une variable aléatoire à densité telle que  $X^n$  est intégrable, c'est-à-dire telle que,  $\int_{-\infty}^{\infty} |x|^n f(x) dx < \infty$ , alors son moment d'ordre  $n$  est égal à

$$\mathbb{E}(X^n) = \int_{-\infty}^{\infty} x^n f(x) dx.$$

### 4.3 INÉGALITÉ DE MARKOV, BIENAYMÉ-TCHEBYCHEV ET AUTRES

Voici deux inégalités classiques.

**Proposition 4.28** (Inégalité de Markov). Soit  $X$  une variable aléatoire admettant un moment d'ordre  $n \geq 1$ . Alors,

$$\forall a > 0, \quad \mathbb{P}(|X| \geq a) \leq \frac{\mathbb{E}[|X|^n]}{a^n}.$$

*Démonstration.* Une manière courte d'écrire cette preuve est d'utiliser une variable aléatoire indicatrice. Remarquer que l'on peut écrire la fonction constante égale à 1 sur  $\Omega$  de la manière suivante : pour tout  $\omega \in \Omega$ , pour tout réel  $a > 0$ ,

$$1(\omega) = \mathbb{I}_{\{|X| \geq a\}}(\omega) + \mathbb{I}_{\{|X| < a\}}(\omega).$$

Ainsi, pour tout  $\omega \in \Omega$ , pour tout  $a > 0$ , on a :

$$\begin{aligned} |X(\omega)|^n &= |X(\omega)|^n \mathbb{I}_{\{|X| \geq a\}}(\omega) + |X(\omega)|^n \mathbb{I}_{\{|X| < a\}}(\omega) \\ &\geq |X(\omega)|^n \mathbb{I}_{\{|X| \geq a\}}(\omega), \quad \text{car } |X(\omega)|^n \mathbb{I}_{\{|X| < a\}}(\omega) \geq 0 \\ &\geq a^n \mathbb{I}_{\{|X| \geq a\}}(\omega), \quad \text{car } a \text{ est positif.} \end{aligned}$$

Autrement dit, on a l'inégalité  $|X|^n \geq a^n \mathbb{I}_{\{|X| \geq a\}}$ . On conclut en utilisant la monotonie de l'espérance :

$$\mathbb{E}[|X|^n] \geq \mathbb{E}[a^n \mathbb{I}_{\{|X| \geq a\}}] = a^n \mathbb{P}[|X| \geq a]. \quad \square$$

L'inégalité suivante formalise le fait que plus la variance est grande, plus les fluctuations de la variable aléatoire autour de sa moyenne sont potentiellement grandes. A noter qu'il s'agit d'une version “déguisée” de l'inégalité de Markov dans le cas où  $X$  est de carré intégrable.

**Proposition 4.29** (Inégalité de Bienaymé-Tchebychev). *Soit  $X$  une variable aléatoire de carré intégrable. Alors,*

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

*Démonstration.* L'inégalité de Bienaymé-Tchebychev est une conséquence de l'inégalité de Markov. La variable aléatoire  $X$  étant de carré intégrable, il en est de même pour la variable aléatoire  $X - \mathbb{E}(X)$ . On applique l'inégalité de Markov avec la variable aléatoire  $Y = X - \mathbb{E}(X)$ , et  $n = 2$ . Pour tout  $a > 0$ ,

$$\begin{aligned} \mathbb{P}(|Y| \geq a) &\leq \frac{\mathbb{E}[|Y|^2]}{a^2} \\ \Leftrightarrow \mathbb{P}(|X - \mathbb{E}(X)| \geq a) &\leq \frac{\mathbb{E}[(X - \mathbb{E}(X))^2]}{a^2} = \frac{\text{Var}(X)}{a^2}. \end{aligned} \quad \square$$

**Proposition 4.30** (Inégalité de Jensen). *Soit  $X$  une variable aléatoire intégrable et soit  $f : \mathbb{R} \rightarrow \mathbb{R}$  une fonction continue, convexe, telle que  $f(X)$  est intégrable. Alors,*

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

*Démonstration.* La fonction  $f$  étant convexe, on a

$$\forall a \in \mathbb{R}, \quad \exists \lambda_a \in \mathbb{R}, \quad \forall x \in \mathbb{R}, \quad f(x) \geq f(a) + \lambda_a(x - a).$$

Ce résultat est vrai en particulier pour  $a = \mathbb{E}(X)$  : il existe  $\lambda_a \in \mathbb{R}$  tel que pour tout  $\omega \in \Omega$ ,

$$f(X(\omega)) \geq f(\mathbb{E}(X)) + \lambda_a(X(\omega) - a) \Leftrightarrow f(X) \geq f(\mathbb{E}(X)) + \lambda_a(X - a).$$

En prenant l'espérance, on obtient :

$$\mathbb{E}(f(X)) \geq \mathbb{E}(f(\mathbb{E}(X)) + \lambda_a(X - \mathbb{E}(X))) \Leftrightarrow \mathbb{E}(f(X)) \geq f(\mathbb{E}(X)),$$

car  $\mathbb{E}(f(X))$  est constante et  $\mathbb{E}(X - \mathbb{E}(X)) = \mathbb{E}(X) - \mathbb{E}(X) = 0$ . □

**Proposition 4.31** (Inégalité de Cauchy-Schwarz). *Soient  $X, Y$  deux variables aléatoires de carré intégrables, alors  $XY$  est intégrable, et :*

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

*Démonstration.* La preuve sera faite en exercices. □

Comme conséquences de l'inégalité de Cauchy-Schwartz, on a :

**Proposition 4.32.** *Soient  $X, Y$  deux variables aléatoires de carré intégrables. La covariance vérifie l'inégalité :*

$$|\text{Cov}(X, Y)| \leq \sigma(X)\sigma(Y).$$

*En conséquence la corrélation satisfait :*

$$-1 \leq \text{Cor}(X, Y) \leq 1.$$

*Démonstration.* Il suffit d'appliquer l'inégalité de Cauchy-Schwartz avec  $X - \mathbb{E}(X)$  et  $Y - \mathbb{E}(Y)$ . La conséquence pour la corrélation est immédiate. □



## CHAPITRE 5

# VECTEURS ALÉATOIRES DISCRETS, INDÉPENDANCE, LOI DES GRANDS NOMBRES

Le but de ce chapitre est de considérer le cas où une variable aléatoire est formée de composantes qui sont chacune des variables aléatoires discrètes ; on parle de vecteur aléatoire discret. De nouvelles questions se posent alors : connaître la loi des composantes suffit-il à connaître la loi du vecteur ? Nous verrons que non. Les composantes se comportent-elles de manières indépendantes ? Nous verrons que pas forcément. Ceci nous amènera aussi naturellement à la notion d'indépendance de variables aléatoires.

Les notions abordées dans ce chapitre s'étendent de manière naturelle au cas des vecteurs aléatoires à valeurs dans  $\mathbb{R}^n$ , où l'espace  $\mathbb{R}^n$  est muni de la *tribu borélienne* de  $\mathbb{R}^n$ , notée  $\mathcal{B}(\mathbb{R}^n)$ , qui est la tribu engendrée par les pavés de la forme  $\prod_{i=1}^n ]-\infty, x_i]$ , avec  $x_i \in \mathbb{R}$ ,  $i = 1, \dots, n$ . Nous nous restreignons cependant au cas discret, car le cas général nécessite la théorie de l'intégration à plusieurs variables, concept que vous êtes entrain d'acquérir. Nous ferons une entorse à ceci pour le calcul de la densité de la somme de deux variables aléatoires indépendantes à densité qui, après une preuve que nous omettrons, ne nécessite que l'intégration uni-dimensionnelle.

### 5.1 VECTEURS ALÉATOIRES DISCRETS

#### 5.1.1 DÉFINITIONS ET PREMIER RÉSULTAT

**Définition 5.1.** Soit  $X = (X_1, \dots, X_n) : \Omega \rightarrow E = E_1 \times \dots \times E_n$  une application, où pour tout  $i \in \{1, \dots, n\}$ ,  $E_i$  est dénombrable ou fini. On munit  $E$  de la tribu  $\mathcal{P}(E)$ <sup>1</sup>. Alors, si  $X$  est une variable aléatoire de  $(\Omega, \mathcal{F})$  dans  $(E, \mathcal{P}(E))$ , on dit que  $X$  est un *vecteur aléatoire discret*.

