

# Notes de Cours d'Inférence Statistique

YEHOR KOROTENKO

31 January 2026

## Table of contents

<b>1 Introduction</b>	<b>2</b>
1.1 Evaluation	2
1.2 Modèle Statistique	2
1.3 Estimateurs	2
1.4 Risque quadratique	3
1.4.1 Exemple : Modèle de Poisson	3
1.5 Consistance	4
1.5.1 Exemple : Retour au modèle de Poisson	4
1.5.2 Méthode “Plug-in”	4
<b>2 Estimateurs</b>	<b>5</b>
2.1 Cadre paramétrique	5
2.1.1 Modèle statistique paramétrique	5
2.2 Méthode des moments	5
2.3 Rendu sur le L.A.C.	7
2.3.1 Variance empirique	8
2.4 Méthode de maximum de vraisemblance	8
2.4.1 Modèle donné	8
2.4.2 En pratique	8
<b>3 Information de Fisher, efficacité</b>	<b>11</b>
3.1 Modèle régulier	11
3.2 Score et Information de Fisher	12
3.3 Inp. de Fisher et dérivé seconde	13
3.4 Inégalité de Cramer - Rao	14

# 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)** – Une 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.

**EXEMPLE 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  $P_\theta$ , donc sur  $\theta$ .

**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  $\theta^*$  la vraie valeur inconnue telle que  $P_{X_i} = P_{\theta^*}$ .

**DEFINITION 1.6 (ESTIMATEUR)** – Un estimateur de  $\theta$  est une fonction de l'échantillon  $(X_1, \dots, X_n)$  mesurable et indépendante de  $\theta$  (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  $\theta$ . 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  $P_\theta$  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  $P_\theta$ . Soit  $\hat{\theta}_n = h(X_1, \dots, X_n)$ .  $\hat{\theta}_n$  est un estimateur consistant (ou convergent) de  $\theta$  ssi :

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

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

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

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

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

**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-paramétré.

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

**EXAMPLE 2.2** – Bernoulli  $(\theta)$ , Exp  $(\theta)$ ,  $\mathcal{N}(\mu, \sigma^2)$ , 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

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

- 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  $\theta$  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

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

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,

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

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

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

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

- $X_1, \dots, X_n$  i.i.d. de la loi  $P_\theta$  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:

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

$$\Rightarrow \hat{\theta}_M = \frac{\bar{X}}{1-\bar{X}}, P_\theta(\bar{X} = 1) = P_\theta(X_1 = X_2 = \dots = X_n = 1) = 0$$

### 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

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

$g(x, y) = \frac{1}{\sqrt{y-x^2}} \Rightarrow \hat{\theta}^M$  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) –**

$$\left( \begin{array}{c} X_n \\ Y_n \end{array} \right) \xrightarrow{P} \left( \begin{array}{c} X \\ Y \end{array} \right) \quad \text{ssi} \quad \begin{cases} X_n \xrightarrow{P} X \\ Y_n \xrightarrow{P} Y \end{cases}$$

**Proof.**

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

- $\Leftarrow$  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 convergence en loi!

□

### 2.3.1 Variance empirique

Si la  $X_i$  admettent une espérance  $\mu$  et une variance  $\sigma^2$ , 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

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

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

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

#### 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  $\mu$  (positive  $\sigma$  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  $\mu$ .

### 2.4.2 En pratique

- soit  $E$  au plus dénombrable:  $\mu$  = 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  $(\theta)$ ,  $X_i = 1$ , 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  $f_\theta$  est la densité usuelle



$f_\theta$  densité de  $P_\theta$

**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:

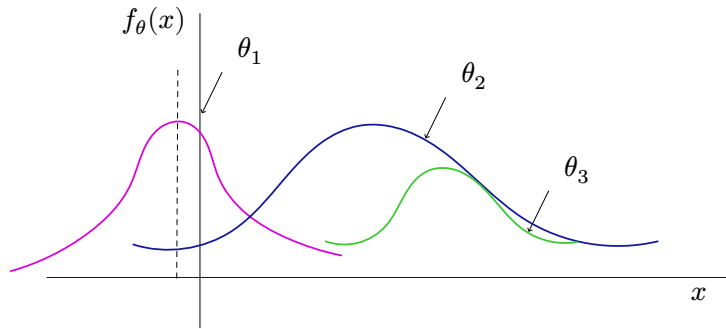
$$\forall \theta \in \Theta, L_n(\theta) \leq L_n(\hat{\theta})$$

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



**EXAMPLE 2.19** –

- Bernoulli( $\theta$ ),  $f_\theta(x) = \theta^x(1-\theta)^{1-x}$ ,  $X_i$  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  $f_\theta$  densité de  $P_\theta$

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

Étant donné  $(X_1, \dots, X_n)$ , i.i.d. de loi  $P_\theta$  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  $\hat{\theta}$  EMV<sup>1</sup> de  $\theta$ ,  $g(\hat{\theta})$  est un EMV de  $g(\theta)$

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

### 3.1 Modèle régulier

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

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

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

est intégrable  $(L, \mu)$  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  $\Theta$ .

**NOTATION 3.3** – On note la dérivée de  $f_{\theta(x)}$  par rapport à  $\theta$ :  $\frac{\partial f_\theta}{\partial \theta}(x)$  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  $C^\infty$  sur  $\Theta = ]0, +\infty[$ ,  $\text{Supp } f_\theta = \mathbb{R}_+$

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

<sup>1</sup>EMV = 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( $\theta$ ),  $x = 0.1$ ,  $f_{\theta(0)} = 1 - \theta$ ,  $f_{\theta(1)} = \theta$ , 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  $P_{\theta}$ ,  $f_{\theta}$

**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  $\theta$ , 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 ( $P_\theta$ ),  $n = 1$

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

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

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

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

**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,

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

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

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

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

### 3.3 Inp. de Fisher et dérivé 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

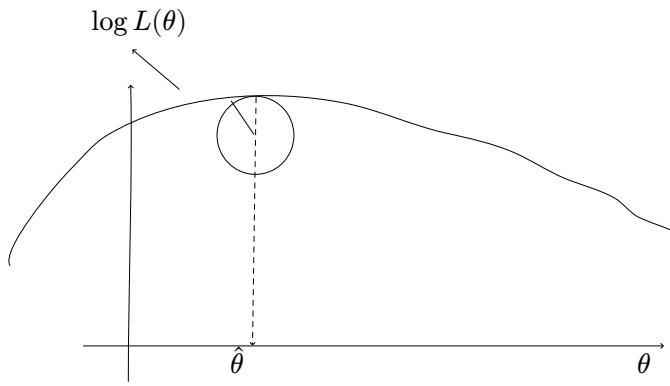
$$I_n(\theta) = -E_\theta \left[ \frac{\partial^2 \log L_n(\theta)}{\partial \theta^2} \right]$$

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

$$n = 1$$

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

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



Si courbe très “piquée” 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  $\theta$ ,  $I(\theta) > 0$ , alors pour tout estimateur  $T = T(X_1, \dots, X_n)$  sans biais,  $E_\theta T^2 < +\infty$ , on a

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

*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

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

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