Prova scritta di Probabilità e Statistica
II appello, I sessione, a.a. 2015/2016
13 Luglio 2016

Esercizio 1. Un dado equilibrato viene lanciato N volte, dove N è un numero aleatorio di legge $\mathrm{Po}(\lambda)$, $\lambda > 0$. Sia X il numero dei lanci che hanno dato esito pari.

a) Calcolare la legge di X.

b) Si osservano esattamente 2 esiti pari: qual è la probabilità di aver lanciato il dado almeno 4 volte?

c) Si osservano esattamente 2 esiti pari: qual è il valor medio del numero di lanci effettuati?

Esercizio 2. Sia (X, Y) un vettore aletorio su $\mathbb{R}^2$ di densità continua

$$f_{X,Y}(x, y) = c\, xy\, 1\!\!1_{0<x<y<1}, \quad \text{dove } c \in \mathbb{R}.$$

a) Verificare che $c = 8$.

b) Scrivere la legge di $Z = \max(|X|, |Y|)$ e calcolare, qualora esistano, media e varianza di Z.

c) Siano $Z_1, Z_2, \dots Z_n$ v.a. i.i.d. tutte con la stessa legge di Z e sia $W_n = \min(Z_1, \dots, Z_n)$. Studiare la convergenza in legge della successione $\{W_n\}_n$.

Esercizio 3. Due campioni casuali $\{X_i\}_{i=1,\dots,n}$ e $\{Y_i\}_{i=1,\dots,n}$ vengono estratti da due popolazioni bernoulliane ed indipendenti, rispettivamente di parametro p_X e p_Y.

a) Supponiamo $n = 200$. Scrivere un intervallo di fiducia (eventualmente approssimato) al 95% per $p = p_X - p_Y$.

b) Supponiamo che $p_X = p_Y = \frac{1}{2}$. Calcolare

$$\lim_{n\to\infty} \mathbb{P}\Big(\sqrt{2}\sum_{i=1}^{n}(X_i + Y_i) + 0.1\sqrt{n} \geq \sqrt{2}\, n\Big).$$

Esercizio 4. Consideriamo una catena di Markov $\{X_n : n \geq 0\}$ su $E = \{1, 2, 3, 4, 5\}$ con matrice di transizione

$$P = \begin{pmatrix} \alpha/2 & (1-\alpha)/2 & \alpha/2 & (1-\alpha)/2 & 0 \\ \beta/2 & (1-\beta)/2 & 0 & (1-\beta)/2 & \beta/2 \\ 0 & 0 & 1/3 & 1/3 & 1/3 \\ 0 & 0 & 1/4 & 1/2 & 1/4 \\ 0 & 0 & 0 & 1/4 & 3/4 \end{pmatrix}$$

per qualche $\alpha, \beta \in (0, 1)$.

a) Classificare gli stati della catena, trovare le classi irriducibili e le distribuzioni stazionarie.

b) Supponiamo che $\mathbb{P}(X_0 = 1) = \mathbb{P}(X_0 = 2) = 1/2$. Trovare i valori di α e β per cui si ha $\mathbb{P}(X_1 = 1) = 1/10$ e $\mathbb{P}(X_1 = 3) = 1/100$.

c) Supponiamo che $\mathbb{P}(X_0 \in \{3, 4, 5\}) = 1$. Calcolare $\lim_{n\to\infty} \mathbb{P}(X_n = j)$ per ogni $j \in E$.

d) Calcolare la probabilità di passaggio in 2 partendo da 1, e la probabilità di passaggio in 1 partendo da 2.

Soluzioni

Esercizio 1. a) Per $x = 0, 1, \ldots$, si ha

$$\mathbb{P}(X = x) = \sum_{n=0}^{\infty} \underbrace{\mathbb{P}(X = x \mid N = n)}_{= 0 \text{ se } x > n} \mathbb{P}(N = n) = \sum_{n=x}^{\infty} \mathbb{P}(X = x \mid N = n) \frac{\lambda^n}{n!} e^{-\lambda}.$$

