Statistica - CPS
Corso di Laurea in Informatica
Compito del 27-10-2022

Esercizio 1. (10 punti) Nella città XY, si stima che, tra coloro che abitualmente utilizzano gli autobus pubblici, il 20% lo faccia sistematicamente senza pagare il biglietto. Tra coloro che non pagano il biglietto, il 70% sono individui di età inferiore a 15 anni, mentre tra coloro che lo pagano chi ha meno di 15 anni rappresenta il 40%.

(i) Si sceglie a caso un individuo (nella popolazione di coloro che utilizzano gli autobus); quanto vale la probabilità che si tratti di un individuo di età inferiore a 15 anni?

(ii) Se l'individuo scelto ha meno di 15 anni, quanto vale la probabilità che egli paghi abitualmente il biglietto?

(iii) Se l'individuo scelto ha 15 anni o più, quanto vale la probabilità che egli abitualmente non paghi il biglietto?

Esercizio 2. (10 punti) Sia p un numero reale fissato, con $0 \leq p \leq 1$, e si consideri la funzione f definita da

$$f(x) = \begin{cases} 0, & x \leq 0; \\ p, & 0 < x \leq 1; \\ (1-p), & 1 < x \leq 2; \\ 0, & x > 2. \end{cases}$$

(i) Dopo aver riconosciuto che la funzione sopra scritta è una densità di probabilità, scrivere l'espressione della funzione di ripartizione e la formula per il β-quantile (con $0 < \beta < 1$) per una v.a. X che abbia quella densità.

(ii) Calcolare i momenti primo e secondo di una v.a. X che abbia densità $f(x)$. Esaminare quando, al variare di p in $[0,1]$, la varianza è massima e quando è minima.

Esercizio 3. (10 punti) Una ditta produce una lozione per la ricrescita dei capelli ed afferma che si nota una ricrescita in almeno il 60% dei casi: tuttavia l'unione consumatori ha effettuato un'indagine ed ha rilevato che su 137 persone che hanno usato quella lozione solo 70 hanno notato una ricrescita dei capelli ed afferma che questa indagine contraddice l'affermazione della ditta.

In termini di un modello statistico, se p è la percentuale (sconosciuta) delle persone sulle quali la lozione ha effetti positivi, l'affermazione della ditta si traduce nell'ipotesi

$$H_0)\ p \geq 0.6 \quad \text{contro} \quad H_1)\ p < 0.6.$$

(i) Si può accettare, al livello 0.05, l'ipotesi sopra indicata (cioè l'affermazione della ditta)?

(ii) A quale livello (approssimativamente) si può accettare l'affermazione della ditta?

ESERCIZIO 1

Consideriamo il sistema di alternative

$B_1 = \{$ chi paga il biglietto $\}$

$B_2 = \{$ chi non paga il biglietto $\}$

e gli eventi $A = \{$ individui di età $< 15 \}$.

I dati del problema sono:

$\mathbb{P}(B_1) = 0.8$

$\mathbb{P}(B_2) = 0.2$

$\mathbb{P}(A \mid B_1) = 0.4$

$\mathbb{P}(A \mid B_2) = 0.7$.

(i) Dobbiamo calcolare $\mathbb{P}(A)$. Usando la formula di fattorizzazione, si ha

$$\mathbb{P}(A) = \mathbb{P}(A \mid B_1)\,\mathbb{P}(B_1) + \mathbb{P}(A \mid B_2)\,\mathbb{P}(B_2) =$$

$$= \frac{4}{10} \cdot \frac{8}{10} + \frac{7}{10} \cdot \frac{2}{10} = \frac{46}{100}$$

Quindi $\boxed{\mathbb{P}(A) = 0.46}$

(ii) La probabilità richiesta è $\mathbb{P}(B_1 \mid A)$. Usando la formula di Bayes

$$\mathbb{P}(B_1 \mid A) = \frac{\mathbb{P}(A \mid B_1)\,\mathbb{P}(B_1)}{\mathbb{P}(A)} = \frac{\frac{4}{10} \cdot \frac{8}{10}}{\frac{46}{100}} = \frac{32}{46}$$

Quindi $\boxed{\mathbb{P}(B_1 \mid A) = \frac{16}{23} \sim 0.696}$

(iii) La probabilità richiesta è $\mathbb{P}(B_2 \mid A^c)$. Usando la formula di Bayes

$$\mathbb{P}(B_2 \mid A^c) = \frac{\mathbb{P}(A^c \mid B_2)\ \mathbb{P}(B_2)}{\mathbb{P}(A^c)}$$

