UNIT 5 RANDOM VARIABLES

Structure

5.1 INTRODUCTION

In the previous units, we have studied the assignment and computation of probabilities of events in detail. In those units, we were interested in knowing the occurrence of outcomes. In the present unit, we will be interested in the numbers associated with such outcomes of the random experiments. Such an interest leads to study the concept of random variable.

In this unit, we will introduce the concept of random variable, discrete and continuous random variables in Sec. 5.2 and their probability functions in Secs. 5.3 and 5.4.

Objectives

A study of this unit would enable you to:

- define a random variable, discrete and continuous random variables;
- specify the probability mass function, i.e. probability distribution of discrete random variable;
- specify the probability density function, i.e. probability function of continuous random variable; and
- define the distribution function.

5.2 RANDOM VARIABLE

Study related to performing the random experiments and computation of probabilities for events (subsets of sample space) have been made in detail in the first four units of this course. In many experiments, we may be interested in a numerical characteristic associated with outcomes of a random experiment. Like the outcome, the value of such a numerical characteristic cannot be predicted in advance.

For example, suppose a die is tossed twice and we are interested in number of times an odd number appears. Let X be the number of appearances of odd number. If a die is thrown twice, an odd number may appear ‘0’ times (i.e. we may have even number both the times) or once (i.e. we may have odd number in one throw and even number in the other throw) or twice (i.e. we may have odd number both the times). Here, X can take the values 0, 1, 2 and is a variable quantity behaving randomly and hence we may call it as 'random variable'. Also notice that its values are real and are defined on the sample space

{(1, 1), (1, 2), (1, 3), (1, 4), (1, 5), (1, 6),
(2, 1), (2, 2), (2, 3), (2, 4), (2, 5), (2, 6),
(3, 1), (3, 2), (3, 3), (3, 4), (3, 5), (3, 6),
(4, 1), (4, 2), (4, 3), (4, 4), (4, 5), (4, 6),
(5, 1), (5, 2), (5, 3), (5, 4), (5, 4), (5, 6),
(6, 1), (6, 2), (6, 3), (6, 4), (6, 5), (6, 6)}

i.e.

$$X = \begin{cases} 0, \text{ if the outcome is } (2,2),(2,4),(2,6),(4,2),(4,4),(4,6),(6,2), \\ \qquad\qquad\qquad\qquad (6,4),(6,6) \\ 1, \text{ if the outcome is } (1,2),(1,4),(1,6),(2,1),(2,3),(2,5),(3,2), \\ \qquad\qquad\qquad\qquad (3,4),(3,6),(4,1),(4,3),(4,5),(5,2),(5,4), \\ \qquad\qquad\qquad\qquad (5,6),(6,1),(6,3),(6,5) \\ 2, \text{ if the outcome is } (1,1),(1,3),(1,5),(3,1),(3,3),(3,5),(5,1), \\ \qquad\qquad\qquad\qquad (5,3),(5,5) \end{cases}$$

So, $P[X = 0] = \frac{9}{36} = \frac{1}{4}, P[X = 1] = \frac{18}{36} = \frac{1}{2}, P[X = 2] = \frac{9}{36} = \frac{1}{4}$,

and $P[X = 0] + P\left[X = 1\right] + P\left[X = 2\right] = \frac{1}{4} + \frac{1}{2} + \frac{1}{4} = 1.$

Observe that, a probability can be assigned to the event that X assumes a particular value. It can also be observed that the sum of the probabilities corresponding to different values of X is one.

So, a random variable can be defined as below:

Definition: A random variable is a real-valued function whose domain is a set of possible outcomes of a random experiment and range is a sub-set of the set of real numbers and has the following properties:

i) Each particular value of the random variable can be assigned some probability

ii) Uniting all the probabilities associated with all the different values of the random variable gives the value 1(unity).

Remark 1: We shall denote random variables by capital letters like X, Y, Z, etc. and write r.v. for random variable.

5.3 DISCRETE RANDOM VARIABLE AND PROBABILITY MASS FUNCTION

Discrete Random Variable

A random variable is said to be discrete if it has either a finite or a countable number of values. Countable number of values means the values which can be arranged in a sequence, i.e. the values which have one-to-one correspondence with the set of natural numbers, i.e. on the basis of three-four successive known terms, we can catch a rule and hence can write the subsequent terms. For example suppose X is a random variable taking the values say 2, 5, 8, 11, … then we can write the fifth, sixth, … values, because the values have one-to-one correspondence with the set of natural numbers and have the general term as $3n - 1$ i.e. on taking n = 1, 2, 3, 4, 5, … we have 2, 5, 8, 11, 14,…. So, X in this example is a discrete random variable. The number of students present each day in a class during an academic session is an example of discrete random variable as the number cannot take a fractional value.

Probability Mass Function

Let X be a r.v. which takes the values x_1, x_2, ... and let $P[X = x_i] = p(x_i)$. This function $p(x_i)$, i =1,2, … defined for the values x_1, x_2, … assumed by X is called probability mass function of X satisfying $p(x_i) \geq 0$ and $\sum_i p(x_i) = 1$.

