UNIT 14 CONTINUOUS PROBABILITY DISTRIBUTIONS

Structure

14.1 INTRODUCTION

In Unit 13, you were introduced to some probability distributions of a discrete random variable. Such a variable, as you know, can take discrete values like 0, 1, 2, ... or 0, .5, 1.0, 1.5, 2.0, ... Let us now turn our attention to a variable which takes all the values within a given range or an interval. Such a variable is called a continuous random variable. Measurements of height, temperature, amount of rainfall and waiting time are some of the examples of such a variable. Consider, for instance, the following situation:

A professor of statistics catches a bus every evening to take him home from his university. He has never bothered to find out the exact timings of the departure of the buses run since there is a frequent and punctual service – a bus leaves after every ten minutes. When his work for the day in the University is over, he walks down to the bus stand and catches the next bus. One hot evening as he is waiting for the bus he wonders if he would save much time and discomfort if he could find out the bus timings so that he could time his arrival at the bus stand in a better way. He decides to investigate his problem by setting up a mathematical model. If he continues his usual practice, he is equally likely to wait any time from 0 minute to 10 minutes. He may arrange his waiting time by a discrete random variable which can take the ten values say 0.5, 1.5, 2.5,..., 9.5. If he wants, he can record the timings to the nearest minute or half minute. So he may be justified to use a discrete random variable. But the professor thinks he will do better (or will have a better model) if the waiting-time variable could take any value within a ten minute interval i.e. the interval or range 0-10. This waiting time variable is not a discrete random variable since it can take any value from 0 to 10 say 8.7754. Therefore, the professor decides to choose a continuous random variable which can take continuously values. Such a random variable, as you know from Unit 11, is called a continuous random variable. It, therefore, follows that there are situations where the discrete random variable does not help and we have to bring in the concept of a continuous random variable. What is, then, the corresponding probability distribution of a continuous random variable? How to calculate the mean, variance and other measures of a continuous random variable? We shall try to find answers for such questions in this unit.

Objectives

After reading this unit, you should be able to

- describe a probability distribution of a continuous random variable;
- calculate the mean and variance of a continuous random variable;
- discuss some special types of continuous distributions.

14.2 CONTINUOUS RANDOM VARIABLE

You may recall from Unit 11 that continuous random variables represent **measured data** such as all possible heights, weights, temperatures, rainfalls, distances, life periods, whereas, discrete random variables represent **count data** such as number of children in a family, number of defective bulbs in an electrical firm, or number of accidental deaths in a year and so on. In short, a continuous random variable is one that takes all values within a given range on a continuous scale. The outcomes of an experiment in a continuous case are represented by the points on a line.

You may like to ask, "Are there only two random variables, discrete and continuous"? No. Not all, random variables are either discrete or continuous. In fact, there are random variables which are neither discrete nor continuous. Also, there are random variables which are partly discrete and partly continuous. We will, however, not discuss such variables as the same are beyond the scope of this course.

14.2.1 Probability Distributions

Having explained the meaning of a continuous random variable, let us now discuss the question of assigning or distributing probability values to the values of a continuous random variable. While doing so, you may recall that

i) a probability value p assigned to a value of a random variable always lies between 0 and 1 i.e.,

$$0 < p < 1$$

ii) the sum of all such probability values is equal to 1 (see Unit 12).

In Unit 13, you have learnt the methods of assigning or distributing probability values to the values of a random variable. Take, now, the case of a continuous random variable. For example, consider a random variable whose values are the heights of all persons over 20 years of age. Between any two values, say 168.5 cm and 169.5 cm, there are infinite number of heights, of which only one is 169 centimeters. The probability of selecting a person at random who is exactly 169 cm tall is extremely remote and hence will be zero. In other words, we assign a probability of zero to that event. Thus a continuous random variable has a probability value zero of assuming **exactly** any of its values. This, however, is not the case if we talk about the probability of selecting a person who is atleast 168 cm but not more than 170 cm tall with an interval rather than a point value of a random variable. This example tells us that we cannot define a probability distribution in the same way as for the discrete random variable where we assign non-zero probabilities for a random variable X taking particular values. But we can see that in the case of a continuous random variable X, the probability that X lies between two values is meaningful. That is, there has to be a different method of assigning (distributing) probabilities to the values of a continuous random variable for various intervals such as

$$P\,(a < X < b),\ P\,(X > c),\quad P\,(X < d), \text{ etc.}$$

where X is a continuous random variable taking values in an interval (a, b) or X takes all values greater than a number c or takes all values less than a number d. It does not matter whether we include in the interval the end points a and b or not, that is, to say, the interval may be open or closed or semi-open or semi-closed. This is because of the following reason: Since a continuous random variable has a probability of zero of assuming an exact value, therefore

$$P\,(X = a) = 0,\ P\,(X = b) = 0.$$

Hence

$$\begin{aligned} P\,(a \le X \le b) &= P\,(X = a) + P\,(a < X < b) + P\,(X = b) \\ &= 0 + P\,(a < X < b) + 1 \\ &= P\,(a < X < b). \end{aligned}$$

Thus when X is a continuous random variable, then we have

$$P(a \le X \le b) = P(a < X \le b) = P(a \le X < b) = P(a < X < b)$$

Similarly, $P(X \ge c) = P(X > c)$

$(X \le d) = P(X < d)$

Thus in the case of a continuous random variable X, it makes no difference if X takes values in a closed interval or in an open interval or in a semi-closed or semi-open interval. In view of this, we shall use the interval (a, b) for all types of intervals.

Recall from Unit 12 that the probability distribution of a continuous random variable is expressed as a function. This function is such that it can be represented geometrically as a continuous curve. This function is called a probability density function or simply a density function and is denoted by f(x).

The graph of a continuous distribution can be any continuous curve. The graph may take any one of the several forms. Some of these forms are shown in the Figures 1.

