Statistics for Data Science - 2

Week 1 Graded Assignment

Multiple random variables

1. Three fair coins are tossed. If the first head occurs on the first toss, you score 1 point. If the first head occurs on toss 2 or on toss 3, you score 2 or 3 points, respectively. If no heads appear, you lose 1 point (that is score -1 point). Let X denote the number of heads and Y denote the points scored. What is the probability that fewer than three heads will occur and you will score 1 or less? Write your answer correct to two decimal places.

   Solution:
   Given that X denotes the number of heads and Y denotes the point scored.
   Clearly, $T_X = \{0, 1, 2, 3\}$ and $T_Y = \{-1, 1, 2, 3\}$.
   To find: $P(X < 3, Y \leq 1)$.

Outcome	X	Y
HHH	3	1
HHT	2	1
HTH	2	1
THH	2	2
HTT	1	1
THT	1	2
TTH	1	3
TTT	0	-1

   The outcomes HHT, HTH, HTT, TTT correspond to the event $(X < 3, Y \leq 1)$.
   Therefore,

$$
\begin{aligned}
P(X < 3, Y \leq 1) &= P(\{\text{HHT, HTH, HTT, TTT}\}) \\
&= \frac{4}{8} = \frac{1}{2}
\end{aligned}
$$

2. Contracts for two construction jobs are each assigned uniformly at random to one or more of three firms, A, B, and C. Let X denote the number of contracts assigned to firm A and Y the number of contracts assigned to firm B. Find the value of $f_{X|Y=0}(2)$. Write your answer correct to two decimal places.

Solution:

Given that X denotes the number of contracts assigned to firm A and Y denotes the number of contracts assigned to firm B.

Since each job is randomly assigned to one or more of the three firms, probability of assigning one job to any of the three firms is $\frac{1}{3}$. (Notice that one firm can be assigned either 0 or 1 or 2 jobs).

Clearly, $T_X = T_Y = \{0, 1, 2\}$ Therefore,

$$\begin{aligned} P(X = 2, Y = 0) &= P(\text{Both the jobs are assigned to firm A}) \\ &= \frac{1}{3}.\frac{1}{3} = \frac{1}{9} \end{aligned}$$

Similarly,

$$\begin{aligned} P(X = 1, Y = 0) &= P(\text{one job is assigned to firm A and no job is assigned to firm B}) \\ &= P(\text{one job is assigned to firm A and other job is assigned to firm C}) \\ &= 2\left(\frac{1}{3}.\frac{1}{3}\right) = \frac{2}{9} \end{aligned}$$

and

$$\begin{aligned} P(X = 0, Y = 0) &= P(\text{Both the jobs are assigned to firm C}) \\ &= \frac{1}{3}.\frac{1}{3} = \frac{1}{9} \end{aligned}$$

Therefore,

$$\begin{aligned} P(Y = 0) &= P(X = 0, Y = 0) + P(X = 1, Y = 0) + P(X = 2, Y = 0) \\ &= \frac{1}{9} + \frac{2}{9} + \frac{1}{9} \\ &= \frac{4}{9} \end{aligned}$$

Now,

$$\begin{aligned} f_{X|Y=0}(2) &= \frac{P(X = 2, Y = 0)}{P(Y = 0)} \\ &= \frac{1/9}{4/9} \\ &= \frac{1}{4} \end{aligned}$$

3. 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 1.1.G: Joint distribution of X and Y.

Find the value of $f_{Y|X=1}(2)$.
Solution:
We know that

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

$$\Rightarrow \frac{1}{4} + \frac{1}{8} + \frac{1}{4} + k + 0 + \frac{1}{8} = 1$$
$$\Rightarrow k = 1 - \frac{3}{4} = \frac{1}{4}$$

Now,

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

---

4. 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 \ X	1	2
1	0.4	0.2
2	0.1	0.25
3	0	0.05

Table 1.2.G: Joint distribution of X and Y.

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

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,

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

---

5. 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 k Red, $k+1$ Black and $14-2k$ Green balls. If the drawn card is not a face card, then he draws three balls randomly from bag B which contains t Red, $t+2$ Black and $18-2t$ 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.

Parameters: $k = 4, 5, 6$

$t = 6, 7, 8$

Answer: $\dfrac{k(15-k)}{455}+\dfrac{t(20-t)(19-t)}{2964}$

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 k Red, $k+1$ Black and $14-2k$ Green balls. If the drawn card is not a face card, then he draws three balls randomly from bag B which contains t Red, $t+2$ Black and $18-2t$ Green balls. Random variable Y is the number of Red balls drawn.

