Probabilità e Statistica – Corso di Laurea in Informatica
A.A. 2020/2021

ESERCITAZIONE DI RIPASSO DELLE LEZIONI 11–20

ER.1. Supponiamo che il tasso di evasione negli USA sia di 1 ogni 10 000 detenuti al mese.

(a) In uno stato con 40 000 detenuti, stimare con l'approssimazione di Poisson la probabilità che avvengano 8 o più evasioni in un dato mese (non calcolarne il valore numerico).

(b) Consideriamo lo stesso stato del punto (a). Qual è la probabilità che in un dato anno vi siano almeno 2 mesi nei quali avvengono almeno 8 evasioni (non calcolarne il valore numerico)?

Soluzione.

(a) Costruiamo uno schema di Bernoulli. Per $i = 1, \ldots, 40\,000$, definiamo le variabili aleatorie

$$X_i = \begin{cases} 1 & \text{se l'}i\text{-esimo detenuto evade durante il mese preso in considerazione} \\ 0 & \text{altrimenti} \end{cases}$$

Le variabili $X_1, \ldots, X_{40\,000}$ sono indipendenti (assumiamolo!) e identicamente distribuite con distribuzione $\mathrm{Be}(0.0001)$. Pertanto, nel mese preso in considerazione, il numero di evasi è dato dalla variabile aleatoria $X = \sum_{i=1}^{40\,000} X_i \sim \mathrm{Bin}(40\,000, 0.0001)$. Per l'approssimazione di Poisson, X ha approssimativamente la distribuzione di $V \sim \mathrm{Po}(4)$ e quindi

$$P(X \geq 8) \approx P(V \geq 8) = e^{-4} \sum_{k=8}^{+\infty} \frac{4^k}{k!}.$$

⚠ È lecito utilizzare l'approssimazione di Poisson dal momento che le regole euristiche che ne garantiscono la bontà (cioè $n > 100$, $p < 0.01$ e $np \leq 20$) sono tutte soddisfatte.

(b) Poniamo $p^\star = e^{-4} \sum_{k=8}^{+\infty} \frac{4^k}{k!}$. La probabilità $p^\star$ è la probabilità che in un dato mese ci siano più di 8 evasioni. Possiamo costruire un altro schema di Bernoulli. Per $i = 1, \ldots, 12$, definiamo le variabili aleatorie

$$Y_i = \begin{cases} 1 & \text{se nell'}i\text{-esimo mese ci sono state più di 8 evasioni} \\ 0 & \text{altrimenti} \end{cases}$$

Le variabili $Y_1, \ldots, Y_{12}$ sono indipendenti ed identicamente distribuite con distribuzione $\mathrm{Be}(p^\star)$. Quindi, il numero di mese in cui si verificano più di 8 evasioni è dato dalla variabile aleatoria $Y = \sum_{i=1}^{12} Y_i \sim \mathrm{Bin}(12, p^\star)$ e

$$P(Y \geq 2) = 1 - P(Y = 0) - P(Y = 1) = 1 - (1 - p^\star)^{12} - 12\,p^\star(1 - p^\star)^{11}.$$

ER.2. Siano X e Y due variabili aleatorie $\mathrm{Be}(p)$ indipendenti. Definiamo le variabili aleatorie $V = X + Y$ e $W = |X - Y|$.

(a) Calcolare la distribuzione congiunta e le distribuzioni marginali di V e W.

(b) Dire se V e W sono indipendenti o meno.

Soluzione.

(a) Poiché V e W sono definite come funzioni di X e Y, la densità congiunta di (V, W) si deve calcolare attraverso eventi riguardanti le variabili X e Y, in quanto queste sono gli oggetti noti da cui ricavare tutte le informazioni probabilistiche. Le variabili V e W assumono determinati valori, con determinate probabilità, in base ai valori assunti dalle variabili X e Y e alle probabilità con cui queste ultime li assumono. È forse utile fare una tabella per capire quali sono gli alfabeti di V e W e come i valori che assumono sono legati a quelli di X e Y:

