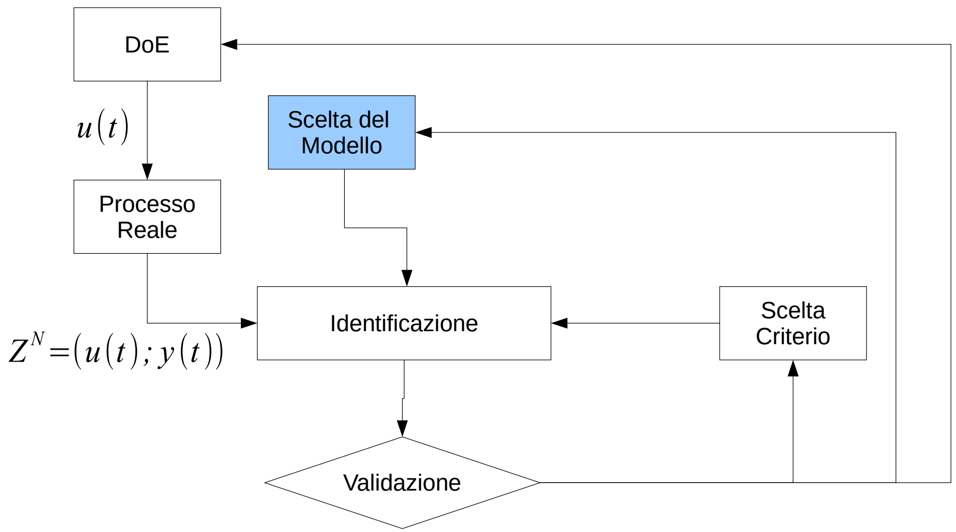
Identificazione

Principali criteri

- Problema di identificazione
- Stima PEM ed LS
- Stima ML
- Errore di stima

Processo di identificazione



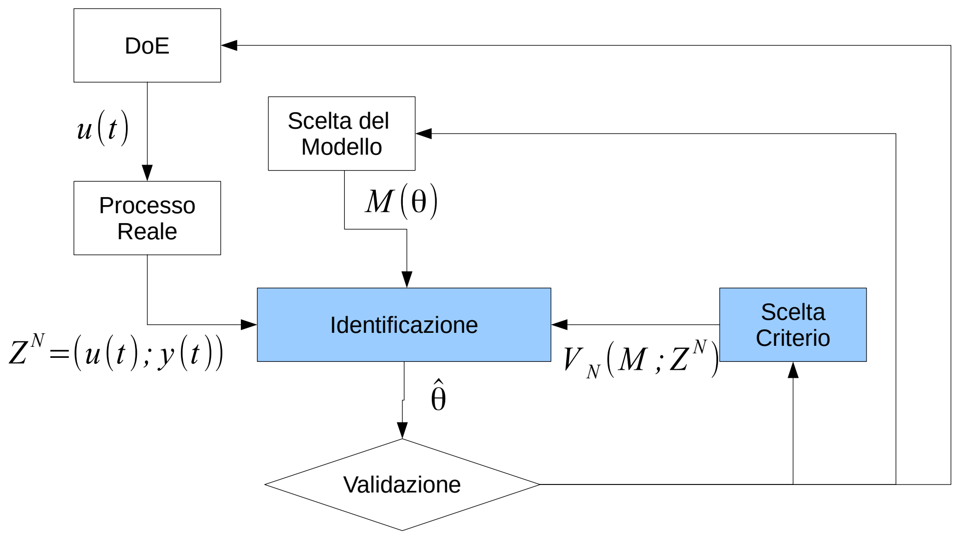
La scelta del Modello

- Scegliere un modello vuol dire:
  - Fissare i polinomi da utilizzare nel modello generale
  - Fissare il grado dei polinomi
- Nella procedura di identificazione implica fissare il numero di parametri (solitamente raccolti in un vettore colonna θ)
- Esempio : scelta di un ARX(2,1)

$$y(t)=-a_1 y(t-1)-a_2 y(t-2)+b u(t-1)+e(t)$$

$$\theta=\begin{bmatrix} a_1 & a_2 & b \end{bmatrix}^T$$

Processo di identificazione



Problema di identificazione - definizione

Dati:

- $M(\theta)$: famiglia di modelli
- Z^N: serie storiche di dati
- $V_N(M,Z^N)$: criterio di aderenza

Determinare

- Il vettore dei parametri “migliore” per dati forniti

- In simboli $\hat{\theta} = arg\,min\, V_N(M(\theta), Z^N)$

- Osservazione l'esito dell'identificazione dipende da Z^N

Criteri PEM

- Scegliere un modello che renda piccolo l'errore di predizione (prediction-error identification methods)

