Tema di Statistica Matematica

Secondo appello

13 febbraio 2019

1) Sia $(X_1, X_2, \ldots, X_n)$ un campione casuale proveniente da una distribuzione la cui densitá sia data da

$$f_X(x;\theta) = \frac{\ln(\theta)}{\theta - 1}\, \theta^x \, \mathbb{I}_{(0,1)}(x), \;\; \theta > 1. \tag{1}$$

a) Verificare se la famiglia di distribuzioni associata alla densitá (1) é regolare.

b) Trovare una statistica sufficiente minimale per l'inferenza su θ.

c) Trovare uno stimatore UMVU per

$$\tau(\theta) = \frac{\theta}{\theta - 1} - \frac{1}{\ln(\theta)}.$$

2) Sia $(X_1, X_2, \ldots, X_n)$ un campione casuale di ampiezza n da una distribuzione continua avente funzione di densitá

$$f(x;\theta) = \frac{\Gamma(2\theta)}{[\Gamma(\theta)]^2}\, x^{\theta-1}\,(1-x)^{\theta-1}\,\mathbb{I}_{(0,1)}(x)$$

con $\theta > 0$ parametro non noto.

a) Individuare una statistica sufficiente minimale per l'inferenza su θ.

b) Con $n > 1$, si supponga di voler risolvere verificare il seguente sistema di ipotesi

$$H_0 : \theta = 1 \;\; vs. \;\; H_1 : \theta > 1$$

utilizzando il test basato sulla statistica

$$T = \sum_{i=1}^{n} \mathbb{I}_{\left(\frac{1}{2}-c,\, \frac{1}{2}+c\right)}(x_i)$$

con $c \in [0.1, 0.2]$ costante fissata che rifiuta H_0 se $t > k$, essendo k un valore soglia da fissare opportunamente.

i) Con $n = 50$ e $c = 0.15$, individuare k in modo che il test abbia livello di significativitá approssimato pari a 0.05.

ii) Fornire un valore approssimato per la potenza del test di cui in i) in corrispondenza dell'alternativa $\theta = 2$.

3) Sia $(X_1, X_2, \ldots, X_n)$ un campione casuale proveniente da una distribuzione Normale di media μ e varianza $\theta^2\mu^2$, con $\mu > 0$ parametro ignoto e $\theta > 0$ costante nota.

a) Trovare la statistica sufficiente minimale per l'inferenza su μ. In questo particolare caso, che cosa posssiamo dedurre dalla forma della statistica sufficiente circa il modello che ha generato i dati campionari?

b) Costruire un intervallo di confidenza esatto per μ di livello $(1-\alpha) = 0.9$.

c) Dimostrare che la statistica sufficiente minimale di cui al punto a) non e completa.

d) Ricavare lo stimatore di massima verosimiglianza $\hat{\mu}_n$ per μ.

4) Sia $(X_1, X_2, \ldots, X_n)$ un campione casuale proveniente da una distribuzione Normale di media μ_0 nota e varianza σ^2 ignota,

a) Stabilire se lo stimatore

$$T_n = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \mu_0^2$$

é stimatore uniformemente a minima varianza nella classe degli stimatori non distorti di σ^2.

b) Posto $\mu_0 = 0$, costruire un test esatto (basato su T_n) di livello $\alpha = 0.05$ per la verifica del sistema di ipotesi $H_0 : \sigma^2 = \sigma_0^2$ vs. $H_1 : \sigma^2 > \sigma_0^2$.

c) Ricavare la funzione di potenza del test di cui al punto precedente.

5) Sia $\{X_n\}_{n=1}^{\infty}$ una successione di variabili casuali indipendenti uniformi sull'intervallo $(0, 3)$ e sia $Z_n = \min(X_1, X_2, \ldots, X_n)$. Stabilire se e a che cosa Z_n converge

a) in distribuzione;

b) in probabilitá.

