12

Probability Distributions

INTRODUCTION

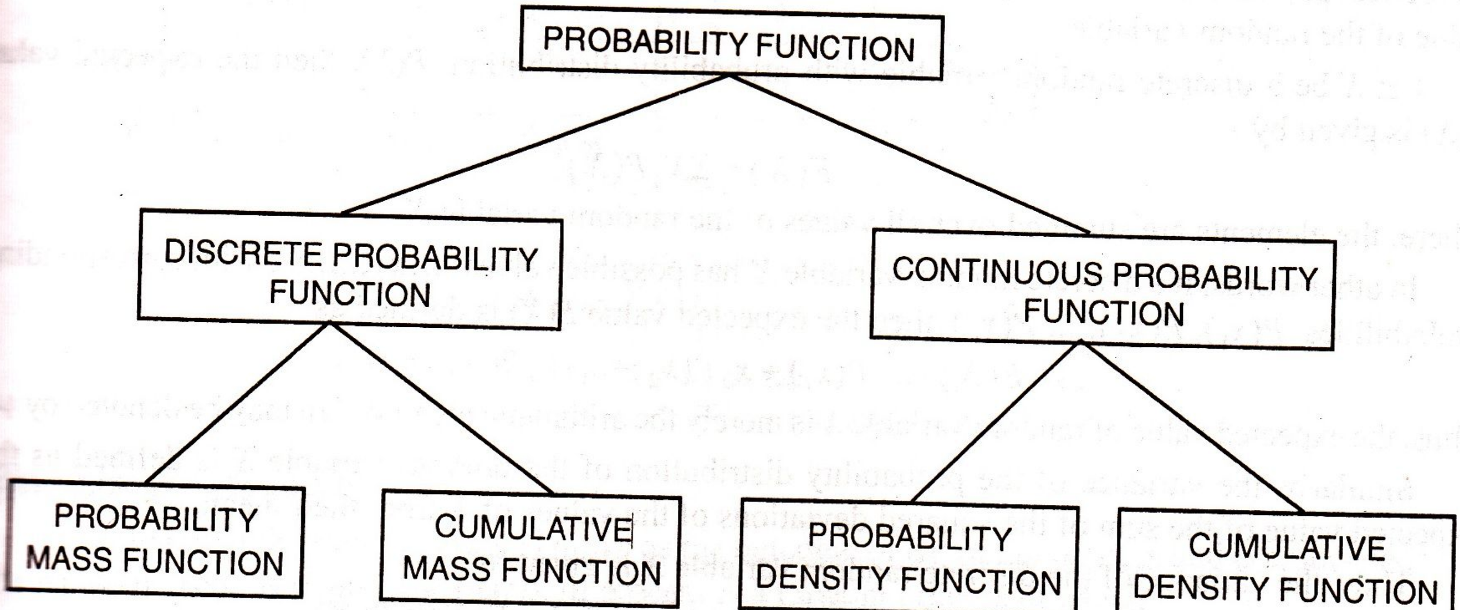
In this chapter, the basic concepts of probability distributions are discussed. In a probability distribution, the variables are distributed according to some definite probability function. We shall discuss some of the most important probability distributions in this chapter. These distributions from their historical interest as well as intrinsic importance, occupy a place of great prominence in business decision-making. Some of the concepts that will facilitate understanding of the topic are given below:

Random Variable

A random variable is a variable which takes specified values with specified probabilities. The probabilities are specified by the way in which the random experiment is conducted and the way in which the random variable is defined and observed on the random experiment. We shall use capital letters to denote a random variable and the corresponding small letters to represent any specific value of the random variable.

Probability Function

If the function permits us to compute the probability for any event that is defined in terms of value of the random variable, then the function is called a probability function. Just as there are discrete and continuous random variables, so there are discrete and continuous probability functions. To emphasize this distinction, we shall draw the diagram as follows:



Discrete Probability Function

A probability function for a discrete random variable is called a discrete probability function since the domain of the function is discrete.

Probability Mass Function

A probability function that specifies the probability that any single value of discrete random variab will occur is called a probability mass function (abbreviated as p.m.f.). If $f(x)$ is the probability ma function of the random variable X, then $f(x) = P(X = x)$ has the following properties :

(*i*) $f(x) \geq 0$ for all values of X; and

(*ii*) $\Sigma f(x) = 1$

Cumulative Mass Function

If X is a discrete random variable with p.m.f. $f(x)$, its cumulative mass function (abbreviated c.m.f.) specifies the probability that an observed value of X will be no greater than x. That is, if $F(x)$ a c.m.f. and $f(x)$ is a p.m.f., then $F(x) = P(X \leq x) = f(X \leq x)$.

Continuous Probability Function

A probability function for a continuous random variable is called a continuous probability functi since the domain of the function is continuous.

Probability Density Function

For a continuous random variable, the corresponding function $f(x)$ is called a probability dens function (abbreviated as p.d.f.). Unlike a p.m.f., a p.d.f. does not specify probabilities for specific individ values of the random variable.

Cumulative Density Function

Corresponding to the cumulative mass function of a discrete random variable, the cumulative dens function (abbreviated as c.d.f.) of a continuous random variable specifies the probability that an observ value of X will be no greater than x.

Expected Value and Variance

The probability distribution provides a model for the theoretical frequency distribution of a rand variable and hence must possess a mean, variance and other descriptive measures associated with t theoretical population which it represents. The average value of a random variable is called the expect value of the random variable.

Let X be a discrete random variable with probability distribution, $P(X)$, then the expected val $E(X)$ is given by

$$E(X) = \Sigma X.P(X)$$

where, the elements are summed over all values of the random variable X.

In other words, if a discrete random variable X has possible values $x_1, x_2, \ldots, x_n$, with correspond probabilities $P(x_1), P(x_2), \ldots, P(x_n)$ then the expected value $E(X)$ is defined as

$$E(X) = x_1 P(x_1) + x_2 P(x_2) + \ldots + x_n P(X_n) = \mu$$

Thus, the expected value of random variable X is merely the arithmetic mean which may be denoted by

Similarly, the variance of the probability distribution of the random variable X is defined as expected value of the sum of the squared deviations of the values of X from their mean.

Thus, the variance of the discrete random variable X is given by

$$\text{Var}(X) = \sigma^2 = E[X - E(X)]^2 = \Sigma[X - E(X)]^2 P(X)$$

$$= E(X^2) - [E(X)]^2.$$

The standard deviation, σ, is the square root of the variance.

Properties of Expected Value and Variance

There are several important properties of expected value and variance which allow computational shortcuts :

1. The expected value of a constant c is equal to the constant.

$$E(c) = c$$

2. The expected value of the product of a constant c and a random variable X is equal to the constant times the expected value of the random variable.

$$E(cX) = cE(X)$$

3. The expected value of the sum of a random variable X and a constant c is the sum of the expected value of the random variable and the constant.

$$E(X + C) = E(X) + c$$

4. The expected value of the product of two independent random variables is equal to the product of thier individual expected values.

$$E(XY) = E(X)\,E(Y)$$

5. The expected value of the sum of the two independent random variables is equal to the sum of their individual expected values.

$$E(X + Y) = E(X) + E(Y)$$

6. The variance of the product of a constant and a random variable X is equal to the constant squared times the variance of the random variable X.

$$\text{Var}(cX) = c^2\,\text{Var}(X)$$

7. The variance of the sum of two independent random variables equal the sum of their individual variances. Also, the variance of the difference of two independent random variables equal the sum of their individual variances.

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) = \text{Var}(X - Y)$$

Illustration 1. Anil company estimates the net profit on a new product, it is launching, to be Rs. 3,000,000 during the first year if it is 'successful' Rs, 1,000,000 if it is moderately successful and a loss of Rs. 1,000,000 if it is 'unsuccessful'. The company assigns the following probabilities to first year prospects for the product, successful : 0.15, moderately successful: 0.25, and unsuccessful: 0.60. What are the expected value and standard deviation of first year net profit for the product? *(MBA, DU, 2003)*

Solution. The probability distribution of net profit (X) of the new product in the first year is given to be

Profit (in million Rs.) X	3	1	–1
Probability $P(X)$	0.15	0.25	0.60

Therefore, expected value of profit is given by

$$\begin{aligned} E(X) &= \Sigma X\,P(X) \\ &= 3 \times 0.15 + 1 \times 0.25 - 1 \times 0.60 \\ &= 0.10 \text{ million Rs.} = \text{Rs. } 1,00,000. \end{aligned}$$

$$\begin{aligned} E(X^2) &= \Sigma X^2\,P(X) \\ &= 9 \times 0.15 + 1 \times 0.25 + 1 \times 0.6 = 2.20 \text{ million Rs.} \end{aligned}$$

$$\text{Var}(X) = E(X^2) - [E(X)]^2 = 2.20 - (0.10)^2 = 2.19$$

$$\text{s.d.} = \sigma = \sqrt{2.19} = 1.48 \text{ (Million Rs.)}$$

Binomial Distribution

The Binomial distribution is also known as the outcome of Bernoulli process and is associated with the name of Jacob Bernoulli. A Bernoulli process is a random process in which :

(*a*) the process is performed under the same conditions for a fixed and finite number of trials, say, n.

(*b*) each trial is independent of other trials, *i.e.*, the probability of an outcome for any particular trial is not influenced by the outcomes of the other trials.

(*c*) each trial has two mutually exclusive possible outcomes, such as "success" or "failure", "good or "defective", "yes" or "no", "hit" or "miss", and so on. The outcomes are usually called success an failure for convenience.

(*d*) the probability of success, p, remains constant from trial to trial (so is the probability of failu q, where, $q = 1 - p$).

These conditions are satisfied if we toss a coin, say, five times. Suppose we are interested in findir the probability of obtaining exactly two heads.

Let us designate head as a success and tail as a failure with corresponding probabilities p and respectively. Find the probability of getting exaclty two heads of five tosses of a fair coin.

Suppose that one of the sequence of outcomes of five tosses of a fair coin showing two heads is:

$$HTHTT$$

The probability of this specific sequence of outcome is found by means of a multiplication rule of probability and is given by

$$pqpqq = p^2q^3$$

Although the resulting probability of obtaining the specific sequence of outcomes in the order shown, we are not interested in the order of occurrence of the successes and failures. Rather, we are interested in the probability of the occurrence of exactly two successes out of five tosses of a coin. In addition to the sequence shown above (call it sequence number 1), two successes and three failures could also occur in any one of the additional sequences shown as follows. Each of the sequences has the same probability of occurring, p^2q^3.

Sequence Number	*Sequence*	*Probability*
2	HHTTT	p^2q^3
3	HTTTH	p^2q^3
4	TTTHH	p^2q^3
5	TTHHT	p^2q^3
⋮	⋮	⋮
⋮	⋮	⋮
10	⋮	p^2q^3

A single sample of five tosses will yield only one sequence of successes and failures. The question then to be answered is: What is the probability of getting sequence number 1 or sequence number 2... or sequence number 10? For finding the answer, the addition rule of probability is used to calculate the sum of the individual probabilities.

To get this, we multiply p^2q^3 by 10, *i.e.*, $10\,p^2q^3$.

Here, $p = 0.5$ and $q = 0.5$

Therefore, the answer is

$$10\,(0.5)^2\,(0.5)^3 = 10 \times 0.25 \times 0.125 = 0.3125$$

As the size of the tosses increases, it becomes more and more difficult to list the number of sequences. An easy method of counting them is required.

We know that the number of combinations of n things taken x at a time is given by

$${}^nC_x = \frac{n!}{x!\,(n-x)!}$$

In our example, $n = 5, x = 2.$

Then $$^5C_2 = \frac{5!}{2!3!} = 10.$$

Therefore, the general model for specifying the probability of obtaining exactly x successes in a given number of n, Bernoulli trials is given by

$$f(x) = P[X = x] = {}^nC_x p^x q^{n-x} \quad \text{for } x = 0, 1, 2, \ldots\ldots., n.$$

where, p = the probability of a success on a single Bernoulli trial
n = the number of Bernoulli trials
x = the number of successes in n trials.

This formula for the probability distribution of the number of successes in series of Bernoulli trials is called the Binomial probability distribution. It gives the probability of obtaining exactly x successes and $(n - x)$ failures in n Bernoulli trials. The Binomial distribution has been extensively tabulated for different values of x and n (See Appendix).

Number of successes x	*Probability f(x)*
0	${}^nC_0 p^0 q^{n-0} = q^n$
1	${}^nC_1 p^1 q^{n-1} = nq^{n-1}p$
2	${}^nC_2 p^2 q^{n-2} = \frac{n(n-1)}{2} q^{n-2} p^2$
:	:
:	:
x	${}^nC_x p^x q^{n-x}$
:	:
:	:
n	${}^nC_n p^n q^{n-n} = p^n$

The binomial distribution satisfies the two essential properties of probability distribution, *viz.*, (*i*) $f(x) \geq 0$; and (*ii*) $\Sigma f(x) = 1$.

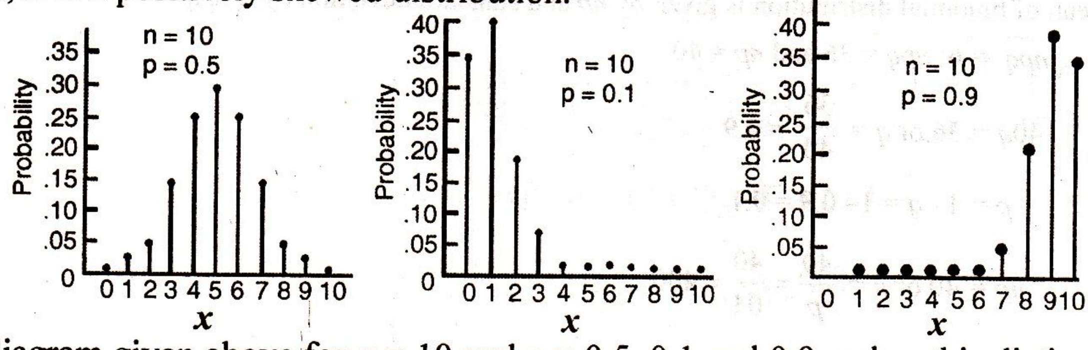
For (*i*), this follows from the fact that both n and p are positive and hence ${}^nC_x$, p^x, q^{n-x} all are positive. Consequently $f(x) \geq 0$.

For (*ii*), we know that binomial expansion of

$$(q + p)^n = q^n + {}^nC_1 q^{n-1} p + {}^nC_2 q^{n-2} p^2 + \ldots\ldots. + {}^nC_x q^{n-x} p^x + \ldots\ldots + p^n$$

Therefore, $\Sigma f(x) = \Sigma\, {}^nC_x p^x q^{n-x} = (q + p)^n = (1 - p + p)^n = 1.$

The binomial distribution is a family of distributions since each different value of n or p specifies a different distribution. In this distribution n and p are called parameters. Regardless of the value of n, the distribution is symmetrical when $p = 0.5$. For small values of n, when p is greater than 0.5, the distribution is asymmetrical, with the peak occurring to the right of centre, *i.e.*, it is a negatively skewed distribution and when p is less than 0.5, the distribution is asymmetrical with the peak occurring to the left of the centre, *i.e.*, it is a positively skewed distribution.



The diagram given above for $n = 10$ and $p = 0.5$, 0.1 and 0.9 makes this distinction clear.

Mean and Variance of Binomial Distribution

The Mean. The mean of binomial random variable X, denoted by μ or $E(X)$, is the theoretic expected number of successes in n trials.

$$\mu = E(X) = \sum_{x=0}^{n} x\, f(x)$$

i.e., the mean of X is the sum of the products of the values that X can assume multiplied by their respecti probabilities.

$$\mu = E(X) = \Sigma x f(x) = \Sigma x\, {}^nC_x\, p^x q^{n-x}$$

$$= \sum_{x=0}^{n} x \frac{n!}{x!\,(n-x)!} p^x q^{n-x}$$

$$= \Sigma x \frac{n(n-1)!}{x(x-1)!\,(n-x)!} p^x q^{n-x}$$

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

$$= np \Sigma\, {}^{n-1}C_{x-1}\, p^{n-1} q^{n-x} = np\,(q+p)^{n-1} = np \qquad [\because (q+p)^{n-1} =$$

Thus, the mean of the binomial distribution is np.

The Variance. The variance of the binomial random variable X measures the variation of the binon distribution and is given by

$$\sigma^2 = E(X^2) - \mu^2 = \sum_{x=0}^{n} x^2 f(x) - \mu^2$$

Here, $\mu = np$

and $\Sigma x^2 f(x) = \Sigma\,[x\,(x-1) + x]\, {}^nC_x\, p^x q^{n-x}$

$$= \Sigma x\,(x-1)\, {}^nC_x\, p^x q^{n-x} + \Sigma x\, {}^nC_x\, p^x q^{n-x}$$

$$= n\,(n-1)\,p^2\,(q+p)^{n-2} + np \qquad \left[\because (q+p)^{n-2} =\right.$$

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

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

$$= np\,[(n-1)\,p + 1 - np]$$

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

Thus, the standard deviation of the binomial distribution is $\sqrt{npq}$ and variance $= npq$.

Illustration 2. The mean of a binomial distribution is 40 and standard deviation 6. Calculate n, p and q.

Solution. The mean of binomial distribution is given by np and standard deviation by $\sqrt{npq}$.

