Statistics for Data Science - 2

Week 3 Graded Assignment

Multiple random variables

1. Joint distribution of two random variables X and Y is given as:

$Y \backslash X$	0	1
1	$\frac{1}{4}$	$\frac{1}{8}$
2	$\frac{1}{4}$	k
3	0	$\frac{1}{8}$

Table 3.1.G: Joint distribution of X and Y.

Find the value of $f_{Y|X=1}(2)$. [1 mark]

Solution:

We know that

$$\sum_{x \in T_X, y \in T_Y} f_{XY}(x, y) = 1$$

$$\begin{aligned}
&\Rightarrow \frac{1}{4} + \frac{1}{8} + \frac{1}{4} + k + 0 + \frac{1}{8} = 1 \\
&\Rightarrow k = 1 - \frac{3}{4} = \frac{1}{4}
\end{aligned}$$

Now,

$$\begin{aligned}
f_{Y|X=1}(2) &= \frac{f_{XY}(1, 2)}{f_X(1)} \\
&= \frac{f_{XY}(1, 2)}{f_{XY}(1, 1) + f_{XY}(1, 2) + f_{XY}(1, 3)} \\
&= \frac{\frac{1}{4}}{\frac{1}{8} + \frac{1}{4} + \frac{1}{8}} \\
&= \frac{\frac{1}{4}}{\frac{1}{2}} = \frac{1}{2}
\end{aligned}$$

2. Customers at a fast-food restaurant buy both sandwiches and drinks. The following joint distribution summarizes the numbers of sandwiches (X) and drinks (Y) purchased by customers.

$Y \backslash X$	1	2
1	0.4	0.2
2	0.1	0.25
3	0	0.05

Table 3.2.G: Joint distribution of X and Y.

Find the probability that a customer will buy two sandwiches given that he has bought three drinks. [1 mark]

Solution:

X denotes the number of sandwiches purchased by a customer and Y denotes the number of drinks purchased by a customer.

To find: $f_{X|Y=3}(2)$

Now,

$$\begin{aligned} f_{X|Y=3}(2) &= = \frac{f_{XY}(2,3)}{f_Y(3)} \\ &= \frac{f_{XY}(2,3)}{f_{XY}(1,3) + f_{XY}(2,3)} \\ &= \frac{0.05}{0 + 0.05} \\ &= 1 \end{aligned}$$

---

3. Consider an experiment of tossing a fair coin twice. Let X be the number of heads that occurs in the two tosses and Y be the number of tails that occurs in the two tosses. Choose the correct statements. [2 marks]

   (a) X and Y are independent random variables.

   (b) X and Y are dependent random variables.

   (c) $f_{XY}(1,1) = \frac{1}{2}$.

   (d) $f_{Y|X=0}(1) = \frac{1}{4}$.

Solution:
X denotes the number of heads that occurs in the two tosses and Y denotes the number of tails that occurs in the two tosses.
First we will make the table of the joint pmf of X and Y.

Y \ X	0	1	2
0	0	0	$\frac{1}{4}$
1	0	$\frac{1}{2}$	0
2	$\frac{1}{4}$	0	0

Joint pmf of X and Y.

From the table, we have
$f_X(0) = 0 + 0 + \dfrac{1}{4} = \dfrac{1}{4}$
$f_Y(0) = 0 + 0 + \dfrac{1}{4} = \dfrac{1}{4}$
and
$f_{XY}(0,0) = 0$

It is clear that
$f_{XY}(0,0) \neq f_X(0).f_Y(0)$
It implies that X and Y are dependent random variables.
So, option (a) is incorrect and option (b) is correct.

Now, from table
$f_{XY}(1,1) = \dfrac{1}{2}$
So, option (c) is correct.

$f_{Y|X=0}(1) = \dfrac{f_{XY}(0,1)}{f_X(0)} = 0$ (Since, $f_{XY}(0,1) = 0$)
So, option (d) is incorrect.

