MSO205: HINTS AND SOLUTIONS TO END SEMESTER EXAMINATION

1. Descriptive Type

Qn 1. (5 marks) It is known that the lifetime of a batch of electric bulbs follows the continuous probability distribution given by the p.d.f

$$f(x) = \begin{cases} \frac{1}{10}\exp(-\frac{x}{10}), & \text{if } x > 0, \\ 0, & \text{otherwise} \end{cases}.$$

In a box containing 100 electric bulbs from that batch, what is the probability (approximately) that 20 bulbs or more have a lifetime less or equal to $10 \ln 3$? Express the approximate value in terms of the DF of $N(0,1)$ distribution.

Answer: For $X \sim Exponential(10)$,

$$\mathbb{P}(X \leq 10\ln 3) = \int_0^{10\ln 3} \frac{1}{10}\exp\left(-\frac{t}{10}\right)\, dt \overset{t=10x}{=} \int_0^{\ln 3} \exp(-x)\, dx = 1 - \exp(-\ln 3) = 1 - \frac{1}{3} = \frac{2}{3}.$$

Let $X_1, \cdots, X_{100}$ denote the lifetime of the electric bulbs. Define

$$Y_i = \begin{cases} 1, & \text{if } X_i \leq 10\ln 3, \\ 0, & \text{otherwise} \end{cases}, \forall i = 1, \cdots, 100.$$

Then, Y_i's are i.i.d. with $Y_1 \sim Bernoulli(\mathbb{P}(X \leq 10\ln 3)) = Bernoulli(\frac{2}{3})$. Here, $\mathbb{E}Y_1 = \frac{2}{3}$ and $Var(Y_1) = \frac{2}{3} \times \frac{1}{3} = \frac{2}{9}$.

By the CLT, for large n the distribution of

$$\sqrt{n}\frac{\frac{1}{n}\sum_{i=1}^n Y_i - \frac{2}{3}}{\sqrt{\frac{2}{9}}}$$

is close to $N(0,1)$ in the sense of convergence in distribution. Putting $n = 100$, the distribution of $\frac{30}{\sqrt{2}}\left(\frac{1}{100}\sum_{i=1}^{100} Y_i - \frac{2}{3}\right)$ is close to $N(0,1)$ in the sense of convergence in distribution. The required probability is

$$\begin{aligned}
\mathbb{P}\left(\sum_{i=1}^{100} Y_i \geq 20\right) &= \mathbb{P}\left(\sum_{i=1}^{100} Y_i > 21\right) \\
&= \mathbb{P}\left(\frac{1}{100}\sum_{i=1}^{100} Y_i - \frac{2}{3} > \frac{21}{100} - \frac{2}{3}\right) \\
&= \mathbb{P}\left(\frac{1}{100}\sum_{i=1}^{100} Y_i - \frac{2}{3} > -\frac{63-200}{300}\right) \\
&= \mathbb{P}\left(\frac{30}{\sqrt{2}}\left\{\frac{1}{100}\sum_{i=1}^{100} Y_i - \frac{2}{3}\right\} > -\frac{137}{10\sqrt{2}}\right)
\end{aligned}$$

$$= 1 - \mathbb{P}\left(\frac{30}{\sqrt{2}}\left\{\frac{1}{100}\sum_{i=1}^{100} Y_i - \frac{2}{3}\right\} \leq -\frac{137}{10\sqrt{2}}\right)$$

Using the above convergence, an approximate value of the required probability is $1 - \Phi(-\frac{137}{10\sqrt{2}})$, where Φ is the DF of $N(0,1)$ distribution.

Alternative:

$$\begin{aligned}
\mathbb{P}\left(\sum_{i=1}^{100} Y_i \geq 20\right) &= \mathbb{P}\left(\frac{1}{100}\sum_{i=1}^{100} Y_i - \frac{2}{3} \geq \frac{1}{5} - \frac{2}{3}\right) \\
&= \mathbb{P}\left(\frac{1}{100}\sum_{i=1}^{100} Y_i - \frac{2}{3} \geq -\frac{7}{15}\right) \\
&= \mathbb{P}\left(\frac{30}{\sqrt{2}}\left\{\frac{1}{100}\sum_{i=1}^{100} Y_i - \frac{2}{3}\right\} \geq -7\sqrt{2}\right) \\
&= 1 - \mathbb{P}\left(\frac{30}{\sqrt{2}}\left\{\frac{1}{100}\sum_{i=1}^{100} Y_i - \frac{2}{3}\right\} < -7\sqrt{2}\right)
\end{aligned}$$

The approximate value $1 - \Phi(-7\sqrt{2})$ shall also be accepted.