Since $\sqrt{npq} = 6$; $npq = 36$ and $np = 40$

Therefore, $40q = 36$ or $q = \dfrac{36}{40} = 0.9$

$$p = 1 - q = 1 - 0.9 = 0.1$$

Given, $np = 40$ or $n = \dfrac{40}{p} = \dfrac{40}{0.1} = 400.$

Hence for the given question, $n = 400$, $p = 0.1$ and $q = 0.9$.

Illustration 3. Suppose that the half of the population of town are consumers of rice. One hundred investigators are appointed to find out its truth. Each investigator interviewed 10 individuals. How many investigators do you expect to report that three or less of the people interviewed are consumers of rice ? *(MBA, Bharathidasan Univ., Nov. 2001)*

Solution. Probability of a person being consumer of rice is $p = 1/2$, $q = 1/2$. Probability that three people or less are consumers of rice is given by

$$P[X \le 3] = P[X=0] + P[X=1] + P[X=2] + P[X=3]$$
$$= q^{10} + {}^{10}C_1\, q^9 p^1 + {}^{10}C_2\, q^8 p^2 + {}^{10}C_3\, q^7 p^3$$
$$= \left(\frac{1}{2}\right)^{10} + 10\left(\frac{1}{2}\right)^{10} + 45\left(\frac{1}{2}\right)^{10} + 120\left(\frac{1}{2}\right)^{10}$$
$$= \left(\frac{1}{2}\right)^{10} (1 + 10 + 45 + 120) = \frac{176}{1024}$$

Therefore, the number of investigators to report that three or less people are consumers of rice is given by $= \frac{176}{1024} \times 100 = 17.2 \simeq 17$ approx.

Illustration 4. The incidence of occupational disease in an industry is such that the workers have a 20% chance of suffering from it. What is the probability that out of six workers 4 or more will contract the disease ? *(MBA, DU, 2002, 2005)*

Solution. The probability of a worker who is suffering from the disease, *i.e.*, $p = \frac{20}{100} = \frac{1}{5}$

The probability of a worker who is not suffering from the disease

i.e., $$q = 1 - p = 1 - \frac{1}{5} = \frac{4}{5}$$

The probability of 4 or more, *i.e.*, 4, 5 or 6 will contract disease is given by

$$P[X \ge 4] = P[4] + P[5] + P[6]$$
$$= \left({}^6C_4\right)\left(\frac{1}{5}\right)^4\left(\frac{4}{5}\right)^2 + \left({}^6C_5\right)\left(\frac{1}{5}\right)^5\left(\frac{4}{5}\right) + \left({}^6C_6\right)\left(\frac{1}{5}\right)^6$$
$$= 15\left(\frac{1}{5}\right)^4\left(\frac{4}{5}\right)^2 + 6\left(\frac{1}{5}\right)^5\left(\frac{4}{5}\right) + \left(\frac{1}{5}\right)^6$$
$$= \frac{15 \times 16}{15625} + \frac{6 \times 4}{15625} + \frac{1}{15625}$$
$$= \frac{1}{15625}[240 + 24 + 1] = \frac{265}{15625} = \frac{53}{3125} = 0.01696.$$

Illustration 5. Assume that on an average one telephone number out of fifteen is busy. What is the probability that if six randomly selected telephone numbers are called

(*a*) not more than three will be busy ?

(*b*) at least three of them will be busy ?

Solution. p = probability that a telephone number is busy $= \frac{1}{15}$

$$q = 1 - p = 1 - \frac{1}{15} = \frac{14}{15}\text{; and } n = 6.$$

(*a*) The probability that out of six randomly selected telephone numbers not more than three numbers are busy is given by

$$P[X \le 3] = P(0) + P(1) + P(2) + P(3)$$
$$= \left({}^6C_0\right)\left(\frac{14}{15}\right)^6 + \left({}^6C_1\right)\left(\frac{14}{15}\right)^5\left(\frac{1}{15}\right)^1 + \left({}^6C_2\right)\left(\frac{14}{15}\right)^4\left(\frac{1}{15}\right)^2 + \left({}^6C_3\right)\left(\frac{14}{15}\right)^3\left(\frac{1}{15}\right)^3$$
$$= \left(\frac{1}{15}\right)^6\left[(14)^6 + 6(14)^5 + 15(14)^4 + 20(14)^3\right]$$
$$= \frac{(14)^3}{(15)^6}\left[(14)^3 + 6(14)^2 + 15(14) + 20\right]$$

$$= \frac{2744}{(15)^6}[2744 + 1176 + 210 + 20]$$

$$= \frac{2744 \times 4150}{11390625} = \frac{11387600}{11390625} = 0.9997$$

(*b*) Probability that at least three telephone numbers are busy is given by

$$P[X \geq 3] = P(3) + P(4) + P(5) + P(6)$$

$$= ({}^6C_3)\left(\frac{14}{15}\right)^3\left(\frac{1}{15}\right)^3 + ({}^6C_4)\left(\frac{14}{15}\right)^2\left(\frac{1}{15}\right)^4 + ({}^6C_5)\left(\frac{14}{15}\right)\left(\frac{1}{15}\right)^5 + ({}^6C_6)\left(\frac{1}{15}\right)^6$$

$$= 0.0051.$$

Fitting a Binomial Distribution. When a binomial distribution is to be fitted to observe data, t following procedure is adopted :

1. Determine the values of p and q. If one of these values is known, the other can be found out by t simple relationship $p = (1 - q)$ and $q = (1 - p)$. When p and q are equal, the distribution is symmetric Then p and q may be interchanged without altering the value of any term, consequently, terms equidista from the two ends of the series are equal. If p and q are unequal, the distribution is skewed. If p is le than 0.5, the distribution is positively skewed and when p is more than 0.5, the distribution is negativ skewed.

2. Expand the binomial $(q + p)^n$. The power n is equal to one less than the number of terms in t expanded binomial. Thus, when $n = 2$ there will be three terms in the binomial. Similarly, when $n =$ there will be five terms.

3. Multiply each term of the expanded binomial by N (the total frequency), in order to obtain expected frequency in each category.

It is convenient to use the following recurrence relation for fitting of binomial distribution :

$$f(x) = P[X = x] = {}^nC_x p^x q^{n-x}$$

$$f(x+1) = P[X = x+1] = {}^nC_{x+1} p^{x+1} q^{n-x-1}$$

$$\frac{f(x+1)}{f(x)} = \frac{p}{q}\frac{n-x}{x+1}$$

$$f(x+1) = \frac{p}{q}\frac{n-x}{x+1}$$

When $x = 0$,

$$f(1) = \frac{p}{q}\frac{n-0}{0+1}f(0) = \frac{p}{q}n f(0)$$

When $x = 1$,

$$f(2) = \frac{p}{q}\frac{n-1}{2}f(1) = \left(\frac{p}{q}\right)^2\frac{n(n-1)}{2!}f(0)$$

When $x = 2$,

$$f(3) = \frac{p}{q}\frac{n-1}{3}f(2) = \left(\frac{p}{q}\right)^3\frac{n(n-1)(n-2)}{3!}f(0)$$

and so on.

This formula provides us a very convenient method for fitting the binomial distribution. The (probability we need to calculate is $f(0)$ which is equal to q^n, where q can be estimated from the gi data.

Illustration 6. The screws produced by a certain machine were checked by examining number of defectives in a sample of 8. The following table shows the distribution of 128 samples according to the number of defective items they contained :

No. of defectives in a sample of 8 :	0	1	2	3	4	5	6	7	Total
No. of samples :	7	6	19	35	30	23	7	1	128

(*a*) Fit a binomial distribution and find the expected frequencies if the chance of machine being defective is ½.

(*b*) Find the mean and standard deviation of the fitted distribution. (*MBA, Delhi Univ., 2003*)

Solution. (*a*) The probability of a defective screw = ½.

$$p = ½, \quad q = ½, \quad N = 128.$$

Since there are 8 terms, therefore, $n = 7$.

The probability that 0, 1, 2, ..., 7 will be defective is given by expansion of :

$$(½ + ½)^7 = (½)^7 + ({}^7C_1)(½)^6(½)^1 + ({}^7C_2)(½)^5(½)^2 + ({}^7C_3)(½)^4(½)^3 + ({}^7C_4)(½)^3(½)^4$$
$$+ ({}^7C_5)(½)^2(½)^5 + ({}^7C_6)(½)^6(½) + ({}^7C_7)(½)^7$$
$$= (½)^7 [1 + 7 + 21 + 35 + 35 + 21 + 7 + 1]$$

In order to obtain the expected frequencies, we shall multiply each term by *N*, *i.e.*, 128.

$$128\left(\frac{1}{2}+\frac{1}{2}\right)^7 = 128 \times \frac{1}{128}(1+7+21+35+35+21+7+1)$$

Thus, the expected frequencies are :

X :	0	1	2	3	4	5	6	7
f :	1	7	21	35	35	21	7	1

(*b*) Mean and standard deviation of the fitted distribution

Mean of the binomial distribution is *np* and standard deviation is $\sqrt{npq}$.

Here $n = 7,$ $p = ½,$ $q = ½$

Mean = *np* $7 \times ½ = 3.5$

$$\text{Standard deviation} = \sqrt{npq} = \sqrt{7 \times \frac{1}{2} \times \frac{1}{2}} = \sqrt{1.75} = 1.32.$$

Poisson Distribution

A second important discrete probability distribution is the Poisson distribution, named after the French mathematician S. Poisson who published its derivation in 1837.

The characteristics of the Poisson distribution are as follows :

1. The occurrence of the events is independent. That is, the occurrence of an event in an interval of space or time has no effect on the probability of a second occurrence of the event in the same, or any other interval.

2. Theoretically, an infinite number of occurrences of the event must be possible in the interval.

3. The probability of single occurrence of the event in a given interval is proportional to the length of the interval.

4. In any infinitesimal (extremely small) portion of interval, the probability of two or more occurrences of the event is negligible.

Poisson distribution differs from the binomial distribution in two important aspects :

(*a*) Rather than consisting of discrete trials, the distribution operates continuously over some given amount of time, distance, area, etc.

(*b*) Rather than producing a sequence of successes and failures, the distribution produces successes, which occur at random points in the specified time, distance, area. These successes are commonly referred to as 'occurrences'.

The Poisson distribution may be used to approximate binomial distribution when *n* is large and *p* is small and, therefore, is regarded as the limit of the binomial distribution.

The Poisson distribution is given by

$$f(x) = P(X = x) = \frac{e^{-m} m^x}{x!}; \quad x = 0, 1, 2......$$

where, m is called the parameter of the distribution and is the average number of occurrences of random event, x is the number of occurrences of the random event and e is the constant whose value is 2.718.

The Poisson distribution satisfies the two essential properties, *i.e.*, $f(x) \geq 0$ and $\Sigma f(x) = 1$.

The Poisson distribution has been extensively tabulated (see the Appendix). It has many application in business and has been widely used in management science and operations research. The following are some of the examples which may be analysed with the use of this distribution :

(*a*) the demand for a product,

(*b*) typographical errors occurring on the pages of a book,

(*c*) the occurrence of accident in a factory,

(*d*) the arrival pattern in a departmental store,

(*e*) the occurrence of flaws in a bolt in a factory, and

(*f*) the arrival of calls at a switch board.

Mean and Variance of the Poisson Distribution

The Mean. The mean of the Poisson distribution is given by

$$\mu = E(X) = \Sigma x f(x) = \sum x \frac{e^{-m} m^x}{x!}$$

$$= 0 + me^{-m} + m^2 e^{-m} + \frac{m^3 e^{-m}}{2!} + \frac{m^4 e^{-m}}{3!} + \ldots\ldots$$

$$= me^{-m}\left[1 + m + \frac{m^2}{2!} + \frac{m^3}{3!} + \ldots\ldots\right]$$

$$= me^{-m} e^{m} = m$$

Thus, the mean of the Poisson distribution is m.

The Variance. The variance of the Poisson distribution is given by

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

$$= \Sigma x^2 \frac{e^{-m} m^x}{x!} - m^2$$

But, $$\Sigma x^2 \frac{e^{-m} m^x}{x!} = e^{-m} \Sigma \frac{[x(x-1) + x]}{x!} m^x$$

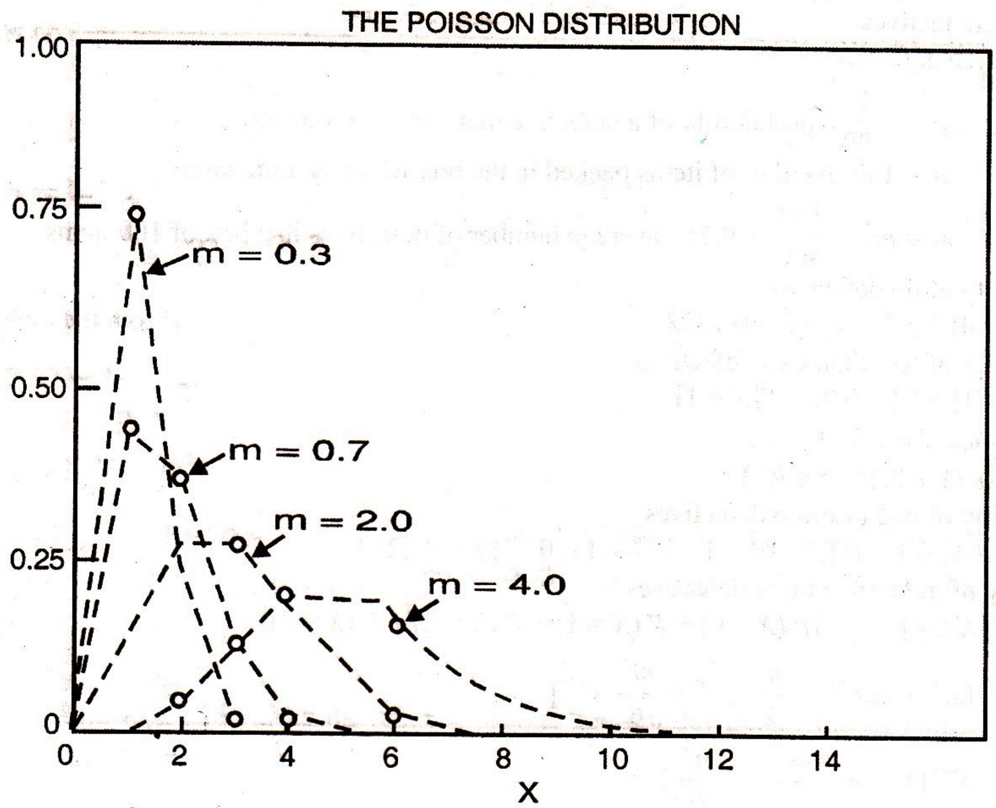
$$= m^2 e^{-m} \Sigma \frac{m^{x-2}}{(x-2)!} + me^{-m} \Sigma \frac{m^{x-1}}{(x-1)!} - m^2$$

$$= m^2 e^{-m} . e^{m} + me^{-m} . e^{m} - m^2$$

$$= m^2 + m - m^2 = m$$

Thus, the variance of the Poisson distribution is also equal to m.

The Poisson distribution is completely defined by the parameter m and is positively skewed. The positive skewness is typical of the Poisson distribution, indicating that, with extremely small probability there is the possibility that distribution will produce an indefinitely large number of occurrences in segment of time or space, even though the mean rate of occurrences may be quite small. As m increase the distribution shift to the right. This is illustrated on the next page, in the diagram for 4 values of from $m = 0.3$ to $m = 4.0$.



The Poisson distribution can frequently be used to approximate the binomial distribution when n is large and p is very small.

Form of the Poisson Distribution

Like binomial distribution, the variate of the Poisson distribution is also discrete one, *i.e.*, it takes only integral values. The probabilities of 0, 1, 2, ... occurrences are given by the successive terms of the expansion.

$$e^{-m}\left(1+m+\frac{m^2}{2!}+\frac{m^3}{3!}+\,........\,+\frac{m^r}{r!}+........\right)$$

This can be written in tabular form as follows :

No. of occurrences (x)	*Probability* $p(x)$	*No. of occurrences* (x)	*Probability* $p(x)$
0	e^{-m}	4	$\frac{m^4 e^{-m}}{4\,!}$
1	me^{-m}	:	:
2	$\frac{m^2 e^{-m}}{2\,!}$	:	:
3	$\frac{m^3 e^{-m}}{3\,!}$	r	$\frac{m^r e^{-m}}{r\,!}$
⋮	⋮	⋮	⋮

where $e = 2.7183$ and m is a constant called the parameter of the distribution, m is the average number of occurrences of an event.

The above table gives probabilities of 0, 1, 2, occurrences respectively. If we want to know the expected number of occurrences, we have to multiply each term by N, *i.e.*, the total number of observations.

Illustration 7. On the average, one in 400 items is defective. If the items are packed in boxes of 100, what is the probability that any given box of items will contain :

(*i*) no defectives ;

(*ii*) less than two defectives ;

(*iii*) one or more defectives ;
(*iv*) more than three defectives. (*MBA, Delhi Univ., 2005*)

Solution. Here, $p = \frac{1}{400}$; probability of a defective item which is very low.

