STATISTICAL CONSIDERATIONS

5.1 Stochastic Activities

The description of activities can be of two types, deterministic and stochastic. The process in which the outcome of an activity can be described completely in terms of its input is deterministic and the activity is said to be deterministic activity. On the other hand, when the outcome of an activity is random, that is there can be various possible outcomes, the activity is said to be stochastic. In case of an automatic machine, the output per hour is deterministic, while in a repair shop, the number of machines repaired will vary from hour to hour, in a random fashion. The terms random and stochastic are used interchangeably.

A stochastic process, can be defined as an ordered set of random variables, the ordering of the set usually being with respect to time. The variation in the ordered set may be discrete or continuous.

A random variable x is called **discrete** if the number of possible values of x (*i.e.*, range space) is finite or countably infinite, i.e., possible values of x may be $x_1, x_2, \ldots, x_n$. The list terminates in the finite case and continues indefinitely in the countable infinite case. The demand of an item can be say 0, 1, 2, 3 or 4 per day, each having its own probability of occurrence. This demand is a discrete activity. Colour of a traffic signal light encountered by a randomly arriving vehicle may be red, amber or green, is a discrete activity. A random variable is called **continuous** if its range space is an interval or a collection of intervals. A continuous variable can assume any value over a continuous range. For example, heights of school children, temperatures and barometric pressures of different cities and velocity of wind are examples of continuous variables.

A stochastic process is described by a probability law, called probability density function.

5.2 Discrete Probability Functions

If a random variable x can take x_i $(i = 1, n)$ countably infinite number of values, with the probability of value x_i being $p(x_i)$, the set of numbers $p(x_i)$ is said to be a probability distribution or probability mass function of the random variable x. The numbers $p(x_i)$ must satisfy the following two conditions:

(*i*) $p(x_i) \geq 0$ for all values of i.

(*ii*) $\sum_{i=1}^{n} p(x_i) = 1$

The probability distribution may be a known set of numbers. For example, in case of a dice, the probability of each of the six faces is 1/6.

x	1	2	3	4	5	6
$p(x)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

If two coins are flipped together, the probabilities of both heads, one head one tail and both tails are $\frac{1}{4}$, $\frac{1}{2}$ and $\frac{1}{4}$ respectively.

Example 5.1: A pair of fair dice is rolled once. The sum of the two numbers on the dice represents the outcome for a random variable x. Determine the probability distribution of x. What is the probability that x is odd?

[P.U. B.E. (ECE) April 2004]

Solution: The sum of the two numbers on the dice, x will range from $1 + 1 = 2$ to $6 + 6 = 12$, or the range of x is 2 to 12 both inclusive.

$x = 2$ can occur in **one** way $(1 + 1)$

$= 3$ can occur in **two** ways $(1 + 2, 2 + 1)$

$= 4$ can occur in **three** ways $(1 + 3, 2 + 2, 3 + 1)$

$= 5$ can occur in **four** ways $(1 + 4, 2 + 3, 3 + 2, 4 + 1)$

$= 6$ can occur in **five** ways $(1 + 5, 2 + 4, 3 + 3, 4 + 2, 5 + 1)$

$= 7$ can occur in **six** ways $(1 + 6, 2 + 5, 3 + 4, 4 + 3, 5 + 2, 6 + 1)$

$= 8$ can occur in **five** ways $(2 + 6, 3 + 5, 4 + 4, 5 + 3, 6 + 2)$

$= 9$ can occur in **four** ways $(3 + 6, 4 + 5, 5 + 4, 6 + 3)$

$= 10$ can occur in **three** ways $(4 + 6, 5 + 5, 6 + 4)$

$= 11$ can occur in **two** ways $(5 + 6, 6 + 5)$

$= 12$ can occur in **one** way $(6 + 6)$

Total number of ways $= 1 + 2 + 3 + 4 + 5 + 6 + 5 + 4 + 3 + 2 + 1 = 36$

Thus the probability of x being 2 is $\frac{1}{36}$, being 3 is $\frac{2}{36}$ etc. The distribution is given in Table 5.1 below:

Table 5.1

x :	2	3	4	5	6	7	8	9	10	11	12
$p(x)$:	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

Probability that x is an odd number is,

$$= p\,(x = 3) + p\,(x = 5) + p\,(x = 7) + p\,(x = 9) + p\,(x = 11)$$

$$= \frac{2}{36} + \frac{4}{36} + \frac{6}{36} + \frac{4}{36} + \frac{2}{36} = \frac{18}{36} = \frac{1}{2} \text{ or } 50\%$$

5.3 Cumulative Distribution Function

It is a function which gives the probability of a random variable being less than or equal to a given value. In the discrete case, the cumulative distribution function is denoted by $P(x_i)$. This function implies that x takes values less than or equal to x_i.

Let us consider the following data, which pertains to the demand of an item by the customers. Total 200 customers demand has been recorded.

Table 5.2

Number of items demanded (x)	*Number of customers*	*Probability Distribution*	*Cumulative Distribution*
0	7	.035	.035
1	25	.125	.160
2	50	.250	.410
3	72	.360	.770
4	33	.165	.935
5	13	.065	1.000

The probability of zero demand is $\frac{7}{200} = .035$, while that of one item is $\frac{25}{200} = .125$, and that of two items is $\frac{50}{200} = .250$, and so on. The cumulative probability, $P(2) = p(0) + p(1) + p(2) = .035 + .125 + .250 = .410$. The probability distribution and the cumulative probability distribution are given in third and fourth columns of Table 5.2.

The probability distribution (p.m.f.) and the cumulative probability distribution (CDF) are commonly represented graphically. The data of Table 5.2, is shown plotted in Fig. 5.1. Since, p.m.f. is defined at discrete points only, the CDF is a step function type of graph.

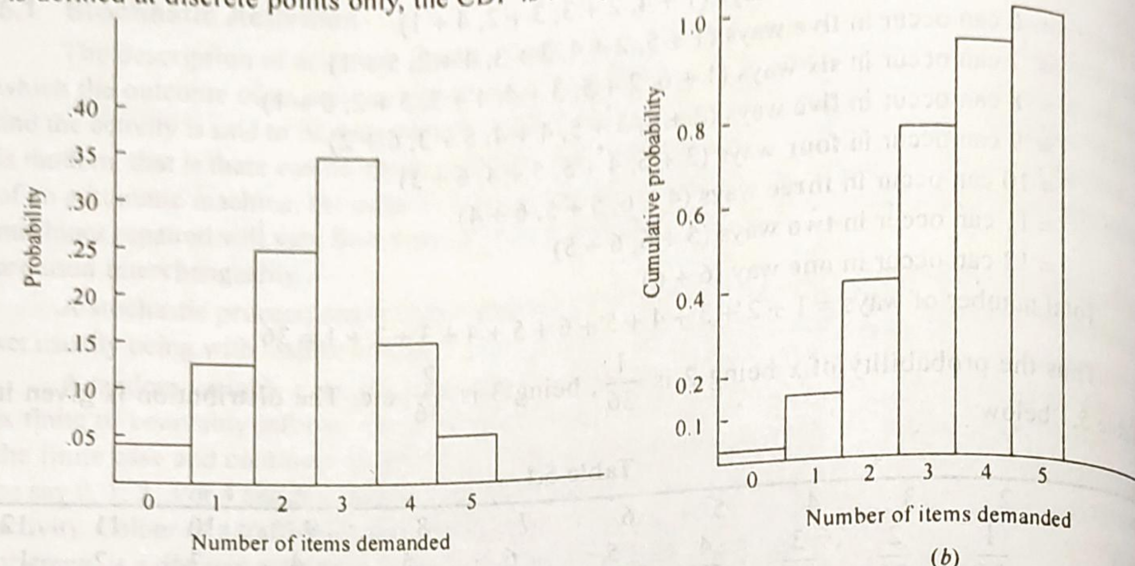


Fig. 5.1

Example 5.2. In Example 5.1, the probability distribution of a variable x, was determined. Represent the same as well as the cumulative distribution in the form of a graph.

Solution: The probability density function obtained in Example 5.1, and the CDF are as follows:

x	2	3	4	5	6	7	8	9	10	11	12
$p(x)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$
$P(x)$	$\frac{1}{36}$	$\frac{3}{36}$	$\frac{6}{36}$	$\frac{10}{36}$	$\frac{15}{36}$	$\frac{21}{36}$	$\frac{26}{36}$	$\frac{30}{36}$	$\frac{33}{36}$	$\frac{35}{36}$	$\frac{36}{36}$

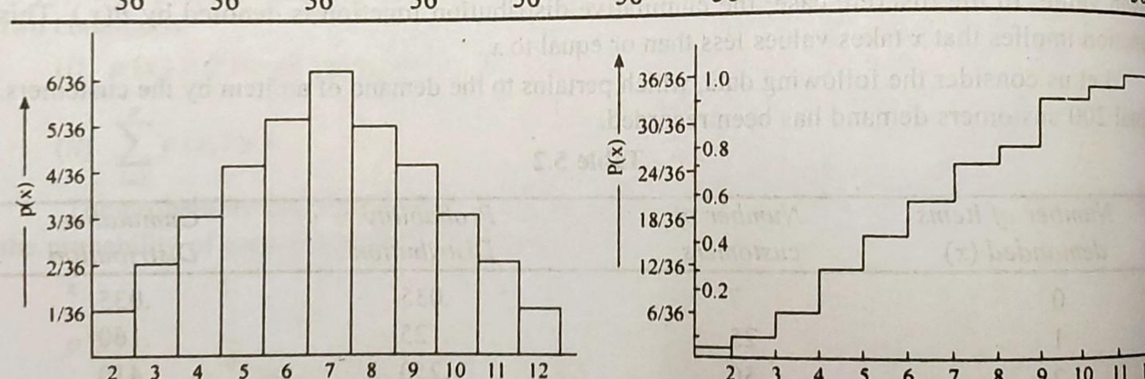


Fig. 5.2

The probability mass function p.m.f. and cummulative density function C.D.F. for the above data are given in Fig. 5.2.