Qn 2. (4 marks) Let X be a non-negative integer valued RV such that $\mathbb{E}X^2$ exists. Find the value of $\lim_{n\to\infty} n^2\ \mathbb{P}(X > n)$.

Answer: For any non-negative integer valued RV Y with finite $\mathbb{E}Y$, we know that $\sum_{n=1}^{\infty} \mathbb{P}(Y > n)$ converges.

As $\{\mathbb{P}(Y > n)\}_n$ is a monotone decreasing sequence of non-negative real numbers, by the Abel-Pringsheim Theorem, a necessary condition is

$$\lim_{n\to\infty} n\mathbb{P}(Y > n) = 0,$$

and in particular, $\lim_{n\to\infty} n^2\mathbb{P}(Y > n^2) = 0$.

Given that X is a non-negative integer valued RV, so is X^2. Applying the above result (putting X^2 instead of Y), we have

$$\lim_{n\to\infty} n^2\mathbb{P}(X > n) = \lim_{n\to\infty} n^2\mathbb{P}(X^2 > n^2) = 0.$$

Qn 3. $(1 + 5$ marks) Let X be a 3-dimensional random vector. State and prove the non-decreasing property for the joint distribution function of X.

Answer: Let $a_1 < b_1, a_2 < b_2, a_3 < b_3$. Then, $\prod_{j=1}^{3}(a_j, b_j]$ is a rectangle in $\mathbb{R}^3$. Observe that the co-ordinates of the vertices are made up of either a_j or b_j for each $j = 1, 2, \cdots, 3$. Let Δ_k^3 denote the set of vertices where exactly k many a_j's appear, for $k = 0, 1, 2, 3$. Then the complete set of vertices is $\cup_{k=0}^{3}\Delta_k^3$. In particular,

$$\begin{aligned}
\Delta_0^3 &= \{(b_1, b_2, b_3)\} \\
\Delta_1^3 &= \{(a_1, b_2, b_3), (b_1, a_2, b_3), (b_1, b_2, a_3)\} \\
\Delta_2^3 &= \{(a_1, a_2, b_3), (a_1, b_2, a_3), (b_1, a_2, a_3)\} \\
\Delta_3^3 &= \{(a_1, a_2, a_3)\}
\end{aligned}$$

For the 3-dimensional random vector $X = (X_1, X_2, X_3) : \Omega \to \mathbb{R}^3$, the non-decreasing property for the joint DF F_X is given by

$$\sum_{k=0}^{3}(-1)^k \sum_{x\in\Delta_k^3} F_X(x) = \mathbb{P}(X \in \prod_{j=1}^{3}(a_j, b_j]) = \mathbb{P}(a_1 < X_1 \leq b_1, a_2 < X_2 \leq b_2, a_3 < X_3 \leq b_3) \geq 0.$$

Since $\mathbb{P}$ is a probability function, we have $\mathbb{P}(X \in \prod_{j=1}^{3}(a_j, b_j]) = \mathbb{P}(a_1 < X_1 \leq b_1, a_2 < X_2 \leq b_2, a_3 < X_3 \leq b_3) \geq 0$.

To complete the proof, it is enough to show

$$\sum_{k=0}^{3}(-1)^k \sum_{x\in\Delta_k^3} F_X(x) = \mathbb{P}(a_1 < X_1 \leq b_1, a_2 < X_2 \leq b_2, a_3 < X_3 \leq b_3)$$

Consider the events

$$A_1 := (-\infty < X_1 \leq a_1, -\infty < X_2 \leq b_2, -\infty < X_3 \leq b_3),$$

$$A_2 := (-\infty < X_1 \leq b_1, -\infty < X_2 \leq a_2, -\infty < X_3 \leq b_3),$$

and

$$A_3 := (-\infty < X_1 \leq b_1, -\infty < X_2 \leq b_2, -\infty < X_3 \leq a_3).$$

