Notes de cours - Statistiques inférentielles

L3 parcours *Mathématiques générales* et *Statistique et applications*

Arnaud Poinas

Année 2023-2024

Table des matières

Chapitre I

Suites aléatoires et théorèmes limites

Cadre : On considère un espace de probabilités $(\Omega, \mathcal{T}, \mathbb{P})$. On rappelle qu'une variable aléatoire réelle (v.a.r) X est une application mesurable de l'espace $(\Omega, \mathcal{T}, \mathbb{P})$ dans l'espace $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Sauf indication contraire, **toutes les variables que l'on considère sont des v.a.r. définies sur cet espace de probabilité.** On s'intéresse dans cette section à des suites de v.a.r. $(X_n)_{n\in\mathbb{N}}$ et leur comportement lorsque $n \to \infty$.

1 Convergence presque sûre

Définition 1

Une suite $(X_n)_{n\in\mathbb{N}}$ de v.a.r. ***converge presque sûrement*** *vers une v.a.r. X si*

$$\mathbb{P}(X_n \underset{n\to+\infty}{\longrightarrow} X) = 1.$$

On note $X_n \xrightarrow[n\to+\infty]{p.s.} X$.

Exemples:

- Soit $U \sim \mathcal{U}([0,1])$, une variable uniforme sur l'intervalle $[0,1]$. On note
$$X_n = \mathbb{1}_{U\leqslant \frac{1}{2}+\frac{1}{n}} \quad \text{et} \quad X = \mathbb{1}_{U<\frac{1}{2}}.$$
On a $X_n - X = \mathbb{1}_{\frac{1}{2}\leqslant U\leqslant \frac{1}{2}+\frac{1}{n}}$ donc $X_n - X$ est décroissant et à valeur dans $\{0,1\}$. X_n ne converge donc pas vers X ssi $X_n - X = 1$ pour tout $n \in \mathbb{N}$ et donc $U \in \bigcap_{n\geqslant 0}[1/2, 1/2 + 1/n] = \{1/2\}$. D'où,
$$\mathbb{P}(X_n \underset{n\to+\infty}{\not\to} X) = \mathbb{P}(U = 1/2) = 0 \;\Rightarrow\; X_n \xrightarrow[n\to+\infty]{p.s.} X.$$
- Soit $(X_n)_{n\in\mathbb{N}}$ une suite de v.a.r. indépendantes vérifiant $X_n \sim b(1/n)$. Comme X_n est à valeur dans $\{0,1\}$ alors $X_n \underset{n\to+\infty}{\longrightarrow} 0$ ssi $\exists N > 0$ tel que $n \geqslant N \Rightarrow X_n = 0$. Soit $N \in \mathbb{N}$,
$$\begin{aligned}\mathbb{P}(\forall n \geqslant N,\ X_n = 0) &= \prod_{n=N}^{+\infty} \mathbb{P}(X_n = 0)\\ &= \prod_{n=N}^{+\infty}\left(1 - \frac{1}{n}\right)\\ &\leqslant \prod_{n=N}^{+\infty} e^{-\frac{1}{n}}\\ &= e^{-\sum_{n=N}^{\infty}\frac{1}{n}} = 0 \text{ car la série } \sum_{n=0}^{\infty}\frac{1}{n} \text{ diverge.}\end{aligned}$$

d'où

$$\mathbb{P}(\exists N > 0 \text{ tel que } n \geqslant N \Rightarrow X_n = 0) = 0 = \mathbb{P}(X_n \underset{n\to+\infty}{\longrightarrow} 0)$$

et donc X_n ne converge pas presque sûrement vers 0.

2 Convergence en probabilités

Définition 2

Une suite $(X_n)_{n\in\mathbb{N}}$ de v.a.r. ***converge en probabilité*** *vers une v.a.r. X si*

$$\forall \varepsilon > 0,\ \mathbb{P}(|X_n - X| > \varepsilon) \underset{n\to+\infty}{\longrightarrow} 0.$$

On note $X_n \overset{\mathbb{P}}{\underset{n\to+\infty}{\longrightarrow}} X$.

Exemple: Soit $X_n \sim b(1/n)$. Comme X_n est à valeur dans $\{0,1\}$ alors, pour tout $\varepsilon \in]0,1[$ on a $|X_n| > \varepsilon \Leftrightarrow X_n = 1$ donc $\mathbb{P}(|X_n| > \varepsilon) = \mathbb{P}(X_n = 1) = \frac{1}{n} \underset{n\to+\infty}{\longrightarrow} 0$ d'où $X_n \overset{\mathbb{P}}{\underset{n\to+\infty}{\longrightarrow}} 0$.

Proposition 3

$$X_n \overset{p.s.}{\underset{n\to+\infty}{\longrightarrow}} X \Longrightarrow X_n \overset{\mathbb{P}}{\underset{n\to+\infty}{\longrightarrow}} X.$$

Démonstration : On suppose $X_n \overset{p.s.}{\underset{n\to+\infty}{\longrightarrow}} X$. Soit $\varepsilon > 0$. Avec probabilité égale à 1 on a $X_n \underset{n\to+\infty}{\longrightarrow} X$ et donc il existe $N \in \mathbb{N}$ tel que $\forall n \geqslant N$, $|X_n - X| \leqslant \varepsilon$. Autrement dit,

$$1 = \mathbb{P}\left(\bigcup_{N\in\mathbb{N}} \{\forall n \geqslant N,\ |X_n - X| \leqslant \varepsilon\}\right) = \mathbb{P}\left(\bigcup_{N\in\mathbb{N}} \bigcap_{n\geqslant N} \{|X_n - X| \leqslant \varepsilon\}\right).$$

En prenant le complémentaire on obtient

$$0 = \mathbb{P}\left(\bigcap_{N\in\mathbb{N}} \bigcup_{n\geqslant N} \{|X_n - X| > \varepsilon\}\right).$$

Si on note $A_N = \bigcup_{n\geqslant N}\{|X_n - X| > \varepsilon\}$ alors ces évènements sont décroissants ($A_1 \supset A_2 \supset \cdots$) d'où

$$0 = \mathbb{P}\left(\bigcap_{N\in\mathbb{N}} A_N\right) = \lim_{N\to\infty} \mathbb{P}(A_N)$$

Donc, pour un N fixé on a

$$\mathbb{P}(|X_N - X| > \varepsilon) \leqslant \mathbb{P}(A_N) \underset{N\to\infty}{\longrightarrow} 0$$

d'où le résultat. ■

Exemples:

- Soit

$$X_n = \begin{cases} n \text{ avec proba } 1/n \\ 0 \text{ avec proba } 1 - 1/n \end{cases}$$

Avec le même raisonnement que précédemment on a $X_n \overset{\mathbb{P}}{\underset{n\to+\infty}{\longrightarrow}} 0$ mais $\mathbb{E}[X_n] = 1$ pour tout $n \in \mathbb{N}$ d'où $\mathbb{E}[X_n] \underset{n\to+\infty}{\not\to} 0$.

- Soit $X_n \sim b(1/2)$ et $X \sim b(1/2)$ indépendant des X_n. Est-ce que $X_n \xrightarrow[n\to+\infty]{\mathbb{P}} X$?

$$\mathbb{P}(|X_n - X| > 1/2) = \mathbb{P}(X_n = 0 \text{ et } X = 1) + \mathbb{P}(X_n = 1 \text{ et } X = 0) = 1/2$$

  donc X_n ne converge pas en proba vers X.

3 Convergence en loi

Définition 4

Une suite $(X_n)_{n\in\mathbb{N}}$ de v.a. ***converge en loi*** *vers une v.a. X si pour toute fonction continue bornée $\phi : \mathbb{R} \to \mathbb{R}$ on a*

$$\mathbb{E}[\phi(X_n)] \underset{n\to+\infty}{\longrightarrow} \mathbb{E}[\phi(X)].$$

On note alors $X_n \xrightarrow[n\to+\infty]{\mathcal{L}} X$.

Exemples:

- Soit $(X_n)_{n\in\mathbb{N}}$, une suite de v.a. de même loi et X une v.a. de même loi que les X_n. Alors $\mathbb{E}[\phi(X_n)] = \mathbb{E}[\phi(X)]$ pour toute fonction continue bornée ϕ donc $X_n \xrightarrow[n\to+\infty]{\mathcal{L}} X$.
- Soit $X_n \sim \mathcal{P}(1 - 1/n)$ et $X \sim \mathcal{P}(1)$. Soit $\phi : \mathbb{N} \to \mathbb{R}$ bornée alors,

$$\mathbb{E}[\phi(X_n)] = \sum_{k\geqslant 0} \phi(k) \frac{e^{-(1-1/n)}(1-1/n)^k}{k!}.$$

  Pour tout $k \in \mathbb{N}$ on a

$$\phi(k) \frac{e^{-(1-1/n)}(1-1/n)^k}{k!} \underset{n\to+\infty}{\longrightarrow} \phi(k) \frac{e^{-1}}{k!}$$

  et

$$\left| \phi(k) \frac{e^{-(1-1/n)}(1-1/n)^k}{k!} \right| \leqslant \|\phi\|_\infty \frac{1}{k!} \text{ qui est sommable et ne dépend pas de } n.$$

  Donc

$$\mathbb{E}[\phi(X_n)] \underset{n\to+\infty}{\longrightarrow} \sum_{k\geqslant 0} \phi(k) \frac{e^{-1}}{k!} = \mathbb{E}[\phi(X)]$$

  et alors $X_n \xrightarrow[n\to+\infty]{\mathcal{L}} X$.

Remarque: Comme la convergence en loi ne s'intéresse qu'à la convergence des lois des variables aléatoires, on peut réécrire l'exemple précédant comme $X_n \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{P}(1)$ ou même $\mathcal{P}(1 - 1/n) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{P}(1)$.

Proposition 5

Soient F_X et F_{X_n} les fonctions de répartition d'une v.a. X et des v.a. d'une suite $(X_n)_{n\in\mathbb{N}}$. On note C l'ensemble des $t \in \mathbb{R}$ tel que F_X est continue en t. Alors

$$X_n \xrightarrow[n\to+\infty]{\mathcal{L}} X \iff \forall t \in C,\ F_{X_n}(t) \underset{n\to+\infty}{\longrightarrow} F_X(t).$$

Démonstration : ADMIS
C.f. Proposition 14.17 de
Ouvrard J. (2009), *Probabilités. Tome II. Master-Agrégation [3e édition]*, Cassini. ■

Corollaire 6

Si $X_n \xrightarrow[n\to+\infty]{\mathcal{L}} X$ *alors pour tout* $a, b \in \mathbb{R} \cup \{\pm\infty\}$ *tels que* $a \leqslant b$,

$$\mathbb{P}(X_n \in [a,b]) \underset{n\to+\infty}{\longrightarrow} \mathbb{P}(X \in [a,b]).$$

Proposition 7

Soit $g : \mathbb{R} \to \mathbb{R}$ *une fonction continue. Alors*

(i) $X_n \xrightarrow[n\to+\infty]{p.s.} X \Longrightarrow g(X_n) \xrightarrow[n\to+\infty]{p.s.} g(X)$

(ii) $X_n \xrightarrow[n\to+\infty]{\mathbb{P}} X \Longrightarrow g(X_n) \xrightarrow[n\to+\infty]{\mathbb{P}} g(X)$

(iii) $X_n \xrightarrow[n\to+\infty]{\mathcal{L}} X \Longrightarrow g(X_n) \xrightarrow[n\to+\infty]{\mathcal{L}} g(X)$

Démonstration : (i) On suppose $X_n \xrightarrow[n\to+\infty]{p.s.} X$. Par continuité de g,

$$X_n \underset{n\to+\infty}{\longrightarrow} X \Rightarrow g(X_n) \underset{n\to+\infty}{\longrightarrow} g(X)$$

donc $\mathbb{P}(g(X_n) \underset{n\to+\infty}{\longrightarrow} g(X)) \geqslant \mathbb{P}(X_n \underset{n\to+\infty}{\longrightarrow} X) = 1$ d'où $g(X_n) \xrightarrow[n\to+\infty]{p.s.} g(X)$.

(ii) On suppose $X_n \xrightarrow[n\to+\infty]{\mathbb{P}} X$. Soit $\varepsilon > 0$, par continuité de g en X, $\exists \delta > 0$ tel que

$$\forall y \in \mathbb{R}, \quad |X - y| \leqslant \delta \Longrightarrow |g(X) - g(y)| \leqslant \varepsilon$$

donc, en prenant $y = X_n$ on a

$$|X - X_n| \leqslant \delta \Rightarrow |g(X) - g(X_n)| \leqslant \varepsilon$$

et alors

$$\begin{aligned} &\mathbb{P}(|g(X) - g(X_n)| \leqslant \varepsilon) \geqslant \mathbb{P}(|X - X_n| \leqslant \delta) \\ \Rightarrow\ &\mathbb{P}(|g(X) - g(X_n)| > \varepsilon) \leqslant \mathbb{P}(|X - X_n| > \delta) \underset{n\to+\infty}{\longrightarrow} 0 \end{aligned}$$

d'où $g(X_n) \xrightarrow[n\to+\infty]{\mathbb{P}} g(X)$

(iii) On suppose $X_n \xrightarrow[n\to+\infty]{\mathcal{L}} X$. Soit $\phi : \Omega \to \mathbb{R}$, une fonction continue bornée, alors $\phi \circ g : \Omega \to \mathbb{R}$ est aussi continue et bornée donc

$$\mathbb{E}[\phi(g(X_n))] \underset{n\to+\infty}{\longrightarrow} \mathbb{E}[\phi(g(X))]$$

d'où $g(X_n) \xrightarrow[n\to+\infty]{\mathcal{L}} g(X)$. ■

Proposition 8

$$X_n \xrightarrow[n\to+\infty]{\mathbb{P}} X \Longrightarrow X_n \xrightarrow[n\to+\infty]{\mathcal{L}} X.$$

Démonstration : On suppose $X_n \xrightarrow[n\to+\infty]{\mathbb{P}} X$. Soit $\phi : \mathbb{R} \to \mathbb{R}$ une fonction continue bornée. Alors, on peut écrire pour tout $\varepsilon > 0$,

$$\begin{aligned}|\mathbb{E}[\phi(X_n)] - \mathbb{E}[\phi(X)]| &\leqslant \mathbb{E}[|\phi(X_n) - \phi(X)|] \\ &= \mathbb{E}[|\phi(X_n) - \phi(X)|\mathbb{1}_{|\phi(X_n)-\phi(X)|\leqslant\varepsilon}] + \mathbb{E}[|\phi(X_n) - \phi(X)|\mathbb{1}_{|\phi(X_n)-\phi(X)|>\varepsilon}] \\ &\leqslant \varepsilon + 2\|\phi\|_\infty \mathbb{E}[\mathbb{1}_{|\phi(X_n)-\phi(X)|>\varepsilon}] \\ &= \varepsilon + 2\|\phi\|_\infty \mathbb{P}(|\phi(X_n) - \phi(X)| > \varepsilon)\end{aligned}$$

Or, $\mathbb{P}(|\phi(X_n) - \phi(X)| > \varepsilon) \xrightarrow[n\to+\infty]{} 0$ car $\phi(X_n) \xrightarrow[n\to+\infty]{\mathbb{P}} \phi(X)$ comme ϕ est continue. Donc $\lim_{n\to\infty} |\mathbb{E}[\phi(X_n)] - \mathbb{E}[\phi(X)]| \leqslant \varepsilon$ pour tout ε d'où $\lim_{n\to\infty} \mathbb{E}[\phi(X_n)] = \mathbb{E}[\phi(X)]$ ce qui prouve la convergence en loi des X_n vers X. ■

Enfin, un résultat que l'on utilisera plusieurs fois dans le cours est le suivant.

Lemme 9 (*Slutsky*)

Soient $(X_n)_{n\in\mathbb{N}}$ et $(Y_n)_{n\in\mathbb{N}}$ deux suites de v.a. vérifiant $X_n \xrightarrow[n\to+\infty]{\mathcal{L}} c$ et $Y_n \xrightarrow[n\to+\infty]{\mathbb{P}} Y$ pour une constante $c \in \mathbb{R}$ et une v.a. Y. Alors $X_nY_n \xrightarrow[n\to+\infty]{\mathcal{L}} cY$.

Démonstration : ADMIS.
C.f. Exercice 14.8 de
Ouvrard J. (2009), *Probabilités. Tome II. Master-Agrégation [3e édition]*, Cassini. ■

4 Théorèmes limites

Théorème 10 (*Loi faible des grands nombres*)

Soit $(X_n)_{n\in\mathbb{N}}$ une suite de v.a.i.i.d. telle que $\mathbb{E}[X_1^2] < \infty$. On note $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ la moyenne empirique des X_i et $\mu = \mathbb{E}[X_1]$. Alors

$$\bar{X}_n \xrightarrow[n\to+\infty]{\mathbb{P}} \mu.$$

Démonstration : On remarque que

$$\mathbb{E}[\bar{X}_n] = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^n X_i\right] = \frac{1}{n}\sum_{i=1}^n \mathbb{E}[X_i] = \mu$$

et

$$\mathrm{Var}(\bar{X}_n) = \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^n X_i\right) = \frac{1}{n^2}\sum_{i=1}^n \mathbb{E}[X_i] = \frac{\mathrm{Var}(X_1)}{n}.$$

Soit $\varepsilon > 0$, par l'inégalité de Bienaimé-Tchébytchev, on a

$$\mathbb{P}\left(|\bar{X}_n - \mu| > \varepsilon\right) \leqslant \frac{\mathrm{Var}(X_n)}{\varepsilon^2} = \frac{\mathrm{Var}(X_1)}{n\varepsilon^2} \xrightarrow[n\to+\infty]{} 0.$$

donc $\bar{X}_n \xrightarrow[n\to+\infty]{\mathbb{P}} \mu$. ■

Théorème 11 (*Loi forte des grands nombres*)

Soit $(X_n)_{n\in\mathbb{N}}$ *une suite de v.a.i.i.d. telle que* $\mathbb{E}[|X_1|] < \infty$. *Alors*

$$\bar{X}_n \xrightarrow[n\to+\infty]{p.s.} \mu.$$

Démonstration : ADMIS.
C.f. Théorème 10.14, Théorème 10.20 et Théorème 10.21 de
Ouvrard J. (2009), *Probabilités. Tome II. Master-Agrégation [3e édition]*, Cassini. ■

Théorème 12 (*Théorème central limite*)

Soit $(X_n)_{n\in\mathbb{N}}$ *une suite de v.a.i.i.d. telle que* $\mathbb{E}[X_1^2] < \infty$. *On note* $\mu = \mathbb{E}[X_1]$ *et* $\sigma^2 = \mathrm{Var}(X_1)$. *Alors*

$$\sqrt{n}\frac{\bar{X}_n - \mu}{\sqrt{\sigma^2}} \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}(0,1).$$

Démonstration : ADMIS. La preuve demande l'utilisation d'un outil pas vu en cours. ■

Remarque: Le TCL peut aussi s'écrire sous la forme

$$\sqrt{n}(\bar{X}_n - \mu) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}(0,\sigma^2).$$

En particulier, ce résultat dit que lorsque n est grand alors on peut approcher la loi de $\bar{X}_n$ par une loi $\mathcal{N}(\mu, \sigma^2/n)$.

Exemple: Vidéo Youtube d'illustration du théorème central limite avec la planche de Galton.

Introduction

Différences entre probabilité et statistiques

Que ce soit en proba ou en stat on part toujours du principe qu'un certain aléa va générer des données.

- Lorsqu'on considère l'aléa connu et que l'on étudie le comportement des données générées par cet aléa on fait des **probabilités**.
- Lorsqu'on observe des données issues d'un aléa inconnu et que l'on cherche à obtenir des informations sur cet aléa on fait des **statistiques**.

Ce point de vue est appelé la **statistique inférentielle**.
⚠ Tout ce qui touche à l'analyse de données en général est souvent appelé des **statistiques**.
Exemples:

- On considère la question suivante :

  On lance un pièce équilibrée 100 *fois de façon indépendantes. Quelle est la probabilité qu'elle tombe sur pile plus de* 75 *fois ?*

  Cette question est un problème de **probabilité** car on part d'un aléa connu (ici 100 lois $b(1/2)$ indépendantes) et on cherche à obtenir des informations sur les valeurs qu'il génère (ici le résultat des 100 lancers).

- On considère la question suivante :

  On lance un pièce 100 *fois de façon indépendantes et on tombe* 75 *fois sur pile et* 25 *fois sur face. Est-ce que la pièce est bien équilibrée ou est-ce qu'elle est biaisée ?*

  Cette question est un problème de **statistiques** car ici l'aléa n'est pas connu (on ne connaît pas la probabilité p que la pièce tombe sur pile) et on cherche à obtenir des informations sur cet aléa (on veut savoir si $p = 1/2$ ou $p \neq 1/2$) à partir d'une observation de données issues de cet aléa (ici les 100 lancers).

⚠ Comparé aux autres domaines des mathématiques, on ne peut quasiment jamais répondre avec certitude à une question de statistique. En effet, considérons le problème suivant :

On lance un dé 100 *fois de façon indépendantes et on tombe à chaque fois sur le chiffre* 6. *Est-ce que le dé est truqué ?*

La probabilité que 100 lancers de dés indépendants tombent tous le temps sur 6 est de $6^{-100} \approx 1.5 \times 10^{-78} \neq 0$. On ne peut donc pas conclure avec certitude que le dé n'est pas équilibré même si les chances que ce soit le cas sont ridiculement faible. En pratique on devra donc se fixer un niveau de confiance dans les réponses.

Chapitre II

Estimation paramétrique

1 Modélisation statistique

Cadre statistique : On suppose à partir de maintenant, et jusqu'à mention du contraire, que l'on observe des données $(x_1, \dots, x_n)$ qui sont la réalisation de n v.a.i.i.d. $X_1, \dots, X_n$. On appelle $(x_1, \dots, x_n)$ un **échantillon** de la loi commune des X_i. On suppose également que cette loi appartient à une famille de lois spécifiée par avance et dépendant d'un paramètre θ inconnu que l'on note

$$\mathcal{F} = \{\mathbb{P}_\theta;\ \theta \in \Theta\},$$

où $\mathbb{P}_\theta$ est la loi de paramètre θ et Θ est l'ensemble des valeurs possibles pour θ. Notre objectif est alors d'estimer θ à partir des données observées $(x_1, \dots, x_n)$.
Notation : On note $\mathbb{E}_\theta$ et Var_θ l'espérance et la variance sous cette loi.

Définition 13 (*Modèle statistique*)

*La famille $\mathcal{F} = \{\mathbb{P}_\theta; \theta \in \Theta\}$ est appelé un **modèle statistique**.*

- *Le modèle est dit **discret** si $\mathbb{P}_\theta$ est une loi sur un ensemble fini ou dénombrable ($\{0,1\}$, $\mathbb{N}$, $\mathbb{Z}$, ...). On note $\mathbb{P}_\theta(X = \cdot\ )$ les probabilités associées.*
- *Le modèle est dit **continu** si $\mathbb{P}_\theta$ est une loi admettant une densité sur $\mathbb{R}$ par rapport à la mesure de Lebesgue. On note f_θ la densité de $\mathbb{P}_\theta$ et F_θ sa fonction de répartition.*

Exemples:

- Modèle de Bernoulli : $\mathcal{F} = \{b(p);\ p \in [0,1]\}$.
  Ici, $\mathbb{P}_p(X = 1) = p$ et $\mathbb{P}_p(X = 0) = 1 - p$. De plus, $\mathbb{E}_p[X] = p$ et $\mathrm{Var}_p(X) = p(1-p)$.
  C'est le modèle utilisé dans la théorie des sondages. On considère que l'on cherche la proportion p de la population qui va voter pour le candidat A. On demande leur opinion à n personnes choisies indépendamment et on note $X_i = 1$ si la i-ème personne va voter pour A et $X_i = 0$ sinon. Estimer le paramètre p revient donc à estimer la proportion de la population qui va voter pour le candidat A.
- Modèle de Poisson : $\mathcal{F} = \{\mathcal{P}(\lambda);\ \lambda \in]0, +\infty[\}$.
  Ici, $\mathbb{P}_\lambda(X = i) = \frac{e^{-\lambda}\lambda^i}{i!}$ pour tout $i \in \mathbb{N}$. De plus, $\mathbb{E}_\lambda[X] = \lambda$ et $\mathrm{Var}_\lambda(X) = \lambda$.
  Modèle utilisé par exemple pour compter le nombre d'accidents de voitures dans une journée, le nombre de clients dans un magasin durant une journée, nombre de buts dans un match de foot...
- Modèle exponentiel : $\mathcal{F} = \{\mathcal{E}(\lambda); \lambda \in]0, +\infty[\}$.
  Ici, $f_\lambda(x) = \lambda e^{-\lambda x}\mathbb{1}_{x \geqslant 0}$ et $F_\lambda(x) = (1 - e^{-\lambda x})\mathbb{1}_{x \geqslant 0}$. De plus, $\mathbb{E}_\lambda[X] = 1/\lambda$ et $\mathrm{Var}_\lambda(X) = 1/\lambda^2$.
  Modélise par exemple le temps de désintégration d'un nucléide radioactif, le temps de

défaillance d'un objet mécanique ou électronique ou le temps entre deux sinistres (pour une compagnie d'assurance).

- Modèle Gaussien : $\mathcal{F} = \{\mathcal{N}(\mu, \sigma^2);\ (\mu, \sigma) \in \mathbb{R}\times]0, +\infty[\}$.
  Ici, $f_{\nu,\sigma}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\nu)^2}{2\sigma^2}}$. De plus, $\mathbb{E}_{\nu,\sigma}[X] = \nu$ et $\mathrm{Var}_{\nu,\sigma}(X) = \sigma^2$.
  Modélise par exemple la répartition de la taille ou du poids de la population d'un âge et d'un sexe donné.

Définition 14 (*Identifiabilité*)

Le modèle est dit ***identifiable*** *si l'application* $\begin{array}{ccc} \Theta & \to & \mathcal{F} \\ \theta & \mapsto & \mathbb{P}_\theta \end{array}$ *est injective (et donc bijective). Autrement dit,* $\mathbb{P}_\theta = \mathbb{P}_{\theta'} \Rightarrow \theta = \theta'$.

⚠ Sans identifiabilité, il est impossible de résoudre le problème d'estimation car deux paramètres différents peuvent correspondre à une même loi. Par la suite, on supposera tout le temps qu'on travaille avec une famille identifiable.
Exemple: $\{\mathcal{N}(\mu, \sigma^2);\ (\mu, \sigma) \in \mathbb{R}\times]0, +\infty[\}$ est identifiable mais $\{\mathcal{N}(\mu, \sigma^2); (\mu, \sigma) \in \mathbb{R} \times \mathbb{R}^*\}$ ne l'est pas vu que $\mathcal{N}(0, 1^2) = \mathcal{N}(0, (-1)^2)$.

