

# Probabilités et introduction à la statistique

# Table des matières

<b>1</b>	<b>Un peu de combinatoire</b>	<b>3</b>
1.1	Trois formules classiques de dénombrement . . . . .	3
1.2	Propriétés des coefficients binomiaux . . . . .	6
<b>2</b>	<b>Probabilités finies</b>	<b>9</b>
2.1	Expériences aléatoires . . . . .	9
2.2	Mesures de probabilité . . . . .	10
2.3	Variables aléatoires finies . . . . .	13
2.3.1	Espérance d'une variable aléatoire finie . . . . .	15
2.3.2	Variance d'une variable aléatoire finie . . . . .	16
<b>3</b>	<b>Introduction aux statistiques descriptives</b>	<b>18</b>
3.1	Un peu de vocabulaire . . . . .	18
3.2	Représentation graphique des variables statistiques . . . . .	19
3.2.1	Diagrammes en barre et circulaires . . . . .	19
3.2.2	Classes et histogrammes . . . . .	19
3.2.3	La courbe des effectifs cumulés . . . . .	24
3.3	Indicateurs statistiques . . . . .	24
3.3.1	Les indicateurs de tendance centrale . . . . .	24
3.3.2	Paramètres de position . . . . .	26
3.3.3	Paramètres de dispersion . . . . .	26
<b>4</b>	<b>Indépendance et conditionnement</b>	<b>28</b>
4.1	Probabilités conditionnelles et indépendance d'événements . . . . .	28
4.2	Indépendance de variables aléatoires finies . . . . .	31
4.3	La loi faible des grands nombres . . . . .	33
<b>5</b>	<b>Probabilités discrètes</b>	<b>35</b>
5.1	Rappels sur la dénombrabilité . . . . .	35
5.2	Rappels sur les séries . . . . .	36
5.3	Probabilités sur les ensembles dénombrables . . . . .	39
5.4	Quelques lois classiques . . . . .	41
5.5	Variables aléatoires sur les espaces dénombrables . . . . .	43
5.5.1	Espérance et variance des variables aléatoires sur les espaces dénombrables . . . . .	43
5.5.2	Variables indépendantes . . . . .	45
<b>6</b>	<b>Variables aléatoires à densité</b>	<b>46</b>
6.1	Définition d'une variable à densité . . . . .	46
6.1.1	Mesures de probabilités sur les ensembles quelconques . . . . .	46
6.1.2	Intégration de fonctions positive . . . . .	47
6.1.3	Densité de probabilité . . . . .	47
6.1.4	Fonction de répartition . . . . .	49
6.2	Espérance et variance d'une variable aléatoire à densité . . . . .	51
6.3	Le théorème central limite . . . . .	54
<b>7</b>	<b>Fluctuation d'échantillonnage et intervalle de confiance</b>	<b>56</b>

# Chapitre 1

## Un peu de combinatoire

### 1.1 Trois formules classiques de dénombrement

#### Nombre de manières d'ordonner un ensemble fini

**Problème concret** : Mon équipe de foot est constituée de 11 joueurs, et j'ai 11 maillots, numérotés de 1 à 11. De combien de manière puis-je distribuer les maillots ?

**Problème mathématique** : Soit  $E$  un ensemble fini, formé de  $n$  éléments :  $E = \{x_1, x_2, \dots, x_n\}$ . De combien de manière peut-on ordonner les éléments de  $E$  ?

Par exemple, considérons l'ensemble à deux éléments  $E = \{x, y\}$ . Il existe deux manières de l'ordonner :  $(x, y)$  et  $(y, x)$ .

L'ensemble à trois éléments  $F = \{a, b, c\}$  peut être ordonné des manières suivantes :

$$(a, b, c); (a, c, b); (b, a, c); (b, c, a); (c, a, b); (c, b, a).$$

Il y a donc 6 manières de faire.

Pour trouver une formule générale, on a besoin de la définition suivante.

**Définition 1.1.** Soit  $n$  un entier positif. Le nombre  $n!$  est défini comme

$$n! := n \times (n - 1) \times (n - 2) \times \dots \times 3 \times 2 \times 1.$$

Par convention, on pose  $0! = 1$ .

**Remarque 1.1.** Cette définition peut aussi s'écrire comme  $n! = \prod_{k=1}^n k$ .

On a le résultat suivant :

**Théorème 1.1.** Soit  $E$  un ensemble fini de cardinal  $n$ . Il existe  $n!$  manières d'ordonner les éléments de  $E$ .

*Démonstration.* Pour ordonner un ensemble à  $n$  éléments, il faut commencer par choisir le premier : on a  $n$  possibilités. On choisit ensuite le second, et on a  $(n - 1)$  possibilité. Pour le troisième, on a  $n - 2$  possibilités, etcætera. En tout, on a donc bien  $n!$  possibilités.  $\square$

**Réponse au problème concret.** Distribuer les maillots, c'est ordonner mon ensemble de 11 joueurs : c'est choisir qui est le joueur numéro 1, qui est le numéro 2, ... Par le théorème précédent, il existe  $11!$  manières de faire, soit 39916800 (d'après ma calculatrice).

**Remarque 1.2.** Si  $E$  est un ensemble de cardinal  $n$ ,  $n!$  est aussi

- le nombre de permutations de  $E$ , c'est-à-dire le nombre de bijections de  $E$  dans  $E$  ;
- le nombre de bijections de  $\{1, \dots, n\}$  dans  $E$ .

## Nombre d'arrangements

**Problème concret** : 10 chevaux participent à une course. Pour jouer au tiercé, on doit prédire quel cheval arrivera premier, quel cheval arrivera deuxième, et quel cheval arrivera troisième. Combien y a-t-il de possibilités ?

**Problème mathématique** : Soit  $E$  un ensemble fini, formé de  $n$  éléments, et soit  $p \leq n$ . De combien de manière peut-on choisir une suite de  $p$  éléments de  $E$  distincts ?

Par exemple, si  $E = \{a, b, c, d\}$  et si  $p = 2$ , il existe 12 manières de faire :

$$(a, b); (a, c); (a, d); (b, a); (b, c); (b, d); (c, a); (c, b); (c, d); (d, a); (d, b); (d, c).$$

**Définition 1.2.** Soit  $E$  un ensemble fini, formé de  $n$  éléments, et soit  $p \leq n$ . Un **arrangement** de  $p$  éléments de  $E$  est le choix d'une suite de  $p$  éléments de  $E$  distincts.

Remarquons qu'un arrangement de  $n$  éléments de  $E$  est simplement une manière d'ordonner  $E$ . On a le résultat suivant :

**Théorème 1.2.** Soit  $E$  un ensemble fini de cardinal  $n$  et soit  $p \leq n$ . Le nombre d'arrangements de  $p$  éléments de  $E$  est donné par :

$$A_n^p := \frac{n!}{(n-p)!} = n \times (n-1) \times \dots \times (n-p+1).$$

*Démonstration.* Pour choisir un arrangement de  $p$  éléments de  $E$ , je commence par choisir le premier élément : j'ai  $n$  possibilités. Je choisis ensuite le 2ème, et j'ai  $n-1$  possibilités. Je continue ainsi jusqu'à choisir le  $p$ -ième élément, pour lequel j'ai  $n-p+1$  possibilités. On a donc bien  $n \times (n-1) \times \dots \times (n-p+1)$  possibilités.

En remarquant que

$$\frac{n!}{(n-p)!} = \frac{1 \times 2 \times \dots \times (n-p-1) \times (n-p) \times (n-p+1) \times \dots \times (n-1) \times n}{1 \times 2 \times \dots \times (n-p-1) \times (n-p)}$$

et en simplifiant les  $(n-p)$  premiers termes, on trouve bien que  $n \times (n-1) \times \dots \times (n-p+1) = \frac{n!}{(n-p)!}$ .  $\square$

**Réponse au problème concret.** Au tiercé, il faut choisir une suite de 3 chevaux (tous distincts) parmi 10. Le nombre de possibilités est donc

$$A_{10}^3 = 10 \times 9 \times 8 = 720.$$

## Nombre de combinaisons

**Problème concret** : Dans certaines variantes du poker, chaque joueur commence avec une main de 2 cartes, sur un jeu de 52 cartes. Combien de mains de départ différentes existe-t-il ?

**Problème mathématique** : Soit  $E$  un ensemble fini, formé de  $n$  éléments, et soit  $p \leq n$ . De combien de manière peut-on choisir  $p$  éléments de  $E$  distincts ?

Attention, ça n'est pas le même problème que précédemment : ici, l'ordre des  $p$  éléments choisis n'est pas pris en compte.

Par exemple, si  $E = \{a, b, c, d\}$  et  $p = 2$ , il existe 6 manières de faire :

$$\{a, b\}; \{a, c\}; \{a, d\}; \{b, c\}; \{b, d\}; \{c, d\}.$$

**Définition 1.3.** Soit  $E$  un ensemble fini, formé de  $n$  éléments, et soit  $p \leq n$ . Une **combinaison** de  $p$  éléments de  $E$  est le choix de  $p$  éléments de  $E$  distincts.

**Théorème 1.3.** Soit  $E$  un ensemble fini de cardinal  $n$  et soit  $p \leq n$ . Le nombre de combinaisons de  $p$  éléments de  $E$  est donné par :

$$\binom{n}{p} := \frac{n!}{p!(n-p)!}.$$

Le nombre  $\binom{n}{p}$  est aussi noté  $C_n^p$ . Ces nombres sont appelés les coefficients binomiaux.

*Démonstration.* À partir d'une combinaison de  $p$  éléments de  $E$ , je peux construire plusieurs arrangements de  $p$  éléments de  $E$  : un pour chaque manière d'ordonner les  $p$  éléments choisis. Par conséquent, chaque combinaison de  $p$  éléments de  $E$  permet de construire  $p!$  arrangements tous différents. De plus, les ensembles d'arrangements pouvant être construits à partir de deux combinaisons différentes sont disjoints. On en déduit que

$$A_n^p = p! \binom{n}{p},$$

et donc que  $\binom{n}{p} := \frac{n!}{p!(n-p)!}$ . □

**Remarque 1.3.** La démonstration ci-dessus, de nature combinatoire, montre que  $\binom{n}{p}$  est un nombre entier. Il n'est pas évident de le montrer à partir de la définition et de propriétés arithmétiques !

**Réponse au problème concret.** Le nombre de mains initiales possibles au poker est le nombre de combinaisons de 2 éléments d'un ensemble à 52 éléments. Par le théorème précédent, cela vaut

$$\binom{52}{2} = \frac{52!}{2!50!} = \frac{51 \times 52}{2} = 1326.$$

## 1.2 Propriétés des coefficients binomiaux

### Symétrie des coefficients

**Proposition 1.1.** Soit  $n \geq 1$  un entier, et soit  $0 \leq p \leq n$  un entier. On a

$$\binom{n}{p} = \binom{n}{n-p}$$

*Démonstration combinatoire.* Choisir  $p$  éléments parmi  $n$ , c'est équivalent à choisir  $(n-p)$  éléments (ceux qu'on ne prend pas). On a donc bien  $\binom{n}{p} = \binom{n}{n-p}$ .  $\square$

*Démonstration calculatoire.* On a

$$\binom{n}{p} = \frac{n!}{p!(n-p)!} = \frac{n!}{(n-p)!p!} = \binom{n}{n-p}.$$

$\square$

### Quelques valeurs à connaître

Pour tout  $n \geq 1$ , on a

$$\begin{aligned}\binom{n}{0} &= \binom{n}{n} = 1 \\ \binom{n}{1} &= \binom{n}{n-1} = n.\end{aligned}$$

### Le triangle de Pascal

**Proposition 1.2** (La formule de Pascal). Soit  $n \geq 2$  un entier, et soit  $0 \leq p \leq n-1$  un entier. On a

$$\binom{n}{p} = \binom{n-1}{p-1} + \binom{n-1}{p}.$$

*Démonstration calculatoire.* On a

$$\begin{aligned}\binom{n-1}{p-1} + \binom{n-1}{p} &= \frac{(n-1)!}{(p-1)!(n-p)!} + \frac{(n-1)!}{p!(n-1-p)!} \\ &= \frac{(n-1)!}{(p-1)!(n-p-1)!(n-p)} + \frac{(n-1)!}{(p-1)!p(n-1-p)!} \\ &= \frac{(n-1)!}{(p-1)!(n-p-1)!} \left( \frac{1}{n-p} + \frac{1}{p} \right) \\ &= \frac{(n-1)!}{(p-1)!(n-p-1)!} \times \frac{n}{(n-p)p} \\ &= \frac{n!}{p!(n-p)!} = \binom{n}{p}.\end{aligned}$$

$\square$

*Démonstration combinatoire.* Soit  $E$  un ensemble à  $n$  éléments, et soit  $x$  un élément quelconque de  $E$ .  $\binom{n}{p}$  compte le nombre de parties de  $E$  ayant  $p$  éléments. Il y a deux types de parties de  $E$  à  $p$  éléments : celles contenant  $x$ , et celles ne contenant pas  $x$ .

Les parties de  $E$  à  $p$  éléments ne contenant pas  $x$  sont aussi des parties de  $E \setminus \{x\}$  contenant  $p$  éléments. Il y en a donc  $\binom{n-1}{p}$ , car  $E \setminus \{x\}$  contient  $n-1$  éléments.

Il y a  $\binom{n-1}{p-1}$  parties de  $E$  à  $p$  éléments contenant  $x$  : elles sont toutes construites en ajoutant  $\{x\}$  à une partie de  $E \setminus \{x\}$  ayant  $p-1$  éléments.

On a donc bien  $\binom{n}{p} = \binom{n-1}{p-1} + \binom{n-1}{p}$ . □

La formule de Pascal nous incite à calculer les coefficients binomiaux en les écrivant dans un triangle comme suit.

- On commence par remplir les cases où  $p = 0$  et celles où  $p = 1$  de 1.
- On remplit les cases où  $p > n$  de zéros (par convention).
- Puis on remplit les autres cases de haut en bas, grâce à la règle suivante : chaque case est la somme de la case au dessus et de la case au dessus à gauche.

n \ P	0	1	2	3	4	5	6
1	1	1	0	0	0	0	0
2	1	2	1	0	0	0	0
3	1	3	3	1	0	0	0
4	1	4	6	4	1	0	0
5	1	5	10	10	5	1	0
6	1	6	15	20	15	6	1

**Un peu d'histoire :** Ce n'est pas Pascal qui a inventé les coefficients binomiaux, ni la « formule de Pascal ». Celles-ci ont été découvertes pour la première fois par les mathématiciens arabes, au début du Moyen-âge, puis ont ensuite été retrouvées par les mathématiciens chinois à la fin du Moyen-âge, et par les mathématiciens européens, au début de la Renaissance. Mais c'est Pascal qui a montré le premier certaines propriétés des coefficients binomiaux par des récurrences rigoureuses.

La combinatoire est née quand les mathématiciens et les grammairiens arabes se sont posé la question suivante : Sachant que l'alphabet arabe comporte 28 consonnes, et que chaque mot arabe est composé de 3 consonnes, combien de mots peut-on former en arabe ?

## La formule du binôme de Newton

Vous connaissez la formule  $(a+b)^2 = a^2 + 2ab + b^2$ . Celle-ci se généralise à des puissances plus grande que 2 : c'est la formule dite du binôme de Newton.

**Théorème 1.4.** Soit  $n \geq 1$  un entier. Pour tout  $x, y \in \mathbb{R}$ , on a

$$(x+y)^n = \sum_{k=0}^n \binom{n}{k} x^k y^{n-k}.$$

*Démonstration.* On a  $(x+y)^n = (x+y) \times (x+y) \times \dots \times (x+y)$ . En développant le produit, je vais obtenir des termes qui sont tous de la forme  $x^k y^{n-k}$ , pour des  $k$  entre 0 et  $n$ . Le nombre de fois que je vais obtenir le terme  $x^k y^{n-k}$  est égal au nombre de manière de choisir  $k$  éléments parmi  $n$ , et donc à  $\binom{n}{k}$ . On en déduit donc bien que  $(x+y)^n = \sum_{k=0}^n \binom{n}{k} x^k y^{n-k}$ . □

## Complément : la formule de Stirling

**Problème mathématique :** Lorsqu'on fait de la combinatoire, on manipule souvent des factorielles de grands nombres. Celles-ci peuvent être très difficiles à calculer, car elles nécessitent de nombreuses étapes de calcul, et car les nombres considérés sont trop grands pour votre calculatrice. Que vaut  $(1000!)$  ? Elle est incapable de répondre ! Elle ne peut même pas vous dire combien de chiffres il y a dans ce nombre...

Heureusement, il existe une formule permettant d'approcher  $(n!)$  lorsque  $n$  est grand.

**Proposition 1.3** (La formule de Stirling). *On a*

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

Nous ne donnerons pas de démonstration de cette formule (peut-être en verrez-vous une dans votre cours d'analyse). Rappelons simplement que, si  $u_n$  et  $v_n$  sont des suites strictement positives,  $u_n \sim v_n$  signifie que  $\frac{u_n - v_n}{u_n} \rightarrow 0$ .



# Chapitre 2

## Probabilités finies

### 2.1 Expériences aléatoires

**Définition 2.1.** Une **expérience aléatoire** est une expérience, ou un phénomène, dont l'issue ne peut pas être prédite avec certitude.

L'**univers** d'une expérience (ou **univers probabiliste**) est l'ensemble des issues possibles de cette expérience aléatoire.

Un univers probabiliste est traditionnellement noté par la lettre  $\Omega$  (oméga majuscule), et ses éléments sont traditionnellement notés  $\omega$  (omega minuscule). Dans tout ce chapitre,  $\Omega$  sera un ensemble fini.

**Exemple 2.1.** • Le lancer d'une pièce est une expérience probabiliste. Son univers est  $\{\text{pile}, \text{face}\}$ .

- Le lancer d'un dé a pour univers  $\{1, 2, 3, 4, 5, 6\}$ .
- La note que vous obtiendrez à l'examen est une expérience probabiliste. Son univers est l'ensemble des entiers (et demi-entiers) compris entre 0 et 20.
- Le temps que mettre un atome radioactif à se désintégrer est aussi une expérience probabiliste. A priori, une telle désintégration peut se produire n'importe quand dans le futur ; son univers est donc  $[0, +\infty[$ . Cet univers n'est pas fini : de telles expériences aléatoires ne seront pas traitées dans ce chapitre.

Il faut parfois réfléchir un peu pour déterminer l'univers d'une expérience probabiliste. Supposons par exemple que je lance 2 dés. Si les deux dés sont identiques et si je les lance en même temps, obtenir 1 sur le premier dé et 3 sur le second, c'est la même chose que d'obtenir 3 sur le premier et 1 sur le troisième. L'univers probabiliste de cette expérience sera donc

$$\Omega = \left\{ \{1, 1\}, \{1, 2\}, \{1, 3\}, \{1, 4\}, \{1, 5\}, \{1, 6\}, \{2, 2\}, \{2, 3\}, \{2, 4\}, \{2, 5\}, \right. \\ \left. \{2, 6\}, \{3, 3\}, \{3, 4\}, \{3, 5\}, \{3, 6\}, \{4, 4\}, \{4, 5\}, \{4, 6\}, \{5, 5\}, \{5, 6\}, \{6, 6\} \right\}.$$

En revanche, si les dés sont différents, ou si je les lance l'un après l'autre, alors obtenir 1 sur le premier dé et 3 sur le second, ça ne sera pas la même chose que d'obtenir 3 sur le premier et 1 sur le troisième. L'univers probabiliste du lancer des deux dés sera alors

$$\Omega = \left\{ (1, 1), (1, 2), (1, 3), (1, 4), (1, 5), (1, 6), (2, 1), (2, 2), (2, 3), (2, 4), (2, 5), (2, 6), \right. \\ (3, 1), (3, 2), (3, 3), (3, 4), (3, 5), (3, 6), (4, 1), (4, 2), (4, 3), (4, 4), (4, 5), (4, 6), \\ \left. (5, 1), (5, 2), (5, 3), (5, 4), (5, 5), (5, 6), (6, 1), (6, 2), (6, 3), (6, 4), (6, 5), (6, 6) \right\}.$$

Remarquons que cet ensemble est  $\{1, 2, 3, 4, 5, 6\} \times \{1, 2, 3, 4, 5, 6\}$  au sens du produit cartésien.

**Définition 2.2.** Soit  $\Omega$  un univers probabiliste fini. Un **événement** est une partie de  $\Omega$ . Autrement dit, un événement est un ensemble d'issues possibles d'une expérience aléatoire.

Si  $A$  et  $B$  sont des parties de  $\Omega$  telles que  $A \cap B = \emptyset$ , on dit que les événements  $A$  et  $B$  sont **incompatibles**.

Dans toute la suite, si  $A$  est un événement de  $\Omega$ , on notera  $A^c$  son complémentaire, c'est-à-dire  $A^c = \Omega \setminus A$ . On notera  $\text{Card}(A)$  le cardinal de  $A$ , c'est-à-dire le nombre d'éléments de  $A$ .

**Exemple 2.2.** Voici quelques exemples d'événements :

- J'obtiens pile en lançant une pièce.
- J'obtiens un résultat pair en lançant un dé à 6 faces.
- Vous obtenez plus que 15/20 à votre examen.
- Nous n'avons défini les événements que pour les univers finis, pas pour les univers infinis. « L'atome radioactif met entre 2 et 3 secondes à se désintégrer » sera sans doute un événement, car il est possible de le mesurer. « Le temps que met l'atome radioactif à se désintégrer est un nombre rationnel » sera-t-il un événement ? En tout cas, ça n'est pas facile à mesurer...

## 2.2 Mesures de probabilité

Une mesure de probabilité est une manière d'associer un nombre entre 0 et 1 à chaque événement.

**Définition 2.3.** Soit  $\Omega$  un ensemble fini. Une **mesure de probabilité** sur  $\Omega$  est une application  $\mathbb{P} : \mathcal{P}(\Omega) \rightarrow [0, 1]$  vérifiant les propriétés suivantes :

- $\mathbb{P}(\Omega) = 1$ .
- Pour tout  $A, B \in \mathcal{P}(\Omega)$  tels que  $A \cap B = \emptyset$ , on a  $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$ .

On dit alors que  $(\Omega, \mathbb{P})$  est un **espace de probabilité (fini)**.

Si  $A \subset \Omega$ ,  $\mathbb{P}(A)$  est alors appelé la **probabilité de l'évènement  $A$** .

La seconde propriété ci-dessus s'appelle la *propriété d'additivité* des mesures de probabilités.

L'exemple le plus important de mesure de probabilité sur un univers fini est la mesure uniforme :

**Définition 2.4.** Soit  $\Omega$  un ensemble fini de cardinal  $N$ . La **mesure uniforme** sur  $\Omega$  est donnée par

$$\forall A \subset \Omega, \quad \mathbb{P}(A) = \frac{\text{Card}(A)}{N}.$$

Cette mesure de probabilité sera parfois notée  $\mathbb{P}_{unif}$ .

Cette loi est utilisée pour modéliser des situations où toutes les issues d'une expérience aléatoire ont la même probabilité de se produire. Par exemple, le lancer d'un dé à 6 faces non truqué sera modélisé par l'espace de probabilité  $(\{1, 2, 3, 4, 5, 6\}, \mathbb{P}_{unif})$ .

### Lien entre mesure uniforme et dénombrement

**Problème concret :** Dans certaines variantes du poker, chaque joueur commence avec une main de 2 cartes, sur un jeu de 52 cartes. Quelle est la probabilité de commencer avec une paire d'as ?

