

MAP 433 : Introduction aux méthodes statistiques. Cours 3

14 février 2014

MAP 433 :
Introduction
aux méthodes
statistiques.
Cours 3

Rappel succinct
du Cours 2

Modélisation
statistique

Méthodes
d'estimation
pour le modèle
de densité



Aujourd'hui

- 1 Rappel succinct du Cours 2
- 2 Modélisation statistique
 - Expérience statistique
 - Expériences dominées
 - Modèle de densité
- 3 Méthodes d'estimation pour le modèle de densité
 - Méthode des moments
 - Z-estimation
 - M-estimation
 - Principe de maximum de vraisemblance

MAP 433 :
Introduction
aux méthodes
statistiques.
Cours 3

Rappel succinct
du Cours 2

Modélisation
statistique

Méthodes
d'estimation
pour le modèle
de densité



Estimation de fonctionnelles régulières

- $X_1, \dots, X_n \sim_{i.i.d.} F$.
- Estimation de $T(F) = h(\int_{\mathbb{R}} g(x)dF(x))$.
- Estimateur par **substitution** ou *plug-in* de $T(F)$:

$$T(\hat{F}_n) = h\left(\frac{1}{n} \sum_{i=1}^n g(X_i)\right)$$

MAP 433 :
Introduction
aux méthodes
statistiques.
Cours 3

Rappel succinct
du Cours 2

Modélisation
statistique

Méthodes
d'estimation
pour le modèle
de densité



Limites de l'approche empirique

L'estimation de $T(F)$ par $T(\hat{F}_n)$ n'est pas toujours possible :

- La fonctionnelle $F \rightsquigarrow T(F)$ n'est pas « régulière »,
- La paramétrisation $F \rightsquigarrow T(F)$ ne donne **pas** lieu à une **forme analytique simple**. \rightarrow autres approches.

Exemple. **Hypothèse** : F admet une densité f par rapport à la mesure de Lebesgue, **continue** (= pp à une fonction continue f).

$$T(F) = f(x_0), \quad x_0 \in \mathbb{R} \text{ (donné).}$$

On ne **peut pas prendre** comme estimateur $\hat{F}'_n(x_0)$ car \hat{F}_n n'est pas différentiable (constante par morceaux...)

MAP 433 :
Introduction
aux méthodes
statistiques.
Cours 3

Rappel succinct
du Cours 2

Modélisation
statistique

Méthodes
d'estimation
pour le modèle
de densité



Limites de l'approche empirique

L'estimation de $T(F)$ par $T(\hat{F}_n)$ n'est pas toujours **souhaitable** :

- Souvent on dispose d'information **a priori** supplémentaire : F appartient à une sous-classe **particulière** de distributions, et il y a des choix plus judicieux que l'estimateur par plug-in.

MAP 433 :
Introduction
aux méthodes
statistiques.
Cours 3

Rappel succinct
du Cours 2

Modélisation
statistique

Méthodes
d'estimation
pour le modèle
de densité



Conclusion

- **L'approche empirique**, basée sur \hat{F}_n permet d'estimer une distribution inconnue F ou une fonctionnelle $T(F) \in \mathbb{R}$ à partir d'un n -échantillon, mais
 - reste très générale, pas toujours adaptée.
 - restreinte à la situation d'un n -échantillon.
- Formalisation de la notion **d'expérience statistique**
 - incorporation d'information de modélisation **supplémentaire**.
 - construction de méthodes d'estimation – de décision – **systematiques**.
 - comparaison et **optimalité** des méthodes.

MAP 433 :
Introduction
aux méthodes
statistiques.
Cours 3

Rappel succinct
du Cours 2

Modélisation
statistique

Méthodes
d'estimation
pour le modèle
de densité



Expérience statistique

Consiste à identifier :

- **Des observations**

$$\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$$

considérées comme des **réalisations** de variables aléatoires $Z = (X_1, \dots, X_n)$ de loi \mathbb{P}^Z .

- **Une famille de lois**

$$\{\mathbb{P}_\vartheta, \vartheta \in \Theta\}.$$

- **Une problématique** : retrouver le paramètre ϑ tel que $\mathbb{P}^Z = \mathbb{P}_\vartheta$ (estimation) ou bien prendre une décision sur une propriété relative à ϑ (test).

MAP 433 :
Introduction
aux méthodes
statistiques.
Cours 3

Rappel succinct
du Cours 2

Modélisation
statistique

Expérience
statistique

Expériences
dominées
Modèle de
densité

Méthodes
d'estimation
pour le modèle
de densité



Expérience statistique

- Approche générale empirique :
 - $\vartheta = F, \Theta$ est l'ensemble de toutes les lois (s'il s'agit de l'estimation de F);
 - $\vartheta = F, \Theta$ est l'ensemble de toutes les lois vérifiant une hypothèse très générale, par exemple, la bornitude d'un moment (s'il s'agit de l'estimation de $T(F)$).
- Approche paramétrique : **on suppose** que F appartient à une **famille de lois connue** indexée par un paramètre ϑ de dimension finie : $\vartheta \in \Theta \subset \mathbb{R}^d$.
 - Exemple : $\Theta = \mathbb{R}$,

$$X_i = \vartheta + \xi_i, \quad i = 1, \dots, n,$$

ξ_i v.a. i.i.d. de densité **connue** f sur \mathbb{R} et $\mathbb{E}(X_i) = \vartheta$.

Question : en utilisant cette information supplémentaire, peut-on construire un estimateur plus performant que l'estimateur \bar{X}_n basé sur l'approche empirique ?

MAP 433 :
Introduction
aux méthodes
statistiques.
Cours 3

Rappel succinct
du Cours 2

Modélisation
statistique

Expérience
statistique

Expériences
dominées
Modèle de
densité

Méthodes
d'estimation
pour le modèle
de densité



Expérience statistique

- En écrivant

$$X_i = \vartheta + \xi_i, \quad i = 1, \dots, n,$$

ξ_i v.a. i.i.d. de densité connue f , nous précisons la forme de la loi \mathbb{P}_ϑ de (X_1, \dots, X_n) :

$$\mathbb{P}_\vartheta [A] = \int_A \left(\prod_{i=1}^n f(x_i - \vartheta) \right) dx_1 \dots dx_n,$$

pour tout $A \in \mathcal{B}(\mathbb{R}^n)$.

MAP 433 :
Introduction
aux méthodes
statistiques.
Cours 3

Rappel succinct
du Cours 2

Modélisation
statistique
Expérience
statistique
Expériences
dominées
Modèle de
densité
Méthodes
d'estimation
pour le modèle
de densité



Expérience statistique

Définition

Une expérience (un modèle) statistique \mathcal{E} est le triplet

$$\mathcal{E} = (\mathfrak{Z}, \mathcal{Z}, \{ \mathbb{P}_\vartheta, \vartheta \in \Theta \}),$$

avec

- $(\mathfrak{Z}, \mathcal{Z})$ espace mesurable (souvent $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$),
- $\{ \mathbb{P}_\vartheta, \vartheta \in \Theta \}$ famille de probabilités définies **simultanément** sur le même espace $(\mathfrak{Z}, \mathcal{Z})$,
- ϑ est le **paramètre inconnu**, et Θ est **l'ensemble des paramètres connu**.

MAP 433 :
Introduction
aux méthodes
statistiques.
Cours 3

Rappel succinct
du Cours 2

Modélisation
statistique
Expérience
statistique
Expériences
dominées
Modèle de
densité
Méthodes
d'estimation
pour le modèle
de densité



Expérience engendrée par (X_1, \dots, X_n)

- **Traitement sur un exemple** : on observe

$$Z = (X_1, \dots, X_n), \quad X_i = \vartheta + \xi_i,$$

ξ_i v.a. i.i.d. de densité connue f .

