Statistics for Data Science - 2

Week 2 Graded Assignment Solution

Discrete random variable

1. Toss a coin 50 times. Let the random variable X be defined as the number of tails observed. Find the average of the values in the range of the random variable.

   Solution:

   Random variable X is defined as the number of tails observed while tossing the coin 50 times.

   So the possible values taken by X is 0, 1, 2, 3 ....48, 49, 50.

   $\Rightarrow$ Range of $X = \{0, 1, 2, 3....., 48, 49, 50\}$

   Average of range values = sum of all values of range/ total number of values

   $\Rightarrow$ Average of range values $= \frac{0+1+2+3+.....+48+49+50}{51} = \frac{1275}{51} = 25$

2. Suppose that 5 fruits are randomly chosen from a basket containing 20 fruits, of which 16 are good and 4 are rotten. Let Y denote the number of rotten fruits chosen. Find the possible values taken by Y.

   a) {1, 2, 3, 4, 5}

   b) {0, 1, 2, 3, 4, 5}

   c) {1, 2, 3, 4}

   d) {0, 1, 2, 3, 4}

   Solution:

   Random variable Y is defined as the number of rotten fruits chosen from the basket while drawing 5 fruits. Since there are only 4 rotten fruits, so Y cannot take values more than 4. Also there are 16 good fruits, so while drawing fruits there can be 0 rotten fruit or 1 rotten fruit or 2 rotten fruits or 3 rotten fruits or 4 rotten fruits.

   Hence, the possible values taken by Y i.e Range $= \{0, 1, 2, 3, 4\}$.

3. Let X be the number of candies present in a box. We have the following information:

   There are at most four candies in the box.

   The probability of having 2 candies in the box is the same as the probability of having one candy.

   The probability of having no candy in the box is the same as the probability of having 3 candies.

   The probability of having four candies is twice of the probability of having three candies and four times of having two candies.

   What will be the PMF of X?

a)

X	0	1	2	3	4
$P(X=x)$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{2}{10}$	$\frac{2}{10}$	$\frac{4}{10}$

b)

X	0	1	2	3	4
$P(X=x)$	$\frac{2}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{2}{10}$	$\frac{4}{10}$

c)

X	0	1	2	3	4
$P(X=x)$	$\frac{1}{10}$	$\frac{2}{10}$	$\frac{1}{10}$	$\frac{2}{10}$	$\frac{4}{10}$

d)

X	0	1	2	3	4
$P(X=x)$	$\frac{4}{10}$	$\frac{2}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{2}{10}$

Solution:
Given that there are at most four candies in the box, so X cannot take values more than 4.
Also given that
$P(X = 2) = P(X = 1), P(X = 0) = P(X = 3), P(X = 4) = 2P(X = 3)$ and $P(X = 4) = 4P(X = 2)$.
Let $P(X = 2) = p$ and $P(X = 0) = q$
$\Rightarrow 2q = 4p$
$\Rightarrow q = 2p$
And we know that $P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4) = 1$
$\Rightarrow q + p + p + q + 2q = 1$
$\Rightarrow 4q + 2p = 1$
Using the above relation, we will get $4 \times 2p + 2p = 1$
$\Rightarrow p = 1/10$ and hence $q = 2/10$.
So, $P(X = 0) = 2/10, P(X = 1) = 1/10, P(X = 2) = 1/10, P(X = 3) = 2/10$, and $P(X = 4) = 4/10$.
Therefore, option b is the correct answer.

4. Let X be a discrete random variable with following probability mass function

X	0	1	2	3	4	5	6
$P(X=x)$	0	k	$4k$	$6k$	$4k$	$10k^2$	$6k^2$

Table 2.1.G: PMF of X

Find the value of $P(X \leq 4)$. Enter your answer correct up to 4 decimals accuracy.
Solution:
We know that $\sum_{x=0}^{6} P(X = x) = 1$
$\Rightarrow P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4) + P(X = 5) + P(X = 6) = 1$
$\Rightarrow 0 + k + 4k + 6k + 4k + 10k^2 + 6k^2 = 1$

$\Rightarrow 16k^2 + 15k - 1 = 0$
$\Rightarrow (16k - 1)(k + 1) = 0$
$\Rightarrow k = -1$ or $k = 1/16$
Since k cannot take negative values, so k must be $1/16$.
Now,

$$\begin{aligned}
P(X \leq 4) &= P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4) \\
&= 0 + k + 4k + 6k + 4k \\
&= 15k \\
&= 15 \times \frac{1}{16} \\
&= 0.9375
\end{aligned}$$

