

Cours de Probabilités et Statistiques (IUYF MAT 8453)  
Telecom SudParis

Xavier ERNY

Année 2023-2024

# Table des matières

<b>A</b>	<b>Probabilités</b>	<b>3</b>
<b>I</b>	<b>Introduction à la théorie des probabilités</b>	<b>4</b>
I.1	Expérience aléatoire, événements . . . . .	4
I.2	Dénombrement, univers fini, loi uniforme . . . . .	5
I.3	Mesure de probabilité (ou loi) . . . . .	6
I.4	Indépendance et probabilités conditionnelles . . . . .	7
<b>II</b>	<b>Variables aléatoires</b>	<b>9</b>
II.1	Premières définitions . . . . .	9
II.2	Caractérisations des lois discrètes . . . . .	10
II.3	Caractérisations des lois continues . . . . .	11
II.4	Espérance, variance, moments . . . . .	12
II.5	Indépendance de variables aléatoires . . . . .	14
II.6	Lois usuelles discrètes . . . . .	15
II.7	Lois usuelles continues . . . . .	17
<b>III</b>	<b>Approximations asymptotiques</b>	<b>19</b>
III.1	Quelques inégalités probabilistes . . . . .	19
III.2	Loi (faible) des Grands Nombres . . . . .	19
III.3	Théorème Central Limite . . . . .	20
<b>B</b>	<b>Statistiques</b>	<b>21</b>
<b>IV</b>	<b>Modèles statistiques et estimateurs</b>	<b>22</b>
IV.1	Modèles statistiques . . . . .	22
IV.2	Qualité des estimateurs . . . . .	23
IV.3	Estimateurs empiriques . . . . .	24
IV.4	Méthode des moments . . . . .	25
<b>V</b>	<b>Estimation par intervalles de confiance</b>	<b>27</b>
V.1	Intervalle de confiance exact . . . . .	27
V.2	Intervalle de confiance asymptotique . . . . .	30
<b>VI</b>	<b>Tests statistiques</b>	<b>32</b>
VI.1	Généralités . . . . .	32
VI.2	Exemple de test . . . . .	33
<b>VII</b>	<b>Vraisemblance</b>	<b>35</b>
VII.1	Estimateur du maximum de vraisemblance . . . . .	36
VII.2	Test du rapport de vraisemblances . . . . .	37

# Introduction

Première partie

Probabilités

# Chapitre I

## Introduction à la théorie des probabilités

### I.1 Expérience aléatoire, événements

On ne définira pas formellement la notion "d'expérience aléatoire" car ce n'est pas une notion mathématique. Globalement, une expérience aléatoire est "quelque que chose qui peut produire des résultats différents de manière non-prévisibles". D'un point de vue de la terminologie mathématique, une expérience qui donne toujours le même résultat peut quand même être qualifiée d'aléatoire.

Intuitivement, ce que l'on appelle l'univers associée à une expérience aléatoire est l'ensemble des résultats possibles.

Formellement, un univers n'est rien d'autre qu'un ensemble non-vide (que l'on notera souvent  $\Omega$ ). Un événement de l'univers  $\Omega$  est une partie de  $\Omega$  (i.e. un élément de  $\mathcal{P}(\Omega)$ ). Dans les énoncés, l'univers sera rarement explicité, et il ne sera souvent pas nécessaire de le faire (le seul cas où il faudra le faire est celui décrit à la Section I.2 suivante).

#### Exemples I.1.1.

- On considère l'expérience "on lance une pièce (équilibrée ou non)" et l'événement  $A :=$  "la pièce fait pile". Un univers possible serait  $\Omega = \{P, F\}$  et l'événement  $A$  serait alors  $A = \{P\}$ .
- On considère l'expérience "on lance un dé à six faces (équilibré ou non)" et l'événement  $A :=$  "le nombre obtenu est impair". Un univers possible serait  $\Omega = \{1, 2, 3, 4, 5, 6\}$  et l'événement  $A$  serait alors  $A = \{1, 3, 5\}$ .

Pour chaque expérience aléatoire, il y a toujours une infinité d'univers possible qui peut lui être associé. Il suffit d'en choisir un suffisamment grand pour dissocier les différents résultats de l'expérience (par exemple, pour un lancé de dé à six faces, vous pouvez prendre n'importe quel ensemble de cardinal supérieur ou égal à six).

**Définition I.1.2** (Événement). Soit  $\Omega$  un univers (i.e.  $\Omega$  est un ensemble non-vide). Alors :

- un événement  $A$  est une partie de  $\Omega$ ,
- si  $\omega \in \Omega$ , alors  $\{\omega\}$  est un événement atomique.

**Remarque I.1.3** (Lien entre ensemble et proposition). Par définition un événement est donc un ensemble, mais dans les énoncés on les définit plutôt par des propositions (énoncés en français) qui peuvent être vraies ou fausses. Intuitivement, ces deux notions sont équivalentes en gardant l'égalité suivante en tête :

$$A = \{\omega \in \Omega : \mathcal{A}(\omega) \text{ est vraie}\},$$

où  $A$  (dans le membre gauche) est l'événement défini comme un ensemble, et  $\mathcal{A}(\omega)$  (dans le membre droit) est l'événement défini comme une proposition qui peut être vraie ou fausse selon  $\omega$ . Réciproquement, on peut aussi écrire (avec les mêmes notations)

$$\mathcal{A}(\omega) = \text{"}\omega \in A\text{"}.$$

**Exemple I.1.4.** Reprenons le deuxième point de l'Exemple I.1.1, où  $\Omega = \{1, 2, 3, 4, 5, 6\}$  et  $A = \{1, 3, 5\}$ . Avec les notations de la Remarque I.1.3, on a

$$\mathcal{A}(1) = \mathcal{A}(3) = \mathcal{A}(5) = \text{vrai}; \quad \mathcal{A}(2) = \mathcal{A}(4) = \mathcal{A}(6) = \text{faux}.$$

**Définition I.1.5** (Opérations ensemblistes). Soient  $A$  et  $B$  deux parties d'un ensemble  $\Omega$  (ou événements d'un univers  $\Omega$ ). On définit les événements suivants :

- la réunion de  $A$  et  $B$  :  $A \cup B = \{\omega \in \Omega : \omega \in A \text{ ou } \omega \in B\}$ ,
- l'intersection de  $A$  et  $B$  :  $A \cap B = \{\omega \in \Omega : \omega \in A \text{ et } \omega \in B\}$ ,
- le complémentaire de  $A$  :  $A^c = \{\omega \in \Omega : \omega \notin A\}$ .

**Lemme I.1.6.** Soient  $A, B, C$  trois événements d'un univers  $\Omega$ . Alors :

- $(A \cap B)^c = A^c \cup B^c$ ,
- $(A \cup B)^c = A^c \cap B^c$ ,
- $(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$ ,
- $(A \cap B) \cup C = (A \cup C) \cap (B \cup C)$ .

*Démonstration.* Cf cours. □

**Définition I.1.7** (Unions et intersections infinies). Soient  $A_1, \dots, A_n, \dots$  une suite infinie d'événements d'un univers  $\Omega$ . On définit les événements suivants :

- la réunion infinie :  $\bigcup_{k=1}^{\infty} A_k = \{\omega \in \Omega : \exists k \in \mathbb{N}^*, \omega \in A_k\}$ ,
- l'intersection infinie :  $\bigcap_{k=1}^{\infty} A_k = \{\omega \in \Omega : \forall k \in \mathbb{N}^*, \omega \in A_k\}$ .

## I.2 Dénombrement, univers fini, loi uniforme

Une formule très courante pour calculer la probabilité d'un événement  $A$  (mais aussi très souvent mal justifiée et parfois mal comprise) est la suivante :

$$\mathbb{P}(A) = \frac{\text{nombre d'issues favorables pour } A}{\text{nombre d'issues totales}}.$$

Il n'y a qu'un seul contexte où cette formule est vraie : l'univers  $\Omega$  est fini et  $\mathbb{P}$  est la mesure de probabilité uniforme sur  $\Omega$  (nous donnerons à la Section I.3 suivante un sens formel à cette phrase, cf Exemple I.3.3). Dans ce contexte, tout événement  $A$  est aussi fini (car inclus dans  $\Omega$ ), et la formule rigoureuse est la suivante

$$\mathbb{P}(A) = \frac{|A|}{|\Omega|}. \tag{I.1}$$

**Exemple I.2.1.** Prenons l'exemple de l'expérience "on lance un dé équilibré à six faces", avec l'événement  $A = \text{"le résultat est impair"}$ . Pour calculer la probabilité de  $A$  avec la formule ci-dessus, il faut considérer (par exemple) l'univers  $\Omega = \{1, 2, 3, 4, 5, 6\}$  (muni de la loi uniforme, car le dé est équilibré) pour lequel  $A = \{1, 3, 5\}$ . On a alors

$$\mathbb{P}(A) = \frac{|A|}{|\Omega|} = \frac{3}{6} = \frac{1}{2}.$$

Les lemmes suivants sont les résultats les plus courants pour calculer des cardinaux d'ensemble dans ce contexte.

**Lemme I.2.2** (Choix simultanés). Soient  $0 \leq k \leq n$  deux entiers naturels. Alors le nombre de parties à  $k$  éléments d'un ensemble à  $n$  éléments est donné par le coefficient binomial :

$$\binom{n}{k} = C_n^k = \frac{n!}{k!(n-k)!},$$

où  $0! := 1$ .

**Exemple I.2.3.** Considérons une urne contenant 8 boules bleues et 3 rouges. On pioche simultanément 2 boules. Calculer la probabilité que les deux boules piochées soient de couleurs différentes (en supposant que toutes les paires ont la même probabilité d'être piochées).

*Indice* : expliciter un univers  $\Omega$  convenable et utiliser le lemme précédent avec la formule (I.1).

**Lemme I.2.4** (Choix successifs). Soient  $0 \leq k \leq n$  deux entiers naturels et  $E$  un ensemble à  $n$  éléments. Alors :

— le nombre de  $k$ -uplets d'éléments de  $E$  sans répétition est :

$$n(n-1)\dots(n-k+1) = \frac{n!}{(n-k)!},$$

— et le nombre de  $k$ -uplets d'éléments de  $E$  (avec possiblement des répétitions) est :  $n^k$ .

**Exemple I.2.5.** Considérons une urne contenant 8 boules bleues et 3 rouges. On pioche successivement 2 boules (sans les remettre). Calculer la probabilité que les deux boules piochées soient de couleurs différentes (en supposant que tous les couples sont équiprobables).

*Indice* : expliciter un univers  $\Omega$  convenable et utiliser le lemme précédent avec la formule (I.1).

**Remarque I.2.6.** En généralisant les deux exemples précédents, il est possible de montrer que les deux expériences suivantes sont équivalentes :

- "piocher simultanément  $k$  objets parmi  $n$ ",
- "piocher successivement et sans remise  $k$  objets parmi  $n$ , sans tenir compte de l'ordre".

### I.3 Mesure de probabilité (ou loi)

**Définition I.3.1** (Mesure de probabilité). Soit  $\Omega$  un ensemble (qui est interprété comme un univers). Une mesure de probabilité (ou loi) sur  $\Omega$  est une application  $\mathbb{P} : \mathcal{P}(\Omega) \rightarrow [0, 1]$  qui vérifie :

- $\mathbb{P}(\emptyset) = 0$ ,
- $\mathbb{P}(\Omega) = 1$ ,
- et, pour toutes parties  $A_1, \dots, A_n, \dots$  de  $\Omega$  disjointes deux-à-deux,

$$\mathbb{P}\left(\bigcup_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} \mathbb{P}(A_k) := \lim_{n \rightarrow \infty} \sum_{k=1}^n \mathbb{P}(A_k).$$

On dit que  $(\Omega, \mathbb{P})$  est un espace de probabilité (ou espace probabilisé).

**Remarque I.3.2.** La Définition I.3.1 est incomplète et en partie fautive : en général, on ne peut pas définir la mesure de probabilités  $\mathbb{P}$  sur toutes les parties de  $\Omega$  pour des raisons techniques qui ne seront pas abordées dans ce cours.

#### Exemples I.3.3.

- Si on s'intéresse au résultat d'un dé à six faces, alors il est naturel de considérer l'univers  $\Omega = \llbracket 1, 6 \rrbracket$ . Si le dé est équilibré, il faut munir  $\Omega$  de la mesure de probabilité uniforme : pour tout  $A \subseteq \Omega$ ,

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

qui définit bien une mesure de probabilité (la vérification de ce fait est laissée en exercice). Si le dé n'est pas équilibré et que l'on note  $p_k$  la probabilité d'obtenir  $k$  (où  $1 \leq k \leq 6$ ), alors la mesure de probabilité à considérer est la suivante

$$\mathbb{P}(A) := \sum_{k \in A} p_k.$$

- Si on pioche une boule dans une urne contenant 5 boules rouges et 3 boules bleues, alors on a vu précédemment que l'on pouvait choisir un univers  $\Omega$  à 8 éléments muni de la probabilité uniforme (où parmi les éléments, on en choisit 5 qui désignent les boules rouges, et les autres les boules bleues), mais on peut aussi choisir un univers à 2 éléments  $\Omega = \{R, B\}$  muni de la mesure de probabilité suivante

$$\mathbb{P} : \begin{cases} \mathcal{P}(\Omega) & \longrightarrow & [0, 1] \\ \emptyset & \longmapsto & 0 \\ \{R\} & \longmapsto & 5/8 \\ \{B\} & \longmapsto & 3/8 \\ \Omega & \longmapsto & 1 \end{cases}$$

**Proposition I.3.4.** Soit  $(\Omega, P)$  un espace de probabilité. Alors, pour toutes parties  $A, B, A_1, \dots, A_n$  de  $\Omega$  :

- si  $A_1, \dots, A_n$  sont disjointes deux-à-deux, alors

$$\mathbb{P} \left( \bigcup_{k=1}^n A_k \right) = \sum_{k=1}^n \mathbb{P}(A_k),$$

- $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$ ,
- si  $A \subseteq B$  alors  $\mathbb{P}(A) \leq \mathbb{P}(B)$ ,
- $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$ .

