CHAPTER – 6
PROBABILITY DISTRIBUTIONS

Introduction

While dealing with random variables and their probabilities it is often found that there exists a functional relationship between the value taken by the random variable and the corresponding probability. This initiates to express the relation between random variables and their probabilities whit the help of mathematical functions. These functions are called as probability distributions. Depending on the nature of the random variable distributions can de either discrete or continuous. If the random variable X takes discrete values only, then its probability distribution is called a discrete probability distribution or probability mass function (*pmf*). However if the random variable X, is such that it can take any value within a given interval them the corresponding distribution is called as continuous probability distribution or probability density function (*pdf*). Binomial distribution, Poisson distribution, geometric distribution and negative binomial distribution are some examples of discrete random variable. Examples of continuous distribution are normal distribution, beta distribution, gamma distribution etc.

Bernoullian Trial

A particular trial having only two outcomes either success or failure in which the probability of success is constant is called as Bernoullian trial. The sample space of a Bernoullian trial has only two sample points, S = {success, failure}. Some examples of Bernoullian trials are:

(a) A toss of a single coin (head or tail)
(b) The throw of a die (even or odd)
(c) Result of a student in an examination (pass or fail)
(d) The selection of items produced in an industry (defective or non–defective)

Bernoulli's Distribution

Let us define a random variable X, which represents the result of a Bernoullian trial. Thus, X takes only two values.

Let, $X = 1$, if the result of the trial is a success.
$= 0$, if the trial results to failure.

Let p be the probability of success. So, we have, X takes the value 1 with probability p and 0 with probability q. Accordingly, we have,

X	0	1
$P(X = x)$	q	p

This can be written functionally as

$P(X = x) = p^x (1-p)^{1-x}, x = 0, 1$ which is called the Bernoulli's distribution. This is also termed as the point Binomial distribution.

Constants of Bernoulli's Distribution

Let us now calculate the mean and variance of the distribution.

We know, Mean = $E(X) = \sum_{i=0}^{1} xP(X = x) = \sum_{i=0}^{1} xp^x (1-p)^{1-x} = 0 \times p^0 (1-p)^{1-0} + 1 \times p^1 (1-p)^{1-1}$
$= p$

Now, $E(X^2) = \sum_{i=0}^{1} x^2 P(X = x) = \sum_{i=0}^{1} x^2 p^x (1-p)^{1-x} = 1 \times p^1 (1-p)^{1-1} = p$

$V(X) = E(X^2) - [E(X)]^2 = p - p^2 = p(1-p)$

Thus, Mean = p and Variance = $p(1-p) = pq$, where, $q = 1-p$

Binomial Distribution

Binomial distribution was discovered by James Bernoulli (1654–1705) in the year 1700. But it was published in the year 1713.

Let a random experiment having only two outcomes either success or failure, be performed a number of times (n, say) under identical conditions. Let X be the random variable that represents the number of successes in n trials, with 'p' the probability of success which remains constant for each time the random experiment is performed. Thus, 'q =1-p' is the probability of failure in any trial. Under the above conditions the random variable X is said to follow binomial distribution if its probability mass function is given by

$$P(X=x) = {}^{n}C_{x}\ p^{x}\ q^{n-x},\ x=0,1,\ldots,n.\ \text{and } 0 \le p \le 1,\ q = 1-p$$

where n and p or q are the parameters of the binomial distribution.

The binomial distribution is a discrete distribution as the random variable can take only the integral values i.e. $0,1,2,\ldots,n$. so, the probability function of the binomial distribution is also called as probability mass function (*p.m.f*). To notation used to denote that a random variable X follows binomial distribution with parameters n and p is $X \sim B(n, p)$.

Mean and Variance of Binomial Distribution

Mean = E(X) = $\sum_{x=0}^{n} x\, P[X=x] = \sum_{x=0}^{n} x\, {}^{n}C_{x}\ p^{x}\ q^{n-x}$

$$= 0.q^{n} + 1.{}^{n}C_{1}pq^{n-1} + 2.{}^{n}C_{2}p^{2}q^{n-2} + 3.{}^{n}C_{3}p^{3}q^{n-3} + \ldots + n.p^{n}$$

$$= 1.npq^{n-1} + 2.\frac{n(n-1)}{2}p^{2}q^{n-2} + 3.\frac{n(n-1)(n-2)}{3!}p^{3}q^{n-3} + \ldots + n.p^{n}$$

$$= npq^{n-1} + n(n-1)p^{2}q^{n-2} + \frac{n(n-1)(n-2)}{2!}p^{3}q^{n-3} + \ldots + n.p^{n}$$

$$= np(q^{n-1} + (n-1)pq^{n-2} + \frac{(n-1)(n-2)}{2!}p^{2}q^{n-3} + \ldots + p^{n-1})$$

$$= np(q^{n-1} + {}^{n-1}C_{1}pq^{n-2} + {}^{n-1}C_{2}p^{2}q^{n-3} + \ldots + p^{n-1})$$

$$= np(q+p)^{n-1} = np \qquad (1) \qquad \text{Since, } p + q = 1$$

So, we have

$E(X) = np$

Variance = $E(X^{2}) - (E(X))^{2}$ (2)

Now,

$$E(X^{2}) = \sum_{x=0}^{n} x^{2}P(X=x) = \sum_{x=0}^{n} x^{2}\times {}^{n}C_{x}p^{x}q^{n-x} = \sum_{x=0}^{n}[x(x-1)+x]\,{}^{n}C_{x}p^{x}q^{n-x}$$

$$= \sum_{x=0}^{n} x(x-1)\,{}^{n}C_{x}p^{x}q^{n-x} + \sum_{x=0}^{n} x\,{}^{n}C_{x}p^{x}q^{n-x} = \sum_{x=0}^{n} x(x-1)\frac{n!}{x!(n-x)!}p^{x}q^{n-x} + np \quad \text{[using (1)]}$$

$$= \sum_{x=0}^{n} x(x-1)\frac{n(n-1)\times(n-2)!}{x(x-1)\times(x-2)!(n-x)!}p^{x}q^{n-x} + np$$

$$= n(n-1)p^{2}\sum_{x=2}^{n}\frac{(n-2)!}{(x-2)!(n-x)!}p^{x-2}q^{n-x} + np = n(n-1)p^{2}(p+q)^{2} + np$$

$$= n(n-1)\,p^{2} + np \quad \text{[since, } p+q = 1]$$

Thus, $E(X^2) = n(n-1)\, p^2 + np$ (3)

Replacing the values of (1) and (3) in (2), we have

Variance = $E(X^2) - (E(X))^2$

$$= n(n-1)\, p^2 + np - (np)^2 = n^2p^2 - np^2 + np - n^2p^2 = np(1-p) = npq$$

Thus, mean of binomial distribution is *np* and variance is *npq*.

