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

- 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. In 10 lanci indipendenti consecutivi di moneta si ottengono 5 teste e 5 croci. Qual è la probabilità di avere una sequenza alternata di teste e croci?

2. 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. Calcolare $\Pr(B = 2)$.

3. Sia $U \sim \mathcal{U}[0, 3]$. Determinare la legge di probabilità di $X = g(U)$, dove

$$g(u) = \begin{cases} 2u & 0 \leq u \leq 1 \\ 2 & 1 \leq u \leq 2 \\ 6 - 2u & 2 \leq u \leq 3 \\ 0 & \text{altrimenti,} \end{cases}$$

e rappresentare graficamente la legge cumulata F_X.

4. 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.

5. 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.

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

Siccome tutti i lanci sono indipendenti e identicamente distribuiti, tutte le sequenze di 5 teste e 5 croci hanno la stessa probabilità di accadere. Lo spazio di probabilità è uniforme discreto. La probabilità cercata è:

$$\Pr(TCTC... \cup CTCTC...) = \frac{10 \cdot 5 \cdot 4 \cdot 4 \cdot 3 \cdot 3 \cdot 2 \cdot 2 \cdot 1 \cdot 1}{10 \cdot 9 \cdot 8 \cdot 7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1} = \frac{10 \cdot 5! \cdot 4!}{10!} = \frac{5!4!}{9!} = \frac{1}{\binom{9}{5}} \approx 0.0178. \tag{1}$$

Problema 2

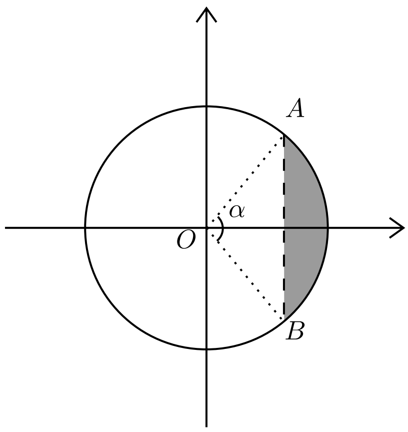
Lo spazio di probabilità è continuo e uniforme dentro il disco di raggio 3. L'evento di interesse può essere reinterpretato come

$$\{B = 2\} = \{2 < X \leq 3\} \tag{2}$$

che equivale all'area evidenziata in figura. L'area può essere calcolata come differenza tra l'area del settore circolare con apertura

$$\alpha = 2 \arccos \frac{2}{3} \tag{3}$$

e l'area del triangolo AOB.



L'area del settore circolare è $\text{area}(\widehat{AB}) = 3^2 \frac{\alpha}{2} = 9 \arccos \frac{2}{3}$, mentre l'area del triangolo è

$$\begin{aligned}
\text{area}(AOB) &= 3 \sin(\frac{\alpha}{2}) \cdot 3 \cos(\frac{\alpha}{2}) && (4)\\
&= \frac{3^2 \cdot 2}{3} \sin\left(\arccos \frac{2}{3}\right) && (5)\\
&= 6\sqrt{1 - \frac{4}{9}} && (6)\\
&= 2\sqrt{5}. && (7)
\end{aligned}$$

La probabilità cercata è dunque

$$\Pr(B = 2) = \frac{\text{area}(\widehat{AB}) - \text{area}(AOB)}{\text{area(disco)}} = \frac{9 \arccos \frac{2}{3} - 2\sqrt{5}}{9\pi} \approx 0.1095 \tag{8}$$

Problema 3

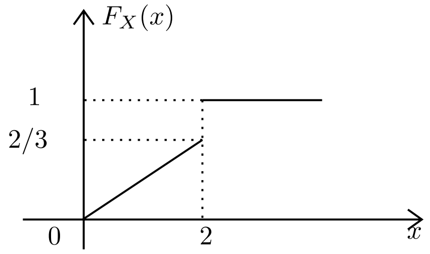
Innanzitutto si noti che

$$F_U(u) = \begin{cases} 0 & u \leq 0 \\ \frac{u}{3} & 0 \leq u \leq 3 \\ 1 & u \geq 3. \end{cases} \tag{9}$$

Usando il metodo della cumulata, si ottiene

$$\begin{aligned}
F_X(x) &= \Pr(g(U) \le x) && (10)\\
&= \Pr(\{2U \le x\} \cup \{6 - 2U \le x\}) && (11)\\
&= \Pr(U \le \frac{x}{2}) + \Pr(U \ge \frac{6-x}{2}) && (12)\\
&= F_U(\frac{x}{2}) + 1 - F_U(\frac{6-x}{2}) && (13)\\
&= \frac{x}{6} + 1 - \frac{6-x}{6} && (14)\\
&= \frac{x}{3}, \qquad 0 \le x < 2, && (15)
\end{aligned}$$

mentre si ha $F_X(x) = 1$ per $x \ge 2$, e $F_X(x) = 0$ per $x \le 0$.



Problema 4

(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 5

Dai dati del problema si sa che $\{X|\Theta = \theta\}$ è una v.a. Geom(θ), 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:

$$\frac{d}{d\theta}(1-\theta)^{X-1}2\theta^2 \overset{!}{=} 0 \tag{16}$$

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 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

$$H(X_1, X_2, \ldots, X_n) = \sum_{i=1}^{n} H(X_i). \tag{17}$$

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

$$\mathsf{E}[L] \geq nH(\mathcal{U}\{1, 2, \ldots, d\}) = n \log_2(d) \text{ bits.} \tag{18}$$