Informazione e stima – 01/02/2024

- 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, e codice persona su ogni foglio consegnato
- Esercizio da escludere dal punteggio finale:

1. Le lettere che compongono la parola STIMA sono scritte su 5 cartoncini. Mischiando e pescando i 5 cartoncini a caso si forma una parola. Si calcoli la probabilità che
   (a) le lettere M e A siano affiancate (MA e AM);
   (b) S sia la prima lettera e M e A siano affiancate.

2. Siano $X \sim \mathcal{N}(1, 1)$ e $Y \sim \mathcal{N}(2, 3)$ due variabili aleatorie tali che $\rho[X, Y] = 0.7$, e sia $W = 2X - Y$. Calcolare $\Pr(0 < W < 2)$.

3. Si supponga che il raggio R di una sfera abbia densità di probabilità $f_R(r) = 12r^2(1 - r)$ per $0 \le r \le 1$ e $f_R(r) = 0$ altrimenti. Determinare la densità di probabilità del volume della sfera.
   Si ricorda che il volume di una sfera di raggio R è $V = \frac{4}{3}\pi R^3$

4. I pacchetti di una certa marca di patatine contengono personaggi giocattolo. Ci sono 5 personaggi diversi da collezionare. Ogni pacchetto è vuoto con probabilità 0.7 oppure contiene uno di questi personaggi, in maniera equiprobabile e indipendente dagli altri pacchetti. Se si compra un pacchetto al giorno, calcolare
   (a) il numero medio di giorni per completare la collezione
   (b) il numero minimo di giorni per ottenere il primo personaggio con una probabilità di almeno il 95%.

5. Si vuole stimare il valore dell'integrale $I = \int_0^1 e^{\tan(x)} dx$ tramite una simulazione Monte Carlo. La stima fatta su n campioni, $\widehat{I}_n$, deve essere tale che $\mathsf{E}[\widehat{I}_n] = I$.
   (a) Scrivere un algoritmo (tramite pseudocodice) che permette di calcolare la stima Monte Carlo di I.
   (b) Qual è la relazione che lega il numero di campioni utilizzati n e la varianza dell'errore di stima?

6. Sia $Y_n = a + X_n$ con $\mathsf{E}[|X_n|] = \frac{1}{n}$ per ogni $n \in \mathbb{N}$. Determinare se la successione di variabili aleatorie $\{Y_n\}_n$ converge in probabilità e, se sì, a quale valore.

Soluzioni

Problema 1

Lo spazio di probabilità è uniforme discreto. Il numero di casi totali è $5! = 120$.

1. Le lettere M e A possono formare un nuovo simbolo unico. Il numero di casi favorevoli consiste nel contare le permutazioni di 4 simboli, e raddoppiare questo numero per considerare sia i casi MA che i casi AM. Dunque la probabilità è
$$\frac{4! \cdot 2}{5!} = \frac{2}{5}. \tag{1}$$
2. Questa volta la lettera S è vincolata a stare nella prima posizione, dunque abbiamo 3 simboli (T, I, MA o AM) che possono essere permutati. La probabilità è:
$$\frac{3! \cdot 2}{5!} = \frac{2}{4 \cdot 5} = \frac{1}{10} \tag{2}$$

Problema 2

La somma di variabili aleatorie Gaussiane è una v.a. Gaussiana, dunque rimane da determinare valore atteso e varianza di W. Per il valore atteso abbiamo

$$\mathsf{E}[W] = \mathsf{E}[2X - Y] = 2 \cdot 1 - 2 = 0 \tag{3}$$

mentre per la varianza abbiamo

$$\begin{aligned}
\mathsf{Var}[W] = \mathsf{Var}[2X - Y] &= \mathsf{Var}[2X] + \mathsf{Var}[-Y] + 2\mathsf{Cov}[2X, -Y] && (4)\\
&= 4\mathsf{Var}[X] + \mathsf{Var}[Y] - 4\mathsf{Cov}[X, Y] && (5)\\
&= 4\mathsf{Var}[X] + \mathsf{Var}[Y] - 4\rho[X, Y]\sqrt{\mathsf{Var}[X]\mathsf{Var}[Y]} && (6)\\
&= 4 \cdot 1 + 3 - 4 \cdot 0.7 \cdot \sqrt{3} \approx 2.1503. && (7)
\end{aligned}$$

Dunque $W \sim \mathcal{N}(0, 2.1503)$. La probabilità si può calcolare come

$$\begin{aligned}
\Pr(0 < W < 2) &= \Pr\left(0 < \frac{W}{\sqrt{2.1503}} < \frac{2}{\sqrt{2.1503}}\right) && (8)\\
&= \Pr(0 < Z < 1.3639) && (9)\\
&= \Phi(1.3639) - \Phi(0) \approx 0.91309 - 0.5 = 0.41309, && (10)
\end{aligned}$$

dove $Z \sim \mathcal{N}(0, 1)$.

