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

Esercizio 1. (10 punti) Un sacchetto contiene 10 monete, di cui 8 "oneste" (ossia una faccia è Testa e una faccia Croce, e nel lancio della moneta le due facce sono equiprobabili) e 2 con entrambe le facce uguali a Testa. L'esperimento consiste nell'estrarre una moneta dal sacchetto e lanciarla, il successo è rappresentato dall'esito Testa. Dopo ogni lancio, la moneta estratta viene reinserita nel sacchetto, e l'esperimento continua fino al primo successo.

(i) Determinare la probabilità che durante l'esperimento vengano eseguiti k lanci, con $k \in \mathbb{N}$.

(ii) Sapendo che l'esito Testa non si è avuto nei primi 10 lanci, determinare la probabilità di successo in meno di 13 lanci.

(iii) Supponiamo che al primo lancio sia uscita Testa, qual è la probabilità che sia stata estratta dal sacchetto una moneta onesta?

Esercizio 2. (10 punti) Consideriamo una variabile aleatoria X la cui densità è la funzione

$$f(x) = \begin{cases} 3x^{-4}, & \text{se } x > 1; \\ 0, & \text{se } x \leq 1. \end{cases}$$

(i) Dire quali momenti possiede X, e calcolare (se sono finite) speranza e varianza.

(ii) Si consideri la v.a. $Y = \log(X)$: dire se Y ha densità (ed in tal caso calcolarla) e calcolare la speranza di Y.

Esercizio 3. (10 punti) Vengono effettuate 25 misurazioni del peso di un piccolo mammifero africano, e si ottengono i risultati $(x_1, \dots, x_{25})$ (espressi in grammi) dai quali si ricava una media campionaria eguale a 648.6 ed una varanza campionaria eguale a 457.66 .

Si vuole verificare l'ipotesi che il peso medio di questo mammifero sia di 640 grammi, e a tal fine si suppone che i dati possano essere rappresentati con 25 variabili gaussiane indipendenti.

(i) Effettuare il test dell'ipotesi

$$H_0\big)\ m = 640 \qquad \text{contro} \qquad H_1\big)\ m \neq 640$$

supponendo che la varianza sia nota ed eguale a 400: dopo aver calcolato il p-value, che cosa si conclude?

(ii) Effettuare lo stesso test sopra scritto, supponendo però che la varianza sia sconosciuta: che cosa si conclude?

Esercizio 1

Calcoliamo la probabilità di ottenere Testa in ogni singolo lancio.

Consideriamo gli eventi $A = \{\text{si estrae una moneta onesta}\}$ e

$B = \{\text{si estrae una moneta truccata}\}$.

Allora

$$\mathbb{P}(\text{Testa}) = \mathbb{P}(\text{Testa} \mid A)\,\mathbb{P}(A) + \mathbb{P}(\text{Testa} \mid B)\,\mathbb{P}(B).$$

e usando $\mathbb{P}(A) = \frac{8}{10} = \frac{4}{5}$, $\mathbb{P}(B) = \frac{2}{10} = \frac{1}{5}$,

$\mathbb{P}(\text{Testa} \mid A) = \frac{1}{2}$, $\mathbb{P}(\text{Testa} \mid B) = 1$

si ha

$$\mathbb{P}(\text{Testa}) = \frac{1}{2} \cdot \frac{4}{5} + 1 \cdot \frac{1}{5} = \frac{3}{5}.$$

(i) Sia X la v.a. che conta il numero di lanci da effettuare per ottenere il primo successo.

X è una v.a. geometrica di parametro $p = 3/5$, quindi

$$\boxed{\mathbb{P}(X = k) = (1-p)^{k-1} \cdot p = \frac{2^{k-1} \cdot 3}{5^k} \qquad \forall\, k \geq 1.}$$

(ii) Una v.a. geometrica ha la proprietà detta "assenza di memoria", ossia per ogni $n, t \geq 1$ si ha $\mathbb{P}(X = n+t \mid X > n) = \mathbb{P}(X = t)$.

Quindi

$$\boxed{\begin{aligned} \mathbb{P}(X < 13 \mid X > 10) &= \mathbb{P}(X = 11 \mid X > 10) + \mathbb{P}(X = 12 \mid X > 10) = \\ &= \mathbb{P}(X = 10+1 \mid X > 10) + \mathbb{P}(X = 10+2 \mid X > 10) = \\ &= \mathbb{P}(X = 1) + \mathbb{P}(X = 2) = \frac{3}{5} + \frac{2 \cdot 3}{25} = \frac{21}{25} \end{aligned}}$$

(iii) Vogliamo calcolare $\mathbb{P}(A \mid X = 1)$, o in maniera equivalente

$\mathbb{P}(A \mid \text{Testa})$. Utilizzando la formula di Bayes si ha

$$\mathbb{P}(A \mid \text{Testa}) = \frac{\mathbb{P}(\text{Testa} \mid A) \cdot \mathbb{P}(A)}{\mathbb{P}(\text{Testa})}$$

e usiamo $\mathbb{P}(A) = \frac{4}{5}$, $\mathbb{P}(\text{Testa}) = \frac{3}{5}$, $\mathbb{P}(\text{Testa} \mid A) = \frac{1}{2}$, per ottenere

$$\boxed{\mathbb{P}(A \mid \text{Testa}) = \frac{\frac{1}{2} \cdot \frac{4}{5}}{\frac{3}{5}} = \frac{2}{3}.}$$

Esercizio 2 Sia data la densità

$$f(x) = \begin{cases} 3x^{-4}, & \text{se } x > 1 \\ 0, & \text{se } x \leq 1 \end{cases}$$

di una v.a. X.

