

Sistemi di Controllo Digitale

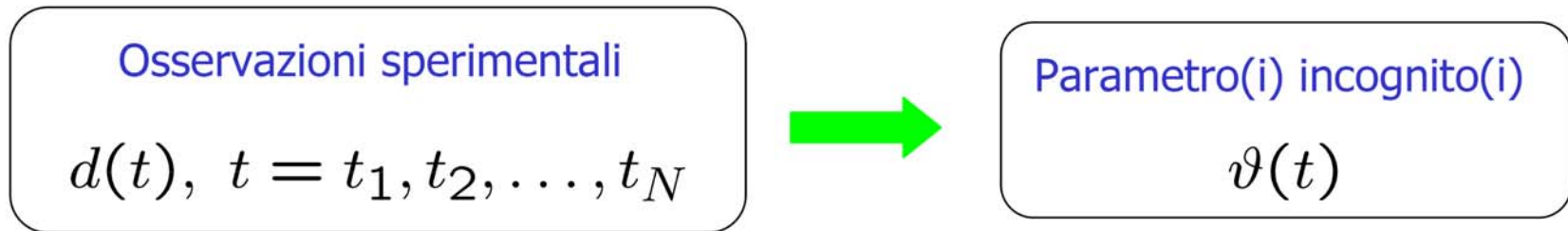
***Identificazione di Modelli
Dinamici***

Contenuto delle Lezioni

- ❑ Il problema della stima e della predizione
- ❑ Teoria della stima e caratteristiche degli stimatori
- ❑ Stima ai minimi quadrati (LMS)

Il Problema della Stima e della Predizione

Il problema della stima nasce quando si vogliono determinare uno o piu` **parametri incogniti a partire da osservazioni sperimentali.**

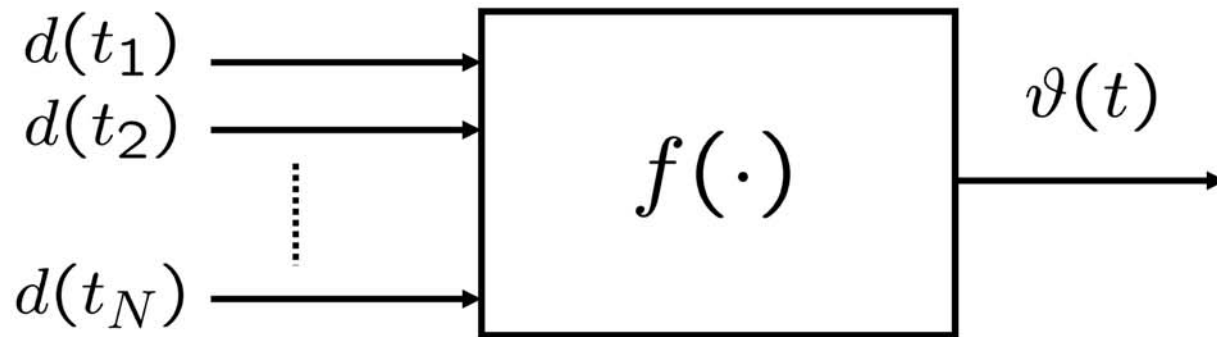


Nella gran maggioranza dei casi i **parametri incogniti sono costanti** $\vartheta(t) = \vartheta$

$T = \{t_1, t_2, \dots, t_N\}$ insieme degli istanti di osservazione

- In generale non e` detto che i t_i debbano avere cadenza regolare
- Se si ha accesso alla generazione dei t_i conviene "addensarli" dove le osservazioni sperimentali sono piu` significative

Stimatore



Lo stimatore e' quindi una **funzione deterministica** che a partire dai dati osservati in ingresso produce in uscita i parametri incogniti.

Stima di parametri costanti

- Se $\vartheta(t) = \bar{\vartheta} = \text{cost}$ si parla di stima o identificazione parametrica.
- La stima prodotta dallo stimatore si indica con $\hat{\vartheta}$ o con $\hat{\vartheta}_T$ qualora si voglia evidenziare l'insieme degli istanti di osservazione.
- Il valore "vero" del parametro incognito si indica con ϑ°

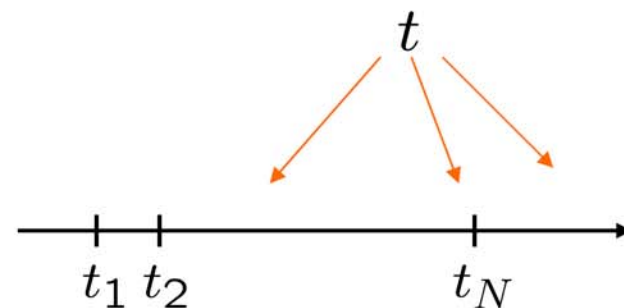
Stima di parametri varianti nel tempo

- La stima prodotta dallo stimatore si indica con $\hat{\vartheta}(t|T)$ o semplicemente con $\hat{\vartheta}(t|N)$ qualora si possa prendere $T = \{1, 2, \dots, N\}$
- A seconda di come si ponga l'istante t rispetto a t_N si hanno i tre casi:

$t > t_N$: problema di **predizione**

$t = t_N$: problema di **filtraggio**

$t < t_N$: problema di **regolarizzazione (smoothing)**



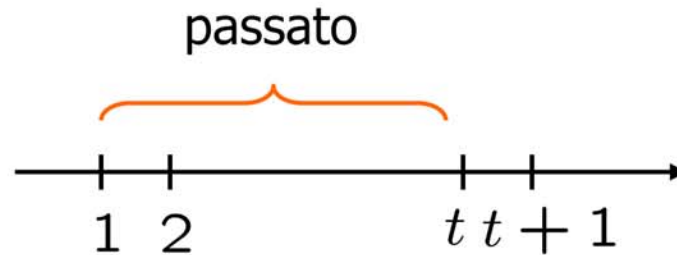
Problema della predizione

Si tratta di un problema fondamentale nell'ambito dell'identificazione dei sistemi dinamici.

