

01AYS - FONDAMENTI DI AUTOMATICA

Stima parametrica classica

proff. Marina Indri e Michele Taragna
Dip. di Automatica e Informatica
Politecnico di Torino

Anno Accademico 2001/2002

Versione 1.1

Identificazione classica (I)

Ipotesi #1: il sistema \mathcal{S} è descritto da un modello parametrico \mathcal{M} che dipende da un vettore di parametri “veri” non noti $p^o \in \mathfrak{R}^n$:

$$\mathcal{S} = \mathcal{M}(p^o)$$

Ipotesi #2: su \mathcal{S} si ha solo un numero limitato N di informazioni sperimentali, corrotte da un rumore di misura non noto $e^N \in \mathfrak{R}^N$, costituito da una sequenza aleatoria con momenti del I e del II ordine noti:

$$\underbrace{E(e^N)}_{\text{valor medio}} = 0 \qquad \underbrace{E(e^N \cdot (e^N)^T)}_{\text{matrice di covarianza}} = \Sigma_e \text{ nota}$$

Identificazione classica (II)

Problema di stima parametrica:

data la relazione:

$$\underbrace{z^N}_{\text{dati noti}} = \underbrace{F(p^o)}_{\text{funzione nota}} + \underbrace{e^N}_{\text{rumore non noto}}$$

trovare una stima \hat{p} di p^o mediante un algoritmo di stima (stimatore) ϕ applicato ai dati z^N :

$$p^o \cong \hat{p} = \phi(z^N)$$

Esempio (I)

Stima della resistenza di un resistore reale

Si effettuano N misure di tensione-corrente supponendo che:

- la caratteristica statica del resistore sia lineare e quindi il modello del dispositivo sia dato dalla legge di Ohm:

$$v_R = R \cdot i_R$$

- le misure siano corrotte da un rumore casuale $e^N = [e_1, \dots, e_N]^T$

Si ottiene così il sistema di equazioni:

$$\begin{aligned} v_{R,1} &= R \cdot i_{R,1} + e_1 \\ v_{R,2} &= R \cdot i_{R,2} + e_2 \\ &\vdots \\ v_{R,N} &= R \cdot i_{R,N} + e_N \end{aligned}$$

Esempio (II)

In forma matriciale:

$$\underbrace{\begin{bmatrix} v_{R,1} \\ v_{R,2} \\ \vdots \\ v_{R,N} \end{bmatrix}}_{z^N} = \underbrace{\begin{bmatrix} i_{R,1} \\ i_{R,2} \\ \vdots \\ i_{R,N} \end{bmatrix}}_L \cdot \underbrace{[R]}_p + \underbrace{\begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_N \end{bmatrix}}_{e^N}$$

è nella forma standard:

$$\underbrace{z^N}_{\text{dati noti}} = \underbrace{F(p)}_{\text{funzione nota}} + \underbrace{e^N}_{\text{rumore non noto}}$$

con:

$$F(p) = L \cdot p = \text{funzione lineare del parametro } p$$

Obiettivo: trovare una stima \hat{R} di R mediante un algoritmo di stima (stimatore) ϕ applicato ai dati z^N :

$$\hat{R} = \phi(z^N) \cong R$$

Proprietà delle stime statistiche (I)

