

Metodi per l'Analisi dei Dati Sperimentali

AA2009/2010



IL CRITERIO DELLA MASSIMA VEROSIMIGLIANZA

Gaetano D'Avino

Esercitazioni

Lezione 6

26/11/09



Massima Verosimiglianza

Esercitazioni
Lezione 6
26/11/09

Sommario

- Introduzione
- La Massima Verosimiglianza
- Esempio 1: una sola misura sperimentale
- Esempio 2: due misure indipendenti
- Esempio 3: un processo Gaussiano
- Esempio 4: il modello lineare
- Esercizio

*La teoria di questa esercitazione si trova nel file
"Lezione06.pdf" sul sito web del Prof. Maffettone*



Introduzione

Esercitazioni
Lezione 6
26/11/09

- Uno stimatore migliore dei MINIMI QUADRATI è la **MASSIMA VEROSIMIGLIANZA (MV)**
- Anzi, è stato mostrato che lo stimatore basato sul criterio della Massima Verosimiglianza è il **migliore stimatore** tra quelli conosciuti
- Infatti si dimostra (vedi teoria) che tale stimatore è **consistente**, il **più efficiente** tra quelli conosciuti ma **non è imparziale**
- In alcuni casi, lo stimatore della Massima Verosimiglianza **coincide** con quello dei minimi quadrati



Massima verosimiglianza

Esercitazioni
Lezione 6
26/11/09

- DEFINIZIONE: la VEROSIMIGLIANZA di avere come esito di n esperimenti $y_1, y_2, \dots, y_n$ è la densità di probabilità di ottenere tutti quei valori fissato il valore dei parametri della distribuzione:

$$L(\theta; y_1, y_2, \dots, y_n) = f_Y(y_1, y_2, \dots, y_n; \theta)$$

- La stima di massima verosimiglianza è quella che si ottiene scegliendo θ tale da **massimizzare** la funzione L :

$$\hat{\theta}_{MV} = \max_{\theta} L(\theta; y_1, \dots, y_n)$$

- Quindi, come nei minimi quadrati, è necessario calcolare un **estremo di una funzione** dei parametri da stimare

L : funzione
massima ve-
rosimiglianza
 f_Y : PDF della Y
 y_i : misure
 θ : vettore dei
parametri

Esempio

Esercitazioni
Lezione 6
26/11/09

- Supponiamo di sapere (o attraverso considerazioni teoriche o precedenti prove sperimentali) che la misura di un processo è caratterizzata da una VA tale che la PDF è:

$$f_Y(y, \theta) = \theta^2 y e^{-\theta y}$$

$$\text{con } y \geq 0$$

NB: Tale distribuzione si chiama *distribuzione di Erlang* e si usa qualche volta per modellare processi dove la variabile dipendente è il tempo

