Probabilità e Statistica [CT0111]
Corso di Laurea triennale in Informatica, a.a. 2021/22

Docente: Isadora Antoniano Villalobos

Esame **Soluzioni**, 7 giugno 2022

Cognome: ______________________ Nome: ______________________

Matricola: ______________________ Firma: ______________________

ISTRUZIONI: DA LEGGERE CON ATTENZIONE!

Questo compito è composto di **6 domande**, per un totale di **30 punti** e dura complessivamente **90 minuti**.

Ai fini della valutazione si terrà conto solo ed esclusivamente di quanto riportato negli appositi spazi. Qualora si avesse bisogno di più spazio, sarà possibile continuare sulla prima e ultima pagina, indicandolo chiaramente.

Si richiede una traccia dello svolgimento di ogni esercizio e dei calcoli effettuati per rispondere alle domande. La sostituzione del calcolo e/o risposta numerica con un opportuno comando di R è consentita senza penalizzazione, soltanto dove indicato esplicitamente. Utilizzare almeno 4 decimali di precisione per tutti i calcoli numerici.

È consentito il solo uso della calcolatrice, della tavola della distribuzione Normale presente nel sito Moodle del corso e di un foglio A4 (entrambi lati) con formule e annotazioni. Non sono ammessi appunti, libri ed esercizi svolti.
Il compito non sarà corretto se ci sono informazioni mancanti: Cognome, Nome, Matricola o Firma

Question:	1	2	3	4	5	6	Total
Points:	5	5	5	5	5	5	30
Score:							

Domanda 1 (5 punti)
Per ognuna delle 5 domande a scelta, leggere attentamente e selezionare un'unica risposta, indicandola chiaramente. Le giustificazioni non sono richieste e, se fornite, non verranno valutate (quindi non perdere tempo).

(a) Sia $W = 3X - 2Y - 1$, con $\mathbb{E}[X] = \mathbb{E}[Y] =\text{Var}[X] =\text{Var}[Y] = 1$, $\text{Cov}(X, Y) = 0$. Quale delle seguenti affermazioni è sicuramente vera

i) $\mathbb{E}[Z] = 0$ e $\text{Var}[Z] = 5$

ii) $\mathbb{E}[Z] = 0$ ma le informazioni fornite sono insufficienti per calcolare $\text{Var}[Z]$

iii) $\mathbb{E}[Z] =\text{Var}[Z] = 0$

iv) $\mathbb{E}[Z] = 0$ e $\text{Var}[Z] = 13$

v) Nessuna delle precedenti opzioni è vera

Soluzione: *iv)*

(b) Quale delle seguenti è una funzione di ripartizione per una variabile continua?

i) Sono tutte funzioni di ripartizione per variabili continue.

ii) $F(x) = 4e^{-4x}, \quad x > 0.$

iii) $F(x) = \int_0^\infty 4e^{-4x}dx.$

iv)
$$F(x) = \begin{cases} 0, & < 0 \\ 4e^{-4x}, & x \geq 0 \end{cases}$$

v)
$$F(x) = \begin{cases} 0, & < 0 \\ 1 - e^{-4x}, & x \geq 0 \end{cases}$$

Soluzione: *v)*

(c) Se A e B sono due eventi con $A \cap B \neq \emptyset$ quale delle seguenti affermazioni è sicuramente vera?

i) $\mathbb{P}[A] > 1/2$ e $\mathbb{P}[B] > 1/2$

ii) $\mathbb{P}[A \cap B] = 0$

iii) $\mathbb{P}[A] = \mathbb{P}[B]$

iv) $\mathbb{P}[A \cup B] > 1$

v) Nessuna delle precedenti.

Soluzione: *v)*

(d) Se X e Y sono due variabili casuali indipendenti con media pari a 0, quale delle seguenti espressioni è sicuramente vera?

i) $\mathbb{E}[aX + bY + c] = 0$ per qualsiasi valore di $a, b, c \in \mathbb{R}$

ii) $\mathbb{E}[X + Y - 1] = 1$

iii) $\mathbb{E}(Y^2) = \mathbb{E}(X^2)$.

iv) $\mathbb{E}[XY] = 0$

v) Nessuna delle precedenti.

Soluzione: *iv)*

(e) Quale delle seguenti matrici può essere la matrice di transizione per una catena di Markov?

i) Tutte.

ii) Nessuna.

iii)

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

iv)

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

v)

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

Soluzione: *ii)*

Domanda 2 (5 punti)

Una data confezione di pasta pesa in media 500 grammi. Si supponga inoltre che il peso delle confezioni sia ben descritto da una variabile casuale normale con varianza pari a 4.

