Informazione e stima – 13/01/2021

- Ogni esercizio viene valutato da 0 a 6 punti.
- Non riportare solo il risultato, ma cerca di argomentare sinteticamente la risposta.
- Riportare il proprio nome, cognome, e codice persona su ogni foglio consegnato.
- Nominare il file da caricare con il proprio codice persona.
- Indicare l'esercizio da escludere dal punteggio finale.

1. Si pescano 5 carte da un mazzo ben mescolato di 52 carte. Qual è la probabilità di ottenere una doppia coppia?

2. Due variabili aleatorie X e Y hanno legge di probabilità $f_{X,Y}(x, y) = c$ nel disco di raggio r centrato nell'origine. Determinare:
   (a) il valore della costante c.
   (b) Le legge di $T = X^2 + Y^2$. (Suggerimento: calcolare la legge cumulata di T, e ragionare sul significato geometrico dell'evento associato alla cumulata)

3. Siano $X_1, X_2, \cdots, X_{1000}$ delle v.a. i.i.d. con $X_i \sim \text{Exp}(1)$, $i = 1, \ldots, 1000$. Calcolare una stima accurata di $\Pr(\sum_{i=1}^{1000} X_i \leq 1017)$.

4. Si consideri un processo di Bernoulli con parametro $p = 1/3$. Ogni 3 arrivi, si smista l'arrivo in un nuovo processo.
   (a) Il nuovo processo è di Bernoulli? Giustificare la risposta.
   (b) Come sono distribuiti i tempi di interarrivo nel nuovo processo?

5. Si vuole determinare numericamente il valore di $I = \Pr(0 \leq X \leq 1)$, dove $X \sim \text{Exp}(\lambda)$. Partendo da un generatore di campioni distribuiti come $U \sim \mathcal{U}[0, 1]$, proporre un algoritmo che produce una stima $\widehat{I}$ di I.

6. Si consideri la v.a. X che conta il numero di lanci di un dado ben bilanciato per ottenere il risultato 1. Mediamente quanti bit di informazione sono prodotti dall'osservazione di X?

Soluzioni

Problema 1

I casi totali sono $\binom{52}{5}$. Per contare i casi favorevoli, procediamo come segue. Essendoci 13 valori diversi nel mazzo di 52, bisogna scegliere:

- i 2 valori tra i 13 disponibili che formeranno la doppia coppia. Le scelte sono $\binom{13}{2}$.
- i semi che formano le coppie. Ogni coppia ha $\binom{4}{2}$ scelte possibili. Dunque in totale si hanno $\binom{4}{2}^2 = 36$ scelte.
- la carta che non farà parte delle coppie. Dobbiamo innanzitutto sceglierne il valore, e poi il seme. Le scelte rimanenti per il valore sono $13 - 2 = 11$, e 4 scelte per il seme. In totale fa 44.

La probabilità cercata è

$$p = \frac{\binom{13}{2} \cdot 36 \cdot 44}{\binom{52}{5}} = 0.0475 \tag{1}$$

Problema 2

1. Siccome la legge congiunta è uniforme nel disco di raggio r, la costante c è pari al reciproco dell'area del disco:
$$f_{X,Y}(x,y) = \frac{1}{\pi r^2}, \qquad (x,y) : \ 0 \leq x^2 + y^2 \leq r^2.$$
2. La legge cumulata di T si calcola come segue
$$\Pr(T \leq t) = \Pr(X^2 + Y^2 \leq t) \tag{2}$$
che geometricamente si può interpretare come il calcolo della probabilità che un punto (X,Y) lanciato casualmente nel disco di raggio r cada nel disco di raggio $\sqrt{t}$. Questa probabilità si calcola facilmente come il rapporto tra aree, essendo lo spazio di probabilità uniforme:
$$\Pr(T \leq t) = \begin{cases} 0 & t < 0 \\ \frac{\pi t}{\pi r^2} = \frac{t}{r^2} & 0 \leq t \leq r^2 \\ 1 & t > r^2 \end{cases}$$
La distribuzione di probabilità si ottiene calcolando la derivata rispetto a t, ottenendo $T \sim \mathcal{U}[0, r^2]$.

Problema 3

In questo caso si può applicare il CLT. Si noti che $\mathsf{E}[X_i] = 1$ e $\mathsf{Var}[X_i] = 1$. Standardizzando l'evento di interesse, si ottiene:

$$\Pr\left(\frac{\sum_{i=1}^{1000} X_i - 1000}{\sqrt{1000}} \leq \frac{1017 - 1000}{\sqrt{1000}}\right) \overset{CLT}{\approx} \Pr\left(Z \leq 0.54\right) \tag{3}$$

$$= \Phi(0.54) \approx 0.7054. \tag{4}$$

Problema 4

1. Il nuovo processo non è di Bernoulli, perché la legge di smistamento è deterministica nel tempo. Ad esempio, se nel processo nuovo c'è un arrivo al tempo t, al tempo $t+1$ non ci può essere un nuovo arrivo; dunque la probabilità di avere un arrivo nel nuovo processo cambia nel tempo.
2. Sia $T_i \sim \text{Geom}(1/3)$ il tempo di interarrivo $i-$esimo nel processo originale, e X_i il tempo di interarrivo $i-$esimo nel nuovo processo. Allora si ha
$$X_i = T_{3i-2} + T_{3i-1} + T_{3i} \sim \text{Pascal}(p = 1/3, k = 3)$$
Da notare che tutti i tempi X_i sono indipendenti.

Problema 5

L'integrale I si può stimare numericamente ricorrendo ad una simulazione Monte Carlo. In particolare, si può notare che:

$$I = \int_0^1 \lambda e^{-\lambda x} dx = \int_0^1 \frac{\lambda e^{-\lambda x}}{f_U(x)} f_U(x) dx = \mathsf{E}\left[\frac{\lambda e^{-\lambda U}}{f_U(U)}\right] = \mathsf{E}\left[\lambda e^{-\lambda U}\right]$$

dove $U \sim \mathcal{U}[0,1]$ e $f_U(u) = 1$ per $0 \leq u \leq 1$. L'algoritmo è fatto come segue:

1. Genero n campioni indipendenti $U_i \sim \mathcal{U}[0,1]$, per $i = 1, \ldots, n$.
2. Calcolo $\widehat{I} = \frac{1}{n} \sum_{i=1}^n \lambda e^{-\lambda U_i}$.

Problema 6

Sappiamo che $X \sim \text{Geom}(1/6)$, e dunque

$$p_X(x) = \frac{1}{6}\left(1 - \frac{1}{6}\right)^{x-1}, \qquad x = 1, 2, \cdots .$$

Il numero medio di bit di informazione dati dall'osservazione di X si può calcolare tramite l'entropia. L'autoinformazione dell'evento $X = x$ è

$$i(x) = \log_2 \frac{1}{p_X(x)} = \log_2\left(6\left(\frac{6}{5}\right)^{x-1}\right) = \log_2(6) + (x-1)\log_2\frac{6}{5}, \qquad x = 1, 2, \cdots .$$

L'entropia di X è:

$$\begin{aligned} H(X) = \mathsf{E}[i(X)] &= \log_2(6) + (\mathsf{E}[X] - 1)\log_2\frac{6}{5} & (5) \\ &= \log_2(6) + (6-1)\log_2\frac{6}{5} & (6) \\ &\approx 3.9001 \text{ bit} & (7) \end{aligned}$$