Secondo compitino di

Probabilità e Statistica

Laurea Triennale in Matematica

17/06/2014

COGNOME e NOME ROSELLI MICHELE

N. MATRICOLA 1069134

Esercizio 1. (V. 2 punti.)
Fornire la definizione di stimatore corretto.

Sia $(\Omega, \mathcal{A}, P_\theta)_{\theta\in\Theta}$ un modello statistico parametrico e f una funzione misurabile
La statistica campionaria $(Y_n)_{n\in\mathbb{N}}$ si dice stimatore corretto per f se:

$$\forall \theta \in \Theta, \forall n\in\mathbb{N} \quad E_\theta(Y_n) = f(\theta).$$

Esercizio 2. (V. 4 punti.)
Siano X e Y due variabili aleatorie (non necessariamente indipendenti) supponiamo che $X \sim Bin(3, \frac{1}{2})$ e $Y \sim exp(1)$. Sia $W := X + Y$. E' possibile che W abbia distribuzione Poissoniana? (`Sugg. Se è possibile fornire un esempio. Se non è possibile fornire una dimostrazione`)

Supponiamo per assurdo che W abbia distribuz. Poissoniana. Allora $\mathbb{P}(W=k) = e^{-\lambda} \frac{\lambda^k}{k!}$ per qualche $\lambda > 0$. Questa probabilità è maggiore stretta di zero. Ma $\mathbb{P}(W=k) = \mathbb{P}(X+Y=k)$.

Ma X può assumere solo 4 valori ed i valori che ~~può~~ deve assumere Y affinché $X+Y=k$ sono dunque al più 4. Ma la probabilità che $exp(1)$ assuma un valore preciso è 0 poiché è ~~ass~~ continua, per ognuno dei 4 valori. Dunque $\mathbb{P}(X+Y=k) = 0 \ \forall k$. $\lightning$

Esercizio 3. (V. 10 punti.)
Sia $(X_n)_{n\in\mathbb{N}}$ una successione di variabili aleatorie indipendenti ed identicamente distribuite con distribuzione uniforme sull'intervallo $(1,7)$, $X_n \sim Unif(1,7)$. Per ogni $n \in \mathbb{N}$ siano definite le seguenti variabili aleatorie:

$$Y_n := \frac{X_1 + \ldots + X_n}{n} \qquad T_n := \frac{Y_1 + \ldots + Y_n}{n}$$

Calcolare media e varianza di Y_n. Le variabili aleatorie $(Y_n)_{n\in\mathbb{N}}$ sono indipendenti? Sono non correlate? Dimostrare che la successione $(Y_n)_{n\in\mathbb{N}}$ soddisfa una legge dei grandi numeri (debole o forte a scelta).

+ 4

$$\forall n \in \mathbb{N}: \mathbb{E}(X_n) = \frac{7-1}{2} = 3, \quad \mathbb{E}(Y_n) = \frac{1}{n}\mathbb{E}(X_1 + \ldots + X_n) =$$

$$= \frac{1}{n}\left[\mathbb{E}(X_1) + \mathbb{E}(X_2) + \ldots + \mathbb{E}(X_n)\right] = \frac{3n}{n} = 3.$$

$$\forall n \in \mathbb{N}: VAR(X_n) = \frac{(7-1)^2}{12} = \frac{36}{12} = 3, \quad VAR(Y_n) = \frac{1}{n^2} VAR(X_1 + \ldots + X_n) =$$

$$= \frac{3n}{n^2} = \frac{3}{n}$$

WLOG $n \geq m$;

$$Cov(Y_m, Y_n) = cov\left(\frac{1}{m}(X_1 + \ldots + X_m); \frac{1}{n}(X_1 + \ldots + X_m) + \frac{1}{n}(X_{m+1} + \ldots + X_n\right)$$

$$= \frac{1}{mn} cov(X_1 + \ldots X_m, X_1 + \ldots X_m) + \frac{1}{nm} cov(X_1 + \ldots + X_m, X_{m+1} + \ldots X$$

$$= \frac{VAR(X_1 + \ldots + X_m)}{mn} + \frac{\sum_{\substack{1 \leq i \leq m \\ m+1 \leq j \leq n}} cov(X_i; X_j)}{mn} =$$

$\neq 0$ $\left(= \frac{3}{n}\right)$

$= 0$

X_i INDIP DA X_j SE $i \neq j$

SONO CORRELATE
QUINDI NON SONO
INDIP.