e^N è una variabile aleatoria

↓

$z^N = F(p^0) + e^N$ è una variabile aleatoria

↓

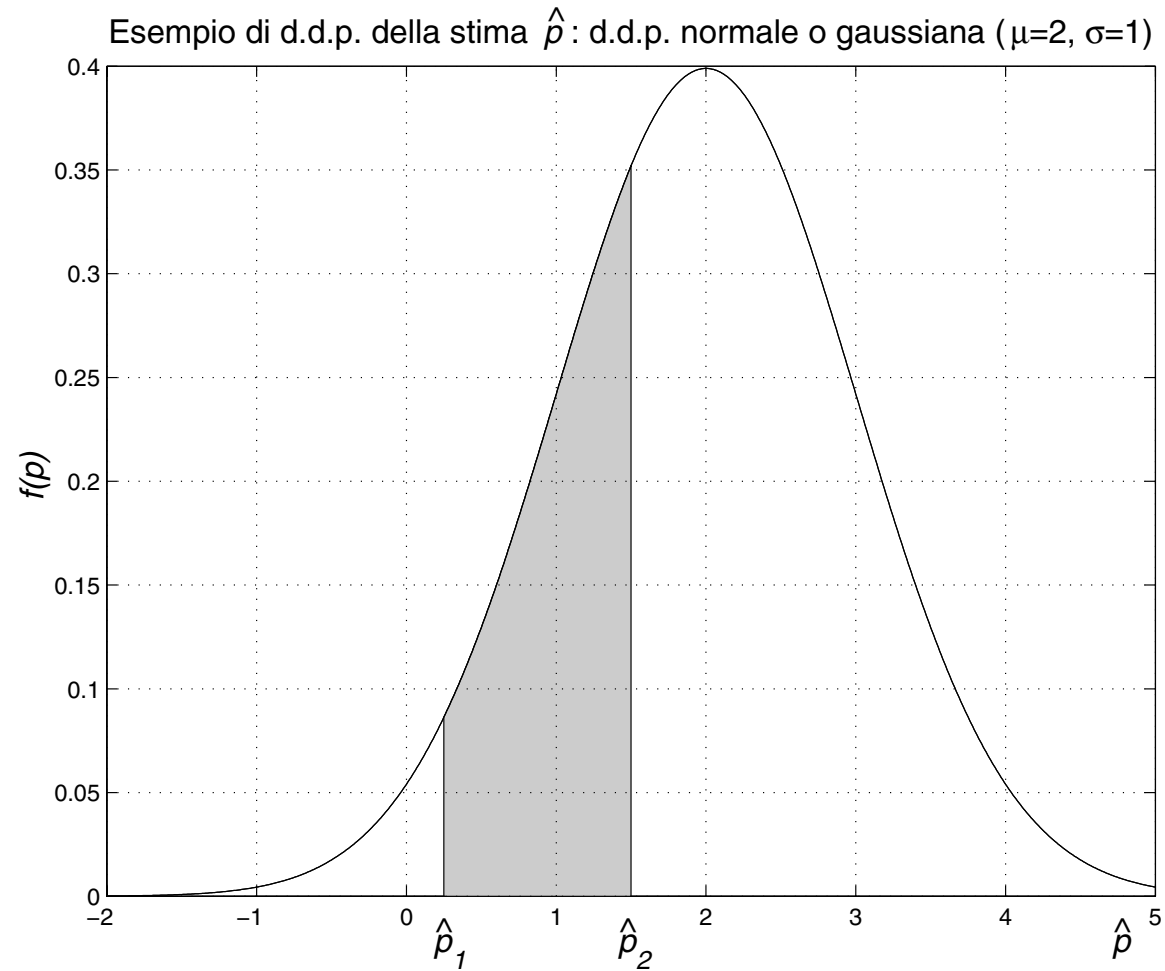
$\hat{p} = \phi(z^N)$ è una variabile aleatoria

Problema: come si misura la “bontà” della stima \hat{p} ?

Per confrontare due differenti algoritmi di stima che portano a due diverse stime \hat{p} e \tilde{p} , paragono le densità di probabilità delle rispettive stime sulla base delle seguenti caratteristiche:

- correttezza;
- efficienza;
- consistenza

Proprietà delle stime statistiche (II)



Area = Probabilità($\hat{p}_1 \leq \hat{p} \leq \hat{p}_2$)

