

# PROBABILITÉS ET STATISTIQUE POUR LE CAPES

Béatrice de Tilière<sup>1</sup>  
Frédérique Petit<sup>2</sup>

31 mai 2016

1. Université Pierre et Marie Curie
2. Université Pierre et Marie Curie



# Table des matières

<b>1</b>	<b>Modélisation de phénomènes aléatoires</b>	<b>7</b>
1.1	Introduction . . . . .	7
1.2	L'espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$ . . . . .	8
1.2.1	Espace des états . . . . .	8
1.2.2	Évènements . . . . .	8
1.2.3	Tribu . . . . .	10
1.2.4	Probabilité . . . . .	10
<b>2</b>	<b>Construction d'espaces probabilisés</b>	<b>15</b>
2.1	Espace des états fini . . . . .	15
2.1.1	Espace probabilisé . . . . .	15
2.1.2	Dénombrement, modèle d'urne . . . . .	17
2.2	Espace des états infini dénombrable . . . . .	22
2.3	Espace des états infini non-dénombrable . . . . .	23
<b>3</b>	<b>Conditionnement et indépendance</b>	<b>25</b>
3.1	Probabilité conditionnelle . . . . .	25
3.1.1	Définition . . . . .	25
3.1.2	Formule des probabilités totales et formule de Bayes . . . . .	27
3.1.3	Arbre de probabilité . . . . .	30
3.2	Indépendance des évènements . . . . .	32
3.2.1	Nombre infini de jets de dés . . . . .	34
<b>4</b>	<b>Variables aléatoires</b>	<b>37</b>
4.1	Définition et loi d'une variable aléatoire . . . . .	37
4.2	Fonction de répartition . . . . .	39
4.3	Variables aléatoires discrètes . . . . .	40
4.3.1	Définitions et exemples à connaître . . . . .	40
4.3.2	Fonction de répartition . . . . .	43
4.3.3	Espérance . . . . .	44

4.3.4	Variance, moments d'ordres supérieurs . . . . .	47
4.3.5	Inégalité de Markov et de Bienaymé Tchebychev . . . . .	48
4.4	Vecteurs aléatoires discrets . . . . .	49
4.4.1	Définition et lois des vecteurs aléatoires . . . . .	49
4.4.2	Espérance, covariance, matrice de covariance . . . . .	51
4.4.3	Variables aléatoires indépendantes . . . . .	54
4.5	Variables aléatoires réelles à densité . . . . .	56
4.5.1	Définitions . . . . .	56
4.5.2	Espérance et moments d'ordres supérieurs . . . . .	57
4.5.3	Exemples de variables aléatoires à densité . . . . .	58
4.6	Couples de variables aléatoires à densité . . . . .	61
4.6.1	Définitions . . . . .	61
4.6.2	Variables aléatoires à densité indépendantes . . . . .	62
4.7	Suites de variables aléatoires réelles . . . . .	66
4.7.1	Loi faible des grands nombres . . . . .	66
4.7.2	Théorème limite central . . . . .	67
4.8	Encore des définitions . . . . .	69
<b>5</b>	<b>Statistique descriptive</b>	<b>71</b>
5.1	Préambule à la statistique . . . . .	71
5.1.1	Un peu de vocabulaire . . . . .	71
5.1.2	Collecte de données . . . . .	71
5.1.3	Deux directions en statistique . . . . .	72
5.1.4	Statistique univariée / multivariée . . . . .	72
5.2	Paramètres de position, dispersion, relation . . . . .	72
5.3	Interprétation des données . . . . .	75
5.3.1	Interprétation probabiliste . . . . .	75
5.3.2	Interprétation vectorielle . . . . .	75
5.4	Méthode des moindres carrés . . . . .	76
<b>6</b>	<b>Statistique inférentielle</b>	<b>87</b>
6.1	Un exemple introductif . . . . .	87
6.2	Modèle statistique paramétrique . . . . .	89
6.3	Estimateur . . . . .	90
6.4	Rappels sur quelques lois utiles en statistique . . . . .	93
6.5	Intervalles de confiance . . . . .	94
6.5.1	Intervalle de confiance à partir de l'inégalité de Bienaymé-Tchebychev . . . . .	95
6.5.2	Construction d'un intervalle de confiance pour la moyenne . . . . .	96

6.6	Tests . . . . .	101
6.6.1	Exemple introductif . . . . .	101
6.6.2	Définitions . . . . .	101
6.6.3	Cas de deux hypothèses simples $\Theta = \{\theta_0, \theta_1\}$ . . . . .	102
6.6.4	Test pour la moyenne . . . . .	103
6.6.5	Comparaison de deux moyennes . . . . .	107
6.6.6	Test d'adéquation à une loi . . . . .	109
6.6.7	À propos de la droite de Henry . . . . .	112
6.6.8	Table de lois du $\chi^2$ . . . . .	114



# Chapitre 1

## Modélisation de phénomènes aléatoires

### 1.1 Introduction

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

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

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

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

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

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

## 1.2 L'espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$

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

### 1.2.1 Espace des états

**Définition.** L'*espace des états* appelé aussi *univers*, noté  $\Omega$ , est l'ensemble des résultats possibles de l'expérience.

EXERCICE 1.1. Déterminer un espace des états possible dans les expériences suivantes.

1. Lancer d'une pièce de monnaie.
2. Deux lancers successifs d'une même pièce de monnaie.
3. Lancer d'un dé.
4. Deux lancers successifs d'un même dé, et on s'intéresse à la somme des nombres obtenus.
5. Lancer d'un même dé indéfiniment.
6. Durée de vie d'un individu.
7. Promenade d'un ivrogne dans une rue (un pas en avant, un pas en arrière).
8. Trajectoire d'une poussière à la surface de l'eau pendant un intervalle de temps  $[0, T]$ .

### 1.2.2 Évènements

**Définition heuristique.** Un *événement* est une propriété dont on peut dire si elle est réalisée ou non, une fois l'issue de l'expérience connue. À chaque événement correspond un sous-ensemble de l'espace des états  $\Omega$ . Un singleton, c'est-à-dire un événement réduit à un seul élément de  $\Omega$ , est appelé un *événement élémentaire*, sinon on parle d'*événement composite*. On note un événement par une lettre majuscule  $A, B, C, \dots$  et l'ensemble de tous les événements de  $\Omega$  par  $\mathcal{A}$ .

*Remarque.* Nous verrons au paragraphe suivant la définition (mathématique) d'un événement. Pour l'instant, essayons de voir à quelles propriétés doivent satisfaire les événements.

EXERCICE 1.2. *Cet exercice est la suite de l'Exercice 1.1. On reprend la numérotation déjà utilisée.* Décrire les événements suivants comme des sous-ensembles de l'espace des états  $\Omega$ .

2. "Le premier jet donne pile".
4. "La somme des résultats obtenus est égale à 4".
5. "Le premier 1 est obtenu au  $N$ -ième lancer".
6. "L'individu atteint au moins 50 ans".



7. “L’ivrogne avance au  $N$ -ième pas”.

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

Une fois fixé le choix de l’univers, un évènement correspond à un *unique* sous-ensemble de ce dernier. Comme il n’y a pas forcément unicité du modèle et qu’alors les évènements peuvent s’écrire en termes de sous-ensembles sous des formes différentes, la phrase qui décrit un évènement permet de se comprendre, quel que soit le modèle choisi, voir par exemple les Exercices 1.1 et 1.2 numéro 4. Remarquons aussi que, étant donné un sous-ensemble d’un univers, il est souvent possible de le décrire par différentes phrases, qui représentent toutes le même évènement. Par exemple l’évènement  $\{PP, PF\}$  de l’Exercice 1.2 numéro 2 peut se traduire par “le premier jet donne pile” ou “le premier jet ne donne pas face”.

A noter que le passage de la phrase au sous-ensemble et réciproquement est rarement expliqué dans les manuels scolaires et les élèves se voient devoir travailler sur des évènements en termes de phrases alors qu’on vient de leur expliquer qu’un évènement est une partie de l’univers, pas facile de s’y retrouver pour eux...

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

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

EXEMPLE. Deux lancers successifs d’une même pièce de monnaie. On considère les évènements  $A = \{PP\}$ ,  $B = \{PF\}$  et  $C = \{FP, FF\}$ . Alors,

- $A \cup B = \{PP, PF\} = C^c$ , est l’évènement “le premier jet donne pile” ;
- $A \cap B = \emptyset$ , est l’évènement impossible,  $A$  et  $B$  sont incompatibles.

**Propriété 1.** *Les opérations sur les évènements satisfont aux règles suivantes. Pour tous évènements  $A, B, C$ , on a*

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

EXERCICE 1.3. Soient  $A, B, C$  trois évènements liés à une même expérience aléatoire. Donner en fonction de  $A, B, C, A^c, B^c, C^c$ , de leurs réunions, intersections, l’expression des évènements suivants :

- $A$  seulement est réalisé ;
- $A$  et  $B$  seulement sont réalisés ;

- au moins un des trois évènements est réalisé ;
- au moins deux des trois évènements sont réalisés ;
- un et un seul des trois évènements est réalisé ;
- au plus deux des trois évènements sont réalisés ;
- aucun des trois évènements n'est réalisé.

### 1.2.3 Tribu

L'ensemble des évènements  $\mathcal{A}$  associés à une expérience aléatoire est donc un sous-ensemble des parties de  $\Omega$ ,  $\mathcal{A} \subset \mathcal{P}(\Omega)$ . Il semblerait naturel de prendre  $\mathcal{A} = \mathcal{P}(\Omega)$ , mais il y a alors des exemples où il est impossible d'associer à chaque évènement une probabilité de façon cohérente. Dans ces cas-là, il est donc nécessaire de se restreindre à un sous-ensemble strict de  $\mathcal{P}(\Omega)$  contenant les évènements "intéressants".

L'ensemble des évènements que l'on considère en probabilité doivent satisfaire à quelques propriétés naturelles, ils doivent former une tribu, dont voici la définition.

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

1.  $\Omega \in \mathcal{A}$  ;
2.  $\forall A \in \mathcal{P}(\Omega), A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$  ;
3.  $\mathcal{A}$  est stable par réunion finie ou dénombrable.

EXEMPLE 1.1.

- $\{\emptyset, \Omega\}$  est une tribu et c'est la plus petite (au sens de l'inclusion).
- $\mathcal{P}(\Omega)$  est une tribu et c'est la plus grande.
- Soit  $\mathcal{C}$  un ensemble arbitraire de parties de  $\Omega$ , alors la plus petite tribu contenant  $\mathcal{C}$ , notée  $\sigma(\mathcal{C})$  est appelée la *tribu engendrée par  $\mathcal{C}$* . On admet l'existence de cette tribu.
- Soit  $A \in \mathcal{P}(\Omega)$ ,  $A \neq \emptyset$ ,  $A \neq \Omega$ , alors la tribu engendrée par  $A$  est  $\sigma(A) = \{\emptyset, A, A^c, \Omega\}$ .
- Sur  $\mathbb{R}$ , on utilise la tribu engendrée par les ouverts de  $\mathbb{R}$ , appelée *tribu borélienne* de  $\mathbb{R}$ . On admet le fait qu'elle soit différente de  $\mathcal{P}(\mathbb{R})$ .
- Dans le cas où l'espace des états est fini ou dénombrable, on prend toujours  $\mathcal{A} = \mathcal{P}(\Omega)$ .

EXERCICE 1.4. Soit  $\Omega = \{1, 2, 3\}$ .

- Quelle est la tribu engendrée par  $A = \{1, 2\}$  ?
- Quelle est la tribu engendrée par  $\mathcal{C} = \{\{1\}, \{2\}\}$  ?

**Définition.** L'ensemble des évènements associé à une expérience est la tribu  $\mathcal{A}$  choisie sur  $\Omega$ .

*Remarque.* Dans le cas où l'espace des états est fini ou dénombrable, puisque  $\mathcal{A} = \mathcal{P}(\Omega)$ , un évènement est donc simplement n'importe quel sous-ensemble de  $\Omega$ . On retrouve bien la définition donnée dans l'enseignement secondaire.

### 1.2.4 Probabilité

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

*Approche utilisant les symétries.* On considère un dé non-pipé. Il est alors naturel de supposer que chacune des issues possibles ait la même probabilité égale à  $1/6$ . Il faut cependant être prudent avec cette approche. En effet, supposons que nous souhaitions déterminer la probabilité du sexe d'un nouveau né. Il n'y a aucune raison de penser qu'il y a plus de chances d'avoir un garçon ou une fille, de sorte qu'il est naturel d'associer une probabilité  $1/2$  à chacun des événements élémentaires. Cependant, les statistiques montrent que la proportion de garçons nouvellement nés en France varie entre  $51,1\%$  et  $51,2\%$  (INED, France métropolitaine, 2003-2013).

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

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

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

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

1. L'événement certain est de probabilité 1 :  $\mathbb{P}(\Omega) = 1$ .
2. *Axiome de  $\sigma$ -additivité* : pour toute famille dénombrable  $(A_n)_{n \geq 0}$  d'événements de  $\mathcal{A}$ , deux-à-deux disjoints, on a

$$\mathbb{P}\left(\bigcup_{n \geq 0} A_n\right) = \sum_{n=0}^{+\infty} \mathbb{P}(A_n).$$

Le triplet  $(\Omega, \mathcal{A}, \mathbb{P})$  est alors appelé un *espace probabilisé*.

On a les conséquences immédiates suivantes.

**Proposition 1.1.** Soit  $(\Omega, \mathcal{A}, \mathbb{P})$  un espace probabilisé et soient deux événements  $A \in \mathcal{A}$ ,  $B \in \mathcal{A}$ .

1. Additivité. Si  $A$  et  $B$  sont disjoints, alors  $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$ . En particulier,  $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$ , et  $\mathbb{P}(\emptyset) = 0$ .
2. Si  $A \subset B$ , alors :  $\mathbb{P}(A) \leq \mathbb{P}(B)$ .
3.  $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$ . Plus généralement, on a la formule de Poincaré : Soit  $(A_n)_{n=1}^N$  une famille d'événements de  $\mathcal{A}$ , alors :

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

EXERCICE 1.5. Démontrer la Proposition 1.1.

Voici une conséquence plus abstraite, qui est fréquemment utilisée.

**Proposition 1.2.** Soit  $(\Omega, \mathcal{A}, \mathbb{P})$  un espace probabilisé.

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

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

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

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

*Remarque 1.1.* Les suites réelles  $(\mathbb{P}(A_n))_{n \geq 1}$  et  $(\mathbb{P}(B_n))_{n \geq 1}$  sont respectivement croissante et décroissante, bornées, donc convergentes, justifiant ainsi l'existence de la limite. Par ailleurs, la suite  $(\bigcup_{n=1}^N A_n)_{N \geq 1}$ , égale à  $(A_N)_{N \geq 1}$ , est une suite croissante au sens de l'inclusion, et l'ensemble  $\bigcup_{n=1}^{+\infty} A_n$  est habituellement noté  $\lim_{n \rightarrow +\infty} A_n$ , de sorte que le résultat de la proposition précédente peut encore s'écrire :

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

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

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

*Démonstration.* Démontrons la proposition dans le cas d'une suite croissante d'évènements. Posons  $C_1 = A_1$  et pour tout  $n \geq 2$ ,  $C_n = A_n \cap A_{n-1}^c$ . Ainsi, pour tout  $n \geq 2$ ,  $A_n = C_n \cup A_{n-1}$ , où la réunion est disjointe, de sorte que :

$$\mathbb{P}(A_n) = \mathbb{P}(C_n) + \mathbb{P}(A_{n-1}).$$

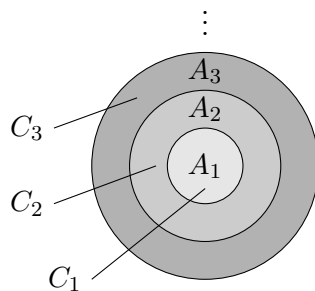


FIGURE 1.1 – Une vision schématique d'une suite croissante d'évènements  $(A_n)_{n \geq 1}$ . Les évènements  $(C_n)_{n \geq 1}$  sont les anneaux successifs.

La famille  $(C_n)_{n \geq 1}$  est une famille dénombrable d'évènements disjoints deux à deux (voir Figure 1.1). Ainsi, d'après la propriété de  $\sigma$ -additivité, on a :

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

Puisque de plus,  $\bigcup_{n \geq 1} C_n = A$ , on déduit que,

$$\begin{aligned} \mathbb{P}(A) &= \lim_{N \rightarrow +\infty} \sum_{n=1}^N \mathbb{P}(C_n) = \lim_{N \rightarrow +\infty} \left( \mathbb{P}(C_1) + \sum_{n=2}^N [\mathbb{P}(A_n) - \mathbb{P}(A_{n-1})] \right) \\ &= \lim_{N \rightarrow +\infty} \left( \mathbb{P}(A_1) + \sum_{n=2}^N [\mathbb{P}(A_n) - \mathbb{P}(A_{n-1})] \right) = \lim_{N \rightarrow +\infty} \mathbb{P}(A_N). \end{aligned}$$

□

*Remarque 1.2.* On rappelle :

$$\omega \in \bigcup_{n \geq 1} A_n \iff \exists n \in \mathbb{N}, \omega \in A_n,$$

et

$$\omega \in \bigcap_{n \geq 1} A_n \iff \forall n \in \mathbb{N}, \omega \in A_n.$$

EXEMPLE 1.2. Un exemple d'application de la Proposition 1.2 est donné dans l'Exercice 3.9.



## Chapitre 2

# Construction d'espaces probabilisés

### 2.1 Espace des états fini

#### 2.1.1 Espace probabilisé

On suppose dans la suite que l'espace des états  $\Omega$  est de cardinal fini. Dans ce cas, la tribu considérée est simplement  $\mathcal{A} = \mathcal{P}(\Omega)$  et la définition d'une probabilité prend la forme suivante, qui est celle utilisée au lycée.

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

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

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

**EXERCICE 2.1.** Montrer que cette définition est équivalente à celle donnée à la Section 1.2.4 dans le cas général. On remarquera que, lorsque l'espace des états est de cardinal fini, toute famille dénombrable d'évènements disjoints deux-à-deux contient un nombre fini d'évènements non vides.

*Remarque.* Reprenons l'approche fréquentiste vue à la Section 1.2.4. Les fréquences  $\frac{n_A(N)}{N}$  de réalisation d'un évènement  $A$  sur les  $N$  premières répétitions d'une expérience vérifient les propriétés suivantes :

- i) pour tout évènement  $A$ ,  $0 \leq \frac{n_A(N)}{N} \leq 1$ ;
- ii)  $\frac{n_\Omega(N)}{N} = 1$ ;
- iii) si  $A$  et  $B$  sont des évènements incompatibles,  $\frac{n_{A \cup B}(N)}{N} = \frac{n_A(N)}{N} + \frac{n_B(N)}{N}$ .

La définition d'une probabilité donnée ci-dessus découle naturellement de ces trois propriétés.

**Proposition 2.1.** Une probabilité sur  $(\Omega, \mathcal{P}(\Omega))$  est caractérisée par sa valeur sur les singletons  $\{\omega\}$ , pour tout  $\omega \in \Omega$ . Réciproquement, à toute famille  $(p_\omega)_{\omega \in \Omega}$  telle que :

1. pour tout  $\omega \in \Omega$ ,  $0 \leq p_\omega \leq 1$ ,
2.  $\sum_{\omega \in \Omega} p_\omega = 1$ ,

on peut associer une unique probabilité  $\mathbb{P}$  sur  $(\Omega, \mathcal{P}(\Omega))$  définie par :  $\mathbb{P}(\{\omega\}) = p_\omega$ . On étend ensuite  $\mathbb{P}$  à  $\mathcal{P}(\Omega)$  par additivité : pour tout  $A \in \mathcal{P}(\Omega)$ ,

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

EXERCICE 2.2. Démontrer la Proposition 2.1.

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

**Corollaire 2.1.** Dans ce cas, pour tout  $\omega \in \Omega$ ,  $\mathbb{P}(\{\omega\}) = \frac{1}{\text{Card}(\Omega)}$ , et, pour tout évènement  $A \in \mathcal{P}(\Omega)$ , on a :

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

*Remarque.* Pour signaler que l'on est en situation d'équiprobabilité, les manuels scolaires ont coutume d'écrire que le phénomène se produit "au hasard" pour permettre à l'élève de comprendre qu'il doit choisir la probabilité uniforme. Cela n'a pourtant aucun sens, puisque n'importe quel modèle probabiliste modélise un phénomène aléatoire... Lorsque l'on veut préciser que l'on est en situation d'équiprobabilité, on mettra donc l'expression "au hasard" entre guillemets, montrant ainsi que l'on est conscient que cette expression ne veut rien dire. Cette remarque sera approfondie en didactique, notamment avec le "paradoxe de Bertrand".

EXEMPLE. Reprenons la question 4 des Exercices 1.1 et 1.2 du Chapitre 1 et calculons la probabilité de l'évènement  $A$  "la somme des dés est égale à 4".

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

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

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

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

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

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

Remarquer que dans le cas de l'univers  $\Omega_3$ , on n'est également pas en situation d'équiprobabilité.

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

Lorsque l'espace des états est fini, les calculs de probabilités se ramènent essentiellement à des problèmes de dénombrement, sujet de la section suivante.



### 2.1.2 Dénombrement, modèle d'urne

Dans cette partie, on va considérer une *population* de taille  $N$ , assimilée à l'ensemble  $\mathcal{S}_N = \{1, \dots, N\}$ .

#### TIRAGES ORDONNÉS

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

- *Le tirage avec remise*. Dans ce cas, chaque élément de l'ensemble peut être choisi à plusieurs reprises. On parle alors d'*échantillon de taille  $r$  avec répétitions*. Soit  $\Omega_1$  l'ensemble de ces échantillons, alors  $\Omega_1 = \{1, \dots, N\}^r$ , et :

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

EXEMPLE. Soit  $\mathcal{S}_4 = \{1, 2, 3, 4\}$  et  $r = 2$ . Alors  $\Omega_1$  peut être représenté par la matrice  $M$  ci-dessous, et  $\text{Card}(\Omega_1) = 16$ .

$$M = \begin{pmatrix} (1, 1) & (1, 2) & (1, 3) & (1, 4) \\ \vdots & & & \vdots \\ (4, 1) & (4, 2) & (4, 3) & (4, 4) \end{pmatrix}$$

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

$$\text{Card}(\Omega_2) = N(N-1) \cdots (N-r+1) = \frac{N!}{(N-r)!}$$

Ce nombre a deux notations usuelles :  $A_N^r$  ou  $(N)_r$  (*symbole de Pochhammer*, qui n'est pas au programme du CAPES).

EXEMPLE. Soit  $\mathcal{S}_4 = \{1, 2, 3, 4\}$  et  $r = 2$ . Alors  $\Omega_2$  peut être représenté par la matrice  $M$  privée de sa diagonale et  $\text{Card}(\Omega_2) = 12$ .

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

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

Donnons quelques applications de ce résultat.

1. On jette un dé six fois de suite. Alors la probabilité d'obtenir six nombres distincts est  $\frac{6!}{6^6} \sim 0,015$ .
2. Supposons que dans une ville, il y ait sept accidents par semaine. Alors la probabilité d'avoir exactement un accident chaque jour de la semaine est  $\frac{7!}{7^7} \sim 0,00612$ .

#### TIRAGES NON ORDONNÉS

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

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

$$\text{Card}(\Omega_3) = \frac{N!}{(N-r)!r!}.$$

Ce nombre, appelé *coefficient binomial*, a deux notations usuelles :  $C_N^r$  (notation qui n'est plus en vigueur dans les lycées) ou  $\binom{N}{r}$  (notation anglo-saxonne).

EXEMPLE. Soit  $\mathcal{S}_4 = \{1, 2, 3, 4\}$  et  $r = 2$ . Alors  $\Omega_3$  peut être représenté par le triangle supérieur de la matrice  $M$ , privé de la diagonale et  $\text{Card}(\Omega_3) = 6$ .

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

EXEMPLE 2.1.

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

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

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

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

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

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

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

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

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

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

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

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

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

Ceci est un exemple de la *loi hypergéométrique*.

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

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

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

- *Partitionnement*

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

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

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

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

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

□

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

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

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

EXEMPLE. Soit  $\mathcal{S} = \{1, 2, 3, 4\}$  et  $r = 2$ . Alors  $\Omega_4$  peut être représenté par le triangle supérieur de la matrice  $M$  et  $\text{Card}(\Omega_4) = 10$ .

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

$$|*****|$$

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

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

représente le tirage : 1, 1, 3, 3, 3, 3, 3. Ainsi, ce problème revient à placer  $N - 1$  cloisons sur  $N + r - 1$  positions, les positions restantes étant occupées par des  $*$ . □

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

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

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

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

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

## EXERCICES

EXERCICE 2.3. Supposons que 23 personnes sont dans une même salle. Quelle est la probabilité qu'au moins deux d'entre elles aient l'anniversaire le même jour ? (On ne considèrera pas les années bissextiles.)

EXERCICE 2.4. Dans une course,  $n$  chevaux sont au départ. On suppose qu'ils ont tous la même chance de gagner. Calculer la probabilité de gagner le tiercé avec un ticket :

1. dans l'ordre,
2. dans l'ordre ou dans un ordre différent,

3. dans un ordre différent ?

EXERCICE 2.5. Un joueur de poker reçoit une main de 5 cartes d'un jeu de 32 cartes. Quelle est la probabilité qu'il reçoive :

1. une seule paire (deux cartes de même hauteur) ;
2. deux paires ;
3. un brelan (trois cartes de même hauteur et pas de paire ni de carré) ;
4. un carré (quatre cartes de même hauteur) ;
5. un full (une paire et un brelan) ?

EXERCICE 2.6. (D'après C. Bouzitat et G. Pagès, En passant par hasard... Chapitre XI. Ed. Vuibert (1999)). Au loto, le joueur doit cocher 6 numéros dans une grille en comportant 49. Un tirage consiste à extraire, sans remise, 6 boules numérotées d'une urne, dont les numéros sont dits gagnants, et une 7-ième boule fournissant le numéro dit complémentaire. Est gagnant du premier rang, toute grille sur laquelle sont cochés les 6 numéros gagnants. Est gagnante du 2-ième rang, toute grille sur laquelle sont cochés 5 des 6 numéros gagnants et dont le 6-ième numéro est le numéro complémentaire. Est gagnant du 3-ième rang, toute grille sur laquelle sont exactement cochés 5 des 6 numéros gagnants.

Considérons une grille validée et notons

$$p_k = \mathbb{P}(\text{la grille est gagnante au } k\text{-ième rang}).$$

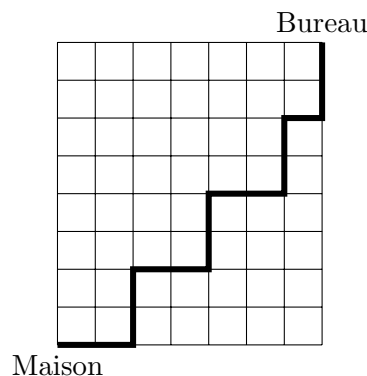
Calculer  $p_k$  pour  $k \in \{1, 2, 3\}$ .

EXERCICE 2.7. On considère la distribution aléatoire de  $r$  boules dans  $n$  urnes. Quelle est la probabilité qu'une urne donnée contienne exactement  $k$  boules ? ( $k \leq r$ )

EXERCICE 2.8. Soit  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  une fonction de  $n$  variables, que l'on suppose de classe  $\mathcal{C}^\infty$ . Quel est le nombre de dérivées partielles distinctes d'ordre  $r$  ?

EXERCICE 2.9. Combien l'équation  $x_1 + x_2 + x_3 = 15$  a-t-elle de solutions entières et non négatives ?

EXERCICE 2.10. (CAPES externe, dossier du 5 juillet 2006). Un homme travaille à Manhattan, dans un quartier où les avenues sont orientées nord-sud et les rues est-ouest. Il travaille à 7 pâtés de maison à l'est et 8 pâtés de maisons au nord de son domicile. Pour aller à son travail chaque jour il parcourt donc la longueur de 15 pâtés de maison (il ne se dirige ni vers le sud, ni vers l'ouest). On suppose qu'il existe une voie le long de chaque pâté de maisons et qu'il peut prendre n'importe lesquelles de ce schéma rectangulaire. La figure ci-dessous illustre la situation ; un exemple de trajet est représenté en ligne grasse.



1. Proposer un codage permettant de décrire le trajet représenté.
2. Combien de trajets différents l'homme peut-il emprunter ?
3. L'homme prétend que le nombre de trajets est aussi le nombre de suites de 8 entiers naturels dont la somme est 8. A-t-il raison ?

Pendant sa préparation, le candidat traitera la question suivante :

1. A quel niveau pensez-vous pouvoir proposer cet exercice ? Quelles indications souhaiteriez-vous ajouter ?
2. La question 3 de l'exercice.

## 2.2 Espace des états infini dénombrable

Lorsque  $\Omega$  est infini dénombrable, on procède de la même manière que dans le cas fini : on prend  $\mathcal{A} = \mathcal{P}(\Omega)$ , et on associe à chaque évènement élémentaire  $\omega \in \Omega$ , sa probabilité,  $\mathbb{P}(\{\omega\}) = p_\omega \in [0, 1]$ , avec :

$$\sum_{\omega \in \Omega} p_\omega = 1.$$

*Remarque 2.1.* La somme ci-dessus est définie de la manière suivante. Comme l'univers  $\Omega$  est dénombrable, il est possible de numéroter les évènements élémentaires, disons  $\Omega = \{\omega_1, \omega_2, \dots\}$ . On pose alors,

$$\sum_{\omega \in \Omega} p_\omega = \sum_{n=1}^{+\infty} p_{\omega_n}.$$

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

On pose ensuite, pour tout  $A \in \mathcal{A}$ ,

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

On vérifie alors de la même façon que dans le cas fini que  $\mathbb{P}$  est bien une probabilité et que toute probabilité sur un univers dénombrable est nécessairement de cette forme.

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

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

Comme  $\mathbb{N}^*$  et  $\{\infty\}$  sont disjoints,  $1 = \mathbb{P}[\Omega] = \mathbb{P}[\{\infty\}] + \mathbb{P}[\bigcup_{k \in \mathbb{N}^*} \{k\}]$ . Ainsi, la probabilité que pile ne sorte jamais est donnée par :

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

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

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

## 2.3 Espace des états infini non-dénombrable

Cette situation est beaucoup plus subtile que dans les cas précédents. On ne peut plus définir une probabilité en définissant celle des singletons (événements élémentaires) car celle-ci est nulle.

La procédure est alors la suivante :

- on détermine une algèbre<sup>1</sup> d'évènements intéressants sur laquelle on définit une probabilité ;
- on utilise un théorème fondamental de la théorie de la mesure, le “théorème d’extension de Carathéodory”, qui affirme qu’une probabilité sur une algèbre s’étend de façon unique en une probabilité sur la tribu engendrée par l’algèbre.

EXEMPLE. Souvenez-vous que lorsque l’espace des états est  $\mathbb{R}$ , on prend comme tribu celle des boréliens (la plus petite tribu engendrée par les ouverts de  $\mathbb{R}$ ). On admettra que cette tribu est engendrée par les intervalles de la forme  $] - \infty, x]$ ,  $x \in \mathbb{R}$ , et qu’une mesure de probabilité  $\mathbb{P}$  sur  $\mathbb{R}$  est entièrement caractérisée par la valeur qu’elle attribue aux intervalles de cette forme.

---

1. Un ensemble  $\mathcal{A}$  de parties de  $\Omega$  est une *algèbre*, si  $\Omega \in \mathcal{A}$ ,  $\mathcal{A}$  est stable par complémentation et par réunions finies. À la différence d’une tribu on ne demande pas la stabilité par réunions dénombrables.





# Chapitre 3

## Conditionnement et indépendance

### 3.1 Probabilité conditionnelle

#### 3.1.1 Définition

Motivons la définition de probabilité conditionnelle sur un exemple.

Soit  $\Omega$  une population partitionnée en  $\Omega = S \cup S^c$ , où  $S$  représente l'ensemble des individus fumeurs. Soit  $F$  l'ensemble des femmes, de sorte que  $F \cup F^c$  représente une autre partition de  $\Omega$ . On suppose que l'on choisit un individu "au hasard", de sorte que l'on munit  $\Omega$  de la probabilité uniforme, notée  $\mathbb{P}$ . Ainsi, la probabilité que l'individu choisi soit fumeur est :

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

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

$$\frac{\text{Card}(S \cap F)}{\text{Card}(F)} = \frac{\text{Card}(S \cap F) / \text{Card}(\Omega)}{\text{Card}(F) / \text{Card}(\Omega)},$$

quantité encore égale, avec les notations précédentes, à

$$\frac{\mathbb{P}(S \cap F)}{\mathbb{P}(F)}.$$

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

$$\mathbb{P}_B(A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}. \quad (3.1)$$

**Lemme 3.1.** *Sous les hypothèses de la définition ci-dessus, l'application  $\mathbb{P}_B$  est une probabilité sur  $(\Omega, \mathcal{A})$ , ainsi  $(\Omega, \mathcal{A}, \mathbb{P}_B)$  est un espace probablisé.*

*Démonstration.* Il faut montrer que  $\mathbb{P}_B$  satisfait aux trois axiomes caractérisant une probabilité.

- Par hypothèse, on a  $\mathbb{P}(B) > 0$ . De plus, pour tout  $A \in \mathcal{A}$ , on a  $A \cap B \subset B$ , d'où l'inégalité  $0 \leq \mathbb{P}(A \cap B) \leq \mathbb{P}(B)$ , et donc :

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

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

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

□

EXERCICE. Refaire la démonstration dans le cas où  $\Omega$  est fini.

*Remarque.*

- On utilise aussi la notation  $\mathbb{P}(A|B)$ . Mais **attention**, cette dernière n'est pas en vigueur dans les lycées. Elle favorise par ailleurs de nombreuses erreurs. La notation  $\mathbb{P}_B(A)$  met en valeur le fait que  $\mathbb{P}_B$  soit une probabilité sur  $(\Omega, \mathcal{A})$ .
- Comme  $(\Omega, \mathcal{A}, \mathbb{P}_B)$  est un espace probabilisé,  $\mathbb{P}_B$  satisfait à toutes les propriétés des Propositions 1.1 et 1.2.
- Si  $\mathbb{P}(A) > 0$  et  $\mathbb{P}(B) > 0$ , on peut réécrire l'équation (3.1) sous la forme :

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

- De manière plus générale, si  $(A_k)_{1 \leq k \leq n+1}$  sont des évènements tels que  $\mathbb{P}\left(\bigcap_{k=1}^n A_k\right) > 0$ , alors :

$$\mathbb{P}\left(\bigcap_{k=1}^{n+1} A_k\right) = \mathbb{P}(A_1) \mathbb{P}_{A_1}(A_2) \mathbb{P}_{A_1 \cap A_2}(A_3) \cdots \mathbb{P}_{A_1 \cap \cdots \cap A_n}(A_{n+1}).$$

EXEMPLE 3.1. On considère deux lancers successifs d'un même dé. Sachant que le premier jet donne 3, on souhaite calculer la probabilité que la somme soit strictement supérieure à 6.

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

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

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

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

### 3.1.2 Formule des probabilités totales et formule de Bayes

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

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

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

**Théorème 3.1.** Soit  $(\Omega, \mathcal{A}, \mathbb{P})$  un espace probabilisé.

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

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

Dans le cas particulier où  $I = \{1, \dots, n\}$ , on a :

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

- **Formule de Bayes.** Soient  $A \in \mathcal{A}$  et  $B \in \mathcal{A}$ , tels que  $0 < \mathbb{P}(A) < 1$  et  $\mathbb{P}(B) > 0$ , alors :

$$\mathbb{P}_B(A) = \frac{\mathbb{P}_A(B)\mathbb{P}(A)}{\mathbb{P}_A(B)\mathbb{P}(A) + \mathbb{P}_{A^c}(B)\mathbb{P}(A^c)}.$$

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

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

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

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

Par définition de la probabilité conditionnelle, on a :

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

On obtient la formule de Bayes en appliquant la formule des probabilités totales au dénominateur avec la partition  $\{A, A^c\}$  puisque  $\mathbb{P}(A) \neq 0$  et  $\mathbb{P}(A^c) \neq 0$ . □

**EXEMPLE 3.2.** Un sondage est effectué dans une société comprenant 40% de cadres et 60% d'ouvriers. On sait que 20% des cadres et 10% des ouvriers de cette société savent parler anglais. On interroge une personne "au hasard". Quelle est la probabilité que ce soit :

- un cadre sachant parler anglais ?
- un ouvrier sachant parler anglais ?
- une personne sachant parler anglais ?

L'individu interrogé sait parler anglais. Quelle est la probabilité que ce soit un ouvrier ?

On prend pour univers  $\Omega$  l'ensemble des employés de la société, que l'on munit de la tribu  $\mathcal{A} = \mathcal{P}(\Omega)$ . Étant donné que l'on interroge une personne "au hasard", on munit  $\Omega$  de la probabilité uniforme, notée  $\mathbb{P}$ .

Soit  $C$  l'évènement "l'employé est un cadre",  $O$  "l'employé est un ouvrier",  $A$  "l'employé parle anglais". Alors, la donnée du problème nous dit que :

$$\mathbb{P}(C) = \frac{4}{10}, \mathbb{P}(O) = \frac{6}{10}.$$

Comme ces deux probabilités sont strictement positives, on peut conditionner par rapport aux évènements  $C$  et  $O$ . D'après la donnée nous savons que :

$$\mathbb{P}_C(A) = \frac{2}{10}, \mathbb{P}_O(A) = \frac{1}{10}.$$

À la première question, on cherche  $\mathbb{P}(C \cap A)$ . D'après la définition des probabilités conditionnelles, on a :

$$\mathbb{P}(C \cap A) = \mathbb{P}_C(A)\mathbb{P}(C) = \frac{2}{10} \frac{4}{10} = 0,08.$$

De manière similaire pour la deuxième question :

$$\mathbb{P}(O \cap A) = \mathbb{P}_O(A)\mathbb{P}(O) = \frac{1}{10} \frac{6}{10} = 0,06.$$

Étant donné que  $\{C, O\}$  forme une partition de  $\Omega$ , d'après la formule des probabilités totales, on a :

$$\mathbb{P}(A) = \mathbb{P}(A \cap C) + \mathbb{P}(A \cap O) = 0,08 + 0,06 = 0,14.$$

Pour répondre à la dernière question, on utilise la formule de Bayes :

$$\mathbb{P}_A(O) = \frac{\mathbb{P}_O(A)\mathbb{P}(O)}{\mathbb{P}(A)} = \frac{1.6}{10.10} \frac{100}{14} = \frac{3}{7}.$$

EXEMPLE 3.3. (Paradoxe de Simpson). Cet exemple réel<sup>1</sup> montre un paradoxe surprenant qui s'explique grâce aux probabilités conditionnelles et à la formule des probabilités totales. Il vous convaincra de l'importance de bien comprendre ce concept pour interpréter correctement les résultats d'études statistiques. Il provient d'une étude médicale sur le succès de deux traitements contre les calculs rénaux. Le traitement A a été effectué dans les années 1972-1980, et le traitement B dans les années 1980-1985.

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

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

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

Résultats en fonction de la taille des calculs

petits calculs		grands calculs	
Traitement A	Traitement B	Traitement A	Traitement B
(81/87) 93%	(234/270) 87%	(192/263) 73%	(55/80) 69%

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

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

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

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

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

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

## EXERCICES

EXERCICE 3.1. On choisit une famille "au hasard" parmi toutes les familles ayant deux enfants.

1. Sachant que la famille choisie a au moins un garçon, quelle est la probabilité qu'elle ait deux garçons ?
2. Sachant que l'aîné de la famille choisie est un garçon, quelle est la probabilité que le plus jeune soit aussi un garçon ?

EXERCICE 3.2. *Un exemple d'urne de Polya.* Une urne contient au départ 5 boules blanches et 7 noires. Chaque fois que l'on tire une boule, on la réintroduit en rajoutant deux nouvelles boules de la même couleur que celle tirée. Quelle est la probabilité que les deux premières boules tirées soient noires ? Que la deuxième boule tirée soit noire ?

*Remarque :* les urnes de Polya peuvent servir pour modéliser la propagation de maladies infectieuses. En effet, chaque réalisation d'un évènement augmente la probabilité des réalisations suivantes.

EXERCICE 3.3. On considère trois cartes à jouer de même forme. Les deux faces de la première carte ont été colorées en noir, les deux faces de la deuxième en rouge tandis que la troisième porte une face noire et une face rouge. On mélange les trois cartes au fond d'un chapeau puis une carte est tirée au hasard et placée sur la table. Si la face apparente est rouge, quelle est la probabilité que l'autre soit noire ?

EXERCICE 3.4. Le test de dépistage d'un certain virus n'est pas infaillible :

- 1 fois sur 100, il est positif, alors que l'individu n'est pas contaminé,
- 2 fois sur 100, il est négatif alors que l'individu est contaminé.

D'autre part, on sait que sur la population totale, la fraction de porteurs est approximativement de  $1/1000$ .

