Statistics for Data Science-1

Week-9 Graded Assignment

1. A discrete random variable X can take the values $1, 2, 3, \cdots, n$. For these values the cumulative distribution function is defined by:
   $F(x) = P(X \leq x) = \dfrac{x^2 + k}{m}$; $x = 1, 2, 3, \cdots, n$
   Find the value of k.
   Answer: $k = m - n^2$
   Solution:
   $F(n) = P(X \leq n) = 1$
   $\implies \dfrac{n^2 + k}{m} = 1$
   Hence, $k = m - n^2$
   Suppose, we substitute values of n and m as 3 and 40 respectively, then
   $$\frac{3^2 + k}{40} = 1$$
   $$k = 31$$

2. An organization in Texas organizes a lucky draw this month. n thousand tickets are sold for m\$ each. Each has an equal chance of winning. x tickets will win a\$, y tickets will win b\$ and z tickets will win c\$. Let, the random variable X denote the net gain from purchase of one ticket. What is the probability that X takes a value less than b? (Enter the answer correct to 4 decimal place)
   Answer: $\dfrac{n \times 1000 - x}{n \times 1000}$
   Solution:
   X can take values $-m$, $c - m$, $b - m$ and $a - m$
   $P(X < b) = P(X = b - m) + P(X = c - m) + P(X = -m)$
   $P(X < b) = \dfrac{y}{n \times 1000} + \dfrac{z}{n \times 1000} + \dfrac{n \times 1000 - x - y - z}{n \times 1000}$
   $P(X < b) = \dfrac{n \times 1000 - x}{n \times 1000}$

   Suppose, we substitute values of n, m, x, a, y, b, z and c as 5, 1, 1, 1000, 2 500, 10 and 100 respectively, then
   $P(X < 500) = P(X = 499) + P(X = 99) + P(X = -1) = \dfrac{2}{5000} + \dfrac{10}{5000} + \dfrac{4987}{5000}$
   Therefore, $P(X < 500) = \dfrac{4999}{5000} = 0.9998$

3. In a group of n people, x are photographers and $n - x$ are journalists. m people are randomly picked from a group of these n people. Let, Y be a random variable which represents the number of photographers. How many possible values can the random variable Y take?

   Answer: $m + 1$

   Solution:

   Possible values of Y are $0, 1, 2, ..., m$.

   Hence, the number of possible values Y can take is $m + 1$.

   Suppose, we substitute values of m, x and n as 8, 240 and 15 respectively, then possible values of Y are $0, 1, 2, ..., 8$

   Hence, the number of possible values Y can take is 9.

4. Which of the following is/are discrete random variables?

   a. Number of tires produced in an automotive tire factory every 30 minutes.

   b. The number of kernels(pieces) of popcorn in a 1 kg container.

   c. The time between customers entering a checkout lane at a retail store.

   d. The amount of rain recorded at an airport one day.

   e. The amount of liquid in a 2 $litres$ bottle of soft drink.

   f. The number of no-shows for every 1000 reservations made with a commercial airline.

   Answer: a, b, f

   Solution:

   The number of tires produced in an automotive tire factory every 30 minutes can have countable possible values, and hence it denotes a discrete random variable.

   Hence, option (a) is correct.

   The number of kernels of popcorn in a 1 kg container also have countable possible values, it cannot take all values between some interval and hence it is a discrete random variable. So option (b) is correct.

   The time between customers entering a checkout lane at a retail store can take any values between some interval. Hence, it is a continuous random variable.

   So, option (c) is incorrect.

   Again, the amount of rain recorded at an airport one day and the amount of liquid in a 2 $litres$ bottle of soft drink can take any values between some interval. Hence, they are continuous random variable.

   So, option (d) and (e) are incorrect.

The number of no-shows for every 1000 reservations made with a commercial airline can have countable possible values, and hence it denotes a discrete random variable. Hence, option (f) is correct.

5. A biased coin with probability of heads 0.75 is tossed three times. Let X be a random variable that represents the number of head runs, a head run being defined as a consecutive occurrence of at least two heads. Then the probability mass function of X is given by:

a.

$$P(X=x)=\begin{cases}0.375 & \text{for } x=0\\ 0.625 & \text{for } x=1\end{cases}$$

b.

$$P(X=x)=\begin{cases}0.297 & \text{for } x=0\\ 0.703 & \text{for } x=1\end{cases}$$

c.

$$P(X=x)=\begin{cases}0.016 & \text{for } x=0\\ 0.140 & \text{for } x=1\\ 0.422 & \text{for } x=2\\ 0.422 & \text{for } x=3\end{cases}$$

d.

$$P(X=x)=\begin{cases}0.016 & \text{for } x=0\\ 0.844 & \text{for } x=1\\ 0.140 & \text{for } x=2\end{cases}$$

Answer: b

Solution:

Possible outcomes	X	$P(X=x)$
HHH	1	0.422
HHT	1	0.141
HTH	0	0.141
HTT	0	0.047
THH	1	0.141
THT	0	0.047
TTH	0	0.047
TTT	0	0.016

Table 9.1

Hence, the probability mass function of X is given by:

$$P(X = x) = \begin{cases} 0.297 & \text{for } x = 0 \\ 0.703 & \text{for } x = 1 \end{cases}$$