- Per fissare le idee consideriamo il caso delle *serie temporali*.
- Si ha una sequenza di osservazioni $y(1), y(2), \dots, y(N)$ di una variabile $y(\cdot)$
- Si vuole stimare $y(t + 1)$
- Si vuole quindi determinare un predittore

$$\hat{y}(t + 1|t) = f[y(t), y(t - 1), \dots, y(1)]$$

- Il predittore esprime quindi una **stima** $\hat{y}(t + 1|t)$ di $y(t + 1)$ in funzione di N valori assunti da $y(\cdot)$ nel **passato**.



- Un predittore si dice **lineare** se

$$\hat{y}(t + 1|t) = a_1(t) y(t) + \cdots + a_t(t) y(1)$$

- Un predittore si dice **a memoria finita** (ovvero utilizza una memoria limitata del passato) se

$$\hat{y}(t + 1|t) = a_1(t) y(t) + \cdots + a_n(t) y(t - n + 1)$$

- Un predittore si dice **lineare tempo-invariante** se

$$\hat{y}(t + 1|t) = a_1 y(t) + \cdots + a_n y(t - n + 1)$$

dove i parametri $a_1, \dots, a_n$ sono **costanti**

- Si definisce il vettore dei parametri $\vartheta^\top = [a_1 \cdots a_n]$

Determinare un buon predittore significa determinare un opportuno vettore ϑ tale che la predizione $\hat{y}(t + 1|t)$ sia il piu' possibile accurata.

Piu` precisamente:

- Si consideri un predittore lineare tempo-invariante a memoria finita

$$\hat{y}(t+1|t) = a_1 y(t) + \dots + a_n y(t-n+1)$$

dove n e` piccolo rispetto al numero di dati misurati fino all'istante t

- La capacita` predittiva del predittore puo` quindi essere valutata rispetto ai dati gia` noti $y(i)$, $i = 1, \dots, t$:

- si calcola

$$\hat{y}(i+1|i) = a_1 y(i) + \dots + a_n y(i-n+1), \forall i > n$$

- si valuta l'errore di predizione

$$\varepsilon(i+1) = y(i+1) - \hat{y}(i+1|i), \forall i > n$$

Il vettore $\vartheta^\top = [a_1 \cdots a_n]$ e' "buono" se ε e' piccolo su tutti i dati disponibili

- definiamo allora una cifra di merito:

$$J(\vartheta) = \sum_{i=n+1}^t \varepsilon(i)^2$$

- quindi

$$\vartheta^\circ = \arg \min_{\vartheta} J(\vartheta)$$

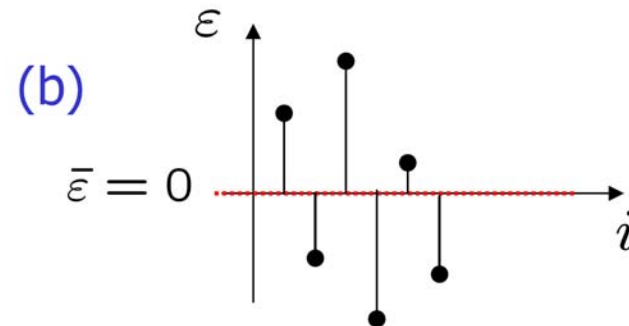
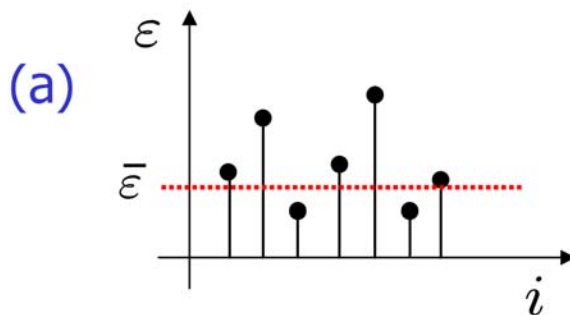
La determinazione di ϑ° e' quindi ricondotta alla soluzione di un problema di ottimizzazione.

Osservazioni

E' molto importante chiarire che cosa si intende per ε "piccolo"



la minimizzazione di $J(\vartheta)$ non e' di per se' un criterio soddisfacente



- Caso (a): non e' soddisfacente perche' l'errore medio $\bar{\varepsilon}$ e' non nullo  errore sistematico
- Caso (b): non e' soddisfacente perche' nonostante l'errore medio $\bar{\varepsilon}$ sia nullo, la sequenza e' alternata, quindi ad ogni passo si sa gia' che segno avra' l'errore al passo successivo

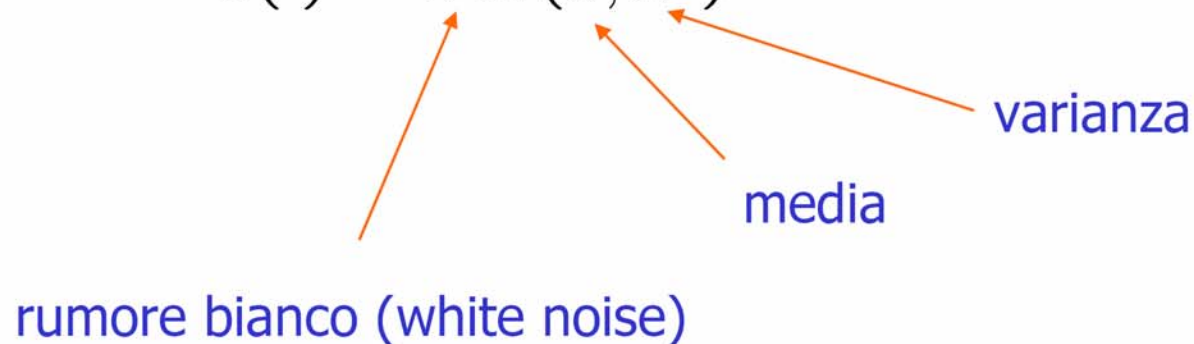
Il predittore non ha acquisito tutta l'informazione



Quindi:

La situazione ideale sarebbe avere un errore ε di valor medio il piu` piccolo possibile e che abbia una natura il piu` possibile **imprevedibile**.