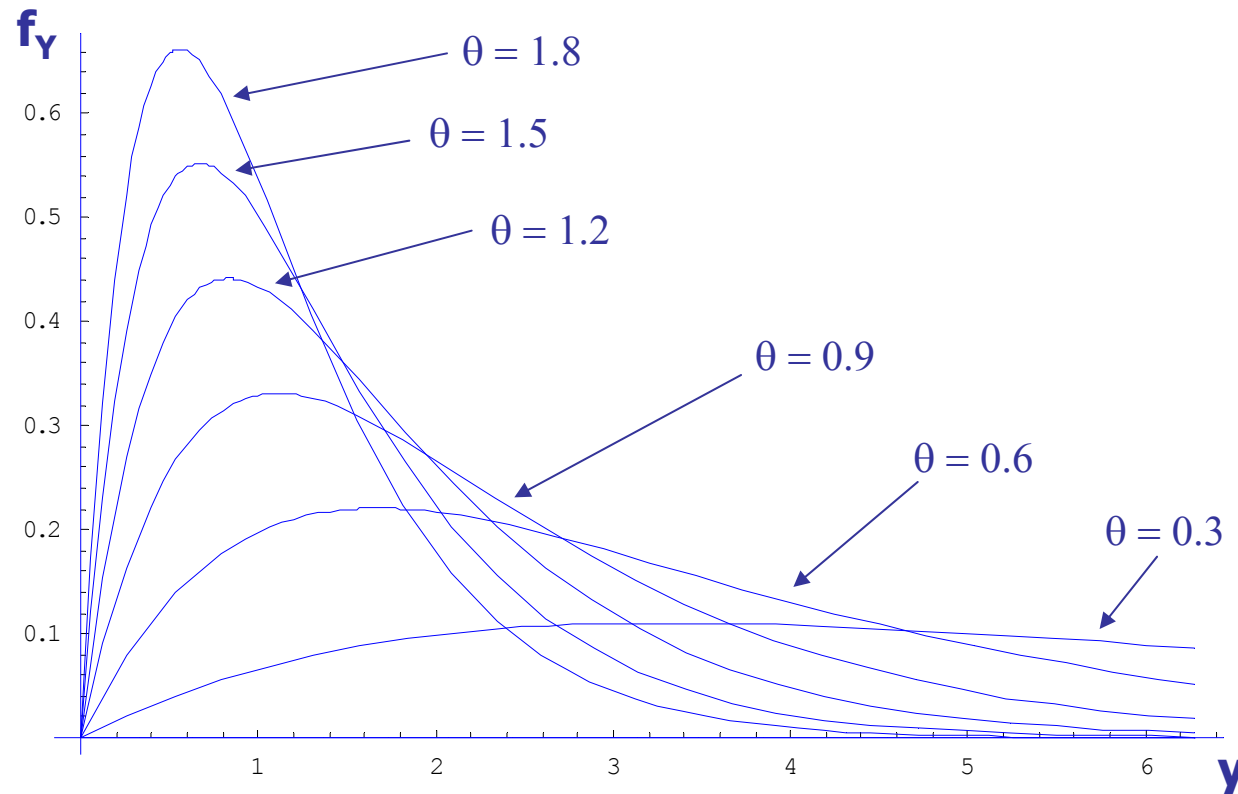
- Il valore di θ **non è noto a priori** ma vogliamo stimarlo sulla base di una campagna sperimentale
- Quindi, dall'espressione precedente della PDF, stiamo considerando una **famiglia di PDF**, parametriche nel parametro θ





Esempio

- L'andamento della PDF al variare di θ è il seguente:



- Quindi, al crescere di θ , il massimo della funzione si alza e si avvicina all'asse delle y



Esempio 1: una sola misura

Esercitazioni
Lezione 6
26/11/09

- Per capire come costruire la funzione MV, ovvero la funzione L , consideriamo il caso più semplice possibile di una sola misura sperimentale y_1
- Abbiamo quindi effettuato un solo esperimento e abbiamo misurato, per esempio, $y_1 = 1.22$
- Vogliamo stimare θ sulla base di questo risultato
- Dalla definizione di MV, la funzione L è la densità di probabilità di avere il valore misurato per un certo vettore di parametri della distribuzione:

