

Consistenza dello stimatore di massima verosimiglianza

Oss

Per definizione lo stimatore di massima verosimiglianza è non distorto per costruzione

Distanza di Kullback-Leibler

Def

Siano f, g funzioni di densità o probabilità

La distanza di Kullback-Leibler tra f e g è definita da

$$D(f, g) = \int \log\left(\frac{f(x, \theta)}{g(x, \theta)}\right) f(x, \theta) dx$$

se f è continua rispetto a g . Altrimenti viene posta a $+\infty$

Per quanto riguarda il caso discreto abbiamo

$$D(p, q) = \sum_{x_i} \log\left(\frac{p(x_i, \theta)}{q(x_i, \theta)}\right) p(x_i, \theta)$$

Proprietà

$D(\cdot, \cdot)$ non è una metrica

non vale la disuguaglianza triangolare né la simmetria

① $D(f, g) \neq D(g, f)$

② Se $f = g \Rightarrow D(f, g) = 0$

③ Se $f \neq g \Rightarrow D(f, g) > 0$

Dim ③

def di valore atteso di una densità

$$\int g(x, \theta) dx = 1$$

Integrale di una densità

$$D(f, g) = E\left[\log\left(\frac{f(x)}{g(x)}\right)\right] = -E\left[\log\left(\frac{g(x)}{f(x)}\right)\right] \geq -\log\left(E\left[\frac{g(x)}{f(x)}\right]\right) = -\log\left(\int \frac{g(x, \theta)}{f(x, \theta)} f(x, \theta) dx\right) = -\log(1) = 0$$

dis. di Jensen

$$E[f(x)] \geq f(E[x])$$

Ricapitolando abbiamo

$$D(f, g) = -E\left[\log\left(\frac{g(x)}{f(x)}\right)\right] \geq 0$$

Consistenza dello stimatore

Visto che lo stimatore è invariante rispetto alle trasformazioni lineari possiamo riscrivere

$$\log L = \sum_{i=1}^n \log f(x_i, \theta)$$

come

$$M_n(\theta) = \frac{1}{n} \sum_{i=1}^n \log f(x_i, \theta) - \frac{1}{n} \sum_{i=1}^n \log f(x_i, \theta_0) = \frac{1}{n} \sum_{i=1}^n \log \frac{f(x_i, \theta)}{f(x_i, \theta_0)}$$

vero valore di θ

Oss

Essendo θ_0 il vero valore di θ , la trasformazione non modifica il valore di $\tilde{\theta} = \arg \max M_n(\theta)$

Sotto alcune condizioni di regolarità vale la legge dei grandi numeri

$$\frac{1}{n} \sum_{i=1}^n \log \frac{f(x_i, \theta)}{f(x_i, \theta_0)} \xrightarrow{P} E \left[\log \frac{f(x, \theta)}{f(x, \theta_0)} \right]$$

Abbiamo poi che

$$E \left[\log \frac{f(x, \theta)}{f(x, \theta_0)} \right] = \int \log \left(\frac{f(x, \theta)}{f(x, \theta_0)} \right) f(x, \theta_0) dx = -D(f_{\theta_0}; f_{\theta})$$

Oss

Possiamo notare che $M_n(\theta)$ converge a $-D(f_{\theta_0}, f_{\theta})$, che per def è sempre negativo tranne per $\theta_0 = \theta$

dove vale 0

Visto che lo stimatore di massima verosimiglianza è definito come il valore di θ che massimizza $M_n(\theta)$

allora questo convergerà al valore che massimizza $-D(f_{\theta_0}, f_{\theta})$ ovvero $\theta = \theta_0$.

Teo

$$\frac{1}{n} \sum_{i=1}^n \log \frac{S(X_i; \theta)}{S(X_i; \theta_0)}$$

Sia $M_n(\theta)$ definito come sopra

$$\text{Sia } M(\theta) = -D(\theta, \theta)$$

Assumendo valgono

a) Legge dei grandi numeri uniformi

$$\sup |M_n(\theta) - M(\theta)| = o_p(1) \quad \text{per } n \rightarrow \infty$$

$$\hookrightarrow \sup |M_n(\theta) - M(\theta)| \xrightarrow{p} 0$$

b) Convessità

$\forall \varepsilon > 0 \exists \eta_\varepsilon > 0$ tale che

$$|\theta_0 - \theta| > \varepsilon \implies M(\theta_0) - M(\theta) > \eta_\varepsilon$$

\hookrightarrow se \exists dei valori θ t.c. $\theta_0 - \theta = 0$ il problema della consistenza sarebbe irrisolvibile

Allora

$$\tilde{\theta}_n \xrightarrow{p} \theta_0$$

Dim

Ricordando che se

$$A \implies B \quad \text{Allora } P_r(A) \leq P_r(B)$$

Vogliamo dimostrare che

$$P_r \{ |M(\theta_0) - M_n(\tilde{\theta}_n)| > \eta_\varepsilon \} \rightarrow 0$$

Grazie alla condizione di convessità abbiamo che

$$P_r \{ |\theta_0 - \theta| > \varepsilon \} \leq P_r \{ |M(\theta_0) - M(\theta)| > \eta_\varepsilon \}$$

Per def di stimatore di massima verosimiglianza abbiamo che

$$M_n(\theta_0) - M_n(\tilde{\theta}_n) > 0$$

$\tilde{\theta}_n$ massimizza M_n

Quindi abbiamo che

$$\begin{aligned} M(\theta_0) - M(\tilde{\theta}_n) &= M(\theta_0) - M_n(\theta_0) + M_n(\theta_0) - M(\tilde{\theta}_n) \leq M(\theta_0) - M_n(\theta_0) + M_n(\tilde{\theta}_n) - M(\tilde{\theta}_n) \leq \\ &\leq |M(\theta_0) - M_n(\theta_0)| + |M(\tilde{\theta}_n) - M_n(\tilde{\theta}_n)| \end{aligned}$$

Possiamo quindi scrivere che

$$P_r \{ M(\theta_0) - M(\tilde{\theta}_n) > \eta_\varepsilon \} \leq P_r \{ |M(\theta_0) - M_n(\theta_0)| > \frac{\eta_\varepsilon}{2} \} + P_r \{ |M(\tilde{\theta}_n) - M_n(\tilde{\theta}_n)| > \frac{\eta_\varepsilon}{2} \}$$

Allora

$$M_n = \frac{1}{n} \sum_{i=1}^n X_i^{(-)} \quad \text{e} \quad M = E[X_1]$$

① $\rightarrow 0$ per la legge dei grandi numeri standard

② $\rightarrow 0$ per la legge dei grandi numeri uniforme posta in Hp