1. Étant donné que le test est positif, quelle est la probabilité que l'individu ne soit pas porteur du virus ?
2. Étant donné que son test est négatif, quelle est la probabilité qu'un individu soit porteur du virus ?

EXERCICE 3.5. Un dé à six faces n'est pas bien équilibré, un échantillonnage a permis d'obtenir le tableau suivant.

Score	1	2	3	4	5	6	Total
Fréquence	0,1	0,2	0,1	0,4	0,1	0,1	1

On cherche à savoir si, avec ce dé, il est plus probable de faire un score d'au moins 5 lorsque le score est pair ou lorsque le score est impair.

1. Déterminer un choix d'espace probabilisé adapté à cette expérience.
2. Calculer la probabilité conditionnelle que le score soit d'au moins 5 sachant que le score est pair. Calculer la probabilité conditionnelle que le score soit d'au moins 5 sachant que le score est impair. Conclure.
3. Calculer la probabilité conditionnelle que le score soit impair sachant qu'il est d'au moins 5. Calculer la probabilité conditionnelle que le score soit pair sachant qu'il est d'au moins 5. Interpréter.

### 3.1.3 Arbre de probabilité

Nous trouvons plus judicieux de traiter ce thème sur un exemple tiré de l'épreuve "dossier" du CAPES externe.

EXEMPLE 3.4. (Dossier CAPES externe, juillet 2007). Une urne contient trois boules blanches et deux boules noires. On tire successivement et "au hasard" trois boules dans cette urne, en respectant le protocole suivant : on remet la boule dans l'urne si elle est noire, on ne la remet pas si elle est blanche.

1. Quelle est la probabilité de n'obtenir aucune boule blanche ?
2. Quelle est la probabilité d'obtenir trois boules blanches ?
3. Quelle est la probabilité d'obtenir exactement une boule blanche ?

Pendant sa présentation, le candidat traitera les questions suivantes.

1. Construire un arbre de probabilité permettant de répondre aux questions de l'exercice.
2. Utiliser un tel arbre pour répondre à la question 3 de l'exercice.

Sur ses fiches, le candidat rédigera et présentera :

- sa réponse à la question 1 ;
- l'énoncé de 1 ou plusieurs exercices se rapportant au thème "Probabilités".

Étant donné que nous sommes intéressés par les trois premiers tirages, nous choisissons comme espace des états :

$$\Omega = \{B, N\}^3 = \{(B, B, B), (B, B, N), \dots, (N, N, N)\}.$$

On munit  $\Omega$  de la tribu  $\mathcal{A} = \mathcal{P}(\Omega)$  et d'une probabilité, notée  $\mathbb{P}$ . Attention, la probabilité  $\mathbb{P}$  n'est pas la probabilité uniforme sur  $(\Omega, \mathcal{A})$ . Pour  $i \in \{1, 2, 3\}$ , définissons  $B_i$  (resp.  $N_i$ ) l'évènement "la  $i$ -ième boule tirée est de couleur blanche (resp. noire)", alors les évènements  $\{B_i, N_i\}$  forment une partition de  $\Omega$ .

### Calcul des probabilités et probabilités conditionnelles

Au moment du premier tirage, il y a 3 boules blanches dans l'urne et 2 boules noires. Comme les boules sont tirées "au hasard", on déduit que :

$$\mathbb{P}(B_1) = \frac{3}{5}, \quad \mathbb{P}(N_1) = \frac{2}{5}.$$

Ces deux probabilités étant strictement positives, il fait sens de conditionner par les évènements correspondants.

Le protocole de tirage nous dit également que les probabilités  $\mathbb{P}(B_1 \cap B_2), \dots, \mathbb{P}(N_1 \cap N_2)$  sont strictement positives, on peut donc conditionner par les évènements correspondants.

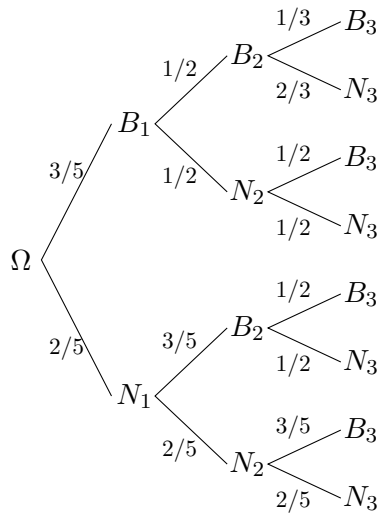
Supposons qu'au premier tirage une boule blanche est tirée. Comme conséquence du protocole de tirage, il reste alors 2 boules blanches et 2 noires. Étant donné que les boules sont tirées "au hasard", on conclut :

$$\mathbb{P}_{B_1}(B_2) = \frac{1}{2}, \quad \mathbb{P}_{B_1}(N_2) = \frac{1}{2}.$$

Supposons que lors des deux premiers tirages une boule blanche est tirée. Alors il reste, à l'issue de ces deux tirages, 1 boule blanche et 2 noires. Étant donné que les boules sont tirées "au hasard", on conclut :

$$\mathbb{P}_{B_1 \cap B_2}(B_3) = \frac{1}{3}, \quad \mathbb{P}_{B_1 \cap B_2}(N_3) = \frac{2}{3}.$$

Les autres probabilités conditionnelles sont calculées de manière similaires, et résumées sur l'arbre de probabilité ci-dessous.



### Interprétation de l'arbre de probabilité

L'arbre de probabilité est un arbre orienté et pondéré, satisfaisant aux conditions suivantes.

- Les sommets de l'arbre sont formés d'évènements de  $\Omega$ , et la racine est  $\Omega$ .
- Si  $A$  est un sommet de l'arbre, alors les feuilles de  $A$  forment une partition de  $A$ .
- Soit  $n \geq 1$ , et  $A_0, A_1, \dots, A_n$  un chemin de l'arbre (où  $A_0 = \Omega$ ). Alors le nombre situé sur la branche de  $A_{n-1}$  vers  $A_n$  est la probabilité conditionnelle de  $A_n$  sachant que  $A_1 \cap \dots \cap A_{n-1}$  est réalisé.

Ainsi, la somme des probabilités des branches issues d'un même sommet est 1.

Soit  $n \geq 1$ , et  $A_0, A_1, \dots, A_n$  un chemin de l'arbre,  $A_0 = \Omega$ . Alors, par définition des probabilités conditionnelles, la probabilité  $\mathbb{P}(A_1 \cap \dots \cap A_n)$  est obtenue en effectuant le produit des nombres situés sur les branches formant le chemin, en effet :

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

En sommant le produit le long de différentes branches, on retrouve différentes formes de la formule des probabilités totales.

### Solutions des questions 1-3

L'évènement "obtenir aucune boule blanche" est le même que l'évènement "obtenir 3 boules noires" qui correspond au sous-ensemble  $N_1 \cap N_2 \cap N_3$  de  $\Omega$ . D'après le point 4 ci-dessus, on déduit que  $\mathbb{P}(N_1 \cap N_2 \cap N_3)$  s'obtient en effectuant le produit des probabilités le long du dernier chemin de l'arbre.

$$\mathbb{P}(N_1 \cap N_2 \cap N_3) = \mathbb{P}(N_1) \mathbb{P}_{N_1}(N_2) \mathbb{P}_{N_1 \cap N_2}(N_3) = \left(\frac{2}{5}\right)^3 = \frac{8}{125}.$$

De manière similaire, la probabilité d'obtenir trois boules blanches s'obtient en effectuant le produit des probabilités le long du premier chemin de l'arbre :

$$\mathbb{P}(B_1 \cap B_2 \cap B_3) = \frac{3}{5.2.3} = \frac{1}{10}.$$

L'évènement "obtenir exactement une boule blanche" correspond au sous-ensemble

$$(B_1 \cap N_2 \cap N_3) \cup (N_1 \cap B_2 \cap N_3) \cup (N_1 \cap N_2 \cap B_3).$$

Comme ces évènements sont disjoints, cette probabilité s'obtient en effectuant le produit le long de la 4-ième, 6-ième et 7-ième branche et en sommant les résultats obtenus. Ainsi :

$$\begin{aligned} & \mathbb{P}((B_1 \cap N_2 \cap N_3) \cup (N_1 \cap B_2 \cap N_3) \cup (N_1 \cap N_2 \cap B_3)) \\ &= \mathbb{P}(B_1 \cap N_2 \cap N_3) + \mathbb{P}(N_1 \cap B_2 \cap N_3) + \mathbb{P}(N_1 \cap N_2 \cap B_3) \\ &= \frac{3}{5.2.2} + \frac{2.3}{5.5.2} + \frac{2.2.3}{5.5.5} = \frac{183}{500}. \end{aligned}$$

EXERCICE 3.6. Donner les arbres de probabilités correspondant à :

1. l'exemple de l'entreprise (cadres/ouvriers, sachant parler anglais ou non) ;
2. l'Exercice 2. sur les urnes de Polya.

Résoudre les exercices en expliquant.

## 3.2 Indépendance des évènements

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

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

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

EXEMPLE 3.5.



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

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

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

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

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

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

*Remarque.*

- L'indépendance n'a rien à voir avec le fait que les évènements soient disjoints ou non. Dans l'Exemple 2 ci-dessus, les évènements sont indépendants, mais non disjoints ( $A \cap B \neq \emptyset$ ).
- Si  $A$  et  $B$  sont deux évènements de probabilité non nulle, alors :

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

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

**Proposition 3.1.** *Si les évènements  $A$  et  $B$  sont indépendants, alors il en est de même des évènements  $A^c$  et  $B$ ,  $A$  et  $B^c$ ,  $A^c$  et  $B^c$ .*

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

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

En remplaçant  $A$  et  $B$  par  $B$  et  $A^c$ , l'implication que l'on vient de prouver entraîne que si  $A^c$  et  $B$  sont indépendants (et donc  $B$  et  $A^c$  indépendants), alors  $B^c$  et  $A^c$  sont indépendants. Et ainsi de suite pour les autres cas.  $\square$

**Définition.**

- Des évènements  $A_1, \dots, A_n$  de  $\mathcal{A}$  sont dits *mutuellement indépendants*, si pour tout entier  $1 \leq k \leq n$  et tout  $k$ -uplet d'entiers  $(i_1, \dots, i_k)$ , tels que,  $1 \leq i_1 < \dots < i_k \leq n$ ,

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

- Les évènements  $A_1, \dots, A_n$  sont dits *deux-à-deux indépendants*, si pour tous entiers  $i \in \{1, \dots, n\}$  et  $j \in \{1, \dots, n\}$ , tels que  $i \neq j$ , on a :

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

**EXERCICE 3.7.** Soit  $\Omega = \{1, 2, 3, 4\}$  muni de la tribu  $\mathcal{A} = \mathcal{P}(\Omega)$  et de la probabilité uniforme, notée  $\mathbb{P}$ . On considère les évènements  $A = \{1, 2\}$ ,  $B = \{2, 3\}$ ,  $C = \{1, 3\}$ .

- Montrer que les évènements  $A, B, C$ , sont deux-à-deux indépendants,

— Montrer que les évènements  $A, B, C$ , ne sont pas mutuellement indépendants. Ainsi l'indépendance deux-à-deux est plus faible que l'indépendance mutuelle.

EXERCICE 3.8. Une urne contient 9 boules indiscernables, numérotées de 1 à 9. On tire une boule "au hasard". Les évènements suivants sont-ils indépendants ?

1.  $A$  : "la boule tirée porte un numéro pair",
2.  $B$  : "le numéro tiré est multiple de 3".

Répondre à la même question lorsque l'urne contient 12 boules.

### 3.2.1 Nombre infini de jets de dés

Dans cette section, nous abordons de manière approfondie un exemple classique qui utilise la notion d'indépendance : la description mathématique d'une expérience aléatoire répétée, dans les mêmes conditions, et de manière indépendante, un nombre fini ou infini de fois. Afin de fixer les idées, nous prendrons comme exemple des jets indépendants d'un dé non pipé, mais vous pourriez aussi imaginer prendre les jets d'une pièce de monnaie, etc.

#### Un jet de dé

On choisit comme espace des états,  $\Omega_1 = \{1, \dots, 6\}$  et comme tribu  $\mathcal{A}_1 = \mathcal{P}(\Omega_1)$ . Étant donné que le dé est non-pipé, on munit  $\Omega_1$  de la probabilité uniforme, notée  $\mathbb{P}_1$ , de sorte que :

$$\forall i \in \{1, \dots, 6\}, \mathbb{P}_1(\{i\}) = \frac{1}{6}.$$

#### Deux jets de dés indépendants

On choisit comme espace des états,

$$\Omega_2 = \{(i, j) \mid i \in \{1, \dots, 6\}, j \in \{1, \dots, 6\}\} = \{1 \dots, 6\}^2.$$

Pour les mêmes raisons que dans le cas d'un dé, on choisit comme tribu  $\mathcal{A}_2 = \mathcal{P}(\Omega_2)$  et l'on munit  $\Omega_2$  de la probabilité uniforme, notée  $\mathbb{P}_2$ , de sorte que :

$$\forall (i, j) \in \Omega_2, \mathbb{P}_2(\{(i, j)\}) = \frac{1}{36}.$$

Une autre manière de voir ceci est la suivante : nous souhaitons modéliser l'indépendance des expériences successives, par conséquent deux évènements portant l'un sur la première expérience, l'autre sur la deuxième, doivent être indépendants.

Pour  $k \in \{1, \dots, 6\}$ ,  $\ell \in \{1, 2\}$ , on considère l'évènement  $E_k^\ell$  "le  $\ell$ -ième jet donne  $k$ ". Lorsque  $\ell = 1$ , cet évènement est le sous-ensemble  $A_k^1(\Omega_2) = \{k\} \times \Omega_1$  de  $\Omega_2$ , et lorsque  $\ell = 2$ , c'est le sous-ensemble  $A_k^2(\Omega_2) = \Omega_1 \times \{k\}$  de  $\Omega_2$ . Remarquons de plus que l'évènement élémentaire  $\{(i, j)\}$  de  $\Omega_2$  s'écrit :

$$\{(i, j)\} = A_i^1(\Omega_2) \cap A_j^2(\Omega_2).$$

Ainsi, on souhaite que :

$$\mathbb{P}_2(\{(i, j)\}) = \mathbb{P}_2[A_i^1(\Omega_2) \cap A_j^2(\Omega_2)] = \mathbb{P}_2[A_i^1(\Omega_2)]\mathbb{P}_2[A_j^2(\Omega_2)].$$

D'autre part, comme l'évènement  $E_i^1$  ne dépend que du premier jet et  $E_j^2$  que du deuxième, on peut les écrire comme des sous-ensembles de  $\Omega_1$  (représentant la première et la deuxième

expérience respectivement) :  $E_i^1$  est le sous-ensemble  $A_i^1(\Omega_1) = \{i\}$  de  $\Omega_1$  et  $E_j^2$  est le sous-ensemble  $A_j^2(\Omega_1) = \{j\}$  de  $\Omega_1$ . Donc, il est naturel de demander que :

$$\begin{aligned} \mathbb{P}_2(\{(i, j)\}) &= \mathbb{P}_2[A_i^1(\Omega_2) \cap A_j^2(\Omega_2)] = \mathbb{P}_2[A_i^1(\Omega_2)]\mathbb{P}_2[A_j^2(\Omega_2)] \\ &= \mathbb{P}_1[A_i^1(\Omega_1)]\mathbb{P}_1[A_j^2(\Omega_1)] = \mathbb{P}_1(\{i\})\mathbb{P}_1(\{j\}) \\ &= \frac{1}{6^2} = \frac{1}{36}. \end{aligned}$$

### $n$ jets de dés indépendants

Ceci se généralise naturellement au cas de  $n$  jets.

$$\Omega_n = \{1, \dots, 6\}^n, \mathcal{A}_n = \mathcal{P}(\Omega_n),$$

Pour tout  $(i_1, \dots, i_n) \in \Omega_n$ ,  $\mathbb{P}_n[\{(i_1, \dots, i_n)\}] = \frac{1}{6^n}$ .

### Répétition infinie de jets de dés indépendants

Ceci sort du programme à proprement parlé, cependant c'est un exemple qui apparaît souvent sous une forme déguisée. On choisit comme espace des états  $\Omega_\infty = \{1, \dots, 6\}^{\mathbb{N}^*}$ . Attention  $\Omega_\infty$  n'est pas dénombrable, donc on ne choisit pas comme tribu  $\mathcal{P}(\Omega_\infty)$ .

La tribu  $\mathcal{A}_\infty$  sur  $\Omega_\infty$  est la tribu engendrée par les évènements ne dépendant que d'un nombre fini de jets. Soit  $n \geq 1$ , et pour tout  $k \in \{1, \dots, n\}$ , soit  $i_k \in \{1, \dots, 6\}$ . On définit l'"évènement"  $E_{i_1, \dots, i_n}$  "l'issue du premier tirage est  $i_1, \dots$ , l'issue du  $n$ -ième tirage est  $i_n$ ". Alors  $E_{i_1, \dots, i_n}$  est le sous-ensemble  $A_{i_1, \dots, i_n}(\Omega_\infty) = \{i_1\} \times \dots \times \{i_n\} \times \Omega_1 \times \Omega_1 \dots$  de  $\Omega_\infty$ , et le sous-ensemble  $A_{i_1, \dots, i_n}(\Omega_n) = \{(i_1, \dots, i_n)\}$  de  $\Omega_n$ .

En utilisant le théorème d'extension de Carthéodory, on montre qu'il existe une unique probabilité  $\mathbb{P}_\infty$  sur  $(\Omega_\infty, \mathcal{A})$ , telle que :

$$\begin{aligned} \mathbb{P}_\infty[A_{i_1, \dots, i_n}(\Omega_\infty)] &= \mathbb{P}_n[A_{i_1, \dots, i_n}(\Omega_n)] = \mathbb{P}_n(\{(i_1, \dots, i_n)\}) \\ &= \mathbb{P}_1(\{i_1\}) \dots \mathbb{P}_1(\{i_n\}) = \frac{1}{6^n}. \end{aligned}$$

Cette probabilité s'appelle la *probabilité produit* sur  $\Omega_\infty$ . Vous ne devez pas connaître les détails de la construction, cependant l'idée à retenir est la suivante : si l'univers est  $\Omega_\infty$ , la probabilité choisie sur  $\Omega_\infty$  permet de calculer la probabilité de tous les évènements qui ne dépendent que d'un nombre fini de jets et cette probabilité est donnée par le produit des probabilités pour chacun des jets.

**EXERCICE 3.9.** On lance un dé à 6 faces non-truqué, indéfiniment.

1. Décrire l'univers  $\Omega_\infty$  associé à cette expérience aléatoire.
2. Pour  $n \in \mathbb{N}^*$ , soit  $A_n$  l'évènement "on obtient 1 pour la première fois au  $n$ -ième jet". Calculer la probabilité de l'évènement  $A_n$ .
3. Pour  $n \in \mathbb{N}^*$ , soit  $B_n$  l'évènement "on obtient 1 aux  $n$  premiers jets, et soit  $B$  l'évènement "on obtient toujours 1". Calculer la probabilité de  $B_n$  et  $B$ .



# Chapitre 4

## Variables aléatoires

### 4.1 Définition et loi d'une variable aléatoire

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

Lorsque  $X$  ne prend qu'un nombre dénombrable de valeurs  $X(\Omega) = \{x_j : j \in J\}$ , où  $J$  est une partie non-vide finie ou dénombrable de  $\mathbb{N}$ , alors  $X$  est appelée une variable aléatoire *discrète*.

*Remarque.*

- La terminologie de variable aléatoire peut être trompeuse, car il s'agit en fait d'une fonction de  $\Omega$  dans  $E$ .
- Afin de pouvoir définir une probabilité de manière cohérente, il faut supposer de plus que pour une classe importante de parties  $B$  de  $E$ , l'ensemble  $\{\omega \in \Omega : X(\omega) \in B\}$  appartient à  $\mathcal{A}$ . Formellement, on munit  $E$  d'une tribu  $\mathcal{E}$ . Une *variable aléatoire*  $X$  est alors une application de  $(\Omega, \mathcal{A})$  dans  $(E, \mathcal{E})$  telle que :

$$\forall B \in \mathcal{E}, \{\omega \in \Omega : X(\omega) \in B\} \in \mathcal{A}.$$

**Hors programme.** Une variable aléatoire  $X$  est en fait une application *mesurable* de  $(\Omega, \mathcal{A})$  dans  $(E, \mathcal{E})$ .

- Soit  $B$  un sous-ensemble de  $E$ . Rappelez-vous les notations :

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

$\{X \in B\}$  est l'évènement " $X$  prend une valeur appartenant à  $B$ ". Attention,  $X^{-1}(B)$  désigne l'image réciproque de la partie  $B$  par l'application  $X$ . Cela ne sous-entend nullement que  $X$  est bijective !

EXEMPLE 4.1.

1. On considère deux lancers successifs d'un même dé, et on note  $S$  la variable aléatoire correspondant à la somme des valeurs obtenues, ainsi  $S$  prend ses valeurs dans  $\{2, 3, \dots, 12\}$ .
2. Soit  $(\Omega, \mathcal{A}, \mathbb{P})$  un espace probabilisé, et  $A \in \mathcal{A}$  un évènement. Alors l'*indicatrice* de  $A$ , notée  $\mathbb{I}_A$ , est la variable aléatoire définie sur  $\Omega$  par :

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

Ainsi l'indicatrice de  $A$  prend ses valeurs dans  $\{0, 1\}$ . En probabilité, il est important de bien savoir manipuler cette notation. En particulier l'indicatrice satisfait aux conditions suivantes. Si  $A$  et  $B$  sont deux évènements de  $\mathcal{A}$ , alors :

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

**Définition.** Soient  $(\Omega, \mathcal{A}, \mathbb{P})$  un espace probabilisé et  $X$  une variable aléatoire définie sur  $\Omega$ , à valeurs dans  $E$ . On définit la *loi de probabilité de  $X$* , notée  $\mathbb{P}^X$  de la manière suivante. Pour tout sous-ensemble  $B$  de  $E$  tel que  $\{X \in B\} \in \mathcal{A}$  :

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

*Remarque.* Reprenons le cadre plus formel, où l'on munit  $E$  d'une tribu  $\mathcal{E}$ . La loi de probabilité de  $X$  est alors une probabilité sur  $(E, \mathcal{E})$ . C'est la transposition de la structure abstraite  $(\Omega, \mathcal{A}, \mathbb{P})$  sur  $(E, \mathcal{E})$ . L'espace  $E$  muni de la tribu  $\mathcal{E}$  et de la probabilité  $\mathbb{P}^X$  devient un espace probabilisé  $(E, \mathcal{E}, \mathbb{P}^X)$ .

**Proposition 4.1.**

— Dans le cas où  $X$  est une variable aléatoire discrète,  $X(\Omega) = \{x_j : j \in J\}$ , où  $J$  est une partie non-vide, finie ou dénombrable de  $\mathbb{N}$ , alors la loi de probabilité de  $X$  est caractérisée par la donnée :

$$\forall j \in J, \mathbb{P}^X(\{x_j\}).$$

— Dans le cas où  $X(\Omega) = \mathbb{R}$ , la loi de probabilité de  $X$  est caractérisée par la donnée de :

$$\forall (a, b) \in \mathbb{R}^2, a < b, \mathbb{P}^X([a, b]).$$

**EXEMPLE 4.2.**

1. Choisissons comme univers  $\Omega = \{1, \dots, 6\}^2$ , que l'on munit de la tribu  $\mathcal{A} = \mathcal{P}(\Omega)$ , et de la probabilité uniforme, notée  $\mathbb{P}$ . Ainsi pour tout sous-ensemble  $A$  de  $\mathcal{P}(\Omega)$ ,  $\mathbb{P}(A) = \frac{\text{Card}(A)}{\text{Card}(\Omega)}$ . Étant donné que  $S$  prend ses valeurs dans  $\{2, \dots, 12\}$ , pour déterminer la loi de probabilité  $\mathbb{P}^S$  de  $S$ , il suffit de calculer, pour tout  $i \in \{2, \dots, 12\}$ ,  $\mathbb{P}^S(\{i\})$  :

$$\begin{aligned} \mathbb{P}^S(\{2\}) &= \mathbb{P}(\{S \in \{2\}\}) = \mathbb{P}(\{\omega \in \Omega : S(\omega) \in \{2\}\}) \\ &= \mathbb{P}(\{\omega \in \Omega : S(\omega) = 2\}) = \mathbb{P}(\{(1, 1)\}) = \frac{1}{36}, \\ \mathbb{P}^S(\{3\}) &= \mathbb{P}(\{\omega \in \Omega : S(\omega) = 3\}) = \mathbb{P}(\{(1, 2), (2, 1)\}) = \frac{1}{18}, \\ &\text{etc.} \end{aligned}$$

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

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

## 4.2 Fonction de répartition

**Définition.** Soit  $X$  une variable aléatoire de loi  $\mathbb{P}^X$  à valeurs dans  $\mathbb{R}$  (ou une partie de  $\mathbb{R}$ ). On appelle *fonction de répartition de la loi  $\mathbb{P}^X$*  ou encore, par abus, *fonction de répartition de  $X$* , l'application  $F_X$  définie sur  $\mathbb{R}$  par :

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

**Propriété 2.** La fonction de répartition de la loi  $\mathbb{P}^X$  satisfait aux propriétés suivantes :

1.  $F_X$  prend ses valeurs dans  $[0, 1]$ ,
2.  $F_X$  est une application croissante,
3.  $F_X$  est continue à droite et admet une limite à gauche,
4.  $\lim_{t \rightarrow -\infty} F_X(t) = 0$ ,  $\lim_{t \rightarrow +\infty} F_X(t) = 1$ .

*Démonstration.* Le point 1. est une conséquence de la définition d'une probabilité. Le point 2. découle de la propriété de croissance des probabilités :

$$s \leq t \Rightarrow ]-\infty, s] \subset ]-\infty, t] \Rightarrow \mathbb{P}^X(]-\infty, s]) \leq \mathbb{P}^X(]-\infty, t]).$$

Étant donné que  $F_X$  est croissante, pour montrer le point 3. il suffit de voir que, pour tout  $t \in \mathbb{R}$  :

$$\lim_{n \rightarrow +\infty} F_X\left(t + \frac{1}{n}\right) = F_X(t), \text{ et } \lim_{n \rightarrow +\infty} F_X\left(t - \frac{1}{n}\right) \text{ existe.}$$

Pour tout  $n \geq 1$ , on définit  $A_n = ]-\infty, t + \frac{1}{n}]$  et  $B_n = ]-\infty, t - \frac{1}{n}]$ . Alors  $(A_n)_{n \geq 1}$  est une suite décroissante d'évènements qui vérifie  $\bigcap_{n \geq 1} A_n = ]-\infty, t]$ ; et  $(B_n)_{n \geq 1}$  est une suite croissante d'évènements, telle que  $\bigcup_{n \geq 1} B_n = ]-\infty, t[$ . Donc, d'après la Proposition 1.2, on sait que :

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

ce qui démontre le point 3. Attention  $\mathbb{P}^X(]-\infty, t[)$  peut-être différent de  $\mathbb{P}^X(]-\infty, t])$ , car

$$\mathbb{P}^X(]-\infty, t]) - \mathbb{P}^X(]-\infty, t[) = \mathbb{P}(X = t),$$

qui peut être non nulle. Mais si  $\mathbb{P}(X = t) = 0$ , alors  $F^X$  est continue au point  $t$ .

De manière analogue, pour démontrer le point 4. il suffit de montrer que :

$$\lim_{n \rightarrow +\infty} F_X(-n) = 0, \text{ et } \lim_{n \rightarrow +\infty} F_X(n) = 1.$$

Pour tout  $n \geq 1$ , on définit  $C_n = ]-\infty, -n]$  et  $D_n = ]-\infty, n]$ . Alors  $(C_n)_{n \geq 1}$  est une suite décroissante d'évènements telle que  $\bigcap_{n \geq 1} C_n = \emptyset$ ; et  $(D_n)_{n \geq 1}$  est une suite croissante telle que  $\bigcup_{n \geq 1} D_n = \mathbb{R}$ . Ainsi, d'après la Proposition 1.2 et la définition d'une probabilité, on déduit :

$$\begin{aligned} \lim_{n \rightarrow +\infty} \mathbb{P}^X(C_n) &= \mathbb{P}^X(\emptyset) = 0 \Leftrightarrow \lim_{n \rightarrow +\infty} F_X(-n) = 0, \\ \lim_{n \rightarrow +\infty} \mathbb{P}^X(D_n) &= \mathbb{P}^X(\mathbb{R}) = 1 \Leftrightarrow \lim_{n \rightarrow +\infty} F_X(n) = 1. \end{aligned}$$

□

La proposition suivante nous dit qu'une fonction de répartition caractérise la loi de la variable aléatoire.

**Proposition 4.2.** *Toute application définie de  $\mathbb{R}$  dans  $[0, 1]$  qui possède les propriétés 2, 3, 4, est la fonction de répartition d'une unique loi de probabilité sur  $\mathbb{R}$ .*

La fonction de répartition permet de calculer les probabilités suivantes :

**Propriété 3.** *Soit  $(\Omega, \mathcal{A}, \mathbb{P})$  un espace probabilisé, et  $X$  une variable aléatoire sur  $\Omega$  de fonction de répartition  $F_X$ , alors :*

- $\mathbb{P}(X > x) = 1 - F_X(x)$ ,
- $\mathbb{P}(x < X \leq y) = F_X(y) - F_X(x)$ ,
- $\mathbb{P}(X < x) = F_X(x^-)$ .

Pour la suite de la théorie, nous allons traiter trois cas séparément, selon que la variable aléatoire soit discrète (finie ou infinie) ou continue et à densité. Ceci pourrait être unifié en utilisant la théorie de la mesure, mais nous amènerait en dehors du programme.

## 4.3 Variables aléatoires discrètes

### 4.3.1 Définitions et exemples à connaître

Rappelons la restriction des définitions générales au cas discret. Soit  $(\Omega, \mathcal{A}, \mathbb{P})$  un espace probabilisé.

**Définition.** Une variable aléatoire  $X$  définie sur  $\Omega$  est dite *discrète* si elle prend ses valeurs dans un ensemble discret :  $X(\Omega) = \{x_j : j \in J\} \subset \mathbb{R}$ , où  $J$  est une partie non-vide finie ou dénombrable de  $\mathbb{N}$ .

*Remarque.* Lorsque  $\Omega$  est fini, toute application définie sur  $\Omega$  est une variable aléatoire discrète.

**Proposition 4.3.** *Si  $X$  est une variable aléatoire discrète définie sur  $\Omega$ , alors la loi de probabilité de  $X$  est caractérisée par la donnée de la famille  $\{(x_j, \mathbb{P}^X(\{x_j\}))\}, j \in J$ , où :*

$$\mathbb{P}^X(\{x_j\}) = \mathbb{P}(\{X = x_j\}).$$

Voici une liste des lois discrètes classiques à connaître.

**Loi uniforme.** Soit  $n \in \mathbb{N}^*$  et une variable aléatoire,  $X : \Omega \rightarrow \{x_1, \dots, x_n\}$ , où pour tout  $i \neq j$ ,  $x_i \neq x_j$ . Supposons que la loi de probabilité de  $X$  est donnée par :

$$\forall j \in \{1, \dots, n\}, \mathbb{P}^X(\{x_j\}) = \frac{1}{n}.$$

Alors la loi de  $X$  est appelée la *loi uniforme discrète* sur  $\{x_1, \dots, x_n\}$ .

**EXEMPLE 4.3.** On lance une pièce équilibrée. Soit  $\Omega = \{P, F\}$  que l'on munit de la probabilité uniforme  $\mathbb{P}$ . On définit la variable aléatoire  $X : \Omega \rightarrow \{-1, 1\}$ , par  $X(\{P\}) = 1$ ,  $X(\{F\}) = -1$ . Ainsi, la loi de probabilité de  $X$  est :

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

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



**Loi de Bernoulli.** Soit  $0 < p < 1$  et une variable aléatoire  $X : \Omega \rightarrow \{0, 1\}$  de loi de probabilité :

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

Alors la loi de  $X$  est appelée *loi de Bernoulli de paramètre  $p$* .

EXEMPLE 4.4. Une *épreuve de Bernoulli de paramètre  $p$*  est une expérience aléatoire admettant deux issues succès/échec, telle que  $p$  est la probabilité d'obtenir un succès. Soit  $\Omega$  représentant les issues de l'expérience,  $\mathbb{P}$  une probabilité sur  $\Omega$  et  $A$  l'évènement représentant le succès. D'après la description de l'expérience, on a  $\mathbb{P}(A) = p$ . Dans ce cas, l'indicatrice de  $A$ ,  $\mathbb{I}_A$ , suit une loi de Bernoulli de paramètre  $p$ . En effet,  $\mathbb{I}_A$  est à valeurs dans  $\{0, 1\}$  et nous avons déjà vu que :

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

On conclut en utilisant le fait que  $\mathbb{P}(A) = p$ .

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

**Loi binomiale.** Soit  $n \in \mathbb{N}^*$ ,  $0 < p < 1$  et une variable aléatoire  $X : \Omega \rightarrow \{0, \dots, n\}$  de loi de probabilité :

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

Alors la loi de  $X$  est appelée *loi binomiale de paramètres  $n$  et  $p$* .

EXEMPLE 4.5. On appelle *schéma de Bernoulli de paramètres  $n$  et  $p$*  l'expérience qui consiste à répéter  $n$  fois de manière indépendante une épreuve de Bernoulli de paramètre  $p$ . On choisit comme univers  $\Omega^n$ , que l'on munit d'une probabilité  $\mathbb{P}_n$ . Pour  $i \in \{1, \dots, n\}$ , on note  $A_i$  l'évènement "obtenir un succès lors de la  $i$ -ième expérience". Soit  $k \in \{0, \dots, n\}$ , comme les expériences sont indépendantes, la probabilité d'obtenir un succès lors des  $k$  premières expériences et un échec lors des  $n - k$  dernières est :

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

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

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

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

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

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

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

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

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

**Loi géométrique.** Soit  $0 < p < 1$  et une variable aléatoire  $X : \Omega \rightarrow \mathbb{N}^*$  de loi de probabilité :

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

Alors la loi de  $X$  est appelée *loi géométrique de paramètre  $p$* . En translatant de 1 les valeurs de la variable aléatoire, on obtient la loi géométrique sur  $\mathbb{N}$ .

EXEMPLE 4.6. Considérons un schéma de Bernoulli où l'expérience est répétée indéfiniment. Si  $X$  est la variable aléatoire égale au temps passé jusqu'au premier succès, alors  $X$  suit une loi géométrique de paramètre  $p$ . En effet, choisissons comme univers  $\Omega^{\mathbb{N}^*}$  et calculons la loi de probabilité de  $X$ . Pour tout  $k \in \mathbb{N}^*$ ,

$$\begin{aligned} \mathbb{P}^X(\{k\}) &= \mathbb{P}_k(A_1^c \cap \dots \cap A_{k-1}^c \cap A_k) \\ &= \mathbb{P}(A_1^c) \dots \mathbb{P}(A_{k-1}^c) \mathbb{P}(A_k), \text{ par indépendance} \\ &= (1 - p)^{k-1}p. \end{aligned}$$

**Loi de Poisson.** Soit  $\theta > 0$  et une variable aléatoire  $X : \Omega \rightarrow \mathbb{N}$  de loi :

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

Alors la loi de  $X$  est appelée *loi de Poisson de paramètre  $\theta$* .

EXEMPLE 4.7. La loi de Poisson modélise le nombre d'autobus passés à un arrêt avant un instant  $T$  donné, et  $\theta$  représente le nombre moyen d'arrivées dans cet intervalle.

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

**Proposition 4.4.** Soit  $(X_n)_{n \geq 1}$  une suite de variables aléatoires. On suppose que, pour tout entier  $n \geq 1$ , la variable aléatoire  $X_n$  suit une loi binomiale de paramètres  $n$  et  $p_n$ , et que  $np_n$  tende vers un nombre réel strictement positif  $\theta$  lorsque  $n$  tend vers  $+\infty$ . Alors, pour tout entier naturel  $k$  :

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

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

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

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

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

Il s'ensuit que

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

d'où le résultat. □

## EXERCICES

**EXERCICE 4.1.** Pour chacune des lois ci-dessus, montrer que la somme des probabilités des événements élémentaires vaut 1.

**EXERCICE 4.2.** Soit  $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5\}$  un espace fini à 5 éléments, de probabilités respectives  $1/4, 1/4, 1/6, 1/6, 1/6$ . On note  $X$  la variable aléatoire définie par :

$$X(\omega_1) = X(\omega_2) = 0, \quad X(\omega_3) = X(\omega_4) = 1, \quad X(\omega_5) = 2.$$

Déterminer la loi de la variable aléatoire  $X$ .

**EXERCICE 4.3.** Une urne contient  $n$  boules numérotées de 1 à  $n$ ,  $n \geq 3$ . On tire 3 boules d'un seul coup. Soit  $X$  la variable aléatoire égale à 1 si on a tiré la boule no 1, égale à 0 dans le cas contraire. Donner la loi de la variable aléatoire  $X$ .

**EXERCICE 4.4.** On lance deux dés équilibrés. On note  $X$  le plus grand des numéros obtenus. Déterminer la loi de la variable aléatoire  $X$ .

**EXERCICE 4.5.** Une urne contient  $N_b$  boules blanches et  $N_n$  boules noires. Posons  $N = N_b + N_n$ . On tire  $r$  boules avec remise dans l'urne. Soit  $X$  la variable aléatoire égale au nombre de boules blanches tirées. Déterminer la loi de la variable aléatoire  $X$ . Reconnaitre la loi de  $X$ .

**EXERCICE 4.6.** On lance un dé à 6 faces non truqué, indéfiniment. Soit  $X$  la variable aléatoire égale au temps passé jusqu'à ce que le premier 1 soit obtenu. Déterminer la loi de la variable aléatoire  $X$ . Reconnaitre la loi de  $X$ .

### 4.3.2 Fonction de répartition

Dans le cas des variables aléatoires discrètes, la fonction de répartition vérifie les propriétés suivantes.

**Propriété 4.** Soit  $(\Omega, \mathcal{A}, \mathbb{P})$  un espace probabilisé et soit  $X$  une variable aléatoire discrète réelle définie sur  $\Omega$ . Alors la fonction de répartition  $F_X$  de la loi  $\mathbb{P}^X$ , vérifie :

1.  $F_X(x) = \sum_{\{y \in X(\Omega) : y \leq x\}} \mathbb{P}^X(\{y\})$  ;

2. si  $x$  et  $y$  sont deux points consécutifs de  $X(\Omega)$ , alors  $F_X$  est constante sur  $[x, y[$ ;
3. la hauteur du saut en  $x \in X(\Omega)$  est donnée par  $\mathbb{P}^X(\{x\})$ .

EXERCICE 4.7. Calculer la fonction de répartition de :

1. la loi uniforme sur  $\{1, \dots, n\}$ ,
2. la loi de Bernoulli de paramètre  $p$ ,
3. la loi binomiale de paramètres 4 et  $\frac{1}{2}$ ,
4. la loi géométrique de paramètre  $p$ .

### 4.3.3 Espérance

L'espérance d'une variable aléatoire représente sa moyenne pondérée par la probabilité de chacune des issues. Lors d'un jeu de hasard par exemple, elle permet de déterminer si le jeu est équitable ou non.

• **Cas où l'univers est fini.**

Soit  $(\Omega, \mathcal{A}, \mathbb{P})$  un espace probabilisé. Supposons l'univers  $\Omega$  fini :  $\Omega = \{\omega_1, \dots, \omega_n\}$ .

**Définition.** Soit  $X$  une variable aléatoire réelle définie sur  $\Omega$ . On appelle *espérance de  $X$* , que l'on note  $\mathbb{E}(X)$ , la quantité :

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

L'espérance satisfait aux propriétés suivantes.

**Propriété 5.**

1. (Linéarité). Si  $X$  et  $Y$  sont deux variables aléatoires définies sur  $\Omega$ , et si  $a, b$  sont deux constantes réelles, alors :

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

2. (Monotonie). Si  $X$  et  $Y$  sont deux variables aléatoires définies sur  $\Omega$  telle que  $X \leq Y$ , alors  $\mathbb{E}(X) \leq \mathbb{E}(Y)$ . En particulier,  $|\mathbb{E}(X)| \leq \mathbb{E}(|X|)$ .
3. Si  $X$  est une variable aléatoire constante,  $X \equiv a$ , alors  $\mathbb{E}(X) = a$ .
4. (Théorème de transfert). Notons  $\{x_1, \dots, x_m\}$  l'ensemble des valeurs prises par la variable aléatoire réelle  $X$ , et soit  $f$  une application définie sur  $X(\Omega)$ , alors :

$$\mathbb{E}(f(X)) = \sum_{k=1}^m f(x_k) \mathbb{P}^X(\{x_k\}).$$

En particulier, si  $f \equiv \text{Id}$ , on obtient :

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

*Démonstration.*

1. Par définition de l'espérance,

$$\begin{aligned}\mathbb{E}(aX + bY) &= \sum_{i=1}^n (aX(\omega_i) + bY(\omega_i))\mathbb{P}(\{\omega_i\}) \\ &= a \sum_{i=1}^n X(\omega_i)\mathbb{P}(\{\omega_i\}) + b \sum_{i=1}^n Y(\omega_i)\mathbb{P}(\{\omega_i\}) \\ &= a\mathbb{E}(X) + b\mathbb{E}(Y).\end{aligned}$$

2. Supposons que  $X \leq Y$ . Alors, comme la probabilité  $\mathbb{P}$  est à valeur positive,

$$\mathbb{E}(X) = \sum_{i=1}^n X(\omega_i)\mathbb{P}(\{\omega_i\}) \leq \sum_{i=1}^n Y(\omega_i)\mathbb{P}(\{\omega_i\}) = \mathbb{E}(Y).$$

En appliquant ceci avec  $-X \leq |X|$  et  $X \leq |X|$ , et en utilisant la linéarité de l'espérance, on obtient,  $\mathbb{E}(X) \leq \mathbb{E}(|X|)$  et  $-\mathbb{E}(X) \leq \mathbb{E}(|X|)$ , d'où on déduit,  $|\mathbb{E}(X)| \leq \mathbb{E}(|X|)$ .

