UNIT 13 DISCRETE PROBABILITY DISTRIBUTIONS

Structure

13.1 INTRODUCTION

Most statistical methods useful to scientists deal with the collection, organisation, analysis and presentation of data. Such analysis of experimental data is used in making reasonable decisions based on the data.

In Unit 11, you have seen that we can organise the data based on a sample result in tabular form with the help of frequency distributions. Using these distributions we can analyse the data. In Unit 12, you studied various rules of probability which enables us to predict how often an event will occur during the entire process. In this unit we combine these ideas and form a **probability distribution.** Probability distribution can theoretically determine the probability of an event depending on the nature of the event and the conditions under which the event is occuring. In this unit you will learn about two such distributions: the **binomial distribution** and the **Poisson distribution.**

You already know from Unit 11 that events are of two types – discrete and continuous. We recall that a discrete random variable can assume values only in a finite set with no possible values located between one value and the next. Whereas, a continuous random variable assumes any one of infinitely large number of values found on a line interval. The probability distribution for discrete variables can be described by binomial and Poisson distributions. These distributions are called **discrete probability distributions.** Here we explain mainly how the binomial and Poisson distributions are generated by taking into consideration the simple laws of probability.

We have also given their means and standard deviations here. In the next unit we shall discuss some probability distributions for continuous variables.

Objectives

After you have completed this unit you should be able to

● describe the characteristics of the binomial and the Poisson distributions;
● choose which distribution to use in a given situation;
● apply the binomial and the Poisson distributions to solve problems;
● obtain their mean and standard deviation.

13.2 COMBINATION OF EVENTS

Sometimes we are concerned with problems which deal with combination of events. For example, instead of tossing one coin, let us toss two coins simultaneously. In this case, each tossing consists of a combination of two events. These two events are: head or tail of one, and head or tail of the other coin. Now can you find out how many different combinations are possible in tossing of two coins simultaneously? To find out the answer you probably think in the following directions: one possible combination is, heads of both the coins; a second possible combination is, head of one coin (say, the first coin) and tail of the other coin (say, the second coin); a third possible combination is, the head of the second coin and the tail of the first coin; and a fourth possible combination is, tails of both the coins. It your coins are identical, then you cannot distinguish between the coins while tossing them simultaneously. So the second and the third combinations are not distinguishable, and therefore, should be considered as one combination. So actually there are only three possible combinations.

Now, we ask the second question: What is the probability of each of these combinations? You already know that when you toss a single coin properly (that is, randomly), the probability of getting a head is 1/2. When you toss two coins, the probability of getting two head is $\frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$. This is because the probability of a combination of two events is obtained by multiplying the probabilities of the individual events, provided the events are independent of each other as in this case (whether you get a head or a tail of one coin does not depend on your getting a head or a tail of the other coin), Similarly, the probability of the second combination is $\left(\frac{1}{2} \times \frac{1}{2} \right) + \left(\frac{1}{2} \times \frac{1}{2} \right) = \frac{1}{2}$. You are probably wondering why the product of the individual probabilities is taken twice in this case. This is because this combination (head + tail) can be obtained in two ways, as we have already seen. Finally, the probability of the third combination is again $\frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$, because the probability of getting a tail in tossing a single coin is $\frac{1}{2}$

We now ask the third question: What is the pattern of distribution of these three combinations? The term, pattern of distribution means, in how many ways can we get each of these combinations. For instance, you have already noticed that the second possible combination (head + tail) can be obtained in two ways: On the other hand, the first and the third combinations are obtained in only one way each. So we can say that the pattern of distribution of the three combinations is (head, head) : (head, tail) : (tail, tail) = 1 : 2 : 1.

To summarise the story told so far: We are concerned with situations where combinations of events take place. Such a situation has three characteristics:

i) The number of possible combinations.
ii) The pattern of distribution of the combinations.
iii) The probability of each combination.

The above example shows the values that are possible outcomes from tossing two coins simultaneously and the probability of each outcome. This is an example of a probability distribution. You must have also noticed that the sum of probabilities in the distribution is 1. Thus, we can say that a probability distribution contains all theoretically possible values that can be assumed by a random variable. It also gives the probabilities of those values, where the sum of these probabilities is equal to 1. In the next section we shall discuss the binomial distribution. But before going to the next section you can try the following exercises.

E1) Consider the mating of parents having the following genotypes (genetic constitution of an individual) for a single gene.

Aa × aa

a) How many different genotypes are expected among the offspring?
b) What is the probability for each of these genotypes?
c) If parents have two children, how many combinations are possible in the genotypes of the children?
d) What is the pattern of distribution of these combinations?
e) What is the probability that both children have genotype Aa?

E2) Suppose you are tossing three identical coins simultaneously.

a) How many different combinations are possible?
b) What is the pattern of distribution of these combinations?.
c) What are the probabilities of having these individual combinations?

Suppose that we are tossing 3 coins and we are interested in the number of heads obtained in each toss. Then, we can take the number of heads obtained, as the random variable. The values taken by the variable are 0, 1, 2 or 3 depending upon the number of heads obtained. This is a discrete random variable. Here, do you see some similarity between the random variable and the term 'event' which we have discussed in Unit 11? If you carefully look at this example of tossing three coins then you will notice that the variable 'number of heads obtained' is the same as the event of getting a head. That means an event is the probabilistic term of a variable. Also the values that the variable takes is the corresponding numerical values that can be assigned to each sample point in the sample space favourable for that event. In Unit 11, you have already learnt about discrete random variables, continuous random variables and their frequency distributions.

We now discuss the binomial distribution.

13.3 THE BINOMIAL DISTRIBUTION

The binomial distribution as we have already mentioned in Sec. 13.1 is a discrete probability distribution i.e. it describes the probability of the event (variable) in case the event is a discrete variable. The binomial distribution is also called · Bernoulli distribution after James Bernoulli (1654-1705), a Swiss mathematician who discovered this distribution early in the eighteenth century.

In Sec. 13.2, we discussed the tossing of two coins. In such experiments we have two possible outcomes for an event namely, successes and failures. A success is simply the outcome favourable for the event. Such probabilities where there are two possible outcomes are called binomial probabilities. The corresponding distribution is called the binomial distribution. Let us consider the formulation of this distribution.

You know that $(p + q)^2 = p^2 + 2pq + q^2$ is a binomial expansion (ref. Unit 3). You will be surprised to see that this particular binomial expansion can be so appropriately used to describe the situation of combination of two events like the tossing of two coins or one of the two children from the cross between Aa × aa. In the case of tossing of two coins, let us take p as the probability of getting head and q as the probability of getting tail. Then $p = \frac{1}{2}$ and $q = \frac{1}{2}$. Now, substitute the values of p in the first term, on the right hand side of the above expansion i.e.,

$$p^2 = p \times p = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$$

This is the probability of getting two heads when the two coins are tossed. Then, take the second term and substitute the values of p and q in it:

$$2 \, pq = pq + pq = \left(\frac{1}{2} \times \frac{1}{2} \right) + \left(\frac{1}{2} \times \frac{1}{2} \right) = \frac{1}{2}$$

This is the probability of getting the head and tail combination in the tossing of two coins. Similarly, substituting the value of q in the third term

$$q^2 = q \times q = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$$

This is the probability of getting the tail and tail combination in the tossing of two coins.