---

4. A fair coin is tossed 4 times. Let X be the total number of heads and Y be the number of heads before the first tail (If there is no tail in all the four tosses, then $Y = 4$). What is the value of $f_{Y|X=2}(0)$? [2 marks]

   (a) $\dfrac{5}{16}$

(b) $\dfrac{1}{8}$

(c) $\dfrac{9}{16}$

(d) $\dfrac{1}{2}$

Solution:
A fair coin is tossed four times. X denotes the number of heads and Y denotes the number of heads before first tail (If there is no tail in all the four tosses, then $Y = 4$). Clearly, $X \sim \text{Binomial}(4, \frac{1}{2})$.

Now,

$$\begin{aligned} f_{Y|X=2}(0) &= \frac{f_{XY}(2,0)}{f_X(2)} \\ &= \frac{f_{X|Y=0}(2).f_Y(0)}{f_X(2)} \qquad ..(1) \end{aligned}$$

Now, event $Y = 0$ shows that there is no head before first tail that is first outcome is tail.
It implies that $f_Y(0) = \frac{1}{2}$

$$\begin{aligned} f_{X|Y=0}(2) &= P(\text{two heads in the next three tosses}) \\ &= {}^3C_2 \left(\frac{1}{2}\right)^3 \end{aligned}$$

And
$f_X(2) = {}^4C_2 \left(\frac{1}{2}\right)^4$
Putting the values in the equation (1), we get

$$\begin{aligned} f_{Y|X=2}(0) &= \frac{{}^3C_2 \left(\frac{1}{2}\right)^3 . \frac{1}{2}}{{}^4C_2 \left(\frac{1}{2}\right)^4} \\ &= \frac{3}{6} = \frac{1}{2} \end{aligned}$$

---

5. Two fair dice are thrown simultaneously. Let X be the outcome on the first die and Y be the sum of the outcomes on both the dice. Find the value of $P(Y - X \geq 6)$. [2 marks]

(a) $\frac{1}{6}$

(b) $\frac{1}{12}$

(c) $\frac{5}{12}$

(d) $\frac{1}{24}$

Solution:
X denotes the outcome on the first die and Y denotes the sum of the outcomes on both the dice.
Notice that $Y - X$ will denote the outcome on the second die.
Let $Z = Y - X$, then $Z \sim \text{Uniform}(\{1, 2, 3, 4, 5, 6\})$

$$P(Y - X \geq 6) = P(Z \geq 6)$$
$$P(Y - X \geq 6) = P(Z = 6)$$
$$P(Y - X \geq 6) = \frac{1}{6}$$

---

6. Let X and Y denote the number of cars and number of bikes reaching a street corner during a certain 15-minute time period, respectively. Joint distribution of X and Y is given as

$$f_{XY}(x, y) = \frac{9}{16(4^{x+y})}$$

Choose the correct option(s). [2 marks]

(a) Marginal pmf of X is $f_X(x) = \frac{3}{4^{x+1}}$.

(b) Marginal pmf of X is $f_X(x) = \frac{3}{4^x}$.

(c) X and Y are independent random variables.

(d) X and Y are dependent random variables.

Solution:
X and Y denote the number of cars and number of bikes reaching a street corner during a certain 15-minute time period, respectively.
Range of X and Y will be $T_X, T_Y = \{0, 1, 2, ..., \infty\}$

Joint distribution of X and Y is given as

$$f_{XY}(x, y) = \frac{9}{16(4^{x+y})}$$

Now,

$$
\begin{aligned}
f_X(x) &= \sum_{y=0}^{\infty} f_{XY}(x,y) \\
&= \sum_{y=0}^{\infty} \frac{9}{16(4^{x+y})} \\
&= \frac{9}{16.4^x} \sum_{y=0}^{\infty} \frac{1}{4^y} \\
&= \frac{9}{16.4^x} \left[1 + \frac{1}{4} + \frac{1}{4^2} + ...\right] \\
&= \frac{9}{16.4^x} \left[\frac{1}{1-\frac{1}{4}}\right] \\
&= \frac{9}{4^2.4^x} \left[\frac{4}{3}\right] \\
&= \frac{3}{4^{x+1}}
\end{aligned}
$$

Therefore, option (a) is correct and option (b) is incorrect.