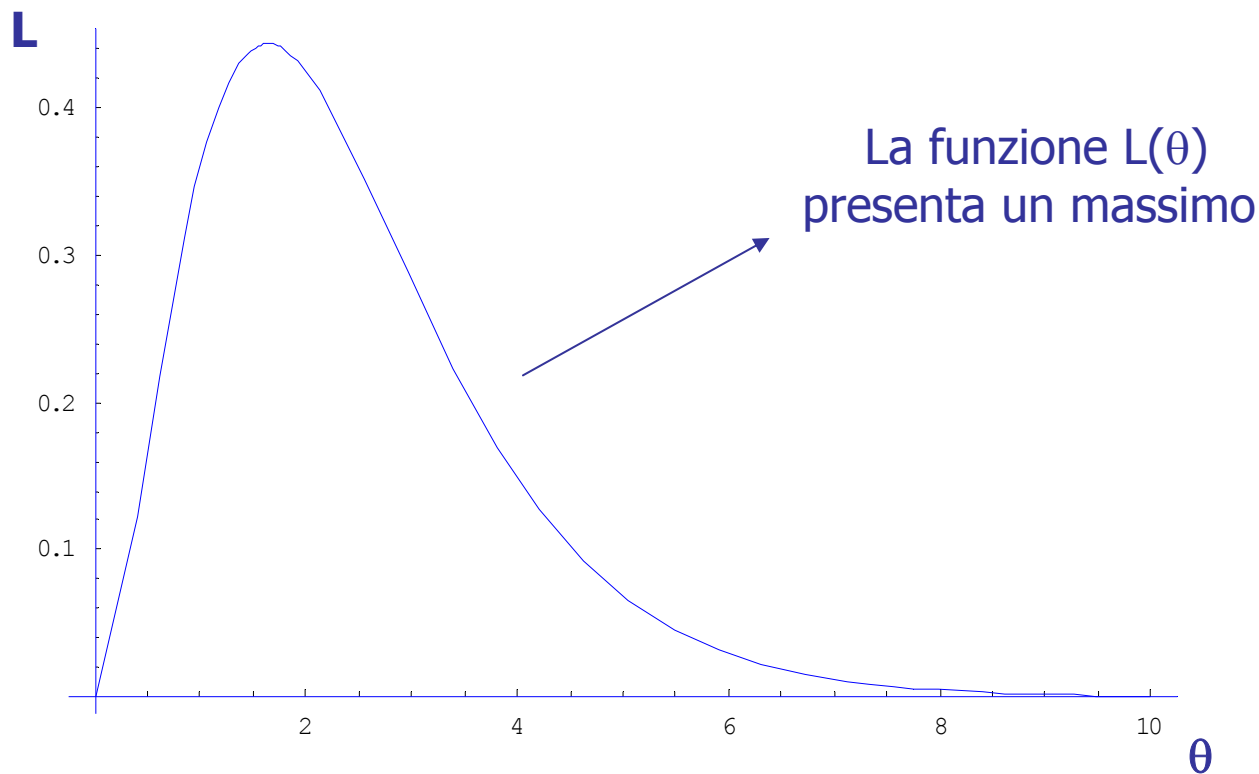
$$L(\theta; y_1) = f_Y(y_1 = 1.22; \theta)$$



Esempio 1: una sola misura

Esercitazioni
Lezione 6
26/11/09

- Ovviamente la funzione L dipenderà solo da θ
- Diagrammando $L(\theta) = f_Y(1.22; \theta)$ si ottiene:





Esempio 1: una sola misura

Esercitazioni
Lezione 6
26/11/09

- Allora la stima di θ si otterrà imponendo:

$$\hat{\theta}_{MV} = \max_{\theta} L(\theta)$$

ovvero calcolando il valore di θ che massimizza $L(\theta)$

- Procedendo per via analitica o utilizzando **"fminsearch"** di Matlab si ottiene $\theta = 1.639$

NB: Il calcolo del massimo di una funzione $f(\mathbf{x})$ equivale alla ricerca del minimo della funzione opposta $-f(\mathbf{x})$

- Il valore di θ calcolato corrisponde alla PDF della famiglia tale che è **MASSIMA** la probabilità che il valore dell'esperimento sia uguale a quello misurato



Esempio 2: due misure

Esercitazioni
Lezione 6
26/11/09

- Consideriamo ora il caso di due misure sperimentali effettuate nelle **STESSE CONDIZIONI** ed **INDIPENDENTI** tra loro
- Il processo è sempre lo stesso quindi entrambe le misure saranno caratterizzate da VA distribuite secondo la distribuzione di Erlang
- Le misure sono: $y_1 = 1.34$ e $y_2 = 1.47$
- Essendo le misure indipendenti tra loro, la congiunta della VA vettoriale $Y = (Y_1, Y_2)$ è data dal prodotto delle marginali:

