Probability

Probability Vs Statistics: Probability is a mathematical method used for statistical analysis.

Probability & statistics is the backbone of machine learning

Probability: Numerical description of how likely an event is to occur

Probability (P) of event (E) to happen: (E.g. Coin, Dice, Card deck) [*0 <= Probability(Event) <= 1*]

$$P(E) = \frac{\text{Number of favorable outcomes}}{\text{Total number of outcomes}}$$

Theoretical Probability

www.hbpatel.in

Q1: While rolling a fair dice, what is a probability of getting 1 or 5?

A1: 2/6 or 1/3

Q2: While rolling a fair dice, what is a probability of getting 2 and 4?

A2: 0

Q3: While rolling a fair dice, what is a probability of getting even number?

A3: 3/6 or 1/2

Q4: Find the probability of pulling a blue ball from a bag of 3 yellow, 4 green, 3 red and 2 blue balls?

A4: 2/12 or 1/6

Q5: Find the probability of pulling anon- blue ball from a bag of 3 yellow, 4 green, 3 red and 2 blue balls?

A5: 10/12 or 5/6

Q6: If a number is chosen randomly from the following list, what is the probability that the number is multiple of 5? [34, 26, 55, 20, 26, 10, 75, 80, 2, 60, 12, 95]

A6: 7/12

Q7: The circumference of a circle C1 is 32∏. A small circle C2 is contained inside C1 with area 12∏. A point P is selected at random from large circle C1. What is the probability that the point P also lies inside small circle P2 also?

A7: 3/64

Probability

If a coin is tosses 10 times and heads is recorded 6 times then the experimental probability of heads is 6/10 (or 3/5)

Sample Space: It is the collection of all possible outcomes (in a random experiments).

Independent Event: The occurrence of one event has no effect on the probability of occurrence of another event.
$P(A \cap B) = P(A).P(B)$

Q: A poll finds that 72% of the Indians consider themselves as Cricket fans. If you randomly pick two people from the population: What is the probability
(a) that first person is a cricket fan and the second as well?
(b) that the first person is and second person isn't?

A: (a) 0.72 x 0.72 (b) 0.72 x 0.28

Probability

Profession	Trouser Color			
	Blue	Black	Brown	Total
Software Programmer	35	25	20	80
Project Manager	7	8	5	20
Total	42	33	25	100

Q1: If an employee is selected at random, what is the probability that s/he is a software programmer?	**A1**: 80 / 100
Q2: If an employee is selected at random, what is the probability that s/he is wearing a blue trouser?	**A2**: 42 / 100
Q2: If an employee is selected at random, what is the probability that s/he is a software programmer and wearing a black trouser?	**A3**: (80/100) x (33/100)

Probability

A survey conducted by a bank revealed that 40% of the accounts are savings accounts and 35% of the accounts are current accounts; and the balance are loan accounts.

1. What is the probability that an account taken at random is a loan account?
2. What is the probability that an account taken at random is not a savings account?
3. What is the probability that an account taken at random is not a current account?
4. What is the probability that an account taken at random is a current account or a loan account?

A1: [(100-40-35) / 100] = 25 /100

A2: [(100-40) / 100] = 60 /100

A3: [(100-35) / 100] = 65 /100

A4: [(35 + 25) / 100] = 60 /100

Conditional Probability

Probability of a card being HEART from a deck of 52 cards? 13 / 52

Probability of second card being HEART if the first card was HEART? 12 / 51

Three persons A, B and C are competing for the post of CEO of a company. The chances of they becoming CEO are 0.2, 0.3 and 0.4 respectively.

The chances of they taking employees' beneficial decisions are 0.5, 0.45 and 0.6 respectively.

What are the chances of having employees' beneficial decisions after having new CEO?

