



* Evaluation: 0.4CC + 0.6 examen
 ↑ 80% partiel, 20% Interro 26/01

1. Modèle Statistique

DEF: un modèle statistique est un espace de proba $(\Omega, \mathcal{F}, P)$ où P est une famille de loi de proba $\{P_\theta; \theta \in \Theta\}$
 si $\exists p \in \mathbb{N}^*$, $\Theta \subset \mathbb{R}^p$: modèle paramétrique

sinon: modèle non paramétrique

Exemples de familles: $\mathcal{P} = \{P(\lambda); \lambda > 0\}$ les lois de poisson

$\mathcal{P} = \{P \text{ loi de probas}\} = \{P; P \text{ dont la densité admet une dérivée seconde bornée}\}$

DEF: une observation est une v.a. dont la loi $\in \{P_\theta, \theta \in \Theta\}$

Notre observation aura une structure de n-échantillons $X_1, \dots, X_n$ des v.a. indep (réelles) identiquement distribuées, de loi commune $\in \{P_\theta, \theta \in \Theta\}$

Rq: $(X_1, \dots, X_n)$ de loi $P_\theta^{\otimes n}$

↳ l'échantillon contient toute l'info sur P_θ , donc sur θ .

DEF: un modèle est identifiable ssi $\theta \mapsto P_\theta$ est injective

Rq:
 observation $\begin{cases} \rightarrow X_i \text{ v.a.} \\ \rightarrow x_i = \text{donnée} \leadsto \text{réalisation de } X_i \end{cases}$

2. Estimateurs

Hyp: 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
 ($\exists \theta^*$ inconnu tq $P_{X_i} = P_{\theta^*}$)

DEF: [estimateur] un estimateur de θ est une fonction de l'échantillon $(X_1, \dots, X_n)$ (mesurable) indep de θ
 (calculable à partir des données)

Notation: $\hat{\theta} = \hat{\theta}_n = h(X_1, \dots, X_n)$. ($\hat{\theta}_n$ est une v.a.)

Ex: $\hat{\theta} = 0$, $\hat{\theta} = X_1 - X_3 \dots$

↳ 1) Comment définir un bon estimateur?

2) Comment construire un bon estimateur?

3. Risque quadratique

Idée: En moyenne, $\hat{\theta}$ doit être proche de $\theta \rightarrow \mathbb{E}(\hat{\theta} - \theta)$

DEF: [biais de $\hat{\theta}$]

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

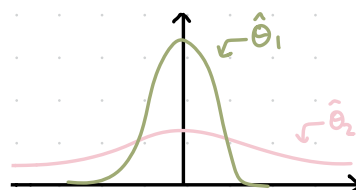
On dit que $\hat{\theta}$ est sans biais si $B(\hat{\theta}, \theta) = 0$

DEF: [Risque quadratique]

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

MSE (mean squared error)

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



Exemple: [modèle de Poisson]

$X_1, \dots, X_n$ i.i.d de loi P_θ de Poisson, $\theta > 0$

Estimateur de θ ? $\theta = \mathbb{E}[X_i]$

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

$$B(\hat{\theta}, \theta) = \mathbb{E} \left[\frac{1}{n} \sum_{i=1}^n X_i \right] - \theta$$

linéarité

$$\downarrow = \frac{1}{n} \sum_{i=1}^n \mathbb{E}[X_i] - \theta$$

$$\stackrel{\text{i.i.d}}{=} \frac{1}{n} \cdot n \mathbb{E}[X_1] - \theta$$

$$= \mathbb{E}[X_1] - \theta = \theta - \theta = 0$$

$$\leadsto \mathbb{E}[\bar{X}] = \theta$$

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

$$= \text{Var}(\bar{X})$$

$$= \text{Var}\left(\frac{1}{n} \sum_{i=1}^n X_i\right)$$

$$= \frac{1}{n^2} \text{Var}\left(\sum_{i=1}^n X_i\right)$$

$$\stackrel{\text{indép}}{=} \frac{1}{n^2} \sum_{i=1}^n \text{Var}(X_i) \stackrel{\text{i.i.d}}{=} \frac{1}{n} \cdot \text{Var}(X_1)$$

PROP: [décomposition biais-variance du risque]

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

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

$$\stackrel{\text{def}}{=} \mathbb{E}[(\hat{\theta} - \mathbb{E}[\hat{\theta}] + \mathbb{E}[\hat{\theta}] - \theta)^2]$$

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

$$= \text{Var}(\hat{\theta}) + (\text{B}(\hat{\theta}, \theta))^2 + 2\text{B}(\hat{\theta}, \theta)(\mathbb{E}[\hat{\theta}] - \mathbb{E}[\hat{\theta}])$$

4. Consistance

↳ propriété asymptotique

on ne considère que des constants.

DEF: $(X_1, \dots, X_n)$ i.i.d de loi P_θ , $\theta \in \Theta \subset \mathbb{R}^p$

Soit $\hat{\theta}_n = h(X_1, \dots, X_n)$. On considère la suite de v.a. $(\hat{\theta}_n)_{n \geq 1}$.

$\hat{\theta}_n$ est un estimateur consistant (ou convergent) de θ ssi : $\hat{\theta}_n \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} \theta$

Rq: $(\hat{\theta}_n)$ est fortement consistant ssi : $\hat{\theta}_n \xrightarrow[n \rightarrow +\infty]{\text{p.s.}} \theta$

Exemple [modèle de poisson]

$$\theta = \mathbb{E}[X_i]; \hat{\theta}_n = \bar{X}$$

* On peut invoquer la LGN : $\bar{X} \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} \theta = \mathbb{E}[X_i]$

$$\begin{aligned} * R(\hat{\theta}, \theta) &= \text{Var}(\bar{X}) = \frac{\theta}{n} \xrightarrow[n \rightarrow +\infty]{} 0 \\ &= \mathbb{E}[(\hat{\theta} - \theta)^2] \end{aligned}$$

$$\text{Ineq. BT: } \forall \varepsilon > 0, \mathbb{P}(|\hat{\theta} - \theta| > \varepsilon) \leq \frac{\mathbb{E}|\hat{\theta} - \theta|^2}{\varepsilon^2} = \frac{R(\hat{\theta}, \theta)}{\varepsilon^2} \xrightarrow{} 0$$

• $(X_1, \dots, X_n)$ i.i.d. Poisson(θ).

