Chapter - 3

DISCRETE PROBABILITY DISTRIBUTION

Study Strategy and Learning Objectives

Please Remember The Following Reading Strategy and Learning Objectives:

Study Strategy:

1. First, read this section with the limited objective of simply trying to understand the important key terms and concepts:
   probability mass function, probability distribution, cumulative distribution functions, discrete probability distributions, means and variance of discrete random variables and discrete probability distributions.
2. Second, try to understand what they accomplish, why they are needed, and develop the ability to calculate them and select appropriate discrete distribution.
3. Third, learn how to interpret them and how discrete probability models are used in engineering and science.
4. Fourth, read the section once again and try to understand the underlying theory.

You will always enjoy much greater success if you understand what you are doing, instead of blindly applying mechanical steps in order to obtain an answer that may or may not make any sense.

Learning Objectives:

After careful study of this chapter, you should be able to do the following:

1. Determine probabilities from probability mass functions and the reverse.
2. Determine probabilities from cumulative distribution functions and cumulative distribution functions from probability mass functions, and the reverse.
3. Calculate means and variances for discrete random variables.
4. Understand the assumptions for each of the discrete probability distributions presented.
5. Select an appropriate discrete probability distribution to probabilities in specific applications.
6. Calculate probabilities; determine means and variances for each of the discrete probability distributions presented.

3.1 Introduction

Whether an experiment yields qualitative or quantitative outcomes, methods of statistical analysis require that we focus on certain numerical aspects of the data (such as a sample proportion X/n, sample mean $\overline{X}$ or sample standard deviation s). The concept of a random variable allows us to pass from the experimental outcomes themselves to a numerical function of the outcomes. For example: In the inspection of a manufactured product we may be interested in the number of defectives; in the analysis of a road test we may be interested only in the average speed and the average fuel consumption; and in the study of the performance of a rotary switch we may be interested only in its lubrication, electrical current, and humidity. All these numbers are values of random variables. There are two fundamentally different types of random variables - ***discrete random variables*** and ***continuous random variables***. In this chapter, we examine the basic properties and discuss the most important examples of discrete random variables.

3.2 Random Variables (or Stochastic variables)

In this section we discuss random variables, probability distributions, procedures for finding the mean and standard deviation for a probability distribution, and methods for distinguishing between outcomes that are likely to occur by chance and outcomes that are "unusual." We begin with the related concepts of ***random variable*** and ***probability distribution***.

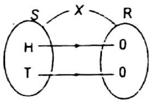
For a given sample space S of some experiment, a random variable (rv) is a function that assigns a numerical value to each possible outcome. Its domain is sample space and range is the set of real numbers. So the value of the random variable is determined by the outcomes of a random experiment.

Definitions: A ***random variable*** is a variable (typically represented by X) that has a single numerical value, determined by chance, for each outcome of a procedure.

A ***probability distribution*** is a graph, table, or formula that gives the probability for each value of the random variable.

Random variables are denoted by capital letters $X, Y, Z, \dots$ and their possible values by lower case letters $x, y, z, \dots$

Examples:



1. If we flip a coin and denote head by 1 and tail by 0 then random variable X takes only two values 1 and 0. In symbol $X(x) = \{x : x = (1,0) \in S)$
2. In toss of two coins, if we define the random variable (X) as the number of heads, the values of X are given below.

Outcomes	*HH*	*HT*	*TH*	*TT*
rv X	2	1	1	0

Similarly in toss of three coins, the values of random variable (X) are given as

Outcomes	*HHH*	*HHT*	*HTT*	*HTH*	*THH*	*THT*	*TTH*	*TTT*
rv (X)	3	2	1	2	2	1	1	0

$X(x) = X(HHH) = 3, X(x) = X(HHT) = 2; \dots, X(x) = X(TTT) = 0$

3. If we measure the height of people, the random variable is
$X(x) = \{x : x \text{ is any real positive number}\}$.
4. If we throw a die, the random variable is $X(x) = \{x : x = 1, 2, 3, 4, 5 \text{ or } 6\}$.
5. The maximum life of electric bulbs is 20000 hours. The random variable is
$X(x) = \{x : 0 \le x \le 2000\}$.
6. If a new type–D flashlight battery has a voltage that is outside certain limits, that battery is characterized as a failure (F); if the battery has a voltage within the prescribed limits, it is a success (S). If we consider the experiment in which batteries were examined until a good one (S) was obtained. Then sample was

$$S = \{S, FS, FFS, \dots\}.$$

Define random variable X by

X = the number of batteries examined before the experiment terminates.

Then $X(S) = 1$, $X(FS) = 2$, $X(FFS) = 3, \dots ; X(FFFFFFS) = 7$ and so on. Any positive integer is possible value of X.

7. Consider the experiment in which a telephone number in a certain area code is dialed using a random number dialer, and define an rv by

$$Y=\begin{cases} 1 \text{ if the selected number is unlisted} \\ 0 \text{ if the selected number is listed in the directory} \end{cases}$$

For example: $Y(4247340) = 0$ if 4247340 appears in the telephone directory, where as $Y(690749) = 1$ tells us that the number 690749 is unlisted.

8. Observe 100 babies to be born in a clinic. The number of boys, which have been born, is a random variable. It may take values from 0 to 100.
9. Number of patients of a clinic daily is a random variable.
10. The weight of babies at birth also is a random variable. It can assume values in the interval, for example, from 800 grams to 6000 grams.

Remarks:

1. The actual value which the event assumes is not a random variable. For example, in toss of three coins the number of heads obtained is a random variable which can take any one of 0, 1, 2 or 3 as long as the coin is not tossed. But after it is tossed and we get two heads, then 2 is not a random variable. So rv X is a real number associated with the outcome of a random experiment.
2. Sometimes we can define and study several different random variables from the same sample space.

3.3 Types of Random Variables:

There are two types of random variables:

(i) Discrete random variable, and (ii) Continuous random variable.

3.3.1 Discrete Random Variables

Many physical systems can be modeled by the same or similar random experiments and random variables. The distribution of the random variables involved in each of these common systems can be analyzed, and the results of that analysis can be used in different applications and examples. In this chapter, we present the analysis of several random experiments and discrete random variables that frequently arise in applications. We often omit a discussion of the underlying sample space of the random experiment and directly describe the distribution of a particular random variable.

Definition: A random variable X defined on a sample space S which assumes only finite or countably infinite set of values, say, $x_1, x_2, x_3, \dots x_m, \dots$ is known as ***discrete random variable***. Such random variable takes only integer (i.e., integral) values.

Examples:

1. A voice communication system for a business contains 48 external lines. At a particular time, the system is observed, and some of the lines are being used. Let the random variable X denote the number of lines in use. Then, X can assume any of the integer values 0 through 48. When the system is observed, if 10 lines are in use, $x = 10$.
2. The number of defective bulbs in a basket.
3. The number of students in a class.
4. The number of interruptions (breakdowns) per day at computer facility.
5. The number of commercial loans processed per day at a bank.
6. The number of telephone calls per unit time etc.

3.3.2 Continuous Random variables

A random variable X, which assumes infinite and uncountable set of values in an interval on the real number line R (or in disjoint union of intervals, e.g. $[0, 10] \cup [20, 30]$). Such a variable is measured experimentally using an interval or ratio scale.

Definitions:

A ***discrete random variable*** has either a finite number of values or a countable number of values, where "countable" refers to the fact that there might be infinitely many values, but they can be associated with a counting process.

A ***continuous random variable*** has infinitely many values, and those values can be associated with measurements on a continuous scale in such a way that there are no gaps or interruptions.

Examples:

1. The height, age or weight of students in a class are continuous *rv*.
2. The maximum life of electric bulbs is 2000 hours. The continuous *rv* is $X(x) = \{x : 0 \le x \le 2000\}$.
3. The duration of a telephone call.
4. The distance required for stopping an automobile travelling at 40 km/hr.
5. Blood pressure measured of *IOE* students in the last medical camp.
6. Range of breakdown voltage of diode. 8. Breaking strength of the cable.
9. Diameter value of a bolt: (1.8 – 2.2) centimeters.
10. Temperature set in a machine: (18 – 22)°C, etc.

3.4 Difference between Discrete and Continuous random variables

1. If we count (cars on a road, death by cancer, tosses until a die shows the first six) we have a discrete random variable and distribution. But if we measure (electric voltage, rainfall, hardness of steel, distance, height, weight, temperature, time length) we have continuous random variable and distribution. So, generally discrete *rv*'s represent counted data while continuous *rv*'s represent measured data.
2. To study basic properties of discrete *rv*'s, only the tools of discrete mathematics–summation and differences are required. The study of continuous *rv*'s require the continuous mathematics of the calculus–Integrals and derivatives.
3. A discrete random variable relates to the numerical events of interest. We can list all possible values of a discrete *rv* and it is meaningful to consider the probability that a particular individual value will be assumed. On the other hand, we cannot list all the values of a continuous *rv* since there is always another possible value between any two of its values. So the meaningful event for a continuous *rv*'s are intervals. The probability that a continuous random variable *X* will assume any particular value is zero.
4. Probability distribution of a discrete *rv* can be presented in a tabular form where as that of continuous *rv* cannot be presented in this form but is has either a formula form or a graphical form.
5. Discrete *rv* takes only integral values whereas continuous *rv* takes any real number in the interval on real line.

Note: If X_1, X_2 are random variables and C, C_1, C_2 are constant then

(a) $CX_1, CX_2, C_1X_2 + C_2X_2, X_1 \pm X_2, X_1X_2$ are also random variables.

(b) max $[X_1, X_2]$ and min $[X_1, X_2]$ are also random variables.

3.5 Discrete Probability Distributions and Probability Mass Functions

Random variables are so important in random experiments that sometimes we essentially ignore the original sample space of the experiment and focus on the probability distribution of the random variable. For example, in Example of a voice communication system for a business containing 48 external lines, our analysis might focus exclusively on the integers {0, 1, . . . , 48} in the range of *X*. In this manner, a random variable can simplify the description and analysis of a random experiment.

The probability distribution of a random variable *X* is a description of the probabilities associated with the possible values of *X*. For a discrete random variable, the distribution is often specified by just a list of the possible values along with the probability of each. In some cases, it is convenient to express the probability in terms of a formula.

Definition: ***Probability mass function (pmf)***

Let *X* be a one dimensional discrete random variable with possible values $x_1, x_2, \ldots$ and the probability of each x_i is

$$P(X = x_i) = P(x_i) = p_i$$

The function $P(x_i)$ is called the ***probability mass function (pmf)*** of discrete random variable *X* if satisfies the following conditions:

$$(i)\ 0 \le P(x_i) \le 0 \text{ for all } i = 1, 2, \ldots$$

$$\text{and} \quad (ii)\ \Sigma P(x_i) = 1.$$

So, *pmf* is responsible for allocating (or distributing) probabilities $P(X = x_i) = P(x_i)$ among the various possible values of *X*.

Remark: If *X* is any discrete random variable, then the probability mass function for *X* is also known as the discrete density function of *X*.

Definition: ***Probability Distribution***

The ***Probability distribution*** of a discrete random variable *X* is a table, graph, or formula that gives the probability of observing each value of *X*

$$P(X = x_i) = P(x_i)$$

and this distribution always satisfies the conditions

$$P(x_i) \ge 0 \text{ and } \Sigma P(x_i) = 1.$$

So probability distribution of a discrete random variable *X* is the set of all ordered pairs $\{x_i, P(x_i)\}$.

Such distribution can be represented in tabular form

X	x_1	x_2	x_3	. . . .	x_n
$P(X = x_i)$	P_1	P_2	P_3	. . . .	P_n

Example 1: If we throw a die, the number we obtain is an example of rv, taking the values 1,2,3,4,5 or 6. The probability distribution is the set of all ordered pairs $\{x, P(x)\}$ i.e.,

$X = x$	1	2	3	4	5	6
$P(X = x)$	1/6	1/6	1/6	1/6	1/6	1/6

Here $P(x) = P(X = x) = 1/6 > 0$ and $\Sigma_{i=1}^{6} P(x_i) = 1$

Example 2: Check whether the following can serve as probability distributions:

(a) $f(x) = (x - 2)/2$ for $x = 1, 2, 3, 4$.

(b) $h(x) = x^2/25$ for $x = 0, 1, 2, 3, 4$.

Solution: (a) $f(1) = 1/2$. Since $f(1)$ is negative the function $f(x)$ cannot serve as probability distribution.

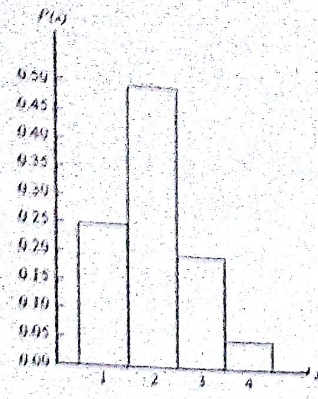
(b) Here $h(0) + h(1) + h(2) + h(3) + h(4) = 6/5$. Since sum of all probabilities is not equal to 1 this function also cannot serve as probability distribution.

Note:

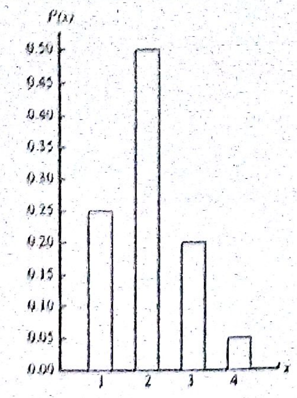
1. Probability distribution of a discrete *rv X*, tells us how the total probability mass of 1 is distributed at various points along the axis of possible values of the *rv X* (where and how much at each *x*).
2. Probability distribution can be represented by means of graphs also. If discrete random variable *X* has *pmf* as shown

$X = x$	1	2	3	4
$P(X=x)$	0.25	0.5	0.2	0.05

Then,



Probability Histogram



Probability bar chart

3.5.1 Cumulative Distribution Function (*cdf*)

Definition: Let X be a one dimensional discrete random variable with *pmf* $P(x)$. Then *cumulative distribution* function (*cdf*) of discrete random variable X denoted by $F(x)$ is defined for any real number x by $F(x) = P(X \le x) = \sum_{x_i \le x} P(x_i)$.

That is, $F(x)$ is the probability that the random variable X takes on a value that will be at most x. It is also known as *probability distribution function* or *distribution function* or *cumulative probability*.

3.5.2 Properties of Cumulative Distribution Function

(1) $0 \le F(x) \le 1$; (2) $F(-\infty) = \lim_{x \to -\infty} F(x) = 0$; (3) $F(+\infty) = \lim_{x \to +\infty} F(x) = 1$

(4) $F(x_1) < F(x_2)$ if $x_1 < x_2$; (5) $P(a < X \le b) = F(b) - F(a)$.

Example 3: Suppose that a day's production of 850 manufactured parts contains parts that do not conform to customer requirements. Two parts are selected at random, without replacement, from the batch. Let the random variable X equal the number of nonconforming parts in the sample. What is the cumulative distribution function of X?

Solution:

The question can be answered by first finding the probability mass function of X.

$$P(X=0) = P(\text{both parts conform}) = \frac{800}{850} \times \frac{799}{849} = 0.886;$$

$P(X=1) = P$(first part selected conforms and the second part selected does not, or the first part selected does not and the second part selected conforms)

$$= 2 \times \frac{800}{850} \times \frac{50}{849} = 0.111$$

$$P(X=2) = P(\text{both parts do not conform}) = \frac{50}{850} \times \frac{49}{849} = 0.003$$

Therefore,

$F(0) = P(X \le 0) = 0.886$

$F(1) = P(X \le 1) = 0.886 + 0.111 = 0.997$

$F(2) = P(X \le 2) = 0.886 + 0.111 + 0.003 = 1$

Note that F(x) is defined for all x from $-\infty < x < \infty$ and not only for 0, 1, and 2.

Example 4: If the random variable X has distribution function

$$F(x) = \begin{cases} 0 & \text{for } x \le 0 \\ 1 - e^{-x^2} & \text{for } x > 0 \end{cases}$$ what is the probability that X exceeds

Solution: The desired probability is

$$P(X > 1) = 1 - P(X \le 1) = 1 - F(1) = e^{-1} = 0.3679$$

Example 5: The discrete *rv* X has *pmf* as shown

$X = x$	1	2	3
$P(X = x)$	1/2	a	1/6

Find: (i) value of 'a' ; (ii) $F(2)$; (iii) $P(1 < X < 3)$; (iv) $P(1 < X \le 3)$ (v) the
(vi) Distribution function of rv X; (vii) $P(1 \le X \le 3)$. Also draw graph of
and $F(x)$.

Solution: (i) Since $\Sigma P(x) = 1$, we have $\frac{1}{2} + a + \frac{1}{6} = 1 \Rightarrow a = \frac{1}{3}$

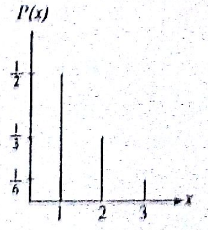
(ii) $F(2) = P(X \le 2) = P(1) + P(2) = \frac{1}{2} + \frac{1}{3} = \frac{5}{6}$. (iii) $P(1 < x < 3) = P(2) = \frac{1}{3}$

(iv) $P(1 < x \le 3) = F(3) - F(1) = \frac{1}{2} + \frac{1}{3} + \frac{1}{6} - \frac{1}{3} = \frac{2}{3}$

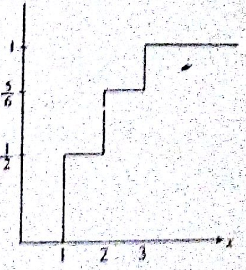
(v) The value of X that has the highest probability is 1. So, the mode = 1

(vi)

X	1	2	3
$P(x)$	1/2	1/3	1/6
$F(x)$	1/2	5/6	1



Graph of $P(x)$



Graph of $F(x)$

Thus distribution function $F(x)$ is a step function.

(vii) $P(1 \le X \le 3) = P(1) + P(2) + P(3) = 1/2 + 1/3 + 1/6 = 1$.

Example 6: A lot of 10 items contains 3 defectives from which a sample of 4 i is drawn without replacement. Le X be the random variable being the numbe defective items in the sample. Find (i) the probability distribution of X, (ii) $P(X<$ and (iii) $P(0 < X < 2)$. (Madras, Oct.B

Solution: (i) Clearly, X takes the values 0,1,2 and 3.

We find the probabilities $P(X = i)$, where $i = 0,1,2,3$.

$$P(X=0) = P(\text{no defective}) = \frac{{}^7C_4}{{}^{10}C_4} = \frac{35}{210} = \frac{1}{6}$$

$$P(X=1) = P(1 \text{ defective}) = \frac{{}^3C_1\,{}^7C_3}{{}^{10}C_4} = \frac{3 \times 35}{210} = \frac{1}{2}$$

$$P(X = 2) = P(2 \text{ defectives}) = \frac{{}^{3}C_{2}\,{}^{7}C_{2}}{{}^{10}C_{4}} = \frac{3 \times 21}{210} = \frac{3}{10}$$

$$P(X = 3) = P(3 \text{ defectives}) = \frac{{}^{3}C_{3}\,{}^{7}C_{1}}{{}^{10}C_{4}} = \frac{1 \times 7}{210} = \frac{1}{30}$$

Hence, the probability distribution of X is given by:

X	0	1	2	3
$f(x)$	1/6	1/2	3/10	1/30

(i) $P(X < 1) = P(X = 0) = \frac{1}{6}$

(ii) $P(0 < X < 2) = P(X = 0) = \frac{1}{6}$

Example 7: If a random variable X takes the values 1,2,3,4 such that
$2P(X = 1) = 3P(X = 2) = P(X = 3) = P(X = 4)$
Find the probability distribution of X.

Solution:

Let, $P(X = 3) = k$. From the given data, it follows that

$$P(X = 1) = \frac{k}{2},\ P(X = 2) = \frac{k}{3} \text{ and } P(X = 4) = \frac{k}{5}$$

Since the probability is equal to 1, we must have

$$P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4) = 1$$

or, $\frac{k}{2} + \frac{k}{3} + k + \frac{k}{5} = 1 \Rightarrow \frac{61k}{30} = 1 \Rightarrow k = \frac{30}{61}$.

Hence, the probability distribution of X is obtained as:

X	1	2	3	4
$f(x)$	15/61	10/61	30/61	5/61

Example 8: A random variable X has the following probability distribution

X	0	1	2	3	4	5	6	7
$f(x)$	0	k	2k	2k	3k	k^2	$2k^2$	$7k^2 + k$

Find, (i) The value of k. (ii) $P(1.5 < X < 4.5 \mid X > 2)$

Solution : (i) Since the total probability is equal to 1,we have

$$\Sigma_{x=0}^{7} P(X = x) = 1$$

or, $0 + k + 2k + 3k + k^2 + 7k^2 + k = 1$ or, $10k^2 + 9k - 1 = 0$

or, $(10k - 1)(k + 1) = 0 \Rightarrow k = \frac{1}{10}$ or, $k = -1$

The value $k = -1$ is inadmissible since probabilities cannot be negative. Hence, $k = 1/10$

(ii) By the definition of conditional probability, we have

$$P(1.5 < X < 4.5 | X > 2) = \frac{P[(1.5 < X < 4.5) \cap (X > 2)]}{P(X > 2)} = \frac{P(2 < X < 4.5)}{1 - P(X \le 2)}$$

Hence, it follows that $P(1.5 < X < 4.5 | X > 2)$

$$= \frac{P(X = 3) + P(X = 4)}{1 - [P(X = 0) + P(X = 1) + P(X = 2)]} = \frac{5k}{1 - 3k} = \frac{5/10}{1-3/10} = \frac{5}{7}$$

Example 9: Let X be a random variable such that
$P(X = -2) = P(X = -1) = P(X = 2)$ and $P(X < 0) = P(X = 0) = P(X < 0)$.
Determine the probability mass function and the distribution of X.

