Statistics for Data Science-1

Week-10 Graded Assignment

1. There are 2^n numbered cards in a deck among which ${}^nC_i$ cards bear the number i ; i=0,1,2,...,n. From the deck, m cards are drawn with replacement. What is the expectation of the sum of their numbers? (Enter the answer correct to one decimal accuracy)

   Answer: $\dfrac{mn}{2}$

   Solution:

   Let, X_j; i=1,2,...,m be the random variable representing the number on the j^{th} card drawn.

X_j	0	1	2	...	n
Number of cards	${}^nC_0$	${}^nC_1$	${}^nC_2$	...	${}^nC_n$
$P(X_j = x)$	$\dfrac{{}^nC_0}{2^n}$	$\dfrac{{}^nC_1}{2^n}$	$\dfrac{{}^nC_2}{2^n}$	...	$\dfrac{{}^nC_n}{2^n}$

Table 10.1

$E(X_j) = \sum_{x=0}^{n} xP(X_j = x) = \sum_{x=0}^{n} x \times \dfrac{{}^nC_x}{2^n}$

$E(X_j) = \dfrac{1}{2^n} \times [(0 \times {}^nC_0) + (1 \times {}^nC_1) + (2 \times {}^nC_2) + ... + (n \times {}^nC_n)]$

$E(X_j) = \dfrac{1}{2^n} \times \left[0 + (1 \times n) + \left(2 \times \dfrac{n(n-1)}{2!}\right) + \left(3 \times \dfrac{n(n-1)(n-2)}{3!}\right) ... + (n \times 1)\right]$

$E(X_j) = \dfrac{n}{2^n} \times \left[1 + (n-1) + \left(\dfrac{(n-1)(n-2)}{2!}\right) + ... + 1\right]$

$E(X_j) = \dfrac{n}{2^n} \times [{}^{n-1}C_0 + {}^{n-1}C_1 + {}^{n-1}C_2 + ... + {}^{n-1}C_{n-1}]$

$E(X_j) = \dfrac{n}{2^n} \times 2^{n-1} = \dfrac{n}{2}$

Therefore, Expectation of the sum of their numbers is given by:

$= \sum_{j=1}^{m} E(X_j)$

$= \sum_{j=1}^{m} \dfrac{n}{2}$

$= \dfrac{mn}{2}$

Suppose, we substitute values of n and m as 4 and 7 respectively, then

Let, X_j; i=1,2,...,7 be the random variable representing the number on the j^{th} card drawn.

X_j	0	1	2	3	4
Number of cards	${}^4C_0$	${}^4C_1$	${}^4C_2$	${}^4C_3$	${}^4C_4$
$P(X_j = x)$	$\frac{{}^4C_0}{2^4}$	$\frac{{}^4C_1}{2^4}$	$\frac{{}^4C_2}{2^4}$	$\frac{{}^4C_3}{2^4}$	$\frac{{}^4C_4}{2^4}$

Table 10.1

$$E(X_j) = \sum_{x=0}^{4} xP(X_j = x) = \sum_{x=0}^{4} x \times \frac{{}^4C_x}{2^4}$$
$$E(X_j) = \frac{1}{2^4} \times [(0 \times {}^4C_0) + (1 \times {}^4C_1) + (2 \times {}^4C_2) + (3 \times {}^4C_3) + (4 \times {}^4C_4)]$$
$$E(X_j) = \frac{1}{2^4} \times [0 + (1 \times 4) + \left(2 \times \frac{4(4-1)}{2!}\right) + \left(3 \times \frac{4(4-1)(4-2)}{3!}\right) + (4 \times 1)]$$
$$E(X_j) = \frac{4}{2^4} \times [1 + (4-1) + \left(\frac{(4-1)(4-2)}{2!}\right) + 1]$$
$$E(X_j) = \frac{4}{2^4} \times [{}^3C_0 + {}^3C_1 + {}^3C_2 + {}^3C_3]$$
$$E(X_j) = \frac{4}{2^4} \times 2^3 = \frac{4}{2} = 2$$

Therefore, Expectation of the sum of their numbers is given by:
$= \sum_{j=1}^{7} E(X_j) = \sum_{j=1}^{7} 2 = 14$

An unbiased die is thrown $n+2$ times. After each throw a '+' is recorded for 2 or 5 and '-' is recorded for 1,3,4 or 6, the signs forming an ordered sequence. To each, except the first and last sign, a random variable X_i; i=1,2,...,n is associated which takes the value 1 if both of its neighbouring sign differs from the one between them and 0 otherwise. If the random variable Y is defined as $Y = aS + b$ where, $S = \sum_{i=1}^{n} X_i$, then use the given information to answer question (2) and (3).

