

RIASSUNTO DI MODELLI STOCASTICI (seconda parte)

Argomenti trattati: (quelli in rossi sono presenti)

- *Analisi di sensitività*
- *Processi stocastici parte 1 (famiglie di modelli AR, MA)*
- *Processi stocastici parte 2 (ARMA, ARMAX, Predizione)*
- *Processi stocastici parte 3 (Validazione)*
- *Modello nello spazio degli stati (filtro di Kalman)*
- *Catene di Markov*
- *Processi a memoria lunga*
- *Laboratorio: Matlab ed R*

Analisi di sensitività

In questa lezione discutiamo come quantificare l'effetto degli input di un Computer Model CM sugli output su base statistica.

– Fonti di incertezza e variabilità:

Indichiamo il "vero" fenomeno di interesse con ζ e supponiamo che siano legati al vettore di input osservabili $x = (x_1, \dots, x_k) \in D$ e ad altri input non-osservabili $x^* : \zeta = \zeta(x, x^*)$.

Il computer model (CM) o "codice" è una funzione computabile $z = f(x)$.

Consideriamo principalmente CM deterministici: se re-eseguiamo il codice con le stesse x , abbiamo lo stesso risultato. Nel caso, semplicistico, ideale il CM è un modello perfetto: $f(x) = \zeta(x, x^*)$ per ogni x .

Incertezza di modello $\rightarrow \zeta - f(x)$, Incertezza degli input $\rightarrow P(x)$, Emulatori $\rightarrow g(x, \beta) \approx f(x)$,

Emulatore stimato $\rightarrow g(x) = g(x, \beta)$.

– Incertezza dell'output

L'incertezza sull'input x si propaga all'output z attraverso il CM, f_x , così che, essendo x vettore casuale con distribuzione p_x , siamo interessati alla distribuzione dell'incertezza sull'output, p_z che è riferito a p_x attraverso il codice f_x . Per esempio nell'analisi dei rischi siamo interessati alla distribuzione cumulativa dell'uscita P_z . Una tipica quantità per stimare l'incertezza quadrata è la varianza dell'uscita, che di solito è calcolata usando la distribuzione di incertezza dell'input p_x :

$$Var(z) = \sigma_z^2 = \int_D (f(x) - f_0)^2 p(x) dx$$

dove $f_0 = E(z)$.

SA basata sulla varianza: locale basata su \rightarrow

$$\frac{\partial f(x_0)}{\partial x_j}$$

e globale.

– Indice di sensitività:

Ricordando la scomposizione $Var(z) = Var(E(z|x)) + E(Var(z|x))$ il rapporto di correlazione di Pearson

$$S_j = \frac{Var(E(z|x))}{\sigma_z^2}$$

è la quota di variabilità di z che può essere eliminata fissando x_j e costituisce un naturale indice di sensitività del prim'ordine per fenomeni a varianza finita.

– GAM:

Consideriamo un modello semplificato ad effetti principali GAM

$$z = f_0 + \sum_{j=1}^{\kappa} f(x_j) + \varepsilon$$

dove $f_0 = E(z)$ e, per esempio, $f_j = E(z|x_j) - f_0$. Se gli input sono indipendenti possiamo decomporre la varianza come:

$$Var(z) = \sum_{j=1}^k Var(f_j) + Var(\varepsilon)$$

con indici di sensitività:

$$S_j = \frac{Var(f_j)}{\sigma_z^2}$$

- Modello con interazioni:

In principio consideriamo un modello con tutte le interazioni

$$z = f_0 + \sum_j f_j + \sum_{i < j} f_{ij} + \dots + f_{j_1, \dots, j_k}$$

dove, per esempio, $f_{ij} = E(z|x_i, x_j) - f_i - f_j - f_0$

In questo caso, per prendere in considerazione l'effetto delle interazioni fra x_j e gli altri input, S_j deve essere aumentato per avere l'effetto totale. Sia x_j il vettore $(k-1)$ -dim corrispondente ad x senza la j^{th} componente, sia $f_{(j)}(x_{(j)}) = E(z|x_j)$ e così via. La decomposizione di cui sopra può essere riscritta $z = f_0 + f_j + f_{(j)} + f_{(j,j)}$

