CdL in Ing. Robotica e dell'Automazione, Prob. e Proc. Stoc. (455AA)
A.A. 2023/24 - Appello 2024-06-25

La durata della prova è di **120 minuti**. *Fornire risposte dettagliate.*

Problema 1

Un'agenzia di viaggi organizza gite per gruppi di clienti. Ci sono due tipologie di gite: gite culturali (C) e gite naturalistiche (N). I clienti dell'agenzia possono essere di due categorie: amanti della cultura e amanti della natura.

Gli amanti della cultura scelgono le gite culturali C con probabilità 90%, mentre gli amanti della natura scelgono le gite naturalistiche N con probabilità 80% (indipendentemente dalle precedenti). Si stima inoltre che il 30% dei clienti siano amanti della cultura, il 70% siano amanti della natura.

Preso un cliente a caso,

1. calcolare la probabilità che in 3 gite abbia scelto due volte la gita culturale C e una volta la naturalistica N;
2. sapendo che su 3 gite ha scelto due volte la gita culturale C e una volta la naturalistica N, calcolare la probabilità che sia un amante della cultura;
3. sapendo che su 3 gite ha scelto due volte la gita culturale C e una volta la naturalistica N, calcolare la probabilità che alla gita successiva scelga la gita culturale.

Una soluzione:

1. Poniamo $X \in \{C, N\}$ una variabile che indica il tipo di cliente: $X = C$ è amante della cultura, $X = N$ se amante della natura, e Y il numero di gite culturali scelte su 3 volte. Vale $P(X = C) = 3/10$, $P(X = N) = 7/10$. Abbiamo che, condizionatamente a $X = C$, Y è binomiale di parametri $(3, 9/10)$, mentre condizionatamente a $X = N$, Y è binomiale di parametri $(3, 2/10)$. Pertanto

$$\begin{aligned} P(Y = 2) &= P(Y = 2|X = C)P(X = C) + P(Y = 2|X = N)P(X = N) \\ &= 3\left(\frac{9}{10}\right)^2 \frac{1}{10} \cdot \frac{3}{10} + 3\left(\frac{2}{10}\right)^2 \frac{8}{10} \cdot \frac{7}{10} \\ &\approx 14\% \end{aligned}$$

2. Per la formula di Bayes, troviamo

$$\begin{aligned} P(X = C|Y = 2) &= \frac{P(Y = 2|X = C)P(X = C)}{P(Y = 2)} \\ &= \frac{3\left(\frac{9}{10}\right)^2 \frac{1}{10} \cdot \frac{3}{10}}{3\left(\frac{9}{10}\right)^2 \frac{1}{10} \cdot \frac{3}{10} + 3\left(\frac{2}{10}\right)^2 \frac{8}{10} \cdot \frac{7}{10}} \\ &\approx 52\%. \end{aligned}$$

3. Poniamo $Z \in \{C, N\}$ la scelta della quarta gita del cliente. Per la formula di decomposizione (tenendo presente l'informazione $I = \{Y = 2\}$ troviamo

$$\begin{aligned} P(Z = C|Y = 2) &= P(Z = C|X = C, Y = 2)P(X = C|Y = 2) + P(Z = C|X = N, Y = 2)P(X = N|Y = \\ &= P(Z = C|X = C)P(X = C|Y = 2) + P(Z = C|X = N)P(X = N|Y = 2) \\ &= \frac{9}{10}P(X = C|Y = 2) + \frac{2}{10}(1 - P(X = C|Y = 2)), \end{aligned}$$

per cui ora basta solo inserire l'espressione per $P(X = C|Y = 2)$ trovata al punto precedente. Con l'approssimazione $P(X = C|Y = 2) \approx 52\%$ troviamo

$$P(Z = C|Y = 2) \approx 56\%.$$

Problema 2

Sia X una variabile aleatoria uniforme (continua) su un intervallo $[a, b] \subseteq \mathbb{R}$ (dove $a < b$ sono parametri) e si ponga $Y = X^2$. Al variare di a e b,

1. descrivere la densità di Y,
2. calcolare la covarianza tra X e Y,
3. dire se X e Y sono indipendenti.

Una soluzione:

1. Si tratta di applicare la formula di cambio di variabile con $g(x) = x^2$ sull'intervallo $[a, b]$. Dobbiamo distingure 3 casi (poi in realtà si riducono a due): $a < b \leq 0$, $0 \leq a < b$ e il terzo $a \leq 0 \leq b$ in cui g non è invertibile. In ogni caso la densità di Y è positiva solo nell'intervallo $g([a, b])$, che è uguale a $[b^2, a^2]$ nel primo caso, $[a^2, b^2]$ nel secondo e $[0, \max(a^2, b^2)]$ nel terzo.

