

Statistiques mathématiques : cours 4

Guillaume Lécué

14 septembre 2016

Rappels du cours 3 : Z et M -estimation, EMV

Z-estimation

- ▶ Situation : on observe $X_1, \dots, X_n \stackrel{i.i.d.}{\sim} \mathbb{P}_\theta$ sur \mathbb{R} et $\theta \in \Theta$
- ▶ Principe : Trouver une application $\phi : \Theta \times \mathbb{R} \rightarrow \mathbb{R}_+$ telle que, pour tout $\theta \in \Theta \subset \mathbb{R}^d$,

$$a \mapsto \mathbb{E}_\theta [\phi(a, X)] = \int \phi(a, x) \mathbb{P}_\theta(dx)$$

admet un zéro en $a = \theta$

Définition

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

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

M-estimation

- ▶ Situation : on observe $X_1, \dots, X_n \stackrel{i.i.d.}{\sim} \mathbb{P}_\theta$ sur \mathbb{R} et $\theta \in \Theta$.
- ▶ Principe : Trouver une application $\psi : \Theta \times \mathbb{R} \rightarrow \mathbb{R}_+$ telle que, pour tout $\theta \in \Theta \subset \mathbb{R}^d$,

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

admet un maximum en $a = \theta$

Définition

On appelle **M-estimateur** associé à ψ tout estimateur $\hat{\theta}_n \in \Theta$ satisfaisant

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

Fonction de vraisemblance

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

$$f(\theta, x) = \frac{d\mathbb{P}_\theta}{d\mu}(x), \quad x \in \mathcal{X}$$

Définition

La *fonction de vraisemblance* du n -échantillon (X_1, \dots, X_n) associée à la famille de densité $\{f(\theta, \cdot) : \theta \in \Theta\}$ est

$$\theta \in \Theta \mapsto \mathcal{L}_n(\theta, X_1, \dots, X_n) = \prod_{i=1}^n f(\theta, X_i)$$

Estimateur du maximum de vraisemblance

Situation :

- ▶ $X_1, \dots, X_n \stackrel{i.i.d.}{\sim} \mathbb{P}_\theta, \{\mathbb{P}_\theta, \theta \in \Theta\}$ dominée, $\Theta \subset \mathbb{R}^d$
- ▶ $\theta \mapsto \mathcal{L}_n(\theta, X_1, \dots, X_n)$ vraisemblance associée

Définition

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

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

Aujourd'hui

l'EMV est un M -estimateur

Asymptotique des Z - et M - estimateurs et de l'EMV

Modèles réguliers et information de Fisher

Construction de l'information de Fisher

Cadre général et interprétation géométrique

Exemples, applications

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

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

(convention $\log(0) = -\infty$)

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

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

Conséquence pour l'EMV

- ▶ On pose

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

(convention $\psi(a, x) = -\infty$ quand $f(a, x) = 0$)

- ▶ La fonction

$$a \in \Theta \mapsto \mathbb{E}_\theta [\psi(a, X)] = \int_{\mathbb{R}} \log f(a, x) f(\theta, x) \mu(dx)$$

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

- ▶ Le M -estimateur associé à ψ maximise la fonction

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

c'est-à-dire la **log-vraisemblance**, donc

**l'estimateur du maximum de vraisemblance est un
 M -estimateur**

- ▶ Si la fonction $\theta \mapsto \log f(\theta, \cdot)$ est régulière, l'EMV est aussi un Z -estimateur associé à la fonction

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

à condition que le maximum de la log-vraisemblance n'est pas atteint sur la frontière de Θ .

Asymptotique des Z et M -estimateurs et de l'EMV

Propriétés statistiques asymptotiques des estimateurs : définitions

Modèle d'échantillonnage : $(X_n)_n$ suite $\overset{i.i.d.}{\sim} \mathbb{P}_\theta \in \{\mathbb{P}_\theta : \theta \in \Theta\}$. Soit $(\hat{\theta}_n)_n$ un estimateur. On dit que :

1. $\hat{\theta}_n$ est **consistant** quand pour tout $\theta \in \Theta$, sous \mathbb{P}_θ ,

$$\hat{\theta}_n \xrightarrow{\mathbb{P}_\theta} \theta$$

($\hat{\theta}_n$ est **fortement consistant** quand la cv est p.s.)

