UNIT 3 PROBABILITY DISTRIBUTIONS

Structure **Page No.**

3.1 INTRODUCTION

In the first two units of this block you learnt a few methods for exploring and describing numerical data. The examples you saw there showed that numerical observations on a characteristic might need careful exploration for any useful and correct inference. In general, when you decide to collect data on a characteristic, you have a specific purpose; you want to either verify a hypothesis or want to estimate a quantity for certain purpose. For example, you may want to know if the district of Jalandhar produces more wheat in a year than the district of Faridkot in the state of Punjab. You may have a hypothesis that in general, Jalandhar district produces more wheat per year than Faridkot district. When you have such a specific question or hypothesis, you will decide to collect the appropriate information or data, in this particular instance, on the annual wheat production in the two districts of Punjab. You may find a secondary source, such as a publication from an agency or a department of the government which has already collected this information and has tabulated the data. Of course, you need to be sure about the reliability of the source from where you have obtained the required data. When you actually look at the data on the amount of annual wheat production in the two districts, for say, the past ten years, what do you think you will find? Do you think the amount produced in any district will be the same from year to year? This is very unlikely as there are a very large number of factors that influence the amount of wheat produced and these factors do not remain constant over the 10 years. Also it is very unlikely that you will find Jalandhar district producing more wheat than the district of Faridkot in each of the 10 years. In such a situation how does one compare the two districts in respect of wheat production?

Firstly, it is clear from the above discussion that a realistic model will be to assume that the annual production of wheat in a district is a random variable (result of a random experiment introduced in the previous unit). The ten year observations may be thought of as the outcome of ten repetitions of a random experiment for a district. Secondly, we need to develop measures to compare the two districts with respect to the annual production of wheat when this is thought of as outcome of a random experiment.

In certain situations or for certain problems, you may find that there is no reliable secondary source where data on the appropriate characteristic are available, and that you have to take measurements or make observations, if necessary after conducting a controlled experiment. As in practice, it is never possible to control or eliminate the influence of all the factors, this experiment may also have to be modelled as a random experiment. This need arises, for instance, when a scientist claims to have developed a variety of wheat which yields more per acre than the existing varieties and you want to verify this claim. To summarise, you notice the following:

1) Given a specific purpose, data are collected on an appropriate characteristic.
2) In most real-life problem situations, the value of the characteristic of interest may depend on a large number of factors which are not constant over space or time. It is better, then, to model the situation as a random experiment and the value of the characteristic of interest as an outcome of a random experiment.
3) Methods have to be developed to draw the correct conclusion or inference from the data collected on a characteristic, which is the outcome of a random experiment.
4) If there is no secondary source of data, you may need to either conduct a controlled experiment, or conduct a survey, to obtain data on the appropriate characteristic.

We start this unit by defining the notion of a random variable and its probability distribution in Section 3.2. The notions of discrete and continuous random variables are introduced next, followed by the notions of expectation and variance. You will see that to compare random variables or to draw inferences about them in a practical application, their probability distributions are important. Also important are certain measures of the distribution such as the mean and variance. These are discussed in Sec.3.2. In the next four sections we have discussed the most commonly used four probability distributions in detail.

Here is a list of what you should be able to do by the end of this unit.

Objectives

After reading this unit, you should be able to

- specify when a variable is a random variable and classify it as discrete or continuous.
- find the probability distribution of discrete and continuous random variables and calculate the mean and variance of these distributions and use these measures to make judgements about the real-life situation.
- describe the following distributions
  a) binomial distribution
  b) Poisson distribution
  c) uniform distribution
  d) normal distribution

  and calculate the mean and variance associated with these distributions.
- distinguish between the real-life situations which can be modelled (or studied) by these distributions - when to use binomial or when to use Poisson and likewise other distributions.

3.2 RANDOM VARIABLE

Let us start with the following examples:

1) The number of telephone calls received by an operator in a specified interval of time.
2) The amount of rainfall on a day.

3) The outcome of throwing a die.
4) The scores of students in an examination
5) The number of misprints on a randomly chosen page of a book.
6) The volume of sale of a certain manufactured item in a given year.
7) Time to failure of a machine.

All these examples have one common feature. They describe a characteristic which associates a numerical value or number to each outcomes of a random experiment. Recall from Unit 2 that **random experiments are experiments the outcomes of which cannot be predicted in advance.** We call this characteristic a variable. This characteristic depends on the outcome of the experiment and its value cannot be predicted in advance. This variable is called a random variable. "*When a variable takes different values according to chance, (which can't be predicted before hand), it is called a random variable.*" In order to make this idea clear, consider the following example.

Example 1: Suppose we are interested in the number X of heads obtained in three tosses of a coin. Let us see what is the variable in this experiment.

For that we first find the set of possible outcomes of the experiments i.e., the sample space S. The outcomes are

$$\begin{array}{lll} a_1 = HHH, & a_2 = HHT, & a_3 = HTH, \\ a_4 = THH, & a_5 = TTH, & a_6 = THT, \\ a_7 = HTT, & a_8 = TTT. & \end{array}$$

Then

$$S = \{a_1, a_2, \ldots, a_7, a_8\}$$

Let us denote by $X(a_j)$ the number of heads obtained when the outcome of our experiment is a_j, where $j = 1, 2, \ldots, 8$. You can easily check that

$$X(a_1) = 3, X(a_2) = X(a_3) = X(a_4) = 2, \quad (1)$$
$$X(a_5) = X(a_6) = X(a_7) = 1, X(a_8) = 0 \quad (2)$$

Then do you agree that the X maps elements of the sample space S to the values 0,1,2,3? i.e. X is a function from the space S to the set $N = \{0, 1, 2, 3\}$.

Also note that corresponding to each value, there is always some sample point or a set of sample points. For example, the set of sample points corresponding to the value '0' is the single point a_8, whereas for 1 the set is: $\{a_5, a_6, a_7\}$. That means corresponding to each value of X, there is a subset of the sample space S.
Now you again recall from Unit 2 that an event is a subset of a sample space. Thus we note that that each value of X is associated with an event.

You can, therefore make the following identification of events corresponding to the values associated by X. Denote the event corresponding to '0' as $[X = 0]$, likewise for other values. Then

$$[X = 0] = \{a_8\}, \ [X = 1] = \{a_5, a_6, a_7\}$$

$$[X = 2] = \{a_2, a_3, a_4\}, [X = 3] = \{a_1\}$$

Assuming that all the sample points are equally likely, we assign probabilities of 1/8 to each of the sample points.

From here, using the law of probability you can calculate the probabilities as follows:

$$\begin{aligned} P[X = 0] &= P\{a_8\} = 1/8, \\ P[X = 1] &= P\{a_5, a_6, a_7\} = P\{a_5\} + P\{a_6\} + P\{a_7\} = 3/8, \\ P[X = 2] &= 3/8 \text{and } P[X = 3] = 1/8, \end{aligned}$$

where we read $P[X = j]$ as "probability that X equals j."

Then P has the following Property:

$$P[X = 0] + P[X = 1] + P[X = 2] + P[X = 3] = 1.$$

* * *

Summing up our discussion above, we find that a random variable X is a function defined on the set of outcomes of a random experiment and it takes on a numerical value corresponding to each outcome of the experiment. Sometimes we use the abbreviation r.v. for a random variable. Now these set of possible values is a numerical representation of the original space. The following figure may help you to see this representation.

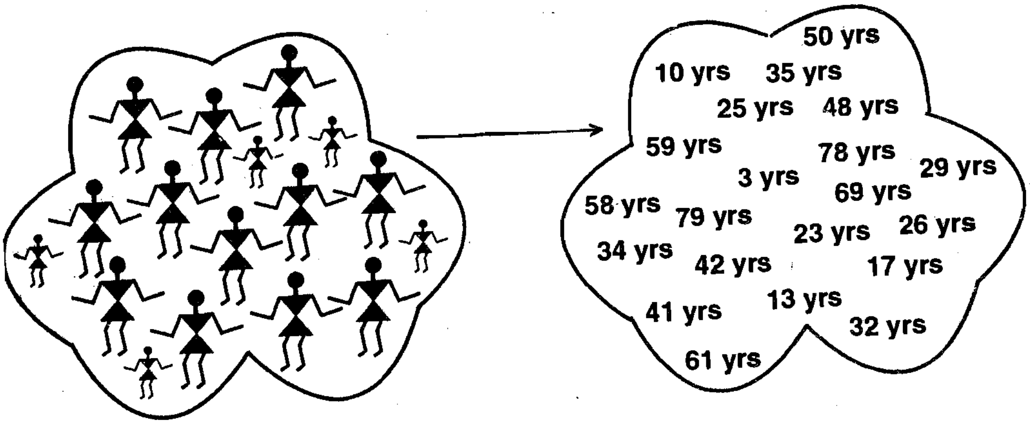


Fig. 1 Representation of a r.v.

Because of this representation, instead of working with an abstract space, we can work with a set of numbers,and this simplifies our problems considerably.

Now let us consider another example.

Example 2: You are sitting in a plane waiting for its take off. The pilot announces a delay until some incoming planes land. Suppose you want to find the following:

i) How long will it be before take off.

ii) How many incoming planes are there.

Let us discuss the random variables for (i) and (ii) of the above example.

We first take (i) In this case we want to find the 'duration of time' before the plane takes off. Note that the variable takes values continuously along a line as given below, say from time duration 'a' to time duration 'b'. No values in between a and b are left out. In other words there is no break in the values assumed by this random variable.

Fig.2

Now let us consider (ii) . Here the random variable is the number of planes. This variable can only take the values 0 or 1 or 2 etc. as shown in Fig. 3. There is no continuity, (see Fig.3) since only non-negative integer values can be assumed.

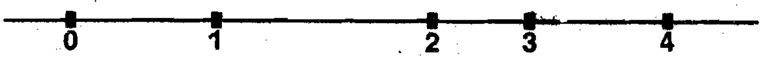


Fig.3

* * *

The above examples show that random variables can be of different types. There are mainly two types of random variables:

1) **discrete random variable**
2) **continuous random variable**

The random variable shown in (ii) of Example 2 is discrete and that of (i) is continuous.

In the next subsection we shall discuss discrete random variables. Before that why don't you try an exercise.

E1) Suppose you take a 50-question multiple-choice exam., guessing every answer, and are interested in the number of correct answers obtained. Then
   a) What is the random variable you will consider for this situation?
   b) What values might this random variable have?
   c) What would $P[X = 40]$ means?

3.2.1 Discrete Random Variables

We first formally define a discrete random variable and familiarise you with some of the properties/ aspects of a discrete r.v.

Definition 1: A random variable **X** is said to be discrete if the. number of values that X can take is finite or countably infinite. These values may be listed as $x_0, x_1, \ldots$, where say, $x_0 < x_1 < \ldots$; these x's are called the jump points. **They need not be equidistant.**

Now let us consider the events associated with the values assigned by it.

Let the events be denoted by $[X = x_i], j = 0, 1, 2, \ldots$. Then, as stated earlier, we can assign probability to these events. We denote $P[X = x_j]$, the probability of the event $[X = x_j]$. For further simplification we denote the probability for each j as $P[X = x_j] = p_j, i = 0, 1, 2, 3, \ldots$.

From the properties of a random variable and by definition of probability it follows that

i) $p_i \geq 0$ (for each i, i.e., each p_i is a non-negative number),

ii) $p_0 + p_1 + \ldots = 1$ (The sum of the probabilities is 1).

Now we have another definition.

