Statistica - CPS
Corso di Laurea in Informatica
Compito del 09-05-2022

Esercizio 1. (8 punti) Un test a risposta multipla prevede 10 domande con 4 possibili risposte per ogni domanda. Il punteggio viene assegnato con le seguenti regole: +1 punto per ogni risposta esatta; -0.25 punti per ogni risposta sbagliata o non data.
Supponendo di rispondere a tutte le domande del test in maniera casuale, con uguale probabilità per ogni risposta, si determini:

(i) la probabilità di ottenere almeno 8 risposte esatte;

(ii) il valore atteso della variabile aleatoria che descrive il punteggio ottenuto al test.

Esercizio 2. (12 punti) Si consideri la funzione

$$F_{a,b}(x) = \begin{cases} 0\,, & \text{se } x \leq 0; \\ a - e^{-x}\,, & \text{se } 0 < x < 1; \\ a - e^{-1} + b(x-1)\,, & \text{se } 1 \leq x \leq 2; \\ 1\,, & \text{se } x > 2; \end{cases}$$

con parametri $a, b \in \mathbb{R}$.

(i) Si determinino i valori di a e b affinché $F_{a,b}$ sia la funzione di ripartizione di una variabile aleatoria X con densità.

(ii) Usando i valori di a e b trovati in (i), scrivere la formula per il calcolo dei β-quantili della variabile aleatoria X per ogni $\beta \in (0,1)$.

(iii) Sia X la variabile aleatoria in (i), scrivere la densità della variabile aleatoria $Y = (X+1)^2$, e calcolare il valore atteso di Y.

Esercizio 3. (10 punti) Si consideri un monitoraggio sulla presenza di sostanze tossiche nell'aria, effettuato in 10 stazioni di monitoraggio vicine. I valori ottenuti restituiscono una concentrazione media di 4.8 mg/dm^3 con una varianza campionaria di 0.49 mg/dm^3.

(i) Fornire una stima della concentrazione delle sostanze tossiche con una fiducia del 90% mediante un intervallo bilatero. Con quale fiducia si ottiene una precisione relativa di $5 \cdot 10^{-2}$?

(ii) Dire se l'ipotesi che la concentrazione non sia superiore a 4.3 mg/dm^3 è plausibile.

Esercizio 1 Un test a risposta multipla prevede 10 domande con 4 possibili risposte per ogni domanda.

Il punteggio viene assegnato con le seguenti regole: +1 punto per ogni risposta esatta; − 0.25 punti per ogni risposta sbagliata o non data.

Supponendo di rispondere a caso alle domande del test, in maniera indipendente tra le domande, si determini:

(i) la probabilità di ottenere almeno 8 risposte esatte;

(ii) il valore atteso della variabile aleatoria che descrive il punteggio ottenuto.

svolgimento Se X_i, $i=1,\dots,10$, è la variabile aleatoria che vale 1 se la risposta data all'i-esima domanda è corretta, e che vale 0 altrimenti, abbiamo $X_i \sim B(1, \frac{1}{4})$ $\forall i$.

Il numero totale di risposte corrette è dunque ottenuto con la v.a. $X = X_1 + X_2 + \dots + X_{10}$. Essendo le $\{X_i\}$ indipendenti, si ha $X \sim B(10, \frac{1}{4})$.

(i) Dobbiamo calcolare $\mathbb{P}(X \geq 8) = \mathbb{P}(X=8) + \mathbb{P}(X=9) + \mathbb{P}(X=10)$.

Si ha
$$\mathbb{P}(X=8) = \binom{10}{8}\left(\frac{1}{4}\right)^8\left(\frac{3}{4}\right)^2$$

$$\mathbb{P}(X=9) = \binom{10}{9}\left(\frac{1}{4}\right)^9\left(\frac{3}{4}\right)^1$$

$$\mathbb{P}(X=10) = \binom{10}{10}\left(\frac{1}{4}\right)^{10}\left(\frac{3}{4}\right)^0$$

quindi
$$\boxed{\mathbb{P}(X \geq 8) = \frac{45 \cdot 9}{4^{10}} + \frac{10 \cdot 3}{4^{10}} + \frac{1}{4^{10}} = \frac{436}{4^{10}} \sim 0.0004}$$

(ii) La v.a. Y che descrive il punteggio si ottiene sommando 1 per ogni domanda corretta, e sottraendo $\frac{1}{4}$ per ogni domanda non data o sbagliata.

