



UNIVERSITÀ
DEGLI STUDI DI TRIESTE



Dipartimento di scienze economiche,
aziendali, matematiche e statistiche
"Bruno de Finetti"

Statistica (c.p.)

2. Proprietà dello SMV

Francesco Pauli

DEAMS

Università di Trieste

A.A. 2017/2018

Problema regolare di stima

Definizione

Un modello $(\mathcal{Y}, p_\theta, \Theta)$ costituisce un **problema regolare di stima** se

- 1 il modello è identificabile;
- 2 Θ è un aperto in \mathbb{R}^d ;
- 3 il supporto di $p_\theta(y)$ non dipende da θ ;
- 4 si ha

$$\int \frac{\partial}{\partial \theta} p_\theta(y) dy = \frac{d}{d\theta} \int p_\theta(y) dy$$

$$\int \frac{\partial^2}{\partial \theta^2} p_\theta(y) dy = \frac{d^2}{d\theta^2} \int p_\theta(y) dy$$

$$\int \frac{\partial^2}{\partial \theta \partial \theta^\top} p_\theta(y) dy = \frac{d^2}{d\theta d\theta^\top} \int p_\theta(y) dy$$

Indice

- 1 Precisione dello SMV
- 2 Informazione
- 3 Proprietà dello SMV
- 4 Proprietà asintotiche dello SMV

Precisione dello stimatore e verosimiglianza

Lo SMV realizza il massimo di L , gli altri valori possibili possono essere confrontati con esso sulla base di quanto diminuisce la verosimiglianza

$$\theta \text{ è } \frac{L(\theta)}{L(\hat{\theta})} \text{ volte meno attendibile di } \hat{\theta}$$

Conviene ragionare con la log-verosimiglianza, sicché il confronto si basa su

$$\log \frac{L(\theta)}{L(\hat{\theta})} = l(\theta) - l(\hat{\theta})$$

Queste quantità permettono di confrontare $\hat{\theta}$ con uno specifico θ vorremmo però, accanto a queste misure specifiche $\hat{\theta}$ contro θ , una misura generica, $\hat{\theta}$ contro $\Theta - \hat{\theta}$, diciamo.

Precisione e curvatura

Approssimiamo $l()$, purché sufficientemente regolare, con uno sviluppo di Taylor in $\hat{\theta}$

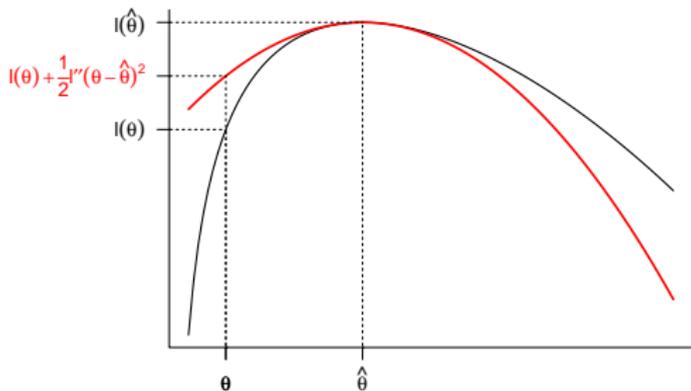
$$l(\theta) \approx l(\hat{\theta}) + l'_*(\hat{\theta})(\theta - \hat{\theta}) + \frac{1}{2}l''(\hat{\theta})(\theta - \hat{\theta})^2$$

assumendo $l'_*(\hat{\theta}) = 0$

$$\approx l(\hat{\theta}) + \frac{1}{2}l''(\hat{\theta})(\theta - \hat{\theta})^2$$

dove $-l''(\hat{\theta})$ prende il nome di informazione osservata, che indichiamo con $J(\hat{\theta})$, sicché

$$\approx l(\hat{\theta}) - \frac{1}{2}J(\hat{\theta})(\theta - \hat{\theta})^2$$



Quanto è buona questa approssimazione?

La bontà dell'approssimazione può essere valutata derivando l'espressione

$$l(\theta) \approx l(\hat{\theta}) - \frac{1}{2} J(\hat{\theta})(\theta - \hat{\theta})^2$$

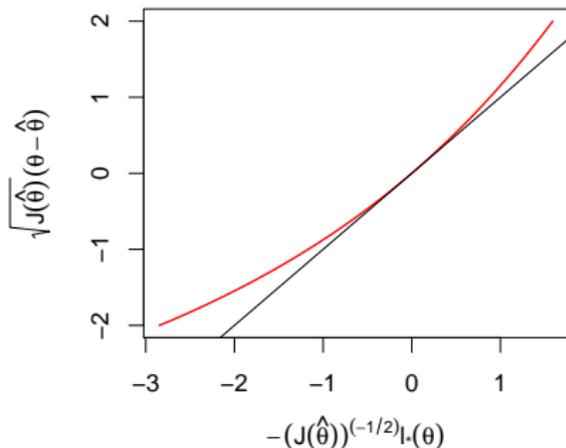
$$l_*(\theta) \approx -J(\hat{\theta})(\theta - \hat{\theta})$$

cioè

$$-(J(\hat{\theta}))^{-1/2} l_*(\theta) \approx \sqrt{J(\hat{\theta})}(\theta - \hat{\theta})$$

Si disegna quindi un grafico con le due quantità sugli assi, se l'approssimazione è buona il grafico è vicino alla bisettrice del I e III quadrante.

Ha senso la verifica per $\sqrt{J(\hat{\theta})}(\theta - \hat{\theta}) \in [-2, 2]$.



Normale, varianza nota

Nel caso della normale con varianza nota si ha

$$l(\theta) = -\frac{n}{2\sigma^2}(\bar{y} - \theta)^2$$

e quindi

$$l^*(\theta) = \frac{n}{\sigma^2}(\bar{y} - \theta)$$

e

$$l''(\theta) = -\frac{n}{\sigma^2}$$

l'informazione è pertanto n/σ^2 e l'approssimazione è

$$l(\theta) = l(\hat{\theta}) - \frac{1}{2} \frac{n}{\sigma^2} (\theta - \hat{\theta})^2$$

che coincide (essendo $\hat{\theta} = \bar{y}$) con la verosimiglianza.

Normale, media nota

Nel caso della normale con media nota ($\mu = 0$)

$$l(\sigma^2) = -\frac{n}{2} \log(\sigma^2) - \frac{n}{2\sigma^2} \sum_i y_i^2$$

e quindi

$$l^*(\sigma^2) = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_i y_i^2 \Rightarrow \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n y_i^2$$

$$l''(\sigma^2) = \frac{n}{2\sigma^4} - \frac{1}{\sigma^6} \sum_i y_i^2$$

$$J(\sigma^2) = -\frac{n}{2\sigma^4} + \frac{n}{\sigma^6} \sum_i y_i^2$$

Normale, media nota

calcolata in $\hat{\sigma}^2 = \sum_i y_i^2/n$ è pari a

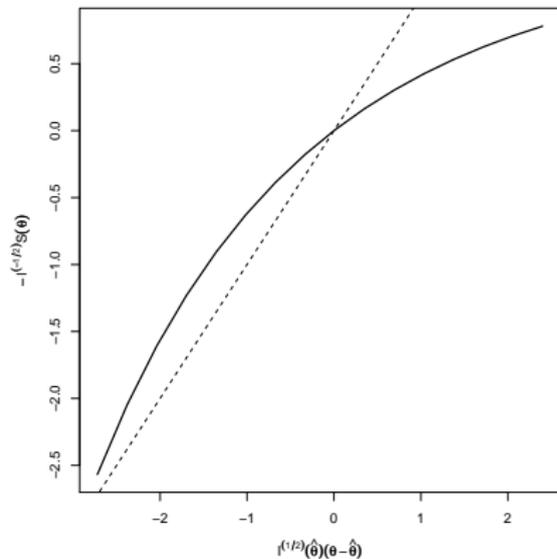
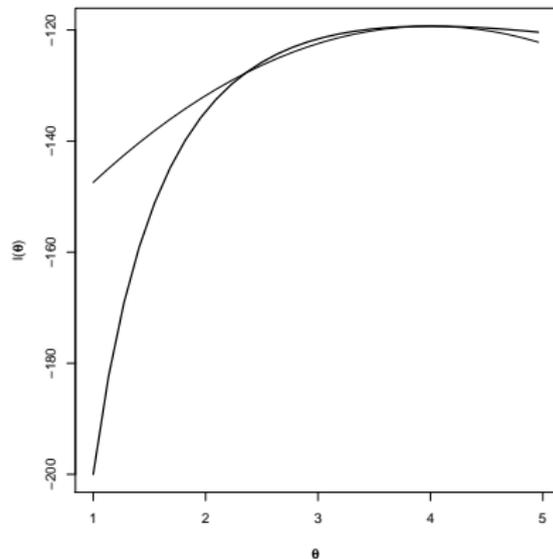
$$J(\hat{\sigma}^2) = -\frac{n}{2\hat{\sigma}^4} + \frac{n}{\hat{\sigma}^4} = \frac{n}{2\hat{\sigma}^4}$$

e quindi l'approssimazione è

$$l(\sigma^2) = l(\hat{\sigma}^2) - \frac{1}{2} \frac{n}{2\hat{\sigma}^4} (\sigma^2 - \hat{\sigma}^2)^2.$$

Normale, media nota

Verosimiglianza e approssimazione quadratica, grafico tra punteggio e parametro standardizzati.



Binomiale

Sia $S = \sum_{i=1}^n y_i$ il numero di successi

$$l(\theta) = S \log \theta + (n - S) \log(1 - \theta)$$

e quindi

$$l^*(\theta) = \frac{S}{\theta} - \frac{n - S}{1 - \theta}$$

e

$$l''(\theta) = -\frac{S}{\theta^2} - \frac{n - S}{(1 - \theta)^2}$$

in $\hat{\theta} = S/n$ si ha

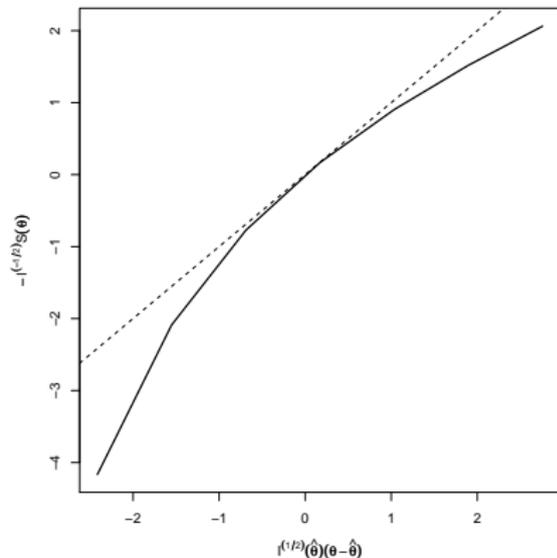
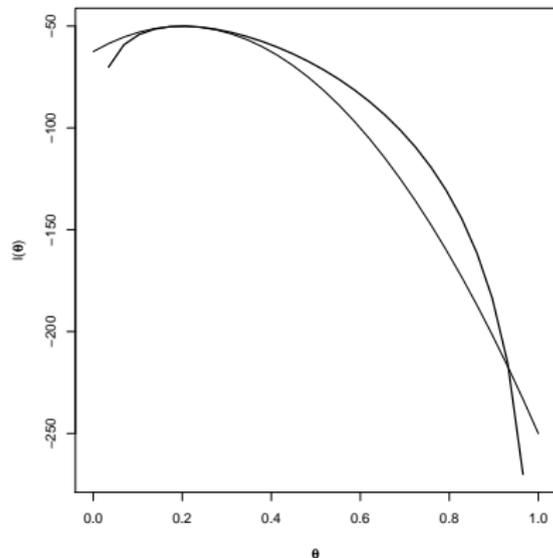
$$J(\hat{\theta}) = \frac{n}{\hat{\theta}(1 - \hat{\theta})}$$

e quindi l'approssimazione è

$$l(\theta) = l(\hat{\theta}) - \frac{1}{2} \frac{n}{\hat{\theta}(1 - \hat{\theta})} (\theta - \hat{\theta})^2.$$

Binomiale

Verosimiglianza e approssimazione quadratica, grafico tra punteggio e parametro standardizzati.



