

Polycopié de cours

Statistique inférentielle

Estimation Ponctuelle et par Intervalle de Confiance

Tests d'Hypothèses

Année
2021

GHEZAL AHMED

Département de Mathématiques-Informatique
Centre Universitaire **Abdelhafid Boussof**-Mila

GHEZAL AHMED

Statistique inférentielle

Estimation Ponctuelle et par Intervalle de Confiance et Tests d'Hypothèses

Département de Mathématiques-Informatique

(Centre Universitaire Abdelhafid Boussouf - Mila)

2021

A decorative border made of watercolor-style flowers and green leaves, framing the central text. The flowers include yellow, pink, and red blooms with green foliage.

على
بركة الله

Avant-Propos

Ce polycopié de cours est destiné principalement aux étudiants de **Licence Mathématiques** mais peut être utile à toute personne souhaitant connaître et surtout utiliser les principales méthodes de la statistique inférentielle.

Le niveau **mathématique** requis est celui de la première année et la deuxième année de **Licence**, avec quelques notions d'analyse et de probabilités, souvent enseignées en deuxième année **Licence**. De ce fait, l'étudiant doit maîtriser les méthodes d'analyse et d'algèbre de base ainsi que les techniques essentielles du calcul des probabilités.

La première partie du polycopié est consacrée à quelques rappels et compléments (Limite inférieure, supérieure, symboles o et O , la convergence des suites de variables aléatoires et les théorèmes limites) pour une utilisation judicieuse des méthodes statistiques. La deuxième partie statistique correspond aux chapitres d'échantillonnage, d'estimation et de tests d'hypothèses. Chaque chapitre se conclut par des exercices corrigés permettant de contrôler l'acquisition des notions essentielles qui ont été introduites.

Enfin, toutes vos remarques, vos commentaires, vos critiques, et même vos encouragements, seront accueillis avec plaisir. Vous pouvez me les communiquer à l'adresse électronique suivante : a.ghezal@centre-univ-Mila.dz

Bon courage !

Mila

Ahmed Ghezal

2021

Table des matières

I	Rappels et compléments	5
1	Limite inférieure, supérieure et symboles o et O	6
1.1	Limite inférieure et supérieure	6
1.2	Quelques lemmes utiles	10
1.3	Les symboles o et O	12
1.4	Exercices	13
2	Modes de convergence	18
2.1	Convergence presque sûre	18
2.2	Convergence en probabilité	21
2.3	Convergence en loi	27
2.4	Convergence en moyenne d'ordre p	31
2.4.1	Inégalités	31
2.4.2	T.C.D et T.C.M	32
2.4.3	Équi-Intégrabilité	32
2.4.4	Propriétés	32
2.5	Exercices	34
3	Théorèmes limites	42
3.1	Lois des grands nombres	42
3.2	Théorèmes centraux limites	45
3.3	Exercices	48

II	Statistique inférentielle	52
4	Échantillonnage. Modèles statistiques	53
4.1	Méthodes d'échantillonnage	53
4.1.1	Échantillonnage par choix raisonné	53
4.1.2	Échantillonnage aléatoire	54
4.2	Échantillon	56
4.3	Quelques statistiques classiques	57
4.3.1	Moyenne, variance, moments empiriques d'un échantillon	57
4.3.2	Lois usuelles continues	59
4.4	Modèle statistique	60
5	Estimation ponctuelle	62
5.1	Estimateur et estimation	62
5.2	Méthodes d'estimation	69
5.2.1	Construction d'estimateur par la méthode du maximum de vraisemblance	69
5.2.2	Construction d'estimateur par la méthode des moments	73
5.3	Exercices	74
6	Estimation par intervalle de confiance	92
6.1	Construction d'un intervalle de confiance	94
6.2	Exemples classiques d'estimation par intervalle	95
6.2.1	Intervalle de confiance pour les paramètres d'une loi normale	95
6.2.2	Intervalle de confiance d'une proportion	101
6.3	Comparaison de moyennes et de variances	102
6.3.1	Intervalle de confiance de la différence de deux moyennes	102
6.3.2	Intervalle de confiance du rapport de deux variances	107
6.4	Exercices	110
7	Tests d'Hypothèses	115
7.1	Définitions	115
7.1.1	Hypothèses nulle et alternative	116

7.1.2	Région critique	118
7.1.3	Risques	119
7.2	Règle de décision	123
7.3	Les tests paramétriques usuels	123
7.3.1	Tests sur la moyenne d'une loi $\mathcal{N}(\mu, \sigma^2)$	123
7.3.2	Tests sur la variance σ^2 d'une loi $\mathcal{N}(\mu, \sigma^2)$	127
7.3.3	Tests de comparaison des moyennes de deux lois de Gauss	128
7.3.4	Tests de comparaison des variances de deux lois de Gauss	130
7.4	Exercices	131
 Annexe		136
8	Annexe : Tables statistiques usuelles	137
8.1	Loi normale	138
8.2	Loi du chi-deux	140
8.3	Loi de Student	141
8.4	Loi de Fisher	142

Première partie

Rappels et compléments

Chapitre 1

Limite inférieure, supérieure et symboles

o et O

Dans toute la suite, nous utiliserons les notations suivantes :

- Ω est un ensemble non vide.
- $P(\Omega)$ désigne l'ensemble des parties d'un ensemble Ω .
- $x \in \Omega$ signifie que x est élément de l'ensemble Ω .
- Soit $(A_n)_{n \geq 1}$ une suite de parties de Ω , la réunion $\bigcup_{n \geq 1} A_n$ de la suite $(A_n)_{n \geq 1}$ est l'ensemble des éléments $x \in \Omega$ ayant la propriété " Il existe un entier n tel que $x \in A_n$ ". De même, l'intersection $\bigcap_{n \geq 1} A_n$ de la suite $(A_n)_{n \geq 1}$ est l'ensemble des éléments $x \in \Omega$ ayant la propriété " $x \in A_n$ pour une infinité d'indices n ".
- Si $(A_n)_{n \geq 1}$ une suite monotone décroissante de parties de Ω , *i.e.*, $A_1 \supset A_2 \supset \dots \supset A_n \supset A_{n+1} \supset \dots$, alors par définition $\lim \downarrow A_n := \bigcap_{n \geq 1} A_n$ et si $(A_n)_{n \geq 1}$ une suite monotone croissante de parties de Ω , *i.e.*, $A_1 \subset A_2 \subset \dots \subset A_n \subset A_{n+1} \subset \dots$, alors $\lim \uparrow A_n := \bigcup_{n \geq 1} A_n$.
- Si $(A_n)_{n \geq 1}$ une suite arbitraire de parties de Ω , alors $\sup_{k \geq n} A_k = \bigcup_{k \geq n} A_k$ et $\inf_{k \geq n} A_k = \bigcap_{k \geq n} A_k$.

1.1 Limite inférieure et supérieure

Cette section est consacrée aux notions de limites, *limite supérieure* et *limite inférieure* d'une suite.

Soit $(A_n)_{n \geq 1}$ une suite de parties de Ω , on définit la *limite supérieure* et la *limite inférieure* de la suite $(A_n)_{n \geq 1}$ notées $\overline{\lim}A_n$ et $\underline{\lim}A_n$, par

$$\begin{aligned}\overline{\lim}A_n & : = \bigcap_{n \geq 1} \bigcup_{k \geq n} A_k = \inf_{n \geq 1} \sup_{k \geq n} A_k, \\ \underline{\lim}A_n & : = \bigcup_{n \geq 1} \bigcap_{k \geq n} A_k = \sup_{n \geq 1} \inf_{k \geq n} A_k,\end{aligned}$$

où la notation $\inf \sup$ et $\sup \inf$ est à prendre au sens de la relation d'ordre partiel \subset dans $P(\Omega)$. Les ensembles $\overline{\lim}A_n$ et $\underline{\lim}A_n$ sont appelés respectivement *limite supérieure* et *limite inférieure* de la suite $(A_n)_{n \geq 1}$. Lorsque $\underline{\lim}A_n = \overline{\lim}A_n$, on définit $\lim A_n$ comme étant l'ensemble commun sur lequel $\underline{\lim}A_n = \overline{\lim}A_n$ et on dit que la suite $(A_n)_{n \geq 1}$ tend vers $A = \lim A_n$ ou converge vers A . Si $\underline{\lim}A_n \neq \overline{\lim}A_n$, alors $\lim A_n$ est indéfinie. Ainsi, *limite supérieure*, *limite inférieure* et *limite* (lorsqu'elles existent) sont calculés à partir de la suite $(A_n)_{n \geq 1}$ toute entière. On peut également écrire $\underline{\lim}A_k$, $\underline{\lim}A_m$ pour $\underline{\lim}A_n$ (mêmes commentaires pour *limite supérieure* et *limite*). Il y a des situations où l'indice de la suite est important de le préciser. Pour l'instant, si $(B_n)_{n \geq 1}$ est une autre suite de parties de Ω , alors $\underline{\lim}(A_n \cap B_k)$ est ambigu. Par contre $\underline{\lim}_n(A_n \cap B_k)$ et $\underline{\lim}_k(A_n \cap B_k)$ désigne deux choses différentes. En effet, pour k fixé $\underline{\lim}_n(A_n \cap B_k)$ désigne la *limite inférieure* de la suite $(A_n \cap B_k)_{n \geq 1}$, par contre $\underline{\lim}_k(A_n \cap B_k)$ désigne la *limite inférieure* de la suite $(A_n \cap B_k)_{k \geq 1}$ où n étant fixé.

Remarque 1.1.1 Soit A la limite d'une suite $(A_n)_{n \geq 1}$ qui converge. Alors A est caractérisé par

$$\begin{cases} \forall x \in A, \exists m, \forall n \geq m, x \in A_n \\ \forall x \notin A, \exists k, \forall n \geq k, x \notin A_n \end{cases}.$$

Les propriétés de base des limites inférieure et supérieure d'une suite $(A_n)_{n \geq 1}$ de parties de Ω sont regroupées dans le lemme suivant

Lemme 1.1.1 On peut aussi caractériser la limite supérieure et la limite inférieure par les assertions suivantes :

- a. $\underline{\lim}A_n = \{x \in \Omega : \#\{n \in \mathbb{N} : x \notin A_n\} < \infty\}$ (c-à-d, $\underline{\lim}A_n$ est définie comme l'ensemble de tous les éléments x de Ω qui appartiennent à tous les A_n sauf à un nombre fini d'entre eux).
- b. $\overline{\lim}A_n = \{x \in \Omega : \#\{n \in \mathbb{N} : x \in A_n\} = \infty\}$ (c-à-d, $\overline{\lim}A_n$ est définie comme l'ensemble de tous les éléments x de Ω qui appartiennent à A_n pour une infinité d'indices n).

c. $\underline{\lim}A_n \subset \overline{\lim}A_n$.

Preuve.

a. Supposons que $\underline{\lim}A_n = \emptyset$, alors $\forall k \in \mathbb{N}$ on a $\bigcap_{n \geq k} A_n = \emptyset$. Mais, alors il n'existe aucun $x \in \Omega : x \in$ à tous les A_n sauf à un nombre fini d'entre eux. Si $\underline{\lim}A_n \neq \emptyset$, alors soit $x \in \underline{\lim}A_n$ donc il existe $k \in \mathbb{N}$, tel que $x \in A_k, A_{k+1}, \dots$. Ainsi $x \in$ à tous les A_n sauf à un nombre fini d'entre eux. Inversement, si $x \in$ à tous les A_n sauf à un nombre fini d'entre eux alors il existe $k \in \mathbb{N}$ tel que $x \in \bigcap_{n \geq k} A_n$ et donc

$$x \in \bigcup_{k \geq 1} \bigcap_{n \geq k} A_n = \underline{\lim}A_n.$$

b. Si $x \in$ à une infinité de parties A_n alors cet élément appartient toujours à la réunion $\bigcup_{n \geq k} A_n$. Par conséquent, $x \in \bigcap_{k \geq 1} \bigcup_{n \geq k} A_n = \overline{\lim}A_n$, ce qui établit la seconde identité.

c. L'assertion est banale si $\underline{\lim}A_n = \emptyset$, sinon $\exists x \in \underline{\lim}A_n$. Ce x est dans tous les A_n sauf un nombre fini d'entre eux, donc il est dans une infinité de A_n , c-à-d, $x \in \overline{\lim}A_n$.

■

Exemple 1.1.1 A et B sont deux parties de Ω , on définit la suite $(C_n)_{n \geq 1}$ de parties de Ω par $C_{2n} = A$ et $C_{2n+1} = B$. Alors on a $\overline{\lim}C_n = A \cup B$, $\underline{\lim}C_n = A \cap B$.

Exemple 1.1.2 Soit $(A_n)_{n \geq 1}$ la suite de partie définie par

$$\begin{cases} A_1 = [0, 1], A_3 = [0, \frac{1}{3}], A_5 = [0, \frac{1}{5}], \dots \\ A_2 = [0, 2], A_4 = [0, 4], A_6 = [0, 6], \dots \end{cases}$$

alors $\underline{\lim}A_n = \{0\}$, $\overline{\lim}A_n = [0, \infty[$ et $\lim A_n$ n'existe pas.

Lemme 1.1.2 Soit $(A_n)_{n \geq 1}$ une suite de parties de Ω , alors on a

1. Toute suite monotone converge.

a. Si $(A_n)_{n \geq 1}$ est croissante, alors $\lim A_n$ existe et égale à $\bigcup_{n \geq 1} A_n$.

b. Si $(A_n)_{n \geq 1}$ est décroissante, alors $\lim A_n$ existe et égale à $\bigcap_{n \geq 1} A_n$.

2. Toute suite de parties deux à deux disjointes est convergente. i.e., si $(A_n)_{n \geq 1}$ est une suite de parties deux à deux disjointes, alors $\lim A_n = \emptyset$.

Preuve.

1. **a.** L'assertion est triviale si $A_n = \emptyset$ pour chaque n . Supposons que $\exists N \in \mathbb{N}$ tel que $A_N \neq \emptyset$. Comme

$(A_n)_{n \geq 1}$ est croissante, $x \in \text{à une infinité de } A_n$ si et seulement si $x \in \text{à tous les } A_n$ sauf un nombre fini d'entre eux. Par conséquent $\underline{\lim} A_n = \overline{\lim} A_n$, donc $\lim A_n$ existe et égale à

$$\underline{\lim} A_n = \bigcup_{n \geq 1} \bigcap_{k \geq n} A_k = \bigcup_{n \geq 1} A_n.$$

b. Si $(A_n)_{n \geq 1}$ est décroissante, alors $\bigcup_{n \geq k} A_n = A_k$, d'où $\overline{\lim} A_n = \bigcap_{n \geq 1} A_n$. D'autre part, la suite

$\left(\bigcap_{k \geq n} A_k \right)_{n \geq 1}$ est croissante, d'où $\underline{\lim} A_n = \bigcap_{n \geq 1} A_n$. Par conséquent $\underline{\lim} A_n = \overline{\lim} A_n$, donc $\lim A_n$

existe et égale à $\bigcap_{n \geq 1} A_n$.

2. Observons que si $(A_n)_{n \geq 1}$ une suite de parties deux à deux disjointes, alors il n'existe aucun $x \in \text{à une infinité de } A_n$, ainsi $\overline{\lim} A_n = \emptyset$, d'après l'assertion **c.** $\underline{\lim} A_n = \emptyset$ et donc $\lim A_n$ existe et égale à \emptyset .

■

Exemple 1.1.3 Soit $C_k = \{(x, y) \in \mathbb{R}^2 : 0 \leq x < k, 0 \leq y < \frac{1}{k}\}$. Alors

$$\begin{aligned} A_n &= \sup_{k \geq n} C_k = \bigcup_{k \geq n} C_k = \left\{ (x, y) \in \mathbb{R}^2 : 0 \leq x < \infty, 0 \leq y < \frac{1}{n} \right\}, \\ B_n &= \inf_{k \geq n} C_k = \bigcap_{k \geq n} C_k = \left\{ (x, y) \in \mathbb{R}^2 : 0 \leq x < n, y = 0 \right\}, \end{aligned}$$

comme $(A_n)_{n \geq 1}$ et $(B_n)_{n \geq 1}$ sont deux suites monotones décroissante et croissante respectivement de parties de Ω , alors

$$\begin{aligned} A &: = \lim \downarrow A_n = \bigcap_{n \geq 1} A_n = \overline{\lim} C_n = \{(x, y) \in \mathbb{R}^2 : 0 \leq x < \infty, y = 0\}, \\ B &: = \lim \uparrow B_n = \bigcup_{n \geq 1} B_n = \underline{\lim} C_n = \{(x, y) \in \mathbb{R}^2 : 0 \leq x < \infty, y = 0\}. \end{aligned}$$

et par conséquent $\lim C_n = \underline{\lim} C_n = \overline{\lim} C_n = \{(x, y) \in \mathbb{R}^2 : 0 \leq x < \infty, y = 0\}$.

Rappelons aussi que si $(a_n)_{n \geq 1}$ est une suite réelle, alors on définit $\underline{\lim} a_n, \overline{\lim} a_n$ par

$$\underline{\lim} a_n := \sup_{n \geq 1} \inf_{k \geq n} a_k, \quad \overline{\lim} a_n := \inf_{n \geq 1} \sup_{k \geq n} a_k,$$

ces deux quantités existent toujours dans $\overline{\mathbb{R}} := [-\infty, \infty] = \mathbb{R} \cup \{-\infty\} \cup \{+\infty\}$. Notons ici qu'une suite réelle est convergente si et seulement si $\underline{\lim} a_n$ et $\overline{\lim} a_n$ existent et sont égales, dans ce cas la quantité commune est notée $\lim a_n$. De même si $(f_n)_{n \geq 1}$ est une suite de fonctions réelles définies sur Ω , alors on définit $\underline{\lim} f_n, \overline{\lim} f_n$ comme suit

$$\forall x \in \Omega : (\underline{\lim} f_n)(x) := \underline{\lim} f_n(x), (\overline{\lim} f_n)(x) := \overline{\lim} f_n(x)$$

ce qui revient aux cas de limite inférieure et limite supérieure d'une suite réelle.

Exemple 1.1.4 Soit $(a_n)_{n \geq 1}$ la suite réelle définie par $a_n = (-1)^n$. Alors $\underline{\lim} a_n = -1$ et $\overline{\lim} a_n = 1$ et par conséquent $\lim a_n$ n'existe pas.

Exemple 1.1.5 Considérons la suite $(a_n)_{n \geq 1}$:

$$a_n = \begin{cases} \frac{n-3}{n} & \text{si } n \equiv 0 [3] \\ 2 & \text{si } n \equiv 1 [3] \\ \frac{n+1}{3} & \text{si } n \equiv 2 [3] \end{cases},$$

alors pour tout $n \in \mathbb{N}$ on a $\sup_{k \geq n} a_k = +\infty$ et

$$\inf_{k \geq n} a_k = \begin{cases} \frac{n-3}{3} & \text{si } n \equiv 0 [3] \\ \frac{n-1}{n+2} & \text{si } n \equiv 1 [3] \\ \frac{n-2}{n+1} & \text{si } n \equiv 2 [3] \end{cases},$$

par conséquent $\underline{\lim} a_n = 1$ et $\overline{\lim} a_n = \infty$ et donc $\lim a_n$ n'existe pas.

1.2 Quelques lemmes utiles

Lemme 1.2.1 [Lemme de Cesàro] Si $(x_n)_{n \geq 1}$ est une suite réelle convergente vers x , alors la suite $\left(\frac{1}{n} \sum_{t=1}^n x_t\right)_{n \geq 1}$ converge aussi vers x .

Remarque 1.2.1 Notons que l'inverse est généralement fausse. En effet, la suite $\left(\frac{1}{n} \sum_{t=1}^n (-1)^t\right)_{n \geq 1}$ converge vers 0, mais $((-1)^n)_{n \geq 1}$ ne converge pas.

Lemme 1.2.2 [Lemme de Toeplitz] Soient $(y_n)_{n \geq 1}$ une suite réelle convergente vers y et $(x_{nt}, t = 1, \dots, k_n)_{n \geq 1}$ une suite triangulaire vérifiant

1. $x_{nt} \longrightarrow 0$ quand $n \rightarrow \infty$ pour tout t .

$$2. \lim_{n \rightarrow \infty} \sum_{t=1}^{k_n} |x_{nt}| \leq K < +\infty.$$

$$3. \lim_{n \rightarrow \infty} \sum_{t=1}^{k_n} x_{nt} = 1.$$

$$\text{Alors } \lim_{n \rightarrow \infty} \sum_{t=1}^{k_n} x_{nt} y_t = y.$$

Preuve. Pour tout $\varepsilon > 0$, $\exists N_\varepsilon \geq 1 : \forall n \geq N_\varepsilon \implies |y_n - y| < \varepsilon/K$. D'autre par

$$\begin{aligned} \lim_{n \rightarrow \infty} \left| \sum_{t=1}^{k_n} x_{nt} y_t - y \right| &\leq \lim_{n \rightarrow \infty} \left| \sum_{t=1}^{k_n} x_{nt} (y_t - y) \right| + \lim_{n \rightarrow \infty} |y| \left| \sum_{t=1}^{k_n} x_{nt} - 1 \right| \\ &\leq \lim_{n \rightarrow \infty} \sum_{t=1}^{k_n} |x_{nt}| |y_t - y| + \lim_{n \rightarrow \infty} |y| \left| \sum_{t=1}^{k_n} x_{nt} - 1 \right| \leq 0. \end{aligned}$$

■

Un cas particulier, mais très fréquent qui satisfait les hypothèses du lemme est la suite (x_{nt}) ; $x_{nt} = x_t / \sum_{t=1}^n x_t$ où $(x_t)_{t \geq 1}$ est une suite réelle telle que $\sum_{t=1}^n x_t \rightarrow +\infty$. Notons aussi que dans le lemme de Toeplitz, si $y = 0$, alors l'hypothèse 3 peut être omise. Aussi si $x_{nt} \geq 0$ pour tout $t = 1, \dots, k_n$ et $n \geq 1$, alors l'hypothèse 2 peut être omise.

Une application immédiate du lemme de Toeplitz est la preuve du lemme suivant

Lemme 1.2.3 [Lemme de Kronecker] Soient $(a_n)_{n \geq 1}$ et $(x_n)_{n \geq 1}$ deux suites réelles positives t.q. $a_n \uparrow \infty$.

Alors lorsque $n \longrightarrow +\infty$ $\sum_{t=1}^n \frac{x_t}{a_t} \rightarrow c \implies \frac{1}{a_n} \sum_{t=1}^n x_t \rightarrow 0$.

Preuve. Posons $s_0 = 0$ et $s_n = \sum_{t=1}^{n-1} \frac{x_t}{a_t}$. Une sommation par partie donne

$$\frac{1}{a_n} \sum_{t=1}^n x_t = \frac{1}{a_n} \sum_{t=1}^n a_t (s_{t+1} - s_t) = s_{n+1} - \frac{1}{a_n} \sum_{t=1}^n (a_t - a_{t-1}) s_t = s_{n+1} - \sum_{t=1}^n a_{nt} s_t,$$

où

$$a_{ni} := \begin{cases} \frac{a_{i+1} - a_i}{a_n} & \text{si } i \leq n \\ 0 & \text{sinon} \end{cases}.$$

La suite $(a_{ni}, 1 \leq i \leq n)_{n \geq 1}$ est une suite triangulaire vérifiant les hypothèses du lemme de Toeplitz, Donc le résultat s'achève en appliquant le lemme de Toeplitz. ■

1.3 Les symboles o et O

Définition 1.3.1 Soient (a_n) et (b_n) deux suites réelles t.q $b_n > 0, \forall n$.

1. On dit que (a_n) est d'ordre plus petit que (ou est négligeable devant) (b_n) lorsque $n \rightarrow \infty$, et on écrit $a_n = o(b_n)$, si $\lim_{n \rightarrow \infty} \frac{a_n}{b_n} = 0$.
2. On dit que (a_n) est au plus du même ordre que (b_n) lorsque $n \rightarrow \infty$, et on écrit $a_n = O(b_n)$, s'il existe une constante $M > 0, \exists N_0, \forall n \geq N_0 : \frac{|a_n|}{b_n} \leq M$.

Exemple 1.3.1 Soit $a_n = 6n^2 + 2n + 4$. Alors, $a_n = O(n^2)$ et $a_n = o(n^{2+\delta}), \forall \delta > 0$.

Exemple 1.3.2 Soit $a_n = (-1)^n$. Alors, $a_n = O(1)$ et $a_n = o(n^\delta), \forall \delta > 0$.

Exemple 1.3.3 Soit $a_n = \exp\{-n\}$. Alors, $a_n = O(n^{-\delta})$ et $a_n = o(n^{-\delta}), \forall \delta > 0$.

Exemple 1.3.4 Soit $a_n = \exp\{n\}$. Alors, $a_n \neq O(n^k), \forall k \in \mathbb{R}$.

Propriétés 1.3.1 Soient $(a_n), (b_n)$ deux suites réelles et soient $(f_n), (g_n)$ deux suites réelles positives.

1. Si $a_n = o(f_n), b_n = o(g_n)$, alors

- $a_n b_n = o(f_n g_n)$,
- $|a_n^s| = o(f_n^s), s > 0$,
- $a_n + b_n = o(\max(f_n, g_n))$.

2. Si $a_n = O(f_n), b_n = O(g_n)$, alors

- $a_n b_n = O(f_n g_n)$,
- $|a_n^s| = O(f_n^s), s \geq 0$,
- $a_n + b_n = O(\max(f_n, g_n))$.

3. Si $a_n = o(f_n), b_n = O(g_n)$, alors

- $a_n b_n = o(f_n g_n)$,
- $a_n + b_n = O(\max(f_n, g_n))$.

Remarque 1.3.1 Une série $\sum_{n=1}^{\infty} x_n$ converge strictement si $\sum_{n=1}^m x_n = O(1)$ et $\sum_{n=m}^{\infty} x_n = o(1)$.

1.4 Exercices

Exercice 1.4.1 Soient $(A_n)_{n \geq 1}$ et $(B_n)_{n \geq 1}$ deux suites de parties de Ω . Montrer les assertions suivantes

1. $\underline{\lim} \mathbb{I}_{A_n} = \mathbb{I}_{\underline{\lim} A_n}$ et $\overline{\lim} \mathbb{I}_{A_n} = \mathbb{I}_{\overline{\lim} A_n}$.
2. Montrer que $\underline{\lim} A_n \subset \overline{\lim} A_n$ sans utiliser la méthode utilisée dans le lemme précédent.
3. $(\underline{\lim} A_n)^c = \overline{\lim} A_n^c$ et $\underline{\lim} A_n^c = (\overline{\lim} A_n)^c$.
4. $\overline{\lim} (A_n \cup B_n) = \overline{\lim} A_n \cup \overline{\lim} B_n$ et $\underline{\lim} (A_n \cap B_n) = \underline{\lim} A_n \cap \underline{\lim} B_n$.
5. $\underline{\lim} A_n \cap \underline{\lim} B_n \subset \overline{\lim} (A_n \cap B_n) \subset \overline{\lim} A_n \cap \overline{\lim} B_n$ et $\underline{\lim} A_n \cup \underline{\lim} B_n \subset \underline{\lim} (A_n \cup B_n) \subset \overline{\lim} (A_n \cup B_n) \subset \overline{\lim} A_n \cup \overline{\lim} B_n$.
6. $\mathbb{I}_{\overline{\lim} (A_n \cup B_n)} = \max \{ \mathbb{I}_{\underline{\lim} A_n}, \mathbb{I}_{\underline{\lim} B_n} \}$ et $\mathbb{I}_{\underline{\lim} (A_n \cap B_n)} = \inf \{ \mathbb{I}_{\overline{\lim} A_n}, \mathbb{I}_{\overline{\lim} B_n} \}$.
7. Si $A_n \rightarrow A$ et $B_n \rightarrow B$ alors $A_n^c \rightarrow A^c$, $A_n \cap B_n \rightarrow A \cap B$, $A_n \cup B_n \rightarrow A \cup B$, $A_n \setminus B_n \rightarrow A \setminus B$ et $A_n \Delta B_n \rightarrow A \Delta B$.
8. $\overline{\lim}_n \underline{\lim}_k (A_n \cap A_k^c) = \emptyset$, $A \setminus \overline{\lim} A_n = \underline{\lim} (A \setminus A_n)$, $A \setminus \underline{\lim} A_n = \overline{\lim} (A \setminus A_n)$.
9. $\overline{\lim} (A \Delta A_n) = (A \setminus \underline{\lim} A_n) \cup (\overline{\lim} (A_n \setminus A))$.
10. Si $A_n \rightarrow A$ alors $\overline{\lim} (A \Delta A_n) = \overline{\lim} (A_n \setminus A)$.
11. $\overline{\lim} A_n \setminus \underline{\lim} A_n = \overline{\lim} A_n \Delta \underline{\lim} A_n = (\underline{\lim} A_n \Delta A) \Delta (\overline{\lim} A_n \Delta A)$.

(Rappel : La différence symétrique $A \Delta B$ des ensembles A et B est définie par $(A \cap B^c) \cup (B \cap A^c)$, la fonction caractéristique ou indicatrice d'une partie A de Ω , est la fonction définie sur Ω à deux éléments $\{0; 1\}$, notée \mathbb{I}_A , donnée par : $\mathbb{I}_A(x) = \begin{cases} 1 & \text{si } x \in A \\ 0 & \text{sinon} \end{cases}$).

Exercice 1.4.2 1. Soit $(A_n)_{n \geq 1}$ la suite de parties de \mathbb{R}^+ définie par :

$$A_{2n} = [0; n] \text{ et } A_{2n+1} = [n, +\infty[.$$

Calculer la limite inférieure $\underline{\lim} A_n$ et la limite supérieure $\overline{\lim} A_n$.

2. Soient $(A_n)_n$ et $(B_n)_n$ deux suites de parties de \mathbb{R} telles que :

$$A_{2n} = \left[-2; 3 + \frac{1}{n} \right], A_{2n+1} = \left[-\frac{1}{n}; 2 \right], B_{2n} = \left[-1; \frac{n+1}{n} \right] \text{ et } B_{2n+1} = \left[-\frac{n+1}{n}; 1 \right].$$

Calculer $\underline{\lim} A_n$, $\overline{\lim} A_n$, $\underline{\lim} B_n$ et $\overline{\lim} B_n$.

Exercice 1.4.3 Soit $(a_n)_{n \geq 1}$ une suite réelle, c une constante. Montrer

1. $\overline{\lim}(c + a_n) = c + \overline{\lim}a_n$.

2. $\underline{\lim}(c + a_n) = c + \underline{\lim}a_n$.

3. $\overline{\lim}(ca_n) = \begin{cases} c\overline{\lim}a_n & \text{if } c \geq 0 \\ c\underline{\lim}a_n & \text{if } c < 0 \end{cases}$.

4. $\underline{\lim}(ca_n) = \begin{cases} c\underline{\lim}a_n & \text{if } c \geq 0 \\ c\overline{\lim}a_n & \text{if } c < 0 \end{cases}$.

5. En déduire que $\overline{\lim}(-a_n) = -\underline{\lim}a_n$ et $-\overline{\lim}(a_n) = \underline{\lim}(-a_n)$.

Exercice 1.4.4 Soient $(a_n)_{n \geq 1}$ et $(b_n)_{n \geq 1}$ deux suites réelles. Montrer que les inégalités suivantes sont vraies, pourvu que les quantités de gauche aient un sens

1. $\overline{\lim}(a_n + b_n) \leq \overline{\lim}a_n + \overline{\lim}b_n$.

2. $\overline{\lim}(a_n - b_n) \leq \overline{\lim}a_n - \underline{\lim}b_n$.

3. $\underline{\lim}(a_n + b_n) \geq \underline{\lim}a_n + \underline{\lim}b_n$.

4. $\underline{\lim}(a_n - b_n) \geq \underline{\lim}a_n - \overline{\lim}b_n$.

Exercice 1.4.5 Soient $(a_n)_{n \geq 1}$ et $(b_n)_{n \geq 1}$ deux suites réelles. On suppose que $\lim a_n$ existe et finie, montrer

1. $\overline{\lim}(a_n + b_n) = \lim a_n + \overline{\lim}b_n$.

2. $\underline{\lim}(a_n + b_n) = \lim a_n + \underline{\lim}b_n$.

Exercice 1.4.6 Supposons que $\lim_{n \rightarrow \infty} a_n = a$, trouver les limites suivantes

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n a_i, \quad \lim_{n \rightarrow \infty} \frac{1}{n^2} \sum_{i=1}^n (n-i+1) a_i, \quad \lim_{n \rightarrow \infty} \sum_{i=1}^n \frac{a_i}{2^{n-i}}.$$

Solutions

Solution 1.4.1 1. Par définition de la limite inf,

$$\begin{aligned} x \in \underline{\lim} A_n &\iff \exists n_0 \in \mathbb{N}, \forall n \geq n_0, x \in A_n \\ &\iff \exists n_0 \in \mathbb{N}, \inf_{n \geq n_0} \mathbb{I}_{A_n}(x) = 1 \\ &\iff \underline{\lim} \mathbb{I}_{A_n}(x) = 1, \end{aligned}$$

donc $\underline{\lim} \mathbb{I}_{A_n} = \mathbb{I}_{\underline{\lim} A_n}$. L'égalité $\overline{\lim} \mathbb{I}_{A_n} = \mathbb{I}_{\overline{\lim} A_n}$ se démontre de façon similaire à $\underline{\lim} \mathbb{I}_{A_n} = \mathbb{I}_{\underline{\lim} A_n}$.

2. En effet $\bigcap_{k \geq n} A_k \subset A_k$ pour tout $k \geq n$, donc $\bigcap_{k \geq n} A_k \subset \bigcup_{k \geq n} A_k$ pour tout n , alors pour tout n , $\bigcap_{k \geq n} A_k \subset \bigcap_{n \geq 1} \bigcup_{k \geq n} A_k = \overline{\lim} A_n$. Finalement $\bigcup_{n \geq 1} \bigcap_{k \geq n} A_k \subset \overline{\lim} A_n$.

3. Il suffit d'écrire $\left(\bigcup_{n \geq 1} \bigcap_{k \geq n} A_k \right)^c = \bigcap_{n \geq 1} \bigcup_{k \geq n} A_k^c$ et $\left(\bigcap_{n \geq 1} \bigcup_{k \geq n} A_k \right)^c = \bigcup_{n \geq 1} \bigcap_{k \geq n} A_k^c$.

4. La première égalité provient du fait qu'un élément appartient à une infinité de $A_n \cup B_n$; si et seulement si; il appartient à une infinité de A_n ou bien à une infinité de B_n , et la seconde se démontre, en passant au complémentaire dans $(\overline{\lim} (A_n \cup B_n) = \overline{\lim} A_n \cup \overline{\lim} B_n)$ en utilisant la question 3.

5. Montrons que $\underline{\lim} A_n \cap \underline{\lim} B_n \subset \overline{\lim} (A_n \cap B_n)$. Soit $x \in \underline{\lim} A_n \cap \underline{\lim} B_n$, puisque $x \in \underline{\lim} A_n$ (resp. $x \in \underline{\lim} B_n$) $\exists n_0 \in \mathbb{N}, \forall n \geq n_0, x \in A_n$ (resp. $\exists n_1 \in \mathbb{N}, \forall n \geq n_1, x \in B_n$). Montrons que $x \in \overline{\lim} (A_n \cap B_n)$. Posons $p = \max(n_0, n_1)$, on a : $n \geq p$ et comme $n \geq n_0$ et $n \geq n_1$ donc $x \in A_n \cap B_n$. Maintenant, montrons que $\overline{\lim} (A_n \cap B_n) \subset \overline{\lim} A_n \cap \overline{\lim} B_n$. Soit $x \in \overline{\lim} (A_n \cap B_n)$ alors $x \in \overline{\lim} A_n$ et $x \in \overline{\lim} B_n$. Le reste, en passant au complémentaire dans $(\underline{\lim} A_n \cap \underline{\lim} B_n \subset \overline{\lim} (A_n \cap B_n) \subset \overline{\lim} A_n \cap \overline{\lim} B_n)$ en utilisant la question 3.

6. On a : $\overline{\lim} (A_n \cup B_n) = \overline{\lim} A_n \cup \overline{\lim} B_n$, donc $\mathbb{I}_{\overline{\lim} (A_n \cup B_n)} = \mathbb{I}_{\overline{\lim} A_n \cup \overline{\lim} B_n}$, si $x \in \overline{\lim} A_n \cup \overline{\lim} B_n$ alors $\mathbb{I}_{\overline{\lim} A_n \cup \overline{\lim} B_n}(x) = 1 = \max\{\mathbb{I}_{\underline{\lim} A_n}(x), \mathbb{I}_{\underline{\lim} B_n}(x)\}$, sinon $\mathbb{I}_{\overline{\lim} A_n \cup \overline{\lim} B_n}(x) = 0 = \max(0, 0) = \max\{\mathbb{I}_{\underline{\lim} A_n}(x), \mathbb{I}_{\underline{\lim} B_n}(x)\}$. De la même façon, on a : $\underline{\lim} (A_n \cap B_n) = \underline{\lim} A_n \cap \underline{\lim} B_n$, donc si $x \in \underline{\lim} A_n \cap \underline{\lim} B_n$ alors $\mathbb{I}_{\underline{\lim} (A_n \cap B_n)}(x) = 1 = \inf\{\mathbb{I}_{\overline{\lim} A_n}(x), \mathbb{I}_{\overline{\lim} B_n}(x)\}$, sinon $\mathbb{I}_{\underline{\lim} (A_n \cap B_n)}(x) = 0 = \inf\{\mathbb{I}_{\overline{\lim} A_n}(x), \mathbb{I}_{\overline{\lim} B_n}(x)\}$, d'où le résultat.

7. Comme $\underline{\lim} A_n^c = (\overline{\lim} A_n)^c$ donc $\underline{\lim} A_n^c = A^c$. De même, $\overline{\lim} A_n^c = (\underline{\lim} A_n)^c = A^c$, donc $A_n^c \rightarrow A^c$. D'après la question 3, $\underline{\lim} (A_n \cap B_n) = \underline{\lim} A_n \cap \underline{\lim} B_n = A \cap B$ et d'après la question 5, $A \cap B \subset$

$\overline{\lim}(A_n \cap B_n) \subset A \cap B$ ce qui montre que $A_n \cap B_n \rightarrow A \cap B$. Passons aux complémentaires. Il suffit de voir que, $\forall n$

$$A_n \cup B_n = (A_n^c \cap B_n^c)^c, A_n \setminus B_n = A_n \cap B_n^c \text{ et } A_n \Delta B_n = (A_n \cap B_n^c) \cup (B_n \cap A_n^c),$$

Les résultats obtenus aux les questions précédentes.

8. On a : $\overline{\lim}_n \underline{\lim}_k (A_n \cap A_k^c) = \overline{\lim}_n A_n \cap \underline{\lim}_k A_k^c = \overline{\lim}_n A_n \cap (\overline{\lim}_k A_k)^c = \emptyset$. Comme $A \setminus \overline{\lim} A_n = A \cap (\overline{\lim} A_n)^c = A \cap \underline{\lim} A_n^c$ donc $A \setminus \overline{\lim} A_n = \underline{\lim} (A \cap A_n^c) = \underline{\lim} (A \setminus A_n)$. De la même façon, on montre que $A \setminus \underline{\lim} A_n = \overline{\lim} (A \setminus A_n)$.

9. D'après les questions 4 et 7, on a :

$$\begin{aligned} \overline{\lim}(A \Delta A_n) &= \overline{\lim}((A \setminus A_n) \cup (A_n \setminus A)) \\ &= (\overline{\lim}(A \setminus A_n)) \cup (\overline{\lim}(A_n \setminus A)) \\ &= (A \setminus \underline{\lim} A_n) \cup (\overline{\lim}(A_n \setminus A)). \end{aligned}$$

10. Si $A_n \rightarrow A$ alors $A \setminus \underline{\lim} A_n = \emptyset$, donc $\overline{\lim}(A \Delta A_n) = \overline{\lim}(A_n \setminus A)$.

11. On a : $\overline{\lim} A_n \Delta \underline{\lim} A_n = (\overline{\lim} A_n \setminus \underline{\lim} A_n) \cup (\underline{\lim} A_n \setminus \overline{\lim} A_n) = \overline{\lim} A_n \setminus \underline{\lim} A_n$, car $\underline{\lim} A_n \setminus \overline{\lim} A_n = \underline{\lim} A_n \cap \underline{\lim} A_n^c = \underline{\lim} (A_n \cap A_n^c) = \emptyset$. Comme $\overline{\lim} A_n \Delta \underline{\lim} A_n = (\overline{\lim} A_n \Delta \underline{\lim} A_n) \Delta (A \Delta A)$ donc $\overline{\lim} A_n \Delta \underline{\lim} A_n = (\underline{\lim} A_n \Delta A) \Delta (\overline{\lim} A_n \Delta A)$.

Solution 1.4.2 1. On a : $A_{2n} = [0; n]$ et $A_{2n+1} = [n, +\infty[$ de sorte que (A_{2n}) est croissante, (A_{2n+1}) est décroissante, pour tout n , $\bigcap_{k \geq n} A_k = \emptyset$ et $\bigcup_{k \geq n} A_k = [0; +\infty[$, donc $\underline{\lim} A_n = \emptyset$ et $\overline{\lim} A_n = [0; +\infty[$.

2. On a : $A_{2n} = [-2; 3 + \frac{1}{n}]$ et $A_{2n+1} = [-\frac{1}{n}; 2]$ de sorte que (A_{2n}) est décroissante, (A_{2n+1}) est croissante et $A_{2n+1} \subset A_{2n}$. Ainsi $\underline{\lim} A_n = \bigcup_{n \geq 1} A_{2n+1} = [0; 2]$ et $\overline{\lim} A_n = \bigcap_{n \geq 1} A_{2n} = [-2; 3]$.

3. On a : $B_{2n} = [-1; \frac{n+1}{n}]$ et $B_{2n+1} = [-\frac{n+1}{n}; 1]$ de sorte que (B_{2n}) est décroissante, (B_{2n+1}) est croissante, pour tout $n \geq 1$, $\bigcap_{k \geq n} B_k = [-1; 1]$ et $\bigcup_{k \geq n} B_k = [-1; 1 + \frac{1}{n}]$, donc $\underline{\lim} B_n = [-1; 1]$ et $\overline{\lim} B_n = [-1; 1]$, par conséquent $\underline{\lim} B_n = \overline{\lim} B_n = \lim B_n$.

Solution 1.4.3 1. Pour tout $A \subset \mathbb{R}$, on sait que $\sup(c + A) = c + \sup(A)$. Donc, pour tout n , $\sup_{k \geq n} (c + a_k) = c + \sup_{k \geq n} (a_k)$. Par conséquent, $\overline{\lim}(c + a_n) = c + \overline{\lim} a_n$.

2. De la même façon.

3. Pour tout $A \subset \mathbb{R}$, on sait que $\sup(cA) = \begin{cases} c \sup(A) & \text{if } c \geq 0 \\ -c \inf(-A) & \text{if } c < 0 \end{cases}$. Donc, pour tout n , $\sup_{k \geq n}(ca_k) =$

$\begin{cases} c \sup_{k \geq n}(a_k) & \text{if } c \geq 0 \\ c \inf_{k \geq n}(-a_k) & \text{if } c < 0 \end{cases}$. Par conséquent, en utilisant la définition de limite inférieure et de limite

supérieure, nous obtenons $\overline{\lim}(ca_n) = \begin{cases} \overline{\lim}a_n & \text{if } c \geq 0 \\ \underline{\lim}a_n & \text{if } c < 0 \end{cases}$.

4. De la même façon.

5. On pose $c = -1 < 0$, d'où le résultat.

Solution 1.4.4 1. Pour tout $n \in \mathbb{N}$, $\{a_k + b_k, k > n\} \subset \{a_m + b_k, m, k > n\}$. Donc, pour tout $n \in \mathbb{N}$

$$\sup_{k > n}(a_k + b_k) \leq \sup_{m, k > n}(a_m + b_k) = \sup_{m > n}(a_m) + \sup_{k > n}(b_k),$$

et par définition de la limite supérieure $\overline{\lim}(a_n + b_n) \leq \overline{\lim}a_n + \overline{\lim}b_n$.

2. Comme $\overline{\lim}b_n = -\underline{\lim}(-b_n)$, donc $\overline{\lim}(a_n - b_n) \leq \overline{\lim}a_n + \overline{\lim}(-b_n) \leq \overline{\lim}a_n - \underline{\lim}b_n$.

3. Par la première question, et depuis $\underline{\lim}b_n = -\overline{\lim}(-b_n)$, on a :

$$\begin{aligned} \underline{\lim}(a_n + b_n) &= -\overline{\lim}(-a_n - b_n) \\ &\geq -\overline{\lim}(-a_n) - \overline{\lim}(-b_n) = \underline{\lim}a_n + \underline{\lim}b_n. \end{aligned}$$

4. Comme $\underline{\lim}b_n = -\overline{\lim}(-b_n)$, donc $\underline{\lim}(a_n - b_n) \geq \underline{\lim}a_n + \underline{\lim}(-b_n) \geq \underline{\lim}a_n - \overline{\lim}b_n$.

Solution 1.4.5 1. On a : $\overline{\lim}(a_n + b_n) \leq \overline{\lim}(a_n) + \overline{\lim}(b_n) = \lim a_n + \overline{\lim}b_n$. Nous voulons montrer l'inégalité inverse,

$$\overline{\lim}(b_n) = \overline{\lim}(a_n + b_n - a_n) \leq \overline{\lim}(a_n + b_n) - \underline{\lim}(a_n),$$

donc $\overline{\lim}(b_n) + \lim(a_n) = \overline{\lim}(b_n) + \underline{\lim}(a_n) \leq \overline{\lim}(a_n + b_n)$. D'où le résultat.

2. De la même façon.

Solution 1.4.6 D'après le lemme de Cesàro, $\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n a_i = a$ et $\lim_{n \rightarrow \infty} \frac{1}{n^2} \sum_{i=1}^n (n-i+1) a_i = 0$.

D'après le lemme de Toeplitz, $\lim_{n \rightarrow \infty} \sum_{i=1}^n \frac{a_i}{2^{n-i}} = a$.

Chapitre 2

Modes de convergence

L'étude de suites de variables aléatoires ainsi que de leur convergence est une question très naturelle en probabilité. Pour cela, nous nous intéresserons à préciser la notion de convergence, et les propriétés qui en découlent. Le but de ce chapitre est d'étudier les différents types de convergence pour une suite $(\underline{X}_n)_{n \geq 1}$ de variables aléatoires de \mathbb{R}^k . Nous distinguerons alors quatre types de convergence, la convergence presque sûre, en probabilité, en loi et en moyenne d'ordre $p > 0$, ainsi que les relations entre ces divers modes de convergence ; la notion d'équi-intégrabilité est introduite à cette fin.

Dans toute la suite, nous supposons que toutes les variables aléatoires sont définies sur un espace probabilisé (Ω, \mathcal{A}, P) et à valeurs dans $(\mathbb{R}^k, \mathcal{B}_{\mathbb{R}^k})$.

2.1 Convergence presque sûre

La première section commence par donner la notion de la convergence stochastique la plus simple et proche à la convergence usuelle qu'est la convergence presque sûre ou la convergence avec probabilité 1 ou forte.

Définition 2.1.1 Soit $(\underline{X}_n, \underline{X})_{n \geq 1}$ une suite de variables aléatoires de \mathbb{R}^k est dite convergente vers $\underline{X} \in \mathbb{R}^k$ presque sûrement s'il existe une partie $A \in \mathcal{A}$ négligeable (i.e., $P(A) = 0$) et $\underline{X}_n(\omega) \xrightarrow[n \rightarrow \infty]{} \underline{X}(\omega), \forall \omega \in \Omega \setminus A$, ou encore $P\left(\underline{X}_n \xrightarrow[n \rightarrow \infty]{} \underline{X}\right) = 1$, ou encore $P\left(\left\{\omega \in \Omega : \underline{X}_n(\omega) \xrightarrow[n \rightarrow \infty]{} \underline{X}(\omega)\right\}\right) = 1$. Dans ce cas, on note $\lim_{n \rightarrow \infty} \underline{X}_n = \underline{X}$ p.s. ou $\underline{X}_n \xrightarrow[n \rightarrow \infty]{} \underline{X}$ p.s. lorsque $n \rightarrow \infty$.

Exemple 2.1.1 Soit $([0, 1], \mathcal{B}_{[0,1]}, \lambda)$ (λ : mesure de Lebesgue) et soit

$$X_n(\omega) = \begin{cases} n & \text{si } \omega \in [0, \frac{1}{n}] \\ 0 & \text{si } \omega \in]\frac{1}{n}, 1] \end{cases} .$$

Alors nous disons $X_n \xrightarrow[n \rightarrow \infty]{p.s.} 0$. En effet, pour $A = \{0\}$, $\forall \omega \in A^c = [0, 1] \setminus A : X_n(\omega) \xrightarrow[n \rightarrow \infty]{} 0$, $X_n(0) \not\rightarrow 0$, parce que $X_n(0) = n$.

Remarque 2.1.1 Notons que l'ensemble $B = \left\{ \underline{X}_n \xrightarrow[n \rightarrow \infty]{} \underline{X} \right\}$ est bien un évènement car la convergence $\underline{X}_n \rightarrow \underline{X}$ p.s. s'écrit pour presque chaque $\omega \in \Omega : \forall \varepsilon > 0, \exists n_0 \in \mathbb{N}, \forall n \geq n_0 : |\underline{X}_n(\omega) - \underline{X}(\omega)| < \varepsilon$. On peut donc écrire $B = \bigcap_{m \geq 1} \bigcup_{k \geq 0} \bigcap_{n \geq k} \{|\underline{X}_n - \underline{X}| < \frac{1}{m}\} \in \mathcal{A}$.

La convergence presque sûre est facile à caractériser. En effet,

Remarque 2.1.2 (Les liens entre la convergence presque sûre et les autres)

Une suite de variables aléatoires $(\underline{X}_n)_{n \geq 1}$ convergente vers \underline{X} presque sûrement si et seulement si $\forall m \geq 1 P(\liminf_n \{|\underline{X}_n - \underline{X}| \leq \frac{1}{m}\}) = 1$ c'est à dire

$$\begin{aligned} \forall \varepsilon > 0, P\left(\liminf_n \{|\underline{X}_n - \underline{X}| \leq \varepsilon\}\right) = 1 &\iff \forall \varepsilon > 0, P\left(\limsup_n \{|\underline{X}_n - \underline{X}| > \varepsilon\}\right) = 0 \\ &\iff \forall \varepsilon > 0, P(\{|\underline{X}_n - \underline{X}| > \varepsilon\}, i.o.) = 0. \end{aligned}$$

Parce que pour une suite d'évènements $(B_n)_{n \geq 1}$, $P(\bigcap_{n \geq 1} B_n) = 1$ est équivalent à $P(B_n) = 1$ pour tout $n \geq 1$. De la même façon, en partant du critère de Cauchy, une suite de variables aléatoires $(\underline{X}_n)_{n \geq 1}$ converge presque sûre si et seulement si

$$\forall \varepsilon > 0, P\left(\liminf_n \{|\underline{X}_n - \underline{X}_m| \leq \varepsilon\}\right) = 1 \iff \forall \varepsilon > 0, P\left(\limsup_n \{|\underline{X}_n - \underline{X}_m| > \varepsilon\}\right) = 0.$$

Exemple 2.1.2 On considère $(X_n)_{n \geq 1}$ une suite de variables aléatoires i.i.d. de loi de Bernoulli de paramètre $p \in (0; 1)$. Alors la suite $Y_n := \sum_{i=1}^n 2^{-i} X_i$ converge presque sûrement d'après le critère de Cauchy. En effet, si $m < n$, $|Y_n - Y_m| \leq \sum_{i=m+1}^n 2^{-i} \leq 2^{-m}$, et donc

$$\forall \varepsilon > 0, P\left(\liminf_n \{\omega, |Y_n(\omega) - Y_m(\omega)| \leq \varepsilon\}\right) \geq P\left(\liminf_n \{\omega, 2^{-m} \leq \varepsilon\}\right) = P\left(\bigcup_{m \geq 1} \{\omega, 2^{-m} \leq \varepsilon\}\right) = P(\Omega).$$

Propriétés

La convergence presque sûre est stable par composition de fonction continue (somme, multiplication par un scalaire, produit, etc), couple, produit scalaire et équivalente à la convergence presque sûre des composantes.

Proposition 2.1.1 Soit (\underline{X}_n) une suite dans \mathbb{R}^k t.q $\underline{X}_n \xrightarrow[n \rightarrow \infty]{p.s} \underline{X}$ et $\underline{X}_n \xrightarrow[n \rightarrow \infty]{p.s} \underline{Y}$, alors $\underline{X} = \underline{Y}$ p.s.

Preuve. Il existe deux parties A, B négligeables $P(A) = P(B) = 0$ et $\forall \omega \in (A \cup B)^c, \underline{X}_n(\omega) \xrightarrow[n \rightarrow \infty]{} \underline{X}(\omega)$ et $\underline{X}_n(\omega) \xrightarrow[n \rightarrow \infty]{} \underline{Y}(\omega)$. D'où, $\forall \omega \in (A \cup B)^c, \underline{X}(\omega) = \underline{Y}(\omega), P(\underline{X} = \underline{Y}) = 1$. ■

Proposition 2.1.2 Soit (\underline{X}_n) une suite dans \mathbb{R}^k t.q $\underline{X}_n \xrightarrow[n \rightarrow \infty]{p.s} \underline{X}$ et soit $f : \mathbb{R}^k \rightarrow \mathbb{R}^p$ une fonction continue sur une partie $S \in \mathcal{B}_{\mathbb{R}^k} : P(\underline{X} \in S) = 1$, alors $f(\underline{X}_n) \xrightarrow[n \rightarrow \infty]{p.s} f(\underline{X})$.