Definition 2: Let $p : X \to R$ be defined as

$$p(x) = \begin{cases} p_i & \text{if } x = x_i, \quad i = 0, 1, 2, \ldots \\ 0, & \text{otherwise} \end{cases}$$

Then p is called the **probability mass function**(p.m.f) of the random variable X. **The collection of pairs** $(x_i, p_i), i = 0, 1, 2, \ldots$ **is called the probability distribution of X.**

For example, suppose X is the r.v. denoting the number of heads obtained in three tosses of a coin, then the probability mass function p is the function $p : X \to R$ such that

$$p(0) = 1/8, p(1) = p(2) = 3/8, p(3) = 1/8.$$

Note that $p(x_i) = p_i \geq 0$ for all x_i and

$$\sum p_i = \frac{1}{8} + \frac{3}{8} + \frac{3}{8} + \frac{1}{8} = 1.$$

Therefore, the probability distribution corresponding to this random variable is the set $\{(0, \frac{1}{8}), (1, \frac{3}{8}), (2, \frac{3}{8}), (3, \frac{1}{8})\}$. This is also expressed in a tabular form as follows:

Table 1: Probability distribution of number of heads in three tosses of a coin

The number of heads (Values of the r.v. X)	Probability
0	1/8
1	3/8
2	3/8
3	1/8

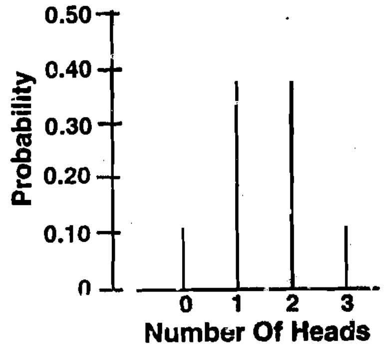


Fig. 4

Suppose you have a random variable X assuming values $x_1, x_2, \ldots$ with probabilities $p_1, p_2, \ldots$, respectively. You may also visualise this as an illustration of a frequency distribution. In fact a frequency distribution tells you how the total probability "one" is distributed over the possible values of the random variable.

Now let us see the graphical representation of this distribution.

Graphically, along the horizontal axis, plot the various possible values x_i of a random variable and on each value erect a vertical line with height proportional to the corresponding probability p_i. (See Fig.4) Recall that in the bar diagram of a frequency distribution, the observed frequencies are graphed along the vertical axis; the total frequency (which is the same as the total number of repetitions of the random experiment) is thus distributed over the possible outcomes.

It is therefore clear that **associated with any random variable, a probability distribution can be defined.** Thus in the study of a random variable it is enough to know the corresponding probability distribution. This is illustrated in the following example.

Example 3: Recall the problem of Sunil, the newspaper boy, which was presented in in Section 1, Unit 2. When Sunil mentioned his dilemma about his 10 irregular customers to his sister Sunita, who is doing a course in statistics at the local college, she advised him, as a first step, to start maintaining a record for each of the 10 customers, showing for each day whether the customer has taken the newspaper from him or not. Following her advise, Sunil had at the end of two months, 61 sets of observations, each set corresponding to a day. Each set of observations was written as a sequence of two numbers, 1's and 0's, 1 at position i showing that customer i has bought the newspaper on that day and a 0 in that position showing that customer i has not bought the newspaper on that day. You will also find that the kth sequence (corresponding to the kth day) is actually an observed sample point of the sample space corresponding to the random experiment performed on day **k** with these 10 customers. Sunil has repeated this random experiment with his 10 customers for 61 days! When Sunil reflected over the mass of observations he has made it suddenly occurred to him that his record may be excessive for solving his problem. Do you also get the same idea as Sunil did? If not, think about it for a while now, and then read ahead.

Sunil reasoned as follows: After all, his daily gain will only depend on how many newspapers he is able to sell on a particular day, irrespective of who among the 10 buys. Therefore, it is enough for his purpose to note down the number of 1's that appear in the kth sequence corresponding to day k, to calculate his gain for that day. So why does he have to maintain a sequence? Just the total number of customers buying on a day should serve the purpose. Do you agree with Sunil? When Sunil showed his diary to Sunita and mentioned to her his new idea, she appreciated his line of thought and told him that the variable he wanted to consider, namely the total number of 1's in a sequence, would be called a random variable by a statistician as the sequences were the observed results of a random experiment. Sunil could forget about the sample space containing the sequences, provided he knew the probability distribution of the random variable chosen by him, namely the number of sales on a day to his irregular customers.

We shall consider the problem of finding the probability distribution of this random variable in the next section.

* * *

In the following sections you will see that there are some standard distributions and most of the real-life problems can be solved by finding the distribution corresponding to a probability model of the given situation. Let us try some exercises.

E2) Which of the random variables given below are discrete? Give reasons for your answer.

1) The daily measurements of snowfall at Shimla
2) The number of industrial accidents in each month.
3) The number of defective goods in a shipment (lot) of goods from a manufacturer.

E3) A box contains twice as many red marbles as green marbles. One marble is drawn at random from the box and replaced; then a second marble is drawn at random from the box. If both marbles are green, you win Rs. 50; if both are red you lose Rs. 10; and if they are of different colours, you will win or lose nothing. Then what is the probability distribution of the amount you win or lose.

Let us now return to the discussion of the three tosses of an unbiased coin. The r.v. X, denoting the number of heads obtained, has the following probability distribution:

$$p_0 = \frac{1}{8},\ p_1 = \frac{3}{8} = p_2,\ p_3 = \frac{1}{8}$$

Suppose you want to calculate the probability of the event $\{X \leq 2\}$.

First note that the event $\{X \leq 2\}$ is the same as the event $\{X = 0\} \cup \{X = 1\} \cup \{X = 2\}]$

Then, since the events are disjoint, we can write,

$$\begin{aligned} P[X \leq 2] &= P[X = 0] + P[X = 1] + P[X = 2] \\ &= \frac{1}{8} + \frac{3}{8} + \frac{3}{8} = \frac{7}{8}. \end{aligned}$$

Therefore there are situations in which you are not only interested in the probability $P[X = x_i] = p_i$, but are also interested in the probability of the $X \leq x_i$, i.e., events of the type $[X \leq x_i]$ which we denote by $P[X \leq x_i]$.

As we already pointed out, the probability distribution of a random variable is analogous to a frequency distribution. It is therefore not surprising that just as we found it useful to calculate the mean of a frequency distribution, we now find it useful to calculate the mean of a probability distribution. Another term often used to denote the mean of a probability distribution is the expected value of the random variable which is defined as follows:

Definition 3: For a discrete random variable X, its **expected value** (or mean), denoted as E(X), is defined as:

$$E(X) = x_0 p_0 + x_1 p_1 + \ldots$$

where x_0, x_1, x_2 are the values assumed by X and p_0, p_1, p_2 are probabilities these values.

Expected value is a fundamental idea in the study of probability distributions. For many years, the concept has been put to considerable practical use in the insurance industry, and in the last twenty years, it has been widely used by many others who must make decisions under conditions of uncertainly.

We shall illustrate this idea with a real-life problem given in Example 4.

Example 4: The Director of a breast cancer screening clinic wants to know how many women will be screened on any one day. If past daily records of the clinic indicate that the number of women screened daily ranges between 100 to 115. The following table illustrates the number of times this level, between 100 to 115, has been reached during the last 100 days.

Table-2 : Number of women screened daily during 100 days

Number screened	Number of days this level was observed	Probability that the random variable will take on this value
100	1	.01
101	2	.02
102	3	.03
103	5	.05
104	6	.06
105	7	.07
106	9	.09
107	10	.10
108	12	.12
109	11	.11
110	9	.09
111	8	.08
112	6	.06
113	5	0.05
114	4	.04
115	2	.02
	Total - 100	Total - 1.00

Let us see how the Director can use this past record to get information about the long-run pattern of daily patient screenings.

We hope you have understood the real-life problem. We will try to model this in statistical language. For that we shall first describe the 'random variable' of interest in this problem.

The random variable here is the number of patients screened, on any given day. Note that this is a discrete random variable, which can assume only non-negative integer values with positive probability. The past record of the clinic indicates that the values of this random variable range between 100 and 115 patients daily. These values are given in the 1st column of the table. The 2nd column contains the number of days each value was observed. For example the value '103' occurred on 5 days. The last column gives the probability/relative frequency for which a particular value is observed. How do you calculate these probabilities? Note that the total number of days is 100 and that the value '100' was observed only one day.

$$\text{Probability that the value} \text{`}100\text{'} \text{is observed} = \frac{1}{100} = .01$$

In this manner you can calculate the other probabilities. (The thumb rule is 'divide each value in the middle column by 100'). This is how the last column is obtained. Notice that the sum of the values in the last column is one. The relative frequencies are taken as probabilities. This is the statistician's empirical approach to assigning probabilities.

Now plot the 'observed values' (i.e. numbers in the 1st column) against the probabilities in a graph. Then you get a graph as given in the next page.

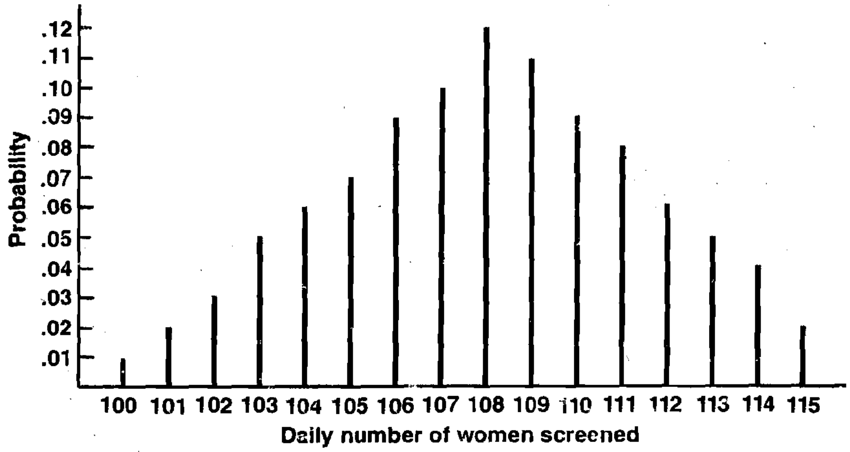


Fig.5: Probability distribution for the discrete random variable 'number screened"

So, this is the graph of the probability distribution of the r.v.

Once we have a frequency distribution, we know that we can find its mean, variance etc. Let us calculate the mean. Here is the mean

$$\text{Mean} = \frac{\sum x_i p_i}{\sum p_i}$$

where x_i's are the observed values in the first column and p_i's are the frequencies

$$\text{Mean} = \frac{100 \times 1 + 101 \times 2 + \ldots + 115 \times 2}{100}$$

This we can rewrite as

$$\begin{aligned} \text{Mean} &= 100 \times \frac{1}{100} + 101 \times \frac{2}{100} + \ldots + 115 \times \frac{2}{100} \\ &= 100 \times .01 + 101 \times .02 + \ldots + 115 \times .02 \\ &= 108.02 \end{aligned}$$

So, essentially what we got is that 'if we multiply each number in the 1st column by the corresponding number in the 3rd column' we will get the mean of the probability distribution. This mean equals the expected value.

The above computation tells us that the expected value of the discrete random variable "number screened" is 108 women. What does this mean? It means that over a long period of time, the number of daily screenings should average about 108.02. This does not mean that 108 women will visit the clinic per day; this only says that over the **long run** and on an average 108 women can visit the clinic. This value is a long run average.

Now based on this expected value (or the mean) the director can decide on what resources/infrastructure is required to get ready for dealing with the expected number of people.

* * *

Why don't you try an exercise now.

E4) A second-hand car dealer has sold as many as five cars in one day, and as few as one. The dealer has tabulated sales records for a large number of days and found that on 5 percent of the days no cars were sold. The dealer took 0.05 as the probability of zero sales in a day, as shown in Table 3 below.Probabilities for sales of 1,2,3,4 and 5 cars were assigned in the same manner (see Table below).

Table 3

Number of cars sold per day	0	1	2	3	4	5
Probability	0.05	0.15	0.35	0.25	0.12	0.18

He wants to find how many cars per day will be sold on the average over a long run. How does he get this number?

3.2.2 Continuous Random Variable

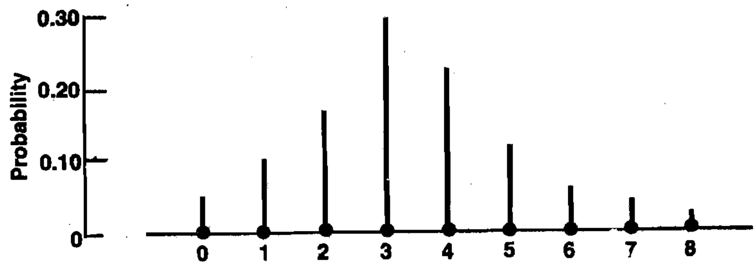
The variables, we discussed in the last section such as 'number of women screened', "number of heads obtained", etc. take on values $0, 1, 2, 3, \ldots$. These are discrete random variables. We saw that the values of a discrete random variable are graphed as separated points and probabilities as lengths of vertical line segments (see Fig.4). We also saw that a probability distribution of such a random variable contains all possible values of the random variable, so the sum of all the probabilities must be 1.

