République Algérienne Démocratique et Populaire
Ministère de l'enseignement supérieur et de la recherche scientifique

Université de Saida - Dr Moulay Tahar.
Faculté des Sciences.
Département de Mathématiques.

Polycopié de cours

Statistique Paramétrique

Dr Ait ouali **N**adia [1]

Année universitaire : 2022/2023

1. e-mail : aitouali.nadai@gmail.com

Avant-Propos

Ce polycopié de cours s'adresse aux étudiants Master 1 spécialité analyse stochastique, statistique des processus et ses applications (ASSPA) mais peut être utile à tous ceux qui seraient désireux d'acquérir ou de revoir les principales méthodes de la statistique paramétrique.

Pour compléter notre polycopié, je me suis appuyé sur plusieurs références différentes. J'ai également présenté les chapitres de manière détaillée et claire, en donnant les preuves, exemples et illustrations nécessaires pour permettre aux étudiants de comprendre facilement le contenu.

Ce cours est subdivisé en deux parties : la première partie est destinée à introduire les bases probabilistes nécessaires à une compréhension minimale des démarches de statistiques paramétrique.

La deuxième partie quant à elle est consacrée aux principaux concepts et approches de l'inférence statistique et comporte trois chapitres. Le premier chapitre introduit la théorie d'échantillonnage, le deuxième traite de l'estimation, le troisième introduit les tests d'hypothèse.

Enfin, toutes vos remarques, vos commentaires, vos critiques, seront accueillis avec plaisir. Vous pouvez me les communiquer à l'adresse électronique suivante : nadia.aitouali@univ-saida.dz

Table des matières

Première partie

Rappels et compléments

Chapitre 1

Généralités

Dans ce chapitre, commençons par quelques rappels et notions de probabilité qui sont trés importantes pour l'étude de la convergence de variables aléatoires . Il s'agit de généralités que nous énonçons sans démonstration .

1.1 Probabilité

1.1.1 Espace d'évènement

Définition 1.1.1.1. *L'ensemble de tous les résultats possible d'une épreuve est dite espace des évènements on le note généralement par* Ω .

1.1.2 Tribu

Définition 1.1.2.1. *On appelle Tribu (σ-algèbre) sur un espace d'évènement* Ω *tout famille qu'on notra* T *vérifiant :*

1. $\Omega \in T$.
2. $\forall A \in T$ *on a* $A^c \in T$ *(T est stable par complémentaire) .*
3. $\forall (A_i)_{i\in I} \in T$ *on a* $\bigcup_{i=0}^{\infty} A_i \in T$ *(T est stable par réunion dénombrable) .*

Remarque 1.1.2.1. *Le couple* (Ω, T) *s'appelle espace probabilisable.*

1.1.3 Espace Probabilisé

Définition 1.1.3.1. *Soit* (Ω, T) *un espace probabilisable, on appelle probabilité sur* (Ω, T) *toute application* $\mathbb{P}$ *de* (Ω, T) *dans l'intervalle* $[0, 1]$ *telle que :*

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

2. $\forall (A_i)_{i\in\mathbb{N}} \in T \ A_i \bigcap A_j = \phi$*(évènement incompatible) .*

3. On a $\mathbb{P}(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} \mathbb{P}(A_i)$ *.*

Définition 1.1.3.2. *On appelle espace de probabilité le triplet* $(\Omega, T, \mathbb{P})$ *où* $\mathbb{P}$ *disigne une probabilité définie sur une tribu* T *de parties de* Ω *.*

1.2 Variables aléatoires

Définition 1.2.0.1. *On appelle variable aléatoire sur l'espace* $(\Omega, T, \mathbb{P})$ *de* $(\Omega, T, \mathbb{P})$ *dans* $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$

$$\begin{array}{rccc} X & : & (\Omega; T; \mathbb{P}) & \to & (\mathbb{R}, \mathcal{B}(\mathbb{R})) \\ & & w & \mapsto & X(w) \end{array}$$

A chaque évènement élémentaire w *de* Ω *correspond un nombre réel* X *associe à la variable aléatoire* X *. La valeur* X *correspond à la réalisation de la variable* X *pour l'évènement élémentaire* w *.*

1.3 Variable aléatoire discrète

Définition 1.3.0.1. *On dit qu'une variable aléatoire* X *définie sur l'espace* $(\Omega, T, \mathbb{P})$ *est discrète lorsque* $X(\Omega)$ *est finie où dénombrable .*

1.3.1 Fonction de répartition

Définition 1.3.1.1. *On appelle fonction de répartition d'une variable aléatoire* x *la fonction* F_X *telle que :*

$$\begin{array}{rccc} F_X & : & \mathbb{R} & \to & \mathbb{R} \\ & & x & \mapsto & F_X(x) = \mathbb{P}(X \leq x) \end{array}$$

1.3.2 Caractéristiques numériques

• Espérance mathématique :

Définition 1.3.2.1. *Soit* X *une variable aléatoire réelle ; l'espérance mathématique de* X *définie par :*
Si X *est une v.a.r discrète finie où dénombrable :* $E(X) = \sum_{x\in\mathbb{R}} x\mathbb{P}(X = x)$

• **Variance :**

Définition 1.3.2.2. *Soit X une variable aléatoire réelle ayant une espérance $E(X)$. On appelle variance de X le réelle .*

$$Var(X) = E[X - E(X)]^2$$

où bien :

$$Var(X) = \sigma^2(X)$$

Remarque 1.3.2.1. *Si X une variable aléatoire ayant une variance $Var(X)$, on appelle écart type de X le réel :*

$$\sigma(X) = \sqrt{Var(X)}.$$

1.3.3 Lois discrète classiques

• **Loi de Bernoulli :**

Définition 1.3.3.1. *La variable aléatoire X suit la loi Brnoulli de probabilité $\mathbb{P}$ si X vaut 1 où 0 avec les probabilités respectives p et 1-p, on a alors :*
la fonction de masse $\mathbb{P}(X)$ de X est donnée par :

$$\mathbb{P}(X = k) = \begin{cases} P^k(1-P)^{1-k} & si \qquad k \in [0,1] \\ 0 & si\ non. \end{cases}$$

on notera $X \rightsquigarrow B(p)$
Donc : $E(X) = p$, $V(X) = p.q$ et $\sigma(X) = \sqrt{p.q}$

• **Loi Binomiale :**

Définition 1.3.3.2. *Si X est une variable aléatoire suit la loi Binomiale de paramètre $n \in \mathbb{N}^*$ et $p \in [0,1]$ on a alors : la fonction de masse de X est donnée par :*

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

On netera $X \rightsquigarrow B(n,p)$. Donc :
$E(X) = n.p$ et
$Var(X) = n.p.q$

• **Loi de Poison :**

Définition 1.3.3.3. *Une variable aléatoire X suit une loi de Poisson de paramètre λ, lorsque son support D est donné par $D = \mathbb{N}$ et sa fonction de*

masse par :

$$\mathbb{P}_X(x) = \begin{cases} \frac{e^{-\lambda}.\lambda^x}{x!} & si\ x \in \mathbb{N} \\ 0 & si\ non \end{cases}$$

On notera $X \rightsquigarrow \mathbb{P}(\lambda)$ *Donc :*
$E(X) = \lambda$ *et* $Var(X) = \lambda$

1.4 Variable aléatoire absolument continue

Définition 1.4.0.1. *Une variable aléatoire X est absolument continue s'il existe $f(x)$ telle que sa fonction de répartition $F(x)$ est égale à*

$$\mathbb{P}(X \leq x) = F(x) = \int_{-\infty}^{x} f(x)dx$$

.

1.4.1 Fonction de densité

Définition 1.4.1.1. *On appelle densité de probabilité toute application continue par morceaux*

$$\begin{array}{rccc} f & : & \mathbb{R} & \to & \mathbb{R} \\ & & x & \mapsto & f(x) \end{array}$$

telle que :

1. $\forall x \in \mathbb{R}\ f(x) \geq 0$
2. $\int_{-\infty}^{+\infty} f(x)dx = 1$

1.4.2 Fonction de répartition

Définition 1.4.2.1. *Soit X une variable aléatoire absolument continue, on définit la fonction de répartition de X :*

$$\begin{array}{rccc} F_X & : & \mathbb{R} & \to & \mathbb{R} \\ & & x & \mapsto & F_X(x) = \mathbb{P}(X \leq x) \end{array}$$

Alors la relation entre la fonction de répartition $F(x)$ et la fonction densité de probabilité $f(x)$ est la suivante :

$$\forall x \in \mathbb{R}\ \ F_X(x) = \mathbb{P}(X \leq x) = \int_{-\infty}^{x} f(x)dx$$

ie :

$$f_X(x) = F'_X(x)$$

1.4.3 Caractéristiques numériques

• Espérance mathématique :

Définition 1.4.3.1. *Si X est une variable aléatoire absolument continue de densité f . On appelle espérance de X, le réel $E(X)$ définie par :*

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

• Variance :

Définition 1.4.3.2. *Si X est une variable aléatoire continue donne par sa densité de probabilité alors variance de X est le nombre réel positif telle que :*

$$Var(X) = \int_{\mathbb{R}} (X - E(X))^2 f(x)dx$$

$$= \int_{\mathbb{R}} x^2 f(x)dx - [E(X)]^2$$

1.4.4 Lois continues classiques

• loi uniforme :

Définition 1.4.4.1. *La variable aléatoire X suit une loi uniforme sur le segment $[a, b]$ avec $a < b$. Si sa densité de probabilité est donnée par :*

$$f(x) = \begin{cases} \frac{1}{b-a} & si\ x \in [a, b] \\ o & si\ non \end{cases}$$

donc :
$E(X) = \frac{b+a}{2}$ *et* $Var(X) = \frac{(b-a)^2}{12}$

• Loi normale (loi de Laplace Gauss) :

Définition 1.4.4.2. *Une variable aléatoire absolument continue X suit une loi normale de paramètre (μ, σ) si sa densité de probabilité est donnée par :*

$$\begin{array}{rccl} f & : & \mathbb{R} & \to & \mathbb{R} \\ & & x & \mapsto & f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2} \end{array}$$

donc :
$E(X) = \mu$ *et* $Var(X) = \sigma^2$

1.4.5 Fonction caractéristique

Définition 1.4.5.1. *On appelle fonction caractéristique de X la fonction φ_X de la variable réelle t définie par :*

$$\varphi_X(t) = E(e^{itX})$$

Expression dans le cas discrèt est :

$$\varphi_X(t) = \sum_{x \in E} e^{itX} \mathbb{P}(X = x).$$

Expression dans le cas absolument continu est :

$$\varphi_X(t) = \int_{\mathbb{R}} e^{itX} f_X(x) dx.$$

Tableau des fonctions caractéristique des lois usuelles :

Loi	Fonction caractéristique
Bernoulli	$pe^{it} + (1-p)$
Binomiale B(n,p)	$(pe^{it} + (1-p))^n$
Poisson $p(\lambda)$	$e^{\lambda(e^{it}-1)}$
Loi normale $\mathcal{N}(\mu, \sigma)$	$e^{it\mu - \frac{t^2\sigma^2}{2}}$
Loi normale $\mathcal{N}(0,1)$	$e^{\frac{-t^2}{2}}$

Chapitre 2

Modes de convergence

Ce chapitre est pour donner la notion de convergence de variables aléatoires . Soit $X_1, X_2, ..., X_n$ une suite de variables aléatoires . Que peut signifier l'expression "la suite X_n converge quand $n \to +\infty$"?, il n'existe pas une seule réponse à cette question . Nous allons définir quelques réponses possibles . En suite nous énonçons les théorèmes limites qui sont la base de la théorie des probabilités et statistiques.

2.1 Les modes de convergence aléatoires

Une suite $(X_n)_{n\in\mathbb{N}^*}$ de variables aléatoires réelles est une suite de fonctions mesurables de Ω dans $\mathbb{R}$. Il existe diverses façons de définir la convergence de (X_n). On considère dans la suite, une suite (X_n) de variables aléatoires réelles et une variable aléatoire X définie sur le même espace probabilisé $(\Omega, T, \mathbb{P})$.

Définition 2.1.0.1. *On dit que* $\{X_n\}$ ***converge en probabilité*** *(ou converge faiblement) vers la v.a. X si, quel que soit*

$$\lim_{n\to+\infty} \mathbb{P}(|X_n - X| < \varepsilon) = 1 \text{ ou } (\lim_{n\to+\infty} \mathbb{P}(|X_n - X| \geq \varepsilon) = 0$$

On note $X_n \xrightarrow{\mathbb{P}} X$

Exemple 2.1.0.1. *Soit* (X_n) *une suite de v.a. On suppose que* $X_n(\Omega) = \{-1, 0, 1\}$ *avec*

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

Cette suite converge en probabilité vers 0, en effet :

$$\forall \varepsilon > 0, \mathbb{P}(|X_n - 0| > \varepsilon) = \mathbb{P}(|X_n| > \varepsilon) = \mathbb{P}(|X_n| = 1) = \frac{1}{2n} \underset{n\to+\infty}{\longrightarrow} 0$$

2.1.1 Inégalité de Markov

Soit X une variable aléatoire réelle g une fonction croissante et positive sur l'ensemble des réels, vérifiant . Alors

$$\forall a > 0, \mathbb{P}(|X| \geq a) \leq \frac{E(g(X))}{g(a)}$$

Démonstration 2.1.1.1.

$$\begin{aligned} E(g(X)) &= \int_{\Omega} g(x)f(x)dx = \int_{|X|<a} g(x)f(x)dx + \int_{|X|\geq a} g(x)f(x)dx \\ &\geq \int_{|X|\geq a} g(x)f(x)dx \ car\ g\ est\ positive \\ &\geq g(x)\int_{|X|\geq a} f(x)dx \ car\ g\ est\ croissante \\ &\geq g(a)\mathbb{P}(|X| \geq a) \end{aligned}$$

Donc

$$g(a)\mathbb{P}(|X| \geq a) \leq E(g(X)) \Leftrightarrow \mathbb{P}(|X| \geq a) \leq \tfrac{E(g(X))}{g(a)},\ puisque\ g(a) > 0$$

2.1.2 l'inégalité de Bienaymé-Tchebychev

Soit X une variable aléatoire telle que $E(X^2)$. Alors pour tout réel $\varepsilon > 0$,

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

Démonstration 2.1.2.1. *On a* $Y = |X - E(X)|, a = \varepsilon,$ *et* $g(a) = \varepsilon^2$ *dans l'inégalité de Markov.*

Théorème 2.1.2.1. *Soit* (X_n) *une suite de variables aléatoires sur le même espace de probabilisé* $(\Omega, T, \mathbb{P})$ *admettant des espérances et des variances vérifiant*

$$\lim_{n\to+\infty} E(X_n) = l \quad \lim_{n\to+\infty} V(X_n) = 0$$

alors les (X_n) *convergent en probabilité vers l.*

Démonstration 2.1.2.2. *Soit* $\varepsilon > 0$. *Posons* $E(X_n) = l + u_n$ *un avec* $\lim_{n\to\infty} u_n = 0$. *Alors il existe* $N \in \mathbb{N}$ *tel que :*

$$n \geq N \Rightarrow |u_n| < \varepsilon/2$$

et donc à partir du rang N,

$$|X_n - E(X_n)| < \varepsilon/2 \Rightarrow |X_n - l| < \varepsilon \tag{2.1}$$

car $|X_n - l| = |X_n - E(X_n) + E(X_n) - l| \leq |X_n - E(X_n)| + |E(X_n) - l|$
L'implication (2.1) peut être encore écrite sous la forme

$$|X_n - l| \geq \varepsilon \Rightarrow |X_n - E(X_n)| \geq \varepsilon/2$$

Par conséquent, en utilisant l'inégalité de Bienaymé-Chebyshev ,

$$\mathbb{P}(|X_n - l| \geq \varepsilon) \leq \mathbb{P}(|X_n - E(X_n)| \geq \varepsilon/2) \leq \frac{V(X_n)}{(\varepsilon/2)^2}$$

qui tend vers 0 quand n tend vers l'infini.

Conséquence : *Pour que* (X_n) *converge en probabilité vers* X *, il suffit que* $E(X_n - X) \to 0$ *et* $V(X_n - X) \to 0$ *lorsque* $n \to \infty$

2.1.3 Loi faible des grands nombres

Théorème 2.1.3.1. *Soit* $(X_i)_{i\geq 1}$ *une suite de variables aléatoires définies sur l'espace de probabilité* $(\Omega, T, \mathbb{P})$ *, deux à deux indépendantes et de même loi telle que* $E(X_i) = \mu$ *et* $V(X_i) = \sigma^2$*. Alors la variable aléatoire :*

$$S_n = \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{\mathbb{P}} \mu$$

i.e $:\forall \varepsilon > 0, \mathbb{P}(|S_n - \mu| > \varepsilon) \to 0$ *quand* $n \to +\infty$

Démonstration 2.1.3.1. *Soit* $(X_n)_{n\geq 1}$ *une suite de variables aléatoires deux à deux indépendantes et telles que* $E(X_n) = \mu$ *et* $Var(X_n) = \sigma^2$*. La suite* $(S_n)_{n\geq 1}$ *où* $S_n = \frac{1}{n}\sum_{i=1}^{n} X_i$*. Alors on a :*

$$E(S_n) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \mu \text{ et } Var(S_n) = \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i) = \frac{\sigma^2}{n}$$

Donc, d'après la suite $(S_n)_{n\geq 1}$ *converge en probabilité vers* μ

2.1.4 Convergence presque sûre (Convergence forte)

Définition 2.1.4.1. *On dit que* $\{X_n\}$ ***converge presque sûrement (ou converge avec probabilité 1, ou converge fortement)*** *vers la v.a. X si,*

$$\mathbb{P}(\lim_{n\to+\infty} X_n = X) = 1 \text{ ou bien } \mathbb{P}(\lim_{n\to+\infty} X_n \neq X) = 0$$

et l'on note $X_n \xrightarrow{p.s} X$
En d'autres termes, l'ensemble des points de divergence est de probabilité nulle.

$$\mathbb{P}\left[w \in \Omega : \lim_{n\to+\infty} X_n(w) = X(w)\right] = 1$$

$$\begin{aligned} X_n \xrightarrow{p.s} X &\Leftrightarrow \forall \varepsilon > 0, \lim_{n\to+\infty} \mathbb{P}(\sup_{m\geq n} |X_m - X| > \varepsilon) = 0 \\ &\Leftrightarrow \forall \varepsilon > 0, \lim_{n\to+\infty} \mathbb{P}(\bigcup_{m\geq n} |X_m - X| > \varepsilon) = 0 \\ &\Leftrightarrow \forall \varepsilon > 0, \lim_{n\to+\infty} \mathbb{P}(\bigcap_{m\geq n} |X_m - X| < \varepsilon) = 1 \end{aligned}$$

On a les évènements $\bigcup_{m\geq n} (|X_m - X| > \varepsilon)$ *et* $(\sup_{m\geq n} |X_n - X| > \varepsilon)$ *sont équivalents et ont comme complémentaire :*

$$\bigcap_{m\geq n} (|X_n - X| < \varepsilon)$$

d'où l'équivalence des 3 conditions :

1. *la première condition signifie que :*

$$\sup_{m\geq n} |X_n - X| \xrightarrow{\mathbb{P}} 0 \Rightarrow |X_n - X| \xrightarrow{\mathbb{P}} 0$$

2. *la convergence P.S est plus forte que la convergence en probabilité :*

$$X_n \xrightarrow{p.s} X \Rightarrow X_n \xrightarrow{\mathbb{P}} X$$

3. *la dernière condition exprime que, à partir d'un certain rang* $N = N(\varepsilon)$*, tous les évènements* $(|X_n - X| < \varepsilon)$ *pour* $n > N$ *sont réalisés avec une probabilité qui tend vers 1.*

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

Lemme 2.1.4.1. *(Borel Cantelli)*

1. *Soit* $(A_n)_{n\in\mathbb{N}}$ *une suite d'évènements telle que :*

$$\sum_{n=0}^{+\infty} \mathbb{P}(A_n) < +\infty$$

alors

$$\mathbb{P}(\limsup_n A_n) = 0$$

2. *Soit* $(A_n)_{n\in\mathbb{N}}$ *une suite d'évènements indépendants telle que :*

$$\sum_{n=0}^{+\infty}\mathbb{P}(A_n)=+\infty$$

alors

$$\mathbb{P}(\limsup_n A_n)=1$$

Théorème 2.1.4.1. *Soient* $(X_n)_{n\geq 1}$ *et* X *des variables aléatoires réelles définies sur le même espace probabilité* $(\Omega,T,\mathbb{P})$ *et vérifiant :*

$$\forall\varepsilon>0,\sum_{n=0}^{+\infty}\mathbb{P}(|X_n-X|\geq\varepsilon)<+\infty. \tag{2.2}$$

Alors X_n *converge presque sûrement vers* X*.*

Démonstration 2.1.4.1. *On pose*

$$A_n=\{|X_n-X|\geq\varepsilon\}$$

$\sum_{n=0}^{+\infty}\mathbb{P}(A_n)<+\infty$ *donc par le premier lemme de Borel Cantelli,*

$$\mathbb{P}(\limsup_n A_n)=0$$

Donc (2.2)la convergence presque sûre de X_n *vers* X*.*

2.1.5 Loi forte des grands nombres

Théorème 2.1.5.1. [5] *Si* (X_n) *est une suite de variables aléatoires indépendantes de même loi et d'espérance* μ*, alors :*

$$X_n\xrightarrow{p.s}\mu$$

2.1.6 Convergence en moyenne d'ordre p

Définition 2.1.6.1. *La suite de variables aléatoires* $(X_n)_{n\geq 1}$ *converge vers* X *une autre variables aléatoires, en moyenne l'ordre* $p\geq 1$ *si :*

$$\lim_{n\to+\infty}E(|X_n-X|^P)=0$$

Remarque 2.1.6.1. *pour* $p=1$ *convergence en moyenne vers une variable aléatoire* X*, si :*

$$\lim_{n\to+\infty}E(|X_n-X|)=0$$

On note $X_n \to X$ en moyenne.
pour $p = 2$ de convergence en moyenne quadratique vers une variable aléatoire X , si :

$$\lim_{n\to+\infty} E(|X_n - X|^2) = 0$$

On note $X_n \xrightarrow{m.q} X$.

2.1.7 Convergence en loi (convergence faible)

La convergence en loi. C'est la plus faible, mais peut-être aussi la plus importante. Elle est souvent utilisée en pratique car elle permet d'approximer la loi de X_n par celle de X.

Définition 2.1.7.1. *On dit que $\{X_n\}$ converge en loi vers la v.a. X si l'on a, en tout x où sa fonction de répartition F_X est continue,*

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$$

et l'on note $X_n \xrightarrow{\mathcal{L}} X$

Remarque 2.1.7.1.

1. *La convergence en loi n'est réalisée qu'aux points de continuité de F. En un point de discontinuié x de F, $F_X(x)$ peut ne pas tendre vers aucune limite, ou tendre vers une limite différente de $F(x)$.*
2. *La convergence en loi ne suppose pas que les variables aléatoires X_n et X soient de même loi. On montre par exemple qu'une suite de variables aléatoires binomiale ou de Poisson converge vers une variable gaussienne.*
3. *La convergence de $F_n(x)$ vers une valeur limite pour tout x ne suffit pas pour avoir la convergence en loi de (X_n). Il faut que la fonction limite soit une fonction de répartition. Ainsi la suite de variables aléatoires $(X_n)_{n\geq 1}$ uniformément réparties sur $[-n, n]$:*

$$F_n(x) = \begin{cases} \frac{x+n}{2n} & si\ x \in [-n, n] \\ 1 & si\ x \geq n \\ o & si\ x \leq -n \end{cases}$$

$\lim_{n\to\infty} F_n(x) = \frac{1}{2}$, $\forall x \in \mathbb{R}$
Cette fonction limite n'est pas une fonction de répartition par conséquent, la suite (X_n) ne converge pas en loi.

Théorème 2.1.7.1. *(théorème de continuité de Lévy) Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires réelles. Si $X_n \xrightarrow{\mathcal{L}} X$ alors : quand $n \to \infty$ $\varphi_{X_n}(t) \to \varphi_X(t)$. Réciproquement si $(X_n)_{n\geq 1}$ est une suite de variables aléatoires réelles tel que : $\varphi_{X_n}(t) \to \varphi_X(t)$ et si φ est continue en 0 alors (X_n) converge en loi vers une variable aléatoire réelle X dont la fonction caractéristique est φ.*

Exemple 2.1.7.1. *Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires réelles de loi $\mathcal{B}(n,p), 0 \leq p \leq 1$.*
$X_n \xrightarrow{\mathcal{L}} X$ avec $X \rightsquigarrow \mathcal{P}(np)$
On a $\varphi_{X_n}(t) = (1-p+pe^{it})^n = \varphi_n(t)$.
Posant $\lambda = np \Rightarrow p = \frac{\lambda}{n} \Rightarrow \varphi_n(t) = (1-\frac{\lambda}{n}+\frac{\lambda}{n}e^{it})^n$

$$\begin{aligned}\lim_{n\to\infty}\varphi_n(t) &= \lim_{n\to\infty}\left[1-\frac{\lambda}{n}(1-e^{it})\right]^n \\ &= \lim_{n\to\infty}\left[1+\frac{\lambda}{n}(e^{it}-1)\right]^n \\ &= e^{\lambda(e^{it}-1)} = \varphi(t) \text{ fonction caractéristique de } X \rightsquigarrow \mathcal{P}(\lambda = np)\end{aligned}$$

Comme φ est continue en 0, alors : $X_n \xrightarrow{\mathcal{L}} X$, i.e :

$$(X_n\mathcal{B}(n,p)) \xrightarrow{\mathcal{L}} X(X \rightsquigarrow \mathcal{P}(np)).$$

2.2 Théorème central limite

Théorème 2.2.0.1. *Soit $X_1, X_2....X_n$ une suite de variables aléatoires définies sur le même espace de probabilité, suivant la même loi D et indépendantes . Supposons que l'espérance μ et l'écart-type σ de D existent et soient finis ($\sigma \neq 0$) .*
Considérons la somme $S_n = X_1 + X_2 + ... + X_n$. Alors l'espérance de S_n est $n\mu$ et son écart-type vaut $\sigma\sqrt{n}$. De plus, quand n est assez grand, la loi normale $\mathcal{N}(n\mu, n\sigma^2)$ est une bonne approximation de la loi de S_n . Afin de formuler mathématiquement cette approximation, nous allons poser :

$$\begin{aligned}\overline{X} = \frac{S_n}{n} &= \frac{(X_1+X_2+....+X_n)}{n} \\ Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}} &= \frac{\overline{X}-\mu}{\sigma/\sqrt{n}}.\end{aligned}$$

De sorte que l'espérance et l'écart-type de Z_n valent respectivement 0 et 1 : la variable est ainsi dite centrée et réduite . Le théorème central limite stipule alors que la loi de Z_n converge vers la loi normale centrée réduite $\mathcal{N}(0,1)$

lorsque n tend vers l'infini . Cela signifie que si Φ *est la fonction de répartition de* $\mathcal{N}(0,1)$*, alors pour tout réel* Z *:*

$$\lim_{n\to+\infty} \mathbb{P}(Z_n \leq z) = \Phi(z)$$

où, de façon équivalente :

$$\lim_{n\to+\infty} \mathbb{P}(\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}) = \Phi(z)$$

Démonstration 2.2.0.1. *On utilise la méthode des fonctions caractéristiques. Pour une variable aléatoire* Y *d'espérance 0 et de variance 1, la fonction caractéristique de la loi Normale centrée réduite est :* $\varphi(t) = e^{-\frac{t^2}{2}}$*.*

$$\varphi_Y(t) = 1 - \frac{t^2}{2} + 0(t^2),\ t \longrightarrow 0.$$

Si Y_i *vaut* $\frac{X_i-\mu}{\sigma}$*, il est facile de voir que la moyenne centrée réduite des observations* $X_1, X_2...., X_n$ *est simplement :*

$$\begin{aligned} Z_n &= \frac{\overline{X}-\mu}{\sigma\sqrt{n}} \\ &= \sum_{i=1}^{n} \frac{Y_i}{\sqrt{n}}. \end{aligned}$$

simplement :

$$\begin{aligned} \varphi_{Z_n}(t) &= E(e^{itZ_n}) \\ &= E(e^{it\sum_{i=1}^{n}\frac{Y_i}{\sqrt{n}}}) \\ &= \prod_{i=1}^{n} E(e^{it\frac{Y_i}{\sqrt{n}}}) \\ &= (E(e^{it\frac{Y_i}{\sqrt{n}}})^n). \end{aligned}$$

Alors :

$$\begin{aligned} \varphi_{Z_n}(t) &= [E(e^{it\frac{Y_i}{\sqrt{n}}})]^n \\ &= (1 - \frac{t^2}{2n} + 0(t^2))^n. \end{aligned}$$

En passant à la limite

$$\begin{aligned} \lim_{n\to+\infty} \varphi_{Z_n}(t) &= \lim_{n\to+\infty} (1 - \frac{t^2}{2n} + 0(t^2))^n \\ &= e^{\frac{-t^2}{2}}. \end{aligned}$$

D'où $Z_n \rightsquigarrow \mathcal{N}(0,1)$*. Mais cette limite est la fonction caractéristique de la loi normale centrée réduite* $\mathcal{N}(0,1)$*, d'où l'on déduit le théorème de la limite centrale grâce au théorème de continuité de Lévy, qui affirme que la convergence des fonctions caractéristiques implique la convergence en loi .*

Exemple 2.2.0.1. *On lance* 15 *fois un dé . Trouver la probabilité que la somme des* 15 *résultats soit compris entre* 40 *et* 50 *?.*
Si on note par X_i *la variable aléatoire égale au résultat du* i^{me} *jet* $i = 1, 2, 3, ..., 15$. *On a :*
$E(X_i) = \frac{7}{2}$ *et* $Var(X_i) = \frac{35}{12}$. *En notant par* $S_n = X_1, X_2, ..., X_{15}$ *; L'application du théorème de la limite central donne :*

$$\begin{aligned}
\mathbb{P}(40 \leq S_n \leq 50) &= \mathbb{P}(\frac{40 - 15.\frac{7}{2}}{\sqrt{15.\frac{35}{12}}} \leq \frac{\sum_{i=1}^{15} X_i - 15.\frac{7}{2}}{\sqrt{15.\frac{35}{12}}} \leq \frac{50 - 15.\frac{7}{2}}{\sqrt{15.\frac{35}{12}}}) \\
&= \Phi(-0.38) - \Phi(-1.89) \\
&= \Phi(1.89) - \Phi(0.38) \\
&= 0.9706 - 0.6480 \\
&= 0.3226
\end{aligned}$$

Chapitre 3

Les liens entre les modes de convergence

Les relations entre les différents type de convergence, sont résumées dans le théorème suivant :

Théorème 3.0.0.1. *Soit $(X_n)_{n\in\mathbb{N}^*}$, une suite de variables aléatoires définies sur un même espace de propabilité $(\Omega, T, \mathbb{P})$.*
On suppose X définie sur ce même espace . Alors, lorsque $n \longrightarrow +\infty$:

1. $X_n \xrightarrow{p.s} X \Rightarrow X_n \xrightarrow{\mathbb{P}} X.$
2. $X_n \xrightarrow{m.q} X \Rightarrow X_n \xrightarrow{\mathbb{P}} X.$
3. $X_n \xrightarrow{\mathbb{P}} X \Rightarrow X_n \xrightarrow{L} X.$

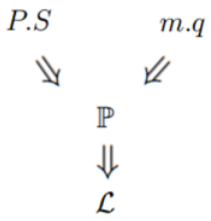


FIGURE 3.1 – Les liens entre les modes de convergence

Les réciproques sont généralement non vérifiées.

