CHAPTER 1

Correlation

1. Introduction

In Statistics so far we have studied problems involving a single variable. Many a time, we come across problems which involve two or more than two variables. Data relating to two variates are called **Bivariate Data**. If we study bivariate data carefully, we may find some relation between the two variables. For example, if a car-owner maintains the record of petrol consumption and mileage, he will find that there is some relation between the two variables. On the other hand, if we compare the figures of rainfall with the production of cars, we may find that there is no relation between the two variables. If there is any relation between two variables *i.e.* if as one variable changes the other also changes in the same or opposite direction, we say that they are correlated. Thus, correlation means **"the study of existence and the magnitude and direction of variation between two or more variables"**.

2. Types of Correlation

Correlation may be classified as :

(1) Positive and negative,

(2) Linear and non-linear.

(1) Positive and Negative Correlations

The distinction between the positive and negative correlation depends upon the direction of change of two variables. If both the variables change in the same direction *i.e.* if, as one variable increases, the other also increases and as one variable decreases, the other also decreases, the correlation is called positive (*e.g.* advertising and sales). If, on the other hand, the variables change in opposite direction *i.e.* if, as one variable increases, the other decreases and vice-versa, then the correlation is called negative (*e.g.* T.V. registrations and cinema attendance).

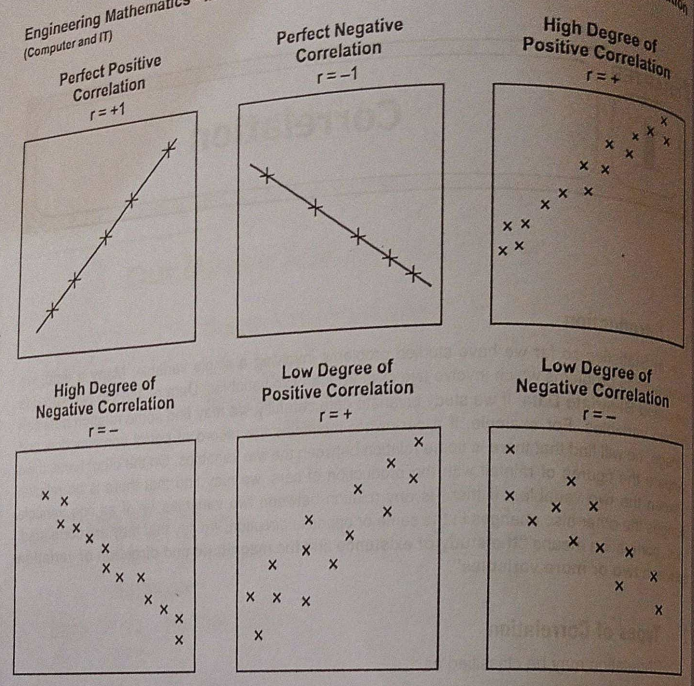
(2) Linear or Non-linear Correlation

This distinction is based upon the nature of the graph of the relation between the variables. If the graph is a straight line the correlation is called linear and if the graph is not a straight line but a curve it is called non-linear or curvi-linear correlation.

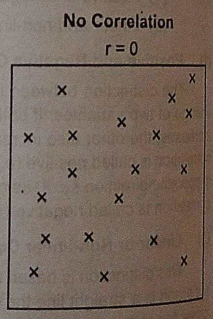
We shall consider the following commonly used methods of studying correlation. (1) Scatter Diagram, (2) Karl Pearson's Coefficient of Correlation, (3) Spearman's Rank-Correlation Coefficient.

3. Scatter Diagram

One of the most simple methods of studying correlation between two variables is to construct a scatter diagram.



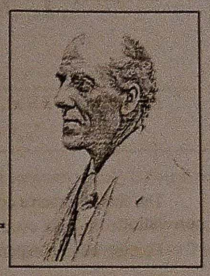
To obtain a scatter diagram, one variable is plotted along the x-axis and the other along the y-axis, on a graph paper. By plotting data in this way, we get points which are generally scattered but which show a pattern. The way in which the points are scattered indicates the degree and direction of correlation. If the points are close to each other we infer that the variables are correlated. If they are spread away from each other, we infer that the variables are not correlated. Moreover, if the points lie in a narrow strip rising from left-hand bottom to the right-hand top, we say that there is positive correlation of high order. If the points lie in a narrow strip, falling from the left-hand top to the right-hand bottom, we say that there is negative correlation of high order. If the points are all spread over, we say that there is zero correlation.



4. Karl Pearson's Coefficient of Correlation

The method of scatter diagram is descriptive in nature and gives only a general idea of correlation. The most commonly used method which gives a mathematical expression for correlation is the one suggested by Karl Pearson (1867-1936) a British Biometrician.

Karl Pearson (1857 - 1936)



Born in London, he went to King's College, Cambridge in 1876 to study mathematics graduating in 1879 as Third Wrangler in the Mathematical Tripos. He then went to Germany to study Physics at the University of Heidelberg. Other subjects he studied in Germany include metaphysics, physiology, Roman Law, German Literature and Socialism. He studied so many subjects because he believed that there was no subject in the universe unworthy of study. Then he returned to London to study Law, although he never practised. In 1881 he returned to mathematics and was appointed as professor of mathematics at University College, London. In 1891 he met Walter Weldon, a zoologist and worked with him in biometry and evolutionary theory. He was introduced to Galton, Darwins cousin and became Galtons statistical heir. He was the first holder of the Galton's Chair of Eugenics. In 1911 he founded the world's first university statistics department at University College, London. He remained with the department until his retirement in 1933 and continued to work until his death in 1936. He thus established the new discipline of mathematical statistics.

His famous book **"The Grammar of Science"** covers several themes that were later to become part of the theories of Einstein and other scientists. He speculated that an observer who travelled at the speed of light would see an eternal now and an observer who travelled faster than light would see time reversal. He also discussed antimatter, fourth dimension and wrinkles in time.

Karl Pearson was awarded many medals including The Darwin Medal, a DSc from university of London. His commitment to socialism and his ideals led him to refuse the honours of being an OBE (Officer of the Order of the British Empire) and knighthood in 1935.

Karl Pearson is known for Karl Pearson's coefficient of correlation, methods of moments, Pearson's system of continuous curves, Chi-distance, Statistical hypothesis testing theory, Statistical decision theory, Pearson's chi-square test, etc.

Just as $\sigma_x^2 = \frac{1}{N}\Sigma(x-\bar{x})^2$ gives us a measure of variation in x and $\sigma_y^2 = \frac{1}{N}\Sigma(y-\bar{y})^2$ gives a measure of variation in y we may expect $\frac{1}{N}\Sigma(x-\bar{x})(y-\bar{y})$ to give the measure of simultaneous variation in x and y. But this will depend upon the units of x and y. To find a ratio which is independent of these units, we divide it by the quantities of the same order that is by $\sigma_x \cdot \sigma_y$. With this view in mind Karl Pearson suggested in 1890 the following coefficient of correlation to measure correlation between x and y. It is denoted by r.

Thus,
$$\boxed{r = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{N\sigma_x\sigma_y}} \quad \text{..... (1)}$$

But $\frac{1}{N}\Sigma(x-\bar{x})(y-\bar{y})$ is called the **covariance** between x and y. Hence, from (1), we have

$$\boxed{r = \frac{\text{cov.}(x,y)}{\sigma_x \cdot \sigma_y}} \quad \text{..... (2)}$$

If we put $\sigma_x = \sqrt{\frac{\Sigma(x-\bar{x})^2}{N}}$, $\sigma_y = \sqrt{\frac{\Sigma(y-\bar{y})^2}{N}}$, then

$$\boxed{r = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{\sqrt{\Sigma(x-\bar{x})^2\,\Sigma(y-\bar{y})^2}}} \quad \text{.....(3)}$$

If we write $x-\bar{x} = x'$, $y-\bar{y} = y'$, then

$$\boxed{r = \frac{\Sigma x'y'}{\sqrt{\Sigma x'^2 \cdot \Sigma y'^2}}} \quad \text{.....(4)}$$

The Karl Pearson's coefficient of correlation is also called the **product moment coefficient correlation.**

Further, we can expand (3) and write

$$r = \frac{\Sigma(xy - x\bar{y} - \bar{x}y + \bar{x}\bar{y})}{\sqrt{\Sigma(x^2 - 2x\bar{x} + \bar{x}^2)\cdot\Sigma(y^2 - 2y\bar{y} + \bar{y}^2)}}$$

$$= \frac{\Sigma xy - \bar{y}\Sigma x - \bar{x}\Sigma y + \bar{x}\bar{y}\Sigma 1}{\sqrt{(\Sigma x^2 - 2\bar{x}\Sigma x + \bar{x}^2\Sigma 1)\cdot(\Sigma y^2 - 2\bar{y}\Sigma y + \bar{y}^2\Sigma 1)}}$$

But $\Sigma x = N\bar{x}$, $\Sigma y = N\bar{y}$ and $\Sigma 1 = N$

$$\therefore \quad \boxed{r = \frac{\Sigma xy - N\bar{x}\bar{y}}{\sqrt{(\Sigma x^2 - N\bar{x}^2)\cdot(\Sigma y^2 - N\bar{y}^2)}}} \quad \text{.....(5)}$$

If $\bar{x}$, $\bar{y}$ are integers we take deviations of x and y from them and use the formula (3). If we have to find r from direct values we use the formula (5). This is the most *commonly used formula.*

(i) Limits for r: $-1 \le r \le 1$

Proof : If we write $E(X) = \mu_X$, $E(Y) = \mu_Y$, then

$$E\left[\left(\frac{X-\mu_X}{\sigma_X}\right) \pm \left(\frac{Y-\mu_Y}{\sigma_Y}\right)\right]^2 \ge 0$$

$$\therefore \quad E\left(\frac{X-\mu_X}{\sigma_X}\right)^2 + E\left(\frac{Y-\mu_Y}{\sigma_Y}\right)^2 \pm 2\frac{E[(X-\mu_X)^2(Y-\mu_Y)]}{\sigma_X\,\Sigma_Y} \ge 0$$

$\therefore \quad 1 + 1 \pm 2r \ge 0 \qquad \therefore \quad 1 \pm r \ge 0$

$\therefore \quad 1 + r \ge 0 \qquad$ or $\qquad 1 - r \ge 0$

$\therefore \quad 1 \ge -r \qquad$ or $\qquad 1 \ge r$

$\therefore \quad -1 \le r,\ r \le 1 \qquad \therefore \quad -1 \le r \le 1 \qquad [\because\ E(X-\mu_X)^2 = \sigma_X^2]$

(ii) Theorems on correlation

Theorem 1 : If x, y are independent variables they are not correlated.

We accept this theorem without proof.

Theorem 2 : Correlation coefficient is independent of change of origin and change of scale.

This means if we write $u_i = \frac{x-a}{h}$, $v_i = \frac{y-b}{k}$, then

$$\boxed{r_{xy} = r_{uv}} \quad \text{.....(6)}$$

i.e., the correlation between x and y is equal to the correlation between u and v.

We accept this theorem without proof.

Remark ....

The above theorem can also be stated as :-

"If $x = au + b$, $y = cv + d$ where a, b, c, d are constants then $r_{xy} = r_{uv}$."

Example : Discuss the statement : "If the coefficient of correlation between x and y is negative then the coefficient of correlation between $(-x)$ and $(-y)$ is positive.

Sol. : If we write $a = 0$ and $b = 0$, $h = -1$ and $k = -1$ in (A), then by the above theorem since $r_{xy} = r_{uv}$, the coefficient of correlation between $-x$ and $-y$ will be also the same in magnitude and sign as the coefficient of correlation between x and y.

Or since the coefficient of correlation does not change under change of scale and since $-x$ and $-y$ mean the change of scale, the coefficient of correlation between $-x$ and $-y$ will be also negative.

Theorem 3 : *If d_x and d_y denote the deviations of x and y from the assumed means A and B then*

$$\boxed{r = \frac{\Sigma d_x d_y - \frac{(\Sigma d_x)(\Sigma d_y)}{N}}{\sqrt{\Sigma d_x^2 - \frac{(\Sigma d_x)^2}{N}}\sqrt{\Sigma d_y^2 - \frac{(\Sigma d_y)^2}{N}}}} \quad \text{.....(7)}$$

We accept this result without proof.

5. Interpretation of the Coefficient of Correlation

1. $r > 0{\cdot}95$: If r is greater than 0·95, it indicates high degree of correlation and the value of one variable can be estimated from a known value of the other fairly accurately.

2. $r > 0{\cdot}75$ but $< 0{\cdot}85$: If r is greater than 0·75 but less than 0·85, there is probably a definite relationship between the variables and the value of one variable can be roughly estimated from a known value of the other.

3. $r > 0{\cdot}40$ but $< 0{\cdot}60$: If r is greater than 0·40 but less than 0·60 there may be some relationship between the two variables. But the value of one variable calculated from a known value of the other cannot be reliable.

4. $r < 0{\cdot}35$: If r is less than 0·35 the correlation is poor and one variable cannot be estimated from the other.

5. r nearly zero : If r is nearly equal to zero, it indicates that there is probably no relation between the two variables *i.e.* they are independent of each other.

6. Computation of Coefficient of Correlation : (Ungrouped Data)

There are three methods of calculating r.

(1) Actual mean method, (2) Step deviation method,

(3) Assumed mean method.

(1) Actual Mean Method

The formula to be used is, $r = \dfrac{\Sigma xy}{\sqrt{\Sigma x^2 \cdot \Sigma y^2}}$

Steps :

(i) Calculate mean $\overline{X}$ and then take deviation x of X from $\overline{X}$ *i.e.*, calculate $x = X - \overline{X}$.

(ii) Calculate mean $\overline{Y}$ and then take deviation y of Y from $\overline{Y}$ *i.e.*, calculate $y = Y - \overline{Y}$.

(iii) Multiply x by y and prepare the column of xy.

(iv) Take the squares of x and prepare the column of x^2.

(v) Take the squares of y and prepare the column of y^2.

(vi) Apply the above formula.

Example 1 : Find from the following values of the demand and the corresponding price of a commodity, the degree of correlation between the demand and price by computing Karl Pearson's coefficient of correlation.

Demand in quintals : 65, 66, 67, 67, 68, 69, 70, 72.

Price in Paise per k.g. : 67, 68, 65, 68, 72, 72, 69, 71. **(S.U. 1997)**

Sol. : Let X denote the demand in Quintals and Y denote the price in paise per kg.

Calculation of r between demand and price

Sr. No.	Demand in Qnt. $X - \overline{X}$, $\overline{X} = 68$			Price per Kg. $Y - \overline{Y}$, $\overline{Y} = 69$			Product
	X	x	x^2	Y	y	y^2	xy
1	65	−3	9	67	−2	4	6
2	66	−2	4	68	−1	1	2
3	67	−1	1	65	−4	16	4
4	67	−1	1	68	−1	1	1
5	68	0	0	72	+3	9	0
6	69	+1	1	72	+3	9	3
7	70	+2	4	69	0	0	0
8	72	+4	16	71	+2	4	8
$N = 8$	$\Sigma X = 544$		$\Sigma x^2 = 36$	$\Sigma Y = 552$		$\Sigma y^2 = 44$	$\Sigma xy = +24$

Now, $\overline{X} = \dfrac{544}{8} = 68$ and $\overline{Y} = \dfrac{552}{8} = 69$.

$\therefore \quad r = \dfrac{\Sigma xy}{\sqrt{\Sigma x^2 \cdot \Sigma y^2}}$

But $\Sigma xy = 24$, $\Sigma x^2 = 36$, $\Sigma y^2 = 44$.

$\therefore \quad r = \dfrac{24}{\sqrt{36 \times 44}} = \dfrac{24}{39 \cdot 80} = 0 \cdot 6030.$

Example 2 : Calculate Karl Pearson's coefficient of correlation for the following bivariate series.

X : 28, 45, 40, 38, 35, 33, 40, 32, 36, 33

Y : 23, 34, 33, 34, 30, 26, 28, 31, 36, 35 **(S.U. 1998)**

Sol. :

Calculationof r between X and Y

Sr. No.	$X - \overline{X}$, $\overline{X} = 36$			$Y - \overline{Y}$, $\overline{Y} = 31$			Product
	X	x	x^2	Y	y	y^2	xy
1	28	−8	64	23	−8	64	+64
2	45	+9	81	34	+3	9	+27
3	40	+4	16	33	+2	4	+8
4	38	+2	4	34	+3	9	+6
5	35	−1	1	30	−1	1	+1
6	33	−3	9	26	−5	25	−15
7	40	+4	16	28	−3	9	−12
8	32	−4	16	31	0	0	0
9	36	0	0	36	+5	25	0
10	33	−3	9	35	+4	16	−12
$N = 10$	$\Sigma X = 360$		$\Sigma x^2 = 216$	$\Sigma Y = 310$		$\Sigma y^2 = 162$	$\Sigma xy = +97$

Now, $\overline{X} = \dfrac{360}{10} = 36$ and $\overline{Y} = \dfrac{310}{10} = 31$.

$\therefore \quad r = \dfrac{\Sigma xy}{\sqrt{\Sigma x^2 \cdot \Sigma y^2}}$

But $\Sigma xy = 97$, $\Sigma x^2 = 216$, $\Sigma y^2 = 162$.

$\therefore \quad r = \dfrac{97}{\sqrt{216 \times 162}} = \dfrac{97}{187 \cdot 1} = 0 \cdot 5186.$

(2) Step-deviation Method

As in the case of mean and standard deviation, to simplify calculations we can use step-deviation method whenever possible for calculating r. But it should be noted that *the result is not to be multiplied by the constant in the final stage*. The reason is the coefficient of correlation is independent of change of origin and change of scale. (See Theorem 2 of r, page 1-4)

Example 1 : Calculate the co-efficient of correlation from the following data.

X : 100, 200, 300, 400, 500.

Y : 30, 40, 50, 60, 70.

Sol. :

Calculations of r between X and Y

Sr. No.	$\overline{X} = 300$, $X - \overline{X}$		$\dfrac{X - \overline{X}}{100}$		$\overline{Y} = 50$, $Y - \overline{Y}$		$\dfrac{Y - \overline{Y}}{10}$		Product
	X	$X - 300$	x	x^2	Y	$Y - 50$	y	y^2	xy
1	100	−200	−2	4	30	−20	−2	4	4
2	200	−100	−1	1	40	−10	−1	1	1
3	300	0	0	0	50	0	0	0	0
4	400	100	1	1	60	10	1	1	1
5	500	200	2	4	70	20	2	4	4
$N = 5$	$\Sigma X = 1500$			$\Sigma x^2 = 10$	$\Sigma Y = 250$			$\Sigma y^2 = 10$	$\Sigma xy = 10$

But, $\Sigma d_x d_y = 186$, $\Sigma d_x = 11$, $\Sigma d_y = 7$
$N = 10$, $\Sigma d_x^2 = 215$, $\Sigma d_y^2 = 163$.

$$\therefore \quad r = \frac{186 - \frac{11 \times 7}{10}}{\sqrt{215 - \frac{(11)^2}{10}}\sqrt{163 - \frac{(7)^2}{10}}}$$

$$= \frac{186 - 77}{\sqrt{215 - 12{\cdot}1}\sqrt{163 - 4{\cdot}9}}$$

$$= \frac{178{\cdot}3}{\sqrt{202{\cdot}9}\sqrt{158{\cdot}1}} = 0{\cdot}9948$$

7. Direct Method of Calculating Coefficient of Correlation

We can find the coefficient of correlation directly without taking the deviations of x and of y from their respective means. In such cases the following formula is used.

$$\boxed{r = \frac{\Sigma xy - \frac{\Sigma x \Sigma y}{n}}{\sqrt{\left(\Sigma x^2 - \frac{(\Sigma x)^2}{n}\right)\left(\Sigma y^2 - \frac{(\Sigma y)^2}{n}\right)}}} \quad \text{.....(8)}$$

where, x and y are the observed values of the variables and $\bar{x}, \bar{y}$ are their respective means.

The formula can also be written as,

$$r = \frac{\Sigma xy - n\bar{x}\bar{y}}{n \cdot \sqrt{\left(\frac{\Sigma x^2}{n} - \bar{x}^2\right)\left(\frac{\Sigma y^2}{n} - \bar{y}^2\right)}}$$

$$\boxed{r = \frac{\Sigma xy - n\bar{x}\bar{y}}{n\sigma_x\sigma_y}} \quad \text{.....(9)}$$

Example 1 : Calculate the coefficient of correlation between X and Y from the following data.

X :	3,	5,	4,	6,	2.
Y :	3,	4,	5,	2,	6.

Sol. :

Calculations of r

x	x^2	y	y^2	xy
3	9	3	9	9
5	25	4	16	20
4	16	5	25	20
6	36	2	4	12
2	4	6	36	12
$\Sigma x = 20$	$\Sigma x^2 = 90$	$\Sigma y = 20$	$\Sigma y^2 = 90$	$\Sigma xy = 73$

Since, $\bar{x} = 4$, $\bar{y} = 4$ putting these values in equation (8).

$$\therefore \quad r = \frac{73 - \frac{20^2}{5}}{\sqrt{90 - \frac{20^2}{5}}\sqrt{90 - \frac{20^2}{5}}} = \frac{73 - 80}{10} = -0{\cdot}7.$$

EXERCISE - I

1. Calculate the coefficient of correlation from the following data. Is there any marked correlation between the production and price of tea ?

Production in crores (kgs) :	34,	27,	31,	38,	38,	36,	39,	40.	
Price in Rs. per kg :	3·75,	4·62,	4·25,	4·12,	4·28,	4·32,	4·21,	4·05.	

[**Ans. :** $r = -0{\cdot}48$]

2. Compute the coefficient of correlation between X and Y from their values given blow.

X :	30,	33,	25,	10,	33,	75,	40,	85,	90,	95.
Y :	68,	65,	80,	85,	70,	30,	55,	18,	15,	10.

[**Ans. :** $r = -0{\cdot}7069$]

3. The following data give the hardness (X) and tensile strength (Y) for some specimens of a material in certain units in a factory. Find the correlation coefficient and interpret your result.

X :	23·3,	17·5,	17·8,	20·7,	18·1,	20·9,	22·9,	20·8.
Y :	4·2,	3·8,	4·6,	3·2,	5·2,	4·7,	4·4,	5·6.

[**Ans. :** $r = -0{\cdot}072$. No correlation]

4. Calculate the product moment coefficient of correlation between the indices of business activity (X) and employment (Y) from the following data.

X :	100,	102,	108,	111,	115,	116,	118.
Y :	110,	100,	104,	108,	112,	116,	120.

[**Ans. :** $r = 0{\cdot}75$]

5. Find Karl Pearson's coefficient of correlation between X and Y.

X :	10,	12,	14,	15,	16,	17,	18,	10,	14,	15
Y :	17,	16,	15,	12,	10,	9,	8,	15,	13,	12

[**Ans. :** $r = -0{\cdot}93$]

6. Compute a coefficient of correlation between X and Y.

X :	3,	6,	4,	5,	7
Y :	2,	4,	5,	3,	6

[**Ans. :** $r = 0{\cdot}7$]

7. Calculate the coefficient of correlation between price and demand by direct method.

Price :	2,	3,	4,	7,	4
Demand :	8,	7,	3,	1,	1

[**Ans. :** $-0{\cdot}81$]

8. Calculate the coefficient of correlation between the X and Y by direct method.

X :	8,	8,	7,	5,	6,	2
Y :	3,	4,	10,	13,	22,	8

[**Ans. :** 0·2646]

9. Soil temperature (x) and Germination interval (y) for winter wheat in 12 places are as follows.

x (in ^oF) :	57,	42,	38,	42,	45,	42,	44,	40,	46,	44,	43,	40.
y (days) :	10,	26,	41,	29,	27,	27,	19,	18,	19,	31,	29,	33.

Calculate the coefficient of correlation between x and y. [**Ans. :** $r = -0{\cdot}74$]

10. Find the Karl Pearson's coefficient of correlation between X and Y from the following.

X : 51, 54, 56, 59, 65, 60, 70

Y : 38, 44, 33, 36, 33, 23, 13 [Ans. : $r = -0.7977$]

Miscellaneous Examples

Example 1 : State true or false with proper justification.

If coefficient of correlation between x and y is negative then the coefficient of correlation between $-x$ and $-y$ is positive.

Sol. : Coefficient of correlation between x and y is given by

$$r = \frac{\Sigma xy - \Sigma x \cdot \Sigma y / N}{\sqrt{\Sigma x^2 - (\Sigma x)^2 / N}\sqrt{\Sigma y^2 - (\Sigma y)^2 / N}}$$

If we change the signs of x and y both, since the product and square terms occur, the sign of r will remain the same.

$\therefore$ The statement is false.

Example 2 : Calculate the correlation coefficient between x and y from the following data.

$N = 10,\ \Sigma x = 140,\ \Sigma y = 150,\ \Sigma (x-10)^2 = 180,$

$\Sigma (y-15)^2 = 215,\ \Sigma (x-10)(y-15) = 60.$

Sol. : With usual notation $\Sigma d_x^2 = 180,\ \Sigma d_y^2 = 215,\ \Sigma d_x d_y = 60.$

Now, $\bar{x} = A + \frac{\Sigma dx}{N} \quad \therefore\ 14 = 10 + \frac{\Sigma d_x}{10} \quad \therefore\ \Sigma d_x = 40$

Similarly, $\bar{y} = B + \frac{\Sigma d_y}{N} \quad \therefore\ 15 = 15 + \frac{\Sigma d_y}{10} \quad \therefore\ \Sigma d_y = 10$

Now,
$$r = \frac{\Sigma d_x d_y - \frac{\Sigma d_x \cdot \Sigma d_y}{N}}{\sqrt{\Sigma d_x^2 - \frac{(\Sigma d_x)^2}{N}}\sqrt{\Sigma d_y^2 - \frac{(\Sigma d_y)^2}{N}}}$$

$$= \frac{60 - \frac{40 \times 0}{10}}{\sqrt{180 - \frac{(40)^2}{10}}\sqrt{215 - \frac{(0)^2}{10}}} = \frac{60}{\sqrt{20}\sqrt{215}} = 0.915.$$

EXERCISE - II

Type I

1. Find the coefficient of correlation (r) between x and y for the following data.

x :	62	64	65	69	70	71	72	74
y :	126	125	139	145	165	152	180	208

[Ans. : 0·9032]

2. Find Karl Pearson's coefficient of correlation of correlation for the following data.

x :	12	17	22	27	32
y :	113	119	117	115	121

Also interpret your result. [Ans. : $r = 0.6$]

3. The following data gave the growth of employment in lakhs in organised sector in India between 1988 and 1995.

Year :	1988,	89,	90,	91,	92,	93,	94,	95.
Public Sector :	98,	101,	104,	107,	113,	120,	125,	128.
Private Sector :	65,	65,	67,	68,	68,	69,	68,	68.

Find the correlation coefficient (r) between the employment in public and private sectors and give your comments. [Ans. : $r = 0.98$]

4. Calculate the coefficient of correlation from the following figures. Is there any marked correlation between the production and price of tea ?

Production in crores of lbs. :	44,	37,	31,	38,	36,	35,	40.	
Price in Rs. per lbs. :	2·75,	3·62,	4·25,	4·12,	4·28,	4·32,	4·05.	

5. Draw a scatter diagram to represent the following data.

X :	2,	4,	5,	6,	8,	11.
Y :	18,	12,	10,	8,	7,	5.

Calculate the coefficient of correlation between X and Y for the above data. [Ans. : $r = -0.92$]

6. Find the coefficient of correlation between height of father and height of son from the following data.

Height of father :	65,	66,	67,	67,	68,	69,	71,	73.
Height of son :	67,	68,	64,	68,	72,	70,	69,	70.

[Ans. : $r = +0.55$]

Type II

1. Find the number of pairs of observations from the following data.

$r = 0.4,\ \Sigma xy = 108,\ \sigma_y = 3,\ \Sigma x^2 = 900.$

where x, y are the deviations of x, y from there respective means. [Ans. : $N = 10$]

2. Coefficient of correlation between two variables is 0·4. Their covariance is 12. The variance of x is 25. Find the standard deviation of y. [Ans. : $\sigma_y = 6$]

3. A computer while calculating the correlation coefficient between two variables x and y, from 25 observations obtained the following results

$N = 25,\ \Sigma x = 125,\ \Sigma y = 100,\ \Sigma x^2 = 650,\ \Sigma y^2 = 960,\ \Sigma xy = 508$

where x, y denote the actual values of the variables. Find the value of r. **(S.U. 1999, 2004)**

[Ans. : 0·067]

4. A sample of 25 pairs of values of x and y lead to the following results.

$\Sigma x = 127,\ \Sigma y = 100,\ \Sigma x^2 = 760,\ \Sigma y^2 = 449,\ \Sigma xy = 500.$

Later on it was found that two pairs of values were taken as (8, 14) and (8, 6) instead of correct values (8, 12) and (6, 8).

Find corrected correlation coefficient between x and y. [Ans. : $r = -0.31$]

5. Given : Number of pairs of observations = 10

X series standard deviation = 22·70. Y series standard deviation = 9·592

Summation of the products of corresponding deviations of X and Y from their respective actual means = – 1439. Find r. [Ans. : $r = -0.66$]

EXERCISE - III

Theory

1. What is meant by correlation ? Describe scatter diagram and interpret.
2. Define - (i) Karl Pearson's coefficient of correlation, (ii) Coefficient of variation.
3. Two variables x and y are connected by the relation $ax + by + c = 0$. Show that the coefficient of correlation is either + 1 or – 1.
4. Define product moment correlation coefficient and show that it is always numerically less than or equal to unity.
5. Define - (i) Covariance between two variables x and y, (ii) Karl Pearson's Product moment correlation coefficient.
6. What is scatter diagram ? How does it help in studying the correlation between two variables. Draw scatter diagrams for $r = +1$, $r = -1$ and $r = 0$.
7. Define the Karl Pearson's coefficient of correlation r between two variables x and y. What is "Spurious Correlation" ?

   Interpret the cases $r = +1$, $r = -1$, $r = 0$. Also draw the scatter diagrams corresponding to these cases.

☆☆☆

CHAPTER 2

Regression

1. Introduction

We have seen in the previous chapter how to examine and measure in magnitude and direction correlation between two variables. After establishing correlation, it is natural to search for a method which will help us to estimate the value of one variable when that of the other is known. This is achieved by the analysis of regression. Regression can be defined as **'a method of estimating the value of one variable when that of the other is known and when the variables are correlated'**.

The term, regression was first used by Galton. He found that although tall fathers have tall sons, and short fathers have short sons, the average height of sons of tall fathers is less than the average height of their fathers and the average height of sons of short fathers is more than the average height of their fathers. In other words the average height of sons of tall fathers or short fathers will regress or go back to the general average height. This phenomenon was described by him as 'regression.'

2. Lines of Regression

We have seen in the previous chapter that if the variables which are highly correlated are plotted on a graph then the points lie in a narrow strip. If the strip is nearly straight, we may draw a line such that all the points are close to it from both the sides. Such a line can be taken as the representative of the ideal variation. It is called the line of best fit. It is a line such that the sum of the distances of the points from the line is minimum. It is also called **'the line of regression'**. But we do not measure the distance by dropping a perpendicular from a point to the line. We measure, the deviations (i) vertically and (ii) horizontally, and get one line when distances are minimised vertically and second line when distances are minimised horizontally. Thus, we get **two lines of regression.**

(i) Line of regression of Y on X

If we minimise the deviations of the points from the line measured along y-axis we get a line which is called **the line of regression of Y on X**. Its equation is written in the from $Y = a + bX$. **This line is used for estimating the value of Y for a given value of X.** [See Fig. 2.1]

The equation of the line of regression of y on x must be written with y on the left hand side and x and the constant term on the right hand side.

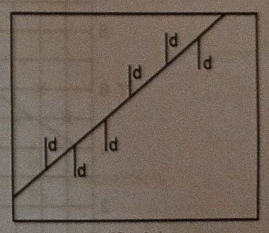


Fig. 2.1
For the Line of Regression of Y on X the distances d are minimised.

(ii) Line of regression of X on Y

If we minimise the deviations of the points from the line measured along x-axis we get a line which is called **the line of regression of X on Y.** Its equation is written in the form $X = a + bY$. **This line is used for estimating the value of X for a given value of Y.** [See Fig. 2.2]

The equation of the line of regression of x on y must be written with x on the left hand side and y and constant term on the right hand side.

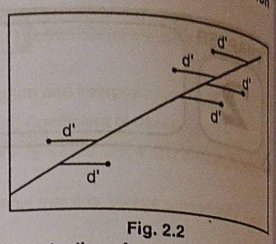


Fig. 2.2
For the line of regression of X on Y the distances d' are minimised.

There are two methods of obtaining the lines of regression. The first is graphical, the other is mathematical. They are :

1. The method of Scatter Diagram,
2. The Method of Least Squares.

3. The Method of Scatter Diagram

It is the simplest method of obtaining the lines of regression. The data are plotted on a graph paper by taking the independent variable on x-axis and the dependent variable on y-axis. We thus get points which are generally scattered. If the correlation is perfect *i.e.* if r is equal to one, positive or negative, the points will lie on a line, which is the line of regression. And there is only one line of regression and not two in such cases. However, in practice we rarely come across problems wherein we have perfect correlation. Usually, the points are scattered in a narrow straight strip and we have to find a line which will best represent all the points of the scatter diagram. We draw a line which will be close to all the points as far as possible.

Example : Given the following pairs of values of X and Y.

X :	1,	2,	3,	4,	5,	6,	7,	8,	9,	10,	11,	12.
Y :	5,	6,	5,	6,	6,	8,	7,	9,	8,	9,	10,	11.

Plot the points on a graph and draw a line of regression.

Sol. : **Scatter Diagram and Line of Regression**

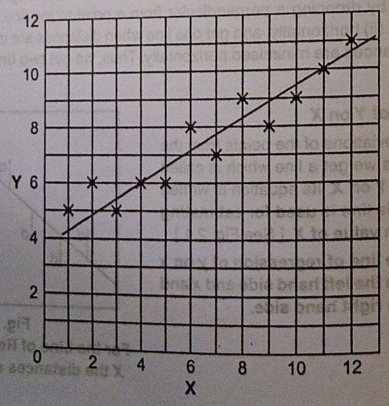


Fig. 2.3

4. The Method of Least Square

This is a mathematical method which gives an objective treatment to find a line of regression.

Let $y = a + bx$ be the equation of the line required. To find the line of regression of y on x we minimise the sum of the absolute distances of the points like $P(x_i, y_i)$ from the line measured along the y-axis. If Q is the point on the line corresponding to $P(x_i, y_i)$ we have to minimise the absolute distance PQ. Since Q lies on $y = a + bx$ its y-coordinate $= a + bx_i$

$$\therefore \quad |PQ| = |y_i - a - bx_i|$$

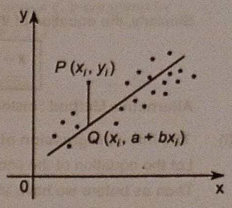


Fig. 2.4

For minimising $|PQ|$ we minimise its squares. Hence, if S denotes the sum of the squares of these distances,

$$S = \Sigma f_i (y_i - a - bx_i)^2 \qquad \text{where } f_i \text{ is the frequency of } (x_i, y_i).$$

We have to find a and b such that S is minimum, the conditions for which are

$$\frac{\partial S}{\partial a} = 2\Sigma f_i (y_i - a - bx_i) = 0 \quad \text{and} \quad \frac{\partial S}{\partial b} = 2\Sigma f_i (y_i - a - bx_i) x_i = 0$$

$$\Sigma f_i (y_i - a - bx_i) = 0 \qquad \text{..... (A)}$$

$$\Sigma x_i f_i (y_i - a - bx_i) = 0 \qquad \text{..... (B)}$$

From (A) we get, $\Sigma f_i y_i - a\Sigma f_i - b\Sigma f_i x_i = 0$

$$\therefore \quad N\bar{y} - aN - bN\bar{x} = 0 \quad \therefore \quad \bar{y} = a + b\bar{x} \qquad \text{..... (C)}$$

which shows that **the line of regression passes through** $(\bar{x}, \bar{y})$.

From (B) we get, $\Sigma f_i x_i y_i - a\Sigma f_i x_i - b\Sigma f_i x_i^2 = 0$ ..... (D)

We now find the values of these expressions in terms of r, σ_x, σ_y.

But since, $r = \dfrac{1}{N} \dfrac{\Sigma f_i (x_i - \bar{x})(y_i - \bar{y})}{\sigma_x \sigma_y}$

$$r = \frac{\Sigma f_i x_i y_i - N\bar{x}\bar{y}}{N\sigma_x \sigma_y} \quad \therefore \quad \Sigma f_i x_i y_i = Nr\sigma_x\sigma_y + N\bar{x}\bar{y}$$

and $\sigma_x^2 = \dfrac{1}{N}\Sigma f_i (x_i - \bar{x})^2 = \dfrac{1}{N}\Sigma f_i x_i^2 - \bar{x}^2 \quad \therefore \quad \Sigma f_i x_i^2 = N\sigma_x^2 + N\bar{x}^2$

Putting the values of $\Sigma f_i x_i y_i$ and $\Sigma f_i x_i^2$ in D, we get

$$Nr\sigma_x\sigma_y + N\bar{x}\bar{y} = aN\bar{x} + bN\sigma_x^2 + bN\bar{x}^2$$

i.e. $\quad r\sigma_x\sigma_y + \bar{x}\bar{y} = a\bar{x} + b\sigma_x^2 + b\bar{x}^2$ ..... (E)

Multiply (C) by $\bar{x}$ and subtract it form (E)

$$r\sigma_x\sigma_y = b\sigma_x^2 \quad \therefore \quad b = r\frac{\sigma_y}{\sigma_x}$$

Since, the line passes through $(\bar{x}, \bar{y})$ and its slope $b = r\dfrac{\sigma_y}{\sigma_x}$ its equation is

$$\boxed{y - \bar{y} = r\frac{\sigma_y}{\sigma_x}(x - \bar{x})} \qquad \text{..... (F)}$$

Similarly, the equation of the line of regression of x on y can be shown to be

$$\boxed{x - \bar{x} = r\frac{\sigma_x}{\sigma_y}(y - \bar{y})} \quad \text{..... (G)}$$

Alternative Method : Instead of calculating $\bar{x}, \bar{y}, \sigma_x, \sigma_y$ and r we may use the following method.

(i) **The line of regression of y on x**

Let the equation of the line of regression of y on x be $y = a + bx$.

Then as before we have to minimise

$$S = \Sigma f_i(y_i - a - bx_i)^2 \quad \text{the conditions of which are}$$

$$\frac{\partial S}{\partial a} = -2\Sigma f_i(y_i - a - bx_i) = 0 \quad \text{and} \quad \frac{\partial S}{\partial b} = -2\Sigma f_i(y_i - a - bx_i)\, x_i = 0$$

i.e. $\Sigma f_i(y_i - a - bx_i) = 0$ and $\Sigma f_i x_i (y_i - a - bx_i) = 0$

i.e. $\Sigma f_i y_i - a\Sigma f_i - b\Sigma f_i x_i = 0$ and $\Sigma f_i x_i y_i - a\Sigma f_i x_i - b\Sigma f_i x_i^2 = 0$.

If the required sums are known, by solving the above two equations simultaneously for a and b we get the required equation.

In particular if all values occur only once *i.e.* if $f_i = 1$ for all i then the above equations take the form

$$\boxed{\Sigma y = aN + b\Sigma x} \quad \text{and} \quad \boxed{\Sigma xy = a\Sigma x + b\Sigma x^2}$$

from which a and b can be calculated.

The above equations are called **normal equations.**

(ii) **The line of regression of x on y**

Let the equation of the line be $x = a + by$.

Proceeding as above we get the normal equations as

$$\boxed{\Sigma x = aN + b\Sigma y} \quad \text{and} \quad \boxed{\Sigma xy = a\Sigma y + b\Sigma y^2}$$

from which we can find the values of a and b.

Example : Find the equations of the lines of regression from the following data.

x: 5 6 7 8 9

y: 2 4 5 6 8 Also find r.

Sol. : **Calculations of regression**

Sr. No.	x	x^2	y	y^2	xy
1	5	25	2	4	10
2	6	36	4	16	24
3	7	49	5	25	35
4	8	64	6	36	48
5	9	81	8	64	72
$N = 5$	35	255	25	145	189

The equation of the line of regression of y on x is $y = a + bx$ where a, b are given by

$$\Sigma y = aN + b\Sigma x \quad \text{and} \quad \Sigma xy = a\Sigma x + b\Sigma x^2$$

Putting the values of $\Sigma x, \Sigma x^2, \Sigma xy$, we get

$25 = 5a + 35b$ ..... (i) and $189 = 35a + 255b$ ..... (ii)

Multiply the first by 7 and subtract it from the second.

$$\begin{aligned} 189 &= 35a + 255b \\ 175 &= 35a + 245b \\ \hline 14 &= \quad\quad 10b \end{aligned} \qquad \therefore\ b = 1{\cdot}4$$

Putting this value of b in (i), we get

$25 = 5a + 35\,(1{\cdot}4)$ $\therefore\ 5a = 25 - 49 = -24$ $\therefore\ a = -4{\cdot}8$

$\therefore$ The equation of the line of regression of y on x is

$y = -4{\cdot}8 + 1{\cdot}4x$

The equation of the line of regression of x on y is $x = a + by$ where a, b are given by

$$\Sigma x = aN + b\Sigma y \quad \text{and} \quad \Sigma xy = a\Sigma y + b\Sigma y^2$$

Putting the values of $\Sigma x, \Sigma y, \Sigma xy, \Sigma y^2$, we get

$35 = 5a + 25b$ ..... (iii) and $189 = 25a + 145b$ ..... (iv)

Multiply (iii) by 5 and subtract it from (iv)

$$\begin{aligned} 189 &= 25a + 145b \\ 175 &= 25a + 120b \\ \hline 14 &= \quad\quad 25b \end{aligned} \qquad \therefore\ b = 0{\cdot}56$$

Putting this value of b in (iii), we get

$35 = 5a + 25\,(0{\cdot}56)$ $\therefore\ 5a = 35 - 14 = 11$ $\therefore\ a = 2{\cdot}2$

$\therefore$ The equation of the line of regression of x on y is

$x = 2{\cdot}2 + 0{\cdot}56y$

Now, $r = \sqrt{b_1 \times b_2} = \sqrt{1{\cdot}4 \times 0{\cdot}56} = 0{\cdot}88$.

5. Calculations of the Equations of the Lines of Regression

There are various methods of calculating the equations of the lines of regression. The choice is yours. We state them below.

(a) **By calculating the coefficient of correlation r and standard deviation σ_x and σ_y.**

The equation of the line of regression of Y on X.

The equation of the line of regression of Y on X is given by,

$$Y - \bar{Y} = b_{yx}(X - \bar{X})$$

where, $\bar{X}$ = the mean of X, $\bar{Y}$ = the mean of Y and $b_{yx} = r\dfrac{\sigma_y}{\sigma_x}$. [See § 6]

$\therefore$ The equation, therefore can be written as

$$\boxed{Y - \bar{Y} = r\frac{\sigma_y}{\sigma_x}(X - \bar{X})} \quad \text{.....(1)}$$

This equation is to be used for calculating the most probable value of Y for a given value of X.

The equation of the line of regression of X on Y.

The equation of the line of regression of X on Y is given by,

$$X - \bar{X} = b_{xy}(Y - \bar{Y})$$

where, $\bar{X}$ = the mean of X, $\bar{Y}$ = the mean of Y and $b_{xy} = r\frac{\sigma_x}{\sigma_y}$.

$\therefore$ The equation, therefore can be written as

$$\boxed{X - \bar{X} = r\frac{\sigma_x}{\sigma_y}(Y - \bar{Y})} \quad \text{.....(2)}$$

This equation is to be used for calculating the most probable value of X for a given value of Y.

When the values of X and Y are known, we can as usual calculate r, σ_x and σ_y and then obtain equations of the lines of regression of X and Y.

(b) By calculating a and b directly

Instead of calculating $\bar{X}, \bar{Y}, \sigma_x, \sigma_y$ and r we may calculate a and b directly as explained below.

The line of regression of Y on X : The constants a and b of the equation of the line of regression of Y on X *i.e.* of,

$$Y = a + bX$$

can be obtained by solving the following simultaneous equations.

$$\boxed{\Sigma Y = aN + \Sigma X} \quad \text{and} \quad \boxed{\Sigma YX = a\Sigma X + b\Sigma X^2} \quad \text{.....(3)}$$

The line of regression of X on Y : The constants a and b of the equation of the line of regression of X on Y *i.e.* of,

$$X = a + bY$$

can be obtained by solving the following simultaneous equations,

$$\boxed{\Sigma X = aN + b\Sigma Y} \quad \text{and} \quad \boxed{\Sigma XY = a\Sigma Y + b\Sigma y^2} \quad \text{.....(4)}$$

The equations, are called 'Normal equations'.

The formulae (3) and (4) are convenient when $\Sigma X, \Sigma Y, \Sigma XY, \Sigma X^2, \Sigma Y^2$ are known.

(c) By taking deviations from the means X and Y

If x and y denote the deviations of X and Y from their means, we know that,

$$b_{yx} = r\frac{\sigma_y}{\sigma_x} = \frac{\Sigma xy}{N\sigma_x\sigma_y}\cdot\frac{\sigma_y}{\sigma_x} = \frac{\Sigma xy}{N\sigma_x^2} = \frac{\Sigma xy}{N\cdot\Sigma x^2/N} = \frac{\Sigma xy}{\Sigma x^2}$$

$$\therefore \boxed{b_{yx} = \frac{\Sigma xy}{\Sigma x^2}} \quad \text{.....(5)}$$

Similarly, we can show that,

$$b_{xy} = r\frac{\sigma_x}{\sigma_y} = \frac{\Sigma xy}{\Sigma y^2} \quad \therefore \quad \boxed{b_{xy} = \frac{\Sigma xy}{\Sigma y^2}} \quad \text{.....(6)}$$

where, $x = X - \bar{X}$, $y = Y - \bar{Y}$.

Hence, the equations of the lines of regression become,

$$\boxed{Y - \bar{Y} = \frac{\Sigma xy}{\Sigma x^2}(X - \bar{X})} \quad \text{.....(7)}$$

and

$$\boxed{X - \bar{X} = \frac{\Sigma xy}{\Sigma y^2}(Y - \bar{Y})} \quad \text{.....(8)}$$

(d) By taking deviations from assumed means

If the deviations of X and Y are taken from assumed means *i.e.* if $d_x = X - A$ and $d_y = Y - B$ then the coefficients b_{yx} and b_{xy} are given by,

$$\boxed{b_{yx} = \frac{\Sigma d_x d_y - \frac{\Sigma d_x \Sigma d_y}{N}}{\Sigma d_x^2 - \frac{(\Sigma d_x)^2}{N}}} \quad \text{.....(9)}$$

$$\boxed{b_{xy} = \frac{\Sigma d_x d_y - \frac{\Sigma d_x \Sigma d_y}{N}}{\Sigma d_y^2 - \frac{(\Sigma d_y)^2}{N}}} \quad \text{.....(10)}$$

(e) By using actual values directly

If X, Y are actual values of the two variates then it can be shown that

$$\boxed{b_{yx} = \frac{\Sigma XY - \frac{\Sigma X\cdot\Sigma Y}{N}}{\Sigma X^2 - \frac{(\Sigma X)^2}{N}}} \quad \text{.....(11)}$$

$$\boxed{b_{xy} = \frac{\Sigma XY - \frac{\Sigma X\cdot\Sigma Y}{N}}{\Sigma Y^2 - \frac{(\Sigma Y)^2}{N}}} \quad \text{.....(12)}$$

6. Regression Coefficients

The slope b of the line of regression of y on x *i.e.* b of the equation $y = a + bx$ is called the **coefficient of regression of y on x.** It represents the increment in y for unit change in the value of x. It is denoted by b_{yx}.

$\therefore$ b_{yx} = Coefficient of Regression of y on x.

$$b_{yx} = r\frac{\sigma_y}{\sigma_x}$$

Similarly, the slope b of the line of regression x on y *i.e.* b of the equation $x = a + by$ is called the **coefficient of regression of x on y**. It represents the increment in x for unit change in y. It is denoted by b_{xy}.

$\therefore$ b_{xy} = Coefficient of Regression of x on y.

$$b_{xy} = r\frac{\sigma_x}{\sigma_y}$$

Putting b_{yx} and b_{xy} which are the slopes of the lines of regression in (F) and (G), we can write the equations of lines of regression as

$$\boxed{y - \bar{y} = b_{yx}(x - \bar{x})} \quad \text{.....(F)}$$

and

$$\boxed{x - \bar{x} = b_{xy}(y - \bar{y})} \quad \text{.....(G)}$$

By putting the values of r and σ_y, σ_x in terms of actual values of x and y or by taking deviations from actual means or assumed means, we get the following formulae for b_{yx} and b_{xy}. (We repeat the first two.)

$$\boxed{b_{yx} = r\frac{\sigma_y}{\sigma_x}} \quad \text{.....(1)}$$

$$\boxed{b_{yx} = \frac{\Sigma xy}{\Sigma x^2}} \quad \text{.....(2)}$$

where $x = X - \bar{x}$, $y = Y - \bar{y}$ *i.e.* x, y are deviations of X, Y from actual means.

$$\boxed{b_{yx} = \frac{\Sigma d_x d_y - \dfrac{\Sigma d_x \cdot \Sigma d_y}{N}}{\Sigma d_x^2 - \dfrac{(\Sigma d_x)^2}{N}}} \quad \text{.....(3)}$$

where $d_x = X - A$, $d_y = Y - B$ *i.e.* d_x, d_y are deviations of X, Y from assumed means A and B.

$$\boxed{b_{yx} = \frac{\Sigma XY - \dfrac{\Sigma X \cdot \Sigma Y}{N}}{\Sigma X^2 - \dfrac{(\Sigma X)^2}{N}}} \quad \text{.....(4)}$$

where X, Y are the actual values of the variables.

Also,

$$\boxed{b_{xy} = r\frac{\sigma_x}{\sigma_y}} \quad \text{.....(1')}$$

$$\boxed{b_{xy} = \frac{\Sigma xy}{\Sigma y^2}} \quad \text{.....(2')}$$

where $x = X - \bar{x}$, $y = Y - \bar{y}$ *i.e.* x, y are the deviations of X, Y from actual means

$$\boxed{b_{xy} = \frac{\Sigma d_x d_y - \dfrac{\Sigma d_x \cdot \Sigma d_y}{N}}{\Sigma d_y^2 - \dfrac{(\Sigma d_y)^2}{N}}} \quad \text{.....(3')}$$

where $d_x = X - A$, $d_y = Y - B$ *i.e.* d_x, d_y are the deviations of X, Y from assumed means A and B.

$$\boxed{b_{xy} = \frac{\Sigma XY - \dfrac{\Sigma X \cdot \Sigma Y}{N}}{\Sigma Y^2 - \dfrac{(\Sigma Y)^2}{N}}} \quad \text{.....(4')}$$

where X, Y are the actual values of the variables.

7. Properties of Coefficients of Regression

1. Coefficient of correlation is the geometric mean between the coefficients of regression.

Proof : From the above results we have

$$b_{yx} \cdot b_{xy} = r\frac{\sigma_y}{\sigma_x} \cdot r\frac{\sigma_x}{\sigma_y}$$

$$\therefore \quad \boxed{b_{yx} \cdot b_{xy} = r^2} \quad \text{Hence, the result.}$$

Remark ....

Since the product of b_{yx} and b_{xy} is positive, if one of them is negative, the other also must be negative. In other words both the coefficients of regression are positive or both the coefficients of regression are negative together.

2. If one coefficient of regression is greater than one, the other must be less than one.

Proof : Since $-1 \le r \le 1$, $r^2 \le 1$.

Hence, from the above result,

$$b_{yx} \cdot b_{xy} \le 1 \quad \therefore \; b_{yx} \le \frac{1}{b_{xy}} \qquad \therefore \text{ If } b_{yx} < 1,\; b_{xy} > 1.$$

[See the values of b_{yx} and b_{xy} in Ex. 4, page 2-16. See that b_{yx} is less than one and b_{xy} is greater than one.]

3. Arithmetic mean of the coefficients of regression is greater than or equal to the coefficient of correlation.

Proof : We have to show that $\dfrac{b_{yx} + b_{xy}}{2} \ge r$

i.e. $\dfrac{1}{2}\left(r\dfrac{\sigma_y}{\sigma_x} + r\dfrac{\sigma_x}{\sigma_y}\right) \ge r$ *i.e.* $\dfrac{\sigma_y}{\sigma_x} + \dfrac{\sigma_x}{\sigma_y} \ge 2$

i.e. $\sigma_x^2 + \sigma_y^2 \ge 2\sigma_x\sigma_y$ *i.e.* $\sigma_x^2 - 2\sigma_x\sigma_y + \sigma_y^2 \ge 0$

i.e. $(\sigma_x - \sigma_y)^2 \ge 0$ which is obviously true.

Remark ....

In other words this means **the sum of the two coefficients of regression is greater than or equal to $2r$.** We shall verify this in Ex. 2, page 2-19; Ex. 5, page 2-17 and Ex. 6, page 2-22. We further note that equality will hold when $\sigma_x = \sigma_y$.

4. Coefficients of regression are independent of change of origin but not of change of scale.

If $u = ax + h$ and $v = by + k$, then

$$\boxed{b_{uv} = \frac{a}{b} \cdot b_{xy} = \frac{\text{Coefficient of } x}{\text{Coefficient of } y} \cdot b_{xy}}$$

We accept this result without proof.

5. If the correlation is perfect then the two coefficients of regression are reciprocals of each other.

Proof : We have $r = \pm 1$ and $r = \sqrt{b_{yx} \cdot b_{xy}}$ $\quad \therefore \quad \pm 1 = \sqrt{b_{yx} \cdot b_{xy}}$.

Squaring $1 = b_{yx} \cdot b_{xy}$ $\quad \therefore \quad b_{yx} = \dfrac{1}{b_{xy}}$.

e.g., if one coefficient of regression is 0·5 and if the correlation is perfect, then the other coefficient of regression is 2.

Example 1 : State whether the following statement is true or false with reasoning : "The regression coefficients between $2x$ and $2y$ are the same as those between x and y".

Sol. : As seen above if $u = ax + h$ and $v = by + k$, $\quad b_{uv} = \dfrac{a}{b} \cdot b_{xy}$.

But by data $u = 2x$ *i.e.* $a = 2$ and $v = 2y$ *i.e.* $b = 2$.

$\therefore \quad b_{vu} = \dfrac{2}{2} \cdot b_{xy} = b_{xy}$. $\quad$ Hence, the statement is true.

Example 2 : State whether the following statement is true or false : "The lines of regression between x and y are parallel to the lines of regression between $2x$ and $2y$.

Sol. : True. Explanation is left to you.

Example 3 : State whether the following statement is true or false : "The coefficients of regression between x and y are the same as the coefficients of regression between $2x + 5$ and $2y - 7$.

Sol. : True. Explanation is left to you.

Example 4 : If the arithmetic mean of regression coefficients is p and their difference is $2q$ find the correlation coefficient.

Sol. : Let the coefficients of regression be b_1 and b_2.

Now by data $\quad \dfrac{b_1 + b_2}{2} = p$ and $b_1 - b_2 = 2q$

$\therefore \quad b_1 + b_2 = 2p$ and $b_1 - b_2 = 2q$

$\therefore \quad b_1 = p + q$ and $b_2 = p - q$

$\therefore$ Coefficient of correlation $= r = \sqrt{b_1 b_2} = \sqrt{p^2 - q^2}$.

Example 5 : State true of false with reasoning : "$2x + y = 3$ and $x = 2y + 3$ cannot be the lines of regression."

Sol. : If the first line is the line of regression of y on x it must written as $y = -2x + 3$ and if the second line is the line of regression of x on y, then it must be written as $x = 2y + 3$.

Hence, the coefficients of regression are $b_{yx} = -2$ and $b_{xy} = 2$ which is not possible as one of the them is negative and the other is positive and both are greater 1 numerically.

Now, we consider the lines in other way round. Let the first line be the line of regression of x on y and let the second line be the line of regression of y on x.

$$\therefore \quad x = -\frac{1}{2}y + \frac{3}{2} \quad \text{and} \quad y = \frac{1}{2}x - \frac{3}{2}.$$

Hence, the coefficients of regression are

$$b_{yx} = -\frac{1}{2} \quad \text{and} \quad b_{xy} = \frac{1}{2}$$

which is again not possible because one is positive and the other is negative.

Hence, the statement is true.

Example 6 : State true or false with justification. If two lines of regression are $x + 3y - 5 = 0$ and $4x + 3y - 8 = 0$ then the correlation coefficient is + 0·5.

Sol. : Let the line $x + 3y - 5 = 0$ be the line of regression of x on y. Writing it as $x = -3y + 5$, we get $b_{xy} = -3$.

Let the line $4x + 3y - 8 = 0$ be the line of regression of y on x. Writing it as $3y = -4x + 8$ *i.e.*, as $y = -\dfrac{4}{3}x + 2$, we get $b_{yx} = -\dfrac{4}{3}$.

$$\therefore \quad r = \sqrt{b_{yx} \cdot b_{xy}} = \sqrt{(-3)(-4/3)} = \sqrt{4} = 2$$

But r cannot be greater than 1.

Hence, our suppositions are wrong.

Now, let the line $x + 3y - 5 = 0$ be the line of regression of y on x. Writing it as

$$3y = -x + 5 \qquad \textit{i.e.,} \quad y = -\frac{1}{3}x + \frac{6}{3}, \text{ we get } \quad b_{yx} = -\frac{1}{3}.$$

Let the line $4x + 3y - 8 = 0$ be the line of regression of x on y. Writing it as

$$4x = -3y + 8 \quad \textit{i.e.,} \quad x = -\frac{3}{4}y + 2, \text{ we get } \quad b_{xy} = -\frac{3}{4}.$$

Now, $\quad r = \sqrt{b_{yx} \cdot b_{xy}} = \sqrt{\left(-\dfrac{1}{3}\right)\left(-\dfrac{3}{4}\right)} = \sqrt{\dfrac{1}{4}} = \dfrac{1}{2} = 0{\cdot}5$

Hence, the statement is true.

6. Angle between the lines of regression

The equation of the lines of regression of y on x is

$$y - \bar{y} = r\frac{\sigma_y}{\sigma_x}(x - \bar{x}). \qquad \text{Hence, its slope } m_1 = r\frac{\sigma_y}{\sigma_x}$$

The equation of the line of regression of x on y is

$$x - \bar{x} = r\frac{\sigma_x}{\sigma_y}(y - \bar{y})$$

$$\textit{i.e.,} \quad y - \bar{y} = \frac{\sigma_y}{r \cdot \sigma_x}(x - \bar{x}) \qquad \text{Hence, its slope } m_2 = \frac{\sigma_y}{r \cdot \sigma_x}.$$

If θ is the angle between the lines of regression

$$\therefore \quad \tan\theta = \frac{m_1 - m_2}{1 + m_1 m_2} = \frac{r\frac{\sigma_y}{\sigma_x} - \frac{\sigma_y}{r\sigma_x}}{1 + \frac{r\sigma_y}{\sigma_x} \cdot \frac{\sigma_y}{r\sigma_x}} = \frac{\frac{(r^2\sigma_y\sigma_x - \sigma_y\sigma_x)}{r\sigma_x^2}}{\frac{\sigma_x^2 + \sigma_y^2}{\sigma_x^2}}$$

$$= \frac{(r^2 - 1)}{r}\left(\frac{\sigma_x\sigma_y}{\sigma_x^2 + \sigma_y^2}\right) = \frac{1 - r^2}{r}\left(\frac{\sigma_x\sigma_y}{\sigma_x^2 + \sigma_y^2}\right)$$

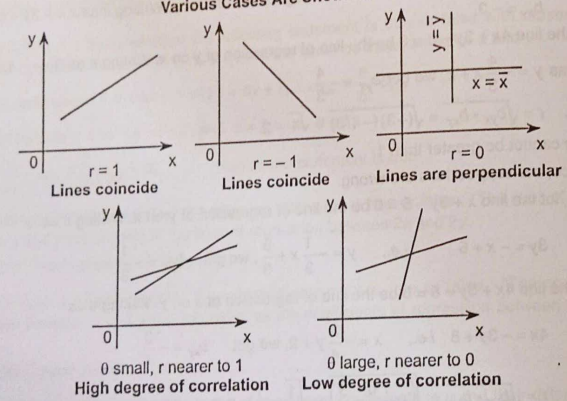
Corollary 1 : If $r = 0$, $\tan\theta = \infty$ $\therefore \theta = \pi/2$.

The lines of regression are perpendicular to each other.

Corollary 2 : If $r = \pm 1$, $\tan\theta = 0$ $\therefore \theta = 0$.

The lines of regression are coincident.

Various Cases Are Shown Below.



Illustrative Examples

Example 1 : A panel of two judges A and B graded dramatic performances by independently awarding marks as follows :

Performance No.	:	1,	2,	3,	4,	5,	6,	7.
Marks by A	:	36,	32,	34,	31,	32,	32,	34.
Marks by B	:	35,	33,	31,	30,	34,	32,	36.

The eighth performance, however, which judge B could not attend, got 38 marks by judge A. If judge B had also been present, how many marks would he be expected to have awarded to the eighth performance ?

Sol. : **Calculations of coefficient of correlation etc.**

Sr. No.	$X - \bar{X}$			$Y - \bar{Y}$			Product xy
	X	x	x^2	Y	y	y^2	
1	36	3	9	35	2	4	6
2	32	– 1	1	33	0	0	0
3	34	1	1	31	– 2	4	– 2
4	31	– 2	4	30	– 3	9	6
5	32	– 1	1	34	1	1	– 1
6	32	– 1	1	32	– 1	1	1
7	35	1	1	36	3	9	3
$N = 7$	$\Sigma X = 231$		$\Sigma x^2 = 18$	$\Sigma Y = 231$		$\Sigma y^2 = 28$	$\Sigma xy = 13$

We have to find the marks that would have been awarded by the judge B. Therefore, let the marks given by the judge B be denoted by Y and those given by A by X.

$$\therefore \quad \bar{X} = \frac{\Sigma X}{N} = \frac{231}{7} = 33, \quad \bar{Y} = \frac{\Sigma Y}{N} = \frac{231}{7} = 33$$

$$\therefore \quad b_{yx} = \frac{\Sigma xy}{\Sigma x^2} = \frac{13}{18} = 0{\cdot}72$$

The equation of the line of regression of Y on X is,

$$Y - \bar{Y} = b_{yx}(X - \bar{X}) \quad i.e. \quad Y - 33 = 0{\cdot}72(X - 33)$$

To find the value of Y when $X = 38$, put $X = 38$ in the above equation.

$$\therefore \quad Y - 33 = 0{\cdot}72(38 - 43) = 0{\cdot}72 \times 5 = 3{\cdot}6$$

$$\therefore \quad Y = 33 + 3{\cdot}76 = 36{\cdot}6 = 37 \text{ approximately.}$$

$\therefore$ The judge B would have given 37 marks to the eighth performannce.

Alternatively : **Calculations of regression**

Sr. No.	x	x^2	y	xy
1	36	1296	35	1260
2	32	1024	33	1056
3	34	1156	31	1054
4	31	961	30	930
5	32	1024	34	1088
6	32	1024	32	1024
7	34	1156	36	1224
$N = 7$	231	7641	231	7636

Let the marks given by A be x and those given by B be y.

Then the line of regression of y on x is $y = a + bx$.

And the normal equations are

$$\Sigma y = Na + b\Sigma x \qquad \therefore \quad 231 = 7a + 231b \quad \text{..... (i)}$$

$$\Sigma xy = a\Sigma x + b\Sigma x^2 \qquad \therefore \quad 7636 = 231a + 7641b \quad \text{..... (ii)}$$

To find the value of Y when $X = 3$ put this value in the above equation,

$\therefore \quad Y - 1800 = 52{\cdot}86\,(3 - 7) = -211{\cdot}44$

$\therefore \quad Y = 1800 - 212.64 = 1588{\cdot}56$

$\therefore$ The cost of maintenance of 3 years old car = ₹ 1588·56.

Example 4 : Find the coefficients of regression and hence the equations of the lines of regression for the following data.

X : 78, 36, 98, 25, 75, 82, 90, 62, 65, 39.

Y : 84, 51, 91, 60, 68, 62, 86, 58, 53, 47.

Draw the lines of regression from your equations on the graph. Estimate the value of Y when $X = 50$ and the value of X when $Y = 90$ from the graph.

What is the significance of the point of intersection of the two lines?

Sol. :

Calculations of coefficients of regression

Sr. No.	$X - \bar{X}$			$Y - \bar{Y}$			Product xy
	X	x	x^2	Y	y	y^2	
1	78	+ 13	169	84	+ 18	324	234
2	36	− 29	841	51	− 15	225	435
3	98	+ 33	1089	91	+ 25	625	825
4	25	− 40	1600	60	− 6	36	240
5	75	+ 10	100	68	+ 2	4	20
6	82	+ 17	289	62	− 4	16	− 68
7	90	+ 25	625	86	+ 20	400	500
8	62	− 3	9	58	− 8	64	24
9	65	0	0	53	− 13	169	0
10	39	− 26	676	47	− 19	361	494
$N = 10$	$\Sigma X = 650$		$\Sigma x^2 = 5398$	$\Sigma Y = 660$		$\Sigma y^2 = 2224$	$\Sigma xy = 2704$

(i) Now $\bar{X} = \dfrac{\Sigma X}{N} = \dfrac{650}{10} = 65, \quad \bar{Y} = \dfrac{\Sigma Y}{N} = \dfrac{660}{10} = 66.$

Coefficient of regression of Y on X is,

$$b_{yx} = \frac{\Sigma xy}{\Sigma x^2} = \frac{2704}{5398} = 0{\cdot}5009$$

Coefficient of regression of X on Y is,

$$b_{xy} = \frac{\Sigma xy}{\Sigma y^2} = \frac{2704}{2224} = 1{\cdot}215$$

The equation of the line of regression of Y on X is,

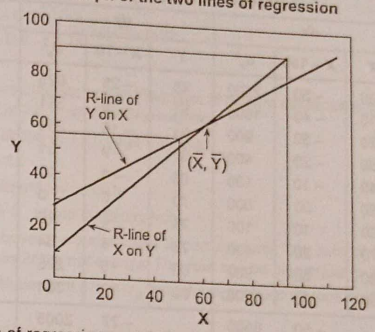
$$Y - \bar{Y} = b_{yx}(X - \bar{X}) \qquad \therefore \quad Y - 66 = 0{\cdot}5\,(X - 65)$$

The equation of the line of regression of X on Y is,

$$X - \bar{X} = b_{xy}(Y - \bar{Y}) \qquad \therefore \quad X - 65 = 1{\cdot}2\,(Y - 66)$$

(ii)

Graph of the two lines of regression



To draw the lines of regression we take two points on each and connect them by a straight line.

From $\quad Y - 66 = 0{\cdot}5\,(X - 65)$

we find that when $\quad X = 65, \quad Y = 66$

and when $\quad X = 0, \quad Y = 27{\cdot}5$

From $\quad X - 65 = 1{\cdot}2\,(Y - 66)$

we find that when $\quad X = 65, \quad Y = 66$

and when $\quad X = 0, \quad Y = 12.$

To estimate Y when $X = 50$ we draw a line at $X = 50$, parallel to y-axis to meet the line of regression of Y on X and read the corresponding value of Y. It is 58. To estimate X when $Y = 90$, we draw a line at $Y = 90$, parallel to x-axis, to meet the line of regression of X on Y and read the corresponding value of X. It is 94.

The two lines intersect at the point for which $X = \bar{X} = 65$ and $Y = \bar{Y} = 66$.

Example 5 : A chemical engineer is investigating the effect of process operating temperature X on product yield Y. The study results in the following data.

X : 100, 110, 120, 130, 140, 150, 160, 170, 180, 190.

Y : 45, 51, 54, 61, 66, 70, 74, 78, 85, 89.

Find the equation of the least square line which will enable to predict yield on the basis of temperature. Find also the degree of relationship between the temperature and the yield.

Also verify that the sum of the coefficients of regression is greater than $2r$.

Sol. :

Calculations of b_{xy}, b_{yx} etc.