5.4 Continuous Probability Functions

If the random variable is continuous and not limited to discrete values, it will have infinite number of values in any interval, howsoever small. Such a variable is defined by a function $f(x)$ called a probability density function (p.d.f.). The probability that a variable x, falls between x and $x + dx$ is expressed as $f(x)\,dx$, and the probability that x falls in the range x_1 to x_2 is given by

$$p(x) = \int_{x_1}^{x_2} f(x)\,dx$$

If x is a continuous variable on the range $-\infty$ to ∞ its p.d.f. $f(x)$ must satisfy the following two conditions :

(*i*) $f(x) \geq 0 \qquad -\infty < x < \infty$ (*ii*) $\int_{-\infty}^{\infty} f(x)\,dx = 1$

Example 5.3. If x is a random variable with the following distribution, $f(x) = xe^{-x}, x \geq 0$ show graphically its p.d.f. and C.D.F.

Solution: Values of the function $f(x)$, corresponding to various values of x are computed below.

Table 5.3

x	$f(x) = xe^{-x}$	p.d.f.	C.D.F.
0	0	0	0
.2	.164	.042	.042
.4	.268	.069	.111
.6	.329	.085	.196
.8	.359	.093	.289
1.0	.368	.095	.384
1.2	.361	.093	.477
1.4	.345	.089	.566
1.6	.323	.084	.650
1.8	.298	.077	.727
2.0	.271	.070	.797
2.4	.218	.056	.853
2.8	.170	.044	.897
3.2	.130	.034	.931
3.6	.098	.025	.956
4.0	.073	.619	.975
4.6	.046	.012	.987
5.2	.029	.008	.995
6.0	.015	.004	.999
	3.865		

The probability density function $f(x)$ and the cumulative distribution function $F(x)$ are given in Figs. 5.3 and 5.4 respectively.

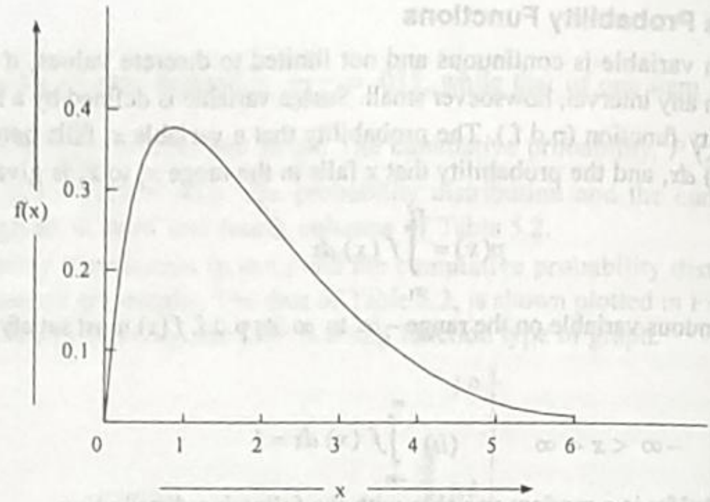


Fig. 5.3

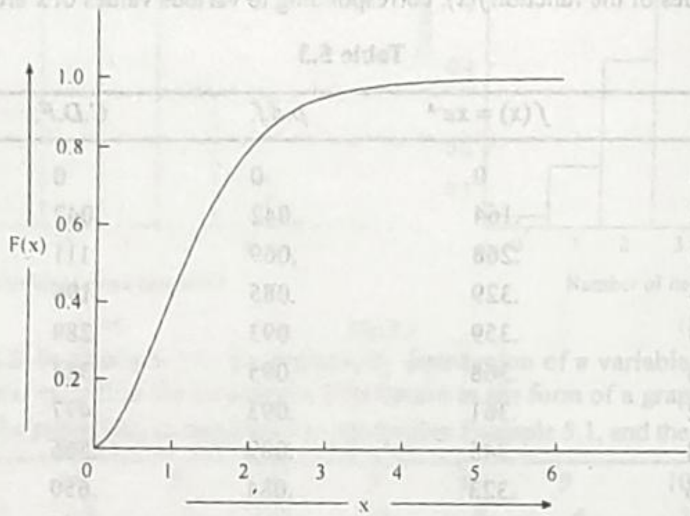


Fig. 5.4

5.5 Measures of Probability Function

The two important characteristics of a probability density function are the central tendency and dispersion.

5.5.1. Central Tendency

Three most important measures of central tendency are mean, mode and median.

Mean

If x is a random variable, then its expected value $E(x)$ itself is called the mean or average value of x, and is identified as $\bar{x}$. In the case of a discrete distribution, if there are N observations, taking the individual values x_i ($i = 1,2, \ldots, N$), the mean is given by

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{i=N} x_i$$

If the N observations are divided into I groups, where the ith group takes the value x_i and has n_i observations, then mean is given by

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{i=I} n_i x_i$$

In case the probability $p(x_i)$ of occurrence of each observation x_i is given, then the mean is given by,

$$\bar{x} = \sum p(x_i)\, x_i$$

For a continuous variable, the mean value is defined as,

$$\bar{x} = \int_{-\infty}^{\infty} x\, f(x)\, dx,$$

where $f(x)$ is the p.d.f. of x *i.e.* $\int_{-\infty}^{\infty} f(x)\, dx = 1$

Mode

Mode represents the most frequently occurring value of a random variable. In other words, when the probability density function has a peak the value of x at which the peak occurs is called mode. A distribution may have more than one peaks, that is may be multimodal (Fig. 5.5). Depending upon the number of peaks, it may be called unimodal, bimodal, trimodal etc. The highest peak, that is the most frequently occurring value is then called the mode of the distribution.

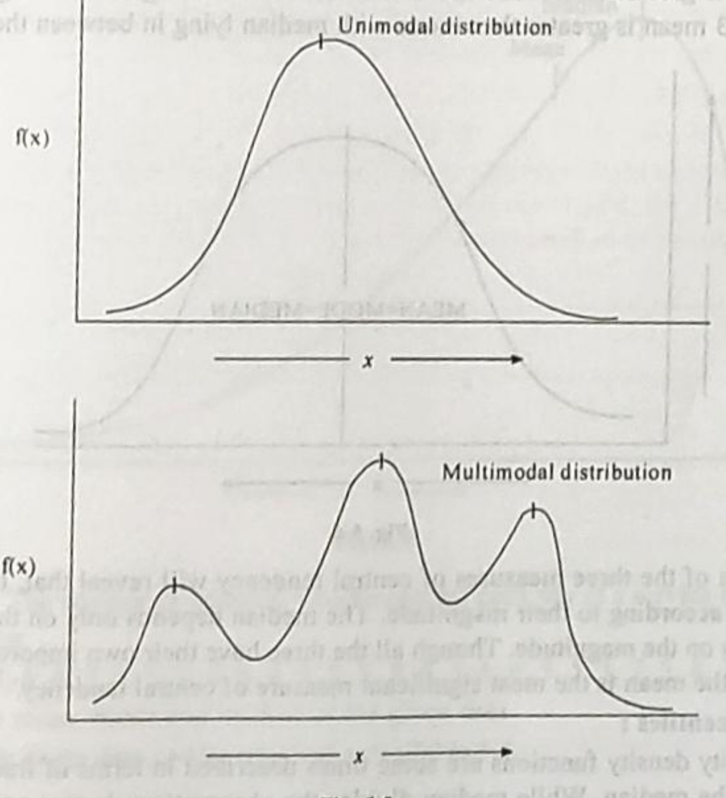


Fig. 5.5

In case of discrete distribution, mode is determined by the following inequations:

$$p\,(x = x_i) \leq p\,(x = \hat{x}),\ x_i \leq \hat{x}$$

and $$p\,(x = x_j) \leq p\,(x = \hat{x}),\ x_j \leq \hat{x}$$

For a continuous distribution, mode is determined as,

$$\frac{d}{dx}[f\,(x)] = 0$$

$$\frac{d^2}{dx^2}[f\,(x)] = 0$$

Median

Median divides the observations of the variable in two equal parts. Half the values of a random variable will fall below the median, and half above the median. For a discrete or continuous distribution of random variable x, if X denotes the median, then

$$p\,(x \leq X) = p\,(x \geq X) = 0.5$$

The median is easily found from the cumulative distribution, since it is the point at which $P(x) = 0.5$. In Fig. 5.3, the median corresponds to $x = 1.28$.

If any two of the mean, mode and median are known, the third can be computed by the following relation.

$$\boxed{\text{Mean} - \text{Mode} = 3\,(\text{Mean} - \text{Median})}$$

The relative values of mean, mode and median depend upon the shape of the probability density function. If the probability distribution curve is symmetric about the center and is unimodal as in Fig. 5.6, then all the three mean, mode and median will coincide. If the curve is skewed to the left as in Fig. 5.7, mean is greater than mode and if it is skewed to the right mode is greater than mean (Fig. 5.8). In Fig. 5.3 mean is greater than mode with median lying in between the two.