P(A) = 0.2
P(B) = 0.3
P(C) = 0.4

P(Decision | A) = 0.5
P(Decision | B) = 0.45
P(Decision | C) = 0.6

P(Decision) = P (Decision ∩ A) + P (Decision ∩ B) + P (Decision ∩ C)

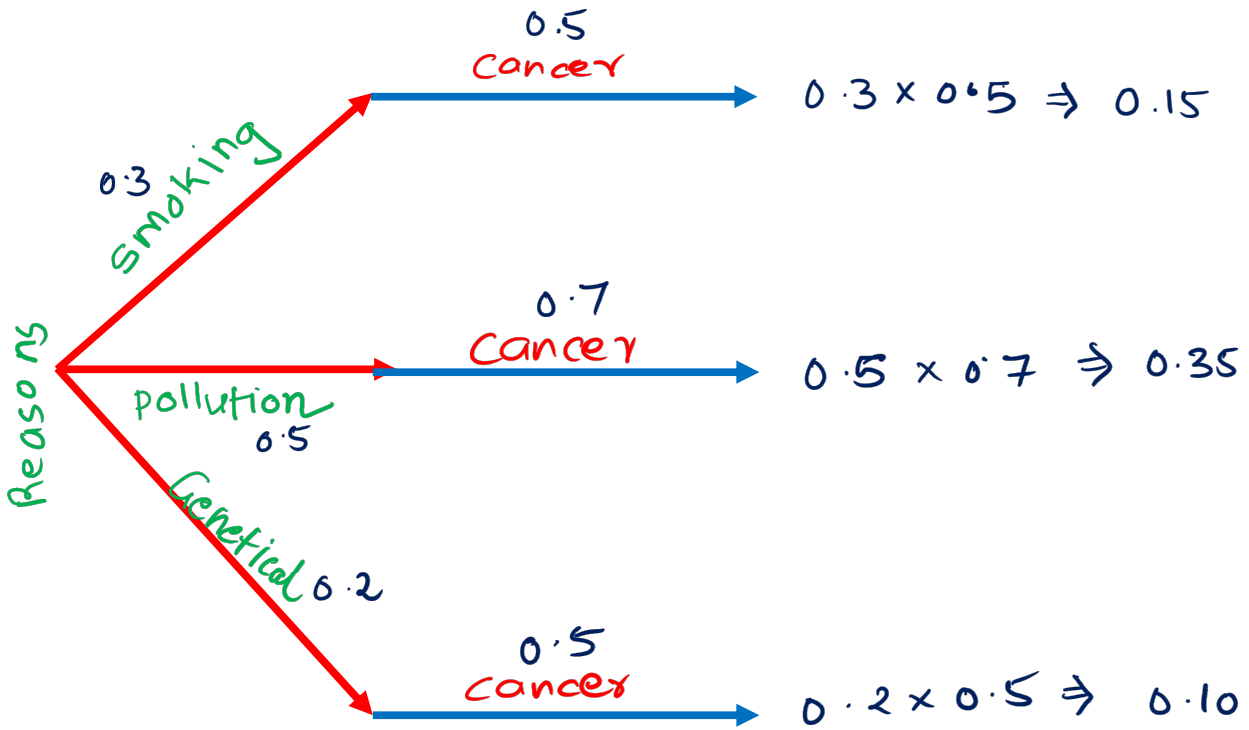
P(Decision) =
P (A) * P (Decision | A) +
P(B) * P (Decision | B) +
P(C) * P (Decision | C)

P(Decision) =
[0.2 x 0.5] + [0.3 x 0.45] + [0.4 x 0.6]

P(Decision) = [0.10] + [0.135] + [0.24]

P(Decision) = [0.475]

Conditional Probability



0.15 + 0.35 + 0.10

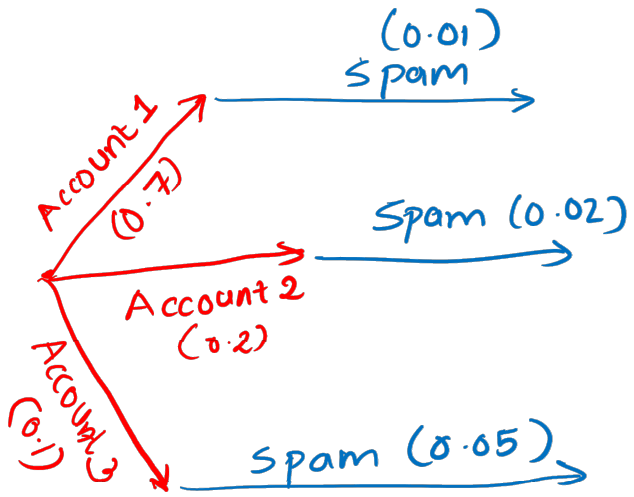
= 0.6 Answer

Conditional Probability

An individual has 3 different email accounts. Most of her messages , in fact 70% come into account #1,whereas 20% come into account#2 and the remaining 10% into account #3.