Solution: Let, $P(X = 0) = \alpha$. Then, it follows that

$P(X < 0) = P(X = 0) = P(X < 0) = \alpha$. Since the total probability is equal to 1, we must have, $P(X < 0) + P(X = 0) + P(X > 0) = 1 \Rightarrow 3\alpha = 1 \Rightarrow \alpha = 1/3$

Next, we note that $P(X = 0) = 1/3 = P(X = -2) + P(X = -1) = 2P(X = -2)$

Hence, we have $P(X = -1) = 1/6$ and $P(X = -2) = 1/6$

Similarly, we note that $P(X > 0) = 1/3 = P(X = 1) + P(X = 2) = 2P(X = 2)$

Hence, we have $P(X = 1) = 1/6$ and $P(X = 2) = 1/6$

Thus, the probability distribution of X is given by

x	−2	−1	0	1	2
$f(x)$	1/6	1/6	1/3	1/6	1/6

Also, the distribution function of X is given by

$$F(x) = P(X \le x) = \begin{cases} 0 \text{ if } x < -2 \\ \frac{1}{6} \text{ if } -2 \le x < -1 \\ \frac{1}{3} \text{ if } -1 \le x < 0 \\ \frac{2}{3} \text{ if } 0 \le x < 1 \\ \frac{5}{6} \text{ if } 1 \le x < 2 \\ 1 \text{ if } x \ge 2 \end{cases}$$

3.6 Expected Value and Variance of a Discrete Random Variable

Since a probability distribution for a random variable X is a model for a population relative frequency distribution, we can describe it with numerical descriptive measures, such as its mean and variance.

These two numbers are often used to summarize a probability distribution for a random variable X. The mean is a measure of the center or middle of the probability distribution, and the variance is a measure of the dispersion, or variability in the distribution. These two measures do not uniquely identify a probability distribution. That is, two different distributions can have the same mean and variance. Still, these measures are simple, useful summaries of the probability distribution of X.

3.6.1 Expected Values of Discrete Random Variables

Many frequently used random variables can be both characterized and dealt with effectively for practical purpose by considering of quantities called their expectation. For example, a gambler might be interested in his average winning at a game, a businessman in his average profits on a product, a physicist in the average charge of a particle and so on. The 'average' value of random phenomenon is also termed as its ***mathematical expectation*** or ***expected value***. In this chapter we will define and study this concept value. In this chapter we will define and study this concept about discrete random variables.

Definition: If X is a discrete random which assumes the values $x_1, x_2, x_3, \ldots$ along with corresponding probabilities $P(x_1), P(x_2), P(x_3), \ldots$ then the ***expectation*** or ***expected value*** or ***Mean*** of X denoted by $E(X)$ is defined by

$$\mu = E(X) = \Sigma x_i\, P(x_i).$$

It is also called ***mathematical expectation*** or ***mean*** of X. In other words, the expected value of X is a weighted average of the possible values that X can take on, each value being weighted by the probability that X assumes it.

(i) What is the probability in a given month of selling 3 or more new computer systems?

(ii) What is the probability of selling less than 2 new computer systems ?

(iii) Calculate the expected value and variance of X. [TU, BE, 2058 Shrawan]

Solution: (i) $P(X \geq 3) = P(X = 3) + P(X = 4) + P(X = 5) = 0.30 + 0.20 + 0.08 = 0.58$

(ii) $P(X < 2) = P(X = 0) + P(X = 1) = 0.05 + 0.12 = 0.17$

(iii) $E(X) = \Sigma x \left(\frac{f}{N}\right) = \Sigma x\, P(x)$

$= 0 + 1 \times 0.12 + 2 \times 0.25 + 3 \times 0.30 + 4 \times 0.20 + 5 \times 0.08 = 2.72$

Again, $E(X^2) = \Sigma x^2 . P(x)$

$= 1 \times 0.12 + 4 \times 0.25 + 9 \times 0.30 + 16 \times 0.20 + 25 \times 0.08 = 9.02$

Variance of $X = V(X) = E(X^2) - [E(X)]^2 = 9.02 - [2.72]^2 = 1.6216$

Example 20: What is the expected number of heads appearing when a fair coin is tossed three times?

Solution: Let X denote the number of heads obtained in a random toss of coin three times. The probability distribution of X is

$X = x$	0	1	2	3
$P(X = x)$	1/8	3/8	3/8	1/8

$$\mu = E(X) = \Sigma x.\ P(x) = 0 \times \frac{1}{8} + 1 \times \frac{3}{8} + 2 \times \frac{3}{8} + 3 \times \frac{1}{8} = 1.5$$

Example 21: Let X = the number of cylinders in the engine of the next car to be turned up at a certain facility. The cost of a tune-up is related to X by $h(X) = 20 + 3X + 0.5X^2$. Since X is a random variable, so is $h(X)$; denote this latter rv by Y. The pmf of X and Y are as follows:

$X = x$	4	6	8
$P(x)$	0.5	0.3	0.2

$Y = y$	40	56	76
$P(Y = y)$	0.5	0.3	0.2

Find (i) $E(Y)$; (ii) $E(4X + 2)$; (iii) $V(X)$;

Solution: (i) $E(Y) = E[h(X)] = \Sigma y.\ P(y) = (40)\,(0.5) + (56)\,(0.3) + (76)\,(0.2)$

$= h(4) \times 0.5 + h(6) \times 0.3 + h(8) \times 0.2 = \Sigma h(x).\ P(x) = 52$

(ii) $E(4X + 2) = 4[\Sigma x.\ P(x)] + 2 = 4[4 \times 0.5 + 6 \times 0.3 + 8 \times 0.2] + 2 = 23.6$

(iii) $V(X) = \sigma^2 = \Sigma(x - \mu)^2 . P(x)$

$= (4 - 5.4)^2\,(0.5) + (6 - 5.4)^2\,(0.3) + (8 - 5.4)^2\,(0.2) = 2.44$

where $\mu = \Sigma x.\ P(x) = 4 \times 0.5 + 6 \times 0.3 + 8 \times 0.2 = 54.$

So, standard deviation of $X = \sqrt{2.44} = 1.562$

Example 22: The probability that there is at least one error in an accounts statement prepared by A, B and C are 0.2, 0.25 and 0.4 respectively. A, B and C prepared 10, 16 and 20 statements respectively. Find the expected number of correct statement in all.

Solution:

Let p_1 = Prob. that there is no error in an accounts statement prepared by A.

= 1 – the prob. that there is at least one error in an accounts statement prepared by A

$= 1 - 0.2 = 0.8.$

Similarly $p_2 = 1 - 0.25 = 0.75$ and $p_3 = 1 - 0.40 = 0.6.$

Let X denote the statements prepared by A, B and C. Then the probability distribution of statements prepared by them is as follows:

$X = x$	10	16	20
$P(X = x)$	0.8	0.75	0.6

$\mu = E(X)$ = the expected number of correct statements in all

$= \Sigma x.\ P(x) = x_1p_1 + x_2p_2 + x_3p_3 = 10 \times 0.8 + 16 \times 0.75 + 20 \times 0.6 = 32.$

3.7 Some Discrete Probability Distributions

As random variables can be either discrete or continuous, we have distributions that are either discrete or continuous. These probability di theoretical distributions provide a relatively easy method for ob probabilities for a random variables. Conceptually these can be co models, which can be used to describe situations that involve outcomes go chance. Here we deal with some more commonly used discrete distributions such as

1. Bernoulli Probability Distribution
2. Binomial Probability Distribution
3. Hypergeometric Probability Distribution
4. Poisson Probability Distribution
5. Negative Binomial Probability Distribution

3.8 Bernoulli Probability Distribution

3.8.1 Introduction

A Bernoulli's trial is an experiment having only two possible outcomes, success or failure. If X is a random variable that takes the value 1 when the Be trial is a success and the value 0 when the Bernoulli trial is a failure, then X is Bernoulli random variable. It originates from the experiment consisting of Bernoulli trial. Examples of Bernoulli trials are: Tossing of a coin (head or firing a target (hit or miss), playing a game (win or lose) etc.

3.8.2 Characteristics (or Model) of Bernoulli random variable and Distribution

1. Only one trial is considered.
2. In the trial, there are only two possible outcomes, that is, success or fai other words, the results of the trial are always dichotomous.
3. The probability of success p remains fixed for the trial.
4. The probability of failure q remains fixed for the trial and $q = 1 - p$.
5. The random variable X is the occurrence of a success in the trial.

<u>**Definition**</u>: ***The probability mass function (pmf)*** of X is given by

$$P(X = x) = \begin{cases} p, & \text{if } x = 1 \\ q, & \text{if } x = 0 \end{cases}$$

where p = probability of success in one trial.

q = probability of failure in one trial and $p + q = 1, q = 1 - p$

Bernoulli distribution: The distribution of a random variable X is sai ***Bernoulli distribution*** with parameter 'p', if its probability mass functio is given by

$$P(X = x) = \begin{cases} p, & \text{if } x = 1 \\ q, & \text{if } x = 0 \end{cases}$$

Mean and Variance of Bernoulli *rv X*

Mean $(\mu) = E(X) = \Sigma x.\ P(x) = 1.\ P(X = 1) + 0 \times P(X = 0) = 1 \times p + 0 \times q$

Variance $= V(X) = E(X^2) - [E(X)]^2 = 0^2 \times P(X = 0) + 1^2 \times P(X = 1) -$

Examples:

1. If we flip a coin, it will show either head or tail on the upper face.
2. The sex of an expected baby will either be male or female exceptions).
3. A manufactured item will either be defective or non-defective.

4. When a student attempts to log on to a computer time - sharing system, either all ports are busy (B), in which case the student will fail to obtain access, or else there is at least one port free (F), in which case the student will be successful in accessing the system.
   With $S = \{B, F\}$, define an *rv* X by $X(B) = 0$, $X(F) = 1$
   The *rv* X indicates whether (1) or not (0) the student can log on.
5. Firing a target (hit or miss), classifying a worker (skilled or unskilled).

$$= p - p^2 = p(1 - p) = pq.$$

Example 23: A die is rolled. The probability that an even number will come is 0.7. Find the mean and the standard deviation of an occurrence of an even number.

Solution: p = Prob. of an even number = 0.7

q = Prob. of an odd number = $1 - 0.7 = 0.3$

Let X denote the occurrence of an even number.

$\mu = \Sigma x.\, P(x) = 0 \times q + 1 \times p = p = 0.7$

$\sigma^2 = V(X) = pq = 0.7 \times 0.3 = 0.21 \quad \Rightarrow \sigma = \sqrt{0.21} = 0.4583$

3.9 Binomial Probability Distribution

3.9.1 Introduction

This distribution was discovered by *James Bernoulli* (1654-1705). Binomial probability distributions are important because they allow us to deal with circumstances in which the outcomes belong to two relevant categories i.e., Bernoulli trials such as *acceptable / defective products* or *yes / no* responses in a survey question or *pass / fail* or *sick / well* or *fill / infill* etc.

Consider the following random experiments and random variables:

1. Flip a coin 10 times. Let X = number of heads obtained.
2. A worn machine tool produces 1% defective parts. Let X= number of defective parts in the next 25 parts produced.
3. Each sample of air has a 10% chance of containing a particular rare molecule. Let X= the number of air samples that contain the rare molecule in the next 18 samples analyzed.
4. Of all bits transmitted through a digital transmission channel, 10% are received in error. Let X= the number of bits in error in the next five bits transmitted.
5. A multiple choice test contains 10 questions, each with four choices, and you guess at each question. Let X = the number of questions answered correctly.
6. In the next 20 births at a hospital, let X the number of female births.
7. Of all patients suffering a particular illness, 35% experience improvement a particular medication. In the next 100 patients administered the medication, let X= the number of patients who experience improvement.

These examples illustrate that a general probability model that includes these experiments as particular cases would be very useful. Each of these random experiments can be thought of as consisting of a series of repeated, random trials: 10 flips of the coin in experiment 1, the production of 25 parts in experiment 2, and so forth. The random variable in each case is a count of the number of trials that meet a specified criterion. The outcome from each trial either meets the criterion that X counts or it does not; consequently, each trial can be summarized as resulting in either a success or a failure. For example, in the multiple choice experiment, for each question, only the choice that is correct is considered a success. Choosing any one of the three incorrect choices results in the trial being summarized as a failure.

<u>Please Remember</u>: The word success as used here is arbitrary and does not necessarily represent something good. Either of the two possible categories may be called the *success S* as long as its probability is identified as p. Once a category has been designated as the success S, be sure that p is the probability of a success and X is the number of *successes*. That is, ***be sure that the values of p and X refer to the same category designated as a success.***

So, the terms success and failure are just labels. Unfortunately, the usual labels can sometimes be misleading. In experiment 2, because X counts defective parts, the production of a defective part is called a success.

A trial with only two possible outcomes is used so frequently as a building block of a random experiment that it is called a *Bernoulli trial.* It is usually assumed that the trials that constitute the random experiment are *independent*. This implies that the outcome from one trial has no effect on the outcome to be obtained from any other trial. Furthermore, it is often reasonable to assume that the probability of a success in each trial is *constant.*

3.9.2 Conditions of using Binomial Distribution

The binomial distribution has the following requirements:

1. The number of Bernoulli trials n should be finite & fixed.
2. Each trial has only two possible outcomes, success or failure, lose or win, head or tail, defective or non defective etc.
3. Probability of a success (or failure) remains same for each trial.
4. The trials are identical and independent.
5. The sum of probabilities of success and failure is one.
6. The probability of success *vis-a-vis* failure is not very low.
7. Random variable X denotes a number of successes in n trials.

Any experiment which satisfies the above conditions is called a ***Binomial experiment.*** If a procedure satisfies these requirements, the distribution of the random variable X is called a ***binomial probability distribution*** (or ***binomial distribution***).

Suppose that n independent trials, each of which results in a "***success***" with probability p and in a "***failure***" with probability $1 - p$, are to be determined.

<u>**Definitions**</u>:

If X represents the number of success that occurs in the n trials, then X is said to be a ***binomial random variable*** with parameters n and p. Its ***probability mass function*** with parameters n and p is given by

$$b(x;\, n, p) = P(X = x) = \begin{cases} {}^nC_x p^x (1 - p)^{n-x} & \text{for } x = 0,\ 1,\ 2, \ldots, n. \\ 0 & \text{otherwise} \end{cases}$$

where $P(X = x)$ = probability of getting exactly 'x' successes in n trials.

n = number of trials, p = probability of success in one trial.

$1 - p = q$, probability of failure in one trial.

Binomial Distribution:

A discrete random variable X is said to have ***binomial distribution*** with parameters n and p if (i) it assumes only non-negative values i.e. $0, 1, 2, 3, \ldots, n$ and

(ii) its probability mass function is given by

$$b(x;\, n, p) = P(X = x) = p(x) = \begin{cases} {}^nC_x p^x (1 - p)^{n-x} & \text{for } x = 0,\ 1,\ 2, \ldots, n. \\ 0 & \text{otherwise} \end{cases}$$

Notation for Binomial Probability Distributions:

S and F (success and failure) denote the two possible categories of all outcomes; p and q will denote the probabilities of S and F, respectively, so

$P(S) = p$ (p = probability of a success)

$P(F) = 1 - q = q$ (q = probability of a failure)

n denotes the fixed number of trials.

X denotes a specific number of successes in n trials, so X can be whole number between 0 and n, inclusive.

p denotes the probability of success in one of the n trials.

q denotes the probability of failure in one of the n trials.

$P(x)$ denotes the probability of getting exactly x successes among the n trials.

$X \sim B(n, p)$ denotes that the random variable X follows Binomial distribution with parameters n and p.

Mean and variance of Binomial Distribution

If $X \sim B(n, p)$, then mean and variance of X are by

$$\textit{Mean} = \mu = E(X) = \Sigma_{x=0}^{n} xP(X = x) = \Sigma_{x=0}^{n} x\,{}^nC_x p^x q^{n-x} = np$$

$$\textit{Variance} = \sigma^2 = V(X) = E(X^2) - [E(X)]^2 = npq.$$

$$\textit{Standard Deviation}\ (s.d.) = \sigma = \sqrt{npq}.$$

3.9.3 Properties of Binomial Distribution

If $X \sim B(n, p)$ then it has following properties

1. It has two parameters n and p.
2. It is discrete probability distribution and the number success assumes the value 0,1,2,3, ... n where n is finite.
3. Its mean and variance are $\mu = np$, $\sigma^2 = npq$. (*iv*) Mean > Variance.
4. It is symmetrical if $p = 0.5 = q$
5. It is skewed to the left (i.e. negatively skewed) if $p > 0.5$ i.e., $p > q$.
6. It is skewed to the right (i.e. positively skewed) if $p < 0.5$ i.e., $p < q$.
7. Binomial distribution tends to normal distribution as n increases. The normal approximation is correct enough if np is greater than 15 for $p = 0.5$.
8. If n and p are known, the binomial distribution is completely determine

Parameters of the Distribution: The constants used in the distribution are called ***parameters of the distribution***. There are only two parameters n and p (or q) in binomial distribution.

Remarks:

1. In practical we use Binomial Distribution if $p \geq 0.5$ and $n \leq 20$.
2. The distribution given by $P(X = x) = {}^nC_x p^x q^{n-x}$ is called binomial distribution since it is the $(x+1)^{th}$ term in the binomial expansion of $(p + q)^n$
3. For $n = 1$, Binomial distribution becomes the Bernoulli distribution.
4. It was discovered by James Bernoulli. It is also called Bernoulli Distribution (by some statisticians)

3.9.4 Cumulative Binomial Distribution Function (*cdf*)

If $X \sim B(n, p)$, then binomial distribution function (*cdf*) is defined by

$$B(x ; n, p) = P\{X \leq x\} = \Sigma_{k=0}^{n}\ {}^nC_k p^k (1-p)^{n-k}$$

$$= \Sigma_{k=0}^{n}\ {}^nC_k p^k q^{n-k} = \Sigma_{k=0}^{n}\ b(k ; n, p) \text{ for } x = 0, 1, 2, \ldots, n.$$

So, $B(x ; n, p) = \Sigma_{k=0}^{n}\ b(k ; n, p)$ for $x = 0,1,2, \ldots, n$.

Note: Cumulative binomial probabilities $B(x; n, p)$ and $B(x-1; n, p)$ differ by the single term $b(x ; n, p)$; i.e., $b(x ; n, p) = B(x ; n, p) - B(x-1; n, p)$

Applications

It is used extensively in statistical quality control and acceptance sampli classify a lot as defective or non-defective; medical application as success or failu surgery, *cure* or *non-cure* of a patient; military application as hit or miss of target e

Example 24: Test for impurities commonly found drinking water from private showed that 30% of all wells in a particular have impurity A. If a random sampl 5 wells is selected from the large number of wells in the country, what is probability that: a) Exactly 3 will have impurity A?

b) At least 3? c) Fewer than 3? [TU, BE, 2067 Mang

Solution: First we confirm that this experiment possesses the characteristics binomial experiment. This experiment consists of $n = 5$ trials, one correspondin each random selected well. Each trial results in an S (the well contains impurity A an F (the well does not contain impurity A). Since the total number of wells in country is large, the probability of drawing a single well and finding that it con impurity A is equal to 0.30 and this probability will remain the same for eacn of 5 selected wells. Further, since the sampling is random, we assume that the outc on any one well is unaffected by the outcome of any other and that the trials independent. Finally, we are interested in the number x of wells in the sample of 5 that contain impurity A. Therefore, the sampling process represents a binon experiment with $n = 5$ and $p = 0.30$.

Let X denote the number of wells that have inpurity A.

(a) The probability of drawing exactly $x = 3$ wells containing inpurity A is

$$P(3) = C(5, 3)\ (0.30)^3\ (0.70)^{5-3} = 0.1323$$

(b) The probability of observing at least 3 wells containing impurity A is

$$P(X \geq 3) = P(3) + P(4) + P(5) = 0.1323 + 0.0284 + 0.00243 = 0.1638$$

(c) Although $P(X < 3) = P(0) + P(1) + P(2)$, we can avoid calculating 3 probabilitie by using the complementary relationship

$$P(X < 3) = 1 - P(X \geq 3) = 1 - 0.16380 = 0.8369.$$

By using binomial table

$$P(X < 3) = P(0) + P(1) + P(2) = \Sigma_{x=0}^{2}\ b(x ; 5, 0.30) = 0.8369.$$

Example 25: A box contains 100 transistors, 20 out of which are defective. If transistors are selected for inspection, find the probability that (*a*) all 10 defectives; (*b*) all 10 are goods; (*c*) at least one is defective; (*d*) at most 3 defective?