Lorsque  $n = 2$ , on parle de *couple aléatoire discret* ; on notera ce couple  $Z = (X, Y)$  et l'espace d'arrivée  $E \times F$ .

*Remarque 5.2.*

1. On peut montrer que  $X = (X_1, \dots, X_n)$  est un vecteur aléatoire discret si et seulement si chacune de ses composantes est une variable aléatoire discrète.

---

1. Comme les  $(E_i)_{i=1}^n$  sont discrets et comme le produit  $E_1 \times \dots \times E_n$  est fini, la tribu  $\mathcal{P}(E)$  est aussi la tribu produit sur  $E_1 \times \dots \times E_n$ . Remarquons aussi qu'en général  $\mathcal{P}(E) \neq \mathcal{P}(E_1) \times \dots \times \mathcal{P}(E_n)$ .

2. Remarquons que la définition de vecteur aléatoire discret rentre dans le cadre de la définition de variable aléatoire discrète de la Section 3.2.1 ; en effet, l'espace d'arrivée  $E = E_1 \times \cdots \times E_n$  est bien dénombrable ou fini car il est produit cartésien fini d'espaces dénombrables ou finis.

Même si les vecteurs aléatoires discrets sont un cas particulier de variables aléatoires discrètes de la Section 3.2.1, ils possèdent une terminologie propre qui permet d'étudier les relations entre le vecteur et ses composantes.

**Définition 5.3** (Proposition).

1. Soit  $X = (X_1, \dots, X_n)$  un vecteur aléatoire discret. Alors, la loi  $\mathbb{P}^X = \mathbb{P}^{(X_1, \dots, X_n)}$  du vecteur aléatoire  $X = (X_1, \dots, X_n)$  est appelée la *loi jointe du vecteur aléatoire*. Pour tout  $i \in \{1, \dots, n\}$ , on appelle *loi marginale de  $X_i$* , la loi  $\mathbb{P}^{X_i}$  de la variable aléatoire  $X_i$ .
2. Comme application de la Proposition 3.5 caractérisant la loi des variables aléatoires discrètes, on a que la loi jointe du vecteur aléatoire discret  $(X_1, \dots, X_n)$  est caractérisée par la donnée des nombres, pour tout  $(x_1, \dots, x_n) \in E_1 \times \cdots \times E_n$ ,

$$\mathbb{P}^{(X_1, \dots, X_n)}(\{(x_1, \dots, x_n)\}) = \mathbb{P}(\{X_1 = x_1, \dots, X_n = x_n\}) = \mathbb{P}(\{X_1 = x_1\} \cap \cdots \cap \{X_n = x_n\}).$$

Voici comment se réécrit la définition ci-dessus dans le cas des couples aléatoires discrets. En utilisant la formule des probabilités totales, on obtient une expression explicite pour les lois marginales. À noter que l'on pourrait aussi faire ceci dans le cas des vecteurs aléatoires de taille supérieure à 2 ; il y aurait simplement des notations plus compliquées au niveau des indices.

**Définition 5.4** (Proposition).

1. Soit  $Z = (X, Y)$  un couple aléatoire discret. Alors la loi  $\mathbb{P}^Z = \mathbb{P}^{(X, Y)}$  du couple aléatoire  $Z = (X, Y)$  est appelée la *loi jointe du couple aléatoire*. On appelle *loi marginale de  $X$* , resp.  $Y$ , la loi  $\mathbb{P}^X$ , resp.  $\mathbb{P}^Y$ , de la variable aléatoire  $X$ , resp.  $Y$ .
2. La loi jointe du couple aléatoire  $(X, Y)$  est caractérisée par la donnée des nombres, pour tout  $(x, y) \in E \times F$ ,

$$\mathbb{P}^{(X, Y)}(\{(x, y)\}) = \mathbb{P}(\{X = x, Y = y\}) = \mathbb{P}(\{X = x\} \cap \{Y = y\}).$$

3. Connaissant la loi jointe du couple aléatoire discret  $(X, Y)$ , on retrouve les lois marginales  $\mathbb{P}^X$ , resp.  $\mathbb{P}^Y$ , des variables aléatoires discrètes  $X$ , resp.  $Y$  :

$$\begin{aligned} \forall x \in E, \quad \mathbb{P}^X(\{x\}) &= \mathbb{P}(\{X = x\}) = \sum_{y \in F} \mathbb{P}^{(X, Y)}(\{(x, y)\}) = \sum_{y \in F} \mathbb{P}(\{X = x, Y = y\}), \\ \forall y \in F, \quad \mathbb{P}^Y(\{y\}) &= \mathbb{P}(\{Y = y\}) = \sum_{x \in E} \mathbb{P}^{(X, Y)}(\{(x, y)\}) = \sum_{x \in E} \mathbb{P}(\{X = x, Y = y\}). \end{aligned}$$

Comme dans le cas des variables aléatoires discrètes réelles, le théorème de transfert est très utile en pratique pour calculer l'espérance d'une fonction d'un vecteur aléatoire. Nous énonçons la version pour les couples, la version pour les vecteurs de taille supérieure à 2 étant tout à fait similaire.

**Théorème 5.5** (Théorème de transfert). *Soit  $(X, Y)$  un couple aléatoire discret, et soit  $h : E \times F \rightarrow \mathbb{R}$  une application telle que  $h$  est positive ou intégrable, c'est-à-dire telle que la série  $\sum_{(x, y) \in E \times F} |h(x, y)| \mathbb{P}^{(X, Y)}(\{(x, y)\})$  converge, alors :*

$$\mathbb{E}(h(X, Y)) = \sum_{(x, y) \in E \times F} h(x, y) \mathbb{P}^{(X, Y)}(\{(x, y)\}) = \sum_{(x, y) \in E \times F} h(x, y) \mathbb{P}(\{X = x, Y = y\}).$$

EXEMPLE 5.6. Soit  $Z = (X, Y)$  un couple aléatoire à valeurs dans

$$\{-1, 1\}^2 = \{(-1, -1), (-1, 1), (1, -1), (1, 1)\},$$

avec probabilités respectives  $\frac{1}{2} - p, p, p, \frac{1}{2} - p$ , où  $0 \leq p \leq \frac{1}{2}$ . Calculons les lois marginales de  $X$  et de  $Y$ . Remarquons que  $X$  et  $Y$  sont chacune à valeurs dans  $\{-1, 1\}$ . De la Définition 5.4, nous déduisons :

$$\begin{aligned}\mathbb{P}(X = 1) &= \mathbb{P}(Z = (1, 1)) + \mathbb{P}(Z = (1, -1)) = \frac{1}{2} - p + p = \frac{1}{2} \\ \mathbb{P}(X = -1) &= \mathbb{P}(Z = (-1, 1)) + \mathbb{P}(Z = (-1, -1)) = p + \frac{1}{2} - p = \frac{1}{2} \\ \mathbb{P}(Y = 1) &= \mathbb{P}(Z = (1, 1)) + \mathbb{P}(Z = (-1, 1)) = \frac{1}{2} - p + p = \frac{1}{2} \\ \mathbb{P}(Y = -1) &= \mathbb{P}(Z = (1, -1)) + \mathbb{P}(Z = (-1, -1)) = p + \frac{1}{2} - p = \frac{1}{2},\end{aligned}$$

donc  $X$  et  $Y$  suivent la loi uniforme sur  $\{-1, 1\}$ ; cependant le couple  $(X, Y)$  ne suit pas, en général, la loi uniforme sur  $\{-1, 1\}^2$ .