The set $\{(x_1, p(x_1)), (x_2, p(x_2)), ...\}$ specifies the probability distribution of a discrete r.v. X. Probability distribution of r.v. X can also be exhibited in the following manner:

X	x_1	x_2	$x_3 \dots$
$p(x)$	$p(x_1)$	$p(x_2)$	$p(x_3)\dots$

Now, let us take up some examples concerning probability mass function:

Example 1: State, giving reasons, which of the following are not probability distributions:

(i)

X	0	1
$p(x)$	$\frac{1}{2}$	$\frac{3}{4}$

(ii)

X	0	1	2
$p(x)$	$\frac{3}{4}$	$-\frac{1}{2}$	$\frac{3}{4}$

(iii)

X	0	1	2
p(x)	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

(iv)

X	0	1	2	3
p(x)	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{1}{4}$	$\frac{1}{8}$

Solution:

(i) Here $p(x_i) \geq 0$, i = 1, 2; but

$$\sum_{i=1}^{2} p(x_i) = p(x_1) + p(x_2) = p(0) + p(1) = \frac{1}{2}+\frac{3}{4}=\frac{5}{4}>1.$$

So, the given distribution is not a probability distribution as $\sum_{i=1}^{2} p(x_i)$ is greater than 1.

(ii) It is not probability distribution as $p(x_2) = p(1) = -\frac{1}{2}$ i.e. negative

(iii) Here, $p(x_i) \geq 0$, i = 1, 2, 3

and $\sum_{i=1}^{3} p(x_i) = p(x_1)+p(x_2)+p(x_3) = p(0)+p(1)+p(2) = \frac{1}{4}+\frac{1}{2}+\frac{1}{4}=1.$

$\therefore$ The given distribution is probability distribution.

(iv) Here, $p(x_i) \geq 0$, i = 1, 2, 3, 4; but

$$\sum_{i=1}^{4} p(x_i) = p(x_1) + p(x_2) + p(x_3) + p(x_4)$$

$$= p(0) + p(1) + p(2) + p(3) = \frac{1}{8}+\frac{3}{8}+\frac{1}{4}+\frac{1}{8}=\frac{7}{8}<1.$$

$\therefore$ The given distribution is not probability distribution.

Example 2: For the following probability distribution of a discrete r.v. X, find

i) the constant c,

ii) $P[X \leq 3]$ and

iii) $P[1 < X < 4]$.

X	0	1	2	3	4	5
p(x)	0	c	c	2c	3c	c

Solution:

i) As the given distribution is probability distribution,

$$\therefore \sum_i p(x_i) = 1$$

$$\Rightarrow 0 + c + c + 2c + 3c + c = 1 \Rightarrow 8c = 1 \Rightarrow c = \frac{1}{8}$$

ii) $P[X \le 3] = P[X = 3] + P[X = 2] + P[X = 1] + P[X = 0]$

$$= 2c + c + c + 0 = 4c = 4 \times \frac{1}{8} = \frac{1}{2}.$$

iii) $P[1 < X < 4] = P[X = 2] + P[X = 3] = c + 2c = 3c = 3 \times \frac{1}{8} = \frac{3}{8}.$

Example 3: Find the probability distribution of the number of heads when three fair coins are tossed simultaneously.

Solution: Let X be the number of heads in the toss of three fair coins.

As the random variable, "the number of heads" in a toss of three coins may be 0 or 1 or 2 or 3 associated with the sample space

{HHH, HHT, HTH, THH, HTT, THT, TTH, TTT},

$\therefore$ X can take the values 0, 1, 2, 3, with

$P[X = 0] = P[TTT] = \frac{1}{8}$

$P[X = 1] = P[HTT, THT, TTH] = \frac{3}{8}$

$P[X = 2] = P[HHT, HTH, THH] = \frac{3}{8}$

$P[X = 3] = P[HHH] = \frac{1}{8}.$

$\therefore$ Probability distribution of X, i.e. the number of heads when three coins are tossed simultaneously is

X	0	1	2	3
p(x)	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

which is the required probability distribution.

Example 4: A r.v. X assumes the values $-2, -1, 0, 1, 2$ such that

$P[X = -2] = P[X = -1] = P[X = 1] = P[X = 2]$,

$P[X < 0] = P[X = 0] = P[X > 0]$.

Obtain the probability mass function of X.

Solution: As $P[X < 0] = P[X = 0] = P[X > 0]$

$\therefore P[X = -1] + P[X = -2] = P[X = 0] = P[X = 1] + P[X = 2]$

$\Rightarrow p + p = P[X = 0] = p + p$

[Letting $P[X = 1] = P[X = 2] = P[X = -1] = P[X = -2] = p$]

$\Rightarrow P[X = 0] = 2p.$

Now, as $P[X < 0] + P[X = 0] + P[X > 0] = 1$,

$\therefore P[X = -1] + P[X = -2] + P[X = 0] + P[X = 1] + P[X = 2] = 1$

$\Rightarrow p + p + 2p + p + p = 1$

$\Rightarrow 6p = 1 \Rightarrow p = \frac{1}{6}$

$\therefore P[X = 0] = 2p = 2 \times \frac{1}{6} = \frac{2}{6}$,

$P[X = -1] = P[X = -2] = P[X = 1] = P[X = 2] = p = \frac{1}{6}$.