Se è noto il numero n di lanci, il numero di esiti pari segue una legge binomiale di parametri n e $\frac{1}{2}$, dunque

$$\begin{aligned}\mathbb{P}(X = x) &= \sum_{n=x}^{\infty} \binom{n}{x} \left(\frac{1}{2}\right)^x \left(\frac{1}{2}\right)^{n-x} \frac{\lambda^n}{n!} e^{-\lambda} = \frac{1}{x!} \left(\frac{\lambda}{2}\right)^x e^{-\lambda} \sum_{n=x}^{\infty} \frac{1}{(n-x)!} \left(\frac{\lambda}{2}\right)^{n-x} \\ &= \frac{1}{x!} \left(\frac{\lambda}{2}\right)^x e^{-\lambda} \sum_{k=0}^{\infty} \frac{1}{k!} \left(\frac{\lambda}{2}\right)^k = \frac{1}{x!} \left(\frac{\lambda}{2}\right)^x e^{-\lambda} e^{\lambda/2} = \frac{(\frac{\lambda}{2})^x}{x!} e^{-\frac{\lambda}{2}},\end{aligned}$$

da cui si ottiene $X \sim \text{Po}(\frac{\lambda}{2})$.

b) Chiede $\mathbb{P}(N \geq 4 \mid X = 2)$:

$$\mathbb{P}(N \geq 4 \mid X = 2) = \sum_{n=4}^{\infty} \mathbb{P}(N = n \mid X = 2).$$

Per $n \geq 2$, si ha

$$\begin{aligned}\mathbb{P}(N = n \mid X = 2) &= \frac{\mathbb{P}(X = 2 \mid N = n)\mathbb{P}(N = n)}{\mathbb{P}(X = 2)} = \frac{\binom{n}{2} \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^{n-2} \frac{\lambda^n}{n!} e^{-\lambda}}{\frac{(\frac{\lambda}{2})^2}{2!} e^{-\frac{\lambda}{2}}} \\ &= \frac{(\frac{\lambda}{2})^{n-2}}{(n-2)!} e^{-\frac{\lambda}{2}}.\end{aligned}$$

Quindi,

$$\begin{aligned}\mathbb{P}(N \geq 4 \mid X = 2) &= \sum_{n=4}^{\infty} \frac{(\frac{\lambda}{2})^{n-2}}{(n-2)!} e^{-\frac{\lambda}{2}} = \underbrace{\sum_{n=2}^{\infty} \frac{(\frac{\lambda}{2})^{n-2}}{(n-2)!} e^{-\frac{\lambda}{2}}}_{=1} - \left(\frac{(\frac{\lambda}{2})^0}{0!} e^{-\frac{\lambda}{2}} + \frac{(\frac{\lambda}{2})^1}{1!} e^{-\frac{\lambda}{2}}\right) \\ &= 1 - e^{-\frac{\lambda}{2}} \left(1 + \frac{\lambda}{2}\right).\end{aligned}$$

c) Chiede $\mathbb{E}(N \mid X = 2)$:

$$\begin{aligned}\mathbb{E}(N \mid X = 2) &= \sum_{n=2}^{\infty} n \frac{(\frac{\lambda}{2})^{n-2}}{(n-2)!} e^{-\frac{\lambda}{2}} = \sum_{n=2}^{\infty} (n-2) \frac{(\frac{\lambda}{2})^{n-2}}{(n-2)!} e^{-\frac{\lambda}{2}} + 2 \underbrace{\sum_{n=2}^{\infty} \frac{(\frac{\lambda}{2})^{n-2}}{(n-2)!} e^{-\frac{\lambda}{2}}}_{=1} \\ &= \underbrace{\sum_{k=0}^{\infty} k \frac{(\frac{\lambda}{2})^k}{k!} e^{-\frac{\lambda}{2}}}_{= \mathbb{E}(Z) \text{ con } Z \sim \text{Po}(\frac{\lambda}{2})} + 2 = \frac{\lambda}{2} + 2.\end{aligned}$$

Esercizio 2. a) Dev'essere

$$1 = \int_{\mathbb{R}^2} cxy \, 1\!\!1_{0<x<y<1} \, dxdy = c\int_0^1 dx \int_x^1 xy \, dy = c\int_0^1 x\Big(\frac{1}{2} - \frac{x^2}{2}\Big) \, dx = \frac{c}{2}\Big(\frac{1}{2} - \frac{1}{4}\Big) = \frac{c}{8},$$

da cui $c = 8$.

