Statistics for Data Science - 2

Week 4 Graded assignment solutions

1. Suppose 1 in 100 products that are coming out of a production line is defective. Suppose we randomly pick and keep aside products from the production line till the first defective item is obtained. Let the random variable X represent the number of products that are kept aside (Assume that the first defective item is also kept aside). Find $\text{Var}(X)$. [1 mark]

   (a) $\dfrac{1}{100}$

   (b) $\dfrac{99}{100}$

   (c) 100

   (d) 9900

   Solution:

   The random variable X represent the number of products that are kept aside (including the first defective item) before the first defective is obtained.

   It is given that 1 out of 100 products are defective.

   Therefore, $X \sim \text{Geometric}\left(\dfrac{1}{100}\right)$

   Now,

   $$\begin{aligned}\text{Var}(X) =& \frac{1-p}{p^2} \\ =& \frac{1-\frac{1}{100}}{(\frac{1}{100})^2} = 9900\end{aligned}$$

   Hence, the correct option is (d).

2. Two coins are tossed. The probabilities of occurrence of tail on the first and the second coin are 0.6 and 0.4, respectively. If the random variable X represents the number of heads obtained, find the expected value of X. (Enter the answer correct to 2 decimal points). [1 mark]

   Answer: 1

   Solution:

Given,

$$P(\text{tail occurs on the first coin}) = 0.6. \tag{1}$$
$$P(\text{tail occurs on the second coin}) = 0.4. \tag{2}$$

Random variable X denote the number of heads obtained after the tossing of two coins. Therefore, X will take the values in $\{0, 1, 2\}$.
Now,

$$\begin{aligned} E(X) =& \sum_{x \in X} xP(X = x) \\ =& 0.P(X = 0) + 1.P(X = 1) + 2.P(X = 2) \end{aligned}$$

$$\begin{aligned} P(X = 1) =& P(\text{H on first coin and T on second coin}) + P(\text{T on first coin and H on second coin}) \\ =& (0.4 \times 0.4) + (0.6 \times 0.6) \\ =& 0.52 \end{aligned}$$

$$\begin{aligned} P(X = 2) =& P(\text{H on both the coins}) \\ &= (0.4 \times 0.6) = 0.24 \end{aligned}$$

Therefore, $E(X) = 0.52 + (2 \times 0.24) = 1$

3. Let the two random variables X and Y be independent with means equal to 10 and 20, and variances equal to 2 and 4, respectively. Find the value of $\text{Var}(XY)$.
Hint: If X and Y are independent, X^2 and Y^2 are also independent. [1 mark]
Answer: 1208

Solution:

Mean and variance of X is 10 and 2, respectively.
Mean and variance of Y is 20 and 4, respectively.

$$\begin{aligned} \text{Var(XY)} =& E[(XY)^2] - (E[XY])^2 \\ =& E[X^2Y^2] - (E[X]E[Y])^2, \quad X \text{ and } Y \text{ are independent.} \\ =& E[X^2]E[Y^2] - E[X]^2E[Y]^2, \quad X \text{ and } Y \text{ are independent.} \\ =& (Var(X) + E[X]^2)(Var(Y) + E[Y]^2) - E[X]^2E[Y]^2 \\ =& (2 + 10^2)(4 + 20^2) - 10^2 20^2 \\ =& (102 \times 404) - 40000 \\ =& 41208 - 40000 \\ =& 1208 \end{aligned}$$

4. Let X and Y be two independent discrete random variables. Define random variables U and V as

$$U = \frac{X - E(X)}{SD(X)}, \qquad V = \frac{Y - E(Y)}{SD(Y)}$$

Find $\text{Cov}(U, V)$. [1 mark]

Answer: 0

Solution:

$\text{Cov}(U, V) = E(UV) - E(U)E(V)$.

