Statistica - CPS
Corso di Laurea in Informatica
Compito del 30-05-2022

Esercizio 1. (10 punti) Un'urna contiene 5 palline, di cui 4 rosse e 1 blu. Si consideri il seguente esperimento. Si estrae una prima pallina dall'urna. Se questa prima pallina estratta è blu, l'esperimento termina. Se invece la prima pallina estratta è rossa, la pallina non viene reinserita nell'urna, ma se ne aggiunge un'altra che è rossa con probabilità $3/4$ e blu con probabilità $1/4$, e successivamente si estrae nuovamente una pallina dall'urna. Dopo la seconda estrazione l'esperimento termina in ogni caso.

(i) Qual è la probabilità che durante l'esperimento venga estratta una pallina blu? Ripetendo 10 volte l'esperimento dall'inizio, determinare la probabilità di estrarre una pallina blu per 3 volte.

(ii) Sapendo che la pallina blu è stata estratta alla seconda estrazione, qual è la probabilità che la pallina inserita nell'urna dopo la prima estrazione fosse blu?

Esercizio 2. (10 punti) Si consideri la funzione

$$f_a(x) = \frac{1}{a\sqrt{2\pi}}\, e^{-\frac{(x-1)^2}{2a^2}}$$

con parametro $a \in (0, +\infty)$.

(i) Sia X una v.a. con densità $f_a(x)$, che tipo di v.a. è X? Determinare $a \in (0, +\infty)$ in modo che $Var(X) = 4$.

(ii) Sia Y una v.a. con densità $N(1, 5)$ e indipendente da X. Usando $a = 2$, scrivere la densità della v.a. $Z = (X - Y)^2$.

(iii) Siano $X_1, \dots, X_{100}$ v.a. indipendenti e con densità $f_a(x)$ con $a = 2$. Trovare un'approssimazione per $\mathbb{P}(X_1 + \dots X_{100} > 90)$.

Esercizio 3. (10 punti) Una ditta produce chiodi di lunghezza dichiarata uguale a 5 cm, e il proprietario della ditta afferma che la deviazione standard delle lunghezze dei chiodi prodotti non supera 0.2 cm.
Analizzando la lunghezza di un campione di 16 pezzi si trova media campionaria di 4.935 cm e varianza campionaria 0.06 cm.

(i) Supponendo che le lunghezze dei chiodi possano essere rappresentate con variabili aleatorie gaussiane, si può accettare al livello 0.05 l'affermazione del proprietario della ditta sulla deviazione standard della lunghezza dei chiodi prodotti?

(ii) Descrivere un test da utilizzare per verificare l'ipotesi che la lunghezza dei chiodi prodotti sia di 5 cm, e usare i dati del campione analizzato per determinare la plausibilità di questa ipotesi.

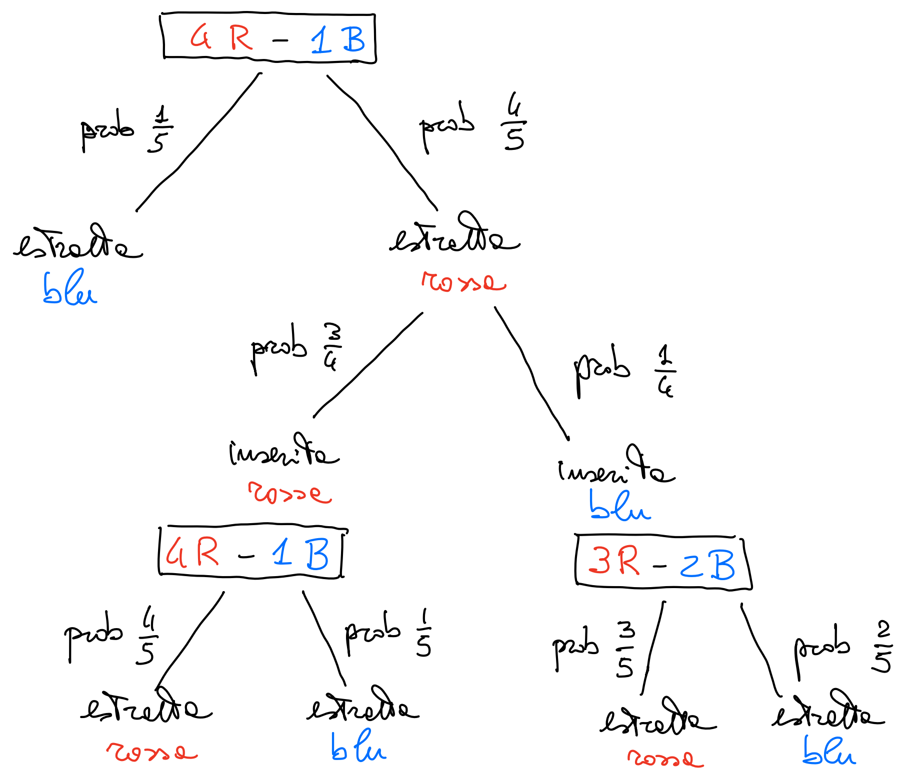
Esercizio 1 Un'urna contiene cinque palline, di cui 4 rosse e 1 blu. Si consideri il seguente esperimento.

