

# Table des matières

<b>Calcul de probabilités</b>	<b>iii</b>
0.1 Rappel de probabilités . . . . .	iii
0.1.1 Espace de probabilité . . . . .	iii
0.1.2 Variables aléatoires . . . . .	v
0.1.3 Loi d'une variable aléatoire . . . . .	vi
0.1.4 Fonction de répartition d'une variable aléatoire . . . . .	vii
0.2 Couples des variables aléatoires . . . . .	viii
0.2.1 Lois associées à un couple de variables aléatoires . . . . .	viii
0.2.2 Lois marginales . . . . .	ix
0.2.3 Lois conditionnelle . . . . .	ix
0.2.4 Moments conditionnels . . . . .	x
0.2.5 Moments associés à un couple . . . . .	xi
0.3 Couple de variables aléatoires continues . . . . .	xii
0.3.1 Loi du couple . . . . .	xii
0.3.2 Lois marginales . . . . .	xiii
0.3.3 Lois conditionnelle . . . . .	xiv
0.3.4 Moments associés à un couple . . . . .	xiv
0.3.5 Régression . . . . .	xvi
<b>Convergences</b>	<b>xvii</b>
0.4 Convergence en probabilité . . . . .	xvii
0.4.1 Inégalité de Markov . . . . .	xvii
0.4.2 Inégalité de Bienaymé-Tchebychev . . . . .	xvii
0.4.3 Convergence en probabilité . . . . .	xvii
0.5 Convergence presque-sûre et loi des grands nombres . . . . .	xviii
0.6 Convergence en loi . . . . .	xxi

<b>Statistique inférentielle</b>	<b>xxiii</b>
0.7 Echantillonnage-Estimation . . . . .	xxiii
0.7.1 Distribution d'échantillonnage . . . . .	xxv
0.7.2 Distribution d'échantillonnage des moyennes . . . . .	xxv
0.7.3 Valeurs caractéristiques de $\bar{X}_n$ . . . . .	xxv
0.7.4 Distribution d'échantillonnage des fréquences . . . . .	xxvii
0.8 Estimation de Paramètres . . . . .	xxix
0.8.1 Estimateur optimal (efficace) . . . . .	xxx
0.8.2 Qualité d'un estimateur . . . . .	xxx
0.8.3 Estimateur efficace . . . . .	xxxi
0.8.4 Inégalité de Fréchet-Darmois-Cramer-Rao(F.D.C.R) . . . . .	xxxii
0.8.5 Estimateur efficace . . . . .	xxxiv
0.8.6 Méthodes de construction d'un estimateur . . . . .	xxxiv
0.8.7 Méthode du maximum de vraisemblance . . . . .	xxxiv
0.8.8 Estimation d'une moyenne par intervalle de confiance . . . . .	xxxvi
<b>Tests d'hypothèses</b>	<b>xxxviii</b>
0.9 Catégories de tests . . . . .	xxxix
0.9.1 Région critique et d'acceptation d'une hypothèse . . . . .	xxxix
0.9.2 Test entre deux hypothèses simples (méthode de Ney- mane et Pearson) . . . . .	xl
0.9.3 Test d'homogénéité . . . . .	xli
0.9.4 Test d'homogénéité de deux moyennes . . . . .	xli
0.9.5 Test d'homogénéité de deux proportion . . . . .	xliii

# Calcul de probabilités

## 0.1 Rappel de probabilités

### 0.1.1 Espace de probabilité

Un espace de probabilité est un triplet  $(\Omega, F, P)$  où :

- $\Omega$  est un ensemble,

- $F$  est une tribu ( ou  $\sigma$ -algèbre) sur  $\Omega$ ,

- $P$  est une (mesure de) probabilité sur  $(\Omega, F)$ .

**Définition 0.1.1** Une tribu (ou  $\sigma$ -algèbre) sur  $\Omega$  est une famille  $F$  de sous ensemble de  $\Omega$  (appelés “événements”) tels que

$$\begin{cases} i) \Phi \in F \\ ii) A \in F \Rightarrow A^c \in F \\ iii) (A_n)_{n=1}^{\infty} \subset F \Rightarrow \cup_{n=1}^{\infty} A_n \in F \end{cases}$$

En particulier :  $A, B \in F \Rightarrow A \cup B \in F$ . De même,  $A, B \in F \Rightarrow A \cap B \in F$ .

**Exemple 0.1.1** Soit  $\Omega = \{1, 2, \dots, 6\}$ . On peut définir plusieurs tribus sur  $\Omega$  :  $F = P(\Omega) = \{\Phi, \{1\}, \{2\}, \dots, \{1, 2\}, \dots, \Omega\}$  =tribu complète (la plus grande),  $F_0 = \{\Phi, \Omega\}$  =tribu triviale ( la plus petite) ( $\Phi$  = évènement impossible,  $\Omega$  = évènement arbitraire ou certain).  $F_1 = \{\Phi, \{1\}, \{2, \dots, 6\}, \Omega\}$ ,  $F_2 = \{\Phi, \{1, 3, 5\}, \{2, 4, 6\}, \Omega\}$ , etc. Soit  $\Omega = [0, 1]$  et  $I_1, I_2, \dots, I_n$  une famille d'intervalles formant une partition de  $\Omega$ . la famille de sous-ensembles définie par :

$$G = \{\Phi, I_1, I_2, \dots, I_1 \cup I_2, \dots, I_1 \cup I_2 \cup I_3, \dots, \Omega\} \text{ est une tribu sur } \Omega.$$

**Définition 0.1.2** Soit  $\mathcal{A} = \{A_i, i \in I\}$  une famille de sous-ensembles de  $\Omega$ . Alors la tribu engendrée  $\mathcal{A}$  est la plus petite tribu sur  $\Omega$  qui contient tous les sous-ensembles  $A_i, i \in I$ . Elle est notée  $\sigma(\mathcal{A})$ . (NB : L'ensemble  $I$  n'est pas forcément dénombrable)

**Exemple 0.1.2** Reprenons l'exemple 1. Soit  $\Omega = \{1, \dots, 6\}$ . Si  $\mathcal{A}_1 = \{\{1\}\}$ , alors  $\sigma(\mathcal{A}_1) = F_1$ . Si  $\mathcal{A}_2 = \{\{1, 3, 5\}\}$ , alors  $\sigma(\mathcal{A}_2) = F_2$ . Et si  $\mathcal{A} = \{\{1, 3, 5\}, \{1, 2, 3\}\}$ , alors  $\sigma(\mathcal{A}) = ?$  Soit  $\Omega = [0, 1]$ . Si  $\mathcal{A} = \{I_1, I_2, \dots, I_n\}$ , alors  $\sigma(\mathcal{A}) = G$ .

**Définition 0.1.3** Soit  $\Omega = [0, 1]$ . La tribu borélienne sur  $[0, 1]$  est la tribu engendrée par la famille de sous-ensembles

$$\mathcal{A} = \{]a, b[ : 0 \leq a < b \leq 1\} = \{\text{intervalles ouverts dans } [0, 1]\}$$

Elle est notée  $\mathcal{B}([0, 1])$ . Elle contient un très grand nombre de sous-ensembles de  $[0, 1]$ , mais pas tous.

**Remarque 0.1.1** \*Pour  $\Omega$  ensemble fini, on choisit souvent  $\mathcal{F} = \mathcal{P}(\Omega)$ . \*\*Pour  $\Omega \subset \mathbb{R}^n$ , on choisit souvent  $\mathcal{F} = \mathcal{B}(\Omega)$ .

**Définition 0.1.4** Une sous-tribu de  $\mathcal{F}$  est une tribu  $G$  telle que. Si  $A \in G$  alors  $A \in \mathcal{F}$ . On note  $G \subset \mathcal{F}$  :

**Remarque importante :**

Il est toujours vrai que  $A \in G$  et  $G \subset \mathcal{F} \Rightarrow A \in \mathcal{F}$ . Mais il est faux de dire que  $A \subset B$  et  $B \in \mathcal{F} \Rightarrow A \in \mathcal{F}$ .

**Contre exemple**

$$\{1\} \subset \{\{1, 3, 5\}, \{1, 3, 5\}\} \subset \mathcal{F}_2, \text{ mais } \{1\} \notin \mathcal{F}_2.$$

**Définition 0.1.5** Soit  $\mathcal{F}$  une tribu sur  $\Omega$ . Une (mesure de) probabilité sur  $(\Omega, \mathcal{F})$  est une application  $P : \mathcal{F} \rightarrow [0, 1]$  telle que

$$\begin{cases} i) P(\emptyset) = 0 \text{ et } P(\Omega) = 1, \\ ii) (A_n)_{n=1}^\infty \subset \mathcal{F} \text{ disjoints (i.e } A_n \cap A_m = \emptyset, \forall n \neq m) \Rightarrow P(\cup_{i=1}^\infty A_n) = \sum_{n=1}^\infty P(A_n) \end{cases}$$

En particulier :  $A, B \in \mathcal{F}$  et  $A \cap B = \emptyset \Rightarrow P(A \cup B) = P(A) + P(B)$ . De plus :

$$\begin{cases} i) \text{ Si } (A_n)_{n=1}^\infty \subset \mathcal{F}, A_n \subset A_{n+1} \text{ et } \cup_{n=1}^\infty A_n = A, \text{ alors } \lim_{n \rightarrow \infty} P(A_n) = P(A). \\ ii) \text{ Si } (A_n)_{n=1}^\infty \subset \mathcal{F}, A_n \supset A_{n+1} \text{ et } \cap_{n=1}^\infty A_n = A, \text{ alors } \lim_{n \rightarrow \infty} P(A_n) = P(A). \end{cases}$$

## 0.1. RAPPEL DE PROBABILITÉS

---

**Définition 0.1.6** Soient  $\Omega = [0, 1]$  et  $\mathcal{F} = \mathcal{B}([0, 1])$ . On appelle mesure de Lebesgue sur  $[0, 1]$ , la mesure de probabilité définie par

$$P(]a, b]) = b - a, \forall 0 \leq a < b \leq 1$$

$P$  n'est pas définie a priori que sur les intervalles, mais est uniquement extensible à tous ensemble borélien  $B \in \mathcal{B}([0, 1])$ . Elle est notée

$$P(B) = |B|, B \in \mathcal{B}([0, 1]).$$

En utilisant la propriété (ii) ci-dessus, on obtient que pour tout  $x \in [0, 1]$  :

$$\{x\} = \lim_{n \rightarrow \infty} \left[ \left[ x - \frac{1}{n}, x + \frac{1}{n} \right[ \right] = \lim_{n \rightarrow \infty} \frac{2}{n} = 0$$

### Généralisation à $n$ dimensions

Soit  $\Omega = [0, 1]^n$ .

\* Tribu borélienne :  $\mathcal{B}(\Omega) = \sigma(\mathcal{A})$ , où  $\mathcal{A} = \{]a_1, b_1[ \times ]a_2, b_2[ \times \dots \times ]a_n, b_n[, 0 \leq a_i < b_i \leq 1\}$ .  $\mathcal{A}$  est la famille des "rectangles" dans  $\Omega$ .

\* Mesure de Lebesgue :  $P(]a_1, b_1[ \times ]a_2, b_2[ \times \dots \times ]a_n, b_n]) = (b_1 - a_1)(b_2 - a_2) \dots (b_n - a_n)$ .

## 0.1.2 Variables aléatoires

**Définition 0.1.7** Soit  $(\Omega, \mathcal{F}, P)$  un espace de probabilité. Une variable aléatoire (souvent abrégé **v.a** par la suite) est une application

$$X : \Omega \rightarrow \mathbb{R} \text{ telle que } \{\omega \in \Omega : X(\omega) \in B\} = \{X \in B\} \in \mathcal{F}, \forall B \in \mathcal{B}(\mathbb{R})$$

**Proposition 0.1.1**  $X$  est une **v.a** ssi  $\{\omega \in \Omega : X(\omega) \leq t\} \in \mathcal{F}, \forall t \in \mathbb{R}$ .

**Remarque 0.1.2**  $X$  est aussi dite une fonction ou (**v.a**)  $\mathcal{F}$ -mesurable. -Si  $\mathcal{F} = \mathcal{P}(\Omega)$ , alors  $X$  est toujours  $\mathcal{F}$ -mesurable. -Si  $\Omega = \mathbb{R}$  et  $\mathcal{F} = \mathcal{B}(\mathbb{R})$ , alors  $X$  est une fonction borélienne.

**Définition 0.1.8** Pour  $A \subset \Omega$ , on pose  $\chi_A(\omega) = \begin{cases} 1 & \text{si } \omega \in A \\ 0 & \text{sinon} \end{cases}$  On vérifie que la v.a  $\chi_A$  est  $\mathcal{F}$ -mesurable ssi  $A \in \mathcal{F}$ .