Since U and V are the standardized form of random variables X and Y, respectively,

$$E(U) = E(V) = 0 \quad \text{and} \quad Var(X) = Var(Y) = 1$$

Now,

$$\begin{aligned}
\text{Cov}(U, V) &= E(UV) \\
&= E\left[\left(\frac{X - E(X)}{SD(X)}\right)\left(\frac{Y - E(Y)}{SD(Y)}\right)\right] \\
&= \frac{1}{SD(X)SD(Y)} E[(X - E(X))(Y - E(Y))] \\
&= E[XY - XE(Y) - YE(X) + E(X)E(Y)] \\
&= E[XY] - E[X]E(Y) - E[Y]E(X) + E(X)E(Y)
\end{aligned}$$

Since X and Y are independent, $E(XY) = E(X)E(Y)$
Therefore, $\text{Cov}(U, V) = E[X]E(Y) - E[X]E(Y) = 0$.

Use the following information to answer questions (5) and (6).
Number of people (X) who make a reservation in a restaurant a day is a random variable with mean equal to 10 and variance equal to 2.

5. Using Markov's inequality, find a bound on the probability that on a particular day, the number of reservations will exceed 30. [1 mark]

   (a) $P(X > 30) \leq \frac{1}{4}$

   (b) $P(X > 30) \geq \frac{1}{3}$

   (c) $P(X > 30) \leq \frac{10}{31}$

   (d) $P(X > 30) > \frac{10}{31}$

Solution:

Random variable X represents the number of people who make reservation in a restaurant. It is given that

$$E(X) = 10 \tag{3}$$

Using Markov's inequality, we know that

$$P(X \geq c) \leq \frac{\mu}{c}$$

Therefore, $P(X > 30) = P(X \geq 31) \leq \dfrac{10}{31}$.
Therefore, the correct option is (c).

6. Find a bound on the probability that on a particular day, number of reservations made will lie in between 6 and 14 using Chebyshev's inequality. [2 marks]

   (a) $P(6 < X < 14) \leq \dfrac{7}{8}$

   (b) $P(6 < X < 14) \geq \dfrac{7}{8}$

   (c) $P(6 < X < 14) > \dfrac{7}{8}$

   (d) $P(6 < X < 14) \leq \dfrac{1}{8}$

Solution:

Using the Chebyshev's inequality, we know that

$$P(\mid X - \mu \mid \geq k\sigma) \leq \frac{1}{k^2} \tag{4}$$

$$P(\mu - k\sigma < X < \mu + k\sigma) \geq 1 - \frac{1}{k^2} \tag{5}$$

Given $\mu = 10$ and $\sigma^2 = 2$
Now, we can write $P(6 < X < 14)$ as
$P(10 - k\sigma < X < 10 + k\sigma) \geq 1 - \dfrac{1}{k^2}$. Using (5)
Now, let

$$10 - k\sigma = 6 \tag{6}$$

$$10 + k\sigma = 14 \tag{7}$$

Solving (6) and (7), we get $k\sigma = 4$

$\Rightarrow k = \dfrac{4}{\sigma} \Rightarrow k^2 = \dfrac{16}{2} = 8$

Therefore, $P(6 < X < 14) \geq 1 - \dfrac{1}{8} = \dfrac{7}{8}$

Hence, the correct option is (b).

7. The joint probability mass function of three discrete random variables X, Y and Z is given as

   $p(0,1,2) = p(0,2,3) = p(1,0,-2) = \dfrac{1}{3}$

   Calculate $\text{Var}(XY + 2Z)$. [2 mark]

   (a) $\dfrac{52}{9}$

   (b) $\dfrac{32}{9}$

   (c) $\dfrac{80}{3}$

   (d) $\dfrac{56}{3}$

   Solution:

t_1	t_2	t_3	$t_1t_2 + 2t_3$	$f_{XYZ}(t_1, t_2, t_3)$
0	1	2	4	1/3
0	2	3	6	1/3
1	0	-2	-4	1/3

Joint PMF of X, Y and Z.