Preuve. Par définition, $\exists A \in \mathcal{A}, P(A) = 0$ et $\underline{X}_n(\omega) \xrightarrow[n \rightarrow \infty]{} \underline{X}(\omega), \forall \omega \in A^c$ et $\exists B \in \mathcal{A}, P(B) = 0. \forall \omega \in B, f$ est continue au point $\underline{X}(\omega)$, donc $f(\underline{X}_n(\omega)) \xrightarrow[n \rightarrow \infty]{} f(\underline{X}(\omega)) \forall \omega \in (A \cup B)^c$, i.e., $f(\underline{X}_n) \xrightarrow[n \rightarrow \infty]{p.s} f(\underline{X})$. ■

Exemple 2.1.3 Soit $g : \mathbb{R} \rightarrow \mathbb{R}$ t.q

$$g(t) = \begin{cases} t - 1 & \text{si } t < 0 \\ t + 1 & \text{si } t \geq 0 \end{cases}$$

et soient $X_n(\omega) = -\frac{1}{n}$ et $X(\omega) = 0$. Alors, d'une part $g(X_n) = -\frac{1}{n} - 1 \rightarrow -1$ et d'autre part $g(0) = 1$.

Proposition 2.1.3 Si (\underline{X}_n) et (\underline{Y}_n) deux suites dans \mathbb{R}^k et dans \mathbb{R}^l respectivement, alors

$$\left(\begin{pmatrix} \underline{X}_n \\ \underline{Y}_n \end{pmatrix} \xrightarrow[n \rightarrow \infty]{p.s} \begin{pmatrix} \underline{X} \\ \underline{Y} \end{pmatrix} \right) \iff \left(\begin{array}{c} \underline{X}_n \xrightarrow[n \rightarrow \infty]{p.s} \underline{X} \\ \underline{Y}_n \xrightarrow[n \rightarrow \infty]{p.s} \underline{Y} \end{array} \right).$$

Proposition 2.1.4 Si (\underline{X}_n) et (\underline{Y}_n) deux suites dans \mathbb{R}^k t.q $\underline{X}_n \xrightarrow[n \rightarrow \infty]{p.s} \underline{X}$ et $\underline{Y}_n \xrightarrow[n \rightarrow \infty]{p.s} \underline{Y}$, alors

- $\underline{X}_n \mp \underline{Y}_n \xrightarrow[n \rightarrow \infty]{p.s} \underline{X} \mp \underline{Y}$.
- $\underline{X}'_n \underline{Y}_n \xrightarrow[n \rightarrow \infty]{p.s} \underline{X}' \underline{Y}$.
- Si A est une matrice, alors $A \underline{X}_n \xrightarrow[n \rightarrow \infty]{p.s} A \underline{X}$.
- Si $k = 1, \frac{\underline{X}_n}{\underline{Y}_n} \xrightarrow[n \rightarrow \infty]{p.s} \frac{\underline{X}}{\underline{Y}}$, avec $P(\underline{Y}_n = 0) = P(\underline{Y} = 0) = 0$.

Un des outils classiques pour montrer la convergence presque sûre est le lemme de Borel-Cantelli.

Lemme 2.1.1 [Borel – Cantelli] Soit (A_n) une suite d'évènements. Si $\sum_n P(A_n) < +\infty$ alors $P(A_n, i.o.) = 1$. Autrement dit, avec une probabilité égale à 1, au plus un nombre fini d'évènements A_n se réalise.

De plus, si les évènements A_n sont indépendants et si $\sum_n P(A_n) = +\infty$ alors $P(A_n, i.o.) = 1$. Autrement dit, avec une probabilité égale à 1, une infinité d'évènements A_n se réalise.

Preuve. $P(\limsup A_n) = \lim_{n \rightarrow \infty} P\left(\bigcup_{m \geq n} A_m\right) \leq \lim_{n \rightarrow \infty} \sum_{m \geq n} P(A_m) = 0$. Supposons que les évènements A_n sont indépendants et que $\sum_n P(A_n) = +\infty$, comme $(\limsup A_n)^c = \left(\bigcap_{n \geq 1} \bigcup_{m \geq n} A_m\right)^c$ alors

$$P((\limsup A_n)^c) = \lim_{n \rightarrow \infty} P\left(\bigcap_{m \geq n} A_m^c\right) = \lim_{n \rightarrow \infty} \prod_{m=n}^{\infty} (1 - P(A_m)) \leq \lim_{n \rightarrow \infty} e^{-\sum_{m=n}^{\infty} P(A_m)} = 0,$$

donc $P(\limsup A_n) = 1$. ■

Corollaire 2.1.1 Soit $(\underline{X}_n, \underline{X})_{n \geq 1}$ une suite de variables aléatoires de \mathbb{R}^k .

1. Si pour tout $\varepsilon > 0$, $\sum_n P(|\underline{X}_n - \underline{X}| > \varepsilon) < +\infty$, alors $\underline{X}_n \xrightarrow[n \rightarrow \infty]{p.s.} \underline{X}$.
2. Si les $(\underline{X}_n)_{n \geq 1}$ sont mutuellement indépendantes, alors $\underline{X}_n \xrightarrow[n \rightarrow \infty]{p.s.} \underline{Q}$ si et seulement si pour tout $\varepsilon > 0$, $\sum_n P(|\underline{X}_n| > \varepsilon) < +\infty$.

Preuve. Si pour tout $\varepsilon > 0$, $\sum_n P(|\underline{X}_n - \underline{X}| > \varepsilon) < +\infty$, alors le résultat découle du lemme de Borel Cantelli. Le reste se démontre de façon analogue à partir de la partie indépendante du lemme de Borel-Cantelli. ■

Proposition 2.1.5 Soit $(\underline{X}_n, \underline{X})_{n \geq 1}$ une suite de variables aléatoires de \mathbb{R}^k alors

$$\underline{X}_n \xrightarrow[n \rightarrow \infty]{p.s.} \underline{X} \iff (\underline{X}_n)_{n \geq 1} \text{ est de Cauchy p.s.}$$

2.2 Convergence en probabilité

La convergence en probabilité, appelée aussi convergence en mesure, ou dans $\mathbb{L}_0(\Omega, \mathcal{A}, P)$, est définie comme suit

Définition 2.2.1 Soit (X_n) une suite de v.a.r. et soit X une v.a.r., définies sur le même espace probabilisé (Ω, \mathcal{A}, P) . On dit que (X_n) converge en probabilité vers X et on écrit $X_n \xrightarrow[n \rightarrow \infty]{P} X$ ($p \lim_{n \rightarrow \infty} X_n = X$ ou

$X_n - X = o_p(1)$ si $\forall \varepsilon > 0, \lim_{n \rightarrow \infty} P(|X_n - X| > \varepsilon) = 0$ (ou $\forall \varepsilon > 0, \lim_{n \rightarrow \infty} P(|X_n - X| \leq \varepsilon) = 1$). i.e.,

$$\forall \delta > 0, \forall \varepsilon > 0, \exists n_0, \forall n \geq n_0 : P(|X_n - X| > \varepsilon) < \delta.$$

Exemple 2.2.1 Soit (X_n) une suite de v.a. On suppose que $X_n(\Omega) = \{-1; 0; 1\}$ avec

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

alors $X_n \xrightarrow[n \rightarrow \infty]{P} 0$. En effet, $\forall \varepsilon > 0, P(|X_n| > \varepsilon) = P(|X_n| = 1) = \frac{1}{n} \xrightarrow[n \rightarrow \infty]{} 0$.

Définition 2.2.2 Soient (X_n) et (Y_n) deux suites de v.a.r positives, définies sur le même espace probabilisé (Ω, \mathcal{A}, P) .

1. On dit que (X_n) est négligeable par rapport à (Y_n) en probabilité et on écrit $X_n = o_p(Y_n)$ si $p \lim_{n \rightarrow \infty} \frac{X_n}{Y_n} = 0$.
2. On dit que (X_n) est au plus du même ordre que (Y_n) en probabilité et on écrit $X_n = O_p(Y_n)$ si

$$\forall \varepsilon > 0, \exists M_{(\varepsilon)} > 0, P(|X_n| > M_{(\varepsilon)} Y_n) \leq \varepsilon, \forall n.$$

On dit souvent X_n est bornée en probabilité par Y_n .

Propriétés

La convergence en probabilité est stable par les opérations algébriques usuelles.

Proposition 2.2.1 Soient (X_n) et (Y_n) deux suites de v.a.r. et (f_n) et (g_n) sont deux suites réelles positives.

1. Si $X_n = o_p(f_n)$ et $Y_n = o_p(g_n)$, alors
 - a. $X_n Y_n = o_p(f_n g_n)$,
 - b. $|X_n^s| = o_p(f_n^s), s > 0$,
 - c. $X_n + Y_n = o_p(\max(f_n, g_n)) = o_p(f_n + g_n)$.
2. Si $X_n = O_p(f_n)$ et $Y_n = O_p(g_n)$, alors
 - a. $X_n Y_n = O_p(f_n g_n)$,
 - b. $|X_n^s| = O_p(f_n^s), s > 0$,

c. $X_n + Y_n = O_p(\max(f_n, g_n)) = O_p(f_n + g_n)$.

3. Si $X_n = O_p(f_n)$ et $Y_n = o_p(g_n)$ alors $X_n Y_n = o_p(f_n g_n)$.

Définition 2.2.3 Soit (\underline{X}_n) une suite de k -vecteurs aléatoires et soit (Y_n) une suite réelle positive.

1. $\underline{X}_n = O_p(Y_n) \Leftrightarrow X_n(j) = O_p(Y_n)$. i.e.,

$$\forall \varepsilon > 0, \exists M_{(\varepsilon)} > 0, P(X_n(j) > M_{(\varepsilon)} Y_n) < \varepsilon, j = 1, 2, \dots, k.$$

2. $\underline{X}_n = o_p(Y_n) \Leftrightarrow X_n(j) = o_p(Y_n)$. i.e.,

$$\forall \varepsilon > 0, \forall \delta > 0, \exists n_0(\varepsilon), \forall n \geq n_0(\varepsilon) : P(|X_n(j)| > Y_n \varepsilon) < \delta.$$

Définition 2.2.4 Soit (X_n) une suite de $k \times l$ -matrices aléatoires et soit (Y_n) une suite réelle positive.

Alors,

1. $X_n = O_p(Y_n) \Leftrightarrow X_n(i, j) = O_p(Y_n)$,

2. $X_n = o_p(Y_n) \Leftrightarrow X_n(i, j) = o_p(Y_n)$.

Lemme 2.2.1 Soit $(\underline{X}_i)_{1 \leq i \leq n}$ une suite de k -vecteurs aléatoires. Alors,

$$\forall \varepsilon > 0, P\left(\left\|\sum_{i=1}^n X_i\right\| > \varepsilon\right) \leq \sum_{i=1}^n P\left(\|X_i\| > \frac{\varepsilon}{n}\right).$$

Preuve. On a : $\sum_{i=1}^n \|X_i\| \geq \varepsilon \Rightarrow \exists i_0 \in \{1, \dots, n\} : \|X_{i_0}\| \geq \frac{\varepsilon}{n}$, donc

$$P\left(\left\|\sum_{i=1}^n X_i\right\| \geq \varepsilon\right) \leq P\left(\sum_{i=1}^n \|X_i\| \geq \varepsilon\right) \leq \sum_{i=1}^n P\left(\|X_i\| > \frac{\varepsilon}{n}\right).$$

■

Lemme 2.2.2 Soit (\underline{X}_n) une suite de k -vecteurs aléatoires t.q $X_n(j) \xrightarrow[n \rightarrow \infty]{P} X(j), j = 1, \dots, k$. Alors, $\underline{X}_n \rightarrow \underline{X} = (X(1), \dots, X(k))'$.

Preuve. Par hypothèse $\forall j \in \{1, \dots, k\}, \forall \delta > 0, \forall \varepsilon > 0, \exists N_j > 0, \forall n \geq N_j : P\left(|X_n(j) - X(j)| > \frac{\sqrt{\varepsilon}}{\sqrt{k}}\right) < \frac{\delta}{k}$.

Soit $N_0 = \max\{N_1, \dots, N_k\}$, alors on a :

$$\begin{aligned} P(|\underline{X}_n - \underline{X}| > \varepsilon) &= P\left(\sum_{j=1}^k |X_n(j) - X(j)|^2 > \varepsilon^2\right) \\ &\leq \sum_{j=1}^k P\left(|X_n(j) - X(j)|^2 > \frac{\varepsilon^2}{k}\right) \\ &\leq \sum_{j=1}^k \frac{\delta}{k} = \delta. \end{aligned}$$

■

L'inégalité de Bienaymé-Tchebychev donne une majoration de la probabilité que l'écart soit grand.

Théorème 2.2.1 [L'inégalité de Bienaymé-Tchebychev] Soient $r > 0$ et X une v.a. t.q. $E\{|X|^r\} < +\infty$.

Alors,

$$\forall \varepsilon > 0, \forall a \in \mathbb{R} : P(|X - a| > \varepsilon) \leq \frac{E\{|X - a|^r\}}{\varepsilon^r}.$$

Preuve. Soit $S = \{x \in \mathbb{R} : |x - a| > \varepsilon\}$. $F_X(\cdot)$ la fonction de répartition de X , alors on a :

$$E\{|X - a|^r\} = \int_{\mathbb{R}} |x - a|^r dF_X(x) \geq \varepsilon^r \int_S dF_X(x) = \varepsilon^r P(S).$$

■

Corollaire 2.2.1 Soit (X_n) une suite de v.a. et soit (a_n) une suite de nombres réels positifs. Si $E\{X_n^2\} = O(a_n^2)$, alors $X_n = O_p(a_n)$.

Preuve. Par hypothèse, $\exists M_1 > 0 : E\{X_n^2\} \leq M_1^2 a_n^2$. D'après l'inégalité de Tchebychev on a :

$$P(|X_n| > M_1 a_n) \leq \frac{E\{X_n^2\}}{M_1^2 a_n^2}.$$

Soit $\varepsilon > 0$, $M_2 \geq \frac{M_1}{\sqrt{\varepsilon}}$ alors

$$P(|X_n| > M_2 a_n) \leq P\left(|X_n| > \frac{M_1 a_n}{\sqrt{\varepsilon}}\right) \leq \frac{E\{X_n^2\}}{M_1^2 a_n^2} \varepsilon \leq \varepsilon.$$

■

Corollaire 2.2.2 Soit (X_n) une suite de v.a.r. et soit (a_n) une suite de nombres réels positifs. Alors

$$\left. \begin{array}{l} Var(X_n) = O(a_n^2) \\ E\{X_n\} = o(a_n) \end{array} \right\} \Rightarrow X_n = O_p(a_n).$$

Preuve. On a : $E\{X_n^2\} = Var(X_n) + (E\{X_n\})^2 = O(a_n^2)$, donc $X_n = O_p(a_n)$. ■

La limite en probabilité d'une suite de variables aléatoires est presque sûrement unique, c'est à dire

Proposition 2.2.2 [Unicité p.s. de la limite en probabilité] Soient (X_n) et (Y_n) deux suites de v.a.r. t.q. $|X_n - Y_n| = o_p(1)$. S'il existe une v.a.r. t.q. $X_n - X = o_p(1)$, alors $Y_n - X = o_p(1)$.

Preuve. Soit $\varepsilon > 0, \delta > 0, \exists n_0 > 0, \forall n \geq n_0$:

$$P\left(|X_n - Y_n| > \frac{\varepsilon}{2}\right) < \frac{\delta}{2} \text{ et } P\left(|X_n - X| > \frac{\varepsilon}{2}\right) < \frac{\delta}{2}.$$

On a :

$$\begin{aligned} P(|Y_n - X| > \varepsilon) &= P(|Y_n - X_n + X_n - X| > \varepsilon) \\ &\leq P\left(|Y_n - X_n| > \frac{\varepsilon}{2}\right) + P\left(|X_n - X| > \frac{\varepsilon}{2}\right) < \frac{\delta}{2} + \frac{\delta}{2} = \delta. \end{aligned}$$

■

Proposition 2.2.3 Soit (\underline{X}_n) une suite de vecteurs aléatoires de \mathbb{R}^k telle que $\underline{X}_n \xrightarrow[n \rightarrow \infty]{P} \underline{X}$ où $\underline{X} \in \mathbb{R}^k$. Soit $g : \mathbb{R}^k \rightarrow \mathbb{R}^p$ une fonction continue. Alors, $g(\underline{X}_n) \xrightarrow[n \rightarrow \infty]{P} g(\underline{X})$.

Preuve. [Cas scalaire]

$$g \text{ continue en } x_0 \Leftrightarrow \forall \varepsilon > 0, \exists \delta(\varepsilon) > 0 : (|g(x) - g(x_0)| \geq \varepsilon \Rightarrow |x - x_0| \geq \delta).$$

Pour n assez grand :

$$P(|g(X_n) - g(X)| \geq \varepsilon) \leq P(|X_n - X| \geq \delta).$$

D'où le résultat. ■

Exemple 2.2.2 Soit $(X_i)_{1 \leq i \leq n}$ une suite de v.a.r. i.i.d. t.q $E\{X_1\} = \mu$ et $Var(X_1) = \sigma^2$. Alors, $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i \xrightarrow[n \rightarrow \infty]{P} \mu$. En effet, soit $\varepsilon > 0$, on a

$$P(|\bar{X}_n - \mu| > \varepsilon) \leq \frac{1}{\varepsilon^2} E\left\{(\bar{X}_n - \mu)^2\right\} = \frac{\sigma^2}{n} \rightarrow 0.$$

Donc $\bar{X}_n \xrightarrow[n \rightarrow \infty]{P} \mu$.

Proposition 2.2.4 Soient (X_n) et (Y_n) deux suites de v.a.r. convergeant en probabilité vers X et Y respectivement. Alors,

1. $X_n \pm Y_n \xrightarrow[n \rightarrow \infty]{P} X \pm Y$;
2. $X_n Y_n \xrightarrow[n \rightarrow \infty]{P} XY$,
3. $\frac{X_n}{Y_n} \xrightarrow[n \rightarrow \infty]{P} \frac{X}{Y}$ si $Y_n \neq 0$ p.s, $Y \neq 0$ p.s.

On a un premier résultat sur le lien entre les modes de convergence.

Proposition 2.2.5 *La convergence presque sûre implique la convergence en probabilité.*

Preuve. Par le lemme de Fatou, on a pour tout $\varepsilon > 0$:

$$\liminf_n P(|X_n - X| < \varepsilon) \geq P\left(\liminf_n \{|X_n - X| < \varepsilon\}\right) = 1.$$

■

La réciproque de ce résultat est fautive comme le montre le contre-exemple suivant.

Exemple 2.2.3 [*Contre-exemple*] *On considère une suite de variables aléatoires réelles indépendantes (X_n) de loi de Bernoulli de paramètre $\frac{1}{n}$. Alors puisque pour $0 < \varepsilon < 1$,*

$$P(|X_n| > \varepsilon) = P(X_n = 1) = \frac{1}{n} \xrightarrow{n \rightarrow \infty} 0,$$

donc la suite (X_n) converge en probabilité vers 0, mais (X_n) ne converge pas presque sûrement vers 0. En effet,

$$\sum_{n \geq 1} P(|X_n| > \varepsilon) = \sum_{n \geq 1} \frac{1}{n} = +\infty.$$

Comme nous venons de le voir la convergence en probabilité n'implique pas la convergence presque sûre. Toutefois, on a le résultat important suivant.

Corollaire 2.2.3 *Si la suite $(X_n)_{n \geq 1}$ converge en probabilité vers X , on peut en extraire une sous-suite $(X_{n_i})_{i \geq 1}$ qui converge presque sûrement.*

Il est parfois bien utile de disposer de critères de type Cauchy pour établir la convergence d'une suite de variables aléatoires.

Définition 2.2.5 *On dit que (X_n) est une suite de variables aléatoires réelles vérifie le critère de Cauchy en probabilité si*

$$\forall \varepsilon > 0, \exists n_0, \forall n \geq m \geq n_0 : P(|X_n - X_m| > \varepsilon) < \varepsilon.$$

Proposition 2.2.6 *Une suite de variables aléatoires (X_n) converge en probabilité si et seulement si X_n satisfait le critère de Cauchy en probabilité.*

2.3 Convergence en loi

Nous passons à présent à une notion de convergence d'un autre type : la convergence en loi. C'est la plus faible, mais peut-être aussi la plus importante. Elle est souvent utilisée dans les applications.

Définition 2.3.1 Soit (X_n, X) une suite de v.a.r. définie sur (Ω, \mathcal{A}, P) , de fonction de répartition $F_n(\cdot)$ (resp. $F(\cdot)$) de X_n (resp. de X) ($\forall x \in \mathbb{R}, F_n(x) = P(X_n \leq x)$). La suite (X_n) est dite convergente en loi vers X , et on écrit $X_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} X$, si et seulement si $F_n(x) \xrightarrow[n \rightarrow \infty]{} F(x), \forall x$ point de continuité de F .

Exemple 2.3.1 Soit $(X_i)_{1 \leq i \leq n}$ une suite de v.a.r. i.i.d. de loi uniforme sur $]0, \theta[$. Soit $t_n = \inf_{1 \leq i \leq n} X_i$, alors on sait que la densité de t_n est $g_n(t) = nt^{n-1}\theta^n \mathbb{I}_{\{0 < t < \theta\}}$ et de fonction de répartition $G_n(t) = (\frac{t}{\theta})^n \mathbb{I}_{\{0 < t < \theta\}} + \mathbb{I}_{\{t \geq \theta\}}$. Posons $U_n = n(\frac{\theta - t_n}{\theta})$. Alors, $F_n(u) = P(U_n \leq u) = \begin{cases} 0 & \text{si } u \leq 0 \\ 1 - (1 + \frac{u}{n})^n & \text{si } u > 0 \end{cases}$

et $\lim_{n \rightarrow \infty} F_n(u) = \begin{cases} 0 & \text{si } u \leq 0 \\ 1 - e^{-u} & \text{si } u > 0 \end{cases}$. Soit U la v.a de densité $f_U(u) = e^{-u} \mathbb{I}_{\{u > 0\}}$ et de fonction de répartition $F_U(u) = (1 - e^{-u}) \mathbb{I}_{\{u > 0\}}$. Il est clair que $\forall u \in]-\infty, 0[\cup]0, +\infty[, F_n(u) \xrightarrow[n \rightarrow \infty]{} F(u)$.

Définition 2.3.2 On appelle fonction caractéristique d'une v.a.r. X et on la note $\varphi_X(\cdot)$, la fonction $\varphi_X(\cdot)$ définie par $\forall t \in \mathbb{R} : \varphi_X(t) = E \{e^{itX}\} = \int_{\mathbb{R}} e^{itx} dF_X(x)$.

Remarque 2.3.1 Si \underline{X} est un vecteur de \mathbb{R}^k alors

$$\forall t \in \mathbb{R}^k : \varphi_{\underline{X}}(t) = E \left\{ e^{i \langle t, \underline{X} \rangle} \right\} = \int_{\mathbb{R}^k} e^{i \langle t, \underline{x} \rangle} dF(\underline{x}) = \int_{\mathbb{R}^k} e^{i \sum_{i=1}^k t_i x_i} dF(x_1, \dots, x_k).$$

Exemple 2.3.2 1. Si X suit une loi discrète, alors la fonction caractéristique φ_X de X est définie par $\varphi_X(t) = \sum_{x \in E} e^{itx} P(X = x)$. En particulier, si $X \rightsquigarrow \mathcal{B}(p)$ alors $\varphi_X(t) = 1 - p + pe^{it}$, si $X \rightsquigarrow \mathcal{B}(n, p)$ alors $\varphi_X(t) = (1 - p + pe^{it})^n$ et si $X \rightsquigarrow P(\lambda)$ alors $\varphi_X(t) = \exp \{ \lambda (e^{it} - 1) \}$.

2. Lorsque X admet une densité f sur \mathbb{R} , alors $\forall x \in \mathbb{R} : \varphi_X(t) = \int_{\mathbb{R}} e^{itx} f(x) dx$. En particulier si $X \rightsquigarrow \mathcal{N}(0, 1)$ alors $\varphi_X(t) = e^{-t^2/2}$.

Nous relierons à présent la convergence en loi à celle des fonctions caractéristiques.

Théorème 2.3.1 Soit (X_n) une suite de v.a.r. et soit X une v.a.r, définies sur le même espace probabilisé (Ω, \mathcal{A}, P) . Si $\varphi_{X_n}(t) \xrightarrow[n \rightarrow \infty]{} \varphi_X(t)$ pour tout $t \in \mathbb{R}$, alors $X_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} X$.

On donne maintenant un critère de convergence en loi pour les variables aléatoires discrètes.

Proposition 2.3.1 Soit (X_n, X) une suite de v.a.r. On a l'équivalence

$$X_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} X \iff \forall k, P(X_n = k) \xrightarrow[n \rightarrow \infty]{} P(X = k).$$

Exemple 2.3.3 Soit (X_n) une suite de v.a. de loi binomiale de paramètre $\mathcal{B}(n; \lambda/n)$ et X une v.a. de Poisson de paramètre λ . On a : $X_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} X$. En effet, pour k ,

$$P(X_n = k) = \frac{\lambda^k}{k!} \frac{n!}{(n-k)!n^k} \left(1 - \frac{\lambda}{n}\right)^{n-k} \xrightarrow[n \rightarrow \infty]{} \frac{e^{-\lambda} \lambda^k}{k!} = P(X = k),$$

car, le second terme du produit tend vers 1, le 3^{ème} vers $e^{-\lambda}$.

Propriétés

Théorème 2.3.2 [Recette de Cramer] Soit (\underline{X}_n) une suite dans \mathbb{R}^k , alors

$$\underline{X}_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \underline{X} \iff \forall \underline{\lambda} \in \mathbb{R}^k, \underline{\lambda}' \underline{X}_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \underline{\lambda}' \underline{X}.$$

Preuve. On a : pour tout $t \in \mathbb{R}$, $\varphi_{\underline{\lambda}' \underline{X}_n}(t) = E \left\{ e^{it \underline{\lambda}' \underline{X}_n} \right\} = \varphi_{\underline{X}_n}(\underline{\lambda} t) \xrightarrow[n \rightarrow \infty]{} \varphi_{\underline{X}}(\underline{\lambda} t) = \varphi_{\underline{\lambda}' \underline{X}}(t)$. Ceci implique que $\underline{\lambda}' \underline{X}_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \underline{\lambda}' \underline{X}$. Inversement, si $\underline{\lambda}' \underline{X}_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \underline{\lambda}' \underline{X}, \forall \underline{\lambda} \in \mathbb{R}^k$,

$$\varphi_{\underline{X}_n}(\underline{\lambda}) = E \left\{ e^{i \underline{\lambda}' \underline{X}_n} \right\} = \varphi_{\underline{\lambda}' \underline{X}_n}(1) \xrightarrow[n \rightarrow \infty]{} \varphi_{\underline{\lambda}' \underline{X}}(1) = \varphi_{\underline{X}}(\underline{\lambda}).$$

■

Corollaire 2.3.1 Si $\underline{X}_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \underline{X}$, alors $X_n(j) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} X(j), j = 1, \dots, k$.

La proposition suivante montre que la convergence en loi est plus faible que tous les autres modes de convergence.

Proposition 2.3.2 Soit (X_n, X) une suite de v.a. Alors si $X_n \xrightarrow{P} X$ alors $X_n \xrightarrow{\mathcal{L}} X$.

Preuve. Soit $F_n(\cdot)$ (resp. $F(\cdot)$) la fonction de répartition de X_n (resp. de X). Soit $\varepsilon > 0, x \in \mathbb{R}$,

$$\begin{aligned} P(X_n \leq x) &\leq P(X \leq x + \varepsilon) + P(X_n - X \leq -\varepsilon) \\ &\leq P(X \leq x + \varepsilon) + P(|X_n - X| > \varepsilon). \end{aligned}$$

Un raisonnement analogue partant de l'inégalité

$$P(X_n \leq x) \geq P(X \leq x - \varepsilon) + P(|X_n - X| > \varepsilon),$$

réécrite sous la forme

$$F(x - \varepsilon) + P(|X_n - X| > \varepsilon) \leq F_n(x) \leq F(x + \varepsilon) + P(|X_n - X| > \varepsilon),$$

puis lorsque $n \rightarrow \infty$

$$F(x - \varepsilon) \leq \liminf_n F_n(x) \leq \limsup_n F_n(x) \leq F(x + \varepsilon),$$

on choisit ε assez petit, alors $F_n(x)$ converge vers $F(x)$, en x point de continuité de F . ■

Remarque 2.3.2 *La réciproque est fautive en général : la convergence en loi n'entraîne pas la convergence en probabilité.*

Nous avons seulement la réciproque lorsque la limite est déterministe.

Proposition 2.3.3 *Soit (X_n) une suite de v.a. Si $X_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} b$, $b \in \mathbb{R}$, alors $X_n \xrightarrow[n \rightarrow \infty]{P} b$.*

Preuve On a : $F_{X_n}(x) \rightarrow \mathbb{I}_{[b, +\infty[}(x)$. Donc $\forall \varepsilon > 0$,

$$\begin{aligned} P(|X_n - b| \leq \varepsilon) &= P(b - \varepsilon \leq X_n \leq b + \varepsilon) \\ &= F_{X_n}(b + \varepsilon) - F_{X_n}(b - \varepsilon) \\ &\rightarrow \mathbb{I}_{[b, +\infty[}(b + \varepsilon) - \mathbb{I}_{[b, +\infty[}(b - \varepsilon) = 1. \end{aligned}$$

Théorème 2.3.3 *Si (\underline{X}_n) et (\underline{Y}_n) sont deux suites de \mathbb{R}^k t.q $\underline{X}_n - \underline{Y}_n = O_p(1)$ et si $\underline{X}_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \underline{X}$, alors $\underline{Y}_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \underline{X}$.*

Théorème 2.3.4 [Slutsky] *Soit (U_n, V_n) une paire de suite de v.a.r. Soit U une v.a.r, $a \in \mathbb{R}$. Supposons que $U_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} U$ et $V_n \xrightarrow[n \rightarrow \infty]{P} a$, alors*

1. $U_n \mp V_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} U \mp a$,
2. $U_n V_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} aU$,

$$3. \frac{U_n}{V_n} \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \frac{U}{a}, V_n \neq 0 \text{ p.s, } a \neq 0.$$

Théorème 2.3.5 (*Méthode Delta*) Soient (X_n) une suite de variables aléatoires et (a_n) une suite de réels tendant vers $+\infty$. Supposons qu'il existe un réel θ et une variable X tels que : $a_n(X_n - \theta) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} X$. Si g est une fonction dérivable au point θ , alors : $a_n(g(X_n) - g(\theta)) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} g'(\theta) X$.

Preuve. D'après le Théorème de Slutsky, on sait que : $X_n \xrightarrow[n \rightarrow \infty]{p} \theta$ (car $X_n - \theta = \frac{a_n(X_n - \theta)}{a_n} \xrightarrow[n \rightarrow \infty]{\mathcal{L}} 0$ et $\frac{1}{a_n} \xrightarrow[n \rightarrow \infty]{p} 0$).

– **1er cas :** Si $g \in \mathcal{C}^1(\mathcal{V}(\theta))$, nous avons donc le développement limité de Taylor d'ordre 1,

$$g(X_n) = g(\theta) + g'(\theta_n^*)(X_n - \theta),$$

où (θ_n^*) est une suite de v.a. t.q. $|\theta_n^* - \theta| \leq |X_n - \theta|$ et $\theta_n^* \xrightarrow[n \rightarrow \infty]{p} \theta$. Comme $g' \in \mathcal{C}(\mathcal{V}(\theta))$ alors $g'(\theta_n^*) \xrightarrow[n \rightarrow \infty]{p} g'(\theta)$ et $a_n(g(X_n) - g(\theta)) = g'(\theta_n^*) a_n(X_n - \theta)$, d'où le résultat d'après le Théorème de Slutsky.

– **2ème cas :** Si g' n'est pas nécessairement continue, donc $g(X_n) = g(\theta) + g'(\theta)(X_n - \theta) + R_n$, où $R_n = o_p(|X_n - \theta|)$. Comme $a_n R_n = a_n(X_n - \theta) \frac{R_n}{X_n - \theta} \xrightarrow[n \rightarrow \infty]{p} 0$ et $a_n(g(X_n) - g(\theta)) = g'(\theta) a_n(X_n - \theta) + a_n R_n$. D'où le résultat.

■

Proposition 2.3.4 Soit $(\underline{X}_n, \underline{X})$ une suite dans \mathbb{R}^s t.q $\underline{X}_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \underline{X}$.

1. Si $f : \mathbb{R}^s \rightarrow \mathbb{R}^k$ est une fonction continue, alors $f(\underline{X}_n) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} f(\underline{X})$.
2. Si (\underline{Y}_n) est une suite dans \mathbb{R}^s t.q $\|\underline{Y}_n\| \xrightarrow[n \rightarrow \infty]{p} 0$, alors $\underline{X}_n + \underline{Y}_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \underline{X}$.
3. Soit M une matrice $r \times s$ et soit (M_n) une suite de matrices aléatoires dans $\mathcal{M}(\mathbb{R}^r \times \mathbb{R}^s)$. Si $\|M_n - M\| \xrightarrow[n \rightarrow \infty]{p} 0$, alors $M_n \underline{X}_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} M \underline{X}$.

On a le théorème suivant, très utile pour les preuves.

Théorème 2.3.6 [*Portmanteau*] Soit $(\underline{X}_n, \underline{X})$ une suite dans \mathbb{R}^k . Les assertions suivantes sont équivalentes :

1. $\underline{X}_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \underline{X}$.
2. $E\{f(\underline{X}_n)\} \rightarrow E\{f(\underline{X})\}$, pour toute fonction bornée et continue.

3. $E \{f(\underline{X}_n)\} \longrightarrow E \{f(\underline{X})\}$, pour toute fonction lipschitzienne.
4. $\liminf E \{f(\underline{X}_n)\} \geq E \{f(\underline{X})\}$.
5. $P(\underline{X}_n \in G) > P(\underline{X} \in G), \forall G$ un ouvert.
6. $P(\underline{X}_n \in F) \leq P(\underline{X} \in F), \forall F$ un fermé.

2.4 Convergence en moyenne d'ordre p

Dans toute la suite on note $\mathbb{L}_p(\Omega, \mathcal{A}, P)$ l'espace de classes d'équivalence des *v.a.r.*, X définies sur (Ω, \mathcal{A}, P) vérifiant $\int_{\Omega} |X|^p dP < +\infty$.

Définition 2.4.1 Soient (\underline{X}_n) une suite dans \mathbb{L}_p et $\underline{X} \in \mathbb{L}_p$. On dit que \underline{X}_n converge en moyenne d'ordre p vers \underline{X} et on écrit $\underline{X}_n \xrightarrow[n \rightarrow \infty]{\mathbb{L}_p} \underline{X}$ si et seulement si

$$E \{\|\underline{X}_n - \underline{X}\|^p\} = \int_{\Omega} \|\underline{X}_n - \underline{X}\|^p dP \xrightarrow[n \rightarrow \infty]{} 0.$$

Remarque 2.4.1 1. En appliquant l'inégalité de Tchebychev on montre que la convergence dans \mathbb{L}_p implique la convergence en probabilité.

2. L'espace \mathbb{L}_p ($p \in [1, +\infty[$) est un espace vectoriel normé par la norme définie par

$$\|X\|_p = (E \{\|X\|^p\})^{\frac{1}{p}} = \left(\int_{\Omega} \|X\|^p dP \right)^{\frac{1}{p}}.$$

3. \mathbb{L}_p est un espace de Banach pour $p \in [1, +\infty[$.

4. Lorsque $p = 1$ on parle de convergence en moyenne et pour $p = 2$ de convergence en moyenne quadratique.

2.4.1 Inégalités

1. [Hölder] Soient p et q deux nombres conjugués : $\frac{1}{p} + \frac{1}{q} = 1$. Si $X \in \mathbb{L}_p$ et $Y \in \mathbb{L}_q$ alors $XY \in \mathbb{L}_1$ et

$$E \{|XY|\} \leq (E \{\|X\|^p\})^{\frac{1}{p}} (E \{\|Y\|^q\})^{\frac{1}{q}} = \|X\|_p \|Y\|_q.$$

2. [Minkowski] Si $p \in]1, +\infty[$ et $X, Y \in \mathbb{L}_p$, alors

$$\|X + Y\|_p \leq \|X\|_p + \|Y\|_p.$$

2.4.2 T.C.D et T.C.M

Signalons également que le théorème de convergence dominée permet de passer de la convergence presque sûre à la convergence en moyenne.

Théorème 2.4.1 [T.C.D dans \mathbb{L}_p] Soit (X_n) une suite de v.a.r. dans \mathbb{L}_p qui converge p.s. vers X . Supposons qu'il existe une v.a.r. positive $Y \in \mathbb{L}_p$ t.q $|X_n| < Y$ p.s. $\forall n$. Alors,

$$X \in \mathbb{L}_p \text{ et } \|X_n - X\|_p \xrightarrow[n \rightarrow \infty]{} 0,$$

$$\left(\text{et } \lim_{n \rightarrow \infty} \|X_n\|_p = \|X\|_p \right).$$

Théorème 2.4.2 [T.C.M] Si (X_n) est une suite de v.a.r. avec $X_n \geq 0, \forall n$, alors $\lim_{n \rightarrow \infty} E \{X_n\} = E \left\{ \lim_{n \rightarrow \infty} X_n \right\}$.

Théorème 2.4.3 [Lemme de Fatou] Si $X_n \geq 0$ p.s. alors $E \left\{ \liminf_{n \rightarrow \infty} X_n \right\} \leq \liminf_{n \rightarrow \infty} E \{X_n\}$.

2.4.3 Équi-Intégrabilité

Définition 2.4.2 Une suite de v.a. (X_n) est dite équi-intégrable d'ordre p si et seulement si

$$\lim_{a \rightarrow \infty} \left\{ \sup_n \int_{\{|X_n| > a\}} |X_n|^p dP \right\} = 0.$$

2.4.4 Propriétés

Proposition 2.4.1 Soit (X_n) une suite de v.a.r. dans \mathbb{L}_q qui converge vers X dans \mathbb{L}_q alors (X_n) converge vers X dans \mathbb{L}_p pour tout $p \leq q$.

Preuve. D'après l'inégalité de Hölder, on a : $E \{|X_n - X|^p\} \leq E \{|X_n - X|^q\}^{\frac{p}{q}}$. Le résultat découle immédiatement de cette inégalité. ■

Proposition 2.4.2 Une condition nécessaire et suffisante pour que (X_n) soit équi-intégrable est que

1. $\forall \varepsilon > 0, \exists \eta(\varepsilon) > 0, \forall A \subset \mathbb{Q} : P(A) < \eta(\varepsilon) \implies \sup_n \int_A |X_n|^p dP < \varepsilon,$
2. $\sup_n \int_{\Omega} |X_n|^p dP < +\infty.$

Preuve. $\forall A \subset \mathbb{Q}$,

$$\begin{aligned} \int_A |X_n|^p dP &= \int_{A \cap \{|X_n|^p \leq a\}} |X_n|^p dP + \int_{A \cap \{|X_n|^p > a\}} |X_n|^p dP \\ &\leq aP(A) + \int_{A \cap \{|X_n|^p > a\}} |X_n|^p dP. \end{aligned}$$

Donc

$$\sup_n \int_A |X_n|^p dP \leq aP(A) + \sup_n \int_{\{|X_n|^p > a\}} |X_n|^p dP.$$

Si (X_n) est équi-intégrable, $\exists a_0 : \sup_n \int_{\{|X_n|^p > a_0\}} |X_n|^p dP < \frac{\varepsilon}{2}$. Alors, $\forall A : P(A) < \frac{\varepsilon}{2a}$, $\sup_n \int_{\{|X_n|^p > a\}} |X_n|^p dP < \varepsilon$. De plus, pour $A = \Omega$, on a : $\sup_n \int_{\Omega} |X_n|^p dP < +\infty$. Pour la condition suffisante : $\forall a > 0$,

$$\int_{\Omega} |X_n|^p dP > a^p P(|X_n| \geq a).$$

Donc

$$\sup_n P(|X_n| \geq a) \leq \frac{1}{a^p} \sup_n \int_{\Omega} |X_n|^p dP.$$

ce qui implique que $\lim_{a \rightarrow \infty} \sup_n P(|X_n| \geq a) = 0$. Soit $\varepsilon > 0$, étant fixé pour a assez grand on a :

$$\sup_n P(|X_n| \geq a) < \eta(\varepsilon) \implies \sup_n \int_{\{|X_n|^p \geq a\}} |X_n|^p dP < \varepsilon.$$

■

Théorème 2.4.4 Une suite (X_n) converge en moyenne d'ordre p vers X si et seulement si

- (X_n) est équi-intégrable d'ordre p .
- (X_n) converge en probabilité vers X .

Preuve. On a :

$$\forall \varepsilon > 0, P(|X_n - X| \geq \varepsilon) \leq \frac{1}{\varepsilon^p} \int_{\Omega} |X_n - X|^p dP.$$

La convergence en moyenne d'ordre p implique la convergence en probabilité. D'autre part, soit $n \in \mathbb{N}$ t.q

$$\int_{\Omega} |X_n - X|^p < \frac{\varepsilon}{2^{p-1}}, \quad (|x + y|^p \leq 2^{p-1} |x|^p + 2^{p-1} |y|^p), \text{ donc}$$

$$|x_n|^p \leq 2^{p-1} |x_n|^p + 2^{p-1} |x_n - x_n|^p \implies \int_A |X_n|^p dP \leq 2^{p-1} \int_A |X_n|^p dP + 2^{p-1} \int_A |X_n - X_n|^p dP.$$

■

Proposition 2.4.3 Soit $\alpha > 0$, X_1, \dots, X_n, \dots une suite de v.a. indépendantes t.q : $P(X_n = 0) = 1 - \frac{1}{n^\alpha}$, $P(X_n = n) = \frac{1}{n^\alpha}$, $n \geq 1$. On a :

- $X_n \xrightarrow[n \rightarrow \infty]{P} 0$, même sans l'hypothèse d'indépendance.
- $X_n \xrightarrow[n \rightarrow \infty]{p.s.} 0$ si et seulement si $\alpha > 1$.
- $X_n \xrightarrow[n \rightarrow \infty]{\mathbb{L}_r} 0$ si et seulement si $\alpha > r$.

Preuve. En effet, pour $\varepsilon > 0$, $P(|X_n| \geq \varepsilon) = P(X_n = n) = \frac{1}{n^\alpha} \rightarrow 0$. La convergence p.s. est une conséquence du lemme de Borel Cantelli

$$\sum_{n=1}^{\infty} P(|X_n| \geq \varepsilon) = \sum_{n=1}^{\infty} \frac{1}{n^\alpha} \begin{cases} < +\infty \text{ si } \alpha > 1 \\ = +\infty \text{ si } \alpha \leq 1 \end{cases} .$$

La convergence dans \mathbb{L}_r : $E\{|X_n|^r\} = 0^r \left(1 - \frac{1}{n^\alpha}\right) + n^r \frac{1}{n^\alpha} = n^{r-\alpha} \rightarrow \begin{cases} +\infty, & r > \alpha \\ 1, & r = \alpha \\ 0, & r < \alpha \end{cases} . \blacksquare$

2.5 Exercices

Exercice 2.5.1 Soit (X_n) une suite de variables aléatoires réelles t.q. pour tout $n \geq 1$, $f_{n, X_n}(x) = \frac{n^2}{2} e^{-n^2|x|}$, $\forall x \in \mathbb{R}$.

1. Calculer, pour tout $n \geq 1$, $P(|X_n| > n^{-3/2})$.
2. Déterminer $P(|X_n| > n^{-3/2}, i.o.)$.
3. En déduire que, presque sûrement, la série $\sum_{n \geq 1} X_n$ converge absolument.

Exercice 2.5.2 Soit (X_n) une suite de variables aléatoires indépendantes telle que pour tout $n \geq 1$,

$$P(X_n = -1) = 1 - \frac{1}{n^2}, \quad P(X_n = n^2 - 1) = \frac{1}{n^2}.$$

1. Montrer que la suite (X_n) converge en probabilité vers -1 .
2. Montrer que la suite (X_n) converge presque sûrement vers -1 .

Exercice 2.5.3 Soit (X_n) une suite de variables aléatoires positives ; pour tout $n \geq 1$, X_n suit la loi exponentielle de paramètre $\mu_n > 0$.

1. Donner une condition nécessaire et suffisante pour que (X_n) converge vers 0 :

(a) en probabilité ;

(b) dans \mathbb{L}_1 .

2. On suppose que les (X_n) soient indépendantes. Donner un exemple dans lequel (X_n) converge vers 0 dans \mathbb{L}_1 mais pas presque sûrement.

Exercice 2.5.4 Soient (X_n) et (Y_n) deux suites de v.a.r sur (Ω, \mathcal{A}, P) convergeant en loi vers X et Y respectivement.

1. On suppose que pour tout n , X_n et Y_n sont indépendantes et que X et Y sont aussi indépendantes. Montrer que $X_n + Y_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} X + Y$. Donner un contre-exemple montrant que l'hypothèse d'indépendance est indispensable.

2. On suppose que $Y = 0$. Montrer que $X_n + Y_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} X$ et $X_n Y_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} 0$.

Exercice 2.5.5 Soit (X_n) une suite de v.a. gaussiennes, centrées, de variance (σ_n^2) convergeant en loi vers une v.a. X .

1. Montrer que la suite (σ_n^2) est convergente et en déduire que X suit une loi gaussienne.

2. On suppose que $X_n \xrightarrow[n \rightarrow \infty]{P} X$. Démontrer que (X_n) converge vers X dans \mathbb{L}_p .

Exercice 2.5.6 Soit (X_n) une suite de v.a., de loi uniforme sur $[0; 1]$. Soit N_n une v.a. de loi binomiale $\mathcal{B}(n, p)$ et indépendante des X_n .

Montrer que $n \min_{1 \leq i \leq N_n} (X_i)$ converge en loi, lorsque $n \uparrow \infty$, vers une v.a. exponentielle de moyenne p .

Exercice 2.5.7 Soit (α_n) une suite de nombres appartenant à $[0; 1]$; on lui associe une suite (X_n) de variables aléatoires indépendantes sur un espace probabilisé (Ω, \mathcal{A}, P) t.q. $P(X_n \leq t) = (\alpha_n + (1 - \alpha_n)t^n) \mathbb{I}_{[0;1]}(t) + \mathbb{I}_{]1;+\infty[}(t)$.

À quelles conditions sur (α_n) , la suite (X_n) converge-t-elle en loi, en probabilité, presque sûrement ?

Exercice 2.5.8 Soit (X_n) une suite de variables aléatoires positives de carré intégrable. On suppose que $\lim E\{X_n\} = +\infty$ et que la suite $\left(\frac{\text{Var}(X_n)}{E\{X_n\}}\right)$ est bornée.

1. Montrer que la suite $\left(\frac{X_n}{E\{X_n\}}\right)$ converge vers 1 dans \mathbb{L}_2 .

2. On suppose que, pour tout $n \geq 1$, X_n suit la loi de Poisson de paramètre $\mu_n > 0$ et que $\sum_{n \geq 1} \mu_n^{-1} < +\infty$.

(a) La suite $\left(\frac{X_n}{\mu_n}\right)$ converge-t-elle vers 1 dans \mathbb{L}_2 ?

(b) Montrer que $\left(\frac{X_n}{\mu_n}\right)$ converge vers 1 presque sûrement.

Solutions

Solution 2.5.1 1. On a, pour tout $n \geq 1$,

$$P(|X_n| > n^{-3/2}) = \int_{\{|x| > n^{-3/2}\}} f_{n, X_n}(x) dx = \int_{\{x > n^{-3/2}\}} \frac{n^2}{2} e^{-n^2 x} dx + \int_{\{x < -n^{-3/2}\}} \frac{n^2}{2} e^{n^2 x} dx = e^{-\sqrt{n}}.$$

2. Puisque $\sum_{n \geq 1} e^{-\sqrt{n}} < +\infty$, le lemme de Borel-Cantelli donne $P(|X_n| > n^{-3/2}, i.o.) = 0$.

3. Notons $A = \{|X_n| > n^{-3/2}, i.o.\}$, donc $A^c = \liminf \{|X_n| \leq n^{-3/2}\}$. Soit $\omega \in A^c$, alors il existe un entier n_ω tel que, pour tout $n \geq n_\omega$, $\omega \in \{|X_n| \leq n^{-3/2}\}$. Soit $|X_n(\omega)| \leq n^{-3/2}$ pour tout $n \geq n_\omega$. Il résulte du critère de Riemann ($3/2 > 1$), que la série $\sum_{n \geq 1} X_n(\omega)$ est absolument convergente pour tout $\omega \in A^c$.

Solution 2.5.2 1. Pour tout $\varepsilon > 0$, pour tout $n \geq 1$, on a :

$$P(|X_n + 1| > \varepsilon) = P(X_n = n^2 - 1) = \frac{1}{n^2} \longrightarrow 0,$$

donc la suite (X_n) converge en probabilité vers -1 .

2. On a, pour tout $n \geq 1$, $P(X_n \neq -1) = P(X_n = n^2 - 1) = \frac{1}{n^2}$, donc $\sum_{n \geq 1} P(X_n \neq -1) < +\infty$. D'après le lemme de Borel-Cantelli $P(\liminf \{X_n = -1\}) = P(\limsup \{X_n = -1\}) = 1$. En particulier, avec probabilité 1, la suite (X_n) converge vers -1 . Alors, la suite (X_n) converge presque sûrement vers -1 .

Solution 2.5.3 1. (a) Soit $\varepsilon > 0$, on a :

$$P(|X_n| > \varepsilon) = P(X_n > \varepsilon) = \int_{\{x > \varepsilon\}} f_{X_n}(x) dx = e^{-\mu_n \varepsilon},$$

qui tend vers 0 si et seulement si $\mu_n \longrightarrow 0$.

(b) Pour tout entier $n \geq 1$, $E\{|X_n|\} = E\{X_n\} = \mu_n^{-1}$ qui tend vers 0 si et seulement si $\mu_n \longrightarrow 0$.

2. La suite (X_n) converge vers 0 presque sûrement si et seulement si, pour tout réel $\varepsilon > 0$, $P(X_n > \varepsilon, i.o.) = 0$. Comme les (X_n) sont indépendantes, cette condition est équivalente à $\sum_{n \geq 1} P(X_n > \varepsilon) < +\infty$ pour tout $\varepsilon > 0$ c-à-d., $\forall \varepsilon > 0$, $\sum_{n \geq 1} e^{-\mu_n \varepsilon} < +\infty$. Il suffit de prendre $\mu_n = \ln n$ pour obtenir l'exemple demandé.

Solution 2.5.4 1. On utilise les fonctions caractéristiques :

$$\begin{aligned} \forall t, \varphi_{X_n+Y_n}(t) &= \varphi_{X_n}(t) \varphi_{Y_n}(t) \text{ car } X_n \text{ et } Y_n \text{ indépendants} \\ &\longrightarrow \varphi_X(t) \varphi_Y(t) \\ &= \varphi_{X+Y}(t) \text{ car } X \text{ et } Y \text{ indépendants.} \end{aligned}$$

Donc $X_n + Y_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} X + Y$.

Pour se convaincre de l'importance de l'hypothèse d'indépendance, il suffit de considérer une variable aléatoire X suivant une loi normale $\mathcal{N}(0, 1)$ et poser : $X_n = X$ et $Y_n = -X$. On a ainsi : $X_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} X$, $Y_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} X$ et $X_n + Y_n = 0$.

2. Pour tout $x \in \mathbb{R}$ et tout $\varepsilon > 0$, $\{X_n \leq x - \varepsilon\} \cap \{|Y_n| \leq \varepsilon\} \subset \{X_n + Y_n \leq x\}$. En considérant les événements contraires, puis les probabilités respectives, on obtient :

$$F_{X_n}(x - \varepsilon) \leq F_{X_n+Y_n}(x) + P(|Y_n| > \varepsilon).$$

De même : $\{X_n > x + \varepsilon\} \cap \{|Y_n| > \varepsilon\} \subset \{X_n + Y_n > x\}$, puis

$$F_{X_n+Y_n}(x) \leq F_{X_n}(x + \varepsilon) + P(|Y_n| > \varepsilon).$$

De ces deux inégalités, on obtient

$$F_{X_n}(x - \varepsilon) - P(|Y_n| > \varepsilon) \leq F_{X_n+Y_n}(x) \leq F_{X_n}(x + \varepsilon) + P(|Y_n| > \varepsilon).$$

La fonction F_{X_n} étant croissante, on déduit l'encadrement :

$$|F_{X_n+Y_n}(x) - F_{X_n}(x)| \leq F_{X_n}(x + \varepsilon) - F_{X_n}(x - \varepsilon) + P(|Y_n| > \varepsilon).$$

On considère alors x point de continuité de F_X . On peut choisir ε aussi petit que l'on veut avec de plus $x - \varepsilon$ et $x + \varepsilon$ points de continuité de F_X et $F_X(x + \varepsilon) - F_X(x - \varepsilon)$ arbitrairement petit. Pour de tels x et ε , on a :

$$\overline{\lim} |F_{X_n+Y_n}(x) - F_{X_n}(x)| \leq F_X(x + \varepsilon) - F_X(x - \varepsilon).$$

On déduit que : $F_{X_n+Y_n}(x) \longrightarrow F_{X_n}(x)$ et $X_n + Y_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} X$.