3. Par linéarité de l'espérance, il suffit de montrer que si  $X \equiv 1$ , alors  $\mathbb{E}(1) = 1$ .

$$\mathbb{E}(1) = \sum_{i=1}^n X(\omega_i) \mathbb{P}(\{\omega_i\}) = \sum_{i=1}^n \mathbb{P}(\{\omega_i\}) = \mathbb{P}(\Omega) = 1.$$

4. Démontrons ce point pour  $f \equiv \text{Id}$ . Par définition, on a  $\mathbb{E}(X) = \sum_{i=1}^n X(\omega_i)\mathbb{P}(\{\omega_i\})$ . Comme  $X$  prend les valeurs  $\{x_1, \dots, x_m\}$ , on peut réécrire ceci sous la forme :

$$\begin{aligned}\mathbb{E}(X) &= \sum_{k=1}^m \sum_{\substack{1 \leq i \leq n \\ X(\omega_i) = x_k}} X(\omega_i) \mathbb{P}(\{\omega_i\}) \\ &= \sum_{k=1}^m x_k \sum_{\substack{1 \leq i \leq n \\ X(\omega_i) = x_k}} \mathbb{P}(\{\omega_i\}) \\ &= \sum_{k=1}^m x_k \mathbb{P}(\{\omega \in \Omega : X(\omega) = x_k\}) = \sum_{k=1}^m x_k \mathbb{P}^X(\{x_k\}).\end{aligned}$$

□

### • Cas où l'espace des états est quelconque et la variable aléatoire discrète

Soit  $(\Omega, \mathcal{A}, \mathbb{P})$  un espace probabilisé et  $X$  une variable aléatoire discrète définie sur  $\Omega$ , à valeurs dans  $X(\Omega) = \{x_j : j \in J\}$ , où  $J$  est une partie non vide, finie ou dénombrable de  $\mathbb{N}$ .

#### Définition.

- La variable aléatoire discrète  $X$  est dite *intégrable*, si la série de terme général  $|x_j|p_j$  converge.
- Si  $X$  est une variable aléatoire discrète intégrable, on définit son *espérance*, notée  $\mathbb{E}(X)$ , par :

$$\mathbb{E}(X) = \sum_{j \in J} x_j \mathbb{P}^X(\{x_j\}).$$

*Remarque.*

- Lorsque l'univers  $\Omega$  est fini, toute variable aléatoire définie sur  $\Omega$  est intégrable et la définition coïncide avec celle donnée précédemment.
- Si  $X$  n'est pas intégrable, elle n'admet pas d'espérance.
- L'espérance de  $X$ , lorsqu'elle existe, ne dépend que de la loi de  $X$ , c'est-à-dire des couples  $\{(x_j, \mathbb{P}^X(\{x_j\}))\}$ ,  $j \in J$ , et représente le barycentre de l'ensemble de ces couples.

On retrouve les propriétés vues lorsque l'espace des états est fini, mais elles sont souvent plus difficile à montrer dans ce cas.

### Propriété 6.

1. Une variable aléatoire discrète  $X$  est intégrable si et seulement si  $|X|$  est intégrable.
2. Si la variable aléatoire  $X$  est bornée, alors elle est intégrable.
3. (Linéarité). Si  $X$  et  $Y$  sont deux variables aléatoires discrètes intégrables, et si  $a, b$  sont deux constantes réelles, alors  $aX + bY$  est intégrable et on a :

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

4. (Monotonie). Si  $X$  et  $Y$  sont deux variables aléatoires discrètes intégrables telles que  $X \leq Y$ , alors  $\mathbb{E}(X) \leq \mathbb{E}(Y)$ . En particulier si  $X$  est intégrable, alors  $|\mathbb{E}(X)| \leq \mathbb{E}(|X|)$ .
5. Si  $X$  est une variable aléatoire constante,  $X \equiv a$ , alors  $X$  est intégrable et  $\mathbb{E}(X) = a$ .
6. (Théorème de transfert). Si  $X$  est une variable aléatoire discrète et si  $f$  est une application définie sur  $X(\Omega)$  telle que la série de terme général  $|f(x_j)|\mathbb{P}^X(\{x_j\})$  converge, alors  $f(X)$  est une variable aléatoire discrète intégrable et

$$\mathbb{E}(f(X)) = \sum_{j \in J} f(x_j)\mathbb{P}^X(\{x_j\}).$$

**Définition.** Soit  $X$  une variable aléatoire discrète intégrable. Si  $X$  est d'espérance nulle, on dit que  $X$  est centrée.

### EXERCICES

EXERCICE 4.8. Soit  $X$  une variable aléatoire discrète intégrable. Montrer que la variable  $X - \mathbb{E}(X)$  est centrée.

EXERCICE 4.9. Calculer, si elle existe, l'espérance des variables aléatoires ayant pour loi :

1. loi uniforme sur  $\{1, \dots, n\}$ ,
2. loi de Bernoulli de paramètre  $p$ . Soit  $A \in \mathcal{A}$  un évènement de  $\Omega$ , déduire que

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

3. loi binomiale de paramètres  $n$  et  $p$ ,
4. loi géométrique de paramètre  $p$ ,  $0 < p < 1$ ,
5. loi de Poisson de paramètre  $\theta > 0$ .

### 4.3.4 Variance, moments d'ordres supérieurs

Soient  $(\Omega, \mathcal{A}, \mathbb{P})$  un espace probabilisé et  $X$  une variable aléatoire discrète définie sur  $\Omega$ , à valeurs dans  $\{x_j : j \in J\}$ , où  $J$  est une partie non-vide, finie ou dénombrable de  $\mathbb{N}$ .

**Définition.** La variable aléatoire  $X$  est dite de *carré intégrable*, si  $X^2$  admet une espérance. La quantité  $\mathbb{E}(X^2)$  est alors bien définie et est appelée *moment d'ordre 2* de  $X$ .

*Remarque.*

1. Grâce au théorème de transfert, la variable aléatoire  $X$  est de carré intégrable si et seulement si la série de terme général  $x_j^2 \mathbb{P}^X(\{x_j\})$  converge.
2. Si la variable aléatoire  $X$  est de carré intégrable, alors elle est intégrable.

En effet, supposons que la variable aléatoire  $X$  prenne les valeurs  $\{x_i : i \in \mathbb{N}\}$ . Fixons  $n \in \mathbb{N}^*$ , et notons  $I_n$  l'ensemble des entiers compris entre 0 et  $n$ . Alors,

$$\begin{aligned} \sum_{i \in I_n} |x_i| \mathbb{P}^X(\{x_i\}) &= \sum_{\{i \in I_n : |x_i| \leq 1\}} |x_i| \mathbb{P}^X(\{x_i\}) + \sum_{\{i \in I_n : |x_i| > 1\}} |x_i| \mathbb{P}^X(\{x_i\}) \\ &\leq \sum_{\{i \in I_n : |x_i| \leq 1\}} 1 \cdot \mathbb{P}^X(\{x_i\}) + \sum_{\{i \in I_n : |x_i| > 1\}} x_i^2 \mathbb{P}^X(\{x_i\}) \\ &\leq 1 + \sum_{i \in I_n} x_i^2 \mathbb{P}^X(\{x_i\}). \end{aligned}$$

On conclut en utilisant le critère de comparaison des séries à termes positifs.

3. Soit  $\lambda \in \mathbb{R}$ . Si la variable aléatoire  $X$  est de carré intégrable, il en est de même pour les variables aléatoires  $(X + \lambda)$  et  $(|X| + \lambda)$ .

**Propriété 7.** Si  $X$  est une variable aléatoire discrète de carré intégrable, alors :

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

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

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

Ainsi,  $f$  est un polynôme de degré 2, positif ou nul. Son discriminant est donc négatif ou nul, ce qui implique le résultat.  $\square$

**Définition.** Si  $X$  est une variable aléatoire discrète de carré intégrable, on définit sa *variance*, notée  $\text{Var}(X)$ , par :

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

L'*écart-type*, noté  $\sigma(X)$ , est défini par  $\sigma(X) = \sqrt{\text{Var}(X)}$ .

*Remarque.*

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

**Propriété 8.** Soit  $X$  une variable aléatoire discrète de carré intégrable, alors :

1.  $\text{Var}(X) \geq 0$ ,

2.  $\forall a \in \mathbb{R}, \text{Var}(X + a) = \text{Var}(X),$
3.  $\forall a \in \mathbb{R}, \text{Var}(aX) = a^2 \text{Var}(X),$
4.  $\text{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2,$
5.  $\text{Var}(X) = 0 \Leftrightarrow \forall j \in J \text{ tel que } \mathbb{P}^X(\{x_j\}) > 0, x_j = \mathbb{E}[X].$

*Démonstration.*

1. Découle de la propriété de monotonie de l'espérance et du fait que  $(X - \mathbb{E}(X))^2 \geq 0.$
2. 3. et 4. En utilisant la définition de la variance et la linéarité de l'espérance, on a :

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

5. D'après le théorème de transfert,

$$\text{Var}(X) = \sum_{j \in J} (x_j - \mathbb{E}(X))^2 \mathbb{P}^X(\{x_j\}).$$

C'est une somme de termes positifs. Ainsi, cette somme est nulle si et seulement si chacun de ses termes l'est, c'est-à-dire si et seulement si

$$\forall j \in J, \text{ tel que } \mathbb{P}^X(\{x_j\}) > 0, x_j = \mathbb{E}(X).$$

□

**Définition.** Soit  $X$  une variable aléatoire discrète de carré intégrable. Si  $X$  est de variance égale à 1, on dit que  $X$  est *réduite*.

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

*Remarque.* D'après le théorème de transfert, la variable aléatoire  $X^n$  est intégrable si et seulement si la série de terme général  $|x_j|^n \mathbb{P}^X(\{x_j\})$  converge.

## EXERCICES

EXERCICE 4.10. Soit  $X$  une variable aléatoire discrète de carré intégrable et de variance non nulle. Montrer que la variable aléatoire  $\frac{X}{\sigma(X)}$  est réduite, et que la variable aléatoire  $\frac{X - \mathbb{E}(X)}{\sigma(X)}$  est centrée réduite.

EXERCICE 4.11. Calculer, si elle existe, la variance de chacune des lois discrètes classiques.

### 4.3.5 Inégalité de Markov et de Bienaymé Tchebychev

Voici deux inégalités classiques.

**Proposition 4.5** (Inégalité de Markov). *Soit  $X$  une variable aléatoire admettant un moment d'ordre  $n \geq 1$ . Alors,*

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



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

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

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

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

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

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

□

**Proposition 4.6** (Inégalité de Bienaymé-Tchebychev). *Soit  $X$  une variable aléatoire discrète de carré intégrable. Alors,*

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

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

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

□

## 4.4 Vecteurs aléatoires discrets

Soit  $(\Omega, \mathcal{A}, \mathbb{P})$  un espace probabilitisé, et soient  $X, Y$  deux variables aléatoires définies sur  $\Omega$ . Alors la loi de probabilité de  $X$  et  $Y$  encode toute l'information pour chacune des variables aléatoires, par contre, elle n'encode aucune information sur les propriétés relativement l'une à l'autre. Une façon de résoudre ce problème est de considérer  $X$  et  $Y$  non pas comme deux variables aléatoires, mais comme les composantes d'un *vecteur aléatoire*  $(X, Y)$  prenant ses valeurs dans  $\mathbb{R}^2$ .

### 4.4.1 Définition et lois des vecteurs aléatoires

Soit  $(\Omega, \mathcal{A}, \mathbb{P})$  un espace probabilitisé.

**Définition.** Soient  $X, Y$  deux variables aléatoires définies sur  $\Omega$ . Le couple aléatoire  $Z = (X, Y)$  est dit *discret* si chacune des variables aléatoires  $X$  et  $Y$  est discrète. Plus généralement, soient  $X_1, \dots, X_n$  des variables aléatoires définies sur  $\Omega$ . Le vecteur aléatoire  $\mathbf{X} = (X_1, \dots, X_n)$  est dit *discret*, si chacune des variables aléatoires  $X_1, \dots, X_n$  l'est.

*Notations.* Pour un couple de variable aléatoire  $(X, Y)$ , on note :

$$E = X(\Omega), F = Y(\Omega).$$

Pour un vecteur aléatoire  $\mathbf{X} = (X_1, \dots, X_n)$ , on note :

$$\forall i \in \{1, \dots, n\}, E_i = X_i(\Omega).$$

Le vecteur  $\mathbf{X} = (X_1, \dots, X_n)$  est donc un vecteur aléatoire défini sur  $\Omega$ , à valeurs dans l'ensemble discret  $E_1 \times \dots \times E_n$ .

**Proposition 4.7** (Définitions).

1. Loi des vecteurs aléatoires

— Si  $Z = (X, Y)$  est un couple aléatoire discret défini sur  $\Omega$ , alors la loi de probabilité de  $Z$  est caractérisée par la donnée des nombres

$$(\mathbb{P}^Z(\{(x, y)\}))_{(x, y) \in E \times F}, \text{ définis par } \forall (x, y) \in E \times F :$$

$$\begin{aligned} \mathbb{P}^Z(\{(x, y)\}) &= \mathbb{P}(\{Z = (x, y)\}) = \mathbb{P}(\{X = x\} \cap \{Y = y\}) \\ &= \mathbb{P}(\{\omega \in \Omega : X(\omega) = x \text{ et } Y(\omega) = y\}). \end{aligned}$$

— Si  $\mathbf{X} = (X_1, \dots, X_n)$  est un vecteur aléatoire discret défini sur  $\Omega$ . Alors la loi de probabilité de  $\mathbf{X}$  est caractérisée par la donnée des nombres

$$(\mathbb{P}^{\mathbf{X}}(\{(x_1, \dots, x_n)\}))_{(x_1, \dots, x_n) \in E_1 \times \dots \times E_n},$$

définis pour  $(x_1, \dots, x_n)$  dans  $E_1 \times \dots \times E_n$  par

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

2. Lois marginales d'un couple aléatoire. Connaissant la loi du couple aléatoire  $Z = (X, Y)$ , on retrouve la loi des variables aléatoires  $X$  et  $Y$ , dites lois marginales de  $Z$ , grâce aux formules suivantes :

$$\forall x \in E, \mathbb{P}^X(\{x\}) = \mathbb{P}(\{X = x\}) = \sum_{y \in F} \mathbb{P}^Z(\{(x, y)\}) = \sum_{y \in F} \mathbb{P}(\{X = x\} \cap \{Y = y\}),$$

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

3. Loi conditionnelle. Soit  $x \in E$  tel que  $\mathbb{P}(\{X = x\}) > 0$ . La loi conditionnelle de  $Y$  sachant que  $X$  prend la valeur  $x$ , est caractérisée par la donnée des nombres :

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

*Remarque.*

— La loi du vecteur aléatoire  $Z$  permet de connaître la loi des variables aléatoires  $X$  et  $Y$ , mais la réciproque est fautive ! La connaissance de chacune des lois  $X$  et  $Y$  n'entraîne pas la connaissance de la loi du couple. Voir l'exemple ci-dessous.

— Comme conséquence du point 3. on sait que si  $x \in E$  et  $y \in F$  sont tels que  $\mathbb{P}(\{X = x\}) > 0$  et  $\mathbb{P}(\{Y = y\}) > 0$ , alors :

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

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

EXEMPLE 4.8. Soit  $Z = (X, Y)$  un couple aléatoire à valeurs dans

$$\{(-1, -1), (-1, 1), (1, -1), (1, 1)\},$$

respectivement avec les probabilités  $\frac{1}{2} - p, p, p, \frac{1}{2} - p$ , où  $0 \leq p \leq \frac{1}{2}$ . Calculons les lois marginales du couple aléatoire  $Z$  et la loi conditionnelle de  $Y$  sachant  $\{X = 1\}$ . Remarquons que  $X$  et  $Y$  sont à valeurs dans  $\{-1, 1\}$ . De la Proposition 4.7, nous déduisons :

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

donc  $X$  et  $Y$  suivent une loi uniforme sur  $\{-1, 1\}$ . D'après la Proposition 4.7, comme  $\mathbb{P}(X = 1) = \frac{1}{2}$  est strictement positif, la loi conditionnelle est bien définie et est donnée par :

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

#### 4.4.2 Espérance, covariance, matrice de covariance

Soit  $(\Omega, \mathcal{A}, \mathbb{P})$  un espace probabilisé. Toutes les variables aléatoires et vecteurs aléatoires que l'on considère sont définis sur  $\Omega$ .

##### Proposition 4.8.

- (Théorème de transfert pour les couples). Si  $Z = (X, Y)$  est un couple aléatoire discret et que  $f$  est une application réelle définie sur  $E \times F$ , telle que la série

$$\sum_{(x,y) \in E \times F} |f(x, y)| \mathbb{P}^Z(\{(x, y)\})$$

converge, alors  $f(X, Y)$  est intégrable et :

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

- (Théorème de transfert pour les vecteurs). Si  $\mathbf{X} = (X_1, \dots, X_n)$  est un vecteur aléatoire discret et que  $f$  est une application réelle définie sur  $E_1 \times \dots \times E_n$ , telle que la série

$$\sum_{(x_1, \dots, x_n) \in E_1 \times \dots \times E_n} |f(x_1, \dots, x_n)| \mathbb{P}^{\mathbf{X}}(\{(x_1, \dots, x_n)\})$$

converge, alors  $f(x_1, \dots, x_n)$  est intégrable et :

$$\mathbb{E}(f(X_1, \dots, X_n)) = \sum_{(x_1, \dots, x_n) \in E_1 \times \dots \times E_n} f(x_1, \dots, x_n) \mathbb{P}^{\mathbf{X}}(\{(x_1, \dots, x_n)\}).$$

*Remarque.* Avec le théorème de transfert pour les couples, nous avons les outils nécessaires pour démontrer la propriété de linéarité de l'espérance.

EXEMPLE 4.9. Vérifions sur l'exemple que  $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$ . D'après le théorème de transfert :

$$\begin{aligned}\mathbb{E}(X + Y) &= (1 + 1)\mathbb{P}(Z = (1, 1)) + (1 - 1)\mathbb{P}(Z = (1, -1)) + \\ &\quad + (-1 + 1)\mathbb{P}(Z = (-1, 1)) + (-1 - 1)\mathbb{P}(Z = (-1, -1)) \\ &= 2\left(\frac{1}{2} - p\right) - 2\left(\frac{1}{2} - p\right) \\ &= 0.\end{aligned}$$

D'autre part, nous avons montré que  $X$  et  $Y$  suivent une loi uniforme sur  $\{-1, 1\}$ , donc :

$$\mathbb{E}(X) = 1\frac{1}{2} - 1\frac{1}{2} = 0 = \mathbb{E}(Y),$$

d'où  $\mathbb{E}(X) + \mathbb{E}(Y) = 0$ .

**Proposition 4.9** (Inégalité de Cauchy-Schwarz). *Soient  $X$  et  $Y$  deux variables aléatoires discrètes. Si  $X$  et  $Y$  sont de carré intégrable, alors  $XY$  est intégrable, et :*

$$|\mathbb{E}(XY)| \leq \mathbb{E}(X^2)^{\frac{1}{2}} \mathbb{E}(Y^2)^{\frac{1}{2}}.$$

*Démonstration.* L'idée de la preuve est la même que pour montrer qu'une variable aléatoire de carré intégrable est intégrable.  $\square$

**Définition.** Soient  $X, Y$  deux variables aléatoires discrètes de carré intégrable, alors on définit la *covariance de  $X$  et  $Y$* , notée  $\text{Cov}(X, Y)$ , par :

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

EXEMPLE 4.10. Calculons la covariance des variables aléatoires  $X$  et  $Y$  de l'exemple. Nous avons déjà calculé  $\mathbb{E}(X) = \mathbb{E}(Y) = 0$ . D'après le théorème de transfert, nous savons que :

$$\begin{aligned}\mathbb{E}(XY) &= 1.1\mathbb{P}(Z = (1, 1)) + 1(-1)\mathbb{P}(Z = (1, -1)) \\ &\quad + (-1)(1)\mathbb{P}(Z = (-1, 1)) + (-1)(-1)\mathbb{P}(Z = (-1, -1)) \\ &= 2\left(\frac{1}{2} - p\right) - 2p = 1 - 4p.\end{aligned}$$

Ainsi,

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

**Propriété 9.** *Soient  $X, Y$  deux variables aléatoires discrètes, de carré intégrable.*

1.  $\text{Cov}(X, X) = \text{Var}(X)$  et  $\text{Cov}(X, Y) = \text{Cov}(Y, X)$ .
2. Si  $a, b, c, d$ , sont des constantes réelles, alors :

$$\text{Cov}(aX + b, cY + d) = ac \text{Cov}(X, Y).$$

3. La variance et la covariance sont reliées par l'égalité :

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

4. La covariance vérifie l'inégalité :

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

*Démonstration.*

2. Par définition de la covariance, et en utilisant la linéarité de l'espérance,

$$\begin{aligned} \text{Cov}(aX + b, cY + d) &= \mathbb{E}[(aX + b)(cY + d)] - \mathbb{E}[aX + b]\mathbb{E}[cY + d] \\ &= ac\mathbb{E}(XY) + bc\mathbb{E}(Y) + ad\mathbb{E}(X) + bd - ac\mathbb{E}(X)\mathbb{E}(Y) - ad\mathbb{E}(X) - bc\mathbb{E}(Y) - bd \\ &= ac[\mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)] = ac \text{Cov}(X, Y). \end{aligned}$$

La covariance est une forme bilinéaire symétrique semi-définie positive sur les variables aléatoires de carré intégrable.

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

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

4. Il suffit d'appliquer l'inégalité de Cauchy-Schwarz avec  $X - \mathbb{E}(X)$  et  $Y - \mathbb{E}(Y)$ . □

**Définition.** Soient  $X, Y$  deux variables aléatoires discrètes de carré intégrable, de variances non nulles. Alors, le *coefficient de corrélation de  $X$  et  $Y$* , noté  $\rho(X, Y)$ , est défini par :

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

*Remarque.* Le coefficient de corrélation est sans unité et est très utilisé en statistique. Comme conséquence de la Propriété 4. de la covariance, nous savons que :

$$-1 \leq \rho(X, Y) \leq 1.$$

**Définition.** Si  $\mathbf{X} = (X_1, \dots, X_n)$  est un vecteur aléatoire discret, telle que chacune des composantes est une variable aléatoire de carré intégrable, alors on définit la *matrice de covariance du vecteur  $\mathbf{X}$* , notée  $V(\mathbf{X})$ , par ses coefficients :

$$\forall (i, j) \in \{1, \dots, n\}^2, \quad (V(\mathbf{X}))_{i,j} = \text{Cov}(X_i, X_j).$$

**Propriété 10.**

1. La matrice de covariance  $V(\mathbf{X})$  est une matrice réelle symétrique, dont la diagonale est formée des variances des variables aléatoires  $X_1, \dots, X_n$ .

2. 
$$\text{Var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \text{Var}(X_i) + 2 \sum_{1 \leq i < j \leq n} \text{Cov}(X_i, X_j).$$

*Démonstration.* Le point 1. est une conséquence directe de la Propriété 9. Le point 2. se démontre par récurrence sur le nombre de variables aléatoires considérées. □

### 4.4.3 Variables aléatoires indépendantes

Soit  $(\Omega, \mathcal{A}, \mathbb{P})$  un espace probabilisé. Toutes les variables aléatoires et tous les vecteurs aléatoires que l'on considère sont définis sur  $\Omega$ .

#### Définition.

- Soit  $(X, Y)$  un couple de variables aléatoires discrètes. On dit que les variables aléatoires  $X$  et  $Y$  sont *indépendantes*, si pour tout sous-ensemble  $A$  de  $E$  et tout sous-ensemble  $B$  de  $F$ , tels que  $\{X \in A\} \in \mathcal{A}$  et  $\{Y \in B\} \in \mathcal{A}$  :

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

- Soit  $\mathbf{X} = (X_1, \dots, X_n)$  un vecteur aléatoire discret. On dit que  $(X_1, \dots, X_n)$  est un *n-uplet de variables aléatoires indépendantes*, ou un *vecteur aléatoire indépendant* si, pour tout sous-ensemble  $(A_1, \dots, A_n)$  de  $E_1 \times \dots \times E_n$  tel que pour tout  $i \in \{1, \dots, n\}$ ,  $\{X_i \in A_i\} \in \mathcal{A}$  :

$$\mathbb{P}(\{X_1 \in A_1\} \cap \dots \cap \{X_n \in A_n\}) = \mathbb{P}(\{X_1 \in A_1\}) \dots \mathbb{P}(\{X_n \in A_n\}).$$

**Proposition 4.10.** *Soit  $(X, Y)$  un couple de variables aléatoires discrètes. Les assertions suivantes sont équivalentes.*

1. Les variables aléatoires  $X$  et  $Y$  sont indépendantes.
2.  $\forall (x, y) \in E \times F$ ,  $\mathbb{P}(\{X = x\} \cap \{Y = y\}) = \mathbb{P}(\{X = x\})\mathbb{P}(\{Y = y\})$ .
3.  $\forall (x, y) \in E \times F$ ,  $\mathbb{P}(\{X = x\}) > 0 \Rightarrow \mathbb{P}_{\{X=x\}}(\{Y = y\}) = \mathbb{P}(\{Y = y\})$ .
4.  $\forall (x, y) \in E \times F$ ,  $\mathbb{P}(\{Y = y\}) > 0 \Rightarrow \mathbb{P}_{\{Y=y\}}(\{X = x\}) = \mathbb{P}(\{X = x\})$ .
5. Pour toutes fonctions  $f$  et  $g$ , respectivement définies sur  $E$  et  $F$ , telles que  $f(X)$  et  $g(Y)$  soient intégrables et telles que le produit  $f(X)g(Y)$  est intégrable, et on a :

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

Soit  $(X_1, \dots, X_n)$  un vecteur aléatoire discret. Les assertions suivantes sont équivalentes.

1.  $(X_1, \dots, X_n)$  est un *n-uplet de variables aléatoires indépendantes*.
2.  $\forall (x_1, \dots, x_n) \in E_1 \times \dots \times E_n$ ,

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

3. Pour toutes fonctions  $f_1, \dots, f_n$  respectivement définies sur  $E_1, \dots, E_n$ , telles que les variables  $f_1(X_1), \dots, f_n(X_n)$  soient intégrables et telles que le produit  $f_1(X_1) \dots f_n(X_n)$  est intégrable, on a :

$$\mathbb{E}[f_1(X_1) \dots f_n(X_n)] = \mathbb{E}[f_1(X_1)] \dots \mathbb{E}[f_n(X_n)].$$

*Remarque 4.1.* On déduit de la Proposition 4.10 le fait suivant : si les variables aléatoires  $X$  et  $Y$  sont indépendantes, alors pour toutes fonctions  $f$  et  $g$  suffisamment régulières, les variables aléatoires  $f(X)$  et  $g(Y)$  sont aussi indépendantes.

**EXEMPLE 4.11.** Étudions l'indépendance des variables aléatoires  $X$  et  $Y$  de l'exemple. Les variables aléatoires  $X$  et  $Y$  sont indépendantes, si et seulement si :

$$\begin{aligned} \mathbb{P}(\{X = 1\} \cap \{Y = 1\}) &= \mathbb{P}(X = 1)\mathbb{P}(Y = 1), \text{ et} \\ \mathbb{P}(\{X = 1\} \cap \{Y = -1\}) &= \mathbb{P}(X = 1)\mathbb{P}(Y = -1), \text{ et} \\ \mathbb{P}(\{X = -1\} \cap \{Y = 1\}) &= \mathbb{P}(X = -1)\mathbb{P}(Y = 1), \text{ et} \\ \mathbb{P}(\{X = -1\} \cap \{Y = -1\}) &= \mathbb{P}(X = -1)\mathbb{P}(Y = -1), \end{aligned}$$

$$\Leftrightarrow \frac{1}{2} - p = \frac{1}{4} \text{ et } p = \frac{1}{4} \Leftrightarrow p = \frac{1}{4}.$$

**Proposition 4.11.**

— Soient  $X$  et  $Y$  deux variables aléatoires discrètes de carré intégrable. Alors, si  $X$  et  $Y$  sont indépendantes,

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

— Soit  $(X_1, \dots, X_n)$  un  $n$ -uplet de variables aléatoires discrètes de carré intégrable et indépendantes, alors :

1.  $\mathbb{E}\left(\prod_{i=1}^n X_i\right) = \prod_{i=1}^n \mathbb{E}(X_i)$ ,
2.  $\forall (i, j) \in \{1, \dots, n\}^2, i \neq j, \text{Cov}(X_i, X_j) = 0$ . Autrement dit, la matrice de covariance est diagonale.
3.  $\text{Var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \text{Var}(X_i)$ .

*Démonstration.* Montrons la proposition dans le cas des couples de variables aléatoires. D'après le théorème de transfert,

$$\begin{aligned} \mathbb{E}(XY) &= \sum_{(x,y) \in E \times F} xy \mathbb{P}(\{X = x\} \cap \{Y = y\}) \\ &= \sum_{(x,y) \in E \times F} xy \mathbb{P}(X = x) \mathbb{P}(Y = y), \text{ (par indépendance)} \\ &= \left( \sum_{x \in E} x \mathbb{P}(X = x) \right) \left( \sum_{y \in F} y \mathbb{P}(Y = y) \right) \\ &= \mathbb{E}(X) \mathbb{E}(Y). \end{aligned}$$

Ainsi, si  $X$  et  $Y$  sont indépendantes,

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

et

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

□

EXEMPLE 4.12. Recalculons la variance d'une variable aléatoire qui suit une loi binomiale de paramètres  $n$  et  $p$ . Alors  $X$  a même loi que  $\sum_{k=1}^n X_k$ , où les  $(X_k)_{k=1}^n$  sont des variables aléatoires indépendantes de Bernoulli de paramètre  $p$ . Ainsi, par la Proposition 4.11 :

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

*Remarque.* Attention, si  $\text{Cov}(X, Y) = 0$ , cela n'implique pas que les variables aléatoires  $X$  et  $Y$  sont indépendantes. Par exemple, considérons le couple  $Z = (X, Y)$  de loi uniforme sur  $\{(1, 0), (-1, 0), (0, 1), (0, -1)\}$ . D'après la Proposition 4.7, nous avons :

$$\begin{aligned} \mathbb{P}(X = 1) &= \frac{1}{4}, \quad \mathbb{P}(X = -1) = \frac{1}{4}, \quad \mathbb{P}(X = 0) = \frac{1}{2}, \\ \mathbb{P}(Y = 1) &= \frac{1}{4}, \quad \mathbb{P}(Y = 0) = \frac{1}{2}, \quad \mathbb{P}(Y = -1) = \frac{1}{4}. \end{aligned}$$

Ainsi :

$$\mathbb{E}(X) = 0, \mathbb{E}(Y) = 0, \mathbb{E}(XY) = 0, \\ \text{d'où } \text{Cov}(X, Y) = 0.$$

Mais,  $X$  et  $Y$  ne sont pas indépendantes. En effet,

$$\mathbb{P}(X = 1, Y = 0) = \frac{1}{4} \text{ et } \mathbb{P}(X = 1)\mathbb{P}(Y = 0) = \frac{1}{8}, \\ \text{d'où } \mathbb{P}(X = 1, Y = 0) \neq \mathbb{P}(X = 1)\mathbb{P}(Y = 0).$$

**Définition.** Soit  $(X_k)_{k \geq 1}$  une suite de variables aléatoires discrètes. On dit que la suite  $(X_k)_{k \geq 1}$  est une *suite de variables aléatoires discrètes indépendantes*, si pour tout entier  $n$ ,  $(X_1, \dots, X_n)$  est un  $n$ -uplet de variables aléatoires indépendantes.

## EXERCICES

EXERCICE 4.12. Soient  $X$  et  $Y$  deux variables aléatoires indépendantes, à valeurs dans  $\mathbb{N}$ .

1. Calculer la loi de  $Z = X + Y$ .
2. Calculer la loi de  $T = \min(X, Y)$ .

EXERCICE 4.13.

1. Montrer que la loi de la somme de  $n$  variables aléatoires indépendantes de Bernoulli de paramètre  $p$  est une loi binomiale de paramètres  $n$  et  $p$ .
2. Montrer que la loi de la somme de  $n$  variables aléatoires indépendantes de Poisson de paramètres  $\lambda_1, \dots, \lambda_n$ , respectivement est une loi de Poisson de paramètre  $\sum_{i=1}^n \lambda_i$ .
3. Montrer que la variable aléatoire  $T = \min(X, Y)$ , où  $X$  et  $Y$  sont indépendantes de même loi géométrique de paramètres  $p$  ( $p \in ]0, 1[$ ), suit une loi géométrique de paramètre  $1 - (1 - p)^2$ .

EXERCICE 4.14. Soient  $X$  et  $Y$  deux variables aléatoires indépendantes de loi uniforme sur  $\{1, \dots, n\}$ . Calculer  $\mathbb{P}[X = Y]$ .

EXERCICE 4.15. Soient  $X$  et  $Y$  deux variables aléatoires indépendantes, de même loi de Bernoulli de paramètre  $p$ ,  $0 < p < 1$ . On pose  $S = X + Y$  et  $D = XY$ .

1. Déterminer la loi du couple  $(S, D)$ .
2. En déduire les lois marginales de  $S$  et  $D$ .
3. Calculer de trois manières différentes  $\mathbb{E}(S)$  et  $\mathbb{E}(D)$ .
4. Calculer  $\text{Cov}(S, D)$ . Les variables aléatoires  $S$  et  $D$  sont-elles indépendantes ?

## 4.5 Variables aléatoires réelles à densité

### 4.5.1 Définitions

Soit  $(\Omega, \mathcal{A}, \mathbb{P})$  un espace probabilisé et soit  $X$  une variable aléatoire définie sur  $\Omega$  à valeurs dans  $\mathbb{R}$ .



**Définition.** La variable aléatoire  $X$  est dite à *densité*, s'il existe une fonction réelle positive  $p$  n'ayant qu'un nombre fini de points de discontinuité, telle que la fonction de répartition de la loi de probabilité de  $X$  s'écrit :

$$\forall t \in \mathbb{R}, F_X(t) = \int_{-\infty}^t p(x) dx.$$

La fonction  $p$  est alors appelée *densité de la loi de probabilité* de  $X$ .

**Propriété 11.** On a les propriétés suivantes :

1. l'application  $p$  est intégrable sur  $\mathbb{R}$  et  $\int_{-\infty}^{+\infty} p(x) dx = 1$  ;
2. la fonction de répartition  $F_X$  est continue sur  $\mathbb{R}$  ;
3. pour tout intervalle  $I$  de  $\mathbb{R}$ , non réduit à un point :

$$\mathbb{P}[X \in I] = \int_{-\infty}^{+\infty} \mathbb{I}_I(x) p(x) dx = \int_I p(x) dx;$$

4. pour tout réel  $t$ ,  $\mathbb{P}[X = t] = 0$  ;
5. la fonction de répartition est dérivable partout où  $p$  est continue et, en ces points,  $F'_X = p$ .

*Remarque.*

1. Toute application positive  $p$ , continue sauf en un nombre fini de points, pour laquelle l'intégrale  $\int_{-\infty}^{+\infty} p(x) dx$  existe et vaut 1 est une densité de probabilité sur  $\mathbb{R}$  ; c'est-à-dire qu'il existe une variable aléatoire  $X$  telle que, pour tout  $t \in \mathbb{R}$ ,  $F_X(t) = P(X \leq t) = \int_{-\infty}^t p(x) dx$ .
2. Soit  $X$  une variable aléatoire de fonction de répartition  $F_X$ . Si l'application  $F_X$  est continue sur  $\mathbb{R}$  et  $\mathcal{C}^1$ -par morceaux, alors  $X$  est une variable aléatoire à densité, de densité égale à  $F'_X$  partout où elle est dérivable.
3. On parle de **la** densité de la loi d'une variable aléatoire, alors qu'on devrait parler d'**une** densité. En effet, si l'on modifie  $p$  en un nombre fini de points en lui donnant d'autres valeurs positives, on obtient une autre densité.

#### 4.5.2 Espérance et moments d'ordres supérieurs

**Définition.** Soit  $X$  une variable aléatoire réelle à densité, de densité  $p$ . On dit que la variable aléatoire  $X$  est *intégrable* ou encore qu'elle *admet une espérance* si l'intégrale  $\int_{-\infty}^{+\infty} |x| p(x) dx$  existe. Dans ce cas, on définit son *espérance*, notée  $\mathbb{E}(X)$ , par :

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

**Définition.** Soit  $n$  un entier naturel non nul. On dit que la variable aléatoire  $X$  admet un *moment d'ordre  $n$*  si la variable aléatoire  $X^n$  est intégrable. Dans ce cas, son  *$n$ -ième moment* est  $\mathbb{E}(X^n)$  par définition.

**Théorème 4.1** (Théorème de transfert). *Soit  $X$  une variable aléatoire réelle. Alors  $X$  est une variable aléatoire à densité, si et seulement si il existe une fonction positive  $p$  n'ayant qu'un nombre fini de points de discontinuité, telle que pour toute fonction  $\phi$  continue bornée de  $\mathbb{R}$  dans  $\mathbb{R}$ , on a :*

$$\mathbb{E}[\phi(X)] = \int_{\mathbb{R}} \phi(x) p(x) dx.$$

*Dans ce cas, la fonction  $p$  est la densité de la variable aléatoire  $X$ .*

*Remarque.*

1. Lorsque la variable aléatoire  $\phi(X)$  est intégrable, son espérance est encore donnée par  $\int_{\mathbb{R}} \phi(x)p(x) dx$ , même si  $\phi$  n'est pas continue bornée.
2. Ce théorème donne une autre caractérisation des variables aléatoires à densité. On retrouve celle impliquant la fonction de répartition en considérant les fonctions indicatrices des intervalles  $] -\infty, t]$ . Nous ne démontrons pas l'implication dans l'autre sens.
3. Grâce au théorème de transfert, on déduit que  $X$  admet un moment d'ordre  $n$  si et seulement si l'intégrale  $\int_{\mathbb{R}} |x|^n p(x) dx$  existe. Si c'est le cas, alors  $\mathbb{E}(X^n) = \int_{\mathbb{R}} x^n p(x) dx$ . (Voir exercice ci-dessous dans le cas du moment d'ordre 2).

*Remarque.* Les propriétés énoncées pour l'espérance (linéarité, monotonie, etc.), la définition de la variance, l'inégalité de Markov, Bienaymé-Tchebychev, restent valables dans le cas des variables aléatoires à densité.

## EXERCICES

EXERCICE 4.16. Soit  $X$  une variable aléatoire réelle de densité  $p_X$ . Déterminer la densité de la variable aléatoire  $Y$  dans les cas suivants :

1.  $Y = aX + b$ , où  $a$  et  $b$  sont des nombres réels,  $a \neq 0$  ;
2.  $Y = X^2$  ;
3.  $Y = \exp(X)$ .

Résoudre cet exercice de deux manières : en utilisant soit la fonction de répartition, soit le théorème de transfert.

EXERCICE 4.17. Soit  $X$  une variable aléatoire réelle de densité  $p_X$ .

1. À l'aide de la densité de  $X^2$ , traduire le fait que  $X$  admet un moment d'ordre deux. Montrer alors que  $X$  admet un moment d'ordre deux si et seulement si l'intégrale  $\int_{-\infty}^{+\infty} x^2 p_X(x) dx$  existe.
2. On suppose que  $X$  admet un moment d'ordre deux. Montrer, de deux façons différentes, que  $X$  admet une espérance. Indication : pour l'une des méthodes, on utilisera l'inégalité  $|x| \leq 1 + x^2$ .

### 4.5.3 Exemples de variables aléatoires à densité

#### Loi de Gauss ou loi normale

**Définition.** Une variable aléatoire réelle  $X$  suit une loi de *Gauss* ou *loi normale centrée réduite* si elle admet pour densité l'application  $p_X$  définie sur  $\mathbb{R}$  par :

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

Plus généralement, on définit la loi de *Gauss* ou *loi normale, de moyenne  $m$  et de variance  $\sigma^2$* ,  $\sigma^2 > 0$ , notée  $\mathcal{N}(m, \sigma^2)$ , comme la loi d'une variable aléatoire  $X$  de densité définie sur  $\mathbb{R}$  par :

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

## EXERCICES

EXERCICE 4.18. Soit  $X$  une variable aléatoire de loi normale centrée réduite.

