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

Isadora Antoniano Villalobos

Facsimile Esame **Soluzioni**

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	4	6	5	6	4	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) Siano X_1 e X_2 due variabili casuali indipendenti con la stessa media pari a 0 e varianze σ_1^2 e σ_2^2, rispettivamente. Se $Y_1 = X_1 + X_2$, $Y_2 = X_1 - X_2$, quale delle seguenti affermazioni è sicuramente vera?

i) $\mathbb{E}[Y_1] > \mathbb{E}[Y_2]$ e $\mathrm{Var}[Y_1] > \mathrm{Var}[Y_2]$

ii) $Y_1 > Y_2$

iii) $\mathbb{E}[Y_1] = \mathbb{E}[Y_2] = 0$ e $\mathrm{Var}[Y_1] > \mathrm{Var}[Y_2]$

iv) $\mathrm{Cov}[Y_1, Y_2] = \sigma_1^2 - \sigma_2^2$

v) $\mathrm{Cov}[Y_1, Y_2] \leq 0$

Soluzione: *iv)*

(b) Siano $X_1, \ldots, X_n$ variabili casuali con valori nel intervallo $(0, 80)$, quale delle seguenti affermazioni è sicuramente vera?

i) $\mathbb{E}[\overline{X}_n] \leq 80$

ii) $n > 80$

iii) Nessuna delle altre affermazioni è sicuramente vera

iv) $\mathrm{Var}[\sum_{i=1}^n X_i] = \sum_{i=1}^n \mathrm{Var}[X_i]$

v) $\overline{X}_n \sim N(\mu, \sigma^2/n)$

Soluzione: *i)*

(c) Se A e B sono due eventi disgiunti, quale delle seguenti affermazioni è sicuramente vera?

i) A e B sono indipendenti

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

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

iv) $\mathbb{P}[A \cup B] = 1/2$

v) $\mathrm{Cov}[A, B] = 0$

Soluzione: *ii)*

(d) Se $A = \{4, \dagger, \heartsuit, \clubsuit\}$ e $B = \{2, 1, 4, 3\}$, allora:

i) $A \cap B = 4$

ii) $A \cup B = \{4, \dagger, \heartsuit, 1, 2, 3\}$

iii) $A \cap B = \{4, \dagger\}$

iv) $A \cup B = \emptyset$

v) $B \cap \bar{A} = \{\dagger, \heartsuit, \clubsuit\}$

Soluzione: *i)*

(e) Un giocatore ad ogni giocata vince 1 euro con probabilità 0.54 e perde 1 euro con probabilità 0.46. Comincia con 5 euro e smette di giocare quando dupplica i suoi soldi o quando li perde tutti. Si consideri la catena di Markov che rappresenta i soldi (in euro) del giocatore.

Si scelga, tra le seguenti opzioni, quella che descrive in modo giusto come simulare una traiettoria del processo per sapere se il giocatore vince o perde.

i) Si tratta di una passeggiata aleatoria con punto di partenza in 0 e barriere assorbenti in 5 e 10.
Si può simulare una traiettoria della catena usando il seguente codice

```
n <- 100
x <- c(5, rbinom(n-1,1,0.54)*2-1)
y <- cumsum(x)
t1<-min(which(y==0 | y==10))
y[t1:n]<- y[t1]
```

Il valore di y[t1] ci indica se il giocatore ha duplicato i suoi soldi o se li ha persi tutti.

ii) Si tratta di una passeggiata aleatoria con punto di partenza in 5 e barriere assorbenti in 0 e 10.
Si può simulare una traiettoria della catena usando il seguente codice

```
n <- 100
x <- c(5, rbinom(n-1,1,0.54)*2-1)
y <- cumsum(x)
t1<-min(which(y==0 | y==10))
```

Il valore di t1 ci indica se il giocatore ha duplicato i suoi soldi o se li ha persi tutti.

iii) Si tratta di una passeggiata aleatoria con punto di partenza in 0 e barriere assorbenti in 5 e 10.
Si può simulare una traiettoria della catena usando il seguente codice

```
n <- 100
x <- c(0, rbinom(n-1,1,0.54)*2-1)
y <- cumsum(x)
t1<-min(which(y==5 | y==10))
```

Il valore di t1 ci indica se il giocatore ha duplicato i suoi soldi o se li ha persi tutti.

iv) Nessuna delle altre opzioni descrive in modo giusto come simulare una traietoria del processo per sapere se il giocatore vince o perde.

v) Si tratta di una passeggiata aleatoria con punto di partenza in 5 e barriere assorbenti in 0 e 10.
Si può simulare una traiettoria della catena usando il seguente codice

```
n <- 100
x <- c(5, rbinom(n-1,1,0.54)*2-1)
y <- cumsum(x)
t1<-min(which(y==0 | y==10))
y[t1:n]<- y[t1]
```

Il valore di y[t1] ci indica se il giocatore ha duplicato i suoi soldi o se li ha persi tutti.