$XY + 2Z$ will take the values in {-4, 6, 4} with the probabilities $\dfrac{1}{3}$ each.

$$\begin{aligned} E(XY + 2Z) &= \frac{1}{3}[-4 + 6 + 4] \\ &= \frac{6}{3} = 2 \end{aligned}$$

$$\begin{aligned} E[(XY + 2Z)^2] &= \frac{1}{3}[(-4)^2 + 6^2 + 4^2] \\ &= \frac{1}{3}[16 + 36 + 16] \\ &= \frac{68}{3} \end{aligned}$$

Now,

$$\begin{aligned}\text{Var}(XY+2Z) =& E[(XY+2Z)^2] - [E(XY+2Z)]^2 \\ =& \frac{68}{3} - 2^2 \\ =& \frac{56}{3}\end{aligned}$$

Hence, the correct option is (d).

8. An urn contains 5 white balls and 5 red balls. 2 balls are selected at random. Let X denote the number of red balls drawn and let Y denote the number of white balls drawn. Find the correlation coefficient between X and Y. [2 marks]

   (a) $\rho(X, Y) = 1$
   (b) $\rho(X, Y) = -1$
   (c) $\rho(X, Y) = 0$
   (d) $\rho(X, Y) = -0.5$

   Solution:

   Two balls are selected at random from the urn containing 5 white and 5 red balls.

   Random variable X represent the number of red balls drawn.
   Therefore, X will take values in {0, 1, 2}.
   Random variable Y represent the number of white balls drawn.
   Therefore, Y will take values in {0, 1, 2}.

   Joint probability distribution of X and Y is given by

Y \ X	0	1	2
0	0	0	$\frac{10}{45}$
1	0	$\frac{25}{45}$	0
2	$\frac{10}{45}$	0	0

   Joint distribution of X and Y.

   Now,

$$\begin{aligned}E(X) =& \left(0 \times \frac{10}{45}\right) + \left(1 \times \frac{25}{45}\right) + \left(2 \times \frac{10}{45}\right) \\ =& 1\end{aligned}$$

Similarly, $E(Y) = 1$.

$$E(X^2) = \left(0 \times \frac{10}{45}\right) + \left(1 \times \frac{25}{45}\right) + \left(2^2 \times \frac{10}{45}\right)$$
$$= \frac{65}{45}$$

Similarly, $E(Y^2) = \dfrac{65}{45}$.

Now, $\text{Var}(X) = \text{Var}(Y) = \dfrac{65}{45} - (1)^2 = \dfrac{20}{45}$

$$E(XY) = \left(0 \times \frac{10}{45}\right) + \left(1 \times \frac{25}{45}\right) + (2 \times 0)$$
$$= \frac{25}{45}$$

Correlation coefficient between X and Y is given by

$$\begin{aligned}\rho(X,Y) =& \frac{\text{Cov}(X,Y)}{SD(X)SD(Y)}\\ =& \frac{E(XY) - E(X)E(Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}\\ =& \frac{(\frac{25}{45} - 1)}{\sqrt{(\frac{20}{45}) \times (\frac{20}{45})}}\\ =& -1\end{aligned}$$

Therefore, the correct option is (b).

9. Five students each from class 8, 9 and 10 have been nominated for the formation of the school committee. The number of boys and girls who are selected from each of the classes is given in Table 4.1.A.

	Class 8	Class 9	Class 10
Girls	2	2	3
Boys	3	3	2

Table 4.1.A: Total number of boys and girls selected.

If the committee comprises of two students from each class, find the expected number of girls in the committee. (Enter the answer correct to 1 decimal point) [2 marks]

Answer: 2.8

Solution:

Let X_1 represent the number of girls from class eight in the school committee.
Let X_2 represent the number of girls from class nine in the school committee.
Let X_3 represent the number of girls from class ten in the school committee.

We need to find $E(X_1 + X_2 + X_3)$.

We know that $E(X_1 + X_2 + X_3) = E(X_1) + E(X_2) + E(X_3)$.

Since total number of girls selected from class eight is 2, therefore, the committee can comprise of either 0 girl or 1 girl or 2 girls from class eight.
i.e. X_1 will take values in $\{0, 1, 2\}$.

