Informazione e stima – 21/06/2022 – Compito A

- Ogni esercizio viene valutato da 0 a 6 punti
- Verranno valutate solo le parti scritte in penna – non usare correttori
- Non riportare solo il risultato, ma cerca di argomentare sinteticamente la risposta.
- Riportare il proprio nome, cognome, codice persona, e tipo di compito svolto su ogni foglio consegnato
- Esercizio da escludere dal punteggio finale:

1. Sfruttando l'approssimazione Gaussiana della Binomiale data dal teorema di De Moivre-Laplace, dare una stima accurata del coefficiente binomiale $\binom{300}{140}$.

2. Si consideri un processo di Poisson con intensità λ, e si costruisca un nuovo processo considerando solo gli arrivi dispari del processo originale.
   (a) Il nuovo processo è di Poisson? Giustificare la risposta.
   (b) Determinare le leggi di T_1 e T_2, primo e secondo tempo di interarrivo nel processo nuovo.

3. Si vuole stimare il bilanciamento ignoto $\Pr(\text{Testa}) = \theta$ di una moneta osservando il numero di lanci necessari per ottenere una testa. L'unica informazione a priori disponibile è $f_\Theta(\theta) = 2\theta$ per $\theta \in [0, 1]$. Ipotizzando di osservare la prima testa al lancio X, trovare lo stimatore MAP di Θ.

4. Si considerino punti di coordinate (X, Y) distribuiti uniformemente in un disco di raggio 3 e centro nell'origine di un sistema Cartesiano. Sia $B = \lfloor X \rfloor$ l'intero più grande minore o uguale a X.

   Partendo da un generatore di campioni indipendenti U_i distribuiti come $U \sim \mathcal{U}[0, 1]$, descrivere un algoritmo (pseudocodice) per campionare dalla legge di B.

   Mediamente ogni quanti tentativi si genera un campione valido di B?

5. Ad un seggio elettorale gli uomini e le donne arrivano secondo due processi di Poisson indipendenti con intensità $\lambda_1 = 1$ e $\lambda_2 = 2$ al minuto, rispettivamente. Sapendo che ci sono 10 persone in coda, qual è la probabilità che la coda sia aperta e chiusa da due donne?

6. Gli n studenti del Politecnico rispondono ad un sondaggio con una domanda che ha d opzioni (esclusive) di risposta. Si vogliono memorizzare in un file tutte le risposte al quesito. Nel peggiore dei casi, quale sarà la lunghezza media minima del file (in bit)?
   Si può assumere che gli studenti rispondano in maniera indipendente.

Soluzioni

Problema 1

Si consideri $X \sim \text{Bin}(300, 1/2)$, con $\mathsf{E}[X] = 300 \cdot \frac{1}{2} = 150$ e $\mathsf{Var}[X] = 300 \cdot \frac{1}{4} = 75$. Si può notare che

$$\Pr(X = 140) = \binom{300}{140} \frac{1}{2^{300}}, \tag{1}$$

e che

$$\begin{aligned}
\Pr(X = 140) &= \Pr(139.5 \le X \le 140.5) &(2)\\
&= \Pr\left(\frac{139.5 - 150}{\sqrt{75}} \le \frac{X - 150}{\sqrt{75}} \le \frac{140.5 - 150}{\sqrt{75}}\right) &(3)\\
&\overset{(CLT)}{\approx} \Pr(-1.2124 \le Z \le -1.0969) &(4)\\
&= \Pr(1.0969 \le Z \le 1.2124) &(5)\\
&= \Phi(1.2124) - \Phi(1.0969) &(6)\\
&\approx 2.36 \cdot 10^{-2} &(7)
\end{aligned}$$

dove (5) deriva dalla simmetria della Gaussiana standard Z. In (4) possiamo applicare il CLT perché la Binomiale X è somma di 300 v.a. iid. Usando (1) e (7), l'approssimazione del coefficiente binomiale risulta:

$$\binom{300}{140} \approx 2^{300} \cdot 2.36 \cdot 10^{-2}. \tag{8}$$

Problema 2

(a) Il nuovo processo non è di Poisson, perché lo splitting degli arrivi del processo di Poisson originario non è fatto secondo prove indipendenti.

(b) Il tempo T_1 è quello che intercorre dal tempo 0 al primo arrivo, che corrisponde al primo arrivo nel processo di Poisson originale. Dunque $T_1 \sim \text{Exp}(\lambda)$.

