Soluzioni Esame di Calcolo delle Probabilità e Statistica [2959]

Corso di Studi di Ingegneria Gestionale (D.M.270/04) (L)

Dipartimento di Meccanica, Matematica e Management
Politecnico di Bari

A.A.: 2020/2021
Docente: Gianluca Orlando
Appello: VII Appello
Data: 27/09/2021

Esercizio 1. Due candidati alle elezioni, T e C, competono per la presidenza di un piccolo paese di 5 abitanti (T e C inclusi). Da un sondaggio risulta che:

- T (elettore 1) voterà per T;
- C (elettore 2) voterà per C;
- l'elettore 3 voterà per C;
- gli elettori 4 e 5 discutono di politica e sono ancora indecisi: si sa che ciascuno di loro voterà per T con il 40% di probabilità. Ma il voto dell'elettore 4 influenza il voto dell'elettore 5: se 4 decide di votare per C, 5 farà lo stesso con $\frac{2}{3}$ di probabilità.

Rispondere ai seguenti quesiti, descrivendo lo spazio campione.

1. Qual è la probabilità che C vinca le elezioni?
2. Qual è la probabilità che C ottenga esattamente il 40% dei voti degli elettori nel paese?
3. Sono finite le elezioni. Supponiamo di sapere che l'elettore 5 abbia votato per T, ma di non conoscere ancora il voto dell'elettore 4. Con che probabilità l'elettore 4 ha votato per C?

Soluzione. Lo spazio campione è dato dalle possibili coppie formate dal voto dell'elettore 4 e del voto dell'elettore 5, ovvero

$$\Omega = \{(C,C),(T,T),(T,C),(C,T)\}\,.$$

Scriviamo le ipotesi sugli elettori 4 e 5 in termini di eventi. Chiamiamo

$$A = \{\text{“4 vota per } C\text{”}\} = \{(C,C),(C,T)\}$$

e osserviamo che $\mathbf{P}(A) = 1 - \mathbf{P}(\text{“4 vota per } T\text{”}) = \frac{3}{5}$. Inoltre

$$B = \{\text{“5 vota per } C\text{”}\} = \{(T,C),(C,C)\}\,.$$

e osserviamo che $\mathbf{P}(B) = 1 - \mathbf{P}(\text{“5 vota per } T\text{”}) = \frac{3}{5}$. L'ipotesi sull'influenza del voto di 4 su 5 diventa allora $\mathbf{P}(B|A) = \frac{2}{3}$.

Calcoliamo le probabilità degli eventi elementari. Per la definizione di probabilità condizionata abbiamo che:

$$\mathbf{P}(C,C) = \mathbf{P}(A \cap B) = \mathbf{P}(A)\mathbf{P}(B|A) = \frac{3}{5}\frac{2}{3} = \frac{2}{5}$$

$$\mathbf{P}(C,T) = \mathbf{P}(A \setminus \{(C,C)\}) = \mathbf{P}(A) - \mathbf{P}(C,C) = \frac{3}{5} - \frac{2}{5} = \frac{1}{5}$$

$$\mathbf{P}(T,C) = \mathbf{P}(B \setminus \{(C,C)\}) = \mathbf{P}(B) - \mathbf{P}(C,C) = \frac{3}{5} - \frac{2}{5} = \frac{1}{5}$$

e quindi

$$\mathbf{P}(T,T) = 1 - \mathbf{P}(C,C) - \mathbf{P}(C,T) - \mathbf{P}(T,C) = 1 - \frac{2}{5} - \frac{1}{5} - \frac{1}{5} = \frac{1}{5}\,.$$

1. Osserviamo che C vince se ottiene almeno un voto. Quindi

$$\mathbf{P}(\text{“}C \text{ vince le elezioni”}) = \mathbf{P}(\{(C,C),(C,T),(T,C)\}) = \frac{2}{5} + \frac{1}{5} + \frac{1}{5} = \frac{4}{5} = 80\%\,.$$

Questo risponde alla prima domanda.

2. Poiché il 40% di 5 è $\frac{40}{100}5 = 2$, la probabilità che C ottenga il 40% dei voti coincide con la probabilità che T vinca. Quindi $\mathbf{P}(\text{“}C \text{ ottiene il 40\% dei voti”}) = 1 - 80\% = 20\%$. Osserviamo che questo è uguale a $\mathbf{P}(T,T)$ perché T vince solo se ottiene 2 voti.

3. Ci viene chiesta la probabilità:

$$\mathbf{P}(A|\overline{B}) = \frac{\mathbf{P}(A \setminus B)}{(1 - \mathbf{P}(B))} = \frac{\mathbf{P}(A \setminus (A \cap B))}{(1 - \mathbf{P}(B))} = \frac{\mathbf{P}(C,T)}{(1 - \mathbf{P}(B))} = \frac{1/5}{2/5} = \frac{1}{2}\,.$$

Esercizio 2. In una certa regione, i terremoti si susseguono secondo un processo di Poisson con una media di 5 terremoti all'anno.