Démonstration. Cf cours. □

**Proposition I.3.5.** Soient  $(\Omega, P)$  un espace de probabilité et  $(A_k)_{k \geq 1}$  une suite de parties de  $\Omega$ . Alors :

- si  $A_1 \subseteq A_2 \subseteq \dots \subseteq A_n \subseteq \dots$ , alors

$$\mathbb{P} \left( \bigcup_{k=1}^{\infty} A_k \right) = \lim_{k \rightarrow \infty} \mathbb{P}(A_k),$$

- si  $\dots \subseteq A_n \subseteq \dots \subseteq A_2 \subseteq A_1$ , alors

$$\mathbb{P} \left( \bigcap_{k=1}^{\infty} A_k \right) = \lim_{k \rightarrow \infty} \mathbb{P}(A_k).$$

Démonstration. Cf cours. □

## I.4 Indépendance et probabilités conditionnelles

**Définition I.4.1** (Indépendance). Soit  $(\Omega, \mathbb{P})$  un espace de probabilité. Deux événements  $A, B$  sont indépendants ssi :

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

**Remarque I.4.2.** Il ne faut pas confondre "A et B sont disjoints" (i.e.  $A \cap B = \emptyset$ ) avec "A et B sont indépendants" (i.e.  $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$ ).

**Définition I.4.3** (Probabilité conditionnelle). Soient  $(\Omega, \mathbb{P})$  un espace de probabilité et  $A$  un événement de probabilité non-nulle (i.e.  $\mathbb{P}(A) > 0$ ). Alors la probabilité conditionnellement à  $A$ , notée  $\mathbb{P}_A$  ou  $\mathbb{P}(\bullet|A)$  est définie par

$$\begin{aligned} \mathbb{P}_A : \mathcal{P}(\Omega) &\longrightarrow [0, 1] \\ B &\longmapsto \mathbb{P}_A(B) := \mathbb{P}(B|A) := \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)}. \end{aligned}$$

**Remarque I.4.4.** Les notations  $\mathbb{P}_A(B)$  et  $\mathbb{P}(B|A)$  se lisent "probabilité de  $B$  sachant  $A$ ".

**Lemme I.4.5.** Soient  $(\Omega, \mathbb{P})$  un espace de probabilité, et  $A, B$  deux événements de probabilités non-nulles. Alors les assertions suivantes sont équivalentes :

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

*Démonstration.* Cf cours. □

Il faut faire attention à un point qui peut sembler contre-intuitif à première vue : le fait que deux événements sont indépendants ne signifie pas que la réalisation de l'un n'affecte pas la réalisation de l'autre, mais plutôt que ça n'affecte pas la probabilité de sa réalisation (voir Lemme I.4.5 ci-dessus). L'exemple suivant illustre ce point.

**Exemple I.4.6.** On lance deux pièces équilibrées. Soient  $A$  l'événement "la première pièce fait pile" et  $B$  l'événement "les deux pièces donnent le même résultat". Considérons la question "A et B sont-ils indépendants?". Un raisonnement faux (qui consisterait à dire que la réalisation de  $A$ , impose une réalisation particulière de  $B$  (i.e. "double pile")) pourrait faire penser que  $A$  et  $B$  ne sont pas indépendants.

**Théorème I.4.7** (Formule des probabilités totales). Soient  $(\Omega, \mathbb{P})$  un espace de probabilité et  $A_1, \dots, A_n$  des événements de probabilités non-nulles, disjoints deux-à-deux tels que

$$\bigcup_{k=1}^n A_k = \Omega.$$

Alors, pour tout événement  $B$ ,

$$\mathbb{P}(B) = \sum_{k=1}^n \mathbb{P}(B|A_k) \mathbb{P}(A_k).$$

*Démonstration.* Cf cours. □

# Chapitre II

## Variables aléatoires

### II.1 Premières définitions

Informellement, une variable aléatoire réelle est un nombre réel qui dépend du résultat d'une expérience aléatoire.

**Définition II.1.1** (Variable aléatoire réelle). *Soit  $(\Omega, \mathbb{P})$  un espace de probabilité. Une variable aléatoire réelle (abrégée en v.a.r.)  $X$  est une fonction de  $\Omega$  dans  $\mathbb{R}$  :*

$$X : \omega \in \Omega \mapsto X(\omega) \in \mathbb{R}.$$

**Remarque II.1.2.** *De même que ce qui est expliqué à la Remarque I.3.2, la définition ci-dessus est incomplète : pour qu'une fonction de  $\Omega$  dans  $\mathbb{R}$  soit une v.a.r. il faut aussi qu'elle soit mesurable (cette notion ne sera pas abordée dans ce cours).*

**Définition II.1.3** (Ensemble des valeurs d'une v.a.r.). *Soit  $X$  une v.a.r. définie sur un espace  $(\Omega, \mathbb{P})$ . L'ensemble des valeurs prises par  $X$  est :*

$$X(\Omega) := \{X(\omega) : \omega \in \Omega\} \subseteq \mathbb{R}.$$

**Exemple II.1.4.** *En considérant l'expérience "on lance un dé à six faces", que l'on peut modéliser avec l'univers  $\Omega = \{1, 2, 3, 4, 5, 6\}$ , le résultat du dé  $X(\omega) = \omega$  est une variable aléatoire à valeurs dans  $\{1, 2, 3, 4, 5, 6\}$ .*

**Remarque II.1.5.** *Si  $X$  est une v.a.r. sur un espace  $(\Omega, \mathbb{P})$ , et si  $I$  est une partie de  $\mathbb{R}$ , alors l'ensemble*

$$\{\omega \in \Omega : X(\omega) \in I\}$$

*est une partie de  $\Omega$ , et est donc un événement. Cet événement est noté  $\{X \in I\}$ .*

**Définition II.1.6** (Loi d'une v.a.r.). *Soit  $X$  une v.a.r. définie sur un espace  $(\Omega, \mathbb{P})$ . La loi de  $X$  est la mesure de probabilité  $\mathbb{P}_X$  définie sur l'ensemble  $\mathbb{R}$  par :*

$$\mathbb{P}_X : I \in \mathcal{P}(\mathbb{R}) \mapsto \mathbb{P}(X \in I) \in [0, 1].$$

**Définition II.1.7** (Fonction de répartition). *Soit  $X$  une v.a.r. définie sur un espace  $(\Omega, \mathbb{P})$ . La fonction de répartition de  $X$  est définie par*

$$F_X : t \in \mathbb{R} \mapsto F_X(t) := \mathbb{P}(X \leq t) \in [0, 1].$$

**Lemme II.1.8.** *Soient  $X$  une v.a.r. et  $a < b$  deux réels. Alors :*

- $\mathbb{P}(X \in ]a, b]) = F_X(b) - F_X(a)$ ,
- $\mathbb{P}(X = a) = F_X(a) - F_X(a-)$ .

Démonstration. Cf cours. □

**Proposition II.1.9.** Soit  $X$  une v.a.r. définie sur un espace  $(\Omega, \mathbb{P})$ . La fonction de répartition de  $X$  vérifie les propriétés suivantes :

- $F_X$  est croissante sur  $\mathbb{R}$ ,
- $\lim_{t \rightarrow -\infty} F_X(t) = 0$ ,
- $\lim_{t \rightarrow +\infty} F_X(t) = 1$ ,
- $F_X$  est continue à droite en tout point :  $\forall t \in \mathbb{R}, F_X(t+) := \lim_{s \rightarrow t+} F_X(s) = F_X(t)$ .

**Remarque II.1.10.** Si une fonction  $F$  vérifie les quatre conditions de la Proposition II.1.9 ci-dessus, alors  $F$  est nécessairement la fonction de répartition d'une v.a.r. (i.e. il existe une v.a.r.  $X$  sur un certain espace  $(\Omega, \mathbb{P})$  telle que  $F = F_X$ ).

**Proposition II.1.11.** Soient  $X, Y$  deux v.a.r. (qui ne sont pas nécessairement définies sur le même espace de probabilité). Alors :

1. si  $P_X = P_Y$ , alors  $F_X = F_Y$ ,
2. si  $F_X = F_Y$ , alors  $P_X = P_Y$ .

Démonstration. Pour le premier point voir cours. Le second point est admis (la preuve étant trop technique pour ce cours). □

## II.2 Caractérisations des lois discrètes

**Définition II.2.1** (V.a.r. discrète). Une v.a.r.  $X$  est dite discrète si  $X(\Omega)$  est fini ou dénombrable.

**Remarque II.2.2.** Il est équivalent de dire qu'une v.a.r. est discrète ou que sa loi est discrète.

**Lemme II.2.3.** Soit  $X$  une v.a.r. discrète. Alors la loi de  $X$  est caractérisée par la suite (finie ou dénombrable) :

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

**Exemples II.2.4.**

- Soit  $X$  une v.a.r. modélisant la résultat d'un jet de dé à six faces équilibré. Alors  $X(\Omega) = \{1, 2, 3, 4, 5, 6\}$  est finie, et pour tout  $x \in \{1, 2, 3, 4, 5, 6\}$ ,  $\mathbb{P}(X = x) = 1/6$ .
- On considère un lancer de pièce, et on définit  $X = 1$  si obtient pile, et  $X = 0$  si on obtient face. Alors  $X(\Omega) = \{0, 1\}$  est fini, et si on suppose que la pièce n'est pas équilibré (avec par exemple la probabilité d'obtenir pile égale à  $2/3$ ), alors  $\mathbb{P}(X = 0) = 1/3$  et  $\mathbb{P}(X = 1) = 2/3$ .

**Théorème II.2.5.** Soit  $X$  une v.a.r. telle que  $X(\Omega) = \{x_1, \dots, x_n\}$  est fini, où  $x_1 < x_2 < \dots < x_n$ . Alors :

- $\sum_{k=1}^n \mathbb{P}(X = x_k) = 1$ ,
- $\forall t < x_1, F_X(t) = 0$ ,
- $\forall t \geq x_n, F_X(t) = 1$ ,
- $\forall t \in [x_k, x_{k+1}[, F_X(t) = \sum_{j=1}^k \mathbb{P}(X = x_j)$ .

Démonstration. Cf cours. □

On pourrait énoncé un théorème similaire à celui ci-dessus dans le cas où  $X(\Omega)$  est infini dénombrable, mais nous ne l'utiliserons pas. Ce qui est important à retenir est plutôt l'utilisation de ce résultat que l'énoncé en lui-même (i.e. Remarque II.2.7).

**Corollaire II.2.6.** Soit  $X$  une v.a.r. telle que  $X(\Omega) = \{x_1, \dots, x_n\}$  est fini, où  $x_1 < x_2 < \dots < x_n$  (resp.  $X(\Omega) = \{x_1, \dots, x_k, \dots\}$  infini dénombrable avec  $x_1 < \dots < x_k \dots$ ). Alors, pour tout  $1 \leq k \leq n$  (resp.  $k \geq 1$ ),

$$\mathbb{P}(X = x_k) = F_X(x_k) - F_X(x_{k-}).$$

La remarque suivante est une application simple, mais importante du Théorème II.2.5 et du Corollaire II.2.6 ci-dessus.

**Remarque II.2.7.** Soit  $X$  une v.a.r. discrète. Alors sa fonction de répartition  $F_X$  est une fonction en escalier croissante à valeurs dans  $[0, 1]$  telle que les réels  $t$  où  $F_X$  saute sont exactement les réels  $t$  tels que  $\mathbb{P}(X = t) > 0$  (et dans ce cas  $\mathbb{P}(X = t)$  est l'amplitude du saut).

Un point important de cette section est que, si vous considérer une v.a.r.  $X$  qui ne prend qu'un nombre fini de valeurs  $x_1 < \dots < x_n$  (dans le cas infini dénombrable, la méthode est la même, mais le résultat ne sera en général pas explicite), il est très simple de passer de la suite  $(\mathbb{P}(X = x_k))_{1 \leq k \leq n}$  à la fonction de répartition  $F_X$  (cf Théorème II.2.5 ci-dessus), et inversement (voir Exemple ci-dessous).

**Exemple II.2.8.** Soit  $X$  une v.a.r. dont la fonction de répartition  $F_X$  est définie par :

$$F_X(t) = \begin{cases} 0 & \text{si } t < 2 \\ 1/5 & \text{si } 2 \leq t < 5 \\ 1/2 & \text{si } 5 \leq t < 6 \\ 1 & \text{si } 6 \leq t \end{cases}$$

Alors  $X(\Omega) = \{2, 5, 6\}$  et

$$\begin{array}{c|c|c|c} x & 2 & 5 & 6 \\ \hline \mathbb{P}(X = x) & 1/5 & 1/2 - 1/5 = 3/10 & 1 - 1/2 = 1/2 \end{array}.$$

## II.3 Caractérisations des lois continues

**Définition II.3.1** (V.a.r. continue (ou à densité)). Une v.a.r.  $X$  est dite continue (ou à densité) s'il existe une fonction  $f_X : \mathbb{R} \rightarrow \mathbb{R}_+$  continue par morceaux telle que, pour tout  $-\infty \leq a \leq b \leq +\infty$ ,

$$\mathbb{P}(X \in [a, b]) = \mathbb{P}(X \in ]a, b]) = \mathbb{P}(X \in [a, b[) = \mathbb{P}(X \in ]a, b[) = \int_a^b f_X(x) dx.$$

On dit que  $f_X$  est la densité de  $X$  (ou de sa loi).

**Remarque II.3.2.** Pour qu'une fonction  $f$  soit la densité d'une v.a.r., elle doit vérifier les trois conditions suivantes :

- $f_X$  est continue par morceaux (cette condition peut être généralisée),
- $f_X$  est à valeurs dans  $\mathbb{R}_+$ ,
- et  $\int_{-\infty}^{+\infty} f_X(x) dx = 1$ .