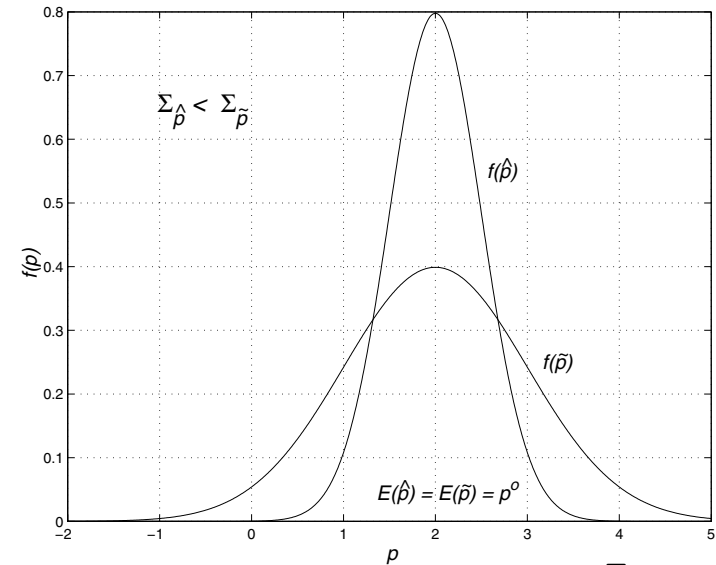
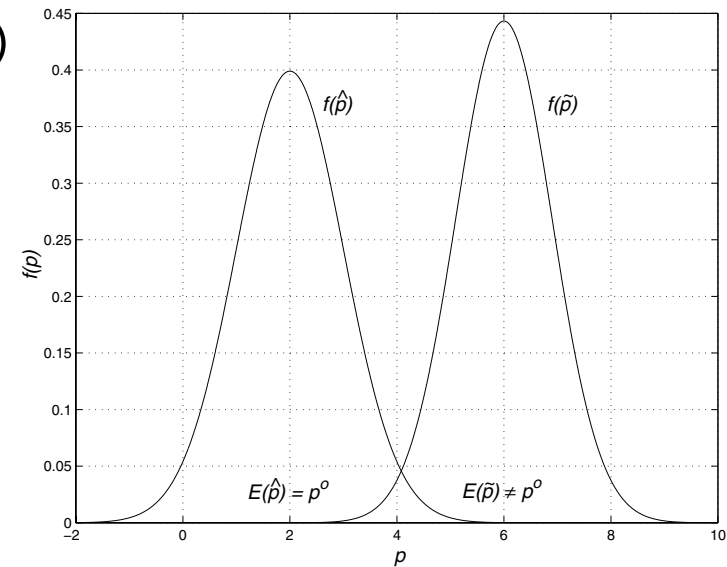
Proprietà delle stime statistiche (III)

- **Correttezza (o non polarizzazione)**
confronto i momenti del I ordine e scelgo \hat{p} in modo che:

$$\underbrace{E(\hat{p})}_{\text{valor medio della stima}} = \underbrace{p^o}_{\text{valore vero dei parametri}}$$

- **Efficienza**
se entrambe le stime sono corrette, confronto i momenti del II ordine e scelgo \hat{p} in modo che:

$$\Sigma_{\hat{p}} \leq \Sigma_{\tilde{p}}$$

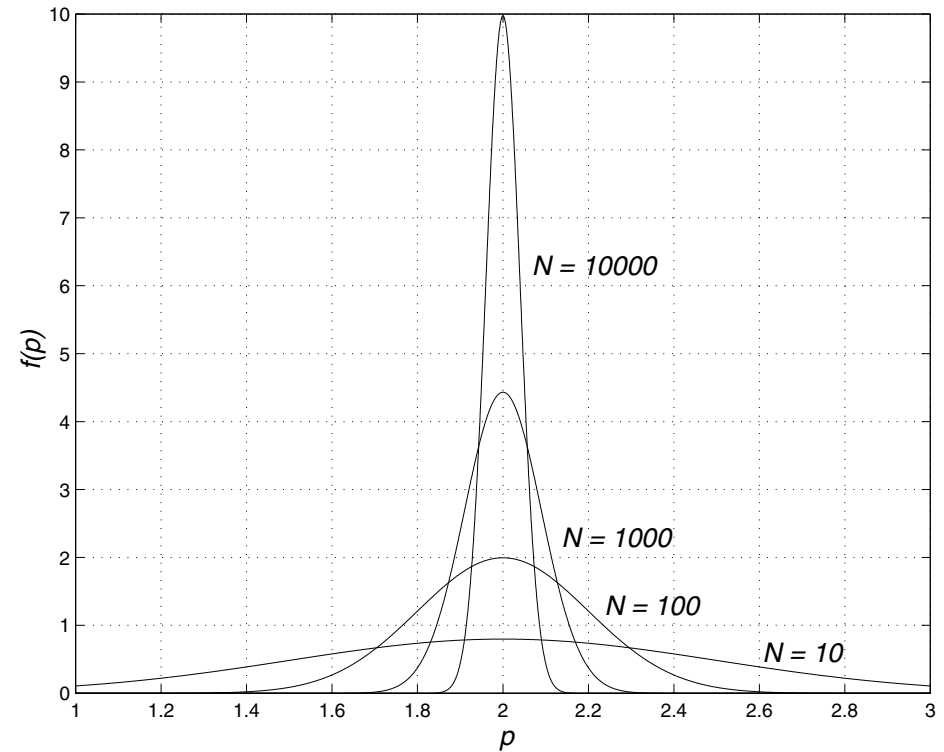


Proprietà delle stime statistiche (IV)

- **Consistenza**

sia per stime efficienti
sia per stime non efficienti,
guardo se vale la seguente
proprietà asintotica:

$$\lim_{N \rightarrow \infty} \sum \hat{p} = 0$$



Stima di massima verosimiglianza (I)

Si riescono a trovare stimatori corretti, efficienti, consistenti?

