8.Comparació d'estimadors

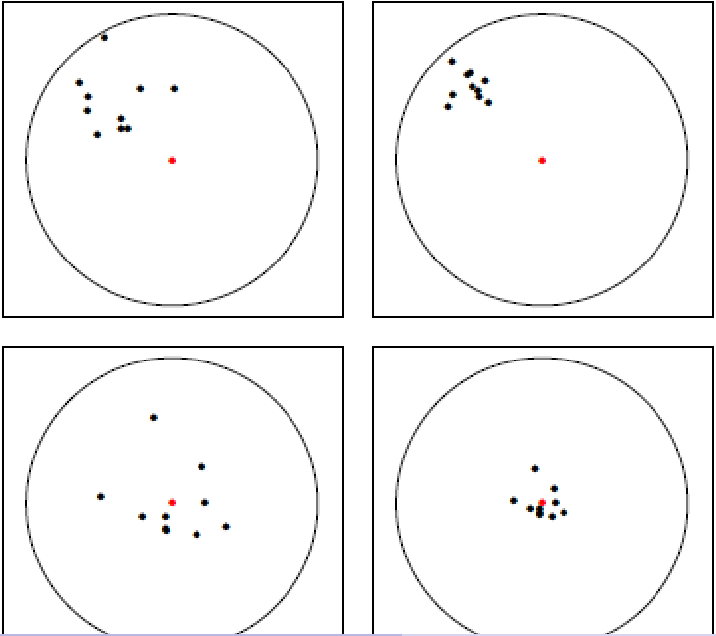
Estadística

Grau en Matemàtiques

Josep A. Sanchez
Dept. Estadística i I.O.(UPC)

UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH

Propietats dels estimadors: Exactitud i Precisió



Biaix d'un estimador

- Sigui W un estimador del paràmetre θ
- Es defineix com a **biaix** de l'estimador: $B_\theta(W) = E(W) - \theta$
- El biaix pot ser positiu (sobrestimació) o negatiu (infraestimació)
- Els estimadors W amb $B_\theta(W) = 0$ són estimadors no esbiaxats (*unbiased*)

Error Quadràtic Mig

- **Error Quadràtic Mig** d'un estimador (*Mean Square Error, MSE*)
$$MSE_\theta(W) = E_\theta[(W-\theta)^2]$$

- Es una mesura de la distància entre l'estimador i el valor del paràmetre que es vol estimar

$$E_\theta[(W-\theta)^2] = E_\theta[(W - E_\theta(W) + E_\theta(W) - \theta)^2] =$$

$$E_\theta[(W-E_\theta(W))^2]-2E_\theta[(W-E_\theta(W))(E_\theta(W)-\theta)]+E_\theta[(E_\theta(W)-\theta)^2]$$
$$= E_\theta[(W-E_\theta(W))^2] + (E_\theta(W)-\theta)^2 = V(W) + B_\theta(W)^2$$

Millor estimador sense biaix

- W^* és un **millor estimador sense biaix** o **estimador sense biaix uniformement de mínima variància** de $\tau(\theta)$ si:
  - $E(W^*) = \tau(\theta) \quad \forall \theta \in \Theta$
  - Si W és un altre estimador sense biaix de $\tau(\theta)$ llavors, $V(W^*) \leq V(W)$
- En anglès es coneixen com *UMVUE* (*Uniform Minimum Variance Unbiased Estimator*)

Informació de Fisher

- Sigui X una variable aleatòria amb densitat $f(x|\theta)$
- Es considera que el suport de θ és un interval obert a la recta real
- Definició: **Informació de Fisher de θ continguda en la variable X**:

$$I_X(\theta) = E\left[\left(\frac{\partial}{\partial\theta}\log f_X(X|\theta)\right)^2\right]$$

Informació de Fisher d'una mostra

- Sigui $X_1, \ldots, X_n$ una m.a.s. amb

$$L(\theta; \underset{\sim}{X}) = f(\underset{\sim}{X}|\theta) = \prod_{i=1}^{n} f(x_i|\theta)$$

- La **informació de Fisher** continguda en la mostra pel paràmetre θ es defineix en base a la densitat conjunta:

$$I_{\underset{\sim}{X}}(\theta) = E\left[\left(\frac{\partial}{\partial\theta} \log f_{\underset{\sim}{X}}(\underset{\sim}{X}|\theta)\right)^2\right]$$