b) Calcoliamo la f.r. F_Z di Z. Dalla legge congiunta si vede che $0 < X < Y < 1$ q.c. e quindi $Z = \max(X, Y) = Y \in (0, 1)$, da cui segue che $F_Z = F_Y$, dove F_Y denota la f.r. di Y. In particolare, $F_Z(z) = 0$ per $z \leq 0$ e $F_Z(z) = 1$ per $z \geq 1$. Se invece $z \in (0, 1)$, si ha

$$\begin{aligned} F_Z(z) = F_Y(y) = \mathbb{P}(X \in \mathbb{R}, Y \leq z) &= \int_{\mathbb{R}\times(-\infty,z]} 8\,xy \, 1\!\!1_{0<x<y<1} dxdy \\ &= \int_0^z dx \int_x^z 8xydy = z^4. \end{aligned}$$

Derivando, otteniamo la densità di Z:

$$f_Z(z) = \frac{d}{dz}F_Z(z) = 4z^3 \, 1\!\!1_{0<z<1}.$$

Z ha tutti i momenti perché $Z \in [0, 1]$ q.c. Calcoliamo media e varianza:

$$\begin{aligned} \mathbb{E}(Z) &= \int_{\mathbb{R}} z f_Z(z)dz = \int_{\mathbb{R}} 4z^4 \, 1\!\!1_{0<z<1}dz = 4\int_0^1 z^4 \, dz = \frac{4}{5}, \\ \mathrm{Var}(Z) &= \mathbb{E}(Z^2) - \mathbb{E}(Z)^2 = \int_{\mathbb{R}} 4z^5 \, 1\!\!1_{0<z<1}dz - \frac{16}{25} = 4\int_0^1 z^5 \, dz - \frac{16}{25} = \frac{4}{6} - \frac{16}{25} = \frac{2}{75}. \end{aligned}$$

c) Ancora una volta, W_n è a valori in $(0, 1)$, quindi $F_{W_n}(w) = 0$ per $w \leq 0$ e $F_{W_n}(w) = 1$ per $w \geq 1$. Se invece $w \in (0, 1)$,

$$\begin{aligned} F_{W_n}(w) &= \mathbb{P}(\min(Z_1, \ldots, Z_n) \leq w) = 1 - \mathbb{P}(\min(Z_1, \ldots, Z_n) > w) \\ &= 1 - \mathbb{P}(Z_1 > w, \ldots, Z_n > w) = 1 - \mathbb{P}(Z_1 > w) \cdots \mathbb{P}(Z_n > w) = 1 - \mathbb{P}(Z > w)^n \\ &= 1 - (1 - \mathbb{P}(Z \leq w))^n = 1 - (1 - w^4)^n \to 1. \end{aligned}$$

Detta G la f.r. della v.a. $W = 0$, cioè $G(w) = 0$ per $w < 0$ e $G(w) = 1$ per $w \geq 0$, si ha $F_{W_n}(w) \to G(w)$ per ogni w tale che $\Delta G(w) \neq 0$, da cui segue che $W_n \to 0$ in legge.

Esercizio 3. a) La differenza dei parametri $p = p_X - p_Y$ risulta la media di $Z_i = X_i - Y_i$, quindi $Z_1, \ldots, Z_{200}$ è un campione lungo 200 per la stima di p. Osserviamo che, essendo le X_i indipendenti tra loro, le Y_i idem ed essendo i due campioni indipendenti, otteniamo che le Z_i sono indipendenti ed inoltre (poiché $X_i \perp\!\!\!\perp Y_i$) Z_i ha media (incognita) p e varianza

$$\mathrm{Var}(Z_i) = \mathrm{Var}(X_i) + \mathrm{Var}(Y_i) = p_X(1 - p_X) + p_Y(1 - p_Y) \leq \frac{1}{4} + \frac{1}{4} = \frac{1}{2}.$$

Quindi, un IF approssimato per p a livello richiesto è dato dall'intervallo aleatorio I di estremi

$$\bar{Z}_{200} \pm \sqrt{\frac{1/2}{200}}\, 1.96 = \bar{Z}_{200} \pm \frac{1}{20}\, 1.96 = \bar{Z}_{200} \pm 0.098,$$

dove $\bar{Z}_{200} = \frac{1}{200}\sum_{i=1}^{200} Z_i$.