On the contrary, a continuous random variable can take on any value in an interval.

For example, it can take all values x in the interval $0 \leq x \leq 1$ of the form $0, 0.01, 0.0002, \text{etc.} \ldots, 0.98, 0.99, 1.00$. As we have seen in the discrete case, we have to assign probabilities to each value of the variable. How do we do this? Since the possible values of X are uncountable, we cannot really speak of the i^{th} value of X, hence $p(x_i)$ becomes meaningless. So, what we shall do is to replace x_i by any interval of the type of (x_{i-1}, x_i) where $0 \leq x_{i-1} < x_i \leq 1$. and define probability for such intervals. To define this we consider a function defined on this interval $0 \leq x \leq 1$, which assumes non-negative values such that the total area under the graph of this function and over the interval is 1. Then we define the probability of any interval (x_{i-1}, x_i) as the area over this interval and under the graph of f. (See Fig.6(b)) At present you don't have to worry about these functions. Wherever the need arises we will specify these functions.

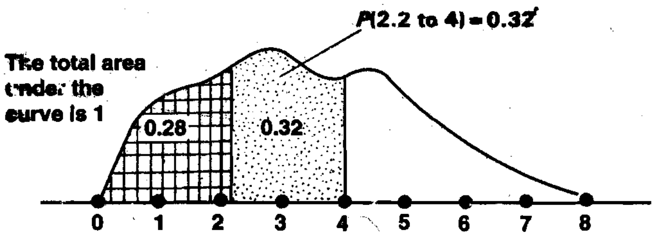
The above discussion may appear a little vague to you. You need not worry about this at this stage. The ideas will be clear to you once we discuss particular cases of continuous random variables and their probability distributions.

Now let us compare the graphs of probability distributions drawn for a discrete and continuous random variable.



Probability distribution for a discrete random variable

Fig.6(a)



Probability distribution for a continous random variable
(the probability of a horizontal interval is the area over the interval)

Fig.6(b)

Note that in Fig.6(b) we did not put vertical lines because a vertical distance tells us what the height of the curve is, and that is not what we want. We want areas under the portion of curves. You will see later that there are tables available for certain continuous random variables where these areas are given for different segments. For examples in Fig. 6(b), the probability of the line segment $2.2 \leq x \leq 4$ is

$$P[2.2 \leq x \leq 4] = 0.32$$

and that of $0 \leq x \leq 2.2$ is

$$P[0 \leq x \leq 2.2] = 0.28$$

It is important to remember that a line segment has no area (zero area), because a line has no width. Thus, the vertical line segment at 4 in the distribution of Figure 6 (b) has an area of zero. This means the probability of a single value 4 is zero. In general, the **probability for an exact (single) value of a continuous random variable is zero.** Consequently, the probability of an interval is the same whether the endpoints are included or not — because the endpoints have probability zero.

Now we formalise the above discussion and make the following definition.

Definition 4: Let X be a continuous random variable which takes on values in the interval (a,b). [i.e. all values between a and b, $a < b$)]. A function f(x) defined on X is called the **probability density function of** X if

(i) f(x) is nonnegative for $a \leq x \leq b$ i.e., $f(x) \geq 0$ for all x lying between a and b.

(ii) the area under the graph and above the interval $(a, b)$is 1.

(iii) For any two real numbers c and d between a and b, the probability that the random variable **adopt** a value between c and d is equal to the area under the graph and above the interval (c, d) i.e.

$$P[c \leq x \leq d] = \text{ area under the graph over the interval}(c, d)$$

Note: Those who are familiar with the mathematical concept of integration can easily see that the above mentioned area is given by the integral of the function i.e.

$$P[c \leq x \leq d] = \int_c^d f(x)dx$$

and (ii) says that

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

We shall now see how we can use the graph of the distribution of a continuous random variable to study real-life problems.

Problem 1: Suppose the Director of a training programme wants to conduct a programme to upgrade the supervisory skills of production line supervisors. Because the programme is self-administered, supervisors require different numbers of hours to complete the programme. Based on a past study of participants, the following distribution (see Fig.7) showing the time spent by candidates is available which shows that average time spent is 500 hours.

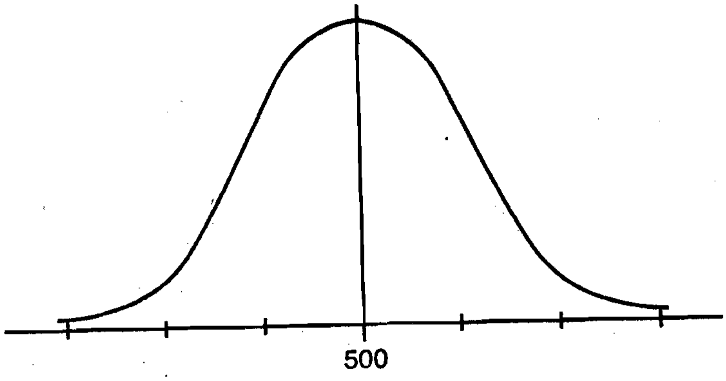


Fig.7

How can she use the graph of distribution, to find the following: What is the chance that a participant selected at random will require

i) more than 500 hours to complete the programme?

ii) less than 500 hours to complete the programme?

Solution:

i) From the figure, we see that half of the area under the curve is located on either side of the mean of 500 hours. Thus, we get that the probability that the random variable will take on value higher than 500 is one half, or 0.5.

ii) A similar argument shows that the chance is 0.5.

———— × ————

Why don't you try some exercise now.

E5) Suppose X is a continuous random variable defined on [2, 4] and f is a function of [2, 4] such that

$$f(x) = \begin{cases} \frac{1}{2} & \text{for } 2 < x < 4 \\ 0, & \text{elsewhere} \end{cases}$$

i) Draw a rough sketch of f(x),

ii) Does f define a probability density function of f? if so why?

E6) Classify the r.v.'s given at the beginning of Sec. 3.1 as discrete or continuous.

In the following sections we shall discuss some standard distribution.

3.3 BINOMIAL DISTRIBUTION

One of the important discrete random variables (or, discrete distributions) is the binomial variable. In this section we shall discuss this random variable and its probability distribution.

Many times we have to deal with experiments where there are only two possible outcomes. For example, when a coin is tossed, either head or tail comes up, seed either generates or fails to generate, a newborn is either a girl or boy.

Let us consider such an experiment. For example, consider the experiment of tossing a fair coin 3 times. This experiment has certain characteristic. First of all, it involves **repetition of three identical experiments (trials).** Each trial has **only two possible outcomes** - a head or tail. We call outcome "head" success and outcome "tail" a failure. All trials are independent of each other. We also know that probability of getting a head in a trial and probability of getting a tail in a trial are both $\frac{1}{2}$.

$$P(H) = P(\text{success}) = \frac{1}{2}$$

and

$$P(T) = P(\text{failure}) = \frac{1}{2}$$

This shows that the **probability of a "success" and of a "failure" do not change from one trial to another.**

If X denotes the total number of heads, obtained in 3 trials, then X is a random variables which takes values from $\{0, 1, 2, 3\}$.

Suppose that p denote the probability of a success (i.e. getting a head) and q denote the probability of failure (i.e. getting a tail).

Then regarding the above experiment, we have observed the following:

1) It involves a repetition of n identical trials (Here $n = 3$).
2) The trials are independent of each other.
3) Each trials has two possible outcomes
3) The probabilities of a "success" (p) and of a "failure" (q) do not change.

If you go back for a moment to Sec. 3.1, you will see that we have already obtained the probability distribution of this in Example 3. Let us look at the probabilities once again.

$$\begin{aligned} P[X = 0] &= P[\text{getting three tails}] \\ &= P[T, T, T] = q \times q \times q \\ &= q^3 = \frac{1}{8} \end{aligned}$$

Similarly,

$$\begin{aligned} P[X = 1] &= P[[TTH], [THT], HTT]] \\ &= P[TTH] + P[THT] + P[HTT] \\ &= q^2p + q^2p + q^2p = 3q^2p \end{aligned}$$

Similarly,

$$\begin{aligned} P[X = 2] &= P[THH] + P[THH] + P[HHT] \\ &= 3p^2q \end{aligned}$$

and

$$P[X = 3] = p^3$$

In fact the probability $P[X = r], r = 0, 1, 2, 3$ gives that if we toss a coin three times, how many ways, or combinations, will yield r heads and $n - r$ tails.

Now you recall from your school mathematics that the number of combinations of n objects taken r at a time is calculated by the formula

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

In the case of tossing of three coins, n=3. Therefore, we rewrite the probabilities as

$$\begin{aligned} P[X = 0] &= C(3, 0)p^0q^{3-0} = q^3 \\ P[X = 1] &= C(3, 1)p^1q^{3-1} = 3pq^2 \\ P[X = 2] &= C(3, 2)p^2q^{3-2} = 3p^2q \\ P[X = 3] &= C(3, 3)p^3q^0 = p^3 \end{aligned}$$

This suggests that the probability $p[X = r] = p_r$ for a given r can be calculated using the formula

$$p_r = C(n, r)p^rq^{n-r} \tag{3}$$

where r = number of successes
n = number of trials made
p = probability of success in a trial
$q = 1 - p$ = probability of failure in a trial.

Why don't you check this formula for n=5, i.e. tossing of a coin 5 times. For example, try this exercise.

E7) In the experiment of tossing a coin 5 times find the probability of getting 3 heads and 2 tails. Verify that this probability is given by the formula given in Equation (3).

Let us now sum up the points we have observed in the example, above.

An experiment consisting of n trials is performed such that

i) **each trial has two possible outcomes, viz., a 'success'(p) and a "failure"(q);**

ii) **the probability of success, p, is the same for any trial;**

iii) **the outcomes of different trials are statistically independent (i.e. the trials are independent).**

These trials are called Bernoulli trials.

The **sample space** of this experiment consists of elements like "**SSFSSF.......**" of length n of 'S's and 'F's where S stands for the success and F stands for the failure.

Let X represent the number of successes (in any order whatsoever) in the set of n trials. Then **X is a discrete random variable taking integral values** $0, 1, \ldots, n$. **The probability of the event** $P[X = r]$ **is given by**

Binomial Distribution

$$P[X = r] = p_r = C(n, r)p^r q^{n-r} \tag{4}$$

where r = exact number of successes
n = number of trials made
p = probability of success on a trial
$q = 1 - p$ = probability of failure on a trial and
$C(n, r) = \dfrac{n!}{(n-r)!r!}$ (The earlier example illustrates how we got this formula for p_r).

James Bernoulli was a seventeenth century swiss mathematician who performed some of the early work on binomial distribution.

Such a random variable X is called a **binomial random variable** and its probability distribution is called **binomial distribution** and is given by Eqn.(2).

Problem 2: A sales representative, calls on four potential clients. The probability that she will obtain an order from each of them is $\frac{1}{2}$ and whether or not she obtains an order from one of them is statistically independent of whether or not she obtains an order from any of the others. What is the probability distribution of the number of orders she will receive?

Solution: We note that there are two mutually exclusive events (obtaining an order or no order) each time she makes a call and the probability of an order $1/2$ each time. Also the outcomes of the calls are statistically independent. Therefore this is a situation where there are four Bernoulli trials and where the probability of a success (an order) equals $1/2$. Substituting $n = 4$ and $p = \dfrac{1}{2}$ in Eqn.(4), we get that

$$\begin{aligned}
P(0) &= \frac{4!}{0!4!}\left(\frac{1}{2}\right)^0\left(\frac{1}{2}\right)^4 = \frac{1}{16},\\
P(1) &= \frac{4!}{1!3!}\left(\frac{1}{2}\right)^1\left(\frac{1}{2}\right)^3 = \frac{1}{4},\\
P(2) &= \frac{4!}{2!2!}\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^2 = \frac{3}{8},\\
P(3) &= \frac{4!}{3!1!}\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right)^1 = \frac{1}{4},\\
P(4) &= \frac{4!}{4!0!}\left(\frac{1}{2}\right)^4\left(\frac{1}{2}\right)^0 = \frac{1}{16},
\end{aligned}$$

Thus, the probability of no orders is $1/16$, of one order is $1/4$, of two order is $3/8$, of three orders is $1/4$, and of four orders is $1/16$.