$$\varepsilon(\cdot) \sim WN(0, \lambda^2)$$



Predittore come sistema dinamico

$$\hat{y}(t|t-1) = a_1 y(t-1) + \dots + a_n y(t-n)$$

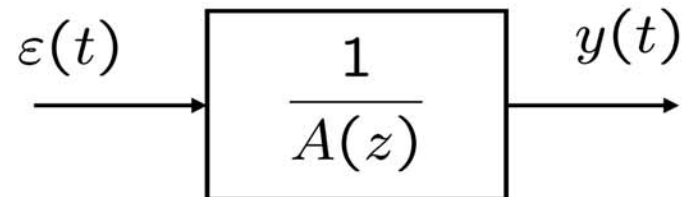
$$\varepsilon(t) = y(t) - \hat{y}(t|t-1) \quad \longrightarrow \quad y(t) = \varepsilon(t) + \hat{y}(t|t-1)$$

$$\hookrightarrow y(t) = a_1 y(t-1) + \dots + a_n y(t-n) + \varepsilon(t)$$

$$\hookrightarrow y(t) = (a_1 z^{-1} + \dots + a_n z^{-n}) y(t) + \varepsilon(t)$$

$$\hookrightarrow A(z) y(t) = \varepsilon(t) \quad \text{con} \quad A(z) = 1 - a_1 z^{-1} - \dots - a_n z^{-n}$$

$$\hookrightarrow y(t) = \frac{1}{A(z)} \varepsilon(t)$$



Teoria della Stima e Caratteristiche degli Stimatori

Generalita`

- In generale abbiamo:

$$d = d(s, \vartheta^\circ)$$

dove

- d sono i dati osservati
 - ϑ° e` la quantita` da stimare
 - s e` l'esito dell'esperimento casuale
- Lo stimatore e` una funzione:

$$\hat{\vartheta} = f [d(s, \vartheta^\circ)]$$



Lo stimatore e` variabile aleatoria in quanto il suo valore dipende dall'esito dell'esperimento casuale s

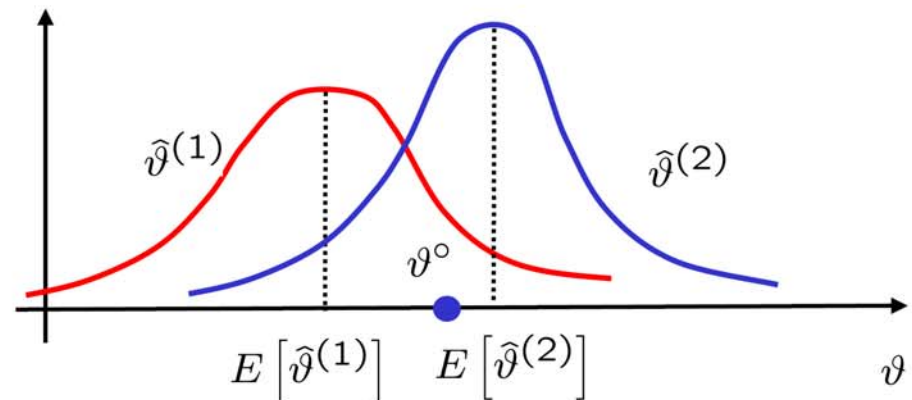
Polarizzazione

- In generale lo stimatore $\hat{\vartheta} = f[d(s, \vartheta^\circ)]$ si dice **non polarizzato** se

$$E(\hat{\vartheta}) = \vartheta^\circ$$

- La non polarizzazione di uno stimatore e` evidentemente una proprieta` che bisogna cercare di assicurare.

Nel caso in figura gli stimatori sono ambedue polarizzati ma lo stimatore $\hat{\vartheta}^{(2)}$ ha una polarizzazione minore.



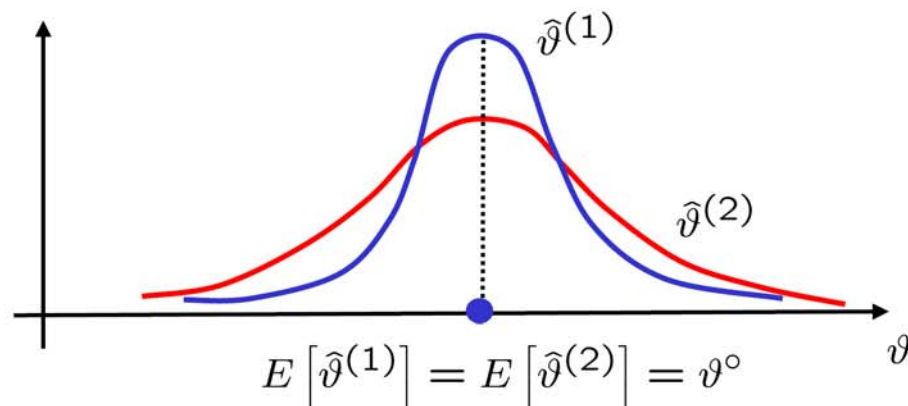
Minima varianza

- La non polarizzazione (o correttezza) non e' l'unico criterio con cui valutare la qualita' di uno stimatore.

Nel caso in figura gli stimatori sono ambedue non polarizzati.

Pero'

$$\text{var} [\hat{\vartheta}^{(1)}] \ll \text{var} [\hat{\vartheta}^{(2)}]$$



- Quindi lo stimatore $\hat{\vartheta}^{(1)}$ ha probabilita' maggiore di produrre valori della stima vicini al valore vero ϑ^0 rispetto allo stimatore $\hat{\vartheta}^{(2)}$
- Si vuole quindi che la varianza dello stimatore sia la piu' piccola possibile.

