MAP 433 : Introduction aux méthodes statistiques. Cours 4

18 Septembre 2015

Aujourd'hui

1. Rappels
   - Principe de maximum de vraisemblance
2. Loi asymptotique des Z- et M- estimateurs
   - Approche asymptotique
3. Modèles réguliers et information de Fisher
   - Construction de l'information de Fisher
   - Modèle régulier
   - Cadre général et interprétation géométrique

M-estimation

- Situation : on observe $X_1, \ldots, X_n$ de loi $\mathbb{P}_\theta$ sur $\mathbb{R}$ et $\theta \in \Theta$.
- Principe : Se donner une application $\psi : \Theta \times \mathbb{R} \rightarrow \mathbb{R}_+$ telle que, pour tout $\theta \in \Theta \subset \mathbb{R}^d$,

$$\alpha \rightsquigarrow \mathbb{E}_\theta\left[\psi(\alpha, X)\right] = \int \psi(\alpha, x)\, \mathbb{P}_\theta(dx)$$

admet un maximum ou un minimum en $\alpha = \theta$.

Definition

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

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

Au lieu de maximiser, on peut aussi minimiser

Paramètre de localisation

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

$$\boxed{\psi(\alpha, x) = (\alpha - x)^2}$$

- La fonction

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

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

- *M*-estimateur associé :

$$\widehat{\theta}_n = \operatorname{argmin}_{\alpha\in\mathbb{R}} \sum_{i=1}^n (X_i - \alpha)^2 = n^{-1}\sum_{i=1}^n X_i.$$

Paramètre de localisation

- C'est aussi un Z-estimateur associé à $\phi(\alpha, x) = 2(x - \alpha)$: on résout
$$\sum_{i=1}^{n}(X_i - \alpha) = 0 \;\text{ d'où }\; \widehat{\theta}_n = \overline{X}_n.$$
- Dans cet exemple très simple, tous les points de vue coïncident.
- Si, dans le même contexte, $\int_{\mathbb{R}} x^2\,\mathbb{P}_\theta(dx) = +\infty$ et $f(x) = f(-x)$, on peut utiliser Z-estimateur avec $\phi(\alpha, x) = \mathrm{Arctg}(x - \alpha)$. Méthode robuste, mais est-elle optimale ? Peut-on faire mieux si f est connue ? A suivre...

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

$$\phi(\alpha, x) = \partial_\theta \psi(\alpha, x),$$

on a

$$\boxed{\sum_{i=1}^{n} \partial_\theta \psi(\alpha, X_i)\big|_{\alpha=\widehat{\theta}_n} = \sum_{i=1}^{n} \phi(\widehat{\theta}_n, X_i) = 0}.$$

Fonction de vraisemblance

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

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

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

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

Remarques

- Log-vraisemblance :

$$\theta \rightsquigarrow \ell_n(\theta, X_1, \ldots, X_n) = n^{-1} \log \mathcal{L}_n(\theta, X_1, \ldots, X_n)$$
$$= n^{-1} \sum_{i=1}^{n} \log f(\theta, X_i).$$

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

Max. vraisemblance = max. log-vraisemblance.

- L'estimateur du maximum de vraisemblance ne dépend pas du choix de la mesure dominante μ.
- Racine de l'équation de vraisemblance : tout estimateur $\widehat{\theta}_n^{\mathtt{rv}}$ vérifiant

$$\nabla_\theta \ell_n(\widehat{\theta}_n^{\mathtt{rv}}, X_1, \ldots, X_n) = 0.$$

l'EMV est un M-estimateur

- On pose

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

(on suppose que $f(\alpha, \cdot) > 0$.)

- La fonction

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

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

- Le M-estimateur associé à ψ maximise la fonction

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

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

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

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

  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.)

Asymptotique des Z- et M-estimateurs

- Problème général délicat. Dans ce cours : conditions suffisantes.
- Convergence : critère simple pour les M-estimateurs.
- Vitesse de convergence : technique simple pour les Z-estimateurs, à condition de savoir que l'estimateur est convergent.
- Sous des hypothèses de régularité, un M-estimateur est un Z-estimateur.