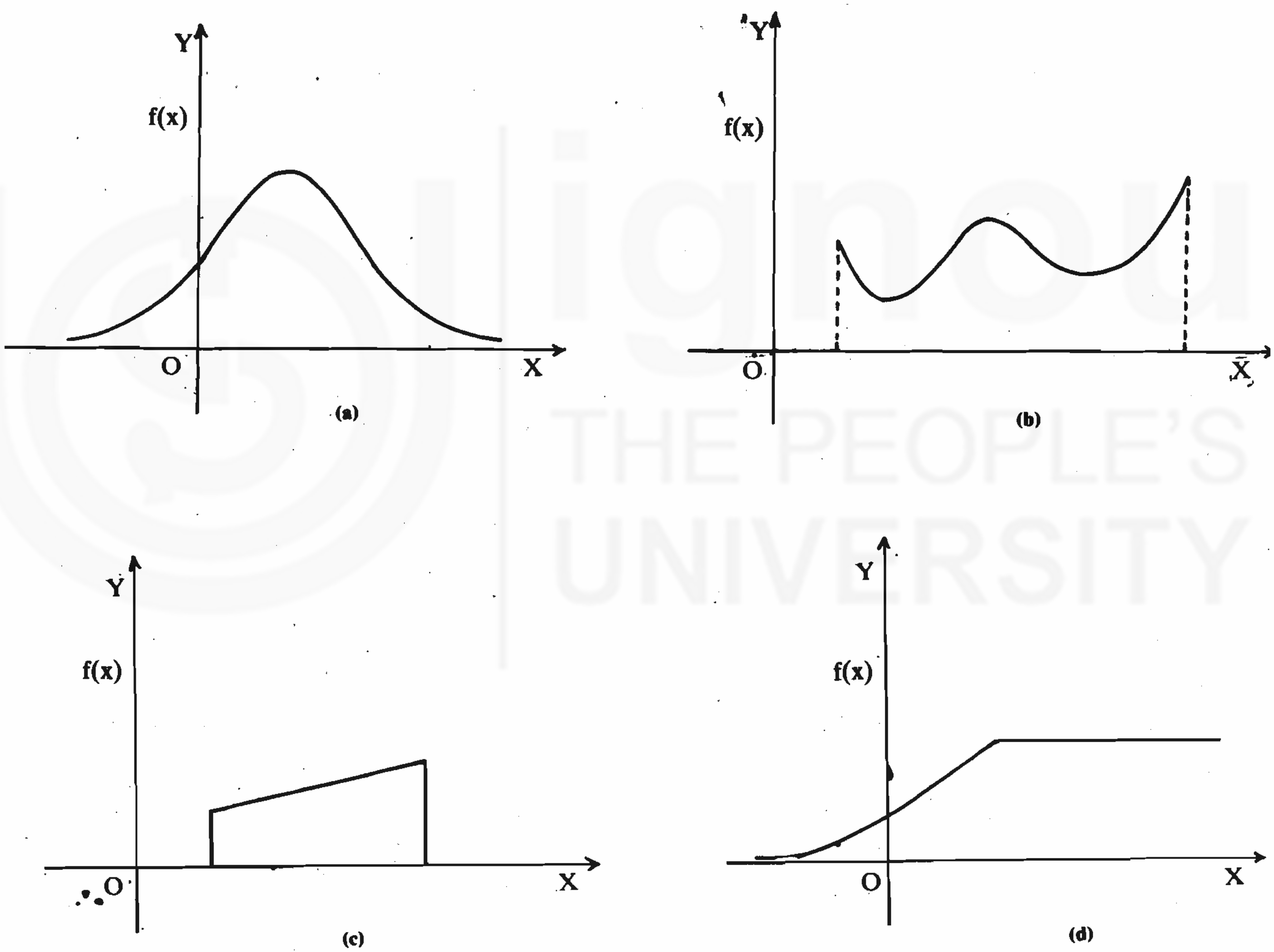

Fig. 1

The probabilities associated with values of a continuous random variable will be represented by the areas bounded by the lines/curves like the ones given in Figure 1. Also, these probabilities are positive numerical values. Therefore, the graph of the density function f(x) must lie above the X-axis and between an inerval (a, b). Moreover, (Unit 12) a probability distribution is such that the sum of the probabilities in the distribution is 1 (see Unit 13). Therefore, the total area under its curve bounded by the X-axis is equal to 1. For example, you can see Fig. 2.

From Figure 2, probability that X assumes a value between a and b is equal to the shaded area under the distribution function between the ordinates $x = a$ and $x = b$. Thus, now we are in a position to define a probability density function.

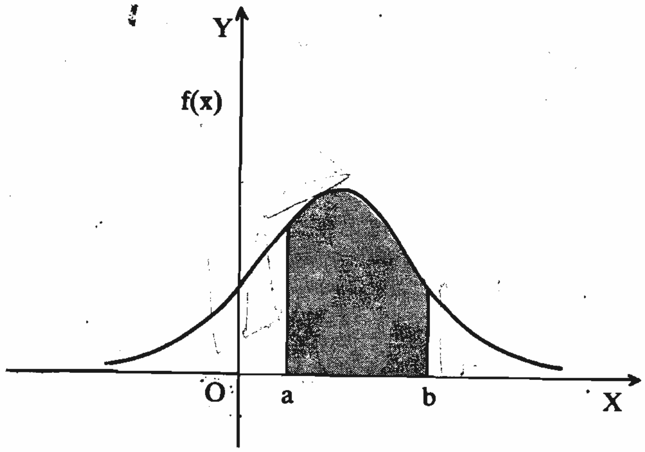

Fig. 2: P (a < X < b)

I. Probability Density Function (PDF)

A function with values f(x) is called a probability density function (PDF) for the continuous random variable X if

i) $f(x) \geq 0$.

ii) the total area under its curve and above the X-axis is equal to 1,

iii) the area under the curve between any two ordinates $x = a$ and $x = b$ and the X-axis gives the probability that X lies between a and b i.e. $P(a \leq x \leq b)$ = area under the curve between $x = a$ and $x = b$ and the X-axis.

Consider the following example of a density function:

Example 1: A continuous random variable X that can assume values between $x = 2$ and $x = 4$ has a density function f(x) given by.

$$f(x) = \frac{x+1}{8}.$$

Find $P(2 < X < 3)$.

Solution: As soon as you draw the graph of the given function, you will find that the region under the given restrictions namely $2 < X < 3$ is the region as shaded in the Figure 3.

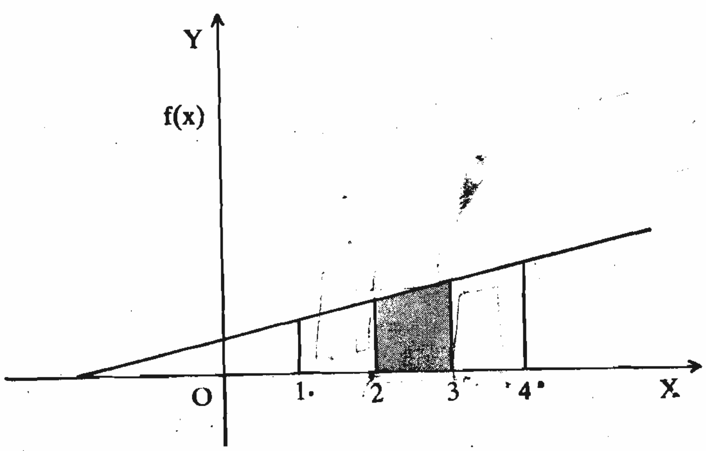

Fig. 3

This shaded region represents $P(2 < X < 3)$ and hence
$P(2 < X < 3)$ = Area of the shaded region.