5. I roll two fair six sided dice and observe the two outcomes. Let the random variables Y and Z denote the outcomes observed on the two dice and let $X = Y + Z$. Find $P(Y = 3|X = 6)$.
   Solution:
   Y and Z denotes the outcomes observed on the two dice.
   Given $X = Y + Z$, so the favourable outcomes for $X = 6$ will be {(1,5),(2,4),(3,3),(4,2),(5,1)}.
   From the reduced sample space the favourable outcomes for $(Y = 3|X = 6)$ will be {(3,3)}.
   Hence, $P(Y = 3|X = 6) = \frac{1}{5} = 0.2$

6. Let X be a discrete random variable with following probability mass function

$$P(X = k) = \begin{cases} 0.2 & \text{for } k = 0 \\ 0.3 & \text{for } k = 1 \\ 0.4 & \text{for } k = 2 \\ 0.1 & \text{for } k = 3 \\ 0 & \text{otherwise.} \end{cases}$$

   Define $Y = (X - 1)(X + 1)(X + 3)$. Find $P(Y \leq 32)$.
   Solution:
   Given that X is taking values 0, 1, 2 and 3 and $Y = (X - 1)(X + 1)(X + 3)$.
   Now we will calculate the values taken by Y corresponding to every value of X.
   At $X = 0$
   $Y = (0 - 1)(0 + 1)(0 + 3) = -3$
   At $X = 1$
   $Y = (1 - 1)(1 + 1)(1 + 3) = 0$
   At $X = 2$
   $Y = (2 - 1)(2 + 1)(2 + 3) = 15$

At $X = 3$
$Y = (3-1)(3+1)(3+3) = 48$
This implies that Y is taking values -3, 0, 15, and 48.
So,

$$\begin{aligned}
P(Y \leq 32) &= P(Y = -3) + P(Y = 0) + P(Y = 15) \\
&= P(X = 0) + P(X = 1) + P(X = 2) \\
&= 0.2 + 0.3 + 0.4 \\
&= 0.9
\end{aligned}$$

7. A shopkeeper sells mobile phones. The demand for mobile phone follows a Poisson distribution with mean 4.6 per week. The shopkeeper has 5 mobile phones in his shop at the beginning of a week. Find the probability that this will not be enough to satisfy the demand for mobile phones in that week. Enter your answer correct up to two decimals accuracy.
Solution:
The shopkeeper has 5 mobile phones in his shop at the beginning of a week. The shopkeeper will not be able to satisfy the demand for mobile phones in that week only if the demand of mobile phone is more than 5 phones. So, we need to find the value of $P(X > 5)$.
Also given that demand for mobile phone follows a Poisson distribution with mean 4.6 per week. i.e. $\lambda = 4.6$

$$\begin{aligned}
P(X > 5) &= 1 - P(X \leq 5) \\
&= 1 - [P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4) + P(X = 5)] \\
&= 1 - \left[\frac{e^{-4.6}(4.6)^0}{0!} + \frac{e^{-4.6}(4.6)^1}{1!} + \frac{e^{-4.6}(4.6)^2}{2!} + \frac{e^{-4.6}(4.6)^3}{3!} + \frac{e^{-4.6}(4.6)^4}{4!} + \frac{e^{-4.6}(4.6)^5}{5!}\right] \\
&= 1 - e^{-4.6}[1 + 4.6 + 10.58 + 16.22 + 18.66 + 17.16] \\
&= 1 - 0.68 \\
&= 0.32
\end{aligned}$$

8. Suppose that in the end semester paper of Statistics there are 18 multiple-choice questions (only one option is correct for each question). Each question has 4 possible options. You know the answer to 8 questions, but you have no idea about the other 10 questions and choose answers randomly and independently. Your score X of the exam is the total number of correct answers. Find the value of $P(X \geq 12)$. Enter your answer correct up to 2 decimals accuracy.
Solution:
Since your score is the total number of correct answers and you know the answer to 8 questions.

So, instead of finding the value of $P(X \geq 12)$, define a new random variable Y and find the value of $P(Y \geq 4)$ from the set of 10 questions for which you do not know the answer.
Also there are four options to each question and only one is correct. That means probability of getting an answer correct is $1/4$ and each question is independent of other.
So we can use binomial distribution with $n = 10$ and $p = 0.25$
Now,