3.1 les relations entre les différents types de convergence

3.1.1 La convergence p.s ⇒ La convergence en $\mathbb{P}$

Proposition 3.1. *La convergence presque sûre entraîne la convergence en probabilité . la réciproque est en générale fausse.*

Démonstration 3.1.1.1. *Soit $\varepsilon > 0$ fixé. On pose $f_n = 1_{|X_n - X| > \varepsilon}$. Les fonctions f_n sont mesurables, positives et majorées par 1. Par ailleurs, d'après la convergence presque sûre, pour presque tout $\omega \in \Omega$, il existe n_0 tel que pour tout $n > n_0$,*

$$|X_n(\omega) - X(\omega)| \leq \varepsilon$$

C'est à dire $f_n(\omega) = 0$ pour tout $n \geq n_0$. On a donc écrit que la suite f_n converge presque sûrement vers 0. D'aprés le théorème de convergence dominée

$$\mathbb{P}(|X_n - X| \geq \varepsilon) = E[f_n] \longrightarrow E[\lim_{n \longrightarrow \infty}(f_n)] = E[0] = 0.$$

la convergence en $\mathbb{P}$ $\nRightarrow$La convergence P.S

Exemple 3.1.1.1. *Soit $(X_n)_{n \in \mathbb{N}^*}$, une suite de variables aléatoires indépendantes, définies sur un espace de probabilité $(\Omega, T, \mathbb{P})$, telles que :*

$$\mathbb{P}\{X_n = 0\} = 1 - \frac{1}{n}$$

$$\mathbb{P}\{X_n = 1\} = \frac{1}{n}$$

La suite (X_n) converge en probabilité vers 0 . En effet :

$$\mathbb{P}\{|\ X_n - 0\ | > \varepsilon\} = \mathbb{P}\{X_n = 1\} = \frac{1}{n} \longrightarrow 0$$

quand $n \to +\infty$ Mais, cette suite ne converge pas presque sûrement vers 0, car :

$$\mathbb{P}\{sup_{n \geq \mathbb{N}} |\ X_n\ | > \varepsilon\} = \mathbb{P}(\bigcup_{n \geq \mathbb{N}} \{|\ X_n\ | > \varepsilon\})$$

$$= 1 - \mathbb{P}(\bigcap_{n \geq \mathbb{N}} \{|\ X_n\ | \leq \varepsilon\})$$

$$= 1 - \prod_{n \geq \mathbb{N}} \mathbb{P}\{|\ X_n\ | \leq \varepsilon\}$$

$$= 1 - \prod_{n\geq\mathbb{N}}(1 - \frac{1}{n})$$

La série de terme générale $\frac{1}{N}$*, étant divergente,lorsque* $N \longrightarrow +\infty, \prod_{n\geq\mathbb{N}}(1-\frac{1}{n}) \longrightarrow 0$ *Alors :*

$$\mathbb{P}\{sup_{n\geq\mathbb{N}} \mid X_n \mid > \varepsilon\} \longrightarrow 1.$$

3.1.2 La convergence m.q ⇒ La convergence en $\mathbb{P}$

Proposition 3.2. *La convergence en moyenne quadratique entraîne la convergence en probabilité . La réciproque est en générale fausse .*

Démonstration 3.1.2.1. *La preuve est une conséqunce immédiate de l'inégalité de Tchébychev :*

$$\mathbb{P}\{\mid X_n - X \mid > \varepsilon\} \leq \frac{E\{\mid X_n - X \mid^2\}}{\varepsilon^2}, \forall\varepsilon > 0.$$

La convergence en $\mathbb{P}$ ⇏ La convergence m.q

Exemple 3.1.2.1. *Soit* $(X_n)_{n\in\mathbb{N}^*}$*, une suite de variables aléatoires indépendantes, définies sur un espace de probabilité* $(\Omega, T, \mathbb{P})$ *par :*

$$\mathbb{P}\{X_n = -\frac{1}{n}\} = \frac{n^2}{1+n^2}$$

$$\mathbb{P}\{X_n = n\} = \frac{1}{1+n^2}$$

La suite (X_n) *converge en probabilité vers 0 .*

$$\mathbb{P}\{\mid X_n - 0 \mid > \varepsilon\} = \mathbb{P}\{\mid X_n \mid > \varepsilon\}$$

$$= \frac{1}{1+n^2}$$

dés que $n > \frac{1}{\varepsilon}$*. Alors :*

$$\mathbb{P}\{\mid X_n \mid > \varepsilon\} \longrightarrow 0$$

quand $n \to +\infty$ *La suite* (X_n) *ne converge pas en moyenne quadratique vers 0.*

$$E(X_n^2) = \frac{1}{n^2}.\frac{n^2}{1+n^2} + n^2.\frac{1}{1+n^2} = 1$$

$E(X_n^2)$ *ne tend pas vers* $E(X) = 0$ *quand* $n \to +\infty$.

3.1.3 La convergence en $\mathbb{P} \Rightarrow$ La convergence en Loi

Proposition 3.3. *La convergence en probabilité entraîne la convergence en loi . La réciproque est en générale fausse .*

Démonstration 3.1.3.1. *Sopposons d'abord que $(X_n)_{n\geq 0}$ converge dans L^1 vers X_∞. Soit $\varepsilon > 0$. On a alors, par l'inégalité de Markov,*

$$\mathbb{P}(|X_n - X_\infty| > \varepsilon) < \frac{1}{\varepsilon} E(|X_n - X_\infty|)$$

et ce dernier terme tend vers 0, donc la suite convergence en probabilités vers X_∞.
Si $(X_n)_{n\geq 0}$ converge p.s. vers X_∞, $(1_{|X_n - X_\infty|>\varepsilon})_{n\geq 0}$ converge p.s. vers 0, et est dominée par 1. Par application du théorème de convergence dominée, $\mathbb{P}(|X_n - X_\infty| > \varepsilon) \to 0$.

Enfin, sopposons que $(X_n)_{n\geq 0}$converge en probabilités vers X_∞. Soit f une fonction continue à support compact sur $\mathbb{R}$. f est alors uniformément continue sur $\mathbb{R}$: pour $\varepsilon > 0$ donné, il existe $\alpha > 0$ tel que $|x - y| < \alpha \Rightarrow |f(x) - f(y)| < \varepsilon$. On a alors

$$\begin{aligned} |E(f(X_n) - f(X_\infty))| &= E(|f(X_n) - f(X_\infty)|1_{|X_n - X_\infty|<\alpha}) \\ &+ E(|f(X_n) - f(X_\infty)|1_{|X_n - X_\infty|} \geq \alpha) \\ &\leq \varepsilon + 2\| f \|_\infty \mathbb{P}(|X_n - X_\infty| \geq \alpha) \\ &\leq 2\varepsilon \end{aligned}$$

pour n assez grand. On a donc $E(f(X_n)) \to E(f(X_\infty))$.

La convergence en loi $\not\Rightarrow$ La convergence en $\mathbb{P}$

Exemple 3.1.3.1. *Soit la suite $X_n = (-1)^n X$ où X est une variable aléatoire normale centrée réduite $(\mathcal{N}(0, 1))$.*
$*$ Montrer que X_n converge en loi vers X et ne converge pas en probabilité vers X.

Solution :

$X_n \longrightarrow X$ si pour tout x : $F_{Xn}(x) \longrightarrow F_X(x)$ F_{Xn} est la fonction de répartition de X_n et F_X la fonction de répartition de X.

$$F_{Xn}(x) = \mathbb{P}(X_n \leq x) = \mathbb{P}((-1)^n X \leq x)$$

Si n est pair : $(-1)^n = 1$ *et* : $X_n = X$

$$F_{Xn}(x) = \mathbb{P}(X \leq x) = F_X(x)$$

* *Si n est impair* : $(-1)^n = -1$ *et* :$X_n = -X$

$$F_{Xn}(x) = \mathbb{P}(-X \leq x) = \mathbb{P}(X \geq -x) = \mathbb{P}(X \leq x)$$

ceci en raison de la symétrie de la densité de X D'où si n est impair :

$$F_{Xn}(x) = F_X(x)$$

ainsi pour tout n : $F_{Xn} = F_X$, X_n converge en loi vers X Toutfois X_n ne converge pas en probabilité vers X car X_n ne tend pas vers X : $X_n = -X$ si n est impair .

Corollaire 3.1.1. [5] *Soit $(X_n)_{n\in\mathbb{N}^*}$ est une suite de variables aléatoire et soit c une constante.*
Si X_n converge en loi vers la constante c alors elle converge en probabilité vers la constante c.

Proposition 3.4. [5] *Soient $(X_n)_{n\in\mathbb{N}^*}$ une suite de variables aléatoires et X une variable aléatoire . Donc si (X_n) converge en moyenne d'orde p vers X , alors elle converge en loi vers la même variable X.*

Proposition 3.5. [5] *Soient $(X_n)_{n\in\mathbb{N}^*}$ une suite de variables aléatoires et X une variable aléatoire . Donc si (X_n) converge presque sûre moyenne vers X, alors elle converge en loi vers la même variable X.*

Deuxième partie

Statistique Paramétrique

Chapitre 4

Echantillonnage

l'échantillonnage statistique consiste à prédire, à partir d'une population connue, les caractéristiques des échantillons qui en seront prélevés. Le choix de l'échantillon est une étape délicate. En effet, pour étendre à une population entière les résultats obtenus sur un échantillon, il faut que cet échantillon reproduise le mieux possible les caractéristiques de la population.

Plusieurs méthodes de sélection d'échantillon existent (aléatoire simple, en grappes, stratification). On distingue les méthodes probabilistes où les éléments tirés de la population ont une probabilité connue de faire partie de l'échantillon et les méthodes dites non probabilistes.

L'échantillonnage statistique consiste à utiliser l'information obtenue sur un échantillon pour pouvoir déduire de l'information sur la population (ou l'univers) d'intérêt (illustration à la figure 4) : on extrait un échantillon de la population, on l'analyse et on infère sur la population.

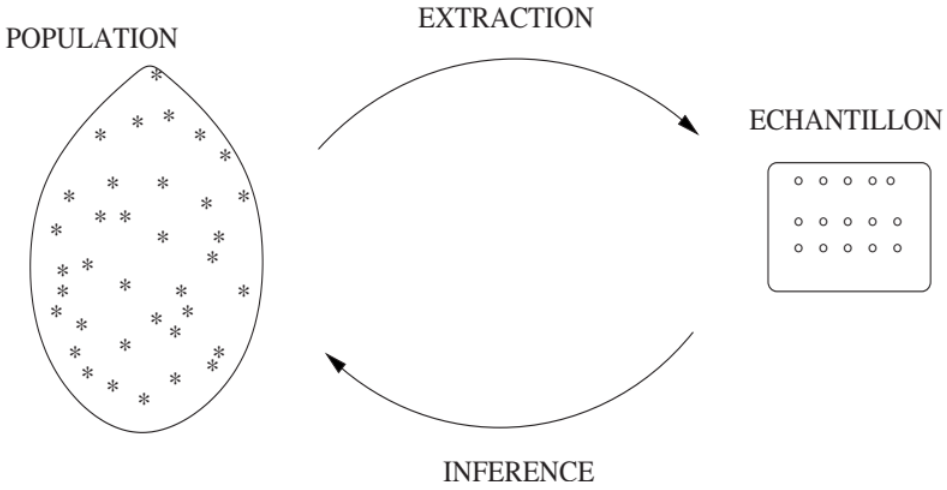


FIGURE 4.1 – Illustration de l'échantillonnage

4.1 Notions fondamentales

4.1.1 Échantillonnage

Définition 4.1.1.1. *On appelle échantillon aléatoire de taille n (en bref n-échantillon) d'une variable aléatoire X une suite de n variables aléatoires indépendantes et de même loi (ou v.a. i.i.d). Cette loi est appelée* ***la loi mère*** *de l'échantillon.*

Remarque 4.1.1.1. *On distinguera la notion d'échantillon aléatoire* $(X_1, X_2, .., X_n)$ *dont on peut dire qu'elle se réfère à des résultats potentiels avant expérience, de celle d'échantillon réalisé* $(x_1, x_2, ..., x_n)$ *correspondant aux valeurs observées après expérience.*

Définition 4.1.1.2. *Soit* $X_1, X_2, ..., X_n$ *un n-échantillon, on appelle statistique toute v.a.* $T_n = h(X_1, X_2, ..., X_n)$*, prend la valeur* $h(x_1, x_2, .., x_n)$

Exemple 4.1.1.1. $T_1 = \frac{1}{n}\sum_{i=1}^{n} X_i, T_2 = \sum_{i=1}^{n} X_i^2, T_3 = \sqrt{\sum_{i=1}^{n} X_i^3}$ *sont des statistiques.*

4.2 Quelques statistiques classiques

4.2.1 Moyenne, variance, moments empiriques

Définition 4.2.1.1. *On appelle* ***moyenne*** *de* ***l'échantillon*** *ou* ***moyenne empirique*** *la statistique,notée* $\overline{X}$*,définie par :*

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

Définition 4.2.1.2. *On appelle* ***variance empirique*** *la statistique, notée* S^2*, définie par :*

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

.

Définition 4.2.1.3. *On appelle* ***variance empirique corrigée*** *la statistique, notée* S^{*2}*, définie par :*

$$S^{*2} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

.

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

Proposition 4.1. *Soit* μ *et* σ^2 *, respectivement la moyenne et la variance de la loi mère on a :*

$$E(\overline{X}) = \mu \quad V(\overline{x}) = \frac{\sigma^2}{n}$$

Démonstration 4.2.1.1.

$$E(\frac{1}{n}\sum_{i=1}^{n} X_i) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{1}{n}\sum_{i=1}^{n} \mu = \mu$$

Puis,en raison de l'indépendance des X_i

$$V(\frac{1}{n}\sum_{i=1}^{n} X_i) = \frac{1}{n^2}\sum_{i=1}^{n} V(X_i) = \frac{1}{n^2}\sum_{i=1}^{n} \sigma^2 = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$$

Proposition 4.2.1.1. *La moyenne de la loi de la variance empirique est :*

$$E(S^2) = \frac{n-1}{n}\sigma^2$$

Démonstration 4.2.1.2.

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2 &= \frac{1}{n}\sum_{i=1}^{n}[(X_i-\mu)-(\overline{X}-\mu)]^2\\ &= \frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2-2(\overline{X}-\mu)\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)+(\overline{X}-\mu)^2\\ &= \frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2-2(\overline{X}-\mu)^2+(\overline{X}-\mu)^2\\ &= \frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2-(\overline{X}-\mu)^2\end{aligned}$$

D'où

$$E(S^2)=\frac{1}{n}\sum_{i=1}^{n}V(X_i)-V(\overline{X})=\sigma^2-\frac{\sigma^2}{n}=\frac{n-1}{n}\sigma^2$$

Remarque 4.2.1.1. *Quand on abordera l'estimation ponctuelle (chapitre 3) on dira que S^2 est un estimateur biaisé de σ^2 son biais vaut $\frac{\sigma^2}{n}$. on utilise fréquemment la variance corrigée dont l' espérance vaut exactement σ^2*

$$S^{*2}=\frac{n}{n-1}S^2$$

4.3 Cas des échantillons gaussiens

$\overline{X}$ combinaison linéaire de variables gaussiennes indépendantes est aussi de gaussienne et

$$\overline{X}\longrightarrow\mathcal{N}(\mu,\frac{\sigma^2}{n})$$

.

Proposition 4.2. [7] *Si la loi mère est gaussienne, $\overline{X}$ et S^2 sont des v.a. indépendantes.*

4.4 Moments d'un échantillon

Définition 4.4.0.1. *Soit $(X_1,X_2,...,X_n)$ un n-échantillon aléatoire issu d'une variable alèatoire X et r un nombre entier ($r\in\mathbb{N}$). On appelle* ***moment d'ordre r****, de l'échantillon et on note M_r, la quantité*

$$M_r=\frac{1}{n}\sum_{i=1}^{n}X_i^r$$

Définition 4.4.0.2. *On appelle* ***moment centré empirique*** *d'ordre r, noté* $\grave{M}_r$*, la statistique*

$$\grave{M}_r = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^r$$

.

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

4.5 Loi du Khi-deux

Définition 4.5.0.1. *Soit* $Z_1, Z_2, ..., Z_n$ *une suite de variables aléatoires i.i.d. de loi* $\mathcal{N}(0,1)$*. Alors la v.a.* $\sum_{i=1}^{n} Z_i^2$ *suit une loi appelée loi du Khi-deux à n degrés de liberté, notée* $\mathcal{X}(n)$*.*

Proposition 4.3. *La densité de la loi du Khi-deux à n degrés de liberté est :*

$$f_n(x) = \frac{1}{2^{n/2}\Gamma(n/2)} x^{\frac{n}{2}-1} e^{-\frac{x}{2}} \text{ pour } x > 0 (0 \text{ si non}).$$

Démonstration 4.5.0.1. *calculons la fonction génératrice de* Z^2 *où* $Z \rightsquigarrow \mathcal{N}(0,1)$*. On a :*

$$\begin{aligned}
\Psi_{Z^2(t)} &= E(e^{tz^2}) = \int_{-\infty}^{+\infty} e^{tz^2} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{1}{2}(1-2t)z^2} dz \\
&= \frac{1}{\sqrt{1-2t}} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{1}{2}u^2} du \text{ en posant } u = \sqrt{1-2t}z \\
&= \frac{1}{\sqrt{1-2t}}
\end{aligned}$$

qui est définie pour $t < \frac{1}{2}$*.*

Pour la somme des Z_i^2 *indépendantes on a donc*

$$\Psi_{\sum_{i=1}^{n} Z_i^2}(t) = \left(\frac{1}{1-2t}\right)^{\frac{n}{2}}$$

qui n'est autre que la fonction génératrice d'une loi $\Gamma(\frac{n}{2}, \frac{1}{2})$*. On voit donc, au passage, qu'une loi du Khi-deux est un cas particulier de loi de gamma.*

Proposition 4.4. *La moyenne de la loi* $\chi^2(n)$ *est égale au nombre de degrés de liberté n, sa variance est* $2n$*.*

Démonstration 4.5.0.2. *Comme* $Z_i \rightsquigarrow \mathcal{N}(0,1)$ *on a :*

$$E(Z_i^2) = V(Z_i) = 1 \text{ d'où } E(\sum_{i=1}^{n} Z_i^2) = n$$

$$V(Z_i^2) = E(Z_i^4) - (E(Z_i^2))^2 = \mu_4 - 1$$

$\mu_4 = 3$, *d'où* $V(Z_i^2) = 2$ *et* $V(\sum_{i=1}^{n} Z_i^2) = 2n$

Théorème 4.5.0.1. *Soit un n-èchantillon* $X_1, X_2, ..., X_n$ *de loi* $\mathcal{N}(\mu, \sigma^2)$ *on a :*

$$\frac{(n-1)S^{*2}}{\sigma^2} \rightsquigarrow \chi^2(n-1)$$

.

Démonstration 4.5.0.3. *D'aprés la décomposition de* S^2 *:*

$$\sum_{i=1}^{n}(X_i - \mu)^2 = \sum_{i=1}^{n}(X_i - \overline{X})^2 + n(\overline{X} - \mu)^2$$

En divisant par σ^2 *:*

$$\begin{aligned}\sum_{i=1}^{n}\left(\frac{X_i - \mu}{\sigma}\right)^2 &= \sum_{i=1}^{n}(\frac{X_i - \overline{X}}{\sigma})^2 + \frac{n}{\sigma^2}(\overline{X} - \mu)^2 \\ &= \frac{(n-1)S^{*2}}{\sigma^2} + \left(\frac{\overline{X} - \mu}{\sigma/\sqrt{n}}\right)^2\end{aligned}$$

Les deux termes de droite sont, respectivement, $\frac{(n-1)S^{*2}}{\sigma^2}$ *et* $\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}$ *qui est une gaussienne centré-réduite. Ces termes aléatoires étant indépendants comme fonctions de* S^{*2} *et de* $\overline{X}$*, et les v.a.* $\frac{X_i-\mu}{\sigma}$ *étant indépendantes de loi* $\mathcal{N}(0,1)$ *on a, en termes de fonctions génératrices :*

$$\left(\frac{1}{1-2t}\right)^{\frac{n}{2}} = \Psi_{\frac{(n-1)S^{*2}}{\sigma^2}}(t).\left(\frac{1}{1-2t}\right)^{\frac{1}{2}} \text{ si } t < \frac{1}{2}$$

Finalement :

$$\Psi_{\frac{(n-1)S^{*2}}{\sigma^2}}(t) = \left(\frac{1}{1-2t}\right)^{\frac{n-1}{2}}.$$

ce qui prouve le théorème.

4.6 Loi de Student

Définition 4.6.0.1. *Soit Z et Q deux v.a. indéndantes telles que $Z \rightsquigarrow \mathcal{N}(0,1)$ et $\rightsquigarrow \chi^2(n)$. Alors la v.a.*

$$T = \frac{Z}{\sqrt{\frac{Q}{n}}}$$

*suit une loi appelée **loi de Student** à n degrés de liberté, notée $t(n)$.*
La densité de la loi de Student à n degrés de liberté est :

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

Proposition 4.5. [10] *Soit $T \to t(n)$ alors $E(T) = 0$ si $n \geq 2$ et $V(T) = \frac{n}{n-2}$ si $n \geq 3$.*

Théorème 4.6.0.1. *Soit $X_1, X_2, ..., X_n$ un n-échantillon de loi mère $\mathcal{N}(\mu, \sigma^2)$. Alors :*

$$\frac{\overline{X} - \mu}{S^*/\sqrt{n}} \rightsquigarrow t(n-1)$$

Démonstration 4.6.0.1. *La démonstration est immédiate, il suffit d'écrire la v.a.*

$$\frac{\overline{X} - \mu}{S^*/\sqrt{n}} = \frac{\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}}{\sqrt{\frac{(n-1)S^{*2}/\sigma^2}{n-1}}}$$

4.7 Loi de Fisher-Snedecor

Définition 4.7.0.1. *Soit U et V deux v.a. indépendantes telles que $U \rightsquigarrow \chi^2(n_1)$ et $V \rightsquigarrow \chi^2(n_2)$. Alors la v.a.*

$$F = \frac{U/n_1}{V/n_2}$$

***suit une loi de Fisher-Snedecor** à n_1 degrés de liberté au numérateur et n_2 degrés de liberté au dénominateur, notée $F(n_1, n_2)$. En bref on l'appellera loi de Fisher.*
La densité de la loi $F(n_1, n_2)$ est :

$$f_{n_1,n_2}(x) = \frac{\Gamma(\frac{n_1+n_2}{2})}{\Gamma(\frac{n_1}{2})\Gamma(\frac{n_2}{2})}\left(\frac{n_1}{n_2}\right)^{\frac{n_1}{2}} \frac{x^{\frac{n_1-2}{2}}}{(1+\frac{n_1}{n_2}x)^{\frac{n_1+n_2}{2}}} \text{ si } x > 0 (0 \text{ sinon}).$$

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 $\frac{2n_2^2(n_1+n_2-2)}{n_1(n_2-2)^2(n_2-4)}$.