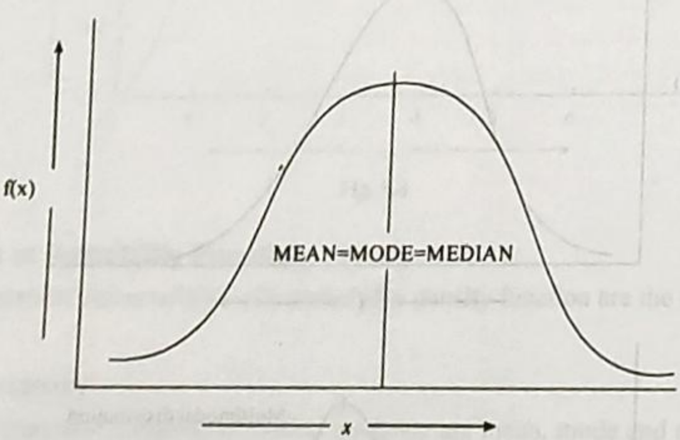


Fig. 5.6

A comparison of the three measures of central tendency will reveal that, the mean involves weighting the data according to their magnitude. The median depends only on the order, while the mode depends only on the magnitude. Though all the three have their own importance in describing a distribution, but the mean is the most significant measure of central tendency.

Fractiles and Percentiles :

The probability density functions are some times described in terms of fractiles, which are a generalization of the median. While median divides the observations in two equal parts, fractiles divide the data into a number of parts. For example, ten fractiles, divide the data into ten parts as the first tenth, second tenth and so on. The fractiles taken as percentage are called percentiles.

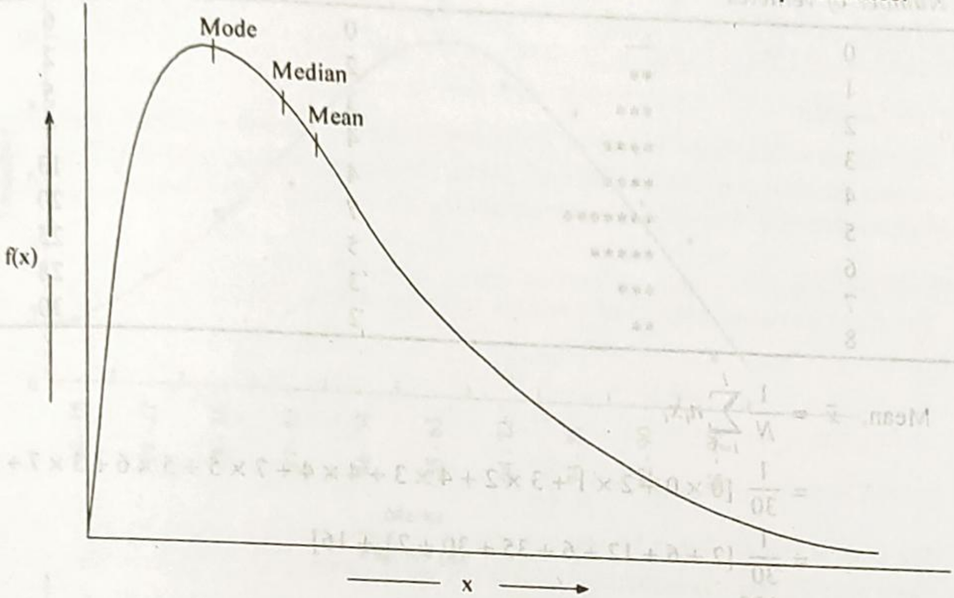


Fig. 5.7

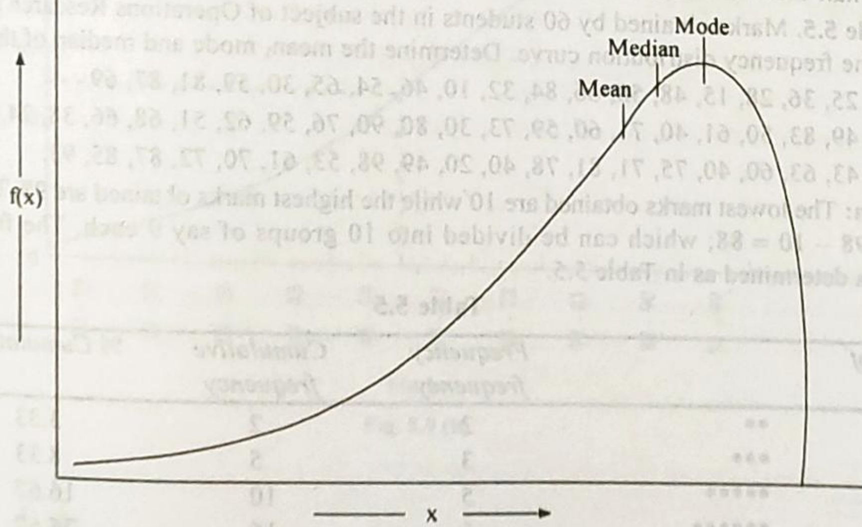


Fig. 5.8

Example 5.4. The number of vehicles, which reported each day at a service station during June 1999, is given below.

3, 2, 6, 1, 5, 4, 6, 8, 7, 5, 3, 2, 4, 6, 4, 5, 2, 5, 3, 6, 5, 4, 6, 5, 5, 1, 7, 3, 8, 7.

Determine the mean, mode and median of the given data.

Solution : The given data can be grouped as in Table 5.4.

Table 5.4

Number of vehicles		*Frequency*	*Cumulative frequency*
0	—	0	0
1	**	2	2
2	***	3	5
3	****	4	9
4	****	4	13
5	*******	7	20
6	*****	5	25
7	***	3	28
8	**	2	30

$$\text{Mean, } \bar{x} = \frac{1}{N}\sum_{i=1}^{I} n_i x_i$$

$$= \frac{1}{30}[0 \times 0 + 2 \times 1 + 3 \times 2 + 4 \times 3 + 4 \times 4 + 7 \times 5 + 5 \times 6 + 3 \times 7 + 2 \times 8]$$

$$= \frac{1}{30}[2 + 6 + 12 + 6 + 35 + 30 + 21 + 16]$$

$$= \frac{138}{30} = 4.6$$

Mode, the most frequently occurring value is 5 or the mode $\hat{x} = 5$.

Median: It is the value of the number of vehicles corresponding to the cumulative frequency of 15, that is half the total number of vehicles, which in this case is slightly more than 4.

Example 5.5. Marks obtained by 60 students in the subject of Operations Research are given below. Plot the frequency distribution curve. Determine the mean, mode and median of the marks.

55, 25, 36, 28, 15, 48, 58, 66, 84, 32, 10, 46, 54, 65, 30, 59, 81, 87, 69
78, 49, 83, 50, 61, 40, 71, 60, 59, 73, 30, 80, 90, 76, 59, 62, 51, 68, 66, 38, 24,
38, 43, 63, 60, 40, 75, 71, 81, 78, 40, 20, 49, 98, 53, 61, 70, 72, 87, 85, 93.

Solution: The lowest marks obtained are 10 while the highest marks obtained are 98. The range of marks is 98 – 10 = 88; which can be divided into 10 groups of say 9 each. The frequency distribution is determined as in Table 5.5.

Table 5.5

Range of marks		*Frequency frequency*	*Cumulative frequency*	*% Cumulative*
10–18	**	2	2	3.33
19–27	***	3	5	8.33
28–36	*****	5	10	16.67
37–45	******	6	16	26.67
46–54	********	8	24	40.00
55–63	***********	11	35	58.33
64–72	*********	9	44	73.33
73–81	********	8	52	86.67
82–90	******	6	58	96.67
91–99	**	2	60	100.00

The probability density function is plotted in Fig. 5.9

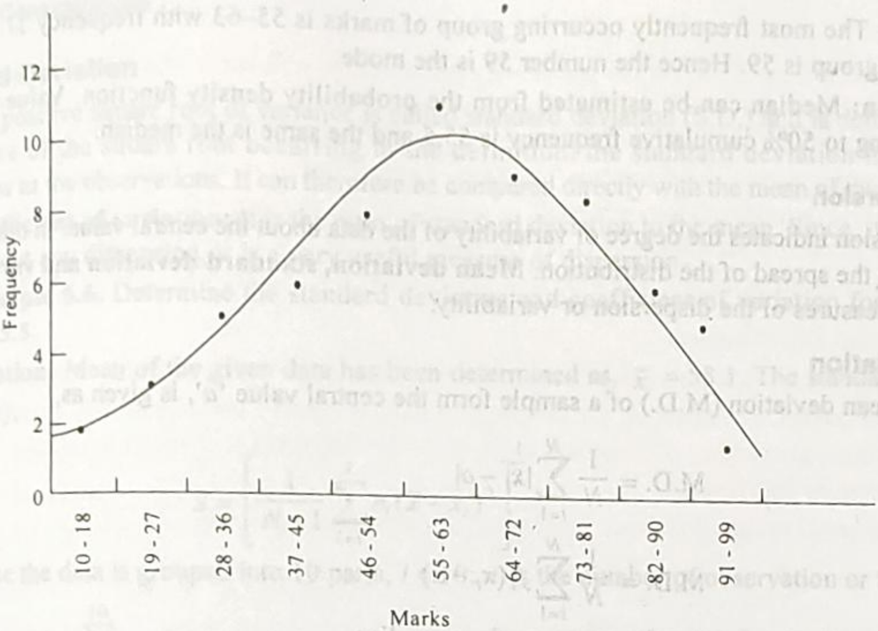


Fig. 5.9 (a)

