UNIT- 4
PROBABILITY DISTRIBUTIONS

Let ‘S‘ be a sample space of a random experiment. Suppose to each element s of ‘S’ ,a unique real number X is associated according to some rule.

Then X, is called a random variable on S.

Example:

Consider a random experiment of tossing three coins together.

The corresponding sample space is
S = {HHH,HHT,HTH,THH,HTT,THT,TTH,TTT}
which has 8 possible outcomes.

Suppose we define the mapping f: S→R by f(s)= number of heads in an outcome s i.e.,

f(HHH)=3,f(HHT)=2,f(HTH)=2,f(THH)=2,f(HTT)=1,f(TTH)=1,f(TTT)=0

As s varies over the set S ,X varies over the set {0,1,2,3} belongs to R.

Note: One can define infinitely many random variable on a given sample space.

Discrete Random Variables:

A random variable which can take some specified values only is called as Discrete Random Variables.
(Varying only over integral values)

Ex: Tossing a coin and observing the number of heads turning up.

Continuous Random Variables:

A random variable which can take any value in a specified range is called Continuous Random Variable.

(can assume any value in the interval of real numbers)

Example: Speed ,time etc…..

Discrete Probability Distributions:

If for each value x_i of a discrete random variable X, a real number p(x_i) is assigned such that a)p(x_i)≥ 0

b)$\sum_i p(x_i) = 1$

Then the function p(x) is called Probabilty Function

The set of values $[x_i, p(x_i)]$ is called a discrete probability distribution of discrete random variable X.
The function p(x) is called the probability density function(pdf).

The distribution function f(x) is defined by $f(x) = P(X \leq x) = \sum_{i=1}^{x} p(x_i)$, x being an integer is called the cumulative distribution function(cdf).

Note:

$$Mean(\mu) = \sum_i x_i . p(x_i)$$

$$Variance\ (V) = \sum_i (x_i - \mu)^2 . p(x_i)$$

$$= \sum x_i{}^2 . p(x_i) - \mu^2$$

$$Standard\ deviation(\sigma) = \sqrt{V}$$

Discrete Probability Distribution

Binomial distribution:

It is concerned with trails of repetitive nature in which only the occurrence or non-occurrence, success or failure, acceptance or rejection ,yes or no of a particular event is of interest.

It is a discrete distribution which satisfy the following conditions:

i) Each trail has two mutually exclusive possible outcomes.
ii) Each trail is independent of the other trails.
iii) The probability of success p or failure q remains constant from trail to trail.
iv) The trails are performed under the same conditions for fixed number of times ,say ‘n’.

If a series of independent trails are performed such that for each trail, p is the probability of success and q is the probability of failure, then the probability of x successes in a series of n trails is given by,

$nC_x p^x q^{n-x}$, where x=0,1,2,3……n

That is., p(x)= $nC_x p^x q^{n-x}$

Mean, Standard deviation and Variance of Binomial Distributions:

$$Mean(\mu) = np$$
$$Variance(V) = npq$$
$$Standard\ deviation(\sigma) = \sqrt{npq}$$

Poission Distribution

It is a distribution related to the probability of events which are extremely rare, but which have a large number of independent oppurtunities for occurance.

This can be derived as a limiting case of the Binomial Distribution by making n very large and p very small keeping np fixed.

Poission Probability Function:

$$p(x) = \frac{m^x e^{-m}}{x!}$$ for x=0,1,2 ,3………

Mean and Standard Deviation:

$$Mean(\mu) = m$$
$$Variance(V) = m$$
$$Standard\ deviation(\sigma) = \sqrt{m}$$

Continuous Probability Distribution:

Definition:

If for every 'x' belonging to the range of continuous random variable X, we assign a real number f(x) satisfying the conditions

$$a)\ f(x) \geq 0 \qquad b)\ \int_{-\infty}^{\infty} f(x)dx = 1$$

Then f(x) is called a continuous probability function or probability density function.

If (a, b) is a sub interval of the range space of X then the probability that ‘x’ lies in (a,b) is defined to be a integral of f(x) between a and b.

$$that\ is.,P(a \leq x \leq b) = \int_a^b f(x)\, dx$$

Cumulative Distribution Function:

If X is a continuous random variable with pdf f(x) then the function F(x) defined by $\boldsymbol{F(x) = P(X \leq x) = \int_{-\infty}^{x} f(x)dx}$ is called cumulative distribution function (c d f) of X.

