

Notes de Cours d'Inférence Statistique

YEHOR KOROTENKO

17 February 2026

Table of contents

1 Introduction	3
1.1 Evaluation	3
1.2 Modèle Statistique	3
1.3 Estimateurs	3
1.4 Risque quadratique	4
1.4.1 Exemple : Modèle de Poisson	4
1.5 Consistance	5
1.5.1 Exemple : Retour au modèle de Poisson	5
1.5.2 Méthode “Plug-in”	5
2 Estimateurs	6
2.1 Cadre paramétrique	6
2.1.1 Modèle statistique paramétrique	6
2.2 Méthode des moments	6
2.3 Rendu sur le L.A.C.	8
2.3.1 Variance empirique	9
2.4 Méthode de maximum de vraisemblance	9
2.4.1 Modèle donné	9
2.4.2 En pratique	9
3 Information de Fisher, efficacité	12
3.1 Modèle régulier	12
3.2 Score et Information de Fisher	13
3.3 Inv. de Fisher et dérivé seconde	14
3.4 Inégalité de Cramer - Rao	15
4 Étude asymptotique des estimateurs	17
4.1 Convergences	17
4.2 Consistance des estimateurs	18
4.3 Normalité asymptotique	18
4.4 δ -méthode	20
5 Fonction de répartition empirique	21
5.1 Estimation empirique	22
5.2 Inverse généralisé	22
5.3 Quantile empirique	24
6 Intervalles de confiance	25
6.1 Définitions	25

6.2 Interprétation	25
6.3 Méthode pivotale	25
7 Compléments (avant partiel)	27
7.1 Propriétés asymptotiques d'une suite d'estimateurs $(\hat{\theta}_n)_{n \geq 1}$	27
7.2 Pivot (asymptotique) ou statistique pivotale	29

Introduction

§1

1.1 Evaluation

- 0.4 CC +0.6 Examen.
- Répartition : 80% partiel, 20% Interro (prévue le 26/01).

1.2 Modèle Statistique

DEFINITION 1.1 (MODÈLE STATISTIQUE) – *Un modèle statistique est un espace de probabilité $(\Omega, \mathcal{A}, \mathcal{P})$ où \mathcal{P} est une famille de lois de probabilité $\{P_\theta; \theta \in \Theta\}$.*

- Si $\exists p \in \mathbb{N}^*, \Theta \subset \mathbb{R}^p$: modèle paramétrique.
- Sinon : modèle non paramétrique.

EXAMPLE 1.2 (FAMILLES DE LOIS) –

- *Lois de Poisson* : $\mathcal{P} = \{P(\lambda); \lambda > 0\}$.
- *Densité régulière* : $\mathcal{P} = \{\mathbb{P}; \mathbb{P} \text{ dont la densité admet une dérivée seconde bornée}\}$.

DEFINITION 1.3 (OBSERVATION) – *Une observation est une variable aléatoire (v.a.) dont la loi appartient à $\{P_\theta, \theta \in \Theta\}$. Notre observation aura une structure de n -échantillons X_1, \dots, X_n i.i.d. (indépendants et identiquement distribués) de loi commune $\in \{P_\theta, \theta \in \Theta\}$.*

REMARK 1.4 – (X_1, \dots, X_n) est de loi $P_\theta^{\otimes n}$. L'échantillon contient toute l'information sur P_θ , donc sur θ .

DEFINITION 1.5 (IDENTIFIABILITÉ) – *Un modèle est identifiable si et seulement si (ssi) l'application $\theta \mapsto P_\theta$ est injective.*

1.3 Estimateurs

Hypothèse : On observe X_1, \dots, X_n i.i.d. de loi commune $\in \{P_\theta, \theta \in \Theta \subset \mathbb{R}^p\}$ (modèle paramétrique identifiable). Soit θ^* la vraie valeur inconnue telle que $P_{X_i} = P_{\theta^*}$.

DEFINITION 1.6 (ESTIMATEUR) – *Un estimateur de θ est une fonction de l'échantillon (X_1, \dots, X_n) mesurable et indépendante de θ (calculable à partir des données).*

Notation : $\hat{\theta} = \hat{\theta}_n = h(X_1, \dots, X_n)$. C'est une variable aléatoire.
Exemples : $\hat{\theta} = \bar{X}$, $\hat{\theta} = X_1 - X_3$, etc.

Questions fondamentales :

1. Comment définir un bon estimateur ?
2. Comment construire un bon estimateur ?

1.4 Risque quadratique

Idée : En moyenne, $\hat{\theta}$ doit être proche de θ . On regarde $\mathbb{E}[\hat{\theta} - \theta]$.

DEFINITION 1.7 (BIAIS) – Le biais de $\hat{\theta}$ est défini par :

$$B(\hat{\theta}, \theta) = \mathbb{E}[\hat{\theta}] - \theta$$

On dit que $\hat{\theta}$ est sans biais si $B(\hat{\theta}, \theta) = 0$.

DEFINITION 1.8 (RISQUE QUADRATIQUE / MSE) –

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

C'est la Mean Squared Error (MSE) en anglais.

On dit que $\hat{\theta}_1$ est meilleur que $\hat{\theta}_2$ ssi $R(\hat{\theta}_1, \theta) \leq R(\hat{\theta}_2, \theta)$.

1.4.1 Exemple : Modèle de Poisson

Soit X_1, \dots, X_n de loi P_θ de Poisson, $\theta > 0$. On cherche un estimateur de $\theta = \mathbb{E}[X_i]$.

Proposons : $\hat{\theta} = \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$.

Calcul du Biais :

$$\begin{aligned} B(\hat{\theta}, \theta) &= \mathbb{E}\left[\frac{1}{n} \sum_{i=1}^n X_i\right] - \theta \\ &= \frac{1}{n} \sum_{i=1}^n \mathbb{E}[X_i] - \theta \quad (\text{par linéarité}) \\ &= \frac{1}{n} \cdot n \cdot \mathbb{E}[X_1] - \theta \\ &= \theta - \theta = 0 \end{aligned}$$

Donc $\mathbb{E}[\bar{X}] = \theta$, est l'estimateur sans biais.

Calcul du Risque :

$$\begin{aligned} R(\hat{\theta}, \theta) &= \mathbb{E}[(\bar{X} - \theta)^2] = \mathbb{E}[(\bar{X} - \mathbb{E}[\bar{X}])^2] \\ &= \text{Var}(\bar{X}) = \text{Var}\left(\frac{1}{n} \sum X_i\right) \\ &= \frac{1}{n^2} \sum \text{Var}(X_i) \quad (\text{car i.i.d}) \\ &= \frac{1}{n^2} \cdot n \cdot \text{Var}(X_1) = \frac{\text{Var}(X_1)}{n} = \frac{\theta}{n} \end{aligned}$$

THEOREM 1.9 (DÉCOMPOSITION BIAIS-VARIANCE DU RISQUE) –

$$R(\hat{\theta}, \theta) = (B(\hat{\theta}, \theta))^2 + \text{Var}(\hat{\theta})$$

Proof.

$$\begin{aligned}
 R(\hat{\theta}, \theta) &= \mathbb{E}[(\hat{\theta} - \theta)^2] \\
 &= \mathbb{E}[(\hat{\theta} - \mathbb{E}[\hat{\theta}] + \mathbb{E}[\hat{\theta}] - \theta)^2] \\
 &= \mathbb{E}[(\hat{\theta} - \mathbb{E}[\hat{\theta}])^2] + \mathbb{E}[(\mathbb{E}[\hat{\theta}] - \theta)^2] + 2\mathbb{E}[(\hat{\theta} - \mathbb{E}[\hat{\theta}])(\mathbb{E}[\hat{\theta}] - \theta)] \\
 &= \text{Var}(\hat{\theta}) + (B(\hat{\theta}, \theta))^2 + 2(\mathbb{E}[\hat{\theta}] - \theta)\underbrace{\mathbb{E}[\hat{\theta} - \mathbb{E}[\hat{\theta}]]}_0 \\
 &= \text{Var}(\hat{\theta}) + B(\hat{\theta}, \theta)^2
 \end{aligned}$$

□

1.5 Consistance

Propriété asymptotique. On ne considère que des estimateurs consistants.

DEFINITION 1.10 (CONSISTANCE) — Soit (X_1, \dots, X_n) i.i.d. de loi P_θ . Soit $\hat{\theta}_n = h(X_1, \dots, X_n)$. $\hat{\theta}_n$ est un estimateur consistant (ou convergent) de θ ssi :

$$\hat{\theta}_n \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} \theta$$

REMARK 1.11 — $\hat{\theta}_n$ est fortement consistantssi $\hat{\theta}_n \xrightarrow[n \rightarrow +\infty]{\text{p.s.}} \theta$.

1.5.1 Exemple : Retour au modèle de Poisson

$\Theta = \mathbb{R}_+^*, \hat{\theta}_n = \bar{X}$.

- On peut invoquer la Loi des Grands Nombres (LGN) : $\bar{X} \xrightarrow{\mathbb{P}} \mathbb{E}[X_i] = \theta$.
- Via le risque quadratique :

$$R(\hat{\theta}_n, \theta) = \text{Var}(\bar{X}) = \frac{\theta}{n} \xrightarrow[n \rightarrow +\infty]{} 0$$

D'après l'inégalité de Bienaymé-Tchebychev :

$$P(|\hat{\theta}_n - \theta| > \varepsilon) \leq \frac{\mathbb{E}[(\hat{\theta}_n - \theta)^2]}{\varepsilon^2} = \frac{R(\hat{\theta}_n, \theta)}{\varepsilon^2} \rightarrow 0$$

1.5.2 Méthode “Plug-in”

Soit (X_1, \dots, X_n) i.i.d. Poisson(θ). On veut estimer $\beta = P(X_i = 0) = e^{-\theta}$.

$$\hat{\beta} = e^{-\hat{\theta}} = e^{-\bar{X}}$$

$\hat{\beta}$ est consistant pour estimer β .