The shaded region is a trapezium whose area is found by summing up the parallel heights, multiplying the sum by the length of the base and dividing by 2 i.e.

$$P(2 < X < 4) = \frac{(\text{sum of parallel sides}) \times \text{base}}{2}$$

$$= \frac{[f(2) + f(3)] \times 2}{2} = \frac{\left(\frac{3}{8} + \frac{4}{8}\right) \times 2}{2} = \frac{7}{8}$$

E1) Find P (2.4 < X < 3.5) in Example 1.

In Example 1, it was convenient for us to find the area of the shaded region because it happened to be a well-known geometrical figure like a trapezium whose formula for the area is known to us. However, it becomes difficult to calculate the area of a region when it is not a familiar geometrical figure. Then, what to do in such a situation? For this, we normally take the help of integral calculus which you have studied in Unit 8 and 9.

As you know that the probability of the variable X taking any value between the interval (a, b) is equal to the area under the curve of the density function f(x) between $x=a$ and $x=b$, $(a < b)$. This area, in the language of integral calculus (refer to Unit 8), is equal to the integral

$$\int f(x)\, dx.$$

Therefore, we can restate the definition of the probability density function f(x) of a continuous random variable X in the language of calculus in the following way:

Definition: The probability density function f(x) of a continuous random variable X is a function whose integral from $x=a$ to $x=b$ $(a < b)$ gives the probability that X takes a value in the interval [a, b]. In other words,

$$P(a \le X \le b) = \int_a^b f(x)\, dx.$$

Note that

$$P(X=a) = P(a \le X \le a) = \int_a^a f(x)\, dx = 0.$$

What conditions must be satisfied by the continuous function f(x) if it has to be a probability density function? Obviously, the answer is that these conditions should correspond to axioms of probability which you have learnt in Unit 12. These are two, namely,

i) $f(x) \ge 0$ for all values of x
corresponding to this axiom of probability (Unit 12), probability is always non-negative

ii) $\int_{-\infty}^{\infty} f(x)\, dx = \int R\, f(x)\, dx = 1$, R being the set of real numbers
corresponding to the probability axiom that the total sum of the probabilities is equal to 1.

Both the conditions, in fact, correspond to the law that the probability p is non-negative and is less than or equal to 1 i.e.

$$0 \le p \le 1.$$

Note that condition (ii) implies that the area under the curve over the whole region R of the variable X is 1 since the total sum of the probability is 1. The range of integration is written as $-\infty$ to $+\infty$ so that all cases are covered.

Also, note that if X takes only positive values in the interval (a, b) and is zero otherwise, then the condition (ii) reduces to

$$\int_a^b f(x)\,dx = 1.$$

We may verify Example 2 by this method

$$\int_2^4 f(x)\,dx = \int_2^4 \frac{x+1}{8}\,dx = \frac{1}{8}\left|\frac{x^2}{2} + x\right|_2^4 = 1.$$

Finally, note that

$$P(X \le b) = \int_{-\infty}^{b} f(x)\,dx = P(X < b)$$

$$P(X \ge a) = \int_{a}^{\infty} f(x)\,dx = P(X > a)$$

Let us take another example:

Example 2. Let the PDF of X be

$$f(x) = 12\,x^2\,(1 - x),\ 0 < x < 1$$
$$= 0, \text{ otherwise.}$$

Find $P\left(\frac{1}{3} < X < \frac{1}{2}\right)$.

Solution: The probability that X lies between 1/3 and 1/2 is given by

$$P\left(\frac{1}{3} < X < \frac{1}{2}\right) = \int_{1/3}^{1/2} 12\,x^2\,(1 - x)\,dx$$

$= .2013$ (verify the answer).

E2) Solve E1) by the method of integration.

II. The Cumulative Distribution Function (CDF)

Let X be a continuous random variable with PDF f(x). Then the probability that X has a value less than or equal to x is given by.

$$F(x) = P(X \le x) = \int_{-\infty}^{x} f(x)\,dx$$

The function F(x) is called the cumulative distribution function (CDF) of X. Since F(x) is the integral of f(x), therefore, we have

$$\frac{d\,F(x)}{dx} = f(x) \text{ or } F'(x) = f(x)$$

Obviously, then it follows from (ii) that

$$F(-\infty) = 0$$
$$F(\infty) = 1$$

and F(x) is a non-decreasing function of x, that is, if $y \ge x$,

then

$$F(y) \ge F(x)$$

Also

$$F'(x) = f(x) \ge 0$$

For any continuous random variable X, the graph of F(x) will look as given in Figure 4.

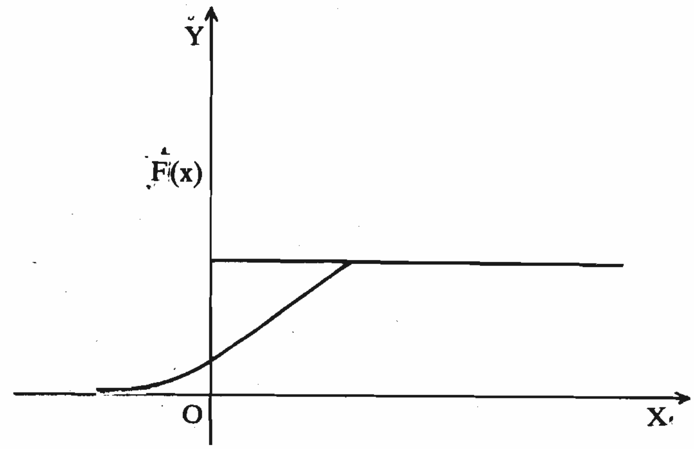

Fig. 4

Example 3: Find the CDF in Example 2

Solution: For the PDF given in Example 2, the CDF is

$$F(x) = P(-\infty < X < x)$$
$$= 0 \quad \text{for } x \le 0$$
$$= 4x^3 - 3x^4 \text{ for } 0 < x < 1$$
$$= 1 \quad \text{for } x \ge 1.$$

Your should, similarly, attempt the following exercise.

E3) A continuous random variable X takes values between 0 and 3 only. Its probability density function is given by

$$f(x) = 4x^2 (3 - x) \text{ for } 0 < x < 3,$$
$$= 0 \qquad \text{otherwise.}$$

Find the CDF of f(x).

14.2.2 Mean and Variance

You are familiar with the measures of central tendency which you have learnt in Unit 12. These are popularly known as the median, mean, mode. The measures of dispersion, as you know are, variance and standard deviation. We shall however, confine our discussion to mean and variance.

1. The Mean of a Continuous Random Variable

The mean μ_x or the expected value E (X) of a continuous random variable with PDF f(x) is defined as follows:

$$\bar{x} = \mu_x = E(x) = \int_{-\infty}^{\infty} x\, f(x)\, dx \text{ for } -\infty < x < +\infty$$

while $\int_{-\infty}^{\infty} f(x)\, dx = 1$

or

$$\bar{x} = \mu_x = E(x) = \int_a^b x\, f(x)\, dx, \text{ for } a < x < b$$

while $\int_a^b f(x)\, dx = 1$

Example 4. Find the expected value or mean of the variable X with PDF as defined in Example 2.