b) Le v.a. $X_i + Y_i$, $i = 1, \dots, n$, sono indipendenti, di media $m = \frac{1}{2} + \frac{1}{2} = 1$ e varianza $\sigma^2 = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$. Allora,

$$\begin{aligned}\mathbb{P}\Big(\sqrt{2}\sum_{i=1}^n (X_i + Y_i) + 0.1\sqrt{n} > \sqrt{2}\,n\Big) &= \mathbb{P}\Big(\sqrt{2}\sum_{i=1}^n (X_i + Y_i) - \sqrt{2}\,n > -0.1\sqrt{n}\Big) \\ &= \mathbb{P}\Big(\frac{\sum_{i=1}^n [(X_i + Y_i) - m]}{\sqrt{n\sigma^2}} > -0.1\Big).\end{aligned}$$

Usando il TLC, ed indicando con W una v.a. gaussiana standard,

$$\begin{aligned}\lim_{n\to\infty} \mathbb{P}\Big(\sqrt{2}\sum_{i=1}^n (X_i + Y_i) + 0.1\sqrt{n} > \sqrt{2}\,n\Big) &= \lim_{n\to\infty} \mathbb{P}\Big(\frac{\sum_{i=1}^n [(X_i + Y_i) - m]}{\sqrt{n\sigma^2}} > -0.1\Big) \\ &= \mathbb{P}(W > -0.1) = \Phi(0.1) = 0.53983.\end{aligned}$$

Esercizio 4. a) L'insieme degli stati transitori è $T = \{1, 2\}$; infatti, per ogni $i \in T$, lo stato i comunica con lo stato 4 e non vale il viceversa. Gli stati 3, 4 e 5 sono ricorrenti perché $C = \{3, 4, 5\}$ è una classe chiusa irriducibile. Una distribuzione stazionaria è del tipo

$$\pi = (\pi_1, \pi_2, \pi_3, \pi_4, \pi_5)$$

con $\pi_1 = \pi_2 = 0$. Inoltre $\pi = (\pi_3, \pi_4, \pi_5)$ è l'unica distribuzione stazionaria della sottocatena ristretta a C, dove l'unicità segue dalla irriducibilità. Quindi si deve considerare la relazione matriciale

$$(\pi_3, \pi_4, \pi_5)\begin{pmatrix} 1/3 & 1/3 & 1/3 \\ 1/4 & 1/2 & 1/4 \\ 0 & 1/4 & 3/4 \end{pmatrix} = (\pi_3, \pi_4, \pi_5)$$

da cui segue il sistema

$$\begin{cases} \pi_3 = \frac{1}{3}\pi_3 + \frac{1}{4}\pi_4 \\ \pi_4 = \frac{1}{3}\pi_3 + \frac{1}{2}\pi_4 + \frac{1}{4}\pi_5 \\ \pi_5 = \frac{1}{3}\pi_3 + \frac{1}{4}\pi_4 + \frac{3}{4}\pi_5. \end{cases}$$

Quindi si ottiene

$$\pi_4 = \frac{8}{3}\pi_3$$

dalla prima equazione; poi, sostituendo la prima equazione nella terza, si ha $\pi_5 = \pi_3 + \frac{3}{4}\pi_5$, e quindi

$$\pi_5 = 4\pi_3.$$

Allora, poiché si ha anche la condizione $\pi_3 + \pi_4 + \pi_5 = 1$, si ottiene

$$1 = \pi_3 + \frac{8}{3}\pi_3 + 4\pi_3 = \frac{3 + 8 + 12}{3}\pi_3 = \frac{23}{3}\pi_3$$

da cui segue

$$\pi_3 = \frac{3}{23},\ \pi_4 = \frac{8}{3}\cdot\frac{3}{23} = \frac{8}{23},\ \pi_5 = 4\cdot\frac{3}{23} = \frac{12}{23}.$$

In conclusione

$$\pi = (0, 0, 3/23, 8/23, 12/23)$$

è l'unica distribuzione stazionaria.

b) Si deve considerare la seguente relazione matriciale

$$(1/2,1/2,0,0,0)\begin{pmatrix} \alpha/2 & (1-\alpha)/2 & \alpha/2 & (1-\alpha)/2 & 0 \\ \beta/2 & (1-\beta)/2 & 0 & (1-\beta)/2 & \beta/2 \\ 0 & 0 & 1/3 & 1/3 & 1/3 \\ 0 & 0 & 1/4 & 1/2 & 1/4 \\ 0 & 0 & 0 & 1/4 & 3/4 \end{pmatrix} = (p_1,p_2,p_3,p_4,p_5),$$

dove $p_i = \mathbb{P}(X_1 = i)$ per ogni $i \in E$. Allora si ottiene

$$(p_1,p_2,p_3,p_4,p_5) = \left(\frac{\alpha+\beta}{4}, \frac{2-(\alpha+\beta)}{4}, \frac{\alpha}{4}, \frac{2-(\alpha+\beta)}{4}, \frac{\beta}{4}\right).$$

