UNIT 14 PROBABILITY DISTRIBUTIONS

STRUCTURE

14.0 OBJECTIVES

After studying this unit, you should be able to:

- differentiate between frequency distribution and probability distribution,
- become aware of the concepts of random variable and probability distribution,
- appreciate the usefulness of probability distributions in decision-making,
- identify situations where discrete probability distributions can be applied,
- fit a binomial distribution and poisson distribution to the given data,
- identify situations where continuous probability distributions can be applied, and
- appreciate the usefulness of continuous probability distributions in decision-making.

14.1 INTRODUCTION

A probability distribution is essentially an extension of the theory of probability which we have already discussed in the previous unit. This unit introduces the concept of a probability distribution, and to show how the various basic probability distributions (binomial, poisson, and normal) are constructed. All these probability distributions have immensely useful applications and explain a wide variety of business situations which call for computation of desired probabilities.

By the theory of probability

$$P(H_1) + P(H_2) + \ldots\ldots P(H_n) = 1$$

This means that the unity probability of a certain event is distributed over a set of disjointed events making up a complete group. In general, a tabular recording of the probabilities of all the possible outcomes that could result if random (chance) experiment is done is called **"Probability Distribution"**. It is also termed as theoretical frequency distribution.

Frequency Distribution and Probability Distribution

One gets a better idea about a probability distribution by comparing it with a frequency distribution. It may be recalled that the frequency distributions are based on observation and experimentation. For instance, we may study the profits (during a particular period) of the firms in an industry and classify the data into two columns with class intervals for profits in the first column, and corresponding classes' frequencies (No. of firms) in the second column.

The probability distribution is also a two-column presentation with the values of the random variable in the first column, and the corresponding probabilities in the second column. These distributions are obtained by expectations on the basis of theoretical or past experience considerations.Thus, probability distributions are related to theoretical or expected frequency distributions.

In the frequency distribution, the class frequencies add up to the total number of observations (N), where as in the case of probability distribution the possible outcomes (probabilities) add up to 'one'. Like the former, a probability distribution is also described by a curve and has its own mean, dispersion, and skewness.

Let us consider an example of probability distribution. Suppose we toss a fair coin twice, the possible outcomes are shown in Table 14.1 below.

Table 14.1: Possible Outcomes from Two-toss Experiment of a Fair Coin

No. of possible outcomes	Ist toss	2nd toss	No. of Heads on two tosses	Probability of the possible outcomes
1	Head	Head	2	$0.5 \times 0.5 = 0.25$
2	Head	Tail	1	$0.5 \times 0.5 = 0.25$
3	Tail	Head	1	$0.5 \times 0.5 = 0.25$
4	Tail	Tail	0	$0.5 \times 0.5 = 0.25$
				Total = 1.00

Now we are intestered in framing a probability distribution of the possible outcomes of the number of Heads from the two-toss experiment of a fair coin. We would begin by recording any result that did not contain a head, i.e., only the fourth outcome in Table 14.1. Next, those outcomes containing only one head, i.e., second and third outcomes (Table 14.1), and finally, we would record that the first outcome contains two heads (Table 14.1). We recorded the same in Table 14.2 to highlight the number of heads contained in each outcome.

Table 14.2: Probability Distribution of the Possible No. of Heads from Two-toss Experiment of a Fair Coin

No. of Heads (H)	Tosses	Probability of outcomes P (H)
0	(T, T)	1/4 = 0.25
1	(H, T) + (T, H)	1/2 = 0.50
2	(H, H)	1/4 = 0.25

We must note that the above tables are not the real outcome of tossing a fair coin twice. But, it is a theoretical outcome, i.e., it represents the way in which we expect our two-toss experiment of an unbaised coin to behave over time.

14.2 TYPES OF PROBABILITY DISTRIBUTION

Probability distributions are broadly classified under two heads: (i) Discrete Probability Distribution, and (ii) Continuous Probability Distribution.

i) **Discrete Probability Distribution:** The discrete probability is allowed to take on only a limited number of values. Consider for example that the probability of having your birthday in a given month is a discrete one, as one can have only 12 possible outcomes representing 12 months of a year.

ii) **Continuous Probability Distribution:** In a continuous probability distribution, the variable of interest may take on any values within a given range. Suppose we are planning to release water for hydropower generation. Depending on how much water we have in the reservoir viz., whether it is above or below the normal level, we decide on the amount and time of release. The variable indicating the difference between the actual reservoir level and the normal level, can take positive or negative values, integer or otherwise. Moreover, this value is contingent upon the inflow to the reservoir, which in turn is uncertain. This type of random variable which can take an infinite number of values is called a continuous random variable, and the probability distribution of such a variable is called a continuous probability distribution.

Before we attempt discrete and continuous probability distributions, the concept of random variable which is central to the theme, needs to be elaborated.

14.3 CONCEPT OF RANDOM VARIABLES

A random variable is a variable (numerical quantity) that can take different values as a result of the outcomes of a random experiment. When a random experiment is carried out, the totality of outcomes of the experiment forms a set which is known as **sample space** of the experiment. Similar to the probability distribution function, a random variable may be discrete or continuous.

The example given in the Introduction, we have seen that the outcomes of the experiment of two-toss of a fair coin were expressed in terms of the number of heads. We found in the example, that H (head) can assume values of 0, 1 and 2 and corresponding to each value, a probability is associated. This uncertain real variable H, which assumes different numerical values depending on the outcomes of an experiment, and to each of whose value a possibility assignment can be made, is known as a **random variable**. The resulting representation of all the values with their probabilities is termed as the **probability distribution of H.**

It is customary to present the distribution as shown in Table 14.3 below.

Table 14.3: Probability Distribution of No. of Heads

H:	0	1	2
P (H:	0.25	0.50	0.25

In this case, as we find that H takes only discrete values, the variable H is called a **discrete random variable**, and the resulting distribution is a **discrete probability distribution**. The function that specifies the probability distribution of a discrete random variable is called the **probability mass function** (p.m.f.).

In the above situations, we have seen that the random variable takes a limited number of values. There are certain situations where the variable under consideration may have infinite values. Consider for example, that we are interested in ascertaining the probability distribution of the weight of one kg. coffee packs. We have reasons to believe that the packing process is such that a certain percentage of the packs slightly below one kg., and some packs are above one kg. It is easy to see that it is essentially by chance that the pack will weigh exactly 1 kg., and there are an infinite number of values that the random variable 'weight' can take. In such cases, it makes sense to talk of the probability that the weight will be between two values, rather than the probability of the weight taking any specific value. These types of random variables which can take an infinitely large number of values are called **continuous random variables**, and the resulting distribution is called a **continuous probability distribution**. The function that specifies the probability distribution of a continuous random variable is called the **probability density function** (p.d.f.).

Sometimes, for the sake of convenience, a discrete situation with a large number of outcomes is approximated by a continuous distribution. For example, if we find that the demand of a product is a random variable taking values of 1, 2, 3, …to 1,000, it may be worthwhile to treat it as a continuous variable.

In a nutshell, if the random variable is restricted to take only a limited number of values, it is termed as discrete random variable and if it is allowed to take any value within a given range it is termed as continuous random variable.