Sr. No.	dx			dy			$d_x d_y$
	X	$X-150$	d_x^2	Y	$X-70$	d_y^2	
1	100	– 50	2500	45	– 25	625	1250
2	110	– 40	1600	51	– 19	361	760
3	120	– 30	900	54	– 16	256	480
4	130	– 20	400	61	– 9	81	180
5	140	– 10	100	66	– 4	16	40
6	150	00	000	70	0	0	00
7	160	10	100	74	4	16	40
8	170	20	400	78	8	64	160
9	180	30	900	85	15	225	450
10	190	40	1600	89	19	361	760
$N = 10$		– 50	8500		– 27	2005	4120

$$\bar{X} = A + \Sigma \frac{dx}{N} = 150 - \frac{50}{10} = 145; \qquad \bar{Y} = B + \Sigma \frac{dy}{N} = 70 - \frac{27}{10} = 67{\cdot}3$$

$$b_{yx} = \frac{\Sigma d_x d_y - \dfrac{\Sigma d_x \Sigma d_y}{N}}{\Sigma d_x^2 - \dfrac{(\Sigma d_x)^2}{N}} = \frac{4120 - \dfrac{(-50)(-27)}{10}}{8500 - \dfrac{(-50)^2}{10}} = \frac{4120 - 135}{8500 - 250} = \frac{3985}{8250} = 0{\cdot}483$$

$$b_{xy} = \frac{\Sigma d_x d_y - \dfrac{\Sigma d_x \Sigma d_y}{N}}{\Sigma d_y^2 - \dfrac{(\Sigma d_y)^2}{N}} = \frac{4120 - \dfrac{(-50)(-27)}{10}}{2005 - \dfrac{(-27)^2}{10}} = \frac{4120 - 135}{2005 - 72{\cdot}9} = \frac{3985}{1932{\cdot}1} = 2{\cdot}06$$

The line of regression of Y on X is

$$Y - \bar{Y} = b_{yx}(X - \bar{X})$$

$$\therefore \quad Y - 67{\cdot}3 = 0{\cdot}483\,(X - 145) \qquad \therefore \quad Y = 0{\cdot}483X - 2{\cdot}735$$

The coefficient of correlation

$$r = \sqrt{b_{yx} \times b_{xy}} = \sqrt{0{\cdot}483 \times 2{\cdot}06} = 0{\cdot}9975$$

Now, $b_{xy} + b_{yx} = 2{\cdot}060 + 0{\cdot}483 = 2{\cdot}543$ and $2r = 2 \times 0{\cdot}9975 = 1{\cdot}995$

Hence, we see that $b_{yx} + b_{xy} > 2r$.

Miscellaneous Examples

Example 1 : Find the angle between the lines of regression using the following data.

$n = 10$, $\Sigma x = 270$, $\Sigma y = 630$, $\sigma_x = 4$, $\sigma_y = 5$, $r_{xy} = 0{\cdot}6$.

Sol. : The angle between the lines of regression is given by

$$\tan\theta = \left(\frac{1 - r^2}{r}\right)\left(\frac{\sigma_x \cdot \sigma_y}{\sigma_x^2 + \sigma_y^2}\right)$$

Putting the given values

$$\tan\theta = \left(\frac{1 - 0{\cdot}6^2}{0{\cdot}6}\right)\left(\frac{4 \times 5}{16 + 25}\right) = 0{\cdot}52.$$

Example 2 : Discuss the statement "The sum of the two coefficients of regression is always greater than $2r$ where r is the coefficient of correlation".

Sol. : We have proved above (§ 7, 3, page 2-9) that

$$b_{yx} + b_{xy} \geq 2r.$$

Because this statement leads us to

$$(\sigma_x - \sigma_y)^2 \geq 0 \quad \text{which is always true.} \qquad \text{..... (A)}$$

This means the above given statement is partially true. The sum of the two coefficients of regression is greater than $2r$ but not always. The sum can be also equal to $2r$.

The condition (A) shows that if $\sigma_x = \sigma_y$, then the sign of equality will hold and then $b_{yx} + b_{xy}$ will be equal to $2r$.

This is also clear otherwise. From (1) and (1') (page 2-8), we have

$$b_{yx} + b_{xy} = r\frac{\sigma_y}{\sigma_x} + r\frac{\sigma_x}{\sigma_y}$$

If $\sigma_x = \sigma_y$, we will get $b_{yx} + b_{xy} = 2r$.

Example 3 : Given the following results of weights X and heights Y of 1000 men.

$\bar{X} = 150$ lbs. $\quad \sigma_x = 20$ lbs.

$\bar{Y} = 68$ inches, $\quad \sigma_y = 2{\cdot}5$ inches, $\quad r = 0{\cdot}6$.

where $\bar{X}$ and $\bar{Y}$ are means of X and Y, σ_x and σ_y are standard deviations of X and Y and r is the correlation coefficient between X and Y.

John weighs 200 lbs., Smith is five feet tall. Estimate the height of John and weight of Smith. From the value of height of John estimate his weight. Why is it different from 200 ?

Sol. : With the given notation the line of regression of Y on X is

$$Y - \bar{Y} = r\frac{\sigma_y}{\sigma_x}(X - \bar{X})$$

Substituting the given values,

$$Y - 68 = 0{\cdot}6 \times \frac{2{\cdot}5}{20}(X - 150) = \frac{15}{200}(X - 150).$$

Put $X = 200$,

$$\therefore \quad Y - 68 = \frac{15}{200}(200 - 150) = \frac{15}{4} = 3{\cdot}75$$

$$\therefore \quad Y = 68 + 3{\cdot}75 = 71{\cdot}75 \text{ inches.}$$

Now the line of regression of X of Y is

$$X - \bar{X} = r\frac{\sigma_x}{\sigma_y}(Y - \bar{Y})$$

Substituting the given values,

$$X - 150 = 0{\cdot}6 \times \frac{20}{2{\cdot}5}(Y - 68) = \frac{24}{5}(Y - 68)$$

Put $Y = 5$ feet $= 60$ inches.

$$\therefore \quad X - 150 = \frac{24}{5}(60 - 68) = -\frac{192}{5}$$

$$\therefore \quad X = 150 - \frac{192}{5} = 111{\cdot}6 \text{ lbs.}$$

Hence, height of John = 71·75 inches and weight of Smith = 111·6 lbs.

To estimate the weight of John from his height 71·25 we have to use the equation of line of regression of X on Y (and not Y on X).

i.e., $\quad X - 150 = \frac{24}{5}(Y - 68)$

Putting $Y = 71{\cdot}75$, we get

$$X - 150 = \frac{24}{5}(3{\cdot}75) = 18 \quad \therefore \quad X = 168.$$

The difference is due to the fact that for estimating Y we use one equation and for estimating X we use another equation.

Example 4 : Given $6Y = 5X + 90$, $15X = 8Y + 130$, $\sigma_x^2 = 16$.

Find (i) $\overline{X}$ and $\overline{Y}$, (ii) r and (iii) σ_y^2. **(S.U. 1997, 98, 2002)**

Sol. : (i) To find $\overline{X}$ and $\overline{Y}$: We solve the given equations simultaneously. Multiply the first equation by 3.

$\therefore \quad -15X + 18Y = 270$ and add $15X - 8Y = 130$

$\therefore \quad 10Y = 40 \qquad \therefore \quad \overline{Y} = 40$

Putting this value in any of the given equations.

$$6 \times 40 = 5X + 90 \quad \therefore \quad X = 30 \quad \therefore \quad \overline{X} = 30.$$

(ii) To find r : Suppose the first equation represents, the line of regression of X on Y.

Writing it as $X = \frac{6}{5}Y - 18$, we find $b_{xy} = \frac{6}{5}$.

Suppose the second equation represents the line of regression of Y on X.

Writing it as $Y = \frac{15}{8}X - \frac{130}{8}$, we find $b_{yx} = \frac{15}{8}$.

$$\therefore \quad r = \sqrt{b_{xy} \times b_{yx}} = \sqrt{\frac{6}{5} \times \frac{15}{8}} = \sqrt{\frac{9}{4}} = \sqrt{2{\cdot}25} = 1{\cdot}5.$$

But the value of r can never be greater than 1 numerically. Hence, our supposition is wrong. Now treating the first equation as representing the line of regression of Y on X, we write it as,

$$Y = \frac{5}{6}X + 15 \quad \therefore \quad b_{yx} = \frac{5}{6}.$$

Treating the second equation as representing the line of regression of X on Y, we write it as,

$$X = \frac{8}{15}Y + \frac{130}{15} \quad \therefore \quad b_{xy} = \frac{8}{15}$$

$$\therefore \quad r = \sqrt{b_{xy} \times b_{yx}} = \sqrt{\frac{8}{15} \times \frac{5}{6}} = \sqrt{\frac{4}{9}} = \frac{2}{3} = 0{\cdot}667.$$

(iii) To find σ_y. Consider, $\quad b_{yx} = r\frac{\sigma_y}{\sigma_x} \quad \therefore \quad \frac{5}{6} = \frac{2}{3} \times \frac{\sigma_y}{4} \quad \therefore \sigma_y = 5.$

Example 5 : The equations of the two regression lines are $3x + 2y = 26$ and $6x + y = 31$.

Find : (i) the means of x and y,

(ii) coefficient of correlation between x and y,

(iii) σ_y if $\sigma_x = 3$.

Sol. : (i) To find $\overline{X}$ and $\overline{Y}$

We solve the equations simultaneously. Multiply the second by 2 and subtract from the first.

$$\begin{aligned} \therefore \quad 3x + 2y &= 26 \\ 12x + 2y &= 62 \\ \hline 9x \quad &= 36 \qquad \therefore \quad x = 4. \end{aligned}$$

Putting this value of x in the second equation, we get $24 + y = 31 \qquad \therefore \quad y = 7.$

$\therefore \quad \overline{X} = 4, \quad \overline{Y} = 7.$

(ii) To find r : Suppose the first equation represents the line of regression of X on Y.

Writing it as $3x = -2y + 26$. $\qquad \therefore \quad x = -\frac{2}{3}y + \frac{26}{3}$

$\therefore$ We find that $b_{xy} = -\frac{2}{3}$.

Suppose the second equation represents the line of regression of Y on X.

Writing it as $y = -6x + 31 \qquad \therefore \quad b_{yx} = -6.$

$$\therefore \quad r = \sqrt{b_{xy} \cdot b_{yx}} = \sqrt{(-2/3)(-6)} = \sqrt{4} = 2$$

But the value of r can never be greater than 1. Hence, our supposition is wrong.

Now, treating the first equation as representing the line of regression on of Y on X, we write it as

$$2y = -3x + 26 \quad \therefore \quad y = -\frac{3}{2}x + 13 \quad \therefore \quad b_{yx} = -\frac{3}{2}.$$

Treating the second equation as representing the line of regression of X on Y, we write it as

$$6x = -y + 31 \quad \therefore \quad x = -\frac{1}{6}y + \frac{31}{6} \quad \therefore \quad b_{xy} = -\frac{1}{6}.$$

$$\therefore \quad r = \sqrt{b_{yx} \cdot b_{xy}} = \sqrt{\left(-\frac{3}{2}\right)\left(-\frac{1}{6}\right)} = \sqrt{\frac{1}{4}} = \frac{1}{2} = 0{\cdot}5$$

Since, both b_{xy} and b_{yx} are negative r is negative. $\qquad \therefore \quad r = -0{\cdot}5.$

(iii) To find σ_y : Consider $b_{yx} = r\frac{\sigma_y}{\sigma_x}$.

But $b_{yx} = -\frac{3}{2}$, $r = -\frac{1}{2}$, and $\sigma_x = 3$.

$$\therefore \quad \sigma_y = b_{yx} \cdot \frac{\sigma_x}{r} = \left(-\frac{3}{2}\right) \cdot \frac{3}{(-1/2)} = 9$$

Example 6 : The regression lines of a sample are $x + 6y = 6$, and $3x + 2y = 10$. Find (i) sample means $\bar{x}$ and $\bar{y}$, (ii) coefficient of correlation between x and y. Also estimate y when $x = 12$.

Also verify that the sum of the coefficients of regressions is greater than $2r$.

Sol. : (i) Mean $\bar{x}$ and $\bar{y}$ are obtained by solving the two given equations.

$$3x + 18y = 18 \quad \therefore \quad y = 1/2$$
$$3x + 2y = 10 \quad \therefore \quad x = 3$$
$$16y = 8$$

(ii) If the line $x + 6y = 6$ is the line of regression of y on x, then

$$6y = -x + 6 \quad i.e. \quad y = -\frac{1}{6}x + 1 \quad \therefore \quad b_{yx} = -\frac{1}{6}$$

If the line $3x + 2y = 10$ is the line of regression of x on y, then

$$3x = -2y + 10 \quad i.e. \quad x = -\frac{2}{3}y + \frac{10}{3} \quad \therefore \quad b_{xy} = -\frac{2}{3}$$

$$\therefore \quad r = \sqrt{b_{yx} \times b_{xy}} = \sqrt{\left(-\frac{1}{6}\right) \times \left(-\frac{2}{3}\right)} = \sqrt{\frac{1}{9}} = \frac{1}{3}$$

Since b_{yx} and b_{xy} are both negative, r is negative $\quad \therefore \quad r = -1/3$.

Since $\quad b_{yx} + b_{xy} = \frac{1}{6} + \frac{2}{3} = \frac{5}{6}$ (Numerically)

and $\quad 2r = \frac{2}{3}$, we see that $b_{yx} + b_{xy} > 2r$.

(iii) To estimate y when $x = 12$, we use the line of regression of y on x *i.e.* $y = -\frac{1}{6}x + 1$, when $x = 12$, $y = -2 + 1 = -1$.

Example 7 : If the tangent of the angle made by the line of regression of y on x is 0.6 and $\sigma_y = 2\,\sigma_x$, find the correlation coefficient between x and y.

Sol. : If the equation of the line of regression of y on x is $y - \bar{y} = b_{yx}(x - \bar{x})$ then we know that b_{yx} is the slope of the line of regression. We are thus, given $b_{yx} = 0{\cdot}6$.

But $\quad b_{yx} = r\frac{\sigma_y}{\sigma_x}$ and $\sigma_y = 2\sigma_x$

Putting these value, $\quad 0{\cdot}6 = r \cdot \frac{2\sigma_x}{\sigma_x} = 2r \quad \therefore \quad r = \frac{0{\cdot}6}{2} = 0{\cdot}3.$

Example 8 : If $\sigma_x = \sigma_y = \sigma$ and the angle between the lines of regression is $\tan^{-1} 3$, find the coefficient of correlation.

Sol. : We have

$$\tan\theta = \frac{1 - r^2}{r}\left(\frac{\sigma_x\,\sigma_y}{\sigma_x^2 + \sigma_y^2}\right) \quad \therefore \quad 3 = \frac{1 - r^2}{r} \cdot \left(\frac{\sigma^2}{\sigma^2 + \sigma^2}\right) = \frac{1 - r^2}{2r}$$

$$\therefore \quad \frac{1 - r^2}{r} = 6 \quad \therefore \quad r^2 + 6r - 1 = 0 \quad \therefore \quad r = \frac{-6 \pm \sqrt{36 - 4}}{2} = -3 \pm 2\sqrt{2}$$

Since r cannot be numerically greater than 1, $r = -3 + 2\sqrt{2} = -0{\cdot}17$.

Example 9 : The following data regarding the heights (y) and weights (x) of 100 college students are given

$$\Sigma x = 15000,\ \Sigma x^2 = 2272500,\ \Sigma y = 6800,\ \Sigma y^2 = 463025,\ \Sigma xy = 1022250.$$

Find the coefficient of correlation between height and weight and also the equation of regression of height and weight.

Sol. : The coefficients of regression are given by

$$b_{yx} = \frac{\Sigma xy - \frac{\Sigma x \cdot \Sigma y}{N}}{\Sigma x^2 - \frac{(\Sigma x)^2}{N}} = \frac{1022250 - \frac{15000 \times 6800}{100}}{2272500 - \frac{15000^2}{100}} = \frac{2250}{22500} = 0{\cdot}1.$$

$$b_{xy} = \frac{\Sigma xy - \frac{\Sigma x \cdot \Sigma y}{N}}{\Sigma y^2 - \frac{(\Sigma y)^2}{N}} = \frac{1022250 - \frac{15000 \times 6800}{100}}{463025 - \frac{6800^2}{100}} = \frac{2250}{625} = 3{\cdot}6.$$

$$\therefore \quad r = \sqrt{b_{yx} \times b_{xy}} = \sqrt{0{\cdot}1 \times 3{\cdot}6} = 0{\cdot}6.$$

The equation of the lines of regression of y on x is

$$y - \bar{y} = b_{yx}(x - \bar{x}) \quad \therefore \quad y - 68 = 0{\cdot}1(x - 1500)$$
$$\therefore \quad y = 0{\cdot}1x - 82.$$

Example 10 : It is given that the means of x and y are 5 and 10. If the line of regression of y on x is parallel to the line $20y = 9x + 40$, estimate the value of y for $x = 30$.

Sol. : The line of regression of y on x is $y - \bar{y} = b_{yx}(x - \bar{x})$.

Its slope is b_{yx}. But this line is parallel to $20y = 9x + 40$

i.e. $\quad y = \frac{9}{20}x + 2$ whose slope is $\frac{9}{20}$. $\quad \therefore \quad b_{yx} = \frac{9}{20}$.

But by data $\bar{x} = 5$ and $\bar{y} = 10$. Hence, the equation of the line of regression of y on x is

$$y - 10 = \frac{9}{20}(x - 5) \quad i.e. \quad y = \frac{9}{20}x + \frac{155}{20}$$

When $x = 30$, $\quad y = \frac{270}{20} + \frac{155}{20} = \frac{425}{20} = 21{\cdot}25.$

EXERCISE - I

Type I

State true or false with proper reasoning.

1. If $r = 0$, the lines of regression are parallel to each other. **[Ans. : False]**
2. The values of r and R can never be equal. **[Ans. : False]**
3. In a regression analysis it was found that $b_{yx} = 0{\cdot}87$, $b_{xy} = 1{\cdot}55$. These values are not consistent. **[Ans. : True]**
4. The two regression coefficients are both positive or both negative. **[Ans. : True]**
5. $3x + y = 5$ and $2x - 3y = 7$ cannot be lines of regression for any set of values of x and y. **[Ans. : True]**

Type II

1. The following table gives the age of car of a certain make and annual maintenance cost. Obtain the equation of the line of regression of cost on age.

Age of a car :	2	4	6	8
Maintenance :	1	2	2·5	3

[**Ans. :** $y = 0{\cdot}325x + 0{\cdot}5$]

2. Obtain the equation of the line of regression of y on x from the following data and estimate y for $x = 73$.

x :	70,	72,	74,	76,	78,	80.
y :	163,	170,	179,	188,	196,	220.

[**Ans. :** $y = 5{\cdot}31x - 212.57$; $y = 175{\cdot}37$]

3. The heights in cms of fathers (x) and of the eldest sons (y) are given below.

x :	165	160	170	163	173	158	178	168	173	170	175	180
y :	173	168	173	165	175	168	173	165	180	170	173	178

Estimate the height of the eldest son if the height of the father is 172 cms. and the height of the father if the height of the eldest son is 173 cm.

Also find the coefficient of correlation between the heights of fathers and sons.

[**Ans. : (i)** $y = 1{\cdot}016x - 5{\cdot}123$, **(ii)** $x = 0{\cdot}476y + 98{\cdot}98$, **(iii)** 169·97, 173·45, **(iv)** $r = 0{\cdot}696$]

4. Find (i) the lines of regression, (ii) coefficient of correlation for the following data.

x :	65	66	67	67	68	69	70	72
y :	67	68	65	66	72	72	69	71

[**Ans. :** $y = 19{\cdot}79 + 0{\cdot}72x$, **(ii)** $x = 33{\cdot}29 + 0.5y$; $r = 0{\cdot}6$]

5. Find the line of regression for the following data and estimate y corresponding to $x = 15{\cdot}5$.

x :	10	12	13	16	17	20	25
y :	19	22	24	27	29	33	37

[**Ans. :** $y = 0{\cdot}8x + 13{\cdot}23$; 25·63]

Type III

1. Given

	x series	y series
Mean	18	100
S.D.	14	20

$r = 0{\cdot}8$.

Find the most probable value of y when $x = 70$ and most probable value of x when $y = 90$.

[**Ans. :** $y = 159{\cdot}3$, $x = 12{\cdot}4$]

2. Given the following information about marks of 60 students.

	Mathematics	English
Mean	80	50
S.D.	15	10

Coefficient of correlation $r = 0{\cdot}4$. Estimate the marks of the student in mathematics who scored 60 marks in English. [**Ans. :** 86]

3. You are supplied with the following information. The equation of the lines regression are $2x + 3y + 8 = 0$ and $x + 2y - 5 = 0$.

Find the means of x and y and the coefficient of correlation between them.

[**Ans. :** $\bar{x} = -31$, $\bar{y} = 18$, $r = -0{\cdot}87$]

4. From 8 observations the following results were obtained :

$\Sigma x = 59$, $\Sigma y = 40$, $\Sigma x^2 = 524$, $\Sigma y^2 = 256$, $\Sigma xy = 364$.

Find the equation of the line of regression of x on y and the coefficient of correlation.

[**Ans. :** $x = 1{\cdot}5y - 0{\cdot}5$, $r = 0{\cdot}98$]

5. The equations of the two lines of regression are $x = 19{\cdot}13 - 0{\cdot}87y$ and $y = 11{\cdot}64 - 0{\cdot}50x$. Find (i) the means of x and y, (ii) the coefficient of correlation between x and y.

[**Ans. : (i)** $\bar{x} = 15{\cdot}79$, $\bar{y} = 3{\cdot}74$; **(ii)** $r = -0{\cdot}66$, $b_{yx} = -0{\cdot}50$, $b_{xy} = -0{\cdot}87$]

6. Out of the two equations given below which can be a line of regression of x on y and why ?

$x + 2y - 6 = 0$ and $2x + 3y - 8 = 0$.

[**Ans. :** $2x + 3y - 8 = 0$]

7. In a partially destroyed laboratory record of analysis of correlation data the following results are legible. Variance of $x = 9$, equations of the lines of regression

$4x - 5y + 33 = 0$, $20x - 9y - 107 = 0$.

Find (i) the mean values of x and y, (ii) the standard deviation of y, (iii) coefficient of correlation between x and y.

[**Ans. : (i)** $\bar{x} = 13$, $\bar{y} = 17$, **(ii)** $\sigma_y = 4$, **(iii)** $r = 0{\cdot}6$]

8. Given : var (x) = 25. The equations of the two lines of regression are $5x - y = 22$ and $64x - 45y = 24$.

Find (i) $\bar{x}$ and $\bar{y}$, (ii) r, (iii) σ_y. [**Ans. : (i)** $\bar{x} = 6$, $\bar{y} = 8$, **(ii)** $r = 1{\cdot}87$, **(iii)** $\sigma_y = 1/5$]

9. Find the regression coefficients and the coefficient of correlation from the following data where x, y denote the actual values.

$N = 12$, $\Sigma x = 120$, $\Sigma y = 432$, $\Sigma xy = 4992$, $\Sigma x^2 = 1392$, $\Sigma y^2 = 18252$.

[**Ans. :** $b_{yx} = 3{\cdot}5$, $b_{xy} = 0{\cdot}249$, $r = 0{\cdot}93$]

EXERCISE - II

Theory

1. Distinguish between correlation and regression.
2. Explain "the line of regression". Why there are two lines of regression ?
3. Explain what you understand by regression. What are lines of regression ? Why are there in general two lines of regression ? When do they coincide, when are they perpendicular ?
4. Explain the method of scatter diagram to obtain a line of regression.
5. Obtain the equations of lines of regression.
6. Prove that the sum of the coefficients of regression is greater than or equal to $2r$ where r is the coefficient of correlation. [See 3, page 2-9]
7. Find the expression for the acute angle between the lines of regression.
8. With usual notation prove that

   (i) $r = \sqrt{b_{yx} \cdot b_{xy}}$, **(ii)** $b_{xy} + b_{yx} \geq 2r$.

9. Examine whether the following statement is correct

   $b_{xy} = 3{\cdot}2$, $b_{yx} = 0{\cdot}7$. [**Ans. :** No]

10. If θ is the angle between the two lines of regression, prove that

$$\tan\theta = \left(\frac{1-r^2}{r}\right)\left(\frac{\sigma_x - \sigma_y}{\sigma_x^2 + \sigma_y^2}\right)$$

☆☆☆

CHAPTER 3

Curve Fitting

1. Introduction

In many social, economical, engineering and physical problems we have a set of values of x and y although we do not know the functional relationship between them. Fitting of a curve to a given set of values means finding a functional relationship between x and y whose curve is the closest possible curve to the given values. The curve so obtained does not pass through all the given points but is close to them to the maximum extent. Finding such a curve for a given set of values is called **Curve Fitting**. The relation in general is assumed to be a linear function $y = a + bx$ or a parabolic function $y = a + bx + cx^2$ or even exponential or logarithmic. The method used for fitting the curve is based on the **principle of least squares**.

2. Fitting a Straight Line by the Method of Least Squares

Suppose we have a set of values (x_i, y_i). Suppose further that we want to fit a straight line $y = a + bx$ to these values. The straight line must be close to the given points to the maximum extent. **The principle of least square states that the straight line should be such that the distances of the given points from the straight line measured along the y-axis must be minimum.** The line obtained in this way is called **the line of best fit.**

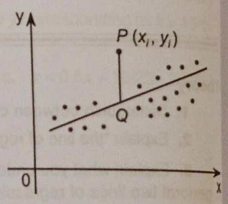


Fig. 3.1

Suppose $P(x_i, y_i)$ is a given point and suppose the equation of the line of best fit be $y = a + bx$. Suppose further that the line parallel to the y-axis through P intersects the line in Q. Now, the coordinates of Q are $(x_i, a + bx_i)$. We want to find a, b such that the distance $|PQ|$ is minimum. But the distance $|PQ|$ is minimum when the square of the distance $(y_i - a - bx_i)^2$ is minimum. This must be true for all points. This means we should have $S \equiv \Sigma (y_i - a - bx_i)^2$ minimum. Since we find a, b such that the sum of the squares S is least, the method is known as the method of least squares.

Now, for S to be minimum the conditions are

$$\frac{\partial S}{\partial a} = 0 \quad \text{and} \quad \frac{\partial s}{\partial b} = 0$$

$$\therefore \quad \Sigma (y_i - a - bx_i) = 0 \quad \text{and} \quad \Sigma (y_i - a - bx_i)\, x_i = 0 \qquad \therefore \quad \Sigma y_i = \Sigma a - b \Sigma x_i$$

$$\therefore \quad \Sigma y_i = Na + b \Sigma x_i \quad \text{and} \quad \Sigma y_i x_i = a \Sigma x_i + b \Sigma x_i^2$$

We shall drop the suffix i and write the equations as

$$\boxed{\begin{aligned} \Sigma y &= Na + b \Sigma x \\ \Sigma xy &= a \Sigma x + b \Sigma x^2 \end{aligned}}$$

These equations are known as **Normal Equations.**

Methods of finding the constants a

1. Direct Method : If we find Σx, Σy, etc. and substitute their values in the above equations, we get two equations in two unknowns a and b. Solving these two equations we can find the constants a and b.

2. Assumed Mean Method : We may assume a certain mean of x and of y, *i.e.*, we may put $X = x - m$ and $Y = y - n$ and find ΣX, ΣY, etc. put these values in the above two equations.

If necessary, we may divide the deviations by a common factor, if there is any, *i.e.*, we may put $X = \dfrac{x - m}{h}$ and $Y = \dfrac{y - n}{k}$.

Solving these equations, we get a and b and resubstituting the values of X and Y, we get the equation of the line.

3. Actual Mean Method : If actual means of x and y are round figures, we put $X = x - \bar{x}$ and $Y = y - \bar{y}$ $\left(\text{or } X = \dfrac{x - \bar{x}}{h},\ Y = \dfrac{y - \bar{y}}{k}\right)$, find ΣX, ΣY, etc. and put these values in the above two equations.

Solving these equations as above and resubstituting the values of X and Y, we get the equation of the line.

Of these three methods we shall see below that the last method is the simplest one.

Example 1 : Fit a straight line to the following data :

$$(x, y) : (1, 1),\ (2, 5),\ (3, 11),\ (4, 8),\ (5, 14)$$

Sol. : We shall obtain the line from the above normal equations.

Calculations of Σx, Σx^2 etc.

x	y	x^2	xy
1	1	1	1
2	5	4	10
3	11	9	33
4	8	16	32
5	14	25	70
$N = 5$, 15	39	55	146

The normal equations are

$$\Sigma y = Na + b \Sigma x \qquad \therefore \quad 39 = 5a + 15b \qquad \text{..... (1)}$$

$$\text{and} \quad \Sigma xy = a \Sigma x + b \Sigma x^2 \qquad \therefore \quad 146 = 15a + 55b \qquad \text{..... (2)}$$

Multiply (1) by 3,

$$\therefore \quad 117 = 15a + 45b \quad \text{and} \quad 146 = 15a + 55b$$

By subtraction $\quad 10b = 29 \qquad \therefore \quad b = 2{\cdot}9.$

Putting this value of b in (1),

$$5a = 39 - 15 \times 2{\cdot}9 = -4{\cdot}5$$

$$\therefore \quad a = -\frac{4{\cdot}5}{5} = -0{\cdot}9.$$

Hence, the equation of the line is $\quad y = -0{\cdot}9 + 2{\cdot}9x.$

Example 2 : Fit a straight line $y = a + bx$ to the following data.

x :	0	5	10	15	20	25
y :	12	15	17	22	24	30

Sol. : (1) Direct Method

Calculations of Σx, Σx^2 etc.

x	y	x^2	xy
0	12	0	0
5	15	25	75
10	17	100	170
15	22	225	330
20	24	400	480
25	30	625	750
$N = 6, \Sigma x = 75$	$\Sigma y = 120$	$\Sigma x^2 = 1375$	$\Sigma xy = 1805$

Let the equation be $y = a + bx$.

The normal equations are

$\Sigma y = Na + b\Sigma x$ $\quad\therefore\ 120 = 6a + 75b$ ..... (1)

and $\Sigma xy = a\Sigma x + b\Sigma x^2$ $\quad\therefore\ 1805 = 75a + 1375b$ ..... (2)

Now, to find a, b multiply (1) by 50 and (2) by 4 and subtract.

$$\begin{aligned} 6000 &= 300\,a + 3750\,b \\ 7220 &= 300\,a + 5500\,b \\ \hline 1220 &= \qquad\quad 1750\,b \end{aligned}$$

$\therefore\ b = \dfrac{1220}{1750} = 0{\cdot}6971 = 0{\cdot}7$

Putting this value of b in (1), we get

$120 = 6a + 75\,(0{\cdot}7)$ $\quad\therefore\ 6a = 120 - 52{\cdot}5 = 67{\cdot}5$ $\quad\therefore\ a = 11{\cdot}25.$

$\therefore$ The equation of the line is $y = 11{\cdot}25 + 0{\cdot}7x$.

(2) Assumed Mean Method : We shall assume 10 and 17 as means of x and y, *i.e.*, we shall put $X = x - 10$ and $Y = y - 17$.

Calculations of ΣX, ΣX^2 etc.

x	y	$X = x - 10$	$Y = y - 17$	X^2	XY
0	12	− 10	− 5	100	50
5	15	− 5	− 2	25	10
10	17	0	0	0	0
15	22	5	5	25	25
20	24	10	7	100	70
25	30	15	13	225	195
$N = 6$		15	18	475	350

Let the equation be $Y = a + bX$.

The normal equations are

$\Sigma Y = Na + b\Sigma X$ $\quad\therefore\ 18 = 6a + 15b$ ..... (1)

and $\Sigma XY = a\Sigma X + b\Sigma X^2$ $\quad\therefore\ 350 = 15a + 475$ ..... (2)

Multiply (1) by 5 and (2) by 2 and subtract.

$$\begin{aligned} \therefore\quad 90 &= 30\,a + 75\,b \\ 700 &= 30\,a + 950\,b \\ \hline 610 &= \qquad\quad 875\,b \end{aligned}$$

$\therefore\ b = \dfrac{610}{875} = 0{\cdot}7$

Putting this value of b in (1), we get

$18 = 6a + 15\,(0{\cdot}7)$ $\quad\therefore\ 6a = 18 - 10{\cdot}5$ $\quad\therefore\ a = 1{\cdot}25.$

$\therefore$ The equation is $Y = 1{\cdot}25 + 0{\cdot}7X$.

Resubstituting $Y = y - 17$ and $X = x - 10$, we get

$y - 17 = 1{\cdot}25 + 0{\cdot}7\,(x - 10)$ $\quad\therefore\ y = 0{\cdot}7x + 11{\cdot}25.$

Example 3 : Fit a straight line $y = a + bx$ to the following data.

x :	1	2	3	4	5	6
y :	49	54	60	73	80	86

Sol. : We shall solve this example by all the three methods.

(a) Direct method :

Calculations of Σx, Σx^2 etc.

x	y	x^2	xy
1	49	1	49
2	54	4	108
3	60	9	180
4	73	16	292
5	80	25	400
6	86	36	516
$N = 6$, 21	402	91	1545

Now, $y = a + bx$

$\Sigma y = Na + b\Sigma x$ $\quad\therefore\ 402 = 6a + 21b$ ..... (1)

and $\Sigma xy = a\Sigma x + b\Sigma x^2$ $\quad\therefore\ 1545 = 21a + 91b$ ..... (2)

Multiply (1) by 7 and (2) by 2 and subtract

$$\begin{aligned} \therefore\quad 2814 &= 42\,a + 147\,b \\ 3090 &= 42\,a + 182\,b \\ \hline 276 &= \qquad\quad 35\,b \end{aligned}$$

$\therefore\ b = 7{\cdot}89$

Putting this value of b in (1), we get

$402 = 6a + 21\,(7{\cdot}89)$ $\quad\therefore\ 6a = 402 - 165{\cdot}69 = 236{\cdot}31$ $\quad\therefore\ a = 39{\cdot}38$

$\therefore\ y = 39{\cdot}38 + 7{\cdot}89x.$

(b) Assumed Mean Method : We shall assume 3 and 60 as means of x and y and find the deviations.

Calculations of ΣX, ΣX^2 etc.

x	y	$X = x - 3$	$Y = y - 60$	X^2	XY
1	49	– 2	– 11	4	22
2	54	– 1	– 6	1	6
3	60	0	0	0	0
4	73	1	13	1	13
5	80	2	20	4	40
6	86	3	26	9	78
$N = 6$		3	42	19	159

Now, $Y = a + bX$

$\Sigma Y = Na + b\Sigma X$ $\quad\therefore\ 42 = 6a + 3b$ ..... (1)

and $\Sigma XY = a\Sigma X + b\Sigma X^2$ $\quad\therefore\ 159 = 3a + 19b$ ..... (2)

Multiply (2) by 2 and subtract (1) from the result.

$$\begin{aligned} \therefore\ 318 &= 6a + 38b \\ 42 &= 6a + 3b \\ \hline 276 &= 35b \end{aligned}$$

$\therefore\ b = 7{\cdot}89$

Putting this value of b in (1),

$42 = 6a + 3(7{\cdot}89)$ $\quad\therefore\ 6a = 18{\cdot}33$ $\quad\therefore\ a = 3{\cdot}05$

$\therefore\ Y = 3.05 + 7.89\,X.$

Putting $X = x - 3$, $Y = y - 60$, we get

$y - 60 = 3{\cdot}05 + 7{\cdot}89\,(x - 3)$ $\quad\therefore\ y = 39{\cdot}38 + 7{\cdot}89\,x.$

(c) Actual Mean method : We shall take deviations from actual means of x and y.

Now, $\bar{x} = \dfrac{21}{6} = 3{\cdot}5$, $\bar{y} = \dfrac{402}{6} = 67.$

Calculations of ΣX, ΣX^2 etc.

x	y	$X = x - 3{\cdot}5$	$Y = y - 67$	X^2	XY
1	49	– 2·5	– 18	6·25	45·0
2	54	– 1·5	– 13	2·25	19·5
3	60	– 0·5	– 7	0·25	3·5
4	73	0·5	6	0·25	3·0
5	80	1·5	13	2·25	19·5
6	86	2·5	19	6·25	47·5
$N = 6$		0	0	17·5	138·0

Now, $Y = a + bX$

$\therefore\ \Sigma Y = Na + b\Sigma X$ and $\Sigma XY = a\Sigma X + b\Sigma X^2$

Putting the above values

$0 = 6a + b\,(0)$ $\quad\therefore\ a = 0$

and $138 = a\,(0) + 17{\cdot}5\,b$ $\quad\therefore\ b = 7{\cdot}89$

Hence, the equation is $Y = 7{\cdot}89X.$

Putting the values of X, Y, we get

$y - 67 = 7{\cdot}89\,(x - 3{\cdot}5)$ $\quad\therefore\ y = 39{\cdot}38 + 7{\cdot}89x.$

Note ....

Thus, we see that if we take deviations from $\bar{x}$ and $\bar{y}$, we get $a = 0$ and we get the value of b directly. In this way the last method is more convenient.

Example 4 : Fit a straight line to the following data.

Year x :	1951	1961	1971	1981	1991
Production y (000 tons) :	10	12	8	10	15

Also estimate the production in 1987.

Sol. : Now the means of x and y are $\bar{x} = 1971$ and $\bar{y} = 11$.

Since $\bar{x}$ and $\bar{y}$ are round figures, we shall take deviations from 1971 and 11.

Calculations of ΣX, ΣY etc.

x	y	$X = x - 1971$	$Y = y - 11$	X^2	XY
1951	10	– 20	– 1	400	20
1961	12	– 10	1	100	– 10
1971	8	0	– 3	0	0
1981	10	10	– 1	100	– 10
1991	15	20	4	400	80
$N = 5$		0	0	1000	80

Now, $Y = a + bX$

$\therefore\ \Sigma Y = Na + b\Sigma X$ and $\Sigma XY = a\Sigma X + b\Sigma X^2$

Putting the above values

$0 = 5a + 0$ $\quad\therefore\ a = 0$

and $80 = 0 + 1000\,b$ $\quad\therefore\ b = 0{\cdot}08$

Hence, the equation is $Y = 0{\cdot}08X.$

Putting the values of X, Y, we get

$y - 11 = 0.08\,(x - 1971)$ $\quad\therefore\ y = -146{\cdot}68 + 0{\cdot}08x.$

Putting $x = 1987$, $\quad y = -146{\cdot}68 + 0{\cdot}08\,(1987) = 12{\cdot}28.$

$\therefore$ The production in 1987 is 12·28.

Example 5 : Fit a straight line of the form $y = a + bx$ to the following data and estimate the value of y for $x = 3{\cdot}5$.

x :	0	1	2	3	4
y :	1	1·8	3·3	4·5	6·3

Sol. : We shall solve this problem by both the methods *i.e.*, by direct method and also by deviation method.

(a) Direct method :

Calculations of Σx, Σx^2 etc.

x	y	x^2	xy
0	1·0	0	0·0
1	1·8	1	1·8
2	3·3	4	6·6
3	4·5	9	13·5
4	6·3	16	25·2
$N = 5$, 10	16·9	30	47·1

Let the equation of the line be $y = a + bx$.

Then the normal equations are

$\Sigma y = Na + b\Sigma x$ $\quad\therefore\ 16{\cdot}9 = 5a + 10b$ ..... (1)

and $\Sigma xy = a\Sigma x + b\Sigma x^2$ $\quad\therefore\ 47{\cdot}1 = 10a + 30b$ ..... (2)

Now, multiply (1) by 2 and subtract the result from (2).

$\therefore\ 47{\cdot}1 = 10a + 30b$

$33{\cdot}8 = 10a + 20b$

$13{\cdot}3 = 10b$ $\quad\therefore\ b = 1{\cdot}33$

Putting this value of b in (1), we get

$16{\cdot}9 = 5a + 13{\cdot}3$ $\quad\therefore\ 5a = 3{\cdot}6$ $\quad\therefore\ a = 0{\cdot}72$

Hence, the equation of the line is

$y = 0{\cdot}72 + 1{\cdot}33x$.

(b) Deviation method : We shall take deviations of x from the mid-point of the values of x.

x	$X = x - 2$	$Y = y$	X^2	XY
0	– 2	1.0	4	– 2·0
1	– 1	1.8	1	– 1·8
2	0	3.3	0	0·0
3	1	4.5	1	4·5
4	2	6.3	4	12·6
$N = 5$	0	16.9	10	13·3

Let the equation of the line be $Y = a + bX$.

The normal equations are

$\Sigma Y = Na + b\Sigma X$ and $\Sigma XY = a\Sigma X + b\Sigma X^2$

Putting the values from the table, we get,

$16{\cdot}9 = 5a + b(0)$ $\quad\therefore\ a = \dfrac{16{\cdot}9}{5} = 3{\cdot}38.$

And $13{\cdot}3 = a(0) + 10b$ $\quad\therefore\ b = 1{\cdot}33.$

Hence, the equation of the line is $Y = 3{\cdot}38 + 1{\cdot}33X$.

Putting $X = x - 2$ and $Y = y$ we get

$y = 3{\cdot}38 + 1{\cdot}33(x - 2) = 3{\cdot}38 - 2{\cdot}66 + 1{\cdot}33x$

$\therefore\ y = 0{\cdot}72 + 1{\cdot}33x$

Thus, the equation of the line is $y = 0{\cdot}72 + 1{\cdot}33x$.

Putting $x = 3{\cdot}5$, $y = 0{\cdot}72 + 4{\cdot}655 = 5{\cdot}375$.

Hence, when $x = 3{\cdot}5$, $y = 5{\cdot}375$.

Example 6 : Fit a straight line to the following data, with x as independent variable.

x :	1965	1966	1967	1968	1969
y :	125	140	165	195	200

Sol. : The means of x and y are $\bar{x} = 1967$, $\bar{y} = 165$.

Since $\bar{x}$ and $\bar{y}$ are round figures, we shall take deviations from 1967 and 165.

Calculations of ΣX, ΣY etc.

x	y	$X = x - 1967$	$Y = y - 165$	X^2	XY
1965	125	– 2	– 40	4	80
1966	140	– 1	– 25	1	25
1967	165	0	0	0	0
1968	195	1	30	1	1
1969	200	2	35	4	70
$N = 5$		0	0	10	176

Now, $Y = a + bX$

$\therefore\ \Sigma Y = Na + b\Sigma X$ and $\Sigma XY = a\Sigma X + b\Sigma X^2$

Putting the above values

$0 = 5a + 0$ $\quad\therefore\ a = 0$

and $176 = 0 + 10b$ $\quad\therefore\ b = 17{\cdot}6$

Hence, the equation of line is $Y = 17{\cdot}6X$.

Putting the values of X and Y, we get

$y - 165 = 17{\cdot}6(x - 1967)$

$= 165 + 17{\cdot}6x - 34619{\cdot}2$

$\therefore\ y = 17.6x - 34454{\cdot}2.$

Example 7 : Fit a straight line to the following data.

x :	1	2	3	4	5	6
y :	83	92	71	90	160	191

Sol. : Now, the mean of $x = 3{\cdot}5$. Since the deviations will be in decimals, we multiply them by 2, *i.e.*, we put $2(x - 35) = X$.

The mean of $y = 114{\cdot}5$.

Since, the deviations will be in decimals, we multiply them by 2, *i.e.*, we put $2(y - 114{\cdot}5) = Y$.

Calculations of ΣX, ΣY etc.

x	y	$X = 2(x - 3{\cdot}5)$	$Y = 2(y - 114.5)$	X^2	XY
1	83	– 5	– 63	25	315
2	92	– 3	– 45	9	135
3	71	– 1	– 87	1	87
4	90	1	– 49	1	– 49
5	160	3	91	9	273
6	191	5	153	25	765
$N = 6$		0	0	70	1526

Now, $Y = a + bX$ and $\Sigma XY = a\Sigma X + b\Sigma X^2$

$\therefore \Sigma Y = Na + b\Sigma X$

Putting the above values

$0 = 6a + 0 \quad \therefore a = 0$

and $1526 = 0 + 70\,b \quad \therefore b = 21{\cdot}8$

Hence, the equation of line is $Y = 21{\cdot}8X$.

Putting the values of X and Y, we get

$2(y - 114{\cdot}5) = 21{\cdot}8\,(2)\,(x - 3{\cdot}5)$

$\therefore y - 114{\cdot}5 = 21{\cdot}8\,(x - 3{\cdot}5) \quad \therefore y = 38{\cdot}2 + 21{\cdot}8x.$

Note ....

We shall consider fitting of a second degree curve and an exponential curve on page 3-10 and 3-16.

EXERCISE - I

Class (a) : 4 Marks

1. Fit a straight line to the following data :

x :	0	1	2	3	4	5
y :	1	2	3	4·5	6	7·5

[**Ans. :** $y = 0{\cdot}70 + 1{\cdot}32x$]

2. Fit a straight line to the following data :

x :	100	120	140	160	180	200
y :	0·45	0·55	0·60	0·70	0·80	0·85

[**Ans. :** $y = 0{\cdot}041 + 0{\cdot}0041x$]

3. Fit a first degree curve to the following data and estimate the value of y when $x = 73$.

x :	10	20	30	40	50	60	70	80
y :	1	3	5	10	6	4	2	1

[**Ans. :** $y = 4 - 0{\cdot}071u$ where $u = (x - 45)/5$, $y = 3{\cdot}596$, when $x = 73$]

4. Fit a straight line to the following data and estimate y when $x = 12$

x :	1	2	3	4	5	6	7	8	9	10
y :	52·5	58·7	65·0	70·2	75·4	81·1	87·2	95·5	102·2	108·4

[**Ans. :** $Y = 79{\cdot}62 + 3{\cdot}08X$ where $X = (x - 5{\cdot}5)2$, when $x = 12$, $y = 119{\cdot}66$]

5. Fit a straight line to the following data.

$(x, y) = (-1, -5), (1, 1), (2, 4), (3, 7), (4, 10)$. Estimate y when $x = 7$.

[**Ans. :** $y = -2 + 3x$; 19]

3. Fitting A Parabola

Suppose we have a set of values (x_i, y_i). Suppose further that we want to fit a parabola $y = a + bx + cx^2$ to these values. The parabola must be close to the given points as much as possible. The principle of least squares states that the parabola should be such that the distances of the given points from the parabola measured along the y axis must be minimum.

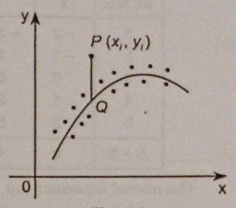


Fig. 3.2

Suppose $P(x_i, y_i)$ is a given point and a line through P parallel to the y axis intersects the curve $y = a + bx + cx^2$ in Q. Then Q is $(x_i, a + bx_i + cx_i^2)$. We find a, b, c such that the distance $|PQ|$ is minimum. But distance $|PQ|$ is minimum when the square of the distance $(y_i - a - bx_i - cx_i^2)^2$ is minimum. This must be true for all points. This means we should have $s = \Sigma(y_i - a - bx_i - cx_i^2)^2$ minimum. Since we find a, b, c such that the sum of the squares S is least the method as you know is knowm as the **method of least squares.**

Now for s to be minimum, the conditions are

$$\frac{\partial s}{\partial a} = 0, \quad \frac{\partial s}{\partial b} = 0, \quad \frac{\partial s}{\partial c} = 0$$

Differentiating s partially w.r.t. a, b, c, we get

$$\therefore \Sigma(y_i - a - bx_i - cx_i^2) = 0 ; \qquad \Sigma(y_i - a - bx_i - cx_i^2)\,x_i = 0 ;$$

$$\Sigma(y_i - a - bx_i - cx_i^2)\,x_i^2 = 0.$$

i.e., $\Sigma y_i = Na + b\Sigma x_i + c\Sigma x_i^2 ; \qquad \Sigma x_i y_i = a\Sigma x_i + b\Sigma x_i^2 + c\Sigma x_i^3 ;$

$$\Sigma x_i^2 y_i = a\Sigma x_i^2 + b\Sigma x_i^3 + c\Sigma x_i^4.$$

where N is the number of observations.

Dropping the suffix, i, we get the equations as

$$\boxed{\begin{aligned} \Sigma y &= Na + b\Sigma x + c\Sigma x^2 \\ \Sigma xy &= a\Sigma x + b\Sigma x^2 + c\Sigma x^3 \\ \Sigma x^2 y &= a\Sigma x^2 + b\Sigma x^3 + c\Sigma x^4 \end{aligned}}$$

These equations are called **normal equations.**

As in the case of straight line for fitting of parabola also we have two methods. (i) Direct Method and (ii) Deviation Method (or Change of Origin Method).

In direct method we use the given values of x and y and find the constants a, b, c.

If the values of x (or y) are evenly spaced we change the origin to the mid-point of x and use the deviations as in Ex. 3 or 4 below.

Example 1 : By the method of least squares find the best values of a, b, c in the law $y = a + bx + cx^2$ to fit the following data.

x :	-2	-1	0	1	2
y :	$-3{\cdot}150$	$-1{\cdot}390$	$0{\cdot}620$	$2{\cdot}880$	$5{\cdot}378$

Sol. :

Calculations of $\Sigma x, \Sigma y, \Sigma x^2$

Sr. No.	x	y	x^2	x^3	x^4	xy	x^2y
1	-2	$-3{\cdot}150$	4	-8	16	$6{\cdot}300$	$-12{\cdot}600$
2	-1	$-1{\cdot}390$	1	-1	1	$1{\cdot}390$	$-1{\cdot}390$
3	0	$0{\cdot}620$	0	0	0	$0{\cdot}000$	$0{\cdot}000$
4	1	$2{\cdot}880$	1	1	1	$2{\cdot}880$	$2{\cdot}880$
5	2	$5{\cdot}378$	4	8	16	$10{\cdot}756$	$21{\cdot}512$
$N = 5$	0	$4{\cdot}338$	10	0	34	$21{\cdot}326$	$10{\cdot}402$

The normal equations are

$$\Sigma y = Na + b\Sigma x + c\Sigma x^2 \qquad \therefore \quad 4{\cdot}338 = 5a + 10c$$

$$\Sigma xy = a\Sigma x + b\Sigma x^2 + c\Sigma x^3 \qquad \therefore \quad 21{\cdot}325 = 10b$$

$$\Sigma x^2y = a\Sigma x^2 + b\Sigma x^3 + c\Sigma x^4 \qquad \therefore \quad 10{\cdot}402 = 10a + 34c$$

Solving these three equations, we get

$$a = 0{\cdot}621, \quad b = 2{\cdot}1326, \quad c = 0{\cdot}1233$$

Hence, the law is

$$y = 0{\cdot}621 + 2{\cdot}1326x + 0{\cdot}1233x^2.$$

Example 2 : Fit a second degree curve *i.e.*, parabola to the following data.

x :	0	1	2	3	4
y :	$1{\cdot}0$	$1{\cdot}8$	$1{\cdot}3$	$2{\cdot}5$	$6{\cdot}3$

Sol. : Let the equation of the parabola be $y = a + bx + cx^2$.

Calculations of $\Sigma x, \Sigma x^2$, etc.

Sr. No.	x	y	x^2	x^3	x^4	xy	x^2y
1	0	$1{\cdot}0$	0	0	0	0	0
2	1	$1{\cdot}8$	1	1	1	$1{\cdot}8$	$1{\cdot}8$
3	2	$1{\cdot}3$	4	8	16	$2{\cdot}6$	$5{\cdot}2$
4	3	$2{\cdot}5$	9	27	81	$7{\cdot}5$	$22{\cdot}5$
5	4	$6{\cdot}3$	16	64	256	$25{\cdot}2$	$100{\cdot}8$
$N = 5$	10	$12{\cdot}9$	30	100	354	$37{\cdot}1$	$130{\cdot}3$

The normal equations are

$$\Sigma y = Na + b\Sigma x + c\Sigma x^2$$

$$\Sigma xy = a\Sigma x + b\Sigma x^2 + c\Sigma x^3$$

$$\Sigma x^2y = a\Sigma x^2 + b\Sigma x^3 + c\Sigma x^4$$

Putting the values from the table,

$$12{\cdot}9 = 5a + 10b + 30c$$

$$37{\cdot}1 = 10a + 30b + 100c \qquad \text{..... (1)}$$

$$130{\cdot}3 = 30a + 100b + 354c \qquad \text{..... (2)}$$

We first eliminate a. For this multiply (1) by 2 and subtract the result from the second. ..... (3)

$$\therefore \quad 37{\cdot}1 = 10a + 30b + 100c$$

$$25{\cdot}8 = 10a + 20b + 60c$$

$$11{\cdot}3 = 10b + 40c \qquad \text{..... (4)}$$

Now, multiply (1) by 6 and subtract the result from (3),

$$\therefore \quad 130{\cdot}3 = 30a + 100b + 354c$$

$$77{\cdot}4 = 30a + 60b + 180c$$

$$52{\cdot}9 = 40b + 174c \qquad \text{..... (5)}$$

Now, multiply (4) by 4 and subtract the result from (5).

$$\therefore \quad 52{\cdot}9 = 40b + 174c$$

$$45{\cdot}2 = 40b + 160c$$

$$7{\cdot}7 = 14c \qquad \therefore \quad c = \frac{7{\cdot}7}{14} = 0{\cdot}55.$$

Putting this value of c in (4), we get,

$$11{\cdot}3 = 10b + 40\,(0{\cdot}55)$$

$$\therefore \quad 10b = 11{\cdot}3 - 22 = -10{\cdot}7 \qquad \therefore \quad b = -1{\cdot}07$$

Putting the values of b and c in (1), we get

$$12{\cdot}9 = 5a + 10\,(-1{\cdot}07) + 30\,(0{\cdot}55)$$

$$\therefore \quad 5a = 12{\cdot}9 + 10{\cdot}7 - 16{\cdot}5 = 7{\cdot}1 \qquad \therefore \quad a = 1{\cdot}42$$

Hence, the equation of the parabola is

$$y = 1{\cdot}42 - 1{\cdot}07x + 0{\cdot}55x^2.$$

Example 3 : Fit a non-linear trend of the form $y = a + bx + cx^2$ to the following data.

x :	0	1	2	3	4
y :	$1{\cdot}0$	$1{\cdot}5$	$1{\cdot}5$	$2{\cdot}5$	$3{\cdot}5$

Sol. :

Calculations of $\Sigma X, \Sigma Y, \Sigma X^2$

Sr. No.	x	$Y = y$	Deviations from mid point $X = x - 2$	X^2	X^3	X^4	XY	X^2Y
1	0	$1{\cdot}0$	-2	4	-8	16	$-2{\cdot}0$	$4{\cdot}0$
2	1	$1{\cdot}5$	-1	1	-1	1	$-1{\cdot}5$	$1{\cdot}5$
3	2	$1{\cdot}5$	0	0	0	0	$0{\cdot}0$	$0{\cdot}0$
4	3	$2{\cdot}5$	1	1	1	1	$2{\cdot}5$	$2{\cdot}5$
5	4	$3{\cdot}5$	2	4	8	16	$7{\cdot}0$	$14{\cdot}0$
$N = 5$	10	$10{\cdot}0$	0	10	0	34	$6{\cdot}0$	$22{\cdot}0$

Thus, we have

$N = 5,\ \Sigma X = 0,\ \Sigma Y = 10,\ \Sigma X^2 = 10,\ \Sigma X^3 = 0,\ \Sigma X^4 = 34,\ \Sigma XY = 6,\ \Sigma X^2 Y = 22.$

Since, we have taken mid-point of X as origin $\Sigma X = 0$ and $\Sigma X^3 = 0$. The normal equations now reduce to

$$\Sigma Y = Na + b\Sigma X^2$$
$$\Sigma XY = a\Sigma X^2$$
$$\Sigma X^2 Y = a\Sigma X^2 + c\Sigma X^4$$

Putting the above values, we get

$$\therefore \quad 22 = 10a + 34c \quad \text{and} \quad 6 = 10b \quad \therefore\ b = \frac{6}{10}$$
$$20 = 10a + 20c$$
$$2 = 14c \quad \therefore\ c = \frac{2}{14} = \frac{1}{7}$$

$\therefore\ 10 = 5a + \frac{10}{7} \quad \therefore\ a = \frac{12}{7}.\quad \therefore$ The equation is $Y = \frac{12}{7} + \frac{6}{10}X + \frac{1}{7}X^2.$

Now putting $Y = y$ and $X = x - 2$, we get,

$$y = \frac{12}{7} + \frac{6}{10}(x-2) + \frac{1}{7}(x-2)^2 = \frac{12}{7} + \frac{6}{10}x - \frac{12}{10} + \frac{1}{7}x^2 - \frac{4}{7}x + \frac{4}{7}$$
$$\therefore\ y = \frac{38}{35} + \frac{1}{35}x + \frac{1}{7}x^2.$$

Example 4 : Fit a second degree parabolic curve to the following data.

x : 1, 2, 3, 4, 5, 6, 7, 8, 9.

y : 2, 6, 7, 8, 10, 11, 11, 10, 9.

Sol. : Since the values of x are odd and are equally spaced we change x to X by the relation $X = x - 5$ and put $y = Y$.

Let the equation of the parabola be $Y = a + bX + cX^2$. Then the normal equations are

$$\Sigma Y = Na + b\Sigma X + c\Sigma X^2$$
$$\Sigma XY = a\Sigma X + b\Sigma X^2 + c\Sigma X^3$$
$$\Sigma X^2 Y = a\Sigma X^2 + b\Sigma X^3 + c\Sigma X^4$$

Calculations of ΣX, ΣX^2 etc.

Sr. No.	x	$X = x - 5$	$Y = y$	X^2	X^3	X^4	XY	X^2Y
1	1	−4	2	16	−64	256	−8	32
2	2	−3	6	9	−27	81	−18	54
3	3	−2	7	4	−8	16	−14	28
4	4	−1	8	1	−1	1	−8	8
5	5	0	10	0	0	0	0	0
6	6	1	11	1	1	1	11	11
7	7	2	11	4	8	16	22	44
8	8	3	10	9	27	81	30	90
9	9	4	9	16	64	256	36	144
$N = 9$		0	74	60	0	708	51	411

Now the normal equations become

$$74 = 9a + 60c \qquad \text{..... (1)}$$
$$51 = 60b \qquad \therefore\ b = 0{\cdot}85 \qquad \text{..... (2)}$$
$$411 = 60a + 708c \qquad \text{..... (3)}$$

Now to find a and c multiply (1) by 60 and (3) by 9.

$$\therefore\ 4440 = 540a + 3600c$$
$$3699 = 540a + 6372c$$
$$741 = -2772c \quad \therefore\ c = -0{\cdot}2673$$

Now, $9a = 74 - 60c = 74 + 16{\cdot}038 = 90{\cdot}038$

$\therefore\ a = 10{\cdot}0042 = 10$

$\therefore$ The equation of the parabola is $Y = 10 + 0{\cdot}85X - 0{\cdot}27X^2$.

i.e. $y = 10 + 0{\cdot}85\,(x-5) - 0{\cdot}27\,(x-5)^2 \quad \therefore\ y = -1 + 3{\cdot}55x - 0{\cdot}27x^2.$

Example 5 : Fit a second degree parabolic curve to the following data and estimate the production in 1982.

Year (x) :	1974	75	76	77	78	79	80	91
Production (y) : (in tons)	12	14	26	42	40	50	52	53

Sol. : Since the values of X though not odd are equispaced we change x to X by the relation $X = (x - 1977{\cdot}5)2$ and we put $y = Y$.

Let the equation of the parabola be $Y = a + bX + cX^2$. Then the normal equations are

$$\Sigma Y = Na + b\Sigma X + c\Sigma X^2$$
$$\Sigma XY = a\Sigma X + b\Sigma X^2 + c\Sigma X^3$$
$$\Sigma X^2 Y = a\Sigma X^2 + b\Sigma X^3 + c\Sigma X^4$$

Calculations of ΣX, ΣX^2 etc.

Sr. No.	x	X	Y	X^2	X^3	X^4	XY	X^2Y
1	1974	−7	12	49	−343	2401	−84	588
2	1975	−5	14	25	−125	625	−70	350
3	1976	−3	26	9	−27	81	−78	234
4	1977	−1	42	1	−1	1	−42	42
5	1978	1	40	1	1	1	40	40
6	1979	3	50	9	27	81	150	450
7	1980	5	52	25	125	625	260	1300
8	1981	7	53	49	343	2401	371	2597
$N = 9$		0	289	168	0	6216	547	5601

Now, the normal equations become

$$289 = 8a + 168c \qquad \text{..... (1)}$$
$$547 = 168b \qquad \therefore\ b = 3{\cdot}2559 \qquad \text{..... (2)}$$
$$5601 = 168a + 6216c \qquad \text{..... (3)}$$

Multiply (1) by 21 and subtract (3) from the result,

$$6069 = 168a + 3528c$$
$$5601 = 168a + 6216c$$
$$468 = \quad -2688c \qquad \therefore\ c = -0{\cdot}1741$$

Now $289 = 8a + 168c$

$\therefore\ 289 = 8a + (168)(-0{\cdot}1741) \qquad \therefore\ a = 39{\cdot}7811$

The equation to the curve is $Y = 39{\cdot}78 + 3{\cdot}26X - 0{\cdot}17X^2$.

Putting $Y = y$ and $X = (x - 1977{\cdot}5)2$, the equation of the parabola is

$$y = 39{\cdot}78 + 3{\cdot}26\,(x - 1977{\cdot}5)2 - 0{\cdot}17\,(x - 1977{\cdot}5)^2 \cdot 4$$
$$y = -2671997{\cdot}77 + 2695{\cdot}92x - 0{\cdot}68x^2.$$

Putting $x = 1982$, we get $y = 55{\cdot}35$.

EXERCISE - II

1. Fit a parabola to the following data and estimate the value of y for $x = 6$.

x :	1	2	3	4	5
y :	25	28	33	39	46

[**Ans. :** $Y = -0{\cdot}086 + 5{\cdot}3X + 0{\cdot}643X^2$ where $X = x - 3$, $Y = y - 33$, $y = 54{\cdot}6$]

2. Fit a second parabola to the following data taking x as the independent variable and shifting the origin to 2 for x.

x :	0	1	2	3	4
y :	1	1·8	1·3	2·5	6·3

Find the difference between the actual value and estimated value of y when $x = 2$.

[**Ans. :** $Y = 1{\cdot}48 + 1{\cdot}13X + 0{\cdot}55X^2$ where $X = x - 2$. Difference $= -0{\cdot}18$]

3. Fit a parabola to the following data :

x :	−2	−1	0	1	2
y :	1·0	1·8	1·3	2·5	6·3

[**Ans. :** $y = 1{\cdot}48 + 1{\cdot}13x + 0{\cdot}55x^2$]

4. Fit a parabola to the following data :

x :	−2	−1	0	1	2
y :	−3·150	−1·390	0·620	2·880	5·378

[**Ans. :** $y = 0{\cdot}621 + 2{\cdot}1326x + 0{\cdot}1233x^2$]

5. Fit a second degree curve to the following data and estimate the value of y when $x = 80$

x :	10	20	30	40	50	60	70
y :	20	60	70	80	90	100	100

[**Ans. :** Put $X = (x - 40) / 10$, $Y = y / 10$, $Y = 8{\cdot}381 + 1{\cdot}2143X - 0{\cdot}2381X^2$ when $x = 80$, $y = 94{\cdot}286$]

6. Fit a parabola to the following data and estimate y when $x = 10$.

x :	1	2	3	4	5	6	7	8	9
y :	2	6	7	8	10	11	11	10	9

[**Ans. :** $Y = 3 + 0{\cdot}85X - 0{\cdot}27X^2$ where $X = x - 5$, $Y = y - 7$, $y = 7{\cdot}5$]

7. Fit a second degree curve to the following data and estimate the production in 1975.

Year :	1921,	31,	41,	51,	61,	71,	81.
Production (in tons) :	3,	5,	9,	10,	12,	14,	15.

[**Ans. :** $Y = 10{\cdot}33 + 2{\cdot}0357X - 0{\cdot}1547X^2$ where $X = (x - 1951) / 10$. Production $= 14{\cdot}32$]

8. Fit a second degree curve to the following data

Year :	1965,	66,	67,	68,	69,	70,	71,	72.
Profit (in Crores Rs.) :	125,	140,	165,	195,	200,	215,	220,	230.

Also estimate the profit in 1973.

[**Ans. :** $Y = 194{\cdot}68 + 7{\cdot}68X - 0{\cdot}40X^2$ where $X = (x - 1968{\cdot}5) / 2$. And in 1973 profit = 231·4 crores of Rs.]

4. Fitting Exponential Curve

If we have to fit an exponential curve of the form $y = ax^b$ or $y = ab^x$ or $y = ae^x$, we take logarithm of both sides. Then the given exponential equation transforms into a linear equation. We then use the above technique.

Suppose the law as $y = ab^x$.

Taking logarithm of both sides to the base 10, we get

$$\log_{10} y = \log_{10} a + x \log_{10} b$$

Putting $\log_{10} y = Y$, $\log_{10} a = A$, $x = X$, $\log_{10} b = B$, we get

$$Y = A + BX$$

This is, a linear equation in X and Y and we can obtain the constants A and B by the method explained in § 2 and resubstituting for A and B, we can obtain the equation $y = ab^x$.

If the law is $y = ax^b$ taking logarithm of both sides, we get

$$\log y = \log a + b \log x.$$

Putting $\log y = Y$, $\log a = A$, $b = B$ and $\log x = X$, we get

$$Y = A + BX.$$

Again using the technique of § 2, we can find A and B and then a and b.

Example 1 : Fit a curve $y = ab^x$ to the following data, using the method of least squares.

x :	2	3	4	5	6
y :	144	172·8	207·4	248·8	298·5

Sol. : Taking logarithms of $y = ab^x$, we get $\log y = \log a + (\log b)\, x$. Let us put $\log y = Y$, $\log a = A$, $\log b = B$ and $x = X$. Then the law becomes $Y = A + BX$.

Calculations of ΣX, ΣY, etc.

Sr. No.	x	y	X	Y	X^2	XY
1	2	144·0	2	2·1584	4	4·3168
2	3	172·8	3	2·2375	9	6·7125
3	4	207·4	4	2·3168	16	9·2672
4	5	248·8	5	2·3959	25	11·9795
5	6	298·5	6	2·4749	36	14·8494
$N = 5$			20	11·5835	90	47·1254

Now, we have

$\sum Y = NA + B\sum X$ $\quad\therefore\ 11{\cdot}5835 = 5A + 20B$

$\sum XY = A\sum X + B\sum X^2$ $\quad\therefore\ 47{\cdot}1254 = 20A + 90B$

Solving these equations, we get,

$B = 0{\cdot}07914$ and $A = 2{\cdot}0001$

Hence, b = antilog $0{\cdot}07914 = 1{\cdot}199$, a = antilog $2{\cdot}0001 = 100$.

$\therefore$ The law is $y = 100\,(1{\cdot}199)^x$.

Example 2 : Find the law of the form $y = ab^x$ to the following data.

x :	1	2	3	4	5	6	7	8
y :	1	1·2	1·8	2·5	3·6	4·7	6·6	9·1

Sol. : Taking logarithm of both sides of $y = ab^x$, we get

$\log y = \log a + x \log b$.

Putting $\log y = Y$, $\log a = A$, $x = X$, $\log b = B$, we get $Y = A + BX$.

We, now, form the following table.

Calculations of $\sum X$, $\sum Y$, etc.

Sr. No.	$X = x$	y	$Y = \log y$	XY	X^2
1	1	1·0	0·0000	0·0000	1
2	2	1·2	0·0792	0·1584	4
3	3	1·8	0·2553	0·7689	9
4	4	2·5	0·3979	1·5916	16
5	5	3·6	0·5563	2·7815	25
6	6	4·7	0·6721	4·0326	36
7	7	6·6	0·8195	5·7365	49
8	8	9·1	0·9590	7·6720	64
$N = 8$	36		3·7393	22·7385	204

Now, $\quad Y = A + BX$

$\therefore\ \sum Y = NA + B\sum X$

$\sum XY = A\sum X + B\sum X^2$

Putting $N = 8$, $\sum X = 36$, $\sum Y = 3{\cdot}7393$, $\sum XY = 22{\cdot}7385$, $\sum X^2 = 204$, we get

$3{\cdot}7393 = 8A + 36B$

$22{\cdot}7385 = 36A + 204B$

Solving the two equations, we get,

$A = -0{\cdot}1657$, $B = 0{\cdot}1407$

$\therefore\ a$ = anti-log $(-0{\cdot}1657) = 0{\cdot}6828$

b = anti-log $(0{\cdot}1407) = 1{\cdot}3828$

Hence, the required law correct to two places of decimals is

$y = 0{\cdot}68\,(1.38)^x$.

Example 3 : Fit a curve of the form $y = ax^b$ to the following data.

x :	1	2	3	4	5	6
y :	120	90	60	20	11	5

Sol. : Taking logarithms of both sides of $y = ax^b$, we get

$\log y = \log a + b \log x$.