LEMMA 1.12 (LEMME DE L'APPLICATION CONTINUE) — Si $Z_n \xrightarrow{\mathbb{P}} Z$, alors $h(Z_n) \xrightarrow{\mathbb{P}} h(Z)$ pour toute fonction continue h .

Estimateurs

§2

2.1 Cadre paramétrique

2.1.1 Modèle statistique paramétrique

On dispose d'une observation (X_1, \dots, X_n) , un échantillon de variable aléatoire i.i.d (indépendantes, identiquement distribuées) de loi commune P appartenant à une famille de lois de probabilités paramétrée $\{P_{\theta, \theta \in \Theta \subset \mathbb{R}^p}\}$.

REMARK 2.1 — Si $\Theta \subset$ espace de dimension infinie \rightarrow modèle non-paramétrique.

Estimer P c'est estimer $\theta \in \mathbb{R}^p$.

EXAMPLE 2.2 — Bernoulli (θ), Exp (θ), $\mathcal{N}(\mu, \sigma^2)$, loi de densité $f_\theta(x) = \theta x^{\theta-1} 1_{x \in [0,1]}$

NOTATION 2.3 — $E_{\theta_n}[h(X_1, \dots, X_n)]$, $\Theta[h(X_1, \dots, X_n)]$
 Loi de $(X_1, \dots, X_n) \rightarrow P_\theta^{\otimes n}$

DEFINITION 2.4 (ESTIMATEUR) —

$$\hat{\theta} = \hat{\theta}_n = h(X_1, \dots, X_n)$$

DEFINITION 2.5 (QUALITÉ) —

- Risque

$$R(\hat{\theta}, \theta) = E_\theta \left[(\hat{\theta} - \theta)^2 \right]$$

- Consistance

$$\hat{\theta}_n \xrightarrow[n \rightarrow +\infty]{P} \theta$$

DEFINITION 2.6 (MODÈLE IDENTIFIABLE) —

$$\theta \rightarrow P_\theta \text{ injective}$$

2.2 Méthode des moments

DEFINITION 2.7 — On appelle *moment théorique* de la loi de X_i d'ordre k :

$$\mu_k = E[X_i^k], \quad k \geq 1$$

DEFINITION 2.8 — On appelle *moment empirique* de la loi des X_i d'ordre k :

$$\hat{\mu}_k = \frac{1}{n} \sum_{i=1}^n X_i^k$$

Par la loi des grands nombres $\hat{\mu}_k \xrightarrow[n \rightarrow +\infty]{P} \mu_k$.

La méthode des moments: si on peut écrire θ ou $g(\theta)$ paramètre d'intérêt comme une fonction des k premiers moments théoriques.

$$\theta = \mathcal{L}(\mu_1, \dots, \mu_k)$$

alors l'estimateur

$$\hat{\theta} = \mathcal{L}(\hat{\mu}_1, \dots, \hat{\mu}_k)$$

est obtenu par la méthode.

EXAMPLE 2.9 (DES CALCULS DES ESTIMATEURS EN UTILISANT LA MÉTHODE DES MOMENTS) –

- $X_i \sim \text{Bernoulli}(\theta)$ à valeurs 0-1,

$$\theta = P(X_i = 1) = E[X_i] \rightarrow \frac{1}{n} \sum_{i=1}^n X_i = \bar{X}$$

- $X_i \sim \text{Exp}(\theta)$, $f_\theta(x) = \theta e^{-\theta x} 1_{x \geq 0}$, $E[X] = \frac{1}{\theta} \iff \theta = \frac{1}{\mu_1}$, par la méthode des moments,

$$\hat{\theta} = \frac{1}{\hat{\mu}_1} = \frac{1}{\bar{X}}$$

$$\begin{aligned} \Theta(X_i) = \frac{1}{\theta^2} &\iff \theta^2 = \frac{1}{E[X_i^2] - E[X_i]^2} \\ &\iff \theta = \frac{1}{\sqrt{\mu_2 - \mu_1^2}} \\ &\implies \hat{\theta}_2 = \frac{1}{\sqrt{\frac{1}{n} \sum_{i=1}^n X_i^2 - (\bar{X})^2}} \end{aligned}$$

- X_1, \dots, X_n i.i.d. de la loi P_θ de densité

$$f_\theta(x) = \theta x^{\theta-1} 1_{x \in [0,1]}$$

$$E_\theta[X_i] = \theta \int_0^1 x^\theta dx = \frac{\theta}{\theta+1}$$

Méthode des moments:

$$\begin{aligned} (\theta+1)\mu_1 = \theta &\iff \theta(1-\mu_1) = \mu_1 \iff \theta = \frac{E[X_i]}{1-E[X_i]} \\ &\implies \hat{\theta}_M = \frac{\bar{X}}{1-\bar{X}}, P_\theta(\bar{X} = 1) = P_\theta(X_1 = X_2 = \dots = X_n = 1) = 0 \end{aligned}$$

2.3 Rendu sur le L.A.C.

(L.A.C = lemme des applications continues) $(X_n)_{n \geq 1}$ suite de variables aléatoires. Si X_n converge vers X , que peut-on dire de $g(X_n)_{n \geq 1}$? Si g continue, LAC.

- si $X_n \xrightarrow{P} X$ alors $g(X_n) \xrightarrow{P} g(X)$
- si $X_n \xrightarrow{\mathcal{L}} X$ alors $g(X_n) \xrightarrow{\mathcal{L}} g(X)$

REMARK 2.10 (CONDITION SUFFISANTE) –

$$D_g = \{\text{points de discontinuité de } g\}$$

si $P(X \in D_g) = 0$, le LAC est vrai.

EXAMPLE 2.11 –

$$g(x) = \frac{x}{1-x}$$

- LGN: $\bar{X} \xrightarrow{P} E[X]$
- LAC: $g(\bar{X}) = \hat{\theta}_n \xrightarrow[n \rightarrow +\infty]{P} g(E[X]) = \theta$

LAC pour des couples de suites de variables aléatoires:

- si $(X_n, Y_n) \xrightarrow{P} (X, Y)$, alors $g(X_n, Y_n) \xrightarrow{P} g(X, Y)$, si $g : \mathbb{R}^2 \rightarrow \mathbb{R}$ ou \mathbb{R}^2 continue
- si $(X_n, Y_n) \xrightarrow{\mathcal{L}} (X, Y)$, alors $g(X_n, Y_n) \xrightarrow{\mathcal{L}} g(X, Y)$

EXAMPLE 2.12 –

$$\hat{\theta}_2 = \frac{1}{\sqrt{\frac{1}{n} \sum_{i=1}^n X_i^2 - (\bar{X})^2}} \quad \text{consistant?}$$

LGN:

- $\bar{X} \xrightarrow{P} \mu_1$
- $\frac{1}{n} \sum_{i=1}^n X_i^2 \xrightarrow{P} \mu_2$

donc

$$\left(\frac{1}{n} \sum_{i=1}^n X_i^2 \right) \xrightarrow{P} \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$$

$g(x, y) = \frac{1}{\sqrt{y-x^2}} \Rightarrow \hat{\theta}^M \text{ constant de } \theta$, g continue sauf en $\{(x, y) \in \mathbb{R}^2, y = x^2\}$ de mesure nulle.

Mais c'est faux pour une converge en loi.

PROPOSITION 2.13 (CONVERGENCE DE COUPLES) –

$$\begin{pmatrix} X_n \\ Y_n \end{pmatrix} \xrightarrow{P} \begin{pmatrix} X \\ Y \end{pmatrix} \quad \text{ssi} \quad \begin{cases} X_n \xrightarrow{P} X \\ Y_n \xrightarrow{P} Y \end{cases}$$

Proof.

- \Rightarrow alors LAC $g(x, y) = x$ continue donc $X_n \rightarrow X$ et $Y_n \rightarrow Y$

- \iff convergence du couple?

$$\forall \varepsilon > 0, P(|X_n - X| + |Y_n - Y| > \varepsilon) \leq \underbrace{P\left(|X_n - X| > \frac{\varepsilon}{2}\right)}_{\rightarrow 0} + \underbrace{P\left(|Y_n - Y| > \frac{\varepsilon}{2}\right)}_{\rightarrow 0}$$

Cette réciproque est fausse pour la converge en loi!

□

2.3.1 Variance empirique

Si la X_i admettent une esperance μ et une variance σ^2 , on appelle variance empirique

$$\begin{aligned}\hat{\sigma}_n^2 &= \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 + \frac{1}{n} \sum_{i=1}^n \bar{X}^2 - \frac{2}{n} \sum_{i=1}^n X_i \bar{X} \\ &= \frac{1}{n} \sum_{i=1}^n X_i^2 + \bar{X}^2 - 2\bar{X}\bar{X} = \tilde{\sigma}^2\end{aligned}$$

estimateur des moments:

$$\sigma^2 = E[X_i^2] - E[X_i]^2$$

On remplace les moments théoriques par les moments empiriques

$$\rightarrow \tilde{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 - (\bar{X})^2$$

Consistance: $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 - (\bar{X})^2$,

$$\begin{cases} \bar{X} \xrightarrow{P} E[X] \\ \frac{1}{n} \sum_{i=1}^n X_i^2 \xrightarrow{P} E[X^2] \end{cases} \xrightarrow{\text{cv en proba}} \left(\frac{1}{n} \sum_{i=1}^n X_i^2 \right) \xrightarrow{\text{LAC}} \hat{\sigma}^2 \text{ consistant de } \text{Var}(X) = E[X^2] - E[X]^2$$

EXAMPLE 2.14 –

- calculer le biais de $\hat{\sigma}_n^2$
- calculer le risque de $\hat{\sigma}_n^2$

2.4 Méthode de maximum de vraisemblance

2.4.1 Modèle donné

$(P_\theta)_{\theta \in \Theta}$ est donné s'il existe une mesure μ (positive σ définie $\rightarrow X_i$ à valeurs dans E , $E = \cup E_n$ avec $\mu(E_n)$ finie) telle que $\forall \theta, P_\theta$ admet une densité par rapport à μ .

