

Statistiques mathématiques : cours 5

Guillaume Lécué

18 septembre 2017

Aujourd'hui

Comparaison d'estimateurs

Optimalité de l'EMV

Borne de Cramer-Rao

Approche non-asymptotique de comparaison d'estimateurs

Comparaison d'estimateurs dans les modèles paramétriques dominés

Modèle d'échantillonnage dominé paramétrique : on observe

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

où $\mathbb{P}_\theta = f(\theta, \cdot) \cdot \mu$

Plusieurs estimateurs : moments, Z - et M -estimateurs, **EMV**

Question : Quel estimateur choisir ? comment comparer des estimateurs ?

Vitesse de convergence et régions de confiance

$\hat{\theta}_n$ estimateur de θ . Il y a deux types de résultats :

- ▶ vitesse de convergence **non-asymptotique** : à n fixé

$$\mathbb{E} [\|\hat{\theta}_n - \theta\|^2] \leq c_n(\theta)^2, \text{ (où } c_n(\theta) \downarrow 0)$$

- ▶ vitesse de convergence **asymptotiques** : quand $n \rightarrow +\infty$,

$$v_n(\hat{\theta}_n - \theta) \xrightarrow{d} Z_\theta, \text{ (où } v_n^{-1} \downarrow 0)$$

Ces deux résultats permettent de construire des intervalles / régions de confiance (non-asymptotique / asymptotique).

Idee : La taille de ces régions de confiance caractérise la qualité d'estimation de θ par $\hat{\theta}_n$

Régions de confiance : définition formelle

Définition

Dans le modèle d'échantillonnage $\{\mathbb{P}_\theta : \theta \in \Theta\}$ pour $\Theta \subset \mathbb{R}^d$. Soit $\alpha \in (0, 1)$. Pour tout $n \geq 1$, on considère un sous-ensemble observable $\mathcal{C}_{n,\alpha} = \mathcal{C}_{n,\alpha}(X_1, \dots, X_n)$ de \mathbb{R}^d . $\mathcal{C}_{n,\alpha}$ est appelé :

1. *région de confiance de niveau $1 - \alpha$ quand*

$$\forall \theta \in \Theta : \mathbb{P}_\theta [\theta \in \mathcal{C}_{n,\alpha}] \geq 1 - \alpha$$

2. *région de confiance de niveau asymptotique $1 - \alpha$ quand*

$$\forall \theta \in \Theta : \liminf_{n \rightarrow \infty} \mathbb{P}_\theta [\theta \in \mathcal{C}_{n,\alpha}] \geq 1 - \alpha.$$

Comparaison d'estimateurs

Dans le modèle d'échantillonnage, comment **construire** le **meilleur** estimateur ? Dans quel sens ?

- ▶ Intuitivement : $\hat{\theta}_n$ fournit une précision d'estimation de θ optimale si on peut lui associer une **région de confiance de longueur / volume minimale** (à un niveau $1 - \alpha$ donné).
- ▶ Différence entre point de vue **asymptotique** et **non-asymptotique**. Dans ce cours, nous étudions le point de vue asymptotique :
 - ▶ **vitesse de convergence asymptotique**
 - ▶ **variance asymptotique**

et on donne une brève initiation à la notion d'optimalité en non-asymptotique.

Approche asymptotique

Hypothèse : $\theta \in \Theta \subset \mathbb{R}$ et on se restreint aux estimateurs $\hat{\theta}_n$
asymptotiquement normaux :

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} \mathcal{N}(0, v(\theta))$$

(théorèmes limites pour les Z -, M -estimateurs, EMV, plug-in)

Intuitivement : on a

$$\hat{\theta}_n \approx \theta + \sqrt{\frac{v(\theta)}{n}} \mathcal{N}(0, 1)$$

Donc la variance asymptotique mesure :

1. l'"incertitude" de l'estimation de θ par $\hat{\theta}_n$
2. la taille des régions de confiance asymptotiques

efficacité asymptotique

Définition

Soit $\hat{\theta}_{n,1}$ et $\hat{\theta}_{n,2}$ deux estimateurs asymptotiquement normaux de variance asymptotique respectives $v_1(\theta)$ et $v_2(\theta)$. On dit que $\hat{\theta}_{n,1}$ est **plus efficace** que $\hat{\theta}_{n,2}$ quand :

$$v_1(\theta) \leq v_2(\theta) \text{ pour tout } \theta \in \Theta$$

Un estimateur est **asymptotiquement efficace** s'il n'existe pas d'autre estimateur (dans la classe considérée) plus efficace que lui.