Le domande corrette sono X, quelle non corrette sono $10-X$, quindi $Y = (+1)\cdot X + \left(-\frac{1}{4}\right)(10-X) = X + \frac{1}{4}X - \frac{10}{4} = \frac{5X-10}{4}$

Il valore atteso di Y è quindi

$$\boxed{E[Y] = \frac{1}{4}(5\cdot E[X]-10) = \frac{1}{4}\left(5\cdot 10\cdot \frac{1}{4} - 10\right) = = \frac{5}{2}\cdot\frac{1}{4} = \frac{5}{8}.}$$

Esercizio 2 Si consideri la funzione

$$F_{a,b}(x) = \begin{cases} 0, & x \le 0 \\ a - e^{-x}, & 0 < x < 1 \\ a - e^{-1} + b(x-1), & 1 \le x \le 2 \\ 1, & x > 2 \end{cases}$$

con parametri $a, b \in \mathbb{R}$.

(i) Si determinino i valori di a e b affinché $F_{a,b}$ sia la funzione di ripartizione di una v.a. X con densità.

(ii) Scrivere la formula per il calcolo dei β-quantili della v.a. X per ogni $\beta \in (0,1)$.

(iii) Scrivere la densità della v.a. $Y = (X+1)^2$, e calcolare il valore atteso di Y.

svolgimento (i) La funzione $F_{a,b}$ deve soddisfare le seguenti proprietà:

- $\lim_{x\to-\infty} F_{a,b}(x)=0$, $\lim_{x\to+\infty} F_{a,b}(x)=1$

- $F_{a,b}$ è debolmente crescente

- $F_{a,b}$ è continua

Imponendo la continuità, ottenimo che deve valere

$$F_{a,b}(0)=\lim_{x\to0^+} F_{a,b}(x) \iff 0=a-1 \implies a=1$$

$$F_{a,b}(1)=\lim_{x\to1^-} F_{a,b}(x) \iff a-e^{-1}=a-e^{-1}\ \text{, ok } \forall a$$

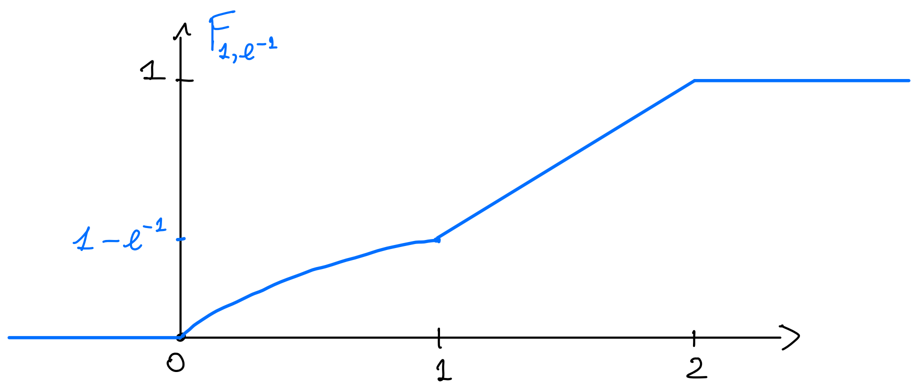
$$F_{a,b}(2)=\lim_{x\to2^+} F_{a,b}(x) \iff a-e^{-1}+b=1 \implies b=e^{-1}+1-a$$

Da cui ricaviamo $\boxed{a=1}$ e $\boxed{b=e^{-1}}$.

Le altre condizioni sono verificate per questi valori di a e b.

Per esempio basta verificare che $F'_{a,b}\geq 0$ $\forall x\neq\{0,1,2\}$.

