Tema di Statistica Matematica

Sessione estiva

21 giugno 2019

1) Sia $(X_1, X_2, \ldots, X_n)$ un campione casuale proveniente da una distribuzione cui é associata la seguente funzione di densitá

$$f(x;\theta) = 2\,\theta\,x\,\exp\{-\theta x^2\}\,\mathbb{I}_{\mathbb{R}^+}(x), \quad \theta > 0$$

a) Verifcare se la famiglia di distribuzioni in questione é regolare e, laddove esista, individuare una statistica sufficiente per θ.

b) Trovare lo stimatore di massima verosimiglianza $\hat{\theta}_n$ di θ.

c) Determinare lo stimatore di massima verosimiglianza per $\gamma = 2/\theta$.

d) Calcolare le stime di massima verosimiglianza di θ e di γ ipotizzando di aver estratto dalla popolazione un campione casuale di ampiezza $n = 100$ cosí costituito

x_i	0.01	0.05	0.5	1	2	4	**Totale**
Frequenza	35	25	20	10	5	5	**100**

e) Lo stimatore di massima verosimiglianza di θ é anche stimatore efficiente per θ?

f) Trovare la distribuzione asintotica di $\hat{\theta}_n$ e costruire un intervallo di confidenza di livello 0.95 per θ.

2) Un individuo lavora a circa 50 Km di distanza dalla cittá in cui risiede e tutti i giorni si reca al lavoro in automobile compiendo lo stesso tragitto. Durante i primi 4 mesi dell'anno si é recato al lavoro per 91 volte e nell'arco di tale periodo ha osservato che lo scarto quadratico medio del consumo giornaliero di carburante é stato pari a 0.7 litri. Nell'arco dei successivi 3 mesi si é recato al lavoro per 61 volte osservando che, in questo periodo, lo scarto quadratico medio del consumo giornaliero di carburante e stato pari a 1.5 litri. Supponendo che il consumo giornaliero di carburante sia distribuito come una v.c. Normale,

a) costruire gli intervalli di confidenza di livello 0.95 per la varianza del consumo di carburante per ciascuno dei due periodi di tempo considerati;

b) fissato $\alpha = 0.05$, verificare l'ipotesi che la varianza del consumo di carburante sia uguale a 1.8 contro l'alternativa che sia diversa da tale valore per ciascuno dei due periodi considerati;

c) supponendo le misurazioni nei due periodi di tempo siano indipendenti e fissato in α il livello di significativitá del test, costruire la regola di decisione per la verifica dell'ipotesi di uguaglianza delle varianze nei due periodi considerati.

3) Sia $(X_1, X_2, \ldots, X_n, \ldots)$ una successione di variabili casuali i.i.d. aventi ciascuna distribuzione $U(0,1)$ e sia $M_n = \max(X_1, X_2, \ldots, X_n)$. Dimostrare che

$$n(1 - M_n) \xrightarrow{D} Exp(1)$$

4) Supponiamo di essere interessati a studiare la relazione causale che intercorre tra le variabili x e Y che supponiamo essere lineare; di conseguenza, possiamo scrivere il seguente modello che descrive Y in termini di x

$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i \tag{1}$$

per $i = 1, 2, \ldots, 9$. Sulla base dei seguenti riassunti

$$\sum_{i=1}^{9} x_i = 309, \ \sum_{i=1}^{9} x_i^2 = 11009, \ \sum_{i=1}^{9} Y_i = 465, \ \sum_{i=1}^{9} Y_i^2 = 16522, \ \sum_{i=1}^{9} x_i Y_i = 16552,$$

e sapendo che il coefficiente di determinazione $R^2 = 0.917$, dopo aver discusso le ipotesi classiche alla base del modello,

a) stimare a minimi quadrati i parametri β_0 e β_1 del modello di regressione lineare (1).

Assumendo l'errore ϵ seguire una distribuzione Normale $N(0, \sigma^2)$,

b) fissato $\alpha = 0.05$, verificare la significativitá del modello nel suo complesso;

c) fissato $\alpha = 0.05$, verificare la significativitá del coefficiente di regressione β_1;

d) costruire un intervallo di confidenza per il coefficiente di regressione β_1 a livello 0.99.

5 Sia $(X_1, X_2, \ldots, X_n, \ldots)$ un campione casuale proveniente da una distribuzione di Poisson $\mathcal{P}(\theta)$, $\theta > 0$. Trovare uno stimatore $UMVU$ di $e^{-3\theta}$.