Ici, l'expérience aléatoire considérée est la pioche de deux cartes dans un jeu de 52 cartes. L'univers probabiliste est donc l'ensemble  $\Omega$  de toutes les combinaisons de deux cartes pouvant être piochées dans un jeu de 52 cartes. Aucune combinaison de deux cartes n'étant a priori plus probable que les autres (personne ne triche!),  $\Omega$  est muni de la probabilité uniforme. L'évènement  $A$  est la pioche de deux as dans un jeu de 52 cartes. On a donc

$$\mathbb{P}(A) = \frac{\text{Card}(A)}{\text{Card}(\Omega)}.$$

On a vu au chapitre précédent que  $\text{Card}(\Omega) = \binom{52}{2} = 1326$ . Quant à  $A$ , il correspond à choisir 2 éléments parmi les 4 as du jeu. Son cardinal est donc  $\text{Card}(A) = \binom{4}{2} = 6$ . On a donc

$$\mathbb{P}(A) = \frac{6}{1326} = \frac{1}{221} \approx 0.0045$$

**Problème concret** : Quelle est la probabilité pour que, dans une classe de 30 élèves, au moins deux élèves aient la même date de naissance ?

S'il y a  $n$  élèves dans la classe, notons  $\Omega = \{1, \dots, 365\}^n$ , c'est-à-dire l'ensemble des  $n$ -uplets possibles donnant la liste des dates de naissance des élèves. En première approximation, on munit cet ensemble de la mesure uniforme.<sup>1</sup> On a donc, pour tout événement  $A$

$$\mathbb{P}(A) = \frac{\text{Card}(A)}{\text{Card}(\Omega)} = \frac{\text{Card}(A)}{365^n}.$$

L'événement  $A$  correspond à ce que au moins deux coordonnées du  $n$ -uplet aient la même valeur. Son complémentaire est donc l'ensemble  $A^c$  où toutes les coordonnées du  $n$ -uplet sont différentes. Cela correspond à choisir choisir une suite de  $n$  éléments distincts parmi 365. Par le théorème 1.2, on a  $\text{Card}(A^c) = A_{365}^n = \frac{365!}{(365-n)!}$ . On a donc, pour  $n = 30$ ,

$$\mathbb{P}(A^c) = \frac{365!}{(335!) \times 365^{30}} = \frac{365}{365} \times \frac{364}{365} \times \dots \times \frac{336}{365} \approx 0,29.$$

On a donc

$$\mathbb{P}(A) = 1 - \mathbb{P}(A^c) \approx 0,71.$$

## Propriétés des mesures de probabilité

On peut déduire les propriétés suivantes de la définition d'une mesure de probabilité.

**Proposition 2.1** (Propriétés élémentaires des mesures de probabilité). *Soit  $(\Omega, \mathbb{P})$  un espace de probabilité fini.*

1. On a  $\mathbb{P}(\emptyset) = 0$ .
2. Pour tout  $A \subset \Omega$ , on a  $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$ .
3. Pour tout  $A \subset \Omega$  et tout  $B \subset A$ , on a  $\mathbb{P}(B) = \mathbb{P}(A) - \mathbb{P}(A \setminus B)$ .
4. Pour tout  $A, B \subset \Omega$ , on a  $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$ .

*Démonstration.* 1. En prenant  $A = B = \emptyset$ , on a  $\emptyset \cap \emptyset = \emptyset$ ,  $\emptyset \cup \emptyset = \emptyset$ , donc  $\mathbb{P}(\emptyset) = \mathbb{P}(\emptyset \cup \emptyset) = \mathbb{P}(\emptyset) + \mathbb{P}(\emptyset)$ . On en déduit que  $\mathbb{P}(\emptyset) = 0$ .

2. On a  $A \cap A^c = \emptyset$ , donc  $\mathbb{P}(A) + \mathbb{P}(A^c) = \mathbb{P}(A \cup A^c) = \mathbb{P}(\Omega) = 1$ . On a donc bien  $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$ .

3. Soit  $A \subset \Omega$  et  $B \subset A$ . On a  $B \cup (A \setminus B) = A$  et  $B \cap (A \setminus B) = \emptyset$ , donc  $\mathbb{P}(A) = \mathbb{P}(B) + \mathbb{P}(A \setminus B)$ , d'où le résultat découle.

4. Soient  $A, B \subset \Omega$ . On a  $A \cup B = B \cup (A \setminus (A \cap B))$  et  $B \cup (A \setminus (A \cap B)) = \emptyset$ , donc

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

par le point 3. Le résultat en découle. □

**Théorème 2.1.** *Soit  $(\Omega, \mathbb{P})$  un espace de probabilité fini, et soit  $A \subset \Omega$ . On a*

$$\mathbb{P}(A) = \sum_{\omega \in A} \mathbb{P}(\{\omega\}). \tag{2.1}$$

*Démonstration.* Montrons ce résultat par récurrence sur le cardinal de  $A$ . Si  $|A| = 0$ , alors  $A = \emptyset$ , et on a  $\mathbb{P}(A) = 0$ , donc le résultat est vrai. Si  $|A| = 1$ , alors  $A$  est un singleton : il existe  $\omega_0 \in \Omega$  tel que  $A = \{\omega_0\}$ . On a donc bien  $\mathbb{P}(A) = \mathbb{P}(\omega_0) = \sum_{\omega \in A} \mathbb{P}(\{\omega\})$ . (2.1) est donc bien vérifiée pour tous les ensembles  $A$  de cardinal 0 ou 1.

1. En fait, plus de gens naissent le printemps et l'été que l'hiver, mais cela reste du même ordre de grandeur.

Supposons que (2.1) est vérifiée pour tous les ensembles de cardinal  $n$ , et soit  $A$  un ensemble de cardinal  $n + 1$ . Choisissons un élément  $\omega_0 \in A$ , et notons  $B = A \setminus \{\omega_0\}$ . On a alors  $A = B \cup \{\omega_0\}$ , et  $B \cap \{\omega_0\} = \emptyset$ . Par la propriété d'additivité, on a donc  $\mathbb{P}(A) = \mathbb{P}(B) + \mathbb{P}(\{\omega_0\})$ . Mais, par hypothèse de récurrence, on a

$$\mathbb{P}(B) = \sum_{\omega \in B} \mathbb{P}(\{\omega\}) = \sum_{\omega \in A \setminus \{\omega_0\}} \mathbb{P}(\{\omega\}).$$

On a donc bien  $\mathbb{P}(A) = \sum_{\omega \in A \setminus \{\omega_0\}} \mathbb{P}(\{\omega\}) + \mathbb{P}(\{\omega_0\}) = \sum_{\omega \in A} \mathbb{P}(\{\omega\})$ , et on déduit le résultat par récurrence.  $\square$

## Lois de probabilité

Plutôt que de travailler avec des mesures de probabilité, c'est-à-dire d'associer une valeur à chaque évènement, il est souvent plus facile d'associer une valeur à chaque issue de l'expérience aléatoire, c'est-à-dire à chaque élément de  $\Omega$ . C'est l'objet de la définition suivante.

**Définition 2.5.** Soit  $\Omega$  un ensemble fini. Une loi de probabilité sur  $\Omega$  est une application  $q : \Omega \rightarrow [0, 1]$  telle que  $\sum_{\omega \in \Omega} q(\omega) = 1$ .

La proposition suivante nous explique comment passer d'une mesure de probabilité à une loi de probabilité, et réciproquement ; lorsque  $\Omega$  est un ensemble fini, ces deux notions sont tellement proches qu'on les identifie souvent.

**Proposition 2.2** (Lien entre mesure de probabilité et loi de probabilité). Soit  $\Omega$  un ensemble fini.

1. Si  $\mathbb{P}$  est une mesure de probabilité sur  $\Omega$ , alors l'application  $q : \Omega \rightarrow [0, 1]$  donnée par  $q(\omega) = \mathbb{P}(\{\omega\})$  est une loi de probabilité.
2. Réciproquement, si  $q$  est une loi de probabilité sur  $\Omega$ , alors l'application  $\mathbb{P} : \mathcal{P}(\Omega) \rightarrow [0, 1]$  donnée par  $\mathbb{P}(A) := \sum_{\omega \in A} q(\omega)$  est une mesure de probabilité sur  $\Omega$ .

*Démonstration.* 1. On applique (2.1) à  $A = \Omega$ , et on a bien

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

2. Avec cette définition de  $\mathbb{P}$ , on a  $\mathbb{P}(\Omega) = \sum_{\omega \in \Omega} q(\omega) = 1$ , car  $q$  est une loi de probabilité. Ensuite, si  $A, B \subset \Omega$  avec  $A \cap B = \emptyset$ , on a

$$\begin{aligned} \mathbb{P}(A \cup B) &= \sum_{\omega \in A \cup B} q(\omega) \\ &= \sum_{\omega \in A} q(\omega) + \sum_{\omega \in B} q(\omega) \quad \text{car } A \cap B = \emptyset \\ &= \mathbb{P}(A) + \mathbb{P}(B). \end{aligned}$$

$\mathbb{P}$  est donc bien une mesure de probabilité.  $\square$

**Exemple 2.3.** Soit  $\Omega$  un ensemble fini de cardinal  $N$ . La loi de probabilité associée à la mesure uniforme est donnée par

$$\forall \omega \in \Omega, \quad q(\omega) = \frac{1}{N}.$$

## La loi de Bernoulli et la loi binomiale

**Définition 2.6.** Soit  $p \in [0, 1]$ . La loi de Bernoulli de paramètre  $p$  est la loi sur  $\{0, 1\}$  donnée par

$$q(1) = p, \quad q(0) = 1 - p.$$

**Problème concret :** Je lance 10 fois une pièce. Quelle est la probabilité que j'obtienne 4 fois face ?

La réponse est fournie par la loi suivante :

**Définition 2.7.** Soit  $N \in \mathbb{N}$  et soit  $p \in [0, 1]$ . La loi binomiale de paramètre  $p$  sur  $\{0, \dots, N\}$  est donnée par

$$\forall k \in \{0, \dots, N\}, \quad q(k) = \binom{N}{k} p^k (1-p)^{N-k}.$$

Il s'agit bien d'une loi de probabilité : par la formule du binôme de Newton, on a

$$\sum_{k=0}^N \binom{N}{k} p^k (1-p)^{N-k} = (p + (1-p))^N = 1.$$

Vérifions que cette loi répond bien à la question posée précédemment. Supposons que je lance  $N$  fois la même pièce, et que la probabilité d'obtenir face sur cette pièce vaut  $p$ . Par exemple,  $p = 1/2$  si la pièce est équilibrée.

L'univers décrivant cette expérience est alors  $\{(P, F)\}^N$ . Soit  $\{x_1, \dots, x_N\} \in \{(P, F)\}^N$ , et soit  $k$  le nombre de  $F$  dans  $\{x_1, \dots, x_N\}$ . Le nombre de  $P$  dans  $\{x_1, \dots, x_N\}$  est donc  $N - k$ , et la probabilité d'obtenir  $\{x_1, \dots, x_N\}$  est  $p^k (1-p)^{N-k}$ .

L'espace  $\{(P, F)\}^N$  est donc muni de la mesure de probabilité  $\mathbb{P}$  telle que

$$\mathbb{P}(\{x_1, \dots, x_N\}) = p^k (1-p)^{N-k}, \quad \text{où } k \text{ est le nombre de fois où apparaît } F \text{ dans } \{x_1, \dots, x_N\}.$$

Revenons à notre problème initial, qui était de calculer la probabilité d'obtenir  $k$  fois face. Il existe  $\binom{N}{k}$   $N$ -uplets  $\{x_1, \dots, x_N\}$ , et chacun a une probabilité  $p^k (1-p)^{N-k}$  d'être obtenu. La probabilité d'obtenir  $k$  fois face est donc bien

$$\binom{N}{k} p^k (1-p)^{N-k},$$

comme annoncé.

## 2.3 Variables aléatoires finies

Dans tout ce chapitre,  $\Omega$  sera un ensemble fini, muni d'une mesure de probabilité  $\mathbb{P}$ .

**Définition 2.8.** Une variable aléatoire sur  $(\Omega, \mathbb{P})$  est une application  $X : \Omega \rightarrow \mathbb{R}$ . Le support  $X(\Omega)$  d'une variable aléatoire  $X$  est l'image de  $\Omega$  par  $X$ , c'est-à-dire l'ensemble des valeurs prises par  $X(\omega)$  lorsque  $\omega$  parcourt  $\Omega$ .

$X(\Omega)$  est donc un sous-ensemble fini de  $\mathbb{R}$ .

**Définition 2.9.** *On notera*

$$\{X = x\} := \{\omega \in \Omega; X(\omega) = x\}.$$

*On notera aussi*

$$\mathbb{P}(X = x) := \mathbb{P}\{\omega \in \Omega; X(\omega) = x\}.$$

*Plus généralement, si  $A \subset X(\Omega)$ , on notera*

$$\{X \in A\} := \{\omega \in \Omega; X(\omega) \in A\}, \quad \mathbb{P}(X \in A) := \mathbb{P}\{\omega \in \Omega; X(\omega) \in A\}.$$

*Enfin, si  $X_1, \dots, X_n$  sont des variables aléatoires, et si  $x_1 \in X_1(\Omega), \dots, x_n \in X_n(\Omega)$ , on notera*

$$\{X_1 = x_1, \dots, X_n = x_n\} = \{\omega \in \Omega; X_1(\omega) = x_1 \text{ et } X_2(\omega) = x_2 \text{ et } \dots \text{ et } X_n(\omega) = x_n\}.$$

**Proposition 2.3.** *L'application  $X(\Omega) \ni x \mapsto \mathbb{P}(X = x) \in [0, 1]$  est une loi de probabilité sur  $X(\Omega)$ .*

Cette loi est appelée la **loi de  $X$** .

*Démonstration.* On a

$$\begin{aligned} \sum_{x \in X(\Omega)} \mathbb{P}(X = x) &= \sum_{x \in X(\Omega)} \mathbb{P}(\{\omega \in \Omega; X(\omega) = x\}) \\ &= \mathbb{P}\left(\bigcup_{x \in X(\Omega)} \{\omega \in \Omega; X(\omega) = x\}\right) \quad \text{car les événements } \{X = x\} \text{ sont deux à deux disjoints} \\ &= \mathbb{P}(\Omega) = 1. \end{aligned}$$

□

**Exemple 2.4.** *Considérons le lancer de deux dés à 6 faces équilibrés et discernables. L'espace de probabilité associé est  $\Omega = \{1, \dots, 6\}^2$  muni de la mesure uniforme.*

- Pour  $i = 1, 2$ , la variable aléatoire  $X_i$  donnant le résultat du dé numéro  $i$  est une variable aléatoire de support  $\{1, \dots, 6\}$ , et la loi de cette variable est la loi uniforme.
- La variable aléatoire  $S = X_1 + X_2$  donnée par la somme des résultats des deux dés est une variable aléatoire de support  $\{2, \dots, 12\}$ . On a

$$\begin{aligned} \{S = 2\} &= \{(1, 1)\}, & \{S = 3\} &= \{(1, 2); (2, 1)\}, & \{S = 4\} &= \{(1, 3); (2, 2); (3, 1)\}, \\ \{S = 5\} &= \{(1, 4); (2, 3); (3, 2); (4, 1)\} & \{S = 6\} &= \{(1, 5); (2, 4); (3, 3); (4, 2); (5, 1)\}, \\ \{S = 7\} &= \{(1, 6); (2, 5); (3, 4); (4, 3); (5, 2); (6, 1)\}, \\ \{S = 8\} &= \{(2, 6); (3, 5); (4, 4); (5, 3); (6, 2)\}, & \{S = 9\} &= \{(3, 6); (4, 5); (5, 4); (6, 3)\} \\ \{S = 10\} &= \{(4, 6); (5, 5); (6, 4)\}, & \{S = 11\} &= \{(5, 6); (6, 5)\}, & \{S = 12\} &= \{(6, 6)\}. \end{aligned}$$

*On a donc*

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

**Définition 2.10.** *Soient  $(\Omega; \mathbb{P})$  et  $(\Omega', \mathbb{P}')$  deux univers probabilistes finis, et soient  $X : \Omega \rightarrow \mathbb{R}$ ,  $X' : \Omega' \rightarrow \mathbb{R}$  deux variables aléatoires. On dit que  $X$  et  $X'$  sont **égales en loi**, ce que l'on note  $X \stackrel{\mathcal{L}}{=} X'$  si  $X(\Omega) = X'(\Omega')$  et si, pour tout  $x \in X(\Omega)$ , on a  $\mathbb{P}(X = x) = \mathbb{P}'(X' = x)$ .*

- Soit  $0 < p < 1$ . On dit que  $X$  suit une loi de Bernoulli de paramètre  $p$ , ce que l'on note  $X \sim \mathcal{B}(p)$ , si  $X(\Omega) = \{0, 1\}$  et  $\mathbb{P}(X = 1) = p$ ,  $\mathbb{P}(X = 0) = 1 - p$ .
- Soit  $0 < p < 1$ ,  $n \in \mathbb{N}$ . On dit que  $X$  suit une loi binomiale de paramètre  $(n, p)$ , ce que l'on note  $X \sim \mathcal{B}(n, p)$ , si  $X(\Omega) = \{0, \dots, n\}$  et  $\mathbb{P}(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$ .

### 2.3.1 Espérance d'une variable aléatoire finie

**Définition 2.11.** Soit  $X$  une variable aléatoire sur  $\Omega$ . On définit son *espérance*  $\mathbb{E}(X)$  par

$$\mathbb{E}(X) = \sum_{x \in X(\Omega)} x \mathbb{P}(X = x).$$

**Remarque 2.1.** — Soit  $X$  la variable aléatoire donnant le résultat d'un dé équilibré à 6 faces. Son espérance vaut

$$\mathbb{E}(X) = \frac{1}{6} + \frac{2}{6} + \frac{3}{6} + \frac{4}{6} + \frac{5}{6} + \frac{6}{6} = 3,5.$$

- L'espérance d'une variable aléatoire ne dépend que de sa loi :  $X \stackrel{\mathcal{L}}{=} Y \implies \mathbb{E}(X) = \mathbb{E}(Y)$ .
- Soit  $A \subset \Omega$  un événement. Sa fonction indicatrice  $\mathbf{1}_A : \Omega \rightarrow \mathbb{R}$ , définie par  $\mathbf{1}_A(\omega) = 1$  si  $\omega \in A$ , et 0 si  $\omega \in \Omega \setminus A$ , est une variable aléatoire de support  $\{0, 1\}$ . Son espérance vaut

$$\mathbb{E}(\mathbf{1}_A) = 1 \times \mathbb{P}(\mathbf{1}_A = 1) + 0 \times \mathbb{P}(\mathbf{1}_A = 0) = \mathbb{P}(A).$$

**Proposition 2.4.** Soient  $X, Y : \Omega \rightarrow \mathbb{R}$  des variables aléatoires, et soit  $\lambda \in \mathbb{R}$ . Alors

$$\mathbb{E}(\lambda X + Y) = \lambda \mathbb{E}(X) + \mathbb{E}(Y). \quad (2.2)$$

**Lemme 2.1.** Soient  $X, Y$  des variables aléatoires. Pour tout  $x \in X(\Omega)$ , on a

$$\mathbb{P}(X = x) = \sum_{y \in Y(\Omega)} \mathbb{P}(X = x, Y = y).$$

*Démonstration du lemme.* On a  $\{X = x\} = \bigcup_{y \in Y(\Omega)} \{X = x, Y = y\}$ , et les événements de la réunion sont deux à deux distincts. On peut donc conclure en utilisant la deuxième propriété dans la définition d'une mesure de probabilité.  $\square$

*Démonstration de la proposition.* Notons  $Z = \lambda X + Y$ . On a  $Z(\Omega) = \{\lambda x + y \mid (x, y) \in X(\Omega) \times Y(\Omega)\}$ . Soit  $z \in Z(\Omega)$ . On a

$$\{Z = z\} = \bigcup_{\substack{(x,y) \in X(\Omega) \times Y(\Omega) \\ \lambda x + y = z}} \{X = x, Y = y\}.$$

Ces événements étant deux à deux disjoints, on a

$$\mathbb{P}(Z = z) = \sum_{\substack{(x,y) \in X(\Omega) \times Y(\Omega) \\ \lambda x + y = z}} \mathbb{P}(X = x, Y = y) = \sum_{(x,y) \in X(\Omega) \times Y(\Omega)} \mathbf{1}_{\lambda x + y = z} \mathbb{P}(X = x, Y = y).$$

On a

$$\begin{aligned}
\mathbb{E}(Z) &= \sum_{z \in Z(\Omega)} z \mathbb{P}(Z = z) = \sum_{z \in Z(\Omega)} \sum_{(x,y) \in X(\Omega) \times Y(\Omega)} \mathbf{1}_{\lambda x + y = z} (\lambda x + y) \mathbb{P}(X = x, Y = y) \\
&= \sum_{(x,y) \in X(\Omega) \times Y(\Omega)} \left( \sum_{z \in Z(\Omega)} \mathbf{1}_{\lambda x + y = z} \right) (\lambda x + y) \mathbb{P}(X = x, Y = y) \\
&= \sum_{(x,y) \in X(\Omega) \times Y(\Omega)} (\lambda x + y) \mathbb{P}(X = x, Y = y) \\
&= \lambda \sum_{x \in X(\Omega)} \sum_{y \in Y(\Omega)} x \mathbb{P}(X = x, Y = y) + \sum_{x \in X(\Omega)} \sum_{y \in Y(\Omega)} y \mathbb{P}(X = x, Y = y) \\
&= \lambda \sum_{x \in X(\Omega)} x \mathbb{P}(X = x) + \sum_{y \in Y(\Omega)} y \mathbb{P}(X = x, Y = y) \quad \text{par le lemme 2.1} \\
&= \lambda \mathbb{E}(X) + \mathbb{E}(Y).
\end{aligned}$$

□

**Remarque 2.2.** On a vu que, si  $X_1, \dots, X_n$  étaient des variables aléatoires indépendantes (voir la section ?? pour la définition précise de l'indépendance) suivant une loi de Bernoulli de paramètre  $p$ , alors  $X_1 + \dots + X_n$  suivait la loi  $\mathcal{B}(n, p)$ . On déduit de la proposition précédente que si  $X \sim \mathcal{B}(n, p)$ , alors

$$\mathbb{E}(X) = np.$$

**Proposition 2.5.** 1. Soient  $X \rightarrow [0, +\infty[$  une variable aléatoire positive. Alors  $\mathbb{E}(X) \geq 0$ . De plus, si  $\mathbb{E}(X) = 0$ , alors  $\mathbb{P}(X = 0) = 1$ . (On dit alors que  $X$  est presque sûrement nulle.)  
2. Soient  $X, Y$  deux variables aléatoires telles que  $X \leq Y$ . Alors  $\mathbb{E}(X) \leq \mathbb{E}(Y)$ .

*Démonstration.* 1. On a  $\mathbb{E}(X) = \sum_{x \in X(\Omega)} x \mathbb{P}(X = x)$ . Cette somme ne comporte que des termes positifs, donc elle est positive. Si  $\mathbb{E}(X) = 0$ , alors tous les termes de la somme doivent être nuls. Par conséquent, on doit avoir  $\mathbb{P}(X = x) = 0$  pour tous les  $x > 0$ , et donc  $\mathbb{P}(X = 0) = 1$ .