Si estrae una prima pallina dall'urna. Se questa prima pallina estratta è blu, l'esperimento termina. Se invece la prima pallina estratta è rossa, la pallina non viene reinserita nell'urna, ma se ne aggiunge un'altra che è rossa con probabilità $3/4$ e blu con probabilità $1/4$, e successivamente si estrae nuovamente una pallina dall'urna. Dopo la seconda estrazione l'esperimento termina in ogni caso.

(1) Qual è la probabilità che durante l'esperimento venga estratta una pallina blu? Ripetendo 10 volte l'esperimento dall'inizio, determinare la probabilità di estrarre una pallina blu per 3 volte.

$\longleftrightarrow$

Utilizziamo un diagramma per rappresentare l'esperimento e le probabilità di ciascuna possibilità.



Le probabilità lungo ciascun ramo si possono moltiplicare per la formula del condizionamento ripetuto, quindi

$\mathbb{P}(\text{estrarre blu}) = \mathbb{P}(\text{prima estratta blu}) + \mathbb{P}(\text{seconda estratta blu})$

dove

$\mathbb{P}(\text{prima estratta blu}) = \frac{1}{5}$

$\mathbb{P}(\text{seconda estratta blu}) = \mathbb{P}(\{\text{seconda estratta blu}\} \cap \{\text{inserita rossa}\} \cap \{\text{prima estratta rossa}\}) + \mathbb{P}(\{\text{seconda estratta blu}\} \cap \{\text{inserita blu}\} \cap \{\text{prima estratta rossa}\}) =$

$= \frac{1}{5} \cdot \frac{3}{4} \cdot \frac{4}{5} + \frac{2}{5} \cdot \frac{1}{4} \cdot \frac{4}{5} = \frac{3}{25} + \frac{2}{25} = \frac{1}{5}$

$\Longrightarrow \boxed{\mathbb{P}(\text{estrarre blu}) = \frac{1}{5} + \frac{1}{5} = \frac{2}{5}}$

Siano $X_1, \dots, X_{10}$ v.a. che descrivono l'estrazione di una pallina blu durante la ripetizione dell'esperimento. Quindi le $\{X_i\}$ sono v.a. indipendenti ed equidistribuite, e se poniamo $X_i = 1$ se all'i-esima ripetizione viene estratta una pallina blu, e $X_i = 0$ altrimenti, otteniamo che

- $X_i \sim B(1, \frac{2}{5})$
- la v.a. $Y = X_1 + \dots + X_{10}$ conta il numero di volte in cui viene estratta la pallina blu, quindi $Y \sim B(10, \frac{2}{5})$ e

$$\boxed{\mathbb{P}(Y=3) = \binom{10}{3}\left(\frac{2}{5}\right)^3 \cdot \left(\frac{3}{5}\right)^7 = \frac{10 \cdot 9 \cdot 8}{3 \cdot 2} \cdot \frac{2^3 \cdot 3^7}{5^{10}} = \frac{2^6 \cdot 3^8}{5^9} \sim 0.215}$$

(ii) Sapendo che la pallina blu è stata estratta alla seconda estrazione, qual è la probabilità che la pallina inserita nell'urna dopo la prima estrazione fosse blu?

$\longleftrightarrow$

Siamo nella situazione $\boxed{3R - 1B}$, inseriamo una pallina nell'urna e ne estraiamo una blu. Indichiamo con $\{B_1, B_2\}$ un sistema di alternative date da

$$B_1 = \{\text{inseriamo pallina rossa}\}, \quad B_2 = \{\text{inseriamo pallina blu}\}$$

di cui sappiamo che $\mathbb{P}(B_1) = 3/4$, $\mathbb{P}(B_2) = 1/4$.

