MA 589 | MID-SEMESTER EXAMINATION
STATISTICAL FOUNDATION FOR DATA SCIENCE | 09:00 – 11:00 IST
IIT GUWAHATI | SEPTEMBER 21, 2025

MODEL ANSWERS OF MID-SEMESTER EXAMINATION (TOTAL POINTS:30)

1. (5 points) Consider four coding machines M_1, M_2, M_3, and M_4 producing binary codes 0 and 1. The machine M_1 produces codes 0 and 1 with respective probabilities $\frac{1}{4}$ and $\frac{3}{4}$. The code produced by machine M_k is fed into machine M_{k+1}, ($k = 1, 2, 3$), which may either leave the received code unchanged or may change it. Suppose that each of the machines M_2, M_3, and M_4 change the code with probability $\frac{3}{4}$. Given that the machine M_4 has produced code 1, find the conditional probability that the machine M_1 produced code 0.

   Solution: Let A_{ij} denotes the event that jth machine produces the code i, $i = 0, 1$ and $j = 1, 2, 3, 4$. Using Bayes theorem, the required probability

$$P(A_{01}|A_{14}) = \frac{P(A_{14}|A_{01})P(A_{01})}{P(A_{14}|A_{01})P(A_{01}) + P(A_{14}|A_{11})P(A_{11})}.$$

   Now, each code can change maximum 3 times. If M_1 produce code 0 and M_4 produce code 1, then the code has changed odd number times, *i.e.* either once or thrice. Similarly, if both M_1 and M_4 produce code 1, then the code has changed even number times, *i.e.* either twice or none. Hence

$$P(A_{14}|A_{01}) = \binom{3}{1}\left(\frac{3}{4}\right)\left(\frac{1}{4}\right)^2 + \binom{3}{3}\left(\frac{3}{4}\right)^3\left(\frac{1}{4}\right)^0 = \frac{9}{16},$$
$$P(A_{14}|A_{11}) = \binom{3}{0}\left(\frac{3}{4}\right)^0\left(\frac{1}{4}\right)^3 + \binom{3}{2}\left(\frac{3}{4}\right)^2\left(\frac{1}{4}\right) = \frac{7}{16}.$$

   Hence the required probability is $P(A_{01}|A_{14}) = \frac{3}{10}$.

2. (5 points) At a movie theater, the manager announces that a free ticket will be given to the first person whose birthday is same as someone who has already bought a ticket. You have the option of getting in the queue at any time, *i.e.*, at any position in the queue. Assuming that you don't know the birthday of anyone else and that the birthdays are uniformly distributed through out 365 days of the year, which position in the queue gives you the best chance of winning the free ticket?

   Solution: Let $P(n)$ denote the probability of winning the free ticket standing at the nth position. Note that there are $n-1$ persons before me, if I am standing at the n-th position of the queue. To win a free ticket by standing at n-th position, all $n-1$ persons before me must have distinct birthdays and my birthday has to match with one of the $n-1$ persons standing before me. Thus,

$$P(n) = \frac{365 \times 364 \times \ldots \times (365-(n-2))}{(365)^{n-1}} \times \frac{n-1}{365}.$$

   We need to find n for which $P(n)$ is maximum. Now

$$\frac{P(n+1)}{P(n)} = \frac{365 \times 364 \times \ldots \times (366-n))}{(365)^n} \times \frac{n}{365} \times \frac{(365)^{n-1}}{365 \times 364 \times \ldots (367-n)} \times \frac{365}{n-1}$$
$$= \frac{n(366-n)}{365(n-1)}.$$

Thus $P(n+1) > P(n) \iff n^2 - n - 365 < 0 \iff (n-0.5)^2 < 365.25 \iff n < \sqrt{365.25} + 0.5 = 19.611$. Similarly, $P(n+1) < P(n) \iff n > 19.611$. Thus, the mode of the distribution is at $n = 20$. Hence, the best position to stand to win the ticket is 20th position.

3. Prove or disprove the following statements. Note that you have to provide appropriate logic to prove a statement and a counter example to disprove a statement.

(a) (3 points) Let $\mathcal{F}_1$ and $\mathcal{F}_2$ are two σ-fields on the sample space $\mathcal{S}$. Then, $\mathcal{F}_1 \cap \mathcal{F}_2$ is a σ-field on $\mathcal{S}$.

Solution: $\mathcal{S} \in \mathcal{F}_i$ for $i = 1, 2 \implies \mathcal{S} \in \mathcal{F}_1 \cap \mathcal{F}_2$.
Let $A \in \mathcal{F}_1 \cap \mathcal{F}_2$. Then $A \in \mathcal{F}_1$ and $A \in \mathcal{F}_2 \implies A^c \in \mathcal{F}_1$ and $A^c \in \mathcal{F}_2 \implies A^c \in \mathcal{F}_1 \cap \mathcal{F}_2$.
Let $A_1, A_2, \ldots \in \mathcal{F}_1 \cap \mathcal{F}_2$. Then,

$$\begin{aligned} A_i \in \mathcal{F}_1 \text{ and } A_i \in \mathcal{F}_2 \text{ for all } i = 1, 2, \ldots &\implies \cup_{i=1}^{\infty} A_i \in \mathcal{F}_1 \text{ and } \cup_{i=1}^{\infty} A_i \in \mathcal{F}_2 \\ &\implies \cup_{i=1}^{\infty} A_i \in \mathcal{F}_1 \cap \mathcal{F}_2. \end{aligned}$$

Therefore, $\mathcal{F}_1 \cap \mathcal{F}_2$ is a σ-field on $\mathcal{S}$, and the given statement is correct.