Calculons maintenant la covariance de  $X$  et  $Y$ . On a

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

Comme  $X$  et  $Y$  suivent chacune la loi uniforme sur  $\{-1, 1\}$ , on a  $\mathbb{E}(X) = \mathbb{E}(Y) = 0$ . De plus, d'après le théorème de transfert,

$$\begin{aligned}\mathbb{E}(XY) &= 1 \cdot 1\mathbb{P}(X = 1, Y = 1) + 1 \cdot (-1)\mathbb{P}(X = 1, Y = -1) + \\ &\quad + (-1) \cdot 1\mathbb{P}(X = -1, Y = 1) + (-1) \cdot (-1)\mathbb{P}(X = -1, Y = -1) \\ &= 2\left(\frac{1}{2} - p\right) - 2p = 1 - 4p.\end{aligned}$$

Ainsi,  $\text{Cov}(X, Y) = 1 - 4p$ .

### 5.1.2 LOIS ET ESPÉRANCE CONDITIONNELLES

On peut de plus s'intéresser à la loi de la variable aléatoire  $X$  (ou  $Y$ ) lorsque l'on a une information supplémentaire sur les réalisations de la variable aléatoire  $Y$  (ou  $X$ ). Il est donc naturel de définir les lois conditionnelles. A nouveau, nous donnons les définitions dans le cas des couples, le cas général étant facilement déductible.

**Définition 5.7** (Proposition). Soit  $(X, Y)$  un couple aléatoire discret. Soit  $x \in E$  tel que  $\mathbb{P}(X = x) > 0$  et soit  $y \in F$  tel que  $\mathbb{P}(Y = y) > 0$ . Alors :

- la *loi conditionnelle de  $Y$  sachant  $\{X = x\}$*  est la probabilité  $\mathbb{P}^Y(\cdot | X = x) : \mathcal{P}(F) \rightarrow [0, 1]$  caractérisée par la donnée des nombres

$$\forall y \in F, \quad \mathbb{P}(\{Y = y\} | \{X = x\}) = \frac{\mathbb{P}(\{Y = y, X = x\})}{\mathbb{P}(\{X = x\})} = \frac{\mathbb{P}^{(X, Y)}(\{(x, y)\})}{\mathbb{P}^X(\{x\})}.$$

- la *loi conditionnelle de  $X$  sachant  $\{Y = y\}$*  est la probabilité  $\mathbb{P}^X(\cdot | Y = y) : \mathcal{P}(E) \rightarrow [0, 1]$  caractérisée par la donnée des nombres

$$\forall x \in E, \quad \mathbb{P}(\{X = x\} | \{Y = y\}) = \frac{\mathbb{P}(\{X = x, Y = y\})}{\mathbb{P}(\{Y = y\})} = \frac{\mathbb{P}^{(X, Y)}(\{(x, y)\})}{\mathbb{P}^Y(\{y\})}.$$



*Remarque 5.8.*

1. La loi du couple aléatoire  $(X, Y)$  permet de connaître la loi des variables aléatoires  $X$  et  $Y$ , mais la réciproque est fausse ! La connaissance de chacune des lois  $X$  et  $Y$  n'entraîne pas la connaissance de la loi du couple. Voir l'exemple ci-dessous.
2. Comme conséquence de la Définition 5.7, on sait que si  $x \in E$  et  $y \in F$  sont tels que  $\mathbb{P}(\{X = x\}) > 0$  et  $\mathbb{P}(\{Y = y\}) > 0$ , alors :

$$\mathbb{P}^{(X,Y)}(\{(x, y)\}) = \mathbb{P}(\{Y = y\}|\{X = x\})\mathbb{P}(\{X = x\}) = \mathbb{P}(\{X = x\}|\{Y = y\})\mathbb{P}(\{Y = y\}).$$

Ainsi, dans ce cas, la connaissance de la loi de  $(X, Y)$  est équivalente à la connaissance de celle de  $X$  et de la loi conditionnelle de  $Y$  sachant  $X$ , ou encore est équivalente à la connaissance de la loi de  $Y$  et de la loi conditionnelle de  $X$  sachant  $Y$ .

Nous pouvons maintenant définir l'espérance conditionnelle.

**Définition 5.9.**

1. Soit  $X$  et  $Y$  deux variables aléatoires discrètes telle que  $Y$  est intégrable, et soit  $x \in E$  tel que  $\mathbb{P}(X = x) > 0$ . L'*espérance conditionnelle de  $Y$  sachant  $\{X = x\}$*  est l'espérance de la loi conditionnelle de  $Y$  sachant  $\{X = x\}$ , c'est-à-dire,

$$\mathbb{E}(Y|X = x) = \sum_{y \in F} y \mathbb{P}(Y = y|X = x).$$

2. On appelle *espérance conditionnelle de  $Y$  sachant  $X$*  la variable aléatoire

$$\mathbb{E}(Y|X) = \psi(X), \text{ où } \psi(x) = \begin{cases} \mathbb{E}(Y|X = x) & \text{si } \mathbb{P}(X = x) > 0 \text{ et} \\ 0 & \text{sinon.} \end{cases}$$

*Remarque 5.10.* Attention, contrairement l'espérance qui est un nombre réelle, l'espérance conditionnelle est une variable aléatoire, dépendant de l'aléa à travers la variable aléatoire  $X$ .

L'espérance conditionnelle possède la propriété fondamentale suivante.

**Théorème 5.11.** *Si la variable aléatoire  $Y$  est intégrable, alors l'espérance conditionnelle  $\mathbb{E}(Y|X)$  est intégrable et*

$$\mathbb{E}[\mathbb{E}(Y|X)] = \mathbb{E}(Y).$$

*Démonstration.* Nous ne démontrons pas l'intégrabilité qui découle du théorème de Fubini pour les séries. Posons  $\Psi(X) = \mathbb{E}(Y|X)$ . Alors, d'après le théorème de transfert,

$$\begin{aligned} \mathbb{E}[\Psi(X)] &= \sum_{x \in E} \psi(x) \mathbb{P}(X = x) \\ &= \sum_{x \in E} \mathbb{E}(Y|X = x) \mathbb{P}(X = x), \text{ par définition de l'espérance conditionnelle} \\ &= \sum_{x \in E} \sum_{y \in F} y \mathbb{P}(Y = y|X = x) \mathbb{P}(X = x), \text{ par définition de } \mathbb{E}(Y|X = x) \\ &= \sum_{y \in F} y \sum_{x \in E} \mathbb{P}(Y = y|X = x) \mathbb{P}(X = x), \text{ car les séries sont convergentes} \\ &= \sum_{y \in F} y \mathbb{P}(Y = y), \text{ d'après la formule des probabilités totales.} \end{aligned}$$

□

L'espérance conditionnelle étant définie comme l'espérance de la loi conditionnelle, elle hérite des propriétés de l'espérance.

**Théorème 5.12.** *Soit  $Y, Z$  des variables aléatoires intégrables. Alors,*

1. **Linéarité.** *Pour tout  $a, b \in \mathbb{R}$ ,*

$$\mathbb{E}(aY + bZ|X) = a\mathbb{E}(Y|X) + b\mathbb{E}(Z|X).$$

2. **Positivité.**  $Y \geq 0 \implies \mathbb{E}(Y|X) \geq 0$ .

3.  $\mathbb{E}(1|X) = 1$ .