2. $\hat{\theta}_n$ est **asymptotiquement normal** quand, pour tout $\theta \in \Theta$, sous \mathbb{P}_θ , il existe une suite croissante de réels positifs $(a_n) \uparrow \infty$ et V_θ une variable aléatoire telles que :

$$a_n(\hat{\theta}_n - \theta) \xrightarrow{d} V_\theta$$

Quand $V_\theta \sim \mathcal{N}(0, v(\theta))$, $v(\theta)$ est appelée la **variance asymptotique** ;
 $1/a_n$ est la vitesse de convergence asymptotique (généralement,
 $a_n = \sqrt{n}$)

Asymptotique des Z - et M -estimateurs

- ▶ Problème général délicat. Dans ce cours : **conditions suffisantes**
- ▶ résultats établis pour $\Theta \subset \mathbb{R}$ mais généralisable à \mathbb{R}^d
- ▶ application directe à l'EMV (vu comme un M -estimateur)

Consistance des Z-estimateurs

- ▶ Situation : on observe $X_1, \dots, X_n \stackrel{i.i.d.}{\sim} \mathbb{P}_\theta, \theta \in \Theta$.
- ▶ $\phi : \Theta \times \mathbb{R} \rightarrow \mathbb{R}$
- ▶ **Loi des grands nombres** : pour tout $\theta, a \in \Theta$, sous \mathbb{P}_θ ,

$$Z_n(a) = \frac{1}{n} \sum_{i=1}^n \phi(a, X_i) \xrightarrow{\mathbb{P}_\theta} Z(a, \theta) = \mathbb{E}_\theta [\phi(a, X)]$$

qui s'annule en $a = \theta$

- ▶ à montrer : pour tout $\theta \in \Theta$, sous \mathbb{P}_θ ,

$$\hat{\theta}_n \text{ (zéro de } Z_n(\cdot)) \xrightarrow{\mathbb{P}_\theta} \theta \text{ (zéro de } Z(\cdot, \theta))$$

Consistance des Z-estimateurs

Proposition

On suppose que :

a) $\sup_{a \in \Theta} |Z_n(a) - Z(a, \theta)| \xrightarrow{\mathbb{P}_\theta} 0,$

b) $\forall \varepsilon > 0, \inf_{|a - \theta| \geq \varepsilon} |Z(a, \theta)| > 0$

alors tout Z-estimateur $\hat{\theta}_n$ (càd tq $Z_n(\hat{\theta}_n) = 0$) est consistant.

1. "a)" se montre grâce aux techniques de processus empiriques
2. "b)" est une condition (déterministe) sur le zéro θ de $Z(\cdot, \theta)$

Consistance des M -estimateurs

- ▶ Situation : on observe $X_1, \dots, X_n \stackrel{i.i.d.}{\sim} \mathbb{P}_\theta, \theta \in \Theta$
- ▶ $\psi : \Theta \times \mathbb{R} \rightarrow \mathbb{R}$ fonction de contraste
- ▶ **Loi des grands nombres** : pour tout $\theta, a \in \Theta$, sous \mathbb{P}_θ ,

$$M_n(a) = \frac{1}{n} \sum_{i=1}^n \psi(a, X_i) \xrightarrow{\mathbb{P}_\theta} M(a, \theta) = \mathbb{E}_\theta [\psi(a, X)]$$

qui atteint son maximum en $a = \theta$

- ▶ à montrer : pour tout $\theta \in \Theta$,

$$\hat{\theta}_n = \arg \max_{a \in \Theta} M_n(a) \xrightarrow{\mathbb{P}_\theta} \arg \max_{a \in \Theta} \mathbb{E}_\theta [\psi(a, X)] = \theta$$

Consistance des M -estimateurs

Proposition

On suppose que :

- a) $\sup_{a \in \Theta} |M_n(a) - M(a, \theta)| \xrightarrow{\mathbb{P}_\theta} 0,$
- b) $\forall \varepsilon > 0, \sup_{|a - \theta| \geq \varepsilon} M(a, \theta) < M(\theta, \theta)$

alors tout M -estimateur $\hat{\theta}_n \in \arg \max_{a \in \Theta} M_n(a)$ est consistant.

