Notes de Cours d'Inférence Statistique

Yehor KOROTENKO

02 February 2026

Table of contents

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 $X_1, ..., X_n$ i.i.d. (indépendants et identiquement distribués) de loi commune $\in \{P_\theta, \theta \in \Theta\}$.*

Remark 1.4 – *$(X_1, ..., X_n)$ 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 $X_1, ..., X_n$ 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 $(X_1, ..., X_n)$ mesurable et indépendante de θ (calculable à partir des données).*

Notation : $\hat{\theta} = \hat{\theta}_n = h(X_1, ..., X_n)$. C'est une variable aléatoire.
Exemples : $\hat{\theta} = \overline{X}$, $\hat{\theta} = X_1 - X_3$, 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 $\mathbb{E}\left[\hat{\theta}-\theta\right]$.

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

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

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

DEFINITION 1.8 (RISQUE QUADRATIQUE / MSE) –

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

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

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

1.4.1 Exemple : Modèle de Poisson

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

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

Calcul du Biais :

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

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

Calcul du Risque :

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

THEOREM 1.9 (DÉCOMPOSITION BIAIS-VARIANCE DU RISQUE) –

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

Proof.

$$
\begin{aligned}
R\left(\hat{\theta}, \theta\right) &= \mathbb{E}\left[\left(\hat{\theta} - \theta\right)^2\right] \\
&= \mathbb{E}\left[\left(\hat{\theta} - \mathbb{E}\left[\hat{\theta}\right] + \mathbb{E}\left[\hat{\theta}\right] - \theta\right)^2\right] \\
&= \mathbb{E}\left[\left(\hat{\theta} - \mathbb{E}\left[\hat{\theta}\right]\right)^2\right] + \mathbb{E}\left[\left(\mathbb{E}\left[\hat{\theta}\right] - \theta\right)^2\right] + 2\mathbb{E}\left[\left(\hat{\theta} - \mathbb{E}\left[\hat{\theta}\right]\right)\left(\mathbb{E}\left[\hat{\theta}\right] - \theta\right)\right] \\
&= \mathrm{Var}\left(\hat{\theta}\right) + \left(B\left(\hat{\theta}, \theta\right)\right)^2 + 2\left(\mathbb{E}\left[\hat{\theta}\right] - \theta\right)\underbrace{\mathbb{E}\left[\hat{\theta} - \mathbb{E}\left[\hat{\theta}\right]\right]}_{0} \\
&= \mathrm{Var}\left(\hat{\theta}\right) + B\left(\hat{\theta}, \theta\right)^2
\end{aligned}
$$

□

1.5 Consistance

Propriété asymptotique. On ne considère que des estimateurs consistants.

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

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

REMARK 1.11 – *$\hat{\theta}_n$ est fortement consistant ssi $\hat{\theta}_n \xrightarrow[n \to +\infty]{\text{p.s.}} \theta$.*

1.5.1 Exemple : Retour au modèle de Poisson

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

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

$$R\left(\hat{\theta}_n, \theta\right) = \mathrm{Var}\left(\overline{X}\right) = \frac{\theta}{n} \xrightarrow[n \to +\infty]{} 0$$

  D'après l'inégalité de Bienaymé-Tchebychev :

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

1.5.2 Méthode "Plug-in"

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

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

$\hat{\beta}$ 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 $(X_1, ..., X_n)$, 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-paramétré.*

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 – $E_{\theta_n}[h(X_1, ..., X_n)]$, $\Theta[h(X_1, ..., X_n)]$
Loi de $(X_1, ..., X_n) \rightarrow P_\theta^{\otimes n}$

Definition 2.4 (Estimateur) –

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

Definition 2.5 (Qualité) –

- *Risque*

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

- *Consistance*

$$\hat{\theta}_n \underset{n\to+\infty}{\overset{P}{\rightarrow}} \theta$$

Definition 2.6 (Modèle identifiable) –

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

2.2 Méthode des moments

Definition 2.7 – *On appelle **moment théorique** de la loi de X_i d'ordre k:*

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

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 $\hat{\mu}_k \xrightarrow[n\to+\infty]{P} \mu_k$.