Convergence des M-estimateurs

- Situation : on observe $X_1, \ldots, X_n$ i.i.d. de loi dans la famille $\{\mathbb{P}_\theta, \theta \in \Theta\}$.
- $\psi : \Theta \times \mathbb{R} \to \mathbb{R}$ fonction de contraste.
- Loi des grands nombres :

$$M_n(\alpha) = \frac{1}{n}\sum_{i=1}^n \psi(\alpha, X_i)$$

converge en $\mathbb{P}_\theta$-probabilité vers

$$M(\alpha, \theta) = \mathbb{E}_\theta\left[\psi(\alpha, X)\right]$$

qui atteint son maximum en $\alpha = \theta$

- « à montrer » :

$$\widehat{\theta}_n = \arg\max_{\alpha \in \Theta} M_n(\alpha) \xrightarrow{\mathbb{P}_\theta} \arg\max_{\alpha \in \Theta} \mathbb{E}_\theta\left[\psi(\alpha, X)\right] = \theta.$$

Exemple

- $(X_i)_{i\geq 1}$ variables i.i.d. de loi de densité $(f(x-\theta), \theta \in \mathbb{R})$ où $\int xf(x)dx = 0$ et $\int x^2 f(x)dx = \sigma^2 < \infty$.
- On pose $\psi(\alpha, x) = (\alpha - x)^2$ et donc

$$M_n(\alpha) = n^{-1}\sum_{i=1}^n (X_i - \alpha)^2 = n^{-1}\sum_{i=1}^n (X_i - \bar{X}_n)^2 + (\bar{X}_n - \alpha)^2$$

$$M(\alpha, \theta) = (\theta - \alpha)^2 + \sigma^2$$

Une illustration

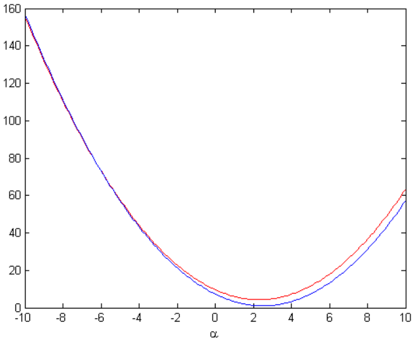

FIGURE – $n = 50$, rouge : $\alpha \mapsto M_n(\alpha)$, bleu : $\alpha \mapsto M(\alpha, \theta)$

Convergence des M-estimateurs

Proposition

Si le M-estimateur $\widehat{\theta}_n$ associé à la fonction de contraste est bien défini et si

- $\sup_{\alpha \in \Theta} |M_n(\alpha) - M(\alpha, \theta)| \xrightarrow{\mathbb{P}_\theta} 0$,
- *pour tout* $\varepsilon > 0$, $\sup_{|\alpha - \theta| \geq \varepsilon} M(\alpha, \theta) < M(\theta, \theta)$ *(condition de maximum)*

alors

$$\widehat{\theta}_n \xrightarrow{\mathbb{P}_\theta} \theta \,.$$

- La condition 1 (convergence uniforme) peut être délicate à montrer...

Consistance des M-estimateurs

- Comme $M(\theta,\theta) \geq M(\alpha,\theta)$ et $M_n(\widehat{\theta}_n) \geq M_n(\alpha)$,

$$\begin{aligned} 0 \leq M(\theta,\theta) - M(\widehat{\theta}_n,\theta) &\leq M(\theta,\theta) - M_n(\theta) + M_n(\widehat{\theta}_n) - M(\widehat{\theta}_n,\theta) \\ &\leq 2 \sup_{\alpha\in\Theta} \left|M_n(\alpha) - M(\alpha,\theta)\right| \xrightarrow{\mathbb{P}_\theta} 0 \end{aligned}$$