1. Qual è la probabilità che si verifichi almeno 1 terremoto nella prima metà del 2022?
2. Assumendo che si sia verificato almeno 1 terremoto nella prima metà del 2022, qual è la probabilità che si verifichino esattamente 3 terremoti nei primi 9 mesi del 2022?

Soluzione. È opportuno considerare la media di terremoti al mese $\lambda = \frac{5}{12}$. Allora la v.a. X data dal numero di terremoti in un mese è di Poisson con parametro $\lambda = \frac{5}{12}$ e la sua funzione di probabilità è data da

$$\mathbf{P}(X = i) = \frac{\lambda^i}{i!}e^{-\lambda}\,, \quad i = 0, 1, 2, \ldots$$

Consideriamo la v.a. $X = X_1 + X_2 + \cdots + X_6$, dove X_h è il numero di terremoti nell'h-esimo mese del 2022. Per la riproducibilità della distribuzione di Poisson, X è una v.a. di Poisson con parametro $6\lambda = \frac{5}{2}$.

1. Possiamo calcolare

$$\begin{aligned}\mathbf{P}(X \geq 1) &= 1 - \mathbf{P}(X < 1) = 1 - \mathbf{P}(X = 0)\\ &= 1 - e^{-\frac{5}{2}} \sim 0.918\end{aligned}$$

2. Consideriamo ora anche X_7, X_8, X_9 date dal numero di terremoti nei mesi 7, 8 e 9. Per comodità chiamiamo $Y = X_7 + X_8 + X_9$, che è una v.a. di Poisson con parametro $3\lambda = \frac{5}{4}$. Siamo interessati alla probabilità

$$\begin{aligned}
&\mathbf{P}(X_1 + \cdots + X_9 = 3 | X_1 + \cdots + X_6 \geq 1) \\
&= \mathbf{P}(X + Y = 3 | X \geq 1) = \frac{\mathbf{P}(X + Y = 3\,, X \geq 1)}{\mathbf{P}(X \geq 1)} \\
&= \frac{\mathbf{P}(Y = 2\,, X = 1) + \mathbf{P}(Y = 1\,, X = 2) + \mathbf{P}(Y = 0\,, X = 3)}{\mathbf{P}(X \geq 1)} \\
&= \frac{\mathbf{P}(Y = 2)\mathbf{P}(X = 1) + \mathbf{P}(Y = 1)\mathbf{P}(X = 2) + \mathbf{P}(Y = 0)\mathbf{P}(X = 3)}{\mathbf{P}(X \geq 1)} \\
&= \frac{1}{1 - e^{-\frac{5}{2}}}\left(\frac{\left(\frac{5}{4}\right)^2}{2!}e^{-\frac{5}{4}}\frac{\left(\frac{5}{2}\right)}{1!}e^{-\frac{5}{2}} + \frac{\left(\frac{5}{4}\right)}{1!}e^{-\frac{5}{4}}\frac{\left(\frac{5}{2}\right)^2}{2!}e^{-\frac{5}{2}} + e^{-\frac{5}{4}}\frac{\left(\frac{5}{2}\right)^3}{3!}e^{-\frac{5}{2}}\right) \sim 0.317\,.
\end{aligned}$$

Esercizio 3. Sia (X, Y) una variabile aleatoria continua bidimensionale con la seguente densità congiunta:

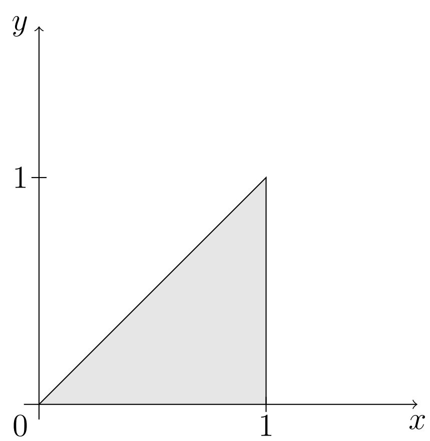
$$f(x, y) = \begin{cases} C & \text{se } 0 < y < x < 1\,, \\ 0 & \text{altrimenti.} \end{cases}$$

1. Si determini C in modo che f sia una densità di probabilità.
2. Si calcoli la funzione di densità di probabilità marginale g di X.
3. Si calcoli la funzione di densità di probabilità marginale h di Y.
4. Le variabili aleatorie X e Y sono indipendenti?

Soluzione.
1. La funzione f è una densità di probabilità se $f \geq 0$ e $\int\int f(x, y)\,\mathrm{d}x\,\mathrm{d}y = 1$. Imponiamo la seconda condizione

$$1 = \int\int_{\{0<y<x<1\}} C\,\mathrm{d}x\,\mathrm{d}y = C|\{0 < y < x < 1\}| = C\frac{1}{2}\,.$$

L'insieme $\{0 < y < x < 1\}$ di cui è stata calcolata l'area è il triangolo raffigurato in grigio.