Il tempo T_2 è quello che intercorre tra il primo e il terzo arrivo del processo originale di Poisson. Dunque $T_2 = X_2 + X_3$, con $X_2 \sim X_3 \sim \text{Exp}(\lambda)$. Da ciò segue che $T_2 \sim \text{Erlang-2}(\lambda)$.

Problema 3

Dai dati del problema si sa che $\{X|\Theta = \theta\}$ è una v.a. $\text{Geom}(\theta)$, ovvero $p_{X|\Theta}(x|\theta) = (1-\theta)^{x-1}\theta$, e $f_\Theta(\theta) = 2\theta$ per $\theta \in [0, 1]$. Lo stimatore MAP si trova applicando la definizione:

$$\begin{aligned}
\widehat{\Theta}_{\text{MAP}}(X) &= \arg\max_{\theta\in[0,1]} f_{\Theta|X}(\theta|X)\\
&= \arg\max_{\theta\in[0,1]} p_{X|\Theta}(X|\theta) f_\Theta(\theta)\\
&= \arg\max_{\theta\in[0,1]} (1-\theta)^{X-1} 2\theta^2.
\end{aligned}$$

Per $X \ge 1$ la funzione $\theta \mapsto (1-\theta)^{X-1}2\theta^2$ è differenziabile, e possiamo trovare i punti stazionari impostando:

$$\frac{d}{d\theta}(1-\theta)^{X-1}2\theta^2 \overset{!}{=} 0 \tag{9}$$

che dà come unico risultato $\theta = \frac{2}{X+1}$. Siccome la derivata seconda è sempre negativa, il punto stazionario trovato è un massimo. Dunque, $\widehat{\Theta}_{\text{MAP}}(X) = \frac{2}{X+1}$ per $X \ge 1$.

Problema 4

Per generare i punti uniformemente distribuiti dentro il disco di raggio 3 si potrebbero generare punti dentro un quadrato di lato 6 che circoscrive il disco e poi rigettare i punti che escono dal disco. Un possibile algoritmo per campionare dalla legge di B è il seguente:

1. Genero due campioni $U_i \sim \mathcal{U}[0, 1]$ in maniera indipendente e pongo $X = 6U_1 - 3$ e $Y = 6U_2 - 3$.
2. Se $X^2 + Y^2 \le 3^2$ allora pongo $B = \lfloor X \rfloor$, altrimenti torno al punto 1.

Un campione valido di B viene generato quando il punto (X,Y) cade dentro il disco, e ciò accade con probabilità:

$$\Pr(B \text{ valido}) = \frac{\text{area(disco)}}{\text{area(quadrato)}} = \frac{\pi 3^2}{6^2} = \frac{\pi}{4}. \tag{10}$$

Il numero medio di tentativi per avere un valore valido di B è pari alla media di una v.a. Geometrica di parametro $\frac{\pi}{4}$, cioè $4/\pi \approx 1.27$.

Problema 5

Si può immaginare un processo unione dove gli arrivi sono indipendenti e di tipo "uomo" con probabilità $\frac{\lambda_1}{\lambda_1+\lambda_2} = \frac{1}{3}$ e di tipo "donna" con probabilità $\frac{\lambda_2}{\lambda_1+\lambda_2} = \frac{2}{3}$.

La probabilità di avere due donne in posizione 1 e 10 nella coda è

$$\Pr(D_1 \cap D_{10}) = \Pr(D_1)\Pr(D_{10}) = \frac{4}{9} = 0.444. \tag{11}$$

Problema 6

Sia $X_i \in \{1, 2, \ldots, d\}$ la v.a. che registra la risposta dello studente i-esimo alla domanda. Per il teorema della codifica di sorgente, sappiamo che la mediamente la lunghezza minima del file sarà

$$H(X_1, X_2, \ldots, X_n) = \sum_{i=1}^{n} H(X_i). \tag{12}$$

Si noti che non conosciamo la legge di probabilità delle X_i, ma solo che possono assumere d valori diversi. Nel peggiore dei casi, le entropie $H(X_i)$ sono massimizzate quando la legge di X_i è uniforme. Pertanto, nel peggiore dei casi, la lunghezza media minima del file sarà di

$$\mathsf{E}[L] \geq nH(\mathcal{U}\{1, 2, \ldots, d\}) = n \log_2(d) \text{ bits}. \tag{13}$$