Consistance des M-estimateurs

- Comme $M(\theta,\theta) \geq M(\alpha,\theta)$ et $M_n(\widehat{\theta}_n) \geq M_n(\alpha)$,

$$\begin{aligned} 0 \leq M(\theta,\theta) - M(\widehat{\theta}_n,\theta) &\leq M(\theta,\theta) - M_n(\theta) + M_n(\widehat{\theta}_n) - M(\widehat{\theta}_n,\theta) \\ &\leq 2 \sup_{\alpha\in\Theta} |M_n(\alpha) - M(\alpha,\theta)| \xrightarrow{\mathbb{P}_\theta} 0 \end{aligned}$$

- Pour tout $\varepsilon > 0$, il existe $\eta > 0$ tel que $M(\alpha,\theta) \leq M(\theta,\theta) - \eta$ pour tout α tel que $|\alpha - \theta| > \varepsilon$.

Consistance des M-estimateurs

- Comme $M(\theta,\theta) \geq M(\alpha,\theta)$ et $M_n(\widehat{\theta}_n) \geq M_n(\alpha)$,

$$0 \leq M(\theta,\theta) - M(\widehat{\theta}_n,\theta) \leq M(\theta,\theta) - M_n(\theta) + M_n(\widehat{\theta}_n) - M(\widehat{\theta}_n,\theta)$$
$$\leq 2 \sup_{\alpha \in \Theta} |M_n(\alpha) - M(\alpha,\theta)| \xrightarrow{\mathbb{P}_\theta} 0$$

- Pour tout $\varepsilon > 0$, il existe $\eta > 0$ tel que $M(\alpha,\theta) \leq M(\theta,\theta) - \eta$ pour tout α tel que $|\alpha - \theta| > \varepsilon$.
- Par conséquent $\{|\widehat{\theta}_n - \theta| \geq \epsilon\} \subset \{M(\widehat{\theta}_n,\theta) \leq M(\theta,\theta) - \eta\}$.

Loi limite des Z-estimateurs

- Situation : on observe $X_1, \ldots, X_n$ i.i.d. de loi dans la famille $\{\mathbb{P}_\theta, \theta \in \Theta\}$, $\Theta \subset \mathbb{R}$.
- $\widehat{\theta}_n$: Z-estimateur associé à $\phi : \Theta \times \mathbb{R} \to \mathbb{R}$ vérifie

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

- On suppose la suite $(\widehat{\theta}_n)_n$ est consistante, *i.e.* $\widehat{\theta}_n \xrightarrow{\mathbb{P}_\theta} \theta$. Que dire de sa loi limite ?

Loi limite des Z-estimateurs : principe

- Loi des grands nombres

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

- Principe. Développement de Taylor autour de θ :

$$0 = Z_n(\widehat{\theta}_n) = Z_n(\theta) + (\widehat{\theta}_n - \theta)Z_n'(\theta) + \tfrac{1}{2}(\widehat{\theta}_n - \theta)^2 Z''(\widetilde{\theta}_n).$$

- On néglige le reste :

$$\sqrt{n}(\widehat{\theta}_n - \theta) \approx \frac{-\sqrt{n}Z_n(\theta)}{Z_n'(\theta)}$$

Loi limite des Z-estimateurs : principe

- Convergence du numérateur

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

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

- Convergence du dénominateur

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

$\neq 0$ (à supposer).

- $+$ hypothèses techniques pour contrôler le reste (besoin de la convergence de $\widehat{\theta}_n$).

Loi limite des Z-estimateurs

Proposition (Convergence des Z-estimateurs)

- *Soit Θ un ouvert de $\mathbb{R}$. Pour tout $\theta \in \Theta$, $\widehat{\theta}_n \xrightarrow{\mathbb{P}_\theta} \theta$, $\mathbb{E}_\theta\left[\phi(\theta, X)^2\right] < +\infty$ et*

$$\mathbb{E}_\theta\left[\phi(\theta, X)\right] = 0,\ \mathbb{E}_\theta\left[\partial_\theta \phi(\theta, X)\right] \neq 0.$$

