Statistics for Data Science-1

Week-11 Graded Assignment

1. A match predictor claims that he can predict the result of a match correctly $x\%$ of the time. It is agreed that his claim will be accepted if he correctly predicts the results of at least m of n matches. What is the probability that his claim gets rejected?

   (a). $\sum_{i=0}^{m-1} {}^nC_i \; p^i(1-p)^{n-i}$

   (b). $\sum_{i=0}^{m} {}^nC_i \; p^i(1-p)^{n-i}$

   (c). $\sum_{i=1}^{m-1} {}^nC_i \; p^i(1-p)^{n-i}$

   (d). $\sum_{i=1}^{m} {}^nC_i \; p^i(1-p)^{n-i}$

   Answer: a

   Solution:

   P(Correct match prediction)$= p = \dfrac{x}{100}$

   Let, X be a random variable representing the number of correct match predictions.

   Hence, $X \sim Binomial(n, p)$

   The probability that his claim gets rejected is given by:

$$\begin{aligned} P(X < m) &= \sum_{i=0}^{m-1} P(X = i) \\ &= \sum_{i=0}^{m-1} {}^nC_i \; p^i(1-p)^{n-i} \end{aligned}$$

   Suppose we substitute the values of x, m and n as 75, 5 and 6 respectively, then

   P(Correct match prediction)$= p = 0.75$

   Let, X be a random variable representing the number of correct match predictions.

   Hence, $X \sim Binomial(6, 0.75)$

   The probability that his claim gets rejected is given by:

$$\begin{aligned} P(X < 5) &= \sum_{i=0}^{5-1} P(X = i) \\ &= \sum_{i=0}^{4} {}^6C_i \; 0.75^i(1-0.75)^{6-i} \\ &= 0.466 \end{aligned}$$

   Therefore, the probability that his claim gets rejected is 0.47

2. If $X \sim Binomial(n, p)$, then which of the following statement/s is/are always true? ($n > 0$ and $0 < p < 1$)

   a). $E(X) \leq Var(X)$

   b). $E(X) < Var(X)$

   c). $E(X) \geq Var(X)$

   d). $E(X) > Var(X)$

   e). $Var(X) \leq S.D(X)$

   f). $Var(X) \geq S.D(X)$

   Answer: d

   Solution:

   If $X \sim Binomial(n, p)$, then

   $E(X) = np$ and $Var(X) = np(1 - p)$

   When, $0 < p < 1$ and $n > 0$, $np > np(1 - p)$(always)

   Thus, option(d) is always true.

   For example: $n = 1$ **and** $p = \frac{1}{2}$

   $np = 1 \times \frac{1}{2} = \frac{1}{2}$ and $np(1 - p) = 1 \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$

   Since, $\frac{1}{2} > \frac{1}{4}$, thus $E(X) > Var(X)$

   Option (e) and (f) are not always true.

   For example: $n = 3$ and $p = \frac{1}{2}$

   $Var(X) = np(1 - p) = 3 \times \frac{1}{2} \times \frac{1}{2} = \frac{3}{4} = 0.75$

   $S.D(X) = \sqrt{Var(X)} = \sqrt{0.75} = 0.87$, which is greater than variance in this particular example.

   But, for $n = 16$ and $p = \frac{1}{2}$

   $Var(X) = np(1 - p) = 16 \times \frac{1}{2} \times \frac{1}{2} = \frac{16}{4} = 4$

   $S.D(X) = \sqrt{Var(X)} = \sqrt{4} = 2$, which is less than variance in this particular example.

   Hence, we can say that option (e) and (f) are not always true.

3. Two friends (say 'A' and 'B') could not decide whether to play a racing game or a boxing game on Xbox. They decide to play a card-game first. If 'A' wins at least x rounds out of the y rounds of the card game played, then the boxing game will be played. The chances of 'A' winning in any round of the card game is $a : b$. Find the probability that the boxing game will be played on Xbox? (Enter the answer correct

to 2 decimal places)
Hint: If the chances of happening of an event is $x : y$ then, the probability equals $\dfrac{x}{x+y}$