- ▶ on ne compare que les estimateurs asymptotiquement normaux (de vitesse de convergence en $1/\sqrt{n}$)
- ▶ cela exclut les estimateurs pathologique comme $\hat{\theta}_n = \theta_0$

Efficacité asymptotique de l'EMV

- ▶ Cadre : modèle d'échantillonnage dominé paramétrique

$$X_1, \dots, X_n \stackrel{i.i.d.}{\sim} \mathbb{P}_\theta = f(\theta, \cdot) \cdot \mu \text{ pour } \theta \in \Theta \subset \mathbb{R}$$

- ▶ Hypothèse : la quantité

$$\mathbb{I}(\theta) = \mathbb{E}_\theta \left[(\partial_\theta \log f(\theta, X_1))^2 \right]$$

est bien définie. C'est l'**information de Fisher** du modèle $\{\mathbb{P}_\theta, \theta \in \Theta\}$ en θ .

Définition

La famille de densités $\{f(\theta, \cdot), \theta \in \Theta\}$, par rapport à la mesure dominante μ , et $\Theta \subset \mathbb{R}$ 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 tout $a \in \mathcal{V}_\theta$

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

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

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

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

Résultat principal

Theorem

Dans le modèle d'échantillonnage dominé (pour $\Theta \subset \mathbb{R}$) tel que $\{\mathbb{P}_\theta, \theta \in \Theta\}$ est *régulier* on a :

- ▶ l'EMV est asymptotiquement normal :

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

- ▶ Si $\hat{\theta}_n$ est un Z-estimateur associé à ϕ "régulière", c-à-d tel que $\hat{\theta}_n$ est a.n. de variance asymptotique

$$v(\theta) = \frac{\mathbb{E}_\theta[\phi(\theta, X)^2]}{(\mathbb{E}_\theta[\partial_\theta \phi(\theta, X)])^2} \text{ alors } \boxed{\forall \theta \in \Theta, v(\theta) \geq \frac{1}{\mathbb{I}(\theta)}}$$

Donc l'EMV est asymptotiquement efficace parmi les Z-estimateurs réguliers

Preuve de l'efficacité asymptotique de l'EMV parmi les Z-estimateurs réguliers

Soit $\hat{\theta}_n$ un Z-estimateur régulier associé à la fonction ϕ , alors, sa variance asymptotique vaut

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

A montrer : pour toute fonction ϕ (régulière) :

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

Preuve de l'inégalité (1/2)

- ▶ Par construction de ϕ , on a $\forall \theta \in \Theta, \mathbb{E}_\theta \phi(\theta, X) = 0$ alors

$$F'(\theta) = 0 \text{ où } F(\theta) = \mathbb{E}_\theta [\phi(\theta, X)] = 0$$

- ▶ càd

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

- ▶ Conclusion

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

Preuve de l'inégalité (2/2)

- ▶ On a

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

- ▶ Cauchy-Schwarz :

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

c'est-à-dire

$$\boxed{v_\phi(\theta) = \frac{\mathbb{E}_\theta [\phi(\theta, \mathbf{X})^2]}{(\mathbb{E}_\theta [\partial_\theta \phi(\theta, \mathbf{X})])^2} \geq \frac{1}{\mathbb{I}(\theta)}}$$

Information de Fisher et résultats non-asymptotiques : borne de Cramer-Rao

Borne de Cramer-Rao

Soit Z une observation de l'expérience $(\mathfrak{Z}, \mathcal{Z}, \{\mathbb{P}_\theta : \theta \in \Theta\})$ où

1. Θ est un ouvert de \mathbb{R}
2. $\mathbb{P}_\theta = f(\theta, \cdot) \cdot \mu$ (modèle dominé par μ)
3. le modèle $\{\mathbb{P}_\theta : \theta \in \Theta\}$ est régulier et on note l'information de Fisher par :

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

Alors pour tout estimateur $\hat{\theta}$, on a

$$\mathbb{E}_\theta (\hat{\theta} - \theta)^2 \geq \frac{(1 + b'(\theta))^2}{\mathbb{I}(\theta)} + b(\theta)^2$$

où $b(\theta) = \mathbb{E}_\theta \hat{\theta}(Z) - \theta$ est le biais de $\hat{\theta}$.