$$\begin{aligned}
P(Y \geq 4) &= 1 - P(Y < 4) \\
&= 1 - [P(Y = 0) + P(Y = 1) + P(Y = 2) + P(Y = 3)] \\
&= 1 - \left[{}^{10}C_0\left(\frac{1}{4}\right)^0\left(\frac{3}{4}\right)^{10} + {}^{10}C_1\left(\frac{1}{4}\right)^1\left(\frac{3}{4}\right)^9 + {}^{10}C_2\left(\frac{1}{4}\right)^2\left(\frac{3}{4}\right)^8 + {}^{10}C_3\left(\frac{1}{4}\right)^3\left(\frac{3}{4}\right)^7\right] \\
&= 1 - \left(\frac{3}{4}\right)^7\left[\left(\frac{3}{4}\right)^3 + 10\left(\frac{1}{4}\right)^1\left(\frac{3}{4}\right)^2 + 45\left(\frac{1}{4}\right)^2\left(\frac{3}{4}\right)^1 + 120\left(\frac{1}{4}\right)^3\right] \\
&= 1 - \left(\frac{3}{4}\right)^7\left[\frac{372}{64}\right] \\
&= 1 - 0.78 \\
&= 0.22
\end{aligned}$$

This implies that $P(X \geq 12) = 0.22$.

9. A fruit owner sells fruit in a lot that contains 50 fruits. A customer selects 5 fruits at random from a lot and rejects the lot (will not purchase) if one of the 5 selected fruits is rotten. What is the probability that the customer will purchase the lot if there are 4 rotten fruits in the lot? Enter your answer correct up to 2 decimals accuracy.
Solution:
Given that there are 4 rotten fruits in the lot that contains 50 fruits.
Customer will purchase the lot if out of 5 selected fruits there is no rotten fruit.
Probability that there will not be any rotten fruit in 5 selected fruits will be
$\dfrac{{}^{4}C_0 \, {}^{46}C_5}{{}^{50}C_5} = \dfrac{1370754}{2118760} = 0.6469$
Accepted range: 0.61 - 0.67

10. Suppose the probability that any given person will independently believe a tale about the existence of a parallel universe is 0.6. What is the probability that the eighth person to hear this tale about existence of a parallel universe is the fifth one to believe it?
    a) ${}^{8}C_5(0.6)^5(0.4)^3$
    b) ${}^{7}C_4(0.6)^5(0.4)^3$
    c) ${}^{8}C_5(0.6)^3(0.4)^5$
    d) ${}^{7}C_4(0.6)^3(0.4)^5$

Solution:
Given that the probability that any given person will believe a tale about the existence of parallel universe is 0.6.
We need to find the probability that the eighth person to hear this tale about existence of parallel universe is the fifth one to believe it.
We can put this into other words as out of 7 trials we need 4 successes and 8th trial also a success.(Here success is considered as the probability that the person will believe the tale about the existence of parallel universe)
Probability of getting 4 successes out of 7 will be ${}^7C_4(0.6)^4(0.4)^3$
Combining that 8th trial also, success will be ${}^7C_4(0.6)^4(0.4)^3 \times 0.6$.
This implies that the probability that the eighth person to hear this tale about existence of parallel universe is the fifth one to believe it is ${}^7C_4(0.6)^5(0.4)^3$.

11. Suppose the number of visitors arriving at a zoo can be modeled to be Poisson distributed. On an average 20 visitors arrive per hour. Let X be the number of visitors arriving from 2pm to 4pm. Then the probability that at least 35 visitors will arrive in the given duration is

a) $\sum_{k=35}^{k=\infty} \frac{e^{-20}(20)^k}{k!}$

b) $1 - \sum_{k=0}^{k=34} \frac{e^{-20}(20)^k}{k!}$

c) $\sum_{k=35}^{k=\infty} \frac{e^{-40}(40)^k}{k!}$

d) $1 - \sum_{k=0}^{k=34} \frac{e^{-40}(40)^k}{k!}$

Solution:
Given that on an average 20 visitors arrive per hour and X is the number of visitors arriving from 2pm to 4pm. So, here $\lambda = 20 \times 2 = 40$
Now we have to find the probability that at least 35 visitors will arrive in the given duration, that is from 2pm to 4pm.

$$\begin{aligned} P(X \geq 35) &= P(X = 35) + P(X = 36) + P(X = 37) + ... \\ &= \sum_{k=35}^{k=\infty} \frac{e^{-\lambda}(\lambda)^k}{k!} \\ &= \sum_{k=35}^{k=\infty} \frac{e^{-40}(40)^k}{k!} \end{aligned}$$

Also we can write

$$
\begin{aligned}
P(X \geq 35) &= 1 - P(X < 35) \\
&= 1 - [P(X = 0) + P(X = 1) + P(X = 2) + ... + P(X = 34)] \\
&= 1 - \sum_{k=0}^{k=34} \frac{e^{-40}(40)^k}{k!}
\end{aligned}
$$