CdL in Ing. Robotica e dell'Automazione, Prob. e Proc. Stoc. (455AA)
A.A. 2024/25 - Prova in itinere 2024-11-28

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

Problema 1

Un'urna contiene una frazione $p \in [0,1]$ di palline rosse, e le rimanenti blu. Si effettuano estrazioni con rimpiazzo finché non si osservano, in tutta la sequenza delle palline estratte, due palline di colore uguale (non necessariamente consecutive). Si pone T il numero di estrazioni effettuate.

1. Al variare del parametro p, determinare la densità discreta di T.
2. Al variare del parametro p, calcolare il valore atteso di T.
3. Fornire una stima di massima verosimiglianza per p supponendo di aver osservato $T = 2$. Cosa cambia se invece si osserva $T = 3$?

Una soluzione:

1. La variabile T assume solamente valori in $\{2,3\}$. Infatti dopo tre estrazioni sicuramente avremo visto almeno due palline dello stesso colore (quindi non sarà necessario farne 4 o piu). Poniamo $X_1, X_2, X_3 \in \{R, B\}$ gli esiti di tre estrazioni con rimpiazzo dall'urna. Abbiamo che

$$\{T = 2\} = \{X_1 = X_2 = R\} \text{ oppure } \{X_1 = X_2 = B\}$$

e quindi

$$P(T = 2) = p^2 + (1-p)^2.$$

D'altra parte vale quindi $P(T = 3) = P(T \neq 2) = 1 - P(T = 2) = 1 - p^2 - (1-p)^2$.

3. Usiamo la densità trovata per il calcolo del valore atteso:

$$\mathbb{E}[T] = 2P(T = 2) + 3P(T = 3) = 2(p^2 + (1-p)^2) + 3(1 - p^2 - (1-p)^2) = 3 - (p^2 + (1-p)^2).$$

4. Si tratta di massimizzare $L(p; T = 2) = P(T = 2|p) = p^2 + (1-p)^2$. Osserviamo che per $p \in \{0,1\}$ la verosimiglianza vale 1, quindi è massima in quei casi (la stima di massima verosimiglianza non è unica). Invece osserviamo che $L'(p; T = 2) = 2p - 2(1-p)2(2p-1) = 0$ se $p = 1/2$, ma in tal punto la funzione è minima (vale 1/2). Se si osserva $T = 3$, invece la verosimiglianza è $L(p; T = 3) = 1 - L(p; T = 2)$ che quindi è massima se $p = 1/2$, che è quindi la stima di massima verosimiglianza in questo caso.

Problema 2

Siano T_1, T_2 variabili aleatorie continue, entrambe aventi densità esponenziale di parametro 1 e indipendenti fra loro.

1. Dire se la variabile $U = \min\{T_1, T_2\}$ ha densità esponenziale e calcolarne il valore atteso.
2. Dire se la variabile $V = \max\{T_1, T_2\}$ ha densità esponenziale e calcolarne il valore atteso.
3. Le variabili aleatorie U e V sono correlate? sono indipendenti?

Una soluzione:

1. Calcoliamo la funzione di sopravvivenza di U:

$$\mathrm{SUR}_U(t) = P(U > t) = P(T_1 > t, T_2 > t) = \mathrm{SUR}_{T_1}(t)\,\mathrm{SUR}_{T_2}(t).$$

ricordando che per una variabile esponenziale si ha $\mathrm{SUR}_T(t) = e^{-t}$ per $t \geq 0$ (mentre $\mathrm{SUR}_T(t) = 1$ per $t < 0$) troviamo che per $t \geq 0$

$$\mathrm{SUR}_U(t) = e^{-2t}$$

da cui la densità di U (derivando e cambiando di segno) è esponenziale di parametro 2, quindi $\mathbb{E}[U] = 1/2$