**Exemple 0.1.3** Soit  $(\Omega, \mathcal{F}, P)$  l'espace de probabilités du dé équilibré

$$X_1(\omega) = \omega : P(\{\omega \in \Omega; X_1(\omega) = i\}) = P(\{i\}) = \frac{1}{6}.$$

$$X_2(\omega) = \chi_{\{1,3,5\}}(\omega) : P(\{\omega \in \Omega; X_2(\omega) = 1\}) = P(\{1, 2, 3\}) = \frac{3}{6} = \frac{1}{2}.$$

Soit  $\mathcal{F} = \mathcal{P}(\Omega)$ .  $X_1$  et  $X_2$  sont toutes deux  $\mathcal{F}$ -mesurables. Soit  $\mathcal{F}_2 = \{\emptyset, \{1, 3, 5\}, \{2, 4, 6\}, \Omega\}$ . Seule  $X_2$  est  $\mathcal{F}_2$ -mesurable.  $X_1$  ne l'est pas. En effet :

$$\begin{aligned} \{\omega \in \Omega; X_2(\omega) = 1\} &= \{1, 2, 3\} \in \mathcal{F}_2 \text{ et} \\ \{\omega \in \Omega; X_2(\omega) = 0\} &= \{2, 4, 6\} \in \mathcal{F}_2 \end{aligned}$$

tandis que

$$\{\omega \in \Omega; X_1(\omega) = 1\} = \{1\} \notin \mathcal{F}_2.$$

**Définition 0.1.9** La tribu engendrée par une famille de v.a  $\{X_i, i \in I\}$  sur  $(\Omega, \mathcal{F}, P)$  est définie par

$$\sigma(X_i, i \in I) = \sigma(\{X_i \in B\}, i \in I, B \in \mathcal{B}(\mathbb{R})) = \sigma(\{X_i \leq t\}, i \in I, t \in \mathbb{R})$$

**Exemple 0.1.4** Reprenons l'exemple précédent :  $\sigma(X_1) = \mathcal{F} = \mathcal{P}(\Omega)$ ,  $\sigma(X_2) = \mathcal{F}_2 \neq \mathcal{P}(\Omega)$ .

**Proposition 0.1.2** Si  $g : \mathbb{R} \rightarrow \mathbb{R}$  est borélienne et  $X : \Omega \rightarrow \mathbb{R}$  est une v.a, alors  $g(X)$  est une v.a.

### 0.1.3 Loi d'une variable aléatoire

**Définition 0.1.10** La loi d'une v.a  $X$  est l'application  $\mu_X : \mathcal{B}(\mathbb{R}) \rightarrow [0, 1]$  définie par

$$\mu_X(B) = P(\{X \in B\}), B \in \mathcal{B}(\mathbb{R})$$

**NB :**

$(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mu_X)$  forme un nouvel espace de probabilité

**Exemple 0.1.5** Soit

$$\Omega = \{1, \dots, 6\}, \mathcal{F} = \mathcal{P}(\Omega), P(\{i\}) = \frac{1}{6} \forall i,$$

Soit

$$\Omega = \{1, \dots, 6\} \times \{1, \dots, 6\}, \mathcal{F} = \mathcal{P}(\Omega), P(\{i, j\}) = \frac{1}{36}, \forall (i, j) \in \mathcal{P}(\Omega).$$

$$X(\omega) = X(\omega_1, \omega_2) = \omega_1 + \omega_2.$$

On a alors. p.ex :

$$\mu_X(7) = P(\{X = 7\}) = P(\{(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)\}) = 6 \times \frac{1}{36} = \frac{1}{6}$$

#### 0.1.4 Fonction de répartition d'une variable aléatoire

**Définition 0.1.11** *La fonction de répartition d'une v.a est l'application  $F_X : \mathbb{R} \rightarrow [0, 1]$  définie par*

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

**Proposition 0.1.3** *La donnée de  $F_X$  équivaut à celle de  $\mu_X$ .*

## 0.2 Couples des variables aléatoires

**Définition 0.2.1** On appelle couple de variables aléatoires réelles toute application

$$\begin{cases} Z : \Omega \rightarrow \mathbb{R}^2 \\ \omega \mapsto (X(\omega), Y(\omega)) \end{cases}$$

où  $X$  et  $Y$  sont des variables aléatoires sur  $(\Omega, \mathcal{F}, P)$ . On note  $Z = (X, Y)$

**Exemple 0.2.1** Une urne contient 4 boules **blanches**, 3 **rouges** et 2 **verte**. On tire 3 boules et on note  $X$  le nombre de boules blanches,  $Y$  le nombre de boules **rouges** et on a :

$$Z(\Omega) = \{(i, j) \in \{1, 2, 3, 4, 5, 6\}\}, i \leq j$$

**Théorème 0.2.1** Soit  $Z = (X, Y)$  un couple de variables aléatoires réelles telles que  $X(\Omega) = \{x_1, \dots, x_n\}$  et  $Y(\Omega) = \{y_1, \dots, y_p\}$ . Alors, la famille d'événements

$$([X = x_i] \cap [Y = y_j])_{\substack{1 \leq i \leq n \\ 1 \leq j \leq p}}$$

forme un système complet d'événement de  $\Omega$ .

**Remarque 0.2.1** L'événement  $[X = x_i]$  et  $[Y = y_j]$  peut également se noter  $[X = x_i, Y = y_j]$  ou encore  $[(X, Y) = (x_i, y_j)]$ .

### 0.2.1 Lois associées à un couple de variables aléatoires

Le couple de variable aléatoire réelle  $(X, Y)$  est confondu avec une (**loi de**) probabilité  $P_{(X,Y)}$  sur  $\Omega = \mathbb{R}^2$ .  $\mathcal{F}$  est dans ce cas construit avec les "**pavés**"  $(a, b) \times (c, d)$ .  $(X, Y)$  représente une expérience aléatoire dont le résultat est un couple de réels. La loi du couple  $(X, Y)$  permet de caractériser les liaisons et influences mutuelles des deux caractères de l'expérience. On note

$$P_{(X,Y)}(\Delta) = P((X, Y) \in \Delta)$$

La probabilité pour que les résultats de l'expérience appartiennent à  $\Delta$ . Dans le cas où  $\Delta = \Delta_1 \times \Delta_2$  on pourra noter la quantité précédente

$$P_{(X,Y)}(\Delta) = P(X \in \Delta_1, Y \in \Delta_2)$$

### 0.2.2 Lois marginales

A la loi d'un couple sont associées deux lois marginales qui sont les lois de chacun des éléments du couples pris séparément, définies par l'ensemble des valeurs possibles et les probabilités associées obtenues par sommation, soit :

$$P_X(X = x_i) = \sum_{j \in J} P(X = x_i, Y = y_j) = \sum_{j \in J} P_{ij} = P_{i.}$$

$$P_Y(Y = y_j) = \sum_{i \in I} P(X = x_i, Y = y_j) = \sum_{i \in I} P_{ij} = P_{.j}$$

Si la loi du couple est présentée dans un tableau, ces lois sont obtenues dans les marges, par sommation de ligne ou de colonne.

$Y \setminus X$	$x_i$	
.	.	.
.	.	.
$y_j$	$P_{ij}$	$P_{.j}$
.	.	.
.	.	.
	$P_{i.}$	1

### 0.2.3 Lois conditionnelle

On peut également associer deux lois conditionnelles à la loi d'un couple, c'est-à-dire la loi d'une variable, l'autre ayant une valeur fixée (loi dans une ligne ou dans une colonne donnée). Par exemple, pour  $Y = y_j$  fixé, la loi conditionnelle de  $X$  est définie par l'ensemble des valeurs possibles et les probabilités associées :

$$P(X = x_i | Y = y_j) = \frac{P(X = x_i, Y = y_j)}{P(Y = y_j)} = \frac{P_{ij}}{P_{.j}} = P_i^j$$

On vérifie que c'est bien une loi de probabilité sur

$$\Omega_X = \{x_i; i \in I\} : \sum_{i \in I} P_i^j = \frac{1}{P_{.j}} \sum_{i \in I} P_{ij} = 1$$

**Exemple 0.2.2** La loi d'un couple  $(X, Y)$  est donnée par le tableau suivant :

$Y \setminus X$	-2	0	2	
-1	0,1	0,2	0,1	0,4
2	0,2	0,2	0,2	0,6
	0,3	0,4	0,3	1

La loi conditionnelle de  $X$  pour  $Y = -1$  figure dans le tableau ci-après :

$X   Y = -1$	-2	0	2	
	$\frac{0,1}{0,4}$	$\frac{0,2}{0,4}$	$\frac{0,1}{0,4}$	1

Rappelons que les deux variables aléatoires  $X$  et  $Y$  sont indépendantes si pour tout  $i \in I$  et  $j \in J$  :

$$P(X = x_i, Y = y_j) = P(X = x_i) P(Y = y_j) \quad (P_{ij} = P_i \cdot P_j)$$

Dans ce cas, bien entendu, les lois conditionnelles sont confondues avec les lois marginales ; par exemples :

$$P(X = x_i | Y = y_j) = P_i^j = \frac{P_i \cdot P_j}{P_j} = P_i.$$

C'est l'un des seuls cas où la donnée des lois marginales permet de reconstruire la loi du couple.

### 0.2.4 Moments conditionnels

Aux lois conditionnelles sont associés des moments, comme par exemple l'espérance de la loi définie par les couples  $\{(y_j, P_j^i; j \in J)\}$ , soit :

$$\mathbb{E}(Y | X = x_i) = \sum_{j \in J} y_j P(Y = y_j | X = x_i) = \sum_{j \in J} y_j P_j^i$$

Le graphe de cette espérance conditionnelle en fonction de  $x_i$  s'appelle courbe de régression (non linéaire) de  $Y$  en  $X$ .

**Exemple 0.2.3** Dans l'exemple précédent, la loi conditionnelle de  $Y$  pour  $X = 2$  est donnée par le tableau suivant :

$Y \setminus X = 2$	-1	2	
	$\frac{0,1}{0,3}$	$\frac{0,2}{0,3}$	1

On peut calculer, à partir de ce tableau, l'espérance conditionnelle :

$$\mathbb{E}(Y|X=2) = (-1)\frac{1}{3} + 2\left(\frac{2}{3}\right) = 1.$$

**Remarque 0.2.2** Notons que  $\mathbb{E}(Y|X)$  est une fonction de  $X$ , donc est une variable aléatoire discrète dont la loi de probabilité est définie par l'ensemble des valeurs possibles, soit ici  $\{\mathbb{E}(Y|X=x_i); i \in I\}$  et les probabilités associées  $P_i = P(X=x_i)$ . On peut donc calculer la valeur moyenne de cette variable aléatoire, soit :

$$\begin{aligned} \mathbb{E}[\mathbb{E}(Y|X)] &= \sum_{i \in I} P_i \mathbb{E}(Y|X=x_i) \\ &= \sum_{i \in I} P_i \sum_{j \in J} y_j P(Y=y_j|X=x_i) \\ &= \sum_{i \in I} P_i \sum_{j \in J} y_j P_j^i \\ &= \sum_{i \in I} \sum_{j \in J} y_j P_i \frac{P_{ij}}{P_i} \\ &= \sum_{j \in J} y_j \sum_{i \in I} P_{ij} \\ &= \sum_{j \in J} y_j P_j = \mathbb{E}(Y). \end{aligned}$$

### 0.2.5 Moments associés à un couple

Si  $h : \mathbb{R}^2 \rightarrow \mathbb{R}$  est une application continue, elle définit une variable aléatoire réelle dont on peut calculer les moments, comme par exemple l'espérance :

$$\mathbb{E}(h(X, Y)) = \sum_{i \in I} \sum_{j \in J} P_{ij} h(x_i, y_j)$$

Dans le cas particulier où  $h(X, Y) = [X - \mathbb{E}(X)][Y - \mathbb{E}(Y)]$ , on définit ainsi la covariance de  $X$  et  $Y$  :

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

Si les variables aléatoires  $X$  et  $Y$  sont indépendantes, alors  $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$  et par conséquent  $Cov(X, Y) = 0$ .

