MA589
Statistical Foundation for Data Science
IIT GUWAHATI

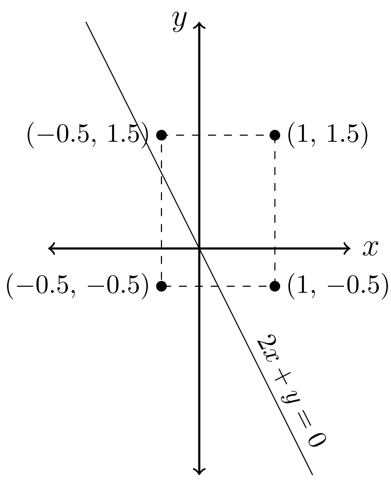
Mid-semester Examination
9:00–11:00 IST
September 17, 2024

Model Answers

1. (2 points) Is the following function is a joint cumulative distribution function of 2-dimensional random vector?

$$G(x,\, y) = \begin{cases} 1 & 2x + y \geq 0 \\ 0 & 2x + y < 0. \end{cases}$$

Solution:



Consider $a_1 = -0.5$, $a_2 = 1$, $b_1 = -0.5$, and $b_2 = 1.5$. Then $a_1 < a_2$ and $b_1 < b_2$. Now,

$$\begin{aligned} & G(a_2,\, b_2) - G(a_2,\, b_1) - G(a_1,\, b_2) + G(a_1,\, b_1) \\ &= G(1,\, 1.5) - G(1,\, -0.5) - G(-0.5,\, 1.5) + G(-0.5,\, -0.5) \\ &= 1 - 1 - 1 + 0 \\ &= -1 < 0 \end{aligned}$$

Therefore, $G(\cdot,\, \cdot)$ is not a joint CDF of 2-dimensional random vector.

2. (4 points) Let $F(\cdot)$ be a cumulative distribution function of a random variable. Let a function $H : \mathbb{R} \to \mathbb{R}$ is defined by

$$H(x) = \alpha F(x) + (1 - \alpha)\left(F(x)\right)^2,$$

where $0 < \alpha < 1$. Then prove that $H(\cdot)$ is a cumulative distribution function of some random variable.

Solution: Let $x_1 < x_2$. Then

$$H(x_2) - H(x_1) = \alpha\left(F(x_2) - F(x_1)\right) + (1 - \alpha)\left((F(x_2))^2 - (F(x_1))^2\right) \geq 0,$$

as $F(x_1) \leq F(x_2)$. Thus, $H(\cdot)$ is non-decreasing.

$$\lim_{x \to \infty} H(x) = 1 \text{ as } \lim_{x \to \infty} F(x) = 1,$$
$$\lim_{x \to -\infty} H(x) = 0 \text{ as } \lim_{x \to -\infty} F(x) = 0.$$

Finally, $H(\cdot)$ is a right continuous function as

$$\lim_{h \to 0+} F(x+h) = F(x) \implies \lim_{h \to 0+} H(x+h) = H(x) \text{ for all } x \in \mathbb{R}.$$

Thus, $H(\cdot)$ is a CDF of a random variable.

3. (4 points) Of the $2n$ people in a given collection of n couples, exactly m die. Assuming that m have been picked at random, find the expected number of surviving couples.

Solution: Let us define for $i = 1, 2, \ldots, n$,

$$X_i = \begin{cases} 1 & \text{if } i\text{-th couple survive} \\ 0 & \text{otherwise.} \end{cases}$$

Then the number of surviving couple $X = \sum_{i=1}^{n} X_i$. Now,

$$P(X_i = 1) = \frac{\binom{2n-2}{m}}{\binom{2n}{m}} \implies E(X_i) = \frac{\binom{2n-2}{m}}{\binom{2n}{m}} \implies E(X) = \sum_{i=1}^{n} E(X_i) = \frac{n\binom{2n-2}{m}}{\binom{2n}{m}}.$$

4. (5 points) The joint probability density function of random variables X and Y is

$$f(x, y) = \begin{cases} e^{-(x+y)} \left[1 + \frac{1}{2}(2e^{-x} - 1)(2e^{-y} - 1)\right] & \text{if } x > 0,\ y > 0 \\ 0 & \text{otherwise.} \end{cases}$$

Find correlation coefficient between X and Y.