1. Démontrer que pour tout  $k \in \mathbb{N}$ , l'intégrale  $\int_0^{+\infty} x^k e^{-\frac{x^2}{2}} dx$  est convergente. En déduire que la variable aléatoire  $X$  admet des moments de tout ordre.
2. Démontrer que l'application  $x \mapsto p_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$  est bien une densité de probabilité sur  $\mathbb{R}$ .
3. Montrer que  $X$  admet 0 comme espérance et 1 comme variance.
4. Montrer que la variable aléatoire  $Y = \sigma X + m$ , suit une loi normale  $\mathcal{N}(m, \sigma^2)$ . En déduire que  $Y$  admet des moments de tout ordre,  $m$  comme espérance et  $\sigma^2$  comme variance.

*Remarque.*

- La loi normale centrée réduite a une grande importance en théorie des probabilités, car elle apparaît comme limite dans certains théorèmes, comme le théorème limite central.
- Nous avons montré que si  $X$  suit une loi normale  $\mathcal{N}(0, 1)$ , alors  $Y = \sigma X + m$  suit une loi  $\mathcal{N}(m, \sigma^2)$ . De manière analogue, si  $Y$  suit une loi  $\mathcal{N}(m, \sigma^2)$ , alors  $X = \frac{Y-m}{\sigma}$  suit une loi  $\mathcal{N}(0, 1)$ .
- La fonction de répartition de  $X$  ne se calcule pas à l'aide des fonctions usuelles. Si on note  $\Pi$  la fonction de répartition de la loi gaussienne centrée réduite, des tables donnent des valeurs approchées de  $\Pi(t)$  pour  $t > 0$  (voir à la fin de cette première partie). En remarquant que :  $\Pi(t) + \Pi(-t) = 1$  on peut en déduire des valeurs approchées de  $\Pi(t)$  pour les valeurs négatives de  $t$ .

EXERCICE 4.19. Soit  $X$  une variable aléatoire gaussienne de moyenne  $m$  et de variance  $\sigma^2$ .

1. Expliciter à l'aide de la fonction  $\Pi$  la quantité  $\mathbb{P}[X \in [m - k\sigma, m + k\sigma]]$ ,  $k > 0$ .
2. En donner une valeur approchée lorsque  $k = 1$ ,  $k = 2$ ,  $k = 3$ ,  $k = 4$ .
3. Montrer que, si  $m - 4\sigma > 0$ , la probabilité que  $X$  soit négative est inférieure à  $10^{-4}$ .
4. En déduire une condition pour que l'on puisse raisonnablement modéliser un phénomène positif (durée de vie, poids, longueur, ...) par une variable gaussienne.

### Loi uniforme

**Définition.** Soit  $I = [a, b]$  un intervalle de  $\mathbb{R}$ , où  $a < b$ . La variable aléatoire  $X$  suit la loi *uniforme sur  $I$*  si elle est à valeurs dans  $I$  et pour tout intervalle  $J$  inclus dans  $I$ , la probabilité que  $X$  appartienne à  $J$  est proportionnelle à la longueur de  $J$ .

Comme la variable aléatoire  $X$  est à valeurs dans  $[a, b]$ , pour tout  $t < a$ ,  $\mathbb{P}(X \leq t) = 0$  et pour tout  $t \geq b$ ,  $\mathbb{P}(X \leq t) = 1$ . Soit maintenant  $t \in [a, b]$ , alors  $\mathbb{P}(X \leq t) = \mathbb{P}(X \in [a, t])$ . Par définition, il existe une constante  $c > 0$  telle que,  $\mathbb{P}(X \in [a, t]) = c(t - a)$ . De plus, comme  $X$  est à valeurs dans  $[a, b]$ ,  $\mathbb{P}(X \in [a, b]) = c(b - a) = 1$ , d'où  $c = \frac{1}{b-a}$ . Ainsi, pour tout  $t \in \mathbb{R}$  :

$$F_X(t) = \begin{cases} 0 & \text{si } t \leq a \\ \frac{t-a}{b-a} & \text{si } t \in [a, b[, \\ 1 & \text{si } t \geq b \end{cases}$$

autrement dit, pour tout  $t \in \mathbb{R}$ ,  $F_X(t) = \int_{-\infty}^t \frac{1}{b-a} \mathbb{I}_{[a,b]}(u) du$ . Étant donné que l'application  $x \mapsto p_X(x) = \frac{1}{b-a} \mathbb{I}_{[a,b]}(x)$  est positive et continue sauf en un nombre fini de points, on déduit que la variable aléatoire  $X$  admet comme densité  $p_X$ . De plus, si  $X$  admet pour densité  $p_X$ , on montre facilement qu'elle suit une loi uniforme au sens de la définition. On en déduit donc le corollaire suivant.

**Corollaire 4.2.** La variable aléatoire  $X$  suit la loi *uniforme sur l'intervalle*  $I = [a, b]$  si et seulement si elle admet pour densité l'application  $p_X$  définie par :

$$\forall x \in \mathbb{R}, p_X(x) = \frac{1}{b-a} \mathbb{I}_{[a,b]}(x).$$

EXERCICE 4.20. Soit  $X$  une variable aléatoire de loi uniforme sur l'intervalle  $[a, b]$ . Montrer que  $X$  admet des moments de tout ordre. Calculer son espérance et sa variance.

### Loi exponentielle

**Définition.** La variable aléatoire  $X$  suit une *loi exponentielle de paramètre*  $\lambda > 0$ , si elle admet pour densité l'application  $p_X$  définie par :

$$\forall x \in \mathbb{R}, p_X(x) = \mathbb{I}_{[0,+\infty[}(x) \lambda e^{-\lambda x}.$$

EXERCICE 4.21. Soit  $X$  une variable aléatoire de loi exponentielle de paramètre  $\lambda > 0$ .

1. Vérifier que l'application  $x \mapsto p_X(x) = \mathbb{I}_{[0,+\infty[}(x) \lambda e^{-\lambda x}$  est bien une densité de probabilité sur  $\mathbb{R}$ .
2. Montrer que pour tout entier naturel  $n$ , la variable aléatoire  $X$  admet un moment d'ordre  $n$  et que :

$$\mathbb{E}[X^n] = \frac{n!}{\lambda^n},$$

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

3. Montrer que sa fonction de répartition est, pour tout  $t \in \mathbb{R}$  :

$$F_X(t) = \mathbb{I}_{[0,+\infty[}(t) (1 - e^{-\lambda t}).$$

EXERCICE 4.22. *Loi sans mémoire*

Soit  $X$  une variable exponentielle de paramètre  $\lambda > 0$ . Montrer que, pour tous nombres réels positifs  $s$  et  $t$  :

$$\mathbb{P}_{\{X>s\}}[X > s + t] = \mathbb{P}[X > t].$$

Cette propriété traduit le fait que les variables aléatoires exponentielles sont sans mémoire. Par exemple, la probabilité qu'un évènement se produise après l'instant  $s + t$  sachant qu'il ne s'est pas produit jusqu'à l'instant  $s$  est la même que la probabilité qu'il se passe après l'instant  $t$ . La propriété de 'sans mémoire' caractérise en fait les lois exponentielles (parmi les lois à densité). Pour cette raison, elles modélisent souvent des durées de vie (d'un atome radioactif par exemple), des temps d'attente. Elle est aussi intimement liée à la loi de Poisson. L'analogie dans le cas discret est la loi géométrique.

## Loi de Cauchy

**Définition.** La variable aléatoire  $X$  suit une *loi de Cauchy standard* si elle admet pour densité l'application  $p_X$  définie par :

$$\forall x \in \mathbb{R}, p_X(x) = \frac{1}{\pi(1+x^2)}.$$

*Remarque.*

- Si la variable aléatoire  $X$  suit une loi de Cauchy standard, elle n'est pas intégrable. En effet,  $|x|p_X(x) \sim \frac{1}{\pi x}$  au voisinage de l'infini, et  $x \mapsto \frac{1}{x}$  n'est pas intégrable à l'infini (critère de Riemann avec  $\alpha = 1$ ). La variable aléatoire  $X$  n'admet donc pas d'espérance ni de moment d'ordre deux.
- La loi de Cauchy apparaît comme la loi d'une variable aléatoire qui est le quotient de variables aléatoires normales centrées indépendantes, de même variance (voir exercice 4.24). L'inverse d'une variable aléatoire qui suit une loi de Cauchy suit également une loi de Cauchy.

EXERCICE 4.23. Soit  $X$  une variable aléatoire de loi de Cauchy standard. Déterminer la fonction de répartition de :

1. la variable aléatoire  $X$ ,
2. la variable aléatoire  $Y = \frac{1}{X^2}$ . La variable aléatoire  $Y$  admet-elle une espérance ?

## 4.6 Couples de variables aléatoires à densité

Soit  $(\Omega, \mathcal{A}, \mathbb{P})$  un espace probabilisé et soit  $(X, Y)$  un couple de variables aléatoires défini sur  $\Omega$  à valeurs dans  $\mathbb{R}^2$ .

### 4.6.1 Définitions

**Définition.** Le couple aléatoire  $(X, Y)$  est dit à *densité*, s'il existe une application réelle positive  $p$  définie sur  $\mathbb{R}^2$ , "suffisamment régulière", telle que pour tout "bon" sous-ensemble  $B$  de  $\mathbb{R}^2$  (*i.e.* tel que  $\{(X, Y) \in B\} \in \mathcal{A}$ ) on ait :

$$\mathbb{P}[(X, Y) \in B] = \int_{\mathbb{R}^2} \mathbb{I}_B(x, y) p(x, y) dx dy = \int_B p(x, y) dx dy.$$

L'application  $p$  est alors appelée la *densité de la loi de probabilité* du couple  $(X, Y)$ .

*Remarque.* Toute application positive  $p$  définie sur  $\mathbb{R}^2$ , "suffisamment régulière", telle que  $\int_{\mathbb{R}^2} p(x, y) dx dy$  existe et vaut 1, est une densité de probabilité sur  $\mathbb{R}^2$ .

Il est difficile de donner une définition mathématique précise d'une densité de probabilité sur  $\mathbb{R}^2$  avec les outils au programme du CAPES. C'est pourquoi nous parlons de fonction "suffisamment régulière". L'exemple classique de densité de probabilité sur  $\mathbb{R}^2$  est celui où l'application  $p$  est continue sur un domaine borné  $D$  et nulle en dehors, comme dans l'exemple qui suit.

EXEMPLE. (**Loi uniforme sur une partie bornée de  $\mathbb{R}^2$** ).

Soit  $D$  une partie bornée de  $\mathbb{R}^2$  dont on peut calculer l'aire  $\mathcal{A}(D)$  supposée strictement positive (par exemple, le disque unité).

Le couple  $(X, Y)$  suit la *loi uniforme sur  $D$*  si, pour toute partie  $A$  de  $\mathbb{R}^2$  incluse dans  $D$  et dont on peut calculer l'aire, la probabilité que  $(X, Y)$  appartienne à  $A$  est proportionnelle à l'aire de  $A$ .

De manière analogue au cas unidimensionnel, on peut montrer que le couple  $(X, Y)$  suit la loi uniforme sur  $D$ , si et seulement si il admet pour densité l'application  $p$  définie, pour tout  $(x, y) \in \mathbb{R}^2$ , par :

$$p(x, y) = \begin{cases} \frac{1}{\mathcal{A}(D)}, & \text{si } (x, y) \in D, \\ 0, & \text{si } (x, y) \notin D. \end{cases}$$

Le théorème de caractérisation de la densité d'une variable aléatoire réelle à densité se généralise de la façon suivante.

**Théorème 4.3** (Théorème de transfert). *Soit  $(X, Y)$  un couple de variables aléatoires réelles. Alors le couple  $(X, Y)$  est un couple de variables aléatoires à densité, si et seulement si il existe une application positive  $p$  définie sur  $\mathbb{R}^2$ , "suffisamment régulière", telle que pour toute fonction  $\psi$  continue bornée de  $\mathbb{R}^2$  dans  $\mathbb{R}$ , on a :*

$$\mathbb{E}[\psi(X, Y)] = \int_{\mathbb{R}^2} \psi(x, y) p(x, y) dx dy.$$

L'application  $p$  est la densité du couple  $(X, Y)$ .

Remarque.

1. En toute rigueur, on devrait là encore parler d'une densité de probabilité du couple  $(X, Y)$ .
2. Si  $\psi : \mathbb{R}^2 \rightarrow \mathbb{R}$  est telle que la variable aléatoire  $\psi(X, Y)$  est intégrable, alors son espérance est égale à  $\int_{\mathbb{R}^2} \psi(x, y) p(x, y) dx dy$  (que  $\psi$  soit continue bornée ou non).

**Proposition 4.12** (Densités marginales). *Connaissant la densité  $p_{X,Y}$  du couple  $(X, Y)$ , on retrouve les densités dites marginales de  $X$  et  $Y$  par :*

$$\text{pour tout } x \in \mathbb{R}, p_X(x) = \int_{\mathbb{R}} p_{X,Y}(x, y) dy,$$

$$\text{pour tout } y \in \mathbb{R}, p_Y(y) = \int_{\mathbb{R}} p_{X,Y}(x, y) dx.$$

EXERCICE. Démontrer la proposition 4.12.

## 4.6.2 Variables aléatoires à densité indépendantes

**Définition.** Soit  $(X, Y)$  un couple de variables aléatoires à densité. On dit que les variables aléatoires  $X$  et  $Y$  sont *indépendantes*, si pour tout sous-ensembles  $A$  et  $B$  de  $\mathbb{R}$  tels que  $\{X \in A\} \in \mathcal{A}$ ,  $\{Y \in B\} \in \mathcal{A}$ , les évènements  $\{X \in A\}$  et  $\{Y \in B\}$  sont indépendants, c'est-à-dire :

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

**Proposition 4.13.** *Soit  $(X, Y)$  un couple de variables aléatoires réelles à densité sur  $\mathbb{R}^2$ . Les assertions suivantes sont équivalentes :*

1. les variables aléatoires  $X$  et  $Y$  sont indépendantes ;

2. le couple  $(X, Y)$  admet une densité de probabilité  $p_{X,Y}$  de la forme :  
 $\forall (x, y) \in \mathbb{R}^2, p_{X,Y}(x, y) = p_X(x)p_Y(y)$ .
3. pour toutes fonctions  $F$  et  $G$  de  $\mathbb{R}$  dans  $\mathbb{R}$  telles que  $F(X)$  et  $G(Y)$  soient intégrables et telles que le produit  $F(X)G(Y)$  est intégrable, on a :

$$\mathbb{E}[F(X)G(Y)] = \mathbb{E}[F(X)]\mathbb{E}[G(Y)];$$

*Remarque.* Lorsque la densité du couple  $(X, Y)$  s'écrit :

$$\forall (x, y) \in \mathbb{R}^2, p_{X,Y}(x, y) = f(x)g(y),$$

il existe une constante  $c$  non nulle telle que  $\forall x \in \mathbb{R}, p_X(x) = cf(x)$  et  $\forall y \in \mathbb{R}, p_Y(y) = \frac{1}{c}g(y)$ .

## EXERCICES

**EXERCICE 4.24.** Soient  $N_1$  et  $N_2$  deux variables gaussiennes centrées réduites indépendantes. Déterminer la fonction de répartition  $F_C$  de la variable aléatoire  $C = \frac{N_1}{|N_2|}$ . En déduire que  $C$  suit une loi de Cauchy standard.

**EXERCICE 4.25.** Soient  $X_1$  et  $X_2$  deux variables aléatoires indépendantes qui suivent une loi exponentielle de paramètres  $\lambda_1$  et  $\lambda_2$  respectivement. Déterminer la loi de  $Z = \min(X_1, X_2)$ .

*Remarque.* En raisonnant par récurrence on peut montrer que si  $X_1, \dots, X_n$  sont  $n$  variables aléatoires indépendantes de loi exponentielle de paramètres  $\lambda_1, \dots, \lambda_n$  respectivement, alors la variable aléatoire  $\min(X_1, \dots, X_n)$  suit une loi exponentielle de paramètre  $\lambda_1 + \dots + \lambda_n$ .

**Proposition 4.14.** Soient  $X$  et  $Y$  deux variables aléatoires réelles indépendantes de densités respectives  $p_X$  et  $p_Y$ . La variable aléatoire  $X + Y$  admet une densité  $p_{X+Y}$  sur  $\mathbb{R}$  égale à :

$$\forall t \in \mathbb{R}, p_{X+Y}(t) = \int_{\mathbb{R}} p_X(s) p_Y(t-s) ds. \quad (4.1)$$

La densité  $p_{X+Y}$  s'appelle le produit de convolution de  $p_X$  et  $p_Y$ .

*Démonstration.* Les variables aléatoires  $X$  et  $Y$  étant indépendantes, la densité du couple  $(X, Y)$  est :  $\forall (x, y) \in \mathbb{R}^2, p_{X,Y}(x, y) = p_X(x)p_Y(y)$ . Posons  $Z = X + Y$  et calculons la fonction de répartition de  $Z$ . Soit  $t \in \mathbb{R}$ . Alors,

$$\begin{aligned} F_Z(t) &= \mathbb{P}(Z \leq t) = \mathbb{P}(X + Y \leq t) \\ &= \int_{\mathbb{R}^2} \mathbb{I}_{\{x+y \leq t\}} p_X(x) p_Y(y) dx dy, \text{ par définition et indépendance} \\ &= \int_{\mathbb{R}} p_X(x) \left( \int_{\mathbb{R}} \mathbb{I}_{\{x+y \leq t\}} p_Y(y) dy \right) dx, \text{ d'après le théorème de Fubini.} \end{aligned}$$

À  $x$  fixé, on effectue le changement de variables  $u = x + y$ . Il s'ensuit :

$$\begin{aligned} F_Z(t) &= \int_{\mathbb{R}} p_X(x) \left( \int_{-\infty}^t p_Y(u-x) du \right) dx, \\ &= \int_{-\infty}^t \left( \int_{\mathbb{R}} p_X(x) p_Y(u-x) dx \right) du, \text{ d'après le théorème de Fubini.} \end{aligned}$$

On en déduit que la variable aléatoire  $X + Y$  admet pour densité (4.1). □

EXERCICES

EXERCICE 4.26. Soient  $X$  et  $Y$  deux variables aléatoires réelles indépendantes de lois gaussiennes de moyennes respectives  $m_1$  et  $m_2$ , et de variances respectives  $\sigma_1^2$  et  $\sigma_2^2$ . Alors  $X + Y$  suit une loi gaussienne de moyenne  $m_1 + m_2$ , et de variance  $\sigma_1^2 + \sigma_2^2$ .

*Remarque.* En raisonnant par récurrence, on peut montrer que si  $X_1, \dots, X_n$  sont  $n$  variables aléatoires gaussiennes indépendantes de moyenne  $m_1, \dots, m_n$  et variance  $\sigma_1^2, \dots, \sigma_n^2$  respectivement, alors  $\sum_{k=1}^n X_k$  suit une loi gaussienne de paramètres  $(\sum_{k=1}^n m_k, \sum_{k=1}^n \sigma_k^2)$ . Ce résultat est indispensable pour le cours de statistique.

EXERCICE 4.27.

1. Soit  $N_1$  une variable aléatoire gaussienne centrée réduite. Montrer que la variable  $Y_1 = N_1^2$  admet pour densité l'application  $p_{Y_1}$  définie par :

$$\forall t \in \mathbb{R}, p_{Y_1}(t) = \mathbb{I}_{[0, +\infty[}(t) \frac{1}{\sqrt{2\pi}} t^{-\frac{1}{2}} e^{-\frac{t}{2}}.$$

On dit que  $Y_1$  suit une *loi du  $\chi^2$  (khi-deux) à 1 degré de liberté*.

*Remarque.* Si  $N_1, \dots, N_d$  sont  $d$  variables gaussiennes centrées réduites indépendantes, alors  $Y_d = \sum_{i=1}^d X_i^2$  suit une *loi du  $\chi^2$  à  $d$  degrés de liberté*, dont la densité  $p_{Y_d}$  est donnée par :

$$\forall t \in \mathbb{R}, p_{Y_d}(t) = \mathbb{I}_{[0, +\infty[}(t) \frac{2^{-\frac{d}{2}}}{\Gamma(\frac{d}{2})} t^{\frac{d}{2}-1} e^{-\frac{t}{2}}.$$

La fonction  $\Gamma$  s'appelle la *fonction gamma* et est définie pour tout  $t \in \mathbb{R}_+^*$  par :

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

Elle satisfait aux propriétés suivantes :

- (a)  $\forall t \in \mathbb{R}_+^*, \Gamma(t + 1) = t \Gamma(t)$ ,
- (b)  $\forall n \in \mathbb{N}^*, \Gamma(n) = (n - 1)!$ ,
- (c)  $\Gamma(\frac{1}{2}) = \sqrt{\pi}$ .

Un des problèmes que nous étudierons en M2 portera sur ce sujet.

2. Soit la variable aléatoire  $T_d = N \sqrt{\frac{d}{Y_d}}$  où  $N$  et  $Y_d$  sont indépendantes. On suppose que  $N$  est une gaussienne centrée réduite et que  $Y_d$  suit la loi du  $\chi^2$  à  $d$  degrés de liberté ( $d \geq 1$ ). Déterminer la densité  $p_{T_d}$  de  $T_d$ . On dit que  $T_d$  suit la *loi de Student à  $d$  degrés de liberté*.

On peut montrer que pour tout  $x$  réel, la limite de  $p_{T_d}(x)$  lorsque  $d$  tend vers  $+\infty$  est la densité d'une gaussienne centrée réduite évaluée en  $x$ .

*Remarque.* Les définitions de covariance, matrice de covariance ainsi que leurs propriétés et l'inégalité de Cauchy-Schwarz, restent valables dans le cas des variables aléatoires à densité.



### Table de Gauss

$$\forall t \in \mathbb{R}, \quad \Pi(t) = \int_{-\infty}^t \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx$$

Valeurs approchées à  $10^{-4}$  près.

$t$	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2,6	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974
2,8	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,9980	0,9981
2,9	0,9981	0,9982	0,9982	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986

Cas des grandes valeurs de  $t$

$t$	3,0	3,1	3,2	3,3	3,4
$\Pi(t)$	0,998650	0,999032	0,999313	0,999517	0,999663
$t$	3,5	3,6	3,8	4,0	4,5
$\Pi(t)$	0,999767	0,999841	0,999928	0,999968	0,999997

## 4.7 Suites de variables aléatoires réelles

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

**Définition.** Soit  $(X_n)_{n \geq 1}$  une suite de variables aléatoires. On dit que la suite  $(X_n)_{n \geq 1}$  converge en probabilité vers la variable aléatoire  $X$  si :

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

Dans ce cas, on note :  $X_n \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} X$ .

**Proposition 4.15.** Soit  $(X_n)_{n \geq 1}$  une suite de variables aléatoires convergeant en probabilité vers  $X$ , et soit  $f$  une fonction continue définie sur  $\mathbb{R}$ .

1. La suite  $(f(X_n))_{n \geq 1}$  converge en probabilité vers  $f(X)$ .
2. Si de plus  $f$  est bornée, on a :  $\lim_{n \rightarrow +\infty} \mathbb{E}[f(X_n)] = \mathbb{E}[f(X)]$ .

### 4.7.1 Loi faible des grands nombres

**Théorème 4.4** (Loi faible des grands nombres.). Soit  $(X_n)_{n \geq 1}$  une suite de variables aléatoires indépendantes et de même loi. On suppose que ces variables sont de carré intégrable, d'espérance commune  $m$ . Alors la suite  $\left(\frac{1}{n} \sum_{k=1}^n X_k\right)_{n \geq 1}$  converge en probabilité vers  $m$ .

*Démonstration.* Notons  $m$  l'espérance et  $\sigma^2$  la variance commune des variables aléatoires  $X_k$ ,  $k \geq 1$ .

Rappelons tout d'abord l'inégalité de Bienaymé-Tchebychev. Si  $Z$  est une variable aléatoire de carré intégrable, alors pour tout  $a > 0$  :

$$\mathbb{P}[|Z - \mathbb{E}[Z]| > a] \leq \frac{\text{Var}(Z)}{a^2}.$$

Soit  $\varepsilon > 0$  et définissons  $Z = \frac{1}{n} \sum_{k=1}^n X_k$ . Alors la variable aléatoire  $Z$  est de carré intégrable car somme finie de variables aléatoires de carré intégrable. De plus :

$$\begin{aligned} \mathbb{E}[Z] &= \mathbb{E}\left[\frac{1}{n} \sum_{k=1}^n X_k\right] = \frac{1}{n} \sum_{k=1}^n \mathbb{E}[X_k] = m, \text{ par linéarité de l'espérance} \\ \text{Var}[Z] &= \text{Var}\left[\frac{1}{n} \sum_{k=1}^n X_k\right] = \frac{1}{n^2} \text{Var}\left[\sum_{k=1}^n X_k\right] \\ &= \frac{1}{n^2} \sum_{k=1}^n \text{Var}[X_k] = \frac{\sigma^2}{n}, \text{ car les variables } X_k \text{ sont indépendantes.} \end{aligned}$$

En appliquant l'inégalité de Bienaymé-Tchebychev à  $Z$ , on obtient :

$$0 \leq \mathbb{P}\left[\left|\left(\frac{1}{n} \sum_{k=1}^n X_k\right) - m\right| > \varepsilon\right] \leq \frac{\sigma^2}{n \varepsilon^2}.$$

D'après le théorème d'encadrement, pour tout  $\varepsilon > 0$  :

$$\lim_{n \rightarrow +\infty} \mathbb{P} \left[ \left| \left( \frac{1}{n} \sum_{k=1}^n X_k \right) - m \right| > \varepsilon \right] = 0,$$

et on en déduit que la suite  $\left( \frac{1}{n} \sum_{k=1}^n X_k \right)_{n \geq 1}$  converge en probabilité vers  $m$ .  $\square$

*Remarque.*

1. On peut démontrer que ce résultat reste vrai si on suppose seulement les variables aléatoires intégrables (au lieu de carré intégrable).
2. La loi faible des grands nombres et la notion de convergence en probabilité ne semblent plus être explicitement au programme du concours du CAPES. Néanmoins, les programmes de classe de première S proposent “un énoncé vulgarisé de la loi des grands nombres : pour une expérience donnée, dans le modèle défini par une loi de probabilité  $\mathbb{P}$ , les distributions des fréquences obtenues sur des séries de taille  $n$  se rapprochent de  $\mathbb{P}$  quand  $n$  devient grand.” Il vaut mieux connaître la théorie pour pouvoir expliquer ce que cela veut dire...
3. Il existe également une version “forte” de la loi des grands nombres.

EXEMPLE. Considérons ce que l'on pourrait appeler un schéma de Bernoulli de paramètres  $\infty$  et  $p$ , *i.e.* une expérience qui consiste à répéter indéfiniment une même expérience admettant deux issues : succès/échec, telle que  $p$  est la probabilité d'obtenir un succès. Si on note  $A$  l'évènement “obtenir un succès”, alors  $\mathbb{P}(A) = p$ . Par exemple, on lance indéfiniment un dé non pipé et on considère l'évènement  $A$  “obtenir un 6”, alors  $p = 1/6$ .

Pour  $k \geq 1$ , on définit  $A_k$  l'évènement “ $A$  est réalisé lors de la  $k$ -ième expérience” et  $X_k = \mathbb{I}_{A_k}$ . Alors la suite  $(X_k)_{k \geq 1}$  est une suite de variables aléatoires indépendantes de même loi de Bernoulli de paramètre  $p$ , de carré intégrable étant donné qu'elles ne prennent que deux valeurs. De plus, pour tout  $k \geq 1$ ,  $\mathbb{E}(X_k) = \mathbb{P}(A_k) = \mathbb{P}(A) = p$ . D'après la loi des grands nombres :

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

Interprétons ce résultat. Pour tout  $N \in \mathbb{N}$ , la variable aléatoire  $\sum_{k=1}^N X_k = \sum_{k=1}^N \mathbb{I}_{A_k} = n_N(A)$

représente le nombre de fois où  $A$  est réalisé lors des  $N$  premières épreuves, et  $\frac{1}{N} \sum_{k=1}^N X_k$  représente donc la fréquence de  $A$  lors des  $N$  premières épreuves. Ainsi la loi des grands nombres peut se réécrire :

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

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

#### 4.7.2 Théorème limite central

Soit  $(X_n)_{n \geq 1}$  une suite de variables aléatoires indépendantes de même loi, de moyenne et variance commune  $m$  et  $\sigma^2$  respectivement. On définit  $S_n = \sum_{k=1}^n X_k$ . Alors d'après la loi des grands

nombres,

$$\frac{S_n - nm}{n} \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} 0.$$

Supposons que l'on divise par quelque chose de plus petit que  $n$ , peut-on alors obtenir une limite? La réponse est oui, si on normalise par une quantité proportionnelle à  $\sqrt{n}$ . Cela précise en particulier la vitesse de convergence dans la loi faible des grands nombres.

**Théorème 4.5** (Théorème limite central). *Soit  $(X_n)_{n \geq 1}$  une suite de variables aléatoires de carré intégrables, indépendantes, de même loi, de moyenne et variance commune  $m$  et  $\sigma^2$  respectivement. On pose  $S_n = \sum_{k=1}^n X_k$ . Alors pour tout réel  $t$  :*

$$\lim_{n \rightarrow +\infty} \mathbb{P} \left[ \frac{S_n - nm}{\sigma \sqrt{n}} \leq t \right] = \lim_{n \rightarrow +\infty} \mathbb{P} \left[ \frac{S_n - \mathbb{E}[S_n]}{\sqrt{\text{Var}[S_n]}} \leq t \right] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^t e^{-\frac{x^2}{2}} dx.$$

En posant  $\tilde{S}_n = \frac{S_n - \mathbb{E}[S_n]}{\sqrt{\text{Var}[S_n]}}$ , cela peut s'écrire :

$$\forall t \in \mathbb{R}, \quad \lim_{n \rightarrow +\infty} F_{\tilde{S}_n}(t) = \Pi(t),$$

où  $\Pi$  est la fonction de répartition de la loi normale centrée réduite.

*Remarque.*

1. On dit que la variable aléatoire  $\tilde{S}_n$  converge en loi vers une variable aléatoire gaussienne centrée réduite. A noter que cette notion n'est pas explicitement au programme du CAPES.
2. De manière équivalente, on a que pour tous réels  $a$  et  $b$ ,  $a \leq b$ ,

$$\begin{aligned} \lim_{n \rightarrow +\infty} \mathbb{P}[a \leq \tilde{S}_n \leq b] &= \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}} dx \\ \Leftrightarrow \lim_{n \rightarrow +\infty} F_{\tilde{S}_n}(b) - F_{\tilde{S}_n}(a) &= \Pi(b) - \Pi(a). \end{aligned}$$

3. Ce théorème a d'abord été démontré lorsque la loi commune des variables est une loi de Bernoulli de paramètre  $p$  (le théorème est dans ce cas connu sous le nom de théorème de Moivre-Laplace). Dans ce cas, la variable  $S_n$  suit une loi binomiale de paramètres  $n$  et  $p$ . L'approximation de cette loi binomiale par la loi de Gauss est déjà très bonne lorsque  $n \geq 10$ , et que  $np$  et  $n(1-p)$  dépassent quelques unités.

Plus précisément dans ce cas, notons  $F_n$  la fonction de répartition de la variable aléatoire centrée réduite  $\tilde{S}_n = \frac{S_n - \mathbb{E}[S_n]}{\sqrt{\text{Var}[S_n]}}$ . Alors, on a :

$$\sup_{x \in \mathbb{R}} |F_n(x) - \Pi(x)| \leq \frac{p^2 + q^2}{\sqrt{npq}},$$

où  $q = 1 - p$ ; (cf. Shiryaev, Probability, Springer).

On a donc vu deux façons d'approcher la loi binomiale : par la loi de Gauss, et par la loi de Poisson.

4. Corrections de continuité (programme BTS).

Si  $a$  et  $b$  sont des nombres entiers compris entre 0 et  $n$  (avec  $a \leq b$ ), la probabilité que la variable  $S_n$  de loi binomiale de paramètres  $n$  et  $p$  soit comprise entre  $a$  et  $b$  est égale à  $\mathbb{P}[a - \varepsilon \leq S_n \leq b + \varepsilon]$  pour tout  $\varepsilon \in ]0, 1[$  puisqu'elle ne prend que des valeurs entières.  $\Pi$

se trouve que l'approximation que l'on obtient (pour les grandes valeurs de  $n$ ) à l'aide du théorème limite central est la meilleure pour  $\varepsilon = \frac{1}{2}$ . On approchera donc  $\mathbb{P}[a \leq S_n \leq b]$  par l'approximation de

$$\mathbb{P}\left[a - \frac{1}{2} \leq S_n \leq b + \frac{1}{2}\right] = \mathbb{P}\left[\frac{a - \frac{1}{2} - np}{\sqrt{npq}} \leq \frac{S_n - \mathbb{E}[S_n]}{\sqrt{\text{Var}[S_n]}} \leq \frac{b + \frac{1}{2} - np}{\sqrt{npq}}\right],$$

c'est-à-dire par  $\Pi\left(\frac{b + \frac{1}{2} - np}{\sqrt{npq}}\right) - \Pi\left(\frac{a - \frac{1}{2} - np}{\sqrt{npq}}\right)$  (cf. Feller W., On the normal approximation to the binomial distribution, *Ann. Math. Statist.*, vol. 16, p. 319–329, 1945).

5. Le nom de “théorème limite central” est une traduction de l'allemand “Zentraler Grenzwertsatz”.

Le terme “Grenzwertsatz” désigne un “théorème limite”, c'est-à-dire un résultat de convergence d'une suite de variables aléatoires. Le nom de théorème limite *central* signifie ici que l'on a affaire à un théorème limite particulièrement important.

On trouve parfois d'autres traductions, comme “théorème central limite” ou “théorème de la limite centrée”.

EXERCICE 4.28 (*D'après le polycopié de cours de P. Priouret.*). Un joueur pense qu'un dé (à six faces) est pipé. Il le lance 720 fois, et obtient le six 150 fois. Quelle conclusion le joueur peut-il en tirer ?

## 4.8 Encore des définitions

**Définition.** Soit  $X$  une variable aléatoire réelle de fonction de répartition  $F$ . On appelle *valeur médiane*, toute valeur  $a$  telle que  $\mathbb{P}(X \geq a) \geq \frac{1}{2}$  et  $\mathbb{P}(X \leq a) \geq \frac{1}{2}$ .

*Remarque.*

1. Dans le cas où  $F$  est continue (par exemple si  $X$  est une variable aléatoire à densité),  $a$  est une valeur médiane si et seulement si  $F(a) = \frac{1}{2}$ .
2. Lorsque  $X$  est une variable aléatoire discrète, il existe une autre définition de *médiane*, utilisée dans l'enseignement secondaire, sur laquelle nous reviendrons au Chapitre 5.

EXERCICE 4.29. Soit  $X$  une variable aléatoire de carré intégrable. Soient  $\phi$  et  $\psi$  les fonctions définies sur  $\mathbb{R}$  par :  $\forall t \in \mathbb{R}, \phi(t) = \mathbb{E}[(X - t)^2]$  et  $\psi(t) = \mathbb{E}[|X - t|]$ .

1. Montrer que la fonction  $\phi$  admet un minimum. En quel point ce minimum est-il atteint ?
2. Montrer que  $\psi$  est minimale en toute valeur médiane de  $X$ .
3. Application. Sur une route, sont disposés  $n$  magasins, aux abscisses  $x_1, \dots, x_n$ . Chaque jour, on doit se rendre dans chacun des magasins ; après avoir visité un magasin, on revient au point de départ. De quel point de la route doit-on partir pour effectuer le plus court déplacement ? On suppose :  $x_1 < x_2 < \dots < x_n$ . On pourra commencer par étudier les cas  $n = 1, n = 2$  puis  $n = 3$ , avant de traiter le cas général.

**Définition.** Soit  $X$  une variable aléatoire réelle de fonction de répartition  $F$ . On définit la fonction *fractile* ou encore *quantile* associée à  $X$  par :

$$\forall u \in ]0, 1[, Q(u) = \inf\{x \in \mathbb{R}; F(x) \geq u\}.$$



# Chapitre 5

## Statistique descriptive

### 5.1 Préambule à la statistique

“*La statistique*” : méthode scientifique qui consiste à observer et à étudier une/plusieurs particularité(s) commune(s) chez un groupe de personnes ou de choses.

“La statistique” est à différencier d’“une statistique”, qui est un nombre calculé à propos d’une population.

#### 5.1.1 Un peu de vocabulaire

◦ *Population* : collection d’objets à étudier ayant des propriétés communes. Terme hérité des premières applications de la statistique qui concernait la démographie.

Exemple : ensemble de parcelles sur lesquelles on mesure un rendement, un groupe d’insectes...

◦ *Individu* : élément de la population étudiée.

Exemple : une des parcelles, un des insectes...

◦ *Variable* ou *caractère* : propriété commune aux individus de la population, que l’on souhaite étudier. Elle peut être

- *qualitative* : couleur de pétales,

- *quantitative* : (numérique). Par exemple la taille, le poids, le volume. On distingue encore les variables

- *continues* : toutes les valeurs d’un intervalle de  $\mathbb{R}$  sont acceptables. Par exemple : le périmètre d’une coquille de moule.

- *discrètes* : seul un nombre discret de valeurs sont possibles. Par exemple : le nombre d’espèces recensées sur une parcelle.

Les valeurs observées pour les variables s’appellent les *données*.

◦ *Échantillon* : partie étudiée de la population.

#### 5.1.2 Collecte de données

La collecte de données (obtention de l’échantillon) est une étape clé, et délicate. Nous ne traitons pas ici des méthodes possibles, mais attirons l’attention sur le fait suivant.

Hypothèse sous-jacente en statistique : l'échantillon d'individus étudié est choisi "au hasard" parmi tous les individus qui auraient pu être choisis.

⇒ Tout mettre en œuvre pour que cela soit vérifié.

### 5.1.3 Deux directions en statistique

#### 1. Statistique descriptive

Elle a pour but de décrire, c'est-à-dire de résumer ou représenter, par des statistiques, les données disponibles quand elles sont nombreuses. Questions typiques :

- (a) Représentation graphique : diagramme en bâtons, diagramme circulaire, voir exercice 5.1.
- (b) Paramètres de position, de dispersion, de relation : voir paragraphe 5.2.
- (c) Régression linéaire : voir paragraphe 5.4.
- (d) Questions liées à des grands jeux de données : hors programme.

#### 2. Statistique inférentielle

Les données ne sont pas considérées comme une information complète, mais une information partielle d'une population infinie. Il est alors naturel de supposer que les données sont des réalisations de variables aléatoires, qui ont une certaine loi de probabilité.

Nécessite des outils mathématiques plus pointus de théorie des probabilités.

Questions typiques :

- (a) Estimation de paramètres.
- (b) Intervalles de confiance.
- (c) Tests d'hypothèse.
- (d) Modélisation : exemple (régression linéaire).

### 5.1.4 Statistique univariée / multivariée

Lorsque l'on observe une seule variable pour les individus de la population, on parle de *statistique univariée*, et de *statistique multivariée* lorsqu'on en observe au moins deux. Pour chacune des catégories, on retrouve les deux directions ci-dessus.

EXEMPLE. Univarié. Population : iris. Variable : longueur des pétales.

Multivarié. Population : iris. Variable 1 : longueur des pétales. Variable 2 : largeur des pétales.

## 5.2 Paramètres de position, dispersion, relation

On considère un échantillon de  $n$  individus et on souhaite étudier une variable (caractère) quantitative de cette population. Pour tout  $i \in \{1, \dots, n\}$ ,  $x_i$  représente la donnée de la variable (caractère) pour le  $i$ -ième individu. Les données sont représentées dans un vecteur  $(x_1, \dots, x_n)$ , qui s'appelle une *série statistique quantitative*. Sans perte de généralité, supposons que  $x_1 \leq x_2 \leq \dots \leq x_n$  (puisque les  $x_i$  sont des nombres réels).

EXEMPLE. On étudie les résultats de 10 étudiants à un test de statistique :

$$x = (2, 5, 5, 8, 10, 12, 12, 12, 15, 20).$$

L'*effectif total* est  $n$ . Dans l'exemple,  $n = 10$ .



Notons  $(z_1, \dots, z_m)$  les valeurs différentes des données de la série. On a donc  $m \leq n$ . L'effectif  $n_j$  d'une valeur  $z_j$  est le nombre de répétitions de  $z_j$  dans la série. Remarquer que  $\sum_{j=1}^m n_j = n$ .

La fréquence  $f_j$  d'une donnée  $z_j$  est :  $f_j = \frac{\text{effectif}(z_j)}{\text{effectif total}} = \frac{n_j}{n}$ .