2. Si  $X \leq Y$ , alors  $Y - X \geq 0$ , donc  $\mathbb{E}(Y - X) \geq 0$ . Par la proposition précédente, on a  $\mathbb{E}(Y - X) = \mathbb{E}(Y) - \mathbb{E}(X)$ , donc on en déduit bien que  $\mathbb{E}(X) \leq \mathbb{E}(Y)$ . □

**Théorème 2.2** (Formule de transfert). Soit  $X$  une variable aléatoire, et soit  $f : \mathbb{R} \rightarrow \mathbb{R}$ .

$$\mathbb{E}(f(X)) = \sum_{x \in X(\Omega)} f(x) \mathbb{P}(X = x). \quad (2.3)$$

*Démonstration.* Notons  $Y = f(X)$ . On a  $Y(\Omega) = f(X(\Omega)) = \{f(x); x \in X(\Omega)\}$ . On a

$$\begin{aligned}
\mathbb{E}(Y) &= \sum_{y \in Y(\Omega)} y \mathbb{P}(Y = y) = \sum_{y \in Y(\Omega)} y \sum_{x \in X(\Omega); f(x)=y} \mathbb{P}(X = x) \\
&= \sum_{x \in X(\Omega)} f(x) \mathbb{P}(X = x).
\end{aligned}$$

□

### 2.3.2 Variance d'une variable aléatoire finie

**Définition 2.12.** Soit  $X : \Omega \rightarrow \mathbb{R}$  une variable aléatoire. On définit sa **variance**  $\text{Var}(X)$  par

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

L'**écart-type** de  $X$  est la racine carrée de la variance de  $X$ .



**Remarque 2.3.** — La variance (et l'écart-type) mesurent l'étalement, la dispersion de la variable aléatoire  $X$  autour de son espérance  $E(X)$ .

— On a

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

La démonstration de cette égalité sera faite en TD.

**Exemple 2.5.**

Si  $X \sim \mathcal{B}(p)$ , alors  $\mathbb{E}(X) = p$ , et donc  $\text{Var}(X) = (1-p)(0-p)^2 + p(1-p)^2 = p(1-p)(p+(1-p)) = p(1-p)$ .

**Un peu d'histoire (et de philosophie) :** La théorie des probabilités a été inventée par Fermat et Pascal, au XVIIème siècle. Mais, notre définition d'une mesure de probabilité est celle donnée par Kolmogorov, au début du XXème siècle. Que dit-elle ? Qu'à chaque événement, on associe un nombre compris entre zéro et un, en vérifiant la propriété d'additivité et en donnant la valeur un à l'univers tout entier. Et c'est tout !

Pourquoi avoir attendu tant de siècles pour donner une définition si simple ? Par ce que personne n'est d'accord sur ce que signifie la probabilité d'un événement ! Qu'est-ce que ça veut dire, qu'une pièce lancée en l'air a une probabilité  $\frac{1}{2}$  de tomber sur pile ? Pour certains, cela traduit la *symétrie* de la pièce. Pour d'autres, cela signifie que quand on lance un très grand nombre de fois cette pièce, on obtiendra environ autant de pile que de face : la *fréquence* des piles est de  $\frac{1}{2}$ . Certains considèrent qu'en l'absence d'information, il est également *plausible* que la pièce donne pile ou face au prochain lancer (mais, si elle a donné trois fois pile, il est peut-être plus plausible qu'elle soit truquée, et donc qu'elle donne pile au lancer suivant). Pour d'autres enfin, la probabilité désigne la *propension* d'un événement à advenir, une propriété a priori de tous les événements ; une jolie notion, mais difficile à définir proprement...

Toutes ces interprétations des probabilités ont un dénominateur commun : elles suivent toutes la définition de Kolmogorov ! Donc, quelles que soient vos convictions philosophiques, ce cours pourra vous être utile.

## Chapitre 3

# Introduction aux statistiques descriptives

### 3.1 Un peu de vocabulaire

**Définition 3.1.** — En statistiques, on considère une **population**, que l'on peut modéliser mathématiquement comme un ensemble fini  $\Omega$ . Un élément de cet ensemble est appelé un **individu** ou une **unité statistique**.

- Une **variable statistique** est une observation réalisée sur une population (on parle aussi d'un **caractère** d'un individu).
- On dit qu'une variable est **quantitative** si le résultat de l'observation est un nombre. Une variable quantitative est dite **discrète** si son résultat est un nombre entier (ou appartient à un petit ensemble de nombres réels), et elle est dite **continue** si son résultat peut a priori être n'importe quel nombre dans un intervalle.
- On dit qu'une variable est **qualitative** si le résultat n'est pas un nombre. On parle de variable qualitative **ordinaire** si ses résultats peuvent être classés, et **nominale** ou **catégorielle** sinon.
- Les valeurs pouvant être prises par une variable statistique sont appelées ses **modalités** (ou juste ses valeurs, quand il s'agit d'une variable quantitative).

Mathématiquement, on peut donc voir une variable statistique comme une application  $X$  à valeurs dans  $\mathbb{R}$ , dans  $\mathbb{N}$  ou dans un ensemble fini.

**Exemple 3.1.** — La population considérée peut être la population d'un pays, l'ensemble des étudiants d'une classe, l'ensemble des résultats d'une expérience de physique que l'on a répété plusieurs fois...

- Des exemples de variables **qualitatives nominales** sont : la couleur des yeux d'une personne, la ville d'origine d'un étudiant...
- Des exemples de variables **qualitatives ordinaires** sont : la mention au baccalauréat, le résultat de l'arrachage des pétales d'une marguerite (pas du tout, un peu, beaucoup, passionément, à la folie),...
- Des exemples de variables **quantitatives discrètes** sont : le nombre de frères et sœurs, la note obtenue à l'examen...
- Des exemples de variables **quantitatives continues** sont : le revenu annuel, la superficie du logement...

**Remarque 3.1.** En pratique, les variables quantitatives continues ne peuvent pas toujours prendre toutes les valeurs dans un intervalle, car on fait souvent un arrondi. Par exemple, une somme d'argent est toujours arrondie au centime près.

De même, une variable quantitative discrète ne prend pas toujours pour valeur un nombre entier : par exemple, une note à un examen peut contenir des demi-points.

En pratique, ce qui distingue une variable continue d'une variable discrète, c'est que la première peut prendre un très grand nombre de valeurs différentes, tandis que la seconde ne prend qu'un nombre restreint de valeurs.

**Définition 3.2.** L'effectif d'une modalité d'une variable statistique est le nombre d'individus possédant cette modalité

Sa fréquence est son effectif divisé par la population totale.

Autrement dit, si  $X : \Omega \rightarrow \{x_1, \dots, x_n\}$  est une variable statistique, pour chaque  $x \in \{x_1, \dots, x_n\}$  sa fréquence est définie par

$$f(x) = \frac{\text{Card}(\{\omega \in \Omega; X(\omega) = x\})}{\text{Card}(\Omega)}.$$

**Remarque 3.2.** Pour une variable quantitative continue, l'effectif de chaque modalité sera en général de 1. Les notions d'effectifs et de fréquences ne sont donc pas très intéressantes pour de telles variables aléatoires.

## Lien entre probabilités et statistiques

La statistique descriptive que nous voyons dans ce chapitre possède de nombreux liens avec les probabilités.

En effet, la population considérée peut être vue comme un univers probabiliste, que l'on équipe avec la mesure uniforme. Une variable statistique quantitative est alors la même chose qu'une variable aléatoire, et la fréquence, qui associe à chaque modalité un nombre entre 0 et 1, peut être vue comme la loi de cette variable aléatoire.

La moyenne d'une variable statistique, que nous définirons dans la section 3.3.1, est la même chose que l'espérance d'une variable aléatoire.

## 3.2 Représentation graphique des variables statistiques

### 3.2.1 Diagrammes en barre et circulaires

**Problème concret :** Je connais la note au deuxième contrôle continu, le sexe, et le revenu mensuel brut du père <sup>a</sup> de mes étudiants de l'an passé. Comment représenter ces données efficacement ?

<sup>a</sup>. Cette dernière donnée est complètement fantaisiste, je n'ai évidemment pas accès à cette information.

Remarquons que les trois variables considérées ici ne sont pas de même nature : la note au CC2 est une variable quantitative discrète, le sexe est une variable qualitative nominale, et le revenu du père est une variable quantitative continue (même si, ici, elle est arrondie à l'euro près).

Une première manière de représenter toutes ces données et de faire un **tableau** (voir ci-dessous). Le problème, c'est que si la population considérée est plus grande (la population d'un pays, par exemple), le tableau devient immense, et illisible !

Une manière plus efficace de représenter une variable qualitative est de faire un **diagramme en barres** (ou **diagramme en bâtons**), comme dans les figures 3.1 et 3.3, où on représente en abscisse les différentes modalités prises par la variable statistique, et, en ordonnées, les effectifs de chaque modalité. Plutôt que les effectifs, on peut aussi représenter les fréquences des différentes modalités en abscisse.

Pour représenter les fréquences d'une variable qualitative, on peut aussi faire un **diagramme circulaire** (ou **camembert**), comme dans la figure 3.2.

**Remarque 3.3.** Ce n'est pas une bonne idée de faire des diagrammes en bâtons ou en camembert pour des variables continues (car en général, toutes les modalités ont un effectif de 1, et il y a énormément de bâtons ou de parts de camembert). Voyez par exemple la figure 3.4, qui est assez illisible.

### 3.2.2 Classes et histogrammes

Il est difficile, à partir de la figure 3.3 de se rendre compte si les notes ont été bonnes ou non. Une solution commode est de regrouper la population en **classes**.

Ici, on considérera les classes « avoir une note comprise entre 0 et 4 », « entre 4 et 8 », « entre 8 et 12 », « entre 12 et 16 », et « entre 16 et 20 ». On peut alors réaliser un **histogramme**, comme dans

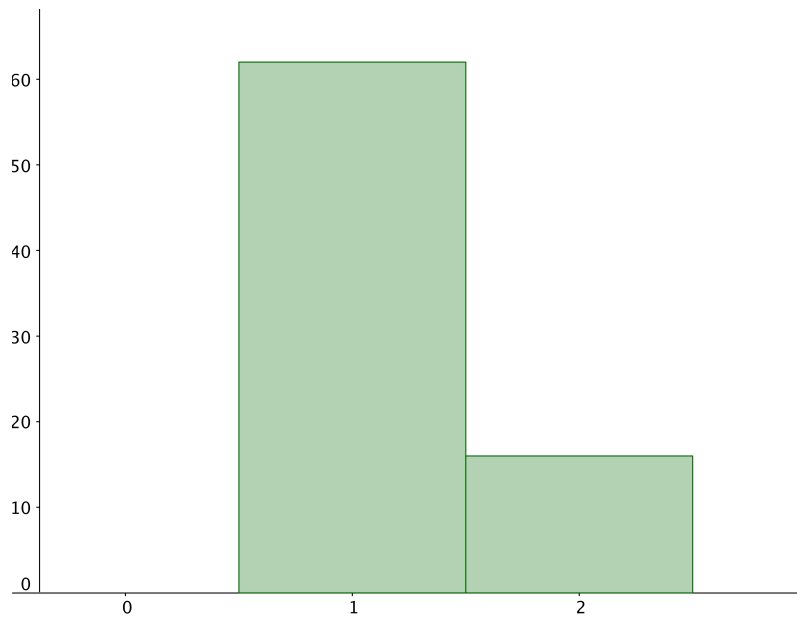


FIGURE 3.1 – Diagramme en barres des garçons et filles au cours de probas l’an dernier

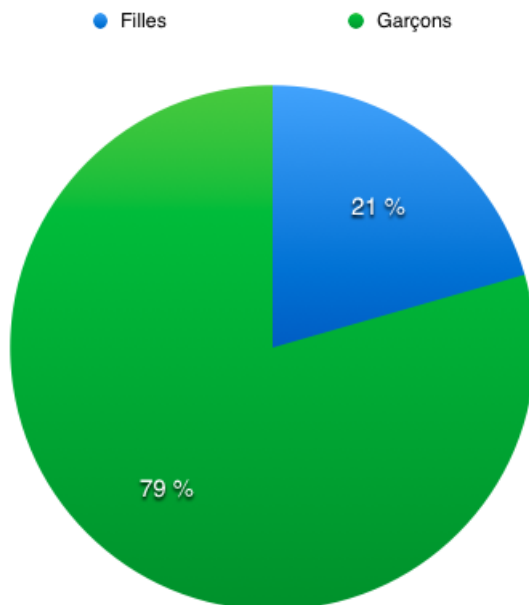


FIGURE 3.2 – Diagramme circulaire des garçons et filles au cours de probas l’an dernier

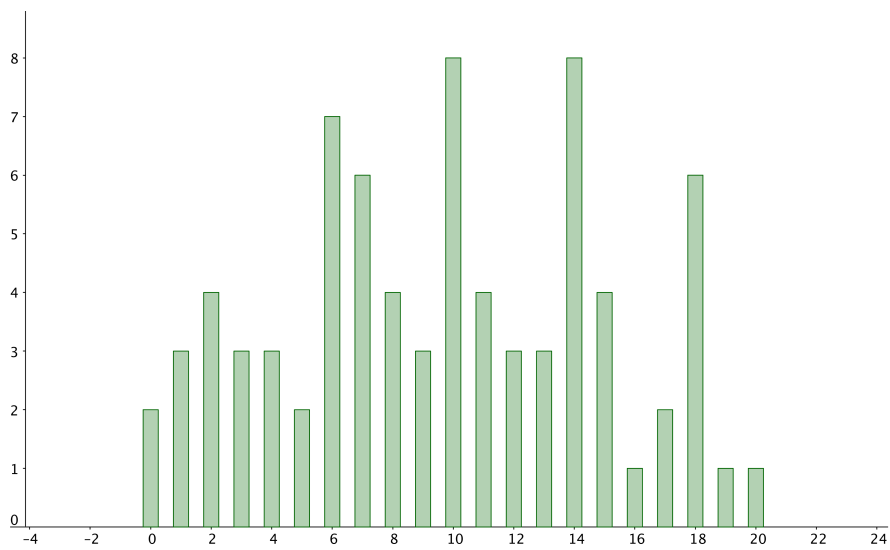


FIGURE 3.3 – Diagramme en barre des notes au 2ème contrôle continu de probas l’an dernier

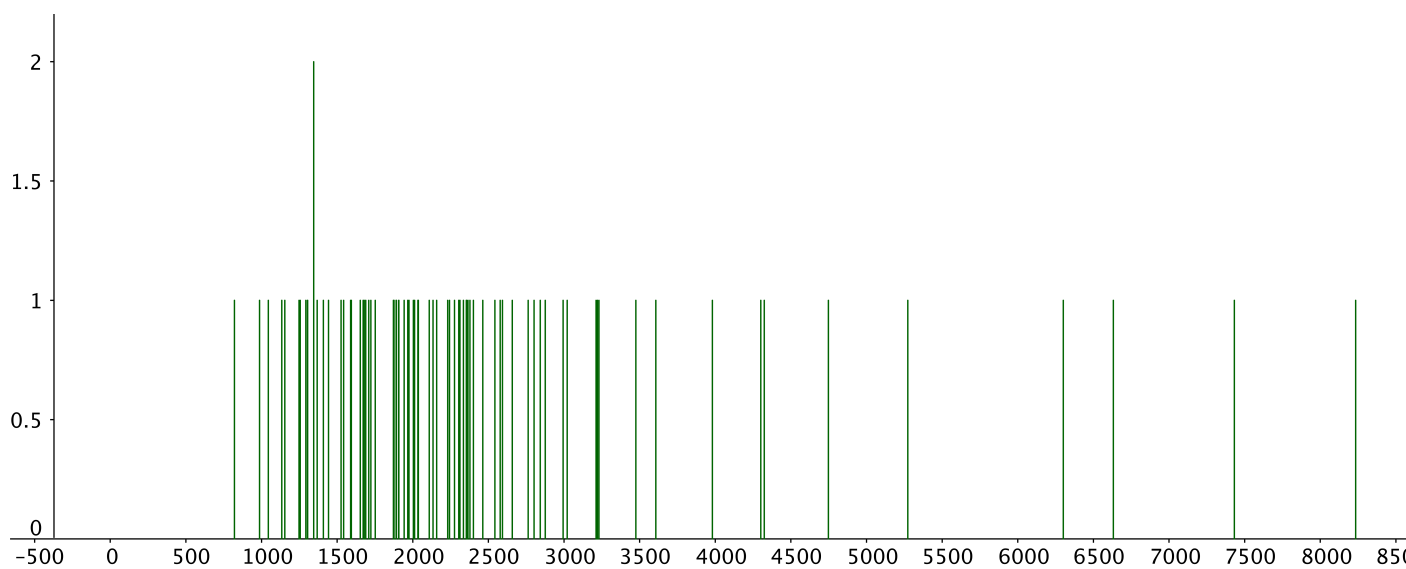


FIGURE 3.4 – Un diagramme en bâtons pour une variable continue : ça n’est pas très lisible!

Note au CC2	Sexe	Revenu mensuel brut du père
8	H	1526
7	H	2335
18	H	1154
6	F	4748
7	H	2994
0	H	2463
15	H	5273
15	H	1256
10	H	6632
4	H	8234
10	H	3221
1	F	1443
14	F	2543
18	H	7432
13	H	2354
12	F	4324
15	H	3231
15	H	1345
10	H	3212
5	H	2231
6	H	1134
9	F	2312
13	F	1892
6	H	1943
10	H	2004
18	F	2876
18	F	2134
11	F	1672
4	H	1689
2	H	1709
18	H	1543
8	H	821
6	H	2658
10	H	4301
6	H	1354
20	H	1872
12	H	2356
14	F	2578
11	H	1908

Note au CC2	Sexe	Revenu mensuel brut du père
14	H	1907
14	H	732
2	H	2109
8	F	2243
3	H	2843
9	H	9021
7	H	2378
10	H	1890
6	H	1975
12	H	3475
2	H	2401
8	H	2365
5	H	1679
14	H	1045
4	H	987
11	H	2158
11	H	2276
19	H	2593
9	H	2012
7	H	1876
3	H	1594
17	H	1409
14	H	1368
3	H	1723
14	F	981
1	H	2763
18	F	2034
14	H	6301
6	F	2802
2	H	1967
1	H	1653
10	H	1589
13	F	1752
7	F	2307
16	H	1304
0	H	1305
0	H	1293
7	H	1248
10	H	2037

la figure 3.5. La figure obtenue est alors beaucoup plus simple à lire : par exemple, elle est à peu près symétrique autour de la moyenne.

**Attention :** Quand on réalise un histogramme, on perd toujours de l'information. Le choix des classes en lesquels on regroupe les résultats est arbitraire. Il y a donc plusieurs histogrammes possibles pour une variable statistique !

Les histogrammes sont particulièrement adaptés pour représenter les variables quantitatives continues, comme dans la figure 3.6.

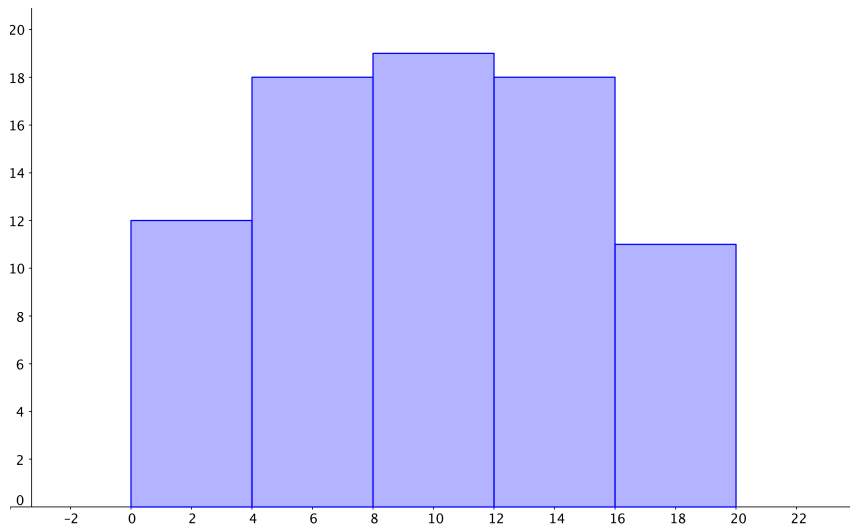


FIGURE 3.5 – Histogramme des notes au 2ème contrôle continu de probas l’an dernier

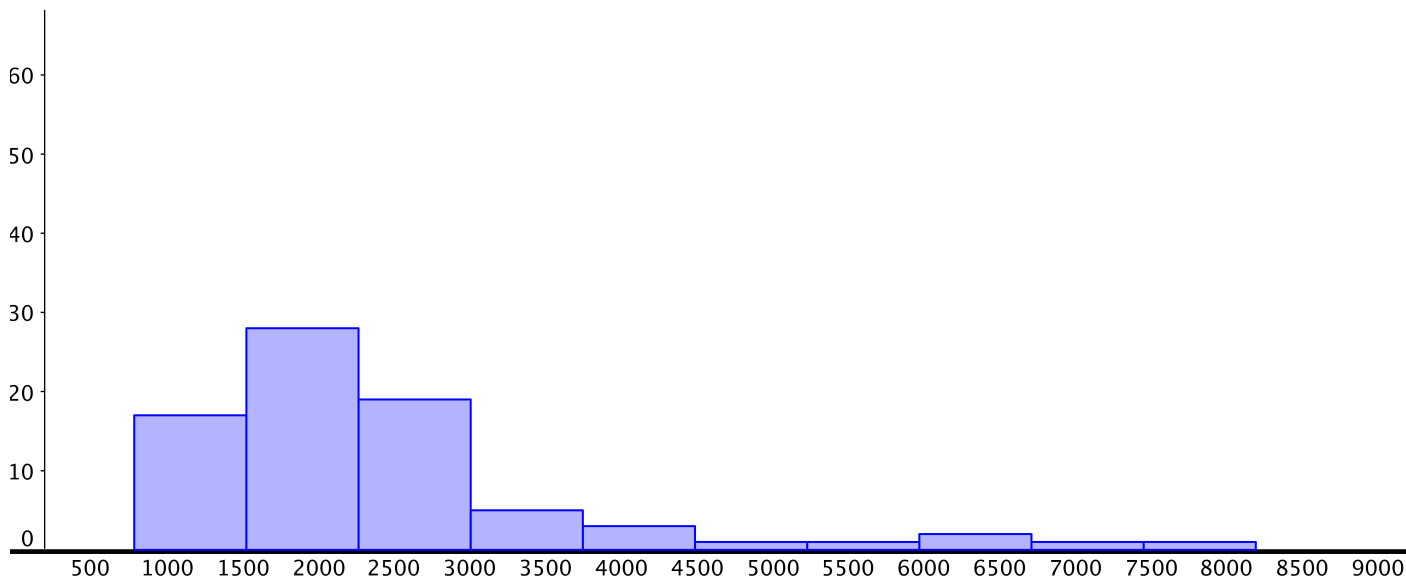


FIGURE 3.6 – Histogramme des revenus du père

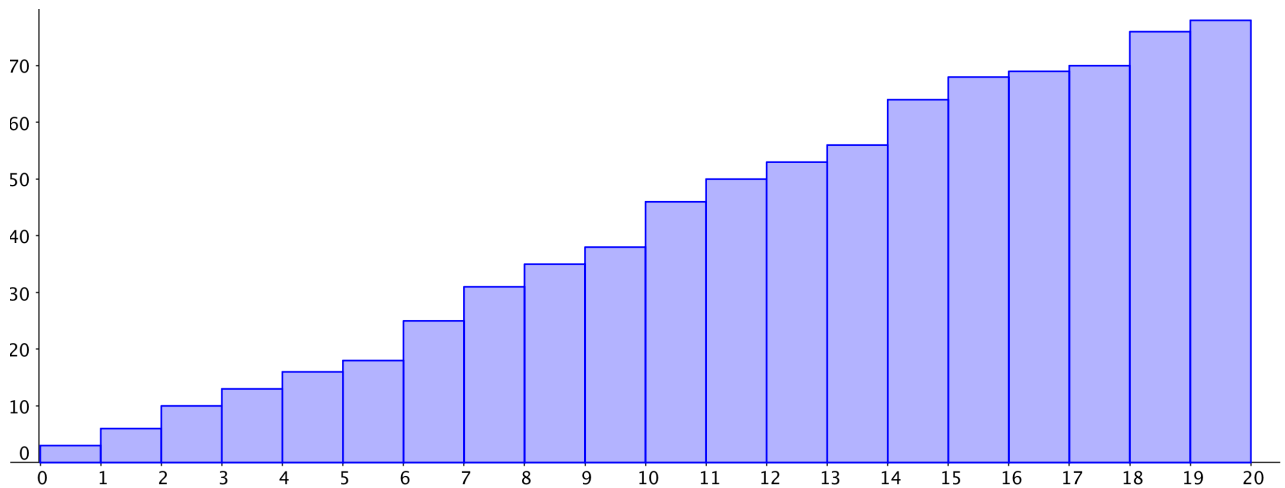


FIGURE 3.7 – Courbe des effectifs cumulés des notes

### 3.2.3 La courbe des effectifs cumulés

**Définition 3.3.** Si  $X$  est une variable quantitative (ou une variable qualitative ordinale), on définit sa **fonction des fréquences cumulées** comme

$$F(x) := \sum_{y \leq x} f(y).$$

On peut aussi définir la **fonction des effectifs cumulés** comme

$$F_{eff}(x) := \text{Card}(\Omega)F(x) = \sum_{y \leq x} \text{Card}(\{\omega \in \Omega; X(\omega) = y\}).$$

En représentant graphiquement cette fonction on obtient la **courbe des effectifs cumulés**, comme dans les figures 3.7 et 3.8.

