

Identificazione dei Modelli e Analisi dei Dati

Prof. G. De Nicolao

II Prova scritta - 29 Giugno 2007

Cognome..... **Nome**.....

Matricola..... **Firma**.....

- Compilare a penna questo foglio all'inizio della prova.
- Durante lo svolgimento della prova, non è consentito l'uso di materiale diverso dai comuni strumenti di calcolo, scrittura e disegno.
- Le risposte devono essere scritte in modo chiaramente leggibile nello spazio immediatamente seguente ogni domanda (se necessario, a seguito di cancellature, passare sul retro).
- Le uniche risposte valide sono quelle riportate nel presente fascicolo, che va consegnato, senza fogli addizionali, al termine della prova.

1.
2.
3.
4.

1. Si consideri l'intervallo di confidenza della media campionaria dei dati X_i , $i = 1, \dots, n$, i.i.d., $X_i \sim N(m, \sigma^2)$. Si indichi con A l'ampiezza dell'intervallo (=limite superiore - limite inferiore). Si considerino i seguenti casi:

1. $A = 2, \quad \sigma^2 = 1$

2. $A = 1, \quad \sigma^2 = 4$

3. $A = 2, \quad \sigma^2 = 9$

4. $A = 1, \quad \sigma^2 = 16$

5. $A = 2, \quad \sigma^2 = 25$

6. $A = 1, \quad \sigma^2 = 36$

Scrivere, in corrispondenza dei seguenti valori di n , il numero del caso corretto.

$n = 62 \quad \dots\dots$

$n = 246 \quad \dots\dots$

$n = 554 \quad \dots\dots$

$n = 35 \quad \dots\dots$

$n = 97 \quad \dots\dots$

$n = 4 \quad \dots\dots$

2. Si considerino i seguenti dati

$$\begin{array}{cccc} x_1 = 0 & x_2 = \pi/2 & x_3 = \pi & x_4 = 3\pi/2 \\ y_1 = 2 & y_2 = 0.3 & y_3 = -3 & y_4 = -0.1 \end{array}$$

Per il modello

$$Y_k = \theta_1 \sin(x_k) + \theta_2 \cos(x_k) + V_k, \text{Var}[V] = \Psi$$

vi sono le seguenti scelte per la matrice Ψ :

- (a) $\Psi = 0.5I$
- (b) $\Psi = 2I$
- (c) $\Psi = \text{diag}\{2, 1, 2, 1\}$
- (d) $\Psi = \text{diag}\{1, 0.5, 1, 0.5\}$

Scrivere accanto alle matrici che forniscono $\text{Var}[\theta^{BLUE}]$ la lettera della corrispondente matrice Ψ .

$$\text{Var}[\theta^{BLUE}] = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad \dots$$

$$\text{Var}[\theta^{BLUE}] = \begin{bmatrix} 0.25 & 0 \\ 0 & 0.25 \end{bmatrix} \quad \dots$$

$$\text{Var}[\theta^{BLUE}] = \begin{bmatrix} 0.25 & 0 \\ 0 & 0.5 \end{bmatrix} \quad \dots$$

$$\text{Var}[\theta^{BLUE}] = \begin{bmatrix} 0.5 & 0 \\ 0 & 1 \end{bmatrix} \quad \dots$$

3. Dire se le seguenti affermazioni sono vere o false: (Punteggio: risposta esatta =1, errore=-1, non risponde =0)

V F

- (a) Se X e Y sono V.C. congiuntamente gaussiane, allora $E[XY] = E[X]E[Y]$.

□ □

- (b) Date delle V.C. i.i.d. $X_i, i = 1, \dots, n$, il momento campionario del terzo ordine M_3 è non polarizzato, consistente e asintoticamente gaussiano.

□ □

- (c) Si considerino n prove di Bernoulli con probabilità di successo pari a p . Allora, per $n \rightarrow \infty$, il numero di successi converge in probabilità a np .

□ □

- (d) Sia MSE l'errore quadratico medio associato ad uno stimatore polarizzato $\hat{\theta}$. Si consideri un secondo stimatore $\hat{\theta}^{new}$ non polarizzato e tale che $Var[\hat{\theta}^{new}] = Var[\hat{\theta}]$. Allora $MSE^{new} = MSE - (E[\hat{\theta}] - \theta^o)^2$.

□ □

- (e) Se $X_i, i = 1, \dots, n$, sono V.C. identicamente distribuite, ma non indipendenti, allora $Var[M_1] > \sigma^2/n$.

□ □

- (f) La non polarizzazione di uno stimatore è condizione necessaria per la consistenza.

□ □

- (g) Sotto l'ipotesi I2 con $Var[V] = \sigma^2 I$ e $\Phi = \Phi^V$, risulta $E[E[SSR^V]] = (n + q)\sigma^2$, dove $\Phi^V = \Phi$ è la matrice di sensitività per i dati di validazione e SSR^V è la somma dei quadrati dei residui in validazione (i due valori medi sono rispetto ai dati di identificazione e di validazione).

□ □

- (h) Sotto l'ipotesi I2 con $Var[V] = \sigma^2 I$, risulta $E[SSR] = (n - q)\sigma^2$.

□ □

- (i) Se $X_i, i = 1, \dots, n$, sono V.C. identicamente distribuite, ma non indipendenti, allora la media campionaria M_1 può essere polarizzata.

□ □

- (j) Sotto l'ipotesi I1, lo stimatore BLUE è gaussiano anche se σ^2 è sconosciuta.

□ □

4. Si consideri il modello

$$y_k = \theta_1 + \theta_2 x_k + v_k, k = 1, \dots, n$$

Si supponga che sia nota una stima $\hat{\theta}$ insieme alla sua matrice varianza

$$Var[\hat{\theta}] = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix}$$

In corrispondenza di un nuovo valore x^{new} si usa il predittore

$$\hat{y}^{new} = \hat{\theta}_1 + \hat{\theta}_2 x^{new}$$

Ricavare, riportando i passaggi, l'espressione di $Var[\hat{y}^{new}]$.

(Suggerimento: definendo opportunamente A è possibile scrivere $\hat{y}^{new} = A\hat{\theta}$)

5. Il seguente codice MatLab implementa il calcolo di FPE per uno stimatore BLUE. Il modello è:

$$Y_k = \theta_1 + \theta_2 x_k^2 + V_k$$

Si ipotizza $\Psi = I$. Le osservazioni e le variabili indipendenti sono contenute nei vettori colonna Y e x . Evidenziare e correggere eventuali errori.

```
n=length(Y);
q=2;
Phi=[ones(n,1); x.*x];
thetaBLUE = Phi \ Y;
epsilon = Y - Phi*thetaBLUE;
SSR = epsilon*epsilon';
FPE=SSR*(n+q)/(n-q);
```