Indice

- 1 Precisione dello SMV
- 2 Informazione**
- 3 Proprietà dello SMV
- 4 Proprietà asintotiche dello SMV

Informazione osservata

Definizione

Si dice *informazione osservata*

$$J(\hat{\theta}) = -l''(\hat{\theta}).$$

Nel caso multivariato J è una matrice quadrata di dimensione pari alla dimensione di θ

$$J(\hat{\theta}) = -\frac{d^2}{d\theta d\theta^T} l(\hat{\theta}) \Big|_{\theta=\hat{\theta}}$$

Altra informazione

L'informazione osservata è legata a un particolare campione

$$J(\hat{\theta}) = -l''(\hat{\theta}; y)$$

cioè $J(\hat{\theta})$ è la curvatura in $\hat{\theta}$ della particolare funzione di verosimiglianza associata al campione (osservato) y .

È utile però definire un analogo di J che non sia legato ad un particolare campione, questo si può ottenere considerando una media sui possibili campioni, si perviene allora a

Informazione attesa

Definizione

Si dice *informazione attesa di Fisher*

$$\mathcal{I}(\theta) = E_{\theta}(-l''(\theta)).$$

Nel caso multivariato si ha la matrice di informazione attesa

$$\mathcal{I}(\theta) = -E_{\theta} \left(\frac{\partial^2}{\partial \theta \partial \theta^T} l(\theta) \right)$$

(dove si intende che la speranza matematica è fatta rispetto alla distribuzione $f(y; \theta)$).

L'informazione attesa è una misura media delle curvature delle verosimiglianze associate ai possibili campioni per quel particolare modello.

Informazione osservata e attesa

- l'informazione osservata è più specifica per il particolare problema che si sta affrontando,
- l'informazione attesa misura quanto difficile è stimare un parametro in relazione al particolare modello (non al particolare modello e campione)

Informazione per il modello normale

Sia $Y_1, \dots, Y_n \sim IID\mathcal{N}(\mu, \sigma^2)$, si ha allora il vettore punteggio

$$\frac{\partial l}{\partial \mu} = \frac{n}{\sigma^2}(\bar{y} - \mu)$$

$$\frac{\partial l}{\partial(\sigma^2)} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_i (y_i - \mu)^2$$

e le derivate seconde

$$\frac{\partial^2 l}{\partial \mu^2} = -\frac{n}{\sigma^2}$$

$$\frac{\partial^2 l}{\partial(\sigma^2)^2} = \frac{n}{2\sigma^4} - \frac{1}{\sigma^6} \sum_i (y_i - \mu)^2$$

$$\frac{\partial l}{\partial \mu \partial(\sigma^2)} = \frac{\partial l}{\partial(\sigma^2) \partial \mu} = -\frac{n}{\sigma^4}(\bar{y} - \mu)$$

Informazione per il modello normale

L'hessiano è quindi la matrice

$$-\frac{d^2}{d\theta d\theta^T} l(\theta) = \begin{bmatrix} \frac{n}{\sigma^2} & \frac{n}{\sigma^4}(\bar{y} - \mu) \\ \frac{n}{\sigma^4}(\bar{y} - \mu) & \frac{n}{2\sigma^4} - \frac{1}{\sigma^6} \sum_i (y_i - \mu)^2 \end{bmatrix}$$

otteniamo allora la matrice di informazione osservata $J(\hat{\theta})$ sostituendo ai parametri le SMV cioè

$$\hat{\mu} = \bar{y}; \quad \hat{\sigma}^2 = \frac{1}{n} \sum_i (y_i - \bar{y})^2$$

e

$$J(\hat{\theta}) = \begin{bmatrix} \frac{n}{\hat{\sigma}^2} & 0 \\ 0 & \frac{n}{2\hat{\sigma}^4} \end{bmatrix}$$

Informazione per il modello normale

L'hessiano è quindi la matrice

$$-\frac{d^2}{d\theta d\theta^T} l(\theta) = \begin{bmatrix} \frac{n}{\sigma^2} & \frac{n}{\sigma^4}(\bar{y} - \mu) \\ \frac{n}{\sigma^4}(\bar{y} - \mu) & \frac{n}{2\sigma^4} - \frac{1}{\sigma^6} \sum_i (y_i - \mu)^2 \end{bmatrix}$$

Si ottiene invece la matrice di informazione attesa calcolando le speranze matematiche e si ha allora, ricordando che

- $E(\bar{y}) = \mu$ e quindi $E\left(\frac{n}{\sigma^4}(\bar{y} - \mu)\right) = 0$;
- $E((y_i - \mu)^2) = \sigma^2$ e quindi

$$E\left(\frac{n}{2\sigma^4} - \frac{1}{\sigma^6} \sum_i (y_i - \mu)^2\right) = \frac{n}{2\sigma^4}$$

È dunque

$$\mathcal{I}(\theta) = \begin{bmatrix} \frac{n}{\sigma^2} & 0 \\ 0 & \frac{n}{2\sigma^4} \end{bmatrix}.$$

Media della funzione punteggio

Si ha

$$\begin{aligned} E(I^*(\theta); \theta) &= E\left(\frac{\partial}{\partial \theta} \log f(y; \theta); \theta\right) \\ &= E\left(\frac{1}{f(y; \theta)} \frac{\partial}{\partial \theta} f(y; \theta); \theta\right) \\ &= \int \frac{1}{f(y; \theta)} \left(\frac{\partial}{\partial \theta} f(y; \theta)\right) f(y; \theta) dy \end{aligned}$$

supponendo di essere in un problema regolare di stima

$$= \frac{d}{d\theta} \int f(y; \theta) dy = \frac{d}{d\theta} 1 = 0$$

ne consegue anche che

$$\text{var}(I^*(\theta); \theta) = E((I^*(\theta))^2; \theta)$$

Informazione e varianza della funzione punteggio

Si ha poi

$$\begin{aligned} E((I^*(\theta))^2; \theta) &= \int \left(\frac{\partial}{\partial \theta} \log f(y; \theta) \right)^2 f(y; \theta) dy \\ &= \int \left(\frac{1}{f(y; \theta)} \frac{\partial}{\partial \theta} f(y; \theta) \right)^2 f(y; \theta) dy \end{aligned}$$

Informazione e varianza della funzione punteggio

Si ha poi

$$\begin{aligned} E((I^*(\theta))^2; \theta) &= \int \left(\frac{\partial}{\partial \theta} \log f(y; \theta) \right)^2 f(y; \theta) dy \\ &= \int \left(\frac{1}{f(y; \theta)} \frac{\partial}{\partial \theta} f(y; \theta) \right)^2 f(y; \theta) dy \end{aligned}$$

qui notiamo che

$$\frac{\partial^2}{\partial \theta^2} \log f = \frac{\partial}{\partial \theta} \left(\frac{1}{f} \frac{\partial}{\partial \theta} f \right) = -\frac{1}{f^2} \left(\frac{\partial}{\partial \theta} f \right)^2 + \frac{1}{f} \frac{\partial^2}{\partial \theta^2} f$$

e quindi

$$\frac{1}{f^2} \left(\frac{\partial}{\partial \theta} f \right)^2 = -\frac{\partial^2}{\partial \theta^2} \log f + \frac{1}{f} \frac{\partial^2}{\partial \theta^2} f$$

Informazione e varianza della funzione punteggio

Si ha poi

$$\begin{aligned} E((I^*(\theta))^2; \theta) &= \int \left(\frac{\partial}{\partial \theta} \log f(y; \theta) \right)^2 f(y; \theta) dy \\ &= \int \left(\frac{1}{f(y; \theta)} \frac{\partial}{\partial \theta} f(y; \theta) \right)^2 f(y; \theta) dy \\ &= - \int \frac{\partial^2}{\partial \theta^2} \log f(y; \theta) f(y; \theta) dy + \\ &\quad + \int \left(\frac{1}{f(y; \theta)} \frac{\partial^2}{\partial \theta^2} f(y; \theta) \right) f(y; \theta) dy \\ &= -E \left(\frac{\partial^2}{\partial \theta^2} \log f(y; \theta) \right) + \frac{d^2}{d\theta^2} \int f(y; \theta) dy \\ &= -E \left(\frac{\partial^2}{\partial \theta^2} \log f(y; \theta) \right) = \mathcal{I}(\theta) \end{aligned}$$

Informazione e varianza della funzione punteggio

Si ha quindi

$$\text{var}(I^*(\theta); \theta) = E((I^*(\theta))^2; \theta) = \mathcal{I}(\theta).$$

Nel caso multivariato la formula si scrive

$$\mathcal{I}(\theta) = E \left((I^*(\theta))(I^*(\theta))^T; \theta \right).$$

$\mathcal{I}(\hat{\theta})$ e $\mathcal{I}(\theta_0)$

Detto θ_0 il vero valore del parametro la quantità che interessa è

$$\mathcal{I}(\theta_0)$$

che ovviamente non è nota, si ha però che

se $\hat{\theta}$ è consistente

allora (teorema di Slutsky)

$$\mathcal{I}(\hat{\theta}) \rightarrow \mathcal{I}(\theta_0),$$

cioè $\mathcal{I}(\hat{\theta})$ è uno stimatore consistente per $\mathcal{I}(\theta_0)$.

Informazione e indipendenza

Se le osservazioni Y_i sono indipendenti

$$l_*(\theta; y) = \frac{\partial}{\partial \theta} \sum_i \log f(y_i; \theta) = \sum_i \frac{\partial}{\partial \theta} \log f(y_i; \theta)$$

e quindi

$$\mathcal{I}(\theta) = \sum_i E \left(-\frac{\partial^2}{\partial \theta^2} \log f(y_i; \theta) \right) = \sum_i \mathcal{I}_i(\theta)$$

cioè l'informazione, in caso di indipendenza, è la somma delle informazioni.

Se in aggiunta si ha identica distribuzione

$$\mathcal{I}(\theta) = nE \left(-\frac{\partial^2}{\partial \theta^2} \log f(y_1; \theta) \right) = ni(\theta)$$

dove $i(\theta)$ è l'informazione associata a un'unica osservazione.