On veut estimer $\beta = \mathbb{P}(X_i = 0) = e^{-\theta}$

↳ $\hat{\beta} = e^{-\hat{\theta}} = e^{-\bar{X}}$ [méthode "plug-in"]

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

LEMME [de l'application continue]

Si $Z_n \xrightarrow{\mathbb{P}} Z$, alors $h(Z_n) \xrightarrow{\mathbb{P}} h(Z)$ $\forall h$ fonction continue

Résumé Cours 1:

Cadre paramétrique

Modèle statistique paramétrique:

On dispose d'une observation $(X_1, \dots, X_n)$, un échantillon de v.a. i.i.d. de loi commune P appartenant à une famille de lois de probabilités paramétrée $\{P_\theta, \theta \in \mathcal{M} \subset \mathbb{R}^p\}$

Rq: si $\mathcal{M} \subset \mathbb{R}^p$ espace de dim infinie $\rightarrow$ modèle non-paramétrique

Question: estimer P , c'est estimer $\theta \in \mathbb{R}^p$

Exemples: Bernoulli (θ) , Exponentielle (θ) , gaussienne (μ, σ^2) , lois de densité $f_\theta(x) = \theta x^{\theta-1} \mathbb{1}_{x \in [0,1]}$

Modèle identifiable: $\theta \rightarrow P_\theta$ injective

Notation: $\mathbb{E}_\theta[h(X_1, \dots, X_n)]$, $\text{Var}_\theta[h(X_1, \dots, X_n)]$.
loi de $(X_1, \dots, X_n) \rightarrow P_\theta^{\otimes n}$

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

Qualité: • risque: $R(\hat{\theta}, \theta) = \mathbb{E}((\hat{\theta} - \theta)^2)$

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

1 Méthode des moments

a) Définitions

On appelle moment théorique de la loi des X_i d'ordre k : $\mu_k = \mathbb{E}(X_i^k)$, $k \geq 1$

// 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 LGN, $\hat{\mu}_k \xrightarrow[n \rightarrow +\infty]{P} \mu_k$

La méthode des moments: si on peut écrire θ ou $g(\theta)$ paramètre d'intérêt comme une fonction des K premiers moments théorique: $\theta = \psi(\mu_1, \mu_2, \dots, \mu_K)$. Alors l'estimateur $\hat{\theta} = \psi(\hat{\mu}_1, \hat{\mu}_2, \dots, \hat{\mu}_K)$ est obtenu par la méthode des moments.

b) Exemples:

• $X_i \sim \text{Bernoulli}(\theta)$ à valeurs 0-1, $\theta = P(X_i = 1) = \mathbb{E}[X_i]$
 $\hookrightarrow \hat{\theta} = \frac{1}{n} \sum_{i=1}^n X_i = \bar{X}$

• $X_i \sim \text{Exp}(\theta)$, $f_\theta(x) = \theta e^{-\theta x} \mathbb{1}_{x \geq 0}$. $\mathbb{E}[X_i] = 1/\theta \Leftrightarrow \theta = \frac{1}{\mu_1}$
par la méthode des moments, $\hat{\theta}^{\text{MM}} = \frac{1}{\hat{\mu}_1} = \frac{1}{\bar{X}}$

Rq: $\text{Var}(X_i) = \frac{1}{\theta^2} \Leftrightarrow \theta^2 = \frac{1}{\mathbb{E}X_i^2 - (\mathbb{E}X_i)^2}$
 $\Leftrightarrow \theta = \frac{1}{\sqrt{\mu_2 - \mu_1^2}} \Rightarrow \hat{\theta}_2^{\text{MM}} = \frac{1}{\sqrt{\frac{1}{n} \sum X_i^2 - (\bar{X})^2}}$

• $X_1, \dots, X_n$ i.i.d. de loi P_θ de densité $f_\theta(x) = \theta x^{\theta-1} \mathbb{1}_{x \in [0,1]}$, $\theta > 0$
 $\mathbb{E}_\theta[X_i] = \theta \int_0^1 x^\theta dx = \frac{\theta}{\theta+1}$

Méthode des moments $(\theta+1)\mu_1 = 0 \Leftrightarrow \theta(1-\mu_1) = \mu_1 \Leftrightarrow \theta = \frac{\mu_1}{1-\mu_1} = \frac{\mathbb{E}(X_i)}{1-\mathbb{E}(X_i)}$

$\Rightarrow \hat{\theta}_{\text{MM}} = \frac{\bar{X}}{1-\bar{X}}$, $P_\theta(\bar{X} = 1) = 0$

hyper-moments finis + indépendance
 $\hookrightarrow$ pas nécessairement constant (+LGN)

c) retour sur le LAC:

$(X_n)_{n \geq 1}$ suite de v.a. r. Si X_n converge vers X , que peut-on dire de $(g(X_n))_{n \geq 1}$?

si g continue si $X_n \xrightarrow{P} X$, alors $g(X_n) \xrightarrow{P} g(X)$

si $X_n \xrightarrow{d} X$, alors $g(X_n) \xrightarrow{d} g(X)$

rg: [condition suffisante]

$Dg = \{ \text{points de discontinuité de } g \}$. si $P(X \in Dg) = 0$, le LAC est vrai

ex: $g(x) = \frac{x}{1-x}$ L.G.N: $\bar{X} \xrightarrow{P} E[X]$ étape à faire avant LAC
 LAC: $g(\bar{X}) \xrightarrow{P} g(E[X]) = 0$
 $= \hat{\theta}_n$

LAC pour des couples de suites de v.a.:

si $(X_n, Y_n) \xrightarrow{P} (X, Y)$,

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

(← exemple) $\hat{\theta}_2 = \frac{1}{\sqrt{\frac{1}{n} \sum X_i^2 - (\bar{X})^2}}$ constant. 1?

L.G.N: $\bar{X} \xrightarrow{P} \mu_1$, $\frac{1}{n} \sum X_i^2 \xrightarrow{P} \mu_2$] donc $\begin{pmatrix} \bar{X} \\ \frac{1}{n} \sum X_i^2 \end{pmatrix} \xrightarrow{P} \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$ $g(x, y) = \frac{1}{\sqrt{y - x^2}}$
 faux pour la convergence en loi continue sauf en $\{(x, y) \in \mathbb{R}^2, y = x^2\}$ de mesure nulle

LAC $\Rightarrow \hat{\theta}_2^{MM}$ constant de 0

PROP: $\begin{pmatrix} X_n \\ Y_n \end{pmatrix} \xrightarrow{P} \begin{pmatrix} X \\ Y \end{pmatrix}$ ssi $X_n \xrightarrow{P} X$ et $Y_n \xrightarrow{P} Y$

pv: ($\Rightarrow$) alors LAC $g(x, y) = x$ continue donc $X_n \xrightarrow{P} X$ et $Y_n \xrightarrow{P} Y$

($\Leftarrow$) $\forall \varepsilon > 0$, $P(|X_n - X| + |Y_n - Y| > \varepsilon) \leq P(|X_n - X| > \frac{\varepsilon}{2}) + P(|Y_n - Y| > \frac{\varepsilon}{2})$
 $\rightarrow 0 \quad \rightarrow 0$ (sauf cas indépendance)

Cette réciproque est fautive sur la convergence en loi !!

d) Variance Empirique

Si les X_i admettent une espérance μ et une variance σ^2 ,

on appelle variance empirique: $\hat{\sigma}_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$ pas hypothèse d'indépendance ~ on ne peut pas appliquer L.G.N

$= \frac{1}{n} \sum X_i^2 + \frac{1}{n} \sum \bar{X}^2 - \frac{2}{n} \sum X_i \bar{X}$

$= \frac{1}{n} \sum X_i^2 + \bar{X}^2 - 2\bar{X} \bar{X} = \tilde{\sigma}^2$

(Estimateur des moments: $\sigma^2 = E[X_i^2] - (E[X_i])^2$).