Solution: Total number of transistors = 100, Number of defective transistors = 20

p = Prob. of success (of defective transistors) = 20/100 = 0.20

q = Prob. of failure (or non-defective transistors) = 0.80

n = Number of transistors selected in the sample = 10

Let X denote the number of defective transistors. Then by Binom distribution the probability of getting 'x' defective transistors is given by

$$P(X = x) = C(n, x)\, p^x q^{n-x} ; x = 0,1,2, \ldots, n$$

$$= C(10, x)\ (0.20)^x\ (0.80)^{10-x} ; x = 0,1,2, \ldots, 10.$$

(a) Probability of all defective transistors is

$$P(X = 10) = C(10,10)\ (0.20)^{10}\ (0.80)^{10} = 1 \times (0.20)^{10} \times 1 = 0.00000001024.$$

(b) Probability of all 10 good transistors is

$$P(X = 0) = C(10, 0)\ (0.20)^0\ (0.80)^{10-0} = 0.1074.$$

(c) Probability of at least one defective transistor is

$P(X \geq 1) = 1 - P(X < 1) = 1 - P(X = 0) = 1 - 0.1074 = 0.8926$

(d) Probability of at most three defective transistors is

$P(X \leq 3) = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3)$

$= 0.1074 + 0.2684 + 0.302 + 0.2013 = 0.8791$

Or, using table $P(X \leq 3) = B(3;\ 10, 0.2) = \Sigma_{x=0}^{n}\ b(x;\ 10, 0.2) = 0.8791$

Example 26: If 20% of the bolts produced by machine are defective, determine the probability that out of 4 bolts chosen at random

(i) exactly one bolt will be defective; (ii) No defective bolt;

(iii) less than 2 bolts will be defective. *[TU, BE, 2061 Ashwin/2067 Mangsir]*

Solution: Let X denote the number of detective bolts.

Here, $n = 4, p = 20\% = 0.2, q = 0.8$

Required probability is given by $P(x) = {}^{n}C_{x}\, p^{x}\, q^{n-x}$

So, (i) $P(1) = {}^{4}C_{1}(0.2)^{1}\ (0.8)^{4-1} = 0.4096$

(ii) $P(0) = {}^{4}C_{0}\ (0.2)^{0}\ (0.8)^{4-0} = 0.4096$; (iii) $P(X < 2) = P(0) + P(1) = 0.82$

Example 27: If 75% of all purchases at a certain store are made with a credit card and X is the number among ten randomly selected purchases made with credit card, find $E(X)$ and $V(X)$.

Solution: Here $X \sim B(10, 0.75)$. So, $E(X) = np = (10)\ (0.75) = 7.5$;

$V(X) = npq = 1.875.$

Even though X can take on only integer values, $E(X)$ need not be an integer. If we perform a large number of independent binomial experiments, each with $n = 10$ trials and $p = 0.75$, then the average number of success per experiment will be close to 7.5.

Example 28: If the probability is 0.05 that a certain wide - flange column will fail under a given axial load, what is the probability that among 16 such columns (a) at most two will fail; (b) at least four will fail ?

[TU, BIE 2066 Magh/2062 Bhadra/2068 Magh]

Solution: p = probability of success = 0.05; q = probability of failure = 0.95

n = number of columns = 16

(a) $P(X \leq 2) = B(2;\ 16, 0.05) = \Sigma_{x=0}^{n}\ b(x\ ;\ 16, 0.05) = 0.9771$ [Using Table]

(b) $P(X \geq 4) = \Sigma_{x=4}^{16}\ b(x;\ 16, 0.05) = 1 - \Sigma_{x=0}^{3}\ b(x\ ;\ 16, 0.05)$

$= 1 - B(3;\ 16, 0.05) = 1 - 0.9930 = 0.0070.$

Example 29: If the probability is 0.20 that any person will dislike the taste of a new toothpaste, what is the probability that 5 of 18 randomly selected persons will dislike it?

Solution: Using table $P(X = 5) = b(5;\ 18, 0.20) = B(5; 18, 0.20) - B(4; 18, 0.20)$

$= 0.8671 - 0.7164 = 0.5107.$

Example 30: (*A binomial probability to guide decision making*)

A manufacturer of external hard drives claims that only 10% of his drives requires repair within the warranty period of 12 months. If 5 of 20 of his drives required repair within the first year does this tend to support or refute the claim?

Solution: Here, Total drives = 20

p = Prob. that anyone will require repairs within a year = 0.10

Let us first find the probability that 5 or more of 20 of the hard drives will require repairs within a year. Using Table, we get

$$\Sigma_{x=5}^{20}\ b(x;\ 20, 0.10) = 1 - B(4; 20, 0.10) = 1 - 0.9568 = 0.0432$$

Since this probability is very small, it would seem reasonable to reject the hard drives manufacturer's claim.

Example 31: The probability that the noise level of a wide–band amplifier will exceed 2 *dB* is 0.05. Find the probability that among 12 such amplifiers the noise level of (a) one will exceed 2 *dB*; (b) at most two will exceed 2 *dB*;

(c) two or more will exceed 2 *dB*. *[TU, BE, 2063 Kartik]*

Solution: Given: $p = 0.05, n = 12$

(a) $P(X = 1) = b(1; 12, 0.05) = B(1;\ 12, 0.05) - B(0; 12, 0.05)$ [Using Table]

$= 0.8816 - 0.5404 = 0.3412$

Or, $P(X = 1) = C(12,\ 1)(0.05)^{1}(0.95)^{12-1} = 0.3412$

(b) $P(X \leq 2) = B(2; 12.0.05) = \Sigma_{x=0}^{2}\ b(x;\ 12, 0.05) = 0.9804$

(c) $P(X \geq 2) = \Sigma_{x=2}^{12}\ b(x;\ 12, 0.05) = 1 - \Sigma_{x=0}^{1}\ b(x;\ 12, 0.05) = 1 - B(1; 12, 0.05)$

$= 1 - 0.8816 = 0.1184$ [Using Table]

✓ **Example 32:** 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 at least one third in (a) four of five installations?

(b) at least four of five installations.

Solution: Given : $p = 0.60, n = 5$

(a) $P(X = 4) = b(4; 5, 0.60) = C(5,4)(0.60)^{4}(0.40)^{5-4} = 0.259$

(b) $P(X \geq 4) = \Sigma_{x=4}^{5}\ b(x;\ 5, 0.60) = 1 - \Sigma_{x=0}^{3}\ b(3;\ 5, 0.60)$

$= 1 - 0.6630 = 0.337$ [*Using Table* A-1]

Example 32: If 90% of all students taking a beginning computer course fail to get their first program to run on first submission, what is the probability that among 15 randomly chosen such students *[TU, BE, 2063 Kartik]*

(a) at least 12 will fail on first submission;

(b) between 10 and 13 inclusive will fail on the first submission?

Solution: Given: $p = 0.90, n = 15$

(a) $P(X \geq 12) = \Sigma_{x=2}^{15}\ b(x;\ 15, 0.90) = 1 - \Sigma_{x=0}^{11}\ b(x;\ 15, 0.90)$

$= 1 - 0.0556 = 0.9444$ [Using Table]

(b) $P(10 \leq X \leq 13) = \Sigma_{x=10}^{13}\ b(x;\ 15, 0.90) = \Sigma_{x=0}^{13}\ b(x;\ 15, 0.90) - \Sigma_{x=0}^{9}\ b(x;\ 15, 0.90)$

$= 0.4510 - 0.0022 = 0.4488$ [Using Table]

Example 34: An agricultural cooperative claims that 90% of the watermelons shipped out are ripe and ready to eat. Find the probability that among 18 watermelons shipped out (a) all 18 are ripe and ready to eat;(b) at least 16 are ripe and ready to eat; (c) at most 14 are ripe and ready to eat. *[TU, BE.2068 Magh(Back)]*

Solution: Given : $p = 0.90, n = 18, q = 0.10.$

Let X denote the number of watermelons shipped out that are ripe and ready to eat.

(a) $P(X = 18) = b(18;\ 18, 0.90) = C(18, 18)(0.90)^{18}(0.10)^{0} = 0.1501$

(b) $P(X \geq 16) = \Sigma_{x=16}^{18}\ b(x;\ 18, 0.90) = 1 - \Sigma_{x=0}^{15}\ b(x;\ 18, 0.90)$

$= 1 - 0.2262 = 0.7338$ [Using Table 1]

(c) $P(X \leq 14) = \sum_{x=0}^{14}\ b(x;\ 18, 0.90) = 0.0982$ [Using Table]

Example 35: If on an average one vessel in every ten is wrecked, find the probability that out of 5 vessels, at least 4 vessels expected to arrive safely.

Solution: Given: q = Probability of vessel is wrecked (or failure) = 1/10 = 0.10

p = Prob. of vessel arrives safely (or success) = 9/10 = 0.90

n = Number of vessels in sample = 5

Let X denote the number of vessel that arrives safely. Then probability that out of 5 vessels, at least 4 vessels will arrive safely is

$P(X \geq 4) = \Sigma_{x=4}^{5}\ (x;\ 5, 0.90) = 1 - \Sigma_{x=0}^{3}\ b(x;\ 5, 0.90)$

$= 1 - 0.0815 = 0.9185$ [Using Table]

Example 36: A student selects his answers on a true/ false examination by tossing a coin (so that any particular answer has a 0.50 probability of being correct). He must answer at least 70% correctly in order to pass. Find the probability of passing when the number of question is (i) 10; (ii) 20.

Solution: Here $p = 0.50$, $q = 0.50$. Let X denote no. of questions.

(i) n = number of questions in the sample = 10

$$P(X \geq 70\% \text{ of } 10) = P(X \geq \frac{70}{100} \times 10) = \Sigma_{x=7}^{10} b(x; 10, 0.50)$$

$$= 1 - \Sigma_{x=0}^{6} b(x; 10, 0.50) = 1 - 0.8281 = 0.1719 \quad \text{[Using Table]}$$

(ii) $n = 20$, so $P(X \geq 70\% \text{ of } 20) = P(X \geq 14) = \Sigma_{x=14}^{20} b(x; 20, 0.50)$

$$= 1 - \Sigma_{x=0}^{13} b(x; 20, 0.50) = 1 - 9423 = 0.0577 \quad \text{[Using Table]}$$

Example 37: The normal rate of infection of a certain disease in animal is known to be 25%. In an experiment with 6 animals injected with a new vaccine it was observed that none of the animals caught infection. Calculate the probability of the observed result.

Solution: p = Prob. of animal being infected (or success) = 0.25

q = Prob. of animal not being infected (or failure) = 0.75

n = Number of animals injected = 6

Let X denote the number of animals caught infection. So, the probability that none of the animals caught infection is $P(X = 0) = C(6,0)(0.25)^0(0.75)^6 = 0.1780$

Example 38: In a precision bombing attack there is a 50% chance that any one bomb will strike the target. One direct hit is required to destroy the target completely. How many bombs must be dropped to give a 99% change or better of completely destroying the target?

Solution: p = Prob. of bomb hitting the target (or success) = 0.50

q = Prob. of bomb not hitting the target (or failure) = 0.50

Since one direct hit is required to destroy the target completely, the probability that the target is destroyed completely is the probability that at least one bomb hitting the target.

Let X denote the number of bomb hitting the target.

By question: $P(X \geq 1) = 0.99 \quad \Rightarrow \quad 1 - P(X < 1) = 0.99 \quad$ or, $1 - P(0) = 0.99$

or, $\quad 1 - C(n, 0)(0.50)^0(.050)^{n-0} = 0.99 \quad \Rightarrow \quad (0.50)^n = 0.01$

Taking *log* on both sides we get $\quad n \log(0.50) = \log(0.01)$

or, $\quad n = \dfrac{\log(0.01)}{\log(050)} = \dfrac{-2}{-0.301} = 6.64 \simeq 7$

Example 39: Suppose a warship takes 10 shots at a target, and it takes at least 4 hits to sink it. If the warship has a record of hitting with 20% of its shots in the long run, what is the probability (or chance) of sinking the target?

Solution: p = Prob. of hitting the target (or success) = 0.20

q = Prob. of not hitting the target (or failure) = 0.80

n = Number of shots targeted = 10

Let X denote number of shots hitting the target. According to question, in order to sink the target at least 4 hits are required; therefore the prob. of sinking the target is given by $\quad P(X \geq 4) = 1 - P(X < 4)$

$$= 1 - [P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3)]$$

$$= 1 - [0.1074 + 0.2684 + 0.302 + 0.2013] = 0.1209$$

Or, $\quad$ Using Table $P(X \geq 4) = 1 - P(X < 4) = 1 - \sum_{x=0}^{3} b(x; 10, 0.20) = 0.1209$

Example 40: Multiple–choice test consists of 10 questions and 4 answers question. If each question is answered by shuffling 4 tags tabled 1,2, 3 and 4 one and making the alternative whose number is drawn. Find the probability getting 3; (b) at least one; (c) at least 7; (d) at most 3 of these questions correctly.

Solution: Let X be a *rv* which denote no. of questions answered correctly. To of questions $(n) = 10$

In multiple choice, one answer is correct out of 4 answers.

So p = Prob. of correct answering in each question = $1/4 = 0.25$; $q = 1 - 1/4 = 0.75$

(a) Prob. of getting 3 of these questions answered correctly is

$$P(X = 3) = C(10,3)\left(\frac{1}{3}\right)^3\left(\frac{3}{4}\right)^7 = 0.2502$$

(b) $P(X \geq 1) = 1 - P(X < 1) = 1 - P(X = 0) = 0.\ 9436$

(c) $P(X \geq 7) = 1 - P(X \leq 6) = 1 - \sum_{x=0}^{6} b(x; 10, 0.25) = 1 - 0.9965 = 0.0035$

(d) $P(X \leq 3) = B(3; 10, 0.25) = \sum_{x=0}^{3} b(x; 10, 0.20) = 0.7759 \quad$ [Using Table]

Example 41: A man tosses a fair coin 10 times. Find the probability that he have (a) heads on the first 5 tosses and tails on the next 5 (b) heads on tosses 1, 3, 5, 7, 9 and tails on tosses 2, 4, 6, 8, 10; (c) 5 heads tails; (d) at least 5 heads.

Solution: Total no. of coin tosses $(n) = 10$

p = Prob. of getting head in each toss = $1/2 = 0.5$, $q = 0.5$.

(a) Prob. of getting heads on first 5 tosses and tails on next 5 tosses

$= p^5 q^{n-5} = (0.5)^5(0.5)^5 = 0.000977$

(b) $P(HT\,HT\,HT\,HT\,HT) = P(H)P(T)P(H)P(T)P(H)P(T)P(H)P(T)P(H)P(T)$

$= (0.5)^{10} = 0.000977$

(c) The probability of getting 5 heads and the probability of getting 10–5 tails equal i.e. $b(x; n, p) = b(n-x\,; n, q)$

So, prob. of getting 5 heads = $P(X = 5) = C(10,5)\,p^5 q^{10-5} = 0.2461$

(d) $P(\text{at least 5 heads}) = P(5H \text{ and } 5T) + P(6H \text{ and } 4T) + P(7H \text{ and } 3T)$
$+ P(8H \text{ and } 2T) + P(9H \text{ and } 1T) + P(10H)$

$\therefore \quad P(X \geq 5) = P(X = 5) + P(X = 6) + P(X = 7) + P(X = 8) + P(X = 9) + P(X = 10) = 0.623$

Or, $\quad$ Using Table

$P(X \geq 5) = \Sigma_{x=5}^{10} b(x; 10, 0.5) = 1 - \Sigma_{x=0}^{4} b(x; 10, 0.5) = 1 - 0.3770 = 0.623$

Example 42: Out of 1000 families with 4 children each, what percentage would expect to have (a) 2 boys and 2 girls; (b) 1 boy and 3 girls; (c) at least one boy; no girls; (e) at most 2 girls; Assume equal probabilities for boys and girls.

Solution: $p = q = 1/2 = 0.50$, $n = 4$

Let X denote the number of boys in the family with 4 children each.

(a) $P(2 \text{ boys and } 2 \text{ girls}) = P(X = 2) = C(4,2)(0.5)^2(0.5)^2 = 0.375$

Hence, percentage of families expected to have 2 boys and 2 girls
$= 0.375 \times 100\% = 37.5\%$

(b) $P(1 \text{ boy and } 3 \text{ girls}) = P(X = 1) = C(4,1)(0.5)^1(0.5)^3 = 0.25\%$

Required percentage = $0.25 \times 100\% = 25\%$

(c) $P(\text{at least one boy}) = P(X \geq 1) = 1 - (X = 0) = 1 - C(4,0)(0.5)^0(0.5)^4 = 0.9375$

Required percentage = $0.9357 \times 100\% = 93.75\%$

(d) $P(\text{no girls}) = 1 - P(\text{at least one boys}) = 1 - 0.9357 = 0.0625$

Required percentage = $0.0625 \times 100\% = 6.25\%$

(e) P(at most 2 girls) = 1 − P(at least 3 boys) = $1 - P(X \geq 3) = 1 - [P(X=3) + P(X=4)]$

$= 1 - [C(4,3)(0.5)^3(0.5)^3(0.5)^1 + C(4,4)(0.5)^4(0.5)^0]$

$= 1 - [0.25 + 0.0625] = 0.6875$

Required percentage = 0.6875 × 100% = 68.75%

3.9.5 Fitting of Binomial Distribution

If a random experiment consisting of n trials satisfying the conditions of binomial distribution is repeated N times then the *expected frequencies* or *theoretical frequencies* of getting r successes is given by

$$f(r) = N \times C(n, r)\, p^r q^{n-r}\ ;\ r = 0, 1, 2, \ldots, n$$

Note: If p is unknown, we find $np = \bar{X} = \frac{\Sigma fx}{N}$. So, $p = \frac{\bar{X}}{n}$

Example 43: Out of 9000 families 4 children each, how many families would you expect to have (a) 2 boys and 2 girls; (b) at least one boy

(c) no girls (d) 3 boys and 1 girl [TU, 2063/BE, 2058 shrawan]

(e) at most 2 girls, (f) at least one boy and one girl [TU, BE, 2058 shrawan]

if boys and girls are equally probable.

Solution: With usual notations, we have $p = 0.50$, $q = 0.50$, $N = 9000$Let r be the number of successes (boys or girls) in the experiment. Using $P(X = r) = C(n, r)\, p^r q^{n-r}$; $r = 0, 1, 2, \ldots, n$, we get [we can also solve this problem by defining X as the rv denoting the number of 38 boys in the family with 4 children each. See example 36]

(a) No. of families expected to have 2 boys and 2 girls

$= N \times P(X = 2) = 9000 \times C(4, 2)(0.5)^2(0.5)^{4-2} = 9000 \times 0.375 = 3375$

(b) No. of families to have at least one boys $= N \times P(X \geq 1)$

$= N[1 - P(X=0)] = 9000 \times [1 - 0.0625] = 9000 \times 0.9375 = 8437.5 = 8438$

(c) No. of families expected to have no girl is

$= N \times P(X = 0) = N \times 0.0625 = 562.5 = 563$

(d) No. of families expected to have 3 boys and 1 girl

$= N \times P(X = 3) = 9000 \times 0.25 = 2250$

(e) No. of families expected to have at most 2 girls $= N \times P(X \leq 2)$

$= N \times B(2; 4, 0.5) = 9000 \times \Sigma_{x=0}^{2} (x; 4, 0.5) = 9000 \times 0.6875$

$= 6187.5 = 6188$ [Using Table]

(f) $N \times P(X \geq 1) = N \times [1 - P(X = 0) = 8438$

Example 44: Fit a binomial distribution for the following assuming the coin unbiased. [TU 2065]

No. of heads	0	1	2	3	4
Frequency	28	62	46	20	4

Solution: Since coin is unbiased, $p = q = 1/2 = 0.5$

Sample size $(n) = 4$ [$\because$ only four samples are success]

Total frequency $(N) = 160$. Let r be the number of successes (or heads). Then, the expected frequency of getting r successes is given by

$f(r) = N \times P(X = r) = N \times C(n, r)\, p^r q^{n-r}\ ;\ r = 0, 1, 2, \ldots, n$

$= 160 \times C(4, r)(0.5)^r (0.5)^{4-r} = 160 \times C(4, r)(0.5)^4 = 10 \times C(4, r)$

Fitting of Binomial Distribution

R	$f(r) = 10 \times C(4, r)$
0	$f(0) = 10 \times C(4, 0) = 10$
1	$f(1) = 10 \times C(4,1) = 40$
2	$f(2) = 10 \times C(4, 2) = 60$
3	$f(3) = 10 \times C(4, 3) = 40$
4	$f(4) = 10 \times C(4, 4) = 10$

Hence fitted binomial distribution is:

x	0	1	2	3	4
f	10	40	60	40	10

Example 45: Fit a binomial distribution to the following data:

x	0	1	2	3	4
f	5	10	20	30	15

Solution: $n = 4$, $N = \Sigma f = 80$, $np = \bar{X} = \frac{\Sigma fx}{N} = \frac{200}{80} = \frac{5}{2} \Rightarrow p = \frac{5}{8} = 0.625$

Then, proceeding as in example 40, we get fitted binomial distribution as

X	0	1	2	3	4
f	2	11	26	29	12

Example 46: In an experiment, a fair die is thrown 6 times. The event of occurring number greater than 4 is 'success' of the experiment. If the experiment is repeated 3645 times, find the binomial frequency distribution. Also calculate the mean number of successes and S.D.

Solution: $P(X > 4) = P(5) + P(6) = 1/6 + 1/6 = 1/3$, $q = 2/3$, $n = 6$, $N = 3645$.