2.4.2 En pratique

- soit E au plus dénombrable: μ = mesure de comptage. Si $\exists, \{a_1, a_2, \dots\}$ tq $\sum_{k \geq 1} P_\theta(X_i = a_k) = 1$, alors $\mu = \sum_{k \geq 1} \delta_{a_k}$ avec $\delta_a(\{a\}) = 1$ mesure de dirac.

EXAMPLE 2.15 – Bernoulli (θ), $X_i = 1$, probas $\theta \rightarrow \mu = \delta_0 + \delta_1$ On écrira

$$f_\theta(x) = \underbrace{P_\theta(\{x\})}_{=1-\theta} - P_\theta(X_i = x) \text{ avec } x \in \{a_1, a_2, \dots\}$$

- soit $E = \mathbb{R}^p$, alors f_θ est la densité usuelle

f_θ densité de P_θ

DEFINITION 2.16 – On appelle vraisemblance de l'échantillon (X_1, \dots, X_n) la fonction

$$\theta \rightarrow L_n(\theta) = \prod_{i=1}^n f_\theta(X_i) \quad (\text{variable aléatoire})$$

DEFINITION 2.17 – Un estimateur du max de vraisemblance $\hat{\theta}_{MV}$ est définie par:

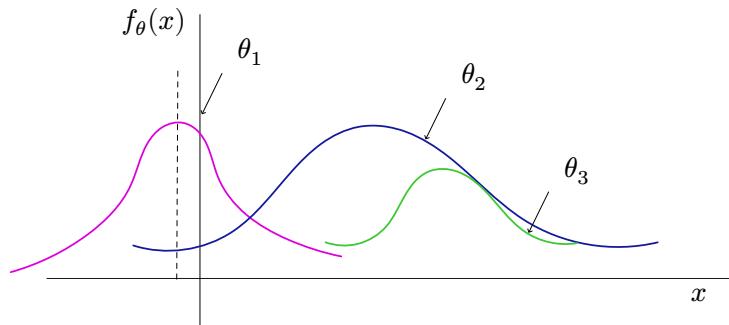
$$\forall \theta \in \Theta, L_n(\theta) \leq L_n(\hat{\theta})$$

On travaille souvent avec la **log-vraisemblance**

$$\log L_n(\theta) = \sum_{i=1}^n \ln f_\theta(X_i) \quad \text{somme de variables aléatoires}$$

$$\log L_n(\hat{\theta}) = \sup_{\theta \in \Theta} \log L_n(\theta)$$

REMARK 2.18 – $\hat{\theta}$ est une variable aléatoire



fdsa

EXAMPLE 2.19 –

- Bernoulli(θ), $f_\theta(x) = \theta^x(1-\theta)^{1-x}$, X_i à valeurs 0-1

$$L_n(\theta) = \prod_{i=1}^n \theta^{X_i}(1-\theta)^{1-X_i} = \theta^{\sum_{i=1}^n X_i} (1-\theta)^{n-\sum_{i=1}^n X_i}$$

$$\log L_n(\theta) = \left(\sum_{i=1}^n X_i \right) \ln \theta + \left(n - \sum_{i=1}^n X_i \right) \ln(1-\theta)$$

$$(\log L_n)'(\theta) = \frac{\sum_{i=1}^n X_i}{\theta} - \frac{n - \sum_{i=1}^n X_i}{1-\theta} = \frac{\sum_{i=1}^n X_i - n\theta}{\theta(1-\theta)} (\bar{X} - \theta)$$

Equation de vraisemblance:

$$\begin{aligned} (\log L_n)'(\theta) = 0 &\iff (1 - \theta) \sum_{i=1}^n X_i = \left(n - \sum_{i=1}^n X_i \right) \theta \\ &\iff \sum_{i=1}^n X_i = n\theta \implies \theta = \frac{\sum_{i=1}^n X_i}{n} \end{aligned}$$

le point critique, est-il un maximum?

La dérivée change de signe en $\bar{X} \rightarrow$ on a bien un max $\rightarrow \hat{\theta}^{MV} = \bar{X}$

Condition du 2nd ordre, si $(\log L_n)''(\theta) < 0$ pour tout $\theta \implies \log L_n$ est concave \implies max global

$$(\log L_n)''(\theta) = -\frac{\sum_{i=1}^n X_i}{\theta^2} - \frac{n - \sum_{i=1}^n X_i}{(1 - \theta)^2} < 0, \forall \theta$$

Information de Fisher, efficacité

§3

Soit $(P_\theta)_{\theta \in \Theta}$, $\Theta \subset \mathbb{R}^p$ (identifiable, donnée). On note f_θ densité de P_θ

$$\text{Supp } f_\theta = \{x \in E, f_{\theta(x)} > 0\}$$

Étant donné (X_1, \dots, X_n) , i.i.d. de loi P_θ et $\theta \mapsto L(\theta) = \prod_{i=1}^n f_{\theta(X_i)}$ la vraisemblance de l'échantillon. Sur $\text{Supp } f_\theta$ on peut calculer

$$\log L_{n(\theta)} = \sum_{i=1}^n \log f_{\theta(X_i)}$$

$$\hat{\theta} = \text{argmax}_{\theta \in \Theta} \log L_{n(\theta)}$$

PROPOSITION 3.1 – Si $\hat{\theta}$ EMV¹ de θ , $g(\hat{\theta})$ est un EMV de $g(\theta)$

Objectif: que peut-on avoir de “mieux” comme estimateur? → modèle régulier

3.1 Modèle régulier

DEFINITION 3.2 – Le modèle $(P_\theta)_{\theta \in \Theta}$ est dit régulier si

1. Θ est un ouvert et $\theta \mapsto f_{\theta(x)}$ est C^1
2. $\text{Supp } f_\theta$ ne dépend pas de θ : $S = \{x, f_{\theta(x)} > 0\}$
3. Pour tout θ , l'application

$$x \mapsto \frac{\frac{\partial f_\theta}{\partial \theta}(x)}{f_{\theta(x)}} \mathbb{1}_{f_{\theta(x)} > 0}$$

est intégrable (L, μ) et l'intégrale

$$I(\theta) = \int_S \frac{\frac{\partial f_\theta}{\partial \theta}(x)}{f_{\theta(x)}} \mathbb{1}_{f_{\theta(x)} > 0} dx$$

est continue sur Θ .

NOTATION 3.3 – On note la dérivée de $f_{\theta(x)}$ par rapport à θ : $\frac{\partial f_\theta}{\partial \theta}(x)$. La quantité $I(\theta)$ est appelée **Information de Fisher du modèle**.

EXAMPLE 3.4 –

- $f_{\theta(x)} = \theta e^{-x\theta}$ densité par rapport à $\mu(dx) = \mathbb{1}_{x \geq 0} dx$
- $\theta \mapsto \theta e^{-x\theta}$ est C^∞ sur $\Theta =]0, +\infty[$, $\text{Supp } f_\theta = \mathbb{R}_+$

$$\frac{\partial f_\theta}{\partial \theta}(x) = (1 - x\theta)e^{-x\theta}$$

¹EMV = Estimateur de Maximum de Vraisemblance

$$\frac{(1-x\theta)^2(e^{-x\theta})^2}{\theta e^{-x\theta}} = \frac{(1-x\theta)^2}{\theta} e^{-x\theta}$$

$$\begin{aligned} I(\theta) &= \int_{\theta}^{\infty} \frac{(1-x\theta)^2}{\theta^2} \theta e^{-x\theta} dx \\ &= \frac{1}{\theta^2} E_{\theta}(1-X\theta)^2 \\ &= \frac{1}{\theta^2} [1 - 2\theta E(X) + \theta^2 E(X^2)] = \frac{1}{\theta^2} \end{aligned}$$

continue sur $]0, +\infty[$

EXAMPLE 3.5 – Bernoulli(θ), $x = 0, 1$, $f_{\theta(0)} = 1 - \theta$, $f_{\theta(1)} = \theta$, densité par rapport à $\delta_0 + \delta_1$

Pour tout $x \in \{0, 1\}$, $\theta \mapsto f_{\theta(x)}$ est C^1

$$\frac{\left(\frac{\partial f_{\theta(0)}}{\partial \theta}\right)^2}{f_{\theta(0)}} = \frac{1}{1-\theta}$$

$$\frac{\left(\frac{\partial f_{\theta(1)}}{\partial \theta}\right)^2}{f_{\theta(1)}} = \frac{1}{\theta} \Rightarrow I(\theta) = \frac{1}{1-\theta} + \frac{1}{\theta} = \frac{1}{\theta(1-\theta)}$$

continue sur $]0, 1[$

EXAMPLE 3.6 – $f_{\theta(x)} = \frac{1}{\theta} \mathbb{1}_{[0, \theta]}(x) = \frac{1}{\theta} \mathbb{1}_{[x, +\infty[}(\theta)$ modèle non régulier

3.2 Score et Information de Fisher

(X_1, \dots, X_n) i.i.d de loi de P_{θ} , f_{θ}

DEFINITION 3.7 – On appelle *score* ou *vecteur de score* la dérivée de la log vraisemblance $\frac{\partial}{\partial \theta} \log L_{n(\theta)} = S_{n(\theta)} = \sum_{i=1}^n \frac{\partial}{\partial \theta} \log f_{\theta}(X_i)$