It should be clear, from the above discussion, that a probability distribution is defined only in the context of a random variable or a function of random variable. Thus in any situation, it is important to identify the relevant random variable and to find the probability distribution to facilitate decision making.

Expected Value of a Random Variable

Expected value is the fundamental idea in the study of probability distributions. For finding the expected value of a discrete random variable, we multiply each value that the random variable can assume by its corresponding probability of occurrence and then sum up all the products. For example to find out the expected value of the discrete random variable (RV) of "Daily Visa Cleared" given in the following table.

Table 14.4

Possible Nos. of the RV	Probability	Product
100	0.3	30
110	0.6	66
120	0.1	12

Hence the expected values of RV "Daily Visa Cleared" = 108

Now, we will examine situations involving discrete random variables and discuss the methods for assessing them.

14.4 DISCRETE PROBABILITY DISTRIBUTION

In the previous sections, we have seen that a representation of all possible values of a discrete random variable together with their probabilities of occurrence. It is called a discrete probability distribution. There are two kinds of distributions in the discrete probability distribution. i) Binomial Distribution, and (ii) Poisson Distribution. Let us discuss these two distributions in detail.

14.4.1 Binomial Distribution

It is the basic and the most common probability distribution. It has been used to describe a wide variety of processes in business. For example, a quality control manager wants to know the probability of obtaining defective products in a random sample of 10 products. If 10 per cent of the products are defective, he/she can quickly obtain the answer, from tables of the binomial probability distributions. It is also known as **Bernoulli Distribution**, as it was originated by Swiss Mathematician James Bernoulli (1654-1705).

The binomial distribution describes discrete, not continuous, data resulting from an experiment known as Bernoulli Process. Binomial distribution is a probability distribution expressing the probability of one set of dichotomous alternatives, i.e., success or failure.

As per this distribution, the probability of getting 0, 1, 2, …n heads (or tails) in n tosses of an unbiased coin will be given by the successive terms of the expansion of $(q + p)^n$, where p is the probability of success (heads) and q is the probability of failure (i.e. = 1– p).

Binomial law of probability distribution is applicable only when:

a) A trial results in either success or failure of an event.

b) The probability of success ‘p’ remains constant in each trial.

c) The trials are mutually independent i.e., the outcome of any trial is neither affected by others nor affects others.

Assumptions i) Each trial has only two possible outcomes either Yes or No, success or failure, etc.

ii) Regardless of how many times the experiment is performed, the probability of the outcome, each time, remains the same.

iii) The trials are statistically independent.

iv) The number of trials is known and is 1, 2, 3, 4, 5, etc.

Binomial Probability Formula:

$$P(r) = {}^nC_r \, p^r \, q^{n-r}$$

where, P (r) = Probability of r successes in n trials; p = Probability of success; q = Probability of failure = 1–p; r = No. of successes desired; and n = No. of trials undertaken.

The determining equation for ${}^{n}C_{r}$ can easily be written as:

$${}^{n}C_{r} = \frac{n!}{r!\,(n-r)!}$$

n! can be simplified as follows:

n! = n (n–1)! = n (n–1) (n–2) ! = n (n–1) (n–2) (n–3) ! and so on.

Hence the following form of the equations, for carrying out computations of the binomial probability is perhaps more convenient.

$$P(r) = \frac{n!}{r!\,(n-r)!}\,p^{r}q^{n-r}$$

The symbol '!' means 'factorial', which is computed as follows: 5! means 5 × 4 × 3 × 2 × 1 = 120. Mathematicians define 0! as 1.

If n is large in number, say, ${}^{50}C_{3}$, then we can write (with the help of the above explanation)

$${}^{50}C_{3} = \frac{50!}{3!(50-3)!} = \frac{(50)\ (49)\ (48)\ (47)!}{3!\ (47)!}$$

$$= \frac{50 \times 49 \times 48}{3 \times 2 \times 1}$$

Similarly,

$${}^{75}C_{5} = \frac{75\ !}{5\ !\ (75-5)\ !} = \frac{(75)\ (74)\ (73)\ (72)\ (71)\ (70)!}{5!\ (70)!}$$

$$= \frac{75 \times 74 \times 73 \times 72 \times 71}{5 \times 4 \times 3 \times 2 \times 1},\ \text{and so on.}$$

Characteristics of a Binomial Distribution

i) The form of the distribution depends upon the parameters 'p' and 'n'.

ii) The probability that there are 'r' successes in 'n' no. of trials is given by

$$P\ (r) = {}^{n}C_{r}p^{r}q^{n-r} = \frac{n\,!}{r\,!\,(n-n)\,!}\ p^{r}q^{n-r}$$

iii) It is mainly applied when the population being sampled is infinite.

iv) It can also be applied to a finite population, if it is not very small or the units sampled are replaced before the next trial is attempted. The point worth noting is 'p' should remain unchanged.

Let us consider the following illustration to undertand the application of binomial distribution.

Illustration 1

A fair coin is tossed six times. What is the probability of obtaining four or more heads?

Solution: When a fair coin is tossed, the probabilities of head and tail in case of an unbiased coin are equal, i.e.,

$p = q = ½$ or 0.5

∴ The probabilities of obtaining 4 heads is : $P(4) = {}^6C_4\ (1/2)^4\ (1/2)^{6-4}$

$$P(r) = \frac{n!}{r!\ (n-r)!}\ p^r q^{n-r}$$

$$P(4) = \frac{6!}{4!(6-4)!}\ (0.5)^4\ (0.5)^2$$

$$= \frac{6 \times 5 \times 4 \times 3 \times 2 \times 1}{(4 \times 3 \times 2 \times 1)\ (2 \times 1)} (0.625)\ (0.25)$$

$$= \frac{720}{(24)\ (2)}\ (0.625)\ (0.25) = 15 \times 0.625 \times 0.25$$

= 0.234

The probability of obtaining 5 heads is :

$P(5) = {}^6C_5(1/2)^5\ (1/2)^{6-5}$

$$P(5) = \frac{6!}{5!(6-5)!}\ (0.5)^5\ (0.5)^1$$

$$= \frac{6 \times 5 \times 4 \times 3 \times 2 \times 1}{5 \times 4 \times 3 \times 2 \times 1\ (1 \times 1)}\ (0.03125)\ (0.5)$$

= 6 × (0.03125) (0.5)

= 0.094

The probability of obtaining 6 heads is : $P(6) = {}^6C_6\ (1/2)^6\ (1/2)^{6-6}$

$$P(6) = \frac{6!}{6!(6-6)!}\ (0.5)^2\ (0.5)^0$$

$$= \frac{6 \times 5 \times 4 \times 3 \times 2 \times 1}{6 \times 5 \times 4 \times 3 \times 2 \times 1\ (1)}\ (0.015625)\ (1)$$

= 1 × 0.015625 × 1

= 0.016

∴ The probability of obtaining 4 or more heads is :