The binomial frequency distribution is

x	0	1	2	3	4	5	6
f	320	960	1200	800	300	60	5

Mean $(np) = 6 \times \frac{1}{3} = 2$, s.d. $= \sqrt{npq} = 1.15$

3.10 The Hypergeometric Distribution

3.10.1 Introduction

If each unit selected for the sample is replaced before the next one is drawn, then this guarantees independence of trials and leads to the binomial distribution. So, we apply binomial distribution when we do ***sampling with replacement*** i.e. sampling with replacement is associated with binomial distribution. On the other hand, ***sampling without replacement*** is associated with Hypergeometric Distribution. So to solve the problem of sampling without replacement we apply Hypergeometric distribution.

So Binomial distribution is applicable in which the probability of success is same for all trials. However, if it varies from trial to trial, hypergeometric distribution bution is more suitable.

Hypergeometric and binomial distribution are related to each other. Whereas the binomial distribution is the approximate probability model for sampling without replacement from a finite dichotomous (S–F) population, the hypergeometric distribution is the exact probability model for the number of successes (S) in the same sample.

3.10.2 Conditions of Using Hypergeometric Distribution

The assumptions leading to the hypergeometric distribution are:

1. The population or set to be sampled consists of N individuals, objects or elements (a finite population).
2. Each individual can be characterized as a success (S) or a failure (F), and there are M successes in the population.
3. A sample of n individuals is selected without replacement in such a way that each subset of size n is equally likely to be chosen.

Here, the random variable of interest is X = the number of successes in the sample. The probability distribution of X depends on the parameters n, M and N. So we wish to obtain

$$P(X=x) = h(x;\, n, M, N).$$

Hence, when the population is finite and the sampling isdone without replacement so that the events are stochastically dependent, although random, we obtain hypergeometric distribution. Consider a basket contains N items, M of which are defective and $N-M$ are non defective.

Here, the x successes (defectives) can be chosen in $\binom{M}{x}$ ways, the $n-x$ failures (non defectives) can be chosen in $\binom{N-M}{n-x}$ ways. Hence x successes and $n-x$ failures can be chosen in $\binom{M}{x}\binom{N-M}{n-x}$ ways.

Also, n objects can be chosen from a set of N objects in $\binom{N}{n}$ ways. If we consider all probabilities equally likely, it follows that for sampling without replacement the probability of getting 'x successes in n trial' is

$$P(X=x) = h(x\,;\, n, M, N) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}, x = 0, 1, 2, \ldots, n$$

Definition: A set of N objects contains M objects classified as successes and $N-M$ objects classified as failures.

A sample of size n objects is selected randomly (without replacement) from the N objects, where $M \le N$ and $n \le N$.

Let the random variable X denote the number of successes in the sample. Then X is a *hypergeometric random variable*.

The distribution of a discrete random variable X is said to be *hypergeometric distribution* with parameters n, M, N if it assumes only non-negative values and its probability mass function is given by

$$P(X=x) = h(x;\, n, M, N) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}, x = 0, 1, 2, \ldots, n$$

where $x \le M$ and $n - x \le N - M$,

n = sample size, N = the *lot* (or *population*) *size*,

M = the number of *successes* in the lot.

Mean and variance of hypergeometric distribution

The mean and variance of the hypergemotric rv X having *pmf* $h(x;\, n, M, N)$ are

$$Mean = E(X) = n\frac{M}{N} = np;$$

$$Variance = V(X) = \left(\frac{N-n}{N-1}\right).n.\frac{M}{N}.\left(1-\frac{M}{N}\right) = \left(\frac{N-n}{N-1}\right)np(1-p), \text{ where } p = \frac{M}{N}$$

and $\left(\frac{N-n}{N-1}\right)$ is called the finite population correction factor.

Remark:

1. This shows that the 'means' of the binomial and hypergeometric whereas the variances of two *rv's* differ by the factor $(N-n)/(N-1)$, the *finite population correction factor*. This factor is less than 1, hypergeometric variable has smaller variance than does the binom variable.
2. Variance of hypergeometric distribution

$$= \left(\frac{N-n}{N-1}\right) np(1 - p) = \left(1-\frac{N-n}{N-1}\right) \times \text{ variance of}$$

distribution.

It should be noted that, for fixed p, as N increases to ∞, var(X) to $np(1-p)$, which is the variance of Binomial *rv* with parameters

3.10.3 Generalization of hyper geometric distribution

1. If there are two categories N_1 = good and $N_2 = N - N_1$ = defective geometric distribution is

$$P(X=x) = h(x;\, n. N_1, N) = \frac{\binom{N_1}{x}\binom{N_2}{n-x}}{\binom{N}{n}}. x = 0, 1. 2. 3, \ldots, n.$$

2. If there are m categories instead of two (good and defective), the hypergeometric distribution may be generalized to the form,

$$\frac{\binom{N_1}{x_1}\binom{N_2}{x_2}\binom{N_3}{x_3}\cdots\binom{N_m}{x_m}}{\binom{N}{n}}$$

3.10.4 Difference between binomial distribution and hypergeometric distribution

In binomial distribution, since items are replaced after each trial, the tr performed under identical conditions and probability of success/failure is c throughout experiment. Size of population remains constant after each trial.

But in hypergeometric distribution (*i*) Item are not replaced after each t Entire population is divided into two groups: one with success characteris other with failure characteristics. (*iii*) Successive trials are not performed identical conditions. (*iv*)The probability of failure/success does not remain co throughout the experiment. (*v*) The population of size decreases after each trial.

Example: We use hypergeometric model to estimate the number of fish in a lake.

3.10.5 The binomial approximation to hypergeometric distribu

Let the population size N, and the number of population successes M, g with the ratio M/N approaching p. Then $h(x\,;\, n, M, N)$ approaches $b(x\,;\, n, p)$; so i small, the two are approximately equal provided that p is not too near either Hypergeometric distribution tends to Binomial distribution as $N \to \infty$ and $M/N \to p$

So, for hypergeometric distribution if the sampling fraction n/N is sma less than 0.1, then the binomial distribution with parameters $p = M/N$ and n pr a good approximations. The smaller the ration n/N, the better the approximation

Example 47: (*Calculating a probability using the hypergeometric distribution*) An Internet-based company that sells discount accessories for cell phones often an excessive number of defective products. The company needs better cont quality. Suppose it has 20 identical car chargers on hand but that 5 are defective. If the company decides to randomly select 10 of these items, what is the probability that 2 of the 10 will be defective?

Solution: Substituting $x = 2$, $n = 10$, $M = 5$, and $N = 20$ into the formula for the hypergeometric distribution

$$P(X = x) = h(x; n, M, N) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}, x = 0, 1, 2, \dots, n$$

$$\text{we get } h\,(2; 10, 5, 20) = \frac{\binom{5}{2}\binom{15}{8}}{\binom{20}{10}} = \frac{10 \times 6{,}435}{184{,}756} = 0.348.$$

Note: In the preceding example, n was not small compared to N, and if we had made the mistake of using the binomial distribution with $n = 10$ and $p = 5/20 = 0.25$ to calculate the probability of two defectives, the result would have been 0.282, which is much too small. However, when n is small compared to N, less than $N/10$, the composition of the lot is not seriously affected by drawing the sample without replacement, and the binomial distribution with the parameters n and $p = M/N$ will yield a good approximation.

Example 48: A batch of parts contains 100 parts from a local supplier of tubing and 200 parts from a supplier of tubing in the next state. If four parts are selected randomly and without replacement,(**a**) what is the probability they are all from the local supplier? (**b**) What is the probability that two or more parts in the sample are from the local supplier? (**c**) What is the probability that at least one part in the sample is from the local supplier? (**d**) Find the mean and variance.

Solution: Let X equal the number of parts in the sample from the local supplier. Then, X has a hypergeometric distribution.

(**a**) The requested probability is $P(X= 4)$. Consequently,

$$P(X = 4) = \frac{\binom{100}{4}\binom{200}{0}}{\binom{300}{4}} = 0.0119.$$

(**b**) The probability that two or more parts in the sample are from the local supplier is

$$P(X \geq 2) = \frac{\binom{100}{2}\binom{200}{2}}{\binom{300}{4}} + \frac{\binom{100}{3}\binom{200}{1}}{\binom{300}{4}} + \frac{\binom{100}{4}\binom{200}{0}}{\binom{300}{4}}$$

$$= 0.298 + 0.098 + 0.0119 = 0.408$$

(**c**) The probability that at least one part in the sample is from the local supplier is

$$P(X \geq 1) = 1 - P(X = 0) = \frac{\binom{100}{0}\binom{200}{4}}{\binom{300}{4}} = 0.084$$

(**d**) The sample size is $n = 4$. The random variable X is the number of parts in the sample from the local supplier. Then, $p = 100/300 = 1/3$. Therefore,

$$\text{Mean} = E(X) = np = 4\times(1/3) = 1.33.$$

$$\text{Variance} = V(X) = \left(\frac{N-n}{N-1}\right) np(1-p) = \left(\frac{100-4}{100-1}\right) 4(1/3)(2/3) = 0.862.$$

Example 49: (***Calculate a probability using the hypergeometric distribution***):
A shipment of 20 digital voice recorders contains 5 that are defective. If 10 of them are randomly chosen for inspection, what is the probability that 2 of the 10 will be defective? Also, find mean and variance of the distribution.

[TU, BE, 2063 Ashadh/2066 Magh/2066 Magh/2067 Shrawan/2068 Bhadra]

Solution: By using Hypergeometric distribution

$$x = 2, n = 10, M = 5, N = 20$$

Let X denote the number of defective digital voice recorders.

Probability of 2 out of n is given by

$$P(X = 2) = h(2;10,5,20) = \frac{\binom{5}{2}\binom{20-5}{10-2}}{\binom{20}{10}} = \frac{10 \times 6\,435}{184756} = 0.3483$$

$$\text{Also, Mean} = \mu = E(X) = n \times \frac{M}{N} = 10 \times \frac{5}{20} = 2.5$$

$$\sigma^2 = \text{variance } V(X) = \left(\frac{N-n}{N-1}\right) npq = \left(\frac{20-10}{20-1}\right) 10 \times 0.25 \times 0.75 = 0.9868$$

$$\text{So, } \sigma = \sqrt{\frac{75}{76}} = \sqrt{0.9868} = 0.9934$$

In the preceding example, n was not small compared to N, and if we had made the mistake of using the binomial distribution with $n = 10$ and $p = 5/20 = 0.25$ to calculate the probability of two defectives, the result would have been 0.282, which is much too small.

[By using Binomial distribution: $P(X = 2) = b(x; n, p) = C(n, x)\, p^x\, q^{n-x}$

where $n = 10$, $x = 2$, $p = \frac{M}{N} = \frac{5}{20} = 0.25$

So, probability of getting 2 out of $n = 10$ is given by

$P(X = 2) = C(10,2)(0.25)^2(0.75)^{10-2} = 0.282$]

However, when n is small compared to N, less then $N/10$, the composition of the lot is not seriously affected by drawing the sample without replacement, and the binomial distribution with the parameters n and $p = M / N$ will yield a good approximation.

Example 50: (**A numerical comparison of the hypergeometric and binomial distributions**)

Repeat the preceding example for a lot of 100 digital voice recorders, of which 25 are defective; by using

(**a**) the formula for the hypergeometric distribution ;

(**b**) the formula for the binomial distribution as an approximation.

Solution: (**a**) N = lot size = 100, n = sample size = 10

M = number of 'successes' in the lot = 25

So, probability of getting 2 defectives out of 10 is obtained by using

$$P(X=x)=h(x\,;n,M,N)=\frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}$$

$$\text{So, } P(X=2)=h(2;\,10,25,100)=\frac{\binom{25}{2}\binom{100-25}{10-2}}{\binom{100}{10}}=0.292$$

(b) Substituting $x=2$, $n=10$, and $p=\frac{M}{N}=\frac{25}{100}=0.25$ into the formula for the binomial distribution. $P(X=x)=b(x;n,p)=C(n,x)\,p^x\,q^{n-x}$ we get, $P(X=2)=b(2;\,10,0.25)=C(10,2)\,(0.25)^2\,(0.75)^{10-2}=0.282.$
Here the difference between the two values is only 0.010. In general, it can be shown that $h(x;n,M,N)$ approaches $b(x;n,p)$ with $p=M/N$ when $N\to\infty$.
So, we can use binomial distribution as an approximation to the hypergeometric distribution if $n\le\frac{N}{10}$ i.e. $\frac{n}{N}\le\frac{1}{10}$.

Example 51: A carton of 24 hand grenades contains 4 that are defective. If 3 hand grenades are randomly selected from this carton, what is the probability that exactly 2 of them are defective?
[TU, BIE, 2065 Chaitra]

Solution: Let X be a rv denoting the number of defective hand grenades.
Given N = lot size = 24, n = sample size = 3
M = no. of success (of defective hand grenades) = 4.
Probability of getting 2 hand grenades out of 3 is

$$P(X=x)=h(x\,;n,M,N)=\frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}$$

$$\text{i.e., } P(X=2)=h(2;\,3,4,24)=\frac{\binom{4}{2}\binom{24-4}{3-2}}{\binom{24}{3}}=0.0593.$$

Example 52: From a lot containing 25 items, 5 of which are defective, 4 are chosen at random. Obtain the probability distribution of the number of defective items drawn. Also find the probability of 3 or 4 defective items.

Solution: Given, N = lot size = 25, n = sample size = 4
M = no. of successes (i.e. defectives) in the lot = 5
Let X be the rv denoting the number of defective items drawn. Probability of "x success out of n" is given by

$$P(X=x)=h(x\,;n,M,N)=\frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}$$

The distribution is

$X=x$	$P(X=x)=h(x;n,M,N)$
0	$P(X=0)=0.383$
1	$P(X=1)=0.4506$
2	$P(X=2)=0.1502$
3	$P(X=3)=0.158$
4	$P(X=4)=0.0004$

$$P(3\text{ or }4\text{ defectives})=P(X=3\text{ or }X=4)=P(X=3\cup x=4)=P(X=3)+P(x=4)$$
$$=0.0158+0.0004=0.0162$$

Example 53: A quality control engineer inspects a random sample of 3 batte from each lot of 24 car batteries ready to be shipped. If such a lot contains 6 batte with slight defects, what are the probabilities that the inspector's sample will cont. (a) none of the batteries with defects; (b) only one of the batteries with defects; (c least two of the batteries with defects?
[TU, BE, 2062 Baishal

Solution: Given, N = lot size = 24, n = sample size = 3
M = no. of successes (i.e. defective batteries) = 6
Let X be the rv which denote the number of the batteries with defects.

(a) Probability of none of the batteries with defects is given

$$P(X=0)=h(0;\,3,6,24)=\frac{\binom{6}{0}\binom{24-6}{3-0}}{\binom{24}{3}}=\frac{816}{2024}=0.4032$$

(b) $$P(X=1)=h(1;\,3,6,24)=\frac{\binom{6}{1}\binom{24-6}{3-1}}{\binom{24}{3}}=\frac{918}{2024}=0.4536$$

(c) $P(X\ge 2)=P(X=2)+P(X=3)=h(2;\,3,6,24)+h(3;\,3,6,24)$

$$=\frac{\binom{6}{2}\binom{24-6}{3-2}}{\binom{24}{3}}+\frac{\binom{6}{3}\binom{18}{0}}{\binom{24}{3}}=\frac{270}{2024}+\frac{20}{2024}=0.1433$$

Example 54: Five individuals from an animal population thought to be n extinction in a certain region have been caught, tagged, and released to mix into t population. After they have had an opportunity to mix a random sample of 10 these animals is selected. Let X = the number of tagged animals in the seco sample. If there are actually 25 animals of this type in the region, what is t probability that (a) $X=2$; (b) $X\le 2$?

Solution: The parameter values are $n=10$
$M=5$ (5 tagged animals in the population), $N=25$

$$\text{So, } P(X=x)=h(x\,;n,M,N)=\frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}},\quad x=0,1,2,3,4,5.$$

(a) $$P(X=2)=h(2;\,10,5,25)=\frac{\binom{5}{2}\binom{20}{8}}{\binom{25}{10}}=0.3854$$

(b) $P(X\le 2)=P(X=0,1\text{ or }2)=\sum_{x=0}^{2}h(x;\,10,5,25)$
$=0.0565\ 6\ 0.2569+0.3854=0.6988.$

(c) Find also mean and variance in the above example.

Scanned by CamScanner

Here, $n = 10, M = 5, N = 25, p = \frac{M}{N} = \frac{5}{25} = 0.2,\ E(X) = np = 10 \times 0.20 = 2$

$$V(X) = \left(\frac{N-n}{N-1}\right) \cdot np(1-p) = \frac{15}{24}(10)(0.20)(0.80) = (0.625)(1.6) = 1$$

In Binomial distribution $V(X) = np(1-p) = npq$.

So, if the sampling was carried out with replacement, then $V(X) = 1.6$.

Example 55: If 6 of 18 new buildings in a city violate the building code, what is the probability that a building inspector, who randomly selects 4 of the new buildings for inspection, will catch (a) none of the buildings that violate the building code; (b) 1 of the new buildings that violate the building code; *[TU, BIE, 2068 Bhadra]*

(c) 2 of the new buildings that violate the building code;

(d) at least 3 of the new buildings that violate the building code;

Solution: $N = 18, M = 6, n = 4$

Let X denote number of buildings that violate the building code in second sample.

(a) $$P(X = 0) = h(0; 4, 6, 18) = \frac{\binom{6}{0}\binom{18-6}{4-0}}{\binom{18}{4}} = 0.1618$$

(b) $P(X = 1) = h(1; 4, 6, 18) = 0.4314$, (c) $P(X = 2) = h(2; 4, 6, 18) = 0.3235$

(d) $P(X \geq 3) = \sum_{x=3}^{4} h(x; n, M, N) = h(3; n, M, N) + h(4; n, M, N) = 0.0835$

- **Example 56:** A taxi cab company has 12 Ambassador and 8 Fiats. If 5 of these taxi-cabs are in the shop for repairs and Ambassador is as likely to be in for repairs as a Fiat, what is the probability that (a) 3 of them are Ambassadors and 2 are Fiats; (b) 3 of them are Fiats and 2 are Ambassadors; (c) at least 3 of them are Ambassadors; (d) all 5 of them are of the same make? *[Pokhara Uni. BE 2002]*

Solution: Let X be a random variable which denote the number of Ambas-sador cars that are in the shop for repairs.

(a) Prob. that 3 of them are Ambassadors and 2 are Fiats is

$$P(X = 3) = \frac{\binom{12}{3}\binom{8}{2}}{\binom{20}{5}} = 0.3973$$

(b) Prob. that 3 of them are Fiats and 2 of them are Ambassadors is

$$P(X = 2) = \frac{\binom{12}{2}\binom{8}{3}}{\binom{20}{5}} = 0.2384$$

(c) $$P(X \geq 3) = P(X = 3) + P(X = 4) + P(X = 5) = \frac{\binom{12}{3}\binom{8}{2}}{\binom{20}{5}} + \frac{\binom{12}{4}\binom{8}{1}}{\binom{20}{5}} + \frac{\binom{12}{5}\binom{8}{0}}{\binom{20}{5}}$$

$= 0.3973 + 0.2554 + 0.0511 = 0.7038.$

(d) Let Y be a rv which denote the number of Fiat cars that are in the shop for repairs. Since X and Y are mutually exclusive,

$$P(X = 5 \text{ or } Y = 5) = P(X = 5) + P(Y = 5) = \frac{\binom{12}{5}}{\binom{20}{5}} + \frac{\binom{8}{5}}{\binom{20}{5}} = 0.0511 + 0.0036 = 0.0547$$

Example 57: The components of a 6- component system are to be randomly chosen from a bin of 20 used components. The resulting will be functional if at least 4 of its 6 components are in working condition. If 15 of the 20 components in the bin are in working condition, what is the probability that the resulting system will be functional?

Solution: Let X denote the number of working components chosen. Then X is hypergeometric with parameters $n = 6, M = 15, N = 20$. The probability that the system will be functional is

$$P(X \geq 4) = \sum_{x=4}^{6} h(x; n, M, N) = \frac{\binom{15}{4}\binom{5}{2} + \binom{15}{5}\binom{5}{1} + \binom{15}{6}\binom{5}{0}}{\binom{20}{6}} = 0.8687.$$

3.11 The Poisson Probability Distribution

3.11.1 Introduction:

When the probability of success p of an event is very small [so that $q = (1 - p)$ is almost equal to unity] and n, the number of trials, is very large such that the average number of series is a finite number, the resulting distribution is known as Poisson distribution. This discrete probability distribution was discovered by French Mathematician **Simeon Denis Poisson** (1781–1840) in 1837.

This distribution deals with the evaluation of probabilities of rare events such as "number of car accidents on road", "number of earthquakes in a year", "number of transistors that fail on their first day of use", "number of misprints on a page" etc.

It is particularly important because it is often used as a mathematical model describing behavior such as radioactive decay, arrivals of people in a line, planes arriving at an airport, cars pulling into a gas station, diners arriving at a restaurant, students arriving at a bookstore line, Internet users logging onto a Web site.

It is used to describe a number of processes, including the distribution of telephone calls going through a switchboard system, the demand of patients for service at a health institution, the arrivals of trucks and cars at a tollbooth.