4. *Si  $g : E \rightarrow \mathbb{R}$  est telle que  $g(X)$  est intégrable, alors*

$$\mathbb{E}[Yg(X)|X] = g(X)\mathbb{E}(Y|X),$$

*c'est-à-dire que  $g(X)$  doit être considérée comme une constante du point de vue de l'espérance conditionnelle sachant  $X$ .*

*Démonstration.* Démontrons uniquement le Point 3. Pour tout  $x \in E$  tel que  $\mathbb{P}(X = x) > 0$ , posons  $\Psi_1(X) = \mathbb{E}(Yg(X)|X)$ , et  $\Psi_2(X) = g(X)\mathbb{E}(Y|X)$ . L'égalité à montrer revient à montrer que  $\Psi_1$  et  $\Psi_2$  sont égales en tout point. Par définition de l'espérance conditionnelle, on a

$$\begin{aligned} \Psi_1(x) &= \mathbb{E}(Yg(X)|X = x) \\ &= \sum_{(x', y) \in E \times F} yg(x')\mathbb{P}(X = x', Y = y|X = x), \text{ d'après le théorème de transfert} \\ &= \sum_{(x', y) \in E \times F} yg(x') \frac{\mathbb{P}(X = x', Y = y, X = x)}{\mathbb{P}(X = x)}, \text{ d'après la définition de probabilité conditionnelle} \\ &= \sum_{y \in F} yg(x)\mathbb{P}(Y = y|X = x), \text{ car } \{X = x', X = x\} = \emptyset \text{ si } x \neq x' \\ &= g(x) \sum_{y \in F} y\mathbb{P}(Y = y|X = x) = g(x)\mathbb{E}(Y|X = x) = \Psi_2(x). \end{aligned} \quad \square$$

**EXEMPLE 5.13.** Reprenons l'Exemple 5.6 et calculons maintenant la loi conditionnelle de  $Y$  sachant  $\{X = 1\}$ . De la Définition 5.7, comme  $\mathbb{P}(X = 1) = \frac{1}{2} > 0$ , la loi conditionnelle est bien définie et est caractérisée par :

$$\begin{aligned} \mathbb{P}(Y = 1|X = 1) &= \frac{\mathbb{P}(Y = 1, X = 1)}{\mathbb{P}(X = 1)} = \frac{\mathbb{P}(Z = (1, 1))}{\mathbb{P}(X = 1)} = \frac{\frac{1}{2} - p}{\frac{1}{2}} = 1 - 2p. \\ \mathbb{P}(Y = -1|X = 1) &= \frac{\mathbb{P}(Y = -1, X = 1)}{\mathbb{P}(X = 1)} = \frac{\mathbb{P}(Z = (1, -1))}{\mathbb{P}(X = 1)} = \frac{p}{\frac{1}{2}} = 2p. \end{aligned}$$

On peut également calculer l'espérance conditionnelle de  $Y$  sachant que  $\{X = 1\}$ . D'après la Définition 5.9,

$$\mathbb{E}(Y|X = 1) = 1 \cdot \mathbb{P}(Y = 1|X = 1) - 1 \cdot \mathbb{P}(Y = -1|X = 1) = (1 - 2p) + (-1)(2p) = 1 - 4p.$$

De manière analogue, on peut montrer que

$$\mathbb{E}(Y|X = -1) = 4p - 1.$$

Ainsi, l'espérance conditionnelle de  $Y$  sachant  $X$  est la variable aléatoire  $\Psi(X) = \mathbb{E}(Y|X)$  qui prend les valeurs

$$\Psi(x) = \begin{cases} 1 - 4p & \text{si } x = 1 \\ 4p - 1 & \text{si } x = -1. \end{cases}$$

On retrouve sur un exemple que  $\mathbb{E}[\mathbb{E}(Y|X)] = \mathbb{E}(\Psi(X)) = (1 - 4p)\frac{1}{2} + (4p - 1)\frac{1}{2} = 0 = \mathbb{E}(Y)$ , ceci étant le contenu du Théorème 5.11.

## 5.2 INDÉPENDANCE

Pour aborder cette section, il faut se souvenir de la notion d'indépendance des évènements introduite à la Section 2.2.

### Définition 5.14.

- Soit  $X, Y$  deux variable aléatoires discrètes. On dit que  $X$  et  $Y$  sont *indépendantes*, si pour tout  $A \in \mathcal{P}(E)$ ,  $B \in \mathcal{P}(F)$ ,

$$\begin{aligned} \mathbb{P}(\{X \in A, Y \in B\}) &= \mathbb{P}(\{X \in A\} \cap \{Y \in B\}) = \mathbb{P}(\{X \in A\})\mathbb{P}(\{Y \in B\}) \\ \Leftrightarrow \mathbb{P}^{(X,Y)}(A \times B) &= \mathbb{P}^X(A)\mathbb{P}^Y(B). \end{aligned}$$

- Soit  $(X_i)_{i \in I}$  une famille de variables aléatoires, avec  $I$  fini ou dénombrable. Alors les variables aléatoires  $(X_i)_{i \in I}$  sont dites *mutuellement indépendantes*, ou simplement *indépendantes* si, pour tout partie finie  $K$  de  $I$ , et pour tout  $A_i \in \mathcal{P}(E_i)$ ,  $i \in K$ ,

$$\mathbb{P}(\cap_{i \in K} \{X_i \in A_i\}) = \prod_{i \in K} \mathbb{P}(X_i \in A_i).$$

- Les variables aléatoires  $(X_i)_{i \in I}$  sont dites *indépendantes et identiquement distribuées*, abrégé *i.i.d.*, si elles sont indépendantes, et si toutes les variables aléatoires  $X_i$  ont même loi.

*Remarque 5.15.* Comme conséquence de la définition d'indépendance, nous avons que si  $(X_i)_{i \in I}$  est une famille de variables aléatoires indépendantes, alors il en est de même pour :

1. toute famille  $(X_i)_{i \in J}$ , avec  $J \subset I$ ,
2. toute famille  $(Y_i)_{i \in I}$  où, pour tout  $i \in I$ ,  $Y_i = h_i(X_i)$  avec  $h_i : E_i \rightarrow \mathbb{R}$  et  $h_i$  mesurable.
3. toute famille  $(Y_j)_{j \in J}$  obtenue en posant  $Y_j = h_j(X_i : i \in I_j)$ , où les  $(I_j)_{j \in J}$  sont des parties 2-à-2 disjointes de  $I$  et où, pour tout  $j \in J$ ,  $h_j : \mathbb{R}^{|I_j|} \rightarrow \mathbb{R}$  est une fonction mesurable.

Nous avons les propriétés fondamentales suivantes qui permettent de caractériser l'indépendance de variables aléatoires.

**Théorème 5.16.** *Soient  $X, Y$  deux variables aléatoires discrètes. Alors, les conditions suivantes sont équivalentes.*

1. Les variables aléatoires  $X$  et  $Y$  sont indépendantes.
2. Pour tout  $x \in E, y \in F$ ,

$$\begin{aligned} \mathbb{P}(\{X = x, Y = y\}) &= \mathbb{P}(\{X = x\})\mathbb{P}(\{Y = y\}) \\ \Leftrightarrow \mathbb{P}^{(X,Y)}(x, y) &= \mathbb{P}^X(x)\mathbb{P}^Y(y) \end{aligned}$$

3. Pour tout  $x \in E, \mathbb{P}(X = x) > 0 \Rightarrow \mathbb{P}^Y(\cdot | \{X = x\}) = \mathbb{P}^Y(\cdot)$ .
4. Pour tout  $y \in F, \mathbb{P}(Y = y) > 0 \Rightarrow \mathbb{P}^X(\cdot | \{Y = y\}) = \mathbb{P}^X(\cdot)$ .
5. Pour toutes fonction  $f$  et  $g$  définies respectivement sur  $E$  et  $F$ , telles que  $f(X)$  et  $g(Y)$  sont bornées, alors

$$\mathbb{E}[f(X)g(Y)] = \mathbb{E}[f(X)]\mathbb{E}[g(Y)].$$

*Remarque 5.17.*

1. L'implication 1.  $\Rightarrow$  5. reste vraie dès que les fonctions  $f, g$  sont telles que  $f(X), g(Y)$  et  $f(X)g(Y)$  sont intégrables.
2. Dans le Point 5. du Théorème 5.16, on peut remplacer l'ensemble des fonctions continues bornées par l'ensemble des fonctions boréliennes positives par exemple.

