CdL in Ing. Robotica e dell'Automazione, Prob. e Proc. Stoc. (455AA)
A.A. 2023/24 - Prova in itinere 2024-01-10

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

Problema 1

Un'azienda vuole acquistare un algoritmo per il riconoscimento automatico della presenza di incendi tramite immagini. Nel mercato sono a disposizione proposte simili, che però dichiarano capacità di riconoscimento leggermente diverse. Per esempio, la casa di software *IAlpha* dichiara che in un esperimento in cui sono state presentate 100 immagini di principio di incendio, il proprio algoritmo ne ha rilevato la presenza in 99 casi. La casa di software *IBeta* dichiara invece che in un esperimento simile, ma su 1000 immagini, il proprio algoritmo ne ha rilevato la presenza in 990 casi.

Poniamo $\alpha \in (0, 1)$ la probabilità che l'algoritmo di *IAlpha* rilevi correttamente la presenza di fuoco in un'immagine di un principio di incendio, e similmente $\beta \in (0, 1)$ per l'algoritmo di *IBeta*. Supponiamo che i due algoritmi lavorino indipendentemente e che immagini diverse siano classificate le une indipendentemente dalle altre.

1. Fornire una stima di massima verosimiglianza per α.
2. Si vuole stimare α usando un approccio bayesiano e si suppone la densità a priori per α uniforme (continua) su $(0, 1)$. Calcolare esplicitamente la densità a posteriori.
3. Ripetere i due punti precedenti per β (ma indicare la densità a posteriori a meno di una costante moltiplicativa). A parità di costo, quale tra le due proposte suggerireste all'azienda di scegliere?

Una soluzione:

1. Scriviamo la verosimiglianza (basta considerare soltanto l'informazione relativa ad α per l'ipotesi di indipendenza). Sapendo α, si tratta di 100 esperimenti indipendenti ripetuti, ciascuno cor probabilità α di successo. Quindi la verosimiglianza è data dalla densità binomiale

$$L(\alpha; 99 \text{ su } 100) = P(99 \text{ successi su } 100|\alpha) = \binom{100}{99}\alpha^{99}(1-\alpha). \tag{1}$$

Passando al logaritmo e derivando si trova

$$\frac{\partial}{\partial\alpha}\log L(\alpha) = \frac{99}{\alpha} - \frac{1}{1-\alpha}. \tag{2}$$

ponendo la derivata nulla si trova con facili calcoli che $\alpha_{MLE} = 99\%$.

2. Stavolta supponiamo che $\alpha \in (0, 1)$ sia una variabile aleatoria con distribuzione uniforme a priori (indichiamola con X). Usando la formula di Bayes si trova la densità a posteriori

$$p(X = \alpha|99 \text{ su } 100) \propto p(X = \alpha)L(\alpha; 90 \text{ su } 100) = c\alpha^{99}(1-\alpha), \tag{3}$$

dove la costante c è data da

$$c = \left(\int_0^1 \alpha^{99}(1-\alpha)d\alpha\right)^{-1}. \tag{4}$$

Calcoliamo facilmente

$$\int_0^1 \alpha^{99}d\alpha = \frac{1}{100}, \quad \int_0^1 \alpha^{100}d\alpha = \frac{1}{101}, \tag{5}$$

da cui

$$c = 100 \cdot 101. \tag{6}$$

3. Se ripetiamo la stessa cosa per β troviamo $\beta_{MLE} = 990/1000 = 99\%$ e

$$p(\beta|990 \text{ su } 1000) = c\beta^{990}(1-\beta)^{10}, \tag{7}$$

per una costante c opportuna (non serve calcolarla esplicitamente). Vediamo che se ci limitiamo alla stima di massisma verosimiglianza i due algoritmi sembrerebbero equivalenti, tuttavia dovrebbe essere intuitivo che *IBeta* fornisce una indicazione più precisa – e quindi l'azienda dovrebbe optare per questo. Per giustificarlo, basterebbe disegnare un grafico delle due densità a posteriori, osservando che la seconda è molto più concentrata intorno al valore 99% rispetto alla prima. Per vederlo analiticamente, possiamo ad esempio calcolare l'approssimazione di Laplace intorno a 99% in entrambi i casi. Troviamo infatti che la derivata seconda del logaritmo densità di α è