Dans la plupart des cas, les deux premières conditions sont immédiates à vérifier, mais pour autant, il ne faut pas les oublier.

**Proposition II.3.3.** Si deux densités sont égales partout sauf sur un ensemble fini ou dénombrable de points, alors elles définissent la même loi. Autrement dit, si  $X, Y$  sont deux v.a.r. à densité et s'il existe une partie  $A \subseteq \mathbb{R}$  finie ou dénombrable telle que

$$\forall x \in \mathbb{R} \setminus A, f_X(x) = f_Y(x),$$

alors  $P_X = P_Y$ .

Démonstration. Admis. □

**Lemme II.3.4.** Soit  $X$  une v.a.r. continue de densité  $f_X$ . Alors la fonction de répartition de  $X$  est continue et vérifie : pour tout  $t \in \mathbb{R}$ ,

$$F_X(t) = \int_{-\infty}^t f_X(x)dx.$$

En particulier, si  $t$  est un point de continuité de  $f_X$ , alors  $F_X$  est dérivable en  $t$ , et  $F'_X(t) = f_X(t)$ .

Démonstration. Cf cours. □

**Remarque II.3.5.** Le lemme précédent est un des résultats principaux de cette section. Il permet de passer facilement de la densité à la fonction de répartition (par intégration), et inversement (par dérivation). Il arrive souvent que la fonction de répartition d'une v.a.r. continue ne soit pas dérivable partout. Dans ce cas, vous pouvez donner les valeurs de la densité  $f_X$  où  $F_X$  est dérivable, et pour les points de non-dérivabilité de  $F_X$ , vous pouvez donner à  $f_X$  n'importe quelle valeur (cela définira toujours la même loi, peu importe les valeurs choisies : cf Proposition II.3.3).

**Lemme II.3.6.** Soit  $X$  une v.a.r. continue. Alors : pour tout  $\alpha \in \mathbb{R}$ ,

$$\mathbb{P}(X = \alpha) = 0.$$

## II.4 Espérance, variance, moments

La notion d'espérance peut être définie de la même manière pour n'importe quelle v.a.r. Toutefois, cette définition générale étant hors-programme, nous la définirons plutôt en distinguant le cas des v.a.r. discrètes et des v.a.r. continues.

**Définition II.4.1** (Espérance). Si  $X$  une v.a.r. discrète, son espérance est définie par

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

Si  $X$  est une v.a.r. continue, son espérance est définie par

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

**Remarque II.4.2.** Il peut arriver que l'espérance d'une v.a.r. n'existe pas, ou qu'elle existe mais soit infinie.

L'étude rigoureuse et exhaustive des cas où l'espérance est bien définie serait trop technique pour ce cours (cela nécessiterait d'étudier la théorie de la mesure et de l'intégration au sens de Lebesgue). Dans la suite du cours, nous nous contenterons du lemme suivant qui donne quelques conditions suffisantes qui permettent de garantir que l'espérance est bien définie (cela sera suffisant pour les v.a.r. que l'on étudiera).

**Lemme II.4.3.** Soit  $X$  une v.a.r. :

- Si  $X(\Omega)$  est inclus dans  $\mathbb{R}_+$  (ou  $\mathbb{R}_-$ ), alors l'espérance de  $X$  est bien définie (éventuellement infinie).
- Si  $X(\Omega)$  est inclus dans un intervalle borné (en particulier si  $X(\Omega)$  est fini par exemple), alors l'espérance de  $X$  est bien définie et finie.

- Si  $\mathbb{E}[|X|] < \infty$  (qui a bien sens d'après le premier point de ce lemme), alors l'espérance de  $X$  est bien définie et finie.

**Lemme II.4.4** (Linéarité de l'espérance). Soient  $X, Y$  deux v.a.r. dont les espérances sont bien définies et finies, et  $\alpha$  un réel. Alors :

- $\mathbb{E}[\alpha] = \alpha$ ,
- $\mathbb{E}[\alpha X] = \alpha \mathbb{E}[X]$ ,
- $\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$ .

*Démonstration.* Cf cours. □

**Lemme II.4.5** (Positivité de l'espérance). Soient  $X, Y$  deux v.a.r. dont les espérances sont bien définies et finies. Alors :

- si  $X \geq 0$ , alors  $\mathbb{E}[X] \geq 0$ ,
- si  $X \geq Y$ , alors  $\mathbb{E}[X] \geq \mathbb{E}[Y]$ .

*Démonstration.* Cf cours. □

**Théorème II.4.6** (Théorème de transfert). Soient  $X$  une v.a.r. et  $g$  une fonction continue par morceaux telles que l'espérance  $g(X)$  est bien définie. Alors :

- si  $X$  est une v.a.r. discrète,

$$\mathbb{E}[g(X)] = \sum_{x \in X(\Omega)} g(x) \mathbb{P}(X = x),$$

- et, si  $X$  est une v.a.r. continue, alors

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

*Démonstration.* Admis. □

Notons que si la fonction  $g$  est continue par morceaux et, soit positive, soit bornée, alors le théorème de transfert est utilisable pour calculer  $\mathbb{E}[g(X)]$  (cf Lemme II.4.3).

**Définition II.4.7** (Variance, covariance, écart-type). Soient  $X, Y$  deux v.a.r. telles que  $\mathbb{E}[X^2] < \infty$  et  $\mathbb{E}[Y^2] < \infty$ , on définit les quantités suivantes :

- la variance de  $X$  :  $\mathbb{V}[X] = \mathbb{E}[(X - \mathbb{E}[X])^2]$ ,
- la covariance de  $X$  et  $Y$  :  $\text{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X] \mathbb{E}[Y]$ ,
- l'écart-type de  $X$  :  $\sigma(X) = \sqrt{\mathbb{V}[X]}$ .

**Remarque II.4.8.** Si  $X$  est une v.a.r. telle que  $\mathbb{E}[X^2] < \infty$ , alors sa variance existe nécessairement, est finie et positive. Et si  $\mathbb{E}[X^2] = \infty$ , la variance ne peut pas être définie. En particulier, dès que la variance a un sens, elle est nécessairement finie (ce qui n'est pas le cas de l'espérance).

**Lemme II.4.9** (Formule de König-Huygens). Soit  $X$  une v.a.r. telle que  $\mathbb{E}[X^2] < \infty$ , alors

$$\mathbb{V}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$$

*Démonstration.* Cf cours. □

**Lemme II.4.10.** Soient  $X, Y, Z$  des v.a.r. de carré intégrables, et  $\alpha, \beta$  des réels. Alors :

- $\text{Cov}(X, Y) = \text{Cov}(Y, X)$ ,
- $\text{Cov}(\alpha X + \beta Y, Z) = \alpha \text{Cov}(X, Z) + \beta \text{Cov}(Y, Z)$ ,
- $\mathbb{V}[\alpha] = 0$ ,
- $\mathbb{V}[X] = \text{Cov}(X, X)$ ,
- $\mathbb{V}[\alpha X] = \alpha^2 \mathbb{V}[X]$ ,

$$- \mathbb{V}[X + Y] = \mathbb{V}[X] + \mathbb{V}[Y] + 2\text{Cov}(X, Y).$$

*Démonstration.* Cf cours □

**Définition II.4.11** (Corrélation). *Deux v.a.r.  $X, Y$  sont non-corrélées si  $\text{Cov}(X, Y) = 0$ .*

La non-corrélation est une notion importante en statistiques, car lorsque l'on considère  $n$  v.a.r.  $X_1, \dots, X_n$  deux-à-deux non-corrélés, alors (d'après le dernier point du lemme ci-dessus)

$$\mathbb{V}\left[\sum_{k=1}^n X_k\right] = \sum_{k=1}^n \mathbb{V}[X_k].$$

Cette relation sera beaucoup utilisée dans la suite.

**Définition II.4.12** (Moments). *Soit  $X$  une v.a.r. et  $p$  un entier. Le moment d'ordre  $p$  de  $X$  est l'espérance de  $X^p$  (si elle est bien définie).*

## II.5 Indépendance de variables aléatoires

**Définition II.5.1** (Indépendance entre deux v.a.r.). *Soient  $X, Y$  deux v.a.r. On dit que  $X$  et  $Y$  sont indépendantes si pour toute parties  $A, B$  de  $\mathbb{R}$*

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

*autrement dit, les événements  $\{X \in A\}$  et  $\{Y \in B\}$  sont indépendants.*

**Remarque II.5.2.** *Encore une fois, la définition ci-dessus est en partie fautive. La relation ne doit pas être vraie pour toutes les parties de  $\mathbb{R}$ , mais seulement pour les parties dites mesurables (nous n'étudierons pas cette notion dans ce cours). En pratique, pour que deux v.a.r. soient indépendantes, il suffit de vérifier la relation ci-dessus pour les parties  $A, B$  qui sont des intervalles.*

**Définition II.5.3** (Indépendance mutuelle de v.a.r.). *Soient  $X_1, \dots, X_n$  des v.a.r. On dit que  $X_1, \dots, X_n$  sont mutuellement indépendantes si pour toute parties  $A_1, \dots, A_n$  de  $\mathbb{R}$*

$$\mathbb{P}\left(\bigcap_{k=1}^n \{X_k \in A_k\}\right) = \prod_{k=1}^n \mathbb{P}(X_k \in A_k).$$

**Remarque II.5.4.** *L'indépendance mutuelle de v.a.r.  $X_1, \dots, X_n$  implique toujours l'indépendance deux-à-deux de ces v.a.r., mais la réciproque est fautive en général.*

**Lemme II.5.5.** *Soient  $X, Y$  des v.a.r. discrètes. Alors dire que  $X$  et  $Y$  sont indépendantes équivaut à dire que : pour tout  $x \in X(\Omega)$  et  $y \in Y(\Omega)$ ,*

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

*Démonstration.* Cf cours. □

Il est possible d'énoncer un lemme équivalent au précédent concernant l'indépendance mutuelle de v.a.r.

**Lemme II.5.6.** *Soient  $X, Y$  deux v.a.r. de carré intégrable. Alors l'indépendance entre  $X$  et  $Y$  implique leur non-corrélation.*

*Démonstration.* Admis. □

## II.6 Lois usuelles discrètes

**Définition II.6.1** (Loi uniforme sur un ensemble fini  $E : \mathcal{U}(E)$ ). Soit  $E = \{x_1, \dots, x_n\}$  une partie finie de  $\mathbb{R}$  de cardinal  $n$ . Une v.a.r.  $X$  qui suit la loi uniforme sur  $E$  est une v.a.r. discrète telle que  $X(\Omega) = E$  et, pour chaque  $1 \leq k \leq n$ ,  $\mathbb{P}(X = x_k) = 1/n$ .

**Exemple II.6.2.** Les exemples typiques où on modélise une expérience aléatoire avec la loi uniforme sur un ensemble fini est le lancé d'un dé équilibré (où  $n$  est le nombre de faces, et les  $x_k$  les numéros des faces), ou le tirage de boules dans une urne (où  $n$  est le nombre de boules). Ces exemples ont été détaillés à la Section I.2.

**Lemme II.6.3.** Soit  $X$  une v.a.r. uniforme sur  $\{x_1, \dots, x_n\}$  avec  $x_1 < x_2 < \dots < x_n$ . Alors la fonction de répartition de  $X$  vérifie

$$F_X(t) = \begin{cases} 0 & \text{si } t < x_1 \\ k/n & \text{si } x_k \leq t < x_{k+1} \\ 1 & \text{si } t \geq x_n \end{cases}$$

*Démonstration.* Cf cours. □

**Définition II.6.4** (Loi de Bernoulli de paramètre  $p : \mathcal{B}(p)$ ). Soit  $p \in [0, 1]$ . Une v.a.r.  $X$  qui suit la loi de Bernoulli de paramètre  $p$  est une v.a.r. discrète telle que  $X(\Omega) = \{0, 1\}$ ,  $\mathbb{P}(X = 1) = p$  et  $\mathbb{P}(X = 0) = 1 - p$ .

**Exemple II.6.5.** Les lois de Bernoulli peuvent être utilisées pour modéliser un lancer de pièce (qui peut être équilibrée si  $p = 1/2$ , ou non-équilibrée sinon). Plus généralement, elles peuvent être utilisées pour modéliser n'importe quelle expérience qui n'a que deux issues (que l'on appelle souvent "réussite" et "échec") : on appelle ce type d'expérience une épreuve de Bernoulli.

**Définition II.6.6** (Loi binomiale de paramètres  $(n, p) : \mathcal{B}(n, p)$ ). Soient  $n \in \mathbb{N}^*$  et  $p \in [0, 1]$ . Une v.a.r.  $X$  qui suit la loi binomiale de paramètres  $(n, p)$  est une v.a.r. discrète telle que  $X(\Omega) = \llbracket 0, n \rrbracket$  et pour tout  $0 \leq k \leq n$ ,

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

À première vue, les lois binomiales peuvent paraître compliquées. En fait, elles permettent naturellement de compter le nombre de réussites quand on répète des expériences de Bernoulli indépendantes et de même paramètre. Cet énoncé est formel dans la proposition ci-dessous. En particulier, si vous voulez utiliser ce résultat, vous avez besoin de la proposition formelle suivante (plutôt que donner un énoncé vague).

**Proposition II.6.7.** Soient  $p \in [0, 1]$  et  $X_1, \dots, X_n$  des v.a.r. i.i.d. (c'est-à-dire indépendantes et identiquement distribuées) qui suivent la loi de Bernoulli de paramètre  $p$ . Alors  $\sum_{k=1}^n X_k$  est une v.a.r. binomiale de paramètres  $(n, p)$ .

*Démonstration.* Cf cours. □

**Exemple II.6.8.** On lance un dé équilibré à six faces cinq fois d'affilée. Alors le nombre de six obtenus suit une loi binomiale de paramètres  $(5, 1/6)$  (pourquoi ?).