Similarly, we can show that
$f_Y(y) = \dfrac{3}{4^{y+1}}$

Now, Choose two arbitrary points x and y in the range of X and Y, respectively, then

$f_X(x).f_Y(y) = \dfrac{3}{4^{x+1}}.\dfrac{3}{4^{y+1}}$
$\Rightarrow f_X(x).f_Y(y) = \dfrac{9}{16(4^{x+y})}$
$\Rightarrow f_X(x).f_Y(y) = f_{XY}(x,y)$

Hence, X and Y are independent random variables.
Therefore, option (c) is correct and option (d) is incorrect.

---

7. Which of the following option(s) is (are) always correct? [2 marks]

   (a) $f_{XYZ}(x,y,z) = f_{X|(Y=y,Z=z)}(x).f_{YZ}(y,z)$

   (b) $f_{XYZ}(x,y,z) = f_{X|(Y=y,Z=z)}(x).f_X(x)$

   (c) $f_X(x) = \sum_{y \in R_Y} f_{XY}(x,y)$ where R_Y is the range of Y.

(d) $f_{XY}(x,y) = f_X(x).f_Y(y)$

Solution:
We know that $f_{X|(Y=y,Z=z)}(x) = \dfrac{f_{XYZ}(x,y,z)}{f_{YZ}(y,z)}$
$\Rightarrow f_{XYZ}(x,y,z) = f_{X|(Y=y,Z=z)}(x).f_{YZ}(y,z)$
Hence, option (a) is correct and option (b) is incorrect.

We know by the definition of marginal pmf that
$f_X(x) = \sum\limits_{y\in R_Y} f_{XY}(x,y)$ where R_Y is the range of Y.

Hence, option (c) is correct.

$f_{XY}(x,y) = f_X(x).f_Y(y)$ is true only when X and Y are independent. Therefore, option (d) need not to be always true.

---

8. Two random variables X and Y are jointly distributed with joint pmf

$$f_{XY}(x,y) = a(bx+y)$$

, where x and y are integers in $0 \le x \le 2$ and $0 \le y \le 3$ such that $P(X \ge 1, Y \le 2) = \frac{4}{7}$. Find the value of $f_{XY}(2,1)$. [2 marks]

1. $\dfrac{1}{21}$
2. $\frac{5}{42}$
3. $\frac{1}{42}$
4. $\frac{9}{42}$

Solution: We know that

$$\sum_{x\in T_X, y\in T_Y} f_{XY}(x,y) = 1$$

$$\Rightarrow f_{XY}(0,0) + f_{XY}(0,1) + f_{XY}(0,2) + f_{XY}(0,3) + f_{XY}(1,0) + f_{XY}(1,1) + f_{XY}(1,2) + f_{XY}(1,3) + f_{XY}(2,0) + f_{XY}(2,1) + f_{XY}(2,2) + f_{XY}(2,3) = 1$$

$$\Rightarrow a + 2a + 3a + ab + (ab+a) + (ab+2a) + (ab+3a) + (2ab) + (2ab+a) + (2ab+2a) + (2ab+3a) = 1$$
$$\Rightarrow 18a + 12ab = 1 \qquad ...(1)$$

Now, using the given condition,

$$P(X \geq 1, Y \leq 2) = \frac{4}{7}$$

$$\Rightarrow P(X=1,Y=0) + P(X=1,Y=1) + P(X=1,Y=2) + P(X=2,Y=0) + P(X=2,Y=1) + P(X=2,Y=2) = \frac{4}{7}$$

$$\Rightarrow ab + ab + a + ab + 2a + 2ab + 2ab + a + 2ab + 2a = \frac{4}{7}$$

$$\Rightarrow 6a + 9ab = \frac{4}{7} \qquad ....(2)$$

Solving equation (1) and (2), we get
$ab = \frac{1}{21}$ and $a = \frac{1}{42}$
It implies that
$a = \frac{1}{42}$ and $b = 2$
Therefore, the joint pmf of X and Y will be

$$f_{XY}(x,y) = \frac{1}{42}(2x+y)$$

Now, $f_{XY}(2,1) = \frac{1}{42}(4+1) = \frac{5}{42}$.