Answer: $\sum_{i=x}^{y} {}^{y}C_i \; p^i(1-p)^{y-i}$
Solution:
Number of rounds of card game played; $n = y$
P('A' winning a round of card game)$= p = \dfrac{a}{a+b}$
Let, X be a random variable representing the number of rounds of card game won by 'A'.
Hence, $X \sim Binomial(n, p)$
Now, the probability that the boxing game will be played is given by:

$$\begin{aligned} P(X \geq x) &= \sum_{i=x}^{y} P(X = i) \\ &= \sum_{i=x}^{y} {}^{y}C_i \; p^i(1-p)^{y-i} \end{aligned}$$

Suppose we substitute the values of x, y, a and b as 3, 5, 3 and 2 respectively, then
Number of rounds of card game played; $n = 5$
P('A' winning a round of card game)$= p = \dfrac{3}{3+2} = 0.6$
Let, X be a random variable representing the number of rounds of card game won by 'A'.
Hence, $X \sim Binomial(5, 0.6)$
Now, the probability that the boxing game will be played is given by:

$$\begin{aligned} P(X \geq 3) &= \sum_{i=3}^{5} P(X = i) \\ &= \sum_{i=3}^{5} {}^{5}C_i \; 0.6^i(1-0.6)^{5-i} \\ &= 0.68 \end{aligned}$$

Therefore, the probability that the boxing game will be played is 0.68.

4. Let $X \sim Binomial(n, p)$. If the probabilities of x and $x + 1$ successes are a and, b respectively, then find the parameter 'p' of the distribution.(Enter the answer correct to 2 decimal places)

   Answer: $\dfrac{b(x+1)}{a(n-x)+b(x+1)}$

Solution:

$$
\begin{aligned}
\frac{a}{b} &= \frac{P(X=x)}{P(X=x+1)} \\
\frac{a}{b} &= \frac{{}^nC_x p^x (1-p)^{n-x}}{{}^nC_{x+1} p^{x+1} (1-p)^{n-x-1}} \\
\frac{a}{b} &= \frac{{}^nC_x (1-p)}{{}^nC_{x+1} p} \\
ap \times {}^nC_{x+1} &= b(1-p) \times {}^nC_x \\
\frac{ap}{x+1} &= \frac{b(1-p)}{n-x} \\
ap(n-x) &= b(1-p)(x+1) \\
p &= \frac{b(x+1)}{a(n-x)+b(x+1)}
\end{aligned}
$$

Suppose we substitute the values of n, x, a and b as 5, 1, 0.4096 and 0.2048 respectively, then

$$
\begin{aligned}
p &= \frac{0.2048(1+1)}{0.4096(5-1)+0.2048(1+1)} \\
p &= \frac{0.2048(2)}{0.4096(4)+0.2048(2)} \\
p &= \frac{0.4096}{0.4096(4)+0.4096} \\
p &= \frac{1}{5} = 0.2
\end{aligned}
\tag{1}
$$

5. If the expected number of sixes hit by a batsman on n balls is e and the variance for the same is v, then what is the probability of him hitting at least one six on any randomly selected n balls? (Enter the answer correct to 4 decimal places)

   Answer: $1-(1-p)^n$

   Solution:

   Let, X be a random variable representing the number of sixes hit by a batsman in an over.

   Hence, $X \sim Binomial(n, p)$

   $E(X) = np$ and $Var(X) = np(1-p)$

Now,

$$\begin{aligned}
Var(X) &= E(X) \times (1-p) \\
(1-p) &= \frac{v}{e} \\
p &= 1 - \frac{v}{e}
\end{aligned}$$

Also, $e = np$
$\implies n = \dfrac{e}{p}$
Therefore, $P(X \geq 1) = 1 - P(X = 0) = 1 - (1-p)^n$

Suppose we substitute the values of e and v as 4 and $\frac{4}{3}$ respectively, then

$$\begin{aligned}
Var(X) &= E(X) \times (1-p) \\
(1-p) &= \frac{\frac{4}{3}}{4} \\
p &= 1 - \frac{1}{3} = \frac{2}{3}
\end{aligned}$$

Also, $4 = \dfrac{2n}{3}$
$\implies n = 6$
Therefore,
$P(X \geq 1) = 1 - P(X = 0) = 1 - \left(\dfrac{1}{3}\right)^6 = 0.9986$

6. The probability of a student clearing a competitive exam is $\frac{1}{m}$. If he gives the exam n times, then what is the probability of him clearing the exam at least twice? (Enter the answer correct to 2 decimal places)