Of the messages into account #1,only 1% are spam , whereas the corresponding percentages for accounts #2 and #3 are 2% and 5% respectively.

What is the probability that a randomly selected message is a spam?



$0.7 \times 0.01 = 0.007$

$0.2 \times 0.02 = 0.004$

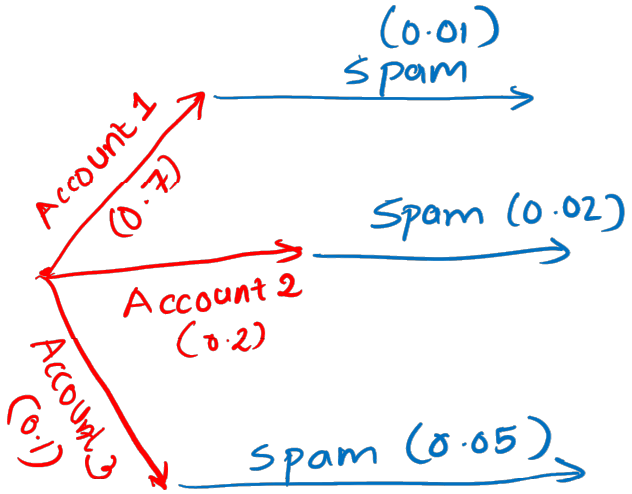
$0.1 \times 0.05 = 0.005$

$0.007 + 0.004 + 0.005$

0.016 (Answer)

Conditional Probability

What is the probability that a randomly selected message is a spam is from Account #2?



$0.7 \times 0.01 = 0.007$

$0.2 \times 0.02 = 0.004$

$0.1 \times 0.05 = 0.005$

$0.007 + 0.004 + 0.005 = 0.016$ (Answer)

$$\frac{P(\text{Email being spam from Account \#2})}{P(\text{Email being spam from all accounts})} = \frac{0.004}{0.016} = 0.25$$

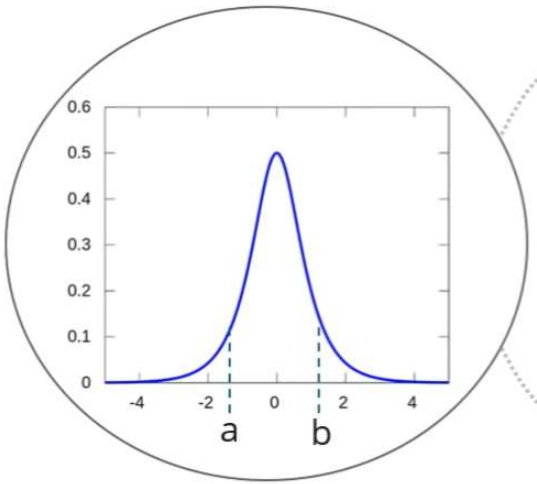
Probability Distribution

Probability Density Function (PDF): The equation describing a continuous probability distribution (between range A and range B)

Normal Distribution: It is a probability distribution that associates the normal random variable X with cumulative probability

Central Limit Theorem: It states that the sampling distribution of the mean of any independent, random variable will be normal and nearly normal, if the sample size is large enough

Probability Distribution Function (PDF)



Property 01

Graph of a PDF will be continuous over a range

Property 02

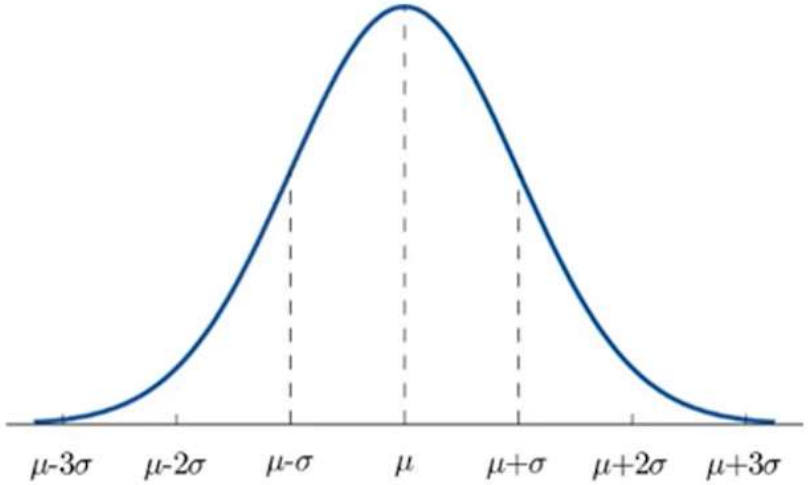
Area bounded by the curve of density function and the x-axis is equal to 1

