ST1- 1 ESONERO: 16-4-2010 *(Orlandi)*

Esercizio 1 (14 punti) Siano (X_1, X_2) un campione di lunghezza $n = 2$ estratto da una popolazione con distribuzione uniforme nell'intervallo $[0, 1]$.

(1) Mostrare che $\mu = \frac{1}{2}$ é la media della popolazione e $\sigma^2 = \frac{1}{12}$ la varianza.
(2) Verificare che la varianza campionaria é data da $S^2 = \frac{1}{2}(X_1 - X_2)^2$.
(3) Mostrare che $E[S^4] = \frac{1}{60}$.
(4) Sia $\bar{X}$ la media campionaria. Determinare il valore numerico di

$$P[\bar{X} \leq \frac{1}{4}].$$

(5) Determinare il valore numerico di

$$P[S^2 \geq \frac{1}{8}].$$

(6) Determinare la funzione generatrice dei momenti di X_1.
(7) Dimostrare che $\bar{X}$ e S^2 non sono stocasticamente indipendenti.

Soluzione

$$E[X] = \int_0^1 x dx = \frac{1}{2}, \quad E[X^2] = \int_0^1 x^2 dx = \frac{1}{3},$$

quindi $\mu = \frac{1}{2}$ e $\sigma^2 = \frac{1}{3} - \frac{1}{4} = \frac{1}{12}$.

Dalla definizione

$$S^2 = (X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 = \frac{1}{2}(X_1 - X_2)^2.$$

$$E[S^4] = \frac{1}{4}E[(X_1 - X_2)^4] = \frac{1}{60}$$

Il risultato si ottiene sviluppando il polinomio, usando l'indipendenza di X_1 e X_2 e svolgendo semplici integrali.

$$P[\bar{X} \leq \frac{1}{4}] = P[X_1 + X_2 \leq \frac{1}{2}] = \int_0^{\frac{1}{2}} dx_1 \int_0^{\frac{1}{2} - x_1} dx_2$$
$$= \int_0^{\frac{1}{2}} (\frac{1}{2} - x_1) dx_1 = \frac{1}{4} - \frac{1}{8} = \frac{1}{8}.$$

$$P[S^2 \geq \frac{1}{8}] = P[(X_1 - X_2)^2 \geq \frac{1}{4}] = 2P[(X_1 - X_2) \geq \frac{1}{2}].$$

$$P[(X_1 - X_2) \geq \frac{1}{2}] = \int_{\frac{1}{2}}^1 dx_1 \int_0^{x_1 - \frac{1}{2}} dx_2 = \int_{\frac{1}{2}}^1 (x_1 - \frac{1}{2}) dx_1 = \frac{1}{2}[1 - \frac{1}{4}] - \frac{1}{4} = \frac{1}{8}$$

Quindi

$$P[S^2 \geq \frac{1}{8}] = \frac{1}{4}.$$

La funzione generatrice dei momenti di X_1 é

$$m(t) = E[e^{tX_1}] = \frac{e^t - 1}{t}$$

Per mostrare che $\bar{X}$ e S^2 non sono stocasticamente indipendenti si puó utilizzare il metodo delle funzione generatrice dei momenti. Dimostriamo che $Y_1 = X_1 + X_2$ e $Y_2 = X_1 - X_2$ non sono stocasticamente indipendenti. Poiché $\bar{X}$ e S^2 sono funzioni rispettivamente di Y_1 e Y_2 otteniamo che anche $\bar{X}$ e S^2 non sono stocasticamente indipendenti. Si ha

$$E[e^{tY_1}] = E[e^{t(X_1+X_2)}] = E[e^{tX_1}]E[e^{tX_2}] = \left(\frac{e^t - 1}{t}\right)^2$$

$$E[e^{tY_2}] = E[e^{t(X_1-X_2)}] = E[e^{tX_1}]E[e^{-tX_2}] = \left(\frac{e^t - 1}{t}\right)\left(\frac{1 - e^{-t}}{t}\right)$$

Le variabili aleatorie Y_1 e Y_2 sono stocasticamente indipendenti se e solo se per ogni $s \in R$ e $t \in R$

$$E[e^{tY_1}e^{sY_2}] = E[e^{tY_1}]E[e^{sY_2}].$$

Si calcola facilmente che

$$E[e^{tY_1}e^{sY_2}] = E[e^{t(X_1+X_2)}e^{s(X_1-X_2)}] = E[e^{(t+s)X_1}e^{(t-s)X_2}]$$
$$= E[e^{(t+s)X_1}]E[e^{(t-s)X_2}] = \left(\frac{e^{t+s} - 1}{t+s}\right)\left(\frac{e^{t-s} - 1}{t-s}\right)$$

Poiché

$$E[e^{tY_1}e^{sY_2}] \neq E[e^{tY_1}]E[e^{sY_2}],$$

$\bar{X}$ e S^2 non sono stocasticamente indipendenti.