NOTE: $F(x) = P(X \leq x) = P(-\infty \leq x \leq x)$

and $\frac{d[F(x)]}{dx} = f(x)$

NOTE:

1) $\int_{-\infty}^{\infty} f(x)dx = 1$; geometrically means that the area bounded by the curve f(x) and x axis is equal to unity.

2) $P(a \leq x \leq b)$ $is\ equal\ to\ area\ of\ the\ region\ bounded\ by\ the\ curve\ f(x), the\ x - axis\ and\ the\ co - ordinates\ x = a\ and\ x = b.$

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

4) If r is any real no. then,

$$P(x \geq r) = \int_r^\infty f(x)dx;$$

$$P(x < r) = 1 - P(x \geq r)$$

Mean and Variance:

Mean(μ) $= \int_{-\infty}^{\infty} xf(x)dx$

Variance(v)=

$$\int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx = \int_{-\infty}^{\infty} x^2 f(x)dx - \mu^2$$

Exponential Distribution

Probability density function is

$$f(x) = \begin{cases} \alpha e^{-\alpha x} & \text{for } x > 0 \\ 0 & \text{otherwise, where } \alpha > 0. \end{cases}$$

Mean & S.D :

$$\text{Mean } (\mu) = \int_{-\infty}^{\infty} x f(x) dx = \int_{-\infty}^{\infty} x \cdot \alpha e^{-\alpha x} dx.$$

Using Bernoulli's rule

$$\mu = \alpha \left[x \cdot \left(\frac{e^{-\alpha x}}{-\alpha} \right) - 1 \left(\frac{e^{-\alpha x}}{\alpha^2} \right) \right]_0^{\infty}$$

$$= \alpha \left[0 \cdot \frac{1}{\alpha^2} (0 - 1) \right] = \frac{1}{\alpha}$$

$$\therefore \mu = \frac{1}{\alpha}.$$

$$\text{Var} = \int_{-\infty}^{\infty} (x-\mu)^2 f(x) dx$$

$$= \int_{0}^{\infty} (x-\mu)^2 x \cdot \alpha e^{-\alpha x} dx$$

$$= \alpha \int_0^{\infty} (x-\mu)^2 e^{-\alpha x} dx$$

Applying Bernoulli's rule

$$= \alpha \left[(x-\mu)^2 \left[\frac{e^{-\alpha x}}{-\alpha} \right] - 2(x-\mu) \cdot \left[\frac{e^{-\alpha x}}{\alpha^2} \right] + 2 \left[\frac{e^{-\alpha x}}{-\alpha^3} \right] \right]_0^{\infty}$$

$$= \alpha \left[-\frac{1}{\alpha} (0-\mu)^2 - \frac{2}{\alpha^2} (0 - (-\mu)) - \frac{2}{\alpha^3} (0 - 1) \right]$$

$$= \alpha \left\{ \frac{\mu^2}{\alpha} - \frac{2\mu}{\alpha^2} + \frac{2}{\alpha^3} \right\} \qquad \text{But } \mu = 1/\alpha$$

$$\sigma^2 = \alpha \left\{ \frac{1}{\alpha^3} - \frac{2}{\alpha^3} + \frac{2}{\alpha^3} \right\} = \frac{1}{\alpha^2}$$

$$\sigma = \frac{1}{\alpha}$$

Mean $(\mu) = \frac{1}{\alpha}$; S.D $(\sigma) = \frac{1}{\alpha}$.

Normal Distribution:

The pdf $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{(x-\mu)^2}{2\sigma^2}}$

Where variable x assumes all the values from -∞ to ∞;

μ *and* σ are called the parameters of the equation, respectively the mean and the standard deviation of the distribution.

Basic Properties:

a) The area under the normal curve is1
b) The normal distribution is symmetric about the mean.
c) The graph of the normal distribution is called the normal curve. It is bell shaped and symmetric about the mean. The two tails of the curve extend to -∞ and ∞ towards positive and negative directions of x-axis and gradually approach the x-axis without ever meeting it.