(b) (5 points) Let X be a continuous random variable. Also, assume that the second order raw moment of X exists. Then, the first order moment of X exists.

Solution: Note that the existence of the second order raw moment implies $E\left(X^2\right) < \infty$. We have to show that $E\left(|X|\right) < \infty$. Now, notice that for $|x| > 1$, $|x| \leq |x|^2 = x^2$. Denoting the PDF of X by $f_X(\cdot)$, we have

$$\begin{aligned} E\left(|X|\right) &= \int_{-\infty}^{+\infty} |x| f_X(x) dx \\ &= \int_{-\infty}^{-1} |x| f_X(x) dx + \int_{-1}^{+1} |x| f_X(x) dx + \int_{+1}^{+\infty} |x| f_X(x) dx \\ &\leq \int_{-\infty}^{-1} x^2 f_X(x) dx + \int_{-1}^{+1} f_X(x) dx + \int_{+1}^{+\infty} x^2 f_X(x) dx \\ &\leq \int_{-\infty}^{-1} x^2 f_X(x) dx + \int_{-1}^{+1} x^2 f_X(x) dx + \int_{+1}^{+\infty} x^2 f_X(x) dx + \int_{-1}^{+1} f_X(x) dx \\ &\qquad \text{as } \int_{-1}^{+1} x^2 f_X(x) dx \geq 0 \\ &\leq E\left(X^2\right) + 1 \quad \text{as } \int_{-1}^{+1} f_X(x) dx \leq 1 \\ &< \infty. \end{aligned}$$

Thus, the given statement is correct.

(c) (5 points) Let $f(\cdot)$ be a probability density function of a continuous random variable. Then $\lim_{x \to \infty} f(x) = 0$. (Hint: You may use the fact that $\sum_{n=1}^{\infty} \frac{1}{2^n}$ converges.)

Solution: Define the function $f : \mathbb{R} \to \mathbb{R}$ by

$$f(x) = \begin{cases} 1 & \text{if } n \leq x \leq n + \frac{1}{2^n} \text{ for } n = 1, 2, \ldots \\ 0 & \text{otherwise,} \end{cases}$$

Then $f(x) \geq 0$ for all $x \in \mathbb{R}$. Moreover,

$$\int_{-\infty}^{+\infty} f(x)dx = \sum_{n=1}^{\infty} \int_{n}^{n+\frac{1}{2^n}} dx = \sum_{n=1}^{\infty} \frac{1}{2^n} = 1.$$

Hence, $f(x)$ is a PDF of some continuous random variable. But, $\lim_{x\to\infty} f(x)$ does not exists. Therefore, given statement is not correct.

4. Let X and Y be two independent random variables with probability density functions

$$f_1(x) = \begin{cases} e^{-x} & \text{if } x > 0 \\ 0 & \text{otherwise} \end{cases} \qquad \text{and} \qquad f_2(y) = \begin{cases} e^{-2y} & \text{if } y > 0 \\ 0 & \text{otherwise,} \end{cases}$$

respectively. Let $U = \max\{0, \min\{X, Y\} - 1\}$.

(a) (5 points) Derive the cumulative distribution function of U.

Solution: The CDF of U is

$$\begin{aligned} F_U(u) &= P(U \leq u) \\ &= P\left(\max\{0, \min\{X, Y\} - 1\} \leq u\right) \\ &= \begin{cases} 0 & \text{if } u < 0 \\ P\left(\min\{X, Y\} - 1 \leq u\right) & \text{if } u \geq 0. \end{cases} \end{aligned}$$

Now, for $u \geq 0$,

$$\begin{aligned} P\left(\min\{X, Y\} - 1 \leq u\right) &= 1 - P\left(\min\{X, Y\} > u + 1\right) \\ &= 1 - P\left(X > u + 1\right) P\left(Y > u + 1\right) \\ &= 1 - e^{-3(u+1)}. \end{aligned}$$

Thus, the CDF of U is

$$F_U(u) = \begin{cases} 0 & \text{if } u < 0 \\ 1 - e^{-3(u+1)} & \text{if } u \geq 0. \end{cases}$$

(b) (2 points) Determine the type (discrete or continuous) of the random variable U.

Solution: The set of discontinuity of $F_U(\cdot)$ is $\{0\}$. Now, $P(U = 0) = 1 - e^{-3} \in (0, 1)$. Thus, U is neither discrete nor continuous random variable.

Additional Pages for Answers