**Exemple 0.2.4** *Considérons le couple  $(X, Y)$  dont la loi est définie par le tableau suivant :*

$Y \backslash X$	-1	0	1
-1	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$
0	$\frac{1}{16}$	$\frac{1}{8}$	$\frac{1}{16}$
1	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$

*Les lois de  $X$  et  $Y$  sont symétriques par rapport à 0, donc  $\mathbb{E}(X) = 0 = \mathbb{E}(Y)$  et :*

$$\text{Cov}(X, Y) = \mathbb{E}(XY) = 1 \times \frac{1}{8} + (-1) \frac{1}{8} + (-1) \frac{1}{8} + 1 \times \frac{1}{8} = 0$$

*et cependant ces deux variables ne sont pas indépendantes puisque par exemple :*

$$P(X = -1, Y = -1) = \frac{1}{8} \neq P(X = -1)P(Y = -1) = \frac{5}{16} \times \frac{3}{8}$$

On appelle coefficient de corrélation linéaire de deux variables aléatoires  $X$  et  $Y$  le nombre réel :

$$\rho = \text{Corr} = \text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{V(X)}\sqrt{V(Y)}}.$$

C'est un nombre tel que :

$$-1 \leq \rho \leq 1, \text{ avec : } |\rho| = 1 \Leftrightarrow \exists a \in \mathbb{R}^*, \exists b \in \mathbb{R} : Y = aX + b.$$

## 0.3 Couple de variables aléatoires continues

### 0.3.1 Loi du couple

Si  $X$  et  $Y$  sont deux variables aléatoires réelles continues, la loi de probabilité du couple  $(X, Y)$  est déterminée par sa fonction de répartition  $F$ , définie sur  $\mathbb{R}^2$  par :

$$F(x, y) = P(X < x, Y < y).$$

L'application  $F$  est croissante au sens large par rapport à chacune des deux variables et telle que  $0 \leq F(x, y) \leq 1$ , avec pour valeurs limites  $F(-\infty, y) = F(x, -\infty) = 0$  pour tout  $x$  réel et pour tout  $y$  réel et  $F(+\infty, +\infty) = 1$ .

### 0.3. COUPLE DE VARIABLES ALÉATOIRES CONTINUES

---

Si  $F$  est deux fois dérivable par rapport aux deux variables, alors la loi de  $(X, Y)$  est dite absolument continue, de densité  $f$  définie par :

$$f(x, y) = \frac{\partial^2 F(x, y)}{\partial x \partial y}$$

**Exemple 0.3.1** La loi de  $(X, Y)$  est définie par la densité :

$$f(x, y) = \begin{cases} \exp(-x) & \text{si } 0 \leq x \leq y \\ 0 & \text{sinon} \end{cases}$$

Calculon la fonction de répartition :

$$\begin{cases} \text{si } x_0 \leq 0 \text{ ou } y_0 \leq 0 & F(x_0, y_0) = 0 \\ \text{Si } 0 \leq x_0 \leq y_0 & F(x_0, y_0) = \int_{-\infty}^{x_0} \int_{-\infty}^{y_0} f(x, y) dx dy \\ & = \int_{x=0}^{x_0} \left( \int_{y=x}^{y=y_0} \exp(-y) dy \right) dx \\ & = 1 - \exp(-x_0) - x_0 \exp(-y_0) \\ \text{En fin si } 0 \leq y_0 \leq x_1 & F(x_1, y_0) = F(y_0, y_0) \\ & = 1 - \exp(-y_0) - y_0 \exp(-y_0). \end{cases}$$

C'est-à-dire :

$$F(x, y) = \begin{cases} 0 & \text{si } x \leq 0 \text{ ou } y \leq 0 \\ 1 - \exp(-x) - x \exp(-y) & \text{si } 0 \leq x \leq y \\ 1 - \exp(-y) - y \exp(-y) & \text{si } 0 \leq y \leq x \end{cases}$$

#### 0.3.2 Lois marginales

Les fonctions de répartitions marginales de  $X$  et  $Y$  sont définies à partir de la fonction de répartition du couple par :

$$\begin{aligned} F_X(x) &= P(X < x) = F(x, +\infty) \\ F_Y(y) &= P(Y < y) = F(+\infty, y) \end{aligned}$$

Cependant, si la loi du couple est définie par sa densité, les densités marginales sont obtenues plutôt par intégration :

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{+\infty} f(x, y) dy \\ f_Y(y) &= \int_{-\infty}^{+\infty} f(x, y) dx \end{aligned}$$

**Exemple 0.3.2** *Si nous reprenons la fonction de répartition de l'exemple précédent, en faisant tendre  $y$  vers plus l'infini, on obtient :*

$$F_X(x) = F(x, +\infty) = 1 - \exp(-x) \text{ pour } x > 0$$

*et par dérivation :*

$$f_X(x) = \exp(-x) \text{ pour } x > 0, \text{ densité de la loi exponentielle ou } \Gamma(1)[X \rightarrow \exp(1)].$$

*Cette densité peut également être obtenue par intégration de la densité du couple.*

### 0.3.3 Lois conditionnelle

Si l'une des deux variables  $X$  ou  $Y$  a une valeur fixée, on peut définir la loi conditionnelle de l'autre variable. Pour des lois absolument continues, les lois conditionnelles sont définies par les densités conditionnelles :

$$\begin{aligned} f_X(x|Y=y) &= \frac{f(x,y)}{f_Y(y)} \text{ si } f_Y(y) > 0 \\ f_Y(y|X=x) &= \frac{f(x,y)}{f_X(x)} \text{ si } f_X(x) > 0 \end{aligned}$$

L'indépendance des variables aléatoires  $X$  et  $Y$  se définit alors par :

$$f(x,y) = f_X(x) f_Y(y) \quad \forall x \in \mathbb{R}, \forall y \in \mathbb{R}$$

On a bien entendu dans ce cas :

$$f_X(x|Y=y) = f_X(x) \text{ et } f_Y(y|X=x) = f_Y(y) \quad \forall x \in \mathbb{R}, \forall y \in \mathbb{R}$$

### 0.3.4 Moments associés à un couple

Si  $h : \mathbb{R}^2 \rightarrow \mathbb{R}$  est une application continue, l'espérance de  $h(x,y)$  se calcule pour une loi de densité  $f$  par l'intégration :

$$\mathbb{E}[h(x,y)] = \int \int_{\mathbb{R}^2} h(x,y) f(x,y) dx dy$$

### 0.3. COUPLE DE VARIABLES ALÉATOIRES CONTINUES

---

Dans le cas particulier où  $h(x, y) = [X - \mathbb{E}(X)][Y - \mathbb{E}(Y)]$ , ceci définit la covariance :

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

Dans le cas particulier où les variables aléatoires  $X$  et  $Y$  sont indépendantes :

$$\mathbb{E}(XY) = \int_{\mathbb{R}^2} \int xyf(x, y) dx dy = \int_{\mathbb{R}} xf(x) dx \int_{\mathbb{R}} yf(y) dy = \mathbb{E}(X)\mathbb{E}(Y)$$

et par conséquent :  $Cov(X, Y) = 0$ .

**Remarque 0.3.1** *La réciproque, généralement fautive, c'est-à-dire que si deux variables aléatoires ont une covariance nulle ne sont pas forcément indépendantes, sauf dans le cas particulier où  $(X, Y)$  est un couple gaussien.*

**Exemple 0.3.3** *Soit  $X$  une variable aléatoire de loi  $N(0, 1)$  et définissons  $Y = X^2$ . On a :*

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

*donc les variables aléatoires  $X$  et  $X^2$  ont une covariance nulle alors qu'elles sont dépendantes (mais non linéairement) puisque la seconde est fonction de la première.*

On peut définir également le coefficient de corrélation linéaire par :

$$\rho = Corr(X, Y) = \frac{Cov(X, Y)}{\sigma_X \sigma_Y}$$

ayant posé :

$$Var(X) = \sigma_X^2 \text{ et } Var(Y) = \sigma_Y^2$$

Nous allons établir que ce coefficient est compris entre  $-1$  et  $+1$ , en raison de l'inégalité de **Schwarz** :

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

que l'on obtient en considérant le polynôme en  $\lambda$ , toujours positif :

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

ce qui implique  $\mathbb{E}^2(XY) - \mathbb{E}(X^2)\mathbb{E}(Y^2) \leq 0$ , soit en appliquant cette inégalité aux variables centrées :

$$|Cov(X, Y)| \leq \sigma_X \sigma_Y$$

et par conséquent  $|\rho| \leq 1$ . Le cas  $|\rho| = 1$  correspond à l'existence d'une relation affine entre les deux variables :

$$|\rho| = 1 \iff \exists a \in \mathbb{R}^*, \exists b \in \mathbb{R}; Y = aX + b.$$

Cette relation affine quand elle existe, s'écrit d'ailleurs précisément :

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

### 0.3.5 Régression

Les densités conditionnelles permettent de calculer les moments conditionnels, comme par exemple les espérances ; on peut définir notamment la fonction :

$$x \mapsto \mathbb{E}(Y | X = x)$$

qui s'appelle fonction de régression (non linéaire), son graphe étant la courbe de régression de  $Y$  en  $X$ . Pour une loi absolument continue, on obtient :

$$\begin{aligned} \mathbb{E}(Y | X = x) &= \int_{-\infty}^{+\infty} y f(Y = y | X = x) dy \\ &= \int_{-\infty}^{+\infty} y \frac{f(x, y)}{f_X(x)} dy \end{aligned}$$

**Exemple 0.3.4** *En reprenant toujours l'exemple de la loi exponentielle on obtient ici :*

$$\begin{aligned} \mathbb{E}(Y | X = x) &= \int_x^{+\infty} y \frac{\exp(-y)}{\exp(-x)} dy = \int_x^{+\infty} y \exp(x - y) dy \\ &= \exp(x) \int_x^{+\infty} y d(-\exp(-y)) \text{ par partie, on obtient donc} \\ &= x + 1. \end{aligned}$$

# Convergences

## 0.4 Convergence en probabilité

La définition de la convergence en probabilité fait intervenir une suite numérique de probabilité dont la convergence sera souvent établie grâce à l'inégalité de **Bienaymé-Tchebychev**, qui lie une probabilité et une variance.

### 0.4.1 Inégalité de Markov

Si  $X$  une variable aléatoire (**v.a**) positive dont l'espérance existe, l'inégalité de Markov établit que pour tout  $\lambda > 0$  :

$$P(X \geq \lambda \mathbb{E}(X)) \leq \frac{1}{\lambda} \text{ ou } P(X \geq \lambda) \leq \frac{\mathbb{E}(X)}{\lambda}$$

### 0.4.2 Inégalité de Bienaymé-Tchebychev

On obtient l'inégalité de Bienaymé-Tchebychev en appliquant l'inégalité de Markov (pour tout  $k$  tel que  $\mathbb{E}(X^k)$  existe :  $P(|X^k| \geq \epsilon) \leq \frac{\mathbb{E}|X|^k}{\epsilon^k}$ ;  $\epsilon > 0$ ) à la **v.a**  $X - \mathbb{E}(X)$  pour  $k = 2$ , donc pour une v.a dont la variance existe, soit pour tout  $\epsilon > 0$  fixé :

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

### 0.4.3 Convergence en probabilité

Si  $(X_n)$  une suite de v.a qui converge vers une v.a  $X$ , cela signifie que  $X_n$  se "rapproche" de  $X$  quand  $n$  augmente. On mesure la distance entre

$X_n$  et  $X$  par  $|X_n - X|$  qui sera d'autant plus petite que  $n$  sera grand ; mais, s'agissant de v.a, il faut considérer l'événement  $|X_n - X| < \epsilon$  qui sera réalisé avec une probabilité d'autant plus élevée que  $n$  sera grand.

On va donc associer à la suite aléatoire  $(X_n)$  la suite numérique des probabilités de ces événements, qui devra converger vers  $un$ .

**Définition 0.4.1** *On dit que la suite de variable aléatoire  $(X_n)$  converge en probabilité vers une variable  $X$  si, pour tout  $\epsilon > 0$  :*

$$P(|X - \mathbb{E}(X)|) < \epsilon \rightarrow 1 \text{ quand } n \rightarrow \infty$$

ou, de façon équivalente :

$$P(|X - \mathbb{E}(X)|) > \epsilon \rightarrow 0 \text{ quand } n \rightarrow \infty$$

on écrit :  $X_n \xrightarrow{P} X$ .