- La famille de lois $\{ \mathbb{P}_\vartheta^n, \vartheta \in \Theta = \mathbb{R} \}$ est définie sur $\mathfrak{Z} = \mathbb{R}^n$ par

$$\mathbb{P}_\vartheta^n [A] = \int_A \left(\prod_{i=1}^n f(x_i - \vartheta) \right) dx_1 \dots dx_n,$$

pour $A \in \mathcal{Z} = \mathcal{B}(\mathbb{R}^n)$ (et \mathbb{P}^Z est l'une des \mathbb{P}_ϑ^n).

- Expérience **engendrée par l'observation Z** :

$$\mathcal{E}^n = (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \{ \mathbb{P}_\vartheta^n, \vartheta \in \Theta \}).$$

MAP 433 :
Introduction
aux méthodes
statistiques.
Cours 3

Rappel succinct
du Cours 2

Modélisation
statistique
Expérience
statistique
Expériences
dominées
Modèle de
densité
Méthodes
d'estimation
pour le modèle
de densité



Expérience (modèle) paramétrique, non-paramétrique

- Si Θ peut être « pris » comme un sous-ensemble de \mathbb{R}^d : **expérience (=modèle) paramétrique**.
- Sinon (par exemple si le paramètre ϑ est un élément d'un espace fonctionnel) : **expérience (=modèle) non-paramétrique**.

MAP 433 :
Introduction
aux méthodes
statistiques.
Cours 3

Rappel succinct
du Cours 2

Modélisation
statistique
Expérience
statistique
Expériences
dominées
Modèle de
densité
Méthodes
d'estimation
pour le modèle
de densité



Expériences dominées

- On fait une hypothèse minimale de « complexité » sur le modèle statistique. **But** : ramener l'étude de la famille

$$\{\mathbb{P}_\vartheta, \vartheta \in \Theta\}$$

à l'étude d'une famille de fonctions

$$\{z \in \mathfrak{Z} \rightsquigarrow f(\vartheta, z) \in \mathbb{R}_+, \vartheta \in \Theta\}.$$

- Via la notion de **domination**. Si μ, ν sont deux mesures σ -finies sur \mathfrak{Z} , alors μ **domine** ν (notation $\nu \ll \mu$) si

$$\mu[A] = 0 \Rightarrow \nu[A] = 0.$$

MAP 433 :
Introduction
aux méthodes
statistiques.
Cours 3

Rappel succinct
du Cours 2

Modélisation
statistique
Expérience
statistique
Expériences
dominées
Modèle de
densité
Méthodes
d'estimation
pour le modèle
de densité



Théorème de Radon-Nikodym

Théorème

Si $\nu \ll \mu$, il existe une fonction positive

$$z \rightsquigarrow p(z) \stackrel{\text{notation}}{=} \frac{d\nu}{d\mu}(z),$$

définie μ -p.p., μ -intégrable, telle que

$$\nu[A] = \int_A p(z) \mu(dz) = \int_A \frac{d\nu}{d\mu}(z) \mu(dz), \quad A \in \mathfrak{Z}.$$

MAP 433 :
Introduction
aux méthodes
statistiques.
Cours 3

Rappel succinct
du Cours 2

Modélisation
statistique
Expérience
statistique
Expériences
dominées
Modèle de
densité
Méthodes
d'estimation
pour le modèle
de densité



Expérience dominée

Définition

Une expérience statistique $\mathcal{E} = (\mathfrak{Z}, \mathfrak{Z}, \{\mathbb{P}_\vartheta, \vartheta \in \Theta\})$ est **dominée** par la mesure σ -finie μ définie sur \mathfrak{Z} si

$$\forall \vartheta \in \Theta : \mathbb{P}_\vartheta \ll \mu.$$

On appelle **densités** de la famille $\{\mathbb{P}_\vartheta, \vartheta \in \Theta\}$ la famille de fonctions (définies μ -p.p.)

$$z \rightsquigarrow \frac{d\mathbb{P}_\vartheta}{d\mu}(z), \quad z \in \mathfrak{Z}, \vartheta \in \Theta.$$

MAP 433 :
Introduction
aux méthodes
statistiques.
Cours 3

Rappel succinct
du Cours 2

Modélisation
statistique
Expérience
statistique
Expériences
dominées
Modèle de
densité
Méthodes
d'estimation
pour le modèle
de densité



Densité, régression

Deux classes d'expériences statistiques **dominées** fondamentales :

- Le modèle de **densité**
- Le modèle de **régression**

MAP 433 :
Introduction
aux méthodes
statistiques.
Cours 3

Rappel succinct
du Cours 2

Modélisation
statistique
Expérience
statistique
Expériences
dominées
Modèle de
densité
Méthodes
d'estimation
pour le modèle
de densité



Modèle de densité (paramétrique)

- On observe un n -échantillon de v.a.r. X_1, \dots, X_n .
- La loi des X_i appartient à $\{\mathbb{P}_\vartheta, \vartheta \in \Theta\}$, famille de **probabilités sur \mathbb{R} , dominée** par une mesure (σ -finie) $\mu(dx)$ sur \mathbb{R} .
- La loi de (X_1, \dots, X_n) s'écrit

$$\begin{aligned} \mathbb{P}_\vartheta^n(dx_1 \cdots dx_n) &= \mathbb{P}_\vartheta(dx_1) \otimes \cdots \otimes \mathbb{P}_\vartheta(dx_n) \\ &\ll \mu(dx_1) \otimes \cdots \otimes \mu(dx_n) \\ &\stackrel{\text{notation}}{=} \mu^n(dx_1 \cdots dx_n) \end{aligned}$$

◀ ▶ ⏪ ⏩ ⏴ ⏵ ⏶ ⏷ ⏸ ⏹ ⏺ ⏻ ⏼ ⏽ ⏾ ⏿ 🔍 ↻

MAP 433 :
Introduction
aux méthodes
statistiques.
Cours 3

Rappel succinct
du Cours 2

Modélisation
statistique

Expérience
statistique
Expériences
dominées
Modèle de
densité

Méthodes
d'estimation
pour le modèle
de densité

Modèle de densité (paramétrique)

- **Densité du modèle** : on part de

$$f(\vartheta, x) = \frac{d\mathbb{P}_\vartheta}{d\mu}(x), \quad x \in \mathbb{R}$$

et

$$\frac{d\mathbb{P}_\vartheta^n}{d\mu^n}(x_1, \dots, x_n) = \prod_{i=1}^n f(\vartheta, x_i), \quad x_1, \dots, x_n \in \mathbb{R}.$$

- **L'expérience statistique** engendrée par (X_1, \dots, X_n) s'écrit :

$$\mathcal{E}^n = \left(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \{ \mathbb{P}_\vartheta^n, \vartheta \in \Theta \} \right), \quad \Theta \subset \mathbb{R}^d.$$

◀ ▶ ⏪ ⏩ ⏴ ⏵ ⏶ ⏷ ⏸ ⏹ ⏺ ⏻ ⏼ ⏽ ⏾ ⏿ 🔍 ↻

MAP 433 :
Introduction
aux méthodes
statistiques.
Cours 3

Rappel succinct
du Cours 2

Modélisation
statistique

Expérience
statistique
Expériences
dominées
Modèle de
densité

Méthodes
d'estimation
pour le modèle
de densité

Exemple 1 : modèle de densité gaussienne univariée

- $X_i \sim \mathcal{N}(m, \sigma^2)$, avec

$$\vartheta = (m, \sigma^2) \in \Theta = \mathbb{R} \times \mathbb{R}_+ \setminus \{0\}.$$

$$\begin{aligned} \mathbb{P}_\vartheta(dx) = f(\vartheta, x)dx &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right) dx \\ &\ll \mu(dx) = dx. \end{aligned}$$

- Puis

$$\begin{aligned} \frac{d\mathbb{P}_\vartheta^n}{d\mu^n}(x_1, \dots, x_n) &= \prod_{i=1}^n f(\vartheta, x_i) \\ &= (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - m)^2\right), \end{aligned}$$

avec $x_1, \dots, x_n \in \mathbb{R}$.