$n = 100$; number of items packed in the box which is quite large

$m = np = \frac{100}{400} = 0.25$; average number of defectives in a box of 100 items

(*i*) Probability of no defective

$= P(X = 0) = e^{-m} = e^{-0.25} = 0.7788$ [From the table given in the appendix]

(*ii*) Probability of less than two defectives

$$= P[X \leq 1] = P[X = 0] + P[X = 1]$$
$$= e^{-m} + me^{-m} = e^{-m}(1 + m)$$
$$= 0.7788\,(1 + 0.25) = 0.9735$$

(*iii*) Probability of one or more defectives

$$= P\,[X \geq 1] = 1 - P\,[X = 0] = 1 - e^{-m} = 1 - 0.7788 = 0.2212$$

(*iv*) Probability of more than three defectives

$$= P\,[X \geq 4] = 1 - [P\,(X = 0) + P\,(X = 1) + P\,(X = 2) + P\,(X = 3)]$$
$$= 1 - [e^{-m} + me^{-m} + \frac{m^2}{2}\,e^{-m} + \frac{m^3}{6}\,e^{-m}]$$
$$= 1 - e^{-m}\,[1 + m + \frac{m^2}{2} + \frac{m^3}{6}]$$
$$= 1 - 0.7788[1 + 0.25 + 0.03125 + 0.0026]$$
$$= 1 - 0.7788[1.28385]$$
$$= 1 - 0.99986 = 0.00014$$

Illustration 8. A factory produces blades in packets of 10. The probability of a blade to be defective is 0.2%. Find the number of packets having two defective blades in a consignment of 10,000 packets.

Solution. $m = np = 10 \times 0.002 = 0.02$

$$P\,[X = 2 \text{ defective blades}] = \frac{e^{-0.02}(0.02)^2}{2}$$
$$= \frac{0.9802 \times 0.0004}{2} = 0.4901 \times 0.0004 = 0.000196$$

Therefore, the total number of packets having two defective blades in a consignment of 10,000 packets is

$10{,}000 \times 0.000196 = 1.96$ or 2.

Illustration 9. What probability model is appropriate to describe a situation where 100 misprints are distributed randomly throughout the 100 pages of a book ? For this model, what is the probability that a page observed at random will contain at least three misprints ?

Solution. Since there are 100 misprints in 100 pages, it implies that there is only one mistake on the average in a page. Therefore, the probability of being a misprint is very small as a page contains large number of words and *n* the number of words in 100 pages will be very large. So, in this case probability of being a misprint is small and *n* is very large, therefore, Poisson distribution is best suited here.

Average or expected number of misprints in one page is

$$m = np = 100 \times 0.01 = 1$$
$$e^{-m} = e^{-1} = 0.3679$$

Probability of at least three misprints in a page is

$$= P[X \geq 3] = 1 - P(X < 3) = 1 - [P(X = 0) + P(X = 1) + P(X = 2)]$$
$$= 1 - [e^{-1} + e^{-1} + \frac{e^{-1}}{2\,!}]$$
$$= 1 - e^{-1}\,[2.5] = 1 - 0.3679\,(2.5)$$
$$= 1 - 0.9198 = 0.0802.$$

Fitting a Poisson Distribution. The process of fitting a Poisson distribution is simple. We have to obtain the values of *m* and calculate the probability of zero occurrences. The other probabilities can be easily calculated by the recurrence relation as follows :

$$f(x) = \frac{e^{-m} m^x}{x\,!}, \text{ and } f(x+1) = \frac{e^{-m} m^{x+1}}{(x+1)\,!}$$

$$\frac{f(x+1)}{f(x)} = \frac{m}{x+1} \text{ or } f(x+1) = \frac{m}{x+1} f(x)$$

when $x = 0, \quad f(1) = m\,f(0)$

when $x = 1, \quad f(2) = \frac{m}{2} f(1) = \frac{m^2}{2} f(0)$

when $x = 2, \quad f(3) = \frac{m}{3} f(2) = \frac{m^3}{6} f(0)$

and so on.

This recurrence relation provides a very convenient method for fitting the Poisson distribution. The only probability we need to know is $f(0) = e^{-m}$, where m is the only parameter of the Poisson distribution.

Multiplying by N (the total frequency) each probability of the Poisson distribution, we get the expected frequencies for respective probabilities.

Illustration 10. The following table gives the number of days in a 50-day period during which automobile accidents occurred in a city. Fit Poisson distribution to the data :

No. of accidents	:	0	1	2	3	4
No. of days	:	21	18	7	3	1

(*MBA, Kumaun Univ., 2006*)

Solution. FITTING OF POISSON DISTRIBUTION

X	f	fX
0	21	0
1	18	18
2	7	14
3	3	9
4	1	4
	$N = 50$	$\Sigma fX = 45$

$$m = \overline{X} = \frac{\Sigma fX}{N} = \frac{45}{50} = 0.9$$

$f(0) = e^{-m} = e^{-0.9} = 0.4066$

$f(1) = m\,f(0) = (0.9)\,(0.4066) = 0.3659$

$$f(2) = \frac{m}{2} f(1) = \frac{0.9}{2}\,(0.3659) = 0.1647$$

$$f(3) = \frac{m}{3} f(2) = \frac{0.9}{3}\,(0.1647) = 0.0494$$

$$f(4) = \frac{m}{4} f(3) = \frac{0.9}{4}\,(0.0494) = 0.0111$$

In order to fit Poisson distribution, we shall multiply each probability by N, *i.e.*, 50.

Hence, the expected frequencies are :

X:	0	1	2	3	4
f:	0.4066 × 50 = 20.33	0.3659 × 50 = 18.30	0.1647 × 50 = 8.24	0.0494 × 50 = 2.47	0.0111 × 50 = 0.56

Negative Binomial Distribution

Where as the binomial distribution describes the probabilities of the number of successes likely to appear in a sequence of *fixed* number of Bernoulli trials, the negative binomial distribution, as the name itself implies, describes the probabilities of the number of trials likely to be required in order to obtain a *fixed* number of successes.

To derive the probability mass function of the negative binomial distribution, we proceed as follows: suppose that x trials are required to obtain exactly k successes ($x \geq k$; *i.e.*, $x = k, k+1, k+2, ...$). Then clearly, the xth trial should yield the kth success, the previous $(x-1)$ trials should give the remaining $(k-1)$ successes and $(x-1)-(k-1)=(x-k)$ failures in some sequence of successes and failures.

Clearly, this can happen in $\binom{x-1}{k-1}$ distinct ways. Now, observe that each particular sequence has k successes and $(x-k)$ failures in itself and hence its probability of occurrence is $p^k(1-p)^{x-k}$. As the sequence of trials are Bernoullian, the probability that exactly x trials will be required to obtain k successes becomes

$$\binom{x-1}{k-1} p^k (1-p)^{x-k}; \qquad x = k, k+1, ...$$

Thus, if x is the random variable corresponding to the number of trials required for observing exactly k successes in an indefinite sequence of Bernoullian trials, its distribution is said to be negative binomial and has the following probability mass function.

$$P[x\,;\,k,p] = \binom{x-1}{k-1} p^k (1-p)^{x-k} \qquad \text{if } x = k, k+1, ...$$

$$= 0, \text{ otherwise}$$

where $0 < p < 1$ and $k \geq 1$ is a fixed integer.

The negative binomial distribution arises in practice, where observation of successes takes place as a waiting-time phenomenon.

For the negative binomial variate, it is easy to prove that :

$$\text{Mean} = \frac{k}{p} \text{ and variance} = \frac{k(1-p)}{p^2}$$

Illustration 11. A market research agency that conducts interviews by telephone has found, from past experience, that there is a 0.40 probability that a call made between 2.30 and 5.30 P.M. will be answered. Assuming a Bernoullian process :

(*a*) What is the probability that an interviewer's tenth answer comes on his twentieth call?

(*b*) What is the expected number of calls necessary to obtain seven answers ?

(*c*) What is the probability that an interviewer will receive his first answer on his third call?

Solution. Let success denote an answer to a call. Then $p = 0.40$.

(*a*) Prob. (10th answer comes on 20th call) $= \binom{x-1}{k-1} p^k (1-p)^{x-k}$

where $x = 20$ and $k = 10$

$$= \binom{20-1}{10-1} (0.4)^{10} (0.6)^{10} = 0.05856$$

(*b*) Expected number of calls for seven answers $= \dfrac{k}{p}$

where $k = 7$ and $p = 0.4$

$$= \frac{7}{0.4} = \frac{70}{4} = 17.5$$

(*c*) Prob. (1st answer on 3rd call) $= \binom{3-1}{1-1} (0.4)^2 (0.6)^2 = 0.1440.$

Multinomial Distribution

The multinomial distribution is the multivariate analogue of the binomial distribution. It is one of the simplest but most important discrete multivariate distributions. The binomial distribution arises from a random experiment in which a finite sequence of n repeated and independent Bernoulli trials are conducted, each trial resulting in only one or two possible outcomes, success or failure. The multinomial distribution arises when the number of possible outcomes of a single trial is generalised from 2 to k.

Let $E_1, E_2, \ldots\ldots, E_k$ denote the k possible outcomes in a single trial of the experiment. Let n repeated trials be conducted and the various outcomes noted. Let X_1 denote the number of times the outcome E_1 is observed with corresponding probabilities p_i $(i = 1, 2, \ldots, k)$. As the outcomes are mutually exclusive,

$$X_1 + X_2 + \cdots\cdots + X_k = n$$

and $$p_1 + p_2 + \cdots\cdots + p_k = 1$$

Then, the joint probability mass function of $X_1, X_2, \ldots\ldots, X_k$ is called the multinomial distribution and is given by

$$P[X_1 = x_1, X_2 = x_2, \ldots, X_k = x_k] = \frac{n!}{x_1!\, x_2! \ldots x_k!}\, p_1^{x_1}\, p_2^{x_2} \cdots\cdots p_k^{x_k}$$

where $X_1, X_2, \ldots\ldots, X_k$ are non-negative integers such that

$$\sum_{i=1}^{k} X_i = n$$

Illustration 12. Consider the production of ball bearings of a certain type, the diameter of which should be 0.2500 inch. Because of the inherent variability in the manufacturing process, the bearings are classified as 'undersize' 'oversize' and 'acceptable' if they measure less than 0.2495 inch, more than 0.2505 inch, and between 0.2495 and 0.2505 inch, respectively. Suppose the production process for three bearings is such that 4% of the bearings are undersize, 6% are oversize, and 90% are acceptable. If 100 of these bearings are picked at random, what is the probability of getting x_1 undersize, x_2 oversize, and x_3 acceptable bearings?

Solution. Here, $n = 100$, $p_1 = 0.04$, $p_2 = 0.06$ and $p_3 = 0.90$

The required probability is given by

$$P(x_1, x_2, x_3) = \frac{100!}{x_1!x_2!x_3!}(0.04)^{x_1}(0.06)^{x_2}(0.90)^{x_3}$$

where $0 \le x_i \le 100;\ i = 1, 2, 3$

and $$\sum_{i=1}^{k} X_i = 100.$$

Illustration 13. In the above illustration, suppose 6 bearings are sampled from the process.

(*i*) What is the probability that there will be two bearings of each category?

(*ii*) What is the probability that all of them will be accepted?

Solution. (*i*) Here, $n = 6, x_1 = 2, x_2 = 2$, and $x_3 = 2$

The required probability is given by

$$= \frac{6!}{2!\,2!\,2!}(0.04)^2\,(0.06)^2\,(0.9)^2$$

$$= 90\,(0.00000576)\,(0.81) = 0.0004199$$

(*ii*) Here, $n = 6, x_1 = 0, x_2 = 0$, and $x_3 = 6$

$$\text{Prob. [all will be acceptable]} = \frac{6}{0!\,0!\,6!}(0.04)^0\,(0.06)^6\,(0.90)^6 = 0.5314.$$

Hypergeometric Distribution

The Hypergeometric distribution arises when a simple random sample without replacement is drawn from a dichotomous population (*i.e.*, one whose elements can be divided into two mutually exclusive categories) consisting of finite number of elements in each of the two categories and the random variable observed is the number of items of one particular category present in the sample. Suppose a box contains numbers of a manufactured product, of which n_1 are good and $n - n_1$ are defective. Suppose a sample of size R is drawn, then the number X of good items in the sample is a random variable taking values 0, 1, 2......, K with probabilities,

$$P(X) = \frac{\binom{n_1}{X}\binom{n-n_1}{K-X}}{\binom{n}{K}} \qquad \text{if } X = 0,1,.......K$$

$= 0$, otherwise.

This probability mass function is called hypergeometric distribution.

It is easy to verify that the mean and the variance of the above distribution are given by :

$$\text{Mean} = \frac{n_1}{n} \text{ and Variance} = K\frac{n_2}{n}\frac{n-n_1}{n}\frac{n-K}{n-1}$$

Illustration 14. Certain missile components are shipped in lots of 12. Three components are selected from each lot and a particular lot is accepted if none of the three components selected is defective.

(*a*) What is the probability that a lot will be accepted if it contains 5 defectives?

(*b*) What is the probability that a lot will be rejected if it contains 9 defectives?

(*c*) Let X be a random variable denoting the number of defectives in a sample of 3 components selected randomly from one of the above lots. If the lot contains 4 defectives, specify the probability function $f(x)$. Present the probability distribution (*i*) as a mathematical expression, and (*ii*) in the form of a table.

(*d*) Under the conditions stated in (c) above, what is the expected number of defectives in a sample of 3 components?

Solution. (*a*) Here, $n = 12$, $n_1 = 5$.

The lot will be accepted if the 3 components randomly selected from this lot has no defectives. The required probability is, therefore,

$$\frac{\binom{7}{3}\binom{5}{0}}{\binom{12}{5}} = \frac{7}{44} = 0.1591$$

(*b*) Here, $n = 12$, $n_1 = 9$.

The lot will be rejected if at least one of the components randomly chosen from this lot is defective. This probability is given by

$$1 - \text{Prob [none is defective]} = 1 - \frac{\binom{3}{3}\binom{9}{0}}{\binom{12}{3}} = 1 - \frac{1}{220} = 1 - 0.0045 = 0.9955$$

(c) The lot contains 4 defectives, therefore,

Number of defectives = 4

Number of non-defectives = 8

X = random variable denoting the number of defectives in a sample of 3 components from the above lot.

The required probability function is

$$f(x) = \frac{\binom{4}{X}\binom{8}{3-X}}{\binom{12}{3}}; X = 0, 1, 2, 3$$

$$= 0, \text{ otherwise}$$

The required table of probabilities is

X	$f(x)$
0	0.2545
1	0.5091
2	0.2182
3	0.0182
	$\Sigma f(x) = 1$

(ii) The expected number of defectives in a sample of 3 components

$$= K.\frac{n_1}{n} \text{ where } K = 3, n_1 = 4 \text{ and } n = 12$$

$$= \frac{3 \times 4}{12} = 1.$$

Normal Distribution

The Normal* distribution was discovered by De Moivre as the limiting case of Binomial model in 1733. It was also known to Laplace no later than 1774, but through a historical error it has been credited to Gauss, who first made reference to it in 1809. Throughout the 18th and 19th centuries, various efforts were made to establish the normal model as the underlying law ruling all continuous random variables—thus, the name Normal. These efforts failed because of the false premises. The normal model has, nevertheless, become the most important probability model in statistical analysis.

The normal distribution is approximation to binomial distribution. Whether or not p is equal to q, the binomial distribution tends to the form of the continuous curve when n becomes large at least for the material part of the range. As a matter of fact, that correspondence between binomial and the normal curve is surprisingly close even for low values of n provided p and q are fairly near equality. The limiting frequency curve, obtained as n, becomes large and is called the normal frequency curve or simply the normal curve.

A random variable X is said to have a normal distribution with parameters μ (mean) and σ^2 (variance) if the density function is given by :

$$y = f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{X-\mu}{\sigma}\right)^2} \qquad -\infty < X < \infty$$

where
$y =$ the computed height of an ordinate at a distance of X from the mean,
$\sigma =$ Standard deviation of the given normal distribution,
$\pi =$ the constant = 3.1416 ; $\sqrt{2\pi} = 2.5066$,
$e =$ the constant = 2.7183 (the base of the system of natural logarithm),
$\mu =$ Mean of the given normal distribution.

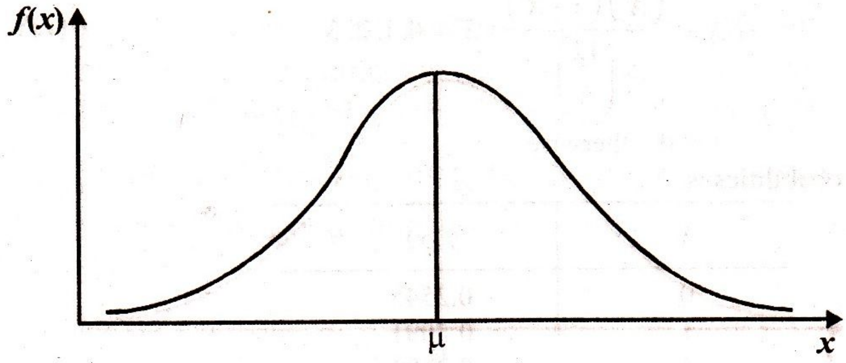
In symbols, it can be expressed as

$$X \sim N(\mu, \sigma)$$

This is read as : The random variable X follows normal distribution with mean μ and standard deviation σ.

*The word normal does not simply mean that the other distributions are abnormal. It is simply the name customarily given to this distribution. Sometimes, the name *Gaussian* is used instead of *normal* to avoid confusion.

If we draw the graph of normal distribution, the curve obtained will be known as normal curve and is given below.



The graph of $y = f(x)$ is a famous 'bell shaped' curve. The top of the bell is directly above the mean μ. For large values of σ, the curve tends to flatten out and for small values of σ, it has a sharp peak.