La derivata di g è $g'(x) = 2x$ e l'immagine inversa di $y \geq 0$ è data da $x = \left\{\pm\sqrt{y}\right\}$. Tenendo presente i tre casi sopra si trovano le formule:

$$p(Y = y) = \begin{cases} \frac{1}{(b-a)2\sqrt{y}} & \text{per } b^2 < y < a^2, \\ 0 & \text{altrimenti.} \end{cases}$$

nel primo caso. Nel secondo caso,

$$p(Y = y) = \begin{cases} \frac{1}{(b-a)2\sqrt{y}} & \text{per } a^2 < y < b^2, \\ 0 & \text{altrimenti.} \end{cases}$$

Nel terzo caso,

$$p(Y = y) = \begin{cases} \frac{1}{(b-a)\sqrt{y}} & \text{per } 0 \leq y \leq \min\left\{a^2, b^2\right\}, \\ \frac{1}{(b-a)2\sqrt{y}} & \text{per } \min\left\{a^2, b^2\right\} < y \leq \max\left\{a^2, b^2\right\}, \\ 0 & \text{altrimenti.} \end{cases}$$

2. Basta usare la formula $\operatorname{Cov}(X, Y) = \mathbb{E}XY - \mathbb{E}X\mathbb{E}Y$, per cui $\mathbb{E}X = (a + b)/2$,

$$\mathbb{E}Y = \mathbb{E}X^2 = \int_a^b x^2 \frac{dx}{b-a} = \frac{b^3 - a^3}{3(b-a)} = \frac{b^2 + ab + a^2}{3}$$

e

$$\mathbb{E}XY = \mathbb{E}X^3 = \int_a^b x^3 \frac{dx}{b-a} = \frac{b^4 - a^4}{4(b-a)} = \frac{(b^2 + a^2)(a + b)}{4}.$$

Mettendo insieme troviamo

$$\begin{aligned} \operatorname{Cov}(X, Y) &= \frac{b^4 - a^4}{4} - \frac{(a + b)(b^3 - a^3)}{6} = \frac{(b^2 + a^2)(a + b)}{4} - \frac{(b^2 + ab + a^2)(a + b)}{6} \\ &= \frac{1}{12}(a + b)\left(b^2 + a^2 - 2ab\right) = \frac{(a - b)^2(a + b)}{12}. \end{aligned}$$

3. Dalla formula sopra vediamo che la covarianza non è mai nulla a meno che $a = -b$. Se la covarianza non è nulla le variabili non sono indipendenti. Nel caso $a = -b$ comunque le variabili non sono indipendenti. Ad esempio basta osservare che

$$P(Y > a^2/4) = P(|X| > a/2) = 1/4$$

ma

$$P(Y > a^2/4||X| \leq a/2) = P(|X| > a/2||X| \leq a/2) = 0.$$

Problema 3

Si consideri una catena di Markov $(X_n)_{n\geq 0}$ sull'insieme degli stati $\{1,2,3\}$ con matrice di transizione

$$Q = \begin{pmatrix} 0 & p & (1-p) \\ 1/4 & 1/2 & 1/4 \\ (1-p) & p & 0 \end{pmatrix}$$

dove $p \in [0,1]$ è un parametro.

1. Al variare di p, classificare gli stati della catena e calcolarne tutte le distribuzioni invarianti.
2. Supponendo che $X_0 = 1$ e $X_2 = 2$, dire se è più probabile che sia $X_1 = 2$ o $X_1 \neq 2$ (la risposta può variare in funzione di p).
3. Supponendo che la catena sia stazionaria, si osserva che $X_2 = 2$. È possibile fornire una stima per p?

Una soluzione:

1. La catena è chiaramente invariata se sostituiamo lo stato 1 con lo stato 3. Ne segue che $\pi_1 = \pi_3$ per la distribuzione invariante. Imponiamo il bilancio di flusso nello stato 1:

$$\pi_1 = \frac{1}{4}\pi_2 + (1-p)\pi_3$$

da cui usando che $\pi_1 = \pi_3$ troviamo

$$p\pi_1 = \frac{1}{4}\pi_2,$$

e quindi $\pi_2 = 4p\pi_1$. Troviamo quindi un vettore $\propto (1, 4p, 1)$ e quindi esplicitamente $\pi = \frac{1}{2(1+2p)}(1, 4p, 1)$.

2. Supponendo $X_0 = 1$ basta contare i pesi dei cammini lunghi due che terminano in 2, che sono solamente $1 \to 2 \to 2$, con probabilità $p/2$ e $1 \to 3 \to 2$, con probabilità $(1-p)p$. La probabilità che $X_1 = 2$ vale quindi

$$P(X_1 = 2|X_0 = 1, X_2 = 2) = \frac{p/2}{p/2 + (1-p)p}$$

Poiché $p/2 > (1-p)p$ se $1/2 > 1-p$, ossia $p > 1/2$, segue che è più probabile che sia $X_1 = 2$ nel caso $p > 1/2$.

3. Supponendo la catena stazionaria, la probabilità dell'evento osservato è $2p/(1+2p)$, che quindi può servire da verosimiglianza. Massimizzando la verosimiglianza troviamo quindi che $p = 1$.