CdL in Ing. Robotica e dell'Automazione, Prob. e Proc. Stoc. (455AA)

A.A. 2022/23 - Appello del 2023-09-15

La durata della prova è di **150 minuti**. *Fornire risposte dettagliate.*

Problema 1

Il numero di persone che prendono l'autobus alla fermata principale della città in una data ora di un giorno feriale ha una distribuzione Poisson di un parametro λ che dipende dall'orario. Precisamente, si suppone che sia $\lambda = 100$ in una della quattro ore di punta del giorno (precisamente le ore 7:00-8:00, 13:00-14:00, 16:00-17:00 e 18:00-19:00) mentre si ha $\lambda = 60$ in tutte le altre 20 ore (anche quelle notturne). Si assuma che il numero di passeggeri osservati in ore diverse siano indipendenti tra loro.

1. Calcolare il valor medio e la varianza del numero totale di persone che prendono l'autobus alla fermata, in un intero giorno feriale.
2. Qual è la probabilità che alla fermata salgano esattamente 80 persone in un'ora presa a caso di un giorno feriale?
3. Un addetto al controllo qualità del servizio dei trasposti ha registrato che in un'ora di un giorno feriale sono salite 80 persone, però non ricorda più se era un'ora di punta oppure no. Sapendo che l'addetto prende sempre servizio alle 9:00 e lavora fino alle 17:00, è più probabile che abbia registrato il dato in un'ora di punta oppure no?

Una soluzione:

1. Il numero totale di ore in un giorno feriale è 24. Le ore di punta sono 4, con $\lambda = 100$. Le altre ore hanno $\lambda = 60$. Il valore medio totale del numero X di persone salite è dato dalla somma dei valori medi orari:

$$\mathbb{E}[X] = 4 \cdot 100 + 20 \cdot 60 = 400 + 1200 = 1600.$$

La varianza totale è data dalla somma delle varianze orarie (essendo le variabili indipendenti):

$$\operatorname{Var}(X) = 4 * 100 + 20 \cdot 60 = 1600,$$

che si può anche ottenere ricordando che la somma di Poisson indipendenti tra loro è una variabile Poisson (con parametro dato dalla somma dei parametri).

2. La probabilità che in un'ora qualsiasi salgano esattamente 80 persone (evento A) è data da una densità Poisson con $\lambda = 60$ o 100 a seconda dell'ora. Precisamente

$$P(A|\lambda) = e^{-\lambda}\frac{\lambda^{80}}{80!} \approx \begin{cases} 0,005 & \text{se } \lambda = 100, \\ 0,002 & \text{se } \lambda = 60. \end{cases}$$

Poiché la probabilità richiesta è per un'ora presa a caso, la risposta è la media (pesata) delle due probabilità e vale quindi

$$P(A) = 20 \cdot e^{-60}\frac{60^{80}}{80!} + 4 \cdot e^{-100}\frac{100^{80}}{80!} \approx 20 \cdot 0,002 + 4 \cdot 0,005 \approx 6\%.$$

3. L'addetto lavora dalle 9:00 alle 17:00 (8 ore). In questo intervallo ci sono le ore 13:00-14:00 e 16:00-17:00 che sono ore di punta con $\lambda = 100$. Per la formula di Bayes, posto $B =$ "ha osservato un'ora di punta" avremo

$$\frac{P(B|A)}{P(\text{non } B|A)} = \frac{P(A|B)P(B)}{P(A|\text{non } B)P(\text{non } B\ )}.$$

Dato l'orario di lavoro dell'addetto abbiamo

$$\frac{P(B)}{P(\text{non } B))} = \frac{\frac{2}{8}}{\frac{6}{8}} = \frac{1}{3}$$

mentre abbiamo calcolato prima sia la probabilità di registrare 80 passeggeri in un'ora di punta ($\approx 0,005$) che la probabilità di registrare 80 passeggeri in un'ora non di punta ($\approx 0,002$). Ne segue che

$$\frac{P(B|A)}{P(\text{non } B|A)} \approx \frac{1}{3} \cdot \frac{e^{-100}\frac{100^{80}}{80!}}{e^{-60}\frac{60^{80}}{80!}} = \frac{1}{3} \cdot e^{-40}\left(\frac{5}{3}\right)^{80}.$$

Per sapere se questo rapporto è più grande o più piccolo di 1 basta osservare che

$$\frac{1}{3}e^{-40}\left(\frac{5}{3}\right)^{80} = \frac{1}{3}\left(\frac{5}{3\cdot\sqrt{e}}\right)^{80} \approx \frac{1}{3}1,01^{80} \approx \frac{2}{3} < 1.$$

Quindi è più probabile che l'addetto abbia registrato un'ora *non* di punta.

Problema 2

Si consideri un triangolo T avente base di lunghezza B e altezza (relativa alla base) H, dove B ed H sono variabili aleatorie esponenziali aventi media 5 cm e indipendenti tra loro. Si denoti con X la variabile aleatoria che descrive l'area del triangolo T.