Grazie all'indipendenza fra gli input abbiamo: $Var(z) = Var(f_0) + Var(f_j) + Var(f_{(j)}) + Var(f_{(j,j)})$

e l'indice di sensitività totale è dato da

$$S_{T_j} = \frac{Var(f_j) + Var(f_{(j,j)})}{\sigma_z^2} = 1 - S_{(j)}$$

Dettagli di calcolo:

Siano D_j e $D_{(j)}$ i domini di x_j e $x_{(j)}$ rispettivamente. Allora la risposta media ad x_j è data dall'integrale $(k-1)$ dimensionale:

$$f_j = E(z|x_j) = \int_{D_{(j)}} f(x) p(x_{(j)}) dx_{(j)}$$

e la sua varianza è data dall'integrale unidimensionale:

$$Var(E(z|x_j)) = \int_{D_j} E(z|x_j)^2 p(x_j) dx_j - f_0^2$$

Inoltre, $Var(f_{(j)})$, che prende parte alla sensitività totale S_{T_j} , è data da:

$$Var(E(z|x_{(j)})) = \int_{D_{(j)}} E(z|x_{(j)})^2 p(x_{(j)}) dx_{(j)} - f_0^2$$

Queste quantità possono essere stimate con poche repliche MC in taluni casi.

SA via emulatore:

Quando la risposta è lineare

$$g(x, \beta) = \beta_0 + \sum_{j=1}^k x_j \beta_j$$

con errori $NID(0,2)$, gli indici di sensitività S_j "nella popolazione" derivano dalla seguente decomposizione della varianza:

$$\sigma_z^2 = \sum_{j=1}^k \sigma_{x_j}^2 \beta_j^2 + \sigma_\varepsilon^2$$

Così se abbiamo un campione Monte Carlo abbastanza ampio e usiamo le stime LS di β abbiamo:

$$\hat{S}_j = \frac{\hat{\sigma}_{x_j}^2 \hat{\beta}_j^2}{\hat{\sigma}_y^2}$$

$$\sum \hat{S}_j = R^2 = 1 - \frac{\hat{\sigma}_\varepsilon^2}{\hat{\sigma}_y^2}$$

che sommano

Questo approccio si estende facilmente al caso polinomiale con input trasformati: $z = g(x, \beta) = h(x)' \beta + \varepsilon$.

Processi stocastici parte 1

Introduzione

La teoria dei processi stocastici riguarda lo studio di sistemi che evolvono nel tempo (ma anche più in generale nello spazio) secondo leggi probabilistiche. Tali sistemi o modelli descrivono fenomeni complessi del mondo reale che hanno la possibilità di essere aleatori. Tali fenomeni sono più frequenti di quanto si possa credere e si incontrano tutte quelle volte in cui le quantità alle quali siamo interessati non sono prevedibili con certezza. Quando però tali fenomeni mostrano una variabilità di possibili esiti che può essere in qualche modo spiegata, allora possiamo introdurre il modello probabilistico del fenomeno.

In **teoria della probabilità** un **processo stocastico** è una generalizzazione dell'idea di **variabile casuale**, e può euristicamente essere interpretato come una variabile casuale che prenda valori in spazi più generali dei **numeri reali** (come ad esempio, \mathbb{R}^n , o **spazi funzionali**, o **successioni** di numeri reali). Pertanto, è in genere possibile identificare un processo stocastico come una famiglia ad un parametro di variabili casuali reali. Supponiamo ad esempio di voler modellizzare matematicamente la dinamica di un punto che si muove su di una retta con una legge probabilistica. Possiamo introdurre un processo stocastico come la collezione delle variabili casuali, dove per ogni valore della variabile *tempo* t , X_t è semplicemente la variabile casuale (reale) che esprime la legge probabilistica del punto considerato al tempo t .

Processi Stocastici a tempo discreto:

Famiglia di variabili casuali y_t (con $t = 0, \pm 1, \pm 2, \dots$) caratterizzata dalle sue distribuzioni congiunte $f(y_{t_1}, y_{t_2}, \dots, y_{t_k})$; per esempio il random walk: $y_t = y_{t-1} + \epsilon_t$ con $\epsilon_t \text{ NID}(0, \sigma^2)$.