Property 03

Probability that a random variable assumes a value between a & b is equal to the area under the PDF bounded by a & b

Image Source: www.edureka.co/data-science

Normal Distribution



$$Y = [\ 1/\sigma * \text{sqrt}(2\pi)\ ] * e^{-(x - \mu)2/2\sigma2}$$

Where,

- *X* is a normal random variable
- μ is the mean and
- σ is the standard deviation

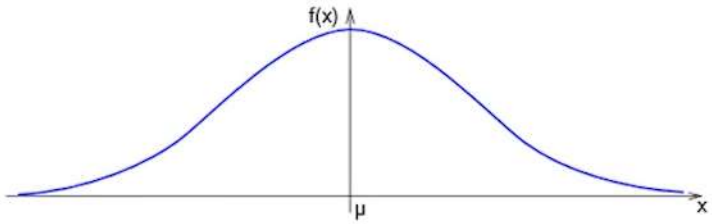
Note: Normal Random variable is variable with mean at 0 and variance equal to 1

Image Source: www.edureka.co/data-science

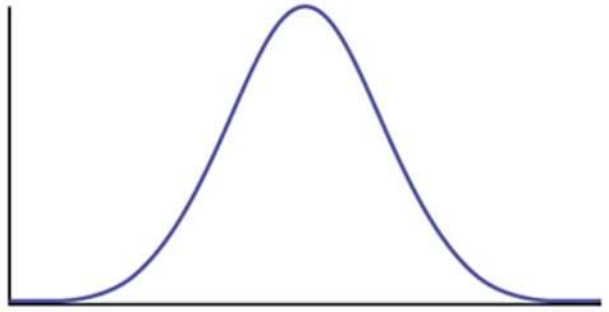
Standard Deviation and Curve

The graph of the Normal Distribution depends on two factors: the *Mean* and the *Standard Deviation*

- **Mean:** *Determines the location of center of the graph*
- **Standard Deviation:** *Determines the height of the graph*



If the standard deviation is large, the curve is short and wide.



If the standard deviation is small, the curve is tall and narrow.

Image Source: www.edureka.co/data-science

Central Limit Theorem

The ***Central Limit Theorem*** states that the sampling distribution of the mean of any independent, random variable will be normal or nearly normal, if the sample size is large enough

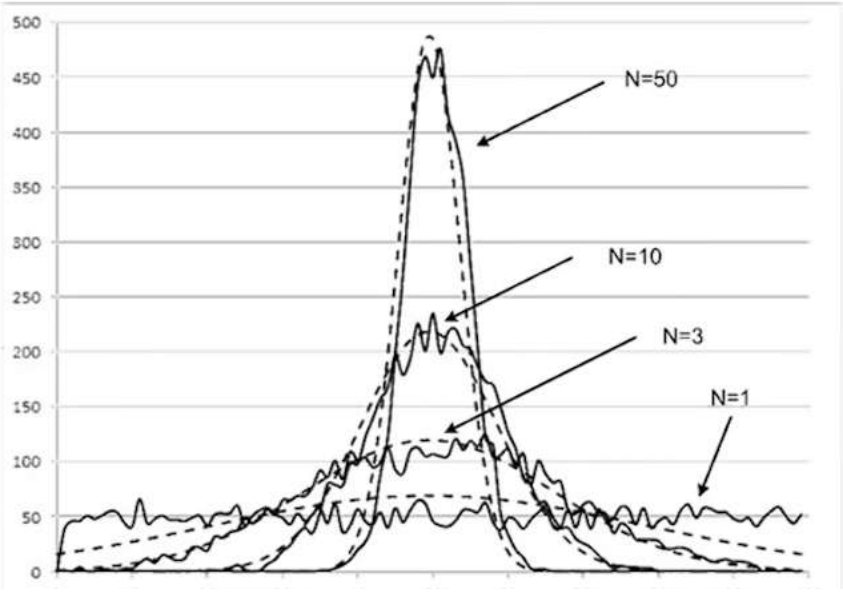
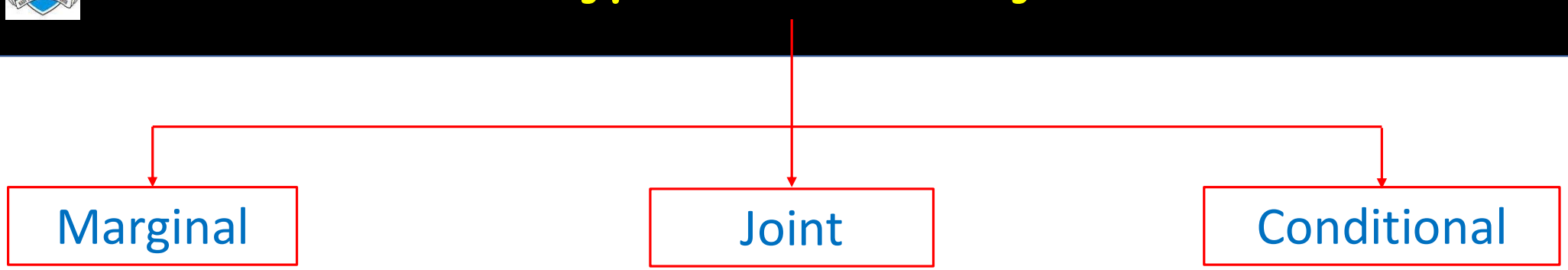