En particulier, si l'estimateur $\hat{\theta}$ est sans biais alors son risque quadratique est plus grand que $\mathbb{I}(\theta)^{-1}$.

Mise en perspective de Cramer-Rao

Le résultat de l'efficacité asymptotique de l'EMV parmi les Z -estimateurs réguliers dit que si $\hat{\theta}_n$ est un Z -estimateur régulier alors

$$\hat{\theta}_n \approx \theta + \sqrt{\frac{v_\phi(\theta)}{n}} \mathcal{N}(0, 1) \text{ où } v_\phi(\theta) \geq \frac{1}{\mathbb{I}(\theta)}$$

alors, intuitivement,

$$\mathbb{E}_\theta (\hat{\theta}_n - \theta)^2 \approx \frac{v_\phi(\theta)}{n} \geq \frac{1}{n\mathbb{I}(\theta)} = \frac{1}{\mathbb{I}(\theta|\mathcal{E}^n)}$$

On retrouve donc la version "asymptotique" de Cramer-Rao pour les Z -estimateurs réguliers (qui sont "asymptotiquement sans biais" car consistant).

Preuve de Cramer-Rao

On note :

- ▶ $R_\theta(\hat{\theta}) = \mathbb{E}_\theta (\hat{\theta} - \theta)^2$: le risque quadratique de $\hat{\theta}$
- ▶ $b(\theta) = \mathbb{E}_\theta \hat{\theta}(Z) - \theta$: le biais de $\hat{\theta}$
- ▶ $\ell(\theta, z) = \log f(\theta, z)$: le score de θ en z

Montrer que :

1. $R_\theta(\hat{\theta}) = b(\theta)^2 + \text{var}_\theta(\hat{\theta})$ (décomposition biais/variance)
2. $\mathbb{E}_\theta \partial_\theta \ell(\theta, Z) = 0$
3. $b'(\theta) = \mathbb{E}_\theta [\hat{\theta}(Z) \partial_\theta \ell(\theta, Z)] - 1$
4. $1 + b'(\theta) = \mathbb{E}_\theta [(\hat{\theta} - \mathbb{E}_\theta \hat{\theta}) \partial_\theta \ell(\theta, Z)]$
5. en déduire la borne de Cramer-Rao

Approche non-asymptotique de comparaison d'estimateurs

Risque quadratique

Définition

Soit Z une observation de l'expérience $(\mathfrak{Z}, \mathcal{Z}, \{\mathbb{P}_\theta : \theta \in \Theta\})$ pour $\Theta \subset \mathbb{R}$.

1. Soit $\hat{\theta} = \hat{\theta}(Z)$ un estimateur. On appelle *risque quadratique* de $\hat{\theta}$ au point $\theta \in \Theta$:

$$\mathcal{R}(\hat{\theta}_n, \theta) = \mathbb{E}_\theta [(\hat{\theta} - \theta)^2]$$

2. Soient $\hat{\theta}_1, \hat{\theta}_2$ deux estimateurs. On dit que $\hat{\theta}_1$ est *préférable au sens du risque quadratique* à $\hat{\theta}_2$ quand

$$\forall \theta \in \Theta, \mathcal{R}(\hat{\theta}_1, \theta) \leq \mathcal{R}(\hat{\theta}_2, \theta)$$

Absence d'optimalité (1/3)

Question : Existe-t-il un estimateur **optimal** θ^* au sens où

$$\forall \theta \in \Theta, \mathcal{R}(\theta^*, \theta) \leq \inf_{\hat{\theta}} \mathcal{R}(\hat{\theta}_n, \theta) ?$$