◀ ▶ ⏪ ⏩ ⏴ ⏵ ⏶ ⏷ ⏸ ⏹ ⏺ ⏻ ⏼ ⏽ ⏾ ⏿ 🔍 ↻

MAP 433 :
Introduction
aux méthodes
statistiques.
Cours 3

Rappel succinct
du Cours 2

Modélisation
statistique

Expérience
statistique
Expériences
dominées
Modèle de
densité

Méthodes
d'estimation
pour le modèle
de densité

Exemple 2 : modèle de Bernoulli

- $X_i \sim \text{Bernoulli}(\vartheta)$, avec $\vartheta \in \Theta = [0, 1]$.

$$\begin{aligned} \mathbb{P}_\vartheta(dx) &= (1 - \vartheta) \delta_0(dx) + \vartheta \delta_1(dx) \\ &\ll \mu(dx) = \delta_0(dx) + \delta_1(dx) \quad (\text{mesure de comptage}). \end{aligned}$$

- Puis

$$\frac{d\mathbb{P}_\vartheta}{d\mu}(x) = (1 - \vartheta) 1_{\{x=0\}} + \vartheta 1_{\{x=1\}} = \vartheta^x (1 - \vartheta)^{1-x}$$

avec $x \in \{0, 1\}$ (et 0 sinon), et

$$\frac{d\mathbb{P}_\vartheta^n}{d\mu^n}(x_1 \cdots x_n) = \prod_{i=1}^n \vartheta^{x_i} (1 - \vartheta)^{1-x_i},$$

avec $x_i \in \{0, 1\}$ (et 0 sinon).

◀ ▶ ⏪ ⏩ ⏴ ⏵ ⏶ ⏷ ⏸ ⏹ ⏺ ⏻ ⏼ ⏽ ⏾ ⏿ 🔍 ↻

MAP 433 :
Introduction
aux méthodes
statistiques.
Cours 3

Rappel succinct
du Cours 2

Modélisation
statistique

Expérience
statistique
Expériences
dominées
Modèle de
densité

Méthodes
d'estimation
pour le modèle
de densité

Exemple 3 : temps de panne « arrêtés »

- On observe X_1, \dots, X_n , où $X_i = Y_i \wedge T$, avec Y_i lois exponentielles de paramètre ϑ et T temps fixe (censure).
- Cas 1 : $T = \infty$ (pas de censure). Alors $\vartheta \in \Theta = \mathbb{R}_+ \setminus \{0\}$ et

$$\mathbb{P}_\vartheta(dx) = \vartheta \exp(-\vartheta x) \mathbf{1}_{\{x \geq 0\}} dx \ll \mu(dx) = dx$$

et

$$\frac{d\mathbb{P}_\vartheta^n}{d\mu^n}(x_1, \dots, x_n) = \vartheta^n \exp\left(-\vartheta \sum_{i=1}^n x_i\right),$$

avec $x_i \in \mathbb{R}_+$ (et 0 sinon).

- Cas 2 : Comment s'écrit le modèle dans la cas où $T < \infty$ (présence de censure) ? Comment choisir μ ?

⏪ ⏩ ⏴ ⏵ ⏶ ⏷ ⏸ ⏹ ⏺ ⏻ ⏼ ⏽ ⏾ ⏿ 🔍 ↻

MAP 433 :
Introduction
aux méthodes
statistiques.
Cours 3

Rappel succinct
du Cours 2

Modélisation
statistique
Expérience
statistique
Expériences
dominées
Modèle de
densité
Méthodes
d'estimation
pour le modèle
de densité

Exemple : temps de panne « arrêtés »

- Loi $\mathbb{P}_\vartheta(dx)$ de $X = Y \wedge T$: $Y \sim$ exponentielle de paramètre ϑ :

$$X = Y \mathbf{1}_{\{Y < T\}} + T \mathbf{1}_{\{Y \geq T\}}$$

d'où

$$\begin{aligned} \mathbb{P}_\vartheta(dx) &= \vartheta e^{-\vartheta x} \mathbf{1}_{\{0 \leq x < T\}} dx + \left(\int_T^{+\infty} \vartheta e^{-\vartheta y} dy \right) \delta_T(dx) \\ &= \vartheta e^{-\vartheta x} \mathbf{1}_{\{0 \leq x < T\}} dx + e^{-\vartheta T} \delta_T(dx) \\ &\ll \mu(dx) = dx + \delta_T(dx) \quad (\text{par exemple}). \end{aligned}$$

⏪ ⏩ ⏴ ⏵ ⏶ ⏷ ⏸ ⏹ ⏺ ⏻ ⏼ ⏽ ⏾ ⏿ 🔍 ↻

MAP 433 :
Introduction
aux méthodes
statistiques.
Cours 3

Rappel succinct
du Cours 2

Modélisation
statistique
Expérience
statistique
Expériences
dominées
Modèle de
densité
Méthodes
d'estimation
pour le modèle
de densité

Exemple : temps de panne « arrêtés » (fin)

- Alors, pour ce choix de mesure dominante

$$\frac{d\mathbb{P}_\vartheta}{d\mu}(x) = \vartheta e^{-\vartheta x} \mathbf{1}_{\{0 \leq x < T\}} + e^{-\vartheta T} \mathbf{1}_{\{x=T\}}$$

- Finalement,

$$\mathbb{P}_\vartheta^n(dx_1, \dots, dx_n) \ll \mu^n(dx_1, \dots, dx_n) = \bigotimes_{i=1}^n [dx_i + \delta_T(dx_i)]$$

et

$$\begin{aligned} \frac{d\mathbb{P}_\vartheta^n}{d\mu^n}(x_1, \dots, x_n) &= \prod_{i=1}^n (\vartheta e^{-\vartheta x_i} \mathbf{1}_{\{0 \leq x_i < T\}} + e^{-\vartheta T} \mathbf{1}_{\{x_i=T\}}) \\ &= \vartheta^{N_n(T)} e^{-\vartheta \sum_{i=1}^n x_i \mathbf{1}_{\{x_i < T\}}} e^{-\vartheta T(n - N_n(T))}, \end{aligned}$$

avec $0 \leq x_i \leq T$ et 0 sinon, et $N_n(T) = \sum_{i=1}^n \mathbf{1}_{\{x_i < T\}}$.

⏪ ⏩ ⏴ ⏵ ⏶ ⏷ ⏸ ⏹ ⏺ ⏻ ⏼ ⏽ ⏾ ⏿ 🔍 ↻

MAP 433 :
Introduction
aux méthodes
statistiques.
Cours 3

Rappel succinct
du Cours 2

Modélisation
statistique
Expérience
statistique
Expériences
dominées
Modèle de
densité
Méthodes
d'estimation
pour le modèle
de densité

Méthodes d'estimation

- Méthode de substitution (ou des moments)
- Z-estimation
- M-estimation
- Le principe du maximum de vraisemblance

⏪ ⏩ ⏴ ⏵ ⏶ ⏷ ⏸ ⏹ ⏺ ⏻ ⏼ ⏽ ⏾ ⏿ 🔍 ↻

MAP 433 :
Introduction
aux méthodes
statistiques.
Cours 3

Rappel succinct
du Cours 2

Modélisation
statistique
Méthodes
d'estimation
pour le modèle
de densité
Méthode des
moments
Z-estimation
M-estimation
Principe de
maximum de
vraisemblance

Méthode des moments : dimension 1

- $X_1, \dots, X_n \sim_{\text{i.i.d.}} \mathbb{P}_\vartheta$, avec $\vartheta \in \Theta \subset \mathbb{R}$.
- **Principe** : trouver $g : \mathbb{R} \rightarrow \mathbb{R}$ (en général $g(x) = x^k$) et $h : \mathbb{R} \rightarrow \mathbb{R}$ régulières de sorte que

$$\vartheta = h(\mathbb{E}_\vartheta[g(X)]) = h\left(\int_{\mathbb{R}} g(x) dF_\vartheta(x)\right) = T(F_\vartheta)$$

et T fonctionnelle régulière de la distribution inconnue F_ϑ .