Recurrence Relation

The recurrence relation of the probabilities of binomial distribution is given by the following expression:

$$P(X = x+1) = \frac{n-x}{x+1} \times \frac{p}{q} \times P(X = x)$$

We know that, the binomial probability is given by

$P(X=x) = {}^nC_x\, p^x\, q^{n-x}$ where $x = 0,1,2,\ldots,n$ and $p + q = 1$

Now,

$$\frac{P(X = x+1)}{P(X = x)} = \frac{{}^nC_{x+1}\, p^{x+1} q^{n-x-1}}{{}^nC_x\, p^x q^{n-x}} = \frac{\frac{n!}{(x+1)!(n-x-1)!}}{\frac{n!}{x!(n-x)!}} \times \frac{p}{q}$$

$$\Rightarrow \frac{P(X = x+1)}{P(X = x)} = \frac{x!(n-x)!}{(x+1)!(n-x-1)!} \times \frac{p}{q} = \frac{x!(n-x)\times(n-x-1)!}{(x+1)\times x!(n-x-1)!} \times \frac{p}{q}$$

$$\Rightarrow \frac{P(X = x+1)}{P(X = x)} = \frac{n-x}{(x+1)} \times \frac{p}{q}$$

Thus,

$$P(X = x+1) = \frac{n-x}{(x+1)} \times \frac{p}{q} \times P(X = x)$$

✍**Note**

The recurrence relation is used to calculate the probability of $P(X=x+1)$ if the value of the parameters of the distribution and $P(X = x)$ is known.

Properties of Binomial Distribution

1. The binomial distribution is a discrete distribution where the random variable *X* takes the values $0,1,2,\ldots,n$
2. The binomial distribution has two parameters n and p or q.
3. The mean of the binomial distribution is *np* and variance is *npq* . So standard deviation is $\sqrt{npq}$
4. The mean is greater than the variance.
5. Skewness and kurtosis of binomial distribution are $(q-p)/\sqrt{npq}$ and $(1-6pq)/npq$ respectively.
6. The binomial distribution may have either one or two modes.
7. The binomial distribution is a symmetrical distribution if $p = q = ½$. Otherwise it is a skewed distribution.
8. The distribution is said to be mesokurtic if $p\times q = 1/6$
9. The binomial distribution may be obtained as a limiting case of Hypergeometric distribution.
10. If *X* follows binomial distribution with parameters n_1 and *p* and *Y* follows binomial distribution with parameters n_2 and *p* , then *X+Y* follows binomial distribution with parameters n_1+n_2 and *p*. This property is known as additive property of binomial distribution.

Applications of Binomial Distribution

Binomial Distribution is used to determine the probability of –

- occurrence of heads or tails when a number of coins are tossed.
- number of males or females in a given population
- number of defective or non–defective items in a production process
- number of successes or failures of a gambler in a particular game which is repeated a number of times.

SOLVED ILLUSTRATIONS (BINOMIAL DISTRIBUTION)

Illustration 1: Five fair coins are tossed. Find the probability of (i) Exactly three heads. (ii) Atleast three heads.

Solution: In a fair coin we have probability of a head to occur = ½
Let X be a random variable that denotes the number of heads that occurs when five coins are tossed.
Here, n = no of coins tossed = 5
So using the binomial distribution we have,

$$P\,[\text{Exactly three heads}] = P[X = 3] = {}^5C_3 \left(\tfrac{1}{2}\right)^3 \left(1 - \tfrac{1}{2}\right)^{5-3}$$

$$= \frac{5\times4}{2} \times \frac{1}{8} \times \left(\frac{1}{2}\right)^2 = 5/16$$

$$P\,[\text{Atleast three heads}] = P[X = 3 \text{ or } X = 4 \text{ or } X = 5]$$

$$= P[X = 3] + P[X = 4] + P[X = 5]$$

$$= {}^5C_3 \left(\tfrac{1}{2}\right)^3 \left(1 - \tfrac{1}{2}\right)^{5-3} + {}^5C_4 \left(\tfrac{1}{2}\right)^4 \left(1 - \tfrac{1}{2}\right)^{5-4} + {}^5C_5 \left(\tfrac{1}{2}\right)^5 \left(1 - \tfrac{1}{2}\right)^{5-5}$$

$$= \frac{5\times4}{2} \times \frac{1}{8} \times \left(\frac{1}{2}\right)^2 + 5 \times \frac{1}{16} \times \frac{1}{2} + \frac{1}{32} = \frac{10}{32} + \frac{5}{32} + \frac{1}{32} = \frac{16}{32} = \frac{1}{2}$$

Illustration 2: The mean of a binomial distribution is 6 and the standard deviation is given by √(3/2). Find the distribution.

Solution: We, know that for a binomial distribution with parameters n and p. We have *mean = np* and Variance = *npq*
Also, Variance = (Standard Deviation)^2 = *3/2*
Thus, np = 6 and npq = 3/2
Now, q = npq/np = (3/2)/6 = ¼
p = 1 – q = 1- ¼ = ¾
Also, np = 6
$\Rightarrow n \times \frac{3}{4} = 6$
$\Rightarrow n = 8$
So, the required distribution is
$P(X=x) = {}^8C_x \left(\tfrac{3}{4}\right)^x \left(\tfrac{1}{4}\right)^{n-x}$ where $x = 0,1,2,\ldots,n$

Illustration 3: A random variable X follows binomial distribution with mean 5/3 and P(X=2) = P (X = 1). Find variance, P(X = at least 1) and P(X = at most 1).