Solution:

$$\mu_x = \int_{-\infty}^{\infty} x\, f(x)\, dx = \int_0^1 x\, [12x^2 (1 - x)]\, dx$$

$$= \left| \frac{12x^4}{4} \; \frac{-12x^5}{5} \right|_0^1$$

$$= 0.6$$

Now, try the following exercise.

E4) Find the mean of a variable X with PDF given by f(x) as

$$f(x) = 5 - x, \; 0 \leq x \leq 5$$
$$= 0, \quad \text{otherwise}$$

II. The Variance of a Continuous Random Variable

The variance V(x) of a continuous random variable X with PDF f(x) is defined as follows:

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

$$= \int_{-\infty}^{\infty} x^2 f(x)\, dx - \mu_x^2 \text{ (check how?).}$$

The variance is a measure of the dispersion (or scatter) of the values of X. Its square root $\sqrt{V(x)}$ is called the standard deviation of X and is denoted by σ_x.

Example 5. Calculate the variance of the variable defined in Example 2.

Solution:

$$V(X) = \sigma_x^2 = \int_0^1 x^2 [12x^2 (1 - x)]\, dx - (\mu_x)^2$$

$$= \left| \frac{12x^5}{5} - 12\frac{x^6}{6} \right|_0^1 - (.6)^2$$

$$= \frac{12}{5} - 2 - (.6)^2 = .4 - .36 = .04.$$

Its standard deviation σ_x is therefore $\sqrt{(.04)} = .2$

E5) Find the Var (X) in E4).

Example 6: A continuous random variable X has the PDF

$$f(x) = 2\, x^k, \; 0 < x \leq 1$$
$$= 0, \quad \text{otherwise}$$

Find (a) the value of the constant k

(b) CDF of X

(c) Mean of X

(d) Variance and standard deviation of X.

Solution: For (a), we, must have

$$\int_0^1 f(x)\, dx = 1,$$

so that $\int_0^1 x^k \, dx = 1$

which implies $2 \left| \frac{x^{k+1}}{k+1} \right|_0^1 = 1$

This gives

$k = 1.$

for (b), we know that CDF is given by F(x), where $F'(x) = f(x)$

$$\text{or } F(x) = \int_{-\infty}^{x} f(x) \, dx$$

$$= \int_{-\infty}^{x} 2x \, dx = x^2$$

For the sake of convenience we shall write μ instead of μ_x σ instead of σ_x

for $0 < x \le 1$ and $F(x) = 0$, otherwise (check why ?).

Hence CDF is given by

$$F(x) = x^2, \quad 0 < x \le 1$$
$$= 0, \quad \text{otherwise}$$

For (c), we have

$$\text{Mean} = \mu = \int_0^1 f(x) \, dx$$

$$= \int_0^1 x.2x \, dx = 2 \int_0^1 x^2 \, dx$$

$$= 2. \left| \frac{x^3}{3} \right|_0^1 = 2.\frac{1}{3} = \frac{2}{3}$$

For (d), we have

$$\text{Var}(x) = \sigma^2 = \int_0^1 (x - \mu)^2 \quad f(x) \, dx \int_0^1 \left(x - \frac{2}{3} \right)^2 .2x \, dx$$

$$= \int_0^1 \left(x^2 + \frac{4}{9} - \frac{4}{3} x \right) 2x \, dx$$

$$= \int_0^1 \left(2x^3 + \frac{8}{3} x^2 + \frac{8x}{9} \right) dx$$

$$= 2 \left| \frac{x^4}{4} \right|_0^1 - \frac{8}{3} \left| \frac{x^3}{3} \right|_0^1 + \frac{8}{9} \left| \frac{x^2}{2} \right|_0^1$$

$$= 2 \frac{1}{4} - \frac{8}{3} . \frac{1}{3} + \frac{8}{9} . \frac{1}{2}$$

$$= \frac{1}{2} - \frac{8}{9} + \frac{4}{9} = \frac{9 - 16 + 8}{18} = \frac{1}{18}$$

Hence

$$SD = \frac{1}{\sqrt{18}} = \frac{1}{3\sqrt{2}}$$

Note that the negative root here, is ruled out because it is outside the range of x, $(0 \le x \le 1)$.

E 6) A continuous random variable X has the PDF as

$$f(x) = k(1 - x), \ 0 \le x < 1,$$
$$= 0, \text{ otherwise.}$$

Find (a) value of the constant k

(b) CDF

(c) mean

(d) variance and S.D.

The variance may often be worked out if we use the expectation E (X) called the expectation operator. We know that

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

Generalising it, we have expectation for any function g(x) as

$$E[g(x)] = \int_{-\infty}^{\infty} f(x)\, g(x)\, dx$$

Then

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

To illustrate this method, let us take the following example:

Example 7: The PDF of a continuous random variable X is given by

$$f(x) = .02(10 - x), \ 0 \le x < 10,$$
$$= 0, \quad \text{otherwise}$$

Solution: Here

$$E(X) = \int_{0}^{10} .02(10 - x)\, x\, dx$$

$$= .02 \left| 5x^2 - \frac{x^3}{3} \right|_0^{10} = \frac{10}{3}$$

$$E(X^2) = \int_{0}^{10} x^2 (.02)(10 - x)\, dx$$

$$= .02 \left| 10\frac{x^3}{3} - \frac{x^4}{4} \right|_0^{10} = \frac{50}{3}$$

Therefore

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

$$= \frac{50}{3} - \left(\frac{10}{3}\right)^2 = \frac{50}{9}$$

Hence

$$\text{S.D.} = \sqrt{\frac{50}{9}} = \frac{5\sqrt{2}}{3}$$

E7) A random variable X has the PDF as

$$f(x) = cx(6 - x)^2 \ 0 < x < 6$$
$$= 0, \quad \text{elsewhere.}$$

Calculate the mean, variance and S.D.

(**Hint:** First find the value of the constant c by using $\int f(x)dx = 1$ and then use the method of Example 7.)

14.3 TYPES OF CONTINUOUS DISTRIBUTIONS

In Section 14.2, you have seen some of the most important **Continuous Probability Distributions.** In fact, there are several types of continuous probability distributions. The graphs of some of these distributions are sometimes of geometrical shapes (rectangular figures say) or show some amount of skewness or in some cases, these graphs may be perfectly symmetric curves, as shown in the Figure 5.