———— × ————

Problem 3: It has been claimed that in 60% of all solar heat installations, the utility bill is reduced by at least one-third. Accordingly, what are the probabilities that the utility bill will be reduced by one-third in

i) four of five installations?

ii) at least four of the five installations?

Solution Here the random variable follows binomial distribution with $p = 0.6, r = 4$ and n=5.

To find (i), we have to calculate $P[X = 4]$, which is given by

$$\begin{aligned} P[X = 4] &= C(5,4)(0.6)^4(0.4) \\ &= 0.259 \end{aligned}$$

Now to find (ii), we have to find the probability that X is at least 4. This probability is the sum of the probabilities that X = 4 and X = 5 because 'at least 4 means 4 or more'. Thus we have to find $p[X = 4] + P[X = 5]$.

$$\begin{aligned} P[X = 5] &= (5,5)(0.6)^5 \\ &= 0.078 \end{aligned}$$

$\therefore$ the required probability = 0.259 + 0.078 = 0.337.

———— × ————

Binomial distribution is very applicable in situations where we have to decide whether to accept a lot of goods (items) coming out of of a manufacturing process. This decision is based on how many defective items are in the lot. Companies (or firms) will generally return the entire items if there is evidence that more than certain items is defective. To make such decision, let us see how we can make use of the binomial distribution.

An item coming out of a manufacturing process can either be defective or non defective. Consider a lot of **N** items produced by the manufacturing process. Let **m** of these be defective. Suppose a quality control inspector draws a random sample of **n** items from the lot, one by one, with replacement (i.e. an item drawn is put back in the lot, after noting down whether it is defective or non defective, before the next item is drawn at random). Let X be the number of defective items drawn by the inspector. Note that there are **n** trials and in each trial the probability that a defective item is picked remains the same, namely $\frac{m}{N}$, as the drawing is done with replacement and at random. Also note that the trials are independent. Therefore, the random variable X defined above is distributed as a Binomial (n,p) where $p = \frac{m}{N}$.

By now you must have got some idea for recognising those situations where we can apply binomial formula. If we can apply binomial distribution to study a situation, then we say that the situation can be modelled by binomial distribution.

Here are some exercises for you.

E8) A farmer buys a quantity of cabbage seeds from a company that claims that approximately 80 % of the seeds will germinate if planted properly. If four seeds are planted, what is the probability that exactly two will germinate?

E9) Consider again the data collected by Sunil, the newspaper boy. When Sunita, the statistics student, saw the data, she started wondering if the number of customers from among his ten irregular customers, who actually buy from him on a given day, will follow a binomial distribution? What do you think? Under what conditions will this random variable follows a binomial distribution.

E10) Sunita was still glancing through Sunil's diary wondering to herself if she could think of the 10 customers as 'ten identical coins', when she noticed something significant. She noticed that a lot more of the sequences had a 1 in the third position than in the 8th position. Sunil remembered that customer 3 was the management trainee whom he called 'Alka Didi'. She was from a neighbouring town and was undergoing training in a software company. She was interested in news about software companies, science and environmental issues. She would often buy from Sunil but not always. Customer 8, Sunil told his sister, was a mysterious young man by name Kapil, who was rumoured to be working for a detective agency. One could rarely find him in the morning hours and if he was at home in the morning hours, he would certainly buy from Sunil.

Given the situation above, do you still think that the number of sales on a day can be modelled as a binomially distributed random variable? Give reasons for your answer.

Once we have the probability distribution, we naturally ask what is the 'expected value'. We shall see that now.

Expected Value of a Binomial Variable

We have already seen in Sec.3.2 that for a discrete random variable X, the 'Expected Value' E(X) is

$$E(X) = x_0p_0 + x_1p_1 + \dots$$

where $x_0, x_2, \dots$ are the values assumed by X and $p_0, p_1, \dots$ are the probabilities associated with these values i.e.

$$P[X = x_i] = p_i, \qquad i = 0, 1, 2, \dots$$

If X is a binomial r.v., taking values $0, 1, \dots, n$, then, we know that

$$P[X = i] = C(n, i)p^i(1 - p)^{n-i}$$

$$\therefore E(X) = n\times C(n, n)p^n(1-p)^0 + (n-1)C(n, n-1)p^{n-1}(1-p)^1 + \dots + 0\times C(n, 0)p^0(1-p)^n$$

We rewrite this expression in the sum notation $\sum$ (called sigma) as

$$\begin{aligned} E(X) &= \sum_{j=0}^{n} j\, C(n, j)p^j(1 - p)^{n-j} \\ &= np \sum_{j=1}^{n-1} C(n - 1, j - 1)\, p^j(1 - p)^{n-j} \end{aligned}$$

Those who are familiar with binomial expansion can recognise that the second expression on the R.H.S. is $1 - p + p)^{n-1}$. Therefore we have

$$\begin{aligned} E(X) &= np\,[1 - p + p]^{n-1} \\ &= np. \end{aligned}$$

This means that the expected number of successes is np.

Let us do a problem.

Problem 4: An oil exploration firm plans to drill six holes. It is believed that the probability that each hole will yield oil is 0.1. Since the holes are in quite different locations, the outcome of drilling one hole is statistically independent of that of drilling any of the other holes.

(a) If the firm will be able to stay in business only if two or more holes produce oil, what is the probability of its staying in business?

(b) Give the expected value of the number of holes that result in oil.

Solution: (a) If the firm can stay in business only if two or more holes produce oil, it follows that the probability that it will stay in business equals 1 minus the probability that the number of holes resulting in oil is 0 or 1. Each hole drilled can be viewed as a Bernoulli trial where the probability of success is .1. Thus, the probability that the number of successes is 0 or 1 equals:

$$\begin{aligned} P(0 \text{ or} 1) &= P(0) + P(1) = \frac{6!}{0!6!}(.9^6) + \frac{0!6!}{1!5!}(.1)(.9^5) \\ &= .531 + .354 = .885. \end{aligned}$$

Consequently, the probability that the firm will be able to stay in business is $1 - .885 = .115$.

(b) The expected value of the number of holes yielding oil is $6 \times 0.1 = 0.6$, since $n = 6$ and $p = .1$.

———— × ————

A problem with the binomial distribution is that if the number trials 'n' is very large and probability 'p' is very small, computation of $P[X = r]$ is cumbersome.

The distribution which we introduce in the next section may be useful in such a situation.

3.4 POISSON DISTRIBUTION

Poisson
A nineteenth century swiss mathematician.

In this section we introduce you to another discrete distribution called 'Poisson distribution'. We will familiarise you with different situations where we can apply this distribution. Let us try to understand this distribution through an example.

Suppose it is the the busy Friday noon hour at a bank, and we are interested in the number of customers who might arrive during that hour, or during a 5-minute or a 10-minute interval in that hour;.

In statistical terms, we want to find the probabilities for the number of arrivals in a time interval.

As in the case of binomial , here also we make some assumptions.

1) The average arrival rate at any unit time remains the same over the entire noon hour.

2) The number of arrivals in a time interval does not depend on what happened in previous time intervals.

3) It is extremely unlikely that there will be more than one arrival in a very short interval of time. That means that it is impossible for more than one customer to get through the revolving entrance door in a fraction of a second.

Under these assumption we find the required probability. For this we make use of the following formula known as **Poisson formula**, given by

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

where λ is the Greek letter lambda which denotes the average arrival rate per unit of time and **t** is the number of units of time is the number of arrivals in t units of time

Also we know that $\lambda = 72$ arrivals per hour is a constant for this situation. Since in the question λ is given in 'hour', to standardise the unit, we have to find 't' in hour.

$$\begin{aligned} \text{i.e. } 60 \text{ minutes} &= 1\text{hour} \\ \therefore 3\text{minutes} &= \frac{1}{20}\text{hours} \\ \therefore t &= \frac{1}{20}\text{hours} \end{aligned}$$

Then

$$p(4) = \frac{e^{-72\times\frac{1}{20}}\left(72\times\frac{1}{20}\right)^4}{4!} = \frac{e^{-3.6}(3.6)^4}{4!}$$

To find P(4), we use the Table 2, given in the Appendix. This table shows p(x) for selected values of λ.

From the table, we get

$$p(4) = 0.191$$

What does this value 0.191 specify? This tells us that if the arrivals are arrivals of customers at a bank, there is 19.1% chance that exactly four customers will arrive in the next 3 minutes.

If we vary the values of x and t, we can get different probabilities. This gives the probability distribution which is called Poisson probability distribution.

In the above discussion we saw that the Poisson formula is applicable only if certain conditions are specified. We re-state the formula now.

Poisson Formula

Poisson Distribution

The Poisson Formula is given by

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

where λ is used to compute probabilities for the number of occurances in an interval of time, if the occurrences have the following characteristic.

1) **the average occurrence rate per unit of time is constant**
2) **occurrence in an interval is independent of what happened previously.**
3) **It rarely happens that there will be more than one occurrence in a very short time interval**

A distribution having probabilities given by Poisson Formula is called Poisson distribution.

Now let us see some situations where we can apply Poisson distribution . Here is an example.

Problem 5: Calls at a telephone switch board occur at an average rate of six calls per 10 minutes. Suppose the operator leaves for a 5-minute coffee break, what is the probability that exactly two calls come in (and so go unanswered)while the operator is away?

Solution : Here you can check that the conditions 1,2 and 3 of the Poisson formula are satisfied in this case. Therefore we can use the formula. Now that here $\lambda = \frac{6}{10}$, In this case $t = 5$ so that $\lambda t = 3$. Hence the required probability P(2) is given by

$$p(2) = \frac{e^3 3^2}{2!} = 3e^3 = 0.2240$$

That means there is 0.2240 chance that two calls go unanswered.

———— × ————

Here are some exercises for you.

E11) If a bank receives on an average $\lambda = 6$ bad checks per day, what is the probability that it will receive 4 bad checks on any given day.

E12) A hospital has 20 kidney dialysis machines and that the chance of any one of them malfunctioning during any day is .02. We want to find the probability that exactly 3 machines will be out of service on the same day. Then,

i) can we use the binomial formula to find this probability? If yes, calculate the probability.

ii) can we use the Poisson formula to find this? If yes calculate the probability.

In the above exercise we have seen that the difference between the two calculations is very small.

The Poisson formula can be used to approximate the binomial probability of r successes in n binomial trials in the situations where n is large and probability of success 'p' is small.

For instance, suppose we are interested in number of road accidents in a metropolitan city or daily number of machine breakdown in a work shop etc., during a specified interval of time. Each of these subintervals is so small that at best one and no more occurrence happens within it. Thus we may look upon each subinterval as a trial. Each trial leads to a "success" if the occurrence happens during that subinterval and to a "failure" if the occurrence does not happen.Assume that the occurrences are independent of each other. Hence, the total number of occurrences can be constructed to be distributed binomially, the total number of trials being equal to the number of subintervals which we have ensured to be large; also, the length for each subinterval being small, the probability of an arrival (success) is likely to be small.

Thus we have seen that there are situations where both binomial and Poisson are applied. The rule of thump followed by most statisticians is that if $n \geq 20$ and $p \leq 0.05$, then Poisson formula can be used to calculate binomial probability.

It is clear that the Poisson calculation is simpler than the binomial calculation. An advantage of the Poisson distribution. , if it is applicable, is that it has only one parameter, λ, whereas the binomial distribution has two parameters, n and p; consequently, Poisson probabilities can be tabulated more compactly than binomial probabilities. For example, the Poisson probability P(3) is the same for $n = 200$, $p = 0.01$ as it is for $n = 100$, $p = 0.02$, and for any other pair of n and p values whose product is $\lambda = np = 2$.

By now you must have got a fairly good idea where the Poisson formula can be used.

In all the situation we have considered so far, we have calculated the probability over an interval of time. But there are situations where we need to calculate probability over a region (or space) or something else as our physical reference. In the following example we given such a situation and illustrated how to use Poisson distribution to calculate the probability.

Example 5: During second world war,a v-2 rocket hit in South London. Later a study was conducted on'what are regions not affected by the rocket hit.Let us see how they used Poisson distribution for this study.