L'avantage d'une courbe des effectifs cumulés est qu'elle peut-être utilisée pour des variables discrètes aussi bien que continues, et permet de calculer rapidement la médiane et autres quantiles (voir la section suivante).

**Remarque 3.4.** La fonction des fréquences cumulées est croissante, et prend ses valeurs dans  $[0, 1]$ . Dans le chapitre sur les variables aléatoires continues, on verra son analogue probabiliste, qui est appelé la **fonction de répartition**.

## 3.3 Indicateurs statistiques

Un indicateur statistique est un nombre réel donnant une information sur une variable statistique. On les range en général en trois catégories : les indicateurs de tendance central, de position, et de dispersion.

### 3.3.1 Les indicateurs de tendance centrale

Les indicateurs de tendance centrale sont des nombres qui décrivent le comportement d'un individu typique. Nous en verrons trois.

#### Le mode

**Définition 3.4.** Le **mode** d'une variable statistique est la modalité ayant l'effectif le plus grand.

Remarquons que le mode n'est pas nécessairement unique.



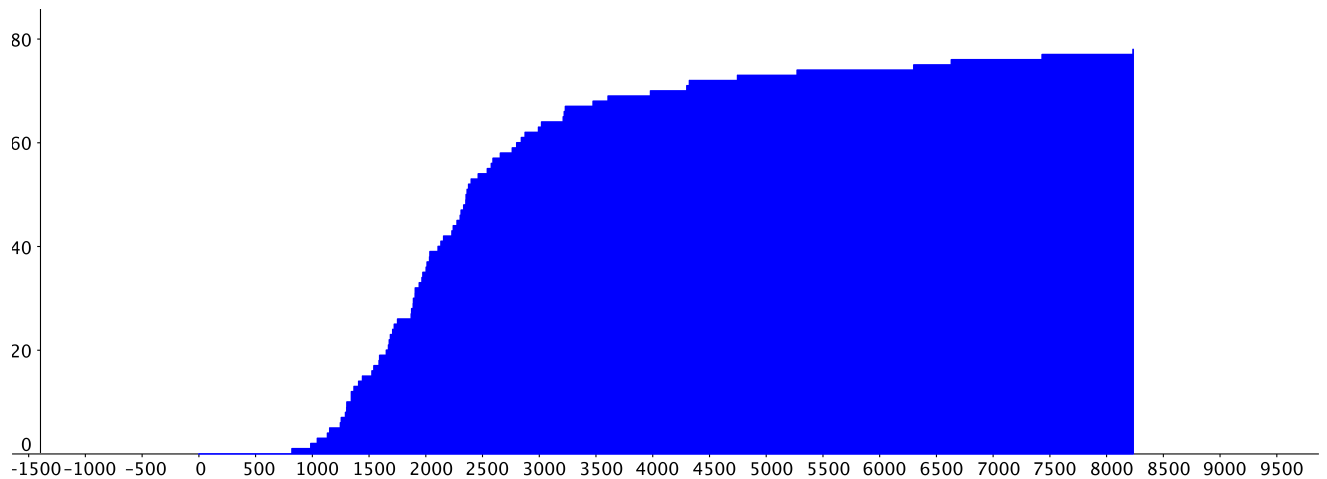


FIGURE 3.8 – Courbe des effectifs cumulés des revenus

**Exemple 3.2.** Si la population est la classe de l'an dernier, le mode de la variable « sexe » est homme, tandis que la variable « note au CC2 » possède deux modes, 10 et 14.

Pour une variable quantitative à densité, la notion de mode n'est pas du tout pertinente. Cependant, on peut regrouper les résultats par classes, appeler **classe modale** la classe qui possède le plus grand effectif.

### La moyenne

**Définition 3.5.** Si  $X$  est une variable quantitative, sa **moyenne**, notée  $\bar{X}$  est

$$\sum_{x \in X(\Omega)} x f(x).$$

Remarquons que cette définition est exactement la même que celle de l'espérance d'une variable aléatoire.

### La médiane

**Définition 3.6.** Si  $X$  est une variable quantitative, sa **médiane** est

$$\text{Med}(X) := \inf \left\{ x \in X; F(x) \geq \frac{1}{2} \right\},$$

où  $F(x)$  est la fonction des effectifs cumulés.

### Avantages et Inconvénients des différents paramètres de tendance centrale

- Le mode :
  - simple à calculer
  - ne dépend pas des valeurs extrêmes
  - non nécessairement unique
  - ne donne pas de renseignement sur l'ensemble des données
- La moyenne :
  - tient compte de l'ensemble des données
  - très dépendante des valeurs extrêmes
  - la plus populaire des trois même si parfois pas représentative
- La médiane :

- ne dépend pas des valeurs extrêmes
  - plus représentative que la moyenne si le jeu de données présente des valeurs extrêmes
  - un grand écart entre la médiane et la moyenne peut indiquer la présence de données extrêmes.
- Attention :** La réciproque est fausse.
- ne donne pas de renseignement sur l'ensemble des données

### 3.3.2 Paramètres de position

Les indicateurs (ou mesure, ou paramètres) de position donnent des informations sur la manière dont les données sont situées les unes par rapport aux autres.

#### Quantiles

**Définition 3.7.** Soit  $X$  est une variable statistique quantitative, et soit  $\alpha \in ]0, 1[$ . Le **quantile d'ordre  $\alpha$**  de  $X$ , noté  $q_\alpha(X)$ , est défini par

$$q_\alpha(X) := \inf \{x \in \mathbb{R}; F(x) \geq \alpha\},$$

où  $F$  est la fonction des fréquences cumulées.

**Remarque 3.5.** — La médiane est le quantile d'ordre  $\frac{1}{2}$ .

- On appelle **quartiles**  $Q_1$ ,  $Q_2$  et  $Q_3$  les quantiles d'ordre  $\alpha = \frac{1}{4}$ ,  $\alpha = \frac{1}{2}$  et  $\alpha = \frac{3}{4}$  respectivement.
- Les **déciles**  $D_1, \dots, D_9$  correspondent à  $\alpha = 0, 1, \dots, \alpha = 0, 9$ .
- Les **centiles**  $C_1, \dots, C_{99}$  correspondent à  $\alpha = 0, 01, \dots, \alpha = 0, 99$ .

#### Avantages et inconvénients des quantiles

- Valeurs non dépendantes des valeurs extrêmes.
- Les écarts entre les quartiles renseignent sur la concentration des données et sur leur symétrie.
- Il est difficile de détecter les quantiles intéressants.

**En économie :** Il existe de nombreux autres indicateurs statistiques, que nous ne présenterons pas dans ce cours, mais qui sont utiles dans d'autres sciences. Par exemple, en économie, on utilise souvent la **fonction de Lorenz**  $L$  pour représenter l'inégalité de répartition des richesses. Si  $x$  est un nombre entre 0 et 100,  $L(x)$  désigne le pourcentage de la richesse détenue par les  $x$ -pourcent les plus pauvres. Ainsi, on a toujours  $L(x) \leq x$ , et, si la richesse était répartie également entre tous, on aurait  $L(x) = L_{eg}(x) := x$  pour tout  $x$ . L'aire entre les courbes  $L(x)$  et  $L_{eg}(x)$ , multiplié par deux, est appelé le **coefficient de Gini**. Il vaut donc 1 quand la répartition des richesses dans la société est très inégale, et 0 quand elle est parfaitement égalitaire.

### 3.3.3 Paramètres de dispersion

Dans toute cette section,  $X$  sera une variable statistique quantitative.

Les indicateurs de dispersion permettent de mesurer l'étalement d'une variable statistique.

#### L'étendue et l'écart inter-quartile

**Définition 3.8.** L'**étendue** d'une variable est la différence entre sa plus grande modalité et sa plus petite modalité.

**Définition 3.9.** L'**écart inter-quartile** d'une variable est la différence entre son troisième et son premier quartile :

$$IQ(X) := Q_3(X) - Q_1(X).$$

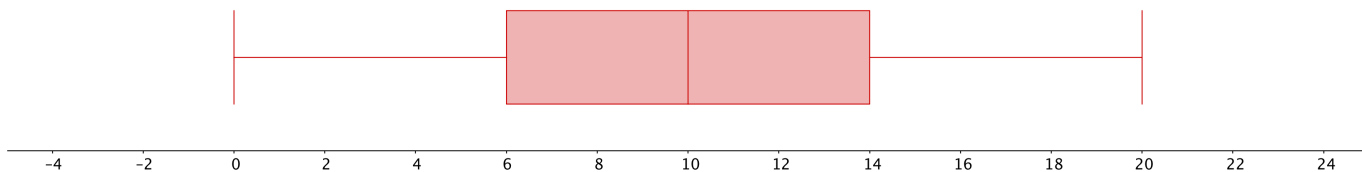


FIGURE 3.9 – Diagramme en boîte des notes

## La variance et l'écart-type

**Définition 3.10.** La variance d'une variable statistique est

$$s_X^2 := \sum_{x \in X(\Omega)} f(x) (x - \bar{X})^2$$

et son écart-type est

$$s_X := \sqrt{s_X^2}$$

Remarquons que ces définitions sont exactement les mêmes que pour les variables aléatoires.

## Avantages et Inconvénients

- L'étendue :
  - simple à calculer
  - ne donne aucune information sur la dispersion des valeurs intermédiaires
- La variance (ou l'écart-type) :
  - prend en compte toutes les données
  - mesure de dispersion vis à vis de la moyenne
  - difficilement interprétable si elle est prise seule
- L'écart inter-quartile :
  - joue le rôle de l'écart-type lorsque la médiane est plus représentative que la moyenne
  - ne se calcule que dans certains cas (variable avec une échelle de rapport et que la valeur 0 signifie absence de)
  - qualifie l'homogénéité d'une distribution
  - quantifie la fiabilité de la médiane

## Représentation graphique de certains paramètres : les « boîtes à moustaches »

Il est possible de synthétiser l'information de l'étendue et des trois quartiles sous la forme d'une **boîte à moustaches**, aussi appelée **diagramme en boîte** (voir la figure 3.9). On représente une boîte centrale, ayant pour extrémités le premier et le troisième quartile. Au milieu, on trace un trait pour la médiane (le deuxième quartile). Enfin, on trace deux traits horizontaux (les moustaches), qui s'arrêtent aux valeurs extrêmes prises par la variable statistique.

Sur la figure 3.9, on voit que la médiane des notes est de 10, que les notes vont de 0 à 20, et que les trois quarts des étudiants ont eu entre 6 et 14. L'écart interquartile est donc de 8.

# Chapitre 4

## Indépendance et conditionnement

Dans tout ce chapitre,  $(\Omega, \mathbb{P})$  sera un espace de probabilité. Le fait qu'il soit fini ne joue aucun rôle ici, et toutes les définitions et les résultats présentés ici s'adaptent sans problème au cas des espaces de probabilités infinis que nous considérerons dans les chapitres 5 et 6.

### 4.1 Probabilités conditionnelles et indépendance d'événements

#### Probabilités conditionnelles

**Définition 4.1.** Soient  $A, B \subset \Omega$  des événements. La **probabilité conditionnelle de  $A$  sachant  $B$**  est notée  $\mathbb{P}(A|B)$ , et est définie par

$$\mathbb{P}(A|B) = \begin{cases} \mathbb{P}(A \cap B) / \mathbb{P}(B) & \text{si } \mathbb{P}(B) > 0 \\ \mathbb{P}(A) & \text{sinon.} \end{cases}$$

La quantité  $\mathbb{P}(A|B)$  est parfois notée  $\mathbb{P}_B(A)$ .

En général, on n'utilisera cette définition que dans le cas où  $\mathbb{P}(B) > 0$ . Remarquons qu'elle alors peut se réécrire comme

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

**Exemple 4.1.** On lance 2 dés équilibrés à 6 faces. Sachant que la somme des résultats des deux dés vaut 4, quelle est la probabilité pour que le résultat du premier dé soit un 3 ?

Cette situation est modélisée par  $\Omega = \{1, \dots, 6\}^2$ , muni de la probabilité uniforme.

Notons  $A$  l'événement « le résultat du premier dé est un 3 », et  $B$  l'événement « la somme des résultats des dés vaut 4 ». On a  $A = \{3\} \times \{1; \dots; 6\}$ , de sorte que  $\text{Card}(A) = 6$ . On a donc  $\mathbb{P}(A) = \frac{6}{36} = \frac{1}{6}$ . On a  $B = \{(1, 3); (2, 2); (3, 1)\}$ , de sorte que  $\mathbb{P}(B) = \frac{3}{36} = \frac{1}{12}$ . Enfin,  $A \cap B = \{(3, 1)\}$ , donc  $\mathbb{P}(A \cap B) = \frac{1}{36}$ . On a donc

$$\mathbb{P}(A|B) = \frac{1/36}{1/12} = \frac{1}{3}.$$

Connaissant  $\mathbb{P}(A|B)$ ,  $\mathbb{P}(A|\bar{B})$  et  $\mathbb{P}(B)$ , on peut retrouver  $\mathbb{P}(A)$  : c'est la formule dite des *probabilités totales*.

**Lemme 4.1.** Soit  $A$  un événement.

1. Soit  $B$  un événement. Alors on a

$$\mathbb{P}(A) = \mathbb{P}(A|B) \times \mathbb{P}(B) + \mathbb{P}(A|\bar{B}) \times \mathbb{P}(\bar{B}).$$

2. Plus généralement, si  $B_1, \dots, B_n$  sont des événements tels que  $\bigcup_{i=1}^n B_i = \Omega$ , et  $B_i \cap B_j = \emptyset$  pour  $i \neq j$ , alors

$$\mathbb{P}(A) = \sum_{i=1}^n \mathbb{P}(A|B_i)\mathbb{P}(B_i).$$

*Démonstration.* Le premier point est bien une conséquence du second, car on a  $B \cup B^c = \Omega$ , et  $B \cap B^c = \emptyset$ .  
 Montrons maintenant le second point :

$$\begin{aligned} \sum_{j=1}^m \mathbb{P}(A|B_j)\mathbb{P}(B_j) &= \sum_{j=1}^m \mathbb{P}(A \cap B_j) \\ &= \mathbb{P}\left(\bigcup_{j=1}^m A \cap B_j\right) \quad \text{car les } A \cap B_j \text{ sont deux à deux disjoints} \\ &= \mathbb{P}\left(A \cap \left(\bigcup_{j=1}^m B_j\right)\right) \\ &= \mathbb{P}(A). \end{aligned}$$

□

Souvent, on connaît  $\mathbb{P}(A|B)$ , et on veut en déduire  $\mathbb{P}(B|A)$ . Ceci peut être fait grâce à la formule suivante, appelée *formule de Bayes*

**Proposition 4.1.** Soient  $A, B$  des événements tels que  $\mathbb{P}(A) > 0$ ,  $\mathbb{P}(B) > 0$ . On a

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

*Démonstration.* On a

$$\begin{aligned} \mathbb{P}(A|B) \times \frac{\mathbb{P}(B)}{\mathbb{P}(A)} &= \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \times \frac{\mathbb{P}(B)}{\mathbb{P}(A)} \\ &= \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)} \\ &= \mathbb{P}(B|A), \end{aligned}$$

ce qui nous donne bien le résultat annoncé. □

**Exemple 4.2.** On dispose de deux urnes contenant 100 boules chacune. Dans la première urne, il y a 50 boules blanches et 50 boules noires. Dans la seconde urne, il y a 20 boules blanches et 80 boules noires.

On choisit une urne au hasard, et on tire une boule au hasard dans cette urne. La boule tirée est noire. Quelle est la probabilité pour que l'urne qui a été choisie soit la seconde ?

Notons  $A$  l'événement « l'urne choisie est la seconde », et  $B$  l'événement « une boule noire est tirée. » On a  $\mathbb{P}(A) = \frac{1}{2}$ , et, par hypothèse,  $\mathbb{P}(B|A) = \frac{80}{100}$ ,  $\mathbb{P}(B|\bar{A}) = \frac{50}{100}$ . Par la formule des probabilités totales, on a donc

$$\mathbb{P}(B) = \mathbb{P}(B|A) \times \mathbb{P}(A) + \mathbb{P}(B|\bar{A}) \times \mathbb{P}(\bar{A}) = \frac{80}{100} \times \frac{1}{2} + \frac{50}{100} \times \frac{1}{2} = \frac{65}{100}.$$

On peut alors appliquer la formule de Bayes pour obtenir

$$\mathbb{P}(A|B) = \mathbb{P}(B|A) \times \frac{\mathbb{P}(A)}{\mathbb{P}(B)} = \frac{80}{100} \times \frac{1/2}{65/100} = \frac{80}{130} = \frac{8}{13} \approx 0,62.$$

Il y a donc une probabilité d'environ 0,62 pour que la boule provienne de la seconde urne.

Lorsqu'il y a plus de deux urnes, on applique plutôt la formule suivante, dite *formule de Bayes généralisée*.

**Proposition 4.2.** Soit  $B_1, \dots, B_m$  une partition de  $\Omega$  (c'est à dire que  $\bigcup_{i=1}^m B_i = \Omega$  et  $\forall 1 \leq i, j \leq m$ ,  $i \neq j \implies B_i \cap B_j = \emptyset$ ), et soit  $A \subset \Omega$  un événement tel que  $\mathbb{P}(A) > 0$ . Alors, pour tout  $1 \leq i \leq m$ , on a

$$\mathbb{P}(B_i|A) = \frac{\mathbb{P}(A|B_i)\mathbb{P}(B_i)}{\sum_{j=1}^m \mathbb{P}(A|B_j)\mathbb{P}(B_j)}.$$

*Démonstration.* Le numérateur est égal à  $\mathbb{P}(A \cap B_i)$ . Quant au dénominateur, il est égal à  $\mathbb{P}(A)$ , par la formule des probabilités totales. On a donc bien

$$\frac{\mathbb{P}(A|B_i)\mathbb{P}(B_i)}{\sum_{j=1}^m \mathbb{P}(A|B_j)\mathbb{P}(B_j)} = \frac{\mathbb{P}(A \cap B_j)}{\mathbb{P}(A)} = \mathbb{P}(B_i|A).$$

□

## Indépendance d'événements

**Définition 4.2.** Soit  $(\Omega, \mathbb{P})$  un espace de probabilité (fini).

1. Soient  $A, B \subset \Omega$  des événements. On dit que  $A$  et  $B$  sont **indépendants** si

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

2. Plus généralement, si  $A_1, \dots, A_n$  sont des événements, on dit qu'ils sont **indépendants** (ou que **la famille  $(A_i)_{i \in \{1, \dots, n\}}$  est indépendante**) si, pour toute sous-famille  $A_{i_1}, \dots, A_{i_k}$ , on a

$$\mathbb{P}(A_{i_1} \cap \dots \cap A_{i_k}) = \mathbb{P}(A_{i_1}) \times \dots \times \mathbb{P}(A_{i_k}).$$

**Attention!** Pour qu'une famille d'événements soit indépendante, il ne suffit pas que les événements soient indépendants deux à deux.

Par exemple, on lance deux pièces équilibrées. Cette expérience est modélisée par la probabilité uniforme sur l'univers  $\{PP, PF, FP, FF\}$ . On note  $A$  l'événement « la première pièce est pile »,  $B$  l'événement « la seconde pièce est pile » et  $C$  l'événement « les deux pièces donnent le même résultat ».

On a  $\mathbb{P}(A) = \mathbb{P}(B) = \mathbb{P}(C) = \frac{1}{2}$ , et  $A \cap B = B \cap C = A \cap C = \{PP\}$ , de sorte que  $\mathbb{P}(A \cap B) = \mathbb{P}(A \cap C) = \mathbb{P}(B \cap C) = \frac{1}{4}$ , mais  $\mathbb{P}(A \cap B \cap C) = \mathbb{P}(\{PP\}) = \frac{1}{4} \neq \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C)$ .

## 4.2 Indépendance de variables aléatoires finies

**Définition 4.3.** 1. Deux variables aléatoires  $X$  et  $Y$  de supports respectifs  $X(\Omega)$  et  $Y(\Omega)$  sont dites **indépendantes** si

$$\forall x \in X(\Omega), \forall y \in Y(\Omega), \quad \mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x) \times \mathbb{P}(Y = y).$$

2. Plus généralement, une famille de  $n$  variables indépendantes  $X_1, \dots, X_n$  de supports respectifs  $X_1(\Omega), \dots, X_n(\Omega)$  est dite **indépendante** si

$$\forall x_1 \in X_1(\Omega), \dots, \forall x_n \in X_n(\Omega), \quad \mathbb{P}(X_1 = x_1, \dots, X_n = x_n) = \prod_{i=1}^n \mathbb{P}(X_i = x_i).$$

3. Enfin, une famille quelconque de variables aléatoires est dite **indépendante** si toute-sous famille finie est indépendante.