- In generale, a parità di caratteristiche di polarizzazione, diremo che lo stimatore $\hat{\vartheta}^{(1)}$ è migliore dello stimatore $\hat{\vartheta}^{(2)}$ se

$$\text{var} [\hat{\vartheta}^{(1)}] \leq \text{var} [\hat{\vartheta}^{(2)}]$$

ovvero se la matrice (ϑ può essere un vettore)

$$\text{var} [\hat{\vartheta}^{(2)}] - \text{var} [\hat{\vartheta}^{(1)}] \geq 0$$

- Ricordiamo che $A \geq 0 \implies \det A \geq 0, \lambda_i \geq 0, a_{ii} \geq 0$

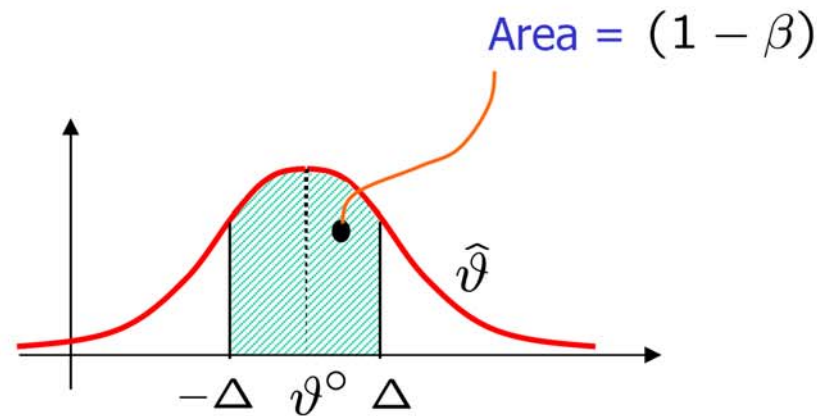
Quindi

$$\text{var} [\hat{\vartheta}^{(2)}] - \text{var} [\hat{\vartheta}^{(1)}] \geq 0 \implies \text{var} [\hat{\vartheta}_i^{(2)}] \geq \text{var} [\hat{\vartheta}_i^{(1)}]$$

dove $\hat{\vartheta}_i^{(1)}, \hat{\vartheta}_i^{(2)}$ rappresentano le componenti i -esime dei vettori $\hat{\vartheta}^{(1)}, \hat{\vartheta}^{(2)}$

Confidenza della stima

Consideriamo uno stimatore $\hat{\vartheta}$:

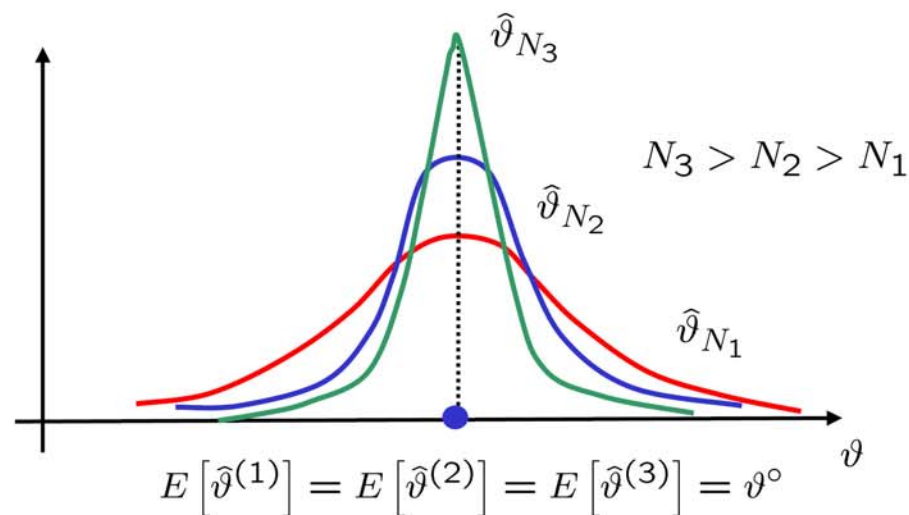


Si dice che la stima $\hat{\vartheta}$ è entro l'intervallo $(-\Delta, \Delta)$ attorno a ϑ^0 con confidenza $(1 - \beta) \cdot 100\%$

Caratteristiche asintotiche

- Se il numero N di dati a disposizione cresce nel tempo
 - ➔ aumenta l'informazione a disposizione per effettuare la stima
 - ➔ diminuisce l'incertezza
- Da questo punto di vista uno stimatore $\hat{\vartheta}_N$ è buono se

$$\lim_{N \rightarrow \infty} \text{var} [\hat{\vartheta}_N] = 0$$



- Un'altra caratterizzazione della bontà di uno stimatore $\hat{\vartheta}_N$ in cui la stima viene determinata in base ad un numero N di dati crescente nel tempo è

$$\lim_{N \rightarrow \infty} E \left[\left\| \hat{\vartheta}_N - \vartheta^\circ \right\|^2 \right] = 0 \quad (\star)$$

Se vale la $(\star)$ si dice che la stima $\hat{\vartheta}_N$ converge in media quadratica a ϑ°

- Osserviamo che $\hat{\vartheta}_N$ è un vettore casuale, ϑ° è un vettore costante e $\left\| \hat{\vartheta}_N - \vartheta^\circ \right\|^2$ è una variabile aleatoria scalare per cui l'operazione "valore atteso" è perfettamente legittima

Convergenza quasi certa

- Ricordiamo che lo stimatore basato su N dati e'

$$\hat{\vartheta}_N(s, \vartheta^\circ) = f[d(s, \vartheta^\circ)]$$

- Fissato $\bar{s} \in S$, si avra' una sequenza

$$\hat{\vartheta}_1(\bar{s}, \vartheta^\circ), \hat{\vartheta}_2(\bar{s}, \vartheta^\circ), \dots, \hat{\vartheta}_N(\bar{s}, \vartheta^\circ), \dots$$

- Potrebbe allora succedere che:

$$\bar{s} \in S \quad \longrightarrow \quad \lim_{N \rightarrow \infty} \hat{\vartheta}_N(\bar{s}, \vartheta^\circ) = \vartheta^\circ$$

$$\tilde{s} \in S \quad \longrightarrow \quad \lim_{N \rightarrow \infty} \hat{\vartheta}_N(\tilde{s}, \vartheta^\circ) \neq \vartheta^\circ$$

- Introduciamo l'insieme di esiti

$$A \subset S, \quad A = \left\{ s \in S : \lim_{N \rightarrow \infty} \hat{\vartheta}_N(s, \vartheta^\circ) = \vartheta^\circ \right\}$$

- Se $A = S$  Convergenza certa
- Se $A \subset S$ e $P(A) = 1$  Convergenza quasi-certa

Notiamo che se la misura dell'insieme $S \setminus A$ è nulla cioè
implica $P(A) = 1$ e quindi convergenza quasi-certa

- Evidentemente $A = S$  $P(A) = 1$
 Convergenza certa  Convergenza quasi-certa
- Uno stimatore per cui si abbia convergenza quasi-certa si dice **consistente**