Proposition 4.6. *Soit H une v.a. suit une loi de Fisher telle que $H \rightsquigarrow F(n_1, n_2)$ alors $\frac{1}{H} \rightsquigarrow F(n_2, n_1)$*

Démonstration 4.7.0.1. *La démonstration est évidente de par la définition de la loi de Fisher.*

4.8 Statistique d'ordre d'un échantillon

Pour une série de nombres réels $(x_1, x_2, ..., x_n)$ notons $max\{x_1, x_n, ..., x_n\}$ la fonction de $\mathbb{R}^n$ dans $\mathbb{R}$ qui lui associe le nombre maximal de cette série. On peut donc définir une v.a., notée $X_{(n)}$, fonction de $(X_1, X_2, ..., X_n)$ par :

$$X_{(n)} = max\{X_1, X_2, ..., X_n\}$$

a) Loi de $X_{(n)} = max\{X_1, X_2, ..., X_n\}$:

$$\begin{aligned} F_{X_{(n)}}(x) &= P(X_1 \leq x, X_2 \leq x, ..., X_n \leq x) \\ &= P(X_1 \leq x).P(X_2 \leq x).....P(X_n \leq x) \text{ (indépendance.)} \\ &= [F(x)]^n \text{ même loi.} \end{aligned}$$

b) Loi de $X_{(1)} = min\{X_1, X_2, ..., X_n\}$:

$$\begin{aligned} P(X_{(1)} > x) &= P(X_1 > x).P(X_2 > x).....P(X_n > x) \\ &= [1 - F(x)]^n \end{aligned}$$

d'où

$$F_{X_{(1)}}(x) = 1 - [1 - F(x)]^n$$

Définition 4.8.0.1. *Soit h_k la fonction de $\mathbb{R}^n$ dans $\mathbb{R}$ qui à $(x_1, x_2, ..., x_n)$ fait correspondre la k-ième valeur parmi $x_1, x_2, ..., x_n$ lorsqu'on les range dans l'ordre croissant.*
On appelle alors statistique d'ordre k, la v.a. notée $X_{(k)}$, définie par :

$$X_{(k)} = h_k(X_1, X_2, ...X_n)$$

4.9 Fonction de répartition empirique d'un échantillon

Définition 4.9.0.1. *Pour tout $x \in \mathbb{R}$, on appelle valeur de la fonction de répartition empirique en x, la statistique, notée $F_n(x)$, définie par :*

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{]-\infty,x]}(X_i)$$

$F_n(x)$ est la v.a «proportion» des n observations $X_1, X_2, ..., X_n$ prenant une valeur inférieure ou égale à x, $n.F_n(x)$ suit une loi binomiale $\mathcal{B}(n, F(x))$ et par conséquence, $F_n(x)$ est une v.a discrète prenant les valeurs $\frac{k}{n}$, 0ù $k = 0, 1, 2...., n$ avec probabilités :

$$\mathbb{P}(F_n(x) = \frac{k}{n}) = \mathbb{P}(nF_n(x) = k) = C_n^k (F(x))^k (1 - F(x))^{n-k}$$

4.9.1 Convergence de $F_n(x)$ vers $F(x)$

Théorème 4.9.1.1. *Soit $(X_1, X_2, ..., X_n)$ un n-échantillon de fonction de répartition empirique $F_n(x)$ et $F(x)$ fonction de répartition de X (variable aléatoire parente). Alors : $F_n(x) \overset{p.s}{\to} F(x)$. i.e : $\mathbb{P}\left(\lim_{x\to+\infty} F_n(x) = F(x)\right) = 1$*

Démonstration 4.9.1.1. *$F_n(x)$ étant une moyenne empirique de variable aléatoire réelle indépendante (puisque les X_i le sont), d'après la loi forte des grands nombres :*

$$F_n(x) \overset{p.s}{\to} E[\mathbb{1}_{(X_i<x)}] = E[\mathbb{1}_{X<x}] = F(x)$$

Convergence en loi de $F_n(x)$

On a : $\mathbb{1}_{(X_i<x)}$ est une variable aléatoire de Bernoulli $\mathcal{B}(F(x))$

$$\mathbb{P}(F_n(x) = \frac{k}{n}) = \mathbb{P}(nF_n(x) = k) = C_n^k (F(x))^k (1 - F(x))^{n-k}$$

donc $nF_n(x) \rightsquigarrow \mathcal{B}(n, F(x))$. Alors d'après le théorème central limite :

$$\frac{nF_n(x) - nF(x)}{\sqrt{nF(x)(1 - F(x))}} \overset{\mathcal{L}}{\to} \mathcal{N}(0, 1)$$

ie

$$\frac{\sqrt{n}(F_n(x) - F(x))}{\sqrt{F(x)(1 - F(x))}} \overset{\mathcal{L}}{\to} \mathcal{N}(0, 1)$$

4.10 Modèle statistique

Définition 4.10.0.1. *Un modèle statistique est défini par la donnée d'une caractéristique vectorielle* $(X_1, X_2, ..., X_n)$ *et d'une famille de lois de probabilité de* X *notée* $(\mathbb{P}_\theta)_{\theta\in\Theta}$.

Lorsque la famille de lois de probabilité $(\mathbb{P}_\theta)_{\theta\in\Theta}$ *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* $\mathbb{R}^n$ *où* n *est la dimension du paramètre* θ*, le modèle est appelé modèle paramétrique. Dans le cas contraire, le modèle est appelé modèle non paramétrique.*

Remarque 4.10.0.1. *Le modèle statistique d'échantillonnage paramétrique est noté le triplet* $(E, \mathcal{E}, \mathbb{P}_\theta, \theta \in \Theta)^n$

Remarque 4.10.0.2. *La famille de lois de probabilités* $\{\mathbb{P}_\theta, \theta \in \Theta\}$ *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* $(x_1, x_2, ..., x_n.)$ *de l'échantillon aléatoire* $(X_1, X_2, ..., X_n)$*, où les* X_n *sont des variables aléatoires réelles indépendantes et de même loi que* X*. Naturellement, l'estimation du ou des paramètres inconnus doit être aussi précise que possible. Les trois grands types d'estimation paramétriques sont :*

1. *L'estimation ponctuelle : on estime directement par un ou des réels le ou les paramètres inconnus*
2. *L'estimation par intervalles de confiance : on détermine des intervalles de réels, les moins étendus possible, qui ont de fortes chances de contenir un paramètre inconnu*
3. *Les tests statistiques : démarches qui consistent à accepter ou non une hypothèse mettant en jeu un ou plusieurs paramètres inconnus, avec un faible risque de se tromper.*

Chapitre 5

Estimation ponctuelle

5.1 Introduction :

La théorie de l'échantillonnage consistait à déterminer des propriétés sur des échantillons tirés au hasard parmi une population dont on connait les propriétés.
Le principe de l'estimation est de faire exactement l'inverse, c'est-à-dire que l'on a accès à des informations sur des échantillons et l'on souhaite déterminer certaines propriétés sur la population entière.

La distribution exacte de la variable X qui nous intéresse est généralement partiellement connue. En effet le type de distribution est très souvent connu, mais il nous manque les paramètres de cette loi. Par exemple, on sait que X suit une loi normale, mais on ne connait pas les paramètres : espérance et variance.

L'estimation consiste à rechercher une valeur pour chacun des paramètres inconnus, à partir des observations. Lorsqu'on attribue à un paramètre une valeur unique, on dit que l'on fait une estimation ponctuelle. L'estimateur choisi est une fonction des observations, c'est une variable aléatoire, c'est-à-dire la valeur de l'estimateur dépend des observations. On le choisit suivant des critères qui assurent sa proximité avec le paramètre à estimer .
En résumé :

Estimer un paramètre, c'est donner une valeur approchée de ce paramètre, à partir des résultats obtenus sur un échantillon aléatoire extrait de la population.

5.2 Définition d'une statistique

Soit X est une variable aléatoire dont la fonction de répartition $F(x, \theta)$ et la densité $f(x, \theta)$ dépendent du paramètre θ, Θ est l'ensemble des valeurs possibles de ce paramètre. On considère un échantillon de taille n de cette variable $X = (X_1, ..., X_n)$. Une statistique est une fonction mesurable T des variables aléatoires X_i :

$$T = h(X_1, ..., X_n)$$

À un échantillon, on peut associer différentes statistiques. La théorie de l'estimation consiste à définir des statistiques particulières, appelées **estimateurs**. Une fois l'échantillon effectivement réalisé, l'estimateur prend une valeur numérique, appelée **estimation** du paramètre θ.
On notera $\hat{\theta}$ l'estimateur du paramètre θ.

5.2.1 Exemples élémentaires

$\overline{x}, S^2$ sont **des estimations** de μ et σ^2 (resp).
Les variables aléatoires $\overline{X}, S^2$, sont **les estimateurs** de μ et σ^2 (resp).

Remarque 5.2.1.1. *Le même paramètre peut être estimé à l'aide d'estimateurs différents*

Exemple 5.2.1.1. *Le paramètre λ d'une loi de Poisson peut-être estimé par $\overline{X}$ et S^2.*

5.3 Principales qualités d'un estimateur

5.3.1 Estimateur convergent

La première condition imposée à un estimateur est d'être convergent : (consistant $\lim_{n\to\infty} T = 0$). Deux estimateurs convergents ne convergent cependant pas nécessairement à la même vitesse, ceci est lié à la notion de précision d'un estimateur.
Deux estimateurs convergents ne convergent cependant pas nécessaireme la même vitesse, ceci est lié à la notion de précision d'un estimateur. On mesure généralement la précision d'un estimateur par **l'erreur quadratique moyenne**.

— **Perte quadratique :** C'est l'écart au carré entre le paramètre et son estimateur :

$$l(T, \theta) = (T - \theta)^2$$

— **Risque d'un estimateur :** C'est la moyenne des pertes :
$eqm(T,\theta) = \mathbb{E}[(T-\theta)]^2$ est le risque quadratique moyen.

Un estimateur T est dit convergent si $\lim\limits_{n\to\infty} eqm(T,\theta) = 0$

5.3.2 Estimateur sans biais

L'erreur d'estimation est mesurée par la quantité $T-\theta$ qui peut s'écrire :

$$T-\theta = T - E(T) + E(T) - \theta$$

- $T - E(T)$ représente les fluctuations de l'estimateur T autour de sa valeur moyenne $E(T)$ (espérance mathématique).
- $E(T) - \theta$ est une erreur systématique car l'estimateur T varie autour de son espérance mathématique $E(T)$ et non autour de la valeur θ du paramètre sauf si $E(T) = \theta$.

La quantité $E(T) - \theta$ est **le biais** de l'estimateur.
Un estimateur est **sans biais** si $E(T) = \theta$.
Un estimateur est **biaisé** si $E(T) \neq \theta$.
Un estimateur est **asymptotiquement sans biais** si $E(T) \longrightarrow \theta$, quand la tailleé n de l'échantillon tend vers l'infini.

Exemple 5.3.2.1. *$\overline{X}$ est sans biais pour μ, mais S^2 est biaisé pour σ^2*
En effet,
$E(S^2) = \frac{n-1}{n}\sigma^2 = (1-\frac{1}{n})\sigma^2$, *le biais est égal à* $\frac{\sigma^2}{n}$
$\lim\limits_{n\to+\infty} E(S^2) = \sigma^2$ *donc S^2 est asymptotiquement sans biais.*
En revanche, la statistique S^{*2} définie par $S^{*2} = \frac{n}{n-1}S^2 = \frac{1}{n-1}\sum\limits_{i=1}^{n}(X_i - \overline{X})^2$ *est un estimateur sans biais pour la variance.*
En effet :

$$E(S^{*^2}) = \frac{n}{n-1}E(S^2) = \sigma^2$$

5.3.3 Variance et erreur quadratique moyenne d'un estimateur

Définition 5.3.3.1. *On appelle erreur quadratique moyenne de T par rapport à θ, la valeur, note $eqm(T)$, définie par :*

$$eqm(\theta) = E[(T-\theta)^2]$$

et l'on a :

$$eqm(\theta) = V(T) + [b(T)]^2$$

En effet,

$$\begin{aligned} E[(T-\theta)^2] &= E[\{T - E(T) + E(T) - \theta\}^2] \\ &= E[\{T - E(T)\}^2] - [(E(T) - \theta)]^2 + 2E[T - E(T)][E(T) - \theta] \\ &= V(T) + [b(T)^2] \ car\ E[T - E(T)] = 0 \end{aligned}$$

Pour rendre l'erreur quadratique moyenne la plus petite possible, il faut que :

1. $E(T) = \theta$, *donc choisir un estimateur sans biais*
2. *Var(T) soit petite.*

Remarque 5.3.3.1. *Si T est sans biais alors $eqm(T, \theta) = Var(T)$ et T convergent si $\lim_{n\to+\infty} Var(T) = 0$. Entre deux estimateurs sans biais, le plus précis est donc celui de variance minimale.*

Meilleur estimateur

Soient T_1, T_2 deux estimateur de θ. On dit que T_1 est meilleur que T_2 si :

$$eqm(T_1, \theta) < eqm(T_2, \theta), \quad \forall \theta \in \Theta$$

5.4 Statistique suffisante (exhaustive)

Dans un problème statistique où figure un paramètre θ inconnu, un échantillon nous apporte certaine information sur ce paramètre. Lorsque l'on résume cet échantillon par une statistique, il s'agit de ne pas perdre cette information. Une statistique qui conserve l'information sera dite suffisante (exhaustive).

Définition 5.4.0.1. *On dit que T est une statistique exhaustive pour $\theta \in \Theta \subset \mathbb{R}^k$ si la loi conditionnelle de $(X_1, X_2, ..., X_n)$ sachant T ne dépend pas de θ.*

Exemple 5.4.0.1. *. Soit $X_1, X_2, ..., X_n$ issu de $X \to \mathcal{P}(\theta)$, θ inconnu. La statistique $T = \sum_{i=1}^{n} X_i$ est-elle exhaustive pour θ ?*

On a :

$$f(x, \theta) = e^{-x}\frac{\theta^x}{x!}$$

la loi de Poisson est stable donc $T = \sum_{i=1}^{n} X_i \longrightarrow \mathcal{P}(n\theta)$.

Notons P_θ *la quantité* $\mathbb{P}_\theta(X_1 = x_1, X_2 = x_2, ..., X_n = x_n | T = t)$

$$\begin{aligned}
P_\theta &= \frac{\mathbb{P}_\theta(X_1 = x_1, X_2 = x_2, ..., X_n = x_n, T = t)}{\mathbb{P}_\theta(T = t)} \\
&= \frac{\mathbb{P}_\theta(X_1 = x_1, X_2 = x_2, ..., X_n = x_n, X_n = t - \sum_{i=1}^{n-1} x_i)}{\mathbb{P}_\theta(T = t)} \\
&= \frac{\mathbb{P}_\theta(X_1 = x_1), ..., \mathbb{P}_\theta(X_n = x_n)}{\mathbb{P}_\theta(T = t)} \\
&= \frac{\prod_{i=1}^{n} f(x_i, \theta)}{f(t, \theta)} \\
&= \frac{e^{-n\theta\theta^{\sum_{i=1}^{n} x_i}}}{\prod_{i=1}^{n} x_i!} \cdot \frac{t!}{e^{-n\theta}(n\theta)^t} \\
&= \frac{t!}{\prod_{i=1}^{n} x_i! n^t} = \frac{(\sum_{i=1}^{n} x_i)!}{\prod_{i=1}^{n} x_i!} \cdot (\frac{1}{n})^{\sum_{i=1}^{n} x_i} \quad \textit{indépendante de } \theta
\end{aligned}$$

Donc T est exhaustive pour θ.

Théorème 5.4.0.1. ***Théorème de Factorisation*** [10]
La statistique $T = t(X_1, X_2, ..., X_n)$ *est exhaustive pour* θ *si et seulement si la densité de probabilité (ou fonction de probabilité) conjointe s'écrit, pour tout* $(x_1, x_2, ..., x_n) \in \mathbb{R}^n$*, sous la forme :*

$$L(x_1, x_2, ..., x_n, \theta) = g(t, \theta)h(x_1, x_2, ..., x_n)$$

Exemple 5.4.0.1. *Reprenons l'exemple précédent où :*

$$\begin{aligned}
L(x, \theta) = \prod_{i=1}^{n} f(x_i, \theta) = \prod_{i=1}^{n} \left(\frac{\theta^{x_i} e^{-\theta}}{x_i!} \right) &= \frac{\theta^t e^{-n\theta}}{\prod_{i=1}^{n} x_i!} \\
&= \frac{(n\theta)^t e^{-n\theta}}{t!} \frac{t!}{n^t (\prod_{i=1}^{n} x_i!)} \\
&= g(t, \theta)h(x)
\end{aligned}$$

D'après le principe de factorisation, T est exhaustive pour θ.

Résultat :

Soit T une statistique exhaustive et $\grave{T}$ une statistique telle que T soit une fonction de $\grave{T}$. Alors $\grave{T}$ est également exhaustive.

Définition 5.4.0.2. *On dit que la statistique T^* est **exhaustive minimale** si elle est exhaustive et si, pour toute statistique exhaustive T, on peut trouver une fonction u telle que $T^* = u(T)$*

5.5 La classe exponentielle de lois

Définition 5.5.0.1. *Soit une famille paramétrique de lois admettant des fonctions de densité (cas continu) ou des fonctions de probabilité (cas discret) $\{f(x,\theta); \theta \in \Theta \subseteq \mathbb{R}^k\}$. On dit qu'elle appartient à la classe exponentielle de lois si $f(x,\theta)$ peut s'écrire :*

$$f(x,\theta) = a(\theta)b(x)\exp\{c_1(\theta)d_1(x) + c_2(\theta)d_2(x) + ... + c_n(\theta)d_k(x)\}$$

pour tout $x \in \mathbb{R}$.
En particulier, si θ est de dimension 1, on a :

$$f(x,\theta) = a(\theta)b(x)\exp\{c(\theta)d(x)\}$$

Exemple 5.5.0.1. 1. *Loi $\mathcal{B}(n,p)$ avec n connu.*

$$\begin{aligned} f(x,n,p) &= C_n^x p^x (1-p)^{n-x}, \text{ pour } x \in= 0,1,2,..,n \\ \ln f(x,n,p) &= \ln C_n^x + x\ln p + (n-x)\ln(1-p) \\ &= \ln C_n^x + n\ln(1-p) + x\ln\frac{p}{n-p} \\ f(x,n,p) &= C_n^x (1-p)^n \exp\left\{x\ln\frac{p}{1-p}\right\}. \end{aligned}$$

d'où $d(x) = x$. Le cas de la loi de Bernoulli $\mathcal{B}(p)$ est identique avec n=1.

2. *Loi $\mathcal{N}(\mu,\sigma^2)$.*

$$\begin{aligned} f(x,\mu,\sigma^2) &= \frac{1}{\sigma\sqrt{2\pi}}\exp\{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\}, x \in \mathbb{R} \\ &= \frac{1}{\sigma\sqrt{2\pi}}\exp\left\{-\frac{\mu^2}{2\sigma^2}\right\}\exp\left\{-\frac{1}{2\sigma^2}x^2 + \frac{\mu}{\sigma^2}x\right\}. \end{aligned}$$

d'où $d_1(x) = x^2$ et $d_2(x) = x$

loi	paramètre	$d_1(x)$	$d_2(x)$
$\mathcal{B}(p)$	p	x	-
$\mathcal{B}(n,p)$	p(n connu)	x	-
$\mathcal{BN}(r,p)$	p(r connu)	x	-
$\mathcal{P}(\lambda)$	λ	x	-
$\mathcal{E}(\lambda)$	λ	x	-
$\Gamma(r,\lambda)$	λ(r connu)	x	-
$\mathcal{N}(\mu,\sigma^2)$	(μ,σ^2)	x^2	x
Pareto (a,θ)	θ(a connu)	$\ln x$	-
Beta (a,θ)	(α,β)	$\ln x$	$\ln(1-x)$

TABLE 5.1 – Principales lois usuelles appartenant à la classe exponentielle

Remarque 5.5.0.1. *En revanche les lois hypergéométriques (M inconnu), Weibull et Gumbel n'appartiennent pas à la classe exponentielle.*

Proposition 5.1. [10] *Soit une loi mère appartenant à une famille paramétrique de la classe exponentielle, avec un paramètre de dimension k. Alors , la statistque de dimension k :*

$$\left(\sum_{i=1}^{n} d_1(X_i), \sum_{i=1}^{n} d_2(X_i), ..., \sum_{i=1}^{n} d_k(X_i)\right)$$

est exhaustive pour le paramètre inconnu.

Exemple 5.5.0.1. *1. Soit la loi de Bernoulli $\mathcal{B}(p)$. On peut écrire sa fonction de probabilité :*

$$\begin{aligned} f(x,p) &= p^x(1-p)^{1-x}, x \in \{0,1\} \\ \ln f(x,p) &= x \ln p + (1-x)\ln(1-p) \\ &= x \ln \frac{p}{1-p} + \ln(1-p) \\ f(x,p) &= (1-p)\exp x \ln \frac{p}{1-p} \\ &\Rightarrow d(x) = x \end{aligned}$$

d'après la proposition, $\sum_{i=1}^{n} x_i$ est une statistique exhaustive minimale pour p.

2. Soit $(X_1, X_2, ..., X_n)$ un n-échantillon issu de $X \to \gamma(\theta)$

$$f(x,\theta) = \frac{1}{\Gamma(\theta)} e^{-x} x^{\theta-1}, x > 0$$

$$\begin{aligned} \ln f(x,\theta) &= -x + (\theta-1)\ln x - \ln \Gamma(\theta) \\ &= -x e^{-x} \Gamma(\theta) \exp\{\theta \ln x\} \end{aligned}$$

d'où $d(x) = \ln x$, donc $\sum_{i=1}^{n} \ln Xi = \ln \prod_{i=1}^{n} X_i$ est exhaustive minimale.

Remarque 5.5.0.2. *Lorsque le domaine de X dépend de θ, Cette proposition ne s'applique pas, ce qui n'empêche pas de trouver une statistique exhaustive.*

Exemple 5.5.0.2. $X \rightsquigarrow U_{[0,\theta]} \Rightarrow f(x,\theta) = \frac{1}{\theta}, x \in [0,\theta]$.
$T = \sup X_i$ *est exhaustive pour θ.*

$$\begin{aligned} L(x_1, ..., x_n, \theta) &= \prod_{i=1}^{n} f(x_i, \theta) \\ &= \prod_{i=1}^{n} \frac{1}{\theta} \mathbb{1}_{[0,\theta]} \\ &= \frac{1}{\theta^n} \mathbb{1}_{[\sup x_i, +\infty[}(\theta) \end{aligned}$$

Comme $G(t) = \mathbb{P}(\sup X_i \leq t) = (F(t))^n = (\frac{t}{n})^n$

$$g(t,\theta) = n(\frac{t}{\theta})^{n-1}\frac{1}{\theta} = n\frac{t^{n-1}}{\theta^n}$$

$$L(x_1,..,x_n,\theta) = \frac{1}{\theta^n} \Rightarrow \frac{L}{g} = \frac{1}{nt^{n-1}} \textit{indépendant de } \theta.$$

donc $T = \sup X_i$ *est exhaustive pour* θ.

5.5.1 Recherche des meilleurs estimateurs sans biais

Définition 5.5.1.1. *On dit que l'estimateur* T^* *est* ***UMVUE*** *pour* θ *(uniformly minimum variance unbiased estimator) s'il est sans biais pour* θ *et si pour tout autre estimateur* T *sans biais on a :*

$$Var_\theta(T^*) \leq Var_\theta(T) \textit{ pour tout } \theta \in \Theta$$

Proposition 5.2. [14] *Si la famille de la loi mère appartient à la classe exponentielle avec paramètre de dimension 1* ($\Theta \subset \mathbb{R}$) *et s'il existe une statistique fonction de la statistique exhaustive minimale* $\sum_{i=1}^{n} d(X_i)$ *qui soit sans biais pour* θ*, alors elle est unique et elle est UMVUE pour* θ.

Exemple 5.5.1.1.

1. *Bernoulli* $\mathcal{B}(p)$, $T = \overline{X}$ *est UMVUE.*
2. *Poisson* $\mathcal{P}(\lambda)$, $T = \overline{X}$ *est UMVUE.*
3. *Pour* $f(x,\lambda) = \lambda e^{-\lambda x}$ *si* $x \geq 0$ *(0 sinon).* $T = \frac{n-1}{\sum_{i=1}^{n} X_i}$ *est l'estimateur UMVUE de* λ.
4. *Pour la loi de Pareto avec a connu on a* $d(x) = \ln x$ *et* $T = \frac{n-1}{\sum_{i=1}^{n} \ln(\frac{X_i}{a})}$ *est l'estimateur UMVUE pour* θ.

5.5.2 Statistique complète

Définition 5.5.2.1. *On dit qu'une statistique* T *est complète pour une famille de loi de probabilité* $(\mathbb{P}_\theta)_{\theta\in\Theta}$ *si pour toute fonction borellienne* h *on ait :*

$$\forall \theta \in \Theta : E(h(T)) = 0 \Rightarrow h = 0, \mathbb{P}_\theta.p.s$$

En particulier la statistique exhaustive de famille exponentielle est complète.

Exemple 5.5.2.1. *Pour $X \to \mathcal{P}(\lambda)$, λ inconnu.*
$T = \sum_{i=1}^{n} X_i$ *est complète.*

$$\mathbb{E}(h(T)) = \sum_{t=0}^{\infty} h(t) e^{-n\lambda} \frac{(n\lambda)^t}{t!} \ car\ T \to \mathcal{P}(n\lambda)$$

$$\mathbb{E}(h(T)) = 0, \forall \lambda \Rightarrow h(t) = 0, \forall t \in \mathbb{N}$$

5.6 L'information de Fisher

Définition 5.6.0.1. *On appelle quantité d'information de Fisher $I_n(\theta)$ apportée par un échantillon sur le paramètre θ, la quantité suivante positive ou nulle (si elle existe) :*

$$I_n(\theta) = \mathbb{E}\left[\left(\frac{\partial \ln L(\underline{X}, \theta)}{\partial \theta}\right)^2\right]$$

avec $L(\underline{X}, \theta) = \prod_{i=1}^{n} f(X_i, \theta)$

Théorème 5.6.0.1. *Si le domaine de définition de X ne dépend pas de θ alors :*

$$I_n(\theta) = -E\left[\frac{\partial^2 \ln L(\underline{X}, \theta)}{\partial \theta^2}\right]$$

Démonstration 5.6.0.1. *On a $L(\underline{x}, \theta)$ est une densité de probabilité :*

$$\int_{\mathbb{R}^n} L(\underline{x}, \theta) d\underline{x} = 1$$

On a

$$\frac{\partial \ln L(\underline{x}, \theta)}{\partial \theta} = \frac{\partial L(\underline{x}, \theta)}{\partial \theta L(\underline{x}, \theta)}$$

donc

$$\frac{\partial L(\underline{x}, \theta)}{\partial \theta} = L(\underline{x}, \theta). \frac{\partial \ln L(\underline{x}, \theta)}{\partial \theta}$$

$$\int_{\mathbb{R}^\times} \frac{\partial L(\underline{x}, \theta)}{\partial \theta} d\underline{x} = 0 \Rightarrow \int_{\mathbb{R}^\times} L(\underline{x}, \theta) \frac{\partial \ln L(\underline{x}, \theta)}{\partial \theta} d\underline{x} = 0$$

Ce qui prouve que la variable aléatoire $\frac{\partial \ln L(\underline{X}, \theta)}{\partial \theta}$ est centrée i.e :

$$E\left[\left(\frac{\partial \ln L(\underline{X}, \theta)}{\partial \theta}\right)\right] = 0$$

Donc

$$I_n(\theta) = E\left[\left(\frac{\partial \ln L(\underline{x}, \theta)}{\partial \theta}\right)^2\right] = Var\left[\left(\frac{\partial \ln L(\underline{x}, \theta)}{\partial \theta}\right)\right]$$

En dérivant la deuxième fois :

$$\int_{\mathbb{R}^n} L(\underline{x},\theta)\frac{\partial^2 \ln L(\underline{x},\theta)}{\partial^2\theta}d\underline{x} + \int_{\mathbb{R}^n} \frac{\partial \ln L(\underline{x},\theta)}{\partial\theta}\frac{\partial L(\underline{x},\theta)}{\partial\theta}d\underline{x} = 0$$

$$\int_{\mathbb{R}^n} L(\underline{x},\theta)\frac{\partial^2 \ln L(\underline{x},\theta)}{\partial^2\theta}d\underline{x} + \int_{\mathbb{R}^n} \left(\frac{\partial \ln L(\underline{x},\theta)}{\partial\theta}\right)^2 L(\underline{x},\theta)d\underline{x} = 0$$

$$I_n(\theta) = -\int_{\mathbb{R}^n} \frac{\partial^2 \ln L(\underline{x},\theta)}{\partial\theta^2} L(\underline{x},\theta)dx$$

$$I_n(\theta) - E\left[\frac{\partial^2 \ln L(\underline{X},\theta)}{\partial\theta^2}\right]$$

5.7 Propriétés de la quantité d'information

5.7.1 L'additivité

Soit Y une variable aléatoire indépendante de X dont la loi dépend du même paramètre θ. Soit $f(x,\theta), g(y,\theta), h(x,y,\theta)$ les densités de X, Y et du couple (X,Y) (resp.), d'information de Fisher $I_X(\theta), I_Y(\theta), I(\theta)$ (resp.)
Si les domaines de définition de X, Y ne dépendent pas de θ alors :

$$I(\theta) = I_X(\theta) + I_Y(\theta)$$

Démonstration 5.7.1.1. *Puisque X, Y sont indpendantes :*

$$h(x,y,\theta) = f(x,\theta)g(y,\theta)$$

$$\ln h(x,y,\theta) = \ln f(x,\theta) \ln g(y,\theta)$$

$$\frac{\partial^2 \ln h(x,y,\theta)}{\partial\theta^2} = \frac{\partial^2 \ln f(x,\theta)}{\partial\theta^2}.\frac{\partial^2 \ln g(y,\theta)}{\partial\theta^2}$$

$$E\left(\frac{\partial^2 \ln h(X,Y,\theta)}{\partial\theta^2}\right) = E\left(\frac{\partial^2 \ln f(X,\theta)}{\partial\theta^2}\right).E\left(\frac{\partial^2 \ln g(Y,\theta)}{\partial\theta^2}\right)$$

$$-I(\theta) = -I_X(\theta) - I_Y(\theta) \Rightarrow I(\theta) = I_X(\theta) + I_Y(\theta)$$

Conséquence : Si le domaine de définition ne dépend pas de θ alors : $I_n(\theta) = nI_1(\theta)$ avec

$$I_1(\theta) = -E\left[\frac{\partial^2 \ln f(X,\theta)}{\partial\theta^2}\right]$$

i.e : que chaque observation a la même importance. Ce qui n'est pas le cas pour la loi uniforme sur $[0,\theta]$ où la plus grande observation est la plus intéressante

Exemple 5.7.1.1. *Soit $X \rightsquigarrow \mathcal{N}(\mu, \sigma^2)$. Calculer l'information de Fisher apportée par un n-échantillon issu de X sur le paramètre μ.*

$$I_n(\mu) = n.I_1(\mu)$$

$$I_1(\mu) = -E\left[\frac{\partial^2 \ln f(X,\mu)}{\partial \mu^2}\right]$$

$$f(x,\mu) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$$

$$\frac{\partial \ln f(x,\mu)}{\partial \mu} = \frac{1}{\sigma^2}(x-\mu)$$

$$\frac{\partial^2 \ln f(x,\mu)}{\partial \mu^2} = \frac{-1}{\sigma^2} \Rightarrow I_n(\mu) = \frac{n}{\sigma^2}$$

5.7.2 Dégradation de l'information

On montre que l'information portée par une statistique est inférieure ou égale à celle apportée par un échantillon.
En effet :
soit T la statistique de densité $g(t,\theta)$ que l'on substitue à l'échantillon.
On a
$L(x,\theta) = g(t,\theta)h(x,\theta/t)$ où $h(x,\theta/t)$ est la densité conditionnelle de l'échantillon.

$$\ln L(x,\theta) = \ln g(t,\theta) + \ln h(x,\theta/t)$$

$$\frac{\partial^2 \ln L(x,\theta)}{\partial \theta^2} = \frac{\partial^2 \ln g(t,\theta)}{\partial \theta^2} + \frac{\partial^2 \ln h(x,\theta/t)}{\partial \theta^2}$$

En prennant l'espérance mathématique :

$$I_n(\theta) = I_T(\theta) - E\left(\frac{\partial^2 \ln h(x,\theta/t)}{\partial \theta^2}\right) = I_T(\theta) + I_{n/T}(\theta)$$

D'où

$$I_T(\theta) \leq I_n(\theta) \text{ car } I_{n/T}(\theta) \geq 0$$

Si T est suffisante $I_n(\theta) = I_T(\theta)$.
La réciproque est vraie si le domaine de X est indépendant de θ.