Problema 3

Innanzitutto si noti che $V \in \left[0, \frac{4}{3}\pi\right]$, perchè il raggio massimo è unitario. Applicando il metodo della cumulata si ha che

$$\begin{aligned}
F_V(v) = \Pr(V \le v) &= \Pr\left(\frac{4}{3}\pi R^3 \le v\right) && (11)\\
&= \Pr\left(R \le \sqrt[3]{\frac{3}{4\pi}v}\right) && (12)\\
&= F_R\left(\sqrt[3]{\frac{3}{4\pi}v}\right), \qquad v \in \left[0, \frac{4}{3}\pi\right]. && (13)
\end{aligned}$$

La densità di probabilità è:

$$\begin{aligned}
f_V(v) = \frac{d}{dv}F_R\left(\sqrt[3]{\frac{3}{4\pi}v}\right) &= \frac{dF_R}{dv}\left(\sqrt[3]{\frac{3}{4\pi}v}\right) \cdot \frac{d}{dv}\left(\sqrt[3]{\frac{3}{4\pi}v}\right) && (14)\\
&= f_R\left(\sqrt[3]{\frac{3}{4\pi}v}\right) \cdot \sqrt[3]{\frac{3}{4\pi}}\frac{v^{-\frac{2}{3}}}{3} && (15)\\
&= 12\left(\frac{3}{4\pi}v\right)^{\frac{2}{3}}\left(1 - \sqrt[3]{\frac{3}{4\pi}v}\right) \cdot \sqrt[3]{\frac{3}{4\pi}}\frac{v^{-\frac{2}{3}}}{3} && (16)\\
&= \frac{3}{\pi}\left(1 - \sqrt[3]{\frac{3}{4\pi}v}\right), \qquad v \in \left[0, \frac{4}{3}\pi\right], && (17)
\end{aligned}$$

e $f_V(v) = 0$ altrimenti.

Problema 4

Possiamo modellizzare il problema con un collage di processi di Bernoulli. Il primo processo di Bernoulli si ferma al giorno in cui si trova il primo personaggio. Il secondo processo inizia il giorno successivo alla fine del primo e termina quando si trova il secondo personaggio, e così via. Il parametro (la probabilità di successo in un giorno) dell'i-esimo processo di Bernoulli è

$$p_i = (1 - 0.7) \cdot \frac{6-i}{5}, \qquad i = 1, 2, 3, 4, 5. \tag{18}$$

Per rispondere alle domande:

1. sia X_i la lunghezza dell'i-esimo processo di Bernoulli. Si noti che $X_i \sim \text{Geom}(p_i)$. Allora il numero di giorni medi per completare la collezione è:

$$\sum_{i=1}^{5} \mathsf{E}[X_i] = \sum_{i=1}^{5} \frac{1}{p_i} = \frac{5}{0.3}\left(\frac{1}{5} + \frac{1}{4} + \frac{1}{3} + \frac{1}{2} + 1\right) \approx 38.0556 \tag{19}$$

2. Si sta chiedendo di trovare il più piccolo valore (intero) di x che soddisfa la disequazione:

$$\Pr(X_1 \le x) \overset{!}{\geq} 0.95. \tag{20}$$

Siccome $X_1 \sim \text{Geom}(0.3)$, dobbiamo risolvere

$$1 - (0.7)^x \overset{!}{\geq} 0.95, \tag{21}$$

ovvero

$$(0.7)^x \overset{!}{\leq} 0.05 \Rightarrow x \overset{!}{\geq} \left\lceil \frac{\log 0.05}{\log 0.7} \right\rceil = 9. \tag{22}$$

Dunque ci vogliono almeno 9 giorni.

Problema 5

Innanzitutto, si noti che l'integrale può essere reinterpretato nel modo seguente:

$$I = \int_0^1 e^{\tan x} dx = \int_0^1 e^{\tan u} f_U(u) du = \mathsf{E}[e^{\tan U}] \tag{23}$$

dove $U \sim \mathcal{U}[0, 1]$. La media campionaria ci dà una stima di I tale che $\mathsf{E}[\widehat{I}_n] = I$, infatti:

$$\mathsf{E}[\widehat{I}_n] = \mathsf{E}\left[\frac{1}{n}\sum_{i=1}^{n} e^{\tan(U_i)}\right] = \frac{1}{n}\sum_{i=1}^{n} \mathsf{E}\left[e^{\tan(U_i)}\right] = \frac{1}{n}\sum_{i=1}^{n} I = I, \tag{24}$$

dove le U_i sono v.a. iid distribuite come $\mathcal{U}[0, 1]$. Lo pseudocodice dell'algoritmo Monte Carlo è come segue:

1. Genero n campioni $U_i \sim \mathcal{U}[0, 1]$ in maniera indipendente
2. Calcolo $\widehat{I}_n = \frac{1}{n}\sum_{i=1}^{n} e^{\tan(U_i)}$

La varianza dell'errore di stima è

$$\mathsf{Var}[I - \widehat{I}_n] = \mathsf{Var}[\widehat{I}_n] = \mathsf{Var}\left[\frac{1}{n}\sum_{i=1}^{n} e^{\tan(U_i)}\right] = \frac{1}{n}\mathsf{Var}[e^{\tan U_1}]. \tag{25}$$

Siccome $\mathsf{Var}[e^{\tan U_1}]$ è un valore indipendente da n, si può notare una relazione di proporzionalità inversa tra la varianza dell'errore di stima e il numero di campioni n utilizzati nella stima Monte Carlo.

Problema 6

Testiamo la convergenza della successione $\{Y_n\}_n$ al valore a tramite la definizione:

$$\Pr(|Y_n - a| > \varepsilon) = \Pr(|X_n| > \varepsilon) \le \frac{\mathsf{E}[|X_n|]}{\varepsilon} = \frac{1}{n\varepsilon} \overset{n\to\infty}{\longrightarrow} 0, \qquad \forall \varepsilon > 0, \tag{26}$$

dove abbiamo usato la diseguaglianza di Markov. Dunque abbiamo che $Y_n \xrightarrow{P} a$.