Esempio 1

- Si considerino N dati scalari $d(1), d(2), \dots, d(N)$ tali che

$$E [d(i)] = \vartheta^\circ, \quad i = 1, \dots, N$$

- Si supponga che i dati siano mutuamente scorrelati, cioè`

$$E \{ [d(i) - \vartheta^\circ] [d(j) - \vartheta^\circ] \} = 0, \quad \forall i \neq j$$

- Si consideri lo stimatore

$$\hat{\vartheta}_N = \frac{1}{N} \sum_{i=1}^N d(i)$$

Stimatore a media campionaria

- Polarizzazione:

$$E [\hat{\vartheta}_N] = E \left\{ \frac{1}{N} \sum_{i=1}^N d(i) \right\} = \frac{1}{N} \sum_{i=1}^N E [d(i)] = \frac{1}{N} \sum_{i=1}^N \vartheta^\circ = \vartheta^\circ$$



lo stimatore non e' polarizzato

- Varianza:

$$\begin{aligned} \text{var} (\hat{\vartheta}_N) &= E \left\{ [\hat{\vartheta}_N - E(\hat{\vartheta}_N)]^2 \right\} = E \left\{ \left[\frac{1}{N} \sum_{i=1}^N d(i) - \frac{1}{N} \sum_{i=1}^N \vartheta^\circ \right]^2 \right\} \\ &= E \left\{ \frac{1}{N^2} \left[\sum_{i=1}^N d(i) - \sum_{i=1}^N \vartheta^\circ \right]^2 \right\} = \frac{1}{N^2} \sum_{i=1}^N E \left\{ [d(i) - \vartheta^\circ]^2 \right\} \\ &= \frac{1}{N^2} \sum_{i=1}^N \text{var} [d(i)] \end{aligned}$$

(i termini "misti" sono nulli per l'ipotesi di dati mutuamente scorrelati)

Se $\text{var} [d(i)] \leq \sigma, \quad 1 = 1, \dots, N$



$$\lim_{N \rightarrow \infty} \text{var} (\hat{\vartheta}_N) \leq \lim_{N \rightarrow \infty} \frac{\sigma}{N} = 0$$



lo stimatore converge in media quadratica

Esempio 2

- Si considerino N dati scalari $d(1), d(2), \dots, d(N)$ tali che

$$E [d(i)] = \vartheta^\circ, \quad i = 1, \dots, N$$

- Si supponga che i dati siano mutuamente scorrelati, cioè`

$$E \{ [d(i) - \vartheta^\circ] [d(j) - \vartheta^\circ] \} = 0, \quad \forall i \neq j$$

- Si consideri lo stimatore

$$\hat{\vartheta}_N = \sum_{i=1}^N \alpha(i) d(i)$$

- Polarizzazione:

$$E [\hat{\vartheta}_N] = E \left\{ \sum_{i=1}^N \alpha(i) d(i) \right\} = \sum_{i=1}^N \alpha(i) E [d(i)] = \vartheta^o \sum_{i=1}^N \alpha(i)$$

 lo stimatore non e' polarizzato  $\sum_{i=1}^N \alpha(i) = 1 \quad (*)$

N.B. Nel caso precedente $\alpha(i) = \frac{1}{N}$ per cui $(*)$ e' soddisfatta

La condizione $(*)$ e' un vincolo da soddisfare affinche' lo stimatore non sia polarizzato e caratterizza una classe di infiniti stimatori possibili

- Determiniamo lo stimatore migliore tra quelli non polarizzati (quindi che soddisfano il vincolo $(\star)$) andando a scegliere quello di **minima varianza**

$$\begin{cases} \min \text{var} (\hat{\vartheta}_N) = \min \sum_{i=1}^N [\alpha(i)]^2 \text{var} [d(i)] \\ 1 - \sum_{i=1}^N \alpha(i) = 0 \end{cases}$$

dati scorrelati

Usando il metodo dei moltiplicatori di Lagrange

$$J(\hat{\vartheta}) = \sum_{i=1}^N [\alpha(i)]^2 \text{var} [d(i)] + \lambda \left(1 - \sum_{i=1}^N \alpha(i) \right)$$

$$\begin{aligned} \text{L} \rightarrow \frac{\partial J}{\partial \alpha(i)} = 0 & \iff 2 \alpha(i) \text{var} [d(i)] - \lambda = 0 & \iff \alpha(i) = \frac{\lambda}{2 \text{var} [d(i)]} \end{aligned}$$

- Imponendo il vincolo di non polarizzazione (★)

$$\sum_{i=1}^N \alpha(i) = 1 \quad \longleftrightarrow \quad \frac{\lambda}{2} \sum_{i=1}^N \frac{1}{\text{var}[d(i)]} = 1 \quad \longleftrightarrow \quad \lambda = \frac{2}{\sum_{i=1}^N \frac{1}{\text{var}[d(i)]}}$$

$$\hookrightarrow \alpha(i) = \frac{1}{\text{var}[d(i)]} \alpha \quad \text{con} \quad \alpha = \frac{1}{\sum_{i=1}^N \frac{1}{\text{var}[d(i)]}}$$

Quindi $\alpha(i)$ viene scelto inversamente proporzionale alla varianza del dato $\text{var}[d(i)]$: più è grande e meno peso gli viene associato.

- Calcoliamo ora la varianza dello stimatore:

$$\begin{aligned}\text{var}(\hat{\vartheta}_N) &= E \left\{ [\hat{\vartheta}_N - E(\hat{\vartheta}_N)]^2 \right\} = E \left\{ \left[\sum_{i=1}^N \alpha(i) d(i) - \vartheta^0 \sum_{i=1}^N \alpha(i) \right]^2 \right\} \\ &= E \left\{ \left[\sum_{i=1}^N \alpha(i) [d(i) - \vartheta^0] \right]^2 \right\} = \sum_{i=1}^N \alpha(i)^2 E \{ [d(i) - \vartheta^0]^2 \} \\ &= \sum_{i=1}^N \alpha(i)^2 \text{var} [d(i)] = \alpha^2 \sum_{i=1}^N \frac{1}{\text{var} [d(i)]} = \frac{1}{\sum_{i=1}^N \frac{1}{\text{var} [d(i)]}}\end{aligned}$$

Se $\text{var} [d(i)] \leq \sigma$, $i = 1, \dots, N$

 $\lim_{N \rightarrow \infty} \text{var}(\hat{\vartheta}_N) \leq \lim_{N \rightarrow \infty} \frac{\sigma}{N} = 0$  lo stimatore converge in media quadratica

Generalizzazione

- Nel caso in cui le quantità da stimare siano **varianti nel tempo** e' necessario modificare gli indici di bonta' degli stimatori
- Si indichi con $\hat{\vartheta}(t|t-1)$ la stima di $\vartheta^o(t)$ in base ai dati fino all'istante $t-1$
- Evidentemente, siccome $\vartheta^o(t)$ varia nel tempo, non e' sensato parlare di convergenza asintotica rispetto alla disponibilita' di dati in quanto i dati nel passato possono non essere piu' significativi.