$\boldsymbol{V}$		$X=0$	$X=1$
Y	0	0	1
	1	1	2

$\boldsymbol{W}$		$X=0$	$X=1$
Y	0	0	1
	1	1	0

Quindi l'alfabeto di V è $\mathcal{V} = \{0, 1, 2\}$ e l'alfabeto di W è $\mathcal{W} = \{0, 1\}$. Il vettore aleatorio (V, W) assume pertanto valori in $\mathcal{V} \times \mathcal{W}$. Determiniamo la densità congiunta:

$$
\begin{aligned}
p_{VW}(0,0) &= P(V = 0, W = 0) \\
&= P(X = Y = 0) && \text{(se guardiamo le tabelle, vediamo che } X = Y = 0 \text{ è l'unico caso per cui sia } V \text{ che } W \text{ risultano zero)} \\
&= P(X = 0)P(Y = 0) && \text{(indipendenza di } X \text{ e } Y) \\
&= p_X(0)p_Y(0) \\
&= (1-p)^2
\end{aligned}
$$

$$
p_{VW}(0,1) = P(V = 0, W = 1) = 0 \qquad \text{(se guardiamo le tabelle, vediamo che non ci sono coppie } (x, y) \text{ per cui, allo stesso tempo, } V = 0 \text{ e } W = 1\text{; questi due eventi sono pertanto incompatibili)}
$$

$$
\begin{aligned}
p_{VW}(1,1) &= P(V = 1, W = 1) \\
&= P(X = 0, Y = 1) + P(X = 1, Y = 0) && \text{(se guardiamo le tabelle, vediamo che } \{X = 0, Y = 1\} \text{ e } \{X = 1, Y = 0\} \text{ sono due esiti favorevoli all'evento } V = W = 1) \\
&= P(X = 0)P(Y = 1) + P(X = 1)P(Y = 0) && \text{(indipendenza di } X \text{ e } Y) \\
&= p_X(0)p_Y(1) + p_X(1)p_Y(0) \\
&= 2p(1-p)
\end{aligned}
$$

e procedendo in modo simile calcoliamo anche $p_{VW}(1, 0) = 0$, $p_{VW}(2, 0) = p_X(1)p_Y(1) = p^2$ e $p_{VW}(2, 1) = 0$. Inoltre, le densità marginali sono date da

$$
\text{marginale di } V \left|
\begin{aligned}
&p_V(0) = p_{VW}(0,0) + p_{VW}(0,1) = (1-p)^2 \\
&p_V(1) = p_{VW}(1,0) + p_{VW}(1,1) = 2p(1-p) \\
&p_V(2) = p_{VW}(2,0) + p_{VW}(2,1) = p^2
\end{aligned}
\right.
$$

e

$$
\text{marginale di } W \left|
\begin{aligned}
&p_W(0) = p_{VW}(0,0) + p_{VW}(1,0) + p_{VW}(2,0) = 1 - 2p(1-p) \\
&p_W(1) = p_{VW}(0,1) + p_{VW}(1,1) + p_{VW}(2,1) = 2p(1-p).
\end{aligned}
\right.
$$

(b) Le variabili V e W non possono essere indipendenti. Infatti, si ha

$$p_{VW}(0,1) = 0 \quad \neq \quad 2p(1-p)^3 = p_V(0)\,p_W(1).$$

ER.3. Siano X e Y due variabili aleatorie discrete con densità congiunta p_{XY} parzialmente illustrata dalla seguente tabella

$\boldsymbol{p_{XY}}$		Y: 0	1	2	p_X
X	-1	$\dots$	$\dots$	$\dots$	$\frac{1}{2}$
	1	$\dots$	$\frac{1}{2}$	$\dots$	$\frac{1}{2}$
	p_Y	$\frac{1}{6}$	$\frac{2}{3}$	$\frac{1}{6}$	1

(a) Completare la tabella. (b) Dire se X e Y sono indipendenti o meno. (c) Dire se X e Y sono correlate o meno. (d) Calcolare $Var(X+Y)$.

Soluzione.

(a) Sfruttiamo i vincoli tra densità congiunta e marginali. Dall'equazione

$$p_X(1) = p_{XY}(1,0) + \underbrace{p_{XY}(1,1)}_{=\frac{1}{2}} + p_{XY}(1,2) = \tfrac{1}{2},$$

otteniamo $p_{XY}(1,0) = p_{XY}(1,2) = 0$. Di conseguenza, la coppia di equazioni

$$p_Y(0) = p_{XY}(-1,0) + \overbrace{p_{XY}(1,0)}^{=0} = \tfrac{1}{6}$$

$$p_Y(2) = p_{XY}(-1,2) + \underbrace{p_{XY}(1,2)}_{=0} = \tfrac{1}{6}$$

implica $p_{XY}(-1,0) = p_{XY}(-1,2) = \frac{1}{6}$. Per finire, da

$$p_Y(1) = p_{XY}(-1,1) + \underbrace{p_{XY}(1,1)}_{=\frac{1}{2}} = \tfrac{2}{3},$$

ricaviamo $p_{XY}(-1,1) = \frac{1}{6}$. Riassumendo e mettendo in tabella, abbiamo ottenuto

