

Remise à niveau - Statistique

Comparaison d'estimateurs

Claire Boyer

1. Généralités

2. Information de Fisher

3. Borne de Cramér-Rao

4. Efficacité asymptotique et comparaison des estimateurs

Contexte : Estimation d'un paramètre $\theta \in \Theta$ d'une variable aléatoire X , avec Θ un intervalle ouvert de $\mathbb{R}$

Objectifs :

- ▶ Comparer différents estimateurs de θ
- ▶ Quid de leur optimalité

Comparaison via EQM/risque : on peut dire que $\hat{\theta}_n^{(1)}$ est un meilleur estimateur de θ que $\hat{\theta}_n^{(2)}$ si

$$\mathcal{R}(\hat{\theta}_n^{(1)}) \leq \mathcal{R}(\hat{\theta}_n^{(2)}) \quad \text{càd} \quad \mathbb{E} \left[\|\hat{\theta}_n^{(1)} - \theta\|^2 \right] \leq \mathbb{E} \left[\|\hat{\theta}_n^{(2)} - \theta\|^2 \right] \quad \ddagger$$

Cela signifie que $\hat{\theta}_n^{(1)}$ réalise un meilleur compromis biais-variance

‡. en version multivariée

Biais d'un estimateur On donne trop souvent trop d'importance au biais d'un estimateur. Débiaiser un estimateur ne donne pas forcément de meilleurs résultats.

Exemple : Soit $\theta > 0$ et $X \sim \mathcal{U}([0, \theta])$. Comparer les risques quadratiques des estimateurs suivants :

$$\hat{\theta}_n = X_{(n)}, \quad \tilde{\theta}_n = \frac{n+1}{n} X_{(n)}, \quad \theta_{\alpha,n} = \alpha X_{(n)} \quad (\alpha > 0 \text{ choisi judicieusement}).$$

Comparaison asymptotique Soient $\hat{\theta}_n$ et $\tilde{\theta}_n$ des estimateurs asymptotiquement normaux, i.e., tels qu'il existe σ_1^2, σ_2^2

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, \sigma_1^2), \quad \sqrt{n}(\tilde{\theta}_n - \theta) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, \sigma_2^2).$$

Si $\sigma_1^2 \leq \sigma_2^2$, on préférera alors l'estimateur $\hat{\theta}_n$.

1. Généralités

2. Information de Fisher

3. Borne de Cramér-Rao

4. Efficacité asymptotique et comparaison des estimateurs

Modèle paramétrique : Soit $X \sim P_\theta$ de densité f_θ et de log-densité $\ell_\theta = \log f_\theta$

Score (de Fisher) = Gradient de la log-densité *

$$s(x) = \nabla_\theta \ell_\theta(x) = \begin{pmatrix} \partial_{\theta_1} \ell_\theta(x) \\ \vdots \\ \partial_{\theta_d} \ell_\theta(x) \end{pmatrix}$$

Information de Fisher

$$I(\theta) = \mathbb{E}_\theta [\nabla_\theta \ell_\theta(X) (\nabla_\theta \ell_\theta(X))^\top] \quad (\text{multi-d})$$

$$= \mathbb{E}_\theta \left[\left(\frac{\partial}{\partial \theta} \log f_\theta(X) \right)^2 \right] \quad (\text{uni-d})$$

Si le modèle est **régulier**, alors le score est **centré** et

1. $I(\theta) = \text{Var}(\partial_\theta \ell_\theta(X)) = \mathbb{E} [(\partial_\theta \ell_\theta(X))^2]$
2. $I(\theta) = -\mathbb{E} [\partial_\theta^2 \ell_\theta(X)]$

*. différent du score utilisé dans les modèles génératifs à base de diffusion

Le modèle $(P_\theta)_{\theta \in \Theta}$ qui admet une densité g_θ par rapport à une mesure de référence μ est dit **régulier** si :

- ▶ pour μ -presque tout $x \in E$, l'application $\theta \mapsto g_\theta(x)$ est absolument continue[†] sur Θ ;
- ▶ pour tout $\theta_0 \in \Theta$, il existe $E_0 \subseteq E$ avec $\mu(E \setminus E_0) = 0$ tel que, pour tout $x \in E_0$, l'application $\theta \mapsto \partial_\theta g_\theta(x)$ est continue en θ_0 ;
- ▶ pour tout $\theta \in \Theta$, le score doit admettre un moment d'ordre 2 et l'**information de Fisher**

$$\theta \mapsto I(\theta) = \mathbb{E}_\theta[(\partial_\theta \ell_\theta(X))^2] = \int_E \frac{(\partial_\theta g_\theta(x))^2}{g_\theta(x)} 1_{g_\theta(x) > 0} \mu(dx)$$

doit être continue sur Θ .