(a) Qual è la probabilità di produrre una confezione con almeno 506.18 grammi di pasta?

Soluzione: Sia $X \sim$N$(500, 4)$ il peso della donfezione. Dalla tavola della normale, la probabilità richiesta è

$$\mathbb{P}[X > 506, 18] = \mathbb{P}\left[\frac{X - 500}{2} > \frac{506, 18 - 500}{2}\right] = \mathbb{P}[Z > 3.09] = 1 - \mathbb{P}[Z \leq 3.09] = 0.001$$

Alternativamente, usando R, $\mathbb{P}[X > 506, 18] = 1 - \text{pnorm}(506.18, 500, 2)$

(b) Se scelgono a caso confezioni di pasta da pesare,x una alla volta, fino a trovarne una con un peso superiore a 506.18 grammi, quante confezioni ci si attende di doverer pesare?

Soluzione: Sia $W \sim$Geo(0.001) il numero di confezioni pesate prima di trovare una con più di 506.18 grammi. Allora, la quantità richiesta è $E(W) = 1/0.001 = 1000$.

(c) Supponendo di spedire un lotto di 20000 confezioni, quante confezioni ci si attende possano avere un peso maggiore di 506.18 grammi?

Soluzione: Sia $Y \sim$Bin$(20000, 0.001)$ il numero di confezioni nel lotto con un un peso maggiore di 506.18 grammi. Allora, la quantità richiesta è $E(Y) = 20000 \cdot 0.001 = 20$.

(d) Qual è la probabilità che il peso medio di 3 confezioni sia almeno 506.18 grammi?

Soluzione: Il peso medio di 3 confezioni è $\bar{X}_3 \sim$N$(500, 4/3)$. Perciò, la probabilità richiesta è

$$\mathbb{P}[\bar{X}_3 > 506,18] = \mathbb{P}\left[\frac{\bar{X}_3 - 500}{2/\sqrt{3}} > \frac{506,18 - 500}{2/\sqrt{3}}\right] = \mathbb{P}[Z > 5.352] = 1 - \mathbb{P}[Z \leq 5.352] \approx 0$$

Alternativamente, usando R, $\mathbb{P}[X > 506,18] = 1 - \text{pnorm}(506.18, 500, 2/\text{sqrt}(3))$

Domanda 3 (5 punti)

Un'urna contiene 3 dadi equilibrati e 5 dadi truccati in modo che per ciascuno di essi la probabilità di ottenere 3 sia 1/2 mentre ogni altro risultato ha probabilità 1/10.

(a) Un dado viene estratto a caso e lanciato. Qual è la probabilità di ottenere 3?

Soluzione: Definiamo gli eventi $A =$"esce il numero 3" e $T =$"il dado è truccato". Per la formula della probabilità totale, la probabilità richiesta è,

$$P(A) = P(A|T) \cdot P(T) + P(A|\bar{T}) \cdot P(\bar{T}) = \frac{1}{2} \cdot \frac{5}{8} + \frac{1}{6} \cdot \frac{3}{8} = 0.375.$$

(b) Un dado estratto a caso viene lanciato ottenendo 3. Qual è la probabilità che si tratti di un dado truccato?

Soluzione: La probabilità richiesta si calcola con il teorema di Bayes:

$$P(T|A) = \frac{P(A|T) \cdot P(T)}{P(A)} = \frac{\frac{1}{2} \cdot \frac{5}{8}}{0.375} = 0.833.$$

(c) Dall'urna si estraggono in blocco 3 dadi. Qual è la probabilità che almeno 2 di questi siano truccati? (è possibile sostituire il calcolo con un opportuno comando di R).

Soluzione: La probabilità richiesta si calcola utilizzando la distribuzione ipergeometrica, adatta alle estrazioni in blocco. Sia Y la variabile che conta il numero di dadi truccati su 3 estratti in blocco:

$$\begin{aligned} P(Y \geq 2) &= P(Y = 2) + P(Y = 3) \\ &= \frac{\binom{3}{1}\binom{5}{2}}{\binom{8}{3}} + \frac{\binom{3}{0}\binom{5}{3}}{\binom{8}{3}} = 0.536 + 0.178 = 0.714. \end{aligned}$$

Alternativamente, in R: $P(Y \geq 2) = 1 - P(Y \leq 1) = 1 - \text{phyper}(1, 5, 3, 3)$

Domanda 4 (5 punti)

Siano X e Y due variabili aleatorie discrete con funzione di probabilità congiunta

$$p(x, y) = k(y + x), \qquad x = 1, 2 \quad y = 0, 1.$$

(a) Determinare il valore di k.