**Définition II.6.9** (Loi géométrique de paramètre  $p : \mathcal{G}(p)$ ). Soit  $p \in ]0, 1]$ . Une v.a.r.  $X$  qui suit la loi géométrique de paramètre  $p$  est une v.a.r. discrète telle que  $X(\Omega) = \mathbb{N}^*$  et pour tout  $k \geq 1$ ,

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

Les lois géométriques permettent de modéliser la première réussite lorsque l'on répète des épreuves de Bernoulli indépendantes et de même paramètres (voir Lemme ci-dessous).

**Lemme II.6.10.** Soient  $p \in ]0, 1]$  et  $X_1, \dots, X_n, \dots$  des v.a.r. i.i.d. qui suivent la loi de Bernoulli de paramètre  $p$ . Alors la variable

$$\min \{k \in \mathbb{N}^* : X_k = 1\}$$

suit la loi géométrique de paramètre  $p$ .

*Démonstration.* Cf cours. □

**Exemple II.6.11.** On lance un dé équilibré à six faces une infinité de fois. Le premier lancé où on obtient six suit une loi géométrique de paramètre  $1/6$ .

**Lemme II.6.12.** Soit  $X$  une v.a.r. géométrique de paramètre  $p \in ]0, 1]$ . Alors la fonction de répartition de  $X$  vérifie

$$F_X(t) = \begin{cases} 0 & \text{si } t < 1 \\ 1 - (1 - p)^{\lfloor t \rfloor} & \text{sinon} \end{cases}$$

**Définition II.6.13** (Loi de Poisson de paramètre  $\lambda : \mathcal{P}(\lambda)$ ). Soit  $\lambda > 0$ . Une v.a.r.  $X$  qui suit la loi de Poisson de paramètre  $\lambda$  est une v.a.r. discrète telle que  $X(\Omega) = \mathbb{N}$  et pour tout  $k \geq 0$ ,

$$\mathbb{P}(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}.$$

On introduit souvent les lois de Poisson comme des lois d'événements rares : elles permettent de modéliser le nombre de réussites quand il y a beaucoup d'épreuves ( $\approx \infty$ ) mais que la probabilité de réussite est très faible ( $\approx 0$ ). D'une certaine manière, ces lois sont proches des lois binomiales (pour lesquelles le nombre d'épreuve est fini  $n$  et dont la probabilité de réussite est fixé  $p$  et non-nulle dans les cas non-triviaux). Ce lien est formalisé dans la proposition suivante.

**Proposition II.6.14.** Soit  $\lambda > 0$ , et pour tout  $n \in \mathbb{N}^*$ , soit  $X_n$  une v.a.r. binomiale de paramètres  $(n, \lambda/n)$ . Alors, pour tout  $k \in \mathbb{N}$ ,

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

*Démonstration.* Cf cours. □

**Remarque II.6.15.** La proposition précédente implique que, lorsque  $n$  est grand, la loi binomiale de paramètres  $(n, \lambda/n)$  est une approximation de la loi de Poisson de paramètre  $\lambda$ .

**Exemple II.6.16.** Un exemple classique d'utilisation des lois de Poisson est la modélisation du nombre d'appels reçus par un standard téléphonique pendant un intervalle de temps donné. Le paramètre  $\lambda$  représente le nombre d'appels reçus en moyenne dans cet intervalle de temps.

**Proposition II.6.17.** Soient  $X, Y$  deux v.a.r. indépendantes de Poisson de paramètres respectifs  $\lambda_X$  et  $\lambda_Y$ . Alors  $X + Y$  suit une loi de Poisson de paramètre  $\lambda_X + \lambda_Y$ .

*Démonstration.* Cf cours. □

Les informations récapitulées dans le tableau ci-dessous sont à connaître (ou encore mieux, à savoir retrouver pour les deux dernières lignes).

$\mathbb{P}_X$	$\mathcal{U}(\llbracket 1, n \rrbracket)$	$\mathcal{B}(p)$	$\mathcal{B}(n, p)$	$\mathcal{G}(p)$	$\mathcal{P}(\lambda)$
$X(\Omega)$	$\llbracket 1, n \rrbracket$	$\{0, 1\}$	$\llbracket 0, n \rrbracket$	$\mathbb{N}^*$	$\mathbb{N}$
$\mathbb{P}(X = k)$ (si $k \in X(\Omega)$ )	$1/n$	$p$ si $k = 1$ $1 - p$ si $k = 0$	$\binom{n}{k} p^k (1 - p)^{n-k}$	$p(1 - p)^{k-1}$	$e^{-\lambda} \lambda^k / k!$
$\mathbb{E}[X]$	$(n + 1)/2$	$p$	$np$	$1/p$	$\lambda$
$\mathbb{V}[X]$	$(n^2 - 1)/12$	$p(1 - p)$	$np(1 - p)$	$(1 - p)/p^2$	$\lambda$

## II.7 Lois usuelles continues

**Définition II.7.1** (Loi uniforme sur un intervalle borné  $I : \mathcal{U}(I)$ ). Soit  $I$  un intervalle bornée. Une v.a.r.  $X$  qui suit la loi uniforme sur  $I$  est une v.a.r. continue dont la densité est définie par

$$f_X(x) = \frac{1}{\text{longueur}(I)} 1_{\{x \in I\}} := \begin{cases} \frac{1}{\text{longueur}(I)} & \text{si } x \in I \\ 0 & \text{sinon} \end{cases}$$

Les lois uniformes peuvent être utilisées pour modéliser la valeur d'un réel aléatoire où la seule information que l'on possède est qu'il est dans un intervalle borné.

**Exemple II.7.2.** Quelqu'un vous a prévenu qu'il vous appellerait entre 12h et 13h sans précision supplémentaire. Alors le temps d'attente de l'appel (à partir de 12h) peut être modélisé par une variable uniforme sur  $[0, 1]$  (en comptant le temps en heures).

**Lemme II.7.3.** Soit  $X$  une v.a.r. uniforme sur un intervalle  $[a, b]$  (avec  $a < b$ ). Alors la fonction de répartition de  $X$  vérifie

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

*Démonstration.* Cf cours. □

**Lemme II.7.4.** Soit  $X$  une v.a.r. de loi  $\mathcal{U}(]a, b])$  (avec  $a < b$  deux réels). Soient  $\alpha \in \mathbb{R}$  et  $\lambda > 0$ . Alors

$$\lambda X + \alpha \sim \mathcal{U}(] \lambda a + \alpha, \lambda b + \alpha ]).$$

*Démonstration.* Cf cours. □

**Remarque II.7.5.** Une conséquence intéressante du lemme précédent est la suivante : si  $X \sim \mathcal{U}(]a, b])$ , alors

$$\frac{X - a}{b - a} \sim \mathcal{U}(]0, 1]).$$

**Définition II.7.6** (Loi exponentielle de paramètre  $\lambda : \mathcal{E}(\lambda)$ ). Soit  $\lambda > 0$ . Une v.a.r.  $X$  qui suit la loi exponentielle de paramètre  $\lambda$  est une v.a.r. continue dont la densité est définie par

$$f_X(x) = \lambda e^{-\lambda x} 1_{\{x > 0\}} := \begin{cases} \lambda e^{-\lambda x} & \text{si } x > 0 \\ 0 & \text{sinon} \end{cases}$$

Les lois exponentielles permettent de modéliser des temps d'attentes pour qu'un événement aléatoire se produise. Informellement, cette interprétation vient du fait que si on répartit le plus uniformément possible un nuage dénombrable de points sur  $\mathbb{R}_+$  (la formalisation de ces différents termes est très largement hors-programme), et que l'on interprète ces points comme des instants où des événements se produisent, alors les durées entre les événements consécutifs sont des variables exponentielles i.i.d. de paramètre 1. Une manière (plus faible) de formuler cet énoncé est la suivante (même si la preuve est techniquement faisable avec nos outils, nous ne la ferons pas). Ce résultat n'est écrit qu'à titre purement indicatif, il ne sera pas utilisé (et ne sera pas utile pour valider ce cours).

**Théorème II.7.7.** Soient  $N_k$  ( $k \in \mathbb{N}$ ) une suite de v.a.r. de Poisson i.i.d. de paramètre  $\lambda > 0$ . Pour chaque  $k \in \mathbb{N}$ , soient  $X_1^k, \dots, X_{N_k}^k$  des v.a.r. i.i.d. uniformes sur  $[k, k + 1[$ . Définissons

$$K := \min\{k \in \mathbb{N} : N_k \geq 1\}$$

et

$$T := \min\{X_j^K : 1 \leq j \leq N_K\}.$$

Alors  $T$  est une v.a.r. exponentielle de paramètre  $\lambda$  (les  $X_j^k$  sont les instants des événements qui ont lieu dans l'intervalle de temps  $[k, k+1]$ , et  $T$  est donc le temps d'attente du tout premier événement).

**Lemme II.7.8.** Soit  $X$  une v.a.r. exponentielle de paramètre  $\lambda > 0$ . Alors la fonction de répartition de  $X$  vérifie

$$F_X(t) = \begin{cases} 0 & \text{si } t \leq 0 \\ 1 - e^{-\lambda t} & \text{si } t \geq 0 \end{cases}$$

**Proposition II.7.9.** Soient  $X, Y$  deux v.a.r. exponentielles indépendantes de paramètres respectifs  $\lambda_X$  et  $\lambda_Y$ , alors  $\min(X, Y)$  est une v.a.r. exponentielle de paramètre  $\lambda_X + \lambda_Y$ .

*Démonstration.* Cf cours. □

**Proposition II.7.10.** Soient  $X$  une v.a.r. de loi  $\mathcal{E}(\lambda)$ . Alors  $\lambda X$  suit la loi  $\mathcal{E}(1)$ .

*Démonstration.* Cf cours. □

**Définition II.7.11** (Loi gaussienne de paramètres  $(\mu, \sigma^2) : \mathcal{N}(\mu, \sigma^2)$ ). Soient  $\mu$  un réel et  $\sigma$  un réel strictement positif. Une v.a.r.  $X$  qui suit la loi gaussienne (ou loi normale) de paramètres  $(\mu, \sigma^2)$  est une v.a.r. continue dont la densité est définie par

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

Il n'y a, à ma connaissance, pas d'interprétation directe et naturelle des lois gaussiennes. Toutefois, elles sont fondamentales en probabilités en statistiques, car, d'une certaine manière, elles permettent d'approcher n'importe quelle loi avec des moments d'ordre deux finis (même les lois discrètes). Nous donnerons un sens à cet énoncé dans le Théorème III.3.1 (i.e. le Théorème Central Limite).

**Proposition II.7.12.** Soient  $X, Y$  deux gaussiennes indépendantes de paramètres respectifs  $(\mu_X, \sigma_X^2)$  et  $(\mu_Y, \sigma_Y^2)$ . Alors :

- $X + Y$  est une v.a.r. gaussienne de paramètres  $(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$ ,
- pour tout  $\alpha \in \mathbb{R}^*, \beta \in \mathbb{R}$ , alors  $\alpha X + \beta$  est une v.a.r. gaussienne de paramètres  $(\alpha\mu_X + \beta, \alpha^2\sigma_X^2)$ .

**Remarque II.7.13.** Une conséquence importante de la proposition précédente est la suivante : si  $X$  est une v.a.r. gaussienne de paramètres  $(\mu, \sigma^2)$ , alors

$$Y := \frac{X - \mu}{\sigma}$$

est une v.a.r. gaussienne de paramètres  $(0, 1)$ . On dit que  $Y$  est une gaussienne centrée (i.e. d'espérance nulle) et réduite (i.e. de variance égale à un).

De même que pour les lois discrètes, les informations récapitulées dans le tableau ci-dessous sont à connaître.

$\mathbb{P}_X$	$\mathcal{U}([a, b])$	$\mathcal{E}(\lambda)$	$\mathcal{N}(\mu, \sigma^2)$
$X(\Omega)$	$[a, b]$	$\mathbb{R}_+^*$	$\mathbb{R}$
$f_X(x)$	$\frac{1}{b-a} 1_{\{a \leq x \leq b\}}$	$\lambda e^{-\lambda x} 1_{\{x > 0\}}$	$\frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]$
$\mathbb{E}[X]$	$(b-a)/2$	$1/\lambda$	$\mu$
$\mathbb{V}[X]$	$(b-a)^2/12$	$1/\lambda^2$	$\sigma^2$

# Chapitre III

## Approximations asymptotiques

### III.1 Quelques inégalités probabilistes

**Proposition III.1.1** (Inégalité de Markov). *Soient  $X$  une v.a.r. positive et  $\alpha > 0$ . Alors*

$$\mathbb{P}(X \geq \alpha) \leq \frac{\mathbb{E}[X]}{\alpha}.$$

*Démonstration.* Cf cours. □

**Proposition III.1.2** (Inégalité de Bienaymé-Tchebychev). *Soient  $X$  une v.a.r. de carré intégrable et  $\alpha > 0$ . Alors*

$$\mathbb{P}(|X - \mathbb{E}[X]| \geq \alpha) \leq \frac{\mathbb{V}[X]}{\alpha^2}.$$

*Démonstration.* Cf cours. □

### III.2 Loi (faible) des Grands Nombres

**Théorème III.2.1** (Loi faible des grands nombres (cas  $L^2$ )). *Soient  $X_1, \dots, X_n, \dots$  une suite de v.a.r. i.i.d. de carré intégrables, dont on note  $m$  l'espérance et  $\sigma^2$  la variance. Alors, pour tout  $\varepsilon > 0$ ,*

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

*Démonstration.* Cf cours. □

**Remarque III.2.2.** *La formulation du théorème ci-dessus est classique, mais peut se généraliser facilement de deux manières :*

- les v.a.r.  $X_k$  peuvent être seulement non-corrélées deux-à-deux plutôt que mutuellement indépendantes,
- il suffit que les v.a.r.  $X_k$  aient la même espérance et la même variance, mais il n'est pas nécessaire qu'elles aient la même loi.

