ST1- 1 ESONERO: 3-11-2010 *(Orlandi)*

Esercizio 1 (6 punti) Siano X and Y due variabili aleatorie indipendenti con distribuzione $f(x,\theta) = \frac{1}{\theta} e^{-\frac{x}{\theta}} \chi_{(0,\infty)}(x)$.

(1) Si determini la distribuzione della variabile aleatoria $Z = \frac{X}{Y}$.

(2) $\sin X$ e Y^2 sono stocasticamente indipendenti ?

Soluzione Per determinare la distribuzione della variabile aleatoria $Z = \frac{X}{Y}$ si consideri per $z > 0$

$$P[Z \le z] = P[X \le zY] = \int_0^\infty dy \frac{1}{\theta} e^{-\frac{y}{\theta}} \int_0^{zy} dx \frac{1}{\theta} e^{-\frac{x}{\theta}}$$
$$= \int_0^\infty dy \frac{1}{\theta} e^{-\frac{y}{\theta}} \left[1 - e^{-\frac{zy}{\theta}} \right] = 1 - \frac{1}{(1+z)}$$

Quindi

$$f_Z(z) = \frac{d}{dz} P[Z \le z] = \frac{1}{(1+z)^2}.$$

Le due variabili aleatorie $\sin X$ e Y^2 sono stocasticamente indipendenti poiché X e Y sono indipendenti.

Esercizio 2 (18 punti) Siano $(X_1, \dots, X_n)$ indipendenti estratti da una popolazione con distribuzione

$$f(x,\theta) = \frac{1}{\theta} x^{\frac{(1-\theta)}{\theta}}, \quad x \in (0,1). \quad \theta > 0$$

.

(1) Determinare la distribuzione di $Y = -\log X$, dove X e' distribuita secondo $f(x,\theta)$. Determinare media e varianza di Y.

(2) Trovare lo stimatore di massima verosimiglianza di θ.

(3) É uno stimatore non distorto? (motivare).

(4) Determinare l' errore quadratico medio.

(5) Si determini la distribuzione di T_n.

(6) Si consideri la successione degli stimatori T_n al variare della lunghezza del campione n, $\{T_n\}_n$. Si dica cosa si intende per successione di stimatori *semplicemente consistenti* e si verifichi che $\{T_n\}_n$ lo sia.

Soluzione La distribuzione di $Y = -\log X$ si determina facilmente per $y \in (0,\infty)$

$$P[Y \le y] = P[\log X > -y] = P[X > e^{-y}] = \int_{e^{-y}}^1 dy \frac{1}{\theta} y^{\frac{(1-\theta)}{\theta}} = 1 - e^{-\frac{y}{\theta}}$$

Derivando

$$f_Y(y,\theta) = \frac{1}{\theta} e^{-\frac{y}{\theta}}, \qquad y \in (0,\infty).$$

Quindi Y é una variabile aleatoria distribuita come una esponenziale di parametro $\frac{1}{\theta}$. Si ricorda che $E[Y] = \theta$ e $var(Y) = \theta^2$.

La funzione di verosimiglianza é data da

$$L(x_1, \dots, x_n; \theta) = \frac{1}{\theta^n} \prod_{i=1}^n x_i^{\frac{(1-\theta)}{\theta}}, \qquad x_i \in (0,1).$$

Per minimizzare conviene considerare

$$\log L(x_1, \dots, x_n; \theta) = -n \log\theta + \sum_{i=1}^n \frac{(1-\theta)}{\theta} \log x_i, \qquad x \in (0,1).$$

Derivando rispetto alla variabile θ si ottiene

$$\frac{\partial \log L}{\partial \theta} = -\frac{n}{\theta} - \frac{1}{\theta^2}\sum_{i=1}^{n} \log x_i.$$