On va montrer que le produit $X_n Y_n$ converge en probabilité vers 0. Pour tout entier k , $\{|X_n| \leq k\} \cap \{|Y_n| \leq \frac{1}{k^2}\} \subset \{|X_n Y_n| \leq \frac{1}{k}\}$ et donc

$$P\left(|X_n Y_n| > \frac{1}{k}\right) \leq P(|X_n| > k) + P\left(|Y_n| > \frac{1}{k^2}\right).$$

Soit $\varepsilon > 0$. La suite (X_n) étant convergente en loi, elle est tendue. Donc quel que soit n , $P(|X_n| > k) < \varepsilon$ si k est suffisamment grand. D'autre part la suite (Y_n) convergente en loi vers une constante, converge en probabilité vers cette constante, donc $P(|Y_n| > \frac{1}{k^2}) < \varepsilon$ si n est suffisamment grand. Finalement $\forall k, P(|X_n Y_n| > \frac{1}{k}) \xrightarrow[n]{n} 0$. Alors la variable $(X_n Y_n)$ converge en probabilité, et donc en loi, vers 0.

Solution 2.5.5 1. On a : $X_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} X$. Donc, la suite des fonctions caractéristiques $(\varphi_{X_n}(t))$ converge simplement sur \mathbb{R} vers $\varphi_X(t)$, alors $\forall t, e^{-\frac{1}{2}\sigma_n^2 t^2} \longrightarrow \varphi_X(t)$. On déduit que la suite (σ_n) est convergente vers un réel σ positif. Dans le cas où $\sigma > 0$, $\varphi_X(t) = e^{-\frac{1}{2}\sigma^2 t^2}$ et la variable X suit donc la loi gaussienne $\mathcal{N}(0, \sigma^2)$. En revanche, le cas $\sigma = 0$ donne une convergence en loi vers la variable constante égale à 0 qui n'est pas gaussienne.

2. Par le résultat du 1), X est gaussienne centrée et de variance σ^2 . Donc, il suffit de montrer que la suite $(E\{|X_n|^p\})$ est majorée. On pose $X_n = \sigma_n Y_n$ où Y_n suit une loi normale centrée réduite. De plus

$$E\{|X_n|^p\} = \sigma_n^p E\{|Y_n|^p\} = \sigma_n^p E\{|Y_0|^p\} \leq K,$$

où K est une constante indépendante de n dont l'existence est assurée par la convergence de la suite (σ_n) . Donc, la suite (X_n) converge vers X dans \mathbb{L}_p .

Solution 2.5.6 Soit $t \in [0; 1]$. On a :

$$\begin{aligned} \left\{n \min_{1 \leq i \leq N_n} (X_i) > t\right\} &= \bigcup_{k=0}^n \left(\left\{n \min_{1 \leq i \leq N_n} (X_i) > t\right\} \cap \{N_n = k\} \right) \\ &= \bigcup_{k=0}^n \left(\left\{ \min_{1 \leq i \leq k} (X_i) > \frac{t}{n} \right\} \cap \{N_n = k\} \right) \\ &= \bigcup_{k=0}^n \left(\left\{ X_1 > \frac{t}{n}, \dots, X_k > \frac{t}{n} \right\} \cap \{N_n = k\} \right). \end{aligned}$$

Les X_i et N_n étant indépendantes, il s'en suit :

$$\begin{aligned} P\left(n \min_{1 \leq i \leq N_n} (X_i) > t\right) &= \sum_{k=0}^n \left(1 - \frac{t}{n}\right)^k C_n^k p^k (1-p)^{n-k} = \left(\left(1 - \frac{t}{n}\right)p + (1-p)\right)^n \\ &= \left(1 - \frac{tp}{n}\right)^n \xrightarrow[n]{n} e^{-tp}. \end{aligned}$$

Pour $t \notin [0; 1]$, le calcul est trivial, et finalement : $\forall t, P\left(n \min_{1 \leq i \leq N_n} (X_i) \leq t\right) \xrightarrow[n]{n} P(Y \leq t)$ où $Y \sim$ exponentielle de moyenne p

Solution 2.5.7 Pour que la suite (X_n) converge en loi il faut qu'il existe un $t \in]0; 1[$ pour lequel la suite $(P(X_n \leq t))$, soit convergente.

1er cas Si la suite (α_n) ne tend pas vers 0, alors quel que soit $t \in]0; 1[$, $P(X_n \leq t) = \alpha_n + (1 - \alpha_n)t^n \sim \alpha_n$. Dans ce cas, il est nécessaire que (α_n) soit convergente. Si $\alpha_n \rightarrow \alpha$, la suite (X_n) converge en loi vers X où $X \sim \mathcal{B}(\alpha)$.

2eme cas Si la suite (α_n) tend vers 0, alors la suite (X_n) converge en loi vers $X = 1$.

En conclusion, pour que (X_n) converge en loi, il faut et il suffit que α_n soit convergente vers un réel α , et (X_n) converge alors en loi vers $X \sim \mathcal{B}(\alpha)$.

Pour pouvoir affirmer que la convergence soit une convergence en probabilité, il faut et il suffit que la limite X soit constante presque sûrement, c'est-à-dire $\alpha = 0$ ou $\alpha = 1$.

De même pour pouvoir affirmer que $X_n \xrightarrow[n]{p.s.} 0$ (resp. 1), il faut et il suffit que $\sum_n P(X_n > \varepsilon) < \infty$ (resp. $\sum_n P(1 - X_n > \varepsilon) < \infty$) pour tout ε (voir Lemme de Borel-Cantelli), c'est-à-dire si $\sum_n (1 - \alpha_n) < \infty$ (resp. $\sum_n \alpha_n < \infty$).

Solution 2.5.8 1. Puisque $\lim E\{X_n\} = +\infty$, $E\{X_n\} > 0$ pour n assez grand. On a :

$$E\left\{\left(\frac{X_n}{E\{X_n\}} - 1\right)^2\right\} = \frac{Var(X_n)}{(E\{X_n\})^2} \rightarrow 0,$$

parce que la suite $\left(\frac{Var(X_n)}{E\{X_n\}}\right)$ est bornée et $\lim E\{X_n\} = +\infty$. D'où le résultat.

(a) On a $Var(X_n) = E\{X_n\} = \mu_n$, donc la suite $\left(\frac{Var(X_n)}{E\{X_n\}}\right)$ est constante. De plus, $\lim \mu_n = +\infty$ car $\sum_{n \geq 1} \mu_n^{-1} < +\infty$. La question précédente donne le résultat.

(b) D'après l'inégalité de Bienaymé-Tchebychev, pour tout $\varepsilon > 0$, on a :

$$P\left(\left|\frac{X_n}{\mu_n} - 1\right| > \varepsilon\right) \leq \frac{\text{Var}\left(\frac{X_n}{\mu_n}\right)}{\varepsilon^2} = \frac{\text{Var}(X_n)}{\varepsilon^2 \mu_n^2} = \frac{1}{\varepsilon^2 \mu_n}.$$

Comme $\sum_{n \geq 1} \mu_n^{-1} < +\infty$, alors le lemme de Borel-Cantelli donne $P\left(\left\{\left|\frac{X_n}{\mu_n} - 1\right| > \varepsilon\right\}, i.o.\right) = 0$.

Donc la suite $\left(\frac{X_n}{\mu_n}\right)$ converge presque sûrement vers 1.

Chapitre 3

Théorèmes limites

Dans ce chapitre, nous allons donner deux exemples de théorèmes limites pour des suites de v.a. réelles indépendantes : la loi des grands nombres (en abrégé, *LGN*) et le théorème de la limite centrale ou théorème central limite (en abrégé, *TCL*).

3.1 Lois des grands nombres

La loi des grands nombres consiste à étudier la limite de la suite $\left(\frac{1}{n} \sum_{i=1}^n X_i\right)_{n \in \mathbb{N}}$. On parle de loi faible lorsqu'on étudie la convergence en probabilité et de loi forte lorsqu'on s'intéresse à la convergence presque sûre.

Nous mènerons cette étude dans le cas d'une suite de variables aléatoires réelles $(X_n)_{n \in \mathbb{N}}$ indépendantes et identiquement distribuées. Le premier théorème important et très simple à montrer est la loi faible des grands nombres.

Théorème 3.1.1 (Loi faible des grands nombres) Soit $(X_n)_n$ une suite de variables aléatoires indépendantes et identiquement distribuées admettant un moment d'ordre deux. On note $\mu = E\{X_1\}$ et $\sigma^2 = \text{Var}(X_1)$. Alors $\left(\frac{1}{n} S_n := \frac{1}{n} \sum_{i=1}^n X_i\right)_n$ converge en probabilité vers μ .

Preuve. On remarque que, pour tout $n \geq 1$, par linéarité $E\left\{\frac{1}{n} S_n\right\} = \mu$ et par indépendance $\text{Var}\left(\frac{1}{n} S_n\right) = \frac{1}{n} \sigma^2$. On a alors, par inégalité de Tchebychev, pour $\varepsilon > 0$ et $n \geq 1$,

$$P\left(\left|\frac{1}{n} S_n - \mu\right| > \varepsilon\right) \leq \frac{\text{Var}\left(\frac{1}{n} S_n\right)}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2} \longrightarrow 0,$$

d'où le résultat. ■

Théorème 3.1.2 (Loi faible des grands nombres) Soit $(X_n)_n$ une suite de variables aléatoires indépendantes et identiquement distribuées admettant un moment d'ordre un. On note $\mu = E\{X_1\}$. Alors $(\frac{1}{n}S_n)_n$ converge en probabilité vers μ .

Preuve. Quitte à centrer les variables X_n , on peut supposer que $E\{X_n\} = 0$. Puisque $X \in \mathbb{L}_1$, la fonction caractéristique φ_X est dérivable et de plus $\varphi'_X(0) = iE\{X_1\} = 0$. La formule de Taylor donne $\varphi_X(t) = 1 + o(t)$. Donc $\varphi_{\frac{1}{n}S_n}(t) = \varphi_X^n(\frac{t}{n}) = (1 + o(n^{-1}))^n = 1 + o(1)$. Donc $\frac{1}{n}S_n$ converge en loi vers la constante 0, donc en probabilité vers 0. ■

En pratique, comment montrer une convergence presque sûre ?

Parfois, il suffit d'appliquer le résultat général suivant, qui garantit une convergence presque sûre vers une variable aléatoire constante.

Théorème 3.1.3 (Loi forte des grands nombres) Soit $(X_n)_n$ une suite de variables aléatoires indépendantes et identiquement distribuées admettant un moment d'ordre deux. On note $\mu = E\{X_1\}$. Alors $(\frac{1}{n}S_n)_n$ converge presque sûrement vers μ .

Preuve. La démonstration consiste à prouver dans un premier temps le résultat sous l'hypothèse plus forte que $E\{X_1^4\} < +\infty$ et $E\{X_1\} = 0$ (on peut remplacer X_i par $X_i - E\{X_i\}$ et supposer les variables aléatoires centrées). Dans ce cas, l'inégalité de Markov montre que pour tout $n \geq 1$ et tout $\varepsilon > 0$, $P(|S_n| > n\varepsilon) \leq \frac{E\{S_n^4\}}{n^4\varepsilon^4}$. Donc, par linéarité de l'espérance, indépendance et centrage des X_i , $E\{S_n^4\} = E\{X_1^4\} + 3n(n-1)(E\{X_1^2\})^2$. Donc $\sum_n P(|S_n| > n\varepsilon) < +\infty$, ce qui démontre la loi forte des grands nombres dans ce cas d'après le lemme de Borel-Cantelli. ■

L'inégalité de Kolmogorov, due à Andreï Kolmogorov, est une étape essentielle de sa démonstration de la loi forte des grands nombres.

Théorème 3.1.4 (L'inégalité de Kolmogorov) Soit (X_n) une suite finie de variables aléatoires indépendantes, de moyenne nulle et $Var(X_n) < +\infty$. Alors

$$P\left(\max_{k=1,\dots,n} |S_k| \geq \varepsilon\right) \leq \frac{\sum_{k=1}^n Var(X_k)}{\varepsilon^2}.$$

Preuve. Soit $F = \{\max_{k=1,\dots,n} |S_k| \geq \varepsilon\} = \bigcup_{k=1}^n F_k$ où $F_k := \{|S_k| \geq \varepsilon, |S_j| < \varepsilon, j = 1, \dots, k-1\}$, Les F_k sont disjoints par construction. Donc

$$\sum_{k=1}^n \int_{F_k} S_n^2 dP = \int_F S_n^2 dP \leq \int_{\Omega} S_n^2 dP = \text{Var}(S_n) = \sum_{k=1}^n \text{Var}(X_k).$$

Calculons $\int_{F_k} S_n^2 dP$:

$$\begin{aligned} \int_{F_k} S_n^2 dP &= \int_{F_k} \left(S_k + \sum_{j=k+1}^n X_j \right)^2 dP = \int_{F_k} S_k^2 dP + \int_{F_k} \mathbb{I}_{F_k} \left(\sum_{j=k+1}^n X_j \right)^2 dP \\ &= \int_{F_k} S_k^2 dP + P(F_k) \int \left(\sum_{j=k+1}^n X_j \right)^2 dP, \end{aligned}$$

ce qui implique

$$\int_{F_k} S_n^2 dP \geq \int_{F_k} S_k^2 dP \geq \varepsilon^2 P(F_k).$$

Par conséquent $\varepsilon^2 P(F) = \varepsilon^2 \sum_{k=1}^n P(F_k) \leq \sum_{k=1}^n \text{Var}(X_k)$. D'où le résultat. ■

Application a une condition suffisante de convergence presque sûre.

Théorème 3.1.5 Soit (X_n) une suite de variables aléatoires indépendantes de moyenne nulle et $\sum_{n=1}^{+\infty} \text{Var}(X_n) < +\infty$. Alors la série $\sum_{n=1}^{+\infty} X_n$ converge presque sûrement.

Preuve. Posons : $A_m(\omega) = \sup_{k \geq 0} |S_{m+k}(\omega) - S_m(\omega)|$ et $A(\omega) = \inf_m A_m(\omega)$. D'après le critère de Cauchy pour que la suite des nombres $S_n(\omega)$ converge presque sûrement il faut et il suffit que $A(\omega) = 0$.

Or $\{A(\omega) \neq 0\} = \bigcup_n \{A(\omega) > \frac{1}{n}\}$,

$$\begin{aligned} \left\{ A(\omega) > \frac{1}{n} \right\} &\iff \left\{ \inf_m A_m(\omega) > \frac{1}{n} \right\} \implies \left\{ A_m(\omega) > \frac{1}{n} \right\} \forall m, \text{ i.e.,} \\ \left\{ A(\omega) > \frac{1}{n} \right\} &\subset \bigcap_m \left\{ A_m(\omega) > \frac{1}{n} \right\}, \end{aligned}$$

ceci entraîne

$$\forall m, P\left(A(\omega) > \frac{1}{n}\right) \leq P\left(A_m(\omega) > \frac{1}{n}\right).$$

D'autre part $\{A_m(\omega) > \frac{1}{n}\} = \bigcup_r \{\sup_{0 \leq k \leq r} |S_{m+k}(\omega) - S_m(\omega)|\}$, donc

$$\begin{aligned} P\left(A_m(\omega) > \frac{1}{n}\right) &= \lim_{r \rightarrow \infty} P\left(\sup_{0 \leq k \leq r} |S_{m+k}(\omega) - S_m(\omega)|\right) \\ &\leq \lim_{r \rightarrow \infty} n^2 \sum_{j=m+1}^{m+r} \text{Var}(X_j) = n^2 \sum_{j=m+1}^{+\infty} \text{Var}(X_j), \forall m \text{ (D'après l'inégalité de Kolmogorov)} \end{aligned}$$

Or, par hypothèse $\sum_{n=1}^{+\infty} \text{Var}(X_n) < +\infty$ est convergente; donc $\sum_{j=m+1}^{+\infty} \text{Var}(X_j)$, reste d'une série convergente, donc $P(A_m(\omega) > \frac{1}{n}) = 0 \forall m$, ce qui entraîne

$$P(A(\omega) \neq 0) = P\left(\bigcup_n \left\{A(\omega) > \frac{1}{n}\right\}\right) \leq \sum_n P\left(A(\omega) > \frac{1}{n}\right) = 0,$$

ce qui implique $P(A(\omega) = 0) = 1 \iff A(\omega) = 0$ presque sûrement $\implies \sum_{n=1}^{+\infty} X_n$ converge presque sûrement.

■

Théorème 3.1.6 Soit (X_n) une suite de variables aléatoires indépendantes de variance $\text{Var}(X_n)$ finie;

1. $E\{X_n\} = \mu_n \longrightarrow \mu.$

2. $\sum_{n=1}^{+\infty} \frac{\text{Var}(X_n)}{n^2} < +\infty.$

Alors $(\frac{1}{n}S_n)_n$ converge presque sûrement vers $\mu.$

Preuve. On a : $\frac{1}{n}S_n - \frac{1}{n} \sum_{l=1}^n \mu_k = \frac{1}{n} \sum_{l=1}^n (X_k - \mu_k)$, comme $E\{X_k - \mu_k\} = 0$ et $\text{Var}(X_k - \mu_k) = \text{Var}(X_k)$,

donc $\sum_{k=1}^{+\infty} \text{Var}\left(\frac{X_k - \mu_k}{k}\right) = \sum_{k=1}^{+\infty} \frac{\text{Var}(X_k)}{k^2} < +\infty$ par hypothèse. Donc, d'après le théorème 3.1.5 : $\sum_{k=1}^{+\infty} \frac{X_k - \mu_k}{k}$

converge presque sûrement et maintenant d'après le lemme de **Kronecker** : $\lim_n \frac{1}{n} \sum_{k=1}^n (X_k - \mu_k) = 0$

presque sûrement, c'est-à-dire $\lim_n \left(\frac{1}{n}S_n - \frac{1}{n} \sum_{k=1}^n \mu_k\right) = 0$ presque sûrement. On conclue grâce à l'hypothèse 1) et le lemme de **Cesàro**, on obtient $(\frac{1}{n}S_n)_n$ converge presque sûrement vers $\mu.$ ■

3.2 Théorèmes centraux limites

Rappelons ici qu'un vecteur \underline{X} dans \mathbb{R}^k est de distribution $\mathcal{N}(\underline{\mu}, \Sigma)$, si sa fonction caractéristique est

$$\varphi_{\underline{X}}(\underline{t}) = \exp\left\{-it' \underline{\mu} - \frac{1}{2} \underline{t}' \Sigma \underline{t}\right\}.$$

Proposition 3.2.1 Soit (a_n) une suite réelle convergente vers 0 et soit $\varphi_{\underline{X}}(\cdot)$ la fonction caractéristique d'un vecteur $\underline{X} \in \mathbb{R}^k$. Alors, $\lim_{n \rightarrow \infty} \frac{\log \varphi_{\underline{X}}(a_n \underline{t})}{\varphi_{\underline{X}}(a_n \underline{t}) - 1} = 1$.

Proposition 3.2.2 1. Soit \underline{X} un vecteur aléatoire de fonction caractéristique $\varphi_{\underline{X}}(\cdot)$. Si $E\{|\underline{X}|\} < +\infty$, alors pour toute suite (a_n) positive convergente vers 0 et pour tout \underline{t} fixé,

$$\lim_{n \rightarrow \infty} \frac{\varphi_{\underline{X}}(a_n \underline{t}) - 1}{a_n} = i \underline{t}' \underline{\mu}.$$

2. Supposons que $E\{\underline{X}\} = 0$ et $E\{\underline{X}^2\} < +\infty$, alors

$$\lim_{n \rightarrow \infty} \frac{\varphi_{\underline{X}}(a_n \underline{t}) - 1}{a_n^2} = -\frac{1}{2} \underline{t}' \Sigma \underline{t} = -\frac{1}{2} \underline{t}' \text{Cov}(\underline{X}, \underline{X}) \underline{t}.$$

Preuve.

1. On a : $|\exp\{ia_n \underline{t}' \underline{X}\} - 1| \leq a_n |\underline{t}' \underline{X}| \implies \frac{|\exp\{ia_n \underline{t}' \underline{X}\} - 1|}{a_n} \leq |\underline{t}' \underline{X}|$. D'autre part,

$$\lim_{n \rightarrow \infty} \frac{\exp\{ia_n \underline{t}' \underline{X}\} - 1}{a_n} \longrightarrow i \underline{t}' \underline{X}.$$

D'après le T.C.D. on a :

$$\lim_{n \rightarrow \infty} \frac{\varphi_{\underline{X}}(a_n \underline{t}) - 1}{a_n} = \lim_{n \rightarrow \infty} \frac{E\{\exp\{ia_n \underline{t}' \underline{X}\}\} - 1}{a_n} = E\left\{\lim_{n \rightarrow \infty} \frac{\exp\{ia_n \underline{t}' \underline{X}\} - 1}{a_n}\right\} = i \underline{t}' \underline{\mu}.$$

2. On a : $\lim_{n \rightarrow \infty} \frac{\varphi_{\underline{X}}(a_n \underline{t}) - 1}{a_n^2} = E\left\{\lim_{n \rightarrow \infty} \frac{\exp\{ia_n \underline{t}' \underline{X}\} - 1 - i \underline{t}' \underline{X}}{a_n^2}\right\} = -\frac{1}{2} E\{(\underline{t}' \underline{X})^2\} = -\frac{1}{2} \underline{t}' \Sigma \underline{t}$.

■

Proposition 3.2.3 Soit (\underline{X}_n) une suite de vecteurs aléatoires à valeurs dans \mathbb{R}^k i.i.d. On suppose que $E\{\underline{X}_1\} = \underline{0}$ et $E\{|\underline{X}_1|^2\} < +\infty$. Soit $\underline{S}_n = \sum_{i=1}^n \underline{X}_i$. Alors, $\frac{\underline{S}_n}{\sqrt{n}} \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(\underline{0}, \Sigma)$, où $\Sigma = E\{\underline{X}_1 \underline{X}_1'\} = \text{Cov}(\underline{X}_1, \underline{X}_1)$.

Preuve. La fonction caractéristique de $\frac{\underline{S}_n}{\sqrt{n}}$ est $\left(\varphi_{\underline{X}_1}\left(\frac{1}{\sqrt{n}} \underline{t}\right)\right)^n$. Pour n assez grand $|\varphi_{\underline{X}}\left(\frac{1}{\sqrt{n}} \underline{t}\right) - 1| \leq \rho <$

1. D'où, $\frac{n \log \varphi_{\underline{X}}\left(\frac{1}{\sqrt{n}} \underline{t}\right)}{n \left\{\varphi_{\underline{X}}\left(\frac{1}{\sqrt{n}} \underline{t}\right) - 1\right\}} \xrightarrow[n \rightarrow \infty]{} 1$, et $n \left\{\varphi_{\underline{X}}\left(\frac{1}{\sqrt{n}} \underline{t}\right) - 1\right\} \xrightarrow[n \rightarrow \infty]{} -\frac{1}{2} \underline{t}' \Sigma \underline{t}$. Donc

$$\lim_{n \rightarrow \infty} E\left\{\exp\left\{i \underline{t}' \frac{\underline{S}_n}{\sqrt{n}}\right\}\right\} = \lim_{n \rightarrow \infty} \exp\left\{n \log \varphi_{\underline{X}}\left(\frac{1}{\sqrt{n}} \underline{t}\right)\right\} = \exp\left\{-\frac{1}{2} \underline{t}' \text{Cov}(\underline{X}, \underline{X}) \underline{t}\right\}.$$

■

Proposition 3.2.4 Soit (\underline{X}_n) une suite de vecteurs aléatoires à valeurs dans \mathbb{R}^k i.i.d. On suppose que $E\{\underline{X}_1\} = \underline{\mu}$ et $E\{|\underline{X}_1|^2\} < +\infty$. Alors,

$$\frac{1}{\sqrt{n}} (\underline{S}_n - n\underline{\mu}) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(\underline{0}, \Sigma),$$

$$\sqrt{n} (\overline{X}_n - \underline{\mu}) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(\underline{0}, \Sigma) \quad (\text{où } \overline{X}_n = \frac{1}{n} \underline{S}_n).$$

Remarque 3.2.1 Si $(X_j)_{1 \leq j \leq n}$ est une suite i.i.d de moyenne μ et de variance σ^2 , alors

$$\sqrt{n} \frac{(\overline{X}_n - \mu)}{\sigma} \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, 1) \implies n \frac{(\overline{X}_n - \mu)^2}{\sigma^2} \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \chi^2(1).$$

Théorème 3.2.1 (Cramèr) Soit (λ_n) une suite réelle convergente vers $+\infty$ et $(\underline{X}_n, \underline{a})$ une suite de vecteurs aléatoires à valeurs dans \mathbb{R}^k t.q : $\lambda_n (\underline{X}_n - \underline{a}) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \underline{X}, \underline{a} \in \mathbb{R}^k$. Soit $f : \mathbb{R}^k \rightarrow \mathbb{R}^m$, une fonction de classe $C^1(v_{\{\underline{a}\}})$, alors

$$\lambda_n (f(\underline{X}_n) - f(\underline{a})) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \overset{\circ}{f}(\underline{a}) \underline{X},$$

où $\overset{\circ}{f}(\underline{x}) = \frac{\partial}{\partial \underline{x}} f(\underline{x})$.

Preuve. Remarquons d'abord que : $\underline{X}_n - \underline{a} = \frac{1}{\lambda_n} \lambda_n (\underline{X}_n - \underline{a}) \xrightarrow[n \rightarrow \infty]{P} 0$. D'autre part, $\exists t_n \in]0, 1[: f(\underline{X}_n) - f(\underline{a}) = f(\underline{a} + t_n (\underline{X}_n - \underline{a})) (\underline{X}_n - \underline{a})$. Comme $\overset{\circ}{f}$ est continue au point \underline{a} on a

$$f(\underline{a} + t_n (\underline{X}_n - \underline{a})) \xrightarrow[n \rightarrow \infty]{P} \overset{\circ}{f}(\underline{a}) \implies \lambda_n (f(\underline{X}_n) - f(\underline{a})) = f(\underline{a} + t_n (\underline{X}_n - \underline{a})) \lambda_n (\underline{X}_n - \underline{a}) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \overset{\circ}{f}(\underline{a}) \underline{X}.$$

■

Proposition 3.2.5 Soit (\underline{X}_n) une suite de vecteurs aléatoires dans \mathbb{R}^k t.q : $\sqrt{n} (\underline{X}_n - \underline{\mu}) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, \Sigma)$. Soit $f : \mathbb{R}^k \rightarrow \mathbb{R}^m$, une fonction continûment différentiable sur un voisinage de $\underline{\mu}$. Alors,

$$\sqrt{n} (f(\underline{X}_n) - f(\underline{\mu})) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}\left(\underline{0}, \overset{\circ}{f}(\underline{\mu}) \Sigma \overset{\circ}{f}'(\underline{\mu})\right).$$

Exemple 3.2.1 1. Soit (X_n) une suite i.i.d. de moyenne μ_1 et de variance σ^2 où $\mu_1 = E\{X_1\} \neq 0$

avec $\mu_4 = E\{X_1^4\} < +\infty$. Trouver la distribution asymptotique de $\frac{\sqrt{\frac{1}{n} \sum_{t=1}^n (X_t - \overline{X}_n)^2}}{\overline{X}_n}$ (où $\overline{X}_n = \frac{1}{n} \sum_{t=1}^n X_t$). On a :

$$\sqrt{n} \left[\begin{pmatrix} \sum_{i=1}^n X_i \\ \sum_{i=1}^n X_i^2 \end{pmatrix} - \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \right] \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, \Sigma),$$

où

$$\Sigma = \begin{pmatrix} \text{Var}(X_1) & \text{Cov}(X_1, X_1^2) \\ \text{Cov}(X_1, X_1^2) & \text{Var}(X_1^2) \end{pmatrix}.$$

Comme $\frac{\sqrt{\frac{1}{n} \sum_{t=1}^n (X_t - \bar{X}_n)^2}}{\bar{X}_n} = \sqrt{n} \frac{\sqrt{\frac{1}{n} \sum_{t=1}^n X_t^2 - (\bar{X}_n)^2}}{\bar{X}_n} = \sqrt{n} f(\bar{X}_n, \bar{X}_n^2)$, où $f(x, y) = \frac{\sqrt{y - x^2}}{x}$,
donc

$$\begin{aligned} \overset{\circ}{f}(x, y) &= \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y} \right) = \left(-\frac{y}{x^2 \sqrt{y - x^2}}, \frac{1}{2x \sqrt{y - x^2}} \right), \\ \overset{\circ}{f}(\mu_1, \mu_2) &= \left(\frac{\sigma^2 + \mu_1^2}{\sigma \mu_1^2}, \frac{1}{2\mu_1 \sigma} \right). \end{aligned}$$

Le théorème Delta montre que

$$\frac{\sqrt{\frac{1}{n} \sum_{t=1}^n (X_t - \bar{X}_n)^2}}{\bar{X}_n} \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N} \left(0, \overset{\circ}{f}(\mu_1, \mu_2) \Sigma \overset{\circ}{f}'(\mu_1, \mu_2) \right).$$

2. Le coefficient d'asymétrie (skewness) empirique,

$$\frac{\mu_3}{\sigma^3} = \beta = \frac{\mu_3}{(\sigma^2)^{\frac{3}{2}}} \xleftarrow{P} \hat{\beta}_n = \frac{\frac{1}{n} \sum_{t=1}^n (X_t - \bar{X}_n)^3}{\left(\frac{1}{n} \sum_{t=1}^n (X_t - \bar{X}_n)^2 \right)^{\frac{3}{2}}} \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, \sigma).$$

Le coefficient d'aplatissement (kurtosis) empirique,

$$\alpha = \frac{\mu_4}{(\sigma^2)^2} \xleftarrow{P} \hat{\alpha}_n = \frac{\frac{1}{n} \sum_{t=1}^n (X_t - \bar{X}_n)^4}{\left(\frac{1}{n} \sum_{t=1}^n (X_t - \bar{X}_n)^2 \right)^2}.$$

Notons que $\mu_4 = 3\sigma^4$ pour la loi $\mathcal{N}(0, \sigma^2)$.

3.3 Exercices

Exercice 3.3.1 Soit (X_n) une suite i.i.d. centrée de variance σ^2 et de moment d'ordre 4 fini. Montrer que

$$\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2 - E\{X_1^2\} \right) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, \text{Var}(X_1^2)).$$

Exercice 3.3.2 Soit (X_n) une suite i.i.d. de moyenne μ et de variance σ^2 . Trouver $\lambda_n \rightarrow \infty$ et a t.q

$$\lambda_n \left(\frac{1}{\bar{X}_n} - a \right) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, 1).$$

Exercice 3.3.3 Soient $a > 0$ et (X_n) une suite de variables aléatoires positives i.i.d. suivant la loi de Poisson de paramètre a . On note, pour tout $n \geq 1$, $S_n = X_1 + \dots + X_n$.

1. Montrer que $\left(\frac{\sqrt{S_n}}{\sqrt{n}} \right)$ converge presque sûrement vers \sqrt{a} .
2. Montrer que la suite de terme général $T_n = \sqrt{n} \left(\frac{S_n}{n} - a \right)$ converge en loi vers $T \sim \mathcal{N}(0, a)$

Exercice 3.3.4 Soit (X_n) une suite de variables aléatoires réelles i.i.d. suivant la loi gaussienne $\mathcal{N}(0, 1)$.

On note, pour tout $n \geq 1$, $S_n = X_1 + \dots + X_n$, $T_n = e^{\left(\frac{S_n - \frac{n}{2}}{n} \right)}$.

1. Justifier la convergence presque sûre de $\left(\frac{S_n}{n} \right)$ et préciser la limite.
2. En déduire que (T_n) converge presque sûrement vers 0.

Exercice 3.3.5 Soit (X_n) une suite de variables aléatoires réelles i.i.d. suivant la loi gaussienne $\mathcal{N}(0, 1)$.

1. Calculer, pour $t \in \mathbb{R}$, $E \{ e^{tX_1} \}$.

2. Montrer que la suite de terme général $T_n = \frac{\sum_{k=1}^n X_k^2}{\sum_{k=1}^n e^{X_k}}$ converge presque sûrement et préciser sa limite.

Exercice 3.3.6 Appliquer le théorème central limite à une suite (X_n) de variables aléatoires indépendantes de même loi de Poisson de paramètre 1 pour trouver la limite de la suite $\left(e^{-n} \sum_{j=0}^n \frac{n^j}{j!} \right)$.

Exercice 3.3.7 Soit $(X_i)_{1 \leq i \leq n}$ une suite de v.a.r. et i.i.d. de loi $\mathcal{N}(0, \sigma^2)$. Posons $\widetilde{S}_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$.

Montrer que $\widetilde{S}_n^2 \xrightarrow[n \rightarrow \infty]{P} \sigma^2$.

Exercice 3.3.8 Soit (X_n) une suite de v.a.r. et i.i.d de loi $\mathcal{N}(0, \sigma^2)$. Soit $T_n = \frac{\sqrt{n} \bar{X}_n}{\sqrt{\widetilde{S}_n^2}}$, où $\widetilde{S}_n^2 =$

$\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$. Trouver la distribution asymptotique de T_n .

Solutions

Solution 3.3.1 Notons que $\frac{1}{n} \sum_{t=1}^n (X_t - \bar{X}_n)^2 = \frac{1}{n} \sum_{t=1}^n X_t^2 - \bar{X}_n^2$. Donc $\sqrt{n} \left(\frac{1}{n} \sum_{t=1}^n (X_t - \bar{X}_n)^2 - E\{X_1^2\} \right) = \sqrt{n} \left(\frac{1}{n} \sum_{t=1}^n X_t^2 - \bar{X}_n^2 - E\{X_1^2\} \right)$. Par le T.C.L appliqué à la suite (X_n^2) , on a :

$$\sqrt{n} \left(\frac{1}{n} \sum_{t=1}^n X_t^2 - E\{X_1^2\} \right) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, \text{Var}(X_1^2)),$$

et

$$\sqrt{n}\bar{X}_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, \sigma^2), \quad \bar{X}_n \xrightarrow[n \rightarrow \infty]{P} 0 \implies \sqrt{n}\bar{X}_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} 0.$$

En appliquant le théorème de Slutsky, on obtient le résultat.

Solution 3.3.2 On sait que $\sqrt{n}(\bar{X}_n - \mu) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, \sigma^2)$. Soit $f : \mathbb{R}^* \rightarrow \mathbb{R}$, $f(x) = \frac{1}{x}$. D'après le théorème de Cramer

$$\sqrt{n} \left(\frac{1}{\bar{X}_n} - \frac{1}{\mu} \right) \longrightarrow -\frac{1}{\mu^2} \mathcal{N}(0, \sigma^2) \equiv \mathcal{N} \left(0, \left(\frac{\sigma}{\mu^2} \right)^2 \right) \equiv \frac{\sigma}{\mu^2} \mathcal{N}(0, 1).$$

Donc

$$\lambda_n \left(\frac{1}{\bar{X}_n} - a \right) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, 1),$$

où $\lambda_n = \frac{\mu^2}{\sigma} \sqrt{n}$ et $a = \frac{1}{\mu}$.

Solution 3.3.3 1. Comme la suite (X_n) est i.i.d. et X_1 est intégrable (car $E\{X_1\} = a$), alors d'après la loi forte des grands nombres $\frac{S_n}{n} \xrightarrow{p.s.} E\{X_1\} = a$. Puisque $x \mapsto \sqrt{x}$ est continue au point a , donc $\frac{\sqrt{S_n}}{\sqrt{n}} \xrightarrow{p.s.} \sqrt{a}$.

2. La suite (X_n) est i.i.d. et X_1 est de carré intégrable, donc le théorème central limite implique que T_n converge en loi vers une v.a.r. $T \sim \mathcal{N}(0, \text{Var}(X_1) = a)$

Solution 3.3.4 1. Puisque les (X_n) sont i.i.d. et que X_1 est intégrable, la loi forte des grands nombres donne la convergence presque sûre de $\left(\frac{S_n}{n} \right)$ vers $E\{X_1\} = 0$.

2. Pour tout $n \geq 1$, $\log T_n = S_n - \frac{n}{2} = n \left(\frac{S_n}{n} - \frac{1}{2} \right)$. Puisque $\left(\frac{S_n}{n} \right)$ converge presque sûrement vers 0, $\log T_n \xrightarrow{p.s.} -\infty$ et (T_n) converge presque sûrement vers 0.

Solution 3.3.5 1. On a, pour tout réel t ,

$$E \{ e^{tX_1} \} = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{tx} e^{-\frac{x^2}{2}} dx = e^{\frac{t^2}{2}} \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-\frac{(x-t)^2}{2}} dx = e^{\frac{t^2}{2}},$$

puisqu'on reconnaît la densité de la loi gaussienne $\mathcal{N}(t, 1)$.

2. Comme les variables (X_n) sont i.i.d., alors les variables (X_n^2) et (e^{X_n}) sont i.i.d. et comme X_1^2 et e^{X_1} étant intégrables, donc la loi forte des grands nombres donne la convergence presque sûre des suites de terme général $\frac{1}{n} \sum_{k=1}^n X_k^2$ et $\frac{1}{n} \sum_{k=1}^n e^{X_k}$ respectivement vers $E \{ X_1^2 \} = 1$ et $E \{ e^{X_1} \} = e^{\frac{1}{2}}$. La suite (T_n) converge presque sûrement vers $e^{-\frac{1}{2}}$.

Solution 3.3.6 Comme (X_n) est une suite de variables aléatoires indépendantes suivant la même loi de Poisson $\mathcal{P}(1)$, alors, on sait que $S_n = \sum_{j=1}^n X_j \sim \mathcal{P}(n)$ avec, $E \{ S_n \} = n$ et $\text{Var} (S_n) = n$. On applique le théorème central limite $\frac{S_n - n}{\sqrt{n}} \xrightarrow[n \rightarrow \infty]{\mathcal{L}} S \sim \mathcal{N}(0, 1)$. Donc,

$$P \left(\frac{S_n - n}{\sqrt{n}} \leq 0 \right) \xrightarrow[n \rightarrow \infty]{} P(S \leq 0) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^0 e^{-\frac{x^2}{2}} dx = \frac{1}{2},$$

d'autre part,

$$P \left(\frac{S_n - n}{\sqrt{n}} \leq 0 \right) = P(S_n \leq n) = \sum_{j=0}^n P(S_n = j) = \sum_{j=0}^n \frac{n^j}{j!} e^{-n} = e^{-n} \sum_{j=0}^n \frac{n^j}{j!}.$$

D'où le résultat : $e^{-n} \sum_{j=0}^n \frac{n^j}{j!} \xrightarrow[n \rightarrow \infty]{} \frac{1}{2}$.

Solution 3.3.7 On a : pour tout n ,

$$\widetilde{S}_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2 = \frac{1}{n-1} \left(\sum_{i=1}^n X_i^2 - n\bar{X}_n^2 \right) = \frac{n}{n-1} \left(\frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}_n^2 \right),$$

donc, d'après la loi faible des grands nombres $\widetilde{S}_n^2 \xrightarrow[n \rightarrow \infty]{P} 1 (E \{ X_1^2 \} - 0^2) = \sigma^2$.

Solution 3.3.8 On a : pour tout n , $T_n = \frac{\sqrt{n} \frac{\bar{X}_n}{\sigma}}{\sqrt{\frac{\widetilde{S}_n^2}{\sigma^2}}}$, alors on applique le théorème central limite $\sqrt{n} \frac{\bar{X}_n}{\sigma} \xrightarrow[n \rightarrow \infty]{\mathcal{L}}$

$\mathcal{N}(0, 1)$ et comme $\widetilde{S}_n^2 \xrightarrow[n \rightarrow \infty]{P} \sigma^2$ (voir., Exercice 3.3.7) Donc, $T_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} T \sim \mathcal{N}(0, 1)$. D'où le résultat.

Deuxième partie

Statistique inférentielle

Chapitre 4

Échantillonnage. Modèles statistiques

On veut, à partir d'un échantillon de la population, déduire des informations sur cette population. Le problème qui se pose alors est le suivant : comment choisir une partie de la population qui reproduit le plus fidèlement possible ses caractéristiques. C'est le problème de l'échantillonnage. Le but de ce chapitre est de présenter les différentes méthodes d'échantillonnage.

4.1 Méthodes d'échantillonnage

On distingue deux grandes catégories de méthodes d'échantillonnage :

- l'échantillonnage par choix raisonné ;
- l'échantillonnage aléatoire.

4.1.1 Échantillonnage par choix raisonné

Les méthodes **d'échantillonnage par choix raisonné** incluent diverses techniques qui consistent à construire l'échantillon sur la base d'informations connues relatives à la population étudiée. Ces méthodes comportent une part d'arbitraire ne permettant pas d'évaluer la précision des estimations, mais elles présentent dans certains cas des avantages de coût et de rapidité par rapport à la méthode de l'échantillonnage aléatoire.

L'échantillonnage par choix raisonné est aussi appelé échantillonnage empirique. La méthode principale est celle des quotas. Selon cette méthode, l'enquêteur sélectionne les unités, en fonction de quotas qui lui sont donnés. Dans le cas d'une enquête auprès des ménages ou d'individus, ces quotas portent géné-

ralement sur des critères socio-démographiques tels que le sexe, l'âge ou la catégorie socio-professionnelle. Ils sont établis à partir de statistiques officielles et visent à constituer un échantillon possédant la même structure que la population. Dans la limite des quotas, le choix des unités physiques qui feront partie de l'échantillon est laissé à la discrétion de l'enquêteur dans la zone géographique attribuée. Le hasard intervient donc d'une façon limitée dans la sélection des unités de la population qui feront partie de l'échantillon.

La méthode des quotas est très fréquemment utilisée par les entreprises privées en raison de ses avantages pratiques. En effet, sa mise en oeuvre est rapide car il n'y a pas besoin de tester tous les éléments de la population pour effectuer l'échantillonnage. Elle ne nécessite pas de base de sondage, c'est-à-dire une liste exhaustive des éléments de la population considérée. En permettant un gain de temps, elle est moins coûteuse que les échantillonnages probabilistes. Toutefois, la sélection de l'échantillon n'étant pas basée sur des méthodes aléatoires, il devient difficile d'évaluer objectivement à quel point l'échantillon est représentatif et de ce fait, il n'est pas possible de connaître la marge d'erreur des résultats obtenus à partir de l'échantillon même.

4.1.2 Échantillonnage aléatoire

L'**échantillonnage aléatoire** correspond à des méthodes de tirage de l'échantillon où chaque unité de la population a une probabilité positive et connue d'être sélectionnée. Ces méthodes permettent non seulement d'estimer les paramètres de la population, mais encore d'obtenir une mesure de l'erreur susceptible d'avoir été commise.

Les trois types d'échantillonnage aléatoire les plus courants sont : l'échantillonnage aléatoire simple, l'échantillonnage stratifié et l'échantillonnage par grappes.

Échantillonnage aléatoire simple

L'échantillonnage aléatoire simple, ou échantillonnage probabiliste simple est basé sur le principe que tous les éléments de la population ont une probabilité égale (non nulle) de faire partie de l'échantillon. La population considérée est généralement finie. Soit N le nombre d'unités qui composent la population considérée. Au cours d'un tirage aléatoire, on attribuera à chaque unité de la population la même probabilité d'être choisie soit N^{-1} . En prélevant au hasard un échantillon de taille n d'une population de N unités, les valeurs obtenues pour les n tirages sont aléatoires. Si l'extraction est réalisée sans remettre les unités tirées dans la population, il s'agit d'un échantillon sans remplacement. Si, en revanche, l'extraction est faite avec

remise, l'échantillon est avec remplacement.

L'échantillonnage avec remise est utilisé très rarement en pratique, car il y a peu d'intérêt de détenir une même unité deux fois dans l'échantillon. Dans certaines situations, cependant, comme le cas d'échantillonnage d'une faune, l'utilisation d'un échantillonnage avec remise est pratiquement inévitable.

Pour effectuer un échantillonnage aléatoire simple, il faut d'une part, avoir accès au préalable à une liste complète des éléments de la population et d'autre part, utiliser une méthode de tirage qui garantisse la même probabilité de sélection à tous les éléments de la liste.

Ainsi, pour effectuer le tirage en s'assurant que le choix de l'échantillon se fait au hasard, on utilise généralement des tables de nombres aléatoires ou des programmes de génération de nombres aléatoires.

Échantillonnage stratifié

L'échantillonnage stratifié consiste à découper la population en strates ou classes homogènes par rapport à l'ensemble de la population puis à réaliser dans chaque strate un échantillonnage aléatoire simple. La méthode d'échantillonnage stratifié est généralement utilisée lorsque la population étudiée est hétérogène à certains égards. La stratification nécessite donc une connaissance préalable de la structure de cette dernière.

Échantillonnage par grappes

L'échantillonnage par grappes consiste à tirer au hasard des ensembles d'unités de la population, ou grappes, et ensuite à mener l'enquête sur toutes les unités de ces grappes. Les grappes sont souvent constituées par des unités de type géographique comme les quartiers d'une ville. La méthode consiste à diviser une ville en quartiers, puis à sélectionner les quartiers qui feront partie de l'échantillon. On mènera ensuite l'enquête sur toutes les personnes ou ménages, habitant dans les quartiers choisis.

Dans la suite de ce chapitre, nous allons étudier comment se comporte un échantillon (éléments pris au hasard) dans une population dont on connaît les caractéristiques statistiques (lois,...) d'une variable considérée X à valeurs dans un espace (E, \mathcal{E}) . Dans ce cas, prendre un échantillon aléatoire de taille n consiste à considérer n réalisations de X ou encore considérer n variables aléatoires X_1, X_2, \dots, X_n indépendantes, de même loi que X .

4.2 Échantillon

Une étude statistique portant sur tous les éléments d'une population étant impossible à réaliser (trop grand nombre d'individus à étudier), il faut obtenir des résultats fiables sur les caractéristiques d'une population en se limitant à l'étude des éléments ou unités d'un échantillon.

Définition 4.2.1 *On appelle échantillon aléatoire de taille n (en bref n -échantillon) d'une variable aléatoire X une suite finie de n variables aléatoires indépendantes et de même loi (ou v.a. i.i.d.) suivant la même loi que X . Cette loi est appelée la loi mère de l'échantillon et la v.a. X appelée variable aléatoire parente.*

Remarque 4.2.1 *Pour des raisons de commodité, nous avons supposé que les X_k sont mutuellement indépendantes. Dans certains cas, l'indépendance deux à deux sera suffisante.*

Remarque 4.2.2 *On distinguera la notion d'échantillon aléatoire (X_1, X_2, \dots, X_n) dont on peut dire qu'elle se réfère à des résultats potentiels avant expérience, de celle d'échantillon réalisé (x_1, x_2, \dots, x_n) correspondant aux valeurs observées après expérience.*

Pour obtenir un échantillon observé (x_1, x_2, \dots, x_n) , on effectue n épreuves identiques et indépendantes, pour lesquelles la variable aléatoire X_k (associée à la k^e épreuve) a pris la valeur x_k .

L'objectif de ce chapitre est d'étudier certaines caractéristiques de l'échantillon aléatoire, essentiellement sa moyenne et sa variance, en relation avec celles de la loi mère. Une telle caractéristique est une v.a. qui prend le nom de «statistique» dans le contexte de l'échantillonnage, selon la définition suivante.

Définition 4.2.2 *Soit (X_1, X_2, \dots, X_n) un n -échantillon, on appelle statistique de l'échantillon toute v.a. $T_n = h(X_1, X_2, \dots, X_n)$, fonction de X_1, X_2, \dots, X_n .*

On peut concrétiser la loi d'une statistique (donc d'une caractéristique, telle la moyenne de l'échantillon) en imaginant une simulation en très grand nombre d'échantillons de taille n , en calculant pour chacun d'eux la valeur prise par la statistique et en étudiant la distribution de ces valeurs. De façon imagée on peut dire qu'il s'agit de la distribution d'échantillonnage de la statistique sur «l'univers» de tous les échantillons possibles. Notons qu'une statistique peut être une fonction à valeurs dans \mathbb{R} , \mathbb{R}^2 ou \mathbb{R}^r . En particulier les moments empiriques ci-après sont à valeurs dans \mathbb{R} . Les définitions qui suivent se rapportent toutes à un échantillon aléatoire noté (X_1, X_2, \dots, X_n) .

4.3 Quelques statistiques classiques

4.3.1 Moyenne, variance, moments empiriques d'un échantillon

Définition 4.3.1 On appelle *moyenne de l'échantillon aléatoire* (X_1, X_2, \dots, X_n) ou **moyenne empirique**, la statistique, notée \bar{X}_n , définie par : $\bar{X}_n := \frac{1}{n} \sum_{k=1}^n X_k$. Sa réalisation est $\bar{x}_n := \frac{1}{n} \sum_{k=1}^n x_k$.

Définition 4.3.2 Soit (X_1, X_2, \dots, X_n) un n -échantillon d'une population de moyenne μ et de variance σ^2 . Si μ est connu, on appelle *variance empirique*, la statistique, notée S_n^2 , définie par : $S_n^2 := \frac{1}{n} \sum_{k=1}^n (X_k - \mu)^2$.

Si non¹, on appelle **variance empirique**, la statistique, notée \overline{S}_n^2 , définie par : $\overline{S}_n^2 := \frac{1}{n} \sum_{k=1}^n (X_k - \bar{X}_n)^2$. Sa réalisation est $\overline{s}_n^2 := \frac{1}{n} \sum_{k=1}^n (x_k - \bar{x}_n)^2$.

Remarque 4.3.1 La notation S_n^2 ne désigne pas un carré, l'exposant 2 désigne la somme de carrés.

Définition 4.3.3 On appelle **variance empirique corrigée**, la statistique, notée \widetilde{S}_n^2 , définie par : $\widetilde{S}_n^2 := \frac{1}{n-1} \sum_{k=1}^n (X_k - \bar{X}_n)^2$. Sa réalisation est $\widetilde{s}_n^2 := \frac{1}{n-1} \sum_{k=1}^n (x_k - \bar{x}_n)^2$.

Nous commençons maintenant à établir certaines relations entre les lois de ces statistiques et la loi mère.

Proposition 4.3.1 Soient μ et σ^2 , respectivement la moyenne et la variance de la loi mère. On a : $E\{\bar{X}_n\} = \mu$ et $Var(\bar{X}_n) = \frac{\sigma^2}{n}$.

Preuve. En effet : $E\{\bar{X}_n\} = E\left\{\frac{1}{n} \sum_{k=1}^n X_k\right\} = \frac{1}{n} \sum_{k=1}^n E\{X_k\} = \frac{1}{n} \sum_{k=1}^n \mu = \mu$. Puis, en raison de l'indépendance des X_k :

$$Var(\bar{X}_n) = Var\left(\frac{1}{n} \sum_{k=1}^n X_k\right) = \frac{1}{n^2} \sum_{k=1}^n Var(X_k) = \frac{1}{n^2} \sum_{k=1}^n \sigma^2 = \frac{\sigma^2}{n}.$$

■

Proposition 4.3.2 Soient μ et σ^2 , respectivement la moyenne et la variance de la loi mère. On a : $E\{S_n^2\} = \sigma^2$ et $Var(S_n^2) = \frac{\mu_4 - \sigma^4}{n}$.

¹En pratique, μ est rarement connu exactement : on le remplace par un estimateur, la moyenne empirique, et on introduit la variance empirique.

Preuve. En effet : $E\{S_n^2\} = E\left\{\frac{1}{n}\sum_{k=1}^n (X_k - \mu)^2\right\} = \frac{1}{n}\sum_{k=1}^n E\{(X_k - \mu)^2\} = \frac{1}{n}\sum_{k=1}^n \sigma^2 = \sigma^2$. Puis, en raison de l'indépendance des X_k :

$$\begin{aligned} \text{Var}(S_n^2) &= \text{Var}\left(\frac{1}{n}\sum_{k=1}^n (X_k - \mu)^2\right) = \frac{1}{n^2}\sum_{k=1}^n \text{Var}\left((X_k - \mu)^2\right) \\ &= \frac{1}{n}\text{Var}\left((X_1 - \mu)^2\right) = \frac{\mu_4 - \sigma^4}{n}, \end{aligned}$$

où $\mu_4 = E\{(X_1 - \mu)^4\}$. ■

Proposition 4.3.3 *La moyenne de la loi de la variance empirique (resp. la variance empirique corrigée) est : $E\{\overline{S_n^2}\} = \frac{n-1}{n}\sigma^2$ (resp. $E\{S_n^2\} = E\{\widetilde{S_n^2}\} = \sigma^2$).*

Preuve. En effet :

$$\frac{1}{n}\sum_{k=1}^n (X_k - \overline{X}_n)^2 = \frac{1}{n}\sum_{k=1}^n ((X_k - \mu) - (\overline{X}_n - \mu))^2 = \frac{1}{n}\sum_{k=1}^n (X_k - \mu)^2 - (\overline{X}_n - \mu)^2.$$

D'où : $E\{\overline{S_n^2}\} = \frac{1}{n}\sum_{k=1}^n \text{Var}(X_k) - \text{Var}(\overline{X}_n) = \sigma^2 - \frac{\sigma^2}{n} = \frac{n-1}{n}\sigma^2$. Comme $n\overline{S_n^2} = (n-1)\widetilde{S_n^2}$ donc $E\{\widetilde{S_n^2}\} = \frac{n}{n-1}\frac{n-1}{n}\sigma^2 = \sigma^2$. ■

Exemple 4.3.1 *Si la loi mère est $\mathcal{N}(\mu, \sigma^2)$ alors \overline{X}_n est gaussienne, en tant que combinaison linéaire de gaussiennes indépendantes. Par conséquent : $\overline{X}_n \sim \mathcal{N}(\mu, \frac{\sigma^2}{n})$.*

Dans ce cas, \overline{X}_n et $\widetilde{S_n^2}$ sont des v.a. indépendantes.