Thus,

i)   There are three possible combinations in the case of tossing of the coins, and there are three terms in this particular binomial expansion.

ii)   Each term of the binomial expansion gives the value of the probability of the corresponding combination.

iii) The pattern of distribution of the combinations is 1 : 2 : 1, and the coefficients of the three terms of the binomial expansion also show the distribution pattern 1 : 2 : 1.

Therefore, the binomial expansion of $(p+q)^2$ completely describes the three characteristics of a combination of two events as in the case of tossing two coins.

The case of two children from the cross As $\times$ aa is also completely described by this particular binomial expansion. We leave it to you as an exercise to satisfy yourself.

E3)   Show that all the three characteristics of the genotypic combinations of two children from a cross between A $\times$ aa are completely described by the binomial expansion:

$$(p + q)^2 = p^2 + 2pq + q^2$$

Let us now extend the application of binomial expansion in the case of tossing **three** coins. You have already solved this problem in E2) using simple concepts of probability. You must have noticed there that there are **four** possible combinations and, on calculating the probabilities of these combinations, the pattern of distribution of these four combinations turns out to be 1 : 3 : 3 : 1.

If you consider the binomial expansion of $(p+q)^3$, then you will get all the information that you have obtained in E2), because

$$(p+q)^3 = p^3 + 3p^2q + 3pq^2 + q^3$$

Firstly there are four terms on the right hand side of this expansion. **Secondly**, if you substitute, $p = 1/2$ and $q = 1/2$ and calculate the values of these four terms, you will get the probabilities of the four combinations. **Thirdly**, the distribution of the coefficients of the four terms is 1 : 3 : 3 : 1.

You must remember that whether you toss two or three or any number of coins, you are dealing with only **two** events, head or tail. So the binomial expansion is applicable, but the power of the binomial will be different depending on the number of coins tossed.

Following the logic of binomial expansion, you can now apply a suitable binomial to describe situations where higher combinations of two events are involved. Using a suitable binomial you can find the probability of any particular type of combination without considering the other type of possible combinations. Consider the following example.

Suppose, you are asked to find the probability of having three children from the cross Aa $\times$ aa so that two children have genotype Aa and one has genotype aa. How will you proceed?

Firstly, you have to choose a binomial, not a trinomial or any higher polynomial, because only **two** types are possible from the cross. If, from a cross three types of

genotypes are possible, then an appropriate trinomial should be chosen. In the case of tossing coins, you always choose a binomial, because in tossing a coin there are only two possibilities head or tail. If you throw a dice which has six faces, then you have to use an appropriate polynomial with six terms.

In this unit we are concerned with binomials only; nevertheless, the above statements help you to determine where binomials cannot be used. Coming back to our problem, you know that you have to use a binomial only, because there are only two possible genotypes from the cross Aa $\times$ aa. Then you have to find out what exponent should be used for the binomial. Since it is a combination of **three** events (three children in this case), it should be a binomial of power 3, that is, $(p+q)^3$. This expansion has four terms which correspond to four different genotype combinations of three children. Now the question is: Which term should be chosen to solve the given problem? In other words, we want the term which corresponds to the combination of two children with genotypes Aa and one child with genotype aa. If we assume that p and q are the probabilities for Aa and for aa, respectively, then the required term is $3 \times p \times p \times q$, i.e., $3p^2 q$ of the binomial $(p+q)^3$. Once the required term is obtained, we calculate the probability by putting the values of p and q in the term. In this particular problem $p = 1/2$ and $q = 1/2$. Therefore the probability of having three children from the cross Aa $\times$ aa so that two children have genotype Aa and one child has genotype aa is

$$3p^2q = 3 \times \left(\frac{1}{2}\right)^2 \times \left(\frac{1}{2}\right) = \frac{3}{8}.$$ Thus, the probability of a given combination

is determined directly using an appropriate binomial.

Before proceeding further you may try the following exercises in order to make sure that you have understood the above discussion.

E4) Consider the case of tossing three coins:

    a) Should you use a binomial or a trinomial to describe the distribution of the possible combinations?

    b) What should be the power (exponent) of this bionomial/trinomial?

    c) Which terms of the expansion should be used to find the probabilities of the following combinations:

        i)   two heads and one tail

        ii) three tails

        iii) one head and two tails,

        iv) three heads

    d) What are the probabilities of the combinations mentioned in (c)?

E5) Which power of the binomial would you use to describe the various possible combinations? If

    a) you are tossing four coins

    b) you are tossing seven coins

Now you have learnt that a term of a binomial expansion gives the probability of a given combination of two events. In tossing of a coin, head and tail are the two events. In a cross between Aa $\times$ aa, the two events are the offspring of genotype Aa and the offspring of genotype aa. The number of possible combinations depends on the total number of coins tossed or the total number of offsprings in the family. These combinations correspond to the terms of the appropriate binomial. It is therefore, necessary to know some general rule which gives these terms very easily.

13.3.1 Terms

So far, you have learnt to find out the value of any of the terms of a binomial of power 2 and of power 3, that is for $(p+q)^2$ and $(p+q)^3$, respectively. Now if you have done E 5), you will agree that if you toss 7 coins, then you have to deal with the binomial of order $(p+q)^7$, and you should also know all the eight terms of this

expansion. As the order of the binomial increases its expansion to get all the terms also becomes difficult. But here, we can use the binomial theorem, which you have studied in Sec. 3.5, Unit 3. It gives the expansion of $(p+q)^n$ for all integral powers of n. The general formula for a xth term in this expansion of $(p+q)^n$ is given by

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

Thus, we arrive at a pattern for binomial distribution.

We now give the **binomial distribution formula.**

Consider a binomial experiment that has two possible outcomes, success or failure. Let P (success) = p and P (failure) = q. If this experiment is performed n times, then the probability of getting x successes out of the n trial is

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

$$=\ \frac{n!}{x!\,w!}\,p^x\,q^w,\ where\ w\ =\ n-x.$$

We now illustrate this formula with the help of an example.

Example 1: Suppose seven coins have been tossed. Calculate the probability of a combination which contains 4 heads and 3 tails.

Solution: To find out the required probability, we use the formula

$\dfrac{n!}{w!\,x!}\,p^x\,q^w$, where in this case n = 7, w = 4,

x = 3, and we already know that p $=\ \dfrac{1}{2}$ and q $=\ \dfrac{1}{2}$. Substituting all these values, the required probability is,

$$\frac{7!}{(4!)\,(3!)}\left(\frac{1}{2}\right)^4\times\left(\frac{1}{2}\right)^3$$

$$=\ \frac{7\times6\times5\times4\times3\times2\times1}{(4\times3\times2)\times(3\times2)}\times\frac{1}{16}\times\frac{1}{8}$$

$$=\ (7\times5)\times\frac{1}{16}\times\frac{1}{8}\ =\ 35\times\frac{1}{16}\times\frac{1}{8}.$$

$$=\ 0.273$$

Hence, in tossing 7 coins, the probability of getting a combination of 4 heads and 3 tails is 0.273. In other words, if you toss seven coins one thousand times, you may expect the combination of 4 heads + 3 tails to appear 1000×0.273 = 273 times.