$J(\hat{\theta})$ e $i(\theta_0)$

Se il campione è IID

$$J_n(\theta) = - \sum_{i=1}^n \frac{\partial^2}{\partial \theta^2} \log f(y_i; \theta)$$

dove le v.a. $\frac{\partial^2}{\partial \theta^2} \log f(y_i; \theta)$ sono IID, quindi,

$$\frac{1}{n} J_n(\theta) \xrightarrow{q.c.} -E \left(\frac{\partial^2}{\partial \theta^2} \log f(y_1; \theta) \right) = i(\theta)$$

Sotto opportune condizioni si può anche mostrare che

$$\frac{1}{n} J_n(\hat{\theta}) \rightarrow i(\theta_0)$$

Informazione e riparametrizzazioni

Sia θ unidimensionale, consideriamo la riparametrizzazione

$$\psi = \psi(\theta)$$

da $(\mathcal{Y}, f(y; \theta), \Theta)$ a $(\mathcal{Y}, g(y; \psi), \Psi)$

$$g(y; \psi(\theta)) = f(y; \theta)$$

allora

$$I_*(\psi) = \frac{\partial}{\partial \psi} \ln g(y; \psi) = \frac{\partial}{\partial \psi} \ln f(y; \theta(\psi)) = \frac{d\theta}{d\psi} \left(\frac{\partial}{\partial \theta} \ln f(y; \theta) \right)$$

e quindi l'informazione attesa è

$$\mathcal{I}(\psi) = \text{var} I_*(\psi) = \left(\frac{d\theta}{d\psi} \right)^2 \mathcal{I}(\theta) = \left(\frac{d\psi}{d\theta} \right)^{-2} \mathcal{I}(\theta)$$

Informazione e riparametrizzazioni, caso di parametro multidimensionale

Nel caso di parametro q -dimensionale, dove la trasformazione è definita dalle

$$\psi_1(\theta_1, \dots, \theta_q)$$

$$\vdots$$

$$\psi_q(\theta_1, \dots, \theta_q)$$

si ha

$$\mathcal{I}(\psi) = \Delta^T \mathcal{I}(\theta) \Delta$$

dove Δ è una matrice $q \times q$ con elemento i, j

$$\Delta_{i,j} = \frac{d\theta_i}{d\psi_j}$$

Informazione e varianza

Un legame generico tra informazione e varianza di uno stimatore è suggerito dalla disuguaglianza di Rao-Cramer, secondo la quale, per uno stimatore non distorto $\tilde{\theta}$ si ha

$$\text{var}(\tilde{\theta}) \geq \mathcal{I}(\theta)^{-1}$$

che, nel caso di campione casuale semplice, diventa

$$\text{var}(\tilde{\theta}) \geq \frac{1}{ni(\theta)}.$$

Modello con statistica ancillare

Sia (a_i, y_i) campione di v.a. indipendenti con

- $a_i \in \{0, 1\}$ e $P(a_i = 0) = 1/2$,
- $y_i | a_i = k \sim \mathcal{N}(\theta, \sigma_{a_i}^2)$

si ha allora

$$l(\theta) = - \sum_{i=1}^n \log \sigma_{a_i}^2 - \frac{1}{2} \sum_{i=1}^n \frac{1}{\sigma_{a_i}^2} (y_i - \theta)^2$$

e la funzione punteggio è

$$l_*(\theta) = \sum_{i=1}^n \frac{1}{\sigma_{a_i}^2} (y_i - \theta)$$

da cui lo SMV $\hat{\theta} = \sum \frac{y_i}{\sigma_{a_i}^2} \left(\sum \frac{1}{\sigma_{a_i}^2} \right)^{-1}$.

Modello con statistica ancillare

La derivata seconda della log-verosimiglianza è allora

$$l''(\theta) = - \sum \frac{1}{\sigma_{a_i}^2}$$

e si ha

$$\mathcal{I}(\theta) = E(-l''(\theta)) = \sum E\left(\frac{1}{\sigma_{a_i}^2}\right) = \frac{n}{2} \left(\frac{1}{\sigma_0^2} + \frac{1}{\sigma_1^2} \right)$$

$$J(\hat{\theta}) = -l''(\hat{\theta}) = \frac{n - \sum a_i}{\sigma_0^2} + \frac{\sum a_i}{\sigma_1^2}$$

si noti che, in generale, $\mathcal{I}(\hat{\theta}) \neq J(\hat{\theta})$.

Esempio: normale con coefficiente di variazione unitario

Sia $y_1, \dots, y_n \sim IID(\mathcal{N}(\theta, \theta^2))$, $\theta > 0$, allora

$$L(\theta) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\theta}} \exp\left\{-\frac{1}{2\theta^2}(y_i - \theta)^2\right\} \propto \theta^{-n} \exp\left\{-\frac{1}{2\theta^2}\left(\sum_{i=1}^n y_i^2 - 2\theta n\bar{y} + n\theta^2\right)\right\}$$

$$l(\theta) = -n \log \theta - \frac{\sum y_i^2}{2\theta^2} + \frac{n\bar{y}}{\theta}$$

e quindi

$$l_*(\theta) = -\frac{n}{\theta} + \frac{\sum y_i^2}{\theta^3} - \frac{n\bar{y}}{\theta^2} = \theta^{-3} \left(-n\theta^2 - n\bar{y}\theta + \sum y_i^2\right)$$

le radici sono

$$\theta_{+,-} = \frac{1}{2n} \left(-n\bar{y} \pm \sqrt{n^2\bar{y}^2 + 4n \sum y_i^2}\right)$$

Esempio: normale con coefficiente di variazione unitario

dove $\theta_- < 0$ è al di fuori dello spazio parametrico mentre θ_+ è all'interno (si noti che $\sqrt{n^2\bar{y}^2 + 4n\sum y_i^2} > |n\bar{y}|$), studiando il segno della derivata prima si evince che θ_+ è un massimo, insomma

$$\hat{\theta} = \frac{1}{2n} \left(-n\bar{y} + \sqrt{n^2\bar{y}^2 + 4n\sum y_i^2} \right).$$

La derivata seconda della log-verosimiglianza è

$$l''(\theta) = \frac{n}{\theta^2} - \frac{3\sum y_i^2}{\theta^4} + \frac{2n\bar{y}}{\theta^3}$$

e si ha quindi

$$\mathcal{I}(\theta) = -\frac{n}{\theta^2} + \frac{3nE(y_i^2)}{\theta^4} - \frac{2nE(\bar{y})}{\theta^3} = \frac{3n}{\theta^2}$$

Indice

- 1 Precisione dello SMV
- 2 Informazione
- 3 Proprietà dello SMV**
- 4 Proprietà asintotiche dello SMV

Invarianza (equivarianza) rispetto a riparametrizzazioni

Consideriamo una riparametrizzazione $\psi = \psi(\theta) : \Theta \rightarrow \Phi$ biiettiva

$$(\mathcal{Y}, \Theta, p_\theta) \rightarrow (\mathcal{Y}, \Psi, g_\psi)$$

Proprietà

Se $\hat{\theta}$ è la SMV per il primo modello allora

$$\hat{\psi} = \psi(\hat{\theta})$$

Si ha infatti, qualunque sia $\theta \in \Theta$

$$g_{\psi(\hat{\theta})}(y) = p_{\hat{\theta}}(y) > p_\theta(y) = g_{\psi(\theta)}(y)$$

siccome ψ è biunivoca, al variare di θ in Θ , $\psi(\theta)$ “copre” tutto Ψ ,

$$g_{\psi(\hat{\theta})}(y) > g_\psi(y) \quad \forall \psi \in \Psi$$

e quindi $\psi(\hat{\theta})$ è SMV del secondo modello.

Invarianza (equivarianza) rispetto a riparametrizzazioni

Consideriamo una riparametrizzazione $\psi = \psi(\theta) : \Theta \rightarrow \Phi$ biiettiva

$$(\mathcal{Y}, \Theta, p_\theta) \rightarrow (\mathcal{Y}, \Psi, g_\psi)$$

Proprietà

Se $\hat{\theta}$ è la SMV per il primo modello allora

$$\hat{\psi} = \psi(\hat{\theta})$$

Si ha poi

$$\mathcal{I}_\psi(\theta) = \Delta^T \mathcal{I}(\theta_0) \Delta$$

dove Δ è la matrice con elemento i, j $\frac{\partial \theta_i}{\partial \psi_j}$.

Esponenziale, due parametrizzazioni

Consideriamo $y_1, \dots, y_n \sim IID$ $p_\theta(y) = \theta e^{-\theta y}$ e sia

$$\psi = h(\theta) = 1/\theta \rightarrow p_\psi(y) = e^{-y/\psi}/\psi$$

la log-verosimiglianza per θ è

$$l(\theta) = n \log \theta - \theta \sum_i y_i$$

In virtù dell'invarianza si ha

$$\hat{\theta} = 1/\bar{y} \Rightarrow \hat{\psi} = h(\hat{\theta}) = \bar{y}$$

Infatti la log-verosimiglianza di ψ è

$$l(\psi) = -n \log \psi - \psi^{-1} \sum_i y_i$$

e la funzione punteggio si annulla per $\hat{\psi} = 1/\bar{y}$

$$l_*(\psi) = -n\psi^{-1} + \psi^{-2} \sum_i y_i$$

Riparametrizzazioni: informazione

Per quanto riguarda l'informazione di Fisher si scriva, nel caso di un parametro reale, la funzione di log-verosimiglianza per il parametro ψ come

$$l_{\Psi}(\psi) = l_{\Theta}(h^{-1}(\psi))$$

derivando questa due volte si ottiene

$$\begin{aligned} \frac{d^2 l_{\Psi}(\psi)}{d\psi^2} &= \frac{d}{d\theta} \left(\frac{dl_{\Theta}(\theta)}{d\theta} \Big|_{\theta=h^{-1}(\psi)} \frac{dh^{-1}(\psi)}{d\psi} \right) \\ &= \frac{d^2 l_{\Theta}(\theta)}{d\theta^2} \Big|_{\theta=h^{-1}(\psi)} \left(\frac{dh^{-1}(\psi)}{d\psi} \right)^2 + \frac{dl_{\Theta}(\theta)}{d\theta} \Big|_{\theta=h^{-1}(\psi)} \frac{d^2 h^{-1}(\psi)}{d\psi^2} \end{aligned}$$

e quindi

$$\mathcal{I}_{\Psi}(\psi) = E \left(\frac{d^2 l_{\Psi}(\psi)}{d\psi^2} \right) = \mathcal{I}_{\Theta}(h^{-1}(\psi)) \left(\frac{dh^{-1}(\psi)}{d\psi} \right)^2 + 0$$

in virtù del fatto che il punteggio ha media nulla.

(Segue) Esponenziale

Nel caso dell'esponenziale con parametro θ : $Y_1, \dots, Y_n \sim_{IID} \theta e^{-\theta y}$ si ha

$$\mathcal{I}_{\Theta}(\theta) = n/\theta^2$$

Da questo si ricava

$$\mathcal{I}_{\Psi}(\psi) = \mathcal{I}_{\Theta}(h^{-1}(\psi)) \left(\frac{dh^{-1}(\psi)}{d\psi} \right)^2 = n/(1/\psi^2)(1/\psi^2)^2 = n/\psi^2$$

Considerando d'altra parte la derivata seconda della log-verosimiglianza

$$l''_{\Psi}(\psi) = n\psi^{-2} - 2n\bar{y}\psi^{-3}$$

si ha, ricordando che $E(\bar{y}) = \psi$

$$\mathcal{I}_{\Psi}(\Psi) = -E(l''_{\Psi}(\psi)) = n\psi^{-2}$$

Riparametrizzazioni: informazione

Nel caso di parametro multidimensionale

$$\psi = \psi(\theta) : \mathbb{R}^d \rightarrow \mathbb{R}^d$$

(ψ_i per $i = 1, \dots, d$ è l' i -mo elemento di ψ), $\theta = \theta(\psi)$ indichiamo la sua inversa allora

$$\mathcal{I}_\psi(\theta) = \Delta^T \mathcal{I}(\theta_0) \Delta$$

dove Δ è la matrice con elemento i, j $\frac{\partial \theta_i}{\partial \psi_j}$

Da ciò deriva che la distribuzione asintotica per $\hat{\psi}$ è

$$\sqrt{n}(\hat{\psi} - \psi_0) \rightarrow_d \mathcal{N}(0, (\Delta^T \mathcal{I}(\theta_0) \Delta)^{-1})$$

È bene notare che questo non è lo stesso risultato che si otterrebbe considerando la normale come distribuzione asintotica di $\hat{\theta}$ e valutando la distribuzione del trasformato $\hat{\psi} = \psi(\hat{\theta})$.

