ST410- Esame 3-2-2011 *(Orlandi)*

Esercizio 1 (16 punti)
Sia $(X_1,\ldots,X_n)$ un campione estratto da $= f(x,\theta) = \theta x^{\theta-1}\chi_{(0,1)}(x)$, $\theta > 0$.

(1) Determinare lo stimatore T di $\frac{1}{\theta}$ con il metodo della massima verosimiglianza.
(2) Determinare se lo stimatore trovato é o non é distorto.
(3) Si determini la distribuzione di T.
(4) Determinare l' errore quadratico medio.
(5) 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.
(6) Si definisca cosa si intende per statistica sufficiente. É T una statistica sufficiente?
(7) É T un UMVUE (stimatore non distorto a varianza uniformemente minina).
(8) Determinare con la disuguaglianza di Cramer -Rao il limite inferiore della varianza di T

Soluzione Per determinare lo stimatore T di θ con il metodo della massima verosimiglianza si consideri la funzione di verosimiglianza

$$F(x_1,\ldots,x_n;\theta) = \prod_{i=1}^{n}\left(\theta x_i^{\theta-1}\chi_{(0,1)}(x_i)\right).$$

Per determinare il massimo di $F(x_1,\ldots,x_n;\theta)$ per $\theta > 0$ si consideri

$$L(x_1,\ldots,x_n;\theta) = \sum_{i=1}^{n}\left(\log\theta + (\theta-1)\log(x_i\chi_{(0,1)}(x_i))\right)$$

$$L'(x_1,\ldots,x_n;\theta) = \frac{n}{\theta} + \sum_{i=1}^{n}\log(x_i\chi_{(0,1)}(x_i)).$$

Quindi lo stimatore é

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

Sia $Z = -\log X$. La funzione di distribuzione di probabilitá di Z é

$$f_Z(z) = \theta e^{-z\theta}\chi_{(0,\infty)}(z).$$

Immediatamente si calcola

$$E[T] = \frac{1}{\theta}$$

Da ció si deduce che T é uno stimatore non distorto di $\frac{1}{\theta}$. Per determinare la distribuzione di T si puó usare il metodo della funzione generatrice dei momenti. Sia $Z_i = -\log X_i$ e $m(t)$ la funzione generatrice dei momenti di T . Per $t < n\theta$ otteniamo

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

Quindi la distribuzione di T_n é una distribuzione Γ di parametri n e $n\theta$.

L'errore quadratico medio si ottiene, considerando che $E[Z] = \frac{1}{\theta}$ and $var(Z) = \frac{1}{\theta^2}$. Si ottiene

$$E[(T-\frac{1}{\theta})^2] = E\left[\frac{1}{n}\sum_{i=1}^{n}(Z_i-\frac{1}{\theta})\right]^2 = \frac{1}{n}var(Z) = \frac{1}{\theta^2 n}.$$

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

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

Poiché

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

Applicando la diseguaglianza di Chebyshev

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

Quindi $-\frac{1}{n}\sum_{i=1}^n \log X_i$ é una successione al variare di n semplicemente consistente. La densitá appartiene alla famiglia esponenziale. Infatti

$$f(x,\theta) = a(\theta)b(x)e^{c(\theta)d(x)}$$

con $a(\theta) = \theta$, $b(x) = \chi_{(0,1)}(x)$, $c(\theta) = \theta - 1$, $d(x) = \log x$.

Una statistica S é sufficiente se e solo se la distribuzione condizionata di $(X_1, \dots, X_n)$ dato $\{S = t\}$ non dipende da θ per qualsiasi valore di t.

Un criterio per stabilire se T é una statistica sufficiente é dato dal teorema 7.4 del testo. La distribuzione appartiene alla famiglia esponenziale e quindi $S = \sum_{i=1}^n d(X_i)$ é una statistica sufficiente. Ma $T = -\frac{1}{n}S$ é una funzione di S e quindi é una statistica sufficiente.

Dalla diseguaglianza di Cramer Rao, la stima inferiore per la varianza di T é data da

$$var[T] \geq \frac{1}{nE[\left(\frac{\partial}{\partial\theta}\ln f(X,\theta)\right)^2]}.$$

Si ha che

$$\frac{\partial}{\partial\theta}\ln f(x,\theta) = \frac{1}{\theta} + \log x$$

$$E[(\frac{1}{\theta} + \log X)^2] = E[(\frac{1}{\theta} - Z)^2] = \frac{1}{\theta^2} = var(Z)$$

Quindi il limite inferiore di Cramer Rao coincide con $var[T]$. Quindi T é un UMVUE di $\frac{1}{\theta}$.

Esercizio 2 (8 punti)
Sia (X_1, X_2) un campione estratto da $f(x,\theta) = \theta x^{\theta-1}\chi_{(0,1)}(x)$, $\theta > 0$. Si vuole verificare

$$H_0 : \theta = 2, \quad \text{contro} \quad H_1 : \theta > 2.$$

Si costruisca un test piú potente di ampiezza $\alpha = 0,05$.
Soluzione La densitá appartiene alla famiglia esponenziale. Infatti

$$f(x,\theta) = a(\theta)b(x)e^{c(\theta)d(x)}$$

con $a(\theta) = \theta$, $b(x) = \chi_{(0,1)}(x)$, $c(\theta) = \theta - 1$, $d(x) = \log x$. Inoltre, poiché $c(\theta)$ é una funzione crescente di θ, applicando il teorema 9.5 del testo abbiamo che il test Y uniformente piú potente é determinato dalla regione critica $C = C_Y$

$$C = \{(x_1, x_2) \in (R^+)^2 : \sum_{i=1}^{2} \log X_i \geq k\},$$

con $k < 0$ scelto in modo

$$P_{\theta=2}[C] = \alpha = 0,05.$$

Per determinare k si osservi che $Z = -\log X$ ha una distribuzione esponenziale $f(z) = \theta e^{-\theta z}$. Quindi $T = -\sum_{i=1}^{2} \log X_i$ ha una distribuzione $\Gamma(2, \theta)$. Si ponga $k = -k^*$ con $k^* > 0$ e tale

$$\alpha = \int_0^{k^*} \Gamma(2,2)(z)dz.$$

Analoghe conclusioni si possono dedurre applicando il criterio di verosimiglianza generalizzato e studiando la funzione potenza.

Esercizio 3 (6 punti)
Siano X e Y due variabili casuali indipendenti che assumono valori interi positivi, aventi funzione di distribuzione $f(x) = 2^{-x}$ per $x = 1, 2, \ldots$ Trovare:

(1) $P[min\{X, Y\} \leq a]$ con $a > 0$.
(2) $P[Y > X]$.

Soluzione Sia $Z = min\{X, Y\}$ e $[a]$ parte intera di a. Si ponga $[a] \geq 1$.

$$P[Z \leq a] = 1 - P[Z > a] = 1 - P[X > a, Y > a]$$

$$P[X > a, Y > a] = (P[X > a])^2$$
$$P[X > a] = 1 - P[X \leq [a]]$$

Quindi

$$P[X > a] = 1 - \sum_{x=1}^{[a]} 2^{-x} = 1 - (1 - 2^{-[a]}) = 2^{-[a]}.$$

Quindi

$$P[Z \leq a] = 1 - 2^{-2[a]}.$$

Inoltre

$$P[Y > X] = \sum_{i=1}^{\infty} P[Y > i]P[X = i] = \sum_{i=1}^{\infty} P[Y > i]2^{-i} = \sum_{i=1}^{\infty} \sum_{j=i+1}^{\infty} 2^{-j}2^{-i}.$$