Putting $\log y = Y$, $\log a = A$, $b = B$, $\log x = X$, we get

$Y = A + BX$.

Calculations of $\sum X$, $\sum Y$, etc.

Sr. No.	x	y	$X = \log x$	$Y = \log y$	XY	X^2
1	1	120	0·0000	2·0792	0·0000	0·0000
2	2	90	0·3010	1·9542	0·5882	0·0906
3	3	60	0·4771	1·7782	0·8484	0·2276
4	4	20	0·6021	1·3010	0·7833	0·3625
5	5	11	0·6990	1·0414	0·7279	0·4886
6	6	5	0·7781	0·6990	0·5439	0·6054
$N = 6$			2·8573	8·8530	3·4917	1·7747

Now, the normal equations are

$\sum Y = NA + B\sum X$

$\sum XY = A\sum X + B\sum X^2$

Putting the values from the table

$8{\cdot}8530 = 6A + 2{\cdot}8573B$ ..... (1)

$3{\cdot}4917 = 2{\cdot}8573A + 1{\cdot}7747B$ ..... (2)

Multiply (1) by $2{\cdot}8573$ and (2) by 6 and subtract.

$$\begin{aligned} 25{\cdot}2957 &= 17{\cdot}1438A + 8{\cdot}1642B \\ 20{\cdot}9502 &= 17{\cdot}1438A + 10{\cdot}6482B \\ \hline 4.3455 &= \qquad\quad - 2.4840B \end{aligned}$$

$\therefore\ B = -1{\cdot}7494$

Putting this value of B in (1), we get

$6A = 8{\cdot}8530 + 4{\cdot}9991$ $\quad\therefore\ A = 2{\cdot}3087$

$\therefore\ a$ = antilog A = antilog $2{\cdot}3087 = 203{\cdot}56$

and $\quad b = B = -1{\cdot}7494$

Hence, we get $y = 203{\cdot}86x^{-1{\cdot}494}$.

Example 4 : Fit a curve of the form $y = ab^x$ to the following data.

x :	1	2	3	4	5	6
y :	151	100	61	50	20	8

Sol. : Let the equation of the curve be $y = ab^x$.

Taking logarithms of both sides, we get

$\log y = \log a + x \log b$

Let $\log y = Y$, $\log a = A$, $\log b = B$, $X = x$.

$\therefore \quad Y = A + BX$.

Calculations of ΣX, ΣY, etc.

Sr. No.	$X = x$	y	$Y = \log y$	X^2	XY
1	1	151	2·1790	1	2·1790
2	2	100	2·0000	4	4·0000
3	3	61	1·7853	9	5·3559
4	4	50	1·6990	16	6·7960
5	5	20	1·3010	25	6·5050
6	6	8	0·9031	36	5·4186
$N = 6$	21		9·8674	91	30·2545

Now, the normal equations are

$$\Sigma Y = NA + B\Sigma X$$

$$\Sigma XY = A\Sigma X + B\Sigma X^2$$

Putting the above values, we get,

$$9{\cdot}8674 = 6A + 21B \qquad \text{———(1)}$$

$$30{\cdot}2545 = 21A + 91B \qquad \text{———(2)}$$

Multiply (1) by 7 and (2) by 2 and subtract.

$$\therefore \quad 69{\cdot}0718 = 42A + 147B$$

$$60{\cdot}5090 = 42A + 182B$$

$$8{\cdot}5628 = \quad -35B \qquad \therefore \quad B = -0{\cdot}2447$$

Putting the value of B in (1), we get

$$9{\cdot}8674 = 6A + 21\,(-0{\cdot}2447)$$

$$\therefore \quad 6A = 9{\cdot}8674 + 5{\cdot}1387$$

$$6A = 15{\cdot}0061 \qquad \therefore \quad A = 2{\cdot}5010$$

$$\therefore \quad a = \text{antilog } A = 316{\cdot}96$$

$$b = \text{antilog } B = 0{\cdot}5692$$

$$\therefore \quad y = 316{\cdot}96\,(0{\cdot}5692)^x.$$

EXERCISE - III

Class (c) : 8 Marks

1. Fit a curve of the form $y = ab^x$ to the following data.

   x : 0 2 4

   y : 5·012 10·000 31·620 [**Ans.** : $y = 4{\cdot}68\,(1{\cdot}2$

2. Fit a curve of the form $y = ax^b$ to the following data by the method of least squares.

   x : 1 2 3 4

   y : 2·50 8·00 19·00 50·00 [**Ans.** : $y = 2{\cdot}22\,x$

3. Fit a curve of the form $y = ax^b$ to the following data by the method of least squares.

   x : 1 2 3 4

   y : 76 210 390 605 [**Ans.** : $y = 75{\cdot}2\,x^{1{\cdot}61}$]

4. Fit a curve of the form $y = ab^x$ to the following data using the method of least squares.

   x : 2 3 4 5 6

   y : 144 172·8 207·4 248·8 298·5 [**Ans.** : $y = 100\,(1{\cdot}999)^x$]

5. Fit a curve of the form $y = ab^x$ to the following data using the method of least squares.

   x : 2 3 4 5 6

   y : 34·385 79·0855 181·90 418·36 962·23 [**Ans.** : $y = 6{\cdot}5\,(2{\cdot}3)^x$]

☆ ☆ ☆

CHAPTER 4

Probability Distributions

1. Introduction

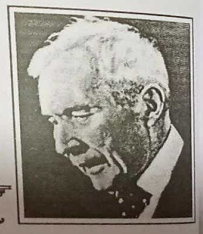
You were introduced probability in the X-th standard and you have studied laws of probability, Bayes' theorem probability distributions, particularly Binomial distribution in XII-th standard. In this chapter we shall briefly review probability and then study once again Bays' Theorem in detail. Then we shall study probability distributions in general.

2. Terms used in Axiomatic Theory

The limitations of classical theory of probability are removed by putting the theory in axiomatic form. This approach was suggested by **A. N. Kolmogorov**, a Russian mathematician in 1933.

It should be noted that as in any other axiomatic theory (*e.g.* Geometry) we start with certain undefined terms, state certain axioms about them and then deduce theorems about these terms strictly from the axioms by rules of logic. Before developing the axiomatic probability theory we explain below certain peculiar terms used therein.

Andrey Nikolaevich Kolmogorov (1903 - 1987)



A well-known Russian mathema-tician known for his great contributions to the fields of probability theory, topology, intuitionistic logic, classical mechanics and others. In 1922 he constructed a Fourier Series that diverges almost everywhere. His pioneering work "About The Analytical Methods Of Probability Theory" was published in 1931, in which year he become professor at Moscow University. In 1933 he published "Foundations of The Theory of Probability" laying the foundation of modern axiomatic theory. In 1939 he become a member of U.S.S.R. Academy of Sciences. He along with British mathematician Chapman developed "The Chapman-Kolmogorov Equations" in Random processes. He was a founder of algorithm complexity theory known as "Kolmogorov Complexity Theory". A quotation of Kolmogorov. "Every mathematician believes that he is ahead of all others. The reason why they do not say this in public is they are intelligent people".

(1) Sample space

The aggregate or the set of all possible outcomes of an experiment is called *sample space*. the outcomes *i.e.* the members of the set themselves being called *sample points*. Sample space is denoted by S or Ω and sample points by $w_1, w_2, w_3, \ldots$. Thus, $\Omega = \{w_1, w_2, w_3, \ldots\}$.

head (H) and a tail (T) and the sample space is

Example 1 : In a random experiment of tossing a coin the sample points are its outcomes a

$$\Omega = \{H, T\}.$$

Example 2 : In a random experiment of tossing a dice the sample points are 1, 2, 3, 4, 5 and 6 the sample space is

$$\Omega = \{1, 2, 3, 4, 5, 6\}.$$

Example 3 : In a throw of two coins the sample points are $(H, H), (H, T), (T, H), (T, T)$, and the sample space is

$$\Omega = \{(H, H), (H, T), (T, H), (T, T)\}.$$

(2) Event

A subset of a sample space is called an *event*.

(i) In a throw of a coin $\Omega = \{H, T\}$. We have shown two events

$A_1 = \{H\}$,

$A_2 = \{T\}$.

(ii) In a throw of a die $\Omega = \{1, 2, 3, 4, 5, 6\}$, there are a number of events.

$A_1 = \{1\}$, $A_2 = \{2\}$, $A_3 = \{3, 4, 5, 6\}$, ...

$B_1 = \{1, 2\}$, $B_2 = \{1, 3\}$ .....

$C_1 = \{1, 2, 3\}$, $C_2 = \{1, 2, 4\}$, ... etc.

We have shown the first three.

(iii) In a throw of a coin and a die there are again a number of events

$A_1 = \{(H, 1), (H, 2)\}$

$A_2 = \{(H, 1), (H, 3), (H, 5),\}$ .....

$B_1 = \{(T, 1), (T, 2)\}$

$B_2 = \{(T, 1), (T, 3), (T, 5)\}$ ..... etc.

3. Axiomatic Definition of Probability

Let E be a random experiment and S be the sample space. We define the probability $P(A)$ for every element A of S (*i.e.* for every subset A of the sample space S) satisfying the following axioms :

1. $P(A) \geq 0$ (Axiom 1)
2. $P(S) = 1$ (Axiom 2)
3. $P(A \cup B) = P(A) + P(B)$ (Axiom 3)

if A and B are any two exclusive events (*i.e.* they are disjoint sets).

Explanation : The first axiom states that the probability of any event is greater than zero, which means the probability of any event cannot be negative.

The second axiom stales that the **sum** of all the probabilities is equal to 1. This together with the first axiom states that the probability of an event must be less than or equal to 1 *i.e.* $P(A) \leq 1$.

The third axiom states that if two events are mutually exclusive, then the probability that either of them will occur is equal to the some of their probabilities.

For our study the classical definition is sufficient. We give that definition in terms of sets again for ready reference.

Definition : If a sample space S has n points which are equally likely and mutually exclusive and an event A has m points then the ratio m/n is called probability of A and is denoted by $P(A)$. [Fig. 4.1 (a)]

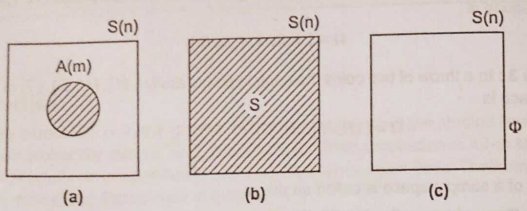


Fig. 4.1

Thus,
$$\boxed{P(A) = \frac{m}{n} = \frac{\text{number of points in } A}{\text{number of point in } S}}$$

Note ....

We shall be concerned with only such random experiments in which all sample points are equally likely. And to find the probability of an event in such a random experiment we need to know the number of points in sample space S and the number of points in the event A.

4. Laws of Probability

There are two laws. The probability of **A or B** *i.e.* $P(A \cup B)$ is given by the *law of addition* and the probability of **A and B** *i.e.* $P(A \cap B)$ is given by the *law of multiplication*.

Theorem 1 : For any two events A and B the probability that exactly B will occur is given by

$$\boxed{P(B \cap \overline{A}) = P(B) - P(A \cap B)}$$

and that exactly A will occur is given by

$$\boxed{P(A \cap \overline{B}) = P(A) - P(A \cap B)}$$

where $\overline{A}$ denotes the complement of A and $\overline{B}$ denotes the complement of B.

Proof : To prove this result we first express the event B as the union of two exclusive events $A \cap B$ and $\overline{A} \cap B$.

$\therefore \quad B = (A \cap B) \cup (\overline{A} \cap B)$

Since the events on the r.h.s. are exclusive

$P(B) = P(A \cap B) + P(\overline{A} \cap B)$ [By Axiom 3]

$\therefore \quad P(B \cap \overline{A}) = P(B) - P(A \cap B)$

Similarly, we can prove that

$P(A \cap \overline{B}) = P(A) - P(A \cap B)$

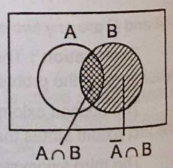


Fig. 4.2

Theorem 2 : (Additive Theorem) (Two Events)

Probability that atleast one of the events A and B will occur is given by

$P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Proof : We first express the event $A \cup B$ is the union of two exclusive events A and $\overline{A} \cap B$.

$A \cup B = A \cup (\overline{A} \cap B)$

$\therefore \quad P(A \cup B) = P[A \cup (\overline{A} \cap B)]$

But the events on the r.h.s. are mutually exclusive.

$\therefore \quad P(A \cup B) = P(A) + P(\overline{A} \cap B)$ [By Axiom 3]

$= P(A) + P(B) - P(A \cap B)$ [By Theorem 2]

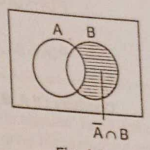


Fig. 4.3

Corollary 1 : If A and B are two mutually exclusive events then the probability that either A or B will happen is the sum of the probabilities of A and B *i.e.*

$$\boxed{P(A \cup B) = P(A) + P(B)}$$

Proof : Since events are exclusive $A \cap B = \Phi$.

$\therefore \quad P(A \cap B) = P(\Phi) = 0$ [By Theorem 1]

Hence, from the above theorem,

$P(A \cup B) = P(A) + P(B)$

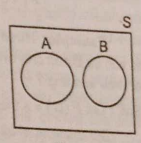


Fig. 4.4

Corollary 2 : If $A, B, C, \ldots, K$ are mutually exclusive events such that their union is the whole of sample space then

$$\boxed{P(A) + P(B) + P(C) + \ldots + P(K) = 1}$$

Proof : Since $A \cup B \cup C \cup \ldots \cup K = S$

$\therefore \quad P(A \cup B \cup C \cup \ldots \cup K) = P(S)$

Since, events are mutually exclusive, we get, from above,

$P(A) + P(B) + P(C) + \ldots + P(K) = P(S)$

But $P(S) = 1$ [By Axiom 2]

$\therefore \quad P(A) + P(B) + P(C) + \ldots + P(K) = 1$

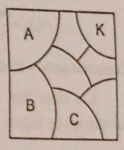


Fig. 4.5

Note ....

The group of all possible events $A, B, C, \ldots, K$ of the sample space S such that

$P(A) + P(B) + P(C) + \ldots + P(K) = 1$

is called **exhaustive.** *e.g.* in the toss of a coin H, T; in the toss of a dice 1, 2, 3, 4, 5, 6 are exhaustive events.

Theorem 3 : *Addition Theorem (Three Events)* : If A, B, C are any three events then the probability that *at least one* of them will occur is given by

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(B \cap C) - P(C \cap A) + P(A \cap B \cap C).$$

Proof : We prove this theorem by considering the union of A and $(B \cup C)$ and applying the above theorem.

$\therefore \quad P(A \cup (B \cup C)) = P(A) + P(B \cup C) - P(A \cap (B \cup C))$

But $P(B \cup C) = P(B) + P(C) - P(B \cap C)$

And $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$

$$\therefore \quad P(A \cap (B \cup C)) = P[(A \cap B) \cup (A \cap C)]$$
$$= P(A \cap B) + P(A \cap C) - P[(A \cap B) \cap (A \cap C)]$$
$$\therefore \quad P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(B \cap C)$$
$$- P(B \cap C) - P(A \cap B) - P(A \cap C) + P(A \cap B \cap C)$$

Corollary : If A, B, C are three events which are pairwise exclusive then

$$\boxed{P(A \cup B \cup C) = P(A) + P(B) + P(C)}$$

Proof : Since, A, B are exclusive $A \cap B = \Phi$. Since, $P(\Phi) = 0$, $P(A \cap B) = 0$. Similarly, $P(B \cap C) = 0$, $P(C \cap A) = 0$, $P(A \cap B \cap C) = 0$, $P(C \cap A) = 0$, $P(A \cap B \cap C) = 0$.

Hence, from the above result,

$$P(A \cup B \cup C) = P(A) + P(B) + P(C).$$

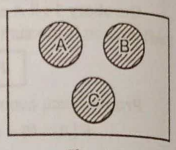


Fig. 4.6

Example : Three horses A, B and C are in a race. A is twice as likely to win as B and B is twice as likely to win as C. What are the probabilities of their winning ?

Sol. : Let $P(C) = x$ then $P(B) = 2x$ and $P(A) = 4x$.

But the sum of all probabilities is 1.

$$\therefore \quad P(A) + P(B) + P(C) = 1 \quad \therefore \quad 4x + 2x + x = 1 \quad \therefore \quad x = \frac{1}{7}$$

$$\therefore \quad P(A) = \frac{4}{7}, \quad P(B) = \frac{2}{7}, \quad P(C) = \frac{1}{7}.$$

Complementary Events

It is clear that $A \cup \overline{A} = S$ *i.e.* the union of the event and its complement is the whole of sample space S and $\overline{A} \cap A = \Phi$ *i.e.* the event A and its complement are exclusive.

Theorem 4 : $\boxed{P(\overline{A}) = 1 - P(A)}$

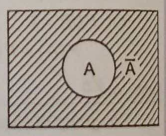


Fig. 4.7

Proof : Since $A \cup \overline{A} = S$, $P(A \cup \overline{A}) = P(S)$. But $P(S) = 1$ and A, $\overline{A}$ are exclusive.

By Axiom 2, $\quad P(A \cup \overline{A}) = P(A) + P(\overline{A})$

$\therefore \quad P(A) + P(\overline{A}) = 1 \quad \therefore \quad P(\overline{A}) = 1 - P(A)$

Note ....

Since A and $\overline{A}$ are exclusive, we have $A \cup \overline{A} = S$ and $A \cap \overline{A} = \Phi$.

Corollary : Probability of an event is always less than or equal to one.

i.e. $\boxed{P(A) \le 1}$

Proof : $P(A) = 1 - P(\overline{A})$. But $P(\overline{A}) \ge 0$ by axiom 1.

$\therefore \quad P(A) \le 1.$

De Morgan's Laws

Since an event is a subset of sample space, De Morgan's are laws are applicable to events. Thus, we get

$$\boxed{\begin{aligned} P(\overline{A \cup B}) &= P(\overline{A} \cap \overline{B}) \\ P(\overline{A \cap B}) &= P(\overline{A} \cup \overline{B}) \end{aligned}}$$

Augustus De Morgan (1806 - 1871)

Augustus De Morgan was born in Madurai, Tamil Nadu. His father was a colonel in the Indian army. His family returned to England when he was 7 months old. When in schools he mastered Latin, Greek and Hebrew and developed strong interest in mathematics.

He was a fellow of the Astronomical Society and a founder of London Mathematical Society. De Morgan greatly influenced the development of mathematics in the 19th century. He was a prolific writer and wrote over 1000 articles in more than 15 journals, in addition to a number of books known for clarity, logical presentation and minute details. He made original contributions to analysis and logic. He coined the term mathematical induction and gave first precise definition of limit in his book "The Differential And Integral Calculus".

Example 1 : Find the probability that a card drawn will be black or picture ?

Sol. : Let A = the card is picture and B = the card is black.

Now A and B are not mutually exclusive events. There are 6 black pictures

$$P(A) = \frac{12}{52}, \quad P(B) = \frac{26}{52}, \quad P(A \cap B) = \frac{6}{52}$$

$$\therefore \quad P(A \cup B) = P(A) + P(B) - P(A \cap B)$$
$$= \frac{12}{52} + \frac{26}{52} - \frac{6}{52} = \frac{32}{32} = \frac{8}{13}.$$

Note ....

Many examples of this type can be more easily solved by counting the number of points in the desired event. In the above example $A \cup B$ = {Black or Picture Cards} has 26 + 12 – 6 = 32 points.

$$\therefore \quad P(A \cup B) = \frac{32}{52} = \frac{8}{13}.$$

Example 2 : Two cards are drawn from a pack of cards. Find the probability that they will be both red or both pictures.

Sol. : Let A = { both red } and B = { both pictures }

$$P(A) = \frac{{}^{26}C_2}{{}^{52}C_2}, \quad P(B) = \frac{{}^{12}C_2}{{}^{52}C_2}$$

But A and B are not exclusive. There are six red picture cards.

$$\therefore \quad P(A \cap B) = \frac{{}^{6}C_2}{{}^{52}C_2},$$

$$\therefore \quad P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$= \frac{{}^{26}C_2}{{}^{52}C_2} + \frac{{}^{12}C_2}{{}^{52}C_2} - \frac{{}^{6}C_2}{{}^{52}C_2} = \frac{203}{663}.$$

5. Conditional Probability

Consider general case of conditional probability (See Fig. 4.8 on page 4-8). Let n = the number of points in S, m_1 = number of points in A, m_2 = number of points in B, m_{12} = number of points in A and B both.

Then, $P(A) = \frac{m_1}{n}$, $P(B) = \frac{m_2}{n}$ and $P(A \cap B) = \frac{m_{12}}{n}$.

Now, suppose in a trial we know the result partially *i.e.* we know that A has occurred. What is the probability now that B has occurred along with A ?

Since we know that A has occurred the outcome of the trial is one of those points in A *i.e.* one of m_1 (and not n) And of these, m_{12} are in B. Hence, the probability that B will occur, when A has already occurred is (m_{12} / m_1). This is called conditional probability of B under the condition that A has occurred. It is denoted by $P(B / A)$.

Thus, $\quad P(B / A) = \frac{m_{12}}{m_1}$.

Definition : Let A and B be any two events in a sample space S. The probability that B will occur, given that A has already occurred is called the **conditional probability of B** and is denoted by $P(B / A)$.

Similarly the probability that A will occur given that B has already occurred is called the **conditional probability of A** and is denoted by $P(A / B)$.

Example : Suppose there are 100 students in a class and the results of an examination of the class are given in the following table.

	Passed	Failed	Total
Boys	28	32	60
Girls	26	14	40
Total	54	46	100

Let us define two events A, B as follows :

A = a student has passed, $\quad B$ = a student is a male student.

Suppose a student is selected at random and is known to be a male student. What is the probability that this student has passed ? In symbols, we want to find $P(A / B)$.

Since the student is a male student, the sample space of B has 60 points. Of these 28 have passed *i.e.* A has now 28 points. Hence,

$$\therefore \quad P(A / B) = \frac{28}{60} = \frac{7}{15}$$

But, from the table we see that 28 is the number of points in $A \cap B$ and 60 is the number of points in B

$$\boxed{P(A / B) = \frac{\text{No. of points in } A \cap B}{\text{No. of points in } B}} \quad \text{.....(1)}$$

or

$$\boxed{P(A / B) = \frac{\text{No. of points in } A \text{ out of } B}{\text{No. of points in } B}} \quad \text{.....(2)}$$

Similarly, we can find that

$$\boxed{P(B / A) = \frac{\text{No. of points in } B \cap A}{\text{No. of points in } A}} \quad \text{.....(3)}$$

or

$$\boxed{P(B / A) = \frac{\text{No. of points in } B \text{ out of } A}{\text{No. of points in } A}} \quad \text{.....(4)}$$

Further, we see that, (1) can be written as

$$P(A / B) = \frac{(\text{No. of points in } A \cap B) / \text{No. of points in } S}{(\text{No. of points in } B) / \text{No. of points in } S}$$

$$\therefore \quad P(A / B) = \frac{P(A \cap B)}{P(B)}$$

Similarly, from (3) we can get $\quad P(B / A) = \frac{P(A \cap B)}{P(A)}$.

From these we get $\quad P(A \cap B) = P(A / B) \times P(B)$ and $P(A \cap B) = P(B / A) \times P(A)$ which is called the law of multiplication of probability.

Theorem 5 : *Multiplication Theorem*

If A and B are two events and neither is null then the probability that both of them will occur is given by

$$\boxed{P(A \cap B) = P(A) \times P(B / A)} \quad \text{or} \quad \boxed{P(A \cap B) = P(B) \times P(A / B)}$$

where $P(A / B)$ and $P(B / A)$ denote the conditional probabilities which are greater than zero.

Proof : Let the number of points in A be m_1 and those in B be m_2. Let the number of points in $(A \cap B)$ be m_{12} and let n be the total number of points in S.

$$P(A) = \frac{m_1}{n}, \quad P(B) = \frac{m_2}{n}$$

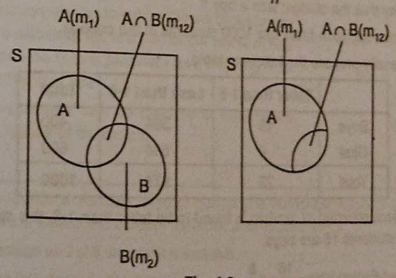


Fig. 4.8

To find the probability of B when A has happened, we have to consider the sample space of A which has m_1 points. In this sample space only B can occur (along with A) which has m_{12} points.

$$\therefore \quad P(B/A) = \frac{\text{No. of points in } A \cap B}{\text{No. of points in } A} = \frac{m_{12}}{m_1}$$

$$\text{Now,} \quad P(A \cap B) = \frac{m_{12}}{n} = \frac{m_1}{n} \times \frac{m_{12}}{m_1}$$

$$\therefore \quad P(A \cap B) = P(A) \times P(B/A)$$

Similarly, we can prove that $P(A \cap B) = P(B) \times P(A/B)$.

Example 1 : There are 11 tickets in a box bearing numbers 1 to 11. Three tickets are drawn one after the other without replacement. Find the probability that they are drawn in the order bearing (i) even, odd, even number, (ii) odd, odd, even number. **(M.U. 1997)**

Sol. : This is an example on conditional probability. Out of 11 tickets, 5 are even and 6 are odd.

$$P(\text{even, odd, even}) = P(\text{even}) \cdot P(\text{odd}) \cdot P(\text{even})$$

$$= \frac{5}{11} \cdot \frac{6}{10} \cdot \frac{4}{9} = \frac{4}{33}$$

$$P(\text{odd, odd, even}) = P(\text{odd}) \cdot P(\text{odd}) \cdot P(\text{even})$$

$$= \frac{6}{11} \cdot \frac{5}{10} \cdot \frac{5}{9} = \frac{5}{33}$$

Example 2 : In a bag there are 4 white and 3 black balls. If four balls are drawn one by one at random without replacement, what is probability that the balls so drawn are alternately of different colours ?

Sol. : Try.

$$P(A) = \frac{4}{7} \cdot \frac{3}{6} \cdot \frac{3}{5} \cdot \frac{2}{4} = \frac{3}{35}; \qquad P(B) = \frac{3}{7} \cdot \frac{4}{6} \cdot \frac{2}{5} \cdot \frac{3}{4} = \frac{3}{35}$$

$$P(\text{Alternate colour}) = \frac{3}{35} + \frac{3}{35} = \frac{6}{35}.$$

Example 3 : In a certain college 4% of the boys and 1% of the girls are taller than 1·8 mts. Furthermore 60% of the students are girls. If a student selected at random is taller than 1·8 mts, what is the probability that the student was a boy ? **(M.U. 2002, 03)**

Sol. : For convenience suppose there are 1000 students in the college.

Then we can easily prepare the following table.

	Taller than 1·8	Less than 1·8	Total
Boys	16	384	400
Girsl	6	594	600
Total	22	978	1000

Since the student selected at random is found to be taller than 1·8, the student is one of 22. But out of these 22 students 16 are boys.

$$\therefore \quad \text{Required Probability} = \frac{16}{22} = \frac{8}{11}.$$

6. Random Variable

In the previous chapter we learnt how to find the probability of an outcome and the laws of probability. We saw that the outcome of an experiment can be anything : it may be colour (black-white-red .....) of a ball, a gender (male-female) of a child, a suit (club-diamond-spade-heart) of a card or a number (1 – 2 – 3 – 4 – 5 – 6) of a die or a logical answer (yes - no) of a question or result of a toss (head - tail) of a coin. In most of the problems the outcome of an experiment is a number *e.g.* the salary of a person, the height of a student, the temperature at a place, the rainfall on a particular day. However, when the outcome is not a number we can express these outcomes in numbers by agreeing to denote,

(a) head by 1 and tail by 0 (b) boy by 1 and girl by 0

(c) yes by 1 and no by 0 (d) club by 1, diamond by 2, spade by 3 and heart by 4

(e) red by 1, white by 2 and black by 3, etc.

In probability problems it is found convenient to think of a variable and consider the values of the variable which describe the outcomes of the experiment. In the toss of a coin the variable takes values 1 and 0; in the selection of a child it takes values 1 and 0, in the answers of the question it takes the values 1 and 0, in drawing a card it takes values 1, 2, 3, 4, in drawing a ball it takes values 1, 2, 3 etc. This variable may take discrete values or may take any value in a range continuously *e.g.* the arrival time of a bus at a stop, say, between 9 a.m. and 9.10 a.m. may be any number between the range. Such a variable is called a **random variable**. Actually random variable is a misnomer. It is a function which assigns a real number to the outcome of an experiment. A random variable is denoted by X and a particular value of X is denoted by x. In the toss of a coin X assigns the value 1 to H and 0 to T, in the selection of a ball X assigns value 1 to red, 2 to white and 3 to black.

In these cases X takes discrete values and is called a discrete random variable. But in the case of arrival time of a bus X takes continuously any value between 9 am. and 9.10 am. and hence X is a continuous random variable. In this way, we consider X as a function from sample space S to the set of real numbers R. Thus we get the following definition.

(a) Definition

Let E be an experiment and S be the sample space associated with it. A function X assigning to every element s of S one and only one real number $x = X(s)$ of R is called a *random variable.*

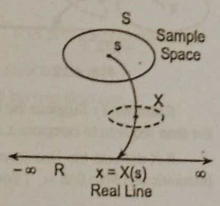


Fig. 4.9

Since X is a function whose domain is the set of outcomes of an experiment and whose range is a part or the whole of real line $(-\infty < x < \infty)$, it can be shown pictorially as a mapping from the sample space to the real line.

The random variable X can be discrete or continuous depending upon the nature of its domain.

Remarks ....

1. In simple words a variable used to denote the numerical value of the outcome of an experiment is called the random variable, abbreviated as r.v.
2. X is a *function* and still we call it a variable.
3. We are not interested in the *functional nature* of X but in the *values of X*.

4. X must be single valued *i.e.* for every s of S there corresponds *exactly one value* of X. Different elements of S may lead to the same value of X (See Example 2). But two values of X cannot be assigned to the same sample point.

5. We shall denote *random variables* by capital letters $X, Y, Z, \ldots$ and shall denote the *unknown values* of these random variables by small letters $x, y, z, \ldots, x_1, x_2, \ldots y_1, y_2, \ldots z_1, z_2, \ldots$ etc. This is an important distinction and students should note it carefully. With this notation it is meaningless to write $P(x \geq 10)$ say since x being a value of X either is or is not ≥ 10. Instead we should write $P(X \geq 10)$.

Example 1 : Suppose the experiment E is to toss a fair coin.

Then $S = \{H, T\}$. If X is the random variable denoting the number of heads then we have $X(H) = 1$ and $X(T) = 0$. [See Fig. 4.10]

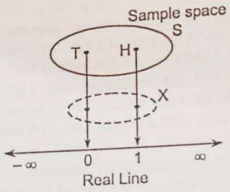


Fig. 4.10

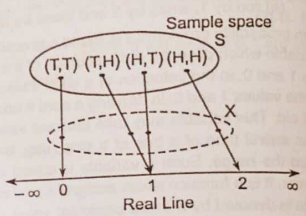


Fig. 4.11

Example 2 : Suppose the experiment E is to toss two fair coins.

Then $S = \{(H, H), (H, T), (T, H), (T, T)\}$. If X is the random variable denoting the number of heads then $X(H, H) = 2$, $X(H, T) = 1$, $X(T, H) = 1$, $X(T, T) = 0$. [See Fig. 4.11]

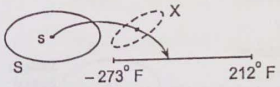


Fig. 4.12

Example 3 : Suppose the experiment is to record the temperature at a place.

If X denoted the temperature then X can take any value from -273° F to 212° F say. [See Fig. 4.12]

Example 4 : Suppose the experiment is to record the time required to complete a software project.

If X denotes the time then X can take any value (theoretically) from 0 to ∞. [See Fig. 4.13]

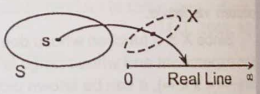


Fig. 4.13

In examples 1 and 2, X takes discrete values and in examples 3 and 4, X takes continuous all values between a specified interval. In the first case X is called a ***discrete random variable*** the second case it is called ***continuous random variable.***

(b) Definition

Let X be a random variable. If X takes finite or countably infinite values $x_0, x_1, x_2, \ldots$ then X called a ***discrete random variable.***s

(c) Definition

Let X be a random variable. If X takes uncountably infinite values in a given interval then X is called a ***continuous random variable.***

7. Probability Distribution of a Discrete Random Variable

As we know already, with every possible outcome of an experiment there will be associated its probability. We shall be interested in the values of the random variable X along with their probabilities. If x_i is the value of X and $P(x_i)$ is the probability of x_i then set of pairs $\{x_i, P(x_i)\}$ is called the **probability distribution.**

Definition : Let X be a **discrete random variable.** Let $x_1, x_2, \ldots x_n \ldots$ be the possible values of X. With each possible outcome x_i we associate a number $p(x_i) = P(X = x_i)$ called the probability of x_i. The numbers $p(x_i)$, $i = 1, 2, \ldots n \ldots$ must satisfy the following conditions :

1. $p(x_i) \geq 0$ for all i
2. $\sum_{i=1} p(x_i) = 1$

The function p is called the **probability function** or **probability mass function (p.m.f)** or **probability density function (p.d.f.)** of the random variable X and the set of pairs (x_i, p_i) is called the probability distribution of X.

The probability distribution of a discrete random variable X taking values $x_1, x_2, x_3 \ldots x_n \ldots$ with probabilities $p_1, p_2, p_3 \ldots p_n \ldots$ where $p_i \geq 0$ and $\Sigma p_i = 1$ can be given in tabular form as

X	x_1	x_2	x_3	.....	x_n	.....
$P(x_i)$	p_1	p_2	p_3	.....	p_n	.....

Example 1 : State true or false with justification :

(a) A random variable X takes values 0, 1, 2 and 3 then $p(X = x) = \frac{x-1}{2}$ can be its probability distribution.

(b) A random variable takes values 0, 1, 2 and $p(x) = \frac{x+1}{3}$ is its probability distribution.

Sol. : As seen above for a probability distribution, two conditions must be satisfied.

(i) Each probability must be equal to or greater than zero but less than one.

(ii) The sum of all probabilities must be equal to unity.

Putting $x = 0, 1, 2, 3$ in (a), we get

$$P(0) = -\frac{1}{2},\ P(1) = 0,\ P(2) = \frac{1}{2},\ P(3) = 1$$

Since, the probability cannot be negative, $P(X = x) = \frac{x-1}{2}$ cannot be a probability distribution.

Putting $x = 0, 1, 2, 3$, in (b), we get

$$P(0) = \frac{1}{3},\ P(1) = \frac{2}{3},\ P(2) = 1.$$

Although all probabilities are positive, the sum of all the probabilities is 2, (greater than 1). Hence, $P(X = x) = \frac{x+1}{3}$ also cannot be a probability distribution.

Example 2 : From the past experience it was found that the daily demand at an autogarage was as under.

Daily Demand	:	5	6	7
Probability	:	0·25	0·65	0·10

Check if this is a probability distribution. Find also the probability that over a period of two days the number of demands would be 11 or 12.

Sol. : Since the sum of all probabilities = 0·25 + 0·65 + 0·10 = 1, it is a probability distribution.

P(11 requests over two days)

$= P$(5 requests on the first day and 6 on the second)
$\quad + P$(6 request on the first day and 5 on the second)

$= (0{\cdot}25 \times 0{\cdot}65) + (0{\cdot}65 \times 0{\cdot}25)$

$= 0{\cdot}1625 + 0{\cdot}1625 = 0{\cdot}325$

P(12 requests over two days)

$= P$(5 requests on the first day and 7 on the second)
$\quad + P$(6 requests on the first day and 6 on the second)
$\quad + P$(7 requests on the first day and 5 on the second)

$= (0{\cdot}25 \times 0{\cdot}10) + (0{\cdot}65 \times 0{\cdot}65) + (0{\cdot}10 \times 0{\cdot}25)$

$= 0{\cdot}025 + 0{\cdot}4224 + 0{\cdot}025 = 0{\cdot}7975.$

Example 3 : Find the probability distribution of number of heads (X) obtained when a fair coin is tossed 4 times.

Sol. : When a coin is tossed 4 times, there are $2^4 = 16$ outcomes which are listed below.

HHHH, HHHT, HHTH, HHTT, HTHH, HTHT, HTTH, HTTT,
THHH, THHT, THTH, THTT, TTHH, TTHT, TTTH, TTTT.

(Write first *H, T, H, T* alternately for 16 times. Then 2 *H*'s and 2 *T*'s alternately, then 4 *H*'s and 4 *T*'s alternatively and lastly 8 *H*'s and 8 *T*'s alternately). Hence, the probability distribution of X is

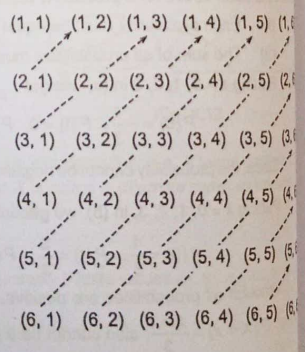
X	:	0	1	2	3	4
$P(X = x)$	:	1 / 16	4 / 16	6 / 16	4 / 16	1 / 16

Example 4 : Write down the probability distribution of the sum of numbers appearing on the toss of two unbiased dice.

Sol. : When two dice are thrown, we get the sum of numbers as shown by slant dotted lines.

See adjoint figure.

It is easy to see that the sum 2 appears once, 3 twice, 4 thrice etc. The probability of each single event is 1/ 36.



The probability distribution obtained is shown below.

X	2	3	4	5	6	7	8	9	10	11	12
$P(X = x)$	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

Example 5 : For the above distribution, (i) find the probability that X is an odd number, (ii) find the probability that X lies between 3 and 9.

Sol. : $P(X \text{ is odd}) = P(X = 3, 5, 7, 9 \text{ or } 11)$

$= P(X = 3) + P(X = 5) + P(X = 7) + P(X = 9) + P(X = 11)$

$\therefore\ P(X \text{ is odd}) = \frac{2}{36} + \frac{4}{36} + \frac{6}{36} + \frac{4}{36} + \frac{2}{36} = \frac{18}{36} = \frac{1}{2}$

$P(3 \le X \le 9) = P(X = 3, 4, 5, 6, 7, 8 \text{ or } 9)$

$= P(X = 3) + P(X = 4) + \ldots\ldots + P(X = 9)$

$= \frac{2}{36} + \frac{3}{36} + \ldots\ldots + \frac{4}{36} = \frac{29}{36}.$

Example 6 : Two unbiased dice are thrown. Let x_1 and x_2 denote the scores obtained and y denote the maximum of them *i.e.* $y = \max.(x_1, x_2)$. Find the probability distribution of Y.

Sol. : Refer to the chart of Ex. 4 above. It is easy to see that y takes values, 2, 3, 4, 5, 6 with probabilities 1/36, 3/36, 5/36, 7/36, 9/36, 11/36 respectively as shown by thick arrows.

Y	1	2	3	4	5	6
$P(Y = y)$	1/36	3/36	5/36	7/36	9/36	11/36

Example 7 : For the above distribution, (i) find the probability that X is an odd number, (ii) find the probability that X lies between 3 and 9.

Sol. : $P(X \text{ is odd}) = P(X = 3, 5, 7, 9 \text{ or } 11)$

$= P(X = 3) + P(X = 5) + P(X = 7) + P(X = 9) + P(X = 11)$

$= \frac{2}{36} + \frac{4}{36} + \frac{6}{36} + \frac{4}{36} + \frac{2}{36} = \frac{18}{36} = \frac{1}{2}$

$P(3 \le X \le 9) = P(X = 3, 4, 5, 6, 7, 8 \text{ or } 9)$

$= P(X = 3) + P(X = 4) + \ldots\ldots + P(X = 9)$

$= \frac{2}{36} + \frac{3}{36} + \ldots\ldots + \frac{4}{36} = \frac{29}{36}.$

Example 8 : The probability mass function of a random variable X is zero except at the points $X = 0, 1, 2$. At these points it has the values $P(0) = 3c^3$, $P(1) = 4c - 10c^2$, $P(2) = 5c - 1$.

(i) Determine c, (ii) Find $P(X < 1)$, $P(1 < X \le 2)$, $P(0 < X \le 2)$.

Sol. : Since $\Sigma\, p_i = 1$, we have, $P(0) + P(1) + P(2) = 1$.

$\therefore\ 3c^3 - 10c^2 + 4c + 5c - 1 = 1 \qquad \therefore\ 3c^3 - 10c^2 + 9c - 2 = 0$

$(3c - 1)(c - 2)(c - 1) \qquad \therefore\ c = 1/3$

(The other values are not admissible. Why?)

$\therefore$ The probability distribution is

X	0	1	2
$P(X = x)$	1/9	2/9	2/3

$\therefore\ P(X<1) = P(X=0) = \frac{1}{9}$; $\quad P(1 < X \le 2) = P(X=2) = \frac{2}{3}$;

$$P(0 < X \le 2) = P(X=1) + P(X=2) = \frac{2}{9} + \frac{2}{3} = \frac{8}{9}.$$

Example 9 : A random variable X has the following probability distribution

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

(i) Find k, (ii) $P\left(\frac{1.5 < X < 4.5}{X > 2}\right)$, (iii) The smallest value of λ for which $P(X \le \lambda) > \frac{1}{2}$.

Sol. : (i) Since $\sum p(x_i) = 1$, we get

$10k^2 + 9k = 1$ $\qquad \therefore\ 10k^2 + 9k - 1 = 0$

$\therefore\ 10k^2 + 10k - k - 1 = 0$ $\qquad \therefore\ 10k(k+1) - 1(k+1) = 0$

$\therefore\ (10k-1)(k+1) = 0$ $\qquad \therefore\ k = 1/10.$ k cannot be -1.

$\therefore$ The probability distribution of X is

X	:	0	1	2	3	4	5	6	7
$P(X=x)$	:	0	1/10	2/10	2/10	3/10	1/100	2/100	17/100

(ii) Now, $P(A/B) = \frac{P(A \cap B)}{P(B)}$

$$\therefore\ P\left(\frac{1.5 < X < 4.5}{X > 2}\right) = \frac{P[(1.5 < X < 4.5) \cap (X > 2)]}{P(X > 2)} = \frac{P(2 < X < 4.5)}{P(X > 2)}$$

$$= \frac{P(X = 3, 4)}{P(X = 3, 4, 5, 6, 7)} = \frac{5/10}{70/100} = \frac{5}{7}.$$

Now from the table we find that

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

$$= 0 + \frac{1}{10} + \frac{2}{10} + \frac{2}{10} = \frac{5}{10} = \frac{1}{2}.$$

Hence, $P(X \le 4) = \frac{8}{10} > \frac{1}{2}$. Hence, $\lambda = 4$.

Example 10 : An urn contains 4 white and 3 red balls. Find the probability distribution of the number of red balls in three draws made successively with replacement from the urn.

Sol. : We get the following probabilities.

No. of red balls	Probability
0	$\frac{4}{7}\cdot\frac{4}{7}\cdot\frac{4}{7}$
1	$\frac{3}{7}\cdot\frac{4}{7}\cdot\frac{4}{7}$, $\frac{4}{7}\cdot\frac{3}{7}\cdot\frac{4}{7}$, $\frac{4}{7}\cdot\frac{4}{7}\cdot\frac{3}{7}$

No. of red balls	Probability
2	$\frac{3}{7}\cdot\frac{3}{7}\cdot\frac{4}{7}$, $\frac{3}{7}\cdot\frac{4}{7}\cdot\frac{3}{7}$, $\frac{4}{7}\cdot\frac{3}{7}\cdot\frac{3}{7}$
3	$\frac{3}{7}\cdot\frac{3}{7}\cdot\frac{3}{7}$

$\therefore$ Probability distribution is

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

Example 11 : A random variable X has the following probability function :

X	:	1	2	3	4	5	6	7
$P(X=x)$	:	k	$2k$	$3k$	k^2	k^2+k	$2k^2$	$4k^2$

Find (i) k, (ii) $P(X<5)$, (iii) $P(X>5)$, (iv) $P\left(\frac{X<5}{2<X\le 6}\right)$, (v) $P\left(\frac{X=4}{3\le X\le 5}\right)$.

Sol. : Since $\sum p(x_i) = 1$,

$$k + 2k + 3k + k^2 + k^2 + k + 2k^2 + 4k^2 = 1$$

$\therefore\ 8k^2 + 7k - 1 = 0$ $\quad \therefore\ (8k-1)(k+1) = 0$

$\therefore\ k = 1/8$ or $k = -1$ which is impossible (why ?)

Thus, we have the following probability distribution.

X	:	1	2	3	4	5	6	7
$P(X=x)$	:	1/8	2/8	3/8	1/64	9/64	2/64	4/64

(i) $P(X<5) = P(X = 1, 2, 3, 4,)$

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

$$= \frac{1}{8} + \frac{2}{8} + \frac{3}{8} + \frac{1}{64} = \frac{49}{64}.$$

(ii) $P(X>5) = P(X = 6, 7) = P(X=6) + P(X=7)$

$$= \frac{2}{64} + \frac{4}{64} = \frac{6}{64} = \frac{3}{32}.$$

(iii) $$P\left(\frac{X<5}{2<X\le 6}\right) = \frac{P(X<5 \cap 2<X\le 6)}{P(2<X\le 6)} = \frac{P(2<X<5)}{P(2<X\le 6)}$$

$$= \frac{P(X = 3, 4)}{P(X = 3, 4, 5, 6)} = \frac{25/64}{36/64} = \frac{25}{36}.$$

(iv) $$P\left(\frac{X=4}{3\le X\le 5}\right) = \frac{P(X=4 \cap 3\le X\le 5)}{P(3\le X\le 5)}$$

$$= \frac{P(X=4)}{P(X = 3, 4, 5)} = \frac{1/64}{34/64} = \frac{1}{64}.$$

Example 12 : The probability of a man hitting the target is 1/4. How many times must he fire so that the probability of his hitting the target at least once is greater than 2/3?

Sol. : We are given that the probability of hitting the target $p = 1/4$.

Probability of not hitting the target, $q = 1 - \frac{1}{4} = \frac{3}{4}$.

$\therefore$ Probability of not hitting the target in n trials

$$= \left(\frac{3}{4}\right)\left(\frac{3}{4}\right).....(n \text{ times}) = \left(\frac{3}{4}\right)^n$$

$\therefore$ Probability of hitting the target at least once in n trials

$$= 1 - \left(\frac{3}{4}\right)^n$$

We want this probability to be greater 2/3.

$$\therefore \quad 1 - \left(\frac{3}{4}\right)^n > \frac{2}{3} \qquad \therefore \quad 1 - \left(\frac{2}{3}\right) > \left(\frac{3}{4}\right)^n$$

$$\therefore \quad \frac{1}{3} > \left(\frac{3}{4}\right)^n \qquad \therefore \quad \left(\frac{3}{4}\right)^n < \frac{1}{3}$$

Taking logarithms of both sides

$$n(\log 3 - \log 4) < -\log 3$$

$$\therefore \quad n(0{\cdot}4771 - 0{\cdot}6021) < -\log 3 \qquad \therefore \quad -n(0{\cdot}1250) < -0{\cdot}4771$$

$$\therefore \quad n > \frac{0{\cdot}4771}{0{\cdot}1250} = 3{\cdot}81 \qquad \therefore \quad n > 4$$

$\therefore$ He must fire at least 4 times, so that he will hit the target at least once, with probability of success greater 2/3.

EXERCISE - I

1. If $P(X = x) = x/25$, $x = 1, 3, 5, 7, 9$, find $P(X = 1 \text{ or } 3)$ and $P(4 < X < 8)$.

[**Ans. : (i)** 4 / 25, **(ii)** 12 / 25]

2. Verify whether the following functions can be considered as p.m.f. and if so find $P(X = 1 \text{ or } 3)$. Give reasons.

(i) $P(X = x) = \frac{1}{5}$, $x = 0, 1, 2, 3, 4$. **(ii)** $P(X = x) = \frac{x^2 + 1}{18}$, $x = 0, 1, 2, 3$.

(iii) $P(X = x) = \frac{x^2 - 2}{8}$, $x = 1, 2, 3$. **(iv)** $P(X = x) = \frac{2x + 1}{18}$, $x = 0, 1, 2, 3$.

[**Ans. : (i)** Yes, 2 / 5 ; **(ii)** Yes, 2 / 3 ; **(iii)** No ; **(iv)** No]

3. If the p.m.f. $P(X = x)$ of a discrete random variate which assumes values x_1, x_2, x_3 such that $P(x_1) = 2P(x_2) = 3P(x_3)$, obtain the probability distribution of X.

[**Ans. : (i)** $P(x_1) = 2/11$, **(ii)** $P(x_2) = 3/11$, **(iii)** $P(x_3) = 6/11$]

4. Presuming the daily demand to be independent, find the probability that over a two-day period the number of requests at a service station will be (i) 9, (ii) 10, if the past record show that the demand was 4, 5 or 6 with probabilities 0·50, 0·40 or 0·10 respectively.

(**Hint :** The event can occur as **(i)** (4, 5), (5, 4); **(ii)** (4, 6), (6, 4), (5, 5).)

5. Find the probability distribution and the probability mass function of the number of points obtained when a fair die is tossed.

[**Ans. : (i)** 0.40, **(ii)** 0·26]

6. Find the probability distribution and the p.m.f. of the number of heads obtained when an unbiased coin is tossed three times.

[**Ans. :** $P(X = x) = \frac{{}^3C_x}{2^3}$, $x = 0, 1, 2, 3$]

7. The probability density function of a random variable X is

X	0	1	2	3	4	5	6
$P(X = x)$	k	$3k$	$5k$	$7k$	$9k$	$11k$	$13k$

Find $P(X < 4)$, $P(3 < X \le 6)$.

[**Ans. :** $k = 1/49, 16/49, 33/49$]

8. A random variable X has the following probability function

X	1	2	3	4	5	6	7
$P(X = x)$	k	$2k$	$3k$	k^2	$k^2 + k$	$2k^2$	$4k^2$

Find **(i)** k, **(ii)** $P(X < 5)$, **(iii)** $P(X > 5)$, **(iv)** $P(0 \le X \le 5)$.

[**Ans. : (i)** $k = 1/8$, **(ii)** 49 / 64, **(iii)** 3 / 32, **(iv)** 29 / 32]

9. A discrete random variable X has the following probability distribution

X	−2	−1	0	1	2	3
$P(X = x)$	0·1	k	0·2	$2k$	0·3	$3k$

Find **(i)** k, **(ii)** $P(X \ge 2)$, **(iii)** $P(-2 < X < 2)$. [**Ans. : (i)** $k = 1/15$, **(ii)** 1 / 2; **(iii)** 2 / 5]

10. Given the following probability function of a discrete random variable X

X	0	1	2	3	4	5	6	7
$P(X = x)$	0	c	$2c$	$2c$	$3c$	c^2	$2c^2$	$7c^2 + c$

(i) Find c, **(ii)** Find $P(X \ge 6)$, **(iii)** $P(X < 6)$, **(iv)** Find k if, $P(X \le k) > 1/2$, where k is a positive integer, **(v)** $P(1{\cdot}5 < X < 4{\cdot}5 / X > 2)$. [**Ans. : (i)** $c = 0{\cdot}1$, **(ii)** 0·19, **(iii)** 0·81, **(iv)** $k = 4$, **(v)** 5 / 7]

11. In the example 9 above, find

(i) $P(-1 \le X \le 1 / -2 \le X \le 3)$, **(ii)** $P(X \le 2 / 0 \le X \le 4)$, **(iii)** $P(X \le 1 / X \le 2)$.

[**Ans. : (i)** $\frac{1}{2}$, **(ii)** $\frac{19}{26}$, **(iii)** $\frac{15}{24}$.]

12. A random variable X takes values −2, −1, 0, 1, 2 such that $P(X > 0) = P(X = 0) = P(X < 0)$; $P(X = -2) = P(X = -1)$, $P(X = 1) = P(X = 2)$. Obtain the probability distribution of X.

Also find (i) $P\left(\frac{-1 \le X \le 1}{-2 \le X \le 0}\right)$, (ii) $P\left(\frac{X = 1}{0 \le X \le 2}\right)$.

[**Ans. :**

X :	−2	−1	0	1	2
$P(X = x)$:	1/6	1/6	1/3	1/6	1/6

(i) 4/5, **(ii)** 1/5]

13. A random variable X takes values 0, 1, 2, ... n with probabilities proportional to ${}^nC_0$, ${}^nC_1$, ${}^nC_2$, ... ${}^nC_n$. Find the proportionality constant.

(**Hint :** $k[{}^nC_0 + {}^nC_1 + {}^nC_2 + ... + {}^nC_n] = 1 \quad \therefore \quad k \cdot 2^n = 1$) [**Ans. :** $k = 2^{-n}$]

14. A random variable X assumes four values with probabilities $(1 + 3x)/4$, $(1 - x)/4$, $(1 + 2x)/4$ and $(1 - 4x)/4$. For what value of x do these values represent the probability distribution of X?

[**Ans. :** $\Sigma p_i = 1$. But $\frac{1+3x}{4} \geq 0$ if $x \geq -\frac{1}{3}$ and also $\frac{1-4x}{4} \geq 0$ if $x \leq \frac{1}{4}$ $\therefore$ $-\frac{1}{3} \leq x \leq \frac{1}{4}$]

15. The amount of bread X (in hundred kgs) that a certain bakery is able to sell in a day is a random variable with probability density function is given by

$$f_X(x) = \begin{cases} Ax, & 0 < x < 5 \\ A(10 - x), & 5 < x < 10 \\ 0, & \text{elsewhere} \end{cases}$$

Find (i) A, (ii) the probabilities of the events : B, the amount of bread sold in a day is more than 500 kgs, C : the amount of bread sold in a day is less than 500 kgs, D : the amount of bread sold in a day is between 250 kgs and 750 kgs. (iii) Are the events B and C exclusive ? (iv) Are the events B and D exclusive. [**Ans. : (i)** $A = 1/25$, **(ii)** $1/2$, $1/2$, 0.75, **(iii)** Yes, **(iv)** No]

8. Distribution Function of a Discrete Random Variable X

Probability distribution of X gives us the probability $p(x)$ that X will take a particular value x. Sometimes we need to know the probability that X will take a value less than or equal to a given value x. This probability is obtained by adding the probabilities of all values less than or equal to x.

Suppose, X is a discrete random variable taking values $x_1, x_2 \ldots\ldots x_n$ with probabilities $p(x_i)$, $i = 1, 2, \ldots\ldots n \ldots\ldots$ such that

(i) $p(x_i) \geq 0$ for all i,

(ii) $\Sigma\, p(x_i) = 1$ and consider the following table.

X	x_1	x_2	x_3	.....	x_n	.....
$F(x_i) = F(X = x_i)$	$p(x_i)$	$\sum_1^2 p(x_i)$	$\sum_1^3 p(x_i)$	.....	$\sum_1^n p(x_i)$	.....

The table states that

$$F(x_1) = P(a \leq X < x) = p(x_1)$$

$$F(x_2) = P(X \leq x_2) = p(a \leq X \leq x_1) + p(x_1 \leq X \leq x_2)$$
$$= p(x_1) + p(x_2)$$

$$F(x_3) = P(X \leq x_3) = p(a \leq X \leq x_1) + p(x_1 \leq X \leq x_2) + p(x_2 \leq X \leq x_3)$$
$$= p(x_1) + p(x_2) + p(x_3).$$

And so on.

The graph shown in the Fig. 4.15 shows this diagrammatically.

The function F is called the **distribution function.** We have more precise definition as follows.

(a) Definition

Let X be a discrete random variable taking values $x_1, x_2, \ldots\ldots$ such that $x_1 < x_2 < x_3 \ldots\ldots$ with probabilities $p(x_1), p(x_2), \ldots\ldots$ such that $p(x_i) \geq 0$ for all i and $\Sigma\, p(x_i) = 1$.

Consider F defined by $\quad F(x_i) = p(X \leq x_i),\ i = 1, 2, 3, \ldots\ldots$

i.e. $\quad F(x_i) = p(x_1) + p(x_2) + \ldots\ldots + p(x_i)$

then the function F is called the **cumulative distribution function** or simply **distribution function** and the set of pairs $\{x_i, F(x_i)\}$ is called the **cumulative probability distribution.**

Note ....

The distribution function is to the probability mass function as cumulative frequency distribution is to frequency distribution.

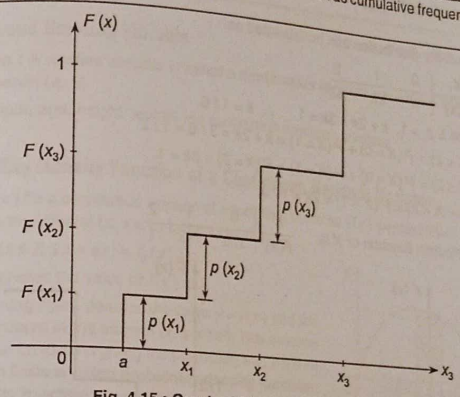


Fig. 4.15 : Graph of Distribution Function.

(b) Important Properties of Distribution function

The distribution function F of a random variable X has the following important properties.

1. $0 \leq F(x) \leq 1$

Proof : Since, $F(x_1) = p(x_1)$, $F(x_2) = p(x_1) + p(x_2)$

$\ldots\ldots\ldots\ldots F(x_n) = p(x_1) + p(x_2) + \ldots\ldots + p(x_n)$

and $0 \leq p(x_i) \leq 1$ for every i and $\Sigma\, p(x_i) = 1$, it is clear that $0 \leq F(x) \leq 1$.

2. $F(x) = 0$ **for** $x < a$ **and** $F(x) = 1$ **for** $x > b$ where $a < x_1 < x_2 < \ldots\ldots x_n < b$.

Proof : If $x < a$ where $a < x_1$, $p(x) = 0$ and hence, $F(x) = 0$ for $x < a$.

If $x > b$ then $F(x) = p(x_1) + p(x_2) + \ldots\ldots + p(x_n) = 1$.

3. $F(x)$ **is a step function**

Proof : By definition, $F(x_1) = F(X < x_1) = p(x_1)$

$F(x_2) = F(X < x_2) = p(x_1) + p(x_2)$

i.e. $F(X)$ has the same value $p(x_1)$ for $x_1 \leq x \leq x_2$,

and the same value $p(x_1) + p(x_2)$ for $x_2 \leq x \leq x_3 \ldots\ldots$

Hence, the graph of $F(x)$ is made up of horizontal line segments taking "jumps" at the possible values x_i of X. The jump is of magnitude $p(x_i) = P(X = x_i)$.

Hence, $F(x)$ is a step functions.

Example 1 : A random variable X has the probability function given below :

$f(x) = k$ if $x = 0$; $f(x) = 2k$ if $x = 1$;

$f(x) = 3k$ if $x = 2$; $f(x) = 0$ otherwise.

(i) Determine the value of k, (ii) Evaluate $P(X < 2)$, $P(X \le 2)$, $P(0 < X < 2)$, (iii) Obtain the distribution function.

Sol. : The probability distribution can be tabulated as

X	0	1	2
$p(x)$	k	$2k$	$3k$

(i) Since $\Sigma\, p_i = 1$, $k + 2k + 3k = 1$ $\therefore$ $k = 1/6$.

(ii) $P(X < 2) = P(X = 0) + P(X = 1) = k + 2k = 3/6 = 1/2$

$P(X \le 2) = P(X = 0) + P(X = 1) + P(X = 2) = 6k = 1$.

$P(0 < X < 2) = P(X = 1) = 2k = 1/3$.

(iii) Distribution function of X is

X	0	1	2
$F(x)$	1/6	1/2	1

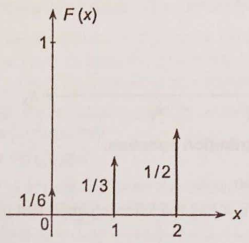


Fig. 4.16 (a)
Probability Density Function.

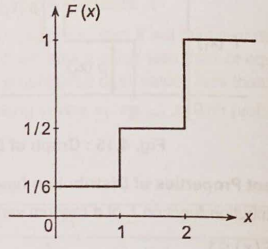


Fig. 4.16 (b)
Distribution Function.

EXERCISE - II

1. A random variable takes values 1, 2, 3, 4 such that $2P(X = 1) = 3P(X = 2) = P(X = 3) = 5P(X = 4)$. Find the probability distribution and the cumulative distribution function.

[**Ans. :**

x	1	2	3	4
$P(x)$	15/61	10/61	30/61	6/61
$F(x)$	15/61	25/61	55/61	1

2. A shipment of 8 computers contains 3 that are defective. If a college makes a random purchase of 2 of these computers, find the probability distribution of the defective computers. Find also distribution function.

[**Ans. :**

x	0	1	2
$P(x)$	10/28	15/28	3/28
$F(x)$	10/28	25/28	1

3. The probability density function of a random variable X is

X	:	0	1	2	3	4	5	6
$P(X = x)$	:	k	$3k$	$5k$	$7k$	$9k$	$11k$	$13k$

Find $P(X < 4)$, $P(3 < X \le 6)$.

[**Ans. :** $k = 1/49, 16/49, 33/49$]

9. Continuous Random Variable

Definition : A random variable is called a **continuous random variable** if it takes all values between an interval (a, b).

For example, age, height, weight are continuous random variables.

10. Probability Density Function of a Continuous Random Variable

Let $y = f(x)$ be a continuous function of x such that the area $f(x)\,\delta x$ represents the probability that X will lie in the interval $(x, x + \delta x)$. Symbolically,

$$P(x \le X \le x + \delta x) = f_x(x)\,\delta x$$

where, $f_x(x)$ denotes the value of $f(x)$ at x.

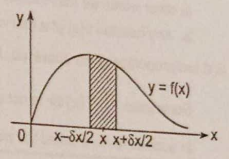


Fig. 4.17

The adjoining figure denotes the curve $y = f(x)$ and the area under the curve in the interval $(x, x + \delta x)$. The function satisfying certain conditions giving the probability that x will lie between certain limits is called **probability density function**, or simply **density function** of a continuous random variable X and is abbreviated as p.d.f. The curve given by $y = f(x)$ is called the **probability density curve** or simply **probability curve**. The expression $f(x)\,dx$ is usually denoted by $df(x)$ and is known as **probability differential**.

Definition : A continuous function $y = f(x)$ such that

(i) $f(x)$ is integrable. (ii) $f(x) \ge 0$

(iii) $\int_a^b f(x)\,dx = 1$ if X lies in $[a, b]$ and

(iv) $\int_\alpha^\beta f(x)\,dx = P(\alpha \le X \le \beta)$ where $a < \alpha < \beta < b$

is called **probability desnity function** of a continuous random variable X.

Thus, for a continuous random variale X,

$$\boxed{P(\alpha \le X \le \beta) = \int_\alpha^\beta f(x)\,dx}$$

Clearly $\int_\alpha^\beta f(x)\,dx$ represents the area under the curve $y = f(x)$, the x-axis and the ordinates at $x = \alpha$ and $x = \beta$. Further since, the total probability is one, if X lies in the interval $[a, b]$ then $\int_a^b f(x)\,dx = 1$. For a continuous random variable X the range may be finite $[a, b]$ or infinite $[-\infty, \infty]$.

Properties of Probability Density Function

The probability density function $f(x)$ has the following properties.

(i) $f(x) \geq 0, -\infty < x < \infty$ (*i.e.* the curve $y = f(x)$ lies above the x-axis in the first and second quadrants only)

(ii) $\int_{-\infty}^{\infty} f(x)\,dx = 1$ (*i.e.* the total area under the curve and the x-axis is one.)

(iii) The probability that $\alpha \leq X \leq \beta$ is given by $P(\alpha \leq X \leq \beta) = \int_{\alpha}^{\beta} f(x)\,dx$.

Notes ....

1. The property (1) and the property (2) can be used to verify whether a given function $f(x)$ can be a probability density function.

2. You know that for A discrete random variable the probability at $X = c$ may not be zero. But, in a continuous random variable $P(X = c)$ is always zero because $P(X = c) = \int_c^c f_X(x)\,dx$ and this definite integral is zero. Hence, for a continuous random variable X,

$$P(\alpha \leq X \leq \beta) = P(\alpha < X < \beta) = P(\alpha < X \leq \beta) = P(\alpha < X < B)$$

In other words we may include or may not include the end-points in the interval.

3. Any function $f(x)$ of a real variable x can be a probability density function if it satisfies the first two properties given above *viz.* $f(x)$ is non-negative for all value of x and $\int_{-\infty}^{\infty} f(x)\,dx = 1$.

Sometimes $\int_{-\infty}^{\infty} f(x)\,dx$ is not equal to 1 but $\int_{-\infty}^{\infty} k f(x) = 1$ for some value of k.

In such cases k is called the ***normalising factor*** or ***normalisation constant.***

Example 1 : Find the normalising factor, k if the following function is a probability density function

$$f(x) = \begin{cases} k(1 - x^2) & 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

Also find $P(0{\cdot}1 < X < 0{\cdot}2)$ and $P(X > 0{\cdot}5)$.

Sol. : Since $0 < x < 1$, $f(x) \geq 0$ for all x.

Now, $\int_0^1 f(x)\,dx = k\int_0^1 (1 - x^2)\,dx = k\left[x - \frac{x^3}{3}\right]_0^1 = k\frac{2}{3}$

But this must be equal to 1.

$$\therefore \frac{2k}{3} = 1 \quad \therefore k = \frac{3}{2}$$

(i) $P(0{\cdot}1 < X < 0{\cdot}2) = \int_{0{\cdot}1}^{0{\cdot}2} \frac{3}{2}(1 - x^2)\,dx = \frac{3}{2}\left[x - \frac{x^3}{3}\right]_{0{\cdot}1}^{0{\cdot}2}$

$$= \frac{3}{2}\left[\left\{0{\cdot}2 - \frac{(0{\cdot}2)^3}{3}\right\} - \left\{0{\cdot}1 - \frac{(0{\cdot}1)^3}{3}\right\}\right]$$

$$= \frac{3}{2}[0{\cdot}0977] = 0{\cdot}146$$

(ii) $P(X > 0{\cdot}5) = \int_{0{\cdot}5}^{1} \frac{3}{2}(1 - x^2)\,dx = \frac{3}{2}\left[x - \frac{x^3}{3}\right]_{0{\cdot}5}^{1}$

$$= \frac{3}{2}\left[\left\{1 - \frac{1}{3}\right\} - \left\{0{\cdot}5 - \frac{(0{\cdot}5)^3}{3}\right\}\right]$$

$$= \frac{3}{2}[0{\cdot}667 - 0{\cdot}458] = 0{\cdot}313$$

Example 2 : A continuous random variable X has the following probability law

$$f(x) = kx^2, \quad 0 \leq x \leq 2$$

Determine k and find the probabilities that (i) $0.2 \leq X \leq 0.5$, (ii) $X \geq 3/4$ given that $X \geq 1/2$.

Sol. : Since the total probability *i.e.*, the total area is unity

$$\int_a^b f(x)\,dx = \int_0^2 kx^2 dx = 1$$

$$k\left[\frac{x^3}{3}\right]_0^2 = 1 \quad \therefore k \cdot \frac{8}{3} = 1 \quad \therefore k = \frac{3}{8}$$

(i) $P(0{\cdot}2 \leq X \leq 0{\cdot}5) = \frac{3}{8}\int_{0{\cdot}2}^{0{\cdot}5} x^2 dx = \frac{3}{8}\left[\frac{x^3}{3}\right]_{0{\cdot}2}^{0{\cdot}5}$

$$= \frac{1}{8}\left[0{\cdot}5^3 - 0{\cdot}3^3\right] = 0{\cdot}0123$$

(ii) Let $A = (X \geq 1/2)$, $B = (X \geq 3/4)$

$$\therefore P(A) = P(X \geq 1/2) = \frac{3}{8}\int_{0{\cdot}5}^{2} x^2 dx = \frac{3}{8}\left[\frac{x^3}{3}\right]_{0{\cdot}5}^{2}$$

$$= \frac{1}{8}\left[2^3 - 0{\cdot}5^3\right] = 0{\cdot}984$$

$$P(B) = P(X \geq 3/4) = \frac{3}{8}\int_{0{\cdot}75}^{2} x^2 dx = \frac{3}{8}\left[\frac{x^3}{3}\right]_{0{\cdot}75}^{2}$$

$$= \frac{1}{8}\left[2^3 - 0{\cdot}75^3\right] = 0{\cdot}947$$

$$P(A \cap B) = P(B) = 0{\cdot}947$$

$$\therefore P(B/A) = \frac{P(A \cap B)}{P(A)} = \frac{0{\cdot}947}{0{\cdot}984} = 0{\cdot}96$$

Example 3 : Let X be a continuous random variable with probability distribution

$$p(x) = \begin{cases} \frac{x}{6} + k & \text{if } 0 \leq x \leq 3 \\ 0 & \text{elsewhere} \end{cases}$$

Evaluate k and find $P(1 \leq x \leq 2)$.