**Exemple 4.3.** — Reprenons l'exemple du jet de deux dés, où  $X_i$  est la variable aléatoire donnant le résultat du  $i$ -ème dé. Les variables  $X_1$  et  $X_2$  sont indépendantes. En effet, si  $j_1, j_2 \in \{1, \dots, 6\}$ , on a

$$\mathbb{P}(X_1 = j_1, X_2 = j_2) = \frac{1}{36},$$

et

$$\mathbb{P}(X_1 = j_1)\mathbb{P}(X_2 = j_2) = \frac{1}{6} \times \frac{1}{6} = \frac{1}{36}.$$

— En revanche, les variables  $S$  et  $X_1$  ne sont pas indépendantes. En effet,  $\{S = 2, X_1 = 3\} = \emptyset$ , donc  $\mathbb{P}(S = 2, X_1 = 3) = 0$ , tandis que  $\mathbb{P}(S = 2)\mathbb{P}(X_1 = 3) = \frac{1}{36} \times \frac{1}{6}$ .

**Remarque 4.1.** Si  $X_1, \dots, X_n$  est une famille indépendante de  $n$  variables aléatoires, alors  $X_1, \dots, X_{n-1}$  est une famille indépendante de  $n - 1$  variables aléatoires.

En effet, si  $x_1 \in X_1(\Omega), \dots, x_{n-1} \in X_{n-1}(\Omega)$ , on a

$$\{X_1 = x_1, \dots, X_{n-1} = x_{n-1}\} = \bigcup_{x_n \in X_n(\Omega)} \{X_1 = x_1, \dots, X_{n-1} = x_{n-1}, X_n = x_n\},$$

et le membre de droite est formé d'ensembles deux à deux disjoints. Par additivité, on en déduit que

$$\begin{aligned}
\mathbb{P}(X_1 = x_1, \dots, X_{n-1} = x_{n-1}) &= \sum_{x_n \in X_n(\Omega)} \mathbb{P}(X_1 = x_1, \dots, X_{n-1} = x_{n-1}, X_n = x_n) \\
&= \sum_{x_n \in X_n(\Omega)} \mathbb{P}(X_1 = x_1) \times \dots \times \mathbb{P}(X_{n-1} = x_{n-1}) \times \mathbb{P}(X_n = x_n) \\
&= \mathbb{P}(X_1 = x_1) \times \dots \times \mathbb{P}(X_{n-1} = x_{n-1}) \times \sum_{x_n \in X_n(\Omega)} \mathbb{P}(X_n = x_n) \\
&= \mathbb{P}(X_1 = x_1) \times \dots \times \mathbb{P}(X_{n-1} = x_{n-1}).
\end{aligned}$$

Par une récurrence immédiate, on en déduit que si  $X_1, \dots, X_n$  est une famille indépendante de  $n$  variables aléatoires et si  $m \leq n$ , alors  $X_1, \dots, X_m$  est une famille indépendante de  $m$  variables aléatoires.

## Indépendance et espérance

Le théorème suivant fait le lien entre indépendance et calcul d'espérances.

**Théorème 4.1.** 1. Soient  $X$  et  $Y$  des variables aléatoires indépendantes, et soient  $f, g : \mathbb{R} \rightarrow \mathbb{R}$ . Alors on a

$$\mathbb{E}(f(X)g(Y)) = \mathbb{E}(f(X))\mathbb{E}(g(Y)). \quad (4.1)$$

2. Réciproquement, si pour tout  $f, g : \mathbb{R} \rightarrow \mathbb{R}$ , on a  $\mathbb{E}(f(X)g(Y)) = \mathbb{E}(f(X))\mathbb{E}(g(Y))$ , alors  $X$  et  $Y$  sont indépendantes.

*Démonstration.* 1. Les variables  $X$  et  $Y$  étant supposées indépendantes, on a

$$\begin{aligned}
\mathbb{E}(f(X)g(Y)) &= \sum_{(x,y) \in X(\Omega) \times Y(\Omega)} f(x)g(y)\mathbb{P}(X = x, Y = y) \\
&= \sum_{(x,y) \in X(\Omega) \times Y(\Omega)} f(x)g(y)\mathbb{P}(X = x)\mathbb{P}(Y = y) \\
&= \left( \sum_{x \in X(\Omega)} f(x)\mathbb{P}(X = x) \right) \times \left( \sum_{y \in Y(\Omega)} g(y)\mathbb{P}(Y = y) \right).
\end{aligned}$$

2. Soient  $x \in X(\Omega), y \in Y(\Omega)$ . Posons  $f(z) = \mathbf{1}_{z=x}, g(t) = \mathbf{1}_{t=y}$ . On a

$$\mathbb{E}(f(X)g(Y)) = \sum_{(z,t) \in X(\Omega) \times Y(\Omega)} \mathbf{1}_{z=x} \mathbf{1}_{t=y} \mathbb{P}(X = z, Y = t) = \mathbb{P}(X = x, Y = y),$$

tandis que

$$\mathbb{E}(f(X))\mathbb{E}(g(Y)) = \mathbb{P}(X = x)\mathbb{P}(Y = y).$$

On déduit donc de notre hypothèse que  $\mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x) \times \mathbb{P}(Y = y)$ , et donc que  $X$  et  $Y$  sont indépendantes.  $\square$

## Indépendance et variance

**Proposition 4.3.** Soient  $X_1, \dots, X_n$  des variables aléatoires indépendantes. Alors

$$\text{Var}(X_1 + \dots + X_n) = \text{Var}(X_1) + \dots + \text{Var}(X_n).$$



Démonstration de la proposition. On a

$$\begin{aligned}
 \text{Var}(X_1 + \dots + X_n) &= \mathbb{E}\left(\left(X_1 + \dots + X_n - \mathbb{E}(X_1 + \dots + X_n)\right)^2\right) \\
 &= \mathbb{E}\left(\left(\sum_{i=1}^n X_i - \mathbb{E}(X_i)\right)^2\right) \quad \text{par linéarité de l'espérance} \\
 &= \mathbb{E}\left(\left(\sum_{i=1}^n X_i - \mathbb{E}(X_i)\right)\left(\sum_{j=1}^n X_j - \mathbb{E}(X_j)\right)\right) \\
 &= \mathbb{E}\left(\sum_{i=1}^n (X_i - \mathbb{E}(X_i))^2 + \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n (X_i - \mathbb{E}(X_i))(X_j - \mathbb{E}(X_j))\right) \\
 &= \sum_{i=1}^n \text{Var}(X_i) + \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n \mathbb{E}\left((X_i - \mathbb{E}(X_i))(X_j - \mathbb{E}(X_j))\right) \\
 &= \sum_{i=1}^n \text{Var}(X_i) + \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n \mathbb{E}(X_i - \mathbb{E}(X_i))\mathbb{E}(X_j - \mathbb{E}(X_j)) \quad \text{car } X_i \text{ et } X_j \text{ sont indépendantes} \\
 &= \sum_{i=1}^n \text{Var}(X_i),
 \end{aligned}$$

car  $\mathbb{E}(X_i - \mathbb{E}(X_i)) = 0$ . □

### 4.3 La loi faible des grands nombres

**Théorème 4.2.** Soient  $X_1, \dots, X_n$  des variables aléatoires indépendantes, de même loi. Alors pour tout  $\varepsilon > 0$ , on a

$$\mathbb{P}\left(\left|\frac{1}{n} \sum_{i=1}^n X_i - \mathbb{E}(X_1)\right| > \varepsilon\right) \leq \frac{\text{Var}(X_1)}{\varepsilon^2 n}.$$

Intuitivement, le théorème nous dit que, quand on fait la moyenne d'un grand nombre de variables aléatoires indépendantes et de même loi, on obtient avec très grande probabilité un résultat proche de l'espérance de cette loi. Cela correspond à la définition intuitive de l'espérance : c'est le résultat moyen qu'on obtient, en répétant de nombreuses fois la variable aléatoire.

Il est parfois utile de réécrire le résultat comme

$$\mathbb{P}\left(\left|\sum_{i=1}^n X_i - n\mathbb{E}(X_1)\right| > n\varepsilon\right) \leq \frac{\text{Var}(X_1)}{\varepsilon^2 n}.$$

**Exemple 4.4.** Je jette 1000 fois une pièce équilibrée, et je note  $X$  le nombre de face. On a alors  $X = \sum_{i=1}^n X_i$ , où les  $X_i$  sont des variables de Bernoulli de paramètre  $1/2$ , indépendantes. On a vu que l'espérance d'une telle variable était de  $\frac{1}{2}$ , et son espérance de  $\frac{1}{4}$ , donc le théorème nous dit que

$$\mathbb{P}\left(\left|X - 500\right| > 1000\varepsilon\right) \leq \frac{1}{4000\varepsilon^2}.$$

Par exemple, pour  $\varepsilon = \frac{1}{10}$ , on obtient  $\mathbb{P}\left(\left|X - 500\right| > 100\right) \leq \frac{1}{40}$  : la probabilité que j'obtienne un nombre de piles qui n'est pas entre 400 et 600 est inférieure à 2,5%.

Pour prouver ce théorème, nous aurons besoin du lemme suivant appelé inégalité de Bienaymé-Chebychev (ou Tchebychev).

**Lemme 4.2.** Soit  $X$  une variable aléatoire. Pour tout  $a > 0$ , on a

$$\mathbb{P}(|X| > a) \leq \frac{\mathbb{E}(X^2)}{a^2}.$$

*Démonstration du lemme.* On a  $X^2 \geq X^2 \mathbf{1}_{|X|>a} \geq a^2 \mathbf{1}_{|X|>a}$ , donc, en prenant l'espérance, on a

$$\mathbb{E}(X^2) \geq \mathbb{E}(X^2 \mathbf{1}_{|X|>a}) \geq a^2 \mathbb{E}(\mathbf{1}_{|X|>a}) = a^2 \mathbb{P}(|X| > a),$$

et le résultat découle en divisant par  $a^2$ . □

*Démonstration du théorème.* Posons  $X = \frac{1}{n} \sum_{i=1}^n X_i - \mathbb{E}(X_i) = \frac{1}{n} \sum_{i=1}^n (X_i - \mathbb{E}(X_i))$ . On a  $\mathbb{E}(X) = 0$ , donc  $\text{Var}(X) = \mathbb{E}(X^2)$ . On en déduit que

$$\begin{aligned} \mathbb{E}(X^2) = \text{Var}(X) &= \frac{1}{n^2} \text{Var}\left(\sum_{i=1}^n (X_i - \mathbb{E}(X_i))\right) \\ &= \frac{1}{n^2} \sum_{i=1}^n \text{Var}(X_i) \quad \text{par la Proposition 2.5} \\ &= \frac{\text{Var}(X_1)}{n}. \end{aligned}$$

On peut alors appliquer le lemme à  $X$  et à  $a = \varepsilon$  pour conclure. □

# Chapitre 5

## Probabilités discrètes

**Problème mathématique :** Il arrive souvent qu'une expérience aléatoire puisse avoir pour résultat n'importe quel entier positif. Par exemple, une expérience aléatoire peut être le nombre de tours que durera une partie de Monopoly, ou encore le nombre d'e-mails que je recevrai aujourd'hui.  
Comment faire des probabilités dans un espace infini ?

### 5.1 Rappels sur la dénombrabilité

**Définition 5.1.** Un ensemble  $E$  est dit **dénombrable** s'il existe une application bijective  $\varphi : \mathbb{N} \rightarrow E$ . Autrement dit, il existe une manière de numérotter les éléments de  $E$ . Dit encore autrement, on peut écrire

$$E = \{x_1, x_2, \dots, x_n, \dots\}.$$

On dit qu'un ensemble est **au plus dénombrable** s'il est fini ou dénombrable.

**Remarque 5.1.** En fait, pour montrer qu'un ensemble  $E$  est au plus dénombrable, il suffit de montrer qu'il existe une application injective  $E \rightarrow \mathbb{N}$ , ou qu'il existe une application surjective  $\mathbb{N} \rightarrow E$ .

Il est souvent un peu technique de montrer qu'un ensemble est dénombrable. Toutefois, un dessin permet parfois de se convaincre qu'un ensemble est dénombrable, en indiquant une manière de numérotter ses éléments : voir par exemple les figures 5.1 et 5.2.

Le théorème suivant est admis.

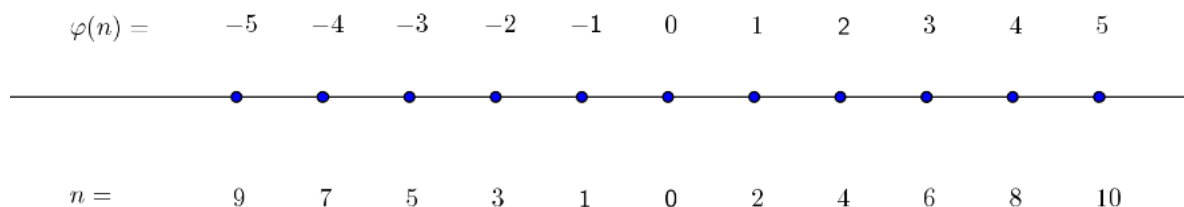


FIGURE 5.1 –  $\mathbb{Z}$  est dénombrable

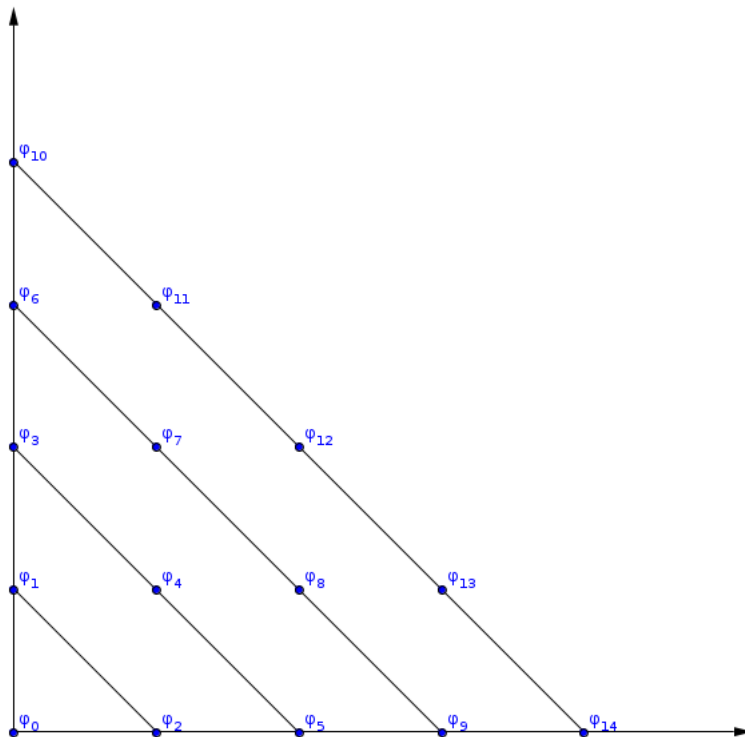


FIGURE 5.2 –  $\mathbb{N}^2$  est dénombrable

- Théorème 5.1.**
1. Les ensembles  $\mathbb{N}$ ,  $\mathbb{Z}$  et  $\mathbb{Q}$  sont dénombrables.
  2. Si  $E$  est un ensemble dénombrable, alors  $E \times E$  est un ensemble dénombrable, et, plus généralement,  $E^n$  est un ensemble dénombrable pour tout  $n \in \mathbb{N}$ . En particulier,  $\mathbb{N}^n$ ,  $\mathbb{Z}^n$  et  $\mathbb{Q}^n$  sont dénombrables pour tout  $n$ .
  3. Si  $E$  est dénombrable et si  $F \subset E$ , alors  $F$  est au plus dénombrable.
  4. Les intervalles de  $\mathbb{R}$  non vides et non réduits à des singletons ne sont jamais dénombrables. L'ensemble  $\{0, 1\}^{\mathbb{N}}$  des suites à valeurs dans  $\{0, 1\}$  n'est pas dénombrable.
  5. Si  $E$  est un ensemble infini,  $\mathcal{P}(E)$  n'est jamais dénombrable.

**Un peu d'histoire :** La notion d'ensemble dénombrable a été inventée par Cantor en 1874, dans un article où il prouve que  $\mathbb{R}$  n'est pas un ensemble dénombrable. Cet article, qui a fondé la théorie des ensembles, semblait très abstrait à ses contemporains. Pourtant, la dénombrabilité se retrouvera partout dans les mathématiques du XXème siècle : en topologie, en analyse, en probabilités...

## 5.2 Rappels sur les séries

### Séries à termes positifs

Soit  $(u_n)_{n \in \mathbb{N}}$  une suite de nombres positifs :  $\forall n \in \mathbb{N}, u_n \geq 0$ . La suite  $S_n = \sum_{k=0}^n u_k$  est alors une suite croissante. On sait donc qu'elle possède toujours une limite, qui est un nombre réel ou  $+\infty$ .

**Définition 5.2.** On note

$$\sum_{k=0}^{+\infty} u_k := \lim_{n \rightarrow \infty} \sum_{k=0}^n u_k.$$

Cette quantité est parfois aussi notée  $\sum_{n \in \mathbb{N}} u_n$ .

- Si  $\sum_{k=0}^{+\infty} u_k < +\infty$ , on dit que la **série**  $\sum u_k$  est **convergente**.
- Si  $\sum_{k=0}^{+\infty} u_k = +\infty$ , on dit que la **série**  $\sum u_k$  est **divergente**.

Plus généralement, si  $p \in \mathbb{N}$ , on note  $\sum_{k=p}^{+\infty} u_k := \lim_{n \rightarrow \infty} \sum_{k=p}^n u_k$ . Cette quantité est aussi parfois notée  $\sum_{k \geq p} u_k$ . Le fait que la série soit convergente ou divergente ne dépend pas de la valeur de  $p$  choisie.

**Exemple 5.1.** — Vous savez que, si  $q \geq 0$ ,  $q \neq 1$ , on a  $\sum_{k=0}^n q^k = \frac{q^{n+1}-1}{q-1}$ . Par conséquent, la série  $\sum q^n$  est convergente si et seulement si  $q < 1$ , et on a alors  $\sum_{k \geq 0} q^k = \frac{1}{1-q}$ .

- On peut montrer que la série  $\sum \frac{1}{n}$  est divergente.
- On peut montrer que la série  $\sum \frac{1}{n^2}$  est convergente, et qu'on a  $\sum_{n=1}^{+\infty} \frac{1}{n^2} = \frac{\pi^2}{6}$ .

Le théorème suivant nous donne un critère extrêmement utile pour montrer qu'une série converge. Vous avez vu (ou vous verrez) sa démonstration dans votre cours d'analyse.

**Théorème 5.2** (Critère de Riemann). Soit  $(u_n)_{n \in \mathbb{N}}$  une suite de nombres positifs. S'il existe  $\alpha > 1$  tel que  $n^\alpha u_n \rightarrow 0$ , alors  $\sum u_n$  converge.

**Exemple 5.1.** La série  $\sum e^{-n}$  est convergente. En effet, par croissance comparée, on a  $n^2 e^{-n} \rightarrow 0$ .

## Séries dont le terme est de signe quelconque

**Définition 5.3.** Soit  $u_n$  une suite de nombre réels. On dit que la série  $\sum u_n$  est convergente si la suite  $\sum_{k=0}^n u_k$  est convergente. Sinon, on dit que la série est divergente. Lorsque la série  $\sum u_n$  est convergente, on note

$$\sum_{k=0}^{+\infty} u_k := \lim_{n \rightarrow \infty} \sum_{k=0}^n u_k.$$

Le théorème suivant est très utile pour montrer qu'une série dont le terme général est de signe quelconque est convergente. Vous verrez (ou avez vu) sa démonstration dans votre cours d'analyse.

**Théorème 5.3.** Soit  $(u_n)_{n \in \mathbb{N}}$  une suite de nombres réels tels que  $\sum |u_k|$  est convergente. Alors la série  $\sum u_k$  est convergente. On dit alors que la série  $\sum u_k$  est **absolument convergente**. On a alors

$$\left| \sum_{k=0}^{+\infty} u_k \right| \leq \sum_{k=0}^{+\infty} |u_k|.$$

**Exemple 5.2.** La série  $\sum e^{-n}$  étant convergente, la série  $\sum (-1)^n e^{-n}$  est aussi convergente.

## Échanges de sommations

**Proposition 5.1** (Pour une série absolument convergente, l'ordre de sommation est sans importance). Soit  $(u_n)_{n \in \mathbb{N}}$  une suite de nombres réels, et soit  $\varphi : \mathbb{N} \rightarrow \mathbb{N}$  une bijection. Alors  $\sum |u_n|$  est convergente si et seulement si  $\sum |u_{\varphi(n)}|$  est convergente. Si c'est le cas, on a alors

$$\sum_{n \in \mathbb{N}} u_n = \sum_{n \in \mathbb{N}} u_{\varphi(n)}. \quad (5.1)$$

En particulier, le résultat s'applique lorsque  $u_n$  est une série à termes positifs convergente.

**Attention!** Le résultat n'est en général pas vrai lorsque  $\sum |u_n|$  n'est pas convergente!

*Démonstration.* Pour tout  $n \in \mathbb{N}$ , on a  $\sum_{k=0}^n |u_{\varphi(k)}| \leq \sum_{k=0}^{+\infty} |u_k|$ , donc si  $\sum |u_k|$  est convergente, alors  $\sum |u_{\varphi(k)}|$  est aussi convergente. De même,  $\sum_{k=0}^n |u_k| \leq \sum_{k=0}^{+\infty} |u_{\varphi(k)}|$ , donc si  $\sum |u_{\varphi(k)}|$  est convergente, alors  $\sum |u_k|$  est aussi convergente.

Dans ce cas, par le Théorème 5.3,  $\sum u_k$  et  $\sum u_{\varphi(k)}$  sont bien convergents.

Montrons maintenant l'égalité (5.1). Soit  $\varepsilon > 0$ . Comme  $\sum |u_k|$  converge, il existe  $n_0 \in \mathbb{N}$  tel que  $\sum_{k=n_0+1}^{+\infty} |u_k| < \varepsilon$ .

$\varphi$  étant une bijection, il existe  $n_1 \geq n_0$  tel que  $\{0, 1, \dots, n_0\} \subset \{\varphi(0), \varphi(1), \dots, \varphi(n_1)\}$ . On a donc

$$\begin{aligned} \left| \sum_{k=0}^{+\infty} u_k - \sum_{k=0}^{+\infty} u_{\varphi(k)} \right| &= \left| \sum_{k=0}^{+\infty} u_k - \sum_{k=0}^{n_0} u_k + \sum_{k=0}^{n_0} u_k - \sum_{k=0}^{n_1} u_{\varphi(k)} + \sum_{k=0}^{n_1} u_{\varphi(k)} - \sum_{k=0}^{+\infty} u_{\varphi(k)} \right| \\ &\leq \left| \sum_{k=0}^{+\infty} u_k - \sum_{k=0}^{n_0} u_k \right| + \left| \sum_{k=0}^{n_0} u_k - \sum_{k=0}^{n_1} u_{\varphi(k)} \right| + \left| \sum_{k=0}^{n_1} u_{\varphi(k)} - \sum_{k=0}^{+\infty} u_{\varphi(k)} \right| \\ &\leq \sum_{k=n_0+1}^{+\infty} |u_k| + \sum_{k=n_0+1}^{+\infty} |u_k| + \sum_{k=n_1+1}^{+\infty} |u_{\varphi(k)}| \\ &\leq 3 \sum_{k=n_0+1}^{+\infty} |u_k| \leq 3\varepsilon. \end{aligned}$$

Ceci étant vrai pour tout  $\varepsilon > 0$ , on en déduit bien le résultat.  $\square$

**Proposition-Definition 5.1.** Soit  $\Omega$  un ensemble dénombrable, et soit  $f : \Omega \rightarrow [0, +\infty[$  une fonction toujours positive.

