Informazione e stima – 19/07/2021

- 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
- Se consegnato online, nominare il file solo con il proprio codice persona
- Esercizio da escludere dal punteggio finale:

1 Si lancia 100 volte un dado a 6 facce non truccato. Si calcoli la probabilità di ottenere esattamente 30 volte un risultato maggiore o uguale 5, sia in modo esatto sia utilizzando un'appropriata approssimazione.

2 X e Y sono variabili casuali Gaussiane con coefficiente di correlazione $\rho[X,Y]$ non nullo, valore medio nullo e varianza 4 e 9, rispettivamente. Si calcolino il valore massimo e minimo possibili della varianza della v.a. $Z = X + Y$.

3 Sia $X \sim \text{Exp}(\lambda)$. Determinare la densità di probabilità di $Y = \log(X)$.

4 Sia $X_n \sim \mathcal{N}\left(3, \frac{1}{n}\right)$, per $n \in \mathbb{N}$, una successione di v.a. indipendenti, e sia $Y_n = (-1)^n X_n$. Determinare se la successione $\{Y_n\}_{n=1}^{\infty}$ converge in probabilità e, se sì, a quale valore.

5 La distribuzione standard di Cauchy è data da $f_X(x) = \frac{1}{\pi(1+x^2)}$ per $x \in \mathbb{R}$. Partendo da un generatore di campioni indipendenti U_i distribuiti come $U \sim \mathcal{U}[0,1]$, descrivere un algoritmo per generare campioni distribuiti come X. Quant'è l'efficienza media dell'algoritmo proposto?
Si ricorda che $\frac{d}{dx} \arctan(x) = \frac{1}{1+x^2}$.

6 Si vogliono salvare su file i primi 1000 istanti di arrivo di un processo di Bernoulli con parametro $p = 0.3$. In particolare, si decide di salvare su file il valore di tutti i primi 1000 tempi di interarrivo. Mediamente quanto sarà lungo (in bit) il più piccolo file che contiene queste informazioni?

Soluzioni

Problema 1

La probabilità di ottenere un numero maggiore o uguale a 5 in un lancio di dado è $\frac{1}{3}$. La probabilità di avere 30 successi in 100 prove indipendenti è una prob Binomiale:

$$\binom{100}{30}\left(\frac{1}{3}\right)^{30}\left(\frac{2}{3}\right)^{70} \approx 0.0673. \tag{1}$$

Utilizzando l'approssimazione di De Moivre-Laplace per la Binomiale, si ha:

$$\begin{aligned}
\Pr(X=30) &= \Pr(29.5 \le X \le 30.5) && (2)\\
&= \Pr\left(\frac{29.5-\frac{100}{3}}{\sqrt{\frac{100}{3}\frac{2}{3}}} \le \frac{X-\frac{100}{3}}{\sqrt{\frac{100}{3}\frac{2}{3}}} \le \frac{30.5-\frac{100}{3}}{\sqrt{\frac{100}{3}\frac{2}{3}}}\right) && (3)\\
&\approx \Pr\left(-0.8132 \le Z \le -0.601\right) && (4)\\
&= \Pr\left(0.601 \le Z \le 0.8132\right) && (5)\\
&= \Phi(0.8132) - \Phi(0.601) && (6)\\
&\approx 0.791 - 0.7291 = 0.0619. && (7)
\end{aligned}$$

Problema 2

Poiché

$$\begin{aligned}
\mathsf{E}\left[Z^2\right] &= \mathsf{E}\left[(X+Y)^2\right] = \mathsf{E}\left[X^2\right] + \mathsf{E}\left[Y^2\right] + 2\mathsf{E}\left[XY\right] && (8)\\
\mathsf{E}\left[XY\right] &= \mathsf{Cov}[X,Y] + \mathsf{E}[X]\mathsf{E}[Y] = \rho[X,Y]\sqrt{\mathsf{Var}[X]\mathsf{Var}[Y]}, && (9)
\end{aligned}$$

dove l'ultimo passaggio è dovuto a $\mathsf{E}[X] = \mathsf{E}[Y] = 0$, si ottiene

$$\mathsf{Var}[Z] = \mathsf{E}\left[X^2\right] + \mathsf{E}\left[Y^2\right] + 2\rho[X,Y]\sqrt{\mathsf{Var}[X]\mathsf{Var}[Y]} = 13 + 12\rho[X,Y]. \tag{10}$$