0.234 + 0.094 + 0.016 = 0.344

Illustration 2

The incidence of a certain disease is such that on an average 20% of workers suffer from it. If 10 workers are selected at random, find the probability that

i) Exactly 2 workers suffer from the disease
ii) Not more than 2 workers suffer from the disease
iii) At least 2 workers suffer from the disease

Solution: Probability that a worker suffering from the disease $= \frac{20}{100} = \frac{1}{5}$

i.e., $p = \frac{1}{5}$, and

The probability of a worker not suffering from the disease i.e.,

$$q = \left[1-\frac{1}{5}\right] = \frac{4}{5}$$

By binomial probability law, the probability that out of 10 workers, 'x' workers suffer from a disease is given by:

$$P(r) = {}^nC_r\ p^r\ q^{n-r}$$

$${}^{10}C_r \cdot \frac{1}{5}r.\ \frac{4}{5}^{10-r} \quad ; r = 0, 1, 2, \ldots 10$$

i) The required probability that exactly 2 workers will suffer from the disease is given by :

$$P(2) = {}^{10}C_2 \left(\frac{1}{5}\right)^2 \left(\frac{4}{5}\right)^{10-2}$$

$$= \frac{10\,!}{2\,!(10-2)\,!}\ (0.2)^2\ \ (0.8)^8 = \frac{(10)\ (9)\ (8)\,!}{(2\ \times\ 1)\ (8)\,!}\ (0.04)\ \ (0.16777)$$

$$= 45\ (0.04)\ (0.16777) = 0.302$$

ii) The required probability that not more than 2 workers will suffer from the disease is given by :

$$P\ (0) + P(1) + P(2)$$

$$P\ (0) = {}^{10}C_0 \left(\frac{1}{5}\right)^0 \left(\frac{4}{5}\right)^{10-0} = 0.107$$

$$P\ (1) = {}^{10}C_1 \left(\frac{1}{5}\right)^1 \left(\frac{4}{5}\right)^{10-1} = 0.269$$

$$P\ (2) = {}^{10}C_2 \left(\frac{1}{5}\right)^2 \left(\frac{4}{5}\right)^{10-2} = 0.302$$

Probability of not more than 2 workers suffering from the disease

$= 0.107 + 0.269 + 0.302 = 0.678$

iii) We have to find $P\ (r \geq 2)$

i.e., $P\ (r \geq 2) = 1–P\ (0) - P\ (1)$

$= 1 - 0.107 - 0.269 = 0.624$

Thus, the probability at least two workers suffering from the disease is 0.624.

Measures of Central Tendency and Dispersion for the Binomial Distribution

As discussed in the Introduction, the binomial distribution has expected values of mean (μ) and a standard deviation (σ). We now see the computation of both these statistical measures.

We can represent the mean of the binomial distribution as :

Mean (μ) = np.

where, n = Number of trials; p = probability of success

And, we can calculate the standard deviation by :

$$\sigma = \sqrt{npq}$$

where, n = Number of trials; p = probability of success; and q = probability of failure = 1–p

Illustration 3

If the probability of defective bolts is 0.1, find the mean and standard deviation for the distribution of defective bolts in a total of 50.

Solution: P = 0.1, n = 500

$\therefore$ Hence (μ) = np = 500 × 0.1 = 50

Thus, we can expect 50 bolts to be defective.

Standard Deviation (σ) $= \sqrt{npq}$

n = 500, p = 0.1, q = 1 – p = 1 – 0.1 = 0.9

$\therefore \sigma = \sqrt{500 \times .1 \times .9} = 6.71$

Fitting a Binomial Distribution

When a binomial distribution is to be fitted to observed data, the following procedure is adopted:

i) Determine the values of 'p' and 'q'. If one of these values is known, the other can be found out by the simple relationship p = 1–q and q = 1–p. If p and q are equal, we can say, the distribution is symmetrical. On the other hand if 'p' and 'q' are not equal, the distribution is skewed. The distribution is positively skewed, in case 'p' is less than 0.5, otherwise it is negatively skewed.

ii) Expand the binomial $(p + q)^n$. The power 'n' is equal to one less than the number of terms in the expanded binomial. For example, if 3 coins are tossed (n = 3) there will be four terms, when 5 coins are tossed (n = 5) there will be 6 terms, and so on.

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

Let us consider an illustration for fitting a binomial distribution.

Illustration 4

Eight coins are tossed at a time 256 times. Number of heads observed at each throw is recorded and the results are given below. Find the expected frequencies. What are the theoretical values of mean and standard deviation? Also calculate the mean and standard deviation of the observed frequencies.

No. of Heads at a throw	f	No. of heads at a throw	f
0	2	5	56
1	6	6	32
2	30	7	10
3	52	8	1
4	67		

Solution: The chance of getting a head is a single throw of one coin is ½. Hence, p = ½, q = ½, n = 8, N = 256

By expanding 256 $(½ + ½)^8$. We shall get the expected frequencies of 1, 2, … 8 heads (successes).

No. of Head (X)	**Expected Frequency = N × ^nC_r p^r q^{n-r} (Frequencies approximated)**
0	$256 \times {}^8C_0 (0.5)^0 (0.5)^8 = 1$
1	$256 \times {}^8C_1 (0.5)^1 (0.5)^7 = 8$
2	$256 \times {}^8C_2 (0.5)^2 (0.5)^6 = 28$
3	$256 \times {}^8C_3 (0.5)^3 (0.5)^5 = 56$
4	$256 \times {}^8C_4 (0.5)^4 (0.5)^4 = 70$
5	$256 \times {}^8C_5 (0.5)^5 (0.5)^3 = 56$
6	$256 \times {}^8C_6 (0.5)^6 (0.5)^2 = 28$
7	$256 \times {}^8C_7 (0.5)^7 (0.5)^1 = 8$
8	$256 \times {}^8C_8 (0.5)^8 (0.5)^0 = 1$
Total	**= 256**

If we compare the above expected frequencies with the observed frequencies, given in the illustration, we find that the two frequencies are in close agreement. This provides the basis to conclude that the observed distribution will fits the expected distribution.
The mean of the above distribution is:

$$\mu = np = 8 \times \frac{1}{2} = 4$$

The Standard Deviation is $(\sigma) = \sqrt{npq}$

$$= \sqrt{8 \times \frac{1}{2} \times \frac{1}{2}} = \sqrt{2} = 1.414$$

If we compute the mean and standard deviation of the observed frequencies, we will obtain the following values

$$\overline{X} = 4.062; \quad \text{S.D.} = 1.462$$

Note: The procedure for computation of mean and standard deviation of the observed frequencies has been already discussed in Units 8 and 9 of this course. Check these values by computing on your own.

Remark: To determine binomial probabilities quickly we can use the Binomial Tables given at the end of this block (Appendix Table 1).

Self Assessment Exercise A

1) State whether the following statements are true or false:

   a) By the theory of probability $P(H_1) + P(H_2) + \ldots P(H_n) = 1$

b) Frequency distribution is obtained by expectations on the basis of theoretical considerations.

c) In a continuous probability distribution the variable under consideration can take on any value within a given range.

d) Binomial distribution is a probability distribution expressing the probability of one set of dichotomous alternatives.

e) Binomial distribution may not be applied, when the population being sampled is infinite.

f) Random variable is a numerical quantity whose value is determined by the outcome of a random experiment.