Lema 1

- Sigui $X_1, \ldots, X_n$ una m.a.s, llavors,

$$I_{\underset{\sim}{X}}(\theta) = nI_X(\theta)$$

- **Demostració:**

$$\frac{\partial}{\partial\theta}\log f(\underset{\sim}{X}|\theta) = \frac{\partial}{\partial\theta}\log\prod_{i=1}^{n} f(x_i|\theta) = \sum_{i=1}^{n}\frac{\partial}{\partial\theta}\log f(x_i|\theta)$$

Llavors,

$$I_{\underset{\sim}{X}}(\theta) = E\left[\left(\frac{\partial}{\partial\theta}\log f_{\underset{\sim}{X}}(\underset{\sim}{X}|\theta)\right)^2\right] = E\left[\left(\sum_{i=1}^{n}\frac{\partial}{\partial\theta}\log f(x_i|\theta)\right)^2\right] =$$

$$= E\left[\sum_{i=1}^{n}\left(\frac{\partial}{\partial\theta}\log f(x_i|\theta)\right)^2 + \sum_{i\neq j}\left(\frac{\partial}{\partial\theta}\log f(x_i|\theta)\frac{\partial}{\partial\theta}\log f(x_j|\theta)\right)\right] =$$

$$= nE\left[\left(\frac{\partial}{\partial\theta}\log f(x|\theta)\right)^2\right] + \sum_{i\neq j} E\left[\frac{\partial}{\partial\theta}\log f(x_i|\theta)\right] E\left[\frac{\partial}{\partial\theta}\log f(x_j|\theta)\right]$$

- **Demostració** (cont.) Només queda demostra que $E\left[\frac{\partial}{\partial\theta}\log f(x_i|\theta)\right]=0$ i haurem demostrat que $I_{\underset{\sim}{X}}(\theta)=nI_X(\theta)$

$$E\left[\frac{\partial}{\partial\theta}\log f(x_i|\theta)\right]=\int\left(\frac{\partial}{\partial\theta}\log f(x_i|\theta)\right)f(x_i|\theta)dx_i=\int\left(\frac{\frac{\partial}{\partial\theta}f(x_i|\theta)}{f(x_i|\theta)}\right)f(x_i|\theta)dx_i=$$

$$=\int\left(\frac{\partial}{\partial\theta}f(x_i|\theta)\right)dx_i=\frac{\partial}{\partial\theta}\int f(x_i|\theta)dx_i=\frac{\partial}{\partial\theta}1=0$$

- Necessitem la condició de regularitat que contempla que la derivada respecte el paràmetre i la integral de la variable conmuten:
- Hipòtesi H1: $\frac{\partial}{\partial\theta}\int f(x|\theta)dx_i=\int\frac{\partial}{\partial\theta}f(x|\theta)dx$

Funció Score

- La funció $S(\theta; \underset{\sim}{X}) = \frac{\partial}{\partial\theta} \log f(\underset{\sim}{X}|\theta)$ es denomina **funció Score**
- Per buscar els punts singulars, busquem els valors de θ on la funció de Score s'anul · la.
- En l'anterior demostració hem vist que $E(S(\theta; \underset{\sim}{X})) = 0$
- D'altra banda, la variància d'aquesta funció és la informació de Fisher:
$$I_{\underset{\sim}{X}}(\theta) = E(S(\theta; \underset{\sim}{X})^2) = V(S(\theta; \underset{\sim}{X}))$$
- La informació de Fisher mesura la curvatura de la funció $\log L(\theta; \underset{\sim}{X})$: els models amb molta curvatura tenen molta informació (més precisió en l'estimació)

Lema 2

Sota la hipòtesis (H2): $\frac{\partial^2}{\partial\theta^2}\int f(x|\theta)dx_i = \int \frac{\partial^2}{\partial\theta^2} f(x|\theta)dx$

Es pot obtenir una expressió alternativa per a la informació de Fisher pel paràmetre θ:

- Continguda a la variable x:

$$I_X(\theta) = E\left[-\frac{\partial^2}{\partial\theta^2}\log f(X|\theta)\right]$$