✗ Si le support de g_θ dépend de θ , ciao le premier point

✓ Quand pour tout $x \in E$, $\theta \mapsto g_\theta(x)$ est C^1 , deux premiers points automatiquement vérifiés

[†]. f est absolument continue sur un intervalle ouvert I de $\mathbb{R}$, s'il existe f' telle que $f(b) - f(a) = \int_a^b f'(t)dt$

✓ **Loi normale** $X \sim \mathcal{N}(\theta, 1)$.

$$\log f_{\theta}(x) = -\frac{1}{2}(x - \theta)^2 - \frac{1}{2} \log(2\pi) \quad \frac{\partial}{\partial \theta} \log f_{\theta}(x) = (x - \theta)$$

$$I(\theta) = \mathbb{E}_{\theta}[(X - \theta)^2] = 1$$

✓ **Loi exponentielle** $X \sim \mathcal{E}(\theta)$ avec densité $f_{\theta}(x) = \theta e^{-\theta x} 1_{x \geq 0}$.

$$\frac{\partial}{\partial \theta} \log f_{\theta}(x) = \frac{1}{\theta} - x \quad \text{et} \quad I(\theta) = \mathbb{E}_{\theta} \left[\left(\frac{1}{\theta} - X \right)^2 \right] = \frac{1}{\theta^2}$$

✓ **Loi Bernoulli** $X \sim \mathcal{B}(\theta) \in (0, 1)$, alors $\mu(dx) = \delta_0 + \delta_1$ et $g_{\theta}(0) = 1 - \theta$, $g_{\theta}(1) = \theta$. Le modèle est régulier et

$$I(\theta) = \frac{1}{\theta(1 - \theta)}$$

✗ **Loi uniforme** $X \sim \mathcal{U}([0, \theta])$ avec $\theta \in \Theta = \mathbb{R}_{+}^{*}$, $\mu(dx) = 1_{x \geq 0} dx$,
 $g_{\theta}(x) = \frac{1}{\theta} 1_{[0, \theta]}(x) = \frac{1}{\theta} 1_{[x, +\infty[}(\theta)$ ($\theta \mapsto g_{\theta}(x)$ est discontinue en x)
 $\Rightarrow$ Le modèle n'est pas régulier

- ▶ Si $X_1, \dots, X_n$ i.i.d. $\sim P_\theta$ de densité f_θ , alors

$$I_n(\theta) = \text{Var}(\partial_\theta \log f_\theta) = nI(\theta).$$

- ▶ Exemple Gaussien

Plus l'écart-type σ est petit, plus la variable $X \sim (\mu, \sigma^2)$ a des chances de tomber près de la moyenne μ que l'on cherche, donc plus on aura "d'information" sur celle-ci grâce à celle-là : ceci est cohérent avec le fait que $I(\mu) = 1/\sigma^2$. Avec cette interprétation, il est tout aussi logique que $I(\mu)$ ne dépende pas de μ : que la moyenne vaille 0 ou 50, l'information sur cette moyenne apportée par une observation est clairement la même. Ceci est en fait vrai pour tous les modèles de translation réguliers

- ▶ $I(\theta)$ mesure la quantité d'information sur θ contenue dans l'observation de X .
- ▶ Plus $I(\theta)$ est grand, plus on peut espérer estimer θ précisément.

1. Généralités

2. Information de Fisher

3. Borne de Cramér-Rao

4. Efficacité asymptotique et comparaison des estimateurs

Proposition

Soit (f_θ, μ) un modèle régulier, et $\hat{\theta}_n$ un estimateur de θ , alors le risque quadratique satisfait

$$\text{Risque}(\hat{\theta}_n) = \mathbb{E}_\theta \left[(\hat{\theta}_n - \theta)^2 \right] \geq \text{Biais}^2(\hat{\theta}_n, \theta) + \frac{\left(1 + \partial_\theta \text{Biais}(\hat{\theta}_n, \theta)\right)^2}{I_n(\theta)}$$

avec $\text{Biais}(\hat{\theta}_n, \theta) = \mathbb{E}_\theta[\hat{\theta}_n] - \theta$

- ▶ N'est pas utilisable en tant que tel pour comparer des estimateurs
- ▶ Atteindre la borne inf n'indique en rien l'optimalité de l'estimateur

Par exemple, $\hat{\theta} = 0$, on a $\text{Risque}(\hat{\theta}) = \text{Biais}^2(\hat{\theta}, \theta) + \frac{(1 + \partial_\theta \text{Biais}(\hat{\theta}, \theta))^2}{I_n(\theta)}$ mais l'estimateur n'est pas bon

Borne de Cramer-Rao Soient un modèle régulier et $T = T(X_1, \dots, X_n)$ un **estimateur sans biais** de θ . Alors,

$$\text{Risque}(T) = \text{Var}_{\theta}(T) \geq \frac{1}{I_n(\theta)} = \frac{1}{nI(\theta)}.$$