Normale, stima della varianza e dello s.q.m.

Poco sopra si è considerato il modello $\mathcal{N}(\mu, \sigma^2)$ dove $\theta = (\mu, \sigma^2)$, consideriamo la riparametrizzazione

$$(\mu, \sigma^2) \rightarrow (\mu, \sigma)$$

si ha allora

$$l(\mu, \sigma) = -n \log \sigma - \frac{1}{2\sigma^2} \sum_i (y_i - \mu)^2$$

e

$$l_*(\mu, \sigma) = \left[-\frac{n}{\sigma} + \frac{1}{\sigma^3} \sum_i (y_i - \mu)^2 \right]$$

quindi

$$J((\mu, \sigma)) = \begin{bmatrix} -n/\sigma^2 & -2 \sum_i (y_i - \mu)/\sigma^3 \\ -2 \sum_i (y_i - \mu)/\sigma^3 & \frac{n}{\sigma^2} - \frac{3}{\sigma^4} \sum_i (y_i - \mu)^2 \end{bmatrix}$$

e

$$J(\hat{\mu}, \hat{\sigma}) = \begin{bmatrix} n/\hat{\sigma}^2 & 0 \\ 0 & -n/(2\hat{\sigma}^2) \end{bmatrix} \quad \mathcal{I}(\mu, \sigma) = \begin{bmatrix} n/\sigma^2 & 0 \\ 0 & n/(2\sigma^2) \end{bmatrix}$$

Normale, stima della varianza e dello s.q.m.

$$J(\hat{\mu}, \hat{\sigma}) = \begin{bmatrix} n/\hat{\sigma}^2 & 0 \\ 0 & -n/(2\hat{\sigma}^2) \end{bmatrix} \quad \mathcal{I}(\mu, \sigma) = \begin{bmatrix} n/\sigma^2 & 0 \\ 0 & n/(2\sigma^2) \end{bmatrix}$$

Usando la formula sopra riportata per la trasformazione di variabile, essendo la nuova parametrizzazione ψ legata alla vecchia, θ , da

$$g(\psi) = (\theta_1, \sqrt{\theta_2})$$

Si ha allora

$$\Delta^T \mathcal{I}(\theta) \Delta = \begin{bmatrix} 1 & 0 \\ 0 & 2\sigma \end{bmatrix} \begin{bmatrix} n/\sigma^2 & 0 \\ 0 & n/2\sigma^4 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 2\sigma \end{bmatrix} = \begin{bmatrix} n/\sigma^2 & 0 \\ 0 & n/(2\sigma^2) \end{bmatrix}$$

Indice

- 1 Precisione dello SMV
- 2 Informazione
- 3 Proprietà dello SMV**
 - Confronto tra SMV e stimatore dei momenti
- 4 Proprietà asintotiche dello SMV

Esempio: uniforme, continua

Sia $(Y_1, \dots, Y_n) \sim IID \text{ Unif}(0, \theta)$, $\hat{\theta} = M_n = Y_{(n)}$, ricaviamo la distribuzione dello stimatore.

Ciascuna Y_i ha funzione di densità

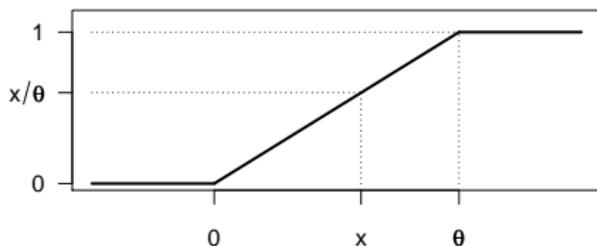
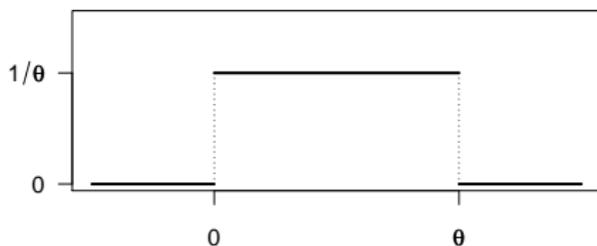
$$f_X(x; \theta) = \begin{cases} 1/\theta & \text{se } 0 \leq x \leq \theta \\ 0 & \text{altrimenti} \end{cases}$$

e quindi funzione di ripartizione

$$F_X(x; \theta) = \begin{cases} 0 & \text{se } x < 0 \\ \int_0^x 1/\theta dt = \frac{x}{\theta} & \text{se } 0 \leq x < \theta \\ 1 & \text{se } x > \theta. \end{cases}$$

Esempio: uniforme, continua

Sia $(Y_1, \dots, Y_n) \sim IID \text{ Unif}(0, \theta)$, $\hat{\theta} = M_n = Y_{(n)}$, ricaviamo la distribuzione dello stimatore.



Uniforme, distribuzione dello SMV

La funzione di ripartizione di M_n è dunque

$$\begin{aligned} F_M(y) &= P(M_n \leq y) = P\left(\bigcap_{i=1}^n (Y_i \leq y)\right) = \prod_{i=1}^n P(Y_i \leq y) = [P(Y_i \leq y)]^n \\ &= [F_Y(y)]^n = \begin{cases} 0 & \text{se } y < 0 \\ \frac{y^n}{\theta^n} & \text{se } 0 \leq x \leq \theta \\ 1 & \text{se } y > \theta. \end{cases} \end{aligned}$$

Possiamo calcolare anche la corrispondente funzione di densità che risulta

$$f_M(y) = \frac{d}{dy} F_M(y) = \begin{cases} \frac{ny^{n-1}}{\theta^n} & \text{se } 0 \leq y \leq \theta \\ 0 & \text{altrimenti.} \end{cases}$$

Uniforme, diversi stimatori

Consideriamo la stima col metodo dei momenti

$$\bar{Y} = E(Y) = \int_{-\infty}^{+\infty} y f_Y(y) dy = \int_0^{\theta} \frac{y}{\theta} dy = \frac{\theta}{2} \Rightarrow \tilde{\theta} = 2\bar{Y}.$$

Confrontiamo gli stimatori sotto tre aspetti,

- a. appartenenza della stima allo spazio parametrico
- b. correttezza
- c. errore quadratico medio

Confronto stimatori: appartenenza della stima allo spazio parametrico

- Lo stimatore di massima verosimiglianza è tale per cui $Y_i \in [0, \hat{\theta}] \forall i$ per costruzione.
- Lo stimatore del metodo dei momenti non rispetta invece questa condizione, se ad esempio il campione fosse $(0.5, 1.5, 10)$ si avrebbe $\tilde{\theta} = 2\bar{Y} = 2 \times 4 = 8$ e $x_3 > \tilde{\theta}$.

Questa circostanza farebbe preferire lo stimatore di massima verosimiglianza.

Confronto stimatori: correttezza

$$\begin{aligned} E(\tilde{\theta}) &= E(2\bar{Y}) = 2E(\bar{Y}) = 2E\left(\frac{1}{n}\sum_i Y_i\right) = 2\frac{1}{n}E\left(\sum_i Y_i\right) \\ &= 2\frac{1}{n}\sum_i E(Y_i) = 2\frac{1}{n}\sum_i E(Y_1) = 2\frac{1}{n}nE(Y_1) = 2E(Y_1) = \theta \end{aligned}$$

quindi lo stimatore $\tilde{\theta}$ ottenuto col metodo dei momenti è corretto.

Confronto stimatori: correttezza

Per quanto concerne lo stimatore di massima verosimiglianza si ha

$$E(\hat{\theta}) = E(M_n) = \int_{-\infty}^{+\infty} tf_M(t)dt$$

ricordando quanto visto al primo punto otteniamo

$$= \int_0^{\theta} t \frac{nt^{n-1}}{\theta^n} dt = \frac{n}{\theta^n} \int_0^{\theta} t^n dt = \frac{n}{\theta^n} \frac{\theta^{n+1}}{n+1} = \frac{n}{n+1} \theta$$

quindi lo SMV non è corretto

$$\text{bias} = E(\hat{\theta}) - \theta = -\frac{1}{n+1} \theta < 0$$

è comunque asintoticamente corretto in quanto $E(M_n) \rightarrow \theta$.

Sotto l'aspetto della correttezza, dunque, lo stimatore $\tilde{\theta}$ risulta preferibile.

Confronto stimatori: EQM

Per lo stimatore $\tilde{\theta}$ l'EQM è uguale alla varianza, poiché lo stimatore è corretto, quindi

$$\begin{aligned}EQM(\tilde{\theta}) &= V(\tilde{\theta}) = V(2\bar{Y}) = 4V(\bar{Y}) = 4V\left(\frac{1}{n}\sum_i Y_i\right) = 4\frac{1}{n^2}V\left(\sum_i Y_i\right) \\ &= 4\frac{1}{n^2}\sum_i V(Y_i) = 4\frac{1}{n^2}\sum_i V(Y_1) = 4\frac{1}{n^2}nV(Y_1) = \frac{\theta^2}{3n}\end{aligned}$$

poiché $V(Y_1) = \theta^2/12$.

Confronto stimatori: EQM

Per lo stimatore $\hat{\theta}$ invece dobbiamo tenere conto anche della distorsione, abbiamo allora

$$V(\hat{\theta}) = V(M_n) = E(M_n^2) - E^2(M_n)$$

dove

$$E(M_n^2) = \int_{-\infty}^{+\infty} t^2 f_M(t) dt = \int_0^{\theta} t^2 \frac{nt^{n-1}}{\theta^n} dt = \frac{n}{\theta^n} \frac{\theta^{n+2}}{n+2} = \frac{n}{n+2} \theta^2$$

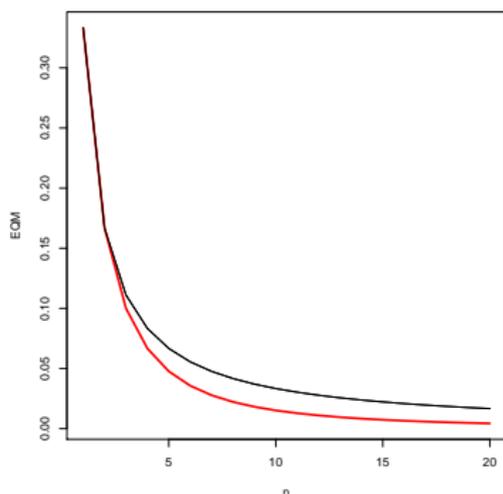
e quindi

$$V(\hat{\theta}) = \frac{n}{n+2} \theta^2 - \left(\frac{n}{n+1} \theta \right)^2 = \theta^2 \frac{n}{(n+2)(n+1)^2}$$

l'errore quadratico medio è invece

$$EQM(\hat{\theta}) = \text{bias}^2 + V(\hat{\theta}) = \frac{\theta^2}{(n+1)^2} + \frac{n\theta^2}{(n+2)(n+1)^2} = \frac{2\theta^2}{(n+2)(n+1)}$$

Confronto stimatori: EQM



EQM di $\tilde{\theta}$ (nero) e $\hat{\theta}$ (rosso).

Il criterio dell'errore quadratico medio fa preferire lo stimatore di massima verosimiglianza

Nel complesso dunque che stimatore scegliamo?

L'EQM è migliore per $\hat{\theta}$, che garantisce anche l'appartenenza della stima allo spazio parametrico, l'unico punto a favore di $\tilde{\theta}$, la correttezza, è trascurabile. La conclusione è in favore dello stimatore di massima verosimiglianza.