2. Find the expected value of Y .

   (a) $\left(a \times \frac{2n}{9}\right) + b$

   (b) $\left(a \times \frac{2n}{9}\right)$

   (c) $\frac{2n}{9} + b$

   (d) $\frac{2n}{9}$

Answer: a

Solution:

$$X_i = \begin{cases} 1 & \text{if the pattern is either '+ - +' or '- + -'} \\ 0 & \text{otherwise} \end{cases}$$

$E(X_i) = 1 \times P(X_i = 1) + 0 \times P(X_i = 0) = P(X_i = 1)$

Now,

$P(X_i = 1) =$P(Pattern = '+ - +') + P(pattern='- + -')

$P(X_i = 1) = \left(\frac{4}{6}\right)^2 \times \left(\frac{2}{6}\right) + \left(\frac{4}{6}\right) \times \left(\frac{2}{6}\right)^2 = \frac{2}{9}$

$E(S) = E(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} E(X_i)$

$E(S) = \frac{2n}{9}$

Therefore, $E(Y) = aE(S) + b$

$E(Y) = \left(a \times \frac{2n}{9}\right) + b$

Suppose, we substitute values of n, a and b as 10, 9 and 1 respectively then,

$E(S) = E(\sum_{i=1}^{10} X_i) = \sum_{i=1}^{10} E(X_i)$

$E(S) = \sum_{i=1}^{10} \frac{2}{9} = \frac{20}{9}$

Therefore, $E(Y) = 9E(S) + 1 = 9 \times \frac{20}{9} + 1$

$E(Y) = 20 + 1 = 21$

3. Which of the following statement(s) is/are true?

   a. $V(Y) = a^2V(S) + b$
   b. $V(Y) = a^2V(S)$
   c. $V(Y) \neq a^2V(S)$
   d. $E(Y) = a^2E(S) + b$
   e. $E(Y) = aE(S) + b$

Answer: b,e

By the property of Expectation and Variance, we get:

$V(Y) = a^2V(S)$

$E(Y) = aE(S) + b$ (always)

Hence, option (b) and (e) is correct.

Suppose, we substitute values of a and b as 9 and 1 respectively then,

$V(Y) = (9)^2 \times V(S) = 81V(S)$
$E(Y) = 9E(S) + 1$ (always)

Amandeep is in the middle of a bridge of infinite length. He takes the unit step to the right with probability p and to the left with probability $1 - p$. Assume that the movements are independent of each other.
Hint: Consider the random variable X_i associated with the i^{th} step defined as:

$$X_i = \begin{cases} 1 & \text{if the step of Amandeep is towards the right} \\ -1 & \text{if the step of Amandeep is towards the left} \end{cases}$$

Using this information, answer question (4) and (5).

4. What is the expected distance between the starting point and end point of Amandeep after n steps?

   (a). $2p - 1$

   (b). $n(2p - 1)$

   (c). $1 - 2p$

   (d). $n(1 - 2p)$

   Answer: *b*
   Solution:
   Let us associate a random variable X_i with the i^{th} step.

$$X_i = \begin{cases} 1 & \text{if } i^{th} \text{ step of Amandeep is towards the right} \\ -1 & \text{if } i^{th} \text{ step of Amandeep is towards the left} \end{cases}$$

   $E(X_i) = 1 \times P(X_i = 1) + (-1) \times P(X_i = -1)$
   $E(X_i) = 1 \times p - 1 \times (1 - p) = 2p - 1$
   $S = X_1 + X_2 + ... + X_n$ represents the random distance moved from the starting point after n steps.
   Therefore, $E(S) = \sum_{i=1}^{n} E(X_i) = n(2p - 1)$
   Hence, expected distance between the starting point and end point of Amandeep after n steps is $n(2p - 1)$

   For example: p=0.6 , n=7
   Let us associate a random variable X_i with the i^{th} step.

$$X_i = \begin{cases} 1 & \text{if } i^{th} \text{ step of Amandeep is towards the right} \\ -1 & \text{if } i^{th} \text{ step of Amandeep is towards the left} \end{cases}$$

   $E(X_i) = 1 \times P(X_i = 1) + (-1) \times P(X_i = -1)$
   $E(X_i) = 1 \times 0.6 - 1 \times 0.4 = 0.2$

$S = X_1 + X_2 + ... + X_7$ represents the random distance moved from the starting point after 7 steps.
Therefore, $E(S) = \sum_{i=1}^{7} E(X_i) = 7(0.2) = 1.4$
Hence, the expected distance between the starting point and end point of Amandeep after 7 steps is 1.4.