- **Estimateur** : « plug-in »

$$\hat{\vartheta}_n = h\left(\frac{1}{n} \sum_{i=1}^n g(X_i)\right).$$

Navigation icons

MAP 433 :
Introduction
aux méthodes
statistiques.
Cours 3

Rappel succinct
du Cours 2

Modélisation
statistique

Méthodes
d'estimation
pour le modèle
de densité

Méthode des
moments

Z-estimation

M-estimation

Principe de
maximum de
vraisemblance

Méthode des moments

- **Précision d'estimation** via les techniques empiriques :

$$\sqrt{n}(\hat{\vartheta}_n - \vartheta) \xrightarrow{d} \mathcal{N}(0, h'(\mathbb{E}_\vartheta[g(X)])^2 \text{Var}_\vartheta[g(X)])$$

en loi sous \mathbb{P}_ϑ et la variance asymptotique dépend en général de $\vartheta \rightarrow$ élimination par estimation préliminaire licite via le lemme de Slutsky.

- **Exemple** : $X_1, \dots, X_n \sim_{\text{i.i.d.}}$ exponentielle de paramètre ϑ .
On a

$$\mathbb{E}_\vartheta[X] = \frac{1}{\vartheta},$$

l'estimateur par moment associé s'écrit

$$\hat{\vartheta}_n = \frac{1}{\bar{X}_n}.$$

Navigation icons

MAP 433 :
Introduction
aux méthodes
statistiques.
Cours 3

Rappel succinct
du Cours 2

Modélisation
statistique

Méthodes
d'estimation
pour le modèle
de densité

Méthode des
moments

Z-estimation

M-estimation

Principe de
maximum de
vraisemblance

Exemple en dimension $d > 1$

- $X_1, \dots, X_n \sim_{\text{i.i.d.}}$ Béta(α, β), de densité

$$x \rightsquigarrow \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1} \mathbf{1}_{\{0 < x < 1\}},$$

- Le paramètre est $\vartheta = (\alpha, \beta) \in \Theta = \mathbb{R}_+ \setminus \{0\} \times \mathbb{R}_+ \setminus \{0\}$.
- On a

$$\mathbb{E}_\vartheta[X] = \frac{\alpha}{\alpha + \beta}, \quad \mathbb{E}_\vartheta[X^2] = \frac{\alpha(\alpha + 1)}{(\alpha + \beta + 1)(\alpha + \beta)}$$

Navigation icons

MAP 433 :
Introduction
aux méthodes
statistiques.
Cours 3

Rappel succinct
du Cours 2

Modélisation
statistique

Méthodes
d'estimation
pour le modèle
de densité

Méthode des
moments

Z-estimation

M-estimation

Principe de
maximum de
vraisemblance

Exemple en dimension $d > 1$

- L'estimateur par moment $\hat{\vartheta}_n = (\hat{\vartheta}_n^{(1)}, \hat{\vartheta}_n^{(2)})$ associé est défini par

$$\begin{cases} \bar{X}_n &= \frac{\hat{\vartheta}_n^{(1)}}{\hat{\vartheta}_n^{(1)} + \hat{\vartheta}_n^{(2)}} \\ \frac{1}{n} \sum_{i=1}^n X_i^2 &= \frac{\hat{\vartheta}_n^{(1)}(\hat{\vartheta}_n^{(1)} + 1)}{(\hat{\vartheta}_n^{(1)} + \hat{\vartheta}_n^{(2)} + 1)(\hat{\vartheta}_n^{(1)} + \hat{\vartheta}_n^{(2)})}. \end{cases}$$

- **Etude asymptotique** via le TCL multidimensionnel et la méthode « delta » multidimensionnelle.

Navigation icons

MAP 433 :
Introduction
aux méthodes
statistiques.
Cours 3

Rappel succinct
du Cours 2

Modélisation
statistique

Méthodes
d'estimation
pour le modèle
de densité

Méthode des
moments

Z-estimation

M-estimation

Principe de
maximum de
vraisemblance

Limites de la méthode des moments

- Méthode **non systématique**
- Représentation pas toujours explicite
- Choix de la fonction g , notion d'optimalité parmi une classe d'estimateurs...
- **Généralisation** : Z-estimation (ou estimation par méthode des moments généralisés, GMM= *generalized method of moments*).

MAP 433 :
Introduction
aux méthodes
statistiques.
Cours 3

Rappel succinct
du Cours 2

Modélisation
statistique

Méthodes
d'estimation
pour le modèle
de densité

Méthode des
moments
Z-estimation
M-estimation
Principe de
maximum de
vraisemblance

⏪ ⏩ ⏴ ⏵ ⏶ ⏷ ⏸ ⏹ ⏺ ⏻ ⏼ ⏽ ⏾ ⏿ 🔍 ↻

Z-estimation

- La méthode des moments (en dimension 1) est basée sur l'inversibilité de la fonction

$$m_g(\vartheta) = \int_{\mathbb{R}} g(x) \mathbb{P}_{\vartheta}(dx)$$

i.e. pour tout $\vartheta \in \Theta$

$$\int_{\mathbb{R}} (m_g(\vartheta) - g(x)) \mathbb{P}_{\vartheta}(dx) = 0.$$

- Principe de construction d'un Z-estimateur : **remplacer** $m_g(\vartheta) - g(x)$ par une fonction $\phi(\vartheta, x) : \Theta \times \mathbb{R} \rightarrow \mathbb{R}$ **arbitraire** telle que

$$\forall \vartheta \in \Theta, \int_{\mathbb{R}} \phi(\vartheta, x) \mathbb{P}_{\vartheta}(dx) = 0.$$

⏪ ⏩ ⏴ ⏵ ⏶ ⏷ ⏸ ⏹ ⏺ ⏻ ⏼ ⏽ ⏾ ⏿ 🔍 ↻

MAP 433 :
Introduction
aux méthodes
statistiques.
Cours 3

Rappel succinct
du Cours 2

Modélisation
statistique

Méthodes
d'estimation
pour le modèle
de densité

Méthode des
moments
Z-estimation
M-estimation
Principe de
maximum de
vraisemblance

Z-estimation

- Résoudre l'équation **empirique** associée :

$$\frac{1}{n} \sum_{i=1}^n \phi(a, X_i) = 0 \text{ pour } a \in \Theta.$$

Définition

On appelle **Z-estimateur** associé à ϕ tout estimateur $\hat{\vartheta}_n$ satisfaisant

$$\sum_{i=1}^n \phi(\hat{\vartheta}_n, X_i) = 0$$

- Il n'y a pas unicité de $\hat{\vartheta}_n$ (à ce niveau).
- Programme **Etablir des conditions** sur ϕ et sur la famille $\{\mathbb{P}_{\vartheta}, \vartheta \in \Theta\}$ pour obtenir la convergence et les performances asymptotiques de $\hat{\vartheta}_n$.

MAP 433 :
Introduction
aux méthodes
statistiques.
Cours 3

Rappel succinct
du Cours 2

Modélisation
statistique

Méthodes
d'estimation
pour le modèle
de densité

Méthode des
moments
Z-estimation
M-estimation
Principe de
maximum de
vraisemblance

⏪ ⏩ ⏴ ⏵ ⏶ ⏷ ⏸ ⏹ ⏺ ⏻ ⏼ ⏽ ⏾ ⏿ 🔍 ↻

Z-estimation : à quoi ça sert ?

- **Exemple.** $\Theta = \mathbb{R}$, $\mathbb{P}_{\vartheta}(dx) = f(x - \vartheta)dx$, et f symétrique : $f(-x) = f(x)$, $\forall x \in \mathbb{R}$.
- **Il n'y a pas de bornitude des moments !**
- On pose

$$\phi(a, x) = \text{Arctg}(x - a).$$

- La fonction

$$a \rightsquigarrow \mathbb{E}_{\vartheta} [\phi(a, X)] = \int_{\mathbb{R}} \text{Arctg}(x - a) f(x - \vartheta) dx$$

est strictement décroissante et s'annule seulement en $a = \vartheta$.