Now, if you want to know only the **pattern** of distribution, you can calculate the coefficient in this expression for the general term. You do not have to know the values of p and q to find out the values of the coefficients. Do you agree that the pattern in our example of tossing 7 coins will be given by the coefficients

$$\frac{n!}{x!\,w!},\ w\ =\ 0,\ 1,\ 2,\ .....,\ 7,$$

that is, by 1, 7, 21, 35, 21, 7, 1?

How about doing some exercises now?

E6) Consider the case of tossing four coins simultaneously.

    a) Find the pattern of distribution of the combinations.

    b) Find the probabilities of all possible combinations.

d) If there are 1000 such families of five children each how many families will have all the five children of genotype Aa?

E8) Suppose you survey 256 families, each having six children.
a) How many families are expected to have all boys?
b) How many families are expected to have four boys and two girls?

E7) There are five children in a family of parents Aa × aa.
a) How many types of combinations are possible in the genotypes of the children?
b) What is the probability that two of the children have genotype Aa and the three others have genotype aa?
c) What is the pattern of distribution of these combinations?

13.3.2 Graphical Representation

You know that there are only two types (or classes) of events in a binomial distribution. Each of the two events occur in all the terms of the distribution with varying probabilities. Let us consider the problem as given is E8) once again. There you might have seen that the distribution of girls in all the seven terms, that is, seven possible combinations of the two events is as follows:

$$(p + q)^6 = p^6 + 6p^5q + 15p^4q^2 + 20p^3q^3 + 15p^2q^4 + 6pq^5 + q^6$$

Assuming the probability of a girl $= p = \dfrac{1}{2}$ and the probability of a boy $= q = \dfrac{1}{2}$, the values of the seven terms are

$$\frac{1}{64}, \frac{6}{64}, \frac{15}{64}, \frac{20}{64}, \frac{15}{64}, \frac{6}{64}, \frac{1}{64}$$
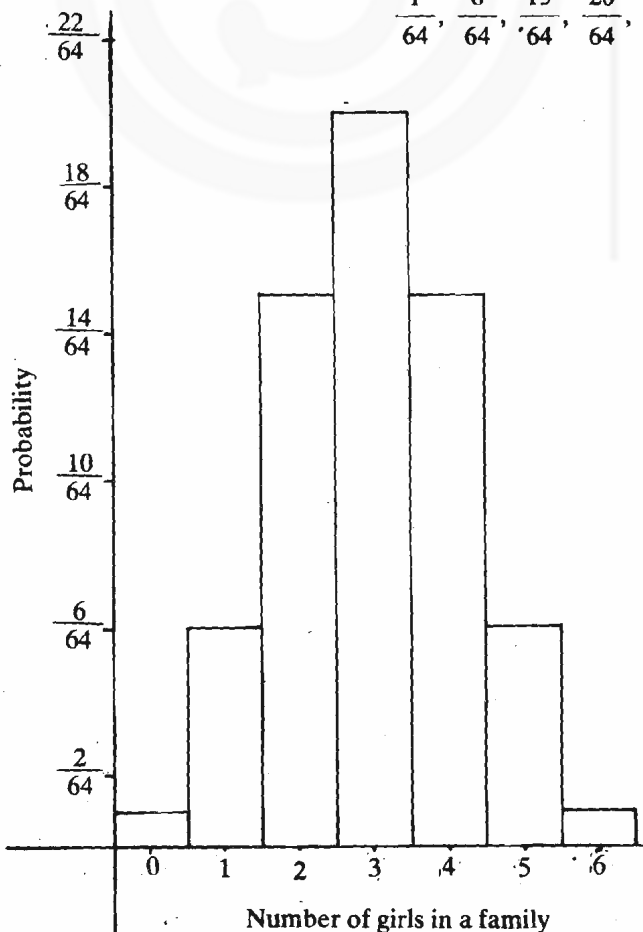
Number of girls in the family	Probability
6	$\dfrac{1}{64}$
5	$\dfrac{6}{64}$
4	$\dfrac{15}{64}$
3	$\dfrac{20}{64}$
2	$\dfrac{15}{64}$
1	$\dfrac{6}{64}$
0	$\dfrac{1}{64}$

Fig. 1: Probability distribution (in the form of a bar diagram) of the number of girls in a family with six children.

These seven terms, starting from the left side, are the probabilities of having 6 girls, 5 girls, 4 girls, 3 girls, 2 girls, 1 girl and no girl in a family of six children. This data can be plotted with probabilities along the y-axis and the corresponding number of girls along the x-axis to obtain the **probability distribution curve** (Fig. 1). For convenience, we can make a table in the following manner before plotting the data on a graph paper.

Note : that we have drawn the probability distribution curve shown in Fig. 1, in the form of a bar diagram. This curve can also be drawn in the form of a polygon. We leave it to you as an exercise.

E9) Draw the probability distribution curve shown in Fig. 1 in the form of a polygon.

The binomial distribution curve in Fig. 1 has been drawn using the probabilities on the y-axis. It can also be drawn using the frequencies, provided the total number of trials is knwon. For example, in E9), 256 families have been surveyed. So the expected numbers (frequencies) of families having 6, 5, 4, 3, 2, 1 and zero girls are obtained by multiplying the respective probabilities by 256, that is, the expected frequencies are 4, 24, 60, 80, 60, 24, 4 and zero, respectively. Now, if we plot the frequencies along the y-axis, then we obtain the **frequency polygon of the binomial distribution,** as shown in Fig. 2.
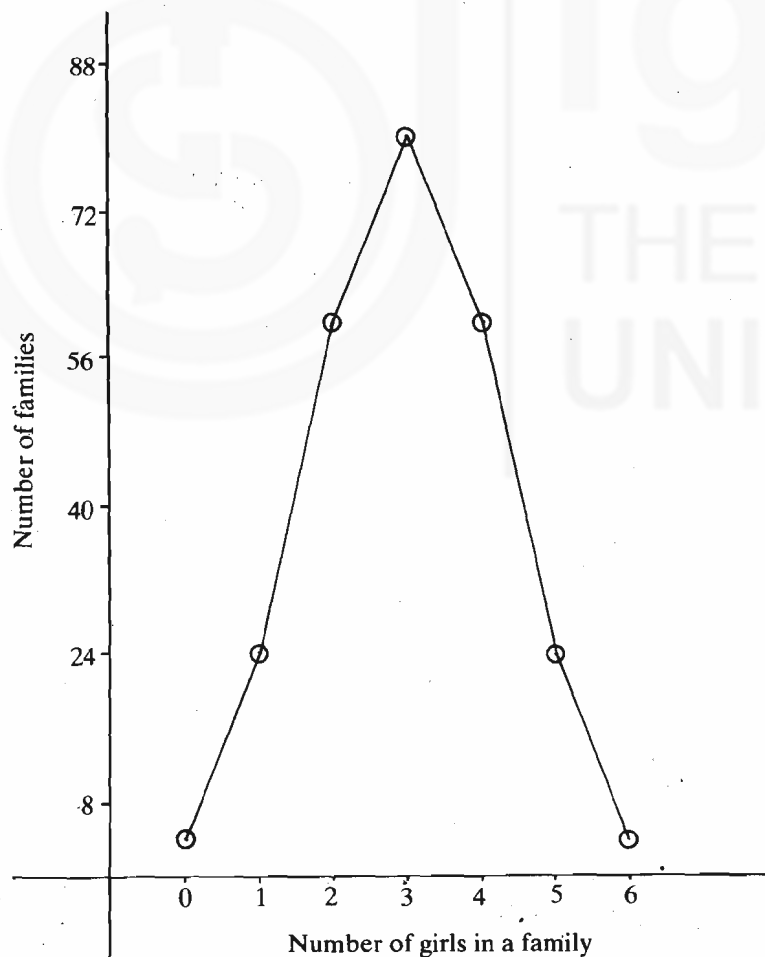