EXAMPLE 3.8 – $X_i \sim \mathcal{E}(\theta)$, $L_{n(\theta)} = \theta^n e^{-\theta \sum_i X_i}$, $\log L_n(\theta) = n \log \theta - \theta \sum_i X_i$, donc $S = n(\theta) = \frac{n}{\theta} - \sum_{i=1}^n X_i - n$

REMARK 3.9 –

$$E(S_n(\theta)) = E\left[n\left(\frac{1}{\theta} - \frac{\sum X_i}{n}\right)\right]$$

Hyp supplémentaire de régularité: (H) pour tout estimateur $h(X)$ et tout θ , les intégrales suivantes existent et sont égales:

$$\frac{\partial}{\partial \theta} \int_S h(x) f_{\theta}(x) dx = \int_S h(x) \frac{\partial f_{\theta}}{\partial \theta}(x) dx$$

REMARK 3.10 – condition d'application du thm de dérivation de Lebesgue.

$$h \sup_{\theta \in V_\theta} \left| \frac{\partial f_\theta}{\partial \theta}(x) \right| \in L_1(\mu)$$

PROPOSITION 3.11 – Sous (H) , le score est centré (P_θ) , $n = 1$

$$E_\theta \left[\frac{\partial}{\partial \theta} \log L_1(\theta) \right] = \int_S \frac{\partial}{\partial \theta} \log f_\theta(x) dx = \int_S \frac{\partial}{\partial \theta} \frac{f_\theta(x)}{f_\theta(x)} dx = \int_S \frac{\partial f_\theta}{\partial \theta}(x) dx \stackrel{(H)}{\equiv} \frac{\partial}{\partial \theta} \overline{\int f_\theta(x) dx} = 0$$

DEFINITION 3.12 – L'information de Fisher associé à (X_1, \dots, X_n)

$$I_n(\theta) \stackrel{\text{def}}{\equiv} E_\theta \left[\left(\frac{\partial}{\partial \theta} \log L_{n(\theta)} \right)^2 \right] \stackrel{\text{cor. de la prop 1}}{\equiv} = \text{Var}_\theta \left[\frac{\partial \log L_n(\theta)}{\partial \theta} \right]$$

$$(*) E_\theta \left[\frac{\partial}{\partial \theta} \log f_\theta(X_1) \right]^2 = \int_S \left(\frac{\partial}{\partial \theta} \frac{f_\theta(x)}{f_\theta(x)} \right)^2 f_\theta(x) dx = \int_S \frac{(\frac{\partial}{\partial \theta} f_\theta(x))^2}{f_\theta(x)} = \text{"expression de la definition 1"}$$

EXAMPLE 3.13 – $(X_1, \dots, X_n) \sim \mathcal{E}(\theta)$, $\frac{\partial}{\partial \theta} \log L_n(\theta) = \frac{n}{\theta} - \sum_{i=1}^n X_i$

$$I_n(\theta) = E \left(\left(\frac{n}{\theta} - \sum X_i \right)^2 \right) = n^2 E \left[\left(\frac{1}{\theta} - \frac{\sum X_i}{n} \right)^2 \right] = n^2 \text{Var}(\bar{X}) = n^2 \frac{1}{n} \frac{1}{\theta^2} = \frac{n}{\theta^2}$$

PROPOSITION 3.14 –

$$I_n(\theta) = nI(\theta)$$

en effet,

$$\begin{aligned} I_n(\theta) &= \text{Var} \left(\frac{\partial}{\partial \theta} \log L_n(\theta) \right) = \text{Var} \left(\sum_{i=1}^n \frac{\partial}{\partial \theta} \log f_\theta(X_i) \right) \stackrel{\text{independance}}{\equiv} = \sum_{i=1}^n \text{Var} \left(\frac{\partial}{\partial \theta} \log f_\theta(X_i) \right) = \\ &= n \underbrace{\text{Var} \left(\frac{\partial}{\partial \theta} \log f_\theta(X_1) \right)}_{= I(\theta)} = nI(\theta) \end{aligned}$$

EXAMPLE 3.15 – (X_1, \dots, X_n) i.i.d $\mathcal{P}(\theta)$, $f_\theta(x) = e^{-\theta} \frac{\theta^x}{x!}$

$$\log L_n(\theta) = -n\theta + (\sum X_i) \log \theta - \log \prod_{i=1}^n X_i!$$

$$\frac{\partial}{\partial \theta} \log L_n(\theta) = -n + \frac{\sum X_i}{\theta} \Rightarrow I_n(\theta) = \text{Var} \left(\frac{\sum X_i}{\theta} \right) = \frac{1}{\theta^2} n\theta = \frac{n}{\theta}$$

3.3 Inp. de Fisher et dérivé seconde

PROPOSITION 3.16 – En ajoutant que $\theta \mapsto f_\theta(x)$ est C^2 et que (H) vrai pour $\frac{\partial^2}{\partial \theta^2}$ alors l'info de Fisher s'écrit encore

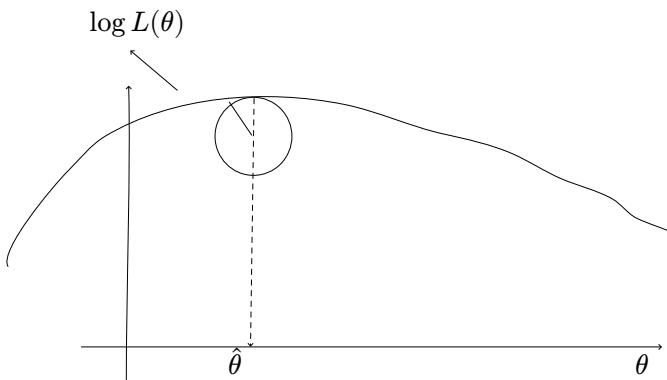
$$I_n(\theta) = -E_\theta \left[\frac{\partial^2 \log L_n(\theta)}{\partial \theta^2} \right]$$

si $\hat{\theta}$ EMV, $I_{n(\hat{\theta})} > 0$

$n = 1$

$$\frac{\partial^2}{\partial \theta^2} \log f_\theta(x) = \frac{\left(\frac{\partial^2 f_\theta(x)}{\partial \theta^2} \right)^2}{f_\theta(x)} - \frac{\left(\frac{\partial f_\theta(x)}{\partial \theta} \right)^2}{f_\theta^2(x)}$$

$$E \left[\frac{\partial^2}{\partial \theta^2} \log f_\theta(X_1) \right] = \int_S \frac{\partial^2 f_\theta(x)}{\partial \theta^2} f_\theta(x) dx - \underbrace{\int_S \frac{\left(\frac{\partial f_\theta(x)}{\partial \theta} \right)^2}{f_\theta^2(x)} dx}_{I(\theta)}$$



Si courbe très “piqué” en l’EMV (i.e. info. Fisher est grande) alors l’EMV est localisé de façon précise

3.4 Inégalité de Cramer - Rao

Soit $g(\theta)$ le paramètre d'intérêt où $g : \Theta \rightarrow \mathbb{R}$

PROPOSITION 3.17 – *Sous les hypothèses d'un modèle régulier, si pour tout θ , $I(\theta) > 0$, alors pour tout estimateur $T = T(X_1, \dots, X_n)$ sans biais, $E_\theta T^2 < +\infty$, on a*

$$\forall \theta \in \Theta, \text{Var}_\theta(T) \geq \frac{(g'(\theta))^2}{I_n(\theta)} = \frac{(g'(\theta))^2}{n I(\theta)}$$

Proof.

$$\begin{aligned}
 & \forall \theta \quad E_{\theta(T)} = g(\theta) \\
 \Rightarrow & \frac{\partial}{\partial \theta} E_{\theta}(T) = g'(\theta) \\
 T=T(X_1) \Leftrightarrow & \frac{\partial}{\partial \theta} \int_S T(x) f_{\theta}(x) dx = g'(\theta) \\
 (H) \Leftrightarrow & \int_S T(x) \frac{\frac{\partial}{\partial \theta} f_{\theta}(x)}{f_{\theta}(x)} f_{\theta}(x) dx = g'(\theta) \\
 \Leftrightarrow & \int_S (T(x) - g(\theta)) \frac{\frac{\partial}{\partial \theta} f_{\theta}(x)}{f_{\theta}(x)} f_{\theta}(x) dx = g'(\theta)
 \end{aligned}$$

□

Inégalité de Cauchy-Schwarz pour $\langle h_1, h_2 \rangle = \int h_1(x) h_2(x) f_{\theta}(x) dx$ avec $h_1(X)$ et $h_2(X)$ centrées

$$\left(\langle T(X) - g(\theta), \frac{\frac{\partial}{\partial \theta} f_{\theta}(x)}{f_{\theta}(x)} \rangle_{\theta} \right)^2 = (g'(\theta))^2 \stackrel{\text{c.s}}{=} \underbrace{\int (T(x) - g(\theta))^2 f_{\theta}(x) dx}_{= \text{Var}_{\theta}(T)} \times \underbrace{\int \left(\frac{\frac{\partial}{\partial \theta} f_{\theta}(x)}{f_{\theta}(x)} \right)^2 f_{\theta}(x) dx}_{= I(\theta)}$$

DEFINITION 3.18 – Si T réalise l'égalité, alors T est dit *efficace*.

Étude asymptotique des estimateurs

§4

Dans un modèle paramétrique régulier, si $\hat{\theta}_n$ estimateur de θ , alors

$$\text{Var}(\hat{\theta}_n) \geq \frac{1}{I_{n(\theta)}} = \frac{1}{nI(\theta)}$$

si $\text{Var}(\hat{\theta}_n) = \frac{1}{nI(\theta)}$ sans biais, $\hat{\theta}_n$ est efficace efficace

Asymptotique: $n \rightarrow +\infty$,

$$n \text{Var}(\hat{\theta}_n) \xrightarrow{n \rightarrow +\infty} \frac{1}{I(\theta)}$$

4.1 Convergences

$(X_n)_{n \geq 0}$ suite de variables aléatoires réelles (\mathbb{R}^d)

- convergence en loi: $X_n \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} X$ ssi $P(X_n \leq x) \rightarrow P(X \leq x)$ en tout point de continuité de x .