Definisco **funzione di verosimiglianza** la d.d.p. condizionata

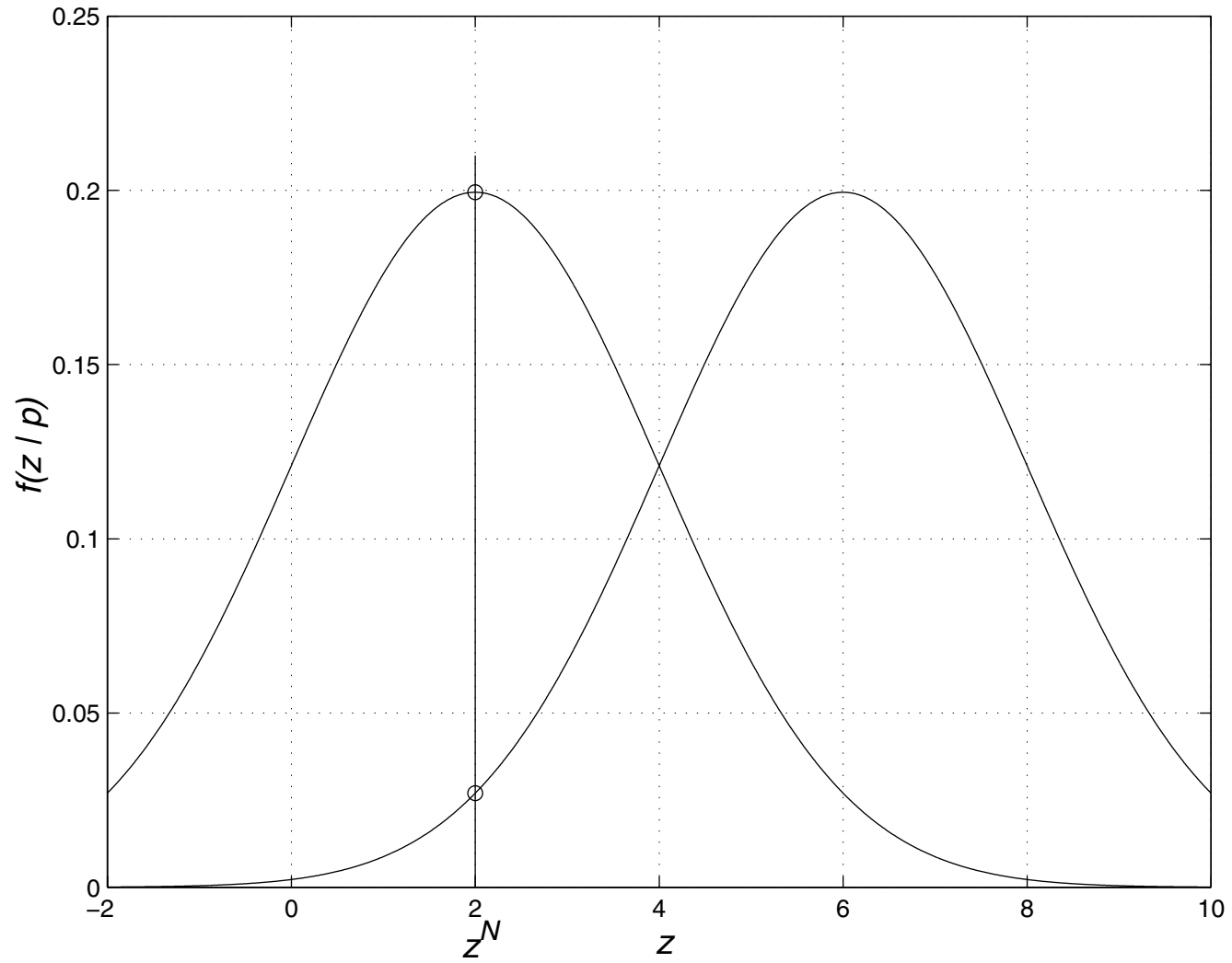
$f(z|p) =$ d.d.p. dei dati z se fossero generati dal sistema $\mathcal{M}(p)$

e definisco **stima di massima verosimiglianza**:

$$\hat{p}^{MV} = \arg \max_{p \in \mathcal{R}^n} f(z = z^N | p)$$

la stima corrispondente al valore dei parametri per cui i dati misurati z^N sono i più probabili