Il punto critico é $\hat{\theta} = -\frac{1}{n}\sum_{i=1}^{n} \log x_i$. Per verificare se esso é un punto di massimo si calcola la derivata seconda rispetto a θ di $\log L(x_1, \dots, x_n; \theta)$. Si ottiene

$$\frac{\partial^2 \log L}{\partial^2 \theta} = \frac{n}{\theta^2} + \frac{2}{\theta^3}\sum_{i=1}^{n} \log x_i.$$

Calcolando in $\hat{\theta}$ si ottiene

$$\frac{\partial^2 \log L}{\partial^2 \theta}|_{\hat{\theta}} = -\frac{n^3}{(\sum_{i=1}^{n} \log x_i)^2} < 0.$$

Quindi $\hat{\theta}$ é un punto di massimo. Lo stimatore di θ é quindi

$$T_\theta = -\frac{1}{n}\sum_{i=1}^{n} \log X_i.$$

Direttamente o tenendo conto che $Y = -\log X$ é un esponenziale di parametro $\frac{1}{\theta}$ si ottiene: $E[-logX] = \theta$.

$$E[T_\theta] = \theta$$

Quindi T_θ é uno stimatore corretto di θ.

L'errore quadratico medio si ottiene, considerando che $E[Y] = \theta$, calcolando

$$E[(T_\theta - \theta)^2] = E\left[\frac{1}{n}\sum_{i=1}^{n}(Y_i - \theta)\right]^2 = \frac{1}{n}var(Y) = \frac{\theta^2}{n}.$$

Per determinare la distribuzione di T_n si puó usare il metodo della funzione generatrice dei momenti. Sia $Y_i = -\log X_i$ e $m(t)$ la funzione generatrice dei momenti di T_n . Per $t < \frac{n}{\theta}$ otteniamo

$$m(t) = E[e^{tT_n}] = \prod_{i=1}^{n} E[e^{\frac{t}{n}Y_i}] = \left(\frac{\theta^{-1}}{\theta^{-1} - \frac{t}{n}}\right)^n = \left(\frac{n\theta^{-1}}{n\theta^{-1} - t}\right)^n.$$

Quindi la distribuzione di T_n é una distribuzione Γ di parametri n e $n\theta^{-1}$.

Uno successione di stimatori $\{T_n\}_n$ é semplicemente consistente se per ogni $\epsilon > 0$

$$\lim_{n\to\infty} P[|T_n - \theta| \leq \epsilon] = 1.$$

Poiché

$$P[|T_n - \theta| \leq \epsilon] = 1 - P[|T_n - \theta| > \epsilon],$$

Applicando la diseguaglianza di Cheybishev

$$P[|T_n - \theta| \geq \epsilon] \leq \frac{1}{\epsilon^2}var(T_n) = \frac{\theta^2}{n\epsilon^2},$$

immediatamente si verifica che $-\frac{1}{n}\sum_{i=1}^{n} \log X_i$ é una successione al variare di n semplicemente consistente.

Esercizio 3 (6 punti)

Siano $(X_1, \dots, X_n)$ indipendenti estratti da una popolazione con distribuzione uniforme nell'intervallo $(\theta - 1, \theta + 1)$.

(1) Trovare lo stimatore di θ con il metodo dei momenti.
(2) É uno stimatore non distorto? (motivare).
(3) Date le seguenti 5 osservazioni di X dare una stima puntuale di θ:

$$6,61 \qquad 7,70 \qquad 6,98 \qquad 8,36 \qquad 7,26$$

Soluzione La media campionaria e' il momento campionaria di ordine uno, $M_1 = \frac{1}{n}\sum_{i=1}^{n} X_i$. Si ottiene immediatamente che

$$E[M_1] = \int_{\theta-1}^{\theta+1} \frac{x}{2} dx = \theta.$$

Quindi lo stimatore di θ con il metodo dei momenti é

$$T_\theta = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

E' uno stimatore corretto. La stima puntuale di θ con i dati dell'esercizio é

$$\hat{\theta} = \frac{1}{5}(6,61 + 7,70 + 6,98 + 8,36 + 7,26) = \frac{36,91}{5} = 7,382.$$