5. What is the variance distance between the starting point and end point of Amandeep after n steps?

   (a). $4np(1-p)$

   (b). $4p(1-p)$

   (c). 1

   (d). $n^2(2p-1)^2 - 1$

   Answer: a
   Solution:
   $E(X_i)^2 = 1^2 \times P(X_i = 1) + (-1)^2 \times P(X_i = -1)$
   $E(X_i)^2 = p + (1-p) = 1$
   $V(X_i) = E(X_i)^2 - [E(X_i)]^2$
   $V(X_i) = 1 - (2p-1)^2 = 4p(1-p)$
   Therefore, $V(S) = \sum_{i=1}^{n} V(X_i)$ (because, movements of steps are independent)
   Hence, $V(S) = \sum_{i=1}^{n} 4p(1-p) = 4np(1-p)$

   For Example: p=0.6 , n=7
   $E(X_i)^2 = 1^2 \times P(X_i = 1) + (-1)^2 \times P(X_i = -1)$
   $E(X_i)^2 = 0.6 + 0.4 = 1$
   $V(X_i) = E(X_i)^2 - [E(X_i)]^2$
   $V(X_i) = 1 - (0.2)^2 = 0.96$
   Therefore, $V(S) = \sum_{i=1}^{7} V(X_i)$
   Because, movements of steps are independent.
   Hence, $V(S) = \sum_{i=1}^{7}(0.96) = 7 \times (0.96) = 6.72$

6. A box contains a white and b black balls. c balls are drawn at random without replacement. Find the expected value of the number of white balls drawn?(Enter the answer correct to 2 decimal places).
   Solution:
   Let X denote the number of white balls drawn. The probability distribution of X is obtained as follows:

X	0	1	2	...	c
$p(x)$	$\frac{^bC_c}{^{a+b}C_c}$	$\frac{^aC_1 \times {^bC_{c-1}}}{^{a+b}C_c}$	$\frac{^aC_2 \times {^bC_{c-2}}}{^{a+b}C_c}$	...	$\frac{^aC_c}{^{a+b}C_c}$

Then expected number of white balls drawn is :
$$E(X) = 0 \times \frac{{}^{b}C_c}{{}^{a+b}C_c} + 1 \times \frac{{}^{a}C_1 \times {}^{b}C_{c-1}}{{}^{a+b}C_c} + 2 \times \frac{{}^{a}C_2 \times {}^{b}C_{c-2}}{{}^{a+b}C_c} + ... + c \times \frac{{}^{a}C_c}{{}^{a+b}C_c}$$

For example: a=7, b=4, c=2
Let X denote the number of white balls drawn. The probability distribution of X is obtained as follows:

X	0	1	2
$p(x)$	$\frac{{}^{4}C_2}{{}^{11}C_2} = \frac{6}{55}$	$\frac{{}^{7}C_1 \times {}^{4}C_1}{{}^{11}C_2} = \frac{28}{55}$	$\frac{{}^{7}C_2}{{}^{11}C_2} = \frac{21}{55}$

Then expected number of white balls drawn is :
$E(X) = 0 \times \frac{6}{55} + 1 \times \frac{28}{55} + 2 \times \frac{21}{55} = \frac{70}{55} = 1.27.$

7. Rohit wants to open his door with 5 keys(out of which 1 will open the door) and tries the keys independently and at random. If unsuccessful keys are eliminated from further selection, then Find the expected number of trials required to open the door.
(Hint: Suppose Rohit gets the first success at x^{th} trial, i.e., he is unable to open the door in the first $(x-1)$ trials. And, P(he gets first success at second trial)=$(1-\frac{1}{5}) \times \frac{1}{4}$)

(a). 1

(b). 9

(c). 3

(d). 2

Answer:c
Solution:

If unsuccessful keys are eliminated from further selection, then the random variable X will take the values from 1 to n. In this case, we have
Probability of success at the first trial=$\frac{1}{5}$

Probability of success at the second trial=$\frac{1}{4}$

Probability of success at the third trial=$\frac{1}{3}$

Probability of success at the fourth trial=$\frac{1}{2}$

Probability of success at the fifth trial=1 and so on.