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

$$|\partial_\theta^2 \phi(\alpha, x)| \leq g(x),\ \ \mathbb{E}_\theta\left[g(X)\right] < +\infty.$$

Alors

$$\sqrt{n}(\widehat{\theta}_n - \theta) \xrightarrow{d} \mathcal{N}\Big(0, \frac{\mathbb{E}_\theta[\phi(\theta, X)^2]}{\big(\mathbb{E}_\theta[\partial_\theta \phi(\theta, X)]\big)^2}\Big).$$

Approche asymptotique

- Hypothèse simplificatrice : $\theta \in \Theta \subset \mathbb{R}$. On se restreint aux estimateurs asymptotiquement normaux c'est-à-dire vérifiant
$$\sqrt{n}\big(\widehat{\theta}_n - \theta\big) \xrightarrow{d} \mathcal{N}\big(0, v(\theta)\big)$$
cf. théorèmes limites obtenus pour les Z-,M-estimateurs.
- Si $\widehat{\theta}_{n,1}$ et $\widehat{\theta}_{n,2}$ as. normaux de variance asymptotique $v_1(\theta) \leq v_2(\theta)$, alors la précision de $\widehat{\theta}_{n,1}$ est asymptotiquement meilleure que celle de $\widehat{\theta}_{n,2}$ au point θ :
$$\widehat{\theta}_{n,1} = \theta + \sqrt{\frac{v_1(\theta)}{n}}\xi^{(n)}$$
$$\widehat{\theta}_{n,2} = \theta + \sqrt{\frac{v_2(\theta)}{n}}\zeta^{(n)}$$
où $\xi^{(n)}$ et $\zeta^{(n)} \xrightarrow{d} \mathcal{N}(0,1)$.

Comparaison d'estimateurs : cas asymptotique

- Si $v_1(\theta) < v_2(\theta)$, et si $\theta \rightsquigarrow v_i(\theta)$ est continue, on pose

$$\mathcal{C}_{n,\alpha}(\widehat{\theta}_{n,i}) = \left[\widehat{\theta}_{n,i} \pm \sqrt{\frac{v_i(\widehat{\theta}_{n,i})}{n}}\Phi^{-1}(1-\alpha/2)\right], \quad i = 1,2$$

  où $\alpha \in (0,1)$ et $\Phi(\cdot)$ est la fonction de répartition de la loi normale standard.
- $\mathcal{C}_{n,\alpha}(\widehat{\theta}_{n,i})$, $i = 1,2$ sont deux intervalles de confiance asymptotiquement de niveau $1-\alpha$ et on a

$$\frac{|\mathcal{C}_{n,\alpha}(\widehat{\theta}_{n,1})|}{|\mathcal{C}_{n,\alpha}(\widehat{\theta}_{n,2})|} \xrightarrow{\mathbb{P}_\theta^n} \sqrt{\frac{v_1(\theta)}{v_2(\theta)}} < 1.$$

- La notion de longueur minimale possible d'un intervalle de confiance est en général difficile à manipuler.

Conclusion provisoire

- Il est difficile en général de comparer des estimateurs.
- Cadre asymptotique + normalité asymptotique $\rightarrow$ comparaison de la variance asymptotique $\theta \rightsquigarrow v(\theta)$.
- Sous des hypothèses de régularité du modèle $\{\mathbb{P}_\theta^n, \theta \in \Theta\}$

alors

- Il existe une variance asymptotique $v^\star(\theta)$ minimale parmi les variances de la classe des M-estimateurs as. normaux.
- Cette fonction est associée à une quantité d'information intrinsèque au modèle.
- La variance asymptotique de l'EMV est $v^\star(\theta)$.
- Ceci règle partiellement le problème de l'optimalité.

Régularité d'un modèle statistique et information

- Cadre simplificateur : modèle de densité

$$X_1, \ldots, X_n \text{ i.i.d. de loi } \mathbb{P}_\theta$$

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

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

est bien définie.