On remplace les moments théoriques par les moments empiriques

$\Rightarrow \hat{\sigma}_n^{MM} = \frac{1}{n} \sum X_i^2 - (\bar{X})^2$

Constant: $\hat{\sigma}^2 = \frac{1}{n} \sum X_i^2 - (\bar{X})^2$ LAC, $g(x, y) = y - x^2$ continue $\mathbb{R}^2 \rightarrow \mathbb{R}$
 $\begin{pmatrix} \bar{X} \xrightarrow{P} EX \\ \frac{1}{n} \sum X_i^2 \xrightarrow{P} EX^2 \end{pmatrix} \Rightarrow \begin{pmatrix} \bar{X} \\ \frac{1}{n} \sum X_i^2 \end{pmatrix} \xrightarrow{P} \begin{pmatrix} EX \\ EX^2 \end{pmatrix} \rightarrow \hat{\sigma}^2 \text{ constant de Var } X = EX^2 - (EX)^2$

(exercice: calculer le biais de $\hat{\sigma}_n^2$ + le risque de $\hat{\sigma}_n^2$)

2. Méthode du maximum de vraisemblance

a) Modèle dominé

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

En pratique:

• soit E au plus dénombrable: $\mu =$ mesure de comptage

si $\exists \{a_1, a_2, a_3, \dots\}$ 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.

ex: Bernoulli(θ), $X_i = 1$ proba $\theta \rightarrow \mu = \delta_0 + \delta_1$
 $0 \parallel 1 - \theta$

On écrit $f_\theta(x) = P_\theta(\{x\}) = P_\theta(X_i = x)$
($x \in \{a_1, a_2, \dots\}$)

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

b) Définitions

$\hookrightarrow f_\theta$ densité de P_θ

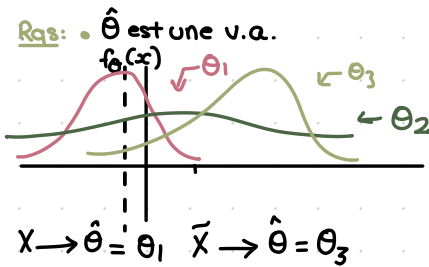
On appelle vraisemblance de l'échantillon $(X_1, \dots, X_n)$ la fonction $\theta \rightarrow L_n(\theta) = \prod_{i=1}^n f_\theta(X_i)$
(variable aléatoire)

Un estimateur du max de vraisemblance $\hat{\theta}_{MV}$ est défini par: $\forall \theta \in \Theta, L_n(\theta) \leq L_n(\hat{\theta})$

On travaille souvent avec la log-vraisemblance $\log L_n(\theta) = \sum_{i=1}^n \ln f_\theta(X_i)$ somme de v.a.

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

Rqs: • $\hat{\theta}$ est une v.a.



Exemples:

• Bernoulli(θ)

X_i à valeurs 0-1, $f_\theta(x) = \theta^x (1-\theta)^{1-x}$
 $\rightarrow L_n(\theta) = \prod_{i=1}^n \theta^{X_i} (1-\theta)^{1-X_i} = \theta^{(\sum X_i)} (1-\theta)^{(n - \sum X_i)}$

continue (C^∞) en θ

$\log L_n(\theta) = (\sum X_i) \ln \theta + (n - \sum X_i) \ln(1-\theta)$

$\rightarrow (\log L_n)'(\theta) = \frac{\sum X_i}{\theta} - \frac{n - \sum X_i}{1-\theta}$ (*)

équation de vraisemblance: $(\log L_n)'(\theta) = 0 \Leftrightarrow (1-\theta) \sum X_i = (n - \sum X_i) \theta$
 $\Leftrightarrow \sum X_i = n\theta \Rightarrow \theta = \frac{\sum X_i}{n}$

pt critique est-il un maximum?

(*) = $\frac{\sum X_i - n\theta}{\theta(1-\theta)}$ la dérivée s'annule et change de signe en $\bar{X}$
 $= \frac{n}{\theta(1-\theta)} (\bar{X} - \theta)$

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

- condition du second ordre:

si $(\log L_n)''(\theta) < 0$ pour tout $\theta \Rightarrow \log L_n$ est concave $\Rightarrow$ max global

$$(\log L_n)''(\theta) = \frac{-\sum x_i}{\theta^2} - \frac{n - \sum x_i}{(1-\theta)^2} < 0 \quad \forall \theta$$

- condition du 2nd ordre "locale"

si $(\log L_n)''(\hat{\theta}) < 0$, $\hat{\theta}$ est un max local. mais par unicite' du point critique, c'est L'EMV

Ex: $f_{\theta}(x) = \theta x^{\theta-1} \mathbb{1}_{x \in [0,1]}$ EMV?

Information de Fisher, efficacité

Soit $(P_\theta)_{\theta \in \Theta}$, $\Theta \subset \mathbb{R}^p$ (identifiable, domine)
↳ toutes les lois admettent une densité

On note f_θ densité de P_θ

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

Étant donné $(X_1, \dots, X_n)$ iid de loi P_θ

$\theta \rightarrow L_n(\theta)$ vraisemblance de l'échantillon

$$= \prod_{i=1}^n f_\theta(X_i) \rightarrow \text{on peut calculer } \log L_n(\theta) = \sum_{i=1}^n \log f_\theta(X_i)$$

$$\hat{\theta} = \underset{\theta \in \Theta}{\operatorname{argmax}} \log L_n(\theta)$$

Propriété d'invariance de l'EMV

si $\hat{\theta}$ EMV de θ , $g(\hat{\theta})$ est un EMV de $g(\theta)$

Objectif: que peut on avoir de "mieux" comme estimateur?

↳ modèle régulier

i) Modèle régulier

DEF 1: Le modèle $(P_\theta)_{\theta \in \Theta}$ est dit régulier si :

a) Θ est un ouvert et pour tout $x \in E$ $\theta \rightarrow f_\theta(x)$ est C^1

b) $\text{Supp } f_\theta$ ne dépend pas de θ : $S = \{x; f_\theta(x) > 0\}$

c) Pour tout θ , l'application $x \mapsto \frac{\left(\frac{\partial f_\theta}{\partial \theta}(x)\right)^2}{f_\theta(x)} \mathbb{1}_{f_\theta(x) > 0}$

est intégrable ($L_1(\mu)$) et l'intégrale $I(\theta) = \int_S \left(\frac{\partial f_\theta}{\partial \theta}(x)\right)^2 \frac{1}{f_\theta(x)} dx$ est continue sur Θ

Notation: on note la dérivée de $f_\theta(x)$ par rapport à θ $\frac{\partial f_\theta}{\partial \theta}(x)$