*Démonstration.*

- Implication 1.  $\Rightarrow$  2. Il suffit de prendre  $A = \{x\}, B = \{y\}$  dans la définition.
- Implication 2.  $\Rightarrow$  1. Soit  $A \in \mathcal{P}(E), B \in \mathcal{P}(F)$ . Alors,

$$\begin{aligned} \mathbb{P}(\{X \in A, Y \in B\}) &= \sum_{(x,y) \in A \times B} \mathbb{P}(\{X = x, Y = y\}) \\ &= \sum_{(x,y) \in A \times B} \mathbb{P}(\{X = x\})\mathbb{P}(\{Y = y\}), \text{ d'après le Point 2.} \\ &= \left( \sum_{x \in A} \mathbb{P}(\{X = x\}) \right) \left( \sum_{y \in B} \mathbb{P}(\{Y = y\}) \right) \\ &= \mathbb{P}(\{X \in A\})\mathbb{P}(\{Y \in B\}). \end{aligned}$$

- Equivalence 2.  $\Leftrightarrow$  3.  $\Leftrightarrow$  4. Provient du Point 2. de la Remarque 5.8. Noter que si  $\mathbb{P}(\{X = 0\}) = 0$ , alors  $0 \leq \mathbb{P}(\{X = x, Y = x\}) \leq \mathbb{P}(\{X = x\}) = 0$ , de sorte que l'on a bien  $\mathbb{P}(\{X = x, Y = x\}) = \mathbb{P}(\{X = x\})\mathbb{P}(\{Y = y\})$ .
- Implication 2.  $\Rightarrow$  5. D'après le théorème de transfert

$$\begin{aligned} \mathbb{E}[f(X)g(Y)] &= \sum_{(x,y) \in E \times F} f(x)g(y)\mathbb{P}(\{X = x, Y = y\}) \\ &= \sum_{(x,y) \in E \times F} f(x)g(y)\mathbb{P}(\{X = x\})\mathbb{P}(\{Y = y\}) \text{ d'après le Point 2.} \\ &= \left( \sum_{x \in E} f(x)\mathbb{P}(\{X = x\}) \right) \left( \sum_{y \in Y} g(y)\mathbb{P}(\{Y = y\}) \right) \\ &= \mathbb{E}[f(X)]\mathbb{E}[g(Y)]. \end{aligned}$$

- Implication 5.  $\Rightarrow$  2. Il suffit de prendre  $f = \mathbb{I}_{\{X=x\}}, g = \mathbb{I}_{\{Y=y\}}$ . A noter que comme nous sommes dans le cas discret, nous ne devons pas nous soucier de la continuité de  $f$  et de  $g$ , tel que c'était le cas dans la démonstration du Théorème 4.8.  $\square$

### 5.3 CONSÉQUENCES

L'indépendance a des conséquences sur le calcul de la variance et de la covariance.

**Proposition 5.18.** *Soient  $X, Y$  deux variables aléatoires de carré intégrables indépendantes. Alors,*

1.  $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$ .
2.  $\text{Cov}(X, Y) = 0$ .
3.  $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$ .

*Soient  $X_1, \dots, X_n$  des variables aléatoires de carré intégrables indépendantes. Alors,*

4.  $\mathbb{E}(\prod_{i=1}^n X_i) = \prod_{i=1}^n \mathbb{E}(X_i)$ .
5.  $\forall i, j \in \{1, \dots, n\}, i \neq j, \text{Cov}(X_i, X_k) = 0$ , autrement dit, la matrice de covariance  $\text{Cov}$  est diagonale.
6.  $\text{Var}(\sum_{i=1}^n X_i) = \sum_{i=1}^n \text{Var}(X_i)$ .

*Démonstration.*

1. Conséquence directe du Théorème 5.16 et de la Remarque 5.17.
2. Découle de la définition du fait que  $\text{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$  et du Point 1.
3. Découle du Point 5. de la Proposition 4.21.
- 4.-6. Les Points 4. et 6. se montrent par récurrence sur le nombre de variables aléatoires considérées. Le Point 5. est une conséquence du Point 2.  $\square$

*Remarque 5.19.*

1. A noter que la réciproque du Point 2. est fautive : si la covariance de  $X$  et  $Y$  est nulle, ceci n'implique pas que  $X$  et  $Y$  sont indépendantes. Prenons par exemple une variable aléatoire  $Z$  de loi uniforme sur  $\{-1, 0, 1\}$ , et posons  $X = Z, Y = Z^2$ . Alors

$$\text{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) = \mathbb{E}(Z^3) - \mathbb{E}(Z)\mathbb{E}(Z^2) = 0,$$

car  $Z^3 = Z$  et  $\mathbb{E}(Z) = -\frac{1}{3} + \frac{1}{3} = 0$ . Cependant  $X$  et  $Y$  ne sont pas indépendantes, car on a

$$\mathbb{P}(X = -1, Y = 1) = \mathbb{P}(Z = -1, Z^2 = 1) = \mathbb{P}(Z = -1) = \frac{1}{3}$$

$$\mathbb{P}(X = -1)\mathbb{P}(Y = 1) = \mathbb{P}(Z = -1)\mathbb{P}(Z^2 = 1) = \mathbb{P}(Z = -1)\mathbb{P}(Z \in \{-1, 1\}) = \frac{1}{3} \frac{2}{3} = \frac{2}{9},$$

et ces quantités ne sont pas égales.

Une application utile en pratique est le calcul de la loi de la somme de variables aléatoires et sa spécification lorsque  $X$  et  $Y$  sont indépendantes.

**Proposition 5.20.** *Soient  $X, Y$  deux variables aléatoires discrètes réelles. Alors la variable aléatoire  $X + Y$  est à valeurs dans  $G = \{x + y : x \in E, y \in F\}$  et la loi de  $S = X + Y$  est caractérisée par*

$$\forall s \in G, \quad \mathbb{P}^S(s) = \mathbb{P}(X + Y = s) = \sum_{x \in E} \mathbb{P}(X = x, Y = s - x) = \sum_{y \in F} \mathbb{P}(X = s - y, Y = y).$$

Si de plus  $X$  et  $Y$  sont indépendantes, on a

$$\forall s \in G, \quad \mathbb{P}^S(s) = \mathbb{P}(X + Y = s) = \sum_{x \in E} \mathbb{P}(X = x) \mathbb{P}(Y = s - x) = \sum_{y \in F} \mathbb{P}(X = s - y) \mathbb{P}(Y = y).$$

*Démonstration.* On a, pour tout  $s \in G$ ,

$$\begin{aligned} \mathbb{P}(X + Y = s) &= \sum_{x \in E} \mathbb{P}(X + Y = s, X = x), \text{ d'après la formule des probabilités totales} \\ &= \sum_{x \in E} \mathbb{P}(Y = s - x, X = x), \text{ car } \{X + Y = s, X = x\} = \{X = x, Y = s - x\}. \end{aligned}$$

Le raisonnement pour la deuxième égalité est symétrique. Le cas indépendant se déduit du cas général et de la définition d'indépendance.  $\square$

Même si nous n'avons pas étudié le cas des vecteurs aléatoires à densité, vous trouverez ci-dessous la formule pour la densité de la somme de deux variables aléatoires à densité indépendantes. En effet, cette formule est très utile en pratique, et ne nécessite pas d'intégration à plusieurs variables (même si sa preuve oui). Commençons par donner la définition d'indépendance dans le cas des variables aléatoires réelles.

**Définition 5.21.** Soit  $X, Y : \Omega \rightarrow \mathbb{R}$  deux variables aléatoires réelles<sup>2</sup>. Alors  $X$  et  $Y$  sont dites *indépendantes* si, pour tout  $A \in \mathcal{B}(\mathbb{R}), B \in \mathcal{B}(\mathbb{R})$ ,

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

Voici maintenant le résultat sur la densité de la somme de deux variables aléatoires indépendantes.

**Proposition 5.22.** Soient  $X, Y$  deux variables aléatoires réelles à densité indépendantes, de densités respectives  $f^X, f^Y$ . Alors, la variable aléatoire  $Z = X + Y$  est aussi une variable aléatoire réelle à densité, de densité  $f^Z$  donnée par

$$\forall z \in \mathbb{R}, \quad f^Z(z) = \int_{\mathbb{R}} f^X(w) f^Y(z - w) dw = \int_{\mathbb{R}} f^X(z - w) f^Y(w) dw.$$

*Démonstration.* Hors programme.  $\square$

**Définition 5.23.** Les formules pour la loi de  $Z$  (cas discret) et la densité de  $Z$  (cas à densité) s'appellent le *produit de convolution* des lois de  $X$  et de  $Y$ .