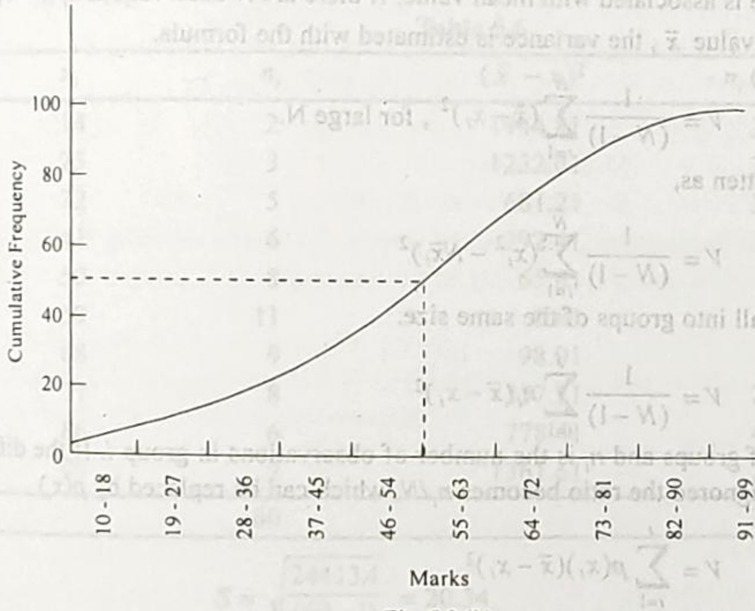


Fig. 5.9 (b)

Mean

$$\bar{x} = \frac{1}{60}\sum_{i=1}^{I} x_i n_i$$

$$\bar{x} = \frac{1}{60}\sum_{i=1}^{I} x_i n_i =$$

$$[14 \times 2 + 23 \times 3 + 32 \times 5 + 41 \times 6 + 50 \times 8 + 59 \times 11 + 68 \times 9 + 77 \times 8 + 86 \times 6 + 95 \times 2]$$

$$= \frac{3486}{60} = 58.1$$

Mode: The most frequently occurring group of marks is 55–63 with frequency 11. The mid point of the group is 59. Hence the number 59 is the mode.

Median: Median can be estimated from the probability density function. Value of marks corresponding to 50% cumulative frequency is 55.5 and the same is the median.

5.5.2 Dispersion

Dispersion indicates the degree of variability of the data about the central value. In other words, it represents the spread of the distribution. **Mean deviation, standard deviation and variance** are important measures of the dispersion or variability.

Mean Deviation

The mean deviation (M.D.) of a sample form the central value '*a*', is given as,

$$\text{M.D.} = \frac{1}{N}\sum_{i=1}^{N}|x_i - a|$$

$$\text{M.D.} = \frac{1}{N}\sum_{i=1}^{N}f_i(x_i - a)$$

The central value '*a*' may be mean, mode or median.

Variance: Variance is associated with mean value. If there are N observations x_i ($i = 1, 2 \ldots\ldots, N$) and having the mean value $\bar{x}$, the variance is estimated with the formula.

$$V = \frac{1}{(N-1)}\sum_{i=1}^{N}(\bar{x} - x_i)^2 \text{, for large N.}$$

which can further be written as,

$$V = \frac{1}{(N-1)}\sum_{i=1}^{N}(x_i^2 - N\bar{x}_i)^2$$

when the observations fall into groups of the same size.

$$V = \frac{1}{(N-1)}\sum_{i=1}^{I}n_i(\bar{x} - x_i)^2$$

where I is the number of groups and n_i is the number of observations in group i. If the difference between N and $N - 1$ is ignored the ratio becomes n_i/N, which can be replaced by $p(x_i)$

$$V = \sum_{i=1}^{I}p(x_i)(\bar{x} - x_i)^2$$

Since $\sum_{i=1}^{I}p(x_i) = 1$ and $\sum_{i=1}^{I}p(x_i)\,x_i = \bar{x}$

$$V = \sum_{i=1}^{I}p(x_i)\,x_i^2 - \bar{x}^2$$

For a continuous variable,

$$V = \int_{-\infty}^{\infty}f(x)\,x^2dx - \bar{x}^2$$

Variance is also called **second moment of dispersion.**

Standard Deviation

The positive square root of variance is called standard deviation (S.D.) and is denoted by σ or S. Because of the square root occurring in the definition, the standard deviation has the same dimensions as the observations. It can therefore be compared directly with the mean of the distribution.

Coefficient of variation: It is the ratio of standard deviation to the mean. Since, it is a relative term without any dimension, it is a very useful measure of dispersion.

Example 5.6. Determine the standard deviation and coefficient of variation for the data of Example 5.5.

Solution: Mean of the given data has been determined as, $\bar{x} = 58.1$. The standard deviation is given by,

$$S = \left[\frac{1}{N-1}\sum_{i=1}^{I} n_i(\bar{x}-x_i)^2\right]^{\frac{1}{2}}$$

Since the data is grouped into 10 parts, $I = 10$; n_i is the number of observation or frequency in each group. $N = \sum_{i=1}^{10} n_i$. The computations are given in Table 5.6.

Table 5.6

x_i	n_i	$(\bar{x}-x_i)^2$	$n_i(\bar{x}-x_i)^2$
14	2	1944.81	3889.62
23	3	1232.01	3696.03
32	5	681.21	3406.05
41	6	292.41	1754.46
50	8	65.61	524.88
59	11	0.81	8.91
68	9	98.01	882.09
77	8	357.21	2857.68
86	6	778.41	4670.46
95	2	1361.61	2723.22
	60		24413.4

$$S = \sqrt{\frac{24413.4}{(60-1)}} = 20.34$$

$$\text{Coefficient of variation} = \frac{S}{\bar{x}} = \frac{20.34}{58.1} = 0.35$$

5.6 Generation of Random Variates

In Gambling game of Example 2.2, we have discussed the generation of random observations from discrete distribution and in Example 2.3 of Numerical Integration, from a continuous distribution. There are a large number of probability density functions, which describe the various random phenomena. Here we will discuss the generation of random variates from some continuous and discrete distributions, which are commonly encountered in simulation.

5.7 Bernoulli Trial

Bernoulli Trial is a trial which results in two outcomes, usually called 'a success' and 'a failure'. Bernoulli Process is an experiment performed repeatedly, which has only two outcomes, say success (S) and failure (F). Here the probability that an event will occur (success) remains constant. In rolling a die, the probability of face i (1, 2, ..., 6) coming up is always $\frac{1}{6}$. The probability of a head while tossing a coin is 0.5 in each flip. Firing a target (hit or miss), contesting an election (win or lose) etc. are examples of Bernoulli trials. It is very easy to simulate a Bernoulli case. For example, if the probability that a machine will fail in the next one hour is 0.3, we just draw a random number, if it is less than or equal to 0.3, machine will fail otherwise it will not fail in the next hour.

5.8 Binomial Distribution

If 'n' Bernoulli trials are made at a time, then the distribution of the number of successes, is the Binomial distribution. For example, if a sample of N balls is drawn from an infinitely large population of balls, having a proportion p of say red balls, then the distribution of red balls in each sample is given by Binomial distribution.

The Binomial probability mass function of a variable x is given by,

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0, 1, 2, ..., n.$$

where p is the probability of success in an independent trial and thus $1 - p$ is the probability of failure. x, a positive integer represents the number of successes in n independent trials. To generate a random variate x, n random numbers are generated. A random number less than or equal to p i.e., $r \leq p$ gives a success and the total numbers of such successes gives the value of x.

The properties of Binomial distribution are

$$\text{Mean } \bar{x} = n.p$$

$$\text{Variance } \sigma^2 = n.p.(1-p)$$

A computer program for generating the random variates (x) from a Binomial distribution with parameters n (number of trials) and p (probability of success in a trial) is given below.

```
#include<stdio.h>
#include<stdlib.h>
#include<math.h>
main()
  { /* To generate variates from a BINOMIAL distribution having
       parameters - n (number of trials) and p (probability of success
       in a trial)*/
       int n,i,x,k,m,nn,nt;
       float y,p;
       printf("\n The parameter n (number of trials) nt=");
       scanf("%d",&nt);
      printf("\n Enter the value of p(<1.0) the probability of succss=");
      scanf("%f",&p);
      printf("\n Number of variates to be generated (nn)=");
      scanf("%d",&nn);
      printf("\n Number of trials=%d\n Probability of success=
%4.2f",nt,p);
      printf("\n Number of variates= %d \n",nn);
      printf("\n Values of variate x:");
      for(m=1;m<=nn;++m) {
```

```
        x=0; n=nt;
        for(i=1;i<=n;++i) {
        y=rand()/32768.0;
        if(y < p) x=x+1;
            }
            printf(" %d",x);
            }

    }
```

The output of this program is given below:

Number of trials : 10

Probability of success = 0.6

Number of values : 20

Values of x : 8 6 3 7 7 7 8 9 6 6 8 4 3 7 3 6 5 5 4 5

5.9 Negative Binomial Distribution

In Binomial distribution, the number of trials N is fixed and the number of successes (or failures) determines the random variate. In negative Binomial distribution, the random variable is given by the number of independent trials to be carried out until a given number of successes occur. If X denotes the number of trials to obtain the fixed number K of successes then,

Probability of X trials until K successes occur

= Probability of $(K-1)$ successes in $(X-1)$ trials × probability of a success in Xth trial.

$$p(X) = \binom{X-1}{K-1} P^{K-1}((1-P)^{X-K}P)$$

$$= \binom{X-1}{K-1} P^{K}(1-P)^{X-K}$$

where P is the probability of success in a trial.