Définition 4.3.4 *On appelle moment empirique d'ordre r , la statistique, notée M_r , définie par : $M_r := \frac{1}{n}\sum_{k=1}^n X_k^r$.*

Définition 4.3.5 *On appelle moment empirique centré d'ordre r , la statistique, notée \overline{M}_r , définie par : $\overline{M}_r := \frac{1}{n}\sum_{k=1}^n (X_k - \overline{X}_n)^r$.*

Nous abordons maintenant trois lois omniprésentes en statistique car liées aux distributions d'échantillonnage de moyennes et de variances dans le cas gaussien.

4.3.2 Lois usuelles continues

Définition 4.3.6 (Loi du Khi-deux) Soit (X_n) une suite de variables aléatoires i.i.d. de loi $\mathcal{N}(0, 1)$.

Alors la v.a. $Y = \sum_{k=1}^n X_k^2$ suit une loi appelée loi du Khi-deux à n degrés de liberté, notée $\chi^2(n) \equiv \Gamma\left(\frac{n}{2}, \frac{1}{2}\right)$.

La densité de la loi du Khi-deux à n degrés de liberté est

$$f_Y(x) = \frac{1}{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)} x^{\frac{n}{2}-1} e^{-\frac{x}{2}} \mathbb{I}_{\{x>0\}} \quad \forall x \in \mathbb{R},$$

où $\Gamma(a) = \int_0^{+\infty} x^{a-1} e^{-x} dx$. La moyenne de Y est égale au nombre de degrés de liberté n , sa variance est $2n$.

Proposition 4.3.4 Si $Y_1 \sim \chi^2(n_1)$, $Y_2 \sim \chi^2(n_2)$, Y_1 et Y_2 indépendantes, alors $Y_1 + Y_2 \sim \chi^2(n_1 + n_2)$.

Cette proposition est évidente de par la définition de la loi du Khi-deux.

Nous revenons maintenant sur la loi de \widetilde{S}_n^2 dans le cas d'un échantillon de loi mère gaussienne.

Théorème 4.3.1 Soit un n -échantillon (X_1, X_2, \dots, X_n) de loi $\mathcal{N}(\mu, \sigma^2)$ on a : $\frac{(n-1)\widetilde{S}_n^2}{\sigma^2} \sim \chi^2(n-1)$.

Par conséquent : $E\left\{\widetilde{S}_n^2\right\} = \sigma^2$ et $Var\left(\widetilde{S}_n^2\right) = \frac{2\sigma^4}{n-1}$.

Preuve. En reprenant les développements qui suivent l'énoncé de la proposition 4.3.3, on établit que :

$$\begin{aligned} \sum_{k=1}^n (X_k - \mu)^2 &= \sum_{k=1}^n (X_k - \bar{X}_n)^2 + n(\bar{X}_n - \mu)^2, \\ \sum_{k=1}^n \left(\frac{X_k - \mu}{\sigma}\right)^2 &= \frac{\sum_{k=1}^n (X_k - \bar{X}_n)^2}{\sigma^2} + \left(\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}\right)^2 \\ &= \frac{(n-1)\widetilde{S}_n^2}{\sigma^2} + \left(\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}\right)^2. \end{aligned}$$

Les deux termes de droite sont, respectivement, $\frac{(n-1)\widetilde{S}_n^2}{\sigma^2}$ et le carré de $\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}$ qui est une gaussienne centrée-réduite. Ces termes aléatoires étant indépendants et les v.a. $\frac{X_k - \mu}{\sigma}$ étant indépendantes de loi $\mathcal{N}(0, 1)$ on a, en termes de fonctions caractéristiques :

$$\left(\frac{1}{1-2t}\right)^{n/2} = \varphi_{\frac{(n-1)\widetilde{S}_n^2}{\sigma^2}}(t) \left(\frac{1}{1-2t}\right)^{1/2} \quad \text{si } r < 1/2.$$

Finalement : $\varphi_{\frac{(n-1)\widetilde{S}_n^2}{\sigma^2}}(t) = \left(\frac{1}{1-2t}\right)^{(n-1)/2}$, ce qui prouve le théorème. ■

Définition 4.3.7 (Loi de Student) Soient X et Y deux v.a. indépendantes telles que $X \sim \mathcal{N}(0, 1)$ et $Y \sim \chi^2(n)$. Alors la v.a. $T = \frac{X}{\sqrt{\frac{Y}{n}}}$ suit une loi appelée loi de Student à n degrés de liberté, notée $\mathcal{T}(n)$.

La densité de la loi de Student à n degrés de liberté est :

$$f_T(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{\pi n} \Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}, \quad \forall x \in \mathbb{R}.$$

La moyenne de T est 0 si $n \geq 2$, sa variance est $\frac{n}{n-2}$ si $n \geq 3$.

Théorème 4.3.2 Soit un n -échantillon (X_1, X_2, \dots, X_n) de loi $\mathcal{N}(\mu, \sigma^2)$ on a : $\frac{\bar{X}_n - \mu}{\sqrt{\widetilde{S}_n^2/n}} \sim \mathcal{T}(n-1)$.

Preuve. La démonstration est immédiate, il suffit d'écrire la v.a. $\frac{\bar{X}_n - \mu}{\sqrt{\widetilde{S}_n^2/n}} = \frac{\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}}{\sqrt{\frac{(n-1)\widetilde{S}_n^2/\sigma^2}{n-1}}}$. ■

Définition 4.3.8 (Loi de Fisher-Snedecor) Soient X et Y deux v.a. indépendantes telles que $X \sim \chi^2(n_1)$ et $Y \sim \chi^2(n_2)$. Alors la v.a. $F = \frac{X/n_1}{Y/n_2}$ suit une loi de Fisher-Snedecor à n_1 degrés de liberté au numérateur et n_2 degrés de liberté au dénominateur, notée $\mathcal{F}(n_1, n_2)$. En bref on l'appellera loi de Fisher. La densité de la loi $\mathcal{F}(n_1, n_2)$ est :

$$f_F(x) = \frac{\Gamma\left(\frac{n_1+n_2}{2}\right)}{\Gamma\left(\frac{n_1}{2}\right)\Gamma\left(\frac{n_2}{2}\right)} \left(\frac{n_1}{n_2}\right)^{\frac{n_1}{2}} x^{\frac{n_1}{2}-1} \left(1 + \frac{n_1}{n_2}x\right)^{-\frac{n_1+n_2}{2}} \mathbb{I}_{\{x>0\}}, \quad \forall x \in \mathbb{R}.$$

Si $n_2 \geq 3$ sa moyenne existe et est égale à $\frac{n_2}{n_2-2}$. Si $n_2 \geq 5$ sa variance existe et est égale à $\frac{2n_2^2(n_1+n_2-2)}{n_1(n_2-2)^2(n_2-4)}$.

La proposition suivante permet une économie de tables.

Proposition 4.3.5 Soit F une v.a. suit une loi de Fisher telle que $F \sim \mathcal{F}(n_1, n_2)$, alors $\frac{1}{F} \sim \mathcal{F}(n_2, n_1)$.

Preuve. La démonstration est évidente de par la définition de la loi de Fisher. ■

4.4 Modèle statistique

Définition 4.4.1 Un modèle statistique est défini par la donnée d'une caractéristique vectorielle (X_1, X_2, \dots, X_n) et d'une famille de lois de probabilité de X notée $(P_\theta)_{\theta \in \Theta}$. Lorsque la famille de lois de probabilité $(P_\theta)_{\theta \in \Theta}$

peut être indexée par un paramètre θ dont l'ensemble des valeurs possibles, noté Θ espace des paramètres (ou espace paramétrique), est un sous-ensemble de \mathbb{R}^r où r est la dimension du paramètre θ , le modèle est appelé modèle paramétrique. Dans le cas contraire, le modèle est appelé modèle non paramétrique.

Remarque 4.4.1 Le modèle statistique d'échantillonnage paramétrique est noté le triplet $(E, \mathcal{E}, P_\theta; \theta \in \Theta)^n$.

Remarque 4.4.2 La famille de lois de probabilités $\{P_\theta, \theta \in \Theta\}$ à laquelle appartient la loi de X est supposée connue; seul le paramètre θ est inconnu.

On considère un phénomène aléatoire et une variable aléatoire réelle X qui lui est lié. Le type de la loi X est supposé connu et dépend d'un paramètre θ inconnu qui varie dans un ensemble Θ . L'objectif est de donner une estimation de la valeur du paramètre θ à partir d'un échantillon observé (x_1, x_2, \dots, x_n) de l'échantillon aléatoire (X_1, X_2, \dots, X_n) , où les X_k sont des variables aléatoires réelles indépendantes et de même loi que X . Il y a deux types d'estimation :

1. **l'estimation ponctuelle, on cherche à trouver une valeur approchée de θ ,**
2. **l'estimation par intervalle de confiance, on cherche à déterminer un intervalle dans lequel θ à une certaine probabilité de se trouver.**

Chapitre 5

Estimation ponctuelle

Le but de la théorie de l'estimation est de choisir, parmi toutes les statistiques possibles, le meilleur estimateur, c'est-à-dire celui qui donnera une estimation ponctuelle la plus proche possible du paramètre et ceci, quel que soit l'échantillon.

5.1 Estimateur et estimation

La théorie générale de l'estimation repose sur la notion d'estimateur.

Définition 5.1.1 Si (X_1, X_2, \dots, X_n) est un échantillon aléatoire de la variable aléatoire parente X , alors nous appelons **estimateur** du paramètre θ toute fonction h de l'échantillon aléatoire (X_1, X_2, \dots, X_n) , cet estimateur, noté $\hat{\theta}_n$, est défini par : $\hat{\theta}_n = h(X_1, X_2, \dots, X_n)$.

Bien évidemment, cette fonction ou cette «formule» h n'est pas choisie au hasard. L'idée est de trouver une fonction qui combine les réalisations de l'échantillon de sorte à révéler de l'information sur le paramètre d'intérêt θ . Nous verrons comment déduire cette fonction, c'est-à-dire comment construire un estimateur, dans la sous-section 5.2.1 consacrée aux méthodes d'estimation. Mais à ce stade, considérons quelques exemples d'estimateurs.

Exemple 5.1.1 La moyenne empirique est un estimateur de $E\{X\}$. En effet, \bar{X}_n est une fonction des variables X_1, X_2, \dots, X_n telle que :

$$\bar{X}_n = \frac{1}{n} \sum_{k=1}^n X_k = h(X_1, X_2, \dots, X_n).$$

Exemple 5.1.2 La variance empirique est un estimateur de la variance σ^2 .

Remarque 5.1.1 À priori l'estimateur $\hat{\theta}_n$ est à valeurs dans un ensemble Θ , contenant l'ensemble des valeurs possibles du paramètre θ .

Remarque 5.1.2 $\hat{\theta}_n$ est une v.a. de loi de probabilité qui dépend du paramètre θ .

Remarque 5.1.3 $\hat{\theta}_n$ peut être univarié ou multivarié.

Si l'estimateur $\hat{\theta}_n$ est une variable aléatoire (continue ou discrète), elle est nécessairement caractérisée par une fonction de distribution (fonction de densité dans le cas continu ou fonction de masse dans le cas discret).

Définition 5.1.2 La distribution de probabilité d'un estimateur (ou d'une statistique) est appelée **distribution d'échantillonnage**.

Exemple 5.1.3 Si la loi mère est $\mathcal{N}(\mu, \sigma^2)$ alors \bar{X}_n est un estimateur du paramètre μ . La loi exacte de l'estimateur \bar{X}_n est alors la suivante : $\bar{X}_n = \frac{1}{n} \sum_{k=1}^n X_k \sim \mathcal{N}(\mu, \frac{\sigma^2}{n})$.

De même pour $\hat{\theta} = \frac{1}{2} \sum_{k=1}^2 X_k$, $\hat{\theta}$ est un estimateur du paramètre μ et $\hat{\theta} \sim \mathcal{N}(\mu, \frac{\sigma^2}{2})$.

Comme pour toute variable aléatoire, on doit distinguer la variable aléatoire elle-même, de sa réalisation. Cette réalisation correspond à une estimation.

Définition 5.1.3 Une fois l'échantillon prélevé, nous disposons de n valeurs observées x_1, x_2, \dots, x_n , ce qui nous fournit une valeur $h(x_1, x_2, \dots, x_n)$ qui est une réalisation de $\hat{\theta}_n$ et que nous appelons **estimation (ponctuelle)** du paramètre θ .

Remarque 5.1.4 Nous distinguons la variable aléatoire $\hat{\theta}_n$ de sa valeur observée, notée $\hat{\theta}_n(x_1, x_2, \dots, x_n)$.

Remarque 5.1.5 L'estimation $\hat{\theta}_n(x_1, x_2, \dots, x_n)$ ne dépend que de l'échantillon observé (x_1, x_2, \dots, x_n) . L'estimation $\hat{\theta}_n(x_1, x_2, \dots, x_n)$ ne dépend pas de θ .

Reprenons l'exemple précédent.

Exemple 5.1.4 $\hat{\theta}$ est un estimateur de l'espérance μ . Pour une réalisation $(x_1, x_2) = (10; 2)$ de l'échantillon, on obtient une estimation (ponctuelle) du paramètre μ égale à : $\hat{\theta}(x_1, x_2) = \frac{10+2}{2} = 6$.

Ainsi à ce stade du chapitre, il convient de bien distinguer la notion d'estimateur de la notion d'estimation (réalisation) :

- Estimateur (variable aléatoire) : $\widehat{\theta}_n$.
- Estimation (constante) : $\widehat{\theta}_n(x_1, x_2, \dots, x_n)$.

Propriétés d'un estimateur

Le choix d'un estimateur va reposer sur ses qualités. Le premier défaut possible concerne la possibilité de comporter un biais.

Définition 5.1.4 (Biais d'un estimateur) Si pour tout θ de Θ , $\widehat{\theta}_n$ admet une espérance alors le biais de $\widehat{\theta}_n$ se définit par $b_\theta(\widehat{\theta}_n) = E_\theta\{\widehat{\theta}_n\} - \theta$.

Remarque 5.1.6 Notons que la moyenne de $\widehat{\theta}_n$, $E_\theta\{\widehat{\theta}_n\}$, est indicée par θ pour rappeler qu'elle est liée à la valeur inconnue de θ .

Définition 5.1.5 $\widehat{\theta}_n$ est un **estimateur sans biais** (ou non biaisé) du paramètre θ si pour tout θ de Θ , $b_\theta(\widehat{\theta}_n) = 0$, i.e., $E_\theta\{\widehat{\theta}_n\} = \theta$.

Exemple 5.1.5 Comme $E_\mu\{\overline{X}_n\} = \mu$, alors la moyenne empirique \overline{X}_n est un estimateur sans biais de l'espérance μ . Une estimation de μ est la moyenne observée \overline{x}_n .

Exemple 5.1.6 Si μ est connu et comme $E\{S_n^2\} = \sigma^2$, alors S_n^2 est un estimateur sans biais de σ^2 .

Exemple 5.1.7 Si μ est inconnu et comme $E\{\widetilde{S}_n^2\} = \sigma^2$, alors la variance empirique corrigée \widetilde{S}_n^2 est un estimateur sans biais de la variance σ^2 .

Définition 5.1.6 Un estimateur $\widehat{\theta}_n$ est **asymptotiquement sans biais** pour θ si pour tout θ de Θ , $\lim_n b_\theta(\widehat{\theta}_n) = 0$, i.e., $\lim_n E_\theta\{\widehat{\theta}_n\} = \theta$.

Exemple 5.1.8 Si μ est inconnu et comme $E\{\overline{S}_n^2\} = (1 - \frac{1}{n})\sigma^2 \neq \sigma^2$, alors la variance empirique \overline{S}_n^2 est un estimateur biaisé de la variance σ^2 , le biais est égal à $\frac{\sigma^2}{n}$, mais asymptotiquement sans biais de σ^2 (parce que $(1 - \frac{1}{n})\sigma^2 \rightarrow \sigma^2$).

Exemple 5.1.9 Soit la famille des lois continues $\mathcal{U}_{[0,\theta]}$. Montrons que $M_n = \max\{X_1, X_2, \dots, X_n\}$ est biaisé pour θ . Pour une loi de fonction de répartition $F_{\theta, X_1}(x)$, la fonction de répartition du maximum

de l'échantillon M_n est $F_{\theta, X_1}^n(x)$. Dans la situation particulière considérée $\forall x, F_{\theta, X_1}(x) = \frac{x}{\theta} \mathbb{I}_{[0, \theta]}(x)$ et la densité de M_n est donc : $\forall x, f_{\theta, M_n}(x) = \frac{n}{\theta} \left(\frac{x}{\theta}\right)^{n-1} \mathbb{I}_{[0, \theta]}(x)$, d'où

$$E_{\theta} \{M_n\} = \int_0^{\theta} x \frac{n}{\theta} \left(\frac{x}{\theta}\right)^{n-1} dx = \int_0^{\theta} n \left(\frac{x}{\theta}\right)^n dx = \frac{n}{n+1} \frac{\theta^{n+1}}{\theta^n} = \frac{n}{n+1} \theta.$$

La qualité d'un estimateur ne dépend pas seulement de la proximité de son espérance avec la vraie valeur du paramètre à estimer, mais aussi de la dispersion des valeurs qu'il prend autour de cette valeur à estimer.

Définition 5.1.7 Soit $\hat{\theta}_n$ un estimateur de θ . Si pour tout θ de Θ , $\hat{\theta}_n$ admet une variance, nous mesurons la précision de $\hat{\theta}_n$ par l'écart quadratique moyen, noté $r_{\theta}(\hat{\theta}_n)$: $r_{\theta}(\hat{\theta}_n) = E_{\theta} \left\{ \left(\hat{\theta}_n - \theta \right)^2 \right\}$.

Proposition 5.1.1 Si pour tout θ de Θ , $\hat{\theta}_n$ admet une variance, alors on a : $r_{\theta}(\hat{\theta}_n) = \text{Var}_{\theta}(\hat{\theta}_n) + b_{\theta}^2(\hat{\theta}_n)$.

Preuve. On écrit : $\hat{\theta}_n - \theta = \hat{\theta}_n - E_{\theta} \{ \hat{\theta}_n \} + b_{\theta}(\hat{\theta}_n)$. Par linéarité de l'espérance, on a :

$$\begin{aligned} r_{\theta}(\hat{\theta}_n) &= E_{\theta} \left\{ \left(\hat{\theta}_n - E_{\theta} \{ \hat{\theta}_n \} \right)^2 \right\} + b_{\theta}^2(\hat{\theta}_n) \\ &= \text{Var}_{\theta}(\hat{\theta}_n) + b_{\theta}^2(\hat{\theta}_n), \end{aligned}$$

car $E_{\theta} \left\{ \hat{\theta}_n - E_{\theta} \{ \hat{\theta}_n \} \right\} = 0$. ■

Remarque 5.1.7 Le critère d'écart quadratique moyen n'est pas la panacée mais il est préféré parce qu'il s'exprime en fonction des notions simples de biais et de variance. D'autres critères peuvent paraître tout aussi naturels, en particulier l'écart absolue moyen $E_{\theta} \left\{ \left| \hat{\theta}_n - \theta \right| \right\}$, mais celle-ci est beaucoup plus difficile à manipuler analytiquement.

Remarque 5.1.8 Si $\hat{\theta}_n$ est un estimateur sans biais, i.e., si pour tout θ de Θ , $b_{\theta}(\hat{\theta}_n) = 0$ et s'il admet une variance, alors $r_{\theta}(\hat{\theta}_n) = \text{Var}_{\theta}(\hat{\theta}_n)$.

Entre deux estimateurs de θ , nous choisissons celui dont l'écart quadratique moyen ou le risque est le plus faible.

Définition 5.1.8 Un estimateur $\hat{\theta}_{1,n}$ est **relativement plus efficace** qu'un estimateur $\hat{\theta}_{2,n}$ s'il est plus précis que le second, i.e., $r_\theta(\hat{\theta}_{1,n}) \leq r_\theta(\hat{\theta}_{2,n})$.

Comment comparer deux estimateurs non biaisés ? Cette comparaison se fait sur la base de leur variance.

Propriété 5.1.1 Soient deux estimateurs sans biais $\hat{\theta}_{1,n}$ et $\hat{\theta}_{2,n}$. L'estimateur $\hat{\theta}_{1,n}$ domine l'estimateur $\hat{\theta}_{2,n}$, si $Var_\theta(\hat{\theta}_{1,n}) \leq Var_\theta(\hat{\theta}_{2,n})$.

Exemple 5.1.10 Soit un n -échantillon (X_1, X_2, \dots, X_n) tel que $E\{X_1\} = \mu$ et $Var(X_1) = \sigma^2$. Comparons les deux estimateurs $\hat{\theta}_{1,n} = \bar{X}_n$ et $\hat{\theta}_{2,n} = X_1$ de l'espérance μ . Tout d'abord, nous savons que ces deux estimateurs sont «sans biais». Par ailleurs : $Var_\theta(\hat{\theta}_{1,n}) = Var_\theta(\bar{X}_n) = \frac{\sigma^2}{n}$ et $Var_\theta(\hat{\theta}_{2,n}) = Var_\theta(X_1) = \sigma^2$. On obtient $Var_\theta(\hat{\theta}_{1,n}) \leq Var_\theta(\hat{\theta}_{2,n})$, dès lors que la taille d'échantillon n est supérieure ou égale à un, l'estimateur $\hat{\theta}_{1,n}$ est préféré à $\hat{\theta}_{2,n}$.

Remarque 5.1.9 Seuls des estimateurs non biaisés peuvent être comparés sur la base de leur variance.

Ainsi, nous savons comparer deux estimateurs non biaisés. Mais existe-t-il un estimateur sans biais qui soit plus efficace que tous les autres ? C'est la notion d'estimateur **optimal**.

Définition 5.1.9 Nous appelons **estimateur sans biais optimal** parmi les estimateurs sans biais, un estimateur $\hat{\theta}_n$ préférable à tout autre au sens de la variance, i.e., l'estimateur le plus efficace parmi tous les estimateurs sans biais.

La question qui se pose ici est de savoir comment se comporte l'estimateur $\hat{\theta}_n$ lorsque la taille d'échantillon n tend vers l'infini. Pourquoi étudier le comportement asymptotique de $\hat{\theta}_n$? Dans ce contexte, on cherche à caractériser le comportement de la variable aléatoire $\hat{\theta}_n$ dans le cas d'un échantillon de taille infinie à l'aide des différentes notions de convergence (en probabilité, presque sûre, en moyenne quadratique ou en loi, voir les chapitres 2 et 3). On étudie généralement

- la convergence de l'estimateur $\hat{\theta}_n$;
- la distribution asymptotique de $\hat{\theta}_n$, généralement établie à partir du théorème central limite (chapitre 3).

Définition 5.1.10 Un estimateur $\hat{\theta}_n$ est un estimateur convergent au sens faible s'il converge en probabilité vers θ quand n tend vers l'infini.

On parle d'estimateur fortement convergent lorsqu'on a convergence presque sûre.

Remarque 5.1.10 *Lorsqu'un estimateur est qualifié de convergent sans plus de précision (consistant en anglais), cela signifie qu'il est convergent au sens faible.*

Remarque 5.1.11 *Un estimateur convergent s'écarte donc du paramètre θ avec une probabilité faible, lorsque la taille de l'échantillon est assez grande (généralement $n > 30$).*

Condition suffisante de convergence d'un estimateur

Proposition 5.1.2 *Si pour tout θ de Θ , $\hat{\theta}_n$ admet une variance, si $\lim_n r_\theta(\hat{\theta}_n) = 0$ alors $\hat{\theta}_n$ est un estimateur convergent de θ .*

Preuve. Pour $\varepsilon > 0$, on a : $\{|\hat{\theta}_n - \theta| > \varepsilon\} = \{(\hat{\theta}_n - \theta)^2 > \varepsilon^2\}$. Comme $(\hat{\theta}_n - \theta)^2$ est une variable aléatoire positive admettant une espérance, on applique l'inégalité de Bienaymé-Chebyshev,

$$P\left((\hat{\theta}_n - \theta)^2 > \varepsilon^2\right) \leq \frac{E_\theta\left\{(\hat{\theta}_n - \theta)^2\right\}}{\varepsilon^2} = \frac{r_\theta(\hat{\theta}_n)}{\varepsilon^2} \longrightarrow 0,$$

d'où le résultat. ■

Exemple 5.1.11 *On a : la moyenne empirique \bar{X}_n est un estimateur sans biais de l'espérance μ et $\lim_n r_\theta(\bar{X}_n) = \lim_n \text{Var}_\theta(\bar{X}_n) = \lim_n \frac{\sigma^2}{n} = 0$, donc la moyenne empirique est un estimateur convergent de l'espérance¹.*

Exemple 5.1.12 *Soit un n -échantillon (X_1, X_2, \dots, X_n) de loi $\mathcal{N}(\mu, \sigma^2)$, on a : \widetilde{S}_n^2 est un estimateur sans biais de σ^2 et $\lim_n r_\theta(\widetilde{S}_n^2) = \lim_n \text{Var}_\theta(\widetilde{S}_n^2) = \lim_n \frac{2\sigma^4}{n-1} = 0$, donc la variance empirique corrigée est un estimateur convergent de la variance.*

Condition suffisante de convergence d'un estimateur asymptotiquement sans biais

Proposition 5.1.3 *Si l'estimateur $\hat{\theta}_n$ est sans biais (ou asymptotiquement sans biais) et s'il admet une variance, si $\lim_n \text{Var}_\theta(\hat{\theta}_n) = 0$ alors $\hat{\theta}_n$ est un estimateur convergent de θ .*

Preuve. D'après la décomposition biais-variance, on a : $r_\theta(\hat{\theta}_n) = \text{Var}_\theta(\hat{\theta}_n) + b_\theta^2(\hat{\theta}_n)$. Or, d'après les hypothèses, on a : $\lim_n b_\theta(\hat{\theta}_n) = 0$ et $\lim_n \text{Var}_\theta(\hat{\theta}_n) = 0$. Donc $\lim_n r_\theta(\hat{\theta}_n) = 0$. Par conséquent, d'après la condition suffisante de convergence d'un estimateur, $\hat{\theta}_n$ est un estimateur convergent de θ . ■

¹L'estimation d'une proportion p est un cas particulier du précédent, au sens où les v.a.r. X_k considérées sont de Bernoulli de paramètre p .

Exemple 5.1.13 Soit un n -échantillon (X_1, X_2, \dots, X_n) de loi $\mathcal{N}(\mu, \sigma^2)$ on a : $\overline{S_n^2}$ est un estimateur asymptotiquement sans biais de σ^2 et $\lim_n \text{Var}_\theta \left(\overline{S_n^2} \right) = \lim_n \text{Var}_\theta \left(\frac{n-1}{n} \widetilde{S_n^2} \right) = \lim_n \frac{2(n-1)\sigma^4}{n^2} = 0$, donc la variance empirique est un estimateur convergent de la variance.

Remarque 5.1.12 Une autre façon de démontrer la convergence en probabilité consiste à utiliser la loi faible des grands nombres. Dans le cadre d'un échantillon i.i.d., nous savons que la moyenne empirique \overline{X}_n des variables de l'échantillon converge vers l'espérance. Il suffit alors d'exprimer $\widehat{\theta}_n$ comme une fonction de cette moyenne empirique, i.e., sous la forme $\widehat{\theta}_n = h(\overline{X}_n)$. En utilisant le théorème de Slutsky, on en déduit la convergence en probabilité de $\widehat{\theta}_n$, et son éventuel caractère convergent ou non.

Dans de nombreux cas, les estimateurs que nous étudierons, convergent en distribution vers une loi normale.

Définition 5.1.11 Un estimateur $\widehat{\theta}_n$ est asymptotiquement normalement distribué dès lors que : $\sqrt{n}(\widehat{\theta}_n - \theta_0) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Sigma)$. Sa distribution asymptotique est définie par : $\widehat{\theta}_n \overset{asy}{\approx} \mathcal{N}\left(\theta_0, \frac{\Sigma}{n}\right)$.

Exemple 5.1.14 La moyenne empirique est un estimateur sans biais et convergent de μ (voir, l'exemple 5.1.11). D'après la loi forte des grands nombres, la moyenne empirique est même fortement convergente. Il est possible de déterminer la loi asymptotique de la moyenne empirique, si n est assez grand on peut utiliser l'approximation normale (lorsque X admet un moment d'ordre 2), $\overline{X}_n \overset{\mathcal{L}}{\sim} \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$. C'est une conséquence du TCL qui nous assure que $\sqrt{n}(\overline{X}_n - \mu) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2)$.

Exemple 5.1.15 On a S_n^2 est un estimateur sans biais de σ^2 et comme $\text{Var}(S_n^2) = \frac{\mu_4 - \sigma^4}{n} \rightarrow 0$, donc S_n^2 est un estimateur convergent. La loi forte des grands nombres appliquée aux variables $(X_k - \mu)^2$ entraîne même la convergence presque sûre vers σ^2 . Comme dans le cas de la moyenne empirique le TCL nous permet de déterminer la loi asymptotique de S_n^2 ; on a lorsque n est assez grand : $S_n^2 \overset{\mathcal{L}}{\sim} \mathcal{N}\left(\sigma^2, \frac{\mu_4 - \sigma^4}{n}\right)$.

Remarque 5.1.13 Jusqu'ici, nous avons considéré le cas où le paramètre à estimer, θ , était un scalaire. Nous allons à présent étendre les définitions précédentes au cas où $\underline{\theta}$ est un vecteur de m paramètres : $\underline{\theta} = (\theta_1, \theta_2, \dots, \theta_m)'$, où le symbole $'$ correspond à la transposée. À cet effet, on considère un estimateur $\widehat{\underline{\theta}}_n$, défini par un vecteur de dimension $m \times 1$, tel que : $\widehat{\underline{\theta}}_n = (\widehat{\theta}_{1,n}, \widehat{\theta}_{2,n}, \dots, \widehat{\theta}_{m,n})'$, et son espérance $E_{\underline{\theta}}\{\widehat{\underline{\theta}}_n\}$ est un vecteur de $m \times 1$ valeurs : $E_{\underline{\theta}}\{\widehat{\underline{\theta}}_n\} = \left(E_{\underline{\theta}}\{\widehat{\theta}_{1,n}\}, E_{\underline{\theta}}\{\widehat{\theta}_{2,n}\}, \dots, E_{\underline{\theta}}\{\widehat{\theta}_{m,n}\}\right)'$.

5.2 Méthodes d'estimation

On peut concevoir une méthode d'estimation comme une sorte de recette de cuisine qui permet d'obtenir un estimateur $\hat{\theta}_n$ à partir des ingrédients X_1, X_2, \dots, X_n . Plus formellement, on définit une méthode d'estimation de la façon suivante.

Définition 5.2.1 Une *méthode d'estimation* est une méthode mathématique qui permet de dériver la forme fonctionnelle d'un estimateur $\hat{\theta}_n = h(X_1, X_2, \dots, X_n)$ à partir des variables aléatoires de l'échantillon X_1, X_2, \dots, X_n .

Pour un même problème, on peut parfois appliquer plusieurs méthodes d'estimation. À chaque méthode d'estimation correspond un estimateur particulier. Si l'on se restreint aux seules méthodes d'estimation paramétriques, il existe de nombreuses méthodes suivant le problème étudié et les hypothèses retenues. Citons par exemple :

- la méthode des moindres carrés ordinaires ;
- la méthode des moindres carrés généralisés ;
- la méthode du maximum de vraisemblance ;
- la méthode des moments ;

Construction d'estimateurs

Nous abordons maintenant deux méthodes générales qui apportent des solutions dans des situations variées : l'approche par le maximum de vraisemblance et l'approche des moments. Nous commençons par celle du maximum de vraisemblance qui est la plus universelle (y compris pour des modèles complexes) pour deux raisons :

1. Elle est facile à mettre en oeuvre, se ramenant à un problème classique de résolution numérique.
2. Elle est optimale et même «efficace» asymptotiquement, i.e. quand la taille de l'échantillon tend vers l'infini. D'un point de vue pratique, pour un échantillon suffisamment grand (disons $n > 30$ pour fixer les idées), elle fournit des estimateurs de très bonne qualité.

5.2.1 Construction d'estimateur par la méthode du maximum de vraisemblance

Nous commencerons par présenter le concept de fonction de vraisemblance

Définition 5.2.2 On appelle **fonction de vraisemblance (likelihood)** de θ pour une réalisation donnée (x_1, \dots, x_n) de l'échantillon aléatoire (X_1, X_2, \dots, X_n) la loi de probabilité de ce n -uplet, notée $L_\theta(x_1, \dots, x_n)$, et définie par : si X est une v.a. discrète

$$\theta \in \Theta \longmapsto L_\theta(x_1, \dots, x_n) = P_\theta(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) = \prod_{k=1}^n P_\theta(X_k = x_k),$$

et si X est une v.a. continue de densité $f_{\theta, X}$ par

$$\theta \in \Theta \longmapsto L_\theta(x_1, \dots, x_n) = f_{\theta, (X_1, X_2, \dots, X_n)}(x_1, \dots, x_n) = \prod_{k=1}^n f_{\theta, X_k}(x_k).$$

Remarque 5.2.1 L'expression de la fonction de vraisemblance est donc la même que celle de la densité (ou fonction de probabilité) conjointe mais le point de vue est différent.

Définition 5.2.3 On dira que la valeur θ_1 est plus vraisemblable que la valeur θ_2 si $L_{\theta_1}(x_1, \dots, x_n) > L_{\theta_2}(x_1, \dots, x_n)$.

En ce sens il devient naturel de choisir pour θ la valeur la plus vraisemblable, disons $\hat{\theta}_{MV}$,

Définition 5.2.4 On appelle **estimation du maximum de vraisemblance** une valeur $\hat{\theta}_{MV}$, s'il en existe une, telle que :

$$L_{\hat{\theta}_{MV}}(x_1, \dots, x_n) = \max_{\theta \in \Theta} L_\theta(x_1, \dots, x_n),$$

ou le programme de maximisation suivant² :

$$\hat{\theta}_{MV} = \arg \max_{\theta \in \Theta} L_\theta(x_1, \dots, x_n).$$

Une telle solution est fonction de (x_1, \dots, x_n) , soit $\hat{\theta}_{MV} = h(x_1, \dots, x_n)$. Cette fonction h induit la statistique, $\tilde{\theta}_{MV} = h(X_1, X_2, \dots, X_n)$ appelée **estimateur du maximum de vraisemblance (EMV)**.

Remarque 5.2.2 $\tilde{\theta}_{MV}$ est une fonction de \mathbb{R}^n dans \mathbb{R}^r , associant à tout échantillon particulier une valeur particulière de θ .

Remarque 5.2.3 Généralement l'EMV existe et il est unique, i.e. quel que soit (x_1, \dots, x_n) il y a un et un seul maximum pour L_θ .

²Le terme $\arg \max$ signifie l'argument qui maximise. En effet, $\hat{\theta}_{MV}$ est défini comme l'argument de la fonction de vraisemblance qui maximise cette fonction.

Remarque 5.2.4 En pratique, la recherche de ce maximum se fait par dérivation de L_θ relativement à θ . La condition nécessaire de ce programme est la suivante : $\frac{\partial}{\partial \theta} L_\theta(x_1, \dots, x_n) \Big|_{\hat{\theta}_{MV}} = 0^3$. À présent, il convient de vérifier que l'on a bien un maximum. Pour ce faire, on considère la condition suffisante du programme de maximisation : $\frac{\partial^2}{\partial \theta^2} L_\theta(x_1, \dots, x_n) \Big|_{\hat{\theta}_{MV}} < 0$. Cette quantité étant négative, on a bien un maximum.

Remarque 5.2.5 Quand les densités (resp. fonctions de probabilité) conjointes sont des produits de fonctions puissances et exponentielles, ce qui est le cas la plupart du temps, on a plutôt intérêt à maximiser $\log L_\theta$, appelée \log -**vraisemblance**, ce qui est équivalent puisque la fonction logarithmique est strictement croissante. Dans les cas «réguliers» où L_θ est continûment dérivable et le support pour la famille de lois considérée est indépendant de θ , l'estimation par le maximum de vraisemblance (MV) vérifie (pour $\Theta \subset \mathbb{R}$) :

$$\begin{aligned} \frac{\partial}{\partial \theta} \log L_\theta &= 0 \text{ ou } \frac{\partial}{\partial \theta} \log \left(\prod_{k=1}^n f_{\theta, X_k}(x_k) \right) = 0 \text{ ou } \sum_{k=1}^n \frac{\partial}{\partial \theta} \log f_{\theta, X_k}(x_k) = 0, \\ (\text{resp. } \frac{\partial}{\partial \theta} \log L_\theta &= 0 \text{ ou } \frac{\partial}{\partial \theta} \log \left(\prod_{k=1}^n P_\theta(X_k = x_k) \right) = 0 \text{ ou } \sum_{k=1}^n \frac{\partial}{\partial \theta} \log P_\theta(X_k = x_k) = 0). \end{aligned}$$

Cette dernière égalité s'appelle l'**équation de \log -vraisemblance**. La résolution de cette équation en $\hat{\theta}_{MV}$ permet d'obtenir l'estimation du maximum de vraisemblance en fonction des réalisations de l'échantillon (données) (x_1, \dots, x_n) . De cette forme fonctionnelle, on déduira ensuite l'estimateur du maximum de vraisemblance. Mais avant cela, il convient de s'assurer que la solution $\hat{\theta}_{MV}$ est un maximum en vérifiant la condition suffisante du programme de maximisation : $\frac{\partial^2}{\partial \theta^2} \log L_\theta(x_1, \dots, x_n) \Big|_{\hat{\theta}_{MV}} < 0$. Dans le cas où θ possède k dimensions $(\theta_1, \theta_2, \dots, \theta_k)$, on résout un système de k équations obtenues en dérivant par rapport à chacune des composantes.

Exemple 5.2.1 On considère une variable aléatoire réelle, discrète X , supposée suivre une loi de Poisson $\mathcal{P}(\lambda)$. On rappelle que la loi de probabilité de X est définie par : $P_\lambda(X = x) = e^{-\lambda} \frac{\lambda^x}{x!}$. Soit un n -échantillon (X_1, \dots, X_n) où les variables X_k sont i.i.d. de même loi que X . Alors les fonctions de vraisemblance et de

³La condition nécessaire du programme de maximisation de la vraisemblance correspond à l'équation de vraisemblance.

log-vraisemblance associées à l'échantillon sont respectivement définies par :

$$L_\lambda(x_1, \dots, x_n) = \prod_{k=1}^n P_\lambda(X_k = x_k) = \prod_{k=1}^n e^{-\lambda} \frac{\lambda^{x_k}}{x_k!} = e^{-n\lambda} \frac{\lambda^{\sum_{k=1}^n x_k}}{\prod_{k=1}^n x_k!},$$

$$\log L_\lambda(x_1, \dots, x_n) = -n\lambda + \left(\sum_{k=1}^n x_k \right) \log \lambda - \sum_{k=1}^n \log(x_k!).$$

L'estimateur du maximum de vraisemblance $\hat{\lambda}$ est la solution du programme :

$$\hat{\lambda} = \arg \max_{\lambda \in \Theta} \log L_\lambda(x_1, \dots, x_n).$$

En dérivant par rapport à λ , on obtient l'équations de log-vraisemblance (la condition nécessaire du

programme de maximisation) : $-n + \frac{\sum_{k=1}^n x_k}{\lambda} = 0$, s'annule pour $\hat{\lambda} = \frac{1}{n} \sum_{k=1}^n x_k = \bar{x}_n$. On vérifie que cette so-

lution est un maximum (la condition suffisante du programme de maximisation) : $\left. \frac{\partial^2}{\partial \lambda^2} \log L_\lambda(x_1, \dots, x_n) \right|_{\hat{\lambda}} =$

$-\frac{\sum_{k=1}^n x_k}{\lambda^2} \Big|_{\hat{\lambda}} < 0$. Nous avons bien un maximum. Par conséquent, l'estimateur du maximum de vraisemblance

du paramètre λ est égal à \bar{X}_n . Sa réalisation (estimation du maximum de vraisemblance) est égale à : \bar{x}_n .

Exemple 5.2.2 On considère une variable aléatoire réelle, continue X , supposée suivre une loi $\mathcal{N}(\mu, \sigma^2)$.

On rappelle que la fonction de densité de X est définie par :

$$\forall x, f_{\theta, X}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{1}{2\sigma^2} (x - \mu)^2 \right\}.$$

Soit un n -échantillon (X_1, \dots, X_n) où les variables X_k sont i.i.d. de même loi que X . Si l'on définit un

vecteur de paramètres $\theta = (\mu, \sigma^2)'$, alors les fonctions de vraisemblance et de log-vraisemblance associées

à l'échantillon sont respectivement définies par :

$$L_\theta(x_1, \dots, x_n) = \prod_{k=1}^n f_{\theta, X_k}(x_k) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{k=1}^n (x_k - \mu)^2 \right\},$$

$$\log L_\theta(x_1, \dots, x_n) = -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log(\sigma^2) - \frac{1}{2\sigma^2} \sum_{k=1}^n (x_k - \mu)^2.$$

L'estimateur du maximum de vraisemblance $\hat{\theta}_{MV}$ est la solution du programme :

$$\hat{\theta}_{MV} = \arg \max_{\theta \in \Theta} \log L_\theta(x_1, \dots, x_n).$$

En dérivant par rapport à μ d'une part et par rapport à σ^2 d'autre part, on obtient le système d'équations de log-vraisemblance (la condition nécessaire du programme de maximisation) :

$$\begin{cases} \frac{1}{\sigma^2} \sum_{k=1}^n (x_k - \mu) = 0 \\ -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{k=1}^n (x_k - \mu)^2 = 0 \end{cases}$$

d'où la solution $\hat{\mu} = \frac{1}{n} \sum_{k=1}^n x_k = \bar{x}_n$ et $\hat{\sigma}^2 = \frac{1}{n} \sum_{k=1}^n (x_k - \hat{\mu})^2 = \overline{s_n^2}$. On vérifie que cette solution est un maximum (la condition suffisante du programme de maximisation) :

$$\begin{cases} \left. \frac{\partial^2}{\partial \mu^2} \log L_{\theta}(x_1, \dots, x_n) \right|_{\hat{\mu}} = \frac{-n}{\sigma^2} < 0 \\ \left. \frac{\partial^2}{\partial \sigma^4} \log L_{\theta}(x_1, \dots, x_n) \right|_{\hat{\sigma}^2} = \frac{n}{2\sigma^4} - \frac{1}{\sigma^6} \sum_{k=1}^n (x_k - \mu)^2 \Big|_{\hat{\sigma}^2} = \frac{n}{2\hat{\sigma}^4} - \frac{n\hat{\sigma}^2}{\hat{\sigma}^6} = -\frac{n}{2\hat{\sigma}^4} < 0 \end{cases}$$

Nous avons bien un maximum. Par conséquent, l'estimateur du maximum de vraisemblance du paramètre $(\mu, \sigma^2)'$ est donc $(\overline{X}_n, \overline{S_n^2})'$. Sa réalisation (estimation du maximum de vraisemblance) est égale à : $(\bar{x}_n, \overline{s_n^2})'$.

5.2.2 Construction d'estimateur par la méthode des moments

C'est la méthode la plus naturelle. L'idée de base est d'estimer une espérance mathématique par une moyenne empirique, une variance par une variance empirique, etc...

Définition 5.2.5 La méthode des moments consiste à poser l'égalité entre les moments théoriques de la variable aléatoire X et les moments de l'échantillon.

Explications : Si le paramètre à estimer est l'espérance de la loi des X_k , alors on peut l'estimer par la moyenne empirique de l'échantillon. Autrement dit, si $\theta = E_{\theta}\{X\}$, alors l'estimateur de θ par la méthode des moments (EMM) est $\tilde{\theta}_{MM} = \overline{X}_n$. De même, pour estimer le paramètre $\theta = Var_{\theta}(X)$, variance de la loi, nous retenons logiquement comme estimateur la variance empirique $\overline{S_n^2}$ (ou la variance empirique corrigée $\widetilde{S_n^2}$). En général, on pose : $E\{X^m\} = \frac{1}{n} \sum_{k=1}^n X_k^m$ pour $m \geq 1$ (le cas non-centré) (ou $E\{(X - \mu)^m\} = \frac{1}{n} \sum_{k=1}^n (X_k - \overline{X}_n)^m$ pour $m \geq 1$ (le cas centré)), La solution de l'équation retenue, si elle existe et unique, sera appelée estimateur obtenu par la méthode des moments.

Exemple 5.2.3 On considère une variable aléatoire réelle, discrète X , supposée suivre une loi de Bernoulli $\mathcal{B}(p)$. On rappelle que $E_p\{X\} = p$. Soit un n -échantillon (X_1, \dots, X_n) où les variables X_k sont i.i.d. de même loi que X . Alors l'estimateur de p par la méthode des moments est $\tilde{p}_n = \overline{X}_n$.

Exemple 5.2.4 On considère une variable aléatoire réelle, continue X , supposée suivre une loi exponentielle $\mathcal{E}(\lambda)$. On rappelle que $E_\lambda\{X\} = \frac{1}{\lambda}$. Soit un n -échantillon (X_1, \dots, X_n) où les variables X_k sont i.i.d. de même loi que X . Alors l'estimateur de λ par la méthode des moments est $\tilde{\lambda}_n = \frac{1}{\bar{X}_n}$.

Exemple 5.2.5 On considère une variable aléatoire réelle, continue X , supposée suivre une loi normale $\mathcal{N}(\mu, \sigma^2)$. On rappelle que $E_\mu\{X\} = \mu$ et $\text{Var}_{\sigma^2}(X) = \sigma^2$. Soit un n -échantillon (X_1, \dots, X_n) où les variables X_k sont i.i.d. de même loi que X . Alors les estimateurs de μ et σ^2 par la méthode des moments sont $\tilde{\mu}_n = \bar{X}_n$ et $\tilde{\sigma}_n^2 = \widetilde{S}_n^2$.

Exemple 5.2.6 On considère une variable aléatoire réelle, continue X , supposée suivre une loi uniforme $\mathcal{U}_{[0, \theta]}$. On rappelle que $E_\theta\{X\} = \frac{\theta}{2}$. Soit un n -échantillon (X_1, \dots, X_n) où les variables X_k sont i.i.d. de même loi que X . Alors l'estimateur de θ par la méthode des moments est $\tilde{\theta}_n = 2\bar{X}_n$.

Exemple 5.2.7 On considère une variable aléatoire réelle, continue X , supposée suivre une loi uniforme $\mathcal{U}_{[-\theta, \theta]}$. On rappelle que $E_\theta\{X\} = 0$. Soit un n -échantillon (X_1, \dots, X_n) où les variables X_k sont i.i.d. de même loi que X . Alors, on remarque que l'équation $0 = \bar{X}_n$ n'a pas de solution. Donc, dans ce cas, l'estimateur par la méthode des moments n'existe pas. En revanche, on peut essayer avec le moment d'ordre 2. On a : $E_\theta\{X^2\} = \frac{\theta^2}{3}$. Alors, l'estimateur de θ par la méthode des moments est $\tilde{\theta}_n = \sqrt{\frac{3}{n} \sum_{k=1}^n X_k^2}$.

5.3 Exercices

Exercice 5.3.1 Soient X_1, \dots, X_n n -variables aléatoires indépendantes, telles que $E\{X_k\} = \mu$ et $\text{Var}(X_k) = \sigma^2$ pour $k = 1, \dots, n$.

1. Montrer que \bar{X}_n est un estimateur sans biais pour μ .
2. Montrer que \bar{X}_n^2 n'est pas un estimateur sans biais de μ^2 . Déterminer son biais.
3. Déterminer m tel que $\bar{X}_n^2 - m\widetilde{S}_n^2$ soit un estimateur sans biais de μ^2 .

Exercice 5.3.2 Soient X_1, \dots, X_n n -variables aléatoires indépendantes et identiquement distribuées selon la loi Bernoulli $\mathcal{B}(p)$. On désire estimer $\text{Var}(X_1)$ c'est-à-dire $\theta = p(1-p)$. Pour ce faire, on propose de s'inspirer de l'estimateur $\hat{p}_n = \bar{X}_n$ de p et de considérer $\hat{\theta}_n = \bar{X}_n(1 - \bar{X}_n)$.

1. Calculer le biais de cet estimateur.

2. Comment proposez-vous de corriger ce biais ?

Exercice 5.3.3 Soit un n -échantillon (X_1, \dots, X_n) où les variables X_k sont i.i.d. de même loi que X . On dispose de deux estimateurs de $\mu = E\{X_1\}$,

$$T_n = \sum_{k=1}^2 \frac{X_k}{2} \text{ et } S_n = \sum_{k=1}^n \frac{X_k}{n} - \frac{X_1 - X_2}{2}.$$

Déterminer lequel des deux estimateurs est le plus efficace.

Exercice 5.3.4 Soit X une variable aléatoire réelle, continue, supposée suivre une loi exponentielle $\mathcal{E}(\lambda)$. Soit un n -échantillon (X_1, \dots, X_n) où les variables X_k sont i.i.d. de même loi que X . Comparer en écart quadratique moyen les estimateurs \bar{X}_n et $T_n = \sum_{k=1}^n \frac{X_k}{n+1}$ pour estimer $\frac{1}{\lambda}$.

Exercice 5.3.5 Soient X_1, \dots, X_n distribués selon une loi $\mathcal{N}(\mu, \sigma^2)$. On considère les deux estimateurs suivants de σ^2 (si μ inconnu) : \widetilde{S}_n^2 et \overline{S}_n^2 .

Comparer les deux estimateurs du point de vue de l'écart quadratique moyen.

Exercice 5.3.6 Soit T_1 et T_2 deux estimateurs sans biais et indépendants, d'un paramètre θ , de variances respectives σ_1^2, σ_2^2 , deux nombres réels strictement positifs. Soit a un réel appartenant à $[0; 1]$.

1. Montrer que la variable aléatoire $T = aT_1 + (1 - a)T_2$ est aussi un estimateur sans biais de θ .
2. Pour quelle valeur du paramètre a la variance de la variable aléatoire T est-elle minimale ?

Exercice 5.3.7 Soient X_1, \dots, X_n des variables aléatoires indépendantes qui sont distribuées uniformément sur l'intervalle $[0, \theta]$. Soient T_n et S_n t.q.

$$T_n = 2\bar{X}_n \text{ et } S_n = \frac{n+1}{n} \max\{X_1, \dots, X_n\}.$$

1. Trouver l'espérance et la variance de T_n .
2. Trouver l'espérance et la variance de S_n .
3. Comparer les deux estimateurs du point de vue de l'écart quadratique moyen.

Exercice 5.3.8 Soit X une variable aléatoire réelle, continue, positive et caractérisée par une fonction de densité $f_{\sigma^2, X}$ telle que :

$$\forall x \in \mathbb{R}^+, f_{\sigma^2, X}(x) = \frac{x}{\sigma^2} \exp\left(-\frac{x^2}{2\sigma^2}\right).$$

où σ^2 est un paramètre inconnu. Afin d'estimer ce paramètre, on dispose d'un n -échantillon (X_1, \dots, X_n) de variables i.i.d. de même loi que X . Déterminons l'estimateur du maximum de vraisemblance du paramètre σ^2 .

Exercice 5.3.9 Soit X une variable aléatoire réelle, continue, positive et caractérisée par une fonction de densité $f_{\theta, X}$ telle que :

$$\forall x, f_{\theta, X}(x) = \frac{1}{\theta} \mathbb{I}_{[0, \theta]}(x).$$

où $\theta > 0$ est un paramètre inconnu. Afin d'estimer ce paramètre, on dispose d'un n -échantillon (X_1, \dots, X_n) de variables i.i.d. de même loi que X . Déterminons l'estimateur du maximum de vraisemblance du paramètre θ .

Exercice 5.3.10 Soit X une variable aléatoire réelle, continue, positive et caractérisée par une fonction de densité $f_{\theta, X}$ telle que :

$$\forall x, f_{\theta, X}(x) = \theta(\theta + 1)x(1 - x)^{\theta - 1} \mathbb{I}_{[0, 1]}(x),$$

où $\theta > 0$ est un paramètre inconnu. Afin d'estimer ce paramètre, on dispose d'un n -échantillon (X_1, \dots, X_n) de variables i.i.d. de même loi que X . Déterminons l'estimateur du maximum de vraisemblance du paramètre θ .

