

Università di Siena

Corso di STATISTICA

Parte seconda: Teoria della stima

Andrea Garulli, Antonello Giannitrapani, Simone Paoletti

Master E^2C

Centro per lo Studio dei Sistemi Complessi

Università di Siena

email: statistica@csc.unisi.it

2.1 Teoria della stima

- ✓ Stima parametrica
  - ◇ Stima puntuale
  - ◇ Stima bayesiana
- ✓ Problema della stima puntuale
- ✓ Proprietà degli stimatori
  - ◇ Correttezza
  - ◇ Consistenza
  - ◇ Efficienza
- ✓ Stime a minima varianza
- ✓ Limite di Cramer-Rao
- ✓ Stima di massima verosimiglianza
- ✓ Intervalli di confidenza

Variabili aleatorie scalari

Sia X una variabile aleatoria (v.a.) scalare

$$X : \Omega \to \mathbb{R}$$

definita sull'insieme di eventi elementari Ω.
La notazione

$$X \sim F_X(x),\ f_X(x)$$

denota che:

- $F_X(x)$ è la funzione *distribuzione di probabilità* di X

$$F_X(x) = P\{X \leq x\}, \quad \forall x \in \mathbb{R}$$

- $f_X(x)$ è la funzione *densità di probabilità* di X

$$F_X(x) = \int_{-\infty}^{x} f_X(\sigma)\, d\sigma, \quad \forall x \in \mathbb{R}$$

Variabili aleatorie vettoriali

Sia $X = (X_1, \ldots, X_n)$ una variabile aleatoria vettoriale

$$X : \Omega \to \mathbb{R}^n$$

definita sull'insieme di eventi elementari Ω.
La notazione

$$X \sim F_X(x),\ f_X(x)$$

denota che:

- $F_X(x)$ è la funzione *distribuzione di probabilità congiunta* di X

$$F_X(x) = P\{X_1 \leq x_1, \ldots, X_n \leq x_n\}, \quad \forall x = (x_1, \ldots, x_n) \in \mathbb{R}^n$$

- $f_X(x)$ è la funzione *densità di probabilità congiunta* di X

$$F_X(x) = \int_{-\infty}^{x_1} \ldots \int_{-\infty}^{x_n} f_X(\sigma_1, \ldots, \sigma_n)\, d\sigma_1 \ldots d\sigma_n, \quad \forall x \in \mathbb{R}^n$$

Momenti di una distribuzione

- primo momento (*media*)

$$m_X = E[X] = \int_{-\infty}^{+\infty} x\, f_X(x)\, dx$$

- secondo momento centrato (*varianza*)

$$\sigma_X^2 = \mathrm{Var}(X) = E\left[(X - m_X)^2\right] = \int_{-\infty}^{+\infty} (x - m_X)^2\, f_X(x)\, dx$$

Esempio Si definisce *densità normale*, e si indica con $N(m, \sigma^2)$, la densità

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}}$$

in cui m è la media della distribuzione e σ^2 è la varianza.

Campionamento di una variabile aleatoria

Si considerino n ripetizioni indipendenti dello stesso esperimento casuale. L'osservazione è dunque costituita da una successione $X_1, \ldots, X_n$ di v.a. indipendenti ed aventi la stessa densità di probabilità $f(\cdot)$.

Definizione 1 *Una successione $X_1, \ldots, X_n$ di v.a. indipendenti e identicamente distribuite* (i.i.d.) *si dice* campione *di dimensione n di densità $f(\cdot)$.*

Si definisca la v.a. vettoriale $X=(X_1, \ldots, X_n)$. *Qual è la densità di probabilità congiunta di X?*

Dato che le v.a. $X_1, \ldots, X_n$ sono indipendenti, risulta:

$$f_X(x) = \prod_{i=1}^{n} f(x_i), \quad x = (x_1, \ldots, x_n)$$

Stima parametrica

Problema. Stimare il valore del parametro $\theta \in \mathbb{R}^p$ sulla base di un'osservazione x della v.a. $X \in \mathbb{R}^n$.

Due possibili scenari:

a. *Stima puntuale*
   La densità di probabilità di X dipende dal parametro incognito θ

b. *Stima bayesiana*
   Il parametro incognito θ è una variabile aleatoria, ed è nota la densità di probabilità congiunta di X e θ

Stima puntuale

- La distribuzione (o la densità) di probabilità della v.a. X ha una *forma funzionale nota*, che dipende da un vettore θ di *parametri incerti*

$$X \sim F_X^\theta(x),\ f_X^\theta(x)$$

- $\Theta \subseteq \mathbb{R}^p$ denota lo *spazio dei parametri*, in cui assume valori il vettore dei parametri θ
- $\mathcal{X} \subseteq \mathbb{R}^n$ denota lo *spazio delle osservazioni*, in cui assume valori la variabile aleatoria X

Problema della stima puntuale

Il problema della *stima puntuale* consiste nello stimare il parametro incognito θ sulla base di un'osservazione x della v.a X.

Definizione 2 *Uno* stimatore *del parametro* θ *è una funzione*

$$T : \mathcal{X} \longrightarrow \Theta$$

Dare uno stimatore $T(\cdot)$ corrisponde a fissare la regola che, se si osserva x, allora si stima θ con la quantità $\hat{\theta} = T(x)$.

In base alla definizione data, la classe dei possibili stimatori è infinita!

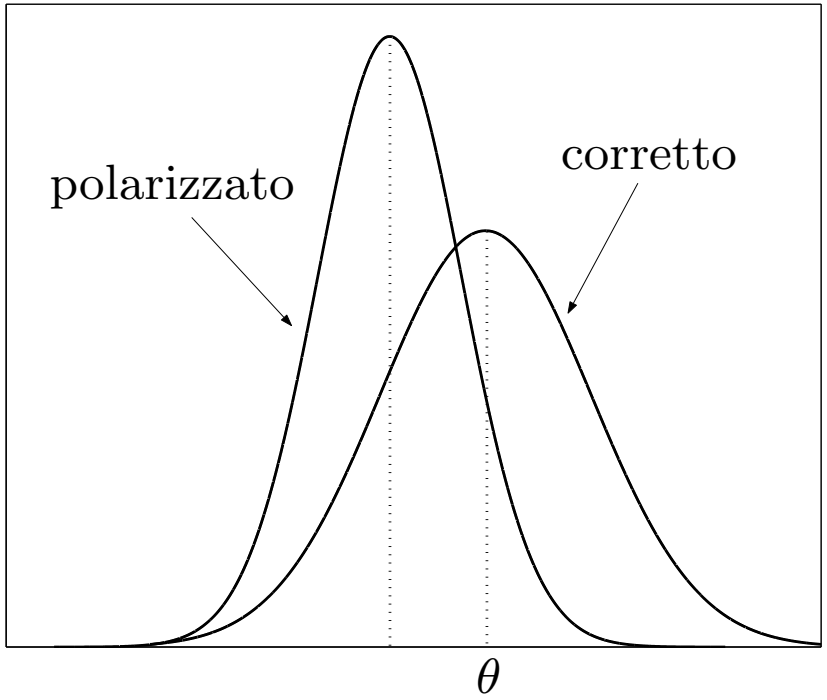
Una prima questione consiste quindi nello stabilire dei criteri per decidere quali stimatori siano "buoni" e quali no, ovvero per confrontare due stimatori.

