ST1- ESAME: Giugno 7 -2010 *(Orlandi)*

Esercizio 1 (10 punti) Siano $(X_1, \ldots, X_n)$ indipendenti estratti da una popolazione $N(\mu, \sigma^2)$ con μ e σ^2 incogniti. Si vuole verificare $H_0 : \sigma^2 = 1$ contro l'ipotesi $H_1 : \sigma^2 > 1$. Si calcoli la zona di rifiuto R come funzione di $s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$ per un test di ampiezza α usando il rapporto di verosimiglianza generalizzato.

Si dica inoltre cosa si intende per test uniformemente piú potente di ampiezza α. Il test trovato é un test uniformemente piú potente? Motivare.

Soluzione Chiamiamo $\theta = (\mu, \sigma^2)$. Dobbiamo determinare un test per

$$H_0 : \theta \in \Theta_0 := R \times \{1\} \quad \text{in alternativa a} \quad H_1 : \theta \in \Theta := R \times (0, \infty).$$

Il rapporto di verosimiglianza generalizzato si definisce come

$$\lambda(x_1, \ldots, x_n) := \frac{\sup_{\theta \in \Theta_0} L(\theta)}{\sup_{\theta \in \Theta} L(\theta)},$$

dove

$$L(\theta) := \prod_{i=1}^n f(x_i, \theta), \quad \text{and} \quad f(x, \theta) := \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

Le stime di massima verosimiglianza sono

$$\hat{\mu} := \bar{x} := \frac{1}{n} \sum_{i=1}^n x_i, \qquad s^2 := \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2.$$

Inoltre

$$\frac{\partial \log L}{\partial \sigma^2} = \frac{n}{2\sigma^2}(-1 + \frac{s^2}{\sigma^2}).$$

Quindi L é crescente per $s^2 > \sigma^2$, decresce se $s^2 < \sigma^2$ ed ha un massimo per $\sigma^2 = s^2$. quindi

$$\lambda(x_1, \ldots, x_n) = \begin{cases} \dfrac{L(\bar{x}, 1)}{L(\bar{x}, 1)} & \text{quando} \quad s^2 \leq 1, \\ \dfrac{L(\bar{x}, 1)}{L(\bar{x}, s^2)} & \text{quando} \quad s^2 > 1. \end{cases}$$

Quando $s^2 > 1$, $\lambda(x_1, \ldots, x_n) < 1$

$$\lambda(x_1, \ldots, x_n) = (s^2 e^{-s^2+1})^{\frac{n}{2}}.$$

Sia $0 < \lambda_0 < 1$, allora un test di rapporto di verosimiglianza generalizzato é dato da: si rifiuti H_0 se se $s^2 > 1$ e $\lambda(x_1, \ldots, x_n) \leq \lambda_0$. E' facile verificare che $\lambda(x_1, \ldots, x_n)$ é una funzione decrescente di s^2 quindi esisterá un k il cui valore dipende da λ_0 tale che la regione R di rifiuto é data da

$$R = \{(x_1, \ldots, x_n) : s^2 > k\} = \{(x_1, \ldots, x_n) : n\frac{s^2}{\sigma^2} > k'\}.$$

Per derminare k' si ricorda che

$$\frac{nS^2}{\sigma^2} = \sum_{i=1}^n \frac{(X_i - \bar{X})^2}{\sigma^2} \simeq \chi^2_{n-1}$$

che sotto H_0 diventa

$$nS^2 \simeq \chi^2_{n-1}.$$

Il valore di k' é quindi dato dalla condizione

$$\alpha = P[n\frac{S^2}{\sigma^2} > k'|\sigma^2 = 1] = P[nS^2 > k'] = P[nS^2 > \chi^2_{n-1,1-\alpha}].$$

Allora

$$k' = \chi^2_{n-1,1-\alpha}.$$

Per la definizione di test uniformemente piú potente si rimanda a pagina 423, del libro di testo Mood-Graybill-Boes.

Il test trovato é uniformente piú potente poiché α non dipende da $\theta = (\mu, \sigma^2)$.

Esercizio 2 (8 punti) Siano $(X_1, \ldots, X_n)$ indipendenti estratti da una popolazione con distribuzione esponenziale $f(x) = \lambda e^{-\lambda x}$ per $x \geq 0$. Si vuole stimare λ. Si stimi λ con lo stimatore

$$\hat{\lambda}_n := \frac{1}{\bar{X}}$$

con $\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i$.

Si definisca cosa si intende per stimatore non distorto.

Si verifichi se $\hat{\lambda}_n$ é uno stimatore distorto o non distorto di λ.

Qualora $\hat{\lambda}_n$ sia uno stimatore distorto si proponga uno stimatore non distorto. Valutare inoltre la distorsione di $\hat{\lambda}_n$ al variare di n.