When we say that curve has unit area, we mean that the total frequency N is equated to 1. To obtain ordinates for particular distribution the ordinates given by the above formula are multiplied by N. The equation to a normal curve corresponding to a particular distribution is given by

$$y = \frac{N}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{X-\mu}{\sigma}\right)^2}$$

The quantity $\frac{N}{\sigma\sqrt{2\pi}}$ in the above formula is equal to the maximum ordinate of the normal curve which will always occur at the mean of the distribution.

A random variable with any mean and standard deviation can be transformed to a standardized normal variable by subtracting the mean and dividing by the standard deviation. For a normal distribution with mean μ and standard deviation σ, the standardized variable z is obtained as

$$z = \frac{X - \mu}{\sigma}$$

A value z represents the distance, expressed as a multiple of the standard deviation, that the value X lies away from the mean. The standardized variable z is called a *standard normal variate* which has mean zero and standard deviation one. In symbols, if

$$X \sim N(\mu, \sigma)$$

then

$$z \sim N(0, 1)$$

The probability density function of the standard normal variate z is given by

$$y = f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} \qquad -\infty < z < \infty$$

Relation between Binomial, Poisson and Normal Distribution

The three distributions, namely Binomial, Poisson and Normal, are very closely related to each other. As explained earlier, when n is large and the probability p of occurrence of an event is close to zero so that np remains a finite constant, then the binomial distribution tends to Poisson distribution.

Similarly, there is a relation between binomial and normal distributions. Normal distribution is a limiting form of binomial distribution under the following conditions :

(1) n, the number of trials is very large, *i.e.*, $n \to \infty$, and

(2) neither p nor q is very small.

In fact, it can be proved that the binomial distribution approaches a normal distribution with standardized normal variable, *i.e.*,

$$z = \frac{X - np}{\sqrt{npq}} \sim N(0, 1)$$

or, $\frac{X - np}{\sqrt{npq}}$ will follow the normal distribution with mean zero and variance one.

Similarly, Poisson distribution also approaches a normal distribution with standardized normal variable, *i.e.*,

$$z = \frac{X - m}{\sqrt{m}} \sim N(0, 1)$$

In other words, $\frac{X - m}{\sqrt{m}}$ will follow the normal distribution with mean zero and variance one.

The Standard Deviation and the Normal Curve

In any normal curve, an exact percentage of observations in the distribution falls within ranges established by the standard deviation in conjunction with the mean. If we cut through the distribution at one, two, and three standard deviation away from mean, on both sides, we obtain the areas of the distribution which contain certain percentages of all the observations.

In a normal curve, between the range of arithmetic mean plus 1 standard deviation and minus 1 standard deviation, *i.e.*, $\mu \pm 1\sigma$ covers 68.27% of the observations in the distribution and 34.13% of observations will fall on either side of mean. Similarly, $\mu \pm 2\sigma$ covers 95.45% of observations and $\mu \pm 3\sigma$ covers 99.73% of observations.

The table shows the area of the normal curve between mean ordinate and ordinates at various sigma (σ) distances from the mean (μ) as percentage of the total area.

AREA RELATIONSHIP

Distance from the mean ordinate	*Percentage of total area*
0.5 σ	19.146
1.0 σ	34.135
1.5 σ	43.319
1.96 σ	47.500
2.00 σ	47.725
2.5 σ	49.379
2.5758 σ	49.500
3.0 σ	49.865

Thus, the two ordinates at distance 1.96 σ from the mean on either side would enclose 47.5 + 47.5 = 95% of the total area and two ordinates at 2.5758 σ distance from the mean on either side would enclose 49.5 + 49.5 = 99% of the total area. The area enclosed between ordinates at 3σ distance from the mean on either side would be 49.865 + 49.865 = 99.73% of the total area.

Moments of the Normal Distribution

For the normal distribution, all odd order moments about mean are zero and given by the relation

$$\mu_1 = \mu_3 = \mu_5 = ... = \mu_{2n+1} = 0$$

and the even order moments about mean is given by the relation

$$\mu_{2n} = 1.3.5 \ldots\ldots\ldots\ldots (2n-1)\sigma^{2n}$$

In particular, we have

$$\mu_2 = 1.\sigma^2 = \sigma^2;\ \mu_4 = 1.3.\sigma^4 = 3\sigma^4.$$

Therefore, moment coefficient of skewness β_1 is given by

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} = 0 \qquad [\because \mu_3 = 0]$$

and moment coefficient of kurtosis is given by

$$\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{3\sigma^4}{\sigma^4} = 3$$

Hence, $\beta_1 = 0$ shows that normal distribution is perfectly symmetrical and $\beta_2 = 3$ indicates that the normal curve is mesokurtic in shape. If $\beta_2 < 3$, the curve is platykurtic and if $\beta_2 > 3$, the curve is leptokurtic.

Properties of the Normal Distribution

The following are the important properties of the normal curve and the normal distribution:

1. The normal curve is symmetrical about the mean (skewness = 0). If the curve is folded along its vertical axis, the two halves will coincide. The number of cases below the mean in a normal distribution is equal to the number of cases above the mean, which makes the mean and median coincide. The height of the curve for a positive deviation of 3 units is the same as the height of the curve for a negative deviation of 3 units.

2. The height of the normal curve is at its maximum at the mean. Hence, the mean and mode of the normal distribution coincide. Thus for a normal distribution mean, median and mode are all equal.

3. There is one maximum point of the normal curve which occurs at the mean. The height of the curve declines as we go in either direction from the mean. The curve approaches nearer and nearer to the base but it never touches it. In other words, the curve is asymptotic to the base on either direction. Hence, its range is unlimited or infinite on both directions.

4. Since there is only one maximum point, therefore, the normal curve is unimodal, *i.e.*, it has only one mode.

5. The points of inflection occur at

$$x = \mu \pm \sigma,\ y = f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}}$$

6. As distinguished from binomial and Poisson distributions where variable is discrete, the variable distributed according to the normal curve is a continuous one.

7. First and third quartiles are equidistant from the median.

8. Mean deviation about mean = $\frac{4}{5}\sigma$ or, more precisely, 0.7979 times of standard deviation.

9. Linear combination of independent normal variates is also a normal variate, *i.e.*, if X_1 and X_2 are two independent normal variates and a_1 and a_2 are given constants, then the linear combination $a_1X_1 + a_2X_2$ will also follow a normal distribution.

10. All odd moments of the normal distribution are zero.

i.e., $\mu_{2n+1} = 0 \qquad (n = 0, 1, 2, \ldots)$

11. $\beta_1 = 0$ and $\beta_2 = 3$

since $\beta_1 = 0$, therefore, the normal distribution is perfectly symmetrical and $\beta_2 = 3$ implies that normal curve is neither leptokurtic nor platykurtic.

12. Mean $\pm\ \sigma$, mean $\pm\ 2\sigma$, and mean $\pm\ 3\sigma$ covers 68.27%, 95.45%, and 99.73% area respectively.

Importance of Normal Distribution

The normal distribution has great significance in statistical work because of the following reasons :

1. The normal distribution has the remarkable property stated in the so-called central limit theorem*, which asserts that certain statistics, most important of which is the sample mean and sample variance, tends to be normally distributed as the sample size becomes large.

2. Even if a variable is not normally distributed, it can sometimes be brought to normal form by simple transformation of variable. For example, if distribution of X is skewed, the distribution of $\sqrt{X}$ might come out to be normal.

3. Many of the sampling distributions like Student's t, F, etc., also tend to normal distribution.

4. The sampling distribution and tests of hypothesis are based upon the assumption that samples have been drawn from a normal population with mean μ and variance σ^2.

5. Normal distribution find large applications in Statistical Quality Control.

6. As n becomes large, the normal distribution serves as a good approximation for many discrete distributions (such as Binomial, Poisson, etc.).

7. In theoretical statistics, many problems can be solved only under the assumption of a normal population. In applied work, we often find that methods developed under the normal probability law yield satisfactory results, even when the assumption of a normal population is not fully met, despite the fact that the problem can have a formal solution only if such a premise is hypothesized.

8. The normal distribution has numerous mathematical properties which make it popular and comparatively easy to manipulate. For example, the moments of the normal distribution are expressed in simple form. The normal curve is reasonably close to many distributions of the humped type. If, therefore, we are ignorant of the exact nature of a humped distribution, or know the form but find it mathematically interactive, we may assume as a first approximation that the distribution is normal and see where this assumption leads us.

The admiration of normal distribution has been beautifully expressed by a well-known statistician W.J. Youden in the following words :

THE
NORMAL
LAW OF ERROR
STANDS OUT IN
THE EXPERIENCE OF
MANKIND AS ONE OF THE
BROADEST GENERALISATION OF
NATURAL PHILOSOPHY. IT SERVES AS THE
GUIDING INSTRUMENT RESEARCHES IN THE
PHYSICAL AND SOCIAL SCIENCES AND IN MEDICINE,
AGRICULTURE AND ENGINEERING. IT IS AN INDISPENSABLE
TOOL FOR THE ANALYSIS AND THE INTERPRETATION OF
THE BASIC DATA OBTAINED BY OBSERVATION AND EXPERIMENT.

Artistically enough, it gives us the shape of the normal curve also.

Area under the Normal Curve

Since for different values of μ and σ, we shall have different normal curves, therefore, it may be very difficult to find areas under the normal curves for different pair of values of μ and σ. Hence, the areas for a normal curve are tabulated in terms of the standardized normal variate z. As any normally distributed random variate X with parameter μ and σ can be transformed to the standardized normally distributed random variate z, therefore, the table given in the appendix under the heading of area under the normal curve may be used.

---

*See Chapter on Sampling and Sampling Distributions.

This table in the appendix contains the probabilities for the area under the normal curve between mean $z = 0$ and any other specified value of z. As the normal curve is symmetrical, therefore, the area under the normal curve is given only for half of the positive side of the curve.

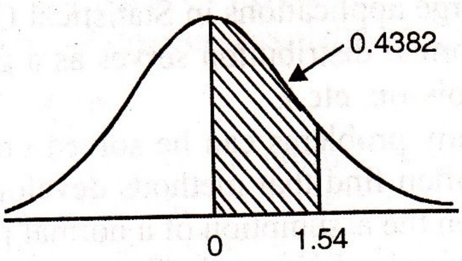
For example, corresponding to $z = 1$, the area under the normal curve is given as 0.3413 (from the table in the appendix). Therefore, for $z = -1$, the area is also 0.3413.

Hence, $Pr[-1 \leq z \leq +1] = 0.3413 + 0.3413 = 0.6826$.

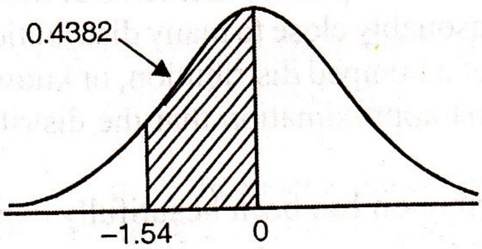
The following are some of the examples to illustrate how tables are to be used in order to obtain area under the normal curve.

Illustration 15. Find out the area under the normal curve for $z = 1.54$.

Solution. If we look to the table, the entry corresponding to $z = 1.54$ is 0.4382 and this gives the area in the shaded region in following figure between $z = 0$ and $z = 1.54$.



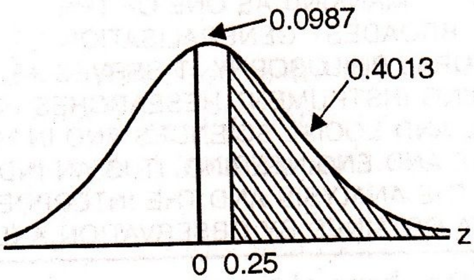
The table given at the end of the book does not contain entries corresponding to negative values of Z. But, since the curve is symmetrical, we can find the area between $z = 0$ and $z = -1.54$ by looking up the area corresponding to $Z = 1.54$.



If we wish to cut the area under normal curve to the right of a positive value of z, we should subtract the tabular value from 0.5000. The reason is that the normal curve is symmetrical, the area to the right of the mean is 0.5000 and the area to the right of a positive value of z is 0.5000 minus the tabular value given for z.

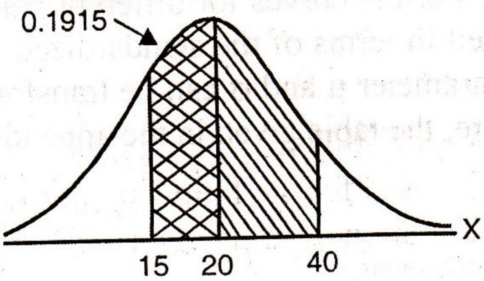
Illustration 16. Find the area to the right of $z = 0.25$.

Solution. Subtract 0.0987 (the entry given in the table for $z = 0.25$) from 0.5000 getting $(0.5000 - 0.0987) = 0.4013$ as shown below :



If we wish to find out the area to the left of a positive value of z, we add 0.5000 to the tabular value given for z.

Illustration 17. A normal curve has $\mu = 20$ and $\sigma = 10$. Find the area between $X_1 = 15$ and $X_2 = 40$.



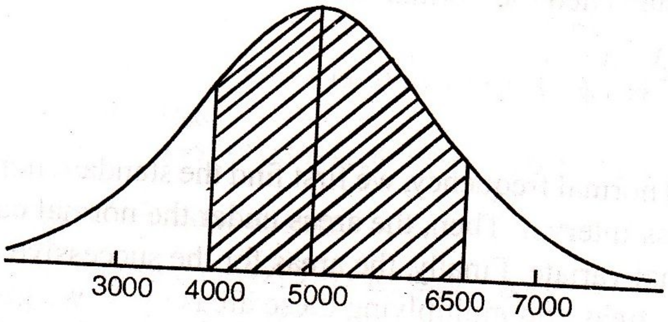
Solution. $$z_1 = \frac{X-\mu}{\sigma} = \frac{15-20}{10} = -0.5$$

$$z_2 = \frac{40-20}{10} = +2.0.$$

Consulting the table, we find the areas corresponding to the z's are 0.1915 and 0.4772 and thus the desired area between $X_1 = 15$ and $X_2 = 40$ is $(0.1915 + 0.4772) = 0.6687$.

Illustration 18. How many workers have a salary between Rs. 4000 and Rs. 6500, if the arithmetic mean is Rs. 5000, standard deviation is Rs. 1000 and number of worker is 15,000, if the salary of the worker is assumed to follow the normal law ?

Solution.



$$z_1 = \frac{4000-5000}{1000} = -1$$ (left of the mean)

$$z_2 = \frac{6500-5000}{1000} = 1.5$$ (right of the mean)

From the table, we find that 34.13% of workers fall between Rs. 4000 and Rs. 5000 and 43.32% fall between Rs. 5000 and Rs. 6500.

$\therefore$ 34.13 + 43.32 = 77.45% of workers have a salary between Rs. 4000 and Rs. 6500.

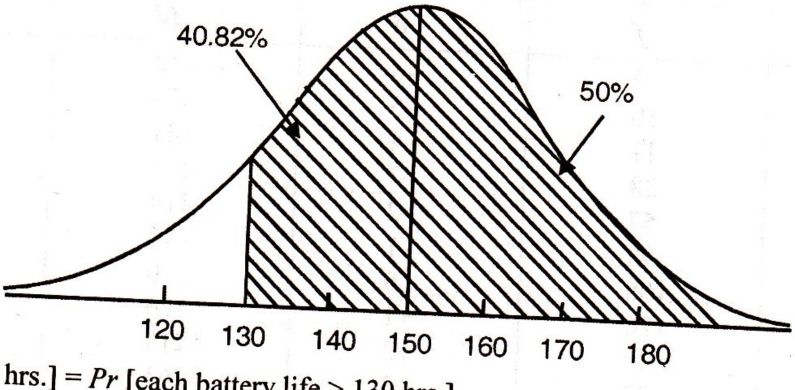
$\therefore$ Number of workers getting a salary between Rs. 4000 and Rs. 6500 is given by

$$0.7745 \times 15{,}000 = 11{,}618$$

Illustration 19. A large flashlight is powered by 5 batteries. Suppose that the life of a battery is normally distributed with $\mu = 150$ hours and $\sigma = 15$ hours. The flashlight will cease functioning if one or more of its batteries go dead. Assuming the lives of batteries are independent, what is the probability that flashlight will operate more than 130 hours ?

Solution. $$z = \frac{X-\mu}{\sigma} = \frac{130-150}{15} = -1.33$$

From the table, we find the 40.82% of the batteries will operate more than 130 hours.



Pr [Flashlight life > 130 hrs.] = *Pr* [each battery life > 130 hrs.]

Since the life of batteries are independent

$$= Pr\,[\text{each battery life} > 130 \text{ hrs.}]^5 = [0.9082]^5$$

Hence, the probability that the flashlight will operate more than 130 hours is given by $(0.9082)^5$.

Applications of the Normal Distribution

The normal distribution is mostly used for the following purposes :

1. To approximate or "fit" a distribution of measurement under certain conditions.

2. To approximate the binomial distribution and other discrete or continuous probability distributions under suitable conditions.

3. To approximate the distributions of means and certain other statistic calculated from samples, especially large samples.

Fitting of Normal Distribution

In order to fit a normal distribution to the given data, we first calculate the mean μ and standard deviation σ from the given data. Then the normal curve fitted to the given data is given by

$$y = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{X-\mu}{\sigma}\right)^2} \qquad -\infty < X < \infty$$

To calculate the expected normal frequency, we first find the standard normal variates corresponding to the lower units of each class-interval. Then, the areas under the normal curve are computed from the tables for each standard normal variate. Finally, the areas for the successive class-intervals are obtained by subtracting the successive areas and multiplying these areas by N, we get the normal frequencies.

