ASSIGNMENT : PROBABILITY

1) A bag contains fifteen balls distinguishable only by their colours ; ten are blue and five are red. I reach into the bag with both hands and pull out two balls (one with each hand) and record their colours.

Given :

A bag containing 15 balls.

Blue = 10

Red = 5

a) What is the random phenomenon ?

The random phenomenon is the act of reaching into the bag with both hands and pulling out two balls and recording their colours.

b) What is the sample space ?

Sample space is all the possible combinations of colors for the 2 balls.

$S = \{(Blue, Blue), (Blue, Red), (Red, Blue), (Red, Red)\}$

c) Express the event that the ball in my left hand is red as a subset of the sample space.

The event that the ball in my left hand is red as a subset of the sample space is

Event $= \{(Red, Blue), (Red, Red)\}$

2) Three unbiased coins are tossed. what is the probability of getting at most 2 heads?

$P(\text{at most 2 heads}) = P(0 \text{ heads}) + P(1 \text{ head}) + P(2 \text{ heads})$

$P(0 \text{ heads}) = P(TTT) = \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{8}$

$P(1 \text{ head}) = P(HTT) + P(THT) + P(TTH)$

$= \left[\frac{1}{2} \times \frac{1}{2} \times \frac{1}{2}\right] + \left[\frac{1}{2} \times \frac{1}{2} \times \frac{1}{2}\right] + \left[\frac{1}{2} \times \frac{1}{2} \times \frac{1}{2}\right] = \frac{3}{8}$

$P(2 \text{ heads}) = P(HHT) \times P(HTH) = P(THH)$

$= \left[\frac{1}{2} \times \frac{1}{2} \times \frac{1}{2}\right] + \left(\frac{1}{2} \times \frac{1}{2} \times \frac{1}{2}\right) + \left[\frac{1}{2} \times \frac{1}{2} \times \frac{1}{2}\right] = \frac{3}{8}$

$\therefore P(\text{at most 2 heads}) = \frac{1}{8} + \frac{3}{8} + \frac{3}{8} = \frac{7}{8}$

3) Throw two dice. A and B events are defined below.
$A = \{\max \text{ is } 2\}$, $B = \{\min \text{ is } 2\}$. Are A and B independent?
Provide answer with explanation.

To determine whether events A and B are independent when throwing 2 dice, need to check if the occurrence of one event affects the probability of the other event.

Two events, A and B, are considered independent if and only if: $P(A \cap B) = P(A) \times P(B)$

$A = \{(1,2), (2,1), (2,2)\}$

$B = \{(2,1), (1,2), (2,2), (2,3), (3,2), (2,4), (4,2), (2,5), (5,2), (2,6), (6,2)\}$

$P(A) = \frac{3 \text{ (Outcomes in A)}}{36 \text{ (Total Outcomes)}} = \frac{1}{12}$

$P(B) = \frac{11 \text{ (outcomes in B)}}{36 \text{ (Total outcomes)}} = \frac{11}{36}$

$P(A \cap B) = \frac{1}{36}$ (which is $(2,2)$)

$P(A) \times P(B) = \frac{1}{12} \times \frac{11}{36} = \frac{11}{432}$

$P(A \cap B) \neq P(A) \times P(B)$

$\therefore$ Events A and B are not independent when throwing 2 dice because the probability of both events occuring together $(P(A \cap B))$ does not equal the product of their individual probabilities $(P(A) \times P(B))$

4) You call 2 Uber and 3 Ola cars. If the time that each takes to reach you is IID, what is the probability that Uber arrives first?

Since the time it takes for each car to reach is independent and identically distributed (IID), need to calculate the probability that the earliest arrival time corresponds to one of the 2 Ubers.

Let U1 and U2 represent the arrival times of the 2 Uber cars, and O1, O2, O3 represent the arrival times of the 3 Ola cars.

There are 5! possible arrival orderings.

P(Both Uber cars first) $= \frac{2}{5} \times \frac{1}{4} = \frac{1}{10}$

P(1 Uber car first, then the other, then other Ola cars)

$= \frac{2}{5} \times \frac{2}{4} \times \frac{3}{3} = \frac{1}{10}$

$\therefore$ P(Uber arrives first) = P(Both Uber cars first) + P(one Uber car first, then other, then other Ola cars)

P (Uber arrives first) $= \frac{1}{10} + \frac{1}{10} = \frac{1}{5}$

5) By using NLP, I can detect spam e-mails in my inbox. Assume that the word 'offer' occurs in 80% of the spam messages in my account. Also, let's assume 'offer' occurs in 10% of my desired emails. If 30% of the received e-mails are considered as a scam, and I will receive a new message which contains 'offer', what is the probability that it is spam?