They took λ as the average number of hits per unit area (Note that earlier in the formula λ was average rate per unit time).Instead of the variable't' they replace the variable 'v', and x denotes the number of hits per unit area. Then they assumed that all the conditions to satisfy the Poisson formula holds in this case.With all this assumptions, they calculated the probability using the formula

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

According the problem stated, they have to calculate the probability of 'no hit' per unit area. That is, the $x = 0$ and $v = 1$, so that $\lambda v = \lambda$. Now, to calculate λ, what they did was, they divided the area into 576 areas of equal size (the number 576 is chosen based on some other study and they found that they were 537 hits).

$$\therefore \text{ the average number of hits per unit area } \lambda = \frac{537}{576} = 0.9323$$

Then the required Probability is

$$P[x = 0] = e^{-0.9323} = 0.3936$$

This means that if we take one region, then the probability that the region is not hit by the rocket is 0.3936. Hence, out of 576 regions, the number of regions not hit by the rocket is given by

$$576 \times 0.3936 = 226$$

* * *

Now, the actual number got from the record was that there are 229regions not hit by the rocket. This number is quite close to 226. This shows that the values got using Poisson formula are very close to the actual values.

Thus we saw that the Poisson distribution is very effective in studying various real-life problems where the occurrence is very rare.

One of the main disadvantages of this distribution is that it is applicable only in situation where the outcomes are independent i.e. each outcome is independent of what happened previously.

In the next section we shall discuss another standard distribution.

3.5 UNIFORM DISTRIBUTION

The uniform distribution is the simplest of a few well-known continuous distributions which occur often.

As we have seen in Sec.3.2 in the continuous case we are interested in behaviour of the variable in the subintervals of the sample space, rather than at single points. If for example, the sample space is what we call the unit interval [0, 1], and we set the random variable X as a value selected from this interval, then we are no longer interested in the outcome of the kind $\{X = a\}$, but rather outcome of events of the kind $\{a < x < b\}$ i.e. values lying between the two numbers a and b, where $0 \leq a \leq b \leq 1$.

Suppose X is a random variable such that if we take any subinterval of the sample space, then the probability of this interval is the same as the probability of any other subinterval of the same length. The distribution corresponding to this r;v. is called a uniform distribution. As the name suggests the probability is uniform along subintervals.

Let us see some examples of such sample spaces.

Example 6: A train is likely to arrive at a station at any time between 6.10 p.m. and 6.40 p.m. The time the train reaches, measured in minutes, after 6 p.m. is a random variable X. Here X can take any value between 10 and 40 minutes. Therefore the sample space is the interval (10, 40). It is reasonable to assume that the likelihood for X taking any value between 10 and 40 is equal. So if we take subintervals of equal lengths, then the probability will be the same. The distribution corresponding to this r.v. is uniform over the interval (10, 40).

* * *

Example 7: An office fire drill is scheduled for a particular day, and the fire alarm is likely to ring at any time between 9 a.m. and 5 p.m. The time the fire alarm starts, measured in minutes, after 9 a.m. is therefore a random variable which takes any value between 0 and 480 ($= 8$ hours $= 8 \times 60 = 480$minutes) equally. The distribution corresponding to this r.v. is uniform.

* * *

Now, why don't you look for such sample spaces on your own. Try this exercise now.

E13) Verify whether the following situations can be described by uniform distribution or not?

a) The average life span of a life bulb produced by a manufacturing company.

b) The number of defective items produced by an assembly process.

Next we will see how we can define (calculate) the probabilities for this distribution. As we have seen in Sec. 3.2 , in the case of a continuous distribution, the probabilities are calculated using a function called 'probability density function' (p.d.f.). The p.d.f. for uniform distribution is given as follows.

Definition 5: The **pd.f. of a random variable X** which is distributed uniformly in the interval [a,b], where a < b is given by

$$f(x) = \begin{cases} \dfrac{1}{b-a}, & \text{if } a \leq x \leq b \\ 0, & \text{otherwise} \end{cases}$$

We can easily draw the graph of this distribution. It is given in Fig.8.

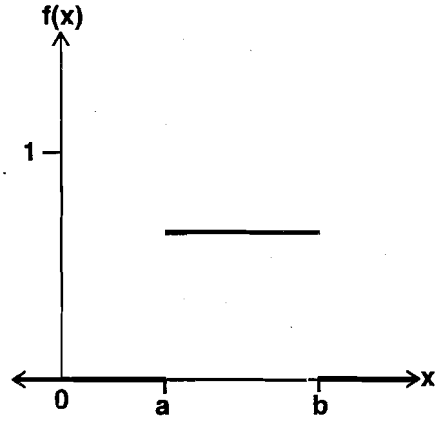


Fig. 8: Graph of P.d.f. of a uniform distribution

Now let us see how we calculate different probabilities for this distribution. As stated in Sec.3.2, for a continuous r.v., we calculate the probability of an interval rather than a point. For example, what will be P[c < X < d] where a < c < d < b? We have seen that it is given by the area above this interval and under the graph. The area is shown in Fig.9.

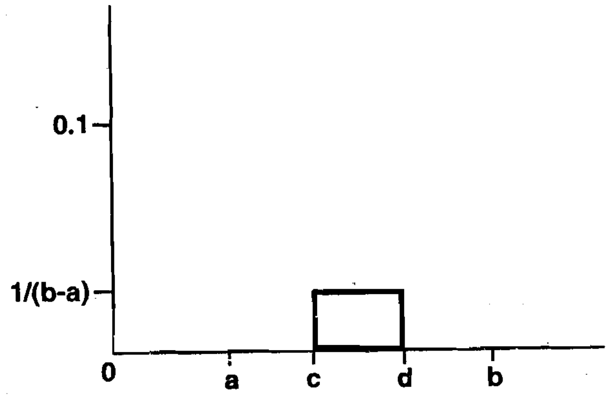


Fig.9 P [c < X < d] = **Area of the rectangle shown**

So, essentially it is the area of the rectangle with length d − c and height $= \dfrac{1}{b-a}$ i.e.

$$P[c < X < d] = (d - c) \times \frac{1}{b-a}$$

For example, if we take the situation in Example 4, let us find the probability that the alarm sounds between 1 p.m. and 2 p.m. Here the pdf,

$$f(x) = \frac{1}{480}, 0 \le x \le 480$$

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

To find the required probability, you have to find time elapsed in minutes between 9 a.m. and 1 p.m. and between 9 a.m. and 2 p.m.

For 1 p.m. this is $4 \times 60 = 240$ minutes.

Similarly, for 2 p.m., it is $5 \times 60 = 300$ minutes.

Therefore you have to calculate the probability $P[240 < X < 300]$. This is given by the shaded area given below in Fig.10.

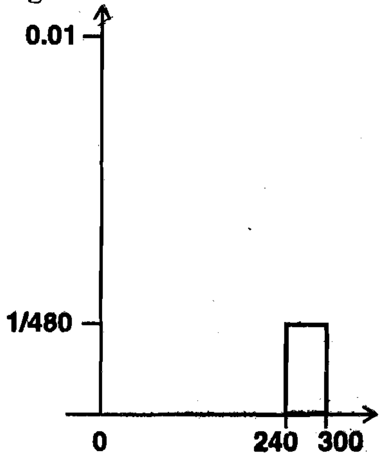


Fig.10

This area is the rectangle with base $60(= 300 - 240)$ and height $\frac{1}{480}$.

$$P[240 < X < 300] = \frac{1}{8} = 0.125$$

That is there is 12.5% chance that the alarm sounds between 1 p.m. and 5 p.m. [Some of you may think that this fact was rather obvious from the statement of the problem itself. But we have given this situation as an illustrative example. There are situations which are complicated, where we can easily calculate the probability using this distribution.]

Next we state below the **expected value of this distribution.**

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

You can try this exercise now.

E14) Suppose that the weight of sugar obtained by processing a tank of sugar cane juice is uniformly distributed with a mean of 10 kg. and range of 1.8 kg. Then

i) What are the largest and smallest weights of sugar obtained from a tank of sugar can juice?

ii) What is the probability that a tank of juice will yield sugar weighing between 9 kg. and 10.5 kg.?

E15) A train is due to arrive at 5.30 p.m. but in practise is equally likely to arrive at any time between 2 minutes early and 30 minutes late. Let the time of arrival (expressed as minutes from due time) be X. Sketch the pdf f(x) of the r.v. X and shade the areas given bellow

1) The probability that the train is less than 10 minutes late.

2) The probability that the train is late, but less than 16 minutes late.

Next we shall discuss another continuous distribution which is widely used in statistical problems.

3.6 NORMAL DISTRIBUTION

Normal distribution' is a class of distribution which can be used to study the probability distribution occurring frequently in real-life situations, of biology, manufacturing machines, psychology etc.

A particular form of this distribution was found by seventeenth - eighteenth century mathematicians Abraham De Moivre and Pierre Laplace, while they were working on various problems in probability. They found that the distribution corresponding to certain random variables had got special property that when graphed, a bell-shaped curve is obtained and came to be called the normal pattern. The graph of the pattern became known as normal curve. Later this class of distribution was studied extensively by another mathematician Karl Friedrich Gauss and therefore this became known as 'Gaussian distribution'.

We shall now state the distribution.

Definition 6: We say that a random variable X is normally distributed with parameters μ and σ if the probability density function f(x) of X is given by,

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2} \text{where} - \infty < x < \infty$$

where μ is a real number lying between $-\infty$ and ∞ and σ is a real number lying between 0 and ∞.

The function f(x) may look rather formidable to you at first sight. At this stage we just ask you to notice that it involves two parameters, σ and μ. Corresponding to each pair (μ, σ), we get a distribution. Therefore there is a whole family of distributions, each one specified by a particular pair of values for σ and μ.

The most important characteristic of this distribution is that the graph of pdf, f(x) for a particular value of μ and σ is bell-shaped as shown in Fig.11.

The probability density function, pdf is also symmetrical about the mean μ. The word symmetrical means that the two halves of the curve are mirror images (see Fig.11). In Fig.11 you note that if we place a mirror on the dashed vertical line (which occurs at 75 in Fig.11) then the mirror image of the portion on the left is the same as the portion on the right side.

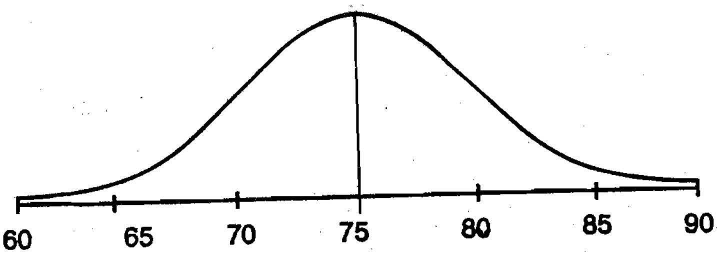


Fig.11

Both μ and σ have a 'nice' interpretation. We have already said that the pdf is symmetric about μ, so it is no surprise that μ **is the mean of the distribution**. The other constant, σ^2 dictates how spread out and flat the 'bell-shape' is and in fact σ^2 **is the variance of the normal distribution**.

As an illustration, the following figure shows that the normal pdfs for μ and σ are given as follows:

A $\mu = 10, \sigma = 1$

B $\mu = 10, \sigma = 2$

C $\mu = 10, \sigma = 3$

D $\mu = 15, \sigma = 1$

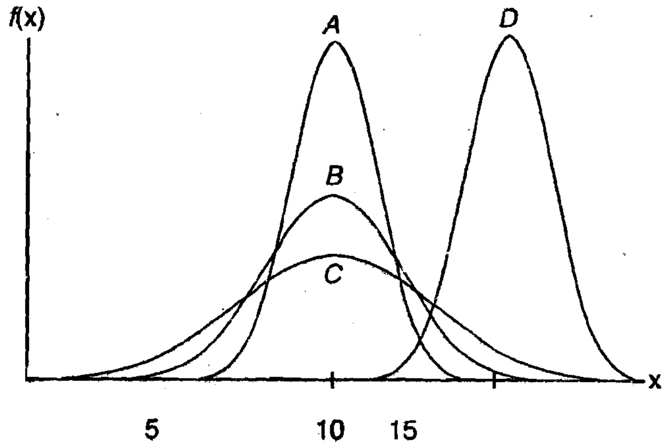


Fig.12

Pdfs A, B and C all have the mean 10 and so they are all centred at $x = 10$. Of these three curves, C has the largest variance and so is the most 'spread out'. Curve B has a smaller variance and so is less spread out, and curve A has the smallest variance and so is the most 'squeezed in'. Curves A and D have the same variance and so they have exactly the same shape, but they have different means so they are centred at $x = 10$ and $x = 15$ respectively.