- **Demostració:**

$$I_X(\theta) + E\left[\frac{\partial^2}{\partial\theta^2}\log(f(x|\theta))\right] =$$

$$= \int \left[\left(\frac{\partial}{\partial\theta}\log(f(x|\theta))\right)^2 + \left(\frac{\partial^2}{\partial\theta^2}\log(f(x|\theta))\right)\right] f(x|\theta)dx$$

$$= \int \left(\frac{\partial}{\partial\theta}\log(f(x|\theta))\right) \frac{\frac{\partial}{\partial\theta}f(x|\theta)}{f(x|\theta)} f(x|\theta)dx + \int \left(\frac{\partial^2}{\partial\theta^2}\log(f(x|\theta))\right) f(x|\theta)dx$$

$$= \int \frac{\partial}{\partial\theta}\log(f(x|\theta))\frac{\partial}{\partial\theta}f(x|\theta)dx + \int \left(\frac{\partial^2}{\partial\theta^2}\log(f(x|\theta))\right) f(x|\theta)dx$$

$$= \int \frac{\partial}{\partial\theta}\left[\left(\frac{\partial}{\partial\theta}\log(f(x|\theta))\right) f(x|\theta)\right] dx = \int \frac{\partial}{\partial\theta}\left[\left(\frac{\frac{\partial}{\partial\theta}f(x|\theta)}{f(x|\theta)}\right) f(x|\theta)\right] dx$$

$$= \int \frac{\partial}{\partial\theta}\left[\frac{\partial}{\partial\theta}f(x|\theta)\right] dx = \frac{\partial^2}{\partial\theta^2}\int f(x|\theta)dx = 0$$

Lema 2

- Generalització: Informació pel paràmetre continguda en una mostra $\underset{\sim}{X}$:

$$I_{\underset{\sim}{X}}(\theta) = E\left[-\frac{\partial^2}{\partial\theta^2}\log f(\underset{\sim}{X}|\theta)\right] = E\left[-\frac{\partial^2}{\partial\theta^2}\log L(\theta;\underset{\sim}{X})\right]$$

Desigualtat de Cramer-Rao

Teorema:

Sigui $X_1, \ldots, X_n$ m.a.s amb funció de densitat conjunta $f(\underset{\sim}{X}|\theta)$ i sigui $W(\underset{\sim}{X})$ un estimador sense biaix de $\tau(\theta)$ amb

- $V(W(\underset{\sim}{X})) < \infty$
- $\frac{\partial}{\partial\theta} E[W(\underset{\sim}{X})] = \int \frac{\partial}{\partial\theta} W(\underset{\sim}{X}) f(\underset{\sim}{X}|\theta) dx$

Llavors,

$$V(W(\underset{\sim}{X})) \geq \frac{\left(\frac{d}{d\theta}\tau(\theta)\right)^2}{E_\theta\left[\left(\frac{\partial}{\partial\theta}\log f(\underset{\sim}{X}|\theta)\right)^2\right]} = \frac{\tau'(\theta)^2}{I_{\underset{\sim}{X}}(\theta)}$$

Desigualtat de Cramer-Rao

Demostració: (Desigualtat de Cauchy-Schwartz: $(x'y) \leq \|x\|^2 \|y\|^2$)

Estadísticament,

$$Cov(X, Y)^2 \leq V(X)V(Y) \Rightarrow V(X) \geq \frac{Cov(X, Y)^2}{V(Y)}$$

Considerem $X \equiv W(\underset{\sim}{X})$ i $Y \equiv \frac{\partial}{\partial \theta} \log(f(\underset{\sim}{X}|\theta))$

Notem que

$$\frac{d}{d\theta} E(W(X)) = \int W(x) \frac{\partial}{\partial \theta} f(x|\theta) dx = E\left(W(X) \frac{\frac{\partial}{\partial \theta} f(X|\theta)}{f(X|\theta)} \right)$$

$$= E\left(W(X) \frac{\partial}{\partial \theta} \log(f(X|\theta)) \right)$$

Aquesta expressió correspon a la covariància entre les dues variables (Def: $Cov(X, Y) = E(XY) - E(X)E(Y)$), ja que

$$E\left(\frac{\partial}{\partial \theta} \log(f(X|\theta)) \right) = \int \frac{\frac{\partial}{\partial \theta} f(x|\theta)}{f(X|\theta)} f(X|\theta) dx = \frac{\partial}{\partial \theta} \int f(x|\theta) dx = 0$$

Demostració(cont.):

$$Cov\left(W(X), \frac{\partial}{\partial\theta}\log(f(X|\theta))\right) = E\left(W(X)\frac{\partial}{\partial\theta}\log(f(X|\theta))\right) = \frac{d}{d\theta}E(W(X))$$