EXEMPLE 5.24. En utilisant les produits de convolution, vous montrerez en exercices que la somme de deux variables aléatoires indépendantes de lois respectives :

- $\text{Bin}(n_1, p)$  et  $\text{Bin}(n_2, p)$  suit une loi binomiale  $\text{Bin}(n_1 + n_2, p)$ ,
- $\mathcal{N}(\mu, \sigma^2)$  et  $\mathcal{N}(\nu, \tau^2)$  suit une loi normale  $\mathcal{N}(\mu + \nu, \sigma^2 + \tau^2)$ .

---

2. Souvenez-vous que cela veut dire que l'on munit  $\mathbb{R}$  de la tribu borélienne  $\mathcal{B}(\mathbb{R})$

## 5.4 LOI FAIBLE DES GRANDS NOMBRES

Le but de cette section est d'aborder un résultat fondamental concernant la somme de variables aléatoires indépendantes et identiquement distribuées, qui valide l'intuition de l'approche fréquentiste de la définition d'une probabilité.

Soit  $(\Omega, \mathcal{F}, \mathbb{P})$  un espace probabilisé. Toutes les variables aléatoires considérées dans ce paragraphe sont définies sur  $\Omega$ , à valeurs réelles, discrètes ou à densité.

Dans un premier temps, définissons deux modes de convergence de suites de variables aléatoires. Vous en verrez d'autres dans votre cours de probabilités l'année prochaine.

**Définition 5.25.** Soit  $(X_n)_{n \geq 1}$  une suite de variables aléatoires. On dit que la suite  $(X_n)_{n \geq 1}$  :

1. *converge presque sûrement* vers la variable aléatoire  $X$  si,

$$\mathbb{P}[\omega \in \Omega : \lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)] = 1.$$

Dans ce cas, on note :  $X_n \rightarrow X$  p.s.

2. *converge en probabilité* vers la variable aléatoire  $X$  si,

$$\forall \varepsilon > 0, \lim_{n \rightarrow +\infty} \mathbb{P}[|X - X_n| \geq \varepsilon] = 0.$$

Dans ce cas, on note :  $X_n \xrightarrow{\mathbb{P}} X$ .

*Remarque 5.26.* La convergence presque-sûre est la plus proche de la convergence simple des fonctions mais, vu que l'on est dans un cadre probabiliste, on se permet d'exclure les ensembles de probabilité nulle. Aussi, la convergence presque sûre est plus forte que la convergence en probabilité comme vous le verrez l'année prochaine.

Nous sommes maintenant prêts pour la loi des grands nombres. Énonçons et prouvons sa version faible, nous ferons ensuite une remarque sur sa version forte.

**Théorème 5.27** (Loi faible des grands nombres.). *Soit  $(X_n)_{n \geq 1}$  une suite de variables aléatoires indépendantes et identiquement distribuées, de carré intégrable, de moyenne  $\mu$  et de variance  $\sigma^2$ . Pour tout  $n \geq 1$ , définissons la moyenne empirique, notée  $Z_n$ , par*

$$Z_n = \frac{X_1 + \cdots + X_n}{n}.$$

*Alors, la suite  $(Z_n)_{n \geq 1}$  des moyennes empiriques converge en probabilité vers  $\mu$ . Plus précisément, on a :*

$$\forall \varepsilon > 0, \quad \mathbb{P}(|Z_n - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2 n}.$$

*Démonstration.* D'après la linéarité de l'espérance on a

$$\mathbb{E}(Z_n) = \frac{\mathbb{E}(X_1) + \cdots + \mathbb{E}(X_n)}{n} = \mu.$$

De plus, comme les variables aléatoires  $(X_n)_{n \geq 1}$  sont indépendantes, on a aussi

$$\text{Var}(Z_n) = \frac{\text{Var}(X_1) + \cdots + \text{Var}(X_n)}{n^2} = \frac{\sigma^2}{n}.$$

Comme les  $(X_n)_{n \geq 1}$  sont de carré intégrable, il en est de même pour les  $(Z_n)_{n \geq 1}$ . Ainsi, d'après l'inégalité de Bienaymé-Tchebychev, on a

$$\forall \varepsilon > 0, \quad \mathbb{P}(|Z_n - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2 n}.$$

La convergence en probabilité s'obtient en prenant la limite lorsque  $n$  tend vers l'infini.  $\square$

*Remarque 5.28.*

1. On peut démontrer que ce résultat reste vrai si on suppose seulement les variables aléatoires  $(X_n)_{n \geq 1}$  intégrables (au lieu de carré intégrable).
2. Il existe également une version *forte* de la loi des grands nombres, qui démontre la convergence presque sûre de la suite  $(Z_n)_{n \geq 1}$  des moyennes empiriques vers  $\mu$ , lorsque les variables aléatoires  $(X_n)_{n \geq 1}$  sont supposées intégrables.

EXEMPLE 5.29. Considérons un schéma de Bernoulli de paramètre  $p$  où l'expérience est répétée indéfiniment. Soit  $A$  l'évènement qui représente le succès, on a donc  $\mathbb{P}(A) = p$ . Par exemple, on lance indéfiniment un dé non pipé, le succès est l'évènement  $A$  "obtenir un 6", de sorte que  $p = 1/6$ .

Soit  $(\Omega, \mathcal{F}, \mathbb{P})$  l'espace probabilisé correspondant à une épreuve et, pour tout  $n \geq 1$ , soit  $A_n$  l'évènement "obtenir un succès lors de la  $n$ -ième répétition"; alors  $\mathbb{P}(A_n) = p$ . Définissons  $X_n = \mathbb{I}_{A_n}$ . Alors la suite  $(X_n)_{n \geq 1}$  est une suite de variables aléatoires indépendantes de même loi de Bernoulli de paramètre  $p$ , de carré intégrable étant donné qu'elles ne prennent que deux valeurs. De plus, pour tout  $n \geq 1$ ,  $\mathbb{E}(X_n) = \mathbb{P}(A_n) = p$ . D'après la loi des grands nombres :

$$\frac{1}{N} \sum_{n=1}^N X_n \xrightarrow[N \rightarrow +\infty]{\mathbb{P}} \mathbb{E}(X_1) = p.$$

Interprétons ce résultat. Pour tout  $N \geq 1$ , la variable aléatoire  $\sum_{n=1}^N X_n = \sum_{n=1}^N \mathbb{I}_{A_n} = n_N(A)$  représente le nombre de fois où  $A$  est réalisé lors des  $N$  premières épreuves, et  $\frac{1}{N} \sum_{n=1}^N X_n$  représente donc la fréquence de  $A$  lors des  $N$  premières épreuves. Ainsi la loi des grands nombres peut se réécrire :

$$\frac{n_N(A)}{N} \xrightarrow[N \rightarrow +\infty]{\mathbb{P}} p.$$

Cela justifie l'approche intuitive dont nous avons parlé au début de ce cours (cf. chapitre 1). Dans l'exemple du lancer de dé, cela signifie que la fréquence du nombre de 6 obtenus tend vers  $1/6$  en probabilité.





# CHAPITRE 6

## INTRODUCTION À LA MARCHE ALÉATOIRE

Le but de ce chapitre est de faire une introduction à la marche aléatoire. Il s'agit d'une suite de variables aléatoires décrivant une marche au hasard sur un graphe. Elle revêt une importance capitale en probabilités car elle est sous-jacente à la modélisation de nombreux phénomènes tels que : le déplacement d'une particule, les cours de la bourse, les réseaux électriques, l'évolution d'une population, et de nombreux autres phénomènes.

### 6.1 DÉFINITION

**Définition 6.1.** Une *marche aléatoire* est une suite  $(S_n)_{n \geq 0}$  de variables aléatoires définie par :

$$S_n = x + \sum_{k=1}^n X_k,$$

où  $x = S_0$  est la position de la marche aléatoire à l'instant initial 0, et les variables aléatoires  $(X_k)_{k \geq 1}$  sont indépendantes et identiquement distribuées.