Then,

$$\begin{aligned}
&\mathbb{P}(a_1 < X_1 \leq b_1, a_2 < X_2 \leq b_2, a_3 < X_3 \leq b_3)\\
&= \mathbb{P}(-\infty < X_1 \leq b_1, -\infty < X_2 \leq b_2, -\infty < X_3 \leq b_3) - \mathbb{P}(A_1 \cup A_2 \cup A_3)\\
&= F_X(b_1, b_2, b_3) - \mathbb{P}(A_1 \cup A_2 \cup A_3).
\end{aligned}$$

But, by the inclusion-exclusion principle,

$$\begin{aligned}
&\mathbb{P}(A_1 \cup A_2 \cup A_3)\\
&= \sum_{j=1}^{3}\mathbb{P}(A_j) - \sum_{j,l\in\{1,2,3\}, j\neq l}\mathbb{P}(A_j \cap A_k) + \mathbb{P}(A_1 \cap A_2 \cap A_3)\\
&= \sum_{k=1}^{3}(-1)^{k-1} \sum_{x\in\Delta_k^3} F_X(x)
\end{aligned}$$

and hence,

$$\mathbb{P}(a_1 < X_1 \leq b_1, a_2 < X_2 \leq b_2, a_3 < X_3 \leq b_3) = F_X(b_1, b_2, b_3) + \sum_{k=1}^{3}(-1)^k \sum_{x\in\Delta_k^3} F_X(x).$$

This completes the proof.

2. Short Answer Type

Qn 4 (i) (2 marks) Compute the mode for $Poisson(\sqrt{5})$ distribution.

Answer: 2

Qn 4 (ii) (1 + 1 + 1 + 1 marks) An RV X has the MGF $M_X(t) = \frac{e^t}{4t}(e^{4t} - 1)$ for $t \neq 0$, and 1 if $t = 0$. Write the values of $\mathbb{P}(X > 2|X < 3), \mathbb{E}X$ and $Var(X)$. Is $\mathbb{E}X > Var(X)$?

Answer: $\mathbb{P}(X > 2|X < 3) = \frac{1}{2}, \mathbb{E}X = 3$ and $Var(X) = \frac{4}{3}$. Yes, $\mathbb{E}X > Var(X)$. (Hint: $X \sim Uniform(1, 5)$)

Qn 4 (iii) (3 marks) Fix $p_1 = p_2 = p_3 = \frac{1}{8}, p_4 = \frac{5}{8}$. Let the random vector $(X_1, X_2, X_3)^t$ have the joint p.m.f. given by

$$\mathbb{P}(X_1 = x_1, X_2 = x_2, X_3 = x_3) = \frac{20!}{x_1!x_2!x_3!(20 - \sum_{j=1}^{3} x_j)!} p_1^{x_1} p_2^{x_2} p_3^{x_3} p_4^{20-\sum_{j=1}^{3} x_j},$$

if $x_1 + x_2 + x_3 \leq 20$ and 0 otherwise, where x_1, x_2, x_3 are non-negative integers. Write the value of correlation between X_2 & X_3.

Answer: $-\left(\frac{p_2 p_3}{(1-p_2)(1-p_3)}\right)^{\frac{1}{2}} = -\frac{1}{7}$.

Qn 4 (iv) (1 + 1 marks) There are 25 keys in a bunch and these are numbered $1, 2, \cdots, 25$. Except the numbers, the keys are identical. A person is told to open a lock, whose key is known to be there in the bunch. However, exactly one of the keys opens the lock and the person does not know the correct key number. The person tries the keys one by one by choosing one key at random, in each attempt, from one of the unused keys. Unsuccessful keys are not used for future attempts. Let X denote the number of attempts required to open the lock. Identify the distribution of X and write the value of $\mathbb{E}X$.

Answer: $X \sim Uniform(\{1, 2, \cdots, 25\})$ with the p.m.f.

$$f_X(x) = \begin{cases} \frac{1}{25}, & \text{if } x \in \{1, 2, \cdots, 25\} \\ 0, & \text{otherwise.} \end{cases}$$

Then, $\mathbb{E}X = 13$.

Qn 4 (v) (1.5 + 1.5 marks) Let $X \sim Poisson(2)$ and $Y \sim Poisson(3)$ be independent. Write the values of $\mathbb{E}[X \mid X + Y = 3]$ and $Var[X \mid X + Y = 3]$.