5.8 Borne de Cramer-Rao et estimateurs efficaces

Sous certaines conditions de régularité, à la fois pour la famille étudiée et pour l'estimateur sans biais considéré, on peut montrer que sa variance ne

peut descendre au-dessous d'un certain seuil qui est fonction de θ. Ce seuil, appelé **borne de Cramer-Rao**, est intrinsèque à la forme de la densité (ou de la fonction de probabilité) $f(x, \theta)$. L'intérêt de ce résultat est que, si l'on trouve un estimateur sans biais dont la variance atteint ce seuil, alors il est le meilleur possible (UMVUE) parmi les estimateurs sans biais «réguliers».

Théorème 5.8.0.1. ***(Inégalité de Cramer-Rao ou de Fréchet)***
Soit T un estimateur sans biais pour θ de dimension 1. Sous certaines conditions de régularité on a nécessairement, pour tout $\theta \in \Theta$:

$$V_\theta(T) \geq \frac{1}{I_n(\theta)}$$

Si T est un estimateur sans biais de $h(\theta)$:

$$V_\theta(T) \geq \frac{[h^{'}(\theta)]^2}{I_n(\theta)}$$

Remarque 5.8.0.1. *Les conditions de régularité, dans le cas continu, sont les suivantes :*

1. *$I(\theta)$ existe pour tout $\theta \in \Theta$.*
2. *la dérivée par rapport à θ d'une intégrale sur la densité conjointe*
$$\int \cdots \int f(x_1, x_2, ..., x_n, \theta) dx_1 dx_2 dx_n$$
peut s'obtenir en dérivant à l'intérieur de l'intégrale.
3. *la dérivée par rapport à θ de $E_\theta(T)$ peut s'obtenir en dérivant à l'intérieur de l'intégrale correspondante.*
4. *le support de $f(x, \theta)$ est indépendant de θ.*

Dans le cas discret les conditions portent sur les sommations en lieu et place des intégrations.

Démonstration 5.8.0.1.

$$\begin{aligned}
Cov(T, \frac{\partial \ln L(\underline{x}, \theta)}{\partial \theta}) &= E\left(T \frac{\partial \ln L(\underline{x}, \theta)}{\partial \theta}\right) \text{ puisque } \frac{\partial \ln L(\underline{x}, \theta)}{\partial \theta} \text{ est centrée} \\
&= \int_{\mathbb{R}^n} t \frac{\partial \ln L(\underline{x}, \theta)}{\partial \theta} L(\underline{x}, \theta) d\underline{x} \\
&= \int_{\mathbb{R}^n} t \frac{\partial L(\underline{x}, \theta)}{\partial \theta} d\underline{x} \\
&= \frac{d}{d\theta} \int_{\mathbb{R}^n} t L(\underline{x}, \theta) d\underline{x} \\
&= \frac{d}{d\theta} E(T) \\
&= h^{'}(\theta)
\end{aligned}$$

D'autre part l'inégalité de Schwartz donne :

$$\left[Cov(T, \frac{\partial \ln L(\underline{x}, \theta)}{\partial \theta})\right]^2 \leq Var(T)Var(\frac{\partial \ln L(\underline{x}, \theta)}{\partial \theta})$$

c'est à dire :

$$\left[h^{'}(\theta)\right]^2 \leq Var(T)I_n(\theta)$$

5.9 Estimateur efficace

Définition 5.9.0.1. *On dit qu'un estimateur T est efficace si sa variance est égale à la borne de Cramer Rao.*

$$V_\theta(T) = \frac{1}{I_n(\theta)}$$

Si T est un estimateur sans biais de $h(\theta)$:

$$V_\theta(T) = \frac{[h^{'}(\theta)]^2}{I_n(\theta)}$$

Propriété :

Un estimateur sans biais efficace est convergent.

En effet :

T efficace $\Rightarrow V(T) = \frac{[h^{'}(\theta)]^2}{I_n(\theta)}$ or $I_n(\theta) = n.I_1(\theta)$ et T sans biais.

$$\begin{aligned}
\lim_{n\to\infty} eqm(T,\theta) &= \lim_{n\to\infty} Var(T,\theta) \text{ puisque T est san biais} \\
&= \lim_{n\to\infty} \frac{[h^{'}(\theta)]^2}{I_n(\theta)} \\
&= \lim_{n\to\infty} \frac{[h^{'}(\theta)]^2}{n.I_1(\theta)} \\
&= 0
\end{aligned}$$

Donc T convergent.

Remarque 5.9.0.1. *Un estimateur efficace T est un estimateur sans biais de variance minimale.*

Proposition 5.3. [10] *La borne de Cramer-Rao n'est atteinte que :*

- *Si la famille de lois est dans la classe exponentielle.*
- *et pour l'estimation d'une fonction de reparamétrisation particulière de θ, à savoir $h(\theta) = E_\theta(\sum_{i=1}^{n} d(X_i))$.*

Nous admettrons cette proposition. Ainsi il n'existe qu'une fonction de θ qui puisse être estimée de façon «efficace». Pour déterminer cette fonction il suffit de calculer l'espérance mathématique de $\sum_{i=1}^{n} d(X_i)$ qui en est donc l'estimateur efficace.

Exemple 5.9.0.1. *Soit à estimer le paramètre λ dans la famille des lois $E(\lambda)$ de densités $f(x,\lambda) = \lambda e^{-\lambda x}$ pour $x \geq 0$. Déterminons la borne de Cramer-Rao. On a :*

$$\ln f(x,\lambda) = \ln\lambda - \lambda x$$
$$\frac{\partial}{\partial\lambda}\ln f(x,\lambda) = \frac{1}{\lambda} - x$$
$$\frac{\partial^2}{\partial\lambda^2}\ln f(x,\lambda) = -\frac{1}{\lambda^2}$$

D'où

$$I_1(\lambda) = -E(-\frac{1}{\lambda^2}) = \frac{1}{\lambda^2}$$
$$I_n(\lambda) = nI_1(\lambda) = \frac{n}{\lambda^2}$$

et la borne de Cramer-Rao est donc égale à $\frac{\lambda^2}{n}$.

La borne de Cramer-Rao ne peut être atteinte que pour estimer la fonction de λ pour laquelle $\sum_{i=1}^{n} d(X_i) = \sum_{i=1}^{n} X_i$ est sans biais, à savoir $E(\sum_{i=1}^{n} X_i) = nE(X) = \frac{n}{\lambda}$. En reparamétrant avec $\theta = h(\lambda) = \frac{1}{\lambda}$, θ est alors la moyenne de la loi et $\overline{X}$ est l'estimateur qui atteint la borne de Cramer-Rao pour θ. Celle-ci est

$$\frac{[h'(\lambda)]^2}{I_n(\lambda)} = \frac{(-1/\lambda^2)^2}{n(1/\lambda^2)} = \frac{1}{n\lambda^2} = \frac{\theta^2}{n}$$

On a

$$V(\overline{X}) = \frac{1}{n}V(X) = \frac{\theta^2}{n} \text{ puisque } \theta^2 = \frac{1}{\lambda^2} \text{ est la variance de la loi.}$$

et donc $\overline{X}$ est un estimateur efficace pour θ.

5.10 Généralisation (Cas multidimensionnel)

Statistique exhaustive

Soit le modèle statistique paramétrique : $(\mathcal{X}, (\mathbb{P}_\theta)_{\theta\in\Theta})$, avec $\theta \in \Theta \subset \mathbb{R}^p$. Soit $(X_1, X_2, ..., X_n)$ un n-échantillon issu de $X \rightsquigarrow f(x,\theta)$, on suppose que

$f(x,\theta)$ appartienne à la famille exponentielle :

$$f(x,\theta) = a(\theta)b(x).\exp\sum_{j=1}^{p} c_j(\theta)d_j(x)$$

Si $X(\Omega)$ ne dépend pas de θ.
La statistique $(\sum_{i=1}^{n} d_1(X_i), \sum_{i=1}^{n} d_2(X_i), \cdots, \sum_{i=1}^{n} d_p(X_i))$ est exhaustive pour θ.

Information de Fisher

Soit $X \rightsquigarrow f(x,\theta), \theta \in \Theta \subset \mathbb{R}^p$. On note $\theta = (\theta_1, \theta_2, ..., \theta_p)$
On fait les hypothèses de régularité suivante :

$H_1) \forall x, \forall\theta, f(x,\theta) > 0.$

$H_2) \nabla_\theta f$ existe , *i.e* : on peut dériver f par rapport à $\theta_i, \forall i = 1, \cdots, p.$ $\mathbb{P}_\theta.p.s.$

$H_3)$ On peut dériver au moins deux fois $f(x,\theta)$ par rapport à $\theta_i, \forall i = 1, p$ dériver $\int f(x,\theta)dx$ et permuter entre dérivation et intégration.

Remarque 5.10.0.1. $\nabla_\theta f = \left(\frac{\partial f}{\partial\theta_1}, \frac{\partial f}{\partial\theta_2}, \cdots, \frac{\partial f}{\partial\theta_n}\right)$ *vecteur ligne* $(1,p)$. *On appelle score* $S(x,\theta)$ *le vecteur* $(1,p)$ *défini par* $\nabla_\theta \log f$.

Définition 5.10.0.1. *On appelle information en* θ *la matrice (p, p) de variance-covariance de* $\nabla_\theta \log f(x,\theta)$.

$$I(\theta) = E(S^{'}S)$$

Sous les hypothèses H_1, H_2, H_3 *on obtient,* $I(\theta)$ *: la matrice dont le terme général* $I_{i,j}(\theta)$,

$$I_{i,j}(\theta) = -E\left(\frac{\partial^2 \ln f(x,\theta)}{\partial\theta_i\partial\theta_j}\right)$$

$I(\theta)$ *est une matrice définie positive.*

Exemple 5.10.0.1. *Prenons le cas de la loi* $\mathcal{N}(\mu,\sigma^2)$, *où* (μ,σ^2) *inconnus.On a, en posant* $\upsilon = \sigma^2$ *:*

$$f(x,\mu,\upsilon) = \frac{1}{\sqrt{2\pi\upsilon}} e^{-\frac{1}{2\upsilon}(x-\mu)^2}$$

$$\ln f(x,\mu,\upsilon) = -\frac{1}{2}\ln 2\pi - \frac{1}{2}\ln\upsilon - \frac{1}{2\upsilon}(x-\mu)^2$$

$\frac{\partial}{\partial\mu}\ln f(x,\mu,\upsilon) = \frac{1}{\upsilon}(x-\mu)$ $\qquad$ $\frac{\partial}{\partial\upsilon}\ln f(x,\mu,\upsilon) = -\frac{1}{2\upsilon} + \frac{(x-\mu)^2}{2\upsilon^2}$
En position (1,1) de la matrice $I(\mu,\sigma^2)$ *on trouve :*

$$-E\left(\frac{\partial^2}{\partial\mu^2}\ln f(x,\mu,\upsilon)\right) = \frac{1}{\upsilon} = \frac{1}{\sigma^2}$$

et en position (2,2) :

$$\frac{\partial}{\partial \upsilon} \ln f(x,\mu,\upsilon) = -\frac{1}{2\upsilon} + \frac{1}{2\upsilon^2}(x-\mu)^2$$

$$\frac{\partial^2}{\partial \upsilon^2} \ln f(x,\mu,\upsilon) = \frac{1}{2\upsilon^2} - \frac{1}{\upsilon^3}(x-\mu)^2$$

$$-E\left(\frac{\partial^2}{\partial \upsilon^2} \ln f(x,\mu,\upsilon)\right) = \frac{-1}{2\upsilon^2} + \frac{1}{\upsilon^2} E\left(\frac{(x-\mu)^2}{\upsilon}\right) = -\frac{1}{2\upsilon^2} + \frac{1}{\upsilon^2} = \frac{1}{2\upsilon^2} = \frac{1}{2\sigma^4}$$

En position (1,2) ou (2,1) on a :

$$\frac{\partial}{\partial \mu} \ln f(x,\mu,\upsilon) = \frac{1}{\upsilon}(x-\mu)$$

$$\frac{\partial^2}{\partial \mu \partial \upsilon} \ln f(x,\mu,\upsilon) = \frac{-1}{\upsilon^2}(x-\mu)$$

$$-E\left(\frac{\partial^2}{\partial \mu \partial \upsilon} \ln f(x,\mu,\upsilon)\right) = 0$$

$$I_1(\theta) = \begin{pmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{1}{2\sigma^4} \end{pmatrix} \Rightarrow I_n(\theta) = \begin{pmatrix} \frac{n}{\sigma^2} & 0 \\ 0 & \frac{n}{2\sigma^4} \end{pmatrix}$$

5.11 Méthodes d'estimation

Les paramètres les plus fréquents à estimer sont la moyenne et la variance, qui sont respectivement estimés par la moyenne et la variance empiriques. Dans le cas où il n'y a pas d'estimateur évident, on en utilise quelque méthodes, en prend comme exemples la méthode du maximum de vraisemblance et la méthode des moments.

5.11.1 Méthode du maximum de vraisemblance

Définition 5.11.1.1. *Soit un échantillon aléatoire $(X_1, X_2, ..., X_n)$ dont la loi mère appartient à une famille paramétrique de densités (ou fonctions de probabilité) $\{f(x,\theta), \theta \in \Theta\}$ où $\Theta \subset \mathbb{R}^k$. On appelle* ***fonction de vraisemblance*** *de θ pour une réalisation donnée $(x_1, x_2, ..., x_n)$ de l'échantillon, la fonction de θ :*

$$L(x_1, x_2, .., x_n, \theta) = \prod_{i=1}^{n} f(x_i, \theta)$$

Remarque 5.11.1.1. *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 5.11.1.2. *On appelle* ***estimation du maximum de vraisemblance*** *une valeur* $\hat{\theta}^{MV}$ *, s'il en existe une, telle que :*

$$L(\hat{\theta}^{MV}) = \sup_{\theta \in \Theta} L(\theta)$$

Une telle solution est fonction de $(x_1, x_2, ..., x_n)$*, soit* $\hat{\theta}^{MV} = h(x_1, x_2, ..., x_n)$*. Cette fonction* h *induit la statistique* $\hat{\theta}^{MV} = h(X_1, X_2, ..., X_n)$ *appelée estimateur du maximum de vraisemblance (EMV).*

Remarque 5.11.1.2. *Quand les densités conjointes sont des produits de fonctions puissances et exponentielles, on a plutôt intérêt à maximiser* $\ln L(\theta)$*, appelée log-vraisemblance.*

Cette méthode consiste à résoudre :

1.

$$\frac{\partial}{\partial \theta} \ln L(\theta) = 0$$

$$où \ \frac{\partial}{\partial \theta} \ln \prod_{i=1}^{n} f(x_i, \theta) = 0$$

$$où \ \sum_{i=1}^{n} \frac{\partial}{\partial \theta} \ln f(x_i, \theta) = 0$$

Cette dernière égalité s'appelle ***l'équation de vraisemblance****.*

2.

$$\frac{\partial^2}{\partial \theta^2} \ln L(\theta) < 0$$

pour assurer l'existence du maximum de vraisemblance (EMV).

5.11.2 Exemples et propriétés

Exemple 5.11.2.1. *1. On souhaite estimer le paramètre d'une loi de Poisson à partir d'un n échantillon. On a*

$$\mathbb{P}(x, \lambda) = \mathbb{P}_\lambda(X = x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

La fonction de vraisemblance s'écrit :

$$L(x_1, x_2, ..., x_n, \lambda) = \prod_{i=1}^{n} e^{-\lambda} \frac{\lambda^{x_i}}{x_i!} = e^{-n\lambda} \prod_{i=1}^{n} \frac{\lambda^{x_i}}{x_i!}$$

$$\begin{aligned}\ln L(x_1, x_2, ..., x_n, \lambda) &= -\lambda n + \sum_{i=1}^{n} \ln \frac{\lambda^{x_i}}{x_i!} \\ &= -\lambda n + \ln \lambda \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} \ln(x_i!)\end{aligned}$$

La dérivée première du logarithme de la vraisemblance est

$$\begin{aligned}\frac{\partial \ln L(x_1, x_2, ..., x_n, \lambda)}{\partial \lambda} &= -n + \frac{\sum_{i=1}^{n} x_i}{\lambda} \\ &\Rightarrow \quad n + \frac{\sum_{i=1}^{n} x_i}{\lambda} = 0 \\ &\Rightarrow \quad \lambda = \frac{1}{n} \sum_{i=1}^{n} x_i\end{aligned}$$

La dérivée seconde

$$\frac{\partial^2 \ln L(x_1, x_2, ..., x_n, \lambda)}{\partial \lambda^2} = \frac{-\sum_{i=1}^{n} x_i}{\lambda^2} \leq 0$$

Donc l'EMV

$$\hat{\lambda}^{MV} = \frac{1}{n} \sum_{i=1}^{n} x_i = \overline{X}$$

2. *Soit $(X_1, X_2, ..., X_n)$ un n-échantillon issu de $X \to \mathcal{N}(\mu, \sigma^2)$, trouver l'estimateur EMV de μ et σ^2.*

La densité de X est :

$$\begin{aligned}
f(x,\mu,\sigma^2) &= \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-1}{2\sigma^2}(x-\mu)^2}, x>0 \\
L(x_1,x_2,....x_n,\mu,\sigma^2) &= \prod_{i=1}^{n} f(x_i,\mu,\sigma^2) \\
&= \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n e^{\frac{-1}{2}\sum_{i=1}^{n}\left(\frac{(x_i-\mu)}{\sigma}\right)^2} \\
\ln L(x,\mu,\sigma^2) &= -n\ln(\sigma\sqrt{2\pi}) - \frac{1}{2}\sum_{i=1}^{n}(\frac{x_i-\mu}{\sigma})^2 \\
\frac{\partial \ln L(x,\mu,\sigma^2)}{\partial\mu} &= \sum_{i=1}^{n}(\frac{x_i-\mu}{\sigma})^2 = 0 \\
&\Rightarrow \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu)=0 \\
&\Rightarrow \sum_{i=1}^{n}(x_i-\mu)=0 \Rightarrow n\mu=\sum_{i=1}^{n}x_i \\
&\Rightarrow \hat{\mu}=\frac{1}{n}\sum_{i=1}^{n}x_i \\
\frac{\partial^2 \ln L(x,\mu,\sigma^2)}{\partial\mu^2} &= \frac{-1}{\sigma^2}<0
\end{aligned}$$

d'où l'EMV

$$\hat{\mu}^{MV}=\frac{1}{n}\sum_{i=1}^{n}x_i=\overline{X}$$

Proposition 5.4. *Si T est une statistique exhaustive pour θ alors $\hat{\theta}^{MV}$, s'il existe, est fonction de T.*

Ceci résulte immédiatement du théorème de factorisation.
En effet :
$L(x,\theta)=g(t,\theta)h(x)$ et résoudre $\frac{\partial}{\partial\theta}\ln L(x,\theta)=0$ revient à résoudre $\frac{\partial}{\partial\theta}\ln g(t,\theta)=0$, donc $\hat{\theta}=f(t)$

Proposition 5.5. [14] *Si la famille de lois considérée répond à certaines conditions de régularité et si elle admet un estimateur sans biais efficace pour θ, alors l'EMV existe et est cet estimateur.*

5.11.3 Caractéristiques de l'EMV

1. $L(\underline{x},\theta)$ **n'a aucune raison d'être différentiable en** θ.

Exemple 5.11.3.1. *Soit* $X = (X_1, X_2, ..., X_n) \rightsquigarrow \mathcal{U}[0, \theta]$, *alors :*

$$\begin{aligned} L(x,\theta) &= \prod_{i=1}^{n} \frac{1}{\theta} \mathbb{1}_{[0,\theta]}(x_i) \\ &= \frac{1}{\theta^n} \mathbb{1}_{\inf x_i \geq 0} \mathbb{1}_{\sup x_i \leq \theta} \\ \ln L(x,\theta) &= -n \ln \theta \mathbb{1}_{[0,\theta]}(x_i) \end{aligned}$$

$+$ *Alors*

$$\frac{\partial \ln L(x_1, x_2, ..., \theta)}{\partial \theta} = -\frac{n}{\theta} = 0 \tag{5.1}$$

Mais l'équation (5.1) n'a pas de sens on peut écrire la fonction de vraisemblance en fonction de θ *sous la forme*

$$\frac{1}{\theta^n} \mathbb{1}_{[\sup x_i, \infty[}(\theta)$$

Qui admet une maximum pour $\theta = \sup x_i$ *donc*

$$\hat{\theta} = \sup x_i$$

Est l'estimateur du maximum de vraisemblance de θ

2. **L'EMV n'est pas forcement sans biais, donc pas forcement efficace.**

Exemple 5.11.3.2. *Reprenons l'exemple précédent :*
$Y_n = \sup X_i$ *où :*

$$\begin{aligned} F_{Y_n}(y) &= \mathbb{P}(Y_n \leq y) = \frac{y^n}{\theta^n}, \quad y \in [0, \theta] \\ f_{Y_n}(y) &= \frac{n y^{n-1}}{\theta^n} \\ E(Y_n) &= \int_0^\theta \frac{n y^n}{\theta^n} dy = \frac{n}{n+1}\theta \neq \theta. \end{aligned}$$

Y_n *est baisé.*
On en déduit un estimateur sans biais :

$$\hat{Y}_n = \frac{n+1}{n} \sup X_i$$

$$\begin{aligned} Var(\widehat{Y}_n) &= Var(\frac{n+1}{n} \sup X_i) = \frac{(n+1)^2}{n^2} Var(\sup X_i) \\ Var(\sup X_i) &= Var(Y_n) = E(Y_n^2) - E^2(Y_n) \\ E(Y_n^2) &= \frac{n}{\theta^n}\eta \int_0^\theta y^{n+1} dy = \frac{n}{n+1}\theta^2 \\ Var(Y_n) &= \frac{n}{n+2}\theta^2 - \frac{(n+1)^2}{n^2}\theta^2 \\ &= \frac{n}{(n+2)(n+1)^2}\theta^2 \end{aligned}$$

$$Var(\frac{n+1}{n}\sup X_i) = \frac{\theta^2}{n(n+2)} \to 0 \ quand \ n \to \infty$$

3. **L'EMV n'est pas unique.**

   Exemple 5.11.3.3. $\mathcal{U}_{[\theta,\theta+1]}, \theta > 0$. *On considère un n-échantillon issu de X :*

$$\begin{aligned} L(\underline{x},\theta) &= \mathbb{1}_{[\theta \leq \inf x_i \leq \sup x_i \leq \theta+1]} \\ &= \mathbb{1}_{[\theta \leq \inf x_i]} \mathbb{1}_{[\theta \geq \sup x_i - 1]} \end{aligned}$$

θ			
$\mathbb{1}_{[\theta \leq \inf x_i]}$	1	1	0
$\mathbb{1}_{[\theta \geq \sup x_i - 1]}$	0	1	1
L	0	1	0

   Tout estimateur $\hat{\theta}$ compris entre $\hat{\theta}_1 = supX_i - 1$ et $\hat{\theta}_2 = infX_i -$ est un EMV. $\hat{\theta}$ est unique si $supX_i = infX_i + 1$.

Comportement asymptotique de l'EMV

Proposition 5.6. *Soit l'échantillon $X_1, X_2, ..., X_n$ issu de la densité (ou fonction de probabilité) $f(x,\theta)$ où $\Theta \subset \mathbb{R}$, répondant à certaines conditions de régularité qui garantissent notamment l'existence d'un EMV $\hat{\theta}_n^{MV}$ pour tout n. On considère la suite $\{\hat{\theta}_n^{MV}\}$ quand n croît à l'infini. Alors cette suite est telle que :*

$$\sqrt{n}(\hat{\theta}_n^{MV} - \theta) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \frac{1}{I(\theta)})$$

Ceci résulte immédiatement d'une application indirecte de la loi des grands nombres et du théorème central limite.

Le résultat énoncé dans cette proposition implique les propriétés suivantes :

1. $\hat{\theta}_n^{MV}$ est asymptotiquement sans biais, i.e. $E(\hat{\theta}_n^{MV}) \longrightarrow_{n \longrightarrow \infty} \theta$.
2. pour n tendant vers l'infini, la variance de $\hat{\theta}_n^{MV}$ se rapproche de $1/(nI(\theta))$. On dit que $\hat{\theta}_n^{MV}$ est asymptotiquement efficace.
3. $\hat{\theta}_n^{MV}$ converge vers θ en moyenne quadratique (avec un choix adéquat de conditions de régularité on démontre que $\widehat{\theta}_n^{MV}$ converge presque s$\widehat{u}$rement p.s).

4. $\hat{\theta}_n^{MV}$ tend à devenir gaussien quand n s'accroît.