2 Définition d'un estimateur et critères de performance

Définition 15 (*Statistique*)

Une ***statistique*** t *est une fonction mesurable des observations* $x_1, \dots, x_n$. $t(x_1, \dots, x_n)$ *est alors la réalisation d'une variable aléatoire* $T_n = t(X_1, \dots, X_n)$ *qu'on appellera aussi une* ***statistique****.*

Exemple: La **moyenne** $t(x_1, \dots, x_n) = \frac{1}{n} \sum_{i=1}^{n} x_i$ est une réalisation de la **moyenne empirique** $\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$. C'est une statistique utilisée en général pour estimer l'espérance de $\mathbb{P}_\theta$.

Définition 16 (*Estimateur*)

Soit $\Theta \subset \mathbb{R}^d$. *Un* ***estimateur*** *de* θ *est une statistique (aléatoire)* $\hat{\theta}_n$ *à valeurs dans* $\mathbb{R}^d$. *Une réalisation de l'estimateur* $\hat{\theta}_n$ *est appelée une* ***estimation****.*

⚠ Un estimateur est donc une variable aléatoire, tandis qu'une estimation est une valeur déterministe.
Exemple: On considère qu'on a $n = 10$ données : $1, 9, 2, 3, 7, 4, 3, 0, 0, 7$ que l'on modélise par un modèle de Poisson $\mathcal{F} = \{\mathcal{P}(\lambda);\ \lambda \in]0, +\infty[\}$. Un **estimateur** de λ est $\hat{\lambda}_n = \bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ et l'**estimation** correspondante est

$$\frac{1}{10}(1 + 9 + 2 + 3 + 7 + 4 + 3 + 0 + 0 + 7) = 3.6.$$

Définition 17 (*Biais*)

Soit $\hat{\theta}_n$, *un estimateur de* θ *à valeurs dans* $\mathbb{R}$.

- *On défini son* ***biais*** *par* $b(\hat{\theta}_n, \theta) = \mathbb{E}_\theta[\hat{\theta}_n] - \theta$.
- *On dit que* $\hat{\theta}_n$ *est* ***non biaisé*** *si* $b(\hat{\theta}_n, \theta) = 0$ *pour tout* $\theta \in \Theta$.
- *On dit que* $\hat{\theta}_n$ *est a****symptotiquement non biaisé*** *si* $b(\hat{\theta}_n, \theta) \underset{n\to+\infty}{\longrightarrow} 0$ *pour tout*

$\theta \in \Theta$.

Exemple: On reprends $\hat{\lambda}_n = \bar{X}_n$ pour le modèle de Poisson. Son biais est

$$b(\hat{\lambda}_n, \lambda) = \mathbb{E}_\lambda[\hat{\lambda}_n] - \lambda = \frac{1}{n}\sum_{i=1}^n \mathbb{E}_\lambda[X_i] - \lambda = \frac{1}{n}\sum_{i=1}^n \lambda - \lambda = 0.$$

$\hat{\lambda}_n$ est donc non biaisé.

Définition 18 (*Risque quadratique/ Erreur moyenne quadratique*)

*Soit $\hat{\theta}_n$ un estimateur de θ à valeurs dans $\mathbb{R}$. On définit son **risque quadratique** par*

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

*Si $\hat{\theta}_n$ et $\hat{\theta}'_n$ sont deux estimateurs de θ à valeurs dans $\mathbb{R}$, on dit que $\hat{\theta}_n$ est **préférable** à $\hat{\theta}'_n$ lorsque*

$$\forall \theta \in \Theta,\ R(\hat{\theta}_n, \theta) \leqslant R(\hat{\theta}'_n, \theta).$$

Proposition 19

$$R(\hat{\theta}_n, \theta) = \mathrm{Var}_\theta(\hat{\theta}_n) + b(\hat{\theta}_n, \theta)^2.$$

Démonstration :

$$\begin{aligned}
R(\hat{\theta}_n, \theta) &= \mathbb{E}_\theta\left[(\hat{\theta}_n - \mathbb{E}_\theta[\hat{\theta}_n] + \mathbb{E}_\theta[\hat{\theta}_n] - \theta)^2\right] \\
&= \mathbb{E}_\theta\left[(\hat{\theta}_n - \mathbb{E}_\theta[\hat{\theta}_n])^2\right] + \mathbb{E}_\theta\left[(\mathbb{E}_\theta[\hat{\theta}_n] - \theta)^2\right] + 2\mathbb{E}_\theta\left[(\hat{\theta}_n - \mathbb{E}_\theta[\hat{\theta}_n])(\mathbb{E}_\theta[\hat{\theta}_n] - \theta)\right] \\
&= \mathbb{E}_\theta\left[(\hat{\theta}_n - \mathbb{E}_\theta[\hat{\theta}_n])^2\right] + (\mathbb{E}_\theta[\hat{\theta}_n] - \theta)^2 + 2\mathbb{E}_\theta[\hat{\theta}_n - \mathbb{E}_\theta[\hat{\theta}_n]](\mathbb{E}_\theta[\hat{\theta}_n] - \theta) \\
&= \mathrm{Var}_\theta(\hat{\theta}_n) + b(\hat{\theta}_n, \theta)^2 + 2(\mathbb{E}_\theta[\hat{\theta}_n] - \mathbb{E}_\theta[\hat{\theta}_n])(\mathbb{E}_\theta[\hat{\theta}_n] - \theta) \\
&= \mathrm{Var}_\theta(\hat{\theta}_n) + b(\hat{\theta}_n, \theta)^2.
\end{aligned}$$

■

Exemple: On reprends $\hat{\lambda}_n = \bar{X}_n$ pour le modèle de Poisson. Son biais est nul donc

$$R(\hat{\lambda}_n, \lambda) = \mathrm{Var}_\lambda(\hat{\lambda}_n) = \mathrm{Var}_\lambda\left(\frac{1}{n}\sum_{i=1}^n X_i\right) = \frac{1}{n^2}\sum_{i=1}^n \mathrm{Var}_\lambda(X_i) = \frac{1}{n^2}\sum_{i=1}^n \lambda = \frac{\lambda}{n},$$

où on a utilisé que la variance d'une somme de variables **indépendantes** est égale à la somme de leurs variances.

Définition 20 (*Consistance*)

*Un estimateur $\hat{\theta}_n$ est dit **consistant** si $\hat{\theta}_n \xrightarrow[n\to+\infty]{\mathbb{P}} \theta$ pour tout $\theta \in \Theta$. Il est dit **fortement consistant** si $\hat{\theta}_n \xrightarrow[n\to+\infty]{p.s.} \theta$ pour tout $\theta \in \Theta$.*

Exemple: On reprends $\hat{\lambda}_n = \bar{X}_n$ pour le modèle de Poisson. Les variables X_i sont i.i.d. de loi $\mathcal{P}(\lambda)$ et vérifient $\mathbb{E}_\lambda[|X_1|] < +\infty$ donc, par la loi forte des grands nombres,

$$\hat{\lambda}_n = \frac{1}{n}\sum_{i=1}^n X_i \xrightarrow[n\to+\infty]{p.s.} \mathbb{E}_\lambda[X_1] = \lambda.$$

$\hat{\lambda}_n$ est donc fortement consistant.

Proposition 21

Soit $\hat{\theta}_n$, un estimateur de θ à valeurs dans $\mathbb{R}$. Si $R(\hat{\theta}_n, \theta) \underset{n\to+\infty}{\longrightarrow} 0$ pour tout $\theta \in \Theta$ alors $\hat{\theta}_n$ est consistant.

Démonstration : Soit $\varepsilon > 0$, alors

$$\begin{aligned}\mathbb{P}_\theta(|\hat{\theta}_n - \theta| \geqslant \varepsilon) &= \mathbb{P}_\theta(|\hat{\theta}_n - \theta|^2 \geqslant \varepsilon^2)\\ &\leqslant \frac{\mathbb{E}_\theta(|\hat{\theta}_n - \theta|^2)}{\varepsilon^2}\\ &= \frac{R(\hat{\theta}_n, \theta)}{\varepsilon^2} \underset{n\to+\infty}{\longrightarrow} 0\end{aligned}$$

■

Proposition 22

Soient Θ, Θ' deux espaces de paramètre et $\hat{\theta}_n$ un estimateur consistant (resp. fortement consistant) de θ. Soit $\phi : \Theta \to \Theta'$ une fonction continue, alors $\phi(\theta_n)$ est un estimateur consistant (resp. fortement consistant) de $\phi(\theta)$.

Démonstration : Conséquence directe de la proposition 7 sur la stabilité de la convergence en proba et presque sûre par composition avec une fonction continue. ■

⚠ $R(\hat{\theta}_n, \theta)$ et $b(\hat{\theta}_n, \theta)$ sont en général différent de $R(\phi(\hat{\theta}_n), \phi(\theta))$ et $b(\phi(\hat{\theta}_n), \phi(\theta))$. Du coup, ce n'est pas parce que θ_n est un "bon" estimateur de θ que $\phi(\theta_n)$ sera aussi un "bon" estimateur de $\phi(\theta)$.

3 Propriétés de la moyenne empirique

Proposition 23

*On suppose que $\theta = \mathbb{E}_\theta[X_1]$ et on estime θ par $\hat{\theta}_n = \bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ appelé la **moyenne empirique**. Alors :*

- *$\hat{\theta}_n$ est un estimateur sans biais de θ.*
- *Si $\forall \theta \in \Theta$, $\mathbb{E}_\theta[|X_1|] < +\infty$, alors $\hat{\theta}_n$ est fortement consistant.*
- *Si $\forall \theta \in \Theta$, $\mathbb{E}_\theta[X_1^2] < +\infty$, alors $\hat{\theta}_n$ a pour risque quadratique*

$$\forall \theta \in \Theta,\ R(\hat{\theta}_n, \theta) = \frac{\mathrm{Var}_\theta(X_1)}{n}.$$

Démonstration : Voir les exemples de la partie précédente. ■

Définition 24

*On suppose qu'il existe une fonction $g : \mathbb{R} \to \Theta$ telle que $\theta = g(\mathbb{E}_\theta[X_1])$. On appelle estimateur par la **méthode des moments d'ordre 1** (EMM) l'estimateur $\hat{\theta}_n = g(\bar{X}_n)$.*

Exemples:

- Modèle statistique $\{\mathcal{U}([0,\theta]), \theta \in]0, +\infty[\}$:
  Ici, $\mathbb{E}_\theta[X_1] = \frac{\theta}{2}$ donc $\theta = 2\mathbb{E}_\theta[X_1]$. L'EMM est donc $\hat{\theta}_n = 2\bar{X}_n$.
- Modèle statistique $\{\mathcal{E}(\theta), \theta \in]0, +\infty[\}$:
  Ici, $\mathbb{E}_\theta[X_1] = \frac{1}{\theta}$ donc $\theta = \frac{1}{\mathbb{E}_\theta[X_1]}$. L'EMM est donc $\hat{\theta}_n = \frac{1}{\bar{X}_n}$.

Proposition 25

Si $\mathbb{E}_\theta[|X_1|] < +\infty$ pour tout $\theta \in \Theta$ et g est une fonction continue sur $\{\mathbb{E}_\theta[X_1],\ \theta \in \Theta\}$ alors l'EMM $\hat{\theta}_n = g(\bar{X}_n)$ est un estimateur fortement consistant de θ.

Démonstration : Comme les X_i sont des variables aléatoires i.i.d. vérifiant $\mathbb{E}_\theta[|X_1|] < +\infty$ alors par la loi forte des grands nombres on a

$$\bar{X}_n \xrightarrow[n\to+\infty]{p.s.} \mathbb{E}_\theta[X_1]$$

et par continuité de g on a

$$\hat{\theta}_n = g(\bar{X}_n) \xrightarrow[n\to+\infty]{p.s.} g(\mathbb{E}_\theta[X_1]) = \theta.$$

■

4 Propriétés de la variance empirique

Définition 26

On appelle ***variance empirique*** *des X_i, notée S_n^2, la quantité*

$$S_n^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X}_n)^2 = \frac{1}{n}\sum_{i=1}^n X_i^2 - (\bar{X}_n)^2$$

Proposition 27

On suppose que $\theta = \mathrm{Var}_\theta(X_1)$ alors l'estimateur $\hat{\theta}_n = S_n^2$ est fortement consistant.

Démonstration : On a $\mathbb{E}_\theta[|X_1|] \leqslant \mathbb{E}_\theta[X_1^2] < +\infty$ car $\mathrm{Var}_\theta(X_1) = \theta < +\infty$. Comme les X_i sont iid alors on peut appliquer la loi forte des grands nombre aux X_i et X_i^2 ce qui donne

$$\bar{X}_n \xrightarrow[n\to+\infty]{p.s.} \mathbb{E}_\theta[X_1] \text{ et } \frac{1}{n}\sum_{i=1}^n X_i^2 \xrightarrow[n\to+\infty]{p.s.} \mathbb{E}_\theta[X_1^2].$$

En conséquence,

$$S_n^2 = \frac{1}{n}\sum_{i=1}^n X_i^2 - (\bar{X}_n)^2 \xrightarrow[n\to+\infty]{p.s.} \mathbb{E}_\theta[X_1]^2 - \mathbb{E}_\theta[X_1]^2 = \mathrm{Var}_\theta(X_1) = \theta$$

d'où S_n^2 est fortement consistant. ■

Proposition 28

On suppose que $\theta = \mathrm{Var}_\theta(X_1)$. L'estimateur $\hat{\theta}_n = S_n^2$ est biaisé car

$$\mathbb{E}_\theta[S_n^2] = \frac{n-1}{n}\theta.$$

Pour estimer θ sans biais on utilise

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

appelé la ***variance corrigée****.*

Démonstration : Tout d'abord, on a $\mathbb{E}_\theta\left[\frac{1}{n}\sum_{i=1}^n X_i^2\right] = \frac{1}{n}\sum_{i=1}^n \mathbb{E}_\theta[X_i^2] = \mathbb{E}_\theta[X_1^2]$ et

$$\begin{aligned}
\mathbb{E}_\theta\left[(\bar{X}_n)^2\right] &= \mathbb{E}_\theta\left[\left(\frac{1}{n}\sum_{i=1}^n X_i\right)^2\right] \\
&= \mathbb{E}_\theta\left[\frac{1}{n^2}\sum_{i,j=1}^n X_iX_j\right] \\
&= \mathbb{E}_\theta\left[\frac{1}{n^2}\left(\sum_{i=1}^n X_i^2 + \sum_{\substack{i,j=1 \\ i\neq j}}^n X_iX_j\right)\right] \\
&= \frac{1}{n^2}(n\mathbb{E}_\theta[X_1^2] + (n^2-n)\mathbb{E}_\theta[X_1]^2) \\
&= \frac{1}{n}\mathbb{E}_\theta[X_1^2] + \frac{n-1}{n}\mathbb{E}_\theta[X_1]^2
\end{aligned}$$

d'où

$$\mathbb{E}_\theta[S_n^2] = \mathbb{E}_\theta\left[\frac{1}{n}\sum_{i=1}^n X_i^2\right] - \mathbb{E}_\theta[(\bar{X}_n)^2] = \frac{n-1}{n}\left(\mathbb{E}_\theta[X_1^2] - \mathbb{E}_\theta[X_1]^2\right) = \frac{n-1}{n}\mathrm{Var}_\theta(X_1) = \frac{n-1}{n}\theta$$

et donc $\mathbb{E}_\theta[\widetilde{S}_n^2] = \frac{n}{n-1}\mathbb{E}_\theta[S_n^2] = \theta$. ■

Remarque: S'il existe une fonction $g : \mathbb{R} \times \mathbb{R}_+ \to \Theta$ telle que $\theta = g(\mathbb{E}_\theta[X_1], \mathrm{Var}_\theta(X_1))$ alors on appelle estimateur par la **méthode des moments d'ordre 2** l'estimateur $\hat{\theta}_n = g(\bar{X}_n, S_n^2)$. Ce principe peut se généraliser lorsqu'on peut exprimer θ comme n'importe quelle fonction des moments de la loi $\mathbb{P}_\theta$.

5 Méthode du maximum de vraisemblance

Exemple: On considère le modèle $\{b(p), p \in \{1/6, 5/6\}\}$ et on observe l'échantillon $x_1 = 0, x_2 = 0, x_3 = 0, x_4 = 0, x_5 = 1, x_6 = 0$ de cette loi. Afin d'estimer p regardons pour chaque valeur possible de p quelle était la probabilité d'obtenir l'échantillon obtenu.
Si $p = 1/6$ alors

$$\mathbb{P}(X_1 = x_1, X_2 = x_2, \cdots, X_6 = x_6) = \mathbb{P}(X_1 = x_1)\mathbb{P}(X_2 = x_2)\cdots\mathbb{P}(X_6 = x_6) = \left(\frac{5}{6}\right)^5 \times \frac{1}{6} = \frac{5^5}{6^6}.$$

Si $p = 5/6$ alors

$$\mathbb{P}(X_1 = x_1, X_2 = x_2, \cdots, X_6 = x_6) = \mathbb{P}(X_1 = x_1)\mathbb{P}(X_2 = x_2)\cdots\mathbb{P}(X_6 = x_6) = \left(\frac{1}{6}\right)^5 \times \frac{5}{6} = \frac{5}{6^6}.$$

Vu qu'il y avait une probabilité plus élevé que notre échantillon soit généré si $p = 1/6$ alors on a envie d'estimer p par cette quantité. C'est l'idée derrière le principe du **maximum de vraisemblance**.

Définition 29 (*Fonction de vraisemblance*)

On appelle ***fonction de vraisemblance*** *l'application $L : \theta \in \Theta \mapsto \mathcal{L}(\theta|x_1, \ldots, x_n)$ définie par $L(\theta|x_1, \ldots, x_n) = \prod_{i=1}^n f_\theta(x_i)$ où :*

- *f_θ est la densité de $\mathbb{P}_\theta$ si le modèle est continu ;*
- *$f_\theta(x) = \mathbb{P}_\theta(X = x)$ si le modèle est discret.*

Remarques:

- Si les X_i ont une loi discrete alors

$$L(\theta|x_1, \cdots, x_n) = \mathbb{P}_\theta(X_1 = x_1, \cdots, X_n = x_n).$$

- Si la loi des X_i est continue alors $L(\theta|x_1, \cdots, x_n)$ est la valeur de la densité de probabilité de la variable $(X_1, \cdots, X_n)$ prise en $(x_1, \cdots, x_n)$.

Exemples:

- Modèle de Poisson : $\mathcal{F} = \{\mathcal{P}(\theta), \theta \in \mathbb{R}_+^*\}$.

$$L(\theta|x_1, \dots, x_n) = \prod_{i=1}^n \mathbb{P}_\theta(X = x_i) = \prod_{i=1}^n e^{-\theta} \frac{\theta^{x_i}}{x_i!} = e^{-n\theta} \frac{\theta^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n x_i!}.$$

- Modèle exponentiel : $\mathcal{F} = \{\mathcal{E}(\theta), \theta \in \mathbb{R}_+^*\}$.

$$L(\theta|x_1, \dots, x_n) = \prod_{i=1}^n f_\theta(x_i) = \prod_{i=1}^n \theta e^{-\theta x_i} \mathbb{1}_{x_i > 0} = \theta^n e^{-\theta \sum_{i=1}^n x_i} \prod_{i=1}^n \mathbb{1}_{x_i > 0}.$$

Définition 30 (*Estimateur du maximum de vraisemblance*)

Toute statistique $\hat{\theta}_n$ à valeurs dans Θ telle que

$$L(\hat{\theta}_n|X_1, \dots, X_n) = \max_{\theta \in \Theta} L(\theta|X_1, \dots, X_n)$$

est appelée ***estimateur du maximum de vraisemblance*** *(EMV).*

Remarques:

- L'EMV peut ne pas exister ou ne pas être unique.
- Si la fonction $\mathcal{L} = \log L$ est bien définie, on l'appelle la **log-vraisemblance**. Elle est souvent plus facile à manipuler et possède les mêmes maxima que L.
- l'EMV est souvent préférable à l'EMM.
- La vraisemblance s'appelle "likelihood" en anglais d'où la notation avec un L.

Exemples:

- Pour le modèle exponentiel, on a

$$L(\theta|X_1, \dots, X_n) = \theta^n e^{-\theta \sum_{i=1}^n X_i}$$

car $X_i \geqslant 0$ presque sûrement quelque soit la loi $\mathbb{P}_\theta$. La log vraisemblance est donc

$$\mathcal{L}(\theta|X_1, \dots, X_n) = \log(\theta^n e^{-\theta \sum_{i=1}^n X_i}) = n \log(\theta) - \theta \sum_{i=1}^n X_i.$$

Sa dérivée est

$$\frac{\mathrm{d}\mathcal{L}(\theta|X_1, \dots, X_n)}{\mathrm{d}\theta} = \frac{n}{\theta} - \sum_{i=1}^n X_i,$$

qui s'annule pour $\theta = \frac{n}{\sum_{i=1}^n X_i} = \frac{1}{\bar{X}_n}$. Sa dérive seconde est

$$\frac{\mathrm{d}^2\mathcal{L}(\theta|X_1, \dots, X_n)}{\mathrm{d}\theta^2} = -\frac{n}{\theta^2} < 0.$$

En conclusion, $\mathcal{L}(\theta|X_1, \dots, X_n)$ (et donc $L(\theta|X_1, \dots, X_n)$) est strictement concave et possède donc un unique maximum pour $\theta = \frac{1}{\bar{X}_n}$ d'où $\hat{\theta}_n = \frac{1}{\bar{X}_n}$.

- Pour le modèle de Poisson, on a

$$L(\theta|X_1,\dots,X_n) = \prod_{i=1}^n e^{-\theta}\frac{\theta^{X_i}}{X_i!} = e^{-n\theta}\frac{\theta^{\sum_{i=1}^n X_i}}{\prod_{i=1}^n X_i!}$$

donc la log-vraisemblance s'écrit

$$\mathcal{L}(\theta|X_1,\dots,X_n) = -n\theta + \log(\theta)\sum_{i=1}^n X_i - \log\left(\prod_{i=1}^n X_i!\right).$$

Sa dérivée est

$$\frac{\mathrm{d}\mathcal{L}(\theta|X_1,\dots,X_n)}{\mathrm{d}\theta} = -n + \frac{\sum_{i=1}^n X_i}{\theta},$$

qui s'annule pour $\theta = \frac{\sum_{i=1}^n X_i}{n} = \bar{X}_n$. Sa dérive seconde est

$$\frac{\mathrm{d}^2\mathcal{L}(\theta|X_1,\dots,X_n)}{\mathrm{d}\theta^2} = -\frac{\sum_{i=1}^n X_i}{\theta^2} < 0.$$

En conclusion, $\mathcal{L}(\theta|X_1,\dots,X_n)$ est strictement concave et possède donc un unique maximum pour $\theta = \bar{X}_n$ d'où $\hat{\theta}_n = \bar{X}_n$.

6 Théorèmes limites pour l'EMM et l'EMV

Cadre : On suppose dans cette partie que $\Theta \subset \mathbb{R}$ et on s'intéresse à la loi limite de $\hat{\theta}_n - \theta$ lorsque $\hat{\theta}_n$ est un EMM ou un EMV.

Définition 31 (*Normalité asymptotique*)

Soit $\hat{\theta}_n$ un estimateur de θ. Si $\hat{\theta}_n$ satisfait

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}(0, v(\theta)),$$

*pour une fonction $v : \Theta \to \mathbb{R}_+^*$ alors on dit que $\hat{\theta}_n$ est **asymptotiquement normal** de variance asymptotique $v(\theta)$.*

Remarques:

- Si $\hat{\theta}_n$ est asymptotiquement normal alors, quand n est assez grand on peut approximer sa loi par une loi $\mathcal{N}(\theta, \frac{v(\theta)}{n})$.
- Si $\hat{\theta}_n$ est asymptotiquement normal alors en général

$$R(\hat{\theta}_n, \theta) \underset{n\to\infty}{\sim} \frac{v(\theta)}{n}.$$

Donc, si $\hat{\theta}'_n$ est un autre estimateur asymptotiquement normal dont la variance asymptotique $v'(\theta)$ vérifie

$$\forall\theta\in\Theta,\ v'(\theta) \leqslant v(\theta),$$

alors $\hat{\theta}'_n$ est en général asymptotiquement préférable à $\hat{\theta}_n$.