Distribuzioni di un PS:

1. Distribuzioni Marginali: $f(y_t) = f_t(y_t)$ $t=0, \pm 1, \pm 2, \dots$; dipende solo da sé stessa in un istante tempo.
2. Distribuzioni Doppie: $f(y_t, y_{t+h}) = f_{t, t+h}(y_t, y_{t+h})$, \dots , $f(y_t, y_{t+h}) = f_{t, t+h}(y_t, y_{t+h})$, \dots $t, h=0, \pm 1, \pm 2, \dots$; h è un intervallo di tempo.
3. Distribuzioni Triple: $f(y_t, y_{t+h}, y_{t+k}) = f_{t, t+h, t+k}(y_t, y_{t+h}, y_{t+k})$, \dots $t, h, k=0, \pm 1, \pm 2, \dots$.

Quantità caratteristiche di un processo stocastico $y(t)$:

✓ **Valor medio** $m(t)$: è il valore atteso della variabile casuale $y(t)$ al tempo t : $m(t) = E[y(t)]$

✓ **Funzione di covarianza**: $\gamma(t_1, t_2)$ è il valore atteso del prodotto tra la variabile casuale depolarizzata ($y(t) - m(t)$) considerata

in due diversi istanti di tempo (t_1, t_2): $\gamma(t_1, t_2) = E[(y(t_1) - m(t_1))(y(t_2) - m(t_2))]$. Caso particolare della funzione

di covarianza: per $t_1 = t_2$, abbiamo che $\gamma(t) = E[(y(t) - m(t))^2]$ e $\gamma(t)$ diventa la varianza del processo stocastico.

Processi Markoviani: $f(y_t | y_{t-1}, y_{t-2}, \dots) = f(y_t | y_{t-1})$: l'osservazione al tempo t , condizionata a tutto il passato, dipende solo dalla singola osservazione al tempo $t-1$.

Allora $f(y_t | y_{t-1}, y_{t-2}, \dots, y_0) = \prod_{j=1}^t f(y_j | y_{j-1})$.

Teoria della Previsione:

Previsore a k passi di y_{t+k} : $\hat{y}_{t+k} = f_k(y_t, y_{t-1}, \dots) = f_k(S_t)$, dove S_t è l'insieme delle informazioni (storia) fino a t . La previsione ottima (a minima varianza) è $\hat{y}_{t+k} = E(y_{t+k} | S_t)$. In particolare è importante la previsione ad un passo $\hat{y}_{t+1} = E(y_{t+1} | S_t)$. L'errore di previsione $e_t = y_t - \hat{y}_t$:

- ha media nulla;
- è incorrelato (ortogonale) col passato $E(e_t | y_{t-1}) = 0$, con $j=1, 2, \dots$; • La sua varianza determina la precisione della previsione: $\text{var}(e_t) = E(e_t^2)$.

Supporremo nel seguito omoschedasticità: $\text{var}(e_t) = \text{var}(e_{t-1}) = E(e_t^2 | S_{t-1})$.

Esempio AR(1): Previsione a un passo $\hat{y}_t = a y_{t-1}$. Previsione a k passi $\hat{y}_t = a^k y_{t-k}$ (deriva dalla lunga divisione con $C=1$ - vedi IMAD).

NB: Memoria breve: $a \rightarrow 1 \Rightarrow \hat{y}_t = y_t$ per $k \rightarrow \infty$ dopo un certo numero di passi, il predittore ottimo diventa la media, che è 0.

Stazionarietà: Un processo stocastico si dice stazionario se le sue distribuzioni dipendono solo dall'intervallo di tempo:

la correlazione tra "oggi e domani" è uguale a quella tra "un mese e un mese + 1 giorno". Conseguenze:

$f_{y_t, y_{t+h}} = f_{y_t, y_{t+h}}$ per ogni t ; $E[y_t] = \mu$ per ogni t ; $Var[y_t] = \sigma^2$ per ogni t ; $E[y_t y_{t+h}] = \gamma(h)$.

Un processo stocastico si dice stazionario (in senso "debole") se valgono le seguenti proprietà:

- $m_{y_t} = m \forall t$;
- $\gamma(t_1, t_2)$ deve essere solo funzione di $\tau = |t_1 - t_2|$, cioè: $\gamma(t_1, t_2) = \gamma(t_3, t_4)$ se $|t_1 - t_2| = |t_3 - t_4|$ per ogni t_1, t_2, t_3, t_4 .