**Remarque III.2.3.** *Dans l'énoncé du Théorème III.2.1, on dit que la suite de v.a.r.  $\frac{1}{n} \sum_{k=1}^n X_k$  converge en probabilités vers  $m$ .*

### III.3 Théorème Central Limite

**Théorème III.3.1** (Théorème Central Limite). *Soient  $X_1, \dots, X_n, \dots$  une suite de v.a.r. i.i.d. de carré intégrables, dont on note  $m$  l'espérance et  $\sigma^2$  la variance. Alors, pour tout intervalle  $I$ ,*

$$\mathbb{P} \left( \frac{\sqrt{n}}{\sigma} \left( \frac{1}{n} \sum_{k=1}^n X_k - m \right) \in I \right) \xrightarrow{n \rightarrow \infty} \mathbb{P}(Y \in I),$$

où  $Y$  est une v.a.r. de loi  $\mathcal{N}(0, 1)$ .

*Démonstration.* Admis. □

Deuxième partie

Statistiques

# Chapitre IV

## Modèles statistiques et estimateurs

Ce chapitre est consacré à l'étude des modèles statistiques. Dans ce cours, nous nous restreindrons à l'étude de l'estimation paramétrique. Pour comprendre le principe de base, il faut imaginer que l'on peut répéter une expérience aléatoire plusieurs fois (disons  $n \in \mathbb{N}^*$ ), et que l'on note  $X_k$  le résultat de la  $k$ -ième expérience ( $1 \leq k \leq n$ ). Dans le contexte de l'estimation paramétrique, la loi des variables  $X_k$  n'est pas explicite, elle dépend d'un paramètre inconnu  $\theta$  (qui sera souvent un réel pour simplifier, mais qui peut aussi être un élément de  $\mathbb{R}^d$  dans le cadre de l'estimation paramétrique). Le but sera de déterminer une valeur approchée  $\theta$  (qui est inconnue mais déterministe), en utilisant l'information que l'on a obtenue : i.e. les valeurs des v.a.r.  $X_1, \dots, X_n$  (qui sont aléatoires mais connues). C'est cette valeur approchée que nous appellerons estimateur.

**Exemple IV.0.1.** *On a une pièce dont on ne sait pas si elle est équilibrée. Dans ce cas, si on lance  $n$  fois la pièce et que l'on définit  $X_k = 1$  si le  $k$ -ième lancé fait pile,  $X_k = 0$  s'il fait face, on sait que les  $X_k$  sont des v.a.r. i.i.d. qui suivent une loi de Bernoulli dont le paramètre est une inconnue que l'on peut noter  $\theta \in [0, 1]$  (où  $\theta$  est donc la probabilité d'obtenir pile). Nous nous intéressons donc aux questions suivantes : est-il possible de donner une valeur approchée de  $\theta$  qui ne dépend que des v.a.r.  $X_1, \dots, X_n$  ? Et en quel sens cette approximation est correcte ?*

### IV.1 Modèles statistiques

**Définition IV.1.1.** *Soient  $X_1, \dots, X_n$  des v.a.r. i.i.d., notons  $\mathcal{X}$  le domaine des  $X_k$ ,  $T = (X_1, \dots, X_n)$  et  $P_\theta$  la loi des  $X_k$ , où  $\theta$  est inconnu, mais appartient à un ensemble  $\Theta$  qui lui est connu. Alors  $(\mathcal{X}, \Theta, (P_\theta)_{\theta \in \Theta})$  est le modèle statistique associé à  $T$ .*

Il faut bien comprendre que la seule inconnue du modèle statistique est  $\theta$  (et donc  $P_\theta$  aussi). En particulier, si vous vous donner un certain  $\theta_0 \in \Theta$ , la mesure de probabilité  $P_{\theta_0}$  est une loi connue.

**Exemple IV.1.2.** *Reprenons l'Exemple IV.0.1 ci-dessus et les notations de la Définition IV.1.1. Dans ce cas,  $\mathcal{X} = \{0, 1\}$ ,  $\Theta = [0, 1]$  et  $P_\theta = \mathcal{B}(\theta)$ .*

Attention aux notations et à la différence entre  $\mathbb{P}$  et  $P_\theta$  : par exemple, si  $I$  est un intervalle, alors, avec les notations ci-dessus,

$$\mathbb{P}(X_1 \in I) = P_\theta(I).$$

**Remarque IV.1.3.** *Dans ce cours, nous n'étudierons que les cas où les v.a.r.  $X_k$  sont i.i.d. (comme ce qui est supposé dans la définition ci-dessus). Dans ce cas on dit que le modèle statistique associé est un modèle d'échantillonnage. De plus, le modèle est dit paramétrique quand  $\Theta$  est une partie de  $\mathbb{R}^d$  (où  $d \in \mathbb{N}^*$ ). Dans la suite, pour simplifier, nous étudierons principalement le cas où  $\Theta \subseteq \mathbb{R}$ .*

**Définition IV.1.4.** Soit  $(\mathcal{X}, \Theta, (P_\theta)_{\theta \in \Theta})$  un modèle statistique associé à  $T = (X_1, \dots, X_n)$ . Alors :

- une statistique est une v.a.r.  $Y$  qui s'écrit comme une fonction de  $T : Y = g(X_1, \dots, X_n)$ ,
- un estimateur est une statistique à valeurs dans  $\Theta$ .

Dans la suite, nous verrons différents exemples d'estimateurs. Un estimateur que nous verrons souvent est la moyenne empirique du modèle statistique :

$$Y = \frac{1}{n} \sum_{k=1}^n X_k.$$

Pour le moment, nous avons seulement défini la notion d'estimateur, mais nous n'avons pas décrit ce qu'est un bon estimateur. En effet la définition précédente d'un estimateur ne prend pas en compte le paramètre  $\theta$  que l'on cherche à estimer. Le but de la section suivante est de donner quelques éléments pour savoir si un estimateur est "bon" (dans un certain sens).

## IV.2 Qualité des estimateurs

**Définition IV.2.1** (Biais). Soit  $\hat{\theta}$  un estimateur pour un paramètre  $\theta$  que l'on cherche à estimer. Le biais d'estimation pour  $\hat{\theta}$  par rapport à  $\theta$  est

$$\mathbb{B}_{\hat{\theta}}(\theta) := \mathbb{E} \left[ \hat{\theta} - \theta \right] = \mathbb{E} \left[ \hat{\theta} \right] - \theta.$$

On dit que l'estimateur est sans biais si le biais est nul, et qu'il est biaisé sinon.

Soit  $(\hat{\theta}_n)_n$  une suite d'estimateur pour  $\theta$ . On dit que la suite  $(\hat{\theta}_n)_n$  est asymptotiquement sans biais si

$$\mathbb{B}_{\hat{\theta}_n}(\theta) \xrightarrow[n \rightarrow \infty]{} 0.$$

**Définition IV.2.2** (Risque quadratique). Soit  $\hat{\theta}$  un estimateur pour un paramètre  $\theta$  que l'on cherche à estimer. Le risque quadratique pour  $\hat{\theta}$  par rapport à  $\theta$  est

$$\mathbb{R}_{\hat{\theta}}(\theta) = \mathbb{E} \left[ \left( \hat{\theta} - \theta \right)^2 \right].$$

**Définition IV.2.3.** Étant donné deux estimateurs  $\hat{\theta}_1$  et  $\hat{\theta}_2$  d'un même paramètre, on dit que  $\hat{\theta}_1$  est meilleur que  $\hat{\theta}_2$  en terme de risque quadratique si

$$\mathbb{R}_{\hat{\theta}_1}(\theta) \leq \mathbb{R}_{\hat{\theta}_2}(\theta).$$

**Lemme IV.2.4.**

$$\mathbb{R}_{\hat{\theta}}(\theta) = \mathbb{V} \left[ \hat{\theta} \right] + \left( \mathbb{B}_{\hat{\theta}}(\theta) \right)^2.$$

Il existe des conditions générales sur les modèles statistiques  $(T, \Theta, (\mathbb{P}_\theta)_{\theta \in \Theta})$  telles que, si ces conditions sont vérifiées, le risque quadratique de n'importe quel estimateur sans biais est supérieure à une borne explicite. Il s'agit de la borne de Cramer-Rao et des conditions de Cramer-Rao.

**Définition IV.2.5** (Consistance). On dit qu'une suite d'estimeurs  $(\hat{\theta}_n)_n$  est consistante (ou que les estimateurs sont consistants) pour  $\theta$  si la suite converge en probabilités vers  $\theta$ . C'est-à-dire si : pour tout  $\varepsilon > 0$ ,

$$\mathbb{P} \left( \left| \hat{\theta}_n - \theta \right| > \varepsilon \right) \xrightarrow[n \rightarrow \infty]{} 0.$$

## IV.3 Estimateurs empiriques

Dans cette section, nous allons voir quelques estimateurs "naturels" en fonction du paramètre que l'on cherche à estimer. Dans toute cette section, nous supposons que le modèle statistique que nous étudions  $(T, \Theta, (\mathbb{P}_\theta)_{\theta \in \Theta})$  est un modèle d'échantillonnage : on écrit  $T = (X_1, \dots, X_n)$  où les  $X_k$  ( $1 \leq k \leq n$ ) sont des v.a.r. i.i.d.

**Définition IV.3.1** (Moyenne empirique). *La moyenne empirique de  $n$  v.a.r.  $X_1, \dots, X_n$  est*

$$\bar{X}_n := \frac{1}{n} \sum_{k=1}^n X_k.$$

La moyenne empirique est un estimateur naturel si le paramètre  $\theta$  à estimer est l'espérance de la loi  $\mathbb{P}_\theta$  (i.e.  $\theta = \mathbb{E}[X_k]$ ). Si  $\theta = \mathbb{E}[g(X_k)]$  (où  $g$  est une fonction à valeurs réelles continue en  $\theta$ ), c'est la moyenne empirique des v.a.r.  $Y_k := g(X_k)$  qui est naturelle.

### Exemples IV.3.2.

- Si  $X_1, \dots, X_n$  sont des v.a.r. de loi  $\mathcal{P}(\theta)$ , alors  $\bar{X}_n := \frac{1}{n} \sum_{k=1}^n X_k$  est un estimateur naturel pour  $\theta$ .
- Si  $X_1, \dots, X_n$  sont des v.a.r. de loi  $\mathcal{E}(\sqrt{2/\theta})$ , alors  $\frac{1}{n} \sum_{k=1}^n X_k^2$  est un estimateur naturel pour  $\theta$ .

**Lemme IV.3.3.** *Soient  $X_1, \dots, X_n, \dots$  des v.a.r. Alors :*

- si les v.a.r.  $X_k$  ont la même espérance  $\mathbb{E}[X_k] = \theta$ , alors, pour tout  $n$ ,  $\bar{X}_n$  est un estimateur sans biais pour  $\theta$ ,
- si de plus les v.a.r.  $X_k$  sont deux-à-deux non-corrélées, alors  $\bar{X}_n$  est consistant pour  $\theta$ .

*Démonstration.* Cf cours. □

**Définition IV.3.4** (Variance empirique). *La variance empirique de  $n$  v.a.r.  $X_1, \dots, X_n$  d'espérance  $m$  connue est*

$$\bar{\sigma}_n^2 := \frac{1}{n} \sum_{k=1}^n (X_k - m)^2.$$

La variance empirique est un estimateur naturel si le paramètre  $\theta$  à estimer est l'espérance de la loi  $\mathbb{P}_\theta$  (i.e.  $\theta = \mathbb{V}[X_k]$ ), mais n'est utilisable que si l'espérance de la loi est connue.

**Exemple IV.3.5.** *Si  $X_1, \dots, X_n$  sont des v.a.r. de loi  $\mathcal{N}(0, \theta)$ , alors  $\bar{\sigma}_n^2$  est un estimateur naturel pour  $\theta$ .*

**Lemme IV.3.6.** *Soient  $X_1, \dots, X_n, \dots$  des v.a.r. Alors :*

- si les v.a.r.  $X_k$  ont la même espérance et la même variance, et si cette variance est  $\theta$ , alors, pour tout  $n$ ,  $\bar{\sigma}_n^2$  est un estimateur sans biais pour  $\theta$ ,
- si de plus les v.a.r.  $X_k$  sont deux-à-deux non-corrélées, alors  $\bar{\sigma}_n^2$  est consistant pour  $\theta$ .

Si l'espérance est inconnue et que le paramètre à estimer est la variance, il est possible d'utiliser la variance empirique modifiée (qui consiste simplement à remplacer l'espérance inconnue par la moyenne empirique de l'échantillon dans la définition de la variance empirique) :

**Définition IV.3.7** (Variance empirique modifiée). *La variance empirique modifiée de  $n$  v.a.r.  $X_1, \dots, X_n$  d'espérance  $m$  connue est*

$$\hat{\sigma}_n^2 := \frac{1}{n-1} \sum_{k=1}^n (X_k - \bar{X}_n)^2.$$

**Remarque IV.3.8.** *Le  $n - 1$  dans la définition de la variance empirique modifiée n'est pas une erreur. Cela permet d'avoir un estimateur sans biais de la variance.*

**Exemples IV.3.9.**

- Si  $X_1, \dots, X_n$  sont des v.a.r. de loi  $\mathcal{P}(\theta)$ , alors  $\hat{\sigma}_n^2$  est un estimateur naturel pour  $\theta$  (tout comme  $\bar{X}_n$ ).
- Si  $X_1, \dots, X_n$  sont des v.a.r. de loi  $\mathcal{E}(1/\sqrt{\theta})$ , alors  $\hat{\sigma}_n^2$  est un estimateur naturel pour  $\theta$ .
- Si  $X_1, \dots, X_n$  sont des v.a.r. de loi  $\mathcal{N}(\mu, \theta)$  (où  $\mu$  et  $\theta$  sont inconnus), alors  $\hat{\sigma}_n^2$  est un estimateur naturel pour  $\theta$  (et  $\bar{\sigma}_n^2$  n'est pas utilisable, car il dépend de  $\mu$  qui est inconnu).

**Lemme IV.3.10.** *Soient  $X_1, \dots, X_n, \dots$  des v.a.r. i.i.d de variance  $\theta$ . Alors :*

- $\hat{\sigma}_n^2$  est un estimateur sans biais pour  $\theta$ ,
- $\bar{\sigma}_n^2$  est consistant pour  $\theta$ .