**Exemple 0.4.1** *Soit  $(U_n)$  une suite de variables aléatoires indépendantes v.a.i, de même loi uniforme sur  $[0, 1]$  ; on lui associer la suite  $(X_n)$  définie pour tout entier  $n$  par*

$$X_n = \min \{U_1, \dots, U_n\}. \text{ Nous allons établir que : } X_n \xrightarrow{P} 0 \text{ quand } n \rightarrow \infty.$$

*Pour cela, nous allons calculer, pour un  $\epsilon > 0$  fixé, la probabilité  $P(|X_n|) > \epsilon = P(X_n > \epsilon)$  (puisque  $X_n$  est une v.a positive). Pour  $\epsilon \geq 1$  :  $P(X_n > \epsilon) = 0$ . Pour  $\epsilon < 1$ . Toutes les v.a  $U_n$  étant indépendantes, de même loi que  $U$  avec  $P(U < x) = x$  pour  $0 \leq x \leq 1$ , on obtient :*

$$\begin{aligned} P(|X_n|) > \epsilon &= P\left[\bigcap_{i=1}^n (U_i > \epsilon)\right] = \prod_{i=1}^n P(U_i > \epsilon) = (1 - P(U < \epsilon))^n \\ &= (1 - \epsilon)^n \xrightarrow{n \rightarrow \infty} 0, \text{ d'où } X_n \xrightarrow{P} 0. \end{aligned}$$

## 0.5 Convergence presque-sûre et loi des grands nombres

La convergence presque-sûre est la convergence simple d'une suite de fonction de  $\Omega$  dans  $\mathbb{R}$ , c'est-à-dire la convergence pour chaque éventualité  $\omega$ .

0.5. CONVERGENCE PRESQUE-SÛRE ET LOI DES GRANDS  
NOMBRES

**Définition 0.5.1** On dit que la suite de variables aléatoires  $(X_n)$  converge presque sûrement vers la variable aléatoire  $X$  quand  $n$  tend vers l'infini si :

$$P(\{\omega \in \Omega, X_n(\omega) \rightarrow X(\omega)\}) = 1$$

on note :

$$X_n \xrightarrow[n \rightarrow \infty]{P.S} X \left[ \lim_{n \rightarrow \infty} X_n(\omega) = X(\omega) \quad \forall \omega \in A; P(A) = 1 \right]$$

Le théorème suivant établit des propriétés équivalentes à la définition de la convergence presque-sûre, et souvent d'utilisation préférable.

**Théorème 0.5.1**  $(X_n)$  et  $(X)$  étant des variables aléatoires réelles, les quatre propriétés suivantes sont équivalentes :

$$\begin{aligned} a) & X_n \xrightarrow{PS} X \\ b) & \forall \epsilon > 0, \lim_n P \left[ \sup_{m \geq n} \{|X_m - X| > \epsilon\} \right] = 0 \\ c) & \forall \epsilon > 0, \lim_n P \left[ \bigcup_{m \geq n} \{|X_m - X| > \epsilon\} \right] = 0 \\ d) & \forall \epsilon > 0, \lim_n P \left[ \bigcap_{m \geq n} \{|X_m - X| \leq \epsilon\} \right] = 1 \end{aligned}$$

Pour la démonstration, on démontre que :

$$\begin{aligned} 1) & c \Leftrightarrow d \text{ (par passage au complémentaire)} \\ 2) & b \Leftrightarrow c \text{ (on remarque que } \omega \in \left\{ \sup_{m \geq n} \{|X_m - X| > \epsilon\} \right\} \\ & \Leftrightarrow \sup_{m \geq n} \{|X_m(\omega) - X(\omega)| > \epsilon\} \\ & \Leftrightarrow \exists m \geq n; |X_m(\omega) - X(\omega)| > \epsilon \Leftrightarrow \omega \in \bigcup_{m \geq n} \{|X_m - X| > \epsilon\} \\ 3) & a \Leftrightarrow d \end{aligned}$$

**Remarque 0.5.1** Supposons que la série de terme général  $U_n = P\{|X_n - X| > \epsilon\}$  soit convergente, i.e :

$$\sum_{n=1}^{\infty} P\{|X_n - X| > \epsilon\} < \infty,$$

pour tout  $\epsilon > 0$ . Alors :

$$P \left[ \bigcup_{m \geq n} \{|X_m - X| > \epsilon\} \right] \leq \underbrace{\sum_{m=n}^{\infty} P \{|X_m - X| > \epsilon\}}_R$$

$R$  est le reste d'une série convergente ; on en déduit :

$$\lim_n P \left[ \bigcup_{m \geq n} \{|X_m - X| > \epsilon\} \right] = 0$$

c'est une condition suffisante de convergence presque-sûre de  $X_n$  vers  $X$ , que l'on peut utiliser comme définition d'une convergence plus forte que la convergence presque-sûre.

A titre d'illustration de cette convergence, nous pouvons énoncer le théorème suivant :

**Théorème 0.5.2 (Loi des grands nombres :LGN)** Soient  $(X_n)_{n=1}^{\infty}$  une suite de v.a indépendantes et identiquement distribuée (i.i.d) et  $S_n = X_1 + \dots + X_n$ . Si  $\mathbb{E}(|X_1|) < \infty$ , alors :

$$\begin{aligned} \text{a) Loi faible} & : \frac{S_n}{n} \xrightarrow[n \rightarrow \infty]{P} \mathbb{E}(X_1) \\ \text{b) Loi forte} & : \frac{S_n}{n} \xrightarrow[n \rightarrow \infty]{PS} \mathbb{E}(X_1) \end{aligned}$$

**Exemple 0.5.1** Soit  $(X_n)$  une suite de pile (1) ou face (0) indépendants et équilibrés. Du fait que  $\mathbb{E}(X_1) = \frac{1}{2} < \infty$ , on a par la loi forte des grands nombres :  $\frac{S_n}{n} \rightarrow \frac{1}{2}$  P.S, i.e :  $S_n \simeq \frac{n}{2}$ .

**Question :** Quelle est la déviation moyenne de  $S_n - \frac{n}{2}$ .

On a :  $\mathbb{E}(X_1) = \frac{1}{2}$  et  $Var(X_1) = \frac{1}{4}$ , donc donc théorème central limite **TCL** donne :

$$\frac{S_n - \frac{n}{2}}{\sqrt{n} \frac{1}{2}} \rightarrow Z \rightarrow N(0, 1), \text{ i.e : } S_n \simeq \frac{n}{2} + \frac{\sqrt{n}}{2} Z$$

donc la déviation est d'ordre  $\sqrt{n}$ .

## 0.6 Convergence en loi

La convergence en loi est une convergence des lois des variables aléatoires, sans prendre en compte le comportement de la suite pour une éventualité fixée :

**Définition 0.6.1** *On dit que la suite de variables aléatoires  $(X_n)_{n \geq 1}$  converge en loi vers la variable aléatoire  $X$  quand  $n$  tend vers l'infini, et on écrit  $X_n \xrightarrow[n \rightarrow \infty]{\text{loi}} X$ , si pour toute fonction  $f : \mathbb{R} \rightarrow \mathbb{R}$  continue bornée ;*

$$\mathbb{E}(f(X_n)) \xrightarrow[n \rightarrow \infty]{} \mathbb{E}(f(X))$$

*cette définition n'est pas forcément la plus simple à manipuler, et il existe beaucoup de définitions équivalentes de cette convergence.*

### Propriétés

La convergence en loi de  $X_n$  vers  $X$  est équivalente à chacune des propriétés suivantes :

1. Convergence simple des fonctions caractéristiques :

$$\varphi_{X_n}(t) \xrightarrow[n \rightarrow \infty]{} \varphi_X(t) \quad \forall t \in \mathbb{R}$$

2. Convergence simple des fonctions de répartition aux points de continuité :

$$F_{X_n}(x) \rightarrow F_X(x)$$

pour tout  $x$  tel que  $F_X$  est continue en  $x$ .

3.  $P(X_n \in A) \xrightarrow[n \rightarrow \infty]{} P(X \in A) \quad \forall A \in C(X)$  où  $C(X)$  est l'ensemble des sous-ensembles  $A$  de  $\mathbb{R}$  qui vérifient  $P(X \in \partial A) = 0$  ( $\partial A$  désigne la frontière de  $A$ ). Le lemme suivant découle immédiatement de la définition de la convergence en loi :

**Lemme 0.6.1** *Si la suite  $(X_n)_{n \geq 1}$  converge en loi vers  $X$  et si  $f : \mathbb{R} \rightarrow \mathbb{R}$  est une fonction continue, alors  $f(X_n)$  converge en loi vers  $f(X)$ .*

**Théorème 0.6.1** *Soit  $(X_n)_{n \geq 1}$  une suite de variables aléatoires réelles indépendantes et de même loi, de moyenne  $m$  et de variance  $\sigma^2$  finies, soit :  $S_n = X_1 + \dots + X_n$ . Alors :*

$$\frac{S_n - nm}{\sigma\sqrt{n}} \xrightarrow[n \rightarrow \infty]{\text{loi}} S \xrightarrow[n \rightarrow \infty]{\text{loi}} N(0, 1)$$

*i.e :*

$$P\left(\frac{S_n - nm}{\sigma\sqrt{n}} \leq t\right) \rightarrow \int_{-\infty}^t \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right) dx = \phi(t) \quad \forall t \in \mathbb{R}$$

**NB :** La fonction de répartition  $\phi$  de la loi normale  $N(0, 1)$  sur tout  $\mathbb{R}$ .

# Statistique inférentielle

## 0.7 Echantillonnage-Estimation

### I-Introduction

L'étude des caractéristiques de tous éléments d'une population est souvent impossible à réaliser en raison de contraintes de coûts et de temps.

Cette impossibilité conduit à étudier un sous-ensemble issu de la population mère : L'échantillon.

L'échantillonnage consiste à déduire de la connaissance supposée connue des caractéristiques d'une population, les caractéristiques des échantillons prélevés dans cette population.

L'estimation est le problème inverse. Il s'agit d'estimer, à partir des caractéristiques calculées sur un ou plusieurs échantillons, la valeur des caractéristiques de la population mère.

	Population mère $\Omega$	Echantillons $E_i$
<b>Effectif ou taille</b>	$N$	$n$
<b>Moyenne</b>	$m$	$\bar{X}_i$
<b>Fréquence ou proportion</b>	$P$	$f_i$
<b>Variance</b>	$\sigma^2$	$\sigma_i^2$
<b>Ecart-type</b>	$\sigma$	$\sigma_i$

Les méthodes probabilistes de constitutions des échantillons consistent à prélever au hasard des éléments de la population et sont les seuls à respecter les lois statistiques. le prélèvements des éléments de l'échantillon peut être effectué.

### Avec remise :

L'élément prélevé est immédiatement remis dans la population mère avant de prélevé le suivant. Un élément pouvant être éventuellement prélevé plusieurs fois, les tirages sont indépendants et l'échantillon est dit **non-exhaustif**.

**Sans remise :**

L'échantillon est exhaustif, mais les tirages ne sont pas indépendants puisque la composition de la population mère est modifiée à chaque tirage.

Dans la suite, pour pouvoir appliquer les règles du calcul des probabilités, les échantillons seront supposés être constitués avec remise, ou être des échantillons sans remise dont la taille est négligable par rapport à celle de la population qui est grande taille ou infinie (le tirage est alors assimilable à un tirage avec remise).

Indiquons la procédure à mettre en oeuvre pour constituer un échantillon à l'aide d'une table de nombre aléatoires.

Cette table est constituée des nombres 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 et chacune des valeurs entières a la même probabilité d'apparition.

N'importe lequel nombre de la table n'a aucune relation avec le nombre au-dessus, en-dessous, à la droite ou à la gauche de lui.

les nombres sont éparpillés au hasard dans la table que nous utilisons, les nombres sont regroupés en colonnes de 5 chiffres, chaque ligne comporte 50 nombres (10 groupes de 5). Pour choisir des nombres de la table, il s'agit simplement :

a) De choisir un point d'entrée dans la table.

b) De choisir itinéraire de lecture. On peut lire les nombres en ligne ( de gauche à droite ou de bas en haut).

On pourrait également sauter un nombre sur deux, etc...

**Exemple 0.7.1** *Tirage au sort d'un échantillon de taille  $n = 10$  à l'aide de la table de nombres aléatoires. Le délégué d'une association des étudiants d'une université veut tirer au sort un échantillon de 10 individus faisant partie de l'association. Supposons que l'AEU comporte 300 individus listés sur un fichier. On a donc ce qui suit :*

Population : Les individus membres de l'association. Base de sondage : La liste des noms des individus sur le fichier. Unité statistique : Les individus. Taille de la population :  $N = 300$ . Taille requise de l'échantillon :  $N = 10$ . Mode de tirage : Sans remise (tirage exhaustif). On débute par numéroter chaque individu dans la base de sondage de 001 à 300. Puisque la base de sondage comporte 300 individus, nous allons choisir de la table des nombres de 3 chiffres. Pour lire dans la table, nous proposons la règle suivante : Partir de la 3<sup>ème</sup> lignes en ne considérant que les 3 derniers chiffres de la 4<sup>ème</sup> colonne (et des suivantes s'il ya lieu) avec lecture de haut en bas,

ne retenir que les résultats de lecture qui soient compris entre 001 et 300. Puisqu'on effectue un tirage sans remise, on rejette tout nombre déjà sorti qui apparait à nouveau dans la procédure de sélection. On obtient alors les 10 nombres suivants : Les individus portant les numéros suivants dans la base de sondage vont constituer l'échantillon de taille  $n = 10$ .