Considero T_n, ovvero $\bar{Y}_n$ ovvero $\frac{Y_1 + \dots - Y_n}{n}$. Per la linearità del valor medio $\mathbb{E}(T_n) = 3, \forall n \in \mathbb{N}$.

Per la disuguaglianza di Chebyshev:

$$\forall n \in \mathbb{N}, \forall \varepsilon > 0, \ \mathbb{P}(|T_n - \mathbb{E}(T_n)| > \varepsilon) \leq \frac{VAR(T_n)}{\varepsilon^2} = \frac{\frac{1}{n^2} \cancel{COV}(Y_1 + \dots Y_n ; Y_1 + \dots)}{\varepsilon^2}$$

$$= \frac{1}{n^2 \varepsilon^2} \cdot \sum_{i,j \in \{1, \dots n\}} COV(Y_i ; Y_j) = \frac{3}{n^2 \varepsilon^2} \left(\frac{1}{1} + \frac{3}{2} + \frac{5}{3} + \frac{7}{4} + \dots \frac{2n-1}{n} \right)$$

$$\uparrow$$

$$COV(Y_i ; Y_j) = \frac{3}{\max(i;j)}$$

(PUNTO PRECEDENTE)

$$= \frac{3}{n^2 \varepsilon^2} \left(\sum_1^n \frac{2n-1}{n} \right) = \frac{3}{n^2 \varepsilon^2} \left(\sum_1^n 2 - \frac{1}{n} \right) \leq \frac{3}{n^2 \varepsilon^2} \sum_1^n 2 =$$

$$= \frac{6n}{n^2 \varepsilon^2} = \frac{6}{n \varepsilon^2}.$$

Passo al limite perché vale $\forall n$:

$$\lim_{n \to +\infty} \mathbb{P}(|T_n - 3| > \varepsilon) \leq \lim_{n \to +\infty} \frac{6}{n \varepsilon^2} = 0$$

$\square$

Esercizio 4. (V. 6 punti.)
Sia $(\Omega, \mathcal{A}, P_\theta)_{\theta\in\Theta}$ un modello statistico con $\Theta = (0, \infty)$. Sia $(X_n)_{n\in\mathbb{N}}$ un campione tale che per ogni $\theta > 0$, le v.a. X_n abbiano distribuzione assolutamente continua con densità f_θ data da:

$$f_\theta(x) = \begin{cases} \alpha_\theta x & x \in (0, \theta) \\ 0 & \text{altrimenti} \end{cases}$$

Sia infine $T_n := \beta \frac{X_1 + \ldots + X_n}{n}$ per ogni $n \in \mathbb{N}$.
(a) Calcolare α_θ.
(b) Dimostrare che per $\beta = \frac{3}{2}$ si ha: T_n stimatore corretto per θ.
(c) Sia $\beta = \frac{3}{2}$ dimostrare che per ogni $\theta > 0$ vale:
$P_\theta(\omega \in \Omega | \lim_{n\to\infty} T_n(\omega) = \theta) = 1$

a) $$\int_{-\infty}^{+\infty} f_\Theta(x) = 1 = \int_0^\Theta \alpha_\Theta x = \left|\frac{\alpha_\Theta}{2} x^2\right|_0^\Theta = \frac{\Theta^2 \alpha_\Theta}{2} - 0 \rightarrow \alpha_\Theta = \frac{2}{\Theta^2}$$

b) Basta mostrare che $\mathbb{E}\left(\frac{3(X_1 + .. X_n)}{2n}\right) = \Theta$.