   Answer: $1 - \left[{}^nC_0 \left(\frac{1}{m}\right)^0 \left(1-\frac{1}{m}\right)^{n-0} + {}^nC_1 \left(\frac{1}{m}\right)^1 \left(1-\frac{1}{m}\right)^{n-1}\right]$

   Solution:
   p = Probability of clearing the exam = $\dfrac{1}{m}$
   1-p = $1 - \dfrac{1}{m}$
   $p(x)$ = Probability of clearing the exam x times out of a total of n
   $= {}^nC_x \left(\dfrac{1}{m}\right)^x \left(1-\dfrac{1}{m}\right)^{n-x}; x = 0, 1, ..., n$
   Thus, Probability of clearing the exam at least twice

$P(X \geq 2) = 1 - P(X < 2) = 1 - [P(X = 0) + P(X = 1)] = 1 - [p(0) + p(1)]$

$$=1 - \left[{}^nC_0 \left(\frac{1}{m}\right)^0 \left(1 - \frac{1}{m}\right)^{n-0} +^n C_1 \left(\frac{1}{m}\right)^1 \left(1 - \frac{1}{m}\right)^{n-1} \right]$$

For example: m=4 and n=7

p = Probability of clearing the exam = $\frac{1}{4}$

1-p = $1 - \frac{1}{4} = \frac{3}{4}$

p(x) = Probability of clearing the exam x times in total of 7

$= {}^7C_x \left(\frac{1}{4}\right)^x \left(\frac{3}{4}\right)^{7-x}$; $x = 0, 1, ..., 7$

Thus, Probability of clearing the exam at least twice

$P(X \geq 2) = 1 - P(X < 2) = 1 - [P(X = 0) + P(X = 1)] = 1 - [p(0) + p(1)]$

$$=1 - \left[{}^7C_0 \left(\frac{1}{4}\right)^0 \left(\frac{3}{4}\right)^{7-0} +^7 C_1 \left(\frac{1}{4}\right)^1 \left(\frac{3}{4}\right)^{7-1} \right] = \frac{4547}{8192} = 0.56$$

7. Choose the correct condition/s about binomial distribution.

   (a). The probability of success p keeps varying for each trial.

   (b). The number of trials n is finite.

   (c). The trials are dependent on each other.

   (d). The trials are independent of each other.

   Answer: b, d

   Solution:

   As we know, if a random variable, $X\ B(n, p)$ then:

   - The total number of trials (n) is fixed.
   - Each of the trials is independent and identically distributed.

   Independent trial means that each of the trials is independent of the other trials. And, identical trials mean that probability of success is same for each of the trials.

8. Rithika wants to test whether the coin she has is a fair coin or not. To test this, she conducted an experiment of tossing the coin 5 times. Binomial random variable X is defined as the total number of heads(i) after 5 tosses. The probability distribution of the binomial random variable is given in Table 11.1.G.

$X=i$	$P(X=i)$
0	${}^5C_0p^0(1-p)^5$
1	${}^5C_1p^1(1-p)^4$
2	${}^5C_2p^2(1-p)^3$
3	${}^5C_3p^3(1-p)^2$
4	${}^5C_4p^4(1-p)^1$
5	${}^5C_5p^5(1-p)^0$

Table 11.1.G

What is the approximate probability of getting a head in tossing the given coin?(Enter the answer correct to one decimal place)

Answer: p

Solution:

Given the coin is tossed 5 times, $n=5$

Given $P(X=5)=0.0009765625 \implies P(X=5)={}^5C_5\times p^5\times(1-p)^0=0.0009765625$
$\implies p^5=0.0009765625 \implies p=0.25$

Thus, probability of head: $p=0.25$ or $\dfrac{1}{4}$

9. At a school function, It is noticed that $a\%$ of the students are not wearing polished shoes and $b\%$ of students are not wearing school ties. It is announced that the students who have committed any of the infractions will be punished, and that these two infractions are independent of one another. If a teacher selects 5 students at random, then find the probability that exactly three of the students will be punished for any of the infractions?