$\boldsymbol{p_{XY}}$		Y: 0	1	2	p_X
X	-1	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{2}$
	1	0	$\frac{1}{2}$	0	$\frac{1}{2}$
	p_Y	$\frac{1}{6}$	$\frac{2}{3}$	$\frac{1}{6}$	1

(b) Le variabili X e Y non possono essere indipendenti. Infatti, si ha

$$p_{XY}(1,0) = 0 \quad \neq \quad \frac{1}{12} = \frac{1}{2}\cdot\frac{1}{6} = p_X(1)\,p_Y(0).$$

(c) Ricordiamo che $Cov(X,Y) = E(XY) - E(X)E(Y)$. Calcoliamo

$$E(XY) = (-1)\cdot 1 \cdot \frac{1}{6} + (-1)\cdot 2 \cdot \frac{1}{6} + 1 \cdot 1 \cdot \frac{1}{2} = 0$$

e

$$E(X) = -1 \cdot \frac{1}{2} + 1 \cdot \frac{1}{2} = 0 \quad \text{e} \quad E(Y) = 1 \cdot \frac{2}{3} + 2 \cdot \frac{1}{6} = 1.$$

Quindi $Cov(X,Y) = 0$ e le variabili X e Y risultano scorrelate.

⚠ Osserviamo: X e Y sono scorrelate, ma non indipendenti!

(d) In generale, $Var(X+Y) = Var(X) + Var(Y) + 2Cov(X,Y)$. Poiché nel nostro caso X e Y sono scorrelate, otteniamo $Var(X+Y) = Var(X) + Var(Y)$. Calcoliamo

$$Var(X) = (-1)^2 \cdot \frac{1}{2} + 1^2 \cdot \frac{1}{2} = 1 \quad \text{e} \quad Var(Y) = 1^2 \cdot \frac{2}{3} + 2^2 \cdot \frac{1}{6} - 1^2 = \frac{1}{3}$$

e, di conseguenza, $Var(X+Y) = 1 + \frac{1}{3} = \frac{4}{3}$.

ER.4. Un distributore viene rifornito di carburante una volta alla settimana. Il carburante venduto in una settimana (in migliaia di litri) è descritto da una variabile aleatoria con densità

$$f(x) = \begin{cases} 5(1-x)^4 & \text{se } 0 < x < 1 \\ 0 & \text{altrimenti,} \end{cases}$$

Quale dev'essere la capacità massima della stazione di rifornimento affinché la probabilità di finire le scorte in una settimana sia 0.01?

Soluzione. Siano X la variabile aleatoria che descrive la quantità di carburante venduto in una settimana (in migliaia di litri) e $0 < C < 1$ la capacità (in migliaia di litri) della stazione di rifornimento. In una data settimana, le scorte terminano se il carburante venduto supera la capacità del distributore, cioè se $X > C$. Per determinare C dobbiamo risolvere $P(X > C) = 0.01$. Poiché si ha

$$P(X > C) = \int_C^1 5(1-x)^4 = -(1-x)^5\Big|_C^1 = (1-C)^5,$$

otteniamo $(1-C)^5 = 0.01$, da cui segue $C = 1 - \sqrt[5]{0.01}$.

ER.5. Un gioco si svolge come segue. Sia U una variabile aleatoria uniforme su $[0,1]$. Se $U > 0.1$ si vince il primo round, se $U > 0.2$ si vince il secondo; mentre, se $U > 0.3$, si vince il terzo round. Calcolare la probabilità (a) di vincere il primo round; (b) di vincere il secondo round, sapendo di aver vinto il primo; (c) di vincere il terzo round, sapendo di aver vinto il primo e il secondo.

Soluzione. Otteniamo

(a) $P(U > 0.1) = 1 - F_U(0.1) = \dfrac{9}{10}$;

(b) $P(U > 0.2 | U > 0.1) = \dfrac{P(U > 0.2, U > 0.1)}{P(U > 0.1)} = \dfrac{P(U > 0.2)}{P(U > 0.1)} = \dfrac{1 - F_U(0.2)}{1 - F_U(0.1)} = \dfrac{8}{9}$;

(c) $P(U > 0.3 | U > 0.2, U > 0.1) = \dfrac{P(U > 0.3, U > 0.2, U > 0.1)}{P(U > 0.2, U > 0.1)} = \dfrac{P(U > 0.3)}{P(U > 0.2)} = \dfrac{1 - F_U(0.3)}{1 - F_U(0.2)} = \dfrac{7}{8}$.