Exercice 5.3.11 Soit (X_1, \dots, X_n) un échantillon d'une v.a. X à valeurs entières, de loi définie par :

$$P(X = k) = \frac{\theta^k}{(\theta + 1)^{k+1}}, k \geq 0,$$

où θ est un paramètre positif. Déterminer l'estimateur du maximum de vraisemblance de θ et étudier ses propriétés.

Exercice 5.3.12 Soit (X_1, \dots, X_n) un échantillon d'une v.a. X , de loi définie par :

$$\forall x, f_{\theta, X}(x) = (\theta - 1)x^{-\theta} \mathbb{I}_{[1, +\infty[}(x),$$

où θ est un paramètre inconnu. Déterminer l'estimateur du maximum de vraisemblance de θ et montrer qu'il est convergent.

Exercice 5.3.13 On suppose que X_1, \dots, X_n sont indépendants et identiquement distribués selon une loi $\mathcal{N}(\mu, \sigma^2)$. Pour μ connu, donner l'estimateur du maximum de vraisemblance de σ^2 .

Exercice 5.3.14 Soit X une variable aléatoire réelle, continue et caractérisée par une fonction de densité $f_{\theta, X}$ telle que :

$$\forall x \in \mathbb{R}, f_{\theta, X}(x) = \frac{1}{2} (1 + \theta x) \mathbf{I}_{[-1,1]}(x).$$

Soit (X_1, \dots, X_n) un échantillon d'une v.a. X .

1. Calculer l'estimateur des moments $\hat{\theta}_{MM}$ de θ .
2. Calculer le biais et la variance de $\hat{\theta}_{MM}$.

Exercice 5.3.15 Soient X_1, \dots, X_n des variables aléatoires indépendantes suivant une loi avec fonction de densité

$$\forall x \in \mathbb{R}, f_{\theta, X_1}(x) = (1 + \theta) x^\theta \mathbf{I}_{[0,1]}(x).$$

Trouver l'estimateur des moments de θ .

Exercice 5.3.16 On considère le modèle statistique de la loi Gamma $(\mathbb{R}^+, \mathcal{B}_{\mathbb{R}^+}, \Gamma(a, b); a, b > 0)$. On rappelle que la densité d'une v.a. X de loi $\Gamma(a, b)$ est :

$$\forall x, f_{a,b,X}(x) = \frac{b^a}{\Gamma(a)} x^{a-1} e^{-bx} \mathbf{I}_{\mathbb{R}^+}(x).$$

1. Calculer $E_{a,b}\{X\}$ et $\text{Var}_{a,b}(X)$.
2. Par la méthode des moments, donner un estimateur du paramètre bidimensionnel (a, b) .

Exercice 5.3.17 Si X_1, \dots, X_n sont issus d'une loi uniforme $\mathcal{U}_{[a,b]}$, calculer \hat{a}_{MM} et \hat{b}_{MM} , les estimateurs des moments de a et de b .

Exercice 5.3.18 On suppose qu'une v.a.r. X suit une loi exponentielle $\mathcal{E}(\lambda)$. Soit un n -échantillon (X_1, \dots, X_n) où les variables X_k sont i.i.d. de même loi que X .

1. Calculer $E_\lambda\{X^2\}$.
2. Par la méthode des moments, donner un estimateur du paramètre λ .

Exercice 5.3.19 Soit (X_1, \dots, X_n) un échantillon d'une v.a. X , de loi définie par :

$$\forall x, f_{\theta, X}(x) = \frac{2}{\theta} \left(1 - \frac{x}{\theta}\right) \mathbf{I}_{[0,\theta]}(x),$$

où θ est un paramètre strictement positif. Déterminer par la méthode des moments un estimateur du paramètre θ et étudier ses propriétés.

Exercice 5.3.20 On considère deux modèles :

- le modèle statistique de la loi de Poisson $(\mathbb{N}, \mathcal{P}(\mathbb{N}), \mathcal{P}(\lambda); \lambda > 0)$.
- le modèle statistique de la loi exponentielle $(\mathbb{R}^+, \mathcal{B}_{\mathbb{R}^+}, \mathcal{E}(\lambda); \lambda > 0)$.

Pour chacun de ces modèles, répondre à l'ensemble des questions suivantes. On considérera à chaque fois l'observation d'un n -échantillon (X_1, \dots, X_n) .

1. Rappeler l'expression de l'estimateur du maximum de vraisemblance dans ce modèle.
2. Étudier la consistance, le biais et le risque quadratique de cet estimateur.
3. Si cet estimateur est biaisé, est-il asymptotiquement sans biais ? Donner un estimateur sans biais.

Solutions

Solution 5.3.1 1. L'espérance de \bar{X}_n vaut $E\{\bar{X}_n\} = E\{X_1\} = \mu$. Donc l'estimateur \bar{X}_n est sans biais pour μ .

2. On calcule l'espérance de \bar{X}_n^2 ,

$$E\{\bar{X}_n^2\} = \text{Var}(\bar{X}_n) + (E\{\bar{X}_n\})^2 = \frac{\sigma^2}{n} + \mu^2 \text{ et } b_{\mu^2}(\bar{X}_n^2) = \frac{\sigma^2}{n}.$$

3. On veut obtenir un nouvel estimateur de μ^2 de la forme $\bar{X}_n^2 - m\widetilde{S}_n^2$. Déterminons m ,

$$E\{\bar{X}_n^2 - m\widetilde{S}_n^2\} = E\{\bar{X}_n^2\} - mE\{\widetilde{S}_n^2\} = \frac{\sigma^2}{n} + \mu^2 - m\sigma^2 = \left(\frac{1}{n} - m\right)\sigma^2 + \mu^2.$$

Pour que cet estimateur soit sans biais, il faut que $m = \frac{1}{n}$.

Solution 5.3.2 Soit l'estimateur $\hat{\theta}_n = \bar{X}_n(1 - \bar{X}_n)$.

1. Pour calculer son biais, on calcule d'abord son espérance

$$\begin{aligned} E\{\hat{\theta}_n\} &= E\{\bar{X}_n(1 - \bar{X}_n)\} = E\{\bar{X}_n\} - E\{\bar{X}_n^2\} \\ &= E\{\bar{X}_n\} - \left(\text{Var}(\bar{X}_n) + (E\{\bar{X}_n\})^2\right) \\ &= p - \left(\frac{p(1-p)}{n} + p^2\right) \\ &= \frac{n-1}{n}p(1-p) = \frac{n-1}{n}\theta. \end{aligned}$$

Donc son biais est

$$b_{\theta}(\hat{\theta}_n) = \frac{n-1}{n}\theta - \theta = \frac{-1}{n}\theta.$$

2. Pour corriger ce biais, on prendra l'estimateur $\tilde{\theta}_n = \frac{n}{n-1}\bar{X}_n(1 - \bar{X}_n)$.

Solution 5.3.3 L'estimateur le plus efficace est celui qui à l'écart quadratique moyen la plus petite. Commençons par calculer l'espérance des deux estimateurs pour en avoir le biais,

$$\begin{aligned} E\{T_n\} &= E\left\{\sum_{k=1}^2 \frac{X_k}{2}\right\} = \frac{1}{2} \sum_{k=1}^2 E\{X_k\} = \mu, \\ E\{S_n\} &= E\left\{\sum_{k=1}^n \frac{X_k}{n} - \frac{X_1 - X_2}{2}\right\} = \sum_{k=1}^n \frac{E\{X_k\}}{n} - \frac{E\{X_1\} - E\{X_2\}}{2} \\ &= \sum_{k=1}^n \frac{\mu}{n} - \frac{\mu - \mu}{2} = \mu. \end{aligned}$$

Les deux estimateurs sont des estimateurs sans biais de μ . Le plus efficace des deux sera, par conséquent, celui dont la variance est la plus petite. Le calcul des variances donne

$$\text{Var}(T_n) = \text{Var}\left(\sum_{k=1}^n \frac{X_k}{2}\right) = \frac{1}{4} \sum_{k=1}^n \text{Var}(X_k) = \frac{\sigma^2}{2},$$

où $\sigma^2 = \text{Var}(X_1)$. Dans le cas de $\text{Var}(S_n)$, on a

$$\begin{aligned} \text{Var}(S_n) &= \text{Var}\left(\frac{1}{n} \sum_{k=1}^n X_k - \frac{1}{2}X_1 + \frac{1}{2}X_2\right) \\ &= \frac{1}{n^2} \sum_{k=1}^n \text{Var}(X_k) + \frac{1}{4}(\text{Var}(X_1) + \text{Var}(X_2)) \\ &\quad - \frac{1}{n} \sum_{k=1}^n \text{Cov}(X_k, X_1) + \frac{1}{n} \sum_{k=1}^n \text{Cov}(X_k, X_2) \\ &= \frac{\sigma^2}{n} + \frac{\sigma^2}{2}. \end{aligned}$$

On en déduit que

$$\text{Var}(S_n) - \text{Var}(T_n) = \frac{\sigma^2}{n} + \frac{\sigma^2}{2} - \frac{\sigma^2}{2} = \frac{\sigma^2}{n} > 0.$$

La variance de S_n est donc supérieure à celle de T_n , ce qui implique que T_n est l'estimateur le plus efficace, malgré le fait qu'il n'utilise que l'information de X_1 et X_2 .

Solution 5.3.4 \bar{X}_n est sans biais pour $\frac{1}{\lambda}$ puisque $E_{1/\lambda}\{\bar{X}_n\} = \frac{1}{\lambda}$, $\text{Var}_{1/\lambda}(\bar{X}_n) = \frac{1}{n^2} \sum_{k=1}^n \text{Var}_{1/\lambda}(X_k) = \frac{1}{n^2} n \frac{1}{\lambda^2} = \frac{1}{n\lambda^2}$ et $r_{1/\lambda, \bar{X}_n}(1/\hat{\lambda}_n) = \frac{1}{n\lambda^2}$. On a : $T_n = \frac{n}{n+1}\bar{X}_n$, donc $E_{1/\lambda}\{T_n\} = \frac{n}{n+1}\frac{1}{\lambda}$ et T_n a un biais $b_{1/\lambda}(T_n) = E_{1/\lambda}\{T_n\} - \frac{1}{\lambda} = -\frac{1}{n+1}\frac{1}{\lambda}$, $\text{Var}_{1/\lambda}(T_n) = \frac{1}{(n+1)^2} \sum_{k=1}^n \text{Var}_{1/\lambda}(X_k) = \frac{n}{(n+1)^2} \frac{1}{\lambda^2}$. D'où

$$r_{1/\lambda}(T_n) = \left(\frac{-1}{n+1} \frac{1}{\lambda}\right)^2 + \frac{n}{(n+1)^2} \frac{1}{\lambda^2} = \frac{1}{n+1} \frac{1}{\lambda^2} < r_{1/\lambda}(\bar{X}_n).$$

En écart quadratique moyen T_n est meilleur, le gain de variance étant supérieur à la perte due au biais.

Solution 5.3.5 Soit Y une variable aléatoire distribuée selon une loi χ^2 à $(n-1)$ degrés de liberté. Son espérance est $(n-1)$ et sa variance $2(n-1)$. Calculons l'espérance de l'estimateur \widetilde{S}_n^2 ,

$$\begin{aligned} E\left\{\widetilde{S}_n^2\right\} &= E\left\{\frac{1}{n-1} \sum_{k=1}^n (X_k - \bar{X})^2\right\} = \frac{\sigma^2}{n-1} E\left\{\sum_{k=1}^n \frac{(X_k - \bar{X})^2}{\sigma^2}\right\} \\ &= \frac{\sigma^2}{n-1} E\{Y\} = \sigma^2, \end{aligned}$$

La variance de l'estimateur est

$$\begin{aligned} \text{Var} \left(\widetilde{S}_n^2 \right) &= \text{Var} \left(\frac{1}{n-1} \sum_{k=1}^n (X_k - \bar{X})^2 \right) = \frac{\sigma^4}{(n-1)^2} \text{Var} \left(\sum_{k=1}^n \frac{(X_k - \bar{X})^2}{\sigma^2} \right) \\ &= \frac{\sigma^4}{(n-1)^2} \text{Var}(Y) = \frac{2\sigma^4}{n-1}, \end{aligned}$$

et, par conséquent, son écart quadratique moyen est $r_{\sigma^2} \left(\widetilde{S}_n^2 \right) = \frac{2\sigma^4}{n-1}$. L'estimateur \overline{S}_n^2 s'écrit comme une fonction de \widetilde{S}_n^2 , $\overline{S}_n^2 = \frac{n-1}{n} \widetilde{S}_n^2$. On déduit que son espérance, son biais, sa variance et son écart quadratique moyen

$$\begin{aligned} E \left\{ \overline{S}_n^2 \right\} &= \frac{n-1}{n} \sigma^2, b_{\sigma^2} \left(\overline{S}_n^2 \right) = \frac{-1}{n} \sigma^2, \\ \text{Var} \left(\overline{S}_n^2 \right) &= \frac{2(n-1)\sigma^4}{n^2}, \\ r_{\sigma^2} \left(\overline{S}_n^2 \right) &= \frac{2(n-1)\sigma^4}{n^2} + \frac{\sigma^4}{n^2} = \frac{2n-1}{n^2} \sigma^4. \end{aligned}$$

Comparons les écarts quadratiques moyens

$$r_{\sigma^2} \left(\widetilde{S}_n^2 \right) - r_{\sigma^2} \left(\overline{S}_n^2 \right) = \frac{2\sigma^4}{n-1} - \frac{2n-1}{n^2} \sigma^4 = \frac{3n-1}{n^2(n-1)} \sigma^4.$$

Le dernier résultat est toujours positif car la taille de l'échantillon n est obligatoirement plus grande que 1. Cela implique que \overline{S}_n^2 est plus efficace que \widetilde{S}_n^2 du point de vue de l'écart quadratique moyen, et ceci malgré son biais.

Solution 5.3.6 1. T est une variable aléatoire réelle comme combinaison linéaire de deux variables aléatoires réelles T_1 et T_2 . Comme T_1 et T_2 sont des estimateurs sans biais de θ , T_1 et T_2 admettent une espérance. Nous avons, pour $a \in [0; 1]$,

$$\begin{aligned} E_{\theta} \{T\} &= E_{\theta} \{aT_1 + (1-a)T_2\} = aE_{\theta} \{T_1\} + (1-a)E_{\theta} \{T_2\} \\ &= a\theta + (1-a)\theta = \theta. \end{aligned}$$

Donc T est un estimateur sans biais de θ .

2. Calculons la variance de T . Comme T_1 et T_2 sont indépendants, nous avons

$$\begin{aligned} \text{Var}_{\theta}(T) &= \text{Var}_{\theta}(aT_1 + (1-a)T_2) = a^2 \text{Var}_{\theta}(T_1) + (1-a)^2 \text{Var}_{\theta}(T_2) \\ &= a^2 \sigma_1^2 + (1-a)^2 \sigma_2^2 \\ &= (\sigma_1^2 + \sigma_2^2) a^2 - 2a\sigma_2^2 + \sigma_2^2. \end{aligned}$$

Posons, pour $a \in [0; 1]$, $f(a) = (\sigma_1^2 + \sigma_2^2)a^2 - 2a\sigma_2^2 + \sigma_2^2$ est un polynôme du second degré dont le coefficient du terme de degré 2 est strictement positif. Pour tout $a \in [0; 1]$, nous avons $f'(a) = 2(\sigma_1^2 + \sigma_2^2)a - 2\sigma_2^2$. $f'(a_0) = 0$ équivaut à $a_0 = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2}$. Cette valeur a_0 appartient à l'intervalle $]0; 1[$ et par conséquent la variance de T est minimale pour $a = a_0 = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2}$.

Solution 5.3.7 Les variables aléatoires X_1, \dots, X_n suivent une loi uniforme $\mathcal{U}_{[0, \theta]}$. Leur espérance est $\frac{\theta}{2}$ et leur variance est $\frac{\theta^2}{12}$.

1. On a : $E_\theta \{T_n\} = 2E_\theta \{\bar{X}_n\} = 2E_\theta \{X_1\} = \theta$ et

$$\text{Var}_\theta(T_n) = 4\text{Var}_\theta(\bar{X}_n) = \frac{4}{n}\text{Var}_\theta(X_1) = \frac{4}{n} \frac{\theta^2}{12} = \frac{\theta^2}{3n}.$$

2. Déterminons la fonction de répartition de $U_n = \max\{X_1, \dots, X_n\}$,

$$\begin{aligned} \forall u, F_{U_n}(u) &= P(\max\{X_1, \dots, X_n\} \leq u) = P(X_1 \leq u, \dots, X_n \leq u) \\ &= (P(X_1 \leq u))^n = F_{X_1}^n(u) = \left(\frac{u}{\theta}\right)^n \mathbf{I}_{[0, \theta]}(u) + \mathbf{I}_{] \theta, +\infty[}. \end{aligned}$$

Ainsi, la fonction de densité de U_n est

$$\forall u, f_{U_n}(u) = \frac{n}{\theta^n} u^{n-1} \mathbf{I}_{[0, \theta]}(u).$$

Le calcul de l'espérance de U_n donne

$$E\{U_n\} = \int_0^\theta u \frac{n}{\theta^n} u^{n-1} du = \frac{n}{\theta^n} \int_0^\theta u^n du = \frac{n}{\theta^n} \frac{u^{n+1}}{n+1} \Big|_0^\theta = \frac{n}{n+1} \theta,$$

et implique $E\{S_n\} = \frac{n+1}{n} E\{U_n\} = \theta$. L'estimateur S_n de θ est sans biais. Nous cherchons à présent sa variance.

$$E\{U_n^2\} = \int_0^\theta u^2 \frac{n}{\theta^n} u^{n-1} du = \frac{n}{\theta^n} \int_0^\theta u^{n+1} du = \frac{n}{\theta^n} \frac{u^{n+2}}{n+2} \Big|_0^\theta = \frac{n}{n+2} \theta^2,$$

alors la variance de U_n est

$$\text{Var}(U_n) = E\{U_n^2\} - (E\{U_n\})^2 = \frac{n}{n+2} \theta^2 - \left(\frac{n}{n+1} \theta\right)^2 = \frac{n}{(n+2)(n+1)^2} \theta^2,$$

et celle de S_n ,

$$\text{Var}(S_n) = \left(\frac{n+1}{n}\right)^2 \text{Var}(U_n) = \frac{1}{n(n+2)} \theta^2.$$

3. Les deux estimateurs sont sans biais. Donc le plus efficace est celui qui a la variance la plus petite.

Comparons-les

$$\text{Var}(T_n) - \text{Var}(S_n) = \frac{1}{3n}\theta^2 - \frac{1}{n(n+2)}\theta^2 = \frac{n-1}{3n(n+2)}\theta^2.$$

Ce dernier résultat est toujours positif, n étant forcément plus grand que 1. Donc l'estimateur S_n est le plus efficace du point de vue de l'écart quadratique moyen.

Solution 5.3.8 Puisque les variables X_k sont indépendantes, alors les fonctions de vraisemblance et de log-vraisemblance associées à l'échantillon sont respectivement définies par :

$$\begin{aligned} L_{\sigma^2}(x_1, \dots, x_n) &= \prod_{k=1}^n f_{\sigma^2, X_k}(x_k) = \frac{1}{\sigma^{2n}} \left\{ \prod_{k=1}^n x_k \right\} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{k=1}^n x_k^2 \right\}, \\ \log L_{\sigma^2}(x_1, \dots, x_n) &= -n \log(\sigma^2) + \sum_{k=1}^n \log x_k - \frac{1}{2\sigma^2} \sum_{k=1}^n x_k^2. \end{aligned}$$

L'estimateur du maximum de vraisemblance $\hat{\sigma}_{MV}^2$ est la solution du programme :

$$\hat{\sigma}_{MV}^2 = \arg \max_{\sigma^2 \in \Theta} \log L_{\sigma^2}(x_1, \dots, x_n).$$

En dérivant par rapport à σ^2 , on obtient l'équations de log-vraisemblance (la condition nécessaire du programme de maximisation) : $-\frac{n}{\sigma^2} + \frac{1}{2\sigma^4} \sum_{k=1}^n x_k^2 = 0$. On en déduit que : $\hat{\sigma}_{MV}^2 = \frac{1}{2n} \sum_{k=1}^n x_k^2$. On vérifie que cette solution est un maximum (la condition suffisante du programme de maximisation) :

$$\left. \frac{\partial^2}{\partial \sigma^4} \log L_{\sigma^2}(x_1, \dots, x_n) \right|_{\hat{\sigma}_{MV}^2} = \left. \frac{n}{\sigma^4} - \frac{1}{\sigma^6} \sum_{k=1}^n x_k^2 \right|_{\hat{\sigma}_{MV}^2},$$

puisque $\sum_{k=1}^n x_k^2 = 2n\hat{\sigma}_{MV}^2$, cette expression peut se réécrire sous la forme suivante :

$$\left. \frac{\partial^2}{\partial \sigma^4} \log L_{\sigma^2}(x_1, \dots, x_n) \right|_{\hat{\sigma}_{MV}^2} = \frac{n}{\hat{\sigma}_{MV}^4} - \frac{2n\hat{\sigma}_{MV}^2}{\hat{\sigma}_{MV}^6} = -\frac{n}{\hat{\sigma}_{MV}^4} < 0.$$

Nous avons bien un maximum. Par conséquent, l'estimateur du maximum de vraisemblance du paramètre σ^2 est défini par : $\tilde{\sigma}_{MV}^2 = \frac{1}{2n} \sum_{k=1}^n X_k^2$. Sa réalisation (estimation du maximum de vraisemblance) est égale à : $\hat{\sigma}_{MV}^2 = \frac{1}{2n} \sum_{k=1}^n x_k^2$.

Solution 5.3.9 Puisque les variables X_k sont indépendantes, alors la fonction de vraisemblance associée à l'échantillon est définie par :

$$L_{\theta}(x_1, \dots, x_n) = \prod_{k=1}^n f_{\theta, X_k}(x_k) = \frac{1}{\theta^n} \mathbb{I}_{[0, +\infty[}(x_{(1)}) \mathbb{I}_{]-\infty, \theta]}(x_{(n)}),$$

où $x_{(1)} = \min_{k=1, \dots, n} (x_k)$, $x_{(n)} = \max_{k=1, \dots, n} (x_k)$. Elle contient $\theta \mapsto \frac{1}{\theta^n}$ qui est une fonction décroissante de θ , mais à partir du moment où $\mathbb{I}_{]-\infty, \theta]}(x_{(n)}) = 1$. i.e., $\theta \geq x_{(n)}$. Par conséquent, le maximum est atteint pour $\theta = x_{(n)}$, puisque pour $\theta < x_{(n)}$ la fonction de vraisemblance est nulle. Donc l'EMV est $X_{(n)} = \max(X_1, X_2, \dots, X_n)$.

Solution 5.3.10 Puisque les variables X_k sont indépendantes, alors les fonctions de vraisemblance et de log-vraisemblance associées à l'échantillon sont respectivement définies par :

$$L_\theta(x_1, \dots, x_n) = \prod_{k=1}^n f_{\theta, X_k}(x_k) = \theta^n (\theta + 1)^n \left\{ \prod_{k=1}^n x_k \right\} \left\{ \prod_{k=1}^n (1 - x_k)^{\theta-1} \right\} \mathbb{I}_{\{x_{(1)} \geq 0\}} \mathbb{I}_{\{x_{(n)} \leq 1\}},$$

$$\log L_\theta(x_1, \dots, x_n) = n \log \theta + n \log (\theta + 1) + \log \left(\prod_{k=1}^n x_k \right) + (\theta - 1) \sum_{k=1}^n \log (1 - x_k),$$

où $x_{(1)} = \min_{k=1, \dots, n} (x_k)$, $x_{(n)} = \max_{k=1, \dots, n} (x_k)$. L'estimateur du maximum de vraisemblance $\hat{\theta}_{MV}$ est la solution du programme :

$$\hat{\theta}_{MV} = \arg \max_{\theta \in \Theta} \log L_\theta(x_1, \dots, x_n).$$

En dérivant par rapport à θ , on obtient l'équations de log-vraisemblance (la condition nécessaire du programme de maximisation) : $\frac{n}{\theta} + \frac{n}{\theta+1} + \sum_{k=1}^n \log(1 - x_k) = 0$ ou $\frac{2\theta + 1}{\theta(\theta + 1)} = -\frac{1}{n} \sum_{k=1}^n \log(1 - x_k)$. Posons :

$c = -\frac{1}{n} \sum_{k=1}^n \log(1 - x_k)$. Il faut résoudre en θ l'équation : $c\theta^2 + (c - 2)\theta - 1 = 0$. Les solutions sont $\frac{2 - c - \sqrt{c^2 + 4}}{2c}$ et $\frac{2 + c + \sqrt{c^2 + 4}}{2c}$. Comme c ne prend que des valeurs positives, la seule solution dans

\mathbb{R}^+ est $\hat{\theta}_{MV} = \frac{2 + \frac{1}{n} \sum_{k=1}^n \log(1 - x_k) + \sqrt{\frac{1}{n^2} \left(\sum_{k=1}^n \log(1 - x_k) \right)^2 + 4}}{-\frac{2}{n} \sum_{k=1}^n \log(1 - x_k)}$. On vérifie que cette solution est un

maximum (la condition suffisante du programme de maximisation) :

$$\left. \frac{\partial^2}{\partial \theta^2} \log L_\theta(x_1, \dots, x_n) \right|_{\hat{\theta}_{MV}} = -\frac{n}{\theta^2} - \frac{n}{(\theta + 1)^2} \Big|_{\hat{\theta}_{MV}} < 0.$$

Nous avons bien un maximum. Par conséquent, l'estimateur du maximum de vraisemblance du paramètre θ est défini par :

$$\tilde{\theta}_{MV} = \frac{2 + \frac{1}{n} \sum_{k=1}^n \log(1 - X_k) + \sqrt{\frac{1}{n^2} \left(\sum_{k=1}^n \log(1 - X_k) \right)^2 + 4}}{-\frac{2}{n} \sum_{k=1}^n \log(1 - X_k)}.$$

Solution 5.3.11 Puisque les variables X_k sont indépendantes, alors les fonctions de vraisemblance et de log-vraisemblance associées à l'échantillon sont respectivement définies par :

$$L_\theta(x_1, \dots, x_n) = \prod_{k=1}^n P_\theta(X_k = x_k) = \frac{\theta^{\sum_{k=1}^n x_k}}{(\theta + 1)^{\sum_{k=1}^n (x_k + 1)}},$$

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

L'estimateur du maximum de vraisemblance $\hat{\theta}_{MV}$ est la solution du programme :

$$\hat{\theta}_{MV} = \arg \max_{\theta \in \Theta} \log L_\theta(x_1, \dots, x_n).$$

En dérivant par rapport à θ , on obtient l'équations de log-vraisemblance (la condition nécessaire du programme de maximisation) : $\frac{\sum_{k=1}^n x_k}{\theta} - \frac{n + \sum_{k=1}^n x_k}{\theta + 1} = 0$ ou $\frac{\theta + 1}{\theta} = \frac{n + \sum_{k=1}^n x_k}{\sum_{k=1}^n x_k}$. On en déduit que : $\hat{\theta}_{MV} = \frac{1}{n} \sum_{k=1}^n x_k$.

On vérifie que cette solution est un maximum (la condition suffisante du programme de maximisation) :

$$\left. \frac{\partial^2}{\partial \theta^2} \log L_\theta(x_1, \dots, x_n) \right|_{\hat{\theta}_{MV}} = - \left. \frac{\sum_{k=1}^n x_k}{\theta^2} + \frac{n + \sum_{k=1}^n x_k}{(\theta + 1)^2} \right|_{\hat{\theta}_{MV}} = - \left. \frac{n}{\hat{\theta}_{MV}} + \frac{n}{\hat{\theta}_{MV} + 1} \right| < 0.$$

Nous avons bien un maximum. Par conséquent, l'estimateur du maximum de vraisemblance du paramètre θ est défini par : $\tilde{\theta}_{MV} = \frac{1}{n} \sum_{k=1}^n X_k$.

Nous allons calculer l'espérance de $\tilde{\theta}_{MV}$,

$$\begin{aligned} E_\theta \left\{ \tilde{\theta}_{MV} \right\} &= E_\theta \{ X_1 \} = \sum_{k \geq 1} k \frac{\theta^k}{(\theta + 1)^{k+1}} = \frac{\theta}{(\theta + 1)^2} \sum_{k \geq 1} k \left(\frac{\theta}{\theta + 1} \right)^{k-1} \\ &= \frac{\theta}{(\theta + 1)^2} (\theta + 1)^2 = \theta, \end{aligned}$$

(parce que, la somme de la série géométrique $\sum_{k \geq 0} x^k = \frac{1}{1-x}$, si $|x| < 1$, et de sa dérivée $\sum_{k \geq 1} kx^{k-1} = \frac{1}{(1-x)^2}$. La valeur de la somme précédente est obtenue pour $x = \frac{\theta}{\theta+1}$). On sait que cet estimateur est sans biais, convergent d'après la loi des grands nombres.

Solution 5.3.12 Puisque les variables X_k sont indépendantes, alors les fonctions de vraisemblance et de log-vraisemblance associées à l'échantillon sont respectivement définies par :

$$L_\theta(x_1, \dots, x_n) = \prod_{k=1}^n f_{\theta, X_k}(x_k) = (\theta - 1)^n \left\{ \prod_{k=1}^n x_k \right\}^{-\theta} \mathbb{I}_{\{x_{(1)} \geq 1\}},$$

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

où $x_{(1)} = \min_{k=1, \dots, n}(x_k)$. L'estimateur du maximum de vraisemblance $\hat{\theta}_{MV}$ est la solution du programme :

$$\hat{\theta}_{MV} = \arg \max_{\theta \in \Theta} \log L_\theta(x_1, \dots, x_n).$$

En dérivant par rapport à θ , on obtient l'équation de log-vraisemblance (la condition nécessaire du programme de maximisation) : $\frac{n}{\theta - 1} - \sum_{k=1}^n \log x_k = 0$. On en déduit que : $\hat{\theta}_{MV} = 1 + \frac{n}{\sum_{k=1}^n \log x_k}$. On vérifie

que cette solution est un maximum (la condition suffisante du programme de maximisation) :

$$\frac{\partial^2}{\partial \theta^2} \log L_\theta(x_1, \dots, x_n) \Big|_{\hat{\theta}_{MV}} = - \frac{n}{(\theta - 1)^2} \Big|_{\hat{\theta}_{MV}} < 0.$$

Nous avons bien un maximum. Par conséquent, l'estimateur du maximum de vraisemblance du paramètre θ est défini par : $\tilde{\theta}_{MV} = 1 + \frac{n}{\sum_{k=1}^n \log X_k}$.

On peut appliquer la loi faible des grands nombres à $\left(\frac{1}{n} \sum_{k=1}^n \log X_k\right)$, nous allons calculer l'espérance de $\log X_1$:

$$E\{\log X_1\} = (\theta - 1) \int_1^{+\infty} x^{-\theta} \log x dx = \left[-x^{-\theta+1} \log x\right]_1^{+\infty} + \int_1^{+\infty} x^{-\theta} dx = \frac{1}{\theta - 1},$$

et $Var(\log X_1) = \frac{1}{(\theta - 1)^2}$. Par conséquent, la loi faible des grands nombres permet d'obtenir : $\frac{1}{n} \sum_{k=1}^n \log X_k \xrightarrow{p} E\{\log X_1\} = \frac{1}{\theta - 1}$. On en déduit, du théorème de Slutsky, la convergence de l'estimateur :

$$\tilde{\theta}_{MV} = 1 + \frac{n}{\sum_{k=1}^n \log X_k} \xrightarrow{p} 1 + (\theta - 1) = \theta.$$

Solution 5.3.13 Si μ connu alors l'estimateur du maximum de vraisemblance $\tilde{\sigma}_{MV}^2$ de σ^2 est S_n^2 (voir l'exemple 5.2.2).

Solution 5.3.14 1. L'espérance de la variable aléatoire X_1 est

$$\begin{aligned} E_\theta\{X_1\} &= \int_{-1}^1 x \frac{1}{2} (1 + \theta x) dx = \frac{1}{2} \int_{-1}^1 x (1 + \theta x) dx \\ &= \frac{1}{2} \left[\frac{x^2}{2} + \theta \frac{x^3}{3} \right]_{-1}^1 = \frac{\theta}{3}. \end{aligned}$$

Donc l'estimateur des moments $\hat{\theta}_{MM}$ de θ est $\hat{\theta}_{MM} = 3\bar{X}_n$.

2. C'est un estimateur sans biais

$$E\{\hat{\theta}_{MM}\} = 3E\{\bar{X}_n\} = 3\frac{\theta}{3} = \theta.$$

Pour trouver sa variance, il est nécessaire de calculer d'abord $E_{\theta}\{X_1^2\}$,

$$\begin{aligned} E_{\theta}\{X_1^2\} &= \int_{-1}^1 x^2 \frac{1}{2} (1 + \theta x) dx = \frac{1}{2} \int_{-1}^1 x^2 (1 + \theta x) dx \\ &= \frac{1}{2} \left[\frac{x^3}{3} + \theta \frac{x^4}{4} \right]_{-1}^1 = \frac{1}{3}, \end{aligned}$$

donc $Var_{\theta}(X_1) = E_{\theta}\{X_1^2\} - (E_{\theta}\{X_1\})^2 = \frac{1}{3} - \left(\frac{\theta}{3}\right)^2 = \frac{3 - \theta^2}{9}$. Alors la variance de l'estimateur $\hat{\theta}_{MM}$ est

$$Var_{\theta}(\hat{\theta}_{MM}) = \frac{9}{n} Var_{\theta}(X_1) = \frac{3 - \theta^2}{n}.$$

Solution 5.3.15 Pour trouver l'estimateur des moments $\hat{\theta}_{MM}$ de θ , calculons l'espérance de la variable aléatoire X_1 ,

$$\begin{aligned} E_{\theta}\{X_1\} &= \int_0^1 x (1 + \theta) x^{\theta} dx = (1 + \theta) \int_0^1 x^{\theta+1} dx \\ &= (1 + \theta) \left[\frac{x^{\theta+2}}{2 + \theta} \right]_0^1 = \frac{1 + \theta}{2 + \theta}. \end{aligned}$$

On en déduit l'estimateur des moments

$$\frac{1 + \hat{\theta}_{MM}}{2 + \hat{\theta}_{MM}} = \bar{X}_n \iff \hat{\theta}_{MM} = \frac{2\bar{X}_n - 1}{1 - \bar{X}_n}.$$

Solution 5.3.16 1. On a :

$$E_{a,b}\{X\} = \int_0^{+\infty} x \frac{b^a}{\Gamma(a)} x^{a-1} e^{-bx} dx = \frac{b^a}{\Gamma(a)} \int_0^{+\infty} x^a e^{-bx} dx.$$

En effectuant le changement de variable $y = bx$ et en notant que $\Gamma(a+1) = a\Gamma(a)$, il vient :

$$E_{a,b}\{X\} = \frac{1}{b\Gamma(a)} \int_0^{+\infty} y^a e^{-y} dy = \frac{\Gamma(a+1)}{b\Gamma(a)} = \frac{a}{b}.$$

De la même manière on montre que l'on a : $E_{a,b}\{X^2\} = \frac{\Gamma(a+2)}{b^2\Gamma(a)} = \frac{a(a+1)}{b^2}$. Ainsi, $Var_{a,b}(X) = E_{a,b}\{X^2\} - (E_{a,b}\{X\})^2 = \frac{a(a+1)}{b^2} - \left(\frac{a}{b}\right)^2 = \frac{a}{b^2}$.

2. On a vu que : $E_{a,b}\{X\} = \frac{a}{b}$ et $Var_{a,b}(X) = \frac{a}{b^2}$. Les estimateurs empiriques des moments d'ordre 1 et 2 de X nous donnent ainsi des fonctions des estimateurs de a et b sous la forme :

$$\frac{\tilde{a}_n}{\tilde{b}_n} = \bar{X}_n = \frac{1}{n} \sum_{k=1}^n X_k \quad \text{et} \quad \frac{\tilde{a}_n}{\tilde{b}_n^2} = \bar{S}_n^2 = \frac{1}{n} \sum_{k=1}^n (X_k - \bar{X}_n)^2.$$

De ces équations on tire

$$\tilde{b}_n = \frac{\bar{X}_n}{\bar{S}_n^2} = \frac{\sum_{k=1}^n X_k}{\sum_{k=1}^n (X_k - \bar{X}_n)^2}, \quad \tilde{a}_n = \tilde{b}_n \bar{X}_n = \frac{\left(\sum_{k=1}^n X_k\right)^2}{n \sum_{k=1}^n (X_k - \bar{X}_n)^2}.$$

Solution 5.3.17 L'espérance et la variance d'une variable aléatoire X_1 suivant une loi uniforme de paramètres a, b sont

$$E_{a,b}\{X_1\} = \frac{a+b}{2} \quad \text{et} \quad Var_{a,b}(X_1) = \frac{(b-a)^2}{12}.$$

Alors le 2^{ème} moment de X_1 est

$$E_{a,b}\{X_1^2\} = Var_{a,b}(X_1) + (E_{a,b}\{X_1\})^2 = \frac{a^2 + ab + b^2}{3}.$$

Les estimateurs des moments \hat{a}_{MM} et \hat{b}_{MM} de a et b sont les solutions du système

$$\begin{cases} \frac{\hat{a}_{MM} + \hat{b}_{MM}}{2} = \bar{X}_n \\ \frac{\hat{a}_{MM}^2 + \hat{a}_{MM}\hat{b}_{MM} + \hat{b}_{MM}^2}{3} = \frac{1}{n} \sum_{k=1}^n X_k^2 \end{cases}.$$

La 1^{er} équation donne $\hat{a}_{MM} = 2\bar{X}_n - \hat{b}_{MM}$, ce qui conduit par substitution dans la 2^{ème} équation $\hat{b}_{MM} = \bar{X}_n \pm \sqrt{\frac{3}{n} \sum_{k=1}^n X_k^2 - 3\bar{X}_n}$.

Solution 5.3.18 1. On a :

$$E_\lambda\{X^2\} = \int_0^{+\infty} x^2 \lambda e^{-\lambda x} dx = \frac{1}{\lambda^2} \int_0^{+\infty} y^2 e^{-y} dy = \frac{1}{\lambda^2} \Gamma(3) = \frac{2}{\lambda^2},$$

où la deuxième égalité est obtenue par changement de variable $y = \lambda x$.

2. On rappelle que $E_\lambda\{X^2\} = \frac{2}{\lambda^2}$. Alors, la méthode des moments nous donne, $\frac{2}{\tilde{\lambda}_n} = \frac{1}{n} \sum_{k=1}^n X_k^2$. Donc

l'estimateur de λ par la méthode des moments est $\tilde{\lambda}_n = \sqrt{\frac{2n}{\sum_{k=1}^n X_k^2}}$.

On a vu en cours que la méthode des moments permet d'obtenir un estimateur du paramètre λ dans un modèle de la loi exponentielle : $\tilde{\lambda}_n = \frac{1}{\bar{X}_n}$ basé sur la relation $E_\lambda \{X\} = \frac{1}{\lambda}$. L'intérêt de cet exercice est de montrer que cette méthode permet la construction de plusieurs estimateurs de ce même paramètre λ .

Solution 5.3.19 Le calcul de l'espérance donne :

$$E_\theta \{X_1\} = \frac{2}{\theta} \int_0^\theta x \left(1 - \frac{x}{\theta}\right) dx = \frac{2}{\theta} \left[\frac{x^2}{2} - \frac{x^3}{3\theta} \right]_0^\theta = \frac{\theta}{3}.$$

Donc, on cherche à résoudre l'équation $\bar{X}_n = \frac{\theta}{3}$, de solution immédiate $\theta = 3\bar{X}_n$, ce qui fournit l'estimateur de la méthode des moments $\tilde{\theta}_n = 3\bar{X}_n$, qui va être sans biais par construction :

$$E_\theta \left\{ \tilde{\theta}_n \right\} = 3E_\theta \left\{ \bar{X}_n \right\} = 3E_\theta \{X_1\} = \theta,$$

et convergent d'après la loi faible des grands nombres :

$$\tilde{\theta}_n = 3\bar{X}_n \xrightarrow{p} 3E_\theta \{X_1\} = \theta.$$

Solution 5.3.20 – **Modèle de Poisson**

1. On a vu que l'estimateur du maximum de vraisemblance est : $\tilde{\lambda}_n = \bar{X}_n$.
2. Par la loi forte des grands nombres, on a :

$$\tilde{\lambda}_n = \frac{1}{n} \sum_{k=1}^n X_k \xrightarrow{p.s.} E_\lambda \{X_1\} = \lambda,$$

donc $\tilde{\lambda}_n$ est un estimateur fortement consistant. De plus, on a : $E_\lambda \left\{ \tilde{\lambda}_n \right\} = E_\lambda \left\{ \bar{X}_n \right\} = \lambda$ et cet estimateur est également sans biais. Calculons maintenant son risque quadratique. Comme l'estimateur est sans biais, on a :

$$r_\lambda \left(\tilde{\lambda}_n \right) = \text{Var}_\lambda \left(\tilde{\lambda}_n \right) = \text{Var}_\lambda \left(\bar{X}_n \right) = \frac{1}{n} \text{Var}_\lambda \left(X_1 \right) = \frac{\lambda}{n}.$$

Ainsi, on a : $r_\lambda \left(\tilde{\lambda}_n \right) = E_\lambda \left\{ \left(\tilde{\lambda}_n - \lambda \right)^2 \right\} = \frac{\lambda}{n} \longrightarrow 0$, donc aussi $\tilde{\lambda}_n \xrightarrow{\text{L}_2} \lambda$.

– **Modèle de la loi exponentielle**

1. On a vu que l'estimateur du maximum de vraisemblance est : $\tilde{\lambda}_n = \frac{1}{\bar{X}_n} = \frac{n}{\sum_{k=1}^n X_k}$.

2. Par la loi forte des grands nombres, on a : $\bar{X}_n \xrightarrow{p.s.} E_\lambda \{X_1\} = \frac{1}{\lambda}$ ce qui implique que $\tilde{\lambda}_n \xrightarrow{p.s.} \lambda$, donc $\tilde{\lambda}_n$ est un estimateur fortement consistant.

Pour déterminer le biais de l'estimateur, rappelons que si X_1, \dots, X_n sont des v.a.r. indépendantes et de même loi exponentielle de paramètre λ , alors $\sum_{k=1}^n X_k$ est de loi Gamma(n, λ). Ainsi,

$$\begin{aligned} E_\lambda \left\{ \tilde{\lambda}_n \right\} &= E_\lambda \left\{ \frac{n}{\sum_{k=1}^n X_k} \right\} = \int_0^{+\infty} \frac{n}{x} \frac{\lambda^n}{\Gamma(n)} x^{n-1} e^{-\lambda x} dx = \frac{n\lambda^n}{\Gamma(n)} \int_0^{+\infty} x^{n-2} e^{-\lambda x} dx \\ &= \frac{n\lambda^n}{\Gamma(n) \lambda^{n-1}} \int_0^{+\infty} y^{n-2} e^{-y} dy = \frac{n\lambda \Gamma(n-1)}{\Gamma(n)} = \frac{n}{n-1} \lambda, \end{aligned}$$

où la quatrième égalité est obtenue par le changement de variable $y = \lambda x$. Donc l'estimateur $\tilde{\lambda}_n$ est un estimateur biaisé.

Évaluons maintenant l'écart quadratique moyen de notre estimateur. On a : $r_\lambda(\tilde{\lambda}_n) = Var_\lambda(\tilde{\lambda}_n) + b_\lambda^2(\tilde{\lambda}_n)$, où $b_\lambda(\tilde{\lambda}_n)$ est le biais de notre estimateur. Le calcul précédent nous donne

$$b_\lambda(\tilde{\lambda}_n) = E_\lambda \left\{ \tilde{\lambda}_n \right\} - \lambda = \frac{\lambda}{n-1}.$$

De plus, la variance de $\tilde{\lambda}_n$ est donnée par : $Var_\lambda(\tilde{\lambda}_n) = E_\lambda \left\{ \tilde{\lambda}_n^2 \right\} - \left(E_\lambda \left\{ \tilde{\lambda}_n \right\} \right)^2$. On a :

$$\begin{aligned} E_\lambda \left\{ \tilde{\lambda}_n^2 \right\} &= E_\lambda \left\{ \frac{n^2}{\left(\sum_{k=1}^n X_k \right)^2} \right\} = \int_0^{+\infty} \frac{n^2}{x^2} \frac{\lambda^n}{\Gamma(n)} x^{n-1} e^{-\lambda x} dx = \frac{n^2 \lambda^n}{\Gamma(n)} \int_0^{+\infty} x^{n-3} e^{-\lambda x} dx \\ &= \frac{n^2 \lambda^n}{\Gamma(n) \lambda^{n-2}} \int_0^{+\infty} y^{n-3} e^{-y} dy = \frac{n\lambda^2 \Gamma(n-2)}{\Gamma(n)} = \frac{n^2}{(n-1)(n-2)} \lambda^2, \end{aligned}$$

où la quatrième égalité est obtenue par le changement de variable $y = \lambda x$. D'où : $Var_\lambda(\tilde{\lambda}_n) = \frac{n^2}{(n-1)(n-2)} \lambda^2 - \left(\frac{n}{n-1} \lambda \right)^2$, et

$$\begin{aligned} r_\lambda(\tilde{\lambda}_n) &= \frac{n^2}{(n-1)(n-2)} \lambda^2 - \frac{n^2}{(n-1)^2} \lambda^2 + \left(\frac{\lambda}{n-1} \right)^2 \\ &= \frac{n^2 + n - 2}{(n-1)^2 (n-2)} \lambda^2. \end{aligned}$$

3. Nous avons vu que cet estimateur est biaisé. Il apparaît clairement asymptotiquement sans biais.

Un estimateur sans biais de λ est donné par : $\hat{\lambda}_n = \frac{n-1}{n} \tilde{\lambda}_n = \frac{n-1}{\sum_{k=1}^n X_k}$. Son risque quadratique

est donné par :

$$r_\lambda(\hat{\lambda}_n) = \text{Var}_\lambda(\hat{\lambda}_n) = \left(\frac{n-1}{n}\right)^2 \text{Var}_\lambda(\tilde{\lambda}_n) = \frac{\lambda^2}{n-2}.$$

Chapitre 6

Estimation par intervalle de confiance

Dans le chapitre précédent, l'objectif était de donner une valeur unique pour estimer le paramètre inconnu θ , et le principal inconvénient de l'estimation ponctuelle, c'est qu'elle ne rend pas compte de l'incertitude autour de l'estimation. Une façon de rendre compte de cette incertitude est de proposer un intervalle de confiance sur la valeur du paramètre θ . Dans ce chapitre, nous souhaitons donner un ensemble de valeurs plausibles pour θ essentiellement sous forme d'un intervalle.

Définition 6.0.1 On appelle *intervalle aléatoire* tout intervalle dont une extrémité au moins une variable aléatoire.

Exemple 6.0.1 Soit $X \sim \mathcal{N}(0, 1)$, $]-\infty, X]$ est un *intervalle aléatoire*. De même, $]X - 1, X + 2]$ est un *intervalle aléatoire*.

Définition 6.0.2 Un *intervalle de confiance* sur le paramètre θ pour un niveau de confiance de $(1 - \alpha)^1$ (ou un niveau de risque α), avec $\alpha \in]0, 1[$, est un encadrement du type :

$$P(A(X_1, \dots, X_n) \leq \theta \leq B(X_1, \dots, X_n)) = 1 - \alpha,$$

où $A(X_1, \dots, X_n)$ et $B(X_1, \dots, X_n)$ sont des variables aléatoires, fonctions des variables de l'échantillon (X_1, \dots, X_n) .

Une réalisation de cet intervalle de confiance est notée : $IC_{1-\alpha}(\theta) = [a, b]$ où a et b sont les réalisations respectives des variables $A(\cdot)$ et $B(\cdot)$, obtenues à partir de la réalisation de l'échantillon (x_1, \dots, x_n) .

¹ $(1 - \alpha)$ est le seuil de confiance ou la quasi-certitude.

Remarque 6.0.1 Un intervalle de confiance est un **intervalle aléatoire** car les bornes de cet intervalle sont des variables aléatoires.

Remarque 6.0.2 Un intervalle $[A(X_1, \dots, X_n), B(X_1, \dots, X_n)]$ est appelé une estimation par intervalle de confiance (**bilatéral**) du paramètre θ ou estimation ensembliste du paramètre θ . Des intervalles de confiance **unilatéraux** avec borne inférieure et borne supérieure sont donnés respectivement par $[A(X_1, \dots, X_n), +\infty[$ et $]-\infty, B(X_1, \dots, X_n)]^2$, où maintenant

$$P(A(X_1, \dots, X_n) \leq \theta) = P(\theta \leq B(X_1, \dots, X_n)) = 1 - \alpha.$$

Notons que, si l'on désire construire un intervalle de confiance avec borne supérieure pour une variance σ^2 , par exemple, alors l'intervalle sera plutôt de la forme $[0, B(X_1, \dots, X_n)]$.

Remarque 6.0.3 Les intervalles de confiance les plus courts sont souvent les intervalles de confiance bilatéraux symétriques³; i.e., ceux pour lesquels on a :

$$P(\theta < A(X_1, \dots, X_n)) = P(\theta > B(X_1, \dots, X_n)) = \frac{\alpha}{2}.$$

Remarque 6.0.4 $l = B(\cdot) - A(\cdot)$ s'appelle la longueur de l'intervalle de confiance. Si l est petit alors le meilleur intervalle de confiance est $[A(X_1, \dots, X_n), B(X_1, \dots, X_n)]$.

Remarque 6.0.5 Il convient d'éviter les notations du type : $P(a \leq \theta \leq b) = 1 - \alpha$. Par exemple, la notation $P(1,25 \leq \theta \leq 1,35) = 0,95$ n'a pas de sens, car θ n'est pas une variable aléatoire. Il n'y aucune raison d'utiliser la probabilité dans ce cas puisque le paramètre θ est supposé constant.

Remarque 6.0.6 Le seuil α étant donné. Les valeurs usuelles de α sont 1%, 5% ou 10%. Dans la pratique, on peut prendre par exemple $\alpha = 5\%$, ce qui nous donne un $IC_{0,95}(\theta)$ à 95%. Cela signifie qu'il y a 95% de chance que la valeur inconnue θ soit comprise entre $A(x_1, \dots, x_n)$ et $B(x_1, \dots, x_n)$.

Remarque 6.0.7 La probabilité $(1 - \alpha)$ représente le niveau de confiance de l'intervalle; ce niveau de confiance est associé à l'intervalle et non à la valeur inconnue du paramètre.

Remarque 6.0.8 Pour définir un intervalle de confiance, il faut connaître un estimateur ponctuel du paramètre ainsi que sa loi de distribution.

²La plupart du temps, un intervalle de confiance s'écrit sous la forme d'un intervalle (unilatère ou bilatère).

³D'une manière générale, on choisit un intervalle en divisant le risque d'erreur en 2 parties égales $\frac{\alpha}{2}$ et $\frac{\alpha}{2}$.

6.1 Construction d'un intervalle de confiance

Comment obtenir un intervalle de confiance ? Il n'existe pas de méthode générale, mais la procédure suivante peut être utilisée dans de nombreux cas :

- **Étape 1.** On considère un estimateur $\hat{\theta}_n$, sans biais et convergent, du paramètre θ . On cherche à caractériser soit (i) sa loi exacte, si cela est possible (ou celle d'une variable transformée), (ii) soit sa loi asymptotique. Cette loi dépend nécessairement de θ puisque l'estimateur est sans biais.
- **Étape 2.** On transforme la variable $\hat{\theta}_n$ de sorte à ce que la loi de la variable transformée ne dépende plus de θ , ni d'autres paramètres inconnus. Cette variable transformée dépend naturellement de θ (paramètre à estimer) et de $\hat{\theta}_n$, mais elle ne doit pas dépendre d'autres paramètres inconnus. Soit $m(\theta, \hat{\theta}_n)$ la variable transformée. On cherche à obtenir un résultat du type :

$$m(\theta, \hat{\theta}_n) \sim \text{loi connue (ne dépendant pas de paramètres inconnus)}.$$

En général, on utilise ici une R -transformée du type :

$$R = \frac{\hat{\theta}_n - E\{\hat{\theta}_n\}}{\sqrt{\text{Var}(\hat{\theta}_n)}} = \frac{\hat{\theta}_n - \theta}{\sqrt{\text{Var}(\hat{\theta}_n)}}.$$

Dans certains cas, cette variable transformée dépend d'autres paramètres inconnus que θ . Il faut alors chercher à les remplacer par leurs estimateurs.

- **Étape 3.** À partir de la loi de la variable aléatoire transformée $m(\theta, \hat{\theta}_n)$, on construit un encadrement du type :

$$P(a \leq m(\theta, \hat{\theta}_n) \leq b) = 1 - \alpha.$$

Pour cela, on cherche deux constantes réelles a et b , telles que $a < b$ et que la distance $b - a$ soit la plus petite possible. On peut obtenir les valeurs a et b de la façon suivante :

$$P(m(\theta, \hat{\theta}_n) < a) = \alpha_1 \text{ et } P(m(\theta, \hat{\theta}_n) > b) = \alpha_2,$$

où $\alpha_1 + \alpha_2 = \alpha$.