- Un criterio tipico e'

$$E \left[\left\| \hat{\vartheta}(t|t-1) - \vartheta^o(t) \right\|^2 \right] \leq c$$

dove c e' una costante ragionevolmente piccola

- **Non si chiede convergenza ma si chiede "non divergenza"**

Stima ai Minimi Quadrati (LSM)

Regressione lineare

- Si tratta del tipico contesto in cui si usa lo stimatore ai minimi quadrati (MQ)
- Si hanno $q + 1$ variabili $y(t), u_1(t), \dots, u_q(t)$ sull'arco temporale $t = 1, 2, \dots, N$
- Si vogliono calcolare (se possibile) q parametri $\vartheta_1, \vartheta_2, \dots, \vartheta_q$ tali che

$$y(t) = \vartheta_1 u_1(t) + \dots + \vartheta_q u_q(t), \quad t = 1, \dots, N \quad (\star)$$

($\star$) viene definita come la regressione lineare della variabile $y(t)$ sulle variabili $u_1(t), \dots, u_q(t)$

- Il problema può essere formulato in maniera vettoriale

$$\vartheta = \begin{bmatrix} \vartheta_1 \\ \vdots \\ \vartheta_q \end{bmatrix} \quad \varphi(t) = \begin{bmatrix} u_1(t) \\ \vdots \\ u_q(t) \end{bmatrix}$$



$$y(t) = \varphi(t)^\top \vartheta$$

- Nei problemi reali in realtà si avrà un certo errore

$$\varepsilon(t) = y(t) - \varphi(t)^\top \vartheta$$

- Obiettivo del problema delle regressione lineare sarà quindi quello di minimizzare l'errore $\varepsilon(t)$ determinando un vettore ϑ° per cui questo minimo venga raggiunto
- Si definisce la **funzione di costo quadratica**

$$J(\vartheta) = \sum_{t=1}^N [\varepsilon(t)]^2 = \sum_{t=1}^N [y(t) - \varphi(t)^\top \vartheta]^2$$



$$\vartheta^\circ = \arg \min_{\vartheta} J(\vartheta)$$

Stimatore ai minimi quadrati

- Indicando con ϑ_i la componente i-esima del vettore ϑ


$$\begin{aligned}\frac{\partial J}{\partial \vartheta_i} &= \frac{\partial}{\partial \vartheta_i} \left\{ \sum_{t=1}^N \left[y(t) - \varphi(t)^\top \vartheta \right]^2 \right\} \\ &= -2 \sum_{t=1}^N \left[y(t) - \varphi(t)^\top \vartheta \right] u_i(t), \quad i = 1, 2, \dots, q\end{aligned}$$

ed osservando che

$$\frac{\partial J}{\partial \vartheta} = \left[\frac{\partial J}{\partial \vartheta_1} \quad \frac{\partial J}{\partial \vartheta_2} \quad \dots \quad \frac{\partial J}{\partial \vartheta_q} \right]$$


$$\frac{\partial J}{\partial \vartheta} = -2 \sum_{t=1}^N \left[y(t) - \varphi(t)^\top \vartheta \right] \varphi(t)^\top$$

- Imponendo $\frac{\partial J}{\partial \vartheta} = [0 \ 0 \ \dots \ 0]$

$$\rightarrow -2 \sum_{t=1}^N [y(t) - \varphi(t)^\top \vartheta] \varphi(t)^\top = [0 \ 0 \ \dots \ 0]$$

$$\rightarrow \sum_{t=1}^N y(t) \varphi(t)^\top = \sum_{t=1}^N \varphi(t)^\top \vartheta \varphi(t)^\top$$

convertendo l'uguaglianza tra vettori riga in un'uguaglianza tra vettori colonna si può mettere in evidenza il vettore ϑ :

$$\sum_{t=1}^N \varphi(t) y(t) = \left[\sum_{t=1}^N \varphi(t) \varphi(t)^\top \right] \vartheta$$

Equazioni normali dei minimi quadrati (q eqz. in q incognite)

- Se $\sum_{t=1}^N \varphi(t) \varphi(t)^\top$ è non singolare

Formula dei minimi quadrati

$$\rightarrow \hat{\vartheta}_N = \left[\sum_{t=1}^N \varphi(t) \varphi(t)^\top \right]^{-1} \sum_{t=1}^N \varphi(t) y(t)$$

Interpretazione geometrica delle eq. normali

Poniamo

$$\varepsilon_{\vartheta}^N = \begin{bmatrix} \varepsilon_{\vartheta}(1) \\ \vdots \\ \varepsilon_{\vartheta}(N) \end{bmatrix} \quad y^N = \begin{bmatrix} y(1) \\ \vdots \\ y(N) \end{bmatrix} \quad \Phi_N = \begin{bmatrix} \varphi(1)^\top \\ \vdots \\ \varphi(N)^\top \end{bmatrix}$$

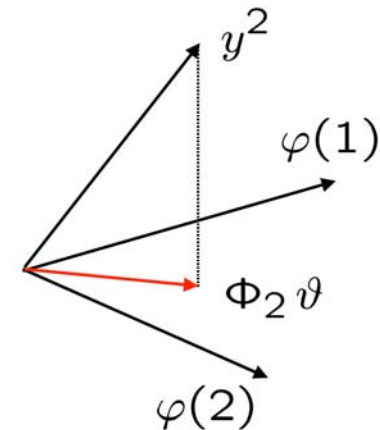
Scriviamo quindi

$$J(\vartheta) = \sum_{t=1}^N [y(t) - \varphi(t)^\top \vartheta]^2 = \|y^N - \Phi_N \vartheta\|^2$$