EX. 1)

a) Si può facilmente provare che la famiglia di distribuzioni in questione è una FAM. ESPONENZIALE a $k=1$ parametri poiché

$$L(\theta;\underline{x}) = \prod_{i=1}^{n} 2\theta x_i \, e^{-\theta x_i^2} \mathbb{1}_{\mathbb{R}^+}(x_i), \quad \theta > 0$$

$$= \underbrace{\prod_{i=1}^{n}\left[\mathbb{1}_{\mathbb{R}^+}(x_i)\cdot x_i\right]}_{C(\underline{x})} \underbrace{(2\theta)^n}_{[D(\theta)]^n} \exp\{\underbrace{-\theta}_{A(\theta)} \underbrace{\sum_{i=1}^{n} x_i^2}_{\sum_{i=1}^{n} B(x_i)}\}$$

sicché è REGOLARE e la STATISTICA

$$T_n(X_1,\dots,X_n) = \sum_{i=1}^{n} X_i^2$$

è STATISTICA SUFFICIENTE (minimale) e COMPLETA per θ.

b) $\ell(\theta;\underline{x}) = \ln L(\theta;\underline{x})$

$$= \ln(C(\underline{x})) + n \ln(2\theta) - \theta \sum_{i=1}^{n} x_i^2$$

sicché la SOLUZIONE dell'EQUAZIONE ($\equiv$ SCORE EQUATION) in θ

$$\frac{d}{d\theta}\ell(\theta;x) = \frac{n}{\cancel{2}\theta}\cdot\cancel{2} - \sum_{i=1}^{n} x_i^2 = 0$$

restituisce lo STIMATORE di MASSIMA VEROSIMIGLIANZA di θ

$$\hat{\theta}_n = \frac{n}{\sum_{i=1}^{n} X_i^2}$$

tenuto conto del fatto che $\frac{d^2}{d\theta^2}\ell(\theta;x) = -\frac{n}{\theta^2} < 0$.

c) In virtù della PROPRIETÀ di INVARIANZA degli STIMATORI di MASSIMA VEROSIMIGLIANZA, considerato $\gamma = \frac{2}{\theta} \equiv \tau(\theta)$

$$\hat{\gamma}_n = \tau(\hat{\theta}_n) = \frac{2}{\hat{\theta}_n}$$

d) Si tratta ora di calcolare, sulla base dei dati in tabella, il valore della STATISTICA

$$\sum_{i=1}^{100} x_i^2 = 0{,}01^2 \cdot 35 + 0{,}05^2 \cdot 25 + 1^2 \cdot 10 + 2^2 \cdot 5 + 4^2 \cdot 5$$

$$= 115{,}066$$

sicché

$$\hat{\theta}_n = \frac{100}{\sum_{i=1}^{100} x_i^2} = \frac{100}{115{,}066} = 0.8691$$

e perciò

$$\hat{\gamma}_n = \frac{2}{\hat{\theta}_n} = \frac{2}{0.8691} = 2.3013$$

e) Si tratta di trovare la VARIANZA dello STIMATORE di MASSIMA VEROSIMIGLIANZA e confrontarla con il LIMITE INFERIORE della VARIANZA di uno STIMATORE NON DISTORTO fornito da RAO-CRAMÈR. Ciò a patto che $\hat{\theta}_n$ sia STIMATORE NON DISTORTO di θ.

Cominciamo con il verificare quest'ultimo aspetto.

Consideriamo la trasformazione

$$Y_i = X_i^2$$

sicché per il TEOREMA di TRASFORMAZIONE

$$f_{Y_i}(y_i;\theta) = f_{X_i}(\sqrt{y_i};\theta)\cdot\left|\frac{d}{dy_i}\sqrt{y_i}\right|$$

$$= 2\theta\, y_i^{1/2}\, e^{-\theta y_i}\cdot\left|\frac{1}{2y_i^{1/2}}\right| \mathbb{1}_{\mathbb{R}^+}(y_i)$$

$$= \theta\, e^{-\theta y_i}\, \mathbb{1}_{\mathbb{R}^+}(y_i)$$

ossia

$$Y_i \sim \mathcal{E}xp(\theta), \quad i=1,2,\dots,n$$

Ma allora, per la proprietà di RIPRODUCIBILITÀ

$$W = \sum_{i=1}^{n} X_i^2 = \sum_{i=1}^{n} Y_i \sim \mathcal{G}(n,\theta)$$

e dunque, essendo $\hat\theta_n = n / \sum_{i=1}^{n} X_i^2 = \frac{n}{W}$,

$$\mathbb{E}_\theta(\hat\theta_n) = \mathbb{E}_\theta\left(\frac{n}{W}\right) = \int_0^\infty \frac{n}{w}\,\frac{\theta^n}{\Gamma(n)}\, w^{n-1} e^{-\theta w}\, dw$$

$$= \frac{n}{n-1}\,\theta \int_0^\infty \frac{\theta^{n-1}}{\Gamma(n-1)}\, w^{n-1} e^{-\theta w}\, dw$$

$$= \frac{n}{n-1}\,\theta \neq \theta$$

essendo $\Gamma(n) = (n-1)\Gamma(n-1)$ e l'argomento nell'ultimo integrale, la funzione di densità di una Gamma di parametri $\alpha = n-1$ e $\beta = 1/\theta$ che quindi, INTEGRA a 1.