$$f_Y(y_1, y_2; \theta) = f_{Y_1}(y_1; \theta) \cdot f_{Y_2}(y_2; \theta)$$

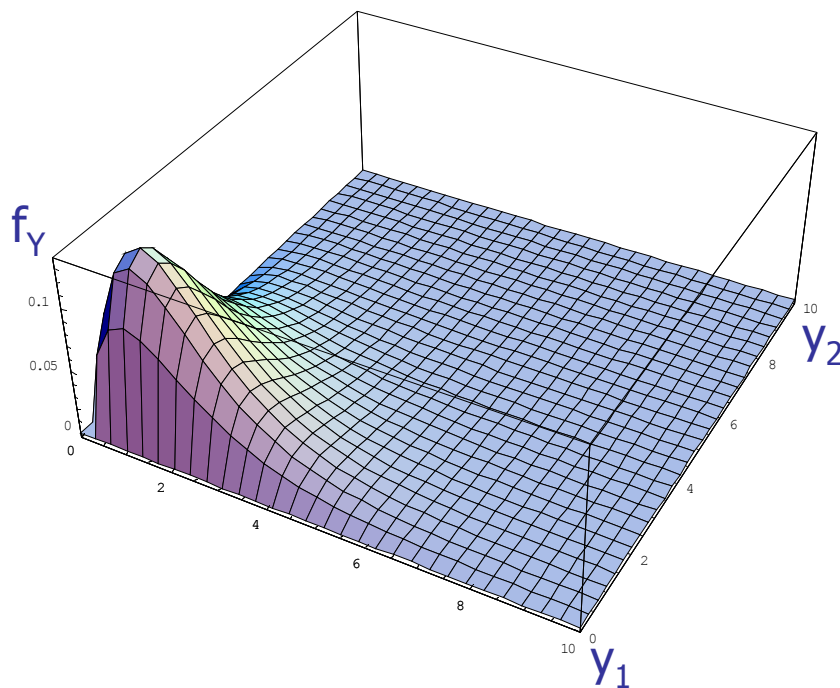
- Anche in questo caso la PDF è una famiglia di curve (3D)



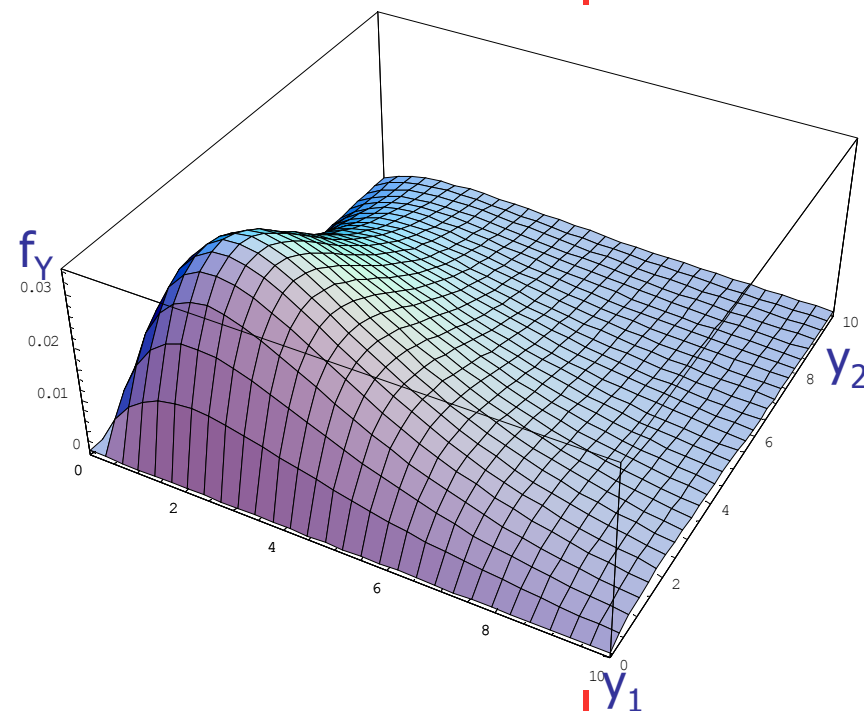
Esempio 2: due misure

Esercitazioni
Lezione 6
26/11/09

- Diagrammando la congiunta per due diversi valori di θ :