Résumé :
L'EMV est un estimateur BAN (Best Asymptotically Normal).

5.12 Méthode des moments

Soit à considérer le cas d'un paramètre à une dimension. Pour une réalisation $x_1, x_2, ..., x_n$ de l'échantillon la méthode consiste alors à choisir pour estimation de θ la valeur telle que la moyenne théorique $\mu(\theta)$ (ou premier moment de la loi) coincide avec la moyenne empirique $\bar{x}$.

Exemple 5.12.0.1. *1. Pour la loi $E(\lambda)$, par exemple, l'estimation de λ sera $\hat{\lambda}^M$ telle que $\frac{1}{\hat{\lambda}^M} = \bar{x}$, soit $\hat{\lambda}^M = \frac{1}{\bar{x}}$.*

2. Pour la loi $\mathcal{BN}(r,p)$ avec r connu, l'estimation de p sera $\hat{p}^M$ telle que

$$\frac{r(1-\hat{p}^M)}{\hat{p}^M} = \bar{x} \text{ d'où } \hat{p}^M = \frac{r}{r+\bar{x}}$$

Pour un paramètre de dimension 2 l'estimation résulte de la résolution de deux équations, l'une reposant sur le premier moment, l'autre sur le moment d'ordre 2.

Exemple 5.12.0.2. *1. Prenons le cas de la loi de Gauss avec (μ, σ^2) comme paramètre, dont le premier moment est μ lui-même et le moment d'ordre 2 est $E(X^2) = \mu^2 + \sigma^2$.*
On résout donc

$$\begin{cases} \mu = \bar{X} \\ \mu^2 + \sigma^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 \end{cases}$$

d'où $\hat{\mu}^M = \bar{X}$ et $\hat{\sigma^2}^M = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}^2 = S^2$.
La moyenne et la variance théoriques sont donc estimées naturellement par la moyenne et la variance empiriques.

2. Prenons le cas de la loi de Gumbel de paramètre (α, β), dont la moyenne est $\alpha + \gamma\beta$, où γ est la constante d'Euler, et la variance est $\pi^2\beta^2/6$.
On résout :

$$\begin{cases} \alpha + \gamma\beta = \bar{X} \\ (\alpha+\gamma\beta)^2 + \frac{\pi^2\beta^2}{6} = \frac{1}{n}\sum_{i=1}^{n} X_i^2 \end{cases}$$

ou, de façon équivalente,

$$\begin{cases} \alpha + \gamma\beta = \bar{X} \\ \frac{\pi^2\beta^2}{6} = S^2 \end{cases}$$

ce qui donne la solution $\hat{\beta}^M = \frac{\sqrt{6}}{\pi}S$ *et* $\hat{\alpha}^M = \bar{X} - \frac{\gamma\sqrt{6}}{\pi}S$

Définition 5.12.0.1. *Soit un n-éhantillon aléatoire* $(X_1, X_2, ..., X_n)$ *dont la loi mère appartient à une famille paramétrique de paramètre inconnu* $\theta \in \Theta$*, où* $\Theta \subseteq \mathbb{R}^k$*, et telle que pour tout* $\theta \in \Theta$ *il existe un moment* $\mu_k(\theta)$ *l'ordre k. Si, pour toute réalisation* $(x_1, x_2, ..., x_n)$ *de* $(X_1, X_2, ..., X_n)$ *le système à k équation*

$$\begin{cases} \mu_1(\theta) = m_1 \\ \mu_2(\theta) = m_1 \\ . \\ . \\ . \\ \mu_k(\theta) = m_k \end{cases}$$

(où m_r *dénote la réalisation du moment empirique d'ordre r :* $m_r = \frac{1}{n}\sum_{i=1}^{n} x_i^r$*) admet une solution unique, cette solution est appelée estimation des moments de* θ*. La fonction(de* $\mathbb{R}^n$ *dans* $\mathbb{R}^k$*) qui à toute réalisation* $(x_1, x_2, ..., x_n)$ *fait correspondre cette solution définit, en s'appliquant à* $(X_1, X_2, ..., X_n)$*, une statistique à valeurs dans* $\mathbb{R}^k$ *appelée* ***estimateur des moments*** *de* θ*.*

Chapitre 6

Estimation paramétrique par intervalle de confiance

Dans le chapitre précédent, l'objectif était de donner une valeur unique pour estimer le paramètre inconnu θ par l'estimation ponctuelle. mais elle n'apporte aucune information sur la précision des résultats, c'est-à-dire qu'elle ne tient pas compte des erreurs dues aux fluctuations d'échantillonnage.

Pour évaluer la confiance que l'on peut avoir en une estimation, il est nécessaire de lui associer un intervalle qui contient, avec une certaine probabilité, la vraie valeur du paramètre, c'est l'estimation par intervalle de confiance.

6.1 Définition et principe de construction

Définition 6.1.0.1. *Soit $X_1, X_2, , X_n$ un échantillon aléatoire issu d'une loi de densité (ou fonction de probabilité) $f(x, \theta)$ où $\theta \in \Theta$ est un paramètre inconnu de dimension 1. On appelle procédure d'intervalle de confiance de niveau β tout couple de statistiques (T_1, T_2) tel que, quel que soit $\theta \in \Theta$, on ait :*

$$\mathbb{P}(T_1 \leq \theta \leq T_2) \geq \beta$$

En pratique on choisira β assez élevé : couramment $\beta = 0,95$. Ainsi, il y a une forte probabilité pour que l'intervalle à bornes aléatoires $[T_1, T_2]$ contienne la vraie valeur de θ.

Exemple 6.1.0.1. . *Soit $X \rightsquigarrow \mathcal{N}(\mu, 1)$. On a :*

$$\frac{\overline{X} - \mu}{\frac{1}{\sqrt{n}}} \longrightarrow \mathcal{N}(0, 1)$$

d'où

$$\mathbb{P}(-1.96 < \frac{\overline{X} - \mu}{\frac{1}{\sqrt{n}}} < 1.96) = 0.95$$

$$\mathbb{P}(\frac{-1.96}{\sqrt{n}} < \overline{X} - \mu < \frac{1.96}{\sqrt{n}}) = 0,95$$

$$\mathbb{P}(\overline{X} - \frac{1.96}{\sqrt{n}} < \mu < \overline{X} + \frac{1,96}{\sqrt{n}}) = 0.95$$

et donc l'intervalle $[\overline{X} - \frac{1.96}{\sqrt{n}} < \mu < \overline{X} + \frac{1.96}{\sqrt{n}}]$ *constitue une procedure d'intervalle de confiance (IC) de niveau 0.95.*

Définition 6.1.0.2. *Dans le contexte de la définition (6.1.0.1), soit* $x_1, x_2, , x_n$ *une réalisation de* $X_1, X_2, , X_n$ *conduisant à la réalisation* (t_1, t_2) *de* (T_1, T_2). *Alors l'intervalle* $[t_1, t_2]$ *est appelé intervalle de confiance de niveau* β *pour* θ *et l'on note :*

$$IC_\beta(\theta) = [t_1, t_2].$$

L'intervalle de confiance est donc l'application numérique de la procédure suite à la réalisation de l'échantillon. Supposons que dans l'exemple précédent, avec un échantillon de taille 9, on ait observé la valeur 6 pour la moyenne de cet échantillon, alors :

$$IC_{0.95}(\mu) = \left[6 - \frac{1.96}{\sqrt{n}}, 6 + \frac{1.96}{\sqrt{n}}\right] = [5.35, 6.65]$$

.

Remarque 6.1.0.1. *S'il s'agit d'estimer une fonction* $h(\theta)$ *bijective, par exemple strictement croissante, du paramètre* θ, *il suffit de prendre l'intervalle* $[h(t_1), h(t_2)]$

6.2 Construction des IC classiques

6.2.1 Intervalle de confiance pour les paramètres d'une loi normale

Intervalle de l'espérance μ d'une loi $\mathcal{N}(\mu, \sigma^2)$

a) Cas où σ^2 est connue :

le cas où σ^2 est connu a été traité dans l'exemple 6.1.0.1 où, par commodité, on a supposé $\sigma^2 = 1$. L'IC obtenu est donc :

$$IC_{0,95}(\mu) = [\overline{x} - 1,96\frac{\sigma}{\sqrt{n}}, \overline{x} + 1,96\frac{\sigma}{\sqrt{n}}]$$

b) Cas où σ^2 est inconnue :

On a $S^{*2} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$ est un estimateur sans biais de σ^2 et

$$\frac{X_i - \overline{X}}{\sigma} \rightsquigarrow \mathcal{N}(0,1) \Rightarrow \sum_{i=1}^{n}\frac{(X_i - \overline{X})^2}{\sigma^2} \rightsquigarrow \chi^2_{n-1}$$

D'aprés la décomposition de S^2 :

$$\sum_{i=1}^{n}(X_i - \mu)^2 = \sum_{i=1}^{n}(X_i - \overline{X})^2 + n(\overline{X} - \mu)^2$$

en dévisant par σ^2 :

$$\sum_{i=1}^{n}\left(\frac{X_i - \mu}{\sigma}\right)^2 = \frac{nS^2}{\sigma^2} + \left(\frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}}\right)^2$$

On a : $\sum_{i=1}^{n}\left(\frac{X_i - \mu}{\sigma}\right)^2 \rightsquigarrow \chi^2_n$. Comme somme de n carrés de variables aléatoires centrées réduites.

$$\left(\frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}}\right)^2 \rightsquigarrow \chi^2_1$$

on en déduit que

$$\frac{nS^2}{\sigma^2} \to \chi^2_{n-1}$$

$$\frac{nS^2}{\sigma^2} = \frac{(n-1)S^{*2}}{\sigma^2} \rightsquigarrow \chi^2_{n-1}$$

Alors la variable aléatoire :

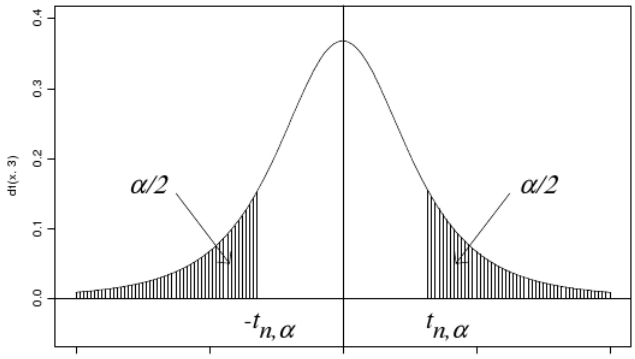
$$\frac{\frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}}}{\sqrt{\frac{(n-1)S^{*2}}{\sigma^2(n-1)}}} = \frac{\overline{X} - \mu}{S^*}\sqrt{n} \longrightarrow T_{n-1} \text{ Student}$$

L'intervalle de probabilité est :

$$\begin{aligned}
\mathbb{P}\left(-t_{\frac{\alpha}{2}} \leq \frac{\bar{X} - \mu}{S^*}\sqrt{n} \leq t_{\frac{\alpha}{2}}\right) &= 1 - \alpha \\
\mathbb{P}(|T_{n-1}| \leq t_\alpha) &= 1 - \alpha \\
\text{ou bien } \mathbb{P}(|T_{n-1}| > t_\alpha) &= \alpha \ (t_\alpha \text{ est lue de la table de Student à n-1 d.d.l})
\end{aligned}$$

D'où l'intervalle de confiance pour μ si la variance est inconnue est

$$\overline{x} - t_{\frac{\alpha}{2}}\frac{S^*}{\sqrt{n}} \leq \mu \leq \overline{x} + t_{\frac{\alpha}{2}}\frac{S^*}{\sqrt{n}}$$



Exemple 6.2.1.1. *Pour un niveau de risque $\alpha = 5\%$, une taille d'échantillon $n = 10$, une réalisation de la variance empirique corrigée $S^{*2} = 15.21$ et une réalisation de la moyenne empirique $\overline{x}_n = 5.6$, on obtient une réalisation de l'intervalle de confiance égale à :*

$$\begin{aligned} IC_{0.95}(\mu) &= \left[\bar{x} - t_{\frac{\alpha}{2}}\frac{S^*}{\sqrt{n}}, \bar{x} + t_{\frac{\alpha}{2}}\frac{S^*}{\sqrt{n}}\right] \\ &= \left[5.6 - 2.2622\frac{\sqrt{15.21}}{\sqrt{10}}, 5.6 + 2.2622\frac{\sqrt{15.21}}{\sqrt{10}}\right] \\ &= [5.3059, 5.8940] \end{aligned}$$

Intervalle de confiance pour σ^2 d'une loi $\mathcal{N}(\mu, \sigma^2)$

a) μ connu :

On a : $T = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2$ est le meilleur estimateur de σ^2.

$\frac{nT}{\sigma^2} = \sum_{i=1}^{n}\left(\frac{X_i - \mu}{\sigma}\right)^2 \rightsquigarrow \chi^2_n$ comme somme de n carrés de variable aléatoire $\mathcal{N}(0, 1)$.

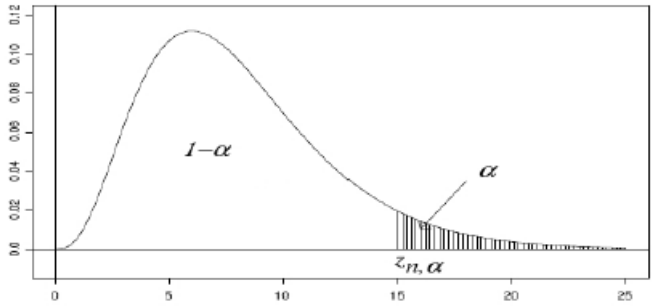
Un intervalle de probabilité pour la variable aléatoire $\chi^2(n)$ est (les bornes de l'intervalle sont lues sur la table loi du chi-deux) :

$\mathbb{P}(\chi^2_{\alpha/2}(n) \leq \chi^2(n) \leq \chi^2_{1-\alpha/2}(n)) = 1 - \alpha$.

L'intervalle de confiance pour σ^2 est :

$$\frac{nt}{\chi^2_{1-\alpha/2}} \leq \sigma^2 \leq \frac{nt}{\chi^2_{\alpha/2}}$$

La figure suivante réprésente la densité de la loi de χ^2 de paramètre n et définition du quantile $z_{n,\alpha}$ tel que $z_{n,\alpha} \approx \chi^2_{n,1-\alpha}$



Exemple 6.2.1.2. *Soit X une variable aléatoire suivant la loi normale $\mathcal{N}(40, \sigma)$. Pour estimer la variance, on prélève un échantillon de taille n=25 et on calcule la valeur de la statistique T (définie précédemment) pour cet échantillon. On obtient t=12.*
-Intervalle de confiance bilatéral, à risques symétriques, pour la variance.
-Niveau de confiance : $1 - \alpha = 0.95$.
- La statistique $\frac{nT}{\sigma^2}$ *suit une loi du chi-deux à n=25 degrés de liberté. On obtient successivement :*

$$\begin{aligned} \mathbb{P}(\chi^2_{0.025}(25) \leq \chi^2(25) \leq \chi^2_{0.975}(25)) &= \mathbb{P}(13.120 \leq \chi^2(25) \leq 40.644) \\ \mathbb{P}(13.120 \leq \frac{25 \times 12}{\sigma^2} \leq 40.644) &= \mathbb{P}(\frac{25 \times 12}{40.644} \leq \sigma^2 \leq \frac{25 \times 12}{0.025}) \\ &= \mathbb{P}(7.381 \leq \sigma^2 \leq 22.866) = 0.95. \end{aligned}$$

L'IC obtenu est donc : [7.381, 22.866].

b) μ **inconnu** :

On a

$$\frac{(n-1)S^{*2}}{\sigma^2} \longrightarrow \chi^2_{n-1}$$

d'où

$$\mathbb{P}\left(\chi^2_{\frac{\alpha}{2}}(n-1) \leq \frac{(n-1)S^{*2}}{\sigma^2} \leq \chi^2_{1-\frac{\alpha}{2}}(n-1)\right) = 1 - \alpha$$

où $\chi^2_\alpha(n-1)$ dénote le quantile d'ordre α de la loi $\chi^2(n-1)$. On peut directement isoler σ^2 pour obtenir :

$$\mathbb{P}\left(\frac{(n-1)S^{*2}}{\chi^2_{1-\frac{\alpha}{2}}} \leq \sigma^2 \leq \frac{(n-1)S^{*2}}{\chi^2_{\frac{\alpha}{2}}}\right) = 1 - \alpha$$

et

$$IC_{1-\alpha}(\sigma^2) = \left[\frac{(n-1)S^{*2}}{\chi^2_{1-\frac{\alpha}{2}}}, \frac{(n-1)S^{*2}}{\chi^2_{\frac{\alpha}{2}}}\right]$$

Exemple 6.2.1.3. *Pour un niveau de risque $\alpha = 5\%$, une taille d'échantillon $n = 16$, une réalisation de $S^{*2} = 15.246$, on obtient une réalisation de l'intervalle de confiance égale à :*

$$\begin{aligned} IC_{0.95}(\sigma^2) &= [\frac{(n-1)S^{*2}}{\chi^2_{1-\frac{\alpha}{2}}}, \frac{(n-1)S^{*2}}{\chi^2_{\frac{\alpha}{2}}}] \\ &= [\frac{15 \times 15.246}{27.488}, \frac{15 \times 15.246}{6.262}] \\ &= [8.319, 36.520] \end{aligned}$$

6.2.2 Estimation et intervalle de confiance d'une proportion

Suite au théorème central limite nous avons que la loi $\mathcal{B}(n,p)$ de $S_n = \sum_{i=1}^{n} X_i$ peut être approchée convenablement par la loi de Gauss $\mathcal{N}(np, np(1-p))$ pourvu que $np > 5$ et $n(1-p) > 5$. D'où le résultat :

$$\frac{S_n - np}{\sqrt{np(1-p)}} \rightsquigarrow \mathcal{N}(0,1)$$

qui met en évidence une v.a. de loi asymptotiquement indépendante de p et aussi

$$\mathbb{P}\left(-\mathcal{U}_{\alpha/2} < \frac{S_n - np}{\sqrt{np(1-p)}} < \mathcal{U}_{\alpha/2}\right) = 1 - \alpha$$

L'intervalle de confiance, en fonction de la fréquence relative observée de $\widehat{p} = \frac{S_n}{n}$ est

$$IC(P) = \left[\widehat{P} - \mathcal{U}_{\alpha/2}\sqrt{\frac{\widehat{P}(1-\widehat{P})}{n}}, \widehat{P} + \mathcal{U}_{\alpha/2}\sqrt{\frac{\widehat{P}(1-\widehat{P})}{n}}\right]$$

$\mathcal{U}_{\alpha/2}$ est lue sur la table $\mathcal{N}(0,1)$

Exemple 6.2.2.1. *Dans un échantillon pris au hasard de 100 automobilistes, on constate que 25 d'entre eux possèdent une voiture de cylindrée supérieure à 1 600 cc.*
Quel est l'intervalle de confiance pour la proportion d'automobilistes possédant une voiture de cylindrée supérieure à 1 600 cc (intervalle bilatéral, à risques symétriques, seuil de confiance 95 %) ?
Un estimateur non biaisé pour la proportion p est donné par la fréquence

$$\hat{p} = \frac{25}{100} = 0.25.$$

Intervalle de confiance pour p, on utilise l'approximation normale :

$$\begin{aligned} IC(P) &= \left[\hat{p} - \mathcal{U}_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + \mathcal{U}_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right] \\ &= \left[0.25 - \sqrt{\frac{0.25(1-0.25)}{100}}, 0.25 + 1.96\sqrt{\frac{0.25(1-0.25)}{100}}\right] \\ &= [0.165, 0.335] \end{aligned}$$

Chapitre 7

Les tests statistiques paramétriques

7.1 Introduction

Un test d'hypothèses est un mécanisme qui permet de trancher entre deux hypothèses complémentaires au vu des observations réalisées sur un échantillon.
Un test statistique est une mise à l'épreuve d'une hypothèse concernant une population sur la base de données fournie à partir d'un échantillon représentatif de population, qui permet de prendre la décision d'accepter ou de rejeter les hypothèses.

Dans la théorie des tests d'hypothèses et en pratique, on distingue deux catégories de tests :

1. *Tests paramétriques : tester certaine hypothèse relative à un ou plusieurs paramètres d'une variable aléatoire de densité de probabilité connue.*
2. *Tests non paramétriques : utilisés lorsqu'on a à faire à un échantillon aléatoire dont la loi est complètement inconnue.*

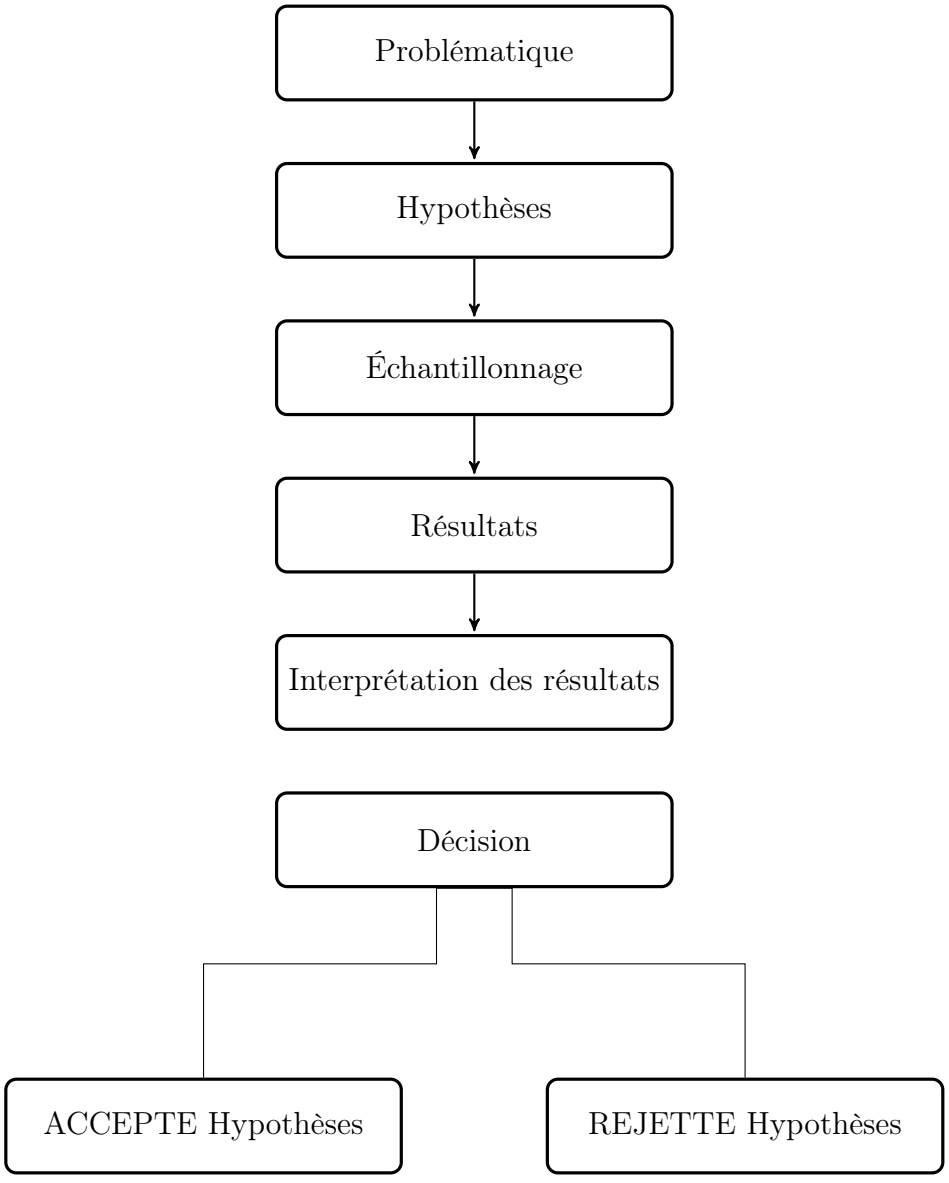


FIGURE 7.1 – La démarche scientifique des tests statistiques

7.2 Notions générales

Définition 7.2.0.1. *Un test statistique est une procédure de décision entre deux hypothèses concernant un ou plusieurs échantillons.*

Un test statistique permet de trancher entre deux hypothèses :
Hypothèse nulle notée H_0 : c'est l'hypothèse à tester.
Hypothèse alternative notée H_1 : hypothèse contraire.

Définition 7.2.0.2. *Une hypothèse statistique est une assertion (affirmation) sur la distribution d'une ou plusieurs variables ou sur les paramètres des distributions*

Il existe deux sorte d'hypothèses : hypothèse simple et composite.
Si l'hypothèse est réduite à un seul paramètre de l'espace des paramètres, on parlera d'hypothèse simple sinon on parlera d'hypothèse composite (multiple).
$H_0 : \theta = \theta_0$ hypothèse ***simple****.*
$H_1 : \theta > \theta_0, H_1 : \theta < \theta_0, H_1 : \theta \neq \theta_0$ sont ***composites (multiples)****.*

La nature de H_0 détermine la façon de formuler H_1 par conséquent la nature unilatérale ou bilatérale du test.

Définition 7.2.0.3. *Un* ***test unilatéral gauche*** *est un test de la forme $H_0 : \theta = \theta_0$ contre $H_1 : \theta < \theta_0$.* ***Un test unilatéral droit*** *est un test de la forme $H_0 : \theta = \theta_0$ contre $H_1 : \theta > \theta_0$*

Définition 7.2.0.4. *Un* ***test bilatéral*** *est un test de la forme $H_0 : \theta = \theta_0$ contre $H_1 : \theta \neq \theta_0$*

7.3 Test d'une hypothèse simple avec alternative simple

L'espace paramétrique Θ ne comprend donc que deux valeurs θ_0 et θ_1, la valeur θ_0 étant la valeur spécifiée à tester, i.e. :

$$H_0 : \theta = \theta_0 \text{ contre } H_1 : \theta = \theta_1.$$

Un test pour H_0 est une règle de décision fondée sur la valeur réalisée t d'une statistique T appelée statistique de test. Sauf exception la statistique T sera à valeurs dans $\mathbb{R}$, nous le supposerons implicitement. La règle est comme suit :

- *si $t \in A$ (une partie de $\mathbb{R}$) on accepte H_0 ,*
- *si $t \in \overline{A}$ (partie complémentaire) on rejette H_0 . La région A, qui est généralement un intervalle, sera appelée **région d'acceptation** et $\overline{A}$ **région de rejet.***

Une telle règle de décision recèle deux types d'erreur possibles du fait que la vraie valeur du paramètre est inconnue :

- *rejeter H_0 alors qu'elle est vraie (i.e. $\theta = \theta_0$) : **erreur de première espèce.***
- *accepter H_0 alors qu'elle est fausse (i.e. $\theta = \theta_1$) : **erreur de deuxième.** espèce*

Définition 7.3.0.1. *On appelle risque de **première espèce** la valeur α telle que :*

$$\mathbb{P}_{\theta_0} = (T \in \overline{A})$$

c'est à dire la probabilité de rejeter H_0 alors qu'elle est vraie, α s'appelle aussi niveau de signification.

Remarque 7.3.0.1. *Il est usuel de noter cette probabilité $\mathbb{P}(T \in \overline{A}|H_0)$ même s'il ne s'agit pas là d'une probabilité conditionnelle et, dorénavant, nous adopterons cette notation commode.*

Définition 7.3.0.2. *On appelle risque de **deuxième espèce** la valeur β telle que :*

$$\mathbb{P}_{\theta_1} = (T \in A)$$

c'est à dire la probabilité de rejeter H_1 alors qu'elle est vraie β est notée aussi $\mathbb{P}(T \in A|H_1)$.

Remarque 7.3.0.2. *Il est essentiel de garder à l'esprit que dans une procédure de test on contrôle le risque α mais pas le risque β. En d'autres termes, dans un test, on souhaite avant tout limiter à un faible niveau le risque de rejeter à tort la spécification H_0, se souciant moins d'accepter à tort, quand H_1 est vraie, cette même spécification.*

***Question :** Comment choisir la variable de décision ? Il ne suffit pas de choisir n'importe quelle statistique de loi connue sous H_0. Il est naturel de poser comme exigence que la statistique ait une plus forte propension à tomber dans la région de rejet quand H_1 est la bonne hypothèse, ce que nous transcrivons mathématiquement par la condition que la probabilité de rejeter H_0 soit plus élevée sous H_1 que sous H_0, et si possible nettement plus élevée. Toute la recherche, intuitive ou non, d'une bonne statistique de test*

repose sur ce principe que nous allons maintenant formaliser avec la notion de puissance.