2) Determine the following by using binomial probability formula.

a) If n = 4 and P = 0.12, then what is P (0) ?

b) If n = 10 and P = 0.40, then what is p (9) ?

c) If n = 6 and P = 0.83, then what is P (5)?

....

....

....

....

....

3) The following data shows the result of the experiment of throwing 5 coins at a time 3,100 times and the number of heads appearing in each throw. Find the expected frequencies and comment on the results. Also calculate mean and standard deviation of the theoretical values.

No. of heads:	0	1	2	3	4	5
frequency:	32	225	710	1,085	820	228

....

....

....

....

....

....

....

....

....

....

....

....

14.4.2 Poisson Distribution

Poisson distribution, developed by a French mathematician Simeon Poisson, is so known by his name. It deals with counting the number of occurances of a particular event in a specific **time interval or region of space**. It is used in practice where there are infrequently occuring events with respect to time, volume (similar units), area, etc. For instance, the number of deaths or accidents occuring in a specific time, the number of defects in production, the number of workers absent per day etc.

The binomial distribution, as discussed above, is determined by two parameters ‘p’ and ‘n’. In a number of cases ‘p’ (the probability of success) may happen to be very small (even less than 0.01) and the ‘n’ (the no. of trials) is large enough (like more than 50) so that their product “np” remains a constant, the situation is termed as “Poisson Distribution”, and it gives an approximation for binomial probability distribution formula, i.e., $P(r) = {}^nC_r\, p^r\, q^{n-r}$

The Poisson distribution process corresponds to a Bernoulli process with a very large number of trials (n) and a very low probability of success.

This would comparatively be simpler in dealing with and is given by the Poisson distribution formula as follows:

$$p\,(r) = \frac{m^r e^{-m}}{r!},$$

where, p (r) = Probability of successes desired

r = 0, 1, 2, 3, 4, … ∞ (any positive integer)

e = a constant with value: 2.7183 (the base of natural logarithms)

m = The mean of the Poisson Distribution, i.e., np or the average number of occurrences of an event.

Characteristics of the Poisson Distribution

a) It is also a discrete probability distribution and it is the limiting form of the binomial distribution.

b) The range of the random variable is $0 \le r < \infty$

c) It consists of a single parameter “m” only. So, the entire distribution can be obtained by knowing this value only.

d) It is a positively skewed distribution. The skewness, therefore, decreases when ‘m’ increases.

Measures of Central Tendency and Dispersion for Poisson Distribution

In poisson distribution, the mean (m) and the variance (s^2) represent the same value, i.e.,

Mean = variance = np = m

$$\text{S.D. } (\sigma) = \sqrt{\text{Variance}} = \sqrt{np}$$

Let us consider the following illustrations to understand the application of the poisson distribution.

Illustration 5

2% of the electronic toys produced in a certain manufacturing process turnout to be defective. What is the probability that a shipment of 200 toys will contain exactly 5 defectives ? Also find the mean and standard deviation.

Solution: In the given illustration n = 200;

Probabiliity of a defective toy $(P) = \frac{2}{100} = 0.02$

Since, n is large and p is small, the poisson distribution is applicable. Apply the formula:

$$p(r) = \frac{m^r e^{-m}}{r!}$$

The probability of 5 defective pieces in 200 toys is given by:

$$p(5) = \frac{m^5 e^{-m}}{5!}, \quad \text{where } m = np = 200 \times 0.02 = 4;$$

$$e = 2.7183 \text{ (constant)}$$

$$\therefore P(5) = \frac{4^5 \; 2.7183^{-4}}{5\times4\times3\times2\times1} = \frac{(1024)\;\frac{1}{2.7183^4}}{120}$$

$$= \frac{(1024)\;0.0183}{120} = 0.156$$

Mean = np = 200 × 0.02 = 4; $\sigma = \sqrt{np} = \sqrt{4} = 2$

Illustration 6

Find the probability of exactly 4 defective tools in a sample of 30 tools chosen at random by a certain tool producing firm by using i) Binomial distribution and ii) Poisson distribution. The probability of defects in each tool is given to be 0.02.

Solution: i) When binomial distribution is used, the probability of 4 defectives in 30 tools is given by:

$$P(4) = {}^{30}C_4 \; (0.02)^4 \; (0.98)^{26}$$

$$= 27405 \times 0.00000016 \times 0.59 = 0.00259$$

ii) When poisson distribution is used, the probability of 4 defectives in 30 tools is given by:

$$P(4) = \frac{m^4 e^{-m}}{4!}, \quad \text{where, } m = np = 30\,(0.02) = 0.6$$

$$e = 2.7183 \text{ (constant)}$$

$$\therefore P(4) = \frac{0.6^2 \; 2.7183^{-0.6}}{4\times3\times2\times1}$$

$$\frac{\text{Reciprocal [anti log } (0.6 \times \log. 2.7183)]\; 0.1296}{24}$$

$$= \frac{\text{Rec. [anti log } 0.2606]\; 0.1296}{24} = \frac{0.5485 \times 0.1296}{24}$$

= 0.00296

Remark: In general the Poisson distribution can be used as an approximation to binomial with parameter m = np, is good if n ≥20 and p ≤ 0.05.

Fitting of a Poisson Distribution

To fit a poisson distribution to a given observed data (frequency distribution), the procedure is as follows:

1) We must obtain the value of its mean i.e., m = np

2) The probabilities of various values of the random variables (r) are to be computed by using p.m.f. i.e., $p\,(r) = \dfrac{m^r e^{-m}}{r!}$

3) Each probability so obtained in step 2 is then multiplied by N (the total frequency) to get expected frequencies.

Let us consider an illustration to understand for fitting poisson distribution.

Illustration 7

The number of defects per unit in a sample of 330 units of manufactured product was found as follows:

No. of defects	No. of units
0	214
1	92
2	20
3	3
4	1

Fit a poisson distribution to the above given data.

Solution: The mean of the given frequency distribution is:

$$m = \frac{(0\times 214)+(1\times 92)+(2\times 20)+(3\times 3)+(4\times 1)}{214+92+20+3+1} = \frac{145}{330} = 0.439$$

We can write $P(r) = \dfrac{0.439^r \times e^{-0.439}}{r\,!}$. Substituting r = 0, 1, 2, 3, and 4, we get the probabilities for various values of r, as shown below:

$$(P_0) = \frac{m^r\ e^{-m}}{r!} = \frac{0.439^0 \times 2.7183^{-0.439}}{0\,!}$$

$$= \frac{1\,(0.6443)}{1} = 0.6443$$

$$N(P_0) = (P_0) \times N = 0.6443 \times 330 = 212.62$$

$$N(P_1) = (P_0) \times m/1 = 212.62 \times 0.439/1 = 93.34$$

$$N(P_2) = (P_1) \times m/2 = 93.34 \times 0.439/2 = 20.49$$

$$N(P_3) = (P_2) \times m/3 = 20.49 \times 0.439/3 = 3.0$$

$$N(P_4) = (P_3) \times m/4 = 3 \times 0.439/4 = 0.33$$

Thus, the expected frequencies as per poisson distribution are :

No. of defects (x)	0	1	2	3	4
Expected frequencies (No. of units) (f)	212.62	93.34	20.49	3.0	0.33

Note: We can use Appendix Table-2, given at the end of this block, to determine poisson probabilities quickly.