251 045 075 157 199  
267 278 026 238 051

### 0.7.1 Distribution d'échantillonnage

Soit dans une population mère  $\Omega$  de taille  $N$ , une variable aléatoire  $X$  pour laquelle l'espérance mathématique  $m$ , la proportion  $P$  et l'écart-type sont connues.

De cette population sont issus  $k$  échantillons  $E_1, E_2, \dots, E_k$  de taille  $n$  qui auront des moyennes et des écarts-types différents. La notion de distribution d'échantillonnage peut être résumé et schématisée :

Population mère : $\Omega$	Echantillon 1	Echantillon 2	Echantillon k
Taille : $N$	Taille : $n$	Taille : $n$	Taille : $n$
Moyenne : $m$ ( <i>connue</i> )	Moyenne : $\bar{X}_1$	Moyenne : $\bar{X}_2$	Moyenne : $\bar{X}_k$
proportion : $P$ ( <i>connue</i> )	proportion : $f_1$	proportion : $f_2$	proportion : $f_k$
Écart-type $\sigma$ ( <i>connue</i> )	Écart-type : $\sigma_1$	Écart-type : $\sigma_2$	Écart-type : $\sigma_k$

Déduire les caractéristiques d'un échantillon de la connaissance des caractéristiques de la population mère

### 0.7.2 Distribution d'échantillonnage des moyennes

Les moyennes  $\bar{X}_i$  de chaque échantillon varient d'un échantillon à l'autre et représentent la distribution des moyennes de la variable aléatoire  $\bar{X}_i$  qui associe à tout échantillon de taille  $n$  la moyenne de cet échantillon.

La variable aléatoire  $\bar{X}_n$  prend donc les valeurs :  $\bar{X}_1, \bar{X}_2, \dots, \bar{X}_k$ .

### 0.7.3 Valeurs caractéristiques de $\bar{X}_n$

· L'espérance mathématique de la variable aléatoire  $\bar{X}_n$  est égale à celle de la population mère :

$$\mathbb{E}(\bar{X}_n) = m$$

· La variance de la variable aléatoire  $\bar{X}_n$  est égale à celle de la population mère rapportée à la taille de l'échantillon :

$$V(\bar{X}_n) = \frac{\sigma^2}{n}$$

· L'écart-type de la variable aléatoire  $\bar{X}_n$  se déduit de la variance :

$$\sigma(\bar{X}_n) = \frac{\sigma}{\sqrt{n}}$$

**Remarque 0.7.1** *Si les échantillons sont issus d'une population mère finie et sont constituée sans remise. L'espérance mathématique de  $\bar{X}_n$  est toujours égale à  $m$ , mais l'écart-type est corrigé par le facteur d'exhaustivité (facteur de correction)*

$$\sigma(\bar{X}_n) = \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}} \approx \frac{\sigma}{\sqrt{n}} \sqrt{1 - \frac{n}{N}} \text{ tel que } \frac{n}{N} \text{ représente le taux de sondage.}$$

**Justification mathématique : Espérance mathématique et variance de  $\bar{X}$ , la moyenne d'échantillon**

On prélève au hasard un échantillon de taille  $n$  (tirage avec remise) dont les éléments possèdent un caractère mesurable  $X$  suivant une distribution de probabilité de moyenne  $\mathbb{E}(X) = m$  et de variance  $Var(X) = \sigma^2$ .

En prélevant au hasard un échantillon de taille  $n$  de cette population, on crée une suite de  $n$  variables aléatoires indépendantes  $X_1, X_2, \dots, X_n$  dont chacune a la même distribution que  $X$ .

a) La moyenne d'échantillon  $\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n}$  est une variable aléatoire dont l'espérance mathématique est  $\mathbb{E}(\bar{X}) = m$ .

**Justification**

$$\begin{aligned} \mathbb{E}(\bar{X}) &= \mathbb{E}\left(\frac{X_1 + X_2 + \dots + X_n}{n}\right) = \frac{1}{n} \mathbb{E}(X_1 + X_2 + \dots + X_n) \\ &= \frac{1}{n} [\mathbb{E}(X_1) + \mathbb{E}(X_2) + \dots + \mathbb{E}(X_n)] = \frac{n}{n} m. \end{aligned}$$

b) La variance de  $\bar{X}$  est égale à la variance  $\sigma^2$  de la population divisée par  $n$ , la taille de l'échantillon :

$$Var(\bar{X}) = \sigma^2(\bar{X}) = \mathbb{E}(\bar{X} - m)^2 = \frac{\sigma^2}{n}$$

**Justification**

$$\text{Var}(\bar{X}) = \text{Var}\left(\frac{X_1 + X_2 + \dots + X_n}{n}\right) = \frac{1}{n^2} [\text{Var}(X_1 + X_2 + \dots + X_n)]$$

Puisque  $X_1, X_2, \dots, X_n$  sont indépendantes, on peut écrire :

$$\begin{aligned}\text{Var}(\bar{X}) &= \frac{1}{n^2} [\text{Var}(X_1) + \text{Var}(X_2) + \dots + \text{Var}(X_n)] \\ &= \frac{1}{n^2} (\sigma^2 + \sigma^2 + \dots + \sigma^2) \\ &= \frac{1}{n^2} n\sigma^2 = \frac{\sigma^2}{n}.\end{aligned}$$

**0.7.4 Distribution d'échantillonnage des fréquences**

La probabilité de réalisation d'un événement  $A$  est égale à  $P$ . On considère les échantillons aléatoires de tailles  $n$  extraits d'une population de taille  $N$ . Pour chaque échantillon on détermine la proportion  $f$  de réalisation de l'événement  $A$ .

La population et les échantillons suivent des lois binomiales  $B(N, P)$  et  $B(n, f)$  respectivement.

La moyenne  $m_f$  et l'écart-type  $\sigma_f$  de la distribution d'échantillonnage des fréquences valent :

$$\begin{aligned}m_f &= P \text{ et } \sigma_f = \sqrt{\frac{P(1-P)}{n}} \text{ si le tirage est non exhaustif ou} \\ &\quad \text{si la population est infinie.} \\ \sigma_f &= \sqrt{\frac{P(1-P)}{n}} \sqrt{\frac{N-n}{N-1}} \text{ si le tirage est exhaustif.}\end{aligned}$$

\* Si la taille  $n$  des échantillons est assez grande (en pratique  $n \geq 30$ ) la distribution d'échantillonnage de la moyenne approche la distribution normale quelle que soit la distribution de la population, c'est-à-dire :  $\bar{X} \rightarrow N\left(m, \frac{\sigma}{\sqrt{n}}\right)$ .

\* Si la population est normalement distribuée, la distribution d'échantillonnage de la moyenne est une loi normale quelle que soit la valeur  $n$  de la taille des échantillons.

\* Si la population parente possède une distribution pratiquement symétrique, il semble qu'un échantillon d'au moins **15** observations soit convenable pour que la distribution de la moyenne soit approximativement normale.

## 0.8 Estimation de Paramètres

### 1. Introduction

Les variables aléatoires  $X_1, X_2, \dots, X_n$  sont définies par leurs fonction de répartition :

$$F(x_1, x_2, \dots, x_n) = P(X < x_1, X_2 < x_2, \dots, X_n < x_n)$$

dependant d'un paramètre  $\theta$  appartenant à un ensemble  $\Theta$  inclus dans  $\mathbb{R}$ . Le problème de l'estimation est celui de la mesure de  $\theta$  à partir de l'observation des variables aléatoires  $(X_1, X_2, \dots, X_n)$ .

Un estimateur de  $\theta$  est une variable aléatoire, notée  $\hat{\theta} = \hat{\theta}(X_1, X_2, \dots, X_n)$ , fonction suffisamment régulière des variables aléatoires  $X_1, X_2, \dots, X_n$  et uniquement de ces variables aléatoires (c'est-à-dire que  $\theta$  ne doit pas intervenir dans  $\hat{\theta}$  explicitement).

On demande à  $\hat{\theta}$  les deux qualités suivantes :

$$\left\{ \begin{array}{l} 1) \text{ absence de biais : } \mathbb{E}(\hat{\theta}) = \theta \text{ (biais} = b(\theta) = \mathbb{E}(\hat{\theta}) - \theta) \\ 2) \text{ Convergence : } \forall \varepsilon > 0; \lim_{n \rightarrow \infty} P\left[|\hat{\theta} - \theta| > \varepsilon\right] = 0 \end{array} \right.$$

La première propriété est à rapprocher de l'absence d'erreur systématique tandis que la seconde est la convergence en probabilité.

**Définition 0.8.1** *Un estimateur  $\hat{\theta}$  asymptotiquement sans biais*

$\lim_{n \rightarrow \infty} \mathbb{E}(\hat{\theta}) = \theta$  et dont la variance vérifie  $\lim_{n \rightarrow \infty} \text{Var}(\hat{\theta}) = 0$  est convergent.

**Théorème 0.8.1** *Tout estimateur sans biais dont la variance tend vers zéro est convergent.*

**Preuve.** Ce résultat se déduit directement de l'inégalité de Bienaymé-Tchebychev :

$$P_{\theta}\left[|\hat{\theta}_n - \theta| > \varepsilon\right] \leq \frac{\text{Var}_{\theta}(\hat{\theta})}{\varepsilon^2} \rightarrow 0 \text{ quand } n \rightarrow \infty$$

■

**Exemple 0.8.1** *Estimation de la moyenne*

**Exemple 0.8.2** Lorsque  $X_1, X_2, \dots, X_n$  sont des variables aléatoires indépendantes (v.a.i), de même moyenne  $m$  et de même variance  $\sigma^2$ , l'estimateur de la moyenne,  $\hat{m} = \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$  est tel que,

$$\begin{cases} \mathbb{E}(\bar{X}) = m \\ \text{Var}(\bar{X}) = \frac{\sigma^2}{n} \rightarrow 0 \text{ quand } n \rightarrow \infty \end{cases}$$

Donc  $\hat{m} = \bar{X}$  est un estimateur sans biais et convergent de  $m$ .

**Exemple 0.8.3**  $S^2 = \sigma_{\text{échantillon}}^2 = \sigma_{\text{écha}}^2$  est un estimateur sans biais de  $\sigma^2$ .

$$\text{Car : } S^2 = \frac{\sum_i (X_i - \bar{X})^2}{n - 1}; \quad \mathbb{E}(S^2) = \sigma^2$$

On peut écrire

$$(n - 1)S^2 = \sum_i (X_i - \bar{X})^2 = \sum_i [(X_i - m) - (\bar{X} - m)]^2$$

qui donne, en développement le membre de droite :

$$(n - 1)S^2 = \sum_i (X_i - m)^2 - n(\bar{X} - m)^2$$

### 0.8.1 Estimateur optimal (efficace)

### 0.8.2 Qualité d'un estimateur

La qualité d'un estimateur va se mesurer à l'aide d'une distance au paramètre qui peut être par exemple  $|\hat{\theta}_n - \theta|$  ou  $(\hat{\theta}_n - \theta)^2$ . Pour obtenir un indicateur numérique on peut alors déterminer la valeur moyenne de cette distance.

L'indicateur généralement retenu, car il se prête facilement aux calculs, est l'erreur quadratique moyenne définie pour tout  $\theta$  par

$$E_Q(\hat{\theta}_n) = \mathbb{E}_\theta (\hat{\theta}_n - \theta)^2 = \text{Var}(\hat{\theta}_n) + b_n^2(\theta)$$

Dans le cas particulier d'un estimateur sans biais, cette erreur quadratique se confond avec la variance de l'estimateur.

### 0.8.3 Estimateur efficace

Un estimateur sans biais est plus efficace (ou simplement efficace) si sa variance est la plus faible parmi les variances des autres estimateurs sans biais. Ainsi si  $\hat{\theta}_1$  et  $\hat{\theta}_2$  sont deux estimateurs sans biais du paramètre  $\theta$ , l'estimateur  $\hat{\theta}_1$  est plus efficace si

$$\text{Var}(\hat{\theta}_1) < \text{Var}(\hat{\theta}_2) \text{ et } \mathbb{E}(\hat{\theta}_1) = \mathbb{E}(\hat{\theta}_2) = \theta$$

La notion d'estimateur efficace peut s'illustrer de la façon suivantes.