Indice

- 1 Precisione dello SMV
- 2 Informazione
- 3 Proprietà dello SMV
- 4 Proprietà asintotiche dello SMV**

Indice

- 1 Precisione dello SMV
- 2 Informazione
- 3 Proprietà dello SMV
- 4 Proprietà asintotiche dello SMV
 - Premesse: convergenze in probabilità
 - Proprietà asintotiche degli stimatori
 - Consistenza dello SMV
 - Distribuzione asintotica dello SMV

Convergenze stocastiche

Sia Y_n successione di v.a

- Y_n converge in probabilità a Y , $Y_n \xrightarrow{p} Y$ sse

$$\lim_{n \rightarrow +\infty} P(\|Y_n - Y\| \geq \varepsilon) = 0 \quad \forall \varepsilon > 0$$

- $Y_n \sim F_n$ converge in legge a $Y \sim F$, $Y_n \xrightarrow{d} Y$ sse

$$\lim_{n \rightarrow +\infty} F_n(y) = F(y) \quad \forall y \in \mathbb{R}^d$$

- Y_n converge quasi certamente a Y , $Y_n \xrightarrow{q.c.} Y$ sse

$$P\left(\lim_{n \rightarrow \infty} Y_n = Y\right) = 1$$

- Y_n converge in media p -ma a Y , $Y_n \xrightarrow{L_p} Y$ sse $\lim_{n \rightarrow \infty} E(|X_n - X|^p) = 0$.

Risultati notevoli sulle convergenze

- Legge dei grandi numeri: $\{Y_i\}$ IID, $E(Y_1^2) < +\infty$, allora

$$\frac{1}{n} \sum_{i=1}^n Y_i \xrightarrow{q.c.} E(Y_1)$$

- Teorema del limite centrale: Y_n IID, $E(Y_1^2) < +\infty$, allora

$$\frac{\sum_{i=1}^n Y_i - nE(Y_1)}{\sqrt{n\text{var}(Y_1)}} \xrightarrow{d} \mathcal{N}(0, 1)$$

- Relazioni tra convergenze

$$Y_n \xrightarrow{q.c.} Y \Rightarrow Y_n \xrightarrow{p} Y \Rightarrow Y_n \xrightarrow{d} Y$$

$$Y_n \xrightarrow{L_2} Y \Rightarrow Y_n \xrightarrow{p} Y$$

$$Y_n \xrightarrow{d} c \Rightarrow Y_n \xrightarrow{p} c$$

Convergenze e operazioni

Considerando due successioni X_n e Y_n che convergono in probabilità a X e Y , allora

- $X_n + Y_n \xrightarrow{P} X + Y$
- $cX_n \xrightarrow{P} cX$
- $X_n Y_n \xrightarrow{P} XY$
- $X_n / Y_n \xrightarrow{P} X / Y$
- $g : \mathbb{R}^k \rightarrow \mathbb{R}^m$ continua, X_n a valori in \mathbb{R}^k , allora

$$g(X_n) \xrightarrow{P} g(X)$$

le relazioni valgono anche sostituendo convergenza in probabilità con convergenza in legge.

Teorema di Slutsky

Teorema

Siano X_n, Y_n successioni di v.a. tali che

$$X_n \xrightarrow{p} c \text{ e } Y_n \xrightarrow{d} Y$$

e sia

$$T_n = g(X_n, Y_n)$$

con $g(x, y) : \mathbb{R}^2 \rightarrow \mathbb{R}$ funzione

- continua di y per ogni x e
- continua in $x = c$ per ogni y

allora

$$T_n \rightarrow_d g(c, Y)$$

Teorema di Slutsky

Conseguenza: la t di Student converge a una normale

$$T_n = \frac{Z}{\sqrt{U_n/n}} \rightarrow \frac{Z}{1}$$

dove $U_n \sim \chi_n^2$ e $Z \sim \mathcal{N}(0, 1)$ e poiché

- $E(U_n/n) = 1$ e $\text{var}(U_n/n) = 1/n \rightarrow 0$ quindi $E(U_n/n - 1)^2 \rightarrow 0$
- Slutsky con $g(x, y) = x/\sqrt{y}$

Ordini di infinito e infinitesimo

Date due successioni $\{a_n\}$ e $\{b_n\}$ di reali si ha

$$b_n = o(a_n) \Leftrightarrow \lim_{n \rightarrow +\infty} \frac{b_n}{a_n} = 0,$$

$$b_n = O(a_n) \Leftrightarrow \lim_{n \rightarrow +\infty} \left| \frac{b_n}{a_n} \right| < +\infty.$$

Analogamente per una successione di v.a. $\{X_n\}$ e una successione di reali a_n si ha

$$X_n = o_p(a_n) \Leftrightarrow \frac{X_n}{a_n} \xrightarrow{p} 0,$$

$$X_n = O_p(a_n) \Leftrightarrow \forall \varepsilon > 0 \exists M_\varepsilon \text{ t.c. } P\left(\left|\frac{X_n}{a_n}\right| > M_\varepsilon\right) < \varepsilon \forall n > N_\varepsilon.$$

Proprietà degli ordini di infinito e infinitesimo

Si danno di seguito alcune immediate conseguenze e alcune proprietà

- $X_n = o_p(1) \Leftrightarrow X_n \xrightarrow{P} 0$
- $X_n = O_p(1)$ la probabilità di osservare valori elevati in valore assoluto di X_n è limitata.
- se $X_n = o_p(a_n)$ e $Y_n = o_p(b_n)$, allora
 - $X_n Y_n = o_p(a_n b_n)$;
 - $X_n + Y_n = o_p(\max\{a_n, b_n\})$.
- se $X_n = O_p(a_n)$ e $Y_n = O_p(b_n)$, allora
 - $X_n Y_n = O_p(a_n b_n)$;
 - $X_n + Y_n = O_p(\max\{a_n, b_n\})$.

Proprietà degli ordini di infinito e infinitesimo

- se $X_n = O_p(a_n)$ e $Y_n = o_p(b_n)$, allora
 - $X_n Y_n = o_p(a_n b_n)$.
- Se X_n è tale che $E(X_n^2) = a_n^2 < +\infty$ e b_n è una successione di reali positivi, allora
 - $a_n^2 = O(b_n^2) \Rightarrow X_n = O_p(b_n)$;
 - $a_n^2 = o(b_n^2) \Rightarrow X_n = o_p(b_n)$.
- Dati X_n e a_n successioni di v.a. e reali tali che

$$a_n(X_n - c) \xrightarrow{d} Z$$

con c reale e Z v.a. non degenere, allora

$$X_n = c + O_p(a_n^{-1})$$

Sviluppo in serie stocastico

Consideriamo una successione X_n tale che

$$X_n = c + O_p(a_n), \quad a_n \rightarrow 0$$

Sia poi $f(\cdot) \in \mathcal{C}^k$ si ha lo sviluppo

$$f(x) = f(c) + \sum_{i=1}^k \frac{1}{i!} f^{(i)}(c)(x-c)^i + \frac{1}{k!} (f^{(k)}(v) - f^{(k)}(c))(x-c)^k$$

con v compreso tra c e x , e di conseguenza con resto

$$\frac{1}{k!} (f^{(k)}(v) - f^{(k)}(c))(x-c)^k = o((x-c)^k)$$

Sostituendo X_n a x si ha un'espressione per $f(X_n)$ sulla falsariga di quella sopra,

$$f(X_n) = f(c) + \sum_{i=1}^k \frac{1}{i!} f^{(i)}(c)(X_n - c)^i + \frac{1}{k!} (f^{(k)}(V_n) - f^{(k)}(c))(X_n - c)^k$$

dove V_n è compreso tra c e X_n .

Sviluppo in serie stocastico

$$f(X_n) = f(c) + \sum_{i=1}^k \frac{1}{i!} f^{(i)}(c)(X_n - c)^i + \frac{1}{k!} (f^{(k)}(V_n) - f^{(k)}(c))(X_n - c)^k$$

dove V_n è compreso tra c e X_n .

Essendo $V_n \xrightarrow{P} c$ ed essendo $f^{(k)}$ continua si ha

$$f^{(k)}(V_n) \xrightarrow{P} f^{(k)}(c) \quad \text{cioè} \quad f^{(k)}(V_n) - f^{(k)}(c) = o_p(1).$$

D'altra parte si ha $(X_n - c)^k = O_p(a_n^k)$ e quindi

$$\frac{1}{k!} (f^{(k)}(V_n) - f^{(k)}(c))(X_n - c)^k = o_p(a_n^k)$$

e per $f(X_n)$ si può scrivere

$$f(X_n) = f(c) + \sum_{i=1}^k \frac{1}{i!} f^{(i)}(c)(X_n - c)^i + o_p(a_n^k). \quad (1)$$

Metodo delta

Un'immediata conseguenza di ciò è la relazione, nota anche come metodo delta, per cui se

$$\sqrt{n}(X_n - c) \xrightarrow{d} Z \sim N(0, \sigma^2)$$

e f è una funzione continua allora

$$\sqrt{n}(f(X_n) - f(c)) \xrightarrow{d} f'(c)Z.$$

cioè

$$\sqrt{n}(f(X_n) - f(c)) \xrightarrow{d} \mathcal{N}(0, [f'(c)]^2 \sigma^2)$$

Indice

- 1 Precisione dello SMV
- 2 Informazione
- 3 Proprietà dello SMV
- 4 Proprietà asintotiche dello SMV**
 - Premesse: convergenze in probabilità
 - Proprietà asintotiche degli stimatori**
 - Consistenza dello SMV
 - Distribuzione asintotica dello SMV

Successione di stimatori

Uno stimatore è una funzione del campione a valori nello spazio parametrico

$$T(Y_1, \dots, Y_n) : \mathcal{Y} \rightarrow \Theta$$

Si parla di proprietà asintotiche degli stimatori quando si considera il loro comportamento al crescere della dimensione campionaria.

Indichiamo allora lo stimatore con

$$T_n(Y) : \mathcal{Y} \rightarrow \Theta$$

per evidenziare il ruolo di n .

Consistenza: definizione

Definizione

Lo stimatore T_n si dice **consistente in senso debole** se e solo se

$$T_n(Y) \xrightarrow{P} \theta_0$$

dove θ_0 è il vero valore del parametro.

Definizione

Lo stimatore T_n si dice **consistente in senso forte** se e solo se

$$T_n(Y) \xrightarrow{q.c.} \theta_0$$

dove θ_0 è il vero valore del parametro.

Consistenza della media aritmetica

Se Y_1, \dots, Y_n sono IID e

$$E(Y_i) = \mu, \quad E(Y_i^2) = \sigma^2 < +\infty$$

La statistica

$$\bar{Y}_n = \frac{1}{n} \sum_i Y_i$$

è uno stimatore consistente della media μ .

Questo risultato si ottiene banalmente dalla legge dei grandi numeri ricordata sopra.

Distribuzione uniforme

Dato un campione IID $Y_i \sim \text{Unif}(0, \theta)$, si ha $\hat{\theta}_n = \max\{Y_1, \dots, Y_n\}$, questo è consistente in quanto è asintoticamente corretto e

$$\text{var}(\hat{\theta}_n) = \frac{\theta^2}{3n} \xrightarrow{n \rightarrow \infty} 0$$

da ciò infatti segue che $\hat{\theta}_n \xrightarrow{L_2} \theta_0$.