La quantité $\theta \mapsto I(\theta)$ est appelée information de Fisher du modèle

Exemple:

• $f_\theta(x) = \theta e^{-x\theta}$ densité par rapport à $\mu(dx) = \mathbb{1}_{x \geq 0} dx$

$\theta \mapsto \theta e^{-x\theta}$ est C^∞ sur $\Theta =]0, +\infty[$

$\text{Supp } f_\theta = \mathbb{R}_+$

(si $X \sim \mathcal{E}(\theta)$, $\mathbb{E}[X^k] = \frac{k!}{\theta^k}$)

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

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

$$I(\theta) = \int_0^{+\infty} \frac{(1-x\theta)^2}{\theta} \theta e^{-x\theta} dx = \frac{1}{\theta^2} \mathbb{E}_\theta (1-X\theta)^2$$

$$= \frac{1}{\theta^2} \left(1 - 2\theta \mathbb{E}[X] + \theta^2 \mathbb{E}[X^2] \right)$$

$$= \frac{1}{\theta^2} \text{ continue sur }]0, +\infty[$$

Exemple 2: Bernoulli θ

$$x=0,1 \quad f_\theta(0)=1-\theta$$

$$f_\theta(1)=\theta \quad \text{densité par rapport à } \delta_0 + \delta_1$$

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

$$\left(\frac{\partial f_\theta(0)}{\partial \theta}\right)^2 = \frac{1}{(1-\theta)^2}, \quad \left(\frac{\partial f_\theta(1)}{\partial \theta}\right)^2 = \frac{1}{\theta^2} \Rightarrow I(\theta) = \frac{1}{1-\theta} + \frac{1}{\theta} = \frac{1}{\theta(1-\theta)} \quad \text{continue sur }]0,1[$$

$\uparrow$ l'espace des paramètres

Exemple 3: $f_\theta(x) = \frac{1}{\theta} \mathbb{1}_{[0,\theta]}(x) = \frac{1}{\theta} \mathbb{1}_{[x,+\infty]}(\theta)$ modèle non régulier
 $\uparrow$ (non dérivable en x)

2) Score et information de Fisher

DEF2: On appelle

score
vecteur de score

$$\text{la dérivée de la log-vraisemblance } \frac{\partial}{\partial \theta} \log L_n(\theta) = S_n(\theta)$$

ex: $X_i \sim \mathcal{E}(\theta)$, $L_n(\theta) = \theta^n e^{-\theta \sum X_i}$

$$\log L_n(\theta) = n \log(\theta) - \theta \sum X_i$$

$$\hookrightarrow S_n = \frac{n}{\theta} - \sum_{i=1}^n X_i$$

$$\text{rq: } \mathbb{E}(S_n(\theta)) = \mathbb{E}\left[n\left(\frac{1}{\theta} - \frac{\sum X_i}{n}\right)\right] = 0$$

Hypothèse supplémentaire de régularité

(H) pour tout estimateur $h(x)$ et tout θ , les intégrales suivantes sont égales

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

Remarque: condition d'application du thm de dérivation de Lebesgue

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

PROP: sous (H), le score est centré

$$\text{ex: } \mathbb{E}\left[\frac{\partial}{\partial \theta} \log L_1(\theta)\right] = \int \frac{\partial}{\partial \theta} \log f_\theta(x) f_\theta(x) dx = \int_S \frac{\frac{\partial}{\partial \theta} f_\theta(x)}{f_\theta(x)} f_\theta(x) dx = \int_S \frac{\partial f_\theta(x)}{\partial \theta} dx \stackrel{(H)}{=} \frac{\partial}{\partial \theta} \int_S f_\theta(x) dx \stackrel{=1}{=} 0$$

DEF: L'information de Fisher associée à $(X_1, \dots, X_n)$

est définie par

$$I_n(\theta) = \mathbb{E}_\theta \left[\left(\frac{\partial}{\partial \theta} \log L_n(\theta) \right)^2 \right] \stackrel{\text{def}}{=} \text{var}_\theta \left[\frac{\partial \log L_n(\theta)}{\partial \theta} \right]$$

cor. de la prop 1

$$* \mathbb{E}_\theta \left(\left(\frac{\partial}{\partial \theta} \log f_\theta(X_1) \right)^2 \right) = \int_S \left(\frac{\frac{\partial}{\partial \theta} f_\theta(x)}{f_\theta(x)} \right)^2 f_\theta(x) dx = \int_S \frac{\left(\frac{\partial}{\partial \theta} f_\theta(x) \right)^2}{f_\theta(x)} = \text{expression de la définition 1}$$

ex: $X_1, \dots, X_n \sim \mathcal{E}(\theta)$

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

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

PROP2: $I_n(\theta) = n I(\theta)$

$$\text{En effet: } I_n(\theta) = \text{Var} \left(\frac{\partial}{\partial \theta} \log L_n(\theta) \right) = \text{Var} \left(\sum_{i=1}^n \frac{\partial}{\partial \theta} \log f_\theta(X_i) \right) \stackrel{\text{indép}}{=} \sum_{i=1}^n \text{Var} \left(\frac{\partial}{\partial \theta} \log f_\theta(X_i) \right)$$

$$= n \text{Var} \left(\frac{\partial}{\partial \theta} \log f_\theta(X_1) \right) = n I(\theta)$$

Ex2: $(X_1, \dots, X_n)$ i.i.d. Poisson(θ) $f_\theta(x) = e^{-\theta} \frac{\theta^x}{x!}$

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

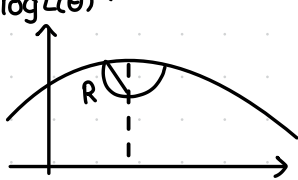
$$\frac{\partial}{\partial \theta} \log L_n(\theta) = -n + \frac{\sum X_i}{\theta} \Rightarrow I_n(\theta) = \text{Var}\left(\frac{\sum X_i}{\theta}\right) = \frac{1}{\theta^2} n\theta = \frac{n}{\theta}$$

3. Info de Fisher et dérivée seconde

PROP3: En ajoutant que $\theta \mapsto f_\theta(x)$ est C^2 et que (H) vraie pour $\frac{\partial^2}{\partial \theta^2}$, alors l'info de Fisher s'écrit encore

$$I_n(\theta) = -\mathbb{E}_\theta \left[\frac{\partial^2}{\partial \theta^2} \log L_n(\theta) \right] \quad (\text{si } \hat{\theta} \in \text{EMV}, I_n(\hat{\theta}) > 0)$$

$\log L(\theta)$ (si $\hat{\theta} \in \text{EMV}, I_n(\hat{\theta}) > 0$)



$$\frac{\partial}{\partial \theta^2} \log(\theta) \propto \frac{1}{R}$$

si courbe très "piquée" en l'EMV (i.e. info Fisher est grande) alors l'EMV est localisé de façon précise.