Esercizio 2 (12 punti) Siano $(X_1, \ldots, X_n)$ indipendenti estratti da una popolazione con distribuzione normale di media μ e varianza σ^2.

(1) Trovare lo stimatore di massima verosimiglianza di μ supponendo nota la varianza.
(2) Si determini la sua distribuzione campionaria.
(3) É uno stimatore non distorto? (motivare).
(4) Determinare l' errore quadratico medio.
(5) É lo stimatore trovato T_μ una statitistica sufficiente? (Si scriva cosa si intende per statistica sufficiente).
(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.
(7) Si trovi il limite inferiore di Cramer-Rao per lo stimatore trovato.
(8) É lo stimatore un UMVUE?

Soluzione

(1) Si trova facilmente che lo stimatore di massima verosimiglianza é

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

Attenzione: si deve sempre verificare che é il massimo della funzione di massima verosimiglianza!!!

(2) La distribuzione campionaria é una normale con media μ e varianza $\frac{\sigma^2}{n}$. Infatti si calcola la funzione generatrice dei momenti di $\frac{1}{n}\sum_{i=1}^{n} X_i$

$$E[e^{t\frac{1}{n}\sum_{i=1}^{n} X_i}] = \prod_{i=1}^{n} E[e^{t\frac{1}{n}X_i}] = \left(E[e^{t\frac{1}{n}X}]\right)^n.$$

Si ottiene che

$$E[e^{t\frac{1}{n}X}] = e^{\frac{t}{n}\mu + \frac{t^2}{n^2}\sigma^2}.$$

Quindi

$$E[e^{t\frac{1}{n}\sum_{i=1}^n X_i}] = e^{t\mu + t^2\frac{\sigma^2}{n}}.$$

(3)

$$E[T_\mu] = \mu,$$

quindi lo stimatore non é distorto.

(4) L' errore quadratico medio é

$$E[(T_\mu - \mu)^2] = \frac{\sigma^2}{n}.$$

(5) Lo stimatore trovato é una statistica sufficiente. Infatti la distribuzione normale appartiene alla *famiglia esponenziale*. vedi pag 318 del libro di testo. La densitá congiunta di $(X_1, \dots, X_n)$ si fattorizza come una funzione della statistica sufficiente

$$\prod_{i=1}^n f(x_i, \mu, \sigma) = g(T_\mu(x_1, x_2, \dots x_n))h(x_1, x_2, \dots x_n)$$

(6) Uno successione di stimatore é semplicemente consistente se per ogni $\epsilon > 0$

$$\lim_{n\to\infty} P[|T_\mu - \mu| \le \epsilon] = 1.$$

Poiché

$$P[|T_\mu - \mu| \le \epsilon] = 1 - P[|T_\mu - \mu| \ge \epsilon],$$

e applicando la diseguaglianza di Cheybishev

$$P[|T_\mu - \mu| \ge \epsilon] \le \frac{1}{\epsilon^2} var(T_\mu) = \frac{\sigma^2}{n\epsilon^2},$$

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

(7) Il limite inferiore Cramer-Rao per lo stimatore trovato si trova facilmente. E' dato da

$$\frac{\sigma^2}{n}.$$

(8) Quindi coincide con la varianza di T_μ. Lo stimatore trovato e' un UMVUE.

Esercizio 3 (4 punti)

Si consideri la funzione di densitá di una coppia di variabili aleatorie (X, Y),

$$f(x,y) = \begin{cases} 24xy, & x \ge 0, y \ge 0, x + y \le 1 \\ 0 & \text{altrove.} \end{cases}$$

Verificare se X e Y sono tra loro dipendenti o indipendenti.

Soluzione Si calcolano facilmente le distribuzioni marginali.

$$f_X(x) = 24x \int_0^{1-x} y dy = 12x(1-x)^2$$

$$f_Y(y) = 24y \int_0^{1-y} x dy = 12y(1-y)^2.$$

Si verifica immediatamente che

$$f_X(x) f_Y(y) \neq f(x,y).$$

Quindi le variabili aleatorie X e Y sono tra loro dipendenti.