Solution: Let n and p be the parameters of the distribution. Thus, we have
$P(X=x) = {}^nC_x\, p^x q^{n-x}$ where $x = 0,1,2,\ldots,n$ and $p + q = 1$
By the question, we have

$$\text{Mean} = np = 5/3 \qquad (1)$$

And
P(X=2) = P (X = 1)
$\Rightarrow {}^nC_2\, p^2 q^{n-2} = {}^nC_1\, p^1 q^{n-1}$

$$\Rightarrow \frac{n(n-1)}{2!} p^2 q^{n-2} = npq^{n-1}$$

$\Rightarrow (n - 1)\, p = 2\, q$
$\Rightarrow\; n\,p - p = 2\, q$
$\Rightarrow 5/3 - p = 2\, q$ [Using (1)]
$\Rightarrow 5 - 3\,p = 6\, q$
$\Rightarrow 6q + 3\,p = 5$
$\Rightarrow 6\,(1 - p) + 3\,p = 5$
$\Rightarrow -3\,p = -1$
$\Rightarrow\; p = 1/3$ (2)
So, $q = 1 - p = 1 - 1/3 = 2/3$
Putting (2) in (1) we have,
$n = 5$
Now, variance of the distribution = $npq = 5\times(1/3)\times(2/3) = 10/9$
P(X= at least 1) = P(X ≥ 1)
= 1 – P(X ≤ 0) = 1 – P(X = 0) [Since, X cannot be less than zero
= $1 - {}^5C_0\,(1/3)^0\,(2/3)^5 = 1 - (2/3)^5 = 211/243$
$P(X = \text{at most } 1)$ = P(X=0) + P(X=1)
= $(2/3)^5 + {}^5C_1\,(1/3)^1\,(2/3)^4$
= $(32/243) + 5\times(16/243) = (32+80)/243 = 112/253$

Illustration 4: The chances of a bomb hitting a target to that it will not are 3:2. Find the probability that the target will hit atleast once in five shots.

Solution: Let X be a random variable which represents the number of shots required to hit a target out of five shots.
By the question we have, n = 5, p = 3/5 and q = 1– p = 1 – (3/5) = 2/5
Now,
P(The target is hit atleast once in 5 shots) = P (X ≥ 1)
= 1 – P(X ≤ 0) = 1 – P(X = 0) [Since, X cannot be less than zero
= $1 - {}^5C_0\,(3/5)^0\,(2/5)^5 = 1 - (2/5)^5 = 3093/3125$

Poisson Distribution

Poisson distribution was discovered by Simeon Denis Poisson (1781–1840) and it was published in the year 1837. A random variable X is said to follow Poisson distribution if it assumes only non–negative values and if its probability mass function is given by

$$P(X = x) = \frac{e^{-\lambda}\lambda^{x}}{x!}, \; x = 0, 1, 2, \ldots, \infty \text{ and } \lambda > 0$$

Where λ is the parameter of the Poisson distribution.

Deriving Poisson Distribution from Binomial Distribution

The Poisson distribution is a limiting form of the binomial distribution. The conditions under which the binomial distribution tends to Poisson distribution are :

(i) When the number of trials (n) is very large i.e. when $n \to \infty$

and

(ii) when the probability of success in each trial (p) is very small i.e. when $p \to 0$, such that $np\ (= \lambda)$ a reasonable finite quantity.

The derivation of the limiting form is as follows:

Let X be a random variable which follows binomial distribution with parameters n and p, so we have,

$$P(X=x) = {}^{n}C_{x}\, p^{x} q^{n-x}$$

Now,

$$\underset{n\to\infty}{Lt}\, P(X=x) = \underset{n\to\infty}{Lt}\, {}^{n}C_{x}\, p^{x} q^{n-x} = \underset{n\to\infty}{Lt}\, \frac{n(n-1)(n-2)\ldots(n-x+1)}{x!}\, p^{x} q^{n-x}$$

$$= \underset{n\to\infty}{Lt}\, \frac{n^{x}(1-\frac{1}{n})(1-\frac{2}{n})\ldots(1-\frac{x-1}{n})}{x!}\, p^{x}(1-p)^{n-x}$$

$$= \underset{n\to\infty}{Lt}\, \frac{(1-\frac{1}{n})(1-\frac{2}{n})\ldots(1-\frac{x-1}{n})}{x!}(np)^{x}\left(1-\frac{np}{n}\right)^{n-x}$$

$$= \frac{1}{x!}\underset{n\to\infty}{Lt}\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\ldots\left(1-\frac{x-1}{n}\right)\lambda^{(}\left(1-\frac{\lambda}{n}\right)^{n-x}$$

Mean and Variance of Poisson Distribution

$$\text{Mean} = E(X) = \sum_{x=0}^{\infty} xP[X=x] = \sum_{x=0}^{\infty} x\frac{e^{-\lambda}\lambda^{x}}{x!} = 0 + 1\times\frac{e^{-\lambda}\lambda^{1}}{1!} + 2\times\frac{e^{-\lambda}\lambda^{2}}{2!} + 3\times\frac{e^{-\lambda}\lambda^{3}}{3!} + \ldots$$

$$= e^{-\lambda}[\,\lambda + \frac{\lambda^{2}}{1!} + \frac{\lambda^{3}}{2!} + \ldots] = \lambda e^{-\lambda}[\,1 + \frac{\lambda}{1!} + \frac{\lambda^{2}}{2!} + \ldots] = \lambda e^{-\lambda}e^{\lambda} = \lambda \qquad (1)$$