The *binomial*, *hypergeometric*, and *negative binomial distributions* are all derived by starting with an experiment consisting of trials or draws and applying the laws of probability to various outcomes of the experiment. There is no simple experiment on which the Poisson distribution is based. In fact, it is based on the empirical results of past experience relating to the problem under study. So, past experiences provide the knowledge about the average or mean occurrences of the events.

3.11.2 Conditions of using Poisson distribution

The Poisson distribution has the following requirements:

1. The discrete random variable X is the number of occurrences of an event over some interval.
2. The occurrences of the events must be random.
3. The occurrences of the events must be rare and independent of each other.
4. The occurrences must be uniformly distributed over the interval being used.

3.11.3 Poisson Random Variable and Poisson Probability Function

The Poisson distribution is a discrete probability distribution that applies to occurrences of some event over a specified interval. The random variable X is the number of occurrences of the event in an interval. The interval can be time, distance, area, volume, or some similar unit. The probability of the event occurring X times over an interval is given by the following Formula.

Definitions: (Poisson distribution):

A random variable X is said to have ***Poisson distribution*** with parameter λ, $\lambda > 0$ if its *pmf* is given by

$$P(X = x) = f(x\,;\lambda) = \frac{e^{-\lambda}\lambda^{x}}{x!}\,;x = 0, 1, 2, 3, \ldots, \infty \qquad \text{———} (1)$$

Here, λ = mean of distribution

= average number of occurrences per specified interval.

$$e = \lim_{n \to \infty}\left(1 + \frac{1}{n}\right)^{n} = 2.7182818284 \ldots$$ is a constant in mathematics, named after the Swiss mathematician *L. Euler*.

$P(X = x)$ = the probability of exactly x successes.

Poisson random variable:

A random variable X, taking on one of the values 0, 1, 2, ... is said to be a ***Poisson random variable*** with parameter λ, $\lambda > 0$, if its probability mass function is given by (1). If X is ***Poisson random variable*** and the probability of getting x successes is given by (1) then this function is called the ***Poisson probability function***.

Remark: If we consider the length of interval as d instead of unit length, the average number of occurrences in d length of interval is λd. Thus the ***probability mass function*** in this situation is

$$P(X = x) = \frac{e^{-\lambda d}(\lambda d)^{x}}{x\,!}, x = 0, 1, 2, 3, \ldots .$$

3.11.4 Difference between Poisson distribution and binomial distribution

1. The binomial distribution is affected by the sample size n and the probability p, whereas the Poisson distribution is affected only by the mean μ.
2. In a binomial distribution, the possible values of the random variable X are 0, 1, ... , n, but a Poisson distribution has possible X values of 0, 1, 2, ... with no upper limit. .

3.11.5 The Poisson distribution as approximation to the Binomial distribution

The use of binomial distribution very tedious when very large and probability is very low. In this case Poisson distribution is used to approximate the binomial distribution. That is, binomial problem can be solved by using Poisson distribution. So, Poisson distribution is the limiting case of the Binomial distribution under the following conditions:

1. n, the number of trials is indefinitely large, i.e., $n \to \infty$.
2. p, the probability of success for each trial is indefinitely small, i.e., $p \to 0$.
3. $np = \lambda$, (say) is finite.

So, for smaller values of p and larger values of n, one may use either binomial probabilities or Poisson probabilities. The calculated probabilities will be almost

equal. As n gets larger and p gets smaller, both the probability distribution y same probabilities.

Please Remember: In the practical point of view, ***Poisson distribution is approximation of the Binomial distribution when the number of trials (n) i*** than or equal to 20 (i.e., $n \geq 20$) and probability of success p is less than or 0 05 (i.e., $p \leq 0.05$). If $n \geq 100$, the approximation is generally excellent so long $np \leq 10$.

In this case $P(X = x) = \dfrac{e^{-\lambda}\lambda^{x}}{x!} = \dfrac{e^{-np}(np)^{x}}{x!}$

Remarks:

1. We shall use the notation $X - P(\lambda)$ to denote that the random va follows Poisson distribution with parameter λ.
2. Poisson distribution is free of the number of trials (n).
3. The value of λ is frequently a rate per unit time or per unit area
4. (*i*) Because λ must be positive, $f(x;\lambda) > 0$ for all possible x values.

   (*ii*) Since by Maclaurin's infinite series

   $$e^{\lambda} = 1 + \lambda + \frac{\lambda^2}{2!} + \frac{\lambda^3}{3!} + \ldots = \Sigma_{x=0}^{\infty}\frac{\lambda^{x}}{x!} \quad \text{So, } \Sigma_{x=0}^{\infty} e^{-\lambda}\frac{\lambda^{x}}{x!} = e^{-\lambda}e^{\lambda} =$$

   Thus $f(x;\lambda)$ fulfills both conditions necessary for specifying a *pmf*.
5. The Poisson distribution is specially used when there are events which occur as outcomes of a definite number of trials in an experiment, rathe randomly in nature.

Some Examples of Poisson Distribution

We can apply position distribution for the following cases:

1. The number of wrong telephone numbers that are dialed in a day.
2. The number of α–particles discharged in a fixed period of time fro radioactive particle.
3. The number of suicides in a year in Kathmandu.
4. The number of deaths from cancer, heart attack, snake bite in a year.
5. The number of defective items manufactured by a reputed company.
6. The number of twin births in a city hospital.
7. The number of machines breaking down during any one day.
8. The number of defects in the insulation of a fifty meter length of wire.
9. The number of customers entering a post office on a given day.
10. The number of bacteria in a unit.
11. The number of cars parked at a place in an hour, say between 9:00 a.m 10:00 a.m.;
12. The number of fragments received by a surface area 'A' from a fragmen bomb etc

3.11.6 The Mean and Variance of Poisson distribution

Since $b(x;\,n,\,p) \to f(x;\,\lambda)$ as $n \to \infty$, $np \to \lambda$, the mean and variance binomial variable should approach those of a Poisson variable. These limits are

$np \to \lambda$ and $np(1 - p) \to \lambda$.

So, if X has a Poisson distribution with parameter λ, then

(i) $\mu = E(X) = \lambda$, (ii) $V(X) = E(X^2) - [E(X)]^2 = \lambda$, (iii) $\sigma_X = \sqrt{\lambda}$.

The mean and variance of a Poisson random variable are equal. For ex particle counts follow a Poisson distribution with a mean of 25 particles per centimeter, the variance is also 25 and the standard deviation of the counts is 5 per square centimeter. Consequently, information on the variability is very easily obtained. Conversely, if the variance of count data is much greater than the mean of the same data, the Poisson distribution is not a good model for the distribution of the random variable.

Theorem: In Poisson distribution mean and variance are the same.

[Purbanchal Uni. BE 2003]

Proof: For the Poisson distribution

$$P(X=x)=\frac{e^{-\lambda}\lambda^x}{x!}, x=0, 1, 2, \ldots.$$

$$\therefore \text{ Mean} = E(X) = \Sigma_{x=0}^{\infty} x\, P(X=x) = 0e^{-\lambda} + 1\,\lambda e^{-\lambda} + 2\,\frac{\lambda^2 e^{-\lambda}}{2!} + 3\,\frac{\lambda^3 e^{-\lambda}}{3!} + \ldots.$$

$$= \lambda e^{-\lambda}\left(1 + 2.\frac{\lambda}{2!} + 3.\frac{\lambda^3}{3!} + \ldots.\right) = \lambda 3^{-\lambda}\left(1 + \frac{\lambda}{1!} + \frac{\lambda^2}{2!} + \ldots.\right) = \lambda e^{-\lambda}e^{\lambda} = \lambda.$$

$$\text{Variance} = V(X) = x^2\, P(X=x) - \left[\Sigma_{x=0}^{\infty} x.P(X=x)\right]^2$$

$$\text{Now, } \Sigma_{x=0}^{\infty} x^2 P(X=x) = 0 + 1^2.\frac{\lambda e^{-\lambda}}{1!} + 2^2.\frac{\lambda^2 e^{-\lambda}}{2!} + 3.\frac{\lambda^3 e^{-\lambda}}{3!} + \ldots.$$

$$= \lambda e^{-\lambda}\left\{1 + \frac{2\lambda}{1!} + \frac{3\lambda^2}{2!} + \ldots.\right\} = \lambda e^{-\lambda}\left\{1 + \frac{(1+1)}{1!}\lambda + \frac{(1+2)}{2!}\lambda + \ldots.\right\}$$

$$= \lambda e^{-\lambda}\left[\left\{1 + \frac{\lambda}{1!} + \frac{\lambda^2}{2!} + \frac{\lambda^3}{3!} + \ldots.\right\}\right] + \left\{\lambda + \frac{2\lambda^2}{2!} + \frac{3\lambda^3}{3!} + \ldots.\right\} = \lambda e^{-\lambda}[e^{\lambda} = \lambda e^{\lambda}]$$

$$\therefore \text{ Variance} = \lambda e^{-\lambda}[e^{\lambda} + \lambda e^{\lambda}] - (\lambda)^2 = \lambda; \text{ Hence } E(X) = V(X) = \lambda.$$

3.11.7 Properties of Poisson distribution

The Poisson distribution is used to describe behavior of rare events where probability of success is very low. It is also called the "*Law of improbable events*". Poisson distribution possesses the following properties:

1. It is the only distribution known so far, of which the mean and variance are equal. So it is uniparametric distribution.
2. Poisson distribution possesses only one parameter λ.
3. Since it depends up on the λ, the average number of success per unit, λ is called the parameter of Poisson Distribution.
4. It is independent of number of trials (n).
5. It is always positively skewed and it tends to become symmetric for high value of λ.
6. It follows the additive properties i.e. if $X \sim P(\lambda_1)$ and $Y \sim P(\lambda_2)$ then $X + Y \sim P(\lambda_1 + \lambda_2)$
7. Poisson distribution is completely known when the value of λ is known.
8. Like binomial distribution, Poisson distribution could be also unimodal or bimodal depending upon the value of λ.
9. The number of successes that occur in any interval is independent of the number of successes that occur in any other interval.
10. The probability that a success will occur in an interval is the same for all intervals of equal size and is proportional to the size of the interval.
11. The probability that two or more successes will occur in an interval approaches zero as the interval becomes smaller.

3.11.8 Poisson Distribution Function:

Definition: If X is a Poisson random variable with *pmf*

$$P(X=x)=f(x;\lambda)=e^{-\lambda}\frac{\lambda^x}{x!}, \quad \lambda>0,\ x=0, 1, 2, 3, \ldots.$$

Then *Poisson distribution function* is defined as

$$P(X\le x)=F(x;\lambda)=\Sigma_{k=0}^{x} f(k;\lambda)=\Sigma_{k=0}^{x}\frac{e^{-\lambda}\lambda^k}{k!}$$

3.11.9 Fitting of Poisson Distribution

If a series of trials is repeated N times and satisfies the condition of Poisson distribution, then expected or theoretical frequency of x successes is given by

$$f(x)=NP(x)=N\frac{e^{-\lambda}\lambda^x}{x!}, x=0, 1, 2, 3, \ldots.$$

Note: If λ is not known, we find $\bar{X}$ of distribution and equate λ i.e. $\lambda = \bar{X}$

Example 58: From past experience it is known that there are on an average 2 traffic accidents per week. Find the probability that in a given week there will be (i) less than 2 accident; (ii) exactly 2 accidents, (iii) more than 2 accidents. *[TU BE 2068 Bhadra]*

Solution: Here, λ = Average no. of traffic accidents per week = 2

Let X denotes the number of traffic accidents per week. Then using Poisson distribution, the probability of exactly x traffic accidents in a given week is given by

$$P(X=x)=\frac{e^{-\lambda}\lambda^x}{x!},\ x=0, 1, 2.\ldots\ (1)$$

(i) The probability of less then 2 accidents in a given week is

$$P(X<2)=P(X=0)+P(X=1)=\frac{e^{-2}2^0}{0!}+\frac{e^{-2}2^1}{1!}=e^{-2}(1+2)=3e^{-2}=0.4060$$

(ii) The probability of exactly 2 accidents in a given week is

$$P(X=2)=\frac{e^{-2}2^2}{2!}=\frac{4e^{-2}}{2}=0.2707$$

(iii) The probability of more than two accidents in a given week is

$$P(X>2)=1-P(X\le 2)=1-[P(X=0)+P(X=1)+P(X=2)]$$

$$=1-\left[\frac{e^{-2}2^0}{0!}+\frac{e^{-2}2^1}{1!}+\frac{e^{-2}2^2}{2!}\right]=1-e^{-2}[1+2+2]=1-0.6767=0.3233.$$

Example 59: Suppose that we are investigating the safety of a dangerous intersection. Past police records indicate a mean of 5 accidents per month at this intersection. Suppose the number of accidents is distributed according to a Poisson distribution. Calculate the probability in any month of exactly 0, 1, 2, 3 or 4 accidents.

Solution: Since the number of accidents is distributed according to a Poisson distribution and the mean number of accidents per month is 5, we have the probability of happening accidents in any month

$$P(X=x)=\frac{e^{-\lambda}\lambda^x}{x!},\ x=0, 1, 2, 3, 4.$$

By this formula we can calculate

$P(0) = 0.00674$, $P(1) = 0.3370$, $P(2) = 0.08425$, $P(3) = 0.14042$, $P(4) = 0.17552$.

Example 60: (***World War II Bombs***): In analyzing hits by V-1 buzz bombs nor interest, but the number has in World War II, South London was subdivided into 576 regions, each with an area of 0.25 km^2. A total of 535 bombs hit the combined area of 576 regions. (a) If a region is randomly selected, find the probability that it was hit exactly twice. (b) Based on the probability found in part (a), how many of the 576 regions are expected to be hit exactly twice?

Solution: (a) The Poisson distribution applies because we are dealing with the occurrences of an event (bomb hits) over some interval (a region with area of 0.25 km^2). The mean number of hits per region is

$$\lambda = \frac{number\ of\ bomb\ hits}{number\ of\ regions} = \frac{535}{576} = 0.929.$$

Because we want the probability of exactly two hits in a region, we let $x = 2$ and use

formula $\quad P(X = x) = f(x\ ;\ \lambda) = \frac{e^{-\lambda}\lambda^x}{x!}$;

$$P(2) = \frac{e^{-0.929}0.929^2}{2!} = 0.170.$$

The probability of a particular region being hit exactly twice is $P(2) = 0.170$.

(b) Because there is a probability of 0.170 that a region is hit exactly twice, we expect that among the 576 regions, the number that are hit exactly twice is 576 × 0.170 = 97.9.

In the preceding example, we can also calculate the probabilities and expected values for 0, 1, 3, 4, and 5 hits. (We stop at $X = 5$ because no region was hit more than five times, and the probabilities for $X > 5$ are 0.000 when rounded to three decimal places.) Those probabilities and expected values are listed in Table.

V-1 Buzz Bomb Hits for 576 Regions in South London

Number of Bomb Hits	*Probability*	*Expected Number of Regions*	*Actual Number of Regions*
0	0.395	227.5	229
1	0.170	211.4	211
2	0.170	97.9	93
3	0.053	30.5	35
4	0.012	6.9	7
5	0.002	1.2	1

The fourth column of Table describes the results that actually occurred during World War II. There were 229 regions that had no hits, 211 regions that were hit once, and so on. We can now compare the frequencies predicted with the Poisson distribution (third column) to the actual frequencies (fourth column) to conclude that there is very good agreement. In this case, the Poisson distribution does a good job of predicting the results that actually occurred.

Example 61: An office switchboard receives telephone calls at a rate of 3 calls per minute on an average. Find the probability of receiving (i) no calls in one minute interval; (ii) at most 2 call in a five minute interval. *[TU BIE 2064 Bhadra]*

Solution: (i) λ = Average no. of calls per minute = 3

Let X denote the number of calls on an office switchboard receives per minute. Then, using Poisson distribution, the probability of receiving exactly x calls per minute is given by

$$P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}, x = 0, 1, 2, \ldots.$$

$\therefore$ The probability of receiving no calls in one minute interval is

$$P(X = 0) = \frac{e^{-3}\lambda^0}{0!} = e^{-3} = 0.0498$$

(ii) Since average no. of calls per minute = 3

So, average no. of calls in 5 minute interval $= \lambda d = 3 \times 5 = 15 \qquad [\therefore d = 5]$

Let X denote the number of calls an office switchboard receives in 5 minutes. Then, using Poisson distribution, the probability of receiving exactly x calls in 5 minutes is given by

$$P(X = x) = \frac{e^{-\lambda}\lambda^x}{x\,!}, x = 0, 1, 2, \ldots.$$

$\therefore$ The probability of receiving at most 2 calls in 5 minutes interval is

$P(X \le 2) = P(X = 0) + P(X = 1) + P(X = 2)$

$$= \frac{e^{-15}(15)^0}{0!} + \frac{e^{-15}(15)^1}{1!} + \frac{e^{-15}(15)^2}{2!} = e^{-15}(1+15+\frac{225}{2}) = 0.0000393$$

Example 62: Suppose the Highway Safety Division wants to investing of the safety of a dangerous intersection. Past police records indicate a mean of five accidents per month at this intersection. The number of accidents is distributed according to Poisson distribution. Find the probability of accidents (i) exactly 0 per month; (ii) $P(0 < X \le 3)$ per month; (iii) more than 3 per month; (iv) $P(2 \le X \le 3)$

[TU, BE, 2064 shrawan/2067 Mangsir/2068Magh]

Solution: Here, λ = Average no. accidents per month = 5

Let, X denote the number of accidents per month. Then

(i) $P(X = 0) = \frac{e^{-5}5^0}{0!} = e^{-5} = 0.0067$

(ii) $P(0 < X \le 3) = P(1) + P(2) + P(3)$

$$= \frac{e^{-5}5^1}{1!} + \frac{e^{-5}5^2}{2!} + \frac{e^{-5}5^3}{3!} = e^{-5}\left(5 + \frac{25}{2} + \frac{125}{6}\right) = 0.2583$$

(iii) $P(X > 3) = 1 - P(X \le 2) = 1 - \Sigma_{k=0}^{2} \frac{e^{-\lambda}\lambda^k}{k\,!}$

$$= 1 - [P(0) + P(1) + P(2)] = 1 - \left[e^{-5}\{1+5+\frac{25}{2}\}\right] = 1 - 0.1247 = 0.8753$$

(iv) $P(2 \le X \le 3) = P(2) + P(3) = e^{-5}\left\{\frac{5^2}{2!} + \frac{5^3}{3!}\right\} = e^{-5}\left\{\frac{25}{2} + \frac{125}{6}\right\} = 0.2246$

Example 63: Let X = the number of automobiles of particular year and model that will at some time in the future suffer a catastrophic failure in the steering mechanism causing complete loss of control at high speed. Suppose X has a Poisson distribution with parameter $\lambda = 10$. *[TU, BE, 2056 Bhadra/2062 Bhadra]*

(a) What is the probability that (i) at least 10 cars suffer such a failure? (ii) at most 10 cars suffer such a failure?

(b) What is the probability that between 10 and 15 inclusive cars will suffer such a failure?

Solution: Here λ = Average no. of cars that suffer a failure = 10

Let X denote the number of cars that will suffer a failure. Then using Poisson distribution, the probability of suffering a failure of exactly x cars is given by

$$P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}, x = 0, 1, 2, 3, \ldots.$$

(a) (i) P(at least 10 cars will suffer a failure) $= P(x \ge 10)$

$$= \Sigma_{k=10}^{\infty} \frac{e^{-10}(10)^k}{k\,!} = 1 - \Sigma_{k=0}^{9} \frac{e^{-10}(10)^k}{k\,!} = 1 - F(9; 10) = 1 - 0.458 = 0.542$$

(ii) $P(X \ge 10) = ?$

Using $P(X \le x) = F(x; \lambda) = \Sigma_{k=10}^{x} \frac{e^{-\lambda}\lambda^x}{x!}$, we get

$$P(X \le 10) = F(10;10) = \Sigma_{k=10}^{10} \frac{e^{-10}(10)^k}{k\,!} = 0.458$$

(iii) $P(10 \le X \le 15) = \Sigma_{k=10}^{15} \frac{e^{-10}(10)^k}{k\,!}$

$$= \Sigma_{k=10}^{15} \frac{e^{-10}(10)^k}{k\,!} - \Sigma_{k=0}^{9} \frac{e^{-10}(10)^k}{k\,!} = 0.951 - 0.458 = 0.493 \quad \text{[Using Table]}$$

Example 64: It is known that the average number of suicides per week in Kathmandu valley is 1.5. Express the no. of suicide in a month. Find the probability that there will be 5 or more suicides in a month.

Solution: Here, Average no. of suicides per week = 1.5

So, average no of suicides per month = $1.5 \times 4 = 6 = \lambda$

Let X denote the no. of suicides per month. Then using Poisson distribution $P(X=x) = \frac{e^{-\lambda}\lambda^x}{x!}$, we get

$$P(X \ge 5) = \Sigma_{k=5}^{\infty} \frac{e^{-6}6^k}{k\,!} = 1 - \Sigma_{k=0}^{4} \frac{e^{-6}6^k}{k\,!} = 1 - 0.285 = 0.715$$

Example 65: At a checkout counter customers arrive at an average of 1.5 per minute. Find the probability that (a) at most 4 will arrive in any given minute: (b) at least 3 will arrive during an interval of 2 minutes;

(c) at most 15 will arrive in an interval of 6 minutes

[Pokhara uni, BE 2007 spring; 2009 Fall]

Solution: (a) $\lambda = 1.5$. Let X denote the no. of customers who arrive in a minute.

$$\therefore\ P(X \le 4) = \Sigma_{k=0}^{4} e^{-1.5}\frac{(1.5)^k}{k\,!} = 0.981 \text{ [Using Table]}$$

(b) λ = average no. of customers who arrive in 2 minutes = $1.5 \times 2 = 3$.