Hence, the probability distribution of X is given by

X	−2	−1	0	1	2
p(x)	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{2}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

Now, here are some exercises for you.

E1) 2 bad articles are mixed with 5 good ones. Find the probability distribution of the number of bad articles, if 2 articles are drawn at random.

E2) Given the probability distribution:

X	0	1	2	3
p(x)	$\frac{1}{10}$	$\frac{3}{10}$	$\frac{1}{2}$	$\frac{1}{10}$

Let $Y = X^2 + 2X$. Find the probability distribution of Y.

E3) An urn contains 3 white and 4 red balls. 3 balls are drawn one by one with replacement. Find the probability distribution of the number of red balls.

Let us define and explain a continuous random variable and its probability function in the next section.

5.4 CONTINUOUS RANDOM VARIABLE AND PROBABILITY DENSITY FUNCTION

In Sec. 5.3 of this unit, we have defined the discrete random variable as a random variable having countable number of values, i.e. whose values can be arranged in a sequence. But, if a random variable is such that its values cannot be arranged in a sequence, it is called continuous random variable. Temperature of a city at various points of time during a day is an example of continuous random variable as the temperature takes uncountable values, i.e. it can take fractional values also. So, a random variable is said to be continuous if it can take all possible real (i.e. integer as well as fractional) values between two certain limits. For example, let us denote the variable, "Difference between the rainfall (in cm) of a city and that of another city on every rainy day in a rainy reason", by X, then X here is a continuous random variable as it can take any real value between two certain limits. It can be noticed that for a continuous random variable, the chance of occurrence of a particular value of the variable is very small, so instead of specifying the probability of taking a particular value by the variable, we specify the probability of its lying within an interval. For example, chance that an athlete will finish a race in say exactly 10 seconds is very-very small, i.e. almost zero as it is very rare to finish the race in a fixed time. Here, the probability is specified for an interval, i.e. we may be interested in finding as to what is the probability of finishing the race by the athlete in an interval of say 10 to 12 seconds.

So, continuous random variable is represented by different representation known as **probability density function** unlike the discrete random variable which is represented by probability mass function.

Probability Density Function

Let $f(x)$ be a continuous function of x. Suppose the shaded region ABCD shown in the following figure represents the area bounded by $y = f(x)$, x-axis and the ordinates at the points x and $x + \delta x$, where δx is the length of the interval $(x, x + \delta x)$.

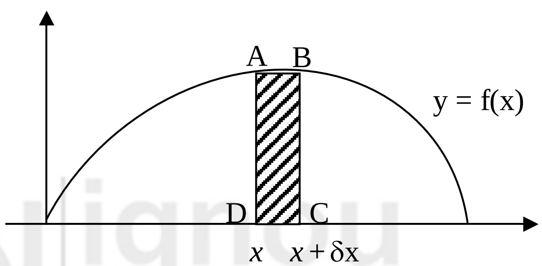


Fig. 5.1

Now, if δx is very-very small, then the curve AB will act as a line and hence the shaded region will be a rectangle whose area will be AD × DC i.e. $f(x)\,\delta x$ [$\because$ AD = the value of y at x i.e. $f(x)$, DC = length δx of the interval $(x, x+\delta x)$]

Also, this area = probability that X lies in the interval $(x, x+\delta x)$

$$= P[x \le X \le x + \delta x]$$

Hence,

$$P[x \le X \le x + \delta x] = f(x)\,\delta x$$

$$\Rightarrow \frac{P[x \le X \le x + \delta x]}{\delta x} = f(x), \text{ where } \delta x \text{ is very-very small}$$

$$\Rightarrow \lim_{\delta x \to 0} \frac{P[x \le X \le x + \delta x]}{\delta x} = f(x).$$

$f(x)$, so defined, is called probability density function.

Probability density function has the same properties as that of probability mass function. So, $f(x) \ge 0$ and sum of the probabilities of all possible values that the random variable can take, has to be 1. But, here, as X is a continuous random variable, the summation is made possible through 'integration' and hence

$$\int_R f(x)\,dx = 1,$$

where integral has been taken over the entire range R of values of X.

Remark 2

i) Summation and integration have the same meanings but in mathematics there is still difference between the two and that is that the former is used in case of discrete values, i.e. countable values and the latter is used in continuous case.

ii) An essential property of a continuous random variable is that there is zero probability that it takes any specified numerical value, but the probability that it takes a value in specified intervals is non-zero and is calculable as a definite integral of the probability density function of the random variable and hence the probability that a continuous r.v. X will lie between two values a and b is given by

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

Example 5: A continuous random variable X has the probability density function:

$f(x) = Ax^3,\ 0 \le x \le 1.$

Determine

i) A

ii) $P[0.2 < X < 0.5]$

iii) $P[X > \frac{3}{4} \text{ given } X > \frac{1}{2}]$

Solution:

(i) As $f(x)$ is probability density function,

$$\therefore \int_R f(x)\,dx = 1$$

$$\Rightarrow \int_0^1 f(x)\,dx = 1 \Rightarrow \int_0^1 Ax^3 dx = 1$$

$$\Rightarrow A\left[\frac{x^4}{4}\right]_0^1 = 1 \Rightarrow A\left(\frac{1}{4} - 0\right) = 1 \Rightarrow A = 4$$

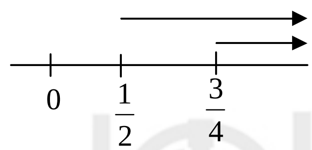
(ii) $P[0.2 < X < 0.5] = \int_{0.2}^{0.5} f(x)\,dx = \int_{0.2}^{0.5} Ax^3 dx = 4\left[\frac{x^4}{4}\right]_{0.2}^{0.5} = [(0.5)^4 - (0.2)^4]$

$$= 0.0625 - 0.0016 = 0.0609$$

(iii) $P[X > \frac{3}{4}$ given $X > \frac{1}{2}] = P[X > \frac{3}{4} \mid X > \frac{1}{2}]$

$$= \frac{P\left[X > \frac{3}{4} \cap X > \frac{1}{2}\right]}{P\left[X > \frac{1}{2}\right]} \qquad \left[\because P(A|B) = \frac{P(A \cap B)}{P(B)}\right]$$

$$= \frac{P\left[X > \frac{3}{4}\right]}{P\left[X > \frac{1}{2}\right]} \qquad \left[\begin{array}{l}\because \text{the common portion for} \\ X > \frac{3}{4} \text{ and } X > \frac{1}{2} \text{ is } X > \frac{3}{4}\end{array}\right]$$



Now, $P\left[X > \frac{3}{4}\right] = \int_{3/4}^{1} f(x)\,dx$

$$\left[\begin{array}{l}\because \text{ lower limit is } \frac{3}{4} \text{ and upper limit} \\ \text{is given in the problem which is } 1\end{array}\right]$$

$$= \int_{3/4}^{1} 4x^3 dx = 4\left[\frac{x^4}{4}\right]_{3/4}^{1} = (1)^4 - \left(\frac{3}{4}\right)^4 = 1 - \frac{81}{256} = \frac{175}{256}, \text{ and}$$

$$P\left[X > \frac{1}{2}\right] = \int_{1/2}^{1} f(x) = [x^4]^1_{1/2} = 1 - \frac{1}{16} = \frac{15}{16}.$$

$$\therefore \text{ the required probability } = \frac{P\left[X > \frac{3}{4}\right]}{P\left[X > \frac{1}{2}\right]} = \frac{175}{256} \times \frac{16}{15} = \frac{35}{16 \times 3} = \frac{35}{48}.$$

Example 6: The p.d.f. of the different weights of a "1 litre pure ghee pack" of a company is given by:

$$f(x) = \begin{cases} 200(x-1) & \text{for } 1 \le x \le 1.1 \\ 0, & \text{otherwise} \end{cases}$$

Examine whether the given p.d.f. is a valid one. If yes, find the probability that the weight of any pack will lie between 1.01 and 1.02.

Solution: For $1 \le x \le 1.1$, we have $f(x) \ge 0$, and

$$\int_1^{1.1} f(x)\,dx = \int_1^{1.1} 200(x-1)\,dx = 200\left[\frac{x^2}{2} - x\right]_1^{1.1} = 200\left[\left\{\frac{(1.1)^2}{2} - 1.1\right\} - \left\{\frac{1}{2} - 1\right\}\right]$$

$$= 200\left[\left(\frac{1.21-2.2}{2}\right) - \left(\frac{1-2}{2}\right)\right] = 200\left[-\frac{0.99}{2} + \frac{1}{2}\right] = 200\frac{(0.01)}{2} = 1.$$

$\therefore$ f(x) is p.d.f.

$$\text{Now, } P[1.01 < X < 1.02] = \int_{1.01}^{1.02} 200(x-1)\,dx = 200\left[\frac{x^2}{2} - x\right]_{1.01}^{1.02}$$

$$= 200\left[\left\{\frac{(1.02)^2}{2} - 1.02\right\} - \left\{\frac{(1.01)^2}{2} - 1.01\right\}\right]$$

$$= 200\left[\frac{1.0404}{2} - 1.02 - \frac{1.0201}{2} + 1.01\right]$$

$$= 200[0.5202 - 1.02 - 0.51005 + 1.01]$$

$$= 200[0.00015] = 0.03.$$

Now, you can try the following exercise.

E4) The life (in hours) X of a certain type of light bulb may be supposed to be a continuous random variable with p.d.f.:

$$f(x) = \begin{cases} \dfrac{A}{x^3}, & 1500 < x < 2500 \\ 0, & \text{elsewhere} \end{cases}$$

Determine the constant A and compute the probability that $1600 \le X \le 2000$.

5.5 DISTRIBUTION FUNCTION

A function F defined for all values of a random variable X by $F(x) = P[X \le x]$ is called the distribution function. It is also known as the cumulative distribution function (c.d.f.) of X since it is the cumulative probability of X up to and including the value x. As X can take any real value, therefore the domain of the distribution function is set of real numbers and as F(x) is a probability value, therefore the range of the distribution function is [0, 1].