Now,

$$E(X^{2}) = \sum_{x=0}^{\infty} x^{2}P[X=x] = \sum_{x=0}^{\infty} x^{2}\frac{e^{-\lambda}\lambda^{x}}{x!} = \sum_{x=0}^{\infty}[x(x-1)+x]\;\frac{e^{-\lambda}\lambda^{x}}{x!}$$

$$= \sum_{x=0}^{\infty} x(x-1)\frac{e^{-\lambda}\lambda^{x}}{x!} + \sum_{x=0}^{\infty} x\frac{e^{-\lambda}\lambda^{x}}{x!} = \sum_{x=0}^{\infty} x(x-1)\frac{e^{-\lambda}\lambda^{x}}{x(x-1)\times(x-2)!} + \lambda \qquad \text{[Using (1)]}$$

$$= e^{-\lambda}\lambda^2 \sum_{x=2}^{\infty} \frac{\lambda^{x-2}}{(x-2)!} + \lambda = \lambda^2 e^{-\lambda} e^{\lambda} + \lambda = \lambda^2 + \lambda \qquad (2)$$

Thus,

variance = $E(X^2) - [E(X)]^2 = E(X^2) - \lambda^2$ Using (i)

$= \lambda^2 + \lambda - \lambda^2 = \lambda$ Using (ii)

Thus, the mean and variance of a Poisson distribution are equal and their value is equal to λ.

Recurrence relation of the Poisson Distribution

We know that for a Poisson distribution with parameter λ

$x = 0,1,2,\ldots, \infty$ *and* $\lambda > 0$

Thus, $P(X = x+1) = \dfrac{e^{-\lambda}\lambda^{x+1}}{(x+1)!}$

So,

$$\frac{P(X = x+1)}{P(X = x)} = \frac{e^{-\lambda}\lambda^{x+1}}{(x+1)!} \Big/ \frac{e^{-\lambda}\lambda^{x}}{x!} = \frac{\lambda}{x+1}$$

$\Rightarrow P(X = x+1) = \dfrac{\lambda}{x+1} P(X = x)$ gives the required recurrence relation.

Properties of Poisson Distribution

1. The Poisson distribution is a discrete distribution where the random variable X takes the values $0,1,2,\ldots \propto$
2. The Poisson distribution has one parameter i.e., λ.
3. The mean and variance of the Poisson distribution are equal that is, mean = variance = λ .
4. The standard deviation is equal to $\sqrt{\lambda}$.
5. Skewness and kurtosis of Poisson distribution are $1/\sqrt{\lambda}$ and $1/\lambda$ respectively.
6. The Poisson distribution may have either one or two modes.
7. The Poisson distribution is a positively skewed distribution as $1/\sqrt{\lambda}$ is always positive.
8. The distribution is said to be leptokurtic as $1/\lambda$ is always positive.
9. The Poisson distribution may be obtained as a limiting case of Binomial distribution.
10. If X and Y are two independent Poisson variates with parameters λ_1 and λ_2 then X+Y is also a Poisson variate with parameter $\lambda_1 + \lambda_2$.

Applications of Poisson Distribution

Poisson Distribution can be used in the following cases –

- Number of cars passing through a certain point per minute during a busy hour of the day.
- Number of suicides reported per week in a particular town.
- Number of printing mistakes per page of a standard book.
- Number of persons born blind per year in a particular village.

SOLVED ILLUSTRATIONS (POISSON DISTRIBUTION)

Illustration 1: Find the mean of a Poisson distribution such that we have
P(X = 2) = P(X = 1).

Solution: From the recurrence relation of the Poisson distribution we have

$$P(X = x+1) = \frac{\lambda}{x+1} P(X = x)$$

So, putting $x = 1$ in the above expression, we have

$$P(X = 2) = \frac{\lambda}{2} P(X = 1)$$

$\Rightarrow 1 = \lambda/2$ Since, $P(X = 2) = P(X = 1)$

$\Rightarrow \lambda = 2$

Thus, the mean of the Poisson distribution is 2.

Illustration 2: In a Poisson distribution we have the probability that X takes the value 0 is 0.1. Find the mean of the distribution.

Solution: We know that the mean of the Poisson distribution is equal to its parameter λ.
Now,
So, putting $x = 0$ we have

P (X = 0) =

$\Rightarrow 0.1 = e^{-\lambda} \Rightarrow e^{\lambda} = 10 \Rightarrow \lambda = \log_e 10$

$\Rightarrow \lambda = 2.3026$ *(obtained from table)*

Thus the mean of the Poisson Distribution is *2.3026.*

Illustration 3: In a factory manufacturing blades it is found that on an average 2% of the blades are defective. What is the probability that atmost 5 defective blades will be found in a box of 200 blades.

Solution: Here, we have

n = total number of blades = 200

p = probability of a defective blade = 2/100 = 0.02

Thus, $\lambda = n \times p = 200 \times 0.02 = 4$

Let X be the random variable which represents the number of defective blades in a packet of 200 blades.

Thus, by the question we are to find the probability of $X \le 5$.

So, $P(X \le 5)$

$= P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4) + P(X = 5)$

$$= \sum_{x=0}^{5} P(X = x)$$

$$= \sum_{x=0}^{5} \frac{e^{-4} 4^x}{x!} = e^{-4} \sum_{x=0}^{5} \frac{4^x}{x!} = e^{-4}\left[\frac{4^0}{0!} + \frac{4}{1!} + \frac{4^2}{2!} + \frac{4^3}{3!} + \frac{4^4}{4!} + \frac{4^5}{5!}\right]$$

= 0.0183 (1 + 4 + 8 + 10.6667+ 10.6667+ 8.5333)

= 0.785

Thus, required probability is 0.785

Illustration 4: Vijay Lodge, of Guwahati has three rooms only. The number of demands for a room is Poisson distributed variate with mean 1.5. Calculate the proportion of days on which (i) neither room is demanded (ii) some demand for rooms are refused because of non-availability of rooms.