Théorème 32 (*Méthode Delta*)

Soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires réelles i.i.d. telles que $\mathbb{E}[X_1^2] < \infty$. On pose $\mu = \mathbb{E}[X_1]$ et $\sigma^2 \mathrm{Var}(X_1)$. Soit $g : \mathbb{R} \to \mathbb{R}$ une fonction dérivable en μ. Alors

$$\sqrt{n}(g(\bar{X}_n) - g(\mu)) \underset{n\to+\infty}{\overset{\mathcal{L}}{\longrightarrow}} \mathcal{N}\Big(0, g'(\mu)^2\sigma^2\Big).$$

Démonstration : Comme les X_i sont i.i.d. et $\mathbb{E}[|X_1|] \leqslant \mathbb{E}[X_1^2] < +\infty$ alors, par la loi forte des grands nombres, on a $\bar{X}_n \underset{n\to+\infty}{\overset{p.s.}{\longrightarrow}} \mu$ donc $\mathbb{P}(\bar{X}_n \underset{n\to+\infty}{\longrightarrow} \mu) = 1$. Comme g est dérivable en μ on en déduit alors que

$$\mathbb{P}\left(\frac{g(\bar{X}_n) - g(\mu)}{\bar{X}_n - \mu} \underset{n\to+\infty}{\longrightarrow} g'(\mu)\right) = 1$$

donc

$$\frac{g(\bar{X}_n) - g(\mu)}{\bar{X}_n - \mu} \underset{n\to+\infty}{\overset{p.s.}{\longrightarrow}} g'(\mu) \;\Rightarrow\; \frac{g(\bar{X}_n) - g(\mu)}{\bar{X}_n - \mu} \underset{n\to+\infty}{\overset{\mathbb{P}}{\longrightarrow}} g'(\mu).$$

De plus, par le TCL on a

$$\sqrt{n}(\bar{X}_n - \mu) \underset{n\to+\infty}{\overset{\mathcal{L}}{\longrightarrow}} \mathcal{N}(0, \sigma^2).$$

Au final, en appliquant le Lemme de Slutsky aux deux identités précédentes, on obtient

$$\frac{g(\bar{X}_n) - g(\mu)}{\bar{X}_n - \mu} \times \sqrt{n}(\bar{X}_n - \mu) \underset{n\to+\infty}{\overset{\mathcal{L}}{\longrightarrow}} g'(\mu) \times \mathcal{N}(0, \sigma^2) \;\Leftrightarrow\; \sqrt{n}(g(\bar{X}_n) - g(\mu)) \underset{n\to+\infty}{\overset{\mathcal{L}}{\longrightarrow}} \mathcal{N}(0, \sigma^2 g'(\mu)^2).$$ ■

Corollaire 33

On suppose que $\mathbb{E}_\theta[X_1^2] < \infty$ et qu'il existe une fonction $g : \mathbb{R} \to \mathbb{R}$ telle que $\theta = g(\mathbb{E}_\theta[X_1])$ pour tout $\theta \in \Theta$. On note $\hat{\theta}_n = g(\bar{X}_n)$ l'EMM.
Si g est dérivable en $\mathbb{E}_\theta[X_1]$ pour tout $\theta \in \Theta$ alors $\hat{\theta}_n$ est asymptotiquement normal de variance asymptotique

$$v(\theta) = g'(\mathbb{E}_\theta[X_1])^2 \mathrm{Var}_\theta(X_1)$$

Démonstration : Comme toutes les hypothèses de la méthode delta sont satisfaites pour chaque loi $\mathbb{P}_\theta$, on en déduit

$$\sqrt{n}(g(\bar{X}_n) - g(\mathbb{E}_\theta[X_1])) \underset{n\to+\infty}{\overset{\mathcal{L}}{\longrightarrow}} \mathcal{N}(0, g'(\mathbb{E}_\theta[X_1])^2 \mathrm{Var}_\theta(X_1)).$$

On peut donc conclure car $\hat{\theta}_n = g(\bar{X}_n)$ et $\theta = g(\mathbb{E}_\theta[X_1])$. ■

Exemple: Modèle exponentiel $\{\mathcal{E}(\theta), \theta \in \mathbb{R}_+^*\}$:
On a vu que l'EMM est $\hat{\theta}_n = \frac{1}{\bar{X}_n}$. Donc, si on applique la méthode delta pour la fonction $g(x) = \frac{1}{x}$ qui est bien dérivable en $1/\theta > 0$ pour tout $\theta \in \mathbb{R}_+^*$ avec $g'(1/\theta) = -\theta^2$, on obtient

$$\sqrt{n}\left(\hat{\theta}_n - \theta\right) \underset{n\to+\infty}{\overset{\mathcal{L}}{\longrightarrow}} \mathcal{N}\left(0, \theta^4 \times \frac{1}{\theta^2}\right).$$

Théorème 34

Sous certaines conditions de régularité sur la famille $\{\mathbb{P}_\theta, \theta \in \Theta\}$, l'EMV $\hat{\theta}_n$ existe, est unique et fortement consistant. De plus,

$$\forall \theta \in \Theta,\ \sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}\left(0, \frac{1}{I(\theta)}\right)$$

*où $I(\theta)$ est appelé l'**information de Fisher** et est défini par*

$$I(\theta) = -\mathbb{E}_\theta\left[\frac{\partial^2 \log(L(\theta|X))}{\partial\theta^2}\right].$$

Démonstration : ADMIS (c.f. compléments) ■

<u>*Exemple*</u>: On a vu que pour le modèle exponentiel $\{\mathcal{E}(\theta), \theta \in \mathbb{R}_+^*\}$ on a

$$L(\theta|X) = \theta e^{-\theta X} \implies L(\theta|X) = \log(\theta) - \theta X$$

donc

$$\frac{\partial L(\theta|X)}{\partial\theta} = \frac{1}{\theta} - X \ \text{ et } \ \frac{\partial^2 L(\theta|X)}{\partial\theta^2} = -\frac{1}{\theta^2}.$$

L'information de Fisher du modèle est donc

$$I(\theta) = -\mathbb{E}_\theta\left[-\frac{1}{\theta^2}\right] = \frac{1}{\theta^2}.$$

On retrouve donc le résultat

$$\sqrt{n}\left(\hat{\theta}_n - \theta\right) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}\left(0, \theta^2\right).$$

Chapitre III

Propriétés des lois normales

On rappelle qu'une variable aléatoire X suit une loi $\mathcal{N}(\mu, \sigma^2)$ pour $\mu \in \mathbb{R}$ et $\sigma \in]0, +\infty[$ si c'est une variable réelle de densité

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

1 Stabilité par linéarité

Proposition 35

Soit $X \sim \mathcal{N}(\mu, \sigma^2)$ et $c \in \mathbb{R}$, alors

$$X + c \sim \mathcal{N}(\mu + c, \sigma^2) \quad \text{et} \quad cX \sim \mathcal{N}(c\mu, c^2\sigma^2).$$

Démonstration : Soit $Y = X + c$ et $\phi : \mathbb{R} \to \mathbb{R}$ une fonction continue bornée. Alors

$$\mathbb{E}[\phi(Y)] = \mathbb{E}[\phi(X + c)] = \int_{\mathbb{R}} \phi(x + c) \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \mathrm{d}x$$

On fait le changement de variable $y = x + c$ qui est une bijection croissante de $\mathbb{R}$ dans $\mathbb{R}$ avec $\mathrm{d}y = \mathrm{d}x$ ce qui donne

$$\mathbb{E}[\phi(Y)] = \int_{\mathbb{R}} \phi(y) \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-c-\mu)^2}{2\sigma^2}} \mathrm{d}y$$

donc Y est une variable aléatoire de densité

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

donc $Y \sim \mathcal{N}(c + \mu, \sigma^2)$.
Soit $Y = cX$ et $\phi : \mathbb{R} \to \mathbb{R}$ une fonction continue bornée. Alors

$$\mathbb{E}[\phi(Y)] = \mathbb{E}[\phi(cX)] = \int_{-\infty}^{+\infty} \phi(cx) \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \mathrm{d}x$$

On fait le changement de variable $y = cx$ avec $\mathrm{d}y = c\mathrm{d}x$. Lorsque c est positif c'est une bijection croissante d'où

$$\mathbb{E}[\phi(Y)] = \int_{-\infty}^{+\infty} \phi(y) \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(\frac{y}{c}-\mu)^2}{2\sigma^2}} \frac{\mathrm{d}y}{c} = \int_{-\infty}^{+\infty} \phi(y) \frac{1}{\sqrt{2\pi}\sigma c} e^{-\frac{(y-c\mu)^2}{2c^2\sigma^2}} \mathrm{d}y.$$

Lorsque c est négatif c'est une bijection décroissante d'où

$$\mathbb{E}[\phi(Y)] = \int_{+\infty}^{-\infty} \phi(y) \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(\frac{y}{c}-\mu)^2}{2\sigma^2}} \frac{\mathrm{d}y}{c} = \int_{-\infty}^{+\infty} \phi(y) \frac{1}{\sqrt{2\pi}\sigma(-c)} e^{-\frac{(y-c\mu)^2}{2c^2\sigma^2}} \mathrm{d}y$$

donc Y est une variable aléatoire de densité

$$f(y) = \frac{1}{\sqrt{2\pi}\sigma|c|} e^{-\frac{(y-c\mu)^2}{2c^2\sigma^2}}$$

d'où $Y \sim \mathcal{N}(c\mu, c^2\sigma^2)$. ■

Corollaire 36

Si $X \sim \mathcal{N}(\mu, \sigma^2)$ alors

$$\frac{X-\mu}{\sigma} \sim \mathcal{N}(0,1).$$

La loi $\mathcal{N}(0,1)$ est appelée la ***loi normale centrée réduite****.*

Proposition 37

Soient $\mu, \mu' \in \mathbb{R}$, $\sigma, \sigma' \in]0, +\infty[$, $X \sim \mathcal{N}(\mu, \sigma^2)$ et $Y \sim \mathcal{N}(\mu', \sigma'^2)$. Si X et Y sont indépendant alors

$$X + Y \sim \mathcal{N}(\mu + \mu', \sigma^2 + \sigma'^2)$$

Démonstration : On fait la preuve dans le cas $\mu = \mu' = 0$ et $\sigma = \sigma' = 1$. Dans ce cas, X et Y ont pour densité $f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$. Comme $X \perp\!\!\!\perp Y$ alors $X + Y$ a pour densité

$$\begin{aligned}
f * f(x) &= \int_{\mathbb{R}} f(y) f(x-y) \mathrm{d}y \\
&= \frac{1}{2\pi} \int_{\mathbb{R}} e^{-\frac{y^2}{2}} e^{-\frac{(x-y)^2}{2}} \mathrm{d}y \\
&= \frac{1}{2\pi} \int_{\mathbb{R}} e^{-\frac{x^2+2y^2-2xy}{2}} \mathrm{d}y \\
&= \frac{e^{-\frac{x^2}{2}}}{2\pi} \int_{\mathbb{R}} e^{-(y^2-xy)} \mathrm{d}y \\
&= \frac{e^{-\frac{x^2}{2}}}{2\pi} \int_{\mathbb{R}} e^{-\left(y-\frac{x}{2}\right)^2 - \frac{x^2}{4}} \mathrm{d}y \\
&= \frac{e^{-\frac{x^2}{2}} e^{\frac{x^2}{4}}}{2\pi} \int_{\mathbb{R}} e^{-z^2} \mathrm{d}z \text{ avec le CDV } z = y - x/2 \\
&= \frac{e^{-\frac{x^2}{4}}}{2\sqrt{\pi}}
\end{aligned}$$

ce qui correspond bien à la densité de la loi $\mathcal{N}(0,2)$. ■

Corollaire 38

Toute combinaison linéaire de variables aléatoires normales indépendantes suit une loi normale.

Proposition 39

Soient $X_1, \cdots, X_n$ des v.a.i.i.d. de loi $\mathcal{N}(\mu, \sigma^2)$. Alors,

$$\bar{X}_n \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right) \quad \text{et donc} \quad \sqrt{n}\frac{\bar{X}_n - \mu}{\sigma} \sim \mathcal{N}(0,1).$$

Démonstration : $\bar{X}_n$ est une combinaison linéaire de variables indépendantes de loi normale donc $\bar{X}_n$ suit une loi normale. De plus, on a déjà montré que $\mathbb{E}[\bar{X}_n] = \mathbb{E}[X_1] = \mu$ et $\mathrm{Var}(\bar{X}_n) = \frac{\mathrm{Var}(X_1)}{n}$ d'où

$$\bar{X}_n \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right).$$

■

2 Estimation par maximum de vraisemblance

On considère le modèle statistique $\mathcal{F} = \{\mathcal{N}(\mu, \sigma^2),\ (\mu, \sigma) \in \mathbb{R}\times]0, +\infty[\}$.
La vraisemblance du modèle est

$$L(\mu, \sigma^2|X_1, \cdots, X_n) = \prod_{i=1}^{n} f_{\mu,\sigma}(X_i) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(X_i-\mu)^2}{2\sigma^2}} = (2\pi\sigma^2)^{-n/2} e^{-\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{2\sigma^2}}.$$

Sa log vraisemblance est alors

$$\begin{aligned}
\mathcal{L}(\mu, \sigma|X_1, \cdots, X_n) &= \log\left((2\pi\sigma^2)^{-n/2} e^{-\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{2\sigma^2}}\right) \\
&= -\frac{n}{2}\log(2\pi\sigma^2) - \frac{\sum_{i=1}^{n}(X_i-\mu)^2}{2\sigma^2} \\
&= -\frac{n}{2}\log(2\pi) - n\log(\sigma) - \frac{\sum_{i=1}^{n}(X_i-\mu)^2}{2\sigma^2}.
\end{aligned}$$

On commence d'abord à optimiser L par rapport à μ pour un σ fixé. L est bien deux fois dérivable sur $\mathbb{R}$ par rapport à μ de dérivées :

$$\frac{\partial \mathcal{L}}{\partial \mu} = \frac{\sum_{i=1}^{n}(X_i-\mu)}{\sigma^2} \quad \text{et} \quad \frac{\partial^2 \mathcal{L}}{\partial^2 \mu} = \frac{\sum_{i=1}^{n} -1}{\sigma^2} = -\frac{n}{\sigma^2} < 0.$$

Comme la dérivée seconde est tout le temps strictement négative alors L est strictement convexe par rapport à μ et admet donc un unique maximum au point $\hat{\mu}_n$ qui annule la dérivée :

$$\frac{\sum_{i=1}^{n}(X_i-\hat{\mu}_n)}{\sigma^2} = 0 \ \Leftrightarrow\ \sum_{i=1}^{n} X_i - n\hat{\mu}_n = 0 \ \Leftrightarrow\ \sum_{i=1}^{n} X_i - n\hat{\mu}_n = 0 \ \Leftrightarrow\ \hat{\mu}_n = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}_n.$$

On remarque alors que quel que soit la valeur de σ, L est maximisé pour $\mu = \bar{X}_n$ qui ne dépend pas de σ. L est donc maximisé en la valeur de σ qui maximise

$$\mathcal{L}(\bar{X}_n, \sigma|X_1, \cdots, X_n) = -\frac{n}{2}\log(2\pi) - n\log(\sigma) - \frac{\sum_{i=1}^{n}(X_i-\bar{X}_n)^2}{2\sigma^2}.$$

Cette fonction est deux fois dérivable par rapport à σ sur $]0, +\infty[$ de dérivées :

$$\frac{\partial \mathcal{L}}{\partial \sigma} = \frac{-n}{\sigma} + \frac{\sum_{i=1}^{n}(X_i-\bar{X}_n)^2}{\sigma^3} \text{ et } \frac{\partial^2 \mathcal{L}}{\partial \sigma^2} = \frac{n}{\sigma^2} - \frac{3\sum_{i=1}^{n}(X_i-\bar{X}_n)^2}{\sigma^4}.$$

Regardons le signe de la dérivée première :

$$\begin{aligned}
\frac{-n}{\sigma} + \frac{\sum_{i=1}^{n}(X_i-\bar{X}_n)^2}{\sigma^3} > 0 \ &\Leftrightarrow\ \sum_{i=1}^{n}(X_i-\bar{X}_n)^2 > -n\sigma^2 \\
&\Leftrightarrow\ \sum_{i=1}^{n}(X_i-\bar{X}_n)^2 > n\sigma^2 \\
&\Leftrightarrow\ \sigma^2 < \frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X}_n)^2
\end{aligned}$$

On obtient donc que $\mathcal{L}$ est strictement croissante pour σ allant de 0 à $\hat{\sigma}_n = \sqrt{\frac{1}{n}\sum_{i=1}^n (X_i - \bar{X}_n)^2}$ et strictement décroissante de $\hat{\sigma}_n$ à $+\infty$. La vraisemblance possède donc un unique maximum en

$$(\hat{\mu}_n, \hat{\sigma}_n) = \left(\bar{X}_n, \sqrt{\frac{1}{n}\sum_{i=1}^n (X_i - \bar{X}_n)^2}\right).$$

Remarques:

- $\hat{\sigma}_n^2$ correspond à la variance empirique non corrigée S_n^2.
- Si μ est supposé connu et que l'on cherche simplement à estimer σ alors, avec un raisonnement similaire, on obtient que l'EMV de σ est

$$\hat{\sigma}_n = \sqrt{\frac{1}{n}\sum_{i=1}^n (X_i - \mu)^2}$$

3 Loi du χ^2

Définition/Proposition 40

Soient $X_1, \cdots, X_n$ des v.a.i.i.d. de loi $\mathcal{N}(0,1)$. On appelle loi du χ^2 à n degrés de libertés, notée $\chi^2(n)$, la loi de la variable

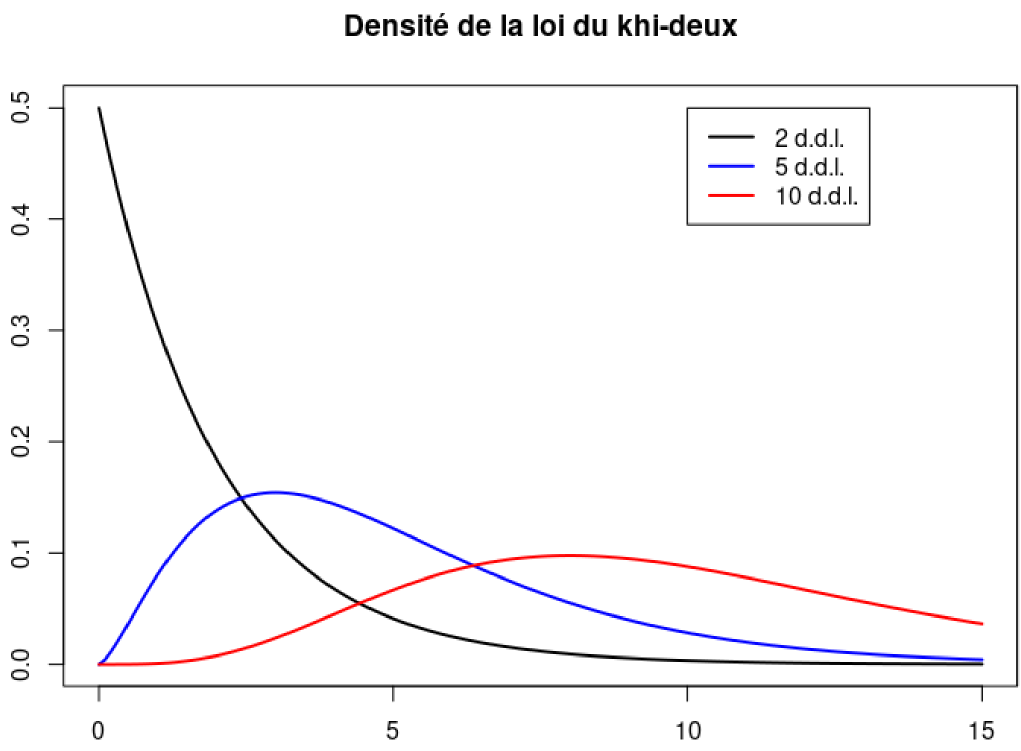
$$X_1^2 + \cdots + X_n^2.$$

C'est une loi à densité sur $\mathbb{R}_+$ de densité

$$\forall x \in \mathbb{R}_+,\ f(x) = \frac{x^{n/2-1}e^{-x/2}}{2^{n/2}\Gamma(n/2)}$$

où Γ est la foncion Gamma définie par

$$\forall x \in]0, +\infty[,\ \Gamma(x) = \int_0^{+\infty} t^{x-1}e^{-t}\mathrm{d}t.$$



Démonstration : On commence par montrer les propriétés suivantes de la fonction Γ :

Lemme 41

- $\Gamma(1/2) = \sqrt{\pi}$.
- $\Gamma(1) = 1$.
- $\forall x > 1,\ \Gamma(x) = (x-1)\Gamma(x-1)$.

Démonstration :

- $$\Gamma(1/2) = \int_0^{+\infty} t^{\frac{1}{2}-1} e^{-t} \mathrm{d}t = \int_0^{+\infty} \frac{1}{\sqrt{t}} e^{-t} \mathrm{d}t$$

  En utilisant le changement de variable $u = \sqrt{t}$ qui est bien une bijection croissante de $[0, +\infty[$ dans $[0, +\infty[$ avec $\mathrm{d}u = \frac{1}{2\sqrt{t}}\mathrm{d}t = \frac{1}{2u}\mathrm{d}t$ on obtient

  $$\Gamma(1/2) = \int_0^{+\infty} \frac{1}{u} e^{-u^2} (2u\mathrm{d}u) = 2\int_0^{+\infty} e^{-u^2} \mathrm{d}u = \int_{-\infty}^{+\infty} e^{-u^2} \mathrm{d}u,$$

  où on a utilisé le fait que la fonction $u \mapsto e^{-u^2}$ est paire dans la dernière égalité. Maintenant, on sait que

  $$\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi} \times \sqrt{\frac{1}{2}}} e^{-u^2} \mathrm{d}u = 1$$

  car la fonction à l'intérieur de l'intégrale est la densité de la loi $\mathcal{N}(0, \frac{1}{2})$ d'où

  $$\Gamma(1/2) = \int_{-\infty}^{+\infty} e^{-u^2} \mathrm{d}u = \sqrt{\pi}.$$

- $$\Gamma(1) = \int_0^{+\infty} e^{-t} \mathrm{d}t = [-e^{-t}]_0^{+\infty} = 1.$$

- Par intégration par partie pour $u(x) = t^{x-1}$ et $v'(x) = e^{-t}$ (et donc $u'(x) = (x-1)t^{x-1}$ et $v(x) = -e^{-t}$) on obtient :

  $$\Gamma(x) = \int_0^{+\infty} t^{x-1}e^{-t}\mathrm{d}t = [-t^{x-1}e^{-t}]_0^{+\infty} - \int_0^{+\infty} -(x-1)t^{x-2}e^{-t}\mathrm{d}t = (x-1)\Gamma(x-1),$$

  où on a utilisé le fait que $t^{x-1}e^{-t}$ vaut 0 quand $t = 0$ car $x > 1$. ■

On revient maintenant à la preuve de l'expression de la densité des lois du χ^2 que l'on effectue par récurrence.

- Preuve si $n = 1$:
  Soit ϕ une fonction continue bornée et $Y = X_1^2$. Alors

  $$\mathbb{E}[\phi(Y)] = \mathbb{E}\left[\phi\left(X_1^2\right)\right] = \int_{-\infty}^{\infty} \phi(x^2)\frac{e^{-x^2/2}}{\sqrt{2\pi}}\mathrm{d}x = 2\int_0^{\infty} \phi(x^2)\frac{e^{-x^2/2}}{\sqrt{2\pi}}\mathrm{d}x.$$

  On fait le changement de variable $y = x^2$ qui est bien une bijection croissante de $[0, +\infty[$ dans $[0, +\infty[$ d'inverse $x = \sqrt{y}$. On a $\mathrm{d}x = \mathrm{d}y/(2\sqrt{y})$ d'où

  $$\mathbb{E}[\phi(Y)] = 2\int_0^{\infty} \phi(y)\frac{e^{-y/2}}{\sqrt{2\pi}}\frac{\mathrm{d}y}{2\sqrt{y}} = \int_0^{\infty} \phi(y)\frac{y^{-1/2}e^{-y/2}}{2^{1/2}\Gamma(1/2)}\mathrm{d}y$$

  ce qui conclue le cas $n = 1$.
- Preuve si $n = 2$:
  Soient $X \sim \chi^2(1)$ et $Y \sim \chi^2(1)$ deux variables indépendantes et $Z = X + Y$. Si on pose f la densité de la loi $\chi^2(1)$ alors Z a pour densité

  $$\begin{aligned} f * f(x) &= \int_0^x f(t, x-t)\mathrm{d}t \\ &= \int_0^x \frac{e^{-t/2}}{\sqrt{2\pi t}}\frac{e^{-(x-t)/2}}{\sqrt{2\pi(x-t)}}\mathrm{d}t \\ &= \frac{e^{-x/2}}{2\pi}\int_0^x \frac{1}{\sqrt{t(x-t)}}\mathrm{d}t \end{aligned}$$

  On fait maintenant le changement de variable de t en $y = \frac{2t-x}{x}$, d'inverse $t = x\frac{y+1}{2}$, qui est une bijection croissante de $[0, x]$ en $[-1, 1]$ avec $\mathrm{d}y = \frac{2}{x}\mathrm{d}t$. Cela donne

  $$\begin{aligned} f * f(x) &= \frac{e^{-x/2}}{2\pi}\int_{-1}^1 \frac{1}{\sqrt{x\frac{y+1}{2}\left(x - x\frac{y+1}{2}\right)}} \times \frac{x}{2}\mathrm{d}y \\ &= \frac{e^{-x/2}}{2\pi}\int_{-1}^1 \frac{1}{\sqrt{1-y^2}}\mathrm{d}y \\ &= \frac{e^{-x/2}}{2\pi}\left[\arcsin(y)\right]_{y=-1}^{y=1} \\ &= \frac{e^{-x/2}}{2}. \end{aligned}$$

  C'est bien la densité de la loi $\chi^2(2)$ car $\Gamma(1) = 1$ donc $X + Y \sim \chi^2(2)$.
- Hérédité : Soient $X \sim \chi^2(n)$ et $Y \sim \chi^2(2)$ deux variables indépendantes et $Z = X + Y$. Avec le même raisonnement que pour le cas précédant on en déduit que la densité g de

Z vérifie :

$$\begin{aligned} g(x) &= \int_0^x \frac{t^{n/2-1}e^{-t/2}}{2^{n/2}\Gamma(n/2)} \frac{e^{-(x-t)/2}}{2} \mathrm{d}t \\ &= \frac{e^{-x/2}}{2^{n/2+1}\Gamma(n/2)} \int_0^x t^{n/2-1} \mathrm{d}t \\ &= \frac{x^{n/2}e^{-x/2}}{2^{(n+2)/2}\Gamma(n/2)(n/2)} \\ &= \frac{x^{(n+2)/2-1}e^{-x/2}}{2^{(n+2)/2}\Gamma((n+2)/2)}, \end{aligned}$$

où on a utilisé le fait que $\Gamma((n+2)/2) = \Gamma(n/2+1) = (n/2)\Gamma(n/2)$ pour la dernière égalité. g est donc bien la densité de la loi $\chi^2(n+2)$ d'où $X+Y \sim \chi^2(n+2)$ ce qui prouve la propriété par récurrence double. ■

Une conséquence immédiate est le résultat suivant :

Proposition 42

Soient $X_1, \cdots, X_n$ des v.a.i.i.d. de loi $\mathcal{N}(\mu, \sigma^2)$. Alors

$$\frac{1}{\sigma^2}\sum_{i=1}^n (X_i - \mu)^2 \sim \chi^2(n)$$

On se demande maintenant ce qui se passe si on cherche à remplacer μ par $\bar{X}_n$ dans ce résultat afin de pouvoir déterminer la loi de $\hat{\sigma}_n^2$.

Théorème 43 (*Cochran*)

Soient $X_1, \cdots, X_n$ des v.a.i.i.d. de loi $\mathcal{N}(0,1)$ et P une matrice de projection orthogonale de taille $n \times n$ et de rang r. On note $Y_1, \cdots, Y_n$ et $Z_1, \cdots, Z_n$ les variables aléatoires définies par

$$\begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} = P \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} \quad et \quad \begin{pmatrix} Z_1 \\ \vdots \\ Z_n \end{pmatrix} = (I_n - P) \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}.$$

Alors, les Y_i et Z_i suivent des lois normales, les variables Y_i sont indépendantes des variables Z_i et

$$Y_1^2 + \cdots + Y_n^2 \sim \chi^2(r) \quad et \quad Z_1^2 + \cdots + Z_n^2 \sim \chi^2(n-r)$$

Démonstration : Comme chaque Y_i et Z_i peut s'écrire comme combinaison linéaire des X_i qui sont indépendants et de loi normale alors ils suivent aussi une loi normale.