EXEMPLE. Reprenons l'exemple précédent.

$$\begin{aligned} m &= 7, \\ (z_1, \dots, z_7) &= (2, 5, 8, 10, 12, 15, 20) \\ n_1 &= 1, n_2 = 2, \dots \\ f_1 &= \frac{1}{10}, f_2 = \frac{1}{5}, \dots \end{aligned}$$

Nous souhaitons ensuite étudier une deuxième variable quantitative de cette population. Pour tout  $i \in \{1, \dots, n\}$ ,  $y_i$  représente la donnée de la deuxième variable pour le  $i$ -ième individu.

On appelle *statistique double* la donnée  $\{(x_i, y_i) : i \in \{1, \dots, n\}\}$  des deux variables pour chacun des individus.

EXEMPLE. Considérons les deux séries représentant les résultats de 10 étudiants à l'examen de statistique et d'algèbre, respectivement :

$$\begin{aligned} x &= (2, 5, 5, 8, 10, 12, 12, 12, 15, 20), \\ y &= (3, 4, 6, 7, 8, 10, 12, 14, 16, 19). \end{aligned}$$

**Définition (Paramètres de position).** Chacune des quantités définies ci-dessous se calcule pour la donnée d'une des variables.

— La *moyenne arithmétique* ou simplement la *moyenne*, notée  $\bar{x}$  est :

$$\begin{aligned} \bar{x} &= \frac{1}{n} \sum_{k=1}^n x_k \\ &= \frac{1}{n} \sum_{j=1}^m n_j z_j = \sum_{j=1}^m f_j z_j. \end{aligned}$$

— La *moyenne élaguée* est la moyenne de la série privée de ses valeurs extrêmes lorsque celles-ci semblent aberrantes (à interpréter selon les cas).

— On appelle *médiane* ou *valeur médiane* tout nombre  $a$  tel qu'au moins la moitié de l'effectif de la série soit inférieur ou égal à  $a$ , et au moins la moitié de l'effectif soit supérieur ou égal à  $a$ . Lorsque  $n = 2p + 1$  est impair,  $x_{p+1}$  est la médiane de la série. Lorsque  $n = 2p$  est pair, tout nombre de l'intervalle  $[x_p; x_{p+1}]$  (appelé dans ce cas *intervalle médian* car c'est l'ensemble des valeurs médianes) convient.

**Attention!** Dans le secondaire on utilise la définition suivante, légèrement différente, pour la médiane. L'avantage est que la médiane est unique, on la note  $\mu$  : si  $n = 2p + 1$  est impair,  $\mu = x_{p+1}$ , et si  $n = 2p$  est pair,  $\mu = \frac{x_p + x_{p+1}}{2}$ . On peut alors démontrer que la médiane  $\mu$  possède la propriété suivante : au moins la moitié de l'effectif de la série est inférieur ou égal à  $\mu$ , et au moins la moitié de l'effectif est supérieur ou égal à  $\mu$ .

— Un *mode* est une valeur de plus grand effectif. Un mode n'est pas forcément unique. Les modes n'ont d'intérêt que si leurs effectifs (égaux) sont nettement supérieurs à ceux des autres caractères.

— Soit  $\alpha \in ]0, 1[$ . Le *fractile d'ordre  $\alpha$*  est le plus petit nombre réel  $a$  de la série, tel qu'au moins  $100\alpha\%$  de l'effectif est inférieur ou égal à  $a$ . C'est donc une valeur de la série.

Lorsque  $\alpha$  vaut  $\frac{1}{4}$ ,  $\frac{1}{2}$  ou  $\frac{3}{4}$ , on parle de quartiles : premier quartile noté  $Q_1$ , deuxième quartile  $Q_2$ , troisième quartile  $Q_3$ . L'intervalle  $[Q_1, Q_3]$  s'appelle l'intervalle interquartile. Lorsque  $\alpha$  vaut  $0,1$  ou  $0,2, \dots$ , ou  $0,9$ , on parle de déciles : premier, deuxième, ..., neuvième.

**Attention.** La médiane et le deuxième quartile ne sont pas forcément égaux. En effet, le deuxième quartile est par définition une valeur de la série, ce qui n'est pas le cas pour la médiane.

EXEMPLE. Nous traitons la première série de l'exemple.

- La moyenne est :  $\bar{x} = \frac{2+2.5+8+10+3.12+15+20}{10} = 10,1$ .
- Tout nombre de l'intervalle  $[10, 12]$  est valeur médiane. Avec la définition utilisée dans le secondaire, la médiane vaut  $\mu = 11$ .
- Le premier quartile est  $Q_1 = 5$ , le deuxième  $Q_2 = 10$ , le troisième  $Q_3 = 12$ . Ainsi médiane ( $\mu = 11$ ) et deuxième quartile ( $Q_2 = 10$ ) ne coïncident pas nécessairement.

*Remarque.* Attention, les calculatrices et les tableurs ne calculent généralement pas correctement les fractiles des séries statistiques, et confondent deuxième quartile et médiane. Ainsi, pour calculer le premier quartile de la série 1, 2, 3 (qui vaut  $Q_1 = x_1 = 1$  car la série comporte trois termes), le tableur fait une approximation affine entre les différentes valeurs de la série, comme s'il s'agissait d'une série "continue uniforme", et retourne des valeurs qui ne sont pas nécessairement des valeurs de la série, ici 1,5 (ne pas oublier que, par définition, un quartile, et plus généralement un fractile, est une valeur de la série).

**Définition (Paramètres de dispersion).** Chacune des quantités définies ci-dessous se calcule pour la donnée d'une des variables.

- Les *valeurs extrêmes* sont le minimum et le maximum de la série, ici  $x_1$  et  $x_n$ .
- L'*étendue* est la différence entre les valeurs extrêmes de la série, ici  $x_n - x_1$ . On parle de *faible dispersion* si l'étendue est considérée comme petite, de *forte dispersion* si elle est considérée comme grande, ce qui est une appréciation subjective.
- La *variance*, notée  $\sigma_x^2$ , permet de mesurer la dispersion des données en tenant compte de toutes les valeurs. Elle est définie par :

$$\begin{aligned}\sigma_x^2 &= \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = \left( \frac{1}{n} \sum_{i=1}^n x_i^2 \right) - \bar{x}^2 \\ &= \frac{1}{n} \sum_{j=1}^m n_j (z_j - \bar{z})^2 = \left( \frac{1}{n} \sum_{j=1}^m n_j z_j^2 \right) - \bar{x}^2.\end{aligned}$$

L'*écart-type* est  $\sigma_x$ , ( $\sigma_x \geq 0$ ). Remarquons que  $\sigma_x = 0$  équivaut à dire que, pour tout  $i$ ,  $(x_i - \bar{x})^2 = 0$ , c'est-à-dire que tous les  $x_i$  sont égaux.

EXEMPLE. Valeurs extrêmes : 2, 20. Étendue :  $20 - 2 = 18$ .

**Définition (Paramètres de relation).** Les quantités définies ci-dessous permettent de déterminer des relations entre la donnée des différentes variables.

- La *covariance* entre les données  $x$  et  $y$ , notée  $\sigma_{x,y}$  est définie par :

$$\sigma_{x,y} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \left( \frac{1}{n} \sum_{i=1}^n x_i y_i \right) - \bar{x} \bar{y}.$$

- Si  $\sigma_x \neq 0$  et  $\sigma_y \neq 0$ , le coefficient de *corrélation*, noté  $\rho_{x,y}$ , est défini par :

$$\rho_{x,y} = \frac{\sigma_{x,y}}{\sigma_x \sigma_y}.$$

*Remarque.* Si  $x$  et  $y$  ont des unités (kg, m, ...),  $\sigma_x$  a pour unité celle de  $x$ ,  $\sigma_y$  celle de  $y$ , et  $\sigma_{x,y}$  a pour unité le produit des unités de  $x$  et de  $y$ . Mais le coefficient de corrélation  $\rho_{x,y}$  est un nombre, sans unité donc. Nous verrons plus tard qu'il vérifie :  $-1 \leq \rho_{x,y} \leq 1$ .

## 5.3 Interprétation des données

Considérons une statistique double  $\{(x_i, y_i) : i \in \{1, \dots, n\}\}$ , décrivant la donnée de deux variables quantitatives pour  $n$  individus.

### 5.3.1 Interprétation probabiliste

Soit  $\Omega = \{\omega_1, \dots, \omega_n\}$ . La série statistique double peut-être vue comme l'ensemble des valeurs prises par un couple de variables aléatoires  $(X, Y)$  défini sur l'univers  $\Omega$  :

$$\forall i \in \{1, \dots, n\}, (X, Y)(\omega_i) = (x_i, y_i).$$

Si on munit  $\Omega$  de la probabilité uniforme, alors :

$$\forall i \in \{1, \dots, n\}, \mathbb{P}(X = x_i, Y = y_i) = \frac{n_i}{n},$$

où  $n_i$  représente la multiplicité du couple  $(x_i, y_i)$  dans la série ( $n_i = 1$  si tous les couples sont distincts), et

$$\begin{aligned} \mathbb{E}(X) &= \bar{x}, \quad \mathbb{E}(Y) = \bar{y}, \quad \text{Var}(X) = \sigma_x^2, \quad \text{Var}(Y) = \sigma_y^2, \\ \text{Cov}(X, Y) &= \sigma_{x,y}, \quad \text{Cor}(X, Y) = \rho_{x,y}. \end{aligned}$$

### 5.3.2 Interprétation vectorielle

Notons “ $\cdot$ ” le produit scalaire usuel sur  $\mathbb{R}^n$  et  $\|\cdot\|$  la norme associée (la norme euclidienne). La donnée de chacune des variables peut être interprétée comme un vecteur de  $\mathbb{R}^n$ , noté respectivement  $\vec{x}$ ,  $\vec{y}$ . Notons  $\vec{u}$  le vecteur formé de 1 de  $\mathbb{R}^n$ . Ainsi, nous considérons :

$$\vec{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \quad \vec{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \quad \vec{u} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}.$$

Alors, la moyenne, variance et coefficient de corrélation peuvent s'écrire :

$$\begin{aligned} \bar{x} &= \frac{1}{n} \vec{x} \cdot \vec{u}, \quad \bar{y} = \frac{1}{n} \vec{y} \cdot \vec{u}, \\ \sigma_x^2 &= \frac{1}{n} (\vec{x} - \bar{x}\vec{u}) \cdot (\vec{x} - \bar{x}\vec{u}) = \frac{1}{n} \|\vec{x} - \bar{x}\vec{u}\|^2, \quad \sigma_y^2 = \frac{1}{n} \|\vec{y} - \bar{y}\vec{u}\|^2, \\ \sigma_{x,y} &= \frac{1}{n} (\vec{x} - \bar{x}\vec{u}) \cdot (\vec{y} - \bar{y}\vec{u}), \\ \rho_{x,y} &= \frac{(\vec{x} - \bar{x}\vec{u}) \cdot (\vec{y} - \bar{y}\vec{u})}{\|\vec{x} - \bar{x}\vec{u}\| \|\vec{y} - \bar{y}\vec{u}\|} = \cos \angle(\vec{x} - \bar{x}\vec{u}, \vec{y} - \bar{y}\vec{u}). \end{aligned}$$

Le coefficient de corrélation est un cosinus, il est donc compris entre  $-1$  et  $1$ .

## 5.4 Méthode des moindres carrés

Considérons une statistique double  $\{(x_i, y_i); 1 \leq i \leq n\}$ , vue comme un ensemble de  $n$  points dans  $\mathbb{R}^2$ . Représentons-les par un nuage de points dans un repère orthogonal  $(O, \vec{i}, \vec{j})$ , et notons  $M_i$  le point de coordonnées  $(x_i, y_i)$  (si un point apparaît plusieurs fois, on note sa multiplicité sur le graphe).

Le point  $G$  de coordonnées  $(\bar{x}, \bar{y})$  est appelé le *point moyen du nuage*; c'est "l'isobarycentre" des points  $\{(x_i, y_i); 1 \leq i \leq n\}$ . En fait, c'est le barycentre des points  $M_i$  comptés avec leur multiplicité, donc l'isobarycentre lorsque tous les points sont distincts.

Parfois, le nuage semble intuitivement "proche" d'une droite, même si cela n'a a priori aucun sens mathématique. On cherche quelle droite serait la "meilleure" pour approcher le nuage, dans le but d'estimer des données manquantes ou de faire des prévisions par exemple. On appelle cela faire un ajustement affine du nuage de points. Mais il faut bien sûr définir cette notion de "meilleur ajustement".

Considérons une droite  $\Delta$  qui ne soit pas parallèle à l'axe des ordonnées, d'équation  $y = ax + b$ , et, pour tout  $i$ , notons  $H_i$  le projeté de  $M_i$  sur  $\Delta$  parallèlement à  $(O, \vec{j})$ , voir Figure 5.1.

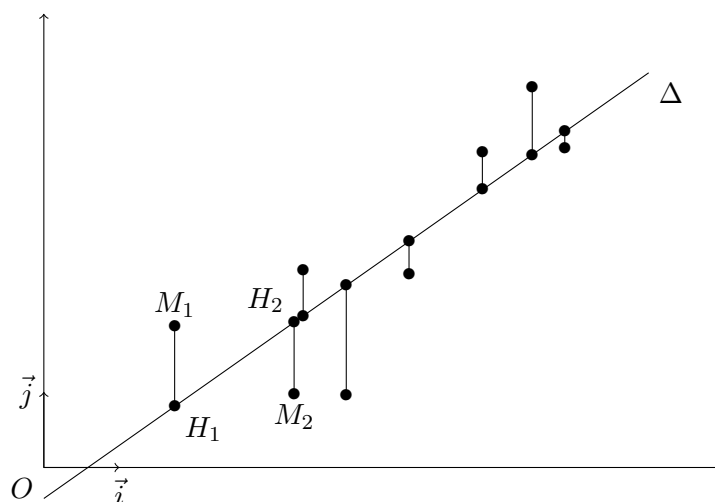


FIGURE 5.1 – Statistique double et droite de régression.

La *méthode des moindres carrés* consiste à dire que la meilleure approximation est celle qui minimise la somme des distances au carré des  $H_i$  aux  $M_i$ . Elle consiste donc à trouver  $a$  et  $b$  qui minimisent la quantité  $T(a, b)$ , appelée *somme des résidus*, où :

$$T(a, b) = \sum_{i=1}^n (y_i - ax_i - b)^2.$$

On suppose dans la suite que  $\sigma_x^2 \neq 0$ , ce qui signifie que les points du nuage ne sont pas alignés sur une droite verticale (s'ils sont alignés sur une droite verticale, on ne va pas essayer de faire mieux!). On peut aussi supposer, même si cela ne sert à rien dans la suite, que le nuage comporte au moins trois points distincts (si on n'a que deux points, ils sont alignés!).

**Théorème 5.1.** *Soit  $(x_i, y_i)_{1 \leq i \leq n}$  un nuage de points tel que  $\sigma_x^2 \neq 0$ . Alors il existe une unique droite d'équation  $y = ax + b$  telle que la somme des résidus soit minimale. C'est la droite*

d'équation

$$y = \frac{\sigma_{x,y}}{\sigma_x^2}x + \bar{y} - \bar{x}\frac{\sigma_{x,y}}{\sigma_x^2},$$

appelée droite de régression de  $y$  en  $x$ .

*Remarque.*

1. Cette droite passe par le point moyen du nuage.
2. Cette méthode d'ajustement, appelée méthode des moindres carrés, fut introduite par Gauss qui, en tant qu'astronome, s'intéressait aux orbites de petits astéroïdes comme Cérès, découvert en 1801. Après avoir observé l'astéroïde Cérès pendant plusieurs jours, celui-ci avait été perdu dans l'éclat du soleil. Gauss en calcula alors l'orbite en utilisant la méthode des moindres carrés (avec une ellipse), ce qui lui permit de prévoir la position et la date de réapparition de l'astéroïde.

*Démonstration. Première méthode*

En utilisant l'interprétation en terme de variables aléatoires de la statistique double, nous devons minimiser l'espérance :  $T(a, b) = n \mathbb{E}((Y - aX - b)^2)$ . D'après les propriétés de l'espérance et de la variance, nous pouvons écrire :

$$\begin{aligned} \frac{1}{n}T(a, b) &= \text{Var}(Y - aX - b) + [\mathbb{E}(Y - aX - b)]^2 \\ &= \text{Var}(Y - aX) + [\bar{y} - a\bar{x} - b]^2 \\ &= \text{Var}(Y) + a^2 \text{Var}(X) - 2a \text{Cov}(X, Y) + [\bar{y} - a\bar{x} - b]^2 \\ &= \sigma_y^2 + a^2 \sigma_x^2 - 2a \sigma_{x,y} + [\bar{y} - a\bar{x} - b]^2 \\ &= \sigma_y^2 + \left(a\sigma_x - \frac{\sigma_{x,y}}{\sigma_x}\right)^2 - \frac{\sigma_{x,y}^2}{\sigma_x^2} + [\bar{y} - a\bar{x} - b]^2. \end{aligned}$$

On en déduit l'équivalence :

$$T(a, b) \text{ minimal} \Leftrightarrow \left(a\sigma_x - \frac{\sigma_{x,y}}{\sigma_x} = 0 \text{ et } \bar{y} - a\bar{x} - b = 0\right).$$

Ainsi, on obtient la droite de régression de  $y$  en  $x$  :

$$y = \frac{\sigma_{x,y}}{\sigma_x^2}x + \bar{y} - \bar{x}\frac{\sigma_{x,y}}{\sigma_x^2},$$

et la borne inférieure suivante pour la somme des résidus :

$$\forall (a, b) \in \mathbb{R}^2, T(a, b) \geq n \left( \sigma_y^2 - \frac{\sigma_{x,y}^2}{\sigma_x^2} \right).$$

□

*Remarque.* Supposons que de plus  $\sigma_y \neq 0$ , ce qui signifie que les points ne sont pas alignés non plus horizontalement.

— La démonstration précédente permet d'obtenir l'équivalence suivante :

$$\exists (a, b) \in \mathbb{R}^2, T(a, b) = 0 \Leftrightarrow \sigma_y^2 - \frac{\sigma_{x,y}^2}{\sigma_x^2} = 0 \Leftrightarrow |\rho_{x,y}| = 1.$$

C'est-à-dire que les points du nuage sont alignés si et seulement si la valeur absolue du coefficient de corrélation vaut 1.

— Notons  $(a_0, b_0)$  le point qui minimise  $T(a, b)$ . Alors,

$$\frac{1}{n}T(a_0, b_0) = \sigma_y^2 - \frac{\sigma_{x,y}^2}{\sigma_x^2}, \text{ d'où } 1 - \rho_{x,y}^2 = \frac{1}{n\sigma_y^2}T(a_0, b_0).$$

— On aurait pu faire la même chose en projetant sur l'axe  $(0, \vec{i})$ .

**Démonstration. Deuxième méthode**

On considère, pour tout  $(a, b) \in \mathbb{R}^2$  :

$$\begin{aligned} \frac{1}{n}T(a, b) &= \mathbb{E}[(Y - (aX + b))^2] \\ &= a^2 \mathbb{E}[X^2] + \mathbb{E}[Y^2] + b^2 - 2a \mathbb{E}[XY] - 2b \mathbb{E}[Y] + 2ab \mathbb{E}[X] \\ &= a^2(\sigma_x^2 + \bar{x}^2) + \sigma_y^2 + \bar{y}^2 + b^2 - 2a(\sigma_{x,y} + \bar{x}\bar{y}) - 2b\bar{y} + 2ab\bar{x}, \end{aligned}$$

comme une application polynomiale en  $a$  et  $b$ . Ainsi, l'application  $T$  est  $\mathcal{C}^\infty$ , de sorte que si elle admet un minimum en un point, alors ses dérivées partielles sont nulles en ce point.

$$\begin{cases} \frac{\partial T}{\partial a}(a_0, b_0) = 0 \\ \frac{\partial T}{\partial b}(a_0, b_0) = 0 \end{cases} \Leftrightarrow \begin{cases} a_0(\sigma_x^2 + \bar{x}^2) + b_0\bar{x} = \sigma_{x,y} + \bar{x}\bar{y} \\ a_0\bar{x} + b_0 = \bar{y} \end{cases} \Leftrightarrow \begin{cases} a_0 = \frac{\sigma_{x,y}}{\sigma_x^2} \\ b_0 = \bar{y} - \bar{x} \frac{\sigma_{x,y}}{\sigma_x^2} \end{cases}$$

Montrons que  $(a_0, b_0)$  réalise effectivement un minimum pour  $T$ . Comme  $T$  est une application polynomiale de degré 2, et que les dérivées partielles de premier ordre s'annulent en  $(a_0, b_0)$ , la formule de Taylor à l'ordre 2 autour de  $(a_0, b_0)$  donne que  $T(a_0 + h, b_0 + k)$  est égal à :

$$\begin{aligned} &= T(a_0, b_0) + \frac{1}{2}h^2 \frac{\partial^2 T}{\partial a^2}(a_0, b_0) + hk \frac{\partial^2 T}{\partial a \partial b}(a_0, b_0) + \frac{1}{2}k^2 \frac{\partial^2 T}{\partial b^2}(a_0, b_0) + 0 \\ &= T(a_0, b_0) + h^2(\sigma_x^2 + \bar{x}^2) + 2hk\bar{x} + k^2 \\ &= T(a_0, b_0) + (k + h\bar{x})^2 + h^2\sigma_x^2 \geq T(a_0, b_0). \end{aligned}$$

Il s'ensuit :  $\forall h \in \mathbb{R}, \forall k \in \mathbb{R}, T(a_0 + h, b_0 + k) \geq T(a_0, b_0)$ . C'est-à-dire que  $(a_0, b_0)$  est l'unique point réalisant le minimum de  $T$ .  $\square$

**Démonstration. Troisième méthode**

Considérons les trois vecteurs de  $\mathbb{R}^n$  :  $\vec{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ ,  $\vec{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$ ,  $\vec{u} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$ .

Ainsi,  $T(a, b) = \|\vec{y} - (a\vec{x} + b\vec{u})\|^2$ . De plus, l'hypothèse  $\sigma_x \neq 0$  équivaut à dire que  $\|\vec{x} - \bar{x}\vec{u}\| \neq 0$ , autrement dit que les vecteurs  $\vec{x}$  et  $\vec{u}$  ne sont pas colinéaires. En conséquence  $\text{Vect}(\vec{x}, \vec{u})$  est un plan vectoriel. La quantité  $T(a, b)$  est donc minimale lorsque le vecteur  $a\vec{x} + b\vec{u}$  est la projection orthogonale de  $\vec{y}$  sur le plan  $\text{Vect}(\vec{x}, \vec{u})$ , voir figure 5.2.

Ceci équivaut à dire que le vecteur  $\vec{y} - (a\vec{x} + b\vec{u})$  est orthogonal à  $\vec{x}$  et à  $\vec{u}$  :

$$\begin{aligned} &\begin{cases} (\vec{y} - (a\vec{x} + b\vec{u})) \cdot \vec{x} = 0 \\ (\vec{y} - (a\vec{x} + b\vec{u})) \cdot \vec{u} = 0 \end{cases} \Leftrightarrow \begin{cases} a\|\vec{x}\|^2 + b\vec{u} \cdot \vec{x} = \vec{y} \cdot \vec{x} \\ a\vec{x} \cdot \vec{u} + bn = \vec{y} \cdot \vec{u} \end{cases} \\ \Leftrightarrow &X \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} \vec{y} \cdot \vec{x} \\ n\bar{y} \end{pmatrix}, \text{ où } X = \begin{pmatrix} \|\vec{x}\|^2 & n\bar{x} \\ n\bar{x} & n \end{pmatrix} \end{aligned}$$

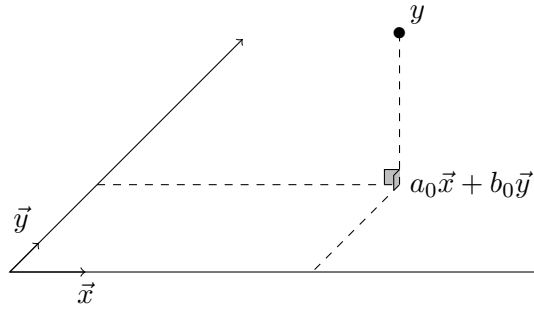


FIGURE 5.2 –  $T(a, b)$  est minimal lorsque  $a\vec{x} + b\vec{u}$  est la projection orthogonale de  $\vec{y}$  sur le plan  $\text{Vect}(\vec{x}, \vec{u})$ .

Le déterminant de la matrice  $X$  est :

$$\det(X) = n\|\vec{x}\|^2 - n^2\bar{x}^2 = n^2\sigma_x^2 \neq 0, \text{ par hypothèse.}$$

En conséquence :

$$\begin{aligned} \begin{pmatrix} a \\ b \end{pmatrix} &= \frac{1}{n^2\sigma_x^2} \begin{pmatrix} n & -n\bar{x} \\ -n\bar{x} & \|\vec{x}\|^2 \end{pmatrix} \begin{pmatrix} \vec{y} \cdot \vec{x} \\ n\bar{y} \end{pmatrix} \\ &= \frac{1}{n^2\sigma_x^2} \begin{pmatrix} n\vec{y} \cdot \vec{x} - n^2\bar{x}\bar{y} \\ n\bar{y}\|\vec{x}\|^2 - n\bar{x}\vec{y} \cdot \vec{x} \end{pmatrix} = \frac{1}{\sigma_x^2} \begin{pmatrix} \sigma_{x,y} \\ \sigma_x^2\bar{y} - \bar{x}\sigma_{x,y} \end{pmatrix}. \end{aligned}$$

□

### EXERCICE 5.1. Série statistique à une variable

On veut contrôler la qualité des livraisons d'une coopérative agricole de production de pommes de terre. En principe, elle produit des sacs de 5 kg, avec une tolérance de 0,2 kg en moins. On pèse 50 sacs pris au hasard. Cela donne la série statistique suivante :

5,6 – 4,9 – 5,5 – 5,0 – 4,4 – 5,3 – 5,0 – 5,1 – 5,8 – 5,3 – 5,3 – 5,3 – 5,0 – 5,5 –  
 5,0 – 5,1 – 4,3 – 4,5 – 5,2 – 4,4 – 5,3 – 4,9 – 5,4 – 4,5 – 4,6 – 4,4 – 5,6 – 5,6 –  
 5,7 – 4,4 – 5,1 – 5,1 – 4,9 – 5,4 – 5,2 – 5,1 – 4,8 – 5,0 – 3,8 – 5,4 – 4,3 – 5,2 –  
 5,3 – 5,1 – 5,4 – 4,0 – 4,9 – 5,5 – 4,5 – 5,5.

1. Donner le tableau des effectifs, des effectifs cumulés croissants et des effectifs cumulés décroissants. Même chose avec les fréquences. Quelle erreur des élèves peut-on anticiper dans ce type de problème, erreur qu'on a peu de chance de voir ici compte tenu du nombre de données ?
2. Donner le maximum, le minimum de cette série, son étendue, son ou ses modes.
3. Donner diverses représentations statistiques de ces données.
4. Calculer la moyenne  $\bar{x}$  de cette série, ainsi que son écart-type  $s$ . En déterminer la médiane.
5. Déterminer le premier et le troisième quartile de la série. Calculer l'intervalle interquartile. Déterminer le premier et le neuvième décile de cette série.
6. Tracer la boîte à moustaches correspondante.
7. Classer cette série en classes d'égale étendue de 0,2 kg. Indiquer quelles sont les classes modale et médiane. Donner une représentation graphique de cette série (effectifs et fréquences cumulées). Parler de la médiane de la série classée a-t-il un sens ?

8. À partir de la série classée, donner une estimation de la moyenne de la série de départ. Parler de la moyenne de la série classée a-t-il un sens ?

SOLUTION 5.1.

1. Notons  $(x_1, \dots, x_{50})$  la série statistique, et  $(z_1, \dots, z_{17})$  les valeurs différentes des données de la série. Notons encore  $n = 50$  et  $m = 17$ .

L'effectif  $n_j$  d'une valeur  $z_j$  est le nombre de répétitions de cette valeur dans la série.

La fréquence  $f_j$  d'une donnée  $z_j$  est  $f_j = \frac{\text{effectif}(z_j)}{\text{effectif total}} = \frac{n_j}{n}$ .

Ces quantités sont résumées dans le tableau ci-dessous, ainsi que les effectifs cumulés et les fréquences cumulées.

Classe (kg)	Effectif	Effectifs cumulés croissants	Effectifs cumulés décroissants	Fréquence (en %)	Fréquences cumulées croissantes
3,8	1	1	50	2	2
4,0	1	2	49	2	4
4,3	2	4	48	4	8
4,4	4	8	46	8	16
4,5	3	11	42	6	22
4,6	1	12	39	2	24
4,8	1	13	38	2	26
4,9	4	17	37	8	34
5,0	5	22	33	10	44
5,1	6	28	28	12	56
5,2	3	31	22	6	62
5,3	6	37	19	12	74
5,4	4	41	13	8	82
5,5	4	45	9	8	90
5,6	3	48	5	6	96
5,7	1	49	2	2	98
5,8	1	50	1	2	100

Il s'agit d'erreurs d'arrondis. Par exemple, si la première fréquence est  $a/b$  dont l'élève donne une valeur approchée  $x$  avec deux décimales, puis la deuxième est  $c/d$  dont il donne une valeur approchée  $y$  avec deux décimales, dans le tableau des fréquences cumulées croissantes, il risque d'additionner  $x$  et  $y$  au lieu d'additionner  $a/b$  et  $c/d$  et de donner une valeur approchée de cette somme, de sorte qu'à la fin, l'élève risquent d'obtenir une fréquence cumulée croissante différente de 1.

2. Le maximum est 5,8 et le minimum est 3,8. L'étendue est la différence entre la valeur maximale et la valeur minimale, soit 2. Le mode est le ou les valeurs de plus grands effectifs. Dans notre cas, on a deux modes : 5,1 et 5,3.
3. Voici le diagramme en bâtons des effectifs. La hauteur de chacun des "bâtons" représente l'effectif.

Voici le diagramme en bâton des fréquences cumulées croissantes.

On aurait aussi pu tracer le diagrammes en bâton des effectifs cumulés ou celui des fréquences.

On peut également représenter les données sous forme d'un *diagramme circulaire*, plus connu sous le nom de "camembert", où la mesure angulaire de chaque secteur correspond



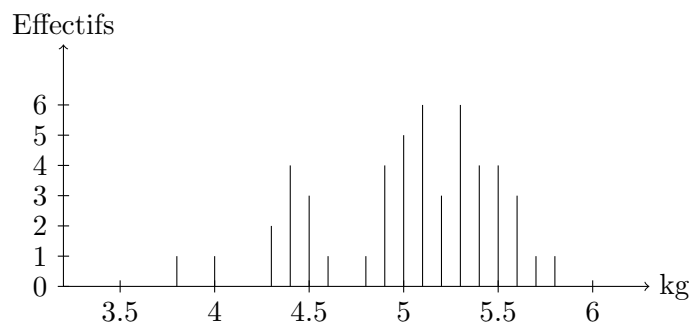


FIGURE 5.3 – Diagramme en bâtons des effectifs.

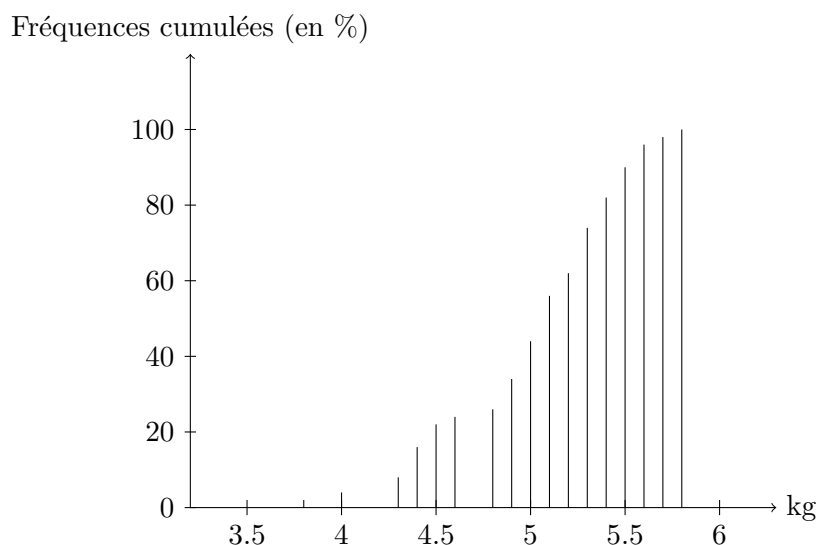


FIGURE 5.4 – Diagramme en bâtons des fréquences cumulées croissantes.

à la fréquence de chaque caractère. C'est une simple question de proportionnalité. Ce genre de représentation est adapté lorsque le nombre de données différentes dans la série n'est pas trop grand.

On représente parfois les données sur un diagramme semi-circulaire. Il suffit alors de calculer les mesures angulaires par rapport à  $180^\circ$ .

4. Par définition, la moyenne est :

$$\bar{x} = \frac{1}{50} \sum_{k=1}^{50} x_k = \frac{1}{50} \sum_{j=1}^{17} n_j z_j,$$

d'où :

$$\begin{aligned} \bar{x} &= \frac{1.3,8+1.4+2.4,3+4.4,4+3.4,5+1.4,6+1.4,8+4.4,9+5.5+6.5,1+3.5,2+6.5,3+4.5,4+4.5,5+3.5,6+1.5,7+1.5,8}{50} \\ &= \frac{251,4}{50} \simeq 5,0 \text{ kg.} \end{aligned}$$

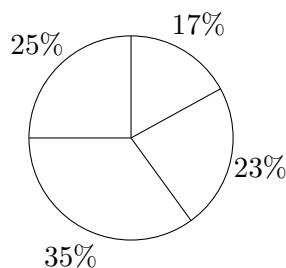


FIGURE 5.5 – Exemple de diagramme circulaire. Attention, il ne correspond pas à cette série statistique.

L'écart type est :

$$\begin{aligned} \sigma_x &= \sqrt{\frac{1}{50} \sum_{i=1}^{50} (x_i - \bar{x})^2} = \sqrt{\frac{1}{50} \sum_{i=1}^{50} x_i^2 - (\bar{x})^2} \\ &= \sqrt{\frac{1274,42}{50} - \left(\frac{251,4}{50}\right)^2} \simeq 0,46 \text{ kg.} \end{aligned}$$

Rappelons qu'on appelle valeur médiane toute valeur  $\mu$  telle qu'au moins la moitié de l'effectif est supérieur ou égal à  $\mu$ , et au moins la moitié de l'effectif est inférieur ou égal à  $\mu$ . On constate que plus de la moitié de l'effectif (28 sur 50) est inférieur ou égal à 5,1 alors que plus de la moitié de l'effectif (28 sur 50) est supérieur ou égal à 5,1. On en déduit que 5,1 est une valeur médiane. Si l'on choisit pour définition du mot "médiane" d'une série  $(x_1, x_2, \dots, x_{2p})$  d'un nombre pair de valeurs classées par ordre croissant, le nombre  $\frac{x_p + x_{p+1}}{2}$ , on trouve encore ici que 5,1 est la médiane de cette série ( $n = 2p = 50$ ,  $x_{25} = x_{26} = 5,1$ ).

5. Le quartile  $Q_1$  (resp.  $Q_3$ ) est le plus petit nombre  $q$  de la série tel qu'au moins 25% (resp. 75%) de l'effectif soit inférieur ou égal à  $q$ . On en déduit :  $Q_1 = 4,8$  et  $Q_3 = 5,4$ .

L'intervalle interquartile est alors :  $Q_3 - Q_1 = 5,4 - 4,8 = 0,6$ .

Le premier (resp. neuvième) décile  $D_1$  (resp.  $D_9$ ) est le plus petit nombre  $d$  de la série tel qu'au moins 10% (resp. 90%) de l'effectif soit inférieur ou égal à  $d$ . Ainsi,  $D_1 = 4,4$  et  $D_9 = 5,5$ .

6. La *boîte à moustache* consiste à tracer un rectangle qui va du premier quartile au troisième quartile et coupé par la médiane. On ajoute ensuite des segments aux extrémités menant jusqu'au premier décile et au neuvième décile. Ainsi :

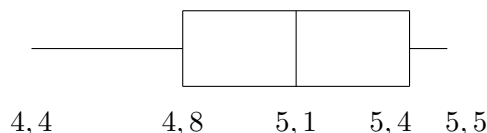


FIGURE 5.6 – Boîte à moustache.

7. Il n'y a bien sûr pas unicité du choix du regroupement par classes. Néanmoins, le plus logique est de choisir des classes centrées sur les valeurs de la série initiale. D'où le tableau :

Classe	Effectif	Fréquence en %	Fréquence cumulées
[3, 7; 3, 9[	1	2	2
[3, 9; 4, 1[	1	2	4
[4, 1; 4, 3[	0	0	4
[4, 3; 4, 5[	6	12	16
[4, 5; 4, 7[	4	8	24
[4, 7; 4, 9[	1	2	26
[4, 9; 5, 1[	9	18	44
[5, 1; 5, 3[	9	18	62
[5, 3; 5, 5[	10	20	82
[5, 5; 5, 7[	7	14	96
[5, 7; 5, 9[	2	4	100

Une classe modale est une classe de fréquence maximale ; ici, il n'y a qu'une seule classe modale, la classe [5, 3; 5, 5[.

La médiane appartient à la classe [5, 1; 5, 3[ qui est donc la classe médiane. Cela ne fait pas de sens de parler de médiane de la série classée.

On notera que 62% de l'effectif est strictement inférieur à 5, 3 alors que 44% de l'effectif est strictement inférieur à 5, 1 (i.e. 56% de l'effectif est supérieur ou égal à 5, 3).

Lorsque les données sont réparties dans des classes (ce qui est en particulier le cas lorsque l'on étudie une variable continue) la représentation appropriée est l'*histogramme*. L'effectif de chaque classe est représenté par un rectangle dont la base est l'amplitude de la classe et l'aire est proportionnelle à l'effectif (ou la fréquence). Il faut faire attention à la construction lorsque les classes ne sont pas régulières, ce qui n'est pas le cas ici.

Voici l'histogramme des effectifs.

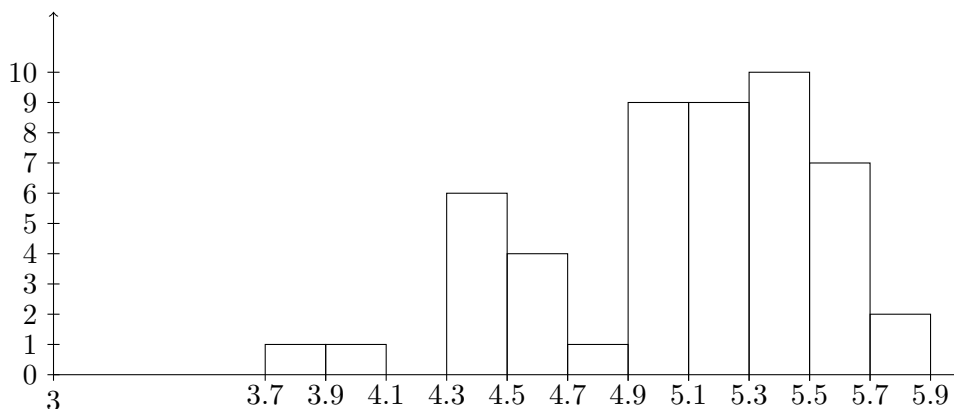


FIGURE 5.7 – Histogramme des effectifs.

Voici l'histogramme des fréquences cumulées croissantes.

- La moyenne  $\bar{x}$  de la série est comprise entre la moyenne  $\mu_i$  de la série pondérée des valeurs inférieures de chaque classe et la moyenne  $\mu_s$  de la série pondérée des valeurs supérieures de chaque classe.