Usando che $\mathbb{P}(A^c) = 1 - \mathbb{P}(A) = 0.54$ e

$$\mathbb{P}(A^c \mid B_2) = 1 - \mathbb{P}(A \mid B_2) = 0.3$$

otteniamo

$$\mathbb{P}(B_2 \mid A^c) = \frac{\frac{3}{10} \cdot \frac{2}{10}}{\frac{54}{100}} = \frac{6}{54}$$

Quindi $\boxed{\mathbb{P}(B_2 \mid A^c) = \frac{1}{9} \sim 0.111}$

ESERCIZIO 2

Sia

$$f(x) = \begin{cases} 0, & x \le 0 \\ p, & 0 < x \le 1 \\ (1-p), & 1 < x \le 2 \\ 0, & x > 2 \end{cases} \qquad \text{con } p \in [0,1]$$

(i) La funzione $f(x)$ è una densità di probabilità perché:

- $f(x) \ge 0 \quad \forall\, x \in \mathbb{R}$;
- f è integrabile su $\mathbb{R}$ (essendo continua a meno di un numero finito di punti);
- $\int_{-\infty}^{+\infty} f(x)\, dx = \int_0^1 p\, dx + \int_1^2 (1-p)\, dx = 1$.

Se X è una v.a. con densità $f(x)$, la funzione di ripartizione è $F(x) = \mathbb{P}(X \le x) = \int_{-\infty}^{x} f(t)\, dx$. Si trova quindi

se $x \le 0$, $\quad F(x) = \int_{-\infty}^{x} f(t)\, dx = \int_{-\infty}^{x} 0\, dt = 0$

se $0 < x \le 1$, $\quad F(x) = \int_{-\infty}^{x} f(t)\, dx = F(0) + \int_0^x p\, dt = p x$

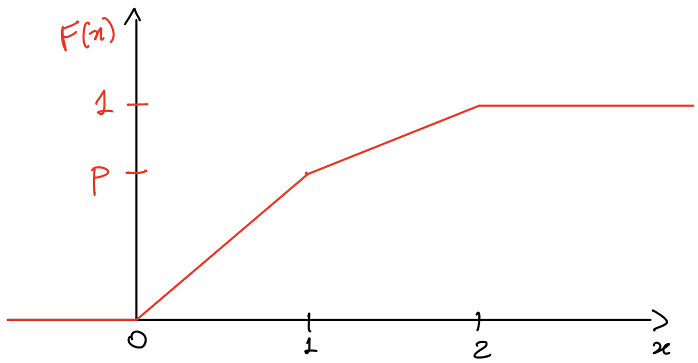
se $1 < x \le 2$, $\quad F(x) = \int_{-\infty}^{x} f(t)\,du = F(1) + \int_1^x (1-p)\,dt = p + (1-p)(x-1) =$

$= (1-p)x + 2p - 1$

se $x > 2$, $\quad F(x) = \int_{-\infty}^{x} f(t)\,du = F(2) + \int_2^x 0\,dt = 1.$

quindi
$$\boxed{F(x) = \begin{cases} 0, & x \le 0 \\ px, & 0 \le x \le 1 \\ (1-p)x + 2p - 1, & 1 \le x \le 2 \\ 1, & x \ge 2 \end{cases}}$$

Disegniamo il grafico di F.



Per trovare una formula per il β-quantile dobbiamo distinguere due casi:

se $\beta \in (0, p]$ allora r_β è l'unica soluzione di

$$F(r_\beta) = \beta \iff p r_\beta = \beta \iff \boxed{r_\beta = \frac{\beta}{p}}$$

se $\beta \in [p, 1)$ allora r_β è l'unica soluzione di

$$F(r_\beta) = \beta \iff (1-p)r_\beta + 2p - 1 = \beta \iff \boxed{r_\beta = \frac{\beta + 1 - 2p}{1-p}}$$

Questa è la formula nel caso $p \in (0,1)$. Se invece

- $p = 0$. Allora
$$F(x) = \begin{cases} 0, & x \le 1 \\ x-1, & 1 \le x \le 2 \\ 1, & x \ge 2 \end{cases}$$

e $\quad r_\beta = \beta + 1 \quad \forall \beta \in (0,1)$;