$$\frac{\partial^2}{\partial \alpha^2} \log L(\alpha) = -\frac{99}{\alpha^2} - \frac{1}{(1-\alpha)^2}, \tag{8}$$

che valutata in $\alpha = 99\%$ darebbe il valore $-100^2(1+1/99)$. Possiamo quindi approssimare α con una Gaussiana di media 99% e varianza $\approx 1/100$ (quindi una deviazione standard di circa 1/10) Similmente la derivata seconda della densità di β in 99% darebbe il valore $-1000^2 * (1+1/990)$ quindi approssimiamo β con una Gaussiana con la stessa media ma varianza $\approx 1/1000$, quindi una deviazione standard di circa 1/32, quindi più piccola.

Problema 2

Dato $m \in \mathbb{R}$, sia X una variabile aleatoria gaussiana reale di parametri $\mathcal{N}(m,1)$, avente quindi funzione caratteristica $\mathbb{E}\left[e^{itX}\right] = e^{itm - t^2/2}$.

1. Calcolare, in funzione di m, $\mathbb{E}[\cos(X)]$ e $\mathbb{E}[\operatorname{sen}(X)]$. *(Sugg: usare la funzione caratteristica)*
2. Calcolare, in funzione di m, $\operatorname{Var}(\cos(X))$ e $\operatorname{Var}(\operatorname{sen}(X))$. *(Sugg: usare un'identità per* $\cos(2x)$*)*
3. Determinare per quali valori di m le variabili $\cos(X)$, $\operatorname{sen}(X)$ sono positivamente correlate.

Una soluzione:

La funzione caratteristica indicata è un suggerimento, perché troviamo

$$\mathbb{E}\left[e^{itX}\right] = \mathbb{E}[\cos(tX)] + i\mathbb{E}[\operatorname{sen}(tX)] = e^{itm-t^2/2} = e^{-t^2/2}\cos(tm) + ie^{-t^2/2}\operatorname{sen}(tm) \tag{9}$$

e quindi eguagliando rispettivamente le parte reali e le parti immaginarie scopriamo che

$$\mathbb{E}[\cos(tX)] = e^{-t^2/2}\cos(tm) \quad \mathbb{E}[\operatorname{sen}(tX)] = e^{-t^2/2}\operatorname{sen}(tm). \tag{10}$$

Usiamo allora queste identità per rispondere alle tre domande.

1. Basta porre $t=1$ nelle identità e troviamo che

$$\mathbb{E}[\cos(X)] = e^{-1/2}\cos(m) \quad \mathbb{E}[\operatorname{sen}(X)] = e^{-1/2}\operatorname{sen}(m). \tag{11}$$

2. Per calcolare la varianza, calcoliamo prima il momento secondo. Poiché $\cos^2(X)+\text{sen}^2(X) = 1$, basterà calcolare il momento secondo di $\cos(X)$. D'altra parte si ha l'identità trigonometrica $\cos^2(x) = \frac{1+\cos(2x)}{2}$ e quindi

$$\mathbb{E}\left[\cos^2(X)\right] = \mathbb{E}\left[\frac{1+\cos(2X)}{2}\right] = \frac{1+e^{-2}\cos(2m)}{2} \tag{12}$$

avendo usato l'identità preliminare con $t = 2$. Troviamo poi

$$\mathbb{E}\left[\text{sen}^2(X)\right] = 1 - \mathbb{E}\left[\cos^2(X)\right] = \frac{1-e^{-2}\cos(2m)}{2}, \tag{13}$$

e quindi

$$\text{Var}\left(\cos(X)\right) = \frac{1+e^{-2}\cos(2m)}{2} - e^{-1}\cos^2(m), \quad \text{Var}\left(\text{sen}(X)\right) = \frac{1-e^{-2}\cos(2m)}{2} - e^{-1}\,\text{sen}^2(m). \tag{14}$$