La distribution de  $\hat{\theta}_1$  est plus concentrée autour de  $\theta$  que celle de  $\hat{\theta}_2$ . Si nous échantillonons une population normale  $\bar{X}$  et  $M_e$  sont des estimateurs sans biais de  $m$  car :

$$\mathbb{E}(\bar{X}) = \mathbb{E}(M_e) = m$$

D'autre part, la variance de  $\bar{X}$  est plus petite que celle de la médiane puisque

$$\text{Var}(\bar{X}) = \frac{\sigma^2}{n} \text{ et } \text{Var}(M_e) = 1,57 \frac{\sigma^2}{n}$$

Pour la même taille d'échantillon,  $\bar{X}$  est plus efficace que  $M_e$  pour estimer  $m$ ;  $\text{Var}(\bar{X}) < \text{Var}(M_e)$ .

**Comparaison de deux estimateurs de  $m$ .**

**Exemple 0.8.4** Soit  $X_1, X_2, X_3$  un échantillon aléatoire prélevé d'une population infinie avec  $\mathbb{E}(X_i) = m$  et  $\text{Var}(X_i) = \sigma^2$ . Montrer que

$\hat{m} = \frac{X_1 + 2X_2 + 3X_3}{6}$  est un estimateur sans biais de  $m$  mais qu'il est moins efficace que

$$\bar{X} = \frac{\sum_i X_i}{3}$$

C'est-à-dire que

$$\mathbb{E}(\hat{m}) = m \text{ et } \text{Var}(\hat{m}) > \text{Var}(\bar{X})$$

Vérifions que  $\hat{m}$  est un estimateur sans biais de  $m$ .

$$\mathbb{E}(\hat{m}) = \mathbb{E}\left(\frac{X_1 + 2X_2 + 3X_3}{6}\right) = \frac{1}{6} [\mathbb{E}(X_1) + 2\mathbb{E}(X_2) + 3\mathbb{E}(X_3)] = \frac{6m}{6} = m$$

Déterminons maintenant  $Var(\hat{m})$ .

$$Var(\hat{m}) = \frac{14}{36}\sigma^2$$

Or

$$Var(\bar{X}) = \frac{\sigma^2}{n} = \frac{\sigma^2}{3}, \text{ d'où } 0,333\sigma^2 < \frac{14}{36}\sigma^2.$$

Par conséquent  $\bar{X}$  est plus efficace que  $\hat{m}$  pour estimer  $m$ .

#### 0.8.4 Inégalité de Fréchet-Darmois-Cramer-Rao(F.D.C.R)

Nous allons voir dans certaines conditions il existe une borne inférieure pour l'ensemble des variances des estimateurs sans biais, ce qui va constituer un butoir ne permettant pas d'améliorer sans cesse les estimateurs. D'autre part, si cette borne est atteinte par un estimateur, il deviendra le meilleur et sera qualifié d'optimal dans la classe des estimateurs sans biais.

**Définition 0.8.2** On appelle vraisemblance de l'échantillon  $(X_1, X_2, \dots, X_n)$  la loi de probabilité de ce  $n$ -uple, notée  $L(X_1, X_2, \dots, X_n; \theta)$  et, définie par

$L(x_1, x_2, \dots, x_n; \theta) = \prod_{i=1}^n P(X_i = x_i; \theta)$  si  $X$  est une variable aléatoire discrète et par

$L(x_1, x_2, \dots, x_n; \theta) = \prod_{i=1}^n f(x_i; \theta)$  si est une variable aléatoire continue de densité  $f(x, \theta)$

Le théorème suivant précise la borne inférieure pour la variance des estimateurs sans biais.

**Théorème 0.8.2** Sous les hypothèses de Cramer-Rao, en particulier si  $E = X(\Omega)$  est indépendant du paramètre à estimer  $\theta$ , Pour tout estimateur sans biais  $\hat{\theta}$  de  $\theta$  on a :

$$Var_{\theta}(\hat{\theta}) \geq \frac{1}{I_n(\theta)} = B_F(\theta)$$

Tel que  $I_n(\theta)$  est la quantité d'information de Fisher qui est définie par :

$$I_n(\theta) = \mathbb{E}\left(-\frac{\partial^2 \ln l}{\partial \theta^2}\right)$$

et  $B_F(\theta)$  est la borne inférieure de **F.D.C.R**

**Remarque 0.8.1** Si  $E = X(\Omega)$  dépend du paramètre à estimer  $\theta$ , on obtient  $I_n(\theta) = \mathbb{E} \left( \frac{\partial \ln l}{\partial \theta} \right)^2$ .

**Exemple 0.8.5** Soit  $X$  une variable aléatoire de loi exponentielle de paramètre  $\frac{1}{\theta}$ , ou loi Gamma notée  $\Gamma \left( 1, \frac{1}{\theta} \right)$ , avec  $\theta > 0$  de densité pour  $x > 0$  :

$$f(x, \theta) = \frac{1}{\theta} \exp \left( -\frac{x}{\theta} \right)$$

La vraisemblance admet ici pour expression :

$$L(x_1, x_2, \dots, x_n; \theta) = \prod_{i=1}^n f(x_i; \theta) = \frac{1}{\theta^n} \exp \left( -\frac{1}{\theta} \sum_{i=1}^n x_i \right)$$

Pour calculer la quantité d'information de Fisher nous écrivons

$$\ln L(x_1, x_2, \dots, x_n; \theta) = -n \ln \theta - \frac{1}{\theta} \sum_{i=1}^n x_i.$$

$$\frac{\partial \ln l}{\partial \theta} = -\frac{n}{\theta} + \frac{1}{\theta^2} \sum_{i=1}^n x_i$$

comme  $X(\Omega) = \mathbb{R}_+$  est indépendant de  $\theta$  on a :

$$\frac{\partial^2 \ln l}{\partial \theta^2} = \frac{n}{\theta^2} - \frac{2}{\theta^3} S_n \text{ tq } S_n = \sum_{i=1}^n x_i$$

D'où :

$$I_n(\theta) = \mathbb{E} \left( -\frac{\partial^2 \ln l}{\partial \theta^2} \right) = \frac{n}{\theta^2} + \frac{2n\theta}{\theta^3} = \frac{n}{\theta^2}.$$

**Exemple 0.8.6** Si nous prenons maintenant l'exemple de la loi exponentielle sur  $[\theta, +\infty[$ , de densité :

$$f(x, \theta) = \begin{cases} \exp(-(x - \theta)); & \text{si } x \geq \theta \\ 0; & \text{sinon.} \end{cases}$$

La vraisemblance s'écrit :

$$L(x_1, x_2, \dots, x_n; \theta) = \exp \left( -\sum_{i=1}^n (x_i - \theta) \right)$$

si tous les  $x_i$  sont plus grands que  $\theta$ , c'est-à-dire si

$$\min \{x_i/1 \leq i \leq n\} \geq \theta$$

on a alors :

$$\ln L(x_1, x_2, \dots, x_n; \theta) = - \sum_{i=1}^n x_i + n\theta.$$

d'où

$$\frac{\partial \ln l}{\partial \theta} = n \text{ et } I_n(\theta) = \mathbb{E} \left( \frac{\partial \ln l}{\partial \theta} \right)^2 = n^2.$$

### 0.8.5 Estimateur efficace

**Définition 0.8.3** Un estimateur sans biais  $\hat{\theta}$  est dit efficace si sa variance est égale à la borne inférieure de (F.D.C.R) :

$$\text{Var}(\hat{\theta}) = \frac{1}{I_n(\theta)}$$

**Exemple 0.8.7** Si nous reprenons l'exemple de la loi exponentielle de paramètre  $\frac{1}{\theta}$ , comme  $\mathbb{E}_\theta(X) = \theta$ , on sait que  $\hat{\theta}_n = \bar{X}$  est un estimateur sans biais est convergent. De plus :

$$\text{Var}_\theta(\hat{\theta}_n) = \text{Var}_\theta(\bar{X}_n) = \frac{\text{Var}_\theta(X)}{n} = \frac{\theta^2}{n} = \frac{1}{I_n(\theta)}$$

donc cet estimateur est aussi efficace.

### 0.8.6 Méthodes de construction d'un estimateur

#### 0.8.7 Méthode du maximum de vraisemblance

La vraisemblance  $L(x_1, x_2, \dots, x_n; \theta)$  représente la probabilité d'observer le  $n$ -uple  $(x_1, \dots, x_n)$  pour une valeur fixée de  $\theta$ . Dans la situation inverse ici où on a observé  $(x_1, \dots, x_n)$  sans connaître la valeur de  $\theta$ , on va attribuer à  $\theta$  la valeur qui paraît la plus vraisemblable, compte tenu de l'observation dont on dispose, c'est-à-dire celle qui va lui attribuer la plus forte probabilité. On se fixe donc la règle suivante : à  $(x_1, \dots, x_n)$  fixé on considère la vraisemblance  $L$  comme une fonction de  $\theta$  et on attribue à  $\theta$  la valeur qui maximise cette fonction. D'où la définition suivante :

**Définition 0.8.4** On appelle estimateur du maximum de vraisemblance (e.m.v) toute fonction  $\hat{\theta}$  de  $(x_1, \dots, x_n)$  qui vérifie :

$$L(x_1, x_2, \dots, x_n; \hat{\theta}) = \max_{\theta \in \Theta} L(x_1, x_2, \dots, x_n; \theta)$$

Cette définition ne renseigne en aucune façon, ni sur l'existence, ni sur l'unicité, d'un tel estimateur. La recherche d'e.m.v peut se faire directement par recherche du maximum de  $L$ , ou dans le cas particulier où la fonction  $L$  est deux fois dérivable par rapport à  $\theta$ , comme solution de l'équation  $\frac{\partial \ln l}{\partial \theta} = 0$  qui vérifie aussi  $\frac{\partial^2 \ln l}{\partial \theta^2} < 0$ .

**Exemple 0.8.8** Cherchons l'emv pour la famille de lois exponentielle de paramètre  $\frac{1}{\theta}$ . La log-vraisemblance est indéfiniment dérivable pour  $\theta > 0$  et nous avons obtenu dans l'exemple 1

$$\frac{\partial \ln l}{\partial \theta} = -\frac{n}{\theta} + \frac{1}{\theta^2} \sum_{i=1}^n x_i = 0$$

$$\Rightarrow \theta = \frac{1}{n} \sum_{i=1}^n x_i = \bar{X}_n \text{ avec :}$$

$$\frac{\partial^2 \ln l}{\partial \theta^2} = \frac{n}{\theta^2} - \frac{2}{\theta^3} \sum_{i=1}^n x_i = \frac{n}{\theta^3} (\theta - 2\bar{X}_n) \text{ soit pour } \theta = \bar{X}_n$$

$$\left( \frac{\partial^2 \ln l}{\partial \theta^2} \right)_{\theta = \bar{X}_n} = -\frac{n}{\bar{X}_n^2} < 0, \text{ donc l'emv est } \theta = \bar{X}_n.$$

### 0.8.8 Estimation d'une moyenne par intervalle de confiance

On se propose d'estimer, par intervalle de confiance, la moyenne  $m$  d'un caractère mesurable d'une population. Il s'agit de calculer, à partir de la moyenne  $\bar{X}$  de l'échantillon, un intervalle dans lequel il est vraisemblable que la vraie valeur de  $m$  s'y trouve. On obtient cet intervalle en calculant deux limites auxquelles est associée une certaine assurance de contenir la vraie valeur de  $m$ . Cet intervalle se définit d'après l'équation suivante

$$P(\bar{X} - k \leq m \leq \bar{X} + k) = 1 - \alpha$$

et les limites prendront, après avoir prélevé l'échantillon et calculé l'estimation  $\bar{X}$ , la forme suivante

$$\bar{X} - k \leq m \leq \bar{X} + k.$$

où  $k$  sera déterminé à l'aide de l'écart-type de la distribution d'échantillonnage de  $\bar{X}$  et de niveau de confiance  $1 - \alpha$  choisi a priori.

Nous savons que si nous prélevons un échantillon aléatoire de taille  $n$  d'une population normale de variance connue,

$$\bar{X} \rightarrow N\left(m, \frac{\sigma^2}{n}\right)$$

Si la distribution du caractère mesurable (la population) est inconnue ou si la variance de la population est inconnue, un échantillon de taille  $n \geq 30$  nous permet, d'après le théorème central limite de considérer que  $\bar{X}$  suit approximativement une loi normale. Par conséquent, la quantité

$$Z = \frac{\bar{X} - m}{\frac{\sigma}{\sqrt{n}}} \text{ ou } \left( \frac{\bar{X} - m}{\frac{S}{\sqrt{n}}} \text{ selon le cas } \right)$$

suit une loi normale centrée réduite.