Self Assessment Exercise B

1) What are the features of binomial and poisson distributions?

....

....

....

....

....

2) Suppose on an average 2% of electric bulbs manufactured by a company are defective. If they produce 100 bulbs in a day, what is the probability that 4 bulbs will have defects on that day ?

....

....

....

3) Four hundred car air-conditioners are inspected as they come off the production line and the number of defects per set is recorded below. Find the expected frequencies by assuming the poisson model.

No. of defects :	0	1	2	3	4	5
No. of ACs:	142	156	69	27	5	1

....

....

....

....

....

....

....

....

....

....

....

14.5 CONTINUOUS PROBABILITY DISTRIBUTION

In the previous sections, we have examined situations involving discrete random variables and the resulting probability distributions. Let us now consider a situation, where the variable of interest may take any value within a given range. Suppose that we are planning to release water for hydropower generation and irrigation. Depending on how much water we have in the reservoir, viz., whether it is above or below the 'normal' level, we decide on the quantity of water and time of its release. The variable indicating the difference between the actual level and the normal level of water in the reservoir, can take positive or negative values, integer or otherwise. Moreover, this value is contingent upon the inflow to the reservoir, which in turn is uncertain. This type of random variable which can take an infinite number of values is called a **continuous random variable**, and the probablity distribution of such a variable is called a **continuous probability distribution**.

Now we present one important probability density function (p.d.f), viz., the normal distribution.

14.5.1 Normal Distribution

The normal distribution is the most versatile of all the continuous probability distributions. It is useful in statistical inferences, in characterising uncertainities in many real life situations, and in approximating other probability distributions.

As stated earlier, the normal distribution is suitable for dealing with variables whose magnitudes are continuous. Many statistical data concerning business problems are displayed in the form of normal distribution. Height, weight and dimensions of a product are some of the continuous random variables which are found to be normally distributed. This knowledge helps us in calculating the probability of different events in varied situations, which in turn is useful for decision-making.

To define a particular normal probability distribution, we need only two parameters i.e., the mean (μ) and standard deviation (σ).

Now we turn to examine the characteristics of normal distribution with the help of the figure 14.1, and explain the methods of calculating the probability of different events using the distribution.

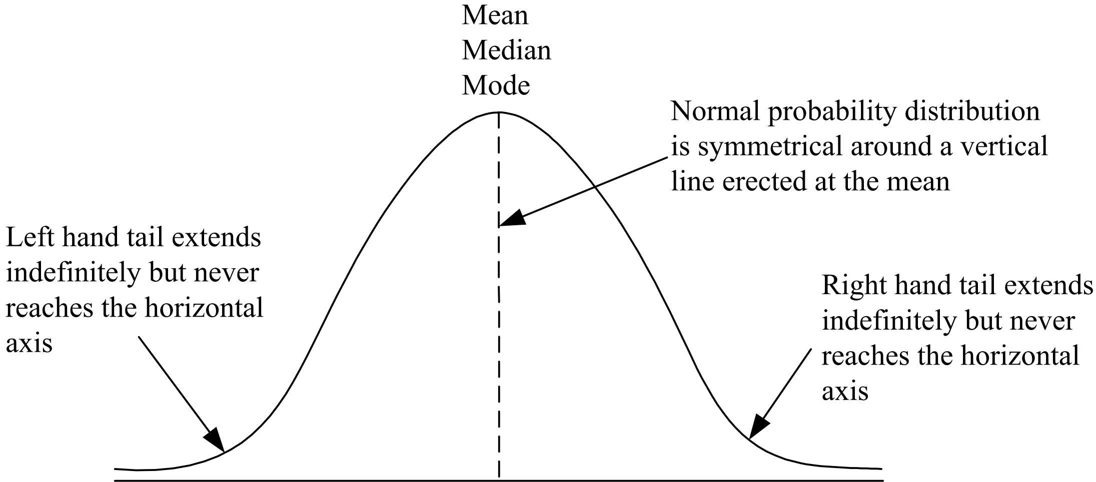


Figure 14.1: Frequency Curve for the Normal Probability Distribution

14.5.2 Characteristics of Normal Distribution

1) The curve has a single peak, thus it is unimodal i.e., it has only one mode and has a bellshape.

2) Because of the symmetry of the normal probability distribution (skewness = 0), the median and the mode of the distribution are also at the centre. Thus, for a normal curve, the mean, median and mode are the same value.

3) The two tails of the normal probability distribution extend indefinitely but never touch the horizontal axis.

Areas Under the Normal Curve

The area under the normal curve (Fig. 14.1) gives us the proportion of the cases falling between two numbers or the probability of getting a value between two numbers.

Irrespective of the value of mean (μ) and standard deviation (σ), for a normal distribution, the total area under the curve is 1.00. The area under the normal curve is approximately distributed by its standard deviation as follows:

$\mu \pm 1\ \sigma$ covers 68% area, i.e., 34.13% area will lie on either side of μ.

$\mu \pm 2\sigma$ covers 95.5% area, i.e., 47.75% will lie on either side of μ.

$\mu\sigma \pm 3\sigma$ covers 99.7% area, i.e., 49.85% will lie on either side of μ.

Using the Standard Normal Table

The areas under the normal curve are shown in the Appendix Table-3 at the end of this block. To use the standard normal table to find normal probability values, we follow two steps. They are:

Step 1: Convert the normal distribution to a standard normal distribution.

The standard random variable Z, can be computed as follows:

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

Where,

X = Value of the random variable with which we are concerned.

μ = mean of the distribution of this random variable

σ = standard deviation of this distribution.

Z = Number of standard deviations from X to the mean of this distribution.

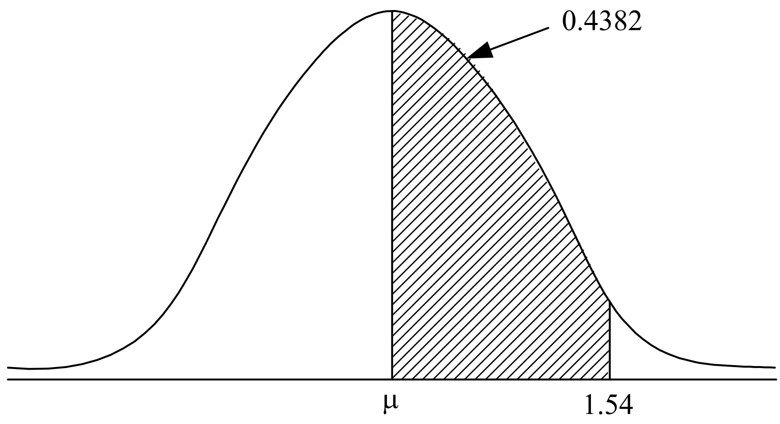
Step 2: Look up the probability of z value from the Appendix Table-3, given at the end of this block, of normal curve areas. This Table is set up to provide the area under the curve to any specified value of Z. (The area under the normal curve is equal to 1. The curve is also called the **standard probability curve**).