Alors, si  $\varphi : \mathbb{N} \rightarrow \Omega$  est une bijection, la quantité  $\sum_{n \in \mathbb{N}} f(\varphi(n))$  ne dépend pas du choix de la bijection  $\varphi$ , et cette quantité est notée

$$\sum_{\omega \in \Omega} f(\omega).$$

*Démonstration.* Soient  $\varphi_1, \varphi_2$  des bijections de  $\mathbb{N}$  dans  $\Omega$ . Alors  $\varphi := (\varphi_1)^{-1} \circ \varphi_2$  est une bijection de  $\mathbb{N}$  dans  $\mathbb{N}$ . En notant  $u_n := f(\varphi_1(n))$ , on a  $u_{\varphi(n)} = f(\varphi_2(n))$ , et par la proposition 5.1, on a bien  $\sum_{n \in \mathbb{N}} f(\varphi_2(n)) = \sum_{n \in \mathbb{N}} u_{\varphi(n)} = \sum_{n \in \mathbb{N}} u_n = \sum_{n \in \mathbb{N}} f(\varphi_1(n))$ , d'où le résultat.  $\square$

Dans la suite, on utilisera souvent le principe (un peu vague) suivant, dans le même esprit que la Proposition 5.1 :

**Principe général :** Quand on fait une somme dénombrable d'éléments **positifs**, l'ordre dans lequel on les somme n'a pas d'importance.

**Attention!** Ce principe est faux si on somme des nombres de signe quelconque. Par exemple, supposons que l'on cherche à calculer la somme  $S = 1 - 1 + 1 - 1 + 1 - 1 + 1 - 1 + 1 - 1 + \dots$ . Si mon principe était vrai, je pourrais regrouper les termes deux à deux et obtenir  $S = (1 - 1) + (1 - 1) + (1 - 1) + \dots = 0 + 0 + 0 + \dots = 0$ . Mais je pourrais aussi regrouper les termes comme  $S = 1 + (-1 + 1) + (-1 + 1) + (-1 + 1) + \dots = 1 + 0 + 0 + 0 + \dots = 1$ .  $S$  vaudrait donc à la fois zéro et un : une contradiction!

### 5.3 Probabilités sur les ensembles dénombrables

**Définition 5.4.** Soit  $\Omega$  un ensemble au plus dénombrable. Une **mesure de probabilité** sur  $\Omega$  est une application  $\mathbb{P} : \mathcal{P}(\Omega) \rightarrow [0, 1]$  vérifiant les propriétés suivantes.

- (1)  $\mathbb{P}(\Omega) = 1$ .
- (2) Si  $A, B \subset \Omega$  vérifient  $A \cap B = \emptyset$ , alors  $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$ .
- (3) Pour toute suite  $(A_n)_{n \in \mathbb{N}}$  d'éléments deux à deux incompatibles (c'est-à-dire telle que,  $\forall i, j \in \mathbb{N}, i \neq j \implies A_i \cap A_j = \emptyset$ ), on a

$$\mathbb{P}\left(\bigcup_{n \in \mathbb{N}} A_n\right) = \sum_{n \in \mathbb{N}} \mathbb{P}(A_n). \quad (5.2)$$

On dit alors que  $(\Omega, \mathbb{P})$  est un **espace de probabilité au plus dénombrable**.

**Remarque 5.2.** Remarquons que si  $(A_n)_{n \in \mathbb{N}}$  est une suite d'événements deux à deux incompatibles (c'est-à-dire que  $A_n \cap A_m = \emptyset$  si  $n \neq m$ ), alors pour tout  $N \in \mathbb{N}$ , une récurrence immédiate utilisant la propriété 2 ci-dessus nous dit que

$$\mathbb{P}\left(\bigcup_{n=0}^N A_n\right) = \sum_{n=0}^N \mathbb{P}(A_n). \quad (5.3)$$

Le terme de gauche dans (5.3) est toujours borné par 1. Par conséquent, la somme  $\sum \mathbb{P}(A_n)$  est bien convergente.

On rappelle que l'ensemble  $\bigcup_{n \in \mathbb{N}} A_n = \{\omega \in \Omega; \exists n \in \mathbb{N} \text{ tel que } \omega \in A_n\} \subset \Omega$ .

Les deux termes de (5.2) sont donc toujours bien défini. Toutefois, on ne peut pas déduire des hypothèses 1 et 2 qu'ils sont égaux : l'hypothèse 3 est vraiment une hypothèse supplémentaire qui complète la définition.

**Proposition 5.2.** Soit  $\Omega$  un ensemble au plus dénombrable, et soit  $\mathbb{P} : \mathcal{P}(\Omega) \rightarrow [0, 1]$  une application vérifiant les propriétés 1. et 2. ci-dessus. La propriété 3. est alors équivalente à la propriété suivante.

- (3') Pour toute suite  $(A_n)_{n \in \mathbb{N}}$  croissante d'événements (c'est-à-dire que  $\forall n \in \mathbb{N}, A_n \subset A_{n+1}$ ), on a

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

*Démonstration.* **Supposons (3) et montrons (3').**

Soit  $(A_n)_{n \in \mathbb{N}}$  une suite croissante d'événements. On pose  $B_0 = A_0$ , et, pour tout  $n \geq 1$ ,  $B_n = A_n \setminus A_{n-1}$ . Montrons que les événements  $(B_n)_{n \in \mathbb{N}}$  sont deux à deux disjoints. Soit  $n < n'$ . On a  $B_n \subset A_n \subset A_{n'-1}$ , car la suite  $(A_n)$  est croissante. Or  $B_{n'} \cap A_{n'-1} = \emptyset$ . On a donc bien  $B_n \cap B_{n'} = \emptyset$ .

Vérifions que l'on a  $\bigcup_{n=0}^N B_n = \bigcup_{n=0}^N A_n$ . Pour  $N = 0$ , on a bien  $A_0 = B_0$ . Supposons le résultat vrai au rang  $N$  et montrons-le au rang  $N + 1$ . Soit  $\omega \in A_{N+1}$ . Si  $\omega \in A_N$ , alors, par hypothèse de récurrence,  $\omega \in \bigcup_{n=0}^N B_n$ . Si  $\omega \notin A_N$ , alors  $\omega \in A_{N+1} \setminus A_N = B_{N+1}$ , donc  $\omega \in \bigcup_{n=0}^{N+1} B_n$ . On a donc  $A_{N+1} \subset \bigcup_{n=0}^{N+1} B_n$ , donc  $\bigcup_{n=0}^{N+1} A_n \subset \bigcup_{n=0}^{N+1} B_n$ . Réciproquement, pour tout  $n \in \mathbb{N}$ , on a  $B_n \subset A_n$ , donc  $\bigcup_{n=0}^{N+1} B_n \subset \bigcup_{n=0}^{N+1} A_n$ . On a donc bien le résultat par récurrence.

En prenant la limite  $N \rightarrow \infty$ , on obtient que  $\bigcup_{n \in \mathbb{N}} A_n = \bigcup_{n \in \mathbb{N}} B_n$ .

On peut donc appliquer le résultat (3) pour obtenir que

$$\begin{aligned}
\mathbb{P}\left(\bigcup_{n \in \mathbb{N}} A_n\right) &= \mathbb{P}\left(\bigcup_{n \in \mathbb{N}} B_n\right) \\
&= \sum_{n \in \mathbb{N}} \mathbb{P}(B_n) \\
&= \lim_{n \rightarrow \infty} \sum_{k=0}^n \mathbb{P}(B_k) \\
&= \lim_{n \rightarrow \infty} \mathbb{P}\left(\bigcup_{k=0}^n B_k\right) \\
&= \lim_{n \rightarrow \infty} \mathbb{P}\left(\bigcup_{k=0}^n A_k\right).
\end{aligned}$$

Or, la suite  $A_n$  étant croissante, on a  $\bigcup_{k=0}^n A_k = A_n$ . On déduit donc bien (3').

**Supposons (3') et montrons (3).** Soit  $(A_n)_{n \in \mathbb{N}}$  une suite d'événements deux à deux incompatibles. On pose  $C_n := \bigcup_{k=0}^n A_k$ .  $(C_n)_{n \in \mathbb{N}}$  est alors une suite croissante d'événements, et, par (2), on a  $\mathbb{P}(C_n) = \sum_{k=0}^n \mathbb{P}(A_k)$ . On a  $\bigcup_{n \in \mathbb{N}} A_n = \bigcup_{n \in \mathbb{N}} C_n$ , donc en appliquant (3'), on a

$$\begin{aligned}
\mathbb{P}\left(\bigcup_{n \in \mathbb{N}} A_n\right) &= \mathbb{P}\left(\bigcup_{n \in \mathbb{N}} C_n\right) \\
&= \lim_{n \rightarrow \infty} \mathbb{P}(C_n) \\
&= \lim_{n \rightarrow \infty} \sum_{k=0}^n \mathbb{P}(A_k) \\
&= \sum_{n \in \mathbb{N}} \mathbb{P}(A_n).
\end{aligned}$$

□

La propriété (3) s'appelle la propriété de  $\sigma$ -additivité des mesures, tandis que la propriété (3') s'appelle la propriété de *continuité par réunion croissante*.

Sur les univers dénombrables, on peut encore parler de loi de probabilité, comme dans le cas fini.

**Définition 5.5.** Soit  $\Omega$  un ensemble dénombrable. Une **loi de probabilité** sur  $\Omega$  est une application  $q : \Omega \rightarrow [0, 1]$  telle que  $\sum_{\omega \in \Omega} q(\omega) = 1$ .

La proposition suivante nous explique comment passer d'une mesure de probabilité à une loi de probabilité, et réciproquement ; là encore, lorsque  $\Omega$  est un ensemble dénombrable, ces deux notions sont tellement proches qu'on les identifie souvent.

**Proposition 5.3** (Lien entre mesure de probabilité et loi de probabilité). Soit  $\Omega$  un ensemble dénombrable.

1. Si  $\mathbb{P}$  est une mesure de probabilité sur  $\Omega$ , alors l'application  $q : \Omega \rightarrow [0, 1]$  donnée par  $q(\omega) = \mathbb{P}(\{\omega\})$  est une loi de probabilité.
2. Réciproquement, si  $p$  est une loi de probabilité sur  $\Omega$ , alors l'application  $\mathbb{P} : \mathcal{P}(\Omega) \rightarrow [0, 1]$  donnée par  $\mathbb{P}(A) := \sum_{\omega \in A} q(\omega)$  est une mesure de probabilité sur  $\Omega$ .

*Démonstration.* 1.  $\Omega$  étant dénombrable, il existe une bijection  $\varphi : \mathbb{N} \rightarrow \Omega$ , de sorte que  $\Omega = \{\varphi(0), \varphi(1), \dots\}$ . On a donc  $\Omega = \bigcup_{n \in \mathbb{N}} \{\varphi(n)\}$ , et, si  $n \neq n'$ , on a  $\{\varphi(n)\} \cap \{\varphi(n')\} = \emptyset$ . On peut donc appliquer (5.2) pour en déduire que

$$1 = \mathbb{P}(\Omega) = \mathbb{P}\left(\bigcup_{n \in \mathbb{N}} \{\varphi(n)\}\right) = \sum_{n \in \mathbb{N}} \mathbb{P}(\{\varphi(n)\}) = \sum_{\omega \in \Omega} \mathbb{P}(\{\omega\}) = \sum_{\omega \in \Omega} q(\omega).$$

$q$  est donc bien une loi de probabilité.



2. Le calcul précédent nous dit que, avec cette définition de  $\mathbb{P}$ , on a  $\mathbb{P}(\Omega) = \sum_{\omega \in \Omega} q(\omega)$ . Par conséquent, si  $q$  est une loi de probabilité, on a bien  $\mathbb{P}(\Omega) = 1$ .

Ensuite, si  $A, B \subset \Omega$  avec  $A \cap B = \emptyset$ , on a

$$\begin{aligned} \mathbb{P}(A \cup B) &= \sum_{\omega \in A \cup B} q(\omega) \\ &= \sum_{\omega \in A} q(\omega) + \sum_{\omega \in B} q(\omega) \quad \text{car } A \cap B = \emptyset \\ &= \mathbb{P}(A) + \mathbb{P}(B). \end{aligned}$$

$\mathbb{P}$  est donc bien une mesure de probabilité.

Soit  $(A_n)_{n \in \mathbb{N}}$  une suite d'éléments deux à deux incompatibles (c'est-à-dire telle que,  $\forall i, j \in \mathbb{N}, i \neq j \implies A_i \cap A_j = \emptyset$ ). On a

$$\mathbb{P}\left(\bigcup_{n \in \mathbb{N}} A_n\right) = \sum_{\omega \in \bigcup_{n \in \mathbb{N}} A_n} q(\omega)$$

Les  $A_n$  étant deux à deux disjoints, on a

$$\sum_{\omega \in \bigcup_{n \in \mathbb{N}} A_n} q(\omega) = \sum_{n \in \mathbb{N}} \sum_{\omega \in A_n} q(\omega) = \sum_{n \in \mathbb{N}} \mathbb{P}(A_n),$$

donc la propriété de  $\sigma$ -additivité est vérifiée. □

## 5.4 Quelques lois classiques

### La loi géométrique

**Problème concret :** Je lance une pièce jusqu'à obtenir face. Quelle est la probabilité que j'obtienne face pour la première fois au bout de 5 lancers ?

La réponse est fournie par la loi suivante :

**Définition 5.6.** Soit  $p \in ]0, 1[$ . La loi géométrique de paramètre  $p$  est la loi sur  $\mathbb{N}^*$  donnée par

$$q(n) = p(1-p)^{n-1}.$$

On note  $X \sim \text{Geo}(p)$ , si  $X$  suit la loi géométrique de paramètre  $p$ .

Vérifions tout d'abord que cette loi est bien une loi de probabilité. On a

$$\begin{aligned} \sum_{n=1}^{+\infty} q(n) &= \lim_{N \rightarrow \infty} \sum_{n=1}^N p(1-p)^{n-1} \\ &= p \lim_{N \rightarrow \infty} \sum_{n=0}^{N-1} (1-p)^n \\ &= p \lim_{N \rightarrow \infty} \frac{1 - (1-p)^N}{1 - (1-p)} \\ &= p \times \frac{1}{p} = 1. \end{aligned}$$

Vérifions que cette loi répond bien à la question posée. Notons  $X_i$  le résultat du  $i$ -ème lancer de la pièce. On veut déterminer la probabilité de l'événement  $A = \{X_1 = \text{pile}, X_2 = \text{pile}, X_3 = \text{pile}, X_4 = \text{pile}, X_5 = \text{face}\}$ . Notons  $p$  la probabilité d'obtenir face (qui vaut  $\frac{1}{2}$  si la pièce est équilibrée). Les lancers étant considérés comme indépendants, la probabilité de  $A$  est

$$\mathbb{P}(A) = \left( \prod_{i=1}^4 \mathbb{P}(X_i = \text{pile}) \right) \times \mathbb{P}(X_5 = \text{face}) = (1-p)^4 \times p = q(5).$$

De façon générale,  $q(n)$  représente la probabilité pour que, en lançant des pièces ayant une probabilité  $p$  de faire face, le premier face arrive au bout de  $n$  lancers.

**Attention!** Dans le raisonnement précédent, on a travaillé sur l'espace des suites de lancers de dés, qui n'est pas dénombrable! Le raisonnement précédent, bien que correct, ne rentre pas dans le cadre des probabilités définies dans ce cours. Vous verrez en TD une autre manière de retrouver la loi géométrique à partir de probabilités finies.

## La loi de Poisson

**Définition 5.7.** Soit  $\lambda > 0$ . La loi de Poisson de paramètre  $\lambda$  est la loi sur  $\mathbb{N}$  donnée par

$$q(n) = \exp(-\lambda) \frac{\lambda^n}{n!}$$

On note  $X \sim \mathcal{P}(\lambda)$  si  $X$  suit la loi de Poisson de paramètre  $\lambda$ .

Le fait que cette expression définisse bien une loi de probabilité découle du développement en série entière de l'exponentielle, que vous avez vu ou que vous verrez en cours d'analyse :

$$e^\lambda = \sum_{n=0}^{+\infty} \frac{\lambda^n}{n!}.$$

La loi de Poisson est très importante à cause du fait suivant, dont la démonstration sera donnée en TD.

Si  $\lambda > 0$ ,  $n \in \mathbb{N}^*$ , on considère des variables aléatoires  $X_\lambda$  et  $Y_n$  avec  $X_\lambda \sim \mathcal{P}(\lambda)$ , et  $Y \sim \mathcal{B}(n, \lambda/n)$ . Alors on a

$$\forall k \in \mathbb{N}, \mathbb{P}(Y_n = k) \longrightarrow \mathbb{P}(X_\lambda = k).$$

Autrement dit, si on regarde un grand nombre  $n$  d'expériences aléatoires indépendantes ayant chacune une probabilité de succès très faible de  $\lambda/n$ , alors le nombre de succès obtenu sera correctement décrit par la loi de Poisson de paramètre  $\lambda$ .

## 5.5 Variables aléatoires sur les espaces dénombrables

Comme précédemment, une variable aléatoire sur un espace dénombrable  $\Omega$  est simplement une application  $X : \Omega \rightarrow \mathbb{R}$ , et son support est  $X(\Omega)$ .

### 5.5.1 Espérance et variance des variables aléatoires sur les espaces dénombrables

**Définition 5.8.** Soit  $X$  une variable aléatoire sur  $\Omega$ . On dit que  $X$  est **intégrable** si  $\sum_{x \in X(\Omega)} |x| \mathbb{P}(X = x) < +\infty$ . Dans ce cas, on définit son espérance  $\mathbb{E}(X)$  par

$$\mathbb{E}(X) = \sum_{x \in X(\Omega)} x \mathbb{P}(X = x).$$

#### Quelques critères pour qu'une variable soit intégrable

**Remarque 5.3.** — Une variable aléatoire de support fini est toujours intégrable.

- Le caractère intégrable d'une variable aléatoire, ainsi que son espérance, ne dépendent que de sa loi :  $X \stackrel{\mathcal{L}}{=} Y \implies \mathbb{E}(X) = \mathbb{E}(Y)$ .
- Toutes les variables aléatoires ne sont pas intégrables ! Par exemple, sur  $\mathbb{N}^*$ , on peut définir une mesure de probabilité par  $\mathbb{P}(\{n\}) = \frac{6}{\pi^2 n^2}$ . La variable aléatoire  $X(n) = n$  n'est alors pas intégrable :  $\sum_{n \in \mathbb{N}^*} n \mathbb{P}(X = n) = \frac{6}{\pi^2} \sum_{n \in \mathbb{N}^*} \frac{1}{n} = +\infty$ .
- Soient  $X, Y : \Omega \rightarrow \mathbb{R}$  des variables aléatoires intégrables, et soit  $\lambda \in \mathbb{R}$ . Alors  $\lambda X + Y$  est intégrable, et on a  $\mathbb{E}(\lambda X + Y) = \lambda \mathbb{E}(X) + \mathbb{E}(Y)$ . Cela se prouve comme la proposition 2.4.

**Lemme 5.1.** Soit  $Y$  une variable aléatoire intégrable, et soit  $X$  une variable aléatoire telle que  $|X| \leq |Y|$ . Alors  $X$  est intégrable.

*Démonstration.* Par le lemme 2.1, on a

$$\begin{aligned} \sum_{x \in X(\Omega)} |x| \mathbb{P}(X = x) &= \sum_{(x,y) \in X(\Omega) \times Y(\Omega)} |x| \mathbb{P}(X = x, Y = y) \\ &\leq \sum_{(x,y) \in X(\Omega) \times Y(\Omega)} |y| \mathbb{P}(X = x, Y = y) \\ &= \sum_{y \in Y(\Omega)} |y| \mathbb{P}(Y = y), \end{aligned}$$

qui est fini, car  $Y$  est intégrable.  $X$  est donc bien intégrable.  $\square$

Rappelons la formule de transfert, pour les variables sur les espaces dénombrables.

**Théorème 5.4** (Formule de transfert). Soit  $X$  une variable aléatoire, et soit  $f : \mathbb{R} \rightarrow \mathbb{R}$ . La variable aléatoire  $f(X)$  est intégrable si et seulement si  $\sum_{x \in X(\Omega)} |f(x)| \mathbb{P}(X = x) < \infty$ , et on a alors

$$\mathbb{E}(f(X)) = \sum_{x \in X(\Omega)} f(x) \mathbb{P}(X = x). \quad (5.4)$$

*Démonstration.* Notons  $Y = f(X)$ . On a  $Y(\Omega) = f(X(\Omega)) = \{f(x); x \in X(\Omega)\}$ . On a

$$\begin{aligned} \sum_{y \in Y(\Omega)} |y| \mathbb{P}(Y = y) &= \sum_{y \in Y(\Omega)} |y| \sum_{x \in X(\Omega); f(x)=y} \mathbb{P}(X = x) \\ &= \sum_{x \in X(\Omega)} |f(x)| \mathbb{P}(X = x), \end{aligned}$$

donc  $Y$  est intégrable si et seulement si  $\sum_{x \in X(\Omega)} |f(x)| \mathbb{P}(X = x) < \infty$ . L'équation (2.3) découle alors du même calcul, en enlevant les valeurs absolues.  $\square$

**Remarque 5.4.** Si  $f : \mathbb{R} \rightarrow \mathbb{R}$  est bornée, alors  $f(X)$  est intégrable. En effet,

$$\sum_{x \in X(\Omega)} |f(x)| \mathbb{P}(X = x) \leq \sup_{y \in \mathbb{R}} |f(y)| \sum_{x \in X(\Omega)} \mathbb{P}(X = x) = \sup_{y \in \mathbb{R}} |f(y)| < +\infty.$$

### Deux calculs classiques d'espérances

Si  $X \sim \mathcal{Geo}(p)$ , alors

$$\begin{aligned} \mathbb{E}(X) &= \sum_{n=1}^{+\infty} np(1-p)^{n-1} \\ &= p \sum_{n=0}^{+\infty} n(1-p)^{n-1} \quad \text{car le terme } n=0 \text{ est nul} \\ &= -p \sum_{n=0}^{+\infty} \frac{d}{dp} (1-p)^n \\ &= -p \frac{d}{dp} \sum_{n=0}^{+\infty} (1-p)^n \\ &= -p \frac{d}{dp} \frac{1}{1-(1-p)} \\ &= \frac{1}{p}. \end{aligned}$$

Pour passer de la troisième à la quatrième ligne, on a utilisé le théorème de *dérivation des séries de fonctions* que vous avez vu dans votre cours d'analyse.

Si  $X \sim \mathcal{P}(\lambda)$ , alors

$$\begin{aligned} \mathbb{E}(X) &= \sum_{n=0}^{+\infty} e^{-\lambda} n \frac{\lambda^n}{n!} \\ &= e^{-\lambda} \sum_{n=1}^{+\infty} \frac{\lambda^n}{(n-1)!} \\ &= \lambda e^{-\lambda} \sum_{n=1}^{+\infty} \frac{\lambda^{n-1}}{(n-1)!} \\ &= \lambda e^{-\lambda} \sum_{n=0}^{+\infty} \frac{\lambda^n}{n!} \\ &= \lambda e^{-\lambda} e^\lambda = \lambda. \end{aligned}$$

### Variables de carré intégrable

**Définition 5.9.** Soit  $X : \Omega \rightarrow \mathbb{R}$  une variable aléatoire. On dit que  $X$  est **de carré intégrable** si  $X^2$  est intégrable, c'est-à-dire si  $\sum_{x \in X(\Omega)} |x|^2 \mathbb{P}(X = x) < +\infty$ . Dans ce cas, on définit sa **variance**  $\text{Var}(X)$  par

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

L'écart-type de  $X$  est la racine carrée de la variance de  $X$ .