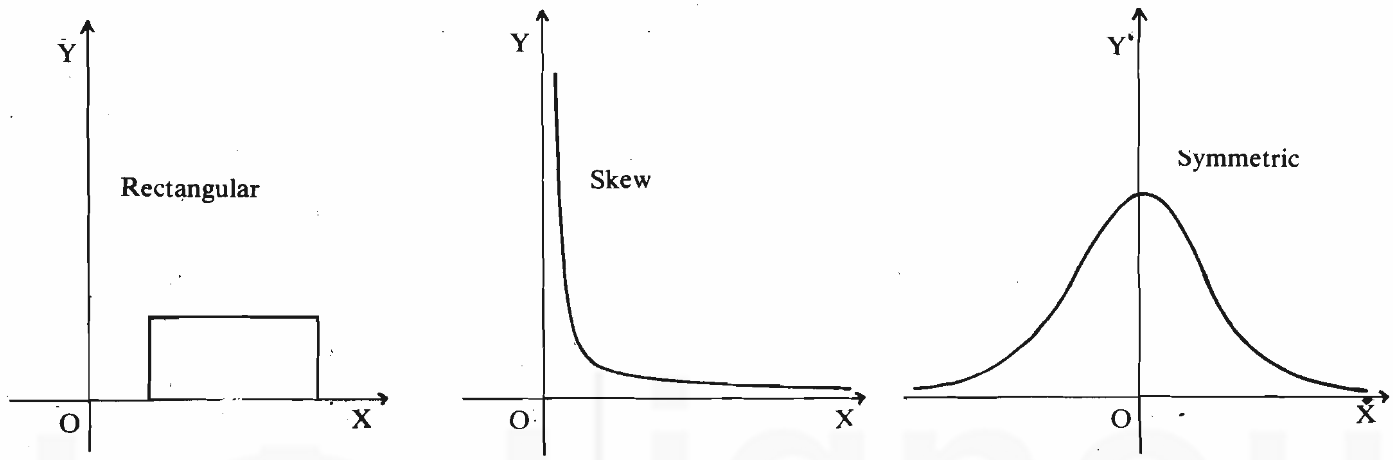

Fig. 5

In this unit, we would like to discuss only the following three continuous distributions.

I Uniform Continuous Distribution
II Exponential Distribution
III Normal Distribution.

In this section, we shall confine our discussion to uniform and exponential distributions. The normal distribution shall be discussed in Section 14.4.

14.3.1 Uniform Continuous Distribution

A uniform distribution is one in which the probability values (frequencies) are equally or uniformly distributed. Continuous random variable is said to be uniformly distributed in the interval (a, b) if it is equally likely to lie anywhere in this interval but cannot lie outside it. In other words, it is the probability distribution in which the CRV assumes all its values in an interval with equal probabilities.

For example, if we measure the height in inches of a group of men and consider only the fractional part of each height ignoring the integral part, then the result will be a random variable which must take all values between 0 and 1. The probability of this variable will be nearly evenly or equally distributed over this interval. Moreover, in this example the density function of ~~such~~ a random variable is always a constant which must be 1 since the area under the density function must be 1. Hence,

$$f(x) = 1, \quad 0 \le x \le 1$$
$$= 0, \quad \text{otherwise}$$

is the density function in this case.

The integration of this function with respect to x gives us the cumulative distribution function F(x) as

$$F(x) = x, \quad 0 < x \le 1$$
$$= 0, \quad x \le 0$$
$$= 1, \quad x \ge 1$$

In general, the density function of a continuous random variable X with uniform probability distribution is given by

$$f(x) = \frac{1}{(b-a)} \quad a \le x \le b, \quad a \ne b.$$

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

The graph of f(x) is as shown in the Figure 6.

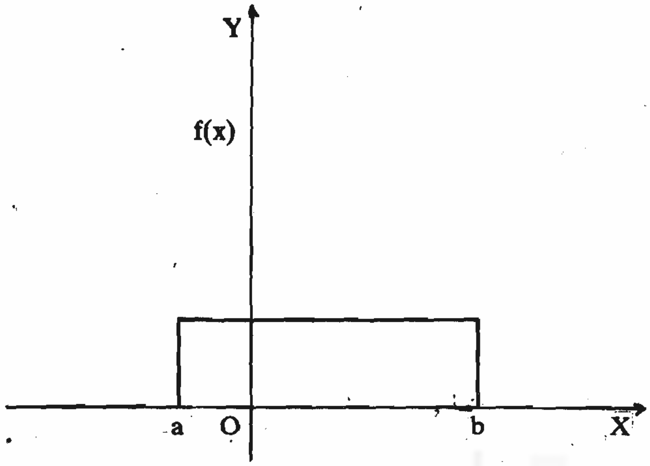

Fig. 6

The CDF of X is given by

$$F(x) = \int_a^x \frac{1}{b-a}\,dx \qquad \text{for } a \le x \le b, \quad a \ne b$$

$$\text{When } x = a, \text{ then } F(x) = \int_a^a \frac{1}{b-a}\,dx = 0 \left(\because \int_a^a f(x)\,dx = 0\right)$$

$$\text{When } x = b, \text{ then } F(x) = \int_a^b \frac{1}{b-a}\,dx = 1 \left(\because \int_a^b f(x)\,dx = 1\right)$$

Hence, we have CDF of X as

$$F(x) = 0 \quad , \quad x \le a$$

$$= \int_a^b \frac{1}{b-a}\,dx, \quad a < x < b$$

$$= 1 \quad , \quad x \ge b$$

The graph of F(x) is as shown in Figure 7.

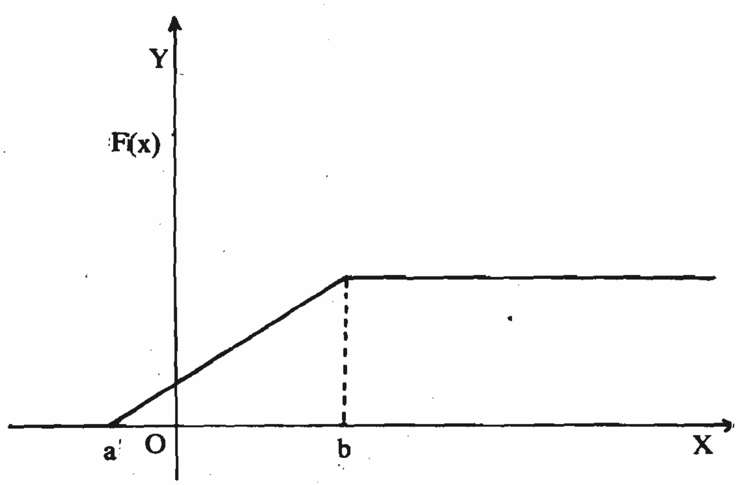

Fig. 7

The CRV with uniform distribution is also called a uniform random variable. The uniform random variable is also called a rectangular variable and its corresponding probability distribution is called a rectangular distribution. It may be noted that the rectangular variable has the following property. If X is rectangular in (a, b) and we consider two subintervals (c, d), (c′, d′) of (a, b), (i.e. $a \le c < d \le b$, $a \le c' < d' \le b$ such that $|d - c| = |d-'c|$, then the probability of its having a value in (c, d) is the same as the probability of having a value in (c′, d′). If $|d - c| = 2|d' - c'|$. Then the former probability is exactly twice the latter probability.