Quindi si cercano le soluzioni del seguente sistema

$$\begin{cases} \frac{\alpha+\beta}{4} = \frac{1}{10} \\ \frac{\alpha}{4} = \frac{1}{100}. \end{cases}$$

Dalla seconda equazione si ottiene subito

$$\alpha = \frac{4}{100} = \frac{1}{25}$$

e, sostituendo nella prima, si ha

$$\beta = \frac{4}{10} - \alpha = \frac{2}{5} - \frac{1}{25} = \frac{10-1}{25} = \frac{9}{25}.$$

c) Per l'ipotesi sulla distribuzione di X_0 abbiamo una catena di Markov $\{X_n : n \geq 0\}$ a valori in $C = \{3,4,5\}$. Quindi $\mathbb{P}(X_n = 1) = \mathbb{P}(X_n = 2) = 0$ per ogni n. Inoltre la sottocatena ristretta a C è regolare perché è irriducibile e la corrispondente matrice di transizione

$$\begin{pmatrix} 1/3 & 1/3 & 1/3 \\ 1/4 & 1/2 & 1/4 \\ 0 & 1/4 & 3/4 \end{pmatrix}$$

(ottenuta restringendosi agli stati di C) ha almeno un elemento positivo sulla diagonale principale (in realtà tutti gli elementi sulla diagonale principale sono positivi ...). Quindi, se consideriamo i valori $(\pi_3,\pi_4,\pi_5) = (3/23, 8/23, 12/23)$ calcolati prima, per il teorema di Markov applicato alla sottocatena ristretta a C si ha

$$\lim_{n\to\infty} p_{ij}^{(n)} = \pi_j \text{ per ogni } i,j \in C,$$

da cui segue

$$\lim_{n\to\infty} \mathbb{P}(X_n = j) = \lim_{n\to\infty} \sum_{i\in C} \mathbb{P}(X_0 = i) p_{ij}^{(n)} = \sum_{i\in C} \mathbb{P}(X_0 = i)\pi_j = \pi_j \text{ per ogni } j \in C.$$

In conclusione si ha

$$\lim_{n\to\infty} \mathbb{P}(X_n = j) = \begin{cases} 0 & \text{per } j \in \{1,2\} \\ 3/23 & \text{per } j = 3 \\ 8/23 & \text{per } j = 4 \\ 12/23 & \text{per } j = 5. \end{cases}$$

d) Vogliamo considerare le probabilità di passaggio $\{\lambda_i : i \in D_C\}$ per $C = \{2\}$ e le probabilità di passaggio $\{\tilde{\lambda}_i : i \in D_{\tilde{C}}\}$ per $\tilde{C} = \{1\}$. Come vedremo in entrambi i casi i due insiemi di probabilità di passaggio sono ridotti ad un unico elemento.
Nel primo caso abbiamo $D_C = \{1\}$; quindi $\{\lambda_i : i \in D_C\} = \{\lambda_1\}$, dove λ_1 è soluzione della seguente equazione

$$\lambda_1 = p_{12} + p_{11}\lambda_1 = \frac{1-\alpha}{2} + \frac{\alpha}{2}\lambda_1.$$

In corrispondenza si ha

$$\left(1 - \frac{\alpha}{2}\right)\lambda_1 = \frac{1-\alpha}{2}, \text{ da cui segue } \lambda_1 = \frac{1-\alpha}{2-\alpha}.$$

Nel secondo caso abbiamo $D_{\tilde{C}} = \{2\}$; quindi $\{\tilde{\lambda}_i : i \in D_{\tilde{C}}\} = \{\tilde{\lambda}_2\}$, dove $\tilde{\lambda}_2$ è soluzione della seguente equazione

$$\tilde{\lambda}_2 = p_{21} + p_{22}\tilde{\lambda}_2 = \frac{\beta}{2} + \frac{1-\beta}{2}\tilde{\lambda}_2.$$

In corrispondenza si ha

$$\left(1 - \frac{1-\beta}{2}\right)\tilde{\lambda}_2 = \frac{\beta}{2}, \text{ da cui segue } \tilde{\lambda}_2 = \frac{\beta}{1+\beta}.$$