Now

$$P(X_1 = 0) = \frac{{}^3C_2}{{}^5C_2} = \frac{3}{10}$$
$$P(X_1 = 1) = \frac{{}^3C_1 \times {}^2C_1}{{}^5C_2} = \frac{6}{10}$$
$$P(X_1 = 2) = \frac{{}^2C_2}{{}^5C_2} = \frac{1}{10}$$

Therefore, $E(X_1) = \left(0 \times \frac{3}{10}\right) + \left(1 \times \frac{6}{10}\right) + \left(2 \times \frac{1}{10}\right) = \frac{8}{10}$

Similarly, total number of girls selected from class nine is 2, therefore, the committee can comprise of either 0 girl or 1 girl or 2 girls from class nine.
i.e. X_2 will take values in $\{0, 1, 2\}$, hence $E(X_2) = \frac{8}{10}$.

Total number of girls selected from class ten is 3 and we have to select 2 students from each class, therefore, the committee can comprise of either 0 girl or 1 girl or 2 girls from class ten.
i.e. X_3 will take values in $\{0, 1, 2\}$.

Now

$$P(X_3 = 0) = \frac{{}^2C_2}{{}^5C_2} = \frac{1}{10}$$
$$P(X_3 = 1) = \frac{{}^3C_1 \times {}^2C_1}{{}^5C_2} = \frac{6}{10}$$
$$P(X_3 = 2) = \frac{{}^3C_2}{{}^5C_2} = \frac{3}{10}$$

Therefore, $E(X_3) = \left(0 \times \frac{1}{10}\right) + \left(1 \times \frac{6}{10}\right) + \left(2 \times \frac{3}{10}\right) = \frac{12}{10}$

Now

$$\begin{aligned} E(X_1 + X_2 + X_3) =& E(X_1) + E(X_2) + E(X_3) \\ =& \frac{8}{10} + \frac{8}{10} + \frac{12}{10} \\ =& 2.8 \end{aligned}$$

Hence, expected number of girls in the class committee is 2.8.

10. A share of a company costs ₹1000 today. Suppose today's share price increases by 50% with probability 0.6 and decreases by 50% with probability 0.4. Independent of today, suppose that tomorrow's share price increases by 20% with probability 0.2, and decreases by 30% with probability 0.8. If you decide to buy 3 shares today, find the expected profit (in ₹) at the end of 2 days. [2 marks]

    (a) -120
    (b) 360
    (c) 120
    (d) -360

    Solution:

    The cost price of a share of the company is ₹1000.
    Let the random variable X represent the price of the share at the end of 2 days.

    Price can either go up by 50% with probability 0.6 or can go down by 50% with probability 0.4 on the first day.
    Independent of today, the share price can either go up by 20% with probability 0.2 or can go down by 30% with probability 0.8.

    i.e. If the share price increases by 50% on the first day, the price of the share will become ₹1500.
    And the price of the share at the end of two days if the share prices increases by 20% is ₹$\left(1500 \times \frac{20}{100} + 1500\right)$=₹1800 with probability $(0.6 \times 0.2) = 0.12$.
    Similarly, the price of the share at the end of two days if the share prices decreases by 30% is ₹$\left(1500 - 1500 \times \frac{30}{100}\right)$=₹1050 with probability $(0.6 \times 0.8) = 0.48$.

    Again, if the share price decreases by 50% on the first day, the price of the share will become ₹500.
    And the price of the share at the end of two days if the share prices increases by 20%

is ₹$\left(500 \times \dfrac{20}{100} + 500\right)$=₹600 with probability $(0.4 \times 0.2) = 0.08$.
Similarly, the price of the share at the end of two days if the share prices decreases by 30% is ₹$\left(500 - 500 \times \dfrac{30}{100}\right)$=₹350 with probability $(0.4 \times 0.8) = 0.32$.

Therefore, X will take values in $\{1800, 1050, 600, 350\}$, where

$P(X = 1800) = 0.12$
$P(X = 1050) = 0.48$
$P(X = 600) = 0.08$
$P(X = 350) = 0.32$

Now,

$$\begin{aligned} E(X) =& (1800 \times 0.12) + (1050 \times 0.48) + (600 \times 0.08) + (350 \times 0.32) \\ =& 880 \end{aligned}$$

The expected gain at the end of two days if you buy one share is ₹(880-1000) = -₹120.
Therefore, if you buy 3 shares of the company, expected gain will be -₹360.