En effet, si on désigne par  $m_j$  et  $M_j$  les valeurs inférieures et supérieures de la  $j$ -ième

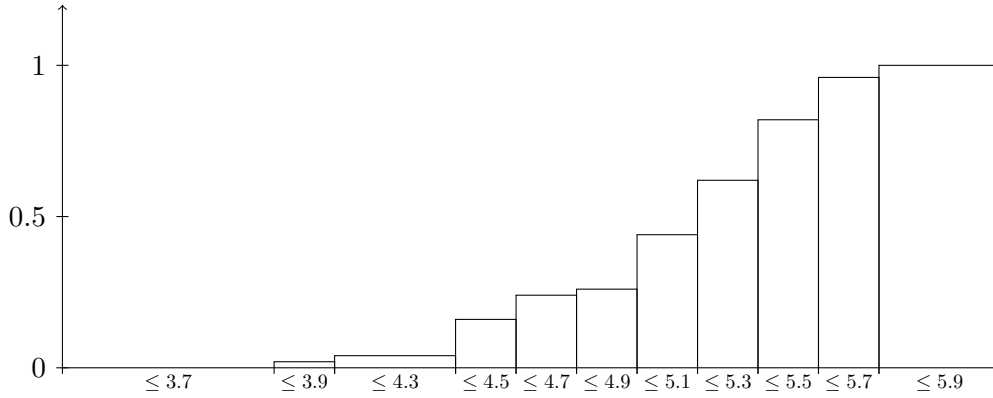


FIGURE 5.8 – Histogramme des effectifs.

classe, on a :

$$\forall j, \forall k \text{ tel que } x_k \in [m_j, M_j[, \quad m_j \leq x_k < M_j,$$

$$\text{d'où, } n_j m_j \leq \sum_{k \text{ tel que } x_k \in [m_j, M_j[} x_k < n_j M_j,$$

$$\text{ce qui entraîne, } \frac{1}{n} \sum_{j=1}^m n_j m_j \leq \frac{1}{n} \sum_{k=1}^N x_k < \frac{1}{n} \sum_{j=1}^m n_j M_j,$$

$$\text{autrement dit, } \mu_i \leq \bar{x} \leq \mu_s.$$

Lorsque les classes sont de même amplitude  $a$ , il est commode de calculer la moyenne  $\bar{x}'$  de la série pondérée  $(\frac{m_j + M_j}{2})_{1 \leq j \leq m}$  des milieux des classes :

$$\bar{x}' = \frac{1}{n} \sum_{j=1}^m n_j \frac{M_j + m_j}{2}$$

ce qui donne sur cet exemple

$$\begin{aligned} \bar{x}' &= \frac{1.3,8 + 1.4 + 6.4,4 + 4.4,6 + 1.4,8 + 9.5 + 9.5,2 + 10.5,4 + 7.5,6 + 2.5,8}{50} \\ &= \frac{254}{50} = 5,08. \end{aligned}$$

L'erreur faite alors est égale à la moitié de l'amplitude commune des classes :

$$|\bar{x} - \bar{x}'| \leq \frac{1}{2}(\mu_s - \mu_i) = \frac{1}{2n} \sum_{j=1}^m n_j (M_j - m_j) = \frac{1}{2n} \sum_{j=1}^m n_j a = \frac{a}{2},$$

et l'on peut écrire  $\bar{x}' - \frac{a}{2} \leq \bar{x} \leq \bar{x}' + \frac{a}{2}$ . On en déduit ici :  $5,07 \leq \bar{x} \leq 5,09$ .

On trouve souvent dans les manuels scolaires la valeur  $\bar{x}'$  considérée comme étant **égale** à la moyenne de la série classée, ce qui est **faux**. Pour que cela soit vrai, il faudrait que les valeurs de la série soient uniformément réparties à l'intérieur des classes (ce qui entraîne que l'on connaît en fait toutes les valeurs de la série!), ce qui est quasiment toujours faux. Parfois, cette hypothèse est ajoutée dans les énoncés, même si elle est totalement irréaliste. Cela est pourtant inutile, puisque l'on peut écrire facilement des

choses toujours justes, en procédant comme ci-dessus. On ne parlera donc **jamais** de **la valeur** de la moyenne d'une série classée, mais on en donnera systématiquement un encadrement.

**EXERCICE 5.2. Séries statistiques à deux variables**

*D'après Méthodes statistiques. Ed. DUNOD.*

Le tableau ci-dessous donne les indices des prix des produits alimentaires et des produits énergétiques pour les douze pays de la Communauté européenne et les États-Unis en 1990 (base 100 en 1985).

Pays	Produits énergétiques $x_i$	Produits alimentaires $y_i$
Belgique	82,1	107,7
Danemark	116,4	111,4
R. F. A.	85,5	104,9
Grèce	172,4	215,9
Espagne	99,1	136,8
France	91,9	116,2
Irlande	97,9	116,8
Italie	116	127,1
Luxembourg	77,6	108,8
Pays-Bas	80,5	100,3
Portugal	135,1	158,6
Royaume Uni	114,8	125,7
États Unis	100,4	126,9

- Utiliser la calculatrice ou un logiciel pour :
  - représenter le nuage de points associé à la série statistique  $(x_i, y_i)_{1 \leq i \leq n}$  ;
  - déterminer les coordonnées du point moyen  $G$  du nuage ; le placer ;
  - déterminer l'écart-type de chacune des deux séries, ainsi que la covariance ;
  - déterminer le coefficient de corrélation linéaire en  $x$  et en  $y$  au millième près ;
  - déterminer une équation de la droite de régression  $\mathcal{D}$  de  $y$  en  $x$  par la méthode des moindres carrés ;
  - représenter la droite  $\mathcal{D}$ .
- On constate qu'à part la Grèce, le nuage de points est assez compact. Reprendre alors les questions précédentes après avoir éliminé la Grèce de la population considérée.
- Quel est l'effet d'un changement de repère et d'un changement d'échelle sur l'équation de la droite de régression ?

**EXERCICE 5.3. Séries statistiques à deux variables**

*D'après Déclic, Terminale ES, Édition 2006.*

Le tableau ci-dessous donne le trafic aérien intérieur français, exprimé en milliards de voyageurs-kilomètres.

Année	1985	1986	1987	1988	1989	1990	1991
rang $x_i$	1	2	3	4	5	6	7
trafic $y_i$	7,4	8,3	8,9	9,6	11	11,4	11,7

Année	1992	1993	1994	1995	1996	1997	1998
rang $x_i$	8	9	10	11	12	13	14
trafic $y_i$	12, 2	12, 3	12, 7	12, 7	13, 8	13, 8	14, 5

Source : Direction générale de l'Aviation civile.

1. Représenter le nuage de points  $M_i(x_i, y_i)$  dans un repère orthogonal.  
Calculer les coordonnées du point moyen  $G$  du nuage. Le placer.
2. Ajustement par la droite de Mayer
  - (a) Déterminer les coordonnées du point moyen  $G_1$  des 7 premiers points du nuage et du point moyen  $G_2$  des 7 derniers.
  - (b) Déterminer l'équation réduite de la droite  $(G_1, G_2)$  sous la forme  $y = ax + b$ . Vérifier que  $G$  est un point de  $(G_1, G_2)$ .
3. Ajustement par une droite choisie.  
Déterminer l'équation réduite de la droite  $\mathcal{D}$  passant par  $G$ , de coefficient directeur 0,5. Tracer  $\mathcal{D}$ .
4. Ajustement par la droite des "extrêmes".  
Calculer l'accroissement moyen annuel du trafic entre 1985 et 1988. En déduire l'équation réduite de la droite  $(M_1, M_{14})$ . Le point  $G$  appartient-il à cette droite ?
5. Ajustement affine par la méthode des moindres carrés  
À l'aide de la calculatrice, déterminer l'équation réduite de la droite de régression  $\Delta$  de  $y$  en  $x$  par la méthode des moindres carrés.
6. Comparaison
  - (a) À l'aide des listes de la calculatrice, ou d'un tableur, calculer la somme des résidus :

$$S = \sum_{i=1}^{14} (y_i - ax_i - b)^2,$$

pour chacune des droites précédentes d'équation réduite de la forme  $y = ax + b$  ( $a$  et  $b$  donnés à  $10^{-3}$  près).

- (b) Quelle est la droite pour laquelle cette somme est minimale ?

# Chapitre 6

## Statistique inférentielle

Ce chapitre est fortement inspiré du cours de Pierre Priouret (Université Pierre et Marie Curie, 1983).

### 6.1 Un exemple introductif

On veut estimer la proportion inconnue  $\theta$  de chauves dans une population  $\mathcal{S} = \{s_1, \dots, s_N\}$  de  $N$  individus. Pour cela, on va faire des expériences, des sondages, en interrogeant  $n$  personnes.

On note  $\Omega = \mathcal{S}^n$  l'ensemble des échantillons de taille  $n$  avec répétitions de  $\mathcal{S}$ . Un élément  $\omega = (\omega_1, \dots, \omega_n)$  de  $\Omega$  est donc un  $n$ -uplet constitué de  $n$  personnes, choisies "au hasard" et avec répétition dans  $\mathcal{S}$ . Étant donné qu'il n'y a pas de raison de choisir un échantillon de personnes sondées  $\omega$  plutôt qu'un autre, on munit  $\Omega$  de la probabilité uniforme  $\mathbb{P}$  (c'est ce que l'on sous-entend en disant que les personnes sont choisies "au hasard"). Ainsi,

$$\forall \omega \in \Omega, \mathbb{P}(\{\omega\}) = \frac{1}{\text{Card}(\Omega)} = \frac{1}{N^n}.$$

Faire un sondage, c'est se donner un élément  $\omega = (\omega_1, \dots, \omega_n)$  de  $\Omega$  et demander à chacun des individus  $\omega_1, \dots, \omega_n$ , s'il est chauve ou pas. Pour tout  $i$  compris entre 1 et  $n$ , on définit la variable aléatoire  $X_i$  de la façon suivante : si  $\omega = (\omega_1, \dots, \omega_n)$ , on pose

$$X_i(\omega) = \begin{cases} 1 & \text{si l'individu } \omega_i \text{ est chauve,} \\ 0 & \text{si l'individu } \omega_i \text{ n'est pas chauve.} \end{cases}$$

Alors, pour tout échantillon  $\omega$  de personnes sondées,  $\sum_{k=1}^n X_k(\omega)$  représente le nombre de personnes chauves de l'échantillon, et la quantité  $M_n(\omega) = \frac{\sum_{k=1}^n X_k(\omega)}{n}$  représente la proportion d'individus chauves dans cet échantillon.

#### Propriété 12.

1. Pour tout  $i \in \{1, \dots, n\}$ , la variable aléatoire  $X_i$  suit une loi de Bernoulli de paramètre  $\theta$ .
2. Les variables aléatoires  $X_1, \dots, X_n$  sont indépendantes.
3. La variable aléatoire  $nM_n = \sum_{k=1}^n X_k$  suit une loi binomiale de paramètres  $n$  et  $\theta$ .

4. Pour tout  $\varepsilon > 0$ ,  $\mathbb{P}(\theta \in ]M_n - \varepsilon, M_n + \varepsilon[) \geq 1 - \frac{1}{4n\varepsilon^2}$ .

*Démonstration.*

1. Les variables aléatoires  $X_i$  suivent la même loi de Bernoulli de paramètre  $\theta$ . En effet, en désignant par  $C$  l'ensemble des individus chauves de  $\mathcal{S}$ , on a, puisque  $\mathbb{P}$  est la probabilité uniforme sur  $\Omega$  :

$$\mathbb{P}[X_i = 1] = \mathbb{P}[\{(\omega_1, \dots, \omega_n) \in \Omega : \omega_i \text{ est chauve}\}] = \frac{\text{Card}(A_i)}{\text{Card}(\Omega)},$$

où

$$\begin{aligned} A_i &= \{(\omega_1, \dots, \omega_n) \in \Omega : \omega_i \text{ est chauve}\} \\ &= \underbrace{\mathcal{S} \times \dots \times \mathcal{S}}_{i-1 \text{ termes}} \times C \times \underbrace{\mathcal{S} \times \dots \times \mathcal{S}}_{n-i \text{ termes}} = \mathcal{S}^{i-1} \times C \times \mathcal{S}^{n-i}, \end{aligned}$$

avec la convention que, dans le produit cartésien,  $\mathcal{S}^{i-1}$  n'est pas écrit si  $i = 1$  ou que  $\mathcal{S}^{n-i}$  n'est pas écrit si  $n = i$ . On en déduit :

$$\mathbb{P}[X_i = 1] = \frac{\text{Card}(\mathcal{S}^{i-1} \times C \times \mathcal{S}^{n-i})}{[\text{Card}(\mathcal{S})]^n} = \frac{[\text{Card}(\mathcal{S})]^{n-1} \text{Card}(C)}{[\text{Card}(\mathcal{S})]^n} = \frac{\text{Card}(C)}{\text{Card}(\mathcal{S})} = \theta,$$

puisque, par définition,  $\theta$  représente la proportion de chauves dans la population.

2. Exercice.

3. La variable aléatoire  $\sum_{k=1}^n X_k$  suit une loi binomiale de paramètres  $n$  et  $\theta$  car elle est somme de  $n$  variables aléatoires indépendantes de Bernoulli de paramètre  $\theta$ .
4. La variable aléatoire  $M_n$  étant bornée ( $0 \leq M_n \leq 1$ ), elle admet des moments de tout ordre et nous déduisons du point précédent que :

$$\mathbb{E}(M_n) = \mathbb{E}\left(\frac{1}{n} \sum_{k=1}^n X_k\right) = \frac{1}{n} \sum_{k=1}^n \mathbb{E}(X_k) = \theta$$

et, puisque les variables sont indépendantes,

$$\text{Var}(M_n) = \text{Var}\left(\frac{1}{n} \sum_{k=1}^n X_k\right) = \frac{1}{n^2} \sum_{k=1}^n \text{Var}(X_k) = \frac{\theta(1-\theta)}{n}.$$

Les hypothèses étant vérifiées, on peut appliquer l'inégalité de Bienaymé-Tchebychev. Ainsi :

$$\forall \varepsilon > 0, \quad \mathbb{P}[|M_n - \theta| \geq \varepsilon] \leq \frac{\text{Var}(M_n)}{\varepsilon^2} = \frac{\theta(1-\theta)}{n\varepsilon^2}.$$

Or, l'application  $x \mapsto x(1-x)$  est majorée sur  $[0, 1]$  par sa valeur au point  $x = \frac{1}{2}$ , à savoir  $\frac{1}{4}$  (on a un morceau de parabole, à concavité tournée vers le bas, et dont le sommet est au point d'abscisse  $\frac{1}{2}$ ). On en déduit, que pour tout  $\varepsilon > 0$  :

$$\begin{aligned} \mathbb{P}[|M_n - \theta| \geq \varepsilon] &\leq \frac{1}{4n\varepsilon^2} \\ \Leftrightarrow \mathbb{P}[|M_n - \theta| < \varepsilon] &\geq 1 - \frac{1}{4n\varepsilon^2}, \text{ en passant au complémentaire} \\ \Leftrightarrow \mathbb{P}[\theta \in ]M_n - \varepsilon, M_n + \varepsilon[ &\geq 1 - \frac{1}{4n\varepsilon^2}. \end{aligned}$$



□

Fixons  $\varepsilon$  de sorte que, par exemple,  $1 - \frac{1}{4n\varepsilon^2} = 0,95$ , *i.e.*  $\varepsilon = \sqrt{\frac{5}{n}}$ . Faisons un sondage de taille  $n$ , c'est-à-dire choisissons  $\omega$  et observons

$$x_1 = X_1(\omega), \dots, x_n = X_n(\omega).$$

Calculons alors  $M_n(\omega) = \frac{1}{n} \sum_{k=1}^n x_k$ . La proportion  $\theta$  inconnue se trouve dans l'intervalle  $]M_n(\omega) - \varepsilon, M_n(\omega) + \varepsilon[$ , c'est-à-dire dans l'intervalle

$$\left] \frac{1}{n} \sum_{k=1}^n x_k - \sqrt{\frac{5}{n}}, \frac{1}{n} \sum_{k=1}^n x_k + \sqrt{\frac{5}{n}} \right[$$

avec une confiance au moins égale à 0,95. Ceci veut dire que la proportion  $\theta$  appartient à 95% des intervalles de la forme  $\left( \frac{1}{n} \sum_{k=1}^n x_k - \sqrt{\frac{5}{n}}, \frac{1}{n} \sum_{k=1}^n x_k + \sqrt{\frac{5}{n}} \right[ : (x_1, \dots, x_n) \in \{0, 1\}^n$ .

On dit que les variables aléatoires  $X_1, \dots, X_n$ , sont un échantillon de taille  $n$  de loi de Bernoulli de paramètre  $\theta$ .

## 6.2 Modèle statistique paramétrique

On suppose qu'un certain phénomène suit une loi de probabilité connue, dépendant d'un ou plusieurs paramètres, en général inconnu. On note  $\theta$  la collection de paramètres,  $\Theta$  l'ensemble des valeurs que peut prendre  $\theta$  et  $\mu_\theta$  la loi correspondante. On se restreint au cas où  $\mu_\theta$  est :

1. soit à valeurs dans  $\mathbb{R}$ , à densité notée  $p_\theta$  ;
2. soit à valeurs dans un ensemble fini ou dénombrable  $E$ , de loi caractérisée par la donnée des nombres  $\{p_\theta(x)\}_{x \in E}$ .

Afin de déterminer la collection des paramètres avec une certitude partielle, vraie avec une certaine probabilité, on réalise un certain nombre d'expériences aléatoires indépendantes de loi  $\mu_\theta$ .

### Définition.

- Soit  $\mu_\theta$  une loi de probabilité sur un ensemble fini ou dénombrable  $E$  ou à densité sur  $\mathbb{R}$ . Soit  $n \geq 1$ . Un *échantillon de taille  $n$  de loi  $\mu_\theta$*  est un  $n$ -uplet  $(X_1, \dots, X_n)$  de variables aléatoires indépendantes et de même loi  $\mu_\theta$ .
- Une *observation* ou *réalisation*  $x = (x_1, \dots, x_n)$  est formée des valeurs de réalisation  $X_1(\omega), \dots, X_n(\omega)$  des variables aléatoires  $X_1, \dots, X_n$ , en un point  $\omega$ .

### EXEMPLE.

1. Dans l'exemple introductif, la loi  $\mu_\theta$  est une loi de Bernoulli de paramètre  $\theta$  inconnu, et  $\Theta = ]0, 1[$ . Nous avons introduit les variables aléatoires  $X_1, \dots, X_n$ , indépendantes, de même loi de Bernoulli de paramètre inconnu  $\theta$ , et avons observé des réalisations de ces variables aléatoires.
2. On pourrait imaginer un exemple où la loi  $\mu_\theta$  est une loi gaussienne de moyenne  $m$  et de variance  $\sigma^2$  inconnues. Dans ce cas,  $\theta = (m, \sigma^2)$ , et  $\Theta = \mathbb{R} \times \mathbb{R}_+^*$ .

### Définition.

On appelle *modèle statistique paramétrique* un triplet  $(\mathcal{X}, \mathcal{A}, (\mathbb{P}_\theta)_{\theta \in \Theta})$  où  $\mathcal{X}$  s'appelle l'*espace des observations*,  $\mathcal{A}$  est une tribu sur  $\mathcal{X}$  et  $(\mathbb{P}_\theta)_{\theta \in \Theta}$  est une famille de probabilités sur  $(\mathcal{X}, \mathcal{A})$ .

*Remarque.* Si  $X = (X_1, \dots, X_n)$  est un échantillon de taille  $n$  de loi  $\mu_\theta$ , alors  $\mathcal{X}$  est l'espace dans lequel  $X$  prend ses valeurs, celui dans lequel "vivent" les observations  $x = (x_1, \dots, x_n)$ , la loi  $\mathbb{P}_\theta$  est la loi  $\mathbb{P}_\theta^X$  du vecteur aléatoire  $X$ . Comme les variables aléatoires  $X_1, \dots, X_n$  sont indépendantes, la loi de probabilité  $\mathbb{P}_\theta$  est entièrement déterminée par la loi  $\mu_\theta$ . Plus particulièrement :

1. Si la loi  $\mu_\theta$  est à densité  $p_\theta$  sur  $\mathbb{R}$ , alors  $\mathcal{X} = \mathbb{R}^n$ ,  $\mathcal{A}$  est la tribu borélienne de  $\mathbb{R}^n$  et pour tout "bon" sous-ensemble  $B$  de  $\mathcal{X}$  :

$$\mathbb{P}_\theta[B] = \mathbb{P}_\theta^X[B] = \int_{\mathbb{R}^n} \mathbb{I}_B(x_1, \dots, x_n) p_\theta(x_1) \cdots p_\theta(x_n) dx_1 \cdots dx_n.$$

2. Si  $\mu_\theta$  est définie sur un ensemble fini ou dénombrable  $E$ , alors  $\mathcal{X} = E^n$ ,  $\mathcal{A} = \mathcal{P}(\mathcal{X})$  et pour tout  $x = (x_1, \dots, x_n)$  appartenant à  $\mathcal{X}$  :

$$\mathbb{P}_\theta[\{(x_1, \dots, x_n)\}] = \mathbb{P}_\theta^X[\{(x_1, \dots, x_n)\}] = p_\theta(x_1) \cdots p_\theta(x_n).$$

On note  $\mathbb{E}_\theta$  et  $\text{Var}_\theta$  l'espérance et la variance par rapport à la probabilité  $\mathbb{P}_\theta$ .

L'objectif de la statistique paramétrique est de répondre à l'une des questions suivantes :

1. problèmes d'estimation : estimer le paramètre inconnu  $\theta$ , ou une fonction réelle  $f(\theta)$  de ce paramètre, soit par une valeur unique (estimation ponctuelle), soit par un intervalle de confiance, *i.e.* en répondant à la question "avec quelle probabilité  $f(\theta)$  appartient-il à  $[a, b]$ " ?
2. test d'hypothèse : étant donné un sous-ensemble  $H_0$  de  $\Theta$ , il s'agit de décider si  $\theta$  appartient ou non à  $H_0$  ; on dit que l'on teste l'hypothèse  $H_0$  contre l'hypothèse contraire  $H_1$ .

## 6.3 Estimateur

Soit  $(X_1, \dots, X_n)$  un échantillon de taille  $n$  de loi  $\mu_\theta$ . On note  $\mathbb{P}_\theta$  la loi jointe du vecteur  $(X_1, \dots, X_n)$  et  $x = (x_1, \dots, x_n)$ ,  $x \in \mathcal{X}$ , une observation.

**Définition.** Soit  $f$  une fonction définie sur  $\Theta$  à valeurs réelles. Un *estimateur de  $f(\theta)$*  est une variable aléatoire  $T_n(X_1, \dots, X_n)$  à valeurs dans  $f(\Theta)$ , utilisée pour estimer  $f(\theta)$ . Un *estimateur ponctuel* est la valeur de cet estimateur en un point  $\omega$ .

Cette définition est assez générale et ne précise que peu le choix d'un estimateur pour  $f(\theta)$ . Il est souhaitable qu'il ait les propriétés supplémentaires suivantes : consistance, absence de biais (voir définitions ci-dessous), minimisation de l'erreur moyenne. À noter que le fait d'être sans biais n'est pas essentiel, car il arrive que respecter cette contrainte entraîne la perte d'autres propriétés importantes. Il est également utile que l'estimateur suive une loi connue.

**Définition.** Un estimateur  $T_n(X_1, \dots, X_n)$  de  $f(\theta)$  est *consistant* ou *convergent* s'il converge en probabilité vers  $f(\theta)$  :

$$\forall \theta \in \Theta, \forall \varepsilon > 0, \lim_{n \rightarrow +\infty} \mathbb{P}_\theta[|T_n(X_1, \dots, X_n) - f(\theta)| > \varepsilon] = 0.$$

*Remarque.* Cette notion n'est pas explicitement au programme. Néanmoins, il n'est pas inutile de la mentionner car elle permet de mieux comprendre certains théorèmes étudiés.

**Définition.** Si l'estimateur  $T_n(X_1, \dots, X_n)$  de  $f(\theta)$  est intégrable, son *biais* est défini par :

$$\forall \theta \in \Theta, \mathbb{E}_\theta[T_n(X_1, \dots, X_n)] - f(\theta).$$

On dit que  $T_n(X_1, \dots, X_n)$  est un estimateur *sans biais* si,

$$\forall \theta \in \Theta, \mathbb{E}_\theta[T_n(X_1, \dots, X_n)] = f(\theta).$$

*Remarque.* La notion de biais d'un estimateur apparaît dans les programmes de BTS. Elle ne peut être exigible d'un élève, mais doit par contre être connue du professeur...

EXEMPLE. Soit  $(X_1, \dots, X_n)$  un échantillon de taille  $n$  de loi  $\mu_\theta$ . Notons  $m$  et  $\sigma^2$  la moyenne et la variance commune de l'échantillon.

1. **Moyenne empirique.** On souhaite estimer  $m = \mathbb{E}_\theta(X_1)$ . Un estimateur naturel est la *moyenne empirique*, notée  $M_n$  ou  $\bar{X}$ , définie par :

$$M_n = \bar{X} = \frac{1}{n} \sum_{k=1}^n X_k.$$

La moyenne empirique observée,  $\frac{1}{n} \sum_{k=1}^n x_k$ , est un estimateur ponctuel de la moyenne de l'échantillon (rappel :  $x_k = X_k(\omega)$ ).

EXERCICE. Montrer que la moyenne empirique est un estimateur sans biais et consistant de la moyenne  $m$ .

2. **Variance empirique.** On souhaite estimer  $\sigma^2 = \mathbb{E}_\theta[(X_1 - \mathbb{E}_\theta(X_1))^2]$ .
  - Si la moyenne  $m = \mathbb{E}_\theta(X_1)$  de l'échantillon est connue. Alors un estimateur naturel est  $V_n$  défini par :

$$V_n = \frac{1}{n} \sum_{k=1}^n (X_k - m)^2.$$

La valeur observée  $\frac{1}{n} \sum_{k=1}^n (x_k - m)^2$  est un estimateur ponctuel de la variance de l'échantillon.

EXERCICE. Montrer que  $\frac{1}{n} \sum_{k=1}^n (x_k - m)^2$  est également un estimateur sans biais de la variance de l'échantillon.

- Si la moyenne de l'échantillon est inconnue. Un estimateur naturel est la variance empirique, notée  $S_n^2$ , et définie par :

$$S_n^2 = \frac{1}{n-1} \sum_{k=1}^n (X_k - M_n)^2.$$

Alors la quantité  $\frac{1}{n-1} \sum_{k=1}^n \left( x_k - \left( \frac{1}{n} \sum_{i=1}^n x_i \right) \right)^2$  est un estimateur ponctuel de la variance de l'échantillon.

PROPRIÉTÉ 13. La variable aléatoire  $S_n^2 = \frac{1}{n-1} \sum_{k=1}^n (X_k - M_n)^2$  est un estimateur sans biais de la variance de l'échantillon.

*Démonstration.* Comme les variables aléatoires  $X_1, \dots, X_n$  sont identiquement distribuées :

$$\mathbb{E}_\theta[S_n^2] = \frac{n}{n-1} \mathbb{E}_\theta[(X_1 - M_n)^2].$$

Calculons donc  $\mathbb{E}_\theta[(X_1 - M_n)^2]$ . Par linéarité de l'espérance :

$$\begin{aligned} \mathbb{E}_\theta[(X_1 - M_n)^2] &= \mathbb{E} \left[ \left( X_1 - \frac{1}{n} \sum_{i=1}^n X_i \right)^2 \right] \\ &= \mathbb{E}_\theta[X_1^2] - \frac{2}{n} \sum_{i=1}^n \mathbb{E}_\theta[X_1 X_i] + \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \mathbb{E}_\theta[X_i X_j]. \end{aligned}$$

Comme les variables aléatoires sont indépendantes et identiquement distribuées,

$$\mathbb{E}(X_i X_j) = \begin{cases} \mathbb{E}(X_1^2) & \text{si } i = j \\ \mathbb{E}(X_1)^2 & \text{si } i \neq j, \end{cases}$$

ainsi :

$$\begin{aligned} -\frac{2}{n} \sum_{i=1}^n \mathbb{E}_\theta(X_1 X_i) &= -\frac{2}{n} [\mathbb{E}_\theta(X_1^2) + (n-1)\mathbb{E}_\theta(X_1)^2] \\ \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \mathbb{E}_\theta(X_i X_j) &= \frac{1}{n^2} \sum_{i=1}^n [\mathbb{E}_\theta(X_1^2) + (n-1)\mathbb{E}_\theta(X_1)^2] \\ &= \frac{1}{n} [\mathbb{E}_\theta(X_1^2) + (n-1)\mathbb{E}_\theta(X_1)^2], \end{aligned}$$

$$\begin{aligned} \text{d'où } \mathbb{E}_\theta[(X_1 - M_n)^2] &= \mathbb{E}_\theta[X_1^2] - \frac{1}{n} [\mathbb{E}_\theta(X_1^2) + (n-1)\mathbb{E}_\theta(X_1)^2] \\ &= \frac{n-1}{n} (\mathbb{E}_\theta[X_1^2] - \mathbb{E}_\theta[X_1]^2). \end{aligned}$$

On conclut que  $\mathbb{E}_\theta[S_n^2] = \mathbb{E}_\theta[X_1^2] - \mathbb{E}_\theta[X_1]^2 = \sigma^2$  et que la variance empirique est un estimateur sans biais de la variance de l'échantillon. □

**EXERCICE 6.1.** Pour tout nombre réel  $\theta$ , on définit l'application  $p_\theta$  sur  $\mathbb{R}$  par :

$$p_\theta(x) = \begin{cases} 0 & \text{si } x < \theta \\ e^{-(x-\theta)} & \text{si } x \geq \theta. \end{cases}$$

1. Montrer que, pour tout nombre réel  $\theta$ , l'application  $p_\theta$  est une densité de probabilité sur  $\mathbb{R}$ .
2. Soit  $X_0$  une variable aléatoire de densité  $p_\theta$ . Montrer que  $X_0$  admet un moment d'ordre deux, et expliciter son espérance et sa variance.
3. Déterminer la fonction de répartition  $F_0$  de  $X_0$ .
4. Soit  $n \geq 1$  et  $(X_1, \dots, X_n)$  un échantillon de taille  $n$  de la loi de  $X_0$ .

On pose  $U_n = \min(X_1, \dots, X_n)$  et  $M_n = \frac{1}{n} \sum_{i=1}^n X_i$ .

- (a) Pour tout nombre réel  $t$ , calculer  $\mathbb{P}[U_n > t]$ . En déduire la loi de  $U_n$ .
- (b) Montrer que  $M_n - 1$  et  $U_n - \frac{1}{n}$  sont des estimateurs sans biais de  $\theta$ .

*Remarque.* On ne peut se contenter des estimations ponctuelles de la moyenne et de la variance de notre échantillon, même si ces estimations sont obtenues à partir d'estimateurs sans biais de ces paramètres. En effet, nous n'avons aucune indication sur la façon dont ces valeurs s'écartent des "vraies" moyenne et variance de l'échantillon. Voir paragraphe 6.5.

## 6.4 Rappels sur quelques lois utiles en statistique

Voici des définitions et propriétés que nous avons déjà vues, mais que nous rappelons pour leur utilité en statistiques. On se réfère aux exercices 4.26 et 4.27 de la section 4.6.

**Rappel.** Si  $X_1, \dots, X_n$  sont des variables aléatoires indépendantes, gaussiennes de moyennes  $m_1, \dots, m_n$ , et de variance  $\sigma_1^2, \dots, \sigma_n^2$ , respectivement, leur somme  $\sum_{k=1}^n X_k$  suit une loi gaussienne de moyenne  $\sum_{k=1}^n m_k$  et de variance  $\sum_{k=1}^n \sigma_k^2$ .

**Définition.** Soit  $d \geq 1$ . On dit que la variable aléatoire  $Y_d$  suit la loi du  $\chi^2$  à  $d$  degrés de liberté si elle est de la forme  $Y_d = \sum_{i=1}^d N_i^2$ , où  $N_1, \dots, N_d$  sont  $d$  variables aléatoires gaussiennes centrées réduites indépendantes.

**Propriété 14.** Soit  $Y_d$  une variable aléatoire qui suit une loi du  $\chi^2$  à  $d$  degrés de liberté. Alors :

1. elle admet une densité  $p_{Y_d}$  de la forme

$$\forall x \in \mathbb{R}, p_{Y_d}(x) = k_d x^{\frac{d}{2}-1} e^{-\frac{x}{2}} \mathbb{I}_{[0,+\infty[}(x),$$

où  $k_d = \frac{1}{2^{\frac{d}{2}} \Gamma(\frac{d}{2})}$ , et  $\Gamma$  désigne la fonction gamma ;

2. elle est intégrable et admet comme espérance  $\mathbb{E}[Y_d] = d$  ;
3. la fonction de répartition de  $Y_d$  n'a pas d'expression simple, elle est tabulée ;
4. si  $d > 30$ , pour tout intervalle  $I$  de  $\mathbb{R}$ , on a

$$\mathbb{P}[\sqrt{Y_d} - \sqrt{2d-1} \in I] \simeq \mathbb{P}[N \in I],$$

où  $N$  est une variable aléatoire gaussienne centrée réduite.

*Démonstration.* (Remarques)

1. Dans l'exercice 4.27, nous avons calculé la densité d'une variable aléatoire de loi du  $\chi^2$  à 1 degré de liberté.
2. Le fait que  $Y_d$  est intégrable découle de l'existence de la variance d'une loi gaussienne et du fait que  $Y_d$  est une somme finie de gaussiennes au carré. Le calcul découle de la linéarité de l'espérance.
4. Résultat admis, ainsi que le fait que cette approximation est excellente.

□

**Définition.** Soit  $d \geq 1$ . On dit que la variable aléatoire  $T_d$  suit la loi de Student à  $d$  degrés de liberté, si elle est de la forme  $N \sqrt{\frac{d}{Y_d}}$ , où  $N$  est une variable gaussienne centrée réduite, et où  $Y_d$  suit la loi du  $\chi^2$  à  $d$  degrés de liberté et est indépendante de  $N$ .

**Propriété 15.** Soit  $T_d$  une variable aléatoire qui suit la loi de Student à  $d$  degrés de liberté. Alors :

1. elle admet une densité  $p_{T_d}$  de la forme

$$\forall x \in \mathbb{R}, p_{T_d}(x) = c_d \left(1 + \frac{x^2}{d}\right)^{-\frac{d+1}{2}},$$

où  $c_d = \frac{\Gamma(\frac{d+1}{2})}{\sqrt{\pi d}}$  et  $\Gamma$  désigne la fonction gamma ;

2. la fonction de répartition de  $T_d$  n'est pas simple même si elle se calcule; on a alors recours à des tables qui permettent de calculer des quantités du type  $\mathbb{P}[|T_d| \geq c]$ ;
3. pour tout réel  $x$ ,  $p_{T_d}(x)$  tend vers  $\frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}$  lorsque  $d$  tend vers  $+\infty$ ;  
si  $d \geq 30$ , pour tout intervalle  $I$  de  $\mathbb{R}$ , on a

$$\mathbb{P}[T_d \in I] \simeq \mathbb{P}[N \in I],$$

où  $N$  est une variable aléatoire gaussienne centrée réduite.

*Démonstration.* (Remarques) Dans l'exercice 4.27, nous avons calculé la densité d'une variable aléatoire de Student à  $d$  degrés de liberté. Les autres résultats sont admis.  $\square$

**Proposition 6.1.** Soit  $(X_1, \dots, X_n)$  un échantillon de taille  $n$  de loi normale de moyenne  $m$  et de variance  $\sigma^2$ . Rappelons que  $M_n$ ,  $V_n$  et  $S_n^2$  sont respectivement la moyenne empirique, la variance empirique dans le cas où  $m$  est connu, et la variance empirique dans le cas où la moyenne et la variance sont inconnues. On a alors :

1.  $\sqrt{n} \frac{M_n - m}{\sigma}$  suit une loi normale centrée réduite;
2.  $\frac{n}{\sigma^2} V_n = \sum_{k=1}^n \left( \frac{X_k - m}{\sigma} \right)^2$  suit une loi du  $\chi^2$  à  $n$  degrés de liberté;
3.  $\frac{n-1}{\sigma^2} S_n^2 = \sum_{k=1}^n \left( \frac{X_k - M_n}{\sigma} \right)^2$  suit une loi du  $\chi^2$  à  $n-1$  degrés de liberté et est indépendante de  $M_n$  (admis);
4.  $\sqrt{n} \frac{M_n - m}{S_n}$  suit une loi de Student à  $n-1$  degrés de liberté.

*Démonstration.* 4. Posons  $N = \sqrt{n} \frac{M_n - m}{\sigma}$  et  $Y = \frac{n-1}{\sigma^2} S_n^2$ . Alors  $N$  est une gaussienne centrée réduite,  $Y$  suit une loi du  $\chi^2$  à  $d = n-1$  degrés de liberté. De plus, d'après le point 3,  $N$  et  $Y$  sont indépendantes. Il s'ensuit que  $N \sqrt{\frac{d}{Y}} = \sqrt{n} \frac{M_n - m}{\sigma} \frac{\sigma}{S_n} = \sqrt{n} \frac{M_n - m}{S_n}$  suit une loi de Student à  $d$  degrés de liberté.  $\square$

## 6.5 Intervalles de confiance

On reprend les notations du paragraphe 6.2 :  $(X_1, \dots, X_n)$  est un échantillon de taille  $n$  de loi  $\mu_\theta$ ,  $\mathbb{P}_\theta$  est la loi jointe du vecteur  $(X_1, \dots, X_n)$ ,  $x = (x_1, \dots, x_n)$ ,  $x \in \mathcal{X}$ , est une observation.

Soit  $f$  une fonction réelle définie sur l'ensemble des paramètres  $\Theta$ . Plutôt que d'estimer ponctuellement  $f(\theta)$ , estimation qui est probablement voisine de  $f(\theta)$  mais pratiquement jamais égale à  $f(\theta)$ , on peut envisager de trouver un intervalle  $I(x_1, \dots, x_n)$  qui dépende de l'observation  $x = (x_1, \dots, x_n)$ , et qui permette de :

- répondre à la question “ $f(\theta)$  appartient-il à l'intervalle  $I(x_1, \dots, x_n)$ ” ?
- donner une précision numérique à cette réponse en disant que  $f(\theta)$  appartient à l'intervalle  $I(x_1, \dots, x_n)$  avec une probabilité supérieure à 0,9 ou 0,95, ou 0,99, ...

Cela conduit à la définition suivante :

**Définition.** Soit  $\alpha \in ]0, 1[$ . On appelle *intervalle de confiance au niveau  $\alpha$  pour  $f(\theta)$* , une famille d'intervalles  $(I(x_1, \dots, x_n) : (x_1, \dots, x_n) \in \mathcal{X})$  telle que :

$$\forall \theta \in \Theta, \mathbb{P}_\theta [f(\theta) \in I(X_1, \dots, X_n)] \geq \alpha.$$

*Remarque.* Attention ! Un intervalle de confiance est **une famille** d'intervalles **déterministes**, et dans la quantité  $\mathbb{P}_\theta [f(\theta) \in I(X_1, \dots, X_n)]$ , l'intervalle  $I(X_1, \dots, X_n)$  est aléatoire.

### 6.5.1 Intervalle de confiance à partir de l'inégalité de Bienaymé-Tchebychev

Supposons que  $\phi(X_1, \dots, X_n)$  soit un estimateur sans biais de  $f(\theta)$ . Nous souhaitons chercher, quand cela est possible, un intervalle de confiance au niveau  $\alpha$  pour  $f(\theta)$  de la forme :

$$\begin{aligned} \{I_\alpha(x_1, \dots, x_n) : (x_1, \dots, x_n) \in \mathcal{X}\} = \\ = \{[\phi(x_1, \dots, x_n) - C_\alpha, \phi(x_1, \dots, x_n) + C_\alpha] : (x_1, \dots, x_n) \in \mathcal{X}\}. \end{aligned}$$

Si l'estimateur  $\phi(X_1, \dots, X_n)$  satisfait à quelques hypothèses, on peut trouver un tel intervalle de confiance en utilisant l'inégalité de Bienaymé-Tchebychev.

**Proposition 6.2.** *Soit  $(X_1, \dots, X_n)$  un échantillon de taille  $n$  de loi  $\mu_\theta$ . Soit  $\mathbb{P}_\theta$  la loi jointe du vecteur aléatoire  $(X_1, \dots, X_n)$ . On considère  $\phi(X_1, \dots, X_n)$  un estimateur sans biais de  $f(\theta)$ . On suppose que, pour tout  $\theta \in \Theta$ ,  $\phi(X_1, \dots, X_n)$  admet une variance sous  $\mathbb{P}_\theta$ , et qu'il existe un nombre réel  $r$  tel que :*

$$\forall \theta \in \Theta, \text{Var}_\theta[\phi(X_1, \dots, X_n)] \leq r.$$

Soit  $\alpha \in ]0, 1[$ . Alors,

$$\left( \left[ \phi(x_1, \dots, x_n) - \sqrt{\frac{r}{1-\alpha}}; \phi(x_1, \dots, x_n) + \sqrt{\frac{r}{1-\alpha}} \right] : (x_1, \dots, x_n) \in \mathcal{X} \right)$$

est un intervalle de confiance pour  $f(\theta)$  au niveau  $\alpha$ .

*Démonstration.* Soit  $\alpha \in ]0, 1[$ . Notons  $X$  le vecteur aléatoire  $(X_1, \dots, X_n)$ .