Indice

- 1 Precisione dello SMV
- 2 Informazione
- 3 Proprietà dello SMV
- 4 Proprietà asintotiche dello SMV
 - Premesse: convergenze in probabilità
 - Proprietà asintotiche degli stimatori
 - **Consistenza dello SMV**
 - Distribuzione asintotica dello SMV

Consistenza dello SMV

La consistenza dello SMV è garantita sotto condizioni abbastanza generali, qui ne diamo una dimostrazione in un caso tutto sommato particolare ma abbastanza generico ai nostri fini.

Intuitivamente parlando la consistenza dello SMV deriva dal fatto che la funzione di verosimiglianza $l_n(\theta)$ è, asintoticamente, incommensurabilmente più grande in corrispondenza al vero valore del parametro, si osservi infatti che, per un campione iid.

Limite di $l(\cdot)$ nel vero valore

Per un campione IID si ha in virtù della legge dei grandi numeri

$$\frac{l_n(\theta)}{n} = \frac{1}{n} \sum_{i=1}^n l(\theta; y_i) \rightarrow_p E_{\theta_0}(l(\theta; y_1))$$

allora

$$\frac{l_n(\theta)}{n} - \frac{l_n(\theta_0)}{n} \rightarrow_p E_{\theta_0}(l(\theta; y_1)) - E_{\theta_0}(l(\theta_0; y_1))$$

per il secondo membro si ha

$$\begin{aligned} E_{\theta_0}(l(\theta; y_1)) - E_{\theta_0}(l(\theta_0; y_1)) &= E_{\theta_0}(\log f(y_1; \theta)) - E_{\theta_0}(\log f(y_1; \theta_0)) \\ &= E_{\theta_0}(\log f(y_1; \theta) - \log f(y_1; \theta_0)) \\ &= E_{\theta_0} \left(\log \frac{f(y_1; \theta)}{f(y_1; \theta_0)} \right) \\ &< \log E_{\theta_0} \left(\frac{f(y_1; \theta)}{f(y_1; \theta_0)} \right) = \log 1 = 0 \end{aligned}$$

Limite di $l(\cdot)$ nel vero valore

Per un campione IID si ha in virtù della legge dei grandi numeri

$$\frac{l_n(\theta)}{n} = \frac{1}{n} \sum_{i=1}^n l(\theta; y_i) \rightarrow_p E_{\theta_0}(l(\theta; y_1))$$

allora

$$\frac{l_n(\theta)}{n} - \frac{l_n(\theta_0)}{n} \rightarrow_p E_{\theta_0}(l(\theta; y_1)) - E_{\theta_0}(l(\theta_0; y_1))$$

per il secondo membro si ha

$$E_{\theta_0}(l(\theta; y_1)) - E_{\theta_0}(l(\theta_0; y_1)) < \log E_{\theta_0} \left(\frac{f(y_1; \theta)}{f(y_1; \theta_0)} \right) = \log 1 = 0$$

e quindi

$$\frac{l_n(\theta)}{n} - \frac{l_n(\theta_0)}{n} \rightarrow_p c < 0$$

Limite di $l(\cdot)$ nel vero valore

Per un campione IID si ha in virtù della legge dei grandi numeri

$$\frac{l_n(\theta)}{n} = \frac{1}{n} \sum_{i=1}^n l(\theta; y_i) \rightarrow_p E_{\theta_0}(l(\theta; y_1))$$

allora

$$\frac{l_n(\theta)}{n} - \frac{l_n(\theta_0)}{n} \rightarrow_p E_{\theta_0}(l(\theta; y_1)) - E_{\theta_0}(l(\theta_0; y_1))$$

per il secondo membro si ha

$$E_{\theta_0}(l(\theta; y_1)) - E_{\theta_0}(l(\theta_0; y_1)) < \log E_{\theta_0} \left(\frac{f(y_1; \theta)}{f(y_1; \theta_0)} \right) = \log 1 = 0$$

e quindi

$$\frac{l_n(\theta)}{n} - \frac{l_n(\theta_0)}{n} \rightarrow_p c < 0 \Rightarrow l_n(\theta) - l_n(\theta_0) \rightarrow -\infty$$

Illustrazione con distribuzione normale

Se $y_1, \dots, y_n \sim IID(\mathcal{N}(\theta, \sigma^2))$ con σ^2 noto si ha

$$l_n(\theta) = -\frac{n}{2\sigma^2}(\bar{y} - \theta)^2$$

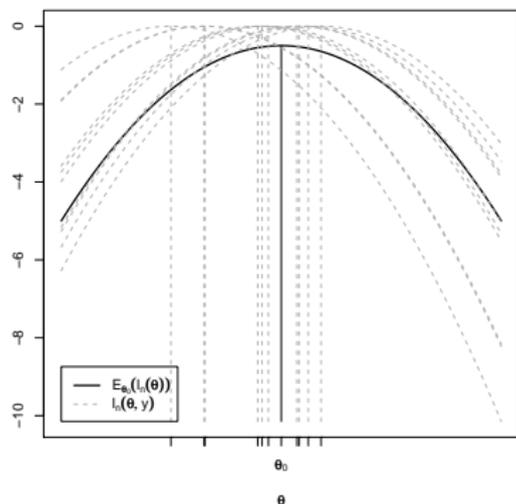
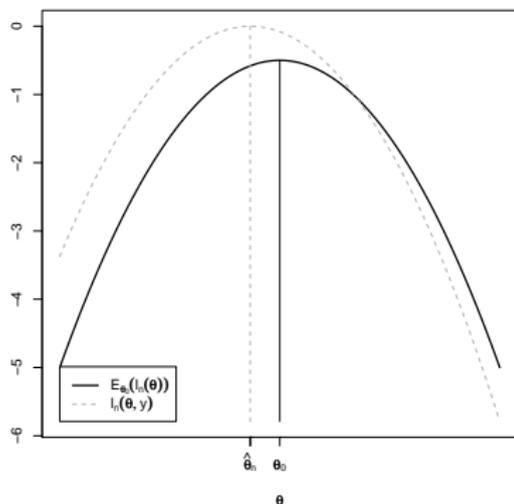
e

$$\begin{aligned} E(l_n(\theta)) &= -\frac{n}{2\sigma^2} E((\bar{y} - \theta)^2) \\ &= -\frac{n}{2\sigma^2} E(\bar{y}^2 - 2\bar{y}\theta + \theta^2) \\ &= -\frac{n}{2\sigma^2} \left(\frac{\sigma^2}{n} + (\theta - \theta_0)^2 \right) \\ &= -\frac{1}{2} - \frac{n}{2\sigma^2} (\theta - \theta_0)^2 \end{aligned}$$

e

$$E(l_n(\theta)) - E(l_n(\theta_0)) = -\frac{n}{2\sigma^2} (\theta - \theta_0)^2 < 0$$

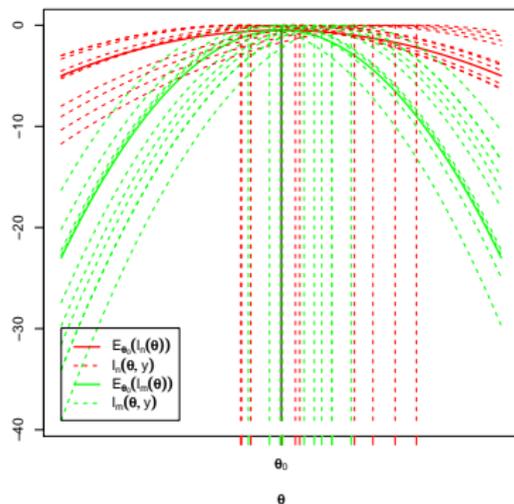
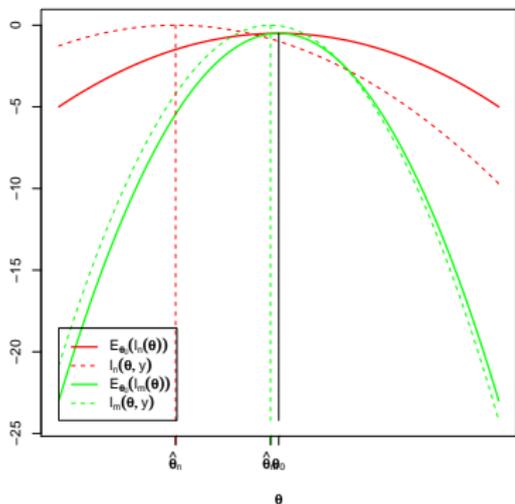
Esempio



funzione $E_{\theta_0}(l_n(\theta; y))$ e, accanto ad essa, 10 determinazioni di $l_n(\theta)$ e relative stime per un campione iid di numerosità n da una normale e suo valore atteso.

Il massimo di $E_{\theta_0}(l_n(\theta; y))$ è in corrispondenza a θ_0 , il massimo di $l_n(\theta)$

Esempio



Confronto delle log-verosimiglianze e relativi valori attesi per due campioni di diversa numerosità ($n = 10$ e $m = 50$, in particolare) si nota come da una parte aumenta $E_{\theta_0}(\theta_0) - E_{\theta_0}(\theta)$ cresce al crescere della numerosità campionaria e, dall'altra le SMV si concentrano intorno al vero valore.

Consistenza dello SMV

Teorema: consistenza dello SMV

Se

- (i) il modello è identificabile;
- (ii) $\text{supp}p_\theta(y)$ non dipende da θ
- (iii) y_1, \dots, y_n sono IID

allora lo SMV $\hat{\theta}_n$ è consistente.

Consistenza dello SMV: dimostrazione

La dimostrazione sul testo è in due parti

- la prima è generale e si arriva al fatto, già visto, che

$$l_n(\theta_0) - l_n(\theta) \xrightarrow{q.c.} +\infty$$

- la seconda è fatta in ipotesi restrittive

Consistenza dello SMV: dimostrazione, II parte

Dal fatto che $I_n(\theta_0)$ è incommensurabilmente più grande di $I_n(\theta)$ vogliamo derivare che $\hat{\theta}_n \rightarrow \theta_0$, si dà la dimostrazione **nel caso particolare in cui Θ è finito**

$$\Theta = \{\theta_0, \theta_1, \dots, \theta_m\}$$

La convergenza q.c. di $\hat{\theta}_n$ a θ_0 equivale a

$$\forall \delta, \varepsilon \exists n_0 \text{ t.c. } P \left(\bigcap_{n > n_0} |\hat{\theta}_n - \theta_0| < \varepsilon \right) > 1 - \delta$$

Consideriamo allora l'evento

$$\bigcap_{n > n_0} |\hat{\theta}_n - \theta_0| < \varepsilon$$

Consistenza dello SMV: dimostrazione, II parte

Dal fatto che $l_n(\theta_0)$ è incommensurabilmente più grande di $l_n(\theta)$ vogliamo derivare che $\hat{\theta}_n \rightarrow \theta_0$, si dà la dimostrazione **nel caso particolare in cui Θ è finito**

$$\Theta = \{\theta_0, \theta_1, \dots, \theta_m\}$$

La convergenza q.c. di $\hat{\theta}_n$ a θ_0 equivale a

$$\forall \delta, \varepsilon \exists n_0 \text{ t.c. } P \left(\bigcap_{n > n_0} |\hat{\theta}_n - \theta_0| < \varepsilon \right) > 1 - \delta$$