$$n=1: \frac{\partial^2}{\partial \theta^2} \log f_\theta(x) = \frac{\frac{\partial^2 f_\theta(x)}{\partial \theta^2}}{f_\theta(x)} - \left(\frac{\frac{\partial f_\theta(x)}{\partial \theta}}{f_\theta(x)} \right)^2$$

$$\mathbb{E}\left(\frac{\partial^2}{\partial \theta^2} \log f_\theta(X_1)\right) = \int_S \frac{\frac{\partial^2 f_\theta(x)}{\partial \theta^2}}{f_\theta(x)} f_\theta(x) dx - \int_S \frac{\frac{\partial f_\theta(x)}{\partial \theta}}{f_\theta(x)} f_\theta(x) dx$$

$I(\theta)$

4) Inégalité de Cramér-Rao

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

PROP4: Sous les hypothèses d'un modèle régulier, si pour tout $\theta, I(\theta) > 0$ alors pour tout estimateur $T = T(X_1, \dots, X_n)$ sans biais, $\mathbb{E}_\theta T^2 < +\infty$ On a $\forall \theta \in \mathbb{M}, \text{Var}_\theta(T) \geq \frac{(g'(\theta))^2}{I_n(\theta)}$

DEF: si T réalise l'égalité, alors T est dit efficace

preuve: $\forall \theta, \mathbb{E}_\theta(T) = g(\theta)$

$$\Rightarrow \frac{\partial}{\partial \theta} \mathbb{E}_\theta(T) = g'(\theta)$$

$$\stackrel{n=1}{T=T(X_1)} \Leftrightarrow \frac{\partial}{\partial \theta} \int_S T(x) f_\theta(x) dx = g'(\theta)$$

$$\Leftrightarrow \int_S T(x) \frac{\frac{\partial f_\theta(x)}{\partial \theta}}{f_\theta(x)} f_\theta(x) dx = g'(\theta)$$

(H)

$$\Leftrightarrow \int_S (T(x) - g(\theta)) \frac{\frac{\partial f_\theta(x)}{\partial \theta}}{f_\theta(x)} f_\theta(x) dx = g'(\theta)$$

Inég 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

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

Étude asymptotique des estimateurs

Dans un modèle paramétrique régulier, si $\hat{\theta}_n$ estimateur de θ ,
alors $\text{Var}(\hat{\theta}_n) \geq \frac{1}{I_n(\theta)} = \frac{1}{n I(\theta)}$

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

Asymptotique: $n \rightarrow +\infty$

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

1) Convergences

$(X_n)_{n \geq 0}$ suite de v.a. réelles ($\mathbb{R}^d$)

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

→ lemme de Portmanteau:

Caractérisations équivalentes:

- Pour toute fonction continue bornée h , $\mathbb{E}[h(X_n)] \rightarrow \mathbb{E}[h(X)]$

⇒ la convergence en loi est stable par passage aux fonctions continues (LAC)

MAIS il est en général faux que si $\begin{matrix} X_n \xrightarrow{\mathcal{L}} X \\ Y_n \xrightarrow{\mathcal{L}} Y \end{matrix}$ 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 $\forall n$, X_n et Y_n sont indépendantes alors convergence en loi de X_n et Y_n
 X et Y sont indépendantes alors cv en loi du couple $\begin{pmatrix} X_n \\ Y_n \end{pmatrix}$
- ② • si $\begin{pmatrix} X_n \xrightarrow{\mathbb{P}} X \\ Y_n \xrightarrow{\mathbb{P}} Y \end{pmatrix} \Rightarrow \begin{pmatrix} X_n \\ Y_n \end{pmatrix} \xrightarrow{\mathbb{P}} \begin{pmatrix} X \\ Y \end{pmatrix} \Rightarrow \begin{pmatrix} X_n \\ Y_n \end{pmatrix} \xrightarrow{\mathcal{L}} \begin{pmatrix} X \\ Y \end{pmatrix}$

③. Lemme de Slutsky

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

en appliquant le LAC:

$$\begin{aligned} h(x, y) &= x + y & X_n + Y_n &\xrightarrow{\mathcal{L}} X + c \\ &xy & X_n Y_n &\xrightarrow{\mathcal{L}} cX \\ &x/y & \text{si } c \neq 0 & X_n/Y_n \xrightarrow{\mathcal{L}} X/c \end{aligned}$$

2) Consistance des estimateurs

DEF1: $\hat{\theta}_n$ asymptotiquement sans biais ssi $\text{Biais}(\hat{\theta}_n, \theta) = \mathbb{E}[\hat{\theta}_n] - \theta \xrightarrow[n \rightarrow +\infty]{} 0$

Rq: la convergence en proba n'implique pas la convergence des espérances!

si $X_n \xrightarrow{\mathbb{P}} X$

$|X_n| \leq Y \in L^1$, alors par convergence dominée $X_n \rightarrow X$ dans L^1 .

exemple: $\hat{\sigma}_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 - (\bar{X})^2$ estimateur des moments de $\sigma^2 = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$
 Δ pas indep.

Biais: $(\hat{\sigma}_n^2, \sigma^2) = \frac{1}{n} \sigma^2$ as. sans biais (as. = asymptotiquement)

outils pour montrer la consistance

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

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

$$\mathbb{E}[X_i^2] < +\infty$$

$$\text{LGN: } \frac{1}{n} \sum_{i=1}^n X_i^2 \xrightarrow{\mathbb{P}} \mathbb{E}[X^2] = \sigma^2 + \mu^2$$

$$\bullet \bar{X} \xrightarrow{\mathbb{P}} \mu \quad (\text{LGN})$$

$$\text{LAC avec } h(x) = x^2; \quad (\bar{X})^2 \xrightarrow{\mathbb{P}} \mu^2$$

$$\bullet \text{ Donc } \begin{pmatrix} \frac{1}{n} \sum_{i=1}^n X_i^2 \\ \frac{1}{n} \sum_{i=1}^n X_i \end{pmatrix} \xrightarrow{\mathbb{P}} \begin{pmatrix} \sigma^2 + \mu^2 \\ \mu \end{pmatrix}$$

$$\bullet \text{ LAC } h(x, y) = x - y^2$$

Donc

$$\frac{1}{n} \sum X_i^2 - \bar{X}^2 \xrightarrow{\mathbb{P}} \sigma^2 + \mu^2 - \mu^2 = \sigma^2$$

3) Normalité asymptotique:

$\hat{\theta}_n$ consistant 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 $\sigma^2(\theta)$ $\hat{\theta} = \bar{X}$

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

DEF: $(\hat{\theta}_n)$ est un estimateur asymptotiquement normal ssi

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

exemple: σ_n^2 est elle as. normale?