**EXEMPLE 6.2.** Un des exemples fameux est la marche aléatoire sur  $\mathbb{Z}^d$ . Dans ce cas s'il on note  $(e_1, \dots, e_d)$  la base canonique de  $\mathbb{Z}^d$ , les variables aléatoires  $X_k$  sont i.i.d. à valeurs dans  $\{e_1, -e_1, \dots, e_d, -e_d\}$ . La marche aléatoire est dite *symétrique* si les variables aléatoires  $X_k$  sont uniformes. Pour tout  $n \geq 1$ ,  $S_n$  représente la position de la marche aléatoire à l'instant  $n$ .

- Lorsque  $d = 1$ , les variables aléatoires  $X_k$  sont à valeurs dans  $\{-1, 1\}$  avec probabilités

$$\mathbb{P}(X_k = 1) = p, \quad \mathbb{P}(X_k = -1) = 1 - p, \quad p \in [0, 1].$$

Si  $p = \frac{1}{2}$ ,  $(S_n)_{n \geq 0}$  est la marche aléatoire symétrique sur  $\mathbb{Z}$ .

- Lorsque  $d = 2$ , les variables aléatoires  $X_k$  sont à valeurs dans  $\{(1, 0), (-1, 0), (0, 1), (0, -1)\}$ . Si l'on a,

$$\mathbb{P}(X_k = (1, 0)) = \mathbb{P}(X_k = (-1, 0)) = \mathbb{P}(X_k = (0, 1)) = \mathbb{P}(X_k = (0, -1)) = \frac{1}{4},$$

alors  $(S_n)_{n \geq 0}$  est la marche aléatoire symétrique sur  $\mathbb{Z}^2$ .

**Notation.** On note aussi  $\mathbb{P}_x(S_n = \cdot)$  pour la loi de la marche aléatoire qui est en  $x$  lors de l'état initial.

## 6.2 RETOURS EN ZÉRO

Une des premières questions qui se pose est de savoir, si la marche aléatoire part de 0, quelle est la probabilité qu'elle retourne en zéro. Une autre question est de calculer l'espérance du nombre de visites en zéro. Ces notions sont sous-jacentes aux notions de récurrence/transience de la marche aléatoire que vous reverrez plus tard.

La proposition suivante donne le calcul de la probabilité de retour en 0 dans le cas de la marche aléatoire sur  $\mathbb{Z}^d$  lorsque  $d = 1, 2$ .

**Proposition 6.3.** *Soit  $(S_n)_{n \geq 0}$  la marche aléatoire symétrique sur  $\mathbb{Z}^d$ , avec  $d = 1, 2$ . Alors, la probabilité de retour en 0 est nulle si le nombre de pas est impair et, pour tout  $n \geq 1$ ,*

$$\mathbb{P}_0(S_{2n} = 0) = \begin{cases} \binom{2n}{n} \frac{1}{2^{2n}} & \text{si } d = 1 \\ \binom{2n}{n}^2 \frac{1}{2^{4n}} & \text{si } d = 2. \end{cases}$$

*Démonstration.* Notons que si  $d = 1$ , pour que la marche aléatoire retourne en 0, elle doit faire autant de pas vers la gauche que vers la droite, ainsi le nombre total de pas doit être pair. De manière similaire, si  $d = 2$ , elle doit faire autant de pas vers le haut que vers le bas, respectivement vers la droite que vers la gauche, ainsi le nombre de pas doit être pair également.

**Preuve pour  $d = 1$ .** Remarquons que, pour tout  $k \geq 1$ ,  $X_k = 2Y_k - 1$  où  $Y_k \sim \text{Ber}(\frac{1}{2})$ . Comme les  $(X_k)_{k \geq 1}$  sont indépendantes, les  $(Y_k)_{k \geq 1}$  le sont aussi et ainsi,  $\sum_{k=1}^n Y_k$  suit une loi binomiale de paramètres  $n, \frac{1}{2}$ . En conséquence, pour tout  $n \geq 1$ ,

$$\begin{aligned} \mathbb{P}_0(S_{2n} = 0) &= \mathbb{P}\left(\sum_{k=1}^{2n} X_k = 0\right) = \mathbb{P}\left(\sum_{k=1}^{2n} (2Y_k - 1) = 0\right) = \mathbb{P}\left(\sum_{k=1}^{2n} Y_k = n\right) \\ &= \binom{2n}{n} \frac{1}{2^{2n}}, \text{ car } \sum_{k=1}^{2n} Y_k \sim \text{Bin}(2n, 1/2). \end{aligned}$$

**Preuve pour  $d = 2$ .** La marche aléatoire  $(S_n)_{n \geq 0}$  sur  $\mathbb{Z}^2$  est en bijection avec la marche aléatoire  $(\tilde{S}_n)_{n \geq 0}$  sur  $\tilde{\mathbb{Z}}^2$ , où  $\tilde{\mathbb{Z}}^2$  est le réseau  $\mathbb{Z}^2$  tourné de  $45^\circ$  avec longueur des arêtes multipliée par  $\sqrt{2}$ . Si la marche débute en  $0 = (0, 0)$  on a, pour tout  $n \geq 1$ ,  $\tilde{S}_n = \sum_{k=1}^n \tilde{X}_k$  où les couples aléatoires  $(\tilde{X}_k)_{k \geq 1}$  sont i.i.d. à valeurs dans  $\{(1, 1), (-1, 1), (1, -1), (-1, -1)\}$  de loi uniforme.

Soit  $k \geq 1$ , alors  $\tilde{X}_k = (\tilde{X}_k^1, \tilde{X}_k^2)$ , où chacune des composantes est à valeurs dans  $\{-1, 1\}$ . Calculons la loi marginale de  $\tilde{X}_k^1$ . On a,

$$\begin{aligned} \mathbb{P}(\tilde{X}_k^1 = 1) &= \mathbb{P}(\tilde{X}_k = (1, 1)) + \mathbb{P}(\tilde{X}_k = (1, -1)) = \frac{1}{2}, \\ \mathbb{P}(\tilde{X}_k^1 = -1) &= \mathbb{P}(\tilde{X}_k = (-1, 1)) + \mathbb{P}(\tilde{X}_k = (-1, -1)) = \frac{1}{2}, \end{aligned}$$

et de manière analogue pour  $\tilde{X}_k^2$ . Ainsi,  $\tilde{X}_k^1$ , resp.  $\tilde{X}_k^2$ , suit une loi uniforme sur  $\{-1, 1\}$ . De plus, pour tout  $(j, \ell) \in \{-1, 1\}^2$ ,

$$\mathbb{P}(\tilde{X}_k = (j, \ell)) = \mathbb{P}(\tilde{X}_k^1 = j) \mathbb{P}(\tilde{X}_k^2 = \ell) = \frac{1}{4},$$

et les composantes  $\tilde{X}_k^1, \tilde{X}_k^2$  sont donc indépendantes.

Notons  $\tilde{S}_n^1$ , resp.  $\tilde{S}_n^2$ , la première, resp. deuxième composante de  $\tilde{S}_n$ . Alors, pour tout  $n \geq 1$ ,

$$\tilde{S}_n = (\tilde{S}_n^1, \tilde{S}_n^2) = \left( \sum_{k=1}^n \tilde{X}_k^1, \sum_{k=1}^n \tilde{X}_k^2 \right).$$

Comme pour tout  $k \geq 1$ ,  $\tilde{X}_k^1, \tilde{X}_k^2 \sim \mathcal{U}(\{-1, 1\})$ , chacune des composantes  $\tilde{S}_n^1, \tilde{S}_n^2$  est une marche aléatoire symétrique sur  $\mathbb{Z}$  débutant en 0. De l'indépendance des  $(\tilde{X}_k)_{k \geq 1}$  et de  $\tilde{X}_k^1, \tilde{X}_k^2$  pour tout  $k \geq 1$ , on déduit de plus que les composantes  $\tilde{S}_n^1, \tilde{S}_n^2$  sont indépendantes. Ainsi, pour tout  $n \geq 1$ ,

$$\begin{aligned} \mathbb{P}_0(S_{2n} = (0, 0)) &= \mathbb{P}_0(\tilde{S}_{2n} = (0, 0)) \\ &= \mathbb{P}_0(\tilde{S}_{2n}^1 = 0, \tilde{S}_{2n}^2 = 0) \\ &= \mathbb{P}_0(\tilde{S}_{2n}^1 = 0) \mathbb{P}_0(\tilde{S}_{2n}^2 = 0), \text{ par indépendance de } \tilde{S}_n^1, \tilde{S}_n^2 \\ &= \binom{2n}{n}^2 \frac{1}{2^{4n}}, \text{ d'après le calcul en dimension 1.} \end{aligned} \quad \square$$

Notons  $N_0 = \sum_{n=1}^{\infty} \mathbb{I}_{\{S_n=0\}}$  le nombre de visite en 0 de la marche aléatoire. Le corollaire suivant montre que l'espérance de  $N_0$  pour la marche aléatoire symétrique sur  $\mathbb{Z}^d$  est infinie lorsque  $d = 1, 2$ . A noter que cette espérance devient finie lorsque  $d \geq 3$ . L'intuition est la suivante : lorsque la dimension augmente, il devient de plus en plus difficile pour la marche aléatoire de revenir en 0, une fois qu'elle est partie.