Let X denote the no. of customers who arrive in 2 minutes

$$\therefore\ P(X \ge 3) = 1 - P(X \le 2) = 1 - 0.423 = 0.577$$

(c) λ = average no. of customers who arrive in 6 minute interval = $1.5 \times 6 = 9$

Let X denote the no. of customers who arrive in 6 minute interval.

$$\therefore\ P(X \le 15) = \Sigma_{k=0}^{15} \frac{e^{-1}1^k}{k\,!} = \Sigma_{k=0}^{15} \frac{e^{-9}9^k}{k\,!} = 0.978 \text{ [Using Table 2]}$$

Example 66: It is known that 5% of the books bound at a certain bindery have defective bindings. Find the probability that 2 out of 100 books bound by this bindery will have defective bindings using

(a) the formula for the binomial distribution

(b) the Poisson approximation to the binomial distribution.

Solution: (a) $x = 2$, $n = 100$, $p = 0.05$

Using $b(x;\, n, p) = C(n, x)\, p^x\, q^{n-x}$

We get $b(2;\, 100, 0.05) = C(100, 2)\, (0.05)^2\, (0.95)^{98} = 0.081$

(b) $x = 2$, $\lambda = np = 100\,(0.05) = 5$

Using $P(X = x) = f(x;\, \lambda) = \frac{e^{-\lambda}\lambda^x}{x!}$, we get

$$f(2;\, 5) = \frac{e^{-5}\,5^2}{2!} = 0.084$$

Since $n = 100$, approximation is excellent.

Here the difference between two values we obtained (the error we would make by using the Poisson approximation) is only 0.003.

Example 67: (*Probability of Imperfection*): In the inspection of tin plate produce by continuous electrolytic process, 0.2 imperfection is spotted per minute, o average. Find the probabilities of spotting

(a) one imperfection in 3 minutes; (b) at least two imperfections in 5 minutes; (c) a most one imperfection in 15 minutes;

Solution: (a) $\lambda = (0.2) \times 3 = 0.6$. Using $F(x;\, \lambda) = \Sigma_{k=0}^{\infty} \frac{e^{-\lambda}\lambda^k}{k\,!}$, we get

$$\therefore\ P(X = 2) = F(1;\, 0.6) - F(0;\, 0.6)$$

$$= 0.878 - 0.549 = 0.329 \text{ [Using table]}$$

(b) $\lambda = (0.2) \times 5 = 1\ P(X \ge 2) = 1 - P(X \le 1) = 1 - F(1;\, 1) = 1 - 0.7636 = 0.264$

(c) $\lambda = 0.2 \times 15 = 3;\ P(X \le 1) = F(1;\, 3) = 0.199$

Example 68: Assume that the probability of an individual coal miner being killed in a mine accident during a year is 1/2400. Use appropriate statistical distribution to calculate the probability that in a mine employing 200 miners there will be (a) at least one fatal accident; (b) exactly 2 fatal accidents; (c) at least two fatal accidents in a year.

[TU BE, 2066 Magh/2067 Shrawan/2067 Mangsir]

Solution: Here, we have

$$n = 200,\ p = \frac{1}{2400},\ \text{So } \lambda = np = 200 \times \frac{1}{2400} = 0.083.$$

Now the probability of x fatal accidents is given by

$$P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!},\ x = 0, 1, 2, 3, \ldots$$

(a) The probability of at least one fatal accidents is

$$P(X \ge 1) = 1 - P(0)$$

$$= 1 - \frac{e^{-\lambda}\lambda^0}{0!} = 1 - e^{-0.083} = 1 - 0.9204 = 0.0769$$

(b) $P(X = 2) = 1 - P(X \le 1) = 1 - [P(0) + P(1)]\ 1 - e^{-\lambda}[1 + 0.083]$

$$= 1 - 0\,.977 = 0.003$$

Or, Using table 2 $P(X \ge 2) = 1 - P(X \le 1) = 1 - F(1;\, 0.083) = 1 - 0.977 = 0.003$

Example 69: In a certain factory turning out blades, there is a 0.2% probability for any blade to be defective. Blades are supplied in packets of 10. Using Poisson distribution, calculate the approximate number of packets containing no defective, one defective, two defective, at least one defective respectively in a consignment of 20,000 packets.

[TU, BE, 2057 Bhadra]

Solution: Here, $N = 20{,}000$, $n = 10$ and p = probability for any blade to be defective = 0.2% = 0.002. So, $\lambda = np = 10 \times 0.002 = 0.02$

Let X be a random variable which denote the number of defective blades in a packet of 10. Let x denote the no. of defective blades per packet. Then the probability of getting x defective blades is given by $P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}$, $x = 0, 1, 2, \ldots$

(i) The expected no. of packets containing no defective blades

$$= f(0) = N \times P(X = 0) = 20{,}000 \times \frac{e^{-0.02}(0.02)^0}{0!} = 19604$$

(ii) Approximate no. of packets containing one defective blade

$$= f(1) = N \times P(X = 1) = 20{,}000 \times \frac{e^{-0.02}(0.02)^1}{1!} = 392.08 \approx 392$$

(iii) The expected no. of packets containing two defective blades

$= f(2) = N \times P(X = 2) = 20,000 \times \frac{e^{-0.02}(0.02)^2}{2!} = 3.9208 \approx 4.$

(iv) Expected no. of packets containing three defective blades

$= f(3) = N \times P(X = 3) = 20,000 \times \frac{e^{-0.02} \times (0.02)^3}{3!} = 0.0261 \approx 0.$

Note: [19604 + 392 + 4 + 0 = 20,000]

(v) Expected no. of packets containing at least one defective blades

$= N \times P(X \geq 1) = N \times [1 - P(0)] = 20,000 \times [1 - 0.9802] = 396$

[i.e. 20,000 – 19604 = 396]

(vi) Expected no. of packets containing at most two defective blades

$= N \times [P(X = 0) + P(X = 1) + P(X = 2)] = 19604 + 392 + 4 = 20,000.$

Example 70: The probability that an individual suffers a bad reaction from a particular injection is 0.001. Find the probability that out of 1500 individuals (a) exactly two; (b) none; (c) at least one; (d) at most two individuals will suffer a bad reaction.

[TU, BE, 2065 Chaitra]

Solution: *Here*, $p = 0.001$, $n = 1500$

$\lambda = np = 1.5$

According to Poisson's distribution

$P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}, x = 0, 1, 2, 3, \ldots$

(a) $P(X = 2) = \frac{e^{-1.5}(1.5)^2}{2!} = 0.251$

(b) $P(X = 0) = \frac{e^{-1.5}(1.5)^0}{0!} = 0.2231$

(c) $P(X \geq 1) = 1 - p(X = 0) = 1 - 0.2231 = 0.7769$

(d) $P(X \leq 2) = ?$ [Using formula]

$P(X \leq x) = F(x; \lambda) = \Sigma_{k=0}^{x} \frac{e^{-\lambda}\lambda^k}{k!}$

We get, $P(X \leq 2) = F(2; 1.5) = \Sigma_{k=0}^{2} \frac{e^{-1.5}(1.5)^k}{k!} = 0.809.$

Example 71: Over a 10- minute period, a counter records an average of 1.3 gamma particles per million second coming from a radioactive substance. To a good approximation, the distribution of the count, X, of gamma particles during the next millisecond is Poisson. Determine (a) λ, (b) the probability of one or more gamma particles, and (c) the variance.

Solution: (a) The mean of the Poisson distribution is λ, so we approximate that λ = 1.3 to agree with the long run average.

(b) $P(X \geq 1) = 1 - P(X = 0) = 1 - e^{-1.3} = 0.7275$

(c) Variance $(\sigma^2) = \lambda = 1.3$

Example 72: Consider the customers arriving at a cafeteria at an average rate of 0.3 per minute. Find the probability that

(a) exactly 2 customers arrive in a 10-minute span;

(b) 2 or more customers arrive in a 10-minute span;

(c) exactly one customer arrives in a five minute span and one customer arrives in the next 5 minute span.

Solution: (a) $\lambda = 0.3 \times d = 0.3 \times 10 = 3$

$P(X = 2) = f(2, 3) = \frac{e^{-3}3^2}{2!} = 0.2240$

(b) $\lambda = 0.3 \times d = 0.3 \times 10 = 3$, using

$P(X \leq 2) = F(x; \lambda) = \Sigma_{k=0}^{x} \frac{e^{-\lambda}\lambda^x}{x!}$, we get

$P(X \geq 2) = 1 - P(X \leq 1) = 1 - 0.199 = 0.801$

(c) For both cases $\lambda = 0.3 \times 5 = 1.5$.

Since cases are independent, so $P(X = 1) \times P(X = 1) = \frac{e^{-1.5} \times (1.5)^1}{1!} \times \frac{e^{-1.5} \times (1.5)^1}{1!}$

$= 0.3347 \times 0.3347 = .01120$

Example 73: In proof testing of circuit board, the probability that any partic diode will fail is 0.01. Suppose a circuit board contain 2000 diodes. (a) how ma diodes would you expect fail; (b) What is the approximate probability that at le four diode will fail on randomly selected boards ?

[TU, BE, 2062 Jest

Solution: Let X denote the number of board that fail.

Here X is binomial rv with $n = 200$ and $p = 0.01$.

Since $p < 0.05$ and $n > 100$, $X \approx P(\lambda)$.

(a) Expected no. of diodes that will fail = $E(X) = \lambda = np = 2 < 10$.

(b) Since $n = 200 > 100$ *and* $np < 10$

The approximation is excellent.

So, $X \approx P(\lambda)$, using $F(x; \lambda) = \Sigma_{i=0}^{x} \frac{e^{-1}1^x}{x!}$, we get, for $x = 4$

$P(X \leq 4) = F(4; \lambda) = \Sigma_{i=0}^{4} \frac{e^{-2}2^4}{4!} = 0.0902$

Example 74: A car hire firm has two cars which it hires out day by day. The nu of demands for a car on each day is distributed as Poisson distribution with m 1.5. Calculate the proportion of days on which (a) neither car is used; (b) so demand is refused.

Solution: Let X denote the number of demands in a day therefore $X \sim P(1.5)$.

(a) Proportion of days on which neither car is used is

$P(X = 0) = \frac{e^{-1.5}(1.5)^0}{0!} = 0.2231$

(b) Since firm has only two cars so some demand will be refused if demand > 2

So, proportion of days on which some demand is refused is

$P(X > 2) = 1 - P(X \leq 2) = 1 - \{P(0) + P(1) + P(2)\}$

$= 1 - e^{-1.5}\left\{1 + 1.5 + \frac{(1.5)^2}{2!}\right\} = 0.1913.$

Example 75: A manufacturer of pins knows that on the average 3% of its producti is defective. He sells pins in boxes of 100 and guarantees that not more than 2 pin will be defective. What is the probability that a box selected at random (a) will mee the guaranteed quality? (b) will not meet the guaranteed quality?

[TU, MBS, 2063/MBA 20

Solution: Here, p = proportion of defective pins = 0.03

n = Number of pins in a box = 100

λ = Average no. of defective pins = $np = 3$.

Let, X denote the number of defective pins in a box.

(a) The box will meet the guaranteed quality if the number of defective pins in th box is not more than 2.

So, $P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2)$

$$= e^{-3}\left\{1+3+\frac{9}{3}\right\} = e^{-3} \times 8.5 = 0.4232$$

(b) The box will not meet the guaranteed quality if the number of defective pins in the box is more than 2.

So, $P(X > 2) = 1 - P(X \le 2) = 1 - 0.4232 = 0.5768$

Example 76: A distributor of bean seeds determine from extensive tests that 4% of the seeds will not germinate. He sells the seeds in packets of 100 and guaranteed 95% germination. Determine the probability that a particular packet will violate the guarantee.

Solution: Here, p = proportion of non-germinating seeds = 0.04

n = Number of seeds in a packet = 100

λ = Average no of non- germinating seeds = $np = 4$

Let, X denote the number of non germinating seeds in a packet.

Guaranteed number of germinating seeds = 95% = 95 out of 100.

So, the number of non-germinating seeds is not more that 2.

So, P(a particular packet will violate the guarantee)

$$= P(X > 5) = 1 - P(X \le 5)$$

$$= 1 - \{P(0) + P(1) + P(2) + P(3) + P(4) + P(5)\}$$

$$= 1 - e^{-4}\left\{1 + 4 + \frac{16}{2} + \frac{64}{6} + \frac{256}{24} + \frac{1024}{120}\right\} = 0.2155$$

Example 77: The probabilities of a Poisson variate taking the values 3 and 4 are equal. Calculate the probabilities of the variate taking the value of 0 and 1. *[Pokhara Uni. BE, 2001]*

Solution: Let X be a Poisson variate by question $f(3; \lambda) = f(4; \lambda)$

Or, $\dfrac{e^{-\lambda}\lambda^3}{3\,!} = \dfrac{e^{-\lambda}\lambda^4}{4\,!} \Rightarrow \lambda = 4$

The probabilities of the variate taking the value 0 and 1 = $P(X = 0$ and $1)$

$$= P(X = 0) \times P(X = 1) = f(0; \lambda) \times f(1, \lambda) = \frac{e^{-4}4^0}{0!} \times \frac{e^{-4}4^1}{1!} = 0.00134.$$

Example 78: A lottery has a very large number of tickets, one in every 500 of which entitles the purchaser to prize. Calculate the minimum number of tickets the agent must sell to have 90% chance of selling at least one prize winning ticket.

Solution: p = proportion of prize winning tickets = 1/500.

n = Number of tickets sold.

λ = Average number of prize winning tickets = np.

Let, X denote the number of prize winning tickets in the selling of n tickets.

Since, the probability of at least one prize winning tickets = 90%

$\Rightarrow \quad P(X \ge 1) = 0.90 \qquad \Rightarrow (1 - P(X < 1) = 0.90$

$\Rightarrow \quad P(X < 1) = 0.10 \qquad \Rightarrow P(X = 0) = 0.10$

$\therefore \quad \dfrac{e^{-\lambda}\lambda^0}{0!} = 0.10 \qquad \Rightarrow e^{-\lambda} = 0.10$

Taking *log* of both sides we get

$log\, e^{-\lambda} = log\,(0.10)$

$\Rightarrow -\lambda = -2.3026 \quad \Rightarrow \lambda = 2.3026$

Hence $np = \lambda \Rightarrow n = \dfrac{\lambda}{p} = \dfrac{2.3026}{1/500} = 1151.3 \simeq 1151.$

Example 79: An insurance company finds that 0.005% of the population dies from a certain kind of accident each year. What is the probability that the company must pay off no more that 3 of 10,000 insured risks against such accidents in a given year.

Solution: $p = 0.005\% = 0.00005$, $n = 10{,}000$; $\quad \lambda = np = 0.5$

Let X denote the number of pay off's against insured risks.

Now $P(X > 3) = 1 - P(X \le 3)$

$$= 1 - [P(0) + P(1) + P(2) + P(3)] = 1 - 0.9984 = 0.0016$$

Example 80: Fit a Poisson distribution to the following data: *[TU MBS 2057]*

No. of defects	4	3	2	1	0
Frequency	2	3	19	65	11

Solution: Let X denote the number of defects.

X	4	3	2	1	0	Total
F	2	3	19	65	11	$N = 100$
fX	8	9	38	65	0	$\Sigma fX = 120$

Now $\lambda = \bar{X} = \dfrac{\Sigma fx}{N} = \dfrac{120}{100} = 1.2$

Expected frequency is given by

$$f(x) = N\,P(X = x) = N \times \frac{e^{-\lambda}\lambda^x}{x!}, x = 0, 1, 2, \ldots.$$

$$= 100 \times \frac{e^{-1.2}(1.2)^x}{x\,!}, x = 0, 1, 2, 3, 4.$$

Fitting of Poisson Distribution

X	$f(x) = 100 \times e^{-1.2}(1.2)^x$
0	$f(0) = 30.1194 \simeq 30$
1	$f(1) = 36.1433 \simeq 36$
2	$f(2) = 21.6859 \simeq 22$
3	$f(3) = 8.6744 \simeq 9$
4	$f(4) = 2.6023 \simeq 3$

The fitted Poisson distribution is

No. of defects	0	1	2	3	4	Total
Expected frequency	30	36	22	9	3	$N = 100$

Example 81: A discrete random variable X follows Poisson distribution. Find the values of (a) $P(X =$ at least 1) (b) $P(X \le 2|X \ge 1)$

Solution: Given $E(X) = 2.20$, $e^{-2.20} = 0.1108$

(a) $P(X =$ at least 1)

$= P(X \ge 1) = 1 - P(X = 0) = 1 - e^{-\lambda} = 1 - e^{-2.20} \qquad [\because \lambda = E(X)]$

$= 1 - 0.1108 = 0.8892$

(b) $P(X \le 2|X \ge 1) = \dfrac{P(X \le 2) \cap (X \ge 1)}{P(X \ge 1)} \qquad [\because P(A \cap B) = P(A|B)\,P(B)]$

$$= \frac{P(X = 1) + P(X = 2)}{1 - P(X < 1)} = \frac{\frac{e^{-\lambda}\lambda^1}{1!} + \frac{e^{-\lambda}\lambda^2}{2!}}{1 - P(X = 0)} = \frac{e^{-2.20}\lambda + e^{-2.20}\frac{(2.20)^2}{2}}{1 - e^{-2.20}}$$

$$= \frac{0.5119}{0.8892} = 0.5757$$

3.12 Negative Binomial Distribution

3.12.1 Introduction

For the Poisson distribution the mean and variance are equal but for the binomial distribution the mean is always greater than the variance. However, observable phenomena give rise to empirical discrete distribution which show a variance larger than the mean. For example, the frequency distributions of plant density obtained by quadrant sampling when the clustering of plants makes the simple Poisson model inapplicable. In such case we use Negative binomial distribution which provide an excellent model because it has a variance larger than the mean. Bacterial clustering (or contagion), e.g., deaths of insects, number of insect bites; and bating score in games lead to negative binomial distribution. This distribution also arises in inverse sampling from a binomial population or as a weighted average of Poisson distribution. Sometimes this distribution is known as **Pascal distribution** after the French mathematician ***Blaise Pascal*** (1623–62) because he contributed to develop the mathematical model.

Negative binomial distribution basically depends upon the binomial distribution but the last trial must be success in negative binomial experiment. The negative binomial experiment consists of independent trials which are repeated until a desirable number of successes occurs, with the probability of success that remain constant at each trial.

In fact, a negative binomial distribution is the converse of a binomial distribution. In binomial distribution, the number of trials is fixed and the number of successes (or failures) to occur is random variable but in negative binomial distribution the number of successes is fixed and the number of trials is a random variable. Therefore, simpler problems involving negative binomial situation can be translated into problems solvable by the binomial. For example, the probability that 30 throws of a die will suffice to generate 5 even numbers is the same as the probability that at least 5 even numbers will appear in 30 throws.

3.12.2 Conditions of using Poisson distribution

A negative binomial situation arises when

1. The experiment consists of a sequence of independent trials.
2. Each trial has only two possible outcomes: a 'success' or a 'failure'.
3. The probability of success is constant from trial to trial.
4. The experiment continues (trials are performed) until a total of r successes have been observed.
5. We count the number of trials before the r^{th} success occurs.

Remarks:

1. The rv of interest is, X = the number of failures that precede the r^{th} success, X is called ***negative binomial random variable*** because, in contrast to the binomial random variable, the number of successes is fixed and the negative binomial random variable X, the number of failures before the r^{th} success, is found by working out the probability of the r^{th} success at trial number $x + r$.
2. Suppose there are x failures preceding the r^{th} success in $x + r$ trials. Here the last trial must be a success, whose probability is p. In the remaining $(x + r - 1)$ trials we must have $(r - 1)$ successes whose probability is given by binomial probability law by the expression
$$\binom{x+r-1}{r-1} p^{r-1} q^{(x+r-1)-r-1} = \binom{x+r-1}{r-1} p^{r-1} q^{x}$$

Since the last probability must be success whose probability is p. By the compound probability theorem, the probability of r successes out of x+r trials in which the last trial (rth trial) is success, is given by

$$P(X=x) = \binom{x+r-1}{r-1} p^{r-1} q^{x} p = \binom{x+r-1}{r-1} p^{r} q^{x}$$

or, $$P(X=x) = \binom{x+r-1}{x} p^{r} q^{x}$$

$$\therefore\ C(n, r) = C(n, n-r) \Rightarrow \binom{x+r-1}{r-1} = \binom{x+r-1}{x}$$

Definitions:

Negative binomial random variable: In a series of Bernoulli trials (independent trials with probability p of a success), let the random variable X denote the number of trials until r successes occur. Then X is a ***negative binomial random variable*** with parameters p and r.