Quale criterio conviene adottare per la scelta di un buon stimatore?

Stima non polarizzata

Ovviamente, il risultato migliore che uno stimatore può fornire è che la stima coincida con il valore vero del parametro. Dato che la stima è una v.a., è allora ragionevole richiedere che ciò accada in media.

Definizione 3 *Uno stimatore $T(\cdot)$ del parametro θ si dice* corretto *(o* non polarizzato*) se $E^{\theta}[T(X)] = \theta$, $\forall \theta \in \Theta$.*



Esempi

- $X_1, \ldots, X_n$ v.a. *i.d.* con media m. La *media campionaria*

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

è una stima non polarizzata di m. Infatti

$$E\big[\bar{X}\big] = \frac{1}{n} \sum_{i=1}^{n} E[X_i] = m$$

- $X_1, \ldots, X_n$ v.a. *i.i.d.* con varianza σ^2. La *varianza campionaria*

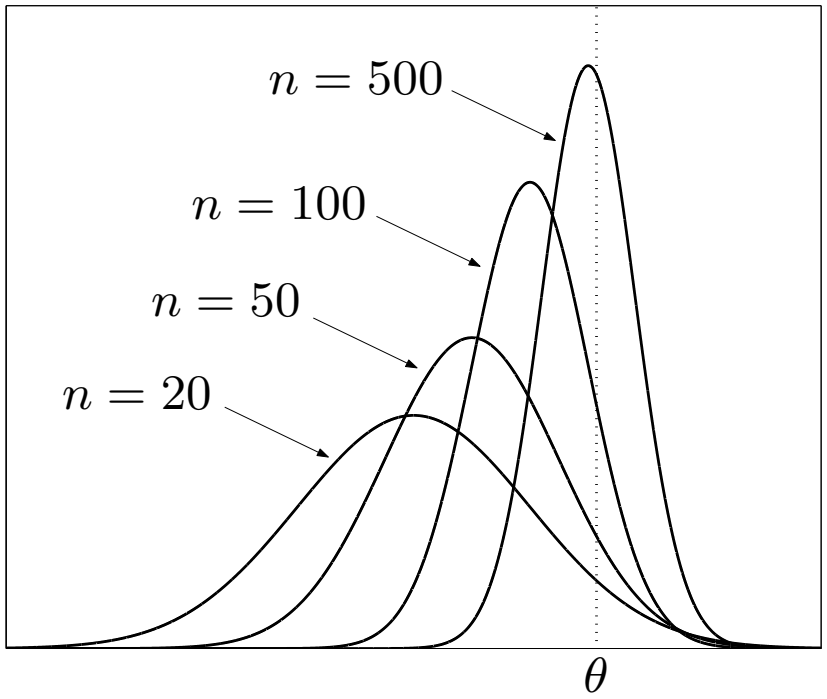
$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$$

è una stima non polarizzata di σ^2.

Stima consistente

Definizione 4 *Sia* $\{X_i\}_{i=1}^{\infty}$ *una successione di v.a.. La successione di stimatori* $T_n{=}T_n(X_1,\ldots,X_n)$ *si dice* consistente *per* θ *se* T_n *converge a* θ *in probabilità per ogni* $\theta \in \Theta$*, cioè*

$$\lim_{n\to\infty} P^{\theta}\left\{\|T_n - \theta\| > \varepsilon\right\} = 0 \quad , \quad \forall \varepsilon > 0 \quad , \quad \forall \theta \in \Theta$$



Esempio

$X_1, \ldots, X_n$ v.a. *i.i.d.* con media m e varianza σ^2. La *media campionaria*

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

è uno stimatore consistente di m.

Vale infatti il seguente teorema.

Teorema 1 (Legge dei grandi numeri) *Sia* $\{X_i\}_{i=1}^{\infty}$ *una successione di v.a. indipendenti e identicamente distribuite con media* m *e varianza finita. Allora la media campionaria* $\bar{X}$ *converge a* m *in probabilità.*

Osservazione Sappiamo che la *media campionaria* è una stima non polarizzata di m. Inoltre, sotto le ipotesi del Teorema 1, risulta

$$\mathrm{Var}(\bar{X}) = \frac{\sigma^2}{n} \to 0 \quad \text{per} \quad n \to \infty$$

Errore quadratico medio

Si consideri uno stimatore $T(\cdot)$ del parametro scalare θ.

Definizione 5 *Si definisce* errore quadratico medio *la quantità*

$$E^{\theta}\left[(T(X)-\theta)^2\right]$$

Se lo stimatore $T(\cdot)$ è corretto, l'errore quadratico medio coincide con la varianza della stima.