**Corollaire 6.4.** *On considère la marche aléatoire symétrique sur  $\mathbb{Z}^d$ , avec  $d = 1, 2$ . Alors, l'espérance du nombre de visites en 0 est infinie, autrement dit :*

$$\mathbb{E}(N_0) = \infty.$$

*Démonstration.* Remarquons que  $N_0 = \lim_{\ell \rightarrow \infty} \sum_{n=1}^{\ell} \mathbb{I}_{\{S_n=0\}}$ . Or, la suite  $(N_0^\ell)_{\ell \geq 1} := (\sum_{n=1}^{\ell} \mathbb{I}_{\{S_n=0\}})_{\ell \geq 1}$  est une suite croissante de variables aléatoires positives, ainsi, d'après le théorème de convergence monotone,

$$\mathbb{E}_0(N_0) = \lim_{\ell \rightarrow \infty} \mathbb{E}_0(N_0^\ell) = \lim_{\ell \rightarrow \infty} \sum_{n=1}^{\ell} \mathbb{P}_0(S_n = 0) = \sum_{n=1}^{\infty} \mathbb{P}_0(S_n = \ell).$$

D'après la Proposition 6.3, les probabilités de la somme sont nulles lorsque  $n$  est impair, d'où  $\mathbb{E}_0(N_0) = \sum_{n=1}^{\infty} \mathbb{P}_0(S_{2n} = \ell)$ . De plus, d'après la formule de Stirling, on a

$$n! \sim \sqrt{2\pi n} \left(\frac{n}{e}\right)^n,$$

d'où

$$\binom{2n}{n} \frac{1}{2^{2n}} = \frac{(2n)!}{n!n!} \frac{1}{2^{2n}} \sim \frac{\sqrt{4\pi n} (2n)^{2n} e^{2n}}{2\pi n \cdot n^{2n} e^{2n}} \frac{1}{2^{2n}} \sim \frac{1}{\sqrt{\pi n}}.$$

Ainsi, lorsque  $d = 1$ , le terme général  $\mathbb{P}_0(S_{2n} = \ell)$  de la série  $\mathbb{E}(N_0)$  est équivalent en l'infini à :  $\frac{1}{\sqrt{\pi n}}$  lorsque  $d = 1$ , et  $\frac{1}{\pi n}$  lorsque  $d = 2$ . Dans les deux cas, il s'agit de termes généraux de séries divergentes, et on en déduit donc le résultat.  $\square$

### 6.3 RUINE DE LA JOUEUSE

Voici un autre problème classique qui utilise les marches aléatoires. Une joueuse  $A$  lance une pièce de manière répétée, la probabilité d'obtenir pile est  $p$ ,  $p \in [0, 1]$ . Quand la pièce tombe sur pile, elle reçoit 1 euro de son adversaire  $B$ , quand la pièce tombe sur face, elle donne 1 euro à son adversaire  $B$ . La fortune initiale de la joueuse  $A$  est  $a$ , avec  $a \in \mathbb{N}^*$ , celle de l'adversaire  $B$  est  $b$ , avec  $b \in \mathbb{N}^*$ . Le jeu s'arrête dès que l'une des joueuses n'a plus d'argent. Modélisons la fortune  $S_n$  de la joueuse  $A$  au temps  $n$  à l'aide d'une marche aléatoire. Pour tout  $n \geq 1$ , on a

$$S_n = a + \sum_{k=1}^n X_k,$$

où les  $(X_k)$  sont i.i.d à valeurs dans  $\{-1, 1\}$ , avec  $\mathbb{P}(X_k = 1) = p$ ,  $\mathbb{P}(X_k = -1) = 1 - p := q$ . A noter que cette formule n'est vraie que jusqu'à la fin du jeu, c'est-à-dire jusqu'à ce que  $S_n = 0$  (la joueuse  $A$  est ruinée) ou  $S_n = a + b$  (l'adversaire  $B$  est ruinée), après quoi  $S_n$  reste constante. On dit que les états 0 et  $a + b$  sont *absorbants*.

Notons  $R$  l'évènement "la joueuse  $A$  est finalement ruinée". Le but est de déterminer  $u_k := \mathbb{P}_k(R)$ , c'est-à-dire la probabilité pour la joueuse  $A$  d'être ruinée lorsqu'elle commence avec une fortune initiale de  $k$ . D'après la formule des probabilités totales, on a :

$$\begin{aligned} \mathbb{P}_k(R) &= \mathbb{P}_k(R|X_1 = 1)\mathbb{P}(X_1 = 1) + \mathbb{P}_k(R|X_1 = -1)\mathbb{P}(X_1 = -1) \\ &= \mathbb{P}_{k+1}(R|X_1 = 1)p + \mathbb{P}_{k-1}(R|X_1 = -1)q. \end{aligned}$$

En effet, si le premier jet donne pile, respectivement face, le jeu se déroule comme si elle était partie de la fortune  $k + 1$ , respectivement  $k - 1$ . Ainsi, les probabilités  $(u_k)$  satisfont la récurrence linéaire suivante :

$$\begin{cases} u_k = pu_{k+1} + qu_{k-1}, & 1 \leq k \leq a + b - 1 \\ u_0 = 1, & u_{a+b} = 0. \end{cases}$$

Afin de résoudre cette récurrence linéaire, considérons l'équation caractéristique associée :

$$pr^2 - r + q = 0,$$

le discriminant vaut  $\Delta = 1 - 4pq = (2p - 1)^2$  et les solutions sont  $r_1 = 1, r_2 = q/p$ . On distingue le cas où

- $p \neq q$ , i.e.  $p \neq \frac{1}{2}$ . Les solutions  $(u_k)$  sont de la forme  $u_k = \alpha r_1^k + \beta r_2^k$ . Les conditions au bord impliquent  $\alpha + \beta = 1$ ,  $\alpha + \beta(\frac{q}{p})^{a+b} = 0$ . Ainsi on trouve,

$$\begin{aligned} \alpha &= -\frac{(q/p)^{a+b}}{1 - (q/p)^{a+b}}, & \beta &= \frac{1}{1 - (q/p)^{a+b}} \\ \Rightarrow \quad \forall 0 \leq k \leq a + b, & \quad u_k &= \frac{(q/p)^{a+b} - (q/p)^k}{(q/p)^{a+b} - 1}. \end{aligned}$$

- $p = q$ , i.e.  $p = \frac{1}{2}$ . Alors  $r_1 = r_2 := r = 1$  et les solutions sont de la forme  $u_k = r^k(\alpha + k\beta) = \alpha + k\beta$ . Les conditions au bord impliquent  $\alpha = 1$ ,  $\alpha + (a + b)\beta = 0$ . Ainsi  $\beta = -1/(a + b)$  et la solution est

$$0 \leq k \leq a + b, \quad u_k = 1 - \frac{k}{a + b}.$$

A noter qu'une approche similaire permet de calculer l'espérance de la durée du jeu avant la ruine d'une des deux joueuses.