Sia A l'evento $A = \{\text{seconda pallina estratta blu}\}$, allora per la formula di probabilità delle cause

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

con $\mathbb{P}(A|B_1) = \frac{1}{5}$, $\mathbb{P}(A|B_2) = \frac{2}{5}$, si ha

$$\boxed{\mathbb{P}(B_2|A) = \frac{\frac{2}{5}\cdot\frac{1}{4}}{\frac{1}{5}\cdot\frac{3}{4} + \frac{2}{5}\cdot\frac{1}{4}} = \frac{2}{3+2} = \frac{2}{5}}$$

Esercizio 2 Si consideri la funzione

$$f_a(x) = \frac{1}{a\sqrt{2\pi}}\, e^{-\frac{(x-1)^2}{2a^2}}$$

con parametro $a \in (0, +\infty)$.

(i) Sia X una v.a. con densità $f_a(x)$, che tipo di v.a. è X?

Determinare $a \in (0,+\infty)$ in modo che $\mathrm{Var}(X) = 4$.

$\longleftrightarrow$

La funzione $f_a(x)$, $a \in (0,+\infty)$, è la densità Gaussiana $N(1, a^2)$ di valor medio 1, e varianza a^2. Quindi $\boxed{X \sim N(1, a^2)}$, e chiaramente $\mathrm{Var}(X) = 4 \iff \boxed{a = 2}$

(ii) Sia Y una v.a. con densità $N(1,5)$ e indipendente da X. Usando $a=2$, scrivere la densità della v.a. $Z=(X-Y)^2$.

Poiché $X \sim N(1,4)$ e $Y \sim N(1,5)$ sono indipendenti, si trova $-Y \sim N(-1,5)$ e quindi $X-Y \sim N(0,9)$.

Se $Z=(X-Y)^2$, possiamo scrivere $Z=h(X-Y)$ con $h(t)=t^2$, ma la funzione $h(t)$ non è invertibile su $\mathrm{Imm}(X-Y)=\mathbb{R}$. Possiamo quindi attraverso la funzione di ripartizione.

$$F_Z(z)=\mathbb{P}(Z \le z)=\mathbb{P}((X-Y)^2 \le z)=\begin{cases} 0, & \text{se } z<0 \\ \mathbb{P}(-\sqrt{z} \le X-Y \le \sqrt{z}), & \text{se } z \ge 0 \end{cases}$$

Inoltre, ponendo $V \sim N(0,1)$, si ha $X-Y=3V$, quindi

$$\mathbb{P}(-\sqrt{z} \le X-Y \le \sqrt{z}) = \mathbb{P}\left(-\frac{\sqrt{z}}{3} \le V \le \frac{\sqrt{z}}{3}\right) = \Phi\left(\frac{\sqrt{z}}{3}\right)-\Phi\left(-\frac{\sqrt{z}}{3}\right) =$$

$$= \Phi\left(\frac{\sqrt{z}}{3}\right)-\left(1-\Phi\left(\frac{\sqrt{z}}{3}\right)\right) = 2\Phi\left(\frac{\sqrt{z}}{3}\right)-1$$

Allora

$$F_Z(z)=\begin{cases} 0, & \text{se } z<0 \\ 2\Phi\left(\frac{\sqrt{z}}{3}\right)-1, & \text{se } z \ge 0 \end{cases}$$

e quindi la densità $f_Z(z)$ soddisfa $f_Z(z)=F_Z'(z)$ $\forall z \ne 0$, ossia

$$\boxed{f_Z(z)=\begin{cases} 0, & \text{se } z \le 0 \\ 2\varphi\left(\frac{\sqrt{z}}{3}\right)\frac{1}{6\sqrt{z}} = \frac{1}{3\sqrt{2\pi}}\frac{1}{\sqrt{z}}e^{-\frac{z}{18}}, & \text{se } z>0 \end{cases}}$$

Allo stesso risultato si poteva arrivare osservando che $X - Y \sim N(0,9)$ implica che $(X-Y)^2 = 9\, C_1$, dove $C_1 \sim \chi^2(1)$ ha densità $\Gamma(\frac{1}{2},\frac{1}{2})$, e usando poi $(X-Y)^2 = g(C_1)$ con $g(t) = 9t$ e la formula per il calcolo della densità nel caso di una funzione g invertibile e derivabile.