**Définition IV.3.11** (maximum et minimum empiriques). *Le maximum empirique de  $n$  v.a.r.  $X_1, \dots, X_n$  est*

$$M_n = \max_{1 \leq k \leq n} X_k.$$

*Le minimum empirique de  $n$  v.a.r.  $X_1, \dots, X_n$  est*

$$I_n = \min_{1 \leq k \leq n} X_k.$$

La maximum (resp. minimum) empirique est un estimateur naturel si le paramètre  $\theta$  à estimer est la borne supérieure (resp. inférieure) de  $X_1(\Omega)$ .

**Exemple IV.3.12.** *Si  $X_1, \dots, X_n$  sont des v.a.r. de loi  $\mathcal{U}([0, \theta])$ , alors  $M_n$  est un estimateur naturel pour  $\theta$ .*

**Proposition IV.3.13.** *Soient  $X_1, \dots, X_n$  des v.a.r. i.i.d. Alors les fonctions de répartition de  $M_n$  et  $I_n$  vérifient : pour tout  $x \in \mathbb{R}$ ,*

$$\begin{aligned} F_{M_n}(x) &= (F_{X_1}(x))^n, \\ F_{I_n}(x) &= 1 - (1 - F_{X_1}(x))^n. \end{aligned}$$

*Démonstration.* Cf cours. □

## IV.4 Méthode des moments

La méthode des moments permet de choisir un estimateur naturel pour un modèle statistique donnée. Le principe est le suivant : si le paramètre  $\theta$  n'est ni l'espérance, ni la variance, ni un moment, ni une borne supérieure, ni une borne inférieure de la loi  $\mathbb{P}_\theta$ , mais que c'est une fonction d'une ou plusieurs des quantités précédents, alors un estimateur naturel est la même fonctions des quantités empiriques correspondantes.

Un point important de cette méthode est que qu'elle conserve la propriété de consistance dans le sens suivant :

**Théorème IV.4.1.** *Soient  $(\hat{\pi}_n)_n$  et  $(\hat{\rho}_n)_n$  des suites de v.a.r. qui convergent en probabilités vers les réels respectifs  $\pi, \rho$  des réels tels que : pour tout  $\varepsilon > 0$ . Alors :*

1. *pour toute fonction  $g$  continue en  $\pi$ ,  $g(\hat{\pi}_n)$  converge en probabilités vers  $g(\pi)$ ,*
2.  *$\pi_n + \rho_n$  convergen en probabilité vers  $\pi + \rho$ ,*
3. *pour tout suite de réel  $\alpha_n$  convergeant vers  $\alpha \in \mathbb{R}$ , les v.a.r.  $\alpha_n \pi_n$  convergent en probabilités vers  $\alpha \pi$ ,*
4. *si  $\rho \neq 0$ , alors  $\pi_n / \rho_n$  converge en probabilités vers  $\pi / \rho$ .*

*Démonstration.* Cf cours. □

Les exemples suivants sont une application directe du premier point du Théorème ci-dessus. Nous utiliserons les autres points pour construire des intervalles de confiance dans le chapitre suivant.

**Exemples IV.4.2.**

- Soient  $X_1, \dots, X_n$  des v.a.r. de loi  $\mathcal{G}(\theta)$  (où  $\theta \in ]0, 1[$ ), alors  $1/\bar{X}_n$  est un estimateur naturel pour  $\theta$ .
- Soient  $X_1, \dots, X_n$  des v.a.r. de loi  $\mathcal{E}(\sqrt{\theta})$  (où  $\theta > 0$ ), alors  $1/\bar{X}_n^2$  et  $1/\hat{\sigma}_n^2$  sont des estimateurs naturels pour  $\theta$ .

# Chapitre V

## Estimation par intervalles de confiance

### V.1 Intervalle de confiance exact

**Définition V.1.1.** Un intervalle de confiance exact de niveau  $1 - \varepsilon$  pour  $\theta$  (où  $0 < \varepsilon < 1$ ) est un intervalle aléatoire  $I$  vérifiant

$$\mathbb{P}(\theta \in I) \geq 1 - \varepsilon.$$

**Remarque V.1.2.** Pour qu'un intervalle de confiance soit intéressant, il faut que  $\varepsilon$  soit "petit" (pour que l'approximation soit vraie avec probabilité élevée) et que la longueur de l'intervalle soit "petite" aussi (pour que l'approximation soit précise).

**Remarque V.1.3.** Une utilisation "naturelle" des intervalles de confiance dans un cas pratique est la suivante : vous cherchez la valeur d'un paramètre inconnu  $\theta$ . Pour faire vos calculs en pratique, vous n'avez besoin que d'une valeur numérique approchée de  $\theta$  et vous voulez que la probabilité que votre approximation soit vraie soit au moins 0.99. Dans ce cas, vous aurez besoin d'un intervalle de niveau 0.99. Si par exemple l'intervalle de confiance suivant est bien de niveau 0.99

$$I := ]\hat{\theta} - 0.001; \hat{\theta} + 0.001[,$$

alors vous saurez que l'estimateur  $\hat{\theta}$  est une valeur approchée de  $\theta$  à moins de 0.001 avec probabilité supérieure (ou égale) à 0.99.

Dans la suite, nous nous intéresserons à deux types d'intervalles de confiance.

**Définition V.1.4.** Un intervalle de confiance  $]U, V[$  est dit symétrique si

$$\mathbb{P}(\theta \leq U) = \mathbb{P}(\theta \geq V).$$

Dans ce cas, si on note  $\varepsilon = \mathbb{P}(\theta \notin ]U, V[)$ , alors les deux probabilités ci-dessus sont égales à  $\varepsilon/2$ .

**Définition V.1.5.** Un intervalle de confiance  $]U, V[$  est dit unilatéral si

$$\mathbb{P}(\theta \leq U) = 0 \text{ ou } \mathbb{P}(\theta \geq V) = 0.$$

**Remarque V.1.6.** Les intervalles de confiance unilatéraux sont intéressants si vous voulez seulement estimer la probabilité que votre paramètre  $\theta$  ne dépasse une certaine borne. Par exemple, si vous voulez prévoir un coût  $\theta$ , vous aurez envie de contrôler la probabilité que le coût ne soit pas trop cher (i.e. que  $\theta$  ne soit pas plus grand qu'une borne supérieure aléatoire connue) sans nécessairement contrôler la probabilité que le coût soit trop faible.

Il existe différentes méthodes pour construire des intervalles de confiance. Pour l'un d'entre elles, nous aurons besoin des fonctions pivotables.

**Définition V.1.7.** Soit  $(\mathcal{X}, \Theta, (P_\theta)_{\theta \in \Theta})$  un modèle statistique associé à  $T$ . Une fonction pivotale est une v.a.r. qui dépend de  $T$  et de  $\theta$  dont la loi est connue (en particulier la loi ne doit pas dépendre de  $\theta$ ).

**Exemple V.1.8.** Soient  $X_1, \dots, X_n$  des v.a.r. i.i.d. de loi  $\mathcal{N}(\theta, 1)$ . Alors

$$Y := \sqrt{n}(\bar{X}_n - \theta) \sim \mathcal{N}(0, 1)$$

est une fonction pivotale.

Pour construire un intervalle de confiance exact de niveau  $1 - \varepsilon$  donné à partir d'une fonction pivotale, il faut exploiter sa loi (qui est connue). À la place de donner une explication théorique et vague de la méthode, illustrons la méthode sur deux exemples. Avant ça, énonçons le lemme suivant qui permet relativement simplement de construire un intervalle de confiance, mais seulement dans certains cas.

**Lemme V.1.9.** Soit  $\hat{\theta}$  un estimateur pour  $\theta$  dont on suppose que le risque quadratique  $\mathbb{R}_{\hat{\theta}}(\theta)$  est connu (notons le  $r$ ). Alors, pour tout  $\varepsilon > 0$ , l'intervalle

$$I := ]\hat{\theta} - \varepsilon, \hat{\theta} + \varepsilon[$$

est un intervalle de confiance exact pour  $\theta$  de niveau  $1 - r/\varepsilon^2$ .

*Démonstration.* Cf cours □

**Remarque V.1.10.** Techniquement, le lemme précédent reste vrai si le risque quadratique est inconnu (ce qui est souvent le cas), mais le résultat est inutile en pratique. En effet, dans ce cas, le niveau de l'intervalle de confiance est inconnu.

## Construction d'un intervalle de confiance symétrique via une fonction pivotale

Reprenons l'Exemple V.1.8 ci-dessus en utilisant la fonction pivotale que l'on a explicitée. Pour construire un intervalle de confiance symétrique de niveau  $1 - \varepsilon$  dans ce cadre, il faut d'abord trouver deux réels  $c_\varepsilon$  et  $d_\varepsilon$  tels que :

$$\mathbb{P}(Y \leq c_\varepsilon) = \mathbb{P}(Y \geq d_\varepsilon) = \varepsilon/2.$$

En particulier

$$F_Y(c_\varepsilon) = \varepsilon/2 \text{ et } F_Y(d_\varepsilon) = 1 - \varepsilon/2.$$

Ce qui donne

$$c_\varepsilon = F_Y^{-1}(\varepsilon/2) \text{ et } d_\varepsilon = F_Y^{-1}(1 - \varepsilon/2). \quad (\text{V.1})$$

Rappelons que la loi de  $Y$  étant connue, sa fonction de répartition  $F_Y$  l'est aussi. En particulier, si  $\varepsilon$  et la loi de  $Y$  sont explicites, vous pouvez donner explicitement les valeurs de  $c_\varepsilon$  de  $d_\varepsilon$  (si les deux quantités sont connues, mais que l'un n'est pas explicite, considérer juste les valeurs de  $c_\varepsilon$  et  $d_\varepsilon$  de (V.1) ci-dessus comme connues sans les expliciter).

Ensuite, puisque,

$$\{Y \in ]c_\varepsilon, d_\varepsilon]\} = \{\sqrt{n}(\bar{X}_n - \theta) \in ]c_\varepsilon, d_\varepsilon]\} = \left\{ \theta \in \left] \bar{X}_n - \frac{d_\varepsilon}{\sqrt{n}}; \bar{X}_n - \frac{c_\varepsilon}{\sqrt{n}} \right[ \right\},$$

l'intervalle suivant

$$I := \left] \bar{X}_n - \frac{d_\varepsilon}{\sqrt{n}}; \bar{X}_n - \frac{c_\varepsilon}{\sqrt{n}} \right[ ,$$

où  $c_\varepsilon$  et  $d_\varepsilon$  sont définies à (V.1), est un intervalle de confiance exact symétrique de niveau  $1 - \varepsilon$ .

**Remarque V.1.11.** La longueur de l'intervalle de confiance ci-dessus est  $(d_\varepsilon - c_\varepsilon)/\sqrt{n}$ . Ce qui implique que, à niveau  $1 - \varepsilon$  fixé, vous pouvez choisir un intervalle de confiance aussi précis que vous voulez quitte à augmenter la taille  $n$  de votre échantillon.

## Construction d'un intervalle de confiance unilatéral via une fonction pivotable

La construction d'un intervalle unilatéral est très similaire à celle d'un intervalle symétrique. Prenons l'exemple d'un échantillon  $X_1, \dots, X_n$  (i.e. les  $X_k$  sont i.i.d.) de loi  $\mathcal{U}(]0, \theta])$  avec  $\theta > 0$  le paramètre inconnu à estimer.

Il faut d'abord trouver une fonction pivotable (de préférence la plus simple possible), et pour cela, commençons par trouver un estimateur naturel pour  $\theta$ , par exemple

$$\hat{\theta} := \max(X_1, \dots, X_n).$$

On peut obtenir explicitement la loi de  $\hat{\theta}$  par la Proposition IV.3.13 qui nous donne sa fonction de répartition :

$$F_{\hat{\theta}}(t) = F_{X_1}(t)^n = \begin{cases} 0 & \text{si } t \leq 0, \\ (t/\theta)^n & \text{si } 0 \leq t \leq \theta, \\ 1 & \text{si } t \geq \theta. \end{cases}$$

La loi de  $\hat{\theta}$  dépend de  $\theta$ , donc cet estimateur ne peut pas servir de fonction pivotable. Définissons donc

$$Y := \hat{\theta}/\theta = \max(X_1/\theta, \dots, X_n/\theta).$$

Comme  $X_k \sim \mathcal{U}(]0, \theta])$ , nous savons que  $Y_k := X_k/\theta \sim \mathcal{U}(]0, 1])$ . Donc la loi des  $Y_k$  ne dépend pas de  $\theta$ , et on peut montrer que la loi de  $Y = \max(Y_1, \dots, Y_n)$  ne dépend pas de  $\theta$  non plus. En effet, en réutilisant la Proposition IV.3.13, on a

$$F_Y(t) = \begin{cases} 0 & \text{si } t \leq 0, \\ t^n & \text{si } 0 \leq t \leq 1, \\ 1 & \text{si } t \geq 1. \end{cases}$$

Si on s'intéresse à garantir une borne inférieure pour  $\theta$  sans se soucier de la borne supérieure, alors cela implique que l'on veut un intervalle de confiance de la forme  $]m_\varepsilon, +\infty[$  (avec  $m_\varepsilon$  une v.a.r. à déterminer). Et donc, comme la fonction pivotable  $Y = \hat{\theta}/\theta$  est une fonction décroissante de  $\theta$ , il faut s'intéresser aux événements de la forme  $\{Y \in ]0, c_\varepsilon[$  (où  $c_\varepsilon$  est une constante déterministe à déterminer). Remarquer que, en toute généralité, il faudrait s'intéresser à l'événement  $\{Y \in ]-\infty, c_\varepsilon[$ , mais comme on sait que la v.a.r.  $Y$  ne prend que des valeurs strictement positives, cela revient à regarder les événements  $\{Y \in ]0, c_\varepsilon[$ .