To find: $f_Y(1)$

We know that

$$\begin{aligned}f_Y(1)&=f_{XY}(0,1)+f_{XY}(1,1)\\ &=f_{Y|X=0}(1).f_X(0)+f_{Y|X=1}(1).f_X(1)\\ &=\frac{{}^kC_1\,{}^{15-k}C_1}{{}^{15}C_2}.\frac{12}{52}+\frac{{}^tC_1\,{}^{20-t}C_2}{{}^{20}C_3}.\frac{40}{52}\\ &=\frac{k(15-k)}{455}+\frac{t(20-t)(19-t)}{2964}\end{aligned}$$

---

6. Which of the following options is/are always correct?

   (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) $f_X(x)=\sum\limits_{y\in R_Y} f_{XY}(x,y)$ 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.

---

7. 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 $f_{XY}(x,y)$. [2 marks]

Parameters: integers in $0 \le x \le 2$ and $0 \le y \le 3$

Answer: $\frac{1}{42}(2x+y)$

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,

$$P(X \ge 1, Y \le 2) = \frac{4}{7}$$
$$\Rightarrow P(X=1,Y=0) + P(X=1,Y=1) + P(X=1,Y=2) + P(X=2,Y=0) +$$
$$P(X=2,Y=1) + P(X=2,Y=2) = \frac{4}{7}$$
$$\Rightarrow ab + ab + a + ab + 2a + 2ab + 2ab + a + 2ab + 2a = \frac{4}{7}$$
$$\Rightarrow 6a + 9ab = \frac{4}{7} \qquad ....(2)$$

Solving equation (1) and (2), we get

$ab = \dfrac{1}{21}$ and $a = \dfrac{1}{42}$

It implies that

$a = \dfrac{1}{42}$ and $b = 2$

Therefore, the joint pmf of X and Y will be

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

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

8. 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) $\frac{5}{16}$

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

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

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

Parameter: $k = 4, 5, 6, 7, 8$

Answer: $\frac{k-2}{k}$

Solution:

A fair coin is tossed k 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 k tosses, then $Y = k$). Clearly, $X \sim \text{Binomial}(k, \frac{1}{2})$.

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 $f_Y(0) = \frac{1}{2}$

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

And

$f_X(2) = {}^{k}C_2 \left(\frac{1}{2}\right)^{k}$

Putting the values in the equation (1), we get

$$f_{Y|X=2}(0) = \frac{{}^{k-1}C_2 \left(\frac{1}{2}\right)^k}{{}^{k}C_2 \left(\frac{1}{2}\right)^k}$$
$$= \frac{k-2}{k}$$

---

9. From a group of k members of party A, $k-1$ members of party B, and $k-2$ member of party C, a committee of two people is to be selected uniformly at random. Let X denote the number of party A members and Y denote the number of party B members on the committee. Find the value of $f_{XY}(1,1)$.
Parameter: $k = 3, 4, 5, 6, 7, 8$
Answer: $\dfrac{2k}{9k-12}$
Solution:
Given that X denotes the number of party A members in selected two member's committee and Y denotes the number of party B members in selected two member's committee.

To find: $f_{XY}(1,1)$

$$f_{XY}(1,1) = P(X=1, Y=1)$$
$$= \frac{{}^{3}C_1 . {}^{2}C_1}{{}^{6}C_2}$$
$$= \frac{3 \times 2}{15}$$
$$= 0.4$$

---

10. A fair coin is tossed k times, and the number of heads, N, is counted. The coin is then tossed N more times. Find the probability that heads will appear for a total of four times in this process. Write your answer correct to two decimal places.
Parameter:$k = 4, 5, 6, 7, 8$
Answer: $\dfrac{1}{2^{k+2}} \left[\dfrac{k(k-1)}{2} + \dfrac{k(k-1)(k-2)}{4} + \dfrac{k(k-1)(k-2)(k-3)}{96} \right]$
Solution:
Given that N denotes the number of heads in five tosses of a coin.
Clearly, $N \sim \text{Binomial}(k, 1/2)$.

Let X denotes the number of heads in N tosses.
Then, $X|(N=n) \sim \text{Binomial}(n, 1/2)$

Heads will appear a total of four times if $(N=2, X=2), (N=3, X=1), (N=4, X=0)$.

It implies that

$$
\begin{aligned}
P(\text{Total four heads will appear}) &= P(N=2, X=2) + P(N=3, X=1) + P(N=4, X=0) \\
&= P(N=2).P(X=2|N=2) + P(N=3).P(X=1|N=3) \\
&\qquad\qquad\qquad + P(N=4).P(X=0|N=4) \\
&= \left({}^5C_2\frac{1}{2^5}\right)\left({}^2C_2\frac{1}{2^2}\right) + \left({}^5C_3\frac{1}{2^5}\right)\left({}^3C_1\frac{1}{2^3}\right) + \left({}^5C_4\frac{1}{2^5}\right)\left({}^4C_0\frac{1}{2^4}\right) \\
&= \frac{10}{2^7} + \frac{30}{2^8} + \frac{5}{2^9} \\
&= \frac{1}{2^7}[10 + 15 + 1.25] \\
&= 0.20
\end{aligned}
$$