LEMMA 4.1 (LEMME DE PORTMANTEAU) – Caractérisations équivalentes:

- Pour toute fonction continue bornée h ,

$$E[h(X_n)] \rightarrow E[h(X)]$$

⇒ la convergence en loi est stable par passage aux fonctions continues (LAC) MAIS il est en général faux que si $X_n \xrightarrow{\mathcal{L}} X$ et $Y_n \xrightarrow{\mathcal{L}} Y$ alors $\begin{pmatrix} X_n \\ Y_n \end{pmatrix} \xrightarrow{\mathcal{L}} \begin{pmatrix} X \\ Y \end{pmatrix}$

Cela est vrai dans 3 cas:

1. si $\begin{cases} \forall n, X_n \text{ et } Y_n \text{ sont indépendantes} \\ X \text{ et } Y \text{ sont indépendantes} \end{cases}$ alors $\begin{cases} \text{convergence en loi de } X_n \text{ et } Y_n \\ \text{convergence en loi du couple } \begin{pmatrix} X_n \\ Y_n \end{pmatrix} \end{cases}$

2.

$$\text{si } \begin{cases} X_n \xrightarrow{P} X \\ Y_n \xrightarrow{P} Y \end{cases} \Rightarrow \begin{pmatrix} X_n \\ Y_n \end{pmatrix} \xrightarrow{P} \begin{pmatrix} X \\ Y \end{pmatrix} \Rightarrow \begin{pmatrix} X_n \\ Y_n \end{pmatrix} \xrightarrow{\mathcal{L}} \begin{pmatrix} X \\ Y \end{pmatrix}$$

3. (Lemme de Slutsky) (le plus important)

$$\text{si } \begin{cases} X_n \xrightarrow{\mathcal{L}} X \\ Y_n \xrightarrow{\mathcal{L}} c \end{cases} \text{ alors } \begin{pmatrix} X_n \\ Y_n \end{pmatrix} \xrightarrow{\mathcal{L}} \begin{pmatrix} X \\ c \end{pmatrix}$$

en appliquant le LAX,

$$\begin{aligned}
 h(x, y) &= x + y & X_n + Y_n &\xrightarrow{\mathcal{L}} X + c \\
 &= xy & X_n Y_n &\xrightarrow{\mathcal{L}} cX \\
 &= \frac{x}{y} & \frac{X_n}{Y_n} &\xrightarrow{\mathcal{L}} \frac{X}{c}
 \end{aligned}$$

4.2 Consistance des estimateurs

DEFINITION 4.2 — $\hat{\theta}_n$ asymptotiquement sans biais si et seulement si

$$\text{Biais}(\hat{\theta}_n, \theta) = E[\hat{\theta}_n] - \theta \xrightarrow{n \rightarrow +\infty} 0$$

REMARK 4.3 — La convergence en proba n'implique pas la convergence des espérances.

Si $X_n \xrightarrow{P} X$, $|X_n| \leq Y \in L'$, alors par convergence dominée $X_n \rightarrow X$ dans L_1

EXAMPLE 4.4 — $\hat{\tau}_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{1}{n} \sum X_i^2 - (\bar{X})^2$ estimateur des moments de $\tau^2 = E[X^2] - (E[X])^2$

$$\text{Biais}(\hat{\tau}_n^2, \tau^2) = -\frac{1}{n} \tau^2 \text{ asymptotiquement sans biais}$$

Consistance de $\hat{\tau}_n^2$?

Outils pour montrer la consistance:

- LGN
- si $R(\hat{\theta}_n, \theta) \rightarrow 0$ alors $\hat{\theta}_n$ consistant car convergence $L^2 \Rightarrow$ convergence en probas
- revenir à la définition de convergence en probas

- si (X_i) sont i.i.d., alors (X_i^2) est i.i.d.

$$E[X_i^2] < +\infty$$

- $\bar{X} \xrightarrow{P} \mu$ (LGN), LAC avec $h(x) = x^2$: $(\bar{X})^2 \xrightarrow{P} \mu^2$
- Donc $\left(\frac{1}{n} \sum_{i=1}^n X_i^2 \right) \xrightarrow{P} \left(\frac{\tau^2 - \mu^2}{\mu} \right)$
- LAC $h(x, y) = x - y^2$

$$\text{Donc } \frac{1}{n} \sum X_i^2 - (\bar{X})^2 \xrightarrow{P} \tau^2 + \mu^2 - \mu^2 = \tau^2$$

4.3 Normalité asymptotique

$\hat{\theta}_n$ pour θ .

→ Question: quelle est la vitesse de convergence de $\hat{\theta}_n$ vers θ ?

(X_1, \dots, X_n) i.i.d., d'espérance θ , $\hat{\theta} = \bar{X}$ de variance $\tau^2(\theta)$

TLC $\sqrt{n}(\bar{X} - \theta) \xrightarrow{\mathcal{L}} Z \sim \mathcal{N}(0, \tau^2(\theta))$ quelle que soit la loi des X_i

DEFINITION 4.5 – $\hat{\theta}_n$ est un estimateur asymptotiquement normal si et seulement si

- vitesse de convergence en \sqrt{n}
- convergence en loi
- loi limite est normale

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{\mathcal{L}} Z \sim \mathcal{N}(0, \tau^2(\theta))$$

EXAMPLE 4.6 – $\hat{\tau}_n^2$ est-elle asymptotiquement normale ?

(X_1, \dots, X_n) i.i.d. d'espérance μ , de variance τ^2

$$\hat{\tau}_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2 + (\bar{X} - \mu)^2 + \underbrace{\frac{2}{n} \sum_{i=1}^n (X_i - \mu)(\mu - \bar{X})}_{=2(\mu - \bar{X})(\bar{X} - \mu)}$$

- TLC: si (X_i) i.i.d., alors les $(X_i - \mu)^2$ sont i.i.d. d'espérance τ^2 ,

$$\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2 - \tau^2 \right) \xrightarrow{\mathcal{L}} Z \sim \mathcal{N}(0, u_4 - \tau^4)$$

$$\text{Var}(X_i - \mu)^2 = E[(X_i - \mu)^4] - \mu^4 = \mu_4 - \tau^4$$

- TLC: $\sqrt{n}(\bar{X} - \mu) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \tau^2)$
- $\sqrt{n}(\hat{\tau}_n^2 - \tau^2) = \sqrt{n} \left(\frac{1}{n} \sum_i (X_i - \mu)^2 - \tau^2 \right) - \underbrace{\sqrt{n}(\bar{X} - \mu)^2}_{\substack{\xrightarrow{\mathcal{L}} \mathcal{N}(0, \tau^2) \\ \xrightarrow{\mathcal{L}, P} 0}}$

$$\begin{aligned} \bar{X} - \mu &\xrightarrow{\mathcal{L}} 0 \\ \sqrt{n}(\bar{X} - \mu) &\xrightarrow{\mathcal{L}} U \sim \mathcal{N}(0, 1) \end{aligned} \left\{ \begin{array}{l} \text{lemme de Slutsky} \\ \implies \end{array} \right. \sqrt{n}(\bar{X} - \mu)^2 \xrightarrow[P]{\mathcal{L}} 0$$

$$\sqrt{n}(\hat{\tau}_n^2 - \tau^2) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} Z + 0$$

Donc $\hat{\tau}_n^2$ est un estimateur asymptotiquement normal

REMARK 4.7 – $\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, \tau^2) \iff \frac{\sqrt{n}(\hat{\theta}_n - \theta)}{\tau} \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, 1)$ Application du lemme de Slutsky: si $\hat{\tau}^2$ est un estimateur consistant de τ^2 , alors on a encore

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

Proof.

$$\frac{\sqrt{n}(\hat{\theta}_n - \theta)}{\hat{\tau}} = \underbrace{\left(\sqrt{n} \frac{(\hat{\theta}_n - \theta)}{\tau} \right)}_{\xrightarrow{\mathcal{L}} Z \sim \mathcal{N}(0, 1)} \times \underbrace{\left(\frac{\tau}{\hat{\tau}} \right)}_{\xrightarrow{P} 1}$$

$$\xrightarrow{\mathcal{L}} 1 \times Z \text{ par Slutsky et consistance de } \hat{\tau}$$

□

4.4 δ -méthode

$\hat{\theta}$ estimateur asymptotiquement normal: quelle est la loi asymptotique de $g(\theta)$?