$(X_1, \dots, X_n)$ i.i.d. d'espérance μ , de variance σ^2 , $\mu_4 = \mathbb{E}((X_i - \mu)^4) < +\infty$

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

$$\frac{1}{n} \sum_{i=1}^n (X_i - \mu)(\mu - \bar{X}) = \frac{1}{n} \sum_{i=1}^n (X_i - \mu)(\mu - \bar{X}) = \frac{1}{n} (\sum_{i=1}^n (X_i - \mu))(\mu - \bar{X}) = \frac{1}{n} (0)(\mu - \bar{X}) = 0$$

$$= 2(\mu - \bar{X})(\bar{X} - \mu) = -2(\bar{X} - \mu)^2$$

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

$$\sqrt{n} \cdot \left(\frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2 - \sigma^2 \right) \xrightarrow{\mathcal{L}} Z \sim \mathcal{N}(0, \mu_4 - \sigma^4)$$

$$\neq \text{Var}(X_i - \mu)^2 = \mathbb{E}((X_i - \mu)^4) - \sigma^4 = \mu_4 - \sigma^4$$

$$\bullet \text{ TLC: } \sqrt{n}(\bar{X} - \mu) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2)$$

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

$$\downarrow \quad \quad \quad \downarrow \quad \quad \quad \downarrow \quad \quad \quad \downarrow$$

$$Z \sim \mathcal{N}(0, \mu_4 - \sigma^4) \quad \quad \quad \mathcal{N}(0, \sigma^2) \quad \quad \quad 0 \text{ constante!}$$

$$\bar{X} - \mu \xrightarrow{\mathcal{L}} 0, \quad \sqrt{n}(\bar{X} - \mu) \xrightarrow{\mathcal{L}} U \sim \mathcal{N}(0, 1) \Rightarrow \sqrt{n}(\bar{X} - \mu)^2 \xrightarrow{\mathcal{L}} 0$$

nouvelle applic du lemme de Slutsky:

$$\sqrt{n}(\hat{\sigma}_n^2 - \sigma^2) \xrightarrow{\mathcal{L}} Z + 0$$

Donc $\hat{\sigma}_n^2$ est un $\hat{\theta}_n$ estimateur asymptotiquement normal.

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

$\Leftrightarrow \frac{\sqrt{n}(\hat{\theta}_n - \theta)}{\sigma} \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, 1)$
 propriété des lois normales

Application du lemme de Slutsky: si $\hat{\sigma}^2$ est un estimateur consistant de σ^2 , alors on a encore

$$\frac{\sqrt{n}(\hat{\theta}_n - \theta)}{\hat{\sigma}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1) \quad (\hat{\sigma} = \hat{\sigma}(X_1, \dots, X_n))$$

preuve (de la remarque):

$$\begin{aligned} \frac{\sqrt{n}(\hat{\theta}_n - \theta)}{\hat{\sigma}} &= \underbrace{\frac{\sqrt{n}(\hat{\theta}_n - \theta)}{\sigma}}_{\xrightarrow{\mathcal{L}} \mathcal{N}(0, 1)} \times \underbrace{\frac{\sigma}{\hat{\sigma}}}_{\xrightarrow{\mathbb{P}} 1} \\ &\quad (\hat{\theta} \text{ as. normal}) \\ &\quad (\text{par consistence } \sigma^2 \xrightarrow{\mathbb{P}} \sigma^2 \\ &\quad \xRightarrow{\text{LAC}} \sigma^2 \xrightarrow{\mathbb{P}} \sigma) \\ &\xRightarrow{\text{par Slutsky}} 1. Z \sim \mathcal{N}(0, 1) \end{aligned}$$

4) δ-méthode:

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

LEMME: [méthode δ]

Soit Z_n une suite de v.a. réelles tq $\sqrt{n}(Z_n - \mu) \xrightarrow{\mathcal{L}} Z \sim \mathcal{N}(0, \sigma^2)$

Soit g une fonction dérivable, $g'(\mu) \neq 0$

Sous ces hypothèses, on a $\sqrt{n}[g(Z_n) - g(\mu)] \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \tilde{Z} \sim \mathcal{N}(0, (g'(\mu))^2 \sigma^2)$

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

pu:

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

$$\begin{aligned} \text{A-t-ons } Z_n \xrightarrow{\mathbb{P}} \mu? \\ \mathbb{P}(|Z_n - \mu| > \varepsilon) &= \mathbb{P}\left(\frac{\sqrt{n}|Z_n - \mu|}{\sigma} > \frac{\sqrt{n}\varepsilon}{\sigma}\right) = \mathbb{P}\left(\frac{\sqrt{n}(Z_n - \mu)}{\sigma} > \frac{\sqrt{n}\varepsilon}{\sigma}\right) + \mathbb{P}\left(\frac{\sqrt{n}(Z_n - \mu)}{\sigma} < -\frac{\sqrt{n}\varepsilon}{\sigma}\right) \\ &\quad (\hookrightarrow \text{erreur}) \approx 1 - \Phi\left(\frac{\sqrt{n}\varepsilon}{\sigma}\right) + \Phi\left(-\frac{\sqrt{n}\varepsilon}{\sigma}\right) = 2\left(1 - \Phi\left(\frac{\sqrt{n}\varepsilon}{\sigma}\right)\right) \end{aligned}$$

Φ f.n. $\frac{\sqrt{n}(Z_n - \mu)}{\sigma}$

Fonction de répartition empirique

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

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

DEF: La fonction de répartition empirique, associée à $(X_1, \dots, X_n)$ est définie par:

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

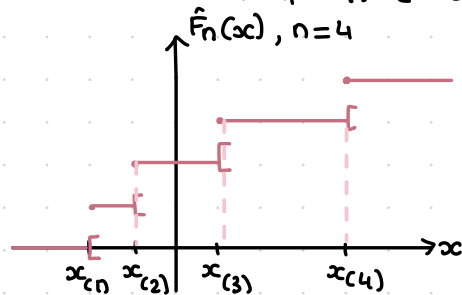
$$x \mapsto \frac{1}{n} \sum_{i=1}^n 1_{X_i \leq x}$$

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

DEF: Loi empirique $P_n = \frac{1}{n} \sum_{i=1}^n \delta_{X_i}$ = 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



si les $x_i \neq$; saut de $1/n$

a) Propriétés immédiates:

• $n \hat{F}_n(x) = \sum_{i=1}^n 1_{X_i \leq x}$ suit la loi binomiale $(n, F(x))$

$$\bullet R(\hat{F}_n(x), F(x)) = 0 + \frac{1}{n^2} \text{Var} \left(\sum_{i=1}^n 1_{X_i \leq x} \right) = \frac{1}{n} F(x) (1 - F(x)) \xrightarrow[n \rightarrow +\infty]{} 0$$

indep

ecdf
→ empirical cumulative
distribution function

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

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

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

y_i

TCL: les X_i sont iid, donc les $\{1_{X_i \leq x}\}$ sont iid

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

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

b) Estimation empirique:

"plug-in" ou méthode de substitution

paramètre d'intérêt $\theta = c(F)$ (caractéristique de la fonction inconnue)

la méthode empirique définit $\hat{\theta}$, estimateur empirique en remplaçant F par $\hat{F}_n$

$$\hookrightarrow \hat{\theta}_n = c(\hat{F}_n)$$

$$\text{ex: } \theta = E_F(X_1) \xrightarrow{=} \hat{\theta}_n = E_{\hat{F}_n}(X_1) = \sum_{i=1}^n X_i \cdot \frac{1}{n} \quad (\text{si les } X_i \text{ distincts})$$