Partons de ce fait pour déduire un intervalle aléatoire ayant, a priori, une probabilité  $1 - \alpha$  de contenir la vraie valeur de  $m$ , ce qui revient à déterminer  $k$  de telle sorte que

$$P(\bar{X} - k \leq m \leq \bar{X} + k) = 1 - \alpha$$

D'où

$$P\left(-Z_{\alpha} \frac{\sigma}{\sqrt{n}} \leq \frac{\bar{X} - m}{\frac{\sigma}{\sqrt{n}}} \leq Z_{\alpha} \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

Alors

$$P\left(\bar{X} - Z_{\alpha} \frac{\sigma}{\sqrt{n}} \leq m \leq \bar{X} + Z_{\alpha} \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

qui est de la forme

$$P(\bar{X} - k \leq m \leq \bar{X} + k) = 1 - \alpha$$

d'où ;

$$k = Z_{\alpha} \frac{\sigma}{\sqrt{n}}.$$

**Exemple 0.8.9** *Un laboratoire indépendant à vérifié pour le compte de l'office de la protection du consommateur, la résistance à l'éclatement (en  $\text{kg lcm}^2$ ) d'un réservoir à essence d'un certain fabricant. Des essais similaires effectués il y a un an, permettent de considérer que la résistance à l'éclatement est distribuée normalement avec une variance de 9. Des essais sur un échantillon de 10 réservoirs conduisent à une résistance moyenne à l'éclatement de  $219 \text{kg lcm}^2$ . Estimer par intervalle de confiance la résistance moyenne à l'éclatement de ce type de réservoir avec un niveau de confiance de 95%.*

Les éléments nécessaires au calcul de l'intervalle de confiance sont indiqués comme suit :

$$\begin{aligned}\bar{X} &= 219 \\ \sigma &= \sqrt{9} = 3 \text{kg lcm}^2\end{aligned}$$

La taille de l'échantillon  $n = 10$ , le niveau de confiance :  $1 - \alpha = 0,95$  donc  $\alpha = 0,05$ .

# Tests d'hypothèses

## Introduction

Soit une hypothèse  $H_0$  concernant une population sur la base des résultats d'échantillons extraits de cette population ont est amené à accepter ou rejeter l'hypothèse  $H_0$ .

Les règles de décisions sont appelées tests statistiques.

$H_0$  désigne l'hypothèse dite hypothèse nulle et par  $H_1$  on note l'hypothèse dite hypothèse alternative on a :

$$\begin{cases} H_0 \text{ vraie} & H_1 \text{ fausse ou bien} \\ H_0 \text{ fausse} & H_1 \text{ vraie} \end{cases}$$

Il ya quatre solutions dont seulement les deux premières sont justes :

- a)  $H_0$  est vraie et on a choisi  $H_0$
- b)  $H_0$  est fausse et on a rejeté  $H_0$
- c)  $H_0$  est vraie et on a rejeté  $H_0$
- d)  $H_1$  est vraie et on a choisi  $H_0$

On distingue deux types d'erreurs :

a) Si  $H_0$  est vraie et on l'a rejetée, on dit que l'on a une erreur de 1<sup>tère</sup> espèce. La probabilité de l'erreur de 1<sup>tère</sup> espèce est notée  $\alpha$ .

b) Si  $H_1$  est vraie et on a accepté  $H_0$ , on dit que l'on a une erreur de 2<sup>e</sup> espèce. La probabilité de l'erreur de 2<sup>e</sup> espèce est notée  $\beta$ .

$\alpha$  est le seuil de signification du test et  $1 - \alpha$  son seuil de confiance.

## 0.9 Catégories de tests

1. Un test est dit d'hypothèse simple si on veut choisir entre deux valeurs d'un paramètre  $\theta$  ( $\theta_0$  et  $\theta_1$ ) on a :

$$\begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta = \theta_1 \end{cases}$$

2. Un test est dit bilatéral si :

$$a) \begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta \neq \theta_1 \end{cases}$$

test unilatéral à droite ou test de supériorité

$$b) \begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta > \theta_1 \end{cases}$$

test unilatéral à gauche ou test d'infériorité

$$c) \begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta < \theta_1 \end{cases}$$

où  $\theta_0$  et  $\theta_1$  sont deux valeurs d'un même paramètre dans deux populations différentes.

3. Un test est dit d'ajustement si :

$$\begin{cases} H_0 : F(x) = F_0(x) \\ H_1 : F(x) \neq F_0(x) \end{cases}$$

Où  $F(x)$  est la fonction de répartition de la variable échantillonnée et  $F_0(x)$  est la fonction de répartition d'une variable aléatoire connue.

5. Un test est dit d'indépendance si :

$$\begin{cases} H_0 : X \text{ et } Y \text{ sont deux variables aléatoires indépendantes} \\ H_1 : X \text{ et } Y \text{ ne sont pas des variables indépendantes.} \end{cases}$$

### 0.9.1 Région critique et d'acceptation d'une hypothèse

La construction d'un test implique la détermination de la région critique  $\omega_0$  de  $\mathbb{R}^n$ . La valeur de  $\alpha$ , erreur de 1<sup>ère</sup> espèce étant fixée (en général  $\alpha = 0.05, 0.01$  ou  $0.1$ ) l'ensemble des valeurs de la variable de décision qui

permettent d'écarter  $H_0$  et de choisir  $H_1$  est dit région critique, le complémentaire  $\bar{\omega}_0$  de cette région critique est dit région d'acceptation. On a :

$$\begin{aligned} P(\omega_0 | H_0) &= \alpha; P(\bar{\omega}_0 | H_0) = 1 - \alpha \\ P(\omega_0 | H_1) &= 1 - \beta; P(\bar{\omega}_0 | H_1) = \beta \end{aligned}$$

on extrait un échantillon aléatoire de la population et on accepte  $H_0$  si la valeur de la variable de décision appartient à la région d'acceptation. Sinon on la rejette et on accepte  $H_1$ .

Pour une valeur  $\alpha$  fixée, on maximise la quantité  $1 - \beta$  dite puissance du test.

### 0.9.2 Test entre deux hypothèses simples (méthode de Neyman et Pearson)

On test

$$\begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta = \theta_1 \end{cases}$$

on fixe le risque de 1<sup>ère</sup> espèce  $\alpha$ .

$$L(x; \theta) = L(x_1, \dots, x_n; \theta)$$

est la fonction de vraisemblance avec  $x = (x_1, \dots, x_n)$ ,  $\omega_0$ , région critique, est définie par :

$$P(\omega_0 | H_0) = \alpha = \int_{\omega_0} L(x; \theta_0) dx.$$

on maximise la quantité :

$$1 - \beta = \int_{\omega_0} L(x; \theta_1) dx = P(\omega_0 | H_1) = 1 - \beta = \int_{\omega_0} \frac{L(x; \theta_1)}{L(x; \theta_0)} L(x; \theta_0) dx$$

pour maximiser  $1 - \beta$ , on cherche l'ensemble des points de  $\mathbb{R}^n$  tels que :

$$A = \frac{L(x; \theta_1)}{L(x; \theta_0)} \geq K_\alpha$$

la constante  $K_\alpha$  est déterminée par :

$$\int_{A \geq K_\alpha} L(x; \theta_0) dx = \alpha.$$

### 0.9.3 Test d'homogénéité

A partir d'un échantillon de taille  $n_1$  extrait d'une population  $P_1$  et d'un échantillon de taille  $n_2$  extrait d'une population  $P_2$ , le test permet de décider entre :

$$\begin{cases} H_0 : \theta_0 = \theta_1 \\ H_1 : \theta_0 \neq \theta_1 \end{cases}$$

$\theta_0$  et  $\theta_1$  sont les deux valeurs d'un même paramètre des deux populations  $P_1$  et  $P_2$ .

### 0.9.4 Test d'homogénéité de deux moyennes

Dans le cas où tailles des échantillons sont élevées ( $n_1, n_2 \geq 30$ ), les variables  $\bar{X}_1$  et  $\bar{X}_2$  (correspondants aux populations  $P_1$  et  $P_2$ ) suivant les lois normales respectives :  $N\left(m_1, \frac{\sigma_1}{\sqrt{n_1}}\right), N\left(m_2, \frac{\sigma_2}{\sqrt{n_2}}\right)$  où  $m_i$  et  $\sigma_i$  sont la moyenne et l'écart-type de la population  $P_i$  ( $i = 1, 2$ ).

La variable aléatoire  $(\bar{X}_1 - \bar{X}_2)$  suit aussi une loi normale  $N\left(m_1 - m_2, \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right)$ , on choisit entre les deux hypothèses :

$$\begin{cases} H_0 : m_1 = m_2 \\ H_1 : m_1 \neq m_2 \end{cases}$$

où

$$\begin{cases} H_0 : m_1 - m_2 = 0 \\ H_1 : m_1 - m_2 \neq 0 \end{cases}$$

si  $H_0$  est vraie,  $m_1 - m_2 = 0$  et  $\bar{X}_1 - \bar{X}_2 \rightarrow N\left(0, \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right)$ . On a alors :

$$P(|Z| \leq U_\alpha) = 1 - \alpha$$

où

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (m_1 - m_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{(\bar{X}_1 - \bar{X}_2) - 0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

suit une loi normale  $N(0, 1)$ . On accepte  $H_0$  si la valeur  $z = \frac{(\bar{X}_1 - \bar{X}_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$  de

$Z$  telle que :

$$-U_\alpha \leq z \leq U_\alpha$$

où la valeur  $U_\alpha$  est obtenue par la lecture de la table de la loi normale  $N(0, 1)$ . On rejette  $H_0$  si  $|Z| > U_\alpha$ . On dira que la différence est significative entre  $\bar{X}_1$  et  $\bar{X}_2$ .

**Remarque 0.9.1** *i) Si  $\sigma_1^2$  et  $\sigma_2^2$  sont inconnues, on les remplace par leurs estimateurs :*

$$S_1^2 = \frac{1}{n_1 - 1} \sum_{i=1}^{n_1} (X_i - \bar{X}_1)^2$$

et

$$S_2^2 = \frac{1}{n_2 - 1} \sum_{i=1}^{n_2} (X_i - \bar{X}_2)^2 \text{ respectivement.}$$

comme les échantillons sont de tailles élevées on considère que

$Z = \frac{(\bar{X}_1 - \bar{X}_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$  suit une loi normale  $N(0, 1)$  (si  $H_0$  est vraie c-à-d  $m_1 - m_2 = 0$ )

*ii) Si les échantillons sont de tailles respectives  $n_1 < 30$  et  $n_2 < 30$ , le test n'est plus valable car le théorème central limite ne s'applique plus. mais pour deux population  $P_1$  et  $P_2$  suivant des lois normales  $N(m_1, \sigma_1)$  et  $N(m_2, \sigma_2)$  respectivement ayant des écarts-type  $\sigma_1$  et  $\sigma_2$  égaux et inconnus, c-à-d  $\sigma_1 = \sigma_2 = \sigma$  on a :*

$t = \frac{\bar{X}_1 - \bar{X}_2}{S \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$  suit une loi de **Student** à  $n_1 + n_2 - 2$  degrés de liberté,  $S^2$  étant l'estimation ponctuelle de  $\sigma^2$ .

on accepte alors  $H_0$  si  $|t| < U_\alpha$  où  $U_\alpha$  est une valeur obtenue par la lecture de la table de la loi  $t$  de **Student-Fisher** (nombre de degrés de liberté  $n_1 + n_2 - 2$ ; seuil de signification  $\alpha$ )

### 0.9.5 Test d'homogénéité de deux proportions

Chaque individu des deux populations  $p_1$  et  $p_2$  peut posséder ou non un certain caractère. On dit ce caractère est présent en proportion  $P_1$  et  $P_2$  dans les populations  $p_1$  et  $p_2$  respectivement : on test au seuil de signification  $\alpha$  :

$$\begin{cases} H_0 : P_1 = P_2 \\ H_1 : P_1 \neq P_2 \end{cases}$$

ou

$$\begin{cases} H_0 : P_1 - P_2 = 0 \\ H_1 : P_1 - P_2 \neq 0 \end{cases}$$

De la population  $P_i$ , on extrait un échantillon de taille  $n_i$ . Il lui correspond une proportion  $f_i$  ( $i = 1, 2$ ).

Si les échantillons sont de tailles élevées ( $n_1 \geq 30$  et  $n_2 \geq 30$ ), le théorème central limite permet d'affirmer que  $f_i$  suit une loi normale  $N\left(P_i, \sqrt{\frac{P_i Q_i}{n_i}}\right)$  avec  $P_i + Q_i = 1$ . La variable aléatoire  $f_1 - f_2$  suit alors une loi normale  $N\left(P_1 - P_2, \sqrt{\frac{P_1 Q_1}{n_1} + \frac{P_2 Q_2}{n_2}}\right)$ .