Consideriamo allora l'evento

$$\bigcap_{n > n_0} |\hat{\theta}_n - \theta_0| < \varepsilon = \bigcap_{n > n_0} l_n(\theta_0) - l_n(\theta_j) > \nu > 0 \quad \forall j \neq 0$$

questo perché Θ è finito! Ponendo $\varepsilon < \min\{|\theta_i - \theta_0|\}$ allora $|\theta_n - \theta_0| < \varepsilon$ se e solo se

$$\hat{\theta}_n = \theta_0$$

Consistenza dello SMV: dimostrazione, II parte

Dal fatto che $l_n(\theta_0)$ è incommensurabilmente più grande di $l_n(\theta)$ vogliamo derivare che $\hat{\theta}_n \rightarrow \theta_0$, si dà la dimostrazione **nel caso particolare in cui Θ è finito**

$$\Theta = \{\theta_0, \theta_1, \dots, \theta_m\}$$

La convergenza q.c. di $\hat{\theta}_n$ a θ_0 equivale a

$$\forall \delta, \varepsilon \exists n_0 \text{ t.c. } P \left(\bigcap_{n > n_0} |\hat{\theta}_n - \theta_0| < \varepsilon \right) > 1 - \delta$$

Consideriamo allora l'evento

$$\bigcap_{n > n_0} |\hat{\theta}_n - \theta_0| < \varepsilon = \bigcap_{n > n_0} l_n(\theta_0) - l_n(\theta_j) > \nu > 0 \quad \forall j \neq 0$$

questo perché Θ è finito! Ponendo $\varepsilon < \min\{|\theta_j - \theta_0|\}$ allora $|\theta_n - \theta_0| < \varepsilon$ se e solo se

$$\hat{\theta}_n = \theta_0 \quad \Leftrightarrow \quad l_n(\theta_0) - l_n(\theta_j) > 0 \quad \forall \theta_j \neq \theta_0$$

Consistenza dello SMV: dimostrazione, II parte

Consideriamo ora l'evento

$$A_j = \{Y | I_n(\theta_0) - I_n(\theta_j) > \nu \forall n > n_0\}$$

dal fatto che $I_n(\theta_0) - I_n(\theta) \xrightarrow{q.c.} +\infty$ discende che

$$P\left(\bigcap_{n>n_0} I_n(\theta_0) - I_n(\theta_j) > \varepsilon\right) \rightarrow 1$$

quindi, qualunque sia $\delta > 0$ esiste n_0 tale che

$$P(A_j) > 1 - \delta$$

Consistenza dello SMV: dimostrazione, II parte

Scriviamo allora

$$P\left(\bigcap_{n>n_0} I_n(\theta_0) - I_n(\theta_j) > \nu > 0 \forall j \neq 0\right) = P\left(\bigcap_j A_j\right)$$

si ha

$$P\left(\bigcap_j A_j\right) = 1 - P\left(\bigcup_j \bar{A}_j\right) \geq 1 - \sum_j P(\bar{A}_j)$$

si ha però che, qualunque sia j e qualunque sia δ si trova n_0 t.c. $P(A_j) > 1 - \delta$, quindi $P(\bar{A}_j) < \delta$ e perciò

$$P\left(\bigcap_j A_j\right) \geq 1 - \sum_j P(\bar{A}_j) > 1 - m\delta$$

essendo m fissato e δ arbitrario si ha la tesi.

Indice

- 1 Precisione dello SMV
- 2 Informazione
- 3 Proprietà dello SMV
- 4 Proprietà asintotiche dello SMV**
 - Premesse: convergenze in probabilità
 - Proprietà asintotiche degli stimatori
 - Consistenza dello SMV
 - Distribuzione asintotica dello SMV**

Distribuzione asintotica dello SMV

Teorema: distribuzione asintotica dello SMV

Se siamo in un problema regolare di stima e

- (i) Y_1, \dots, Y_n è un campione $IID(p)$
- (ii) $0 < i(\theta) < +\infty$
- (iii) $\exists M(Y; \theta)$ t.c.
 - $\left| \frac{\partial^3}{\partial \theta^3} \log f(y; \theta) \right| < M(y; \theta)$
 - $\exists E(M(Y; \theta)) < M_0 < +\infty$

allora

$$\sqrt{n}(\hat{\theta} - \theta_0) \rightarrow_d \mathcal{N}(0, i(\theta_0)^{-1}) \quad (2)$$

dove θ_0 rappresenta il vero valore del parametro

Distribuzione asintotica dello SMV: dimostrazione

Essendo in un problema regolare si ha

$$0 = l'(\hat{\theta}).$$

Se sviluppiamo l' in serie di Taylor intorno a θ_0 otteniamo

$$0 = l'(\hat{\theta}) = l'(\theta_0) + l''(\theta_0)(\hat{\theta} - \theta_0) + \frac{1}{2}l'''(\tilde{\theta})(\hat{\theta} - \theta_0)^2$$

per qualche $\tilde{\theta}$ compreso tra $\hat{\theta}$ e θ_0 . Allora

$$(\hat{\theta} - \theta_0) = \frac{-l'(\theta_0)}{l''(\theta_0) + \frac{1}{2}l'''(\tilde{\theta})(\hat{\theta} - \theta_0)}$$

ossia, moltiplicando ambo i membri per \sqrt{n}

$$\sqrt{n}(\hat{\theta} - \theta_0) = \frac{-\frac{l'(\theta_0)}{\sqrt{n}}}{\frac{l''(\theta_0)}{n} + \frac{1}{2n}l'''(\tilde{\theta})(\hat{\theta} - \theta_0)} = \frac{A_n}{B_n + C_n}$$

Distribuzione asintotica dello SMV: dimostrazione

per A_n : Si ha

$$A_n = -\frac{l''(\theta_0)}{\sqrt{n}} = -\frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{\partial}{\partial \theta} \log f(y_i; \theta) \Big|_{\theta=\theta_0}$$

che è somma di v.a. IID con

$$E \left(\frac{\partial}{\partial \theta} \log f(y_i; \theta) \Big|_{\theta=\theta_0} \right) = 0,$$

$$V \left(\frac{\partial}{\partial \theta} \log f(y_i; \theta) \Big|_{\theta=\theta_0} \right) = i(\theta_0),$$

perciò, in virtù del teorema del limite centrale, si ha

$$A_n = -\frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{\partial}{\partial \theta} \log f(y_i; \theta) \Big|_{\theta=\theta_0} \rightarrow_d \mathcal{N}(0, i(\theta_0))$$

Distribuzione asintotica dello SMV: dimostrazione

per B_n : Si ha

$$B_n = \frac{l'''(\theta_0)}{n} = \frac{1}{n} \sum_{i=1}^n \frac{\partial^2}{\partial \theta^2} \log f(y_i; \theta) \Big|_{\theta=\theta_0} \rightarrow_p i(\theta_0)$$

per la legge dei grandi numeri.

per C_n : Usando l'ipotesi (iv) si ha

$$\begin{aligned} C_n &= \frac{1}{2n} l''''(\tilde{\theta})(\hat{\theta} - \theta_0) \leq \frac{1}{2n} (\hat{\theta} - \theta_0) \sum_{i=1}^n \left| \frac{\partial^3}{\partial \theta^3} \log f(y; \theta) \right| \\ &< \frac{1}{2n} (\hat{\theta} - \theta_0) \sum M(Y_i; \theta) \rightarrow_p 0 \end{aligned}$$

in quanto $(\hat{\theta} - \theta_0)$ è $o_p(1)$ e $\frac{1}{n} \sum M(Y_i; \theta)$ è $O_p(1)$.

Distribuzione asintotica dello SMV: dimostrazione

In sintesi

- $A_n \xrightarrow{d} \mathcal{N}(0, i(\theta_0))$
- $B_n \xrightarrow{q.c.} i(\theta_0)$
- $C_n \xrightarrow{P} 0$

quindi

$$B_n + C_n \xrightarrow{P} i(\theta_0)$$

e il rapporto converge in distribuzione a una $\mathcal{N}((0), i(\theta_0)^{-1})$ (teorema di Slutsky)

$$\sqrt{n}(\hat{\theta} - \theta_0) = \frac{A_n}{B_n + C_n} \xrightarrow{d} \frac{\mathcal{N}(0, i(\theta_0))}{i(\theta_0)} = \mathcal{N}((0), i(\theta_0)^{-1})$$

Versione generale

La normalità asintotica vale anche sotto condizioni più generali, in particolare

- il parametro θ può avere dimensione $d > 1$, il risultato è allora

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} \mathcal{N}((0), i(\theta_0)^{-1})$$

dove $i(\theta_0)$ è la matrice di informazione per la singola osservazione.

- il campione può essere non identicamente distribuito
- il campione può non essere indipendente

Negli ultimi due casi serve che l'informazione portata dal campione aumenti indefinitamente all'aumentare della numerosità campionaria, si fanno poi ipotesi tecniche aggiuntive.

Versioni operative

In pratica, il risultato

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} \mathcal{N}((0), i(\theta_0)^{-1})$$

viene usato dicendo che approssimativamente (n sufficientemente grande ma, ovviamente, finito)

$$\hat{\theta} \sim \mathcal{N}(\theta_0, \mathcal{I}(\theta_0)^{-1})$$

Inoltre, non conoscendo θ_0 , questo viene sostituito dalla stima $\hat{\theta}^{\text{OSS}}$ e allora si ha

$$\hat{\theta} \sim \mathcal{N}(\theta_0, \mathcal{I}(\hat{\theta}^{\text{OSS}})^{-1})$$

Infine, è possibile usare la matrice di informazione osservata invece dell'attesa

$$\hat{\theta} \sim \mathcal{N}(\theta_0, J(\hat{\theta}^{\text{OSS}})^{-1})$$

Efficienza asintotica

Uno stimatore è efficiente se la sua varianza raggiunge il limite inferiore di Rao Cramer.

Si parla di efficienza asintotica con riferimento alla **varianza della distribuzione asintotica** di uno stimatore, se allora T_n e S_n sono due stimatori e

$$\sqrt{n}(T_n - \theta) \rightarrow_d \mathcal{N}(0, \nu(\theta))$$

$$\sqrt{n}(S_n - \theta) \rightarrow_d \mathcal{N}(0, \kappa(\theta))$$

si dice che T_n è asintoticamente più efficiente di S_n se $\nu(\theta) < \kappa(\theta)$.

Uno stimatore che abbia una distribuzione asintotica con varianza pari al limite di Cramer Rao si dice asintoticamente efficiente, lo stimatore di massima verosimiglianza, sotto le condizioni del teorema precedente, raggiunge tale limite.