Fig. 2: Frequency distribution of families with given number of girls in 256 families with six children.

The shape of the binomial distribution curve is quite significant. Its shape depends on the probabilities p and q of the two events under consideration. In the next subsection we shall discuss how the shape of the distribution curve changes with the change in p and q.

13.3.3 The Shape of Curve

The shape of the binomial distribution curve as we have said earlier depends on the values of the probabilities p and q of the two events. When $p = q = 1/2$, the binomial distribution is a symmetrical curve with a peak at the centre. When the values of p and q are not equal, the distribution curve is not symmetrical, but is skewed. Also, the position of the peak is shifted. If $p > q$ the peak is shifted from the centre towards its right (i.e., away from the origin). If $p < q$, the peak is shifted from the centre towards its left (i.e., towards the origin).

Let us consider a familiar example from genetics. In a mating Aa ✕ Aa with, A being dominant over 'a', the offspring have only two phenotypes (phenotype biologically means physical appearance of an individual), so the probability distribution of the phenotypes will follow a binomial distribution. What are these two phenotypes, and what are the probabilities for each of the events (i.e., phenotypes)?

Consider the square

	A	a
A	AA prob. = 1/4	Aa prob. = 1/4
a	aA prob. = 1/4	aa prob. = 1/4

As the square shows, there are four possibilities, each with probability of 1/4. But AA, Aa and aA have the same phenotype because A is assumed to be dominant over a. So this phenotype (let us call it **tall** offspring) has the probability

$\frac{1}{4} + \frac{1}{4} + \frac{1}{4} = \frac{3}{4}$ and the other offspring (**short** offspring) has the probability

of $\frac{1}{4}$. Thus, we can write

$p = \frac{3}{4}$ (probability of tall class)

$q = \frac{1}{4}$ (probability of short class)

Now, let us consider a family of six offspring, and find out the probabilities of having 6, 5, 4, 3, 2, 1 and zero tall offspring. In this case we consider the binomial $(p+q)^6$.

$(p+q)^6 = p^6 + 6p^5q + 15p^4q^2 + 20p^3q^3 + 15p^2q^4 + 6pq^5 + q^6$

$$= \left(\frac{3}{4}\right)^6 + 6\left(\frac{1}{4}\right)\left(\frac{3}{4}\right)^5 + 15\left(\frac{1}{4}\right)^2\left(\frac{3}{4}\right)^4 + 20\left(\frac{3}{4}\right)^3\left(\frac{1}{4}\right)^3$$

$$+ 15\left(\frac{3}{4}\right)^2\left(\frac{1}{4}\right)^4 + 6\left(\frac{3}{4}\right)\left(\frac{1}{5}\right)^5 + \left(\frac{1}{4}\right)^6$$

$$= \frac{729}{4096} + \frac{1458}{4096} + \frac{1215}{4096} + \frac{540}{4096} + \frac{135}{4096} + \frac{18}{4096} + \frac{1}{4096}$$

Thus, the probability of having no tall offspring in a family of six offspring is $\frac{1}{4096}$ the probability of having 3 tall offsprings is $\frac{540}{4096}$, the probability of having 5 tall offsprings is $\frac{1458}{4096}$ and so on. If these probabilities are plotted on a

graph paper, the curve obtained will not be a symmetric one, but a skewed one. Also, the peak position will be different as compared to the peak position for the situation where p=q. You can now draw this graph yourself in order to solve the following exercise.

E10) Plot the graph for the example where p = 3/4 and q = 1/4, considered above in Section 13.3.3. Compare it, with Fig. 2 for which p = q = 1/2, and answer the following:

    a) For what value on the x-axis, do you get the peak position, that is, the maximum value of the probability in Fig. 2.

    b) For what value on the x-axis, do you get the peak position for the curve plotted in E10)?

    c) Compare (a) with (b). In which direction does the peak shift?

    d) Why do you say that the distribution when p = 3/4 and q = 1/4 is not symmetrical?

Recall, that in Unit 11, we discussed the mean and standard deviation of a distribution. In the next subsection, we now give a formula for the mean and standard deviation of the binomial distribution is terms of the three unknown parameters n, p and q (= 1-p).

13.3.4 Mean and Standard Deviation

The formulas for the mean and standard deviation of the binomial distribution can be directly calculated from the expression $(p+q)^n$. But, the computation is tedious. Here we simply state the formula without deriving it.

The **mean of a binomial distribution**, μ, is given by
$\mu = np$, where

n = total number of trials of the experiment
p = probability of success on each trial.

The standard deviation, σ, is given by

$\sigma = \sqrt{npq}$

A success is simply the outcome for which we wish to find the probability distribution

Let us consider on example:

Example 2: In a survey of 100 families, each family having 4 children, what is the expected mean number of boys in each family?

Solution: In this case p = probability of having a boy, $= \dfrac{1}{2}$ and n = the exponent of the binomial = 4.

Therefore, the mean number of boys per family

$= np = 4 \times \dfrac{1}{2} = 2$

This shows that, on an average there are two boys per family.

The standard deviation $= \sqrt{npq}$

$$= \sqrt{4 \times \frac{1}{2} \times \frac{1}{2}} = \sqrt{1} = 1$$
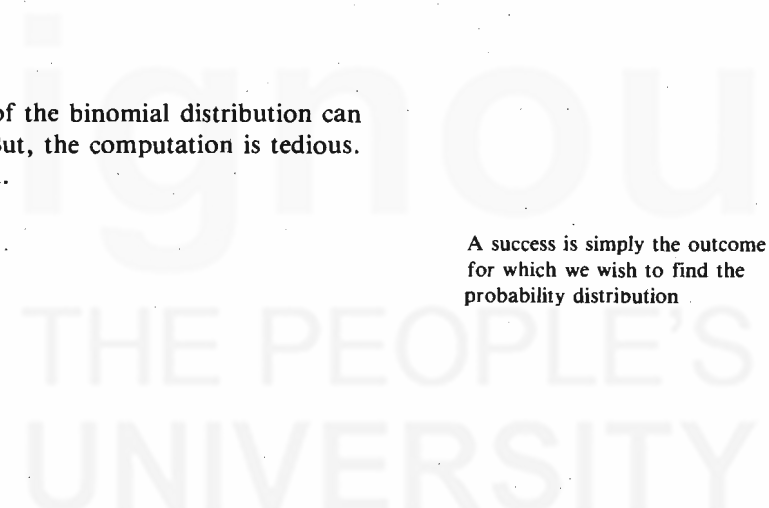
Can you find the mean number of girls per family? To find that out you should take the probability of having a girl. Now, the probability of having a

girl $= q = \dfrac{1}{2}$. Therefore. mean $= nq = 4 \times \dfrac{1}{2} = 2$

That is, on an average, there are two girls per family.