3. Ricordiamo anche l'identità $\text{sen}(2x) = 2\,\text{sen}(x)\cos(x)$, per cui troviamo

$$\mathbb{E}\left[\text{sen}(X)\cos(X)\right] = \frac{1}{2}\mathbb{E}\left[\text{sen}(2x)\right] = \frac{e^{-2}}{2}\,\text{sen}(2m) \tag{15}$$

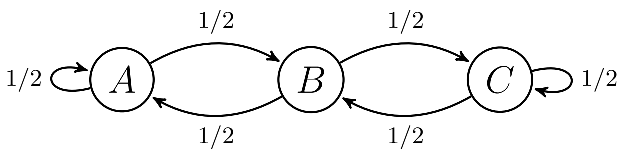
Usando le formule dei valori attesi calcolate al punto 1 abbiamo quindi

$$\text{Cov}\left(\text{sen}(X), \cos(X)\right) = \frac{e^{-2}}{2}\,\text{sen}(2m) - e^{-1}\,\text{sen}(m)\cos(m) = \frac{1}{2}e^{-1}\,\text{sen}(2m)(e^{-1}-1). \tag{16}$$

Poiché l'ultimo fattore è negativo, le variabili sono positivamente correlate solo quando $\text{sen}(2m) < 0$, ossia $m \in ((k+1/2)\pi, (k+1)\pi)$ per qualche $k \in \mathbb{Z}$.

Problema 3

Anna, Bruna e Carla hanno un nuovo videogioco, in cui però si può giocare soltanto una alla volta. Stabiliscono una regola per i turni di gioco secondo la catena di Markov rappresentata in figura, dove gli stati indicano le iniziali della giocatrice cui spetta il turno, e la giocatrice per il primo turno è scelta uniformemente tra le tre.



1. Bruna ritiene di essere sfavorita perché dopo ogni partita deve per forza passare il turno di gioco ad Anna o a Carla (mentre loro no): ha ragione? (*calcolare le distribuzioni invarianti*)
2. Calcolare la probabilità che nei primi 10 turni Bruna non abbia mai giocato.
3. Sapendo che nei primi 10 turni Bruna non ha mai giocato, calcolare la probabilità che giochi all'undicesimo turno.

Una soluzione:

1. Poiché la catena è irriducibile, quindi ha una sola distribuzione invariante. Per calcolarla basta imporre il bilancio di flusso. In A, troviamo l'equazione $\mu_A 2 = \mu_B/2$, quindi $\mu_A = \mu_B$. Similmente, troviamo dal bilancio di flusso in C che $\mu_C/2 = \mu_B/2$ e quindi anche $\mu_B = \mu_C$. Pertanto $\mu_A = \mu_B = \mu_C$ e l'unica distribuzione invariante è quella uniforme sui tre stati. Questa è anche la distribuzione iniziale, quindi la catena è stazionaria. Questo significa che ad ogni istante la probabilità che si trovi in A è $1/3$, lo stesso per B e lo stesso per C. Quindi da questo punto di vista Bruna ha la stessa probabilità di giocare delle altre due, ossia non è sfavorita.

2. Dobbiamo calcolare i cammini che visitano 10 stati ma che non passano mai per B. Vista la struttura della catena, essi sono solo due $A \to A \to A \to \ldots \to A$, con peso 2^{-9} (ci sono 9 transizioni) e $B \to B \to \ldots \to B$, con probabilità 2^{-9}. La probabilità iniziale è uniforme sui tre stati, quindi otteniamo

$$P(\text{Bruna non gioca nei primi 10 turni}) = \frac{1}{3}2^{-9} + \frac{1}{3}2^{-9}. \tag{17}$$

3. Sappiamo per certo che al decimo turno Beatrice non ha giocato, quindi con probabilità 1 la catena sarà in A o in C (notiamo che non è necessario sapere cosa è accaduto prima, per la proprietà di Markov). D'altra parte non è neppure necessario sapere la probabilità con cui la catena si trova in A e quella con cui si trova in C, perché da qualsiasi stato dei due parta, con probabilità $1/2$ la catena al tempo successivo si troverà in B. Pertanto la probabilità richiesta è $1/2$.