

Esercizi di Calcolo delle Probabilità e Statistica

a.a.2023/2024
24/06/2024

Esercizio 1 Una società assicurativa ha $N = 250000$ assicurati, di cui 150000 uomini e 100000 donne. Da una statistica risulta che la probabilità che un uomo abbia un incidente grave in un anno è del 3% mentre questa probabilità scende a 1,5% per le donne.

1. Calcolare la probabilità che un assicurato (scelta a caso) abbia un incidente grave nel prossimo anno.
2. Sapendo che un assicurato ha avuto un incidente grave qual è la probabilità che si tratti di un uomo?
3. Calcolare la media e la varianza del numero complessivo di assicurati che hanno un incidente grave in un anno.

Soluzione: Indichiamo con $U = 15 \cdot 10^4$ il numero di assicurati uomini e con $D := N - U = 10^5$ il numero di assicurati donne $\mathbf{1}_U$ la v.a. che indica che l'assicurato sia un uomo. Sia inoltre G l'evento che un assicurato abbia un incidente grave nel prossimo anno. Per ipotesi si ha che

$$\begin{aligned}\mathbb{E}(\mathbf{1}_U) &= \mathbb{P}\{\text{l'assicurato è un uomo}\} = \frac{U}{N} = \frac{3}{5}, \\ \mathbb{P}(G | \{\mathbf{1}_U = 1\}) &= \mathbb{P}\{\text{un uomo abbia un incidente grave}\} = \frac{3}{10^2}, \\ \mathbb{P}(G | \{\mathbf{1}_U = 0\}) &= \mathbb{P}\{\text{una donna abbia un incidente grave}\} = \frac{3}{2 \cdot 10^2}.\end{aligned}$$

1. La probabilità che un assicurato (scelto a caso) abbia un incidente grave nel prossimo anno, per la formula delle probabilità totali, è pari a

$$\begin{aligned}\mathbb{P}(G) &= \mathbb{P}(G | \{\mathbf{1}_U = 1\}) \mathbb{P}\{\mathbf{1}_U = 1\} + \mathbb{P}(G | \{\mathbf{1}_U = 0\}) \mathbb{P}\{\mathbf{1}_U = 0\} \\ &= \frac{3}{10^2} \mathbb{E}[\mathbf{1}_U] + \frac{3}{2 \cdot 10^2} (1 - \mathbb{E}[\mathbf{1}_U]) \\ &= \frac{3}{10^2} \frac{3}{5} + \frac{3}{2 \cdot 10^2} \frac{2}{5} = \frac{3}{10^2} \frac{4}{5} = \frac{3}{5^3} = 0,024,\end{aligned}$$

2. Per la formula di Bayes

$$\mathbb{P}(\{\mathbf{1}_U = 1\} | G) = \mathbb{P}(G | \{\mathbf{1}_U = 1\}) \frac{\mathbb{P}\{\mathbf{1}_U = 1\}}{\mathbb{P}(G)} = \frac{\frac{3}{2^2 \cdot 5^2} \frac{3}{5}}{\frac{3}{5^3}} = \frac{3}{4}$$

3. Per ogni $i = 1, \dots, D$ sia X_i la v.a. che indica l'evento che l' i -esima donna assicurata avrà un incidente grave in un anno. Allo stesso modo, Per ogni $i = 1, \dots, U$ sia Y_i la v.a. che indica l'evento che l' i -esimo uomo assicurato avrà un incidente grave in un anno. Ponendo $X := \sum_{i=1}^D X_i$ e $Y := \sum_{i=1}^U Y_i$, il numero complessivo di assicurati che avrà un incidente grave in un anno è $Z := X + Y$. Si può supporre che $\{X_i\}_{i=1}^D$ e $\{Y_i\}_{i=1}^U$ siano collezioni di vv.aa. indipendenti e che gli elementi di ciascuna collezione siano indipendenti da quelli dell'altra. Perciò X e Y sono v.a. Binomiali indipendenti rispettivamente di parametri $D, p := \frac{3}{2 \cdot 10^2}$ e $U, q := \frac{3}{10^2}$. Perciò,

