Statistics for Data Science - 2

Week 4 Practice Assignment

Expectation and variance

1. If the expected value and variance of the Binomial random variable X are $\frac{5}{2}$ and $\frac{15}{8}$, respectively, then find the value of $P(X = 10)$. [1 mark]

   (a) $\left(\frac{3}{4}\right)^{10}$

   (b) $10\left(\frac{3}{4}\right)^{10}$

   (c) $\left(\frac{1}{4}\right)^{10}$

   (d) $10\left(\frac{1}{4}\right)^{10}$

   Solution: If $X \sim \text{Binomial}(n, p)$, then expected value and variance of X is given by np and $np(1-p)$, respectively.

   Given that
   $E[X] = np = \frac{5}{2}$ ...(1)
   And
   $\text{Var}(X) = np(1-p) = \frac{15}{8}$ ..(2)
   Putting the value of np in the equation (2) from equation (1), we get
   $(1-p) = \frac{3}{4} \Rightarrow p = \frac{1}{4}$.
   Putting the value of p in equation (1), we get
   $n = 10$
   It implies that $X \sim \text{Binomial}\left(10, \frac{1}{4}\right)$
   Therefore,
   $P(X = 10) = {}^{10}C_{10}\left(\frac{1}{4}\right)^{10}\left(\frac{3}{4}\right)^{0} = \left(\frac{1}{4}\right)^{10}$

---

2. X and Y are two independent geometric random variables with parameters $\frac{1}{2}$ and $\frac{1}{4}$, respectively. Find the value of $\text{Var}(X + 2Y)$. [1 mark]
   Solution:
   We know that if $X \sim \text{Geometric}(p)$, then $\text{Var}(X) = \frac{1-p}{p^2}$

Therefore, $\text{Var}(X) = \dfrac{1-\frac{1}{2}}{\frac{1}{4}} = 2$ ...(1)

$\text{Var}(Y) = \dfrac{1-\frac{1}{4}}{\frac{1}{16}} = 12$ ...(2)

Now, since X and Y are independent, we have

$$\text{Var}(X+2Y) = \text{Var}(X) + 2^2\text{Var}(Y)$$
$$2 + 48 = 50$$

---

3. The number of spam messages (X) sent to a server in a day has Poisson distribution with parameter $\lambda = 21$. Each spam message independently has a probability of $p = \frac{1}{3}$ of not being detected by the spam filter. Let Y denote the number of spam messages detected by the filter in a day. Calculate the expected value of $X + Y$. [2 marks]

   solution:
   X denotes the number of spam messages sent to the server in a day and

   $$X \sim \text{Poisson}(21)$$

   Y denotes the number of spam messages detected by the filter in a day.
   It is given that each spam messages independently has a probability of $\dfrac{1}{3}$ of not being detected. It implies that

   $$Y|X \sim \text{Binomial}(X, \frac{2}{3})$$

   Recall that if $N \sim \text{Poisson}(\lambda)$ and $Z|N \sim \text{Binomial}(N, p)$, then $Z \sim \text{Poisson}(\lambda p)$.

   Therefore, $Y \sim \text{Poisson}(14)$

   $E[X] = 21$ and $E[Y] = 14$
   $\Rightarrow E[X+Y] = E[X] + E[Y] = 35$

---

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

   $$f_{XY}(x,y) = \frac{1}{9}(x+y),$$

   where x and y are integers in $0 \leq x \leq 2$ and $0 \leq y \leq 1$. Let $Z = XY + Y^2$. Find the expected value of Z. [2 marks]

   (a) $\dfrac{1}{3}$

(b) $\frac{4}{3}$

(c) $\frac{2}{3}$

(d) $\frac{14}{9}$

Solution:

$$\begin{aligned}E[Z] &= E[XY + Y^2] \\ &= \sum_{0\le x\le 2; 0\le y\le 1} (xy + y^2) f_{XY}(x, y) \\ &= \frac{1}{9} \sum_{0\le x\le 2; 0\le y\le 1} (xy + y^2)(x + y) \\ &= \frac{1}{9}(1 + 4 + 9) \\ &= \frac{14}{9}\end{aligned}$$

---

5. The distribution of a certain company's employees' monthly salary has mean ₹60000 and standard deviation ₹20000. The probability that a randomly selected employee from that company has a salary either greater than or equal to ₹100000 or less than or equal to ₹20000 is: [2 marks]

(a) at least $\frac{1}{4}$

(b) at most $\frac{1}{4}$

(c) at least $\frac{1}{2}$

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

Solution:
Let X denote the employees' monthly salary.
Given that $E[X] = \mu = 60000$ and SD$= \sigma = 20000$.

$$\begin{aligned}P(X \ge 100000 \text{ or } X \le 20000) &= P(X - 60000 \ge 40000 \text{ or } X - 60000 \le -40000) \\ &= P(|X - 60000| \ge 40000) \\ &= P(|X - \mu| \ge 2\sigma)\end{aligned}$$

By using Chebyshev's inequality

$$\le \frac{1}{4}$$

Hence, probability that a randomly selected employee from that company has a salary either greater than or equal to ₹100000 or less than or equal to ₹20000 is at most $\frac{1}{4}$.