- **Z-estimateur associé** : solution $\hat{\vartheta}_n$ de

$$\sum_{i=1}^n \text{Arctg}(X_i - \hat{\vartheta}_n) = 0$$

(unicité).

⏪ ⏩ ⏴ ⏵ ⏶ ⏷ ⏸ ⏹ ⏺ ⏻ ⏼ ⏽ ⏾ ⏿ 🔍 ↻

MAP 433 :
Introduction
aux méthodes
statistiques.
Cours 3

Rappel succinct
du Cours 2

Modélisation
statistique

Méthodes
d'estimation
pour le modèle
de densité

Méthode des
moments
Z-estimation
M-estimation
Principe de
maximum de
vraisemblance

Le cas multidimensionnel

Si $\Theta \subset \mathbb{R}^d$ avec $d > 1$, la fonction ϕ est remplacée par

$$\Phi = (\phi_1, \dots, \phi_d) : \Theta \times \mathbb{R} \rightarrow \mathbb{R}^d.$$

Definition

On appelle Z-estimateur associé à Φ tout estimateur $\widehat{\vartheta}_n$ satisfaisant

$$\sum_{i=1}^n \phi_\ell(\widehat{\vartheta}_n, X_i) = 0, \quad \ell = 1, \dots, d.$$

MAP 433 :
Introduction
aux méthodes
statistiques.
Cours 3

Rappel succinct
du Cours 2

Modélisation
statistique

Méthodes
d'estimation
pour le modèle
de densité

Méthode des
moments

Z-estimation
M-estimation
Principe de
maximum de
vraisemblance



Z-estimation \rightarrow M-estimation

- En dimension 1 : si

$$\phi(\vartheta, x) = \partial_\vartheta \psi(\vartheta, x)$$

pour une certaine fonction ψ , résoudre $\sum_{i=1}^n \phi(\vartheta, X_i) = 0$ revient à **chercher un point critique** de

$$\vartheta \rightsquigarrow \sum_{i=1}^n \psi(\vartheta, X_i).$$

- En dimension $d \geq 1$, il faut $\phi(\vartheta, x) = \nabla_\vartheta \psi(\vartheta, x)$ (moins facile à obtenir).
- **Invite à généraliser** la recherche d'estimateurs via la maximisation d'un critère \rightarrow M-estimation.

MAP 433 :
Introduction
aux méthodes
statistiques.
Cours 3

Rappel succinct
du Cours 2

Modélisation
statistique

Méthodes
d'estimation
pour le modèle
de densité

Méthode des
moments

Z-estimation
M-estimation
Principe de
maximum de
vraisemblance



M-estimation

- Principe : Se donner une application $\psi : \Theta \times \mathbb{R} \rightarrow \mathbb{R}_+$ telle que, pour tout $\vartheta \in \Theta \subset \mathbb{R}^d$,

$$a \rightsquigarrow \mathbb{E}_\vartheta [\psi(a, X)] = \int \psi(a, x) \mathbb{P}_\vartheta(dx)$$

admet un **maximum en $a = \vartheta$** .

Définition

On appelle M-estimateur associé à ψ tout estimateur $\widehat{\vartheta}_n$ satisfaisant

$$\sum_{i=1}^n \psi(\widehat{\vartheta}_n, X_i) = \max_{a \in \Theta} \sum_{i=1}^n \psi(a, X_i).$$

- Il n'y a pas unicité de $\widehat{\vartheta}_n$ (à ce niveau).

MAP 433 :
Introduction
aux méthodes
statistiques.
Cours 3

Rappel succinct
du Cours 2

Modélisation
statistique

Méthodes
d'estimation
pour le modèle
de densité

Méthode des
moments

Z-estimation
M-estimation
Principe de
maximum de
vraisemblance



Un exemple classique : paramètre de localisation

- $\Theta = \mathbb{R}$, $\mathbb{P}_\vartheta(dx) = f(x - \vartheta)dx$, et $\int_{\mathbb{R}} xf(x)dx = 0$, $\int_{\mathbb{R}} x^2 \mathbb{P}_\vartheta(dx) < +\infty$ pour tout $\vartheta \in \mathbb{R}$. On pose

$$\psi(a, x) = -(a - x)^2$$

- La fonction

$$a \rightsquigarrow \mathbb{E}_\vartheta [\psi(a, X)] = - \int_{\mathbb{R}} (a - X)^2 f(x - \vartheta) dx$$

admet un **maximum en $a = \mathbb{E}_\vartheta [X] = \int_{\mathbb{R}} xf(x - \vartheta)dx = \vartheta$** .

- **M-estimateur associé** :

$$\sum_{i=1}^n (X_i - \widehat{\vartheta}_n)^2 = \min_{a \in \mathbb{R}} \sum_{i=1}^n (X_i - a)^2.$$

MAP 433 :
Introduction
aux méthodes
statistiques.
Cours 3

Rappel succinct
du Cours 2

Modélisation
statistique

Méthodes
d'estimation
pour le modèle
de densité

Méthode des
moments

Z-estimation
M-estimation
Principe de
maximum de
vraisemblance



Paramètre de localisation

- C'est **aussi** un Z-estimateur associé à $\phi(a, x) = 2(x - a)$: on résout

$$\sum_{i=1}^n (a - X_i) = 0 \text{ d'où } \hat{\vartheta}_n = \bar{X}_n.$$

- Dans cet **exemple très simple**, tous les points de vue coïncident.
- Si, dans le même contexte, $\int_{\mathbb{R}} x^2 \mathbb{P}_{\vartheta}(dx) = +\infty$ et $f(x) = f(-x)$, on peut utiliser Z-estimateur avec $\phi(a, x) = \text{Arctg}(x - a)$. Méthode robuste, mais est-elle optimale? Peut-on faire mieux **si f est connue?** A suivre...

MAP 433 :
Introduction aux méthodes statistiques.
Cours 3

Rappel succinct du Cours 2

Modélisation statistique

Méthodes d'estimation pour le modèle de densité

Méthode des moments

Z-estimation

M-estimation

Principe de maximum de vraisemblance

Lien entre Z- et M- estimateurs

- **Pas d'inclusion** entre ces deux classes d'estimateurs en général :

- Si ψ non-régulière, M-estimateur $\not\Rightarrow$ Z-estimateur
- Si une équation d'estimation admet plusieurs solutions distinctes, Z-estimateur $\not\Rightarrow$ M-estimateur (cas d'un extremum local).

- Toutefois, si ψ **est régulière**, les M-estimateurs **sont** des Z-estimateurs : si $\Theta \subset \mathbb{R}$ ($d = 1$), en posant

$$\phi(a, x) = \partial_a \psi(a, x),$$

on a

$$\sum_{i=1}^n \partial_a \psi(\vartheta, X_i) \Big|_{a=\hat{\vartheta}_n} = \sum_{i=1}^n \phi(\hat{\vartheta}_n, X_i) = 0.$$

MAP 433 :
Introduction aux méthodes statistiques.
Cours 3

Rappel succinct du Cours 2

Modélisation statistique

Méthodes d'estimation pour le modèle de densité

Méthode des moments

Z-estimation

M-estimation

Principe de maximum de vraisemblance

Maximum de vraisemblance

- Principe **fondamental** et **incontournable** en statistique. Cas particuliers connus depuis le XVIIIème siècle. Définition générale : Fisher (1922).
- Fournit une première **méthode systématique** de construction d'un M-estimateur (souvent un Z-estimateur, souvent aussi *a posteriori* un estimateur par substitution simple).
- Procédure **optimale** (dans quel sens?) sous des hypothèses de **régularité** de la famille $\{\mathbb{P}_{\vartheta}, \vartheta \in \Theta\}$ (Cours 6).
- Parfois difficile à mettre en oeuvre en pratique \rightarrow **méthodes numériques**, statistique computationnelle.