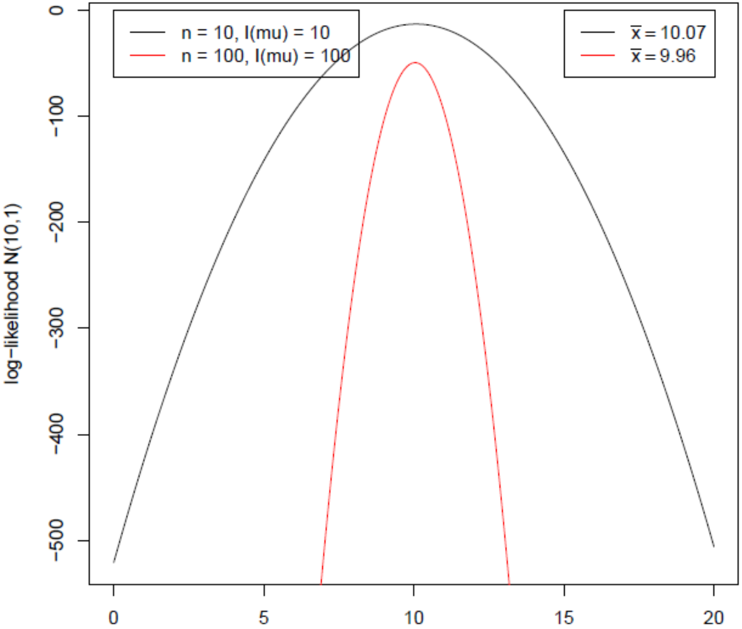
També,

$$V\left(\frac{\partial}{\partial\theta}\log(f(X|\theta))\right) = E\left(\left(\frac{\partial}{\partial\theta}\log(f(X|\theta))\right)^2\right) \quad \text{ja que} \quad E\left(\frac{\partial}{\partial\theta}\log(f(X|\theta))\right) = 0$$

Per tant,

$$V(W(X)) \geq \frac{\left(\frac{d}{d\theta}E(W(X)\right)^2}{E_\theta\left[\left(\frac{\partial}{\partial\theta}\log f(X|\theta)\right)^2\right]} = \frac{\tau'(\theta)^2}{I_X(\theta)}$$

Interpretació de la Informació de Fisher



Notes

- En moltes situacions $\tau(\theta) = \theta$ i la cota de Cramer-Rao correspon a l'invers de la Informació de Fisher:

$$V(W(X)) \geq \frac{1}{I_{\underset{\sim}{X}}(\theta)} = \frac{1}{nI_X(\theta)}$$

- Si un estimador sense biaix té una variància igual a la cota de Cramer-Rao, llavors és UMVUE
- Totes les distribucions de la família exponencial compleixen les condicions del teorema de Cramer-Rao
- Un estimador sense biaix que assoleix la cota de Cramer-Rao s'anomena **eficient**
- Tot estimador eficient és UMVUE. El recíproc no té perquè ser cert.

Notes

- La informació de Fisher pels paràmetres θ i $\tau(\theta)$ estan relacionades:

$$I_n(\theta) = E\left(\left(\frac{\partial l}{\partial \theta}\right)^2\right) = E\left(\left(\frac{\partial l}{\partial \tau}\frac{\partial \tau}{\partial \theta}\right)^2\right) =$$

$$= \tau'(\theta)^2 E\left(\left(\frac{\partial l}{\partial \tau}\right)^2\right) = \tau'(\theta)^2 I_n(\tau(\theta))$$

Per tant,

$$I_n(\tau(\theta)) = \frac{I_n(\theta)}{\tau'(\theta)^2}$$

Assoliment de la Cota de Cramer-Rao

Sigui $X_1, \ldots, X_n$ una m.a.s. amb densitat $f(x|\theta)$, tal que $f(x|\theta)$ satisfà les condicions del teorema de Cramer-Rao.

Sigui $L(\theta|X) = \prod_{i=1}^{n} f(x_i|\theta)$ la funció de versemblança i sigui $W(X)$ un estimador sense bieax de $\tau(\theta)$.

Llavors $W(X)$ assoleix la cota de Cramer-Rao si i només si,

$$a(\theta)(W(X) - \tau(\theta)) = \frac{\partial}{\partial\theta} \log(L(\theta|X))$$

per certa funció $a(\theta)$