You may now try this exercise.

E11) In a mating Aa × Aa, the gene 'A' stands for 'tall' and gene 'a' stands for 'short', and 'A' is dominant over 'a'. If you consider a large number of families each having 8 offspring.

   a) What is the average (mean) number of tall offspring per family?
   b) What is the average (mean) number of short offspring per family?
   c) In (a) and (b) above, what is the value of the standard deviation?

With this exercise we end our discussion about binomial distribution. But, before we go on to the next probability distribution, we summarise all the important features of the binomial distribution.

i) A binomial distribution is a discrete probability distribution. The values of the events (or items or classes) are expressed in whole numbers, not in fractions. For instance, number of boys or girls, number of heads or tails, number of tall plants or short plants, etc.

ii) A binomial distribution is applied to a population which can be divided into only **two** mutually independent classes. Such a population is sometimes called a **dichotomised population.** For instance, a population can be divided into two classes, such as, boys and girls, tall and short, normal and colour blind, diabetic and non-diabetic, black and white, etc.

iii) A binomial distribution can be **plotted graphically** with the numbers of occurrence of one event along the x-axis, and the probabilities (that is, relative frequencies) or the frequencies along the y-axis.

iv) The shape of the binomial curve depends on the values of p and q. When $p = q = 0.5$, the binomial distribution is symmetrical. When p is not equal to q, the curve is skewed.

v) The formula for calculating the mean and the standard deviation of a binomial distribution is

Mean = np or nq

Standard deviation $= \sqrt{npq}$

There are many practical problems where we may be interested in finding the probability that x "successes" will occur over a given interval of time or a region of space. This is especially true when we do not expect many successes to occur over the time interval (which may be of any length, such as a minute, a day, a week, a month, or a year), For example, we may be interested in determining the number of days that a hockey match may be postponed in a given season because of rain or the number of days that school may be closed due to snowstorm. For these and similar problems we use the Poisson probability function formula. We shall now be discussing it in the next section.

13.4 POISSON DISTRIBUTION

The concept of Poisson distribution was developed by a French mathematician, S.D. Poisson (1781 – 1840) in 1837. This distribution gives the probability of an event occurring **rarely** and, of course, **randomly.** How can you define an event as a rare event? In the case of the binomial distribution $(p + q)^n$, p and q are the probabilities of two events. As long as p and q are equal, or p and q do not differ much from each other, we can apply the binomial distribution. But when one of the probabilities becomes very low, say $p = .01$ or even lower, then the event corresponding to probability p becomes a rare event. For a rare event, the distribution is better represented by a Poisson distribution than by a binomial distribution. Of course, it is not the low probability value alone that determines the suitability of the Poisson distribution. The value of n also is important. For the Poisson distribution to be suitable, the value of n should also be small. There

is no strict rule to lay down which events can be considered as rare events. A general convention is that if the product $n \times p$ is less than 5, the event may be taken as a rare event. On the basis of this convention you may try this exercise.

E12) Which of the following events are rare events?

    a) $p = 0.05$, $n = 200$

    b) $p = 0.05$, $n = 20$

    c) $p = 0.04$, $n = 50$

    d) $p = 0.04$, $n = 160$

    c) $p = 0.01$, $n = 200$

In case of binomial distributions, there are two distinct events. If one event does not occur, the other event does occur. But there are many events occurring in nature for which we can only know the number of times the event can occur or has occurred. The number of times that event has not occurred is irrelevant. For example, we can count how many times you blink your eyes every hour, but we cannot count the number of times you do not blink your eyes per hour. Other such examples are: the number of goals scored by a soccer team in various matches, the number of accidents occurring in a city every day, and so on. Under such situations you cannot apply the binomial distribution because the value of n is not known. The Poisson distribution can be applied to study these situations.

You may now yourself think of other such situations.

E13) Give five examples of situations in which only the frequency of occurrence of an event can be measured, not the non-occurrence of the event. (Do not take the examples given in the text).

In the next subsection we give you the formula which gives the Poisson probability of an event. We shall only be stating the formula without giving its derivation since its derivation is beyond the scope of this course.

13.4.1 Equation

The Poisson probability of an event occurring n times in a given time interval or specified region is given by

$$P(\text{n successes}) = P_n = \frac{e^{-\mu} \times \mu^n}{n!}, \quad n = 0,1,2,3,\ldots\ldots$$

where, e = base of the natural logarithm, whose value is approximately equal to 2.7183 correct to four decimal places. ·

    μ = average number of successes occurring in the given time interval.

Thus, to find the probability of an event occurring a given number (n) of times, the only parameter you have to know is 'μ' the average number of times the event occurs. The value of $e^{-\mu}$ can be obtained from a standard table (see appendix). To enable you to handle this equation, let us now discuss various numerical examples showing the use of Poisson distribution in cases of biological and other problems.

13.4.2 Numerical Examples

Consider the following example.

Example 3: A cat catches mice at an average rate of 4 mice per day.

a) What is the probability that on a given day the cat will catch 5 mice?

b) What is the probability that on a given day the cat will catch 4 mice?

c) Following the same method as in (a) and (b) find the probabilities of catching 3 mice, 2 mice, 1 mouse, and none at all.

Solution: $\mu = 4$, the average number of mice caught per day.

a) To find the probability of catching 5 mice per day, we write $n = 5$. Therefore, the required probability is

$$P_5 = \frac{e^{-4} \times 4^5}{5!}$$

Now, $e^{-4} = 0.0183$ (see appendix) and $5! = 5 \times 4 \times 3 \times 2$. So

$$P_5 = \frac{0.0183 \times 4^5}{5 \times 4 \times 3 \times 2} = 0.156$$

Thus, the probability that the cat will catch 5 mice on a given day is 0.156.

b) Following the same method as in (a), we put $n = 4$, and find that the required probability is

$$P_4 = \frac{e^{-4} \times 4^4}{4!}$$

$$= \frac{0.0183 \times 4 \times 4 \times 4 \times 4}{4 \times 3 \times 2} = 0.195$$

So the probability of catching 4 mice on a given day is 0.195. We are leaving part (c) to you to do.

You may now try this exercise.

E14) After calculating all the probabilities in Example 3, plot a graph with the number of mice along the x-axis and the corresponding probabilities along the y-axis. What is the shape of the graph? If you think that you need more points to get a clearer shape of the curve, you may calculate the probabilities of catching 6 mice and 7 mice on a given day.