Esempio: normale

Sia $Y_1, \dots, Y_n \sim IID(\mathcal{N}(\mu, \sigma^2))$, lo SMV è

$$\hat{\mu} = \bar{Y} \quad \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (Y_i - \bar{Y})^2$$

e le matrici di informazione sono

$$\mathcal{I}(\theta) = \begin{bmatrix} n/\sigma^2 & 0 \\ 0 & n/2\sigma^4 \end{bmatrix} \quad J(\hat{\theta}) = \mathcal{I}(\hat{\theta}) = \begin{bmatrix} n/\hat{\sigma}^2 & 0 \\ 0 & n/2\hat{\sigma}^4 \end{bmatrix}$$

Si ha quindi

$$\sqrt{\mathcal{I}(\theta_0)}(\hat{\theta} - \theta_0) = \begin{bmatrix} n/\sigma^2 & 0 \\ 0 & n/2\sigma^4 \end{bmatrix}^{1/2} \begin{bmatrix} \bar{Y} - \mu \\ \hat{\sigma}^2 - \sigma^2 \end{bmatrix} = \begin{bmatrix} \frac{\bar{Y} - \mu}{\sqrt{\sigma^2/n}} \\ \frac{\hat{\sigma}^2 - \sigma^2}{\sqrt{\sigma^4/n}} \end{bmatrix} \xrightarrow{d} \mathcal{N}(0, I)$$

Esempio: normale

Possiamo riscriverlo nella forma

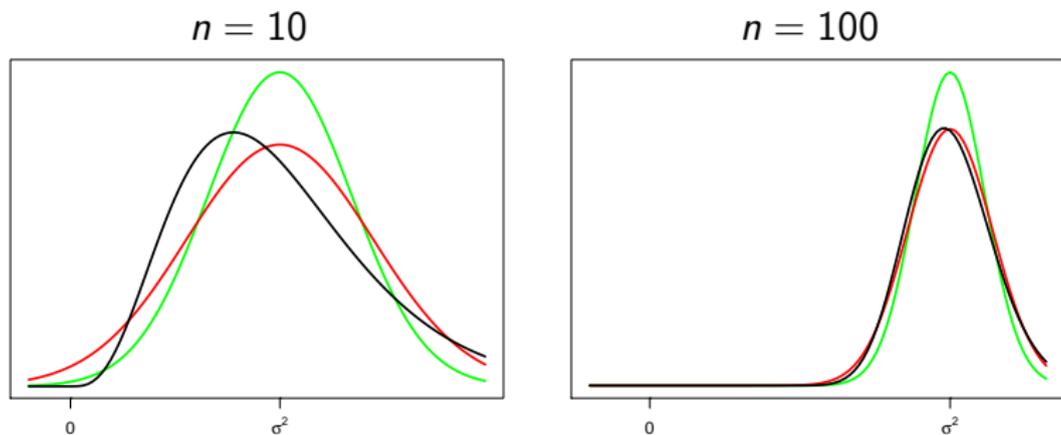
$$\hat{\theta} = \begin{bmatrix} \bar{Y} \\ \hat{\sigma}^2 \end{bmatrix} \dot{\sim} \mathcal{N} \left(\begin{bmatrix} \mu \\ \sigma^2 \end{bmatrix}, \begin{bmatrix} \sigma^2/n & 0 \\ 0 & 2\sigma^4/n \end{bmatrix} \right)$$

che diventa, sostituendo la stima di σ^2

$$\hat{\theta} = \begin{bmatrix} \bar{Y} \\ \hat{\sigma}^2 \end{bmatrix} \dot{\sim} \mathcal{N} \left(\begin{bmatrix} \mu \\ \sigma^2 \end{bmatrix}, \begin{bmatrix} \hat{\sigma}^2/n & 0 \\ 0 & 2\hat{\sigma}^4/n \end{bmatrix} \right)$$

Questa non è la distribuzione esatta dello stimatore, infatti è chiaro che, essendo $\hat{\sigma}^2 > 0$ la sua distribuzione non è una normale (che assegna probabilità positiva a $\hat{\sigma}^2 < 0$).

Esempio normale, distribuzioni esatta e approssimata di $\hat{\sigma}^2$



- nero: distribuzione esatta di $\hat{\sigma}^2$
- rosso: approssimazione normale $\mathcal{N}(\sigma^2, 2\sigma^4/n)$
- verde: approssimazione $\mathcal{N}(\sigma^2, 2\hat{\sigma}^4/n)$

Esempio normale, distribuzioni esatte

La distribuzione esatta dello SMV è

$$\bar{Y} \sim \mathcal{N}(\mu, \sigma^2/n), \quad \frac{n}{\sigma^2} \hat{\sigma}^2 \sim \chi_{n-1}^2$$

e \bar{Y} e $\hat{\sigma}^2$ sono indipendenti.

Definiamo la matrice ortogonale

$$A = \begin{bmatrix} \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \cdots & \cdots & \cdots & \cdots & \frac{1}{\sqrt{n}} \\ \frac{1}{\sqrt{1 \cdot 2}} & -\frac{1}{\sqrt{1 \cdot 2}} & 0 & \cdots & \cdots & \cdots & \cdots & 0 \\ \cdots & \cdots \\ \frac{1}{\sqrt{(i-1)i}} & \frac{1}{\sqrt{(i-1)i}} & \cdots & \frac{1}{\sqrt{(i-1)i}} & -\frac{i-1}{\sqrt{(i-1)i}} & 0 & \cdots & 0 \\ \cdots & 0 \\ \frac{1}{\sqrt{(n-1)n}} & \frac{1}{\sqrt{(n-1)n}} & \cdots & \frac{1}{\sqrt{(n-1)n}} & \cdots & \cdots & \frac{1}{\sqrt{(n-1)n}} & -\frac{n-1}{\sqrt{(n-1)n}} \end{bmatrix}$$

e consideriamo la v.a.

$$Z = AY$$

Esempio normale, distribuzioni esatte

Si ha che

- $Z_1 = \frac{1}{\sqrt{n}} \sum_{i=1}^n Y_i = \sqrt{n}\bar{Y}$
- $Z_2^2 + \dots + Z_n^2 = Z^T Z - Z_1^2 = \sum_{j=1}^n Y_j^2 - n\bar{Y}^2 = n\hat{\sigma}^2$

Per $Z = AY$ si ha che

- $\mu_z = E(Z) = \mu A1_n = (\mu\sqrt{n}, 0, \dots, 0)^T$
- $V(Z) = V(AY) = AV(Y)A^T = \sigma^2 AA^T = \sigma^2 I_n$
- Z è normale in quanto trasformato di un vettore normale

Esempio normale, distribuzioni esatte

Si ha che

- $Z_1 = \frac{1}{\sqrt{n}} \sum_{i=1}^n Y_i = \sqrt{n}\bar{Y}$
- $Z_2^2 + \dots + Z_n^2 = Z^T Z - Z_1^2 = \sum_{j=1}^n Y_j^2 - n\bar{Y}^2 = n\hat{\sigma}^2$

Per $Z = AY$ si ha che

- $\mu_z = E(Z) = \mu A1_n = (\mu\sqrt{n}, 0, \dots, 0)^T$
- $V(Z) = V(AY) = AV(Y)A^T = \sigma^2 AA^T = \sigma^2 I_n$
- Z è normale in quanto trasformato di un vettore normale

$$Z = AY \sim \mathcal{N}_n(\mu_z, \sigma^2 I_n)$$

Esempio normale, distribuzioni esatte

Si ha che

- $Z_1 = \frac{1}{\sqrt{n}} \sum_{i=1}^n Y_i = \sqrt{n}\bar{Y}$
- $Z_2^2 + \dots + Z_n^2 = Z^T Z - Z_1^2 = \sum_{j=2}^n Y_j^2 - n\bar{Y}^2 = n\hat{\sigma}^2$

Per $Z = AY$ si ha che

$$Z = AY \sim \mathcal{N}_n(\mu_Z, \sigma^2 I_n)$$

da questo segue che

$$\frac{n}{\sigma^2} \hat{\sigma}^2 = \frac{1}{\sigma^2} \sum_{j=2}^n Z_j^2 = \frac{1}{\sigma^2} (Z_2^2 + \dots + Z_n^2) \sim \chi_{n-1}^2$$

Inoltre, essendo le Z_j indipendenti, $\hat{\sigma}^2$, che dipende da Z_2, \dots, Z_n , è indipendente da $Z_1 = \sqrt{n}\bar{Y}$.

Distribuzione asintotica per una proporzione campionaria

Sia $Y_1, \dots, Y_n \sim IID \text{ Bernoulli}(\theta)$, posto $S = \sum_{i=1}^n Y_i$ si è già visto che

$$\hat{\theta} = \frac{S}{n}; \quad l''(\theta) = -\frac{S}{\theta^2} - \frac{n-S}{(1-\theta)^2}$$

da cui

$$J(\hat{\theta}) = \frac{n}{\hat{\theta}(1-\hat{\theta})}; \quad \mathcal{I}(\theta) = \frac{n}{\theta(1-\theta)}$$

Si ha perciò

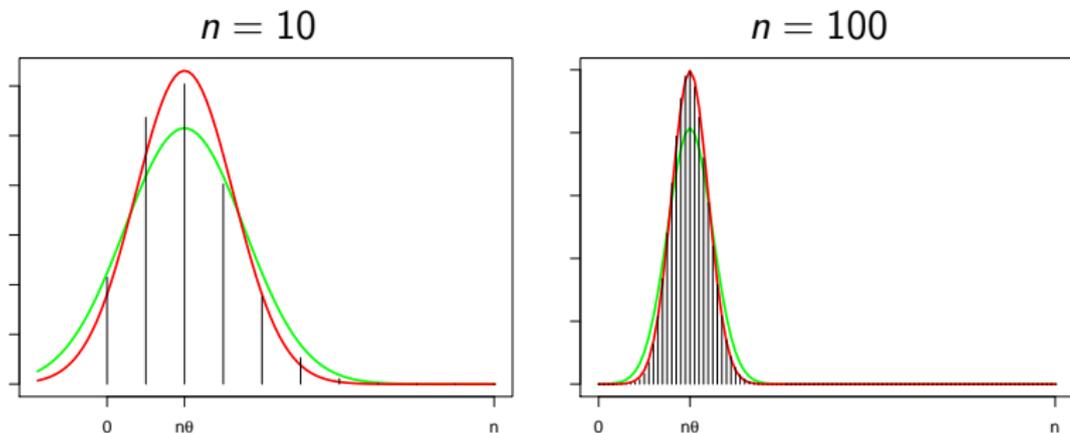
$$\hat{\theta} = \frac{S}{n} \sim \mathcal{N}\left(\theta, \frac{\theta(1-\theta)}{n}\right); \quad \hat{\theta} = \frac{S}{n} \sim \mathcal{N}\left(\theta, \frac{\hat{\theta}(1-\hat{\theta})}{n}\right)$$

$$n\hat{\theta} = S \sim \mathcal{N}(n\theta, n\theta(1-\theta)); \quad n\hat{\theta} = S \sim \mathcal{N}(n\theta, n\hat{\theta}(1-\hat{\theta}))$$

mentre S è una variabile discreta e

$$P(S = s) = \binom{n}{ns} \theta^{nS} (1-\theta)^{n(1-S)}$$

Esempio Bernoulli, distribuzioni esatta e approssimata di S



- nero: distribuzione esatta di S
- rosso: approssimazione normale $\mathcal{N}(n\theta, n\theta(1-\theta))$
- verde: approssimazione $\mathcal{N}(n\theta, nS(1-S))$

Esempio: esponenziale

Riprendiamo l'esempio con l'esponenziale: $Y_1, \dots, Y_n \sim_{IID} \theta e^{-\theta y}$, lo SMV è

$$\hat{\theta} = \bar{Y}^{-1}$$

e

$$\mathcal{I}(\theta) = n\theta^{-2}$$

Applicando il risultato asintotico a $\hat{\theta}$ otteniamo

$$\hat{\theta} = \bar{Y}^{-1} \sim \mathcal{N}(\theta_0, \theta^2/n)$$

Esempio: esponenziale

Riprendiamo l'esempio con l'esponenziale: $Y_1, \dots, Y_n \sim_{IID} \theta e^{-\theta y}$, lo SMV è

$$\hat{\theta} = \bar{Y}^{-1}$$

e

$$\mathcal{I}(\theta) = n\theta^{-2}$$

Applicando il risultato asintotico a $\hat{\theta}$ otteniamo

$$\hat{\theta} = \bar{Y}^{-1} \sim \mathcal{N}(\theta_0, \theta^2/n)$$

Consideriamo ora la riparametrizzazione (già discussa) $\psi = 1/\theta$, con la quale $\hat{\psi} = \bar{Y}$ e $\mathcal{I}(\psi) = n\psi^{-2}$, applicando il risultato asintotico a $\hat{\psi}$ si ha

$$\hat{\psi} = \bar{Y} \sim \mathcal{N}(\psi_0, \psi^2/n)$$