Donc pour avoir un intervalle de niveau  $1 - \varepsilon$ , il faut que  $c_\varepsilon$  vérifie :

$$\mathbb{P}(Y \geq c_\varepsilon) = \varepsilon,$$

Ce qui implique

$$1 - c_\varepsilon^n = \varepsilon$$

et donc

$$c_\varepsilon = (1 - \varepsilon)^{1/n}.$$

Puis, comme

$$\{Y \in ]0, c_\varepsilon[ = \{\hat{\theta}/\theta \in ]0, c_\varepsilon[ = \{\theta \in ]\hat{\theta}/c_\varepsilon, +\infty[ ,$$

on vient de montrer que l'intervalle

$$I := ]\max(X_1, \dots, X_n)/(1 - \varepsilon)^{1/n}, +\infty[$$

est un intervalle de confiance unilatéral de niveau  $1 - \varepsilon$  pour  $\theta$ .

## V.2 Intervalle de confiance asymptotique

Les intervalles de confiance exactes sont généralement assez délicats à construire (en cours, vos enseignants ne font que les cas qui marchent bien). Les intervalles de confiance asymptotique, même s'ils sont moins précis, donnent un cadre plus riche pour l'estimation de paramètres.

**Définition V.2.1.** Une suite d'intervalles de confiance asymptotique de niveau  $1 - \varepsilon$  pour  $\theta$  (où  $0 < \varepsilon < 1$ ) est une suite d'intervalles aléatoires  $(I_n)_n$  vérifiant

$$\lim_{n \rightarrow +\infty} \mathbb{P}(\theta \in I_n) \geq 1 - \varepsilon.$$

**Définition V.2.2.** Un intervalle de confiance asymptotique  $]U_n, V_n[$  est dit symétrique si

$$\lim_{n \rightarrow \infty} \mathbb{P}(\theta \leq U_n) = \lim_{n \rightarrow \infty} \mathbb{P}(\theta \geq V_n).$$

Dans ce cas, si on note  $\varepsilon = \lim_{n \rightarrow \infty} \mathbb{P}(\theta \notin ]U_n, V_n[)$ , alors les deux limites ci-dessus sont égales à  $\varepsilon/2$ .

**Définition V.2.3.** Un intervalle de confiance asymptotique  $]U_n, V_n[$  est dit unilatéral si

$$\lim_{n \rightarrow \infty} \mathbb{P}(\theta \leq U_n) = 0 \text{ ou } \lim_{n \rightarrow \infty} \mathbb{P}(\theta \geq V_n) = 0.$$

**Définition V.2.4.** Soit  $(T, \Theta, (P_\theta)_{\theta \in \Theta})$  un modèle statistique. Une fonction pivotable asymptotique est une suite de v.a.r.  $(Y_n)_n$  qui dépend de  $T$  et de  $\theta$  telle qu'il existe une v.a.r.  $Y$  telle que la loi de  $Y$  est connue (donc ne dépend pas de  $\theta$ ) et telle que, pour tout intervalle  $I$ ,

$$\mathbb{P}(Y_n \in I) \xrightarrow[n \rightarrow \infty]{} \mathbb{P}(Y \in I).$$

Comme pour les intervalles de confiance exacts, il existe une méthode simple mais restrictive qui repose sur l'inégalité de Bienaymé-Chebychev. Pour la version asymptotique, on peut utiliser la Loi Faible des Grands Nombres (dont la preuve repose sur cette inégalité).

**Lemme V.2.5.** Soient  $X_1, \dots, X_n, \dots$  des v.a.r. i.i.d. dont l'espérance est  $\theta$  (le paramètre à estimer) et dont la variance  $\sigma^2$  est finie et connue. Alors, pour tout  $\varepsilon > 0$ , l'intervalle

$$I_n := \left] \frac{1}{n} \sum_{k=1}^n X_k - \varepsilon, \frac{1}{n} \sum_{k=1}^n X_k + \varepsilon \right[$$

est un intervalle de confiance asymptotique pour  $\theta$  de niveau  $1 - \sigma^2/(n\varepsilon^2)$ .

*Démonstration.* Cf cours. □

**Remarque V.2.6.** Si le paramètre  $\theta$  n'est pas l'espérance des v.a.r.  $X_k$ , mais l'espérance d'une fonction de cette espérance (disons  $\theta = g(\mathbb{E}[X_k])$ ), alors le lemme précédent peut encore être utilisé dans certains cas. Par exemple, si la fonction  $g$  est croissante, alors

$$\left] g\left(\frac{1}{n} \sum_{k=1}^n X_k - \varepsilon\right), g\left(\frac{1}{n} \sum_{k=1}^n X_k + \varepsilon\right) \right[$$

est un intervalle de confiance asymptotique pour  $\theta = g(\mathbb{E}[X_k])$  de niveau  $1 - \sigma^2/(n\varepsilon^2)$ .

## Construction d'un intervalle de confiance asymptotique via une fonction pivotable

Nous ne ferons que le cas symétrique (le cas unilatéral est la même variante que celle qui a été vue pour les intervalles exacts). Dans notre exemple, considérons des v.a.r.  $X_1, \dots, X_n, \dots$  i.i.d. de loi  $\mathcal{E}(1/\theta)$ . D'après le Théorème Central Limite, la v.a.r.

$$\frac{\sqrt{n}}{\theta} \left( \frac{1}{n} \sum_{k=1}^n X_k - \theta \right)$$

est une fonction pivotable dont la loi limite est  $\mathcal{N}(0, 1)$ .

Pour construire un intervalle symétrique de niveau  $1 - \varepsilon$ , il faut donc trouver deux réels  $c_\varepsilon, d_\varepsilon$  tels que, si on note  $Y$  une v.a.r. de loi  $\mathcal{N}(0, 1)$  :

$$\mathbb{P}(Y \leq c_\varepsilon) = \varepsilon/2 \text{ et } \mathbb{P}(Y \geq d_\varepsilon) = \varepsilon/2.$$

C'est-à-dire

$$c_\varepsilon = F_Y^{-1}(\varepsilon/2) \text{ et } d_\varepsilon = F_Y^{-1}(1 - \varepsilon/2).$$

Si le  $\varepsilon$  est explicite, il serait possible de donner des valeurs approchées de  $c_\varepsilon$  et  $d_\varepsilon$  à l'aide de la table des valeurs de  $F_Y$ .

Puis, comme

$$\begin{aligned} \left\{ \frac{\sqrt{n}}{\theta} \left( \frac{1}{n} \sum_{k=1}^n X_k - \theta \right) \in ]c_\varepsilon, d_\varepsilon[ \right\} &= \left\{ \sqrt{n} \left( \frac{1}{\theta} \left( \frac{1}{n} \sum_{k=1}^n X_k \right) - 1 \right) \in ]c_\varepsilon, d_\varepsilon[ \right\} \\ &= \left\{ \theta \in \left] \frac{\frac{1}{n} \sum_{k=1}^n X_k}{1 + d_\varepsilon/\sqrt{n}}, \frac{\frac{1}{n} \sum_{k=1}^n X_k}{1 + c_\varepsilon/\sqrt{n}} \right[ \right\}, \end{aligned}$$

on vient de montrer que l'intervalle

$$I_n = \left] \frac{\frac{1}{n} \sum_{k=1}^n X_k}{1 + d_\varepsilon/\sqrt{n}}, \frac{\frac{1}{n} \sum_{k=1}^n X_k}{1 + c_\varepsilon/\sqrt{n}} \right[$$

est un intervalle de confiance asymptotique symétrique pour  $\theta$  de niveau  $1 - \varepsilon$ .

**Remarque V.2.7.** Dans cet exemple, la quantité  $\frac{\sqrt{n}}{\sigma} \left( \frac{1}{n} \sum_{k=1}^n X_k - m \right)$  est particulièrement simple à manipuler car  $\sigma = m = \theta$ . Dans les cas où l'écart-type est plus compliqué, il peut être utile de remplacer la variance exacte  $\sigma^2$  par son estimateur naturel  $\hat{\sigma}_n^2$  introduite dans le chapitre précédent. Dans ce cas, l'intervalle de confiance dépend de la variance empirique, en plus de dépendre de la moyenne empirique.

# Chapitre VI

## Tests statistiques

### VI.1 Généralités

Informellement, un test statistique peut être vu comme un processus pour décider si une hypothèse sur le paramètre inconnu  $\theta$  est vraie (ou plutôt s'il est plausible). L'hypothèse que l'on cherche à tester est l'hypothèse nulle (que l'on notera  $(H_0)$ ), et, dans le contexte le plus général, on testera l'hypothèse contre une hypothèse alternative (notée  $(H_1)$ ), qui sera souvent simplement le complémentaire de l'hypothèse nulle. Le principe du test repose sur ce qu'on appelle une zone de rejet (qui est une partie du domaine des réalisations de l'échantillon  $(X_1, \dots, X_n)$ ) qu'il faut comprendre de la manière suivante : si la réalisation de l'échantillon est dans la zone, alors on rejette  $(H_0)$  (moralement, cela signifie que  $(H_0)$  n'est statistiquement pas plausible) et on accepte  $(H_1)$ , et sinon on accepte  $(H_0)$ .

Dans ce type de test, il y a deux types d'erreurs : accepter  $(H_0)$  quand elle est fausse, et rejeter  $(H_0)$  quand elle est vraie. Dans la "philosophie" des tests statistiques, c'est la deuxième erreur qui est grave : on ne veut surtout pas rejeter  $(H_0)$  si elle est vraie (c'est la raison pour laquelle les définitions et les méthodes ne sont pas symétriques entre  $(H_0)$  et  $(H_1)$ ).

**Exemple VI.1.1.** *Supposons qu'une personne veuille investir toutes ses économies dans un actif à risque. Cette personne est prudente et veut éviter de perdre de l'argent (quitte à rater un bon investissement). Donc elle pourrait faire un test avec l'hypothèse nulle  $(H_0)$  = "l'actif est rentable" et l'hypothèse complémentaire  $(H_1)$  = "l'actif n'est pas rentable".*

Il faut noter que "accepter  $(H_0)$ " ne signifie pas nécessairement que " $(H_0)$  doit être vraie", mais plutôt que "on ne peut pas affirmer que  $(H_0)$  est fausse avec probabilité suffisamment élevée".

Dans toute la suite, notons  $(T, \Theta, (\mathbb{P}_\theta)_{\theta \in \Theta})$  un modèle statistique avec  $T = (X_1, \dots, X_n)$ .

**Définition VI.1.2.** *Un test paramétrique est un triplet  $(\Theta_0, \Theta_1, R)$  (où  $\Theta_0 \cap \Theta_1 = \emptyset$ ) avec :*

- $\Theta_0 \subseteq \Theta$  : l'hypothèse nulle est alors  $(H_0) = "\theta \in \Theta_0"$ ,
- $\Theta_1 \subseteq \Theta$  : l'hypothèse complémentaire est alors  $(H_1) = "\theta \in \Theta_1"$ ,
- $R \subseteq \mathbb{R}^n$  est la zone de rejet : si  $T = (X_1, \dots, X_n) \in R$  on rejette  $(H_0)$ , et sinon on accepte  $(H_0)$ .

Dans la suite, notons  $\mathbb{P}_\theta$  la mesure de probabilité sur  $\Omega$  quand les v.a.r.  $X_1, \dots, X_n$  sont de loi  $P_\theta$ . C'est-à-dire que, par exemple, pour tout intervalle  $I$ ,

$$\mathbb{P}_\theta(X_1 \in I) = P_\theta(I).$$

**Définition VI.1.3.** *Soit  $(\Theta_0, \Theta_1, R)$  un test paramétrique. On appelle :*

- fonction de risque de première espèce, la fonction

$$\alpha : \theta \in \Theta_0 \mapsto \mathbb{P}_\theta((X_1, \dots, X_n) \in R) \in [0, 1],$$

— fonction de risque de seconde espèce, la fonction

$$\beta : \theta \in \Theta_1 \mapsto \mathbb{P}_\theta((X_1, \dots, X_n) \notin R) \in [0, 1],$$

— seuil du test, la quantité suivante

$$\sup_{\theta \in \Theta_0} \alpha(\theta).$$

**Remarque VI.1.4.** Le risque de première espèce correspond à la probabilité de rejeter ( $H_0$ ) à tort (car  $\alpha$  est définie sur  $\Theta_0$ ), et le risque de seconde espèce à la probabilité d'accepter ( $H_0$ ) à tort. Comme expliqué plus tôt, c'est le premier cas que l'on souhaite éviter, c'est pourquoi dans les tests, nous chercherons d'abord à garantir que le seuil du test est petit (i.e. on voudra minimiser la probabilité de rejeter ( $H_0$ ) à tort). Une fois que le seuil du test sera garantie "assez petit" (disons inférieur à une quantité donnée dans les énoncés des exercices), on pourra chercher à minimiser le risque de seconde espèce parmi tous les tests possibles.

**Définition VI.1.5.** On dit que le test est un test d'hypothèses simples si  $\Theta_0$  et  $\Theta_1$  sont des singletons (i.e.  $(H_0) = "\theta = \theta_0"$  et  $(H_1) = "\theta = \theta_1"$  où  $\theta_0, \theta_1$  sont deux éléments de  $\Theta$ ). Si ce n'est pas le cas, on parle de test d'hypothèses composites.

**Remarque VI.1.6.** Dans le cadre d'un test d'hypothèses simples, les risques de première et seconde espèces ne sont pas des fonctions, mais simplement des constantes de  $[0, 1]$  (et le seuil est exactement le risque de première espèce).

Trouver une zone de rejet avec un seuil  $\varepsilon$  donné, est souvent un problème similaire à celui de trouver un intervalle de confiance de niveau  $1 - \varepsilon$  donné. La méthode de base consiste à trouver un estimateur naturel pour  $\theta$ , puis à contrôler la probabilité que l'estimateur ne soit pas proche de  $\theta$ . Donc les méthodes du chapitre précédent sont utiles pour construire des zones de rejet. Illustrons cette remarque dans l'exemple suivant.