$= \bar{X}$

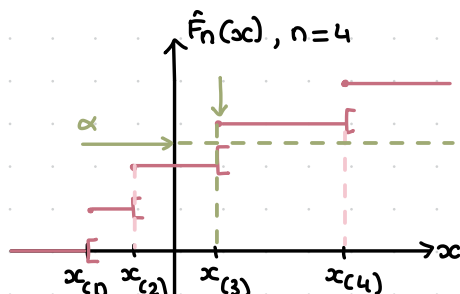
$$\theta = \text{Var}_F(X) \longrightarrow \hat{\theta}_n = \text{Var}_{\hat{F}_n}(X) = \frac{1}{n} \sum (X_i - \bar{X})^2 \quad \text{estimateurs des moments}$$

c) Inverse généralisé

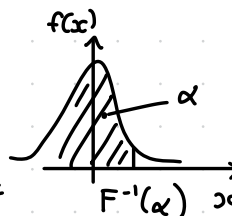
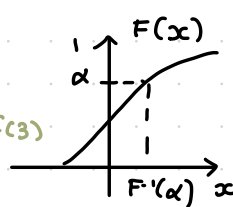
DEF: On définit l'inverse généralisé de F par $F^{-1}: [0,1] \longrightarrow \mathbb{R}$

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

- si F strict $\uparrow$, $\inf x \text{ tq } F(x) \geq \alpha \Leftrightarrow x \geq F^{-1}(\alpha) \leadsto x = F^{-1}(\alpha)$
- si F est la f.r. d'une loi discrète



$$F^{-1}(\alpha) = x_{(3)}$$



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

VOCAB: F^{-1} s'appelle aussi la fonction quantile

$F^{-1}(\alpha)$ = quantile d'ordre α de la loi F

$F^{-1}(1/4)$ = 1^{er} quantile, $F^{-1}(1/2)$ = médiane, $F^{-1}(3/4)$ = 3^{er} quantile

LEMME: U v.a. uniforme sur $[0,1]$

F une f.r.

alors $F^{-1}(U)$ est une v.a. de loi F .

si F est bijective, $P(F^{-1}(U) \leq x) = P(U \leq F(x)) = F(x)$

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

si F est discrète, F^{-1} inverse généralisé, $F^{-1}(y) \leq x \Leftrightarrow y \leq F(x)$ (suite en TD)

d) Quantile empirique

DEF: On définit le quantile empirique (sample quantile) d'ordre α comme étant le quantile de $\hat{F}_n$:

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

Propriétés:

- 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'ech. ordonné des $(X_i)_{1 \leq i \leq n}$
 $\lceil u \rceil$ = + petit entier $\geq u$

ex: $\alpha = \frac{1}{2}$ $\lceil \frac{n}{2} \rceil$ si $n=2k$ médiane $\hat{q}_{n,1/2} = X_{(k)}$

si $n=2k+1$ médiane $\hat{q}_{n,1/2} = X_{(k+1)}$

Dans les logiciels, plusieurs façons de calculer les quantiles empiriques

$n=4$ médiane = $X_{(2)}$

$n=5$ médiane = $X_{(3)} \leadsto \frac{X_{(2)} + X_{(3)}}{2}$

$n=6$ médiane = $X_{(3)}$

- **consistance** si $\alpha \in]0,1[$, si F est strictement croissante au voisinage de α , alors $\hat{q}_{n,\alpha}$ consistant de q_α .

Intervalles de confiance (pas dans le partiel)

$(X_1, \dots, X_n)$ i.i.d. de loi $P \in \{P_\theta, \theta \in \Theta \subset \mathbb{R}^p\}$.

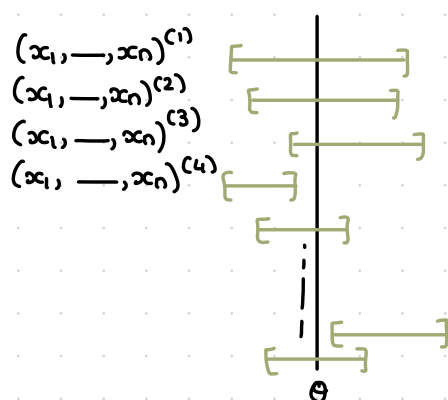
On s'intéresse à $\theta \in \mathbb{R}$ ou $g(\theta): \mathbb{R}^p \rightarrow \mathbb{R}$

1) DEF:

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 dépendent PAS des paramètres inconnus du modèle, et tel que $P([B_{\inf}(X_1, \dots, X_n); B_{\sup}(X_1, \dots, X_n)] \ni \theta) \geq 1 - \alpha$

- 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
 $\downarrow$
 $B_{\inf}(X_1, \dots, X_n)$
- en général, $\alpha = 1\%, 5\%$

Interprétation:



$$IC = [B_{\inf}(X_1, \dots, X_n), B_{\sup}(X_1, \dots, X_n)]$$

formule mathématique qui garantit 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

$\hookrightarrow$ on calcule $IC = [2.3; 5.1]$ de niveau de confiance 95%.

En moyenne, sur 100 intervalles calculées 5% 'rater' θ .

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

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

propriété des lois Gaussiennes

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

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

$\hookrightarrow$ inconnu



$$\text{si } Z \sim \mathcal{N}(0, 1), P(q_{\alpha/2} \leq Z \leq q_{1-\alpha/2}) = 1 - \alpha$$

- 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 $\hat{\sigma}^2$, constant pour estimer $\sigma^2(\theta)$.

$$\text{Si c'est le cas, } \frac{\sqrt{n}(\hat{\theta} - \theta)}{\sigma(\theta)} \times \frac{\sigma^2(\theta)}{\hat{\sigma}^2} \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, 1)$$

$$\begin{array}{ccc} \begin{array}{c} \hat{\theta} \text{ as. normal} \\ \downarrow \mathcal{L} \\ \mathcal{N}(0, 1) \end{array} & \begin{array}{c} \downarrow P(\text{estimateur constant}) \\ 1 \end{array} & \text{lemme de Slutsky,} \end{array}$$

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

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

Compléments: (avant partiel)

- 1) Retour sur normalité a.s.
- 2) Exemple
- 3) Pivot asymptotique
- 4) Exemple 2

1. Propriétés asymptotiques d'une suite d'estimateurs $(\hat{\theta}_n)_{n \geq 1}$ → Consistance $\hat{\theta}_n \xrightarrow{P} \theta$ → Normalité asymptotique: si $\exists \sigma^2 > 0$ $\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, \sigma^2)$ De façon générale, si il existe $U_n \xrightarrow[n \rightarrow +\infty]{} +\infty$