MAP 433 :
Introduction aux méthodes statistiques.
Cours 3

Rappel succinct du Cours 2

Modélisation statistique

Méthodes d'estimation pour le modèle de densité

Méthode des moments

Z-estimation

M-estimation

Principe de maximum de vraisemblance

Fonction de vraisemblance

- La famille $\{\mathbb{P}_{\vartheta}, \vartheta \in \Theta\}$ est dominée par une mesure σ -finie μ . On se donne, pour $\vartheta \in \Theta$

$$f(\vartheta, x) = \frac{d\mathbb{P}_{\vartheta}}{d\mu}(x), \quad x \in \mathbb{R}.$$

Définition

Fonction de vraisemblance du n -échantillon associée à la famille $\{f(\vartheta, \cdot), \vartheta \in \Theta\}$:

$$\vartheta \rightsquigarrow \mathcal{L}_n(\vartheta, X_1, \dots, X_n) = \prod_{i=1}^n f(\vartheta, X_i)$$

- C'est une fonction aléatoire (définie μ -presque partout).

MAP 433 :
Introduction aux méthodes statistiques.
Cours 3

Rappel succinct du Cours 2

Modélisation statistique

Méthodes d'estimation pour le modèle de densité

Méthode des moments

Z-estimation

M-estimation

Principe de maximum de vraisemblance

Exemples

- **Exemple 1 : Modèle de Poisson.** On observe

$$X_1, \dots, X_n \sim_{\text{i.i.d.}} \text{Poisson}(\vartheta),$$

$\vartheta \in \Theta = \mathbb{R}_+ \setminus \{0\}$ et prenons $\mu(dx) = \sum_{k \in \mathbb{N}} \delta_k(dx)$.

- La densité de \mathbb{P}_ϑ par rapport à μ est

$$f(\vartheta, x) = e^{-\vartheta} \frac{\vartheta^x}{x!}, \quad x = 0, 1, 2, \dots$$

- La **fonction de vraisemblance** associée s'écrit

$$\begin{aligned} \vartheta \rightsquigarrow \mathcal{L}_n(\vartheta, X_1, \dots, X_n) &= \prod_{i=1}^n e^{-\vartheta} \frac{\vartheta^{X_i}}{X_i!} \\ &= \frac{1}{\prod_{i=1}^n X_i!} e^{-n\vartheta} \vartheta^{\sum_{i=1}^n X_i}. \end{aligned}$$

◀ ▶ ⏪ ⏩ ⏴ ⏵ ⏶ ⏷ ⏸ ⏹ ⏺ ⏻ ⏼ ⏽ ⏾ ⏿ 🔍 ↻

MAP 433 :
Introduction
aux méthodes
statistiques.
Cours 3

Rappel succinct
du Cours 2

Modélisation
statistique

Méthodes
d'estimation
pour le modèle
de densité

Méthode des
moments
Z-estimation
M-estimation
Principe de
maximum de
vraisemblance

Exemples

- **Exemple 2 Modèle de Cauchy.** On observe

$$X_1, \dots, X_n \sim_{\text{i.i.d.}} \text{Cauchy},$$

$\vartheta \in \Theta = \mathbb{R}$ et $\mu(dx) = dx$ (**par exemple**).

- On a alors

$$\mathbb{P}_\vartheta(dx) = f(\vartheta, x)dx = \frac{1}{\pi(1 + (x - \vartheta)^2)} dx.$$

- La **fonction de vraisemblance** associée s'écrit

$$\vartheta \rightsquigarrow \mathcal{L}_n(\vartheta, X_1, \dots, X_n) = \frac{1}{\pi^n} \prod_{i=1}^n (1 + (X_i - \vartheta)^2)^{-1}.$$

◀ ▶ ⏪ ⏩ ⏴ ⏵ ⏶ ⏷ ⏸ ⏹ ⏺ ⏻ ⏼ ⏽ ⏾ ⏿ 🔍 ↻

MAP 433 :
Introduction
aux méthodes
statistiques.
Cours 3

Rappel succinct
du Cours 2

Modélisation
statistique

Méthodes
d'estimation
pour le modèle
de densité

Méthode des
moments
Z-estimation
M-estimation
Principe de
maximum de
vraisemblance

Principe de maximum de vraisemblance

- Cas d'une famille de lois **restreinte à deux points**

$$\Theta = \{\vartheta_1, \vartheta_2\} \subset \mathbb{R},$$

avec \mathbb{P}_{ϑ_i} discrète et $\mu(dx)$ la mesure de comptage.

- **A priori**, pour tout (x_1, \dots, x_n) , et pour $\vartheta \in \{\vartheta_1, \vartheta_2\}$,

$$\begin{aligned} \mathbb{P}_\vartheta [X_1 = x_1, \dots, X_n = x_n] &= \prod_{i=1}^n \mathbb{P}_\vartheta [X_i = x_i] \\ &= \prod_{i=1}^n f(\vartheta, x_i). \end{aligned}$$

La probabilité d'avoir la réalisation fixée (x_1, \dots, x_n) .

◀ ▶ ⏪ ⏩ ⏴ ⏵ ⏶ ⏷ ⏸ ⏹ ⏺ ⏻ ⏼ ⏽ ⏾ ⏿ 🔍 ↻

MAP 433 :
Introduction
aux méthodes
statistiques.
Cours 3

Rappel succinct
du Cours 2

Modélisation
statistique

Méthodes
d'estimation
pour le modèle
de densité

Méthode des
moments
Z-estimation
M-estimation
Principe de
maximum de
vraisemblance

Principe de maximum de vraisemblance

- **A posteriori**, on observe (X_1, \dots, X_n) . L'événement

$$\left\{ \prod_{i=1}^n f(\vartheta_1, X_i) > \prod_{i=1}^n f(\vartheta_2, X_i) \right\} \quad (\text{Cas 1})$$

ou bien l'événement

$$\left\{ \prod_{i=1}^n f(\vartheta_2, X_i) > \prod_{i=1}^n f(\vartheta_1, X_i) \right\} \quad (\text{Cas 2})$$

est réalisé. (On ignore le cas d'égalité.)

- **Principe de maximum de vraisemblance :**

$$\hat{\vartheta}_n^{\text{mv}} = \vartheta_1 1_{\{\text{Cas 1}\}} + \vartheta_2 1_{\{\text{Cas 2}\}}.$$

◀ ▶ ⏪ ⏩ ⏴ ⏵ ⏶ ⏷ ⏸ ⏹ ⏺ ⏻ ⏼ ⏽ ⏾ ⏿ 🔍 ↻

MAP 433 :
Introduction
aux méthodes
statistiques.
Cours 3

Rappel succinct
du Cours 2

Modélisation
statistique

Méthodes
d'estimation
pour le modèle
de densité

Méthode des
moments
Z-estimation
M-estimation
Principe de
maximum de
vraisemblance

Estimateur du maximum de vraisemblance

- On généralise le principe précédent pour une famille de lois et un ensemble de paramètres **quelconques**.
- **Situation** : $X_1, \dots, X_n \sim_{i.i.d.} \mathbb{P}_\vartheta$, $\{\mathbb{P}_\vartheta, \vartheta \in \Theta\}$ dominée, $\Theta \subset \mathbb{R}^d$, $\vartheta \rightsquigarrow \mathcal{L}_n(\vartheta, X_1, \dots, X_n)$ vraisemblance associée.