Remark 3: Here, X denotes the random variable and x represents a particular value of random variable. $F(x)$ may also be written as $F_X(x)$, which means that it is a distribution function of random variable X.

Discrete Distribution Function

Distribution function of a discrete random variable is said to be discrete distribution function or cumulative distribution function (c.d.f.). Let X be a discrete random variable taking the values $x_1, x_2, x_3, \ldots$ with respective probabilities $p_1, p_2, p_3, \ldots$

Then $F(x_i) = P[X \le x_i] = P[X = x_1] + P[X = x_2] + \ldots + P[X = x_i]$

$$= p_1 + p_2 + \ldots + p_i.$$

The distribution function of X, in this case, is given as in the following table:

X	F(x)
x_1	p_1
x_2	$p_1 + p_2$
x_3	$p_1 + p_2 + p_3$
x_4	$p_1 + p_2 + p_3 + p_4$
.	.
.	.
.	.

The value of F(x) corresponding to the last value of the random variable X is always 1, as it is the sum of all the probabilities. F(x) remains 1 beyond this last value of X also, as it being a probability can never exceed one.

For example, Let X be a random variable having the following probability distribution:

X	0	1	2
p(x)	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

Notice that p(x) will be zero for other values of X. Then, Distribution function of X is given by

X	$F(x) = P[X \le x]$
0	$\frac{1}{4}$
1	$\frac{1}{4} + \frac{1}{2} = \frac{3}{4}$
2	$\frac{1}{4} + \frac{1}{2} + \frac{1}{4} = 1$

Here, for the last value, i.e. for X = 2, we have $F(x) = 1$.

Also, if we take a value beyond 2 say 4, then we get

$F(4) = P[X \le 4]$

$= P[X = 4] + P[X = 3] + P[X \le 2]$

$= 0 + 0 + 1 = 1.$

Example 7: A random variable X has the following probability function:

X	0	1	2	3	4	5	6	7
$p(x)$	0	$\frac{1}{10}$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{3}{10}$	$\frac{1}{100}$	$\frac{1}{50}$	$\frac{17}{100}$

Determine the distribution function of X.

Solution: Here,

$F(0) = P[X \le 0] = P[X = 0] = 0,$

$F(1) = P[X \le 1] = P[X = 0] + P\,[X = 1] = 0 + \frac{1}{10} = \frac{1}{10},$

$F(2) = P[X \le 2] = P[X = 0] + P\,[X = 1] + [X = 2] = 0 + \frac{1}{10} + \frac{1}{5} = \frac{3}{10},$

and so on. Thus, the distribution function $F(x)$ of X is given in the following table:

X	$F(x) = P[X \le x]$
0	0
1	$\frac{1}{10}$
2	$\frac{3}{10}$
3	$\frac{3}{10} + \frac{1}{5} = \frac{1}{2}$
4	$\frac{1}{2} + \frac{3}{10} = \frac{4}{5}$
5	$\frac{4}{5} + \frac{1}{100} = \frac{81}{100}$
6	$\frac{81}{100} + \frac{1}{50} = \frac{83}{100}$
7	$\frac{83}{100} + \frac{17}{100} = 1$

Continuous Distribution Function

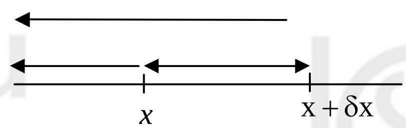
Distribution function of a continuous random variable is called the continuous distribution function or cumulative distribution function (c.d.f.).

Let X be a continuous random variable having the probability density function f(x), as defined in the last section of this unit, then the distribution function F(x) is given by

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

Also, in the last section, we have defined the p.d.f. f(x) as

$$f(x) = \lim_{\delta x \to 0} \frac{P[x \le X \le x + \delta x]}{\delta x},$$



$$\therefore f(x) = \lim_{\delta x \to 0} \frac{P[X \le x + \delta x] - P[X \le x]}{\delta x}$$

$$\Rightarrow \quad f(x) = \lim_{\delta x \to 0} \frac{F(x + \delta x) - F(x)}{\delta x},$$

$\Rightarrow f(x) =$ Derivative of F(x) with respect to x $\begin{bmatrix} \text{By definition of} \\ \text{the derivative} \end{bmatrix}$

$$\Rightarrow f(x) = F'(x)$$

$$\Rightarrow f(x) = \frac{d}{dx}\left(F(x)\right)$$

$$\Rightarrow dF(x) = f(x)\,dx$$

Here, dF(x) is known as the probability differential.

So, $F(x) = \int_{-\infty}^{x} f_X(x)\,dx$ and $F'(x) = f(x)$.

Example 8: The diameter 'X' of a cable is assumed to be a continuous random variable with p.d.f. $f(x) = \begin{cases} 6x(1-x), & 0 \le x \le 1 \\ 0, & \text{elsewhere} \end{cases}$.

Obtain the c.d.f. of X.