Evidentemente $\|y^N - \Phi_N \vartheta\|$ e`

minima quando $y^N - \Phi_N \vartheta$ e`

ortogonale a $\Phi_N \vartheta$



- Verifichiamo che $\hat{\vartheta}_N$ sia un **minimo** valutando la definitezza della matrice simmetrica

$$\left[\frac{d^2 J}{d\vartheta^2} \right]_{i,j} = \frac{\partial^2 J}{\partial \vartheta_i \partial \vartheta_j}, \quad i, j = 1, \dots, q$$

si ha

$$\left(\frac{\partial J}{\partial \vartheta} \right)^\top = 2 \left\{ \left[\sum_{t=1}^N \varphi(t) \varphi(t)^\top \right] \vartheta - \sum_{t=1}^N \varphi(t) y(t) \right\}$$


$$\frac{d^2 J}{d\vartheta^2} = 2 \left[\sum_{t=1}^N \varphi(t) \varphi(t)^\top \right]$$

evidentemente si tratta di una matrice simmetrica e semidefinita positiva

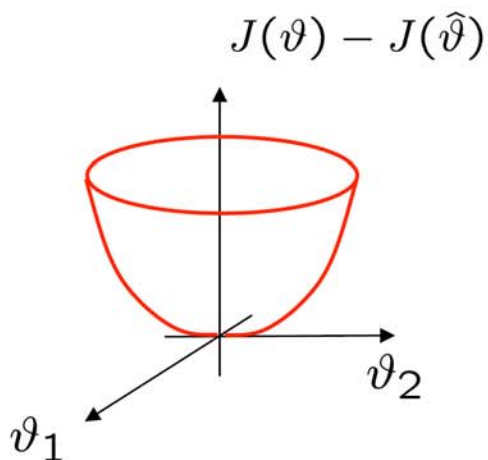


$\hat{\vartheta}_N$ e' un **minimo locale** di $J(\vartheta)$

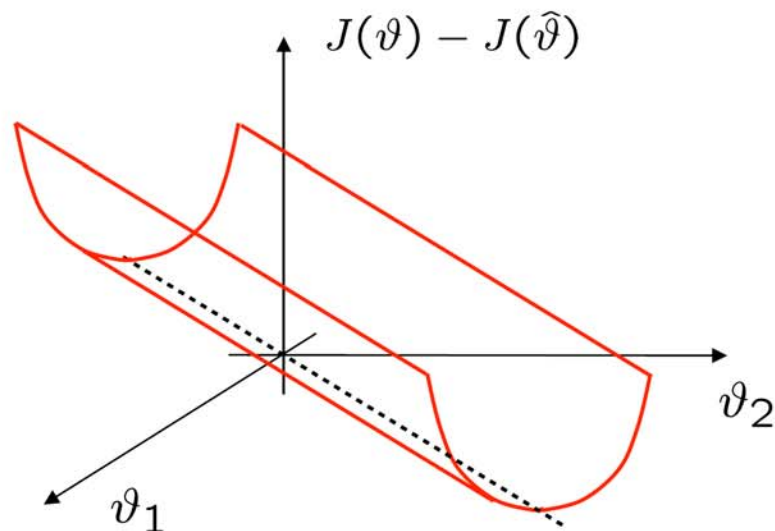
- Considerando quindi la forma quadratica

$$J(\vartheta) - J(\hat{\vartheta}) = \frac{1}{2}(\vartheta - \hat{\vartheta})^\top \left. \frac{d^2 J}{d\vartheta^2} \right|_{\hat{\vartheta}} (\vartheta - \hat{\vartheta})$$

si hanno i due casi possibili:



$$\det \left[\sum_{t=1}^N \varphi(t) \varphi(t)^\top \right] \neq 0$$



$$\det \left[\sum_{t=1}^N \varphi(t) \varphi(t)^\top \right] = 0$$

- Quindi:

- Se $\det \left[\sum_{t=1}^N \varphi(t) \varphi(t)^\top \right] \neq 0 \rightarrow \hat{\vartheta}_N$ unico minimo globale

- Se $\det \left[\sum_{t=1}^N \varphi(t) \varphi(t)^\top \right] = 0 \rightarrow \hat{\vartheta}_N$ e' uno degli infiniti minimi globali

- La condizione $\det \left[\sum_{t=1}^N \varphi(t) \varphi(t)^\top \right] \neq 0$ e' detta
e' detta condizione di identificabilita'

Caratteristiche probabilistiche dello stimatore MQ

- Supponiamo che la condizione di identificabilità sia verificata:

$$\det \left[\sum_{t=1}^N \varphi(t) \varphi(t)^\top \right] \neq 0$$

per cui

$$\hat{\vartheta}_N = \left[\sum_{t=1}^N \varphi(t) \varphi(t)^\top \right]^{-1} \sum_{t=1}^N \varphi(t) y(t)$$

- **Ipotesi:** $y(t) = \varphi(t)^\top \vartheta^\circ + \xi(t)$

dove il processo è scorrelato da $u(\cdot)$ e $E[\xi(t)] = 0$

Si sta quindi ipotizzando che il legame vero tra $y(t)$ e $u_1(t), \dots, u_q(t)$ sia proprio lineare + rumore scorrelato e a media nulla