Soluzione Per verificare se $\hat{\lambda}_n$ é distorto dobbiamo calcolare la sua media. Lo stimatore é

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

Si ricorda che $\sum_{i=1}^n X_i \simeq Gamma(n, \lambda)$. La funzione generatrice dei momenti per una variabile esponenziale é data da $\frac{1}{1-\lambda t}$. Quindi la funzione generatrice dei momenti di $Y := \sum_{i=1}^n X_i$ é data da

$$g_Y(t) = \left(\frac{1}{1-\lambda t}\right)^n,$$

che é appunto la funzione generatrice dei momenti di una $Gamma(n, \lambda)$. Si ricorda che la distribuzione di una $Gamma(n, \lambda)$ é data da

$$f_Y(s) = \frac{\lambda^n s^{n-1} e^{-\lambda s}}{(n-1)!}.$$

Quindi

$$E[\hat{\lambda}_n] = E[\frac{n}{Y}] = n\int_0^\infty ds \frac{1}{s}\frac{\lambda^n s^{n-1} e^{-\lambda s}}{(n-1)!} = \frac{n\lambda^n}{(n-1)!}\int_0^\infty ds s^{n-2} e^{-\lambda s}.$$

Nell'ultimo integrale si riconosce il nucleo di una $Gamma(n-1, \lambda)$. Si ottiene

$$E[\hat{\lambda}_n] = \frac{n\lambda^n}{(n-1)!}\frac{(n-2)!}{\lambda^{n-1}} = \lambda\frac{n}{n-1}.$$

Quindi $\hat{\lambda}_n$ é uno stimatore distorto di λ. Poiché n é noto si ottiene uno stimatore non distorto ponendo

$$\bar{\lambda}_n := \frac{(n-1)}{\sum_{i=1}^n X_i}.$$

Si puó vedere facilmente che

$$\lim_{n\to\infty} E[\hat{\lambda}_n] - \lambda = \lim_{n\to\infty}\frac{\lambda}{(n-1)} = 0.$$

Lo stimatore é dunque asintoticamente non distorto.
Esercizio 3 (8 punti) Siano X e Y due variabili aleatorie con densita di probabilitá congiunta

$$f(x,y) = 2, \qquad 0 \leq x \leq y \leq 1.$$

Determinare

(1)

$$P[0 \leq X \leq \frac{1}{2}; 0 \leq Y \leq \frac{1}{2}].$$

(2) La densitá di probabilitá di X.
(3) La densitá di probabilitá di Y.
(4) La media condizionata $E[Y|x]$, con $0 \leq x \leq 1$.

Soluzione

$$P[0 \leq X \leq \frac{1}{2}; 0 \leq Y \leq \frac{1}{2}] = P[0 \leq X \leq Y; 0 \leq Y \leq \frac{1}{2}] = \int_0^{\frac{1}{2}} dx \int_0^y 2dy = \frac{1}{4}$$

$$f_X(x) = \int_x^1 2dy = 2(1-x), \qquad 0 \leq x \leq 1$$

$$f_Y(y) = \int_0^{\frac{1}{2}} dx = 2y \qquad 0 \leq y \leq 1$$

Si calcola facilmente che

$$f_Y(y|x) = \frac{f(x,y)}{f_X(x)} = \frac{1}{1-x}, 0 \leq x \leq y \leq 1.$$

Quindi

$$E[Y|x] = \int_x^1 y \frac{1}{1-x} dy = \frac{(1+x)}{2}.$$

Esercizio 4 (6 punti) Una bilancia pesa con un margine di errore. Si assuma che l'errore E sia una variabile aleatoria con distribuzione normale di media 0 e varianza 10^{-2} (in milligrammi).

Si assuma che i risultati di 6 pesate successive dello stesso oggetto abbiano dato valori in milligrammi di $4,08 - -3.32 - -4,10 - -4,68 - - -3.15 - -3.14$. Determina un intervallo di confidenza per il peso dell'oggetto ad un livello di confidenza del 95%.

Soluzione

Siano X_i, $i = 1, \ldots 6$ le variabile aleatorie che rappresentano i sei pesi. Per ipotesi le X_i $i = 1, \ldots 6$ sono variabili indipendenti con distribuzione Gaussiana di media μ incognita e varianza $\sigma^2 = 10^{-2}$.

Per determinare un intervallo di confidenza per μ si ponga $\bar{X} = \frac{1}{6}\sum_{i=1}^6 X_i$. Si ottiene che $\frac{\bar{X}-\mu}{\sqrt{\frac{\sigma^2}{6}}} = Z$ é una variabile aleatoria con distribuzione normale standard.

Si determina q in modo che

$$P(-q < Z < q) = \frac{95}{100}.$$

Dalle tavole si ottiene $z = 1,96$. L' intervallo aleatoria richiesto é

$$\left[\bar{X} - 1,96\sqrt{\frac{10^{-2}}{6}}, \bar{X} + 1,96\sqrt{\frac{10^{-2}}{6}}\right].$$

L'intervallo di confidenza con i dati del problema si ottiene sostituendo a $\bar{X}$ la media empirica.