La funzione $F_{a,b}$ è dunque



(ii) La funzione $F_{1,e^{-1}}$ è strettamente crescente nei punti in cui assume valori tra 0 e 1. Quindi $\forall \beta\in(0,1)$ esiste un unico r_β t.c. $F_{1,e^{-1}}(r_\beta)=\beta$.

Se $\beta\in(0,1-e^{-1}]$, $F_{1,e^{-1}}(r_\beta)=\beta \iff 1-e^{-r_\beta}=\beta \iff$

$$\iff \quad r_\beta = \log \frac{1}{1-\beta}$$

Se $\beta \in [1-e^{-1}, 1)$, $\quad F_{1,e^{-1}}(r_\beta) = \beta \iff 1 - e^{-1} + e^{-1}(r_\beta - 1) = \beta$

$$\iff \quad r_\beta = 1 + \frac{\beta - 1 + e^{-1}}{e^{-1}} = 2 + e(\beta - 1)$$

Quindi

$$\boxed{r_\beta = \begin{cases} \log \frac{1}{1-\beta}, & \text{se } \beta \in (0, 1-e^{-1}] \\ e(\beta-1)+2, & \text{se } \beta \in [1-e^{-1}, 1) \end{cases}}$$

(iii) La v.a. $Y = (X+1)^2$ si scrive come $Y = h(X)$ con $h(x) = (x+1)^2$.

Poiché $\operatorname{Imm} X = [0,2]$, e $h(x)$ è derivabile e invertibile su $[0,2]$, possiamo scrivere

$$f_Y(y) = \begin{cases} f_X(h^{-1}(y)) \left| \frac{d}{dy} h^{-1}(y) \right|, & \text{se } y \in h(0,2) = (1,9) \\ 0, & \text{altrimenti} \end{cases}$$

Inoltre

$$f_X(x) = \begin{cases} e^{-x}, & \text{se } 0 < x < 1 \\ e^{-1}, & \text{se } 1 \leq x < 2 \\ 0, & \text{altrimenti} \end{cases}$$

e $h^{-1}(y) = \sqrt{y} - 1 \quad \forall y \in (1,9)$, verifica $h^{-1}(y) \in (0,1)$ se $y \in (1,4)$, $h^{-1}(y) \in [1,2)$ se $y \in [4,9)$, quindi

$$\boxed{f_Y(y) = \begin{cases} e^{1-\sqrt{y}} \cdot \frac{1}{2\sqrt{y}}, & \text{se } y \in (1,4) \\ e^{-1} \cdot \frac{1}{2\sqrt{y}}, & \text{se } y \in [4,9) \\ 0, & \text{altrimenti} \end{cases}}$$

Per calcolare $E[Y]$ possiamo usare la definizione

$$E[Y] = \int_{-\infty}^{+\infty} y\, f_Y(y)\, dy = \int_1^4 \frac{\sqrt{y}}{2} e^{1-\sqrt{y}}\, dy + \int_4^9 \frac{\sqrt{y}}{2} e^{-1}\, dy$$

ma il calcolo è più semplice usando che

$$E[Y] = E[(X+1)^2] = E[X^2] + 2E[X] + 1.$$

Infatti

$$E[X] = \int_{-\infty}^{+\infty} x\, f_X(x)\, dx = \int_0^1 x e^{-x}\, dx + \int_1^2 x \cdot e^{-1}\, dx =$$

$$= \left[-(x+1)e^{-x}\right]\Big|_0^1 + \left[\frac{e^{-1}}{2} x^2\right]\Big|_1^2 =$$

$$= -2e^{-1} + 1 + 2e^{-1} - \frac{1}{2}e^{-1} = 1 - \frac{1}{2}e^{-1}$$

$$E[X^2] = \int_{-\infty}^{+\infty} x^2 f_X(x)\, dx = \int_0^1 x^2 e^{-x}\, dx + \int_1^2 x^2 e^{-1}\, dx =$$

$$= \left[-(x^2+2x+2)e^{-x}\right]\Big|_0^1 + \left[\frac{e^{-1}}{3} x^3\right]\Big|_1^2 =$$

$$= -5e^{-1} + 2 + \frac{8}{3}e^{-1} - \frac{e^{-1}}{3} = -\frac{8}{3}e^{-1} + 2$$