For instance, suppose X has a uniform distribution in the interval (-1, 1) with its PDF as

$f(x) = 1/2, -1/2 < x < 1/2$

$= 0$ otherwise.

Then the probability of X taking a value in (–1/2, 1/2). The probability of X taking a value in (–1/3, 1/6), which has half the length of (–1/2,1/2), is 1/4.

Mean and Variance

The mean and variance of a rectangular variable on (a, b) are respectively

$$\mu = \int_a^b x f(x)\, dx = \frac{1}{b - a} \quad \frac{(b^2 - a^2)}{2}$$

$$= \frac{(b+a)}{2},$$

$$\text{and } V(X) = \sigma^2 = \int_a^b x^2 f(x)\, dx - \frac{(b + a)^2}{2}$$

$$= \frac{b^3 - a^3}{3(b - a)} - \frac{(b + a)^2}{2}$$

$$= \frac{4(b^2 + ab + a^2) - 3(b^2 + 2ab + a^2)}{12}$$

$$= \frac{(b - a)^2}{12}$$

Example 8: Find the mean and variance of X with PDF given as

$f(x) = 1/2, -1/2 < x < 1/2$

$= 0$, otherwise,

$$\text{Mean} = \mu = \int_{-1/2}^{1/2} x f(x)\, dx = \int_{-1/2}^{1/2} \frac{1}{2} x\, dx = \frac{1}{2}\left|\frac{x^2}{2}\right| = 0$$

$$\text{Variance} = V(X) = \int_{-1/2}^{1/2} x^2 f(x)\, dx = \frac{1}{2}\left|\frac{x^3}{3}\right|_{-1/2}^{1/2}$$

$$= \frac{1}{6} \cdot \frac{1}{4} = \frac{1}{24}$$

E8) Let X be a continuous random variable with the following distribution

$f(x) = 1, \quad 0 \le x \le 1$

$= 0$, otherwise

Find the mean, variance and S.D. of X.

14.3.2 Exponential Distribution

A number of real life situations describe another kind of continuous probability distribution. This is called an exponential distribution.

Suppose we have a collection of electric bulbs made in the same factory with the same specification. We may record the length of the time each bulb lasts before it fails. Some of the bulbs will last a short life and some will last a long time. This provides us a model of an exponential distribution.

Consider another quite familiar situation. Suppose an infectious disease like cholera has spread in a village. What is the length of time between successive reports of the disease when it is spreading in a random manner through the population? This is described by a continuous variable that can be modelled using the exponential distribution. The word exponential is derived from the exponential function as is evident from the probability density function of the exponential distribution.

If $f(x) = \alpha e^{-\alpha x}, \alpha > 0, x \geq 0$

$= 0,$ otherwise

where α is some real number called the parameter of the distribution.

We say that a random variable X with PDF f(x) is exponentially distributed with parameter α and hence is sometimes called an exponential variable. The range of the variable is $(0, \infty)$. The graph of f(x) is shown below in Figure 8.

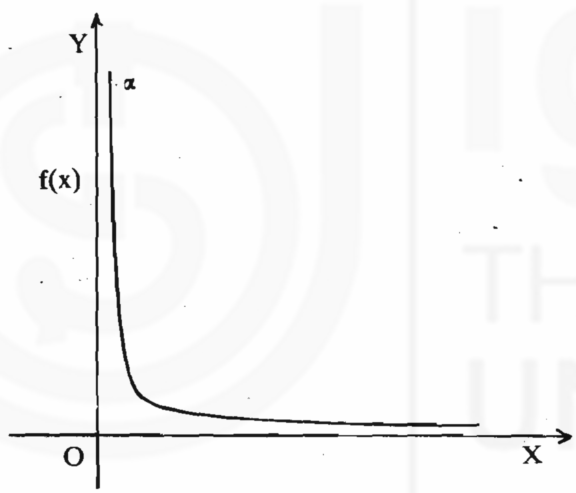

Fig. 8

The CDF of X is given by

$$F(x) = \int_0^x a e^{-\alpha x} dx = 1 - e^{-\alpha x}, x \geq 0$$

$$= 0, \text{ for } x < 0.$$

Mean and Variance

The mean and variance of an exponential variable X are, respectively,

$$\mu = E(X) = \int_0^\infty x \alpha e^{-\alpha x} \alpha x = \frac{1}{\alpha}$$

$$= V(X) = \int_0^\infty x^2 \alpha e^{-\alpha x} dx - \frac{1}{\alpha^2} = \frac{2!}{\alpha^{2+1}} \alpha - \frac{1}{\alpha^2} = \frac{1}{\alpha^2}$$

For evaluation of these integrals, you may use the standard integral

$$\int_0^\infty x^n e^{-\alpha x} dx = \frac{n!}{\alpha^{n+1}}, n = 0, 1, 2, \ldots\ldots$$

Example 9: Suppose that a variable X is exponentially distributed with parameter

$\alpha = \frac{1}{100}$

Find the CDF, mean and variance of X.

Solution: Here $f(x) = \frac{1}{100} e^{-x/100}$

Therefore,

$$F(x) = \int_0^\infty \frac{1}{100} e^{-x/100} \, dx = 1 - e^{-x/100}.$$

$$\mu = \int x^2 . \frac{1}{100} e^{-x/100} \, dx = \frac{1}{(100)}$$

$$V(X) = \int x^2 \frac{1}{100} e^{-x/100} \, dx = \frac{1}{(100)^2}$$

E9) An exponentially distributed random variable is such that its mean is equal to twice its variance. What is the value of the parameter α?

14.4 NORMAL DISTRIBUTION

Suppose a weather expert is to measure the temperature of a day during a particular interval of time say 10 A.M. to 2.00 P.M. Assume that his highest reading of 12 noon is 41°C. All the readings when graphed on a paper frqm a curve like the one given in Figure 9.

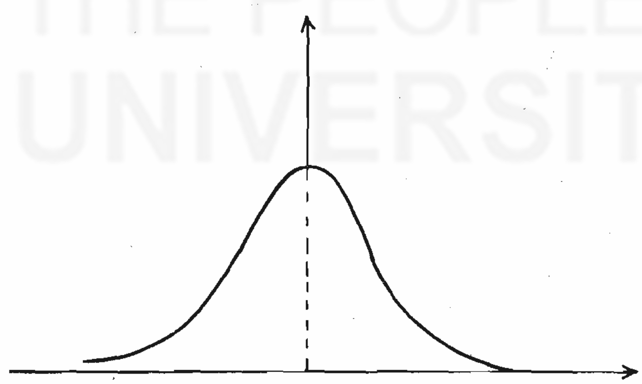

Fig. 9

Such a curve is a bell-shaped curve and is symmetrical about the line at the centre (see Fig. 9). This is called a normal curve and the distribution of the scores (which gave the above curve) is called a **Normal Distribution.**

The word 'normal' is one of the most fascinating words in statistics. We often come across the word 'normal' in our daily life. The doctor tells that the condition of the patient is 'normal'. The situation in the city after some violence is back to normal. We have other usages of the word 'normal' viz normal hydrocarbons in chemistry, normal ranges in medicine, the water level in the river is above 'normal' due to floods, the humidity is below 'normal' etc. Thus the word conveys two senses namely something desirable or something commonly found. We may, however, say that the word 'normal, means the desirability of what is commonly found so that the two senses reinforce each other.