1. "a)" se montre grâce aux techniques de processus empiriques
2. "b)" est une condition sur le maximum de la fonction de contraste

Normalité asymptotique des Z-estimateurs : principe

- ▶ Situation : $X_1, \dots, X_n \stackrel{i.i.d.}{\sim} \mathbb{P}_\theta$ pour $\theta \in \Theta \subset \mathbb{R}$
- ▶ $\hat{\theta}_n$: Z-estimateur associé à $\phi : \Theta \times \mathbb{R} \rightarrow \mathbb{R}$ càd

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

- ▶ Sous certaines conditions $\hat{\theta}_n$ est consistant. Maintenant : **est-il asymptotiquement normal ?** :
 1. Pour quelle vitesse de convergence ?
 2. Pour quelle variance asymptotique ?

- ▶ Principe. Développement de Taylor autour de θ :

$$0 = Z_n(\hat{\theta}_n) = Z_n(\theta) + (\hat{\theta}_n - \theta)Z'_n(\theta) + \frac{1}{2}(\hat{\theta}_n - \theta)^2 Z''(\tilde{\theta}_n)$$

- ▶ On a approximativement (en "négligeant" le reste) :

$$\boxed{\sqrt{n}(\hat{\theta}_n - \theta) \approx \frac{-\sqrt{n}Z_n(\theta)}{Z'_n(\theta)}}$$

Normalité asymptotique des Z-estimateurs : principe

- ▶ Convergence du **numérateur** : par le TCL (sous \mathbb{P}_θ)

$$\sqrt{n}Z_n(\theta) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \phi(\theta, X_i) \xrightarrow{d} \mathcal{N}(0, \mathbb{E}_\theta [\phi(\theta, X)^2])$$

si $\mathbb{E}_\theta [\phi(\theta, X)] = 0$ et $\mathbb{E}_\theta [\phi(\theta, X)^2] < +\infty$.

- ▶ Convergence du **dénominateur** (sous \mathbb{P}_θ)

$$Z'_n(\theta) = \frac{1}{n} \sum_{i=1}^n \partial_1 \phi(\theta, X_i) \xrightarrow{\mathbb{P}_\theta} \mathbb{E}_\theta [\partial_1 \phi(\theta, X)]$$

$\neq 0$ (à supposer).

- ▶ + hypothèses techniques pour **contrôler le reste** (besoin de la consistence de $\hat{\theta}_n$).

Normalité asymptotique des Z -estimateurs

Theorem

Soit Θ un ouvert de \mathbb{R} . On suppose que :

- ▶ le Z -estimateur $\hat{\theta}_n$ associé à ϕ est consistant
- ▶ pour tout $\theta \in \Theta$, $\mathbb{E}_\theta [\phi(\theta, X)] = 0$,

$$\mathbb{E}_\theta [\phi(\theta, X)^2] < +\infty \text{ et } \mathbb{E}_\theta [\partial_1 \phi(\theta, X)] \neq 0$$

- ▶ (*Contrôle reste*) pour tout $\theta \in \Theta$, pour tout a dans un voisinage de θ ,

$$|\partial_1^2 \phi(a, x)| \leq g(x) \text{ où } \mathbb{E}_\theta [g(X)] < +\infty.$$

Alors, pour tout $\theta \in \Theta$, sous \mathbb{P}_θ ,

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} \mathcal{N}\left(0, \frac{\mathbb{E}_\theta[\phi(\theta, X)^2]}{(\mathbb{E}_\theta[\partial_1 \phi(\theta, X)])^2}\right)$$

Normalité asymptotique des M -estimateurs : principe

- ▶ Situation : $X_1, \dots, X_n \stackrel{i.i.d.}{\sim} \mathbb{P}_\theta$ pour $\theta \in \Theta \subset \mathbb{R}$
- ▶ $\hat{\theta}_n$: M -estimateur associé à $\psi : \Theta \times \mathbb{R} \rightarrow \mathbb{R}$ càd

$$\hat{\theta}_n \in \arg \max_{a \in \Theta} \sum_{i=1}^n \psi(a, X_i)$$

- ▶ Sous certaines conditions, les M -estimateurs sont des Z -estimateurs associés à

$$\phi(a, x) = \partial_1 \psi(a, x)$$

On applique le théorème de la normalité asymptotique des Z -estimateurs aux M -estimateurs.