Remarquons que :

$$\begin{aligned} \mathbb{P}_\theta(f(\theta) \in I_\alpha(X)) &= \mathbb{P}_\theta(f(\theta) \in ]\phi(X) - C_\alpha, \phi(X) + C_\alpha]) \\ &= \mathbb{P}_\theta(|\phi(X) - f(\theta)| < C_\alpha). \end{aligned}$$

Ainsi, nous cherchons  $C_\alpha$  telle que :

$$\forall \theta \in \Theta, \mathbb{P}_\theta(|\phi(X) - f(\theta)| < C_\alpha) \geq \alpha.$$

L'estimateur étant sans biais,  $f(\theta) = \mathbb{E}_\theta[\phi(X)]$ , et la variable aléatoire  $\phi(X)$  admettant une variance, nous pouvons utiliser l'inégalité de Bienaymé-Tchebychev. Prenons comme choix de  $\varepsilon$  :

$$\varepsilon = \sqrt{\frac{\text{Var}_\theta[\phi(X_1, \dots, X_n)]}{1-\alpha}}.$$

Ainsi :

$$\begin{aligned} \mathbb{P}_\theta[|\phi(X) - f(\theta)| \geq \varepsilon] &\leq \frac{\text{Var}_\theta[\phi(X)]}{\varepsilon^2} \\ \Leftrightarrow \mathbb{P}_\theta[|\phi(X) - f(\theta)| < \varepsilon] &\geq 1 - \frac{\text{Var}_\theta[\phi(X)]}{\varepsilon^2}, \text{ en passant au complémentaire} \\ \Leftrightarrow \mathbb{P}_\theta[|\phi(X) - f(\theta)| < \varepsilon] &\geq 1 - (1 - \alpha) = \alpha, \text{ par définition de } \varepsilon. \end{aligned}$$

Posons,  $C_\alpha = \sqrt{\frac{r}{1-\alpha}}$ . Alors par hypothèse,  $\varepsilon \leq C_\alpha$ , de sorte que  $\{|\phi(X) - f(\theta)| < \varepsilon\} \subset \{|\phi(X) - f(\theta)| < C_\alpha\}$ . Donc :

$$\mathbb{P}_\theta[|\phi(X) - f(\theta)| < C_\alpha] \geq \mathbb{P}_\theta[|\phi(X) - f(\theta)| < \varepsilon] \geq \alpha.$$

Ainsi, nous avons trouvé  $C_\alpha$  tel que  $\forall \theta \in \Theta, \mathbb{P}_\theta(|\phi(X) - f(\theta)| < C_\alpha) \geq \alpha$ . □

Remarque.

1. C'est exactement de cette façon que l'on a procédé dans l'exemple introductif de ce chapitre (cf. paragraphe 6.1) :
  - $(X_1, \dots, X_n)$  est un échantillon de taille  $n$  de loi de Bernoulli de paramètre  $\theta$  inconnu ;  
 $\Theta = ]0, 1[$ ,  $\mathcal{X} = \{0, 1\}^n$  ;
  - $\forall \theta \in \Theta$ ,  $f(\theta) = \theta$ , et la moyenne empirique,  $\phi(X_1, \dots, X_n) = M_n$ , est un estimateur sans biais de la moyenne  $\theta$  ;
  - nous avons montré que  $\text{Var}[M_n] \leq \frac{1}{4n}$ , d'où  $r = \frac{1}{4n}$  ;
  - posons  $\alpha = 0,95$ , alors

$$\left( \frac{1}{n} \sum_{k=1}^n x_k - \sqrt{\frac{5}{n}}, \frac{1}{n} \sum_{k=1}^n x_k + \sqrt{\frac{5}{n}} \left[ : (x_1, \dots, x_n) \in \{0, 1\}^n \right) \right),$$

est un intervalle de confiance pour  $\theta$  au niveau 0,95. C'est-à-dire que  $\theta$  appartient à 95% de ces intervalles, et que  $\theta$  n'appartient pas à 5% de ces intervalles.

Supposons que l'on ait effectivement fait un sondage sur  $n = 500$  parisiens pour en estimer la proportion de chauves, et que l'on ait trouvé 75 chauves. Ainsi, on obtient

$\frac{1}{n} \sum_{k=1}^n x_k = \frac{75}{500} = 0,15$ . On dira que la moyenne  $\theta$  du nombre de chauves parisiens est comprise entre  $0,15 - \sqrt{\frac{5}{500}} = 0,05$  et  $0,15 + \sqrt{\frac{5}{500}} = 0,25$  avec une *confiance au moins égale* à 0,95.

2. Estimation par intervalle de confiance de la fréquence dans le cas d'une loi de Bernoulli de paramètre inconnu  $\theta$ . Plus généralement, pour  $\alpha \in ]0, 1[$ ,

$$\left( \frac{1}{n} \sum_{k=1}^n x_k - \frac{1}{2\sqrt{n(1-\alpha)}}, \frac{1}{n} \sum_{k=1}^n x_k + \frac{1}{2\sqrt{n(1-\alpha)}} \left[ ; (x_1, \dots, x_n) \in \{0, 1\}^n \right) \right)$$

est un intervalle de confiance au niveau  $\alpha$  pour la moyenne inconnue  $\theta$  de notre échantillon de taille  $n$ .

3. Remarquons que la longueur des intervalles de confiance obtenus dans ce cas est  $\frac{1}{\sqrt{n(1-\alpha)}}$ . Ainsi, pour diviser par 2 la longueur de l'intervalle, il faut multiplier  $n$  par 4. Par ailleurs, plus  $\alpha$  est "proche" de 1, plus la longueur de l'intervalle est grande. Il faut donc un compromis entre le niveau de confiance  $\alpha$  et la longueur de l'intervalle.

EXERCICE 6.2. On reprend les estimateurs  $M_n$  et  $U_n$  (cf. l'exercice 6.1).

1. Construire un intervalle de confiance pour  $\theta$  au niveau 0,95 à partir de  $M_n$ .
2. À l'aide de la variable  $U_n$ , proposer un autre intervalle de confiance pour  $\theta$  au niveau 0,95. Comparer les deux intervalles de confiance obtenus (on comparera leurs longueurs).

### 6.5.2 Construction d'un intervalle de confiance pour la moyenne

Soit  $(X_1, \dots, X_n)$  un échantillon de taille  $n$  et de loi  $\mu_\theta$ . Soit  $\mathbb{P}_\theta$  la loi jointe du vecteur aléatoire  $(X_1, \dots, X_n)$ . On suppose que la variable aléatoire  $X_1$  admet une espérance et on note  $m$  l'espérance commune à  $X_1, \dots, X_n$ .

On considère la moyenne empirique  $M_n = \frac{1}{n} \sum_{k=1}^n X_k$  comme estimateur de la moyenne  $m$  et on fixe un niveau  $\alpha \in ]0, 1[$ . Nous allons chercher, quand cela est possible, un intervalle de confiance



au niveau  $\alpha$  pour  $m$  de la forme :

$$\begin{aligned} & \{I_\alpha(x_1, \dots, x_n) : (x_1, \dots, x_n) \in \mathcal{X}\} \\ &= \left\{ \left[ \frac{1}{n} \sum_{k=1}^n x_k - C_\alpha, \frac{1}{n} \sum_{k=1}^n x_k + C_\alpha \right] : (x_1, \dots, x_n) \in \mathcal{X} \right\}. \end{aligned}$$

En observant que :

$$\mathbb{P}_\theta(m \in I_\alpha(X_1, \dots, X_n)) = \mathbb{P}_\theta(m \in ]M_n - C_\alpha, M_n + C_\alpha]) = \mathbb{P}_\theta(|M_n - m| < C_\alpha),$$

nous cherchons  $C_\alpha$  telle que :

$$\forall \theta \in \Theta, \mathbb{P}_\theta(|M_n - m| < C_\alpha) \geq \alpha. \quad (6.1)$$

Afin de continuer, nous avons besoin d'informations supplémentaires sur la loi de l'échantillon ou sur sa taille.

• **Échantillon de taille  $n$  de loi normale  $\mathcal{N}(m, \sigma^2)$  où  $\sigma$  est connu**

**Proposition 6.3.** Soit  $(X_1, \dots, X_n)$  un échantillon de taille  $n$  de loi  $\mathcal{N}(m, \sigma^2)$  où  $\sigma^2$  est connu et  $m$  est inconnu. Soit  $\alpha \in ]0, 1[$ . Alors,

$$\left( \left[ \frac{1}{n} \sum_{k=1}^n x_k - \frac{\sigma \Pi^{-1}\left(\frac{1+\alpha}{2}\right)}{\sqrt{n}}; \frac{1}{n} \sum_{k=1}^n x_k + \frac{\sigma \Pi^{-1}\left(\frac{1+\alpha}{2}\right)}{\sqrt{n}} \right] : (x_1, \dots, x_n) \in \mathbb{R}^n \right)$$

est un intervalle de confiance pour  $m$  au niveau  $\alpha$ .

*Démonstration.* D'après l'équation (6.1), nous cherchons  $C_\alpha$  tel que :

$$\forall \theta \in \Theta, \mathbb{P}_\theta(|M_n - m| < C_\alpha) \geq \alpha.$$

Dans ce cas  $\Theta = \mathbb{R}$ ,  $\theta = m$ . Or, sous  $\mathbb{P}_\theta$ , la variable aléatoire  $\sqrt{n} \frac{M_n - m}{\sigma}$  suit une loi gaussienne centrée réduite. Notons  $N$  une variable aléatoire de loi  $\mathcal{N}(0, 1)$ . Alors,

$$\begin{aligned} \mathbb{P}_\theta(|M_n - m| < C_\alpha) &= \mathbb{P}_\theta \left[ \sqrt{n} \left| \frac{M_n - m}{\sigma} \right| < \frac{C_\alpha \sqrt{n}}{\sigma} \right] = \mathbb{P} \left[ |N| < \frac{C_\alpha \sqrt{n}}{\sigma} \right] \\ &= 2 \Pi \left( \frac{C_\alpha \sqrt{n}}{\sigma} \right) - 1, \end{aligned}$$

d'où il suffit de choisir  $C_\alpha$  tel que :

$$2 \Pi \left( \frac{C_\alpha \sqrt{n}}{\sigma} \right) - 1 = \alpha \Leftrightarrow \Pi \left( \frac{C_\alpha \sqrt{n}}{\sigma} \right) = \frac{1 + \alpha}{2}.$$

Comme la fonction de répartition  $\Pi$  de la loi normale est bijective de  $\mathbb{R}$  dans  $]0, 1[$ , on peut définir  $\Pi^{-1}\left(\frac{1+\alpha}{2}\right)$ . De la stricte croissance de  $\Pi$ , on déduit que  $\Pi^{-1}\left(\frac{1+\alpha}{2}\right) > \Pi^{-1}\left(\frac{1}{2}\right) = 0$ . Ainsi, il suffit de choisir  $C_\alpha = \frac{\sigma}{\sqrt{n}} \Pi^{-1}\left(\frac{1+\alpha}{2}\right)$ .  $\square$

*Remarque.* Il s'agit d'un cas particulier de la méthode suivante, dite de *la fonction pivotale*. On suppose l'existence d'une application  $g : \mathcal{X} \times \mathbb{R} \rightarrow \mathbb{R}$  telle que, si on note  $X = (X_1, \dots, X_n)$  :

1. pour tout  $\theta \in \Theta$ ,  $g(X, f(\theta))$  suit une loi de densité  $p$ , indépendante de  $\theta$ ;
2. pour tout  $x \in \mathcal{X}$ , l'application  $y \in \mathbb{R} \mapsto g(x, y)$  est continue strictement monotone.

Si  $a$  et  $b$  sont deux nombres réels tels que  $\alpha = \int_a^b p(y) dy$ , alors, pour tout  $\theta \in \Theta$  :

$$\mathbb{P}_\theta [g(X, f(\theta)) \in ]a, b[ ] = \int_a^b p(y) dy = \alpha.$$

Grâce à la deuxième condition, il existe des applications réelles  $A$  et  $B$  définies sur  $\mathcal{X}$  telles que  $\{g(X, f(\theta)) \in ]a, b[ \}$  soit égal à  $\{f(\theta) \in ]A(X), B(X)[ \}$ . La famille  $(]A(x); B(x)[)_{x \in \mathcal{X}}$  est donc un intervalle de confiance pour  $f(\theta)$  au niveau  $\alpha$ . Dans le cas de l'échantillon gaussien de variance connue, l'application  $g : \mathbb{R}^n \times \mathbb{R}$  définie par  $g(x, y) = \frac{1}{\sigma\sqrt{n}} \left( \sum_{k=1}^n x_k - ny \right)$  satisfait aux deux conditions (avec  $f(\theta) = \theta$ ), et  $p$  est la densité de la gaussienne centrée réduite.

EXERCICE 6.3. *D'après Phan-Rowenczyk.*

On modélise la durée de vie d'ampoules électriques par une loi normale de moyenne  $m$  inconnue et d'écart-type  $\sigma = 100$ .

1. On effectue une observation de la durée de vie sur  $n = 50$  lampes. Déterminer un intervalle de confiance pour la moyenne  $m$  au niveau 0,95.
2. Quelle doit être la taille de l'échantillon pour que la longueur de l'intervalle de confiance soit inférieure à 20 heures ?

• **Échantillon de taille  $n$  de loi normale  $\mathcal{N}(m, \sigma^2)$  où  $\sigma$  est inconnu**

**Proposition 6.4.** *Soit  $(X_1, \dots, X_n)$  un échantillon de taille  $n$  de loi  $\mathcal{N}(m, \sigma^2)$  où  $\sigma$  est inconnu. Soit  $\alpha \in ]0, 1[$ . Pour une observation  $(x_1, \dots, x_n)$ , notons l'écart-type empirique observé*

$$s_n = \sqrt{\frac{1}{n-1} \sum_{k=1}^n \left( x_k - \frac{1}{n} \sum_{i=1}^n x_i \right)^2}.$$

Soit  $T_{n-1}$  une variable aléatoire de loi de Student de paramètre  $n-1$ , et  $d_{1-\alpha}$  tel que  $\mathbb{P}[|T_{n-1}| > d_{1-\alpha}] = 1 - \alpha$ . Alors,

$$\left( \frac{1}{n} \sum_{k=1}^n x_k - d_{1-\alpha} \frac{s_n}{\sqrt{n}}; \frac{1}{n} \sum_{k=1}^n x_k + d_{1-\alpha} \frac{s_n}{\sqrt{n}} \right) \left[ : (x_1, \dots, x_n) \in \mathbb{R}^n \right)$$

est un intervalle de confiance pour  $m$  au niveau  $\alpha$ .

*Démonstration.* La démonstration de la proposition précédente repose sur le fait que l'on connaît la loi de la variable aléatoire  $\sqrt{n} \frac{M_n - m}{\sigma}$  sous  $\mathbb{P}_\theta$ . Ici, on ne connaît pas la variance  $\sigma$ . Il est donc naturel de construire une région critique à partir de la variable aléatoire  $\sqrt{n} \frac{M_n - m}{S_n}$  et de chercher une constante  $c_\alpha$  telle que :

$$\forall \theta \in \Theta, \mathbb{P}_\theta \left[ \left| \sqrt{n} \frac{M_n - m}{S_n} \right| < c_\alpha \right] \geq \alpha.$$

Or, sous  $\mathbb{P}_\theta$ , la variable aléatoire  $\sqrt{n} \frac{M_n - m}{S_n}$  suit une loi de Student à  $n-1$  degrés de liberté, où  $S_n^2 = \frac{1}{n-1} \sum_{k=1}^n (X_k - M_n)^2$  est la variance empirique. Notons  $T_{n-1}$  une variable aléatoire de loi de Student à  $n-1$  degrés de liberté. Alors,

$$\begin{aligned} \mathbb{P}_\theta \left[ \left| \sqrt{n} \frac{M_n - m}{S_n} \right| < c_\alpha \right] &= \mathbb{P} [|T_{n-1}| < c_\alpha] \\ &= 1 - \mathbb{P} [|T_{n-1}| \geq c_\alpha]. \end{aligned}$$

Il suffit donc de choisir  $c_\alpha$  tel que :

$$1 - \mathbb{P}[|T_{n-1}| \geq c_\alpha] = \alpha \Leftrightarrow \mathbb{P}[|T_{n-1}| \geq c_\alpha] = 1 - \alpha.$$

Soit  $d_{1-\alpha}$  tel que  $\mathbb{P}[|T_{n-1}| > d_{1-\alpha}] = 1 - \alpha$ . Ainsi, il suffit de choisir  $c_\alpha = d_{1-\alpha}$ .  $\square$

• **Échantillon de taille  $n$ ,  $n > 30$ , de variance connue**

**Proposition 6.5.** Soit  $(X_1, \dots, X_n)$  un échantillon de taille  $n$ ,  $n > 30$ , de loi  $\mu_\theta$  admettant un moment d'ordre 2. On suppose la variance de l'échantillon connue et on la note  $\sigma^2$ . Soit  $\alpha \in ]0, 1[$ . Alors,

$$\left( \frac{1}{n} \sum_{k=1}^n x_k - \frac{\sigma \Pi^{-1}(\frac{1+\alpha}{2})}{\sqrt{n}}; \frac{1}{n} \sum_{k=1}^n x_k + \frac{\sigma \Pi^{-1}(\frac{1+\alpha}{2})}{\sqrt{n}} \right] ; (x_1, \dots, x_n) \in \mathcal{X}$$

est un intervalle de confiance pour  $\theta$  au niveau (environ) égal à  $\alpha$ .

*Démonstration.* La démonstration est analogue à celle faite dans le cas de l'échantillon gaussien. Notons  $m$  la moyenne inconnue de l'échantillon. Sous ces hypothèses, on peut appliquer le théorème central limite et approcher ( $n$  est grand)  $\sqrt{n} \frac{M_n - m}{\sigma}$  par une variable aléatoire gaussienne de moyenne 0 et variance 1.  $\square$

*Remarque.* Il s'agit ici d'une approximation : la probabilité que  $m$  appartienne à l'intervalle  $\left[ M_n - \frac{\sigma \Pi^{-1}(\frac{1+\alpha}{2})}{\sqrt{n}}; M_n + \frac{\sigma \Pi^{-1}(\frac{1+\alpha}{2})}{\sqrt{n}} \right]$  est à peu près égale à  $\alpha$ , d'où la formulation de l'énoncé.

• **Échantillon de taille  $n$  de Bernoulli de moyenne inconnue lorsque la loi binomiale est approchable par une loi normale**

**Proposition 6.6.** Soit  $(X_1, \dots, X_n)$  un échantillon de taille  $n$  de loi de Bernoulli de paramètre inconnu  $\theta$ . On suppose que les paramètres  $n$  et  $\theta$  sont tels que la variable  $S_n = \sum_{k=1}^n X_k$  de loi binomiale est approchable par une loi normale (classiquement,  $n \geq 10$ , et  $n\theta$  et  $n(1-\theta)$  dépassent quelques unités). Soit  $\alpha \in ]0, 1[$ , alors

$$\left( \frac{1}{n} \sum_{k=1}^n x_k - \frac{\Pi^{-1}(\frac{1+\alpha}{2})}{2\sqrt{n}}; \frac{1}{n} \sum_{k=1}^n x_k + \frac{\Pi^{-1}(\frac{1+\alpha}{2})}{2\sqrt{n}} \right] ; (x_1, \dots, x_n) \in \{0, 1\}^n$$

est un intervalle de confiance pour la moyenne  $\theta$  de l'échantillon au niveau (environ) égal à  $\alpha$ .

*Démonstration.* D'après l'équation (6.1), nous cherchons  $C_\alpha$  tel que :

$$\forall \theta \in \Theta, \mathbb{P}_\theta(|M_n - m| < C_\alpha) \geq \alpha.$$

$$\begin{aligned} \mathbb{P}_\theta(|M_n - \theta| < C_\alpha) &= \mathbb{P}_\theta \left( \left| \frac{S_n - n\theta}{\sqrt{n \text{Var}(X_1)}} \right| < \frac{C_\alpha \sqrt{n}}{\sqrt{\text{Var}(X_1)}} \right) \\ &= \mathbb{P}_\theta \left( \left| \frac{S_n - \mathbb{E}(S_n)}{\sqrt{\text{Var}(S_n)}} \right| < \frac{C_\alpha \sqrt{n}}{\sqrt{\theta(1-\theta)}} \right), \text{ car } \text{Var}(X_1) = \theta(1-\theta), \\ &\geq \mathbb{P}_\theta \left( \left| \frac{S_n - \mathbb{E}(S_n)}{\sqrt{\text{Var}(S_n)}} \right| < 2C_\alpha \sqrt{n} \right), \text{ car } \theta(1-\theta) \leq \frac{1}{4} \text{ sur } [0, 1]. \end{aligned}$$

Comme  $n \geq 10$  et que  $n\theta$  et  $n(1 - \theta)$  dépassent quelques unités, on est dans les conditions d'application du théorème central limite. On en déduit :

$$\mathbb{P}_\theta \left( \left| \frac{S_n - \mathbb{E}(S_n)}{\sqrt{\text{Var}(S_n)}} \right| < 2C_\alpha \sqrt{n} \right) \simeq 2 \Pi(2C_\alpha \sqrt{n}) - 1.$$

Ainsi, si on pose  $C_\alpha = \frac{1}{2\sqrt{n}} \Pi^{-1} \left( \frac{\alpha+1}{2} \right)$ , on obtient, pour tout  $\theta \in \Theta$  :

$$\alpha \simeq \mathbb{P}_\theta \left( \left| \frac{S_n - \mathbb{E}(S_n)}{\sqrt{\text{Var}(S_n)}} \right| < 2C_\alpha \sqrt{n} \right) \text{ et } \mathbb{P}_\theta \left( \left| \frac{S_n - \mathbb{E}(S_n)}{\sqrt{\text{Var}(S_n)}} \right| < 2C_\alpha \sqrt{n} \right) \leq \mathbb{P}_\theta(|M_n - \theta| < C_\alpha).$$

□

*Remarque.* On a vu que, sous les hypothèses ci-dessus permettant d'approcher la loi binomiale par la loi de Gauss,  $\mathbb{P}_\theta \left[ \sqrt{n} \left| \frac{M_n - \theta}{\sqrt{\theta(1-\theta)}} \right| < \Pi^{-1} \left( \frac{1+\alpha}{2} \right) \right] \simeq \alpha$ , c'est-à-dire

$$\mathbb{P}_\theta \left[ \theta \in ]M_n - \frac{\Pi^{-1} \left( \frac{1+\alpha}{2} \right) \sqrt{\theta(1-\theta)}}{\sqrt{n}}; M_n + \frac{\Pi^{-1} \left( \frac{1+\alpha}{2} \right) \sqrt{\theta(1-\theta)}}{\sqrt{n}}[ \right] \simeq \alpha.$$

On pourrait donc avoir envie de considérer des intervalles de confiance de la forme

$$\left[ \frac{1}{n} \sum_{k=1}^n x_k - \frac{\Pi^{-1} \left( \frac{1+\alpha}{2} \right) \sqrt{\theta(1-\theta)}}{\sqrt{n}}; \frac{1}{n} \sum_{k=1}^n x_k + \frac{\Pi^{-1} \left( \frac{1+\alpha}{2} \right) \sqrt{\theta(1-\theta)}}{\sqrt{n}} \right],$$

en remplaçant le paramètre inconnu  $\theta$ . On trouve dans les livres deux types d'intervalles de confiance assez curieux :

1. on remplace  $\theta$  par l'estimateur ponctuel de la moyenne d'un échantillon de Bernoulli, à savoir par  $\bar{x} = \frac{1}{n} \sum_{k=1}^n x_k$ ; l'intervalle de confiance proposé devient :

$$I(x_1, \dots, x_n) = \left[ \bar{x} - \frac{\Pi^{-1} \left( \frac{1+\alpha}{2} \right) \sqrt{\bar{x}(1-\bar{x})}}{\sqrt{n}}; \bar{x} + \frac{\Pi^{-1} \left( \frac{1+\alpha}{2} \right) \sqrt{\bar{x}(1-\bar{x})}}{\sqrt{n}} \right];$$

2. on remplace  $\theta(1 - \theta)$  qui est la variance de l'échantillon de Bernoulli, par l'estimateur ponctuel de la variance lorsque la moyenne est inconnue, à savoir par  $s^2 = \frac{1}{n-1} \sum_{k=1}^n (x_k - \bar{x})^2$ ; l'intervalle de confiance proposé devient :

$$I(x_1, \dots, x_n) = \left[ \bar{x} - \frac{s \Pi^{-1} \left( \frac{1+\alpha}{2} \right)}{\sqrt{n}}; \bar{x} + \frac{s \Pi^{-1} \left( \frac{1+\alpha}{2} \right)}{\sqrt{n}} \right].$$

Mais, dans les deux cas, on ne voit pas comment estimer  $\mathbb{P}_\theta [\theta \in I(X_1, \dots, X_n)]$ , encore moins comment cette probabilité peut être égale à  $\alpha$  sans outils plus sophistiqués... Prudence, prudence donc, avant d'écrire n'importe quoi...

Ce type de résultats peuvent être démontrés, mais font appel à des notions plus compliquées. Pour en savoir plus, on pourra par exemple lire le cours de l'ENSTA de Jean-François Delmas, chapitre V.7, th. V.29 (théorème de Slutsky).

## 6.6 Tests

### 6.6.1 Exemple introductif

D'après le cours de Pierre Priouret. Supposons que la probabilité qu'une vache donne naissance à un veau ou à une génisse est la même, à savoir  $\frac{1}{2}$ . Cette hypothèse est vraisemblablement fausse, nous avons vu au paragraphe 1.2.4 que les filles et les garçons ne naissent pas en proportions égales, alors, pourquoi serait-ce le cas chez les ruminants ? Mais peu importe. L'industrie laitière s'intéresse tout particulièrement aux procédés permettant d'obtenir plus de génisses que de veaux. Un biologiste prétend avoir trouvé une méthode pour faire naître plus de génisses que de veaux et la teste sur 20 vaches sélectionnées "au hasard". Il naît 13 génisses et 7 veaux, soit une proportion de  $\frac{13}{20} = 0,65$  de génisses, très supérieure à la proportion naturelle de 0,5. Peut-on conclure à l'efficacité de la méthode ?

Modélisons le nombre de génisses  $X$  par une variable aléatoire qui suit une loi binomiale de paramètres  $n = 20$  et  $\theta$  inconnu (on est dans le cas d'un schéma de Bernoulli), avec  $\theta > 0,5$  si la méthode est efficace,  $\theta = 0,5$  si elle est inefficace : on suppose tout de même que le biologiste a un minimum de compétences et que la méthode ne produit pas l'effet inverse de l'effet recherché !

Considérons l'évènement  $\{X \geq 13\}$  qui s'est produit.

Si la méthode est inefficace, c'est-à-dire si  $\theta = 0,5$ , cet évènement a pour probabilité :

$$\sum_{k=13}^{20} \binom{20}{k} \frac{1}{2^{20}} = \frac{137980}{1048576} \simeq 0,1316.$$

Sous l'hypothèse de l'inefficacité de la méthode, l'évènement qui s'est produit n'est donc pas du tout exceptionnel. Ainsi, on ne rejette pas l'hypothèse  $\theta = 0,5$ , c'est-à-dire qu'on ne rejette pas le fait que la méthode soit inefficace. On ne peut cependant pas, a priori, accepter l'hypothèse que la méthode est efficace.

Un autre biologiste prétend à son tour avoir mis au point un procédé efficace, et on le teste sur  $n = 900$  vaches. Il naît 497 génisses et 403 veaux, soit une proportion de  $\frac{497}{900} \simeq 0,552$  de génisses, supérieure à la proportion naturelle, mais bien inférieure à la proportion obtenue avec la méthode du premier biologiste. Supposons que la méthode soit inefficace. Le nombre de génisses  $X$  est modélisé par une variable aléatoire de loi binomiale cette fois de paramètres  $n = 900$  et  $\theta = 0,5$ . L'évènement  $\{X \geq 497\}$  qui s'est produit s'écrit encore  $\left\{ \frac{X - n/2}{\sqrt{n/4}} \geq \frac{47}{15} \right\}$ , et, comme  $n$  est grand, est de probabilité à peu près égale à  $1 - \Pi\left(\frac{47}{15}\right) \simeq 0,00088$  (grâce au théorème central limite). C'est un évènement de probabilité très faible. On peut donc rejeter avec une très forte probabilité (mais évidemment pas de façon certaine !) le fait que  $\theta = 0,5$ , c'est-à-dire que l'on peut rejeter le fait que la méthode est inefficace.

### 6.6.2 Définitions

La situation générale est la suivante. On se donne  $(\mathcal{X}, \mathcal{A}, (\mathbb{P}_\theta)_{\theta \in \Theta})$  un modèle statistique paramétrique, et deux sous-ensembles  $\Theta_0, \Theta_1$  de  $\Theta$ , tels que  $\Theta_0 \cup \Theta_1 = \Theta$  et  $\Theta_0 \cap \Theta_1 = \emptyset$ .

On appelle *test d'hypothèse* une règle de décision qui permette de décider, à la vue de l'observation  $x = (x_1, \dots, x_n) \in \mathcal{X}$ , entre les hypothèses  $H_0 : "\theta \in \Theta_0"$  et  $H_1 : "\theta \in \Theta_1"$ .  $H_0$  est appelée *hypothèse nulle*, c'est celle que l'on imagine être vraie, c'est-à-dire vraie à moins que l'on ait de fortes preuves qu'elle ne le soit pas ;  $H_1$  s'appelle *hypothèse alternative*, elle est moralement plus osée.

Un test est déterminé par un évènement  $D$  de  $\mathcal{X}$ , appelé *région critique*, tel que si l'observation  $x$  appartient à  $D$ , on refuse l'hypothèse  $H_0$  “ $\theta$  appartient à  $\Theta_0$ ”.

On appelle *erreur de première espèce* le rejet de  $H_0$  à tort. L'erreur de première espèce est mesurée par les probabilités :

$$\{\mathbb{P}_\theta(D), \theta \in \Theta_0\}.$$

Soit  $\alpha \in ]0, 1[$  fixé. On dira qu'un test est de *niveau* ou *risque*  $\alpha$ , respectivement *seuil*  $\alpha$ , si :

$$\sup_{\theta \in \Theta_0} \mathbb{P}_\theta(D) = \alpha, \quad \text{respectivement} \quad \sup_{\theta \in \Theta_0} \mathbb{P}_\theta(D) \leq \alpha.$$

Ainsi, la probabilité de refuser à tort  $H_0$  est majorée par le niveau  $\alpha$  du test.

Si on ne rejette pas  $H_0$ , on ne l'accepte pas pour autant. Accepter  $H_0$  à tort revient à rejeter  $H_1$  à tort, ce qui est commettre une *erreur de seconde espèce*. L'erreur de deuxième espèce est mesurée par les probabilités :

$$\{\mathbb{P}_\theta(D^c), \theta \in \Theta_1\}.$$

On appelle *puissance du test*, la fonction  $\beta : \Theta_1 \rightarrow [0, 1]$ , définie par  $\beta(\theta) = \mathbb{P}_\theta(D)$ .

Le niveau  $\alpha$  du test étant fixé, il s'agit de trouver des régions  $D$  telles que l'erreur de deuxième espèce soit la plus petite possible. Autrement dit, parmi les tests de niveau  $\alpha$ , on cherche ceux de plus grande puissance. Ce n'est pas toujours possible.

Dans la théorie classique des tests, on fixe un seuil maximum à l'erreur de première espèce, à savoir 0,1 ou 0,05 ou 0,01.

*Remarque.* On remarquera, et cela est fondamental, que les deux hypothèses  $H_0$  et  $H_1$  sont traitées de façons dissymétriques. Leur choix n'est donc pas indifférent. Par exemple, lors d'un procès criminel, si l'on prend pour hypothèse  $H_0$ , “l'accusé est innocent”, ce qui est la présomption d'innocence, l'erreur de première espèce qui représente la probabilité de condamner à tort un innocent, peut être considérée comme plus grave que l'erreur de deuxième espèce qui représente la probabilité d'acquitter un coupable.

### 6.6.3 Cas de deux hypothèses simples $\Theta = \{\theta_0, \theta_1\}$

Ce paragraphe n'est pas au programme du concours, mais sa simplicité permet de mieux appréhender la notion de test. Il s'agit de tester des hypothèses simples  $H_0 : \theta = \theta_0$  contre  $H_1 : \theta = \theta_1$ .

**Proposition 6.7** (Lemme de Neyman-Pearson). *Soit  $(\mathcal{X}, \mathcal{A}, (\mathbb{P}_\theta)_{\theta \in \Theta})$  un modèle statistique paramétrique, où  $\Theta = \{\theta_0, \theta_1\}$ . On suppose que les deux lois possibles  $\mathbb{P}_{\theta_0}$  et  $\mathbb{P}_{\theta_1}$  pour l'échantillon  $X = (X_1, \dots, X_n)$  ont des densités, notées respectivement  $p_0$  et  $p_1$ . Soient  $\alpha \in ]0, 1[$  et  $D_\alpha$  l'évènement :*

$$D_\alpha = \{(x_1, \dots, x_n) \in \mathcal{X}; p_1(x_1, \dots, x_n) \geq \lambda_\alpha p_0(x_1, \dots, x_n)\},$$

où  $\lambda_\alpha$  est choisi de sorte que  $\mathbb{P}_{\theta_0}(D_\alpha) = \alpha$ . Alors  $D_\alpha$  est la région critique du test de niveau  $\alpha$  de  $H_0$  contre  $H_1$ , le plus puissant.

*Démonstration.* Afin de simplifier les notations, nous omettons l'indice  $\alpha$ . Soit  $B$  une autre région critique de niveau  $\alpha$ . Nous devons montrer que  $\mathbb{P}_{\theta_1}(B) \leq \mathbb{P}_{\theta_1}(D)$ .

Remarquons d'abord que :

$$\begin{aligned}\mathbb{P}_{\theta_0}(B \cap D^c) &= \mathbb{P}_{\theta_0}(B) - \mathbb{P}_{\theta_0}(B \cap D) = \alpha - \mathbb{P}_{\theta_0}(B \cap D) \\ \text{et } \mathbb{P}_{\theta_0}(D \cap B^c) &= \alpha - \mathbb{P}_{\theta_0}(D \cap B) \\ \text{d'où } \mathbb{P}_{\theta_0}(B \cap D^c) &= \mathbb{P}_{\theta_0}(D \cap B^c).\end{aligned}$$

Comme  $(B \cap D^c) \subset D^c$  et  $(D \cap B^c) \subset D$ , on a :

$$\begin{aligned}\mathbb{P}_{\theta_0}(B \cap D^c) &= \int_{B \cap D^c} p_0(x_1, \dots, x_n) dx_1 \cdots dx_n \\ &\geq \frac{1}{\lambda} \int_{B \cap D^c} p_1(x_1, \dots, x_n) dx_1 \cdots dx_n = \frac{1}{\lambda} \mathbb{P}_{\theta_1}(B \cap D^c), \\ \text{et } \mathbb{P}_{\theta_0}(D \cap B^c) &\leq \frac{1}{\lambda} \mathbb{P}_{\theta_1}(D \cap B^c).\end{aligned}$$

Ainsi,  $\mathbb{P}_{\theta_1}(B \cap D^c) \leq \lambda \mathbb{P}_{\theta_0}(B \cap D^c) = \lambda \mathbb{P}_{\theta_0}(D \cap B^c) \leq \mathbb{P}_{\theta_1}(D \cap B^c)$ , et on conclut :

$$\mathbb{P}_{\theta_1}(B) = \mathbb{P}_{\theta_1}(B \cap D^c) + \mathbb{P}_{\theta_1}(B \cap D) \leq \mathbb{P}_{\theta_1}(D \cap B^c) + \mathbb{P}_{\theta_1}(B \cap D) = \mathbb{P}_{\theta_1}(D).$$

□

#### 6.6.4 Test pour la moyenne

Soit  $(X_1, \dots, X_n)$  un échantillon de taille  $n$  de loi  $\mu_\theta$  et de moyenne commune  $m$ . On note  $\mathbb{P}_\theta$  la loi jointe du vecteur aléatoire  $(X_1, \dots, X_n)$ .

On souhaite tester deux sortes d'hypothèses pour la moyenne.

1.  $H_0$  : “ $m = m_0$ ” contre  $H_1$  : “ $m \neq m_0$ ”, appelé *test bilatère* et non *bilatéral* qui est une mauvaise traduction de l'anglais *bilateral*.
2. Si l'on sait que la moyenne  $m$  satisfait à  $m \geq m_0$ , on teste alors les hypothèses :  $H_0$  : “ $m = m_0$ ” contre  $H_1$  : “ $m > m_0$ ”, appelé *test unilatère*. De manière analogue, si l'on sait que la moyenne  $m$  satisfait à  $m \leq m_0$ , on teste alors les hypothèses :  $H_0$  : “ $m = m_0$ ” contre  $H_1$  : “ $m < m_0$ ”.

Soit  $\alpha \in ]0, 1[$  un seuil fixé. Il est alors naturel de chercher une région critique de la forme :

1.  $D_\alpha = \{x \in \mathcal{X} : |M_n - m| > C_\alpha\}$ ,
2.  $D_\alpha = \{x \in \mathcal{X} : M_n - m > C_\alpha\}$  ou  $D_\alpha = \{x \in \mathcal{X} : M_n - m < C_\alpha\}$ .

Étant donné que l'hypothèse  $H_0$  est simple, le test est de niveau  $\alpha$ , si :

$$\mathbb{P}_{m_0}(D_\alpha) = \alpha.$$

Détaillons ceci dans le cas d'un test bilatère. On cherche  $C_\alpha$  tel que :

$$\alpha = \mathbb{P}_{m_0}(D_\alpha) = \mathbb{P}_{m_0} \left[ (x_1, \dots, x_n) \in \mathcal{X} : \left| \frac{1}{n} \sum_{k=1}^n x_k - m_0 \right| > C_\alpha \right]. \quad (6.2)$$

Afin de continuer, nous avons besoin de plus d'informations sur la loi de l'échantillon ou sur sa taille.

#### • Échantillon de loi gaussienne de paramètres $m$ et $\sigma^2$

Soit  $(X_1, \dots, X_n)$  un échantillon de taille  $n$  de loi gaussienne de moyenne  $m$  et de variance  $\sigma^2$ . Soit  $m_0 \in \mathbb{R}$ .

Dans ce cas, il faut traiter séparément le cas où la variance est connue de celui où elle ne l'est pas. On note  $T_{n-1}$  une variable aléatoire de loi de Student à  $n - 1$  degrés de liberté.

**Proposition 6.8.**

(A) *Cas où la variance  $\sigma^2$  est connue.*

(1) (Test bilatère). Soit  $H_0$  : “ $m = m_0$ ” et  $H_1$  : “ $m \neq m_0$ ”. Posons,

$$D_\alpha = \left\{ (x_1, \dots, x_n) \in \mathbb{R}^n : \left| \frac{1}{n} \sum_{k=1}^n x_k - m_0 \right| > \frac{\Pi^{-1}(1 - \frac{\alpha}{2})\sigma}{\sqrt{n}} \right\}.$$

Alors,  $D_\alpha$  est la région critique d'un test de niveau  $\alpha$  de  $H_0$  contre  $H_1$ .

(2) (Test unilatère). Soit  $H_0$  : “ $m = m_0$ ” et  $H_1$  : “ $m > m_0$ ”. Posons,

$$D_\alpha = \left\{ (x_1, \dots, x_n) \in \mathbb{R}^n : \frac{1}{n} \sum_{k=1}^n x_k - m_0 > \frac{\Pi^{-1}(1 - \alpha)\sigma}{\sqrt{n}} \right\}.$$

Alors,  $D_\alpha$  est la région critique d'un test de niveau  $\alpha$  de  $H_0$  contre  $H_1$ .

(2) (Test unilatère). Soit  $H_0$  : “ $m = m_0$ ” et  $H_1$  : “ $m < m_0$ ”. Posons,

$$D_\alpha = \left\{ (x_1, \dots, x_n) \in \mathbb{R}^n : \frac{1}{n} \sum_{k=1}^n x_k - m_0 < \frac{\Pi^{-1}(\alpha)\sigma}{\sqrt{n}} \right\}.$$

Alors,  $D_\alpha$  est la région critique d'un test de niveau  $\alpha$  de  $H_0$  contre  $H_1$ .

(B) *Cas où la variance  $\sigma^2$  n'est pas connue.*

(1) (Test bilatère). Soit  $H_0$  : “ $m = m_0$ ” et  $H_1$  : “ $m \neq m_0$ ”.

On définit  $d_\alpha$  par  $\alpha = \mathbb{P}[|T_{n-1}| \geq d_\alpha]$ , et on pose :

$$D_\alpha = \left\{ (x_1, \dots, x_n) \in \mathbb{R}^n : \left| \frac{1}{n} \sum_{k=1}^n x_k - m_0 \right| > \frac{d_\alpha \sqrt{\sum_{i=1}^n \left( x_i - \frac{1}{n} \sum_{j=1}^n x_j \right)^2}}{\sqrt{n(n-1)}} \right\}.$$

Alors,  $D_\alpha$  est la région critique d'un test de niveau  $\alpha$  de  $H_0$  contre  $H_1$ .

(2) (Test unilatère). Soit  $H_0$  : “ $m = m_0$ ” et  $H_1$  : “ $m > m_0$ ”.

On définit  $d_\alpha$  par  $\alpha = \mathbb{P}[T_{n-1} \geq d_\alpha]$ , et on pose :

$$D_\alpha = \left\{ (x_1, \dots, x_n) \in \mathbb{R}^n : \frac{1}{n} \sum_{k=1}^n x_k - m_0 > \frac{d_\alpha \sqrt{\sum_{i=1}^n \left( x_i - \frac{1}{n} \sum_{j=1}^n x_j \right)^2}}{\sqrt{n(n-1)}} \right\}.$$

Alors,  $D_\alpha$  est la région critique d'un test de niveau  $\alpha$  de  $H_0$  contre  $H_1$ .