Soluzione: La distribuzione congiunta di X e Y si può scrivere in una tabella:

	Y		
X	0	1	$p_X(x)$
1	k	2k	3k
2	2k	3k	5k
$p_Y(y)$	3k	5k	8k

Perciò $k = 1/8$ e la distribuzione congiunta si può riscrivere come:

	Y		
X	0	1	$p_X(x)$
1	$1/8$	$1/4$	$3/8$
2	$1/4$	$3/8$	$5/8$
$p_Y(y)$	$3/8$	$5/8$	1

(b) Dire, giustificando opportunamente, se le due variabili sono indipendenti.

Soluzione: X e Y non sono indipendenti perché, come si vede dalla tabella, $p_X(1)p_Y(0) = 9/64 \neq p(1,0) = 1/8$.

(c) Calcolare la funzione di probabilità condizionata di $X|Y = 1$.

Soluzione: La distribuzione di $X|Y = 1$ si calcola come $p_{X|Y}(x|1) = p(x,1)/p_Y(1)$:

x	1	2
$p_{X\|Y}(x\|1)$	$2/5$	$3/5$

(d) Calcolare la probabilità $P(X - Y < 1)$.

Soluzione: $P(X - Y < 1) = P(X - Y = 0) = P(X = 1, Y = 1) = 1/4$.

Domanda 5 (5 punti)

Un server viene condiviso da due utenti che lavorano in modo indipendente da remoto. In ogni minuto un utente connesso può disconnettersi con probabilità 0.4, un utente disconnesso può connettersi con probabilità 0.3 ed è impossibile che entrambi si connettano o si disconnettano nello stesso minuto. Sia X_n il numero di utenti connessi al minuto n.

(a) Qual è lo spazio degli stati per la catena di Markov $\{X_n\}$.

Soluzione: X_n rappresenta il numero di uttenti connessi, su un totale di 2, quindi lo spazio degli stati è $\{0, 1, 2\}$

(b) Si scriva la matrice di transizione associata alla catena di Markov $\{X_n\}$.

Soluzione: La matrice di transizione è

$$P = \begin{pmatrix} 0.7 & 0.3 & 0 \\ 0.4 & 0.3 & 0.3 \\ 0 & 0.4 & 0.6 \end{pmatrix}$$

(c) Se entrambi gli utenti sono connessi alle ore 10, qual è la probabilità (marginale) che nessuno sia connesso alle 10:02?

Soluzione: La matrice di transizione a due passi è

$$P^2 = PP = \begin{pmatrix} 0.61 & 0.3 & 0.09 \\ 0.4 & 0.33 & 0.27 \\ 0.16 & 0.36 & 0.48 \end{pmatrix}$$

Il vettore delle probabilità iniziali è $p^{(0)} = (0, 0, 1)$, poiché alle 10 entrambi gli utenti sono connessi, $X_0 = 2$. La distribuzione marginale a due passi è $p^{(2)} = p^{(0)}P^2 = (0.16, 0.36, 0.48)$. Dunque, $P(X_2 = 2) = 0.16$.

Domanda 6 (5 punti)

(a) Si illustri brevemente il metodo di accettazione/rifiuto per la simulazione di valori casuali.

Soluzione: La risposta corretta non è unica...

(b) Si scrivano alcuni comandi di R utili per simulare 1000 valori casuali da una distribuzione con funzione di ripartizione $F(x) = x^4$ nell'intervallo $(0, 1)$, utilizzando il metodo di accettazione/rifiuto.

Soluzione: Per prima cosa bisogna calcolare la densità da cui siamo interessati a simulare:

$$f(x) = \frac{dF(x)}{dx} = 4x^3 I_{(0,1)}(x).$$

Si possono simulare valori casuali da una densità uniforme in $(0, 1)$, $g(x)$. La costante k tale che $f(x) \leq kg(x)$ è $k = 4$. Il codice R per la simulazione è:

```
n<-0
x<-0
while (n<1000) {
    u<-runif(1)
    y<-runif(1,0,4)
    if (y < 4*u^3) {
        n<-n+1 # Si accetta
        x[n]<-y
    }
}
```

Alternativamente (ma i dati simulati potrebbero non essere 1000):

```
x<-runif(10000)
y<-runif(10000,0,4)
x<-x[y <  4*x^3] # Si spera di accettare almeno 1000
x<-x[1:1000]
```