Some notation

As a normal distribution is entirely specified by its parameters μ and σ we denote such distribution by N (μ, σ^2) where μ is the mean and σ^2 is the variance. So, for instance, the curve shown in (A) above is the pdf N(10, 1) the curve in (B) is the pdf N(10, 4) and so on.

The standard normal distribution

The normal distribution with mean $\mu = 0$ and variance $\sigma^2 = 1$, is called the standard normal distribution. Z is the notation usually used for a random variable which has this distribution. A graph of the standard normal pdf, p(z) is shown in Fig.13.

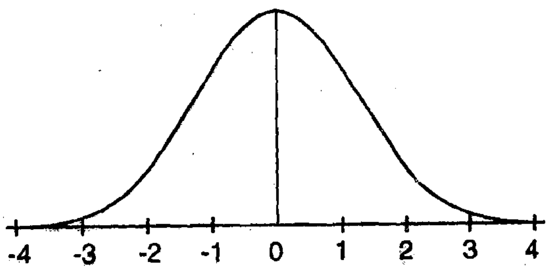


Fig.13

Notice that most of the area under the standard normal curve lies between -3 and $+3$.

Calculating Probabilities

The normal distribution is continuous and so the probability that the random variable X lies between the interval (a, b) is are calculated by obtaining the area under the pdf curve between a and b.

For example, suppose an individual's IQ score, X has a normal distribution with mean $\mu = 100$ and standard deviation $\sigma = 15$. Fig. 14 shows the areas under the pdf which correspond to $P(X < 85)$ and $P(115 < X < 120)$.

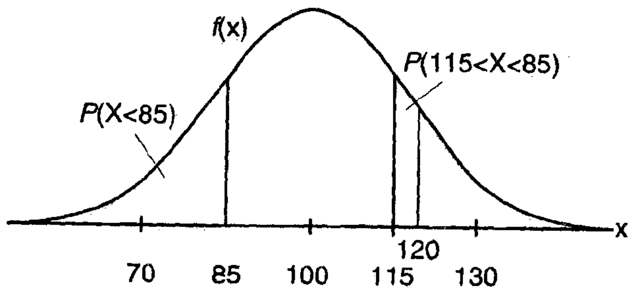


Fig.14

Unfortunately there are no 'nice' formulae for calculating such areas. But there are tables available from which we can find out the area. Statistical software are also available by which we can calculate the area.

Because the number of possible values for μ and σ is unlimited, the number of different normal distributions is unlimited. However, probabilities for every normal distribution can be obtained from a table of probability for standard normal distribution.

We shall first discuss how to use the table for calculating probabilities for a standard normal distribution. Then we shall discuss how to use this to find the probability for any normal distribution.

Using tables to calculate normal probability

We denote by $F(a) = P[Z \leq a]$, the probability that the standard normal variable Z takes values less than or equal to 'a'. The values of F for different values of a are calculated and listed in a table. One such table is given in Sec. 3.9 Appendix.

Note that the entries in the table are the values of z for z=0.00, 0.01, 0.02, 3.49. To find the probability that a random variable having the standard normal distribution will take on a value between a and b, we use the equation

$$P[a < z < b] = F(b) - F(a),$$

and, if either a or b is negative, we also make use of the identity

$$F(-z) = 1 - F(z).$$

In the following example we illustrate how we use the table to calculate different probabilities.

Example 8: Suppose want to find the following probability

i) $P[0.87 < Z < 1.28]$

ii) $P[-0.34 < Z < 0.62]$

iii) $P[Z \geq 0.85]$

iv) $P[Z \geq -0.65]$

We proceed as follows.

i) We know that

$$P[0.87 < Z < 1.28] = F(1.28) - F(0.87)$$

To find F(1.28), we find the row where Z=1.2, then move across that row to the column headed 0.08 and found the entry 0.8997. Similarly we can find that F(0.87) = 0.8078. Then the required Probability is

$$\begin{aligned} P[0.87 < Z < 1.28] &= 0.8997 - 0.8078 \\ &= 0.0919 \end{aligned}$$

ii) Similarly,

$$P[-0.34 < Z < 0.62] = F(0.62) - F(-0.34)$$
$$= F(0.62) - [1 - F(0.34)]$$

by the identify $F(z) = 1 - F(z)$.

$$= 0.7324 - (1 - 0.6331)$$
$$= 0.3655.$$

iii) From the previous unit (Unit 2), you have already learnt that

$$P[Z \geq a] = 1 - P[Z \leq a]$$

Hence we have

$$P[Z > 0.85] = 1 - P[Z \leq 0.85]$$
$$= 1 - F(0.85)$$
$$= 0.1977.$$

v) As in (iii), we can write

$$P[Z > 0.65 = 1 - P[Z \leq -.0.65)]$$
$$= 1 - F(-0.65)$$
$$= 1 - F(1 - F(0.65))$$
$$= F(0.65)$$
$$= 0.7422.$$

* * *

In the following exercise we ask you to find certain probabilities using the normal distribution table.

E16) If a random variable has the standard normal distribution, find the probability that it will take on a value

i) less than 1.50

ii) less than -1.20

iii) greater than - 1.75

E17) A filling machine is set to pour 952 ml (millimetres) of oil into bottles. The amounts of fill are normally distributed with a mean of 952 ml. and a standard deviation of 4 ml. Use the standard normal table to find the probability that a bottle contains oil between 952 and 956 ml.

Next we shall see that how to use the standard normal probability table to calculate probability of any normal distribution.

Standardising

Any normal random variable X, which has mean μ and variance σ^2 can be standardised as follows.

Take the variable X, and

i) subtract its mean, μ and then

ii) divide by its standard deviation, σ.

We will call the result, Z, so

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

For example, suppose, as earlier, that X is an individual's IQ score and that it has a normal distribution with mean $\mu = 100$ and standard deviation $\sigma = 15$. To standardise an individuals IQ score, X, we subtract $\mu = 100$ and divide the result by $\sigma = 15$ to give,

$$Z = \frac{X - 100}{15}$$

In this way every value of X, has a corresponding value of Z. For instance, when $X = 130, Z = \frac{130-100}{15} = 2$ and when $X = 90, Z = \frac{90-100}{15} = -0.67$.

The distribution of standardised normal random variables

The reason for standardising a normal random variable in this way is that a standardised normal random variable

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

has a standard normal distribution.

That is, Z is $N(0, 1)$. So if we take any normal random variable, subtract its mean and then divide by its standard deviation, the resulting random variable will have a standard normal distribution. We are going to use this fact to calculate (non-standard) normal probabilities.

Calculating probabilities

With reference to the problem of IQ score, suppose we want to find the probability that an individual's IQ score is less than 85, i.e. $P[X < 85]$. The corresponding area under the pdf $N(100, 15^2)$ is shown in Fig.15.

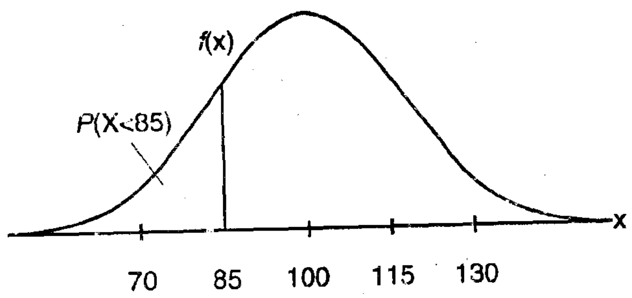


Fig.15

We cannot use normal tables directly because these give $N(0, 1)$ probabilities. Instead, we will convert the statement $X < 85$ into an equivalent statement which involves the standardised score, $Z = \frac{X-100}{15}$ because we know it has a standard normal distribution.

We start with X =85. To turn X into Z we must standardise the X, but to ensure that we preserve the meaning of the statement we must treat the other side of the inequality in exactly the same way. (Otherwise we will end up calculating the probability of another statement, not $X < 85$). 'Standardising' both sides gives, $\frac{X-100}{15} < \frac{85-100}{15}$.

The left hand side is now a standard normal random variable and so we can call it Z, and we have

$$Z < \frac{85 - 100}{15}$$

which is

$$Z < -1.$$

So we have established that the statement we started with, $X < 85$ is equivalent to $Z < -1$. This means that whenever an IQ score, X, is less than 85 the corresponding standardised score, Z will be less than -1 and so the probability we are seeking, $P[X < 85]$ is the same as $P[Z < -1]$.

$P[Z < -1]$. is just a standard normal probability and so we can look it up in Table 1 in the usual way, which gives 0.1587. We get that $P[X < 85] = 0.1587$.

This process of rewriting a probability statement about X, in terms of Z, is not difficult if you are systematically writing down what you are doing at each stage. We would lay out the working we have just done for $P[X < 85]$ as follows.

X has a normal distribution with mean 100 and standard deviation 15. Let us find the probability that X is less than 85.

$$\begin{aligned} P[X < 85] &= P\left[\frac{X-100}{15} < \frac{85-100}{15}\right] \\ &= P[Z < -1] = 0.1587 \end{aligned}$$

Let us do some problems now.

Problem 6: For each of these write down the equivalent standard normal probability.

a) The number of people who visit a historic monument in a week is normally distributed with a mean of 10,500 and a standard deviation of 600. Consider the probability that fewer than 9000 people visit in a week.

b) The number of cheques processed by a bank each day is normally distributed with a mean of 30,100 and a standard deviation of 2450. Consider the probability that the bank processes more than 32,000 cheques in a day.

Solution: Here we want to find the standard normal probability corresponding to the probability $P[X < 9000]$.

a) We have $P[X < 9000] = P\left[\frac{X-10500}{600} < \frac{9000-10500}{600}\right] = P[Z < -2.5]$.

b) Here we want to find the standard normal probability corresponding to the probability $P[X > 32000]$.

$$P[X > 32000] = P\left[\frac{X-30100}{2450} > \frac{32000-30100}{2450}\right] = P[Z > 0.78]$$

———×———

Note Probabilities like $P[a < X < b]$ can be calculated in the same way. The only difference is that when X is standardised, similar operations must be applied to both a and b. That is, $a < X < b$ becomes

$$\frac{a-\mu}{\sigma} < \frac{X-\mu}{\sigma} < \frac{b-\mu}{\sigma}$$

which is

$$\frac{a-\mu}{\sigma} < Z < \frac{b-\mu}{\sigma}$$

Problem 7: An individual's IQ score has a $N(100, 15^2)$ distribution. Find the probability that an individual's IQ score is between 91 and 121.

Solution: We require $P[91 < X < 121]$. Standardising gives

$$P\left[\frac{91-100}{15} < \frac{X-100}{15} < \frac{121-100}{15}\right].$$

The middle term is a standardised normal random variable and so we have,

$$P\left[\frac{-9}{15} < Z < \frac{21}{15}\right] = P[-0.6 < Z < 1.4] = 0.9192 - 0.2743 = 0.6449.$$

———×———

Try these exercises now.

E18) A flight is due at Palam airport at 1800 hours. Its arrival time has a normal distribution with mean 1810 hours and standard deviation 10 minutes.

a) What is the probability that the flight arrives before its due time?

b) Passengers must check in for a connecting flight by 1830 at the latest. What is the probability that passengers from the first flight arrive too late for the connecting flight? (Assume no travelling time from aircraft to check-in.)

E19) The length of metallic strips produced by a machine has mean 100 cm. and variance 2.25 cm. Only strips with a weight between 98 and 103 cm. are acceptable. What proportion of strips will be acceptable? You may assume that the length of a strip has a normal distribution.

With this we come to an end of this unit.

Let us now summarise the points we have covered in this unit.

3.7 SUMMARY

In this unit we have covered the following points

1) A random variable is a variable that takes on different numerical values according to chance outcomes

2) There are two types of random variables - discrete and continuous;

3) A probability distribution gives the probabilities with which the random variables take an various values in their range.