$$\begin{aligned}\mathbb{E}[Z] &= \mathbb{E}[X] + \mathbb{E}[Y] = Dp + Uq \\ &= 10^5 \frac{3}{2 \cdot 10^2} + 3 \cdot 5 \cdot 10^4 \frac{3}{10^2} \\ &= 3 \cdot 5 \cdot 10^2 + 3^2 \cdot 5 \cdot 10^2 = 2 \cdot 3 \cdot 10^3, \\ \text{Var}[Z] &= \text{Var}[X] + \text{Var}[Y] = Dp(1-p) + Uq(1-q) \\ &= 5842,5.\end{aligned}$$

■

Esercizio 2 Sia (X, Y) un vettore aleatorio uniformemente distribuito sull'insieme

$$D := \{(x, y) \in \mathbb{R}^2 : |x| + |y| \geq 1, x^2 + y^2 \leq 1\}.$$

1. Calcolare la costante di normalizzazione della densità di probabilità congiunta.
2. Verificare se le componenti del vettore sono indipendenti e in caso contrario calcolare la densità della probabilità condizionata della componente Y rispetto alla componente X .
3. Calcolare la matrice di covarianza di (X, Y) , ovvero

$$\begin{pmatrix} \text{Cov}(X, X) & \text{Cov}(X, Y) \\ \text{Cov}(Y, X) & \text{Cov}(Y, Y) \end{pmatrix}.$$

Soluzione:

1. La costante di normalizzazione c della densità di probabilità di (X, Y) è pari all'inverso della differenza tra l'area del disco $C := \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1\}$ e quella del rombo $Q := \{(x, y) \in \mathbb{R}^2 : |x| + |y| \leq 1\}$ ovvero $c = \frac{1}{\pi - 2}$.

2. Le densità di probabilità marginali di (X, Y) sono

$$\begin{aligned}
 f_X(x) &= \mathbf{1}_{[0,1]}(x) \left(\int_{-\sqrt{1-x^2}}^{x-1} dy + \int_{1-x}^{\sqrt{1-x^2}} dy \right) \frac{1}{\pi-2} + \\
 &\quad + \mathbf{1}_{[-1,0]}(x) \left(\int_{-\sqrt{1-x^2}}^{-x-1} dy + \int_{x+1}^{\sqrt{1-x^2}} dy \right) \frac{1}{\pi-2} \\
 &= \mathbf{1}_{[0,1]}(x) \left(x-1 + \sqrt{1-x^2} \right) \frac{2}{\pi-2} + \\
 &\quad + \mathbf{1}_{[-1,0]}(x) \left(-x-1 + \sqrt{1-x^2} \right) \frac{2}{\pi-2} \\
 &= \frac{2}{\pi-2} \left(|x| - 1 + \sqrt{1-x^2} \right) \mathbf{1}_{[-1,1]}(x)
 \end{aligned}$$

e, analogamente, per simmetria rispetto allo scambio degli assi coordinati,

$$f_Y(y) = \frac{2}{\pi-2} \left(y-1 + \sqrt{1-y^2} \right) \mathbf{1}_{[-1,1]}(y) .$$

Da cui segue che le vv.aa. X e Y non sono tra loro stocasticamente indipendenti. Infatti, scegliendo per esempio il punto $\left(\frac{1}{4}, \frac{1}{4}\right)$ si ottiene

$$f_X\left(\frac{1}{4}\right) f_Y\left(\frac{1}{4}\right) = \frac{4}{(\pi-2)^2} \left(\frac{-3 + \sqrt{15}}{4} \right)^2 \neq f_{(X,Y)}\left(\frac{1}{4}, \frac{1}{4}\right) = 0 .$$