Polarizzazione

$$\begin{aligned}\hat{\vartheta}_N &= \left[\sum_{t=1}^N \varphi(t) \varphi(t)^\top \right]^{-1} \sum_{t=1}^N \varphi(t) y(t) \\ &= \left[\sum_{t=1}^N \varphi(t) \varphi(t)^\top \right]^{-1} \sum_{t=1}^N \varphi(t) [\varphi(t)^\top \vartheta^\circ + \xi(t)] \\ &= \vartheta^\circ + \left[\sum_{t=1}^N \varphi(t) \varphi(t)^\top \right]^{-1} \sum_{t=1}^N \varphi(t) \xi(t)\end{aligned}$$


$$\hat{\vartheta}_N - \vartheta^\circ = \left[\sum_{t=1}^N \varphi(t) \varphi(t)^\top \right]^{-1} \sum_{t=1}^N \varphi(t) \xi(t)$$


$$E(\hat{\vartheta}_N - \vartheta^\circ) = \left[\sum_{t=1}^N \varphi(t) \varphi(t)^\top \right]^{-1} \sum_{t=1}^N \varphi(t) E[\xi(t)] = 0$$


$$E(\hat{\vartheta}_N) = \vartheta^\circ$$
 Stimatore MQ e' non polarizzato

OSSERVAZIONE IMPORTANTE

Nell'analisi di polarizzazione dello stimatore ai MQ abbiamo considerato il vettore dei regressori $\varphi(t)$ come **noto e fissato**. Se invece l'analisi fosse stata effettuata considerando il vettore $\varphi(s, t)$ come **vettore casuale** (e quindi funzione dell'esito s di un esperimento casuale) allora si otterrebbe che lo stimatore ai MQ e' polarizzato per qualunque valore finito di

$$E \left[\hat{\vartheta}_N - \vartheta^o \right] = E \left\{ \left[\sum_{t=1}^N \varphi(t) \varphi(t)^T \right]^{-1} \sum_{t=1}^N \varphi(t) \xi(t) \right\}$$

Varianza

Ulteriore ipotesi: $\xi(t) \sim WN(0, \lambda^2)$

Introduciamo la matrice simmetrica $S(N) = \sum_{t=1}^N \varphi(t) \varphi(t)^\top$


$$\begin{aligned} \text{var}(\hat{\vartheta}_N) &= E \left[(\hat{\vartheta}_N - \vartheta^\circ) (\hat{\vartheta}_N - \vartheta^\circ)^\top \right] \\ &= E \left\{ \left[S(N)^{-1} \sum_{t=1}^N \varphi(t) \xi(t) \right] \left[S(N)^{-1} \sum_{s=1}^N \varphi(s) \xi(s) \right]^\top \right\} \\ &= E \left\{ \left[S(N)^{-1} \sum_{t=1}^N \varphi(t) \xi(t) \right] \left[\sum_{s=1}^N \xi(s) \varphi(s)^\top S(N)^{-1} \right] \right\} \\ &= S(N)^{-1} E \left[\sum_{t=1}^N \varphi(t) \xi(t) \sum_{s=1}^N \xi(s) \varphi(s)^\top \right] S(N)^{-1} \end{aligned}$$

$S(N)$ simmetrica

Nel prodotto $\sum_{t=1}^N \varphi(t) \xi(t) \sum_{s=1}^N \xi(s) \varphi(s)^\top$ si hanno due tipi di termini:

- $\varphi(t) \xi(t)^2 \varphi(t)^\top$ se $t = s$
- $\varphi(t) \xi(t) \xi(s) \varphi(s)^\top$ se $t \neq s$

Ma $\xi(t) \sim WN(0, \lambda^2) \rightarrow E[\xi(t)\xi(s)] = \begin{cases} \lambda^2 & \text{se } t = s \\ 0 & \text{se } t \neq s \end{cases}$

 $E \left[\sum_{t=1}^N \varphi(t) \xi(t) \sum_{s=1}^N \xi(s) \varphi(s)^\top \right] = \sum_{t=1}^N \lambda^2 \varphi(t) \varphi(t)^\top = \lambda^2 S(N)$

 $\text{var}(\hat{\vartheta}_N) = S(N)^{-1} \lambda^2 S(N) S(N)^{-1} = \lambda^2 S(N)^{-1}$

Interpretazione

Supponiamo che ϑ^o sia scalare per cui anche $\varphi(t)$ e' scalare


$$y(t) = \varphi(t) \vartheta^o + \xi(t) = u(t) \vartheta^o + \xi(t)$$

da cui

$$\hat{\vartheta}_N = \left[\sum_{t=1}^N \varphi(t) \varphi(t)^\top \right]^{-1} \sum_{t=1}^N \varphi(t) y(t) = \frac{\frac{1}{N} \sum_{t=1}^N u(t) y(t)}{\frac{1}{N} \sum_{t=1}^N u(t)^2}$$

Ma:

- $\frac{1}{N} \sum_{t=1}^N u(t) y(t)$ stima campionaria della cross-correlazione $E [u(t) y(t)]$
- $\frac{1}{N} \sum_{t=1}^N u(t)^2$ stima campionaria di $E [u(t)^2]$ (varianza se $E (u) = 0$)

Inoltre

$$\text{var}(\hat{\vartheta}_N) = \lambda^2 S(N)^{-1} = \frac{1}{N} \frac{\lambda^2}{\frac{1}{N} \sum_{t=1}^N u(t)^2}$$

Pertanto:

- $\text{var}(\hat{\vartheta}_N)$ cresce al crescere di λ^2  L'incertezza sulla stima aumenta all'aumentare dell'incertezza sui dati
- a parità di N e λ^2 , $\text{var}(\hat{\vartheta}_N)$ diminuisce all'aumentare della varianza campionaria di u il che è intuitivamente corretto in quanto il rumore viene "sopraffatto" dal segnale contenente informazione utile

○ $\frac{\lambda^2}{\frac{1}{N} \sum_{t=1}^N u(t)^2}$ e' una sorta di rapporto rumore/segnale

○ Se la varianza di u e' limitata  $\lim_{N \rightarrow \infty} \text{var}(\hat{\vartheta}_N) = 0$
e siccome lo stimatore non e' polarizzato si ha:

$$\lim_{N \rightarrow \infty} E \left(\left\| \hat{\vartheta}_N - \vartheta^o \right\|^2 \right) = 0$$

ovvero lo stimatore MQ converge in media quadratica

Riferimenti Bibliografici

- C. Bonivento, C. Melchiorri, R. Zanasi. *Sistemi di controllo digitale*, Esculapio ed., Bologna, 1999
- Gene F. Franklin, J. David Powell, Abbas Emami-Naeini, *CONTROLLO A RETROAZIONE DI SISTEMI DINAMICI*, volume II. EdiSES s.r.l. 2005. ISBN: 88 7959 308 0
- S. Bittanti. *Teoria della Predizione e Filtraggio*. Pitagora, Bologna, 1992
- S. Bittanti. *Identificazione dei Modelli e Controllo adattativo*. Pitagora, Bologna, 1992