- L'errore di predizione è un segnale: serve una norma

$$V_N(M, Z^N) = \frac{1}{N}\sum_{i=1}^{N} \frac{1}{2}\varepsilon^2$$

$$V_N(M, Z^N) = \frac{1}{2\mathrm{N}}\sum_{i=1}^{N} (y(i) - \hat{y}(i))^2$$

Predittori lineari nei parametri

- Si introduce un vettore (detto dei regressori) tale che

$$\hat{y}(t) = \varphi^T(t)\theta$$

- Se $\varphi(t)$ non nasconde dipendenze dai parametri il modello si dice **lineare nei parametri**.
- Esempio:
  - ARX(2,1) $y(t) + a_1 y(t-1) + a_2 y(t-2) = b u(t-1) + e(t)$
  - Predittore $\hat{y}(t|\theta) = -a_1 y(t-1) - a_2 y(t-2) + b u(t-1)$

$$\theta = [a_1 \quad a_2 \quad b]^T \qquad \varphi = [-y(t-1) \quad -y(t-2) \quad u(t-1)]^T$$

Predittori non lineari nei parametri

Esempio:

- ARMAX(1,1,1)

$$y(t) + a\,y(t-1) = bu(t-1) + ce(t-1) + e(t)$$

- Predittore

$$\hat{y}(t|\theta) = -c\,\hat{y}(t-1|\theta) + (c+a)\,y(t-1) + bu(t-1) + e(t)$$

$$\theta = [-c \quad c+a \quad b]^T \qquad \varphi = [\hat{y}(t-1|\theta) \quad y(t-1) \quad u(t-1)]^T$$

Il vettore dei regressori mostra una dipendenza dai parametri

$$\hat{y}(t) = \varphi^T(t|\theta)\theta$$

Criteri PEM – minimi quadrati (LS)

- Nel caso di predittori lineari nei parametri

$$\hat{y}(i)=\varphi(i)\theta$$

- Il criterio PEM diviene

$$V_N(M,Z^N)=\frac{1}{2N}\sum_{i=1}^{N}\left(y(i)-\varphi(i)\theta\right)^2$$

che da il seguente problema di ottimo

$$\hat{\theta}^{PEM}=arg\,min\frac{1}{2N}\sum_{i=1}^{N}\left(y(i)-\varphi(i)\theta\right)^2$$

avente una forma chiusa detta "*ai minimi quadrati*"

$$\hat{\theta}^{LS}=\left[\frac{1}{N}\sum_{i=1}^{N}\varphi(i)\varphi'(i)\right]^{-1}\frac{1}{N}\sum_{i=1}^{N}\varphi(i)y(i)$$

Criteri PEM - Osservazioni

- Definizione criterio semplice ed intuitiva

$$V_N(M,Z^N)=\frac{1}{2N}\sum_{i=1}^{N}\left(y(i)-\hat{y}(i)\right)^2$$

- Nel caso di predittori lineari nei parametri esiste una forma chiusa.

$$\hat{\theta}^{LS}=\left[\frac{1}{N}\sum_{i=1}^{N}\varphi(i)\varphi'(i)\right]^{-1}\frac{1}{N}\sum_{i=1}^{N}\varphi(i)y(i)$$

Non è possibile includere informazioni a priori.

Criteri Maximum Likelihood – Idea

- Idea
  - Definisco i dati come vv. cc. con ddp legata a θ.
  - La stima sarà data dal vettore $\hat{\theta}^{ML}$ che massimizza la probabilità di "ottenere" i dati osservati.

- Esempio semplice non legato ai modelli:

determinare la probabilità p di ottenere testa lanciando una moneta note 4 realizzazioni $R=[r_1,r_2,r_3,r_4]$.

Se dovessi ottenere tutti teste mi aspetterei *p* = *1*.

Criteri ML – Toy Example

determinare la probabilità p di ottenere testa lanciando una moneta note 4 realizzazioni $R=[r_1,r_2,r_3,r_4]$.

- Posto 1 (testa) e 0 (croce) la i-sima realizzazione R_i.

$$P(R_i=r_i)=r_i\,p+(1-r_1)(1-p)$$

- La congiunta diviene

$$f_R=P(R=[r1,r2,r3,r4])=\prod_{i=1}^{4}\left(r_i\,p+(1-r_1)(1-p)\right)$$

- Supposto di avere osservato $R=[1,1,1,1]$

$$\prod_{i=1}^{4}\left(1\,p+(1-1)(1-p)\right)=p^4$$

- Che porta alla seguente stima

$$\hat{\theta}^{ML}=\underset{p\in[0;1]}{arg\ max}\ p^4=1$$

Criteri Maximum Likelihood – Idea - II

- Idea
  - Definisco i dati come vv. cc. con ddp legata a θ.
  - La stima sarà data dal vettore $\hat{\theta}^{ML}$ che massimizza la probabilità di “ottenere” i dati osservati
- Requisiti
  - conoscere la distribuzione dell'errore f_e.
  - calcolare il predittore ottimo

$$\hat{y}(t|\theta)=g_p(u(t),y(y),\theta)$$

- Se il modello e sistema condividono struttura e ordine:

$$y(t)-g_p(u(t),y(y),\theta)=\varepsilon(t|\theta)=e(t)\sim f_e$$

- Osservazione: si includono le informazioni a priori sull'errore

Criteri ML - Verosimiglianza

Dati predittore e modello

$$\hat{y}(t|\theta)=g_p(Z^{t-1},\theta) \qquad y(t)=g_p(Z^{t-1},\theta)+\varepsilon(t|\theta)$$