---

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

$$f_{XY}(x,y) = \frac{1}{27}(xy + x + y + 1),$$

where x and y are integers in $0 \leq x \leq 1$ and $1 \leq y \leq 3$. Find the correlation coefficient of X and Y. [2 marks]

Solution:

$$\begin{aligned}
E[X] &= \sum_{x\in T_X, y\in Y_Y} x f_{XY}(x,y)\\
&= \frac{1}{27}\sum_{x\in T_X, y\in Y_Y} x(xy + x + y + 1)\\
&= \frac{1}{27}(4 + 6 + 8)\\
&= \frac{18}{27} = \frac{2}{3}
\end{aligned}$$

$$\begin{aligned}
E[Y] &= \sum_{x\in T_X, y\in Y_Y} y f_{XY}(x,y)\\
&= \frac{1}{27}\sum_{x\in T_X, y\in Y_Y} y(xy + x + y + 1)\\
&= \frac{1}{27}(2 + 6 + 12 + 4 + 12 + 24)\\
&= \frac{60}{27} = \frac{20}{9}
\end{aligned}$$

$$\begin{aligned}
E[XY] &= \sum_{x\in T_X, y\in Y_Y} xy f_{XY}(x,y)\\
&= \frac{1}{27}\sum_{x\in T_X, y\in Y_Y} xy(xy + x + y + 1)\\
&= \frac{1}{27}(4 + 12 + 24)\\
&= \frac{40}{27}
\end{aligned}$$

$$
\begin{aligned}
\mathrm{Cov}(X,Y) &= E[XY]-E[X]E[Y]\\
&= \frac{40}{27}-\frac{2}{3}.\frac{20}{9}\\
&= 0
\end{aligned}
$$

We know that

$$\text{Correlation coefficient} = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}} = 0$$

---

7. Let X and Y be two independent random variables such that $X \sim \text{Binomial}(4, \frac{1}{2})$ and $Y \sim \text{Uniform}(\{1,2,3\})$. Find the value of $\mathrm{Cov}(2X+Y,\, X+Y^2X)$. [2marks]

   (a) 16.67
   (b) 6.67
   (c) 13.37
   (d) 0

   Solution:

$$
\begin{aligned}
\mathrm{Cov}(2X+Y, X+Y^2X) &= \mathrm{Cov}(2X, X+Y^2X) + \mathrm{Cov}(Y, X+Y^2X)\\
&= \mathrm{Cov}(2X,X)+\mathrm{Cov}(2X,Y^2X)+\mathrm{Cov}(Y,X)+\mathrm{Cov}(Y,Y^2X)\\
&= 2\mathrm{Cov}(X,X)+2\mathrm{Cov}(X,Y^2X)+\mathrm{Cov}(Y,X)+\mathrm{Cov}(Y,Y^2X)\\
&= 2\mathrm{Var}(X)+2(E[X^2Y^2]-E[X]E[Y^2X])+(E[XY]-E[X]E[Y])\\
&\qquad + (E[XY^3]-E[Y]E[Y^2X])
\end{aligned}
$$

Since X and Y are independent random variables, $(X^2,\, Y^2)$, $(X,\, Y^2)$, $(X,\, Y^3)$ are also independent. It implies that
$E[X^2Y^2] = E[X^2]E[Y^2]$
$E[Y^2X] = E[Y^2]E[X]$
$E[XY^3] = E[X]E[Y^3]$

Therefore,

$$
\begin{aligned}
\mathrm{Cov}(2X+Y, X+Y^2X) &= 2\mathrm{Var}(X)+2(E[X^2]E[Y^2]-E[X]^2E[Y^2])+(E[XY]-E[X]E[Y])\\
&\qquad + (E[X]E[Y^3]-E[Y]E[Y^2]E[X])\\
&= 2\mathrm{Var}(X)+2(E[X^2]E[Y^2]-E[X]^2E[Y^2])+E[X]E[Y^3]-E[Y]E[Y^2]E[X]
\end{aligned}
$$