Se il processo stocastico è stazionario, la scrittura della funzione di covarianza può essere semplificata:

$$\gamma(\tau) = E[(y(t) - m)(y(t - \tau) - m)].$$

Proprietà della funzione di covarianza di un p. s. s. :

- (1) $\gamma(0) = E[(y(t) - m)^2] = \sigma^2$, cioè per $\tau=0$, la funzione di covarianza è pari alla varianza.
- (2) $\gamma(0) \geq \gamma(\tau)$, cioè la varianza limita superiormente la funzione di covarianza.
- (3) $\gamma(\tau) = \gamma(-\tau)$, cioè la funzione di covarianza è simmetrica.

Autocovarianza: $\gamma_t(h) = Cov[y_t, y_{t+h}] = E[(y_t - \mu_t)(y_{t+h} - \mu_{t+h})]$.

Se il PS è stazionario allora l' autocovarianza dipende solo dalla distanza $|h|$:

$$\gamma_t(h) = \gamma(|h|) = E[(y_t - \mu)(y_{t+h} - \mu)] = \sigma_y^2 \cdot \rho(h)$$

Autocorrelazione: $\rho_t(h) = \rho(y_t, y_{t+h}) = Cov[y_t, y_{t+h}] / \sqrt{Var[y_t] Var[y_{t+h}]}$

Se il PS è stazionario allora $\rho_t(h) = \rho(|h|) = Cov[y_t, y_{t+h}] / \sigma_y^2$

Inoltre, se $E(y) = 0$: $\rho_t(h) = E[y_t y_{t+h}] / E[y_t^2]$.

Correlogramma: L'autocorrelazione campionaria, è definita da $r(h) = (y_t - \bar{y})(y_{t+h} - \bar{y}) / (y_t^2 - n \bar{y}^2)$.

Se si usano gli scarti dalla media si ha $\bar{y} = 0$ ed $r(h) = y_t y_{t+h} / y_t^2$.

Stima del 2° ordine: Stazionarietà + ergodicità \rightarrow Stimabilità tramite i momenti campionari:

(1) Media campionaria: $\bar{y}_n := (1/n) \sum_{t=1}^n y_t \rightarrow \mu$: è uno stimatore corretto, ma non sempre consistente (per i processi ARMA è comunque sempre consistente).

(2) Funzione di covarianza campionaria: $\hat{\gamma}(h) := (1/n) \sum_{t=1}^{n-h} (y_t - \bar{y})(y_{t+h} - \bar{y}) \rightarrow \gamma(h)$ la sua significatività decresce all'aumentare di τ (quindi si calcola normalmente per $\tau \ll N$); è uno stimatore corretto.

$r(h) := \hat{\gamma}(h) / \hat{\gamma}(0) \rightarrow \rho(h)$; ciò dipende dalla velocità (memoria lunga/corta) con cui $\rho \rightarrow 0$ ed $h \rightarrow \infty$.

Famiglie di modelli: descriviamo la relazione $y(t) \leftrightarrow e(t)$ con delle equazioni alle differenze lineari.

$\rightarrow MA(1) - Moving Average$: $y_t = \epsilon_t + c \epsilon_{t-1}$ (ϵ_t iid $N(0, \sigma_\epsilon^2)$): l'uscita di un processo MA è, dunque, la combinazione lineare degli ultimi $n + 1$ valori del segnale di ingresso $\epsilon(t)$. NB: non è markoviano, infatti, per $|b| < 1$, sostituendo ricorsivamente $\epsilon_t = y_t - b \epsilon_{t-1}$ si ottiene, $\epsilon_t = \sum_{j=0}^{\infty} (-b)^j y_{t-j} + \epsilon_t$.

Si ha memoria finita e moderata; infatti: $E[y_t] = 0$

$$\gamma_{y_t} = Var[y_t] = (1 + c^2) \sigma_\epsilon^2 \quad \gamma_{y_t, t+h} = E[(y_t + c \epsilon_{t-1})(y_{t+h} + c \epsilon_{t+h-1})] = c^2 \sigma_\epsilon^2 \quad \gamma_{y_t, t+h} = 0 \quad \forall h > 1 \quad \rho(h) = c / (1 + c^2)$$

$\rho(h) = 0$ $h > 1$. Un processo MA(n) ha n zeri e n poli ma tutti nell'origine \rightarrow è sempre asintoticamente stabile. Dato che i poli di un MA sono "banali", i processi MA sono detti "all-zeroes".