Information de Fisher

Definition

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

Remarque : on a $\mathbb{P}_\theta\left[f(\theta, X) > 0\right] = 1$, donc la quantité $\log f(\theta, X)$ est bien définie.

Le cas multidimensionnel

Si $\Theta \subset \mathbb{R}^d$ avec $d > 1$, tous les résultats s'étendent en remplaçant dérivation par rapport à α par différentiabilité dans $\mathbb{R}^d$.
L'information de Fisher devient la matrice d'information de Fisher.

Definition

La matrice d'information de Fisher $\mathbb{I}(\theta) = \big(\mathbb{I}(\theta)_{\ell,\ell'}\big)_{1 \leq \ell,\ell' \leq d}$ associée à la famille de densités $\{f(\theta),\, \theta \in \Theta\}$ avec $\theta \in \Theta \subset \mathbb{R}^d$ est définie au point θ par

$$\mathbb{I}(\theta)_{\ell,\ell'} = \mathbb{E}_\theta\left[\partial_{\theta_\ell} \log f(\theta, X) \partial_{\theta_{\ell'}} \log f(\theta, X)\right], \quad 1 \leq \ell, \ell' \leq d,$$

pour peu que cette quantité soit bien définie, avec $\theta = (\theta_1, \ldots, \theta_d)^T$. C'est une matrice symétrique positive.

Information dans quel sens ? Origine de la notion

- Supposons l'EMV $\widehat{\theta}_n^{\,\mathtt{mv}}$ bien défini et convergent.
- Supposons l'application $(\theta, x) \rightsquigarrow f(\theta, x)$ possédant toutes les propriétés de régularité et d'intégrabilité voulues.
- Alors

$$\boxed{\sqrt{n}\Big(\widehat{\theta}_n^{\,\mathtt{mv}} - \theta\Big) \xrightarrow{d} \mathcal{N}\Big(0, \frac{1}{\mathbb{I}(\theta)}\Big)}$$

en loi sous $\mathbb{P}_\theta$, où encore

$$\widehat{\theta}_n^{\,\mathtt{mv}} \overset{d}{\approx} \theta + \frac{1}{\sqrt{n\mathbb{I}(\theta)}}\mathcal{N}(0,1)$$

en loi sous $\mathbb{P}_\theta$.

Construction de l'information + jeu d'hypothèses attenant

- Heuristique : on établira un jeu d'hypothèses justifiant a posteriori le raisonnement.
- Etape 1 : l'EMV $\widehat{\theta}_{\mathrm{n}}^{\,\mathtt{mv}}$ converge :

$$\widehat{\theta}_{\mathrm{n}}^{\,\mathtt{mv}} \xrightarrow{\mathbb{P}_\theta} \theta$$

via le théorème de convergence des M-estimateurs.

- Etape 2 : l'EMV $\widehat{\theta}_{\mathrm{n}}^{\,\mathtt{mv}}$ est un Z-estimateur :

$$0 = \partial_\theta \Big(\sum_{i=1}^{n} \log f(\alpha, X_i) \Big)_{\alpha = \widehat{\theta}_{\mathrm{n}}^{\,\mathtt{mv}}}.$$

Construction de $\mathbb{I}(\theta)$ cont.

- Etape 3 : développement asymptotique autour de θ :

$$0 \approx \sum_{i=1}^n \partial_\theta \log f(\theta, X_i) + (\widehat{\theta}_n^{\,\mathtt{mv}} - \theta) \sum_{i=1}^n \partial_\theta^2 \log f(\theta, X_i),$$

soit

$$\widehat{\theta}_n^{\,\mathtt{mv}} - \theta \approx - \frac{\sum_{i=1}^n \partial_\theta \log f(\theta, X_i)}{\sum_{i=1}^n \partial_\theta^2 \log f(\theta, X_i)}$$

- Etape 4 : le numérateur. Normalisation et convergence de $\sum_{i=1}^n \partial_\theta \log f(\theta, X_i)$?