When K is an integer, this distribution is also called a **Pascal distribution.** When K is equal to one, it is called a **geometric distribution.**

A computer program for generating random variable 'x' from a Negative Binomial distribution with parameters k (numbers of successes) and p (Probability of success in a trial) is given below.

```
#include<stdio.h>
#include<stdlib.h>
#include<math.h>
main()
{     /* To generate  variates  from a NEGATIVE BINOMIAL (Pascal)
      distribution having parameters k (number of successes)
      and p (probability of success in a trial    */

      int x,j,nn,m,s,k;
      float p,r;

      printf("\n Enter the values of p(<1.0)=");
      scanf("%f",&p);
```

```
printf("\n Enter the values of k (>=1)=");
scanf("%d",&k);
printf("\n Value of Papameter k= %d",k);
printf("\n Probability of success in a trial=%4.2f",p);
printf("\n Number of variates to be generated=");
scanf("%d",&nn);
printf("\n Values of variates x:");

for(j=1;j<=nn;++j) {
x=0; s=0 ;

while(s<k) {
r=rand()/32768.0;
if(r<=p)    {s=s+1;}
x=x+1;
/* printf("\n %5.1f %d %d",x,s,n);*/
}
printf("%4d,",x);
}

}
```

An output obtained from the above program is as under :

Value of parameter $k = 3$

Probability of success in a trial = 0.45

Number of variates to be generated = 15

Values of variates x : 3, 3, 6, 5, 18, 7, 4, 8, 6, 4, 3, 5, 3, 4, 3

5.10 Geometric Distribution

It is a special case of Negative Binomial distribution with $K = 1$. Independent Bernoulli trials are performed until a success occurs. If X is the number of trials carried out to get a success, then X is the geometric random variable.

$$p(X) = P.(1-P)^{X-1}$$

Mean $= \dfrac{1}{P}$ and variance $= \dfrac{1-P}{P^2}$, where P is the probability of success in a trial.

5.11 Hypergeometric Distribution

In case of Binomial distribution, the probability of success is same for all trials. In other words, the population with proportion P of the desired events is taken to be of infinite size. However, if this population is finite and samples are taken without replacement, the probability of proportion of desired events in the population will vary from sample to sample. The distribution, which describes the distribution of success X, in such a case, is called the hypergeometric distribution.

If N is sample size and M is the finite size of population to start with, and P is the probability of success in a trial in the beginning, then the hypergeometric probability function is given by,

$$p(X) = M \cdot \frac{P!}{X!}(M \cdot P - X)! \frac{(M \cdot Q)!\,(MQ - N + X)}{(N-X)!\,\dfrac{M!}{N!}(M-N)!}$$

where $Q = 1 - P$

The mean is given by $\bar{x}$ = N.P and variance $\sigma^2 = \frac{M-N}{M-1} N (P \cdot Q)$

5.12 Poisson Distribution

Poisson distribution is the discrete version of the Exponential distribution. Let x be a random variable which takes non-negative integer values only, then the probability mass function,

$$p(x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

is called the Poisson distribution with parameter λ, where $\lambda > 0$

The properties of Poisson distribution are

$$\text{Mean} = \lambda$$
$$\text{Variance} = \lambda$$

One of the first applications of this distribution was in representing the probable number of Prussian cavalry troopers killed each year by being kicked in the head by horses. This distribution finds applications in a wide variety of situations where some kind of event occurs repeatedly but haphazardly. Number of customers arriving at a service station, number of fatal accidents per year per kilometer of highway, number of typographical errors on a page, number of flaws per square yard of cloth, number of α-particles emitted by a radioactive substance etc. are some of the situations.

To generate an observation X, a random number is generated. The cumulative probabilities of X for values 0, 1, 2, ... etc. are computed and each time compared with the random number. As soon as the cumulative probability becomes greater than or equal to the random number, the value of X is returned.

A computer program subroutine for generating the Poisson observations is given below :

```
#include<stdio.h>
#include<stdlib.h>
#include<math.h>
main()
  {  /*  To generate  variates  from a POISSON distribution
     given the 'mean' that is parameter lemda(>0).
     fact - factorial;
     prob - probability and cumprob - cumulative probability  */

     int  k,j,x,nn;
     float lemda,cumprob,fact,prob,y,mean ;
     /* lemda=5.0;*/
     printf("\n Value of mean (lemda)=");
     scanf("%f",&lemda);
     printf("\n Number of variates to be generated=");
     scanf("%d",&nn);

     mean=lemda;

     for(j=1;j<=nn;++j) {
     fact=1.; x=0;
     cumprob=0.;
```

```
    y=rand()/32768.0;
    while(y > cumprob) {
    prob= pow( 2.718282,-mean) * pow(mean,x)/fact  ;
    cumprob=cumprob+prob;
    x=x+1;
    fact=fact*x;
    /*printf("\n %4.3f %4.2f %4.2f %4.2f",y, prob,cumprob,fact);*/
    }
    printf(" %4d",x);
    }

}
```

An output of the above program is given below:

Value of mean (lambda) = 5

Number of variates to be generated = 15

Values of variates x : 2 1 5 2 5 4 6 4 7 10 5 6 3 7 6

5.13 Empirical Distribution

It is a discrete distribution constructed on the basis of experimental observations. The outcomes of an activity are observed over a period of time, and the probability of each outcome is established from the observations. For example, the demand of an item as observed over a long period is 1, 2, 3 or 4 pieces per hour. Out of say 400 observations, 40 times the demand is for one piece, 160 times for 2 pieces, 180 times for three pieces and 20 times for four pieces. In other words the probabilities of demand being 1, 2, 3 or 4 pieces are $\frac{40}{400} = .1$, $\frac{160}{400} = .4$, $\frac{180}{400} = .45$ and $\frac{20}{400} = .05$ respectively. This discrete probability distributions is an empirical distribution.

The general approach of working with such a distribution is called the integral inverse approach. Integration, the process of adding the values is statistically called cumulative distribution. A variate is generated using a random number, that is taking inverse of the law.

If $F(x)$ represents C.D.F. of variable x, and a random number r is such that $F(i-1) \le r \le F_i$ then the discrete variable x assumes the value x_i

i	1	2	3	....	n
x_i	x_1	x_2	x_3	....	x_n
$F(x)$	F_1	F_2	F_3	....	F_n

A computer program in C language for generating a random variable x from a given Empirical distribution is given below.

```
#include<stdio.h>
#include<stdlib.h>
main()
   {
   /* To generate random variables from an EMPIRICAL distribution.*/

     int i,j,k,n,nn;
     float  rnd,y;
     n=4;
```

```
/*  Values of x[i] and cumulative probabilitiesf[i] with i= 0,4 */
    float x[] ={1.5,2.5,4.0,5.5,6.0};
    float p[]= {0.0,0.1,0.25,0.65,0.90,1.0};
    printf("\n x[i] =       ");
    for(i=0;i<=n;++i) {printf("%6.2f",x[i]); }
    printf("\n p[i]=");
    for(i=0;i<=n+1;++i) {printf("%6.2f",p[i]); }
    printf("\n Numbe of variates to be generateds(nn)=");
    scanf("%d",&nn);
    printf("\n Values of variates x:\n");
    for(k=1;k<=nn;++k) {
    rnd=rand()/32768.0;
    for(i=0; i<=n; ++i) {
    j=i+1;
    if((rnd >p[i] ) && ( rnd <= p[j]))  y= x[i];
/*  printf("\n%d %d %5.4f %5.4f x=%5.4f %d",i,j,pr[i],pr[j],x,te);*/
    }

          printf("%5.1f",y);
          }

    }
```

An output obtained from this program is given below.

x[i]	=		1.50	2.50	4.00	5.50	4.00
p[i]	=	0.00	0.10	0.25	0.65	0.90	1.00

Number of variates to be generated = 12

Values of the variate x : 1.5 1.5 4.0 1.5 4.0 2.5 4.0 2.5 5.5 6.0 4.0 4.0

5.14 Continuous Distribution

A random variable can be generated from any of the continuous distributions by making use of the cumulative density function of the variable. If $F(x)$ is the C.D.F. of variable x, then the new random variable,

$$Y = F(x)$$

is uniformly distributed over 0,1 interval. If a random number r is generated, then

$$r = F(x)$$

or

$$x = F^{-1}(r)$$

The generation of random observations from some important continuous distributions is given below.

5.15 Normal Distribution

There are two probability laws, the normal probability and exponential probability, that describe most of the behavior that can be observed in real life systems. There are many other probability laws derived from these two, but they are used only when finer precision is needed in simulation.

The normal distribution is also called the Gaussian distribution after the mathematician who first described it. Because of its bell type shape, normal distribution is also called the bell curve. Normal distribution is used most frequently to describe the distributions as the marks obtained by a class, dimensions of parts made on a machine, number of light bulbs which fuse per time period, heights of male or female adults etc.

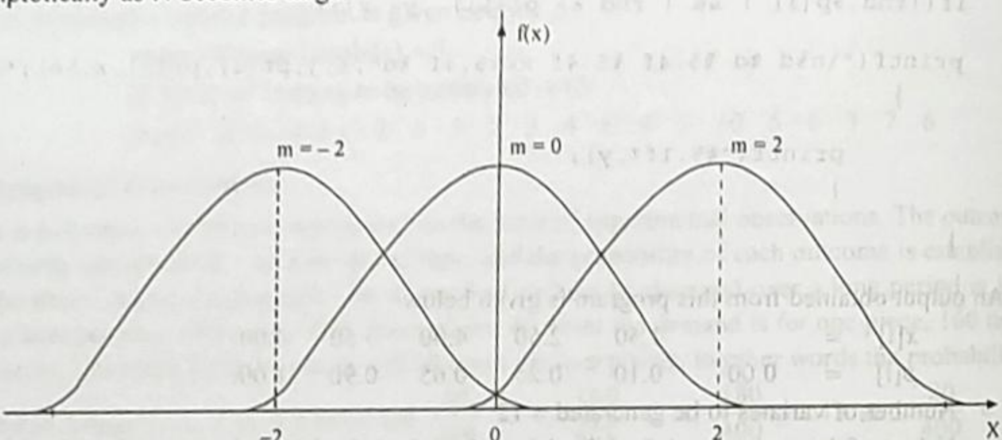
The density function of the normal curve is given by

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right], -\infty < x < \infty$$

where μ is mean of the distribution and σ is standard deviation.