The following illustration shall illustrate the applications of normal distribution :

Illustration 20. The following table gives the distribution of height stature (in cms) among the management trainees in Delhi :

Height (in cms)	*No. of trainees*	*Height (in cms)*	*No. of trainees*
161	2	168	126
162	10	169	109
163	11	170	87
164	38	171	75
165	57	172	23
166	93	173	9
167	106	174	4

Test the normality of the distribution by comparing the proportion of cases lying between $\overline{X} \pm 1\sigma$, $\overline{X} \pm 2\sigma$, $\overline{X} \pm 3\sigma$ for the above distribution.

Solution. CALCULATION OF $\overline{X}$ AND σ

X	*f*	*d*	*fd*	*fd*^{2}
61	2	−6	−12	72
62	10	−5	−50	250
63	11	−4	−44	176
64	38	−3	−114	342
65	57	−2	−114	228
66	93	−1	−93	93
67	106	0	0	0
68	126	+1	+126	126
69	109	+2	+218	436
70	87	+3	+261	783
71	75	+4	+300	1,200
72	23	+5	+115	575
73	9	+6	+54	324
74	4	+7	+28	196
	$N = 750$		$\Sigma fd = 675$	$\Sigma fd^2 = 4801$

$$\overline{X} = A + \frac{\Sigma fd}{N} = 167 + \frac{675}{750} = 167.9$$

$$\sigma = \sqrt{\frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd}{N}\right)^2} = \sqrt{\frac{4801}{750} - \left(\frac{675}{750}\right)^2}$$

$$= \sqrt{6.40 - 0.81} = \sqrt{5.59} = 2.36$$

$\bar{X} \pm 1\sigma = 167.9 \pm 2.36 = 165.54$ and 170.26

Number of trainees having stature in this range

$= 93 + 106 + 126 + 109 + 87 = 521.$

Therefore, Proportion $= \frac{521}{750} = 0.69$ or 69%

$\bar{X} \pm 2\sigma = 167.9 \pm 2\,(2.36) = 167.9 \pm 4.72 = 163.18$ and 172.62

Number of trainees having stature in this range

$= 11+38+57+93+106+126+109+87+75+23=725.$

Therefore, Proportion $= \frac{725}{750} = 0.96$ or 96%

$\bar{X} \pm 3\sigma = 167.9 \pm 3\,(2.36) = 167.9 \pm 7.08 = 160.82$ and 174.98

Number of trainees having stature in this range = 750 and hence 100%. In a normal distribution, the proportion lying between these limits is about 68%, 95%, and 99% respectively. Hence, the given distribution is approximately normal.

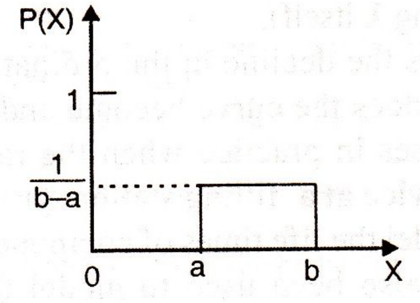
Uniform Distribution

A continuous random variable X is said to have a uniform distribution with parameters a and b if its probability density function is given by

$$P(x) = \frac{1}{b-a}, \quad \text{if } a \le x \le b$$

$$= 0, \text{ otherwise}$$

The uniform distribution is also known as the 'Constant Probability Distribution', as the probability is constant [equal to $1/(b-a)$] at every point of the continuum (a, b) and is independent of whatever value the variable may take within the interval. Yet another name for the uniform distribution is 'rectangular distribution', for the shape of the curve of the probability density function is rectangular, as shown below :



It is easy to check that $P(x)$ defined as above is indeed a probability density function.

The mean and variance of a uniform distribution are given by

$$\text{Mean} = \frac{b+a}{2}; \quad \text{Variance} = \frac{(b-a)^2}{12}$$

The mean, variance and, in fact, all higher order moments of the uniform distribution depend solely on the values of a and b, the lower and upper bounds of the range of values the variable takes. Another interesting property of the uniform distribution is that events described by sub-intervals [of the main interval (a, b)] of equal length have equal probabilities of occurrence. This is a direct consequence, of the *constant* distribution of the total probability over the entire range.

The uniform distribution arises in practice whenever the probability of occurrence of the eve under consideration is constant whatever be the value of the variable, *i.e.*, all possible values of t continuous variable are assumed equally likely. For instance, commuter travel time between specifi points has been considered as a random variable with constant probabilities over a small range of tim

Exponential Distribution

A continuous random variable x is said to have an exponential distribution with parameter λ if i probability density function is given by

$$P(x) = \lambda e^{-\lambda x} \qquad \text{if } 0 \leq x \leq \infty, \lambda >$$

$$= 0, \text{ otherwise.}$$

At the outset, we shall note that whereas the uniform (or rectangular) variable takes values over *finite* range, the exponential variable takes values over an *infinite* range. It is easy to verify that th exponential density function defined as above is indeed a probability density function. We note that th only condition on λ is that it should be a positive real number. Hence by giving different values fo λ, different exponential distributions can be specified. This would help us to understand the natu of the exponential distribution better. The following table gives the ordinates of the exponentia probability density function for $x = 0$ to 6 for $\lambda = 0.2, 0.5, 1$ and 2.

λ / x	0.2	0.5	1	2
0	0.200	0.500	1.000	2.000
1	0.164	0.303	0.368	0.271
2	0.134	0.184	0.135	0.007
3	0.110	0.112	0.050	0.005
4	0.090	0.067	0.018	0.000
5	0.073	0.041	0.007	0.000
6	0.060	0.025	0.002	0.000

It can be proved that :

1. the exponential density function decreases in the range 0 to ∞, the maximum ordinate of the curve occurring at $x = 0$ (the value being λ itself),
2. larger the value of λ, steeper is the decline in the ordinate, even for small values of x,
3. smaller the value of λ, flatter does the curve become and lies closer to X-axis.

The exponential distribution arises in practice when the random variable studied is service time—the time taken to complete service at a filling station, grocery or automobile repair shop. It is also used in reliability theory to model the life times of components subject to wear, *e.g.*, batteries, transistors, tubes, bulbs, etc. It has also been used to model the distribution of length of time between successive random events—the time between arrival of two customers at a service station or the time between breakdowns of a machine.

The mean and variance of the exponential distribution are given by

$$\text{Mean} = \frac{1}{\lambda}; \quad \text{Variance} = \frac{1}{\lambda^2}$$

MISCELLANEOUS ILLUSTRATIONS

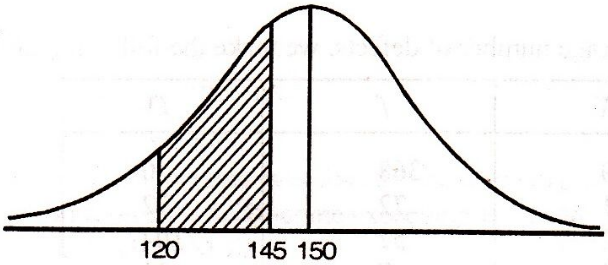
Illustration 21. The average daily sales of 500 branch offices was Rs. 150 thousand and the standard deviation Rs. 15 thousand. Assuming the distribution to be normal, indicate how many branches have sales between :

(*i*) Rs. 120 thousand and Rs. 145 thousand.

(*ii*) Rs. 140 thousand and Rs. 165 thousand.

Solution. (*i*) Standard normal variate corresponding to 120 is

$$z = \frac{X-\mu}{\sigma} = \frac{120-150}{15} = -2$$



and corresponding to 145, the standard normal variate is

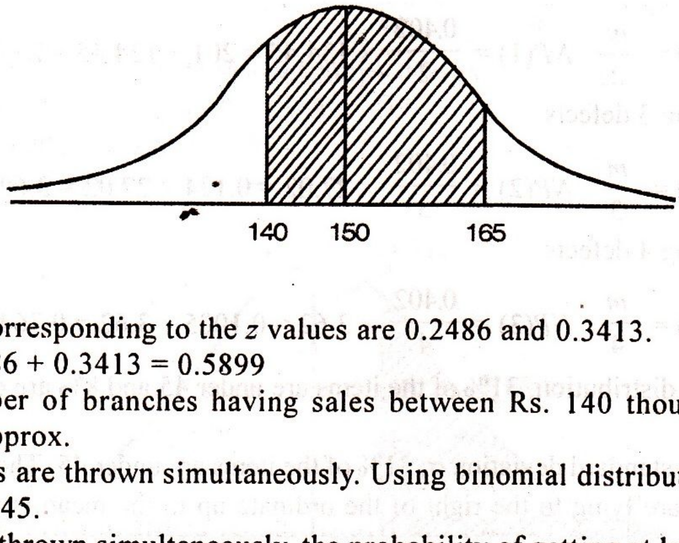
$$z = \frac{145-150}{15} = \frac{-5}{15} = -0.33.$$

From the table, we find the areas corresponding to the values of z are 0.4772 and 0.1293.
Therefore, the desired area between Rs. 120 and Rs. 145 = 0.4772 – 0.1293 = 0.3479.
Hence, the expected number of branches having sales between Rs. 120 thousand and Rs. 145 thousand are 0.3479 × 500 = 173.95 = 174 approx.
(*ii*) Standard normal variate corresponding to 140 is

$$z = \frac{140-150}{15} = \frac{-10}{15} = -0.67.$$

and corresponding to 165, the standard normal variate is

$$z = \frac{165-150}{15} = 1$$



From the table, the areas corresponding to the z values are 0.2486 and 0.3413.
Therefore, the area is 0.2486 + 0.3413 = 0.5899
Hence, the expected number of branches having sales between Rs. 140 thousand and Rs. 165 thousand are 0.5899 × 500 = 294.95 or 295 approx.

Illustration 22. Eight coins are thrown simultaneously. Using binomial distribution, show that the probability of obtaining at least 6 heads is 0.1445. *[MBA, DU, 2003]*

Solution. If eight coins are thrown simultaneously, the probability of getting at least six heads will be given by the separate probabilities of getting 6 heads, 7 heads and 8 heads,

$$= P[x \geq 6] = P[x=6] + P[x=7] + P[x=8]$$

$$n = 8;\ x = 6, 7, 8;\ p = \frac{1}{2} = q$$

$$= {}^8C_6\left(\frac{1}{2}\right)^6\left(\frac{1}{2}\right)^2 + {}^8C_7\left(\frac{1}{2}\right)^7\left(\frac{1}{2}\right)^1 + {}^8C_8\left(\frac{1}{2}\right)^8\left(\frac{1}{2}\right)^0$$

$$= 28\left(\frac{1}{2}\right)^8 + 8\left(\frac{1}{2}\right)^8 + \left(\frac{1}{2}\right)^8$$

$$= \left(\frac{1}{2}\right)^8 [28 + 8 + 1] = \frac{37}{256} = 0.1445$$

Illustration 23. Five hundred television sets are inspected as they come off the production line and the number of defects per set is recorded below :

No. of defects (X):	0	1	2	3	4
No. of sets :	368	72	52	7	1

Estimate the average number of defects per set and expected frequencies of 0, 1, 2, 3 and 4 defects assuming Poisson distribution.

Solution. For finding the average number of defects, we make the following table :

X	f	fX
0	368	0
1	72	72
2	52	104
3	7	21
4	1	4
	$N = 500$	$\Sigma fX = 201$

Therefore, $$\overline{X} = \frac{\Sigma fX}{N} = \frac{201}{500} = 0.402$$

Thus, average number of defects per set is $m = 0.402$

The expected frequency of getting 0 defect

$$= NP(0) = 500 \times e^{-0.402}$$

But $e^{-0.402} = 0.6689$ [From the table]

Therefore, $NP(0) = 500 \times 0.6689 = 334.45$

Expected frequency of getting one defect

$$NP(1) = m.NP(0) = 0.402 \times 334.45 = 134.45$$

Expected frequency of getting 2 defects

$$NP(2) = \frac{m}{2} \cdot NP(1) = \frac{0.402}{2} \times 134.45 = 201 \times 134.45 = 27.02$$

Expected frequency of getting 3 defects

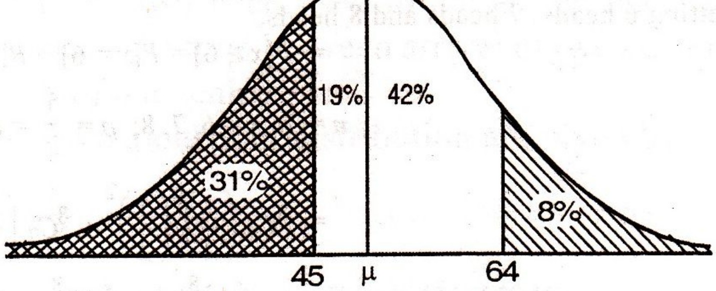
$$NP(3) = \frac{m}{3} \cdot NP(2) = \frac{0.402}{3} \times 27.02 = 0.134 \times 27.02 = 3.62$$

Expected frequency of getting 4 defects

$$NP(4) = \frac{m}{4} \cdot NP(3) = \frac{0.402}{4} \times 3.62 = 0.1005 \times 3.62 = 0.364.$$

Illustration 24. In a normal distribution, 31% of the items are under 45 and 8% are over 64. Find the mean and standard deviation of the distribution. *(MBA, Delhi Univ., 1999)*

Solution. Let mean be μ and standard deviation σ, 31% of the items are under 45. They are lying to the left of the ordinate at $X = 45$ is 0.31, and therefore, are lying to the right of the ordinate up to the mean is $(0.5 - 0.31) = 0.19$. The value of z corresponding to this area is 0.5.



Hence $$z = \frac{45 - \mu}{\sigma} = -0.5 \quad ...(i)$$

8% of the items are above 64. Therefore, area to the right of the ordinate at 64 is 0.08. Area to the left of the ordinate at $X = 64$ up to mean ordinate is $(0.5 - 0.08) = 0.42$ and the value of z corresponding to this area is 1.4.

Hence $$z = \frac{64-\mu}{\sigma} = 1.4 \qquad ...(ii)$$

From equations (*i*) and (*ii*)

$$-\mu + 0.5\sigma = -45$$
$$-\mu - 1.4\sigma = -64$$
$$1.9\sigma = 19 \qquad \text{or } \sigma = 10$$
$$\mu - 0.5 \times 10 = 45 \qquad \text{or } \mu = 50$$

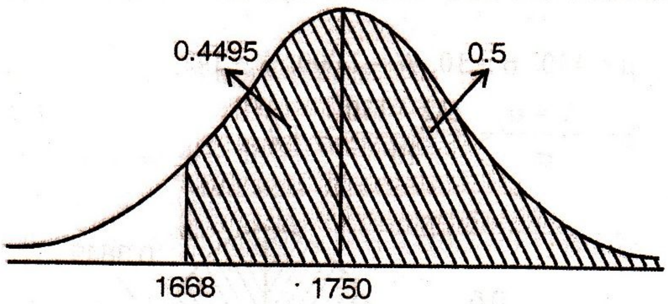
The mean of the distribution is 50 and standard deviation 10.

Illustration 25. The income of a group of 10,000 persons was found to be normally distributed with mean Rs. 1750 p.m. and standard deviation Rs. 50. Show that of this group 95% had income exceeding Rs. 1668 and only 5% had income exceeding Rs. 1832. What was the lowest income among the richest 100 ?

Solution. Standard normal variate is

$$z = \frac{X-\mu}{\sigma}$$

Here $X = 1668, \mu = 1750, \sigma = 50$



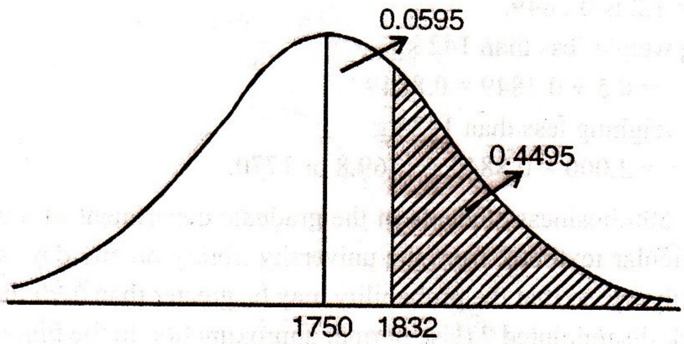
$$z = \frac{1668-1750}{50} = \frac{-82}{50} = -1.64$$

Area to the right of the ordinate at – 1.64 is (0.4495 + 0.5000) = 0.9495.

∴ The expected number of persons getting above Rs. 1668

= 10,000 × 0.9495 = 9495

This is about 95% of the total, *i.e.*, 10,000.



The standard normal variate corresponding to 1832 is

$$z = \frac{1832-1750}{50} = \frac{82}{50} = 1.64$$

Area to the right of ordinate at 1.64 is

0.5000 – 0.4495 = 0.0505

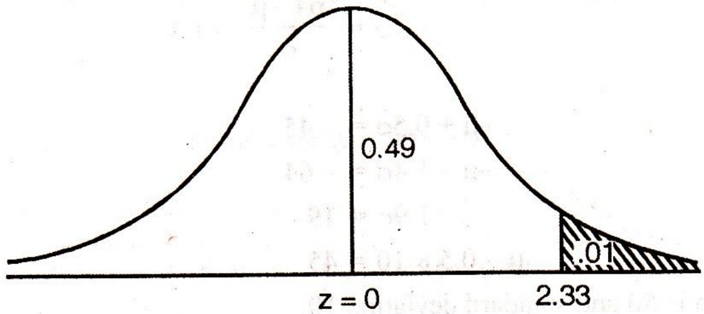
The number of persons getting above Rs. 1832 is

10,000 × 0.0505 = 505

This is 5% approx. of the total, *i.e.*, 10,000.