Définition

On appelle **estimateur du maximum de vraisemblance** tout estimateur $\hat{\vartheta}_n^{\text{mv}}$ satisfaisant

$$\mathcal{L}_n(\hat{\vartheta}_n^{\text{mv}}, X_1, \dots, X_n) = \max_{\vartheta \in \Theta} \mathcal{L}_n(\vartheta, X_1, \dots, X_n).$$

- **Existence, unicité...**

MAP 433 :
Introduction
aux méthodes
statistiques.
Cours 3

Rappel succinct
du Cours 2

Modélisation
statistique

Méthodes
d'estimation
pour le modèle
de densité

Méthode des
moments

Z-estimation

M-estimation

Principe de
maximum de
vraisemblance

◀ ▶ ⏪ ⏩ 🔍 🔄

Remarques

- Log-vraisemblance :

$$\begin{aligned} \vartheta \rightsquigarrow \ell_n(\vartheta, X_1, \dots, X_n) &= \log \mathcal{L}_n(\vartheta, X_1, \dots, X_n) \\ &= \sum_{i=1}^n \log f(\vartheta, X_i). \end{aligned}$$

Bien défini si $f(\vartheta, \cdot) > 0$ μ -pp.

Max. vraisemblance = max. log-vraisemblance.

- L'estimateur du maximum de vraisemblance **ne dépend pas** du choix de la mesure dominante μ .
- Notion de **racine de l'équation de vraisemblance** : tout estimateur $\hat{\vartheta}_n^{\text{rv}}$ vérifiant

$$\nabla_{\vartheta} \ell_n(\hat{\vartheta}_n^{\text{rv}}, X_1, \dots, X_n) = 0.$$

◀ ▶ ⏪ ⏩ 🔍 🔄

MAP 433 :
Introduction
aux méthodes
statistiques.
Cours 3

Rappel succinct
du Cours 2

Modélisation
statistique

Méthodes
d'estimation
pour le modèle
de densité

Méthode des
moments

Z-estimation

M-estimation

Principe de
maximum de
vraisemblance

Exemple : modèle normal

L'expérience statistique est engendrée par un n -échantillon de loi $\mathcal{N}(\mu, \sigma^2)$, le paramètre est $\vartheta = (\mu, \sigma^2) \in \Theta = \mathbb{R} \times \mathbb{R}_+ \setminus \{0\}$.

- **Vraisemblance**

$$\mathcal{L}_n((\mu, \sigma^2), X_1, \dots, X_n) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (X_i - \mu)^2\right).$$

- **Log-vraisemblance**

$$\ell_n((\mu, \sigma^2), X_1, \dots, X_n) = -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (X_i - \mu)^2.$$

MAP 433 :
Introduction
aux méthodes
statistiques.
Cours 3

Rappel succinct
du Cours 2

Modélisation
statistique

Méthodes
d'estimation
pour le modèle
de densité

Méthode des
moments

Z-estimation

M-estimation

Principe de
maximum de
vraisemblance

◀ ▶ ⏪ ⏩ 🔍 🔄

Exemple : modèle normal

Equation(s) de vraisemblance

$$\begin{cases} \partial_{\mu} \ell_n((\mu, \sigma^2), X_1, \dots, X_n) &= \frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \mu) \\ \partial_{\sigma^2} \ell_n((\mu, \sigma^2), X_1, \dots, X_n) &= -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (X_i - \mu)^2. \end{cases}$$

Solution de ces équations (pour $n \geq 2$) :

$$\hat{\vartheta}_n^{\text{rv}} = \left(\bar{X}_n, \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2 \right)$$

et on vérifie que $\hat{\vartheta}_n^{\text{rv}} = \hat{\vartheta}_n^{\text{mv}}$.

◀ ▶ ⏪ ⏩ 🔍 🔄

MAP 433 :
Introduction
aux méthodes
statistiques.
Cours 3

Rappel succinct
du Cours 2

Modélisation
statistique

Méthodes
d'estimation
pour le modèle
de densité

Méthode des
moments

Z-estimation

M-estimation

Principe de
maximum de
vraisemblance

Exemple : modèle de Poisson

Vraisemblance

$$\mathcal{L}_n(\vartheta, X_1, \dots, X_n) = \frac{1}{\prod_{i=1}^n X_i!} e^{-n\vartheta} \vartheta^{\sum_{i=1}^n X_i}.$$

Log-vraisemblance

$$\ell_n(\vartheta, X_1, \dots, X_n) = c(X_1, \dots, X_n) - n\vartheta + \sum_{i=1}^n X_i \log \vartheta.$$

Equation de vraisemblance

$$-n + \sum_{i=1}^n X_i \frac{1}{\vartheta} = 0, \text{ soit } \hat{\vartheta}_n^{\text{rv}} = \frac{1}{n} \sum_{i=1}^n X_i = \bar{X}_n$$

et on vérifie que $\hat{\vartheta}_n^{\text{rv}} = \hat{\vartheta}_n^{\text{mv}}$.

MAP 433 :
Introduction
aux méthodes
statistiques.
Cours 3

Rappel succinct
du Cours 2

Modélisation
statistique

Méthodes
d'estimation
pour le modèle
de densité

Méthode des
moments
Z-estimation
M-estimation
Principe de
maximum de
vraisemblance

◀ ▶ ⏪ ⏩ ⏴ ⏵ ⏶ ⏷ ⏸ ⏹ ⏺ ⏻ ⏼ ⏽ ⏾ ⏿ 🔍 🔄

Exemple : modèle de Laplace

L'expérience statistique est engendrée par un n -échantillon de loi de Laplace de paramètre $\vartheta \in \Theta = \mathbb{R}$. La densité par rapport à la mesure de Lebesgue :

$$f(\vartheta, x) = \frac{1}{2\sigma} \exp\left(-\frac{|x - \vartheta|}{\sigma}\right),$$

où $\sigma > 0$ est **connu**.

Vraisemblance

$$\mathcal{L}_n(\vartheta, X_1, \dots, X_n) = (2\sigma)^{-n} \exp\left(-\frac{1}{\sigma} \sum_{i=1}^n |X_i - \vartheta|\right)$$

Log-vraisemblance

$$\ell_n(\vartheta, X_1, \dots, X_n) = -n \log(2\sigma) - \frac{1}{\sigma} \sum_{i=1}^n |X_i - \vartheta|.$$

MAP 433 :
Introduction
aux méthodes
statistiques.
Cours 3

Rappel succinct
du Cours 2

Modélisation
statistique

Méthodes
d'estimation
pour le modèle
de densité

Méthode des
moments
Z-estimation
M-estimation
Principe de
maximum de
vraisemblance

◀ ▶ ⏪ ⏩ ⏴ ⏵ ⏶ ⏷ ⏸ ⏹ ⏺ ⏻ ⏼ ⏽ ⏾ ⏿ 🔍 🔄

Exemple : modèle de Laplace

Maximiser $\mathcal{L}_n(\vartheta, X_1, \dots, X_n)$ revient à minimiser la fonction $\vartheta \rightsquigarrow \sum_{i=1}^n |X_i - \vartheta|$, dérivable presque partout de dérivée constante par morceaux. **Equation de vraisemblance :**

$$\sum_{i=1}^n \text{sign}(X_i - \vartheta) = 0.$$

Soit $X_{(1)} \leq \dots \leq X_{(n)}$ la statistique d'ordre.

- n pair : $\hat{\vartheta}_n^{\text{mv}}$ **n'est pas unique** ; tout point de l'intervalle $[X_{(\frac{n}{2})}, X_{(\frac{n}{2}+1)}]$ est un EMV.
- n impair : $\hat{\vartheta}_n^{\text{mv}} = X_{(\frac{n+1}{2})}$, l'EMV est unique. Mais $\hat{\vartheta}_n^{\text{rv}}$ n'existe pas.
- pour tout** n , la médiane empirique est un EMV.