Normalité asymptotique des M -estimateurs

Theorem

Soit Θ un ouvert de \mathbb{R} . On suppose que :

- ▶ le M -estimateur $\hat{\theta}_n$ associé à ψ est consistant
- ▶ pour tout $\theta \in \Theta$, $\mathbb{E}_\theta [\partial_1 \psi(\theta, X)] = 0$,

$$\mathbb{E}_\theta [\partial_1 \psi(\theta, X)^2] < +\infty \text{ et } \mathbb{E}_\theta [\partial_1^2 \psi(\theta, X)] \neq 0$$

- ▶ pour tout $\theta \in \Theta$, pour tout a dans un voisinage de θ ,

$$|\partial_1^3 \psi(a, x)| \leq g(x) \text{ où } \mathbb{E}_\theta [g(X)] < +\infty.$$

Alors, pour tout $\theta \in \Theta$, sous \mathbb{P}_θ ,

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} \mathcal{N}\left(0, \frac{\mathbb{E}_\theta[\partial_1 \psi(\theta, X)^2]}{(\mathbb{E}_\theta[\partial_1^2 \psi(\theta, X)])^2}\right)$$

Normalité asymptotique de l'EMV : principe

- ▶ Situation : $X_1, \dots, X_n \stackrel{i.i.d.}{\sim} \mathbb{P}_\theta$ pour $\theta \in \Theta \subset \mathbb{R}$
- ▶ $\{\mathbb{P}_\theta : \theta \in \Theta\}$ est dominé par μ et la vraisemblance associée est

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

- ▶ l'EMV est

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

Normalité asymptotique de l'EMV : principe

- ▶ On pose

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

(convention $\log 0 = -\infty$)

- ▶ La fonction

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

a un maximum en $a = \theta$.

Donc l'EMV est un M -estimateur. On peut alors appliquer le théorème de normalité asymptotique des M -estimateurs.

Normalité asymptotique de l'EMV : principe

Sous certaines conditions, on a

$$\sqrt{n}(\hat{\theta}_n^{\text{mv}} - \theta) \xrightarrow{d} \mathcal{N}(0, \nu(\theta))$$

où la variance asymptotique est

$$\nu(\theta) = \frac{\mathbb{E}_\theta[\partial_1 \psi(\theta, X)^2]}{(\mathbb{E}_\theta[\partial_1^2 \psi(\theta, X)])^2}$$

qui se simplifie quand $\psi(a, x) = \log f(a, x)$ en

$$\nu(\theta) = \frac{1}{\mathbb{I}(\theta)} \text{ où } \mathbb{I}(\theta) = \mathbb{E}_\theta [(\partial_1 \log f(\theta, X))^2]$$

car

$$\mathbb{E}_\theta [(\partial_1^2 f(\theta, X))/f(\theta, X)] = 0$$

Normalité asymptotique de l'EMV : principe

Dans un modèle d'échantillonnage $\{\mathbb{P}_\theta : \theta \in \Theta\}$ dominé par μ , de densités

$$\frac{d\mathbb{P}_\theta}{d\mu}(x) = f(\theta, x)$$

on définit

$$\ell(\theta, x) = \log f(\theta, x)$$

avec la convention ($\log 0 = -\infty$) et quand $\ell(\cdot, x)$ est dérivable pour μ -p.t. x , on appelle :

1. **fonction de score** : à $x \in \mathbb{R}$ fixé,

$$\theta \mapsto \partial_1 \ell(\theta, x)$$

2. on appelle **information de Fisher** en $\theta \in \Theta$,

$$\mathbb{I}(\theta) = \mathbb{E}_\theta [(\partial_1 \ell(\theta, X))^2]$$

Normalité asymptotique de l'EMV : principe

On a donc "sous certaines hypothèses" que :

$$\sqrt{n}(\hat{\theta}_n^{\text{mv}} - \theta) \xrightarrow{d} \mathcal{N}\left(0, \frac{1}{\mathbb{I}(\theta)}\right)$$

Les hypothèses suffisantes pour assurer la **normalité asymptotique de l'EMV de variance asymptotique** $\mathbb{I}(\theta)^{-1}$ sont à l'origine de la définition d'un **modèle régulier**.