Ex. 1)

a) Possiamo subito verificare che la famiglia di distribuzioni avente densità

$$f_X(x;\theta) = \frac{\ln(\theta)}{\theta-1}\,\theta^x\,\mathbb{1}_{(0,1)}(x), \quad \theta>1$$

costituisce una famiglia esponenziale a $k=1$ parametri. Infatti

$$f_X(x;\theta) = \underbrace{\mathbb{1}_{(0,1)}(x)}_{C(x)} \cdot \underbrace{\frac{\ln(\theta)}{\theta-1}}_{D(\theta)} \cdot \exp\{\underbrace{\ln(\theta)}_{A(\theta)}\cdot\underbrace{x}_{B(x)}\}$$

e dunque costituisce una fam. di distribuzioni regolare (soddisfa alle usuali proprietà di regolarità)

b) Per un teorema su famiglie esponenziali e statistiche sufficienti (minimali) si sa che

una serie di teoremi

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

è statistica sufficiente e completa per θ. Nel nostro caso, $B(X_i)=X_i$ sicché la statistica sufficiente (minimale) per θ è $\sum_{i=1}^{n} X_i$.

c) Consideriamo la funzione di verosimiglianza

$$L(\theta;\underline{x}) = \prod_{i=1}^{n} \frac{\ln(\theta)}{\theta-1}\,\theta^{x_i}\,\mathbb{1}_{(0,1)}(x_i)$$

$$= \left[\frac{\ln(\theta)}{\theta-1}\right]^n \theta^{\sum_{i=1}^{n} x_i} \prod_{i=1}^{n}\mathbb{1}_{(0,1)}(x_i)$$

sicchè, lasciando perdere la parte che coinvolge l'indicatrice,

$$l(\theta; \underline{x}) = \frac{d}{d\theta} L(\theta; \underline{x}) = n\left[\ln(\ln(\theta)) - \ln(\theta-1)\right] + \ln(\theta) \sum_{i=1}^{n} x_i$$

e dunque

$$S(\theta; \underline{x}) = \frac{d}{d\theta} l(\theta; \underline{x}) = \frac{n}{\ln(\theta)} \cdot \frac{1}{\theta} - \frac{n}{\theta-1} + \frac{1}{\theta} \sum_{i=1}^{n} x_i$$

$$= n\left[\frac{1}{\theta \ln(\theta)} - \frac{1}{\theta-1} + \frac{\overline{X}_n}{\theta}\right]$$

$$= \frac{n}{\theta}\left[\frac{1}{\ln(\theta)} - \frac{\theta}{\theta-1} + \overline{X}_n\right]$$

$$= \frac{n}{\theta}\left[\overline{X}_n - \left(\frac{\theta}{\theta-1} - \frac{1}{\ln(\theta)}\right)\right]$$

sicchè, essendo la famiglia regolare

$$\mathbb{E}_\theta(S(\theta; \underline{X})) = \mathbb{E}_\theta\left[\frac{n}{\theta}\left(\overline{X}_n - \left(\frac{\theta}{\theta-1} - \frac{1}{\ln(\theta)}\right)\right)\right] = 0$$

$$\cancel{\frac{n}{\theta}}\, \mathbb{E}_\theta(\overline{X}_n) - \cancel{\frac{n}{\theta}}\left(\frac{\theta}{\theta-1} - \frac{1}{\ln(\theta)}\right) = 0$$

$$\mathbb{E}_\theta(\overline{X}_n) = \frac{\theta}{\theta-1} - \frac{1}{\ln(\theta)} \qquad \forall \theta > 0$$

ovvero $\overline{X}_n$ è STIMATORE NON DISTORTO di

$$\tau(\theta) = \frac{\theta}{\theta-1} - \frac{1}{\ln(\theta)}$$

Essendo inoltre FUNZIONE di STATISTICA SUFFICIENTE (E COMPLETA) $\sum_{i=1}^{n} X_i$, $\overline{X}_n$ è STIMATORE UMVU per θ, per il

COROLLARIO DEL TEOREMA di RAO-BLACKWELL.