Hence probability of 1^{st} success at the 2^{nd} trial $= (1 - \frac{1}{5}) \times \frac{1}{4} = \frac{1}{5}$

probability of 1^{st} success at the 3^{rd} trial $= (1-\frac{1}{5}) \times (1-\frac{1}{4}) \times \frac{1}{3} = \frac{1}{5}$

and so on. In general,

$p(x)$= probability of 1^{st} success at the x^{th} trial=$\frac{1}{5}$; $x = 1, 2, 3, 4, 5$

Therefore,

$$E(X) = \sum_{x=1}^{5} xp(x) = 1 \times \frac{1}{5} + 2 \times \frac{1}{5} + 3 \times \frac{1}{5} + 4 \times \frac{1}{5} + 5 \times \frac{1}{5} = 3$$

8. X and Y are independent random variables with means m_1 and m_2, and variances v_1 and v_2 respectively. Find the variance of $aX + bY$?

   Answer: $a^2 \times v_1 + b^2 \times v_2$

   Solution:

   Since X and Y are independent random variables.

   Therefore, $V(aX + bY) = a^2 \times V(X) + b^2 \times V(Y) = a^2 \times v_1 + b^2 \times v_2$

   For example: $m_1 = 10$**,** $m_2 = 20$**,** $v_1 = 2$ **and** $v_2 = 3$

   Since X and Y are independent random variables.

   Therefore, $V(3X + 4Y) = 9 \times V(X) + 16 \times V(Y) = 9 \times 2 + 16 \times 3 = 18 + 48 = 66$

9. Let X be a random variable with the following probability distribution:

X	a	b	c
$P(X = x)$	$\frac{1}{d}$	$\frac{1}{e}$	$\frac{f}{g}$

   Calculate the value of $E(2X + 1)^2$.(Enter the answer correct to 2 decimal places)

   Answer: $4 \times \left(a^2 \times \frac{1}{d} + b^2 \times \frac{1}{e} + c^2 \times \frac{1}{f}\right) + 4 \times \left(a \times \frac{1}{d} + b \times \frac{1}{e} + c \times \frac{1}{f}\right) + 1$

   Solution:

   $E(X) = \Sigma xp(x) = a \times \frac{1}{d} + b \times \frac{1}{e} + c \times \frac{1}{f}$.

   $E(X^2) = \Sigma x^2 p(x) = a^2 \times \frac{1}{d} + b^2 \times \frac{1}{e} + c^2 \times \frac{1}{f}$

   $E(2X + 1)^2 = E(4X^2 + 4X + 1) = 4E(X^2) + 4E(X) + 1$

   For example: $a = -3, b = 6, c = 9, d = 6, e = 2$ and $f = 3$

X	-3	6	9
$P(X = x)$	$\frac{1}{6}$	$\frac{1}{2}$	$\frac{1}{3}$

   $E(X) = \Sigma xp(x) = (-3) \times \frac{1}{6} + 6 \times \frac{1}{2} + 9 \times \frac{1}{3} = \frac{11}{2}$.

$E(X^2) = \Sigma x^2 p(x) = 9 \times \frac{1}{6} + 36 \times \frac{1}{2} + 81 \times \frac{1}{3} = \frac{93}{2}$

$E(2X+1)^2 = E(4X^2 + 4X + 1) = 4E(X^2) + 4E(X) + 1 = 4 \times \frac{93}{2} + 4 \times \frac{11}{2} + 1 = 209$

10. Suppose that X is a random variable for which $E(X) = m$ and $Var(X) = v$. Find the positive values of a and b such that $Y = aX - b$, has expectation 0 and variance 1.

    a. $\frac{-1}{\sqrt{v}}, \frac{-m}{\sqrt{v}}$

    b. $\frac{1}{v}, \frac{m}{\sqrt{v}}$

    c. $\frac{1}{\sqrt{v}}, \frac{m}{\sqrt{v}}$

    d. $1, m$

    Answer: c

    Solution:

    $E(Y) = aE(X) - b = 0 \implies ma - b = 0 \implies ma = b$ ...(1)

    Now,$V(Y) = a^2 V(X) = 1 \implies a^2 \times v = 1 \implies a^2 = \frac{1}{v} \implies a = \frac{1}{\sqrt{v}}$

    Now putting value of a in equation(1) , we get $b = \frac{m}{\sqrt{v}}$

    Hence, option(c) is correct.

    For example: m=10 and v=25

    $E(Y) = aE(X) - b = 0 \implies 10a - b = 0 \implies 10a = b$ ...(1)

    Now,$V(Y) = a^2 V(X) = 1 \implies a^2 \times 25 = 1 \implies a^2 = \frac{1}{25} \implies a = \frac{1}{5}$

    Now putting value of a in equation(1) , we get $b = 2$