Ho che $\mathbb{E}(X_i) = \int_0^\Theta \frac{2}{\Theta^2} x^2 = \left|\frac{2}{3\Theta^2} x^3\right|_0^\Theta = \frac{2\Theta}{3}$

Per la linearità del valor medio: $\mathbb{E}(T_n) = \frac{n \cdot \frac{2\Theta}{3} \cdot 3}{2n} = \Theta$.

c) $\forall \Theta$, La successione $X_n, \ldots X_n$ soddisfa la legge forte dei grandi numeri perché le X_i sono indipendenti e con valor medio uguale a $\frac{2\Theta}{3}$.

Quindi:

$$P_\Theta\left(\omega \in \Omega \,\middle|\, \lim_{n\to\infty} \frac{X_1 + .. + X_n}{n} = \frac{2\Theta}{3}\right) = 1$$

moltiplicando per $\frac{3}{2}$ da entrambe le parti e portando tale scalare

Esercizio 5. (V. 8 punti.)

Una compagnia di assicurazioni auto prevede per i guidatori giovani una polizza più alta, in quanto questo gruppo tende ad avere un numero maggiore di incidenti. La compagnia ha 80000 assicurati divisi in 3 gruppi:

A (sotto i 25 anni, 20% di tutti i suoi assicurati),

B (25-35 anni, 30%),

C (sopra i 35 anni 50%).

I dati mostrano che in media ogni anno le percentuali di assicurati che hanno un incidente sono: 10% per il gruppo A, 5% per il B, 2% per il C. Sia X il numero di assicurati che avranno un incidente il prossimo anno.

(a) Calcolare media e varianza di X.

(b) Scelgo un assicurato a caso (tra gli 80000), qual è la probabilità che abbia un incidente.

(c) Scelgo un assicurato a caso (tra gli 80000), se so che ha avuto un incidente qual è la probabilità che sia del gruppo A?

(d) Stimare (utilizzando l'approssimazione normale) la probabilità che il numero di assicurati che hanno un incidente sia maggiore (strettamente) di 3700.

a) $$\mathbb{E}(X) = \mathbb{E}(A) + \mathbb{E}(B) + \mathbb{E}(C) = 16000 \cdot \frac{1}{10} + 24000 \cdot \frac{1}{20} + 40000 \cdot \frac{1}{50} =$$

$$= 1600 + 1200 + 800 = 3600.$$

SUPPONENDO CHE OGNI PERSONA SIA INDIP. DALLE ALTRE

$$\mathrm{Var}(X) = 16000 \cdot \frac{1}{10} \cdot \frac{9}{10} + 24000 \frac{1}{20} \cdot \frac{19}{20} + 40000 \cdot \frac{1}{50} \cdot \frac{49}{50} =$$

$$= 1440 + 1140 + 784 = 3364$$

b) DISINTEGRANDO: $$\frac{1}{5} \cdot \frac{1}{10} + \frac{3}{10} \cdot \frac{1}{20} + \frac{1}{2} \cdot \frac{1}{50} =$$

$$\frac{1}{50} + \frac{3}{200} + \frac{1}{100} = \frac{4+3+2}{200} = 4,5\%$$

c) $$P(A|I) = \frac{P(I|A) \cdot P(A)}{P(I)} = \frac{\frac{1}{10} \cdot \frac{1}{5}}{\frac{9}{200}} = \frac{200}{450} = \frac{4}{9}$$

d) $$P(X > 3700) \simeq P(X_n \geq 3700,5) = 1 - P(X_n \leq 3700,5)$$

$$= 1 - P\left(Z_n \leq \frac{3700,5 - 3600}{\sqrt{3364}}\right) = 1 - P(Z_n \leq 1,73) = 1 - 0,958 =$$

$$\simeq 0,042.$$