Es. 2)

Si ha

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

$$= \left[\frac{\Gamma(2\theta)}{\Gamma^2(\theta)}\right]^n \prod_{i=1}^{n} [x_i(1-x_i)]^{\theta-1} \prod_{i=1}^{n} \mathbb{1}_{(0,1)}(x_i)$$

Quindi, presi due punti $\underline{x}_1$ e $\underline{x}_2$ dello spazio campionario, il rapporto tra le loro verosimiglianze

$$\frac{L(\theta;\underline{x}_1)}{L(\theta;\underline{x}_2)} = \frac{\prod_{i=1}^{n}[x_{1i}(1-x_{1i})]^{\theta-1}}{\prod_{i=1}^{n}[x_{2i}(1-x_{2i})]^{\theta-1}} = \left[\frac{\prod_{i=1}^{n}[x_{1i}(1-x_{1i})]}{\prod_{i=1}^{n}[x_{2i}(1-x_{2i})]}\right]^{\theta-1}$$

non dipende da θ se e solo se $\underline{x}_1$ e $\underline{x}_2$ sono tali che

$$\prod_{i=1}^{n} x_{1i}(1-x_{1i}) = \prod_{i=1}^{n} x_{2i}(1-x_{2i})$$

Ne segue che la statistica sufficiente minimale per l'inferenza su θ è

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

o, equivalentemente,

$$V_n(X_1,\dots,X_n) = \ln(T_n(X_1,\dots,X_n))$$

$$= \sum_{i=1}^{n} \ln(X_i) + \sum_{i=1}^{n} \ln(1-X_i)$$

b) $n > 1$

i) Sotto H_0 la variabile casuale

$$\mathbb{1}_{(\frac{1}{2}-c,\, \frac{1}{2}+c)}(X)$$

ha distribuzione di BERNOULLI di PARAMETRO $2c$ e di conseguenza, sempre sotto H_0,

$$T = \sum_{i=1}^{n} \mathbb{1}_{(\frac{1}{2}-c,\, \frac{1}{2}+c)}(X_i) \sim b(n, 2c)$$

Usando l'APPROSSIMAZIONE NORMALE alla BINOMIALE si conclude che, sotto H_0,

$$T \underset{a}{\sim} N(2nc,\, 2nc(1-2c))$$

Con i valori forniti, $n=50$ e $c = 0.15$

$$T \underset{a}{\sim} N(15,\, 10.5)$$

Quindi, poichè

$$P\left(\frac{T-15}{\sqrt{10.5}} > 1.645 \mid H_0\right) = P(T > 1.645 \cdot \sqrt{10.5} + 15) = 0.05$$

risulta

$$k = 1.645\sqrt{10.5} + 15 = 20.3.$$

ii) Sotto alternativa $\theta = 2$, la funzione di densità di X è

$$f(x;2) = 6x - 6x^2$$

Quindi

$$P\left(\tfrac{1}{2} - c < X < \tfrac{1}{2} + c \mid \theta = 2\right) = \int_{\frac{1}{2}-c}^{\frac{1}{2}+c} (6u - 6u^2)\, du = 0.4365$$

Pertanto quando $\theta = 2$,

$$T \underset{\sim}{\sim} N(0{,}4365\, n,\ 0{,}4365\,(1-0{,}4365)\cdot n)$$

cioè

$$T \underset{\sim}{\sim} N(21.8\ ,\ 12{,}3)$$

e

$$P(T > 20.3 \mid \theta = 2) = P\left(\frac{T - 21.8}{\sqrt{12.3}} > \frac{20.3 - 21.8}{\sqrt{12.3}} \right)$$

$$= P(Z > -0{,}4277) = 0.665$$

è il valore approssimato della POTENZA del test in questione.

Es. 3)

a) Poiché

$$f(x;\mu) = \frac{1}{\sqrt{2\pi}\,\theta\mu} \exp\left\{-\frac{1}{2\theta^2\mu^2}(x-\mu)^2\right\} \mathbb{1}_{\mathbb{R}}(x)$$

$$= \mathbb{1}_{\mathbb{R}}(x) \quad \frac{e^{-\frac{1}{2\theta^2}}}{\sqrt{2\pi}\,\theta\mu} \cdot \exp\left\{\frac{1}{\theta^2\mu}\cdot x - \frac{1}{2\theta^2\mu^2}x^2\right\}$$

il modello parametrico considerato costituisce una FAMIGLIA ESPONENZIALE NON REGOLARE (poiché l'ORDINE della famiglia è DUE, maggiore della DIMENSIONE del PARAMETRO che è UNO).