Definizione 6 *Dati due stimatori* $T_1(\cdot)$ *e* $T_2(\cdot)$ *del parametro* θ, $T_1(\cdot)$ *si dice* preferibile *a* $T_2(\cdot)$ *se*

$$E^{\theta}\left[(T_1(X)-\theta)^2\right] \leq E^{\theta}\left[(T_2(X)-\theta)^2\right] \quad , \quad \forall \theta \in \Theta$$

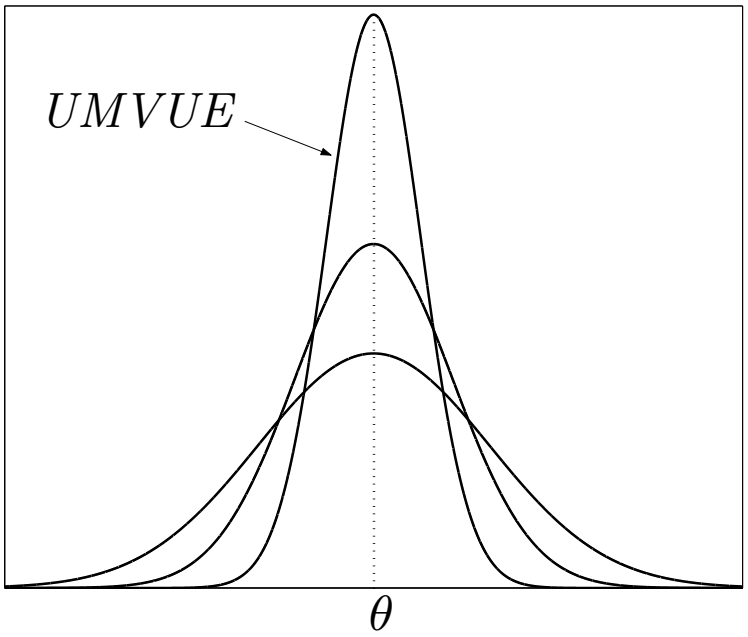
Restringendo l'attenzione agli stimatori corretti, cerchiamo quello, se esiste, con minima varianza per ogni valore di θ.

Stima non polarizzata a minima varianza

Definizione 7 *Uno stimatore corretto $T^*(\cdot)$ del parametro scalare θ viene detto* efficiente *(o* UMVUE*, uniformly minimum variance unbiased estimator) se*

$$E^\theta\left[(T^*(X)-\theta)^2\right] \leq E^\theta\left[(T(X)-\theta)^2\right], \quad \forall\theta\in\Theta$$

per ogni stimatore corretto $T(\cdot)$ di θ.



Migliore stima lineare

Restringiamo l'attenzione alla classe degli stimatori lineari, ossia stimatori del tipo

$$T(x) = \sum_{i=1}^{n} a_i x_i \quad , \quad a_i \in \mathbb{R}$$

Definizione 8 *Uno stimatore lineare corretto $T^*(\cdot)$ del parametro scalare θ viene detto* BLUE *(best linear unbiased estimator) se*

$$E^\theta\left[(T^*(X) - \theta)^2\right] \leq E^\theta\left[(T(X) - \theta)^2\right], \quad \forall \theta \in \Theta$$

per ogni stimatore lineare corretto $T(\cdot)$ di θ.

Esempio X_i v.a. indipendenti con media m e varianza σ_i^2, $i = 1, \ldots, n$.

$$\hat{X} = \frac{1}{\sum_{i=1}^{n} \frac{1}{\sigma_i^2}} \sum_{i=1}^{n} \frac{1}{\sigma_i^2} X_i$$

risulta essere la migliore stima lineare non polarizzata di m.

Limite di Cramer-Rao

Il *limite di Cramer-Rao* stabilisce un limite inferiore per la varianza di ogni stimatore corretto del parametro θ.

Teorema 2 *Sia $T(\cdot)$ uno stimatore corretto del parametro scalare θ, e si supponga che lo spazio delle osservazioni $\mathcal{X}$ sia indipendente da θ. Allora (sotto alcune ipotesi di regolarità...)*

$$E^{\theta}\left[(T(X)-\theta)^2\right] \geq \left[I_n(\theta)\right]^{-1}$$

dove $I_n(\theta)=E^{\theta}\left[\left(\frac{\partial \ln f_X^{\theta}(X)}{\partial \theta}\right)^2\right]$ *(*quantità di informazione di Fisher*).*

Osservazione La valutazione di $I_n(\theta)$ richiede generalmente la conoscenza di θ; quindi il valore del limite di Cramer-Rao è tipicamente sconosciuto all'utente. Esso può comunque essere usato per dimostrare che uno stimatore non polarizzato è efficiente.

Limite di Cramer-Rao

Nel caso in cui il parametro θ sia vettoriale, e $T(\cdot)$ ne è uno stimatore corretto, risulta

$$E^{\theta}\left[(T(X)-\theta)\,(T(X)-\theta)'\right] \geq \left[I_n(\theta)\right]^{-1} \tag{1}$$

dove la disuguaglianza è da intendersi in senso matriciale.

$I_n(\theta)$ denota la *matrice di informazione di Fisher*

$$I_n(\theta) = E^{\theta}\left[\left(\frac{\partial \ln f_X^{\theta}(X)}{\partial \theta}\right)'\left(\frac{\partial \ln f_X^{\theta}(X)}{\partial \theta}\right)\right]$$

La matrice a sinistra nella (1) è la *matrice di covarianza dello stimatore*.

Limite di Cramer-Rao

Se le v.a. $X_1, \ldots, X_n$ sono *i.i.d.*, risulta

$$I_n(\theta) = nI_1(\theta)$$

Dunque, per θ fissato, il limite di Cramer-Rao migliora come $\frac{1}{n}$ all'aumentare della dimensione n del campione.

Esempio $X_1, \ldots, X_n$ v.a. *i.i.d.* con media m e varianza σ^2. Risulta

$$E\left[\left(\bar{X} - m\right)^2\right] = \frac{\sigma^2}{n} \geq \left[I_n(\theta)\right]^{-1} = \frac{\left[I_1(\theta)\right]^{-1}}{n}$$

dove $\bar{X}$ denota la media campionaria. Se le v.a. $X_1, \ldots, X_n$ seguono una densità normale, risulta anche $I_1(\theta) = \frac{1}{\sigma^2}$.

Essendo dunque raggiunto il limite di Cramer-Rao, *nel caso di v.a. normali i.i.d. la media campionaria è uno stimatore efficiente della media.*

Stima di massima verosimiglianza

Si consideri una v.a. $X \sim f_X^\theta(x)$, e una sua osservazione x. Si definisce *funzione di verosimiglianza* la funzione di θ (x è fissato!)

$$L(\theta|x) = f_X^\theta(x)$$

Una stima ragionevole di θ è quel valore del parametro che massimizza la probabilità dell'evento osservato.