LEMMA 4.8 (MÉTHODE DÉLTA) — Soit Z_n suite de variables aléatoires réelles t.q.

$$\sqrt{n}(Z_n - \mu) \xrightarrow{\mathcal{L}} Z \sim \mathcal{N}(0, \tau^2)$$

Soit g une fonction dérivable, $g'(\mu) \neq 0$. Sous ces hypothèses, on a

$$\sqrt{n}[g(Z_n) - g(\mu)] \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \tilde{Z} \sim \mathcal{N}\left(0, (g'(\mu))^2 \tau^2\right)$$

$$g(x) = g(\mu) + g'(\mu)(x - \mu) + (x - \mu)R(x - \mu) \quad \text{où } R(y) \xrightarrow[y \rightarrow 0]{} 0$$

$$\begin{aligned} \sqrt{n}(g(Z_n) - g(\mu)) &= g'(\mu) \underbrace{\sqrt{n}(Z_n - \mu)}_{\xrightarrow{\mathcal{L}} Z \sim \mathcal{N}(0, \tau^2)} + \underbrace{(\sqrt{n})(Z_n - \mu)R(Z_n - \mu)}_{\xrightarrow{\mathcal{L}} \mathcal{N}(0, \tau^2) \xrightarrow{P} 0?} \\ &\xrightarrow{\mathcal{N}(0, (g'(\mu))^2 \tau^2)} \end{aligned}$$

A-t-on $Z_n \xrightarrow{P} \mu$?

$$\begin{aligned} P(|X_n - \mu| > \varepsilon) &= P\left(\frac{\sqrt{n}|Z_n - \mu|}{\tau} > \frac{\sqrt{n}\varepsilon}{\tau}\right) \\ &= P\left(\frac{\sqrt{n}(Z_n - \mu)}{\tau} > \frac{\sqrt{n}\varepsilon}{\tau}\right) + P\left(\frac{\sqrt{n}(Z_n - \mu)}{\tau} < -\frac{\sqrt{n}\varepsilon}{\tau}\right) \\ &\sim 1 - \Phi_n\left(\frac{\sqrt{n}\varepsilon}{\tau}\right) + \Phi_n\left(-\frac{\sqrt{n}\varepsilon}{\tau}\right) = 2\left(1 - \Phi\left(\frac{\sqrt{n}\varepsilon}{\tau}\right)\right) \end{aligned}$$

Fonction de répartition empirique

§5

(X_1, \dots, X_n) échantillon i.i.d. à valeurs réelles de loi F inconnue.

$$\forall x \in \mathbb{R}, F(x) = P(X_1 \leq x) = E[\mathbb{1}_{X_1 \leq x}]$$

DEFINITION 5.1 – La fonction de répartition empirique associée à (X_1, \dots, X_n) est définie par:

$$\hat{F}_n : \mathbb{R} \longrightarrow [0, 1]$$

$$x \mapsto \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{X_i \leq x}$$

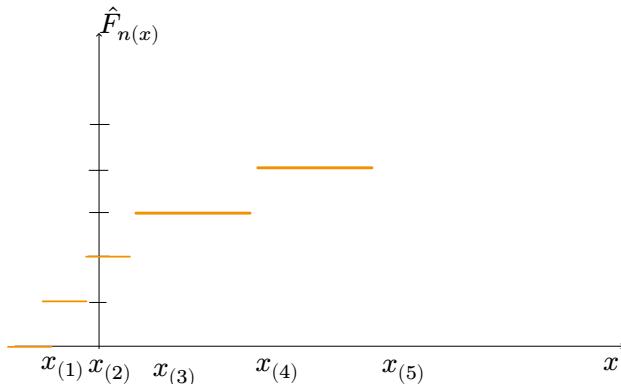
$\forall x \in \mathbb{R}, \hat{F}_n(x)$ est une variable aléatoire, estimateur de $F(x)$.

DEFINITION 5.2 – Loi empirique $P_n = \frac{1}{n} \sum_{i=1}^n \delta_{X_i}$ est une loi discrète uniforme sur $\{X_1, \dots, X_n\}$.

Représentation graphique

Conditionnellement $X_1 = x_1, X_2 = x_2, \dots, X_n = x_n$

$x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$ valeurs ordonnées



fdsa

PROPOSITION 5.3 (PROPRIÉTÉS IMMÉDIATS) –

- $n\hat{F}_n(x) = \sum_{i=1}^n \mathbb{1}_{X_i \leq x}$ suit la loi binomiale($n, F(x)$)
- $R(\hat{F}_n(x), F(x)) = 0 + \frac{1}{n^2} \text{Var}\left(\sum_{i=1}^n \mathbb{1}_{X_i \leq x}\right) \stackrel{\text{indep}}{\equiv} \frac{1}{n} F(x)(1 - F(x)) \xrightarrow{n \rightarrow +\infty} 0$ donc $\forall x \in \mathbb{R}, \hat{F}_n(x) \xrightarrow{P} F(x)$
- ou bien LGN: $\hat{F}_n(x)$ estimateur consistant de $F(x)$.

- On a un résultat de convergence uniforme :

$$\sup_{x \in \mathbb{R}} |\hat{F}_n(x) - F(x)| \xrightarrow[n \rightarrow +\infty]{P} 0 \quad (\text{Théorème de Glivenko-Cantelli})$$

- $\hat{F}_n(x)$ est-il asymptotiquement normal?

$$\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{X_i \leq x}$$

TLC: les X_i sont i.i.d., donc les $\{\mathbb{1}_{X_i \leq x} = Y_i\}$ sont i.i.d.

$$\forall x, F(x) \in]0, 1[, \quad \sqrt{n}(\hat{F}_n(x) - F(x)) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, F(x)(1 - F(x)))$$

$$\iff \frac{\hat{F}_n(x) - F(x)}{\sqrt{\frac{F(x)(1 - F(x))}{n}}} \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, 1)$$

5.1 Estimation empirique

“plug-in” ou méthode de substitution, paramètre d’intérêt $\theta = c(F)$, la méthode empirique définit $\hat{\theta}$, estimateur empirique en remplaçant F par $\hat{F}_n \rightarrow \hat{\theta}_n = c(\hat{F}_n)$.

EXAMPLE 5.4 – $\theta = E_F(X) \rightarrow \hat{\theta}_n = E_{\hat{F}_n}(X) = \sum_{i=1}^n X_i \times \frac{1}{n} = \bar{X}$ si X_i distinctes

$$\theta = \text{Var}_F(X) \rightarrow \hat{\theta}_n = \text{Var}_{\hat{F}_n}(X) = \frac{1}{n} \sum (X_i - \bar{X})^2$$

5.2 Inverse généralisé

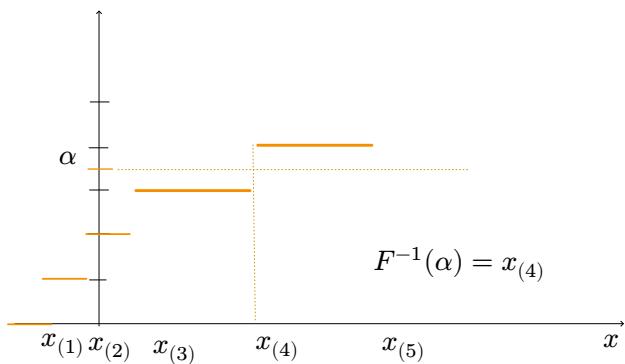
DEFINITION 5.5 – On définit l’inverse généralisé de F par:

$$F^{-1} : [0, 1] \longrightarrow \mathbb{R}$$

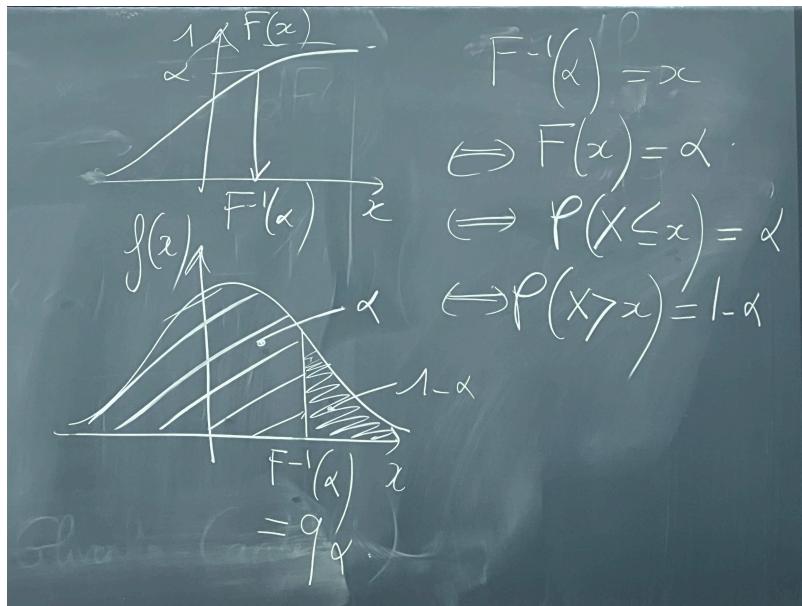
$$\forall \alpha \in [0, 1], F^{-1}(\alpha) = \inf\{x \in \mathbb{R}, F(x) \geq \alpha\}$$

Si F est strictement croissante, $\inf x$ tel que $F(x) \geq a \Leftrightarrow x \geq F^{-1}(a)$, si F est la fonction d’une loi discrète.

²Thm de Glivenko-Cantelli: https://fr.wikipedia.org/wiki/Th%C3%A9or%C3%A8me_de_Glivenko-Cantelli


EXAMPLE 5.6 –

$$F^{-1}(\alpha) = x \Leftrightarrow F(x) = \alpha \Leftrightarrow P(X \leq x) = \alpha \Leftrightarrow P(X > x) = 1 - \alpha$$



Vocab:

- F^{-1} s'appelle aussi la fonction quantile
- $F^{-1}(\alpha)$ = quantile d'ordre α , de la loi F
- $F^{-1}\left(\frac{1}{4}\right)$ = 1er quantile
- $F^{-1}\left(\frac{1}{2}\right)$ = médiane
- $F^{-1}\left(\frac{3}{4}\right)$ = 3eme quantile

LEMMA 5.7 – *U variable aléatoire sur $[0, 1]$, F f.r., alors $F^{-1}(U)$ est une variable aléatoire de loi F*

- Si F bijective:

$$P(F^{-1}(U) \leq x) \underset{F \text{ bijective}}{\equiv} P(U \leq F(x)) \underset{\text{car } P(U \leq x) = x \text{ sur } [0, 1]}{\equiv} F(x)$$

- Si F discrète: F^{-1} inverse généralisé: $F^{-1}(y) \leq x \Leftrightarrow y \leq F(x)$

5.3 Quantile empirique

DEFINITION 5.8 – On définit le quantile empirique (sampe quantile) d'ordre α , comme étant le quantile de \hat{F}_n :

$$\hat{q}_{n,\alpha} = \hat{F}_n^{-1}(\alpha) = \inf\{x, \hat{F}_n(x) \geq \alpha\}$$

PROPOSITION 5.9 –

- On peut montrer que $\hat{q}_{n,\alpha} = X_{(\lceil n\alpha \rceil)}$ où $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$ est l'échantillon ordonné des $(X_i)_{1 \leq i < n}$

$$[u] = \text{le plus petit entier } \geq u$$

EXAMPLE 5.10 – $\alpha = \frac{1}{2}, [\frac{n}{2}],$

$$\begin{cases} \text{si } n = 2k & \text{medianne} = \hat{q}_{n,\frac{1}{2}} = X_{(k)} \\ \text{si } n = 2k + 1 & \text{medianne} = \hat{q}_{n,\frac{1}{2}} = X_{(k+1)} \end{cases}$$

- Consistance

si $\alpha \in]0, 1[, si F est strictement croissante au voisinage de \alpha$

Intervalles de confiance

§6

6.1 Définitions

(X_1, \dots, X_n) i.i.d. de loi $P \in \{P_\theta, \theta \in \Theta \subset \mathbb{R}^p\}$, on s'intéresse à $\theta \in \mathbb{R}$ ou $g(\theta) : \mathbb{R}^p \rightarrow \mathbb{R}$.