Concludiamo che $C = 2$.

2. Dalla definizione si ha che

$$g(x) = \int_{-\infty}^{\infty} f(x,y)\,\mathrm{d}y = \int_0^x 2\,\mathrm{d}y = 2x\,,$$

per $0 < x < 1$ e $g(x) = 0$ altrimenti.

3. Analogamente

$$h(y) = \int_{-\infty}^{\infty} f(x,y)\,\mathrm{d}x = \int_y^1 2\,\mathrm{d}x = 2(1-y)\,,$$

$0 < y < 1$ e $h(y) = 0$ altrimenti.

4. Le due variabili non sono indipendenti perché $f(x,y) \neq g(x)h(y)$.

Esercizio 4. Un produttore afferma che la carica media di un certo tipo di batterie è maggiore o uguale a 240 ampere-ora. Un campione di 18 batterie di questo tipo che è stato analizzato ha fornito i valori seguenti:

237	242	244	262	225	218	242	248	243
234	236	228	232	230	254	220	232	240

Assumendo che la distribuzione della carica sia normale, si può accettare l'affermazione del produttore con un livello di significatività del 5%?

Soluzione. Il campione estratto è $X_1, \dots X_n$ con $n = 18$. Ogni X_i ha distribuzione normale con media μ e varianza σ^2. L'ipotesi che vogliamo testare è la seguente

$$H_0 : \mu \geq 240 \quad H_1 : \mu < 240\,.$$

Poiché media μ e varianza σ^2 della distribuzione normale non sono note, utilizzeremo come statistica test la variabile aleatoria

$$T = \frac{\overline{X} - \mu}{S/\sqrt{n}}\,,$$

dove $\overline{X}$ è la media campionaria e S è la statistica campionaria. Il campione analizzato non è numeroso ($n < 30$), quindi utilizziamo il fatto che T è una variabile aleatoria di tipo t-Student con $n - 1 = 17$ gradi di libertà.
Denotiamo con $t_{\alpha,17}$ i valori della distribuzione t-Student tali che

$$\mathbf{P}(T > t_{\alpha,17}) = \alpha\,, \quad (\text{e } \mathbf{P}(T < -t_{\alpha,17}) = \alpha \text{ per simmetria}).$$

Trattandosi di un test unilaterale sinistro, la regione di accettazione è della forma

$$[-t_{\alpha,17}, +\infty)\,.$$

Con il livello di significatività $\alpha = 5\%$ abbiamo che $t_{0.5,17} \sim 1.740$.
Le realizzazioni del campione forniscono i seguenti dati

$$\overline{X} \sim 237.056 \quad S = 11.280$$

e pertanto

$$\frac{\overline{X} - 240}{S/\sqrt{18}} \sim -1.107\,.$$

L'ipotesi si può accettare con il livello di significatività $\alpha = 5\%$ poiché $-1.107 \in [-1.740, +\infty)$.

Esercizio 5. I valori che seguono sono le percentuali di ceneri residue per dei campioni di carbone trovati in uno stesso sito.

9.2 14.1 9.8 12.4 16.0 12.6 22.7 18.9 21.0 14.5 20.4 16.9

1. Determinare i quartili del campione.
2. Individuare dati anomali o sospetti.
3. Tracciare un box plot.

Soluzione. Ordiniamo il campione

9.2 9.8 12.4 12.6 14.1 14.5 16.0 16.9 18.9 20.4 21.0 22.7

1. Il primo quartile Q_1 è il numero tale che $1/4$ dei dati del campione è a sinistra di Q_1. Poiché $(12+1)/4 = 13/4 = 3 + 1/4$ non è intero, si ottiene Q_1 come media dei valori ai posti 3 e 4, cioè $Q_1 = \frac{12.4+12.6}{2} = 12.5$.
Poiché $13/2 = 6 + 1/2$ si ottiene Q_2 come media dei valori ai posti 6 e 7, cioè $Q_2 = \frac{14.5+16.0}{2} = 15.25$
Infine, poiché $13\frac{3}{4} = 9 + 3/4$ si ottiene Q_3 come media dei valori ai posti 9 e 10, cioè $Q_3 = \frac{18.9+20.4}{2} = 19.65$.

2. L'interquartile range è dato da $IQR = Q_3 - Q_1 = 7.15$.
I dati anomali cadono negli intevalli

$$(-\infty, Q_1 - 3IQR] \cup [Q_3 + 3IQR, +\infty) = (-\infty, -8.95] \cup [41.1, +\infty)$$

quindi non ci sono dati anomali.
I dati sospetti cadono negli intervalli

$$(Q_1 - 3IQR, Q_1 - 1.5IQR] \cup [Q_3 + 1.5IQR, Q_3 + 3IQR) = (-8.95, 1.775] \cup [30.375, 41.1)$$

quindi non ci sono dati sospetti.

3. Segue il box plot.