ER.6. La direzione di un supermercato deve acquistare un nuovo freezer dei surgelati. La durata di funzionamento (in mesi) di un primo modello preso in considerazione è descritta da una variabile aleatoria esponenziale di parametro 0.01.

(a) Si valuti la durata media del freezer.

(b) Si determini la probabilità che la durata del freezer sia superiore a 200 mesi.

(c) Nell'ipotesi che il freezer sia già funzionante da 80 mesi, si calcoli la probabilità che questo funzioni per altri 200 mesi.

(d) Se la direzione deve scegliere fra il freezer di cui sopra ed un secondo modello la cui durata (in mesi) ha densità

$$f(x) = \begin{cases} \dfrac{5000}{x^3} & \text{se } x \geq 50 \\ 0 & \text{altrimenti,} \end{cases}$$

quale modello deve scegliere per avere il maggior valore atteso della durata?

Soluzione.

(a) Sia X la variabile aleatoria che descrive la durata di funzionamento (in mesi) del primo modello di freezer. Abbiamo $X \sim \text{Exp}(0.01)$ e quindi $E(X) = 100$.

(b) Otteniamo $P(X > 200) = 1 - F_X(200) = e^{-2} \approx 0.135$.

(c) Per la proprietà di assenza di memoria, si ha $P(X > 280|X > 80) = P(X > 200)$, come al punto precedente.

(d) Sia Y la variabile aleatoria che descrive la durata di funzionamento (in mesi) del secondo modello di freezer. Calcoliamo

$$E(Y) = \int_{-\infty}^{+\infty} y f(y)\, dy = \int_{50}^{+\infty} \frac{5000}{y^2}\, dy = -\frac{5000}{y}\bigg|_{50}^{+\infty} = 100.$$

Pertanto, in termini di vita media, è indifferente scegliere il primo piuttosto che il secondo freezer.

ER.7. Sia X una variabile aleatoria normale che descrive la portata del fiume Adige, nel mese di giugno, in località Boara Pisani. È noto che in tale località la portata media del fiume in giugno è pari a 243 m^3/s e che $P(X < 256) = 0.9$.

(a) Si determini la varianza di X.

(b) Si supponga che $Y = X + 10$ sia la variabile aleatoria che descrive la portata del fiume Adige in località Boscochiaro. Si calcoli $P(240 < Y < 270)$.

Soluzione.

(a) Abbiamo $X \sim \text{N}(243, \sigma^2)$. Standardizzando, otteniamo

$$P(X < 256) = P\left(Z < \tfrac{256-243}{\sigma}\right) = P\left(Z < \tfrac{13}{\sigma}\right) = \Phi\left(\tfrac{13}{\sigma}\right),$$

dove $Z = \frac{X-243}{\sigma} \sim \text{N}(0, 1)$. Per determinare il valore di σ, risolviamo

$$\Phi\left(\tfrac{13}{\sigma}\right) = 0.9 \quad \Leftrightarrow \quad \tfrac{13}{\sigma} = 1.285 \quad \Leftrightarrow \quad \sigma \approx 10,$$

da cui segue che $Var(X) = \sigma^2 \approx 100$. Pertanto $X \sim \text{N}(243, 100)$.

(b) La variabile Y è una trasformazione affine di X, pertanto risulta anch'essa una variabile aleatoria normale. In particolare, si ha $Y \sim \mathrm{N}(243+10, 100) = \mathrm{N}(253, 100)$. Calcoliamo

$$\begin{aligned} P(240 < Y < 270) &= P\left(\tfrac{240-253}{10} < Z < \tfrac{270-253}{10}\right) = P(-1.3 < Z < 1.7) \\ &= \Phi(1.7) - \underbrace{\Phi(-1.3)}_{=1-\Phi(1.3)} = 0.9554 + 0.9032 - 1 = 0.8586, \end{aligned}$$

dove $Z = \frac{Y-253}{10} \sim \mathrm{N}(0,1)$.

ER.8. Densità χ^2 (chi-quadro). Sia $Z \sim \mathrm{N}(0,1)$. Determinare la densità di $Y := Z^2$.