2. Calcoliamo la funzione di ripartizione di V (per $t \geq 0$)

$$\mathrm{CDF}_V(t) = P(T_1 \leq t, T_2 \leq t) = \mathrm{CDF}_{T_1}(t)^2 = (1 - e^{-t})^2.$$

Derivando si ottiene la densità

$$p(V = t) = 2(1 - e^{-t})e^{-t} = 2e^{-t} - 2e^{-2t}$$

che non è del tipo esponenziale. Si trova il valore atteso

$$\mathbb{E}[V] = \int_0^\infty tp(V = t)dt = 2\int_0^\infty te^{-t} - \int_0^\infty t \cdot 2e^{-2t}dt = 2 - 1/2 = 3/2.$$

3. È facile vedere che le variabili U e V non sono indipendenti: infatti vale banalmente $U \leq V$, quindi

$$P(U > 1 | V \leq 1) = 0 \neq P(U > 1) = e^{-2}.$$

dal punto 1 calcolato prima. Per calcolare la covarianza, osserviamo che vale sempre $UV = T_1 T_2$, quindi

$$\mathbb{E}[UV] = \mathbb{E}[T_1 T_2] = \mathbb{E}[T_1]\,\mathbb{E}[T_2] = 1$$

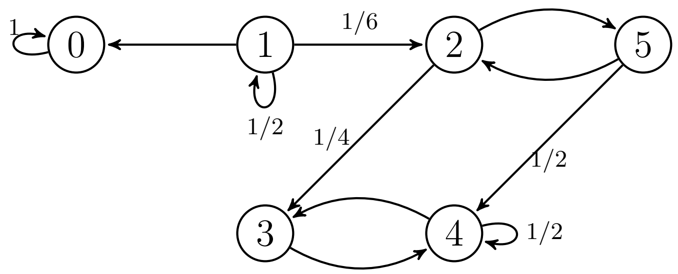
per indipendenza e allora

$$\mathrm{Cov}(U, V) = 1 - \frac{1}{2} \cdot \frac{3}{2} = 1 - \frac{3}{4} = \frac{1}{4}$$

quindi U e V sono positivamente correlate.

Problema 3

Si consideri una catena di Markov $(X_n)_n$ con probabilità di transizione rappresentate in figura



e tale che X_0 sia (a priori) una variabile discreta sull'insieme degli stati avente densità

$$P(X_0 = k) = c(k+1), \quad \text{per } k \in \{0,1,2,3,4,5\},$$

dove $c > 0$ è una opportuna costante.

1. Calcolare c, scrivere la matrice di transizione Q (completare con le probabilità mancanti), classificare gli stati (transitori/ricorrenti) e determinare tutte le distribuzioni invarianti della catena.
2. Avendo osservato che $X_2 = 0$, determinare la densità a posteriori di X_0.
3. Calcolare il limite $\lim_{n\to\infty} P(X_0 = 0|X_n = 0)$ (spiegare anche perché esiste).

Una soluzione:

1. Per calcolare c basta sommare i valori $\sum_{k=0}^{5}(k+1) = 1+2+3+4+5+6 = 21$ e porre $c = 1/21$. Scriviamo la matrice di transizione ordinando gli stati nell'ordine naturale $E = \{0,1,2,3,4,5\}$. Troviamo

$$Q = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1/3 & 1/2 & 1/6 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/4 & 0 & 3/4 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1/2 & 1/2 & 0 \\ 0 & 0 & 1/2 & 0 & 1/2 & 0 \end{pmatrix}$$