Comme P est une matrice de projection orthogonale alors, si on note $(e^{(1)}, \cdots, e^{(r)})$ une base orthonormale de $\mathrm{Im}(P)$ et $(e^{(r+1)}, \cdots, e^{(n)})$ une base orthonormale de $\mathrm{Im}(I_n - P) = \mathrm{Im}(P)^\perp$ et si on note X, Y, Z les vecteur des X_i, Y_i et Z_i, alors on obtient

$$Y = \sum_{i=1}^r \langle X, e^{(i)}\rangle e^{(i)} \quad \text{et} \quad Z = \sum_{i=r+1}^n \langle X, e^{(i)}\rangle e^{(i)}$$

donc, par le théorème de Pythagore,

$$Y_1^2 + \cdots + Y_n^2 = \sum_{i=1}^r \langle X, e^{(i)}\rangle^2 \text{ et } Z_1^2 + \cdots + Z_n^2 = \sum_{i=r+1}^{n-r} \langle X, e^{(i)}\rangle^2$$

On note $U_i = \langle X, e^{(i)} \rangle = X_1 e_1^{(i)} + \cdots + X_n e_n^{(i)}$. U_i suit une loi normale car c'est une combinaison linéaire de loi normales indépendantes. De plus,

$$\mathbb{E}[U_i] = \mathbb{E}[X_1] e_1^{(i)} + \cdots + \mathbb{E}[X_n] e_n^{(i)} = 0$$

et

$$\begin{aligned}
\mathrm{Cov}(U_i, U_j) &= \mathrm{Cov}(X_1 e_1^{(i)} + \cdots + X_n e_n^{(i)}, X_1 e_1^{(j)} + \cdots + X_n e_n^{(j)}) \\
&= \sum_{k,l=1}^{n} \mathrm{Cov}(X_k, X_l) e_k^{(i)} e_l^{(j)} \\
&= \sum_{k=1}^{n} e_k^{(i)} e_k^{(j)} \quad \text{car} \quad \mathrm{Cov}(X_k, X_l) = \mathbb{1}_{k=l} \\
&= \langle e^{(i)}, e^{(j)} \rangle
\end{aligned}$$

On en déduit que $Var(U_i) = \langle e^{(i)}, e^{(i)} \rangle = 1$ et donc tous les U_i suivent une loi $\mathcal{N}(0,1)$. On a ensuite que $\mathrm{Cov}(U_i, U_j) = 0$ si $i \neq j$. Ce n'est pas suffisant pour conclure que les U_i sont indépendant donc on va admettre que c'est le cas. On a donc obtenu que $Y_1^2 + \cdots + Y_n^2$ s'écrit comme une somme de r variables indépendantes de loi $\mathcal{N}(0,1)$ donc ça suit une loi $\chi^2(r)$. De même, $Z_1^2 + \cdots + Z_n^2 \sim \chi^2(n-r)$. De plus, l'indépendance des U_i nous dit que chaque Y_i est indépendant des Z_j et inversement. ■

Proposition 44

Soient $X_1, \cdots, X_n$ des v.a.i.i.d. de loi $\mathcal{N}(\mu, \sigma^2)$.

$$(n-1)\frac{\widetilde{S}_n^2}{\sigma^2} = n\frac{S_n^2}{\sigma^2} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2 \sim \chi^2(n-1)$$

De plus, $\widetilde{S}_n^2$ et S_n^2 sont indépendants de $\bar{X}_n$.

Démonstration : On note Y_i la variable $\frac{X_i - \mu}{\sigma}$ de loi $\mathcal{N}(0,1)$ et on remarque que

$$Y_i - \bar{Y}_n = \frac{X_i - \mu}{\sigma} - \frac{1}{n}\sum_{i=1}^{n}\frac{X_i - \mu}{\sigma} = \frac{1}{\sigma}\left(X_i - \mu - \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)\right) = \frac{1}{\sigma}(X_i - \bar{X}_n)$$

Donc

$$\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2 = \sum_{i=1}^{n}(Y_i - \bar{Y}_n)^2.$$

On définit ensuite les variables $Z_1, \cdots, Z_n$ par

$$\begin{pmatrix} Z_1 \\ \vdots \\ Z_n \end{pmatrix} = P \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} \quad \text{où} \quad P = I_n - \begin{pmatrix} 1/n & \cdots & 1/n \\ \vdots & \ddots & \vdots \\ 1/n & \cdots & 1/n \end{pmatrix}.$$

On obtient donc que $Z_i = Y_i - \frac{1}{n}\sum_{j=1}^{n} X_i = Y_i - \bar{Y}_n$ et comme $\begin{pmatrix} 1/n & \cdots & 1/n \\ \vdots & \ddots & \vdots \\ 1/n & \cdots & 1/n \end{pmatrix}$ est une matrice de projection orthogonale de rang 1 alors P est une matrice de projection orthogonale de rang $n-1$. Par le théorème de Cochran on en déduit donc :

$$\sum_{i=1}^{n}(Y_i - \bar{Y}_n)^2 = \sum_{i=1}^{n} Z_i^2 \sim \chi^2(n-1).$$

De plus, le théorème de Cochran nous dit aussi que les Z_i sont indépendant de

$$(I_n - P)X = \begin{pmatrix} 1/n & \cdots & 1/n \\ \vdots & \ddots & \vdots \\ 1/n & \cdots & 1/n \end{pmatrix} \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} \bar{Y}_n \\ \vdots \\ \bar{Y}_n \end{pmatrix}.$$

Comme $\bar{Y}_n = \frac{\bar{X}_n - \mu}{\sigma}$ alors σ_n^2 et S_n^2 sont bien indépendant de $\bar{X}_n$. ■

Remarque: Comme on a $\frac{1}{\sigma^2}\sum_{i=1}^n (X_i - \mu)^2 \sim \chi^2(n)$ alors le fait de remplacer μ par $\bar{X}_n$ fait "perdre" un degré de liberté.

4 Loi de Student

On a vu que

$$\sqrt{n}\frac{\bar{X}_n - \mu}{\sigma} \sim \mathcal{N}(0,1)$$

et ce résultat nous permet de contrôler l'écrat entre $\bar{X}_n$ et μ lorsque σ est connu. On se demande alors ce qu'il se passe si on remplace σ par $\hat{\sigma}_n$.

Définition/Proposition 45

Soient $X \sim \mathcal{N}(0,1)$ et $Y \sim \chi^2(n)$, pour un entier $n \geqslant 1$, deux variables aléatoires indépendantes. On appelle loi de Student à n degrés de liberté, notée $\mathcal{T}(n)$, la loi de $\sqrt{n}\frac{X}{\sqrt{Y}}$. C'est une loi à densité sur $\mathbb{R}$ de densité

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

Démonstration : Comme X et Y sont indépendants alors la densité du vecteur (X, Y) est

$$f(x,y) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}\frac{y^{n/2-1}e^{-y/2}}{2^{n/2}\Gamma(n/2)}.$$

Alors, pour toutes fonctions continues bornées $\phi : \mathbb{R} \to \mathbb{R}$, si on note Z la variable $\sqrt{n}\frac{X}{\sqrt{Y}}$ on a

$$\mathbb{E}[\phi(Z)] = \mathbb{E}\left[\phi\left(\sqrt{n}\frac{X}{\sqrt{Y}}\right)\right] = \int_{\mathbb{R}_+}\int_{\mathbb{R}} \phi\left(\sqrt{n}\frac{x}{\sqrt{y}}\right) f(x,y)\mathrm{d}x\mathrm{d}y.$$

On fait le changement de variable $x \mapsto z = \sqrt{n}\frac{x}{\sqrt{y}}$ à y fixé. C'est une bijection croissante de $\mathbb{R}$ dans lui-même d'inverse $x = z\sqrt{y/n}$ donc $\mathrm{d}x = \mathrm{d}z\sqrt{y/n}$. On a donc, par changement de variable,

$$\mathbb{E}[\phi(Z)] = \int_{\mathbb{R}_+}\int_{\mathbb{R}} \phi(z) f\left(z\sqrt{\frac{y}{n}}, y\right)\sqrt{\frac{y}{n}}\mathrm{d}z\mathrm{d}y = \int_{\mathbb{R}} \phi(z)g(z)\mathrm{d}z$$

où

$$g(z) = \int_{\mathbb{R}_+} \sqrt{\frac{y}{n}} f\left(z\sqrt{\frac{y}{n}}, y\right)\mathrm{d}y.$$

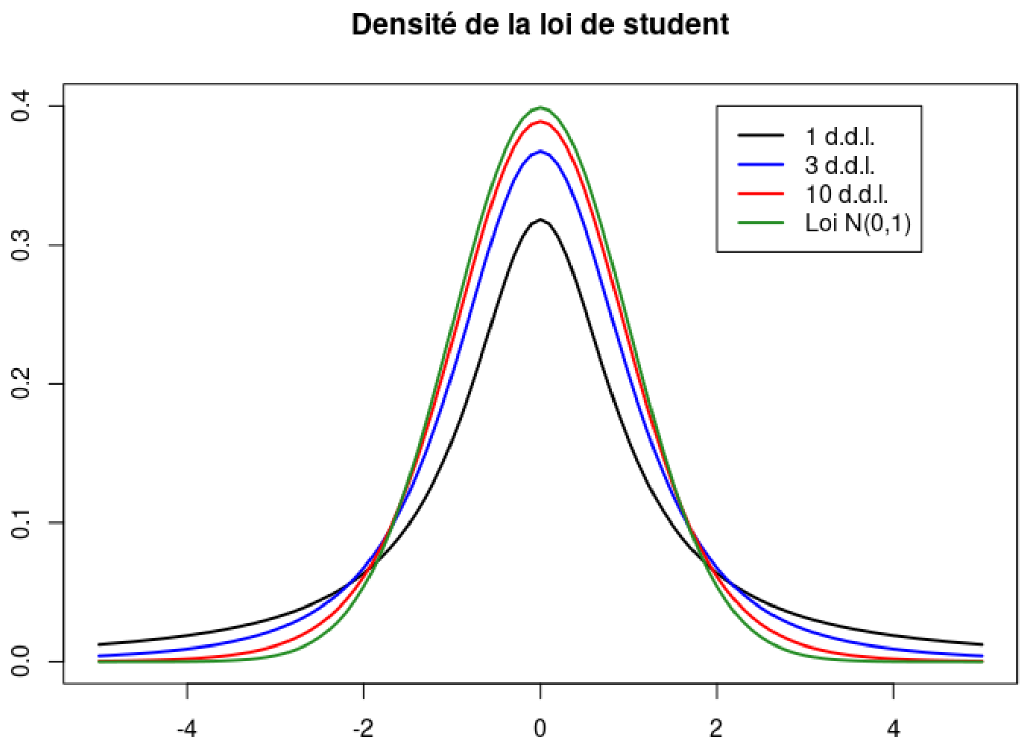
Donc, Z est une variable à densité sur $\mathbb{R}$ de densité g. On cherche alors à simplifier g. On commence par développer g en

$$\begin{aligned} g(z) &= \int_{\mathbb{R}_+} \sqrt{\frac{y}{n}} \times \frac{1}{\sqrt{2\pi}}e^{-z^2y/2n}\frac{y^{n/2-1}e^{-y/2}}{2^{n/2}\Gamma(n/2)}\mathrm{d}y. \\ &= \frac{1}{\sqrt{2\pi n}2^{n/2}\Gamma(n/2)}\int_{\mathbb{R}_+} y^{(n+1)/2-1}e^{-\frac{y}{2}(1+\frac{z^2}{n})}\mathrm{d}y. \end{aligned}$$

On fait maintenant le changement de variable $t = \frac{y}{2}(1 + \frac{z^2}{n})$. C'est une bijection croissante de $\mathbb{R}_+$ dans $\mathbb{R}_+$ d'inverse $y = \frac{2t}{(1+\frac{z^2}{n})}$ d'où $\mathrm{d}y = \frac{2}{(1+\frac{z^2}{n})}\mathrm{d}t$. Donc

$$
\begin{aligned}
g(z) &= \frac{1}{\sqrt{2\pi n}2^{n/2}\Gamma(n/2)} \int_{\mathbb{R}_+} \left(\frac{2t}{1+\frac{z^2}{n}}\right)^{(n+1)/2-1} e^{-t} \left(\frac{2}{1+\frac{z^2}{n}}\right) \mathrm{d}t \\
&= \frac{2^{(n+1)/2}\left(1+\frac{z^2}{n}\right)^{-(n+1)/2}}{\sqrt{2\pi n}2^{n/2}\Gamma(n/2)} \int_{\mathbb{R}_+} t^{(n+1)/2-1}e^{-t}\mathrm{d}t \\
&= \frac{\Gamma(\frac{n+1}{2})}{\sqrt{\pi n}\Gamma(n/2)}\left(1+\frac{z^2}{n}\right)^{-(n+1)/2}
\end{aligned}
$$

■



Proposition 46

Soient $X_1, \cdots, X_n$ des v.a.i.i.d. de loi $\mathcal{N}(\mu, \sigma^2)$. Alors,

$$\sqrt{n-1}\frac{\bar{X}_n - \mu}{\sqrt{S_n^2}} = \sqrt{n}\frac{\bar{X}_n - \mu}{\sqrt{\widetilde{S}_n^2}} \sim \mathcal{T}(n-1).$$

Démonstration : On a vu que $\sqrt{n}\frac{\bar{X}_n-\mu}{\sigma} \sim \mathcal{N}(0,1)$, que $\frac{(n-1)\widetilde{S}_n^2}{\sigma^2} \sim \chi^2(n-1)$ et que $\bar{X}_n$ et $\widetilde{S}_n^2$ sont indépendants. Donc

$$\sqrt{n-1}\frac{\sqrt{n}\frac{\bar{X}_n-\mu}{\sigma}}{\sqrt{(n-1)\frac{\widetilde{S}_n^2}{\sigma^2}}} \sim \mathcal{T}(n-1) \Longrightarrow \sqrt{n}\frac{\bar{X}_n-\mu}{\sqrt{\widetilde{S}_n^2}} \sim \mathcal{T}(n-1)$$

■

Bilan : On a vu quatres résultats essentiels :

- $\sqrt{n}\frac{\bar{X}_n-\mu}{\sigma} \sim \mathcal{N}(0,1) \longrightarrow$ Permet de contrôler $\hat{\mu}_n - \mu$ lorsque σ est connu.
- $\sqrt{n}\frac{\bar{X}_n-\mu}{\sqrt{\tilde{S}_n^2}} \sim \mathcal{T}(n-1) \longrightarrow$ Permet de contrôler $\hat{\mu}_n - \mu$ lorsque σ n'est pas connu.
- $\frac{1}{\sigma^2}\sum_{i=1}^n (X_i-\mu)^2 \sim \chi^2(n) \longrightarrow$ Permet de contrôler $\frac{\hat{\sigma}_n^2}{\sigma^2}$ lorsque μ est connu.
- $n\frac{S_n^2}{\sigma} \sim \chi^2(n-1) \longrightarrow$ Permet de contrôler $\frac{\hat{\sigma}_n^2}{\sigma^2}$ lorsque μ n'est pas connu.

5 Quantiles

Définition 47

*Soit X une variable aléatoire réelle à densité de fonction de répartition F continue et strictement croissante sur un intervalle I. Alors, F est inversible et, pour tout $\alpha \in]0,1[$, on appelle **quantile** d'ordre α la quantité $q_\alpha = F^{-1}(\alpha)$. Autrement dit, q_α est l'unqiue réeel x tel que*

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

Notation : On notera par la suite z_α, $\chi_\alpha^{(n)}$ et $t_\alpha^{(n)}$ les quantiles d'ordre α des lois $\mathcal{N}(0,1)$, $\chi^2(n)$ et $\mathcal{T}(n)$.

Proposition 48

Sous les hypothèses de la définition précédente, on a pour tout $\alpha \in]0,1[$

$$\mathbb{P}(q_{\alpha/2} \leqslant X \leqslant q_{1-\alpha/2}) = 1-\alpha.$$

Si la loi de X est symétrique (i.e. X et $-X$ ont la même loi) alors

$$\mathbb{P}(|X| \leqslant q_{1-\alpha/2}) = 1-\alpha.$$

Démonstration : Par définition des quantiles, on a

$$\mathbb{P}(X \leqslant q_{\alpha/2}) = \alpha/2 \text{ et } \mathbb{P}(X \leqslant q_{1-\alpha/2}) = 1-\alpha/2$$

d'où

$$\mathbb{P}(q_{\alpha/2} \leqslant X \leqslant q_{1-\alpha/2}) = \mathbb{P}(X \leqslant q_{1-\alpha/2}) - \mathbb{P}(X \leqslant q_{\alpha/2}) = 1 - \frac{\alpha}{2} - \frac{\alpha}{2} = 1-\alpha.$$

De plus,

$$\mathbb{P}(|X| \leqslant q_{1-\alpha/2}) = \mathbb{P}(-q_{1-\alpha/2} \leqslant X \leqslant q_{1-\alpha/2}) = 1-\alpha/2 - \mathbb{P}(X \leqslant -q_{1-\alpha/2})$$

Or,

$$\mathbb{P}(X \leqslant -q_{1-\alpha/2}) = \mathbb{P}(-X \geqslant q_{1-\alpha/2}) = \mathbb{P}(X \geqslant q_{1-\alpha/2}) = 1 - \mathbb{P}(X \leqslant q_{1-\alpha/2}) = \alpha/2$$

donc $\mathbb{P}(|X| \leqslant q_{1-\alpha/2}) = 1 - \frac{\alpha}{2} - \frac{\alpha}{2} = 1-\alpha$. ■

Remarques:

- Les lois $\mathcal{N}(0,1)$ et $\mathcal{T}(n)$ sont symétriques mais pas les lois $\chi^2(n)$.
- Si $X \sim \mathcal{N}(\mu, \sigma^2)$ alors

$$\mathbb{P}\left(\left|\frac{X-\mu}{\sigma}\right| \leqslant z_{1-\alpha/2}\right) = 1-\alpha \quad \Leftrightarrow \quad \mathbb{P}\left(|X-\mu| \leqslant z_{1-\alpha/2}\sigma\right) = 1-\alpha.$$

Comme $z_{1-0.05/2} \approx 1.96$, $z_{1-0.01/2} \approx 2.58$ et $z_{1-0.001/2} \approx 3.29$ on en déduit :

— Avec probabilité 95%, X est à valeur dans l'intervalle $[\mu - 1.96\sigma, \mu + 1.96\sigma]$.

— Avec probabilité 99%, X est à valeur dans l'intervalle $[\mu - 2.58\sigma, \mu + 2.58\sigma]$.

— Avec probabilité 99.9%, X est à valeur dans l'intervalle $[\mu - 3.29\sigma, \mu + 3.29\sigma]$.

Chapitre IV

Intervalle de confiance

<u>Cadre :</u> On considère $X_1, \cdots, X_n$, des v.a.i.i.d. de loi $\mathbb{P}_\theta \in \{\mathbb{P}_\theta, \theta \in \Theta\}$. On suppose dans cette section que $\Theta \subset \mathbb{R}$.

1 Définition

Définition 49 (*Intervalle de confiance*)

> *Soit $\alpha \in]0,1[$. Un **intervalle de confiance** de **niveau** $1-\alpha$ (ou de **risque** α) pour θ est un intervalle aléatoire $I_\alpha(X_1, \ldots, X_n) \subset \mathbb{R}$ tel que,*
>
> $$\forall \theta \in \Theta, \ \mathbb{P}_\theta(\theta \in I_\alpha(X_1, \ldots, X_n)) \geqslant 1 - \alpha.$$

⚠ $I_\alpha(X_1, \ldots, X_n)$ ne dépend que de l'échantillon et pas d'éléments inconnus (en particulier, l'intervalle ne doit pas dépendre de θ)

Définition 50 (*Intervalle de confiance asymptotique*)

> *Soit $\alpha \in]0,1[$. Un **intervalle de confiance asymptotique** de niveau $1-\alpha$ pour θ est une suite $\big(I_{n,\alpha}(X_1, \ldots, X_n)\big)_{n \in \mathbb{N}}$ d'intervalles aléatoires tels que,*
>
> $$\forall \theta \in \Theta, \ \lim_{n \to +\infty} \mathbb{P}_\theta(\theta \in I_{n,\alpha}(X_1, \ldots, X_n)) \geqslant 1 - \alpha$$

2 Intervalles de confiance pour les paramètres d'une loi normale

Une première façon de construire des intervalles de confiance exactes est de partir de quantités dont on connaît la loi exacte.

Définition 51

> *Une fonction $g(X_1, \cdots, X_n, \theta)$ est dite **pivotale** si la loi de $g(X_1, \cdots, X_n, \theta)$ est connue et ne dépend pas de θ.*

<u>*Exemple*</u>: On a vu que pour une loi $\mathcal{N}(0,1)$ on a

$$\sqrt{n}\frac{\bar{X}_n - \mu}{\sigma^2} \sim \mathcal{N}(0,1) \text{ et } \sqrt{n}\frac{\bar{X}_n - \mu}{\sqrt{\widetilde{S}_n^2}} \sim \mathcal{T}(n-1)$$

donc $\sqrt{n}\frac{\bar{X}_n-\mu}{\sigma^2}$ est pivotale quand σ est connu et $\sqrt{n}\frac{\bar{X}_n-\mu}{\sqrt{\widetilde{S}_n^2}}$ est pivotale que σ soit connu ou pas.

On peut en général construire des intervalles de confiance exacts à partir de fonctions pivotales. On va regarder des exemples pour les lois normales :

IC pour la moyenne lorsque la variance est connue

On considère le modèle $\{\mathcal{N}(\mu,\sigma^2);\mu\in\mathbb{R}\}$ où σ est connu. On a vu que $\sqrt{n}\frac{\bar{X}_n-\mu}{\sigma}\sim\mathcal{N}(0,1)$ donc

$$\begin{aligned}
&\mathbb{P}_\mu\left(\left|\sqrt{n}\frac{\bar{X}_n-\mu}{\sigma}\right|\leqslant z_{1-\alpha/2}\right)=1-\alpha\\
\Leftrightarrow\ &\mathbb{P}_\mu\left(\left|\bar{X}_n-\mu\right|\leqslant \frac{z_{1-\alpha/2}\sigma}{\sqrt{n}}\right)=1-\alpha\\
\Leftrightarrow\ &\mathbb{P}_\mu\left(\mu\in\left[\bar{X}_n-\frac{z_{1-\alpha/2}\sigma}{\sqrt{n}},\bar{X}_n+\frac{z_{1-\alpha/2}\sigma}{\sqrt{n}}\right]\right)=1-\alpha
\end{aligned}$$

donc $\left[\bar{X}_n-\frac{z_{1-\alpha/2}\sigma}{\sqrt{n}},\bar{X}_n+\frac{z_{1-\alpha/2}\sigma}{\sqrt{n}}\right]$ est un intervalle de confiance pour μ de niveau $1-\alpha$.

IC pour la moyenne lorsque la variance est inconnue

On considère le modèle $\{\mathcal{N}(\mu,\sigma^2);\mu\in\mathbb{R}\}$ où σ est inconnu. On a vu que $\sqrt{n}\frac{\bar{X}_n-\mu}{\sqrt{\widetilde{S}_n^2}}\sim\mathcal{T}(n-1)$ donc

$$\mathbb{P}_\mu\left(\left|\sqrt{n}\frac{\bar{X}_n-\mu}{\sqrt{\widetilde{S}_n^2}}\right|\leqslant t_{1-\alpha/2}^{(n-1)}\right)=1-\alpha$$

$$\Leftrightarrow\ \mathbb{P}_\mu\left(\mu\in\left[\bar{X}_n-\frac{t_{1-\alpha/2}^{(n-1)}\sqrt{\widetilde{S}_n^2}}{\sqrt{n}},\bar{X}_n+\frac{t_{1-\alpha/2}^{(n-1)}\sqrt{\widetilde{S}_n^2}}{\sqrt{n}}\right]\right)=1-\alpha$$

donc $\left[\bar{X}_n-\frac{t_{1-\alpha/2}^{(n-1)}\sqrt{\widetilde{S}_n^2}}{\sqrt{n}},\bar{X}_n+\frac{t_{1-\alpha/2}^{(n-1)}\sqrt{\widetilde{S}_n^2}}{\sqrt{n}}\right]$ est un intervalle de confiance pour μ de niveau $1-\alpha$.

IC pour la variance lorsque la moyenne est inconnue

On considère le modèle $\{\mathcal{N}(\mu,\sigma^2);\sigma\in\mathbb{R}_+^*\}$ où μ est inconnu. On a vu que $n\frac{S_n^2}{\sigma^2}\sim\chi^2(n-1)$ donc

$$\begin{aligned}
&\mathbb{P}_{\sigma^2}\left(\chi_{\alpha/2}^{(n-1)}\leqslant n\frac{S_n^2}{\sigma^2}\leqslant\chi_{1-\alpha/2}^{(n-1)}\right)=1-\alpha\\
\Leftrightarrow\ &\mathbb{P}_{\sigma^2}\left(\frac{1}{\chi_{1-\alpha/2}^{(n-1)}}\leqslant\frac{\sigma^2}{nS_n^2}\leqslant\frac{1}{\chi_{\alpha/2}^{(n-1)}}\right)=1-\alpha\\
\Leftrightarrow\ &\mathbb{P}_{\sigma^2}\left(\sigma^2\in\left[\frac{nS_n^2}{\chi_{1-\alpha/2}^{(n-1)}},\frac{nS_n^2}{\chi_{\alpha/2}^{(n-1)}}\right]\right)=1-\alpha
\end{aligned}$$

donc $\left[\frac{nS_n^2}{\chi_{1-\alpha/2}^{(n-1)}},\frac{nS_n^2}{\chi_{\alpha/2}^{(n-1)}}\right]$ est un intervalle de confiance pour σ^2 de niveau $1-\alpha$.

3 Exemples de construction d'intervalles de confiance

On se place dans le modèle $\{b(p); p \in]0,1[\}$. La loi $b(p)$ est de moyenne p et de variance égale à $p(1-p)$. On estime alors p par $\bar{X}_n$ et on cherche un intervalle de confiance de niveau $1-\alpha$ pour $\alpha \in]0,1[$.

a A partir d'une inégalité

$\bar{X}_n$ est de moyenne p et de variance $\frac{p(1-p)}{n}$ donc, d'après l'inégalité de Bienaymé-Tchebychev, pour tout $\varepsilon > 0$,

$$\mathbb{P}\left(|\bar{X}_n - p| > \varepsilon\right) \leqslant \frac{p(1-p)}{n\varepsilon^2} \leqslant \frac{1}{4n\varepsilon^2}.$$

En choisissant $\varepsilon = \frac{1}{2\sqrt{n\alpha}}$ on obtient

$$\begin{aligned}\mathbb{P}_p\left(|\bar{X}_n - p| > \frac{1}{2\sqrt{n\alpha}}\right) \leqslant \alpha \ &\Leftrightarrow \mathbb{P}_p\left(|\bar{X}_n - p| \leqslant \frac{1}{2\sqrt{n\alpha}}\right) \geqslant 1-\alpha \\ &\Leftrightarrow \mathbb{P}_p\left(p \in \left[\bar{X}_n - \frac{1}{2\sqrt{n\alpha}}, \bar{X}_n + \frac{1}{2\sqrt{n\alpha}}\right]\right) \geqslant 1-\alpha\end{aligned}$$

donc $\left[\bar{X}_n - \frac{1}{2\sqrt{n\alpha}}, \bar{X}_n + \frac{1}{2\sqrt{n\alpha}}\right]$ est un intervalle de confiance pour p de niveau $1-\alpha$.

b Avec un théorème limite

Comme $\mathbb{E}[X_1^2] = \mathbb{E}[X_1] = p < \infty$ alors, par le théorème central limite, on a :

$$\sqrt{n}\frac{\bar{X}_n - p}{\sqrt{p(1-p)}} \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}(0,1).$$

Attention, ici le dénominateur dépend de p que l'on ne connaît pas. Comme on estime p par $\bar{X}_n$, on aimerait donc remplacer $p(1-p)$ par $\bar{X}_n(1-\bar{X}_n)$. D'après la loi des grands nombres, $\bar{X}_n \xrightarrow[n\to+\infty]{p.s.} p$. Donc $\sqrt{\frac{p(1-p)}{\bar{X}_n(1-\bar{X}_n))}} \xrightarrow[n\to+\infty]{p.s.} 1$. Le lemme de Slutsky permet alors de dire que

$$\sqrt{n}\frac{\bar{X}_n - p}{\sqrt{p(1-p)}} \times \sqrt{\frac{p(1-p)}{\bar{X}_n(1-\bar{X}_n)}} \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}(0,1),$$

soit

$$\sqrt{n}\frac{\bar{X}_n - p}{\sqrt{\bar{X}_n(1-\bar{X}_n)}} \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}(0,1).$$

Donc pour tout $\alpha \in]0,1[$ on a

$$\mathbb{P}_p\left(\sqrt{n}\frac{|\bar{X}_n - p|}{\sqrt{\bar{X}_n(1-\bar{X}_n)}} \leqslant z_{1-\alpha}\right) \xrightarrow[n\to+\infty]{} \mathbb{P}(|Y| \leqslant z_{1-\alpha/2}) \text{ pour } Y \sim \mathcal{N}(0,1).$$

d'où

$$\mathbb{P}_p\left(p \in \left[\bar{X}_n - z_{1-\frac{\alpha}{2}}\sqrt{\frac{\bar{X}_n(1-\bar{X}_n)}{n}}, \bar{X}_n + z_{1-\frac{\alpha}{2}}\sqrt{\frac{\bar{X}_n(1-\bar{X}_n)}{n}}\right]\right) \xrightarrow[n\to+\infty]{} 1-\alpha.$$

L'intervalle $\left[\bar{X}_n - z_{1-\frac{\alpha}{2}}\sqrt{\frac{\bar{X}_n(1-\bar{X}_n)}{n}}, \bar{X}_n + z_{1-\frac{\alpha}{2}}\sqrt{\frac{\bar{X}_n(1-\bar{X}_n)}{n}}\right]$ est donc un intervalle de confiance asymptotique pour p de niveau $1-\alpha$.

