Probability and Statistics: MA6.101

Tutorial 3

Topics Covered: Random Variables, Expectation, Functions on Random Variables, Discrete Random Variables.

Q1: Let X be a discrete random variable with the following PMF:

$$p_X(x) = \begin{cases} 0.25 & \text{for } x = 2, \\ 0.15 & \text{for } x = 4, \\ 0.35 & \text{for } x = 7, \\ 0.25 & \text{for } x = 9, \\ 0 & \text{otherwise.} \end{cases}$$

Find and plot the CDF of X.

A:
The CDF is defined by $F_X(x) = P(X \leq x)$. We have:

$$F_X(x) = \begin{cases} 0 & \text{for } x < 2, \\ p_X(2) = 0.25 & \text{for } 2 \leq x < 4, \\ p_X(2) + p_X(4) = 0.40 & \text{for } 4 \leq x < 7, \\ p_X(2) + p_X(4) + p_X(7) = 0.75 & \text{for } 7 \leq x < 9, \\ 1 & \text{for } x \geq 9. \end{cases}$$

Q2: The median of a random variable X is defined as any number m that satisfies both of the following conditions:

$$P(X \geq m) \geq \frac{1}{2} \quad \text{and} \quad P(X \leq m) \geq \frac{1}{2}.$$

Note that the median of X is not necessarily unique. Find the median of X if

(a) The PMF of X is given by

$$P_X(k) = \begin{cases} 0.4 & \text{for } k = 1, \\ 0.3 & \text{for } k = 2, \\ 0.3 & \text{for } k = 3, \\ 0 & \text{otherwise.} \end{cases}$$

(b) X is the result of rolling a fair die.

(c) $X \sim \text{Geometric}(p)$, where $0 < p < 1$.

A:

$$p_X(X \geq m) \geq \frac{1}{2} \quad p_X(X \leq m) \geq \frac{1}{2}$$

$$p_X(X > m) + p_X(X = m) \geq \frac{1}{2} \quad F_X(m) \geq \frac{1}{2}$$

$$1 - F_X(m) + P_X(X = m) \geq \frac{1}{2}$$

$$\boxed{\frac{1}{2} \leq F_X(m) \leq \frac{1}{2} + P_X(X = m)}$$

(a)

$$F_X(x) = \begin{cases} 0 & x < 1 \\ 0.4 & 1 \leq x < 2 \\ 0.7 & 2 \leq x < 3 \\ 1 & x \geq 3 \end{cases}$$

Case 1: m is not a discrete number:

$$p_X(X = m) = 0 \quad \text{(since } X \text{ is a discrete random variable)}$$

$$F_X(m) = \frac{1}{2} \quad \text{for} \quad m \notin \mathbb{Z}^+$$

Case 2: m is discrete:

$$P_X(X = m) \neq 0$$

$$\frac{1}{2} \leq F_X(m) \leq \frac{1}{2} + P_X(X = m)$$

For $m \geq 2$:

$$m \geq 2 \quad \text{will always satisfy condition (2), but}$$

$$m > 3 \quad \text{will not satisfy it}$$

$$\implies m \text{ might lie in } [2, 3]$$

For non-integer values of m in $[2, 3]$:

$$F_X(m) = \frac{1}{2}, \quad \text{which is not true.}$$

$\therefore$ We only consider discrete points where PMF is non-zero and satisfies (1) and (2).

$$\boxed{m = 2}$$

(b)

$$F_X(x) = \begin{cases} 0 & x < 1 \\ \frac{1}{6} & 1 \leq x < 2 \\ \frac{2}{6} & 2 \leq x < 3 \\ \frac{3}{6} & 3 \leq x < 4 \\ \frac{4}{6} & 4 \leq x < 5 \\ \frac{5}{6} & 5 \leq x < 6 \\ 1 & x \geq 6 \end{cases}$$

For $m \geq 3$, we see that:
$$F_X(x) \geq \frac{1}{2}$$
For $m = 3$, we have:
$$F_X(3) \leq \frac{1}{2} + P_X(3)$$
$$\frac{1}{2} \leq \frac{1}{2} + \frac{1}{6} \quad \text{(satisfied)}$$
For $m = 4$, we have:
$$F_X(4) \leq \frac{1}{2} + P_X(4)$$
$$\frac{4}{6} \leq \frac{1}{2} + \frac{1}{6} = \frac{4}{6} \quad \text{(satisfied)}$$
For $m = 5$, we get:
$$F_X(5) \neq \frac{1}{2} + P_X(5)$$
$\therefore$ Discrete points satisfying the condition are$\{3, 4\}$.

To check the interval $(3, 4)$:
$$F_X(x)\, x \in (3, 4) = \frac{1}{2}$$
$\therefore$ All points that are non-integer in $(3, 4)$ will satisfy the condition.
$$\implies m \in (3, 4) \cup \{3, 4\} = m \in [3, 4]$$

(c) Let $X \sim \text{Geometric}(p)$.
$$P_X(X = k) = (1-p)^{k-1}p \quad k = 1, 2, \ldots$$
$$F_X(x = k) = 1 - (1-p)^k$$
For $F_X(m) \geq \frac{1}{2}$:
$$1 - (1-p)^m \geq \frac{1}{2}$$
$$(1-p)^m \leq \frac{1}{2}$$
$$m \ln(1-p) \leq \ln \frac{1}{2}$$
$$m \geq \frac{-\ln 2}{\ln(1-p)}$$
For $F_X(m) \leq \frac{1}{2} + P_X(m)$:
$$1 - (1-p)^m \leq \frac{1}{2} + p(1-p)^{m-1}$$
$$\frac{1}{2} \leq p + (1-p)^{m-1}$$

$$(1-p)^{m-1} \geq \frac{1}{2}$$

$$(m-1)\ln(1-p) \leq \ln\frac{1}{2}$$

$$m \leq 1 + \frac{-\ln 2}{\ln(1-p)}$$

For $p = \frac{1}{2}$, there will be two discrete points that satisfy this:

$$m = \{1, 2\}$$

For the interval $[1, 2)$:

$$F_X(m) = P_X(1) = (1-p)\frac{1}{2} = \frac{1}{2}$$

$$\implies m \in [1, 2) \cup \{2\}$$

$$\boxed{m \in [1, 2]}$$

For $p \neq \frac{1}{2}$, there will be only one discrete point that satisfies this condition.
The answer might not entirely be the range, but only discrete points (just like in part a).