Sotto campionamento casuale la STATISTICA NATURALE

$$T = \left(\sum_{i=1}^n X_i,\ \sum_{i=1}^n X_i^2\right)$$

è STATISTICA (congiuntamente) SUFFICIENTE MINIMALE per il parametro μ.

b) Poiché

$$\frac{X}{\mu} \sim N(1, \theta^2)$$

si ha immediatamente che

$$\frac{\bar{X}_n}{\mu} \sim N\left(1, \frac{\theta^2}{n}\right)$$

sicché $\bar{X}_n$ è STATISTICA PIVOT per l'INFERENZA su μ per il PROBLEMA in questione (NB: θ è QUANTITÀ NOTA)

Di conseguenza

$$\frac{\frac{\overline{X}_n}{\mu} - 1}{\theta/\sqrt{n}} \sim N(0,1)$$

sicché l'INTERVALLO di CONFIDENZA ESATTO per μ di LIVELLO 0.9 si ottiene INVERTENDO la RELAZIONE

$$-1.645 \leq \frac{\bar{x}_n/\mu - 1}{\theta/\sqrt{n}} < 1.645$$

con $1.645 = z_{0.95}$.

Pertanto

$$IC_\mu(0.9) = \left[\frac{\bar{x}_n}{1 + 1.645\, \theta/\sqrt{n}} , \frac{\bar{x}_n}{1 - 1.645\, \theta/\sqrt{n}} \right]$$

c) Dato che $\frac{X}{\mu}$ ha distribuzione che NON DIPENDE da μ, lo STESSO accade per

$$\sum_{i=1}^{n} \frac{X_i}{\mu} \quad e \quad \sum_{i=1}^{n} \frac{X_i^2}{\mu}$$

Quindi anche la FUNZIONE

$$g(T) = \frac{\sum_{i=1}^{n} X_i/\mu}{\sum_{i=1}^{n} X_i^2/\mu}$$

ha distribuzione che NON DIPENDE da μ.

Posto $\gamma = E(g(T))$, γ NON DIPENDE da μ e $E[g(T) - \gamma] = 0$ $\forall \mu > 0$.

Abbiamo quindi individuato una FUNZIONE della STATISTICA T, NON IDENTICAMENTE NULLA, la cui MEDIA è ZERO per OGNI valore del PARAMETRO μ. Di conseguenza, per la DEFINIZIONE data di COMPLETEZZA, la STATISTICA T NON è COMPLETA.

d) La funzione di logverosimiglianza vale

$$l(\mu;\underline{x}) = \frac{\sum_{i=1}^{n} x_i}{\theta^2\mu} - \frac{\sum_{i=1}^{n} x_i^2}{2\theta^2\mu^2} - n\ln(\mu)$$

e quindi

$$\frac{d}{d\mu} l(\mu;\underline{x}) = \frac{1}{2\theta^2\mu^3}\left(-2\mu\sum_{i=1}^{n} x_i + 2\sum_{i=1}^{n} x_i^2 - 2n\theta^2\mu^2\right)$$

Si ha $\frac{d}{d\mu} l(\mu;\underline{x}) = 0$ se

$$h(\mu) = -2\mu\sum_{i=1}^{n} x_i + 2\sum_{i=1}^{n} x_i^2 - 2n\theta^2\mu^2 = 0$$

Tale equazione di secondo grado in μ ha due radici delle quali solo UNA, la più grande, è POSITIVA. Infatti $h(\mu)$ è una PARABOLA con vertice in alto e $h(0) = 2\sum_{i=1}^{n} x_i^2 > 0$.

Pertanto

$$\hat{\mu}_n = \frac{-\sum_{i=1}^{n} x_i + \sqrt{\left(\sum_{i=1}^{n} x_i\right)^2 + 4n\theta^2\sum_{i=1}^{n} x_i^2}}{2n\theta^2}$$

EX. 4)

a) Il modello parametrico considerato costituisce una FAMIGLIA ESPONENZIALE a $K=1$ parametri con STATISTICA NATURALE pari a

$$W_n = \sum_{i=1}^{n} B(X_j) = \sum_{i=1}^{n} (X_j - \mu_0)^2$$

Tale STATISTICA è SUFFICIENTE (minimale) e COMPLETA per σ^2.