Probability of getting richest 100

$$= \frac{100}{10,000} = 0.01$$



Standard normal variate having 0.01 area to its right = 2.33

$$2.33 = \frac{X - 1750}{50}$$

$$X = 2.33 \times 50 + 1750 = \text{Rs. } 1866 \text{ approx.}$$

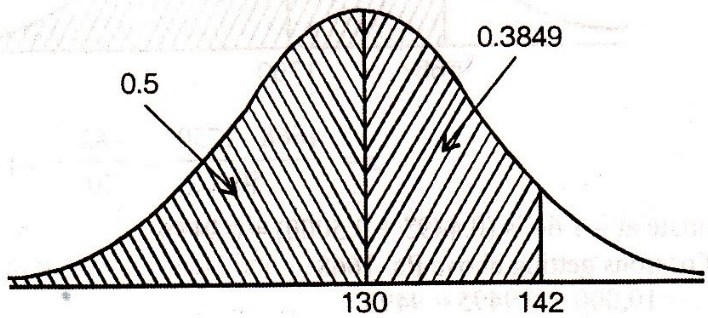
Hence, the lowest income among the richest 10C is Rs. 1866.

Illustration 26. A workshop produces 2000 units per day. The average weight of units is 130 kg with a standard devi... of 10 kg. Assuming normal distribution, how many units are expected to weigh less than 142 kg?

Solution. We are given :

$$\mu = 130, \ \sigma = 10, \ N = 2{,}000, \ X = 142$$

$$z = \frac{X - \mu}{\sigma} = \frac{142 - 130}{10} = 1.2$$



Area between $z = 0$ and $z = 1.2$ is 0.3849.

Probability of units having weight less than 142 kg

$$= 0.5 + 0.3849 = 0.8849$$

Expected number of units weighing less than 142 kg

$$= 2{,}000 \times 0.8849 = 1769.8 \text{ or } 1770.$$

Illustration 27. There are 600 business students in the graduate department of a university, and the probability for ... student to need a copy of a particular textbook from the university library on any day is 0.05. How many copies of the b... should be kept in the university library so that the probability may be greater than 0.90 that none of the students needing a c... from the library has to come back disappointed ? (Use normal approximation to the binomial probability law.)

Solution. Let X represent the number of copies of a textbook required on any day.

$$\text{Mean} = \mu = np = 600 \times 0.05 = 30$$

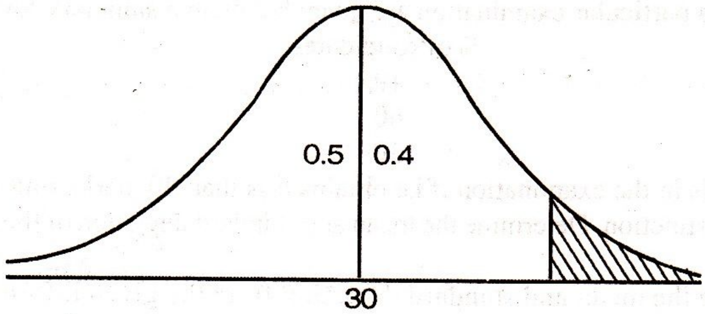
$$\text{s.d.} = \sigma = \sqrt{npq} = \sqrt{600 \times 0.05 \times 0.95} = \sqrt{28.5} = 5.3$$

Therefore, area between mean and ordinate at X is greater than (0.9 – 0.5), *i.e.*, 0.4, corresponding to this area, stan... normal variate $z = 1.28$, and we get

$$\frac{X - \mu}{\sigma} > z$$

$$\frac{X - 30}{5.3} > 1.28$$

or $\quad X - 30 > 6.784$ or $X > 36.784$



Hence, 37 copies of the textbook are required on any day.

Illustration 28. 1,000 tube lights with a mean life of 120 days are installed in a new factory, their length of life is normally ributed with standard deviation 20 days. (*i*) How many tube lights will expire in less than 90 days ? (*ii*) If it is decided to e all the tube lights together, what interval should be allowed between replacements if not more than 10 per cent should e before replacement ?

Solution. (*i*) $\mu = 120, \sigma = 20, X = 90$

Standard normal variate is

$$z = \frac{90-120}{20} = -1.5$$

Area of the curve at ($z = -1.5$) up to the mean ordinate = 0.4332

Area to the left of $-1.5 = 0.5 - 0.4332 = 0.0668$.

Number of tube lights expected to expire in less than 90 days

$$= 0.0668 \times 1000 = 66.8 = 67$$

(*ii*) The value of standard normal variate corresponding to an area 0.4(0.5 – 0.1) is 1.28.

$$\frac{X-120}{20} = -1.28$$

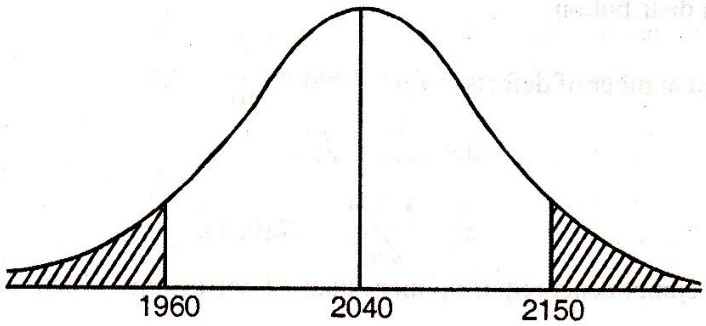
$$X = 120 - 1.28 \times 20 = 120 - 25.6 = 94.4 \text{ or } 94.$$

Hence, 10 per cent of the tube lights will have to be replaced after 94 days.

Illustration 29. As a result of tests on 20,000 electric fans manufactured by a company, it was found that lifetime of the was normally distributed with an average life of 2,040 hours and standard deviation of 60 hours. On the basis of the mation, estimate the number of fans that is expected to run for (*a*) more than 2,150 hours and (*b*) less than 1,960 hours.

Solution. (*a*) $X = 2150, \mu = 2040, \sigma = 60$.

$$z = \frac{X-\mu}{\sigma} = \frac{2150-2040}{60} = \frac{110}{60} = 1.833$$



Area to the right of ordinate at 1.833

$$= 0.5 - 0.4664 = 0.0336$$

The number of fans that is expected to run more than 2,150 hours

$$= 0.0336 \times 20{,}000 = 672.$$

(*b*) $X = 1960, \mu = 2040, \sigma = 60$.

$$z = \frac{1960-2040}{60} = -1.333$$

Area to the left of ordinate at $1.333 = 0.5 - 0.4082 = 0.0918$

∴ The number of fans that is expected to run for less than 1.960 hours

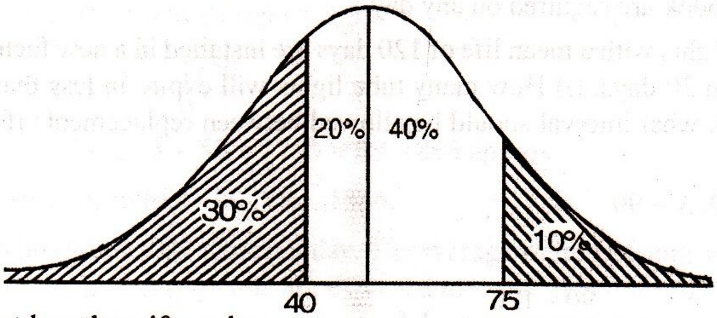
$$= 0.0918 \times 20{,}000 = 1836.$$

Illustration 30. The results of a particular examination are given below in a summary form :

Result	% of candidates
(*i*) Passed with distinction	10
(*ii*) Passed without distinction	60
(*iii*) Failed	30

It is known that a candidate fails in the examination if he obtains less than 40 marks (out of 100) while he must obtain at least 75 marks in order to pass with distinction. Determine the mean and standard deviation of the distribution of marks, assuming this to be normal.

Solution. We have to compute the mean and standard deviation from the given information. The following diagram will help in understanding the question and finding its solution :



We know that 30% students get less than 40 marks.
Therefore, from the table, z value corresponding to

$$0.2(20\% \text{ area}) = -0.524$$

Hence $$\frac{40-\mu}{\sigma} = -0.524 \quad ...(i)$$

Also, 10% students get distinction marks, *i.e.*, 75 or more.
Therefore, from the table, z value corresponding to

$$0.4(40\% \text{ area}) = 1.28$$

Hence $$\frac{75-\mu}{\sigma} = 1.28 \quad ...(ii)$$

Solving equations (*i*) and (*ii*), we get

$$\mu = 50.17 \text{ and } \sigma = 19.4$$

Hence, the mean of the distribution is 50.17 and standard deviation 19.4.

Illustration 31. A complex television component has 1,000 joints by machine which is known to produce on average, one defect in forty. The components are examined, and faulty soldering corrected by hand. If components requiring more than 35 corrections are discarded, what proportion of the components will be thrown away? (Apply normal distribution.)

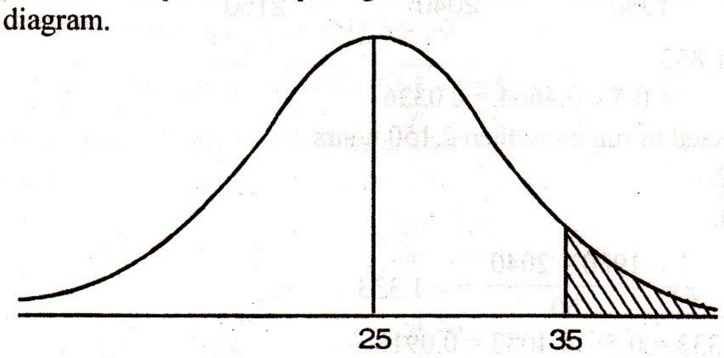
Solution. Since the probability of occurrence of the defect is very small, therefore, it is more appropriate to use normal distribution as a limiting case of Poisson distribution.

Here, $$m = \text{average number of defects} = np = 1000 \times \frac{1}{40} = 25$$

$$\mu = \sqrt{m} = \sqrt{25} = 5$$

We know that $$z = \frac{x-m}{\sqrt{m}} \sim N(0, 1).$$

Since it is required to find out the components requiring more than 35 corrections which have to be discarded is shown below by the shaded area in the diagram.



Therefore, $$z = \frac{35-25}{5} = 2.$$

From the table, the corresponding value of $z = 2$ is 0.4772. The required shaded area will be $0.5 - 0.4772 = 0.0228$. Hence, the number of proportion of the components which will be thrown away is 2.28%.

Illustration 32. A baker has studied his record and notices that for the past 310 working days in the year, the demand for his product (bread) has varied as follows :

Demand ('000 units)	:	5	6	7	8	9	10
Number of days	:	20	60	80	120	20	10

What is the expected demand for his product ?

Solution. Let the demand ('000 units) be denoted by X. Then

$$E(X) = \Sigma X\, P(X)$$

$$= 5 \times \frac{20}{310} + 6 \times \frac{60}{310} + 7 \times \frac{80}{310} + 8 \times \frac{120}{310} + 9 \times \frac{20}{310} + 10 \times \frac{10}{310}$$

$$= \frac{1}{310}\,[100 + 360 + 560 + 960 + 180 + 100]$$

$$= \frac{1}{310}\,[2260] = \frac{226}{31} = 7.29032$$

Therefore, the expected demand = 7.29032 × 1000 = 7290.32.

Illustration 33. In a town, 10 accidents took place in a span of 50 days. Assuming that the number of accidents per day follows Poisson distribution, find the probability that there will be three or more accidents in a day.

Solution. The average number of accidents per day is

$$m = \frac{10}{50} = 0.2$$

Prob. (3 or more accidents) = 1 – Prob. (2 or less accidents)

= 1– [Prob. (0 accident) + Prob. (1 accident) + Prob. (2 accidents)]

$$= 1 - [e^{-m} + me^{-m} + \frac{m^2}{2} e^{-m}]$$

$$= 1 - e^{-m}\,[1 + m + \frac{m^2}{2}]$$

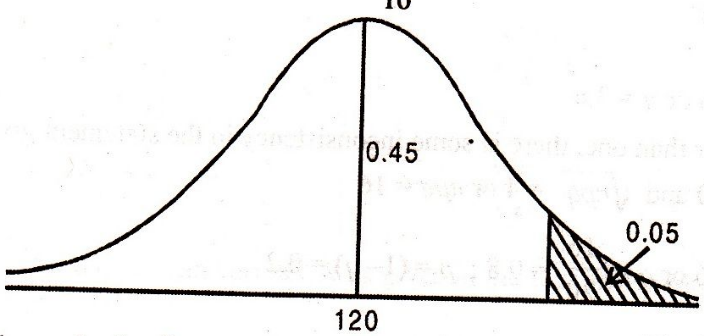
$$= 1 - e^{-0.2}\,[1 + 0.2 + \frac{0.04}{2}]$$

$$= 1 - 0.8187\,[1.22] = 1 - 0.9988 = 0.0012.$$

Illustration 34. A wholesale distributor of fertilizer products finds that the annual demand for one type of fertilizer is normally distributed with a mean of 120 tonnes and standard deviation of 16 tonnes. If he orders only once a year, what quantity should be ordered to ensure that there is only a 5 per cent chance of running short ? *(MBA, Delhi Univ., 2005, 2007)*

Solution. Let the annual demand (in tonnes) be denoted by the random variable X.

Therefore,
$$z = \frac{X - 120}{16}$$



The desired area of 0.05 is shown in the figure. Since the area between the mean and the given value of X is 0.45, therefore, from the table we get this area of 0.45 corresponding to $z = 1.64$.

Substituting this value of $z = 1.64$ in standardized normal variate, we get

$$1.64 = \frac{X - 120}{16}$$

or
$$X = 120 + (1.64)\,(16) = 146.24.$$

If it is necessary to order in whole units, then the wholesale distributor should order 147 tonnes.

Illustration 35. The probability that a bulb will fail before 100 hours is 0.2. Bulbs fail independently. If 15 bulbs are tested for life lengths, what is the probability that the number of failures before 100 hours does not exceed 3?

Solution. The given problem can be assumed to follow binomial distributions.

Here, $n = 15\ ;\ p = 0.2\ ;\ q = 0.8.$

The required probability is given by :

$$P[X \leq 3] = P[X = 0] + P[X = 1] + P[X = 2] + P[X = 3]$$
$$= {}^{15}C_0(0.2)^0(0.8)^{15} + {}^{15}C_1(0.2)^1(0.8)^{14} + {}^{15}C_2(0.2)^2(0.8)^{13} + {}^{15}C_3(0.2)^3(0.8)^{12}$$
$$= (0.8)^{15} + 15(0.2)^1(0.8)^{14} + 105(0.2)^2(0.8)^{13} + 455(0.2)^3(0.8)^{12}$$
$$= (0.8)^{12}[(0.8)^3 + 15(0.2)(0.8)^2 + 105(0.2)^2(0.8) + 455(0.2)^3]$$
$$= 0.0687[0.512 + 1.92 + 3.36 + 3.64]$$
$$= 0.0687(9.432) = 0.648.$$

Illustration 36. A manufacturer who produces medicine bottles, finds that 0.1% of the bottles are defective. The bottles are packed in boxes containing 500 bottles. A drug manufacturer buys 100 boxes from the producer of bottles. Using Poisson distribution, find how many boxes will contain :

(*i*) no defectives,

(*ii*) at least two defectives. (*MBA, Delhi Univ., 2001, 2007*)

(Given $e^{-0.5} = 0.6065$)

Solution. We are given : $p = 0.001$, $n = 500$, $m = np = 500 \times 0.001 = 0.5$

(*i*) $P[X = 0] = e^{-m} = e^{-0.5} = 0.6065$

Therefore, the required number of boxes

$$= 0.6065 \times 100 = 60.65 \text{ or } 61$$

(*ii*)
$$P(X > 2) = 1 - [P_r[X = 0] + P[X = 1]]$$
$$= 1 - [e^{-m} + me^{-m}]$$
$$= 1 - [0.6065 + 0.5(0.6065)]$$
$$= 1 - 0.6065 + 0.30325$$
$$= 1 - 0.90975 = 0.09025.$$

Therefore, the required number of boxes

$$= 100 \times 0.09025 = 9.025 \text{ or } 9.$$

Illustration 37. (*a*) Bring out the fallacy, if any, in the following statement. "The mean of a binomial distribution is 10 and its standard deviation is 6".

(*b*) The mean of a binomial distribution is 20 and the standard deviation is 4. Calculate *n, p* and *q*.

Solution. (*a*) The mean of a binomial distribution is *np* and standard deviation $\sqrt{npq}$.

$$np = 10 \text{ and } \sqrt{npq} = 6 \quad \text{—(i)}$$

Squaring, $npq = 36$

Putting the value of *np* in (*i*)

$$10q = 36 \text{ or } q = 3.6$$

Since the value of *q* is greater than one, there is some inconsistency in the statement given.