$\theta = 1.0$



$\theta = 0.5$

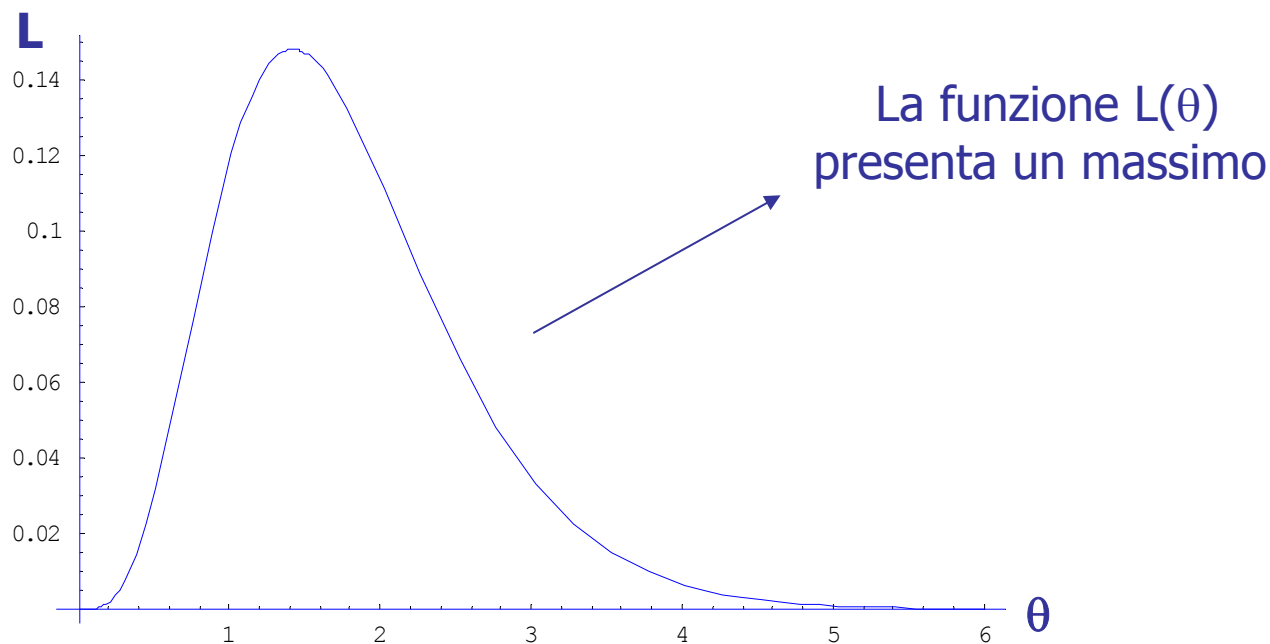


Esempio 2: due misure

- Procedendo in maniera analoga, la funzione L sarà:

$$\begin{aligned} L(\theta; y_1, y_2) &= f_Y(y_1 = 1.34, y_2 = 1.47; \theta) = \\ &= f_{Y_1}(y_1 = 1.34; \theta) \cdot f_{Y_2}(y_2 = 1.47; \theta) \end{aligned}$$

- Ancora la funzione L dipenderà solo da θ





Esempio 2: due misure

Esercitazioni
Lezione 6
26/11/09

- Allora la stima di θ si otterrà imponendo:

$$\hat{\theta}_{MV} = \max_{\theta} L(\theta)$$

ovvero calcolando il valore di θ che massimizza $L(\theta)$

- Utilizzando "fminsearch" di Matlab si ottiene $\theta = 1.423$
- Il calcolo analitico di θ può ancora essere effettuato
- Tuttavia, in questo caso, conviene effettuare una trasformazione sulla funzione $L(\theta)$

NB: E' conveniente effettuare la trasformazione che verrà presentata anche nel caso in cui, per la ricerca del minimo, si ricorre ad una procedura numerica. Infatti, la funzione trasformata è in generale più semplice da gestire.