Numérateur

Lemme

On a

$$\mathbb{E}_\theta\left[\partial_\theta \log f(\theta, X)\right] = 0.$$

Numérateur

Lemme

On a

$$\mathbb{E}_\theta\left[\partial_\theta \log f(\theta, X)\right] = 0.$$

Démonstration.

$$\begin{aligned}
\mathbb{E}_\theta\left[\partial_\theta \log f(\theta, X)\right] &= \int_{\mathbb{R}} \partial_\theta \log f(\theta, x) f(\theta, x)\mu(dx) \\
&= \int_{\mathbb{R}} \frac{\partial_\theta f(\theta, x)}{f(\theta, x)} f(\theta, x)\mu(dx) \\
&= \int_{\mathbb{R}} \partial_\theta f(\theta, x)\mu(dx) = \partial_\theta \int_{\mathbb{R}} f(\theta, x)\mu(dx) = 0.
\end{aligned}$$

Dénominateur

De même $\int_{\mathbb{R}} \partial_\theta^2 f(\theta, x)\mu(dx) = 0$. Conséquence :

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

En effet

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

$$= \int_{\mathbb{R}} \frac{\partial_\theta^2 f(\theta, x) f(\theta, x) - \left(\partial_\theta f(\theta, x)\right)^2}{f(\theta, x)^2} f(\theta, x)\mu(dx)$$

$$= \int_{\mathbb{R}} \partial_\theta^2 f(\theta, x)\mu(dx) - \int_{\mathbb{R}} \frac{\left(\partial_\theta f(\theta, x)\right)^2}{f(\theta, x)}\mu(dx)$$

$$= 0 - \int_{\mathbb{R}} \left(\frac{\partial_\theta f(\theta, x)}{f(\theta, x)}\right)^2 f(\theta, x)\mu(dx) = -\mathbb{E}\left[\left(\partial_\theta \log f(\theta, X)\right)^2\right].$$

Conséquences

- Les $\partial_\theta \log f(\theta, X_i)$ sont i.i.d. et $\mathbb{E}_\theta\big[\partial_\theta \log f(\theta, X)\big] = 0$. TCL :

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \partial_\theta \log f(\theta, X_i) \xrightarrow{d} \mathcal{N}\Big(0, \mathbb{E}_\theta\Big[\big(\partial_\theta \log f(\theta, X)\big)^2\Big]\Big)$$
$$= \mathcal{N}\big(0, \mathbb{I}(\theta)\big).$$

- Les $\partial_\theta^2 \log f(\theta, X_i)$ sont i.i.d. LGN :

$$\frac{1}{n} \sum_{i=1}^{n} \partial_\theta^2 \log f(\theta, X_i) \xrightarrow{\mathbb{P}_\theta} \mathbb{E}_\theta\big[\partial_\theta^2 \log f(\theta, X)\big]$$
$$\overset{\text{conséquence}}{=} -\mathbb{I}(\theta).$$

Conclusion

- En combinant les deux estimations + lemme de Slutsky :

$$\sqrt{n}(\widehat{\theta}_n^{\mathtt{mv}} - \theta) \approx -\frac{\frac{1}{\sqrt{n}}\sum_{i=1}^n \partial_\theta \log f(\theta, X_i)}{\frac{1}{n}\sum_{i=1}^n \partial_\theta^2 \log f(\theta, X_i)}$$

$$\xrightarrow{d} \frac{\mathcal{N}\big(0, \mathbb{I}(\theta)\big)}{\mathbb{I}(\theta)}$$

$$\stackrel{\text{loi}}{=} \mathcal{N}\Big(0, \frac{1}{\mathbb{I}(\theta)}\Big).$$

- Le raisonnement est rigoureux dès lors que : (i) on a la convergence de $\widehat{\theta}_n^{\mathtt{mv}}$, (ii) on peut justifier le lemme et sa conséquence, (iii) $\mathbb{I}(\theta)$ est bien définie et non dégénérée et (iv) on sait contrôler le terme de reste dans le développement asymptotique, partie la plus difficile.