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

on dit que $\hat{\theta}_n$ converge à la vitesse $\frac{1}{V_n}$ Rq: si $\hat{\theta}_n$ as. normal $\Rightarrow \hat{\theta}_n$ consistant

$$(\hat{\theta}_n - \theta) = \underbrace{\frac{1}{\sqrt{n}}}_{\downarrow 0} \underbrace{\sqrt{n}(\hat{\theta}_n - \theta)}_{\downarrow \mathcal{N}(0, \sigma^2)} \xrightarrow[n \rightarrow +\infty]{\text{Slutskey}} 0$$

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

→ δ -méthode

$$\sqrt{n}(X_n - 1) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1)$$

$$\hookrightarrow \sqrt{n}(X_n - 1) \stackrel{\mathcal{L}}{\sim} Z \sim \mathcal{N}(0, 1)$$

$$\hookrightarrow X_n \approx 1 + \frac{1}{\sqrt{n}} Z$$

g dérivable en 1

* si g est dérivable en 1, $g(1+h) = g(1) + h g'(1)$

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

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

→ δ -méthode

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

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

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

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

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

2) Exemple

 $X_1, \dots, X_n$ de loi de densité $f(x) = \frac{1}{\mu} e^{-x/\mu}$; $x \geq 0$, $\mu = \mathbb{E}[X_1] > 0$ μ est estimé par $\hat{\mu} = \bar{X}$

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

sans biais
qui atteint
la borne de
CR

$$\text{Var}(\hat{\mu}) = \frac{1}{n^2} \text{Var}\left(\sum_{i=1}^n X_i\right)$$

$$\stackrel{\text{indep}}{=} \frac{1}{n^2} \sum_{i=1}^n \text{Var}(X_i) \stackrel{\text{i.i.d.}}{=} \frac{1}{n} \text{Var}(X_1) = \frac{\mu^2}{n}$$

$$\text{Var}(X_1) = \mu^2$$

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

$$I_n(\mu) = \text{Var} \left(\frac{-n}{\mu} + \frac{1}{\mu^2} \sum_{i=1}^n X_i \right) = \frac{1}{\mu^4} \text{Var} \left(\sum_{i=1}^n X_i \right) = \frac{n}{\mu^4} \text{Var}(X_i)$$

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

$\hat{\mu}$ sans biais et $\text{Var}(\hat{\mu}) = \frac{1}{I_n(\mu)}$, donc $\hat{\mu}$ est efficace

as. normal, où $\text{var} = \frac{1}{I_n}$

• as. efficace?

TLC: $\sqrt{n}(\hat{\mu} - \mu) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \mu^2)$

$\rightarrow \text{Var}(\hat{\mu}) = \frac{\mu^2}{n} = \text{variance de la loi gaussienne asymptotique} \Rightarrow \hat{\mu}$ est asymptotique efficace

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

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

$$\mathbb{E}[X_i] = \frac{1}{\theta}, \text{Var}(X_i) = \frac{1}{\theta^2}$$

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

$$\frac{\partial}{\partial \theta} (\log L_n)(\theta) = \frac{n}{\theta} - \sum_{i=1}^n X_i \quad \leadsto \quad \hat{\theta}^{MV} = \frac{1}{\bar{X}}$$

$$\hookrightarrow I_n(\theta) = \text{Var} \left(\frac{n}{\theta} - \sum_{i=1}^n X_i \right) = \text{Var} \left(\sum_{i=1}^n X_i \right) = \frac{n}{\theta^2}$$

$$\mathbb{E} \left(\frac{1}{\bar{X}} \right) = ?$$

cf TOI: $n\bar{X} \sim \Gamma(n, \theta)$ et $\mathbb{E} \left(\frac{1}{n\bar{X}} \right) = \frac{\theta}{n-1}$ et $\mathbb{E} \left(\frac{1}{(n\bar{X})^2} \right) = \frac{\theta^2}{(n-1)(n-2)}$

d'où $\mathbb{E}[\hat{\theta}] = \mathbb{E} \left[\frac{1}{\bar{X}} \right] = n \mathbb{E} \left[\frac{1}{n\bar{X}} \right] = \frac{n\theta}{n-1}$ biaisé
 $\hookrightarrow \tilde{\theta} = \frac{n-1}{n} \hat{\theta}$ non biaisé

$$\begin{aligned} \text{Var}(\tilde{\theta}) &= \left(\frac{n-1}{n} \right)^2 \left[\mathbb{E} \left(\frac{1}{\bar{X}^2} \right) - \left(\mathbb{E} \left(\frac{1}{\bar{X}} \right) \right)^2 \right] \\ &= \frac{(n-1)^2}{n^2} \cdot \frac{n^2 \theta^2}{(n-1)(n-2)} - \frac{(n-1)^2}{n^2} \cdot \frac{n^2}{(n-1)^2} \theta^2 \\ &= \theta^2 \frac{n-1}{n-2} - \theta^2 = \frac{\theta^2}{n-2} \end{aligned}$$

CCL: $\text{Var}(\tilde{\theta}) > \frac{1}{I_n(\theta)}$ non efficace

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

$g(x) = \frac{1}{x}$ sur $]0, +\infty[$, $g'(x) = -\frac{1}{x^2} \neq 0$

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

$\Rightarrow \hat{\theta}$ est asymptotiquement efficace

3) Pivot (asymptotique) ou statistique pivotale

DEF: Statistique dont la loi ne dépend pas de paramètres inconnus

ex: $X_1, \dots, X_n$ i.i.d Bernoulli(θ): $\sqrt{n}(\bar{X} - \theta) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \theta(1-\theta))$

$\theta \in]0, 1[$

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

pivot ou stat pivotale

méthode pivotale pour IC:

On estime $\sqrt{\theta(1-\theta)}$ par $\sqrt{\hat{\theta}(1-\hat{\theta})}$ "plug-in"

par le LAC, $g(x) = \sqrt{x(1-x)}$; $x \in]0,1[$, $\sqrt{\hat{\theta}(1-\hat{\theta})}$ est un estimateur consistant de $\sqrt{\theta(1-\theta)}$

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

donc, par Slutsky,

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

4) Exemple 2:

$(X_1, \dots, X_n)$ de densité $\theta > 0$: $f_{\theta}(x) = \frac{3}{\theta} x^2 \exp(-\frac{x^3}{\theta}) \mathbb{1}_{x \geq 0}$ par rapport à mesure $\mathbb{1}_{x \geq 0} dx$
 pour gérer l'indicateur $\mathbb{1}_{x \geq 0}$
 changement de mesure dominante
 (ou dire par hyp, $X_i \geq 0 \forall i$)

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

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

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

TCL: $\frac{\sqrt{n}(\hat{\theta}_n - \theta)}{\sqrt{\hat{\theta}_n(1-\hat{\theta}_n)}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$
 $\Leftrightarrow \frac{\sqrt{n}(\hat{\theta}_n - \theta)}{\hat{\theta}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$
 pivot asymptotique

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

si on note $q_{\alpha/2}$ et $q_{1-\alpha/2}$ quantiles de $\mathcal{N}(0,1)$

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

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

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

$\Rightarrow$ IC(θ) de niveau asymptotique $(1-\alpha)$