Let us consider the following illustration to understand as to how the table should be consulted in order to find the area under the normal curve.

Illustration 8

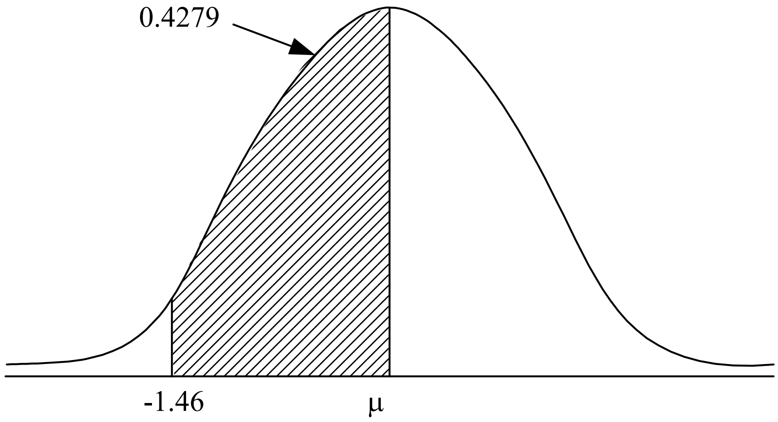
(a) Find the area under the normal curve for Z = 1.54.

Solution: Consulting the Appendix Table-3 given at the end of this block, we find the entry corresponding to Z = 1.54 the area is 0.4382 and this measures the Shaded area between Z = 0 and Z = 1.54 as shown in the following figure.



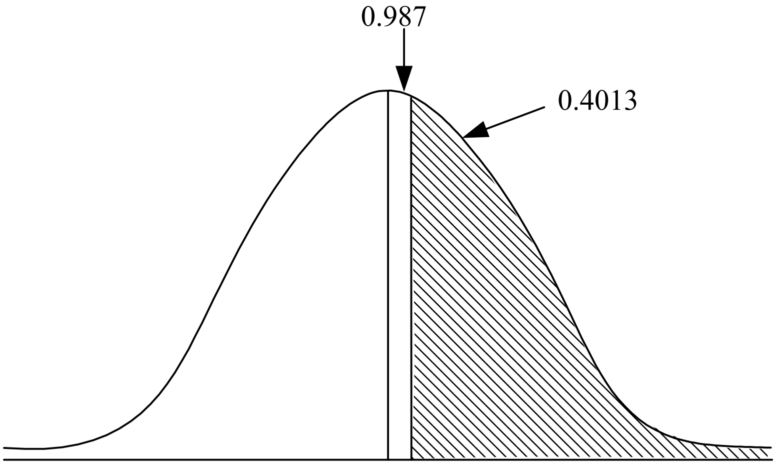
(b) Find the area under normal curve for Z = –1.46

Solution: Since the curve is symmetrical, we can obtain the area between z = –1.46 and Z = 0 by considering the area corresponding to Z = 1.46. Hence, when we look at Z of 1.46 in Appendix Table-3 given at the end of this block, we see the probability value of 0.4279. This value is also the probability value of Z = –1.46 which must be shaded on the left of the μ as shown in the following figure.



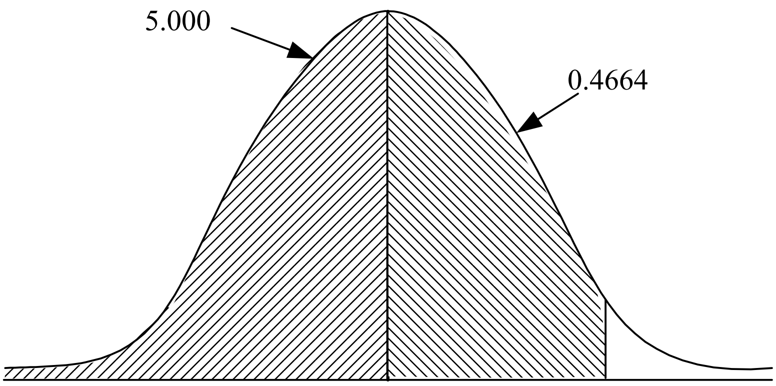
(c) Find the area to the right of Z = 0.25

Solution: If we look up Z = 0.25 in Appendix table, we find the probable area of 0.0987. Subtract 0.0987 (for Z = 0.25) from 0.5 getting 0.4013 (.5–.0987 = 0.4013).



d) Find the area to the left of Z = 1.83.

Solution: If we are interested in finding the area to the left of Z (positive value), we add 0.5000 to the table value given for Z. Here, the table value for Z (1.83) = 0.4664. Therefore, the total area to the left of Z = 0.9664 (0.5000 + 0.4664) i.e., equal to the shaded area as shown below:



Now let us take up some illustrations to understand the application of normal probability distribution.

Illustration 9

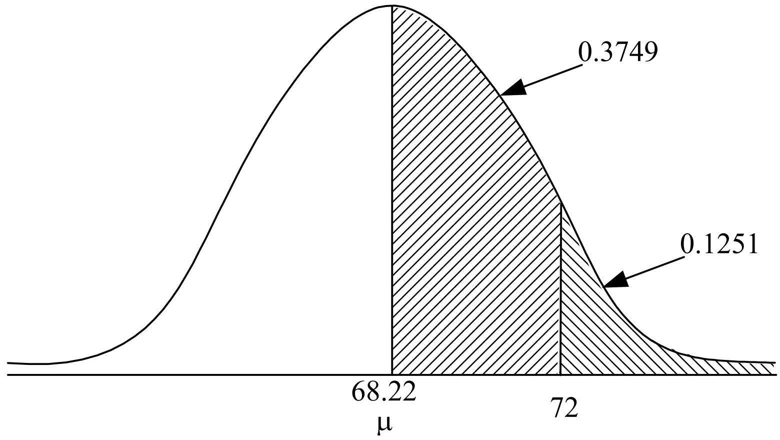
Assume the mean height of soliders to be 68.22 inches with a variance of 10.8 inches. How many soldiers in a regiment of 1,000 would you expect to be over six feet tall ?

Solution: $Z = \dfrac{X - \mu}{\sigma}$

X = 72 inches; μ = 68.22 inches; and $\sigma = \sqrt{10.8} = 3.286$

$\therefore Z = \dfrac{72 - 68.22}{3.286} = 1.15$

for Z = 1.15 the area is .3749 (Appendix Table-3).



Area to the right of the ordinate at 1.16 from the normal table is (0.5–0.3749) = 0.1251. Hence, the probability of getting soldiers above six feet is 0.1251 and out of 1,000 soliders, the expectation is 1,000 × 0.1251 = 125.1 or 125. Thus, the expected number of soldiers over six feet tall is 125.

Illustration 10

(a) 15,000 students appeared for an examination. The mean marks were 49 and the standard deviation of marks was 6. Assuming the marks to be normally distributed, what proportion of students scored more than 55 marks?