Solution: Let X be the random variable which represents the number of demands made for rooms. So the proportion of days on which there are x demands for rooms is given by,

P[x demands for rooms] = P[X = x] = $\dfrac{e^{-\lambda}\lambda^x}{x!} = \dfrac{e^{-1.5}1.5^x}{x!}$ [since, mean = 1.5]

(i) The proportion of days in which neither room is demanded is

$$P[\ X = 0\ ] = \frac{e^{-1.5}1.5^0}{0!} = e^{-1.5} = 0.223$$

(ii) The proportion of days in which some demands are refused

$$\begin{aligned}
P\ [\ X>3\ ] &= 1 - P[\ X\leq 3\ ] \\
&= 1 - \{P[\ X=0\ ] + P[\ X=1\ ] + P[\ X=2\ ] + P[\ X=3\ ]\ \} \\
&= 1 - \left[\frac{e^{-1.5}}{0!} + \frac{1.5\ e^{-1.5}}{1!} + \frac{1.5^2 e^{-1.5}}{2!} + \frac{1.5^3 e^{-1.5}}{3!}\right] \\
&= 1 - e^{-1.5}\ \{\ 1 + 1.5 + (1.5)^2/2! + (1.5)^3/3!\ \} \\
&= 1 - 0.2231 \times 4.1873 = 1 - 0.9342 = 0.0657
\end{aligned}$$

Normal Distribution

The *Normal distribution* also called as the *Gaussian distribution,* is a continuous probability distribution with two parameters μ and σ and is defined by the probability density function (p.d.f.)

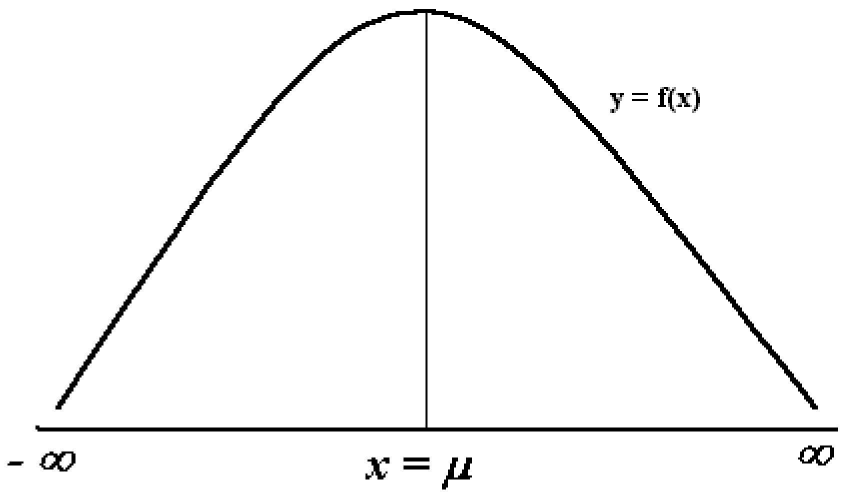
$$P[\, X = x \,] = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

where, $-\infty < x < \infty$ and $-\infty < x < \infty, \sigma > 0$

Here μ is the mean and σ is the standard deviation of the distribution. π and e are two mathematical constants having the approximate values 22/7 and 2.718 respectively.

The history of the distribution is very interesting. It was discovered by an English Mathematician De-Moivre in 1733, used by Laplace, later in the year 1774 but the credit of the distribution was wrongly attributed to Gauss, who first made the reference of the distribution in 1809 to study the errors in the measurement of Astronomy.

This is the most useful distribution in theoretical statistics because of its many important characteristics. Most of the probability distributions of statistics whether discrete or continuous tends to normal distribution especially when the number of observations are large. The probability curve of normal distribution is known as *Normal Curve.* The curve is symmetrical about its mean (μ), bell-shaped and the two tails extend to infinitely on either side.



If a random variable X is normally distributed with mean μ and standard deviation σ, then the random variable $Z = (X-\mu)/\sigma$ is called as the *"standard normal variate"* and the corresponding distribution the standard normal distribution . It has the density function

$$p(\, Z=z \,) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \quad \text{where, } -\infty < z < \infty$$

This is actually a special case of normal distribution with mean 0 and standard deviation 1.

Area under Normal Probability Curve

As in all continuous probability distributions, the total area under the normal curve is 1. For a continuous distribution we cannot calculate the probability at a point but we can calculate probability with in a range. The probability that X lies between c and d, denoted by $P(c \le X \le d)$, is given by the area under the curve between the vertical lines at c and d. This is also equal to the area under *'Standard normal curve'* between the vertical lines at the standardized values of c and d; i.e.

$P(c \le X \le d)$ = Area under 'standard normal curve' between the vertical lines at c′ and d′. Where $c' = (c - \mu)/\sigma$ and $d' = (d - \mu)/\sigma$. Extensive tables showing the areas under standard normal curve are available in this book (see Statistical Tables).

The cumulative distribution function (c.d.f.) of standard normal distribution, viz.

$\Phi(z)$ = Probability that the standard normal variable takes a value less than or equal to z.

= Area under *'standard normal curve'* to the *left* of the ordinate at z.

Mathematically,

$$\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz$$

The values of Φ (z) are given in statistical tables only for positive values of z. For, negative values the relation

$$\Phi(-z) = 1 - \Phi(z)$$

is used to find the area, because of the symmetry of standard normal curve about 0.

Probability Integral of Standard Normal Distribution

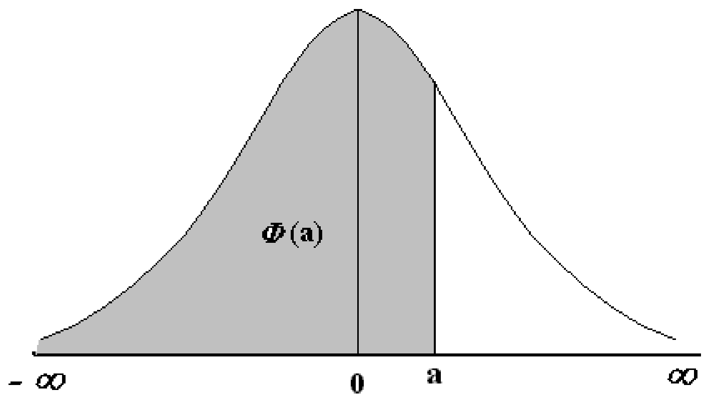
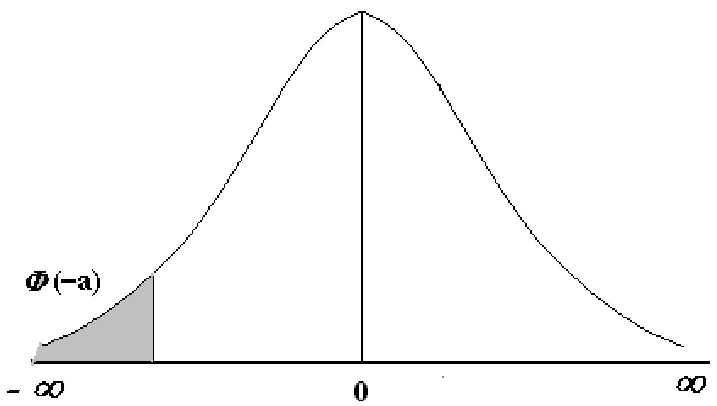