(*b*) Given $np = 20 \text{ and } \sqrt{npq} = 4 \text{ or } npq = 16 \quad \text{—(ii)}$

$$20q = 16 \text{ or } q = \frac{16}{20} = 0.8\ ;\ p = (1 - q) = 0.2$$

Putting the value of *p* and *q* in (*ii*)

$$n(0.8)\,(0.2) = 16 \text{ or } n = 100.$$

Illustration 38. One-fifth per cent of the blades produced by a blade manufacturing company turn out to be defective. The blades are supplied in packets of 10. Use poisson distribution to calculate the approximate number of packets containing no defective, one defective and two defective blades respectively in a consignment of 1,00,000 packets. (*MBA, DU, 2003; 2006*)

Solution. Given $p = \dfrac{1}{500},\ n = 10$

$$np = 10 \times \frac{1}{500} = 0.02 = m$$

(*i*) Probability of no defective blade :

$$f(0) = Pr\,[X = 0] = e^{-m} = e^{-0.02} = 0.9802 \qquad [\text{Table value of } e^{-0.02} = 0.9802]$$

Therefore, number of packets containing no defective blade is given as :

$$Nf(0) = 1{,}00{,}000 \times 0.9802 = 98020.$$

(*ii*) Probability of one defective blade :

$$f(1) = Pr\,[X = 1] = me^{-m} = 0.02 \times 0.9802 = 0.019604$$

Therefore, approximate number of packets containing one defective blade is as :

$$Nf(1) = 1{,}00{,}000 \times 0.019604 = 1960.4 \simeq 1960.$$

(*iii*) Probability of two defective blades :

$$f(2) = Pr\,[X = 2] = \frac{m^2 e^{-m}}{2!} = \frac{(0.02)^2\ (0.02)}{2}$$

$$= (0.0004)\,(0.4901) = 0.00019604$$

Therefore, approximate number of packets containing two defective blades is as :

$$Nf(2) = 1{,}00{,}000 \times 0.00019604 = 19.6 \simeq 20$$

Illustration 39. A market researcher at a major automobile company classified households by car ownership. The relative frequencies of the households for each category of ownership are shown in the table :

Number of cars per household	*Relative frequencies*
0	0.10
1	0.30
2	0.40
3	0.12
4	0.06
5	0.02

(*a*) Establish the probability distribution for the random variable.

(*b*) Calculate the expected value of the random variable, and interpret the result.

(*c*) Compute the values of the variance and standard deviation of the probability distribution. *[MBA, DU, 2003]*

Solution. Let the random variable X denote number of cars per household. Therefore, the table shown below gives the required computation data.

No. of Cars/household X	*Rel. Freq.* $P(X)$	$XP(X)$	$X^2 P(X)$
0	0.10	0.00	0.00
1	0.30	0.30	0.30
2	0.40	0.80	1.60
3	0.12	0.36	1.08
4	0.06	0.24	0.96
5	0.02	0.10	0.50
		1.80	4.44

(*a*) The probability distribution for the random variable X is given in the first two columns of the above table.

(*b*) Expected value of the random variable X :

$E(X) = \Sigma X\ P(X) = 1.80$

This expected value is interpreted as that, on the average, there are 1.8 cars per household.

(*c*)
$$\text{Variance} = \sigma^2 = \text{Var}(X) = E(X^2) - [E(X)]^2$$
$$= \Sigma X^2 P(X) - [1.80]^2 \qquad [E(X) = 1.80]$$
$$= 4.44 - 3.24 = 1.20$$

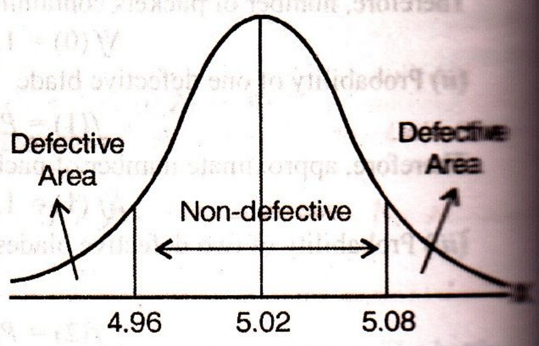
$$\text{Standard deviation} = \sigma = \sqrt{1.20} = 1.095.$$

Illustration 40. The mean inside diameter of a sample of 500 washers produced by a machine is 5.02 mm, and the standard deviation is 0.05 mm. The purpose for which these washers are intended, allows a maximum tolerance in the diameter of 4.96 to 5.08 mm., otherwise the washers are considered defective. Determine the percentage of defective washers produced by the machine assuming the diameters are normally distributed. *(MBA, DU, 2003)*

Solution. Given : $\mu = 5.02, \sigma = 0.05$

Using the Area under the normal curve, the standard normal variate

$$z_1 = \frac{X-\mu}{\sigma}, \text{ therefore } z_1 = \frac{5.08-5.0}{0.05} = 1.2$$



and $$z_2 = \frac{4.96-5.08}{0.05} = -1.2$$

For value of $z = 1.2$, the correspond area from the table is 0.3849. Therefore, the non-defective washer are $0.3849 + 0.3849 = 0.7698$. Hence, the probability of defective washers produced by the machine is $1 - 0.7698 = 0.2302$.

Therefore, the percentage of defective washers is 23.02.

Illustration 41. The demand for a product of a company varies greatly from month to month. The probability distribution in the following table, based on the past 2 years of data, shows the Company's monthly demand :

Unit Demand	*Probability*
300	0.20
400	0.30
500	0.35
600	0.15

(*i*) If the company bases monthly orders on the expected value of the monthly demand, what should monthly order quantity be for this product ?

(*ii*) Assume that each unit demanded generates Rs. 70 in revenue and that each unit ordered costs Rs. 50. How much will the company gain or loss in a month if it places an order based on your answer to part (*i*) and the actual demand for the item is 300 units ?

Solution : (*i*)

Unit Demand X	*Probability* $P(X)$	$XP(X)$
300	0.20	60
400	0.30	120
500	0.35	175
600	0.15	90
		$\Sigma XP(X) = 445$

Expected Value : $E(X) = \Sigma XP(X) = 445$

Therefore, monthly order quantity for the product is 445 units.

(*ii*) Profit = Revenue – Cost = 70 – 50 = Rs. 20

Profit for 445 units = 445 × 20 = Rs. 7250.

Profit for 300 units = 300 × 20 = Rs. 6000

Company Loss = Rs. 7250 – Rs. 6000 = Rs. 1250

Illustration 42. A new automated production process has had an average of 1.5 breakdowns per day. Because of the cost associated with a breakdown, management is concerned about the possibility of having three or more breakdowns, during a day. Assume that breakdowns occur randomly, that the probability of a breakdown is the same for any two time intervals of equal length, and that breakdowns in one period are independent of breakdowns in other periods. What is the probability of having three or more breakdowns during a day ? *(MBA, Delhi Univ., 2003)*

Solution : Given : $m = 1.5, e^{-m} = e^{-1.5} = 0.2231$

$$f(x \geq 3) = 1 - [Pr(x=0) + Pr(x=1) + Pr(x=2)]$$

$$= 1 - [e^{-m} + me^{-m} + \frac{m^2}{2} e^{-m}]$$

$$= 1 - e^{-m}[1 + m + \frac{m^2}{2}]$$

$$= 1 - 0.2231\,[1 + 1.5 + \frac{(1.5)^2}{2}]$$

$$= 1 - 0.2231\,[3.625] = 1 - 0.8088 = 0.1912.$$

45. Assume that the test scores from a college admissions test are normally distributed with a mean of 450 and ... of 100.

(i) ... percentage of the people taking the test score are between 400 and 500 ?

(ii) ... someone received a score of 630. What percentage of the people taking the test score better ? What percentage ... worse ?

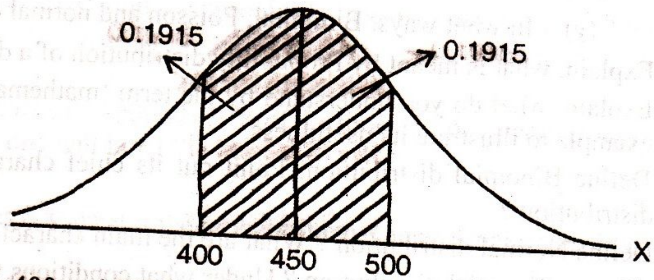
(iii) If a particular university will not admit anyone scoring below 480, what percentage of the persons taking the test would be acceptable to the university ?

Solution : Let X denote the test scores.

$$X \sim N\,(450, 100)$$

(i) $$z_1 = \frac{X-\mu}{\sigma} = \frac{500-450}{100} = 0.5$$

$$z_2 = \frac{400-450}{100} = -0.5$$

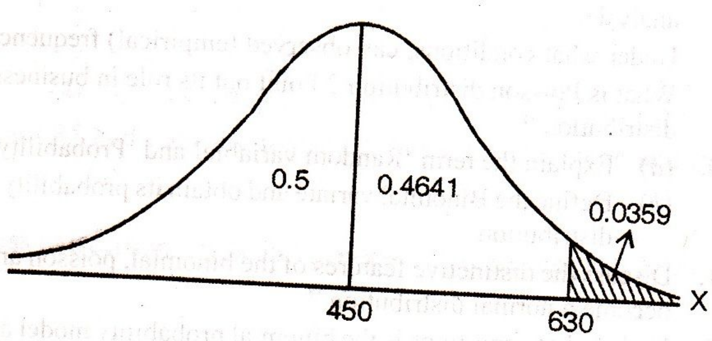


Corresponding to $z = 0.5$, the area is 0.1915.

The required probability $= 0.1915 \times 2 = 0.3830$

Hence, percentage of the people taking the test score between 400 and 500 is 38.30 per cent.

(ii) $$z = \frac{630-450}{100} = 1.8$$

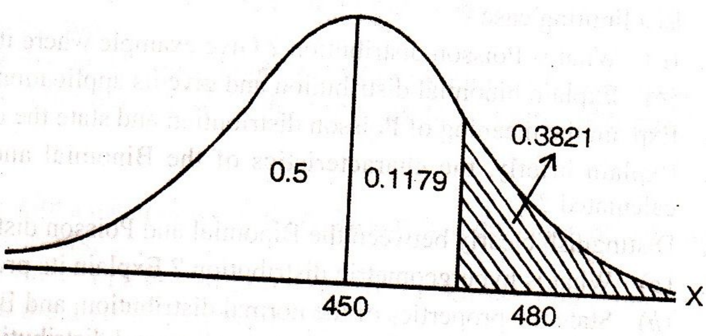


Corresponding to $z = 1.8$, the area is 0.4641.

The required prob $= 0.5 - 0.4641 = 0.0359$ for test score better and required probability $= 1 - 0.0359 = 0.9641$.

Hence, only 3.59 per cent are better and 96.41 per cent are worse.

(iii) $$z = \frac{480-450}{100} = 0.30$$



For $z = 0.30$, the area is 0.1179

required Prob. $= 0.5 - 0.1179 = 0.3821$

Therefore, the per centage of the persons taking the test score acceptable is 38.21.

PROBLEMS

I-A : Answer the following questions, each question carries **one** mark:

(i) Define binomial distribution ?

(ii) What is Bernoulli trial ? *(MBA, Hyderabad Univ., 2006)*

(iii) What is normal distribution ?

(iv) State and explain the constants of binomial distribution ?

(MBA, Madurai-Kamaraj Univ., 2003; M.A. Eco., M.K. Univ., 2003)

(v) Write the parameters of binomial distribution.

(vi) Write the distribution in which mean and variance will be the same.

(vii) Mean value of binomial distribution is ....

(viii) Define normal distribution and list its properties.

(ix) What is normal probability distribution ? *(MBA , UP Tech. Univ., 2006)*

(x) Mean value of Poisson distribution is ....

(xi) What is Poisson distribution ?

(xii) Explain any two properties of normal distribution.

1-B : Answer the following questions, each question carries **four** marks:

(*i*) Distinguish between

(*i*) The mean of a binomial distribution is 20 and standard deviation is 4. Find out n, p, and q.

(*M. Com., M.K. Univ., 2002*)

(*ii*) Define Poisson distribution and state its uses.

(*iii*) Define Binomial distribution. The parameters of a binomial distribution are $n = 10$ and $p = 0.2$. Find the mean and variance of the distribution.

(*MBA, Madras Univ., 2003*)

(*iv*) Give at least five important properties of normal distribution.

(*v*) In what ways, Binomial, Poisson and normal distribution are related.

2. Explain, what is meant by Probability distribution of a discrete random variable.
3. Explain, what do you understand by the term 'mathematical expectation'. How is it useful for a businessman ? Give an example to illustrate its usefulness. (*MBA, Delhi Univ., 2004*)
4. Define Binomial distribution. Point out its chief characteristics and uses. Under what conditions, it tends to Poisson distribution ? (*MBA, Osmania Univ., 2002*)
5. Define Normal distribution ? What are the main characteristics of Normal Distribution?
6. What is Binomial distribution ? Under what conditions will it tend to Normal distribution?
7. What are the chief properties of Normal distribution ? Describe briefly the importance of Normal distribution in statistical analysis.
8. Under what conditions, can observed (empirical) frequency distributions be approximated to binomial distribution ?
9. What is Poisson distribution ? Point out its role in business decision-making. Under what conditions will it tend to normal distribution ? (*MBA, Kumaun Univ., 2006*)
10. (*a*) Explain the term 'Random variable' and 'Probability distribution of a random variable'.

    (*b*) Define the Binomial variate and obtain its probability distribution function. Find the mean and variance of the binomial distribution.
11. Discuss the distinctive features of the binomial, poisson and normal distribution. When does a binomial distribution tend to become a normal distribution ?
12. Under what conditions is the binomial probability model appropriate ? How does it approach the Poisson probability model as a limiting case ?
13. (*a*) What is Poisson distribution ? Give example where it can be applied.

    (*b*) Explain binomial distribution and give its application in business management.
14. Explain the meaning of Poisson distribution and state the conditions under which this distribution is used.
15. Explain briefly, the characteristics of the Binomial and Poisson distributions. How are their means and variances calculated ?
16. Distinguish clearly between the Binomial and Poisson distribution. (*MBA, UP Tech. Univ., 2007*)
17. (*a*) What is hypergeometric distribution ? Explain its properties.

    (*b*) State the properties of the normal distribution, and Binomial distribution. (*MBA, Hyderabad Univ., 2005*)

    (*c*) Describe briefly, the importance of normal distribution in business decision-making. What are its chief properties?
18. (*a*) Briefly describe the characteristics of the normal probability distribution. Why does it occupy such a prominent place in statistics ?

    (*b*) When can Poisson distribution be a reasonable approximation of the binomial?

    (*c*) Fifty per cent of all automobile accidents lead to property damage of Rs. 100. Forty per cent lead to damage of Rs. 500. Ten per cent lead to total loss, a damage of Rs. 1,800. If a car has a 5 per cent chance of being in an accident in a year, what is the expected value of the property damage due to that possible accident ?
19. Suppose that in a lottery 1,000 tickets are sold at Rs. 10 each, and three prizes are to be awarded. The first prize is a television set worth Rs. 12,000, second prize is a short wave radio worth Rs. 1500; and the third prize is a cycle worth Rs. 1300. If you plan to buy one ticket, what is your expected gain or loss from the venture ?
20. Two investment opportunities are open to prospective investor. If opportunity A turns out to be successful, a profit of Rs. 6 lakh will result and the probability of A's success is estimated as 0.75, if A turns out to be a failure there will be a loss of Rs. 1 lakh. If opportunity B succeeds a profit of Rs. 25 lakh will materialize but, if it fails, there will be a loss of Rs. 7 lakh and the probability for B to fail is 0.55. Which investment opportunity should the investor take if the dicision criterion is to maximize profits?

    [Opportunity B]
21. If it rains, a raincoat dealer can earn Rs. 5000 per day. If it is fair, he can lose Rs. 1000 per day. What is his expectation if the probability of rain is 0.4 ?

    [Rs. 1400]

22. A firm plans to bid Rs. 3000 per tonne for a contract to supply 1,000 tonnes of a metal. It has two competitors *A* and *B* and it assumes that the probability that *A* will bid less than Rs. 3000 per tonne is 0.3 and that *B* will bid less than Rs. 3000 per tonne is 0.7. If the lowest bidder gets all the business and the firms bid independently, what is the expected value of the contract to the firm ?
[(*i*) 0.18; (*ii*) 0.32]
23. If the probability that an individual suffers from reaction of a given medicine is 0.001, determine the probability that out of 2,000 individuals (*i*) exactly 3 individuals (*ii*) more than 2 individuals will suffer from reaction.
[(*a*) 0.1353, (*b*) 0.2706, (*c*) 0.2706, (*d*) 0.1804]
24. If 2% of the electric bulbs manufactured by a company are defective, find the probability that in a sample of 100 bulbs (*a*) 0, (*b*) 1, (*c*) 2 and (*d*) 3 bulbs will be defective.
[0.3601]
25. A certain type of plastic bag in the past has burst under a pressure of 10 pounds 30% of the time. If a prospective buyer tests 5 bags chosen at random, what is the probability that exactly one will burst ?
[0.302]
26. The probability that *A* will make a profit on any business deal is 0.8, what is the probability that he will make a profit exactly eight times in ten successive deals ?
27. The distribution of typing mistakes committed by a typist is given below. Assuming a Poisson model, find out the expected frequencies.