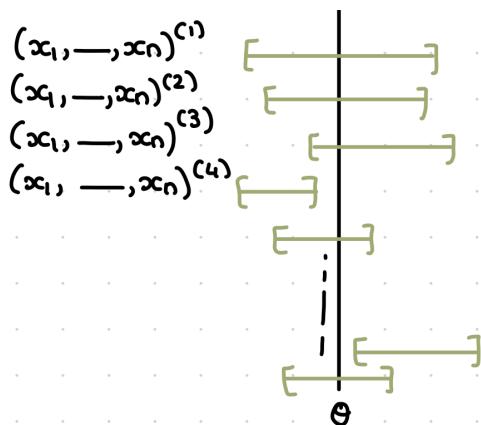
Un intervalle de confiance pour θ , de niveau de confiance $1 - \alpha, \alpha \in]0, 1[$ est un intervalle dont les bornes sont aléatoires, fonctions de l'échantillon et ne dépend PAS des paramètres inconnus du modèle et tel que

$$P([B \inf(X_1, \dots, X_n); B \sup(X_1, \dots, X_n)] \ni \theta) \geq 1 - \alpha$$

³

- Un IC est calculable à partir des données
- si l'inégalité est une égalité = niveau de confiance est exact.
- si on a $P(\theta \in [B \inf, B \sup]) \xrightarrow{n \rightarrow +\infty} 1 - \alpha$, niveau est asymptotique.
- en général $\alpha = 1\%, 5\%$

6.2 Interprétation



$IC = [B \inf(X_1, \dots, X_n), B \sup(X_1, \dots, X_n)]$ formule mathématique qui g'arantit le niveau $1 - \alpha$. On observe $X_1 = x_1, X_2 = x_2, \dots, X_n = x_n$, une réalisation de l'échantillon aléatoire. On calcule $IC = [2.3; 5.1]$ de niveau de confiance 95% ($\alpha = 5\%$).

En moyenne, sur 100 intervalles calculés (avec la même formule), il y a 5 intervalles qui ne contiennent pas θ .

$$P(\theta \in [B \inf, B \sup]) = 1 - \alpha$$

$$\underline{P(\theta \in [2.3, 5.1]) = 95\%} \text{ car } \theta \text{ un nombre}$$

6.3 Méthode pivotale

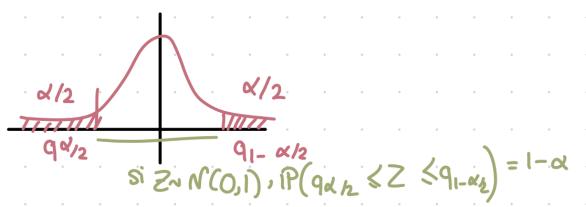
(X_1, \dots, X_n) i.i.d. d'espérance $\theta \in \mathbb{R}$, de variance $\sigma^2(\theta)$. Soit $\hat{\theta}$, asymptotiquement normal:

$$\begin{aligned} \sqrt{n}(\hat{\theta}_n - \theta) &\xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, \sigma^2(\theta)) \\ \Leftrightarrow \frac{\sqrt{n}(\hat{\theta}_n - \theta)}{\sigma(\theta)} &\xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, 1) \end{aligned}$$

Par définition des quantiles gaussiens, $q_\alpha = \Phi^{-1}(\alpha)$ où Φ f.r. de $\mathcal{N}(0, 1)$

³ $B \inf$ pour borne inférieure et $B \sup$ pour borne supérieure

$$P\left(q_{\frac{\alpha}{2}} \leq \frac{\sqrt{n}(\hat{\theta} - \theta)}{\sigma(\theta)} \leq q_{1-\frac{\alpha}{2}}\right) \xrightarrow[n \rightarrow +\infty]{} 1 - \alpha$$



- pivot ou statistique pivotale $= \frac{\sqrt{n}(\hat{\theta} - \theta)}{\hat{\sigma}}$ statistique centrée réduite issue de $\hat{\theta}$, où $\sigma^2(\theta)$ estimé par $\hat{\sigma}^2$, consistant pour estimer $\sigma^2(\theta)$.

Si c'est le cas,

$$\underbrace{\frac{\sqrt{n}(\hat{\theta} - \theta)}{\sigma(\theta)}}_{\hat{\theta} \text{ as. normal}} \times \underbrace{\frac{\sigma^2(\theta)}{\hat{\sigma}^2}}_{\substack{\xrightarrow{\mathcal{L}} \text{estimateur consistant} \\ \xrightarrow{P} 1}} \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, 1) \text{ par lemme de Slutsky}$$

- on en déduit

$$P\left(q_{\frac{\alpha}{2}} \leq \frac{\sqrt{n}(\hat{\theta} - \theta)}{\hat{\sigma}} \leq q_{1-\frac{\alpha}{2}}\right) \xrightarrow[n \rightarrow +\infty]{} 1 - \alpha$$

$$P\left(\hat{\theta} - \frac{1}{\sqrt{n}}\hat{\sigma}q_{1-\frac{\alpha}{2}} \leq \theta \leq \hat{\theta} - \frac{1}{\sqrt{n}}\hat{\sigma}q_{\frac{\alpha}{2}}\right) \rightarrow 1 - \alpha$$

Compléments (avant partiel)

§7

1. Retour sur normalité asymptotique
2. Exemple
3. Pivot asymptotique
4. Exemple 2

7.1 Propriétés asymptotiques d'une suite d'estimateurs $(\hat{\theta}_n)_{n \geq 1}$

- Consistance $\hat{\theta}_n \xrightarrow{P} \theta$
- Normalité asymptotique, s'il existe $\sigma^2 > 0$

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

De façon générale, s'il existe $v_n \xrightarrow{n \rightarrow +\infty} +\infty$

$$v_n(\hat{\theta}_n - \theta) \xrightarrow{\mathcal{L}} Y$$

On dit que $\hat{\theta}_n$ converge à la vitesse $\frac{1}{v_n}$

REMARK 7.1 — Si $\hat{\theta}_n$ asymptotiquement normal $\Rightarrow \hat{\theta}_n$ consistant

$$\hat{\theta}_n - \theta = \underbrace{\frac{1}{\sqrt{n}}}_{\xrightarrow{\rightarrow 0}} \underbrace{\sqrt{n}(\hat{\theta}_n - \theta)}_{\xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2)} \xrightarrow[\text{Slutsky}]{\mathcal{L} \text{ ou } P} 0$$

$$U_n = \frac{1}{\sqrt{n}} \rightarrow 0$$

δ -méthode

$$\begin{aligned} \sqrt{n}(X_n - 1) &\xrightarrow{\mathcal{L}} \mathcal{N}(0, 1) \\ \sqrt{n}(X_n - 1) &\xrightarrow{\mathcal{L}} Z \sim \mathcal{N}(0, 1) \\ X_n &\xrightarrow{\mathcal{L}} 1 + \frac{1}{\sqrt{n}}Z \end{aligned}$$

Si g dérivable en 1,

$$g(1 + h) = g(1) + hg'(1)$$

$$g(X_n) \approx g(1) + \frac{1}{\sqrt{n}}g'(1)Z$$

$$\sqrt{n}(g(X_n) - g(1)) \approx g'(1)Z$$

δ -méthode

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{\mathcal{L}} Z \sim \mathcal{N}(0, 1)$$

g dérivable en θ

$$g(x) = g(\theta) + g'(\theta)[(x - \theta) + r(x)] \text{ où } r(x) \xrightarrow{x \rightarrow 0} 0$$

$$\hat{\theta}_n \xrightarrow{P} \theta \text{ donc (LAC)} r(\hat{\theta}_n) \rightarrow r(\theta) = 0$$

$$g(\hat{\theta}_n) = g(\theta) + (\hat{\theta}_n - \theta)[g'(\theta) + r(\hat{\theta}_n)]$$

$$\sqrt{n}(g(\hat{\theta}_n) - g(\theta)) = \underbrace{\sqrt{n}(\hat{\theta}_n - \theta)}_{\xrightarrow{\mathcal{L}} Z} \underbrace{\left[g'(\theta) + r(\hat{\theta}_n) \right]}_{\xrightarrow{P} g'(\theta)} \xrightarrow{\text{Slutsky}} \sqrt{n}(g(\hat{\theta}_n) - g(\theta)) \xrightarrow{\mathcal{L}} g'(\theta)Z \sim \mathcal{N}(0, (g'(\theta))^2)$$