Ricavo la relazione con l'errore di predizione

$$\varepsilon(t|\theta)=y(t)-g_p(Z^{t-1},\theta)$$

Applicandolo alle N misure Z^N ho la verosimiglianza

$$f_Z(\theta,Z^N)=\prod_{t=1}^{N} f_e(y(t),Z^{t-1},\theta)$$

- Verosimiglianza f_Z: funzione che descrive la ddp congiunta delle realizzazione di ε in funzione di Z^N e θ.
- Osservazione: possibile interpretazione $P(Z^N|\theta)\sim f_Z$

Criteri ML - Criterio

- Il criterio è quello di massimizzare la verosimiglianza.
- il problema di identificazione è un problema di minimo uso

$$V_N(M,Z^N)=-f_Z(\theta,Z^N)$$

- Osservazione: la verosimiglianza è (spesso) un prodotto.
- Spesso si utilizza come criterio la log-verosimiglianza

$$V_N(M,Z^N)=-\ln(f_Z(\theta,Z^N))$$

- Osservazione: nella la verosimiglianza vi sono molti termini noti e di regolarizzazione
- Il problema di minimo (spesso) può essere semplificato.

Criteri ML – Esempio - I

- Modello

$$y(t)=b*u(k-1)+e(t) \qquad e(t)\sim Z\rightarrow f_e=\frac{1}{\sqrt{2\pi}}e^{\frac{-x^2}{2}}$$

- Predittore

$$y(t|b)=b*u(k-1)$$

- Relazione con errore di predizione

$$f_e(y(t)-g_p(Z^{t-1},\theta))=\frac{1}{\sqrt{2\pi}}e^{-\frac{(y(t)-bu(t))^2}{2}}$$

- Verosimiglianza per N dati

$$f_Z=\frac{1}{(2\pi)^{N/2}}\prod_{t=1}^{N} e^{-\frac{(y(t)-bu(t))^2}{2}}$$

Criteri ML - Esempio - II

$$f_Z = \frac{1}{(2\pi)^{N/2}} \prod_{t=1}^{N} e^{-\frac{(y(t)-bu(t))^2}{2}}$$

- Log Verosimiglianza

$$\ln(f_Z) = -\frac{N}{2}\ln(2\pi) - \frac{1}{2}\sum_{t=1}^{N}(y(t)-bu(t))^2$$

Criterio

$$V(Z^N, \theta) = \frac{N}{2}\ln(2\pi) + \frac{1}{2}\sum_{t=1}^{N}(y(t)-bu(t))^2$$

- Problema di minimo

$$\bar{\theta}^{ML} = arg\,min \frac{N}{2}\ln(2\pi) + \frac{1}{2}\sum_{t=1}^{N}(y(t)-bu(t))^2$$

$$\bar{\theta}^{ML} = arg\,min \sum_{t=1}^{N}(y(t)-bu(t))^2$$

Criteri ML - Osservazioni

- Definizione criterio complessa

$$V_N(M, Z^N) = -\ln(f_Z(Z^N, \theta))$$

- Richiede la descrizione completa del modello (f_e)
- E' possibile includere informazioni a priori.
- A volte si riottiene la stessa soluzione dei LS.

Errore di stima

- Quanto son andato bene nel riconoscere il modello?

Criterio naive: errore di stima

- Requisiti
  - il sistema (modello vero) stessa struttura ed ordine del modello scelto

$$S \in M(\theta)$$

  - conosco il modello vero

$$S = M(\theta_0)$$

- Definisco errore di stima

$$\varepsilon_s = \hat{\theta} - \theta_0$$

Errore di stima - metriche

- Sull'errore di stima si possono basare diverse metriche
  - Max: usata per fornire un upper bound $\|\hat{\theta}-\theta_0\|_\infty$
  - Average: una indicazione media $\|\hat{\theta}-\theta_0\|_2$
- Indicatori teorici, difficilmente utilizzabili in pratica
- Nella pratica conviene affidarsi a
  - Grafico delle serie storiche $y(t)$ ed $y(t|t\text{-}1)$
  - Le cifre di merito dell'ottimizzazione
    - SSR (sum of squared errors of prediction)
    - Likelihood