Solution:

$$\begin{aligned} E(XY) &= \int_0^\infty \int_0^\infty xye^{-(x+y)} \left[1 + \frac{1}{2}\left(2e^{-x} - 1\right)\left(2e^{-y} - 1\right)\right] dxdy \\ &= \left(\int_0^\infty xe^{-x} dx\right)^2 + \frac{1}{2}\left(\int_0^\infty xe^{-x}\left(2e^{-x} - 1\right) dx\right)^2 \\ &= 1 + \frac{1}{2}\left(\int_0^\infty 2xe^{-2x} dx - \int_0^\infty xe^{-x} dx\right)^2 \\ &= 1 + \frac{1}{2}\left(\frac{1}{2} - 1\right)^2 \\ &= 1 + \frac{1}{8}. \end{aligned}$$

Now, the marginal PDF of X is

$$f_X(x) = \int_{-\infty}^{+\infty} f(x,\, y) dy = \begin{cases} e^{-x} & \text{if } x > 0 \\ 0 & \text{otherwise.} \end{cases}$$

Similarly, the marginal PDF of Y is

$$f_Y(y) = \int_{-\infty}^{+\infty} f(x,\, y) dx = \begin{cases} e^{-y} & \text{if } y > 0 \\ 0 & \text{otherwise.} \end{cases}$$

Therefore, $E(X) = 1 = E(Y)$ and $E(X^2) = 2 = E(Y^2)$. Thus, $Var(X) = Var(Y) = 1$. Hence, the correlation coefficient between X and Y is

$$\rho(X,\, Y) = \frac{E(XY) - E(X)E(Y)}{\sqrt{Var(X)Var(Y)}} = \frac{1}{8}.$$

5. (5 points) Let X be a continuous random variable such that $E(|X|) < \infty$. Then show that $E(|X|^{\frac{1}{2}}) < \infty$.

Solution: Note that for $|x| > 1$, $|x|^{\frac{1}{2}} \leq |x|$ and for $|x| \leq 1$, $|x|^{\frac{1}{2}} \leq 1$.

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

6. (5 points) Let X_1 and X_2 be independent and identically distributed Geometric random variables with success probability $\frac{1}{2}$. Find the probability mass function of $Y = \max\{X_1, X_2\}$. Note that the probability mass function of a Geometric random variable with success probability $p \in (0, 1)$ is

$$f(k) = \begin{cases} p(1-p)^k & \text{if } k = 0, 1, 2, \ldots \\ 0 & \text{otherwise.} \end{cases}$$

Solution: For $k = 0, 1, 2, \ldots$,

$$\begin{aligned}
P\left(Y = k\right) &= P\left(\max\left\{X_1, X_2\right\} = k\right) \\
&= P\left(X_1 = k, X_2 < k\right) + P\left(X_1 < k, X_2 = k\right) + P\left(X_1 = k, X_2 = k\right) \\
&= P\left(X_1 = k\right) P\left(X_2 < k\right) + P\left(X_1 < k\right) P\left(X_2 = k\right) + P\left(X_1 = k, X_2 = k\right) \\
&= 2\left(\frac{1}{2}\right)^{k+1} \sum_{i=0}^{k-1} \left(\frac{1}{2}\right)^{i+1} + \left(\frac{1}{2}\right)^{2k+2} \\
&= \left(\frac{1}{2}\right)^{k} - \left(\frac{1}{2}\right)^{2k} + \left(\frac{1}{2}\right)^{2k+2}.
\end{aligned}$$