The normal distribution is the most important continuous probability distribution in the entire area of statistics. It is one of the three main theoretical distributions.

The other two important theoretical distributions as you know from Unit 13, are binomial and Poisson. The binominal distribution was proposed in 1700 by Jacob Bernoulli (1654 – 1705), a Swiss mathematician and the Poisson distribution was given in 1837 by S.D. Poisson (1781 – 1840), a French mathematician. The normal distribution is due to De Moivre (1667 – 1754), another French mathematician who is also known for his popular De Moivre's theorem in trigonometry. However, the normal distribution is more commonly associated with the later mathematicians Gauss (1777 – 1855), a German and Laplace (1749 – 1827), a French. That is why physicists or engineers often call it the Gaussian distribution but in France, it is called Laplacean distribution. K. Pearson (1857 – 1936) a British mathematician, appears to have coined the name 'normal'. The normal distribution is, without doubt, the most important distribution in theoretical statistics.

Definition: A continuous random variable X is said to have a normal distribution if its probability density function is of the form

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(x-\mu)^2}{2\sigma^2}} \quad \sigma > 0, \quad -\infty < x < \infty$$

The density function here consists of the two parameters μ (mean) and σ (standard deviation). The variable X is generally called a **normal random variable**. The density function f(x) is also denoted by N (μ, σ^2).

The graph of the density function f(x) is a bell shaped curve or like a cocked hat as shown in the Figure 10.

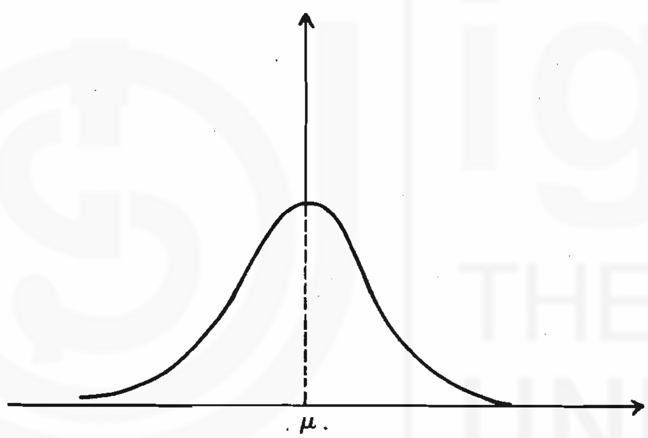

Fig. 10

The curve is generally called a normal curve. The normal curve has the following important properties:

I The mean and variance of a normal variable are μ and σ. Once the mean μ and the standard deviation σ are specified, then the normal curve is completely determined.
For example, if $\mu = 50$, $\sigma = 5$ then the ordinates of f(x) can be easily computed for various values of x and the curve can be drawn.

II The curve attains its maximum at the point $x = \mu$. In other words, the function f(x) is maximum for $x = \mu$, i.e. at the mean.

III The curve is symmetrical about a vertical axis through the point $x = \mu$ i.e. through the mean. For example, if $x = \mu + 3$, then $(x - \mu)^2 = 9$, if $x = \mu - 3$, then also $(x - \mu)^2 = 9$. In both the cases f(x) remains the same.

IV The curve approaches infinity along the horizontal axis in either direction from the mean.

V The total area under the curve and above the horizontal axis is equal to 1. In view of this property, the area under the curve between two ordinates say $X = a$ and $X = b$, $a < b$ represents the probability that X lies between a and b, denoted by $P(a < X < b)$ as shown in the Figure 2.

To find the probability distribution of a normal random variable, we have to calculate the corresponding areas carved by the normal curves. The shapes and hence the areas under the normal curves will be different according to the different values of the mean μ and the standard deviation σ. In this way, it becomes a tedious task to sketch separate curves for every possible values of μ and σ and hence find the required areas. To avoid this, one could suggest the method of integration by using the concept of commulative density function as has been done in the case of other continuous probability distributions. The CDF in this case is given by

$$F(x) = \int \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(x-\mu)^2}{2\sigma^2}} dx$$

But unfortunately, the analytical integration of this function is not possible. One may try to integrate it numerically but then that requires new methods with which you may not be familiar. Thus to avoid all these difficulties, we take the help of the method of transformation of the variable. In other words, we replace the normal variable X with mean μ and standard deviation σ by a new variable Z with mean 0 and standard deviation 1 i.e.

$$N(\mu, \sigma^2) \longrightarrow N(0, 1).$$

This can be done by means of the transformation

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

Then, it is easy to verify that

Mean of Z = 0

Variance of Z = 1

Accordingly, we define a new normal variable Z and call it standard normal variable.

Standard Normal Distribution

A normal variable with mean zero and standard deviation 1 is called a standard normal variable. The distribution of this variable is called a standard normal distribution. Its density function is given by

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

The graph of this standard normal distribution is shown in the Figure 11.

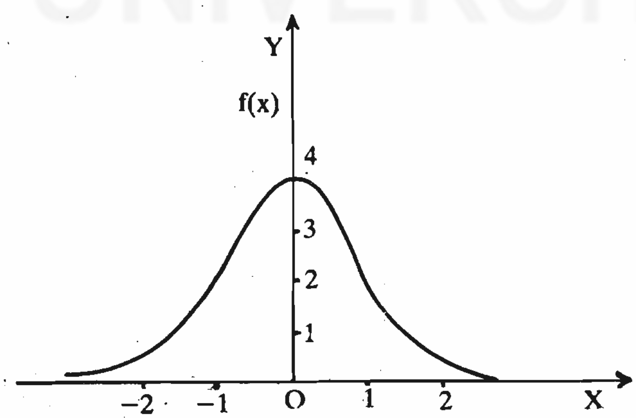

Fig. 11

The commulative distribution function (cdf) of Z is given by $\phi(z)$ where

$$\phi(z) = P(Z \le z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{(-1/2) z^2} dz$$

Note that we use the special notation $\phi(z)$ instead of the usual F(Z) for the normal distribution function.