Sol. : Since the total probability is one

$$\int_{-\infty}^{\infty} p(x)\,dx = \int_0^3 \left(\frac{x}{6}+x\right)dx = \left[\frac{x^2}{12}+kx\right]_0^3 = \frac{3}{4}+3k = 1$$

$$\therefore\ 3\left(\frac{1}{4}+k\right)=1 \quad \therefore\ \frac{1}{4}+k=\frac{1}{3} \quad \therefore\ k=\frac{1}{3}-\frac{1}{4}=\frac{1}{12}$$

$$\therefore\ p(x)=\begin{cases}\dfrac{x}{6}+\dfrac{1}{12} & \text{if } 0\le x\le 3\\ 0 & \text{elsewhere}\end{cases}$$

$$\therefore\ P(1\le x\le 2)=\int_1^2\left(\frac{x}{6}+\frac{1}{12}\right)dx=\left[\frac{x^2}{12}+\frac{x}{12}\right]_1^2$$

$$=\frac{1}{12}[(4+2)-(1+1)]=\frac{1}{12}(4)=\frac{1}{3}.$$

Example 4 : Let X be a continuous random variable with p.d.f. $f(x)=kx(1-x)$, $0\le x\le 1$. Find k and determine a number b such that $P(X\le b)=P(X\ge b)$.

Sol. : Since $\int_{-\infty}^{\infty} f(x)\,dx=1$, we have

$$k\int_0^1 (x-x^2)\,dx=1 \quad \therefore\ k\left[\frac{x^2}{2}-\frac{x^3}{3}\right]_0^1=1$$

$$\therefore\ k\left[\frac{1}{2}-\frac{1}{3}\right]=1 \quad \therefore\ k=6.$$

Since, the total probability is 1 and $P(x\le b)=P(x\ge b)$, $P(x\le b)=1/2$.

$$\therefore\ \int_0^b f(x)\,dx=\frac{1}{2}$$

$$\therefore\ 6\int_0^b (x-x^2)\,dx=\frac{1}{2} \quad \therefore\ \left[\frac{x^2}{2}-\frac{x^3}{3}\right]_0^b=\frac{1}{12}$$

$$\therefore\ \frac{b^2}{2}-\frac{b^3}{3}=\frac{1}{12} \quad \therefore\ 6b^2-4b^3=1 \quad \therefore\ 4b^3-6b^2-1=0$$

$$\therefore\ 4b^3-2b^2-4b^2-2b+2b-1=0$$

$$\therefore\ (2b-1)(2b^2-2b+1)=0 \quad \therefore\ b=1/2.$$

Example 5 : The probability that a person will die in the time interval (t_1, t_2) is given by

$$P(t_1\le t\le t_2)=\int_{t_1}^{t_2} f(t)\,dt$$

where, $$f(t)=\begin{cases}3\times 10^{-9}(100\,t-t^2)^2, & 0<t<100\\ 0, & \text{elsewhere.}\end{cases}$$

Find (i) the probability that Mr. X will die between the ages 60 and 70. (ii) the probability that he will die between the ages 60 and 70, given that he has survived upto age 60.

Sol. : (i) $$P(60\le t\le 70)=\int_{60}^{70} 3\times 10^{-9}(100\,t-t^2)^2\,dt$$

$$=3\times 10^{-9}\int_{60}^{70}(100^2\,t^2-200\,t^3+t^4)\,dt$$

$$=3\times 10^{-9}\left[100^2\frac{t^3}{3}-200\frac{t^4}{4}+\frac{t^5}{5}\right]_{60}^{70}$$

$$=3\times 10^{-9}\left[27{\cdot}89\times 10^8-22{\cdot}75\times 10^8\right]$$

$$=0{\cdot}1542.$$

(ii) $$P\left(\frac{60\le T\le 70}{T\ge 60}\right)=\frac{P(60\le t\le 70\cap t\ge 60)}{P(t\ge 60)}=\frac{P(60\le t\le 70)}{P(60\le t\le 100)}$$

$$=\frac{\int_{60}^{70} f(t)\,dt}{\int_{60}^{100} f(t)\,dt}=\frac{0{\cdot}1542}{0{\cdot}3174}=0{\cdot}4858.$$

EXERCISE - III

1. A function is defined as

$$f(x)=\begin{cases}0 & \text{for } x<2\\ \dfrac{2x+3}{18} & \text{for } 2\le x\le 4\\ 0 & \text{for } x>4\end{cases}$$

Show that $f(x)$ is a probability density function and find the probability that $2<x<3$. [**Ans. :** 5 / 18]

2. A random variable X has the probability density function

$$f(x)=\begin{cases}2e^{-2x} & \text{for } x\ge 0\\ 0 & \text{for } x<0\end{cases}$$

Find $P(1\le X\le 3)$, $P(X\ge 0{\cdot}5)$. [**Ans. :** (i) 0·133, (ii) 0·368]

3. A continuous random variable X has the following probability density function

$$f(x)=\begin{cases}kx & 0\le x\le 1\\ k & 1\le x\le 2\end{cases}$$

Find (i) the value of k, (ii) $P(x\le 1{\cdot}5)$. [**Ans. :** (i) $k=2/3$, (ii) 2 / 3]

4. A continuous random variable has the following probability density function.

$$f(x)=\begin{cases}(x/4)+k & 0\le x\le 2\\ 0 & \text{elsewhere}\end{cases}$$

Evaluate k and $P(1\le X\le 2)$. [**Ans. :** $k=1/4$, 5 / 8]

5. Find the value of k such that the following will be the probability density function. Find also $P(x\le 1{\cdot}5)$.

$$f(x)=\begin{cases}kx & 0\le x\le 1\\ k & 1\le x\le 2\\ k(3-x) & 2\le k\le 3\end{cases}$$

[**Ans. :** $k=\frac{1}{2}, \frac{1}{2}$]

11. Continuous Distribution Function

Probability distribution of X or the probability density function of X helps us to find the probability that X will be within a given interval $[a, b]$ *i.e.* $P(a \le X \le b) = \int_a^b f(x)\,dx$, other conditions being satisfied.

However, sometimes we need to know that probability that X will be less than a given value x. For a continuous random variable X, this probability is obtained by integrating $f(x)$ from $-\infty$ (or the lower limit of the interval) to x. The function so obtained is called **distribution function.**

Definition : If X is a continuous random variable X, having the probability density function $f(x)$ then the function

$$F(x) = P(X \le x) = \int_{-\infty}^{x} f(t)\,dt, \quad -\infty < x < \infty$$

is called **distribution function** or **cumulative distribution function** of the random variable X.

Some Important Properties of Distribution Function $F(x)$ of a Continuous Random Variable

1. The function $F(x)$ is defined for every real number x.
2. Since $F(x)$ denotes probability and probability of X lies between 0 and 1,
   $0 \le F(x) \le 1$.
3. $F(x)$ is a non-decreasing function which means if $x_1 \le x_2$, then $F(x_1) \le F(x_2)$.
4. **The derivative of $F(x)$** *i.e.* $F'(x)$ exists at all points (except perhaps at a finite number of points) and **is equal to the probability density function $f(x)$.**

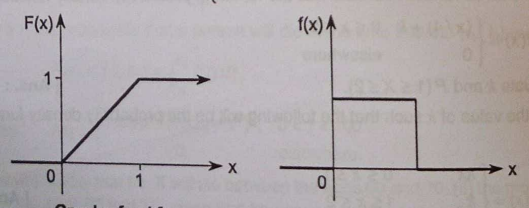
   $$F'(x) = \frac{d}{dx}F(x) = f(x) \ge 0 \text{ provided the derivative exists.}$$
5. If $F(x)$ is a distribution function of a continuous random variable then
   $P(a \le X \le b) = F(b) - F(a)$.

Example 1 : A continuous variable X has the following distribution function

$$F(x) = \begin{cases} 0, & x \le 0 \\ x, & 0 \le x \le 1 \\ 1, & 1 \le x \end{cases} \quad \text{.....(}$$

Find the probability density function and draw the graphs of both p.d.f. and c.d.f.

Sol. : The p.d.f. is $\quad f(x) = F'(x) = \begin{cases} 0 & x < 0 \\ 1 & 0 \le x \le 1 \\ 0 & 1 < x \end{cases}$



Graph of c.d.f. Graph of p.d.f.

Fig. 4.18

From the graph of $F(x)$ we see that $F(x)$ is continuous at all points including $x = 0$ and $x = 1$. $f(x)$ is obtained by differentiating $F(x)$.

Example 2 : For the distribution function given below, find p.d.f.

$$F(x) = \begin{cases} 0 & , \quad x < 0 \\ 1 - e^{-x/4} & , \quad x \ge 0 \end{cases}$$

Also find the probabilities : $P(X \le 4)$, $P(X \ge 8)$, $P(4 \le X \le 8)$.

Sol. : $F(x)$ satisfies all the conditions of a distribution function. If $f(x)$ is the corresponding probability density function

$$f(x) = F'(x) = \begin{cases} \frac{1}{4}e^{-x/4}, & x \ge 0 \\ 0 \quad , & x < 0 \end{cases}$$

Now, we have to verify that $\int_{-\infty}^{\infty} f(x)\,dx = 1$.

$$\therefore \quad \int_{-\infty}^{0} f(x)\,dx + \int_0^{\infty} \frac{1}{4}e^{-x/4}\,dx = 0 + \int_0^{\infty} \frac{1}{4}e^{-x/4}dx = \frac{1}{4}\left[\frac{e^{-x/4}}{-1/4}\right]_0^{\infty}$$

$$= -\left[e^{-x/4}\right]_0^{\infty} = -[0 - 1] = 1$$

Hence, $F(x)$ is a distribution function.

Now, $\quad P(X \le 4) = F(4) = 1 - e^{-1} = 1 - \frac{1}{e} = \frac{e-1}{e}$

$$P(X \ge 8) = 1 - P(x \le 8) = 1 - F(8)$$

$$= 1 - \left[1 - e^{-2}\right] = e^{-2} = 1/e^2$$

$$P(4 \le X \le 8) = F(8) - F(4) = (1 - e^{-2}) - (1 - e^{-1})$$

$$= e^{-1} - e^{-2} = \frac{1}{e} - \frac{1}{e^2} = \frac{e-1}{e^2}$$

Example 3 : If $f(x) = \begin{cases} x e^{-x^2/2} & x \ge 0 \\ 0 & x < 0 \end{cases}$

(i) Show that $f(x)$ is a probability density function. (ii) Find its distribution function.

Sol. : If $f(x)$ is a p.d.f., we must have $f(x) \ge 0$ where $x \ge 0$ and $\int_0^{\infty} f(x)\,dx = 1$.

Clearly $f(x) = xe^{-x^2/2} \ge 0$ for $x \ge 0$.

Now, $\quad \int_0^{\infty} x e^{-x^2/2}\,dx = \int_0^{\infty} e^{-t}\,dt \qquad \left[t = x^2/2\right]$

$$= \left[-e^{-t}\right]_0^{\infty} = -[0 - 1] = 1$$

$\therefore$ $f(x)$ is a probability density function.

Now, its distribution function is given by

$$F(x) = \int_0^x f(x)\,dx = \int_0^x x\,e^{-x^2/2}\,dx = -\left[e^{-x^2/2}\right]_0^x \qquad \text{[As above]}$$

$$= -\left[e^{-x^2/2} - 1\right] = 1 - e^{-x^2/2}, \quad x \ge 0.$$

EXERCISE - IV

1. The distribution function of a random variable X is given by

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

Find the probability density function. Draw the graphs of both p.d.f. and c.d.f.

2. The c.d.f. of a continuous random variable X is given by

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

Find the p.d.f. Draw the graphs of both p.d.f. and c.d.f.

Also find $P\left(\frac{1}{2} \le X \le \frac{4}{5}\right)$. [Ans. : 0·195

3. Find the distribution functions corresponding to the following probability density functions

(i) $f(x) = \begin{cases} \frac{1}{2}x^2 e^{-x}, & 0 \le x < \infty \\ 0, & \text{otherwise} \end{cases}$ (ii) $f(x) = \begin{cases} x, & 0 \le x \le 1 \\ 2 - x, & 1 \le x \le 2 \\ 0, & \text{otherwise} \end{cases}$

(iii) $f(x) = \begin{cases} \lambda(x-1)^4, & 1 \le x \le 3, \lambda > 0 \\ 0, & \text{otherwise} \end{cases}$

[Ans. : (i) $F(x) = \begin{cases} 1 - e^{-x}\left(1 + x + \frac{x^2}{2}\right), & x \ge 0 \\ 0, & \text{otherwise} \end{cases}$

(ii) $F(x) = \begin{cases} 0, & x < 0 \\ x^2/2, & 0 \le x \le 1 \\ 2x - 0{\cdot}5x^2 - 1, & 1 \le x \le 2 \\ 1, & x > 2 \end{cases}$

(iii) $\lambda = \frac{5}{32}$; $F(x) = \begin{cases} 0, & x \le 1 \\ \frac{(x-1)^5}{32}, & 1 \le x \le 3 \\ 1, & x \ge 3 \end{cases}$]

4. A continuous random variable X has the following probability density function

$$f(x) = \frac{a}{x^5}, \quad 2 \le x \le 10$$

Determine the constant a, distribution function of X and find the probability of the event $4 \le x \le 7$.

[Ans. : (i) $a = \frac{2500}{39}$, (ii) $\frac{625}{39}\left[\frac{1}{16} - \frac{1}{x^4}\right]$, (iii) 0·056]

5. The distribution function of a random variable X is given by

$$F(x) = \begin{cases} 1 - (1+x)e^{-x}, & x \ge 0 \\ 0, & x < 0 \end{cases}$$

Find the probability density function and find $P(0 \le X \le 1)$.

[Ans. : (i) $f(x) = x e^{-x}$, (ii) $P(0 \le X \le 1) = F(1) - F(0) = \frac{e-2}{e}$]

EXERCISE - V

Theory

1. Define the following terms giving suitable examples
   (a) Random variable
   (b) Discrete random variable
   (c) Continuous random variable
   (d) Probability density function
   (e) Distribution function
2. State the properties of probability density function.
3. State the properties of distribution function.

☆☆☆

CHAPTER 5

Mathematical Expectation

1. Introduction

Suppose two coins are tossed twenty times. Let X be the number of heads obtained in a toss. X then, takes values 0, 1 and 2. Suppose further that no heads, one head and two heads were obtained 4, 10, 6 times respectively. Then, the average number of heads per toss

$$= \frac{4(0) + 10(1) + 6(2)}{6 + 10 + 4} = 1{\cdot}1$$

This is the average value and is not necessarily a possible outcome of the toss.

The ratios 4 / 20, 10 / 20, 6 / 20 of 0, 1, 2 heads to the total number of tosses are the relative frequencies of $X = 0, 1, 2$. If the experiment is repeated very large number of times, we know that these relative frequencies tend to the probabilities 1 / 4, 1 / 2, 1 / 4 of 0, 1, 2 heads because in the toss of two coins we have the following.

Sample space	:	HH	$\underbrace{HT, TH}$	TT
Probability	:	1/4	1/2	1/4

The average calculated with probabilities in place of relative frequencies above is called **expected value** or **mathematical expectation** and is denoted by $E(X)$. Thus,

$$E(X) = \frac{1}{4}(0) + \frac{1}{2}(1) + \frac{1}{4}(2) = 1$$

$$E(X) = \text{sum of the products of the values and their probabilities}$$

$$= p_1 x_1 + p_2 x_2 + p_3 x_3 + \ldots\ldots$$

This means, a person who throws two coins over and over again will get one head per toss on the average. This suggests us that the expected value of X can be obtained by multiplying the values of X by their respective probabilities and taking the sum. This leads us to the following definition of expectation of a discrete random variable X.

2. Expectation of a Random Variable

(a) Definition : If a **discrete random variable** X assumes values $x_1, x_2, \ldots\ldots x_n \ldots\ldots$ with probabilities $p_1, p_2, \ldots\ldots p_n \ldots\ldots$ respectively then the **mathematical expectation of** X denoted by $E(X)$ (if it exists) is defined by

$$\boxed{E(X) = p_1 x_1 + p_2 x_2 + \ldots\ldots + p_n x_n + \ldots\ldots}$$

i.e. $\boxed{E(X) = \Sigma\, p_i x_i}$ where $\Sigma\, p_i = 1$.

if $\Sigma\, p_i x_i$ is absolutely convergent.

This value is also referred to as *mean value* of X. It is also denoted by μ_1'. $\therefore\ \mu_1' = E(X)$

Notation : In this chapter we shall slightly deviate from our previous notation. Instead of denoting $P(X = x_i)$ by $p(x_i)$ we shall denote it simply by p_i. This will be found more convenient while dealing with expectations.

(b) Definition : Let X be a **continuous random variable** with probability density function $f(x)$. Then the **mathematical expectation of** X, denoted by $E(X)$ (if it exists), is defined by

$$\boxed{E(X) = \int_{-\infty}^{\infty} x \cdot f(x)\, dx} \quad \text{where, } \int_{-\infty}^{\infty} f(x)\, dx = 1$$

if the integral is absolutely convergent.

Notes ....

1. If X assumes only a finite number of values then $E(X) = \Sigma\, p_i x_i$ and can be considered as "weighted average" of the values $x_1, x_2, \ldots\ldots x_n$ with weights $p_1, p_2 \ldots\ldots p_n$.
2. If all values $x_1, x_2, \ldots\ldots x_n$ are equiprobable *i.e.* $p_1 = p_2 = \ldots\ldots = p_n = 1/n$ then $E(X) = (1/n)\, \Sigma\, x_i$ and can be seen to be simple arithmetic mean of the n values $x_1, x_2, \ldots\ldots x_n$.
3. One should guard oneself from being mislead by the term 'expectation'. $E(X)$ does *not* give us the value of X, we can expect in a single trial. In the Example 4 below $E(X) = 7/2$ is not even a possible value of X when a die is tossed.

   $E(X)$ denotes *mathematical* expectation of X in the sense that if we toss a die for a fairly large number of times, observe the frequencies of the outcomes 1, 2, 3, 4, 5, 6 then the average of these values will be closer to 7 / 2 the more often the die were tossed.
4. $E(X)$ is expressed in the *same units* as X.
5. **Expectation of a constant is constant**

   (i) $E(c) = \Sigma\, p_i c = c\, \Sigma\, p_i = c \quad [\because\ \Sigma\, p_i = 1]$

   (ii) $E(c) = \int_{-\infty}^{\infty} c\, f(x)\, dx = c \int_{-\infty}^{\infty} f(x)\, dx = c \quad \left[\because \int_{-\infty}^{\infty} f(x)\, dx = 1\right]$

Example 1 : A fair coin is tossed 3 times. A person received ₹ X^2 if he gets X heads. Find his expectation.

Sol. : When a coin is tossed three times, the sample space is

$$HHH, HHT, HTH, HTT, THH, THT, TTH, TTT$$

The probability distribution of X is

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

Now, X^2 takes the following values

X^2	:	0	1	2	3
$P(X^2)$	:	1/8	3/8	3/8	1/8

$$\therefore \quad E(X) = \Sigma\, p_i x_i = \frac{1}{8} \times 0 + \frac{3}{8} \times 1 + \frac{3}{8} \times 4 + \frac{1}{8} \times 9$$

$$= \frac{3 + 12 + 9}{8} = \frac{24}{8} = 3 \text{ ₹}$$

Example 2 : A fair coin is tossed till a head appears. What is the expectation of the number of tosses required ?

Sol. : Let X denote the order of the toss at which we get the first head. We have

Event	H	TH	TTH	$TTTH$	.....
X	1	2	3	4	.....
$P(X=x)$	$\frac{1}{2}$	$\frac{1}{2}\cdot\frac{1}{2}=\frac{1}{4}$	$\frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2}=\frac{1}{8}$	$\frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2}=\frac{1}{16}$	.....

$$\therefore \quad E(X)=\Sigma p_i x_i = 1\left(\frac{1}{2}\right)+2\left(\frac{1}{4}\right)+3\left(\frac{1}{8}\right)+4\left(\frac{1}{16}\right)+\ldots..$$

Let
$$S=\frac{1}{2}+2\cdot\frac{1}{4}+3\cdot\frac{1}{8}+4\cdot\frac{1}{16}+\ldots..$$

$$\frac{1}{2}S= \quad \frac{1}{4}+2\cdot\frac{1}{8}+3\cdot\frac{1}{16}+\ldots..$$

$$\therefore \quad S-\frac{1}{2}S=\frac{1}{2}+\frac{1}{4}+\frac{1}{8}+\frac{1}{16}+\ldots..$$

$$\therefore \quad \frac{1}{2}S=\frac{1}{2}\cdot\frac{1}{1-(1/2)}=1 \qquad \left[\text{G.P. } S_\infty=\frac{a}{1-r}\right]$$

$$\therefore \quad E(X)=2.$$

Example 3 : A box contains n tickets numbered 1, 2, ..., n. If m tickets are drawn at random from the box. What is the expectation of the sum of the numbers on the tickets drawn ?

Sol. : Let X_i denote the number on the i th ticket drawn.

Then $S = X_1 + X_2 + \ldots.. + X_m$

Now, $E(X_i)=\Sigma p_i x_i=\frac{1}{n}\cdot 1+\frac{1}{n}\cdot 2+\ldots..+\frac{1}{n}\cdot n$

$$=\frac{1}{n}(1+2+\ldots..+n)=\frac{1}{n}\cdot\frac{n}{2}\cdot(n+1)=\frac{n+1}{2}$$

$$\therefore \quad E(S)=\sum_{i=1}^{m}E(X_i)=\frac{m(n+1)}{2}.$$

Example 4 : Three urns contain respectively 3 green and 2 white balls, 5 green and 6 white balls, 2 green and 4 white balls. One ball is drawn from each urn. Find the expected number of white ball drawn.

Sol. : If X denotes the number of white balls drawn from an urn then the expectation of X is as follows.

$$E(X)=p_i x_i = p_1 x_1 + p_2 x_2$$

x takes two values. $X = 1$ if white ball is drawn and $X = 0$ if greenball is drawn.

$\therefore$ From first urn : $\quad E(X_1)=1\times\frac{2}{5}+0\times\frac{3}{5}=\frac{2}{5}$

From second urn : $\quad E(X_2)=1\cdot\frac{6}{11}+0\cdot\frac{5}{11}=\frac{6}{11}$

$$E(X_3)=1\cdot\frac{4}{6}+0\cdot\frac{2}{6}=\frac{2}{3}$$

$\therefore$ The required expectation $= E(X_1) + E(X_2) + E(X_3)$

$$=\frac{2}{5}+\frac{6}{11}+\frac{2}{3}=\frac{266}{165}=1{\cdot}61$$

Example 5 : The daily consumption of electric power (in million kwh) is a random variable X with probability distribution function

$$f(x)=\begin{cases} kx\,e^{-x/3} & \text{for } x>0 \\ 0 & \text{for } x\le 0\end{cases}$$

Find the value of k, the expectation of x and the probability that on a given day the electric consumption is more than expected value.

Sol. : We must have

$$\int_{-\infty}^{\infty} f(x)\,dx = 1 \quad i.e., \quad k\int_0^{\infty} x\,e^{-x/3}\,dx = 1$$

$$\therefore \quad k\left[(x)\left(\frac{e^{-x/3}}{-1/3}\right)-(1)\left(\frac{e^{-x/3}}{1/9}\right)\right]_0^{\infty}=1$$

$$\therefore \quad k[0+9]=1 \quad \therefore \quad 9k=1 \quad \therefore \quad k=1/9$$

$$E(X)=\int_{-\infty}^{x} x\,f(x)\,dx=\frac{1}{9}\int_0^{\infty}x^2\cdot e^{-x/3}\cdot dx$$

$$=\frac{1}{9}\left[x^2\left(\frac{e^{-x/3}}{-1/3}\right)-(2x)\left(\frac{e^{-x/3}}{1/9}\right)+2\left(\frac{e^{-x/3}}{-1/27}\right)\right]_0^{\infty}$$

$$=\frac{1}{9}[0-0+0+54]=6$$

$$\therefore \quad P(X>6)=\frac{1}{9}\int_6^{\infty}x\cdot e^{-x/3}\cdot dx$$

$$\therefore \quad P(X>6)=\frac{1}{9}\left[(x)\left(\frac{e^{-x/3}}{-1/3}\right)-(1)\left(\frac{e^{-x/3}}{1/9}\right)\right]_6^{\infty}$$

$$=\frac{1}{9}\left[(0-0)-(-18\,e^{-2}-9e^{-2})\right]$$

$$=3\,e^{-2}=0{\cdot}406$$

Example 6 : Find k and then $E(X)$ if X has the p.d.f.

$$f(x)=\begin{cases} k\,x(2-x), & 0\le x\le 2,\ k>0 \\ 0, & \text{elsewhere}\end{cases}$$

Sol. : Now $\int_0^2 kx(2-x)\,dx = k\int_0^2 2x-x^2\,dx = k\left[x^2-\frac{1}{3}x^3\right]_0^2 = k\cdot\frac{4}{3}$

$$\therefore \quad k\cdot\frac{4}{3}=1 \quad \therefore \quad k=\frac{3}{4}$$

By definition

$$E(X) = \int_0^2 x\, f(x)\, dx = \int_0^2 x \cdot \frac{3}{4} x(2-x)\, dx = \frac{3}{4}\int_0^2 (2x^2 - x^3)\, dx$$

$$\therefore \quad E(X) = \frac{3}{4}\left[\frac{2x^3}{3} - \frac{x^4}{4}\right]_0^2 = \frac{3}{4}\left[\frac{16}{3} - \frac{16}{4}\right] = \frac{3}{4} \cdot \frac{16}{12} = 1.$$

EXERCISE - I

1. From an urn containing 3 red balls and 2 white balls, a man is to draw 2 balls at random without replacement. He gets ₹ 20 for each red ball and ₹ 10 for each white ball he draws. Find his expectation. [**Ans. :** ₹ 32]
2. Two urns contain respectively 5 white and 3 black balls; 2 white and 3 black balls. One ball is drawn from each urn. Find the expected number of white balls drawn. [**Ans. :** 41 / 40]
3. A and B toss a fair coin alternately. One who gets a head first wins ₹ 12. A starts. Find their mathematical expectations. [**Ans. :** ₹ 8, 4]
4. A, B, and C toss a fair coin. The first one to throw a head wins the game and gets ₹ 28. If A starts, find their mathematical expectations. [**Ans. :** ₹ 16, 8, 4]
5. A, B and C draw a card in that order from a well shuffled pack of 52 cards. The first to draw a diamond wins ₹ 740. If A starts, find their expectations. [**Ans. :** ₹ 320, 240, 180]
6. Two unbiased dice are thrown. Find the expectation of the sum. [**Ans. :** 7]
7. A man with n keys in his pocket wants to open the door of his case by trying the keys independently and randomly one by one. Find the mean and the variance of the number of trials required to open the door if unsuccessful keys are kept aside.

   [**Ans. : (i)** $(n+1)/2$, **(ii)** $(n^2-1)/12$]
8. A player throwing an ordinary die is to receive ₹ $1/2^n$ where n denotes the number of throws required to get first 3. Find his expectation. [**Ans. :** ₹ 1/7]
9. Three fair coins are tossed. Find the expectation and the variance of number of heads.

   [**Ans. :** $\bar{x} = 3/2$, Var $(X) = 3/4$]
10. From a box containing n tickets bearing numbers 1, 2, 3, ..., n a ticket is drawn. If X denotes the number on the ticket drawn, find the mean and variance of X.

    [**Ans. :** $\bar{x} = (n+1)/2$, Var $(X) = (n^2-1)/2$]
11. In a game of chance a man is allowed to throw a fair coin indefinitely. He receives rupees 1, 2, 3, ... if he throws a head at the 1st, 2nd, 3rd, ... trial respectively. If the entry fee to participate in the game is ₹ 2, find the expected value of his net gain. [**Ans. :** Zero]
12. A person draws 3 balls from a bag containing 3 white, 4 red and 5 black balls. He is offered ₹ 10, ₹ 5 and ₹ 2 if he draws 3 balls of the same colour, 2 balls of the same colour and 1 ball of each colour respectively. Find his expectation.

    [**Ans. :** $p_1 = \frac{3}{44}$, $p_2 = \frac{29}{44}$, $p_3 = \frac{12}{44}$; ₹ 4·52]
13. A continuous random variable X has the density function $f(x) = k(1+x)$ where $2 \le x \le 5$. Find k, $P(x \le 4)$ and $E(X)$.

    [**Ans. :** $k = \frac{2}{27}$; $\frac{16}{27}$; $\frac{11}{3}$]

3. Mean and Variance

If we know the probability density function, discrete or continuous, we can find the mean and variance of the random variable as follows.

$$\text{Mean} = E(X) = \Sigma\, p_i x_i$$

We then find $\quad E(X^2) = \Sigma\, p_i x_i^2$

Now,

$$\text{Var}(X) = E(X - \bar{X})^2 = E[X - E(X)]^2$$
$$= E\left[X^2 - 2X\,E(X) + \{E(X)\}^2\right]$$
$$= E(X^2) - 2E(X) \cdot E(X) + [E(X)]^2$$

$$\boxed{\text{Var. } (X) = E(X^2) - [E(X)]^2}$$

We denote $E(X^2)$ by μ_2' and $E(x)$ by μ_1'. Then,

$$\boxed{\text{Mean} = \mu_1' \quad \text{and} \quad \text{Var. } (X) = \mu_2' - \mu_1'^2}$$

Type I : Examples on Mean and Variance of a Discrete Probability Distribution

Example 1 : A discrete random variable has the probability density function given below.

X	:	-2	-1	0	1	2	3
$P(X = x)$	:	0·2	k	0·1	$2k$	0·1	$2k$

Find k, the mean and variance.

Solution : We must have $\Sigma\, p_i = 1$.

$$\therefore \quad 5k + 0{\cdot}4 = 1 \quad \therefore \quad 5k = 0{\cdot}6 \quad \therefore \quad k = \frac{0{\cdot}6}{5} = \frac{3}{25}$$

Hence, the probability distribution is

X :	-2	-1	0	1	2	3
$P(X = x)$:	2 / 10	3 / 25	1 / 10	6 / 25	1 / 10	6 / 25

Now, $\quad \text{Mean} = E(X) = \Sigma\, p_i x_i = -\frac{4}{10} - \frac{3}{25} + 0 + \frac{6}{25} + \frac{2}{10} + \frac{18}{25} = \frac{60}{250} = \frac{6}{25}$

$$E(X^2) = \Sigma\, p_i x_i^2 = \frac{2}{10}(4) + \frac{3}{25}(1) + 0 + \frac{6}{25}(1) + \frac{1}{10}(4) + \frac{6}{25}(9) = \frac{73}{250}$$

$$\therefore \quad \text{Variance} = \sigma^2 = E(X^2) - [E(X)]^2 = \frac{73}{250} - \frac{36}{625} = \frac{293}{625}.$$

Example 2 : Find the value of k from the following data.

X :	0	10	15
$P(x)$:	$\frac{k-6}{5}$	$\frac{2}{k}$	$\frac{14}{5k}$

Also find the distribution function and the expectation of the distribution.

Sol. : Since $\Sigma\, p_i = 1$.

$$\frac{k-6}{5} + \frac{2}{k} + \frac{14}{5k} = 1 \qquad \therefore \quad k^2 - 11k + 14 = 0$$

$$\therefore \quad (k-8)(k-3) = 0 \qquad \therefore \quad k = 8 \text{ or } 3.$$

But when $k = 3$ $P(x = 0) = \frac{3-6}{5} = -\frac{3}{5}$ which is impossible. $\therefore$ $k = 8$.

$\therefore$ The p.d.f. and distribution function are

X	:	0	10	15
$P(x)$	:	2 / 5	1 / 4	7 / 20
$F(x)$	:	2 / 5	13 / 20	1

$$\therefore \quad E(X) = \Sigma p_i x_i = \frac{2}{5}(0) + \frac{1}{4}(10) + \frac{7}{20}(15) = \frac{5}{2} + \frac{21}{4} = \frac{31}{4}.$$

Type II : Examples on Mean and Variance of a Continuous Probability Distribution

Example 1 : A continuous random variable X has the p.d.f. defined by $f(x) = A + Bx, 0 \le x \le 1$. If the mean of the distribution is 1/ 3, find A and B.

Sol. : Since $f(x)$ is a probability distribution

$$\int_{-\infty}^{\infty} f(x)\,dx = 1. \qquad \text{By data } \int_0^1 (A + Bx)\,dx = 1$$

$$\therefore \left[AX + \frac{Bx^2}{2}\right]_0^1 = 1 \quad \therefore A + \frac{B}{2} = 1 \qquad \text{.....(i)}$$

Since the mean is $\frac{1}{3}$. $\int_0^1 x f(x)\,dx = \frac{1}{3}$ $\therefore$ $\int_0^1 (A + Bx)\,x\,dx = \frac{1}{3}$

$$\therefore \int_0^1 (Ax + Bx^2)\,dx = \frac{1}{3} \quad \therefore \left[A\frac{x^2}{2} + \frac{Bx^3}{3}\right]_0^1 = \frac{1}{3}$$

$$\therefore \frac{A}{2} + \frac{B}{3} = \frac{1}{3} \quad \therefore 3A + 2B = 2 \qquad \text{.....(ii)}$$

Solving the equations (i) and (ii), we get $A = 2$, $B = -2$.

$\therefore$ The p.d.f. is $f(x) = 2 - 2x,\ 0 \le x \le 1$

Example 2 : The distribution function of a r.v. X is given by $F_X(x) = 1 - (1 + x)\,e^{-x},\ x \ge 0$. Find the mean and variance.

Sol. : We have

$$f_X(x) = \frac{d F_X(x)}{dx} = (1 + x)\,e^{-x} - e^{-x} = x\,e^{-x}, \quad x \ge 0$$

$$\therefore \quad \text{Mean } \overline{X} = \int_0^{\infty} x \cdot f_X(x)\,dx = \int_0^{\infty} x^2 e^{-x}\,dx$$

$$= \left[x^2(-e^{-x}) - 2x(e^{-x}) + 2(1)\cdot(-e^{-x})\right]_0^{\infty} = 2$$

$$E(X^2) = \int_0^{\infty} x^2 f_X(x)\,dx = \int_0^{\infty} x^3 \cdot e^{-x}\,dx$$

$$\therefore \quad E(X^2) = \left[x^3(-e^{x}) - 3x^2(e^{-x}) + 6x(-e^{-x}) - 6(e^{-x})\right]_0^{\infty} = 6$$

$$\therefore \quad V(X) = E(X^2) - [E(X)]^2 = 6 - 4 = 2.$$

Example 3 : If X is a continuous random variable with probability density function given by

$$f(x) = \begin{cases} k(x - x^3), & 0 \le x \le 1 \\ 0, & \text{otherwise} \end{cases}$$

Find (i) k, (ii) mean, (iii) variance.

Sol. : (i) We have $\int_0^1 k(x - x^3)\,dx = 1$

$$\therefore \quad k\left[\frac{x^2}{2} - \frac{x^4}{4}\right]_0^1 = 1 \quad \therefore \quad k\left[\frac{1}{2} - \frac{1}{4}\right] = 1 \quad \therefore \quad k \cdot \frac{1}{4} = 1 \quad \therefore \quad k = 4$$

(ii) Mean, $\overline{X} = \mu_1' = \int_0^1 x f(x)\,dx = \int_0^1 x \cdot 4(x - x^3)\,dx$

$$= 4\int_0^1 (x^2 - x^4)\,dx = 4\left[\frac{x^3}{3} - \frac{x^5}{5}\right]_0^1 = 4\left[\frac{1}{3} - \frac{1}{5}\right] = \frac{8}{15}$$

$$\mu_2' = \int_0^1 x^2 f(x)\,dx = \int_0^1 x^2 \cdot 4(x - x^3)\,dx = 4\int_0^1 (x^3 - x^5)\,dx$$

$$= 4 \cdot \left[\frac{x^4}{4} - \frac{x^6}{6}\right]_0^1 = 4 \cdot \left[\frac{1}{4} - \frac{1}{6}\right] = 4 \cdot \frac{1}{12} = \frac{1}{3}$$

(iii) Variance (X) $= \mu_2' - \mu_1'^2 = \frac{1}{3} - \frac{64}{225} = \frac{33}{225} = 0{\cdot}1467.$

Example 4 : Find the value of k, if the function

$$f(x) = kx^2(1 - x^3), \quad 0 \le x \le 1$$
$$= 0 \quad \text{otherwise}$$

is a probability density function. Also find $P(0 \le x \le 1/2)$ and the mean and variance.

Sol. : We have $\int_0^1 kx^2(1 - x^3)\,dx = 1$ $\therefore$ $\int_0^1 k(x^2 - x^5)\,dx = 1$

$$\therefore \quad k\left[\frac{x^3}{3} - \frac{x^6}{6}\right]_0^1 = 1 \quad \therefore \quad k\left[\frac{1}{3} - \frac{1}{6}\right] = 1 \quad \therefore \quad k \cdot \frac{1}{6} = 1 \quad \therefore \quad k = 6.$$

Now, $P(0 \le x \le 2) = 6\int_0^{1/2} (x^2 - x^5)\,dx = 6\left[\frac{x^3}{3} - \frac{x^6}{6}\right]_0^{1/2} = 6\left[\frac{1}{8 \cdot 3} - \frac{1}{64 \cdot 6}\right] = \frac{15}{64}.$

Mean $\overline{x} = \mu_1' = \int_0^1 x f(x)\,dx = 6\int_0^1 x \cdot x^2(1 - x^3)\,dx = 6\int_0^1 (x^3 - x^6)\,dx$

$$= 6\left[\frac{x^4}{4} - \frac{x^7}{7}\right]_0^1 = 6\left[\frac{1}{4} - \frac{1}{7}\right] = \frac{18}{28} = \frac{9}{14}$$

$$\mu_2' = \int_0^1 x^2 f(x)\,dx = 6\int_0^1 x^2\left[x^2(1 - x^3)\right]dx = 6\int_0^1 (x^4 - x^7)\,dx$$

$$= 6\left[\frac{x^5}{5} - \frac{x^8}{8}\right]_0^1 = 6\left[\frac{1}{5} - \frac{1}{8}\right] = \frac{18}{40} = \frac{9}{20}$$

$$\text{Variance} = \mu_2' - \mu_1' = \frac{9}{20} - \frac{91}{196} = \frac{441-405}{980} = \frac{36}{980} = \frac{9}{245}.$$

EXERCISE - II

Type - I

1. The probability distribution of a random variable X is given by

X :	-2	-1	0	1	2	3
$P(X=x)$:	$0{\cdot}1$	k	$0{\cdot}2$	$2k$	$0{\cdot}3$	k

Find k, the mean and variance. [**Ans. : (i)** $k = 0{\cdot}1$, **(ii)** $\overline{X} = 0{\cdot}8$, **(iii)** Var. $2{\cdot}16$]

2. If X denotes the larger of the two numbers that appear when a pair of dice is thrown, find the probability distribution of X, and also the mean and variance of X.

[**Ans. :**

X :	1	2	3	4	5	6
$P(X=x)$:	$1/36$	$3/36$	$5/36$	$7/36$	$9/36$	$11/36$

Mean = $4{\cdot}47$, Var. = $1{\cdot}97$]

3. Given the following distribution

X :	-3	-2	-1	0	1	2
$P(X=x)$:	$0{\cdot}01$	$0{\cdot}1$	$0{\cdot}2$	$0{\cdot}3$	$0{\cdot}2$	$0{\cdot}15$

Find (i) $P(X \geq 1)$, (ii) $P(X < 0)$, (iii) $E(X)$, (iv) $V(X)$.

[**Ans. : (i)** $0{\cdot}35$, **(ii)** $0{\cdot}35$, **(iii)** $0{\cdot}05$, **(iv)** $1{\cdot}8475$]

4. A random variable X has the probability distribution

X :	-2	-1	0	1	2	3
$p(x)$:	$0{\cdot}1$	k	$0{\cdot}2$	$2k$	$0{\cdot}3$	k

Find k and then mean and variance of X. [**Ans. :** $k = 0{\cdot}1$, $0{\cdot}8$, $2{\cdot}16$]

Type - II

1. Find the mean and the variance of the following distribution

$$f(x) = \begin{cases} 1-x, & 0 < x < 1 \\ x-1, & 1 < x < 2 \end{cases}$$

[**Ans. : (i)** $\overline{X} = 1$, **(ii)** Var. = $1/2$]

2. If the probability density function is given by

$$f(x) = \begin{cases} kx^2(1-x^3), & 0 \leq x \leq 1 \\ 0, & \text{elsewhere} \end{cases}$$

Find (i) k, (ii) $P(0 < X < 1/2)$, (iii) $\overline{X}$, (iv) σ^2.

[**Ans. : (i)** $k = 6$, **(ii)** $15/64$, **(iii)** $9/14$, **(iv)** $9/245$]

3. If the probability density of a random variable is given by

$$f(x) = \begin{cases} kx, & 0 \leq x \leq 2 \\ 2k, & 2 \leq x \leq 4 \\ 6k - kx, & 4 \leq x \leq 6 \end{cases}$$

Find (i) k, (ii) $P(1 \leq X \leq 3)$, (iii) $\overline{X}$. [**Ans. : (i)** $1/8$, **(ii)** $7/16$, **(iii)** $7/4$]

4. If the probability density of a random variable is given by

$$f(x) = \begin{cases} kx\,e^{-x/3}, & x > 0 \\ 0, & x \leq 0 \end{cases}$$

(i) Find k, (ii) $\overline{X}$, (iii) σ^2.

[**Ans. : (i)** $1/9$, **(ii)** 6, **(iii)** 18]

5. A continuous random variable has probability density function

$$f(x) = \begin{cases} 2e^{-2x} & \text{for } x > 0 \\ 0 & \text{for } x \leq 0 \end{cases}$$

Find (i) $E(X)$, (ii) $E(X^2)$, (iii) Var (X), (iv) S.D. of X.

[**Ans. : (i)** $\frac{1}{2}$, **(ii)** $\frac{1}{2}$, **(iii)** $\frac{1}{4}$, **(iv)** $\frac{1}{2}$.]

6. The length of time (in minutes) a lady speaks on telephone is found to be a random variable with probability density function

$$f(x) = \begin{cases} Ae^{-x/5} & \text{for } x \geq 0 \\ 0 & \text{elsewhere} \end{cases}$$

Find A and the probability that she will speak for (i) more than 10 minutes, (ii) less than 5 minutes, (iii) between 5 and 10 minutes.

[**Ans. : (i)** $A = \frac{1}{5}$, **(ii)** $\frac{1}{e^2}$, **(iii)** $\frac{(e-1)}{e}$, **(iv)** $\frac{(e-1)}{e^2}$.]

7. The probability density function of a random variable is given by

$$f(x) = ke^{-x/\sigma}, \quad 0 < x < \infty$$

Find the mean and standard deviation of X.

[**Ans. : (i)** σ, **(ii)** σ]

EXERCISE - III

Theory

1. Define mathematical expectation.
2. Define expectation of a function of a random variable X.
   (a) when X is a discrete r.v., (b) when X is a continuous r.v.
3. Explain the following terms :
   (i) Expectation of a random variable.
   (ii) Variance of a random variable.
4. Prove that $V(X) = E(X^2) - \{E(X)\}^2$.
5. Prove that $V(aX + b) = a^2\, V(x)$.

☆☆☆

CHAPTER 6

Some Standard Distributions

1. Introduction

In this chapter we shall study some standard probability distributions. Study of such theoretical distributions is the foundation for the development of further topics. The first two distributions are discrete while the third is continuous.

2. Binomial Distribution

This was discovered by James Bernoulli's in 1700 and it expresses probabilities of events of dichotomous (dicho + tomy = two parts) nature *i.e.*, which results in only two ways, success or failure.

(a) To Derive Binomial Distribution : Consider an experiment which results in either success or failure. Let it be repeated n times, the probability p of success remaining constant every time and let $q = 1 - p$, the probability of failure.

The probability of x successes and hence $(n - x)$ failures in a trial in a particular order say, SSS ..... (x times) and then FFF ..... $(n - x)$ times, as given by the multiplication theorem on probability, is

$$= ppp \text{ ..... } (x \text{ times}) \times qqq \text{ ..... } (n - x) \text{ times}$$
$$= p^x q^{n-x}.$$

But x successes can occur in ${}^nC_x$ ways and the probability of each of these ways is the same *viz.* $p^x q^{n-x}$. Hence, the probability of x successes in any order by the addition theorem is ${}^nC_x p^x q^{n-x}$.

Hence, $p(x) = {}^nC_x p^x q^{n-x}, x = 0, 1, 2 \text{ ..... } n.$

This is Binomial distribution.

(b) Definition : A random variable is said to follow **Binomial distribution** if probability of x is given by

$$\boxed{P(X = x) = {}^nC_x p^x q^{n-x}, \quad x = 0, 1, 2, 3 \text{ ..... } n \text{ and } q = 1 - p}$$

The two constants n and p are called the **parameters** of the distribution.

(The meanings of 'success' and 'failure' are quite arbitrary. We may call 'a fatal accident' or 'getting infected by a disease' or 'premature delivery' a success !).

Remarks ....

1. The sum of the probabilities is 1.

$$\sum_{x=0} p(x) = \sum_{x=0} {}^nC_x p^x q^{n-x} = {}^nC_0 q^n + {}^nC_1 q^{n-1} p + {}^nC_2 q^{n-2} p^2 + \ldots\ldots + p^n$$
$$= (q + p)^n = 1.$$

2. Let the experiment of n trials be repeated N times. Then we expect x successes to occur $N \cdot {}^nC_x p^x q^{n-x}$ times. This is called frequency function. The expected frequencies of 0, 1, 2, ...n successes are the successive terms of the binomial expansion $N(q + p)^n$.

3. If x is a binomial variate with parameters n and p, it is denoted as $b(x, n, p)$.

4. The distribution is called **"Binomial Distribution"** because the probabilities ${}^nC_x p^x q^{n-x}$, $x = 0, 1, 2, \ldots\ldots, n$ are the successive terms of the expansion of the binomial expression $(q + p)^n$.

(c) When do we get binomial distribution ?

As is clear from the previous discussion, we get a binomial distribution when the following conditions are satisfied :

(i) A trial is repeated n times where n is a finite number.

(ii) Each trial results only in two ways-success or failure.3

(iii) These possibilities are mutually exclusive, exhaustive but not necessarily equally likely.

(iv) If p and q are the probabilities of success and failure then $p + q = 1$.

(v) The events are independent, *i.e.*, the probability p of success in each trial remains constant in all trials.

(d) Uses : Naturally, we can use binomial distribution when these conditions are satisfied. Thus, in problems involving (i) the tossing of a coin-heads or tails, (ii) the result of an examination-success or failure, (ii) the result of an election-success or failure, (iii) the result of inspection of an article-defective or non-defective, (iv) habit of a person-smoker or non-smoker etc. binomial distribution can be used if other conditions are also satisfied.

(e) Properties of the binomial distribution : We give below some important properties of the binomial distribution without derivation.

(i) If x denotes binomial variate, the mean of the distribution is given by,

$$\boxed{\bar{x} = np}$$

(ii) The variance of the distributions is npq,

i.e.,
$$\boxed{V(x) = npq}$$

$\therefore$
$$\boxed{\text{Mean} = np \text{ and Variance} = npq}$$

(iii) If the experiment, each of n trials, is repeated N times then the expectation of r successes *i.e.*, the expected frequency of r successes in N experiments is $N \cdot {}^nC_r p^r q^{n-r}$

$$\boxed{\text{Frequency} = N \cdot p = N\, {}^nC_r p^r q^{n-r}}$$

Remark ....

Since for a Binomial variate $\frac{V(x)}{\bar{x}} = \frac{npq}{np} = q$ and since $q < 1$, variance of x cannot be greater than the mean.

Example 1 : Find the mean of the probability distribution of the number of heads obtained in three flips of a balanced coin.

Sol. : We have $p = 1/2$, $n = 3$ $\quad\therefore$ Mean $= E(X) = np = 3 \times \frac{1}{2} = 1{\cdot}5$

Example 2 : What is the expectation of heads if an unbiased coin is tossed 12 times ?

Sol. : Since the expectation of x in a binomial distribution is given by $E(X) = np$ and $n = 12$, $p = 1/2$. We could expect $12 \times (1/2)$ *i.e.* 6 heads.

Example 3 : If X is Binomially distributed with $E(X) = 2$ and Var $(X) = 4/3$, find the probability distribution of X.

Sol. : We have $E(X) = np = 2$ and Var $(X) = npq = 4/3$.

$$\therefore \frac{npq}{np} = \frac{4/3}{2} \quad \therefore q = \frac{2}{3} \quad \therefore p = 1 - q = \frac{1}{3}$$

But $np = 2$, $\therefore n \cdot \frac{1}{3} = 2 \quad \therefore n = 6$

Hence, the distribution is

$$P(X = x) = {}^nC_x\, p^x q^{n-x} = {}^6C_x \left(\frac{1}{3}\right)^x \left(\frac{2}{3}\right)^{6-x}$$

Putting $x = 0, 1, 2, \ldots\ldots 6$, we get the following probability distribution of X.

x :	0	1	2	3	4	5	6
$P(X = x)$:	$\frac{64}{729}$	$\frac{192}{729}$	$\frac{240}{729}$	$\frac{160}{729}$	$\frac{60}{729}$	$\frac{12}{729}$	$\frac{1}{729}$

Example 4 : Prove that for all Binomial distributions with the same parameter n, the variance is maximum when $p = 1/2$.

Sol. : We have for Binomial distribution

$$p(x) = {}^nC_x\, p^x q^{n-x}$$

And the variance is given by

$$V = npq \qquad \text{where, } q = 1 - p$$

$$\therefore V = np(1-p) = np - np^2$$

For maxima, $\frac{dV}{dp} = 0$ and $\frac{d^2V}{dp^2} = -\text{ve}$

Now, $\frac{dV}{dp} = n - 2np$ and $\frac{d^2V}{dp^2} = -2n$, always – ve.

$\therefore \frac{dV}{dp} = 0$ gives $\quad n - 2np = 0 \quad \therefore n(1 - 2p) = 0$

But $\quad n \neq 0 \quad \therefore 1 - 2p = 0 \quad \therefore p = \frac{1}{2}$.

Hence, the variance is maximum when $p = \frac{1}{2}$.

Example 5 : The mean and variance of a Binomial distribution are 4 and 4/3. Find the distribution and the probability of atleast one success.

Sol. : We have $\quad E(X) = np = 4, \quad V(X) = npq = 4/3$

$$\therefore \frac{npq}{np} = \frac{4/3}{4} = \frac{1}{3} \quad \therefore q = \frac{1}{3} \quad \therefore p = \frac{2}{3}$$

$$\because np = 4 \quad \therefore n \cdot \frac{2}{3} = 4 \quad \therefore n = 6.$$

$\therefore$ The distribution is

$$P(X = x) = {}^nC_x\, p^x q^{n-x} = {}^6C_x \left(\frac{2}{3}\right)^x \left(\frac{1}{3}\right)^{6-x}$$

x :	0	1	2	3	4	5	6
$P(X = x)$:	$\frac{1}{729}$	$\frac{12}{729}$	$\frac{60}{729}$	$\frac{160}{729}$	$\frac{240}{729}$	$\frac{192}{729}$	$\frac{64}{729}$

P (atleast one success) $= 1 - P$ (0 success) $= 1 - \frac{1}{729} = \frac{728}{729}$.

(Compare Ex. 3 with Ex. 5.)

Example 6 : Find the Binomial distribution if the mean is 4 and variance is 3.

Sol. : We have mean $= np = 4$ and variance $= npq = 4$.

$$\therefore \frac{np}{npq} = \frac{4}{3} \quad \therefore \frac{1}{q} = \frac{4}{3} \quad \therefore q = \frac{3}{4} \quad \therefore p = 1 - q = 1 - \frac{3}{4} = \frac{1}{4}$$

$$\therefore n = 4 \quad \therefore n \cdot \frac{1}{4} = 4 \quad \therefore n = 16$$

$$\therefore P(X = x) = {}^nC_x\, p^x q^{n-x} = {}^{16}C_x \left(\frac{1}{4}\right)^x \left(\frac{3}{4}\right)^{16-x}$$

Example 7 : "In a Binomial distribution the mean is 5 and the standard deviation is 3". Find the fallacy if any in this statement.

Sol. : For a Binomial distribution the mean $= np = 5$ and variance $= npq = 9$.

$$\therefore \frac{np}{npq} = \frac{5}{9} \quad \therefore \frac{1}{q} = \frac{5}{9} \quad \therefore q = \frac{9}{5} = 1{\cdot}8$$

This is the probability of failure and probability of an event cannot be greater than 1.

Hence, the statement is wrong.

Example 8 : With usual notation find p of Binomial distribution if $n = 6$, $9P(X = 4) = P(X = 2)$.

Sol. : We have $\quad P(X = x) = {}^nC_x\, p^x q^{n-x} = {}^6C_x\, p^x q^{6-x}$

Since, $9\, P(X = 4) = P(X = 2)$

$$\therefore 9\, {}^6C_4\, p^4 q^{6-4} = {}^6C_2\, p^2 q^{6-2}$$

$$\because {}^6C_4 = {}^6C_2, \quad 9p^2 = q^2 = (1-p)^2$$

$$\therefore 9p^2 = 1 - 2p + p^2 \quad \therefore 8p^2 + 2p - 1 = 0$$

$$\therefore (4p - 1)(2p + 1) = 0 \quad \therefore p = \frac{1}{4} \text{ or } p = -\frac{1}{2}$$

$$\therefore p = \frac{1}{4} \text{ and } q = \frac{3}{4}.$$

Example 9 : What is the mean and variance of the Binomial Distribution $(0{\cdot}3 + 0{\cdot}7)^{10}$, $q = 0{\cdot}3$?

Sol. : Here, $p = 0{\cdot}7$, $q = 0{\cdot}3$, $n = 10$.

$\therefore$ Mean $= np = 10 \times 0{\cdot}7 = 7$

$\therefore$ Variance $= npq = 10 \times 0{\cdot}3 \times 0{\cdot}7 = 2{\cdot}1$.

Example 10 : A factory turns out an article by mass production methods. From the past experience it is found that 20 articles on an average are rejected out of every batch of 100. Find the mean and the variance of the number of rejected articles.

Sol. : The number of rejected articles in a batch is a Binomial variate with $n = 100$ and $p = \frac{20}{100} = 0{\cdot}2$

Hence, the mean of the distribution $= np = 100 \times \frac{2}{10} = 20$.

$$\text{Variance} = \sqrt{npq} = \sqrt{100 \times 0{\cdot}2 \times 0{\cdot}8} = \sqrt{16} = 4.$$

Example 11 : The 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 ?

Sol. : For a Binomial distribution

$$P(X = x) = {}^nC_x p^x q^{n-x}$$

When $n = 5$, $x = 3$, $\quad P(X = 3) = {}^5C_3\, p^3 q^2$

When $n = 5$, $x = 2$, $\quad P(X = 2) = {}^5C_2\, p^2 q^3$

The ratio of these probabilities is 1/4.

$$\therefore \frac{P(X = 3)}{P(X = 2)} = \frac{{}^5C_3\, p^3 q^2}{{}^5C_2\, p^2 q^3} = \frac{1}{4}$$

Since, ${}^5C_3 = {}^5C_2$, we get

$$\frac{p}{q} = \frac{1}{4} \quad \therefore \frac{p}{1-p} = \frac{1}{4} \quad \therefore 4p = 1 - p$$

$$\therefore 5p = 1 \quad \therefore p = \frac{1}{5} \quad \therefore q = 1 - p = \frac{4}{5}$$

$$\therefore P(X = x) = {}^nC_x \left(\frac{1}{5}\right)^x \left(\frac{4}{5}\right)^{n-x}$$

When $n = 6$ and $x = 4$, $\quad P(X = 4) = {}^6C_4 \left(\frac{1}{5}\right)^4 \cdot \left(\frac{4}{5}\right)^2$.

Example 12 : It has been claimed that in 60% of all solar heat installations, the utility bill is reduced by atleast one third. Accordingly what are the probabilities that the utility bill will be reduced by atleast one third in (i) four of five installations, (ii) atleast four of five installations.

Sol. : For a Binomial distribution $\quad P(X = x) = {}^nC_x p^x q^{n-x}$

We have $n = 5$, $x = 4$, $p = 0{\cdot}6$, $q = 0{\cdot}4$.

(i) $\therefore \quad P(X = 4) = {}^5C_4\, (0{\cdot}6)^4\, (0{\cdot}4)^1 = 0{\cdot}259$

(ii) P (atleast 4 of 5 installations) $= P(X \geq 4) = P(X = 4 \text{ or } 5)$

$= P(X = 4) + P(X = 5)$

$= {}^5C_4\, (0{\cdot}6)^4\, (0{\cdot}4)^1 + {}^5C_5\, (0{\cdot}6)^5\, (0{\cdot}4)^0$

$= 0{\cdot}259 + 0{\cdot}078 = 0{\cdot}337$.

Example 13 : The incidence of an occupational disease in an industry is such that the workers have 20% chance of suffering from it. What is the probability that out of 6 workers chosen at random 4 or more will be suffering from the disease ?

Sol. : We have $p = 20\% = \frac{20}{100} = 0{\cdot}2$, $q = 1 - p = 0{\cdot}8$, $n = 6$

$$\therefore P(X = x) = {}^nC_x\, p^x q^{n-x} = {}^6C_x \left(\frac{1}{5}\right)^x \left(\frac{4}{5}\right)^{6-x}$$

$$\therefore P(X \geq 4) = P(X = 4) + P(X = 5) + P(X = 6)$$

$$= {}^6C_4 \left(\frac{1}{5}\right)^4 \left(\frac{4}{5}\right)^2 + {}^6C_5 \left(\frac{1}{5}\right)^5 \left(\frac{4}{5}\right)^1 + {}^6C_6 \left(\frac{1}{5}\right)^6 \left(\frac{4}{5}\right)^0$$

$$\therefore P(X \geq 4) = \frac{1}{5^6}\left[15{\cdot}4^2 + 6{\cdot}4 + 1\right] = \frac{205}{5^6} = \frac{41}{3125}.$$

Example 14 : The probability that a bomb dropped from a plane will strike the target is 1/5. If six such bombs are dropped, find the probability that (i) exactly two bombs hit the target, (ii) at least two will hit the target.

Sol. : We have $p = \frac{1}{5}$, $q = \frac{4}{5}$, $n = 6$.

By Binomial distribution, $P(x) = {}^nC_x\, p^x q^{n-x} = {}^6C_x \left(\frac{1}{5}\right)^x \left(\frac{4}{5}\right)^{6-x}$

$$\therefore P(2) = {}^6C_2 \left(\frac{1}{5}\right)^2 \left(\frac{4}{5}\right)^4 = \frac{6 \cdot 5}{2} \cdot \frac{1}{25} \cdot \frac{256}{625} = \frac{1536}{6250} = 0{\cdot}24576.$$

P (at least two) $= 1 - [P(1)]$

Now, $\quad P(1) = {}^6C_1 \left(\frac{1}{5}\right)\left(\frac{4}{5}\right)^5 = 6 \cdot \frac{1}{5} \cdot \frac{1024}{3125} = \frac{6144}{15625} = 0{\cdot}3932$

$\therefore$ P (at least two) $= 1 - 0.3932 = 0{\cdot}6068$.

Example 15 : The probability that at any moment one telephone line out of 10 will be busy is 0·2.

(i) What is the probability that 5 lines are busy ?

(ii) Find the expected number of busy lines and also find the probability of this number.

(iii) What is the probability that all lines are busy ?

Sol. : We have, $p = 0{\cdot}2$, $q = 0{\cdot}8$, $n = 10$.

By Binomial distribution, $P(X = x) = {}^nC_x\, p^x q^{n-x}$

$\therefore \quad P(X = x) = {}^{10}C_x\, (0{\cdot}2)^x\, (0{\cdot}8)^{10-x}$

(i) $\therefore \quad P(X = 5) = {}^{10}C_5\, (0{\cdot}2)^5\, (0{\cdot}8)^{0{\cdot}5}$

(ii) Expected number of busy lines = mean = $np = 10 \times \frac{2}{10} = 2$.

$$P(X=2) = {}^{10}C_2\,(0{\cdot}2)^2\,(0{\cdot}8)^{0{\cdot}8}$$

(iii) Probability of all lines busy = $P(X=10) = {}^{10}C_{10}\,(0{\cdot}2)^{10}\,(0{\cdot}8)^0 = 0{\cdot}2^{10}$.

Example 16 : Seven dice are thrown 729 times. How many times do you expect at least four dice to show three or five ?

Sol. : Probability of getting (3 or 5) in a single toss $= \frac{1}{6} + \frac{1}{6} = \frac{1}{3}$.

This is a binomial distribution with $n = 7$, $p = 1/3$, $q = 2/3$.

$$\therefore\ P(X=x) = {}^nC_x\,p^x q^{n-x} = {}^7C_2\left(\frac{1}{3}\right)^x\left(\frac{2}{3}\right)^{7-x}$$

P(at least 4 successes) = $P(x = 4, 5, 6, 7)$

$$= {}^7C_4\left(\frac{1}{3}\right)^4\left(\frac{2}{3}\right)^3 + {}^7C_5\left(\frac{1}{3}\right)^5\left(\frac{2}{3}\right)^2 + {}^7C_6\left(\frac{1}{3}\right)^6\left(\frac{2}{3}\right)^1 + {}^7C_7\left(\frac{1}{3}\right)^7\left(\frac{2}{3}\right)^0$$

$$= \frac{379}{3^7}$$

$\therefore$ The expected number of times of getting (3 or 5) at least 4 times

$$= Np = 729 \times \frac{379}{3^7} = 126{\cdot}3.$$

Example 17 : Out of 1000 families of 3 Children each, how many would you expect to have 2 boys and 1 girl?

Sol. : Here $P(\text{Boy}) = p = \frac{1}{2}$, $P(\text{Girl}) = q = \frac{1}{2}$, $n = 3$, $r = 2$.

$$\therefore\ P(\text{2 boys and 1 girl}) = {}^3C_2\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right) = \frac{3}{8}$$

$$\therefore\ \text{Expected number of families} = Np = 1000 \times \frac{3}{8} = 375$$

Example 18 : If hens of a certain breed lay eggs on 5 days a week on an average; find on how many days during a season of 100 days, a poultry keeper with 5 hens of this breed, will expect to receive at least 4 eggs?

Sol. : Probability of an hen laying an egg, $p = 5/7$

and Probability of not laying an egg, $q = 1 - p = 2/7$.

$$P(X=x) = {}^nC_x\,p^x q^{n-x} \text{ and } n = 5,\ p = \frac{5}{7},\ q = \frac{2}{7}$$

$$\therefore\ P(X \geq 4) = P(X=4) + P(X=5)$$

$$= {}^5C_4\left(\frac{5}{7}\right)^4\left(\frac{2}{7}\right)^1 + {}^5C_5\left(\frac{5}{7}\right)^5\left(\frac{2}{7}\right)^0 = 0{\cdot}5578.$$

Expectation = $Np = 100 \times 0{\cdot}5578 = 55{\cdot}78 = 56$.

Example 19 : If 10% bolts produced by a machine are defective, calculate the probability that out of a sample selected at random of 10 bolts, not more than one bolt will be defective. **(S.U. 1990, 91, 92, 2003, 05)**

Sol. : We have P(defective bolt) = $p = 0{\cdot}1$, $q = 0{\cdot}9$ and $n = 10$

P(not more than one bolt defective) = P(Zero defective) + P(one defective)

By Binomial Distribution

$$P(x) = {}^nC_x\,p^x q^{n-x} = {}^{10}C_x(0{\cdot}1)^x\,(0{\cdot}9)^{10-x}$$

$$\therefore\ P(x=0) = {}^{10}C_0(0{\cdot}1)^0\,(0{\cdot}9)^{10} = 0{\cdot}3487$$

$$P(x=1) = {}^{10}C_1\,(0{\cdot}1)^1\,(0{\cdot}9)^9 = 0{\cdot}3874$$

$\therefore$ Required probability $= P(x=0) + P(x=1) = 0{\cdot}7361$.

Example 20 : From a box containing 100 transistors 20 of which are defective, 10 are selected at random. Find the probability that (i) all will be defective (ii) all are non-defective (iii) at least one is defective. **(S.U. 1985, 88, 93, 98)**

Sol. : P(defective transistor) $= p = \frac{20}{100} = 0{\cdot}2$, $q = 0{\cdot}8$ and $n = 10$

$$P(x\text{ def.}) = {}^nC_x\,p^x q^{n-x} = {}^{10}C_x(0{\cdot}2)^x\,(0{\cdot}8)^{10-x}$$

(i) P(all def.) $= P(x=10) = {}^{10}C_{10}(0{\cdot}2)^{10}\,(0{\cdot}8)^0 = 0{\cdot}0000001$

(ii) P(zero def.) $= P(x=0) = {}^{10}C_0(0{\cdot}2)^0\,(0{\cdot}8)^{10} = 0{\cdot}1074$

(iii) P(at least one defective) $= 1 - P$(zero defective)

$= 1 - 0{\cdot}1074 = 0{\cdot}8926$.

Example 21 : In a sampling the mean number of defective bolts manufactured by a machine in a sample of 20 is 2. Determine the expected number of samples out of such 500 samples to contain at least 2 defective bolts. **(S.U. 1988, 95, 2003, 04)**

Sol. : In Binomial distribution

Mean = np. We have mean = 2 and $n = 20$

$$\therefore\ p = \frac{2}{20} = 0{\cdot}1 \text{ and } q = 1 - p = 0{\cdot}9.$$

If x denotes the number of defective bolts, then

$$P(x) = {}^nC_x\,p^x q^{n-x} = {}^{20}C_x(0{\cdot}1)^x\,(0{\cdot}9)^{20-x}$$

$$P(\text{zero defective}) = {}^{20}C_0(0{\cdot}1)^0\,(0{\cdot}9)^{20} = 0{\cdot}1216$$

$$P(\text{one defective}) = {}^{20}C_1(0{\cdot}1)^1\,(0{\cdot}9)^{19} = 0{\cdot}2702$$

P(At least two defective) $= 1 - [P(x=0) + P(x=1)]$

$= 1 - [0{\cdot}1216 + 0{\cdot}2702]$

$= 0{\cdot}6082$

$\therefore$ Number of samples having at least 2 defective bolts

$= Np = 500 \times 0{\cdot}6082 = 304$ nearly.

Example 22 : Six dice are thrown 729 times. How many times do you expect at least three dice to show a 5 or 6 ? **(S.U. 1987, 2002, 04, 05)**

Sol. : In a toss of a single dice,

$$P(5 \text{ or } 6) = p = \frac{2}{6} = \frac{1}{3} \text{ and } q = \frac{2}{3}$$

If 6 dice are thrown simultaneously $n = 6$

$$\therefore \quad P(x \text{ successes}) = {}^6C_x \left(\frac{1}{3}\right)^x \left(\frac{2}{3}\right)^{6-x}$$

$$P(x=0) = {}^6C_0 \left(\frac{1}{3}\right)^0 \left(\frac{2}{3}\right)^6 = \frac{64}{729}; \qquad P(x=1) = {}^6C_1 \left(\frac{1}{3}\right)^1 \left(\frac{2}{3}\right)^5 = \frac{192}{729};$$

$$P(x=2) = {}^6C_2 \left(\frac{1}{3}\right)^2 \left(\frac{2}{3}\right)^4 = \frac{240}{729}.$$

$$P(\text{at least 3 successes}) = 1 - [\,P(0) + P(1) + P(2)\,] = 1 - \frac{496}{729} = \frac{233}{729}.$$

$$\text{Expected number} = Np = 729 \frac{233}{729} = 233.$$

Example 23 : Seven coins are tossed and the number of heads obtained is noted. The experiment is repeated 128 times and the following distribution is obtained.

No. of heads	:	0,	1,	2,	3,	4,	5,	6,	7 Total
Frequency	:	7,	6,	19,	35,	30,	23,	7,	1 128

Fit a Binomial distribution if (i) the coins are unbiased, (ii) if the nature of the coins is not known.