Application : On a demandé à 1000 personnes, choisies aléatoirement uniformément et de façon indépendante dans la population, quelle était leur main dominante. 200 personnes ont répondu qu'elles étaient gauchères. Donner un intervalle de confiance asymptotique de niveau 95% sur la proportion de gaucher.
Soit p la proportion de personne gauchère dans la population et X_i la variable égale à 1 si la i-ème personne interrogée est gauchère et 0 sinon. Cela donne $X_i \sim b(p)$ et donc on estime p par $\bar{X}_n = \frac{200}{1000} = 0.2$ et on applique la formule précédente avec $n = 1000$, $\alpha = 0.05$ et $z_{0.975} \approx 1.96$ pour avoir l'intervalle de confiance

$$IC = \left[0.2 \pm 1.96 \sqrt{\frac{0.2(1-0.2)}{1000}} \right] = [0.2 \pm 0.025] = [0.175, 0.225].$$

Avec un niveau de certitude de 95% on obtient donc qu'entre 17.5% et 22.5% de la population est gauchère.
⚠ On remarque que la taille de l'intervalle est de l'ordre de $\frac{1}{\sqrt{n}}$. Cela veux dire que lorsqu'on souhaite doubler la précision d'un sondage on a besoin d'interroger quatre fois plus de personnes !

Chapitre V

Tests statistiques

1 Principe général

On considère qu'on observe des données $(x_1, \dots, x_n)$ issues d'un aléa et on cherche à trancher entre deux hypothèses sur l'aléa qui a engendré les données. On note ces hypothèses $\mathcal{H}_0$, appelée hypothèse nulle, et $\mathcal{H}_1$, appelée hypothèse alternative. C'est ce qu'on appelle un test d'hypothèses.
Exemple: On observe plusieurs lancers d'une pièce et on souhaite savoir à partir des résultats des lancers si la pièce est truquée ou pas.

Définition 52 (*Erreurs*)

L'erreur qui consiste à décider de rejeter $\mathcal{H}_0$ alors que $\mathcal{H}_0$ est vraie est appelée ***erreur de première espèce****.*
L'erreur qui consiste à conserver $\mathcal{H}_0$ alors que $\mathcal{H}_1$ est vraie est appelée ***erreur de seconde espèce****.*

Exemple: On cherche à conçevoir un test qui détecte une maladie.

$$\mathcal{H}_0 : \text{Le patient possède la maladie recherchée.}$$
$$\mathcal{H}_1 : \text{Le patient ne possède pas la maladie recherchée.}$$

L'erreur de première espèce correspond à un faux négatif. L'erreur de seconde espèce correspond à un faux positif.

Idéalement, on aimerait minimiser les deux risques d'erreur. En pratique, on doit trouver un compromis entre les deux car diminuer l'une fait souvent augmenter l'autre. On va considérer qu'une de ces deux erreurs est plus dommageable que l'autre (celle de première espèce) et éviter qu'elle se produise. Ceci conduit à un choix de $\mathcal{H}_0$.

2 Formalisation des tests paramétriques

Cadre : On dispose d'observations $(x_1, \dots, x_n)$ qu'on suppose issues de variables aléatoires i.i.d. $X_1, \cdots, X_n$ de loi $\mathbb{P}_\theta$ issue d'une famille $\{\mathbb{P}_\theta; \theta \in \Theta\}$. On sépare l'espace des paramètres Θ en deux espaces disjoints Θ_0 et Θ_1 :

$$\Theta_0 \cup \Theta_1 = \Theta \text{ et } \Theta_0 \cap \Theta_1 = \emptyset.$$

On considère alors que les hypothèses s'expriment sous la forme

$$\mathcal{H}_0 : \theta \in \Theta_0 \quad \text{contre} \quad \mathcal{H}_1 : \theta \in \Theta_1$$

Définition 53 (*Trois types de test*)

- *Un test est dit* ***simple*** *s'il est de la forme* $\mathcal{H}_0 : \theta = \theta_0$ *contre* $\mathcal{H}_1 : \theta = \theta_1$. *Ce test permet de déterminer quelle valeur est la plus probable entre* θ_0 *et* θ_1 *mais ne permet pas d'envisager d'autres valeurs pour* θ.
- *Un test est dit* ***bilatéral*** *s'il est de la forme* $\mathcal{H}_0 : \theta = \theta_0$ *contre* $\mathcal{H}_1 : \theta \neq \theta_0$
- *Un test est dit* ***unilatéral*** *s'il est de la forme* $\mathcal{H}_0 : \theta \leqslant \theta_0$ *contre* $\mathcal{H}_1 : \theta > \theta_0$ *ou* $\mathcal{H}_0 : \theta \geqslant \theta_0$ *contre* $\mathcal{H}_1 : \theta < \theta_0$.

Exemples:

- On souhaite tester si la durée de vie moyenne d'une ampoule est supérieure à 10000 heures comme le prétend le fabriquant. Pour cela, on teste un échantillon de n ampoules et on suppose que leur durée de vie suit une loi $\mathcal{E}(\lambda)$. Ici :

$$\mathcal{H}_0 : \frac{1}{\lambda} \geqslant 10000 \text{ heures} \quad \text{contre} \quad \mathcal{H}_1 : \frac{1}{\lambda} < 10000 \text{ heures.}$$

  C'est un test unilatéral.

- On souhaite tester si une pièce est bien équilibrée ou pas à partir du résultat de n lancés de pièce. On note p la probabilité que la pièce tombe sur face. Ici :

$$\mathcal{H}_0 : p = 1/2 \quad \text{contre} \quad \mathcal{H}_1 : p \neq 1/2.$$

  C'est un test bilatéral.

Définition 54 (*Statistique de test*)

On appelle ***statistique de test***, *une statistique* T *à valeur dans* $\{0, 1\}$ *associée à la stratégie suivante : Si* $T(x_1, \ldots, x_n) = 1$, *le test rejette* $\mathcal{H}_0$ *et accepte* $\mathcal{H}_1$. *Si* $T(x_1, \ldots, x_n) = 0$, *le test ne rejette pas* $\mathcal{H}_0$.

Définition 55

Le ***risque de première espèce*** *est la probabilité de rejeter* $\mathcal{H}_0$ *à tort (Faux négatif). C'est la fonction*

$$\underline{\alpha} : \begin{array}{ccc} \Theta_0 & \to & [0,1] \\ \theta & \mapsto & \mathbb{P}_\theta(T(X_1, \ldots, X_n) = 1). \end{array}$$

Le ***risque de seconde espèce*** *est la probabilité de ne pas rejeter* $\mathcal{H}_0$ *à tort (Faux positif). C'est la fonction*

$$\underline{\beta} : \begin{array}{ccc} \Theta_1 & \to & [0,1] \\ \theta & \mapsto & \mathbb{P}_\theta(T(X_1, \ldots, X_n) = 0). \end{array}$$

Décision \ Vérité	$\mathcal{H}_0$	$\mathcal{H}_1$
$\mathcal{H}_0$ non rejetté	$1 - \underline{\alpha}$	$\underline{\beta}$
$\mathcal{H}_0$ rejetté	$\underline{\alpha}$	$1 - \underline{\beta}$

Définition 56

Soit $\alpha \in]0,1[$,

- *Le test* T *est dit de* ***risque*** α *(ou de* ***niveau*** $1-\alpha$*) si*

$$\forall \theta \in \Theta_0,\ \underline{\alpha}(\theta) \leqslant \alpha \ \Leftrightarrow\ \mathbb{P}_\theta(T(X_1,\dots,X_n)=1) \leqslant \alpha.$$

- *Il est dit* ***asymptotiquement de risque*** α *si*

$$\forall \theta \in \Theta_0,\ \lim_{n\to+\infty} \mathbb{P}_\theta(T(X_1,\dots,X_n)=1) \leqslant \alpha.$$

- *La* ***puissance d'un test*** *est la fonction*

$$1-\underline{\beta} : \begin{array}{ccc} \Theta_1 & \to & [0,1] \\ \theta & \mapsto & \mathbb{P}_\theta(T(X_1,\dots,X_n)=1). \end{array}$$

Définition 57 (*Zone de rejet*)

On appelle ***zone de rejet*** *(ou région crtique) d'un test l'ensemble* R *des valeurs* $(x_1,\dots,x_n)$ *pour lesquelles le test rejette* $\mathcal{H}_0$.

Remarques:

- Soit R la zone de rejet d'une statistique de test T. Alors, $T = \mathbb{1}_{(X_1,\dots,X_n)\in R}$.
- Les intervalles de confiance nous seront très utiles pour construire ces zones de rejet.

Définition 58 (*p-valeur*)

Supposons donné un test T_α *pour un risque* $\alpha \in]0,1[$. *Considérons des observations fixées* $(x_1,\dots,x_n)$. *Il existe un seuil critique* α_c *tel que si* $\alpha \geqslant \alpha_c$, $T_\alpha(x_1,\dots,x_n)=1$ *et si* $\alpha < \alpha_c$, $T_\alpha(x_1,\dots,x_n)=0$. *Ce seuil critique est appelé* ***p-valeur***.

Remarque: La p-valeur est le plus petit risque α tel que le test rejette $\mathcal{H}_0$ pour ces données fixées. Une p-valeur très faible indique qu'on prend peu de risque en rejetant $\mathcal{H}_0$ pour ces données.

3 Tests sur la moyenne et la variance d'une loi normale

a Test sur la moyenne si la variance est connue

On considère que nos données proviennent de la famille $\{\mathcal{N}(\mu,\sigma^2),\ \mu\in\mathbb{R}\}$ avec σ^2 connu.

Test bilatéral

On considère les hypothèses $\mathcal{H}_0 : \mu=\mu_0$ et $\mathcal{H}_1 : \mu \neq \mu_0$ pour un $\mu_0 \in \mathbb{R}$ fixé. Regardons la statistique $Z = \sqrt{n}\frac{\bar{X}_n-\mu_0}{\sigma}$ que l'on peut décomposer en

$$Z = \sqrt{n}\frac{\bar{X}_n-\mu}{\sigma} + \sqrt{n}\frac{\mu-\mu_0}{\sigma}.$$

Le premier terme est de loi connue $\mathcal{N}(0,1)$. Le deuxième terme a un comportement différent selon la valeur de μ :

- Sous $\mathcal{H}_0$, ce terme vaut 0.
- Sous $\mathcal{H}_1$, ce terme diverge vers $\pm\infty$.

On s'attend donc à de grandes valeurs de la statistique sous $\mathcal{H}_1$ et à des valeurs plus faibles, voire très petites sous $\mathcal{H}_0$. On cherche donc une zone de rejet de la forme $|Z| > x$ avec x choisi de sorte que
$$\mathbb{P}_{\mu_0}(|Z| > x) = \alpha.$$
On a vu que cela revient à choisir $x = z_{1-\alpha/2}$. Le test s'écrit alors $\mathbb{1}_{|Z|>z_{1-\alpha/2}}$.
Calcul de la p-valeur associée : Le test rejette $\mathcal{H}_0$ ssi $|Z| = z_{1-\alpha/2} > F^{-1}_{\mathcal{N}(0,1)}(1-\alpha/2)$ où $F_{\mathcal{N}(0,1)}$ est la fonction de répartition de la loi $\mathcal{N}(0,1)$. Par stricte croissance de $F_{\mathcal{N}(0,1)}$ c'est équivalent à
$$1 - \frac{\alpha}{2} < F_{\mathcal{N}(0,1)}(Z) \quad \Leftrightarrow \quad \alpha > 2 - 2F_{\mathcal{N}(0,1)}(Z).$$
La p-valeur est donc égale à $2 - 2F_{\mathcal{N}(0,1)}(Z)$.

Test unilatéral

On considère les hypothèses $\mathcal{H}_0 : \mu \leqslant \mu_0$ et $\mathcal{H}_1 : \mu > \mu_0$ pour un $\mu_0 \in \mathbb{R}$ fixé.
Encore une fois on regarde la statistique $Z = \sqrt{n}\frac{\bar{X}_n-\mu_0}{\sigma}$ que l'on peut décomposer en
$$Z = \sqrt{n}\frac{\bar{X}_n - \mu}{\sigma} + \sqrt{n}\frac{\mu - \mu_0}{\sigma}.$$
Le premier terme est de loi connue $\mathcal{N}(0,1)$. Le deuxième terme a un comportement différent selon la valeur de μ :

- Sous $\mathcal{H}_0$, ce terme vaut 0 ou tend vers $-\infty$.
- Sous $\mathcal{H}_1$, ce terme tend vers $+\infty$.

On s'attend donc à de grandes valeurs de la statistique sous $\mathcal{H}_1$ et à des valeurs plus faibles, voire très petites sous $\mathcal{H}_0$. On cherche donc une zone de rejet de la forme $Z > x$. Sous $\mathcal{H}_0$ ($\mu \leqslant \mu_0$) on a
$$\begin{aligned}\mathbb{P}_\mu(Z > x) &= \mathbb{P}_\mu\left(\sqrt{n}\frac{\bar{X}_n - \mu}{\sigma} + \sqrt{n}\frac{\mu - \mu_0}{\sigma} > x\right)\\ &= \mathbb{P}_\mu\left(\sqrt{n}\frac{\bar{X}_n - \mu}{\sigma} > x - \sqrt{n}\frac{\mu - \mu_0}{\sigma}\right)\\ &\leqslant \mathbb{P}_\mu\left(\sqrt{n}\frac{\bar{X}_n - \mu}{\sigma} > x\right)\end{aligned}$$
Comme $\sqrt{n}\frac{\bar{X}_n-\mu}{\sigma} \sim \mathcal{N}(0,1)$ alors en prenant $x = z_{1-\alpha}$ on obtient $\mathbb{P}_\mu(Z > x) \leqslant \alpha$. Le test s'écrit donc $\mathbb{1}_{Z>z_{1-\alpha}}$.
Calcul de la p-valeur associée : Le test rejette $\mathcal{H}_0$ ssi $Z > z_{1-\alpha} = F^{-1}_{\mathcal{N}(0,1)}(1-\alpha)$. Par stricte croissance de $F_{\mathcal{N}(0,1)}$ c'est équivalent à
$$1 - \alpha < F_{\mathcal{N}(0,1)}(Z) \quad \Leftrightarrow \quad \alpha > 1 - F_{\mathcal{N}(0,1)}(Z).$$
La p-valeur est donc égale à $1 - F_{\mathcal{N}(0,1)}(Z)$.

Test unilatéral dans l'autre sens

On considère les hypothèses $\mathcal{H}_0 : \mu \geqslant \mu_0$ et $\mathcal{H}_1 : \mu < \mu_0$ pour un $\mu_0 \in \mathbb{R}$ fixé.
Encore une fois on regarde la statistique $Z = \sqrt{n}\frac{\bar{X}_n-\mu_0}{\sigma}$ que l'on peut décomposer en
$$Z = \sqrt{n}\frac{\bar{X}_n - \mu}{\sigma} + \sqrt{n}\frac{\mu - \mu_0}{\sigma}.$$
Le premier terme est de loi connue $\mathcal{N}(0,1)$. Le deuxième terme a un comportement différent selon la valeur de μ :

- Sous $\mathcal{H}_0$, ce terme vaut 0 ou tend vers $+\infty$.
- Sous $\mathcal{H}_1$, ce terme tend vers $-\infty$.

On s'attend donc à de petites valeurs de la statistique sous $\mathcal{H}_1$ et à des valeurs plus grandes, voire très grandes sous $\mathcal{H}_0$. On cherche donc une zone de rejet de la forme $Z < x$. Sous $\mathcal{H}_0$ ($\mu \leqslant \mu_0$) on a

$$\begin{aligned}\mathbb{P}_\mu(Z < x) &= \mathbb{P}_\mu\left(\sqrt{n}\frac{\bar{X}_n - \mu}{\sigma} + \sqrt{n}\frac{\mu - \mu_0}{\sigma} < x\right)\\ &= \mathbb{P}_\mu\left(\sqrt{n}\frac{\bar{X}_n - \mu}{\sigma} < x - \sqrt{n}\frac{\mu - \mu_0}{\sigma}\right)\\ &\leqslant \mathbb{P}_\mu\left(\sqrt{n}\frac{\bar{X}_n - \mu}{\sigma} < x\right)\end{aligned}$$

Comme $\sqrt{n}\frac{\bar{X}_n-\mu}{\sigma} \sim \mathcal{N}(0,1)$ alors en prenant $x = z_\alpha$ on obtient $\mathbb{P}_\mu(Z < x) \leqslant \alpha$. Le test s'écrit donc $\mathbb{1}_{Z<z_\alpha}$.
Calcul de la p-valeur associée : Le test rejette $\mathcal{H}_0$ ssi $Z < z_\alpha = F^{-1}_{\mathcal{N}(0,1)}(\alpha)$. Par stricte croissance de $F_{\mathcal{N}(0,1)}$ c'est équivalent à $\alpha > F_{\mathcal{N}(0,1)}(Z)$. La p-valeur est donc égale à $F_{\mathcal{N}(0,1)}(Z)$.

b Test sur la moyenne si la variance est inconnue

On considère que nos données proviennent de la famille $\{\mathcal{N}(\mu,\sigma^2),\ \mu \in \mathbb{R}\}$ avec σ^2 inconnu.

Test bilatéral

On considère les hypothèses $\mathcal{H}_0 : \mu = \mu_0$ et $\mathcal{H}_1 : \mu \neq \mu_0$ pour un $\mu_0 \in \mathbb{R}$ fixé.
Regardons la statistique $T = \sqrt{n}\frac{\bar{X}_n-\mu_0}{\sqrt{\widetilde{S}^2_n}}$ que l'on peut décomposer en

$$T = \sqrt{n}\frac{\bar{X}_n - \mu}{\sqrt{\widetilde{S}^2_n}} + \sqrt{n}\frac{\mu - \mu_0}{\sqrt{\widetilde{S}^2_n}}.$$

Le premier terme est de loi connue $\mathcal{T}(n-1)$. Le deuxième terme a un comportement différent selon la valeur de μ :

- Sous $\mathcal{H}_0$, ce terme vaut 0.
- Sous $\mathcal{H}_1$, ce terme diverge vers $\pm\infty$.

On s'attend donc à de grandes valeurs de la statistique sous $\mathcal{H}_1$ et à des valeurs plus faibles, voire très petites sous $\mathcal{H}_0$. On cherche donc une zone de rejet de la forme $|T| > x$ avec x choisi de sorte que

$$\mathbb{P}_{\mu_0}(|T| > x) = \alpha.$$

On a vu que cela revient à choisir $x = t^{(n-1)}_{1-\alpha/2}$. Le test s'écrit alors $\mathbb{1}_{|T|>t^{(n-1)}_{1-\alpha/2}}$.
Calcul de la p-valeur associée : Le test rejette $\mathcal{H}_0$ ssi $|T| > t^{(n-1)}_{1-\alpha/2} = F^{-1}_{\mathcal{T}(n-1)}(1-\alpha/2)$ où $F_{\mathcal{T}(n-1)}$ est la fonction de répartition de la loi $\mathcal{T}(n-1)$. Par stricte croissance de $F_{\mathcal{T}(n-1)}$ c'est équivalent à

$$1 - \frac{\alpha}{2} < F_{\mathcal{T}(n-1)}(T) \quad \Leftrightarrow \quad \alpha > 2 - 2F_{\mathcal{T}(n-1)}(T).$$

La p-valeur est donc égale à $2 - 2F_{\mathcal{T}(n-1)}(T)$.

Test unilatéral

On considère les hypothèses $\mathcal{H}_0 : \mu \leqslant \mu_0$ et $\mathcal{H}_1 : \mu > \mu_0$ pour un $\mu_0 \in \mathbb{R}$ fixé.
Encore une fois on regarde la statistique $T = \sqrt{n}\frac{\bar{X}_n - \mu_0}{\sqrt{\widetilde{S}_n^2}}$ que l'on peut décomposer en

$$T = \sqrt{n}\frac{\bar{X}_n - \mu}{\sqrt{\widetilde{S}_n^2}} + \sqrt{n}\frac{\mu - \mu_0}{\sqrt{\widetilde{S}_n^2}}.$$

Le premier terme est de loi connue $\mathcal{N}(0,1)$. Le deuxième terme a un comportement différent selon la valeur de μ :

- Sous $\mathcal{H}_0$, ce terme vaut 0 ou tend vers $-\infty$.
- Sous $\mathcal{H}_1$, ce terme tend vers $+\infty$.

On s'attend donc à de grandes valeurs de la statistique sous $\mathcal{H}_1$ et à des valeurs plus faibles, voire très petites sous $\mathcal{H}_0$. On cherche donc une zone de rejet de la forme $T > x$. Sous $\mathcal{H}_0$ ($\mu \leqslant \mu_0$) on a

$$\begin{aligned}
\mathbb{P}_\mu(T > x) &= \mathbb{P}_\mu\left(\sqrt{n}\frac{\bar{X}_n - \mu}{\sqrt{\widetilde{S}_n^2}} + \sqrt{n}\frac{\mu - \mu_0}{\sqrt{\widetilde{S}_n^2}} > x\right) \\
&= \mathbb{P}_\mu\left(\sqrt{n}\frac{\bar{X}_n - \mu}{\sqrt{\widetilde{S}_n^2}} > x - \sqrt{n}\frac{\mu - \mu_0}{\sqrt{\widetilde{S}_n^2}}\right) \\
&\leqslant \mathbb{P}_\mu\left(\sqrt{n}\frac{\bar{X}_n - \mu}{\sqrt{\widetilde{S}_n^2}} > x\right)
\end{aligned}$$

Comme $\sqrt{n}\frac{\bar{X}_n - \mu}{\sigma} \sim \mathcal{T}(n-1)$ alors en prenant $x = t_{1-\alpha}^{(n-1)}$ on obtient $\mathbb{P}_\mu(T > x) \leqslant \alpha$. Le test s'écrit donc $\mathbb{1}_{T > t_{1-\alpha}^{(n-1)}}$.
Calcul de la p-valeur associée : Le test rejette $\mathcal{H}_0$ ssi $T > t_{1-\alpha}^{(n-1)} = F_{\mathcal{T}(n-1)}^{-1}(1-\alpha)$. Par stricte croissance de $F_{\mathcal{T}(n-1)}$ c'est équivalent à

$$1 - \alpha < F_{\mathcal{T}(n-1)}(T) \quad \Leftrightarrow \quad \alpha > 1 - F_{\mathcal{T}(n-1)}(T).$$

La p-valeur est donc égale à $1 - F_{\mathcal{T}(n-1)}(T)$.

Test unilatéral dans l'autre sens

On considère les hypothèses $\mathcal{H}_0 : \mu \geqslant \mu_0$ et $\mathcal{H}_1 : \mu < \mu_0$ pour un $\mu_0 \in \mathbb{R}$ fixé.
Encore une fois on regarde la statistique $T = \sqrt{n}\frac{\bar{X}_n - \mu_0}{\sqrt{\widetilde{S}_n^2}}$ que l'on peut décomposer en

$$T = \sqrt{n}\frac{\bar{X}_n - \mu}{\sqrt{\widetilde{S}_n^2}} + \sqrt{n}\frac{\mu - \mu_0}{\sqrt{\widetilde{S}_n^2}}.$$

Le premier terme est de loi connue $\mathcal{N}(0,1)$. Le deuxième terme a un comportement différent selon la valeur de μ :

- Sous $\mathcal{H}_0$, ce terme vaut 0 ou tend vers $+\infty$.
- Sous $\mathcal{H}_1$, ce terme tend vers $-\infty$.

On s'attend donc à de petites valeurs de la statistique sous $\mathcal{H}_1$ et à des valeurs plus grandes, voire très grandes sous $\mathcal{H}_0$. On cherche donc une zone de rejet de la forme $T < x$. Sous $\mathcal{H}_0$ ($\mu \leqslant \mu_0$) on a

$$\begin{aligned}\mathbb{P}_\mu(T < x) &= \mathbb{P}_\mu\left(\sqrt{n}\frac{\bar{X}_n - \mu}{\sqrt{\widetilde{S}_n^2}} + \sqrt{n}\frac{\mu - \mu_0}{\sqrt{\widetilde{S}_n^2}} < x\right) \\ &= \mathbb{P}_\mu\left(\sqrt{n}\frac{\bar{X}_n - \mu}{\sqrt{\widetilde{S}_n^2}} < x - \sqrt{n}\frac{\mu - \mu_0}{\sqrt{\widetilde{S}_n^2}}\right) \\ &\leqslant \mathbb{P}_\mu\left(\sqrt{n}\frac{\bar{X}_n - \mu}{\sqrt{\widetilde{S}_n^2}} < x\right)\end{aligned}$$

Comme $\sqrt{n}\frac{\bar{X}_n-\mu}{\sigma} \sim \mathcal{T}(n-1)$ alors en prenant $x = t_\alpha^{(n-1)}$ on obtient $\mathbb{P}_\mu(T < x) \leqslant \alpha$. Le test s'écrit donc $\mathbb{1}_{T<t_\alpha^{(n-1)}}$.

Calcul de la p-valeur associée : Le test rejette $\mathcal{H}_0$ ssi $T > t_\alpha^{(n-1)} = F^{-1}_{\mathcal{T}(n-1)}(1-\alpha)$. Par stricte croissance de $F_{\mathcal{T}(n-1)}$ c'est équivalent à $\alpha > F_{\mathcal{T}(n-1)}(T)$. La p-valeur est donc égale à $F_{\mathcal{T}(n-1)}(T)$.

c Tests sur la variance si la moyenne est inconnue

On se place maintenant dans le modèle gaussien $\{\mathcal{N}(\mu,\sigma^2);\ \sigma > 0\}$ où μ est inconnu.