Proposition

Si $\Theta = \{\theta_1, \theta_2\}$ et *s'il n'existe pas d'événement observable* $A \in \mathcal{Z}$ tel que, *simultanément* :

$$\mathbb{P}_{\theta_1} [A] = 0 \text{ et } \mathbb{P}_{\theta_2} [A] = 1,$$

(on dit que \mathbb{P}_{θ_1} et \mathbb{P}_{θ_2} ne sont *pas étrangères*), alors *il n'existe pas d'estimateur optimal*

Absence d'optimalité (2/3)

- ▶ Preuve : Pour tout estimateur θ^* , on a

$$\max \{ \mathcal{R}(\theta^*, \theta_1), \mathcal{R}(\theta^*, \theta_2) \} > 0 \quad (*)$$

- ▶ Supposons θ^* estimateur optimal et $\mathcal{R}(\theta^*, \theta_1) > 0$. Alors $\hat{\theta}^{\text{trivial}} := \theta_1$ vérifie

$$0 = \mathcal{R}(\hat{\theta}^{\text{trivial}}, \theta_1) < \mathcal{R}(\theta^*, \theta_1) \quad \text{contradiction !}$$

ceci contredit l'optimalité de θ^*

Absence d'optimalité (3/3)

Preuve de (\star) : si $\mathcal{R}(\theta^*, \theta_1) = \mathcal{R}(\theta^*, \theta_2) = 0$, alors

$$\theta^* = \theta_1 \quad \mathbb{P}_{\theta_1} \text{-p.s.} \quad \text{et} \quad \theta^* = \theta_2 \quad \mathbb{P}_{\theta_2} \text{-p.s..}$$

Soient

$$A = \{z \in \mathfrak{Z} : \theta^*(z) = \theta_1\} \text{ et } B = \{z \in \mathfrak{Z} : \theta^*(z) = \theta_2\}$$

Alors $\mathbb{P}_{\theta_1}[A] = 1$ et donc $\mathbb{P}_{\theta_2}[A] > 0$. Aussi, $\mathbb{P}_{\theta_2}[B] = 1$. Donc $A \cap B \neq \emptyset$.

Alors il existe $z_0 \in \mathfrak{Z}$ tel que $\theta_1 = \theta^*(z_0) = \theta_2$: **contradiction !**

Notions d'optimalité non-asymptotique

- ▶ **Différentes notions existent.** Deux exemples extrêmes :

Définition (Admissibilité et critère minimax)

- ▶ Un estimateur θ^* est **admissible** s'il **n'existe pas** d'estimateur $\hat{\theta}$ **préférable** à θ^* c'à d tel que, pour **tout** $\theta \in \Theta$,

$$\mathcal{R}(\hat{\theta}, \theta) < \mathcal{R}(\theta^*, \theta).$$

- ▶ Un estimateur θ^* est **minimax** si

$$\sup_{\theta \in \Theta} \mathcal{R}(\theta^*, \theta) = \inf_{\hat{\theta}} \sup_{\theta \in \Theta} \mathcal{R}(\hat{\theta}, \theta)$$

- ▶ **Admissibilité** : permet d'éliminer des estimateurs absurdes (mais pas tous : $\hat{\theta}_n = \theta_0$)
- ▶ **Minimaxité** : notion très **robuste mais conservatrice**

exemple : propriété minimax de la moyenne empirique

cadre : $X_1, \dots, X_n \stackrel{i.i.d.}{\sim} \mathcal{N}(\theta, 1)$ pour $\theta \in \mathbb{R}$

1. la moyenne empirique \bar{X}_n est sans biais et vérifie :

$$\mathcal{R}(\bar{X}_n, \theta) = \frac{1}{n}$$

2. soit $\hat{\theta}_n$ un estimateur sans biais ; par Cramer-Rao, on a :

$$\mathcal{R}(\hat{\theta}_n, \theta) \geq \frac{1}{n}$$

La moyenne empirique est donc **minimax parmi les estimateurs sans biais** dans le modèle $\{\mathcal{N}(\theta, 1) : \theta \in \mathbb{R}\}$