Definizione 9 *Si definisce* stimatore di massima verosimiglianza *del parametro θ lo stimatore*

$$T_{ML}(x) = \arg\max_{\theta \in \Theta} L(\theta|x)$$

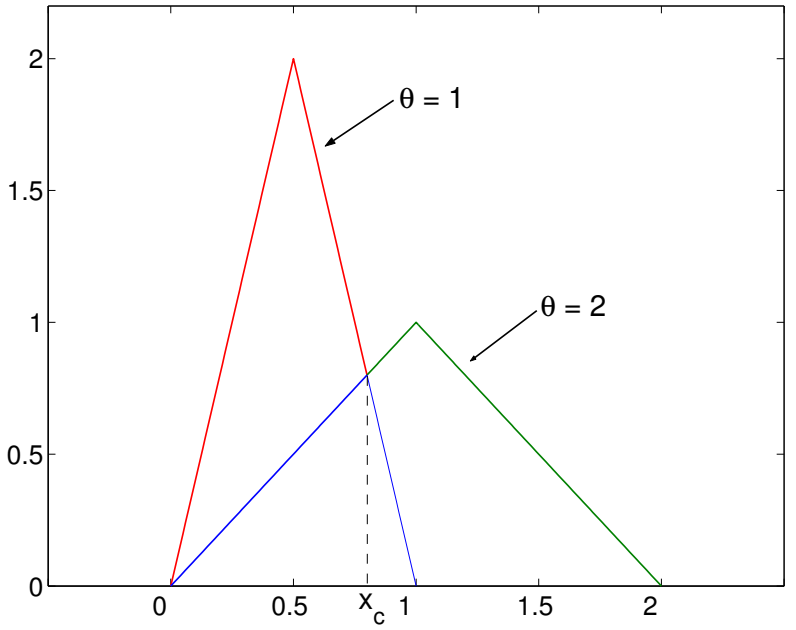
Osservazione I punti di massimo delle funzioni $L(\theta|x)$ e $\ln L(\theta|x)$ coincidono. In alcuni casi può risultare conveniente cercare i punti di massimo di $\ln L(\theta|x)$.

Esempio Si consideri la v.a. X con densità di probabilità triangolare

$$f_X^\theta(x) = \begin{cases} \frac{4}{\theta^2}x & \text{se } 0 \le x \le \frac{\theta}{2} \\ \frac{4}{\theta^2}(\theta - x) & \text{se } \frac{\theta}{2} < x \le \theta \\ 0 & \text{altrimenti} \end{cases}$$

dove $\theta \in \{1, 2\}$. Posto $x_c = \frac{4}{5}$, risulta

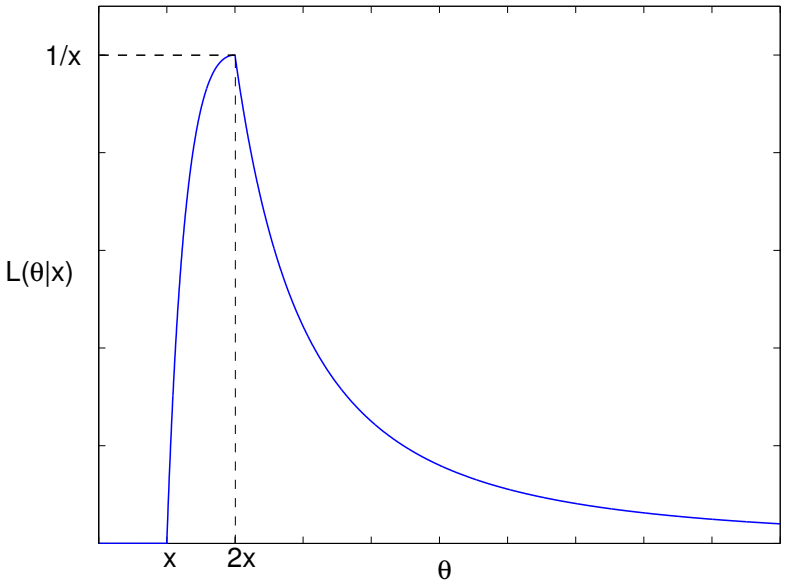
$$\begin{aligned} T_{ML}(x) &= \arg \max_{\theta \in \{1,2\}} f_X^\theta(x) \\ &= \begin{cases} 1 & \text{se } x \le x_c \\ 2 & \text{se } x > x_c \end{cases} \end{aligned}$$

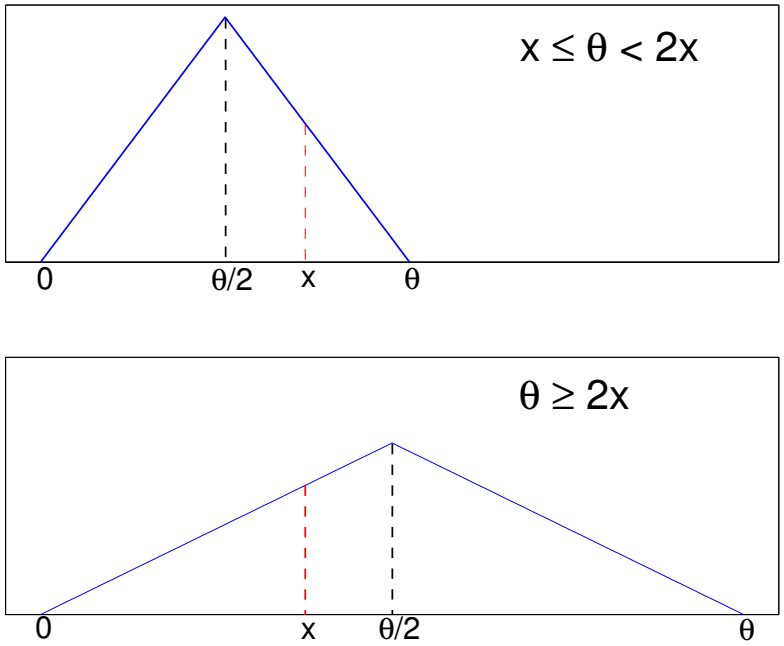


Esempio (continua) Se $\theta \in (0, \infty)$, ed essendo per $x > 0$

$$L(\theta|x) = \begin{cases} 0 & \text{se } 0 < \theta < x \\ \frac{4}{\theta^2}(\theta - x) & \text{se } x \leq \theta < 2x \\ \frac{4}{\theta^2} x & \text{se } \theta \geq 2x \end{cases}$$

risulta $T_{ML}(x) = \arg \max_{\theta \in (0,\infty)} L(\theta|x) = 2x$





Proprietà della stima di massima verosimiglianza

Si consideri il caso di parametro θ scalare.