1. Calcolare il valor medio e la deviazione standard di X.
2. Dire per quali $\alpha \in \mathbb{R}$ si ha $\mathbb{E}[X^\alpha] < \infty$.
3. Mostrare che per ogni $t > 0$ la funzione generatrice dei momenti di X è infinita, $MGF_X(t) = \infty$. *(Sugg: si ricordi per quali t la MGF di una variabile esponenziale è infinita.)*

Una soluzione:

1. L'area del triangolo T è data da $X = BH/2$. Poiché B ed H sono indipendenti con media 5 cm, si ha:

$$\mathbb{E}[X] = \mathbb{E}[BH/2] = \mathbb{E}[B]\,\mathbb{E}[H]/2 = 5\cdot 5/2 = 12.5 \text{ cm}^2$$
$$\operatorname{Var} X = \mathbb{E}\left[B^2H^2/4\right] - \mathbb{E}[X]^2 = \mathbb{E}[B]^2\,\mathbb{E}[H]^2/4 = \frac{3}{4}\cdot 5^4 \text{ cm}^4$$

dove abbiamo usato il fatto noto che $\mathbb{E}\left[B^2\right] = \operatorname{Var} B + \mathbb{E}\left[B^2\right] = 2\cdot 5^2$ (e similmente per H). Pertanto $\sigma_X = \sqrt{\operatorname{Var} X} \approx 21$ cm.

2. Perché $\mathbb{E}[X^\alpha] < \infty$ è necessario che $\alpha > -1$. Infatti, per ogni α, usando il fatto che B e H sono indipendenti e hanno la stessa legge, troviamo

$$\mathbb{E}[X^\alpha] = \mathbb{E}[(BH/2)^\alpha] = 2^{-\alpha}\mathbb{E}[B^\alpha]\,\mathbb{E}[H^\alpha] = 2^{-\alpha}\mathbb{E}[B^\alpha]^2.$$

Usando l'espressione per la densità esponenziale, troviamo infine

$$\mathbb{E}\left[B^{\alpha}\right] = \int_0^{\infty} b^{\alpha} e^{-b/5} \frac{db}{5}.$$

I valori positivi di α non presentano problemi di convergenza, grazie all'esponenziale con esponente negativo. Invece se $\alpha < 0$ dobbiamo tenere conto della singolarità per $b \to 0$. Ricordando fatti noti dell'analisi (per confronto con la funzione x^{α}) si trova che deve essere $\alpha > -1$.

3. Ricordiamo che data una variabile esponenziale U di media m la sua $MGF_U(t)$ è infinita se $mt \geq 1$, perché

$$MGF_U(t) = \int_0^{\infty} e^{tu} e^{-u/m} \frac{du}{m} = \int_0^{\infty} e^{(t-1/m)u} \frac{du}{m} = \infty$$

e l'integrale diverge essendo l'esponente $t - 1/m \geq 0$. Dovendo calcolare la MGF di X, osserviamo che dato un qualsiasi $t > 0$ possiamo sempre decomporre rispetto alle alternative $A = \{Ht/2 \geq 1/5\}$ e non A (osserviamo che $P(A) > 0$ anche se t è molto piccolo, ma strettamente positivo). Scriviamo quindi

$$\begin{aligned} MGF_X(t) &= \mathbb{E}\left[e^{tBH/2}|A\right] P(A) + \mathbb{E}\left[e^{tBH/2}|\text{non } A\right] P(\text{non } A) \\ &\geq \mathbb{E}\left[e^{tBH/2}|A\right] P(A) \geq \mathbb{E}\left[e^{B/5}|A\right] P(A) \\ &= MGF_B(1/5) \cdot P(A) = \infty. \end{aligned}$$

Problema 3

Si consideri una classe universitaria di 100 studenti. Ogni giorno ciascuno studente si può trovare in uno di due stati: presente o assente. Lo studente è presente ad ogni lezione per un numero casuale di giorni (con distribuzione geometrica di parametro $a \in (0,1)$) e poi si assenta per malattia, vacanze ecc. con probabilità $(1-a)^{n-1}a$ di assentarsi il giorno n. Quando lo studente si assenta, rimane a casa per un tempo casuale (con distribuzione geometrica simile, ma di un parametro $b \in (0,1)$) e poi ritorna in classe.

1. Costruire una catena di Markov a due stati per descrivere la presenza/assenza di un singolo studente al variare dei giorni.
2. Calcolare le distribuzioni invarianti della catena introdotta al punto precedente, in funzione di a e b.
3. Supponendo che lo stato di ciascuno studente sia indipendente dagli altri, si osserva in un dato giorno che vi sono 75 studenti presenti su 100. È possibile stimare tramite massima verosimiglianza a e b sulla base di questo dato?

Una soluzione:

1. La catena è molto semplice, con matrice di transizione (ordinando gli stati 0 =*presente*, 1 =*assente*):

$$\begin{pmatrix} 1-a & a \\ b & 1-b \end{pmatrix}.$$

Si può verificare che la probabilità che uno studente presente inizialmente si assenti il giorno n è data da un cammino con peso $(1-a)^{n-1}a$, e similmente per il ritorno in classe.

2. Per trovare le distribuzioni invarianti, impostiamo il bilancio di flusso:

$$\pi_0 a = b\pi_1$$

da cui $\pi_1 = \pi_0 a/b$, e quindi

$$\pi = \pi_0(1, a/b), \quad \Rightarrow \quad \pi = \frac{1}{a+b}(b, a).$$

3. Possiamo supporre che la catena di Markov associata a ciascuno studente sia stazionaria e perciò la probabilità che sia presente è $p = b/(a+b)$. Avendo osservato 100 studenti di cui 75 presenti, per indipendenza la verosimiglianza è

$$L(a,b) = p^{75}(1-p)^{25} = \frac{b^{75}a^{25}}{(a+b)^{100}}.$$

Passando al logaritmo e derivando con i soliti calcoli si ottiene

$$p = 3/4$$

da cui

$$4b = 3(a+b) \Leftrightarrow b = 3a,$$

che è una relazione utile ma non permette di determinare unicamente a e b.