Informazione e stima – 28/06/2021 – 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. In un torneo si sa che ogni ogni squadra di calcio segna in una partita con probabilità 0.6, indipendentemente da tutto. Qual è la legge di probabilità del numero di partite giocate per osservare la quarta partita con risultato 0-0?

2. Alla cassa di un supermercato arrivano clienti paganti in contanti secondo un $PP(\lambda = 10$ clienti/ora$)$ e clienti con pagamenti elettronici secondo un $PP(\lambda = 5$ clienti/ora$)$. Qual è la probabilità che tra i primi 10 clienti ci siano stati più pagamenti elettronici che in contanti?

3. Sia $\{X|M = \mu\} \sim \mathcal{N}(\mu, 1)$, dove μ è un parametro ignoto di cui si conosce solo la distribuzione a priori $M \sim \mathcal{U}[1, 2]$. Determinare lo stimatore MAP $\widehat{M}_{\text{MAP}}$ basato su una osservazione X.

4. Partendo da un generatore di campioni indipendenti U_i distribuiti come $U \sim \mathcal{U}[0, 1]$, descrivere un algoritmo per campionare da una distribuzione $X \sim \text{Geom}(p)$. Mediamente quanti campioni uniformi bisogna generare per avere un campione geometrico?

5. Il numero di Eulero e può essere calcolato come $e = (1 - \int_0^1 e^{-x} dx)^{-1}$. Partendo da un generatore di campioni indipendenti U_i distribuiti come $U \sim \mathcal{U}[0, 1]$, descrivere un algoritmo per stimare il valore di e sempre più accuratamente all'aumentare del numero di campioni U_i generati.

6. Si lancia 20 volte una moneta con $P(T) = 0.2$ e 80 volte una moneta con $P(T) = 0.8$. Tutti i lanci sono indipendenti. Quanti bit di entropia vengono generati?

Soluzioni

Problema 1

La probabilità che una partita termini 0-0 è $p = (1 - 0.6)^2 = 0.16$. Questa può essere interpretata come la probabilità di successo in un processo di Bernoulli. La legge richiesta è quella del tempo al quarto successo. Sia X il numero di partite da giocare fino ad osservare il quarto 0-0, allora:

$$P(X = x) = \binom{x-1}{3} p^4 (1-p)^{x-4}, \qquad x = 4, 5, \cdots. \tag{1}$$

Problema 2

Si consideri il processo unione dei clienti. I tipi di cliente dei primi 10 arrivi sono indipendenti tra loro. La prob di avere un cliente pagante in contanti è $\frac{10}{10+5} = \frac{2}{3}$. Sia C il numero di clienti paganti con contante tra i primi 10 clienti, allora la probabilità cercata è:

$$\Pr(C < 5) = \sum_{i=0}^{4} \binom{10}{i} \left(\frac{2}{3}\right)^i \left(\frac{1}{3}\right)^{10-i}. \tag{2}$$

Problema 3

Dai dati del problema si sa che $f_{X|M}(x|\mu) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2}}$, e $f_M(\mu) = 1$ per $\mu \in [1, 2]$. Lo stimatore MAP si trova applicando la definizione:

$$\begin{aligned}
\widehat{M}_{\mathrm{MAP}}(x) &= \arg \max_{\mu \in [1,2]} f_{M|X}(\mu|x) \\
&= \arg \max_{\mu \in [1,2]} f_{X|M}(x|\mu) f_M(\mu) \\
&= \arg \max_{\mu \in [1,2]} e^{-\frac{(x-\mu)^2}{2}} \\
&= \arg \max_{\mu \in [1,2]} -(x-\mu)^2 \\
&= \arg \min_{\mu \in [1,2]} (x-\mu)^2.
\end{aligned}$$

Si vuole trovare il punto di minimo di una parabola rivolta verso l'alto, nell'intervallo $\mu \in [1, 2]$. Se non ci fossero restrizioni su μ, allora la soluzione sarebbe il punto di minimo della parabola $\mu = x$. Se $x < 1$, allora la parabola tocca il suo punto di minimo in $\mu = 1$, mentre se $x > 2$ la parabola tocca il suo punto di minimo in $\mu = 2$. Quindi lo stimatore MAP sarà dato da

$$\widehat{M}_{\mathrm{MAP}}(x) = \begin{cases} 1 & x \leq 1 \\ x & 1 \leq x \leq 2 \\ 2 & x \geq 2 \end{cases}$$

Problema 4

I campioni geometrici possono essere interpretati come tempi di interarrivo di un processo di Bernoulli BP(p). I campioni uniformi possono essere usati per generare successi e insuccessi del processo di Bernoulli. Un possibile algoritmo per generare un campione X Geometrico è il seguente:

1. Inizializzo $i = 1$
2. Genero un campione $U_i \sim \mathcal{U}[0, 1]$
3. Se $U_i < p$ allora $X = i$, altrimenti $i \leftarrow i + 1$ e torno al punto 1.

Il numero medio di campioni U_i generati per avere un valore valido di X è proprio pari alla media di una v.a. Geometrica di parametro p, cioè $1/p$.

Problema 5

Un possibile approccio è quello di valutare l'integrale con un metodo MonteCarlo. L'integrale si può riscrivere come

$$\int_0^1 e^{-x} dx = \Pr(X < 1) = \mathsf{E}[\mathbb{1}(X < 1)] \tag{3}$$

dove $X \sim \text{Exp}(1)$. Se si avessero a disposizione n v.a. $X_i \sim \text{Exp}(1)$ indipendenti, allora

$$\Pr(X < 1) \approx \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}(X_i < 1). \tag{4}$$

Le variabili esponenziali X_i possono essere generate come $X_i = -\ln(U_i)$. Dunque:

$$e \approx \left(1 - \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}(-\ln(U_i) < 1)\right)^{-1}. \tag{5}$$

Problema 6

Siano X_i i risultati dei lanci della prima moneta, e Y_i i risultati dei lanci della seconda moneta. Allora si ha che

$$\begin{aligned}
H(X_1, X_2, \ldots, X_{20}, Y_1, Y_2, \ldots, Y_{80}) &= \sum_{i=1}^{20} H(X_i) + \sum_{i=1}^{80} H(Y_i) && (6)\\
\text{(identic. distr.)} &= 20H(X_1) + 80H(Y_1) && (7)\\
\text{(Siccome } X_1, Y_1 \text{ sono v.a. binarie)} &= 20H_2(0.2) + 80H_2(0.8) && (8)\\
\text{(Siccome } H_2(x) = H_2(1-x)\text{)} &= 100H_2(0.2) && (9)\\
&= 100\left(0.2\log_2(\frac{1}{0.2}) + 0.8\log_2(\frac{1}{0.8})\right) && (10)\\
&\approx 72.19 \text{ bit} && (11)
\end{aligned}$$

dove il primo passaggio è dovuto all'indipendenza di tutte le v.a.