Soluzione: *v)*

Domanda 2 (4 punti)
In un'indagine sulla soddisfazione dei clienti di una nota marca di computer che produce anche tablet, si è calcolato che il 60% di coloro che possiedono un tablet è soddisfatto, mentre l' 80% di coloro che non possiedono un tablet è soddisfatto. Inoltre si è calcolato che il 70% dei clienti possiede un tablet.

(a) Determinare la probabilità che un cliente sia soddisfatto.

Soluzione: Definiamo due eventi $A =$"cliente soddisfatto" e $B =$"cliente che possiede un tablet".
Si ha che $P(A|B) = 0.6$, $P(A|\bar{B}) = 0.8$, $P(B) = 0.7$ da cui $P(\bar{B}) = 0.3$.
Per la legge delle probabilità totali si trova $P(A) = P(A|B)P(B) + P(A|\bar{B})P(\bar{B}) = 0.6 * 0.7 + 0.8 * 0.3 = 0.66$.

(b) Determinare la probabilità che un cliente che si è dichiarato soddisfatto non possieda un tablet.

Soluzione: Per il Teorema di Bayes si ha

$$P(\bar{B}|A) = \frac{P(A|\bar{B})P(\bar{B})}{P(A)} = \frac{0.8 * 0.3}{0.66} = 0.3636.$$

Domanda 3 (6 punti)
Un'azienda produce cavi ethernet di lunghezza (in cm) distribuita come una normale di media $\mu_c = 100$ e varianza $\sigma_c^2 = 16$. Scelto a caso un cavo ethernet,

(a) si calcoli la probabilità che la lunghezza del cavo sia superiore a 105 cm; (è possibile sostituire il calcolo con un opportuno comando di R)

Soluzione: $X =$ lunghezza cavi d'acciaio $X \sim N(100, 16)$.
La probabilità che un cavo sia superiore a 105 cm è

$$\begin{aligned} P(X > 105) &= 1 - P\left(X \leq 105\right) \\ &= 1 - P\left(Z \leq \frac{105 - 100}{4}\right) = 1 - P(Z \leq 1.25) = \\ &= 1 - 0.89435 = 0.10565 \end{aligned}$$

La probabilità richiesta è 0.10565. Si può anche ottenere in R con il comando `1-pnorm(105,100,4)`.

(b) Quale lunghezza è superata con una probabilità pari a 0.5? (è possibile sostituire il calcolo con un opportuno comando di R)

Soluzione: La risposta è $x_{0.5} = 100$ in quanto X è normale e dunque la media e la mediana coincidono.

(c) Supponendo di scegliere casualmente 6 cavi tra quelli prodotti, qual è la probabilità che almeno uno abbia lunghezza superiore a 105 cm? (è possibile sostituire il calcolo con un opportuno comando di R)

Soluzione: Y = numero di cavi di lunghezza superiore a 105 cm prodotti su 6 scelti a caso tra quelli prodotti $Y \sim Bin(6, 0.10565)$.
La probabilità richiesta corrisponde a

$$P(Y \geq 1) = 1 - P(Y < 1) = 1 - P(Y = 0) = \\ = 1 - 0.5117 = 0.4883;$$

che si ottiene utilizzando la distribuzione binomiale di Y :

$$P(Y = k) = \binom{6}{k} 0.10565^k (1 - 0.10565)^{6-k}.$$

Si può anche calcolare in R con il comando `1-dbinom(0,6,p)`, dove p è la probabilità calcolata al punto (a).

(d) La durata in ore delle batterie prodotte dalla stessa azienda segue una distribuzione normale di media $\mu_b = 500$ ore e varianza σ_b^2 ignota. Quanto vale la varianza se con una probabilità pari a 0.99 la durata media di un campione di $n = 15$ batterie è maggiore di 495?

