Identificazione dei Modelli e Analisi dei Dati

Prof. G. De Nicolao

II prova in itinere - 25 Giugno 2010

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 considerino i dati X_i, $i = 1, \dots, n$ i.i.d., $X_i \sim N(m, \sigma^2)$.

   (a) Riportare la definizione di S_m^2 (varianza campionaria essendo noto il valor medio m).

   (b) Dimostrare, riportando i passaggi, che S_m^2 è uno stimatore non polarizzato di σ^2.

   (c) Dimostrare, riportando i passaggi che S_m^2 è proporzionale ad un χ_n^2 e ricavare il fattore di proporzionalità.

   Suggerimento: riscrivere S_m^2 in funzione delle V.C standardizzate $(X_i - m)/\sigma$

   (d) Sfruttando la risposta data al punto precedente, ricavare $Var[S_m^2]$.

2. A partire dai dati X_i, $i = 1, \ldots, N$ i.i.d., $X_i \sim N(m, \sigma^2)$, è stata calcolata la media campionaria. Si considerino i seguenti casi in cui σ^2 è nota:

   1. $N = 16, \quad \sigma^2 = 1$

   2. $N = 25, \quad \sigma^2 = 4$

   3. $N = 25, \quad \sigma^2 = 9$

   Si considerino inoltre altri tre casi in cui σ^2 è ignota ed è stata calcolata la varianza campionaria S^2:

   4. $N = 16, \quad S_c^2 = 1$

   5. $N = 25, \quad S_c^2 = 4$

   6. $N = 25, \quad S_c^2 = 9$

   Si indichi con $A := U - L$ l'ampiezza dell'intervallo di confidenza al 95%, dove L e U sono, rispettivamente, il limite inferiore e superiore, vale a dire $I_{0.95} = [L, U]$. Scrivere in corrispondenza delle ampiezze, il numero del caso corretto.

   $A = 3.3024$ ......

   $A = 3.9200$ ......

   $A = 2.1310$ ......

   $A = 3.1360$ ......

   $A = 4.1280$ ......

   $A = 1.9600$ ......

3. Si considerino i seguenti dati

   $x_1 = 0 \quad x_2 = \pi/2 \quad x_3 = \pi \quad x_3 = 3\pi/2$
   $y_1 = 2 \quad y_2 = 0.3 \quad y_3 = -3 \quad y_4 = -0.1$

   Per il modello

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

   vi sono le seguenti scelte per la matrice Ψ:

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

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

   $$Var[\theta^{BLUE}] = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \qquad \ldots$$

   $$Var[\theta^{BLUE}] = \begin{bmatrix} 0.25 & 0 \\ 0 & 0.25 \end{bmatrix} \qquad \ldots$$

   $$Var[\theta^{BLUE}] = \begin{bmatrix} 0.25 & 0 \\ 0 & 0.5 \end{bmatrix} \qquad \ldots$$

   $$Var[\theta^{BLUE}] = \begin{bmatrix} 0.5 & 0 \\ 0 & 1 \end{bmatrix} \qquad \ldots$$

4. Il seguente codice MATLAB calcola le medie, la matrice di covarianza e il coefficiente di correlazione per le variabili casuali W de H (peso e altezza), rispettivamente. I vettori colonna h e w contengono altezze e pesi di soggetti estratti a caso dalla popolazione. Trovare gli eventuali errori presenti nel codice e correggerli.

```
medie=mean([h; w])

EW=medie(1);

EH=medie(2);

cov_WH=cov([h; w]);

corr_WH=cov_WH(1,2)/(cov_WH(1,1)*cov_WH(2,2));
```

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

$V \quad F$

(a) Siano date le osservazioni i.i.d. X_i, $i = 1, \dots, n$, $X_i \sim N(m, \sigma^2)$. Allora, S_2 è gaussiano.

□ □

(b) I momenti campionari sono stimatori non polarizzati e consistenti.

□ □

(c) L'ampiezza dell'intervallo di confidenza per la media campionaria di osservazioni i.i.d. gaussiane è inversamente proporzionale al numero n di osservazioni.

□ □

(d) Relativamente alla stima della media di osservazioni i.i.d., l'errore quadratico medio commesso dalla media campionaria pari a σ^2/n.

□ □

(e) Se $X_i \sim N(0, 1)$, $i = 1, \dots, n$, sono V.C. indipendenti, allora M_2 è distribuito come un χ^2_{n-1}.

□ □

(f) Il momento centrale del secondo ordine S_2 è polarizzato, asintoticamente non polarizzato, consistente e asintoticamente gaussiano..

□ □

(g) Se $\Psi = \sigma^2 I$ allora, θ^{LS} coincide con θ^M .

□ □

(h) Se il numero n di dati è pari al numero q di parametri, la condizione di identificabilità è soddisfatta se e solo se $\det \Phi \neq 0$.

□ □

(i) Sotto l'Ipotesi I2 la stima BLUE θ^M è un vettore congiuntamente gaussiano.

□ □

(j) Sotto l'ipotesi I2, quando si identificano modelli gerarchici, risulta che $E[E[SSR^V]] = \sigma^2(2n + q)$ dove q indica il numero di parametri del modello.

□ □