Probability mass function: The *pmf* of the negative binomial random variable X with parameters r = number of success and p = probability of success, is

$$P(X=x) = nb(x\,;r, p) = \binom{x+r-1}{r-1} p^{r} q^{x}\ ;\ x = 0, 1, 2, \ldots.$$

Negative binomial distribution: A discrete rv X is said to have a ***negative binomial distribution*** with parameters r and p if its *pmf* is given by

$$P(X=x) = nb(x\,;r, p) = \binom{x+r-1}{r-1} p^{r} q^{x}\ ;\ x = 0, 1, 2, \ldots.$$

Then mean = $E(X) = \dfrac{r(1-p)}{p}$

Variance = $V(X) = \dfrac{r(1-p)}{p^{2}}$

Since $\binom{x-1}{r-1} p^{r} (1-p)^{x-r}$ is equivalent to $\binom{x+r-1}{r-1} p^{r} q^{x}$ we can define above terms as follows:

Definitions: In a series of Bernoulli trials (independent trials with constant probability p of a success), let the random variable X denote the number of trials until r successes occur. Then X is a ***negative binomial random variable*** with parameters $0 < p < 1$ and r = 1, 2 3, . . . , and *pmf*

$$P(X=x) = nb(x\,;r, p) = \binom{x-1}{r-1} p^{r} (1-p)^{x-r},\ x = r, r+1, r+2, \ldots$$

The distribution *of* X is with this *pmf* called ***negative binomial distribution.***

In the special case $r = 1$, the *pmf* is $nb\,(x\,;1,p) = pq^{x};\ x = 0, 1, 2, \ldots$

Note:

1. Since $\binom{x+r-1}{r-1} = (-1)^{x}\binom{-r}{x}$

   $P(X = x) = \binom{-r}{x} p^{r}(-q)^{x}$ which is the $(x + 1)^{th}$ term in expansion of $(1-q)^{-r}$, a binomial expansion with a negative index. So this distribution is known as negative binomial distribution.

2. Since $\Sigma_{x=0}^{\infty} P(X=x) = p^{r}\, \Sigma_{x=0}^{\infty} \binom{-r}{x} (-q)^{x} = p^{r}\,(1-q)^{-r} = p^{r}\, p^{-r} = 1$

   and $P(X=x) \geq 0$. So, $p(X=x) = \binom{-r}{x}\ p^{r}(-q)^{x}$ is *pmf*.

3. Negative binomial distribution tends to Poisson distribution as $p \to 0$, or $r \to \infty$ such that $rp = \lambda$ (finite). If X is negative binomial *rv* with *pmf*

$$nb(x\,;r,p) = \binom{x+r-1}{r-1} p^r q^x, x = 0, 1, 2, \ldots$$

4. We shall use the notation $X \sim B^-(x\,;r,p)$ to denote the *rv* X follows negative binomial distribution with parameters r and p.
5. If $r = 1$, the negative binomial *pmf* if

$$P(X = x) = nb(x\,;r,p) = pq^x, x = 0, 1, 2, \ldots$$

which is the probability function of geometric distribution. Hence, negative binomial distribution is regarded as the generalization of geometric distribution.

Example 82: The probability is 0.40 a child exposed to certain contagious disease will catch it. What is the probability that the tenth child exposed to disease will be third to catch it?

Solution: Here $x + r = 10$, $r = 3$, $p = 0.40$

Let X be a random variable which denote the number of children exposed to a certain contagious disease will not catch them.

Using, $P(X = x) = nb(x; r, p) = \binom{x+r-1}{r-1} p^r q^x$

We get $nb(7; 3, 0.40) = \binom{9}{2} (0.40)^3 (0.60)^7 = 0.0645$

Note: 1. When the table of binomial probability is available, the determination of negative binomial probabilities can generally be simplified by making use of

the identity $\quad nb(x; ,r\,p) = \dfrac{r}{x+r} b(r; x+r, p)$

which is equivalent to $\quad nb(x; r, p) = \binom{x+r-1}{r-1} p^r q^x; x = 0, 1, 2, 3, \ldots$

Solution: $nb(x; r, p) = \dfrac{r}{x+r} b(r; x+r, p)$

$$\Rightarrow \quad nb(7; 3, 0.40) = \frac{3}{10} b(3; 10, 0.40) = \frac{3}{10}[B(3; 10, 0.40) - B(3-1; 10, 0.40)]$$

$$= \frac{3}{10}[0.3823 - 0.1673] \qquad \text{[Using Table]}$$

$$= \frac{3}{10} \times 0.215 = 0.0645.$$

2. We can show that

$$nb(x; r, p) = \binom{x-1}{r-1} p^r (1-p)^{x-r} \qquad \ldots (i)$$

is equivalent to

$$nb(x; r, p) = \binom{x+r-1}{r-1} p^r q^x \qquad \ldots (ii)$$

Using (i) we can solve above question.

So, $nb(10; 3, 0.40) = \binom{9}{2} (0.40)^3 (0.60)^7 = 0.0645$

Also, $nb(x; r, p) = \dfrac{r}{x} b(r; x, p)$ is equivalent to (ii)

So, $nb(x; r, p) = \dfrac{3}{10} b(3; 10, 0.40) = \dfrac{3}{10} \times 0.215 = 0.0645$

Example 83: If a boy is throwing stone at a target what is the probability that his 10th throw is his 5th hit, if the probability of hitting the target at any trial is 0.5.

[Pokhara Uni. BE, 2009 Fall]

Solution: Here, p = probability of hitting the target in each trial = 0.5

$x + r$ = total number throws = 10

r = number of successes (i.e. hits) = 5.

Let X, denote the number of not hitting hits the target. The probability that his 10th throw is his 5th hit, is obtained by using

$$P(X = x) = nb(x; r, p) = \binom{x+r-1}{r-1} p^r q^x$$

$$\text{So, } P(X = 5) = nb(5; 5, 0.5) = \binom{10-1}{5-1} (0.5)^5 (1-0.5)^5 = 0.1230.$$

$$\text{Or, } \quad nb(x; r, p) = \binom{-r}{x} p^r (-q)^x$$

$$\Rightarrow \quad P(X = 5) = nb(5; 5, 0.5) = \binom{-5}{5} (0.5)^5 (-0.5)^5 = 0.1230$$

$$\text{Mean} = E(X) = \frac{r(1-p)}{p} = \frac{5(1-0.5)}{0.5} = 5, \; V(X) = 10$$

Example 84: A pediatrician wishes to recruit 5 couples, each of whom is expecting their first child, to participate in a new natural child birth regimen, let $p = P(a$ randomly selected couple agrees to participate). If $p = 0.2$, what is the probability that 15 couples must be asked before 5 are found who agree to participate? That is, with S = {agrees to participate}, what is the probability that 10 F's occur before the fifth S?

Solution: Here, p = probability that a selected couple agrees to participate = 0.2

r = number of couples who agree to participate = 5

$x + r$ = total number couples = 15, $x = 15 - 5 = 10$

$$\text{So, } nb(x; r, p) = \binom{x+r-1}{r-1} p^r q^x$$

$$\Rightarrow \quad nb(10; 5, 0.2) = \binom{14}{4} (0.2)^5 (0.8)^{10} = 0.0344.$$

Probability that at most 10 F's are observed (at most 15 couples are asked) is

$$P(X \le 10) = \Sigma_{x=0}^{10}\; nb(x; 5, 0.2) = \Sigma_{x=0}^{10} \binom{x+5-1}{5-1} (0.2)^5 (0.8)^x$$

$$= (0.2)^5\, \Sigma_{x=0}^{10} \binom{x+4}{4} (0.8)^x = 0.164$$

Example 85: An item is produced in large number. The machine is known to produce 5% defective. A quality control inspector is examining the items by taking them at random. What is the probability that at least 4 items are to be examined in order to get 2 defective? Also find the mean and variance. ***[Pokhara uni, BE, 2001]***

Solution: Here, p = probability of producing defective items = 5% = 0.05

$x + r$ = total number of trials = 4, r = number of successes = 2, $x = 4 - 2 = 2$.

If 2 defective items are to be obtained then it can happen in 2 or more trials. The required probability of $P(X \ge x)$ is,

$$P(X + r \ge 4) = P(X \ge 2) = 1 - P(X < 2) = 1 - [P(X = 0) + P(X = 1)]$$

$$= 1 - \left[\binom{-2}{0} (0.05)^2 + \binom{-2}{1} [0.05]^2 (-0.95)\right] = 1 - 0.007 = 0.993$$

$$or, \quad P(X + r \ge 4) = P(X \ge 2) = 1 - P(X < 2) = 1 - \{P(X = 0) + P(X = 1)\}$$

$$= 1 - \left\{\binom{0+2-1}{2-1} (0.05)^2 (0.95)^0 + \binom{1+2-1}{2-1} (0.05)^2 (0.95)^1\right\}$$

$= 1 - 0.0025 - 0.00475 = 0.99275$

Mean $= E(X) = \frac{r(1-p)}{p} = \frac{2(1-0.05)}{0.05} = 38$, $V(X) = \frac{r(1-p)}{p^2} = 760.$

Example 86: A student has taken a 5 answer multiple choice examinations orally. He continues to answer questions until he get 3 correct answers. What is the probability that he gets them at most 7 questions if he guesses at each answer?

Solution:

Here, p = the probability of giving a right answer in each question = 1/5 = 0.2.

$x + r$ = the maximum number of questions = 7,

r = required number of correct answers = 3.

Let X be a rv which denote the number of wrong answers of given questions.

The required probability is given by $P(X \le x) = P(X \le 4)$

$$= P(X=0) + P(X=1) + P(X=2) + P(X=3) + P(X=4)$$

$$= \binom{0+3-1}{3-1}(0.2)^3(0.8) + \binom{1+3-1}{3-1}(0.2)^3(0.8)^1 + \binom{2+3-1}{3-1}(0.2)^3\ (0.8)^2$$

$$+ \binom{3+3-1}{3-1}(0.2)^3(0.8)^3 + \binom{4+3-1}{3-1}(0.2)^3(0.8)^4$$

$$= 0.008 + 0.0192 + 0.03072 + 0.04096 + 0.049152 = 0.14803$$

Example 87: Let X, the number of plants of a certain species found in a particular region, have this distribution with $p = 0.3$ and $r = 3$. What is $P(X = 4)$. What is the probability that at least on plant is found?

Solution: $P(X = x) = \binom{x+r-1}{r-1} p^r q^x,\ x = 0, 1, 2, \ldots.$

$$\Rightarrow \quad P(X=4) = \binom{4+3-1}{3-1}(0.3)^3(0.7)^4 = 0.0972$$

The probability that at least one plant is found is

$$P(X \ge 1) = 1 - P(X < 1) = 1 - P(X = 0)$$

$$= 1 - \binom{0+3-1}{3-1}(0.3)^3(0.7)^0 = 1 - 0.027 = 0.973.$$

Example 88: A Web site contains three identical computer servers. Only one is used to operate the site, and the other two are spares that can be activated in case the primary system fails. The probability of a failure in the primary computer (or any activated spare system) from a request for service is 0.0005. Assuming that each request represents an independent trial, what is the mean number of requests until failure of all three servers? What is the probability that all three servers fail within five requests?

Solution: Let X denote the number of requests until all three servers fail, and let X_1, X_2, and X_3 denote the number of requests before a failure of the first, second, and third servers used, respectively. Now, $X = X_1+X_2+X_3$. Also, the requests are assumed to comprise independent trials with constant probability of failure $p = 0.0005$. Furthermore, a spare server is not affected by the number of requests before it is activated. Therefore, X has a negative binomial distribution with $p = 0.0005$ and $r =$ 3. Consequently, $E(X) = 3/\ 0.0005 = 6000$ requests.

The probability is $P(X \le 5)$ and

$$P(X \le 5) = P(X=3) + P(X=4) + P(X=5)$$

$$= 0.0005^3 + \binom{3}{2}(0.0005)^3\ (0.9995) + \binom{4}{2}\ (0.0005)^3(0.9995)^2$$

$$= 1.25\ (10)^{-10} + 3.75\ (10)^{-10} + 7.49\ \ (10)^{-10}$$

$$= 1.249 \times (10)^{-9}$$

Exercise 3

Theoretical Questions

1. Define random variable. Discuss the difference between discrete and continuous random variables with suitable example. *[TU, BE, 2057 Bhadra/2063 Kartik/2065 Chaitra/2067 Shrawan/BIE 2066 Magh/Purb. Uni, 2006 BE; 2061 Ashwin]*

2. Define discrete and continuous random variables and determine whether each of the following variate is discrete or continuous:

a. (i) The number of defective ball point pens in each carton of twelve
   (ii) The number of interruptions (breakdowns) per day at a computer facility.
   (iii) The distance required for stopping an automobile travelling at 40kph.
   (iv) The number of commercial loans processed per day at a bank.
   (v) The volume of orange juice in each one-liter container. *[TU, BE, 2058 Shrawan]*

b. (i) Range of breakdown voltage of diode.
   (ii) The number of interruptions (breakdowns) per day at a computer facility.
   (iii) Breaking strength of the cable.
   (iv) Number of students admitted in a college.
   (v) Voltage range of automatic voltage guard. *[TU, BE, 2063 Ashadh]*

3. Identify the following as discrete or continuous random variable :
   (i) Increase in length of life attained by a cancer patient as a result of surgery.
   (ii) Quality of ball- pens sold by a stationary store per day during a month.
   (iii) Blood pressure measured of IOE students in the last medical camp.
   (iv) The applications registered for commercial loan during the first fifteen days in the month of June 2008, at Nabil Bank.
   (v) The attendance of the students in the first period of a day's class in various engineering units. *[TU, BE, 2065 Chaitra]*

4. Define discrete random variable, probability distribution and probability distribution function. Support your answer by any suitable example. *[TU, BE, 2062 Chaitra]*

5. Define discrete and continuous random variables and their mean and variance. *[TU, BE, 2066 Magh]*

6. Define discrete random variable and its probability distribution. Explain with two suitable examples. *[TU, BE, 2064 Poush/2063 Kartik]*

7. Classify the following random variable as discrete or continuous *[TU, BIE, 2068 Bhadra]*
   X : the number of automobile accidents per year in Kathmandu.
   Y : the length of time to play 18 holes of golf.
   Z : the amount of milk produced yearly by a particular cow.
   M : the number of eggs laid each month by hen.
   O : the number of building permits issued each month in certain city.
   P : the weight of gain produced per acre.

8. Define mathematical expectation of a random variable. Show that mathematical expectation of a random variable is an arithmetic mean of a random variable. *[Purbanchal Uni, BE, 2005]*

9. State the conditions which must be fulfilled for using the binomial distribution. Also state the parameters of binomial distribution. *[TU, BE, 2056 Bhadra/2058 Shrawan/2067 Shrawan]*

10. Discuss the difference between binomial distribution and hypergeometric distribution. Give suitable example for each distribution. *[TU, BE, 2061 Ashwin]*

11. Write differences between binomial distribution and negative binomial distribution with suitable examples. *[TU, BE, 2062 Baisakh]*

12. Write similarities and differences between Binomial distribution and Negative Binomial distribution. *[TU, BE, 2062 Poush]*

13. Define negative binomial distribution and state the condition of applicability for it. How does negative binomial distribution differ from binomial distribution ? ***[TU, BE, 2065 Kartik/2063 Ashad/2062 Bhadra]***

14. Define hypergeometric distribution. How does it differ from binomial distribution? ***[TU, BE, 2065 Chaitra/BIE 2068 Bhadra]***

15. Define Bernoulli random variable. Discuss the limiting case in which Binomial distribution becomes a Poisson distribution. ***[TU, BE, 2065 Chaitra (Re)]***

16. Write the conditions of negative binomial distribution and define probability distribution function of negative binomial distribution. ***[TU, BE, 2065 Chaitra (Re)/ 2058 Shrawan/2057 Bhadra]***

17. Define binomial distribution. State the conditions underlying binomial distribution. ***[TU, BE, 2065 Chaitra/2068 Magh]***

18. What are the conditions for binomial experiment and also define the Binomial distribution. Write one example of Binomial experiment. ***[TU, BE, 2066 Magh]***

19. Define Hypergeometric distribution. Describe the conditions for Hypergeometric distribution. ***[TU, BE, 2067 Mangsir]***

20. Define the limiting case of Poisson distribution as Binomial distribution. ***[TU, BE, 2067 Mangsir/2062 Bhadra/2062 Jestha/2064 Shrawan/2063 Kartik]***

21. Discuss the properties of Poisson distribution. ***[TU, BE, 2064 Poush]***

22. Compare Poisson distribution with Binomial distribution.

23. Discuss the difference between hypergeometric distribution and negative binomial distribution. ***[TU, BE, 2063 Kartik]***

24. Define hypergeometric distribution. Write down the properties of hyper geometric distribution. ***[Purbanchan Uni. BE, 2005]***

25. Discuss the differences between binomial and hypergeometric distribution. Give one example of hypergeometric distribution and also state the important properties of hypergeometric distribution ***[TU BE, 2068 Bhadra]***

26. Define Poisson probability distribution with the condition for Poisson distribution ***[TU BE, 2068 Bhadra/2068 Magh]***

Numerical Problems

1. What do you mean by probability distribution? Check whether the following can define probability distribution, and explain your answer: ***[TU, BE, 2067 Mangsir]***

(i) $f(x) = \frac{x}{5}$ for $x = 0, 1, 2, 3, 4, 5$; (ii) $f(x) = \frac{5 - x^2}{6}$ for $x = 0, 1, 2, 3$.

[Ans : (i) No, because sum exceeds 1; (ii) No, because $f(3)$ is negative]

2. The discrete *rv* X has *pmf* as shown:

$X = x$	0	1	2	3
$P(X = x)$	0.26	0.50	0.22	0.02

Then (i) draw probability bar chart histogram.

(ii) Find distribution function of *rv* X.

[Ans : (ii)]

$X = x$	0	1	2	3
$P(X = x)$	0.26	0.50	0.22	0.02
$F(X \leq x)$	0.26	0.76	0.98	1

3. The discrete *rv* X has the *pmf* as shown:

$X = x$	–3	–2	–1	0	1
$P(X = x)$	0.1	0.25	0.3	0.15	d

Find **(a)** d; **(b)** $P(-3 \leq X \leq 0)$; **(c)** $P(X > -1)$; **(d)** mode. Also construct the *cdf* of $F(x)$. [Ans : (a) 0.2; (b) 0.65; (c) 0.35; (d) –1]

4. Compute variance of X when X represents the outcome when we roll a fair die. [Ans : 2.92]

5. The probability distribution of a *rv* X is given by

$X = x$	–2	3	1
$P(X = x)$	1/3	1/2	1/6

Find (i) $E[2X + 5]$ (ii) $E[X^2]$ [Ans : (i) 7; (ii) 6]

6. Find the expected sales of Toyota car in Kathmandu city in a week from the following information: ***[TU 2057]***

Days	Sun	Mon	Tue	Wed	Thur	Fri
Sales	90	60	90	50	75	55
Probability	0.25	0.18	0.12	0.05	0.2	0.2

[Ans : 72.6]

7. The information available regarding the frequency distribution of a set of outcomes are

X	0	10	15	20	25
Frequency	10	30	40	15	5

[Ans: 3.25]

8. An industrial salesman wants to know the average number of units he sells per sales call. He checks his past records and comes up with the following probabilities :

Sales in units	0	1	2	3	4	5
Probability	0.15	0.20	0.1	0.05	0.30	0.20

What is the average number (i.e. expectation) of units he sells per sales call? [Ans : 2.75]

9. Define mathematical expectation of a random variable. A dice is rolled at once. What is the expectation of numbers in it? [Ans : 3.5] ***[Purbanchal Uni, BE, 2063]***

10. Define a probability distribution. Let X be the number of tires randomly selected that are under inflated. Which of the following p(x) is *pmf* of X and why are other two not allowed?