Sol. : To fit a distribution to given data means to find the constants of the distribution which will adequately describe the given situation.

(i) *When the coins are unbiased*

$$p = \frac{1}{2}, q = \frac{1}{2} \text{ and by data } n = 7 \qquad \therefore \quad P(X = x) = {}^7C_x \left(\frac{1}{2}\right)^x \left(\frac{1}{2}\right)^{7-x}$$

Putting $x = 0, 1, 2, 3, \ldots, 7$, we get

$$P(0) = \frac{1}{2^7}, \quad P(1) = \frac{7}{2^7}, \quad P(2) = \frac{21}{2^7}, \ \ldots$$

Expected frequency = Np and $N = 128$.

Multiplying the above probabilities by 128 *i.e.* by 2^7 we get the expected frequencies as 1, 7, 21, 35, 35, 21, 7, 1.

(ii) *When the nature of the coins is not known.*

We have $\overline{X} = \dfrac{\sum f_i x_i}{N} = \dfrac{0 \times 7 + 1 \times 6 + 2 \times 19 + \ldots + 7 \times 1}{128} = \dfrac{433}{128} = 3{\cdot}38$

But $\overline{X} = np$

$$\therefore \quad p = \frac{\overline{X}}{n} = \frac{3{\cdot}38}{7} = 0{\cdot}48 \qquad \therefore \quad q = 1 - p = 0{\cdot}52$$

$$\therefore \quad P(X = x) = {}^7C_x (0{\cdot}48)^x (0{\cdot}52)^{7-x}$$

Putting $x = 0, 1, 2, 3, \ldots, 7$ we get

$$P(0) = 0{\cdot}01, \ P(1) = 0{\cdot}066, \ P(2) = 0{\cdot}184, \ldots$$

Multiply these probabilities by 128 we get the expected frequencies as

1, 8, 23, 36, 33, 18, 6, 3.

(Last term = 128 – sum of other terms).

Example 24 : Fit a Binomial distribution to the following data.

X	:	0	1	2	3	4	5	6
Y	:	5	18	28	12	7	6	4

Sol. : We have $n = 6$

$$N = 5 + 18 + 28 + 12 + 7 + 6 + 4 = 80$$

$$\sum f_i x_i = 5 \times 0 + 18 \times 1 + 28 \times 2 + 12 \times 3 + 7 \times 4 + 6 \times 5 + 4 \times 6$$

$$= 0 + 18 + 56 + 36 + 28 + 30 + 24 = 192$$

$$\therefore \quad \overline{X} = \frac{\sum f_i x_i}{N} = \frac{192}{80} = 2{\cdot}4$$

But $\overline{X} = np \qquad \therefore \quad p = \dfrac{\overline{X}}{n} = \dfrac{2{\cdot}4}{6} = 0{\cdot}4$

$$\therefore \quad q = 1 - p = 1 - 0{\cdot}4 = 0{\cdot}6$$

$$\therefore \quad p(x) = {}^nC_x p^n q^{n-x} = {}^6C_x (0{\cdot}4)^x (0{\cdot}6)^{6-x}$$

Putting $x = 0, 1, 2, 3, 4, 5, 6$, we get

$$p(0) = {}^6C_0 (0{\cdot}4)^0 (0{\cdot}6)^6 = 0{\cdot}0466; \qquad p(1) = {}^6C_1 (0{\cdot}4)^1 (0{\cdot}6)^5 = 0{\cdot}1866;$$

$$p(2) = {}^6C_2 (0{\cdot}4)^2 (0{\cdot}6)^4 = 0{\cdot}3110; \qquad p(3) = {}^6C_3 (0{\cdot}4)^3 (0{\cdot}6)^3 = 0{\cdot}2765;$$

$$p(4) = {}^6C_4 (0{\cdot}4)^4 (0{\cdot}6)^2 = 0{\cdot}1382; \qquad p(5) = {}^6C_5 (0{\cdot}4)^5 (0{\cdot}6)^1 = 0{\cdot}0369;$$

$$p(6) = {}^6C_6 (0{\cdot}4)^6 (0{\cdot}6)^0 = 0{\cdot}0041.$$

Multiplying these probabilities by $N = 80$, we get the expected frequencies as

x	:	0	1	2	3	4	5	6
y	:	4	15	25	22	11	3	0

(The frequencies are to be rounded).

Example 25 : Twelve dice were thrown 4096 times and the number of appearance of 6 each time was noted.

No. of successes	:	0	1	2	3	4	5	6 and above
Frequency	:	447	1145	1181	786	380	115	32

Fit a Binomial distribution when the dice are unbiased.

Sol. : We have the probability of getting 6 in a throw of one die $p = \dfrac{1}{6}$.

The probability of not getting 6 is $q = 1 - \dfrac{1}{6} = \dfrac{5}{6}$.

The number of trials, $n = 12$.

The number of repetitions, $N = 4096$.

Now, $P(X = x) = {}^nC_x p^x q^{n-x} = {}^{12}C_x \left(\dfrac{1}{6}\right)^x \left(\dfrac{5}{6}\right)^{12-x}$

Putting $X = 0, 1, 2, 3, 4, 5, ...$

$$P(0) = {}^{12}C_0\left(\frac{1}{6}\right)^0\left(\frac{5}{6}\right)^{12} = 0{\cdot}1122\ ; \qquad P(1) = {}^{12}C_1\left(\frac{1}{6}\right)^1\left(\frac{5}{6}\right)^{11} = 0{\cdot}2692\ ;$$

$$P(2) = {}^{12}C_2\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)^{10} = 0{\cdot}2961\ ; \qquad P(3) = {}^{12}C_3\left(\frac{1}{6}\right)^3\left(\frac{5}{6}\right)^{9} = 0{\cdot}1974\ ;$$

$$P(4) = {}^{12}C_4\left(\frac{1}{6}\right)^4\left(\frac{5}{6}\right)^{8} = 0{\cdot}0888\ ; \qquad P(5) = {}^{12}C_5\left(\frac{1}{6}\right)^5\left(\frac{5}{6}\right)^{7} = 0{\cdot}0284\ ;$$

P (6 and above) = 1 – (0·1122 + 0·2692 + ... + 0·0284) = 0·0079

Expected frequency = $Np = 4096p$

∴ **No. of successes** :	0	1	2	3	4	5	6 and above
Frequency :	460	1103	1213	808	364	116	32

EXERCISE - I

(A) **1.** For a binomial distribution the mean is 2 and the standard deviation is 1. Find all the constants of the distribution. **(S.U. 1999)** [**Ans. :** $n = 4,\ p = 1/2,\ q = 1/2$]

2. For a binomial distribution the mean is 12 and the variance is 4. Find all the constants. **(S.U. 2002)** [**Ans. :** $n = 18,\ p = 2/3,\ q = 1/3$]

3. Find out the fallacy if any in the following statements giving reasons :

(a) The mean of a binomial distribution is 15 and the standard deviation is 5. **(S.U. 1989)**

(b) The mean of a binomial distribution is 12 and the standard deviation is 3.

[**Ans. :** (i) False (ii) True $V(x) \ngtr \bar{x}$]

4. If the probability of defective bulbs is 0·2, find the mean and variance of the distribution of defective bulbs in a lot of 1000 bulbs. [**Ans. :** $\bar{x} = np = 200,\ var = npq = 160$]

5. If the mean of the binomial distribution is 2 and the variance is 4/3 find the probability of (i) two successes, (ii) less than two successes.

(**Hint :** Find n, p, q.) [**Ans. : (i)** 0·3292, **(ii)** 0·3512]

(B) **1.** An examination containing multiple choice questions is designed so that the probability of a correct choice for any question by guessing alone is 0·2. What is the probability that a student will not get more than four questions right out of 20 merely by guessing? **(S.U. 1986)** [**Ans. :** 0·4114]

2. In a hurdle race a player has to cross 10 hurdles. The probability that he will clear each hurdle is 5 / 6. What is the probability that he will knock down less than two hurdles? **(S.U. 1997)** [**Ans. :** 0·4845]

3. The probability that a worker will suffer from an occupational disease is 25%. What is the chance than out of 6 workers 4 or more will suffer from the disease ? **(S.U. 1986)** [**Ans. :** 0·0376]

4. 10% of the tools produced in a certain manufacturing process turn out to be defective.

(a) Find the probability that in a sample of 10 tools chosen at random exactly two will be defective. **(S.U. 1987)** [**Ans. :** 0·1937]

(b) Find the probability that out of 20 tools selected at random there are (i) exactly two defectives (ii) at least two defectives. **(S.U. 1989, 2003)** [**Ans. : (i)** 0·2852, **(ii)** 0·3918]

5. If the probability that a new-born child is male is 0·6, find the probability that in a family of 5 children there will be exactly 3 boys. **(S.U. 1985, 2001)** [**Ans. :** 0·3456]

6. If in a lot of 500 solenoids 25 are defective, find the probability that there will be 0, 1, 2 defective solenoids in a random sample of 20 solenoids. **(S.U. 1985, 2006)** [**Ans. : (i)** 0·3585, **(ii)** 0·3773, **(iii)** 0·1887]

7. In a room there are three lamp-sockets. A bag contains 6 working and 4 non-working bulbs. Three bulbs are selected at random and fitted in the sockets. Find the probability that there will be some light in the room. **(S.U. 1989)** [**Ans. :** 29 / 30]

8. If on an average one candidate out of 10 fails in a certain examination, find the chance that out of 5 candidates that have appeared for the examination at least 4 will be successful. **(S.U. 1985)** [**Ans. :** 0·918]

9. In a locality 20% of people smoke. Find the probability that out of 6 persons chosen at random from this locality 4 or more will we found to be smokers. **(S.U. 1985, 2002)** [**Ans. :** 0·0169]

10. Assume that on an average one telephone number out of fifteen called between 2 p.m. and 3 p.m. on weekdays is busy. What is the probability that if 6 randomly selected numbers are called (i) not more than 3, (ii) at least 3 will be busy ? **(S.U. 2003)** [**Ans. : (i)** 0·9997, **(ii)** 0·0051]

11. In a large consignment of electric bulbs, 10 percent are defective. A random sample of 20 is taken for inspection. Using Binomial distribution find the probability that (i) all are good bulbs, (ii) atmost three are defective bulbs, (iii) exactly three are defective bulbs. **(S.U. 1990, 95)** [**Ans. : (i)** 0·1216, **(ii)** 0·8671, **(iii)** 0·1901]

12. The probability that a man aged 60 will live upto 70 is 0·65. What is the probability that out of 10 men, now 60, at least 8 would live upto 70 ? **(S.U. 1987, 90, 2006)** [**Ans. :** 0·2616]

13. The probability that a student in an evening college will graduate is 0·4. Determine the probability that out of 5 students (i) none, (ii) one, (iii) at least one will graduate. **(S.U. 1987, 2004)** [**Ans. : (i)** 0·0778, **(ii)** 0·2592, **(iii)** 0·9222]

(C) **1.** Out of 320 families with 5 children each, find the expected number of families having 0 boys, 1 boy, 2 boys, ... 5 boys, assuming boys and girls are equally likely. **(S.U. 1986, 2002, 03)** [**Ans. :** 10, 50, 100, 100, 50, 10]

2. On an average a student is present on 5 days a week. Find on how many days in a course of 100 days out of 5 students of this class at least 4 will be present. [**Ans. :** 56]

3. Six fair coins are tossed simultaneously. If 192 such tosses are made find the expected number of tosses showing (i) one and only one heads (ii) no heads (iii) all heads. [**Ans. : (i)** 18, **(ii)** 3, **(iii)** 3]

4. In a sampling of a large number of parts produced by a machine, the mean number of defectives in a sample of 20 is 2. Out of such 1000 samples how many samples would you expect to contain at least three defective ? **(S.U. 1985)** [**Ans. :** 323]

5. Out of 800 families with 5 children each how many would you expect to have (i) 3 boys, (ii) 5 girls ? **(S.U. 1988)** [**Ans. : (i)** 250, **(ii)** 25]

6. The mean of defective blades supplied in packets of 10 is one. In how many packets of this make out of 1000 packets would you expect to find at least 4 non-defective blades. [**Ans. :** 13]

7. Out of 1000 families having 3 children each, how many would you expect to have (i) 2 boys and 1 girl, (ii) 2 girls and 1 boy ? [**Ans. : (i)** 375, **(ii)** 375]

8. Assuming that 20 % the population is literate so that the chance of an individual being literate is 1/5 and assuming that 100 investigators can take a sample of 10 individuals to see whether they are literates, how many investigators would you expect to report that three people or less were literate ? **(S.U. 2006) [Ans. : 88]**

9. Take 100 sets of 10 tosses of an unbiased coin. In how many cases do you expect to get (a) 7 heads and 3 tails (b) 7 heads at least ? **[Ans. : 12, 17]**

10. On an average 3 out of ten students fail in an examination. If 1000 samples each of 10 students are taken in how many would you expect that (i) none has failed, (ii) all have failed ?

[Ans. : (i) 28, (ii) zero]

11. Five coins are tossed 320 times, find the frequency distribution of heads. **(S.U. 1998)**

[Ans. : 10, 50, 100, 100, 50, 10]

(D) 1. Seven coins are tossed at a time, 256 times. Number of heads obtained in each toss are recorded below. Fit a binomial distribution under the hypothesis that the coins are unbiased.

No. of heads :	0,	1,	2,	3,	4,	5,	6,	7.
Frequency :	14,	12,	38,	70,	60,	46,	14,	2.

(S.U. 1999)

[Ans. : $256\left(\frac{1}{2}+\frac{1}{2}\right)^7$ *i.e.*, 2, 14, 42, 70, 70, 42, 14, 2]

2. In an experiment with 500 seeds in groups of 5 the following results were obtained.

x :	0,	1,	2,	3,	4,	5,	**Total**
f :	10,	70,	150,	160,	80,	30,	500.

where f denotes the number of groups in which x seeds germinated. Fit a binomial distribution to the data.

[Ans. : $\bar{x} = 2{\cdot}64,\ n = 5,\ p = \frac{\bar{x}}{n} = 0{\cdot}528,\ q = 0{\cdot}472.$

Frequency 12, 65, 147, 164, 92, 20]

3. Five dice are thrown together 96 times. The number of times 4, 5 or 6 was actually obtained is given below. Fit a binomial distribution if (i) dice are unbiased (ii) the nature of the dice is not known.

No. of dice showing 4, 5, 6 :	0	1,	2,	3,	4,	5.
Frequency :	1,	10,	24,	35,	18,	8.

[Ans. : (i) $p = \frac{3}{6} = \frac{1}{2},\ q = \frac{1}{2}\ \therefore\ 96\left(\frac{1}{2}+\frac{1}{2}\right)^5$ *i.e.* 3, 15, 30, 30, 15, 3.

(ii) $\bar{x} = 2{\cdot}86,\ p = \frac{\bar{x}}{n} = \frac{2{\cdot}86}{5} = 0{\cdot}572,\ q = 1 - p = 0{\cdot}428$ *i.e.*, 1, 9, 25, 33, 22, 6]

4. Four coins were tossed 160 times and the following results were obtained,

No. of heads :	0,	1,	2,	3,	4.
Frequency :	17,	52,	54,	31,	6.

Fit a Binomial distribution. **(S.U. 2000) [Ans. : 17, 51, 58, 29, 5]**

5. Fit a Binomial distribution to the following data,

Value of x :	0,	1,	2,	3,	4,	5,	6.
Frequency f :	6,	20,	28,	12,	8,	6,	0.

(S.U. 1998) [Ans. : 6, 18, 26, 20, 8, 2, 0]

3. Poisson Distribution

Poisson distribution was discovered by the French Mathematician Poisson in 1837. Poisson distribution is **the limiting case of the binomial distribution** under the following conditions :

(i) n, the number of trials is infinitely large *i.e.* $n \to \infty$.

(ii) p, the probability of success in each trial is constant and infinitely small *i.e.* $p \to 0$.

(iii) np, the average success is finite say m, *i.e.* $np = m$.

Siméon Denis Poisson (1781 - 1840)

A French mathematician geometer and physicist was born in Pithiviers, Loiret in France. He was a favourite student of Lagrange and was treated like a son by Laplace. Lagrange and Laplace were his doctoral advisors. In 1806 he became full professor at École Polytechnique, in Paris succeeding Fourier. His notable students were Dirichlet and Liouville. As a teacher of mathematics Poisson is said to have been extremely successful. As a scientific worker, his productivity has rarely, if ever, been equalled. Inspite of his many official duties, he published more than three hundred works, several of them extensive treatises and many of them memoires dealing with the most abtruse branches of pure mathematics, applied mathematics, mathematical physics and rational mechanics. His memoires on the theory of electricity and magnetism created a new branch of mathematical physics. He made important advances in planetary theory. Poisson is known for Poisson distribution, Poisson process, Poisson equation, Poisson Kernel, Poisson regression, Poisson summation formula, Poisson ratio, Euler-Poisson-Darboux equation, Conway-Maxwell-Poisson distribution.

(a) To Derive Poisson Distribution

Consider $p(x) = {}^nC_x p^x q^{n-x}$

$$= {}^nC_x \left(\frac{p}{q}\right)^x q^n = {}^nC_x \left(\frac{p}{1-p}\right)^x (1-p)^n$$

Putting $p = \frac{m}{n}$,

$$p(x) = \frac{n(n-1)(n-2).....(n-x+1)}{x!} \frac{(m/n)^x}{[1-(m/n)]^x}\left[1-\frac{m}{n}\right]^n$$

$$p(x) = \frac{\left[1-\frac{1}{n}\right]\left[1-\frac{2}{n}\right].....\left[1-\frac{x-1}{n}\right]m^x\left[1-\frac{m}{n}\right]^n}{x!\ \left[1-\frac{m}{n}\right]^x}$$

Since $\lim\limits_{n\to\infty}\left[1-\frac{m}{n}\right]^n = e^{-m}$ and $\lim\limits_{n\to\infty}\left[1-\frac{m}{n}\right]^x = 1$

Taking the limits of both sides as $n \to \infty$

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

Thus, the limit of the Binomial random variable is the Poisson random variable.

(b) Definition

A random variable X is said to follow **Poisson distribution** if the probability of x is given by

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

and m (> 0) is called the parameter of the distribution.

Remarks ....

1. The sum of the probabilities is 1.

$$\sum_{x=0} P(X = x) = \sum_{x=0} \frac{e^{-m} m^x}{x!} = e^{-m} \sum \frac{m^x}{x!}$$

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

2. Poisson distribution occurs where the probability of occurrence p is very small and the number of trials n is very large and where the probability of occurrence only can be known e.g. the number of accidents, the number of deaths by a disease, the number of printing mistakes on a page etc. In these cases we can only observe the number of successes but the number of failures cannot be observed. We can observe how many accidents occur; we cannot observe how many times accidents do not occur.

(c) When do we get Poisson distribution ?

As seen before we get a Poisson distribution if the following conditions are satisfied.

1. The number of trials n is infinitely large *i.e.* $n \to \infty$.
2. A trial results in only two ways-success or failure.
3. If p and q are probabilities of success and failure, then $p + q = 1$.
4. These probabilities are mutually exclusive, exhaustive but not necessarily equally likely (*i.e.*, p is not necessarily equal to 1 / 2.)
5. The probability p of success is very small *i.e.* $p \to 0$.
6. $n \to \infty$ and $p \to 0$ such that $np = m$, (> 0) a constant.

(d) Uses

Poisson distribution is used in problems involving :

(i) the number of deaths due to a disease such as heart attack, cancer etc.

(ii) the number of accidents during a week or a month etc.

(iii) the number of phone-calls received at a particular telephone exchange during a period of time.

(iv) the number of cars passing a particular point on a road during a period of time.

(v) the number of printing mistakes on a page of a book etc.

Note ....

Since the Poisson distribution is the limiting case of Binomial distribution, we can calculate binomial probabilities approximately by using Possion distribution whenever n is large and p is small.

(e) Constants of Poisson Distribution

We state below without proof the following two properties of Poisson Distribution.

$$\boxed{\begin{array}{l}\text{(i) The mean} = m \\ \text{(ii) The variance} = m\end{array}}$$

Mean and Variance of Poisson's distribution are both equal to m.

Example 1 : Find out the fallacy if any in the following statement "The mean of a Poisson distribution is 2 and the variance is 3".

Sol. : In a Poisson distribution the mean and variance are same. Hence, the above statement is false.

Example 2 : If the mean of the Poisson distribution is 4, find $P(m - 2\sigma < X < m + 2\sigma)$.

Sol. : For Poisson distribution mean = variance = m.

Hence, $m = 4$ and $\sigma = 2$

$$\therefore \quad P(m - 2\sigma < X < m + 2\sigma) = P(0 < X < 8)$$

$$= P[X = 1, 2, 3, \ldots\ldots, 7]$$

But $\quad P(X) = e^{-m} \dfrac{m^x}{x!} = e^{-4} \dfrac{4^x}{x!}$

$$\therefore \text{ Required probability} = e^{-4}\left[\frac{4}{1!} + \frac{4^2}{2!} + \frac{4^3}{3!} + \frac{4^4}{4!} + \frac{4^5}{5!} + \frac{4^6}{6!} + \frac{4^7}{7!}\right] = 0{\cdot}93$$

Example 3 : If the variance of a Poisson distribution is 2, find the probabilities of $r = 1, 2, 3, 4$ from the recurrence relation of Poisson distribution.

Sol. : We have $P(x) = e^{-m} \dfrac{m^x}{x!}$

Since variance = $m = 2$ by data.

$$P(x) = e^{-2} \frac{2^x}{x!} \quad \text{when } x = 0, \; P(0) = e^{-2}$$

Now, the recurrence relation is $\quad P(x+1) = \dfrac{m}{x+1} P(x)$

Putting $x = 0$, $\quad P(1) = \dfrac{2}{1} P(0) = 2e^{-2}$

Putting $x = 1$, $\quad P(2) = \dfrac{2}{2} P(1) = \dfrac{2}{2} \cdot 2e^{-2} = e^{-2}$

Putting $x = 2$, $\quad P(3) = \dfrac{2}{3} P(3) = \dfrac{2}{3} P(2) = \dfrac{2}{3} e^{-2}$

Putting $x = 3$, $\quad P(4) = \dfrac{2}{4} P(3) = \dfrac{1}{3} e^{-2}$

Example 4 : Find out the fallacy if any in the following statement.

"If X is a Poisson variate such that $P(X=2) = 9P(X=4) + 90P(X=6)$ then mean of $X = 1$."

Sol. : Let m be the mean of X. $\therefore \quad P(X=x) = e^{-m}\dfrac{m^x}{x!}$

By data $e^{-m} \cdot \dfrac{m^2}{2!} = 9 \cdot e^{-m} \cdot \dfrac{m^4}{4!} + 90 \cdot e^{-m} \cdot \dfrac{m^6}{6!}$

$\therefore \quad \dfrac{1}{2} = \dfrac{3m^2}{8} + \dfrac{m^4}{8} \quad \therefore \; m^4 + 3m^2 - 4 = 0$

$\therefore \quad (m^2+4)(m^2-1) = 0 \quad \therefore \; m^2 = -4$ or $m^2 = 1$

$\therefore$ The mean is 1 since $m > 0$.

$\therefore$ The statement is correct.

Example 5 : A car hire firm has two cars which it hires out day by day. The number of demands for a car on each day is distributed as Poisson variate with mean 1·5. Calculate the proportion of days on which (i) neither car is used, (ii) some demand is refused.

Sol. : We have $P(x) = e^{-m}\dfrac{m^x}{x!} = \dfrac{e^{-1\cdot5} \cdot (1\cdot5)^x}{x!}, \quad x = 0, 1, 2, \ldots..$

(i) Probability that there is no demand is

$$P(X=0) = e^{-1\cdot5}\frac{(1\cdot5)^0}{0!} = 0\cdot2231$$

(ii) Probability that some demand is refused means there was demand for more than two cars.

$$\therefore \quad P(X>2) = P(X=3) + P(X=4) + \ldots..$$
$$= 1 - [P(X=0) + P(X=1) + P(X=2)]$$
$$= 1 - \left[e^{-1\cdot5}\frac{(1\cdot5)^0}{0!} + e^{-1\cdot5}\frac{(1\cdot5)^1}{1!} + e^{-1\cdot5}\frac{(1\cdot5)^2}{2!}\right]$$
$$= 1 - [0\cdot2231 + 0\cdot3347 + 0\cdot2510] = 0\cdot1912.$$

$\therefore$ Proportion of days on which (i) neither car is used is 0·2231.

(ii) some demand is refused is 0·1912.

Example 6 : If a random variable X follows Poisson distribution such that $P(X=1) = 2P(X=2)$, find the mean and the variance of the distribution. Also find $P(X=3)$.

Sol. : Let the parameter of the Poisson distribution be m.

$$\therefore \quad P(X=x) = \frac{e^{-m} \times m^x}{x!}$$

We are given that $P(X=1) = 2P(X=2)$

$$\therefore \quad \frac{e^{-m} \times m^1}{1!} = \frac{2e^{-m} \times m^2}{2!} \qquad \therefore \; m = 1$$

$\therefore$ The mean and the variance = 1.

Now, $P(X=3) = \dfrac{e^{-m} \cdot m^x}{x!} = \dfrac{e^{-1} \cdot 1^3}{3!} = 0\cdot0613.$

Example 7 : A hospital switch board receives an average of 4 emergency calls in a 10 minutes interval. What is the probability that (i) there are atleast 2 emergency calls, (ii) there are exactly 3 emergency call in an interval of 10 minutes ?

Sol. : We have $P(x) = \dfrac{e^{-m} \cdot m^x}{x!}$. Here, $m = 4$.

(i) $$P(X \le 2) = P(X=0) + P(X=1) + P(X=2)$$
$$= \frac{e^{-4} \cdot 4^0}{0!} + \frac{e^{-4} \cdot 4^1}{1!} + \frac{e^{-4} \cdot 4^2}{2!}$$
$$= e^{-4}(1+4+8) = 0\cdot238$$

(ii) $$P(X=3) = \frac{e^{-m} \cdot m^x}{x!} = \frac{e^{-4} \cdot 4^3}{3!} = 0\cdot195$$

Example 8 : An insurance company found that only 0·01% of the population is involved in a certain type of accident each year. If its 1000 policy holders were randomly selected from the population, what is the probability that no more than two of its clients are involved in such accident next year ?

Sol. : We have $p = \dfrac{0\cdot01}{100} = 0\cdot0001, \; n = 1000.$

$$\therefore \quad m = np = 1000 \times 0\cdot0001 = 0\cdot1$$

$$\therefore \quad P(X=x) = e^{-0\cdot1}\frac{(0\cdot1)^x}{x!}$$

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

$$\therefore \quad P(X \le 2) = e^{-0\cdot1}\left[\frac{(0\cdot1)^0}{0!} + \frac{(0\cdot1)^1}{1!} + \frac{(0\cdot1)^2}{2!}\right] = 0\cdot9998$$

Example 9 : Find the probability that atmost 4 defective bulbs will be found in a box of 200 bulbs if it is known that 2 percent of the bulbs are defective. (Given $e^{-4} = 0\cdot0183$).

Sol. : Since probability of a defective bulb is small we can use Poisson distribution.

We have, $\therefore \quad m = np = 200 \times 0\cdot02 = 4$

$$\therefore \quad P(X=x) = \frac{e^{-4} \times 4^x}{x!}$$

$$\therefore \quad P(X \le 4) = p(0) + p(1) + p(2) + p(3) + p(4)$$
$$= \frac{e^{-4} \times 4^0}{0!} + \frac{e^{-4} \times 4^1}{1!} + \frac{e^{-4} \times 4^2}{2!} + \frac{e^{-4} \times 4^3}{3!} + \frac{e^{-4} \times 4^4}{4!}$$
$$= e^{-4}\left[1 + \frac{4}{1!} + \frac{16}{2!} + \frac{64}{3!} + \frac{256}{4!}\right]$$
$$= e^{-4} \times \frac{103}{3} = 0\cdot0183 \times \frac{103}{3} = 0\cdot6283.$$

Example 10 : Find the probability that at most 4 defective bulbs will be found in a box of 200 bulbs if it is known that 2 percent of the bulbs are defective. (Given $e^{-4} = 0\cdot0183$).

Sol. : Since probability of a defective bulbs is small we can use Poisson distribution.

We have, $\therefore\ m = np = 200 \times 0{\cdot}02 = 4$

$$\therefore\ P(X = x) = \frac{e^{-4} \times 4^x}{x!}$$

$$\therefore\ P(X \le 4) = p(0) + p(1) + p(2) + p(3) + p(4)$$

$$= \frac{e^{-4} \times 4^0}{0!} + \frac{e^{-4} \times 4^1}{1!} + \frac{e^{-4} \times 4^2}{2!} + \frac{e^{-4} \times 4^3}{3!} + \frac{e^{-4} \times 4^4}{4!}$$

$$\therefore\ P(X \le 4) = e^{-4}\left[1 + \frac{4}{1!} + \frac{16}{2!} + \frac{64}{3!} + \frac{256}{4!}\right]$$

$$= e^{-4} \times \frac{103}{3} = 0{\cdot}0183 \times \frac{103}{3} = 0{\cdot}6283.$$

Example 11 : The number of accidents in a year attributed to taxi drivers in a city follows Poisson distribution with mean 3. Out of 1,000 taxi drivers, find approximately the number of drivers with (i) no accident in a year, (ii) more than 3 accidents in a year. (Given : $e^{-1} = 0{\cdot}3679$, $e^{-2} = 0{\cdot}1353$, $e^{-3} = 0{\cdot}0498$)

Sol. : For a Poisson variate

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

We are given $m = 3$

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

$$\therefore\ P(X = 0) = \frac{e^{-3} \times (3)^0}{0!} = e^{-3} = 0{\cdot}0498$$

$$P(X = 1) = \frac{e^{-3} \times (3)^1}{1!} = 0{\cdot}0498 \times 3 = 0{\cdot}1494$$

$$P(X = 2) = \frac{e^{-3} \times (3)^2}{2!} = 0{\cdot}0498 \times \frac{9}{2} = 0{\cdot}2241$$

$\therefore$ Expected number of drivers with no accidents

$$= N \times p(0) = 1{,}000 \times 0{\cdot}0498 = 49{\cdot}8 = 50 \text{ nearly.}$$

$\therefore\ p(0, 1, 2, 3 \text{ accidents}) = p(0) + p(1) + p(2) + p(3)$

$$= 0{\cdot}0498 + 0{\cdot}1494 + 0{\cdot}2241 = 0{\cdot}4233$$

$\therefore\ p(\text{more than 3 accidents}) = 1 - 0{\cdot}4233 = 0{\cdot}5767.$

Expected number of drivers with more than 3 accidents

$$= Np = 1{,}000 \times 0{\cdot}5767$$

$$= 576{\cdot}7 = 577 \text{ nearly.}$$

Example 12 : In a certain factory turning out blades, there is a small chance 1 / 500 for any blade to be defective. The blades are supplied in packets of 10. Use the Poisson distribution to calculate the approximate number of packets containing no defective, one defective, two defective blades in a consignment of 10,000 packets. (Given $e^{-0{\cdot}02} = 0{\cdot}9802$)

Sol. : We have, $n = 10$, $p = \dfrac{1}{500}$. $\therefore\ m = np = 10 \times \dfrac{1}{500} = 0{\cdot}02$

$$\therefore\ P(X = x) = \frac{e^{-0{\cdot}02} \times (0{\cdot}02)^x}{x!}$$

$$\therefore\ P(X = 0) = \frac{e^{-0{\cdot}02} \times (0{\cdot}02)^0}{0!} = e^{-0{\cdot}02} = 0{\cdot}9802$$

$$P(X = 1) = \frac{e^{-0{\cdot}02} \times (0{\cdot}02)^1}{1!} = e^{-0{\cdot}02} \times 0{\cdot}02 = 0{\cdot}0196$$

$$P(X = 2) = \frac{e^{-0{\cdot}02} \times (0{\cdot}02)^2}{2!} = e^{-0{\cdot}02} \times 0{\cdot}0002 = 0{\cdot}0002$$

$\therefore$ Expected freq. of no defective $= 10000 \times 0{\cdot}9802 = 9802$

Expected freq. of one defective $= 10000 \times 0{\cdot}0196 = 196$

Expected freq. of two defective $= 10000 \times 0{\cdot}0002 = 2.$

Example 13 : Between 2 and 4 p.m. the average number of phone calls per minute coming into a switch board of a company is 2·5. Find the probability during a minute there will be (i) no phone call (ii) exactly 3 calls. **(S.U. 1985)**

Sol. : For a Poisson distribution $P(x) = \dfrac{e^{-m} m^x}{x!}$

We are given that the mean, m = average number of phone calls = 2·5.

(i) $\therefore\ P(x = 0) = \dfrac{e^{-2{\cdot}5}(2{\cdot}5)^0}{0!} = 0{\cdot}08208$

(ii) $P(x = 3) = e^{-2{\cdot}5}\dfrac{(2{\cdot}5)^3}{3!} = 0{\cdot}2138$

Example 14 : A firm has two cars which it hires out day by day. The number of demands for a car on each day is distributed as a Poisson variate with mean 1·5. Calculate the probable number of days in a year on which (i) neither car is in demand, (ii) a demand is refused. **(S.U. 1985, 2005)**

Sol. : Let x be the number of cars in demand. Since x is Poisson with mean $m = 1{\cdot}5$.

$$P(x) = e^{-1{\cdot}5}\frac{(1{\cdot}5)^x}{x!}$$

$$P(x = 0) = e^{-1{\cdot}5}\frac{(1{\cdot}5)^0}{0!} = 0{\cdot}2231; \qquad P(x = 1) = e^{-1{\cdot}5}\frac{(1{\cdot}5)^1}{1!} = 0{\cdot}3347;$$

$$P(x = 2) = e^{-1{\cdot}5}\frac{(1{\cdot}5)^2}{2!} = 0{\cdot}2510$$

$\therefore\ P(\text{demand is refused}) = P(\text{demand greater than 2})$

$$= P(x > 2) = 1 - [P(x = 0) + P(x = 1) + P(x = 2)]$$

$$= 1 - [0{\cdot}2231 + 0{\cdot}3347 + 0{\cdot}2510] = 0{\cdot}1912$$

$\therefore$ No. of days on which there is no demand in a year

$$= Np\ (x = 0) = 365 \times 0{\cdot}2231 = 81 \text{ nearly.}$$

No. of days on which demand is more than 2

$$= Np\ (x > 2) = 365 \times 0{\cdot}1912 = 70 \text{ nearly.}$$

Example 15 : 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 (i) exactly 3, (ii) more than 2 will suffer a bad reaction. **(S.U. 1987, 94, 95, 97, 99)**

Sol. : We are given $n = 2000$, $p = 0{\cdot}001$. But $m = np = 2000 \times 0{\cdot}001 = 2$.

By Poisson Distribution $P(x) = \dfrac{e^{-m} m^x}{x!}$

$$\therefore \quad P(x) = \frac{e^{-2}(2)^x}{x!}$$

$$\therefore \quad P(x=3) = \frac{e^{-2}(2)^3}{3!} = 0{\cdot}1804 \; ; \qquad P(x=0) = \frac{e^{-2}(2)^0}{0!} = 0{\cdot}1353 \; ;$$

$$P(x=1) = \frac{e^{-2}(2)^1}{1!} = 0{\cdot}2707 \; ; \qquad P(x=2) = \frac{e^{-2}(2)^2}{2!} = 0{\cdot}2707 \, .$$

$$\begin{aligned} \therefore P(\text{more than 2 suffer}) &= 1 - [P(x=0) + P(x=1) + P(x=2)] \\ &= 1 - [0{\cdot}1353 + 0{\cdot}2707 + 0{\cdot}2707] \\ &= 0{\cdot}3233. \end{aligned}$$

Example 16 : Using Poisson distribution, find the probability that ace of spade will be drawn from a pack of well shuffled cards at least once in 104 consecutive draws. **(S.U. 1991, 93, 97, 2003)**

Sol. : P (getting ace of spade in a single draw) = 1/ 52

Mean $= n \times p = 104 \times (1/52) = 2$

$\therefore$ By Poisson Distribution $P(x) = e^{-2} \dfrac{(2)^x}{x!}$

$\therefore \quad P(x=0) = 0{\cdot}1353$

$\therefore \quad P$ (at least one) $= 1 - P(x=0) = 1 - 0{\cdot}1353 = 0{\cdot}8647$.

Example 17 : In a certain factory producing cycle tyres there is a small chance 1 in 500 for any tyre to be defective. The tyres are supplied in lots of 20. Using Poisson's' distribution, calculate the approximate number of lots containing no defective, one defective and two defective tyres respectively in a consignment of 20,000 tyres. **(S.U. 1988, 9**

Sol. : For Poisson distribution

Mean $m = np \quad \therefore \quad m = 20 \cdot \dfrac{1}{500} = 0{\cdot}04$

$$\therefore \quad P(x) = \frac{e^{-m} m^x}{x!} = e^{-0{\cdot}04} \frac{(0{\cdot}04)^x}{x!}$$

$$P(\text{zero defective, } x=0) = e^{-0{\cdot}04} \frac{(0{\cdot}04)^0}{0!} = 0{\cdot}9608$$

$$P(\text{one defective, } x=1) = e^{-0{\cdot}04} \frac{(0{\cdot}04)^1}{1!} = 0{\cdot}0384$$

$$P(\text{two defective, } x=2) = e^{-0{\cdot}04} \frac{(0{\cdot}04)^2}{2!} = 0{\cdot}0008$$

Number of lots having

Zero defectives $= Np = 1000 \times 0{\cdot}9608 = 960{\cdot}8 = 961$.

One defective $= Np = 1000 \times 0{\cdot}0384 = 38{\cdot}4 = 38$

Two defectives $= Np = 1000 \times 0{\cdot}0008 = 0{\cdot}8 = 1$.

Example 18 : A firm produces articles of which 0·1 per cent are defective. It packs them in cases each containing 500 articles. If a whole-saler purchases 100 such cases, how many cases can be expected to be free from defectives, how many can be expected to have one defective ? **(S.U. 1985)**

Sol. : We have $p = \dfrac{0{\cdot}1}{100} = 0{\cdot}001$, $n = 500$

$\therefore$ Mean $m = np = 500 \times 0{\cdot}001 = 0.5$

By Poisson distribution

$$P(x) = \frac{e^{-m} m^x}{x!} = e^{-0{\cdot}5} \frac{(0{\cdot}5)^x}{x!}$$

$$\therefore \quad P(\text{no defective}) = P(x=0) = e^{-0{\cdot}5} \frac{(0{\cdot}5)^0}{0!} = 0{\cdot}6065$$

$$P(\text{one defective}) = P(x=1) = e^{-0{\cdot}5} \frac{(0{\cdot}5)^1}{1!} = 0{\cdot}3033$$

$\therefore$ Expected No. of no defective $= Np = 100 \times 0{\cdot}6065 = 61$

Expected No. of one defective $= Np = 100 \times 0{\cdot}3033 = 30$.

Example 19 : Fit a Poisson distribution to the following data :

x :	0,	1,	2,	3,	4,	Total
f :	109,	65,	22,	3,	1,	200.

(S.U. 1997, 98, 2003, 04)

Sol. : Fitting of Poisson's distribution means finding expected frequencies using Poisson's distribution.

Now, $m = \text{mean} = \dfrac{\Sigma fx}{\Sigma f}$

$$\therefore \quad m = \frac{109 \times 0 + 65 \times 1 + 22 \times 2 + 3 \times 3 + 1 \times 4}{200} = \frac{122}{200} = 0{\cdot}61$$

Now, Expected frequency $= N \cdot P(x)$

$$= N \cdot \frac{e^{-m} m^x}{x!} = 200 \cdot \frac{e^{-0{\cdot}61}(0{\cdot}61)^x}{x!}$$

Putting $x = 0, 1, 2, 3, 4$, we get, the expected frequencies as 109, 66, 20, 4, 1.

Example 20 : Fit a Poisson distribution to the following data

No. of deaths :	0,	1,	2,	3,	4.
Frequencies :	123,	59,	14,	3,	1.

Sol. : Fitting Poisson distribution means finding expected frequencies of $X = 0, 1, 2, 3, 4$.

Now, mean $= \dfrac{\Sigma f_i x_i}{\Sigma f_i} = m$

$\therefore$ Mean $= 123(0) + 59(1) + 14(2) + 3(3) + 1(4)$

$$= \frac{100}{200} = 0{\cdot}5$$

$\therefore$ Poisson distribution of X is

$$P(X = x) = \frac{e^{-m} \times m^x}{x!} = \frac{e^{-0{\cdot}5} \times (0{\cdot}5)^x}{x!} \quad \text{.....(1)}$$

Expected frequency $= N \times p(x)$

$$= 200 \times \frac{e^{-0{\cdot}5} \times (0{\cdot}5)^x}{x!}$$

Putting $x = 0, 1, 2, 3, 4$ we get the expected frequencies as 121, 61, 15, 2, 1.

Or putting $x = 0$ in (1).

$$P(X = 0) = e^{-0{\cdot}5} \frac{(0{\cdot}5)^0}{0!} = 0{\cdot}6065$$

$\therefore$ Expected frequency $f(0) = Np = 200 \times 0{\cdot}6065 = 121$.

But $\quad f(x+1) = \dfrac{m}{x+1} \cdot f(x) = \dfrac{0{\cdot}5}{x+1} \cdot f(x)$

Putting $x = 0$, $\quad f(1) = \dfrac{0{\cdot}5}{1} \cdot 121 = 61$. $\qquad$ Putting $x = 1$, $f(2) = \dfrac{0{\cdot}5}{2} \cdot 60 = 15$.

Putting $x = 2$, $\quad f(3) = \dfrac{0{\cdot}5}{3} \cdot 15 = 3$. $\qquad$ Putting $x = 3$, $f(4) = \dfrac{0{\cdot}5}{4} \cdot 3 = 1$.

Example 21 : Letters were received in an office on each of 100 days. Fit a Poisson distribution and find the expected frequencies for $x = 0$ and 1. (Given : $e^{-4} = 0{\cdot}0183$).

Number of letters :	0,	1,	2,	3,	4,	5,	6,	7,	8,	9,	10.
Frequency :	1,	4,	15,	22,	21,	20,	8,	6,	2,	0,	1.

Sol. : $\Sigma f_i = 1 + 4 + 15 + ... + 1 = 100$

$$m = \frac{\Sigma f_i x_i}{\Sigma f_i} = \frac{1 \times 0 + 4 \times 1 + 15 \times 2 + ... + 1 \times 10}{100} = \frac{400}{100} = 4$$

$$\therefore \; P(X = x) = \frac{e^{-m} \cdot m^x}{x!} = \frac{e^{-4} 4^x}{x!} \qquad \therefore \; p(0) = e^{-4} = 0{\cdot}0183$$

Now, expected frequency $f(0) = Np = 100 \times 0{\cdot}0183 = 1{\cdot}83 = 2$

But $\quad f(x+1) = \dfrac{m}{x+1} \cdot f(x) \quad \therefore \; f(1) = \dfrac{4}{1}(1{\cdot}83) = 7{\cdot}32 = 7$

EXERCISE - II

(A) **1.** Find the fallacy if any, in the following statement :

(a) The mean of a Poisson distribution is 2 and the variance is 3.

(b) The mean of a Poisson distribution is 4 and the standard deviation is 2.

[**Ans. : (a)** False, **(b)** True. In a Poisson distribution the mean = variance]

2. If a Poisson distribution is such that

(a) $P(x = 2) = P(x = 3)$ find m and $P(x = 0)$, $P(x = 4)$

(b) $P(x = 0) = 0{\cdot}1$ find m and $P(x = 1)$ [**Ans. : (a)** 3, 0·0498, 0·168, **(b)** 2·3026, 0·23]

(B) **1.** Accidents occur on a particular stretch of highway at an average rate 3 per week. What is the probability that there will be exactly two accidents in a given week ? [**Ans. :** 0·224]

2. Find the probability that at most 4 defective bulbs will be found in a box of 400 bulbs, if it is known that 1 per cent of the bulbs are defective. [**Ans. :** 0·6282]

3. It is 1 in 1000 that an article is defective. There are in a box 100 articles of this type. Assuming Poisson distribution find the probability that the box contains (i) no detective, (ii) two or more detectives. **(S.U. 1998)** [**Ans. :** 0·3679, 0·2642]

4. If the probability that an individual suffers a bad reaction from a particular injection is 0·001, determine the probability that out of 2,000 individuals (i) exactly four, (ii) more than two individuals will suffer a bad reaction. [**Ans. : (i)** 0·0902, **(ii)** 0·3233]

5. The number of accidents on a particular highway in a month is a Poisson variate with parameter 5. Find the probability that more than 2 accidents have occurred on the road in a given month. [**Ans. :** 0·8754]

6. A book contains 100 misprints distributed randomly throughout 100 pages. What is the probability that a page observed at random contains at least two misprints ? **(S.U. 1988)** [**Ans. :** 0·2642]

7. It is known that in a certain plant, there are on an average 4 industrial accidents per month. Find the probability that in a given month there will be less than 4 accidents. [**Ans. :** 0·4335]

8. An insurance company found that only 0·01 per cent of the population is involved in a certain type of accident each year. If its 1000 policy-holders were randomly selected from the population, what is the probability that more than two of its clients are involved in such an accident next year ? **(S.U. 1990)** [**Ans. :** 0·0002]

9. The number of telephone calls received by an operator in a ten-minute interval follows a Poisson distribution with mean 1. What is the probability of receiving (i) no call, (ii) less than four calls, in the interval. What is the most likely number of calls ? **(S.U. 1986)** [**Ans. : (i)** 0·3679, **(ii)** 0·981, **(iii)** 1, the mean]

10. If the chance of being killed by flood during a year is 1/300, using Poisson distribution find the probability that out of 3000 persons living in the area at least one would die in flood in that year. **(S.U. 1989)** [**Ans. :** 0·9999]

11. In a screw making machine there are on the average two defective screws out of 100. The screws are packed in boxes of 100. Find the probability that a box contains (i) no defective, (ii) 5 defective screws. **(S.U. 1997)** [**Ans. :** $m = 2$, (i) 0·1353, (ii) 0·036]

12. If 3% of bulbs manufactured by a company are defective, assuming Poisson distribution find the probability that in a pack of 100 bulbs (i) zero bulbs, (ii) two bulbs are defective. **(S.U. 1988, 99)** [**Ans. : (i)** 0·0498, **(ii)** 0·2240]

13. In a town 10 accidents took place in a period of 50 days. Assuming Poisson distribution find the probability that there will be 3 or more accidents per day. **(S.U. 1987)** [**Ans. :** 0·0012]

14. Assume that the probability of an a coal minor being killed in a mine accident during a year is 1/2400. Use Poisson's distribution to find the probability that in a mine exploying 200 workers there will be at least one fatal accident in a year. **(S.U. 2003)** [**Ans. :** 0·08]

(C) **1.** A manufacturer of pins knows that on an average 5% of his product is defective. He sells pins in boxes of 100 and guarantees that no more than 4 pins will be defective. In how many boxes out of 1000, he will meet the guaranteed quality ? [**Ans. :** 440]

2. In a certain factory turning out blades, there is a small chance 1/250 for any blade to be defective. The blades are supplied in packets of 10. Use the Poisson distribution to calculate the approximate number of packets containing no defective, one defective, two defective blades in a consignment of 10,000 packets. [**Ans. : (i)** 9608, **(ii)** 384, **(iii)** 8]

3. In a certain factory producing certain articles the probability that an article is defective is 1/500. The articles are supplied in packets of 10. Find approximately the number of packets containing no defective, one defective, two defectives in a consignment of 20,000 packets.

[**Ans. :** 19604, 392, 4]

4. A manufacturer finds that the average demand per day for the mechanic to repair his new production is 1·5. Over a period of one year the demand per day is distributed as Poisson distribution. He employs two mechanics. On how many days in one year (a) both mechanics would be free, (b) some demand is refused. [**Ans. : (i)** 84·4, **(ii)** 69·8]

5. In a certain factory turning out blades, there is a small chance of 0·002 for any blade to be defective. The blades are supplied in packets of 10. Use Poisson's distribution to calculate the approximate number of packets containing no defective, one defective and two defective blades in a consignment of 10,000 packets. **(S.U. 1985, 87, 90, 91, 2004)**

[**Ans. : (i)** 9802, **(ii)** 196, **(iii)** 2]

(D) **1.** Fit a Poisson's distribution to the following data :

x :	0,	1,	2,	3,	4,	**Total**
f :	192,	100,	24,	3,	1,	320.

(S.U. 1995)

[**Ans. :** $m = 0.5$ approx. Freq. $320 \dfrac{e^{-0\cdot5}(0\cdot5)^x}{x!}$ *i.e.* 194, 97, 24, 4, 1]

2. Fit a Poisson's distribution to the following data

x :	0,	1,	2,	3,	4,	**Total**
f :	112,	63,	20,	4,	1,	200.

[**Ans. :** $m = 0\cdot595$, Frequency 110, 66, 19, 4, 1]

3. The following mistakes per page were observed in a book

No. of mistakes :	0,	1,	2,	3,	4,	**Total**
No. of pages :	211,	90,	19,	5,	0,	325.

Fit a Poisson distribution. **(S.U. 1999)** [**Ans. :** $m = 0\cdot44$, Frequency : 209, 92, 20, 3, 1]

4. Normal Distribution

Normal distribution is one of the most important and commonly used continuous distribution. It was first developed by DeMoivre but it is credited to Gauss who referred to it first in 1809. A large number of continuous variates follow this distribution, hence, the name 'normal'.

A renowned mathematician Poincare' has remarked that "there must be something mysterious about the normal distribution because mathematicians think it is a law of nature whereas physicists are convinced that it is a mathematical theorem."

1. Definition : A continuous random variable X is said to follow **normal distribution** with parameter m (called mean) and σ^2 (called variance), if its probability density function is given by

$$\boxed{f(x) = \frac{1}{\sqrt{2\pi}\cdot\sigma} e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2}} \qquad \begin{array}{l} -\infty < x < \infty \\ -\infty < m < \infty,\ \sigma^2 > 0 \end{array}$$

Remarks ....

(i) A continuous random variate X following normal distribution with mean m and standard deviation σ is referred to as $X \sim N(m, \sigma)$.

(ii) If X is a normal variate with parameter m, σ, then $\boxed{Z = \dfrac{X-m}{\sigma}}$ is also a normal variate with mean = 0 and standard deviation = 1.

It is called **Standard Normal Variate.**

Abraham De Moivre (1667 - 1754)

A French mathematician who made important contributions to statistics, theory of probability and trigonometry. The concept of statistically independent events was first developed by De Moivre. Through the use of complex number he transformed trigonometry from a branch of geometry to a branch of analysis. His treatise on probability has influenced the development of probability theory.

Karl Friedrich Gauss (1777 - 1855)

Karl Friedrich Gauss was a great German mathematician and scientist. He is called the **"prince of mathematicians"**. He is ranked with Isaac Newton and Anchimedes. It is said that Gauss at the age of three had pointed out an error in calculation made by his father while preparing a payroll.

In his doctoral thesis he gave the first complete proof of the fundamental theorem of algebra that a polynomial of nth degree has n roots. He is considered to have laid the founation of number theory. We are all familiar with Gaussian probability distribution (Normal Distribution). He gave first the geometric interpretation of complex numbers, developed the theory of conformal mapping. He did fundamental work in electromagnetism.

He knew many languages and read extensively. It is said that if Gauss had published all of his discoveries the state of mathematics would have advanced by 50 years.

2. Importance of Normal Distribution

(i) The variables such as height, weight, intelligence etc. follow normal distribution.

(ii) Many other distributions occurring in practice such as Binomial, Poisson etc. can be approximated by normal distribution.

(iii) Many of the distributions of sample statistic *e.g.* Sample Mean, Sample Variance tend to normal distribution for large samples.

(iv) Normal distribution has wide applications in Statistical Quality Control.

(v) Errors in measurements of physical quantities follow normal distribution.

(vi) It is also useful in psychological and educational research.

(a) Mean and Variance of the Normal Distribution

$$\boxed{\begin{aligned} \text{Mean} &= m \\ \text{Var. } (X) &= \sigma^2 \end{aligned}}$$

(b) Area Property

"If X is a normal variate with mean m and variance σ^2 and Z is standard normal variate (with mean zero and variance one) then the area under the normal curve of X between $X = m$ and $X = x_1$ is equal to the area under the S.N. Curve of Z between $Z = 0$ to $Z = z_1$ (say, corresponding to x_1)."

Proof : Consider a normal variate X with mean m and variance σ^2.

Then $$P(m \le X \le x_1) = \frac{1}{\sqrt{2\pi} \cdot \sigma} \int_m^{x_1} e^{-\frac{1}{2}\left(\frac{X-m}{\sigma}\right)^2} dx$$

Now, put $\dfrac{X-m}{\sigma} = Z$

When $X = m$, $Z = 0$; when $X = x_1$, $Z = \dfrac{x_1 - m}{\sigma} = z_1$ say

$$\therefore \quad P(m \le X \le x_1) = P(0 \le Z \le z_1) = \frac{1}{\sqrt{2\pi}} \int_0^{z_1} e^{-\frac{1}{2}z^2} dz$$

Thus, the area under the normal curve from m to x_1 is equal to the area under the standard normal curve from 0 to z_1.

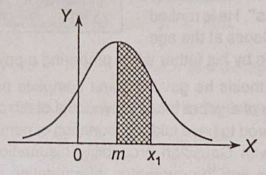


Fig. 6.1

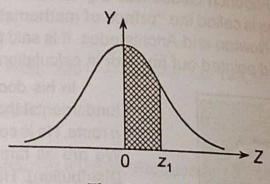


Fig. 6.2

The integral $\dfrac{1}{\sqrt{2\pi}} \int_0^{z_1} e^{-\frac{1}{2}z^2} dz$ is denoted by $\int_0^{z_1} \phi(z)\, dz$ and is known as **normal probability integral**. The areas under standard normal curve from $z = 0$ to various values of z_1 have been calculated and are given in the table at the end of the book.

Reverse Problem : In some problems we know that X is a normal variate with mean m and standard deviation σ. We are required to find the value of $X = x_1$ corresponding to a given probability. Suppose, we want to find of $X = x_1$ such that $P(X > x_1) = \alpha$. In this case also we consider

$$\text{S.N.V. } Z = \frac{X-m}{\sigma}$$

We now consult the area table and find the value of $Z = z_1$ for which area to the right is α *i.e.* area between

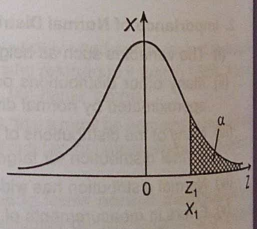


Fig. 6.3

$z = 0$ to $Z = z_1$ is $(0{\cdot}5 - \alpha)$. From this value of z_1 we get X using

$$z_1 = \frac{x_1 - m}{\sigma} \quad \text{i.e.,} \quad x_1 = m + z_1 \sigma.$$

Remarks ....

1. Since standard normal curve is symmetrical about the y-axis it is enough to find the areas to the right. The areas to the left of y-axis at equal distances will be equal.
2. The total area under the curve is unity. Hence, because of symmetry the area under S.N.V. to the right of the y-axis is 0·5.
3. To find the probability that X will lie between x_1 and x_2 $(x_1 < x_2)$, we find the corresponding values of S.N.V. Z (from $Z = \dfrac{X-m}{\sigma}$) say z_1 and z_2 and find the area from z_1 to z_2 and S.N. Curve. The required probability is this area.

$$P(x_1 \le X \le x_2) = P(z_1 \le Z \le z_2)$$
$$= \text{area between } Z = z_1 \text{ and } Z = z_2 \text{ under the S.N. curve.}$$

(c) The Quartile Deviation of a Normal Distribution is (2 / 3) S.D.

Let X be a normal variate with mean m and variance σ^2. Consider the standard normal variate

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

When $X = m$, $Z = 0$ and when $X = Q_3$ the third quartile let $Z = \dfrac{Q_3 - m}{\sigma} = z_1$.

We know that for Normal variate mean = median = mode = m. *i.e.* the mean m divides the area into two equal parts. Hence, the area from Q_1 to m is 0·25. For S.N.V. area from 0 to z_1 is 0·25. For area 0·25 we find from the table that,

$$z_1 = 0{\cdot}6745 = \frac{2}{3}.$$

$$\therefore \quad Q_3 = m + \sigma z_1 = m + \frac{2}{3}\sigma. \quad \text{Similarly, } Q_1 = m - \sigma z_1 = m - \frac{2}{3}\sigma.$$

$$\therefore \quad \boxed{\text{Quartile Deviation} = \frac{Q_3 - Q_1}{2} = \frac{2}{3}\sigma.}$$

(d) Properties of the Normal Distribution

We summaries below important properties of normal distribution which we have proved above. They are given in terms of the properties of the normal curve.

(i) The normal curve is bell-shaped and symmetrical about the maximum ordinate at $x = m$, the *mean*. In other words, the curve is divided into two equal parts by this ordinate. The curve on one side of this ordinate is the mirror image of the curve on the other side.

(ii) The curve has maximum height at $x = m$. Hence the *mode* of the distribution is also m. The ordinate $x = m$ divides the area under the curve into two equal parts. Hence the *median* of the distribution is also m. Thus, for the normal distribution,

$$\boxed{\text{mean} = \text{median} = \text{mode} = m}$$

(iii) The height of the curve goes on decreasing on either side of the ordinate at $x = m$ but never becomes zero. In other words, the curve never intersects the x-axis at any finite point. The x-axis touches it at infinity.

(iv) Since the curve is symmetrical about mean, the first quartile Q_1 and the third quartile Q_3 are at the same distance on the two sides of the mean. The distance of any quartile from the mean is 0·6745 σ units. Hence,

$$Q_1 = m - 0{\cdot}6745\,\sigma \quad \text{or} \quad \boxed{Q_1 = m - \frac{2}{3}\sigma} \quad \text{and}$$

$$Q_3 = m + 0{\cdot}6745 \quad \text{or} \quad \boxed{Q_3 = m + \frac{2}{3}\sigma}$$

Hence, middle 50 % items lie between $m - \frac{2}{3}\sigma$ and $m + \frac{2}{3}\sigma$.

Further, the quartile deviation of a normal distribution is given by

$$\text{Q.D.} = \frac{Q_3 - Q_1}{2} = \frac{\left(m + \frac{2}{3}\sigma\right) - \left(m - \frac{2}{3}\sigma\right)}{2} = \frac{2}{3}\sigma.$$

$$\boxed{\therefore\ \text{Q.D.} = \frac{2}{3}\sigma}$$

(v) The mean of the distribution is m and the standard deviation is σ.

(vi) The mean deviation is given by $\boxed{\text{M.D.} = \frac{4}{5}\sigma}$

(vii) Now, $\dfrac{\text{Q.D.}}{\text{M.D.}} = \dfrac{(2/3)\sigma}{(4/5)\sigma} = \dfrac{2}{3} \times \dfrac{5}{4} = \dfrac{10}{12}$

Also $\text{M.D.} = \dfrac{4}{5}\sigma \quad \therefore \dfrac{\text{M.D.}}{\sigma} = \dfrac{4}{5} = \dfrac{12}{15}$

$\therefore$ $\boxed{\text{Q.D. : M.D. : S.D.} = 10 : 12 : 15}$

(viii) Odd central moments are zero *i.e.*

$$\mu_{2r+1} = 0 \quad \text{for } (r = 0, 1, 2, \ldots..)$$

and $\mu_{2r} = 1 \cdot 3 \cdot 5 \ldots..(2r-1)\,\sigma^{2r} (r = 0, 1, 2, \ldots..)$

i.e., $\mu_2 = \sigma^2,\ \mu_4 = 3\sigma^4,\ \mu_6 = 15\sigma^6$

(ix) The area under the normal curve is distributed as follows,

(a) The area between $x = m - \sigma$ and $x = m + \sigma$ is 68·27 %.

(b) The area between $x = m - 2\sigma$ and $x = m + 2\sigma$ is 95·45 %.

(c) The area between $x = m - 3\sigma$ and $x = m + 3\sigma$ is 99·73 %.

These areas under the normal curve and standard normal curve are shown below.

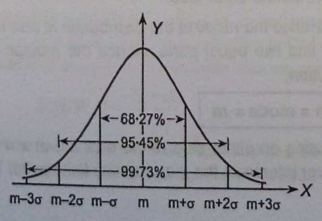


Fig. 6.4 (a)

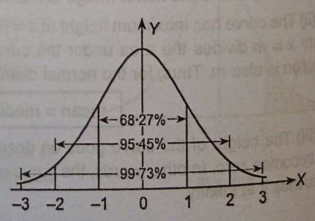


Fig. 6.4 (b)

Type I

Example 1 : For a normal distribution the mean is 50 and the standard deviation is 15. Find (i) Q_1 and Q_3, (ii) mean deviation (Also the interquartile range).

Sol. : (i) For a normal distribution

$$Q_1 = m - \frac{2}{3}\sigma = 50 - \frac{2}{3}15 = 40$$

Again $\quad Q_3 = m + \frac{2}{3}\sigma = 50 + \frac{2}{3}15 = 60$

(ii) The mean deviation of the normal distribution is,

$$\text{M.D.} = \frac{4}{5}\sigma = \frac{4}{5} \times 15 = 12$$

$\therefore$ Interquartile range $= Q_3 - Q_1 = 60 - 40 = 20.$

Example 2 : The first and the third quartiles of a normal distribution are 36 and 44. Find mean, standard deviation and the mean deviation.

Sol. : We have $\quad Q_1 = m - \frac{2}{3}\sigma$ and $Q_3 = m + \frac{2}{3}\sigma$

$\therefore\ 36 = m - \frac{2}{3}\sigma$ and $44 = m + \frac{2}{3}\sigma$

Adding $80 = 2m \quad \therefore\ m = 40$

Then, $36 = 40 - \frac{2}{3}\sigma \quad \therefore \frac{2}{3}\sigma = 4 \quad \therefore \sigma = 6$

And mean deviation $= \frac{4}{5}\sigma = \frac{4}{5} \times 6 = \frac{24}{5}.$

Example 3 : Find the probability that a random variable having the standard normal distribution will take on a value between 0·87 and 1·28.

Sol. : $P(0{\cdot}87 < Z < 1{\cdot}28)$

= area between $Z = 0{\cdot}87$ and $Z = 1{\cdot}28$.

= (area from $Z = 0$ to $Z = 1{\cdot}28$)

– (area from $Z = 0$ to $Z = 0{\cdot}87$)

$= 0{\cdot}3997 - 0{\cdot}3078 = 0{\cdot}0919.$

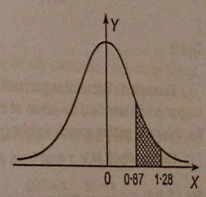


Fig. 6.5

Example 4 : Find the probability that a random variable having standard normal distribution will take a value between (i) 0·87 and 1·28, (ii) – 0·34 and 0·62.

Sol. : (i) As in the above example.

(ii) Area from $Z = -0{\cdot}34$ to 0 is the same as $Z = 0$ to $Z = 0{\cdot}34$ and is 0·1331

Area from $Z = 0$ to $Z = 0{\cdot}62$ is 0·2324.

Required area is the sum of the two

$\therefore\ P(-0{\cdot}34 < Z < 0{\cdot}62) = 0{\cdot}1331 + 0{\cdot}2324$

$= 0{\cdot}3655.$

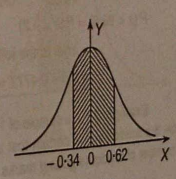


Fig. 6.6

Example 5 : For a normal variate with mean 2·5 and standard deviation 3·5, find the probability that (i) $2 \le X \le 4{\cdot}5$, (ii) $-1{\cdot}5 \le X \le 5{\cdot}5$.

Sol. : We have S.N.V. $Z = \dfrac{X-m}{\sigma} = \dfrac{X-2{\cdot}5}{3{\cdot}5}$

(i) When $X = 2$, $\quad Z = \dfrac{2-2{\cdot}5}{3{\cdot}5} = -0{\cdot}14$

When $X = 4{\cdot}5$, $\quad X = \dfrac{4{\cdot}5-2{\cdot}5}{3{\cdot}5} = 0{\cdot}57$

$P(2 \le X \le 4{\cdot}5) = P(-0.14 \le Z \le 0{\cdot}57)$

= Area between ($Z = -0{\cdot}14$ and $Z = 0{\cdot}57$)

= Area between ($Z = 0$ and $Z = 0{\cdot}14$)

+ Area between ($Z = 0$ and $Z = 0{\cdot}57$)

$= 0{\cdot}0557 + 0{\cdot}2157 = 0{\cdot}2714.$

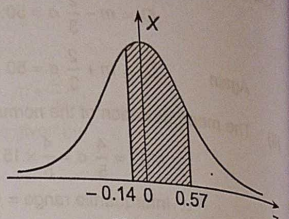


Fig. 6.7 (a)

(ii) When $X = -1{\cdot}5$, $\quad Z = \dfrac{-1{\cdot}5-2{\cdot}5}{3{\cdot}5} = -1{\cdot}14$

When $X = 5{\cdot}3$, $\quad Z = \dfrac{5{\cdot}3-2{\cdot}5}{3{\cdot}5} = 0{\cdot}8$

$P(-1{\cdot}5 \le X \le 5{\cdot}3) = P(-1{\cdot}14 \le Z \le 0{\cdot}8)$

= Area between ($Z = -1{\cdot}14$ and $Z = 0{\cdot}8$)

= Area between ($Z = 0$ and $Z = 1{\cdot}14$)

+ Area between ($Z = 0$ and $Z = 0{\cdot}8$)

$= 0{\cdot}3729 + 0{\cdot}2881 = 0.6610.$

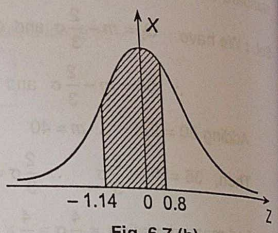


Fig. 6.7 (b)

Type II

Example 1 : Sacks of sugar packed by an automatic loader have an average weight at hundred kilograms with standard deviation of two hundred fifty grams. Assuming a normal distribution, find the chance of getting a sack weighing less than 99·5 kilograms.

(Given : For S.N.V. z area from $z = 0$ to $z = 2$ is 0·4772)

Sol. : S.N.V. $z = \dfrac{x-m}{\sigma} = \dfrac{x-100}{0{\cdot}25}$

When $x = 99{\cdot}5$,

$z = \dfrac{99{\cdot}5-100}{0{\cdot}250} = \dfrac{-0{\cdot}5}{0{\cdot}25} = -2$

$P(x < 99{\cdot}5) = P(z < -2)$

= area to the left of ($z = -2$)

$= 0{\cdot}5 - 0{\cdot}4772 = 0{\cdot}0228.$

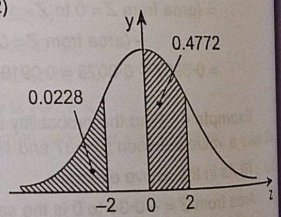


Fig. 6.8

Example 2 : In a sample of 1000 students the mean and standard deviation of marks obtained by the students in a certain test are 14 and 2·5. Assuming the distribution to be normal find the number of students getting marks (i) between 12 and 15, (ii) above 18, (iii) below 8.

(Given : For a S.N.V. z area between $z = 0$ and $z = 0{\cdot}4$ is 0·1554, that between $z = 0$ and $z = 0{\cdot}8$ is 0·2881, that between $z = 0$ and $z = 1{\cdot}6$ is 0·4452, that between $z = 0$ and $z = 2{\cdot}4$ is 0·4918.)