Test bilatéral

On considère les hypothèses $\mathcal{H}_0 : \sigma = \sigma_0$ et $\mathcal{H}_1 : \sigma \neq \sigma_0$ pour un $\sigma_0 > 0$ fixé.
On considère la statistique $S = \frac{(n-1)\widetilde{S}_n^2}{\sigma_0^2}$ que l'on décompose en

$$S = \frac{(n-1)\widetilde{S}_n^2}{\sigma^2} \times \frac{\sigma^2}{\sigma_0^2}$$

Le premier terme est de loi $\chi^2(n-1)$ tandis que le deuxième terme est $= 1$ sous $\mathcal{H}_0$ et $\neq 1$ sous $\mathcal{H}_1$. On s'attend donc à ce que la statistique prenne des valeurs plus grandes ou plus petites que celle d'une loi $\chi^2(n-1)$ sous $\mathcal{H}_1$.
On cherche donc une zone de rejet de la forme $S > x$ ou $S < y$. On choisit alors x et y de sorte que

$$\mathbb{P}_{\sigma_0}(S > x \text{ ou } S < y) = \alpha.$$

Comme $S \sim \chi^2(n-1)$ sous $\mathcal{H}_0$ alors $\mathbb{P}_{\sigma_0}(S > \chi^{(n-1)}_{1-\alpha/2}) = 1 - \alpha/2$ et $\mathbb{P}_{\sigma_0}(S < \chi^{(n-1)}_{\alpha/2}) = \alpha/2$ d'où

$$\mathbb{P}_{\sigma_0}(S > \chi^{(n-1)}_{1-\alpha/2} \text{ ou } S < \chi^{(n-1)}_{\alpha/2}) = \alpha.$$

Le test s'écrit donc

$$\mathbb{1}_{S<\chi^{(n-1)}_{\alpha/2} \text{ ou } S>\chi^{(n-1)}_{1-\alpha/2}}.$$

Calcul de la p-valeur associée : Le test rejette $\mathcal{H}_0$ ssi $S < \chi^{(n-1)}_{\alpha/2} = F^{-1}_{\chi^2(n-1)}(\frac{\alpha}{2})$ ou $S > \chi^{(n-1)}_{1-\alpha/2} = F^{-1}_{\chi^2(n-1)}(1-\frac{\alpha}{2})$ où $F_{\chi^2(n-1)}$ est la fonction de répartition de la loi $\chi^2(n-1)$. Par stricte croissance de $F_{\chi^2(n-1)}$, cette condition est équivalente à

$$\alpha > 2F_{\chi^2(n-1)}(S) \text{ ou } \alpha > 2 - 2F_{\chi^2(n-1)}(S).$$

La p-valeur est donc égale à $\min\left(F_{\chi^2(n-1)}(S), 2 - 2F_{\chi^2(n-1)}(S)\right)$.

Test unilatéral

On considère les hypothèses $\mathcal{H}_0 : \sigma \leqslant \sigma_0$. $\mathcal{H}_1 : \sigma > \sigma_0$ pour un $\sigma_0 > 0$ fixé.
On part de la même statistique de test S que l'on décompose en

$$S = \frac{(n-1)\widetilde{S}_n^2}{\sigma^2} \times \frac{\sigma^2}{\sigma_0^2}.$$

Le premier terme est de loi $\chi^2(n-1)$ tandis que le deuxième terme est $\leqslant 1$ sous $\mathcal{H}_0$ et > 1 sous $\mathcal{H}_1$. On s'attend donc sous $\mathcal{H}_0$ à des valeurs plus petites de S que sous $\mathcal{H}_1$. On cherche par conséquent une zone de rejet de la forme $S > x$. Sous $\mathcal{H}_0$ ($\sigma \leqslant \sigma_0$) :

$$\mathbb{P}_\sigma(S > x) = \mathbb{P}_\sigma\left(\frac{(n-1)\widetilde{S}_n^2}{\sigma^2} > x\frac{\sigma_0^2}{\sigma^2}\right) \leqslant \mathbb{P}_\sigma\left(\frac{(n-1)\widetilde{S}_n^2}{\sigma^2} > x\right).$$

Comme $\frac{(n-1)\widetilde{S}_n^2}{\sigma^2} \sim \chi^2(n-1)$ alors en prenant $x = \chi_{1-\alpha}^{(n-1)}$ on obtient $\mathbb{P}_\sigma(S > x) \leqslant \alpha$. Le test s'écrit donc $\mathbb{1}_{S > \chi_{1-\alpha}^{(n-1)}}$.
Calcul de la p-valeur associée : Le test rejette $\mathcal{H}_0$ ssi $S > \chi_{1-\alpha}^{(n-1)} = F_{\chi^2(n-1)}^{-1}(1-\alpha)$. Par stricte croissance de $F_{\chi^2(n-1)}$, cette condition est équivalente à

$$1 - \alpha > F_{\chi^2(n-1)}(S) \quad \Leftrightarrow \quad \alpha < 1 - F_{\chi^2(n-1)}(S).$$

La p-valeur est donc égale à $1 - F_{\chi^2(n-1)}(S)$.

Test unilatéral dans l'autre sens

On considère les hypothèses $\mathcal{H}_0 : \sigma \geqslant \sigma_0$. $\mathcal{H}_1 : \sigma < \sigma_0$ pour un $\sigma_0 > 0$ fixé.
On part de la même statistique de test S que l'on décompose en

$$S = \frac{(n-1)\widetilde{S}_n^2}{\sigma^2} \times \frac{\sigma^2}{\sigma_0^2}.$$

Le premier terme est de loi $\chi^2(n-1)$ tandis que le deuxième terme est $\geqslant 1$ sous $\mathcal{H}_0$ et < 1 sous $\mathcal{H}_1$. On s'attend donc sous $\mathcal{H}_0$ à des valeurs plus grandes de S que sous $\mathcal{H}_1$. On cherche par conséquent une zone de rejet de la forme $S < x$. Sous $\mathcal{H}_0$ ($\sigma \leqslant \sigma_0$) :

$$\mathbb{P}_\sigma(S < x) = \mathbb{P}_\sigma\left(\frac{(n-1)\widetilde{S}_n^2}{\sigma^2} < x\frac{\sigma_0^2}{\sigma^2}\right) \leqslant \mathbb{P}_\sigma\left(\frac{(n-1)\widetilde{S}_n^2}{\sigma^2} < x\right).$$

Comme $\frac{(n-1)S_n^2}{\sigma^2} \sim \chi^2(n-1)$ alors en prenant $x = \chi_{\alpha}^{(n-1)}$ on obtient $\mathbb{P}_\sigma(S < x) \leqslant \alpha$. Le test s'écrit donc $\mathbb{1}_{S < \chi_{\alpha}^{(n-1)}}$.
Calcul de la p-valeur associée : Le test rejette $\mathcal{H}_0$ ssi $S < \chi_{\alpha}^{(n-1)} = F_{\chi^2(n-1)}^{-1}(\alpha)$. Par stricte croissance de $F_{\chi^2(n-1)}$, cette condition est équivalente à $\alpha < F_{\chi^2(n-1)}(S)$. La p-valeur est donc égale à $1 - F_{\chi^2(n-1)}(S)$.

Remarque: Pour les tests sur la variance si la moyenne est connue alors c'est exactement la même chose en remplaçant S par

$$\frac{1}{\sigma_0^2}\sum_{i=1}^{n}(X_i - \mu)^2$$

et les lois $\chi^2(n-1)$ par des lois $\chi^2(n)$.

On résume alors tous les tests sur les lois normales dans le tableau suivant :

Tests sur la moyenne si la variance est connue			
Statistique de test : $Z = \sqrt{n}\dfrac{\bar{X}_n - \mu_0}{\sigma}$			
$\mathcal{H}_0$	$\mathcal{H}_1$	Zone de rejet	p-valeur
$\mu = \mu_0$	$\mu \neq \mu_0$	$\lvert Z\rvert > z_{1-\alpha/2}$	$2 - 2F_{\mathcal{N}(0,1)}(T)$
$\mu \leqslant \mu_0$	$\mu > \mu_0$	$Z > z_{1-\alpha}$	$1 - F_{\mathcal{N}(0,1)}(T)$
$\mu \geqslant \mu_0$	$\mu < \mu_0$	$Z < z_{\alpha}$	$F_{\mathcal{N}(0,1)}(T)$
Tests sur la moyenne si la variance est inconnue			
Statistique de test : $T = \sqrt{n}\dfrac{\bar{X}_n - \mu_0}{\sqrt{\widetilde{S}_n^2}}$			
$\mathcal{H}_0$	$\mathcal{H}_1$	Zone de rejet	p-valeur
$\mu = \mu_0$	$\mu \neq \mu_0$	$\lvert T\rvert > t_{1-\alpha/2}^{(n-1)}$	$2 - 2F_{\mathcal{T}(n-1)}(T)$
$\mu \leqslant \mu_0$	$\mu > \mu_0$	$T > t_{1-\alpha}^{(n-1)}$	$1 - F_{\mathcal{T}(n-1)}(T)$
$\mu \geqslant \mu_0$	$\mu < \mu_0$	$T < t_{\alpha}^{(n-1)}$	$F_{\mathcal{T}(n-1)}(T)$
Tests sur la variance si la moyenne est connue			
Statistique de test : $S = \dfrac{1}{\sigma_0^2}\sum_{i=1}^{n}(X_i - \mu)^2$			
$\mathcal{H}_0$	$\mathcal{H}_1$	Zone de rejet	p-valeur
$\sigma^2 = \sigma_0^2$	$\sigma^2 \neq \sigma_0^2$	$S < \chi_{\alpha/2}^{(n)}$ ou $S > \chi_{1-\alpha/2}^{(n)}$	$2\min\left(F_{\chi^2(n)}(S), 1 - F_{\chi^2(n)}(S)\right)$
$\sigma^2 \leqslant \sigma_0^2$	$\sigma^2 > \sigma_0^2$	$S > \chi_{1-\alpha}^{(n)}$	$1 - F_{\chi^2(n)}(S)$
$\sigma^2 \geqslant \sigma_0^2$	$\sigma^2 < \sigma_0^2$	$S < \chi_{\alpha}^{(n)}$	$F_{\chi^2(n)}(S)$
Tests sur la variance si la moyenne est inconnue			
Statistique de test : $S = \dfrac{(n-1)\widetilde{S}_n^2}{\sigma_0^2}$			
$\mathcal{H}_0$	$\mathcal{H}_1$	Zone de rejet	p-valeur
$\sigma^2 = \sigma_0^2$	$\sigma^2 \neq \sigma_0^2$	$S < \chi_{\alpha/2}^{(n-1)}$ ou $S > \chi_{1-\alpha/2}^{(n-1)}$	$2\min\left(F_{\chi^2(n-1)}(S), 1 - F_{\chi^2(n-1)}(S)\right)$
$\sigma^2 \leqslant \sigma_0^2$	$\sigma^2 > \sigma_0^2$	$S > \chi_{1-\alpha}^{(n-1)}$	$1 - F_{\chi^2(n-1)}(S)$
$\sigma^2 \geqslant \sigma_0^2$	$\sigma^2 < \sigma_0^2$	$S < \chi_{\alpha}^{(n-1)}$	$F_{\chi^2(n-1)}(S)$

Exemple: On cherche à tester une boîte de médicament en pilule pour savoir si la quantité moyenne d'agent actif dans chaque pilule est de 10mg avec une variance $\leqslant$ 0.4mg^2. Un test sur 30 pilules au hasard à trouvé que la quantité moyenne d'agent actif est de 9.7mg avec une variance corrigée de 0.5mg^2. Est-ce acceptable à un risque 5% ?

Réponse : On note X_i la quantité d'agent actif dans la pilule i et on suppose que les X_i sont i.i.d. de loi $\mathcal{N}(\mu, \sigma^2)$. On considère d'abord les hypothèses $\mathcal{H}_0$: $\mu = 10$mg et $\mathcal{H}_1$: $\mu \neq 10$mg.

Les données nous donnent $n = 30$, $\bar{X}_n = 9.7$mg et $\widetilde{S}_n^2 = 0.5$mg^2. On a vu que l'on rejetait $\mathcal{H}_0$ si

$$\sqrt{n}\frac{|\bar{X}_n - 10|}{\sqrt{\widetilde{S}_n^2}} > t_{1-\alpha/2}^{(n-1)}.$$

Dans notre cas on a $t_{1-\alpha/2}^{(n-1)} = t_{0.975}^{29} \approx 2.04$ et

$$\sqrt{n}\frac{|\bar{X}_n - 10|}{\sqrt{\widetilde{S}_n^2}} = \sqrt{30}\frac{|9.7 - 10|}{\sqrt{0.5}} \approx 2.32,$$

On rejette donc l'hypothèse que la quantité moyenne d'agent actif dans les tablettes est de 10mg au risque 5%. On considère ensuite les hypothèses $\mathcal{H}_0 : \sigma^2 \leqslant 0.4\text{mg}^2$ et $\mathcal{H}_1 : \sigma^2 > 0.4\text{mg}^2$. On a vu qu'on rejette $\mathcal{H}_0$ si

$$\frac{(n-1)S_n^2}{0.4} > \chi_{1-\alpha}^{(n-1)}.$$

Dans notre cas, on a $\chi_{0.95}^{(29)} \approx 42.56$ et

$$\frac{(n-1)S_n^2}{0.4} = \frac{29 \times 0.5}{0.4} \approx 36.25.$$

On ne rejette donc pas l'hypothèse que la variance de la quantité d'agent actifs est $\leqslant 0.4$mg.
⚠ En pratique ce n'est pas cette méthode utilisée pour tester des boites de médicament car ces tests ne sont pas adaptés à la détection de valeurs absurde. On trouvera plus de détail sur le vrai test qui est fait en pratique sur ce lien.

4 Tests à partir d'un estimateur asymptotiquement normal

Dans le cas où on n'est pas dans un modèle gaussien, il est possible de construire un test asymptotique en utilisant la normalité asymptotique des estimateurs. Comme exemple on considère un modèle de Bernoulli $\{b(p), p \in]0,1[\}$ et un test bilatéral :

$$\mathcal{H}_0 : p = p_0 \quad \text{contre} \quad \mathcal{H}_1 : p \neq p_0.$$

On pose la statistique

$$Z_n = \sqrt{n}\frac{\bar{X}_n - p_0}{\sqrt{\bar{X}_n(1-\bar{X}_n)}}.$$

En utilisant la loi forte des grands nombres, le théorème central limite et le lemme de Slutsky, on obtient sous $\mathcal{H}_0$:

$$Z_n = \frac{\sqrt{p(1-p)}}{\sqrt{\bar{X}_n(1-\bar{X}_n)}} \times \sqrt{n}\frac{\bar{X}_n - p_0}{\sqrt{p(1-p)}} \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}(0,1)$$

De plus, par la loi forte des grands nombres on obtient sous $\mathcal{H}_1$:

$$\frac{\bar{X}_n - p_0}{\sqrt{\bar{X}_n(1-\bar{X}_n)}} \xrightarrow[n\to+\infty]{p.s.} \frac{p - p_0}{\sqrt{p(1-p)}} \neq 0 \quad \Rightarrow \quad |Z_n| \xrightarrow[n\to+\infty]{p.s.} +\infty.$$

La statistique Z_n possède donc des valeurs proches de 0 sous $\mathcal{H}_0$ et des valeurs éloignées de 0 sous $\mathcal{H}_0$. On cherche donc une zone de rejet sous la forme $|Z_n| > x$. Or, comme Z_n converge en loi vers $\mathcal{N}(0,1)$ sous $\mathcal{H}_0$ alors on a :

$$\mathbb{P}_{p_0}\left(|Z_n| > z_{1-\alpha/2}\right) \underset{n\to+\infty}{\longrightarrow} \alpha.$$

Le test $\mathbb{1}_{|Z_n|>z_{1-\alpha/2}}$ est donc asymptotiquement de risque α.

Remarque: Ce raisonnement peut être adapté pour obtenir des tests pour n'importe quelle famille paramétrique dont on possède un estimateur asymptotiquement normal.

5 Test de comparaison de moyennes

On considère maintenant deux ensembles de variables : des variables i.i.d. $X_1, \cdots, X_n$ de loi $\mathcal{N}(\mu, \sigma^2)$ et des variables i.i.d. $Y_1, \cdots, Y_m$ de loi $\mathcal{N}(\mu', \sigma'^2)$. On souhaite tester s'il y a une différence entre les moyennes des jeux de données ou pas. Les hypothèses considérées pour un test bilatéral sont alors

$$\mathcal{H}_0 : \mu = \mu' \text{ contre } \mathcal{H}_1 : \mu \neq \mu'.$$

Remarque: On va juste faire les calculs pour le test bilatéral mais ce qu'on va faire peut s'adapter au cas unilatéral.

On considère trois cas différents :

a Le cas apparié

On suppose que $n = m$ et que $X_i - Y_i$ suit une loi $\mathcal{N}(\mu - \mu', \tilde{\sigma}^2)$. On se ramène alors au cas où on teste si la moyenne d'une loi normale est nulle. La statistique de test est donc

$$T = \sqrt{n}\frac{\bar{X}_n - \bar{Y}_n}{\sqrt{\widetilde{S}_n^2}}$$

où $\widetilde{S}_n^2$ est la variance empirique corrigée des $X_i - Y_i$. On rejette alors $\mathcal{H}_0$ au risque $\alpha \in]0,1[$ si $|T| > t_{1-\alpha/2}^{(n-1)}$. C'est ce qu'on appelle le **test de Student apparié**.

b Le cas indépendant à variance égale

On ne suppose pas que $n = m$ mais on suppose que les X_i sont indépendant des Y_j et que $\sigma = \sigma'$. On pose $\widetilde{S}_{n,X}$ et $\widetilde{S}_{m,Y}$ la variance empirique corrigée des X_i et des Y_i et on considère la statistique

$$T = \sqrt{n+m-2}\frac{\bar{X}_n - \bar{Y}_m}{\sqrt{\frac{1}{n}+\frac{1}{m}}\sqrt{(n-1)\widetilde{S}_{n,X}^2 + (m-1)\widetilde{S}_{m,Y}^2}}.$$

Proposition 59

Sous $\mathcal{H}_0$ on a $T \sim \mathcal{T}(m+n-2)$.

Démonstration : Sous $\mathcal{H}_0$ on a $\mu = \mu'$ donc $\bar{X}_n \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$ et $\bar{Y}_m \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{m}\right)$ donc par indépendance de $\bar{X}_n$ et $\bar{Y}_m$ on en déduit que

$$\bar{X}_n - \bar{Y}_m \sim \mathcal{N}\left(0, \sigma^2\left(\frac{1}{n}+\frac{1}{m}\right)\right)$$

d'où

$$\frac{\bar{X}_n - \bar{Y}_m}{\sigma\sqrt{\frac{1}{n}+\frac{1}{m}}} \sim \mathcal{N}(0,1).$$

De plus, $(n-1)\frac{\widetilde{S}^2_{n,X}}{\sigma^2} \sim \chi^2(n-1)$ et $(m-1)\frac{\widetilde{S}^2_{m,Y}}{\sigma^2} \sim \chi^2(m-1)$. Par indépendance de $\widetilde{S}^2_{n,X}$ et $\widetilde{S}^2_{m,Y}$ on en déduit que

$$\frac{(n-1)\widetilde{S}^2_{n,X} + (m-1)\widetilde{S}^2_{m,Y}}{\sigma^2} \sim \chi^2(n+m-2)$$

De plus, on a vu dans le cours que $\widetilde{S}^2_{n,X}$ est indépendant de $\bar{X}_n$ et $\widetilde{S}^2_{m,Y}$ est indépendant de $\bar{Y}_m$. Comme les X_i et les Y_i sont indépendants alors $\widetilde{S}^2_{n,X}$ est aussi indépendant de $\bar{Y}_m$ et $\widetilde{S}^2_{m,Y}$ est aussi indépendant de $\bar{X}_n$ donc

$$\frac{(n-1)\widetilde{S}^2_{n,X} + (m-1)\widetilde{S}^2_{m,Y}}{\sigma^2} \perp\!\!\!\perp \frac{\bar{X}_n - \bar{Y}_m}{\sigma\sqrt{\frac{1}{n}+\frac{1}{m}}}$$

On peut donc écrire

$$\sqrt{n+m-2}\frac{\frac{\bar{X}_n - \bar{Y}_m}{\sigma\sqrt{\frac{1}{n}+\frac{1}{m}}}}{\sqrt{\frac{(n-1)\widetilde{S}^2_{n,X} + (m-1)\widetilde{S}^2_{m,Y}}{\sigma^2}}} \sim \mathcal{T}(n+m-2)$$

et après simplifications on obtient bien

$$\sqrt{n+m-2}\frac{\bar{X}_n - \bar{Y}_m}{\sqrt{\frac{1}{n}+\frac{1}{m}}\sqrt{(n-1)\widetilde{S}^2_{n,X} + (m-1)\widetilde{S}^2_{m,Y}}} \sim \mathcal{T}(n+m-2).$$

■

Sous $\mathcal{H}_1$ la statistique T va prendre des valeurs éloignées de 0. On va donc rejeter $\mathcal{H}_0$ au risque $\alpha \in]0,1[$ lorsque $|T| > t^{(n+m-2)}_{1-\alpha/2}$. C'est ce qu'on appelle le **test de Student non apparié.**

c Le cas indépendant à variance inégale

On suppose toujours que les X_i sont indépendant des Y_j mais maintenant on ne suppose pas que $\sigma^2 = \sigma'^2$. Les calculs étant beaucoup plus compliqués on ne va pas les détailler. On va se contenter de donner le test habituellement utilisé dans ce cas-là appelé le **test de Welch**. La statistique utilisée pour le test est

$$T = \frac{\bar{X}_n - \bar{Y}_m}{\sqrt{\frac{\widetilde{S}^2_{n,X}}{n} + \frac{\widetilde{S}^2_{m,Y}}{m}}}.$$

La loi de T sous $\mathcal{H}_0$ est approximée par une loi $t(\nu)$ de degré de liberté

$$\nu = \frac{\left(\frac{\widetilde{S}^2_{n,X}}{n} + \frac{\widetilde{S}^2_{m,Y}}{m}\right)^2}{\frac{\widetilde{S}^4_{n,X}}{n^2(n-1)} + \frac{\widetilde{S}^4_{m,Y}}{m^2(m-1)}}.$$

⚠ Le degré de liberté ν n'est pas forcement un entier. On avait juste défini les lois de Student $t(n)$ pour des degré de liberté n entier mais on peut remarquer que leur densité ne nécessite pas avoir obligatoirement n entier pour être bien définie. De cette façon on peut généraliser les lois de Student au cas où le degré de liberté n'est pas un entier.

Comme sous $\mathcal{H}_1$ la statistique T va prendre des valeurs éloignées de 0. On va donc rejeter $\mathcal{H}_0$ asymptotiquement au risque $\alpha \in]0,1[$ lorsque $|T| > t^{(\nu)}_{1-\alpha/2}$.

6 Tests du χ^2

a Test d'adéquation à une loi discrète

On se place maintenant dans un cadre non-paramétrique. On considère que l'on observe des données $x_1, \cdots, x_n$ à valeur dans un ensemble fini $\Omega = \{\omega_1, \cdots, \omega_k\}$. On suppose les données issues de v.a.i.i.d. $X_1, \cdots, X_n$ mais on ne fait pas d'hypothèse sur la loi des X_i que l'on note $\mathbb{P}$. On souhaite tester si les données observées sont issues d'une certaine loi $\mathbb{P}_0$. Les hypothèses sont donc

$$\mathcal{H}_0 : \mathbb{P} = \mathbb{P}_0 \text{ contre } \mathcal{H}_1 : \mathbb{P} \neq \mathbb{P}_0.$$

Exemple: On cherche à savoir si un dé est truqué ou pas. On note le résultat de 200 lancers du dé dans le tableau suivant :

Valeur du dé	1	2	3	4	5	6
Nombre de fois tombé sur la valeur	38	30	40	29	36	27

Ici $\Omega = \{1, 2, 3, 4, 5, 6\}$ et $\mathbb{P}_0$ est la loi uniforme sur Ω. Sous $\mathcal{H}_0$, on s'attend à ce que chaque dé soit tiré $\approx 200 \times \frac{1}{6} \approx 33.3$ fois. On a donc envie de rejeter $\mathcal{H}_0$ si le nombre de fois que chaque valeur est tiré n'est pas assez proche de la valeur théorique de 33.3.

Théorème 60

Soit $p_i = \mathbb{P}_0(X_1 = \omega_i)$ et $N_i = \sum_{j=1}^n \mathbb{1}_{X_j=\omega_i} = card\{j : X_j = \omega_i\}$ le nombre de fois que ω_i apparaît parmi les X_i. On définit la statistique

$$\chi_n^2 = \sum_{i=1}^k \frac{(N_i - np_i)^2}{np_i},$$

*appelée **statistique du** χ^2.*

- *Si $\mathbb{P} = \mathbb{P}_0$, alors*
$$\chi_n^2 \xrightarrow[n\to+\infty]{\mathcal{L}} \chi^2(k-1).$$
- *Si $\mathbb{P} \neq \mathbb{P}_0$ alors*
$$\chi_n^2 \xrightarrow[n\to+\infty]{p.s.} +\infty$$

Démonstration :

- Si $\mathbb{P} \neq \mathbb{P}_0$ alors il existe un i tel que $\mathbb{P}(X_1 = \omega_i) = p_i' \neq p_i$. On a donc
$$\chi_n^2 \geqslant \frac{(N_i - np_i)^2}{np_i} = n\frac{(\frac{N_i}{n} - p_i)^2}{p_i}.$$
Or, d'après la loi des grands nombres on a $\frac{N_i}{n} \xrightarrow[n\to+\infty]{p.s.} p_i' \neq p_i$ et donc $\chi_n^2 \xrightarrow[n\to+\infty]{p.s.} +\infty$.
- Si $\mathbb{P} = \mathbb{P}_0$, on se contente de faire la preuve dans le cas où $k = 2$. Dans ce cas-là, on a $N_1 + N_2 = n$ et $p_1 + p_2 = 1$ donc
$$\begin{aligned}\chi_n^2 = \frac{(N_1 - np_1)^2}{np_1} + \frac{(n - N_1 - n(1-p_1))^2}{n(1-p_1)} &= \frac{(N_1 - np_1)^2}{np_1} + \frac{(N_1 - np_1)^2}{n(1-p_1)} \\ &= \frac{(N_1 - np_1)^2}{np_1(1-p_1)} = \left(\sqrt{n}\frac{\frac{N_1}{n} - p_1}{\sqrt{p_1(1-p_1)}}\right)^2.\end{aligned}$$
Or, le TCL nous dit que
$$\sqrt{n}\frac{\frac{N_1}{n} - p_1}{\sqrt{p_1(1-p_1)}} \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}(0,1)$$

donc $\chi_n^2 \xrightarrow[n\to+\infty]{\mathcal{L}} \chi^2(1)$. ■

Sous $\mathcal{H}_0$, la loi de la statistique χ_n^2 est proche d'une loi $\chi^2(k-1)$ alors que sous $\mathcal{H}_1$ cette statistique prends de grandes valeurs positives. Pour un test asymptotiquement de risque α on cherche donc une zone de rejet de la forme $\chi_n^2 > x$. En prenant $x = \chi_{1-\alpha}^{(k-1)}$ on a bien que sous $\mathcal{H}_0$:

$$\mathbb{P}_0(\chi_n^2 > \chi_{1-\alpha}^{(k-1)}) \underset{n\to+\infty}{\longrightarrow} \alpha.$$

Le test s'écrit donc

$$\mathbb{1}_{\chi_n^2 > \chi_{1-\alpha}^{(k-1)}}.$$

Calcul de la p-valeur : le test rejette $\mathcal{H}_0$ ssi $\chi_n^2 > F_{\chi^2(k-1)}^{-1}(1-\alpha)$. Par stricte croissance de $F_{\chi^2(k-1)}$, cette condition est équivalente à

$$\alpha > 1 - F_{\chi^2(k-1)}(\chi_n^2).$$

La p-valeur est donc égale à $1 - F_{\chi^2(k-1)}(\chi_n^2)$.
Exemple: Dans notre exemple de tirage de dé, on a :

Valeur du dé	1	2	3	4	5	6
N_i (Valeurs observées)	38	30	40	29	36	27
$200 \times p_i$ (Valeurs théoriques)	33.3	33.3	33.3	33.3	33.3	33.3

Ici, la statistique de test est :

$$\chi_n^2 = \frac{(38-\frac{200}{6})^2}{\frac{200}{6}} + \frac{(30-\frac{200}{6})^2}{\frac{200}{6}} + \frac{(40-\frac{200}{6})^2}{\frac{200}{6}} + \frac{(29-\frac{200}{6})^2}{\frac{200}{6}} + \frac{(36-\frac{200}{6})^2}{\frac{200}{6}} + \frac{(27-\frac{200}{6})^2}{\frac{200}{6}} \approx 4.3$$

et le quantile d'ordre 0.95 de la loi $\chi^2(5)$ est ≈ 11.07 donc on ne rejette pas l'hypothèse que le dé est bien équilibré.