Pertanto lo STIMATORE di MV $\hat{\theta}_n$ è STIMATORE DISTORTO di θ e NON PUÒ ESSERE STIMATORE EFFICIENTE nel senso di BAHADUR. Fornisce comunque, opportunamente CORRETTO, lo STIMATORE UMVU di θ.

Infatti, partendo da $\hat{\theta}_n$ si può facilmente ottenere

$$U_n = \frac{n-1}{n}\,\hat{\theta}_n \quad \Rightarrow \quad E_\theta(U_n) = \theta, \ \forall \theta > 0$$

e in virtù del COROLLARIO del TEOREMA di RAO-BLACKWELL, U_n è STIMATORE UMVU essendo

- NON DISTORTO per θ
- FUNZIONE di STATISTICA SUFFICIENTE e COMPLETA $T_n = \sum_{i=1}^n X_i^2$

f) Per prima cosa calcoliamo l'INFORMAZIONE di FISHER,

$$I_n(\theta) = -E_\theta\left[\frac{d^2}{d\theta^2}\,\ell(\theta;\underline{x})\right]$$

$$= -E_\theta\left[-\frac{n}{\theta^2}\right] = \frac{n}{\theta^2}$$

e in base a NOTI TEOREMI sulle proprietà degli STIMATORI di MV possiamo affermare che

$$\hat{\theta}_n = \frac{n}{\sum_{i=1}^n X_i^2} \tilde{a} N\left(\theta, \frac{1}{I_n(\theta)}\right)$$

con $\dfrac{1}{I_n(\theta)} = \dfrac{\theta^2}{n}$.

Di conseguenza l' INTERVALLO di CONFIDENZA APPROSSIMATO di livello $(1-\alpha)$ per θ è dato da

$$IC_\theta(1-\alpha) = \left[\hat{\theta}_n - z_{1-\alpha/2} \sqrt{\frac{1}{I_n(\hat{\theta}_n)}} \, , \; \hat{\theta}_n + z_{1-\alpha/2} \sqrt{\frac{1}{I_n(\hat{\theta}_n)}} \right]$$

sicchè

$$IC_\theta(0.95) = [0.8541 \, , \, 0.8841]$$

essendo $\hat{\theta}_n = 0.8691$ e

$$z_{1-0.05/2} = z_{0.975} = 1.96$$

$$I_n(0.8691) = 132.402$$

Es. 2)

Cominciamo con riassumere i dati di cui siamo in possesso

PERIODO 1

$n_1 = 91$

$s_1 = 0.7$

$s_1^2 = 0.49$

PERIODO 2

$n_2 = 61$

$s_2 = 1.5$

$s_2^2 = 2.25$

a) L'intervallo di confidenza per la varianza di una popolazione (normale) è restituito da

$$IC_{\sigma^2}(1-\alpha) = \left[\frac{(n-1)s^2}{\chi^2_{n-1;\alpha/2}} , \frac{(n-1)s^2}{\chi^2_{n-1;1-\alpha/2}} \right]$$

Ora

$$\chi^2_{n_1-1;0.05/2} = \chi^2_{90;0.025} = 118.10$$

$$\chi^2_{n_1-1;1-0.05/2} = \chi^2_{90;0.975} = 65.65$$

sicché

$$IC_{\sigma_1^2}(0.95) = \left[\frac{(91-1)\cdot 0.7^2}{118.1} , \frac{(91-1)\, 0.7^2}{65.65} \right] = [0.373 , 0.672]$$

mentre

$$\chi^2_{n_2-1;0.05/2} = \chi^2_{60;0.025} = 83.30$$

$$\chi^2_{n_2-1;1-0.05/2} = \chi^2_{60;0.975} = 40.48$$

sicché

$$IC_{\sigma_2^2}(0.95) = \left[\frac{(61-1)\,1.5^2}{83,30}, \frac{(61-1)\,1.5^2}{40,48}\right] = [1.621,\ 3.334]$$