Let us look at a few more examples.

Example 4: While walking in a forest an ecologist got 300 insect bites in two and half hours.

a) What is the average number of insect bites per minute?

b) What is the probability of no insect bite in a given minute?

c) For how many one-minute intervals was he free from insect bites?

Solution : a) $2\frac{1}{2}$ hours $= 150$ minutes

Average number of insect bites $= \dfrac{300 \text{ bites}}{150 \text{ minutes}} = 2$ bites/minute

b) The value of 'μ' $= 2$ bites/min. So the probability of zero insect bite is

$$P_n = \frac{e^{-\mu} \times \mu^n}{n!} \text{ with } \mu = 2, n = 0$$

or,

$$P_0 = \frac{e^{-2} \times 2^0}{0!}$$

Now, $e^{-2} = 0.1353$ (from the appendix), $2^0 = 1$ and $0! = 1$

Therefore, $P_0 = \dfrac{0.1353 \times 1}{1} = 0.1353$

So the probability of no insect bite is 0.1353.

c) The ecologist was in the forest for 150 min. Therefore, he was free from insect bites for $150 \times 0.1353 = 20.2$ minutes. So for about 20 one-minute intervals he was free from insect bites.

Example 5: The incidence of cell anemia in a population is 0.5%. In a sample of 200 persons from the population:

a) What is the average incidence of the disease in a sample of 200 persons?

b) What is the probability that 2 persons in this sample have the disease?

Solution:

a) Incidence of the disease is 0.5%. Hence the probability of its occurrence

$$= \frac{0.5}{100} = 0.005.$$

So, **on an average** the number of persons expected to suffer from the disease in a sample of 200 persons is $200 \times 0.005 = 1.0$.

b) We have found out that the average number of persons suffering from the disease is 1. So $\mu = 1$.

We have to find the probability that 2 persons are suffering from the disease. That is n = 2.
So, the required probability is

$$P_2 = \frac{e^{-1} \times 1^2}{2!}$$

Now, $e^{-1} = 0.367$

So, $P_2 = \frac{0.367}{2} = 0.183$

So the probability that two persons are suffering from the disease is 0.183.

13.4.3 Mean and Standard Deviation

As in the case of the binomial distribution, we now give you the formula to calculate the mean and the standard deviation of the Poisson distribution.

Mean $= \mu = np$

Standard deviation $= \sigma = \sqrt{\mu} = \sqrt{np}$

You already know that the value of p is very low (actually much less than 1). Hence the mean, which is equal to the product $n \times p$ is very small as compared to the value of n.
We now summarise our discussion about the Poisson distribution in short and say that

i) The Poisson distribution is a discrete probability distribution.
ii) It is applicable to dichotomised population, provided the probability p of one class is very low, so that $np < 5$. In other words, the event concerned is a rare event.

And now some exercises for you.

E15) In a certain factory turning out fountain pens, there is a small chance, 1/500, for any pen to be defective. The pens are supplied in packets of 10. Calculate the approximate number of packets containing no defective, one defective, two defective and three defective pens respectively in a consignment of 20,000 packets.

E16) In a certain Poisson frequency distribution the frequency corresponding to 2 successes is half the frequency corresponding to 3 successes. Find its mean and standard deviation.

We now conclude this unit by giving a summary of what we have covered in it.

13.5 SUMMARY

In this unit we have covered the following points:

1) The binomial distribution is a frequency distribution of various combinations of two discrete events which are independent of each other.

2) The number and pattern of a binomial distribution of the possible combinations and their probabilities can be obtained by choosing an appropriate binomial expansion $(p + q)^n$, where p, q are the probabilities of the two events with $p + q = 1$ and n is the total number of individuals in each combination.

3) A binomial distribution curve is symmetric for $p = q$ and skewed for $p \neq q$.

4) Mean of the binomial distribution = np and its standard deviation is $\sqrt{npq}$.

5) The Poisson distribution is a frequency distribution of a discrete event occurring rarely.

6) In a Poisson distribution the probability of an event occurring n times is given by the formula $P_n = \dfrac{e^{\pm\mu} \times \mu^n}{n!}$, where μ is the average number of times the event occurs.

7) The mean of the Poisson distribution is μ and standard deviation is $\sqrt{\mu}$.

13.6 SOLUTIONS/ANSWERS

E1) a) Two genotypes i.e., Aa and aa

b) Probability for Aa is 2 out of 4. That is,

Probability $= \dfrac{2}{4} = \dfrac{1}{2} = 0.5$

Probability for aa is also $= \dfrac{2}{4} = \dfrac{1}{2} = 0.5$

c) Three combinations are possible. These are (Aa, Aa) or (Aa, aa) or (aa, aa).

d) The pattern of distribution is (Aa, Aa) : (Aa, aa) : (aa, aa) $= 1 : 2 : 1$. Because, the first and the third combination can occur in one way only. The second combination can occur in two ways: The first child is Aa and the second is aa, or the first is aa and the second is Aa.

e) Probability of Aa is $\dfrac{1}{2}$. So, probability that both children have genotype Aa is $\dfrac{1}{2} \times \dfrac{1}{2} = \dfrac{1}{4} = 0.25$

E2) a) Four different combinations are possible:
i) Three 'head'
ii) Two 'head' and one 'tail'
iii) One 'head' and two 'tail'
iv) Three 'tail'

b) The pattern of distribution of the four possible combinations are
$(3H) : (2H + 1T) : (1H + 2T) : (3T) = 1 : 3 : 3 : 1$

c) The probability of the first combination, i.e., having three 'head' is

$\dfrac{1}{2} \times \dfrac{1}{2} \times \dfrac{1}{2} = \dfrac{1}{8} = 0.125$

The probability of the second combination (2H + 1T) in any one way is

$\dfrac{1}{2} \times \dfrac{1}{2} \times \dfrac{1}{2} = \dfrac{1}{8}$

Since there are three ways of getting this combination, the total
probability is

$$3 \times \frac{1}{8} = \frac{3}{8} = 0.375$$

Similarly, the probability of the third combination (1H and 2T) is also

$$3 \times \frac{1}{8} = \frac{3}{8} = 0.375$$

The probability of the fourth combination (3T) is

$$\frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{8} = 0.125$$

E3) The three characteristics of the genotypic combination of two children from
a cross Aa × aa are

1) There are three possible combinations and the corresponding binomial
   expansion $(p+q)^2$ has three terms.

2) Taking, p = probability of Aa = 1/2 and q = probability of aa = 1/2
   each term of $(p+q)^2 = p^2 + 2pq + q^2$ gives the probability of the
   corresponding combinations, i.e., two Aa, one Aa and one aa
   respectively.

3) The pattern of distribution of the three combinations (two Aa) : (one
   Aa + one aa) : (two aa) = 1:2:1 which is the pattern of the coefficients
   of the corresponding terms of the binomial expansion.