Teorema 3 *Sotto le ipotesi di validità del limite di Cramer-Rao, se esiste uno stimatore $T^*(\cdot)$ che raggiunge il limite di Cramer-Rao, allora esso coincide con lo stimatore di massima verosimiglianza $T_{ML}(\cdot)$.*

Esempio $X_i \sim N(m, \sigma_i^2)$ indipendenti, σ_i^2 nota, $i = 1, \ldots, n$. La stima

$$\hat{X} = \frac{1}{\sum_{i=1}^{n} \frac{1}{\sigma_i^2}} \sum_{i=1}^{n} \frac{1}{\sigma_i^2} X_i$$

di m è corretta e tale che $\mathrm{Var}(\hat{X}) = \dfrac{1}{\sum_{i=1}^{n} \frac{1}{\sigma_i^2}}$, mentre $I_n(m) = \sum_{i=1}^{n} \frac{1}{\sigma_i^2}$.

Essendo raggiunto il limite di Cramer-Rao, $\hat{X}$ risulta lo stimatore di massima verosimiglianza di m.

La stima di massima verosimiglianza ha un buon comportamento asintotico.

Teorema 4 *Se le v.a.* $X_1, \ldots, X_n$ *sono* i.i.d., *allora (sotto alcune ipotesi di regolarità...)*

$$\sqrt{I_n(\theta)}\,(T_{ML}(X) - \theta) \longrightarrow N(0,1)$$

in densità di probabilità, asintoticamente per $n \to \infty$.

Il Teorema 4 ci dice che la stima di massima verosimiglianza è

- asintoticamente corretta
- consistente
- asintoticamente efficiente
- asintoticamente normale

Esempio Sia $X_1, \ldots, X_n$ un campione di densità normale con media m e varianza σ^2. La media campionaria

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

è la stima di massima verosimiglianza di m.
Inoltre $\sqrt{I_n(m)}(\bar{X} - m) \sim N(0,1)$, essendo $I_n(m) = \frac{n}{\sigma^2}$.

Osservazione *La stima di massima verosimiglianza può non essere corretta.* Si consideri il caso di un campione $X_1, \ldots, X_n$ di densità normale con varianza σ^2. La stima di massima verosimiglianza di σ^2 risulta

$$\hat{S}^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2$$

che è non corretta, in quanto $E[\hat{S}^2] = \frac{n-1}{n} \sigma^2$.

Intervalli di confidenza

In molti problemi di statistica si è interessati a costruire, sulla base delle osservazioni, un insieme che contenga con probabilità fissata il valore vero (non noto) del parametro.

Definizione 10 *Si definisce* intervallo di confidenza *di livello* $1-\alpha$, $0<\alpha<1$, *per il parametro scalare* θ *una funzione che ad ogni* $x \in \mathcal{X}$ *fa corrispondere un intervallo* $\mathcal{B}(x) \subseteq \Theta$ *tale che*

$$P^{\theta}\{\theta \in \mathcal{B}(x)\} \geq 1-\alpha \quad , \quad \forall \theta \in \Theta$$

Un intervallo di confidenza di livello $1-\alpha$ per θ è dunque un sottoinsieme di Θ tale che, se il risultato dell'osservazione è x, allora $\theta \in \mathcal{B}(x)$ con probabilità non inferiore a $1-\alpha$, qualunque sia $\theta \in \Theta$.

Esempio Sia $X_1, \ldots, X_n$ un campione di densità normale con media m incognita e varianza σ^2 nota. E' noto che $\frac{\sqrt{n}}{\sigma}(\bar{X} - m) \sim N(0,1)$, dove $\bar{X}$ denota la media campionaria.

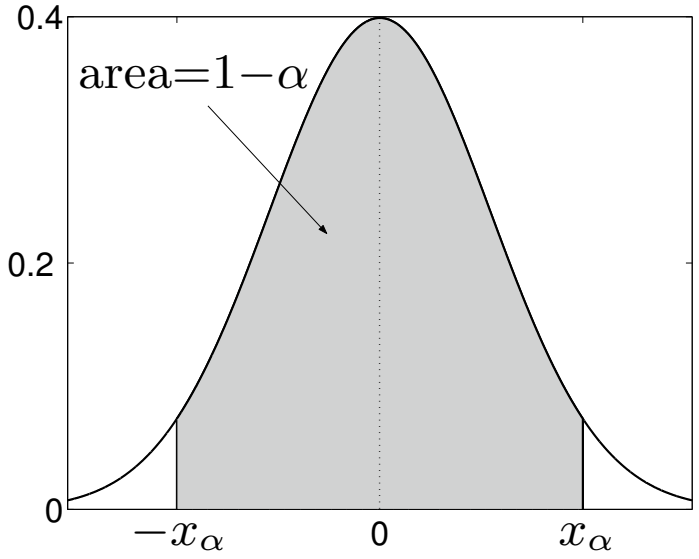
Sia x_α tale che $\int_{-x_\alpha}^{x_\alpha} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = 1 - \alpha$. Risultando dunque

$$1 - \alpha = P\left\{\left|\frac{\sqrt{n}}{\sigma}(\bar{X} - m)\right| \leq x_\alpha\right\} = P\left\{\bar{X} - \frac{\sigma}{\sqrt{n}}\, x_\alpha \leq m \leq \bar{X} + \frac{\sigma}{\sqrt{n}}\, x_\alpha\right\}$$

si ha che

$$\left[\bar{X} - \frac{\sigma}{\sqrt{n}}\, x_\alpha \, , \, \bar{X} + \frac{\sigma}{\sqrt{n}}\, x_\alpha\right]$$

è un intervallo di confidenza di livello $1 - \alpha$ per m.