MAP 433 :
Introduction
aux méthodes
statistiques.
Cours 3

Rappel succinct
du Cours 2

Modélisation
statistique

Méthodes
d'estimation
pour le modèle
de densité

Méthode des
moments
Z-estimation
M-estimation
Principe de
maximum de
vraisemblance

◀ ▶ ⏪ ⏩ ⏴ ⏵ ⏶ ⏷ ⏸ ⏹ ⏺ ⏻ ⏼ ⏽ ⏾ ⏿ 🔍 🔄

Exemple : modèle de Cauchy

Vraisemblance

$$\mathcal{L}_n(\vartheta, X_1, \dots, X_n) = \pi^{-n} \prod_{i=1}^n \frac{1}{1 + (X_i - \vartheta)^2}.$$

Log-vraisemblance

$$\ell_n(\vartheta, X_1, \dots, X_n) = -n \log \pi - \sum_{i=1}^n \log(1 + (X_i - \vartheta)^2)$$

Equation de vraisemblance

$$\sum_{i=1}^n \frac{X_i - \vartheta}{1 + (X_i - \vartheta)^2} = 0$$

pas de solution explicite et admet en général plusieurs solutions.

MAP 433 :
Introduction
aux méthodes
statistiques.
Cours 3

Rappel succinct
du Cours 2

Modélisation
statistique

Méthodes
d'estimation
pour le modèle
de densité

Méthode des
moments
Z-estimation
M-estimation
Principe de
maximum de
vraisemblance

◀ ▶ ⏪ ⏩ ⏴ ⏵ ⏶ ⏷ ⏸ ⏹ ⏺ ⏻ ⏼ ⏽ ⏾ ⏿ 🔍 🔄

Maximum de vraisemblance = M -estimateur

- Une inégalité de convexité : μ mesure σ -finie sur \mathbb{R} ; f, g deux **densités de probabilités** par rapport à μ . Alors

$$\int_{\mathbb{R}} f(x) \log f(x) \mu(dx) \geq \int_{\mathbb{R}} f(x) \log g(x) \mu(dx)$$

(si les intégrales sont finies) avec égalité **ssi** $f = g$ μ -pp.

- Preuve : à montrer

$$\int_{\mathbb{R}} f(x) \log \frac{g(x)}{f(x)} \mu(dx) \leq 0.$$

(avec une convention de notation appropriée)

MAP 433 :
Introduction
aux méthodes
statistiques.
Cours 3

Rappel succinct
du Cours 2

Modélisation
statistique

Méthodes
d'estimation
pour le modèle
de densité

Méthode des
moments
Z-estimation
M-estimation
Principe de
maximum de
vraisemblance



Une inégalité de convexité

- On a $\log(1+x) \leq x$ pour $x \geq -1$ avec égalité ssi $x = 0$.
- Donc

$$\log \frac{g(x)}{f(x)} = \log \left(1 + \left(\frac{g(x)}{f(x)} - 1 \right) \right) \leq \frac{g(x)}{f(x)} - 1$$

(avec égalité ssi $f(x) = g(x)$).

- Finalement

$$\begin{aligned} \int_{\mathbb{R}} f(x) \log \frac{g(x)}{f(x)} \mu(dx) &\leq \int_{\mathbb{R}} f(x) \left(\frac{g(x)}{f(x)} - 1 \right) \mu(dx) \\ &= \int_{\mathbb{R}} g(x) \mu(dx) - \int_{\mathbb{R}} f(x) \mu(dx) \\ &= 0. \end{aligned}$$

MAP 433 :
Introduction
aux méthodes
statistiques.
Cours 3

Rappel succinct
du Cours 2

Modélisation
statistique

Méthodes
d'estimation
pour le modèle
de densité

Méthode des
moments
Z-estimation
M-estimation
Principe de
maximum de
vraisemblance



Conséquence pour l'EMV

- On pose

$$\psi(a, x) := \log f(a, x), \quad a \in \Theta, x \in \mathbb{R}$$

(avec une convention pour le cas où on n'a pas $f(a, \cdot) > 0$.)

- La fonction

$$a \rightsquigarrow \mathbb{E}_{\vartheta} [\psi(a, X)] = \int_{\mathbb{R}} \log f(a, x) f(\vartheta, x) \mu(dx)$$

a un maximum en $a = \vartheta$ d'après **l'inégalité de convexité**.

MAP 433 :
Introduction
aux méthodes
statistiques.
Cours 3

Rappel succinct
du Cours 2

Modélisation
statistique

Méthodes
d'estimation
pour le modèle
de densité

Méthode des
moments
Z-estimation
M-estimation
Principe de
maximum de
vraisemblance



- Le M -estimateur associé à ψ maximise la fonction

$$a \rightsquigarrow \sum_{i=1}^n \log f(a, X_i) = \ell_n(a, X_1, \dots, X_n)$$

c'est-à-dire la **log-vraisemblance**. C'est **l'estimateur du maximum de vraisemblance**.

- C'est aussi un Z-estimateur si la fonction $\vartheta \rightsquigarrow \log f(\vartheta, \cdot)$ est régulière, associé à la fonction

$$\phi(\vartheta, x) = \partial_{\vartheta} \log f(\vartheta, x) = \frac{\partial_{\vartheta} f(\vartheta, x)}{f(\vartheta, x)}, \quad \vartheta \in \Theta, x \in \mathbb{R}$$

lorsque $\Theta \subset \mathbb{R}$, à condition que le maximum de log-vraisemblance n'est pas atteint sur la frontière de Θ . (Se généralise en dimension d .)

MAP 433 :
Introduction
aux méthodes
statistiques.
Cours 3

Rappel succinct
du Cours 2

Modélisation
statistique

Méthodes
d'estimation
pour le modèle
de densité

Méthode des
moments
Z-estimation
M-estimation
Principe de
maximum de
vraisemblance



Choix de modèle statistique

- Le statisticien a le choix de la famille $\{\mathbb{P}_\vartheta, \vartheta \in \Theta\}$. L'EMV dépend de ce choix.

- Exemple : on a l'échantillon ($n = 10$) :

$0.92, -0.20, -1.80, 0.02, 0.49, 1.41, -1.59, -1.29, 0.34, 100.$
tirage de $\mathcal{N}(0,1)$

- On prend $\mathbb{P}_\vartheta(dx) = f(x - \vartheta)dx$ pour deux f différents :
- f densité de la loi normale $\Rightarrow \hat{\vartheta}_n^{\text{mv}} = \bar{X}_n = 9.83.$
- f densité de loi de Laplace \Rightarrow tout point de l'intervalle $[0.02, 0.34]$ est un $\hat{\vartheta}_n^{\text{mv}}$, en particulier, la médiane :

$$\hat{\vartheta}_n^{\text{mv}} = M_n = (0.02 + 0.34)/2 = 0.18.$$

- Autre choix de modèle...

MAP 433 :
Introduction
aux méthodes
statistiques.
Cours 3

Rappel succinct
du Cours 2

Modélisation
statistique

Méthodes
d'estimation
pour le modèle
de densité

Méthode des
moments

Z-estimation

M-estimation
Principe de
maximum de
vraisemblance



Un M -estimateur qui n'est pas un Z -estimateur

- On observe $X_1, \dots, X_n \sim_{\text{i.i.d.}}$ uniformes sur $[0, \vartheta]$,
 $\vartheta \in \Theta = \mathbb{R}_+ \setminus \{0\}$.

- On a

$$\mathbb{P}_\vartheta(dx) = \vartheta^{-1} 1_{[0, \vartheta]}(x) dx$$

et

$$\begin{aligned} \mathcal{L}_n(\vartheta, X_1, \dots, X_n) &= \vartheta^{-n} \prod_{i=1}^n 1_{[0, \vartheta]}(X_i) \\ &= \vartheta^{-n} 1_{\{\max_{1 \leq i \leq n} X_i \leq \vartheta\}} \end{aligned}$$

- La fonction de vraisemblance n'est pas régulière.
- L'estimateur du maximum de vraisemblance est $\hat{\vartheta}_n^{\text{mv}} = \max_{1 \leq i \leq n} X_i$.

MAP 433 :
Introduction
aux méthodes
statistiques.
Cours 3

Rappel succinct
du Cours 2

Modélisation
statistique

Méthodes
d'estimation
pour le modèle
de densité

Méthode des
moments

Z-estimation

M-estimation
Principe de
maximum de
vraisemblance