4) We have discussed three standard distributions:

a) **Binomial Distribution.** The probabilities of an event $P[X = r]$ in this distribution is given by

$$P[X = r] = C(n, r)P^n q^{n-r}$$

b) **Poisson distribution:** The probability of an event $P[X = x]$ in this distribution is given by

$$P[X = x] = \frac{e^{-\lambda}\lambda^x}{x!}$$

where λ is a constant for a particular situation.

c) **Uniform Distribution** The probability density function is defined by

$$f(x) = \begin{cases} \frac{1}{b-a}, & \text{if } a \leq x \leq b \\ 0, & \text{elsewhere} \end{cases}$$

The probability $P[c < X < d] = \frac{d-c}{b-a}$

d) **Normal distribution.** The probability for this distribution is calculated by finding the area, under the curve of a function called probability density function defined by

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

3.8 SOLUTIONS/ANSWERS

E1) a) If X denote the number of correct answers, then X is the random variables for this situation.

b) X can take values $0, 1, 2, \ldots$ up to 50

c) $P[X = 40]$ means the probability that the number of correct answers is 40.

E2) 1 and 2 is not discrete. 2 and 3 are discrete.
(1) is not discrete because it takes values in an interval.

(2) is discrete because the number of accidents is finite. Similarly, you argue for the situation in (3).

E3) Let X denote the amount you win or lose. Then X takes values Rs. .50, 0 or -10 (loss in Rs. 10). The probability that both the marbles are green is $1/9$. The i.e. $P[X = 50] = 1/9$. The probability that both the marbles are red is $4/9$ i.e. $P[X = -10] = 4/9$.

The probability that the marbles are of different colour is $4/9$ is i.e. $P[X = 0] = 4/9$.

Thus the probability distribution is as given in the following table.

Amount (in Rs. won (+) or lost(−)	Probability
50	1/4
0	4/9
−10	4/9

E4) He has to calculate the mean. It is given by

$$\text{Mean} = \frac{0 \times 0.5 + 1 \times 0.15 + 2 \times 0.35 + 4 \times 0.12 \times 5 \times 0.18}{0.05 + 0.15 + 0.35 + 0.25 + 0.12 + 0.18}$$
$$= \frac{2.48}{1} = 2.48.$$

This means that he can expect that on an 2 average cars will be sold per day over long run (or more precisely 5 cars will be sold over (2 days)).

E5) i)

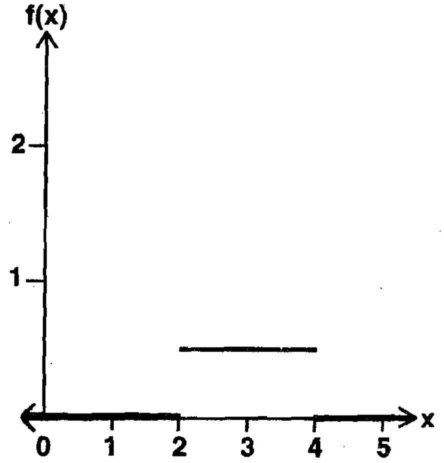


Fig.16

ii) The area under the graph and above the interval $[2, 4]$ is the area of the rectangle shown in Fig.16 which is given by
Area $= 2 \times \frac{1}{2} = 1$.
$\therefore$ f defines a probability density function of X.

E6) 1,3,4,5,6 are discrete. 2 and 7 are continuous.

E7) Here we have to calculate Probability of 3 heads and 2 tails. That is $P[X = 3]$.
The possibility of getting 3 heads and 2 tails in five tosses of a coin is given by [HHHTT], [HHTHT], [HHTTH], [HTHHT], [HTTHH], [THTHH], [TTHHH], [THHHT], [THHTH], [THTHH], [HTHTH].
Each of these events are having probability p^3q^2 and there are 10 such events. Therefore we get

$$P[X = 3] = 10p^3q^2.$$

If we apply the formula in Equation (3) to find $P[X=3]$, then we substitute $r=3, n=5$ in the formula, and we get

$$\begin{aligned} P[X=3] &= C(5,3)p^3q^2 \\ &= \frac{5!}{2! \times 3!}p^3q^3 \\ &= 10p^3q^2. \end{aligned}$$

E8) This situation follows binomial distribution with n=4 and $p = \frac{80}{100} = \frac{4}{5}$. The random variable X is the number of seeds that germinate. We have to calculate the probability that exactly two of the four seeds will germinate. That is $P[X=2]$. By applying binomial formula, we get

$$\begin{aligned} P[X=2] &= C(4,2)\left(\frac{4}{5}\right)^2 \times \left(\frac{1}{5}\right)^2 \\ &= 6 \times \frac{16}{625} \\ &= 0.154 \end{aligned}$$

Therefore the required probability is 0.154.

E9) If X_i denote the random variable that the ith customer buys the paper on a given day, then X_i's may not be identically distributed. Therefore X_i's may not be binomially distributed. But if the customers are having the same business activities or same kind of habits or working nature, then we can expect that X_i's will be identically distributed. In such situation we can expect that X_i's will follow binomial distribution.

E10) The number of sales in a day is actually $X_1 + X_2 + \ldots + X_{10}$ where each X_i is either 0 or 1 depending on whether customer i buys the paper or not on the given day. Now since customer 8 is more likely to buy on a day, than customer 3, X_3 and X_8 are not identically distributed. That is, $P[X_8 = 1] > P[X_3 = 1]$. Therefore $X_1 + X_2 + \ldots + X_{10}$ cannot be thought of as binomially distributed random variable.

E11) Since the problem deals with the receipts of bad cheques which is an event with rare occurrence over an interval of time (a day, in this case), we can apply Poisson distribution.

Since on an average 6 bad cheques are received per day,

Substituting $\lambda = 6$ and $x = 4$ in the Poisson Formula, we get

$$\begin{aligned} P[X=4] &= \frac{6^4e^{-6}}{4!} = \frac{1296 \times (0.0025)}{24} \\ &= 0.135. \end{aligned}$$

E12) Note that here the experiment or trial is 'checking the machine for its functioning. There are 20 trials and each trial is identically distributed with probability 0.2.

i) The trials are independent also. Therefore we can apply binomial formula. We are required to calculate $P[X=3]$. Then

$$\begin{aligned} P[X=3] &= \frac{20!}{3! \times 17!}(0.02)^3(0.98)^{17} \\ &= 0.0065. \end{aligned}$$

ii) Here we have to check whether we can apply Poisson distribution. Note that here the occurrences are 'function of the dialysis machines'. Then the average rate of machines that go out of service in a day is a constant $\lambda = 20 \times 0.02 = 0.4$.

Also note that we can make the subintervals so small that at best only one machine go out of service. Thus conditions (2) and (3) are satisfied. Therefore we can apply Poisson Formula to calculate the required probability P(3). Then

$$P(3) = \frac{(0.4)^3 e^{-0.4}}{3!}$$
$$= \frac{(0.64)(.67032)}{6}$$
$$= .00715.$$

E13) a) We can model it as uniform.

b) We cannot model it as uniform.

E14) i) The mean, 10 is the centre point of a line segment whose length is the range, 1.8 kg. Hence, the line segment extends $\frac{1}{2} \times (1.8) = 0.9$ kg. to the left and to the right of 10 i.e. 9.1 to 10.9 kg. Hence the smallest weight is 9.1 kg. and the largest weight is 10.9 kg.

ii) We are required to calculate $p[9 < X < 10.5]$.

$$p[9 < X < 10.5] = (10.5 - 9) \times \frac{1}{1.8}$$
$$= 0.833.$$

That is the probability that the weight lies between 9.1 kg. and 10.9 kg. is 0.833.

E15) Here the pdf f(x) is given by

$$\begin{cases} \frac{1}{32}, & 5.28 < x < 6 \\ 0, \text{otherwise} & \end{cases}$$

The sketch of f(x) is as given below,

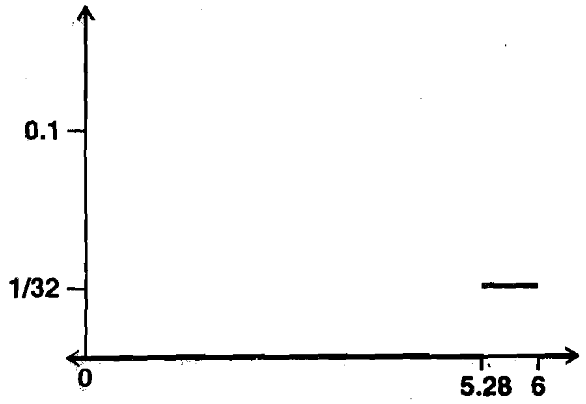


Fig.17

i) To find the required probability we note that X can take values in the interval (5.28, 5.40). Hence the required probability is

$$P[5.28 < X < 5.40] = 0.2 \times \frac{1}{32} = \frac{0.003}{8} = 0.0037$$

E16) a) 0.9332

b) 0,1151

c) 0.9599.

E17) The standard normal probability corresponding to this probability is given by

$$P[952 < X < 956] = P\left[\frac{952 - 952}{4} < \frac{X - 952}{4} < \frac{952 - 956}{4}\right]$$
$$= P[0 < Z < 1]$$
$$= F(1) - F(0)$$
$$= 0.8413 - 0.5$$
$$= 0.343.$$

E18) Let the time of arrival in minutes past 1800hrs be X. Then X follows normal distribution $N(10, 10^2)$.

a) The required probability is $P[X < 18]$. The standard probability corresponding to this is

$$P\left[Z < \frac{18 - 18.10}{10}\right] = P[Z < .0]$$
$$= F(0.01)$$
$$= 0.5040.$$

b) The required probability is $P[X > 30]$. Then

$$P[X > 30] = P\left[\frac{X-10}{10} > \frac{30-10}{10}\right] = P[Z > 2]$$
$$= 1 - P[z \leq 2]$$
$$= 1 - 0.9772 = 0.0228.$$

E19) We have to find the probability $P[98 < X < 103]$. The standard probability is $P\left[\frac{-2}{\sqrt{2.25}} < Z < \frac{3}{\sqrt{2.25}}\right]$.

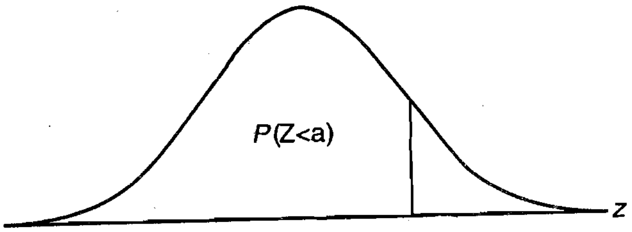
$$\text{i.e.}P\left[\frac{-2}{15} < Z < \frac{3}{15}\right] = P[-0.13 < Z << 0.2]$$
$$= F(0.2) - F(-0.13)$$
$$= F(0.2) - (1 - F(0.13))$$
$$= 0.5793 - (1 - 0.5517)$$
$$= 0.5793 - 0.4483$$
$$= 0.1310$$
$$= \frac{13}{100}\text{approximately}$$

So, only 13 % will be acceptable.

3.9 APPENDIX

Cumulative Standard Normal Probabilities $P[Z < a]$ where $Z \sim N(0, 1)$.

a	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7703	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998
3.6	0.9998	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.7	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.8	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000



POISSON PROBABILITY DISTRIBUTION

This table shows the value of

$$P(x) = \frac{\mu^x e^{-\mu}}{x!}$$

for selected values of x and for $\mu = .005$ to 8.0.