Remarque: On peut écrire la statistique du χ^2 sous la forme

$$\chi_n^2 = \sum \frac{(\text{Observation} - \text{Théorique})^2}{\text{Théorique}}.$$

b Test d'indépendance

On considère que l'on observe une réalisation de n couples de données (x_i, y_i) issues de variables aléatoires discrètes (X_i, Y_i) avec indépendance entre les couples (X_i, Y_i). On note $\Omega = \{\omega_1, \cdots, \omega_k\}$ l'ensemble des k valeurs possibles pour X et $\Omega' = \{\omega'_1, \cdots, \omega'_l\}$ l'ensemble des l valeurs possibles pour Y. On note $\mathbb{P}_X$ la loi commune des X_i et $\mathbb{P}_Y$ la loi commune des Y_i. On cherche alors à tester l'indépendance entre $\mathbb{P}_X$ et $\mathbb{P}_Y$:

$$\mathcal{H}_0 : \mathbb{P}_X \perp\!\!\!\perp \mathbb{P}_Y \quad \text{contre} \quad \mathcal{H}_1 : \mathbb{P}_X \not\perp\!\!\!\perp \mathbb{P}_Y$$

Exemple: Une étude de 2010 (lien) s'intéresse à déterminer si le groupe sanguin influence la gravité du cancer du pancréas. Pour cela, ils ont rapporté les chiffre suivants du groupe sanguin et du stade d'avancement du cancer de 1271 patients atteint de cancer du pancréas :

	A	B	O	AB
Stade I/II (non avancé)	125	169	208	52
Stade III/IV (avancé)	261	180	182	94

La proportion de patients avec le groupe A et un stade non avancé est de $\frac{125}{1271}$. La proportion de patients avec juste le groupe A est de $\frac{125+261}{1271} = \frac{386}{1271}$. La proportion de patients avec juste le stade non avancé est de $\frac{125+169+208+52}{1271} = \frac{554}{1271}$. S'il y a indépendance entre le groupe sanguin et la gravité du cancer du pancréas alors on doit avoir

$$\mathbb{P}(\text{groupe A et stade non avancé}) = \mathbb{P}(\text{groupe A})\mathbb{P}(\text{stade non avancé})$$

et donc la proportion d'individu de groupe A et avec un stade non avancé (ici $\frac{125}{1271}$) devrait être similaire au produit des proportions des individus de groupe A (ici $\frac{386}{1271}$) et la proportion des individus avec un stade non avancé (ici $\frac{554}{1271}$). On devrait donc observer que $\frac{125}{1271} \approx \frac{386}{1271} \times \frac{554}{1271}$ s'il y a indépendance et quelque chose similaire pour toutes les autres paires de modalités. L'idée du test du χ^2 d'indépendance est donc de comparer ces quantités.

Théorème 61

Soit $N_{i,\bullet} = \sum_{k=1}^n \mathbb{1}_{X_k=\omega_i}$, le nombre de fois que ω_i apparaît parmi les X_k. Soit $N_{\bullet,j} = \sum_{k=1}^n \mathbb{1}_{Y_k=\omega'_j}$, le nombre de fois que ω'_j apparaît parmi les Y_k. Soit $N_{i,j} = \sum_{k=1}^n \mathbb{1}_{X_k=\omega_i}\mathbb{1}_{Y_k=\omega'_j}$, le nombre de fois que le couple (ω_i, ω'_j) apparaît parmi les couples (X_k, Y_k). On définit la statistique

$$\chi_n^2 = \sum_{i=1}^k \sum_{j=1}^l \frac{(N_{i,j} - N_{i,\bullet}N_{\bullet,j}/n)^2}{N_{i,\bullet}N_{\bullet,j}/n},$$

aussi appelée ***statistique du*** *χ^2.*

- *Si $\mathbb{P} \perp\!\!\!\perp \mathbb{P}'$ alors*
$$\chi_n^2 \xrightarrow[n\to+\infty]{\mathcal{L}} \chi^2\big((k-1)(l-1)\big).$$
- *Si $\mathbb{P} \not\perp\!\!\!\perp \mathbb{P}'$ alors*
$$\chi_n^2 \xrightarrow[n\to+\infty]{p.s.} \infty.$$

Démonstration : ADMIS ■

Sous $\mathcal{H}_0$, la loi de la statistique χ_n^2 est proche d'une loi $\chi^2\big((k-1)(l-1)\big)$ alors que sous $\mathcal{H}_1$, cette statistique prends de grande valeurs. Le test du χ^2 d'indépendance s'écrit alors

$$\mathbb{1}_{\chi_n^2 > \chi_{1-\alpha}^{((k-1)(l-1))}}.$$

C'est un test asymptotiquement de risque α

Calcul de la p-valeur : le test rejette $\mathcal{H}_0$ ssi $\chi_n^2 < F^{-1}_{\chi^2((k-1)(l-1))}(1-\alpha)$. Par stricte croissance de $F_{\chi^2((k-1)(l-1))}$, cette condition est équivalente à

$$\alpha < 1 - F_{\chi^2((k-1)(l-1))}(\chi_n^2).$$

La p-valeur est donc égale à $1 - F_{\chi^2((k-1)(l-1))}(\chi_n^2)$.

Exemple: On reprend les données :

	A	B	O	AB	Total
Stade I/II (non-avancé)	125	169	208	52	554
Stade III/IV (avancé)	261	180	182	94	717
Total	386	349	390	146	1271

Le tableau des valeurs de $N_{i,\bullet}N_{\bullet,j}/n$ est donc

	A	B	O	AB	Total
Stade I/II (non-avancé)	168.2	152.1	170.0	63.6	554
Stade III/IV (avancé)	217.8	196.9	220.0	82.4	717
Total	386	349	390	146	1271

La statistique de test est alors :

$$\chi_n^2 = \frac{(125 - 168.2)^2}{168.2} + \cdots + \frac{(94 - 82.4)^2}{82.4} \approx 13.96$$

et le quantile d'ordre 0.95 de la loi $\chi^2(3)$ est ≈ 7.8 donc on rejette l'hypothèse qu'il y a indépendance entre le groupe sanguin d'une personne atteint du cancer du pancréas et la gravité de la maladie.

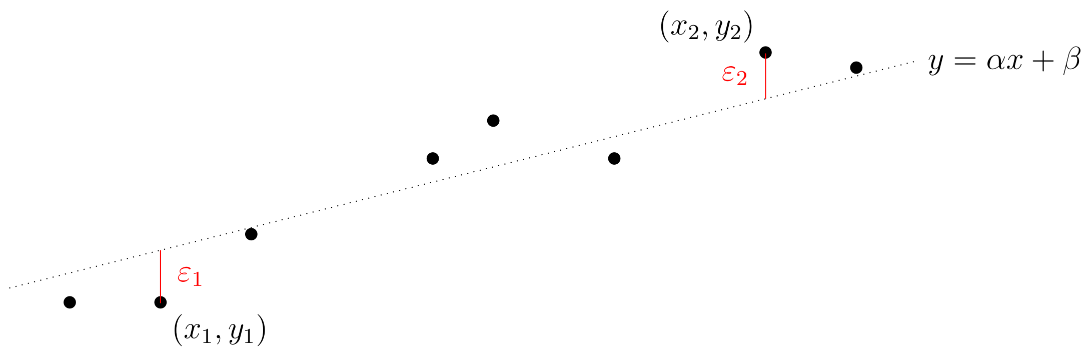
Remarque: Ici aussi on peut écrire la statistique du χ^2 sous la forme

$$\chi_n^2 = \sum \frac{(\text{Observation} - \text{Théorique})^2}{\text{Théorique}}.$$

Chapitre VI

Régression linéaire

<u>Cadre :</u> On considère que l'on observe des couples de données $(x_1, y_1), \cdots, (x_n, y_n)$ où les x_i ne sont pas tous égaux et les $y_1, \cdots, y_n$ sont des réalisations de variables aléatoires $Y_1, \cdots, Y_n$ telles que $Y_i = \alpha x_i + \beta + \varepsilon_i$ avec $\varepsilon_1, \cdots, \varepsilon_n$ des v.a.i.i.d. de loi $\mathcal{N}(0, \sigma^2)$. La variable $x = (x_1, \cdots, x_n)$ est appelée la **variable expliquative** et la variable $y = (y_1, \cdots, y_n)$ est appelée la **variable d'intérêt**.



Ici, le modèle possède 3 paramètres : α, β et σ^2. De plus on a $Y_i \sim \mathcal{N}(\alpha x_i + \beta, \sigma^2)$ donc, comparé aux chapitres précédents, les Y_i ne sont pas tous de même loi.

<u>*Remarque*</u>: Si $\alpha = 0$ alors $Y_i \sim \mathcal{N}(\beta, \sigma^2)$ et on se retrouve dans le problème de l'estimation des paramètres d'une loi normale.

Pour simplifier les notations par la suite, on pose

$$\bar{x}_n = \frac{1}{n}\sum_{i=1}^n x_i, \quad \overline{x_n^2} = \frac{1}{n}\sum_{i=1}^n x_i^2 \quad \text{et} \quad \mathrm{var}(x) = \overline{x_n^2} - \bar{x}_n^2 = \frac{1}{n}\sum_{i=1}^n (x_i - \bar{x}_n)^2.$$

1 Estimation par maximum de vraisemblance

Proposition 62

L'estimateur du maximum de vraisemblance $(\hat{\alpha}, \hat{\beta}, \hat{\sigma})$ de (α, β, σ) existe, est unique et vérifie

$$\begin{cases} \hat{\alpha}_n = \dfrac{\frac{1}{n}\sum_{i=1}^n x_i Y_i - \bar{x}_n \left(\frac{1}{n}\sum_{i=1}^n Y_i\right)}{\mathrm{var}(x)} \\ \hat{\beta}_n = \overline{Y_n} - \hat{\alpha}_n \overline{x_n} \\ \hat{\sigma}_n = \sqrt{\dfrac{1}{n}\displaystyle\sum_{i=1}^n (Y_i - \hat{\alpha}_n x_i - \hat{\beta}_n)^2} \end{cases}$$

Démonstration : On a $Y_i \sim \mathcal{N}(\alpha x_i + \beta, \sigma^2)$ pour tout i. On note $f^{(i)}_{\alpha,\beta,\sigma}$ la densité de probabilité de Y_i. La vraisemblance du modèle s'écrit alors

$$L(\alpha,\beta,\sigma|Y) = \prod_i f^{(i)}_{\alpha,\beta,\sigma}(Y_i) = \prod_i \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(Y_i-\alpha x_i-\beta)^2}{2\sigma^2}} = (2\pi\sigma^2)^{-n/2} e^{-\sum_{i=1}^n \frac{(Y_i-\alpha x_i-\beta)^2}{2\sigma^2}}.$$

La log-vraisemblance du modèle est donc

$$\mathcal{L}(\alpha,\beta,\sigma) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^n (Y_i - \alpha x_i - \beta)^2.$$

A σ fixé, maximiser la vraisemblance revient donc à trouver a et b qui minimise la fonction $g(\alpha,\beta) = \sum_{i=1}^n (Y_i - \alpha x_i - b)^2$. Cette fonction est deux fois différentiable de gradient

$$\nabla g(\alpha,\beta) = \begin{pmatrix} \frac{\partial g}{\partial \alpha} \\ \frac{\partial g}{\partial \beta} \end{pmatrix} = \begin{pmatrix} -2\sum_{i=1}^n x_i(Y_i - \alpha x_i - \beta) \\ -2\sum_{i=1}^n (Y_i - \alpha x_i - \beta) \end{pmatrix},$$

et de Hessienne

$$H(g)(\alpha,\beta) = \begin{pmatrix} \frac{\partial^2 g}{\partial \alpha^2} & \frac{\partial^2 g}{\partial \alpha \partial \beta} \\ \frac{\partial^2 g}{\partial \beta \partial \alpha} & \frac{\partial^2 g}{\partial \beta^2} \end{pmatrix} = \begin{pmatrix} 2\sum_{i=1}^n x_i^2 & 2\sum_{i=1}^n x_i \\ 2\sum_{i=1}^n x_i & 2n \end{pmatrix}.$$

Quel que soit α et β, on a $\mathrm{Tr}(H(g)(\alpha,\beta)) > 0$ et $\det(H(g)(\alpha,\beta)) = 4n\sum_{i=1}^n x_i^2 - 4\left(\sum_{i=1}^n x_i\right)^2 = 4n^2\mathrm{var}(x) > 0$ car les x_i ne sont pas tous constant. $H(g)$ possède alors tout le temps deux valeurs propres strictement positives et est donc définie positive. La fonction g est donc strictement convexe et possède un unique minimum au point $(\hat{\alpha}_n, \hat{\beta}_n)$ qui annule son gradient, c'est à dire :

$$\begin{cases} \sum_{i=1}^n x_i(Y_i - \hat{\alpha}_n x_i - \hat{\beta}_n) = 0 \\ \sum_{i=1}^n (Y_i - \hat{\alpha}_n x_i - \hat{\beta}_n) = 0 \end{cases} \Leftrightarrow \begin{cases} \sum_{i=1}^n x_i Y_i - \hat{\alpha}_n \sum_{i=1}^n x_i^2 - \hat{\beta}_n \sum_{i=1}^n x_i = 0 \\ \sum_{i=1}^n Y_i - \hat{\alpha}_n \sum_{i=1}^n x_i - \hat{\beta}_n n = 0 \end{cases}$$

On tombe sur un système linéaire de deux équations à deux inconnus dont la solution est

$$\hat{\alpha}_n = \frac{\frac{1}{n}\sum_{i=1}^n x_i Y_i - \left(\frac{1}{n}\sum_{i=1}^n x_i\right)\left(\frac{1}{n}\sum_{i=1}^n Y_i\right)}{\frac{1}{n}\sum_{i=1}^n x_i^2 - \left(\frac{1}{n}\sum_{i=1}^n x_i\right)^2} \quad \text{et} \quad \hat{\beta}_n = \overline{Y_n} - \hat{\alpha}_n \overline{x_n}.$$

Il ne reste plus qu'à trouver $\hat{\sigma}_n$. Pour ça, on calcule

$$\frac{\partial \mathcal{L}(\hat{\alpha}_n, \hat{\beta}_n, \sigma)}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^3}\sum_{i=1}^n (Y_i - \hat{\alpha}_n x_i - \hat{\beta}_n)^2.$$

On a donc

$$\begin{aligned} &\frac{\partial \mathcal{L}(\hat{\alpha}_n, \hat{\beta}_n, \hat{\sigma}_n)}{\partial \sigma} = 0 \\ \Leftrightarrow\ & -\frac{n}{\hat{\sigma}_n} + \frac{1}{\hat{\sigma}_n^3}\sum_{i=1}^n (Y_i - \hat{\alpha}_n x_i - \hat{\beta}_n)^2 = 0 \\ \Leftrightarrow\ & -n\hat{\sigma}_n^2 + \sum_{i=1}^n (Y_i - \hat{\alpha}_n x_i - \hat{\beta}_n)^2 = 0 \\ \Leftrightarrow\ & \hat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^n (Y_i - \hat{\alpha}_n x_i - \hat{\beta}_n)^2 \end{aligned}$$

De plus, avec le même raisonnement on en déduit que $\frac{\partial \mathcal{L}(\hat{\alpha}_n, \hat{\beta}_n, \hat{\sigma}_n)}{\partial \sigma} < 0$ quand $\hat{\sigma}_n^2 > \frac{1}{n}\sum_{i=1}^n (Y_i - \hat{\alpha}_n x_i - \hat{\beta}_n)^2$ et $\frac{\partial \mathcal{L}(\hat{\alpha}_n, \hat{\beta}_n, \hat{\sigma}_n)}{\partial \sigma} > 0$ quand $\hat{\sigma}_n^2 < \frac{1}{n}\sum_{i=1}^n (Y_i - \hat{\alpha}_n x_i - \hat{\beta}_n)^2$ ce qui prouve que $\hat{\sigma}_n = \sqrt{\frac{1}{n}\sum_{i=1}^n (Y_i - \hat{\alpha}_n x_i - \hat{\beta}_n)^2}$ est l'unique maximum global de $\sigma \mapsto \mathcal{L}(\hat{\alpha}_n, \hat{\beta}_n, \sigma)$ et donc l'estimateur du maximum de vraisemblance de σ. ■

2 Loi des estimateurs

Comme pour l'estimation des paramètres d'une loi normale on peut déterminer la loi exacte des estimateurs.

Proposition 63

L'estimateur $\hat{\alpha}_n$ vérifie

$$\hat{\alpha}_n \sim \mathcal{N}\left(\alpha, \frac{\sigma^2}{n\text{var}(x)}\right).$$

En particulier, $\hat{\alpha}_n$ est non biaisé et

$$R(\hat{\alpha}_n, \alpha) = \frac{\sigma^2}{n\text{var}(x)}.$$

Démonstration : On rappelle que

$$\hat{\alpha}_n = \frac{\frac{1}{n}\sum_{i=1}^n x_i Y_i - \overline{x}_n \left(\frac{1}{n}\sum_{i=1}^n Y_i\right)}{\text{var}(x)} \quad \text{et} \quad Y_i \sim \mathcal{N}(\alpha x_i + \beta, \sigma^2).$$

On a donc que $\hat{a}_n$ suit une loi normale par combinaison linéaire des Y_i qui sont des loi normales indépendantes. On cherche alors les paramètres de cette loi. Tout d'abord, son espérance :

$$\begin{aligned}
\mathbb{E}[\hat{\alpha}_n] &= \frac{\frac{1}{n}\sum_{i=1}^n x_i \mathbb{E}[Y_i] - \overline{x}_n \left(\frac{1}{n}\sum_{i=1}^n \mathbb{E}[Y_i]\right)}{\text{var}(x)} \\
&= \frac{\frac{1}{n}\sum_{i=1}^n (\alpha x_i^2 + \beta x_i) - \overline{x}_n \left(\frac{1}{n}\sum_{i=1}^n \alpha x_i + \beta\right)}{\text{var}(x)} \\
&= \frac{\alpha \overline{x_n^2} + \beta \overline{x}_n - \alpha \overline{x}_n^2 - \beta \overline{x}_n}{\text{var}(x)} \\
&= \alpha.
\end{aligned}$$

En particulier, $\hat{\alpha}_n$ est non biaisé. Maintenant, cherchons sa variance :

$$\begin{aligned}
\text{Var}(\hat{\alpha}_n) &= \frac{\text{Var}\left(\frac{1}{n}\sum_{i=1}^n x_i Y_i - \overline{x_n}\left(\frac{1}{n}\sum_{i=1}^n Y_i\right)\right)}{\text{var}(x)^2} \\
&= \frac{\text{Var}\left(\frac{1}{n}\sum_{i=1}^n (x_i - \overline{x_n}) Y_i\right)}{\text{var}(x)^2} \\
&= \frac{\frac{1}{n^2}\sum_{i=1}^n (x_i - \overline{x_n})^2 \text{Var}(Y_i)}{\text{var}(x)^2} \\
&= \frac{\frac{\sigma^2}{n} \times \frac{1}{n}\sum_{i=1}^n (x_i - \overline{x_n})^2}{\text{var}(x)^2} \\
&= \frac{\sigma^2}{n\text{var}(x)}.
\end{aligned}$$

■

Remarques:

- $R(\hat{\alpha}_n, \alpha)$ est inversement proportionnelle à la variance des x_i. Cela montre que plus les x_i sont écartés les uns des autres, plus l'estimation est bonne.
- On peut contrôler $\hat{\alpha}_n - \alpha$ avec la formule

$$\sqrt{n\text{var}(x)}\frac{\hat{\alpha}_n - \alpha}{\sigma} \sim \mathcal{N}(0, 1).$$

Proposition 64

L'estimateur $\hat{\beta}_n$ vérifie

$$\hat{\beta}_n \sim \mathcal{N}\left(\beta, \frac{\sigma^2 \overline{x_n^2}}{n\mathrm{var}(x)}\right).$$

En particulier, $\hat{\beta}_n$ est non biaisé et

$$R(\hat{\beta}_n, \beta) = \frac{\sigma^2 \overline{x_n^2}}{n\mathrm{var}(x)}.$$

Démonstration : $\hat{\beta}_n$ suit aussi une loi normale vu qu'il peut s'écrire comme une combinaison linéaire des Y_i. De plus,

$$\mathbb{E}[\hat{\beta}_n] = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^n Y_i\right] - \mathbb{E}[\hat{\alpha}_n]\overline{x_n} = \frac{1}{n}\sum_{i=1}^n (\alpha x_i + \beta) - \alpha \overline{x_n} = \beta,$$

donc $\hat{\beta}_n$ est aussi non-biaisé. Pour la variance, on a

$$\begin{aligned}
\mathrm{Var}(\hat{\beta}_n) &= \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^n Y_i - \frac{\overline{x_n}}{n\mathrm{var}(x)}\sum_{i=1}^n (x_i - \overline{x_n})Y_i\right) \\
&= \mathrm{Var}\left(\frac{1}{n\mathrm{var}(x)}\sum_{i=1}^n (\mathrm{var}(x) - \overline{x_n}(x_i - \overline{x_n}))Y_i\right) \\
&= \frac{1}{n^2\mathrm{var}(x)^2}\sum_{i=1}^n (\overline{x_n^2} - \overline{x_n}x_i)^2 \mathrm{Var}(Y_i)
\end{aligned}$$

Or,

$$\begin{aligned}
\sum_{i=1}^n (\overline{x_n^2} - \overline{x_n}x_i)^2 \mathrm{Var}(Y_i) &= \sigma^2 \sum_{i=1}^n (\overline{x_n^2}^2 - 2\overline{x_n^2}\,\overline{x_n}x_i + \overline{x_n}^2 x_i^2) \\
&= \sigma^2 (n\overline{x_n^2}^2 - 2n\overline{x_n^2}\,\overline{x_n}^2 + n\overline{x_n}^2\overline{x_n^2}) \\
&= n\sigma^2 \overline{x_n^2}\mathrm{var}(x)
\end{aligned}$$

donc

$$\mathrm{Var}(\hat{\beta}_n) = \frac{\sigma^2 \overline{x_n^2}}{n\mathrm{var}(x)}.$$

■

On termine par mentionner les résultats suivants sans les prouver :

Proposition 65

La statistique

$$\hat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^n (Y_i - \hat{\alpha}_n x_i - \hat{\beta}_n)^2$$

est un estimateur biaisé de σ^2. Pour une estimation non-biaisée, on prend

$$\widetilde{S}_n^2 = \frac{1}{n-2}\sum_{i=1}^n (Y_i - \hat{\alpha}_n x_i - \hat{\beta}_n)^2.$$

De plus,

$$\widetilde{S}_n^2 \perp\!\!\!\perp \hat{\alpha}_n \text{ et } \widetilde{S}_n^2 \perp\!\!\!\perp \hat{\beta}_n;$$

$$(n-2)\frac{\widetilde{S}_n^2}{\sigma^2} \sim \chi^2(n-2);$$

$$\sqrt{n\text{var}(x)}\frac{\hat{\alpha}_n - \alpha}{\sqrt{\widetilde{S}_n^2}} \sim \mathcal{T}(n-2);$$

$$\sqrt{n\frac{\text{var}(x)}{\overline{x_n^2}}}\frac{\hat{\beta}_n - \beta}{\sqrt{\widetilde{S}_n^2}} \sim \mathcal{T}(n-2).$$

A partir de ces résultats on peut construire des intervalles de confiance pour $\hat{\alpha}_n$, $\hat{\beta}_n$ et $\hat{\sigma}_n$ ainsi que pour créer des tests d'hypothèses sur ces paramètres.

Annexe A

Preuve de la consistance et la normalité asymptotique de l'EMV

Remarque: Cette preuve est inspirée du polycopié de cours de M2 d'estimation paramétrique de Bernard Delyon.

On rappelle que l'on considère une famille $\mathcal{F} = \{\mathbb{P}_\theta, \theta \in \Theta\}$ de lois toutes discrètes ou continues, à valeur dans un ensemble Ω, avec $\theta \subset \mathbb{R}$. On rappelle que la vraisemblance est définie par

$$L(\theta|X_1, \cdots, X_n) = \prod_{i=1}^{n} f_\theta(X_i)$$

où f_θ est la densité de $\mathbb{P}_\theta$ si $\mathbb{P}_\theta$ est une loi continue et $f_\theta(x) = \mathbb{P}_\theta(X = x)$ si $\mathbb{P}_\theta$ est une loi discrète. L'estimateur du maximum de vraisemblance $\hat{\theta}_n$ vérifie alors

$$\hat{\theta}_n \in \arg\max_{\theta \in \Theta} L(\theta|X_1, \cdots, X_n),$$

c'est à dire que $\hat{\theta}_n$ est choisit comme étant l'une des valeurs de θ (si elle existe) qui maximise la vraisemblance.

Théorème 66

Soit $\theta^* \in \Theta$. On suppose que la famille $\mathcal{F} = \{\mathbb{P}_\theta, \theta \in \Theta\}$ vérifie les hypothèses suivantes :

(i) $\mathcal{F}$ est identifiable.

(ii) Il existe une suite croissante de compacts Θ_p tels que $\Theta \subset \bigcup_{p \in \mathbb{N}} \mathrm{int}(\Theta_p)$.