**Remarque 5.5.** — Si  $X$  est de carré intégrable, alors  $X$  est intégrable. En effet, on a  $(|X| - 1)^2 \geq 0$ , donc  $|X| \leq \frac{X^2 + 1}{2}$ , qui est intégrable.  $X$  est donc bien intégrable, par le dernier point de la proposition 2.5. Si  $X$  est de carré intégrable,  $(X - \mathbb{E}(X))^2 = X^2 - 2\mathbb{E}(X)X + \mathbb{E}(X)^2$  est donc bien intégrable, et  $\text{Var}(X)$  est bien définie.

— Si  $X, Y$  sont des variables de carré intégrable, alors  $X + Y$  est de carré intégrable, car  $(X + Y)^2 \leq 2X^2 + 2Y^2$ .

### 5.5.2 Variables indépendantes

De même que précédemment, on dit qu'une famille de  $n$  variables indépendantes  $X_1, \dots, X_n$  de supports respectifs  $X_1(\Omega), \dots, X_n(\Omega)$  est indépendante si

$$\forall x_1 \in X_1(\Omega), \dots, \forall x_n \in X_n(\Omega), \quad \mathbb{P}(X_1 = x_1, \dots, X_n = x_n) = \prod_{i=1}^n \mathbb{P}(X_i = x_i).$$

On a alors l'analogie du théorème 4.1.

**Théorème 5.5.** 1. Soient  $X$  et  $Y$  des variables aléatoires indépendantes, et soient  $f, g : \mathbb{R} \rightarrow \mathbb{R}$  des fonctions telles que  $f(X)$  et  $g(Y)$  soient intégrables. Alors  $f(X)g(Y)$  est intégrable, et on a

$$\mathbb{E}(f(X)g(Y)) = \mathbb{E}(f(X))\mathbb{E}(g(Y)). \quad (5.5)$$

2. Réciproquement, si pour tout  $f, g : \mathbb{R} \rightarrow \mathbb{R}$  bornées, on a  $\mathbb{E}(f(X)g(Y)) = \mathbb{E}(f(X))\mathbb{E}(g(Y))$ , alors  $X$  et  $Y$  sont indépendantes.

De même, on a l'analogie de la Proposition 4.3.

**Proposition 5.4.** Soient  $X_1, \dots, X_n$  des variables aléatoires de carré intégrable. Alors  $X_1 + \dots + X_n$  est de carré intégrable. Si de plus  $X_1, \dots, X_n$  sont indépendantes, alors

$$\text{Var}(X_1 + \dots + X_n) = \text{Var}(X_1) + \dots + \text{Var}(X_n).$$

# Chapitre 6

## Variables aléatoires à densité

### 6.1 Définition d'une variable à densité

#### 6.1.1 Mesures de probabilités sur les ensembles quelconques

**Définition 6.1** (Définition naïve d'une mesure de probabilité). Soit  $\Omega$  un ensemble. **Dans ce chapitre, on dira de manière imprécise** qu'une mesure de probabilité sur  $\Omega$  est une application  $\mathbb{P} : \mathcal{P}(\Omega) \rightarrow [0, 1]$  vérifiant les mêmes axiomes que dans le cas dénombrable, c'est-à-dire

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

(2) Si  $A, B \subset \Omega$  vérifient  $A \cap B = \emptyset$ , alors  $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$ .

(3) Pour toute suite  $(A_n)_{n \in \mathbb{N}}$  d'éléments deux à deux incompatibles (c'est-à-dire telle que,  $\forall i, j \in \mathbb{N}, i \neq j \implies A_i \cap A_j = \emptyset$ ), on a

$$\mathbb{P}\left(\bigcup_{n \in \mathbb{N}} A_n\right) = \sum_{n \in \mathbb{N}} \mathbb{P}(A_n). \quad (6.1)$$

**Exemple 6.1.** — Sur  $\Omega = [0, 1]$ , on veut définir  $\mathbb{P}(A)$  comme étant la longueur de  $A$ .

— Sur  $\Omega = [0, 1]^2$ , on veut définir  $\mathbb{P}(A)$  comme étant l'aire de  $A$ .

— Sur  $\Omega = [0, 1]^3$ , on veut définir  $\mathbb{P}(A)$  comme étant le volume de  $A$ .

Cette mesure est appelée la mesure uniforme, ou mesure de Lebesgue sur  $[0, 1]^d$  (pour  $d = 1, 2, 3$ ).

**Un peu d'histoire (et de mathématiques) :** Comment définit-on la longueur de  $A$ ? Si  $A$  est un intervalle,  $A = [a, b]$  (ou  $A = ]a, b[$ ), c'est facile, on prendra  $\mathbb{P}(A) = b - a$ . Si  $A$  est une réunion finie d'intervalles disjoints, on prendra la somme des différents intervalles qui le constituent. Mais pour un  $A$  plus général? Définir la longueur de  $A$  en toute généralité, c'est l'objet de la **théorie de la mesure**, que vous verrez peut-être en L3.

Celle-ci a été inventée au début du XX<sup>ème</sup> par Henri Lebesgue. Toutefois, elle ne permet pas de construire la longueur (ou l'aire, ou le volume...) de tous les ensembles, mais seulement d'une très grande famille d'ensembles.

En fait, on peut montrer qu'il est impossible de définir la longueur de toutes les parties de  $[0, 1]^d$  (ou de  $\mathbb{R}^d$ ) : il existe toujours des parties (construites de façon très étrange) qui ne possèdent pas de longueur, d'aire ou de volume.

Un exemple célèbre : le **paradoxe de Banach-Tarski**. Soit  $B$  une boule de volume 1. Il est possible (mathématiquement) de la découper en un nombre fini de morceaux, puis de recoller les morceaux entre eux de manière à former deux boules de volume 1. L'explication de ce paradoxe : les morceaux sont tellement bizarres qu'ils ne possèdent pas de volume!

Ainsi, sur un ensemble non dénombrable  $\Omega$ , une mesure de probabilité n'est en général pas définie sur tout  $\mathcal{P}(\Omega)$ , mais seulement sur un sous-ensemble de  $\mathcal{P}(\Omega)$ . Un tel sous-ensemble est appelé une **tribu**.

Dans ce cours (et dans la vie de tous les jours), nous négligerons toutes ces subtilités, et nous nous contenterons de la définition « naïve » d'une mesure de probabilité. Celle-ci se définit à partir d'intégrales de fonctions positives sur  $\mathbb{R}$  : rappelons donc quelques notions d'intégration sur des intervalles non bornés.

## 6.1.2 Intégration de fonctions positive

Lorsque  $f : \mathbb{R} \rightarrow [0, +\infty[$  est une fonction continue par morceaux, la fonction  $A \mapsto \int_{-A}^A f(x)dx$  est croissante. Elle admet donc toujours une limite quand  $A \rightarrow +\infty$ , qui est égale à un nombre réel, ou à  $+\infty$ . On peut donc bien définir :

$$\int_{-\infty}^{+\infty} f(x)dx := \lim_{A \rightarrow +\infty} \int_{-A}^A f(x)dx \in [0, +\infty].$$

Cette quantité est parfois notée  $\int_{\mathbb{R}} f(x)dx$ . Si  $\int_{\mathbb{R}} f(x)dx < +\infty$ , on dit que  $f$  est *intégrable sur  $\mathbb{R}$* .

**Attention :** cette quantité n'a en général pas de sens quand  $f$  n'est pas à valeurs positives.

Quand  $f$  est continue par morceaux sur  $[a, +\infty[$  ou sur  $]-\infty, a]$  et à valeurs positives, on peut définir de la même manière pour tout  $a \in \mathbb{R}$  les quantités  $\int_a^{+\infty} f(x)dx := \lim_{A \rightarrow +\infty} \int_a^A f(x)dx$  et  $\int_{-\infty}^a f(x)dx := \lim_{A \rightarrow -\infty} \int_A^a f(x)dx$ , qui appartiennent à  $[0, +\infty]$ .

**Exemple 6.2.** Soit  $f(x) = \frac{1}{x^\alpha}$ , avec  $\alpha \neq 1$ . On a

$$\int_1^A \frac{dx}{x^\alpha} = \left[ \frac{1}{(1-\alpha)x^{\alpha-1}} \right]_1^A = \frac{1}{\alpha-1} - \frac{1}{(\alpha-1)A^{\alpha-1}}.$$

Lorsque  $\alpha > 1$ , cette quantité tend vers  $\frac{1}{1-\alpha}$ , tandis qu'elle diverge quand  $\alpha < 1$ . Enfin, quand  $\alpha = 1$ , on a  $\int_1^A \frac{dx}{x} = \ln A \rightarrow +\infty$ . Ainsi,

$$x \mapsto \frac{1}{x^\alpha} \text{ est intégrable sur } [1, +\infty[ \iff \alpha > 1.$$

**Méthode.** Montrer qu'une fonction est intégrable.

Pour montrer qu'une fonction  $f : \mathbb{R} \rightarrow [0, +\infty[$  est intégrable, on peut :

1. Calculer  $\int_{-A}^A f(x)dx$  à l'aide de méthodes classiques (primitives de fonctions usuelles, intégration par parties, changement de variable...), et montrer que cette quantité admet une limite finie quand  $A \rightarrow +\infty$ .
2. Montrer que  $x^2 f(x) \xrightarrow{|x| \rightarrow +\infty} 0$ .
3. Plus généralement, trouver un  $\alpha > 1$  tel que  $|x|^\alpha f(x) \xrightarrow{|x| \rightarrow +\infty} 0$ .

En effet, il existe alors  $M > 1$  tel que pour tout  $|x| > M$ , on a  $|x|^\alpha f(x) \leq 1$ , c'est-à-dire  $f(x) \leq \frac{1}{|x|^\alpha}$ . On a donc, pour  $A > M$ ,

$$\begin{aligned} \int_{-A}^A f(x)dx &= \int_{-A}^{-M} f(x)dx + \int_{-M}^M f(x)dx + \int_M^A f(x)dx \\ &\leq \int_{-A}^{-M} \frac{1}{|x|^\alpha} dx + \int_{-M}^M f(x)dx + \int_M^A \frac{1}{x^\alpha} dx \\ &\leq \int_{-\infty}^{-M} \frac{1}{|x|^\alpha} dx + \int_{-M}^M f(x)dx + \int_M^{\infty} \frac{1}{x^\alpha} dx, \end{aligned}$$

et on a vu dans l'exemple 6.2 que cette quantité était finie.

## 6.1.3 Densité de probabilité

**Définition 6.2.** Soit  $p : \mathbb{R} \rightarrow [0, +\infty[$  une fonction continue par morceaux. On dit que  $p$  est une **densité de probabilité** si l'on a

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

**Exemple 6.3.** —  $p(x) = \mathbf{1}_{[0,1]}(x)$  est une densité de probabilité;  $\mathbf{1}_{]0,1]}(x)$ ,  $\mathbf{1}_{[0,1[}(x)$  et  $\mathbf{1}_{]0,1[}(x)$  le sont aussi.

—  $p(x) = \lambda \mathbf{1}_{[0,+\infty[}(x)e^{-\lambda x}$  est une densité de probabilité. En effet, on a

$$\begin{aligned} \int_{-A}^A p(x) dx &= \lambda \int_0^A e^{-\lambda x} dx \\ &= \lambda \left[ \frac{e^{-\lambda x}}{-\lambda} \right]_0^A \\ &= 1 - e^{-\lambda A}. \end{aligned}$$

En prenant la limite  $A \rightarrow +\infty$ , on en déduit bien que  $\int_{-\infty}^{+\infty} f(x) dx = 1$ .

**Définition 6.3.** Soit  $X : \Omega \rightarrow \mathbb{R}$  une variable aléatoire. On dit que  $X$  est une **variable à densité** s'il existe une densité de probabilité  $p_X$  telle que, pour tout intervalle  $I \subset \mathbb{R}$ ,  $I = [a, b]$ ,  $I = ]a, b[$ ,  $I = [a, b[$  ou  $I = ]a, b]$ , avec  $a$  et  $b$  éventuellement infinis, on a

$$\mathbb{P}(X \in I) = \int_a^b p_X(x) dx.$$

**Remarque 6.1.** En particulier, on a  $\mathbb{P}(a < X < b) = \mathbb{P}(a \leq X < b)$  pour tous  $a, b$ . Ainsi, si  $X$  est une variable à densité, on a  $\mathbb{P}(X = a) = 0$  pour tout  $a \in \mathbb{R}$ .

**Exemple 6.4** (Quelques lois classiques). Les lois suivantes sont à connaître :

— On dit que  $X$  suit la loi uniforme sur  $[a, b]$ , ce que l'on note  $X \sim \mathcal{U}[a, b]$ , si elle a pour densité

$$p(x) = \begin{cases} \frac{1}{b-a} & \text{si } x \in [a, b] \\ 0 & \text{sinon.} \end{cases}$$

— Soit  $\lambda > 0$ . On dit que  $X$  suit la loi exponentielle de paramètre  $\lambda$ , ce que l'on note  $X \sim \mathcal{E}(\lambda)$  si elle a pour densité  $p(x) = \lambda \mathbf{1}_{[0,+\infty[}(x)e^{-\lambda x}$ .

— Soit  $\mu \in \mathbb{R}$ ,  $\sigma > 0$ , on dit que  $X$  suit la loi gaussienne (ou loi normale) de paramètres  $\mu, \sigma^2$ , ce que l'on note  $X \sim \mathcal{N}(\mu, \sigma^2)$ , si  $X$  a pour densité

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right).$$

Lorsque  $\mu = 0$ ,  $\sigma = 1$ , c'est-à-dire quand  $X \sim \mathcal{N}(0, 1)$ , on dit que  $X$  suit une loi normale centrée réduite,

Le fait que  $p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$  définisse bien une densité de probabilité découle du lemme suivant.

**Remarque 6.2.** Comme toute fonction continue par morceaux, la fonction gaussienne  $f(x) = e^{-\frac{x^2}{2}}$  possède une primitive : par exemple, la fonction  $F(x) = \int_0^x e^{-\frac{t^2}{2}} dt$  est bien une primitive de  $f$ .

Cependant, un théorème célèbre dû à Liouville ( $\approx 1840$ ) affirme que  $F(x)$  ne peut pas être exprimée à l'aide de fonctions élémentaires... Il faut donc la calculer numériquement, et retenir le résultat du lemme suivant. Cependant, les valeurs de  $F(x)$  pour différentes valeurs de  $x$  se trouvent dans des tables (et sur internet).

**Lemme 6.1.** 1. On a

$$\int_{\mathbb{R}} e^{-\frac{x^2}{2}} dx = \sqrt{2\pi}.$$

2. Plus généralement, on a

$$\int_{\mathbb{R}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx = \sqrt{2\pi\sigma^2}.$$



*Démonstration.* Le premier point est admis, peut être prouvé de plusieurs manières différentes (voir la page wikipédia « Gaussian Integral »), par des méthodes d'analyse dont nous n'aurons pas besoin dans ce cours.

Vérifions tout de même que  $e^{-\frac{x^2}{2}}$  est intégrable sur  $\mathbb{R}$ . Par croissance comparée, on a  $x^2 e^{-\frac{x^2}{2}} \rightarrow 0$  quand  $|x| \rightarrow +\infty$ , donc on peut appliquer la méthode ci-dessus.

2. En posant  $y = x - \mu$ , on obtient

$$\int_{-A}^A \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx = \int_{-A-\mu}^{A-\mu} \exp\left(-\frac{y^2}{2\sigma^2}\right) dy,$$

de sorte qu'en passant à la limite  $A \rightarrow +\infty$ , on obtient

$$\int_{\mathbb{R}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx = \int_{\mathbb{R}} \exp\left(-\frac{y^2}{2\sigma^2}\right) dy.$$

On pose ensuite  $z = \frac{y}{\sigma}$ , de sorte que  $dy = \sigma dz$ , et on obtient

$$\int_{-A}^A \exp\left(-\frac{y^2}{2\sigma^2}\right) dy = \sigma \int_{-A/\sigma}^{A/\sigma} \exp\left(-\frac{z^2}{2}\right) dz,$$

donc en passant à la limite, on a bien

$$\int_{\mathbb{R}} \exp\left(-\frac{y^2}{2\sigma^2}\right) dy = \sigma \int_{\mathbb{R}} \exp\left(-\frac{z^2}{2}\right) dz = \sigma\sqrt{2\pi}$$

par le premier point. □

### 6.1.4 Fonction de répartition

**Définition 6.4.** Soit  $X$  une variable aléatoire de densité  $p$ . Sa **fonction de répartition** est la fonction  $F_X : \mathbb{R} \rightarrow [0, 1]$  définie par

$$F_X(x) = \int_{-\infty}^x p(t) dt.$$

**Exemple 6.5.** Si  $X \sim \mathcal{U}[a, b]$ , alors on a  $F_X(x) = 0$  si  $x < a$ ,  $F_X(x) = \frac{x-a}{b-a}$  si  $x \in [a, b]$ , et  $F_X(x) = 1$  si  $x > b$ .

**Exemple 6.6.** Si  $X \sim \mathcal{E}(\lambda)$ , alors on a  $F_X(x) = 0$  si  $x < 0$ , et, si  $x > 0$ , on a

$$\begin{aligned} F_X(x) &= \int_{-\infty}^x \lambda e^{-\lambda t} \mathbf{1}_{t \geq 0} dt \\ &= \lambda \int_0^x e^{-\lambda t} dt \\ &= \lambda \left[ -\frac{e^{-\lambda t}}{\lambda} \right]_0^x \\ &= 1 - e^{-\lambda x}. \end{aligned}$$

**Proposition 6.1.** Soit  $X$  une variable aléatoire de densité  $p$ . Alors sa fonction de répartition  $F_X$  est une fonction continue, croissante, et qui vérifie

$$\begin{aligned} \lim_{x \rightarrow -\infty} F_X(x) &= 0 \\ \lim_{x \rightarrow +\infty} F_X(x) &= 1. \end{aligned}$$

*Démonstration.*  $F_X$  est une primitive de  $p$ , qui est une fonction continue par morceaux. La primitive d'une fonction continue par morceaux est continue. De plus, si  $x_2 \geq x_1$ , on a  $F(x_2) - F(x_1) = \int_{x_1}^{x_2} p(x)dx \geq 0$ , car  $p$  est à valeurs positives.  $F$  est donc bien croissante.

Enfin,  $\lim_{x \rightarrow +\infty} F(x) = \int_{-\infty}^{+\infty} p(t)dt = 1$ , car  $p$  est une densité de probabilité, et

$$\lim_{x \rightarrow -\infty} F(x) = \lim_{x \rightarrow -\infty} \int_{-\infty}^x p(t)dt = 0.$$

□

## 6.2 Espérance et variance d'une variable aléatoire à densité

**Définition 6.5.** Soit  $X$  une variable aléatoire de densité  $p$ . On dit que  $X$  est **intégrable** si  $\int_{\mathbb{R}} |x|p(x)dx < \infty$ . On définit alors l'espérance de  $X$  comme

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

**Remarque 6.3.** — L'espérance d'une variable aléatoire à densité ne dépend que de sa densité.

— Soient  $X$  et  $Y$  des variables aléatoires intégrables, et  $\lambda \in \mathbb{R}$ . Alors  $\lambda X + Y$  est intégrable, et on a

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

— **Attention !** Il ne faut pas confondre « une fonction  $f$  est intégrable », et « une variable aléatoire  $X$  est intégrable » (ce qui signifie qu'elle a une densité  $p_X$  telle que  $|x|p_X(x)$  est intégrable).

**Exemple 6.7.** Soit  $X \sim \mathcal{U}[a, b]$ . On a  $\int_{\mathbb{R}} |x|p(x)dx = \frac{1}{b-a} \int_a^b |x|dx < \infty$ , car il s'agit de l'intégrale d'une fonction continue sur un intervalle borné, et

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

**Exemple 6.8.** Soit  $\lambda > 0$ , et soit  $X \sim \mathcal{E}(\lambda)$ . On a

$$\int_{-A}^A |x|p(x)dx = \lambda \int_0^A xe^{-\lambda x} dx.$$

On effectue une intégration par parties, en posant  $u(x) = x$ ,  $v'(x) = e^{-\lambda x}$ ,  $v(x) = -\frac{1}{\lambda}e^{-\lambda x}$ ,  $u'(x) = 1$ . On a

$$\begin{aligned} \lambda \int_0^A xe^{-\lambda x} dx &= [-xe^{-\lambda x}]_0^A + \int_0^A e^{-\lambda x} dx \\ &= -Ae^{-\lambda A} + \frac{1}{\lambda} - \frac{e^{-\lambda A}}{\lambda}. \end{aligned}$$

Par croissance comparée, on a  $\lim_{A \rightarrow +\infty} Ae^{-\lambda A} = 0$ , et  $\lim_{A \rightarrow +\infty} \frac{1}{\lambda}e^{-\lambda A} = 0$ . On en déduit donc que  $X$  est intégrable, et que

$$\mathbb{E}(X) = \frac{1}{\lambda}.$$

**Exemple 6.9.** Soit  $X \sim \mathcal{N}(\mu, \sigma)$ . Vérifions que la variable  $X$  est intégrable. On a  $x^2 \times |x|p(x) = \frac{|x|^3}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$ , qui tend vers zéro quand  $|x| \rightarrow +\infty$ . On en déduit donc que  $X$  est intégrable.

Pour tout  $A > 0$ , on a

$$\begin{aligned} \int_{-A}^A \frac{x}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx &= \int_{-A}^A \frac{x-\mu}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx + \mu \int_{-A}^A \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx \\ &= \int_{-A-\mu}^{A-\mu} \frac{y}{\sqrt{2\pi}\sigma} \exp\left(-\frac{y^2}{2\sigma^2}\right) dy + \mu \int_{-A}^A \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx, \end{aligned}$$

en posant  $y = x - \mu$  dans la première intégrale. La seconde intégrale tend vers

$$\mu \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx = \mu,$$

par le lemme 6.1. Quand à la première, elle tend vers  $\int_{\mathbb{R}} \frac{y}{\sqrt{2\pi}\sigma} \exp\left(-\frac{y^2}{2\sigma^2}\right) dy$ , qui vaut zéro, car la fonction intégrée est impaire. On a donc

$$\mathbb{E}(X) = \mu.$$

**Exemple 6.10** (La loi de Cauchy). Soit  $X$  une variable aléatoire suivant la loi de Cauchy :  $p_X(x) = \frac{1}{\pi} \frac{1}{1+x^2}$ . Il s'agit bien d'une densité de probabilité, car

$$\int_{-A}^A \frac{1}{\pi} \frac{dx}{1+x^2} = \frac{1}{\pi} [\arctan x]_{-A}^A = \frac{2}{\pi} \arctan A,$$

qui tend bien vers 1 quand  $A \rightarrow +\infty$ .

Cette variable n'est pas intégrable (et on ne peut donc pas calculer son espérance). En effet, on a

$$\begin{aligned} \int_{-A}^A \frac{|x|}{1+x^2} dx &= 2 \int_0^A \frac{x}{1+x^2} dx \\ &= \left[ \ln(1+x^2) \right]_0^A \\ &= \ln(1+A^2), \end{aligned}$$

qui tend vers  $+\infty$ .