E4) a) A binomial should be used, because tossing of each coin gives only two
       events, 'head' and 'tail'.

    b) The exponent of the binomial is 3, because 3 coins are tossed each time.

    c) i) $3 p^2q$
       ii) q^3
       iii) $3pq^2$
       iv) p^3

       where p = probability of head and q = probability of tail in $(p+q)^3$
       $= p^3 + 3p^2q + 3pq^2 + q^3$.

    d) Probabilities of (i), (ii), (iii) and (iv) in c) above are 0.375, 0.125, 0.375,
       and 0.125 respectively.

E5) a) For tossing of 4 coins, $(p+q)^4$ should be used.

    b) For tossing of 7 coins, $(p+q)^7$ should be used.

E6) a) The pattern of distribution is

       (4H) : (3H + 1T) : (2H + 2T) : (1H + 3T) : (4T) = 1:4:6:4:1

    b) $(p+q)^4 = p^4 + 4p^3q + 6p^2q^2 + 4pq^3 + q^4$

       with p = q = 1/2 the probabilities of various combinations in the
       above order are

       $\frac{1}{16}, \frac{4}{16}, \frac{6}{16}, \frac{4}{16}$ and $\frac{1}{16}$ respectively.

E7) Since there are only two genotypes possible, a binomial should be used.
    Since there are 5 children in each combination, the exponent of the binomial
    should be 5. So the binomial is $(p+q)^5$ where

    p = probability of Aa = 1/2 and

    q = probability of aa = 1/2

    The expansion of the binomial is

    $(p+q)^5 = p^5 + \frac{5!}{4! \, 1!} p^4q + \frac{5!}{3! \, 2!} p^3q^2 + \frac{5!}{2! \, 3!} p^2q^3 + \frac{5!}{1! \, 4!} pq^4 + 5q^5$

a) Since there are 6 terms in the expansion, 6 combinations are possible. There are (5Aa), (4Aa + 1aa), (3Aa + 2aa), (2Aa + 3aa), (1aa + 4aa) and (5aa).

b) The probability of (2Aa + 3aa) is given by the 4th term of the expansion $(p + q)^5$ that is,

$$\frac{5!}{2! \; 3!} \; p^2 q^3 = \frac{5 \times 4 \times 3 \times 2}{2 \times 3 \times 2} \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^3$$

$$= 10 \times \frac{1}{4} \times \frac{1}{8} = \frac{10}{32}$$

c) To find the pattern of distribution, only the coefficients of the terms are to be evaluated:

So, the pattern of the combinations (5Aa): (4Aa + 1aa) : (3Aa + 2aa): (2Aa + 3aa) : (1Aa + 4aa) : (5aa) = 1:5:10:10:5:1.

d) The probability of the term for 5Aa is $p^5 = (1/2)^5 = 1/32$

So, out of 1000 families with 5 children in every family, $1000 \times 1/32 \approx 31$ i.e. families are expected to have all 5 children of genotype Aa.

E8) Taking p = probability of having a boy = 1/2, and q = probability of having a girl = 1/2, $(p + q)^6$ gives the required information.

a) The probability of having all boys is given by the first term of the expansion, i.e., $p^6 = (1/2)^6 = 1/64$

So out of 256 families, $256 \times 1/64 = 4$ families are expected to have all boys.

b) The probability of having 4 boys and 2 girls is given by the term

$$\frac{6!}{4! \; 2!} \; p^4 q^2 = \frac{6 \times 5 \times 4}{4 \times 2} \left(\frac{1}{2}\right)^4 \left(\frac{1}{2}\right)^2 = \frac{15}{64}$$

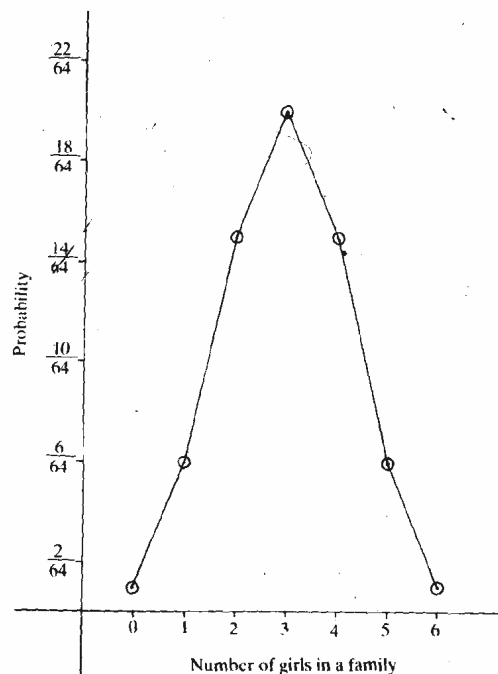So, $256 \times \frac{15}{64} = 60$ families are expected to have 4 boys and 2 girls.

E9

Fig. 3 : Probability distribution of number of girls in a family with six children
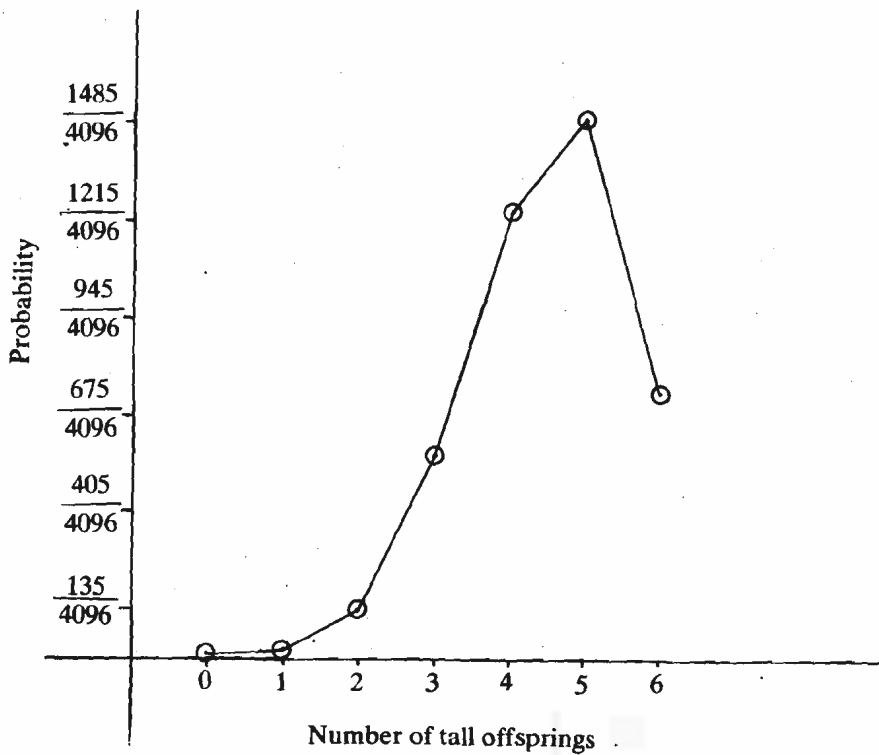
E10)

Fig. 4 Probability distribution of number of tall offsprings in a family with six childrens

a) In Fig. 2, the peak occurs for the value 3 on the x-axis.

b) When $p = 3/4$ and $q = 1/4$, the maximum probability occurs for the value 5 on the x-axis (Fig. 4).

c) The peak is shifted towards the right of the centre.

d) The distribution is not symmetrical because the position of the peak is not at the centre.