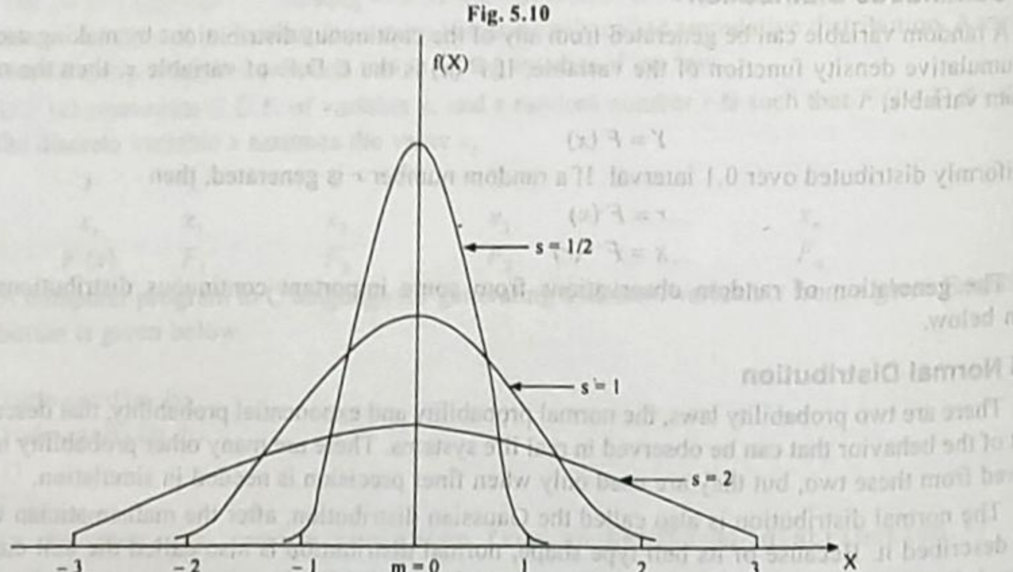
The parameter μ can take any real value, while standard deviation s is always positive. The shape of a normal curve depends upon the values of μ and σ. The effect of variations in the values of μ and σ is illustrated in Fig. 5.10 and Fig. 5.11.

The normal random variates can be generated by a number of methods. The commonly used methods are developed from the central limit theorem, which states that the probability distribution of the sum of n independent and identically distributed random numbers r_i with mean μ_i and variance σ_i^2 ($i = 1, n$), approaches the normal distribution with mean $\sum_{i=1}^{n}\mu_i$ and variance $\sum_{i=1}^{n}\sigma_i^2$ asymptotically as N becomes large.



Normal Distribution with s fixed (s = 1)

Fig. 5.10



Normal Distribution with m fixed (m = 0)

Fig. 5.11

Let $r_1, r_2, \ldots\ldots, r_n$ be the random numbers ($0 < r_i < 1$) and

$$V = \sum_{i=1}^{n} r_i$$

Then $E(V) = \frac{n}{2}$

$$\text{Var}(V) = \sum_{i=1}^{n} \text{var}(r_i) = \frac{n}{12}$$

Thus V is asymptotically normal with mean $\frac{n}{2}$ and variance $\frac{n}{12}$.

Let
$$Z = \frac{V - \frac{n}{2}}{\sqrt{\frac{n}{12}}}$$

which is normal with zero mean and unit variance.

Now any random variate y, corresponding to the above n random numbers can be obtained from the expression.

$$\frac{y-\mu}{\sigma} = \frac{V - \frac{n}{2}}{\sqrt{\frac{n}{12}}}$$

or
$$y = \mu + \frac{\sigma}{\sqrt{\frac{n}{12}}}\left(V - \frac{n}{2}\right) = \mu + \frac{\sigma}{\sqrt{\frac{n}{12}}}\left(\sum_{i=1}^{n} r_i - \frac{n}{2}\right)$$

Since, according to Central Limit Theorem, the normality is approached quite rapidly even for small values of n, the value $n = 12$ is commonly used in practice.

$\therefore$
$$y = \mu + \sigma\left(\sum_{i=1}^{12} r_i - 6\right),\ i = 1, 2, \ldots\ldots, 12$$

A computer program to generate random variable x from a normal distribution with mean 'mue' and standard deviation 'sigma' is given below:

```
#include<stdio.h>
#include<stdlib.h>
#include<math.h>
main()
  {
    /* To generate random variable from a Normal distribution
     having mean mue and standard deviation sigma.*/
     int i,j,m,nn;
     float t,sum,x,mue,sigma;
     printf("\n Enter the values of mue ");
     scanf("%f",&mue);
     printf("\n Enter the values of sigma");
     scanf("%f",&sigma);
     printf("\n Enter the number of variables to be generated nn=");
```

```
        scanf("%d",&nn);
        printf("\n    mue=%4.2f    sigma=%4.2f    nn=%4d",mue,sigma,nn);
        printf("\n Values of variable x:");
        printf("\n");
        for(m=1;m<=nn;++m) {
        sum=0.;
        for(i=1;i<=12;++i) {
        x=rand()/32768.0 ;
        sum=sum+x;
        }
        t=mue+sigma*(sum-6.);
        printf(" %6.2f",t);

        }

}
```

An output of this program is given below.

Mue = 10.0 sigma = 1.5 $nn = 10$

Values of variate x: 7.09 10.42 12.45 8.89 9.33 7.92 7.95 9.11 9.93 10.77

5.16 Exponential Distribution

The exponential distribution is used to represent an activity where most of the events take place in a relatively short time, while there are a few which take very long times. The p.d.f. for this distribution is shown in Fig. 5.12.

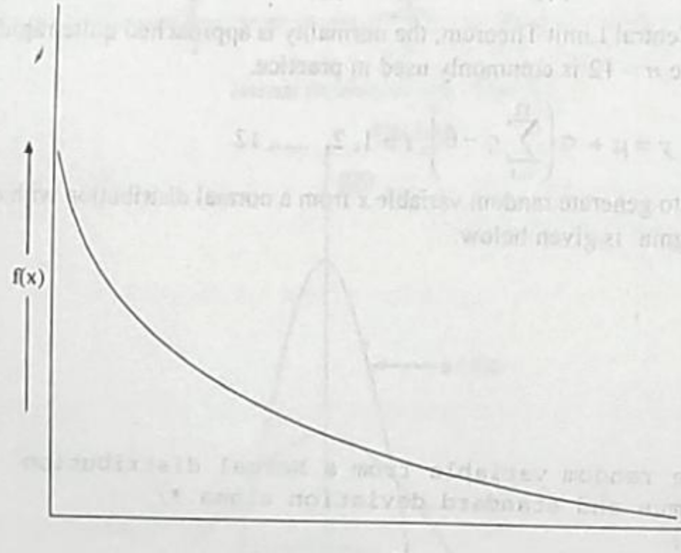


Fig. 5.12

The service times in queuing systems, inter arrival times of vehicles on a highway, the life of some electronic parts, and the orders for a product received per day etc. are some of the examples where exponential distribution can be used. Exponential distribution is analogous to Geometric distribution. While geometric distribution represents a variable as the number of customers arriving up to a specific time, the exponential distribution gives the time of next arrival of the customer.

The exponential p.d.f. is given by

$$f(t) = \begin{cases} \mu e^{-\mu t}, t > 0 \\ 0, \ otherwise \end{cases}$$

The random variate from this distribution is generated by drawing a random number, r.

$$r = f(x) = \int_0^x \mu e^{-\mu x} dx = 1 - e^{-\mu x}$$

$$x = -\frac{1}{\mu} \ln(1-r) = -\frac{1}{\mu} \ln r$$

The above expression is justified since $1 - r$ is also a random number and r and $1 - r$ are equally likely.

5.17 Erlang Distribution

There is a class of distribution functions named after A.K. Erlang, who found these distributions to be representative of certain types of telephone traffic. If there are k independent random variables $v_i (i = 1, 2, \ldots\ldots, k)$, having the same exponential distribution,

$$f(v_i) = \mu k e^{\mu k v_i}$$

with $v_i > 0$, $\mu > 0$ and k a positive integer then

$$V = \sum_{i=1}^{k} v_i \text{ has the Erlang distribution}$$

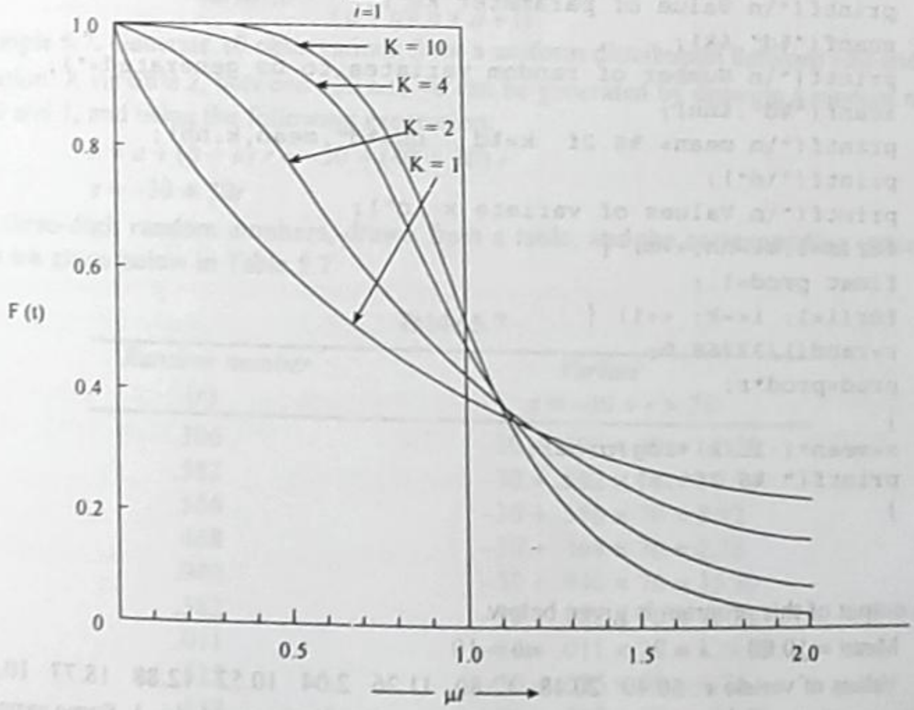


Fig. 5.13

Thus if there are k independent variates having the same exponential distribution, then their combined distribution is called Erlang distribution. Fig. 5.13 illustrates the Erlang distribution for several values of k. The parameter k is called the Erlang shape factor.