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

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

alors l'estimateur

$$\hat{\theta} = \mathcal{L}(\hat{\mu}_1, ..., \mu_k)$$

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,*

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

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

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

$$\Theta(X_i) = \frac{1}{\theta^2} \Longleftrightarrow \theta^2 = \frac{1}{E[X_i^2] - E[X_i]^2}$$

$$\Longleftrightarrow \theta = \frac{1}{\sqrt{\mu_2 - \mu_1^2}}$$

$$\Longrightarrow \hat{\theta}_2 = \frac{1}{\sqrt{\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \left(\overline{X}\right)^2}}$$

- $X_1, ..., X_n$ *i.i.d. de la loi* P_θ *de densité*

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

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

Méthode des moments:

$$(\theta+1)\mu_1 = \theta \Longleftrightarrow \theta(1-\mu_1) = \mu_1 \Longleftrightarrow \theta = \frac{E[X_i]}{1-E[X_i]}$$

$$\Longrightarrow \hat{\theta}_{\mathbb{M}} = \frac{\overline{X}}{1-\overline{X}},, P_\theta\left(\overline{X} = 1\right) = P_\theta(X_1 = X_2 = ... = X_n = 1) = 0$$

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:* $\overline{X} \xrightarrow{P} E[X]$
- *LAC:* $g(\overline{X}) = \hat{\theta}_n \xrightarrow[n\to+\infty]{P} g(E[X]) = \theta$

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 $g : \mathbb{R}^2 \to \mathbb{R}$ 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 –

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

LGN:

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

donc

$$\begin{pmatrix} \overline{X} \\ \frac{1}{n}\sum_{i=1}^n X_i^2 \end{pmatrix} \xrightarrow{P} \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$$

$g(x, y) = \frac{1}{\sqrt{y-x^2}} \implies \hat{\theta}^{\mathbb{M}}$ *constant de* θ, 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.

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

- $\Longleftarrow$ convergence du couple?

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

  <u>Cette réciproque est fausse pour la converge en loi!</u>

$\square$

2.3.1 Variance empirique

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

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

estimateur des moments:

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

On remplace les moments théoriques par les moments empiriques

$$\to \widetilde{\sigma}^{2\mathbb{M}} = \frac{1}{n}\sum_{i=1}^n X_i^2 - \left(\overline{X}\right)^2$$

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

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

> **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 $\to X_i$ 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: $\mu =$ mesure de comptage. Si $\exists, \{a_1, a_2, ...\}$ 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 $\theta \to \mu = \delta_0 + \delta_1$ On écrira*
  >
  > $$f_\theta(x) = \underbrace{P_\theta(\{x\})}_{=1-\theta} - P_\theta(X_i = x) \ \text{ avec } \ x \in \{a_1, a_2, ...\}$$

- 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* $(X_1, ..., X_n)$ *la fonction*

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

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

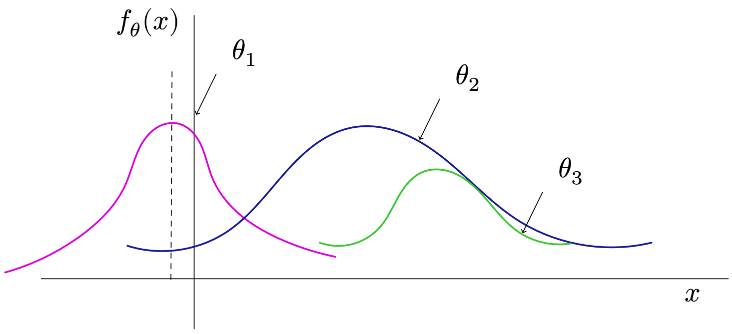
$$\forall \theta \in \Theta, L_n(\theta) \leq L_n\left(\hat{\theta}\right)$$

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

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

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

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



EXAMPLE 2.19 –

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

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

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

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

Equation de vraisemblance:

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

le point critique, est-il un maximum?

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

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

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

Étant donné $(X_1, ..., X_n)$, 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 $\operatorname{Supp} f_\theta$ on peut calculer

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

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

Proposition 3.1 – *Si $\hat{\theta}$ EMV[1] de θ, $g\left(\hat{\theta}\right)$ est un EMV de $g(\theta)$*

Objectif: que peut-on avoir de "mieux" comme estimateur? $\longrightarrow$ 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. *$\operatorname{Supp} f_\theta$ ne dépend pas de θ: $S = \left\{x,\ f_{\theta(x)} > 0\right\}$*
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 $\mu(dx) = \mathbb{1}_{x \geq 0} dx$*

$$\theta \mapsto \theta e^{-x\theta} \text{ est } C^\infty \text{ sur } \Theta =]0, +\infty[,\ \operatorname{Supp} f_\theta = \mathbb{R}_+$$

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

[1] **EMV** = Estimateur de Maximum de Vraisemblance

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

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

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

$(X_1,...,X_n)$ 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* $S-n(\theta)=\frac{n}{\theta}-\sum_{i=1}^n X-i$