Allora,

$$f_{Y|X}(y|x) = \frac{f_{(X,Y)}(x,y)}{f_X(x)} = \frac{1}{2} \frac{\mathbf{1}_{[-\sqrt{1-x^2}, |x|-1]}(y) + \mathbf{1}_{[1-|x|, \sqrt{1-x^2}]}(y)}{(|x| - 1 + \sqrt{1-x^2})} .$$

3. Poiché la funzione identità $(\mathbb{R} \ni x \mapsto x \in \mathbb{R})$ è dispari, mentre f_X è una funzione pari supportata su un intervallo simmetrico,

$$\mathbb{E}[X] = \int_{\mathbb{R}} dx f_X(x) x = \int_{-1}^1 \frac{2}{\pi-2} \left(|x| - 1 + \sqrt{1-x^2} \right) x dx = 0 .$$

Allo stesso modo si ha che $\mathbb{E}[Y] = 0$. Inoltre,

$$\begin{aligned}
 \mathbb{E}[XY] &= \frac{1}{\pi-2} \int_D dx dy xy \\
 &= \frac{1}{\pi-2} \left[\int_C dx dy xy - \int_Q dx dy xy \right] .
 \end{aligned}$$

Notiamo che

$$\int_Q dx dy xy = \int_{-1}^0 dx x \int_{-(x+1)}^{x+1} dy y + \int_{-1}^0 dx x \int_{-(1-x)}^{1-x} dy y$$

è nullo poiché

$$\int_{-(x+1)}^{x+1} dy y = \int_{-(1-x)}^{1-x} dy y = 0$$

integrando funzioni dispari su un intervallo pari. Allora

$$\begin{aligned} \mathbb{E}[XY] &= \frac{1}{\pi-2} \int_C dx dy xy \\ &= \frac{1}{\pi-2} \left[\int_0^1 d\rho \rho^3 \int_0^{2\pi} d\theta \cos \theta \sin \theta \right] \\ &= \frac{1}{\pi-2} \frac{1}{8} \int_0^{2\pi} d\theta \sin 2\theta = \frac{1}{\pi-2} \frac{1}{16} \int_0^{4\pi} dt \sin t \\ &= \frac{1}{\pi-2} \frac{1}{16} \cos t \Big|_0^{4\pi} = 0 . \end{aligned}$$

Perciò

$$Cov(X, Y) = Cov(Y, X) = \mathbb{E}[XY] - \mathbb{E}[X] \mathbb{E}[Y] = 0 .$$

Siccome $Cov(X, X) = \mathbb{E}[X^2] - \mathbb{E}^2[X] = Var[X]$, non resta che calcolare $\mathbb{E}[X^2]$ che per simmetria sarà uguale a $\mathbb{E}[Y^2]$.

$$\begin{aligned} \mathbb{E}[X^2] &= \frac{2}{\pi-2} \int_{-1}^1 dx \left(|x| - 1 + \sqrt{1-x^2} \right) x^2 = \\ &= \frac{2}{\pi-2} \left[\frac{1}{2} - \frac{2}{3} + \int_{-1}^1 dx x^2 \sqrt{1-x^2} \right] \\ &= \frac{2}{\pi-2} \left[\frac{\pi}{8} - \frac{1}{6} \right] = \frac{1}{\pi-2} \left[\frac{\pi}{4} - \frac{1}{3} \right] , \end{aligned}$$

poiché

$$\int_{-1}^1 dx |x| x^2 = \int_0^1 dx x^3 - \int_{-1}^0 dx x^3 = 2 \int_0^1 dx x^3 = \frac{1}{2}$$

e

$$\begin{aligned} \int_{-1}^1 dx x^2 \sqrt{1-x^2} &= \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} dt \sin^2 t \cos^2 t = \frac{1}{4} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} dt \sin^2(2t) \\ &= \frac{1}{8} \int_{-\pi}^{\pi} du \sin^2 u = \frac{1}{16} (x - \sin x \cos x) \Big|_{-\pi}^{\pi} = \frac{\pi}{8} . \end{aligned}$$