(2) (Test unilatère). Soit  $H_0$  : “ $m = m_0$ ” et  $H_1$  : “ $m < m_0$ ”.

On définit  $d_\alpha$  par  $\alpha = \mathbb{P}[T_{n-1} \leq d_\alpha]$ , et on pose :

$$D_\alpha = \left\{ (x_1, \dots, x_n) \in \mathbb{R}^n : \frac{1}{n} \sum_{k=1}^n x_k - m_0 < \frac{d_\alpha \sqrt{\sum_{i=1}^n \left( x_i - \frac{1}{n} \sum_{j=1}^n x_j \right)^2}}{\sqrt{n(n-1)}} \right\}.$$

Alors,  $D_\alpha$  est la région critique d'un test de niveau  $\alpha$  de  $H_0$  contre  $H_1$ .



*Remarque.* On rappelle que la densité d'une variable de Student est paire, de sorte que  $\mathbb{P}[|T_d| \geq c] = 2\mathbb{P}[T_d \geq c]$ . Ainsi :

$$\mathbb{P}[T_d \leq c] = 1 - \mathbb{P}[T_d \geq c] = 1 - \frac{1}{2}\mathbb{P}[|T_d| \geq c].$$

On rappelle encore que, si  $d \geq 30$ ,  $\mathbb{P}[T_d \in I] \simeq \mathbb{P}[N \in I]$ , pour tout intervalle  $I$  de  $\mathbb{R}$ . Lorsque  $d < 30$ , on utilise des tables pour calculer des quantités du type  $\mathbb{P}[|T_d| \geq c]$ .

*Démonstration.*

(A) Cas où la variance  $\sigma^2$  est connue.

Le paramètre inconnu  $\theta$  étant la moyenne  $m$  de l'échantillon, l'ensemble des paramètres est  $\Theta = \mathbb{R}$ . Dans ce cas,  $H_0 : "m = m_0"$ . Le test repose sur le fait que, sous l'hypothèse  $H_0$ , la variable  $\sqrt{n}\frac{M_n - m_0}{\sigma}$  suit une loi gaussienne centrée réduite, où  $M_n = \frac{1}{n} \sum_{k=1}^n X_k$ .

Notons  $N$  une variable aléatoire normale centrée réduite.

Nous prouvons le résultat uniquement dans le cas d'un test bilatère : test de  $m = m_0$  contre  $m \neq m_0$ . D'après l'équation (6.2), nous cherchons  $C_\alpha$  tel que :

$$\alpha = \mathbb{P}_{m_0}[D_\alpha] = \mathbb{P}_{m_0}[|M_n - m_0| > C_\alpha].$$

Or,

$$\begin{aligned} \mathbb{P}_{m_0}[|M_n - m_0| > C_\alpha] &= \mathbb{P}\left[|N| > \frac{C_\alpha \sqrt{n}}{\sigma}\right] \\ &= 2\left(1 - \Pi\left(\frac{C_\alpha \sqrt{n}}{\sigma}\right)\right). \end{aligned}$$

On cherche donc  $C_\alpha$  tel que :  $\alpha = 2\left(1 - \Pi\left(\frac{C_\alpha \sqrt{n}}{\sigma}\right)\right)$ . Il suffit de choisir comme valeur  $C_\alpha = \frac{\sigma}{\sqrt{n}}\Pi^{-1}\left(1 - \frac{\alpha}{2}\right)$ .

(B) Cas où la variance  $\sigma^2$  n'est pas connue.

Le paramètre inconnu  $\theta$  est cette fois le couple moyenne-variance  $(m, \sigma^2)$  de l'échantillon, de sorte que l'ensemble des paramètres est  $\Theta = \mathbb{R} \times \mathbb{R}_+^*$ . Dans ce cas, on choisit

$$H_0 : "(m, \sigma^2) \in \{m_0\} \times \mathbb{R}_+^*"$$

Le test repose sur le fait que, sous l'hypothèse  $H_0$ , la variable  $\sqrt{n}\frac{M_n - m_0}{S_n}$  suit une loi de Student à  $n - 1$  degrés de liberté, où  $S_n = \sqrt{\frac{1}{n-1} \sum_{k=1}^n (X_k - M_n)^2}$ . Notons  $T_{n-1}$  une variable aléatoire de loi de Student à  $n - 1$  degrés de liberté.

Nous prouvons le résultat uniquement dans le cas d'un test bilatère. Comme pour la construction d'un intervalle de confiance, au lieu de travailler à partir de la variable aléatoire  $M_n - m_0$  ou en fait de  $\sqrt{n}\frac{M_n - m_0}{\sigma}$ , nous allons travailler ici à partir de la variable aléatoire  $\sqrt{n}\frac{M_n - m_0}{S_n}$  puisqu'on ne connaît pas la variance  $\sigma$ . Nous cherchons donc une constante  $c_\alpha$  telle que :

$$\alpha = \mathbb{P}_{m_0}[D_\alpha] = \mathbb{P}_{m_0}\left[\left|\sqrt{n}\frac{M_n - m_0}{S_n}\right| > c_\alpha\right].$$

Or, sous  $\mathbb{P}_{m_0}$ , la variable aléatoire  $\sqrt{n} \frac{M_n - m}{S_n}$  suit une loi de Student à  $n - 1$  degré de liberté. Alors,

$$\mathbb{P}_{m_0} \left[ \left| \sqrt{n} \frac{M_n - m}{S_n} \right| > c_\alpha \right] = \mathbb{P}[|T_{n-1}| > c_\alpha]$$

Il suffit donc de choisir  $c_\alpha = d_\alpha$  où  $\mathbb{P}[|T_{n-1}| > d_\alpha] = \alpha$ .

□

*Remarque.*

1. Si on ne rejette pas l'hypothèse  $H_0$ , pourquoi ne peut-on pas en général l'accepter? Accepter  $H_0$  revient en fait à rejeter  $H_1$ . Calculons par exemple la probabilité de commettre une erreur de deuxième espèce, c'est-à-dire de rejeter l'hypothèse  $H_1$  à tort, autrement dit d'accepter  $H_0$  à tort, dans le cas d'un test de  $m = m_0$  contre  $m < m_0$  lorsque la variance de l'échantillon gaussien est connue.

Puisque  $H_1$  est l'hypothèse " $m \in ] - \infty, m_0[$ " dans ce cas, l'erreur de deuxième espèce est majorée par :  $\sup_{m < m_0} \mathbb{P}_m[D_\alpha^c]$ .

Pour  $m$  fixé, calculons  $\mathbb{P}_m[D_\alpha^c]$  :

$$\begin{aligned} \mathbb{P}_m[D_\alpha^c] &= \mathbb{P}_m \left[ M_n - m_0 > \Pi^{-1}(\alpha) \frac{\sigma}{\sqrt{n}} \right] \\ &= \mathbb{P}_m \left[ \sqrt{n} \frac{M_n - m}{\sigma} > \Pi^{-1}(\alpha) + \frac{\sqrt{n}}{\sigma} (m - m_0) \right] \\ &= \mathbb{P}[N > \Pi^{-1}(\alpha) + \frac{\sqrt{n}}{\sigma} (m - m_0)] \\ &= 1 - \Pi \left[ \Pi^{-1}(\alpha) + \frac{\sqrt{n}}{\sigma} (m - m_0) \right], \end{aligned}$$

où  $N$  est une variable aléatoire gaussienne centrée réduite.

Or, l'application  $m \mapsto 1 - \Pi \left( \Pi^{-1}(\alpha) + \frac{\sqrt{n}}{\sigma} (m_0 - m) \right)$  est continue croissante, de sorte que sa borne supérieure sur l'intervalle  $] - \infty, m_0[$  est égale à sa valeur en  $m_0$ , à savoir  $1 - \Pi(\Pi^{-1}(\alpha)) = 1 - \alpha$ .

La probabilité d'accepter à tort l'hypothèse  $H_0$  vaut donc  $1 - \alpha$ ... mais  $\alpha$  est "petit", donc  $1 - \alpha$  est "proche de 1"... On comprend pourquoi cela n'a pas de sens d'accepter l'hypothèse  $H_0$  avec ce type de test!

2. Nous avons énoncé les tests tels qu'ils figurent désormais dans les programmes du CAPES. On remarquera que l'on peut aisément étendre les résultats pour des tests de  $m \leq m_0$  contre  $m > m_0$ , ou de  $m \geq m_0$  contre  $m < m_0$ , qui sont les tests énoncés le plus fréquemment dans la littérature mathématique. On calcule alors l'erreur de première espèce en utilisant la continuité et la croissance de la fonction de répartition  $\Pi$  de la gaussienne centrée réduite, comme pour le calcul de l'erreur de deuxième espèce dans la remarque ci-dessus.

• **Échantillon de loi de Bernoulli de paramètre inconnu  $\theta \in ]0, 1[$ .**

Soit  $(X_1, \dots, X_n)$  un échantillon de taille  $n$  de loi de Bernoulli de paramètre inconnu  $\theta \in ]0, 1[$ .

On suppose que  $n$  et  $\theta$  sont tels que la variable aléatoire  $\sum_{k=1}^n X_k$ , qui suit une loi binomiale de paramètres  $(n, \theta)$ , est approchable par une loi normale (classiquement  $n \geq 10$  et  $n\theta, n(1 - \theta)$  dépassent quelques unités).

Soit  $\theta_0 \in ]0, 1[$ . On obtient alors les résultats suivants :

**Proposition 6.9.** Soit  $\alpha \in ]0, 1[$  le niveau fixé du test.

1. (Test bilatère). Soit  $H_0 : \theta = \theta_0$  et  $H_1 : \theta \in ]0, 1[ \setminus \{\theta_0\}$ . Posons :

$$D_\alpha = \left\{ (x_1, \dots, x_n) \in \{0, 1\}^n; \left| \frac{1}{n} \sum_{k=1}^n x_k - \theta_0 \right| > \frac{\Pi^{-1} \left( 1 - \frac{\alpha}{2} \right) \sqrt{\theta_0(1 - \theta_0)}}{\sqrt{n}} \right\}.$$

Alors,  $D_\alpha$  est la région critique d'un test de niveau (à peu près égal à)  $\alpha$  de  $H_0$  contre  $H_1$ .

2. (Test unilatère). On suppose que l'on sait que  $\theta$  appartient à l'intervalle  $[\theta_0, 1[$ . Soit alors,  $H_0 : \theta = \theta_0$  et  $H_1 : \theta \in ]\theta_0, 1[$ . Posons :

$$D_\alpha = \left\{ (x_1, \dots, x_n) \in \{0, 1\}^n; \frac{1}{n} \sum_{k=1}^n x_k - \theta_0 > \frac{\Pi^{-1} (1 - \alpha) \sqrt{\theta_0(1 - \theta_0)}}{\sqrt{n}} \right\}.$$

Alors,  $D_\alpha$  est la région critique d'un test de niveau (à peu près égal à)  $\alpha$  de  $H_0$  contre  $H_1$ .

3. (Test unilatère). On suppose que l'on sait que  $\theta$  appartient à l'intervalle  $]0, \theta_0[$ . Soit alors,  $H_0 : \theta = \theta_0$  et  $H_1 : \theta \in ]0, \theta_0[$ . Posons :

$$D_\alpha = \left\{ (x_1, \dots, x_n) \in \{0, 1\}^n; \frac{1}{n} \sum_{k=1}^n x_k - \theta_0 < \frac{\Pi^{-1} (\alpha) \sqrt{\theta_0(1 - \theta_0)}}{\sqrt{n}} \right\}.$$

Alors,  $D_\alpha$  est la région critique d'un test de niveau (à peu près égal à)  $\alpha$  de  $H_0$  contre  $H_1$ .

## Conséquences

- Si  $(x_1, \dots, x_n) \in D_\alpha$ , on rejette l'hypothèse  $\theta = \theta_0$ , et la probabilité de se tromper est en gros majorée par  $\alpha$ .
- Si  $(x_1, \dots, x_n) \notin D_\alpha$ , on ne peut pas rejeter l'hypothèse  $\theta = \theta_0$ . On ne peut pas l'accepter non plus car on ne sait pas si la probabilité d'accepter à tort l'hypothèse  $H_0$ , c'est-à-dire  $\sup_{\theta \neq \theta_0} \mathbb{P}_\theta[D_\alpha^c]$  est faible ou non.

EXERCICE 6.4. D'après Dunod ex. 11.1, p. 140.

Dans la population française, le pourcentage d'individus de rhésus négatif est de 15%. Dans un échantillon représentatif de 200 basques français, on observe que 44 personnes sont de rhésus négatif. Peut-on dire, au risque  $\alpha = 0,05$ , que les basques diffèrent du reste de la population française en ce qui concerne le rhésus ? Quelle serait la conclusion si on avait observé seulement 37 basques de rhésus négatif parmi les 200 personnes testées ?

EXERCICE 6.5. D'après Dunod ex. 11.2, p. 140.

Dans une population, le pourcentage d'individus présentant des rides est de 25%. Sur 200 personnes ayant suivi un traitement anti-rides, on observe que 40 personnes ont des rides. Au risque  $\alpha = 0,05$ , peut-on dire que le traitement est efficace ?

## 6.6.5 Comparaison de deux moyennes

**Proposition 6.10.** On considère  $(X_1, \dots, X_n)$  et  $(Y_1, \dots, Y_m)$  deux échantillons gaussiens indépendants de même variance, de lois respectives  $\mathcal{N}(\mu_1, \sigma^2)$  et  $\mathcal{N}(\mu_2, \sigma^2)$ . On veut tester

l'hypothèse  $H_0 : \mu_1 = \mu_2$  contre l'hypothèse  $H_1 : \mu_1 \neq \mu_2$ . On note :

$$\bar{X}_n = \frac{1}{n} \sum_{k=1}^n X_k, \quad \bar{Y}_m = \frac{1}{m} \sum_{k=1}^m Y_k,$$

$$S_1^2 = \frac{1}{n-1} \sum_{k=1}^n (X_k - \bar{X}_n)^2, \quad S_2^2 = \frac{1}{m-1} \sum_{k=1}^m (Y_k - \bar{Y}_m)^2.$$

- Sous l'hypothèse  $H_0 : \mu_1 = \mu_2$ , la variable aléatoire

$$Z = \sqrt{\frac{n+m-2}{\frac{1}{n} + \frac{1}{m}}} \frac{\bar{X}_n - \bar{Y}_m}{(n-1)S_1^2 + (m-1)S_2^2},$$

suit une loi de Student à  $n+m-2$  degrés de liberté.

- On en déduit le test suivant. Soit  $\alpha \in ]0, 1[$ . Fixons  $d_\alpha > 0$  tel que  $\mathbb{P}[|T_{n+m-2}| \geq d_\alpha] = \alpha$ , où  $T_{n+m-2}$  suit une loi de Student à  $n+m-2$  degrés de liberté. On pose

$$F(x, y) = \sqrt{\frac{n+m-2}{\frac{1}{n} + \frac{1}{m}}} \times \frac{\left| \frac{1}{n} \sum_{k=1}^n x_k - \frac{1}{m} \sum_{k=1}^m y_k \right|}{\sqrt{\sum_{i=1}^n \left( x_i - \frac{1}{n} \sum_{k=1}^n x_k \right)^2 + \sum_{i=1}^m \left( y_i - \frac{1}{m} \sum_{k=1}^m y_k \right)^2}}.$$

Alors l'ensemble

$$D_\alpha = \left\{ (x_1, \dots, x_n, y_1, \dots, y_m) \in \mathbb{R}^{n+m} : F(x, y) \geq d_\alpha \right\},$$

est la région critique d'un test de  $H_0 : \mu_1 = \mu_2$  contre  $H_1 : \mu_1 \neq \mu_2$ , de niveau  $\alpha$ .

*Remarque.* Classiquement, si  $x = (x_1, \dots, x_n)$  et  $y = (y_1, \dots, y_m)$  sont les observations, on note

$$\bar{x} = \frac{1}{n} \sum_{k=1}^n x_k, \quad s_1^2 = \frac{1}{n-1} \sum_{k=1}^n (x_k - \bar{x})^2,$$

$$\bar{y} = \frac{1}{m} \sum_{k=1}^m y_k, \quad s_2^2 = \frac{1}{m-1} \sum_{k=1}^m (y_k - \bar{y})^2,$$

de sorte que l'on écrit souvent la région critique sous la forme :

$$D_\alpha = \left\{ (x_1, \dots, x_n, y_1, \dots, y_m) \in \mathbb{R}^{n+m}; \sqrt{\frac{n+m-2}{\frac{1}{n} + \frac{1}{m}}} \frac{|\bar{x} - \bar{y}|}{\sqrt{(n-1)s_1^2 + (m-1)s_2^2}} \geq d_\alpha \right\}.$$

**EXERCICE 6.6.** *Polycopié de Priouret.* Les 148 copies d'un examen sont partagées au hasard entre deux correcteurs. Le correcteur A en corrige 85 et le second B, en corrige 63. Leurs moyennes respectives sont de 10,5 et 9,6, avec, pour les notes de A,  $s_1^2 = 11,25$ , et pour celles de B,  $s_2^2 = 8,4$  (avec les notations classiques rappelées ci-dessus). La différence entre les deux correcteurs est-elle significative au seuil  $\alpha = 0,01$ ? On modélise les notes de chaque correcteur comme les valeurs prises par des variables aléatoires gaussiennes de même variance, dont les moyennes  $\mu_1$  et  $\mu_2$  caractérisent la sévérité du correcteur.

Même question si  $s_1^2 = 4,1$ , et pour celles de B,  $s_2^2 = 3,9$ , les autres données restant inchangées.

### 6.6.6 Test d'adéquation à une loi

Prenons un exemple. On considère un caractère génétique pour lequel on suppose une transmission mendélienne (transmission due à la mutation d'un seul gène) gouvernée par un gène prenant les deux formes A et B. Supposons que l'on sache identifier les individus AA, AB et BB. Si le modèle mendélien est adapté à la situation, les fréquences théoriques des trois possibilités sont respectivement  $\frac{1}{4}$ ,  $\frac{1}{2}$  et  $\frac{1}{4}$ . C'est cette hypothèse que l'on souhaite tester par un test dit du  $\chi^2$ .

Soit une variable aléatoire  $Y$  à valeurs dans l'ensemble fini  $\{y_1, \dots, y_N\}$ . On note  $p = (p_k)_{1 \leq k \leq N}$  où  $p_k$  est la probabilité de l'évènement  $\{Y = y_k\}$ . On souhaite comparer ce vecteur  $p$  à une valeur particulière  $p^0$  (dans l'exemple ci-dessus,  $N = 3$  et  $p^0$  est le vecteur de composantes  $(\frac{1}{4}, \frac{1}{2}, \frac{1}{4})$ ). On désire donc tester l'hypothèse  $H_0$  " $p = p^0$ " contre l'hypothèse  $H_1$  " $p \neq p^0$ ".

On dispose de  $n$  réalisations indépendantes  $Y_1, \dots, Y_n$  de  $Y$ , *i.e.*  $(Y_1, \dots, Y_n)$  est un échantillon de taille  $n$  de la loi de  $Y$ . Notons  $N_1^n, \dots, N_N^n$ , les effectifs de chaque valeur possible pour  $Y$  *i.e.*  $N_i^n = \sum_{j=1}^n \mathbb{I}_{\{Y_j=y_i\}}$  compte le nombre de fois où l'on a obtenu la valeur  $y_i$ . On a bien sûr, puisque chaque variable  $Y_j$  ne peut prendre que les valeurs  $y_1, \dots, y_N$  :

$$\sum_{i=1}^N N_i^n = \sum_{i=1}^N \sum_{j=1}^n \mathbb{I}_{\{Y_j=y_i\}} = \sum_{j=1}^n \sum_{i=1}^N \mathbb{I}_{\{Y_j=y_i\}} = \sum_{j=1}^n 1 = n.$$

Pour tout  $k \in \{1, \dots, N\}$ , on note  $\widehat{p}_k^n$  les fréquences empiriques, et  $\widehat{p}^n = (\widehat{p}_k^n)_{1 \leq k \leq N}$  le vecteur des fréquences empiriques. Ainsi, si  $n_1, \dots, n_N$  sont des entiers naturels tels que  $n_1 + \dots + n_N = n$ , on a :

$$\begin{aligned} & \mathbb{P} [ N_1^n = n_1, \dots, N_N^n = n_N ] \\ &= \mathbb{P} \left[ \bigcap_{i=1}^n \{n_i \text{ variables prennent la valeur } y_i\} \right] \\ &= \binom{n}{n_1} \binom{n-n_1}{n_2} \dots \binom{n-(n_1+\dots+n_{N-1})}{n_N} p_1^{n_1} \dots p_N^{n_N} \\ &= \frac{n!}{n_1! \dots n_N!} p_1^{n_1} \dots p_N^{n_N}. \end{aligned}$$

Définissons la *distance du  $\chi^2$*  entre lois sur  $\{y_1, \dots, y_N\}$ . Si  $p = (p_k)_{1 \leq k \leq N}$  et  $q = (q_k)_{1 \leq k \leq N}$  sont deux vecteurs de composantes strictement positives dont la somme est égale à 1, alors :

$$\chi^2(p, q) = \sum_{k=1}^N \frac{(p_k - q_k)^2}{q_k}.$$

En fait, il ne s'agit pas d'une vraie distance car, par exemple, elle n'est pas symétrique. Cependant, on a bien l'équivalence :  $\chi^2(p, q) = 0 \iff p = q$ .

Par ailleurs :  $n \chi^2(\widehat{p}^n, p^0) = \sum_{k=1}^N \frac{(N_k^n - n p_k^0)^2}{n p_k^0}$  ( $N_k^n$  est l'effectif valant  $y_k$  observé,  $n p_k^0$  est l'effectif théorique valant  $y_k$  ; on compare donc les effectifs observés aux effectifs théoriques).

**Théorème 6.1.** (*admis*). On suppose que :  $\forall k \in \{1, \dots, N\}, p_k^0 \neq 0$ .

i) Soit  $I$  un intervalle de  $\mathbb{R}$ . Sous l'hypothèse  $H_0 : "p = p^0"$ , on a :

$$\lim_{n \rightarrow +\infty} \mathbb{P} \left[ n \chi^2(\widehat{p}^n, p^0) \in I \right] = \mathbb{P} \left[ \chi_{N-1}^2 \in I \right],$$

où  $\chi_{N-1}^2$  est une variable qui suit une loi du  $\chi^2$  à  $N - 1$  degrés de liberté.

ii) Sous l'hypothèse  $H_1 : "p \neq p^0"$ ,  $n \chi^2(\widehat{p}^n, p^0)$  tend en probabilité vers  $+\infty$ .

On en déduit un test asymptotique pour  $n$  grand, appelé *test du  $\chi^2$* .

**Proposition 6.11.** On suppose  $n$  "grand". Soit  $\alpha \in ]0, 1[$ .

Soit  $\beta_\alpha^N$  le fractile d'ordre  $1 - \alpha$  de la loi du  $\chi^2$  à  $N - 1$  degrés de liberté, c'est-à-dire que  $\mathbb{P}[\chi_{N-1}^2 \leq \beta_\alpha^N] = 1 - \alpha$ .

- Si l'observation vérifie  $n \chi^2(\widehat{p}^n, p^0) > \beta_\alpha^N$  alors on rejette l'hypothèse  $H_0 : "p = p^0"$ , et la probabilité de rejeter à tort l'hypothèse est de l'ordre de  $\alpha$ .
- Si l'observation vérifie  $n \chi^2(\widehat{p}^n, p^0) \leq \beta_\alpha^N$ , alors on ne peut pas rejeter l'hypothèse  $H_0 : "p = p^0"$ .

*Remarque.* D'après le théorème précédent, la probabilité de rejeter à tort l'hypothèse  $H_0$  tend vers  $\alpha$  quand  $n$  tend vers l'infini. On parle de test de niveau asymptotique  $\alpha$ . Le problème est de savoir à partir de quelles valeurs de  $n$  l'approximation est justifiée. Il n'y a pas de résultats théoriques précis. À partir de considérations heuristiques reposant sur des simulations (et donc sur l'expérience et non sur la théorie), on considère généralement que l'approximation asymptotique est justifiée dès que :  $\forall k \in \{1, \dots, N\}, np_k > 5$ .

EXERCICE 6.7. D'après Dunod ex. 10.1, p. 127

On a effectué le croisement de balsamines blanches avec des balsamines pourpres. À la première génération, toutes les fleurs sont pourpres, mais à la deuxième, on obtient la répartition suivante :

pourpre	rose	blanc lavande	blanc
1790	547	548	213

On souhaite savoir si la répartition se fait selon les lois de Mendel, c'est-à-dire selon les probabilités  $(\frac{9}{16}; \frac{3}{16}; \frac{3}{16}; \frac{1}{16})$ . Au risque  $\alpha = 0.05$ , peut-on rejeter l'hypothèse de répartition mendélienne ?

EXEMPLE. Les tests d'adéquation à une loi équirépartie sont au programme de la classe de terminale ES, même si le vocabulaire des tests est hors programme. Mais on trouve parfois quelques formulations bien curieuses... Voici l'énoncé d'un exercice donné au baccalauréat ES en juin 2003. Faites-en une analyse critique...

Les guichets d'une agence bancaire d'une petite ville sont ouverts au public cinq jours par semaine : les mardi, mercredi, jeudi, vendredi et samedi. Le tableau ci-dessous donne la répartition journalière des 250 retraits d'argent liquide effectués aux guichets une certaine semaine.

Jour de la semaine	mardi	mercredi	jeudi	vendredi	samedi
Rang $i$ du jour	1	2	3	4	5
Nombre de retraits	37	55	45	53	60

On veut tester l'hypothèse " le nombre de retraits est indépendant du jour de la semaine ". On suppose donc que le nombre des retraits journaliers est égal à  $\frac{1}{5}$  du nombre des retraits de la semaine. On pose  $d_{obs}^2 = \sum_{i=1}^5 (f_i - \frac{1}{5})^2$  où  $f_i$  est la fréquence des retraits du  $i$ -ième jour.

1. Calculer les fréquences des retraits pour chacun des cinq jours de la semaine.

2. Calculer alors la valeur de  $1000 d_{obs}^2$  (la multiplication par 1000 permet d'obtenir un résultat plus lisible).
3. En supposant qu'il y a équiprobabilité des retraits journaliers, on a simulé 2000 séries de 250 retraits hebdomadaires. Pour chaque série, on a calculé la valeur du  $1000 d_{obs}^2$  correspondant. On a obtenu ainsi 2000 valeurs de  $1000 d_{obs}^2$ . Ces valeurs ont permis de construire le diagramme en boîte ci-dessous où les extrémités des "pattes" correspondent respectivement au premier décile et au neuvième décile.



Lire sur le diagramme une valeur approchée du neuvième décile.

4. En argumentant soigneusement la réponse, dire si pour la série observée au début, on peut affirmer, avec un risque d'erreur inférieur à 10%, que "le nombre de retraits est indépendant du jour de la semaine" ?

Reprenons cet exercice avec les notations du cours pour mieux comprendre ce qui se passe. Il s'agit de réaliser un test d'adéquation à une loi dans le cas où  $n = 250$  ( $n$  est grand, on pourra appliquer la règle),  $N = 5$ ,  $p^0 = (\frac{1}{5}, \frac{1}{5}, \frac{1}{5}, \frac{1}{5}, \frac{1}{5})$  (l'hypothèse  $H_0$  est "le nombre de retraits est indépendant du jour de la semaine", d'où la valeur de  $p^0$ ). Le vecteur  $\widehat{p}^n = (\widehat{p}_k^n)_{1 \leq k \leq 5}$  des fréquences empiriques est noté  $(f_k)_{1 \leq k \leq 5}$  dans l'énoncé et vaut  $(\frac{37}{250}, \frac{55}{250}, \frac{45}{250}, \frac{53}{250}, \frac{60}{250})$ . Alors  $n \chi^2(\widehat{p}^n, p^0) = n \sum_{k=1}^N \frac{(p_k - p_k^0)^2}{p_k^0}$  s'écrit, avec les notations de l'énoncé

$$n \chi^2(\widehat{p}^n, p^0) = 250 \sum_{k=1}^5 \frac{(f_k - \frac{1}{5})^2}{\frac{1}{5}} = 1250 d_{obs}^2 = \frac{328}{62500}.$$

On veut construire un test au risque  $\alpha = 0,1$ . La théorie nous dit de chercher  $\beta$  tel que  $\mathbb{P}[\chi_4^2 > \beta] = \alpha = 0,1$ , où  $\chi_4^2$  suit une loi du  $\chi^2$  à  $N - 1 = 4$  degrés de liberté. La lecture dans une table donne  $\beta \simeq 7,78$ .

La règle s'énonce alors ainsi :

- si  $1250 d_{obs}^2 > \beta$ , on refuse l'hypothèse "le nombre de retraits est indépendant du jour de la semaine" ; la probabilité de se tromper est de l'ordre de 0,1 ;
- si  $1250 d_{obs}^2 \leq \beta$ , on ne peut pas refuser l'hypothèse d'indépendance du nombre de retraits par rapport au jour de la semaine.

Comme  $1250 d_{obs}^2 = 6,56$ , on ne peut pas conclure !

Remarquons que l'on ne peut jamais accepter l'hypothèse car on ne sait pas estimer la probabilité de se tromper dans ce cas. L'erreur de 10% correspond à la probabilité de rejeter à tort l'hypothèse, pas celle de l'accepter à tort. Il est donc ridicule de demander si "on peut affirmer, avec un risque d'erreur inférieur à 10%, que le nombre de retraits est indépendant du jour de la semaine", cela n'a aucun sens.

On remarque par ailleurs que l'on a  $\mathbb{P}[\chi_4^2 \leq \beta] = 0,9$ , c'est-à-dire que  $\beta$  est le neuvième décile de  $\chi_4^2$ . Mais les lois du  $\chi^2$  ne sont pas au programme de la terminale ES, et il n'est donc pas question de procéder en appliquant la théorie!... C'est pourquoi l'énoncé donne une simulation de ce qui serait en fait, à quelque chose près, une loi du  $\chi^2$ . Par miracle, le neuvième décile  $D_9$  sur la boîte à moustache vaut 6, ce qui n'est pas loin de  $\frac{1000}{1250}\beta$  (il s'agit d'une simulation de  $1000 d_{obs}^2$  et non de  $1250 d_{obs}^2$  comme la théorie le suggère d'où le facteur multiplicatif). Mais

ces simulations, ne sont que des simulations : elles sont obtenues avec des générateurs pseudo-aléatoires. Cela pose donc un problème sur la validité de ce type de méthode et on peut alors s'interroger sur l'intérêt de présenter ce type de problème...

### 6.6.7 À propos de la droite de Henry

La méthode dite "de la droite de Henry", au programme des classes de BTS, est une méthode pour déterminer les paramètres (moyenne et variance) d'une loi gaussienne qui "approche au mieux" un phénomène étudié.

Prenons un exemple (d'après *Publication 118 de la commission Inter-IREM Lycées techniques*). On étudie la durée de vie, exprimée en heures, de joints spéciaux. Sur un effectif de  $N = 500$  joints, on a obtenu les résultats suivants :

temps de fonctionnement $x_i$	500	700	900	1100	1300	1500	1700
effectifs $n_i$	24	67	108	126	109	51	15

Ensuite, on détermine les valeurs des réels  $t_i$  pour lesquels les fréquences cumulées croissantes  $f_i$  vérifient  $\mathbb{P}[N \leq t_i] = f_i$ , i.e.  $t_i = \Pi^{-1}(f_i)$  (sauf pour  $f_i = 1$  bien sûr!), où  $\Pi$  désigne la fonction de répartition de la variable gaussienne centrée réduite  $N$ . Pour cela, on n'oublie pas que, si  $f < \frac{1}{2}$ ,  $t = \Pi^{-1}(f)$  est strictement négatif et ne se trouve pas dans les tables. Comme alors  $1 - f > \frac{1}{2}$  de sorte que  $\Pi^{-1}(1 - f)$  se trouve dans les tables, et que  $1 = \Pi(t) + \Pi(-t)$ , il faut donc calculer  $t = \Pi^{-1}(f)$  à l'aide de  $t = -\Pi^{-1}(1 - f)$  dans ce cas.

$x_i$ temps de fonctionnement	500	700	900	1100	1300	1500	1700
effectifs $n_i$	24	67	108	126	109	51	15
effectifs cumulés croissants	24	91	199	325	434	485	500
fréquences cumulées croissantes $f_i$	0,048	0,182	0,398	0,65	0,868	0,97	1
$t_i = \Pi^{-1}(f_i)$	-1,6646	-0,9078	-0,2585	0,3853	1,1170	1,8808	$+\infty$

Si les couples  $(x_i, f_i)_{1 \leq i \leq 6}$  provenaient d'une variable gaussienne  $X$  de moyenne  $m$  et de variance  $\sigma^2$ , c'est-à-dire si  $f_i = \mathbb{P}[X \leq x_i]$  pour tout  $i \in \{1, \dots, 6\}$ , alors, puisque  $\frac{X-m}{\sigma}$  serait une gaussienne centrée réduite, les points  $(x_i, t_i)_{1 \leq i \leq 6}$  seraient liés par la relation  $t_i = \frac{x_i - m}{\sigma}$ , c'est-à-dire que les points  $(x_i, t_i)_{1 \leq i \leq 6}$  appartiendraient à la droite d'équation  $t = \frac{1}{\sigma}x - \frac{m}{\sigma}$ .

On appelle alors *droite de Henry*, la droite obtenue par la méthode des moindres carrés, pour le nuage de points  $(x_i, t_i)_{1 \leq i \leq 6}$  (ce qui est possible car par construction les points  $x_i$  ne sont pas tous égaux de sorte que  $\sigma_x \neq 0$ ). La droite de régression de  $t$  en  $x$ , d'équation  $t = \frac{\sigma_{x,t}}{\sigma_x^2}x + \bar{t} - \frac{\sigma_{x,t}}{\sigma_x^2}\bar{x}$  permet alors d'identifier le couple moyenne-variance cherché :  $\frac{1}{\sigma} = \frac{\sigma_{x,t}}{\sigma_x^2}$  et  $\frac{m}{\sigma} = \frac{\sigma_{x,t}}{\sigma_x^2}\bar{x} - \bar{t}$ , ce qui donne

$$\sigma = \frac{\sigma_x^2}{\sigma_{x,t}} \quad \text{et} \quad m = \bar{x} - \bar{t} \frac{\sigma_x^2}{\sigma_{x,t}}.$$

Dans l'exemple considéré, l'équation de la droite de régression de  $t$  en  $x$  est  $t = ax + b$  avec  $a \simeq 0,00349213$  et  $b \simeq -3,4000882$ , de sorte que  $m \simeq 973,643527$  et  $\sigma \simeq 286,358313$ .



*Remarque.* Dans l'exemple ci-dessus, comme  $m - 4\sigma < 0$ , on peut se demander s'il est bien judicieux de vouloir modéliser la durée de vie (positive) de joints par une variable gaussienne. Que l'on se rassure,  $m - 3,4\sigma > 0$ , ce qui entraîne que la probabilité qu'une variable gaussienne de moyenne  $m$  et de variance  $\sigma^2$  soit positive est supérieure à  $2\Pi(3,4) - 1 \simeq 0,99932\dots$

### 6.6.8 Table de lois du $\chi^2$

Si  $\chi_d^2$  est une variable aléatoire qui suit une loi du  $\chi^2$  à  $d$  degrés de liberté, la table suivante donne, pour  $\alpha$  donné, le nombre  $\beta_\alpha$  tel que  $\mathbb{P}[\chi_d^2 \geq \beta_\alpha] = \alpha$ , c'est-à-dire que  $\beta_\alpha$  est le fractile d'ordre  $1 - \alpha$  de la variable  $\chi_d^2$ .

$d \setminus \alpha$	0,99	0,975	0,95	0,90	0,10	0,05	0,025	0,01	0,001
1	0.00	0.00	0.00	0.02	2.71	3.84	5.02	6.63	10.83
2	0.02	0.05	0.10	0.21	4.61	5.99	7.38	9.21	13.82
3	0.11	0.22	0.35	0.58	6.25	7.81	9.35	11.34	16.27
4	0.30	0.48	0.71	1.06	7.78	9.49	11.14	13.28	18.47
5	0.55	0.83	1.15	1.61	9.24	11.07	12.83	15.09	20.52
6	0.87	1.24	1.64	2.20	10.64	12.59	14.45	16.81	22.46
7	1.24	1.69	2.17	2.83	12.02	14.07	16.01	18.48	24.32
8	1.65	2.18	2.73	3.49	13.36	15.51	17.53	20.09	26.12
9	2.09	2.70	3.33	4.17	14.68	16.92	19.02	21.67	27.88
10	2.56	3.25	3.94	4.87	15.99	18.31	20.48	23.21	29.59
11	3.05	3.82	4.57	5.58	17.28	19.68	21.92	24.72	31.26
12	3.57	4.40	5.23	6.30	18.55	21.03	23.34	26.22	32.91
13	4.11	5.01	5.89	7.04	19.81	22.36	24.74	27.69	34.53
14	4.66	5.63	6.57	7.79	21.06	23.68	26.12	29.14	36.12
15	5.23	6.26	7.26	8.55	22.31	25.00	27.49	30.58	37.70
16	5.81	6.91	7.96	9.31	23.54	26.30	28.85	32.00	39.25
17	6.41	7.56	8.67	10.09	24.77	27.59	30.19	33.41	40.79
18	7.01	8.23	9.39	10.86	25.99	28.87	31.53	34.81	42.31
19	7.63	8.91	10.12	11.65	27.20	30.14	32.85	36.19	43.82
20	8.26	9.59	10.85	12.44	28.41	31.41	34.17	37.57	45.31
21	8.90	10.28	11.59	13.24	29.62	32.67	35.48	38.93	46.80
22	9.54	10.98	12.34	14.04	30.81	33.92	36.78	40.29	48.27
23	10.20	11.69	13.09	14.85	32.01	35.17	38.08	41.64	49.73
24	10.86	12.40	13.85	15.66	33.20	36.42	39.36	42.98	51.18
25	11.52	13.12	14.61	16.47	34.38	37.65	40.65	44.31	52.62
26	12.20	13.84	15.38	17.29	35.56	38.89	41.92	45.64	54.05
27	12.88	14.57	16.15	18.11	36.74	40.11	43.19	46.96	55.48
28	13.56	15.31	16.93	18.94	37.92	41.34	44.46	48.28	56.89
29	14.26	16.05	17.71	19.77	39.09	42.56	45.72	49.59	58.30
30	14.95	16.79	18.49	20.60	40.26	43.77	46.98	50.89	59.70

Lorsque  $d > 30$ , la variable  $U = \sqrt{\chi_d^2} - \sqrt{2d-1}$  est telle que, pour tout intervalle  $I$  de  $\mathbb{R}$ ,  $\mathbb{P}[U \in I] \simeq \mathbb{P}[N \in I]$ , où  $N$  est une variable aléatoire centrée réduite.

# Bibliographie

- [CDF07] Françoise Couty, Jean Debord, and Daniel Fredon. *Mini-manuel de Probabilités et Statistique*. Dunod, 2007.
- [Chr] Le site Chronomath. <http://serge.mehl.free.fr/>.
- [Del10] Jean-François Delmas. *Introduction au calcul des probabilités et à la statistique*. Les presses de l'ENSTA, [http://cermics.enpc.fr/~delmas/Enseig/ensta\\_cours.pdf](http://cermics.enpc.fr/~delmas/Enseig/ensta_cours.pdf), 2010.
- [Fel68] William Feller. *An introduction to probability theory and its applications. Vol. I*. Third edition. John Wiley & Sons Inc., New York, 1968.
- [MPDG97] Sylvie Méléard, Claude Piquet, and Annette Decomps-Guilloux. *Analyse et probabilités*. Problèmes de Mathématiques. Écrit du CAPES 1991–1996. Masson, Paris, 1997. Avec rappels de cours.
- [PR12] Thérèse Phan and Jean-Pierre Rowenczyk. *Exercices et problèmes. Statistique et Probabilités*. Sciences sup. Dunod, 2012.
- [Pri05] Pierre Priouret. Polycopié du cours de Probabilités de L3. [http://www.proba.jussieu.fr/cours/proba\\_L\\_priouret.pdf](http://www.proba.jussieu.fr/cours/proba_L_priouret.pdf), 2004-2005.
- [Pub] Publication 118 de la commission inter-irem lycées techniques. la statistique en quatre séances. carnet de stage. [http://dutarte.perso.neuf.fr/statistique/Brochure\\\_118\\\_Statistique\\\_inferentielle.pdf](http://dutarte.perso.neuf.fr/statistique/Brochure\_118\_Statistique\_inferentielle.pdf).
- [Shi96] A. N. Shiryaev. *Probability*, volume 95 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edition, 1996. Translated from the first (1980) Russian edition by R. P. Boas.
- [Vel14] Yvan Velenik. Polycopié du cours de Probabilités et Statistique. <http://www.unige.ch/math/folks/velenik/cours.html>.