$\longleftrightarrow$

(iii) Siano $X_1, \dots, X_{100}$ v.a. indipendenti e con densità $f_a(x)$ con $a=2$. Trovare un'approssimazione per $\mathbb{P}(X_1 + \dots + X_{100} > 90)$

$\longleftrightarrow$

Poiché $X_i \sim N(1,4)$ $\forall i = 1,\dots,100$ e le v.a. sono indipendenti, si ha $X_1 + \dots + X_{100} \sim N(100, 400)$. Quindi se $V \sim N(0,1)$ possiamo scrivere $X_1 + \dots + X_{100} = 20V + 100$ e

$$\boxed{\begin{aligned} \mathbb{P}(X_1 + \dots + X_{100} > 90) &= \mathbb{P}(20V + 100 > 90) = \mathbb{P}\left(V > \frac{90-100}{20}\right) = \\ &= \mathbb{P}(V > -0.5) = 1 - \Phi(-0.5) = \Phi(0.5) \sim 0.69146 \end{aligned}}$$

<u>Esercizio 3)</u> Una ditta produce chiodi di lunghezza dichiarata uguale a 5 cm, e il proprietario della ditta afferma che la deviazione standard delle lunghezze dei chiodi prodotti non supera 0.2 cm. Analizzando un campione di 16 pezzi si trova media campionaria di 4.935 cm e varianza campionaria 0.06 cm.

(i) Supponendo che le lunghezze dei chiodi possano essere rappresentate con variabili aleatorie gaussiane, si può accettare al livello 0.05 l'affermazione del proprietario della ditta sulla deviazione standard della lunghezza dei chiodi prodotti?

Vogliamo testare l'ipotesi $H_0)\ \sigma^2 \le \sigma_0^2 = (0.2)^2 = 0.04$ contro l'alternativa $H_1)\ \sigma^2 > 0.04$.

La regione critica del test al livello α è

$$C = \left\{ \frac{(n-1)S^2}{\sigma_0^2} > \chi^2_{(1-\alpha,\, n-1)} \right\}$$

dove n è la taglia del campione.

Ponendo $\alpha = 0.05$, $n = 16$, $\sigma_0^2 = 0.04$ e $\bar{\sigma}^2 = 0.06$, l'ipotesi è accettabile al livello 0.05 se e solo se

$$\frac{(n-1)\,\bar{\sigma}^2}{\sigma_0^2} = \frac{15 \cdot 0.06}{0.04} \le \chi^2_{(0.95,\, 15)} \sim 24.9958$$

Poiché $\frac{15 \cdot 0.06}{0.04} = \frac{45}{2} = 22.5$, $\boxed{\text{l'ipotesi è accettabile.}}$

(ii) Descrivere un test da utilizzare per verificare l'ipotesi che la lunghezza dei chiodi sia di 5 cm, e usare i dati del campione analizzato per determinare la plausibilità di questa ipotesi.

Adesso vogliamo testare l'ipotesi $H_0)\ \mu = \mu_0 = 5$ per la media della v.a. che descrivono la lunghezza dei chiodi, contro l'alternativa $H_1)\ \mu \ne 5$.

Poiché la varianza della v.a. è ignota, usiamo la regione critica dato al livello α da

$$C = \left\{ \left| \frac{\sqrt{n}}{S} \left(\bar{X} - \mu_0 \right) \right| > t_{(1-\alpha/2,\, n-1)} \right\}$$

dove n è la taglia del campione. Ponendo $n=16$, $\bar{x} = 4.935$, $\bar{\sigma} = \sqrt{0.06}$, l'ipotesi è accettabile al livello α se e solo se

$$\frac{\sqrt{16}}{\sqrt{0.06}} \left| 4.935 - 5 \right| \leq t_{(1-\alpha/2,\, 15)}$$

Per verificare la plausibilità dell'ipotesi, calcoliamo il p-value che in questo caso è dato da

$$\bar{\alpha} = 2\left[1 - F_{T(n-1)}\left(\frac{\sqrt{n}}{\bar{\sigma}} \left| \bar{x} - \mu_0 \right| \right) \right] =$$

$$= 2\left[1 - F_{T(15)}\left(\sqrt{\frac{16}{0.06}} \left| 4.935 - 5 \right| \right) \right] \sim 2\left[1 - F_{T(15)}\left(1.061 \right) \right]$$

$$\sim 2\left[1 - 0.84 \right] \sim 0.32$$

Il p-value è maggiore di 0.3 e dunque l'ipotesi è molto plausibile.