Sapendo dalla teoria $-1 \le \rho[X,Y] \le 1$, i valori massimo e minimo che $\mathsf{Var}[Z]$ può assumere sono

$$\begin{aligned}
\mathsf{Var}[Z] &\le 13 + 12 = 25 && (11)\\
\mathsf{Var}[Z] &\ge 13 - 12 = 1. && (12)
\end{aligned}$$

Problema 3

La cumulata di probabilità di Y è

$$\begin{aligned}
F_Y(y) = \Pr(Y \le y) &= \Pr(\log(X) \le y) && (13)\\
&= \Pr(X \le e^y) && (14)\\
&= 1 - e^{-\lambda e^y}, \qquad y \in \mathbb{R} && (15)
\end{aligned}$$

e la densità di probabilità si ottiene derivando:

$$f_Y(y) = \lambda e^{-\lambda e^y + y}, \qquad y \in \mathbb{R}. \tag{16}$$

Problema 4

La successione $\{X_n\}$ sembra dover convergere al valore 3, dato che è una successione di v.a. Gaussiane con varianza che diminuisce in n e con valore medio fisso a 3. La successione $\{Y_n\}$ sembra essere oscillante in n, a causa del termine $(-1)^n$, dunque non dovrebbe esserci convergenza. Proviamo a verificarlo; per ogni $\varepsilon > 0$ deve succedere che

$$\begin{aligned}
\Pr(|Y_n - a| \le \varepsilon) &= \Pr(|(-1)^n X_n - a| \le \varepsilon) && (17)\\
&= \begin{cases} \Pr(a-\varepsilon \le X_n \le a+\varepsilon) & n \text{ pari} \\ \Pr(-a-\varepsilon \le X_n \le -a+\varepsilon) & n \text{ dispari} \end{cases} && (18)
\end{aligned}$$

deve tendere a 1, per $n \to \infty$. Questo però non è possibile, perché richiederebbe che per n pari X_n tenda al valore a, mentre negli n dispari al valore $-a$.

Problema 5

Un possibile metodo è quello della cumulata inversa. Calcoliamo la cumulata della distribuzione di Cauchy:

$$F_X(x) = \int_{-\infty}^{x} \frac{1}{\pi(1+t^2)} dt = \left[\frac{1}{\pi}\arctan(t)\right]_{-\infty}^{x} = \frac{1}{\pi}\arctan(x) + \frac{1}{2}, \qquad x \in \mathbb{R}. \tag{19}$$

Determiniamo la funzione inversa della cumulata:

$$u = \frac{1}{\pi}\arctan(x) + \frac{1}{2} \tag{20}$$

$$\left(u - \frac{1}{2}\right)\pi = \arctan(x) \tag{21}$$

$$x = \tan\left(\left(u - \frac{1}{2}\right)\pi\right). \tag{22}$$

L'algoritmo funziona come segue:

1. Genero campioni $U_i \sim \mathcal{U}[0,1]$ in maniera indipendente
2. Pongo $X_i = \tan\left(\left(U_i - \frac{1}{2}\right)\pi\right)$.

Per ogni campione uniforme generato, ottengo un campione valido della Cauchy, dunque l'efficienza è del 100%.

Problema 6

Siano $T_i \sim \text{Geom}(p = 0.3)$ i tempi di interarrivo del processo di Bernoulli. L'entropia totale generata è

$$\begin{aligned}
H(T_1, T_2, \ldots, T_{1000}) &= \sum_{i=1}^{1000} H(T_i) && (23)\\
\text{(identic. distr.)} &= 1000 H(T_1) && (24)\\
&= 1000\mathsf{E}[-\log_2 p_{T_1}(T_1)] && (25)\\
&= 1000\mathsf{E}[-\log_2(0.3 \cdot 0.7^{T_1 - 1})] && (26)\\
&= -1000\,[\log_2(0.3) + \log_2(0.7)\mathsf{E}[T_1 - 1]] && (27)\\
&= -1000\left[\log_2(0.3) + \log_2(0.7)\left(\frac{1}{0.3} - 1\right)\right] && (28)\\
&\approx 2937.6 \text{ bit} && (29)
\end{aligned}$$

dove il primo passaggio è dovuto all'indipendenza di tutte le v.a. Dal teorema di codifica di sorgente sappiamo che mediamente il più piccolo file che contiene queste informazione non può essere più piccolo di 2937.6 bit.