Définition 7.3.0.3. *On appelle* ***puissance d'un test*** *la probabilité de rejeter H_0 alors qu'elle est effectivement fausse soit, dans les notations précédentes :*

$$\mathbb{P}(T \in \overline{A}|H_1)$$

Cette probabilité n'est rien d'autre que $1-\beta$

Ces différentes situations sont résumées dans le tableau suivant :

réalité / *décision*	*H_0 est vraie*	*H_0 est fausse*
H_0 est acceptée	*Bonne décision Niveau de confiance $1-\alpha$*	*Mauvaise décision Erreur β de seconde espèce*
H_0 est rejetée	*Mauvaise décision Erreur α de prmière espèce*	*Bonne décision Puissance de test $1-\beta$*

Définition 7.3.0.4. *On dit qu'un test est* ***sans biais*** *si sa puissance est supérieure ou égale à son risque α, soit :*

$$\mathbb{P}(T \in \overline{A}|H_1) \geq \mathbb{P}(T \in \overline{A}|H_0)$$

Remarque 7.3.0.3. *Le terme "sans biais" n'a pas de rapport direct avec la notion de biais d'un estimateur.*

Comparaison de tests :

Définition 7.3.0.5. *On dit que le test τ_1 est plus puissant que le test τ_2 au niveau α s'il est de niveau α, si τ_2 est de niveau égal (ou inférieur) à α et si la puissance de τ_1 est supérieure à celle de τ_2.*

L'objectif sera finalement de rechercher le test le plus puissant parmi tous. Dans le cas où H_0 et H_1 sont des hypothèses simples il existe un tel test, mais cela n'est pas nécessairement vrai dans le cas où l'hypothèse alternative est multiple.

Exemple 7.3.0.1. *Supposons que deux machines A et B produisent le même type de produit, mais la machine A fournit un produit plus cher de qualité supérieure. La qualité d'un produit se mesure à une entité aléatoire qui est*

de loi $\mathcal{N}(5,1)$ pour la machine A et $\mathcal{N}(4,1)$ pour la machine B. Un client achète le produit le plus cher par lots de 10 et désire développer un test pour contrôler qu'un lot donné provient bien de la machine A. Comme accuser le producteur à tort peut avoir de graves conséquences, il doit limiter le risque correspondant et tester $H_0 : \mu = 5$ contre $H_1 : \mu = 4$ à un niveau $0,05$ par exemple.
Il semble naturel d'utiliser comme statistique de test la moyenne $\overline{X}$ du lot. Sous H_0 sa loi est $\mathcal{N}(5,1/10)$ et l'on a alors l'intervalle de probabilité 0,95 :

$$[5 - 1.96/\sqrt{10}, 5 + 1.96\sqrt{10}],$$

Soit $[4.38, 5.62]$. D'où une règle de décision simple :

- *accepter H_0 si la réalisation $\bar{x}$ (moyenne du lot considéré) de $\overline{X}$ est dans $[4.38, 5.62]$,*
- *rejeter sinon.*

Il est possible de calculer la puissance de ce test puisque la loi de $\overline{X}$ est connue sous H_1 : c'est la loi $\mathcal{N}(4,1/10)$. Le risque de deuxième espèce vaut :

$$\begin{aligned}\mathbb{P}(4.38 < \overline{X} < 5.62|H_1) &= \mathbb{P}(\frac{4.38-4}{1/\sqrt{10}} < Z < \frac{5.62-4}{1/\sqrt{10}}) \text{ avec } Z \to \mathcal{N}(0,1) \\ &= \mathbb{P}(1.20 < Z < 5.12) \\ &\approx 0.115\end{aligned}$$

D'où une puissance d'environ 0.885.

7.4 Test du rapport de vraisemblance simple

Définition 7.4.0.1. *On appelle **test du rapport de vraisemblance (RV)** de l'hypothèse $H_0 : \theta = \theta_0$ contre. $H_1 : \theta = \theta_1$ au niveau α, le test défini par la région de rejet de la forme :*

$$\left\{\frac{L(x_1, x_2, ..., x_n, \theta_0)}{L(x_1, x_2, ..., x_n, \theta_1)} < k_\alpha\right\}$$

où k_α est une valeur (positive) déterminée en fonction du risque de première espèce α.

Définition 7.4.0.2. *On appelle **test du rapport de vraisemblance (RV)** de l'hypothèse $H_0 : \theta = \theta_0$ contre. $H_1 : \theta = \theta_1$ au niveau α, le test défini par la région de rejet de la forme :*

$$\left\{\frac{L(x_1, x_2, ..., x_n, \theta_1)}{L(x_1, x_2, ..., x_n, \theta_0)} > k_\alpha\right\}$$

où k_α est une valeur (positive) déterminée en fonction du risque de première espèce α.

7.4.1 La méthode de Neyman et Pearson

Soit X une variable aléatoire de densié $f(x,\theta)$ où θ est un paramètre réel inconnu. Il s'agit de tester :

$$H_0 : \theta = \theta_0 \text{ contre } H_1 : \theta = \theta_1$$

Supposons α connu et soit $\overline{A}$ une région critique de $\mathbb{R}^n$ telle que :

$$\int_{\overline{A}} L(x_1,...,x_n,\theta_0)d(x_1,...,x_n) = \alpha = \mathbb{P}(\overline{A}/H_0)$$

Il s'agit de maximiser : $1-\beta = \int_{\overline{A}} L(x_1,...,x_n,\theta_1)d(x_1,...,x_n) = \mathbb{P}(\overline{A}/H_1)$, on peut écrire

$$1-\beta = \int_{\overline{A}} \frac{L(x_1,...,x_n,\theta_1)}{L(x_1,...,x_n,\theta_0)} L(x_1,...,x_n,\theta_0)d(x_1,...,x_n)$$

Théorème 7.4.1.1. *(Neyman-Pearson)*
La région critique optimale est définie par l'ensemble des points de $\mathbb{R}^n$ tels que :

$$\frac{L(x_1,x_2,....,x_n,\theta_1)}{L(x_1,x_2,....,x_n,\theta_0)} > k_\alpha$$

Démonstration 7.4.1.1. *1. S'il existe une constante k_α, telle que l'ensemble $\overline{A}$ des points de $\mathbb{R}^n$ où :*

$$\frac{L(x_1,x_2,....,x_n,\theta_1)}{L(x_1,x_2,....,x_n,\theta_0)} > k_\alpha$$

Soit de probabilité α sous $H_0 : \mathbb{P}(\overline{A}|H_0) = \alpha$, alors cette région critique réalise le maximum de $1-\beta$.
En effet
Soit $\overline{A}^*$ une autre région de $\mathbb{R}^n$ $|\mathbb{P}(\overline{A}^*|H_0) = \alpha$. $\overline{A}^*$ diffère alors de $\overline{A}$ par des points où :

$$\frac{L(x_1,x_2,....,x_n,\theta_1)}{L(x_1,x_2,....,x_n,\theta_0)} \leq k_\alpha.$$

$$\int_{\overline{A}} \frac{L(x_1,x_2,....,x_n,\theta_1)}{L(x_1,x_2,....,x_n,\theta_0)} L(x_1,x_2,...,\theta_0)d(x_1,...,x_n)$$

diffère de

$$\int_{\overline{A}^*} \frac{L(x_1,x_2,....,x_n,\theta_1)}{L(x_1,x_2,....,x_n,\theta_0)} L(x_1,x_2,...,x_n,\theta_0)d(x_1,...,x_n)$$

pour les parties non communes à $\overline{A}$ et $\overline{A}^*$

$$\mathbb{P}(\overline{A}|H_0) = \mathbb{P}(\overline{A}^*|H_0) = \alpha$$

$$\mathbb{P}(\overline{A} - \overline{A}^*|H_0) = \mathbb{P}(\overline{A}^* - \overline{A}|H_0)$$

On pose $\underline{x} = x_1, x_2, ..., x_n$

$$\int_{\overline{A}-\overline{A}^*} \frac{L(\underline{x},\theta_1)}{L(\underline{x},\theta_0)} L(\underline{x},\theta_0)d\underline{x} > \int_{\overline{A}^*-\overline{A}} \frac{L(\underline{x},\theta_1)}{L(\underline{x},\theta_0)} L(\underline{x},\theta_0)d\underline{x}$$

En effet :

$$\int_{\overline{A}-\overline{A}^*} \frac{L(\underline{x},\theta_1)}{L(\underline{x},\theta_0)} L(\underline{x},\theta_0)d\underline{x} = \frac{L(\grave{\zeta},\theta_1)}{L(\grave{\zeta},\theta_0)} \mathbb{P}(\overline{A} - \overline{A}^*|H_0) \text{ avec } \grave{\zeta} \in \overline{A} - \overline{A}^*$$

$$\int_{\overline{A}^*-\overline{A}} \frac{L(\underline{x},\theta_1)}{L(\underline{x},\theta_0)} L(\underline{x},\theta_0)d\underline{x} = \frac{L(\zeta,\theta_1)}{L(\zeta,\theta_0)} \mathbb{P}(\overline{A}^* - \overline{A}|H_0) \text{ avec } \zeta \in \overline{A}^* - \overline{A}$$

Alors

$$\frac{L(\grave{\zeta},\theta_1)}{L(\grave{\zeta},\theta_0)} \leq k_\alpha \leq \frac{L(\zeta,\theta_1)}{L(\zeta,\theta_0)}$$

D'où 1) est démontré

Proposition 7.1. *Le test du RV est sans biais.*

Démonstration 7.4.1.2. *Puisque* $L(\underline{x},\theta_1) > k_\alpha L(\underline{x},\theta_0)$. *D'où*

$$\int_{\overline{A}} L(\underline{x},\theta_1)d\underline{x} > k_\alpha \int_{\overline{A}} L(\underline{x},\theta_0)d\underline{x}$$

Si $k_\alpha > 1$, *la proposition est triviale.*
Si $k_\alpha < 1$ *: ceci revient à montrer que* $\beta < 1 - \alpha$.
$\beta = \mathbb{P}(A|H_1)$ *et* $1 - \alpha = \mathbb{P}(A|H_0)$.

$$A = \{\underline{x} = (x_1, .., x_n) \in \mathbb{R}^n : \frac{L(\underline{x},\theta_1)}{L(\underline{x},\theta_0)} \leq k_\alpha\}$$

Donc

$$\int_A L(\underline{x},\theta_1)d\underline{x} < k_\alpha \int_A L(\underline{x},\theta_0)d\underline{x}$$

D'où

$$\int_A L(\underline{x},\theta_1)d\underline{x} < \int_A L(\underline{x},\theta_0)d\underline{x}$$

i.e : $\beta < 1 - \alpha$, *d'où le résultat.*
Conclusion *: Les test de RV (Le test N-P) sont sans biais.*

Exemple 7.4.1.1. *Soit* $(X_1, X_2, ..., X_n)$ *un n-échantillon issu de* $X \to \mathcal{N}(\mu, 1)$.

1. *Tester au niveau* $\alpha = 5\%$, $H_0 : \mu = \mu_0$ *contre* $H_1 : \mu = \mu_1, \mu_0, \mu_1$ *données* $\mu_1 > \mu_0$.
2. *Calculer la puissance du test.*

3. *A.N* $\mu_0 = 0, \mu_1 = 2, \bar{x} = 0.6, n = 16.$
4. *Que se passera-t-il si* $H_1 : \mu_1 = 0.5$ *?*

$$\overline{A} = \{(x_1, x_2, ..., x_n) \in \mathbb{R}^n, \frac{L((x_1, x_2, ..., x_n, \mu_1)}{L((x_1, x_2, ..., x_n, \mu_0} > k_\alpha\}$$

$$\begin{aligned}
\frac{L((x_1, x_2, ..., x_n, \mu_1)}{L((x_1, x_2, ..., x_n, \mu_0} &= \frac{\frac{1}{(\sqrt{2\pi})^n} \exp^{\frac{-1}{2}\sum_{i=1}^{n}(x_i-\mu_1)^2}}{\frac{1}{(\sqrt{2\pi})^n} \exp^{\frac{-1}{2}\sum_{i=1}^{n}(x_i-\mu_0)^2}} \\
&= \exp^{\frac{-1}{2}\sum_{i=1}^{n}(x_i-\mu_1)^2} . \exp^{\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu_0)^2} \\
\ln \frac{L((x_1, x_2, ..., x_n, \mu_1)}{L((x_1, x_2, ..., x_n, \mu_0)} > \ln k_\alpha &\Rightarrow -\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu_1)^2 + \frac{1}{2}\sum_{i=1}^{n}(x_i-\mu_0)^2 > \ln k_\alpha \\
&\Rightarrow -\frac{1}{2}\sum_{i=1}^{n}x_i^2 + \mu_1\sum_{i=1}^{n}x_i - \frac{n}{2}\mu_1^2 + \frac{1}{2}\sum_{i=1}^{n}x_i^2 \\
&+ \mu_0\sum_{i=1}^{n}x_i + \frac{n}{2}\mu_0^2 > \ln k_\alpha \\
&\Rightarrow (\mu_1 - \mu_0)\sum_{i=1}^{n}x_i - \frac{n}{2}(\mu_1^2 - \mu_0^2) > \ln k_\alpha \\
&\Rightarrow (\mu_1 - \mu_0)\sum_{i=1}^{n}x_i > \frac{n}{2}(\mu_1^2 - \mu_0^2) + \ln k_\alpha \\
&\Rightarrow \sum_{i=1}^{n}x_i > K_\alpha.
\end{aligned}$$

D'où

$$\overline{A} = \{(x_1, x_2, ..., x_n) \in \mathbb{R}^n, \sum_{i=1}^{n}x_i > k_\alpha\}$$

$\mathbb{P}(\overline{A}|H_0) = \alpha$ *sous* $H_0 : \sum_{i=1}^{n}x_i \to \mathcal{N}(n\mu_0, n)$

$$\mathbb{P}(\sum_{i=1}^{n}X_i > k_\alpha) = \alpha \Rightarrow \mathbb{P}(\frac{\sum_{i=1}^{n}X_i - n\mu_0}{\sqrt{n}} > \frac{k_\alpha - n\mu_0}{\sqrt{n}}) = 1 - \alpha$$

$$\frac{k_\alpha - n\mu_0}{\sqrt{n}} = \phi^{-1}_{\mathcal{N}(0,1)}(1-\alpha)$$

(ϕ fonction de répartition de $\mathcal{N}(0,1)$). Alors :

$$k_\alpha = n\mu_0 + \sqrt{n}\phi^{-1}_{\mathcal{N}(0,1)}(1-\alpha)$$

$A.N : \mu_0 = 0, \mu_2 = 2, \sum_{i=1}^{n} x_i = n\overline{x} = 16.0.6 = 9.6, n = 16$
$\phi^{-1}_{\mathcal{N}(0,1)}(1-\alpha) = \phi^{-1}_{\mathcal{N}(0,1)}(0.95) = 1,64.$
$K_\alpha = 16 \times 0 + \sqrt{16} \times 1.64 = 4 \times 1.64 = 6.56.$
On a $\sum_{i=1}^{n} x_i = 9.6 > K_\alpha = 6.56 \Rightarrow$ *on rejette* H_0.
Puissance du test

$$\begin{aligned}
1-\beta = \mathbb{P}(\overline{A}|H_1) &= \mathbb{P}_{H_1}(\sum_{i=1}^{n} X_i > k_\alpha) \\
&= \mathbb{P}(\frac{\sum_{i=1}^{n} X_i - n\mu_1}{\sqrt{n}} > \frac{k_\alpha - n\mu_1}{\sqrt{n}}) \\
&= 1 - \mathbb{P}(\frac{\sum_{i=1}^{n} X_i - n\mu_1}{\sqrt{n}} \leq \frac{k_\alpha - n\mu_1}{\sqrt{n}}) \\
&= 1 - \phi_{\mathcal{N}(0,1)}(\frac{k_\alpha - n\mu_1}{\sqrt{n}}) \\
&= 1 - \phi_{\mathcal{N}(0,1)}(\frac{n\mu_0 + \sqrt{n}\phi^{-1}(1-\alpha) - n\mu_1}{\sqrt{n}}) \\
&= 1 - \phi_{\mathcal{N}(0,1)}(\frac{n(\mu_0 - \mu_1) + \sqrt{n}\phi^{-1}(1-\alpha)}{\sqrt{n}}) \\
&= 1 - \phi_{\mathcal{N}(0,1)}(\sqrt{n}(\mu_0 - \mu_1) + \phi^{-1}(1-\alpha)) \\
&= 1 - \phi[4(-2) + 1.64] \\
&= 1 - \phi(-6.36) \\
&= 1 - [1 - \phi(6.36)] = \phi(6.36) \\
&= \phi(6.36) = 1 \\
&\Rightarrow \beta = 0
\end{aligned}$$

Si $\mu_1 = 0.5$, *la région critique ne change pas la puissance du test.*
$1 - \beta = \phi(0.36) = 0.64 \Rightarrow \beta = 0.36$ *(grand).*

7.5 Tests d'hypothèses multiples

Définition 7.5.0.1. *Soit* $H_0 : \theta \in \Theta_0$ *une hypothèse nulle multiple et* $\alpha(\theta)$ *le risque de première espèce pour la valeur* $\theta \in \Theta_0$. *On appelle niveau du test*

(ou seuil du test) la valeur α telle que :

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

De même si l'hypothèse alternative $H_1 : \theta \in \Theta_1$ est multiple le risque de deuxième espèce est une fonction $\beta(\theta)$ ainsi que la puissance. On définit alors la fonction puissance du test :

$$h(\theta) = 1 - \beta(\theta) = \mathbb{P}_\theta(T \in \overline{A}) \text{ définie pour tout } \theta \in \Theta_1$$

Définition 7.5.0.2. *On dit qu'un test est **sans biais** si sa fonction puissance reste supérieure ou égale à son niveau α, soit :*

$$\mathbb{P}_\theta(T \in \overline{A}) \geq \alpha \text{ pour tout } \theta \in \Theta_1$$

Définition 7.5.0.3. *On dit que le test τ_1 de niveau α est uniformément plus puissant que le test τ_2 au niveau α s'il est de niveau α, si τ_2 est de niveau égal (ou inférieur) à α et si la fonction puissance de τ_1 reste toujours supérieure ou égale à celle de τ_2, mais strictement supérieure pour au moins une valeur de $\theta \in \Theta_1$ ie $\forall \theta \in \Theta_1$ $h_1(\theta) \geq h_2(\theta)$ et $\exists\, \theta^* \in \Theta_1$ tel que $h_1(\theta^*) > h_2(\theta^*)$, où $h_1(\theta)$ et $h_2(\theta)$ sont les fonctions puissance respectives des tests τ_1 et τ_2.*

7.5.1 Tests d'hypothèses multiples unilatérales

Nous considérons des situations de test du type :

$$H_0 : \theta \leq \theta_0 \text{ contre } H_1 : \theta > \theta_0$$

$$H_0 : \theta \geq \theta_0 \text{ contre } H_1 : \theta < \theta_0$$

Proposition 7.2. [10] *S'il existe une statistique $T = t(X_1, X_2, , X_n)$ exhaustive minimale à valeurs dans $\mathbb{R}$ et si, pour tout couple $(\theta, \grave{\theta})$ tel que $\theta < \grave{\theta}$. le rapport de vraisemblance $L(x_1, x_2, ..., x_n, \theta)/L(x_1, x_2, ..., x_n, \grave{\theta})$ est une fonction monotone de $t(x_1, x_2, , x_n)$, alors il existe un test uniformément le plus puissant pour les situations d'hypothèses unilatérales et la région de rejet est soit de la forme $t(x_1, x_2, , x_n) < k$, soit de la forme $t(x_1, x_2, , x_n) > k$.*

Cas particulier : *Si la loi mère est dans une famille appartenant à la classe exponentielle,i.e $f(x, \theta) = a(\theta)b(x)\exp\{c(\theta)d(x)\}$ et si la fonction $c(\theta)$ est monotone, alors il existe un test uniformément le plus puissant pour les situations d'hypothèses unilatérales et la région de rejet est soit de la forme $\sum_{i=1}^{n} d(x_i) < k_\alpha$, soit de la forme $\sum_{i=1}^{n} d(x_i) > k_\alpha$.*

Exemple 7.5.1.1. *Soit la loi mère $\mathcal{N}(0,1)$ où μ inconnu. Testons $H_0 : \mu \leq \mu_0$ contre $H_1 : \mu > \mu_0$. La densité de la loi mère est :*

$$f(x,\mu) = \frac{1}{\sqrt{2\pi}} \exp\{-\frac{1}{2}(x-\mu)^2\} = \frac{1}{\sqrt{2\pi}} \exp\{-\frac{1}{2}(x^2)\} \exp\{-\frac{1}{2}\mu^2\} \exp\{\mu x\}$$

Ici $c(\mu) = \mu$ et $d(x) = x$. Le test UPP a donc une région de rejet de la forme $\sum_{i=1}^{n} x_i > k$ ou $\sum_{i=1}^{n} x_i < k_\alpha$.

Pour trouver le sens correct de l'inégalité il faut repartir du rapport de vraisemblance

$$\frac{L(x_1, x_2, ..., x_n, \mu)}{L(x_1, x_2, ..., x_n, \grave{\mu})}$$

qui varie comme $\exp\{(\mu - \grave{\mu}) \sum_{i=1}^{n} x_i\}$ et pour $\mu < \grave{\mu}$,

est donc une fonction décroissante de $\sum_{i=1}^{n} x_i$ Ainsi un RV inférieur à k_α est équivalent à $\sum_{i=1}^{n} x_i > \grave{k} \Rightarrow \overline{x} > \grave{\grave{k}}$.

$\bar{X}$ est de loi $\mathcal{N}(\mu_0, \frac{1}{\sqrt{n}})$ pour $\mu = \mu_0$ Pour un niveau 0,05 la constante $\grave{\grave{k}}$ est définie par $\mathbb{P}_{\mu_0}(\overline{X} > \grave{\grave{k}}) = 0,05$ et donc

$$\mathbb{P}(\frac{\overline{X} - \mu_0}{\frac{1}{\sqrt{n}}} > \frac{\grave{\grave{k}} - \mu_0}{\frac{1}{\sqrt{n}}}) = 0.05$$

soit $\grave{\grave{k}} = \mu_0 + 1.645(\frac{1}{\sqrt{n}})$.
La fonction puissance est définie par $h(\mu) = \mathbb{P}_\mu(\overline{X} > \mu_0 + 1.645(\frac{1}{\sqrt{n}})$ pour $\overline{X} \to \mathcal{N}(\mu, \frac{1}{\sqrt{n}})$ et $\mu \in]\mu_0, +\infty[$. En posant $Z = \sqrt{n}(\overline{X} - \mu)$ on obtient :

$$h(\mu) = \mathbb{P}(Z > 1.645 - \sqrt{n}(\mu - \mu_0)) \text{ où } Z \longrightarrow \mathcal{N}(0,1)$$

Exemple 7.5.1.2. *Soit $X \to \mathcal{B}(p)$, $X_1, X_2, ..., X_n$ un n-échantillon issu de X. Tester : $H_0 : p \leq p_0$ contre $H_1 : p > p_0$, on a :*

$$\begin{aligned} f(x,p) &= p^x(1-p)^{1-x}, x \in \{0,1\} \\ \ln f(x,p) &= x \ln p + (1-x)\ln(1-p) \\ &= x \ln \frac{p}{1-p} + \ln(1-p) \\ f(x,p) &= (1-p) \exp x \ln \frac{p}{1-p} \end{aligned}$$

On a

$$\begin{aligned} c(p) &= \ln \frac{p}{1-p} = \ln p - \ln 1 - p \\ \grave{c}(p) &= \frac{1}{p} + \frac{1}{1-p} > 0 \Rightarrow \text{ est croissante en } p \end{aligned}$$

D'où la loi de $\mathcal{B}(p)$ *est à rapport de vraisemblance monotone en*

$$T = \sum_{i=1}^{n} d(x_i) = \sum_{i=1}^{n} x_i$$

Le test UPP a donc une région de rejet de la forme

$$\overline{A} = \{(x_1, x_2, ..., x_n) \in \mathbb{R}^n | \sum_{i=1}^{n} x_i > k_\alpha\}$$

avec sous $H_0 : \sum_{i=1}^{n} x_i \rightsquigarrow \mathcal{B}(n, p_0)$ *tabulée, on trouve* k_α *par :* $\mathbb{P}_{H_0}(\overline{A}) = \alpha$.

7.5.2 Tests d'hypothèses bilatérales

considérons deux situations du type bilatéral :

$$H_0 : \theta = \theta_0 \text{ contre } H_1 : \theta \neq \theta_1$$

ou

$$H_0 : \theta_1 \leq \theta \leq \theta_2 \text{ contre } H_1 : \theta < \theta_1 \text{ et } \theta > \theta_2$$

La première situation est fréquente lorsque θ *représente en fait un écart entre paramètres de deux populations, par exemple entre leurs moyennes. La seconde teste si le paramètre est situé dans un intervalle de tolérance acceptable.*
On ne peut s'attendre dans ces situations à obtenir un test UPP du fait qu'il faut faire face à des alternatives à la fois du type $\theta < \theta_0$ *et du type* $\theta > \theta_0$, *par exemple pour le premier cas.*

Toutefois il pourra y avoir un test uniformément plus puissant dans la classe restreinte des tests sans biais, en bref UPP-sans biais. Ceci est notamment vrai pour les familles de lois de laclasse exponentielle. D'une façon générale la région d'acceptation aura la forme $c_1 < t < c_2$ *ou* $c_1 < c_2$, *pour la réalisation d'une statistique de test* T *appropriée (soit* $\sum_{i=1}^{n} d(X_i)$ *pour la classe exponentielle).*

Remarque 7.5.2.1. *L'usage veut que l'on détermine les valeurs critiques c_1 et c_2 en répartissant $\alpha/2$ sur chaque extrémité. Ainsi, pour le cas $H_0 : \theta = \theta_0$, ces valeurs seront telles que $\mathbb{P}_{\theta_0}(T < c_1) = \mathbb{P}_{\theta_0}(T > c_2) = \alpha/2$. Mais cette règle ne conduit pas au test UPP-sans biais si la loi de T n'est pas symétrique. Dans la classe exponentielle la répartition doit être telle que la dérivée par rapport à θ de $\mathbb{P}_{\theta_0}(T < c_1) + \mathbb{P}_{\theta_0}(T > c_2)$ s'annule en θ_0.*
Pour le cas $H_0 : \theta_1 \leq \theta \leq \theta_2$ la condition est que le seuil α soit atteint à la fois en θ_1 et en θ_2.

7.6 Test du rapport de vraisemblance généralisé

Considérons maintenant les hypothèses paramétriques les plus générales.