(S.U. 1986, 95, 96, 2002)

Sol. : We have $z = \dfrac{x-m}{\sigma} = \dfrac{x-14}{2{\cdot}5}$

(i) When $x = 12$, $\quad z = \dfrac{12-14}{2{\cdot}5} = -\dfrac{2}{2{\cdot}5} = -0{\cdot}8$

When $x = 15$, $\quad z = \dfrac{15-14}{2{\cdot}5} = \dfrac{1}{2{\cdot}5} = 0{\cdot}4$

$P(12 < x < 15) = P(-0{\cdot}8 < z < 0{\cdot}4)$

= area between

($z = -0{\cdot}8$ and $z = 0{\cdot}4$)

$= 0{\cdot}2881 + 0{\cdot}1554$

$= 0{\cdot}4435.$

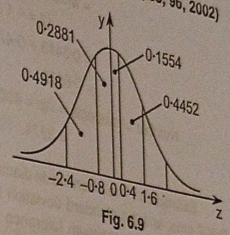


Fig. 6.9

$\therefore$ No. of students $= Np = 1000 \times 0{\cdot}4435 = 443.$

(ii) When $x = 18$, $\quad z = \dfrac{18-14}{2{\cdot}5} = \dfrac{4}{2{\cdot}5} = 1{\cdot}6$

$\therefore \quad P(x > 18) = P(z > 1{\cdot}6)$

= area to the right of ($z = 1{\cdot}6$) $= 0{\cdot}5 - 0{\cdot}4452 = 0{\cdot}0548.$

$\therefore$ No. of students $= Np = 1000 \times 0{\cdot}0548 = 55.$

(iii) When $x = 8$, $\quad z = \dfrac{8-14}{25} = -\dfrac{6}{2{\cdot}5} = -2{\cdot}4$

$P(x < 8) = P(z < -2{\cdot}4)$

= area to the left of ($z = -2{\cdot}4$) $= 0{\cdot}5 - 0{\cdot}4918 = 0{\cdot}0082.$

$\therefore$ No. of students $= Np = 1000 \times 0{\cdot}0082 = 8.$

Example 3 : Weights of 4000 students are found to be normally distributed with mean 50 kgs. and standard deviation 5 kgs. Find the number of students with weights, (i) less than 45 kgs, (ii) between 45 and 60 kgs.

(For a standard normal variate z area under the curve between $z = 0$ and $z = 1$ is 0·3413 and that between $z = 0$ and $z = 2$ is 0·4772).

(S.U. 1990, 2003, 04)

Sol. : We have $z = \dfrac{x-m}{\sigma} = \dfrac{x-50}{5}$

(i) When $x = 45$, $\quad z = \dfrac{45-50}{5} = -1$

$P(x < 45) = P(z < -1)$

= area to the left of ($z = -1$)

$= 0{\cdot}5 - 0{\cdot}4313 = 0{\cdot}0687$

$\therefore$ Number of students $= Np$

$= 4000 \times 0{\cdot}0687$

$= 275$ nearly.

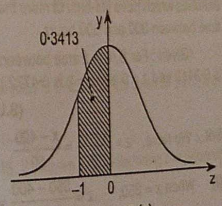


Fig. 6.10 (a)

(ii) When $x = 60$, $z = \dfrac{60 - 50}{5} = 2$

$P(45 < x < 60) = P(-1 < z < 2)$

= area between ($z = -1$ and $z = 2$)

$= 0{\cdot}3413 + 0{\cdot}4772$

$= 0{\cdot}8185$

$\therefore$ Number of students $= Np = 4000 \times 0{\cdot}8185$

$= 3274.$

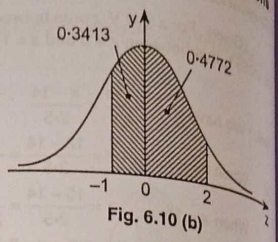


Fig. 6.10 (b)

Example 4 : The mean inside diameter of a sample of 200 washers produced by a machine is 0·502 cms. and the standard deviation is 0·005 cms. The purpose for which these washers are intended allows a maximum tolerance in the diameter of 0·496 to 0·508 cms. Determine the percentage of defective washers produced by the machine assuming the diameters are normally distributed. (Area under the normal curve from $z = 0$ to $z = 1{\cdot}2$ is 0·3849, from $z = 0$ to $z = 1{\cdot}4$ is 0·4192, from $z = 0$ to $z = 1{\cdot}0$ is 0·3413 from $z = 0$ to $z = 1{\cdot}6$ is 0·4452) **(S.U. 1990, 91, 92, 98)**

Sol. : We have, $z = \dfrac{x - m}{\sigma}$

But $m = 0{\cdot}502$, $\sigma = 0{\cdot}005$.

$\therefore \quad z = \dfrac{x - 0{\cdot}502}{0{\cdot}005}$

When $x = 0{\cdot}496$, $z = \dfrac{0{\cdot}496 - 0{\cdot}502}{0{\cdot}005} = -1{\cdot}2$

When $x = 0{\cdot}508$, $z = \dfrac{0{\cdot}508 - 0{\cdot}502}{0{\cdot}005} = 1{\cdot}2$

$P(0{\cdot}496 < x < 0{\cdot}508) = P(-1{\cdot}2 < z < 1{\cdot}2)$

= Area from ($z = -1{\cdot}2$ to $z = 1{\cdot}2$)

$= 2\,(0{\cdot}3849) = 0{\cdot}7698$

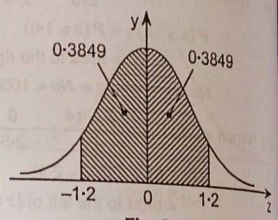


Fig. 6.11

$\therefore$ Percentage of washers between tolerance limits = 76·98 = 77% nearly.

$\therefore$ Percentage of defective washers = 100 – 77 = 23%.

Example 5 : The life-time of certain type of battery has mean life of 400 hours and a standard deviation of 50 hours. Assuming the distribution of life-time to be normal, find (i) the percentage of batteries which have life-time of more than 350 hours, (ii) the proportion of batteries which have life-time between 300 and 500 hours.

(Given : For S.N.V. z area between $z = 0$ and $z = 1$ is 0·3413 and $z = 0$ and $z = 2$ is 0·4772.)

(S.U. 1985, 97, 98)

Sol. : We have, $z = \dfrac{x - m}{\sigma} = \dfrac{x - 400}{50}$

(i) When $x = 350$, $z = \dfrac{350 - 400}{50} = -1$

$\therefore \quad P(x > 350) = P(z > -1)$

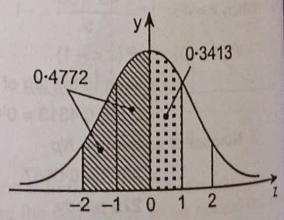


Fig. 6.12

$\therefore \quad P(x > 350)$ = Area to the right of ($z = -1$)

$= 0{\cdot}3413 + 0{\cdot}5 = 0{\cdot}8413$

$\therefore$ Percentage of batteries = 84·13 %

(ii) When $x = 300$, $z = \dfrac{300 - 400}{50} = -2$

When $x = 500$, $z = \dfrac{500 - 400}{50} = 2$

$\therefore \quad P(300 < x < 500) = P(-2 < z < 2)$

= area between ($z = -2$ and $z = 2$)

$= 2\,(0{\cdot}4772) = 0{\cdot}9544$

$\therefore$ Percentage of batteries = 95·44 %.

Example 6 : The income of a group of 10,000 persons is normally distributed with mean ₹ 520 and standard deviation ₹ 60. Find (i) the lowest income of the richest 500 (ii) the highest income of the poorest 500.

(Given : For S.N.V. z area from $z = 0$ to $z = 1{\cdot}645$ is 0·45)

Sol. : This is a reverse problem. So far we have found the probability for a given value of z. Now, we shall find the value of z for a given probability.

If we consider the richest 500 persons, the probability that a person selected at random will be one of them

$$= \frac{500}{10{,}000} = 0{\cdot}05.$$

We have to find the value of z to the right of which the area is 0·05. But the are to the right of $z = 0$ is 0·5.

$\therefore$ Area (from $z = 0$ to z = this value)

$= 0{\cdot}5 - 0{\cdot}05 = 0{\cdot}45.$

We are given that the area from $z = 0$ to $z = 1{\cdot}645$ is 0·45.

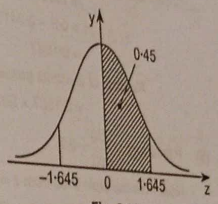


Fig. 6.13

$\therefore$ The required value of $z = 1{\cdot}645$.

But $z = \dfrac{x - 520}{60}$ $\quad \therefore \; 1{\cdot}645 = \dfrac{x - 520}{60}$ $\quad \therefore \; x = 618{\cdot}7$

$\therefore$ The lowest income of the richest 500 is ₹ 619.

Now, to solve the other part of the problem we find that the probability that a person will be one of the poorest 500 $= \dfrac{500}{1000} = 0{\cdot}05$.

We have to find the value of z to the left of which the area is 0·05. But, the area to the left of $z = 0$ is 0·5.

$\therefore$ Area from ($z = 0$ to z = this value) $= 0{\cdot}5 - 0{\cdot}05 = 0{\cdot}45$.

We are given that area from $z = 0$ to $z = 1{\cdot}645$ is 0·45. Since we are considering the poorest persons this value of z must be negative *i.e.* $z = -1{\cdot}645$.

But $z = \dfrac{x - 520}{60}$ $\quad \therefore \; -1{\cdot}645 = \dfrac{x - 520}{60}$ $\quad \therefore \; x = 421$

$\therefore$ The highest income of the poorest 500 person is ₹ 421.

Example 7 : In the above example No. 6 find the minimum income if 7000 persons are to be brought under income tax net. (Given : For S.N.V. z area from $z = 0$ to $z = 0{\cdot}525$ is $0{\cdot}2$)

Sol. : If 7000 persons are to be brought under income tax limits,

$$P\text{(a person paying income tax)} = \frac{7000}{10{,}000} = 0{\cdot}7$$

By data the area to the right of $z = -0{\cdot}525$ is $0{\cdot}5 + 0{\cdot}2 = 0{\cdot}7$.

$$\therefore \quad -0{\cdot}525 = \frac{x - 520}{60} \qquad \therefore \quad x = ₹\ 488.$$

Example 8 : In an examination given by 500 candidates the average and standard deviation of marks obtained are 40 and 10 respectively. Assuming the distribution of marks to be normal find approximately, (i) how many will pass if 50 is fixed as a minimum, (ii) what should be the minimum if 350 candidates are to pass ? **(S.U. 2008)**

Sol. : We have, $z = \dfrac{x - 40}{10}$

(i) When $x = 50$, $z = \dfrac{50 - 40}{10} = 1$

$$\therefore \quad P(x \geq 50) = P(z \geq 1)$$
$$= \text{area to the right of } z = 1$$
$$= 0{\cdot}5 - 0{\cdot}3413$$
$$= 0{\cdot}1587$$

$\therefore$ Number of students passing

$$= 0{\cdot}1587 \times 500 = 79$$

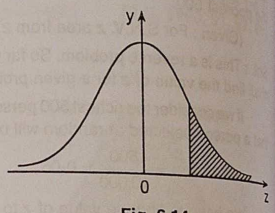


Fig. 6.14

(ii) $P(\text{passing}) = \dfrac{350}{500} = 0{\cdot}7$

Now, from the table area from $z = 0$ to $z = 0{\cdot}525$ is $0{\cdot}2$.

$\therefore$ Area to the right of $z\ (= -0{\cdot}525) = 0{\cdot}7$

$$\therefore \quad -0{\cdot}525 = \frac{x - 40}{10} \qquad \therefore \quad x = 34{\cdot}75.$$

Example 9 : A large number of automobile batteries have average life of 24 months. If 34 percent of them average between 22 and 26 months and 272 of them last longer than 29 months how many were in the group tested ? Assume the distribution to be normal.

(For a normal curve 17% and 36·4% values lie between the mean and respective distances of 0·44 and 1·1 times standard deviation from the mean.)

Sol. : Because of symmetry of the normal distribution and because $m = 24$, the area between 22 and 24 is equal to that between 24 and 26. This area is equal to $34 / 2 = 17\%$ which corresponds to S.N.V. $z = 0{\cdot}44$ by data and $x = 26$.

$$\therefore \quad \frac{26 - 24}{\sigma} = 0{\cdot}44 \quad \therefore \quad \sigma = \frac{2}{0{\cdot}44} = \frac{50}{11}$$

Now, when $x = 29$,

$$z = \frac{x - m}{\sigma} = \frac{29 - 24}{50/11} = \frac{11}{10} = 1{\cdot}1.$$

By data area between $z = 0$ and $z = 1{\cdot}1$ is $0{\cdot}364$. $\therefore$ The area to the right of $z = 1{\cdot}1$ *i.e.* $x = 29$ is $= 0{\cdot}5 - 0{\cdot}364 = 0{\cdot}136$ which is the probability that a battery will last longer than 29 months.

But $p = \dfrac{f}{N}$ and $f = 272$, $p = 0{\cdot}136$

$$\therefore \quad N = \frac{f}{p} = \frac{272}{0{\cdot}136} = 2000$$

Hence, 2000 batteries were tested.

Example 10 : In a normal distribution 31% items are under 45 and 8% are over 64. Find its mean and standard deviation.

(Given : For a normal distribution the area between $z = 0$ and $z = 0{\cdot}5$ is $0{\cdot}19$ and that between $z = 0$ and $z = 1{\cdot}4$ is $0{\cdot}42$) **(S.U. 1987, 90, 99, 2001)**

Sol. : Let m and σ be the mean and the standard deviation of the distribution. Since, 31% items are below 45, $50 - 31 = 19$ % items are between 45 and m. Since, 8% items are above 64. $50 - 8 = 42\%$ items are between m and 64.

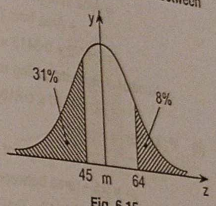


Fig. 6.15

Now, we are given that the area from $z = 0$ to $z = 0{\cdot}5$ is $1{\cdot}9$ *i.e.* the area from $z = 0$ to $z = -0{\cdot}5$ is 0.19. But area in a normal distribution is equal to the probability.

$$\therefore \quad P(45 < x < m) = P(-0{\cdot}5 < z < 0) = 0{\cdot}19.$$

But $z = \dfrac{x - m}{\sigma}$ $\quad\therefore$ When $x = 45$, $z = -0{\cdot}5$.

$$\therefore \quad -0{\cdot}5 = \frac{45 - m}{\sigma} \qquad \therefore \quad -0{\cdot}5\,\sigma = 45 - m \qquad \text{.....(1)}$$

Similarly, $P(45 < x < 64) = P(0 < z < 1{\cdot}4) = 0{\cdot}42$

When $x = 64$, $z = 1{\cdot}4$

$$\therefore \quad 1{\cdot}4 = \frac{64 - m}{\sigma} \qquad \therefore \quad 1{\cdot}4\,\sigma = 64 - m \qquad \text{.....(2)}$$

Solving (i) and (ii), we get, $\quad m = 50$ and $\sigma = 10$.

Type III

Example 1 : Monthly salary X in a big organisation is normally distributed with mean ₹ 3000 and standard deviation of ₹ 250. What should be the minimum salary of a worker in this organisation, so that the probability that he belongs to top 5% workers ?

Sol. : We have $Z = \dfrac{X - m}{\sigma} = \dfrac{X - 3000}{250}$.

We want to find Z_1 such that

$$P(Z > z_1) = \frac{5}{100} = 0{\cdot}05$$

Since $0{\cdot}5 - 0{\cdot}05 = 0{\cdot}45$, and corresponding to $0{\cdot}45$ the entry in the area table is $1{\cdot}64$. $\therefore$ $z_1 = 1{\cdot}64$.

$$\therefore \quad 1{\cdot}64 = \frac{X - 3000}{250}$$

$$\therefore \quad X = 3000 + 250 \times 1{\cdot}64 = ₹\ 3410.$$

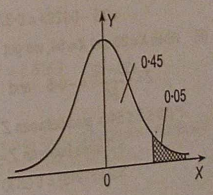


Fig. 6.16

Example 2 : The marks obtained by 1000 students in an examination are found to be normally distributed with mean 70 and standard deviation 5. Estimate the number of students whose marks will be (i) between 60 and 75, (ii) more than 75.

Sol. : We have S.N.V. $Z = \dfrac{X-m}{\sigma} = \dfrac{X-70}{5}$

(i) When $X = 60$, $Z = \dfrac{60-70}{5} = -2$;

When $X = 75$, $Z = \dfrac{75-70}{5} = 1$.

$P(60 \le X \le 75) = P(-2 < Z < 1)$

= Area between ($Z = -2$ and $Z = 1$)

= Area from ($Z = 0$ to $Z = 2$)

+ Area from ($Z = 0$ to $Z = 1$)

$= 0{\cdot}4772 + 0{\cdot}3413 = 0{\cdot}8185$

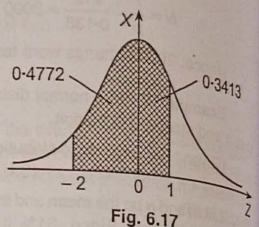


Fig. 6.17

∴ Number of students getting marks between 60 and 75

$= Np = 1000 \times 0{\cdot}8185 = 818$

(ii) $P(X \ge 75) = P(Z \ge 1)$

= Area to the right of $Z = 1$

$= 0{\cdot}5$ – (area between $Z = 0$ and $Z = 1$)

$= 0{\cdot}5 - 0{\cdot}3413 = 0{\cdot}1587$

∴ Number of students getting more than 75 marks

$= Np = 1000 \times 0{\cdot}1587 = 159$

Example 3 : In an intelligence test administered to 1000 students, the average was 42 and standard deviation was 24. Find the number of students (i) exceeding the score 50 and (ii) between 30 and 54.

Sol. : We have S.N.V. $Z = \dfrac{X-m}{\sigma}$

By data, $m = 42$ and $\sigma = 24$.

$\therefore\ Z = \dfrac{X-42}{24}$.

(i) When $X = 50$, $Z = \dfrac{50-42}{24} = \dfrac{1}{3} = 0{\cdot}33$

$P(Z \ge 50)$ = area to the right of 0·33

$= 0{\cdot}5$ – (area between $Z = 0$ and $Z = 0{\cdot}33$)

$= 0{\cdot}5 - 0{\cdot}1293 = 0{\cdot}3707$

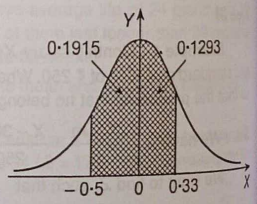


Fig. 6.18

(ii) When $X = 30$ and $X = 54$, we get

$Z = \dfrac{30-42}{24} = -0{\cdot}5$ and $Z = \dfrac{54-42}{24} = 0{\cdot}5$

$P(30 \le Z \le 54)$ = area between $Z = -0{\cdot}5$ to $Z = 0{\cdot}5$

= 2 (area between $Z = 0$ and $Z = 0{\cdot}5$)

$= 2\,(0{\cdot}1915) = 0{\cdot}3830$

Number of students getting more than 50 marks

$= Np = 1000 \times 0{\cdot}3707 = 371$

Number of students getting marks between 30 and 54

$= Np = 1000 \times 0{\cdot}383 = 383$

Type IV

Example 1 : The incomes of a group of 10,000 persons were found to be normally distributed with mean ₹ 520 and standard deviation ₹ 60. Find (i) the number of persons having incomes between ₹ 400 and 550, (ii) the lowest income of the richest 500.

Sol. : S.N.V. $Z = \dfrac{X-m}{\sigma} = \dfrac{X-520}{60}$

(i) When $X = 400$, $Z = \dfrac{400-520}{60} = -2$.

When $X = 550$, $Z = \dfrac{550-520}{60} = 0{\cdot}5$

$\therefore\ P(400 \le X \le 550)$ = Area (from $z = -2$ to $z = 0{\cdot}5$)

But, area (from $z = -2$ to $z = 0$)

= Area (from $z = 0$ to $z = 2$)

$= 0{\cdot}4772$.

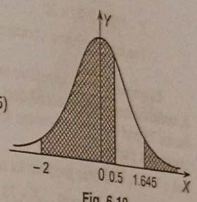


Fig. 6.19

And area (from $z = 0$ to $z = 0{\cdot}5$) = 0·1915.

$\therefore\ P(400 \le x \le 550)$ = Area (from $z = -2$ to $z = 0{\cdot}5$)

$= 0{\cdot}4772 + 0{\cdot}1915 = 0{\cdot}6687$.

∴ The number of persons whose incomes are between ₹ 400 and ₹ 550

$= Np = 10{,}000 \times 0{\cdot}6687 = 6687$.

(ii) If we have to consider the richest 500 persons then the probability that a person selected at random will be one of them

$= \dfrac{500}{10{,}000} = 0{\cdot}05$

This is a reverse problem. So far we have found the probability for a given value of Z. Here, we have to find the value of Z for a given probability. We have to find the value of Z to the right of which the area is 0·05. But area to the right of $Z = 0$ is 0·5.

∴ Area from ($Z = 0$ to Z = this value) $= 0{\cdot}5 - 0{\cdot}05 = 0{\cdot}45$

From the table we find that the area from $Z = 0$ to $Z = 1{\cdot}645$ is 0·45

∴ The required value of $Z = 1{\cdot}645$

But $Z = \dfrac{X-520}{60}$ ∴ $1{\cdot}645 = \dfrac{X-520}{60}$

∴ $X - 520 = 60 \times 1{\cdot}645$ ∴ $X = 520 + 98{\cdot}7 =$ ₹ 618·7.

∴ The lowest income of the richest 500 is ₹ 618·7.

Example 2 : The income of a group of 10,000 persons was found to be normally distributed with mean of ₹ 750 and standard deviation of ₹ 50. What is the lowest income of richest 250 ?

Sol. : S.N.V. $Z = \dfrac{X-m}{\sigma} = \dfrac{X-750}{50}$.

If we have to consider the richest 250 persons then the probability that a person selected at random will be one of them is $= \frac{250}{10,000} = 0{\cdot}025$.

This again is a reverse problem. So far we have found the probability for a given value of X. Here, we have to find the value of Z for a given probability. We have to find the value of Z to the right of which the area is 0·025. But the area to the right of $Z = 0$ is 0·5.

$\therefore$ Area from ($Z = 0$ to $Z =$ this value) $= 0{\cdot}5 - 0{\cdot}025 = 0{\cdot}475$.

From the table we find that the area from $Z = 0$ to $Z = 1{\cdot}96$ is 0·475.

$\therefore$ The required value of $Z = 1{\cdot}96$

When $Z = 1{\cdot}96$, $\quad 1{\cdot}96 = \frac{X - 750}{50}$

$\therefore X - 750 = 1{\cdot}96 \times 50 \qquad \therefore X = 848$

$\therefore$ The lowest income of richest 250 persons = ₹ 848.

Example 3 : In a competitive examination the top 15% of the students appeared will get grade A, while the bottom 20% will be declared fail. If the grades are normally distributed with mean % of marks 75 and S.D. 10, determine the lowest % of marks to receive grade A and the lowest % of marks that passes.

Sol. : This is a reverse problem as above.

We have $Z = \frac{X - m}{\sigma} = \frac{X - 75}{10}$.

(i) Grade A is given for 15%. We have to find the value of Z to the right of which the area is 0·15.

But the area to the right of $Z = 0$ is 0·5.

$\therefore$ Area from ($Z = 0$ to $Z =$ this value)

$= 0{\cdot}5 - 0{\cdot}15 = 0{\cdot}35$.

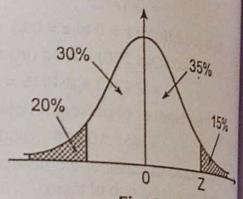


Fig. 6.20

From the table, we find that the area between $Z = 0$ to $Z = 1{\cdot}04$ is 0·35.

$\therefore$ The required value of $Z = 1{\cdot}04$.

But $Z = \frac{X - 75}{10} \qquad \therefore 1{\cdot}04 = \frac{X - 75}{10} \qquad \therefore X = 75 + 10{\cdot}4 = 85{\cdot}4$

(ii) Lowest 20% students are declared fail. We have to find the value of Z to the left of which the area is 0·20. But the area to the left of $Z = 0$ is 0·5.

$\therefore$ Area from ($Z = 0$ to $Z =$ this value) $= 0{\cdot}5 - 0{\cdot}2 = 0{\cdot}3$.

From the table we find that the area between $Z = 0$ and $Z = 0{\cdot}84$ is 0·3.

But this ordinate is on the left and hence negative.

$\therefore$ The required value of $Z = -0{\cdot}84$.

But $Z = \frac{X - 75}{10} \qquad \therefore -0{\cdot}84 = \frac{X - 75}{10} \qquad \therefore X = 75 - 8{\cdot}4 = 66{\cdot}6$.

Type V

Example 1 : Find the mean and the standard deviation of a normal distribution of marks in an examination where 58 % of the candidates obtained marks below 75, 4 % got above 80 and the

between 75 and 80 (For a S.N.V. the area under the curve between $z = \pm 0{\cdot}2$ is 0·16 and between $z = \pm 1{\cdot}8$ is 92).

Sol. : Let m and σ be the mean and the standard deviation of the variate.

Since 58 % students are below 75, $58 - 50 = 8$ % students are between 75 and m.

Since 4 % students are above 80, $50 - 4 = 46$ % students are between m and 80.

We are given that area between $Z = \pm 0{\cdot}2$ is 0·16 and that between $Z = \pm 1{\cdot}8$ is 9·92.

Hence, the area between $Z = 0$ and $Z = 0{\cdot}2$ is $\frac{0{\cdot}16}{2} = 0{\cdot}8$ and that between $Z = 0$ and $Z = +1{\cdot}8$ is $\frac{0{\cdot}92}{2} = 0{\cdot}46$.

In other words for area 0·08 (8 %), $Z = 0{\cdot}2$ and for area 0·46 (46 %), $Z = 1{\cdot}8$.

$\therefore \frac{75 - m}{\sigma} = 0{\cdot}2$ and $\frac{80 - m}{\sigma} = 1{\cdot}8$

$\therefore 75 - m = 0{\cdot}2\,\sigma$ and $80 - m = 1{\cdot}8\,\sigma$

Subtracting $-5 = -1{\cdot}6\,\sigma \quad \therefore \sigma = \frac{5}{1{\cdot}6} = 3{\cdot}125$

$\therefore m = 75 - 0{\cdot}2\,\sigma = 75 - 3{\cdot}125 \times 0{\cdot}2 = 74{\cdot}4$ mark.

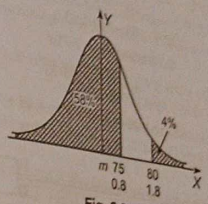


Fig. 6.21

Example 2 : Marks obtained by students in an examination follow normal distribution. If 30% of students got below 35 marks and 10% got above 60 marks, find the mean and standard deviation.

Sol. : Let m and σ be the mean and the standard deviation of the distribution.

Since 30% students are below 35, 20% students are between 35 and m.

Since 10% students are above 60, 40% students are between m and 60.

From the table we find that,

0·2 area corresponds to $Z = 0{\cdot}525$

and 0·4 area corresponds to $Z = 1{\cdot}283$

But 0·2 area is to the left of m hence $Z = -0{\cdot}525$.

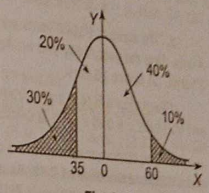


Fig. 6.22

$\therefore \frac{35 - m}{\sigma} = -0{\cdot}525; \qquad \frac{60 - m}{\sigma} = 1{\cdot}283$

$\therefore 35 - m = -0{\cdot}525\,\sigma$

$\therefore 60 - m = 1{\cdot}283\,\sigma$ ..... (1)

By subtraction, we get,

$25 = 1{\cdot}808\,\sigma \qquad \therefore \sigma = \frac{25}{1{\cdot}808} = 13{\cdot}83$.

Putting this value of σ in (1), we get

$35 - m = -0{\cdot}525\,(13{\cdot}83)$

$\therefore m = 35 + 0{\cdot}525\,(13{\cdot}83) = 42{\cdot}26$

Hence, mean = 13·83 and standard deviation, $\sigma = 42{\cdot}26$.

Example 3 : In a distribution exactly normal 7 % of items are under 35 and 89 % are under 63. What are the mean and standard deviation ? **(M.U. 2004)**

Sol. : Since 7 % items are below 35, 50 – 7 = 43 % items are between 35 and m, and since 89 % items are below 63, 89 – 50 = 39 % items are between m, and 63.

For area 0·43, $Z = 1·48$.

Since $35 < m$, $Z = -1·48$ and for area 0·39, $Z = 1·23$.

$$\therefore \quad \frac{35-m}{\sigma} = -1·48 \quad \text{and} \quad \frac{63-m}{\sigma} = 1·23$$

$$\therefore \quad 35 - m = -1·48\,\sigma$$

and $\quad 63 - m = 1·23\,\sigma.$

Subtracting $\quad 28 = 2·71\sigma$

$$\therefore \quad \sigma = \frac{28}{2·71} = 10·33$$

$$\therefore \quad m = 35 + 1·48\sigma = 35 + 1·48 \times 10·33$$

$$= 35 + 15·3 = 50·3.$$

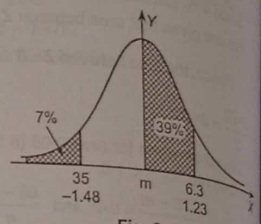


Fig. 6.23

EXERCISE - III

(A) **1.** The weights of 4000 students are found to be normally distributed with mean 50 kilograms and standard deviation 5 kilograms. Find the probability that a student selected at random will have weight (i) less than 45 kilograms (ii) between 45 and 60 kilograms. (For a standard normal variate z, area under the curve between $z = 0$ and $z = 1$ is 0·3413 and that between $z = 0$ and $z = 2$ is 0·4772.)

[Ans. : (i) 0·1587, **(ii)** 0·8185 **]**

2. The sizes of 10,000 items are normally distributed with mean 20 cms and standard deviation 4 cms. Find the probability that an item selected at random will have size between (i) 18 cms and 23 cms. (ii) above 26 cms.

(Given : For a S.N.V. z area from $z = 0$ to $z = 0·5$ is 0·1915 , that from $z = 0$ to $z = 0·75$ is 0·2734 and that from $z = 0$ to $z = 1·5$ is 0·4332)

[Ans. : (i) 0·4649, **(ii)** 0·0668 **]**

3. The customer accounts of a certain departmental store have an average balance of ₹ 120 and a standard deviation of ₹ 40. Assuming that the distribution of account balances is normal, find the proportion of accounts (i) over ₹ 150, (ii) between ₹ 100 and ₹ 150, (iii) between ₹ 60 and ₹ 90.

(Given : For S.N.V. z area from $z = 0$ to $z = 0·75$ is 0·2734, that from $z = 0$ to $z = 0·5$ is 0·1915 and that from $z = 0$ to $z = 1·5$ is 0·4332) **(S.U. 1982, 88)**

[Ans. : (i) 0·2266, **(ii)** 0·4649, **(iii)** 0·1587 **]**

(B) **1.** If the mean and standard deviation of a normal distribution are 50 and 12. Find how many items out of 500 lie between 38 and 62 and also between 26 and 74.

(Given : For S.N.V. z area from $z = 0$ to $z = 1$ is 0·34135 and that from $z = 1$ to $z = 2$ is 0·1359)

[Ans. : (i) 341, **(ii)** 477 **]**

2. The income distribution of a group of 10,000 persons was found to be normal with mean ₹ 7500 and standard deviation ₹ 500. What is the number of persons of this group who have income (i) exceeding ₹ 6680, (ii) exceeding ₹ 8320 ?

(Given : Area under the standard normal curve between ordinates ± 1·64 is 0·8990.)

(S.U. 1987, 89, 99, 2006) [Ans. : (i) 9495, **(ii)** 505 **]**

3. Assuming that the diameter of 1000 plugs taken consecutively from a machine, form a normal distribution with mean 0·7515 and standard deviation 0·002, how many plugs are likely to be rejected if the approved diameter is 0·752 ± 0·004.

(Given : For a S.N.V. z area between $z = 0$ and $z = 1·75$ is 0·4599 and that between $z = 0$ and $z = 2·25$ is 0·4878.) **(S.U. 1985) [Ans. : 52·3]**

4. In a test of 2000 electric bulbs, it was found that the life time of the bulbs was normally distributed with the mean of 2040 hours and standard deviation of 60 hours. Estimate the number of bulbs likely to burn out in less than 1950 hours.

(Given : For S.N.V. z area between $z = 0$ and $z = 1·5$ is 0·4332) **(S.U. 1985) [Ans. : 134]**

5. The incomes of a group of 10,000 persons were distributed normally with mean ₹ 5200 and standard deviation of ₹ 600. Find the number of persons having income between ₹ 4000 and ₹ 5500.

(Given : For S.N.V. z, area between $z = 0$ and $z = 0·5$ is 0·1915 and that from $z = 0$ and $z = 2$ is 0·4772.) **(S.U. 1985) [Ans. : 6687]**

6. The size of hats is normally distributed with mean of 18·5 cms. and standard deviation of 2·5 cms. How many hats in a total of 2000 will have sizes between (i) 18 cms. and 20 and (ii) above 20 cms. ?

(Given : For S.N.V. z, area between $z = 0$ and $z = 0·6$ is 0·2257 and that between $z = 0$ and $z = 0·2$ is 0·0793.) **(S.U. 1985) [Ans. : (i)** 610, **(ii)** 549 **]**

7. In an intelligence test administered to 1000 children the mean score was, 42 with s.d. 24. Find the number of children (i) scoring more than 60 and (ii) between 20 and 40 assuming the distribution to be normal.

(Given : For S.N.V. z area from $z = 0$ to $z = 0·75$ is 0·2734, that between $z = 0$ and $z = 0·9167$ is 0·3202 and that between $z = 0$ to $z = 0·08333$ is 0·0332.) **(S.U. 1988, 99) [Ans. : (i)** 227, **(ii)** 287 **]**

8. In the above problem, find

(i) the number of students exceedign the score 50.

(ii) the number of students getting the score between 30 and 50.

(Area under S.N.V. z, from $z = 0$ to $z = 0·33$ is 0·1293 and from $z = 0$ to $z = 0·5$ is 0·195.)

(S.U. 2006) [Ans. : (i) 371, **(ii)** 324 **]**

(C) **1.** Diameters of ball bearings are normally distributed with mean 0·498 cms. and standard deviation 0·002 cms. If specifications require 0·500 ± 0·004 cm. what percentage of production will be rejected ?

(Given : For S.N.V. z area from $z = 0$ to $z = 1$ is 0·3413 and that from $z = 0$ to $z = 3$ is 0·4987.)

(S.U. 1986) [Ans. : 16 %]

2. The income distribution of a group of 10,000 persons was found to be normal with mean ₹ 750 and standard deviation ₹ 50. What percentage of this group had income (i) exceeding ₹ 668, (ii) exceeding ₹ 832 ?

(Given : Area under Standard Normal Curve between ordinates at ± 1·64 is 0·8990.)

(S.U. 1987, 89) [Ans. : (i) 94·95%, **(ii)** 5·05% **]**

3. The mean I.Q. of a large number of children of age 14 is 100 with s.d. 16. Assuming the distribution of I.Q. to be normal, find the percentage of the children having I.Q. between 70 and 120. (Area for S.N.V. z from $z = 0$ to $z = 1·875$ is 0·4696 and that from $z = 0$ to $z = -1·25$ is 0·3944.)

[Ans. : 86·4%]

4. The daily sales of a certain item are normally distributed with mean ₹ 8000 and variance ₹ 10,000. (i) What is the probability that on a certain day the sales will be less than ₹ 8210 ? (ii) What percentage of the days will the sales be between ₹ 8100 and ₹ 8210 ?

(Given : Area for S.N.V. z from $z = 0$ to $z = 2{\cdot}1$ is 0·4821; that between $z = 0$ and $z = 1$ is 0·3413.) [**Ans. : (i)** 0·9821, **(ii)** 14·06%]

5. The average selling price of houses in a city is ₹ 50,000 with standard deviation of ₹ 10,000. Assuming the distribution of selling price to be normal find (i) the percentage of houses that sell for more than ₹ 55,000, (ii) the percentage of houses selling between ₹ 45,000 and ₹ 60,000. (Area between $t = 0$ and $t = 1$ is 0·3413 and the area between $t = 0$ and $t = 0{\cdot}5$ is 0·1915 where t is S.N.V.) [**Ans. : (i)** 30·85%, **(ii)** 53·28%]

6. If the mean life time and S.D. of battery cells are 12 hrs, and 3 hrs what % of batteries will have life (i) between 10 and 14 hrs., (ii) more than 15 hrs., (iii) less than 6 hrs.

(Given : For S.N.V. z area from $z = 0$to $z = 0{\cdot}67$ is 0·2486, that from $z = 0$ to $z = 1$ is 0·3413 that from $z = 0$ to $z = 2$ is 0·4772.) **(S.U. 1987)** [**Ans. : (i)** 49%, **(ii)** 16%, **(iii)** 2%]

7. The average number of units produced by a manufacturing concern per day is 355 with standard deviation of 50. It makes a profit of ₹ 1·50 per unit.

Determine the percentage of days when its total profit per day is (i) between ₹ 457·50 and ₹ 645·00, (ii) greater than ₹ 682·50. Assume the distribution to be normal.

(Area between $t = 0$ and $t = 1$ is 0·3413 and between $t = 0$ and $t = 1{\cdot}5$ is 0·4332 and that between $t = 0$ and $t = 2$ is 0·4772.)

(**Hint :** Mean of the profit is 532·5 and S. D. of the profit is 75.)

[**Ans. : (i)** 77·45%, **(ii)** 2·27%]

(D) 1. The income distribution of workers in a certain factory was found to be normal with mean of ₹ 500 and standard deviation equal to ₹ 50. There were 228 persons above ₹ 600. How many persons were there in all ? (Area under the S.N. curve between height at 0 and 2 is 0·4772.)

(S.U. 1987) [**Ans. :** 10000]

2. In a factory a large number of workers have average income of ₹ 120. If 38·3% of them have income between ₹ 100 – 140 and 528 of them get more than ₹ 170, how many workers were interrogated ? (Area for S.N.V. between $z = 0$ and $z = 19{\cdot}15$ is 0·5 and that between $z = 0$ and $z = 1{\cdot}25$ is 0·3944. (**Hint :** Find $\sigma = 40$.) [**Ans. :** 5000]

3. The arithmetic mean of purchases per day by a customer in a large store is ₹ 25 with standard deviation of ₹ 10. If on a particular day, 100 customers purchased for ₹ 37·80 or more, estimate the total number of customers who purchased from the store that day.

(Given that the normal area between $t = 0$ and $t = 1{\cdot}28$ is 0·4000 where t is the S.N.V.)

[**Ans. :** 1000]

(E) 1. The heights of 1000 cakes baked with certain mix have a normal distribution with a mean of 5·75 cms, and a standard deviation of 0·75 cms. Find the number of cakes having heights between 5 cms, and 6·25 cms. Also find the maximum height of the flatest 200 cakes.

(For a standard variate t, the area between $t = -1$ and $t = 1$ is 0·6826, that between $t = 0$ and $t = 0{\cdot}67$ is 0·2486 and that between $t = 0$ and $t = 0{\cdot}84$ is 0·3). [**Ans. :** 589·9, 5·12]

2. 1000 light bulbs with an average life of 120 days are installed in streets of Mumbai. Their length of life is normally distributed with variance 400 days. (i) How many will expire in less than 90 days? (ii) If it is decided to replace all the bulbs together what interval should be allowed between replacements if not more than 10 percent should expire before replacement ?

(Area between $z = 0$ and $z = 1{\cdot}5$ is 0·4332 and 80% of the area lies between $z = \pm 1{\cdot}28$).

[**Ans. :** (i) 67, (ii) 94·4]

3. Monthly salaries of 1000 workers have a normal distribution with mean of ₹ 575 and a standard deviation of ₹ 75. Find the number of workers having salaries between ₹ 500 and ₹ 625 p.m. Also find the minimum salary of the highest paid 200 workers.

(Given : For a standard normal variate t (i) area between $t = 0$ and $t = 1$ is 0·3413, (ii) area between $t = 0$ and $t = 0{\cdot}67$ is 0·2486, (iii) area between $t = 0$ and $t = 0{\cdot}84$ is 0·3.)

[**Ans. :** ₹ 5899 , ₹ 638]

4. Incandescent electric bulbs of a certain make have a mean life of 800 hours with standard deviation of 150 hours. Out of 50·000 bulbs of that make used for street lighting how many of these would fuse in first 650 hours, if the distribution of the life of bulbs is assumed to be normal. In how many days will you find that only 7935 bulbs are burning ?

(Area of a normal curve lying between $z = 0$ to $z = 1$ is 0·3413.) [**Ans. :** (i) 7935, (ii) 950]

(F) 1. In a distribution which is exactly normal, 7% of the items are under 35 and 89% items are under 63. Find the mean and standard deviation of the distribution.

(Given that for area 0·43 S.N.V. $z = 1{\cdot}48$ and for area 0·39 S.N.V. $z = 1{\cdot}23$.) **(S.U. 1990)**

[**Ans. :** $m = 50{\cdot}29$, $\sigma = 10{\cdot}33$]

2. Find the mean and the standard deviation of a normal distribution of marks in an examination where 58% of the candidates obtained marks below 75, 4% got above 80 and the rest between 75 and 80.

(For a S.N.V. the area under the curve between $z = \pm 0{\cdot}2$ is, 0·16 and between $z = \pm 1{\cdot}8$ is 0·92.) **(S.U. 1993, 98)** [**Ans. :** $m = 74{\cdot}375$, $\sigma = 3{\cdot}125$]

3. The qualifying marks for a certain examination are 35 and to secure distinction one has to score more than 74. If 25% of the students fail, whereas 6·681% obtained distinction, determine the mean and the standard deviation assuming that the distribution of marks is normal.

(Given : For S.N.V. z area from $z = 0$ to $z = 2/3$ is 0·25 and that between $z = 0$ to $z = 3/2$ is 0·43319.) [**Ans. :** $m = 47$, $\sigma = 18$]

4. At a certain examination 10% students who appeared for the paper in Statistics got less than 30 marks and 97% of the students got less than 62 marks. Assuming the distribution to be normal, find the mean and the s.d. of the distribution.

(Given that 40% of the area of the normal curve is between the ordinates corresponding to $t = 0$ and $t = 1{\cdot}3$ and 7% of the area is between $t = 1{\cdot}3$ and $t = 1{\cdot}9$ where, t is a S.N.V.)

[**Ans. :** $m = 43$, $\sigma = 10$]

☆☆☆

CHAPTER 7

Numerical Integration

1. Introduction

Many a time we come across a function $f(x)$ such $\int_a^b f(x)\,dx$ which is highly difficult to find and in some cases impossible, *e.g.*, $\int \sqrt{\sin x}\,dx$, $\int e^{x^2}\,dx$, etc. But we know that $\int_a^b f(x)\,dx$ = the area bounded by the curve $y = f(x)$, the ordinates at $x = 0$, $x = b$ and the x-axis. If we know the values of $y = f(x)$ at $x = x_0, x_1, \ldots, x_n = b$. Then we can find approximately this area by considering trapeziums through two consecutive points, or parabolas passing through three consecutive points or curves passing through four consecutive points. The sum of all such areas will be approximately equal to the required area *i.e.*, $\int_a^b f(x)\,dx$.

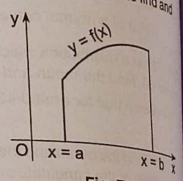


Fig. 7.1

This process of finding the definite integral $\int_a^b f(x)\,dx$ is known as **numerical integration**.

We shall consider here five rules of numerical integration -

(1) Newton Cote's Formulae (2) Trapezoidal Rule,
(3) Simpson's (1/3)rd rule, (4) Simpson's (3/8)th rule,
(5) Weddle's rule.

In trapezoidal rule, we fit straight lines through two consecutive points, in Simpson's (1/3)rd rule we fit parabolas through three consecutive points and in Simpson's (3/8)th rule, we fit cubic curves through three consecutive points.

2. Newton-Cotes Quadrature Formula

Consider $I = \int_a^b f(x)\,dx$

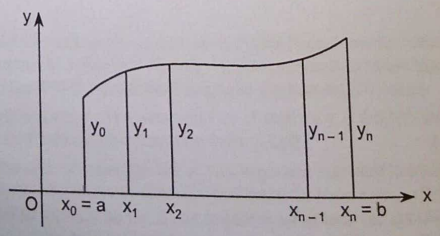
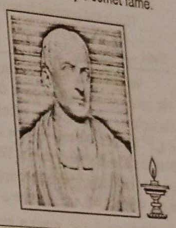


Fig. 7.2

Sir Issac Newton (1643 - 1727)

Newton was born just about one year after Galileo died. His discoveries are so numerous and varied that many consider him to be the **father of modern science**. He graduated from Trinity College, Cambridge. He developed intense interest in mathematics and the laws of nature. This led to his two most famous works "Philosphiae Naturalis Principia Mathematica" and "Optics". He discovered Binomial Theorem when he was a first year student. He cofounded Calculus along with Leibnitz. He explained the laws of light and colour among many other discoveries in Physics and Mathematics. He is best known for the three laws of motion and the law of gravity. A famous story says that Newton uncovered the law of gravity after seeing a falling apple. Newton is often ranked first, second being Albert Eienstein among history's physicists. Newton held the Lucasian Chair of Mathematics at Cambridge - a post later held by Stephen Hawking. He was knighted in 1705. He was the first scientist to be given the honour of burial in Westminster Abbey (burial place of English Monarchs). He was a good friend of Halley - Halley's comet fame.

Roger Cotes (1682 - 1716)



Roger Cotes was an English mathematician and is known for working closely with Newton who was his academic advisor. He invented the formula now known as Newton-Cotes quadrature formula. He is also known for his contribution to integral calculus, logarithms and numerical analysis. He discovered an important theorem on the nth roots of unity, the method of least squares, interpolation methods.

Let the interval (a, b) be divided into n equal subintervals each of length h by points $x_0, x_1, x_2, \ldots, x_n$ such that $x_0 = a$, $x_1 = x_0 + h$, $x_2 = x_0 + 2h, \ldots, x_n = b = x_0 + nh$. Suppose further the corresponding y coordinates are $y_0, y_1, y_2, \ldots, y_n$. Then,

$$I = \int_a^b f(x)\,dx = \int_{x_0}^{x_0+nh} y_r\,dx$$

where y_r is the polynomial passing through these points.

But by Newton's forward difference formula

$$f(a + rh) = y_r = y_0 + r\,\Delta y_0 + \frac{r(r-1)}{2!}\Delta^2 y_0 + \frac{r(r-1)(r-2)}{3!}\Delta^3 y_0 + \frac{r(r-1)(r-2)(r-3)}{4!}\Delta^4 y_0 + \frac{r(r-1)(r-2)(r-3)(r-4)}{5!}\Delta^5 y_0 + \frac{r(r-1)(r-2)(r-3)(r-4)(r-5)}{6!}\Delta^6 y_0 + \ldots$$

Now, let us put $a + rh = x$.

$\therefore$ $h\,dr = dx$. When $x = a = x_0$, $r = 0$

and when $x = x_0 + nh, = a + nh$, $r = n$.

$$\therefore\ I = h\int_0^n \left[y_0 + r\,\Delta y_0 + \frac{r(r-1)}{2!}\Delta^2 y_0 + \ldots\ldots \right] dr$$

$$= h\left[y_0\, r + \frac{r^2}{2}\Delta y_0 + \frac{1}{2!}\left(\frac{r^3}{3} - \frac{r^2}{2}\right)\Delta^2 y_0 + \ldots\ldots \right]_0^n$$

$$= h\left[n\, y_0 + \frac{n^2}{2}\Delta y_0 + \frac{1}{2!}\left(\frac{n^3}{3} - \frac{n^2}{2}\right)\Delta^2 y_0 \right.$$

$$+ \frac{1}{3!}\left(\frac{n^4}{4} - n^3 + n^2\right)\Delta^3 y_0 + \frac{1}{4!}\left(\frac{n^5}{5} - \frac{3n^4}{2} + \frac{11n^3}{3} - 3n^2\right)\Delta^4 y_0$$

$$+ \frac{1}{5!}\left(\frac{n^6}{6} - 2n^5 + \frac{35n^4}{4} - \frac{50n^3}{3} + 12n^2\right)\Delta^5 y_0$$

$$\left. + \frac{1}{6!}\left(\frac{n^7}{7} - \frac{15n^6}{6} + 17n^5 - \frac{225n^4}{4} + \frac{274n^3}{3} - 60n^2\right)\Delta^6 y_0 + \ldots\ldots \right]$$

This is the **general formula** for numerical integration.

Note ....

Various formulae can be derived form the above general formula, by putting $n = 1, 2, \ldots\ldots$ etc. We note that above formula is derived on the assumption that there passes an nth degree curve through the given points. When we put $n = 1$, the polynomial $y = a_0 + a_1 x + a_2 x^2 + \ldots\ldots + a_n x^n$ reduces to $y = a_0 + a_1 x = a + bx$ say. In this case we assume that there passes a straight line through two successive points. When we put $n = 2$, the polynomial reduces to $y = a_0 + a_1 x + a_2 x^2 = a + bx + cx^2$ say. In this case we assume that there passes a parabola through three successive points. When $n = 3$, we assume a cubic curve and so on.

When $n = 1$, *i.e.*, when a straight line is assumed, the resulting rule of integration is called **Trapezoidal rule.** When $n = 2$ *i.e.*, when a parabola is assumed, the rule is called **Simpson's one third rule.** When $n = 3$ *i.e.* when a cubic curve is assumed, the rule is known as **Simpson's three eighth rule.**

3. Trapezoidal Rule

Let the curve from $A_0\ (x_0, y_0)$ to $A_n\ (x_n, y_n)$ be given by $y = f(x)$. Let $A_1\ (x_1, y_1), \ldots\ldots, A_{n-1}\ (x_{n-1}, y_{n-1})$ be the points of sub-division as before.

Let these points on the curve be connected by straight lines. Thus, area under the curve from A_0 to A_1 is approximated to the area of the first trapezium. whose area $= \frac{1}{2}(y_0 + y_1)\cdot h$.

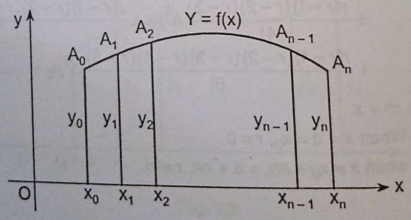


Fig. 7.3

Total area = sum of the areas of all trapeziums.

$$\therefore\ \int_a^b f(x)\,dx = \frac{1}{2}h[(y_0 + y_1) + (y_1 + y_2) + \ldots\ldots + (y_{n-1} + y_n)]$$

$$= \frac{1}{2}h[(y_0 + y_n) + 2(y_1 + y_2 + \ldots\ldots + y_{n-1})]$$

$$\therefore\ \boxed{\int_a^b f(x)\,dx = \frac{1}{2}h[X + 2R]}$$

where X = Sum of the eXtreme ordinates and R = Sum of the Remaining ordinates.

Example : Evaluate $\int_0^1 \sqrt{\sin x + \cos x}\,dx$ by trapezoidal rule by using the following data.

x :	0	0·2	0·4	0·6	0·8	1
y :	1	1·0857	1·1448	1·1790	1·1891	1·1755
Ordinate :	y_0	y_1	y_2	y_3	y_4	y_5

Sol. : We have $h = 0{\cdot}2$.

$$X = y_0 + y_5 = 1 + 1{\cdot}1755 = 2{\cdot}1755$$

$$R = y_1 + y_2 + y_3 + y_4$$

$$= 1{\cdot}0857 + 1{\cdot}1448 + 1{\cdot}1790 + 1{\cdot}1891 = 4{\cdot}5986$$

$$\therefore\ \int_a^b f(x)\,dx = \frac{h}{2}[X + 2R] = \frac{0{\cdot}2}{2}[2{\cdot}1755 + 9{\cdot}1972] = 1{\cdot}13727$$

4. Simpson's (1/3)rd Rule

Let the equation of the parabola passing through the first three points $(x_0, y_0), (x_1, y_1), (x_2, y_2)$ be

$$y = a + b(x - x_1) + c(x - x_1)^2$$

$$y_0 = a + b(x_0 - x_1) + c(x_0 - x_1)^2 \quad \text{when } x = x_0,$$

$$y_1 = a \quad \text{when } x = x_1,$$

$$y_2 = a + b(x_2 - x_1) + c(x_2 - x_1)^2 \quad \text{when } x = x_2,$$

But $x_0 - x_1 = -h$ and $x_2 - x_1 = h$

$$\therefore\ y_0 = a - bh + ch^2$$

$$y_1 = a \qquad \ldots\ldots\ldots\ldots\ldots\ldots \text{(A)}$$

$$y_2 = a + bh + ch^2$$

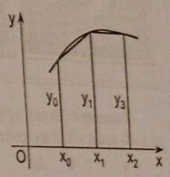


Fig. 7.4

Now, area under this parabola $= \int_{x_1 - h}^{x_1 + h} f(x)\,dx$

$$= \int_{x_1 - h}^{x_1 + h}\left[a + b(x - x_1) + c(x - x_1)^2\right]dx$$

Putting $x - x_1 = u$, $dx = du$. When $x = x_1 - h$, $u = -h$ and when $x = x_1 + h$, $u = h$.

$$\therefore\ \text{Area} = \int_{-h}^{h}(a + bu + cu^2)\,du = \left[au + \frac{bu^2}{2} + \frac{cu^3}{3}\right]_{-h}^{h}$$

$$\therefore\ \text{Area} = \left(ah + \frac{bh^2}{2} + \frac{ch^3}{3}\right) - \left(-ah + \frac{bh^2}{2} - \frac{ch^3}{3}\right) = 2ah + 2\frac{ch^3}{3} = \frac{h}{3}(6a + 2ch^2)$$

But it can be easily seen from (A) that $y_0 + 4y_1 + y_2 = 6a + 2ch^2$.

Assuming that the number of intervals is even and taking the area under all such parabolas.

$$\int_a^b f(x)\,dx = \frac{h}{3}\left[(y_0 + 4y_1 + y_2) + (y_2 + 4y_3 + y_4) + \ldots + (y_{n-2} + 4y_{n-1} + y_n)\right]$$

$$= \frac{h}{3}\left[(y_0 + y_n) + 2(y_2 + y_4 + \ldots + y_{n-2}) + 4(y_1 + y_3 + y_5 + \ldots + y_{n-1})\right]$$

$$\therefore \quad \boxed{\int_a^b f(x)\,dx = \frac{h}{3}[X + 2E + 4O]}$$

where, X = sum of the eXtreme ordinates, E = sum of the Even ordinates,

O = Sum of Odd ordinates.

Note ....

As anywhere else in calculus of finite differences our first point is (x_0, y_0) *i.e.* we start counting from zero and not one. With this convention while using Simpson's rule the order of the last point should be even *i.e.* 2, 4, 6, 8, 10, ... etc. *i.e.*, the number of sub-intervals must be even.

Thomas Simpson (1710 - 1761)

Thomas Simpson was a British mathematician. Now well known Simpson's rule to approximate definite integrals is credited to him, but the rule was first discovered by Johannes Kepler and was in use for 100 years before. He was the son of a weaver and taught himself mathematics. After seeing a solar eclipse he started studying astronomy. He taught mathematics at the Royal Military Academy from 1743. He was a fellow of Royal Society. His published works include Treatise On Fluxions, The Nature and Laws of Chance, A Treatise On Algebra, etc.

Example 1 : Obtain by Simpson's rule the area under the curve bounded by the x-axis, ordinates at x = 0 and x = 1, if the ordinates at different points of the curve are :

x	: 0	0·1	0·2	0·3	0·4	0·5	0·6	0·7	0·8	0·9	1
y	: 1	0·990	0·962	0·917	0·862	0·0800	0·735	0·671	0·610	0·553	0.500
Ordinate	: y_0	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8	y_9	y_{10}

Sol. : Simpson's rule can be applied because $n = 10$ is even. Here $h = 0{\cdot}1$,

X = sum of extreme ordinates

$= 1 + 0{\cdot}50 = 1{\cdot}50$

E = sum of even ordered ordinates

$= 0{\cdot}962 + 0{\cdot}862 + 0{\cdot}735 + 0{\cdot}610 = 3{\cdot}169$

O = sum of odd ordered ordinates

$= 0{\cdot}990 + 0{\cdot}917 + 0{\cdot}800 + 0{\cdot}671 + 0{\cdot}553 = 3{\cdot}931$

$$\therefore \quad \text{Area} = \frac{h}{3}[X + 2E + 4O] = \frac{0{\cdot}1}{3}[1{\cdot}50 + 6{\cdot}338 + 15{\cdot}724]$$

$$= \frac{0{\cdot}1}{3}(23{\cdot}562) = 0{\cdot}7854$$

Example 2 : The water under portion of a water-tank is divided by horizontal planes one metre apart into the following areas : 472, 398, 302, 198, 116, 60, 34, 12 and 4 sq. m. Use the trapezoidal rule to find the volume in cubic metres between the two extreme ends.

Sol. : We prepare the table as follows :

Distance S	:	1	2	3	4	5	6	7	8	9
Area A	:	472	398	302	198	116	60	34	12	4

Since, $V = A \times S$.

By Trapezoidal rule, $V = \frac{h}{2}[X + 2R]$

$h = 1$, $X = 472 + 4 = 476$

$R = 398 + 302 + 198 + 116 + 60 + 34 + 12 = 1120$

$\therefore \quad V = \frac{1}{2}[476 + 2240] = 1358$ cubic metres.

Example 3 : The velocity of a train which starts from rest is given by the following table the time being reckoned in minutes from the start and speed in km / hour.

Time	:	2	4	6	8	10	12	14	16	18	20
Speed	:	10	18	25	29	32	30	11	5	2	0

Find the total distance run in 20 minutes.

Sol. : We know that $\frac{ds}{dt} = v$ *i.e.*, $S = \int v\,dt$ *i.e.*, $S = \int_0^2 v\,dt$

$\therefore$ To use Simpson's Rule we need even number of ordinates. We can take one more ordinate at $t = 0$. By data when $t = 0$, $v = 0$.

Let the ordinates be denoted as

y_0	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8	y_9	y_{10}
0	10	18	25	29	32	30	11	5	2	0

Then by Simpson's (1/3)rd rule

$$S = \frac{h}{3}[X + 2E + 4O]$$

We have $h = \frac{2}{60}$ hrs. $= \frac{1}{30}$ hrs.

$X = 0 + 0$, $E = 18 + 29 + 30 + 5 = 82$.

$O = 10 + 25 + 32 + 11 + 2 = 80$

$\therefore \quad S = \frac{1}{90}[0 + 164 + 320] = 5{\cdot}378$ kms.

Example 4 : In an experiment a quantity G was measured as follows :

$G(20) = 95{\cdot}90$, $G(21) = 96{\cdot}85$, $G(22) = 97{\cdot}77$, $G(23) = 98{\cdot}68$,

$G(24) = 99{\cdot}56$, $G(25) = 100{\cdot}41$, $G(26) = 101{\cdot}24$.

Compute $\int_{20}^{26} G(x)\,dx$ by both Simpson's (1/3)rd rule and Trapezoidal rule.

Sol. : The table can be prepared as

x	:	20	21	22	23	24	25	26
$G(x)$:		95·90	96·85	97·77	98·68	99·56	100·41	101·24
Ordinate	:	y_0	y_1	y_2	y_3	y_4	y_5	y_6

(i) By trapezoidal rule,

$$\int_{20}^{26} G(x)\,dx = \frac{h}{2}[X + 2R]$$

But $h = 1,\ X = 95{\cdot}90 + 101{\cdot}24 = 197{\cdot}14$

$R = [\,96{\cdot}85 + 97{\cdot}77 + 98{\cdot}68 + 99{\cdot}56 + 100{\cdot}41\,] = 493{\cdot}27$

$$\therefore \int_{20}^{26} G(x)\,dx = \frac{1}{2}[197{\cdot}14 + 986{\cdot}54] = 591{\cdot}84$$

(ii) By Simpson's (1/3)rd rule,

$$\int_{20}^{26} G(x)\,dx = \frac{h}{3}[X + 2E + 4O]$$

Now $X = 95{\cdot}90 + 101{\cdot}24 = 197{\cdot}14$

$2E = 2\,(97{\cdot}77 + 99{\cdot}56) = 394{\cdot}66$

$4O = 4\,(96{\cdot}85 + 98{\cdot}68 + 100{\cdot}41) = 1183{\cdot}76$

$$\therefore \int_{20}^{26} G(x)\,dx = \frac{1}{3}[197{\cdot}14 + 394{\cdot}66 + 1183{\cdot}76] = 591{\cdot}85$$

Example 5 : Find the volume of solid of revolution formed by rotating about the x-axis, the area bounded by the lines $x = 0$, $x = 1$, $y = 0$ and the curve passing through the points given below.

x	:	0	0·25	0·50	0·75	1
y	:	1	0·9896	0·9589	0·9089	0·8415

Sol. : The volume is given by $V = \int_a^b \pi y^2\,dx$

We prepare the table of y^2

x	:	0	0·25	0·50	0·75	1
y^2	:	1	0·9793	0·9195	0·8261	0·7081
Ordinate	:	y_0	y_1	y_2	y_3	y_4

(i) By Trapezoidal rule

$$V = \pi \cdot \frac{h}{2}[X + 2R]$$

$$\therefore V = 3{\cdot}14156\left(\frac{0{\cdot}25}{2}\right)[(1 + 0{\cdot}7081) + 2(0{\cdot}9793 + 0{\cdot}9195 + 0{\cdot}8261)] = 2{\cdot}8109$$

(ii) By Simpson's (1/3)rd rule

$$V = \pi \cdot \frac{h}{3}[X + 2E + 4O]$$

$$\therefore V = 3{\cdot}14159\left(\frac{0{\cdot}25}{3}\right)[(1 + 0{\cdot}7081) + 2(0{\cdot}9195) + 4(0{\cdot}9793 + 0{\cdot}8261)] = 2{\cdot}819$$

Example 6 : Evaluate $\int_0^{\pi} \frac{\sin^2\theta}{5 + 4\cos\theta}\,d\theta$ by taking 5 ordinates by Simpson's (1/3)rd rule.

Sol. : We divide the interval (0, 180°) in 4 equal parts by taking $h = 45°$ as follows :

θ	:	0	45°	90°	135°	180°
y	:	0	0·0639	0·2	0·2302	0
Ordinate	:	y_0	y_1	y_2	y_3	y_4

By Simpson's (1/3)rd rule

$$S = \frac{h}{3}[X + 2E + 4O]$$

We have $h = \pi/4,\ X = 0 + 0 = 0,\ E = 2\,(0{\cdot}2) = 0{\cdot}4$

$4O = 4\,(0{\cdot}0639 + 0{\cdot}2302) = 1{\cdot}1764$

$$\therefore S = \frac{\pi}{12}[0 + 0{\cdot}4 + 1{\cdot}1764] = 0{\cdot}4127$$

Example 7 : Find the value of the integral $\int_0^1 \frac{x^2}{1 + x^3}\,dx$, using any two methods.

Compare the errors with the exact value of the integral. Also find log 2.

Sol. : We shall use Simpson's (1/3)rd rule and trapezoidal rule. Since to use Simpson's rule we need even number of ordinates we shall divide the interval (0, 1) into 4 equal parts by taking $h = 0{\cdot}25$. We prepare the following table.

x	:	0	0·25	0·50	0·75	1·0
y	:	0	0·06153	0·22222	0·39560	0·5
Ordinate	:	y_0	y_1	y_2	y_3	y_4

(i) By Simpson's (1/3)rd rule

$$S = \frac{h}{3}[X + 2E + 4O]$$

We have $h = 0{\cdot}25,\ X = 0 + 0{\cdot}5 = 0{\cdot}5,\ 2E = 2\,(0{\cdot}22222) = 0{\cdot}44444$

$4O = 4\,(0{\cdot}06153 + 0{\cdot}39560) = 1{\cdot}82852$

$$\therefore I = \frac{0{\cdot}25}{3}[0{\cdot}5 + 0{\cdot}44444 + 1{\cdot}82852] = 0{\cdot}23108$$

(ii) Also by Trapezoidal rule

$$I = \frac{h}{2}[X + 2R] = \frac{0{\cdot}25}{2}[(0{\cdot}0 + 0{\cdot}5) + 2(0{\cdot}06153 + 0{\cdot}22222 + 0{\cdot}39560)]$$

$$= 0{\cdot}23233$$

Now, by putting $1 + x^3 = t \quad \therefore x^2\,dx = \frac{dt}{3}$

$$\int_0^1 \frac{x^2}{1 + x^3}\,dx = \left[\frac{1}{3}\log(1 + x^3)\right]_0^1 = \frac{1}{3}\log 2 = 0{\cdot}23105$$

$\therefore$ Error in (1) $= 0{\cdot}23108 - 23105 = -\,0{\cdot}00003$

Error in (2) $= 0{\cdot}23233 - 0{\cdot}23105 = -\,0{\cdot}00128$

Thus, Simpson's Rule gives better approximation.

Example 8 : A rocket is launched from the ground. Its acceleration is registered during the first 80 seconds and is given in the following table.

t (sec.) :	0	10	20	30	40	50	60	70	80
a (m/s^2) :	30	31·63	33·34	35·47	37·75	40·33	43·25	46·69	50·67
Ordinate :	y_0	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8

By using Simpson's one-third rule, find the velocity at $t = 80$.

Sol. : By Simpson's (1/3)rd rule

$$I = \frac{h}{3}[X + 2E + 4O]$$

We have $X = 30 + 50·67 = 80·67$

$$E = 33·34 + 37·75 + 43·25 = 114·34$$

$$O = 31·63 + 35·47 + 40·33 + 46·69 = 154·12$$

$$\therefore \quad I = \frac{10}{3}[80·67 + 2(114·34) + 4(154·12)]$$

$$= \frac{10}{3}[925·83] = 3086·1 \text{ cm/sec.}$$

EXERCISE - I

1. Given the following values of e^x.

x :	0·0	0·5	1·0	1·5	2·0	2·5
e^x :	1·0	1·65	2·72	4·48	7·39	12·18

Evaluate $\int_0^{2·5} e^x\,dx$ using Trapezoidal rule. [Ans. : 11·415]

2. Find using the Trapezoidal rule from the following table the area bounded by the curve and x-axis from $x = 7·47$ to $x = 7·52$.

x :	7·47	7·48	7·49	7·50	7·51	7·52
$f(x)$:	1·93	1·95	1·98	2·01	2·03	2·06

[Ans. : 0·0996]

3. A curve is drawn to pass through the points given by

x :	1·0	1·5	2·0	2·5	3·0	3·5	4·0
y :	2·0	2·4	2·7	2·8	3·0	2·6	2·1

Estimate the area bounded by the curve between the x-axis and the lines $x = 1$ and $x = 4$, by Simpson's rule. [Ans. : 7·7833]

4. A river is 90 meter wide. The depth d meters from one bank is given by the following table. Find approximately the area of the cross-section.

a :	0	10	20	30	40	50	60	70	80	90
d :	0	4	7	9	12	15	14	8	3	1

[Ans. : 725]

5. Calculate upto 5 decimal places $\int_4^{5·2} \log_e x \cdot dx$ by using Trapezoidal rule from the following table.

x :	4·0	4·2	4·4	4·6	4·8	5·0	5·2
$\log_e x$:	1·3863	1·4351	1·4816	1·5260	1·5886	1·6094	1·6486

[Ans. : 1·83163]

6. Given the following values of $\sin x$

x :	0	$2\pi/10$	$4\pi/10$	$6\pi/10$	$8\pi/10$	π
$\sin x$:	0·000	0·5878	0·9511	0·9511	0·5878	0·000

Evaluate $\int_0^{\pi} \sin x\,dx$ by any suitable rule and compare it with its exact value. [Ans. : 0·9669, 1]

7. Apply Simpson's (1/3)rd rule to find $\int_0^6 \frac{1}{1+x}\,dx$. [Ans. : 1·9588]

8. Use any rule to evaluate approximately $\int_0^1 \frac{dx}{1+x^2}$ taking five equidistant intervals and find the value of π. [Ans. : 0·7837 ; $\pi = 3.1348$]

9. Calculate by trapezoidal rule an approximate value of $\int_0^1 e^x\,dx$ in steps of 0·20. [Ans. : 1·72399]

10. Evaluate in two ways $\int_4^{5·2} \log_e x\,dx$ by dividing the interval [4, 5·2] into six equal parts. [Ans. : 1·8279]

11. Evaluate $\int_0^1 \frac{dx}{1+x^2}$ by taking suitable sub-intervals and hence, find an approximate value of π by using (i) Trapezoidal rule, (ii) Simpson's (1/3)rd rule. [Ans. : (i) 0·7842, (ii) 0·7854, (iii) 3·1412]

12. Calculate $\int_2^{10} \frac{dx}{1+x}$ upto 4 decimal places by dividing the range into eight equal parts by Simpson's (1/3)rd rule. [Ans. : 1·2996]

13. Apply Simpson's (1/3)rd rule to find the value of $\int_0^6 \frac{dx}{1+x^2}$ upto two places of decimals. [Ans. : 1·37]

14. Evaluate $\int_{-3}^{3} x^4\,dx$ by Trapezoidal rule and compare it with the exact value. [Ans. : 111·27, 97·2]

15. Using Simpson's one third rule obtain $\int_1^2 \frac{dx}{x}$ taking 10 equal intervals. Hence, obtain approximate value of log 2 upto five places of decimals. [Ans. : 0·69315]

16. Evaluate $\int_{0·5}^{0·7} \sqrt{x}\,e^{-x}\,dx$ using Simpson's (1/3)rd rule. [Ans. : 0·08483]

5. Simpson's (3/8)th Rule

Consider a cubic curve passing through the points (x_0, y_0), (x_1, y_1), (x_2, y_2), (x_3, y_3). Let its equation be

$$y = a + b(x - x_1) + c(x - x_1)^2 + d(x - x_1)^3$$

When $x = x_0$, $y_0 = a + b(x_0 - x_1) + c(x_0 - x_1)^2 + d(x_0 - x_1)^3$

When $x = x_1$, $y_1 = a$

When $x = x_2$, $y_2 = a + b(x_2 - x_1) + c(x_2 - x_1)^2 + d(x_2 - x_1)^3$

When $x = x_3$, $y_3 = a + b(x_3 - x_1) + c(x_3 - x_1)^2 + d(x_3 - x_1)^3$

Since $x_0 - x_1 = -h$, $x_2 - x_1 = h$ and $x_3 - x_1 = -2h$

$$y_0 = a - bh + ch^2 - dh^3$$
$$y_1 = a$$
$$y_2 = a + bh + ch^2 + dh^3$$
$$y_3 = a + 2bh + 4ch^2 + 8\,dh^3$$

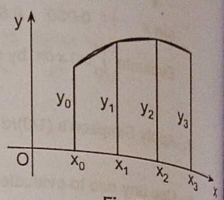


Fig. 7.5

Now, the area under the first cubic curve

$$= \int_{x_1-h}^{x_1+2h} f(x)\,dx$$

$$= \int_{x_1-h}^{x_1+2h} \left[a + b(x - x_1) + c(x - x_1)^2 + d(x - x_1)^3\right] dx$$

Put $x - x_1 = u$, $dx = du$. When $x = x_1 - h$, $-h = u$,

When $x = x_1 + 2h$, $2h = u$

$$\therefore \text{ Area} = \int_{-h}^{2h} (a + bu + cu^2 + du^3)\,du$$

$$= \left[au + \frac{bu^2}{2} + \frac{cu^3}{3} + \frac{du^4}{4}\right]_{-h}^{2h}$$

$$= \left[2ah + \frac{4bh^2}{2} + \frac{8ch^3}{3} + \frac{16ch^4}{4}\right] - \left[-ah + \frac{bh^2}{2} - \frac{ch^3}{3} + \frac{dh^4}{4}\right]$$

$$= 3ah + \frac{3bh^2}{2} + 3ch^3 + \frac{15dh^4}{4}$$

$$= \frac{3}{8}h\left[8a + 4bh^2 + 8ch^3 + 10dh^4\right]$$

It can be seen from (A) that this gives

$$\text{Area} = \frac{3h}{8}\left[y_0 + 3y_1 + 3y_2 + y_3\right]$$

Considering the area under the consecutive cubic curves and taking the sum.

$$\int_a^b f(x)\,dx = \frac{3h}{8}\left[(y_0 + 3y_1 + 3y_2 + y_3) + (y_3 + 3y_4 + 3y_5 + y_6) + \ldots\ldots + (y_{n-3} + 3y_{n-2} + 3y_{n-1} + y_n)\right]$$

$$= \frac{3h}{8}\left[(y_0 + y_n) + 3(y_1 + y_2 + y_4 + y_5 + \ldots\ldots) + 2(y_3 + y_6 + y_9 + \ldots\ldots)\right]$$

$$\therefore \boxed{\int_a^b f(x)\,dx = \frac{3h}{8}\left[X + 2T + 3R\right]}$$

where, X = sum of the eXtremes,

T = sum of the multiples of Three,

R = sum of the Remaining ordinates.