$\rightarrow AR(1) - Auto Regressive$: $y_t = \alpha y_{t-1} + \epsilon_t$ ($\epsilon_t \equiv NID(0, \sigma_\epsilon^2)$): l'uscita di un processo AR è, dunque, la combinazione lineare degli ultimi m "vecchi" valori del processo stesso, più l'ingresso $\epsilon(t)$ allo stesso istante.

Ipotizzando $|\alpha| < 1$ si deduce che $y_t = \sum_{j=0}^{\infty} \alpha^j \epsilon_{t-j}$.

Il Random walk è un particolare tipo di AR(1), con modello di persistenza pari a:

$$y_t = y_{t-1} + \epsilon_t \text{ con } \epsilon_t \text{ NID}(0, \sigma_\epsilon^2). \text{ Segue che: } E y_t = 0 \quad Var[y_t] = t \sigma_\epsilon^2 \rightarrow \infty \quad t \rightarrow \infty$$

$$\rho(y_t, y_{t+h}) = \sqrt{t / (t+h)}$$

$t+h \rightarrow 1, t \rightarrow \infty$ Un AR(m) ha m zeri tutti nell'origine (è sempre a fase minima) e m poli. Dato che gli zeri di un AR sono "banali", i processi AR sono detti "all-poles".

$\rightarrow AR(p)$: $y_t = a_1 y_{t-1} + a_2 y_{t-2} + \dots + a_p y_{t-p} + \epsilon_t$ (ϵ_t iid $N(0, \sigma_\epsilon^2)$) $e_t = 0, \pm 1, \pm 2, \dots$. E' markoviano? Posto

$Y_{t-1} = (y_{t-1}, y_{t-2}, \dots, y_{t-p})'$ e $\epsilon_t = (\epsilon_t, 0, \dots, 0)'$ abbiamo la forma markoviana: $Y_t = A Y_{t-1} + \epsilon_t$, con A matrice con: prima riga con coefficienti di a e per il resto tutti 0 e 1 solo sulla diagonale.

Forma polinomiale di un AR: sia B l'operatore di ritardo: $B y_t = y_{t-1}$.

Allora: $a(B) y_t = \epsilon_t$, con $a(B) = 1 - a_1 B - \dots - a_p B^p$. L'operatore B è un operatore che ritarda di un passo un

segnale a tempo discreto; analogamente l'operatore B^{-1} è l'operatore che anticipa di un passo un segnale a tempo discreto. È un operatore lineare che può essere applicato ricorsivamente.

Stabilità: Dato un sistema, rappresentato da una funzione di trasferimento $W(z)$ e ipotizzando che $W(z)$ sia scritta in potenze positive di z : i poli del sistema sono le radici del denominatore di $W(z)$ e gli zeri del sistema sono le radici del numeratore di $W(z)$.

Sia z appartenente a \mathbb{C} la variabile complessa. Il sistema rappresentato da $W(z)$ è asintoticamente stabile se e solo se tutti i poli del sistema stesso sono strettamente interni al cerchio unitario.

Se il sistema ha tutti gli zeri strettamente interni al cerchio unitario, si dice "a fase minima". Il PS y_t , $AR(p)$ è stazionario se e solo se $a_1/z = 1 - a_1/z - \dots - a_p/z^p \neq 0$ per $|z| = 1$ o equivalentemente le soluzioni dell'eq. caratteristica $a_1/z = 0$ sono interne al cerchio unitario $|z| < 1$.

Forma $MA(\infty)$: se y_t è un AR stazionario allora $y_t = 1/a_1 B^p \epsilon_t = \sum_{j=1}^{\infty} w_j \epsilon_{t-j} + \epsilon_t$.

Memoria breve: con stazionarietà, $\rho(h) = \zeta^h$, con $|\zeta| < 1$, cioè $\rho(h) \rightarrow 0$ con velocità esponenziale per $h \rightarrow \infty$.