- **Étape 4.** En réaménageant les termes de cet encadrement, on cherche à construire un encadrement sur la valeur de θ , tel que :

$$P(m_1(\hat{\theta}_n, a, b) \leq \theta \leq m_2(\hat{\theta}_n, a, b)) = P(A(X_1, \dots, X_n) \leq \theta \leq B(X_1, \dots, X_n)) = 1 - \alpha,$$

où $m_1(\cdot)$ et $m_2(\cdot)$ sont des fonctions, $A(X_1, \dots, X_n) = m_1(\hat{\theta}_n, a, b)$ et $B(X_1, \dots, X_n) = m_2(\hat{\theta}_n, a, b)$ sont des variables aléatoires qui dépendent de l'estimateur $\hat{\theta}_n$ et donc implicitement des variables de l'échantillon (X_1, \dots, X_n) . Elles dépendent en outre des constantes a ou b , suivant les transformations effectuées.

- **Étape 5.** À partir de la réalisation de l'échantillon (x_1, \dots, x_n) et de l'estimation ponctuelle $\hat{\theta}_n(x_1, \dots, x_n)$, on déduit la réalisation de l'intervalle de confiance :

$$IC_{1-\alpha}(\theta) = \left[m_1 \left(\hat{\theta}_n(x_1, \dots, x_n) \right), m_2 \left(\hat{\theta}_n(x_1, \dots, x_n) \right) \right] = [a_1, b_1].$$

6.2 Exemples classiques d'estimation par intervalle

Sous le terme de classique nous présentons les cas de la moyenne et de la variance d'une loi mère gaussienne et le cas du paramètre p d'une loi de Bernoulli.

6.2.1 Intervalle de confiance pour les paramètres d'une loi normale

La variable aléatoire X suit une loi normale $\mathcal{N}(\mu, \sigma^2)$. Soit (X_1, \dots, X_n) un échantillon d'une v.a. X . Les paramètres à estimer sont la moyenne μ et la variance σ^2 .

Intervalle de confiance d'une moyenne

Pour la construction d'un intervalle de confiance de la moyenne, on distingue deux cas : celui où σ^2 est connu et celui où σ^2 est inconnu.

1. La variance σ^2 est connue.

L'estimateur sans biais et convergent de la moyenne μ est la statistique $\bar{X}_n = \frac{1}{n} \sum_{k=1}^n X_k$ qui suit la loi normale $\mathcal{N}(\mu, \frac{\sigma^2}{n})$. Construisons une variable transformée de \bar{X}_n dont la loi ne dépend pas de paramètres inconnus. Ici, il suffit d'utiliser la R -transformée⁴ :

$$R = \frac{\bar{X}_n - E\{\bar{X}_n\}}{\sqrt{Var(\bar{X}_n)}} = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1).$$

⁴On aurait pu utiliser le résultat $\sqrt{n}(\bar{X}_n - \mu) \sim \mathcal{N}(0, \sigma^2)$, puisque la variance σ^2 est connue.

Étant donné un seuil α , on construit un encadrement, pour la moyenne empirique \bar{X}_n de l'échantillon, du type :

$$P\left(a \leq \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \leq b\right) = 1 - \alpha,$$

où les constantes a et b sont telles que :

$$\begin{aligned} P\left(\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} < a\right) &= \Phi_R(a) = \frac{\alpha}{2} \iff a = \Phi_R^{-1}\left(\frac{\alpha}{2}\right), \\ P\left(\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} > b\right) &= 1 - P\left(\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \leq b\right) = \frac{\alpha}{2} \iff b = \Phi_R^{-1}\left(1 - \frac{\alpha}{2}\right), \end{aligned}$$

où Φ_R désigne la fonction de répartition de la loi normale centrée réduite. De l'encadrement précédent, on déduit que :

$$\begin{aligned} P\left(\Phi_R^{-1}\left(\frac{\alpha}{2}\right) \leq \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \leq \Phi_R^{-1}\left(1 - \frac{\alpha}{2}\right)\right) &= 1 - \alpha, \\ \iff P\left(\bar{X}_n - \frac{\sigma}{\sqrt{n}}\Phi_R^{-1}\left(1 - \frac{\alpha}{2}\right) \leq \mu \leq \bar{X}_n - \frac{\sigma}{\sqrt{n}}\Phi_R^{-1}\left(\frac{\alpha}{2}\right)\right) &= 1 - \alpha. \end{aligned}$$

La loi normale étant symétrique, $\Phi_R^{-1}\left(\frac{\alpha}{2}\right) = -\Phi_R^{-1}\left(1 - \frac{\alpha}{2}\right)$, on peut réécrire sous la forme :

$$P\left(\bar{X}_n - \frac{\sigma}{\sqrt{n}}\Phi_R^{-1}\left(1 - \frac{\alpha}{2}\right) \leq \mu \leq \bar{X}_n + \frac{\sigma}{\sqrt{n}}\Phi_R^{-1}\left(1 - \frac{\alpha}{2}\right)\right) = 1 - \alpha.$$

Par conséquent, l'intervalle $\left[\bar{X}_n - \frac{\sigma}{\sqrt{n}}\Phi_R^{-1}\left(1 - \frac{\alpha}{2}\right), \bar{X}_n + \frac{\sigma}{\sqrt{n}}\Phi_R^{-1}\left(1 - \frac{\alpha}{2}\right)\right]$ est un intervalle de confiance au niveau de confiance $1 - \alpha$ pour μ .

Exemple 6.2.1 Puisque $\Phi_R^{-1}(0.975) = 1.96$ l'intervalle de confiance de μ au niveau 95% est

$$\left[\bar{X}_n - 1.96\frac{\sigma}{\sqrt{n}}, \bar{X}_n + 1.96\frac{\sigma}{\sqrt{n}}\right].$$

Exemple 6.2.2 Pour un niveau de risque $\alpha = 5\%$, une taille d'échantillon $n = 9$, une variance $\sigma^2 = 12,96$ et une réalisation de la moyenne empirique $\bar{x}_n = 5.6$, on obtient une réalisation de l'intervalle de confiance égale à :

$$\begin{aligned} IC_{0.95}(\mu) &= \left[\bar{x}_n - \frac{\sigma}{\sqrt{n}}\Phi_R^{-1}\left(1 - \frac{\alpha}{2}\right), \bar{x}_n + \frac{\sigma}{\sqrt{n}}\Phi_R^{-1}\left(1 - \frac{\alpha}{2}\right)\right] \\ &= \left[5.6 - \frac{\sqrt{12,96}}{\sqrt{9}}\Phi_R^{-1}(0.975), 5.6 + \frac{\sqrt{12,96}}{\sqrt{9}}\Phi_R^{-1}(0.975)\right] \\ &= [3.248, 7.952]. \end{aligned}$$

Remarque 6.2.1 Les valeurs de $\Phi_R^{-1}\left(1 - \frac{\alpha}{2}\right)$ peuvent être obtenues à l'aide d'une calculatrice ou trouvées dans une table statistique. Les valeurs les plus utiles sont données dans le Tableau 6.1. De plus, par symétrie, on a : $\Phi_R^{-1}\left(\frac{\alpha}{2}\right) = -\Phi_R^{-1}\left(1 - \frac{\alpha}{2}\right)$.

$\alpha/2$	0.25	0.10	0.05	0.025	0.01	0.005	0.001	0.0005
$\Phi_R^{-1}\left(1 - \frac{\alpha}{2}\right)$	0,674	1,282	1,645	1,960	2,326	2,576	3,090	3,291

Tableau 6.1 : Valeurs de $\Phi_R^{-1}\left(1 - \frac{\alpha}{2}\right)$.

2. **La variance σ^2 est inconnue.**

L'estimateur sans biais et convergent de la moyenne μ est la statistique $\bar{X}_n = \frac{1}{n} \sum_{k=1}^n X_k$ qui suit la loi normale $\mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$. Construisons une variable transformée de \bar{X}_n dont la loi ne dépend pas de paramètres inconnus. Dans ce cas, on ne peut plus utiliser la R -transformée, car cette dernière dépend d'un paramètre inconnu, σ^2 : $\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1)$. Donc nous allons remplacer σ^2 par un estimateur convergent, à savoir la variance empirique corrigées \widetilde{S}_n^2 . Nous savons que dans un échantillon normal :

$$(n - 1) \frac{\widetilde{S}_n^2}{\sigma^2} \sim \chi_{(n-1)}^2.$$

Par ailleurs, on peut démontrer que les deux variables $\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}$ et $(n - 1) \frac{\widetilde{S}_n^2}{\sigma^2}$ sont indépendantes. Rappelons que si X et Y sont deux variables indépendantes telles que $X \sim \mathcal{N}(0, 1)$ et $Y \sim \chi_{(n)}^2$, alors la variable $R = \frac{X}{\sqrt{Y/n}}$ suit une distribution de Student à n degrés de liberté, notée $\mathcal{T}(n)$. Dans notre cas, nous pouvons définir une variable telle que :

$$R = \frac{\bar{X}_n - \mu}{\sqrt{\widetilde{S}_n^2/\sqrt{n}}} = \left(\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}\right) \frac{\sigma}{\sqrt{\widetilde{S}_n^2}} = \frac{\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}}{\sqrt{(n - 1) \frac{\widetilde{S}_n^2}{\sigma^2} / (n - 1)}} \sim \mathcal{T}(n - 1).$$

On observe que cette variable ne dépend pas de paramètres inconnus (hormis μ , le paramètre que l'on souhaite estimer) et sa loi ne dépend pas, elle aussi, de paramètres inconnus. On construit un encadrement du type :

$$P\left(a \leq \frac{\bar{X}_n - \mu}{\sqrt{\widetilde{S}_n^2/\sqrt{n}}} \leq b\right) = 1 - \alpha,$$

où les constantes a et b sont telles que :

$$P\left(\frac{\bar{X}_n - \mu}{\sqrt{\widetilde{S}_n^2}/\sqrt{n}} < a\right) = \Phi_R(a) = \frac{\alpha}{2} \iff a = \Phi_R^{-1}\left(\frac{\alpha}{2}\right),$$

$$P\left(\frac{\bar{X}_n - \mu}{\sqrt{\widetilde{S}_n^2}/\sqrt{n}} > b\right) = 1 - P\left(\frac{\bar{X}_n - \mu}{\sqrt{\widetilde{S}_n^2}/\sqrt{n}} \leq b\right) = \frac{\alpha}{2} \iff b = \Phi_R^{-1}\left(1 - \frac{\alpha}{2}\right),$$

où Φ_R désigne la fonction de répartition de la loi de Student à $n-1$ degrés de liberté. De l'encadrement précédent, on déduit que :

$$P\left(\Phi_R^{-1}\left(\frac{\alpha}{2}\right) \leq \frac{\bar{X}_n - \mu}{\sqrt{\widetilde{S}_n^2}/\sqrt{n}} \leq \Phi_R^{-1}\left(1 - \frac{\alpha}{2}\right)\right) = 1 - \alpha,$$

$$\iff P\left(\bar{X}_n - \frac{\sqrt{\widetilde{S}_n^2}}{\sqrt{n}}\Phi_R^{-1}\left(1 - \frac{\alpha}{2}\right) \leq \mu \leq \bar{X}_n - \frac{\sqrt{\widetilde{S}_n^2}}{\sqrt{n}}\Phi_R^{-1}\left(\frac{\alpha}{2}\right)\right) = 1 - \alpha.$$

La loi de Student étant symétrique, $\Phi_R^{-1}\left(\frac{\alpha}{2}\right) = -\Phi_R^{-1}\left(1 - \frac{\alpha}{2}\right)$, on peut réécrire sous la forme :

$$P\left(\bar{X}_n - \frac{\sqrt{\widetilde{S}_n^2}}{\sqrt{n}}\Phi_R^{-1}\left(1 - \frac{\alpha}{2}\right) \leq \mu \leq \bar{X}_n + \frac{\sqrt{\widetilde{S}_n^2}}{\sqrt{n}}\Phi_R^{-1}\left(1 - \frac{\alpha}{2}\right)\right) = 1 - \alpha.$$

Par conséquent, l'intervalle $\left[\bar{X}_n - \frac{\sqrt{\widetilde{S}_n^2}}{\sqrt{n}}\Phi_R^{-1}\left(1 - \frac{\alpha}{2}\right), \bar{X}_n + \frac{\sqrt{\widetilde{S}_n^2}}{\sqrt{n}}\Phi_R^{-1}\left(1 - \frac{\alpha}{2}\right)\right]$ est un intervalle de confiance au niveau de confiance $1 - \alpha$ pour μ .

Exemple 6.2.3 Pour un niveau de risque $\alpha = 5\%$, une taille d'échantillon $n = 10$, une réalisation de la variance empirique corrigée $\widetilde{s}_n^2 = 15,21$ et une réalisation de la moyenne empirique $\bar{x}_n = 5,6$, on obtient une réalisation de l'intervalle de confiance égale à :

$$\begin{aligned} IC_{0.95}(\mu) &= \left[\bar{x}_n - \frac{\sqrt{\widetilde{s}_n^2}}{\sqrt{n}}\Phi_R^{-1}\left(1 - \frac{\alpha}{2}\right), \bar{x}_n + \frac{\sqrt{\widetilde{s}_n^2}}{\sqrt{n}}\Phi_R^{-1}\left(1 - \frac{\alpha}{2}\right)\right] \\ &= \left[5,6 - \frac{\sqrt{15,21}}{\sqrt{10}}\Phi_R^{-1}(0,975), 5,6 + \frac{\sqrt{15,21}}{\sqrt{10}}\Phi_R^{-1}(0,975)\right] \\ &= \left[5,6 - \frac{\sqrt{15,21}}{\sqrt{10}}(2,2622), 5,6 + \frac{\sqrt{15,21}}{\sqrt{10}}(2,2622)\right] \\ &= [5,3059, 5,8940]. \end{aligned}$$

Remarque 6.2.2 Les valeurs de $\Phi_R^{-1}\left(1 - \frac{\alpha}{2}\right)$ peuvent être obtenues à l'aide d'une calculatrice ou trouvées dans une table statistique. Les valeurs les plus utiles sont données dans le Tableau 6.2. De plus, par symétrie, comme dans le cas de la distribution normale centrée réduite, on a : $\Phi_R^{-1}\left(\frac{\alpha}{2}\right) = -\Phi_R^{-1}\left(1 - \frac{\alpha}{2}\right)$.

$n - 1$	7	9	10	15	25
$\Phi_R^{-1}\left(1 - \frac{0.005}{2}\right)$	2.365	2.262	2.228	2.131	2.060
$\Phi_R^{-1}\left(1 - \frac{0.01}{2}\right)$	3.499	3.250	3.169	2.947	2.787
$\Phi_R^{-1}\left(1 - \frac{0.02}{2}\right)$	2.998	2.821	2.764	2.602	2.485
$\Phi_R^{-1}\left(1 - \frac{0.1}{2}\right)$	1.895	1.833	1.812	1.753	1.708
$\Phi_R^{-1}\left(1 - \frac{0.2}{2}\right)$	1.415	1.383	1.372	1.341	1.316

Tableau 6.2 : Valeurs de $\Phi_R^{-1}\left(1 - \frac{\alpha}{2}\right)$.

Remarque 6.2.3 Si les v.a.r. X_1, \dots, X_n ne sont pas gaussiennes mais que n est assez grand (en pratique supérieur à 30), alors le TCL nous garantit que la moyenne empirique suit approximativement la loi $\mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$. Ainsi, dans le cas où l'on souhaite estimer l'espérance lorsque la variance est connue, l'IC $_{1-\alpha}$ est identique à celui déterminé lorsque les v.a.r. X_1, \dots, X_n suivent la loi $\mathcal{N}(\mu, \sigma^2)$.

Intervalle de confiance d'une variance

Comme précédemment, pour la construction d'un intervalle de confiance de la variance, on distingue deux cas : celui où μ est connu et celui où μ est inconnu.

1. La moyenne μ est connue.

Le meilleur estimateur sans biais et convergent de la variance σ^2 est la statistique $S_n^2 = \frac{1}{n} \sum_{k=1}^n (X_k - \mu)^2$. Construisons une variable transformée de S_n^2 dont la loi ne dépend pas de paramètres inconnus. Ici, il suffit d'utiliser la R -transformée :

$$R = n \frac{S_n^2}{\sigma^2} = \sum_{k=1}^n \left(\frac{X_k - \mu}{\sigma} \right)^2 \sim \chi_{(n)}^2.$$

Étant donné un seuil α , on construit un encadrement, pour S_n^2 de l'échantillon, du type :

$$P\left(a \leq n \frac{S_n^2}{\sigma^2} \leq b\right) = 1 - \alpha,$$

où les constantes a et b sont telles que :

$$\begin{aligned} P(R < a) &= \Phi_R(a) = \alpha_1 \iff a = \Phi_R^{-1}(\alpha_1), \\ P(R > b) &= 1 - P(R \leq b) = \alpha_2 \iff b = \Phi_R^{-1}(1 - \alpha_2), \end{aligned}$$

où Φ_R désigne la fonction de répartition de la loi de χ^2 à n degrés de liberté et $\alpha_1 + \alpha_2 = \alpha$. De l'encadrement précédent, on déduit que :

$$\begin{aligned} P\left(\Phi_R^{-1}(\alpha_1) \leq n \frac{S_n^2}{\sigma^2} \leq \Phi_R^{-1}(1 - \alpha_2)\right) &= 1 - \alpha, \\ \iff P\left(\frac{nS_n^2}{\Phi_R^{-1}(1 - \alpha_2)} \leq \sigma^2 \leq \frac{nS_n^2}{\Phi_R^{-1}(\alpha_1)}\right) &= 1 - \alpha. \end{aligned}$$

Par conséquent, l'intervalle $\left[\frac{nS_n^2}{\Phi_R^{-1}(1 - \alpha_2)}, \frac{nS_n^2}{\Phi_R^{-1}(\alpha_1)}\right]$ est un intervalle de confiance au niveau de confiance $1 - \alpha$ pour σ^2 .

Exemple 6.2.4 Pour un niveau de risque $\alpha = 5\%$ ($\alpha_1 = \alpha_2 = \frac{\alpha}{2}$), une taille d'échantillon $n = 25$, une moyenne $\mu = 40$ et une réalisation de $s_n^2 = 12$, on obtient une réalisation de l'intervalle de confiance égale à :

$$\begin{aligned} IC_{0.95}(\sigma^2) &= \left[\frac{ns_n^2}{\Phi_R^{-1}\left(1 - \frac{\alpha}{2}\right)}, \frac{ns_n^2}{\Phi_R^{-1}\left(\frac{\alpha}{2}\right)}\right] \\ &= \left[\frac{25 \times 12}{40.644}, \frac{25 \times 12}{13.120}\right] \\ &= [7.381, 22.866]. \end{aligned}$$

2. La moyenne μ est inconnue.

Le meilleur estimateur sans biais et convergent de la variance σ^2 est la statistique $\widetilde{S}_n^2 = \frac{1}{n-1} \sum_{k=1}^n (X_k - \bar{X}_n)^2$.

Par conséquent :

$$R = (n-1) \frac{\widetilde{S}_n^2}{\sigma^2} = \sum_{k=1}^n \left(\frac{X_k - \bar{X}_n}{\sigma}\right)^2 \sim \chi_{(n-1)}^2.$$

Étant donné un seuil α , on construit un encadrement, pour \widetilde{S}_n^2 de l'échantillon, du type :

$$P\left(a \leq (n-1) \frac{\widetilde{S}_n^2}{\sigma^2} \leq b\right) = 1 - \alpha,$$

où les constantes a et b sont telles que :

$$\begin{aligned} P(R < a) &= \Phi_R(a) = \alpha_1 \iff a = \Phi_R^{-1}(\alpha_1), \\ P(R > b) &= 1 - P(R \leq b) = \alpha_2 \iff b = \Phi_R^{-1}(1 - \alpha_2), \end{aligned}$$

où Φ_R désigne la fonction de répartition de la loi de χ^2 à $n - 1$ degrés de liberté et $\alpha_1 + \alpha_2 = \alpha$. De l'encadrement précédent, on déduit que :

$$\begin{aligned} P\left(\Phi_R^{-1}(\alpha_1) \leq (n-1) \frac{\widetilde{S}_n^2}{\sigma^2} \leq \Phi_R^{-1}(1 - \alpha_2)\right) &= 1 - \alpha, \\ \iff P\left(\frac{(n-1) \widetilde{S}_n^2}{\Phi_R^{-1}(1 - \alpha_2)} \leq \sigma^2 \leq \frac{(n-1) \widetilde{S}_n^2}{\Phi_R^{-1}(\alpha_1)}\right) &= 1 - \alpha. \end{aligned}$$

Par conséquent, l'intervalle $\left[\frac{(n-1) \widetilde{S}_n^2}{\Phi_R^{-1}(1 - \alpha_2)}, \frac{(n-1) \widetilde{S}_n^2}{\Phi_R^{-1}(\alpha_1)}\right]$ est un intervalle de confiance au niveau de confiance $1 - \alpha$ pour σ^2 .

Exemple 6.2.5 Pour un niveau de risque $\alpha = 5\%$ ($\alpha_1 = \alpha_2 = \frac{\alpha}{2}$), une taille d'échantillon $n = 16$, une réalisation de $\widetilde{s}_n^2 = 15.246$, on obtient une réalisation de l'intervalle de confiance égale à :

$$\begin{aligned} IC_{0.95}(\sigma^2) &= \left[\frac{(n-1) \widetilde{s}_n^2}{\Phi_R^{-1}\left(1 - \frac{\alpha}{2}\right)}, \frac{(n-1) \widetilde{s}_n^2}{\Phi_R^{-1}\left(\frac{\alpha}{2}\right)}\right] \\ &= \left[\frac{15 \times 15.246}{27.488}, \frac{15 \times 15.246}{6.262}\right] \\ &= [8.319, 36.520]. \end{aligned}$$

6.2.2 Intervalle de confiance d'une proportion

Soit (X_1, \dots, X_n) un échantillon aléatoire d'une variable aléatoire X qui présente une distribution de Bernoulli de paramètre p ; i.e., $X \sim \mathcal{B}(1, p)$. On peut écrire que $\sum_{k=1}^n X_k \sim \mathcal{B}(n, p)$. Puisque $\mathcal{B}(n, p) \approx \mathcal{N}(np, np(1-p))$, si n est assez grand, on a : $\bar{X}_n \approx \mathcal{N}\left(p, \frac{p(1-p)}{n}\right)$ (suit approximativement la loi normale). Ainsi,

$$P\left(-a \leq \frac{\bar{X}_n - p}{\sqrt{\frac{p(1-p)}{n}}} \leq a\right) \simeq 1 - \alpha.$$

Maintenant, pour obtenir une formule pour un intervalle de confiance approximatif à $1-\alpha$ pour le paramètre inconnu p , on le remplace par son estimateur $\hat{p}_n = \bar{X}_n$ dans la racine carrée ci-dessus. Alors, on obtient l'intervalle suivant : $\left[\bar{X}_n - a\sqrt{\frac{\bar{X}_n(1-\bar{X}_n)}{n}}, \bar{X}_n + a\sqrt{\frac{\bar{X}_n(1-\bar{X}_n)}{n}} \right]$.

6.3 Comparaison de moyennes et de variances

Soient X et Y deux variables aléatoires indépendantes, suivant les lois normales $\mathcal{N}(\mu_1, \sigma_1^2)$ et $\mathcal{N}(\mu_2, \sigma_2^2)$, respectivement. On considère deux échantillons indépendants de taille n_1 pour la variable X , (X_1, \dots, X_{n_1}) , et de taille n_2 pour la variable Y , (Y_1, \dots, Y_{n_2}) . Nous souhaitons comparer les moyennes, μ_1 et μ_2 , et les variances, σ_1^2 et σ_2^2 , à l'aide de ces échantillons. Pour cela nous allons construire des intervalles de confiance pour $\mu_1 - \mu_2$ et pour $\frac{\sigma_2^2}{\sigma_1^2}$.

6.3.1 Intervalle de confiance de la différence de deux moyennes

Pour la construction d'un intervalle de confiance de la différence de deux moyennes, on distingue deux cas : celui où σ_1^2 et σ_2^2 sont connus et celui où σ_1^2 et σ_2^2 sont inconnus.

1. Si σ_1^2 et σ_2^2 sont connus.

L'estimateur sans biais et convergent de la moyenne μ_1 (resp. μ_2) est la statistique $\bar{X}_{n_1} = \frac{1}{n_1} \sum_{k=1}^{n_1} X_k$ (resp. $\bar{Y}_{n_2} = \frac{1}{n_2} \sum_{k=1}^{n_2} Y_k$) qui suit la loi normale $\mathcal{N}(\mu_1, \frac{\sigma_1^2}{n_1})$ (resp. $\mathcal{N}(\mu_2, \frac{\sigma_2^2}{n_2})$). Par conséquent, l'estimateur sans biais de $\mu_1 - \mu_2$ est la statistique $\bar{D}_{n_1, n_2} = \bar{X}_{n_1} - \bar{Y}_{n_2}$ qui suit la loi normale $\mathcal{N}(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2})$. Construisons une variable transformée de \bar{D}_{n_1, n_2} dont la loi ne dépend pas de paramètres inconnus. Ici, il suffit d'utiliser la R -transformée :

$$R = \frac{\bar{D}_{n_1, n_2} - E\{\bar{D}_{n_1, n_2}\}}{\sqrt{Var(\bar{D}_{n_1, n_2})}} = \frac{\bar{X}_{n_1} - \bar{Y}_{n_2} - \mu_1 + \mu_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim \mathcal{N}(0, 1).$$

Étant donné un seuil α , on construit un encadrement, pour \bar{D}_{n_1, n_2} , du type :

$$P \left(a \leq \frac{\bar{X}_{n_1} - \bar{Y}_{n_2} - \mu_1 + \mu_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \leq b \right) = 1 - \alpha,$$

où les constantes a et b sont telles que :

$$P \left(\frac{\bar{X}_{n_1} - \bar{Y}_{n_2} - \mu_1 + \mu_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} < a \right) = \Phi_R(a) = \frac{\alpha}{2} \iff a = \Phi_R^{-1} \left(\frac{\alpha}{2} \right),$$

$$P \left(\frac{\bar{X}_{n_1} - \bar{Y}_{n_2} - \mu_1 + \mu_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} > b \right) = 1 - P \left(\frac{\bar{X}_{n_1} - \bar{Y}_{n_2} - \mu_1 + \mu_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \leq b \right) = \frac{\alpha}{2} \iff b = \Phi_R^{-1} \left(1 - \frac{\alpha}{2} \right),$$

où Φ_R désigne la fonction de répartition de la loi normale centrée réduite. De l'encadrement précédent, on déduit que :

$$P \left(\Phi_R^{-1} \left(\frac{\alpha}{2} \right) \leq \frac{\bar{X}_{n_1} - \bar{Y}_{n_2} - \mu_1 + \mu_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \leq \Phi_R^{-1} \left(1 - \frac{\alpha}{2} \right) \right) = 1 - \alpha,$$

$$\iff P \left(\bar{X}_{n_1} - \bar{Y}_{n_2} - \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \Phi_R^{-1} \left(1 - \frac{\alpha}{2} \right) \leq \mu_1 - \mu_2 \leq \bar{X}_{n_1} - \bar{Y}_{n_2} - \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \Phi_R^{-1} \left(\frac{\alpha}{2} \right) \right) = 1 - \alpha.$$

La loi normale étant symétrique, $\Phi_R^{-1} \left(\frac{\alpha}{2} \right) = -\Phi_R^{-1} \left(1 - \frac{\alpha}{2} \right)$, on peut réécrire sous la forme :

$$P \left(\bar{X}_{n_1} - \bar{Y}_{n_2} - \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \Phi_R^{-1} \left(1 - \frac{\alpha}{2} \right) \leq \mu_1 - \mu_2 \leq \bar{X}_{n_1} - \bar{Y}_{n_2} + \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \Phi_R^{-1} \left(1 - \frac{\alpha}{2} \right) \right) = 1 - \alpha.$$

Par conséquent, l'intervalle

$$\left[\bar{X}_{n_1} - \bar{Y}_{n_2} - \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \Phi_R^{-1} \left(1 - \frac{\alpha}{2} \right), \bar{X}_{n_1} - \bar{Y}_{n_2} + \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \Phi_R^{-1} \left(1 - \frac{\alpha}{2} \right) \right]$$

est un intervalle de confiance au niveau de confiance $1 - \alpha$ pour $\mu_1 - \mu_2$.

Exemple 6.3.1 Pour un niveau de risque $\alpha = 5\%$, une taille d'échantillon 1, $n_1 = 5$, une variance $\sigma_1^2 = 12.90$ et une réalisation de la moyenne empirique $\bar{x}_{n_1} = 63.516$, et une taille d'échantillon 2, $n_2 = 4$, une variance $\sigma_2^2 = 5.68$ et une réalisation de la moyenne empirique $\bar{y}_{n_2} = 62.567$, on obtient

une réalisation de l'intervalle de confiance égale à :

$$\begin{aligned}
 IC_{0.95}(\mu_1 - \mu_2) &= \left[\bar{x}_{n_1} - \bar{y}_{n_2} - \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \Phi_R^{-1}\left(1 - \frac{\alpha}{2}\right), \bar{x}_{n_1} - \bar{y}_{n_2} + \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \Phi_R^{-1}\left(1 - \frac{\alpha}{2}\right) \right] \\
 &= \left[63.516 - 62.567 - \sqrt{\frac{12.90}{5} + \frac{5.68}{4}} \Phi_R^{-1}(0.975), \right. \\
 &\quad \left. 63.516 - 62.567 + \sqrt{\frac{12.90}{5} + \frac{5.68}{4}} \Phi_R^{-1}(0.975) \right] \\
 &= [-2.971, 4.869].
 \end{aligned}$$

2. Si σ_1^2 et σ_2^2 sont inconnus, mais $\sigma_1^2 = \sigma_2^2 = \sigma^2$ (la variance commune).

L'estimateur sans biais de $\mu_1 - \mu_2$ est la statistique \bar{D}_{n_1, n_2} qui suit la loi normale $\mathcal{N}(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2})$.

Construisons une variable transformée de \bar{D}_{n_1, n_2} dont la loi ne dépend pas de paramètres inconnus.

Dans ce cas, on ne peut plus utiliser la R -transformée, car cette dernière dépend des paramètres inconnus, σ_1^2 et σ_2^2 : $\frac{\bar{X}_{n_1} - \bar{Y}_{n_2} - \mu_1 + \mu_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim \mathcal{N}(0, 1)$. Donc nous allons remplacer σ_1^2 (resp. σ_2^2) par

un estimateur convergent, à savoir la variance empirique corrigées $\widetilde{S}_{n_1}^2$ (resp. $\widetilde{S}_{n_2}^2$). Nous savons que dans un échantillon normal :

$$(n_1 - 1) \frac{\widetilde{S}_{n_1}^2}{\sigma_1^2} \sim \chi_{(n_1-1)}^2 \quad (\text{resp.} \quad (n_2 - 1) \frac{\widetilde{S}_{n_2}^2}{\sigma_2^2} \sim \chi_{(n_2-1)}^2).$$

D'autre part, l'indépendance des deux échantillons entraîne que $(n_1 - 1) \frac{\widetilde{S}_{n_1}^2}{\sigma_1^2} + (n_2 - 1) \frac{\widetilde{S}_{n_2}^2}{\sigma_2^2} \sim \chi_{(n_1+n_2-2)}^2$,

et comme $E \left\{ (n_1 - 1) \frac{\widetilde{S}_{n_1}^2}{\sigma_1^2} + (n_2 - 1) \frac{\widetilde{S}_{n_2}^2}{\sigma_2^2} \right\} = n_1 + n_2 - 2$, donc la statistique

$$\widetilde{S}_{n_1, n_2}^2 = \frac{(n_1 - 1) \widetilde{S}_{n_1}^2 + (n_2 - 1) \widetilde{S}_{n_2}^2}{n_1 + n_2 - 2}$$

est un estimateur sans biais de σ^2 . On va remplacer la variance de \bar{D}_{n_1, n_2} par $\frac{\widetilde{S}_{n_1, n_2}^2}{n_1} + \frac{\widetilde{S}_{n_1, n_2}^2}{n_2}$. Dans

notre cas, nous pouvons définir une variable telle que :

$$\begin{aligned}
 R &= \frac{\bar{X}_{n_1} - \bar{Y}_{n_2} - \mu_1 + \mu_2}{\sqrt{\frac{\widetilde{S}_{n_1, n_2}^2}{n_1} + \frac{\widetilde{S}_{n_1, n_2}^2}{n_2}}} = \left(\frac{\bar{X}_{n_1} - \bar{Y}_{n_2} - \mu_1 + \mu_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \right) \frac{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}{\sqrt{\frac{\widetilde{S}_{n_1, n_2}^2}{n_1} + \frac{\widetilde{S}_{n_1, n_2}^2}{n_2}}} \\
 &= \frac{\bar{X}_{n_1} - \bar{Y}_{n_2} - \mu_1 + \mu_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim T(n_1 + n_2 - 2). \\
 &= \frac{\bar{X}_{n_1} - \bar{Y}_{n_2} - \mu_1 + \mu_2}{\sqrt{(n_1 + n_2 - 2) \frac{\widetilde{S}_{n_1, n_2}^2}{\sigma^2} / (n_1 + n_2 - 2)}}
 \end{aligned}$$

On observe que cette variable ne dépend pas de paramètres inconnus (hormis $\mu_1 - \mu_2$, le paramètre que l'on souhaite estimer) et sa loi ne dépend pas, elle aussi, de paramètres inconnus. On construit un encadrement du type :

$$P \left(a \leq \frac{\bar{X}_{n_1} - \bar{Y}_{n_2} - \mu_1 + \mu_2}{\sqrt{\widetilde{S}_{n_1, n_2}^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} \leq b \right) = 1 - \alpha,$$

où les constantes a et b sont telles que :

$$\begin{aligned}
 P \left(\frac{\bar{X}_{n_1} - \bar{Y}_{n_2} - \mu_1 + \mu_2}{\sqrt{\widetilde{S}_{n_1, n_2}^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} < a \right) &= \Phi_R(a) = \frac{\alpha}{2} \iff a = \Phi_R^{-1} \left(\frac{\alpha}{2} \right), \\
 P \left(\frac{\bar{X}_{n_1} - \bar{Y}_{n_2} - \mu_1 + \mu_2}{\sqrt{\widetilde{S}_{n_1, n_2}^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} > b \right) &= 1 - P \left(\frac{\bar{X}_{n_1} - \bar{Y}_{n_2} - \mu_1 + \mu_2}{\sqrt{\widetilde{S}_{n_1, n_2}^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} \leq b \right) = \frac{\alpha}{2} \iff b = \Phi_R^{-1} \left(1 - \frac{\alpha}{2} \right),
 \end{aligned}$$

où Φ_R désigne la fonction de répartition de la loi de Student à $n_1 + n_2 - 2$ degrés de liberté. De l'encadrement précédent, on déduit que :

$$P \left(\Phi_R^{-1} \left(\frac{\alpha}{2} \right) \leq \frac{\bar{X}_{n_1} - \bar{Y}_{n_2} - \mu_1 + \mu_2}{\sqrt{\widetilde{S}_{n_1, n_2}^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} \leq \Phi_R^{-1} \left(1 - \frac{\alpha}{2} \right) \right) = 1 - \alpha$$

$$\begin{aligned} \Leftrightarrow P \left(\bar{X}_{n_1} - \bar{Y}_{n_2} - \sqrt{\widetilde{S_{n_1, n_2}^2} \left(\frac{1}{n_1} + \frac{1}{n_2} \right)} \Phi_R^{-1} \left(1 - \frac{\alpha}{2} \right) \leq \mu_1 - \mu_2 \leq \right. \\ \left. \bar{X}_{n_1} - \bar{Y}_{n_2} + \sqrt{\widetilde{S_{n_1, n_2}^2} \left(\frac{1}{n_1} + \frac{1}{n_2} \right)} \Phi_R^{-1} \left(\frac{\alpha}{2} \right) \right) = 1 - \alpha. \end{aligned}$$

La loi de Student étant symétrique, $\Phi_R^{-1} \left(\frac{\alpha}{2} \right) = -\Phi_R^{-1} \left(1 - \frac{\alpha}{2} \right)$, on peut réécrire sous la forme :

$$\begin{aligned} P \left(\bar{X}_{n_1} - \bar{Y}_{n_2} - \sqrt{\widetilde{S_{n_1, n_2}^2} \left(\frac{1}{n_1} + \frac{1}{n_2} \right)} \Phi_R^{-1} \left(1 - \frac{\alpha}{2} \right) \leq \mu_1 - \mu_2 \leq \right. \\ \left. \bar{X}_{n_1} - \bar{Y}_{n_2} + \sqrt{\widetilde{S_{n_1, n_2}^2} \left(\frac{1}{n_1} + \frac{1}{n_2} \right)} \Phi_R^{-1} \left(1 - \frac{\alpha}{2} \right) \right) = 1 - \alpha. \end{aligned}$$

Par conséquent, l'intervalle

$$\left[\bar{X}_{n_1} - \bar{Y}_{n_2} - \sqrt{\widetilde{S_{n_1, n_2}^2} \left(\frac{1}{n_1} + \frac{1}{n_2} \right)} \Phi_R^{-1} \left(1 - \frac{\alpha}{2} \right), \bar{X}_{n_1} - \bar{Y}_{n_2} + \sqrt{\widetilde{S_{n_1, n_2}^2} \left(\frac{1}{n_1} + \frac{1}{n_2} \right)} \Phi_R^{-1} \left(1 - \frac{\alpha}{2} \right) \right]$$

est un intervalle de confiance au niveau de confiance $1 - \alpha$ pour $\mu_1 - \mu_2$.

Exemple 6.3.2 Pour un niveau de risque $\alpha = 5\%$, une taille d'échantillon $n_1 = 5$ (resp. $n_2 = 4$), une réalisation de la variance empirique corrigée $\widetilde{s_{n_1}^2} = 0,3409$ (resp. $\widetilde{s_{n_2}^2} = 0,2166$) et une réalisation de la moyenne empirique $\bar{x}_{n_1} = 63.516$ (resp. $\bar{y}_{n_2} = 62.567$), on obtient une réalisation de l'intervalle de confiance égale à :

$$\begin{aligned} IC_{0.95}(\mu_1 - \mu_2) &= \left[\bar{x}_{n_1} - \bar{y}_{n_2} - \sqrt{\widetilde{s_{n_1, n_2}^2} \left(\frac{1}{n_1} + \frac{1}{n_2} \right)} \Phi_R^{-1} \left(1 - \frac{\alpha}{2} \right) \right. \\ &\quad \left. \bar{x}_{n_1} - \bar{y}_{n_2} + \sqrt{\widetilde{s_{n_1, n_2}^2} \left(\frac{1}{n_1} + \frac{1}{n_2} \right)} \Phi_R^{-1} \left(1 - \frac{\alpha}{2} \right) \right] \\ &= \left[63.516 - 62.567 - \sqrt{\frac{4 \times 0,3409 + 3 \times 0,2166}{7} \left(\frac{1}{5} + \frac{1}{4} \right)} \Phi_R^{-1}(0.975), \right. \\ &\quad \left. 63.516 - 62.567 + \sqrt{\frac{4 \times 0,3409 + 3 \times 0,2166}{7} \left(\frac{1}{5} + \frac{1}{4} \right)} \Phi_R^{-1}(0.975) \right] \\ &= [0.10, 1.80], \text{ où } \Phi_R^{-1}(0.975) = 2.365. \end{aligned}$$

Remarque 6.3.1 Qu'en est-il de la condition très restrictive d'égalité des variances ? En fait, on a pu montrer que celle-ci n'est pas si cruciale si les tailles d'échantillons n_1 et n_2 diffèrent peu. Dans ce cas un facteur 2 pour le rapport des variances reste acceptable. En revanche si n_1 et n_2

différent substantiellement la formule ci-dessus s'applique mal quand les variances ne sont pas proches. Alors on peut effectuer les mêmes développements que précédemment en introduisant les variances respectives des deux lois σ_1^2 et σ_2^2 pour obtenir :

$$\frac{\bar{X}_{n_1} - \bar{Y}_{n_2} - \mu_1 + \mu_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \curvearrowright \mathcal{N}(0, 1),$$

et, si les tailles d'échantillons sont élevées, disons au-delà d'une centaine, conserver une approximation raisonnable en substituant à σ_1^2 et σ_2^2 leurs estimations $\widetilde{s}_{n_1}^2$ et $\widetilde{s}_{n_2}^2$, d'où :

$$IC_{0.95}(\mu_1 - \mu_2) = \left[\bar{x}_{n_1} - \bar{y}_{n_2} - 1.960 \sqrt{\frac{\widetilde{s}_{n_1}^2}{n_1} + \frac{\widetilde{s}_{n_2}^2}{n_2}}, \bar{x}_{n_1} - \bar{y}_{n_2} + 1.960 \sqrt{\frac{\widetilde{s}_{n_1}^2}{n_1} + \frac{\widetilde{s}_{n_2}^2}{n_2}} \right].$$

On remarquera que si $n_1 = n_2$ cette formule est identique à celle du cas où $\sigma_1^2 = \sigma_2^2$.

6.3.2 Intervalle de confiance du rapport de deux variances

Comme précédemment, pour la construction d'un intervalle de confiance de la variance, on distingue deux cas : celui où μ_1 et μ_2 sont connus et celui où μ_1 et μ_2 sont inconnus.

1. Si μ_1 et μ_2 sont connus.

Le meilleur estimateur sans biais et convergent de la variance σ_1^2 (resp. σ_2^2) est la statistique $S_{n_1}^2 = \frac{1}{n_1} \sum_{k=1}^{n_1} (X_k - \mu_1)^2$ (resp. $S_{n_2}^2 = \frac{1}{n_2} \sum_{k=1}^{n_2} (Y_k - \mu_2)^2$). Puisque $n_1 \frac{S_{n_1}^2}{\sigma_1^2} \sim \chi_{(n_1)}^2$ et $n_2 \frac{S_{n_2}^2}{\sigma_2^2} \sim \chi_{(n_2)}^2$ sont des variables aléatoires indépendantes, on peut écrire que

$$R = \frac{\frac{S_{n_1}^2}{\sigma_1^2}}{\frac{S_{n_2}^2}{\sigma_2^2}} \sim \mathcal{F}(n_1, n_2),$$

ce résultat permet de déterminer un intervalle de confiance pour le rapport de deux variances. Étant donné un seuil α , on construit un encadrement, du type :

$$P \left(a \leq \frac{\frac{S_{n_1}^2}{\sigma_1^2}}{\frac{S_{n_2}^2}{\sigma_2^2}} \leq b \right) = 1 - \alpha,$$

où les constantes a et b sont telles que :

$$\begin{aligned} P(R < a) &= \Phi_R(a) = \alpha_1 \iff a = \Phi_R^{-1}(\alpha_1), \\ P(R > b) &= 1 - P(R \leq b) = \alpha_2 \iff b = \Phi_R^{-1}(1 - \alpha_2), \end{aligned}$$

où Φ_R désigne la fonction de répartition de la loi de Fisher-Snedecor à n_1 et n_2 degrés de liberté et $\alpha_1 + \alpha_2 = \alpha$. De l'encadrement précédent, on déduit que :

$$\begin{aligned} P\left(\Phi_R^{-1}(\alpha_1) \leq \frac{\frac{S_{n_1}^2}{\sigma_1^2}}{\frac{S_{n_2}^2}{\sigma_2^2}} \leq \Phi_R^{-1}(1 - \alpha_2)\right) &= 1 - \alpha, \\ \iff P\left(\Phi_R^{-1}(\alpha_1) \leq \frac{\sigma_2^2 S_{n_1}^2}{\sigma_1^2 S_{n_2}^2} \leq \Phi_R^{-1}(1 - \alpha_2)\right) &= 1 - \alpha, \\ \iff P\left(\Phi_R^{-1}(\alpha_1) \frac{S_{n_2}^2}{S_{n_1}^2} \leq \frac{\sigma_2^2}{\sigma_1^2} \leq \Phi_R^{-1}(1 - \alpha_2) \frac{S_{n_2}^2}{S_{n_1}^2}\right) &= 1 - \alpha. \end{aligned}$$

Par conséquent, l'intervalle $\left[\Phi_R^{-1}(\alpha_1) \frac{S_{n_2}^2}{S_{n_1}^2}, \Phi_R^{-1}(1 - \alpha_2) \frac{S_{n_2}^2}{S_{n_1}^2}\right]$ est un intervalle de confiance au niveau de confiance $1 - \alpha$ pour $\frac{\sigma_2^2}{\sigma_1^2}$.

Exemple 6.3.3 Pour un niveau de risque $\alpha = 10\%$ ($\alpha_1 = \alpha_2 = \frac{\alpha}{2}$), une taille d'échantillon $n_1 = 24$ (resp. $n_2 = 22$), une moyenne $\mu_1 = 90$ (resp. $\mu_2 = 94$) et une réalisation de $s_{n_1}^2 = 62.80$ (resp. $s_{n_2}^2 = 55.70$), on obtient une réalisation de l'intervalle de confiance égale à :

$$\begin{aligned} IC_{0.95}\left(\frac{\sigma_2^2}{\sigma_1^2}\right) &= \left[\Phi_R^{-1}\left(\frac{\alpha}{2}\right) \frac{s_{n_2}^2}{s_{n_1}^2}, \Phi_R^{-1}\left(1 - \frac{\alpha}{2}\right) \frac{s_{n_2}^2}{s_{n_1}^2}\right] \\ &= \left[0.50 \times \frac{55.70}{62.80}, 2.03 \times \frac{55.70}{62.80}\right] \\ &= [0.44, 1.80]. \end{aligned}$$

2. Si μ_1 et μ_2 sont inconnus.

Le meilleur estimateur sans biais et convergent de la variance σ_1^2 (resp. σ_2^2) est la statistique $\widetilde{S}_{n_1}^2 = \frac{1}{n_1 - 1} \sum_{k=1}^{n_1} (X_k - \bar{X}_{n_1})^2$ (resp. $\widetilde{S}_{n_2}^2 = \frac{1}{n_2 - 1} \sum_{k=1}^{n_2} (Y_k - \bar{Y}_{n_2})^2$). Puisque $(n_1 - 1) \frac{\widetilde{S}_{n_1}^2}{\sigma_1^2} \sim \chi_{(n_1 - 1)}^2$ et $(n_2 - 1) \frac{\widetilde{S}_{n_2}^2}{\sigma_2^2} \sim \chi_{(n_2 - 1)}^2$ sont des variables aléatoires indépendantes, on peut écrire que

$$R = \frac{\frac{\widetilde{S}_{n_1}^2}{\sigma_1^2}}{\frac{\widetilde{S}_{n_2}^2}{\sigma_2^2}} \sim \mathcal{F}(n_1 - 1, n_2 - 1),$$

ce résultat permet de déterminer un intervalle de confiance pour le rapport de deux variances. Étant donné un seuil α , on construit un encadrement, du type :

$$P \left(a \leq \frac{\frac{\widetilde{S}_{n_1}^2}{\sigma_1^2}}{\frac{\widetilde{S}_{n_2}^2}{\sigma_2^2}} \leq b \right) = 1 - \alpha,$$

où les constantes a et b sont telles que :

$$\begin{aligned} P(R < a) &= \Phi_R(a) = \alpha_1 \iff a = \Phi_R^{-1}(\alpha_1), \\ P(R > b) &= 1 - P(R \leq b) = \alpha_2 \iff b = \Phi_R^{-1}(1 - \alpha_2), \end{aligned}$$

où Φ_R désigne la fonction de répartition de la loi de Fisher-Snedecor à $n_1 - 1$ et $n_2 - 1$ degrés de liberté et $\alpha_1 + \alpha_2 = \alpha$. De l'encadrement précédent, on déduit que :

$$\begin{aligned} &P \left(\Phi_R^{-1}(\alpha_1) \leq \frac{\frac{\widetilde{S}_{n_1}^2}{\sigma_1^2}}{\frac{\widetilde{S}_{n_2}^2}{\sigma_2^2}} \leq \Phi_R^{-1}(1 - \alpha_2) \right) = 1 - \alpha, \\ &\iff P \left(\Phi_R^{-1}(\alpha_1) \leq \frac{\sigma_2^2 \widetilde{S}_{n_1}^2}{\sigma_1^2 \widetilde{S}_{n_2}^2} \leq \Phi_R^{-1}(1 - \alpha_2) \right) = 1 - \alpha, \\ &\iff P \left(\Phi_R^{-1}(\alpha_1) \frac{\widetilde{S}_{n_2}^2}{\widetilde{S}_{n_1}^2} \leq \frac{\sigma_2^2}{\sigma_1^2} \leq \Phi_R^{-1}(1 - \alpha_2) \frac{\widetilde{S}_{n_2}^2}{\widetilde{S}_{n_1}^2} \right) = 1 - \alpha. \end{aligned}$$

Par conséquent, l'intervalle $\left[\Phi_R^{-1}(\alpha_1) \frac{\widetilde{S}_{n_2}^2}{\widetilde{S}_{n_1}^2}, \Phi_R^{-1}(1 - \alpha_2) \frac{\widetilde{S}_{n_2}^2}{\widetilde{S}_{n_1}^2} \right]$ est un intervalle de confiance au niveau de confiance $1 - \alpha$ pour $\frac{\sigma_2^2}{\sigma_1^2}$.

Exemple 6.3.4 Pour un niveau de risque $\alpha = 10\%$ ($\alpha_1 = \alpha_2 = \frac{\alpha}{2}$), une taille d'échantillon $n_1 = 25$ (resp. $n_2 = 23$) et une réalisation de $\widetilde{s}_{n_1}^2 = 62.80$ (resp. $\widetilde{s}_{n_2}^2 = 55.70$), on obtient une réalisation de l'intervalle de confiance égale à :

$$\begin{aligned} IC_{0.95} \left(\frac{\sigma_2^2}{\sigma_1^2} \right) &= \left[\Phi_R^{-1} \left(\frac{\alpha}{2} \right) \frac{\widetilde{s}_{n_2}^2}{\widetilde{s}_{n_1}^2}, \Phi_R^{-1} \left(1 - \frac{\alpha}{2} \right) \frac{\widetilde{s}_{n_2}^2}{\widetilde{s}_{n_1}^2} \right] \\ &= \left[0.50 \times \frac{55.70}{62.80}, 2.03 \times \frac{55.70}{62.80} \right] \\ &= [0.44, 1.80]. \end{aligned}$$

6.4 Exercices

Exercice 6.4.1 Soit (X_1, X_2, X_3) un échantillon aléatoire de $X \sim \mathcal{N}(0, 1)$ et (Y_1, Y_2, Y_3) un échantillon aléatoire de $Y \sim \mathcal{N}(12, 4)$. On suppose que X et Y sont des variables aléatoires indépendantes. On définit

$$T = \sum_{k=1}^3 (X_k - \bar{X}_n)^2 \quad \text{et} \quad Z = \sum_{k=1}^3 \frac{(Y_k - 12)^2}{4}.$$

1. Trouver un nombre a tel que $P(T \leq a) \simeq 0.95$.
2. Trouver un nombre b tel que $P\left(\frac{T}{Z} \leq b\right) \simeq 0.025$.

Exercice 6.4.2 Soit $(X_1, X_2, \dots, X_{20})$ un échantillon aléatoire de $X \sim \mathcal{N}(0, 1)$. On pose

$$\begin{aligned} 4\bar{X}_n(1) &= \sum_{k=1}^4 X_k, & 16\bar{X}_n(2) &= \sum_{k=5}^{20} X_k, \\ 3\widetilde{S}_n^2(1) &= \sum_{k=1}^4 (X_k - \bar{X}_n(1))^2 & \text{et} & 15\widetilde{S}_n^2(2) = \sum_{k=5}^{20} (X_k - \bar{X}_n(2))^2. \end{aligned}$$

1. Calculer $P\left(\frac{2\bar{X}_n(1)}{\sqrt{\widetilde{S}_n^2(1)}} \geq -1.638\right)$.
2. Calculer $P\left(\frac{2\bar{X}_n(1)}{\sqrt{\widetilde{S}_n^2(2)}} < 2.555\right)$.

Exercice 6.4.3 Soit X une variable aléatoire $\mathcal{N}(0, 1)$ et Y une variable aléatoire $\mathcal{N}(0, 4)$. On suppose que X et Y sont indépendantes. Calculer $P\left((X + Y)^2 < 25.1\right)$.

Exercice 6.4.4 Soit X une variable aléatoire t.q :

$$P_\theta(X = i) = (\theta - 1)^{i-1} \theta^{-i}, \quad i \geq 1,$$

où $\theta > 1$. On peut montrer que $E\{X\} = \theta$ et $\text{Var}(X) = (\theta - 1)\theta$.

1. Déterminer l'estimateur $\hat{\theta}_{MM}$ de θ par la méthode des moments.
2. Calculer l'écart quadratique moyen de $\hat{\theta}_{MM}$.
3. Obtenir un intervalle de confiance approximatif à 95% pour θ si un échantillon aléatoire de X , de taille $n = 50$, a donné une moyenne \bar{x}_n égale à 2.

Exercice 6.4.5 On veut estimer le rendement d'un engrais pour la culture du blé. Sur douze parcelles expérimentales, on a trouvé les rendements suivants en tonnes par hectare :

7.7; 8.4; 7.8; 8.2; 7.9; 8.5; 8.4; 8.2; 7.6; 7.8; 8.4; 8.3.