(i) Il momento k-esimo, con $k \geq 1$, di X è

$$E[X^k] = \int_{-\infty}^{+\infty} x^k f(x)\,dx = \int_1^{+\infty} 3x^{k-4}\,dx < +\infty \iff k-4 < -1$$
$$\iff k < 3.$$

Dunque $\boxed{E[X^k] < +\infty \iff k \in \{1, 2\}}$

Si calcola dunque che

$$E[X] = \int_1^{+\infty} 3x^{-3}\,dx = -\frac{3}{2}x^{-2}\Big|_1^{+\infty} = \frac{3}{2}$$

$$Var(X) = E[X^2] - (E[X])^2 = \int_1^{+\infty} 3x^{-2}\,dx - \frac{9}{4} = -3x^{-1}\Big|_1^{+\infty} - \frac{9}{4} =$$
$$= 3 - \frac{9}{4} = \frac{3}{4}$$

(ii) Data la funzione $h(t) = \log t$ definita per $t \in (0, +\infty)$, si ha $Y = h(X)$. Poiché X assume valori in $(1,+\infty)$, la Y è ben definita. Inoltre h è su $(1,+\infty)$ una funzione invertibile con inversa $h^{-1}(s) = e^s$ derivabile. Quindi Y ha densità e possiamo scrivere

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

e allora

$$\boxed{f_Y(y) = \begin{cases} 3e^{-4y} e^y = 3e^{-3y} & , \text{ se } y > 0 \\ 0 & , \text{ se } y \le 0 \end{cases}}$$

La speranza di Y si può calcolare come

$$E[Y] = E[\log(X)] = \int_{-\infty}^{+\infty} (\log x) \cdot f(x)\, dx = \int_1^{+\infty} 3x^{-4} \log x \, dx$$

oppure come

$$E[Y] = \int_{-\infty}^{+\infty} y\, f_Y(y)\, dy = \int_0^{+\infty} 3y e^{-3y}\, dy$$

Quindi, usando

$$\int_0^{+\infty} 3y e^{-3y}\, dy = 3 \; y\left(-\frac{1}{3}e^{-3y}\right)\Big|_0^{+\infty} - 3\int_0 1 \cdot \left(-\frac{1}{3}e^{-3y}\right) dy =$$

$$= \int_0^{+\infty} e^{-3y}\, dy = -\frac{1}{3}e^{-3y}\Big|_0^{+\infty} = \frac{1}{3},$$

si ha $\boxed{E[Y] = \frac{1}{3}}$

<u>Esercizio 3</u> Abbiamo $X_1, \ldots, X_{25}$ v.a. gaussiane indipendenti in un campione statistico, con media campionaria $\bar{x} = 648.6$ e varianza campionaria $\overline{\sigma^2} = 457.66$.

(i) Supponiamo che la varianza delle $\{X_k\}$ sia nota e $\sigma^2 = 400$, e si vuole effettuare il Test con ipotesi nulla $H_0)$ $m = 640$ e alternativa $H_1)$ $m \neq 640$.

Effettuiamo quindi un Test sulla media di un campione gaussiano con varianza nota, per il quale il p-value si calcola come

$$\bar{\alpha} = 2\left[1 - \Phi\left(\frac{\sqrt{n}}{\sigma}|\bar{x} - m_0|\right)\right]$$

dove $n = 25$, $\sigma = 20$, $\bar{x} = 648.6$ e $m_0 = 640$. Quindi

$$\boxed{\begin{aligned}\bar{\alpha} &= 2\left[1 - \Phi\left(\frac{\sqrt{25}}{20}|648.6 - 640|\right)\right] = 2\left[1 - \Phi\left(\frac{8.6}{4}\right)\right] = \\ &= 2\left[1 - \Phi(2.15)\right] \sim 2\left[1 - 0.98422\right] \sim 0.03\end{aligned}}$$

e l'ipotesi risulta poco plausibile.

(ii) Se si considera sconosciuta la varianza, utilizziamo il Test sulla media di un campione gaussiano con varianza sconosciuta, per il quale il p-value si calcola come

$$\bar{\alpha} = 2\left[1 - F_{n-1}\left(\frac{\sqrt{n}}{\bar{\sigma}}|\bar{x} - m_0|\right)\right]$$

dove $n = 25$, $\bar{\sigma} = \sqrt{457.66} \sim 21.393$, $\bar{x} = 648.66$, $m_0 = 640$. Quindi

$$\boxed{\begin{aligned}\bar{\alpha} &\sim 2\left[1 - F_{24}\left(\frac{\sqrt{25}}{21.393}|648.66 - 640|\right)\right] \sim 2\left[1 - F_{24}(2.02)\right] \sim \\ &\sim 2\left[1 - 0.972\right] \sim 0.056\end{aligned}}$$

e l'ipotesi va considerata non molto plausibile.