Solution: For $0 \le x \le 1$, the c.d.f. of X is given by

$$F(x) = P[X \le x] = \int_0^x f(x)\,dx = \int_0^x 6x(1-x)\,dx$$

$$= 6\int_0^x (x - x^2)\,dx = 6\left[\frac{x^2}{2} - \frac{x^3}{3}\right]_0^x = 3x^2 - 2x^3$$

$\therefore$ The c.d.f. of X is given by

$$F(x) = \begin{cases} 0, & x < 0 \\ 3x^2 - 2x^3, & 0 \le x \le 1. \\ 1, & x > 1 \end{cases}$$

Remark 4: In the above example, $F(x)$ is taken as 0 for $x < 0$ since $p(x) = 0$ for $x < 0$; and $F(x)$ is taken as 1 for $x > 1$ since $F(1) = 1$ and therefore,

for $x > 1$ also $F(x)$ will remain 1.

Now, you can try the following exercises.

E 5) A random variable X has the following probability distribution:

X	0	1	2	3	4	5	6	7	8
$p(x)$	k	3k	5k	7k	9k	11k	13k	15k	17k

i) Determine the value of k.

ii) Find the distribution function of X.

E 6) Let X be continuous random variable with p.d.f. given by.

$$f(x) = \begin{cases} \frac{x}{2}, & 0 \le x < 1 \\ \frac{1}{2}, & 1 \le x < 2 \\ \frac{1}{2}(3-x), & 2 \le x < 3 \\ 0, & \text{elsewhere.} \end{cases}.$$

Determine $F(x)$, the c.d.f. of X.

5.6 SUMMARY

Following main points have been covered in this unit of the course:

1) A **random variable** is a function whose domain is a set of possible outcomes and range is a sub-set of the set of reals and has the following properties:

   i) Each particular value of the random variable can be assigned some probability.

   ii) Sum of all the probabilities associated with all the different values of the random variable is unity.

2) A random variable is said to be **discrete random variable** if it has either a finite number of values or a countable number of values, i.e. the values can be arranged in a sequence.

3) If a random variable is such that its values cannot be arranged in a sequence, it is called **continuous random variable**. So, a random variable is said to be continuous if it can take all the possible real (i.e. integer as well as fractional) values between two certain limits.

4) Let X be a discrete r.v. which take on the values x_1, x_2, ... and let $P\left[X = x_i\right] = p(x_i)$. The function $p(x_i)$ is called **probability mass function** of X satisfying $p(x_i) \geq 0$ and $\sum_i p\left(x_i\right) = 1$. The set $\left\{\left(x_1, p\left(x_1\right)\right), \left(x_2, p\left(x_2\right)\right), ...\right\}$ specifies the **probability distribution** of discrete r.v. X.

5) Let X be a continuous random variable and $f(x)$ be a continuous function of x. Suppose $(x, x + \delta x)$ be an interval of length δx. Then $f(x)$ defined by $\lim_{\delta x \to 0} \frac{P[x \leq X \leq x + \delta x]}{\delta x} = f\left(x\right)$ is called the **probability density function** of X.

   Probability density function has the same properties as that of probability mass function i.e. $f(x) \geq 0$ and $\int_R f\left(x\right) dx = 1$, where integral has been taken over the entire range R of values of X.

6) A function F defined for all values of a random variable X by $F(x) = P[X \leq x]$ is called the **distribution function**. It is also known as the **cumulative distribution function (c.d.f.)** of X. The domain of the distribution function is a set of real numbers and its range is [0, 1]. Distribution function of a discrete random variable X is said to be **discrete distribution function** and is given by $\left\{\left(x_1, F\left(x_1\right)\right), \left(x_2, F\left(x_2\right)\right), ...\right\}$. Distribution function of a continuous random variable X having the probability density function $f(x)$ is said to be **continuous distribution function** and is given by

   $F(x) = P[X \leq x] = \int_{-\infty}^{x} f\left(x\right) dx$.

   Derivative of $F(x)$ with respect to x is $f\left(x\right)$, i.e. $F'\left(x\right) = f\left(x\right)$.

5.7 SOLUTIONS/ANSWERS

E 1) Let X be the number of bad articles drawn.

$\therefore$ X can take the values 0, 1, 2 with

$P[X = 0] = P[\text{No bad article}]$

$= P[\text{Drawing 2 articles from 5 good articles and zero article from 2 bad articles}]$

$$= \frac{{}^5C_2 \times {}^2C_0}{{}^7C_2} = \frac{5\times4\times1}{7\times6} = \frac{10}{21},$$

P[X = 1] = P[One bad article and 1 good article]

$$= \frac{{}^2C_1 \times {}^5C_1}{{}^7C_2} = \frac{2\times5\times2}{7\times6} = \frac{10}{21}, \text{ and}$$

P[X = 2] = P[Two bad articles and no good article]

$$= \frac{{}^2C_2 \times {}^5C_0}{{}^7C_2} = \frac{1\times1\times2}{7\times6} = \frac{1}{21}$$

$\therefore$ Probability distribution of number of bad articles is:

X	0	1	2
p(x)	$\frac{10}{21}$	$\frac{10}{21}$	$\frac{1}{21}$

E2) As $Y = X^2 + 2X$,

$\therefore$ For X = 0, Y = 0 + 0 = 0;

For X = 1, $Y = 1^2 + 2(1) = 3$;

For X = 2, $Y = 2^2 + 2(2) = 8$; and

For X = 3, $Y = 3^2 + 2(3) = 15$.