EXAMPLE 7.2 – X_1, \dots, X_n de loi de densité $f(x) = \frac{1}{\mu}e^{-\frac{x}{\mu}}, x \geq 0, \mu = E[X_i] > 0$

$$\mu \text{ estimé par } \hat{\mu} = \bar{X} \text{ efficace? } \log L_n(\mu) = -n \log \mu - \frac{1}{\mu} \sum_{i=1}^n X_i$$

$$\text{Var}(\hat{\mu}) = \frac{1}{n^2} \text{Var}\left(\sum_i X_i\right) \underset{\text{indép.}}{\equiv} \frac{1}{n^2} \sum_i \text{Var}(X_i) \underset{\text{i.i.d.}}{\equiv} \frac{1}{n} \text{Var}(X_i) = \frac{\mu^2}{n}, E[\hat{\mu}] = \mu$$

$$\frac{\partial}{\partial \mu}(\log L_n)(\mu) = -\frac{n}{\mu} + \frac{1}{\mu^2} \sum (X_i)$$

$$\begin{aligned} I_{n(\mu)} &= \text{Var}\left(-\frac{n}{\mu} + \frac{1}{\mu^2} \sum X_i\right) \\ &= \frac{1}{\mu^4} \text{Var}\left(\sum X_i\right) \\ &= \frac{n}{\mu^4} \text{Var}(X_i) \end{aligned}$$

$$I_{n(\mu)} = \frac{n}{\mu^2}$$

$\hat{\mu}$ sans biais et $\text{Var}(\hat{\mu}) = \frac{1}{I_{n(\mu)}}$. Donc $\hat{\mu}$ est efficace.

TLC: $\sqrt{n}(\hat{\mu}_n - \mu) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \mu^2) \Rightarrow \text{Var}(\hat{\mu}_n) = \frac{\mu^2}{n} = \text{variance de la loi gaussienne asymptotique}$

$\frac{\sqrt{n}(\hat{\mu}_n - \mu)}{\mu}$ a pour loi asymptotique $\mathcal{N}(0, 1)$

- autre paramétrisation: (X_1, \dots, X_n) i.i.d. $f(x) = \theta e^{-\theta x}, x \geq 0$

$$EX_i = \frac{1}{\theta}, \text{Var } X_i = \frac{1}{\theta^2}$$

$$\log L_n(\theta) = n \log \theta - \theta \sum_{i=1}^n X_i$$

$$\frac{\partial}{\partial \theta}(\log L_n)(\theta) = \frac{n}{\theta} - \sum X_i \hookrightarrow \hat{\theta}^{\text{MV}} = \frac{1}{\bar{X}}$$

$$\hookrightarrow I_{n(\theta)} = \text{Var}\left(\frac{n}{\theta} - \sum X_i\right) = \text{Var}\left(\sum X_i\right) = \frac{n}{\theta^2}$$

REMARK 7.3 – cd TD1: $n\bar{X} \sim \Gamma(n, \theta)$

$$E\left[\frac{1}{n\bar{X}}\right] = \frac{\theta}{n-1} \text{ et } \text{Var}\left(\frac{1}{(n\bar{X})^2}\right) = \frac{\theta^2}{(n-1)(n-2)}$$

$$E\left[\frac{1}{\bar{X}}\right] = n \frac{\theta}{n-1}$$

$$\hookrightarrow \tilde{\theta} = \frac{n-1}{n} \hat{\theta} \text{ non biaisé}$$

$$\begin{aligned} \text{Var}(\tilde{\theta}) &= \left(\frac{n-1}{n}\right)^2 \text{Var}\left(\frac{1}{\bar{X}}\right) = \frac{(n-1)^2}{n^2} \left[E\left[\frac{1}{(\bar{X})^2}\right] - \left(E\left[\frac{1}{\bar{X}}\right]\right)^2 \right] \\ &= \frac{(n-1)^2}{n^2} \times \frac{n^2\theta^2}{(n-1)(n-2)} - \frac{(n-1)^2}{n^2} \frac{n^2}{(n-1)^2} \theta^2 \\ &= \theta^2 \frac{n-1}{n-2} - \theta^2 = \frac{\theta^2}{n-2} \underset{\text{BCR}}{\gtrsim} \frac{1}{I_{n(\theta)}} \text{ non efficace} \end{aligned}$$

$$\sqrt{n}\left(\bar{X} - \frac{1}{\theta}\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \frac{1}{\theta^2}\right)$$

$\hat{\theta}$ est asymptotiquement efficace

\bar{X} asymptotiquement normal (TLC). $g(x) = \frac{1}{x}$ sur $]0, +\infty[$, $g'(x) = -\frac{1}{x^2} \neq 0$, méthode delta:

$$\sqrt{n}\left(\frac{1}{\bar{X}} - \theta\right) \xrightarrow{\substack{\mathcal{L} \\ =\theta^2}} \underbrace{g'\left(\frac{1}{\theta}\right)}_{=\theta^2} \mathcal{N}\left(0, \frac{1}{\theta^2}\right) = \mathcal{N}\left(0, \frac{\theta^4}{\theta^2} = \theta^2\right)$$

7.2 Pivot (asymptotique) ou statistique pivotale

DEFINITION 7.4 – Statistique dont la loi ne dépend pas de paramètres inconnus

EXAMPLE 7.5 – X_1, \dots, X_n i.i.d. Bernoulli(θ) avec $\theta \in]0, 1[$:

$$\begin{aligned} \sqrt{n}\left(\bar{X} - \theta\right) &\xrightarrow{\substack{\mathcal{L} \\ \text{TLC}}} \mathcal{N}(0, \theta(1-\theta)) \\ \Leftrightarrow \sqrt{n} \underbrace{\frac{\bar{X} - \theta}{\sqrt{\theta(1-\theta)}}}_{\text{pivot ou stat. pivotale}} &\xrightarrow{\mathcal{L}} \mathcal{N}(0, 1) \end{aligned}$$

méthode pivotale pour IC: On estime $\sqrt{\theta(1-\theta)}$ par $\sqrt{\hat{\theta}(1-\hat{\theta})}$ “plug-in” par le LAC $g(x) = \sqrt{x(1-x)}$ avec $x \in]0, 1[$, $\sqrt{\hat{\theta}(1-\hat{\theta})}$ estimateur consistant de $\sqrt{\theta(1-\theta)}$

$$\sqrt{n} \frac{\hat{\theta} - \theta}{\sqrt{\hat{\theta}(1-\hat{\theta})}} = \underbrace{\sqrt{n} \frac{\hat{\theta} - \theta}{\sqrt{\theta(1-\theta)}}}_{\xrightarrow{\text{TLC}} \mathcal{N}(0,1)} \times \underbrace{\frac{\sqrt{\theta(1-\theta)}}{\sqrt{\hat{\theta}(1-\hat{\theta})}}}_{\substack{\xrightarrow{P} 1 \\ \text{consistant}}}$$

EXAMPLE 7.6 — (X_1, \dots, X_n) de densité $\theta > 0$. $f_\theta(x) = \frac{3}{\theta} x^2 \exp\left(-\frac{x^3}{\theta}\right) \mathbb{1}_{x \geq 0}$

EMV?

$$\log L_n(\theta) = n(\log^3 - \log \theta) + \sum_{i=1}^n \log(X_i^2) - \frac{1}{\theta} \sum_{i=1}^n X_i^3$$

$$(\log L_n)'(\theta) = -\frac{n}{\theta} + \frac{1}{\theta^2} \sum X_i^3 \Rightarrow \hat{\theta} = \frac{\sum X_i^3}{n}$$

$$(\log L_n)''(\theta) = \frac{n}{\theta^2} - \frac{2}{\theta^3} \sum X_i^3; (\log L_n)''(\hat{\theta}) = \frac{n}{\hat{\theta}^2} - \frac{2}{\hat{\theta}^3} n \hat{\theta} = -\frac{n}{\hat{\theta}^2} < 0 + \text{unicité}$$

\Rightarrow max global

TLC:

$$\begin{aligned} \sqrt{n}(\hat{\theta}_n - \theta) &\xrightarrow{\mathcal{L}} \mathcal{N}(0, \theta^2) \\ \Leftrightarrow \underbrace{\frac{\sqrt{n}(\hat{\theta} - \theta)}{\theta}}_{\substack{\text{pivot asymptotique}}} &\xrightarrow{\mathcal{L}} \mathcal{N}(0, 1) \\ \underbrace{\frac{\sqrt{n}(\hat{\theta} - \theta)}{\hat{\theta}}}_{\substack{\text{Slutsky}}} &\xrightarrow{\mathcal{L}} \mathcal{N}(0, 1) \end{aligned}$$

$q_{\frac{\alpha}{2}}$ et $q_{1-\frac{\alpha}{2}}$ quantiles de $\mathcal{N}(0, 1)$

$$P\left(q_{\frac{\alpha}{2}} \leq \frac{\sqrt{n}(\hat{\theta} - \theta)}{\hat{\theta}} \leq q_{1-\frac{\alpha}{2}}\right) \xrightarrow{n \rightarrow +\infty} 1 - \alpha$$

$$P\left(q_{\frac{\alpha}{2}} \frac{\hat{\theta}}{\sqrt{n}} \leq \hat{\theta} - \theta \leq q_{1-\frac{\alpha}{2}} \frac{\hat{\theta}}{\sqrt{n}}\right) \rightarrow 1 - \alpha$$

$$P\left(\underbrace{\hat{\theta} - q_{1-\frac{\alpha}{2}} \frac{\hat{\theta}}{\sqrt{n}}}_{\Rightarrow IC(\theta) \text{ de niveau asymptotique } (1-\alpha)} \leq \theta \leq \hat{\theta} - q_{\frac{\alpha}{2}} \frac{\hat{\theta}}{\sqrt{n}}\right) \rightarrow 1 - \alpha$$