Donner un intervalle de confiance à 95% pour le rendement moyen de l'engrais (on supposera que le rendement à l'hectare est une v.a. gaussienne).

Exercice 6.4.6 Dans un centre avicole, des études antérieures ont montré que la masse d'un oeuf choisi au hasard peut être considérée comme la réalisation d'une variable aléatoire normale X , de moyenne μ et de variance σ^2 . On admet que les masses des oeufs sont indépendantes les unes des autres. On prend un échantillon de $n = 36$ oeufs que l'on pèse. Les mesures sont données (par ordre croissant) dans le tableau suivant :

50.34	52.62	53.79	54.99	55.82	57.67
51.41	53.13	53.89	55.04	55.91	57.99
51.51	53.28	54.63	55.12	55.95	58.10
52.07	53.30	54.76	55.24	57.05	59.30
52.22	53.32	54.78	55.28	57.18	60.58
52.38	53.39	54.93	55.56	57.31	63.15

Donner un intervalle de confiance au niveau 95% de la masse moyenne μ d'un oeuf.

Solutions

Solution 6.4.1 1. On a : $\widetilde{S}_n^2 = \frac{1}{3-1} \sum_{k=1}^3 (X_k - \bar{X}_n)^2$, alors, on peut écrire que $T = 2\widetilde{S}_n^2 \sim \chi_{3-1}^2$.

Maintenant, on trouve dans le tableau des valeurs de χ_2^2 au 0.05, que $P(X \geq 5.99) \simeq 0.05$ si $X \sim \chi_2^2$.

Il s'ensuit que $a = 5.99$, puisque

$$P(T \leq 5.99) = 1 - P(T > 5.99) \simeq 1 - 0.05 = 0.95.$$

2. On a : $\frac{(Y_k - 12)^2}{4} \sim \chi_1^2$ pour $k = 1, \dots, 3$ et, par indépendance, $Z \sim \chi_3^2$. Alors

$$P\left(\frac{T}{Z} \leq b\right) = P\left(\frac{T/2}{Z/3} \leq \frac{3}{2}b\right).$$

De plus, nous avons vu que si $X \sim \mathcal{F}(m, n)$, alors $\frac{1}{X} \sim \mathcal{F}(n, m)$; il s'ensuit que

$$P\left(\frac{T/2}{Z/3} \leq \frac{3}{2}b\right) = P\left(\frac{Z/3}{T/2} \geq \frac{2}{3b}\right) \simeq 0.025.$$

Maintenant, on trouve dans le tableau des valeurs de $\mathcal{F}(3, 2)$ au 0.025, que $P\left(\frac{Z/3}{T/2} \geq 39.2\right) \simeq 0.025$

si $\frac{Z/3}{T/2} \sim \mathcal{F}(3, 2)$. Il s'ensuit que $\frac{2}{3b} \simeq 39.2$. Donc, on trouve que $b \simeq 0.0255$.

Remarque 6.4.1 Si l'on définit plutôt $Z = \sum_{k=1}^3 \frac{(Y_k - \bar{Y}_n)^2}{4}$, alors on a : $Z \sim \chi_2^2$ et $\frac{T}{Z} \sim \mathcal{F}(2, 2)$. Il s'ensuit que $\frac{1}{b} \simeq 39.0 \iff b \simeq 0.0256$ (car $P\left(\frac{Z}{T} \geq 39.0\right) \simeq 0.025$).

Solution 6.4.2 1. On a : $\frac{\sqrt{4\bar{X}_n(1)}}{\sqrt{\widetilde{S}_n^2(1)}} \sim \mathcal{T}_3$. Alors,

$$P\left(\frac{2\bar{X}_n(1)}{\sqrt{\widetilde{S}_n^2(1)}} \geq -1.638\right) = P\left(\frac{2\bar{X}_n(1)}{\sqrt{\widetilde{S}_n^2(1)}} \leq 1.638\right) \simeq 0.90.$$

2. D'abord, on a : $\frac{\sqrt{4\bar{X}_n(1)}}{\sqrt{\widetilde{S}_n^2(2)}} \sim \mathcal{T}_{15}$

$$P\left(\frac{2\bar{X}_n(1)}{\sqrt{\widetilde{S}_n^2(2)}} < 2.555\right) \simeq 0.99.$$

Solution 6.4.3 On a : $X + Y \sim \mathcal{N}(0 + 0, 1 + 4) \equiv \mathcal{N}(0, 5)$ (car X et Y sont indépendantes), donc $\frac{(X + Y)^2}{5} \sim \chi_1^2$. Alors,

$$P\left((X + Y)^2 < 25.1\right) = P\left(\frac{(X + Y)^2}{5} < 5.02\right) = 0.975.$$

Solution 6.4.4 1. On pose :

$$E\{X\} = \bar{X}_n \iff \theta = \bar{X}_n,$$

donc, l'estimateur de θ par la méthode des moments est $\hat{\theta}_{MM} = \bar{X}_n$.

2. On a :

$$E\{\hat{\theta}_{MM}\} = E\{\bar{X}_n\} = E\{X\} = \theta.$$

Alors on peut écrire que

$$r_{\theta}(\hat{\theta}_{MM}) = \text{Var}(\hat{\theta}_{MM}) = \text{Var}(\bar{X}_n) = \frac{\text{Var}(X)}{n} = \frac{(\theta - 1)\theta}{n}.$$

3. On déduit du théorème central limite et de la partie (2) que

$$\hat{\theta}_{MM} = \bar{X}_n \approx \mathcal{N}\left(\theta, \frac{(\theta - 1)\theta}{n}\right).$$

Il s'ensuit que

$$P\left(-a_{\alpha/2} \leq \frac{\hat{\theta}_{MM} - \theta}{\sqrt{\frac{(\theta - 1)\theta}{n}}} \leq a_{\alpha/2}\right) \simeq 1 - \alpha \implies P\left(-a_{\alpha/2} \leq \frac{\hat{\theta}_{MM} - \theta}{\sqrt{\frac{(\hat{\theta}_{MM} - 1)\hat{\theta}_{MM}}{n}}} \leq a_{\alpha/2}\right) \simeq 1 - \alpha.$$

De là, on trouve que l'intervalle de confiance approximatif à $100(1 - \alpha)\%$ pour θ est donné par

$$\begin{aligned} & \left[\hat{\theta}_{MM} - a_{\alpha/2} \sqrt{\frac{(\hat{\theta}_{MM} - 1)\hat{\theta}_{MM}}{n}}, \hat{\theta}_{MM} + a_{\alpha/2} \sqrt{\frac{(\hat{\theta}_{MM} - 1)\hat{\theta}_{MM}}{n}} \right] \\ \iff & \left[\bar{X}_n - a_{\alpha/2} \sqrt{\frac{(\bar{X}_n - 1)\bar{X}_n}{n}}, \bar{X}_n + a_{\alpha/2} \sqrt{\frac{(\bar{X}_n - 1)\bar{X}_n}{n}} \right]. \end{aligned}$$

Avec $\alpha = 0.05$, $n = 50$, $a_{\alpha/2} = 1.960$ et $\bar{x}_n = 2$, on obtient :

$$\left[2 - 1.960 \sqrt{\frac{(2 - 1)2}{50}}, 2 + 1.960 \sqrt{\frac{(2 - 1)2}{50}} \right] \iff [1.61, 2.39].$$

Solution 6.4.5 On calcule $\bar{x}_n = 8.10$, $\tilde{s}_n^2 = 0.1018$, $\sqrt{\tilde{s}_n^2} = 0.319$ et on lit $t_{0.975}^{(11)} = 2.201$. D'où :

$$\begin{aligned} IC_{0.95}(\mu) &= \left[8.10 - 2.201\sqrt{\frac{0.1018}{12}}, 8.10 + 2.201\sqrt{\frac{0.1018}{12}} \right] \\ &= [7.90, 8.30]. \end{aligned}$$

Solution 6.4.6 IC de niveau de confiance $1 - \alpha = 95\%$ pour μ :

$$\left[\bar{x}_n - a\frac{\sqrt{\tilde{s}_n^2}}{\sqrt{n}}, \bar{x}_n + a\frac{\sqrt{\tilde{s}_n^2}}{\sqrt{n}} \right] = [54.207, 55.96],$$

car $P(X \leq 1.96) = 0.975$ quand X de loi $\mathcal{N}(0, 1)$ donc $a = 1.96$, et $\bar{x}_n = 55.083$, $\tilde{s}_n^2 = 7.195$.

Chapitre 7

Tests d'Hypothèses

On a introduit au chapitre précédent le concept d'intervalle de confiance que l'on a utilisé pour faire de l'inférence sur certains paramètres de la population. Dans ce chapitre, nous présentons une approche alternative pour faire de l'inférence via le concept de test statistique.

La théorie des tests (ou inférence) étudie la construction et les propriétés des tests statistiques. Un test statistique est une règle de décision permettant de rejeter ou de ne pas rejeter une hypothèse, appelée hypothèse nulle, en fonction des observations d'un échantillon. La théorie des tests est la théorie fondamentale de ce que l'on appelle aujourd'hui la statistique décisionnelle ou business intelligence, et de ce que l'on appelait autrefois la statistique mathématique. Elle est le fondement de tous les outils statistiques modernes d'aide à la décision. Au-delà de la règle de décision, le principal avantage d'un test statistique est qu'il permet de mesurer ou de contrôler les risques associés à cette décision. C'est pourquoi ces tests sont très utilisés en pratique.

7.1 Définitions

Définition 7.1.1 *Un **test statistique** (ou **test**) est une règle de décision relative à une hypothèse sur la distribution d'une variable d'intérêt dans la population, qui se fonde sur les observations d'un échantillon.*

Un test statistique est une procédure qui permet de contrôler ou de minimiser, suivant les cas, les risques associés à la décision. C'est pourquoi, un test statistique est toujours associé à trois éléments :

- Une **hypothèse nulle** et une **hypothèse alternative**.

- Une **région critique** fondée sur une statistique de test et une valeur critique.
- Des **risques** de première espèce et de seconde espèce.

Nous allons désormais présenter ces différents éléments.

7.1.1 Hypothèses nulle et alternative

Un test, en tant que règle de décision, se réfère toujours à une hypothèse de référence, dite hypothèse nulle.

Définition 7.1.2 *Une hypothèse est une assertion concernant la population. Un test statistique permet de tester la validité d'une hypothèse de référence (ou de base), dite hypothèse nulle, contre une hypothèse alternative. Ces hypothèses sont respectivement notées \mathbf{H}_0 et \mathbf{H}_1 .*

Considérons l'exemple d'un test concernant l'effet d'un traitement médical. Dans ce cas, on peut construire un test de l'hypothèse nulle \mathbf{H}_0 : « le traitement n'a pas d'effet » contre une hypothèse alternative \mathbf{H}_1 : « le traitement a un effet ». Mais on peut aussi tester l'inverse, c'est-à-dire \mathbf{H}_0 : « le traitement a un effet » contre \mathbf{H}_1 : « le traitement n'a pas d'effet ».

On distingue deux grandes familles de tests statistiques :

- Les tests paramétriques : ces tests portent sur la valeur d'un ou de plusieurs paramètres de la distribution dans la population d'une variable d'intérêt.
- Les tests non-paramétriques : ces tests portent sur la distribution, les moments (espérance, variance, etc.) ou certaines caractéristiques (l'indépendance par exemple) d'une ou de plusieurs variables aléatoires.

Dans ce chapitre nous étudierons tout d'abord les tests paramétriques. Dans le cas des tests paramétriques, l'hypothèse nulle et l'hypothèse alternative portent sur la valeur d'un paramètre θ (ou de plusieurs paramètres) de la distribution d'une variable aléatoire (discrète ou continue) X définie sur un univers probabilisé (Ω, \mathcal{F}, P) et admettant une fonction de densité $f_{X,\theta}(x)$ ou une fonction de masse $P_\theta(X = x)$.

Exemple 7.1.1 *On admet qu'une variable aléatoire discrète X , définie sur \mathbb{N} , suit une loi de Poisson de paramètre θ , où θ est un paramètre réel positif inconnu, telle que :*

$$P_\theta(X = x) = e^{-\theta} \frac{\theta^x}{x!} \forall x \in \mathbb{N}.$$

On cherche à tester l'hypothèse nulle $\mathbf{H}_0 : \theta = 3$ contre une hypothèse alternative $\mathbf{H}_1 : \theta = 4$.

Parmi les tests paramétriques, on distingue les tests d'hypothèses simples et les tests d'hypothèses composites.

Définition 7.1.3 Une *hypothèse simple* caractérise complètement la distribution de la variable d'intérêt. Une *hypothèse composite* ne permet pas de caractériser la distribution de la variable d'intérêt.

Exemple 7.1.2 Considérons une variable aléatoire X distribuée selon une loi de Student $\mathcal{T}(\theta)$ où $\theta > 0$ est un paramètre inconnu. L'hypothèse nulle $\mathbf{H}_0 : \theta = 3$ est une hypothèse simple, car sous \mathbf{H}_0 on connaît exactement la loi de la variable X , i.e. $X \sim \mathcal{T}(3)$. Les hypothèses nulles $\mathbf{H}_0 : \theta > 3$, $\mathbf{H}_0 : \theta < 3$ ou $\mathbf{H}_0 : \theta \neq 3$ sont des hypothèses composites. En effet, on ne sait pas quelle est la loi exacte de X sous \mathbf{H}_0 . Pour $\mathbf{H}_0 : \theta > 3$, cette distribution peut être, par exemple, $X \sim \mathcal{T}(4)$, $X \sim \mathcal{T}(11)$ ou $X \sim \mathcal{T}(111)$.

Remarque 7.1.1 On peut construire des tests d'une hypothèse simple contre une hypothèse simple. Par exemple, $\mathbf{H}_0 : \theta = \theta_0$ contre $\mathbf{H}_1 : \theta = \theta_1$. On peut construire différents tests d'une hypothèse simple contre une hypothèse composite. Par exemple, $\mathbf{H}_0 : \theta = \theta_0$ contre $\mathbf{H}_1 : \theta \neq \theta_0$. On peut aussi construire, même si c'est plus rare, des tests d'une hypothèse composite contre une hypothèse simple ou une hypothèse composite, du type $\mathbf{H}_0 : \theta < \theta_0$ contre $\mathbf{H}_1 : \theta = \theta_0$, ou $\mathbf{H}_0 : \theta < \theta_0$ contre $\mathbf{H}_1 : \theta > \theta_0$.

Parmi les tests d'une hypothèse simple contre une hypothèse composite, on distingue les **tests unilatéraux** des **tests bilatéraux**. Cette distinction sera particulièrement importante pour la définition de la région critique. Le terme unilatéral renvoie au fait que sous \mathbf{H}_1 , la valeur du paramètre θ ne peut être que supérieure (ou inférieure suivant le test) à la valeur de θ sous l'hypothèse nulle \mathbf{H}_0 : la valeur de θ ne prend qu'une seule « direction ». Le terme bilatéral signifie, qu'au contraire, la valeur de θ sous l'hypothèse alternative est différente (inférieure ou supérieure) de la valeur sous l'hypothèse nulle.

Définition 7.1.4 Un *test unilatéral gauche* est un test de la forme $\mathbf{H}_0 : \theta = \theta_0$ contre $\mathbf{H}_1 : \theta < \theta_0$. Un *test unilatéral droit* est un test de la forme $\mathbf{H}_0 : \theta = \theta_0$ contre $\mathbf{H}_1 : \theta > \theta_0$.

Définition 7.1.5 Un *test bilatéral* est un test de la forme $\mathbf{H}_0 : \theta = \theta_0$ contre $\mathbf{H}_1 : \theta \neq \theta_0$.

Enfin, signalons que lorsqu'un test porte sur plusieurs paramètres, on parle de test d'hypothèses jointes ou de **test joint**.

Définition 7.1.6 *Un test d'hypothèses jointes (ou test joint)* est un test dont l'hypothèse nulle porte sur plusieurs paramètres $\theta_1, \dots, \theta_r$ de la distribution de la variable d'intérêt.

$$\mathbf{H}_0 : \theta_1 = a_1, \theta_2 = a_2, \dots, \theta_r = a_r.$$

Par exemple, si la variable d'intérêt X vérifie $X \sim \mathcal{N}(\mu, \sigma^2)$, on peut construire un test joint de la forme $\mathbf{H}_0 : \mu = \mu_0$ et $\sigma^2 = \sigma_0^2$. L'hypothèse alternative peut s'écrire sous la forme $\mathbf{H}_1 : \mu \neq \mu_0$ et $\sigma^2 \neq \sigma_0^2$ ou sous la forme $\mathbf{H}_1 : \mu \neq \mu_0$ ou $\sigma^2 \neq \sigma_0^2$.

7.1.2 Région critique

Supposons que l'on dispose d'un n -échantillon (X_1, \dots, X_n) de variables i.i.d. de même loi que la variable X . Comment tester, à partir de cet échantillon, l'hypothèse nulle d'un test paramétrique portant sur la valeur d'un paramètre θ de sa distribution dans la population? Pour ce faire, nous allons construire une région critique à partir de deux éléments : une statistique de test¹ et une valeur critique. Commençons par définir la notion de statistique de test.

Définition 7.1.7 *Une statistique de test, notée T_n , est une variable aléatoire définie comme une fonction des variables de l'échantillon $(X_1, \dots, X_n) : T_n(X_1, \dots, X_n)$.*

Remarque 7.1.2 *En général, mais pas toujours, la statistique de test correspond à un estimateur $\hat{\theta}$ du paramètre θ ou à une variable transformée de cet estimateur.*

Une statistique de test étant une variable aléatoire, on peut en caractériser la distribution (ou distribution d'échantillonnage). Comme pour un estimateur, on distingue la loi exacte d'une statistique de test, valable pour toute valeur de n (cette loi exacte est généralement difficile à dériver sauf dans des cas simples), de la loi asymptotique, valable pour une taille d'échantillon n suffisamment grande, mais finie.

La région critique correspond à la règle de décision du test statistique. Cette règle est extrêmement simple : si la réalisation de la statistique de test, obtenue à partir des observations (x_1, \dots, x_n) appartient à la région critique, on rejette l'hypothèse nulle \mathbf{H}_0 . La région critique est un ensemble délimité par une ou des valeurs critiques, suivant les cas.

¹Nous avons déjà défini la notion de statistique dans le chapitre 4, consacré à l'estimation. La définition d'une statistique de test est similaire.

Définition 7.1.8 La *région critique* d'un test, notée W , est un ensemble de réalisations de la statistique de test (ou de façon équivalente un ensemble d'échantillon) pour lesquelles l'hypothèse nulle du test est rejetée :

$$W = \{x_1, \dots, x_n : T_n(x_1, \dots, x_n) \in \Delta(c)\},$$

où (x_1, \dots, x_n) désigne un échantillon, $T_n(x_1, \dots, x_n)$ la réalisation associée de la statistique de test, et $\Delta(c)$ un ensemble délimité par une (ou plusieurs) valeur(s) critique(s), notée(s) c .

Exemple 7.1.3 Voici quelques exemples de formes de régions critiques usuelles :

$$W = \{x_1, \dots, x_n : T_n(x_1, \dots, x_n) > c\},$$

$$W = \{x_1, \dots, x_n : c < T_n(x_1, \dots, x_n) < c_1\},$$

$$W = \{x_1, \dots, x_n : |T_n(x_1, \dots, x_n)| > c\},$$

où c et c_1 sont des valeurs critiques, généralement déterminées à partir de tables statistiques.

Ainsi, la procédure d'un test est la suivante : on calcule la réalisation de la statistique de test à partir de l'échantillon d'observations. Si cette réalisation appartient à la région critique, on rejette l'hypothèse nulle \mathbf{H}_0 . Par conséquent, un test statistique est une règle de décision qui spécifie :

- l'ensemble des échantillons pour lesquels on rejette \mathbf{H}_0 ;
- l'ensemble des échantillons pour lesquels on ne peut pas rejeter \mathbf{H}_0 .

Définition 7.1.9 La région complémentaire de la région critique est appelée *région de non-rejet* de l'hypothèse nulle \mathbf{H}_0 , notée \overline{W} , telle que :

$$\overline{W} = \{x_1, \dots, x_n : T_n(x_1, \dots, x_n) \notin \Delta(c)\}.$$

Si la réalisation de la statistique de test appartient à la zone de non-rejet, on conclut au non-rejet de l'hypothèse nulle \mathbf{H}_0 .

7.1.3 Risques

Lorsque l'on considère un test statistique, on ne peut prendre que l'une ou l'autre des deux décisions suivantes : soit on rejette l'hypothèse nulle \mathbf{H}_0 , soit on ne rejette pas l'hypothèse nulle \mathbf{H}_0 . L'avantage

principal d'un test statistique par rapport à une règle de décision heuristique (une décision au hasard par exemple), est qu'il permet de contrôler les risques associés à la décision.

Quels sont ces risques ? On distingue deux types de risque : le risque de première espèce et le risque de deuxième espèce. Dans le tableau ci-dessous, on croise la décision (rejet ou non-rejet de \mathbf{H}_0) et la validité de \mathbf{H}_0 ou de \mathbf{H}_1 dans la population. Ainsi, si l'on rejette \mathbf{H}_0 alors que \mathbf{H}_1 est vraie ou si l'on ne rejette pas \mathbf{H}_0 alors que \mathbf{H}_0 est vraie, on ne commet pas d'erreur. En revanche, si l'on rejette \mathbf{H}_0 alors que \mathbf{H}_0 est vraie, on commet une erreur dite de type *I* ou de première espèce. Si on ne rejette pas \mathbf{H}_0 alors que \mathbf{H}_1 est vraie, on commet une erreur dite de type *II* ou de deuxième espèce. Il y a quatre cas qui sont reproduits dans le tableau ci-dessous

		Décision	
		Non-rejet de \mathbf{H}_0	Rejet de \mathbf{H}_0
Population	\mathbf{H}_0 vraie	Décision correcte	Erreur de type <i>I</i>
	\mathbf{H}_1 vraie	Erreur de type <i>II</i>	Décision correcte

Tableau : Risque *I* et risque *II*.

Définition 7.1.10 *Le risque I ou risque de première espèce, correspond au risque de rejeter l'hypothèse nulle \mathbf{H}_0 alors qu'elle est effectivement vraie dans la population.*

Définition 7.1.11 *Le risque II ou risque de seconde espèce correspond au risque de ne pas rejeter l'hypothèse nulle \mathbf{H}_0 alors que l'hypothèse alternative \mathbf{H}_1 est valide dans la population.*

Pour une règle de décision donnée, c'est-à-dire pour une région critique W , on cherche à quantifier les probabilités associées à ces deux types de risque.

Définition 7.1.12 *Le niveau (ou la taille) d'un test correspond à la probabilité associée au risque de première espèce. Par convention, cette probabilité est notée α : $\alpha = P(W | \mathbf{H}_0)$ où W désigne la région critique du test.*

Ainsi, le niveau correspond à la probabilité de rejeter \mathbf{H}_0 , c'est-à-dire d'être dans la région critique W , sachant que l'hypothèse nulle \mathbf{H}_0 est vraie dans la population. C'est donc la probabilité de rejeter à tort l'hypothèse nulle \mathbf{H}_0 . Bien évidemment, plus le niveau d'un test est faible, plus la probabilité d'erreur de première espèce est faible et mieux c'est. Le symbole $|\mathbf{H}_0$ signifie «sachant que \mathbf{H}_0 est vraie».

Remarque 7.1.3 Pour un test d'hypothèse nulle composite, le niveau du test devient : $\alpha = \sup_{\theta_0 \in \mathbf{H}_0} P(W | \mathbf{H}_0)$.

De façon similaire, nous pouvons définir la probabilité associée au risque de deuxième espèce et la puissance d'un test, définie comme le complémentaire de cette probabilité.

Définition 7.1.13 La **puissance** d'un test correspond à la probabilité de rejet de l'hypothèse nulle \mathbf{H}_0 alors que l'hypothèse alternative \mathbf{H}_1 est vraie :

$$\text{Puissance} = P(W | \mathbf{H}_1) = 1 - \beta,$$

où β correspond à la probabilité de l'erreur de deuxième espèce, i.e., $\beta = P(\bar{W} | \mathbf{H}_1)$.

La puissance correspond à la probabilité d'être dans la région critique (et donc de rejeter l'hypothèse nulle) alors que l'hypothèse alternative \mathbf{H}_1 est vraie dans la population. Par conséquent, plus un test est puissant, plus la probabilité d'erreur de deuxième espèce est faible et mieux c'est.

Exemple 7.1.4 On considère un n -échantillon (X_1, \dots, X_n) , avec $n = 100$, de variables aléatoires i.i.d. telles que $X \sim \mathcal{N}(\mu, 1)$ où μ est un paramètre inconnu. On souhaite tester :

$$\mathbf{H}_0 : \mu = \mu_0 = 1.2 \text{ contre } \mathbf{H}_1 : \mu = \mu_1 = 1.$$

Un économètre vous propose une région critique de la forme : $W = \{x_1, \dots, x_n : \bar{x}_n < c\}$, où \bar{x}_n désigne la réalisation de la moyenne empirique \bar{X}_n et c est une constante (valeur critique) égale à 1.0718. Cette région critique s'interprète de la façon suivante : si la réalisation de la moyenne empirique est inférieure à 1.0718, on rejette l'hypothèse nulle $\mathbf{H}_0 : \mu = 1.2$. Calculons la taille et la puissance de ce test. Sous l'hypothèse nulle $\mathbf{H}_0 : \mu = \mu_0$, la loi exacte de la moyenne empirique \bar{X}_n (statistique de test) est

$$\bar{X}_n \underset{\text{sous } \mathbf{H}_0}{\sim} \mathcal{N}\left(\mu_0, \frac{1}{n}\right) \iff \frac{\bar{X}_n - \mu_0}{1/\sqrt{n}} \underset{\text{sous } \mathbf{H}_0}{\sim} \mathcal{N}(0, 1).$$

Par conséquent, la taille du test est égale à :

$$\begin{aligned} \alpha &= P(W | \mathbf{H}_0) \\ &= P(\bar{X}_n < c | \mathbf{H}_0) \\ &= P\left(\frac{\bar{X}_n - \mu_0}{1/\sqrt{n}} < \frac{c - \mu_0}{1/\sqrt{n}} | \mathbf{H}_0\right) \\ &= F\left(\frac{c - \mu_0}{1/\sqrt{n}}\right), \end{aligned}$$

où F désigne la fonction de répartition de la loi normale centrée réduite. D'après les données de l'énoncé, la taille du test est égale à :

$$\alpha = F\left(\frac{1.0718 - 1.2}{1/\sqrt{100}}\right) = F(-1.2816) = 0.10.$$

Ainsi, avec la règle de décision associée à la région critique $W = \{x_1, \dots, x_n : \bar{x}_n < 1.0718\}$, il y a 10% de chances de rejeter à tort l'hypothèse nulle $\mathbf{H}_0 : \mu = 1.2$ alors qu'elle est vraie. Sous l'hypothèse alternative, $\mathbf{H}_1 : \mu = \mu_1$, la loi exacte de la moyenne empirique \bar{X}_n (statistique de test) est :

$$\bar{X}_n \underset{\text{sous } \mathbf{H}_1}{\sim} \mathcal{N}\left(\mu_1, \frac{1}{n}\right) \iff \frac{\bar{X}_n - \mu_1}{1/\sqrt{n}} \underset{\text{sous } \mathbf{H}_1}{\sim} \mathcal{N}(0, 1).$$

Par conséquent, la puissance du test est égale à :

$$\begin{aligned} \text{Puissance} &= P(W | \mathbf{H}_1) \\ &= P(\bar{X}_n < c | \mathbf{H}_1) \\ &= P\left(\frac{\bar{X}_n - \mu_1}{1/\sqrt{n}} < \frac{c - \mu_1}{1/\sqrt{n}} | \mathbf{H}_1\right) \\ &= F\left(\frac{c - \mu_1}{1/\sqrt{n}}\right). \end{aligned}$$

La probabilité de risque de deuxième espèce est égale à :

$$\beta = 1 - \text{Puissance} = 1 - F\left(\frac{c - \mu_1}{1/\sqrt{n}}\right).$$

D'où :

$$\begin{aligned} \text{Puissance} &= F\left(\frac{1.0718 - 1}{1/\sqrt{100}}\right) = F(0.7184) = 0.7638 \\ \beta &= 1 - \text{Puissance} = 1 - 0.7638 = 0.2362. \end{aligned}$$

Par conséquent, avec la région critique $W = \{x_1, \dots, x_n : \bar{x}_n < 1.0718\}$, il y a 23.62% de chances de ne pas rejeter l'hypothèse nulle $\mathbf{H}_0 : \mu = 1.2$ alors que l'hypothèse alternative $\mathbf{H}_1 : \mu = 1$ est vraie.

Remarque 7.1.4 La quantité β dépend de la taille n de l'échantillon aléatoire : plus n augmente et plus β diminue. De plus, contrairement à α , la valeur de β n'est pas unique dans un problème donné mais dépend de l'hypothèse particulière \mathbf{H}_1 que l'on suppose vraie. Il y a en fait une infinité d'hypothèses particulières \mathbf{H}_1 , tandis que \mathbf{H}_0 est unique (dans le cas où \mathbf{H}_0 est une hypothèse simple).

Remarque 7.1.5 Ces deux probabilités d'erreur sont antagonistes, plus α est grande (respectivement petite), plus β est petite (respectivement grande).

Remarque 7.1.6 Dans la pratique des tests statistiques, il est de règle de se fixer α . Les valeurs les plus courantes sont 1%, 5% ou 10%.

Remarque 7.1.7 On dit que le test T_1 est **plus puissant** que le test T_2 au niveau α s'il est de niveau α , si T_2 est de niveau égal (ou inférieur) à α et si la puissance de T_1 est supérieure à celle de T_2 .

7.2 Règle de décision

L'objectif de cette section est de présenter la règle de décision d'un test pour un niveau de risque de première espèce donné.

Propriété 7.2.1 Si la réalisation de la statistique de test appartient à la région critique, on rejette l'hypothèse nulle \mathbf{H}_0 pour un niveau de risque (ou seuil de significativité) α . Si, au contraire, la réalisation de la statistique de test n'appartient pas à la région critique, on conclut que l'on ne peut pas rejeter l'hypothèse nulle pour un niveau de risque (ou seuil de significativité) α .

Il est donc essentiel de préciser le niveau de risque associé à la décision : on rejette \mathbf{H}_0 au seuil de 5%, de 10%, etc., car la conclusion peut en effet être tout autre pour un niveau de risque de 15% par exemple.

7.3 Les tests paramétriques usuels

Certains de ces tests ont, en fait, déjà été vus dans la théorie générale et nous les indiquerons plus brièvement. La construction des tests usuels découle de la présence d'une statistique de loi connue sous \mathbf{H}_0 et souvent sous \mathbf{H}_1 , s'imposant de façon assez naturelle.

7.3.1 Tests sur la moyenne d'une loi $\mathcal{N}(\mu, \sigma^2)$

On considère un échantillon (X_1, X_2, \dots, X_n) d'une variable aléatoire X qui suit la loi normale $\mathcal{N}(\mu, \sigma^2)$.

1. Cas où σ^2 est connu

On sait que la moyenne empirique \bar{X}_n est un estimateur sans biais de μ , tel que : $\bar{X}_n \sim \mathcal{N}(\mu, \frac{\sigma^2}{n})$ et la statistique $\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}$ de loi connue : $\mathcal{N}(0, 1)$.

Ainsi pour une hypothèse nulle $\mathbf{H}_0 : \mu = \mu_0$ contre $\mathbf{H}_1 : \mu \neq \mu_0$, on a, sous \mathbf{H}_0 :

$$\frac{\bar{X}_n - \mu_0}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1).$$

Comme :

$$\begin{aligned} 1 - \alpha &= P_{\mu_0} \left(-a \leq \frac{\bar{X}_n - \mu_0}{\sigma/\sqrt{n}} \leq a \right) \\ &= P_{\mu_0} \left(\mu_0 - a \frac{\sigma}{\sqrt{n}} \leq \bar{X}_n \leq \mu_0 + a \frac{\sigma}{\sqrt{n}} \right), \end{aligned}$$

donc, une région de non-rejet peut être définie, pour un test de risque α , par :

$$\bar{W} = \left\{ x_1, \dots, x_n : \mu_0 - a \frac{\sigma}{\sqrt{n}} \leq \bar{x}_n \leq \mu_0 + a \frac{\sigma}{\sqrt{n}} \right\}.$$

Pour des hypothèses unilatérales, par exemple $\mathbf{H}_0 : \mu \leq \mu_0$ contre $\mathbf{H}_1 : \mu > \mu_0$, il est naturel de rejeter \mathbf{H}_0 lorsque \bar{x}_n est trop grand car cela reflète une moyenne μ élevée. Pour déterminer la valeur critique on se place en $\mu = \mu_0$ qui est la valeur la plus défavorable pour \mathbf{H}_0 . Comme :

$$\alpha = P_{\mu_0} \left(a < \frac{\bar{X}_n - \mu_0}{\sigma/\sqrt{n}} \right),$$

on a pour région de rejet

$$W = \left\{ x_1, \dots, x_n : \bar{x}_n > \mu_0 + a \frac{\sigma}{\sqrt{n}} \right\}.$$

Pour $\mathbf{H}_0 : \mu \geq \mu_0$ contre $\mathbf{H}_1 : \mu < \mu_0$, la région de rejet sera

$$W = \left\{ x_1, \dots, x_n : \bar{x}_n < \mu_0 - a \frac{\sigma}{\sqrt{n}} \right\}.$$

Exemple 7.3.1 On prélève, au hasard, dans une population suivant une loi normale de variance égale à 25, un échantillon de taille $n = 16$.

- En choisissant un risque de première espèce $\alpha = 0.05$ (risque bilatéral, symétrique), quelle est la règle de décision si l'on veut tester les hypothèses : $\mathbf{H}_0 : \mu = \mu_0 = 45$ contre $\mathbf{H}_1 : \mu \neq 45$? Soient k_1 et k_2 les seuils critiques. La règle de décision est : on accepte l'hypothèse \mathbf{H}_0 si $k_1 < \bar{x}_n < k_2$,

$$\begin{aligned} P_{\mu_0} (k_1 \leq \bar{X}_n \leq k_2) &= 0.95, \\ P_{\mu_0} \left(\frac{k_1 - 45}{5/4} \leq \frac{\bar{X}_n - 45}{5/4} \leq \frac{k_2 - 45}{5/4} \right) &= 0.95, \end{aligned}$$

d'où $\frac{k_1 - 45}{5/4} = -1.96$, $\frac{k_2 - 45}{5/4} = 1.96$, $k_1 = 42.55$, $k_2 = 47.45$.

– On observe une moyenne de l'échantillon égale à 49. Cette valeur est en contradiction avec l'hypothèse \mathbf{H}_0 , on refuse donc l'hypothèse \mathbf{H}_0 et on accepte l'hypothèse \mathbf{H}_1 .

On peut calculer le risque de deuxième espèce β associé à cette valeur $\mu_1 = 49$:

$$\begin{aligned} \beta &= P\left(\frac{42.55 - 49}{5/4} \leq \frac{\bar{X}_n - 49}{5/4} \leq \frac{47.45 - 49}{5/4} \mid \mathbf{H}_1\right) \\ &= P\left(-5.16 \leq \frac{\bar{X}_n - 49}{5/4} \leq -1.24 \mid \mathbf{H}_1\right) \\ &= 0.1075. \end{aligned}$$

La probabilité de refuser l'hypothèse $\mathbf{H}_1 : \mu = \mu_1 = 49$, alors qu'elle est vraie, est égale à 0.1075, la puissance du test est égale à 0.8925.

2. Cas où σ^2 est inconnu

On sait que la moyenne empirique \bar{X}_n et la variance empirique corrigée \widetilde{S}_n^2 sont des estimateurs sans biais de μ et σ^2 respectivement. Ainsi pour une hypothèse nulle $\mathbf{H}_0 : \mu = \mu_0$ contre $\mathbf{H}_1 : \mu \neq \mu_0$, on a, sous \mathbf{H}_0 :

$$\frac{\bar{X}_n - \mu_0}{\sqrt{\widetilde{S}_n^2}/\sqrt{n}} \sim \mathcal{T}(n-1).$$

Comme :

$$\begin{aligned} 1 - \alpha &= P_{\mu_0} \left(-a \leq \frac{\bar{X}_n - \mu_0}{\sqrt{\widetilde{S}_n^2}/\sqrt{n}} \leq a \right) \\ &= P_{\mu_0} \left(\mu_0 - a \frac{\sqrt{\widetilde{S}_n^2}}{\sqrt{n}} \leq \bar{X}_n \leq \mu_0 + a \frac{\sqrt{\widetilde{S}_n^2}}{\sqrt{n}} \right), \end{aligned}$$

donc, une région de non-rejet peut être définie, pour un test de risque α , par :

$$\bar{W} = \left\{ x_1, \dots, x_n : \mu_0 - a \frac{\sqrt{\widetilde{s}_n^2}}{\sqrt{n}} \leq \bar{x}_n \leq \mu_0 + a \frac{\sqrt{\widetilde{s}_n^2}}{\sqrt{n}} \right\}.$$

Pour des hypothèses unilatérales, par exemple $\mathbf{H}_0 : \mu \leq \mu_0$ contre $\mathbf{H}_1 : \mu > \mu_0$, il est naturel de rejeter \mathbf{H}_0 lorsque \bar{x}_n est trop grand car cela reflète une moyenne μ élevée. Pour déterminer la valeur critique on se place en $\mu = \mu_0$ qui est la valeur la plus défavorable pour \mathbf{H}_0 . Comme :

$$\alpha = P_{\mu_0} \left(a < \frac{\bar{X}_n - \mu_0}{\sqrt{\widetilde{S}_n^2}/\sqrt{n}} \right),$$

on a pour région de rejet

$$W = \left\{ x_1, \dots, x_n : \bar{x}_n > \mu_0 + a \frac{\sqrt{\widetilde{s}_n^2}}{\sqrt{n}} \right\}.$$

Pour $\mathbf{H}_0 : \mu \geq \mu_0$ contre $\mathbf{H}_1 : \mu < \mu_0$, la région de rejet sera

$$W = \left\{ x_1, \dots, x_n : \bar{x}_n < \mu_0 - a \frac{\sqrt{\widetilde{s}_n^2}}{\sqrt{n}} \right\}.$$

Exemple 7.3.2 *Un fabricant de téléviseurs achète un certain composant électronique à un fournisseur. Un accord entre le fournisseur et le fabricant stipule que la durée de vie de ces composants doit être égale à 600 heures au moins. Le fabricant qui vient de recevoir un lot important de ce composant veut en vérifier la qualité. Il tire au hasard un échantillon de 16 pièces. Le test de durée de vie pour cet échantillon donne les résultats suivants :*

615	575	560	550	575	605	580	540
595	610	590	530	625	570	620	565

Le fabricant doit-il accepter le lot ? (On choisit un risque de première espèce égal à 5%) Quel est le risque de deuxième espèce ?

On admettra que la durée de vie de ces composants suit une loi normale.

– *Caractéristiques de l'échantillon : $\bar{x}_n = 581.60$ et $\sqrt{\widetilde{s}_n^2} = 27.90$.*

Les hypothèses à tester sont : $\mathbf{H}_0 : \mu = \mu_0 = 600$ heures contre $\mathbf{H}_1 : \mu < 600$ heures

– *La variable de décision est la moyenne de l'échantillon.*

La variable $\frac{\bar{X}_n - \mu_0}{\sqrt{\widetilde{S}_n^2}/\sqrt{n}}$ suit une loi de Student à $(n - 1) = 15$ degrés de liberté.

– *Calcul du seuil critique c . La règle de décision est la suivante :*

$$\bar{x}_n < c \text{ Refus de } \mathbf{H}_0$$

$$\bar{x}_n > c \text{ Acceptation de } \mathbf{H}_0,$$

$$P_{\mu_0}(\bar{X}_n < c) = 0.05,$$

$$P_{\mu_0}\left(\frac{\bar{X}_n - 600}{27.90/4} < \frac{c - 600}{27.90/4}\right) = 0.05,$$

d'où $\frac{c - 600}{27.90/4} = -1.753$ et $c = 587.77$. La valeur de la moyenne arithmétique donnée par l'échantillon étant égale à 581.60, on doit rejeter l'hypothèse \mathbf{H}_0 .

– Risque de deuxième espèce :

La forme de l'hypothèse \mathbf{H}_1 implique que l'on doit calculer le risque β pour toutes les valeurs de $\mu < 600$. On fait le calcul pour $\mu_1 = 575$ par exemple :

$$\begin{aligned}\beta &= P(\bar{X}_n > c | \mathbf{H}_1) \\ &= P\left(\frac{\bar{X}_n - 575}{27.90/4} > \frac{587.77 - 575}{27.90/4} | \mathbf{H}_1\right) \\ &= P\left(\frac{\bar{X}_n - 575}{27.90/4} > 1.83 | \mathbf{H}_1\right) \\ &\simeq 0.05.\end{aligned}$$

7.3.2 Tests sur la variance σ^2 d'une loi $\mathcal{N}(\mu, \sigma^2)$

1. Cas où μ est connu

On sait que la variance empirique corrigée S_n^2 est un estimateur sans biais de σ^2 t.q. : $\frac{nS_n^2}{\sigma^2} \sim \chi_{(n)}^2$. Ainsi pour une hypothèse nulle $\mathbf{H}_0 : \sigma^2 = \sigma_0^2$ contre $\mathbf{H}_1 : \sigma^2 \neq \sigma_0^2$, on a, sous \mathbf{H}_0 :

$$\frac{nS_n^2}{\sigma_0^2} \sim \chi_{(n)}^2.$$

Comme :

$$\begin{aligned}1 - \alpha &= P_{\sigma_0^2}\left(-a \leq \frac{nS_n^2}{\sigma_0^2} \leq a\right) \\ &= P_{\sigma_0^2}\left(-a \frac{\sigma_0^2}{n} \leq S_n^2 \leq a \frac{\sigma_0^2}{n}\right),\end{aligned}$$

donc, une région de non-rejet peut être définie, pour un test de risque α , par :

$$\bar{W} = \left\{x_1, \dots, x_n : -a \frac{\sigma_0^2}{n} \leq s_n^2 \leq a \frac{\sigma_0^2}{n}\right\}.$$

Pour des hypothèses unilatérales, par exemple $\mathbf{H}_0 : \sigma^2 \leq \sigma_0^2$ contre $\mathbf{H}_1 : \sigma^2 > \sigma_0^2$, on a pour région de rejet

$$W = \left\{x_1, \dots, x_n : s_n^2 > a \frac{\sigma_0^2}{n}\right\}.$$

Pour $\mathbf{H}_0 : \sigma^2 \geq \sigma_0^2$ contre $\mathbf{H}_1 : \sigma^2 < \sigma_0^2$, la région de rejet sera

$$W = \left\{x_1, \dots, x_n : s_n^2 < -a \frac{\sigma_0^2}{n}\right\}.$$

2. Cas où μ est inconnu

On sait que la variance empirique corrigée \widetilde{S}_n^2 est un estimateur sans biais de σ^2 t.q. : $\frac{(n-1)\widetilde{S}_n^2}{\sigma^2} \sim \chi_{(n-1)}^2$. Ainsi pour une hypothèse nulle $\mathbf{H}_0 : \sigma^2 = \sigma_0^2$ contre $\mathbf{H}_1 : \sigma^2 \neq \sigma_0^2$, on a, sous \mathbf{H}_0 :

$$\frac{(n-1)\widetilde{S}_n^2}{\sigma_0^2} \sim \chi_{(n-1)}^2.$$

Comme :

$$\begin{aligned} 1 - \alpha &= P_{\sigma_0^2} \left(-a \leq \frac{(n-1)\widetilde{S}_n^2}{\sigma_0^2} \leq a \right) \\ &= P_{\sigma_0^2} \left(-a \frac{\sigma_0^2}{n-1} \leq \widetilde{S}_n^2 \leq a \frac{\sigma_0^2}{n-1} \right), \end{aligned}$$

donc, une région de non-rejet peut être définie, pour un test de risque α , par :

$$\overline{W} = \left\{ x_1, \dots, x_n : -a \frac{\sigma_0^2}{n-1} \leq \widetilde{s}_n^2 \leq a \frac{\sigma_0^2}{n-1} \right\}.$$

Pour des hypothèses unilatérales, par exemple $\mathbf{H}_0 : \sigma^2 \leq \sigma_0^2$ contre $\mathbf{H}_1 : \sigma^2 > \sigma_0^2$, on a pour région de rejet

$$W = \left\{ x_1, \dots, x_n : \widetilde{s}_n^2 > a \frac{\sigma_0^2}{n-1} \right\}.$$

Pour $\mathbf{H}_0 : \sigma^2 \geq \sigma_0^2$ contre $\mathbf{H}_1 : \sigma^2 < \sigma_0^2$, la région de rejet sera

$$W = \left\{ x_1, \dots, x_n : \widetilde{s}_n^2 < -a \frac{\sigma_0^2}{n-1} \right\}.$$

7.3.3 Tests de comparaison des moyennes de deux lois de Gauss

On est en présence de deux échantillons indépendants, l'un de taille n_1 , de moyenne empirique \overline{X}_{n_1} et variance empirique $\widetilde{S}_{n_1}^2$, issu d'une loi $\mathcal{N}(\mu_1, \sigma_1^2)$, l'autre de taille n_2 , de moyenne empirique \overline{Y}_{n_2} et variance empirique $\widetilde{S}_{n_2}^2$, issu d'une loi $\mathcal{N}(\mu_2, \sigma_2^2)$.

1. Si σ_1^2 et σ_2^2 sont connus.

On sait que $\overline{X}_{n_1} - \overline{Y}_{n_2} \sim \mathcal{N}(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2})$ et la statistique $\frac{\overline{X}_{n_1} - \overline{Y}_{n_2} - \mu_1 + \mu_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$ de loi connue :

$\mathcal{N}(0, 1)$. Ainsi pour une hypothèse nulle $\mathbf{H}_0 : \mu_1 - \mu_2 = 0$ contre $\mathbf{H}_1 : \mu_1 - \mu_2 < 0$, on a, sous \mathbf{H}_0 :

$$\frac{\overline{X}_{n_1} - \overline{Y}_{n_2}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim \mathcal{N}(0, 1).$$

Comme :

$$\begin{aligned}
 1 - \alpha &= P \left(-a \leq \frac{\bar{X}_{n_1} - \bar{Y}_{n_2}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \leq a \right) \\
 &= P \left(-a \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \leq \bar{X}_{n_1} - \bar{Y}_{n_2} \leq a \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \right),
 \end{aligned}$$

donc, une région de non-rejet peut être définie, pour un test de risque α , par :

$$\bar{W} = \left\{ x_1, \dots, x_{n_1}, y_1, \dots, y_{n_2} : -a \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \leq \bar{x}_{n_1} - \bar{y}_{n_2} \leq a \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \right\}.$$

Exemple 7.3.3 On dispose de deux échantillons de tubes, construits suivant deux procédés de fabrication A et B. On a mesuré les diamètres de ces tubes et on a trouvé (en millimètres) :

Procédé A	51.90	50.90	52.80	52.90	53.40
Procédé B	51.10	51.30	51.50	52.10	

On suppose que les diamètres sont distribués suivant une loi normale et que les écarts-types sont égaux à : $\sigma_A = 1\text{mm}$ et $\sigma_B = 0.45\text{mm}$.

Peut-on affirmer au niveau 5% qu'il y a une différence significative entre les procédés de fabrication A et B ?

Soient X_A et Y_B , les variables aléatoires « diamètres des tubes fabriqués suivant les procédés A et B ».

On veut tester $\mathbf{H}_0 : \mu_1 - \mu_2 = 0$ contre $\mathbf{H}_1 : \mu_1 - \mu_2 < 0$ avec un seuil critique égal à 5%,

$$P \left(-1.96 \leq \frac{\bar{X}_{n_1} - \bar{Y}_{n_2}}{\sqrt{\frac{1}{5} + \frac{(0.45)^2}{4}}} \leq 1.96 \right) = 0.95,$$

avec les données apportées par les échantillons, on obtient pour la moyenne des deux échantillons $\bar{x}_{n_1} = 52.38$ et $\bar{y}_{n_2} = 51.50$. Comme $\frac{52.38 - 51.50}{\sqrt{\frac{1}{5} + \frac{(0.45)^2}{4}}} = 1.76$, on ne peut rejeter l'hypothèse d'égalité des moyennes.

2. Si σ_1^2 et σ_2^2 sont inconnus, mais $\sigma_1^2 = \sigma_2^2 = \sigma^2$ (la variance commune).

On sait que $\bar{X}_{n_1} - \bar{Y}_{n_2} \sim \mathcal{N}(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2})$ et la statistique $\frac{\bar{X}_{n_1} - \bar{Y}_{n_2} - \mu_1 + \mu_2}{\sqrt{\widetilde{S_{n_1, n_2}^2} \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} \sim \mathcal{T}(n_1 + n_2 - 2)$

où $\widetilde{S_{n_1, n_2}^2} = \frac{(n_1 - 1)S_{n_1}^2 + (n_2 - 1)S_{n_2}^2}{n_1 + n_2 - 2}$. Ainsi pour une hypothèse nulle $\mathbf{H}_0 : \mu_1 - \mu_2 = 0$ contre $\mathbf{H}_1 : \mu_1 - \mu_2 \neq 0$, on a, sous \mathbf{H}_0 :

$$\frac{\bar{X}_{n_1} - \bar{Y}_{n_2}}{\sqrt{\widetilde{S_{n_1, n_2}^2} \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} \sim \mathcal{T}(n_1 + n_2 - 2).$$

Comme :

$$\begin{aligned} 1 - \alpha &= P \left(-a \leq \frac{\bar{X}_{n_1} - \bar{Y}_{n_2}}{\sqrt{\widetilde{S_{n_1, n_2}^2} \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} \leq a \right) \\ &= P \left(-a \sqrt{\widetilde{S_{n_1, n_2}^2} \left(\frac{1}{n_1} + \frac{1}{n_2} \right)} \leq \bar{X}_{n_1} - \bar{Y}_{n_2} \leq a \sqrt{\widetilde{S_{n_1, n_2}^2} \left(\frac{1}{n_1} + \frac{1}{n_2} \right)} \right), \end{aligned}$$

donc, une région de non-rejet peut être définie, pour un test de risque α , par :

$$\bar{W} = \left\{ x_1, \dots, x_{n_1}, y_1, \dots, y_{n_2} : -a \sqrt{\widetilde{S_{n_1, n_2}^2} \left(\frac{1}{n_1} + \frac{1}{n_2} \right)} \leq \bar{x}_{n_1} - \bar{y}_{n_2} \leq a \sqrt{\widetilde{S_{n_1, n_2}^2} \left(\frac{1}{n_1} + \frac{1}{n_2} \right)} \right\}.$$

7.3.4 Tests de comparaison des variances de deux lois de Gauss

Avec les mêmes notations qu'à la sous-section précédente,

1. Si μ_1 et μ_2 sont connus.

On sait que $\frac{\frac{S_{n_1}^2}{\sigma_1^2}}{\frac{S_{n_2}^2}{\sigma_2^2}} \sim \mathcal{F}(n_1, n_2)$. Ainsi pour une hypothèse nulle $\mathbf{H}_0 : \frac{\sigma_1^2}{\sigma_2^2} = 1$ contre $\mathbf{H}_1 : \frac{\sigma_1^2}{\sigma_2^2} \neq 1$, on

a, sous \mathbf{H}_0 :

$$\frac{S_{n_1}^2}{S_{n_2}^2} \sim \mathcal{F}(n_1, n_2).$$

Comme :

$$1 - \alpha = P \left(a \leq \frac{S_{n_1}^2}{S_{n_2}^2} \leq b \right),$$

donc, une région de non-rejet peut être définie, pour un test de risque α , par :

$$\overline{W} = \left\{ x_1, \dots, x_{n_1}, y_1, \dots, y_{n_2} : a \leq \frac{s_{n_1}^2}{s_{n_2}^2} \leq b \right\}.$$

2. Si μ_1 et μ_2 sont inconnus.

On sait que $\frac{\frac{\widetilde{S}_{n_1}^2}{\sigma_1^2}}{\frac{\widetilde{S}_{n_2}^2}{\sigma_2^2}} \sim \mathcal{F}(n_1 - 1, n_2 - 1)$. Ainsi pour une hypothèse nulle $\mathbf{H}_0 : \frac{\sigma_1^2}{\sigma_2^2} = 1$ contre $\mathbf{H}_1 : \frac{\sigma_1^2}{\sigma_2^2} \neq 1$, on a, sous \mathbf{H}_0 :

$$\frac{\widetilde{S}_{n_1}^2}{\widetilde{S}_{n_2}^2} \sim \mathcal{F}(n_1 - 1, n_2 - 1).$$

Comme :

$$1 - \alpha = P \left(a \leq \frac{\widetilde{S}_{n_1}^2}{\widetilde{S}_{n_2}^2} \leq b \right),$$

donc, une région de non-rejet peut être définie, pour un test de risque α , par :

$$\overline{W} = \left\{ x_1, \dots, x_{n_1}, y_1, \dots, y_{n_2} : -a \leq \frac{\widetilde{s}_{n_1}^2}{\widetilde{s}_{n_2}^2} \leq b \right\}.$$

7.4 Exercices

Exercice 7.4.1 Pour déterminer si le contenu effectif de nicotine d'une certaine marque de cigarettes est plus élevé que ce qui est annoncé sur le paquet (1.4 mg), on procède à un test sur un échantillon de taille $n = 25$ cigarettes. Pour cet échantillon, on obtient $\bar{x}_n = 1.79$. On suppose que la variance d'une mesure vaut $\sigma^2 = 1$.