Solution: $Z = \dfrac{X - \mu}{\sigma}$

$X = 55;\ \ \mu = 49;\ \ \sigma = 6$

$\therefore Z = \dfrac{55 - 49}{6} = 1$

For Z = 1, the area is 0.3413 (as per Appendix Table-3).

∴ The proportion of students scoring more than 55 marks is

0.5–0.3413 = 0.1587 or 15.87%

(b) If in the same examination, Grade 'A' is to be given to students scoring more than 70 marks, what proportion of students will receive grade 'A'?

Solution: $Z = \dfrac{X - \mu}{\sigma}$

$X = 70;\ \ \mu = 49;\ \ \sigma = 6$

$\therefore Z = \dfrac{70 - 49}{6} = 3.5$

The table gives the area under the standard normal curve corresponding to Z = 3.5 is 0.4998

Therefore, 0.02% (0.5–0.4998 = 0.0002 × 100) would score more than 70 marks. Since, there are 15,000 candidates, 3 candidates (15,000 × 0.02% = 3) will receive Grade 'A'.

Illustration 11

In a training programme (self-administered) to develop marketing skills of marketing personnel of a company, the participants indicate that the mean time on the programme is 500 hours and that this normally distributed random variable has a standard deviation of 100 hours. Find out the probability that a participant selected at random will take:

i) fewer than 570 hours to complete the programme, and

ii) between 430 and 580 hours to complete the programme.

Solution: (i) To get the Z value for the probability that a candidate selected at random will take fewer than 570 hours, we have

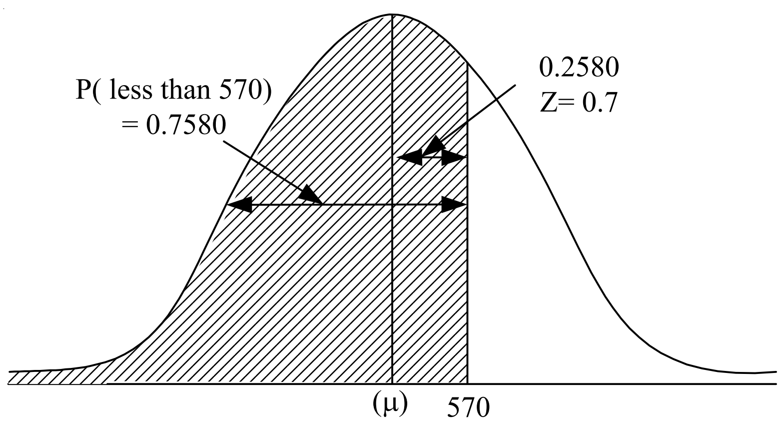
$$Z = \frac{x-\mu}{\sigma} = \frac{570-500}{100}$$

$$= \frac{70}{100} = 0.7$$

Consulting the Appendix Table-3 for a Z value of 0.7, we find a probability of 0.2580 (this probability will lie between the mean, 500 hours and 570 hours. As explained in illustration 8(d), we must add 0.5 to this probability (0.2580) that the random variable will be between the left-hand tail and the mean.

Therefore, we obtain the probability that the random variable will lie between the left-hand tail and 570 hours is 0.7580 (0.5 + 0.2580).

This situation is shown below:



Thus, the probability of a participant taking less than 570 hours to complete the programme, is marginally higher than 75 per cent.

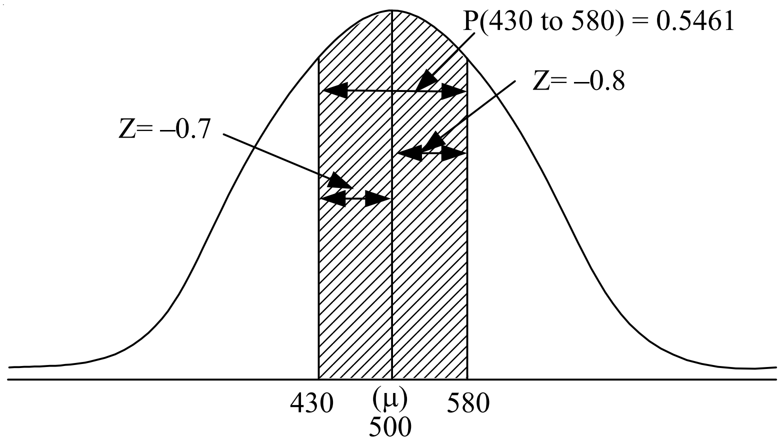
ii) In order to get the probability, of a participant chosen at random, that he will take between 430 and 580 hours to complete the programme, we must, first, compute the Z value for 430 and 580 hours.

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

$$Z \text{ for } 430 = \frac{430-500}{100} = \frac{-70}{100} = -0.7$$

$$Z \text{ for } 580 = \frac{580-500}{100} = \frac{80}{100} = 0.8$$

The table shows the probability values of Z values of –0.7 and 0.8 are 0.2580 and 0.2881 respectively. This situation is shown in the following figure.



Thus, the probability that the random variables lie between 430 and 580 hours is 0.5461 (0.2580 + 0.2881).

14.5.3 Importance and Application of Normal Distribution

This distribution was initially discovered for studying the random errors in measurements, which are encountered during the calculations of orbits of heavenly bodies. It happens because of the fact that the normal distribution follows the basic principle of errors. It is mainly for this quality that the distribution has a wide range of applications in the theory of statistics. To count a few :

- Industrial quality control
- Testing of significance
- Sampling distribution of various statistics
- Graduation of non-normal curve
- length of the leaves observed at particular times of the year.

The main purpose for using a normal distribution are:

(i) To fill a distribution of measurement for same sample data,

(ii) To approximate the distributions like Binomial, Poisson etc.

(iii) To fit sampling distribution of various statistics like mean or variance etc.

Self Assessment Exercise C

1) Given a standardized normal distribution area between the mean and positive value of Z as in Appendix Table 2) what is the probability that:

a) Z is less than +1.08?

b) Z is greater than – 0.21?

c) Z is between the mean and +1.08?

d) Z is less than – 0.27 the mean and greater than +1.06?

e) Z is between – 0.21 and the mean?

2) Give a normal distribution with $\mu = 100$ and $\sigma = 10$, what is the probability that:

   a) $X > 75$?

   b) $X < 70$?

   c) $X > 112$?

   d) $75 < X > 85$?

   e) $X < 80$ or $X > 110$?

14.6 LET US SUM UP

In this unit, we have discussed the meaning of frequency distrbution and probability distribution, and the concepts of random variables and probability distribution. In any uncertain situation, we are often interested in the behaviour of certain quantities that take different values in different outcomes of experiments. These quantities are called random variables and a representation that specifies the possible values a random variable can take, together with the associated probabilities, is called a probability distribution. The distribution of a discrete variable is called a discrete probability distribution and the function that specifies a discrete distribution is termed as a probability mass function (p.m.f.). In the discrete distribution we have considered the binomial and poisson distributions and discussed how these distributions are helpful in decision-making. We have shown the fitting of such distributions to a given observed data.