b) Si tratta di verificare

$$\begin{cases} H_0: & \sigma^2 = 1.8 \\ \text{vs} & \\ H_1: & \sigma^2 \neq 1.8 \end{cases}$$

sulla base della STATISTICA TEST data da

$$W = \frac{(n-1)\,S^2}{\sigma_0^2} \overset{H_0}{\sim} \chi^2_{n-1}$$

cui corrisponde la REGIONE CRITICA $C_\alpha = C'_\alpha \cup C''_\alpha$

dove

$$C'_\alpha = \{\underline{x} \in \mathcal{X} : W < \chi^2_{n-1;\alpha/2}\}$$

$$C''_\alpha = \{\underline{x} \in \mathcal{X} : W > \chi^2_{n-1;1-\alpha/2}\}$$

e che porta alla seguente REGOLA di DECISIONE:

"Si RIFIUTA H_0 a FAVORE di H_1 se $W \in C_\alpha$"

Ora, fissato $\alpha = 0.05$, per il PRIMO PERIODO,

$n_1 = 91$

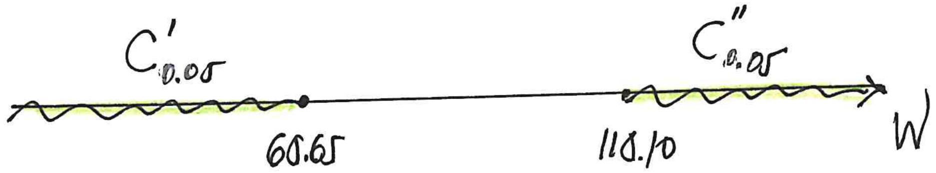
$\chi^2_{90;0.025} = 110.10$

$\chi^2_{90;0.975} = 65.65$

sicchè

$$C_{0.05} = \{ \underline{x} \in \mathcal{X} : W < 65.65 \} \cup \{ \underline{x} \in \mathcal{X} : W > 118.10 \}$$

ossia



e poiché

$$W = \frac{90 \cdot 0.49}{1.8} = 24.5 \in C''_{0.05}$$

si RIFIUTA H_0 a FAVORE di H_1.

Nel SECONDO PERIODO

$n_2 = 61$

$\chi^2_{60;\,0.025} = 83.30$

$\chi^2_{60;\,0.975} = 40.48$

sicchè

$$C_{0.05} = \{ \underline{x} \in \mathcal{X} : W < 40.48 \} \cup \{ \underline{x} \in \mathcal{X} : W > 83.30 \}$$

ossia

e poiché

$$W = \frac{60 \cdot 2.25}{1.8} = 75 \in \overline{C}_{0.05}$$

NON si RIFIUTA H_0.

c) Si tratta di verificare il seguente sistema di ipotesi:

$$\begin{cases} H_0: \sigma_1^2 = \sigma_2^2 \\ \text{vs.} \\ H_1: \sigma_1^2 \neq \sigma_2^2 \end{cases} \equiv \begin{cases} H_0: \sigma_1^2/\sigma_2^2 = 1 \\ \text{vs.} \\ H_1: \sigma_1^2/\sigma_2^2 \neq 1 \end{cases}$$

Ora, per NOTI TEOREMI (vd. CAMPIONAMENTO DA DISTRIBUZIONE NORMALE) sappiamo che le due STATISTICHE

$$\frac{(n_1-1)S_1^2}{\sigma_1^2} \sim \chi^2_{n_1-1} \quad \text{e} \quad \frac{(n_2-1)S_2^2}{\sigma_2^2} \sim \chi^2_{n_2-1}$$

hanno distribuzione CHI-QUADRATO e sono INDIPENDENTI, sicché

$$F = \frac{\frac{(n_1-1)S_1^2}{\sigma_1^2}\Big/(n_1-1)}{\frac{(n_2-1)S_2^2}{\sigma_2^2}\Big/(n_2-1)} \overset{H_0}{=} \frac{S_1^2}{S_2^2} \sim F_{n_1-1,\, n_2-1}$$

poiché sotto H_0 $\sigma_1^2 = \sigma_2^2 = \sigma^2$.