	μ									
x	.005	.01	.02	.03	.04	.05	.06	.07	.08	.09
0	.9950	.9900	.9802	.9704	.9608	.9512	.9418	.9324	.9231	.9139
1	.0050	.0099	.0192	.0291	.0384	.0476	.0565	.0653	.0738	.0823
2	.0000	.0000	.0002	.0004	.0008	.0012	.0017	.0023	.0030	.0037
3	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0001	.0001

	μ									
x	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
0	.9048	.8187	.7408	.6703	.6065	.5488	.4966	.4493	.4066	.3679
1	.0905	.1637	.2222	.2681	.3033	.3293	.3476	.3595	.3659	.3679
2	.0045	.0164	.0333	.0536	.0758	.0988	.1217	.1438	.1647	.1839
3	.0002	.0011	.0033	.0072	.0126	.0198	.0284	.0383	.0494	.0613
4	.0000	.0001	.0002	.0007	.0016	.0030	.0050	.0077	.0111	.0153
5	.0000	.0000	.0000	.0001	.0002	.0004	.0007	.0012	.0020	.0031
6	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0002	.0003	.0005
7	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001

	μ									
x	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	2.0
0	.3329	.3012	.2725	.2466	.2231	.2019	.1827	.1653	.1496	.1353
1	.3662	.3614	.3543	.3452	.3347	.3230	.3106	.2975	.2842	.2707
2	.2014	.2169	.2303	.2417	.2510	.2584	.2640	.2678	.2700	.2707
3	.0738	.0867	.0998	.1128	.1255	.1378	.1496	.1607	.1710	.1804
4	.0203	.0260	.0324	.0395	.0471	.0551	.0636	.0723	.0812	.0902
5	.0045	.0062	.0084	.0111	.0141	.0176	.0216	.0260	.0309	.0361
6	.0008	.0012	.0018	.0026	.0035	.0047	.0061	.0078	.0098	.0120
7	.0001	.0002	.0003	.0005	.0008	.0011	.0015	.0020	.0027	.0034
8	.0000	.0000	.0001	.0001	.0001	.0002	.0003	.0005	.0006	.0009
9	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0001	.0001	.0002

	μ									
x	2.1	2.2	2.3	2.4	2.5	2.6	2.7	2.8	2.9	3.0
0	.1225	.1108	.1003	.0907	.0821	.0743	.0672	.0608	.0550	.0498
1	.2572	.2438	.2306	.2177	.2052	.1931	.1815	.1703	.1596	.1494
2	.2700	.2681	.2652	.2613	.2565	.2510	.2450	.2384	.2314	.2240
3	.1890	.1966	.2033	.2090	.2138	.2176	.2205	.2225	.2237	.2240
4	.0992	.1082	.1169	.1254	.1336	.1414	.1488	.1557	.1622	.1680

Continued .....

5	.0417	.0476	.0538	.0602	.0668	.0735	.0804	.0872	.0940	.1008
6	.0146	.0174	.0206	.0241	.0278	.0319	.0362	.0407	.0455	.0504
7	.0044	.0055	.0068	.0083	.0099	.0118	.0139	.0163	.0188	.0216
8	.0011	.0015	.0019	.0025	.0031	.0038	.0047	.0057	.0068	.0081
9	.0003	.0004	.0005	.0007	.0009	.0011	.0014	.0018	.0022	.0027
10	.0001	.0001	.0001	.0002	.0002	.0003	.0004	.0005	.0006	.0008
11	.0000	.0000	.0000	.0000	.0000	.0001	.0001	.0001	.0002	.0002
12	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001

					μ					
x	3.1	3.2	3.3	3.4	3.5	3.6	3.7	3.8	3.9	4.0
0	.0450	.0408	.0369	.0334	.0302	.0273	.0247	.0224	.0202	.0183
1	.1397	.1304	.1217	.1135	.1057	.0984	.0915	.0850	.0789	.0733
2	.2165	.2087	.2008	.1929	.1850	.1771	.1692	.1615	.1539	.1465
3	.2237	.2226	.2209	.2186	.2158	.2125	.2087	.2046	.2001	.1954
4	.1734	.1781	.1823	.1858	.1888	.1912	.1931	.1944	.1951	.1954
5	.1075	.1140	.1203	.1264	.1322	.1377	.1429	.1477	.1522	.1563
6	.0555	.0608	.0662	.0716	.0771	.0826	.0881	.0936	.0989	.1042
7	.0246	.0278	.0312	.0348	.0385	.0425	.0466	.0508	.0551	.0595
8	.0095	.0111	.0129	.0148	.0169	.0191	.0215	.0241	.0269	.0298
9	.0033	.0040	.0047	.0056	.0066	.0076	.0089	.0102	.0116	.0132
10	.0010	.0013	.0016	.0019	.0023	.0028	.0033	.0039	.0045	.0053
11	.0003	.0004	.0005	.0006	.0007	.0009	.0011	.0013	.0016	.0019
12	.0001	.0001	.0001	.0002	.0002	.0003	.0003	.0004	.0005	.0006
13	.0000	.0000	.0000	.0000	.0001	.0001	.0001	.0001	.0002	.0002
14	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001

					μ					
x	4.1	4.2	4.3	4.4	4.5	4.6	4.7	4.8	4.9	5.0
0	.0166	.0150	.0136	.0123	.0111	.0101	.0091	.0082	.0074	.0067
1	.0679	.0630	.0583	.0540	.0500	.0462	.0427	.0395	.0365	.0337
2	.1393	.1323	.1254	.1188	.1125	.1063	.1005	.0948	.0894	.0842
3	.1904	.1852	.1798	.1743	.1687	.1631	.1574	.1517	.1460	.1404
4	.1951	.1944	.1933	.1917	.1898	.1875	.1849	.1820	.1789	.1755
5	.1600	.1633	.1662	.1687	.1708	.1725	.1738	.1747	.1753	.1755
6	.1093	.1143	.1191	.1237	.1281	.1323	.1362	.1398	.1432	.1462
7	.0640	.0686	.0732	.0778	.0824	.0869	.0914	.0959	.1002	.1044
8	.0328	.0360	.0393	.0428	.0463	.0500	.0537	.0575	.0614	.0653
9	.0150	.0168	.0188	.0209	.0232	.0255	.0280	.0307	.0334	.0363

Continued ....

					μ					
x	4.1	4.2	4.3	4.4	4.5	4.6	4.7	4.8	4.9	5.0
10	.0061	.0071	.0081	.0092	.0104	.0118	.0132	.0147	.0164	.0181
11	.0023	.0027	.0032	.0037	.0043	.0049	.0056	.0064	.0073	.0082
12	.0008	.0009	.0011	.0014	.0016	.0019	.0022	.0026	.0030	.0034
13	.0002	.0003	.0004	.0005	.0006	.0007	.0008	.0009	.0011	.0013
14	.0001	.0001	.0001	.0001	.0002	.0002	.0003	.0003	.0004	.0005
15	.0000	.0000	.0000	.0000	.0001	.0001	0001	.0001	.0001	.0002

					μ					
x	5.1	5.2	5.3	5.4	5.5	5.6	5.7	5.8	5.9	6.0
0	.0061	.0055	.0050	.0045	.0041	.0037	.0033	.0030	.0027	.0025
1	.0311	.0287	.0265	.0244	.0225	.0207	.0191	.0176	.0162	.0149
2	.0793	.0746	.0701	.0659	.0618	.0580	.0544	.0509	.0477	.0446
3	.1348	.1293	.1239	.1185	.1133	.1082	.1033	.0985	.0938	.0892
4	.1719	.1681	.1641	.1600	.1558	.1515	.1472	.1428	.1383	.1339
5	.1753	.1748	.1740	.1728	.1714	.1697	.1678	.1656	.1632	.1606
6	.1490	.1515	.1537	.1555	.1571	.1584	.1594	.1601	.1605	.1606
7	.1086	.1125	.1163	.1200	.1234	.1267	.1298	.1326	.1353	.1377
8	.0692	.0731	.0771	.0810	.0849	.0887	.0925	.0962	.0998	.1033
9	.0392	.0423	.0454	.0486	.0519	.0552	.0586	.0620	.0654	.0688
10	.0200	.0220	.0241	.0262	.0285	.0309	.0334	.0359	.0386	.0413
11	.0093	.0104	.0116	.0129	.0143	.0157	.0173	.0190	.0207	.0225
12	.0039	.0045	.0051	.0058	.0065	.0073	.0082	.0092	.0102	.0113
13	.0015	.0018	.0021	.0024	.0028	.0032	.0036	.0041	.0046	.0052
14	.0006	.0007	.0008	.0009	.0011	.0013	.0015	.0017	.0019	.0022
15	.0002	.0002	.0003	.0003	.0004	.0005	.0006	.0007	.0008	.0009
16	.0001	.0001	.0001	.0001	.0001	.0002	0002	.0002	.0003	.0003
17	.0000	.0000	.0000	.0000	.0000	.0001	.0001	.0001	.0001	.0001

					μ					
x	6.1	6.2	6.3	6.4	6.5	6.6	6.7	6.8	6.9	7.0
0	.0022	.0020	.0018	.0017	.0015	.0014	.0012	.0011	.0010	.0009
1	.0137	.0126	.0116	.0106	.0098	.0090	.0082	.0076	.0070	.0064
2	.0417	.0390	.0364	.0340	.0318	.0296	.0276	.0258	.0240	.0223
3	.0848	.0806	.0765	.0726	.0688	.0652	.0617	.0584	.0552	.0521
4	.1294	.1249	.1205	.1162	.1118	.1076	.1034	.0992	.0952	.0912
5	.1579	.1549	.1519	.1487	.1454	.1420	.1385	.1349	.1314	.1277
6	.1605	.1601	.1595	.1586	.1575	.1562	.1546	.1529	.1511	.1490
7	.1399	.1418	.1435	.1450	.1462	.1472	.1480	.1486	.1489	.1490
8	.1066	.1099	.1130	.1160	.1188	.1215	.1240	.1263	.1284	.1304
9	.0723	.0757	.0791	.0825	.0858	.0891	.0923	.0954	.0985	.1014

Continued .....

x	6.1	6.2	6.3	6.4	6.5	6.6	6.7	6.8	6.9	7.0
10	.0441	.0469	.0498	.0528	.0558	.0588	.0618	.0649	.0679	.0710
11	.0245	.0265	.0285	.0307	.0330	.0353	.0377	.0401	.0426	.0452
12	.0124	.0137	.0150	.0164	.0179	.0194	.0210	.0227	.0245	.0264
13	.0058	.0065	.0073	.0081	.0089	.0098	.0108	.0119	.0130	.0142
14	.0025	.0029	.0033	.0037	.0041	.0046	.0052	.0058	.0064	.0071
15	.0010	.0012	.0014	.0016	.0018	.0020	.0023	.0026	.0029	.0033
16	.0004	.0005	.0005	.0006	.0007	.0008	.0010	.0011	.0013	.0014
17	.0001	.0002	.0002	.0002	.0003	.0003	.0004	.0004	.0005	.0006
18	.0000	.0001	.0001	.0001	.0001	.0001	.0001	.0002	.0002	.0002
19	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0001	.0001

μ

x	7.1	7.2	7.3	7.4	7.5	7.6	7.7	7.8	7.9	8.0
0	.0008	.0007	.0007	.0006	.0006	.0005	.0005	.0004	.0004	.0003
1	.0059	.0054	.0049	.0045	.0041	.0038	.0035	.0032	.0029	.0027
2	.0208	.0194	.0180	.0167	.0156	.0145	.0134	.0125	.0116	.0107
3	.0492	.0464	.0438	.0413	.0389	.0366	.0345	.0324	.0305	.0286
4	.0874	.0836	.0799	.0764	.0729	.0696	.0663	.0632	.0602	.0573
5	.1241	.1204	.1167	.1130	.1094	.1057	.1021	.0986	.0951	.0916
6	.1468	.1445	.1420	.1394	.1367	.1339	.1311	.1282	.1252	.1221
7	.1489	.1486	.1481	.1474	.1465	.1454	.1442	.1428	.1413	.1396
8	.1321	.1337	.1351	.1363	.1373	.1382	.1388	.1392	.1395	.1396
9	.1042	.1070	.1096	.1121	.1144	.1167	.1187	.1207	.1224	.1241
10	.0740	.0770	.0800	.0829	.0858	.0887	.0914	.0941	.0967	.0993
11	.0478	.0504	.0531	.0558	.0585	.0613	.0640	.0667	.0695	.0722
12	.0283	.0303	.0323	.0344	.0366	.0388	.0411	.0434	.0457	.0481
13	.0154	.0168	.0181	.0196	.0211	.0227	.0243	.0260	.0278	.0296
14	.0078	.0086	.0095	.0104	.0113	.0123	.0134	.0145	.0157	.0169
15	.0037	.0041	.0046	.0051	.0057	.0062	.0069	.0075	.0083	.0090
16	.0016	.0019	.0021	.0024	.0026	.0030	.0033	.0037	.0041	.0045
17	.0007	.0008	.0009	.0010	.0012	.0013	.0015	.0017	.0019	.0021
18	.0003	.0003	.0004	.0004	.0005	.0006	.0006	.0007	.0008	.0009
19	.0001	.0001	.0001	.0002	.0002	.0002	.0003	.0003	.0003	.0004
20	.0000	.0000	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0002
21	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0001