Si  $H_0$  est vraie,  $P_1 - P_2 = 0$  et  $f_1 - f_2 \rightarrow N\left(0, \sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}\right)$  car  $P_1 = P_2 = P$ . Mais puisque  $P$  est inconnu, on fait une approximation. On estime  $P$  par  $f = \frac{n_1 f_1 + n_2 f_2}{n_1 + n_2}$  et donc

$$f_1 - f_2 \rightarrow N\left(0, \sqrt{f(1-f)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}\right)$$

et

$$Z = \frac{f_1 - f_2}{\sqrt{f(1-f)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \rightarrow N(0, 1).$$

Soit  $U_\alpha$  la constante telle que  $P(|Z| \geq U_\alpha) = \alpha$ . On a alors :  $P(|Z| < U_\alpha) = 1 - \alpha$  on accepte  $H_0$  si la valeur

$$z = \frac{f_1 - f_2}{\sqrt{f(1-f)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

de  $Z$  est telle que :

$$-U_\alpha \leq z \leq U_\alpha$$

on rejette  $H_0$  si  $|Z| \geq U_\alpha$ .

$U_\alpha$  est obtenue par lecture de la table de la loi normale  $N(0, 1)$ .

**Exercices de travaux dirigés**

**Exercice 1 :** La loi de probabilité conjointe du couple de v.a.r  $(X, Y)$  est donnée par le tableau :

$X/Y \rightarrow$	0	1	2	3
0	$\frac{2}{48}$	$\frac{6}{48}$	$\frac{3}{48}$	$\frac{1}{48}$
2	$\frac{4}{48}$	$\frac{12}{48}$	$\frac{6}{48}$	$\frac{2}{48}$
4	$\frac{2}{48}$	$\frac{6}{48}$	$\frac{3}{48}$	$\frac{1}{48}$

- 1) Déterminer les lois marginales de  $X, Y$ .
- 2) Déterminer l'espérance de  $X$  et  $Y$ .
- 3) Les variables  $X$  et  $Y$  sont-elle indépendantes?

**Exercice 2 :** Soit  $(X, Y)$  un couple de v.a discrètes dont la loi de probabilité est donnée par la tableau ci-après :

$Y/X \rightarrow$	-1	0	1
-1	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$
0	$\frac{1}{16}$	$\frac{1}{8}$	$\frac{1}{16}$
1	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$

- 1) Déterminer les lois marginales de  $X$  et  $Y$ .
- 2) Calculer  $Cov(X, Y)$ .
- 3) Déterminer la loi conditionnelle de  $Y$  sachant que  $X = 0$ .

**Exercice 3:** Soit  $X, Y$  deux v.a indépendantes de lois de Poisson de paramètres respectifs  $\lambda$  et  $\mu$ .

Déterminer la loi conditionnelle de  $X$  lorsque la somme  $S = X + Y$  a une valeur fixée  $S = s$ . En déduire l'expression de la fonction de régression de  $X$  sur  $S$  puis la valeur de  $\mathbb{E}[\mathbb{E}(X/S)]$ .

**Exercice 4 :** Soit  $(X, Y)$  un couple de v.a dont la loi est déterminée par la densité :

$$f(x, y) = \begin{cases} kxy & \text{si } (x, y) \in [0, 2] \times [0, 5] \\ 0 & \text{sinon} \end{cases}$$

- 1) Déterminer la valeur de la constante  $k$ .
- 2) Déterminer les lois marginales de  $X$  et  $Y$ . Ces variables sont-elle indépendantes?

**Exercice 5 :** Soit  $(X, Y)$  un couple de v.a de densité :

$$f(x, y) = \begin{cases} \frac{k}{\sqrt{xy}} & \text{si } 0 < x \leq y < 1 \\ 0 & \text{sinon} \end{cases}$$

1) Déterminer la valeur de la constante  $k$  puis la fonction de répartition  $F$  de ce couple.

2) Déterminer les lois marginales de  $X$  et  $Y$ . Ces variables sont-elles indépendantes ?

3) Déterminer les lois conditionnelles de  $X/Y = y$  et de  $Y/X = x$ . En déduire l'expression de la fonction de régression  $x \mapsto \mathbb{E}(Y/X)$  puis calculer  $\mathbb{E}[\mathbb{E}(Y/X)]$ .

**Exercice 6** : Soit  $X$  et  $Y$  deux v.a indépendantes et de même loi uniforme sur  $] -1, 1[$ . Déterminer la loi de probabilité de la v.a  $Z = Y - X$ .

**Exercice 7**

**I)**

Etudier la convergence en probabilité, en moyenne, puis en moyenne quadratique de la suite de v.a  $(X_n)$  dont la loi de probabilité est définie pour  $n \in \mathbb{N}$  par :

$$P(X_n = -n) = P(X_n = n) = \frac{1}{2n^2} \text{ et } P(X_n = 0) = 1 - \frac{1}{n^2}$$

**II)**

Etudier la convergence en probabilité de la suite de variable aléatoire  $(X_n)_{n \geq 1}$ , telle que la loi de  $X_n$  admette le denticité  $f_n$  par rapport à  $\lambda$  :

$$f_n(x) = \frac{n \exp(-nx)}{(1 + \exp(-nx))^2}$$

**III)**

Soit  $(X_n)_{n \geq 1}$  une suite de variables aléatoires réelles indépendantes et de même loi, de moyenne  $m$  et de variance  $\sigma^2$  finies. Soit  $S_n = X_1 + \dots + X_n$ , alors  $\frac{S_n - nm}{\sigma\sqrt{n}} \xrightarrow[n \rightarrow \infty]{loi} N(0, 1)$ .

**Exercice 8**

1. Supposons qu'une population consiste d'unités statistiques dont le caractère mesurable de chacune est :

$$x_1 = 332, x_2 = 336, x_3 = 340, x_4 = 344, x_5 = 348$$

a) Quelles sont la taille  $N$  de la population, la moyenne et la variance de la population ?

b) On veut prélever de cette population des échantillons de taille  $n = 2$  en effectuant un tirage avec remise. Combien d'échantillons peut-on prélever ?

c) Former tous les échantillons possible de taille  $n = 2$  (tirage non exhaustif) et calculer la moyenne de chacun.

d) Déterminer les paramètres de la distribution d'échantillonnage de  $\bar{X}$ .

### **Exercice 9**

On choisit au hasard sans remise six nombres parmi les nombres entiers de 1 à 9; chacun de ces nombres a la même probabilité d'être choisi. Calculer la moyenne et l'écart-type de la distribution d'échantillonnage des moyennes.

### **Exercice 10**

On suppose que le poids des alevins d'une population est une variable aléatoire d'écart-type égal à **3g**. Dans des conditions données, on a pesé un échantillon de **45** alevins; la moyenne observée est de **8,24g**.

Donner un **intervalle de confiance** du poids moyen des alevins de la population étudiée.

Combien faudrait-il contrôler d'alevins pour que l'intervalle de confiance du poids moyen soit déterminé avec une amplitude de **1g** au maximum? On prendra **95%** pour **seuil de confiance**.

### **Exercice 11**

Le revenu mensuel  $X$  d'une certaine population suit une loi de Pareto de densité :

$$f(x; \theta) = \begin{cases} \frac{1}{x_0^{\frac{\theta}{1+\theta}}} & \text{si } x > x_0 \\ \theta x^{\frac{1+\theta}{\theta}} & \\ 0 & \text{si } x \leq x_0 \end{cases}$$

où  $x_0 > 0$  est le revenu minimum, connu, et  $\theta$  un paramètre strictement positif que l'on se propose d'estimer à partir d'un échantillon  $(X_1, X_2, \dots, X_n)$  de  $X$ .

1) Calculer la vraisemblance  $L(X_1, X_2, \dots, X_n; \theta)$ .

2) Déterminer l'estimateur du maximum de vraisemblance  $\hat{\theta}_n$  de  $\theta$  et étudier ses propriétés.

### **Exercice 12**

Sur **2424** naissances, on a trouvé **1270 garçons** et **1154 filles**.

1) Donner une estimation du pourcentage de garçons à la naissance dans la population, ainsi que l'intervalle de confiance de ce pourcentage aux risques **5%** et **1%**.

2) Combien de naissances doit-on recenser pour connaître le pourcentage de garçons dans la population avec une précision égale à **0,5%** au risque **5%**.

**Examens finaux**

**Exercice 1**

I) Soit  $X$  et  $Y$  deux variables aléatoires indépendantes et de même loi uniforme sur  $] -1, 1[$ . Déterminer la loi de probabilité de la variable aléatoire

$$Z = Y - X.$$

II) Soient  $X$  et  $Y$  deux variables aléatoires indépendantes de loi exponentielle de paramètre  $\lambda > 0$ .

Déterminer la fonction de répartition de  $X$  puis la loi de de la variable

$$Z = \min(X, Y).$$

**Exercice 2**

On effectue  $n$  expériences successives indépendantes où on s'intéresse à chaque fois à la réalisation d'un certain événement  $A$ . On associe donc à chaque expérience  $i$ ,  $1 \leq i \leq n$ , une variable de Bernoulli  $X_i = \begin{cases} 1 & p \\ 0 & q = 1 - p \end{cases}$ .

Montrer que la fréquence empirique, c'est à dire le pourcentage de réalisation de  $A$  est :

$$f_n = \frac{1}{n} \sum_{i=1}^n X_i = \bar{X}$$

converge en probabilité vers  $p(A)$  ( $f_n \xrightarrow{P} p(A) = p$ )

**Exercice 3**

Le revenu mensuel  $X$  d'une certaine population suit une loi de Pareto de densité :

$$f(x; \theta) = \begin{cases} \frac{1}{x_0^{\frac{\theta}{1+\theta}}} & \text{si } x > x_0 \\ \theta x^{1+\frac{\theta}{1+\theta}} & \\ 0 & \text{si } x \leq x_0 \end{cases}$$

où  $x_0 > 0$  est le revenu minimum, connu, et  $\theta$  un paramètre strictement positif que l'on se propose d'estimer à partir d'un échantillon  $(X_1, X_2, \dots, X_n)$  de  $X$ .

1) Calculer la vraisemblance  $L(X_1, X_2, \dots, X_n; \theta)$ .

2) Déterminer l'estimateur du maximum de vraisemblance  $\hat{\theta}_n$  de  $\theta$  et étudier ses propriétés.

**Exercice 4**

Question de cours

Soient  $(X_n)$  et  $X$  étant des variables aléatoires.

Montrer que :

$$X_n \xrightarrow{ps} X \Rightarrow X_n \xrightarrow{p} X.$$

**Exercice 5**

Soit  $X$  une variable aléatoire dont la loi est donnée dans le tableau ci-dessous :

$x_i$	-2	-1	1	2
$P(X = x_i)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

On pose  $Y = X^2$ .

1. Déterminer la loi de la variable aléatoire  $Y$ .
2. Déterminer la loi du couple  $(X, Y)$ . Que peut on conclure ?
3. Déterminer  $Cov(X, Y)$ .

**Exercice 6**

On choisit au hasard **avec remise** six nombres parmi les nombres entiers de **1 à 9** ; chacun de ces nombres a la même probabilité d'être choisi. Calculer la moyenne et l'écart-type de la distribution d'échantillonnage des moyennes.

**Exercice 7**

On suppose que le poids des alevins d'une population est une variable aléatoire d'écart-type égal à **3g**. Dans des conditions données, on a pesé un échantillon de **45** alevins ; la moyenne observée est de **8,24g**.

Donner un **intervalle de confiance** du poids moyen des alevins de la population étudiée.

Combien faudrait-il contrôler d'alevins pour que l'intervalle de confiance du poids moyen soit déterminé avec une amplitude de **1g** au maximum ? On prendra **95%** pour **seuil de confiance**.

# Bibliographie

- [1] Sheldon M. Ross, Initiation aux probabilités. Presses polytechniques romandes, Lausanne, 1987.
- [2] Jim Pitman, Probability. Springer, New York, 1993.
- [3] Michel Lejeune, Statistique La Théorie et ses applications. Springer, Paris, 2004.
- [4] Stephan Morgenthaler, Introduction à la statistique. Presses polytechniques et universitaires romandes, Lausanne, 2ème édition, 2001.
- [5] Yadolah Dodge, Premiers pas en statistique. Springer, Paris, 2003.
- [6] OLIVIER GAUDOIN. Principes et Méthodes Statistiques.
- [7] JULIEN JACQUES. Introduction aux série temporelle, Harner, New York 1965