1. *Soit la famille paramétrique $\{f(x,\theta), \theta \in \Theta\}$, où $\Theta \subseteq \mathbb{R}^k$, et les hypothèses $H_0 : \theta \in \Theta_0$ contre $H_1 : \theta \in \Theta_1$ où $\Theta_1 = \Theta - \Theta_0$ est le complémentaire de Θ_0 par rapport à Θ. On appelle rapport de vraisemblance généralisé (RVG), la fonction $\lambda(x_1, x_2, ..., x_n)$ telle que :*

$$\lambda(x_1, x_2, ..., x_n) = \frac{\sup\limits_{\theta \in \Theta_0} L(x_1, x_2, ..., x_n, \theta)}{\sup\limits_{\theta \in \Theta} L(x_1, x_2, ..., x_n, \theta)}$$

*et **test du RVG**, le test défini par une région de rejet de la forme :*

$$\lambda(x_1, x_2, ..., x_n) < k \leq 1.$$

Il existe une estimation du maximum de vraisemblance $\widehat{\theta}^{MV}$ alors le dénominateur est la valeur de la fonction de vraisemblance en $\widehat{\theta}^{MV}$, soit : $L(x_1, x_2, ..., x_n, \widehat{\theta}^{MV})$.
Le problème est toutefois de connaître la loi de la statistique du RVG $\lambda(X_1, X_2, , X_n)$ pour toute valeur de θ dans Θ_0 afin de définir la valeur de k_α permettant de garantir le niveau α choisi. En effet cette valeur doit être telle que :

$$\sup_{\theta \in \Theta_0} \mathbb{P}_\theta(\lambda(X_1, X_2, , X_n) < k_\alpha) = \alpha.$$

2. *Test $H_0 : \theta = \theta_0$ contre $H_1 : \theta \neq \theta_0$ où θ peut être un vecteur de dimension r. On pose :*

$$\lambda = \frac{L(x_1, ..., x_n, \theta_0)}{\sup\limits_{\theta \in \Theta} L(x_1, .., x_n, \theta)}$$

Remarque 7.6.0.1. *$0 \leq \lambda \leq 1$, plus λ est grand, plus H_0 est vraisemblable, cela revient à remplacer sous H_1, θ par son estimation du maximum de vraisemblance. La région critique du test sera :*

$$\overline{A} = \{(x_1, ...x_n) \in \mathbb{R}^n / \lambda < k\}$$

Théorème 7.6.0.1. *La distribution de $-2\ln\lambda$ est asymptotiquement celle de χ^2_r sous H_0*

Démonstration 7.6.0.1. *Pour $r = 1$, $-2\ln\lambda \underset{n\to+\infty}{\longrightarrow} \chi^2_1$?*

$$\ln\lambda = \ln L(x_1, ...x_n, \theta_0) - \ln L(x_1, ..., x_n, \hat{\theta})$$

Le développement en série de Taylor la log-vraisemblance de $L(x_n, \theta_0)$ en θ_0 autour de l'estimation du maximum de vraisemblance $\hat{\theta}$:

$$\ln L(x_n, \theta_0) = \ln L(x_n, \hat{\theta}) + (\hat{\theta} - \theta_0)\frac{\partial L(x_n, \hat{\theta})}{\partial\theta} + \frac{1}{2}(\hat{\theta} - \theta_0)^2 \frac{\partial^2}{\partial\theta^2}\ln L(x_n, \tilde{\theta})$$

où $\tilde{\theta} \in [\theta_0, \hat{\theta}]$.
Comme, par définition de l'estimation du MV, $\frac{\partial L(x_n, \hat{\theta})}{\partial\theta} = 0\theta$, on a, pour le RVG :

$$-2\ln\lambda = -2[\ln L(x_n, \theta_0) - \ln L(x_n, \hat{\theta})]$$

d'où

$$-2\ln\lambda = -(\hat{\theta} - \theta_0)^2 \frac{\partial^2}{\partial\theta^2}\ln L(x_n, \tilde{\theta})$$

$$\ln\lambda = \frac{1}{2}(\hat{\theta} - \theta_0)^2 \frac{\partial^2}{\partial\theta^2}\ln L(x_n, \tilde{\theta})$$

Sous $H_0 : \theta = \theta_0$, l'EMV $\hat{\theta}$ converge en probabilité vers θ_0 et il en va donc de même pour $\tilde{\theta}$.

$$-\frac{1}{n}\frac{\partial^2 \ln L(x_n, \tilde{\theta})}{\partial\theta^2} = -\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2}{\partial\theta^2}\ln f(x_i, \tilde{\theta})$$

converge en probabilité et donc en loi vers

$$E\left[-\frac{\partial^2}{\partial\theta^2}\ln f(X, \theta_0)\right] = I_1(\theta_0)$$

d'où

$$-2\ln\lambda \xrightarrow{\mathcal{L}} (\hat{\theta} - \theta_0)^2 n.I_1(\theta_0)$$

mais on a

$$\frac{\hat{\theta} - \theta_0}{\sqrt{\frac{1}{I_n(\theta_0)}}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1) \Rightarrow (\hat{\theta} - \theta_0)^2 n.I_1(\theta_0) \to \chi^2_1$$

Exemple 7.6.0.1. *Soit* $(X_1, X_2, ..., X_n)$ *un n-échantillon issu de* $X \to \mathcal{N}(\mu, 1)$, *tester au niveau* α :

$$\begin{cases} H_0 : & \mu = \mu_0 = 5 \\ H_1 : & \mu \neq \mu_0 \end{cases}$$

$$\begin{aligned}
\lambda &= \frac{L(x_1, .., x_n, \mu_0)}{L(x_1, .., x_n, \hat{\mu})} \\
&= \frac{(\frac{1}{\sqrt{2\pi}})^n e^{-\frac{1}{2}\sum_{i=1}^{n}(x_i - \mu_0)^2}}{(\frac{1}{\sqrt{2\pi}})^n e^{-\frac{1}{2}\sum_{i=1}^{n}(x_i - \bar{x})^2}} \\
&= e^{-\frac{1}{2}\left[\sum_{i=1}^{n}(x_i - \mu_0)^2 - \sum_{i=1}^{n}(x_i - \bar{x})^2\right]} \\
-2\ln\lambda &= \sum_{i=1}^{n}(x_i - \mu_0)^2 - \sum_{i=1}^{n}(x_i - \bar{x})^2 \\
&= n(\bar{x}^2 - 2\mu_0\bar{x} + \mu_0^2) \\
&= n(\bar{x} - \mu_0)^2 \\
\lambda < k &\Rightarrow -2\ln\lambda > -2\ln k \\
&\Rightarrow n(\bar{x} - 5)^2 > -2\ln k \\
\mathbb{P}(n(\bar{x} - 5)^2 > -2\ln k) = \alpha &\Rightarrow \mathbb{P}(\chi_1^2 \leq -2\ln k) = 1 - \alpha \\
-2\ln k = \phi_{\chi_1^2}^{-1}(1 - \alpha) &\Rightarrow k = e^{\frac{-1}{2}\phi_{\chi_1^2}^{-1}(1-\alpha)}
\end{aligned}$$

d'où

$$\bar{A} = \{(x_1, ..., x_n) \in \mathbb{R}^n / \lambda < e^{\frac{-1}{2}\phi_{\chi_1^2}^{-1}(1-\alpha)}\}$$

7.7 Les tests paramétriques usuels

7.7.1 Tests sur la moyenne d'une loi $\mathcal{N}(\mu, \sigma^2)$

Cas où σ^2 est connu

$$\begin{cases} H_0 : & \mu = \mu_0 \\ H_1 : & \mu \neq \mu_1 \end{cases}$$

On sait que la moyenne empirique $\overline{X}$ *est un estimateur sans biais de* μ, *tel que :* $\overline{X} \to \mathcal{N}(\mu, \frac{\sigma^2}{n})$ *et la statistique* $\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}$ *de loi connue :* $\mathcal{N}(0, 1)$ *:*
On utilise la variable de décision $\overline{X}$.

Comme

$$\begin{aligned} 1-\alpha &= \mathbb{P}_{\mu_0}(-z_{1-\alpha/2} < \frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}} < z_{1-\alpha/2}) \\ &= \mathbb{P}_{\mu_0}(\mu_0 - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} < \overline{X} < \mu_0 + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}) \end{aligned}$$

donc, une région de non-rejet peut être définie, pour un test de risque α, par :

$$A = \{\mu_0 - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} < \overline{x} < \mu_0 + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\}$$

Exemple 7.7.1.1. *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 : $H_0 : \mu = \mu_0 = 45$ contre $H_1 : \mu = \mu_1 \neq 45$?
Soient k_1 et k_2 les seuils critiques. La règle de décision est : on accepte l'hypothèse H_0 si $k_1 < \overline{x} < k_2$

$$\begin{aligned} \mathbb{P}_{\mu_0}(k_1 \leq \overline{x} \leq k_2) &= 0.95 \\ \mathbb{P}_{\mu_0}(\frac{k_1-45}{5/4} \leq \frac{\overline{X}-45}{5/4} \leq \frac{k_2-45}{5/4}) &= 0.95 \end{aligned}$$

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 H_0, on refuse donc l'hypothèse H_0 et on accepte l'hypothèse H_1
On peut calculer le risque de deuxième espèce associé à cette valeur $\mu_1 = 49$.

$$\begin{aligned} \beta &= \mathbb{P}(\frac{42.55-49}{5/4} \leq \frac{\overline{x}-49}{5/4} \leq \frac{47.45-49}{5/4}|H_1) \\ &= \mathbb{P}(-5.16 \leq \frac{\overline{x}-49}{5/4} \leq -1.24|H_1) \\ &= 0.1075 \end{aligned}$$

La probabilité de refuser l'hypothèse $H_1 : \mu = \mu_1 = 49$, alors qu'elle est vraie, est égale à 0.1075, la puissance du test est égale à 0.8925.

Remarque 7.7.1.1. *Les tests unilatéraux correspondants sont de rejeter H_0 si*

1. $\bar{A} = \{\bar{x} | \bar{x} < \mu_0 - z_{1-\alpha}\frac{\sigma}{\sqrt{n}}\}$ *pour* $\begin{cases} H_0 & \mu \geq \mu_0 \\ H_1 & \mu < \mu_0 \end{cases}$

2. $\bar{A} = \{\bar{x} | \bar{x} > \mu_0 + z_{1-\alpha}\frac{\sigma}{\sqrt{n}}\}$ *pour* $\begin{cases} H_0 & \mu \leq \mu_0 \\ H_1 & \mu > \mu_0 \end{cases}$

Cas où σ^2 est inconnu

σ^2 est estimée par $S^{*2} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$ *On a*

$$\frac{(n-1)S^{*2}}{\sigma^2} \to \chi^2_{n-1}$$

Alors la variable aléatoire :

$$\frac{\frac{\bar{X}-\mu_0}{\frac{\sigma}{\sqrt{n}}}}{\sqrt{\frac{(n-1)S^{*2}}{\sigma^2(n-1)}}} = \frac{\bar{X} - \mu_0}{S^*}\sqrt{n} \longrightarrow T_{n-1} \text{ Student}$$

Comme

$$\begin{aligned} 1-\alpha &= \mathbb{P}_{\mu_0}(-t^{(n-1)}_{1-\alpha/2} \leq \frac{\overline{X} - \mu_0}{S^*/\sqrt{n}} \leq t^{(n-1)}_{1-\alpha/2}) \\ &= \mathbb{P}_{\mu_0}(\mu_0 - t^{(n-1)}_{1-\alpha/2}\frac{S^*}{\sqrt{n}} \leq \overline{X} \leq \mu_0 + t^{(n-1)}_{1-\alpha/2}\frac{S^*}{\sqrt{n}}) \end{aligned}$$

Où $t^{(n-1)}_{1-\frac{\alpha}{2}}$ est lue sur la table de Student à (n-1) d.d.l.
Donc, une région de non-rejet peut être définie, pour un test de risque α, par :

$$A = \{(x_1, x_2, ..., x_n) \in \mathbb{R}^n | \mu_0 - t^{(n-1)}_{1-\alpha/2}\frac{S^*}{\sqrt{n}} \leq \overline{x} \leq \mu_0 + t^{(n-1)}_{1-\alpha/2}\frac{S^*}{\sqrt{n}}\}$$

Remarque 7.7.1.2. *Les tests unilatéraux correspondants sont de rejeter H_0 si*

1. $\bar{A} = \{(x_1, ..., x_n) \in \mathbb{R}^n | \overline{x} < \mu_0 - t^{(n-1)}_{1-\alpha/2}\frac{S^*}{\sqrt{n}}\}$ *pour* $\begin{cases} H_0 & \mu \geq \mu_0 \\ H_1 & \mu < \mu_0 \end{cases}$

2. $\bar{A} = \{(x_1, ..., x_n) \in \mathbb{R}^n | \overline{x} > \mu_0 + t^{(n-1)}_{1-\alpha/2}\frac{S^*}{\sqrt{n}}\}$ *pour* $\begin{cases} H_0 & \mu \leq \mu_0 \\ H_1 & \mu > \mu_0 \end{cases}$

Exemple 7.7.1.1. *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 :

620 570 565 590 530 625 610 595

540 580 605 575 550 560 575 615

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 : $\bar{x} = 581.60$ *et* $S^2 = 778.41$ *et* $S^{*2} = 830.304$
Les hypothèses à tester sont :

$$H_0 : \mu = 600 \text{ heures et } H_1 : \mu < 600 \text{ heures}$$

La variable $\frac{\bar{X}-\mu}{S^*/\sqrt{n}}$ *suit une loi de Student à* $(n-1) = 15$ *degrés de liberté.*
Calcul du seuil critique k*. La règle de décision est la suivante :*

$$\bar{x} < k \text{ Refus de } H_0$$

$$\bar{x} > k \text{ Acceptation de } H_0$$

$$\mathbb{P}_{\mu_0}(\bar{X} < k | H_0) = 0.05$$

$$\mathbb{P}_{\mu_0}(\frac{\bar{x} - 600}{\frac{28.81}{\sqrt{16}}} < \frac{k - 600}{\frac{28.81}{\sqrt{16}}}) = 0.05$$

La valeur $t < t_\alpha^{(15)} = -t(_{1-\alpha}^{(15)}$*. d'où*

$$\frac{\bar{x} - 600}{\frac{28.81}{\sqrt{16}}} = -1.753 \Rightarrow k = 587.38$$

La valeur de la moyenne arithmétique donnée par l'échantillon étant égale à 581.60, on doit ***rejeter l'hypothèse*** H_0*.*
Risque de deuxième espèce :
La forme de l'hypothèse H_1 *implique que l'on doit calculer le risque pour toutes les valeurs de* $\mu < 600$*. On fait le calcul pour* $\mu_1 = 575$ *par exemple :*

$$\begin{aligned}
\beta &= \mathbb{P}(\overline{X} > k | H_1) \\
&= \mathbb{P}(\frac{k - 575}{28.81/4} > \frac{587.38 - 575}{28.81/4} | H_1) \\
&= \mathbb{P}(\frac{k - 575}{28.81/4} > 1.71 | H_1) \\
&\approx 0.05
\end{aligned}$$

Exemple 7.7.1.2. *En un point de captage d'une source on a répété six mesures du taux d'oxygène dissous dans l'eau (en parties par million). On a trouvé :*

$$4.92 \;\; 5.10 \;\; 4.93 \;\; 5.02 \;\; 5.06 \;\; 4.71.$$

La norme en dessous de laquelle on ne doit pas descendre pour la potabilité de l'eau est 5 ppm. Au vu des observations effectuées peut-on avec un faible risque d'erreur affirmer que l'eau n'est pas potable (admettre une distribution quasi-gaussienne des aléas des mesures) ?

On teste

$$H_0 : \mu \geq 5 \text{ et } H_1 : \mu < 5$$

L'alternative H_1 correspondant à l'affirmation (eau non potable) dont le risque doit être contrôlé.
Sous H_0

$$\frac{\bar{X} - 5}{S^*/\sqrt{6}} \to t(5)$$

et l'on doit rejeter H_0 au niveau de risque α si T prend une valeur $t < t_\alpha^{(5)} = -t_{1-\alpha}^{(5)}$ Caractéristiques de l'échantillon : $\bar{x} = 4.9567$ et $S^* = 0.1401$, soit

$$\frac{4.956\bar{7} - 5}{0.1401/\sqrt{6}} = -0.757$$

Pour $\alpha = 0.05, t_{0.05}^{(5)} = -2.105$, on doit accepter H_0

7.7.2 Tests sur la variance σ^2 d'une loi $\mathcal{N}(\mu, \sigma^2)$

Cas où μ est inconnu :

On sait que la variance empirique corrigée S^{*2} est un estimateur sans biais de σ^2 t.q. : n a

$$\frac{(n-1)S^{*2}}{\sigma^2} \to \chi^2_{n-1}$$

Ainsi pour une hypothèse nulle $H_0 : \sigma^2 = \sigma_0^2$ contre $H_1 : \sigma^2 \neq \sigma_0^2$, on a, sous H_0 :

$$\frac{(n-1)S^{*2}}{\sigma_0^2} \to \chi^2_{n-1}$$

Comme

$$\begin{aligned} 1 - \alpha &= \mathbb{P}_{\sigma_0^2}(-\chi_{\alpha/2}^{2(n-1)} \leq \frac{(n-1)S^{*2}}{\sigma_0^2} \leq \chi_{1-\alpha/2}^{2(n-1)}) \\ &= \mathbb{P}_{\sigma_0^2}(-\chi_{\alpha/2}^{2(n-1)} \frac{\sigma_0^2}{n-1} \leq S^{2*} \leq \chi_{1-\alpha/2}^{2(n-1)} \frac{\sigma_0^2}{n-1}) \end{aligned}$$

donc, une région de non-rejet peut être définie, pour un test de risque α, par :

$$A = \{(x_1, x_2, ..., x_n) \in \mathbb{R}^n | -\chi^{2(n-1)}_{\alpha/2} \frac{\sigma_0^2}{n-1} \leq S^{2*} \leq \chi^{2(n-1)}_{1-\alpha/2} \frac{\sigma_0^2}{n-1}\}$$

Remarque 7.7.2.1. *Pour des hypothèses unilatérales, par exemple $H_0 : \sigma^2 \leq \sigma_0^2$ contre $H_1 : \sigma^2 > \sigma_0^2$, il est naturel de On a pour région de rejet*

$$\overline{A} = \{(x_1, ..., x_n) \in \mathbb{R}^n | S^{*2} > \frac{\sigma_0^2}{n-1} \chi^{2(n-1)}_{1-\alpha}\}$$

Pour $H_0 : \sigma^2 \geq \sigma_0^2$ contre $H_1 : \sigma^2 < \sigma_0^2$, la région de rejet sera

$$\overline{A} = \{(x_1, ..., x_n) \in \mathbb{R}^n | S^{*2} < \frac{\sigma_0^2}{n-1} \chi^{2(n-1)}_{\alpha}\}$$

Exemple 7.7.2.1. *On veut tester la précision d'une méthode de mesure du cholestérol émie sur un échantillon sanguin. La précision est définie comme étant égale à deux fois l'écart-type de l'aléa (supposé pratiquement gaussien) de la méthode.*
On partage l'échantillon de référence en 6 éprouvettes que l'on soumet à l'analyse d'un laboratoire. Les valeurs trouvées en g/litre sont :

1.35 1.26 1.48 1.32 1.50 1.44.

Tester l'hypothèse nulle que la précision est inférieure ou égale à 0,1 g/litre au niveau 0,.05.

L'hypothèse à tester est $2\sigma \leq 0.14$, ce qui équivaut à $H_0 : \sigma^2 \leq 0.0025$ contre $H_1 : \sigma^2 > 0.0025$
Sur la base des observations on trouve $S^{*2} = 0.009216$, $\chi^{2(5)}_{0.95} = 11.1$

$$\frac{\sigma_0^2}{n-1} \chi^{2(n-1)}_{1-\alpha} = \frac{0.0025}{5} \times 11.1 = 0.005, \textit{ on doit rejeter } H_0$$

7.7.3 Test pour une proportion

Soit X une variable aléatoire observée sur une population et soit $(X_1, X_2, .., X_n)$ un n-échantillon extrait de cette population. On veut savoir si cet échantillon de fréquence observée $\frac{S_n}{n} = \hat{p}$ estimateur de p, appartient à une population de fréquence p_0 (sous H_0) ou à une autre population inconnu de fréquence p (sous H_1).

$$\begin{cases} p = p_0 & \textit{contre} \\ p \neq p_0 \end{cases}$$

On peut utiliser une approximation gaussienne de la loi binomiale par le théorème de la limite centrale sous la condition que les nombres np_0 et $n(1-p_0)$ soient suffsamment grands.
D'où

$$\hat{p} = \frac{S_n}{n} \rightsquigarrow \mathcal{N}(p_0, \sqrt{\frac{p_0 q_0}{n}}) \text{ sous } H_0$$

Comme

$$\begin{aligned} 1-\alpha &= \mathbb{P}(-z_{1-\frac{\alpha}{2}} \leq \frac{\hat{p}-p_0}{\sqrt{\frac{p_0q_0}{n}}} \leq z_{1-\frac{\alpha}{2}}) \\ &= \mathbb{P}(p_0 - z_{1-\frac{\alpha}{2}}\sqrt{\frac{p_0q_0}{n}} \leq \hat{p} \leq p_0 + z_{1-\frac{\alpha}{2}}\sqrt{\frac{p_0q_0}{n}}) \end{aligned}$$

donc, une région de non-rejet peut être définie, pour un test de risque α, par :

$$A = \{(x_1, x_2, ..., x_n) \in \mathbb{R}^n | p_0 - z_{1-\frac{\alpha}{2}}\sqrt{\frac{p_0q_0}{n}} \leq \hat{p} \leq p_0 + z_{1-\frac{\alpha}{2}}\sqrt{\frac{p_0q_0}{n}}\}$$

où $z_{1-\frac{\alpha}{2}}$ est lue sur la table $\mathcal{N}(0,1)$

Remarque 7.7.3.1. *Pour des hypothèses unilatérales, par exemple $H_0 : p \leq p_0$ contre $H_1 : p > p_0$, il est naturel de On a pour région de rejet*

$$\overline{A} = \{(x_1, ..., x_n) \in \mathbb{R}^n | \hat{p} > p_0 + z_{1-\alpha}\sqrt{\frac{p_0q_0}{n}}\}$$

Pour $H_0 : p \geq p_0$ contre $H_1 : p < p_0$, la région de rejet sera

$$\overline{A} = \{(x_1, ..., x_n) \in \mathbb{R}^n | \hat{p} < p_0 - z_{1-\alpha}\sqrt{\frac{p_0q_0}{n}}\}$$

Exemple 7.7.3.1. *Le fournisseur d'un lot de 100 000 puces affirme que le taux de puces défectueuses ne dépasse pas 4%. Pour tester cette hypothèse 800 puces prises au hasard sont contrôlées et l'on en trouve 40 défectueuses. Effectuer un test de niveau 0.05.*

On doit tester

$$\begin{cases} p \leq 0.04 & \text{contre} \\ p > 0.04 \end{cases}$$

où p est la proportion de pièces défectueuses dan le lot. On applique l'approximation gaussienne car $np_0 = 800 \times 0.04 > 5$ et $n(1-p_0) = 800 \times 0.96 > 5$.
On a trouvé $\hat{p} = \frac{40}{800} = 0.05, z_{1-\alpha} = 1.645$

$$\frac{\hat{p}-p_0}{\sqrt{\frac{p_0q_0}{n}}} = \frac{0.05-0.04}{\sqrt{\frac{(0.04)(0.96)}{800}}} = 1.41$$

On doit accepter H_0

7.8 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 $S^{*2}_{n_1}$, issu d'une loi $\mathcal{N}(\mu_1, \sigma_1^2)$, l'autre de taille n_2, de moyenne empirique $\overline{X}_{n_2}$ et variance empirique $S^{*2}_{n_2}$, issu d'une loi $\mathcal{N}(\mu_2, \sigma_2^2)$

7.8.1 Si σ_1^2 et σ_2^2 sont connus :

On sait que $\overline{X}_{n_1} - \overline{X}_{n_2} \rightsquigarrow \mathcal{N}(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2})$ et la statistique $\frac{\overline{X}_{n_1} - \overline{X}_{n_2} - \mu_1 + \mu_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$ de loi connue $\mathcal{N}(0,1)$.
Ainsi pour une hypothèse nulle $H_0 : \mu_1 - \mu_2 = 0$ contre $H_1 : \mu_1 - \mu_2 \neq 0$, on a, sous H_0 :

$$\frac{\overline{X}_{n_1} - \overline{X}_{n_2}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \rightsquigarrow \mathcal{N}(0,1)$$

Comme

$$\begin{aligned} 1 - \alpha &= \mathbb{P}(-z_{1-\alpha/2} < \frac{\overline{X}_{n_1} - \overline{X}_{n_2}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} < z_{1-\alpha/2}) \\ &= \mathbb{P}(-z_{1-\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} < \overline{X}_{n_1} - \overline{X}_{n_2} < z_{1-\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}) \end{aligned}$$

donc, une région de non-rejet peut être définie, pour un test de risque α, par :

$$A = \{-z_{1-\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} < \overline{X}_{n_1} - \overline{X}_{n_2} < z_{1-\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\}$$

Exemple 7.8.1.1. *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 $\sigma_A^2 = 1mm$ et $\sigma_A^2 = 0.45mm$.
Peut-on jaffirmer au niveau 5% qu'il y a une différence signignificative entre les procédés de fabrication A et B ?

Soient X_A et X_B, les variables aléatoires « diamètres des tubes fabriqués suivant les procédés A et B».
On veut tester $H_0 : \mu_1 - \mu_0 = 0$ contre $H_1 : \mu_1 - \mu_2 < 0$ avec un seuil critique égal à 5%.

$$\mathbb{P}\left(-1.96 < \frac{\overline{X}_{n_1} - \overline{X}_{n_2}}{\sqrt{\frac{1}{5} + \frac{(0.45)^2}{4}}} < 1.96\right)$$

avec les données apportées par les échantillons, on obtient pour la moyenne des deux échantillons $\overline{x}_{n_1} = 52.38$ et $\overline{x}_{n_2} = 51.50$.
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.

7.8.2 Si σ_1^2 et σ_2^2 sont inconnus, mais $\sigma_1^2 = \sigma_2^2 = \sigma^2$

On a :

$$\overline{X}_1 \rightsquigarrow \mathcal{N}(\mu_1, \frac{\sigma^2}{n_1})$$

$$\overline{X}_2 \rightsquigarrow \mathcal{N}(\mu_2, \frac{\sigma^2}{n_2})$$

et

$$\overline{X}_1 - \overline{X}_2 \rightsquigarrow \mathcal{N}(\mu_1 - \mu_2, \sigma^2(\frac{1}{n_1} + \frac{1}{n_2}))$$

$$\frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sigma\sqrt{(\frac{1}{n_1} + \frac{1}{n_2})}} \rightsquigarrow \mathcal{N}(0, 1)$$

Le problème qui se pose est celui de l'estimation de σ que l'on effectue, en fait via σ^2. Sachant que

$$\frac{(n_1 - 1)S_1^{*2}}{\sigma^2} \rightsquigarrow \chi^2_{n_1 - 1}$$

$$\frac{(n_2 - 1)S_2^{*2}}{\sigma^2} \rightsquigarrow \chi^2_{n_2 - 1}$$

L'indépendance des deux échantillons entraîne que :

$$\frac{(n_1 - 1)S_1^{*2} - (n_2 - 1)S_2^{*2}}{\sigma^2} \rightsquigarrow \chi^2_{n_1 + n_2 - 2}$$

En faisant le rapport de la variable aléatoire $\overline{X}_1 - \overline{X}_2$ centrée réduite à la racine carrée de la v.a. ci-dessus divisée par ses degrés de liberté, et en posant :

$$S^{*2}_{n_1 n_2} = \frac{(n_1 - 1)S^{*2}_{n_1} + (n_2 - 1)S^{*2}_{n_2}}{n_1 + n_2 - 2}$$

on obtient :

$$\frac{(\bar{X}_{n_1} - \bar{X}_{n_2}) - (\mu_1 - \mu_2)}{\sqrt{S^{*2}_{n_1 n_2}(\frac{1}{n_1} + \frac{1}{n_2})}} \rightsquigarrow t(n_1 + n_2 - 2)$$

où $S^{*2}_{n_1 n_2} = \frac{(n_1-1)S^{*2}_{n_1} + (n_2-1)S^{*2}_{n_2}}{n_1+n_2-2}$ *est un estimateur sans biais de la variance commune* σ^2.