Modèle régulier

Definition

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

- Θ *ouvert et* $\{f(\theta,\cdot)>0\}=\{f(\theta',\cdot)>0\}$, $\forall\theta,\theta'\in\Theta$.
- μ*-p.p.* $\theta\rightsquigarrow f(\theta,\cdot)$, $\theta\rightsquigarrow\log f(\theta,\cdot)$ *sont* $\mathcal{C}^2$.
- *Pour tout* $\theta\in\Theta$, *il existe* $\mathcal{V}_\theta\subset\Theta$ *t.q. pour* $\alpha\in\mathcal{V}_\theta$

$$|\partial_\theta^2\log f(\alpha,x)|+|\partial_\theta\log f(\alpha,x)|+\big(\partial_\theta\log f(\alpha,x)\big)^2\leq g(x)$$

où

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

- *L'information de Fisher est non-dégénérée : pour tout* $\theta\in\Theta$,

Résultat principal

Proposition

- *Si l'expérience engendrée par l'observation $X_1, \ldots, X_n \sim_{i.i.d.} \mathbb{P}_\theta$ est associée à une famille de probabilités $\{\mathbb{P}_\theta, \theta \in \Theta\}$ sur $\mathbb{R}$ régulière au sens de la définition précédente, alors*

$$\sqrt{n}\big(\widehat{\theta}_{\mathrm{n}}^{\,\mathtt{mv}} - \theta\big) \xrightarrow{d} \mathcal{N}\Big(0, \frac{1}{\mathbb{I}(\theta)}\Big).$$

- *Si $\widehat{\theta}_n$ est un Z-estimateur régulier asymptotiquement normal de variance $v(\theta)$, alors*

$$\forall \theta \in \Theta, \quad v(\theta) \geq \frac{1}{\mathbb{I}(\theta)}.$$

Preuve de la proposition

- Le premier point consiste à rendre rigoureux le raisonnement précédent. Point délicat : le contrôle du terme de reste.
- Optimalité de la variance de l'EMV parmi celle des Z-estimateurs : on a vu que si $\widehat{\theta}_n$ est un Z-estimateur régulier associé à la fonction ϕ, alors, sa variance asymptotique $v(\theta) = v_\phi(\theta)$ vaut

$$v_\phi(\theta) = \frac{\mathbb{E}_\theta\left[\phi(\theta, X)^2\right]}{\left(\mathbb{E}_\theta\left[\partial_\theta \phi(\theta, X)\right]\right)^2}.$$

- A montrer : pour toute fonction ϕ :

$$\boxed{\frac{\mathbb{E}_\theta\left[\phi(\theta, X)^2\right]}{\left(\mathbb{E}_\theta\left[\partial_\theta \phi(\theta, X)\right]\right)^2} \geq \frac{1}{\mathbb{I}(\theta)}}.$$

Preuve de l'inégalité

- Comme pour tout $\theta \in \Theta$, $\mathbb{E}_\theta[\phi(\theta, X)] = 0$,

$$\partial_\theta \, \mathbb{E}_\theta \left[\phi(\theta, X)\right] = 0.$$

- (avec $\dot{\phi}(\theta, x) = \partial_\theta \phi(\theta, x)$)

$$\begin{aligned} 0 &= \int_{\mathbb{R}} \left[\dot{\phi}(\theta, x) f(\theta, x) + \phi(\theta, x) \partial_\theta f(\theta, x)\right] \mu(dx) \\ &= \int_{\mathbb{R}} \left[\dot{\phi}(\theta, x) f(\theta, x) + \phi(\theta, x) \partial_\theta \log f(\theta, x) f(\theta, x)\right] \mu(dx). \end{aligned}$$

- Conclusion

$$\boxed{\mathbb{E}_\theta \left[\dot{\phi}(\theta, X)\right] = - \, \mathbb{E}_\theta \left[\phi(\theta, X) \partial_\theta \log f(\theta, X)\right]}$$

Preuve de l'inégalité (fin)

- On a

$$\mathbb{E}_\theta\left[\dot{\phi}(\theta, X)\right] = -\,\mathbb{E}_\theta\left[\phi(\theta, X)\partial_\theta \log f(\theta, X)\right]$$