Soluzione. Innanzitutto osserviamo che la variabile aleatoria Y assume solamente valori positivi (essendo un quadrato). Pertanto: se $y < 0$, si ha $F_Y(y) \equiv 0$; mentre, se $y \geq 0$, si ottiene

$$F_Y(y) = P(Y \leq y) = P(Z^2 \leq y) = P(-\sqrt{y} \leq Z \leq \sqrt{y}) = \Phi(\sqrt{y}) - \underbrace{\Phi(-\sqrt{y})}_{=1-\Phi(\sqrt{y})} = 2\Phi(\sqrt{y}) - 1.$$

Poiché

$$\frac{d}{dy}F_Y(y) = \frac{d}{dy}\left[2\Phi(\sqrt{y}) - 1\right] = \frac{f_Z(\sqrt{y})}{\sqrt{y}} = \frac{e^{-\frac{y}{2}}}{\sqrt{2\pi y}},$$

possiamo concludere che la funzione di densità di Y (detta densità χ^2) è data da

$$f_Y(y) = \begin{cases} \dfrac{e^{-\frac{y}{2}}}{\sqrt{2\pi y}} & \text{se } y \overset{!}{>} 0^{*} \\ 0 & \text{altrimenti.} \end{cases}$$

ER.9. Un'indagine di mercato viene effettuata intervistando telefonicamente 60 persone. Si sa che il 70% dei chiamati non concede l'intervista.

(a) Se vengono effettuate 180 telefonate, qual è la probabilità (approssimata) di completare l'indagine?

(b) Quante telefonate si dovrebbero fare per completare l'indagine con probabilità 0.95?

Soluzione.

(a) Costruiamo uno schema di Bernoulli. Per $i = 1, \dots, 180$, definiamo le variabili aleatorie

$$X_i = \begin{cases} 1 & \text{se l'}i\text{-esima persona concede l'intervista} \\ 0 & \text{altrimenti} \end{cases}$$

Le variabili $X_1, \dots, X_{180}$ sono indipendenti ed identicamente distribuite con distribuzione $\mathrm{Be}(0.3)$. In particolare, si ha $E(X_i) = 0.3$ e $Var(X_i) = (0.3)(0.7) = 0.21$. Calcoliamo

$$\begin{aligned} P(X_1 + \cdots + X_{180} \geq 60) &= P(\overline{X}_{180} \geq \tfrac{1}{3}) = P\left(\sqrt{180}\cdot\frac{\overline{X}_{180} - 0.3}{0.458} \geq \sqrt{180}\cdot\frac{\frac{1}{3} - 0.3}{0.458}\right) \\ &\approx P(Z \geq 0.98) = 1 - \Phi(0.98) = 1 - 0.8365 = 0.1635, \end{aligned}$$

dove $Z \sim \mathrm{N}(0,1)$.

*Derivando la funzione di distribuzione di Y abbiamo ottenuto una funzione che non è definita per $y = 0$, quindi dobbiamo stare attenti a non includere tale valore nell'intervallo di definizione su cui la densità è non nulla. Dal momento che $F_Y(0) = 0$ possiamo includere, senza problemi, il valore $y = 0$ nell'altro ramo.

(b) Ora l'incognita è il numero n di telefonate che vanno effettuate per riuscire a concludere l'indagine. Usando le stesse notazioni introdotte sopra, per ricavare n dobbiamo risolvere $P(X_1 + \cdots + X_n \geq 60) = 0.95$. Abbiamo

$$\begin{aligned} P(X_1 + \cdots + X_n \geq 60) &= P\left(\overline{X}_n \geq \tfrac{60}{n}\right) = P\left(\sqrt{n} \cdot \frac{\overline{X}_n - 0.3}{0.458} \geq \sqrt{n} \cdot \frac{\frac{60}{n} - 0.3}{0.458}\right) \\ &\approx P\left(Z \geq \frac{60 - 0.3n}{0.458\sqrt{n}}\right) = 1 - \Phi\left(\frac{60 - 0.3n}{0.458\sqrt{n}}\right), \end{aligned}$$

dove $Z \sim \mathrm{N}(0, 1)$. Quindi, otteniamo l'equazione

$$\begin{aligned} 1 - \Phi\left(\frac{60 - 0.3n}{0.458\sqrt{n}}\right) = 0.95 \quad &\Leftrightarrow \quad \Phi\left(-\frac{60 - 0.3n}{0.458\sqrt{n}}\right) = 0.95 \\ &\Leftrightarrow \quad -\frac{60 - 0.3n}{0.458\sqrt{n}} = 1.645 \\ &\Leftrightarrow \quad 0.3n - 0.75341\sqrt{n} - 60 = 0 \\ &\Leftrightarrow \quad \sqrt{n} \approx 15.45 \text{ o } \sqrt{n} \approx -12.94 \text{ (non accettabile)}, \end{aligned}$$

da cui segue $n = 239$.