(iii)

$$\mathbb{P}_{\theta^*}\left(\hat{\theta}_n \text{ est borné}\right) = 1.$$

(iv)

$$\mathbb{P}_{\theta^*}\left(\text{La fonction } \theta \mapsto \log(f_\theta(X)) \text{ est continue en } \theta^*\right) = 1.$$

(v) Pour tout $p \in \mathbb{N}$,

$$\mathbb{E}_{\theta^*}\left[\sup_{\theta \in \Theta_p} |\log(f_\theta(X))|\right] < \infty$$

Alors, l'EMV $\hat{\theta}_n$ vérifie

$$\mathbb{P}_{\theta^*}\left(\hat{\theta}_n \xrightarrow[n \to \infty]{} \theta^*\right) = 1$$

c'est à dire que $\hat{\theta}_n$ converge presque sûrement vers θ^* sous $\mathbb{P}_{\theta^*}$.

Démonstration : On commence par supposer que Θ est un compact. Dans ce cas-là le maximum de vraisemblance est tout le temps bien défini car toute fonction sur Θ admet au moins un

maximum.
Etape 1 : Convergence de la log-vraisemblance
On note $\mathbb{P}_{\theta^*}$ la loi commune des X_i avec $\theta^* \in \Theta$. La log-vraisemblance s'écrit

$$\mathcal{L}(\theta|X_1, \cdots, X_n) = \sum_{i=1}^{n} \log(f_\theta(X_i)).$$

Comme maximiser $\frac{1}{n}\mathcal{L}$ et maximiser L revient au même on en déduit que le maximum de vraisemblance $\hat{\theta}_n$ vérifie

$$\hat{\theta}_n \in \arg\max_{\theta\in\Theta} \frac{1}{n}\sum_{i=1}^{n} \log(f_\theta(X_i))$$

En appliquant la loi des grands nombres aux variables $\log(f_\theta(X_i))$ grâce à l'hypothèse (v) on obtient que pour tout $\theta \in \Theta$,

$$\frac{1}{n}\sum_{i=1}^{n} \log(f_\theta(X_i)) \xrightarrow[n\to+\infty]{p.s.} \mathbb{E}_{\theta^*}[\log(f_\theta(X))].$$

Si on note $h_n(\theta) = \frac{1}{n}\sum_{i=1}^{n} \log(f_\theta(X_i))$ et $h(\theta) = \mathbb{E}_{\theta^*}[\log(f_\theta(X))]$ alors pour tout $\theta \in \Theta$, $h_n(\theta) \xrightarrow[n\to+\infty]{p.s.} h(\theta)$.

Etape 2 : θ^* est un maximum de $h(\theta)$
Soit $\theta \in \Theta$ alors

$$h(\theta^*) - h(\theta) = \mathbb{E}_{\theta^*}[\log(f_{\theta^*}(X))] - \mathbb{E}_{\theta^*}[\log(f_\theta(X))] = \mathbb{E}_{\theta^*}\left[\log\left(\frac{f_{\theta^*}(X)}{f_\theta(X)}\right)\right]$$

En utilisant l'inégalité

$$\forall x > 0, \log(x) \geqslant 1 - \frac{1}{x} \text{ avec égalité ssi } x = 1,$$

on obtient

$$h(\theta^*) - h(\theta) \geqslant 1 - \mathbb{E}_{\theta^*}\left[\frac{f_\theta(X)}{f_{\theta^*}(X)}\right].$$

Or, si $\mathcal{F}$ est une famille continue sur un espace Ω alors

$$\mathbb{E}_{\theta^*}\left[\frac{f_\theta(x)}{f_{\theta^*}(x)}\right] = \int_\Omega \frac{f_\theta(x)}{f_{\theta^*}(x)} f_{\theta^*}(x)\mathrm{d}x = \int_\Omega f_\theta(x)\mathrm{d}x = 1$$

et si $\mathcal{F}$ est une famille discrète sur un espace Ω alors

$$\mathbb{E}_{\theta^*}\left[\frac{f_\theta(x)}{f_{\theta^*}(x)}\right] = \sum_{\omega\in\Omega} \frac{f_\theta(\omega)}{f_{\theta^*}(\omega)}\mathbb{P}_{\theta^*}(X=\omega) = \sum_{\omega\in\Omega} \frac{f_\theta(\omega)}{f_{\theta^*}(\omega)} f_{\theta^*}(\omega) = \sum_{\omega\in\Omega} \mathbb{P}_\theta(X=\omega) = 1.$$

Dans tout les cas on a bien

$$h(\theta^*) - h(\theta) \geqslant 0.$$

Etape 3 : θ^* est l'unique maximum de $h(\theta)$:
En conséquence du raisonnement précédant, on a

$$h(\theta^*) - h(\theta) = 0 \;\Leftrightarrow\; \mathbb{E}_{\theta^*}\left[\log\left(\frac{f_{\theta^*}(X)}{f_\theta(X)}\right) - \left(1 + \frac{f_\theta(X)}{f_{\theta^*}(X)}\right)\right] = 0$$

Comme le terme à l'intérieur de l'espérance est strictement positif si $f_\theta(X) \neq f_{\theta^*}(X)$ et nul si $f_\theta(X) = f_{\theta^*}(X)$ alors

$$h(\theta^*) - h(\theta) = 0 \;\Leftrightarrow\; \mathbb{P}_{\theta^*}(f_{\theta^*}(X) = f_\theta(X)) = 1.$$

C'est une conséquence du lemme suivant :

Lemme 67

Soit X une variable aléatoire à valeur positive. Alors,

$$\mathbb{E}[X] = 0 \;\Leftrightarrow\; \mathbb{P}(X = 0) = 1$$

Démonstration : Si $\mathbb{P}(X > 0) = 0$ alors $\mathbb{P}(X = 0) = 1$ et forcement $\mathbb{E}[X] = 0$. A l'inverse, si $\mathbb{P}(X > 0) > 0$ alors

$$\mathbb{E}[X] = \mathbb{E}[X\mathbb{1}_{X=0}] + \mathbb{E}[X\mathbb{1}_{X>0}] = \mathbb{E}[X|X > 0]\mathbb{P}(X > 0) > 0,$$

ce qui prouve le lemme. ■

On en déduit alors que $h(\theta^*) - h(\theta) = 0$ si et seulement si $f_\theta(x) = f_{\theta^*}(x)$ $\mathbb{P}_{\theta^*}$-presque sûrement. Si on note $\Omega' = \{\omega \in \Omega |\ f_{\theta^*}(\omega) > 0\}$ alors $f_\theta(x) = f_{\theta^*}(x)$ pour tout $x \in \Omega'$. De plus, comme $\int_\Omega f_\theta(x)\mathrm{d}x = \int_\Omega f_{\theta^*}(x)\mathrm{d}x = 1$ dans le cas continu (ou $\sum_{\omega\in\Omega} f_\theta(\omega) = \sum_{\omega\in\Omega} f_{\theta^*}(\omega) = 1$ dans le cas discret) alors on en déduit que $f_\theta(x) = f_{\theta^*}(x) = 0$ pour tout $x \in \Omega\backslash\Omega'$ et donc $\mathbb{P}_\theta = \mathbb{P}_{\theta^*}$. Or, la famille $\mathcal{F}$ est identifiable par l'hypothèse (i) donc $\mathbb{P}_\theta = \mathbb{P}_{\theta^*}$ si et seulement si $\theta = \theta^*$ ce qui prouve que $\hat{\theta}^*$ est l'unique maximum de h.

Etape 4 : h est continue :
Soit $\theta_0 \in \Theta$. On pose la fonction

$$g_{\theta_0}(\eta) = \mathbb{E}_{\theta^*}\left[\sup_{|\theta-\theta_0|\leqslant\eta} |\log(f_\theta(X)) - \log(f_{\theta_0}(X))|\right].$$

On sait que $\theta \mapsto \log(f_\theta(x))$ est continue en θ_0 pour presque tout x par hypothèse (iv). De plus, $\sup_{|\theta-\theta_0|\leqslant\eta} |\log(f_\theta(X)) - \log(f_{\theta_0}(X))|$ est positif et décroissant par rapport à η alors, par la théorème de convergence monotone, on a

$$\lim_{\eta\to 0} g_{\theta_0}(\eta) = \mathbb{E}_{\theta^*}\left[\lim_{\eta\to 0}\sup_{|\theta-\theta_0|\leqslant\eta} |\log(f_\theta(X)) - \log(f_{\theta_0}(X))|\right] = 0.$$

On en déduit alors que

$$\sup_{|\theta-\theta_0|\leqslant\eta} |h(\theta) - h(\theta_0)| \leqslant \sup_{|\theta-\theta_0|\leqslant\eta} \mathbb{E}_{\theta^*}[|\log(f_\theta(X)) - \log(f_{\theta_0}(X))|] \leqslant g_{\theta_0}(\eta) \underset{\eta\to 0}{\longrightarrow} 0$$

et donc h est continu en θ_0.

Etape 5 : h_n converge uniformément presque sûrement vers h :
Soit $\varepsilon > 0$. Pour tout $\theta \in \Theta$ il existe $\eta_\theta > 0$ tel que $g_\theta(\eta_\theta) < \varepsilon$. Soit $B(\theta, \eta_\theta)$ la boule ouverte de centre θ et de rayon η_θ. Comme $B(\theta, \eta_\theta)$ contient θ alors $\Theta \subset \bigcup_{\theta\in\Theta} B(\theta, \eta_\theta)$. Par compacité de Θ on peut donc extraire un sous-recouvrement fini de sorte que $\exists N \in \mathbb{N}$ et $\exists \theta_1, \cdots, \theta_N \in \Theta$ tels que $\Theta \subset \bigcup_{j=1}^N B(\theta_j, \eta_{\theta_j})$. Soit $\theta \in \Theta$, on peut donc choisir un indice j (qui dépend alors de θ) tel que $\theta \in B(\theta_j, \eta_{\theta_j})$. On décompose alors $h_n(\theta) - h(\theta)$ en trois termes :

$$\begin{aligned}|h_n(\theta) - h(\theta)| \leqslant &\frac{1}{n}\sum_{i=1}^n \left|\log(f_\theta(X_i)) - \log(f_{\theta_j}(X_i))\right| \\ &+ \left|\frac{1}{n}\sum_{i=1}^n \log(f_{\theta_j}(X_i)) - \mathbb{E}_{\theta^*}[\log(f_{\theta_j}(X))]\right| + \left|\mathbb{E}_{\theta^*}[\log(f_{\theta_j}(X))] - \mathbb{E}_{\theta^*}[\log(f_\theta(X))]\right|.\end{aligned}$$

Comme $\theta \in B(\theta_j, \eta_{\theta_j})$ alors

$$\left|\log(f_\theta(X_i)) - \log(f_{\theta_j}(X_i))\right| \leqslant g_{\theta_j}(\eta_{\theta_j}) < \varepsilon$$

donc

$$|h_n(\theta) - h(\theta)| \leqslant 2\varepsilon + \left| \frac{1}{n} \sum_{i=1}^{n} \log(f_{\theta_j}(X_i)) - \mathbb{E}_{\theta^*}[\log(f_{\theta_j}(X))] \right|$$

d'où

$$\sup_{\theta\in\Theta} |h_n(\theta) - h(\theta)| \leqslant 2\varepsilon + \sum_{j=1}^{N} \left| \frac{1}{n} \sum_{i=1}^{n} \log(f_{\theta_j}(X_i)) - \mathbb{E}_{\theta^*}[\log(f_{\theta_j}(X))] \right|.$$

Or, avec probabilité 1 on a convergence de $\frac{1}{n}\sum_{i=1}^{n} \log(f_{\theta_j}(X_i))$ vers $\mathbb{E}_{\theta^*}[\log(f_{\theta_j}(X))]$ donc

$$\mathbb{P}_{\theta^*}\left(\lim_{n\to\infty} \sup_{\theta\in\Theta} |h_n(\theta) - h(\theta)| \leqslant 2\varepsilon \right) = 1$$

et par le théorème de convergence monotone on en déduit que

$$\mathbb{P}_{\theta^*}\left(\lim_{n\to\infty} \sup_{\theta\in\Theta} |h_n(\theta) - h(\theta)| = 0 \right) = 1 \quad \Rightarrow \quad \sup_{\theta\in\Theta} |h_n(\theta) - h(\theta)| \xrightarrow[n\to+\infty]{p.s.} 0.$$

Etape 6 : Conclusion

Soit $(\hat{\theta}_n)_{n\in\mathbb{N}}$ une suite de Θ telle que $\hat{\theta}_n \in \arg\max_{\theta\in\Theta} h_n(\theta)$ pour tout $n \in \mathbb{N}$. Comme Θ est compact alors $\arg\max_{\theta\in\Theta} h_n(\theta)$ est tout le temps non-vide donc une telle suite est bien définie. On suppose que $\hat{\theta}_n$ ne converge pas vers θ^*. Il existe donc un $\varepsilon > 0$ tel que pour tout $N \in \mathbb{N}$ il existe un $n \geqslant N$ tel que $|\hat{\theta}_n - \theta^*| > \varepsilon$. On peut donc extraire une sous-suite $\hat{\theta}_{\phi(n)}$ de $\hat{\theta}_n$ dont tous les termes vérifient $|\hat{\theta}_{\phi(n)} - \theta^*| > \varepsilon$. De plus, comme Θ est compact alors on peut extraire une sous-suite $\hat{\theta}_{\psi(n)}$ de $\hat{\theta}_{\phi(n)}$ qui converge vers un $\theta_0 \in \Theta$ qui vérifie donc $|\theta_0 - \theta^*| > \varepsilon$. Or, pour tout $\theta \in \Theta$ on a

$$h(\theta_0)-h(\theta) = h(\theta_0)-h(\hat{\theta}_{\psi(n)})+h(\hat{\theta}_{\psi(n)})-h_{\psi(n)}(\hat{\theta}_{\psi(n)})+h_{\psi(n)}(\hat{\theta}_{\psi(n)})-h_{\psi(n)}(\theta)+h_{\psi(n)}(\theta)-h(\theta).$$

Par définition de $\hat{\theta}_{\psi(n)}$ on a $h_{\psi(n)}(\hat{\theta}_{\psi(n)}) \geqslant h_{\psi(n)}(\theta)$ d'où

$$h(\theta_0) - h(\theta) \geqslant h(\theta_0) - h(\hat{\theta}_{\psi(n)}) + h(\hat{\theta}_{\psi(n)}) - h_{\psi(n)}(\hat{\theta}_{\psi(n)}) + h_{\psi(n)}(\theta) - h(\theta).$$

Par continuité de h on a

$$h(\theta_0) - h(\hat{\theta}_{\psi(n)}) \xrightarrow[n\to+\infty]{} 0.$$

De plus, par convergence uniforme presque sûre des h_n on a

$$|h(\hat{\theta}_{\psi(n)}) - h_{\psi(n)}(\hat{\theta}_{\psi(n)}) + h_{\psi(n)}(\theta) - h(\theta)| \leqslant 2 \sup_{\theta\in\Theta} |h_{\psi(n)}(\theta) - h(\theta)| \xrightarrow[n\to+\infty]{p.s.} 0.$$

On en déduit alors que $h(\theta_0) - h(\theta) \geqslant 0$ presque sûrement pour tout $\theta \in \Theta$. θ_0 est donc un maximiseur de h. Or, on avais démontré que l'unique maximiseur de h est θ^* donc $\theta_0 = \theta^*$ ce qui est une contradiction. On a donc prouvé que $\hat{\theta}_n$ converge presque sûrement vers θ^*.

Etape 7 : Extension au cas où Θ n'est pas compact

On ne considère plus que Θ est compact mais qu'il existe une suite croissante de compacts Θ_p tels que $\Theta \subset \bigcup_{p\in\mathbb{N}} \mathrm{int}(\Theta_p)$ par hypothèse (ii). Par l'hypothèse (iii), la suite $\hat{\theta}_n$ reste confinée à l'intérieur d'un compact de Θ avec probabilité 1 et donc il existe un $p \in \mathbb{N}$ (qui dépend de l'aléa) tel que $\hat{\theta}_n \in \Theta_p$ pour tout $n \in \mathbb{N}$. Pour tout $p \in \mathbb{N}$ on pose A_p l'évènement $\{\forall n \in \mathbb{N}, \hat{\theta}_n \in \Theta_p\}$ alors $A_p \subset A_{p+1}$, $\mathbb{P}_{\theta^*}(\bigcup_{p\in\mathbb{N}} A_p) = 1$ et, conditionnement à A_p, on a montré que $\hat{\theta}_n$ converge presque sûrement vers θ^*. On a donc pour tout $N \in \mathbb{N}$:

$$\begin{aligned}
\mathbb{P}_{\theta^*}\left(\hat{\theta}_n \xrightarrow[n\to+\infty]{} \theta^* \right) &\geqslant \mathbb{P}_{\theta^*}\left(\hat{\theta}_n \xrightarrow[n\to+\infty]{} \theta^* \,\middle|\, \bigcup_{p=1}^{N} A_p \right) \mathbb{P}_{\theta^*}\left(\bigcup_{p=1}^{N} A_p \right) \\
&= \mathbb{P}_{\theta^*}\left(\hat{\theta}_n \xrightarrow[n\to+\infty]{} \theta^* \,|A_N \right) \mathbb{P}_{\theta^*}\left(\bigcup_{p=1}^{N} A_p \right) \\
&= \mathbb{P}_{\theta^*}\left(\bigcup_{p=1}^{N} A_p \right)
\end{aligned}$$

ce qui converge vers 1 quand N tend vers $+\infty$ par convergence monotone. ■

Définition 68

*Pour tout $\theta \in \Theta$ on définit l'**information de Fisher** $I(\theta)$ de la famille $\mathcal{F} = \{\mathbb{P}_\theta,\ \theta \in \Theta\}$ par*

$$I(\theta) = -\mathbb{E}_\theta\left[\frac{\partial^2 \log(f_\theta(X))}{\partial \theta^2}\right].$$

Théorème 69

Soit $\theta^* \in \mathrm{int}(\Theta)$. On suppose que la famille $\mathcal{F} = \{\mathbb{P}_\theta, \theta \in \Theta\}$ vérifie les hypothèses (i) à (v) du Théorème 66. On suppose en plus qu'il existe un voisinage Θ_0 de θ^* dans $\mathrm{int}(\Theta)$ tel que

(vi)

$$\mathbb{P}_{\theta^*}\left(\text{La fonction } \theta \mapsto \log(f_\theta(X)) \text{ est } \mathcal{C}^2 \text{ sur } \Theta_0\right) = 1.$$

(vii) $I(\theta^*)$ est inversible et finie.

Alors,

$$\forall \theta \in \Theta,\ \sqrt{n}(\hat{\theta}_n - \theta^*) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}\left(0, \frac{1}{I(\theta^*)}\right)$$

Démonstration : Comme $\hat{\theta}_n \xrightarrow[n\to+\infty]{p.s.} \theta^*$ par le théorème 66 et $\theta^* \in \mathrm{int}(\Theta)$ alors, avec probabilité 1, il existe un $N \in \mathbb{N}$ tel que pour tout $n \geqslant N$ on a $\hat{\theta}_n \in \Theta_0$. Soit $n \geqslant N$, par le théorème des accroissements finis il existe θ'_n entre $\hat{\theta}_n$ et θ^* tel que

$$h'_n(\hat{\theta}_n) = h'_n(\theta^*) + h''_n(\theta'_n)(\hat{\theta}_n - \theta^*).$$

Comme $\hat{\theta}_n$ est un maximum de h_n dans Θ_0 alors, par l'hypothèse (vi), on en déduit que $h'_n(\hat{\theta}_n) = 0$. On obtient alors que

$$\sqrt{n}(\hat{\theta}_n - \theta^*) = \sqrt{n}h'_n(\theta^*)h''_n(\theta'_n)^{-1}.$$

On commence par montrer le résultat suivant :

Lemme 70

(i) Si la fonction $\theta \mapsto \log(f_\theta(X))$ est dérivable en θ^* presque sûrement alors

$$\mathbb{E}_{\theta^*}\left[\left.\frac{\partial \log(f_\theta(X_i))}{\partial \theta}\right|_{\theta=\theta^*}\right] = 0; \tag{0.1}$$

(ii) De plus, si la fonction $\theta \mapsto \log(f_\theta(X))$ est deux fois dérivable en θ^* presque sûrement alors

$$\mathbb{E}_{\theta^*}\left[\left(\left.\frac{\partial \log(f_\theta(X_i))}{\partial \theta}\right|_{\theta=\theta^*}\right)^2\right] = I(\theta^*). \tag{0.2}$$

Démonstration : (i) On peut réécrire

$$\mathbb{E}_{\theta^*}\left[\left.\frac{\partial \log(f_\theta(X))}{\partial \theta}\right|_{\theta=\theta^*}\right] = \mathbb{E}_{\theta^*}\left[\frac{\left.\frac{\partial f_\theta(X)}{\partial \theta}\right|_{\theta=\theta^*}}{f_{\theta^*}(X)}\right].$$

Si $\mathbb{P}_\theta$ est une loi continue sur Ω alors

$$\begin{aligned}
\mathbb{E}_{\theta^*}\left[\frac{\frac{\partial f_\theta(X)}{\partial\theta}\Big|_{\theta=\theta^*}}{f_{\theta^*}(X)}\right] &= \int_\Omega \frac{\frac{\partial f_\theta(x)}{\partial\theta}\Big|_{\theta=\theta^*}}{f_{\theta^*}(x)} f_{\theta^*}(x)\mathrm{d}x \\
&= \int_\Omega \frac{\partial f_\theta(x)}{\partial\theta}\Big|_{\theta=\theta^*}\mathrm{d}x \\
&= \frac{\partial\int_\Omega f_\theta(x)\mathrm{d}x}{\partial\theta}\Big|_{\theta=\theta^*} = \frac{\partial 1}{\partial\theta}\Big|_{\theta=\theta^*} = 0
\end{aligned}$$

Si $\mathbb{P}_\theta$ est une loi discrète sur Ω alors

$$\begin{aligned}
\mathbb{E}_{\theta^*}\left[\frac{\frac{\partial f_\theta(X)}{\partial\theta}\Big|_{\theta=\theta^*}}{f_{\theta^*}(X)}\right] &= \sum_{\omega\in\Omega} \frac{\frac{\partial\mathbb{P}_\theta(X=\omega)}{\partial\theta}\Big|_{\theta=\theta^*}}{\mathbb{P}_{\theta^*}(X=\omega)}\mathbb{P}_{\theta^*}(X=\omega) \\
&= \sum_{\omega\in\Omega}\frac{\partial\mathbb{P}_\theta(X=\omega)}{\partial\theta}\mathrm{d}x\Big|_{\theta=\theta^*} \\
&= \frac{\partial\sum_{\omega\in\Omega}\mathbb{P}_\theta(X=\omega)}{\partial\theta}\Big|_{\theta=\theta^*} = \frac{\partial 1}{\partial\theta}\Big|_{\theta=\theta^*} = 0.
\end{aligned}$$

(ii) On peut écrire

$$\begin{aligned}
-I(\theta^*) &= \mathbb{E}_{\theta^*}\left[\frac{\partial^2\log(f_\theta(X))}{\partial\theta^2}\Big|_{\theta=\theta^*}\right] \\
&= \mathbb{E}_{\theta^*}\left[\frac{\partial}{\partial\theta}\left(\frac{\frac{\partial f_\theta(X)}{\partial\theta}}{f_\theta(X)}\right)\Big|_{\theta=\theta^*}\right] \\
&= \mathbb{E}_{\theta^*}\left[\frac{\frac{\partial^2 f_\theta(X)}{\partial\theta^2}\Big|_{\theta=\theta^*}}{f_{\theta^*}(X)} - \left(\frac{\frac{\partial f_\theta(X)}{\partial\theta}}{f_\theta(X)}\right)^2\Big|_{\theta=\theta^*}\right] \\
&= \mathbb{E}_{\theta^*}\left[\frac{\frac{\partial^2 f_\theta(X)}{\partial\theta^2}\Big|_{\theta=\theta^*}}{f_{\theta^*}(X)}\right] - \mathbb{E}\left[\left(\frac{\partial\log(f_\theta(X))}{\partial\theta}\right)^2\Big|_{\theta=\theta^*}\right].
\end{aligned}$$

Le premier terme est nul par (0.1) ce qui prouve le lemme. ■

En utilisant les hypothèses (vi), (vii) et l'équation (0.2) on en déduite que

$$\mathbb{E}\left[\left(\frac{\partial\log(f_\theta(X))}{\partial\theta}\right)^2\Big|_{\theta=\theta^*}\right] = I(\theta^*) < +\infty.$$

On peut donc appliquer le théorème central limite aux variables $\frac{\partial\log(f_\theta(X_i))}{\partial\theta}|_{\theta=\theta^*}$ ce qui donne

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n\frac{\partial\log(f_\theta(X_i))}{\partial\theta}\Big|_{\theta=\theta^*} - \mathbb{E}_{\theta^*}\left[\frac{\partial\log(f_\theta(X))}{\partial\theta}\Big|_{\theta=\theta^*}\right]\right) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}\left(0, \mathrm{Var}_{\theta^*}\left(\frac{\partial\log(f_\theta(X))}{\partial\theta}\Big|_{\theta=\theta^*}\right)\right).$$

On remarque que le premier terme correspond à $h_n'(\theta^*)$ et que le deuxième terme est nul par (0.1). De plus, en conséquence de (0.1) et (0.2) la variance de la loi normale vérifie

$$\mathrm{Var}_{\theta^*}\left(\frac{\partial\log(f_\theta(X))}{\partial\theta}\Big|_{\theta=\theta^*}\right) = \mathbb{E}\left[\left(\frac{\partial\log(f_\theta(X))}{\partial\theta}\right)^2\Big|_{\theta=\theta^*}\right] - \mathbb{E}\left[\frac{\partial\log(f_\theta(X))}{\partial\theta}\Big|_{\theta=\theta^*}\right]^2 = I(\theta^*) - 0$$

On peut donc réécrire le résultat du théorème central limite sous la forme

$$\sqrt{n} h_n'(\theta^*) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}(0, I(\theta^*)).$$

Comme θ_n' est entre $\hat{\theta}_n$ et θ^* et $\hat{\theta}_n$ converge presque sûrement vers θ^* alors, par le théorème des gendarmes, θ_n' converge aussi presque sûrement vers θ^*. On en déduit alors que $h_n''(\theta_n')^{-1}$ converge presque sûrement vers $h''(\theta^*)^{-1}$. Or, on remarque que $h''(\theta^*) = -I(\theta^*)$ par définition et on en déduit alors que

$$h_n''(\theta_n')^{-1} \xrightarrow[n\to+\infty]{p.s.} -\frac{1}{I(\theta^*)}.$$

Par le lemme de Slutsky on peut donc conclure que

$$\sqrt{n} h_n'(\theta^*) h_n''(\theta_n')^{-1} \xrightarrow[n\to+\infty]{} -\frac{1}{I(\theta^*)} \times \mathcal{N}(0, I(\theta^*)) = \mathcal{N}\left(0, \frac{1}{I(\theta^*)}\right),$$

ce qui conclue le théorème. ■