- $p = 1$ Allora
$$F(x) = \begin{cases} 0, & x \le 0 \\ x, & 0 \le x \le 1 \\ 1, & x \ge 1 \end{cases}$$

e $\quad r_\beta = \beta \quad \forall \beta \in (0,1)$

(ii) Se X è una v.a. con densità $f(x)$ si ha

momento primo $= \boxed{E[X]} = \int_{-\infty}^{+\infty} x f(x)\,dx = \int_0^1 px\,dx + \int_1^2 (1-p)x\,dx =$

$$= \frac{1}{2}px^2\Big|_0^1 + \frac{1}{2}(1-p)x^2\Big|_1^2 \boxed{= \frac{3}{2} - p}$$

momento secondo $= \boxed{E[X^2]} = \int_{-\infty}^{+\infty} x^2 f(x)\,dx = \int_0^1 px^2\,dx + \int_1^2 (1-p)x^2\,dx =$

$$= \frac{1}{3}px^3\Big|_0^1 + \frac{1}{3}(1-p)x^3\Big|_1^2 \boxed{= \frac{7}{3} - 2p}$$

Per la varianza troviamo

$$\mathrm{Var}(X) = E[X^2] - (E[X])^2 = \frac{7}{3} - 2p - \left(\frac{3}{2} - p\right)^2 = \frac{1}{12} + p - p^2$$

e poiché $\max_{p \in [0,1]} (p - p^2) = \frac{1}{4}$, $\min_{p \in [0,1]} (p - p^2) = 0$, abbiamo che

$$\boxed{\max_{p \in [0,1]} \mathrm{Var}(X) = \frac{1}{3}} \quad \text{e} \quad \boxed{\min_{p \in [0,1]} \mathrm{Var}(X) = \frac{1}{12}}$$

Il risultato è coerente con l'interpretazione intuitiva della varianza. Il massimo si ottiene infatti per $p = \frac{1}{2}$, che corrisponde al caso in cui la densità è uniforme su $[0,2]$, mentre il minimo si ottiene per $p = 0$ oppure $p = 1$, che corrispondono ai due casi in cui la densità è positiva solo su $[0,1]$ o $[1,2]$.

ESERCIZIO 3 | In termini di un modello statistico, possiamo considerare di avere un campione statistico $X_1, \dots, X_n$ di $n = 137$ v.a. con distribuzione $B(1, p)$, su cui facciamo l'ipotesi

$$H_0) \quad p \geq p_0 = 0.6$$

(i) Fissato il livello $\alpha = 0.05$, la regione critica del test è

$$C = \left\{ \overline{X} - p_0 < \frac{\sqrt{p_0(1-p_0)}}{\sqrt{n}} q_\alpha \right\} = \left\{ \overline{X} - 0.6 < \frac{\sqrt{24}}{10\sqrt{137}} q_{0.05} \right\}$$

e l'ipotesi è accettabile al livello $\alpha = 0.05$ se e solo se

$$\bar{x} - p_0 \geq \frac{\sqrt{p_0(1-p_0)}}{\sqrt{n}} q_\alpha \iff \bar{x} - 0.6 \geq \frac{\sqrt{24}}{10\sqrt{137}} q_{0.05}$$

dove $\bar{x} = \frac{70}{137}$ e $q_{0.05} = -q_{0.95} \sim -1.645$

Poiché $\bar{x} - 0.6 \sim -0.089$ e $-\frac{\sqrt{24}}{10\sqrt{137}} 1.645 \sim -0.069$

l'ipotesi $p \geq 0.06$ non si può accettare al livello 0.05.

(ii) Per trovare il livello a cui l'ipotesi diventa accettabile, dobbiamo calcolare il p-value dei dati associato al test. La formula del p-value in questo caso è

$$\bar{\alpha} = \Phi\left(\frac{\sqrt{n}}{\sqrt{p_0(1-p_0)}} (\bar{x} - p_0) \right)$$

Troviamo quindi in questo caso

$$\bar{\alpha} = \Phi\left(\frac{\sqrt{137}}{\sqrt{0.6 \cdot 0.4}} \left(\frac{70}{137} - 0.6 \right) \right) \sim \Phi(-2.126) =$$

$$= 1 - \Phi(2.126) \sim 1 - 0.9832 \sim 0.017 .$$

L'affermazione della ditta è quindi accettabile per livelli $\alpha \leq 0.017$, e l'affermazione risulta non molto plausibile.