Esempio 2: due misure

Esercitazioni
Lezione 6
26/11/09

- Nel caso di misure indipendenti, la funzione $L(\theta)$ si può in generale scrivere come la **PRODUTTORIA** delle singole PDF:

$$L(\theta) = \prod_{i=1}^n f_{Y_i}(y_i; \theta)$$

- Essendo la funzione **"logaritmo"** una funzione crescente, il calcolo del massimo (o del minimo) di una funzione equivale al calcolo del massimo (o del minimo) del suo logaritmo
- Tuttavia, il logaritmo di una produttoria è uguale alla sommatoria dei logaritmi dei termini, ovvero:

$$\ln[L(\theta)] = \ln\left[\prod_{i=1}^n f_{Y_i}(y_i; \theta)\right] = \sum_{i=1}^n \ln[f_{Y_i}(y_i; \theta)]$$

- La ricerca di estremi della funzione sommatoria è più agevole della produttoria



Esempio 2: due misure

Esercitazioni
Lezione 6
26/11/09

- La ricerca degli estremi si ottiene eguagliando a zero la derivata della $L(\theta)$ rispetto a θ :

$$\frac{\partial \ln[L(\theta)]}{\partial \theta} = \frac{\partial}{\partial \theta} \sum_{i=1}^n \ln[f_{Y_i}(y_i; \theta)] = \sum_{i=1}^n \frac{1}{f_{Y_i}(y_i; \theta)} \frac{\partial f_{Y_i}(y_i; \theta)}{\partial \theta} = 0$$

- Nel caso in esame, dopo qualche passaggio algebrico, si ottiene:

$$-y_1 - y_2 + \frac{4}{\theta} = 0$$

ovvero: $\theta = 1.423$

- Questo risultato non sarebbe stato ottenibile analiticamente considerando la produttoria originale



Esempio 3: processo Gaussiano

Esercitazioni
Lezione 6
26/11/09

Dati Processo

y_i

2.87856
2.76360
3.40394
2.97889
3.12673
3.16603
2.86311
3.36514
3.04121
2.86698

- Si consideri ora il caso più realistico di 10 misure **indipendenti** ed effettuate nelle **stesse condizioni sperimentali**

- Si suppone che tali misure siano caratterizzate da VA Gaussiane:

$$y_1 = \mu + \varepsilon$$

$$y_2 = \mu + \varepsilon$$

...

$$y_n = \mu + \varepsilon$$

- Essendo le misure effettuate nelle stesse condizioni sperimentali, tutte le VA avranno la stessa media μ
- Inoltre, essendo le misure indipendenti, l'errore ε avrà sempre la stessa varianza σ^2



Esempio 3: processo Gaussiano

Esercitazioni
Lezione 6
26/11/09

Dati Processo

Y_i

2.87856
2.76360
3.40394
2.97889
3.12673
3.16603
2.86311
3.36514
3.04121
2.86698

- Vogliamo stimare i parametri del modello ovvero μ e σ^2
- Essendo le Y_i Gaussiane, le PDF saranno:

$$f_{Y_i}(y_i) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2} \frac{(y_i - \mu)^2}{\sigma^2}\right]$$

e la congiunta (ovvero la funzione $L(\mu, \sigma)$):

$$L(\mu, \sigma) = \prod_{i=1}^{10} \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2} \frac{(y_i - \mu)^2}{\sigma^2}\right]$$