It is not possible to evaluate this integral by the methods of integration you have learnt in Unit 8. However, numerical approximations for integral of this type can be done. The same have been tabulated and are given in Table 1. The bell-shaped curve in Fig. 11 represents the graph of the PDF of Z and the shaded area represents $\phi(z)$. Because of the symmetry of the standard normal PDF, we have

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

for all real z. In view of this, in the table below we list only those probabilities $\phi(z)$ associated with non-negative values of Z. We denote

$$P(Z \geq b) = P(Z > b) = \phi(b),\ P(Z \leq c) = \phi(c) = P(Z < c),$$

$$P(a \leq Z \leq b) = \phi(b) - \phi(a),\ P(a \geq Z \geq b) = \phi(a) - \phi(b)$$

Example 10 : If Z is N (0, 1), then find $P(0 < Z < 2)$, $P(1.25 < z < 2.75)$ and $P(-1.65 < Z < .70)$.

Solution: $P(0 \leq Z \leq 2) = \phi(2) - \phi(0) = .9772 - .5000 = .4774$ (Refer to Table 1)

$P(1.25 \leq Z \leq 2.75) = \phi(2.75) - \phi(1.25)$

$= .9970 - .8944$ (Refer to Table 1)

$= .1026$

$P(-1.65 \leq Z \leq .70) = \phi(.70) - \phi(-1.65) \quad = \phi(.70) - [1 - \phi(1.65)]$

$= \phi(.70) - 1 + \phi(1.65) = .7580 [-1 + .9506] = .7085$

TABLE 1

Values of z, the standard normal variable, from 0-0 by steps of 0.01 to 3.9 showing the cumulative probability. (Probability correct to 4 decimal places)

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9987	.9890
.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
.2	.9993	.9994	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998
.5	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998
.6	.9998	.9998	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999
.7	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999
.8	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999
.9	1.0000									

E10) If Z is N (0, 1), find the constants a, b, c such that

$$P(0 \le Z \le a) = .4147$$
$$P(Z > b) = .05$$
$$P(|Z| \le c) = .95$$

To find the probabilities about the general normal variable X which is N (μ, σ), we can use the following method.

$$P(a \le X \le b) = P\left(\frac{a-\mu}{\sigma} \le \frac{x-\mu}{\sigma} \le \frac{b-\mu}{\sigma}\right)$$
$$= \phi\left(\frac{b-\mu}{\sigma}\right) - \phi\left(\frac{a-\mu}{\sigma}\right)$$

Let us illustrate it by the following example :

Example 11 : If X is N (3, 16), then ($\mu = 3$, $\sigma = 16$), then find

(a) $P(4 \le X \le 8)$

(b) $P(0 \le X \le 5)$

(c) $P(-2 \le X \le 1)$

Solution : (a) $P(4 \le X \le 8) = P\left(\frac{4-3}{4} \le \frac{X-3}{4} \le \frac{8-3}{4}\right)$

$$= P(.25 \le Z \le 1.25)$$
$$= \phi(1.25) - \phi(.25)$$
$$= .8944 - .5987 = .2957$$

(b) $P(0 \le X \le 5) = P\left(\frac{0-3}{4} \le Z \le \frac{5-3}{4}\right)$

$$= \phi(.5) - \phi(-.75) = .4649$$

(c) $P(-2 \le X \le 1) = P\left(\frac{-2-3}{4} \le Z \le \frac{1-3}{4}\right)$

$$= \phi(-.5) - \phi(-1.25) = .2029$$

E11) If X is N (25, 36), show that $P(|X - 25| < 12) = .9544$

$P(|X - 25| \le 12) = .9544$

Hint : $|x - a| < b \iff a - b \le x \le a + b$

14.5 SUMMARY

Let us now summarise what we have done in this unit.

1 In Section 14.2, we have reviewed the definition of a continuous random variable which has been discussed in Unit 11. We also recalled the meaning of the probability distributions discussed in Unit 13 and defined the probability distributions of a continuous random variable.

2 In Section 14.3, we have discussed various types of continuous probability distributions.

3 Some important continuous probability distributions like normal distributions, standard normal distributions have been discussed in Section 14.4.

14.6 SOLUTIONS/ANSWERS

E1) a) $f(2.4) = \frac{2.4+1}{8} = \frac{3.4}{8}$

$f(3.5) = \frac{3.5+1}{8} = \frac{4.5}{8}$

$$P(2.4 < X < 3.5) = \frac{\left(\frac{3.4}{8} + \frac{4.5}{8}\right)(1.5)}{2} = .54$$

E2) $\int_{2.4}^{3.5} \frac{x+1}{8} dx$. Complete the solution by the methods of integration.

E3) $F(x) = \int_0^3 x f(x) = \int_0^3 4x(3-x) dx$. Complete the solution.

E4) Here $\mu^x = \int_0^5 x f(x) dx = \int_0^5 x(5-x) dx$.

$$= \int_0^5 (5x - x^2) dx = \left| 5\frac{x^2}{2} - \frac{x^3}{2} \right|_0^5$$

E5) $V(X) = \sigma = \int_0^5 x^2 (5-x) dx$. Calculate the integral.

E6) $k = 2, \mu = \frac{1}{3}, V = \frac{1}{18}, \sigma = \frac{1}{\sqrt{18}} = \frac{1}{3\sqrt{2}}$

E7) $\int_0^6 c\, x(6-x)^2 dx = 1 \Rightarrow c \int_0^6 (36x + x^3 - 12x^2) dx = 1$ or

$$c \left| 36.\frac{x^2}{2} + \frac{x^4}{4} - 12 - \frac{x^3}{3} \right|_0^6 = 1$$

$$c = \frac{1}{108}$$

Now you can find the mean, variance and hence the S.D.

E8) Mean $= \frac{1}{2}$, Variance $\frac{1}{3}$. S.D. $= \frac{1}{\sqrt{3}}$.

E9) $\mu = \frac{1}{\alpha}$, $V = \frac{1}{\alpha^2}$. Under the given condition, we have

$\mu = 2V$ i.e. $\frac{1}{\alpha} = \frac{2}{\alpha^2}$. Find α, now.

E10) $.4147 = P(0 \le Z \le a) = \phi(a) - \phi(0)$

$\phi(a) = .4147 + \phi(0) = .4147 + .5000$

$= .9147$

$a = 1.37$ (from the table)

again $P(Z > b) = .05$ gives

$$\phi(b) = .05$$

$$\Rightarrow b = .5199$$

Now

$$P(|Z|) \le c) = (-c \ge Z \ge c)$$

$$(-c \le Z \le c) = .95$$

$$\Rightarrow \quad \phi(c) - \phi(-c) = .95$$

$$\phi(c) - [1 - \phi(c)] = .95$$

$$2\phi(c) = 1.95$$

$$\phi(c) = .9750$$

$$\Rightarrow \quad c = 1.96.$$

E11) Since $|x - a| < b <==> a - b \le x \le a + b$, therefore,

$$|x - 25| \le 12 \Rightarrow 25 - 12 \le x \le 25 + 12$$

$$\Rightarrow 13 \le x \le 37$$

Hence, $P(|X - 25| \le 12) = P(13 \le X \le 37)$

Now complete the solution.