Conclusion : L'étude de la normalité asymptotique des EMV nous amène à introduire les notions suivantes :

1. le **score**
2. l'**information de Fisher**
3. un **modèle régulier**

Régularité d'un modèle statistique et information

- ▶ Cadre simplificateur : modèle dominé (par μ)

$$X_1, \dots, X_n \stackrel{i.i.d.}{\sim} \mathbb{P}_\theta$$

dans la famille $\{\mathbb{P}_\theta, \theta \in \Theta\}$ avec $\Theta \subset \mathbb{R}$ (pour simplifier).

- ▶ Notation :

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

- ▶ Hypothèse : la quantité

$$\mathbb{I}(\theta) = \mathbb{E}_\theta \left[(\partial_1 \log f(\theta, X))^2 \right]$$

est bien définie \mathbb{P}_θ -p.s..

Information de Fisher

Définition

$\mathbb{I}(\theta) = \mathbb{E}_\theta [(\partial_1 \log f(\theta, X))^2]$ s'appelle *l'information de Fisher* de la famille $\{\mathbb{P}_\theta, \theta \in \Theta\}$ au point θ .

- ▶ L'information de Fisher ne dépend pas de la mesure dominante μ
- ▶ Le cadre d'intérêt est celui où

$$0 < \mathbb{I}(\theta) < +\infty.$$

- ▶ $\mathbb{I}(\theta)$ quantifie « l'information » qu'apporte chaque observation X_i sur le paramètre θ .
- ▶ on a $\mathbb{P}_\theta [f(\theta, X) > 0] = 1$, donc la quantité $\log f(\theta, X)$ est bien définie.

Modèle régulier

Définition

La famille de densités $\{f(\theta, \cdot), \theta \in \Theta\}$, par rapport à la mesure dominante μ , $\Theta \subset \mathbb{R}$ ouvert, est **régulière** si

- ▶ $\{f(\theta, \cdot) > 0\} = \{f(\theta', \cdot) > 0\}$, μ -p.p. $\forall \theta, \theta' \in \Theta$
- ▶ μ -p.p. $\theta \mapsto f(\theta, \cdot)$, $\theta \mapsto \log f(\theta, \cdot)$ sont \mathcal{C}^2
- ▶ $\forall \theta \in \Theta, \exists \mathcal{V}_\theta \subset \Theta$ t.q. pour $a \in \mathcal{V}_\theta$

$$|\partial_a^2 \log f(a, x)| + |\partial_a \log f(a, x)| + (\partial_a \log f(a, x))^2 \leq g(x)$$

$$\text{où } \int_{\mathbb{R}} g(x) \sup_{a \in \mathcal{V}(\theta)} f(a, x) \mu(dx) < +\infty$$

- ▶ L'information de Fisher est non-dégénérée :

$$\forall \theta \in \Theta, \mathbb{I}(\theta) > 0$$

Résultat principal

Theorem

*Dans le modèle d'échantillonnage associé à un modèle régulier on a :
pour tout $\theta \in \Theta$,*

$$\sqrt{n}(\hat{\theta}_n^{\text{mv}} - \theta) \xrightarrow{d} \mathcal{N}\left(0, \frac{1}{\mathbb{I}(\theta)}\right)$$

Information de Fisher dans le modèle d'échantillonnage

Dans le modèle d'échantillonnage (sur \mathbb{R}) dominé (par μ), on observe

$$X_1, \dots, X_n \stackrel{i.i.d.}{\sim} \mathbb{P}_\theta, \theta \in \Theta \subset \mathbb{R}$$

et on note les densités (pour tout $\theta \in \Theta$, $x \in \mathbb{R}$)

$$\frac{d\mathbb{P}_\theta}{d\mu}(x) = f(\theta, x)$$

L'expérience statistique associée est

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

qui est dominée par $\mu^{\otimes n}$, de densité : $\forall z = (x_1, \dots, x_n) \in \mathbb{R}^n$,

$$f_n(\theta, z) = \frac{d\mathbb{P}^Z}{d\mu^{\otimes n}}(z) = \prod_{i=1}^n f(\theta, x_i)$$

où $Z = (X_1, \dots, X_n)$ est une observation de l'expérience \mathcal{E}^n .