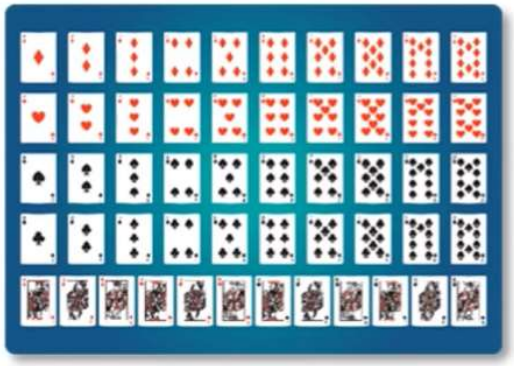
Image Source: www.edureka.co/data-science

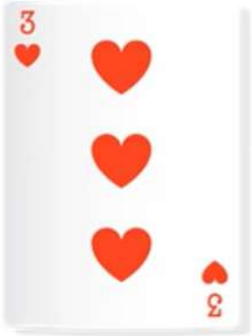
Types of Probability



Marginal Probability

Marginal Probability is the probability of occurrence of a single event.





Marginal Probability = $\frac{13}{52}$

It can be expressed as: P (A) = $\sum_{i=1}^{k} P(x_i)$

Image Source: www.edureka.co/data-science

Joint Probability

Joint Probability is a measure of two events happening at the same time

Example: The probability that a card is an Ace of hearts = P (Ace of hearts)
(There are 13 heart cards in a deck of 52 and out of them one in the Ace of hearts)

Image Source: www.edureka.co/data-science

Conditional Probability

- *Probability of an event or outcome based on the occurrence of a previous event or outcome*
- *Conditional Probability of an event B is the probability that the event will occur given that an event A has already occurred*

If A and B are dependent events then the expression for conditional probability is given by:

$$P(B|A) = P(A \text{ and } B) \; / \; P(A)$$

If A and B are independent events then the expression for conditional probability is given by:

$$P(B|A) = P(B)$$

Image Source: www.edureka.co/data-science

Types of Probability: Use Case

Performance	Training		Total
	Without Training	With Training	
Very Poor	5	0	5
Poor	10	0	10
Average	40	10	50
Good	5	30	35
Excellent	0	5	5
Total	60	45	105

Q 1: Find the probability that a candidate has undergone a training.

Type of probability?	Marginal
Probability?	45/(45+60)=0.42

Q 2: Find the probability that the candidate has undergone a training and ***also*** has a good performance.

Type of probability?	Joint	
Probability?	30/(45+60)=0.28	(45/105) x (30/45)=0.28

Q 3: Find the probability that the candidate has a good performance ***given*** he has not undergone training.

Type of probability?	Conditional
Probability?	5/60=0.08

Image Source: www.edureka.co/data-science