Mean and Standard Deviation of Normal Distribution:

$$Mean(\mu) = \mu$$

That is., the mean of the normal distribution is equal to the mean of the given distribution.

$$Variance(V) = \sigma^2$$

That is., the Variance of the normal distribution is equal to the Variance of the given distribution.

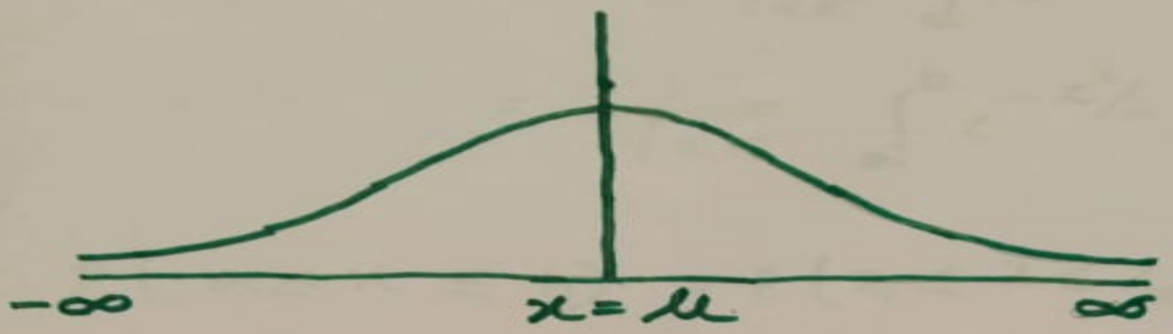
Standard Normal Distribution:

The Normal Distribution for which the mean is zero and the standard deviation is 1 is called the Standard Normal Distribution.

The density function for the standard normal distribution is.,

$$\emptyset(z) = \frac{1}{\sqrt{2\pi}} e^{\frac{z^2}{2}}$$

Normal curve



The graph of the prob. fun^n $f(x)$ is bell shaped curve symmetrical about the line $x = \mu$ and is called normal probability curve.

Standard Normal Distribution

we have
$$P(a \le x \le b) = \int_a^b f(x)\,dx$$
For normal dist^{ly} we have
$$P(a \le x \le b) = \frac{1}{\sigma\sqrt{2\pi}} \int_a^b e^{-(x-\mu)^2/2\sigma^2}\,dx \quad \text{—— (1)}$$
Put $z = \frac{x-\mu}{\sigma}$ or $x = \mu + \sigma z$ then $dx = \sigma\,dz$

Let $z_1 = \frac{a-\mu}{\sigma}$ & $z_2 = \frac{b-\mu}{\sigma}$ be the values of z corresponding to $x = a$ & $x = b$ then

(1) becomes

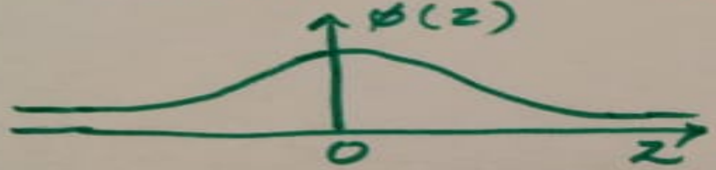
$$P(a \le x \le b) = \frac{1}{\sigma\sqrt{2\pi}} \int_a^b e^{-z^2/2} \cdot \sigma \, dz$$

$$= \frac{1}{\sqrt{2\pi}} \int_a^b e^{-z^2/2} \cdot dz$$

$$\therefore P(a \le x \le b) = P(z_1 \le z \le z_2) = \frac{1}{\sqrt{2\pi}} \int_{z_1}^{z_2} e^{-z^2/2} dz.$$
$$= F(z)$$

$\therefore$ $z = \frac{x-\mu}{\sigma}$ is known as the standard normal variate (SNV) with $\mu = 0, \sigma = 1$.

The $F(z)$ is SNV which is symmetrical about the line $z = 0$



The integral in RHS represent the area bounded by $z = z_1$ and $z = z_2$. Further if $z_1 = 0$ we have

$$\phi(z) = \frac{1}{\sqrt{2\pi}} \int_0^z e^{-z^2/2} dz$$ represent the area from $z = 0$ to z