Thus, the values of Y are 0, 3, 8, 15 corresponding to the values 0, 1, 2, 3 of X and hence

$P[Y = 0] = P[X = 0] = \frac{1}{10}$, $P[Y = 3] = P[X = 1] = \frac{3}{10}$,

$P[Y = 8] = P[X = 2] = \frac{1}{2}$ and $P[Y = 15] = P[X = 3] = \frac{1}{10}$.

$\therefore$ The probability distribution of Y is

Y	0	3	8	15
p(y)	$\frac{1}{10}$	$\frac{3}{10}$	$\frac{1}{2}$	$\frac{1}{10}$

E3) Let X be the number of red balls drawn.

$\therefore$ X can take the values 0, 1, 2, 3.

Let W_i be the event that i^{th} draw gives white ball and R_i be the event that i^{th} draw gives red ball.

$\therefore$ P[X = 0] = P[No Red ball] = $P[W_1 \cap W_2 \cap W_3]$

$= P(W_1).\ P(W_2).\ P(W_3)$

[$\because$ balls are drawn with replacement and hence the draws are independent]

$$=\frac{3}{7}\times\frac{3}{7}\times\frac{3}{7}=\frac{27}{343}$$

P[X = 1] = P[One red and two white]

$$= P\left[\left(R_1 \cap W_2 \cap W_3\right) \text{or} \left(W_1 \cap R_2 \cap W_3\right) \text{or} \left(W_1 \cap W_2 \cap R_3\right)\right]$$

$$= P\left[R_1 \cap W_2 \cap W_3\right] + P\left[W_1 \cap R_2 \cap W_3\right] + P\left[W_1 \cap W_2 \cap R_3\right]$$

$$= P[R_1]P[W_2]P[W_3] + P[W_1]P[R_2]P[W_3] + P[W_1]P[W_2]P[R_3]$$

$$=\frac{4}{7}\times\frac{3}{7}\times\frac{3}{7}+\frac{3}{7}\times\frac{4}{7}\times\frac{3}{7}+\frac{3}{7}\times\frac{3}{7}\times\frac{4}{7}= 3\times\frac{4}{7}\times\frac{3}{7}\times\frac{3}{7}= \frac{108}{343},$$

P[X = 2] = P[Two red and one white]

$$= P\left[\left(R_1 \cap R_2 \cap W_3\right) \text{or } \left(R_1 \cap W_2 \cap R_3\right) \text{or} \left(W_1 \cap R_2 \cap R_3\right)\right]$$

$$= P[R_1]P[R_2]P[W_3] + P[R_1]P[W_2]P[R_3] + P[W_1]P[R_2]P[R_3]$$

$$=\frac{4}{7}\times\frac{4}{7}\times\frac{3}{7}+\frac{4}{7}\times\frac{3}{7}\times\frac{4}{7}+\frac{3}{7}\times\frac{4}{7}\times\frac{4}{7}= 3\times\frac{4}{7}\times\frac{4}{7}\times\frac{3}{7}= \frac{144}{343}\ .$$

P[X = 3] = P [Three red balls]

$$= P[R_1 \cap R_2 \cap R_3] = P(R_1)\ P(R_2)\ P(R_3) = \frac{4}{7}\times\frac{4}{7}\times\frac{4}{7}=\frac{64}{343}.$$

$\therefore$ Probability distribution of the number of red balls is

X	0	1	2	3
$p(x)$	$\frac{27}{343}$	$\frac{108}{343}$	$\frac{144}{343}$	$\frac{64}{343}$

E4) As $f(x)$ is p.d.f.,

$$\therefore \int_{1500}^{2500}\frac{A}{x^3}dx = 1 \Rightarrow A\int_{1500}^{2500}x^{-3}dx = 1 \Rightarrow A\left[\frac{x^{-2}}{-2}\right]_{1500}^{2500} = 1$$

$$\Rightarrow -\frac{A}{2}\left[\frac{1}{(2500)^2}-\frac{1}{(1500)^2}\right] = 1 \Rightarrow -\frac{A}{20000}\left[\frac{1}{625}-\frac{1}{225}\right] = 1$$

$$\Rightarrow -\frac{A}{20000}\left[\frac{9-25}{5625}\right] = 1 \Rightarrow 16A = 5625 \times 20000$$

$$\Rightarrow \quad A = \frac{5625\times 20000}{16} = 5625 \times 1250 = 7031250.$$

Now, $P[1600 \le X \le 2000] = \int_{1600}^{2000} f(x)dx = A\int_{1600}^{2000}\frac{1}{x^3}dx$

$$= -\frac{A}{2}\left[\frac{1}{x^2}\right]_{1600}^{2000} = -\frac{A}{2}\left[\frac{1}{(2000)^2} - \frac{1}{(1600)^2}\right]$$

$$= -\frac{A}{20000}\left[\frac{1}{400} - \frac{1}{256}\right] = -\frac{A}{20000}\left[\frac{16-25}{6400}\right]$$

$$= \frac{9 \times 7031250}{20000 \times 6400} = \frac{2025}{4096}$$

E5) i) As the given distribution is probability distribution,

$\therefore$ sum of all the probabilities = 1

$\Rightarrow$ k + 3k + 5k + 7k + 9k + 11k + 13k + 15k + 17k = 1

$\Rightarrow 81\text{ k} = 1 \Rightarrow \text{k} = \frac{1}{81}$

ii) The distribution function of X is given in the following table:

