

## Consistenza dello stimatore di Massima Verosimiglianza

Oss

Per definizione lo stimatore di massima verosimiglianza e' non distorto per costruzione

## Distanza di Kullback-Leibler

Def

Siano  $f, g$  funzioni di densità o probabilità

La distanza di Kullback-Leibler fra  $f$  e  $g$  e' definita da

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

Se  $f$  e' 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

①  $D(f, g) \neq D(g, f)$

② Se  $f = g \Rightarrow D(f, g) = 0$

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

$D(\cdot, \cdot)$  non e' una metrica  $\rightsquigarrow$  non vale la disegualanza triangolare he' la simmetria

Integrale di una densità

Dim ③

def di valore atteso di una densità

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

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

dis. d. Jensen

$$\mathbb{E}[S(x)] \geq S(\mathbb{E}[x])$$

Riassumendo abbiamo

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

## 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

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

vero valore di  $\theta$

Oss

Essendo  $\theta_0$  il vero valore di  $\theta$ , la trasformazione non modifica il valore di  $\tilde{\theta} = \arg \max M_h(\theta)$

Sotto alcune condizioni di regolarità vale la legge dei grandi numeri

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

Abbiamo poi che

$$E \left[ \log \frac{f(x_i, \theta)}{f(x_i, \theta_0)} \right] = \int \log \left( \frac{f(x_i, \theta)}{f(x_i, \theta_0)} \right) S(x_i, \theta_0) dx = -D(S_{\theta_0}, S_\theta)$$

Oss

Possiamo notare che  $M_h(\theta)$  converga a  $-D(S_{\theta_0}, S_\theta)$ , che per def è sempre negativa tranne per  $\theta_0 = \theta$

dove vale 0

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

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

Teo

$$\int \frac{1}{n} \sum_{i=1}^n \log \frac{s(x_i; \theta)}{s(x_i; \hat{\theta}_0)}$$

Sia  $M_n(\theta)$  definito come sopra

Sia  $M(\theta) = -I(\theta, \vartheta)$

Assumendo valgono

a) Legge dei grandi numeri uniforme

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

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

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  $\theta$  t.c.  $\theta_0 - \theta = 0$  il problema della consistenza sarebbe irrisolvibile

Allora

$$\tilde{\theta}_{nl} \xrightarrow{P} \theta_0$$

Dim

Ricordando che se

$$\lambda \Rightarrow \beta \text{ Allora } P_r(\lambda) \leq P_r(\beta)$$

Vogliamo dimostrare che

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

Grazie alla condizione di convessità abbiamo che

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

Per def di stimatore di massima verosimiglianza abbiamo che

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

Quindi abbiamo che

$\tilde{\theta}_{nl}$  Massimizza  $M_n$

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

Possiamo quindi scrivere che

A

B

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

Allora

$$M_n = \frac{1}{n} \sum_{i=1}^n \log(\cdot) \quad e \quad M = E[M_n]$$

(A)  $\rightarrow 0$  per la legge dei grandi numeri standard

(B)  $\rightarrow 0$  per la legge dei grandi numeri uniforme posta in Hp