Analoghe considerazioni valgono per la v.a. Y . Quindi

$$\begin{aligned} \begin{pmatrix} Cov(X, X) & Cov(X, Y) \\ Cov(Y, X) & Cov(Y, Y) \end{pmatrix} &= \begin{pmatrix} \mathbb{E}[X^2] & 0 \\ 0 & \mathbb{E}[X^2] \end{pmatrix} \\ &= \begin{pmatrix} \frac{1}{\pi-2} \left[\frac{\pi}{4} - \frac{1}{3} \right] & 0 \\ 0 & \frac{1}{\pi-2} \left[\frac{\pi}{4} - \frac{1}{3} \right] \end{pmatrix} . \end{aligned}$$

■

Esercizio 3 Supponiamo di avere una successione di v.v.a.a. $\{X_n\}_{n \geq 0}$ sullo spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ che descrive una catena di Markov con 4 stati, $S = \{1, 2, 3, 4\}$ tale che: dallo stato 1 si transisce con probabilità $\frac{1}{2}$ agli stati 2 e 3. Dallo stato 2, si transisce con probabilità $\frac{1}{3}$ agli stati 1 e 4 e con la stessa probabilità si resta nello stato 2. Dallo stato 3 si transisce con probabilità $\frac{1}{3}$ agli stati 1 e 4 e con la stessa probabilità si resta nello stato 3. Dallo stato 4 si transisce con probabilità $\frac{1}{2}$ agli stati 2 e 3.

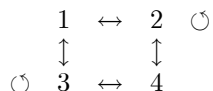
1. Elencare le classi di equivalenza degli stati e calcolarne il periodo.
2. Calcolare $\mathbb{P}(\{X_5 = 2\} | \{X_2 = 3\})$ e, assumendo come distribuzione iniziale quella uniforme, calcolare $\mathbb{E}[X_2]$.
3. Se $\{P_{i,j}\}_{i,j=1,\dots,4}$ indica la matrice delle probabilità di transizione tra gli stati della catena, calcolare se esistono: $\lim_{n \rightarrow \infty} \mathbb{P}\{\omega \in \Omega : X_n(\omega) = 4\}$, $\lim_{n \rightarrow \infty} (P_{1,2}^n + P_{3,4}^n)$, $\lim_{n \rightarrow \infty} P_{2,3}^{2n}$.

Soluzione:

1. La matrice delle probabilità di transizione tra gli stati della catena è

$$P = \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{3} & \frac{1}{3} & 0 & \frac{1}{3} \\ \frac{1}{3} & 0 & \frac{1}{3} & \frac{1}{3} \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 \end{pmatrix}$$

perciò la catena può essere rappresentata mediante il grafo diretto



da cui si deduce che esiste un'unica classe di stati comunicanti aperiodica.

2. Poiché la catena di Markov è omogenea

$$\mathbb{P}(\{X_5 = 2\} | \{X_2 = 3\}) = \mathbb{P}(\{X_3 = 2\} | \{X_0 = 3\}) = P_{3,2}^3.$$

Ma

$$P^2 = \begin{pmatrix} \frac{1}{3} & \frac{1}{6} & \frac{1}{6} & \frac{1}{3} \\ \frac{1}{9} & \frac{1}{9} & \frac{1}{3} & \frac{1}{9} \\ \frac{1}{9} & \frac{1}{3} & \frac{1}{9} & \frac{1}{9} \\ \frac{1}{3} & \frac{1}{6} & \frac{1}{6} & \frac{1}{3} \end{pmatrix},$$

$$P^3 = \begin{pmatrix} \frac{1}{9} & \frac{7}{9} & \frac{7}{9} & \frac{1}{9} \\ \frac{1}{27} & \frac{2}{27} & \frac{2}{27} & \frac{1}{27} \\ \frac{1}{27} & \frac{2}{27} & \frac{1}{27} & \frac{1}{27} \\ \frac{1}{9} & \frac{7}{9} & \frac{7}{9} & \frac{1}{9} \end{pmatrix},$$