Proc. Gaussiani: $\mathbf{y} = (y_1, \dots, y_k) \equiv N_k(\boldsymbol{\mu}, \Gamma_k)$; $E(y_t, y_1, \dots, y_{k-1})$ è lineare in y_1, \dots, y_{k-1} ;

$Var(y_t, y_1, \dots, y_{k-1}) = \sigma_k^2$ è indipendente da y_1, \dots, y_{k-1} .

Processi condizionatamente Gaussiani: $(y_t, y_1, \dots, y_{k-1} = S_t) \equiv N^k(\boldsymbol{\mu}_t, \Sigma_t)$. Lineari / Non-lineari.

Se tutte le distribuzioni sono gaussiane, perché un $AR(p)$ sia gaussiano è sufficiente che l'innovazione sia gaussiana.

Modello nello spazio degli stati

Consideriamo un sistema il cui stato è definito dal vettore m -dim, indicato con X_t , e di cui si dispongono le osservazioni k -dim, indicate con Y_t . Indichiamo con U_t il vettore delle covariate "non-stocastiche".

– Equazione di stato

La dinamica del fenomeno è data da un modello markoviano $X_t = AX_{t-1} + BU_t + R_t \eta_t$ con $\eta_t \sim NID(0, Q_t)$.

NB: l'innovazione del sistema $e_t = R_t \eta_t$ ha matrice di var-cov $V(e_t) = RQR'$ e la coppia (R, Q) non è univocamente determinata senza ulteriori vincoli, tuttavia questa rappresentazione si rivela spesso utile.

– Equazione di misura

Lo stato del sistema X_t non è direttamente osservabile e le osservazioni sono legate agli stati dall'equazione:

$$Y_t = CX_t + DU_t + \epsilon_t \text{ con } \epsilon_t \sim NID(0, H_t)$$

Correlazione fra le due equazioni: Per ora assumiamo $E(\epsilon_t \eta_t') = 0$

Valori iniziali: $E(X_0) = x_0$ e $V(X_0) = P_0$

Esempi: Local level

Consideriamo il modello: $y_t = \mu_t + \epsilon_t$, $\mu_t = \mu_{t-1} + \eta_t$

con $Var(\epsilon_t) = \sigma_\epsilon^2$ e $Var(\eta_t) = \sigma_\eta^2$

Rappresentazione SS di MA(1):

Il modello $MA(1)$ $y_t = c\epsilon_{t-1} + \epsilon_t$ ammette una rappresentazione SS con stato

$$X_t = \begin{pmatrix} y_t \\ c\epsilon_t \end{pmatrix}$$

e equazione di transizione

$$X_t = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} X_{t-1} + \begin{pmatrix} 1 \\ c \end{pmatrix} \varepsilon_t$$

ed equazione di misura

$$y_t = \begin{pmatrix} 1 & 0 \end{pmatrix} X_t + 0$$

Filtro di Kalman:

Supponiamo note le matrici A, B, C, D, H, Q, R , x_0, P_0 e costanti nel tempo. Ci poniamo il problema di **"stimare ricorsivamente"** lo stato X_t partendo dalle osservazioni Y_t (e U_t) in ipotesi di normalità.

$$S_t = (Y_1, \dots, Y_t, U_1, \dots, U_t)$$

Sia $X_t^\wedge = E(X_t | S_t)$ e sia P_{t-1} la matrice di var-cov dell'**errore di stima** $P_t = E(X_t - X_t^\wedge)(X_t - X_t^\wedge)'$,

Noti X_{t-1}^\wedge e P_{t-1} la previsione ottima di X_t è

$$\hat{X}_{t|t-1} = E(X_t | \hat{X}_{t-1}, U_t) = A\hat{X}_{t-1} + BU_t$$

che ha **errore di previsione**

$$X_t - \hat{X}_{t|t-1} = A(X_{t-1} - \hat{X}_{t-1}) + \eta_t$$

con matrice di var-cov

$$P_{t|t-1} = E(X_t - \hat{X}_{t|t-1})(X_t - \hat{X}_{t|t-1})'$$

$$= AP_{t-1}A' + RQR'$$

All'istante successivo osserviamo Y_t e la stima di X_t può essere aggiornata alla luce dell'errore di previsione sull'osservazione Y_t $e_t = Y_t - \hat{Y}_{t|t-1} = Y_t - CX_{t|t-1} - DU_t$

$$= C(X_t - X_{t|t-1} - DU_t) + \varepsilon_t$$

Si ottiene la nuova stima di X_t

$$\hat{X}_t = \hat{X}_{t|t-1} + P_{t|t-1}C'F_t^{-1}e_t$$

e $P_t = P_{t|t-1} - P_{t|t-1}C'F_t^{-1}CP_{t|t-1}$ con, F_t assunta invertibile e data da: $F_t = CP_{t|t-1}C' + H = V_{t|t}(e_t)$

