

# Random variable

## What is RV ?

Random Variable is a variable that holds the value of experiment. Consider the coin toss experiment,

Lets say our RV is number of heads, so RV can be 0 or 1. In case of dice, when RV is the number on dice, it can be  $\{1,2,3,4,5,6\}$ . The RV can be anything, like sum of number on dice by rolling it 5 times. So it can range from  $\{5 \text{ to } 30\}$ . It is generally represented by X.

Hence the RV is a number to which each outcome of a random experiment is associated and the RV may vary with the different outcomes of a random experiment. The RV can be numerical as well as categorical.

## Probability Distribution of RV :

As per name, it is the probability distribution of RV. Lets say in single coin toss, RV is number of head. So, the probability distribution is,

X	0	1
P(X)	1/2	1/2

The probability distribution function is the function that maps the random variable to its probability. There are two types of probability distribution function.

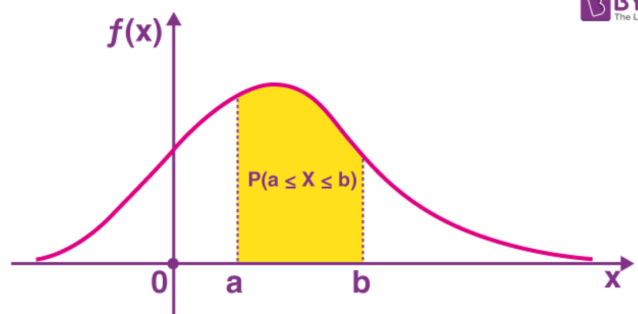
1. Probability density function (PDF) : it is for continuous random variable like mass.
2. Probability mass function (PMF) : it is for discrete random variable like number of people.

## Probability Density Function :

The PDF gives the probability over a continuous range of variable. Like probability of mass between  $[70,80]$  Kg. PDF is also the area under the curve.

The PDF can be,  
 $P(x) = x$  for  $0 < x < 1$   
 $P(x) = 2-x$  for  $1 < x < 2$   
 $P(x) = 0$  for  $x > 2$

$$P(a < x < b) \text{ or } P(a \leq x \leq b) = \int_a^b f(x) dx$$



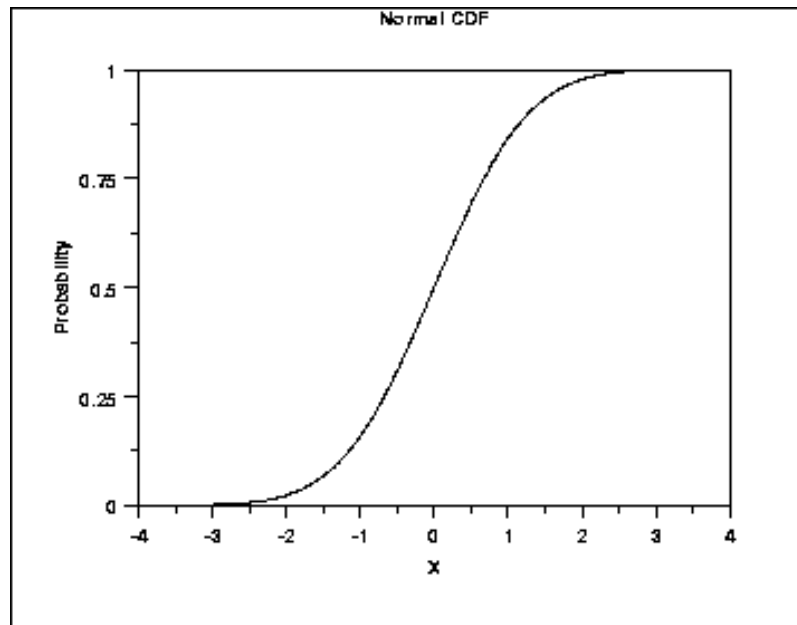
## Cumulative Distribution Function :

The CDF of a random variable  $x$  is defined as the probability that the random variable  $y$  will take value less than or equal to  $x$ . Basically, CDF of heads = 3 means probability that heads  $\leq 3$ . For discrete distribution functions, CDF gives the probability values till what we specify and for continuous distribution functions, it gives the area under the probability density function up to the given value specified.

$$F(x) = \int_{-\infty}^b f(x) dx$$

(Here  $F$  and  $f$  are different functions)

The shape of the curve is not necessary to be sigmoid (S-shaped), it can be of any shape based on the probability distribution. The S-shaped curve is for bell shaped curve, which are symmetric around the centre.



## Mean of RV :

If  $X$  is a random variable, and its possible values are  $x_1, x_2, x_3, \dots, x_n$  associated with the probabilities  $p_1, p_2, p_3, \dots, p_n$ , respectively, then the mean of the random variable  $X$  is given by the formula:

$$E(X) = \sum_x x \cdot P(X = x)$$

The mean of the random variable ( $\mu$ ) is also called the expectation of the random variable  $E(X)$ . It is the expected or the average value of random variable. If we do an experiment then we can expect random variable to be  $E(X)$ .

## Variance of RV :

$$\sigma_x^2 = \sum_{i=1}^n (x_i - \mu)^2 p(x_i)$$

Or

$$\sigma_x^2 = E(X - \mu)^2$$

Or

$$\sigma_x^2 = E(X^2) - [E(X)]^2$$

Where  $E(X) = \mu$ . The variance tells us that how much RV is likely to vary from mean. Like it can give the risk in stock market data that how much stock price can vary from mean.