(a). ${}^5C_3\left(\dfrac{100\times(a+b)-(a\times b)}{10000}\right)^3\left(\dfrac{10000-100\times(a+b)+(a\times b)}{10000}\right)^2$

(b). ${}^5C_3\left(\dfrac{a\times b}{10000}\right)^3\left(\dfrac{10000-a\times b}{10000}\right)^2$

(c). ${}^5C_3\left(\dfrac{a+b}{100}\right)^3\left(\dfrac{100-(a+b)}{100}\right)^2$

(d). ${}^5C_3\left(\dfrac{a}{100}\right)^3\left(\dfrac{100-a}{100}\right)^2+{}^5C_3\left(\dfrac{b}{100}\right)^3\left(\dfrac{100-b}{100}\right)^2$

Answer: a

Solution:

Let X be a random variable representing the number of students punished for any one of the infractions.

Define events as follows:

A: Student is not wearing polished shoes.

B: Student is not wearing a school tie.
Given, $P(A) = \dfrac{a}{100}$, $P(B) = \dfrac{b}{100}$

$P(A \cup B) = P(A) + P(B) - P(A \cap B)$
$P(A \cup B) = \dfrac{a}{100} + \dfrac{b}{100} - \left(\dfrac{a}{100} \times \dfrac{b}{100}\right) = \dfrac{a+b}{100} - \left(\dfrac{a}{100} \times \dfrac{b}{100}\right)$

$X \sim Binomial(5, p)$
Where, $p = \dfrac{a+b}{100} - \left(\dfrac{a}{100} \times \dfrac{b}{100}\right)$ and $1 - p = 1 - \left(\dfrac{a+b}{100}\right) + \left(\dfrac{a}{100} \times \dfrac{b}{100}\right)$
Now,
$P(X = 3) = {}^5C_3(p)^3(1-p)^2 = {}^5C_3 \left[\dfrac{a+b}{100} - \left(\dfrac{a}{100} \times \dfrac{b}{100}\right)\right]^3 \left[1 - \left(\dfrac{a+b}{100}\right) + \left(\dfrac{a}{100} \times \dfrac{b}{100}\right)\right]^2$

For example: a =5 and b=10

Let X be a random variable representing the number of students punished for any one of the infractions.
Define events as follows:
A: Student is not wearing polished shoes.
B: Student is not wearing a school tie.

Given, $P(A) = \dfrac{5}{100}$, $P(B) = \dfrac{10}{100}$

$P(A \cup B) = P(A) + P(B) - P(A \cap B)$
$P(A \cup B) = \dfrac{5}{100} + \dfrac{10}{100} - \left(\dfrac{5}{100} \times \dfrac{10}{100}\right) = \dfrac{5+10}{100} - \left(\dfrac{5}{100} \times \dfrac{10}{100}\right)$
$\implies 0.15 - 0.005 = 0.145$

$X \sim Binomial(5, 0.145)$
Where, $p = \dfrac{5+10}{100} - \left(\dfrac{5}{100} \times \dfrac{10}{100}\right) = 0.15 - 0.005 = 0.145$ and $1-p = 1 - 0.145 = 0.855$
Now,
$P(X = 3) = {}^5C_3(0.145)^3(0.855)^2 = 0.0223$

10. There are x black and y blue pens in a box. A pen is chosen at random, and its colour is noted. If the process repeats independently, n times with replacement, then calculate the expected number of black pens chosen.

(a). $n \times \left(\dfrac{x}{y}\right)$

(b). $n \times \left(\frac{x}{x+y}\right)$

(c). $n \times \left(1 - \frac{x}{y}\right)$

(d). $n \times \left(\frac{y}{x+y}\right)$

Answer: b

Solution:

Let X be defined as number of black pens in n independent trials.
p is the probability of success.
$p = \frac{x}{x+y}$, $1 - p = 1 - \frac{x}{x+y} = \frac{y}{x+y}$
Given, process is repeats n times with replacement.
$E(X) = np = n \times \frac{x}{x+y}$

For example: x=10, y=20 and n=10

Let X be defined as number of black pens in 10 independent trials.
p is the probability of success.
$p = \frac{10}{10+20} = \frac{1}{3}$, $1 - p = 1 - \frac{1}{3} = \frac{2}{3}$
Given, process is repeats 10 times with replacement.
$n = 10$
$E(X) = 10 \times \frac{1}{3} = \frac{10}{3}$