Fig.1: Area between z = $-\infty$ and z = a



(Area between z = $-\infty$ and z = $-$ a)

Thus, probabilities of normal distribution can be calculated by using the relation,

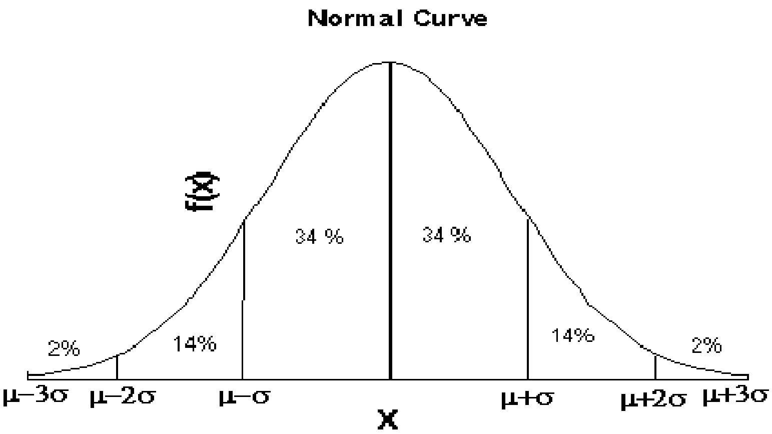
$$P(c \le X \le d) = \Phi(d') - \Phi(c'),$$

where c′ and d′ are the standardized values of *c* and *d*.

Properties of Normal Distribution

The normal distribution has the following properties:

1. The Normal Distribution has two parameters μ the mean of the distribution and σ the standard deviation of the distribution.
2. Normal distribution is a symmetrical distribution with mean = median = mode
3. The quartiles are equidistant from the median, i.e.
   $Q_3 - Q_2 = Q_2 - Q_1$
4. All the odd order moments are equal to 0.
5. The coefficient of skewness $\gamma_1 = 0$ and coefficient of kurtosis $\gamma_2 = 0$.
6. The graph of the normal distribution is called the normal curve. It is a bell shaped curve and is symmetrical about the mean.
7. The linear combination of independent normal variates is also a normal variate.
8. The mean deviation about mean of the normal distribution is $\frac{4}{5}^{th}$ of the standard deviation.
9. For the normal distribution QD: MD: SD : : 10 : 12 : 15
10. The area under the normal curve can be distributed in the following manner:



Importance of Normal Distribution

Normal distribution plays a very important role in statistical theory and its application becomes useful for the following reasons:

1. Most of the distribution occurring in practice, for example binomial, Poisson, hyper−geometric distribution are approximated by Normal Distribution.
2. Even if a variable is not normally distributed it can sometimes be brought to normal form by simple transformation of variable.
3. Many of the distributions of sample statistics tend to normality and as such they can be best studied with the help of normal curve.
4. The proof of all the test of significance in sampling are based upon the fundamental assumption that the population from which the sample has been drawn is normal.
5. The theory of normal curve can be applied to the graduation of the curves, which are not normal.
6. Normal distribution is extensively used in statistical quality control.

✍**Note**

1. If X follows binomial distribution with parameters *n* and *p* then the conditions under which binomial distribution tends to normal distribution are:
   (i) The number of trials (*n*) is infinitely large, i.e. $n \to \infty$
   (ii) Neither p nor q is very small.
2. If X follows Poisson distribution with parameter λ then the conditions under which Poisson distribution tends to normal distribution is that the mean of the distribution i.e. $\lambda \to \infty$.

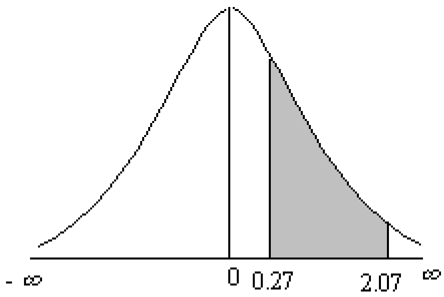
SOLVED ILLUSTRATIONS (NORMAL DISTRIBUTION)

Illustration 1: The mean weight of 500 male students at a certain college is 151 lbs. and the standard deviation is 15 lbs. Assuming that the weights are normally distributed, find how many students weight (i) between 120 and 155 lbs., (ii) more than 155 lbs.
[Given $\Phi(0.27) = 0.6064$ and $\Phi(2.07) = 0.9808$, where $\Phi(t)$ denotes the area under standard normal curve to the left of the ordinate at t.]

Solution: The mean μ and the standard deviation σ are $\mu = 151$ lbs., $\sigma = 15$ lbs.
(i) Proportion of students whose weights lie between 120 & 155 lbs. = Area under standard normal curve between the vertical lines at the standardized values,
viz. $z = (120 - 151)/15 = -2.07$ and $z = (155 - 151)/15 = 0.27$.

$P(120 \le x \le 155)$
$= \Phi(0.27) - \Phi(-2.07)$
$= \Phi(0.27) - [1 - \Phi(2.07)]$
$= 0.6064 - 1 + 0.9808 = 0.5872$ [see Figure].

Thus, the number of students whose weights lie between 120 and 155 lbs. is therefore $0.5872 \times 500 = 294$ (approx.)



(ii) Proportion of students who weigh more than 155 lbs. is
$P(x > 155)$
$= 1 - P(x \le 155)$
$= 1 - \Phi(0.27)$
$= 1 - 0.6064$
$= 0.3936$

The number of students who are expected to weigh more than 155 lbs is $0.3936 \times 500 = 197$ (approx.)