2.2 Stimatori puntuali e Bayesiani

- ✓ Problemi di stima puntuale
  - ◇ Stimatore di Gauss-Markov
  - ◇ Stimatore ai minimi quadrati
- ✓ Stima Bayesiana
  - ◇ Stima ottima a posteriori
  - ◇ Stima a minimo errore quadratico medio
  - ◇ Stima ottima lineare
- ✓ Problemi di stima

Problemi di stima a massima verosimiglianza

Sia $Y \in \mathbb{R}^m$ un vettore di v.a., tali che

$$Y = U(\theta) + \varepsilon$$

dove

- $\theta \in \mathbb{R}^n$ è il parametro incognito da stimare
- $U(\cdot) : \mathbb{R}^n \to \mathbb{R}^m$ è una funzione nota
- $\varepsilon \in \mathbb{R}^m$ è un vettore di v.a., su cui si fa l'ipotesi

$$\varepsilon \sim \mathcal{N}(0, \Sigma_\varepsilon)$$

Problema: determinare la stima a massima verosimiglianza di θ

$$\hat{\theta}_{ML} \;=\; T_{ML}(Y)$$

Stima ai minimi quadrati

La densità di probabilità dei *dati* Y è pari a

$$f_Y(y) \;=\; f_\varepsilon(y - U(\theta)) \;=\; L(\theta|y)$$

Perciò, dalle ipotesi su ε

$$\begin{aligned} \hat{\theta}_{ML} &= \arg \max_\theta \ln L(\theta|y) \\ &= \arg \min_\theta (y - U(\theta))' \Sigma_\varepsilon^{-1} (y - U(\theta)) \end{aligned}$$

Se la covarianza Σ_ε è nota, si ottiene la *stima ai minimi quadrati pesati*

Poichè in generale $U(\theta)$ è una funzione non lineare, la soluzione si calcola tramite metodi numerici:
MATLAB Optimization Toolbox $\rightarrow$ `>> help optim`

Stimatore di Gauss-Markov

Nel caso in cui la funzione $U(\cdot)$ sia *lineare*, ovvero $U(\theta) = U\theta$ con $U \in \mathbb{R}^{m \times n}$ matrice nota, si ha

$$Y = U\theta + \varepsilon$$

e la stima ML coincide con la *stima di Gauss-Markov*

$$\hat{\theta}_{ML} \;=\; \hat{\theta}_{GM} \;=\; (U'\Sigma_\varepsilon^{-1}U)^{-1}U'\Sigma_\varepsilon^{-1}y$$

Nel caso particolare in cui $\varepsilon \sim \mathcal{N}(0, \sigma^2 I)$ (variabili ε_i indipendenti!), si ha la *stima ai minimi quadrati*

$$\hat{\theta}_{LS} \;=\; (U'U)^{-1}U'y$$

Nota: la stima LS non dipende dal valore di σ, ma solo da U

Esempi di stima ai minimi quadrati

Esempio 1.

$Y_i = \theta + \varepsilon_i, \;\; i = 1, \ldots, m$

ε_i variabili aleatorie indipendenti, con media nulla e varianza σ^2

$\Rightarrow \quad E[Y_i] = \theta$

Si vuole stimare il valore di θ sulla base di m osservazioni delle Y_i

Si ha $Y = U\theta + \varepsilon$ con $U = (1 \; 1 \; \ldots \; 1)'$ e

$$\hat{\theta}_{LS} \;=\; (U'U)^{-1}U'y \;=\; \frac{1}{m}\sum_{i=1}^{m} y_i$$

La stima ai minimi quadrati coincide con la media aritmetica (ed è anche la stima a massima verosimiglianza se le ε_i sono Gaussiane)

Esempi di stima ai minimi quadrati

Esempio 2.

Stesso problema dell'Esempio 1, con $E[\varepsilon_i^2] = \sigma_i^2$, $i = 1, \ldots, m$

In questo caso, $E[\varepsilon\varepsilon'] = \Sigma_\varepsilon = \begin{pmatrix} \sigma_1^2 & 0 & \ldots & 0 \\ 0 & \sigma_2^2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \sigma_m^2 \end{pmatrix}$

$\Rightarrow$ La stima lineare ai minimi quadrati è ancora la media aritmetica

$\Rightarrow$ La stima di Gauss-Markov è

$$\hat{\theta}_{GM} = (U'\Sigma_\varepsilon^{-1}U)^{-1}U'\Sigma_\varepsilon^{-1}y = \frac{1}{\sum_{i=1}^{m}\frac{1}{\sigma_i^2}} \sum_{i=1}^{m}\frac{1}{\sigma_i^2}\, y_i$$

e coincide con la stima a massima verosimiglianza se le ε_i sono Gaussiane

Stima Bayesiana

Stima puntuale (parametrica): stimare il valore di un parametro incognito θ sulla base di osservazioni della variabile aleatoria Y, la cui distribuzione ha una forma funzionale nota che dipende da θ, $f_Y^\theta(y)$

$\rightarrow$ stima a massima verosimiglianza

$\rightarrow$ stimatori UMVUE e BLUE

$\rightarrow$ stimatori ai minimi quadrati

Stima Bayesiana : stimare una variabile aleatoria incognita X, sulla base di osservazioni della variabile aleatoria Y, conoscendo la densità di probabilità congiunta $f_{X,Y}(x,y)$

$\Rightarrow$ stima ottima a posteriori

$\Rightarrow$ stima a minimo errore quadratico medio

$\Rightarrow$ stima ottima lineare

Stima Bayesiana: formulazione del problema

Problema:
Data una variabile aleatoria incognita $X \in \mathbb{R}^n$ e una variabile aleatoria $Y \in \mathbb{R}^m$, della quale sono disponibili osservazioni, determinare una stima di X basata sui valori osservati di Y.

Soluzione: occorre individuare uno stimatore $\hat{X} = T(Y)$, dove $T(\cdot) : \mathbb{R}^m \to \mathbb{R}^n$

Per valutare la qualità della stima è necessario definire un opportuno criterio di stima: in generale, si considera il *funzionale di rischio* di Bayes

$$J_r \;=\; E[d(X, T(Y))] \;=\; \int\int d(x, T(y))\, f_{X,Y}(x, y)\, dx\, dy$$

e si *minimizza* J_r rispetto a tutti i possibili stimatori $T(\cdot)$

$d(X, T(Y)) \to$ "distanza" tra la v.a. incognita X e la sua stima $T(Y)$

Stima ottima a posteriori (MAP)

Sia

$$d(X, T(Y)) = \begin{cases} 0, & \|X - T(Y)\| \leq \varepsilon \\ 1, & \text{altrove} \end{cases}$$

con ε "sufficientemente piccolo", in modo tale che $f_{X|Y}(x|y) \approx K$ (costante) nella regione definita da $\|X - T(Y)\| \leq \varepsilon$ (di volume V_ε).