Ainsi pour une hypothèse nulle $H_0 : \mu_1 - \mu_2 = 0$ *contre* $H_1 : \mu_1 - \mu_2 \neq 0$ *on a, sous* H_0 *:*

$$\frac{(\bar{X}_{n_1} - \bar{X}_{n_2})}{\sqrt{S^{*2}_{n_1 n_2}(\frac{1}{n_1} + \frac{1}{n_2})}} \rightsquigarrow t(n_1 + n_2 - 2)$$

ce qui permet de définir une région de non-rejet peut être définie, pour un test de risque α*, par :*

$$A = \{-t^{(n_1+n_2-2)}_{1-\frac{\alpha}{2}} \sqrt{S^{*2}_{n_1 n_2}(\frac{1}{n_1} + \frac{1}{n_2})} \leq \bar{x}_{n_1} - \bar{x}_{n_2} \leq t^{(n_1+n_2-2)}_{1-\frac{\alpha}{2}} \sqrt{S^{*2}_{n_1 n_2}(\frac{1}{n_1} + \frac{1}{n_2})}\}$$

7.9 Tests de comparaison des variances de deux lois de Gauss

7.9.1 Si μ_1 et μ_2 sont inconnus

On considère le rapport $\frac{\sigma_1^2}{\sigma_2^2}$ *pour les lois* $\mathcal{N}(\mu_1, \sigma_1)$ *et* $\mathcal{N}(\mu_1, \sigma_1)$*. Comme :*

$$\frac{(n_1 - 1)S^{*2}}{\sigma_1^2} \rightsquigarrow \chi^2_{n_1-1} \text{ et } \frac{(n_2 - 1)S^{*2}}{\sigma_2^2} \rightsquigarrow \chi^2_{n_2-1}$$

On a

$$\frac{S^{*2}_1/\sigma_1^2}{S^{*2}_2/\sigma_1^2} \rightsquigarrow F(n_1 - 1, n_2 - 1)$$

Ainsi pour une hypothèse nulle

$$\begin{cases} \frac{\sigma_1^2}{\sigma_2^2} = 1 & \text{contre} \\ \frac{\sigma_1^2}{\sigma_2^2} \neq 1 \end{cases}$$

On a sous H_0*,*

$$\frac{S^{*2}_1}{S^{*2}_1} \text{ suit la loi de Fisher } F(n_1 - 1, n_2 - 1)$$

Ce qui permet de définir une région de non-rejet peut être définie,pour un test de risque α *:*

$$A = \{F^{(n_1-1,n_2-1)}_{\frac{\alpha}{2}} \leq \frac{S^{*2}_1}{S^{*2}_1} \leq F^{(n_1-1,n_2-1)}_{1-\frac{\alpha}{2}}\}$$

Remarque 7.9.1.1. *Ce test est basé sur la statistique de Fisher : on a :* $F_{\frac{\alpha}{2}}^{(n_1-1,n_2-1)} = 1/F_{1-\frac{\alpha}{2}}^{(n_2-1,n_1-1)}$

Exemple 7.9.1.1. *Soit* $X_1, X_2, ..., X_{n_1}$ *un* n_1*-échantillon issu de* $X_{n_1} \rightsquigarrow \mathcal{N}(n_1, \sigma_1^2)$ *tel que :*
$n_1 = 7$,$\bar{x}_{n_1} = 14.2$, $S_1^{*2} = 0.0114$. *Soit* $X_1, X_2, ..., X_{n_2}$ *un* n_2*-échantillon issu de* $X_{n_2} \rightsquigarrow \mathcal{N}(n_2, \sigma_2^2)$ *tel que :*
$n_2 = 5$,$\bar{x}_{n_2} = 14.5$, $S_2^{*2} = 0.05$
Peut-on accepter que X_{n_1} *et* X_{n_2} *suivent la même loi normale au niveau* $\alpha = 0.05$ *?*
On veut tester $H_0 : \sigma_1^2 = \sigma_2^2$ *contre* $H_1 : \sigma_1^2 \neq \sigma_2^2$ *Ce test est basé sur la statistique de Fisher :*
on a :

$$\frac{0.0114}{0.05} = 0.228$$

$F_{\frac{\alpha}{2}}^{(n_1-1,n_2-1)} = 1/F_{1-\frac{\alpha}{2}}^{(n_2-1,n_1-1)} = 1/F_{0.975}^{(4,6)} = 1/6.23 = 0.160$
$F_{1-\frac{\alpha}{2}}^{(n_1-1,n_2-1)} = F_{0.975}^{(6,4)} = 9.20$
on ne rejette pas H_0 *car* $0.228 \in [0.160, 9.20]$.

Annexe

Table 1

Loi Normale Centrée Réduite

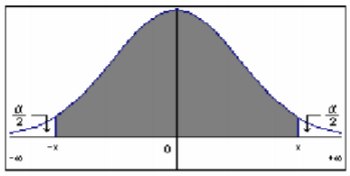
Fonction de répartition F(z)=P(Z<z)

Exemple : P(Z<1.96)= 0.97500 se trouve en ligne 1.9 et colonne 0.06

z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,50000	0,50399	0,50798	0,51197	0,51595	0,51994	0,52392	0,52790	0,53188	0,53586
0,1	0,53983	0,54380	0,54776	0,55172	0,55567	0,55962	0,56356	0,56750	0,57142	0,57535
0,2	0,57926	0,58317	0,58706	0,59095	0,59484	0,59871	0,60257	0,60642	0,61026	0,61409
0,3	0,61791	0,62172	0,62552	0,62930	0,63307	0,63683	0,64058	0,64431	0,64803	0,65173
0,4	0,65542	0,65910	0,66276	0,66640	0,67003	0,67365	0,67724	0,68082	0,68439	0,68793
0,5	0,69146	0,69498	0,69847	0,70194	0,70540	0,70884	0,71226	0,71566	0,71904	0,72241
0,6	0,72575	0,72907	0,73237	0,73565	0,73891	0,74215	0,74537	0,74857	0,75175	0,75490
0,7	0,75804	0,76115	0,76424	0,76731	0,77035	0,77337	0,77637	0,77935	0,78231	0,78524
0,8	0,78815	0,79103	0,79389	0,79673	0,79955	0,80234	0,80511	0,80785	0,81057	0,81327
0,9	0,81594	0,81859	0,82121	0,82382	0,82639	0,82894	0,83147	0,83398	0,83646	0,83891
1,0	0,84135	0,84375	0,84614	0,84850	0,85083	0,85314	0,85543	0,85769	0,85993	0,86214
1,1	0,86433	0,86650	0,86864	0,87076	0,87286	0,87493	0,87698	0,87900	0,88100	0,88298
1,2	0,88493	0,88686	0,88877	0,89065	0,89251	0,89435	0,89617	0,89796	0,89973	0,90148
1,3	0,90320	0,90490	0,90658	0,90824	0,90988	0,91149	0,91309	0,91466	0,91621	0,91774
1,4	0,91924	0,92073	0,92220	0,92364	0,92507	0,92647	0,92786	0,92922	0,93056	0,93189
1,5	0,93319	0,93448	0,93575	0,93699	0,93822	0,93943	0,94062	0,94179	0,94295	0,94408
1,6	0,94520	0,94630	0,94738	0,94845	0,94950	0,95053	0,95154	0,95254	0,95352	0,95449
1,7	0,95544	0,95637	0,95728	0,95819	0,95907	0,95994	0,96080	0,96164	0,96246	0,96327
1,8	0,96407	0,96485	0,96562	0,96638	0,96712	0,96784	0,96856	0,96926	0,96995	0,97062
1,9	0,97128	0,97193	0,97257	0,97320	0,97381	0,97441	0,97500	0,97558	0,97615	0,97670
2,0	0,97725	0,97778	0,97831	0,97882	0,97933	0,97982	0,98030	0,98077	0,98124	0,98169
2,1	0,98214	0,98257	0,98300	0,98341	0,98382	0,98422	0,98461	0,98500	0,98537	0,98574
2,2	0,98610	0,98645	0,98679	0,98713	0,98745	0,98778	0,98809	0,98840	0,98870	0,98899
2,3	0,98928	0,98956	0,98983	0,99010	0,99036	0,99061	0,99086	0,99111	0,99134	0,99158
2,4	0,99180	0,99202	0,99224	0,99245	0,99266	0,99286	0,99305	0,99324	0,99343	0,99361
2,5	0,99379	0,99396	0,99413	0,99430	0,99446	0,99461	0,99477	0,99492	0,99506	0,99520
2,6	0,99534	0,99547	0,99560	0,99573	0,99585	0,99598	0,99609	0,99621	0,99632	0,99643
2,7	0,99653	0,99664	0,99674	0,99683	0,99693	0,99702	0,99711	0,99720	0,99728	0,99736
2,8	0,99744	0,99752	0,99760	0,99767	0,99774	0,99781	0,99788	0,99795	0,99801	0,99807
2,9	0,99813	0,99819	0,99825	0,99831	0,99836	0,99841	0,99846	0,99851	0,99856	0,99861
3,0	0,99865	0,99869	0,99874	0,99878	0,99882	0,99886	0,99889	0,99893	0,99897	0,99900
3,1	0,99903	0,99906	0,99910	0,99913	0,99916	0,99918	0,99921	0,99924	0,99926	0,99929
3,2	0,99931	0,99934	0,99936	0,99938	0,99940	0,99942	0,99944	0,99946	0,99948	0,99950
3,3	0,99952	0,99953	0,99955	0,99957	0,99958	0,99960	0,99961	0,99962	0,99964	0,99965
3,4	0,99966	0,99968	0,99969	0,99970	0,99971	0,99972	0,99973	0,99974	0,99975	0,99976

Table 2

Loi de Student



α	1	0,8	0,6	0,4	0,2	0,1	0,05	0,02	0,01	0,002	0,001
1 − α	0	0,2	0,4	0,6	0,8	0,9	0,95	0,98	0,99	0,998	0,999
ν = ddl											
1	0,0000	0,3249	0,7265	1,3764	3,0777	6,3137	12,706	31,821	63,656	318,29	636,58
2	0,0000	0,2887	0,6172	1,0607	1,8856	2,9200	4,3027	6,9645	9,9250	22,328	31,600
3	0,0000	0,2767	0,5844	0,9785	1,6377	2,3534	3,1824	4,5407	5,8408	10,214	12,924
4	0,0000	0,2707	0,5686	0,9410	1,5332	2,1318	2,7765	3,7469	4,6041	7,1729	8,6101
5	0,0000	0,2672	0,5594	0,9195	1,4759	2,0150	2,5706	3,3649	4,0321	5,8935	6,8685
6	0,0000	0,2648	0,5534	0,9057	1,4398	1,9432	2,4469	3,1427	3,7074	5,2075	5,9587
7	0,0000	0,2632	0,5491	0,8960	1,4149	1,8946	2,3646	2,9979	3,4995	4,7853	5,4081
8	0,0000	0,2619	0,5459	0,8889	1,3968	1,8595	2,3060	2,8965	3,3554	4,5008	5,0414
9	0,0000	0,2610	0,5435	0,8834	1,3830	1,8331	2,2622	2,8214	3,2498	4,2969	4,7809
10	0,0000	0,2602	0,5415	0,8791	1,3722	1,8125	2,2281	2,7638	3,1693	4,1437	4,5868
11	0,0000	0,2596	0,5399	0,8755	1,3634	1,7959	2,2010	2,7181	3,1058	4,0248	4,4369
12	0,0000	0,2590	0,5386	0,8726	1,3562	1,7823	2,1788	2,6810	3,0545	3,9296	4,3178
13	0,0000	0,2586	0,5375	0,8702	1,3502	1,7709	2,1604	2,6503	3,0123	3,8520	4,2209
14	0,0000	0,2582	0,5366	0,8681	1,3450	1,7613	2,1448	2,6245	2,9768	3,7874	4,1403
15	0,0000	0,2579	0,5357	0,8662	1,3406	1,7531	2,1315	2,6025	2,9467	3,7329	4,0728
16	0,0000	0,2576	0,5350	0,8647	1,3368	1,7459	2,1199	2,5835	2,9208	3,6861	4,0149
17	0,0000	0,2573	0,5344	0,8633	1,3334	1,7396	2,1098	2,5669	2,8982	3,6458	3,9651
18	0,0000	0,2571	0,5338	0,8620	1,3304	1,7341	2,1009	2,5524	2,8784	3,6105	3,9217
19	0,0000	0,2569	0,5333	0,8610	1,3277	1,7291	2,0930	2,5395	2,8609	3,5793	3,8833
20	0,0000	0,2567	0,5329	0,8600	1,3253	1,7247	2,0860	2,5280	2,8453	3,5518	3,8496
21	0,0000	0,2566	0,5325	0,8591	1,3232	1,7207	2,0796	2,5176	2,8314	3,5271	3,8193
22	0,0000	0,2564	0,5321	0,8583	1,3212	1,7171	2,0739	2,5083	2,8188	3,5050	3,7922
23	0,0000	0,2563	0,5317	0,8575	1,3195	1,7139	2,0687	2,4999	2,8073	3,4850	3,7676
24	0,0000	0,2562	0,5314	0,8569	1,3178	1,7109	2,0639	2,4922	2,7970	3,4668	3,7454
25	0,0000	0,2561	0,5312	0,8562	1,3163	1,7081	2,0595	2,4851	2,7874	3,4502	3,7251
26	0,0000	0,2560	0,5309	0,8557	1,3150	1,7056	2,0555	2,4786	2,7787	3,4350	3,7067
27	0,0000	0,2559	0,5306	0,8551	1,3137	1,7033	2,0518	2,4727	2,7707	3,4210	3,6895
28	0,0000	0,2558	0,5304	0,8546	1,3125	1,7011	2,0484	2,4671	2,7633	3,4082	3,6739
29	0,0000	0,2557	0,5302	0,8542	1,3114	1,6991	2,0452	2,4620	2,7564	3,3963	3,6595
30	0,0000	0,2556	0,5300	0,8538	1,3104	1,6973	2,0423	2,4573	2,7500	3,3852	3,6460
40	0,0000	0,2550	0,5286	0,8507	1,3031	1,6839	2,0211	2,4233	2,7045	3,3069	3,5510
50	0,0000	0,2547	0,5278	0,8489	1,2987	1,6759	2,0086	2,4033	2,6778	3,2614	3,4960
60	0,0000	0,2545	0,5272	0,8477	1,2958	1,6706	2,0003	2,3901	2,6603	3,2317	3,4602
70	0,0000	0,2543	0,5268	0,8468	1,2938	1,6669	1,9944	2,3808	2,6479	3,2108	3,4350
80	0,0000	0,2542	0,5265	0,8461	1,2922	1,6641	1,9901	2,3739	2,6387	3,1952	3,4164
90	0,0000	0,2541	0,5263	0,8456	1,2910	1,6620	1,9867	2,3685	2,6316	3,1832	3,4019
100	0,0000	0,2540	0,5261	0,8452	1,2901	1,6602	1,9840	2,3642	2,6259	3,1738	3,3905
200	0,0000	0,2537	0,5252	0,8434	1,2858	1,6525	1,9719	2,3451	2,6006	3,1315	3,3398
∞	0,0000	0,2533	0,5244	0,8416	1,2816	1,6449	1,9600	2,3263	2,5758	3,0903	3,2906

Table 3

Loi du χ^2

$$P(\chi^2_\nu \geq \chi^2_{\nu,\alpha}) = \alpha$$

1 – α	0,001	0,005	0,01	0,025	0,05	0,1	0,5	0,9	0,95	0,975	0,99	0,995	0,99
α	0,999	0,995	0,99	0,975	0,95	0,9	0,5	0,1	0,05	0,025	0,01	0,005	0,00
ν = ddl													
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	10,8
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	13,8
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	16,2
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	18,4
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	20,5
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	22,4
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	24,3
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	26,1
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	27,8
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	29,5
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	31,2
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	32,9
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	34,5
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	36,1
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	37,7
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	39,2
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	40,7
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	42,3
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	43,8
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	45,3
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	46,8
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	48,2
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	49,7
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	51,1
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	52,6
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	54,0
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	55,4
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	56,8
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	58,3
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	59,7

Pour ν > 30, La loi du χ^2 peut –être approximée par la loi normale $N(\nu, \sqrt{\nu})$

Table 4

$$P(F_{\nu_1,\nu_2} < f_{\nu_1,\nu_2,\alpha}) = \alpha$$

$\alpha = 0{,}975$

ν_2 \ ν_1	1	2	3	4	5	6	7	8	9	10	15	20	30	50	100	200	500	•
1	648	800	864	900	922	937	948	957	963	969	985	993	1001	1008	1013	1016	1017	1018
2	38,5	39,0	39,2	39,3	39,3	39,4	39,4	39,4	39,4	39,4	39,4	39,4	39,5	39,5	39,5	39,5	39,5	39,5
3	17,4	16,0	15,4	15,1	14,9	14,7	14,6	14,5	14,5	14,4	14,3	14,2	14,1	14,0	14,0	13,9	13,9	13,9
4	12,2	10,6	9,98	9,60	9,36	9,20	9,07	8,98	8,90	8,84	8,66	8,56	8,46	8,38	8,32	8,29	8,27	8,26
5	10,0	8,43	7,76	7,39	7,15	6,98	6,85	6,76	6,68	6,62	6,43	6,33	6,23	6,14	6,08	6,05	6,03	6,02
6	8,81	7,26	6,60	6,23	5,99	5,82	5,70	5,60	5,52	5,46	5,27	5,17	5,07	4,98	4,92	4,88	4,86	4,85
7	8,07	6,54	5,89	5,52	5,29	5,12	4,99	4,90	4,82	4,76	4,57	4,47	4,36	4,28	4,21	4,18	4,16	4,14
8	7,57	6,06	5,42	5,05	4,82	4,65	4,53	4,43	4,36	4,30	4,10	4,00	3,89	3,81	3,74	3,70	3,68	3,67
9	7,21	5,71	5,08	4,72	4,48	4,32	4,20	4,10	4,03	3,96	3,77	3,67	3,56	3,47	3,40	3,37	3,35	3,33
10	6,94	5,46	4,83	4,47	4,24	4,07	3,95	3,85	3,78	3,72	3,52	3,42	3,31	3,22	3,15	3,12	3,09	3,08
11	6,72	5,26	4,63	4,28	4,04	3,88	3,76	3,66	3,59	3,53	3,33	3,23	3,12	3,03	2,96	2,92	2,90	2,88
12	6,55	5,10	4,47	4,12	3,89	3,73	3,61	3,51	3,44	3,37	3,18	3,07	2,96	2,87	2,80	2,76	2,74	2,72
13	6,41	4,97	4,35	4,00	3,77	3,60	3,48	3,39	3,31	3,25	3,05	2,95	2,84	2,74	2,67	2,63	2,61	2,60
14	6,30	4,86	4,24	3,89	3,66	3,50	3,38	3,29	3,21	3,15	2,95	2,84	2,73	2,64	2,56	2,53	2,50	2,49
15	6,20	4,76	4,15	3,80	3,58	3,41	3,29	3,20	3,12	3,06	2,86	2,76	2,64	2,55	2,47	2,44	2,41	2,40
16	6,12	4,69	4,08	3,73	3,50	3,34	3,22	3,12	3,05	2,99	2,79	2,68	2,57	2,47	2,40	2,36	2,33	2,32
17	6,04	4,62	4,01	3,66	3,44	3,28	3,16	3,06	2,98	2,92	2,72	2,62	2,50	2,41	2,33	2,29	2,26	2,25
18	5,98	4,56	3,95	3,61	3,38	3,22	3,10	3,01	2,93	2,87	2,67	2,56	2,44	2,35	2,27	2,23	2,20	2,19
19	5,92	4,51	3,90	3,56	3,33	3,17	3,05	2,96	2,88	2,82	2,62	2,51	2,39	2,30	2,22	2,18	2,15	2,13
20	5,87	4,46	3,86	3,51	3,29	3,13	3,01	2,91	2,84	2,77	2,57	2,46	2,35	2,25	2,17	2,13	2,10	2,09
22	5,79	4,38	3,78	3,44	3,22	3,05	2,93	2,84	2,76	2,70	2,50	2,39	2,27	2,17	2,09	2,05	2,02	2,00
24	5,72	4,32	3,72	3,38	3,15	2,99	2,87	2,78	2,70	2,64	2,44	2,33	2,21	2,11	2,02	1,98	1,95	1,94
26	5,66	4,27	3,67	3,33	3,10	2,94	2,82	2,73	2,65	2,59	2,39	2,28	2,16	2,05	1,97	1,92	1,90	1,88
28	5,61	4,22	3,63	3,29	3,06	2,90	2,78	2,69	2,61	2,55	2,34	2,23	2,11	2,01	1,92	1,88	1,85	1,83
30	5,57	4,18	3,59	3,25	3,03	2,87	2,75	2,65	2,57	2,51	2,31	2,20	2,07	1,97	1,88	1,84	1,81	1,79
40	5,42	4,05	3,46	3,13	2,90	2,74	2,62	2,53	2,45	2,39	2,18	2,07	1,94	1,83	1,74	1,69	1,66	1,64
50	5,34	3,98	3,39	3,06	2,83	2,67	2,55	2,46	2,38	2,32	2,11	1,99	1,87	1,75	1,66	1,60	1,57	1,55
60	5,29	3,93	3,34	3,01	2,79	2,63	2,51	2,41	2,33	2,27	2,06	1,94	1,82	1,70	1,60	1,54	1,51	1,48
80	5,22	3,86	3,28	2,95	2,73	2,57	2,45	2,36	2,28	2,21	2,00	1,88	1,75	1,63	1,53	1,47	1,43	1,40
100	5,18	3,83	3,25	2,92	2,70	2,54	2,42	2,32	2,24	2,18	1,97	1,85	1,71	1,59	1,48	1,42	1,38	1,35
200	5,10	3,76	3,18	2,85	2,63	2,47	2,35	2,26	2,18	2,11	1,90	1,78	1,64	1,51	1,39	1,32	1,27	1,23
500	5,05	3,72	3,14	2,81	2,59	2,43	2,31	2,22	2,14	2,07	1,86	1,74	1,60	1,46	1,34	1,25	1,19	1,14
•	5,02	3,69	3,12	2,79	2,57	2,41	2,29	2,19	2,11	2,05	1,83	1,71	1,57	1,43	1,30	1,21	1,13	1,00

Table 5
Loi de Fisher F (suite)

$$P(F_{\nu_1,\nu_2} < f_{\nu_1,\nu_2,\alpha}) = \alpha$$

$\alpha = 0{,}95$

ν_2 \ ν_1	1	2	3	4	5	6	7	8	9	10	15	20	30	50	100	200	500	•
1	161	200	216	225	230	234	237	239	241	242	246	248	250	252	253	254	254	254
2	18,5	19,0	19,2	19,2	19,3	19,3	19,4	19,4	19,4	19,4	19,4	19,4	19,5	19,5	19,5	19,5	19,5	19,5
3	10,1	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79	8,70	8,66	8,62	8,58	8,55	8,54	8,53	8,53
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,86	5,80	5,75	5,70	5,66	5,65	5,64	5,63
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74	4,62	4,56	4,50	4,44	4,41	4,39	4,37	4,37
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	3,94	3,87	3,81	3,75	3,71	3,69	3,68	3,67
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64	3,51	3,44	3,38	3,32	3,27	3,25	3,24	3,23
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35	3,22	3,15	3,08	3,02	2,97	2,95	2,94	2,93
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14	3,01	2,94	2,86	2,80	2,76	2,73	2,72	2,71
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98	2,85	2,77	2,70	2,64	2,59	2,56	2,55	2,54
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,85	2,72	2,65	2,57	2,51	2,46	2,43	2,42	2,40
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80	2,75	2,62	2,54	2,47	2,40	2,35	2,32	2,31	2,30
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71	2,67	2,53	2,46	2,38	2,31	2,26	2,23	2,22	2,21
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65	2,60	2,46	2,39	2,31	2,24	2,19	2,16	2,14	2,13
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59	2,54	2,40	2,33	2,25	2,18	2,12	2,10	2,08	2,07
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,35	2,28	2,19	2,12	2,07	2,04	2,02	2,01
17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49	2,45	2,31	2,23	2,15	2,08	2,02	1,99	1,97	1,96
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41	2,27	2,19	2,11	2,04	1,98	1,95	1,93	1,92
19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42	2,38	2,23	2,16	2,07	2,00	1,94	1,91	1,89	1,88
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39	2,35	2,20	2,12	2,04	1,97	1,91	1,88	1,86	1,84
22	4,30	3,44	3,05	2,82	2,66	2,55	2,46	2,40	2,34	2,30	2,15	2,07	1,98	1,91	1,85	1,82	1,80	1,78
24	4,26	3,40	3,01	2,78	2,62	2,51	2,42	2,36	2,30	2,25	2,11	2,03	1,94	1,86	1,80	1,77	1,75	1,73
26	4,23	3,37	2,98	2,74	2,59	2,47	2,39	2,32	2,27	2,22	2,07	1,99	1,90	1,82	1,76	1,73	1,71	1,69
28	4,20	3,34	2,95	2,71	2,56	2,45	2,36	2,29	2,24	2,19	2,04	1,96	1,87	1,79	1,73	1,69	1,67	1,65
30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21	2,16	2,01	1,93	1,84	1,76	1,70	1,66	1,64	1,62
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12	2,08	1,92	1,84	1,74	1,66	1,59	1,55	1,53	1,51
50	4,03	3,18	2,79	2,56	2,40	2,29	2,20	2,13	2,07	2,03	1,87	1,78	1,69	1,60	1,52	1,48	1,46	1,44
60	4,00	3,15	2,76	2,53	2,37	2,25	2,17	2,10	2,04	1,99	1,84	1,75	1,65	1,56	1,48	1,44	1,41	1,39
80	3,96	3,11	2,72	2,49	2,33	2,21	2,13	2,06	2,00	1,95	1,79	1,70	1,60	1,51	1,43	1,38	1,35	1,32
100	3,94	3,09	2,70	2,46	2,31	2,19	2,10	2,03	1,97	1,93	1,77	1,68	1,57	1,48	1,39	1,34	1,31	1,28
200	3,89	3,04	2,65	2,42	2,26	2,14	2,06	1,98	1,93	1,88	1,72	1,62	1,52	1,41	1,32	1,26	1,22	1,19
500	3,86	3,01	2,62	2,39	2,23	2,12	2,03	1,96	1,90	1,85	1,69	1,59	1,48	1,38	1,28	1,21	1,16	1,11
•	3,84	3,00	2,60	2,37	2,21	2,10	2,01	1,94	1,88	1,83	1,67	1,57	1,46	1,35	1,24	1,17	1,11	1,00

Bibliographie

[1] *A. Meot. Introduction aux statistiques inférentielles : de la logique à la pratique avec exercices et corrigés. édition (de boeck), 2003.*

[2] *A.Perrut. Cours de Probabilités et Statistiques. Université Claude Bernard Lyon 1, Août 2010.*

[3] *B.Verlant. Statistique et probabilités BTS industriel groupements. Foucher, T2, 2009.*

[4] *D.Foudrinier. Statistique inférentielle. Dunod, Paris 2002.*

[5] *D.Foata, J.Franchi, A.Fuchs. Calcul des probabilités (cours, exercices et problème corrigée). $3^{\text{ème}}$ édition, Dunod, Paris 2012.*

[6] *D. Yadolah. Premiers pas en statistique. Springer, Paris, 2003.*

[7] *G. Saporta. Probabilité, analyse des données et statistique. Editions Technip, 1990.*

[8] *J.P. Lecoutre. Statistique et probabilité, manuel et exercices corrigés. quatrième édition. Masson, 2009.*

[9] *J.P.Michel. Statistical inference : a short course. Wiley, 2012.*

[10] *L. Michel. Statistique : la théorie et ses applications. Springer-Verlag France, Paris, 2010.*

[11] *M. Lefebvre. Probabilités, statistique et applications. Presses Internationales Polytechnique, 2011.*

[12] *M. Jeanblanc. Cours de Calcul stochastique. Master 2IF EVRY , Septembre ,2006.*

[13] *P. Tassi. Méthodes statistiques. Edition Economica, 2004.*

[14] *P.K. Sahu, S.R. Pal, A.K. Das. Estimation and inferential statistics. Springer, 2015.*

[15] *P.J. Bickel , K.Adoksu. Mathematical statistics. . Prentice Hall 1977-2001.*

[16] *R. Veysseyre. Aide-mémoire : Statistique et probabilités pour l'ingénieur. Dunod, Paris, 2006.*

[17] Y.Suhov,M. Kelber. Probability and statistics by example. Cambridge University Press, 2005.

[18] Y.Velenik. Probabilités et Statistique. Université de Genève, Version Préliminaire de 20 Mars 2017.