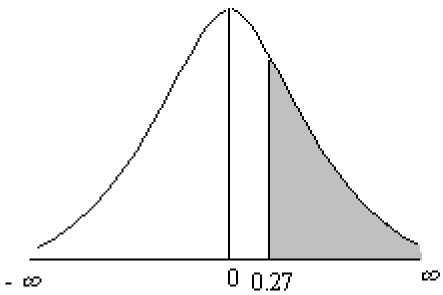
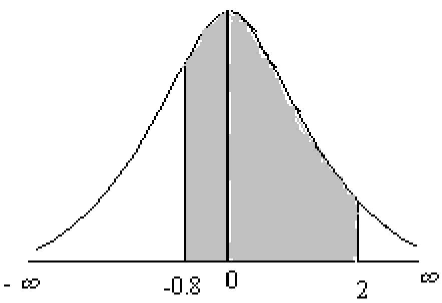


Illustration 2: Let X be a random variable which follows normal distribution with mean 30 and S.D 5. Find the probability that
(i) $26 \le X \le 40$ (ii) $X \ge 45$ (iii) $| X - 30 | > 5$

Solution: Here, the mean μ and the standard deviation σ are 30 and 5 respectively.
(i) When X = 26 we have $Z = (X - \mu)/\sigma = (26 - 30)/5 = -0.8$
and when X = 40 we have $Z = (X - \mu)/\sigma = (40 - 30)/5 = 2$

$\therefore P(26 \le X \le 40)$
$= P(-0.8 \le X \le 2)$
$= P(-0.8 \le X \le 0) + P(0 \le X \le 2)$
$= P(0 \le X \le 0.8) + P(0 \le X \le 2)$
[from symmetry
$= 0.2881 + 0.4772$
[from table
$= 0.7653$



(ii) When X = 45,
We have,

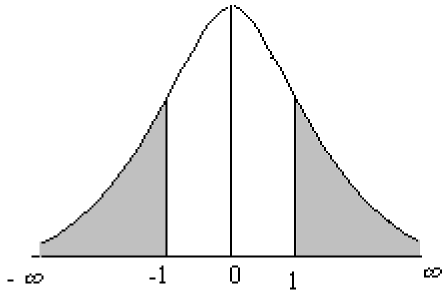
$Z = (X - \mu)/\sigma = (45 - 30)/5 = 3$

Thus, $P(X \geq 45) = P(Z \geq 3) = 1 - P(Z \leq 3)$

$= 1 - [P(-\infty \leq Z \leq 0) + P(0 \leq Z \leq 3)]$

$= 1 - [0.5 + 0.49865] = 0.00135$

(iii) To find $P(|X - 30| \geq 5)$ we proceed in the following way,



$P(|X - 30| \leq 5)$

$= P(-5 \leq X - 30 \leq 5)$

$= P(25 \leq X \leq 35)$

$= P(-1 \leq Z \leq 1)$

$= 2 \times P(0 \leq Z \leq 1)$

$= 2 \times 0.3413$

$= 0.6826$

Thus, $P(|X - 30| \geq 5) = 1 - P(|X - 30| \leq 5) = 1 - 0.6826 = 0.3174$

Illustration 3. In a distribution exactly normal, 7% of the items are under 35 and 89% are under 63. What are the mean and Standard deviation of the distribution?

Solution: Let X be a random variable following normal distribution with mean μ and standard deviation σ. Let $Z = (X - \mu)/\sigma$.

By, the question we have,

$P[X \leq 35] = 0.07$ and $P[X \leq 63] = 0.89$,

Now, $P[X \leq 35] = P[(X - \mu)/\sigma \leq (35 - \mu)/\sigma]$

$= P[Z \leq (35 - \mu)/\sigma] = \Phi[(35 - \mu)/\sigma]$

Thus,

$\Phi[(35 - \mu)/\sigma] = 0.07$

$= 1 - \Phi(1.48)$ [value obtained from table.

$= \Phi(-1.48)$

Thus, $(35 - \mu)/\sigma = -1.48$

$\Rightarrow \mu - 1.48\sigma = 35$ (1)

Similarly,

$P[X \leq 63] = P[(X - \mu)/\sigma \leq (63 - \mu)/\sigma]$

$= P[Z \leq (63 - \mu)/\sigma] = \Phi[(63 - \mu)/\sigma]$

Thus,

$\Phi[(63 - \mu)/\sigma] = 0.89$

$= \Phi(1.23)$ [value obtained from table.

Thus, $(63 - \mu)/\sigma = 1.23$

$\Rightarrow \mu + 1.23\sigma = 63$ (2)

Now, (1) – (2) implies,

$\mu - 1.48\sigma = 35$

(-) $\mu + 1.23\sigma = 63$

$-2.71\sigma = 28$

$\therefore \sigma = 10.33$

Putting the value of $\sigma = 10.33$ in (1) we have

$\mu = 35 + 1.48\sigma = 35 + 15.2884 = 50.2884$

Thus, for the random variable X the mean is 50.2884 and standard deviation is 10.33.

Illustration 4: There are six hundred Economics students in the postgraduate classes of a university and the probability for any student to need a copy of a particular book from the university library on any day is 005. How many copies of the book should be kept in the university library so that the probability may be greater than 090 that none of the students needing a copy from the library has to come back disappointed? (Use normal approximation to the binomial distribution).

Solution: We are given:

$$n = 600,\ p = 0.05,\ \mu = np = 600 \times 0.05 = 30$$

$\sigma^2 = npq = 600\times 0.05\times 0.95 = 28.5$

$\Rightarrow \sigma^2 = \sqrt{(28.5)} = 5.3$

We want x in such a way that

$P(X < x) > 0.90$

$\Rightarrow P(Z<z) > 0.90$ [Here, $z = (x-30)/5.3$

$\Rightarrow P(0 < Z < z) > 0.4$

$\Rightarrow z > 1.28$ (From Normal Probability Tables

$\Rightarrow (x-30)/5.3 > 1.28$

$\Rightarrow x > 30 + 1.28 \times 5.3$

$\Rightarrow x > 36.784 \cong 37$

Hence the university library should keep at least 37 copies of the book.