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.*

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

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

$$E_\theta\left[\frac{\partial}{\partial\theta}\log L_1(\theta)\right] = \int_S \frac{\partial}{\partial\theta}\log f_\theta(x)dx = \int_S \frac{\frac{\partial}{\partial\theta}f_{\theta(x)}}{\cancel{f_\theta(x)}}\cancel{f_\theta(x)}dx = \int_S \frac{\partial f_\theta}{\partial\theta}(x)dx \underbrace{=}_{(H)} \frac{\partial}{\partial\theta}\overbrace{\int f_\theta(x)dx}^{=1} = 0$$

DEFINITION 3.12 – *L'information de Fisher associé à* $(X_1, ..., X_n)$

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

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

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

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

PROPOSITION 3.14 –

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

en effet,

$$I_n(\theta) = \text{Var}\left(\frac{\partial}{\partial\theta}\log L_n(\theta)\right) = \text{Var}\left(\sum_{i=1}^n \frac{\partial}{\partial\theta}\log f_\theta(X_i)\right) \underbrace{=}_{\text{independance}} = \sum_{i=1}^n \text{Var}\left(\frac{\partial}{\partial\theta}\log f_\theta(X_i)\right) =$$

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

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

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

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

3.3 Inp. de Fisher et derivé seconde

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

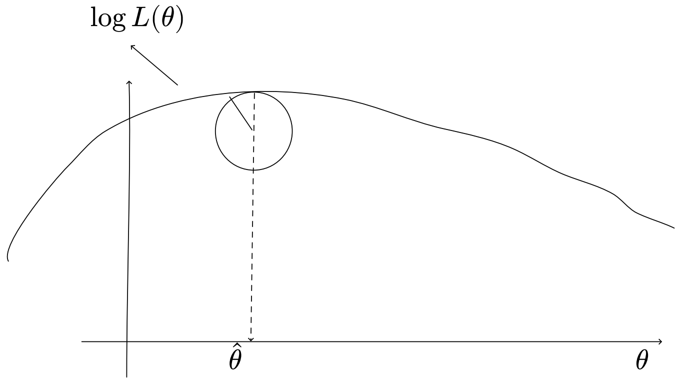
$$I_n(\theta) = -E_\theta\left[\frac{\partial^2 \log L_n(\theta)}{\partial\theta^2}\right]$$

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

$n = 1$

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

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



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 $g : \Theta \to \mathbb{R}$

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

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

Proof.

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

□

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

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

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

Étude asymptotique des estimateurs

§4

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

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

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

Asymptotique: $n \to +\infty$,

$$n \operatorname{Var}\left(\hat{\theta}_n\right) \underset{n\to+\infty}{\longrightarrow} \frac{1}{I(\theta)}$$

4.1 Convergences

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

- convergence en loi: $X_n \underset{n\to+\infty}{\overset{\mathcal{L}}{\longrightarrow}} X$ ssi $P(X_n \leq x) \longrightarrow P(X \leq x)$ en tout point de continuité de x.

Lemma 4.1 (lemme de Portmanteau) – *Caractérisations équivalentes:*

- *Pour toute fonction continue bornée h,*

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

$\Rightarrow$ *la convergence en loi est stable par passage aux fonctions continues (LAC) MAIS il est en général faux que si $X_n \overset{\mathcal{L}}{\longrightarrow} X$ et $Y_n \overset{\mathcal{L}}{\longrightarrow} Y$ alors $\binom{X_n}{Y_n} \overset{\mathcal{L}}{\longrightarrow} \binom{X}{Y}$*

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 $\begin{cases} \text{convergence en loi de } X_n \text{ et } Y_n \\ \text{convergence en loi du couple } \binom{X_n}{Y_n} \end{cases}$

2. $$\text{si } \begin{cases} X_n \overset{P}{\longrightarrow} X \\ Y_n \overset{p}{\longrightarrow} Y \end{cases} \Longrightarrow \begin{pmatrix} X_n \\ Y_n \end{pmatrix} \overset{P}{\longrightarrow} \begin{pmatrix} X \\ Y \end{pmatrix} \Longrightarrow \begin{pmatrix} X_n \\ Y_n \end{pmatrix} \overset{\mathcal{L}}{\longrightarrow} \begin{pmatrix} X \\ Y \end{pmatrix}$$

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

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

en appliquant le LAX,

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

4.2 Consistance des estimateurs

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

$$\text{Biais}\left(\hat{\theta}_n, \theta\right) = E\left[\hat{\theta}_n\right] - \theta \underset{n\to+\infty}{\longrightarrow} 0$$

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* $X_n \longrightarrow X$ *dans* L_1

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

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

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

Outils pour montrer la consistance:

- *LGN*
- *si* $R\left(\hat{\theta}_n, \theta\right) \longrightarrow 0$ *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.*

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

LGN: $\frac{1}{n}\sum_i X_i^2 \xrightarrow{P} E[X^2] = \tau^2 - \mu^2$

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

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

4.3 Normalité asymptotique