## VI.2 Exemple de test

Supposons qu'une personne propose un jeu de pile ou face (en utilisant toujours la même pièce) à des gens dans une rue, et que vous voulez savoir si cette personne est un arnaqueur. Notons  $\theta \in [0, 1]$  la probabilité que la pièce fasse pile. Alors les hypothèses que vous voudrez tester sont  $(H_0) = "\theta = 1/2"$  contre  $(H_1) = "\theta \neq 1/2"$ . Vous avez observé  $n$  lancers de la pièce, et avez noté  $X_k = 1$  si le  $k$ -ième lancer fait pile, et  $X_k = 0$  s'il fait face ( $1 \leq k \leq n$ ). L'hypothèse nulle correspond à "le joueur n'est pas un arnaqueur", cela peut se justifier par le fait que vous ne voulez pas l'accuser de tricher à tort par exemple.

Pour chercher une zone de rejet de seuil au plus égal à un certain  $\varepsilon$  donné, cherchons d'abord un estimateur naturel pour  $\theta$ . Ici, comme les  $X_k$  sont des v.a.r. i.i.d. de loi  $\mathcal{B}(\theta)$ , un estimateur simple est la moyenne empirique  $\bar{X}_n$  de l'échantillon, et une zone de rejet naturelle est la suivante (qui consiste simple à tester si la moyenne empirique est proche ou non de l'espérance)

$$R = \{(x_1, \dots, x_n) \in \{0, 1\}^n : |\bar{x}_n - 1/2| > c_\varepsilon\},$$

où  $c_\varepsilon$  est une constante à déterminer.

Pour que cette zone convienne, il faut qu'elle vérifie

$$\varepsilon \geq \mathbb{P}_{1/2} \left( \left| \bar{X}_n - \frac{1}{2} \right| > c_\varepsilon \right).$$

Or, d'après la Loi Faible des Grands Nombres (ou par l'inégalité de Bienaymé-Chebychev),

$$\mathbb{P}_{1/2} \left( \left| \bar{X}_n - \frac{1}{2} \right| > c_\varepsilon \right) \leq \frac{1}{4nc_\varepsilon^2}.$$

Il suffit donc de choisir n'importe quel  $c_\varepsilon$  vérifiant

$$\frac{1}{4nc_\varepsilon^2} \leq \varepsilon.$$

Prenons (par exemple) le choix optimal

$$c_\varepsilon = \frac{1}{2\sqrt{n\varepsilon}}.$$

Donc votre test paramétrique ici est défini par  $\Theta_0 = \{1/2\}$ ,  $\Theta_1 = [0, 1] \setminus \{1/2\}$  et surtout

$$R = \{(x_1, \dots, x_n) \in \{0, 1\}^n : |\bar{x}_n - 1/2| > 1/(2\sqrt{n\varepsilon})\}.$$

**Applications :**

- vous observez  $n = 100$  lancers avec 70 piles et 30 faces. Si vous voulez un risque de première espèce inférieur à  $\varepsilon = 0.01$ , pouvez-vous accuser la personne de tricher ? C'est-à-dire, s'il ne triche pas, la probabilité de l'accusé à tort doit être inférieure à 0.01.
- même question si vous observez  $n = 10000$  lancers avec 7000 piles et 3000 faces. Que pouvez-vous en conclure ?

# Chapitre VII

## Vraisemblance

**Définition VII.0.1** (Vraisemblance). Soit  $(\mathcal{X}, \Theta, (P_\theta)_{\theta \in \Theta})$  un modèle statistique associé à  $T = (X_1, \dots, X_n)$  où les v.a.r.  $X_k$  sont i.i.d. La vraisemblance du modèle est la fonction  $V : \mathbb{R}^n \times \Theta \rightarrow \mathbb{R}_+$  définie par :

— si les  $P_\theta$  sont des lois discrètes,

$$V((x_1, \dots, x_n), \theta) := \prod_{k=1}^n P_\theta(\{x_k\}) := \prod_{k=1}^n \mathbb{P}_\theta(X_k = x_k),$$

— si les  $P_\theta$  admettent une densité  $f_\theta$ ,

$$V((x_1, \dots, x_n), \theta) := \prod_{k=1}^n f_\theta(x_k).$$

On appelle log-vraisemblance la fonction :

$$L((x_1, \dots, x_n), \theta) := \ln(V(x_1, \dots, x_n), \theta).$$

**Remarque VII.0.2.** L'intérêt du log-vraisemblance est qu'il est parfois plus simple à étudier que la vraisemblance.

**Exemple VII.0.3.** Considérons le modèle statistique associé à l'échantillon  $(X_1, \dots, X_n)$  où les v.a.r.  $X_k$  sont i.i.d. de loi  $\mathcal{B}(\theta)$  avec  $\theta \in ]0, 1[$ . La vraisemblance du modèle est définie par : pour  $x_1, \dots, x_n \in \{0, 1\}$  et  $\theta \in ]0, 1[$ ,

$$V((x_1, \dots, x_n), \theta) = \prod_{k=1}^n \mathbb{P}_\theta(X_k = x_k),$$

où  $\mathbb{P}_\theta(X_k = x_k)$  est égale à  $\theta$  si  $x_k = 1$ ,  $1 - \theta$  si  $x_k = 0$ , et 0 pour toutes les autres valeurs de  $x_k$ . Cela donne donc

$$V((x_1, \dots, x_n), \theta) = \theta^{\sum_{k=1}^n x_k} (1 - \theta)^{\sum_{k=1}^n (1 - x_k)} = \theta^{\sum_{k=1}^n x_k} (1 - \theta)^{n - \sum_{k=1}^n x_k}.$$

Ici la forme de la vraisemblance peut être dure à manipuler pour ce que nous ferons dans la suite à cause des produits et des exponentielles. Dans cet exemple, le log-vraisemblance est plus simple à manipuler : pour  $x_1, \dots, x_n \in \{0, 1\}$  et  $\theta \in ]0, 1[$ ,

$$L((x_1, \dots, x_n), \theta) = \left( \sum_{k=1}^n x_k \right) \ln \theta + \left( n - \sum_{k=1}^n x_k \right) \ln(1 - \theta).$$

Dans ce cours, nous ne traiterons que deux utilisations de la notion de vraisemblance :

- l'estimateur du maximum de vraisemblance,
- et le test du rapport de vraisemblance.

L'avantage de cet estimateur et de ce test (par rapport aux autres estimateurs et tests vus dans les chapitres précédents), est qu'ils sont génériques : il est toujours possible de les définir. Par exemple, il n'est pas nécessaire de trouver un lien entre le paramètre  $\theta$  à estimer et les caractéristiques de la loi à estimer  $P_\theta$  (comme l'espérance ou la variance).

L'inconvénient est que les calculs sont souvent plus difficiles, et les propriétés théoriques plus dures à étudier.

## VII.1 Estimateur du maximum de vraisemblance

**Définition VII.1.1.** Soit  $(\mathcal{X}, \Theta, (P_\theta)_{\theta \in \Theta})$  un modèle statistique associé à  $T = (X_1, \dots, X_n)$ . On appelle estimateur du maximum de vraisemblance de  $\theta$ , toute v.a.r.  $\hat{\theta}$  vérifiant

$$V(T, \hat{\theta}) = \sup_{\theta \in \Theta} V(T, \theta).$$

**Remarque VII.1.2.** De manière équivalente, l'estimateur du maximum de vraisemblance  $\hat{\theta}$  satisfait aussi

$$L(T, \hat{\theta}) = \sup_{\theta \in \Theta} L(T, \theta).$$

**Exemple VII.1.3.** Pour trouver l'estimateur du maximum de vraisemblance de l'Exemple VII.0.3, il faut d'abord étudier les variations de la vraisemblance ou du log-vraisemblance. Comme évoqué précédemment, dans cet exemple, c'est le log-vraisemblance qui est le plus simple à manipuler. Donc, pour  $x_1, \dots, x_n \in \{0, 1\}$  fixés, il faut étudier les variations de la fonction

$$\theta \in ]0, 1[ \mapsto L((x_1, \dots, x_n), \theta).$$

Commençons par dériver cette fonction : pour  $\theta \in ]0, 1[$ ,

$$\frac{\partial}{\partial \theta} L((x_1, \dots, x_n), \theta) = \left( \sum_{k=1}^n x_k \right) \frac{1}{\theta} - \left( n - \sum_{k=1}^n x_k \right) \frac{1}{1-\theta}.$$

Comme

$$\frac{\partial}{\partial \theta} L((x_1, \dots, x_n), \theta) \geq 0 \iff \frac{1-\theta}{\theta} \geq \frac{n}{\sum_{k=1}^n x_k} - 1 \iff \theta \leq \frac{1}{n} \sum_{k=1}^n x_k,$$

on peut dresser le tableau de variation suivant

$\theta$	0	$\frac{1}{n} \sum_{k=1}^n x_k$	1
$\frac{\partial}{\partial \theta} L((x_1, \dots, x_n), \theta)$	+	0	-
$\theta \mapsto L((x_1, \dots, x_n), \theta)$	$\nearrow$		$\searrow$

ce qui implique que la fonction  $\theta \mapsto L((x_1, \dots, x_n), \theta)$  atteint son maximum (dont on ne cherche pas la valeur) au point  $\frac{1}{n} \sum_{k=1}^n x_k$ .

Donc l'estimateur du maximum de vraisemblance de ce modèle est

$$\hat{\theta} = \frac{1}{n} \sum_{k=1}^n X_k,$$

qui est la moyenne empirique de l'échantillon.

## VII.2 Test du rapport de vraisemblances

**Définition VII.2.1.** Le test du rapport de vraisemblance de seuil  $\alpha$  est le test d'hypothèses simples  $(H_0) = "\theta = \theta_0"$  contre  $(H_1) = "\theta = \theta_1"$  définie par la zone de rejet

$$R = \left\{ (x_1, \dots, x_n) \in \mathbb{R}^n : \frac{V((x_1, \dots, x_n), \theta_1)}{V((x_1, \dots, x_n), \theta_0)} > b_\alpha \right\},$$

où  $b_\alpha > 0$  satisfait

$$\mathbb{P}_{\theta_0}((X_1, \dots, X_n) \in R) \leq \alpha.$$

**Remarque VII.2.2.** Parmi les tests dont le risque de première espèce est  $\alpha$ , le test du rapport de vraisemblance est celui qui minimise le risque de seconde espèce.

**Remarque VII.2.3.** De manière équivalente, la zone de rejet du test de rapport de vraisemblances s'écrit aussi

$$R = \{(x_1, \dots, x_n) \in \mathbb{R}^n : L((x_1, \dots, x_n), \theta_1) - L((x_1, \dots, x_n), \theta_0) > c_\alpha\},$$

avec  $c_\alpha \in \mathbb{R}$  vérifiant

$$\mathbb{P}_{\theta_0}((X_1, \dots, X_n) \in R) \leq \alpha.$$

Avec les notations précédentes  $c_\alpha = \ln b_\alpha$ .

**Exemple VII.2.4.** Reprenons le modèle statistique de l'Exemple VII.0.3. Supposons que, comme dans l'exemple donné à la Section VI.2, on suppose que l'on cherche à déterminer si un joueur triche ou non dans un jeu de pile ou face. Mais, contrairement à la Section VI.2, nous ne pouvons pas faire un test d'hypothèses composites pour faire un test de rapport de vraisemblances. Donc considérons l'hypothèse nulle  $(H_0) = "\theta = 1/2"$  (i.e. le cas où le joueur ne triche pas) contre l'hypothèse  $(H_1) = "\theta = 7/10"$  (car dans la partie **Applications** de la fin de la Section VI.2 cette probabilité semble plus naturelle que  $1/2$ ).

La zone de rejet du test de rapport de vraisemblances (de seuil  $\varepsilon$ ) pour ces hypothèses est donc (pour une certaine constante  $c_\varepsilon$  que nous déterminerons à la fin)

$$\begin{aligned} R &= \left\{ (x_1, \dots, x_n) \in \{0, 1\}^n : \left( \sum_{k=1}^n x_k \right) (\ln(7/10) - \ln(1/2)) + \left( n - \sum_{k=1}^n x_k \right) (\ln(3/10) - \ln(1/2)) > c_\varepsilon \right\} \\ &= \left\{ (x_1, \dots, x_n) \in \{0, 1\}^n : \left( \sum_{k=1}^n x_k \right) \ln(7/3) + \ln(3/5) > c_\varepsilon \right\} \\ &= \left\{ (x_1, \dots, x_n) \in \{0, 1\}^n : \sum_{k=1}^n x_k > d_\varepsilon \right\}, \end{aligned}$$

où  $d_\varepsilon = (c_\varepsilon + n \ln(5/3)) / \ln(7/3)$ . Toutefois, l'expression précédente de  $d_\varepsilon$  n'a aucune importance : en effet, pour déterminer  $d_\varepsilon$ , il suffit de savoir que cette constante doit vérifier

$$\mathbb{P}_{1/2}((X_1, \dots, X_n) \in R) \leq \varepsilon.$$

Or, sous  $\mathbb{P}_{1/2}$  les v.a.r.  $X_k$  sont i.i.d. de loi  $\mathcal{B}(1/2)$ , et donc  $\sum_{k=1}^n X_k$  suit la loi  $\mathcal{B}(n, 1/2)$ . En particulier, la zone de rejet est

$$R = \left\{ (x_1, \dots, x_n) \in \{0, 1\}^n : \sum_{k=1}^n x_k > d_\varepsilon \right\},$$

où  $d_\varepsilon$  vérifie

$$F_{\mathcal{B}(n, 1/2)}(d_\varepsilon) \geq 1 - \varepsilon.$$

Pour un  $\varepsilon$  et un  $n$  donnés explicitement, il est possible de donner une valeur explicite de  $d_\varepsilon$ . Si  $n$  est assez petit (disons inférieur à 5) il est possible de faire les calculs à la main, autrement il faudrait les faire numériquement.