E11) In this mating Aa $\times$ Aa as shown in the diagram,

	A	a
A	AA	Aa
a	aA	aa

AA, Aa and aA are tall, aa is short. So, the
probability for tall $= p = 3/4$
probability for short $= q = 1/4$

Number of offsprings in each family $= n = 8$.

a) So, the average number of tall offsprings per family is

$$= np = 8 \times \frac{3}{4} = 6$$

b) The average number of short offsrpings per family is

$$= nq = 8 \times \frac{1}{4} = 2$$

c) Standard deviation $\sqrt{npq} = \sqrt{8 \times \frac{3}{4} \times \frac{1}{4}}$

$$= \sqrt{\frac{3}{2}} = \sqrt{1.50} = 1.22$$

E12) Assuming that if the product $n \times p$ is less than 5, the event is rare,

a) $np = 200 \times 0.05 = 10$

b) $np = 20 \times 0.05 = 1.0$

c) $np = 50 \times 0.04 = 2.0$

d) $np = 160 \times 0.04 = 6.4$

e) $np = 200 \times 0.01 = 2.0$

b), c) and e) are rare events.

E13) i) Number of times one sneezes every day.
   ii) Number of times hailstorms occur every year.
  iii) Number of shooting stars seen every night.
  iv) Number of aeroplanes flying over the village.
   v) Number of marriages taking place in the village every month.
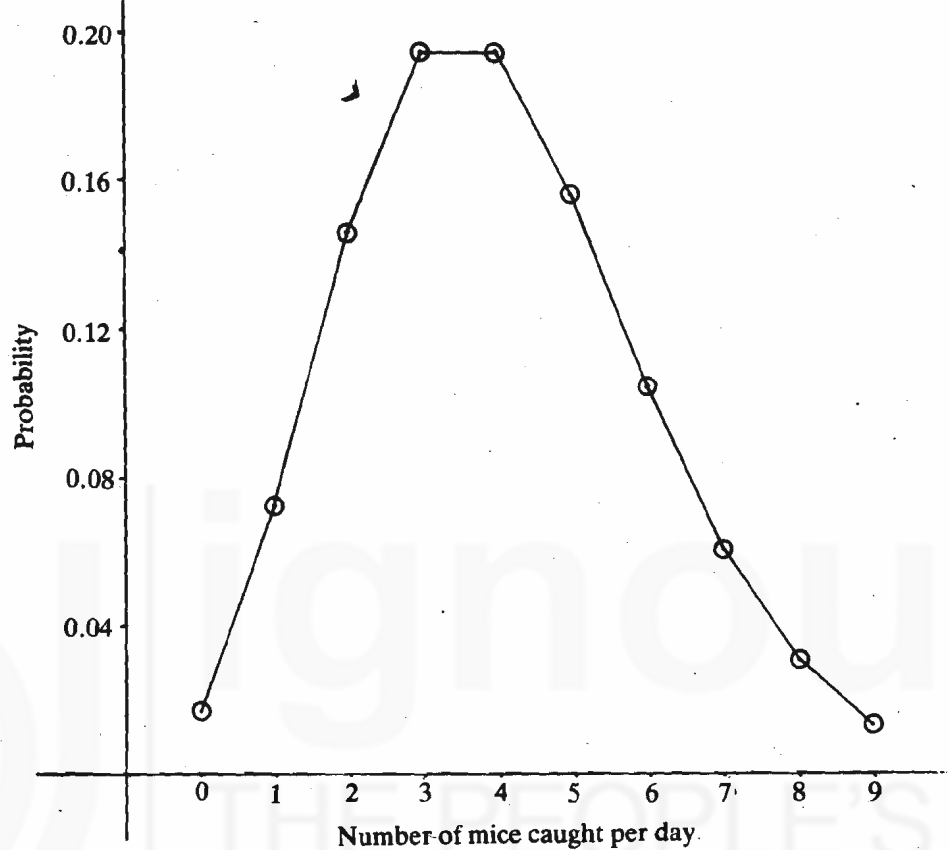
E14) The graph is skewed (see Fig. 5).

Fig. 5 : Probability distribution of the number of mice that a cat catches per day

E15) Here $n = 10$, $p = \dfrac{1}{500}$ therefore

mean $= \mu = np = 10 \times \dfrac{1}{500} = 0.02$ and $e^{-0.02} = 0.9802$.

The probability of 0, 1, 2, 3 defective pens is given by

$P_r = \dfrac{e^{-0.02}\, \mu^r}{r!}$ where $r = 0.\ 1, 2, 3$

The number of packets containing no defective

$= 2000\, P_0 = 1960.4$

Similarly, the number of packets containing one, two and three defectives are 392, 3.9208 $\simeq$ 4, 0.0261 $\simeq$ 0 respectively.

E16) Let the mean of the distribution be 'μ' and total frequency be N.

Frequency corresponding to 2 successes $= \dfrac{e^{-\mu}\mu^2}{2!}\, N$ and

Frequency corresponding to 3 successes $= \dfrac{e^{-\mu}\mu^3}{3!}\, N$

Therefore, we have $\dfrac{N\, e^{-\mu}\mu^2}{2!} = \dfrac{1}{2}\ \dfrac{N\, e^{-\mu}\mu^3}{3!}$ which gives $\mu = 6$

Thus arithmatic mean $= 6$ and standard deviation $= \sqrt{6} = 2.45$ approximately.

Appendix

VALUES OF $e^{-\mu}$

μ	$e^{-\mu}$	μ	$e^{-\mu}$	μ	$e^{-\mu}$	μ	$e^{-\mu}$
.10	.9048	1.90	.1496	3.60	.0273	5.40	.0045
.20	.8187	2.00	.1353	3.70	.0247	5.50	.0041
.30	.7408			3.80	.0224		
.40	.6703	2.10	.1225	3.90	.0202	5.60	.0037
.50	.6065	2.20	.1108	4.00	.0183	5.70	.0033
		2.30	.1003			5.80	.0030
.60	.5488	2.40	.0907	4.10	.0166	5.90	.0027
.70	.4966	2.50	.0821	4.20	.0150	6.00	.0025
.80	.4493			4.30	.0136		
.90	.4066	2.60	.0743	4.40	.0123	6.10	.0022
.100	.3679	2.70	.0672	4.50	.0111	6.20	.0020
		2.80	.0608			6.30	.0018
1.10	.3329	2.90	.0550	4.60	.0101	6.40	.0017
1.20	.3012	3.00	.0498	4.70	.0091	6.50	.0015
1.30	.2725			4.80	.0082		
1.40	.2466	3.10	.0450	4.90	.0074	6.60	.0014
1.50	.2231	3.20	.0408	5.00	.0067	6.70	.0012
		3.30	.0369			6.80	.0011
1.60	.2019	3.40	.0334	5.10	.0061	6.90	.0010
1.70	.1827	3.50	.0302	5.20	.0055	7.00	.0009
1.80	.1653			5.30	.0050		