Now, $X \sim \text{Binomial}(4, \frac{1}{2})$
Therefore, $E[X] = np = 2$
$\text{Var}(X) = np(1-p) = 1$
$E[X^2] = \text{Var}(X) + (E[X])^2 = np(1-p) + (np)^2 = 1 + 4 = 5$

And $Y \sim \text{Uniform}(\{1, 2, 3\})$
$E[Y] = \frac{1}{3}(1 + 2 + 3) = 2$
$E[Y^2] = \frac{1}{3}(1 + 4 + 9) = \frac{14}{3}$
$E[Y^3] = \frac{1}{3}(1 + 8 + 27) = 12$

Therefore,

$$\begin{aligned}\text{Cov}(2X + Y, X + Y^2X) &= 2(1) + 2(\frac{70}{3} - \frac{56}{3}) + 24 - \frac{56}{3} \\ &= 26 - \frac{28}{3} \\ &= 16.67\end{aligned}$$

---

8. The joint distribution of two random variables X and Y is given as:

Y \ X	0	1	2
-1	0	$\frac{2}{17}$	$\frac{5}{17}$
0	$\frac{1}{17}$	$\frac{2}{17}$	0
1	$\frac{3}{17}$	0	$\frac{4}{17}$

Table 4.1.P: Joint distribution of X and Y.

Find the standard deviation of the product of the two random variables. (Write your answer correct up to two decimal points.) [2 marks]
Solution:
To find: $\text{SD}(XY)$

$$
\begin{aligned}
E[XY] &= \sum_{x\in T_X, y\in T_Y} xyf_{XY}(x,y)\\
&= -1(\frac{2}{17}) - 2(\frac{5}{17}) + 2(\frac{4}{17})\\
&= \frac{-4}{17}
\end{aligned}
$$

$$
\begin{aligned}
E[(XY)^2] = &= \sum_{x\in T_X, y\in T_Y} x^2y^2f_{XY}(x,y)\\
&= 1(\frac{2}{17}) + 4(\frac{5}{17}) + 4(\frac{4}{17})\\
&= \frac{38}{17}
\end{aligned}
$$

$$
\begin{aligned}
\text{Var}(XY) &= E[(XY)^2] - [E[XY]]^2\\
&= \frac{38}{17} - \frac{16}{289}\\
&= \frac{630}{289}
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
\text{SD}(XY) &= \sqrt{\text{Var}(XY)}\\
&= \sqrt{\frac{630}{289}} = 1.47
\end{aligned}
$$

---

9. An ice-cream seller sells ice creams at three prices: ₹30, ₹40, and ₹50. A random customer will buy an ice cream of ₹30, ₹40 and ₹50 with probabilities of 0.5, 0.3, and 0.2, respectively. If the number of customers in a day follows Poisson distribution with $\lambda = 60$, what is the expected sales (in ₹) of the seller in a day? [3 marks]

Solution:

Let X denote the number of customers coming to the ice-cream seller in a day, then

$$X \sim \text{Poisson}(60)$$

Let Y denote the price at which the customer buys the ice-cream, then
$E[Y] = 30(0.5) + 40(0.3) + 50(0.2) = 37$

If $X = x$ customers comes at the shop, then expected sale will be $xE[Y]$

But since $X \sim \text{Poisson}(60)$, on an average 60 customers come to the ice-cream seller in a day. It means that expected sale of the day will be

$$60E[Y] = 60(37) = 2220$$

---

10. An urn contains 10 balls numbered from 1 to 10. We remove six balls randomly and add up their numbers. Let X denote the sum of the numbers of the removed balls. Find the expected value of X. [3 marks]

(Hint: Suppose X_i denotes the number of the ith removed ball, then $X = \sum_{i=1}^{6} X_i$)

Solution:
Let X_i, $i = 1, 2, ...6$ denote the number on the ith ball, then

$X = \sum_{i=1}^{6} (X_i)$

$\Rightarrow E[X] = E\left[\sum_{i=1}^{6} (X_i)\right]$

$\Rightarrow E[X] = \left[\sum_{i=1}^{6} E(X_i)\right]$

$\Rightarrow E[X] = 6E(X_i)$ ...(1)

Now, $E[X_i] = \dfrac{1}{10}[1 + 2 + 3 + ...10] = \dfrac{11}{2}$

Putting the value in equation (1), we get

$E[X] = 6 \times \dfrac{11}{2} = 33$