Formulae

Distribution	Type	Functional Form	Range of Variable	Range of Parameter	Mean	Variance
Bernoulli	Discrete	$p^x(1-p)^{1-x}$	$x = 0, 1$	$0 \le p \le 1$	p	pq
Binomial	Discrete	${}^nC_x\, p^x\, q^{n-x}$	$x = 0, 1, 2, \ldots, n$	$0 \le p \le 1$	np	npq
Poisson	Discrete	$\dfrac{e^{-\lambda}\lambda^x}{x!}$	$x = 0, 1, 2, \ldots$	$\lambda > 0$	λ	λ
Normal	Continuous	$\dfrac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$	$-\infty < x < \infty$	$-\infty < \mu < \infty$, $\sigma > 0$	μ	σ^2

EXERCISES

1. Explain the concept of probability distributions. How does probability mass function differ from probability density function?
2. Let X be a random variable with *p.m.f.* $f(x) = (1 / 32)\, {}^5C_2$, where $x = 0, 1, 2, \ldots 5$. Find the mean and standard deviation of the random variable.

[Ans: 5/2 and $\sqrt{5}/2$]

3. Define Binomia1 distribution and state the conditions under which the distribution holds.
4. Derive the mean and variance of binomial distribution.
5. Obtain the recurrence relation for the probabilities of binomial distribution.
6. Four coins are tossed simultaneously. What is the probability of getting 2 heads and 2 tails?

[Ans: 3/8]

7. Find the probability that in a family of 5 children there will be (i) at least one boy, (ii) at least one boy and one girl. (Assume that the probability of a female birth is 1/2). [Ans: (i) 31/32 (ii) 15/16]
8. In a shooting competition, the probability of a man hitting a target is *1/5*. If he fires *5* times, what is the probability of hitting the target at least twice? [Ans: 821/3125]
9. Assume that on the average 30% of the candidates appearing in an examination from a certain college get First Division. What is the probability that out of a group of 4 such candidates not more than two will fail to get a First Division? [Ans: 0.3483]
10. For a binomial distribution, the mean is 3 and the variance is 2. Find the values of *n* and p. Hence find the probability that X (the variable value) is 5. [Ans: 9, 1/3, 224/2187]
11. For a binomial distribution, the mean and S.D. are respectively 4 and calculate the probability of getting a nonzero value from this distribution. . [Ans: $1 - (0.75)^{10}$]
12. Write down the expressions which define Binomial, Poisson and Normal probability distributions. Give 3 physical situations illustrating a Poisson random variable.
13. Derive the mean and variance of Poisson distribution.
14. State and prove the recurrence relation of probabilities of a Poisson variate.
15. State the conditions under which binomial distribution tends to Poisson distribution. Derive the same.
16. State some examples of binomial distribution and Poisson distribution.
17. A random variable x follows Poisson distribution with parameter $m = 2$. Find the probabilities $P(x = 1)$, $P(x \leq 1)$, $P(x < 1)$, $P(x > 1)$, $P(1 \leq x \leq 3)$. Given $e^{-2} = 0.1353$.

[Ans: 0.2706, 0.4059, 0.1353, 0.5941, 0.7216]

18. The standard deviation of a Poisson distribution is 2. Find the probability that x =3. (Given $e^{-4} = .0183$).

[Ans: 0.1952]

19. Is it possible that a Poisson distribution has the same mean and standard deviation? If so, what is the probability that the variable takes the value zero? [Ans: yes, e^{-1}]
20. For a Poisson distribution, Pr(x=0) Pr(x= 1). Find Pr(x>0). [Ans: $1- e^{-1}$]
21. A discrete random variable x follows Poisson distribution such that P(x= 1) = P(x= 2). Find the mean and variance of the distribution. [Ans: 2, 2]
22. The probability that a Poisson variate X takes a positive value is $(1-e^{-2})$. Find the (i) Mean, (ii) Mode, (iii) probability that X lies between −1 and 1.5. [Ans: 2,1 and $3e^{-2}$
23. If 3% of the bolts manufactured by a company are defective, what is the probability that in a sample of 200 bolts, 5 will be defective? (Given $e^{-6} = 0.00248$).
 [Ans: 0.16
24. Suppose that the number of telephone calls an operator receives from 11.00 am. to 11.05 a.m. follows a Poisson distribution with $m = 3$. (i) Find the probability that the operator will receive no calls in that time interval to-morrow. (ii) Find the probability that in the next 3 days the operator will receive a total of 1 call in that time interval. *(e* = 2.7 18). [Ans: 0.05, 0.0011]
25. The average number of misprints per page of a book is 2. Assuming Poisson distribution, what is the probability that a particular page is free from misprints? if the book contains 1000 pages, how many of the pages contain more than 2 misprints'? [Ans: e^{-2}, $1000(1 -5e^{-2})$]

26. State the importance of Normal distribution in statistics.
27. Explain some of the features of normal distribution.
28. Find the areas under the normal curve *(i.e.,* the probabilities) in the following cases using table:
    (i) between $z = 0$ and $z = 1.8$;
    (ii) between $z = -0.78$ and $z = 0$;
    (iii) between $z = 0.85$ and $z = 2.15$;
    where z is a standard normal variate. [Ans: (i) 0.4641, (ii) *0.2823*, (iii) 0.1819]
29. The mean height of 1000 students at a certain college is 165 cms. and SD is 10 cms. Assuming that the height distribution is normal, find the number of students whose heights are (i) less than 172 cms; (ii) between 159 and 178 cms; and (iii) more than 173.2 cms. [Ans: (i) 258; (ii) 629; (iii) 2061]
30. The mean of a normal distribution is 60 and 6% of the values are greater than 70. Find the standard deviation of this distribution. (Given that the area under the standard normal curve between $z = 0$ and $z = 1.56$ is 0.44 or $z = -\infty$ to $1.56 = 0.94$] [Ans. 6.41]

@#@#@#@#@#@#@#@