Therefore, the probability mass function of the random variable Y is

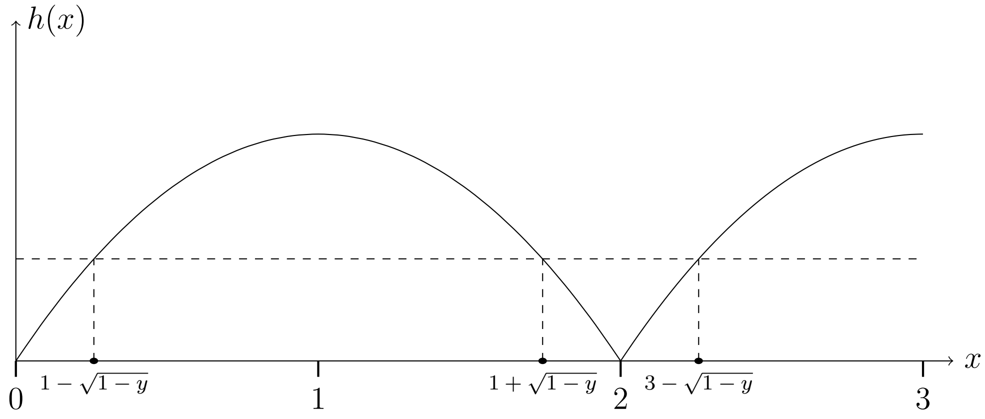
$$f(k) = \begin{cases} \left(\frac{1}{2}\right)^{k} - \left(\frac{1}{2}\right)^{2k} + \left(\frac{1}{2}\right)^{2k+2} & \text{if } k = 0, 1, 2, \ldots \\ 0 & \text{otherwise.} \end{cases}$$

7. (5 points) Let $X \sim U(0, 3)$. Then find the probability density function of $Y = h(X)$, where

$$h(x) = \begin{cases} 1 - (x-1)^2 & \text{if } x \leq 2 \\ 1 - (x-3)^2 & \text{if } x > 2. \end{cases}$$

Solution: The PDF of X is

$$f_X(x) = \begin{cases} \frac{1}{3} & \text{if } 0 < x < 3 \\ 0 & \text{otherwise.} \end{cases}$$



Note that for $x \in (0,\, 3)$, $h(x) \in [0, 1)$. For $y \in (0,\, 1)$, let us find the set $\{x \in (0,\, 3) : h(x) \leq y\}$. Now,

$$\begin{aligned} 1-(x-1)^2 \leq y &\iff x \leq 1-\sqrt{1-y} \text{ or } x \geq 1+\sqrt{1-y}, \\ 1-(x-3)^2 \leq y &\iff x \leq 3-\sqrt{1-y} \text{ or } x \geq 3+\sqrt{1-y}. \end{aligned}$$

Therefore,

$$\begin{aligned} \{x \in (0,\, 3) : h(x) \leq y\} &= (0,\, 1-\sqrt{1-y}] \cup [1+\sqrt{1-y}, 2] \cup [2,\, 3-\sqrt{1-y}] \\ &= (0,\, 1-\sqrt{1-y}] \cup [1+\sqrt{1-y},\, 3-\sqrt{1-y}]. \end{aligned}$$

Thus, for $y \leq 0$, $P(Y \leq y) = 0$ and for $y \geq 1$, $P\left(Y \leq y\right) = 1$. Now, for $0 < y < 1$,

$$\begin{aligned} P\left(Y \leq y\right) &= P\left(h(X) \leq y\right) \\ &= P\left(0 < X \leq 1-\sqrt{1-y}\right) + P\left(1+\sqrt{1-y} \leq X \leq 3-\sqrt{1-y}\right) \\ &= \frac{1}{3}\left(1-\sqrt{1-y}+3-\sqrt{1-y}-1-\sqrt{1-y}\right) \\ &= 1-\sqrt{1-y}. \end{aligned}$$

Therefore, the CDF of Y is

$$F_Y(y) = \begin{cases} 0 & \text{if } y < 0 \\ 1-\sqrt{1-y} & \text{if } 0 \leq y < 1 \\ 1 & \text{if } y \geq 1, \end{cases}$$

and hence, the PDF of Y is

$$f_Y(y) = \begin{cases} \dfrac{1}{2\sqrt{1-y}} & \text{if } 0 < y < 1 \\ 0 & \text{otherwise.} \end{cases}$$