Mistakes per page :	0	1	2	3	4	5
No. of pages :	142	156	69	27	5	1

[147.12, 147.12, 73.56, 24.52, 6.13, 1.22] (*MBA, Sukhadia Univ., 2004; Hyderabad Univ., 2006*)
28. In a normal distribution, 7% of the observations are under 35 and 89% are under 63. What are the mean and the standard deviation of distribution ?
[$\mu = 50.28$, $\sigma = 10.25$]
29. If the average number of rejects in the manufacturing process of a certain article is 4 per cent, what are the probabilities of 0, 1, 2, 3, 4 rejects in a sample of 40 articles ?
[0.6703, 0.2681, 0.0536, 0.00715, 0.000715]
30. A Municipal Corporation had installed 5,000 bulbs in the streets of the city. If these bulbs have an average life of 800 burning hours, with a standard deviation of 200 hours, find :
(*i*) What number of bulbs might be expected to fail in the first 600 burning hours ? (*ii*) The number of bulbs expected to fail between 700 and 900 burning hours, and (iii) the number of bulbs expected to fail after 900 burning hours.
[(*i*) 794, (*ii*) 1915, (*iii*) 1542.5]
31. The weekly wages of 1,000 workers are normally distributed with a mean of Rs. 1700 and a standard deviation of Rs. 150. Estimate the lowest weekly wages of the 100 highest paid workers.
32. Find the probability that at most 5 defective bolts will be found in a box of 200 bolts, if it is known that 2 per cent of such bolts are expected to be defective. (You may take the distribution to be Poisson.)
[0.784]
33. If 20% of the bolts produced by a machine are defective, determine the probability that out of 4 bolts (*i*) 0, (*ii*) 1 and (*iii*) at the most 2 bolts will be defective.
[0.4096, 0.4096, 0.9728]
34. The probability that any customer who enters the store will purchase Colgate toothpaste is 0.3. If 1,000 customers enter the store, what is the minimum number of Colgate toothpastes the store must have on hand, if the probability that it will be out of stock is to be at most 1% ?
[334]
35. Daily demand for a product is approximately normally distributed with mean sales of 12 units per day and standard deviation of 4 units. How many units must be on hand in the morning to assure no more than one chance in 5 of running out of stock during the day ? (*MBA, DU, 2003, 2006*)
[16]
36. The probability that India wins a cricket Test match against England is given to be 1/3. If India and England play three Test matches, use binomial distribution to find the probability that :
(*i*) India will lose all three Test matches ? (*ii*) India will win at least one Test match ?
[(*i*) 0.2963, (*ii*) 0.7037]
37. An individual is offered an opportunity to bet Rs. 500 on the outcome of a roll of a pair of dice. If the dice turn up so that the sum of the faces total 7 or 11, the individual wins Rs. 1500. For any other outcome the bet is lost. What is the expected value of the game for the individual ?

38. The fuel consumption of a fleet of 150 trucks is normally distributed with a mean of 15 km per litre and a standard deviation of 1.5 km per litre. Use normal distribution to find the expected number of trucks that average : (*a*) 13 but less than 14 km per litre, (*b*) 14.5 but less than 15.5 km per litre.
[(*a*) 24, (*b*) 39]

39. Fit a Poisson distribution to the data given below :

X :	0	1	2	3	4
f :	123	59	14	3	1

40. The heights of students in a class are normally distributed with a mean of 62 inches and a standard deviation of 4 inches. What proportion of the students in the class have a height greater than 68 inches ? What is the probability that a student selected at random will have a height between 58 inches and 66 inches ?

41. One hundred car radio sets are inspected as they come off the production line and number of defects per set is recorded below :

No. of defects :	0	1	2	3	4
No. of sets :	79	18	2	1	0

Fit a Poisson distribution to the above data.
[77.88, 19.47, 2.43, 0.2028, 0.1267] ***(MBA, Delhi Univ., 2006)***

42. A machine is supposed to drill holes with a diameter of 1 inch. In fact, the diameters are normally distributed with a mean of 1.0 inches and a standard deviation of 0.02 inch. If there is a tolerance of 0.02 inch, the holes should be between 0.99 and 1.02 inches. What percentage of the holes drilled are within tolerance limits ?

43. A hotel maintains two delux rooms. The demand for these rooms in any day is distributed as a Poisson distribution with mean 1.5. Calculate the proportion of day on which neither of the rooms would be used ; and the proportion of days on which no demand would be refused.
[0.2231, 0.1913]

44. The following frequency table gives the distribution of 1,000 persons according to their income :

Monthly income (Rs.)	*No. of persons*	*Monthly income (Rs.)*	*No. of persons*
Below 5000	16	20000–25000	166
5000–10000	85	25000–30000	100
10000–15000	207	30000–35000	69
15000–20000	346	above 35000	11

Fit a normal distribution to the above frequency table. Also, determine the percentage of persons with an income between 17,500 and 27,500 rupees.

45. A car rental firm has two cars which it rents out day by day. The number of demand for a car on each day is distributed as Poisson distribution with mean 1.5. Calculate the proportion of days on which neither car is used and the proportion of days on which some demand is refused.

46. A dice is thrown 9,000 times and a throw of 3 or 4 is observed 3,240 times. Show that the dice cannot be regarded as an unbiased one and find the limits between which the probability of a throw of 3 or 4 is expected to lie.

47. An automatic detergent packing machine produces packages whose weights are normally distributed with a mean of 8.00 gm and a standard deviation of 0.010 gm.
(*a*) What proportion of packages are between 7.98 gm and 8.000 gm ?
(*b*) What proportion are between 8.005 and 8.0151 gm ?
(*c*) What proportion are between 7.995 and 8.010 gm ?
(*d*) What proportion are above 8.017 gm ?
[(*a*) 47.72 (*b*) 24.30 (*c*) 53.28 (*d*) 4.46.]

48. A manufacturer of electric fuses packs fuses in boxes of 10 each and 2,000 such boxes were sold. The previous experience shows that 5 per cent of the fuses are defective. Using Poisson distribution, find how many boxes will contain (*i*) no defective (*ii*) more than one defective.

49. In a certain factory turning out razor blades, there is a small chance of 1/500 for any blade to be defective. The blades are supplied in packets of 10. Use Poisson distribution to calculate the number of packets containing (*i*) no defective (*ii*) one defective and (*iii*) two defective blades, respectively in a consignment of 10,000 packets.

50. If, on an average 8 ships out of 10 arrive safely at port, find the mean and standard deviation of the number of ships arriving safely out of a total of 1,600 ships.
[$\mu = 1280, \sigma = 16$]

51. Assuming that sex ratio of male children is $\frac{1}{2}$, find the probability that in a family of 5 children, (*i*) all children will be of the same sex, and (*ii*) three of them will be boys and two girls.
[(*i*) 1/16 (*ii*) 5/16]

52. A company has 6 telephones which 10 executives use intermittently. Assume that at any given time each executive has the same probability '*p*' of requiring to use a telephone. If the executives' requirements of telephones are independent, the probability that exactly *k* executives require a phone is $b(k, n, p)$. If on an average, an executive uses the telephone for 10 minutes per hour ($p = 1/6$), find the probability that 7 or more executives need a telephone at the same time.
[0.000267]

53. A machine produces bolts which are 10% defective. Find the probability that in random sample of 400 bolts produced by this machine, the number of defectives found
(*i*) will be at most 30;
(*ii*) will be between 30 and 50;
(*iii*) will exceed 55.
[(*i*) 22.8, (*ii*) 354.32, (*iii*) 3.56]

54. The mean and standard deviation for the life times of a population of light bulbs are 1200 and 150 hours respectively. Assuming these lifetimes are normally distributed, what is the probability that a light bulb will last over 1500 hours ?
[0.0228]

55. An editor of a publishing company, calculates that it requires 11 months on an average to complete the publication process from manuscript to finished books with a standard deviation of 2.4 months. He believes that the distribution of publication times is well described by the normal distribution. Out of 190 books he will handle this year, how many will complete the process in less than a year ?
[126]

56. An analyst predicts that 2.5% of all small companies will file for bankruptcy in the coming year. For a random sample of 200 companies, estimate probability that
(*i*) at least three will file for bankruptcy next year ;
(*ii*) exactly three will file for bankruptcy ;
(*iii*) not more than five will file for bankruptcy.
[(*i*) 0.87 (*ii*) 0.145 (*iii*) 0.6151]

57. Past records show that the average number of accidental drownings at a beach resort is 3 per year for every 100,000 of tourists visiting the resort. If in a year 200,000 tourists visited this resort, find the probabilities that :
(*i*) there will be no drowning accident this year ;
(*ii*) there will be at least 6 accidents this year ;
(*iii*) there will be exactly 5 accidents this year ;
(*iv*) there will be more than 8 accidents this year.
[(*i*) 0.00279 (*ii*) 0.498 (*iii*) 0.1807 (*iv*) 0.0465]

58. Fit a binomial distribution to the following data :

X:	0	1	2	3	4
f:	28	62	28	12	46

[9.34, 40.48, 65.78, 47.51, 12.86]

59. The financial controller of Galaxy Airlines is having some problems with cash flows. Daily revenue fluctuate greatly and are difficult to predict whereas daily expenses remain fairly constant regardless of the daily number of passengers. If daily revenue has a normal distribution with a mean of Rs. 72,000 and 85 per cent of the values lie below Rs. 82,000, what is the standard deviation of the distribution. What is the value above which 5 per cent of the values in the distribution lie ?

60. The following table shows the number of customers returning the products in a marketing territory. The data is for 100 stores :

No. of returns	:	0	1	2	3	4	5	6
No. of stores	:	4	14	23	23	18	9	9

Fit a Poisson distribution.
[4.97, 14.91, 22.36, 22.36, 16.77, 10.06, 5.03]

61. Three fair coins are tossed 300 times. Find the frequencies of the distribution of heads and tails and tabulate the result. Also, calculate the mean and standard deviation of the distribution.
[2.25, 1.06]

62. How many workers have a salary above Rs. 2,675 in the distribution whose average salary is Rs. 2,400 and standard deviation is Rs. 100 and the number of workers in the factory is 15,000, if the salary of workers follows the normal law ?
[4368]

63. The Delhi Municipal Corporation installed 2,000 bulbs in the streets of Kailash Colony. If these bulbs have an average life of 1,000 burning hours, with a standard deviation of 200 hours, what number of bulbs might be expected to fail in the first 700 burning hours?
[134]
(*MBA, HPU*)

64. In 24 trials of an event of small probability, the frequency f of the number of success X is given in the following table :

X	:	0	1	2	3	4	5	6
f	:	3	2	6	5	5	1	2

The mean number of successes is 2.75. Find the expected frequencies of the Poisson distribution with the same total frequency.

65. The appearing of 2 or 3 on a dice is connected as success. Five dice are thrown 729 times and the following results are observed :

Number of successes	:	0	1	2	3	4	5
Frequency	:	45	195	237	132	81	39

Fit the binomial distribution assuming the dice to be unbiased.
[96, 240, 240, 120, 30, 3]

66. Five hundred T.V. sets are inspected as they come off the production line and average number of defects per set is found to be 0.402. Find the expected number of T.V. sets having one or more defects.
[165.5]

67. The time required by bank cashier to deal with a customer has been observed to be normally distributed with mean 25 secs. and a standard deviation 10 secs. Find the probability that a customer arriving at random will have to wait :
(*i*) between 20 and 28 secs. ;
(*ii*) less than 23 secs.
To what value should the mean service time be altered so that only 1 customer in 100 has to wait longer than 50 secs. ?
(*i*) 0.3094 (*ii*) 0.4207. 26.7

68. The marks obtained by the students in an examination are known to be normally distributed. If 10% of the students got less than 40 marks while 15% got over 80, what are the mean and standard deviation of marks ? (*MBA, Delhi Univ., 2006*)
[$\mu = 62.0688$, $\sigma = 17.2413$]

69. In a certain examination, 10% of the students got less than 30 marks and 97% of the students got less than 62 marks. Assuming the distribution to be normal, find the mean and standard deviation of the marks.
[$n = 42.96$, $a = 10.13$]

70. For a binomial distribution the mean is 4 and variance 2. Find the probability of getting
(*i*) at least 2 successes.
(*ii*) at most two successes. (*MBA., DU, 2002, 2007*)
[(*i*) 0.9648 (*ii*) 0.1445]

71. The following table gives the numbers of days in a 50 day period in which automobile accidents occurred in a certain part of a city. Fit a Poisson distribution to the data.

No. of accidents	:	0	1	2	3	4
No. of days	:	19	18	8	4	1

(18.4, 18.4, 9.2, 3.1, 0.8]

72. In a book, the following frequency of mistakes per page was observed. Fit a Poisson distribution.

No. of mistakes per page	:	0	1	2	3	4	5
No. of pages	:	630	160	90	70	30	20

[463, 357, 137, 35, 7, 1]

73. The distribution of monthly income of 4000 employees follows normal distribution with mean Rs. 6000 and standard deviation Rs. 1000 find :
(*i*) Number of employees having income more than Rs. 7,000;
(*ii*) The number of employees having income less than Rs. 5500;
(*iii*) The least monthly income among the highest paid 100 employees.

74. (*a*) Proof reading of 200 pages of a book containing 500 pages gave the following results :

No. of mistakes per page	:	0	1	2	3	4	5
Frequency	:	113	62	20	3	1	1
Cost per page for checking	:	1.0	1.5	2.5	3.0	3.5	4.0

(*a*) Fit a Poisson distribution.
(*b*) Estimate the total cost of correcting the whole book.
[(*a*) 109.76, 65.85, 19.15, 3.95, 0.592, 0.0711 (*b*) 272.139]

75. Which probability distribution is most likely the appropriate one to use for the following data : Binomial, Poisson or Normal ?
(*i*) The life span of a female born in 1957.
(*ii*) The number of autos passing through a toll booth.
(*iii*) The number of defective radios in a lot of 100.
(*iv*) The water level in a reservoir.

76. A book has 700 pages. The number of pages with various number of misprints is recorded below. Fit a Poisson distribution to the given data :

Number of Misprints X:	0	1	2	3	4	5
Number of Pages with X misprints :	616	70	10	2	1	1

(*M.Com., DU, 1999*)

77. An insurance salesman sells policies to 5 men, all of an identical age and in good health. According to actuarial tables, the probability that a man of this particular age will be alive 30 years hence is 2/3. Find the probability that 30 years hence :

(*i*) at least 1 man will be alive.

(*ii*) at least 3 men will be alive. (*MBA, IGNOU, 2002*)

78. (*a*) The local authorities in a certain city install 10,000 electric lamps in the streets of the city. If these lamps have an average life of 1000 burning hours with a standard deviation of 200 hours, assuming normality, what number of lamps might be expected to fail in the first 800 burning hours ? (*MBA, IGNOU, 2004*)

(*b*) In certain organisation out of 400 employees 150 are married. Find the probability that exactly 2 of the 3 randomly chosen employees are unmarried. The purchase department has 10 employees. Find the probability that exactly 4 employees of the department are married. (*MBA, Bharathidasan Univ., April 2003*)

79. In an examination, it is laid down that a student passes if he secures 30% or more marks. He is placed in the first, second or third division according as he secures 60% or more marks, between 45% and 60% marks and marks between 30% and 45% respectively. He gets a distinction in case he secures 80% or more marks. It is noticed from the results that 10% of the students failed in the examination, whereas 5% of them obtained distinction. Calculate the percentage of the students placed in second division (Assume normal distribution). (*MBA, IGNOU, 2003*)

80. A T.V. manufacturer is facing the problem of selecting a supplier of cathode ray tube, which is the most vital component of a T.V. set. Three foreign suppliers, all equally dependable have agreed to supply the tubes. The prices per tube and the expected life of a tube for the three suppliers are as follows :

	Price/Tube	*Expected life of tube*
Supplier 1	Rs. 800	1500 hrs.
Supplier 2	Rs. 1000	2000 hrs.
Supplier 3	Rs. 1500	4000 hrs.

The manufacturer guarantees its customers that it will replace the T.V. set if the tube fails earlier than 1000 hours. Such a replacement would cost him Rs. 1000/tube, over and above the price of the tube. Can you help the manufacturer to select a supplier ? (*MBA, IGNOU, 2006*)

81. (*a*) A duplicating machine maintained for office use is operated by an office assistant who earns Rs. 50 per hour. The time to complete each job varies according to an exponential distribution with mean 6 minutes. Assume a Poisson input with an average arrival rate of 5 jobs per hour. If an 8 hours day is used as a base, determine the percentage idle time of the machine and the average time a job is in the system. (*B.E./ B.Tech., Madras Univ., 2003*)

(*b*) In a survey with a sample of 300 respondents, the monthly income of the respondents follows normal distribution with its mean and standard deviation as Rs. 15,000 and Rs. 3,000 respectively. Answer the following :

(*i*) What is the probability that the monthly income is less than Rs. 12,000 ? Also, find the number of respondent having income less than Rs. 12,000.

(*ii*) What is the probability that the monthly income is more than Rs. 16,000 ? Also, find the number of respondents having income more than Rs. 16,000.

(*iii*) What is the probability that the monthly income, is in between Rs. 10,000 and Rs. 17,000 ? Also, find the number of respondents having income in between Rs. 10,000 and Rs. 17,000. (*MBA, Bharathidasan Univ., 2005*)

(*c*) The mean weight of a lunch rice pack is 0.255 kg with a standard deviation of 0.025. The random variable weights of the pack follows a normal distribution.

(*i*) What is the probability that pack weighs less than 0.280 Kg. ?

(*ii*) What is the probability that pack weighs more than 0.350 Kg. ?

(*iii*) What is the probability that the pack weighs between 0.260 Kg. and 0.340 Kg. ? (*M.A. Eco., M.K. Univ., 2007*)

*****