6. Nina has n music sessions in a week. She attends the sessions n days a week $x\%$ of the time, $n-1$ days $y\%$ of the time, one day $z\%$ of the time, and no days $p\%$ of the time. Let, X be a discrete random variable representing the number of sessions she attends in a week. Suppose one week is randomly selected, what is the probability that the random variable X takes the value at most $n-1$?(Enter the answer correct to 2 decimal places)

   Answer: $1 - \dfrac{x}{100}$

   Solution:

   The pmf of random variable X is given by:

$$P(X = k) = \begin{cases} \dfrac{x}{100} & \text{for } k = n \\[2ex] \dfrac{y}{100} & \text{for } k = n-1 \\[2ex] \dfrac{z}{100} & \text{for } k = 1 \\[2ex] \dfrac{p}{100} & \text{for } k = 0 \end{cases}$$

$$\begin{aligned} P(X \le n-1) &= P(X = 0) + P(X = 1) + P(X = n-1) \\ &= \frac{p}{100} + \frac{z}{100} + \frac{y}{100} \\ &= \frac{p+y+z}{100} \\ &= 1 - \frac{x}{100} \end{aligned}$$

   Suppose, we substitute values of n, x, y, z and p as 5, 50, 20, 10 and 20 respectively, then

   The pmf of random variable X is given by:

$$P(X = k) = \begin{cases} 0.5 & \text{for k=5} \\ 0.2 & \text{for k=4} \\ 0.1 & \text{for k=1} \\ 0.2 & \text{for k=0} \end{cases}$$

$$
\begin{aligned}
P(X \leq 4) &= P(X=0) + P(X=1) + P(X=4) \\
&= 0.2 + 0.1 + 0.2 \\
&= 0.5
\end{aligned}
$$

7. Find the value of k for which $k\left(\frac{m}{n}\right)^x$ (x = 0, 1, 2, ...) is a pmf. (Enter the answer correct up to 2 decimal places)

   Answer: $\frac{n-m}{n}$

   Solution:

   For pmf: $k\left[\left(\frac{m}{n}\right)^0 + \left(\frac{m}{n}\right)^1 + \left(\frac{m}{n}\right)^2 + \cdots\right] = 1$

   $\implies k.\frac{1}{1-\frac{m}{n}} = 1$

   $\implies k.\frac{n}{n-m} = 1$

   Therefore, $k = \frac{n-m}{n}$.

   For example:

   Take $m = 3$ and $n = 8$. For pmf: $k\left[\left(\frac{3}{8}\right)^0 + \left(\frac{3}{8}\right)^1 + \left(\frac{3}{8}\right)^2 + \cdots\right] = 1$

   $\implies k.\frac{1}{1-\frac{3}{8}} = 1$

   $\implies k.\frac{8}{5} = 1$

   Therefore, $k = \frac{5}{8}$.

8. Using the information in the previous question, calculate $P(X = 2)$. (Enter the answer correct up to 2 decimal places)

   Answer: $\frac{(n-m)}{n}.\left(\frac{m}{n}\right)^2$

   Solution:

   $P(X = 2) = \frac{(n-m)}{n}.\left(\frac{m}{n}\right)^2$.

   For example:

   Take $m = 3$ and $n = 8$. For pmf: $k\left[\left(\frac{3}{8}\right)^0 + \left(\frac{3}{8}\right)^1 + \left(\frac{3}{8}\right)^2 + \cdots\right] = 1$

$\implies k.\dfrac{1}{1-\dfrac{3}{8}} = 1$

$\implies k.\dfrac{8}{5} = 1$

Therefore, $k = \dfrac{5}{8}$.

And, $P(X = 2) = \dfrac{5}{8}.\left(\dfrac{3}{8}\right)^2 = 0.09$.

9. From a box A containing 3 white and 6 black balls, 5 balls are transferred into an empty box B. Let X be a random variable that represents the number of white balls which are transferred from A to B. What value of random variable will have the least probability?

   Answer: 0

   Solution:

   Let us define the following cases:
   Transfer of 0 white and 5 black balls.
   Transfer of 1 white and 4 black balls.
   Transfer of 2 white and 3 black balls.
   Transfer of 3 white and 2 black balls.
   Probabilities for all cases:

   (i) $P(X = 0) = \dfrac{{}^6C_5}{{}^9C_5} = 0.048$

   (ii) $P(X = 1) = \dfrac{{}^3C_1\,{}^6C_4}{{}^9C_5} = 0.357$

   (iii) $P(X = 2) = \dfrac{{}^3C_2\,{}^6C_3}{{}^9C_5} = 0.476$

   (iv) $P(X = 3) = \dfrac{{}^3C_3\,{}^6C_2}{{}^9C_5} = 0.119$
   .

   Thus, $X = 0$ has the least probability.

10. The probability mass function of a random variable X is given by:

$$P(X = x) = \begin{cases} 3k^2 - 3k & \text{for } x = 0 \\ 2k^2 - 1 & \text{for } x = 1 \\ 0 & \text{otherwise} \end{cases}$$

   Determine the value of k given $k > 0$.

   Answer:1

   Solution:

From properties of pmf,

$$p(0) + p(1) = 1$$
$$3k^2 - 3k + 2k^2 - 1 = 1$$
$$5k^2 - 3k - 2 = 0$$
$$5k^2 - 5k + 2k - 2 = 0$$
$$5k(k-1) + 2(k-1) = 0$$
$$(5k+2)(k-1) = 0$$
$$k = \frac{-2}{5} \text{ or } k = 1$$

As $k > 0$, therefore $k = 1$.