The probability that an email is spam is,

$$P(S) = 0.3$$

The probability than an email is not spam is,

(desired email) $P(D) = 1 - 0.3 = 0.7$

The probability that the word 'offer' occurs in spam emails is,

$$P(O|S) = 0.8$$

The probability that the word 'offer' occurs in desired emails is,

$$P(O|D) = 0.1$$

By Baye's Theorem,

$$P(S|O) = [P(O|S) \times P(S)] / [P(O|S) \times P(S) + P(O|D) \times P(D)]$$

$$= \frac{0.8 \times 0.3}{[0.8 \times 0.3] + (0.1 \times 0.7)} = \frac{0.24}{0.31} = 0.77$$

∴ The probability that a new message containing the word 'offer' is spam is 0.77.

6) Facebook has a content team that labels pieces of content on the platform as spam or not spam. 90% of them are diligent raters and will correctly label 95% of the time. The remaining 10% are non-diligent raters and will label 50% of the content incorrectly. Assume the pieces of content are labeled independently from one another, for every rater. Given that a piece has been rated as non-spam, what is the probability that it is non-spam?

The probability that a rater is diligent is,

$$P(D) = 0.9$$

The probability that a rater is non-diligent is,

$$P(ND) = 0.1$$

The probability that a diligent rater labels content as spam correctly is,

$$P(S|D) = 0.95$$

The probability that a non-diligent rater labels content as non-spam incorrectly is,

$$P(NS|ND) = 0.5$$

The probability that a non-diligent rater labels content as spam incorrectly is,

$$P(S|ND) = 0.5$$

By Bayes Theorem,

$$P(NS|NS) = \frac{[P(NS) \times P(NS|D)]}{[P(NS) \times P(NS|D)] + [P(NS) \times P(NS|ND)]}$$

$$= \frac{P(NS|D)}{P(NS|D) + P(NS|ND)}$$

$$= \frac{0.95}{0.95 + 0.5} = \frac{0.95}{1.45} = 0.66$$

∴ The probability that a piece of content, rated as non-spam is 0.66.

7) A salesperson from an automobile firm XYZ believes that the probability of making a sale is 38%. If he talks to 5 customers on a particular day, what is the probability that he will make exactly 2 sales? (Please assume independence).

The probability of making a sale (success) is

$$p = 0.38$$

The number of attempts is $n = 5$.

By binomial probability formula,

$$P(X=x) = C(n,x) \times p^x \times (1-p)^{(n-x)}$$

$$P(X=2) = C(5,2) \times (0.38)^2 \times (1-0.38)^3$$

$$= \frac{5!}{2!\,(5-2)!} \times 0.1444 \times 0.3294$$

$$= 10 \times 0.1444 \times 0.3294$$

$$= 0.476$$

∴ The probability that the salesperson will make exactly 2 sales when talking to 5 customers on a particular day is 0.476.

8) A machine produces items on which 1% at random are defective. How many items can be packed in a box while keeping the chance of one or more defectives in the box to be no more than 0.5? What are the expected value and standard deviation of the number of defectives in a box of that size?

The probability of an item being defective is $p = 0.01$

To find the number of items (n) that can be packed in a box while ensuring that the probability of one or more defectives in the box is no more than 0.5,

$$P(x \geq 1) \leq 0.5$$

where x is the no. of defective items in the box.

$$P(x = 0) = (1-p)^n$$

$$(1-p)^n > 0.5 \Rightarrow 1 - P(x=0) \leq 0.5$$

(or)

$$1 - (1-p)^n \leq 0.5$$

Simplifying the inequality,

$$\frac{1}{2} \leq (1-p)^n$$

Taking log on both sides,

$$\log\left(\frac{1}{2}\right) \leq \log\left[(1-p)^n\right]$$

$$\log\left(\frac{1}{2}\right) \leq n \times \log(1-p)$$

$$\therefore n \geq \frac{\log(1/2)}{\log(1-p)} \Rightarrow n \geq \frac{\log(1/2)}{\log(1-0.01)}$$

$$n \geq \frac{\log(1/2)}{\log(0.99)}$$

$$n \geq \frac{(-0.3010)}{(-0.0044)}$$

$$n \geq 68.5$$

$\therefore$ can pack a maximum of 69 items in a box to ensure that the probability of one or more defectives in the box is no more than 0.5.

Expected value (Mean) is,

$$E(x) = n \times p$$

$$= 69 \times 0.01$$

$$\boxed{E(x) = 0.69}$$

Standard deviation is,

$$\sigma = \sqrt{n \times p \times (1-p)}$$

$$= \sqrt{69 \times 0.01 \times 0.99}$$

$$\boxed{\sigma = 0.296}$$