A random variate from Erlang distribution can be obtained by obtaining k random variates from the exponential distribution and adding them.

Thus, if $$v_i = -\frac{1}{\mu k} \ln r_i, \; i = 1, 2 \ldots\ldots k.$$

Then $$v = \sum_{i=1}^{k} v_i = \sum_{i=1}^{k} -\frac{1}{\mu k} \ln r_i = -\frac{1}{\mu k} \ln \prod_{i=1}^{k} r_i$$

A computer program in C language for generating Erlang variates is given below:

```
#include<stdio.h>
#include<stdlib.h>
#include<math.h>
main()
    {
      /*To generate EARLANG distributed variates given the values of
       mean and earlang parameter k.
       If the value of k is taken as 1, than this program generates
       EXPONENTIALLY distributed variates*/

       int  i,j,k,m,nn;
       float x,r,mean;
       printf("\n Value of mean =");
       scanf("%f",&mean);
       printf("\n Value of parameter k=");
       scanf("%d",&k);
       printf("\n Number of random variates to be generated=");
       scanf("%d",&nn);
       printf("\n mean= %6.2f  k=%d   nn=%d",mean,k,nn);
       printf("\n");
       printf("\n Values of variate x:\n");
       for(m=1;m<=nn;++m) {
       float prod=1.;
       for(i=1; i<=k; ++i) {
       r=rand()/32768.0;
       prod=prod*r;
       }
       x=mean*(-1./k)*log(prod);
       printf(" %6.2f",x);
       }
    }
```

An output of this program is given below.

Mean = 10.00 $k = 2$ $nn = 10$

Values of variate x : 50.40 20.48 12.80 11.26 2.04 10.52 12.88 18.77 10.02 3.23

Exponential distribution is a special case of Erlang distribution, with $k = 1$. Some exponentially distributed variates generated by the above program are given below.

Mean = 10.00 $k = 1$ $nn = 10$

45.51 55.30 10.93 34.03 10.34 19.27 6.22 16.31 3.56 0.51

5.18 Uniform Distribution

The uniform probability law states that all the values in a given interval are equally likely to occur. Suppose, we say that service times are uniformly distributed between 30 and 80 seconds. It means that a service will not take less than 30 seconds and more than 80 seconds. On the average service time will be 55 seconds.

Let a variable x be uniformly distributed in the interval (a, b), that is all the values of x between a and b, $(b > a)$ are equally likely, then

$$x = a + (b - a)\, r$$

where r is a random number between 0 and 1. The uniform distribution is also called Rectangular distribution or Homogeneous distribution.

5.19 Beta Distribution

The beta distribution is an extension of Erlang distribution and exists only between the limits 0 and 1. A random variable x is said to have beta distribution with parameters p and q, if its density function is given by:

$$f(x) = \begin{cases} \dfrac{x^{p-1}(1-x)^{q-1}}{\beta(p,q)}, & \text{if } 0 < x < 1 \\ 0, & x \le 0 \text{ and } x \ge 1 \end{cases}$$

$$\text{with } p > 0,\ q > 0.$$

Properties of the beta distribution are:

$$\text{Mean} = \frac{p}{p+q}$$

$$\text{Variance} = \frac{p(p+1)}{(p+q)(p+q+1)}$$

Example 5.7. Generate 10 observations from a uniform distribution between –30 and 40.

Solution: A variable x, between –30 and 40 can be generated by drawing a random number r between 0 and 1, and using the following expression:

$$x = a + (b - a)\, r = -30 + (40 + 30)\, r$$

or

$$x = -30 + 70r$$

Ten three-digit random numbers, drawn from a table, and the corresponding values of the variable x are given below in Table 5.7.

Table 5.7

Random number (r)	*Variate* $x = -30 + r \times 70$
.396	–30 + .396 × 70 = –2.28
.582	–30 + .582 × 70 = 10.74
.556	–30 + .556 × 70 = 8.92
.468	–30 + .468 × 70 = 2.76
.940	–30 + .940 × 70 = 35.80
.382	–30 + .382 × 70 = –3.26
.011	–30 + .011 × 70 = –29.23
.525	–30 + .525 × 70 = 6.75
.117	–30 + .117 × 70 = –21.81
.579	–30 + .579 × 70 = 10.53

Example 5.8. A random variable X has the following empirical distribution.

X	:	1	2	4	6	8	10
$f(x)$	:	.10	.20	.25	.20	.15	.10

Plot the cumulative distribution and find the values of X corresponding to the following two-digit random numbers, 05, 45, 62, 93.

Solution: The cumulative distribution for variable X is computed below in Table 5.8, and is shown in Fig. 5.14.

Table 5.8

X	$f(x)$	$F(x)$
1	.10	0.10
2	.20	0.30
4	.25	0.55
6	.20	0.75
8	.15	0.90
10	.10	1.00

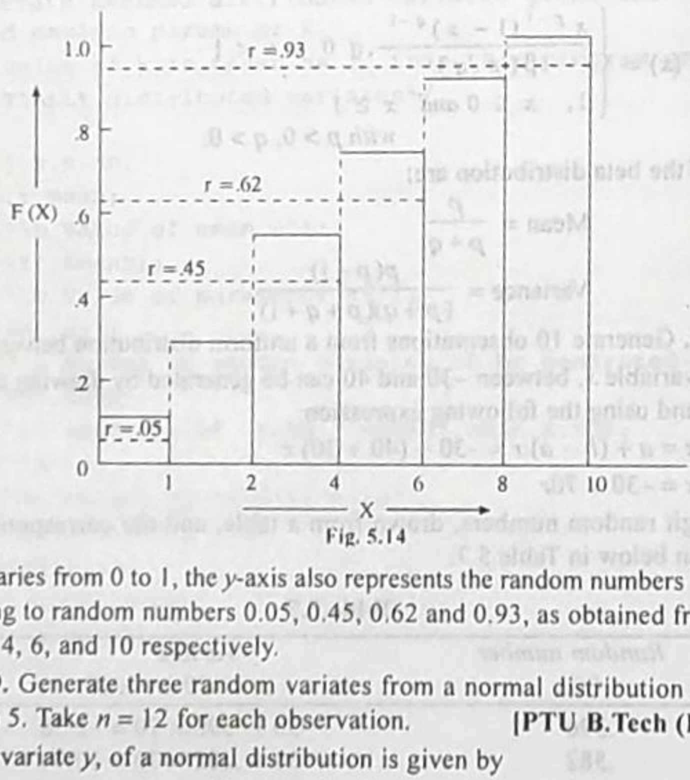


Fig. 5.14

Since CDF varies from 0 to 1, the y-axis also represents the random numbers from 0 to 1. Values of X, corresponding to random numbers 0.05, 0.45, 0.62 and 0.93, as obtained from the cumulative distribution are 1, 4, 6, and 10 respectively.

Example 5.9. Generate three random variates from a normal distribution with mean 20 and standard deviation 5. Take $n = 12$ for each observation. **[PTU B.Tech (Prod.) May 2006]**

Solution : A variate y, of a normal distribution is given by

$$y = \mu + \sigma\left(\sum_{i=1}^{n} r_i - \frac{n}{2}\right)$$

$$y = \mu + \sigma\left(\sum_{1}^{12} r_i - 6\right) \qquad \text{when } n = 12$$

where μ is mean and σ is standard deviation.

Since $n = 12$, we need 12 random numbers for generating one value of the variable.

(*i*) Twelve random numbers taken from a random number table are,

.483, .517, .063, .229, .807, .562, .066, .924, .511, .134, .657, .602

$$\sum_{i=1}^{12} r_i = 5.555$$

$$y_1 = 20 + 5\,(5.555 - 6) = 17.775$$

(*ii*) Twelve random numbers taken for the next value of y are,

.357, .944, .733, .345, .063, .546, .387, .935, .217, .816, .768, .183

$$\sum_{i=1}^{12} r_i = 6.294$$

$$y_2 = 20 + 5\,(6.294 - 6) = 21.47$$

(*iii*) From the same random number sequence, the next 12 random numbers are,

.488, .597, .183, .922, .731, .619, .232, .568, .421, .045, .897, .165

$$\sum_{i=1}^{12} r_i = 5.868$$

$$y_3 = 20 + 5\,(5.868 - 6) = 19.34$$

Thus three random variates from the given normal distribution are 17.775, 21.47 and 19.34.

Example 5.10: Generate three random variates from an exponential distribution having mean value 8.