1. Formuler l'hypothèse nulle et l'hypothèse alternative dans le cas de cette étude.
2. Est-ce que le résultat est significatif au niveau de 5% ?
3. Est-ce que le résultat est significatif au niveau de 1% ?

Exercice 7.4.2 Soient 20 variables aléatoires X_1, \dots, X_{20} indépendantes et identiquement distribuées selon une loi $\mathcal{N}(\mu, \sigma^2)$. On veut tester l'hypothèse $\mathbf{H}_0 : \sigma^2 = \sigma_0^2$ et pour cela on utilise $\widehat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$.

1. Quelle est la distribution de $\frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \bar{X}_n)^2$ sous \mathbf{H}_0 ?
2. Calculer la valeur critique c_α pour un test unilatéral ($\mathbf{H}_1 : \sigma^2 > \sigma_0^2$) au seuil de 5%.

Exercice 7.4.3 On considère un échantillon de $n = 10$ variables aléatoires suivant une même loi normale $\mathcal{N}(\mu, \sigma^2)$.

1. On connaît la variance $\sigma^2 = 2.5$. On a observé la valeur 1.15 de la moyenne empirique \bar{X}_n . Peut-on accepter, au seuil de 95%, l'hypothèse $\mathbf{H}_0 : \mu = 0.1$?
2. En fait, on ne connaît pas la variance σ^2 , mais on observé la valeur $\tilde{s}_n^2 = 2.7$ de la variance empirique corrigée $\tilde{S}_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$. Peut-on accepter, l'hypothèse \mathbf{H}_0 ?

Exercice 7.4.4 La contenance (en grammes) d'une boîte de conserve suit une loi normale $\mathcal{N}(\mu, \sigma^2)$. On prélève au hasard n boîtes dont on note X_1, \dots, X_n les variables aléatoires représentant les contenances. On note x_1, \dots, x_n les valeurs observées de ces variables. A priori, on devrait avoir $\mu = 500$, mais on va faire différents tests, au seuil de risque de 0.05. On donne $n = 25$, $\sum_{k=1}^{25} x_k = 12482.5$.

1. La variance $\sigma^2 = 4$ est connue. Décrire le test de $\mathbf{H}_0 : \mu = 500$ contre l'hypothèse alternative $\mathbf{H}_1 : \mu < 500$.
2. En fait la variance est inconnue. Effectuer le test $\mathbf{H}_0 : \mu = 500$ contre $\mathbf{H}_1 : \mu < 500$ en sachant en plus que $\sum_{k=1}^{25} x_k^2 = 6232598.89$. Même question lorsque $\sum_{k=1}^{25} x_k^2 = 6232618.09$.
3. Tester $\mathbf{H}_0 : \sigma = 1.6$ contre $\mathbf{H}_1 : \sigma > 1.6$ sachant que $\sum_{k=1}^{25} x_k^2 = 6232598.89$, mais que l'on a aucune information sur μ .
4. Tester $\mathbf{H}_0 : \sigma = 1.6$ contre $\mathbf{H}_1 : \sigma > 1.6$ sachant que $\sum_{k=1}^{25} x_k^2 = 6232598.89$ et $\mu = 500$.

Solutions

Solution 7.4.1 1. Les hypothèses formulées sont $\mathbf{H}_0 : \mu = 1.4$ contre $\mathbf{H}_1 : \mu > 1.4$.

2. On définit la statistique de test : $U_n = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}$. La statistique U_n suit une loi normale centrée et réduite. La valeur observée u_{obs} de la statistique sur l'échantillon est $u_{obs} = \frac{1.79 - 1.4}{1/\sqrt{25}} = 1.95$. Région de rejet : $U_n > 1.96$ pour un risque 5%. Le quantile à 95% de la loi normale vaut environ 1.96 et est supérieur à la valeur observée. Donc, on ne rejette pas l'hypothèse nulle.

3. Le quantile à 99% vaut environ 2.49. Encore une fois, l'hypothèse nulle n'est pas rejetée.

Solution 7.4.2 1. Sous \mathbf{H}_0 , $\frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \bar{X}_n)^2$ suit une loi χ_{n-1}^2 à $(n-1)$ degrés de liberté.

2. On cherche c_α tel que : $P_{\mathbf{H}_0}(\hat{\sigma}^2 < c_\alpha) = 0.95$, et, en utilisant le point 1., on trouve,

$$\begin{aligned} P_{\mathbf{H}_0}(\hat{\sigma}^2 < c_\alpha) &= P_{\mathbf{H}_0} \left(\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2 < c_\alpha \right) \\ &= P_{\mathbf{H}_0} \left(\frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \bar{X}_n)^2 < \frac{n-1}{\sigma^2} c_\alpha \right) \\ &= P_{\mathbf{H}_0} \left(D < \frac{n-1}{\sigma^2} c_\alpha \right), \end{aligned}$$

avec $D \sim \chi_{n-1}^2$. À l'aide du quantile à 95% $\chi_{n-1;0.95}^2$, on détermine c_α : $c_\alpha = \frac{\chi_{n-1;0.95}^2}{n-1} \sigma_0^2 \simeq 1.59 \sigma_0^2$.

Solution 7.4.3 1. Sous l'hypothèse \mathbf{H}_0 , chacune des variables aléatoires X_k suit la loi normale $\mathcal{N}(0.1, 2.5)$.

Donc la moyenne \bar{X}_n suit la loi normale $\mathcal{N}(0.1, 0.25)$. Nous utilisons cette moyenne empirique comme variable de test. En l'absence d'autres informations, l'hypothèse alternative \mathbf{H}_1 est $\mu \neq 0.1$.

La v.a. $U_n = \frac{\bar{X}_n - 0.1}{\sqrt{2.5}/\sqrt{10}}$ suit la loi normale centrée réduite. Nous savons que $P(|U_n| < 1.96) = 0.95$.

Sous l'hypothèse \mathbf{H}_0 , la valeur observée de \bar{X}_n doit vérifier

$$P \left(\left| \frac{\bar{X}_n - 0.1}{\sqrt{2.5}/\sqrt{10}} \right| < 1.96 \right) = 0.95.$$

Cela nous donne comme zone d'acceptation de \mathbf{H}_0 l'intervalle $] -0.88, 1.08[$, qui ne contient pas la valeur 1.15 observée. Donc on doit rejeter \mathbf{H}_0 au seuil de 95%.

2. Sous l'hypothèse \mathbf{H}_0 , l'espérance μ est connue, mais la variance est inconnue. On utilise la variance empirique corrigée \tilde{S}_n^2 à la place de la variance σ^2 . La variable $U_n = \frac{\bar{X}_n - 0.1}{\sqrt{\tilde{S}_n^2}/\sqrt{10}}$ suit une loi de Student à $n-1 = 9$ degrés de liberté.

La table de Student nous dit que $P(|T_n| < 2.26) = 0.95$. Donc l'intervalle d'acceptation de \mathbf{H}_0 est défini par

$$P\left(\left|\frac{\bar{X}_n - 0.1}{\sqrt{\tilde{S}_n^2}/\sqrt{10}}\right| < 2.26\right) = 0.95.$$

ce qui nous donne $]-1.03, 1.23[$, l'intervalle qui contient la valeur observée. Au seuil de 95%, on peut maintenant accepter \mathbf{H}_0 .

Solution 7.4.4 1. Notons \bar{X}_n la moyenne empirique des variables contenances, qui est un estimateur sans biais et consistant de μ . Ce sera notre variable de test sur la moyenne μ . Son espérance est μ et son écart-type est $\frac{\sigma}{\sqrt{n}}$. La variable $U_n = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}$ suit une loi normale centrée réduite.

Dans le test $\mathbf{H}_0 : \mu = 500$ contre $\mathbf{H}_1 : \mu < 500$, on va rejeter $\mu = 500$ lorsque l'on observe des faibles valeurs de la contenance, puisque si μ diminue, alors les valeurs observées doivent diminuer. Donc, on fait un test unilatéral, dont la zone de rejet de \mathbf{H}_0 est à gauche. Au seuil de 95%, correspondant à la valeur $t = 1.65$ pour la loi normale centrée réduite, donc la zone de rejet de $\mu = 500$ est définie par la condition $\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} < -t$ soit $\bar{X}_n < -t\frac{\sigma}{\sqrt{n}} + \mu$.

Dans le cas présent, on observe $\bar{X}_n = 499.3$. La valeur limite $-t\frac{\sigma}{\sqrt{n}} + \mu$ est égale à 499.34. Donc on doit rejeter $\mu = 500$ et accepter $\mu < 500$.

2. Nous devons maintenant utiliser la variance empirique corrigée \tilde{S}_n^2 . Sous \mathbf{H}_0 , la variable $T_n = \frac{\bar{X}_n - \mu}{\sqrt{\tilde{S}_n^2}/\sqrt{n}}$ suit une loi de Student à $n - 1 = 24$ degrés de liberté.

Le principe du test est le même, la zone de rejet de \mathbf{H}_0 étant donnée par $\frac{\bar{X}_n - \mu}{\sqrt{\tilde{S}_n^2}/\sqrt{n}} < -t$ soit

$$\bar{X}_n < -t\frac{\sqrt{\tilde{S}_n^2}}{\sqrt{n}} + \mu, \text{ où } t \text{ se lit dans la table de Student, et vaut } t = 1.71.$$

– La valeur observée de la variance empirique corrigée est $\tilde{s}_n^2 = 1.9^2$. Donc la valeur minimale observée pour accepter \mathbf{H}_0 est $500 - 1.71\frac{1.9}{5} = 499.35$. On doit toujours rejeter \mathbf{H}_0 et accepter $\mu < 500$.

– Dans le second cas, la variance empirique corrigée observée est $\tilde{s}_n^2 = 2.1^2$, et la valeur limite est 499.28. Donc la valeur observée de la moyenne est dans la zone d'acceptation du test \mathbf{H}_0 contre \mathbf{H}_1 . On accepte $\mathbf{H}_0 : \mu = 500$.

3. La moyenne étant inconnue, on utilise la variance empirique corrigée \tilde{S}_n^2 dont la valeur observée est encore 1.9^2 . On utilise un χ^2 à $n - 1$ degrés de liberté : la variable de test $V_n = (n - 1) \frac{\tilde{S}_n^2}{\sigma^2}$ suit une loi du χ_{24}^2 à $n - 1 = 24$ degrés de liberté.

La zone de rejet de \mathbf{H}_0 est formée des grandes valeurs de V_n . La table nous donne $P(\chi_{24}^2 < 36.145) = 0.95$. On accepte \mathbf{H}_0 lorsque notre valeur observée $(n - 1) \frac{\tilde{S}_n^2}{\sigma^2}$ est inférieure à cette limite.

On a ici $(n - 1) \frac{\tilde{S}_n^2}{\sigma^2} = 24 \times \frac{1.9^2}{1.6^2} = 33.84$. Cette valeur reste dans la zone d'acceptation de \mathbf{H}_0 .

4. La moyenne étant connue, on utilise la variance empirique S_n^2 dont la valeur observée est 1.988^2 . On utilise un χ_{25}^2 à n degrés de liberté : la variable de test $W_n = n \frac{S_n^2}{\sigma^2}$ suit une loi du χ_{25}^2 à n degrés de liberté.

La zone de rejet de \mathbf{H}_0 est formée des grandes valeurs de W_n . La table nous donne $P(\chi_{25}^2 < 37.652) = 0.95$. On accepte \mathbf{H}_0 lorsque notre valeur observée $n \frac{S_n^2}{\sigma^2}$ est inférieure à cette limite.

On a ici $n \frac{S_n^2}{\sigma^2} = 25 \times \frac{1.988^2}{1.6^2} = 38.62$. Cette valeur est dans la zone de rejet de \mathbf{H}_0 . On doit accepter que $\sigma > 1.6$.

Annexe

Annexe : Tables statistiques usuelles

8.1 Loi normale

Table A-1 : Fonction de répartition de la loi normale centrée et réduite.

Table A-1		Table de la fonction de répartition de la loi normale centrée et réduite								
u	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,00	0,500000	0,503989	0,507978	0,511967	0,515953	0,519939	0,523922	0,527903	0,531881	0,535856
0,10	0,539828	0,543795	0,547758	0,551717	0,555670	0,559618	0,563559	0,567495	0,571424	0,575345
0,20	0,579260	0,583166	0,587064	0,590954	0,594835	0,598706	0,602568	0,606420	0,610261	0,614092
0,30	0,617911	0,621719	0,625516	0,629300	0,633072	0,636831	0,640576	0,644309	0,648027	0,651732
0,40	0,655422	0,659097	0,662757	0,666402	0,670031	0,673645	0,677242	0,680822	0,684386	0,687933
0,50	0,691462	0,694974	0,698468	0,701944	0,705402	0,708840	0,712260	0,715661	0,719043	0,722405
0,60	0,725747	0,729069	0,732371	0,735653	0,738914	0,742154	0,745373	0,748571	0,751748	0,754903
0,70	0,758036	0,761148	0,764238	0,767305	0,770350	0,773373	0,776373	0,779350	0,782305	0,785236
0,80	0,788145	0,791030	0,793892	0,796731	0,799546	0,802338	0,805106	0,807850	0,810570	0,813267
0,90	0,815940	0,818589	0,821214	0,823814	0,826391	0,828944	0,831472	0,833977	0,836457	0,838913
1,00	0,841345	0,843752	0,846136	0,848495	0,850830	0,853141	0,855428	0,857690	0,859929	0,862143
1,10	0,864334	0,866500	0,868643	0,870762	0,872857	0,874928	0,876976	0,878999	0,881000	0,882977
1,20	0,884930	0,886860	0,888767	0,890651	0,892512	0,894350	0,896165	0,897958	0,899727	0,901475
1,30	0,903199	0,904902	0,906582	0,908241	0,909877	0,911492	0,913085	0,914656	0,916207	0,917736
1,40	0,919243	0,920730	0,922196	0,923641	0,925066	0,926471	0,927855	0,929219	0,930563	0,931888
1,50	0,933193	0,934478	0,935744	0,936992	0,938220	0,939429	0,940620	0,941792	0,942947	0,944083
1,60	0,945201	0,946301	0,947384	0,948449	0,949497	0,950529	0,951543	0,952540	0,953521	0,954486
1,70	0,955435	0,956367	0,957284	0,958185	0,959071	0,959941	0,960796	0,961636	0,962462	0,963273
1,80	0,964070	0,964852	0,965621	0,966375	0,967116	0,967843	0,968557	0,969258	0,969946	0,970621
1,90	0,971284	0,971933	0,972571	0,973197	0,973810	0,974412	0,975002	0,975581	0,976148	0,976705
2,00	0,977250	0,977784	0,978308	0,978822	0,979325	0,979818	0,980301	0,980774	0,981237	0,981691
2,10	0,982136	0,982571	0,982997	0,983414	0,983823	0,984222	0,984614	0,984997	0,985371	0,985738
2,20	0,986097	0,986447	0,986791	0,987126	0,987455	0,987776	0,988089	0,988396	0,988696	0,988989
2,30	0,989276	0,989556	0,989830	0,990097	0,990358	0,990613	0,990863	0,991106	0,991344	0,991576
2,40	0,991802	0,992024	0,992240	0,992451	0,992656	0,992857	0,993053	0,993244	0,993431	0,993613
2,50	0,993790	0,993963	0,994132	0,994297	0,994457	0,994614	0,994766	0,994915	0,995060	0,995201
2,60	0,995339	0,995473	0,995603	0,995731	0,995855	0,995975	0,996093	0,996207	0,996319	0,996427
2,70	0,996533	0,996636	0,996736	0,996833	0,996928	0,997020	0,997110	0,997197	0,997282	0,997365
2,80	0,997445	0,997523	0,997599	0,997673	0,997744	0,997814	0,997882	0,997948	0,998012	0,998074
2,90	0,998134	0,998193	0,998250	0,998305	0,998359	0,998411	0,998462	0,998511	0,998559	0,998605
3,00	0,998650	0,998694	0,998736	0,998777	0,998817	0,998856	0,998893	0,998930	0,998965	0,998999
3,10	0,999032	0,999064	0,999096	0,999126	0,999155	0,999184	0,999211	0,999238	0,999264	0,999289
3,20	0,999313	0,999336	0,999359	0,999381	0,999402	0,999423	0,999443	0,999462	0,999481	0,999499
3,30	0,999517	0,999533	0,999550	0,999566	0,999581	0,999596	0,999610	0,999624	0,999638	0,999650
3,40	0,999663	0,999675	0,999687	0,999698	0,999709	0,999720	0,999730	0,999740	0,999749	0,999758
3,50	0,999767	0,999776	0,999784	0,999792	0,999800	0,999807	0,999815	0,999821	0,999828	0,999835
3,60	0,999841	0,999847	0,999853	0,999858	0,999864	0,999869	0,999874	0,999879	0,999883	0,999888
3,70	0,999892	0,999896	0,999900	0,999904	0,999908	0,999912	0,999915	0,999918	0,999922	0,999925
3,80	0,999928	0,999930	0,999933	0,999936	0,999938	0,999941	0,999943	0,999946	0,999948	0,999950
3,90	0,999952	0,999954	0,999956	0,999958	0,999959	0,999961	0,999963	0,999964	0,999966	0,999967

Table A-2 : Quantile de la loi normale centrée et réduite.

Table A-2		Table des quantiles de la loi normale centrée et réduite										
P	0,000	0,001	0,002	0,003	0,004	0,005	0,006	0,007	0,008	0,009	0,010	
0,00	∞	3,0902	2,8782	2,7478	2,6521	2,5758	2,5121	2,4573	2,4089	2,3656	2,3263	0,99
0,01	2,3263	2,2904	2,2571	2,2262	2,1973	2,1701	2,1444	2,1201	2,0969	2,0748	2,0537	0,98
0,02	2,0537	2,0335	2,0141	1,9954	1,9774	1,9600	1,9431	1,9268	1,9110	1,8957	1,8808	0,97
0,03	1,8808	1,8663	1,8522	1,8384	1,8250	1,8119	1,7991	1,7866	1,7744	1,7624	1,7507	0,96
0,04	1,7507	1,7392	1,7279	1,7169	1,7060	1,6954	1,6849	1,6747	1,6646	1,6546	1,6449	0,95
0,05	1,6449	1,6352	1,6258	1,6164	1,6072	1,5982	1,5893	1,5805	1,5718	1,5632	1,5548	0,94
0,06	1,5548	1,5464	1,5382	1,5301	1,5220	1,5141	1,5063	1,4985	1,4909	1,4833	1,4758	0,93
0,07	1,4758	1,4684	1,4611	1,4538	1,4466	1,4395	1,4325	1,4255	1,4187	1,4118	1,4051	0,92
0,08	1,4051	1,3984	1,3917	1,3852	1,3787	1,3722	1,3658	1,3595	1,3532	1,3469	1,3408	0,91
0,09	1,3408	1,3346	1,3285	1,3225	1,3165	1,3106	1,3047	1,2988	1,2930	1,2873	1,2816	0,90
0,10	1,2816	1,2759	1,2702	1,2646	1,2591	1,2536	1,2481	1,2426	1,2372	1,2319	1,2265	0,89
0,11	1,2265	1,2212	1,2160	1,2107	1,2055	1,2004	1,1952	1,1901	1,1850	1,1800	1,1750	0,88
0,12	1,1750	1,1700	1,1650	1,1601	1,1552	1,1503	1,1455	1,1407	1,1359	1,1311	1,1264	0,87
0,13	1,1264	1,1217	1,1170	1,1123	1,1077	1,1031	1,0985	1,0939	1,0893	1,0848	1,0803	0,86
0,14	1,0803	1,0758	1,0714	1,0669	1,0625	1,0581	1,0537	1,0494	1,0451	1,0407	1,0364	0,85
0,15	1,0364	1,0322	1,0279	1,0237	1,0194	1,0152	1,0110	1,0069	1,0027	0,9986	0,9945	0,84
0,16	0,9945	0,9904	0,9863	0,9822	0,9782	0,9741	0,9701	0,9661	0,9621	0,9581	0,9542	0,83
0,17	0,9542	0,9502	0,9463	0,9424	0,9385	0,9346	0,9307	0,9269	0,9230	0,9192	0,9154	0,82
0,18	0,9154	0,9116	0,9078	0,9040	0,9002	0,8965	0,8927	0,8890	0,8853	0,8816	0,8779	0,81
0,19	0,8779	0,8742	0,8706	0,8669	0,8632	0,8596	0,8560	0,8524	0,8488	0,8452	0,8416	0,80
0,20	0,8416	0,8381	0,8345	0,8310	0,8274	0,8239	0,8204	0,8169	0,8134	0,8099	0,8064	0,79
0,21	0,8064	0,8030	0,7995	0,7961	0,7926	0,7892	0,7858	0,7824	0,7790	0,7756	0,7722	0,78
0,22	0,7722	0,7688	0,7655	0,7621	0,7588	0,7554	0,7521	0,7488	0,7454	0,7421	0,7388	0,77
0,23	0,7388	0,7356	0,7323	0,7290	0,7257	0,7225	0,7192	0,7160	0,7128	0,7095	0,7063	0,76
0,24	0,7063	0,7031	0,6999	0,6967	0,6935	0,6903	0,6871	0,6840	0,6808	0,6776	0,6745	0,75
0,25	0,6745	0,6713	0,6682	0,6651	0,6620	0,6588	0,6557	0,6526	0,6495	0,6464	0,6433	0,74
0,26	0,6433	0,6403	0,6372	0,6341	0,6311	0,6280	0,6250	0,6219	0,6189	0,6158	0,6128	0,73
0,27	0,6128	0,6098	0,6068	0,6038	0,6008	0,5978	0,5948	0,5918	0,5888	0,5858	0,5828	0,72
0,28	0,5828	0,5799	0,5769	0,5740	0,5710	0,5681	0,5651	0,5622	0,5592	0,5563	0,5534	0,71
0,29	0,5534	0,5505	0,5476	0,5446	0,5417	0,5388	0,5359	0,5330	0,5302	0,5273	0,5244	0,70
0,30	0,5244	0,5215	0,5187	0,5158	0,5129	0,5101	0,5072	0,5044	0,5015	0,4987	0,4958	0,69
0,31	0,4958	0,4930	0,4902	0,4874	0,4845	0,4817	0,4789	0,4761	0,4733	0,4705	0,4677	0,68
0,32	0,4677	0,4649	0,4621	0,4593	0,4565	0,4538	0,4510	0,4482	0,4454	0,4427	0,4399	0,67
0,33	0,4399	0,4372	0,4344	0,4316	0,4289	0,4261	0,4234	0,4207	0,4179	0,4152	0,4125	0,66
0,34	0,4125	0,4097	0,4070	0,4043	0,4016	0,3989	0,3961	0,3934	0,3907	0,3880	0,3853	0,65
0,35	0,3853	0,3826	0,3799	0,3772	0,3745	0,3719	0,3692	0,3665	0,3638	0,3611	0,3585	0,64
0,36	0,3585	0,3558	0,3531	0,3505	0,3478	0,3451	0,3425	0,3398	0,3372	0,3345	0,3319	0,63
0,37	0,3319	0,3292	0,3266	0,3239	0,3213	0,3186	0,3160	0,3134	0,3107	0,3081	0,3055	0,62
0,38	0,3055	0,3029	0,3002	0,2976	0,2950	0,2924	0,2898	0,2871	0,2845	0,2819	0,2793	0,61
0,39	0,2793	0,2767	0,2741	0,2715	0,2689	0,2663	0,2637	0,2611	0,2585	0,2559	0,2533	0,60
0,40	0,2533	0,2508	0,2482	0,2456	0,2430	0,2404	0,2378	0,2353	0,2327	0,2301	0,2275	0,59
0,41	0,2275	0,2250	0,2224	0,2198	0,2173	0,2147	0,2121	0,2096	0,2070	0,2045	0,2019	0,58
0,42	0,2019	0,1993	0,1968	0,1942	0,1917	0,1891	0,1866	0,1840	0,1815	0,1789	0,1764	0,57
0,43	0,1764	0,1738	0,1713	0,1687	0,1662	0,1637	0,1611	0,1586	0,1560	0,1535	0,1510	0,56
0,44	0,1510	0,1484	0,1459	0,1434	0,1408	0,1383	0,1358	0,1332	0,1307	0,1282	0,1257	0,55
0,45	0,1257	0,1231	0,1206	0,1181	0,1156	0,1130	0,1105	0,1080	0,1055	0,1030	0,1004	0,54
0,46	0,1004	0,0979	0,0954	0,0929	0,0904	0,0878	0,0853	0,0828	0,0803	0,0778	0,0753	0,53
0,47	0,0753	0,0728	0,0702	0,0677	0,0652	0,0627	0,0602	0,0577	0,0552	0,0527	0,0502	0,52
0,48	0,0502	0,0476	0,0451	0,0426	0,0401	0,0376	0,0351	0,0326	0,0301	0,0276	0,0251	0,51
0,49	0,0251	0,0226	0,0201	0,0175	0,0150	0,0125	0,0100	0,0075	0,0050	0,0025	0,0000	0,50
	0,010	0,009	0,008	0,007	0,006	0,005	0,004	0,003	0,002	0,001	0,000	P

8.2 Loi du chi-deux

Table A-3 : Table de la loi du χ^2 , régions unilatérales

Table A-3			Table de la loi du χ^2 (n), régions unilatérales										
α n	0,001	0,005	0,01	0,025	0,05	0,1	0,5	0,9	0,95	0,975	0,99	0,995	0,999
1	0,00	0,00	0,00	0,00	0,00	0,02	0,45	2,71	3,84	5,02	6,63	7,88	10,83
2	0,00	0,01	0,02	0,05	0,10	0,21	1,39	4,61	5,99	7,38	9,21	10,60	13,82
3	0,02	0,07	0,11	0,22	0,35	0,58	2,37	6,25	7,81	9,35	11,34	12,84	16,27
4	0,09	0,21	0,30	0,48	0,71	1,06	3,36	7,78	9,49	11,14	13,28	14,86	18,47
5	0,21	0,41	0,55	0,83	1,15	1,61	4,35	9,24	11,07	12,83	15,09	16,75	20,51
6	0,38	0,68	0,87	1,24	1,64	2,20	5,35	10,64	12,59	14,45	16,81	18,55	22,46
7	0,60	0,99	1,24	1,69	2,17	2,83	6,35	12,02	14,07	16,01	18,48	20,28	24,32
8	0,86	1,34	1,65	2,18	2,73	3,49	7,34	13,36	15,51	17,53	20,09	21,95	26,12
9	1,15	1,73	2,09	2,70	3,33	4,17	8,34	14,68	16,92	19,02	21,67	23,59	27,88
10	1,48	2,16	2,56	3,25	3,94	4,87	9,34	15,99	18,31	20,48	23,21	25,19	29,59
11	1,83	2,60	3,05	3,82	4,57	5,58	10,34	17,28	19,68	21,92	24,73	26,76	31,26
12	2,21	3,07	3,57	4,40	5,23	6,30	11,34	18,55	21,03	23,34	26,22	28,30	32,91
13	2,62	3,57	4,11	5,01	5,89	7,04	12,34	19,81	22,36	24,74	27,69	29,82	34,53
14	3,04	4,07	4,66	5,63	6,57	7,79	13,34	21,06	23,68	26,12	29,14	31,32	36,12
15	3,48	4,60	5,23	6,26	7,26	8,55	14,34	22,31	25,00	27,49	30,58	32,80	37,70
16	3,94	5,14	5,81	6,91	7,96	9,31	15,34	23,54	26,30	28,85	32,00	34,27	39,25
17	4,42	5,70	6,41	7,56	8,67	10,09	16,34	24,77	27,59	30,19	33,41	35,72	40,79
18	4,90	6,26	7,01	8,23	9,39	10,86	17,34	25,99	28,87	31,53	34,81	37,16	42,31
19	5,41	6,84	7,63	8,91	10,12	11,65	18,34	27,20	30,14	32,85	36,19	38,58	43,82
20	5,92	7,43	8,26	9,59	10,85	12,44	19,34	28,41	31,41	34,17	37,57	40,00	45,31
21	6,45	8,03	8,90	10,28	11,59	13,24	20,34	29,62	32,67	35,48	38,93	41,40	46,80
22	6,98	8,64	9,54	10,98	12,34	14,04	21,34	30,81	33,92	36,78	40,29	42,80	48,27
23	7,53	9,26	10,20	11,69	13,09	14,85	22,34	32,01	35,17	38,08	41,64	44,18	49,73
24	8,08	9,89	10,86	12,40	13,85	15,66	23,34	33,20	36,42	39,36	42,98	45,56	51,18
25	8,65	10,52	11,52	13,12	14,61	16,47	24,34	34,38	37,65	40,65	44,31	46,93	52,62
26	9,22	11,16	12,20	13,84	15,38	17,29	25,34	35,56	38,89	41,92	45,64	48,29	54,05
27	9,80	11,81	12,88	14,57	16,15	18,11	26,34	36,74	40,11	43,19	46,96	49,65	55,48
28	10,39	12,46	13,56	15,31	16,93	18,94	27,34	37,92	41,34	44,46	48,28	50,99	56,89
29	10,99	13,12	14,26	16,05	17,71	19,77	28,34	39,09	42,56	45,72	49,59	52,34	58,30
30	11,59	13,79	14,95	16,79	18,49	20,60	29,34	40,26	43,77	46,98	50,89	53,67	59,70
31	12,20	14,46	15,66	17,54	19,28	21,43	30,34	41,42	44,99	48,23	52,19	55,00	61,10
32	12,81	15,13	16,36	18,29	20,07	22,27	31,34	42,58	46,19	49,48	53,49	56,33	62,49
33	13,43	15,82	17,07	19,05	20,87	23,11	32,34	43,75	47,40	50,73	54,78	57,65	63,87
34	14,06	16,50	17,79	19,81	21,66	23,95	33,34	44,90	48,60	51,97	56,06	58,96	65,25
35	14,69	17,19	18,51	20,57	22,47	24,80	34,34	46,06	49,80	53,20	57,34	60,27	66,62
36	15,32	17,89	19,23	21,34	23,27	25,64	35,34	47,21	51,00	54,44	58,62	61,58	67,98
37	15,97	18,59	19,96	22,11	24,07	26,49	36,34	48,36	52,19	55,67	59,89	62,88	69,35
38	16,61	19,29	20,69	22,88	24,88	27,34	37,34	49,51	53,38	56,90	61,16	64,18	70,70
39	17,26	20,00	21,43	23,65	25,70	28,20	38,34	50,66	54,57	58,12	62,43	65,48	72,06
40	17,92	20,71	22,16	24,43	26,51	29,05	39,34	51,81	55,76	59,34	63,69	66,77	73,40
50	24,67	27,99	29,71	32,36	34,76	37,69	49,33	63,17	67,50	71,42	76,15	79,49	86,66
60	31,74	35,53	37,48	40,48	43,19	46,46	59,33	74,40	79,08	83,30	88,38	91,95	99,61
70	39,04	43,28	45,44	48,76	51,74	55,33	69,33	85,53	90,53	95,02	100,43	104,21	112,32
80	46,52	51,17	53,54	57,15	60,39	64,28	79,33	96,58	101,88	106,63	112,33	116,32	124,84
90	54,16	59,20	61,75	65,65	69,13	73,29	89,33	107,57	113,15	118,14	124,12	128,30	137,21
100	61,92	67,33	70,06	74,22	77,93	82,36	99,33	118,50	124,34	129,56	135,81	140,17	149,45

8.3 Loi de Student

Table A-4 : Table des quantiles de la loi de Student

Table A-4		Table des quantiles de la loi de STUDENT									
α	0,6	0,7	0,8	0,9	0,95	0,975	0,99	0,995	0,999	0,9995	
v											
1	0,3249	0,7265	1,3764	3,0777	6,3137	12,7062	31,8210	63,6559	318,2888	636,5776	
2	0,2887	0,6172	1,0607	1,8856	2,9200	4,3027	6,9645	9,9250	22,3285	31,5998	
3	0,2767	0,5844	0,9785	1,6377	2,3534	3,1824	4,5407	5,8408	10,2143	12,9244	
4	0,2707	0,5686	0,9410	1,5332	2,1318	2,7765	3,7469	4,6041	7,1729	8,6101	
5	0,2672	0,5594	0,9195	1,4759	2,0150	2,5706	3,3649	4,0321	5,8935	6,8685	
6	0,2648	0,5534	0,9057	1,4398	1,9432	2,4469	3,1427	3,7074	5,2075	5,9587	
7	0,2632	0,5491	0,8960	1,4149	1,8946	2,3646	2,9979	3,4995	4,7853	5,4081	
8	0,2619	0,5459	0,8889	1,3968	1,8595	2,3060	2,8965	3,3554	4,5008	5,0414	
9	0,2610	0,5435	0,8834	1,3830	1,8331	2,2622	2,8214	3,2498	4,2969	4,7809	
10	0,2602	0,5415	0,8791	1,3722	1,8125	2,2281	2,7638	3,1693	4,1437	4,5868	
11	0,2596	0,5399	0,8755	1,3634	1,7959	2,2010	2,7181	3,1058	4,0248	4,4369	
12	0,2590	0,5386	0,8726	1,3562	1,7823	2,1788	2,6810	3,0545	3,9296	4,3178	
13	0,2586	0,5375	0,8702	1,3502	1,7709	2,1604	2,6503	3,0123	3,8520	4,2209	
14	0,2582	0,5366	0,8681	1,3450	1,7613	2,1448	2,6245	2,9768	3,7874	4,1403	
15	0,2579	0,5357	0,8662	1,3406	1,7531	2,1315	2,6025	2,9467	3,7329	4,0728	
16	0,2576	0,5350	0,8647	1,3368	1,7459	2,1199	2,5835	2,9208	3,6861	4,0149	
17	0,2573	0,5344	0,8633	1,3334	1,7396	2,1098	2,5669	2,8982	3,6458	3,9651	
18	0,2571	0,5338	0,8620	1,3304	1,7341	2,1009	2,5524	2,8784	3,6105	3,9217	
19	0,2569	0,5333	0,8610	1,3277	1,7291	2,0930	2,5395	2,8609	3,5793	3,8833	
20	0,2567	0,5329	0,8600	1,3253	1,7247	2,0860	2,5280	2,8453	3,5518	3,8496	
21	0,2566	0,5325	0,8591	1,3232	1,7207	2,0796	2,5176	2,8314	3,5271	3,8193	
22	0,2564	0,5321	0,8583	1,3212	1,7171	2,0739	2,5083	2,8188	3,5050	3,7922	
23	0,2563	0,5317	0,8575	1,3195	1,7139	2,0687	2,4999	2,8073	3,4850	3,7676	
24	0,2562	0,5314	0,8569	1,3178	1,7109	2,0639	2,4922	2,7970	3,4668	3,7454	
25	0,2561	0,5312	0,8562	1,3163	1,7081	2,0595	2,4851	2,7874	3,4502	3,7251	
26	0,2560	0,5309	0,8557	1,3150	1,7056	2,0555	2,4786	2,7787	3,4350	3,7067	
27	0,2559	0,5306	0,8551	1,3137	1,7033	2,0518	2,4727	2,7707	3,4210	3,6895	
28	0,2558	0,5304	0,8546	1,3125	1,7011	2,0484	2,4671	2,7633	3,4082	3,6739	
29	0,2557	0,5302	0,8542	1,3114	1,6991	2,0452	2,4620	2,7564	3,3963	3,6595	
30	0,2556	0,5300	0,8538	1,3104	1,6973	2,0423	2,4573	2,7500	3,3852	3,6460	
31	0,2555	0,5298	0,8534	1,3095	1,6955	2,0395	2,4528	2,7440	3,3749	3,6335	
32	0,2555	0,5297	0,8530	1,3086	1,6939	2,0369	2,4487	2,7385	3,3653	3,6218	
33	0,2554	0,5295	0,8526	1,3077	1,6924	2,0345	2,4448	2,7333	3,3563	3,6109	
34	0,2553	0,5294	0,8523	1,3070	1,6909	2,0322	2,4411	2,7284	3,3480	3,6007	
35	0,2553	0,5292	0,8520	1,3062	1,6896	2,0301	2,4377	2,7238	3,3400	3,5911	
36	0,2552	0,5291	0,8517	1,3055	1,6883	2,0281	2,4345	2,7195	3,3326	3,5821	
37	0,2552	0,5289	0,8514	1,3049	1,6871	2,0262	2,4314	2,7154	3,3256	3,5737	
38	0,2551	0,5288	0,8512	1,3042	1,6860	2,0244	2,4286	2,7116	3,3190	3,5657	
39	0,2551	0,5287	0,8509	1,3036	1,6849	2,0227	2,4258	2,7079	3,3127	3,5581	
40	0,2550	0,5286	0,8507	1,3031	1,6839	2,0211	2,4233	2,7045	3,3069	3,5510	
50	0,2547	0,5278	0,8489	1,2987	1,6759	2,0086	2,4033	2,6778	3,2614	3,4960	
60	0,2545	0,5272	0,8477	1,2958	1,6706	2,0003	2,3901	2,6603	3,2317	3,4602	
70	0,2543	0,5268	0,8468	1,2938	1,6669	1,9944	2,3808	2,6479	3,2108	3,4350	
80	0,2542	0,5265	0,8461	1,2922	1,6641	1,9901	2,3739	2,6387	3,1952	3,4164	
90	0,2541	0,5263	0,8456	1,2910	1,6620	1,9867	2,3685	2,6316	3,1832	3,4019	
100	0,2540	0,5261	0,8452	1,2901	1,6602	1,9840	2,3642	2,6259	3,1738	3,3905	
200	0,2537	0,5252	0,8434	1,2858	1,6525	1,9719	2,3451	2,6006	3,1315	3,3398	
500	0,2535	0,5247	0,8423	1,2832	1,6479	1,9647	2,3338	2,5857	3,1066	3,3101	
∞	0,2533	0,5244	0,8416	1,2816	1,6449	1,9600	2,3264	2,5758	3,0902	3,2905	

8.4 Loi de Fisher

Les tables A-5a à A-5b donnent les quantiles de la loi de Fisher

Table A-5a		Quantiles de la loi de FISHER $F(n,m)$											$\alpha= 5\%$
n	m	1	2	3	4	5	6	7	8	9	10	12	14
1	1	161,45	199,50	215,71	224,58	230,16	233,98	236,77	238,88	240,54	241,88	243,91	245,36
2	1	18,513	19,000	19,164	19,247	19,296	19,329	19,353	19,371	19,385	19,396	19,412	19,424
3	1	10,128	9,552	9,277	9,117	9,013	8,941	8,887	8,845	8,812	8,786	8,745	8,715
4	1	7,709	6,944	6,591	6,388	6,256	6,163	6,094	6,041	5,999	5,964	5,912	5,873
5	1	6,608	5,786	5,409	5,192	5,050	4,950	4,876	4,818	4,772	4,735	4,678	4,636
6	1	5,987	5,143	4,757	4,534	4,387	4,284	4,207	4,147	4,099	4,060	4,000	3,956
7	1	5,591	4,737	4,347	4,120	3,972	3,866	3,787	3,726	3,677	3,637	3,575	3,529
8	1	5,318	4,459	4,066	3,838	3,687	3,581	3,500	3,438	3,388	3,347	3,284	3,237
9	1	5,117	4,256	3,863	3,633	3,482	3,374	3,293	3,230	3,179	3,137	3,073	3,025
10	1	4,965	4,103	3,708	3,478	3,326	3,217	3,135	3,072	3,020	2,978	2,913	2,865
11	1	4,844	3,982	3,587	3,357	3,204	3,095	3,012	2,948	2,896	2,854	2,788	2,739
12	1	4,747	3,885	3,490	3,259	3,106	2,996	2,913	2,849	2,796	2,753	2,687	2,637
13	1	4,667	3,806	3,411	3,179	3,025	2,915	2,832	2,767	2,714	2,671	2,604	2,554
14	1	4,600	3,739	3,344	3,112	2,958	2,848	2,764	2,699	2,646	2,602	2,534	2,484
15	1	4,543	3,682	3,287	3,056	2,901	2,790	2,707	2,641	2,588	2,544	2,475	2,424
16	1	4,494	3,634	3,239	3,007	2,852	2,741	2,657	2,591	2,538	2,494	2,425	2,373
17	1	4,451	3,592	3,197	2,965	2,810	2,699	2,614	2,548	2,494	2,450	2,381	2,329
18	1	4,414	3,555	3,160	2,928	2,773	2,661	2,577	2,510	2,456	2,412	2,342	2,290
19	1	4,381	3,522	3,127	2,895	2,740	2,628	2,544	2,477	2,423	2,378	2,308	2,256
20	1	4,351	3,493	3,098	2,866	2,711	2,599	2,514	2,447	2,393	2,348	2,278	2,225
21	1	4,325	3,467	3,072	2,840	2,685	2,573	2,488	2,420	2,366	2,321	2,250	2,197
22	1	4,301	3,443	3,049	2,817	2,661	2,549	2,464	2,397	2,342	2,297	2,226	2,173
23	1	4,279	3,422	3,028	2,796	2,640	2,528	2,442	2,375	2,320	2,275	2,204	2,150
24	1	4,260	3,403	3,009	2,776	2,621	2,508	2,423	2,355	2,300	2,255	2,183	2,130
25	1	4,242	3,385	2,991	2,759	2,603	2,490	2,405	2,337	2,282	2,236	2,165	2,111
26	1	4,225	3,369	2,975	2,743	2,587	2,474	2,388	2,321	2,265	2,220	2,148	2,094
27	1	4,210	3,354	2,960	2,728	2,572	2,459	2,373	2,305	2,250	2,204	2,132	2,078
28	1	4,196	3,340	2,947	2,714	2,558	2,445	2,359	2,291	2,236	2,190	2,118	2,064
29	1	4,183	3,328	2,934	2,701	2,545	2,432	2,346	2,278	2,223	2,177	2,104	2,050
30	1	4,171	3,316	2,922	2,690	2,534	2,421	2,334	2,266	2,211	2,165	2,092	2,037
31	1	4,160	3,305	2,911	2,679	2,523	2,409	2,323	2,255	2,199	2,153	2,080	2,026
32	1	4,149	3,295	2,901	2,668	2,512	2,399	2,313	2,244	2,189	2,142	2,070	2,015
33	1	4,139	3,285	2,892	2,659	2,503	2,389	2,303	2,235	2,179	2,133	2,060	2,004
34	1	4,130	3,276	2,883	2,650	2,494	2,380	2,294	2,225	2,170	2,123	2,050	1,995
35	1	4,121	3,267	2,874	2,641	2,485	2,372	2,285	2,217	2,161	2,114	2,041	1,986
36	1	4,113	3,259	2,866	2,634	2,477	2,364	2,277	2,209	2,153	2,106	2,033	1,977
37	1	4,105	3,252	2,859	2,626	2,470	2,356	2,270	2,201	2,145	2,098	2,025	1,969
38	1	4,098	3,245	2,852	2,619	2,463	2,349	2,262	2,194	2,138	2,091	2,017	1,962
39	1	4,091	3,238	2,845	2,612	2,456	2,342	2,255	2,187	2,131	2,084	2,010	1,954
40	1	4,085	3,232	2,839	2,606	2,449	2,336	2,249	2,180	2,124	2,077	2,003	1,948
50	1	4,034	3,183	2,790	2,557	2,400	2,286	2,199	2,130	2,073	2,026	1,952	1,895
60	1	4,001	3,150	2,758	2,525	2,368	2,254	2,167	2,097	2,040	1,993	1,917	1,860
70	1	3,978	3,128	2,736	2,503	2,346	2,231	2,143	2,074	2,017	1,969	1,893	1,836
80	1	3,960	3,111	2,719	2,486	2,329	2,214	2,126	2,056	1,999	1,951	1,875	1,817
90	1	3,947	3,098	2,706	2,473	2,316	2,201	2,113	2,043	1,986	1,938	1,861	1,803
100	1	3,936	3,087	2,696	2,463	2,305	2,191	2,103	2,032	1,975	1,927	1,850	1,792
200	1	3,888	3,041	2,650	2,417	2,259	2,144	2,056	1,985	1,927	1,878	1,801	1,742
500	1	3,860	3,014	2,623	2,390	2,232	2,117	2,028	1,957	1,899	1,850	1,772	1,712
∞	1	3,841	2,996	2,605	2,372	2,214	2,099	2,010	1,938	1,880	1,831	1,752	1,692

Table A-5b		Quantiles de la loi de FISHER F(n,m) (suite)										$\alpha= 5\%$	
m	n	16	18	20	22	24	26	28	30	40	60	80	1000
1		246,46	247,32	248,02	248,58	249,05	249,45	249,80	250,10	251,14	252,20	252,72	254,19
2		19,433	19,440	19,446	19,450	19,454	19,457	19,460	19,462	19,471	19,479	19,483	19,495
3		8,692	8,675	8,660	8,648	8,639	8,630	8,623	8,617	8,594	8,572	8,561	8,529
4		5,844	5,821	5,803	5,787	5,774	5,763	5,754	5,746	5,717	5,688	5,673	5,632
5		4,604	4,579	4,558	4,541	4,527	4,515	4,505	4,496	4,464	4,431	4,415	4,369
6		3,922	3,896	3,874	3,856	3,841	3,829	3,818	3,808	3,774	3,740	3,722	3,673
7		3,494	3,467	3,445	3,426	3,410	3,397	3,386	3,376	3,340	3,304	3,286	3,234
8		3,202	3,173	3,150	3,131	3,115	3,102	3,090	3,079	3,043	3,005	2,986	2,932
9		2,989	2,960	2,936	2,917	2,900	2,886	2,874	2,864	2,826	2,787	2,768	2,712
10		2,828	2,798	2,774	2,754	2,737	2,723	2,710	2,700	2,661	2,621	2,601	2,543
11		2,701	2,671	2,646	2,626	2,609	2,594	2,582	2,570	2,531	2,490	2,469	2,410
12		2,599	2,568	2,544	2,523	2,505	2,491	2,478	2,466	2,426	2,384	2,363	2,302
13		2,515	2,484	2,459	2,438	2,420	2,405	2,392	2,380	2,339	2,297	2,275	2,212
14		2,445	2,413	2,388	2,367	2,349	2,333	2,320	2,308	2,266	2,223	2,201	2,136
15		2,385	2,353	2,328	2,306	2,288	2,272	2,259	2,247	2,204	2,160	2,137	2,072
16		2,333	2,302	2,276	2,254	2,235	2,220	2,206	2,194	2,151	2,106	2,083	2,016
17		2,289	2,257	2,230	2,208	2,190	2,174	2,160	2,148	2,104	2,058	2,035	1,967
18		2,250	2,217	2,191	2,168	2,150	2,134	2,119	2,107	2,063	2,017	1,993	1,923
19		2,215	2,182	2,155	2,133	2,114	2,098	2,084	2,071	2,026	1,980	1,955	1,884
20		2,184	2,151	2,124	2,102	2,082	2,066	2,052	2,039	1,994	1,946	1,922	1,850
21		2,156	2,123	2,096	2,073	2,054	2,037	2,023	2,010	1,965	1,916	1,891	1,818
22		2,131	2,098	2,071	2,048	2,028	2,012	1,997	1,984	1,938	1,889	1,864	1,790
23		2,109	2,075	2,048	2,025	2,005	1,988	1,973	1,961	1,914	1,865	1,839	1,764
24		2,088	2,054	2,027	2,003	1,984	1,967	1,952	1,939	1,892	1,842	1,816	1,740
25		2,069	2,035	2,007	1,984	1,964	1,947	1,932	1,919	1,872	1,822	1,796	1,718
26		2,052	2,018	1,990	1,966	1,946	1,929	1,914	1,901	1,853	1,803	1,776	1,698
27		2,036	2,002	1,974	1,950	1,930	1,913	1,898	1,884	1,836	1,785	1,758	1,679
28		2,021	1,987	1,959	1,935	1,915	1,897	1,882	1,869	1,820	1,769	1,742	1,662
29		2,007	1,973	1,945	1,921	1,901	1,883	1,868	1,854	1,806	1,754	1,726	1,645
30		1,995	1,960	1,932	1,908	1,887	1,870	1,854	1,841	1,792	1,740	1,712	1,630
31		1,983	1,948	1,920	1,896	1,875	1,857	1,842	1,828	1,779	1,726	1,699	1,616
32		1,972	1,937	1,908	1,884	1,864	1,846	1,830	1,817	1,767	1,714	1,686	1,602
33		1,961	1,926	1,898	1,873	1,853	1,835	1,819	1,806	1,756	1,702	1,674	1,589
34		1,952	1,917	1,888	1,863	1,843	1,825	1,809	1,795	1,745	1,691	1,663	1,577
35		1,942	1,907	1,878	1,854	1,833	1,815	1,799	1,786	1,735	1,681	1,652	1,566
36		1,934	1,899	1,870	1,845	1,824	1,806	1,790	1,776	1,726	1,671	1,643	1,555
37		1,926	1,890	1,861	1,837	1,816	1,798	1,782	1,768	1,717	1,662	1,633	1,545
38		1,918	1,883	1,853	1,829	1,808	1,790	1,774	1,760	1,708	1,653	1,624	1,536
39		1,911	1,875	1,846	1,821	1,800	1,782	1,766	1,752	1,700	1,645	1,616	1,526
40		1,904	1,868	1,839	1,814	1,793	1,775	1,759	1,744	1,693	1,637	1,608	1,517
50		1,850	1,814	1,784	1,759	1,737	1,718	1,702	1,687	1,634	1,576	1,544	1,448
60		1,815	1,778	1,748	1,722	1,700	1,681	1,664	1,649	1,594	1,534	1,502	1,399
70		1,790	1,753	1,722	1,696	1,674	1,654	1,637	1,622	1,566	1,505	1,471	1,364
80		1,772	1,734	1,703	1,677	1,654	1,634	1,617	1,602	1,545	1,482	1,448	1,336
90		1,757	1,720	1,688	1,662	1,639	1,619	1,601	1,586	1,528	1,465	1,429	1,314
100		1,746	1,708	1,676	1,650	1,627	1,607	1,589	1,573	1,515	1,450	1,415	1,296
200		1,694	1,656	1,623	1,596	1,572	1,551	1,533	1,516	1,455	1,386	1,346	1,205
500		1,664	1,625	1,592	1,563	1,539	1,518	1,499	1,482	1,419	1,345	1,303	1,138
∞		1,644	1,604	1,571	1,542	1,517	1,496	1,476	1,459	1,394	1,318	1,274	1,075

Bibliographie

- [1] Berge. C. Espaces topologiques-fonctions multivoques. Dunod, 1959.
- [2] Billingsley. P. Probability and measure. Wiley Series in Probability and Mathematical Statistics (2nd ed.). Wiley, 1986.
- [3] Bourbaki. N. Éléments de mathématique. Topologie générale : Chapitres 1 à 4. Springer-Verlag Berlin Heidelberg, 2007.
- [4] Fourdrinier. D. Statistique inférentielle : cours et exercices corrigés. Dunod, Paris, 2002.
- [5] Halmos. P.R. Measure theory. Springer-Verlag New York, 2^e édition, 1950.
- [6] Hurlin. C., Mignon. V. Statistique et probabilités en économie-gestion. Dunod, 2015.
- [7] John A.R. Mathematical statistics and data analysis. Duxbury Press, Belmont (CA), 2^e édition, 1995.
- [8] Lefebvre. M. Probabilités, statistique et applications. Presses Internationales Polytechnique, 2011.
- [9] Lindgren. B.W. Statistical theory. 3^e édition, Macmillan, New York, 1976.
- [10] Michel. L. Statistique : la théorie et ses applications. Springer-Verlag France, Paris, 2010.
- [11] Michel. J. P. Statistical inference : a short course. Wiley, 2012.
- [12] Van der Vaart. A.W. Asymptotic statistics. Cambridge University Press, 1998.
- [13] Verlant. B. Statistique et probabilités BTS industriel groupements. Foucher, T2, 2009.
- [14] Veysseyre. R. Aide-mémoire : Statistique et probabilités pour l'ingénieur. Dunod, Paris, 2001, 2006.
- [15] Randé. B. Procédés sommatoires-développements asymptotiques. Techniques de l'ingénieur, 2004.
- [16] Resnick. S.I. A probability path. Springer Science+Business Media New York, 2014.
- [17] Rudin. W. Principles of mathematical analysis. New York : McGraw-Hil, 1976.
- [18] Sahu. P.K., Pal. S.R., Das. A.K. Estimation and inferential statistics. Springer, 2015.

- [19] Saporta. G. Probabilités. Analyse des données et statistique. Technip. 2006.
- [20] Shao. J. Mathematical statistics : exercises and solutions. Springer Science+Business Media, Inc. 2005.
- [21] Stephan. M. Introduction à la statistique. Presses Polytechniques et Universitaires Romandes, Lausanne, 2^e édition, 2001.
- [22] Suhov. Y., Kelbert. M. Probability and statistics by example. Cambridge University Press, 2005.
- [23] Yadolah. D. Premiers pas en statistique. Springer, Paris, 2003.
- [24] Yves. T. Théorie des sondages : échantillonnage et estimation en population finie : cours et exercices avec solutions. Dunod, Paris, 2001.

- Estimation :

- Estimation Ponctuelle,

- Estimation par Intervalle de Confiance,

- Tests d'Hypothèses.