Q3: Consider a random variable X and another random variable Y defined as a function of X:

$$Y = \begin{cases} 2 & \text{if } X < 2, \\ X & \text{if } X \geq 2. \end{cases}$$

Express Y using the indicator variables of the events $\{X < 2\}$ and $\{X \geq 2\}$.

A: Denote

$$1_{\{X<2\}} \text{ as the indicator variable for the event } \{X < 2\}, \text{ and}$$

$$1_{\{X \geq 2\}} \text{ as the indicator variable for the event } \{X \geq 2\}.$$

Since both of these variables can take values of 0 or 1 for any value of X, and are mutually complementary to each other (i.e., they cannot take the same value at the same time), we can express Y as the sum of a linear combination of these two indicator random variables:

$$Y = 2 \cdot 1_{\{X<2\}} + X \cdot 1_{\{X \geq 2\}}$$

$$\implies Y = 2 \cdot 1_{\{X<2\}} + X \cdot (1 - 1_{\{X<2\}})$$

$$\implies Y = (2 - X) \cdot 1_{\{X<2\}} + X$$

Q4: We roll n dice and sum the highest 3. What is the probability that the sum is 18? Compute formula for general n, and give the value for n=5.

A: The sum will be 18 when there are atleast 3 rolls of 6, since we sum the highest 3 rolls. So the probability will be equal to probability of getting atleast 3 rolls of 6 in n rolls.
Let X be a random variable indicating the number of 6s rolled in n rolls. Then X is a binomial random variable with $p = \frac{1}{6}$
We can calculate the complement i.e. probability of getting less than 3 rolls of 6. The number of 6s is a binomial random variable with $p = \frac{1}{6}$.

$$P(X < 3) = P_X(0) + P_X(1) + P_X(2)$$

where

$$P_X(x) = \binom{n}{x} p^x (1-p)^{n-x}$$

Substituting, we get

$$P(X < 3) = \left(\frac{5}{6}\right)^n + n \cdot \frac{1}{6} \cdot \left(\frac{5}{6}\right)^{n-1} + \frac{n(n-1)}{2} \cdot \left(\frac{1}{6}\right)^2 \cdot \left(\frac{5}{6}\right)^{n-2}$$

And the required probability is the complement of this

Substituting $n = 5$, we get

$$P(X < 3) = 0.94$$

$$P(X >= 3) = 0.06$$

Q5: Two coins are simultaneously tossed until one of them comes up a head and the other a tail. The first coin comes up a head with probability p and the second with probability q. All tosses are assumed independent.

(a) Find the PMF, the expected value, and the variance of the number of tosses.
(b) What is the probability that the last toss of the first coin is a head?

A:

(a) Let X be the number of tosses until the game is over. Noting that X is geometric with the probability of success

$$P(\{HT, TH\}) = p(1-q) + q(1-p),$$

we obtain the probability mass function (PMF):

$$p_X(k) = (1 - p(1-q) - q(1-p))^{k-1} (p(1-q) + q(1-p)), \quad k = 1, 2, \ldots$$

Therefore, the expected value of X is:

$$E[X] = \frac{1}{p(1-q) + q(1-p)}$$

and the variance of X is:

$$\mathrm{Var}(X) = \frac{pq + (1-p)(1-q)}{(p(1-q) + q(1-p))^2}.$$

(b) The probability that the last toss of the first coin is a head is:

$$P(HT \mid \{HT, TH\}) = \frac{p(1-q)}{p(1-q) + (1-p)q}.$$

Q6: Let X be a random variable that takes values from 0 to 9 with equal probability $\frac{1}{10}$.

(a) Find the PMF of the random variable $Y = X \mod 3$.

(b) Find the PMF of the random variable $Y = 5 \mod (X+1)$.

A:

(a) Using the formula

$$p_Y(y) = \sum_{\{x \mid x \mod 3 = y\}} p_X(x),$$

we obtain:

$$p_Y(0) = p_X(0) + p_X(3) + p_X(6) + p_X(9) = \frac{4}{10},$$

$$p_Y(1) = p_X(1) + p_X(4) + p_X(7) = \frac{3}{10},$$

$$p_Y(2) = p_X(2) + p_X(5) + p_X(8) = \frac{3}{10},$$

$$p_Y(y) = 0 \text{ if } y \notin \{0, 1, 2\}.$$

(b) Similarly, using the formula

$$p_Y(y) = \sum_{\{x \mid 5 \mod (x+1) = y\}} p_X(x),$$

we obtain:

$$p_Y(y) = \begin{cases} \frac{2}{10}, & \text{if } y = 0, \\ \frac{2}{10}, & \text{if } y = 1, \\ \frac{1}{10}, & \text{if } y = 2, \\ \frac{5}{10}, & \text{if } y = 5, \\ 0, & \text{otherwise.} \end{cases}$$