**Attention** à ne pas tomber dans le piège suivant : pour tout  $A > 0$ , on a  $\int_{-A}^A \frac{x}{1+x^2} dx = 0$ , car la fonction  $x \mapsto \frac{x}{1+x^2}$  est impaire. On pourrait donc être tentés de conclure que  $\mathbb{E}(X) = 0$ ...

**Définition 6.6.** Soit  $X$  une variable aléatoire de densité  $p$ . On dit que  $X$  est **de carré intégrable** si  $\int_{\mathbb{R}} x^2 p(x) dx < \infty$ . On définit alors la **variance** de  $X$  comme

$$\begin{aligned} \text{Var}(X) &= \mathbb{E}(X - \mathbb{E}(X))^2 \\ &= \int_{\mathbb{R}} \left( x - \int_{\mathbb{R}} yp(y) \right)^2 p(x) dx. \end{aligned}$$

On définit aussi l'écart-type de  $X$  comme  $\sigma(X) = \sqrt{\text{Var}(X)}$ .

**Remarque 6.4.** Comme dans le cas discret, on a  $\text{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$ .

La variance (et l'écart-type) décrivent l'étalement d'une variable aléatoire autour de sa moyenne.

Le lemme suivant est très utile pour calculer la variance d'une variable aléatoire.

**Lemme 6.2.** Soit  $X$  une variable de densité  $p$  et de carré intégrable. On a

$$\mathbb{E}(X^2) = \int_{\mathbb{R}} x^2 p(x) dx.$$

Par conséquent,

$$\text{Var}(X) = \int_{\mathbb{R}} x^2 p(x) dx - \left( \int_{\mathbb{R}} xp(x) dx \right)^2.$$

**Remarque 6.5.** Ce résultat peut-être vu comme un analogue pour les variables à densité de la formule de transfert du théorème 2.2, dans le cas particulier  $f(x) = x^2$ .

*Démonstration.* Soient  $0 < a < b$ . On a

$$\begin{aligned} \mathbb{P}(a < X^2 < b) &= \mathbb{P}(\sqrt{a} < X < \sqrt{b}) + \mathbb{P}(-\sqrt{b} < X < -\sqrt{a}) \\ &= \int_{\sqrt{a}}^{\sqrt{b}} p(x) dx + \int_{-\sqrt{b}}^{-\sqrt{a}} p(x) dx \\ &= \int_a^b \frac{p(\sqrt{y})}{2\sqrt{y}} dy + \int_a^b \frac{p(-\sqrt{y})}{2\sqrt{y}} dy, \end{aligned}$$

Pour la première intégrale, on fait un changement de variable, en posant  $y = x^2$ . On a alors  $x = \sqrt{y}$ , et  $dx = \frac{dy}{2\sqrt{y}}$ . Le paramètre  $x$  allant de  $\sqrt{a}$  à  $\sqrt{b}$ , le paramètre  $y$  va de  $a$  à  $b$ , et donc

$$\int_{\sqrt{a}}^{\sqrt{b}} p(x) dx = \int_a^b \frac{p(\sqrt{y})}{2\sqrt{y}} dy.$$

De même, en posant  $y = -x^2$ , la seconde intégrale devient

$$\int_{-\sqrt{b}}^{-\sqrt{a}} p(x) dx = \int_a^b \frac{p(-\sqrt{y})}{2\sqrt{y}} dy.$$

On a donc

$$\mathbb{P}(a < X^2 < b) = \int_a^b \frac{p(\sqrt{y}) + p(-\sqrt{y})}{2\sqrt{y}} dy.$$

Ainsi,  $X^2$  est une variable à densité, de densité

$$y \mapsto \begin{cases} \frac{p(\sqrt{y}) + p(-\sqrt{y})}{2\sqrt{y}} & \text{si } y > 0 \\ 0 & \text{sinon.} \end{cases}$$

On a donc

$$\int_0^A y \frac{p(\sqrt{y}) + p(-\sqrt{y})}{2\sqrt{y}} dy = \frac{1}{2} \int_0^A \sqrt{y} (p(\sqrt{y}) + p(-\sqrt{y})) dy.$$

On pose  $x = \sqrt{y}$ , de sorte que  $dy = 2x dx$ , et on obtient

$$\begin{aligned} \int_0^A y \frac{p(\sqrt{y}) + p(-\sqrt{y})}{2\sqrt{y}} dy &= \int_0^A x^2 (p(x) + p(-x)) dx \\ &= \int_{-A}^A x^2 p(x) dx, \end{aligned}$$

d'où l'on déduit le résultat en faisant tendre  $A$  vers  $+\infty$ . □

**Exemple 6.11.** Soit  $\lambda > 0$ , et soit  $X \sim \mathcal{E}(\lambda)$ . Montrons que  $X$  est de carré intégrable. On a

$$x^2 \times \lambda x^2 \mathbf{1}_{x \geq 0} e^{-\lambda x} \xrightarrow{|x| \rightarrow +\infty} 0,$$

donc  $\lambda x^2 \mathbf{1}_{x \geq 0} e^{-\lambda x}$  est bien intégrable.

Calculons sa variance en utilisant le lemme 6.2. On a  $\mathbb{E}(X) = \frac{1}{\lambda}$ . D'autre part, on a

$$\int_{-A}^A \lambda x^2 \mathbf{1}_{x \geq 0} e^{-\lambda x} dx = \lambda \int_0^A x^2 e^{-\lambda x} dx.$$

On effectue une intégration par parties, en prenant  $u' = e^{-\lambda x}$ ,  $u = -\frac{e^{-\lambda x}}{\lambda}$ ,  $v = x^2$ ,  $v' = 2x$ , et on obtient

$$\begin{aligned} \lambda \int_0^A x^2 e^{-\lambda x} dx &= [-x^2 e^{-\lambda x}]_0^A + \int_0^A 2x e^{-\lambda x} dx \\ &= -\frac{A^2}{\lambda^2} e^{-\lambda A} + \frac{2}{\lambda} \int_0^A \lambda x e^{-\lambda x} dx. \end{aligned}$$

Par croissance comparée, le premier terme tend vers 0. Quant au second, il tend vers  $\frac{2}{\lambda} \mathbb{E}(X) = \frac{2}{\lambda^2}$ . On a donc

$$\text{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2},$$

et  $\sigma(X) = \frac{1}{\lambda}$ .

**Exemple 6.12.** Soit  $X \sim \mathcal{N}(\mu, \sigma^2)$ . Par croissance comparée,  $x^4 \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$  tend vers zéro quand  $|x| \rightarrow +\infty$ , donc  $X$  est bien de carré intégrable. Calculons maintenant  $\mathbb{E}(X^2)$ . On a

$$\begin{aligned} \int_{-A}^A \frac{x^2}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx &= \int_{-A-\mu}^{A-\mu} \frac{(y+\mu)^2}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{y^2}{2\sigma^2}\right) dy \\ &= \int_{-A-\mu}^{A-\mu} \frac{y^2}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{y^2}{2\sigma^2}\right) dy + \int_{-A-\mu}^{A-\mu} \frac{2y\mu}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{y^2}{2\sigma^2}\right) dy \\ &\quad + \int_{-A-\mu}^{A-\mu} \frac{\mu^2}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{y^2}{2\sigma^2}\right) dy \end{aligned}$$

La seconde intégrale tend vers  $\frac{2\mu}{\sqrt{2\pi\sigma^2}} \int_{\mathbb{R}} ye^{-\frac{y^2}{2\sigma^2}} dy$  qui vaut zéro, car la fonction intégrée est impaire.

La troisième intégrale tend vers  $\mu^2 \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{y^2}{2\sigma^2}\right) dy = \mu^2$ , par le lemme 6.1.

Pour calculer la première intégrale, on effectue une intégration par parties, en posant  $u' = y \exp\left(-\frac{y^2}{2\sigma^2}\right)$ ,  $u = -\sigma^2 \exp\left(-\frac{y^2}{2\sigma^2}\right)$ ,  $v = y$ ,  $v' = 1$ , et on obtient

$$\int_{-A-\mu}^{A-\mu} \frac{y^2}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{y^2}{2\sigma^2}\right) dy = -\frac{\sigma^2}{\sqrt{2\pi\sigma^2}} \left[ y \exp\left(-\frac{y^2}{2\sigma^2}\right) \right]_{-A-\mu}^{A-\mu} + \sigma^2 \int_{-A-\mu}^{A-\mu} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{y^2}{2\sigma^2}\right) dy.$$

La quantité entre crochets tend vers zéro quand  $A$  tend vers  $+\infty$ , par croissance comparée. Quant au deuxième terme, il tend vers  $\sigma^2$  grâce au lemme 6.1. On a donc

$$\mathbb{E}(X^2) = \sigma^2 + \mu^2.$$

Comme  $\mathbb{E}(X) = \mu$ , on en déduit que

$$\text{Var}(X) = \sigma^2,$$

et donc que  $\sigma$  est l'écart-type de  $X$ .

### 6.3 Le théorème central limite

**Problème concret :** À chaque fois que je rencontre quelqu'un, je lui propose de jouer à pile ou face : si je gagne, il me donne dix euros, et s'il gagne, je lui donne dix euros. Depuis que j'ai commencé à jouer, j'ai rencontré 10000 personnes. Quelle est la probabilité pour que je puisse m'acheter une Rolex ?

La loi des grands nombres ne permet pas de répondre à cette question : elle affirme juste que si je fais la somme de  $n$  variables aléatoires de même loi et indépendantes, la valeur obtenue sera proche de  $n$  fois l'espérance de ces variables.

Ici, la variable aléatoire dont je somme  $n$  copies prend la valeur 10 avec probabilité  $\frac{1}{2}$ , et prend la valeur  $-10$  avec probabilité  $\frac{1}{2}$  : son espérance est donc nulle. La somme  $X_1 + \dots + X_n$  sera donc petite devant  $n$ .

En fait, cette somme sera de l'ordre de  $\sqrt{n}$  (ce qui est beaucoup plus petit que  $n$ ), et ressemblera à une gaussienne. C'est ce qu'affirme le théorème suivant, appelé le Théorème Central Limite (TCL), prouvé par De Moivre (1733) dans le cas de variables de Bernoulli de paramètre  $p = \frac{1}{2}$ , puis par Laplace (1809) dans le cas général. Nous ne démontrerons pas ce théorème dans ce cours. Sa démonstration fait appel à des outils d'analyse que vous verrez l'an prochain (transformée de Fourier, produit de convolution...).

**Théorème 6.1** (Théorème central limite). Soit  $(X_k)_{k \in \mathbb{N}}$  une suite de variables aléatoires discrètes, indépendantes et de même loi. Supposons de plus que  $X_1$  est intégrable et est de carré intégrable. Notons  $\sigma = \text{Var}(X_1)$ , et, pour tout  $n \in \mathbb{N}$ ,  $S_n = \frac{X_1 + \dots + X_n}{n}$ . Alors, pour tout  $a < b \in \mathbb{R} \cup \{-\infty, +\infty\}$ , on a

$$\mathbb{P}\left(a \leq \frac{\sqrt{n}}{\sigma}(S_n - \mathbb{E}(X_1)) \leq b\right) \xrightarrow{n \rightarrow +\infty} \mathbb{P}(a \leq Y \leq b),$$

où  $Y \sim \mathcal{N}(0, 1)$ .

**Remarque 6.6.** — Ce résultat peut se réécrire comme

$$\mathbb{P}\left(\frac{X_1 + \dots + X_n}{n} \in \left[\mathbb{E}(X_1) + a\frac{\sigma}{\sqrt{n}}, \mathbb{E}(X_1) + b\frac{\sigma}{\sqrt{n}}\right]\right) \xrightarrow{n \rightarrow +\infty} \mathbb{P}(a \leq Y \leq b),$$

ou encore comme

$$\mathbb{P}(X_1 + \dots + X_n \in [n\mathbb{E}(X_1) + a\sqrt{n}, n\mathbb{E}(X_1) + b\sqrt{n}]) \xrightarrow{n \rightarrow +\infty} \mathbb{P}(a \leq Z \leq b),$$

où  $Z \sim \mathcal{N}(0, \sigma^2)$ .

- En particulier, si les  $X_n$  suivent tous une loi de Bernoulli de paramètre  $p$ , on a  $\mathbb{E}(X_1) = p$ ,  $\text{Var}(X_1) = p(1-p)$ , et donc

$$\mathbb{P}\left(a \leq \frac{\sqrt{n}}{\sqrt{p(1-p)}}(S_n - p) \leq b\right) \xrightarrow{n \rightarrow +\infty} \mathbb{P}(a \leq Y \leq b).$$

- Il existe une version de ce résultat lorsque les  $X_n$  sont des variables aléatoires à densité. Toutefois, nous n'avons pas vu dans ce cours quelle était la définition de variables à densité indépendantes, donc nous ne donnerons pas cette version du théorème.

**Réponse au problème concret.** Soit  $X_i$  une variable aléatoire de loi

$$\mathbb{P}(X_i = 10) = \mathbb{P}(X_i = -10) = \frac{1}{2},$$

qui correspond à l'argent que je gagne (ou que je perds) en rencontrant la  $i$ -ème personne. On a  $\mathbb{E}(X_i) = 0$ , et  $\text{Var}(X_i) = \mathbb{E}(X_i^2) = \frac{1}{2} \times 10^2 + \frac{1}{2} \times 10^2 = 100$ .

Le théorème central limite nous dit que l'on a

$$\mathbb{P}\left(a \leq \frac{\sqrt{10000}}{\sqrt{100}} S_{10000} \leq b\right) \approx \mathbb{P}(a \leq Y \leq b),$$

et donc que

$$\mathbb{P}\left(\frac{X_1 + \dots + X_{10000}}{1000} \geq a\right) = \mathbb{P}(X_1 + \dots + X_{10000} \geq 1000 \times a) \approx \mathbb{P}(Y \geq a).$$

Rappelons que la Rolex la moins chère coûte 3665 euros : il faut donc prendre  $a = 3,7$ . La probabilité que je puisse l'acheter est donc d'environ

$$\mathbb{P}(Y \geq 3,7) \approx 1,08 \times 10^{-4}.$$

## Chapitre 7

# Fluctuation d'échantillonnage et intervalle de confiance

**Problème concret :** La veille d'une élection présidentielle, un institut de sondage interroge 1000 personnes. 600 sondés ont prévu de voter pour Alice, et 400 pour Bernard. L'institut risque-t-il de se tromper en annonçant la victoire d'Alice? Et si seulement 53% des sondés disent avoir l'intention de voter pour Alice?

Quand on effectue un sondage, on ne mesure pas une variable statistique sur toute la population, mais seulement sur une (petite) partie, qu'on appelle un **échantillon**.

Évidemment, un échantillon peut ne pas du tout être représentatif : il est possible d'interroger 1000 personnes qui vont voter pour Bernard, alors qu'en fait, la majorité du pays va voter pour Alice. Néanmoins, si l'échantillon est choisi au hasard, il est très peu probable qu'il soit très peu représentatif.

### Modélisation probabiliste d'un sondage

On modélise le problème de la manière suivante. Pour  $1 \leq i \leq n$ ,  $X_i$  désigne le vote de la  $i$ -ème personne interrogée, qui vaut 1 si la personne souhaite voter pour Alice, et 0 si elle souhaite voter pour Bernard. Ici,  $n = 1000$ .

Les  $X_i$  sont tous indépendants (si le sondage est bien fait<sup>1</sup>), et suivent une loi de Bernoulli de paramètre  $p$ , où  $p$  est la proportion de la population ayant l'intention de voter pour Alice.

Le paramètre  $p$  est inconnu, et il nous faut chercher à le déterminer le plus précisément possible. Pour cela, on peut définir un **estimateur statistique** :

$$p_{est} := \frac{1}{n} \sum_{i=1}^{1000} X_i.$$

**Remarque 7.1.** On peut définir des estimateurs statistiques pour toutes les quantités qui nous intéressent sur une variable statistique : par exemple, on peut définir un estimateur de la variance comme la variance de l'échantillon observé.

Un estimateur statistique n'est pas toujours une information fiable. Par exemple, si un sondage annonce « 40% de la population interrogée va voter pour Alice », (autrement dit,  $p_{est} = 0,4$ ), il s'agit d'une information très peu fiable si seulement 10 personnes ont été interrogées, et très fiable si un million de personnes ont été sondées.

Une information beaucoup plus fiable est la donnée d'un **intervalle de confiance**. On se donne un  $\varepsilon > 0$ , et on estime la probabilité que  $p$  appartienne à  $[p_{est} - \varepsilon, p_{est} + \varepsilon]$ .

Plus  $\varepsilon$  est petit, plus cette probabilité est petite. En pratique, on cherche souvent  $\varepsilon$  suffisamment grand pour que

$$\mathbb{P}(p \in [p_{est} - \varepsilon, p_{est} + \varepsilon]) > 0,95.$$

1. Le problème d'**échantillonnage**, qui consiste à choisir un échantillon de la population « au hasard » est très complexe, et n'est pas purement mathématique (il comporte, par exemple, de nombreux aspects sociologiques). Vous pouvez lire l'article suivant pour en savoir plus : <http://images.math.cnrs.fr/Les-sondages-sont-ils-devenus-fous>



## Détermination d'un intervalle de confiance grâce à la loi des grands nombres

Dans notre problème, on a  $p_{est} = \frac{1}{n} \sum_{i=1}^n X_i$ , et  $p = \mathbb{E}(X_1)$ . Ainsi, par la loi faible des grands nombres, on a

$$\mathbb{P}(|p_{est} - p| > \varepsilon) \leq \frac{\text{Var}(X_1)}{\varepsilon^2 n}.$$

On a  $\text{Var}(X_1) = p(1-p)$ , que l'on ne sait pas a priori déterminer. Néanmoins, pour n'importe quel  $p \in [0, 1]$ , cette quantité est inférieure à  $1/4$ , d'où l'on déduit

$$\mathbb{P}(|p_{est} - p| > \varepsilon) \leq \frac{1}{4\varepsilon^2 n}.$$

Pour  $n = 1000$ , si on veut que le membre de droite soit inférieur à  $0,05$ , il faut donc que  $\frac{1}{4000\varepsilon^2} \leq \frac{1}{20}$ , soit  $\varepsilon \geq \frac{1}{\sqrt{200}} \approx 0,07$ . Ainsi, si 60% des 1000 sondés disent qu'ils vont voter pour Alice, l'institut de sondage peut annoncer qu'avec une probabilité de 95%, Alice obtiendra un score compris entre 0,53 et 0,67. Sa victoire est donc très probable!

## Détermination d'un intervalle de confiance grâce au théorème central limite

Par le théorème central limite, on a

$$\mathbb{P}\left(a \leq \frac{\sqrt{n}}{\sigma}(S_n - \mathbb{E}(X_1)) \leq b\right) \xrightarrow{n \rightarrow +\infty} \mathbb{P}(a \leq Y \leq b),$$

où  $Y \sim \mathcal{N}(0, 1)$ , ce qui peut se réécrire comme

$$\mathbb{P}\left(\frac{a\sigma}{\sqrt{n}} \leq p_{est} - p \leq \frac{b\sigma}{\sqrt{n}}\right) \xrightarrow{n \rightarrow +\infty} \mathbb{P}(a \leq Y \leq b).$$

Ainsi, en prenant  $a = -\frac{c}{\sigma}$ ,  $b = \frac{c}{\sigma}$ ,

$$\mathbb{P}\left(p \in \left[p_{est} - \frac{c}{\sqrt{n}}, p_{est} + \frac{c}{\sqrt{n}}\right]\right) \xrightarrow{n \rightarrow +\infty} \mathbb{P}\left(-\frac{c}{\sigma} \leq Y \leq \frac{c}{\sigma}\right).$$

Là encore,  $\sigma$  n'est pas connu, mais il vaut au plus  $\frac{1}{2}$ . Ainsi,

$$\mathbb{P}\left(p \in \left[p_{est} - \frac{c}{\sqrt{n}}, p_{est} + \frac{c}{\sqrt{n}}\right]\right) \xrightarrow{n \rightarrow +\infty} \mathbb{P}(-2c \leq Y \leq 2c).$$

Le membre de droite est supérieur à  $0,95$  si  $2c \geq 1,96$  (cette valeur est obtenue à l'aide d'un calcul numérique), soit  $c \geq 0,98$ .

Ainsi, pour  $n$  assez grand, on aura

$$\mathbb{P}\left(p \in \left[p_{est} - \frac{0,98}{\sqrt{n}}, p_{est} + \frac{0,98}{\sqrt{n}}\right]\right) \geq 0,95.$$

Ici, pour  $n = 1000$ , on obtient que

$$\mathbb{P}(p \in [p_{est} - 0,03, p_{est} + 0,03]) \geq 0,95.$$

Ainsi, si 600 des 1000 sondés veulent voter pour Alice, il y a une probabilité de 0,95 qu'Alice obtienne un score compris entre 57% et 63 %.

## Comparaison des deux méthodes et remarques

- Pour utiliser le théorème central limite, il faut que  $n$  soit assez grand. En pratique, quand on considère  $\sum_{i=1}^n X_i$  avec  $X_i \sim \mathcal{B}(p)$ , il suffit que  $n \geq 30$ ,  $np \geq 5$  et  $n(1-p) \geq 5$ , comme c'est bien le cas ici.
- Le théorème central limite donne toujours un résultat plus précis.
- La loi des grands nombres peut aussi être utilisée quand  $n$  est petit, ou quand  $p$  ou  $(1-p)$  est proche de 0, contrairement au TCL.

- Avec ces deux méthodes, la précision augmente quand  $n$  augmente. Par contre, le résultat ne dépend pas de la population du pays : un sondage interrogeant 1000 personnes au hasard aura la même précision à Monaco et en Inde !

**En physique : la méthode de Monte Carlo** Les approches probabilistes (ou statistiques) sont souvent utilisées en physique pour calculer la valeur de certaines intégrales.

Par exemple, supposons que nous voulions calculer l'aire d'un cercle de rayon 1, qui vaut donc  $\pi$ , avec une certaine précision. Une première méthode consiste à calculer  $\pi$  d'une autre manière, par exemple, à l'aide d'une série convergente. Une autre manière de faire est d'encadrer le cercle par des polygones dont on sait calculer l'aire.

Une troisième manière est de choisir des suites de points  $(X_n, Y_n)$  au hasard, uniformément dans le carré  $[-1, 1]^2$ . On considère la variable aléatoire  $Z_n = \begin{cases} 1 & \text{si } X_n^2 + Y_n^2 \leq 1 \\ 0 & \text{sinon.} \end{cases}$  Autrement dit,

$Z_n$  vaut 1 si le point  $(X_n, Y_n)$  est dans le disque unité, et zéro sinon.  $Z_n$  est donc une variable de Bernoulli, de paramètre  $p$  valant l'aire du disque divisée par l'aire du carré, soit  $\frac{\pi}{4}$ . En tirant au hasard un grand nombre de points, on peut calculer  $p_{est}$ , et affirmer des choses comme « La probabilité que  $\pi$  soit compris entre 3,1415 et 3,1416 est supérieure à 0,99 ». Si de telles méthodes ont l'inconvénient de ne pas être complètement sûres, elles nécessitent souvent des temps de calcul bien plus courts que les méthodes déterministes. Ces méthodes probabilistes pour calculer la valeur d'une aire (et donc, d'une intégrale) sont appelées **méthodes de Monte Carlo**.