	X	0	1	2	3	4
A	$P(x)$	0.6	0.7	0.4	0.2	0.1
B	$P(x)$	0.4	0.1	0.2	0.1	0.1
C	$P(x)$	0.3	0.5	0.1	0.05	0.05

[Ans: In C, $P(x)$ is legitimate because $\Sigma P(x) = 1$, In other two $p(x)$ is not legitimate because $\Sigma P(x) = 1$] ***[Purbanchal Uni. BE 2004]***

11.(a) A random variable X has following probability function ***[Purbanchal Uni. BE, 2005]***

X	0	1	2	3	4	5	6
$P(x)$	k	$3k$	$5k$	$7k$	$9k$	$11k$	$13k$

(i) Fnd k; (ii) Evaluate $P(X < 3)$ and $P(X \geq 3)$. [Ans : (i) 0.02; (ii) 0.18; 0.82]

(b) A random variable X has following probability function

X	3	2	1	0	–1	–2	–3
$P(x)$	0.1	0.2	$3k$	k	$2k$	0	0.1

(i) Fnd k; (ii) Also find mean and variance of X

[Ans : (i) $k = 0.1$, (ii) $E(X) = \mu = 0.5$, $V(X) = 2.85$] ***[TU BE, 2068 Bhadra]***

12. Define mathematical expectation of a discrete random variable. A probability distribution is given. ***[Pokhara Uni. BE, 2009 Fall]***

$X = x$	0	1	2	3	4	5
$P(X = x)$	0.26	0.25	0.11	0.02	0.25	0.11

Find (a) $P(X = 4)$; (b) $P(0 < X < 4)$; (c) $P(X = 4 \cap X = 5)$; (d) $F(3)$

13. The number of hardware failure of a computer system in a week of 8 operation has following *pmf*.

Number of failure	0	1	2	3	4	5	6
Probability	0.18	0.28	0.25	0.18	0.06	0.04	0.01

(i) Find the expected number of failures in a week. (ii) Find the variance of the number of failures in a week. [Ans : (i) 1.82, (ii) 5.22] *[Pokhara Uni. BE, 2003 Fall]*

Binomial Distribution

14. Human error is given as the reason for 75% of all accidents in a plant. Find the probability that human error will be given as the reason for two of the next four accidents. [Ans : $P(X = 2) = C(4, 2)(0.75)^2(0.25)^{4-2} = 27/128$]

15. If the probability is 0.40 that steam will condense in a thin walled aluminum tube at 10 atm pressure, find the probability that under the stated condition, steam will condense in 4 of 12 such tubes. [Ans : $P(X = 4) = C(12, 4)(0.40)^4(0.60)^8 = 0.2128$]

16. During one stage in the manufacture of integrated circuit chips, a coating must be applied. If 70% of chips receive a thick enough coating, find the probability that among 15 chips

(a) at least 12 will have thick enough coatings;

(b) at most 6 will have thick enough coatings;

(c) exactly 10 will have thick enough coatings;

[Ans : (a) 0.2969; (b) 0.0152 (c) 0.2016]

17. On a very long mathematics test, Ram got 70% of the items right. For a 10 item quiz calculate the probability that Ram will get (i) at least 8 items right; (ii) less than 3 items right. [Ans : (i) 0.3828; (ii) 0.0016] *[TU, MBS model 2056]*

18. It is observed that 80% of television viewers watch entertainment channel. What is the probability that at least 80% of the viewers in a random sample of five watch entertainment channel? [Ans: $P(X \geq 80\%$ of $5) = 0.7373$]

19. How many tosses of a coin are needed so that the probability of getting at least one head is 0.875? [Ans : 3]

20. Ratio of the probability of 3 successes in 5 independent trials to the probability of 2 successes in 5 independent trials is 1/4. What is the probability of 4 successes in 6 independent trials?

[**Hint:** $\frac{c(5,3)\,p^3 q^2}{c(5,2)\,p^2 q^3} = \frac{1}{4} \Rightarrow \frac{p}{4} = \frac{q}{1} \Rightarrow p = k,\ q = 4k,\ p + q = 1,\ p = \frac{1}{5},\ q = \frac{4}{5}$. So, $P(4) = c(6, 4)p^4 q^2 = 0.0154$]

21. The incidence of occupational disease in an industry is such that the workmen have a 20% chance of suffering from it. What is the probability that out of six workmen, 4 or more will contract the disease? [Ans: 0.0169]

22. The probability of a bomb hitting a target is 0.20. Two bombs are enough to destroy a bridge. If six bombs are aimed at the bridge, find the probability that the bridge is destroyed. [Ans: 0.345]

23. The probability of a man hitting a target is 1/4 If he fires 7 limes, what is the probability of his hitting the target at least twice? [Ans: 0.555]

24. Out of 800 families of Kathmandu with 4 children each, what percentage would be expected to have (a) 2 boys and 2 girls; (b) at least one boy; (c) no girls and (d) at most 2 girls. Assume equal probabilities for boys and girls.
[Ans: (a) 37.5% (b) 93.75% (c) 6.25% (d) 68.75%]

25. If the mean and variance of a binomial distribution are 3 and 3/2 find the probability of at least 4 successes. [Ans: 0.34]

26. Is the following statement correct? "The mean of the binomial distribution is 4 and its variance is 9". [Since $q = 2.25$, wrong]

27. Out of 200 families of 4 children each, how many families would you expect to have (a) 2 boys and 2 girls; (b) at least one boy; (c) no girls ; (d) 3 boys and 1 girl; (e) at most 2 girls assuming hat boys and girls are equally likely ?
[Ans : (a) 75; (b) 188; (c) 13; (d) 50; (e) : 138]

28. Calculate the theoretical frequencies from the following by using binomial probability law *[TU, MBS 2057]*

Success	5	4	3	2	1	0
Frequency	190	500	900	960	500	150

Hypergeometric Distribution

29. (i)An Interne-based company that sells discount accessories for cell phones often ships an excessive number of defective products. The company needs better control of needs better control of quality. Suppose it has 20 identical car chargers on hand but that 5 are defective. If the company decides to randomly select 10 of these items, what is the probability that 2 of the 10 will be defective?

(ii) Repeat the preceding example but with 100 car chargers, of which 25 are defective, by using (a) the formula for the hypergeometric distribution;

(b) the formula for the binomial distribution as an approximation.

[Ans: (i) 0.3483; (ii) 0.292; 0.282]

30. What is the probability that an IRS auditor will catch only two income tax returns with illegitimate deductions if she randomly selects 6 returns from among 18 returns, of which 8 contain illegitimate deductions?
[Ans: $P(X = 2) = h(2; 6, 8, 18) = 0.3167$]

31. Among the 12 solar collectors on display at a trade show, 9 are flat-plate collectors and the others are concentrating collectors. If a person visiting the show randomly selects 4 of the solar collectors to check out, what is the probability that 3 of them will be flat-plate collectors? [Ans: $P(X = 3) = h(3: 4, 9, 12) = 0.5091$]

32. Among the 16 cities that a professional society is considerfig for its next 3 annual conventions, 7 are in the western part of the United States. To avoid arguments, the selection is left to chance. If none of the cities can be chosen more than once, what are the probabilities that

(a) None of the conventions will be held in the westem part of the United States:

(b) All of the conventions will be held in the western part of the United States?

[Ans: (a) $h(0; 3,7,16) = 0.15$ (b) $h(3; 3, 7, 16) = 0.0625$]

33. A shipment of 120 burglar alarms contains 5 that are defective. If 3 of these alarms are randomly selected and shipped to a customer, find the probability that the customer will get one bad unit by using

(a) hypergeometric and (b) binomial distributions.

[Ans : (a) $h(1; 3, 5, 120) = 0.1167$; (b) $b(1; 3, 0.0417) = 0.1149$]

34. Among the 300 employees of a company, 240 are union members, while the others are not. If 8 of the employees are chosen by lot to serve on the committee which administers the pension fund, find the probability that 5 of them will be union members while the other not, by using

(a) the formula for hypergeometric distribution.

(b) the formula for binomial distribution as an approximation.

[Ans: (a) $h(5; 8, 240\ 300) = 0.1470$ (b) $b(5; 8, 0.8) = 0.1468$]

Poisson Distribution

35. Given that the switchboard of a consultant's office receives on the average 0.6 calls per minute, find the probability that (a) in a given minute there will be at least 1 call; (b) in a 4-minute interval there will be at least 3 calls.
[Ans: (a) 0.998; (b) 0.779]

36. A telephone exchange receives on an average 4 calls per minute. find the probability on the basis of Poisson distribution of (a) no calls at all; (b) two or less calls per minute; (c) upto four calls per minute; (d) more than four calls per minute.
[TU, BIE, 065 Chaitra]

37. Consider an experiment that consists of counting the number of α particles given off in a one-second interval by one gram of radioactive materials. If we know from past experience that, on the average, 3.2 such α–particles are given off, what is a good approximation to the probability that no more than 2α-particles will appear?
[Ans : $P(X \leq 2) = 0.382$]

38. A bank receives on the average $\alpha = 6$ bad checks per day, what are the probabilities that it will receive (a) 4 bad checks on any given day;
(b) 10 bad checks over any two consecutive days?
[**Hint:** (a) $x = 4, \lambda = \alpha \times T = 6 \times 1 = 6. \quad \therefore P(X = 4) = f(4; 6) = 0.134$
(c) $\lambda = 2 \times d = 12, f(10; 12) = F(10; 12) - F(9; 12) = 0.105$]

39. For the case of the thin copper wire, suppose that the number of flaws follows a Poisson distribution with a mean of 2.3 flaws per millimeter. Determine the probability of (a) exactly 2 flaws in 1 millimeter of wire;
(b) at least 2 flaws in 1 millimeter of wire; (c) at most 2 flows in 1 millimeter of wire.
[Ans : (a) 0.2652; (b) 0.6691; (c) 0.5960] *[TU, BE, 2066 Magh]*

40. In a certain factory turning out optical lenses, there is a small chance, 1/500 for any lens to be defective. The lenses are supplied in packets of 10 each. What is the probability that a packet will contain (a) no defective lens;
(b) exactly two defective lenses; (c) at least one defective lenses;
(d) at most two defective lenses. [Ans: (a) 0.9802; (b) 0.000196; (c) 0.0198; (d) 1]
[TU, BE, 2065 Chaitra/Purbanchal Uni. BE, 2005]

41. Computer terminal repair person is "beeped" each time there is a call for service. The number of beeps per hour is known to occur in accordance with Poisson distribution with mean of $\lambda = 2$ per hour. Find the probability of
(a) exactly three beeps in the next hour; *[TU, BE, 2064 Shrawan]*
(b) less than three beeps in the next hour. [Ans: (a) 0.18045; (b) 0.6767]

42. A typist made 2.8 mistakes on an average per page. Find the probability that in the page typed by him (a) there is no mistake; (b) two or less mistakes; (c) more than 10 mistakes. [Ans: (a) 0.061 (b) 0.469; (c) 0]

43. Let X denote the number of creatures of a particular type captured in a trap during a given time period. Suppose that X has a Poisson distribution with $\lambda = 4.5$, so on average trap will contain 4.5 creatures. Find the probability that (a) a trap will contain exactly five creatures; (b) a trap will contain at most five creatures. Also find the expected number and variance of the trapped creatures.
[Ans: (a) 0.1708; (b) 0.7029; $E(X) = \lambda = 2.12 = \sigma^2$]

44. If the probability that an individual suffers a bad reaction from a certain injection is 0.001, determine the probability that out of 2000 individuals (a) exactly 3; (b) more than two individuals will suffer a bad reaction. [Ans: (a) 0.1804; (b) 0.323]

45. Find the probability of having exactly 2 accidents in a week by using Poisson approximation. Given that variance of the distribution = 1. *[TU, MBS 2057]*

46. In a certain factory manufacturing razor blades, there is small chance 1/50 for any blade to be defective. The blades are in packets each containing 25 blades. Using the approximation probability distribution, calculate the approximate number of pockets containing not more than one defective blades in a consignment of 10,000 packets. [Ans : 9098]

47. In a given city, 6% of all drivers get at least one parking ticket per year. Use the Poisson approximation to the binomial distribution to determine the probabilities that among 80 drivers (randomly) chosen in this city:
(a) 4 will get at least one parking ticket in any given year;
(b) at least 3 will get at least one parking ticket in any given year;
(c) anywhere from 3 to 6, inclusive, will get at least one parking ticket in any given year. [Ans: (a) 0.182; (b) 0.857 ; (c) 0.648]

48. If 0.8% of the fuses delivered to an arsenal are defective, use the Poisson approximation to determine the probability that 4 fuses will be defective in a random sample of 400. [Hint: $\lambda = np = 400 \times 0.008 = 3.2, P(X = 4) = 0.1781$]

49. The number of gamma rays emitted per second by a certain radioactive substance is a random variable having Poisson distribution with $\lambda = 5.8$. If a recording instrument becomes inoperative when there are more then 12 rays per second, what is the probability that this instrument becomes inoperative during any given second? [Ans : 0.007]

50. The arrival of trucks at a receiving dock is a Poisson process with a mean arrival rate of 2 per hour. Find the probability that
(a) exactly 5 trucks arrive in a two- hour period;
(b) 8 or more trucks arrive in a two- hour period;
(c) exactly 2 trucks arrive in a one-hour period and exactly 3 trucks in the next one-hour period. [Ans : (a) 0.156 ; (b) 0.051 ;(c) 0.04885]

51. The number of flaws in fiber optic cable follows a Poisson process with an average of 0.6 per 100 feet. Find the probability of
(a) exactly 2 flaws in a 200 foot cable
(b) exactly 1 flaw in the first 100 feet and exactly 1 flaw in the second 100 feet.
[Ans: (a) 0.216 ; (b) 0.108]

52. If a publisher of non technical books takes great pains to ensure that its books are free of typographical errors, so that the probability of any given page containing at least one such error is 0.005 and errors are independent from page to page, what is the probability that one of its 400- page novels will contain exactly one page with errors? At most three pages with errors? [Ans: 0.270671; 0.8571]

53. A manufacture of matchstick knows that on average 2% of his production is defective. He sells matchsticks in boxes of 100 and guarantees that not more than 2 match sticks will be defective. What is the probability that a match box randomly selected will meet the guaranteed quality. [Ans: 0.6767] *[Purbanchal Uni, BE 2006]*

54. Air corporation having had just two air crashes during its first fifty years of experience wants to make the next decade 'air crash - free'. Assuming that the same trend will continue, what is the probability of the corporation meeting the target?
[Hint: for next decade, length of interval d = 1/5 [$\because$ 10 = 1/5 × 50], so $\lambda = 2 \times d = 2/5 = 0.4$
and $P(X = 0) = \dfrac{e^{-0.4}(0.4)^0}{0!} = 0.6703$]

55. Fit a Poisson distribution to the following data *[TU MBS, 2060]*

Defects	0	1	2	3	4	5
No. of pages	142	156	69	27	5	1

Ans:

Defects	0	1	2	3	4	5	*Total*
Expected frequency	147	147	74	25	6	1	$N = 400$

Additional Questions

56. Name the distribution for which
(a) Mean > variance; (b) Mean < variance; (c) Mean = variance
(d) mean = median = Mode *[Pokhara Uni. BE, 2001]*
[Ans: (a) binomial; (b) Negative Binomial ; (c) Poisson; (d) Normal]

57. Compute the mean and variance of

x	0	1	2	3	4
$f(x)$	0.05	0.20	0.45	0.20	0.10

[Hint: $\mu = \Sigma x, f(x) = 2.1$; $\sigma^2 = \Sigma x^2 f(x) - [\Sigma x f(x)]^2 = \Sigma x^2 f(x) - \mu^2 = 5.4 - (2.1)^2 = 0.990$.]

58. A manufacture of digital phone has the following probability distribution for the number of defects per phone:

x	$F(x)$
0	0.89
1	0.07
2	0.03
3	0.01

(a) Determine the probability of 2 or more defects

(b) Is a randomly selected phone more likely to have 0 defects or 1 more defects?

[**Ans : (a)** 0.04 ; **(b)** 0.]

59. From the DVDs manufactured by Sony, batches of DVDs are randomly selected and the number of defects is found for each batch as given below:

$X = x$	0	1	2	3	4
$P(x)$	0.514	0.385	0.089	0.011	0.001

(a) Identify the random variable X (discrete or continuous)

(b) If in a batch it contains 5000 CD, find the average number of defective DVD's.

[**Ans : (a)** Discrete; **(b)** 3000] *[TU, BE, 2067 Mangsir]*

60. Air America has a policy of routinely over booking flights, because post experience shows that some passengers fail to show up. The random variable X represents the number of passengers who cannot be boarded because there are more passengers than the available seats. *[TU, BE, 2066 Magh]*

X	0	1	2	3	4
$P(X = x)$	0.85	A	0.057	0.009	0.002

(a) Find the value of 'a'

(b) Find the average passengers who cannot board due to over passengers?

[**Ans: (a)** $a = 0.082$; **(b)** $\mu = E(X) = 0.321 \approx 0$]

61. A contractor is required by a country planning department to submit one, two, three, four or five forms (depending on the nature of the project) in applying for building permit. Let Y = the number of forms required of the next applicant. The probability that y forms are required is known to be proportional to y–that is $P(y) = ky$ for $y = 1, 2, 3, 4, 5$. (a) What is the value of k.

(b) What is the probability that between two and four forms (inclusive are required)?

(c) What is the probability that at most three forms are required?

[TU, BE, 2057 Bhadra]

[Hint:

$Y = y$	1	2	3	4	5
$P(y)$	k	$2k$	$3k$	$4k$	$5k$

(a) Since $\Sigma P(y) = 1$. So, $k + 2k + 3k + 4k + 5k = 1 \Rightarrow k = \frac{1}{15}$

(b) $P(2 \le Y \le 4) = 2k + 3k + 4k = 9k = \frac{9}{15} = 0.6$

(c) $P(Y \le 3) = k + 2k + 3k = 6k = \frac{6}{15} = 0.4$

62. Find the mean of the probability distribution of the number of heads obtained in 3 flips of a balanced coin.

[Hint: The probabilities for 0, 1, 2 or 3 heads are $\frac{1}{8}, \frac{3}{8}, \frac{3}{8}$ and $\frac{1}{8}$ as can easily be verified by counting equally likely probabilities or by using the formula for the binomial distribution with $n = 3$, and $p = \frac{1}{2}$.

$\therefore \mu = 0 \times \frac{1}{8} + 1 \times \frac{3}{8} + 2 \times \frac{3}{8} + 3 \times \frac{1}{8} = \frac{3}{2}$

or, Using formula $\mu = np$ we get $\mu = 3 \times \frac{1}{2} = \frac{3}{2}$

63. Find the mean of the probability distributing of the number of heads obtained in five flips of a balanced coin.

[Hint:

X	0	1	2	3	4	5
$f(x)$	1/32	5/32	10/32	10/32	5/32	1/32

$\mu = \Sigma x. f(x) = 2.5$; or $\mu = np = 5 \times 1/2 = 2.5$

$\sigma^2 = \Sigma x^2 f(x) - \mu^2 = 7.5 - 6.25 = 1.25$ or. $\sigma^2 = npq = 5 \times \frac{1}{2} \times \frac{1}{2} = 1.25$]

64. Find the mean and standard deviation of each of the following variables (having binomial distributions) :

(a) The number of heads obtained in 676 flips of a balanced coin.

(b) The number of 4's obtained in 720 rolls of a balanced die.

(c) The number of defectives in sample of 600 parts made by a machine, when the probability is 0.04 that any one of the parts is defective.

(d) The number of students among 800 interviewed who do not like the food served at the university cafeteria, when the probability is 0.65 that any one of them does not like the food. [**Ans: (a)** $\mu = 338$; $\sigma = 13$; **(b)** $\mu = 720 \times 1/6 = 12$, $\sigma = 10$; **(c)** $\mu = 24$; $\sigma = 4.8$; **(d)** $\mu = 520$; $\sigma = 13.49$]

65. If 95% of certain high–performance radial tives last at least 30,000 miles, find the mean and the standard deviation of the distribution of the number of these tires among 20 selected at random, that last at least 30,000 miles.

[**Ans** : $\mu = 19$, $\sigma = 0.9747$]

66. It it known from past experience that 80% of the patient survive a particular heart surgery. In a year 18 patients were operated upon. What is the probability that (a) 15 patients will survive; (b) at least 15 will survive (c) Find the mean and variance.

[**Ans: (a)** $b(15;18,0.8) = 0.230$; **(b)** $B(15;18,0.8) = 0.501$; **(c)** 14.4, 2.88]

67. Six coins are tossed 6400 times. Using Poisson distribution, find the approximate probability of getting six heads in r times.

[**Hint:** The probability of obtaining six heads in one throw of six coins (a single trial), is $p = \left(\frac{1}{2}\right)^6$, assuming that head and tail are equally probable.

$\lambda = np = 6400 \times \left(\frac{1}{2}\right) = 100$. So $P(X = r) = \frac{e^{-\lambda}\lambda^r}{r!} = \frac{e^{-100}(100)^r}{r!}$; $r = 0, 1, 2, \ldots$]

68. Records show that the probability is 0.00004 that a car will have a flat tire while driving through a certain tunnel. Use the formula for the Poisson distribution to approximate the probability that at least 2 of 10,000 cars passing through the tunnel will have a flat tire.

[**Hint:** $\lambda = np = 10000 \times 0.00004 = 0.4$

$P(X \ge 2) = 1 - P(X < 2) = 1 - P(0) - P(1) = 1 - e^{-0.4} - \frac{e^{-0.4} \times (0.4)}{1!} = 0.3224$

69. The number of weekly breakdowns of a computer is a random variable having a Poisson distribution with $\lambda = 0.3$. What is the probability that the computer will operate without a breakdown for 2 consecutive weeks? [**Ans** : 0.5488]

70. It has been established that the number of defective stereos produced daily at a certain plant is Poisson distribution with mean 4. Over a two–day span, what is the probability that the number of defective stereos does not exceed? [**Ans** : 0.04238]

71. Of all the Harry Porter books parched in recent year; about 60% were purchased for readers 14 or older. If 12 Harry Porter who bought books that year are surveyed, find the probability that (a) At least five of them are 14 or older; (b) Exactly nine of them are 14 or older; (c) Less than 3 of them are 14 or older;

[Ans: (a) $P(X \geq 5) = 1 - P(X < 5) = 1 - B(4;\ 12,\ 6) = 1 - 0.0573 = 0.9427$.

(b) $P(X = 9)\ b(9;\ 12, 0.6) = B(9;\ 12, 0.6) - B(8;\ 12, 0.6) = 0.9166 – 0.7747 = 0.1419$.

(c) $P(X < 3) = P(X \leq 1;\ 12, 0.6) = 0.0028$] *[TU, BE 2064 Poush]*

72. From the past experience it is known that in a certain intersection, there are on average 4 traffic accidents per week. Find the probability that, in a given week there will be (i) Less than 2 accidents. (ii) Exactly 2 accidents (iii) More than two accidents.[Ans:(i) 0.09158; (ii) 0.1465; (iii) 0.7619] *[TU,BE,2068Bhadra]*