Allora, fissato $\alpha = 0.05$, si ha

$$F_{90,60;0.025} = 1.611 \quad [\text{valore più prossimo sulle tavole: } 1.58]$$

$$F_{90;60;0.975} = \frac{1}{F_{60,90;0.025}} = 0.635 \quad [\text{valore più prossimo sulle tavole: } 0.65$$

sicché la REGIONE CRITICA di livello $\alpha = 0.05$ è

$$C_{0.05} = C'_{0.05} \cup C''_{0.05} = \{\underline{x} \in \mathcal{X} : F < 0.635\} \cup \{\underline{x} \in \mathcal{X} : F > 1.611\}$$

che conduce alla REGOLA di DECISIONE di

"RIFIUTARE H_0 a FAVORE di H_1 se $F \in C_{0.05}$"

Nel nostro caso

$$F = \frac{s_1^2}{s_2^2} = \frac{0.49}{2.25} = 0.2178 \in C'_{0.05}$$

sicché si RIFIUTA l'IPOTESI NULLA $H_0: \sigma_1^2 = \sigma_2^2$, sulla base dell'INFORMAZIONE CAMPIONARIA disponibile.

Es. 3)

Sia $\{X_n\}_{n\in\mathbb{N}}$ succ.ne di V.C. INDIPENDENTI e ID. DISTRIBUITE secondo $U(0,1)$ sicché

$$f_{X_n}(x) = 1 \, \mathbb{1}_{[0,1]}(x)$$

$$F_{X_n}(x) = x$$

e $M_n = X_{(n)} = \max(X_1, X_2, \dots, X_n)$.

Cominciamo con il trovare la funzione di densità di $X_{(m)}$, m-ma STATISTICA ORDINATA, che sappiamo essere

$$f_{X_{(m)}}(x) = m \binom{n}{m} f(x) \left[F(x)\right]^{m-1} \left[1-F(x)\right]^{n-m}$$

sicché posto $m = n$

$$f_{X_{(n)}}(x) = n \binom{n}{n} 1 \cdot x^{n-1} (1-x)^{n-n}$$
$$= n\, x^{n-1}$$

Ora, ricorrendo al TEOREMA di TRASFORMAZIONE di V.C. (continue) posto

$$W = n(1-M_n) = n(1-X_{(n)})$$

da cui

$$X_{(n)} = \frac{n-W}{n} \quad , \quad \left|\frac{d}{dW} X_{(n)}\right| = \left|-\frac{1}{n}\right| = \frac{1}{n}$$

si ha

$$f_W(w) = n \left[\frac{n-w}{n}\right]^{n-1} \cdot \frac{1}{n} = \left[1 - \frac{w}{n}\right]^{n-1}$$

Ora

$$\lim_{n\to\infty} f_W(w) = \lim_{n\to\infty} \left[1 - \frac{w}{n}\right]^{n-1} = e^{-w}$$

che risulta essere la DENSITÀ di una V.C. $Exp(\theta = 1)$
sicché

$$W = n(1 - M_n) = n(1 - X_{(n)}) \xrightarrow{D} Exp(1)$$

IX. 4)

Omissis (argomento NON in PROGRAMMA per quest'anno)

Ex. 5)

Trovare uno stimatore UMVU per $e^{-3\theta}$ data una statistica suff e completa $X \sim \mathcal{P}(\theta)$ con $\theta > 0$.

a) $$f(x;\theta) = P_\theta(X=x) = \frac{e^{-\theta}\theta^x}{x!} \mathbb{1}_{\mathbb{N}^*}(x), \quad \theta > 0$$

b) $X = \sum_{i=1}^{n} X_i$ è stat. suff e completa per Poisson e

$$e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}$$

Ora abbiamo bisogno di trovare una funzione $g(\cdot)$ di STATISTICA SUFFICIENTE e COMPLETA tale che

$$E[g(X)] = e^{-3\theta}, \quad \forall \theta > 0$$

ossia $g(X) = g(X_1, \dots, X_n)$ è STIMATORE NON DISTORTO di $e^{-3\theta}$. Sicché

$$E(g(X)) = \sum_{k=0}^{\infty} g(k) \frac{\theta^k e^{-\theta}}{k!} = e^{-3\theta}$$

Moltiplicando ambo i membri per e^{θ} si ha

$$\sum_{k=0}^{\infty} g(k) \frac{\theta^k e^{-\theta}}{k!} e^{\theta} = e^{-3\theta} e^{\theta}$$

$$\sum_{k=0}^{\infty} g(k) \frac{\theta^k}{k!} = e^{-2\theta} \equiv \sum_{k=0}^{\infty} \frac{(-2)^k \theta^k}{k!}$$

Dal momento che le due precedenti somme sono uguali $\forall \theta > 0$, otteniamo

$$g(k) = (-2)^k$$

sicché

$$g(X) = (-2)^X = (-2)^{\sum_{i=1}^{n} X_i}$$

è lo stimatore UMVU di $e^{-3\theta}$ dove $X = \sum_{i=1}^{n} X_i$