Gli stati 1, 2 e 5 sono transitori (infatti si raggiunge 4 da questi ma da 4 possiamo raggiungere solo 3 o 4 stesso. Gli altri stati sono ricorrenti. Le classi chiuse irriducibili sono $C_1 = \{0\}$ e $C_2 = \{3,4\}$ (entrambe regolari per il criterio). Per la classe C_1 la distribuzione invariante è banale $\mu_{C_1} = (1,0,0,0,0,0)$, mentre per calcolare μ_{C_2} impostiamo il bilancio di flusso in 3

$$\mu_3 = \frac{1}{2}\mu_4$$

e troviamo quindi $\mu_3 = 1/3$, $\mu_4 = 2/3$ ossia

$$\mu_{C_2} = (0,0,0,\frac{1}{3},\frac{2}{3},0),$$

Tutte le distribuzioni invarianti sono quindi della forma

$$\mu(\alpha) = (1-\alpha)\mu_{C_1} + \alpha\mu_{C_2} = (1-\alpha,0,0,\alpha/3,2\alpha/3,0)$$

al variare di $\alpha \in [0,1]$.

2.Per determinare la densità a posteriori di X_0, basta usare la formula di Bayes:

$$P(X_0 = k|X_2 = 0) \propto P(X_0 = k)P(X_2 = 0|X_0 = k) \propto (k+1)P(X_2 = 0|X_0 = k).$$

Per calcolare la verosimiglianza $P(X_2 = 0|X_0 = k)$, osserviamo che se $k \notin \{0,1\}$, allora questa è nulla (non ci sono cammini che da uno stato $k \geq 2$ portano in 0). Pertanto avremo solo $P(X_2 = 0|X_0 = 0) = 1$, e $P(X_2 = 0|X_0 = 1) = 1/3 \cdot 1 + 1/2 \cdot 1/3 = 1/2$ e la densità a posteriori μ_2 è del tipo

$$\mu_2 = c(1 \cdot 1, 2 \cdot 1/2, 0, 0, 0, 0) = c(1, 1, 0, 0, 0, 0) = (1/2, 1/2, 0, 0, 0, 0)$$

avendo imposto che la somma delle probabilità sia 1.

3. Usiamo nuovamente Bayes

$$P(X_0 = 0|X_n = 0) = P(X_0 = 0)P(X_n = 0|X_0 = 0)/P(X_n = 0).$$

I limiti per $n \to \infty$ di $P(X_n = 0|X_0 = 0)$ e di $P(X_n = 0)$ esistono perché ogni classe chiusa irriducibile è regolare. Infatti il secondo si ottiene per disintegrazione

$$P(X_n = 0) = \sum_{k=0}^{5} P(X_n = 0|X_0 = k)P(X_0 = k).$$

Come osservato nel punto precedente, abbiamo che $P(X_n = 0|X_0 = k) = 0$ se $k \geq 2$ (e quindi anche nel limite pen $n \to \infty$). Inoltre è banalmente vero che $P(X_n = 0|X_0 = 0) = 1$ e quindi anche nel limite per $n \to \infty$. Resta quindi da calcolare il limite $\lim_{n\to\infty} P(X_n = 0|X_0 = 1) = Q^{\infty}_{10}$ (usando la notazione delle lezioni). Impostiamo quindi un'equazione dall'identità $Q^{\infty} = QQ^{\infty}$:

$$Q^{\infty}_{10} = Q_{10}Q^{\infty}_{00} + Q_{11}Q^{\infty}_{10} + Q_{12}Q^{\infty}_{20}$$

ma $Q^{\infty}_{00} = 1$, mentre $Q^{\infty}_{20} = 0$, quindi

$$Q^{\infty}_{10} = \frac{1}{3} + \frac{1}{2}Q^{\infty}_{10}$$

da cui $Q^{\infty}_{10} = 2/3$. Ne segue che

$$\lim_{n\to\infty} P(X_n = 0) = \frac{1}{21} \cdot 1 + \frac{2}{21} \cdot \frac{2}{3} = \frac{7}{3} \cdot \frac{1}{21}$$

e quindi

$$\lim_{n\to\infty} P(X_0 = 0|X_n = 0) = \frac{\frac{1}{21} \cdot 1}{\frac{7}{3} \cdot \frac{1}{21}} = \frac{3}{7}.$$