Stima di massima verosimiglianza (II)



Proprietà della stima di massima verosimiglianza

\hat{p}^{MV} è una stima:

- asintoticamente corretta: $E(\hat{p}^{MV}) \xrightarrow{N \rightarrow \infty} p^o$
- asintoticamente efficiente: $\Sigma_{\hat{p}^{MV}} \leq \Sigma_{\tilde{p}} \quad \forall \tilde{p} \text{ se } N \rightarrow \infty$
- consistente: $\lim_{N \rightarrow \infty} \Sigma_{\hat{p}^{MV}} = 0$
- asintoticamente gaussiana (per $N \rightarrow \infty$)

Calcolo della stima di massima verosimiglianza (I)

Nel caso di errori gaussiani:

$$f(e^N) = \underbrace{\mathcal{N}(0, \Sigma_e)}_{\text{distribuzione normale o gaussiana}} = \frac{1}{\sqrt{(2\pi)^N \cdot \det(\Sigma_e)}} \exp\left(-\frac{1}{2} (e^N)^T \cdot \Sigma_e^{-1} \cdot e^N\right)$$

Caso generale: $F(p) =$ generica funzione non lineare dei parametri

$$z^N = F(p) + e^N$$

$$\Downarrow \quad e^N = z^N - F(p)$$

$$f(z^N | p) = \mathcal{N}(F(p), \Sigma_e) = \frac{1}{\sqrt{(2\pi)^N \cdot \det(\Sigma_e)}} \exp\left(-\frac{1}{2} [z^N - F(p)]^T \cdot \Sigma_e^{-1} \cdot [z^N - F(p)]\right)$$

Calcolo della stima di massima verosimiglianza (II)

$f(z^N | p)$ è una funzione esponenziale in p

⇓

$$\hat{p}^{MV} = \arg \max_{p \in \mathcal{R}^n} f(z^N | p) = \arg \min_{p \in \mathcal{R}^n} \underbrace{\left\{ [z^N - F(p)]^T \cdot \Sigma_e^{-1} \cdot [z^N - F(p)] \right\}}_{R(p)}$$

Nel caso di errori gaussiani, la stima di massima verosimiglianza è la *stima dei minimi quadrati pesati* con la covarianza degli errori

Problema: bisogna trovare il minimo globale di $R(p)$ al variare di p , che però può avere minimi locali se $F(p)$ è una generica funzione non lineare dei parametri; gli usuali algoritmi di ottimizzazione non lineare non garantiscono di arrivare al minimo globale

Calcolo della stima di massima verosimiglianza (III)

Caso particolare: $F(p) =$ funzione lineare dei parametri $= Lp$

$$z^N = Lp + e^N$$

↓

$$R(p) \text{ è quadratica in } p : R(p) = [z^N - Lp]^T \cdot \Sigma_e^{-1} \cdot [z^N - Lp]$$

↓

esiste un solo minimo di $R(p)$

↓

$$\hat{p}^{MV} = (L^T \Sigma_e^{-1} L)^{-1} L^T \Sigma_e^{-1} z^N = \text{stima di Gauss-Markov} = \hat{p}^{GM}$$

Nel caso $\Sigma_e = \sigma_e^2 I_N$ (errori indipendenti identicamente distribuiti):

$$\hat{p}^{MV} = \hat{p}^{GM} = (L^T L)^{-1} L^T z^N = \text{stima dei minimi quadrati}$$

Proprietà della stima di Gauss-Markov

\hat{p}^{GM} è una stima:

- corretta: $E(\hat{p}^{GM}) = p^o$
- efficiente: $\sum_{\hat{p}^{GM}} \leq \sum_{\tilde{p}} \quad \forall \tilde{p}$
- consistente: $\lim_{N \rightarrow \infty} \sum_{\hat{p}^{GM}} = 0$
- gaussiana