Hence, the correct option is (d).

11. A lottery has 500 tickets out of which only 2 tickets contain prizes worth ₹500 and ₹1,000; the rest are worth ₹0. If one has bought 2 tickets, what will be his/her expected gain (in ₹)? [2 marks]
Answer: 6

Solution:

In the lottery, only two tickets out of 500 contain prizes worth ₹500 and ₹1,000.
If one has bought two tickets, one can get the prizes worth ₹0, ₹500, ₹1,000 and ₹1,500.

Let the random variable X represent the worth of the prizes of two tickets.
Therefore, X will take values in $\{0, 500, 1000, 1500\}$.

$P(X = 0) = P(\text{Both the tickets are worth ₹0}) = \dfrac{^{498}C_2}{^{500}C_2}$

$P(X = 500) = P(\text{One of the ticket is worth ₹0 and the other is worth ₹500}) = \dfrac{^{498}C_1\, ^{1}C_1}{^{500}C_2}$

$P(X = 1000) = P(\text{One of the ticket is worth ₹0 and the other is worth ₹1000}) = \dfrac{^{498}C_1\, ^{1}C_1}{^{500}C_2}$

$P(X = 1500) = P(\text{One of the ticket is worth ₹500 and the other is worth ₹1000}) =$

$$\frac{{}^{2}C_2}{{}^{500}C_2}$$

$$
\begin{aligned}
E(X) &= \left(0 \times \frac{{}^{498}C_2}{{}^{500}C_2}\right) + \left(500 \times \frac{{}^{498}C_1\,{}^{1}C_1}{{}^{500}C_2}\right) + \left(1000 \times \frac{{}^{498}C_1\,{}^{1}C_1}{{}^{500}C_2}\right) + \left(1500 \times \frac{{}^{2}C_2}{{}^{500}C_2}\right) \\
&= \frac{1}{{}^{500}C_2}[500 \times {}^{498}C_1 + 1000 \times {}^{498}C_1 + 1 \times 1500] \\
&= \frac{1}{{}^{500}C_2}[249000 + 498000 + 1500] \\
&= \frac{748500}{124750} = 6
\end{aligned}
$$

Therefore, the expected gain is ₹6.

12. Number of cars (X) that visit Garage A each day is a random variable with mean 45 and variance 10 while the number of cars (Y) that visit Garage B each day is a random variable with mean 45 and variance 20. If the arrival of cars in garages A and B are independent, find an upper bound on the probability that the difference in the number of cars arriving in Garage A and Garage B on a particular day is greater than or equal to 10. [3 marks]

(a) $\frac{3}{10}$

(b) $\frac{2}{10}$

(c) $\frac{1}{10}$

(d) $\frac{1}{4}$

Solution:

The random variable X represent the number of cars that come each day in Garage A. Let the mean and variance of X be denoted by μ_X and σ_X^2 respectively.
Given $\mu_X = 45, \sigma_X^2 = 10$.

The random variable Y represent the number of cars that come each day in Garage B. Let the mean and variance of Y be denoted by μ_Y and σ_Y^2 respectively.
Given $\mu_Y = 45, \sigma_Y^2 = 20$.

Arrival of cars in shop A and B are independent. That implies X and Y are independent.

Difference in the number of cars arriving in shop A and shop B is given by $\mid X - Y \mid$.
Let $\mu = E(X - Y) = E(X) - E(Y) = 10 - 10 = 0$
and $\sigma^2 = Var(X - Y) = Var(X) + Var(Y) = 10 + 20 = 30$. (Since X and Y are independent.)

Using Chebysheb's inequality,

$$P\{\mid (X-Y)-0 \mid\geq k\sigma\} \leq \frac{1}{k^2} \tag{8}$$

for $k > 0$.
Substituting $k\sigma = 10$ in equation (8), we get

$$P\{\mid X-Y \mid\geq 10\} \leq \frac{1}{(\frac{100}{30})}$$

$$\Rightarrow P\{\mid X-Y \mid\geq 10\} \leq \frac{3}{10}$$

Therefore, option (a) is correct.