Note .... As anywhere else where calculus of finite difference is used our first point is (x_0, y_0) i.e. we start counting from 0. While using Simpson's (1/3)rd rule, the last ordinate should be even i.e. 2, 4, 6, ... etc. and while using Simpson's (3/8)th rule, the last ordinate should be multiple of 3 i.e. 3, 6, 9, 12, ... etc.

Example 1 : In an experiment a quantity G was measured as follows :

$G(20) = 95{\cdot}90$, $G(21) = 96{\cdot}85$, $G(22) = 97{\cdot}77$, $G(23) = 98{\cdot}68$

$G(24) = 99{\cdot}56$, $G(25) = 100{\cdot}41$, $G(26) = 101{\cdot}24$.

Compute $\int_{20}^{26} G(x)\,dx$ by Simpson's (3/8)th rule.

Sol. : The table can be prepared as

x	:	20	21	22	23	24	25	26
$G(x)$	:	95·90	96·85	97·77	98·68	99·56	100·41	101·24
Ordinate	:	y_0	y_1	y_2	y_3	y_4	y_5	y_6

By Simpson's (3/8)th rule,

$$\int_{20}^{26} G(x)\,dx = \frac{3h}{8}\left[X + 2T + 3R\right]$$

$$\int_{20}^{26} G(x)\,dx = \frac{3h}{8}\left[(y_0 + y_6) + 2(y_3) + 3(y_1 + y_2 + y_4 + y_5)\right]$$

Now, $y_0 + y_6 = 95{\cdot}90 + 101{\cdot}24 = 197{\cdot}14$

$y_1 + y_2 + y_4 + y_5 = 96{\cdot}85 + 97{\cdot}77 + 99{\cdot}56 + 100{\cdot}41 = 394{\cdot}59$

$y_3 = 98{\cdot}68$.

$$\therefore \int_{20}^{26} G(x)\,dx = \frac{3}{8}\left[197{\cdot}14 + 197{\cdot}36 + 1183{\cdot}77\right] = 591{\cdot}85$$

(See also Ex. 4, page 7-6.)

Example 2 : Evaluate $\int_0^{\pi/2} \frac{\sin x}{x}\,dx$ by Simpson's (3/8)th rule.

Sol. : We take $h = \pi / 12$ and prepare the following table.

x	:	0	$\pi/12$	$2\pi/12$	$3\pi/12$	$4\pi/12$	$5\pi/12$	$6\pi/12$
y	:	1	0·9886	0·9579	0·9003	0·8270	0·7379	0·6366
Ordinate	:	y_0	y_1	y_2	y_3	y_4	y_4	y_5

Since $\lim_{x \to 0} \frac{\sin x}{x} = 1$, we take $\frac{\sin x}{x} = 1$ at $x = 0$.

By Simpson's (3/8)th rule, $S = \frac{3h}{8}\left[X + 2T + 3R\right]$

We have $X = y_0 + y_6 = 1{\cdot}6366$, $T = y_3 = 0{\cdot}9003$

$R = y_1 + y_2 + y_4 + y_5 = 3{\cdot}5086$

$$\therefore S = \frac{3}{8} \cdot \frac{\pi}{12}\left[1{\cdot}6366 + 2(0{\cdot}9003) + 3(3{\cdot}.5084)\right] = 1{\cdot}37075$$

Example 3 : Evaluate $\int_0^{\pi} \frac{\sin^2 \theta}{5 + 4\cos\theta} d\theta$ by dividing $(0, \pi)$ into 6 sub-intervals by Simpson's (3/8)th rule.

Sol. : We take $h = \pi / 6$ and prepare the table

x :	0	$\pi/6$	$\pi/3$	$\pi/2$	$2\pi/3$	$5\pi/6$	π
y :	0	0·0295	0·1071	0·2000	0·2500	0·1628	0
Ordinate :	y_0	y_1	y_2	y_3	y_4	y_5	y_6

By Simpson's (3/8)th rule,

$$S = \frac{3h}{8}[X + 2T + 3R]$$

We have $X = y_0 + y_6 = 0$, $T = y_3 = 0{\cdot}2$.

$R = y_1 + y_2 + y_4 + y_5 = 0{\cdot}5494$.

$$\therefore \quad S = \frac{3}{8} \cdot \frac{\pi}{8}[0 + 2(0{\cdot}2) + 3(0{\cdot}5494)] = 0{\cdot}40216$$

(See also Ex. 6, page 7-8)

Example 4 : Evaluate $\int_0^6 \frac{dx}{1+x^2}$ by using Simpson's (3/8)th rule. Also find the error.

Sol. : We divide the interval into 6 equal parts by taking $h = 1$ and prepare the following table.

x :	0	1	2	3	4	5	6
$y = \frac{1}{1+x^2}$:	1	0·5	0·2	0·1	0·0588	0·0385	0·027
Ordinate :	y_0	y_1	y_2	y_3	y_4	y_5	y_6

By Simpson's (3/8)th rule,

$$\int_0^6 \frac{dx}{1+x^2} = \frac{3h}{8}[X + 2T + 3R] = \frac{3h}{8}[(y_0 + y_6) + 2y_3 + 3(y_1 + y_2 + y_4 + y_5)]$$

Now, $X = y_0 + y_6 = 1 + 0{\cdot}027 = 1{\cdot}027$, $T = y_3 = 0{\cdot}1$.

$R = y_1 + y_2 + y_4 + y_5 = 0{\cdot}5 + 0{\cdot}2 + 0{\cdot}0588 + 0{\cdot}0385 = 0{\cdot}7973$

$$\therefore \quad \int_0^6 \frac{dx}{1+x^2} = \frac{3}{8}[1{\cdot}027 + 2(0{\cdot}1) + 3(0{\cdot}7973)] = 1{\cdot}3571$$

Now, $\int_0^6 \frac{dx}{1+x^2} = [\tan^{-1} x]_0^6 = \tan^{-1} 6 = 1{\cdot}4056$

$\therefore$ Error in Simpson's (3/8)th rule,

$$= \int_0^6 \frac{dx}{1+x^2} - \int_0^6 P(x)\, dx = 1{\cdot}4056 - 1{\cdot}3735 = 0{\cdot}0485$$

Example 5 (a) : The velocity of train which starts from rest is given by the following table, the time being reckoned in minutes from the start and speed in km/hour.

Time :	3	6	9	12	15	18
Velocity :	22	29	31	20	4	0
Ordinate :	y_0	y_1	y_2	y_3	y_4	y_5

Estimate approximately the distance covered in 18 minutes by Simpson's (3/8)th rule.

Sol. : We know that $\frac{ds}{dt} = v \quad \therefore \; s = \int v\, dt$ *i.e.* $S = \int_0^{18} v\, dt$.

Since to use Simpson's (3/8) rule, we need 3, 6, 9, ... sub-intervals, we take one more ordinate at $t = 0$.

By data at $t = 0$, $v = 0$. $\therefore$ We have

y_0	y_1	y_2	y_3	y_4	y_5	y_6
0	22	29	31	20	4	0

By Simpson's (3/8)th rule [Note this]

$$S = \frac{3h}{8}[X + 2T + 3R] = \frac{3h}{8}[(y_0 + y_6) + 2y_3 + 3(y_1 + y_2 + y_4 + y_5)]$$

But $h = 3$ min. $= \frac{3}{60} = \frac{1}{20}$ hrs.

$$\therefore \quad S = \frac{3}{8} \cdot \frac{1}{20}[(0 + 0) + 2(31) + 3(22 + 29 + 20 + 4)] = 5{\cdot}38125$$

Example 6 : Evaluate $\int_0^{\pi/2} \sqrt{\sin x + \cos x} \cdot dx$ by Simpson's (3/8)th rule by dividing the interval into six intervals.

Sol. : We first prepare the following table, by taking sub-interval

$$h = \frac{(\pi/2) - 0}{6} = \frac{\pi}{12}$$

x :	0	$\pi/12$	$2\pi/12$	$3\pi/12$	$4\pi/12$	$5\pi/12$	$6\pi/12$
y :	1	1·1067	1·1688	1·1892	1·1688	1·1067	1·00
Ordinate :	y_0	y_1	y_2	y_3	y_4	y_5	y_6

Now, $X = y_0 + y_6 = 2$, $T = y_3 = 1{\cdot}1892$

$R = y_1 + y_2 + y_4 + y_5$

$= 1{\cdot}1067 + 1{\cdot}1688 + 1{\cdot}1688 + 1{\cdot}1067 = 4{\cdot}551$

$$\therefore \quad S = \frac{3h}{8}[X + 2T + 3R] = \frac{3 \times (\pi/12)}{8}[2 + 2(1{\cdot}1892) + 3(4{\cdot}551)]$$

$= 1{\cdot}7702$

Example 7 : Evaluate $\int_0^3 e^{\sqrt{x}}$ by Simpson's (3/8)th rule. Take $h = 0{\cdot}25$.

Sol. : We first prepare the table

x :	0	0·25	0·5	0·75	1·00	1·25	1·50	1·75	2·00	2·25	2·50	2·75	3·00
y :	1	1·6487	2·0281	2·3774	2·7183	3·0588	3·4033	3·7542	4·1133	4·4817	4·8605	5·2504	5·6522
Ordinate :	y_0	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8	y_9	y_{10}	y_{11}	y_{12}

By Simpson's (3/8)th rule,

$$S = \frac{3h}{8}[X + 2T + 3R]$$

We have $X = y_0 + y_{12} = 1 + 5{\cdot}6522 = 6{\cdot}6522$

$T = y_3 + y_6 + y_9 = 2{\cdot}3774 + 3{\cdot}4033 + 4{\cdot}4817 = 10{\cdot}2624$

$R = y_1 + y_2 + y_4 + y_5 + y_7 + y_8 + y_{10} + y_{11}$

$= 1{\cdot}6487 + 2{\cdot}0281 + 2{\cdot}7183 + 3{\cdot}0588 + 3{\cdot}7542 + 4{\cdot}1133 + 4{\cdot}8605 + 5{\cdot}2504$

$= 27{\cdot}4323$

$$\therefore\ S = \frac{3 \times 0{\cdot}25}{8}\left[6{\cdot}6522 + 2(10{\cdot}2624) + 3(27{\cdot}4323)\right] = 10{\cdot}2631$$

Example 8 : A curve is given by the table

x	:	0	1	2	3	4	5	6
y	:	0	2	2·5	2·3	2	1·7	1·5
Ordinate	:	y_0	y_1	y_2	y_3	y_4	y_5	y_6

The x-coordinate of the C.G. of the area bounded by the curve, the end-ordinates and the x-axis is given by $A\bar{x} = \int_0^6 xy\,dx$ where A is the area. Find $\bar{x}$ by using Simpson's (3/8)th rule.

Sol. : By Simpson's (3/8)th rule,

$$A = \frac{3h}{8}\left[X + 2T + 3R\right]$$

We have $X = y_0 + y_6 = 0 + 1{\cdot}5 = 1{\cdot}5$; $T = y_3 = 2{\cdot}3$

$R = y_1 + y_2 + y_4 + y_5 = 2 + 2{\cdot}5 + 2 + 1{\cdot}7 = 8{\cdot}2.$

$$\therefore\ A = \frac{3}{8}(1)\left[1{\cdot}5 + 2(2{\cdot}3) + 3(8{\cdot}2)\right] = \frac{3}{8}(30{\cdot}7) = 11{\cdot}5125$$

To find the integral $\int_0^6 xy\,dx$, we have to prepare the table of xy.

x	:	0	1	2	3	4	5	6
xy	:	0	2	5	6·9	8	8·5	9
Ordinate	:	y_0	y_1	y_2	y_3	y_4	y_5	y_6

$$S = \frac{3h}{8}\left[X + 2T + 3R\right]$$

We have $X = y_0 + y_6 = 0 + 9 = 9$; $T = y_3 = 6{\cdot}9$

$R = y_1 + y_2 + y_4 + y_5 = 2 + 5 + 8 + 8{\cdot}5 = 23{\cdot}5$

$$\therefore\ S = \frac{3}{8}(1)\left[9 + 2(6{\cdot}9) + 3(23{\cdot}5)\right] = 34{\cdot}9875$$

$$\therefore\ \bar{x} = \frac{S}{A} = \frac{34{\cdot}9875}{11{\cdot}5125} = 3{\cdot}0390$$

Example 9 : From the table given above find the area and volume of solid of revolution.

Sol. : (i) We have already calculated the area.

(ii) Since the volume of solid of revolution is given by

$$V = \int \pi\, y^2 dx$$

We prepare the following table for y^2.

x	:	0	1	2	3	4	5	6
y	:	0	4	6·25	5·29	4	2·89	2·25
Ordinate	:	y_0	y_1	y_2	y_3	y_4	y_5	y_6

By Simpson's (3/8)th rule,

$$\int y^2 dx = \frac{3h}{8}\left[X + 2T + 3R\right]$$

We have $X = y_0 + y_6 = 0 + 2{\cdot}25 = 2{\cdot}25$

$T = y_3 = 5{\cdot}29,\ R = y_1 + y_2 + y_4 + y_5$

$\therefore\ R = 4 + 6{\cdot}25 + 4 + 2{\cdot}89 = 17{\cdot}14$

$$\int y^2 dx = \frac{3 \times 1}{8}\left[2{\cdot}25 + 2(5{\cdot}29) + 3(17{\cdot}14)\right] = 24{\cdot}09375$$

$$\therefore\ \text{Volume} = \pi \int y^2 dx = \pi(24{\cdot}09375) = 75{\cdot}692748$$

Example 10 : Find the volume of solid of revolution formed by rotating about the x-axis the area bounded by the lines $x = 0$, $x = 1{\cdot}5$, $y = 0$ and the curve passing through

x	:	0·00	0·25	0·50	0·75	1·00	1·25	1·50
y	:	1·00	0·9826	0·9589	0·9089	0·8415	0·7624	0·7589

Sol. : Since the volume is given by $\int \pi\, y^2 dx$, we first prepare the table of y^2.

x	:	0·00	0·25	0·50	0·75	1·00	1·25	1·50
y	:	1·00	0·9655	0·9195	0·8261	0·7081	0·5812	0·5759
Ordinate	:	y_0	y_1	y_2	y_3	y_4	y_5	y_6

By Simpson's (3/8)th rule,

$$\int y^2 dx = \frac{3h}{8}\left[X + 2T + 3R\right]$$

We have $X = y_0 + y_6 = 1 + 0{\cdot}5759 = 1{\cdot}5789$

$T = y_3 = 0{\cdot}8261,\ R = y_1 + y_2 + y_4 + y_5$

$R = 0{\cdot}9655 + 0{\cdot}9155 + 0{\cdot}7081 + 0{\cdot}5812 = 3{\cdot}1703$

$$\therefore\ \int y^2 dx = \frac{3(0{\cdot}25)}{8}\left[1{\cdot}5759 + 2(0{\cdot}8261) + 3(3{\cdot}1703)\right] = 1{\cdot}19428$$

$$\therefore\ \text{Volume} = \pi \int y^2 dx = \pi(1{\cdot}19428) = 3{\cdot}7519$$

Miscellaneous Examples

Example 1 : Evaluate $\int_{-1}^{1} \frac{dx}{1+x^2}$ by using (i) Trapezoidal Rule, (ii) Simpson's (1/3)rd rule, and (iii) Simpson's (3/8)th rule.

(Also find the exact value.)

Sol. : Our first point is (x_0, y_0). Since for (1/3)rd rule, the last ordinate should be 2, 4, 6, 8, ..., etc. and for the (3/8)th rule the last ordinate should be 3, 6, 9, 12, ..., etc; to apply all the three rules we divide the interval – 1 to 1 into **six equal** parts by taking

$$\text{Subinterval} = \frac{1-(-1)}{6} = \frac{1}{3},\ \textit{i.e.},\ x_0 = -1,\ x_1 = -\frac{2}{3},\ x_2 = -\frac{1}{3},\ \text{etc.}$$

We now prepare the following table by calculating the values of $y_0, y_1, y_2, \ldots\ldots$

x	:	-1	$-\frac{2}{3}$	$-\frac{1}{3}$	0	$\frac{1}{3}$	$\frac{2}{3}$	1
$y=\frac{1}{1+x^2}$	:	$\frac{1}{2}$	$\frac{9}{13}$	$\frac{9}{10}$	1	$\frac{9}{10}$	$\frac{9}{13}$	$\frac{1}{2}$
Ordinate	:	y_0	y_1	y_2	y_3	y_4	y_5	y_6

(i) **By Trapezoidal Rule,**

$$I=\frac{h}{2}[X+2R]$$

Here, $h=\frac{1}{3}$, $X=$ Sum of the extremes $=\frac{1}{2}+\frac{1}{2}=1$

$R=$ Sum of the remaining

$$=\frac{9}{13}+\frac{9}{10}+1+\frac{9}{10}+\frac{9}{13}=\frac{18}{13}+\frac{18}{10}+1=\frac{180+234+130}{130}=\frac{544}{130}$$

$$\therefore\ I=\frac{(1/3)}{2}\left[1+2\left(\frac{544}{130}\right)\right]=\frac{1}{6}\left(1+\frac{544}{65}\right)=\frac{1}{6}\left(\frac{609}{65}\right)=1{\cdot}5615$$

(ii) **By Simpson's (1/3)rd Rule,**

$$I=\frac{h}{3}[X+2E+4O]$$

Here, $h=\frac{1}{3}$, $X=$ Sum of the extremes $=\frac{1}{2}+\frac{1}{2}=1$

$E=$ Sum of the even ordinates $=\frac{9}{10}+\frac{9}{10}=\frac{18}{10}$

$O=$ Sum of the odd ordinates $=\frac{9}{13}+1+\frac{9}{13}=\frac{18}{13}+1=\frac{31}{13}$

$$\therefore\ I=\frac{(1/3)}{3}\left[1+2\cdot\frac{18}{10}+4\cdot\frac{31}{13}\right]=\frac{1}{9}\left[1+\frac{468+1240}{130}\right]$$

$$=\frac{1}{9}\cdot\frac{1838}{130}=1{\cdot}5709$$

(iii) **Simpson's (3/8)th rule,**

$$I=\frac{3h}{8}[X+2T+3R]$$

Here, $h=\frac{1}{3}$, $X=$ Sum of the extremes $=\frac{1}{2}+\frac{1}{2}=1$

$T=$ Sum of the multiples of three $=1$

$R=$ Sum of the remaining $=\frac{9}{13}+\frac{9}{13}+\frac{9}{10}+\frac{9}{10}=\frac{18}{13}+\frac{18}{10}=\frac{414}{130}$

$$\therefore\ I=\frac{3(1/3)}{8}\left[1+2(1)+3\left(\frac{414}{130}\right)\right]=\frac{1}{8}\left[3+\frac{1240}{130}\right]$$

$$=\frac{1}{8}\left[\frac{390+1240}{130}\right]=1{\cdot}5692$$

Now, the exact value is given by

$$I=\int_{-1}^{1}\frac{dx}{1+x^2}=2\int_0^1\frac{dx}{1+x^2}=2\left[\tan^{-1}x\right]_0^1=2\cdot\frac{\pi}{4}=\frac{\pi}{2}=1{\cdot}57079.$$

Example 2 : Evaluate $\int_0^1\frac{dx}{1+x}$ by using (i) Trapezoidal Rule, (ii) Simpson's (1/3)rd rule, and (iii) Simpson's (3/8)th rule.

(Also find the exact value.)

Sol. : As above we divide the interval into **six** parts by taking each subinterval equal to $\frac{1-0}{6}=\frac{1}{6}$

x	:	0	$\frac{1}{6}$	$\frac{2}{6}$	$\frac{3}{6}$	$\frac{4}{6}$	$\frac{5}{6}$	1
$y=\frac{1}{1+x}$	:	1	$\frac{6}{7}$	$\frac{6}{8}$	$\frac{6}{9}$	$\frac{6}{10}$	$\frac{6}{11}$	$\frac{1}{2}$
Ordinate	:	y_0	y_1	y_2	y_3	y_4	y_5	y_6

(i) **By Trapezoidal Rule,**

$$I=\frac{h}{2}[X+2R]$$

Here, $h=\frac{1}{6}$, $X=$ Sum of the extremes $=1+\frac{1}{2}=\frac{3}{2}$

$R=$ Sum of the remaining

$$=6\left(\frac{1}{7}+\frac{1}{8}+\frac{1}{9}+\frac{1}{10}+\frac{1}{11}\right)=6(0{\cdot}56987)=3{\cdot}41922$$

$$\therefore\ I=\frac{1/6}{2}[1{\cdot}5+2(3{\cdot}41922)]=\frac{1}{12}[1{\cdot}5+6{\cdot}83844]=0{\cdot}6949$$

(ii) **By Simpson's (1/3)rd Rule,**

$$I=\frac{h}{3}[X+2E+4O]$$

Here, $h=\frac{1}{6}$, $X=$ Sum of the extremes $=1+\frac{1}{2}=\frac{3}{2}$

$E=$ Sum of the even ordinates $=\frac{6}{8}+\frac{6}{10}=6\cdot\frac{18}{80}=\frac{27}{20}$

$O=$ Sum of the odd ordinates $=\frac{6}{7}+\frac{6}{9}+\frac{6}{11}=6\left(\frac{1}{7}+\frac{1}{9}+\frac{1}{11}\right)=6\left(\frac{239}{693}\right)$

$$\therefore\ I=\frac{1/6}{3}\left[\frac{3}{2}+2\cdot\frac{27}{20}+4\cdot6\left(\frac{239}{693}\right)\right]$$

$$=\frac{1}{18}[1{\cdot}5+2{\cdot}7+8{\cdot}2771]=\frac{1}{18}[12{\cdot}4771]=0{\cdot}6932$$

(iii) **By Simpson's (3/8)th rule,**

$$I=\frac{3h}{8}[X+2T+3R]$$

Here, $h = \frac{1}{6}$, X = Sum of the extremes = $1 + \frac{1}{2} = \frac{3}{2}$

T = Sum of the multiples of three = $\frac{6}{9} = \frac{2}{3}$

R = Sum of the remaining = $\frac{6}{7} + \frac{6}{8} + \frac{6}{10} + \frac{6}{11} = 6(0{\cdot}4588) = 2{\cdot}7528$

$$\therefore\ I = \frac{3(1/6)}{8}\left[\frac{3}{2} + 2\cdot\frac{2}{3} + 3(2{\cdot}7528)\right] = \frac{1}{16}[1{\cdot}5 + 1{\cdot}3333 + 8{\cdot}2584]$$

$$= \frac{1}{16}[11{\cdot}0917] = 0{\cdot}6932$$

Now, the exact value is given by

$$I = \int_0^1 \frac{dx}{1+x} = [\log(1+x)]_0^1 = \log 2 = 0{\cdot}6931.$$

Example 3 : Compute the values of $\int_{0\cdot2}^{1\cdot4} (\sin x - \log_e x + e^x)\,dx$ by using (i) Trapezoidal Rule, (ii) Simpson's (1/3)rd rule, and (iii) Simpson's (3/8)th rule.

Sol. : We divide the interval into **six equal** sub-intervals by taking each sub-interval equal to $\frac{1{\cdot}4 - 0{\cdot}2}{6} = 0{\cdot}2$.

x	:	0·2	0·4	0·6	0·8	1·0	1·2	1·4
y	:	3·02950	2·79753	2·89759	3·16604	3·55975	4·06984	4·70418
Ordinate	:	y_0	y_1	y_2	y_3	y_4	y_5	y_6

(i) By Trapezoidal Rule,

$$I = \frac{h}{2}[X + 2R]$$

Here, $h = 0{\cdot}2$,

X = Sum of the extremes = 3·0295 + 4·70418 = 7·73368

R = Sum of the remaining

= 2·79753 + 2·89759 + 3·16604 + 3·55975 + 4·06984 = 16·49075

$$\therefore\ I = \frac{0{\cdot}2}{2}[7{\cdot}73368 + 2(16{\cdot}49075)] = 4{\cdot}071518$$

(ii) By Simpson's (1/3)rd Rule,

$$I = \frac{h}{3}[X + 2E + 4O]$$

Here, $h = 0{\cdot}2$,

X = Sum of the extremes = 3·0295 + 4·70418 = 7·73368

E = Sum of the even ordinates

= 2·89759 + 3·55975 = 6·45734

O = Sum of the odd ordinates

= 2·79753 + 3·16604 + 4·06984 = 10·03341

$$\therefore\ I = \frac{0{\cdot}2}{3}[7{\cdot}73368 + 2(6{\cdot}45734) + 4(10{\cdot}03341)] = 4{\cdot}05213$$

(iii) By Simpson's (3/8)th rule,

$$I = \frac{3h}{8}[X + 2T + 3R]$$

Here, $h = 0{\cdot}2$,

X = Sum of the extremes = 3·0295 + 4·70418 = 7·73368

T = Sum of the multiples of three = 3·16604

R = Sum of the remaining

= 2·79753 + 2·89759 + 3·55975 + 4·06984 = 13·32471

$$\therefore\ I = \frac{3(0{\cdot}2)}{8}[7{\cdot}73368 + 2(3{\cdot}16604) + 3(13{\cdot}32471)] = 4{\cdot}05299$$

Example 4 : Find the approximate value of $\int_0^6 e^x\,dx$ by using (i) Trapezoidal Rule, (ii) Simpson's (1/3)rd rule, and (iii) Simpson's (3/8)th rule.

Sol. : We divide the interval into **six equal** sub-intervals by taking $h = (6 - 0)/6 = 1$.

x	:	0	1	2	3	4	5	6
$y = e^x$	:	1·0000	2·7183	7·3891	20·0855	54·5981	148·4132	403·4288
Ordinate	:	y_0	y_1	y_2	y_3	y_4	y_5	y_6

(i) By Trapezoidal Rule,

$$I = \frac{h}{2}[X + 2R]$$

Here, $h = 1$,

X = Sum of the extremes = 1 + 403·4288 = 404·4288

R = Sum of the remaining

= 2·7183 + 7·3891 + 20·0855 + 54·5981 + 148·4132 = 233·2042

$$\therefore\ I = \frac{1}{2}[404{\cdot}4288 + 2(233{\cdot}2042)] = 435{\cdot}4186$$

(ii) By Simpson's (1/3)rd Rule,

$$I = \frac{h}{3}[X + 2E + 4O]$$

Here, $h = 1$,

X = Sum of the extremes = 1 + 403·4288 = 404·4288

E = Sum of the even ordinates = 7·3891 + 54·5981 = 61·9872

O = Sum of the odd ordinates = 2·7183 + 20·0855 + 148·4132 = 171·2170

$$\therefore\ I = \frac{1}{3}[404{\cdot}4288 + 2(61{\cdot}9872) + 4(171{\cdot}2170)] = 404{\cdot}4232$$

(iii) By Simpson's (3/8)th rule,

$$I = \frac{3h}{8}[X + 2T + 3R]$$

Here, $h = 1$,

X = Sum of the extremes = 1 + 403·4288 = 404·4288

T = Sum of the multiples of three = 20·0855

R = Sum of the remaining

$= 2{\cdot}7183 + 7{\cdot}3891 + 54{\cdot}5981 + 148{\cdot}4132 = 213{\cdot}1187$

$$\therefore\ I = \frac{3(1)}{8}\,[\,404{\cdot}4288 + 2\,(20{\cdot}0855) + 3\,(213{\cdot}1187)\,] = 406{\cdot}4835$$

Example 5 : Evaluate $\int_0^6 x\,f(x)\,dx$ by using (i) Trapezoidal Rule, (ii) Simpson's (1/3)rd rule using the following table.

x :	0	1	2	3	4	5	6
$f(x)$:	0·146	0·161	0·176	0·190	0·204	0·217	0·230

Sol. : We first prepare the table for $x \cdot f(x)$.

x :	0	1	2	3	4	5	6
$x \cdot f(x)$:	0	0·161	0·352	0·570	0·816	1·085	1·380
Ordinate :	y_0	y_1	y_2	y_3	y_4	y_5	y_6

(i) By Trapezoidal Rule,

$$I = \frac{h}{2}[\,X + 2R\,]$$

Here, $h = 1$,

X = Sum of the extremes = 1 + 1·380 = 1·380

R = Sum of the remaining

$= 0{\cdot}161 + 0{\cdot}352 + 0{\cdot}570 + 0{\cdot}816 + 1{\cdot}085 = 2{\cdot}984$

$$\therefore\ I = \frac{1}{2}\,[\,1{\cdot}380 + 2\,(2{\cdot}984)\,] = 3{\cdot}674$$

(ii) By Simpson's (1/3)rd Rule,

$$I = \frac{h}{3}[\,X + 2E + 4O\,]$$

Here, $h = 1$,

X = Sum of the extremes = 1 + 1·380 = 1·380

E = Sum of the even ordinates = 0·352 + 0·816 = 1·168

O = Sum of the odd ordinates = 0·161 + 0·570 + 1·085 = 1·8160

$$\therefore\ I = \frac{1}{3}\,[\,1{\cdot}380 + 2\,(1{\cdot}168) + 4\,(1{\cdot}8160)\,] = 3{\cdot}579$$

Example 6 : Find the approximate value of $\int_0^4 e^x\,dx$ by using (i) Trapezoidal Rule, (ii) Simpson's (1/3)rd rule.

(Compare yours result with the exact value.)

Sol. : We divide the interval into 4 equal parts by taking $h = (4 - 0) / 4 = 1$.

x :	0	1	2	3	4
$y = e^x$:	1	2·7183	7·3891	20·0855	54·5961
Ordinate :	y_0	y_1	y_2	y_3	y_4

(i) By Trapezoidal Rule,

$$I = \frac{h}{2}[\,X + 2R\,]$$

Here, $h = 1$,

X = Sum of the extremes = 1 + 54·5961 = 55·5961

R = Sum of the remaining = 2·7183 + 7·3891 + 20·0855 = 30·1929

$$\therefore\ I = \frac{1}{2}\,[\,55{\cdot}5961 + 2\,(30{\cdot}1929)\,] = 57{\cdot}9905$$

(ii) By Simpson's (1/3)rd Rule,

$$I = \frac{h}{3}[\,X + 2E + 4O\,]$$

Here, $h = 1$,

X = Sum of the extremes = 1 + 54·5961 = 55·5961

E = Sum of the even ordinates = 7·3891

O = Sum of the odd ordinates = 2·7183 + 20·0855 = 22·8038

$$\therefore\ I = \frac{1}{3}\,[\,55{\cdot}5961 + 2\,(7{\cdot}3891) + 4\,(22{\cdot}8038)\,] = 53{\cdot}8698$$

(Now, the exact value is given by

$$I = \int_0^4 e^x\,dx = \left[e^x\right]_0^4 = e^4 - e^0 = 54{\cdot}5961 - 1 = 53{\cdot}5961\ )$$

EXERCISE - II

1. A curve is drawn to pass through the points given by

x :	1.0	1.5	2.0	2.5	3.0	3.5	4.0
y :	2.0	2.4	2.7	2.8	3.0	2.6	2.1

Estimate the area bounded by the curve between the x-axis, and the lines $x = 1$ and $x = 4$ by Simpson's (3/8)th rule. [**Ans. :** 7·8375]

2. A river is 90 meter wide. The depth d meters from one bank is given by the following table. Find approximately the area of the cross-section by Simpson's (3/8)th rule.

x :	0	10	20	30	40	50	60	70	80	90
d :	0	4	7	9	12	15	14	8	3	1

[**Ans. :** 727·5]

3. Evaluate $\int_4^{5\cdot2} \log x\,dx$ from the following data using Simpson's (3/8)th rule.

x :	4	4·2	4·4	4·6	4·8	5	5·2
$\log x$:	1·3863	1·4351	1·4816	1·5261	1·5686	1·6094	1·6487

Also find the error. [**Ans. :** 1·8326475. Actual value = 1·827847409. The error = 0·0045]

4. Evaluate $\int \frac{dx}{1 + x^2}$ between suitable intervals and hence, find an approximate value of π by Simpson's (3/8)th rule. [**Ans. :** 0·7854, 3·1416]

5. A body is in the form of a solid of revolution. The diameter d in cms. of its sections at various distances x cms. from one end are as given below.

x :	0	2·5	5·0	7·5	10·0	12·5	15·0
d :	5	5·5	6·0	6·75	6·25	5·5	4·0

Estimate the volume of the solid by Simpson's (3/8)th rule.

(Hint : Calculate $\pi d^2 / 4$) [Ans. : 407·58 cub. cms.]

6. Using Simpson's (3/8)th rule, find $\int_0^6 f(x)\,dx$ where

x :	0	1	2	3	4	5	6
$f(x)$:	6·9897	7·4036	7·7815	8·1291	8·4510	8·7506	9·0309

[Ans. : 48·5396]

7. Use Simpson's (3/8)th rule to find $\int_{20}^{26} y\,dx$. Use the tabulated values of y given below.

x :	20	21	22	23	24	25	26
y :	95·20	96·85	97·77	98·68	96·56	100·41	101·24

[Ans. : 588·2137]

8. Using Simpson's (3/8)th rule, find $\int_0^{0\cdot3} \sqrt{1-8x^3}\cdot dx$. [Ans. : 0·2916]

9. Evaluate $\int_{-3}^{3} x^4 dx$ by Simpson's (3/8)th rule and compare it with the exact value. [Ans. : 99. Exact value 97·2]

10. Evaluate $\int_0^1 \frac{dx}{1+x}$ using Simpson's (3/8)th rule. [Ans. : 0·692]

11. Evaluate $\int_0^6 \frac{dx}{1+x}$ using Simpson's (3/8)th rule. Compare it with the exact value. [Ans. : 1·9661]

6. Weddle's Rule

Instead of considering four consecutive values as in Simpson's (3/8)th rule, we may consider six consecutive values, *i.e.*, we consider the intervals, x_0 to x_6, x_6 to x_{12}, ..., etc. Hence, putting $n = 6$ in the general formula (page 7-3), we get (after simplification)

$$\int_{x_0}^{x_6} y\,dx = h\left[6y_0 + 18\,\Delta y_0 + 27\,\Delta^2 y_0 + 24\,\Delta^3 y_0 + \frac{123}{10}\Delta^4 y_0 + \frac{33}{10}\Delta^5 y_0 + \frac{41}{140}\Delta^6 y_0\right]$$

We can replace $\frac{41}{140}$ by $\frac{3}{10}$ since the error can be neglected. Putting the values of $\Delta y_0, \Delta^2 y_0$, ..., $\Delta^6 y_0$ in terms of $y_1, y_2, ..., y_6$ and after simplification, we get,

$$\int_{x_0}^{x_6} y\,dx = \frac{3h}{10}[y_0 + 5\,y_1 + y_2 + 6\,y_3 + y_4 + 5\,y_5 + y_6]$$

Similarly, $$\int_{x_6}^{x_{12}} y\,dx = \frac{3h}{10}[y_6 + 5\,y_7 + y_8 + 6\,y_9 + y_{10} + 5\,y_{11} + y_{12}]$$

.....

.....

$$\int_{x_{n-6}}^{x_n} y\,dx = \frac{3h}{10}[y_{n-6} + 5\,y_{n-5} + y_{n-4} + 6\,y_{n-3} + y_{n-2} + 5\,y_{n-1} + y_n]$$

Adding all these integrals,

$$\int_a^b f(x)\,dx = \int_a^b y\,dx = \frac{3h}{10}[y_0 + 5\,y_1 + y_2 + 6\,y_3 + y_4 + 5\,y_5 + 2\,y_6 + 5\,y_7 + y_8 + ... + y_n]$$

This is known as Weddle's Rule and is more accurate than any rule studied earlier.

Notes ....

1. In deriving the Weddle's Rule, it is assumed that a sixth degree polynomial of the form $y = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + a_4 x^4 + a_5 x^5 + a_6 x^6$ passes through the points (x_0, y_0) (x_1, y_1) .... (x_6, xy_6).

2. Weddle's Rule can be applied when the number of sub-intervals is 6 or a multiple of 6. In other words last ordinate should be 6, 12, 18, ..., etc.

For 6 sub-intervals,

$$\boxed{\int_a^b f(x)\,dx = \frac{3h}{10}[y_0 + 5\,y_1 + y_2 + 6\,y_3 + y_4 + 5\,y_5 + y_6]}$$

For 12 sub-intervals,

$$\boxed{\int_a^b f(x)\,dx = \frac{3h}{10}[(y_0 + 5\,y_1 + y_2 + 6\,y_3 + y_4 + 5\,y_5 + y_6) + (y_6 + 5\,y_7 + y_8 + 6\,y_9 + y_{10} + 5y_{11} + y_{12})]}$$

7. To Find Volume, Distance, Work etc.

If on the x-axis we have distances and also on the y-axis we have distances, the above rules which depend upon the product give total area. If, however, we have areas on the y-axis and distances on the x-axis, the rules give us total volume. If we have velocities on the y-axis and time on the x-axis, the rules give us distance travelled. If we have forces on the y-axis and distances on the x-axis then the rules give us the total work done.

8. Error in Quadrature Formulae

Since we have assumed that a polynomial $P(x)$ passes through the given points instead of the original curve $f(x)$, the error in the quadrature formulae is given by

$$E = \int_a^b f(x)\,dx - \int_a^b P(x)\,dx$$

Example 1 : Evaluate $\int_0^6 \frac{dx}{1+x^2}$ by using (i) Simpson's (3/8)th rule, (ii) Weddle's rule. Also find the error in each case. **(S.U. 1987, 91, 97, 2002)**

Sol. : We divide the interval into 6 equal parts by taking $h = 1$ and prepare the following table.

x	0	1	2	3	4	5	6
$y = \frac{1}{1+x^2}$	1	0·5	0·2	0·1	0·0588	0·0385	0·027

(i) By Simpson's (3/8)th rule

$$\int_0^6 \frac{dx}{1+x^2} = \frac{3h}{8}[X + 2T + 3R]$$

$$= \frac{3h}{8}[(y_0 + y_6) + 2y_3 + 3(y_1 + y_2 + y_4 + y_5)]$$

Now, $y_0 + y_6 = 1 + 0{\cdot}027 = 1{\cdot}027s$

$y_1 + y_2 + y_4 + y_5 = 0{\cdot}5 + 0{\cdot}2 + 0{\cdot}0588 + 0{\cdot}0385 = 0{\cdot}7973$

$$\therefore \int_0^6 \frac{dx}{1+x^2} = \frac{3}{8}[1{\cdot}027 + 2(0{\cdot}1) + 3(0{\cdot}7973)] = 1{\cdot}3571$$

(ii) By Weddle's rule

$$\int_0^6 \frac{dx}{1+x^2} = \frac{3h}{10}[y_0 + 5y_1 + y_2 + 6y_3 + y_4 + 5y_5 + y_6]$$

$$= 0{\cdot}3\,[1 + 5(0{\cdot}5) + 0{\cdot}2 + 6(0{\cdot}1) + 0{\cdot}0588 + 5(0{\cdot}0385) + 0{\cdot}027]$$

$$= 0{\cdot}3\,[4{\cdot}5783] = 1{\cdot}3735$$

Now, $\int_0^6 \frac{dx}{1+x^2} = \left[\tan^{-1} x\right]_0^6 = \tan^{-1} 6 = 1{\cdot}4056$

$\therefore$ Error in Simpson's (3/8)th rule,

$$= \int_0^6 \frac{dx}{1+x^2} - \int_0^6 P(x)\,dx = 1.4056 - 1.3571 = 0.0485$$

Error in Weddle's rule,

$$= \int_0^6 \frac{dx}{1+x^2} - \int_0^6 P(x)\,dx = 1{\cdot}4056 - 1{\cdot}3735 = 0{\cdot}0321.$$

Note ....

Observe that Weddle's rule gives better approximation than the Simpson's (3/8)th rule.

Example 2 : The velocity of train which starts from rest is given by the following data, the time being reckoned in minutes from the start and speed in km/hour.

Time	:	3	6	9	12	15	18
Velocity	:	22	29	31	20	4	0

Estimate approximately the distance covered in 18 minutes by Simpson's rule and also by Weddle's rule.

Sol. : We know that $\frac{dS}{dt} = v \quad \therefore \quad S = \int v\,dt \quad i.e., \quad S = \int_0^{18} v\,dt$

Since to use Simpson's (3/8)th rule and also to use Weddle's rule, we need seven coordinates we take one more ordinate at $t = 0$.

By data at $t = 0$, $v = 0$. Therefore, we have

y_0	y_1	y_2	y_3	y_4	y_5	y_6
0	22	29	31	20	4	0

(i) By Simpson's (3/8)th rule

$$S = \frac{3h}{8}[X + 2T + 3R]$$

$$= \frac{3h}{8}[(y_0 + y_6) + 2y_3 + 3(y_1 + y_2 + y_4 + y_5)]$$

But $h = 3$ min. $= \frac{3}{60} = \frac{1}{20}$ hrs.

$$\therefore \quad S = \frac{3}{8} \cdot \frac{1}{20}[(0 + 0) + 2(31) + 3(22 + 29 + 20 + 4)]$$

$$= 5{\cdot}38125$$

(ii) By Weddle's rule

$$S = \frac{3h}{10}[y_0 + 5y_1 + y_2 + 6y_3 + y_4 + 5y_5 + y_6]$$

$$= \frac{3}{10} \cdot \frac{1}{20}[0 + 5(22) + 29 + 6(31) + 20 + 5(4)\ 0] \qquad \left[\because h = \frac{1}{20}\right]$$

$$\therefore \quad S = 5{\cdot}475$$

Example 3 : Compute the value of $\int_{0\cdot2}^{1\cdot4} (\sin x - \log_e x + e^x)\,dx$ by Simpson's (3/8)th rule and also by Weddle's rule. **(S.U. 1989, 91, 1997, 98)**

Sol. : We prepare the following table

x	:	0·2	0·	0·6	0·8	1·0	1·2	1·4
y	:	3·02950	2·79753	2·89759	3·16604	3·55975	4·06984	4·70418
		y_0	y_1	y_2	y_3	y_4	y_5	y_6

(i) By Simpson's (3/8)th rule

$$S = \frac{3h}{8}[X + 2T + 3R]$$

Here, $X = y_0 + y_6 = 7{\cdot}73368, \qquad T = y_3 = 3{\cdot}16604$

$R = y_1 + y_2 + y_4 + y_5$

$= 2{\cdot}79753 + 2{\cdot}89459 + 3{\cdot}55975 + 4{\cdot}06984$

$= 13{\cdot}32471$

$$\therefore \quad S = \frac{3}{8}(0{\cdot}2)[7{\cdot}7336 + 2(3{\cdot}16604) + 3(13{\cdot}32471)] = 4{\cdot}0530$$

(ii) By Weddle's rule

$$S = \frac{3h}{10}[y_0 + 5y_1 + y_2 + 6y_3 + y_4 + 5y_5 + y_6]$$

$$= \frac{3}{10}(0{\cdot}2)[3{\cdot}02950 + 5(2{\cdot}79753) + 2{\cdot}89759 + 6(3{\cdot}16604) + 3{\cdot}55975 + 5(4{\cdot}06984) + 4{\cdot}70418]$$

$$= 4{\cdot}0514$$

Example 4 : A curve is given by the table

x :	0	1	2	3	4	5	6
y :	0	2	2·5	2·3	2	1·7	1·5

The x-coordinate of the C.G. of the area bounded by the curve, the end-ordinates and the x-axis is given by $A\bar{x} = \int_0^6 xy\,dx$ where A is the area. Find $\bar{x}$ by using Simpson's (3/8)th rule and also by Weddle's rule.

Sol. : (i) By Simpson's (3/8)th rule

$$A = \frac{3h}{8}\,[\,X + 2T + 3R\,]$$

Here, $X = y_0 + y_6 = 0 + 1{\cdot}5 = 1{\cdot}5, \qquad T = y_3 = 2{\cdot}3,$

$R = y_1 + y_2 + y_4 + y_5 = 2 + 2{\cdot}5 + 2 + 1{\cdot}7 = 8{\cdot}2$

$$\therefore\ A = \frac{3}{8}\,(1)\,[\,1{\cdot}5 + 2\,(2{\cdot}3) + 3\,(8{\cdot}2)\,] = \frac{3}{8}\,(30{\cdot}7) = 11{\cdot}5125$$

To find the integral $S = \int_0^6 xy\,dx$, we have to prepare the table of xy.

x :	0	1	2	3	4	5	6
xy :	0	2	5	6·9	8	8·5	9

$$S = \frac{3h}{8}\,[\,X + 2T + 3R\,]$$

Here, $X = y_0 + y_6 = 0 + 9 = 9, \qquad T = y_3 = 6{\cdot}9$

$R = y_1 + y_2 + y_4 + y_5 = 2 + 5 + 8 + 8{\cdot}5 = 23{\cdot}5$

$$\therefore\ S = \frac{3}{8}\,(1)\,[\,9 + 2\,(6{\cdot}9) + 3\,(23{\cdot}5)\,] = 34{\cdot}9875$$

$$\therefore\ \bar{x} = \frac{S}{A} = \frac{34{\cdot}9875}{11{\cdot}5125} = 3.0390.$$

(ii) By Weddle's rule

$$S = \frac{3h}{10}\,[\,y_0 + 5\,y_1 + y_2 + 6\,y_3 + y_4 + 5\,y_5 + y_6\,]$$

Putting the values of $y_0, y_1, \ldots, y_6$ from the given table

$$\therefore\ A = \frac{3}{10}\,[\,0 + 5\,(2) + 2{\cdot}5 + 6\,(2{\cdot}3) + 2 + 5\,(1{\cdot}7) + 1{\cdot}5\,] = 11{\cdot}49$$

To find the integral, we substitute the values of $y_0, y_1, \ldots, y_6$ from the table prepared

$$\therefore\ S = \frac{3h}{10}\,[\,y_0 + 5\,y_1 + y_2 + 6\,y_3 + y_4 + 5\,y_5 + y_6\,]$$

$$\therefore\ S = \frac{3}{10}\,[\,0 + 5\,(2) + 5 + 6\,(6{\cdot}9) + 8 + 5\,(8{\cdot}5) + 9\,] = 34{\cdot}77$$

$$\therefore\ \bar{x} = \frac{S}{A} = \frac{34{\cdot}77}{11{\cdot}49} = 3.0261.$$

EXERCISE - III

1. A curve is drawn to pass through the points given by

x :	1·0	1·5	2·0	2·5	3·0	3·5	4·0
y :	2·0	2·4	2·7	2·8	3·0	2·6	2·1

Estimate the area bounded by the curve between the x-axis, the lines $x = 1$ and $x = 4$ by Simpson's (3/8)th rule and Weddle's rule. **(S.U. 1987, 88, 99)** [**Ans. :** 7·8375, 7·74

2. A river is 90 meter wide. The depth d meters from one bank is given by the following table. Find approximately the area of the cross-section.

x :	0	10	20	30	40	50	60	70	80	90
y :	0	4	7	9	12	15	14	8	3	1

(S.U. 1985, 0, 2000) [**Ans. :** 727.5

3. Evaluate $\int_0^1 \frac{dx}{1+x^2}$ using Simpson's (3/8)th rule and Weddle's rule. **(S.U. 1990, 95)**

[**Ans. :** 0·7854, 0·78535]

4. Evaluate $\int_4^{5\cdot2} \log x\,dx$ from the following data using Simpson's (3/8)th rule and Weddle's rule.

x :	4	4·2	4·4	4·6	4·8	5	5·2
$\log x$:	1·3863	1·4351	1·4816	1·5261	1·5686	1·6094	1·6487

Also find the error in each case. **(S.U. 1987, 95, 97, 99)**

[**Ans. : (i)** 1·8278475, **(ii)** 1·8278474. Actual value = 1·827847409. The error is negligible.]

5. Evaluate $\int \frac{dx}{1+x^2}$ between suitable intervals and hence find an approximate value of π by Simpson's (3/8)th rule. [**Ans. :** 0·7854, 3·1416]

6. A body is in the form of a solid of revolution. The diameter d in cms. of its sections at various distances x cms from one end are as given below :

x :	0	2·5	5·0	7·5	10·0	12·5	15·0
y :	5	5·5	6·0	6·75	6·25	5·5	4·0

Estimate the volume of the solid by Weddle's method. **(S.U. 2001)**

(**Hint :** Calculate $\frac{\pi d^2}{4}$.)

[**Ans. :** 407.58 cub.cms]

7. Using Simpson's (3/8)th rule and Weddle's rule, find $\int_0^6 f(x)\,dx$ where

x :	0	1	2	3	4	5	6
$f(x)$:	6·9897	7·4036	7·7815	8·1291	8·4510	8·7506	9·0309

[**Ans. : (i)** 48·5396, **(ii)** 48·5396]

8. Use Simpson (3/8)th rule to find $\int_{20}^{26} y\,dx$. Use the tabulated values of y given below.

x :	20	21	22	23	24	25	26
y :	95·20	96·85	97·77	98·68	99·56	100·41	101·24

(S.U. 2000) [**Ans. :** 591·5887]

9. Find the value of the integral $\int_0^1 \frac{x^2}{1+x^3}\,dx$ using any two numerical methods and compare each of them with the exact value. **(S.U. 1996, 97)** [**Ans. :** Exact value = 0·2310]

10. Using Simpson's (3/8)th rule find $\int_0^{0\cdot3} \sqrt{1-8x^3}\cdot dx$. **(S.U. 1994, 98,99)**

[**Ans. :** 0·29162]

☆☆☆

CHAPTER 8

Introduction to Fuzzy Sets

Conventions

We denote by

$\mathbb{N}$ the set of all natural numbers 1, 2, 3, ...

$\mathbb{Z}$ the set of all integers ....., – 2, – 1, 0, 1, 2, ...

$\mathbb{Q}$ the set of all fractions

$\mathbb{R}$ the set of all real numbers

(a, b) the set $\{ x \in \mathbb{R} \mid a < x < b \}$

$[a, b]$ the set $\{ x \in \mathbb{R} \mid a \le x \le b \}$

$(a, b]$ the set $\{ x \in \mathbb{R} \mid a < x \le b \}$

$[a, b)$ the set $\{ x \in \mathbb{R} \mid a \le x < b \}$ where a and b are real numbers.

If X is universal set, then we denote by

$p(X)$ the power set of X

$\mathcal{F}(X)$ the fuzzy power set of X

1. Why Fuzzy Set

"As far as laws of Mathematics refer to reality, they are not certain. As far as they are certain, they do not refer to reality."

Albert Einstein

The adjective fuzzy is used to describe an entity which has no sharply demarcated boundary. What we understand through natural languages is usually fuzzy : *fast* car, *hot* water, *cold* weather, *tall* man, *effective* medicine, *complicated* structure are all fuzzy concepts. On the other hand mathematics is always precise. Therefore, when we apply precise mathematics to phenomena described in natural languages, we have problems.

In 1965, Professor of Control Engineering from California University, Lutfi Zadeh, introduced the concept of fuzzy set which offers a reasonably satisfactory mathematical model for fuzzy phenomena. In order to understand what Zadeh achieved through the concept of fuzzy set, we now, see where exactly our set theory falters to represent reality.

We shall call our usual set a 'crisp set' when we need to emphasize the difference between classical set and fuzzy set.

Lotfi A. Zadeh

Professor Lutfi A. Zadeh describes himself as "an American mathematically oriented electrical engineer of Iranian descent born in Russia". Born on February 4, 1921 to Iranian father and Russian Jew mother in Baku, now the capital of Azerbaijan, Zadeh is graduated from Teheran University, M.S. from M.I.T., and Ph.D. from Columbia University. He is best known for his pioneer work in Fuzzy Set Theory, Fuzzy Logic, Z-numbers and Z-transforms.

At the age of 93, he is Professor Emiritus at University of California, Berkeley.

2. Problems With (Crisp) Sets

The concept of (Crisp) set is fundamental in mathematics. We start with the existence of universal set or the set of discourse. We shall denote it by X. A desired set A is, then, constructed by selecting elements x of X which satisfy property P. This we write in symbols as

$$A = \{ x \in X \mid P(x) \}$$

For example, the set of points on a unit circle is written as

$$A = \{ (x, y) \in \mathbb{R} \times \mathbb{R} \mid x^2 + y^2 = 1 \}$$

Here, the universal set is $\mathbb{R} \times \mathbb{R}$, the set of ordered pairs of real numbers (x, y), the property $P(x, y)$ must satisfy is $x^2 + y^2 = 1$.

Now suppose, the Planning Commission of India is interested in finding the set A_1 of poor people in India. Clearly, X will be the set of all citizens of India. But the property "poor" is not precise. Hence, even if we write

$$A_1 = \{ x \in X \mid x \text{ is poor} \},$$

A_1 is mathematically useless, unless we precisely define poverty.

Suppose Government of India passes a law that one who earns ₹ 5000 or less, per month will be considered poor. We introduce a symbol $E(x)$ for monthly earning of Mr. x. Then clearly, the set of the poor in India is

$$A_2 = \{ x \in X \mid E(x) \le 5000 \}$$

Now, A_2 is precise and is fit for Mathematical operations. But does it represent reality?

Suppose E (Ram) = 5000 and E (Laxman) = 5001. Clearly Ram $\in A_2$ and Laxman $\notin A_2$. But this is not an academic question. Ram $\in A_2$, therefore Ram will get all concession which are due for a person of economically backward class. And Laxman will be denied these concessions simply because he earns *one rupee more* than Ram! Usually such problems are solved by condoning one rupee.

We can define condoning as follows :

$$S(x) = S(y) \text{ if and only if } | E(x) - E(y) | \le 1,$$

where, $S(x)$ denotes the financial status of Mr. x.

By this definition the financial status of Ram and Laxman is the same and both will get the same benefits. But this does not solve the problem. Suppose E (Krishna) = 5002. Then clearly,

$$S(\text{Laxman}) = S(\text{Krishna}) \text{ and } S(\text{Ram}) \ne S(\text{Krishna})$$

i.e., the equality relation that we defined is not transitive. Suppose the court of justice decrees that

$$S(x) = S(y),\ S(y) = S(z) \Rightarrow S(x) = S(z).$$

But this will make our definition of poverty ridiculous. For, then every-body will have the same status.

Arrange all elements of A_2 in the order $x_0, x_1, x_2, \ldots, x_n$ such that $|E(x_i) - E(x_{i-1})| \le 1$; for $i = 1, \ldots, n$. Then by the court decree, we have $S(x_0) = S(x_1) = \ldots = S(x_n)$.

Thus, the set of poor people of India is either mathematically useless or practically meaningless.

3. Fuzzy Set : Coping with Reality

Zadeh resolve this problem by assigning grade to element x of X for being a member of A. When A is crisp or well-defined, x has two possible grades : it is 1, if $x \in A$ and 0, if $x \notin A$.

Zadeh assigns intermediate grades when A is not crisply defined. So for A_1, in § 2 above, Zadeh may assign x the grade

$$\begin{array}{lll} 1 & \text{if} & E(x) \le 5000; \\ 0{\cdot}99 & \text{if} & E(x) = 6000; \\ 0{\cdot}90 & \text{if} & E(x) = 8000; \\ 0{\cdot}50 & \text{if} & E(x) = 20{,}000; \\ 0{\cdot}1 & \text{if} & E(x) = 30{,}000. \end{array}$$

More generally, if A is a fuzzy concept which has meaning for every x in the universal set X and if we denote by $A(x)$ the grade of membership of x in A, then $A(x)$ is a real number between 0 and 1.

$A(x) = 0$, if x is not a member of A, and

$A(x) = 1$, if x is a full member of A.

We may then give grades between 0 and 1 based on our subjective judgement. For example, say

$A(x) = 0{\cdot}01$, if x is "less likely" to be a member of A;

. . . . . .

$A(x) = 0{\cdot}50$, if x may or may not be a member of A;

. . . . . .

$A(x) = 0{\cdot}9$, if x "almost surely" a member of A.

We therefore have a definition :

Definition : Let X be a universal set and [0, 1] be the closed unit interval of real numbers between 0 and 1 including 0 and 1. Then a fuzzy set A on X is a function :

$$A : X \to [0, 1].$$

If $x \in X$, then $A(x)$ is called the grade of membership of x in A.

Note ....

By the above definition probability function is also a fuzzy set. But the probability represents objective uncertainty, while fuzzy set theory models subjective 'uncertainty'. If a coin is tossed the probability of getting Head is always 0·5 *provided the coin is unbiased.* But whether the coin is unbiased or not is a subjective uncertainty and hence is in the realm of fuzzy set theory.

EXERCISE - I

1. Write a note on a fuzzy set and its membership function. (S.U. 2008)
2. Describe the concept of fuzzy set in your own words. (S.U. 2009, 11)
3. Explain the need of fuzzy set theory.

4. Basic Terms in Fuzzy Sets

(a) Definition and Notations

Throughout the rest of the book, unless stated otherwise, X stands for the universal set and [0, 1] denotes the set of all real numbers between 0 and 1 including both.

Definition : A *fuzzy* set A on X is a function $A : X \to [0, 1]$.

If $x \in X$, then $A(x)$ is called the **grade of membership of x in A.**

Note ....

To avoid trivial empty statements made for logical exactness we assume that X is non-empty set.

Notations

There are different ways to denote a fuzzy set.

1. Some times a distinction is made between a fuzzy concept and the function representing the concept. In such a case we say, A is a fuzzy set with membership function $\mu_A : X \to [0, 1]$.

The idea behind such notations is the assumption that the same fuzzy concept A may have different membership functions μ_A.

We do not make this distinction.

2. When X has n elements say $x_1, x_2, \ldots, x_n$ (with fixed order) and $A(x_i) = a_i$ for $i = 1, 2, \ldots, n$, then we may write $A = (a_1, a_2, \ldots, a_n)$.

3. If $X = \{x_1, x_2, \ldots, x_n\}$ and $A(x_i) = a_i$, then we also write

$$A = \frac{a_1}{x_1} + \frac{a_2}{x_2} + \ldots + \frac{a_n}{x_n}$$

Note that the sign + here is just a punctuation mark and not any algebraic operation.

We also write $A = \sum_1^n \frac{a_i}{x_i} = \sum_1^n \frac{A(x_i)}{x_i}$

More generally, if $X = \{x_1, x_2, \ldots\}$ is an infinite set, we may write

$$A = \sum \frac{A(x_i)}{x_i}$$

Illustrative Examples

Example 1 : If $X = \{a, b, c, d\}$ and A is a fuzzy set defined on X, by $A(a) = 0$, $A(b) = 0{\cdot}5$, $A(c) = 0{\cdot}8$, $A(d) = 1$, then we write

$$A = \frac{0}{a} + \frac{0{\cdot}5}{b} + \frac{0{\cdot}8}{c} + \frac{1}{d}$$

There is a convention in this notation not to mention the elements of grade membership 0.

If we follow this convention, then

$$A = \frac{0{\cdot}5}{b} + \frac{0{\cdot}8}{c} + \frac{1}{d}.$$

Example 2 : If $A : \mathbb{N} \to [0, 1]$

$A(x) = 1/x,$

then we write $A = \Sigma \dfrac{1/n}{n}.$

When X is "Continuous" subset of real numbers (*i.e.*, union of intervals say), then the fuzzy set

$A : X \to [0, 1]$

is also denoted as $A = \int_X \dfrac{A(x)}{x}.$

The sign $\int$ has no relation with integration operation in calculus.

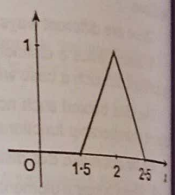
Example 3 : $A : \mathbb{R} \to [0, 1]$

$A(x) = e^{-x^2}$

may be denoted as $A = \int \dfrac{e^{-x^2}}{x}.$

Example 4 : "Almost two" is a fuzzy set $A : \mathbb{R} \to [0, 1]$.

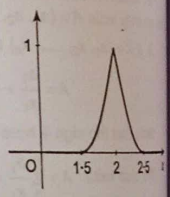
$$\begin{aligned} A(x) &= 0, && \text{for } x \le 1{\cdot}5 \\ &= 2x - 3, && \text{for } 1{\cdot}5 \le x \le 2 \\ &= 5 - 2x, && \text{for } 2 \le x \le 2{\cdot}5 \\ &= 0 && \text{for } x \ge 2{\cdot}5 \end{aligned}$$



The graph of A looks like the adjacent figure.

Example 5 : "A number very close to 2" may be defined as a fuzzy set $B : \mathbb{R} \to [0, 1]$.

$$\begin{aligned} B(x) &= 0, && \text{for } x \le 1{\cdot}5 \text{ or } x \ge 2{\cdot}5 \\ &= (2x - 3)^2, && \text{for } 1{\cdot}5 \le x \le 2 \\ &= (5 - 2x)^2, && \text{for } 2 \le x \le 2.5 \end{aligned}$$



The graph of B looks like the adjacent figure.

5. Crisp Set and Fuzzy Set

Recall that given a (Crisp) set $A = \{ x \in X \mid P(x) \}$, we have $x \in A$, if and only if $P(x)$ is a true statement *i.e.*, $P(x)$ has truth value 1. Similarly, $x \notin A$ if and only if $P(x)$ is a false statement or statement of truth value 0. For this reason we can identify the crisp set A with the fuzzy set

$$\begin{aligned} A &: X \to [0, 1] \\ A(x) &= 1, && \text{if } P(x) \text{ is true} \\ &= 0, && \text{if } P(x) \text{ is false} \end{aligned}$$

(This fuzzy set is usually called the characteristic function of the **crisp set** A and is denoted by χ_A

Thus, every crisp set is a fuzzy set. This fact is to be noted while extending a "crisp property" to a fuzzy set. Then all the mathematical results valid in classical theory will remain valid in fuzzy set theory, when a crisp set is treated as a fuzzy set.

We consider the universal set X as the fuzzy set on X defined by

$$X(x) = 1, \quad \text{for all } x \in X$$

and empty set Φ as a fuzzy set defined by $\Phi(x) = 0$ for all $x \in X$. We denote the crisp sets X and Φ and their counter parts in fuzzy set theory by the same symbols.

Definition : The set of all fuzzy sets on X is called the **fuzzy power set** of X and is denoted by $\mathcal{F}(x)$.

We will denote the power set of X (*i.e.*, the set of all subsets of X including X and Φ) by $p(X)$. Clearly, $p(X) \subseteq \mathcal{F}(x)$.

Note that if $X = \{x_1, x_2, \ldots\ldots, x_n\}$ $p(X)$ has 2^n elements. But $\mathcal{F}(x)$ is never finite when $X \neq \Phi$.

6. Basic Concepts

Definition 1 : Two fuzzy sets A and B on X are said to be equal, if $A(x) = B(x)$ for all $x \in X$. In symbols we, then, write $A = B$.

Definition 2 : Let A and B be two fuzzy sets on X. A is called a subset of B (or A is said to be contained in B), if $A(x) \le B(x)$ for all $x \in X$. In symbols we, then, write $A \subseteq B$.

The following results are obvious :

1. $A = B$, if and only if $A \subseteq B$ and $B \subseteq A$.
2. $A \subseteq B$, $B \subseteq C$, implies $A \subseteq C$.
3. $A \subseteq A$.
4. $\Phi \subseteq A$, for all fuzzy sets A.
5. $A \subseteq X$, for all fuzzy sets A on X.

α-cuts : α-cut or α-level cut or level cut is one of the most basic concepts in the fuzzy set theory. It connects fuzzy set theory with crisp sets.

Definition 3 : Let $\alpha \in [0, 1]$; the crisp set $\{ x \in X \mid A(x) \ge \alpha \}$ is called α-cut or α-level cut of A and is denoted by ${}^{\alpha}A$. Thus,

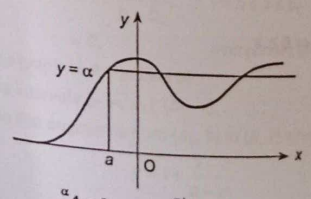
$${}^{\alpha}A = \{ x \in X \mid A(x) \ge \alpha \}$$

The crisp set $\{ x \in A \mid A(x) > \alpha \}$ is called a strict α-cut or strict α-level cut of A and is denoted by ${}^{\alpha+}A$. Thus,

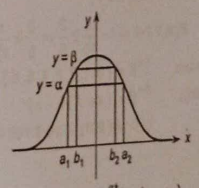
$${}^{\alpha+}A = \{ x \in X \mid A(x) > \alpha \}.$$

Clearly, ${}^{0}A = X$ and ${}^{1+}A = \Phi$, for every fuzzy set A on X.

If A is a fuzzy set on $\mathbb{R}$, then an α-cut of A may look like in the following figure.



${}^{\alpha}A = [a, \infty)$; ${}^{\alpha+}A = (a, \infty)$



${}^{\alpha}A = [a_1, a_2]$; ${}^{\alpha+}A = (a_1, a_2)$

Note that if $\alpha \le \beta$, then ${}^{\beta}A \in {}^{\alpha}A$ we leave this result for readers to prove.

Example 1 : Find ${}^{0\cdot5}A$, if $A(x) = \dfrac{x}{x+2}$ and $X = \{0, 1, 2, 3, 4\}$.