Un estimateur atteignant cette borne est dit **efficace**

Interprétation : aucun estimateur sans biais ne peut avoir une variance inférieure à $1/(nI(\theta))$

👉 précision maximale avec laquelle θ peut être estimé parmi tous les estimateurs sans biais

Remarque : il s'agit en fait de la borne de **Fréchet-Darmois-Cramer-Rao**

Version multi-d : Pour tout estimateur T_n sans biais de $h(\theta)$, pour h une fonction,

$$\text{Var}(T_n) \geq \nabla_{\theta} h(\theta)^{\top} I_n(\theta)^{-1} \nabla_{\theta} h(\theta)$$

Moyenne d'une gaussienne Soit $X_1, \dots, X_n \sim \mathcal{N}(\theta, 1)$, $I(\theta) = 1$ et $I_n(\theta) = n$ La borne de Cramér-Rao donne

$$\mathrm{Var}_\theta(T) \geq \frac{1}{n} \quad \text{et} \quad \mathrm{Var}_\theta(\bar{X}_n) = \frac{1}{n}$$

Donc $\bar{X}_n$ est efficace.

Loi Bernoulli Soient $X_1, \dots, X_n$ des v.a. indépendantes de loi de Bernoulli $\mathcal{B}(\theta)$. Alors $\bar{X}_n$ est efficace.

Loi exponentielle Soit $X_1, \dots, X_n \sim \mathcal{E}(\theta)$, $I(\theta) = \frac{1}{\theta^2}$ et $I_n(\theta) = \frac{n}{\theta^2}$
Borne de Cramér-Rao : $\mathrm{Var}_\theta(T) \geq \frac{\theta^2}{n}$.

Pour $\hat{\theta} = 1/\bar{X}_n$, on a $\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} \mathcal{N}(0, \theta^2)$

Donc $\hat{\theta}$ est *asymptotiquement efficace*.

L'efficacité permet-elle de caractériser l'optimalité d'un estimateur ?

Clairement **non** !

👉 Les estimateurs sans biais, lorsqu'ils existent, ne sont pas nécessairement les plus intéressants en terme d'erreur quadratique

👉 Un estimateur efficace ne peut exister que dans des conditions très particulières et clairement identifiées (estimateur "linéaire" dans un modèle exponentiel)

1. Généralités

2. Information de Fisher

3. Borne de Cramér-Rao

4. Efficacité asymptotique et comparaison des estimateurs

Définition. Soit $\hat{\theta}_n$ une suite d'estimateurs sans biais de θ .
On dit que $\hat{\theta}_n$ est **asymptotiquement efficace** si :

$$n \cdot \text{Var}_{\theta}(\hat{\theta}_n) \xrightarrow{n \rightarrow +\infty} \frac{1}{I(\theta)}$$

Interprétation : l'estimateur atteint asymptotiquement la borne de Cramér-Rao.

Exemple de la loi exponentielle. Soit $X_1, \dots, X_n \sim \mathcal{E}(\theta)$ et $\hat{\theta}_n = 1/\bar{X}_n$.

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} \mathcal{N}(0, \theta^2).$$

Ainsi,

$$n \cdot \text{Var}_{\theta}(\hat{\theta}_n) \longrightarrow \theta^2 = \frac{1}{I(\theta)}.$$

Conclusion : $\hat{\theta}_n$ est *asymptotiquement efficace*.

Théorème. Soit $(f_\theta)_{\theta \in \Theta}$ un modèle régulier d'information de Fisher $I(\theta)$. Soit $\theta_0 \in \Theta$ vérifiant $I(\theta_0) > 0$ et $X_1, \dots, X_n$ des variables aléatoires i.i.d de densité f_{θ_0} . S'il existe un estimateur du maximum de vraisemblance $\hat{\theta}_n$ qui converge vers θ_0 , et s'il existe $h > 0$ tel que

$$\mathbb{E}_{\theta_0} \left[\sup_{\theta_0 - h \leq \theta \leq \theta_0 + h} (\ell'_\theta(X_1))^2 \right] < +\infty,$$

alors

$$\sqrt{n} (\hat{\theta}_n - \theta_0) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, I(\theta_0)^{-1}).$$

Existe-t-il un estimateur optimal, et si oui en quel sens ?

- ☞ D'un point de vue **asymptotique** dans le cadre des modèles **réguliers**, c'est l'estimateur du maximum de vraisemblance qui répond au problème (sous les réserves qui s'imposent : existence d'un EMV consistant, hypothèse de domination, non-nullité de l'information de Fisher)
- ☞ Encore faut-il pouvoir le calculer, ce qui n'est pas toujours chose facile
- ☞ L'EMV peut souffrir d'un manque de robustesse aux données aberrantes ou à une mauvaise spécification du modèle
- ☞ En **non-asymptotique**, la comparaison pratique des estimateurs se fait via le **risque quadratique**