---

9. Let X_1, X_2, X_3 and X_4 be four independent and identically distributed Poisson random variables with $\lambda_i = 4$ for all i. Find the probability that exactly one of the X_i equals 0 and exactly one of the X_i equals 1? [3 marks]

   (a) $24e^{-8}(1 - 25e^{-8})$
   (b) $24e^{-8}(1 - 5e^{-4})^2$
   (c) $48e^{-8}(1 - e^{-8})$
   (d) $48e^{-8}(1 - 5e^{-4})^2$

Solution:
First we will find the probability such that $X_1 = 0$, $X_2 = 1$ and other two random variables do not take value 0 and 1.

Since all four random variable are independent, we have

$$P(X_1 = 0, X_2 = 1, X_3 \neq \{0,1\}), X_4 \neq \{0,1\}) = P(X_1 = 0).P(X_2 = 1).P(X_3 \neq \{0,1\})P(X_4 \neq \{0,1\})....(1)$$

Now,

$$P(X_1 = 0) = \frac{e^{-4}4^0}{0!} = e^{-4}$$

$$P(X_2 = 1) = \frac{e^{-4}4^1}{1!} = 4e^{-4}$$

$$\begin{aligned} P(X_3 \neq \{0,1\}) &= 1 - P(X_3 = \{0,1\}) \\ &= 1 - [P(X_3 = 0) + P(X_3 = 1)] \\ &= 1 - [e^{-4} + 4e^{-4}] = 1 - 5e^{-4} \end{aligned}$$

$$\begin{aligned} P(X_4 \neq \{0,1\}) &= 1 - P(X_4 = \{0,1\}) \\ . \quad &= 1 - [P(X_4 = 0) + P(X_4 = 1)] \\ . \quad &= 1 - [e^{-4} + 4e^{-4}] = 1 - 5e^{-4} \end{aligned}$$

Putting all these values in equation (1), we get

$$P(X_1 = 0, X_2 = 1, X_3 \neq \{0,1\}), X_4 \neq \{0,1\}) = e^{-4}(4e^{-4})(1 - 5e^{-4})^2$$

We can choose such pairs of X_i for which exactly one X_i equals 0 and exactly one X_i equals 1 in ${}^4P_2$ ways.
Therefore,
probability that exactly one of the X_i equals 0 and exactly one of the X_i equals 1 is given by

$$\begin{aligned} &{}^4P_2 e^{-4}(4e^{-4})(1 - 5e^{-4})^2 \\ &= 48e^{-8}(1 - 5e^{-4})^2 \end{aligned}$$

---

10. Akshat draws a card randomly from a well-shuffled pack of 52 cards. If the drawn card is a face card, then he draws two balls randomly from bag A which contains 5 Red, 6 Black and 4 Green balls. If the drawn card is not a face card, then he draws three balls randomly from bag B which contains 7 Red, 8 Black and 5 Green balls. Let two random variables X and Y are defined as:

$$X = \begin{cases} 0 & \text{if the drawn card is a face card} \\ 1 & \text{if the drawn card is not a face card} \end{cases}$$

and Y be the number of Red balls drawn. Find the value of $f_Y(1)$. Write your answer correct up to two decimal places. [3 marks]

Solution:

Akshat draws a card randomly from a well-shuffled pack of 52 cards. Random variable X is defined as

$$X = \begin{cases} 0 & \text{if the drawn card is a face card} \\ 1 & \text{if the drawn card is not a face card} \end{cases}$$

If the drawn card is a face card, then he draws two balls randomly from bag A which contains 5 Red, 6 Black and 4 Green balls. If the drawn card is not a face card, then he draws three balls randomly from bag B which contains 7 Red, 8 Black and 5 Green balls. Random variable Y is the number of Red balls drawn.
To find: $f_Y(1)$

We know that

$$\begin{aligned} f_Y(1) &= f_{XY}(0,1) + f_{XY}(1,1) \\ &= f_{Y|X=0}(1).f_X(0) + f_{Y|X=1}(1).f_X(1) \\ &= \frac{{}^5C_1{}^{10}C_1}{{}^{15}C_2}.\frac{12}{52} + \frac{{}^7C_1{}^{13}C_2}{{}^{20}C_3}.\frac{40}{52} \\ &= 0.109 + 0.368 = 0.47 \end{aligned}$$