X	$F(x) = P[X \le x]$
0	$k = \frac{1}{81}$
1	$k + 3k = 4k = \frac{4}{81}$
2	$4k + 5k = 9k = \frac{9}{81}$
3	$9k + 7k = 16k = \frac{16}{81}$
4	$16k + 9k = 25k = \frac{25}{81}$
5	$25k + 11k = 36k = \frac{36}{81}$
6	$36k + 13k = 49k = \frac{49}{81}$
7	$49k + 15k = 64k = \frac{64}{81}$
8	$64k + 17k = 81k = 1$

E6) For $x < 0$,

$$F(x) = P[X \le x] = \int_{-\infty}^{x} f(x)\,dx = \int_{-\infty}^{x} 0\,dx = 0 \quad [\because f(x) = 0 \text{ for } x < 0].$$

For $0 \le x < 1$,

$$F(x) = P[X \le x] = \int_{-\infty}^{x} f(x)\,dx = \int_{-\infty}^{0} f(x)\,dx + \int_{0}^{x} f(x)\,dx \quad [\because 0 \le x < 1]$$

$$= 0 + \int_{0}^{x} \frac{x}{2}\,dx \qquad \left[\because f(x) = \frac{x}{2} \text{ for } 0 \le x < 1\right]$$

$$= \frac{1}{2}\left[\frac{x^2}{2}\right]_0^x = \frac{x^2}{4}.$$

For $1 \le x < 2$,

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

$$= \int_{-\infty}^{0} f(x)\,dx + \int_{0}^{1} f(x)\,dx + \int_{1}^{x} f(x)\,dx$$

$$= 0 + \int_{0}^{1} \frac{x}{2}\,dx + \int_{1}^{x} \frac{1}{2}\,dx = \frac{1}{4}\left[x^2\right]_0^1 + \frac{1}{2}[x]_1^x$$

$$= \frac{1}{4} + \frac{x}{2} - \frac{1}{2} = \frac{1}{4}(2x - 1)$$

For $2 \le x < 3$,

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

$$= \int_{-\infty}^{0} f(x)\,dx + \int_{0}^{1} f(x)\,dx + \int_{1}^{2} f(x)\,dx + \int_{2}^{x} f(x)\,dx$$

$$= 0 + \int_{0}^{1} \frac{x}{2}\,dx + \int_{1}^{2} \frac{1}{2}\,dx + \int_{2}^{x} \frac{1}{2}(3 - x)\,dx$$

$$= \left[\frac{x^2}{4}\right]_0^1 + \frac{1}{2}[x]_1^2 + \frac{1}{2}\left[3x - \frac{x^2}{2}\right]_2^x$$

$$= \frac{1}{4} + \frac{1}{2} + \frac{1}{2}\left[\left(3x - \frac{x^2}{2}\right) - (6 - 2)\right]$$

$$= \frac{1}{2}\left(3x - \frac{x^2}{2}\right) - \frac{5}{4} = -\frac{x^2}{4} + \frac{3x}{2} - \frac{5}{4}$$

For $3 \le x < \infty$,

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

$$= \int_{-\infty}^{0} f(x)\,dx + \int_{0}^{1} f(x)\,dx + \int_{1}^{2} f(x)\,dx + \int_{2}^{3} f(x)\,dx \int_{3}^{x} f(x)\,dx$$

$$= 0 + \int_{0}^{1} \frac{x}{2}\,dx + \int_{1}^{2} \frac{1}{2}\,dx + \int_{2}^{3} \frac{1}{2}(3-x)\,dx + \int_{3}^{x} 0\,dx$$

$$= \left[\frac{x^2}{4}\right]_0^1 + \left[\frac{x}{2}\right]_1^2 + \frac{1}{2}\left[3x - \frac{x^2}{2}\right]_2^3 + 0$$

$$= \frac{1}{4} + \left(1 - \frac{1}{2}\right) + \frac{1}{2}\left[\left(9 - \frac{9}{2}\right) - (6-2)\right]$$

$$= \frac{1}{4} + \frac{1}{2} + \frac{1}{2}\left(\frac{9}{2} - 4\right)$$

$$= \frac{1}{4} + \frac{1}{2} + \frac{1}{4} = 1$$

Hence, the distribution function is given by:

$$F(x) = \begin{cases} 0, & -\infty < x < 0 \\ \dfrac{x^2}{4}, & 0 \le x < 1 \\ \dfrac{2x-1}{4}, & 1 \le x < 2 \\ -\dfrac{x^2}{4} + \dfrac{3x}{2} - \dfrac{5}{4}, & 2 \le x < 3 \\ 1, & 3 \le x < \infty \end{cases}$$