Answer: $\mathbb{E}[X \mid X + Y = 3] = \frac{6}{5}$ and $Var[X \mid X + Y = 3] = \frac{18}{25}$ (Hint: $X \sim Binomial(3, \frac{2}{2+3})$)

Qn 4 (vi) (2 + 1 marks) Let $X \sim N(1, 4), Y \sim N(3, 9), Z \sim N(-2, 16)$ be independent RVs and set $W = 2Y - X$. Identify the distribution of W. Given that $\beta \frac{Z+2}{W-5} \sim t_1$ for some $\beta \in \mathbb{R}$. Write the value of $\beta^2 + 1$.

Answer: $W \sim N(5, 40)$, $\beta^2 + 1 = \frac{7}{2}$. (Hint: $\frac{\sqrt{40}}{4} \frac{Z+2}{W-5} \sim t_1$)

Qn 4 (vii) (1 + 1 + 2 marks) Let $\{X_n\}_n$ be a sequence of i.i.d. RVs defined on the same probability space.

(a) If $X_1 \sim Bernoulli(\frac{1}{4})$, then define $Y_n := \frac{1}{\sqrt{n}}\left(X_1 + X_2 + \cdots + X_n - \frac{n}{4}\right), \forall n$. It is known that $Y_n \xrightarrow[n\to\infty]{d} Z$. Identify the distribution of Z.

Answer: $Z \sim N(0, \frac{3}{16})$ (Hint: $Var(X_1) = \frac{3}{16}$)

(b) If $X_1 \sim Beta(3,4)$, then define $Y_n := \frac{X_1+X_2+\cdots+X_n}{n}, \forall n$. It is known that $Y_n \xrightarrow[n\to\infty]{\mathbb{P}} V$. Identify the support of V.

Answer: $\{\frac{3}{7}\}$ (Note: Only $\frac{3}{7}$ is not an acceptable answer, as this is a number and NOT a set. Here, V is a degenerate RV with support $\{\frac{3}{7}\}$)

(c) If $X_1 \sim N(-\sqrt{3}, 4)$, then define $Y_n := \sqrt{n}\left[\left(\frac{X_1+X_2+\cdots+X_n}{n}\right)^2 - 3\right], \forall n$. It is known that $Y_n \xrightarrow[n\to\infty]{d} W$. Write the value of $\mathbb{E}W^2$.

Answer: $\mathbb{E}W^2 = 48$

Hint: by Delta method,

$$\frac{\sqrt{n}}{2}\left[\left(\bar{X}_n\right)^2 - (-\sqrt{3})^2\right] \xrightarrow[n\to\infty]{d} U \sim 2(-\sqrt{3})N(0,1),$$

hence $W \sim 4(-\sqrt{3})N(0,1)$.

Qn 4 (viii) (3 marks) Let X be an RV with the distribution function F_X given by

$$F_X(x) = \begin{cases} 0, \text{ if } x < 0 \\ \alpha\, x + \frac{1}{5}, \text{ if } 0 \le x \le \frac{2}{3}, \\ \beta\, x^2, \text{ if } \frac{2}{3} < x \le \frac{2}{\sqrt{3}}, \\ 1, \text{ if } x > \frac{2}{\sqrt{3}}. \end{cases}$$

for some $\alpha, \beta \in \mathbb{R}$. Find the inter-quartile range. (Answers with α, β in the expression will not be accepted.)

Answer: Hint: Using right-continuity of F_X, we have $\beta = \frac{3}{4}, \alpha = \frac{1}{5}$. The inter-quartile range is $1 - \frac{1}{4} = \frac{3}{4}$.

Qn 4 (ix) (3 marks) Suppose the distribution of X is given by the p.d.f.

$$f_X(x) := \frac{1}{\pi(1+x^2)}, \forall x \in \mathbb{R}.$$

Then which of the following statement(s) is/are necessarily true? One or more than one options may be correct. Write down ALL correct option(s) to get full credit. Not selecting a correct option or selecting an incorrect option is taken as incorrect answer. No partial marking applies.

$(a)\ X \sim t_1 \quad (b)\ X^2 \sim F_{1,1} \quad (c)\ \mathbb{E}|X|^{0.99} < \infty \quad (d)\ \mathbb{E}|X|^{1.01} < \infty$

$(e)\ \mathbb{E}|X|^{2.5} < \infty \quad (f)$ The MGF of X exists. $\quad (g)\ Var(X)$ does not exist.

Answer: $(a), (b), (c), (g)$