Information de Fisher dans le modèle d'échantillonnage

L'information de Fisher contenue dans une observation $Z = (X_1, \dots, X_n)$ de \mathcal{E}^n en θ est

$$\mathbb{I}_n(\theta) = \mathbb{I}(\theta | \mathcal{E}^n) = \mathbb{E}_\theta [(\partial_1 \log f_n(\theta, Z))^2]$$

Par ailleurs, pour une seule observation X_1 de l'expérience

$$\mathcal{E} = (\mathbb{R}, \mathcal{B}(\mathbb{R}), \{\mathbb{P}_\theta, \theta \in \Theta\}),$$

l'information de Fisher est

$$\mathbb{I}(\theta) = \mathbb{I}(\theta | \mathcal{E}) = \mathbb{E}_\theta [(\partial_1 \log f(\theta, X_1))^2]$$

Theorem

$$\mathbb{I}(\theta | \mathcal{E}^n) = n\mathbb{I}(\theta | \mathcal{E})$$

Le cas multidimensionnel

Définition

Soit $Z \sim \mathbb{P}_\theta$ avec $\theta \in \Theta \subset \mathbb{R}^d$ de densité $f(\theta, \cdot)$. La **matrice d'information de Fisher** en θ est

$$\mathbb{I}(\theta) = \mathbb{E}_\theta [\nabla_\theta \log f(\theta, Z) \nabla_\theta \log f(\theta, Z)^T]$$

- ▶ $\mathbb{I}(\theta)$ est une **matrice $d \times d$ symétrique positive**
- ▶ Dans le modèle d'échantillonnage dominé, on note $\mathbb{I}(\theta)$ l'information de Fisher pour une observation X_1 , sous des hypothèses de régularité du modèle, on a

$$\sqrt{n}(\hat{\theta}_n^{\text{mv}} - \theta) \xrightarrow{d} \mathcal{N}(0, \mathbb{I}(\theta)^{-1})$$

Région de confiance asymptotique pour l'EMV (1/2)

Dans le modèle d'échantillonnage dominé régulier tel que l'information de Fisher (pour une observation X_1)

$$\theta \in \Theta \subset \mathbb{R}^d \mapsto \mathbb{I}(\theta) \in \mathbb{R}^{d \times d}$$

est continue, on peut appliquer le lemme de Slutsky :

$$\sqrt{n}\mathbb{I}(\hat{\theta}_n^{\text{mv}})^{1/2}(\hat{\theta}_n^{\text{mv}} - \theta) \xrightarrow{d} \mathcal{N}(0, I_d)$$

où I_d est la matrice identité de $\mathbb{R}^{d \times d}$.

Ainsi, pour tout $E \subset \mathbb{R}^d$ (mesurable), quand $n \rightarrow \infty$,

$$\mathbb{P} \left[\sqrt{n}\mathbb{I}(\hat{\theta}_n^{\text{mv}})^{1/2}(\hat{\theta}_n^{\text{mv}} - \theta) \in E \right] \longrightarrow \mathbb{P}[G \in E]$$

Région de confiance asymptotique pour l'EMV (2/2)

Pour $B_2^d = \{x \in \mathbb{R}^d : \|x\|_2 \leq 1\}$, $G \sim \mathcal{N}(0, I_d)$ et $c > 0$, on a :

$$\mathbb{P} \left[\sqrt{n} \mathbb{I}(\hat{\theta}_n^{\text{mv}})^{1/2} (\hat{\theta}_n^{\text{mv}} - \theta) \in c B_2^d \right] \longrightarrow \mathbb{P}[\|G\|_2 \leq c]$$