Dimostrazione

In ipotesi di normalità degli errori, costruendo iterativamente le Y_t e le X_t per $t = 1, 2$, condizionatamente ai valori iniziali ed alle covariate, si deduce che $(X_1, Y_1), \dots, (X_n, Y_n)$ hanno distribuzione congiunta gaussiana.

Pertanto il calcolo di $X_t^\wedge = E(X_t | S_t)$ equivale al calcolo del piano di regressione di una normale multivariata.

A tal fine ricordiamo che, per la normale multivariata

$$\mathbf{x} = \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{pmatrix} \equiv N_{k_1+k_2} \left(\begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \right)$$

il piano di regressione è dato da

$$E(\mathbf{x}_1 | \mathbf{x}_2 = \mathbf{x}) = \boldsymbol{\mu}_1 + \Sigma_{12}\Sigma_{22}^{-1}(\mathbf{x} - \boldsymbol{\mu}_2)$$

con varianza condizionata

$$Var(\mathbf{x}_1 | \mathbf{x}_2) = \Sigma_{1|2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$$

Nel nostro caso interessa

$$\hat{X}_t = E(X_t | \mathbf{S}_t) = E_{t-1}(\hat{X}_{t|t-1} | e_t)$$

ora abbiamo che

$$\Sigma_{12} = \text{Cov}_{t-1}(\hat{X}_{t|t-1}, e_t) = E_{t-1}(X_t - \hat{X}_{t|t-1})e_t' = P_{t|t-1}C'$$

e

$$\Sigma_{22} = V_{t-1}(e_t) = F_t$$

da cui si conclude che

$$\hat{X}_t = \hat{X}_{t|t-1} + P_{t|t-1}C'F_t^{-1}e_t$$

$$e P_t = P_{t/t-1} - P_{t/t-1}C'F_t^{-1}CP_{t/t-1}$$

Mettendo insieme le equazioni di cui sopra si trova il filtro per la forma previsiva

$$\hat{X}_{t+1|t} = (A - K_t C)\hat{X}_{t|t-1} + K_t(Y_t - DU_t) + BU_t$$

Dove $K_t = AP_{t/t-1}C'F_t^{-1}$ è il cosiddetto guadagno del filtro e l'espressione della covarianza dell'errore

$$P_{t+1|t} = A(P_{t|t-1} - P_{t|t-1}C'F_t^{-1}CP_{t|t-1})A' + RQR'$$

è nota come equazione di Riccati.

Osservazione: Se occorre solo la stima si ha invece

$$\hat{X}_t = A\hat{X}_{t-1} + BU_t + P_{t|t-1}C'F_t^{-1}e_t$$

Ottimalità:

in ipotesi di normalità X_t è la media condizionata. Se cadono queste ipotesi vale tuttavia ancora che X minimizza l'MSE fra tutti gli stimatori lineari.

Esercizio: Errori correlati

Se le due equazioni hanno errori correlati

$E(\eta_t \varepsilon_t') = G \neq 0$ mostrare che:

- le equazioni **predittive** che definiscono $X_{t|t-1}$ e $P_{t|t-1}$ restano inalterate mentre
- variano le equazioni di aggiornamento. In particolare
 - $P_{t|t-1}C'$ viene sostituito da $P_{t|t-1}C' + RG$
- mentre $F_t = CP_{t|t-1}C' + CRG + G'R'C' + H$

Osservazioni

- I risultati ottenuti sono condizionati ai valori iniziali x_0 e P_0
- Le matrici coinvolte sono spesso funzioni di un vettore θ di parametri
- $A(\theta), B(\theta), C(\theta), \dots$
 - in tal caso θ può essere stimato tramite MLE o PEM.
- L'estensione del filtro modelli tempo varianti A_t, \dots etc. è immediata.