Soluzione: Siano $B \sim (500, \sigma^2)$ la durata di una batteria e $\bar{B}_{15}$ la media campionaria su un campione di numerosità 15. Allora $\bar{B}_{15} \sim N(500, \sigma^2/15)$. La condizione richiesta è che

$$\begin{aligned} 0.99 &= P\left(\bar{B}_{15} > 495\right) \\ &= 1 - P\left(Z \leq \sqrt{15}\frac{495 - 500}{\sigma_b}\right) = 1 - P\left(Z \leq \frac{-19.365}{\sigma_b}\right) = \\ &= P\left(Z \leq \frac{19.365}{\sigma_b}\right). \end{aligned}$$

Allora $19.365/\sigma_b = z_{0.99} = 2.326$ e $\sigma_b = 8.324$. La varianza richiesta è dunque 69.292.

Domanda 4 (5 punti)
Siano X e Y due variabili aleatorie con densità congiunta $f(x, y) = \lambda^2 e^{-\lambda y} I_{(0,y)}(x) I_{(0,+\infty)}(y)$.

(a) Calcolare la densità marginale di Y. Si tratta di una distribuzione nota?

Soluzione: $f_Y(y) = \int_0^y \lambda^2 e^{-\lambda y} I_{(0,+\infty)}(y) dx = \lambda^2 y e^{-\lambda y} I_{(0,+\infty)}(y)$. Si tratta di una Gamma con parametri $\alpha = 2$ e λ.

(b) Calcolare la densità condizionata di X dato $Y = y$. Si tratta di una distribuzione nota?

Soluzione: $f_{X|Y}(x|y) = f(x, y)/f_Y(y) = 1/y I_{(0,y)}(x)$. E' una distribuzione Uniforme nell'intervallo $(0, y)$.

(c) X e Y sono indipendenti? Giustificare adeguatamente la risposta.

Soluzione: Le due variabili non sono indipendenti. Infatti il loro dominio è una specie di triangolo in cui il dominio di una variabile dipende dai valori che l'altra assume. Inoltre la densità di $X|Y = y$ dipende da y e perciò è sicuramente diversa dalla marginale della X.

Domanda 5 (6 punti)

Sia $\{X_n\}$ una catena di Markov con spazio degli stati $\{0, 1, 2\}$ e con le seguenti probabilità di transizione: $p_{00} = 1$, $p_{10} = 1/4$, $p_{12} = 3/4$, $p_{22} = 1$. Lo stato iniziale è scelto a caso.

(a) Si calcoli la matrice di transizione a due passi.

Soluzione: La matrice di transizione è

$$P = \begin{pmatrix} 1 & 0 & 0 \\ 0.25 & 0 & 0.75 \\ 0 & 0 & 1 \end{pmatrix}$$

e la matrice di transizione a due passi è

$$P^2 = PP = \begin{pmatrix} 1 & 0 & 0 \\ 0.25 & 0 & 0.75 \\ 0 & 0 & 1 \end{pmatrix}$$

ancora uguale a P.

(b) Si calcoli $P(X_2 = 0)$.

Soluzione: Se il vettore delle probabilità iniziali è $p^{(0)} = (1/3, 1/3, 1/3)$, la distribuzione marginale a due passi è $p^{(2)} = p^{(0)}P^2 = (5/12, 0, 7/12)$. Dunque, $P(X_2 = 0) = 5/12$.

(c) Si calcoli una distribuzione stazionaria della catena.

Soluzione: Le distribuzioni stazionarie della catena soddisfano la condizione $\pi = \pi P$ che dà origine ad un sistema lineare con tre equazioni e tre incognite, π_0, π_1, π_2. Aggiungendo la condizione $\sum_i \pi_i = 1$, si ottiene $\pi = (\pi_0, 0, \pi_2)$, con $\pi_0 + \pi_2 = 1$. Ad esempio, $\pi = (0.3, 0, 0.7)$ è stazionaria.

Domanda 6 (4 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 densità $f(x) = 3x^2 I_{(0,1)}(x)$, utilizzando il metodo di accettazione/rifiuto.

Soluzione: 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 $k = 3$. Il codice R per la simulazione è:

```
x<-runif(10000)
y<-runif(10000,0,3)
x<-x[y <  3*x^2]
x<-x[1:1000]
```