Statistiques mathématiques : cours 5

Guillaume Lecué

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

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

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

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

Question : Quel estimateur choisir ? comment comparer des estimateurs ?

Vitesse de convergence et régions de confiance

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

- vitesse de convergence non-asymptotique : à n fixé

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

- vitesse de convergence asymptotiques : quand $n \longrightarrow +\infty$,

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

Ces deux résultats permettent de construire des intervalles / régions de confiance (non-asymptotique / asymptotique).
Idée : La taille de ces régions de confiance caractérise la qualité d'estimation de θ par $\widehat{\theta}_n$

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* $\mathcal{C}_{n,\alpha} = \mathcal{C}_{n,\alpha}(X_1, \ldots, X_n)$ *de* $\mathbb{R}^d$. $\mathcal{C}_{n,\alpha}$ *est appelé :*

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

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

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

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

Comparaison d'estimateurs

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

- Intuitivement : $\widehat{\theta}_n$ 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 $\widehat{\theta}_n$ asymptotiquement normaux :

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

(théorèmes limites pour les Z-,M-estimateurs, EMV, plug-in)
Intuitivement : on a

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

Donc la variance asymptotique mesure :

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

efficacité asymptotique

Définition

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

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

Un estimateur est asymptotiquement efficace s'il n'éxiste 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 $\widehat{\theta}_n = \theta_0$

Efficacité asymptotique de l'EMV

- Cadre : modèle d'échantillonnage dominé paramétrique

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

- Hypothèse : la quantité

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

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* $a \in \mathcal{V}_\theta$

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

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

- *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 :*

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

- *Si $\widehat{\theta}_n$ est un Z-estimateur associé à ϕ "régulière", càd tel que $\widehat{\theta}_n$ est a.n. de variance asymptotique*

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

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

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

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

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

<u>A montrer</u> : pour toute fonction ϕ (regulière) :

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

Preuve de l'inégalité (1/2)

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

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

- càd

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

- <u>Conclusion</u>

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

Preuve de l'inégalité (2/2)

- On a
$$\mathbb{E}_\theta\left[\partial_\theta \phi(\theta, X)\right] = -\,\mathbb{E}_\theta\left[\phi(\theta, X)\partial_\theta \log f(\theta, X)\right]$$
- Cauchy-Schwarz :
$$\left(\mathbb{E}_\theta\left[\partial_\theta \phi(\theta, X)\right]\right)^2 \leq \mathbb{E}_\theta\left[\phi(\theta, X)^2\right] \mathbb{E}_\theta\left[\left(\partial_\theta \log f(\theta, X)\right)^2\right],$$
c'est-à-dire
$$\boxed{v_\phi(\theta) = \frac{\mathbb{E}_\theta\left[\phi(\theta, X)^2\right]}{\left(\mathbb{E}_\theta\left[\partial_\theta \phi(\theta, X)\right]\right)^2} \geq \frac{1}{\mathbb{I}(\theta)}}$$

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. $\mathbb{P}_\theta = f(\theta, \cdot).\mu$ (modèle dominé par μ)
3. le modèle $\{\mathbb{P}_\theta : \theta \in \Theta\}$ est régulier et on note l'information de Fisher par :
$$\mathbb{I}(\theta) = \mathbb{E}_\theta \left[(\partial_1 \log f(\theta, Z))^2 \right]$$

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

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

où $b(\theta) = \mathbb{E}_\theta \, \hat{\theta}(Z) - \theta$ est le bias 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 $\widehat{\theta}_n$ est un Z-estimateur régulier alors

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

alors, intuitivement,

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

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 :

- $R_\theta(\hat\theta) = \mathbb{E}_\theta \left(\hat\theta - \theta\right)^2$: 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. $R_\theta(\hat\theta) = b(\theta)^2 + \mathrm{var}_\theta(\hat\theta)$ (décomposition bias/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. $1 + b'(\theta) = \mathbb{E}_\theta \left[(\hat\theta - \mathbb{E}_\theta \hat\theta)\partial_\theta \ell(\theta, Z)\right]$
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 $\left(\mathfrak{Z}, \mathcal{Z}, \{\mathbb{P}_\theta : \theta \in \Theta\}\right)$ *pour* $\Theta \subset \mathbb{R}$.

1. *Soit* $\hat{\theta} = \hat{\theta}(Z)$ *un estimateur. On appelle* *risque quadratique* *de* $\hat{\theta}$ *au point* $\theta \in \Theta$ *:*
$$\mathcal{R}(\widehat{\theta}_n, \theta) = \mathbb{E}_\theta \left[\left(\hat{\theta} - \theta \right)^2 \right]$$
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*
$$\forall \theta \in \Theta, \ \ \mathcal{R}\left(\widehat{\theta}_1, \theta\right) \leq \mathcal{R}\left(\widehat{\theta}_2, \theta\right)$$

Absence d'optimalité (1/3)

Question : Existe-t-il un estimateur optimal $\theta^\star$ au sens où

$$\forall \theta \in \Theta, \;\; \mathcal{R}\big(\theta^\star, \theta\big) \leq \inf_{\hat{\theta}} \mathcal{R}\big(\,\widehat{\theta}_n, \theta\big) \; ?$$

Proposition

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

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

(on dit que $\mathbb{P}_{\theta_1}$ et $\mathbb{P}_{\theta_2}$ ne sont pas étrangères), alors il n'existe pas d'estimateur optimal

Absence d'optimalité (2/3)

- Preuve : Pour tout estimateur $\theta^\star$, on a

$$\max\left\{\mathcal{R}(\theta^\star,\theta_1),\mathcal{R}(\theta^\star,\theta_2)\right\} > 0 \qquad (\star)$$

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

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

ceci contredit l'optimalité de $\theta^\star$

Absence d'optimalité (3/3)

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

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

Soient

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

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 $\theta_1 = \theta^\star(z_0) = \theta_2$: contradiction !

Notions d'optimalité non-asymptotique

- Différentes notions existent. Deux exemples extrêmes :

Définition (Admissibilité et critère minimax)

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

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

- *Un estimateur $\theta^\star$ est minimax si*

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

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

exemple : propriété minimax de la moyenne empirique

cadre : $X_1, \ldots, X_n \overset{i.i.d.}{\sim} \mathcal{N}(\theta, 1)$ 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 $\widehat{\theta}_n$ un estimateur sans biais ; par Cramer-Rao, on a :

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

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