Informazione e stima – 06/07/2022

- 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 Assumendo di avere X persone in una stanza, con

$$p_X(x) = \begin{cases} 1/4 & x \in \{5, 10\} \\ 1/2 & x = 15. \end{cases}$$

Qual è la probabilità che almeno due persone siano nate nello stesso mese? Si può assumere che le persone siano nate con distribuzione uniforme sui mesi e indipendente dai mesi di nascita delle altre persone.

2 Sia $X \sim \text{Exp}(1)$. Calcolare la legge di probabilità di $Y = \lfloor X \rfloor$, ovvero del più grande intero minore o uguale a X.

3 Si consideri $Z \sim \mathcal{N}(0,1)$ e la successione di v.a. $\{X_n\}$ con $X_n = Z^{2n}$ per $n \in \mathbb{N}$. Dire se la successione $\{X_n\}$ converge in probabilità e, se sì, a quale valore.

4 Due lampadine hanno una vita di durata $X \sim \text{Exp}(1)$ e $Y \sim \text{Exp}(1)$ con X e Y indipendenti. Supponendo di accenderle nello stesso istante, calcolare la probabilità di vederle ancora accese al tempo t.

5 Si consideri una v.a. X di Weibull con legge $f_X(x) = xe^{-x^2/2}$ per $x \geq 0$ e $f_X(x) = 0$ altrimenti. Partendo da un generatore di v.a. uniformi $\mathcal{U}[0,1)$, proporre un algoritmo (pseudocodice) per campionare da f_X. Qual è l'efficienza dell'algoritmo proposto?
Suggerimento: il massimo della funzione $xe^{-\frac{x^2}{2}+x}$ *per* $x \geq 0$ *è* $\frac{1}{2}(1+\sqrt{5})e^{\frac{\sqrt{5}-1}{4}}$

6 Si consideri un esperimento aleatorio con 5 risultati possibili $X \in \{1,2,3,4,5\}$ e con legge $p_X(x) = 1/4$ per $x \in \{1,2,3\}$, e $p_X(x) = 1/8$ per $x \in \{4,5\}$. Proporre un codice prefix free con lunghezza media più piccola possibile. Confrontare la lunghezza media del codice con l'entropia dell'esperimento aleatorio.

Soluzioni

Problema 1

Usiamo il teorema delle probabilità totali per considerare i 3 casi $x = \{5, 10, 15\}$ separatamente.

Il problema è simile al problema del compleanno. Si può partire dall'evento complementare a quello di interesse, dove tutte le persone sono nate in mesi diversi:

$$\Pr(x \text{ persone nate in mesi diversi}) = \frac{12 \cdot 11 \cdot \ldots (12 - x + 1)}{\underbrace{12 \cdot 12 \cdot \ldots 12}_{x \text{ volte}}} = \frac{12!}{(12 - x)! \cdot 12^x} \tag{1}$$

per $2 \leq x \leq 12$, e dunque si ha che

$$\Pr(\text{almeno due tra } x \text{ persone nate stesso mese}) = 1 - \frac{12!}{(12 - x)! \cdot 12^x} \tag{2}$$

per $2 \leq x \leq 12$. Usando il teorema delle probabilità totali si ha:

$$\begin{aligned}
\Pr(\text{almeno due tra } X \text{ persone nate stesso mese}) &= \sum_{x \in \{5,10,15\}} \Pr(\text{almeno due tra } x \text{ persone nate stesso mese}) p_X(x) \quad (3) \\
&= \frac{1}{4} \left(1 - \frac{12!}{(12-5)! \cdot 12^5} + 1 - \frac{12!}{(12-10)! \cdot 12^{10}} \right) + \frac{1}{2} \quad (4)
\end{aligned}$$

dove nell'ultimo passaggio abbiamo riconosciuto che

$$\Pr(\text{almeno due tra 15 persone nate stesso mese}) = 1. \tag{5}$$

Problema 2

Innanzitutto si noti che $\Pr(Y < 0) = 0$, e che Y può assumere solo valori interi non negativi. Si ha che

$$\begin{aligned}
\Pr(Y = y) &= \Pr(y \leq X < y + 1) && (6) \\
&= \int_y^{y+1} f_X(x) dx && (7) \\
&= \int_y^{y+1} e^{-x} dx && (8) \\
&= \left[-e^{-x} \right]_y^{y+1} && (9) \\
&= e^{-y} - e^{-(y+1)} && (10) \\
&= e^{-y}(1 - e^{-1}) && (11)
\end{aligned}$$

per $y \in \{0, 1, 2, \ldots\}$. Si noti che $Y + 1 \sim \text{Geom}(1 - e^{-1})$.