Si ha così

$$\begin{aligned} J_r &= \int f_Y(y) \left\{ \int d(x, T(y))\, f_{X|Y}(x|y)\, dx \right\} dy \\ &\approx \int f_Y(y) \left\{ 1 - V_\varepsilon \cdot f_{X|Y}(T(y)|y) \right\} dy \end{aligned}$$

per cui J_r è minimizzato se si sceglie $\hat{X} = T(Y)$ in modo da massimizzare $f_{X|Y}(x|y)$.

Stima MAP

La *stima a massima densità di probabilità a posteriori* si ottiene quindi risolvendo

$$\hat{X}_{MAP} = \arg \max_x f_{X|Y}(x|y)$$

Dalla regola di Bayes, si ha

$$\hat{X}_{MAP} = \arg \max_x f_{Y|X}(y|x)\, f_X(x)$$

Osservazioni:

- la stima Bayesiana dipende dalla distribuzione a priori di X
- se la densità di probabilità a priori di X è molto più "piatta" di quella dei dati, la stima ottima a posteriori tende a coincidere con la stima a massima verosimiglianza

Stima a minimo errore quadratico medio (MEQM)

Sia $d(X, T(Y)) = \|X - T(Y)\|^2$.
Si ottiene così la stima a *minimo errore quadratico medio (MEQM)*

$$\hat{X}_{MEQM} = T^*(Y)$$

dove

$$T^*(\cdot) = \arg \min_{T(\cdot)} E[\|X - T(Y)\|^2]$$

Osservazioni:

- si deve risolvere un problema di minimo rispetto a tutti i possibili stimatori $T(\cdot) : \mathbb{R}^m \rightarrow \mathbb{R}^n$
- il valore atteso $E[\cdot]$ viene calcolato rispetto a entrambe le variabili aleatorie X e Y $\rightarrow$ è necessario conoscere la densità di probabilità congiunta $f_{X,Y}(x, y)$

Stima MEQM

Risultato

$$\hat{X}_{MEQM} = E[X|Y]$$

Il *valore atteso condizionato* di X rispetto ad Y coincide con la stima a minimo errore quadratico medio di X basata su osservazioni di Y

Generalizzazioni:

- Sia $Q(X, T(Y)) = E[(X - T(Y))(X - T(Y))']$. Allora: $Q(X, \hat{X}_{MEQM}) \leq Q(X, T(Y))$, per ogni possibile $T(Y)$
- $\hat{X}_{MEQM}$ minimizza ogni funzione scalare monotona crescente di $Q(X, T(Y))$, e in particolare trace(Q) (MEQM) e trace(WQ) con $W > 0$ (MEQM pesato)

Stima ottima lineare (LMEQM)

La stima MEQM richiede la conoscenza della distribuzione di X e Y
$\rightarrow$ Stimatori di struttura più semplice

Stimatori lineari:

$$T(Y) = AY + b$$

$A \in \mathbb{R}^{n \times m}$, $b \in \mathbb{R}^{n \times 1}$: coefficienti dello stimatore (da determinare)

La stima lineare a minimo errore quadratico medio è definita da

$$\hat{X}_{LMEQM} = A^* Y + b^*$$

dove

$$A^*, b^* = \arg \min_{A,b} E[\|X - AY - b\|^2]$$

Stima LMEQM

Risultato

Siano X e Y variabili aleatorie tali che:

$$E[X] = m_X \qquad E[Y] = m_Y$$

$$E\left[\begin{pmatrix} X - m_X \\ Y - m_Y \end{pmatrix}\begin{pmatrix} X - m_X \\ Y - m_Y \end{pmatrix}'\right] = \begin{pmatrix} R_X & R_{XY} \\ R'_{XY} & R_Y \end{pmatrix}$$

Allora

$$\hat{X}_{LMEQM} = m_X + R_{XY}R_Y^{-1}(Y - m_Y)$$

ovvero

$$A^* = R_{XY}R_Y^{-1} \qquad b^* = m_X - R_{XY}R_Y^{-1}m_Y$$

Proprietà della stima LMEQM

- La stima LMEQM non richiede la conoscenza della distribuzione di probabilità congiunta di X e Y, ma solo delle covarianze R_{XY}, R_Y (statistiche del secondo ordine)
- La stima LMEQM soddisfa

$$\begin{aligned} E[(X - \hat{X}_{LMEQM})Y'] &= E[\{X - m_X - R_{XY}R_Y^{-1}(Y - m_Y)\}Y'] \\ &= R_{XY} - R_{XY}R_Y^{-1}R_Y = 0 \end{aligned}$$

  $\Rightarrow$ L'errore di stima ottimo lineare è *scorrelato* dai dati Y

- Se X e Y sono congiuntamente Gaussiane si ha

$$E[X|Y] = m_X + R_{XY}R_Y^{-1}(Y - m_Y)$$

  per cui

$$\hat{X}_{LMEQM} = \hat{X}_{MEQM}$$

  $\Rightarrow$ Nel caso Gaussiano, la stima MEQM è funzione *lineare* delle variabili osservate Y, e quindi coincide con la stima LMEQM

Esempio di stima LMEQM (1/2)

Y_i, $i = 1, \ldots, m$, variabili aleatorie definite da

$$Y_i = u_i X + \varepsilon_i$$

dove

- X variabile aleatoria di media m_X e varianza σ_X^2;
- u_i coefficienti noti;
- ε_i variabili aleatorie indipendenti, con media nulla e varianza σ_i^2

Si ha

$$Y = UX + \varepsilon$$

con $U = (u_1 \; u_2 \; \ldots \; u_m)'$ e $E[\varepsilon\varepsilon'] = \Sigma_\varepsilon = \text{diag}\{\sigma_i^2\}$

Si vuole calcolare la stima LMEQM

$$\hat{X}_{LMEQM} = m_X + R_{XY} R_Y^{-1} (Y - m_Y)$$

Esempio di stima LMEQM (2/2)

Si ha:

- $m_Y = E[Y] = Um_X$
- $R_{XY} = E[(X - m_X)(Y - Um_X)'] = \sigma_X^2 U'$
- $R_Y = E[(Y - Um_X)(Y - Um_X)'] = U\sigma_X^2 U' + \Sigma_\varepsilon$

da cui (dopo qualche passaggio...)