In the final section, we have examined situations involving continuous random variables and the resulting probability distributions. The random variable which can take an infinite number of values is called a continuous random variable and the probability distribution of such a variable is called a continuous probability distribution. The function that specifies such distribution is called the probability density function (p.d.f.). One such important distribution, viz., the normal distribution has been presented and we have seen how probability calculations can be done for this distribution.

14.7 KEY WORDS

Binomial Distribution: It is a type of discrete probability distribution function that includes an event that has only two outcomes (success or failure) and all the trials are mutually independent.

Continuous Probability Distribution: In this distribution the variable under consideration can take any value within a given range.

Continuous Random Variable: If the random variable is allowed to take any value within a given range, it is termed as continuous random variable.

Discrete Probability Distribution: A probability distribution in which the variable is allowed to take on only a limited number of values.

Discrete Random Variable: A random variable that is allowed to take only a limited number of values.

Normal Distribution: It is a type of continuous probability distribution with a single peaked, bell-shaped curve. The curve is symmetrical around a vertical line erected at the mean. It is also known as Gaussian distribution.

Poisson Distribution: It is the limiting form of the binomial distribution. Hence, the probability of success is very low and the total number of trials is very high.

Probability: Any numerical value between 0 and 1 both inclusive, telling about the likelihood of occurrence of an event.

Probability Distribution: A curve that shows all the values that the random variable can take and the likelihood that each will occur.

Random Variable: It is a variable, that can take different values as a result of the outcomed of a random experiment.

14.8 ANSWERS TO SELF ASSESSMENT EXERCISES

A) 1. a) True; b) False; c) True

d) True; e) False; f) True

2. a) 0.5997; b) 0.0016; c) 0.4018

3. Expected frequencies approximated

97 484 969 969 484 97

B) 2. 0.11

3. Expected frequencies (No. of ACs) approximated

147, 147, 74, 25, 6, 1.

C) 1. a) 0.8599; b) 0.5832; c) 0.3599

d) 0.4618; e) 0.0832

2. a) 0.9938; b) 0.00135; c) 0.1151

d) 0.0606; e) 0.1815.

14.9 TERMINAL QUESTIONS/EXERCISES

1) Distinguish between frequency distribution and probability distribution.

2) Explain the concept of random variable and probability distribution.

3) Define a binomial probability distribution. State the conditions under which the binomial probability model is appropriate by illustrations.

4) Explain the characteristics of a poisson distribution. Give two examples, the distribution of which will conform to the poisson form.

5) What do you mean by continuous probability distribution? How does it differ from binomial distribution?

6) Explain the procedure involved in fitting binomial and poisson distributions.

7) If the average number of defects items in the manufacturing of certain items is 10%, what is the probability of a) 0, b) 2, c) at most 2 items, d) at least two items are found to be defective in a sample of 12 items taken at random

Ans: a) 0.2824, b) 0.2301, c) 0.8891, d) 0.3410.

8) If the probability of a defective bolt is 0.1, find (a) the mean, and (b) the standard deviation of defective bolts in a total of 900. (Ans. (a) 90; (b)81)

9) Harry Onobr is in charge of the electronics section of a large department store. He has noticed that the probability that a customer, who is just browsing will buy something, is 0.3. Suppose that 15 customers browse in the electronics section each hour.

   a) What is the probability that at least one browsing customer will buy something during a specified hour?

   b) What is the probability that at least 4 browsing customers will buy something during a specified hour?

   c) What is the probability that no browsing customer will buy anything during a specified hour?

   d) What is the probability that not more than 4 browsing customers will buy something during a specified hour?

[Ans. (a) .9953 (b) .7031 (c) .0047 (d) .5155]

10) Given a binomial distribution with $n = 28$ trials and $p = .025$, use the Poisson approximation to the binomial to find:

   (a) $P(r \geq 3)$ (b) $P(r < 5)$

   (c) $p(r = q)$

[Ans. (a) .03414 (b) .99922 (c) 0.0000]

11) The average number of customer arrivals per minute at a departmental stores is 2. Find the probability that during one particular minute:

   a) at least one customer will arrive

   b) exactly three customers will arrive

   c) at the most two customers will arrive

[Ans. a) 0.8646; b) 0.1805; (c) 0.6767]

12) A set of 5 fair coins was thrown 80 times, and the number of heads in each throw was recorded and given in the following table. Estimate the probability of the appearance of head in each throw for each coin and calculate the theoretical frequency of each number of heads on the assumption that the binomial law holds:

No. of heads:	0	1	2	3	4	5
Frequency:	6	20	28	12	8	6

Ans: (7, 19, 24, 18, 9, 3)

13) Fit a poisson distribution to the following observed data and calculate the expected frequencies:

Deaths:	0	1	2	3	4
Frequency:	122	60	15	2	1

14) Given that a random variable X, has a binomial distribution with n = 50 trials and p = .25, use the normal approximation to the Binomial to find:

(a) p (x > 10) (b) p (x > 21)

(c) p (x < 18) (d) p (q < x < 14)

[Ans. (a) .7422 (b) .0016 (c) .9484 (d) .4658]

15) Glenn Howell VP of personnel for the Standard Insurance Company has developed a new training programme that is entirely self-paced. New employees work various stages at their own pace, completion occurs when the material is learned. Howell's programme has been especially effective in speeding up the training process, as an employee's salary, during training, is only 67% of that earned upon completion of the programme. In the last several years, average completion of the programme has been in 44 days. With a standard deviation of 12 days.

a) What is the probability that an employee will finish the programme between 33 and 42 days.

b) What is the probability of finishing the programme in fewer than 30 days?

c) What is the probability of finishing the programme in fewer than 25 or more than 60 days?

[Ans. (a) .2537 (b) .1210 (c) .1489]

16) A project yields an average cash flow of Rs. 500 lakhs with a standard deviation of Rs. 60 lakhs. Calculate the following probabilities:

i) Cash-flow will be more than Rs. 560 lakhs.

ii) Cash-flow will be less than Rs. 420 lakhs.

iii) Cash-flow will be between Rs. 460 lakhs and Rs. 540 lakhs.

iv) Cash-flow will be more than Rs. 680 lakhs.

[Ans. (i) .1587 (b) .4972 (iii) .0918 (iv) .0013]

17) If $\log_{10}x$ is normally distributed with mean 4 and variance 4. Find the probability of $1.202 < x < 83180000$, given that

$\log_{10} 1202 = 3.08$, $\log_{10} 8318 = 3.93$.

(Ans. 95%)

Note: These questions/exercises will help you to understand the unit better. Try to write answers for them. But do not submit your answers to the university for assessment. These are for your practice only.

14.10 FURTHER READING

The following books may be used for more indepth study on the topics dealt within this unit.

Levin, R.I. & Rubin, D.S., 1991, *Statistics for Management*, PHI, New Delhi.

Gupta, S.P. 1999, *Elementary Statistical Methods*, Sultan Chand & Sons, New Delhi.

Bhardwaj, R.S. 2001, *Business Statistics*, Excel Books, New Delhi.

Chandan, J.S. *Statistics for Business and Economics*, Vikas Publishing House Pvt. Ltd., New Delhi.