- Si noti che le y_i sono note e le uniche incognite sono μ e σ



Esempio 3: processo Gaussiano

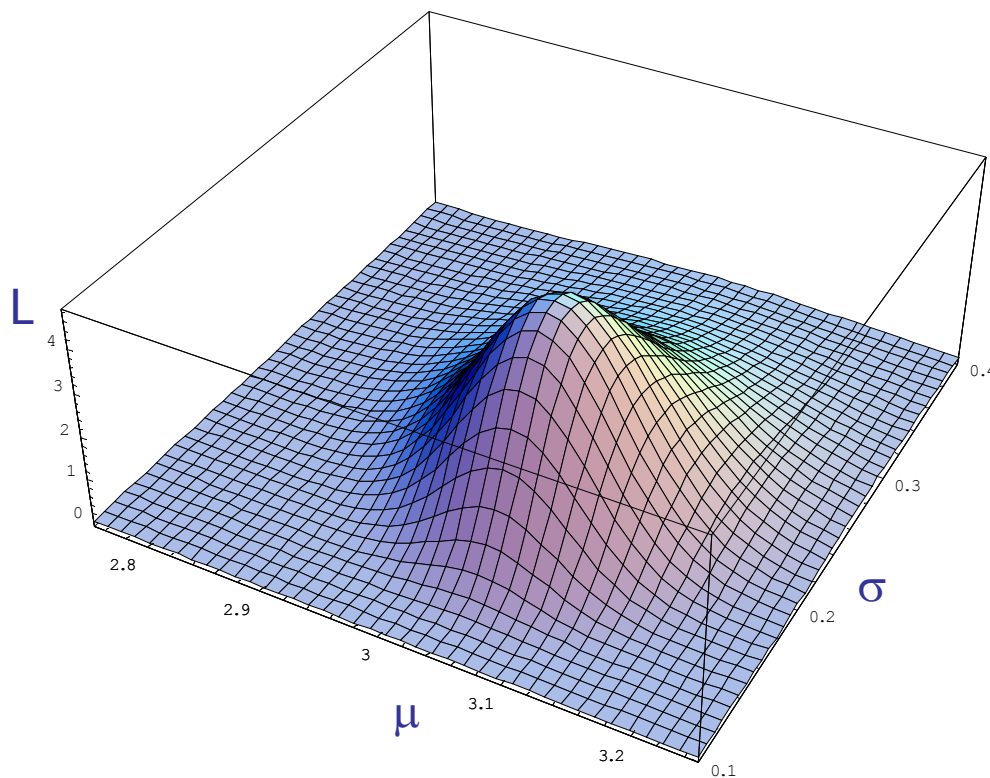
Esercitazioni
Lezione 6
26/11/09

Dati Processo

y_i

2.87856
2.76360
3.40394
2.97889
3.12673
3.16603
2.86311
3.36514
3.04121
2.86698

- Calcolando μ e σ che massimizzano tale funzione, otteniamo la stima di media e deviazione standard per il set di dati y_i
- Innanzitutto possiamo diagrammare la funzione $L(\mu, \sigma)$:





Esempio 3: processo Gaussiano

Esercitazioni
Lezione 6
26/11/09

- Calcolando il massimo della funzione precedente con Matlab otteniamo:
 $\mu = 3.045$ e $\sigma = 0.207$
- E' possibile procedere anche al calcolo analitico, cercando il massimo del logaritmo della funzione $L(\mu, \sigma)$:

$$L(\mu, \sigma^2) = \prod_{i=1}^{10} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2} \frac{(y_i - \mu)^2}{\sigma^2}\right] = (2\pi\sigma^2)^{-10/2} \exp\left[-\frac{1}{2\sigma^2} \sum_{i=1}^{10} (y_i - \mu)^2\right]$$

e passando al logaritmo:

$$\ln[L(\mu, \sigma^2)] = -\frac{10}{2} \ln(2\pi) - \frac{10}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^{10} (y_i - \mu)^2$$



Esempio 3: processo Gaussiano

Esercitazioni
Lezione 6
26/11/09

- Derivando rispetto a μ e σ^2 ed uguagliando le derivate a zero:

$$\frac{\partial}{\partial \mu} \ln[L(\mu, \sigma^2)] = \frac{1}{\sigma^2} \sum_{i=1}^{10} (y_i - \mu) = 0$$

$$\frac{\partial}{\partial \sigma^2} \ln[L(\mu, \sigma^2)] = -\frac{10}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^{10} (y_i - \mu)^2 = 0$$

da cui si ottiene:

$$\hat{\mu} = \frac{1}{10} \sum_{i=1}^{10} y_i \quad \hat{\sigma}^2 = \frac{1}{10} \sum_{i=1}^{10} (y_i - \mu)^2$$