$$\hat{X}_{LMEQM} = \frac{U'\Sigma_\varepsilon^{-1}Y + \frac{1}{\sigma_X^2}m_X}{U'\Sigma_\varepsilon^{-1}U + \frac{1}{\sigma_X^2}}$$

Caso particolare: $U = (1\ 1\ \ldots\ 1)'$ (ovvero $Y_i = X + \varepsilon_i$)

$$\hat{X}_{LMEQM} = \frac{\sum_{i=1}^{m} \frac{1}{\sigma_i^2} Y_i + \frac{1}{\sigma_X^2}m_X}{\sum_{i=1}^{m} \frac{1}{\sigma_i^2} + \frac{1}{\sigma_X^2}}$$

Nota: l'informazione a priori su X è considerata come un dato aggiuntivo

Esercizio sulla stima Bayesiana (1/3)

Si considerino due variabili aleatorie X e Y, la cui pdf congiunta è

$$f_{X,Y}(x,y) = \begin{cases} -\frac{3}{2}x^2 + 2xy & 0 \leq x \leq 1,\ 1 \leq y \leq 2 \\ 0 & \text{altrimenti} \end{cases}$$

Si vogliono determinare le seguenti stime di X, basate su una osservazione della variabile Y:

- $\hat{X}_{MAP}$
- $\hat{X}_{MEQM}$
- $\hat{X}_{L\,MEQM}$

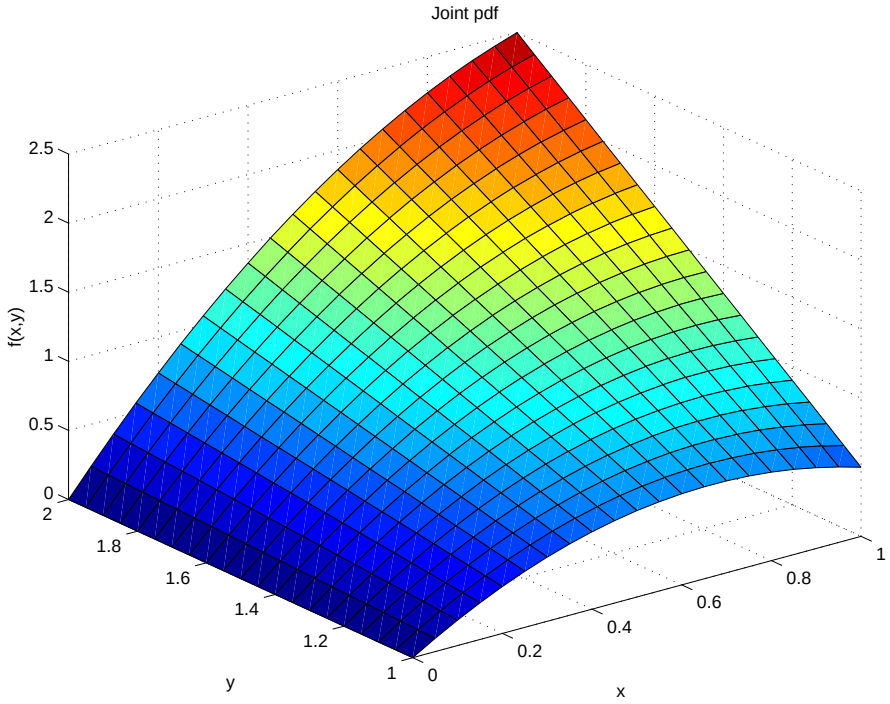
Esercizio sulla stima Bayesiana (2/3)

Soluzioni:

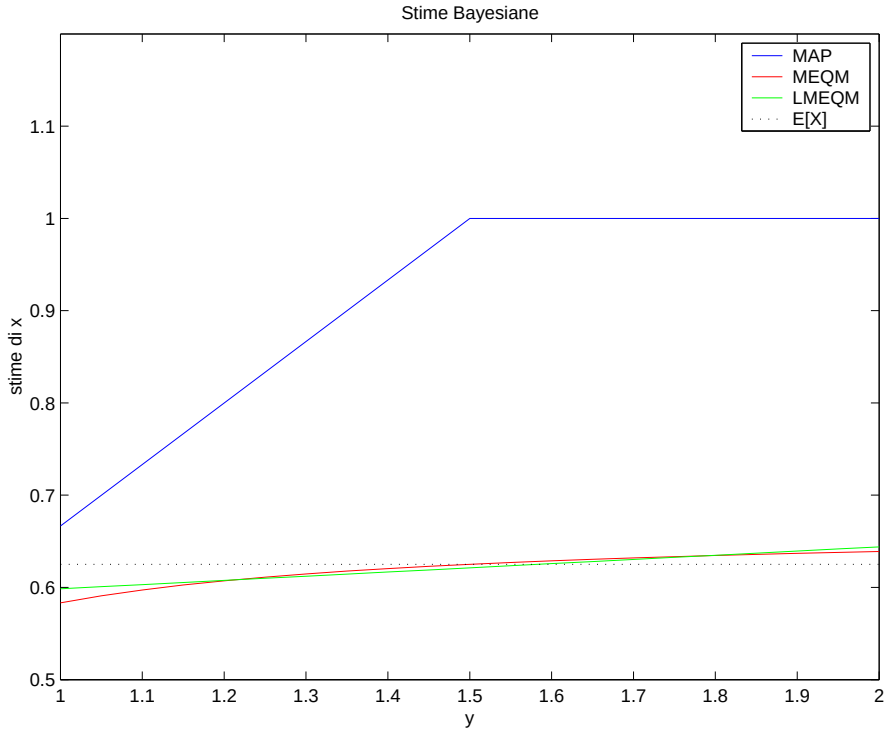
- $\hat{X}_{MAP} = \begin{cases} \frac{2}{3}y & 1 \leq y \leq \frac{3}{2} \\ 1 & \frac{3}{2} \leq y \leq 2 \end{cases}$
- $\hat{X}_{MEQM} = \frac{\frac{2}{3}y - \frac{3}{8}}{y - \frac{1}{2}}$
- $\hat{X}_{LMEQM} = \frac{1}{22}y + \frac{73}{132}$

Vedere file MATLAB: `Es_bayes.m`

Esercizio sulla stima Bayesiana (3/3)



$f_{X,Y}(x, y)$



$\hat{X}_{MAP}(y)$ (blu)

$\hat{X}_{MEQM}(y)$ (rosso)

$\hat{X}_{L\,MEQM}(y)$ (verde)