Quindi, $P_{3,2}^3 = \frac{2}{9}$.

$$\begin{aligned}\mathbb{E}[X_2] &= \sum_{i,j=1}^4 j \mathbb{P}(\{X_2 = j\} | \{X_0 = i\}) \mathbb{P}\{X_0 = i\} = \\ &= \sum_{i,j=1}^4 \frac{1}{4} P_{i,j}^2 j = \frac{1}{4} \left(5 + 2 + \frac{4}{9} + 2 + \frac{5}{9} \right) = \frac{5}{2},\end{aligned}$$

3. Poiché la catena di Markov descritta dalla matrice delle probabilità di transizione P è indecomponibile e aperiodica vale il teorema ergodico. Perciò,

$$\begin{aligned}\lim_{n \rightarrow \infty} P_{2,3}^{2n} &= \pi_3, \quad \lim_{n \rightarrow \infty} P_{1,2}^n = \pi_2, \\ \lim_{n \rightarrow \infty} P_{3,4}^n &= \pi_4 = \sum_{i=1}^4 \mathbb{P}\{\omega \in \Omega : X_0(\omega) = i\} P_{i,4}^n = \lim_{n \rightarrow \infty} \mathbb{P}\{\omega \in \Omega : X_n(\omega) = 4\},\end{aligned}$$

e pertanto $\lim_{n \rightarrow \infty} (P_{1,2}^n + P_{3,4}^n) = \pi_2 + \pi_4$, dove $\pi := (\pi_1, \pi_2, \pi_3, \pi_4)$ è la distribuzione di probabilità invariante per P , ovvero $\pi P = \pi$. Pertanto, risolvendo

$$\begin{cases} \frac{\pi_2}{3} + \frac{\pi_3}{3} = \pi_1 \\ \frac{\pi_1}{2} + \frac{\pi_2}{3} + \frac{\pi_4}{2} = \pi_2 \\ \frac{\pi_1}{2} + \frac{\pi_3}{3} + \frac{\pi_4}{2} = \pi_3 \\ \frac{\pi_2}{3} + \frac{\pi_3}{3} = \pi_4 \\ \pi_1 + \pi_2 + \pi_3 + \pi_4 = 1 \end{cases}$$

si ha $\pi_1 = \pi_4 = \frac{1}{5}, \pi_2 = \pi_3 = \frac{3}{10}$.

■

Esercizio 4 Sia $\{E_i\}_{i \geq 1}$ una successione di eventi stocasticamente indipendenti subordinatamente alla conoscenza di un parametro Θ , che si assume essere una v.a. la cui distribuzione di probabilità a priori è descritta dalla funzione di ripartizione (distribuzione)

$$F_{\Theta}(\theta) := \theta^2 \mathbf{1}_{[0,1]}(\theta),$$

tale che $\mathbb{P}(E_i | \Theta) = \Theta$. Dopo le prime quattro osservazioni si registra che soltanto il primo evento è accaduto.

1. Calcolare la densità di probabilità a posteriori di Θ .
2. Calcolare il valore del parametro α tale che la probabilità a priori che $\Theta \in [\alpha, 1]$ sia maggiore del 95%.
3. Calcolare il valore del parametro α tale che la probabilità a posteriori che $\Theta \in [\alpha, 1]$ sia maggiore del 95%.
4. Calcolare il valore massimo della densità di probabilità a posteriori di Θ .