- La MV fornisce come stimatori per μ e σ^2 la media e la varianza già incontrati nella statistica descrittiva
- Comunque lo stimatore della varianza è PARZIALE (vedi teoria)



Esempio 4: modello lineare

Esercitazioni
Lezione 6
26/11/09

- Il criterio della MV si può applicare anche quando **è noto il modello di un processo** e si vogliono stimare i suoi parametri
- Si consideri l'esempio già visto per i minimi quadrati:

ESEMPIO:

Sono state effettuate 11 misure di un processo (y_i) al variare delle condizioni sperimentali (x_i). Si ipotizza che il processo sia modellabile attraverso un modello lineare nei parametri:

$$y = ax + b$$

Si vogliono stimare i parametri a e b (ed ora anche la varianza σ^2) del modello utilizzando il metodo dei minimi quadrati

- Rispetto all'Esempio 3, questa volta le condizioni sperimentali **sono diverse** da misura a misura (x_i)

Dati Processo

x_i	y_i
0.0	2.23
0.1	2.86
0.2	3.30
0.3	3.88
0.4	4.22
0.5	4.87
0.6	5.48
0.7	5.85
0.8	6.61
0.9	7.07
1.0	7.54



Esempio 4: modello lineare

Esercitazioni
Lezione 6
26/11/09

- Ipotizzando ancora **errori Gaussiani** ed **indipendenza** delle misure:

$$Y_1 = ax_1 + b + \varepsilon$$

$$Y_2 = ax_2 + b + \varepsilon$$

...

$$Y_n = ax_n + b + \varepsilon$$



$$\mathbf{Y} = N(\boldsymbol{\mu}, \sigma^2)$$

$$\boldsymbol{\mu} = \begin{pmatrix} ax_1 + b \\ ax_2 + b \\ \dots \\ ax_n + b \end{pmatrix}$$

- Quindi la PDF congiunta, ovvero la funzione $L(a, b, \sigma^2)$, sarà:

$$L(a, b, \sigma^2) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2} \frac{(y_i - a - bx_i)^2}{\sigma^2}\right]$$

- I valori di a , b e σ^2 che massimizzano questa funzione (o il suo logaritmo) sono le stime che cerchiamo



Esempio 4: modello lineare

Esercitazioni
Lezione 6
26/11/09

- Ancora una volta, per il calcolo del massimo, si può procedere per via numerica o per via analitica
- In quest'ultimo caso è necessario ricorrere al logaritmo della funzione precedente
- Per i calcoli analitici si veda la teoria
- Di nuovo si dimostra che lo stimatore per la varianza è **PARZIALE** mentre quelli per a e b sono **IMPARZIALI**
- E' possibile applicare il criterio della MV anche per modelli lineari a più parametri o modelli non lineari: è sufficiente sostituire alla funzione lineare: $a + b \cdot x_i$ la funzione del modello $g(x_i, \theta)$ (vedi teoria)



Esercizio

Esercitazioni
Lezione 6
26/11/09

Dati

x_i	y_i
1	1.52
2	1.59
3	1.67
4	1.809
5	1.901

1. Si considerino i dati sperimentali riportati a destra (gli stessi dell'esercitazione sui minimi quadrati) e si effettui una stima dei parametri a e b della regressione lineare:

$$y = ax + b$$

Si effettui inoltre anche una stima della varianza dell'errore

A tale scopo si applichi il criterio della Massima Verosimiglianza:

- modificando opportunamente l'M-file: "functionLS2.m"
- procedendo al calcolo numerico (con il comando "fminsearch") del massimo della funzione MV

2. Si verifichino i risultati con le espressioni analitiche ricavate dalla teoria