Problema 3

Si può notare che

$$\lim_{n \to \infty} z^{2n} = \begin{cases} 0 & -1 < z < 1 \\ 1 & z = \pm 1 \\ +\infty & z > 1 \vee z < -1 \end{cases} \tag{12}$$

Ciò suggerisce che la probabilità che la successione $\{X_n\}$ si concentri attorno ad un particolare valore non è mai 1. Ad esempio, testando la convergenza nel punto 0 si ottiene:

$$\begin{aligned}
\lim_{n \to \infty} \Pr(X_n < \varepsilon) &= \lim_{n \to \infty} \Pr(Z^{2n} < \varepsilon) && (13) \\
&= \lim_{n \to \infty} \Pr(-\varepsilon^{\frac{1}{2n}} < Z < \varepsilon^{\frac{1}{2n}}) && (14) \\
&= \lim_{n \to \infty} 2 \left(\Phi(\varepsilon^{\frac{1}{2n}}) - \frac{1}{2} \right) && (15) \\
&= 2 \left(\Phi(1) - \frac{1}{2} \right) < 1. && (16)
\end{aligned}$$

Problema 4

La probabilità cercata è:

$$\begin{aligned} \Pr(X > t, Y > t) &= \Pr(X > t)\Pr(Y > t) && (17)\\ &= e^{-t}e^{-t} && (18)\\ &= e^{-2t}, \qquad t \geq 0. && (19) \end{aligned}$$

dove abbiamo usato l'indipendenza di X e Y e il fatto che

$$\Pr(X > t) = \Pr(Y > t) = 1 - F_X(t) = e^{-t} \qquad t \geq 0. \tag{20}$$

Problema 5

Si potrebbe usare un algoritmo di acceptance-rejection che parte dal campionamento di una v.a. $Y \sim \mathrm{Exp}(1)$, perché c'è bisogno di campionare da una v.a. con supporto $[0, \infty)$. Bisogna trovare il minimo valore di m tale che:

$$m f_Y(x) \geq f_X(x), \qquad x \geq 0 \tag{21}$$

$$m e^{-x} \geq x e^{-x^2/2}, \qquad x \geq 0 \tag{22}$$

$$m \geq x e^{-x^2/2+x}, \qquad x \geq 0. \tag{23}$$

Sfruttando il suggerimento, possiamo assegnare ad m il valore massimo della funzione $xe^{-x^2/2+x}$ per $x \geq 0$:

$$m = \frac{1}{2}(1+\sqrt{5})e^{\frac{\sqrt{5}-1}{4}} \approx 2.204. \tag{24}$$

L'algoritmo acceptance-rejection è:

1. Generare $U \sim \mathcal{U}[0,1)$ e porre $Y = -\ln(U)$.
2. Generare $U' \sim \mathcal{U}[0,1)$
3. Accettare e porre $X = Y$ se $mU' \leq \frac{f_X(Y)}{f_Y(Y)}$, altrimenti tornare al punto 1.

L'efficienza dell'algoritmo è pari a $1/m \approx 45.37\%$.

Alternativamente, si poteva usare il metodo della cumulata inversa. La cumulata della Weibull è:

$$F_X(x) = \int_0^x t e^{-t^2/2} dt = 1 - e^{-x^2/2} \qquad x \geq 0, \tag{25}$$

la cui funzione inversa è:

$$F_X^{-1}(u) = \sqrt{-2\log(1-u)} \qquad u \in [0,1). \tag{26}$$

L'algoritmo è il seguente:

1. Generare $U \sim \mathcal{U}[0,1]$
2. Porre $X = \sqrt{-2\log(1-U)}$.

L'efficienza è del 100%.

Problema 6

Innanzitutto notiamo come le probabilità siano tutte potenze intere negative di 2, dunque le autoinformazioni degli eventi sono numeri interi di bit:

$$i(x) = \log_2 \frac{1}{p_X(x)} = \begin{cases} 2 & x \in \{1,2,3\} \\ 3 & x \in \{4,5\} \end{cases} \tag{27}$$

Ciò suggerisce che sia possibile trovare un prefix free code con tre messaggi lunghi 2 bit e due messaggi lunghi 3 bit. Ad esempio, il codice potrebbe essere:

$$c(1) = 00 \tag{28}$$
$$c(2) = 01 \tag{29}$$
$$c(3) = 10 \tag{30}$$
$$c(4) = 110 \tag{31}$$
$$c(5) = 111. \tag{32}$$

La lunghezza media del codice è:

$$\mathsf{E}[L] = 2 \cdot \frac{3}{4} + 3 \cdot \frac{1}{4} = \frac{9}{4} = 2.25. \tag{33}$$

L'entropia dell'esperimento è:

$$H(X) = \mathsf{E}[i(X)] = 2 \cdot \frac{3}{4} + 3 \cdot \frac{1}{4} = \mathsf{E}[L]. \tag{34}$$

Dato che $\mathsf{E}[L] = H(X)$, non esiste un codice migliore di questo.