Laboratorio: Matlab e R

ESERCIZIO COSTRUZIONE MODELLO ARMA

n=1000;

```

n1=n/2;
wn = randn(n,1);

% costruisco un oggetto AR(2)
m0=ar(wn,2);

% ci metto il polinomio con radici z1 e z2:
z1 = .85 + .28i;
z2 = .85 - .28i;
AA=conv([1,-z1],[1,-z2])
roots(AA)
m0.a=AA;

y=sim(m0,wn);

y1 = y(1:n1);
y2 = y(n1+1:end);
figure
corrf(y,100);
figure

PP=etfe(y,50);
plot(PP.FrequencyData,PP.SpectrumData(:))
%figure
%plot(PP)

m = ar(y,2);
present(m1)

V = arxstruc(y1,y2,[1:5]');
nn = selstruc(V,'mdl'); %best fit to validation data
%m1 = ar(y1,nn)
m1 = ar(y1,1)

figure
res1=resid(y1,m1);
figure
res2=resid(y2,m1);

```



```
figure
hist(res2)
figure
normplot(res2)
```

Esercizio di simulazione Montecarlo sugli ARMA

Si supponga che il Sistema generatore dei dati sia dato dalla seg. Equazione $y_t = \rho y_{t-1} + \epsilon_t$ con $t = 1; 2; \dots$ e $\epsilon_t \text{ NID}(0; 1)$.

Si consideri il problema di stima a minimi quadrati del modello AR(1) $y_t = \rho y_{t-1} + \epsilon_t$ sulla base di $n = 100$ osservazioni e si valuti, su base Montecarlo, la distribuzione di $\hat{\rho}_t$.

Confrontare il risultato di simulazione con quello previsto dalla teoria asintotica per processi stocastici stazionari commentando eventuali discrepanze.

Valutare se i risultati sono stabili con $n = 2$

Soluzione:

```
% dimensione della simulazione MC
m=1000;
% dimensione dei dati
n=100;
a_hat = zeros(m,1);
for j=1:m
% simulazione innovazione del sistema
wn=randn(n,1);
% simulazione "sistema generatore dei dati"
y = .lter(1, [1, -1], wn);
% stima LSE (formula diretta) (usa lag() di matlab_af_lib)
a_hat(j) = sum(y.*lag(y))/sum(y.^2);
% stima LSE usando identification toolbox
%m = ar(y,1,.ls.);
%a_hat(j) = -m.a(2);
end
mean(a_hat)
var(a_hat)
hist(a_hat)
```

Sintesi delle principali funzionalità di R

R: free, specifico per statistica, non un toolbox come matlab

Si utilizza fracdiff.sim, utilizzabile solo nell'ipotesi di dati normalmente distribuiti

Fracdiff.sim(n,ar,ma,d)

N numero simulazioni

Ar parte ar

Ma parte ma

D parametro per filtrare memoria lunga

Effettuato test di normalità dei dati con qqplot e hist e calcolo dei momenti, troviamo che la soluzione più efficace è l'utilizzo della radice quarta.

Effettuate 3 tipi di analisi :

- *fdGPH*: include il vero valore ma l'incertezza è maggiore, metodo poco efficace perché l'IC è troppo grande
Si fa una stima con parametri a memoria lunga d per ARFIMA (p,d,q) differenziando la serie storica per togliere la struttura long memory
- *Sperio*: l'IC è minore, ma non è detto che il valore vero sia incluso. Troviamo che in effetto ic più stretto, ma nelle simulazioni ci si avvicinava solo al valore reale, che era 0,3
- *Fracdiff*: fa una stima di massima verosimiglianza. Indicatore di residuo bianco: la diff. Ha tolto la memoria lunga. E' il migliore dei tre metodi perché usa MLE.

Carte di controllo: si controlla solo un campione perché si hanno troppi dati. Il controllo deve essere molto severo perché si tratta solo di un campione

Definiamo al 99,9° percentile il limite superiore e calcoliamo le bande di allarme.

Viene creato un vettore di disturbo fatto da una retta più del rumore da sommare ai dati e, una volta accodati i dati ok a quelli con disturbo, si calcola il TMR direttamente sui dati.

Con un ciclo viene simulato un gran numero di casi e si calcola la media del TMR.