Solution: A variate y, of an exponential distribution is given as,

$$y = -\frac{1}{\mu}\ \ln\,(1 - r) \quad \text{or} \quad -\frac{1}{\mu}\ \ln\,(r)$$

where $\frac{1}{\mu}$ is the mean of the distribution

$$y = -\,8\ \ln\,(1 - r)$$

Three random numbers taken from a table are

.513, .297 and .728

Then,

$$y_1 = -\,8\ \ln\,(1 - .513) = 5.756$$
$$y_2 = -\,8\ \ln\,(1 - .297) = 2.819$$
$$y_3 = -\,8\ \ln\,(1 - .728) = 10.416$$

The random variates could also be obtained as,

$$y_1 = -\,8\ \ln\,(.513) = 5.340$$
$$y_2 = -\,8\ \ln\,(.297) = 9.712$$
$$y_3 = -\,8\ \ln\,(.728) = 2.540$$

Example 5.11. Generate three random observations from an Erlang distribution having mean 8 and shape factors (*a*) 2, (*b*) 4.

Solution: A variate y of an Erlang distribution is given as

$$y = -\frac{1}{\mu k}\ \ln\ \prod_{i=1}^{k}(1 - r_i)$$

or

$$-\frac{1}{\mu k}\ \ln\ \prod_{i=1}^{k} r_i$$

where $\frac{1}{\mu}$ is the mean and k is shape factor.

(*a*) Since $k = 2$, for each observation we need two random numbers. Six random numbers obtained from a table are,

$$.319, .461, .344, .548, .633, .822$$

$$y_1 = -\frac{8}{2} \ln \{.319 \times .461\} = -\frac{8}{2} \ln 0.14706 = 7.668$$

$$y_2 = -\frac{8}{2} \ln (.344 \times .548) = 6.674$$

$$y_3 = -\frac{8}{2} \ln (.633 \times .822) = 2.613$$

(*b*) Since $k = 4$, for each observation 4 random numbers are required. Twelve random numbers taken from the table are:

$$.320, .147, .638, .182, .527, .823, .627, .471, .220, .618, .292, .504$$

$$Y_1 = -\frac{8}{4} \ln (.320 \times .147 \times .638 \times .182) = 10.420$$

$$Y_2 = -\frac{8}{4} \ln (.527 \times .823 \times .627 \times .471) = 4.110$$

$$Y_3 = -\frac{8}{4} \ln (.220 \times .618 \times .292 \times .504) = 7.823$$

Example 5.12. Generate five random observations from the following distributions.

(*a*) Uniform distribution from 15 to 60

(*b*) Triangular distribution as

$$f(x) = \begin{cases} \dfrac{x^2}{2}, & \text{if } 0 \le x \le 1 \\ 0, & \text{otherwise} \end{cases}$$

(*c*) The distribution having the p.d.f. as

(*i*) $f(x) = \begin{cases} \dfrac{1}{50}(x-5), & \text{if } 10 \le x \le 20 \\ 0, & \text{otherwise} \end{cases}$ **[PTU, B.Tech May 2006]**

(*ii*) $f(x) = \dfrac{5}{7} x^2, \quad 1 \le x \le 2$

Solution: (*a*) If y is the random variable, then

$$y = f(x) = 15 + r \times (60 - 15)$$

where r is random number between 0 and 1.

Table 5.9

Random number	*Random observation* (y)
.526	38.67
.659	44.655
.136	21.12
.712	47.04
.348	30.66

(b) $y = f(x) = \begin{cases} \dfrac{x^2}{2}, & \text{if } 0 \le x \le 1 \\ 0, & \text{otherwise} \end{cases}$

Taking random number r between 0 and 1.

$$y = \frac{r^2}{2}$$

Random number : .181, .963, .652, .892, .137

Random observation : .016, .464, .213, .398, .0094

(c) (i) Again taking random numbers between 0 and 1,

$$y = f(x) = \frac{1}{50}(x-5)$$

where $x = 10 + (20 - 10)\, r$

Table 5.10

Random number	*x*	*y*
.386	13.86	0.1772
.501	15.01	0.2002
.967	19.67	0.2934
.764	17.64	0.2528
.447	14.47	0.1894
.293	12.93	0.1586

(ii) $f(x) = \frac{5}{7}\, x^2,\ 1 \le x \le 2$

Since x lies between 1 and 2, if we take a random number between 0 and 1, then $x = 1 + r$

$$y = f(x) = \frac{5\,(1+r)^2}{7}$$

Random number (r) : .163 .716 .279 .954 .034

Random observation (y) : .966 2.1033, 1.1685, 2.7935, 2.6717

Example 5.13. There are 10 equally reliable semiautomatic machines in a screw-manufacturing unit. Their breakdowns per day follow the **Binomial distribution** with probability of failure 0.15. Generate the number of breakdowns for the next seven days. Determine the mean and variance of the generated observations. What are the theoretical values of the mean and variance?

Solution: For generating one observation, that is the number of machines breaking down per day, 10 random numbers between 0 and 1 are drawn. The random number ≤ 0.15 gives a failure. The random numbers used here in Table 5.11 are drawn form the random number table given in Appendix.

Table 5.11

Day	*Random numbers*	*Breakdowns*
1	48, 51, **06**, 22, 80, 56, **06**, 92, 51, **13**	3
2	65, 60, 51, 50, **13**, 94, 57, 26, 78, 33	1
3	60, 31, **15**, 64, 89, 74, 39, 63, 58, 83	1
4	44, 64, 59, **03**, 59, 30, 16, 57, 87, 21	1
5	36, 60, 82, 37, 72, 33, 90, 76, 29, 66	0
6	40, **11**, 44, 74, 27, 16, 41, 20, 68, 95	1
7	28, 75, 16, **02**, 88, **13**, 74, **07**, 63, 56	3

$$\text{Mean } \bar{x} = \frac{1}{7}\,(3+1+1+1+0+1+3) = \frac{10}{7} = 1.4286$$

$$\text{Variance} = \frac{1}{6}\,[2\,(3-1.4286)^2 + 4\,(1-1.4286)^2 + (0-1.4286)^2]$$

$$\frac{1}{6}\,[4.9386 + .7348 + 2.0409] = \frac{7.7143}{6} = 1.2857$$

Theoretical mean $= np = 10 \times .15 = 1.5$

Theoretical variance $= np\,(1-p) = 10 \times .15 \times .85 = 1.275$

If the observations are carried out for a larger number of days, the observed mean and variance will approach the true mean and variance.

Example 5.14. A variable x has Negative Binomial distribution with parameters $k = 2$ and $p = 0.35$. Generate seven observations of the variable. Determine the observed mean and standard deviation. Also compute the true mean and standard deviation of the observations.

Solution: In Negative Binomial distribution, a random number is drawn to represent an event. If $r \le p$, it is success. The random numbers are drawn until number of successes becomes equal to k. The number of random numbers drawn represents x. The string of two-digit random numbers used below in Table 5.12 is taken from a random number table.

Table 5.12

S.No.	*Random number*	*Success*	*S.No.*	*Random number*	*Success*
1	39	0	1	24	1
2	58	0	2	05	2
3	55	0		$x_4 = 2$	
4	46	0	1	36	0
5	85	0	2	45	0
6	84	0	3	04	1
7	38	0	4	69	1
8	01	1	5	66	1
9	93	1	6	58	1
10	52	1	7	69	1
11	46	1	8	35	2
12	11	2		$x_5 = 8$	0
	$x_1 = 12$		1	29	1
1	57	0	2	29	2
2	75	0		$x_6 = 2$	
3	86	0	1	53	0
4	44	0	2	12	1
5	33	1	3	89	1
6	28	2	4	87	1
	$x_2 = 6$		5	67	1
1	93	0	6	30	2
2	58	0		$x_7 = 6$	
3	18	1			
4	91	1			
5	02	2			
	$x_3 = 5$				

Thus, values of x are 12, 6, 5, 2, 8, 2, 6

$$\text{Mean } \bar{x} = \frac{12+6+5+2+8+2+6}{7} = \frac{41}{7} = 5.857$$

Standard deviation,

$$\text{S.D.} = \left[\frac{1}{6}\sum(x_i - \bar{x})^2\right]^{\frac{1}{2}}$$

$$= \left[\frac{1}{6}\{(12-5.857)^2 + (6-5.875)^2 + (5-5.857)^{2+} + (2-5.857)^2\right.$$

$$\left. + (8-5.857)^2 + (2-5.857)^2 + (6-5.857)^2\}\right]^{\frac{1}{2}}$$

$$= \left[\frac{1}{6}(37.736 + 0.1706 + 0.7344 + 6.5469 + 4.5924 + 14.8764 + 0.0205)\right]^{\frac{1}{2}}$$

$$= \left[\frac{64.6772}{6}\right]^{\frac{1}{2}} = 3.2832$$

$$\text{True mean} = \frac{k}{p} = \frac{2}{.35} = 5.7143$$

$$\text{True S.D.} = \frac{k(1-p)}{p^2} = \frac{2 \times .65}{(.35)^2} = 3.2576$$

Example 5.15. Generate the random variates, if the random variable of Example 5.14 follows geometric distribution with $p = 0.35$. Use the same string of random numbers. Determine the observed mean and S.D. as well as the true mean and S.D.

Solution:

Table 5.13

S.No.	*Random number*	*Success*	*S.No.*	*Random number*	*Success*
1	39	0	1	91	0
2	58	0	2	02	1
3	55	0		$x_6 = 2$	
4	46	0	1	24	1
5	85	0		$x_7 = 1$	
6	84	0	1	05	1
7	38	0		$x_8 = 1$	
8	01	1	1	36	0
	$x_1 = 8$		2	45	0
1	93	0	3	04	1
2	52	0		$x_9 = 3$	
3	46	0	1	69	0
4	11	1	2	66	0
	$x_2 = 4$		3	58	0
1	57	0	4	69	0
2	75	0	5	35	1
3	86	0		$x_{10} = 5$	
4	44	0	1	29	1