Ora $T_n = \frac{1}{n} \sum_{i=1}^{n} X_j^2 - \mu_0^2$ è STIMATORE NON DISTORTO per σ^2 poiché

$$\mathbb{E}_{\sigma^2}[T_n] = \mathbb{E}_{\sigma^2}\left[\frac{1}{n} \sum_{i=1}^{n} X_i^2 - \mu_0^2\right]$$

$$= \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}_{\sigma^2}(X_j^2) - \mu_0^2$$

$$= \frac{1}{n} \sum_{i=1}^{n} (\sigma^2 + \mu_0^2) - \mu_0^2 = \sigma^2 + \mu_0^2 - \mu_0^2 = \sigma^2$$

ma NON è FUNZIONE della STATISTICA SUFFICIENTE e COMPLETA W_n. Quindi lo stimatore NON DISTORTO T_n di σ^2 NON è STIMATORE UMVU.

b) Essendo T_n STIMATORE CONSISTENTE per σ^2, una regola di decisione sensata per risolvere il problema di verifica di ipotesi in questione potrebbe essere quella di

"RIFIUTARE H_0 per GRANDI VALORI di T_n"

Essendo $\mu_0 = 0$, si ha che, sotto H_0,

$$\frac{X}{\sigma_0} \sim N(0,1)$$

Quindi, sotto H_0,

$$\frac{X^2}{\sigma_0^2} \sim \chi_1^2 \overset{\text{PROPR. RIPRODUC.TÀ}}{\Longrightarrow} \frac{1}{\sigma_0^2}\sum_{i=1}^{n} X_i^2 \sim \chi_n^2$$

Ne deriva che

$$\frac{n T_n}{\sigma_0^2} \overset{H_0}{\sim} \chi_n^2$$

e pertanto il TEST ESATTO di livello $\alpha = 0.05$ basato su T_n RIFIUTA l'IPOTESI NULLA H_0 se

$$\frac{n T_n}{\sigma_0^2} > c$$

con c tale che $P\left(\frac{n T_n}{\sigma_0^2} > c\right) = 0.05$.

c) La FUNZIONE di POTENZA associata al test al punto precedente è

$$\eta_{C_\alpha}(\sigma^2) = P\left(\frac{n T_n}{\sigma_0^2} > c \,\middle|\, \sigma^2 > \sigma_0^2\right)$$

$$= P\left(\frac{\sigma^2}{\sigma^2}\frac{n T_n}{\sigma_0^2} > c\right) = P\left(\frac{n T_n}{\sigma^2} > c\,\frac{\sigma_0^2}{\sigma^2}\right)$$

$$= P\left(\chi^2 > c\,\frac{\sigma_0^2}{\sigma^2}\right)$$

Pertanto $\eta_{C_\alpha}(\sigma^2)$ è una FUNZIONE NON DECRESCENTE.

IX.5)

a) $(X_1, X_2, \dots, X_n)$ da $U(0,3)$ sicchè

$$f_X(x;3) = \frac{1}{3} \mathbb{1}_{[0,3]}(x)$$

e

$$F_X(x;3) = \frac{x}{3}$$

Essendo $Z_n = \min(X_1, \dots, X_n) = X_{(1)}$, la DENSITÀ di $X_{(1)}$ è data da

$$f_{X_{(1)}}(x) = 1 \cdot \binom{n}{1} \frac{1}{3} \left(\frac{x}{3}\right)^{1-1} \left(1-\frac{x}{3}\right)^{n-1} = \frac{n}{3}\left(1-\frac{x}{3}\right)^{n-1}$$

ricordando che la densità della m-ma STATISTICA ORDINATA è

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

per $m = 1, 2, \dots$

Inoltre

$$F_{X_{(1)}}(x) = -\left(1-\frac{u}{3}\right)\Big|_0^x = 1-\left(1-\frac{x}{3}\right)^n$$

Dunque

$$F_{X_{(1)}}(x) = \begin{cases} 0 & x<0 \\ 1-\left(1-\frac{x}{3}\right)^n & 0 \le x < 1 \\ 1 & x \ge 1 \end{cases}$$

e di conseguenza

$$\lim_{n\to\infty} F_{X_{(1)}}(x) = \begin{cases} 0 & x<0 \\ 1 & x \ge 0 \end{cases}$$

ossia

$$Z_n = X_{(1)} \xrightarrow{D} 0.$$

b) Convergendo in distribuzione a una COSTANTE, si avrà anche che

$$Z_n = X_{(1)} \xrightarrow{P} 0.$$