Comme $\|G\|_2^2$ suit une loi du $\chi^2(d)$ ("khi2 à d degrés de liberté"), on a pour tout $\alpha \in (0, 1)$, quand $n \rightarrow \infty$,

$$\mathbb{P} \left[\theta \in \mathcal{I}_{n,\alpha} \right] \longrightarrow 1 - \alpha$$

où

$$\mathcal{I}_{n,\alpha} = \hat{\theta}_n^{\text{mv}} + \frac{q_{1-\alpha}^{\chi^2(d)}}{\sqrt{n}} \mathbb{I}(\hat{\theta}_n^{\text{mv}})^{-1/2} B_2^d$$

C'est une ellipse centrée en $\hat{\theta}_n^{\text{mv}}$.

Formules de calcul de l'information de Fisher

Proposition

Dans un modèle régulier :

$$\begin{aligned}\mathbb{I}(\theta) &= \mathbb{E}_\theta [(\partial_\theta \log f(\theta, X))^2] \\ &= -\mathbb{E}_\theta \partial_\theta^2 \log f(\theta, X) \\ &= -\partial_a^2 \mathbb{D}(a, \theta) \Big|_{a=\theta}\end{aligned}$$

où $\mathbb{D}(a, \theta) = \mathbb{E}_\theta [\log f(a, X)]$.

Interprétation géométrique

- ▶ On pose $\mathbb{D}(a, \theta) = \mathbb{E}_\theta [\log f(a, X)]$. On a (par l'inégalité d'entropie) que

$$\begin{aligned}\mathbb{D}(a, \theta) &= \int_{\mathbb{R}} \log f(a, x) f(\theta, x) \mu(dx) \\ &\leq \int_{\mathbb{R}} \log f(\theta, x) f(\theta, x) \mu(dx) = \mathbb{D}(\theta, \theta).\end{aligned}$$

- ▶ On a

$$\mathbb{I}(\theta) = -\partial_a^2 \mathbb{D}(a, \theta) \Big|_{a=\theta}$$

- ▶ $\mathbb{I}(\theta)$ est petite : le **rayon de courbure de $a \mapsto \mathbb{D}(a, \theta)$ est grand** dans un voisinage de θ : la stabilisation d'un maximum empirique (l'EMV) est plus difficile, rendant moins précis l'estimation.
- ▶ $\mathbb{I}(\theta)$ est grande : le **rayon de courbure est petit** et le maximum de l'EMV est mieux localisé.

Exercices classiques

Savoir calculer : 1) la vraisemblance, 2) l'EMV, 3) l'information de Fisher, 4) le comportement asymptotique de l'EMV, 5) un intervalle de confiance pour θ ; pour le modèle d'échantillonnage de loi :

- ▶ Bernoulli $\mathcal{B}(\theta)$
- ▶ Normal $\mathcal{N}(m, \nu)$
- ▶ Uniforme $\mathcal{U}([a, b])$
- ▶ Poisson $\mathcal{P}(\theta)$
- ▶ L'estimation du paramètre d'une loi exponentielle avec ou sans censure.
- ▶ modèle de régression

application : efficacité à un pas

- ▶ Dans un modèle régulier, le **calcul numérique** de l'EMV peut être difficile à réaliser.
- ▶ Si l'on dispose d'un estimateur $\hat{\theta}_n$ **asymptotiquement normal** et si les évaluations

$$\ell'_n(\theta) = \frac{1}{n} \sum_{i=1}^n \partial_1 \log f(\theta, X_i), \quad \ell''_n(\theta) = \frac{1}{n} \sum_{i=1}^n \partial_1^2 \log f(\theta, X_i)$$

sont **faciles**, alors on peut **corriger** $\hat{\theta}_n$ de sorte d'avoir le même comportement asymptotique que l'EMV :

$$\tilde{\theta}_n = \hat{\theta}_n - \frac{\ell'_n(\hat{\theta}_n)}{\ell''_n(\hat{\theta}_n)} \quad (\text{algorithme de Newton})$$

satisfait

$$\sqrt{n}(\tilde{\theta}_n - \theta) \xrightarrow{d} \mathcal{N}\left(0, \frac{1}{\mathbb{I}(\theta)}\right)$$