$$\Longrightarrow \quad E[Y] = -\frac{8}{3}e^{-1} + 2 + 2\left(1 - \frac{1}{2}e^{-1}\right) + 1$$

$$\Longrightarrow \quad \boxed{E[Y] = 5 - \frac{11}{3}e^{-1}}$$

Esercizio 3 Si consideri un monitoraggio sulla presenza di sostanze tossiche nell'aria, effettuato in 10 stazioni di monitoraggio. I valori ottenuti restituiscono una concentrazione media di 4.8 mg/dm^3 con una varianza campionaria di 0.49 mg/dm^3.

(i) Fornire una stima della concentrazione delle sostanze tossiche con una fiducia del 90% tramite un intervallo bilatero.
Con quale fiducia si ottiene una precisione relativa di $5 \cdot 10^{-2}$?

(ii) Dire se l'ipotesi che la concentrazione non sia superiore a 4.3 mg/dm^3 è plausibile.

<u>svolgimento</u>

(i) Il campione statistico si può descrivere come un insieme $X_1, \dots, X_{10}$ di v.a. indipendenti con $X_k \sim N(\mu, \sigma^2)$ $\forall k$ con μ e σ^2 ignote.

Utilizzando i dati $\bar{x} = 4.8$ e $\bar{\sigma} = \sqrt{0.49} = 0.7$, se la fiducia $1-\alpha = 0.9$, e $n = 10$, l'intervallo di fiducia bilatero cercato è

$$I = \left[\bar{x} - \tau_{(1-\alpha/2, n-1)} \frac{\bar{\sigma}}{\sqrt{n}}, \bar{x} + \tau_{(1-\alpha/2, n-1)} \frac{\bar{\sigma}}{\sqrt{n}} \right] =$$

$$= \left[4.8 - \tau_{(0.95, 9)} \frac{0.7}{\sqrt{10}}, 4.8 + \tau_{(0.95, 9)} \frac{0.7}{\sqrt{10}} \right]$$

che, utilizzando $\tau_{(0.95, 9)} \sim 1.8331$, diventa

$$\boxed{I \sim [4.394, 5.206]}.$$

La precisione relativa in questo caso è data da

$$\frac{\tau_{(1-\alpha/2, n-1)} \frac{\bar{\sigma}}{\sqrt{n}}}{|\bar{x}|} = \frac{\tau_{(1-\alpha/2, 9)} \frac{0.7}{\sqrt{10}}}{4.8}$$

quindi $$\frac{\tau_{(1-\alpha/2, 9)} \frac{0.7}{\sqrt{10}}}{4.8} \sim 5 \cdot 10^{-2}$$

$$\iff \quad t_{(1-\alpha/2,\,9)} \sim \frac{4.8 \cdot \sqrt{10}}{0.7} \cdot 5 \cdot 10^{-2} \sim 1.084$$

$$\iff \quad 1-\alpha/2 \sim F_{T_9}(1.084) \sim 0.84$$

$$\iff \quad \alpha/2 \sim 0.16 \quad \iff \quad 1-\alpha \sim 0.68$$

Dunque si ottiene una precisione relativa di $5 \cdot 10^{-2}$ con una fiducia di circa il $\boxed{68\,\%.}$

(ii) Vogliamo analizzare l'ipotesi nulla

$$H_0) \ \mu \leq \mu_0 = 4.3$$

contro l'alternativa $H_1) \ \mu > \mu_0 = 4.3$

Dobbiamo utilizzare il test per la media di un campione gaussiano con media e varianza ignote, dunque il p-value è

$$\bar{\alpha} = 1 - F_{T_{n-1}}\left(\frac{\sqrt{n}}{\sigma}(\bar{x}-\mu_0)\right) = 1 - F_{T_9}\left(\frac{\sqrt{10}}{0.7}(4.8-4.3)\right)$$

$$\sim 1 - F_{T_9}(2.259) \sim 1 - 0.974 \sim 0.026$$

Dunque $\boxed{\text{l'ipotesi è da scartare}}$.