Soluzione:

1. Consideriamo le vv.aa. $\xi_i := \mathbf{1}_{E_i}, i = 1, \dots, 4$. La densità a posteriori di Θ dato il vettore aleatorio $(\xi_1, \xi_2, \xi_3, \xi_4)$ sarà

$$\begin{aligned}\pi_4(\Theta = \theta | E_1 \cap E_2^c \cap E_3^c \cap E_4^c) &= \\ \pi_4(\Theta = \theta | \{\omega \in \Omega : (\xi_1, \xi_2, \xi_3, \xi_4) = (1, 0, 0, 0)\}) &= \\ K \mathbb{P}(\{\omega \in \Omega : (\xi_1, \xi_2, \xi_3, \xi_4) = (1, 0, 0, 0)\} | \Theta = \theta) \pi_0(\theta) &= \\ = K \mathbb{P}(E_1 \cap E_2^c \cap E_3^c \cap E_4^c | \Theta = \theta) \pi_0(\theta) &= \\ = K \mathbb{P}(E_1 | \Theta = \theta) \mathbb{P}(E_2^c | \Theta = \theta) \times & \\ \times \mathbb{P}(E_3^c | \Theta = \theta) \mathbb{P}(E_4^c | \Theta = \theta) \pi_0(\theta) &= \\ K \mathbb{P}(E_1 | \Theta = \theta) (1 - \mathbb{P}(E_2 | \Theta = \theta)) (1 - \mathbb{P}(E_3 | \Theta = \theta)) \times & \\ \times (1 - \mathbb{P}(E_4 | \Theta = \theta)) \pi_0(\theta) = K \theta (1 - \theta)^3 \pi_0(\theta) &= \\ = 2K \theta^2 (1 - \theta)^3 &\end{aligned}$$

dove si è usato che $\pi_0(\theta) = \frac{d}{d\theta} F_\Theta(\theta) = 2\theta$. Poiché

$$\begin{aligned}\int_0^1 d\theta 2K \theta^2 (1 - \theta)^3 &= 2K \frac{\Gamma(3) \Gamma(4)}{\Gamma(7)} = 2K \frac{2!3!}{6!} \\ &= K \frac{1}{30} = 1 \implies K = 30\end{aligned}$$

e

$$\pi_4(\Theta = \theta | E_1 \cap E_2^c \cap E_3^c \cap E_4^c) = 60\theta^2 (1 - \theta)^3.$$

2. La probabilità *a priori* che $\Theta \in [\alpha, 1]$ sia maggiore del 95% per definizione è

$$\int_\alpha^1 d\theta \pi_0(\theta) = F_\Theta(1) - F_\Theta(\alpha) = 1 - \alpha^2 \geq 1 - \frac{5}{10^2},$$

da cui segue $0 \leq \alpha \leq \frac{\sqrt{5}}{10}$

3. La probabilità *a posteriori* che $\Theta \in [\alpha, 1]$ sia maggiore del 95% per definizione sarà

$$\begin{aligned}\int_\alpha^1 d\theta \pi_4(\Theta = \theta | E_1 \cap E_2^c \cap E_3^c \cap E_4^c) &= \int_\alpha^1 d\theta 60\theta^2 (1 - \theta)^3 \\ &= 60 \left(\frac{\theta^3}{3} \Big|_\alpha^1 - 3 \frac{\theta^4}{4} \Big|_\alpha^1 + 3 \frac{\theta^5}{5} \Big|_\alpha^1 - \frac{\theta^6}{6} \Big|_\alpha^1 \right) \\ &= 60 \left(\frac{1}{60} - \frac{20\alpha^3 - 45\alpha^4 + 36\alpha^5 - 10\alpha^6}{60} \right) \\ &= 1 - 20\alpha^3 + 45\alpha^4 - 36\alpha^5 + 10\alpha^6 \geq 1 - \frac{5}{10^2},\end{aligned}$$

pertanto α deve soddisfare la disequazione $20\alpha^3 + 45\alpha^4 - 36\alpha^5 + 10\alpha^6 \leq \frac{5}{10^2}$.

4. Differenziando $\pi_4(\Theta = \theta | E_1^c \cap E_2 \cap E_3 \cap E_4) = 60\theta^2(1-\theta)^3$ otteniamo

$$2\theta(1-\theta)^3 - 3\theta^2(1-\theta)^2 = 0$$

ovvero che il valore massimo della densità di probabilità a posteriori di Θ è raggiunto per $\theta = \frac{2}{5}$.

■