- Cauchy-Schwarz :

$$\left(\mathbb{E}_\theta\left[\dot{\phi}(\theta, X)\right]\right)^2 \leq \mathbb{E}_\theta\left[\phi(\theta, X)^2\right]\mathbb{E}_\theta\left[\left(\partial_\theta \log f(\theta, X)\right)^2\right],$$

c'est-à-dire

$$\boxed{v_\phi(\theta)^{-1} = \frac{\left(\mathbb{E}_\theta\left[\dot{\phi}(\theta, X)\right]\right)^2}{\mathbb{E}_\theta\left[\phi(\theta, X)^2\right]} \leq \mathbb{I}(\theta).}$$

Le cas multidimensionnel

Proposition

- *Si l'expérience engendrée par l'observation $X_1, \ldots, X_n \sim_{i.i.d.} \mathbb{P}_\theta$ est associée à une famille de probabilités $\{\mathbb{P}_\theta, \theta \in \Theta\}$, $\Theta \subset \mathbb{R}^d$ régulière, alors*

$$\sqrt{n}\left(\widehat{\theta}_{\mathtt{n}}^{\mathtt{mv}} - \theta\right) \xrightarrow{d} \mathcal{N}\left(0, \frac{1}{\mathbb{I}(\theta)}\right).$$

- *Si $\widehat{\theta}_n$ est un Z-estimateur régulier asymptotiquement normal de variance $v(\theta)$, alors*

$$\forall \theta \in \Theta, \quad v(\theta) \geq \frac{1}{\mathbb{I}(\theta)}.$$

Information de Fisher dans un modèle général

Definition

- *Situation : suite d'expériences statistiques*

$$\mathcal{E}^n = \left(\mathfrak{Z}^n, \mathcal{Z}^n, \{\mathbb{P}_\theta^n, \theta \in \Theta\}\right)$$

dominées par μ_n, associées à l'observation $Z^{(n)}$,

$$f_n(\theta, z) = \frac{d\,\mathbb{P}_\theta^n}{d\mu^n}(z), \quad z \in \mathfrak{Z}^n, \theta \in \Theta \subset \mathbb{R}\,.$$

- *Information de Fisher (si elle existe) de l'expérience au point θ :*

$$\mathbb{I}^n(\theta) = \mathbb{E}_\theta^n\left[\left(\partial_\theta \log f_n(\theta, Z^{(n)})\right)^2\right]$$

Interprétation géométrique

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

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

- On a

$$\boxed{\mathbb{I}(\theta) = \partial_\alpha^2 \mathbb{D}(\alpha,\theta)\big|_{\alpha=\theta}.}$$

  - Si $\mathbb{I}(\theta)$ est « petite », le rayon de courbure de $\alpha \rightsquigarrow \mathbb{D}(\alpha,\theta)$ est grand dans un voisinage de θ : la stabilisation d'un maximum empirique (l'EMV) est plus difficile, rendant moins précis l'estimation.
  - Si $\mathbb{I}(\theta)$ est « grande », le rayon de courbure est petit et le

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 $\widehat{\theta}_n$ asymptotiquement normal et si les évaluations

$$\ell_n'(\theta) = \tfrac{1}{n}\sum_{i=1}^{n}\partial_\theta \log f(\theta, X_i), \quad \ell_n''(\theta) = \tfrac{1}{n}\sum_{i=1}^{n}\partial_\theta^2 \log f(\theta, X_i)$$

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

$$\widetilde{\theta}_n = \widehat{\theta}_n - \frac{\ell_n'(\widehat{\theta}_n)}{\ell_n''(\widehat{\theta}_n)} \quad \text{(algorithme de Newton)}$$

satisfait

$$\boxed{\sqrt{n}\big(\widetilde{\theta}_n - \theta\big) \xrightarrow{d} \mathcal{N}\Big(0, \frac{1}{\mathbb{I}(\theta)}\Big)}$$