Sol. : By definition $\quad {}^{0\cdot5}A = \left\{ x \in X \,\middle|\, \dfrac{x}{x+2} \ge 0\cdot5 \right\}$

$$\frac{x}{x+2} \ge \frac{1}{2} \Leftrightarrow 2x \ge x+2 \Leftrightarrow x \ge 2$$

But $x \in \{0, 1, 2, 3, 4\}$. Hence, the only possible values for $x \ge 2$ are $x = 2, 3, 4$.

Hence, ${}^{0\cdot5}A = \{2, 3, 4\}$.

Example 2 : Find α-cuts and strong α-cuts of $A(x) = \dfrac{x}{x+1}$, $B(x) = 1 - \dfrac{x}{10}$.

$X = \{0, 1, \ldots\ldots, 10\}$, for $\alpha = 0\cdot2, 0\cdot4$. **(S.U. May 2010)**

Sol. : (i) $\quad {}^{0\cdot2}A = \{ x \in X \mid A(x) \ge 0\cdot2 \}$

$$A(x) \ge 0\cdot2 \Leftrightarrow \frac{x}{x+1} \ge \frac{1}{5} \Leftrightarrow 5x \ge x+1 \Leftrightarrow x \ge \frac{1}{4}$$

But $X = \{0, 1, 2, \ldots, 10\}$ and $x \ge \dfrac{1}{4}$ for $x = 1, 2, \ldots, 10$.

Hence, $\quad {}^{0\cdot2}A = \{1, 2, \ldots\ldots, 10\}$.

(ii) Similarly, $\quad A(x) > 0\cdot2 \Leftrightarrow x > \dfrac{1}{4}$.

Hence, $\quad {}^{0\cdot2+}A = \{1, 2, \ldots\ldots, 10\}$.

(iii) ${}^{0\cdot4}A = \{ x \in X \mid A(x) \ge 0\cdot4 \}$.

$$A(x) \ge 0\cdot4 \Leftrightarrow \frac{x}{x+1} \ge \frac{2}{5} \Leftrightarrow 5x \ge 2x+2 \Leftrightarrow x \ge \frac{2}{3}$$

But $x \ge \dfrac{2}{3}$ for all $x = 1, 2, \ldots\ldots, 10$.

Hence, $\quad {}^{0\cdot4}A = \{1, 2, \ldots\ldots, 10\}$.

(iv) Similarly, $\quad {}^{0\cdot4+}A = \{1, 2, \ldots\ldots, 10\}$.

(v) ${}^{0\cdot2}B = \{ x \in X \mid B(x) \ge 0\cdot2 \}$.

$$B(x) \ge 0\cdot2 \Leftrightarrow 1 - \frac{x}{10} \ge \frac{1}{5} \Leftrightarrow \frac{4}{5} \ge \frac{x}{10} \Leftrightarrow 8 \ge x$$

Therefore, $\quad {}^{0\cdot2}B = \{0, 1, \ldots\ldots, 8\}$.

(vi) Similarly, $\quad x \in {}^{0\cdot2+}B \Leftrightarrow 8 > x$

Hence, $\quad {}^{0\cdot2+}B = \{0, 1, \ldots\ldots, 7\}$

(vii) ${}^{0\cdot4}B = \{ x \in X \mid B(x) \ge 0\cdot4 \}$

$$B(x) \ge 0\cdot4 \Leftrightarrow 1 - \frac{x}{10} \ge \frac{2}{5} \Leftrightarrow \frac{3}{5} \ge \frac{x}{10} \Leftrightarrow 6 \ge x$$

Therefore, $\quad {}^{0\cdot4}B = \{0, 1, 2, 3, 4, 5, 6\}$.

(viii) Similarly, $\quad {}^{0\cdot4}B = \{ x \in X \mid 6 > x \}$

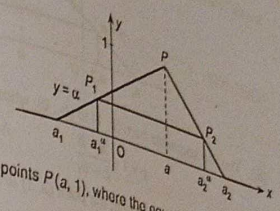
$= \{0, 1, \ldots\ldots, 5\}$.

Example 3 : Find all α-cuts of the fuzzy set A on $\mathbb{R}$, where

$$A(x) = 1 + \frac{x-a}{a-a_1}, \quad \text{for } x \in (a_1, a]$$

$$= 1 + \frac{x-a}{a-a_2}, \quad \text{for } x \in (a, a_2]$$

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



Sol. : The graph of A consists of

(i) infinite interval $(-\infty, a_1]$

(ii) interval $[a_2, \infty)$

(iii) line segment from the point $(a_1, 0)$ to the points $P(a, 1)$, where the equation is

$$y = 1 + \frac{x-a}{a-a_1} \quad \ldots\ldots\ldots\ldots (1)$$

(iv) the line segment from the point $(a, 1)$ to the point $(a_2, 0)$, where the equation is

$$y = 1 + \frac{x-a}{a-a_2} \quad \ldots\ldots\ldots\ldots (2)$$

Therefore, **(1)** ${}^{1}A = \{a\}$, ${}^{0}A = \mathbb{R}$ (as usual)

(2) To find ${}^{\alpha}A$ for $\alpha \in (0, 1)$.

If the line $y = \alpha$, cuts the line (1) at $P_1(a_1^{\alpha}, \alpha)$ and the line (2) at the point $P_2(a_2^{\alpha}, \alpha)$, then

$${}^{\alpha}A = [a_1^{\alpha}, a_2^{\alpha}].$$

To evaluate a_1^{α}, consider $\alpha = 1 + \dfrac{x-a}{a-a_1}$.

This gives $\quad x = \alpha(a - a_1) + a_1$

Therefore, $\quad a_1^{\alpha} = \alpha(a - a_1) + a_1$

Similarly, $\quad \alpha = 1 + \dfrac{x-a}{a-a_2}$ yields

$$x = \alpha(a - a_2) + a_2$$

and therefore, $\quad a_2^{\alpha} = \alpha(a - a_2) + a_2$.

Hence, $\quad {}^{\alpha}A = [\alpha(a - a_1) + a_1, \alpha(a - a_2) + a_2]$ for $\alpha \in (0, 1)$.

Example 4 : Find all the α-cuts of the fuzzy set A on $\mathbb{R}$, where

$$A(x) = 1 + \frac{x-a}{a-a_1}, \quad \text{for } a_1 \le x \le a$$

$$= 1, \quad \text{for } a \le x \le b$$

$$= 1 + \frac{x-b}{b-b_1}, \quad \text{for } b \le x \le b_1$$

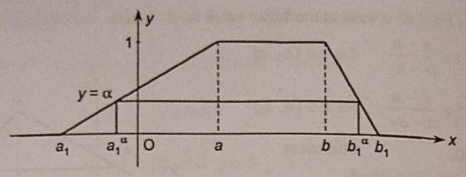
$$= 0, \quad \text{everywhere else.}$$

Sol. : The graph of A consists of

(i) The intervals $(-\infty, a_1]$, $[b_1, \infty)$,

(ii) The line segments from $(a_1, 0)$ to $(a, 1)$ whose points satisfy the equation

$$y = 1 + \frac{x-a}{a-a_1} \quad \ldots\ldots\ldots\ldots (1)$$



(iii) The line segments from $(b, 1)$ to $(b_1, 0)$, whose points satisfy the equation

$$y = 1 + \frac{x-b}{b-b_1} \qquad \text{.....(2)}$$

(iv) And the line segment from the point $(a, 1)$ to $(b, 1)$ whose points satisfy the equation $y = 1$.

$$\text{.....(3)}$$

Therefore, ${}^1A = [a, b]$, ${}^0A = \mathbb{R}$

To find ${}^\alpha A$ for $\alpha \in (0, 1)$. Consider the line $y = \alpha$. Suppose it cuts the line (1) at (a_1^α, α) and the line (2) at (b_1^α, α), then ${}^\alpha A = [\, a_1^\alpha, b_1^\alpha \,]$.

Considering $\alpha = 1 + \dfrac{x-a}{a-a_1}$ as in the previous example, we get

$$a_1^\alpha = \alpha\,(a - a_1) + a_1 \text{ and considering}$$

$$\alpha = 1 + \frac{x-b}{b-b_1}, \quad \text{we get } b_1^\alpha = \alpha\,(b - b_1) + b_1.$$

Therefore, $\quad {}^\alpha A = [\, \alpha\,(a - a_1) + a_1,\ \alpha\,(b - b_1) + b_1 \,]$.

Example 5 : Let A be the fuzzy set on $\mathbb{R}$ given by the following equations

$$A(x) = \frac{3x-1}{2}, \quad \text{for } \frac{1}{3} \le x \le 1$$

$$= \frac{5-3x}{2}, \quad \text{for } 1 \le x \le \frac{5}{3}$$

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

Draw a rough sketch of A and find ${}^{0\cdot5}A$. Indicate ${}^{0\cdot5}A$ on the graph.

Sol. : $\quad {}^{0\cdot5}A = \{\, x \in \mathbb{R} \mid A(x) \ge \frac{1}{2} \,\}$

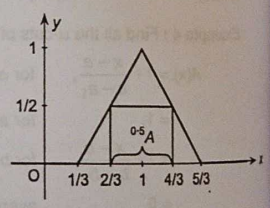
Therefore, $\quad x \in {}^{0\cdot5}A \Leftrightarrow 1 \ge A(x) \ge \dfrac{1}{2}$

$$\Leftrightarrow 1 \ge \frac{3x-1}{2} \ge \frac{1}{2} \quad \text{or} \quad 1 \ge \frac{5-3x}{2} \ge \frac{1}{2}$$

$$\Leftrightarrow 3 \ge 3x \ge 2 \quad \text{or} \quad -3 \ge -3x \ge -4$$

$$\Leftrightarrow 1 \ge x \ge \frac{2}{3} \quad \text{or} \quad 1 \le x \le \frac{4}{3}$$

$$\Leftrightarrow \frac{2}{3} \le x \le \frac{4}{3}. \qquad \text{Hence, } {}^{0\cdot5}A = \left[\frac{2}{3}, \frac{4}{3}\right].$$



Note ....

Compare this problem with Example 3.

Example 6 : Find α-cuts and strong α-cuts of fuzzy sets A and B, for $\alpha = 1, 0{\cdot}99, 0{\cdot}1$, where

(i) $A(x) = \dfrac{100}{1+10x}, \quad x \in \{\, 10, 20, 30, 40, 50 \,\}$

(ii) $B(x) = \dfrac{20x}{20x+5}, \quad x \in \{\, 10, 20, 30, 40, 50 \,\}$

(S.U. 2009)

Sol. : (i) (a) ${}^1A = \left\{ x \in X \,\middle|\, \dfrac{100}{1+10x} = 1 \right\}$

Therefore, $x \in {}^1A \Leftrightarrow 100 = 1 + 10x \Leftrightarrow 9{\cdot}9 = x$.

Clearly, for no $x \in X$, $x = 9{\cdot}9$. Hence, ${}^1A = \Phi$.

(b) ${}^{1+}A = \Phi$, always.

(c) ${}^{0\cdot99}A = \left\{ x \in X \,\middle|\, \dfrac{100}{1+10x} \ge \dfrac{99}{100} \right\}$

Therefore, $x \in {}^{0\cdot9}A \Leftrightarrow \dfrac{100}{1+10x} \ge \dfrac{99}{100} \Leftrightarrow 10{,}000 \ge 99 + 990x$

$$\Leftrightarrow \frac{9901}{990} \ge x \Leftrightarrow 10 + \frac{1}{990} \ge x.$$

The last inequality holds only when $x = 10$ in X. Hence, ${}^{0\cdot99}A = [\,10\,]$.

(d) ${}^{0\cdot99+}A = \left\{ x \in X \,\middle|\, \dfrac{100}{1+10x} \ge \dfrac{99}{100} \right\}$

Using arguments of the previous case $x \in {}^{0.99+}A$ if and only if $10 + \dfrac{1}{990} > x$. This inequality holds only for $x = 10$ in X. Hence, ${}^{0.99+}A = \{10\}$.

(e) ${}^{0\cdot1}A = \left\{ x \in X \,\middle|\, \dfrac{100}{1+10x} \ge \dfrac{1}{10} \right\}$.

Therefore, $x \in {}^{0\cdot1}A \Leftrightarrow \dfrac{100}{1+10x} \ge \dfrac{1}{10} \Leftrightarrow 99{\cdot}9 \ge x$

Clearly $99{\cdot}9 \ge x$ holds for all $x \in X$. Hence, ${}^{0\cdot1}A = X$.

(f) ${}^{0\cdot1+}A = \left\{ x \in X \,\middle|\, \dfrac{100}{1+10x} > \dfrac{1}{10} \right\}$

Therefore, $x \in {}^{0\cdot1+}A \Leftrightarrow 99.9 > x$.

Again, this condition holds for all $x \in X$; therefore, ${}^{0\cdot1+}A = X$.

(ii) (a) ${}^1B = \left\{ x \in X \,\middle|\, \dfrac{20x}{20x+5} = 1 \right\}$.

Therefore, $x \in {}^1B \Leftrightarrow \dfrac{20x}{20x+5} = 1 \Leftrightarrow 20x = 20x + 5$.

Clearly, this equation does not hold for any $x \in X$.

Hence, $\quad {}^1B = \Phi$.

(b) ${}^{1+}B = \Phi$; as always.

(c) ${}^{0\cdot99}B = \left\{ x \in X \,\middle|\, \dfrac{20x}{20x+5} \ge \dfrac{99}{100} \right\}$

Therefore, $x \in {}^{0\cdot99}B \Leftrightarrow 2000x \geq 1980x + 495$

$\Leftrightarrow 20x \geq 495.$

The last inequality holds for $x = 30, 40, 50$ in X.

Hence, ${}^{0\cdot99}B = \{30, 40, 50\}$.

(d) On the same lines we find ${}^{0\cdot99+}B = \{30, 40, 50\}$.

(e) ${}^{0\cdot1}B = \left\{x \in X \middle| \frac{20x}{20x+5} \geq \frac{1}{10}\right\}$

Therefore, $x \in {}^{0\cdot1}B \Leftrightarrow 180x \geq 5$.

The last inequality holds for all $x \in X$. Hence, ${}^{0\cdot1}B = X$.

(f) On similar lines we prove ${}^{0\cdot1+}B = X$.

Example 7 : Find all α-cuts of the fuzzy set A, where A is given by

$$A = \frac{0\cdot1}{[0,1)} + \frac{0\cdot2}{[1,2)} + \frac{0\cdot6}{[2,3)} + \frac{0\cdot7}{[3,4)} + \frac{0\cdot8}{[4,5)} + \frac{0\cdot9}{[5,6)} + \frac{1}{6} + \frac{0\cdot5}{(6,7]} + \frac{0\cdot4}{(7,8]} + \frac{0\cdot2}{(8,9]} + \frac{0\cdot1}{(9,10]}$$

Sol. : If we assume $X = [0, 10]$, then ${}^{0}A = X$. If we assume the convention that elements of grade 0 are not mentioned here, ${}^{0}A = \mathbb{R}$, the set of all real numbers.

For convenience of computation, we consider the decreasing order of α^s which produce distinct α-cuts; namely $1 > 0\cdot9 > 0\cdot8 > 0\cdot7 > 0\cdot6 > 0\cdot5 > 0\cdot4 > 0\cdot2 > 0\cdot1$.

Recall that if $\beta > \alpha$, then ${}^{\beta}A \subseteq {}^{\alpha}A$.

Therefore, ${}^{1}A \subseteq {}^{0\cdot9}A \subseteq {}^{0\cdot8}A \subseteq {}^{0\cdot7}A \subseteq {}^{0\cdot6}A \subseteq {}^{0\cdot5}A \subseteq {}^{0\cdot4}A \subseteq {}^{0\cdot2}A \subseteq {}^{0\cdot1}A$.

Therefore, "adding new elements" to ${}^{\beta}A$, we get ${}^{\alpha}A$.

Clearly, ${}^{1}A = \{6\}$

Therefore, ${}^{0\cdot9}A = [5, 6) \cup \{6\} = [5, 6]$

${}^{0\cdot8}A = [4, 5) \cup {}^{0\cdot9}A = [4, 6]$

${}^{0\cdot7}A = [3, 4) \cup {}^{0\cdot8}A = [3, 6]$

${}^{0\cdot6}A = [2, 3) \cup {}^{0\cdot7}A = [2, 6]$

${}^{0\cdot5}A = (6, 7] \cup {}^{0\cdot6}A = [2, 7]$

${}^{0\cdot4}A = (7, 8] \cup {}^{0\cdot5}A = [2, 8]$

${}^{0\cdot2}A = [1, 2) \cup {}^{0\cdot4}A \cup (8, 9] = [1, 9]$

${}^{0\cdot1}A = [0, 1) \cup {}^{0\cdot2}A \cup (9, 10] = [0, 10]$

EXERCISE - II

1. Let $X = \{a, b, c, d\}$, $A = \left\{\frac{0\cdot8}{a}, \frac{0\cdot3}{b}, \frac{0\cdot1}{c}, \frac{0\cdot01}{d}\right\}$. Find ${}^{0\cdot1}A$ and ${}^{0\cdot1+}A$.

[**Ans. :** ${}^{0\cdot1}A = \{a, b, c\}$; ${}^{0\cdot1+}A = \{a, b\}$]

2. Let $X = \{a, b, c, d, e, f\}$; $A = \frac{0}{a} + \frac{0}{b} + \frac{0\cdot5}{c} + \frac{1}{d} + \frac{0\cdot7}{e} + \frac{0\cdot2}{f}$.

Find α-cuts and strong α-cuts for $\alpha = 0\cdot5, 0\cdot7$.

[**Ans. :** ${}^{0\cdot5}A = \{c, d, e\}$, ${}^{0\cdot5+}A = \{d, e\}$, ${}^{07}A = \{d, e\}$, ${}^{07+}A = \{d\}$]

3. Let A be a fuzzy set on the set of real number $\mathbb{R}$.

$$A(x) = \begin{cases} \frac{x - 1\cdot5}{0\cdot5}, & \text{for } 1\cdot5 \leq x \leq 2 \\ \frac{2\cdot5 - x}{0\cdot5}, & \text{for } 2 \leq x \leq 2\cdot5 \\ 0, & \text{otherwise} \end{cases}$$

Find ${}^{0\cdot5}A$.

[**Ans. :** $[1\cdot75, 2\cdot25]$]

4. (a)

$$A(x) = \begin{cases} 1, & \text{if } x \leq 20 \\ \frac{35 - x}{15}, & \text{if } 20 < x < 35 \\ 0, & \text{otherwise} \end{cases}$$

(b)

$$A(x) = \begin{cases} \frac{x - 20}{15}, & \text{for } 20 < x < 35 \\ \frac{60 - x}{15}, & \text{for } 45 < x < 60 \\ 1, & \text{for } 35 \leq x \leq 45 \\ 0, & \text{otherwise.} \end{cases}$$

(c)

$$A(x) = \begin{cases} \frac{x - 45}{15}, & \text{for } 45 < x < 60 \\ 1, & \text{for } x \geq 60 \\ 0, & \text{otherwise} \end{cases}$$

Find α-cuts and strict α-cuts of A, in each case for $\alpha = 0\cdot1$.

[**Ans. :** **(a)** $(-\infty, 33\cdot5]$, $(-\infty, 33\cdot5)$

(b) $[21\cdot5, 58\cdot5]$, $(21\cdot5, 58\cdot5)$

(c) $[46\cdot5, \infty)$, $(46\cdot5, \infty)$]

5. Determine all α-cuts and strong α-cuts of the following fuzzy set

$$A = \left\{\frac{0\cdot5}{x_1} + \frac{0\cdot4}{x_2} + \frac{0\cdot7}{x_3} + \frac{0\cdot8}{x_4} + \frac{1}{x_5}\right\}$$

(S.U. 2011)

[**Ans. :** ${}^{0}A = X$, ${}^{1+}A = \Phi$, ${}^{0+}A = X$, ${}^{1}A = \{x_5\}$, ${}^{0\cdot8}A = \{x_4, x_5\}$,

${}^{0\cdot8+}A = {}^{1}A$, ${}^{0\cdot7}A = \{x_3, x_4, x_5\}$, ${}^{0\cdot7+}A = {}^{0\cdot8}A$,

${}^{0\cdot5}A = \{x_1, x_3, x_4, x_5\}$, ${}^{0\cdot5+}A = {}^{0\cdot7}A$,

${}^{0\cdot4}A = X$, ${}^{0\cdot4+}A = {}^{0\cdot5}A$, if $X = \{x_1, x_2, \ldots, x_5\}$ is the universe.]

6. Determine all the α-cuts of the fuzzy set

$$A = \frac{0\cdot2}{[0,1)} + \frac{0\cdot6}{[1,2)} + \frac{0\cdot8}{[2,3)} + \frac{0\cdot9}{[3,4)} + \frac{1}{4} + \frac{0\cdot5}{(4,5]} + \frac{0\cdot1}{(5,6]}$$

[**Ans. :** ${}^{1}A = \{4\}$, ${}^{0\cdot9}A = [3, 4]$, ${}^{0\cdot8}A = [2, 4]$, ${}^{0\cdot6}A = [1, 4]$,

${}^{0\cdot5}A = [1, 5]$, ${}^{0\cdot2}A = [0, 5]$, ${}^{0\cdot1}A = [0, 6]$]

7. Algebraic Operations on Fuzzy Sets

Basic algebraic operations like union, intersection and compliment, defined on the power set $p(X)$ of a universal set X, can be easily extended to the fuzzy power set $\mathcal{F}(X)$. This can be done in several ways. We discuss, here, only the most commonly used operations called the standard operations.

Definition

Let A and B be fuzzy sets on a same universal set X.

(a) The fuzzy set A' on X, defined as $A' : X \to [0, 1]$

$$A'(x) = 1 - A(x)$$

is called the (standard) complement of A.

(b) The fuzzy set $A \cup B$ on X defined as $A \cup B : X \to [0, 1]$

$$A \cup B(x) = \max \{ A(x), B(x) \}$$

is called the (standard) union of A and B.

(c) The fuzzy set $A \cap B$ on X, defined as $A \cap B : X \to [0, 1]$

$$A \cap B(x) = \min \{ A(x), B(x) \}$$

is called the (standard) intersection of A and B.

Let A, B, C be fuzzy sets on X. Then the following results always hold :

1. $A \cup A = A,\ A \cap A = A$ [Idempotency]
2. $A \cup B = B \cup A,\ A \cap B = B \cap A$ [Commutativity]
3. $(A \cup B) \cup C = A \cup (B \cup C)$ and
   $(A \cap B) \cap C = A \cap (B \cap C)$ [Associativity]
4. $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
   $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$ [Distributivity]
5. $A \cap X = A,\quad A \cup X = X$
6. $A \cap \Phi = \Phi,\quad A \cup \Phi = A$
7. $(A \cup B)' = A' \cap B'$
   $(A \cap B)' = A' \cup B'$ [DeMorgan's Laws]
8. $(A')' = A$
9. If $A \subseteq B$, then $A \cap B = A$, $A \cup B = B$ and conversely.
10. $A \cap B \subseteq A,\quad A \subseteq A \cup B$
11. $A \cup (A \cap B) = A$ and $A \cap (A \cup B) = A$. [Absorption Laws]

The only laws which hold in $p(X)$ and not in $\mathcal{F}(X)$ are of complementation.

It is, in general, **not true** that $A \cap A' = \Phi$ and $A \cup A' = X$.

Consider the simplest example,

$$A : X \to [0, 1]$$

$$A(x) = 0{\cdot}5$$

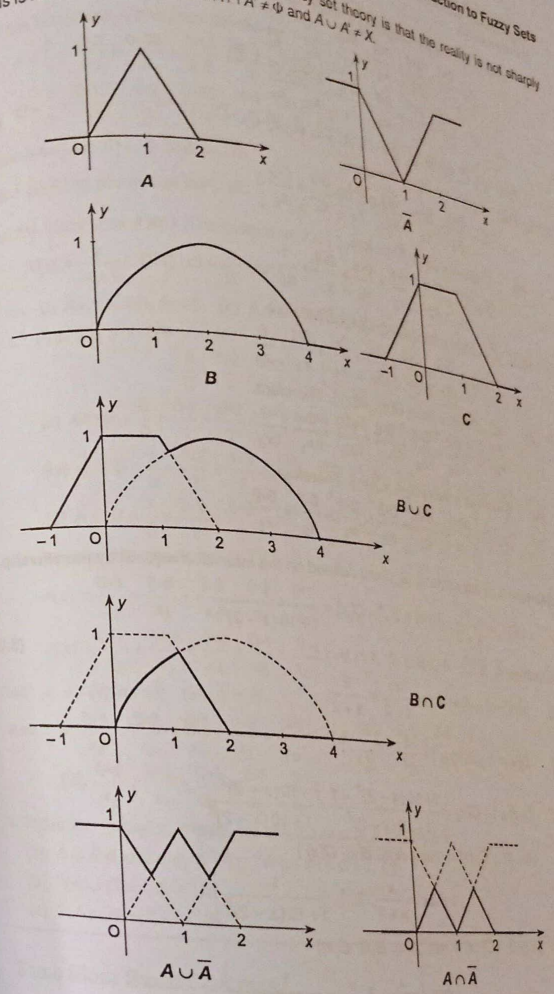
Then, $A'(x) = 1 - A(x) = 0{\cdot}5$

Thus, $A \cap A'(x) = \min \{ A(x), A'(x) \}$

$$= 0{\cdot}5 \neq 0$$

Thus, $A \cap A' \neq \Phi$. Similarly, $A \cup A' \neq X$.

This is not surprising. The basic tenet of fuzzy set theory is that the reality is not sharply defined; this is reflected in the fact that $A \cap A' \neq \Phi$ and $A \cup A' \neq X$.



Illustrations of Standard Operations on Fuzzy Sets

Example 1 : If the fuzzy sets C and D are defined by the following membership function

$$C = \frac{0{\cdot}1}{x_1} + \frac{0{\cdot}6}{x_2} + \frac{0{\cdot}8}{x_3} + \frac{0{\cdot}9}{x_4} + \frac{0{\cdot}7}{x_5} + \frac{0{\cdot}1}{x_6}$$

$$D = \frac{0{\cdot}9}{x_1} + \frac{0{\cdot}7}{x_2} + \frac{0{\cdot}5}{x_3} + \frac{0{\cdot}2}{x_4} + \frac{0{\cdot}1}{x_5} + \frac{0}{x_6}$$

Find : (1) $\overline{C}$, (2) $\overline{D}$, (3) $\overline{C} \cap D$, (4) $C \cup \overline{D}$, (5) $C \cup \overline{C}$. **(S.U. 2008)**

Sol. : (1) $\overline{C}(x) = 1 - C(x)$. Therefore,

$$\overline{C} = \frac{0{\cdot}9}{x_1} + \frac{0{\cdot}4}{x_2} + \frac{0{\cdot}2}{x_3} + \frac{0{\cdot}1}{x_4} + \frac{0{\cdot}3}{x_5} + \frac{0{\cdot}9}{x_6}$$

(2) $\overline{D}(x) = 1 - D(x)$. Therefore,

$$\overline{D} = \frac{0{\cdot}1}{x_1} + \frac{0{\cdot}3}{x_2} + \frac{0{\cdot}5}{x_3} + \frac{0{\cdot}8}{x_4} + \frac{0{\cdot}9}{x_5} + \frac{1}{x_6}$$

(3) $\overline{C} \cap D(x) = \min\{\overline{C}(x), D(x)\}$. Therefore,

$$\overline{C} \cap D = \frac{0{\cdot}9}{x_1} + \frac{0{\cdot}4}{x_2} + \frac{0{\cdot}2}{x_3} + \frac{0{\cdot}1}{x_4} + \frac{0{\cdot}1}{x_5} + \frac{0}{x_6}$$

(4) $(C \cup \overline{D})(x) = \max\{C(x), \overline{D}(x)\}$. Therefore,

$$C \cup \overline{D} = \frac{0{\cdot}1}{x_1} + \frac{0{\cdot}6}{x_2} + \frac{0{\cdot}8}{x_3} + \frac{0{\cdot}9}{x_4} + \frac{0{\cdot}9}{x_5} + \frac{1}{x_6}$$

(5) $C \cup \overline{C}(x) = \max\{C(x), \overline{C}(x)\}$. Therefore,

$$C \cup \overline{C} = \frac{0{\cdot}9}{x_1} + \frac{0{\cdot}6}{x_2} + \frac{0{\cdot}8}{x_3} + \frac{0{\cdot}9}{x_4} + \frac{0{\cdot}7}{x_5} + \frac{0{\cdot}9}{x_6}$$

Example 2 : Fuzzy sets A, B, C are defined on the interval $X = [0, 5]$ by membership function

$$A(x) = \frac{x}{x+2}, \quad B(x) = 2^{-x}, \quad C(x) = \frac{1}{1+10(x-2)^2}$$

Determine $\overline{A}$, $\overline{B}$, $\overline{C}$, $A \cup B \cup C$, $A \cap B \cap C$. **(S.U. 2011)**

Sol. : (i) $\overline{A}(x) = 1 - A(x) = 1 - \dfrac{x}{x+2} = \dfrac{2}{x+2}$

(ii) $\overline{B}(x) = 1 - B(x) = 1 - \dfrac{1}{2^x} = \dfrac{2^x - 1}{2^x}$

(iii) $\overline{C}(x) = 1 - C(x) = \dfrac{1+10(x-2)^2 - 1}{1+10(x-2)^2} = \dfrac{10(x-2)^2}{1+10(x-2)^2}$

(iv) $(A \cup B \cup C)(x) = \max\{A(x), B(x), C(x)\}$

$$= \max\left\{\frac{x}{x+2}, 2^{-x}, \frac{1}{1+10(x-2)^2}\right\}$$

(v) $(A \cap B \cap C)(x) = \min\{A(x), B(x), C(x)\}$

$$= \min\left\{\frac{x}{x+2}, 2^{-x}, \frac{1}{1+10(x-2)^2}\right\}$$

EXERCISE - III

1. Let the fuzzy sets A and B be defined as

$$A = \frac{0}{1} + \frac{0{\cdot}2}{1{\cdot}5} + \frac{0{\cdot}35}{2} + \frac{0{\cdot}15}{2{\cdot}5} + \frac{0{\cdot}5}{3} + \frac{0{\cdot}25}{3{\cdot}5} + \frac{0{\cdot}4}{4}$$

$$B = \frac{1}{1} + \frac{0{\cdot}15}{1{\cdot}5} + \frac{0{\cdot}2}{2} + \frac{0{\cdot}35}{2{\cdot}5} + \frac{0{\cdot}4}{3} + \frac{0{\cdot}15}{3{\cdot}5} + \frac{0}{4}$$

Calculate $(A \cap \overline{B}) \cup A$ and $\overline{A \cup B}$.

[**Ans. :** (i) A (by absorption law), (ii) $\dfrac{0}{1} + \dfrac{0{\cdot}8}{1{\cdot}5} + \dfrac{0{\cdot}65}{2} + \dfrac{0{\cdot}65}{2{\cdot}5} + \dfrac{0{\cdot}5}{3} + \dfrac{0{\cdot}75}{3{\cdot}5} + \dfrac{0{\cdot}6}{4}$] **(S.U. 2012)**

2. Let the fuzzy sets A and B be defined on $X = \{0, 1, 2, \ldots, 10\}$ as follows :

$$A(x) = \frac{x}{x+1}, \quad B(x) = 1 - \frac{x}{10} \quad \text{for } x \in X$$

Find - (i) $\overline{A} \cup \overline{B}$, (ii) $A \cap \overline{B}$, (iii) $\overline{A \cap B}$, (iv) $\overline{A} \cap B$, (v) $\overline{A} \cup B$.

[**Ans. : (i) & (iii)** $\overline{A} \cup \overline{B} = \overline{A \cap B}$ **(S.U. 2010)**

$$= \frac{1}{0} + \frac{1/2}{1} + \frac{1/3}{2} + \frac{0{\cdot}3}{3} + \frac{0{\cdot}4}{4} + \frac{0{\cdot}5}{5} + \frac{0{\cdot}6}{6} + \frac{0{\cdot}7}{7} + \frac{0{\cdot}8}{8} + \frac{0{\cdot}9}{9} + \frac{1}{10}$$

(ii) $A \cap \overline{B} = \dfrac{0}{0} + \dfrac{0{\cdot}1}{1} + \dfrac{0{\cdot}2}{2} + \dfrac{0{\cdot}3}{3} + \dfrac{0{\cdot}4}{4} + \dfrac{0{\cdot}5}{5} + \dfrac{0{\cdot}6}{6} + \dfrac{0{\cdot}7}{7} + \dfrac{0{\cdot}8}{8} + \dfrac{0{\cdot}9}{9} + \dfrac{10/11}{10}$

(iv) $\overline{A} \cap B = \dfrac{1}{0} + \dfrac{1/2}{1} + \dfrac{1/3}{2} + \dfrac{1/4}{3} + \dfrac{1/5}{4} + \dfrac{1/6}{5} + \dfrac{1/7}{6} + \dfrac{1/8}{7} + \dfrac{1/9}{8} + \dfrac{0{\cdot}1}{9} + \dfrac{0}{10}$

(v) $\overline{A} \cup B = \dfrac{1}{0} + \dfrac{0{\cdot}9}{1} + \dfrac{0{\cdot}8}{2} + \dfrac{0{\cdot}7}{3} + \dfrac{0{\cdot}6}{4} + \dfrac{0{\cdot}5}{5} + \dfrac{0{\cdot}4}{6} + \dfrac{0{\cdot}3}{7} + \dfrac{0{\cdot}2}{8} + \dfrac{0{\cdot}1}{9} + \dfrac{1/11}{10}$]

3. Given two fuzzy sets A and B

$$A(x) = \frac{0{\cdot}6}{x_2} + \frac{0{\cdot}8}{x_3} + \frac{0{\cdot}9}{x_4} + \frac{0{\cdot}7}{x_5} + \frac{0{\cdot}1}{x_6}$$

$$B(x) = \frac{0{\cdot}7}{x_2} + \frac{0{\cdot}5}{x_3} + \frac{0{\cdot}2}{x_4} + \frac{0{\cdot}1}{x_5} + \frac{0}{x_6}$$

Find : (i) $\overline{A}$, (ii) $A \cap \overline{B}$, (iii) $\overline{A} \cap B$.

[**Ans. :** (i) $\dfrac{0{\cdot}4}{x_2} + \dfrac{0{\cdot}2}{x_3} + \dfrac{0{\cdot}1}{x_4} + \dfrac{0{\cdot}3}{x_5} + \dfrac{0{\cdot}9}{x_6}$; (ii) $\dfrac{0{\cdot}3}{x_2} + \dfrac{0{\cdot}5}{x_3} + \dfrac{0{\cdot}8}{x_4} + \dfrac{0{\cdot}7}{x_5} + \dfrac{0{\cdot}1}{x_6}$;

(iii) $\dfrac{0{\cdot}4}{x_2} + \dfrac{0{\cdot}2}{x_3} + \dfrac{0{\cdot}1}{x_4} + \dfrac{0{\cdot}1}{x_5} + \dfrac{0}{x_6}$]

4. Prove the following results for any fuzzy sets A and B on X.

(a) $A \subseteq B$ if and only if ${}^{\alpha}A \subseteq {}^{\alpha}B$ for all $\alpha \in [0, 1]$.

(b) ${}^{\alpha}(A \cup B) = {}^{\alpha}A \cup {}^{\alpha}B$

(c) ${}^{\alpha}(A \cap B) = {}^{\alpha}A \cap {}^{\alpha}B$

8. Some More Basic Concepts

In this section we introduce some more basic terms which will be used in our further development of fuzzy set theory. As before X is the universal set and A is a fuzzy set on X.

(a) Support of a Fuzzy Set

The crisp set $\{ x \in X \mid A(x) > 0 \}$ is called the support of the fuzzy set A and is denoted by supp A.

Thus, $\quad \text{supp } A = \{ x \in X \mid A(x) > 0 \}$

Clearly, $\quad \text{supp } A = {}^{0+}A$.

(b) Height, Normality and Subnormality

The suppremum or the least upper bound (lub) of the crisp set $\{ A(x) \mid x \in X \}$ is called the height of the fuzzy set A and is denoted by $h(A)$.

Thus, $\quad h(A) = \sup \{ A(x) \mid x \in X \}$

A is called **normal**, if $h(A) = 1$.

A is called **subnormal**, if $h(A) < 1$.

Illustrative Examples

(i) Let A and B be fuzzy sets defined on $X = \{ 0, 1, 2, \ldots, 10 \}$ such that $A(x) = \dfrac{x}{x+1}$ and $B(x) = 1 - \dfrac{x}{10}$, then $h(A) = \dfrac{10}{11}$ and $h(B) = 1$. Hence, A is subnormal and B is normal.

(ii) Let C be the fuzzy set defined on $\mathbb{N} = \{ 1, 2, 3, \ldots \}$ such that $C(n) = \dfrac{n-1}{n}$.

Then $\sup \{ C(n) \mid n \in \mathbb{N} \} = \sup \left\{ \dfrac{n-1}{n} \,\middle|\, n \in \mathbb{N} \right\} = 1$

Hence, $h(C) = 1$, and C is normal. Note that there is no $n \in \mathbb{N}$ such that $C(n) = 1$. This happen only when X is infinite.

If X is finite, then $h(A) = \max \{ A(x) \mid x \in X \}$ and therefore, there exists $x_0 \in X$, such that

$$h(A) = A(x_0).$$

It is obvious that — (i) $h(\Phi) = 0$, and (ii) $h(A) = 1$, if A is non-empty crisp.

(c) Core of a Fuzzy Set

The crisp set $\{ x \in X \mid A(x) = 1 \}$ is called the core of A and is denoted by Core (A).

Clearly, $\quad \text{Core } (A) = {}^{1}A$.

In case of fuzzy sets A, B, C, defined in the above section (b), we have

$$\text{Core } A = \Phi, \text{ Core } B = \{ 10 \} \text{ and Core } C = \Phi.$$

(d) Scalar Cardinality

Let A be a fuzzy set on X, with finite support Supp $A = \{ x_1, x_2, \ldots, x_n \}$. Then the sum $\sum_{i=1}^{n} A(x_i)$ is called the **scalar cardinality** of A and is denoted by $|A|$. Thus, $|A| = \sum_{i=1}^{n} A(x_i)$.

The scalar cardinality is also called sigma count.

Note that if A is crisp, then $|A|$ = number of elements in A. In particular $|\Phi| = 0$.

Example 1 : Let $X = \{ a, b, c, d, e, f, g, h, i, j \}$ and

$$A = \frac{0}{a} + \frac{0{\cdot}2}{b} + \frac{0{\cdot}5}{c} + \frac{0{\cdot}2}{d} + \frac{0}{e} + \frac{1}{f} + \frac{0}{g} + \frac{0{\cdot}5}{h} + \frac{1}{i} + \frac{0{\cdot}1}{j}, \text{ then}$$

$$|A| = 0{\cdot}2 + 0{\cdot}5 + 0{\cdot}2 + 1 + 0{\cdot}5 + 1 + 0{\cdot}1 = 3{\cdot}5.$$

Example 2 : Find the scalar cardinality of the fuzzy set

$$A(x) = \frac{0{\cdot}5}{x} + \frac{0{\cdot}4}{y} + \frac{1}{z}$$

Sol. : $|A| = 0{\cdot}5 + 0{\cdot}4 + 1 = 1{\cdot}9$.

Example 3 : Find $|A|$ and $|B|$, for the fuzzy sets (S.U. 2009)

$$A(x) = \frac{x}{x+2} \text{ and } B(x) = \frac{x}{x+5}, \text{ on } X = \{ 0, 1, 2, \ldots, 10 \}.$$

Sol. : (i) $|A| = \sum_{x \in X} A(x) = \sum_{n=1}^{10} \dfrac{n}{n+2}$ (S.U. 2009, 10)

(ii) $|B| = \sum_{1}^{10} \dfrac{n}{n+5}$.

Example 4 : For fuzzy sets A and B defined on a finite universe X, prove that

$$|A| + |B| = |A \cap B| + |A \cup B|.$$

Sol. : Let $X = \{ x_1, x_2, \ldots, x_n \}$. Then

(i) $|A| = \sum_{1}^{n} A(x_i)$; $\quad |B| = \sum_{1}^{n} B(x_i)$

(ii) $|A \cap B| = \sum_{1}^{n} (A \cap B)(x_i) = \sum_{1}^{n} \min \{ A(x_i), B(x_i) \}$

(iii) $|A \cup B| = \sum_{1}^{n} (A \cup B)(x_i) = \sum_{1}^{n} \max \{ A(x_i), B(x_i) \}$

Without the loss of generality, suppose

$A(x_i) \le B(x_i)$ for $i = 1, 2, \ldots, k$.

and $A(x_i) > B(x_i)$ for $i = k+1, k+2, \ldots, n$.

Then $\min \{ A(x_i), B(x_i) \} = A(x_i)$, for $i = 1, 2, \ldots, k$

and $= B(x_i)$, for $i = k+1, \ldots, n$

and $\max \{ A(x_i), B(x_i) \} = B(x_i)$, for $i = 1, 2, \ldots, k$

and $= A(x_i)$, for $i = k+1, \ldots, n$

$$\therefore \; |A \cap B| + |A \cup B| = \sum_{1}^{k} A(x_i) + \sum_{k+1}^{n} B(x_i) + \sum_{1}^{k} B(x_i) + \sum_{k+1}^{n} A(x_i)$$

$$= \sum_{1}^{n} A(x_i) + \sum_{1}^{n} B(x_i)$$

$$= |A| + |B|$$

Note ....

We will use this result as a standard formula.

(e) Level Set

Let A be a non-empty fuzzy set. Then the set $\{ A(x) \mid A(x) \neq 0, x \in X \}$ is called the level set of A and is denoted by $L(A)$.

Thus, $\quad L(A) = \{ A(x) \mid A(x) \neq 0, x \in X \}$

In Example 1, above in (d)

$$L(A) = \{0{\cdot}1, 0{\cdot}2, 0{\cdot}5, 1\}$$

In Example 3, above in (d), we have

$$L(A) = \left\{ \frac{n}{n+2} \,\middle|\, 1, 2, \ldots\ldots, 10 \right\}$$

For the fuzzy set B on $\mathbb{R}$, defined by

$$B(x) = \begin{cases} 1 - e^{-x}, & \text{if } x \geq 0 \\ 0, & \text{otherwise} \end{cases}$$

$$L(B) = (0, 1)$$

(t) **Fuzzy Cardinality**

Definition : Let A be a non-empty fuzzy set on a finite set X. Let $\mathbb{N}$ be the set of natural numbers. The fuzzy set $|\tilde{A}|$ defined below is called the fuzzy cardinality of A.

$$|\tilde{A}| : \mathbb{N} \to [0, 1]$$

$$|\tilde{A}|(|{}^{\alpha}A|) = \begin{cases} \alpha, & \text{for } \alpha \in L(A) \\ 0, & \text{otherwise} \end{cases}$$

In summation notation of a fuzzy set, we have

$$|\tilde{A}| = \sum_{\alpha \in L(A)} \frac{\alpha}{|{}^{\alpha}A|}$$

Note ....

$|{}^{\alpha}A|$ is the number of elements in the set ${}^{\alpha}A$, which are necessarily finite as X is finite.

Example 1 : Let $X = \{2n \mid n = 0, 1, 2, \ldots, 40\}$ and

$$A : X \to [0, 1]$$

$$A(x) = 0, \quad \text{for } x \leq 20 \text{ or } x \geq 60$$

$$= \frac{x - 20}{15}, \quad \text{for } 20 < x \leq 35$$

$$= 1, \quad \text{for } 35 \leq x \leq 45$$

$$= \frac{60 - x}{15}, \quad \text{for } 45 < x < 60$$

Find the fuzzy cardinality of the fuzzy set A.

Note ....

To compute $|\tilde{A}|$, it is necessary to count the number of elements in ${}^{\alpha}A$ for each α in $L(A)$. Therefore, it is convenient if we order the elements of $L(A)$, either in increasing order or decreasing order. We arrange them in decreasing order, so that the set of ${}^{\alpha}A$ will be in increasing order. We start with ${}^{1}A$ and go on "adding new elements (if any) in the each succeeding ${}^{\alpha}A$ and find the table for $|{}^{\alpha}A|$.

Sol. : The level set is clearly

$$L(A) = \left\{ \frac{2}{15}, \frac{4}{15}, \frac{6}{15}, \frac{8}{15}, \frac{10}{15}, \frac{12}{15}, \frac{14}{15}, 1 \right\}$$

We rearrange the elements of $L(A)$ in decreasing order and form the following table.

α-level	Elements of degree α	Total member of elements	$\lvert{}^{\alpha}A\rvert$
1	36, 38, 40, 42, 44	5	
14 / 15	34, 46	2	5
12 / 15	32, 48	2	7
10 / 15	30, 50	2	9
8 / 15	28, 52	2	11
6 / 15	26, 54	2	13
4 / 15	24, 56	2	15
2 / 15	22, 58	2	17
		2	19

Since $|\tilde{A}| = \sum \frac{\alpha}{|{}^{\alpha}A|}$, $\alpha \in L(A)$, we get from columns 1 and 4, that

$$|\tilde{A}| = \frac{1}{5} + \frac{14/15}{7} + \frac{12/15}{9} + \frac{10/15}{11} + \frac{8/15}{13} + \frac{6/15}{15} + \frac{4/15}{17} + \frac{2/15}{19}$$

Example 2 : Find the fuzzy cardinality of the fuzzy set given by

$$A(x) = \frac{35 - x}{15} \text{ on } X = \{20, 22, 24, 26, 28, 30, 32, 34\}$$

(S.U. 2011)

Sol. : Level set $L(A) = \left\{ 1, \frac{13}{15}, \frac{11}{15}, \frac{9}{15}, \frac{7}{15}, \frac{5}{15}, \frac{3}{15}, \frac{1}{15} \right\}$

The above arrangement of elements of $L(A)$ is in decreasing order. We, therefore, form the following table :

α-level	Elements of degree α	Total member of elements	$\lvert{}^{\alpha}A\rvert$
1	20	1	1
13 / 15	22	1	2
11 / 15	24	1	3
9 / 15	26	1	4
7 / 15	28	1	5
5 / 15	30	1	6
3 / 15	32	1	7
1 / 15	34	1	8

Since $|\tilde{A}| = \sum \frac{\alpha}{|{}^{\alpha}A|}$, $\alpha \in L(A)$, using columns 1 and 4, we compute

$$|\tilde{A}| = \frac{1}{1} + \frac{13/15}{2} + \frac{11/15}{3} + \frac{9/15}{4} + \frac{7/15}{5} + \frac{5/15}{6} + \frac{3/15}{7} + \frac{1/15}{8}$$

Example 3 : Let A and B be fuzzy set

$$A = \frac{0}{1} + \frac{0{\cdot}2}{1{\cdot}5} + \frac{0{\cdot}35}{2} + \frac{0{\cdot}15}{2{\cdot}5} + \frac{0{\cdot}5}{3} + \frac{0{\cdot}25}{3{\cdot}5} + \frac{0{\cdot}4}{4}$$

$$B = \frac{1}{1} + \frac{0{\cdot}15}{1{\cdot}5} + \frac{0{\cdot}2}{2} + \frac{0{\cdot}35}{2{\cdot}5} + \frac{0{\cdot}4}{3} + \frac{0{\cdot}15}{3{\cdot}5} + \frac{0}{4}$$

Find the fuzzy cardinality of $\overline{A \cup B}$. (S.U. 2012)

Sol. : $A \cup B(x) = \max\{A(x), B(x)\}$, therefore,

$$A \cup B = \frac{1}{1} + \frac{0{\cdot}2}{1{\cdot}5} + \frac{0{\cdot}35}{2} + \frac{0{\cdot}35}{2{\cdot}5} + \frac{0{\cdot}5}{3} + \frac{0{\cdot}25}{3{\cdot}5} + \frac{0{\cdot}4}{4}$$

$\overline{A \cup B}(x) = 1 - (A \cup B)(x)$. Therefore,

$$\overline{A \cup B} = \frac{0}{1} + \frac{0{\cdot}8}{1{\cdot}5} + \frac{0{\cdot}65}{2} + \frac{0{\cdot}65}{2{\cdot}5} + \frac{0{\cdot}5}{3} + \frac{0{\cdot}75}{3{\cdot}5} + \frac{0{\cdot}6}{4}$$

$$L(\overline{A \cup B}) = \{0{\cdot}8, 0{\cdot}75, 0{\cdot}65, 0{\cdot}6, 0{\cdot}5\}$$

We therefore form the following table.

α-level	Elements of degree α	Total member of elements	$\lvert {}^{\alpha}A \rvert$
0·8	1·5	1	1
0·75	3·5	1	2
0·65	2, 2·5	2	4
0·6	4	1	5
0·5	3	1	6

From columns 1st and 4th,

$$|\overline{A \cup B}| = \frac{0{\cdot}8}{1} + \frac{0{\cdot}75}{2} + \frac{0{\cdot}65}{4} + \frac{0{\cdot}6}{5} + \frac{0{\cdot}5}{6}$$

(h) Degree of Subsethood

Definition : Let X be a finite set and A and B be fuzzy subsets on X. Then the number $\frac{1}{|A|}(|A| - \sum_{x \in X} \max[0, A(x) - B(x)])$ is called the degree of subsethood of A in B and is denoted by $S(A, B)$. Thus,

$$S(A, B) = \frac{1}{|A|}(|A| - \sum_{x \in X} \max[0, A(x) - B(x)])$$

Note ....

The number $\sum_{x \in X} \max[0, A(x) - B(x)]$ gives the sum of degrees for which the inequality $A(x) \le B(x)$ is violated. Therefore, the difference $|A| - \sum \max[0, A(x) - B(x)]$ describes, lack of this violation. This number is divided by $|A|$. So that we have $0 \le S(A, B) \le 1$.

Example 1 : Prove that for any fuzzy set A and B on a finite set X.

$$S(A, B) = \frac{|A \cap B|}{|A|}.$$

Sol. : Let $X = \{x_1, x_2, ..., x_n\}$. Without the loss of generality, let

$A(x_i) \le B(x_i)$, for $i = 1, 2, ..., k$

and $A(x_i) > B(x_i)$, for $i = k+1, k+2, ..., n$

Since $(A \cup B)(x_i) = \max\{A(x_i), B(x_i)\}$, we have

$(A \cup B)(x_i) = B(x_i)$, for $i = 1, 2, ..., k$

and $(A \cup B)(x_i) = A(x_i)$, for $i = k+1, k+2, ..., n$

$$\therefore \quad |A \cup B| = \sum_1^k B(x_i) + \sum_{k+1}^n A(x_i)$$

Using the formula $|A| + |B| = |A \cap B| + |A \cup B|$, the definitions of $|A|$, $|B|$ and the result (1), we get ..... (1)

$$|A \cap B| = |A| + |B| - |A \cup B|$$

$$= \sum_1^n A(x_i) + \sum_1^n B(x_i) - \sum_1^k B(x_i) - \sum_{k+1}^n A(x_i)$$

$$= \sum_1^k A(x_i) + \sum_{k+1}^n B(x_i)$$

On the other hand, ... (2)

$\max\{0, A(x_i) - B(x_i)\} = 0$, for $i = 1, 2, ..., k$ .... [Since $A(x_i) - B(x_i) \le 0$]

and $\max\{0, A(x_i) - B(x_i)\} = A(x_i) - B(x_i)$, for $i = k+1, k+2, ..., n$ .... [Since $A(x_i) - B(x_i) > 0$]

Hence, $\sum_{i=1}^n \max\{0, A(x_i) - B(x_i)\} = \sum_{k+1}^n A(x_i) - \sum_{k+1}^n B(x_i)$ ..... (3)

But $$S(A, B) = \frac{1}{|A|}\left(|A| - \sum_1^n \max\{0, A(x_i) - B(x_i)\}\right)$$

$$\therefore \quad |A|\, S(A, B) = \sum_1^n A(x_i) - \sum_{k+1}^n A(x_i) + \sum_{k+1}^n B(x_i) \quad \text{[By (3)]}$$

$$= \sum_1^k A(x_i) + \sum_{k+1}^n B(x_i)$$

$$= |A \cap B| \quad \text{[By (2)]}$$

$$\therefore \quad S(A, B) = \frac{|A \cap B|}{|A|}$$

Note ....

You may use this result as a standard formula.

Example 2 : Find $S(A, B)$ and $S(B, A)$, for A and B defined by

$$A(x) = \frac{x}{x+2} \text{ and } B(x) = \frac{x}{x+5} \text{ on } X = \{0, 1, 2, ..., 10\}$$ (S.U. 2008, 10)

Sol. : (i) Clearly, for $x \in X$, we have

$$A(x) = \frac{x}{x+2} \ge \frac{x}{x+5} = B(x)$$

Hence, $A \cap B = B$.

$$\therefore \quad S(A, B) = \frac{|A \cap B|}{|A|} = \frac{|B|}{|A|}$$

(ii) $$S(B, A) = \frac{|A \cap B|}{|B|} = \frac{|B|}{|B|} = 1.$$

Example 3 : Find the degree of subsethood for the fuzzy sets given by

$$A(x) = \frac{x}{x+1} \text{ and } B(x) = 1 - \frac{x}{10} \text{ for } x \in \{0, 1, 2, ..., 10\}$$

(S.U. 2000)

Sol. : We can write

$$A = \frac{0}{0} + \frac{1/2}{1} + \frac{2/3}{2} + \frac{3/4}{3} + \frac{4/5}{4} + \frac{5/6}{5} + \frac{6/7}{6} + \frac{7/8}{7} + \frac{8/9}{8} + \frac{9/10}{9} + \frac{10/11}{10}$$

$$B = (1, 0{\cdot}9, 0{\cdot}8, 0{\cdot}7, 0{\cdot}6, 0{\cdot}5, 0{\cdot}4, 0{\cdot}3, 0{\cdot}2, 0{\cdot}1, 0)$$

$$\therefore \quad A \cap B = (0, 0{\cdot}5, \frac{2}{3}, 0{\cdot}7, 0{\cdot}6, 0{\cdot}5, 0{\cdot}4, 0{\cdot}3, 0{\cdot}2, 0{\cdot}1, 0)$$

Hence, summation of all numbers in the above bracket gives

$$|A \cap B| = 3{\cdot}3 + \frac{2}{3} = 3{\cdot}96$$

$$|A| = \sum_{n=1}^{10} \frac{n}{n+1} = 7{\cdot}97, \quad |B| = 5{\cdot}5.$$

$$\therefore \quad S(B, A) = \frac{|A \cap B|}{|B|} = \frac{3{\cdot}96}{5{\cdot}5} = 0{\cdot}72.$$

Similarly, $$S(A, B) = \frac{|A \cap B|}{|A|} = \frac{3{\cdot}96}{7{\cdot}97} = 0{\cdot}5.$$

Example 4 : For the fuzzy sets defined by

$$A(x) = \frac{x}{x+2}, \; B(x) = 2^{-x}; \quad x \in \{0, 1, 2, 3, 4, 5\}$$

Calculate $|\tilde{A}|, |\tilde{B}|, S(|\tilde{A}|, |\tilde{B}|), S(|\tilde{B}|, |\tilde{A}|)$.

(S.U. May 2011)

Sol. : We have $$A = \frac{0}{0} + \frac{1/3}{1} + \frac{2/4}{2} + \frac{3/5}{3} + \frac{4/6}{4} + \frac{5/7}{5}$$

Therefore, $$L(A) = \left\{\frac{1}{3}, \frac{2}{4}, \frac{3}{5}, \frac{4}{6}, \frac{5}{7}\right\}$$

We write $L(A)$ in decreasing order in the first column and complete the following table.

α-level	Elements of degree α	Total member of elements	$\|^{\alpha}A\|$
5/7	5	1	1
4/6	4	1	2
3/5	3	1	3
2/4	2	1	4
1/3	1	1	5

Since $|\tilde{A}| = \Sigma \frac{\alpha}{|^{\alpha}A|}$, from columns 1st and 4th, we get

$$|\tilde{A}| = \frac{5/7}{1} + \frac{4/6}{2} + \frac{3/5}{3} + \frac{2/4}{4} + \frac{1/3}{5} \quad \text{.....(1)}$$

Therefore, the scalar cardinality of $|\tilde{A}|$ is

$$||\tilde{A}|| = \frac{5}{7} + \frac{4}{6} + \frac{3}{5} + \frac{2}{4} + \frac{1}{3} = 2{\cdot}81$$

$$B = \frac{1}{0} + \frac{1/2}{1} + \frac{1/4}{2} + \frac{1/8}{3} + \frac{1/16}{4} + \frac{1/32}{5} \quad \text{.....(2)}$$

Therefore, the level set of B is

$$L(B) = \left\{1, \frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}, \frac{1}{32}\right\}$$

This is in decreasing order. We write it as the 1st column and complete the following table.

α-level	Elements of degree α	Total member of elements	$\|^{\alpha}A\|$
1	0	1	
1/2	1	1	1
1/4	2	1	2
1/8	3	1	3
1/16	4	1	4
1/32	5	1	5
			6

Since $|\tilde{B}| = \Sigma \frac{\alpha}{|^{\alpha}B|}$, we get from the columns 1st and 4th

$$|\tilde{B}| = \frac{1}{1} + \frac{1/2}{2} + \frac{1/4}{3} + \frac{1/8}{4} + \frac{1/16}{5} + \frac{1/32}{6} \quad \text{.....(3)}$$

Therefore, the scalar cardinality of $|\tilde{B}|$ is

$$||\tilde{B}|| = 1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} + \frac{1}{32} = 1{\cdot}96 \quad \text{.....(4)}$$

From (1) and (3), we get

$$|\tilde{A}| \cap |\tilde{B}| = \frac{5/7}{1} + \frac{1/2}{2} + \frac{1/4}{3} + \frac{1/8}{4} + \frac{1/16}{5} + \frac{0}{6}$$

Therefore, the scalar cardinality

$$\left||\tilde{A}| \cap |\tilde{B}|\right| = \frac{5}{7} + \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} = 1{\cdot}65 \quad \text{.....(5)}$$

From (2) and (4) and (5), we get

$$S(|\tilde{A}|, |\tilde{B}|) = \frac{||\tilde{A}| \cap |\tilde{B}||}{||\tilde{A}||} = \frac{1{\cdot}65}{2{\cdot}81} = 0{\cdot}58$$

and $$S(|\tilde{B}|, |\tilde{A}|) = \frac{||\tilde{A}| \cap |\tilde{B}||}{||\tilde{B}||} = \frac{1{\cdot}65}{1{\cdot}96} = 0{\cdot}84$$

EXERCISE - IV

1. Find the scalar cardinality of the following fuzzy sets :

$$A = \frac{0{\cdot}5}{x} + \frac{0{\cdot}4}{y} + \frac{1}{z}$$

(S.U. 2009) [Ans. : 1·9]

2. Let C and D be the fuzzy sets defined on $X = \{0, 1, 2, ..., 10\}$ by

$$C(x) = \frac{x}{x+2}, \quad D(x) = \frac{x}{x+5}.$$ Find $|C|$, $|D|$ and $|C \cap D|$. (S.U. 2009)

[**Ans. :** $|C| = \sum_1^{10} \frac{n}{n+2}$, $|D| = \sum_1^{10} \frac{n}{n+5}$, and $|C \cap D| = |D|$]

3. Consider the fuzzy sets

$$A(x) = \frac{x}{x+1}, \quad B(x) = 1 - \frac{x}{10}; \quad x \in \{0, 1, 2, ..., 10\}$$

Calculate $|\tilde{A}|$, $|\tilde{B}|$, $|\tilde{A}| \cup |\tilde{B}|$, $|\tilde{A}| \cap |\tilde{B}|$. (S.U. 2010)

[**Ans. :** $|\tilde{A}| = \frac{10/11}{1} + \frac{9/10}{2} + \frac{8/9}{3} + \frac{7/8}{4} + \frac{6/7}{5} + \frac{5/6}{6} + \frac{4/5}{7} + \frac{3/4}{8} + \frac{2/3}{9} + \frac{1/2}{10}$

$$|\tilde{B}| = \frac{1}{1} + \frac{0{\cdot}9}{2} + \frac{0{\cdot}8}{3} + \frac{0{\cdot}7}{4} + \frac{0{\cdot}6}{5} + \frac{0{\cdot}5}{6} + \frac{0{\cdot}4}{7} + \frac{0{\cdot}3}{8} + \frac{0{\cdot}2}{9} + \frac{0{\cdot}1}{10}$$

$$|\tilde{A}| \cup |\tilde{B}| = \frac{1}{1} + \frac{0{\cdot}9}{2} + \frac{8/9}{3} + \frac{7/8}{4} + \frac{6/7}{5} + \frac{5/6}{6} + \frac{4/5}{7} + \frac{3/4}{8} + \frac{2/3}{9} + \frac{1/2}{10}$$

$$|\tilde{A}| \cap |\tilde{B}| = \frac{10/11}{1} + \frac{9/10}{2} + \frac{0{\cdot}8}{3} + \frac{0{\cdot}7}{4} + \frac{0{\cdot}6}{5} + \frac{0{\cdot}5}{6} + \frac{0{\cdot}4}{7} + \frac{0{\cdot}3}{8} + \frac{0{\cdot}2}{9} + \frac{0{\cdot}1}{10}$$]

4. Find the scalar cardinality of the fuzzy set defined by

$$A(x) = 1 - \frac{x}{5} \text{ on } X = \{0, 1, 2, 3, 4, 5\}$$ [**Ans. :** $|A| = 1 + \frac{4}{5} + \frac{3}{5} + \frac{2}{5} + \frac{1}{5} = 3$]

5. Consider the fuzzy sets given by

$$A(x) = \frac{x}{x+1}, \quad B(x) = 1 - \frac{x}{4}; \quad x \in \{0, 1, 2, 3, 4\}$$

Calculate $|\tilde{A}| \cup |\tilde{B}|$, $|\tilde{A}| \cap |\tilde{B}|$, $|A \tilde{\cup} B|$, $|A \tilde{\cap} B|$ and show that

$$|\tilde{A}| \cup |\tilde{B}| \neq |A \tilde{\cup} B| \quad \text{and} \quad |\tilde{A}| \cap |\tilde{B}| \neq |A \tilde{\cap} B|.$$

[**Ans. :** $|\tilde{A}| \cup |\tilde{B}| = \frac{1}{1} + \frac{3/4}{2} + \frac{2/3}{3} + \frac{1/2}{4}$

$$|A \tilde{\cup} B| = \frac{1}{1} + \frac{4/5}{2} + \frac{3/4}{4} + \frac{2/3}{5}$$

$$|\tilde{A}| \cap |\tilde{B}| = \frac{4/5}{1} + \frac{3/4}{2} + \frac{1/2}{3} + \frac{1/4}{4}$$

$$|A \tilde{\cap} B| = \frac{1/2}{2} + \frac{1/4}{3}.$$]

6. Consider the fuzzy sets defined by

$$A(x) = \frac{x}{x+2} \quad \text{and} \quad B(x) = \frac{x}{x+5}; \quad x \in \{0, 1, 2, ..., 10\}$$

Find $|A|$, $|B|$, $S(A, B)$, $S(B, A)$. (S.U. 2010)

[**Ans. :** $S(A, B) = \dfrac{\sum_1^{10} \frac{n}{n+5}}{\sum_1^{10} \frac{n}{n+2}}$, $S(B, A) = 1$. Refer to Problem 2 above]

☆☆☆

CHAPTER 9

Fuzzy Arithmetic

1. Fuzzy Numbers : Introduction

Arithmetic of fuzzy numbers has the same importance in fuzzy mathematics that arithmetic has in mathematics. Some special types of fuzzy subsets of the set of all real numbers are called **fuzzy numbers**. Fuzzy numbers are defined in several ways. We here state the definition which uses the least jargan and looks natural.

2. Definition and Examples

A fuzzy set on the set of all real numbers $\mathbb{R}$ is called a fuzzy number, if it satisfies the following conditions :

(i) A is normal (*i.e.*, there exists $a \in \mathbb{R}$ such that $A(a) = 1$).

(ii) ${}^{\alpha}A$ is a closed interval for every $\alpha \in (0, 1]$ (*i.e.*, ${}^{\alpha}A = [a_1^{\alpha}, a_2^{\alpha}]$ for each α-cut $\alpha \neq 0$. Note that ${}^{\alpha}A = \mathbb{R}$ for $\alpha = 0$).

(iii) ${}^{0+}A$ is bounded.

Note ....

The condition ${}^{0+}A$ is bounded is not assumed at times to include "unbounded fuzzy numbers. However, we do not consider such numbers. Since ${}^{\alpha}A \subseteq {}^{0+}A$ for $\alpha \in (0, 1]$, it follows that ${}^{\alpha}A = [a_1^{\alpha}, a_2^{\alpha}]$ is a closed bounded interval.

In application we frequently use the following characterization of fuzzy numbers.

Theorem : Let A be a fuzzy set on $\mathbb{R}$. A is a fuzzy number if and only if there exists a closed bounded interval $[a, b] \subset \mathbb{R}$, such that

$$\begin{aligned} A(x) &= 1, && \text{for } x \in [a, b]. \\ &= l(x), && \text{for } x \in (-\infty, a) \\ &= r(x), && \text{for } x \in (b, \infty), \quad \text{where} \end{aligned}$$

(i) $l : (-\infty, a) \to [0, 1]$ is monotonically increasing function continuous from the right and $l(x) = 0$ for $x \in (-\infty, a_1)$ where $a_1 \leq a$.

(ii) $r : (b, \infty) \to [0, 1]$ is monotonically decreasing function continuous from the left and $r(x) = 0$ for $x \in (a_2, \infty)$, where $b \leq a_2$.

Example 1 : Let $c \in \mathbb{R}$. Define

$$\tilde{c} : \mathbb{R} \to [0, 1]$$

$$\tilde{c}(x) = \begin{cases} 1, & \text{if } x = c \\ 0, & \text{if } x \neq c \end{cases}$$

The graph of $\tilde{c}$ is shown in the Fig. 9.1. $\tilde{c}$ is a fuzzy number, because it satisfies the above theorem, for $a = c$, $b = c$, $a_1 = c$, $a_2 = c$; $l(x) = 0$ for

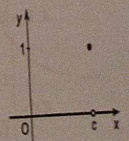


Fig. 9.1

all $x \in (-\infty, c)$. Being a constant function, l is trivially, monotonically increasing; similarly $r(x) = 0$, for all $x \in (c, \infty)$ and being a constant function it is trivially, monotonically decreasing.

Note that $\bar{c}$ can be identified with the real number $c \in \mathbb{R}$.

Thus, every real number is a fuzzy (real) number.

Example 2 : Consider the fuzzy set

$$A : \mathbb{R} \to [0, 1]$$

$$A(x) = \begin{cases} 1, & \text{for } x \in [a, b] \\ 0, & \text{everywhere else} \end{cases}$$

Here, $l(x) = 0$, for $x \in (-\infty, a)$

$r(x) = 0$, for $x \in (b, 0)$.

Therefore, A is a fuzzy number.

Thus every closed bounded interval of $\mathbb{R}$ is a fuzzy number.

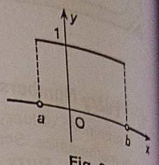


Fig. 9.2

Example 3 : Consider the fuzzy set

$$A : \mathbb{R} \to [0, 1]$$

$$A(x) = \begin{cases} 1 + \dfrac{x - c}{c - a_1} & \text{for } x \in [a_1, c] \\ 1 + \dfrac{x - c}{c - a_2} & \text{for } x \in [c, a_2] \\ 0, & \text{otherwise} \end{cases}$$

Here, $a = b = c$,

$$l : (-\infty, c) \to [0, 1]$$

$$l(x) = \begin{cases} 1 + \dfrac{x - c}{c - a_1}, & \text{for } x \in [a_1, c] \\ 0, & \text{otherwise} \end{cases}$$

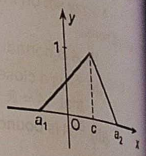


Fig. 9.3

is an increasing function. It is continuous. Moreover $l(x) = 0$, for $x \in (-\infty, a_1)$, where $a_1 < c$.

$$r : (c, \infty) \to [0, 1]$$

$$r(x) = \begin{cases} 1 + \dfrac{x - c}{c - a_2}, & \text{for } x \in [c, a_2] \\ 0, & \text{otherwise} \end{cases}$$

r is decreasing and is a continuous function. Also $r(x) = 0$, for all $x \in (a_2, \infty)$.

Hence, A is a fuzzy number. This is a very common and useful fuzzy number. It is called a triangular fuzzy number because of the shape of its graph. It is usually denoted as (a_1, c, a_2).

Example 4 : Consider the fuzzy set

$$A : \mathbb{R} \to [0, 1]$$

$$A(x) = \begin{cases} 1, & \text{for } x \in [a, b] \\ 1 + \dfrac{x - a}{a - a_1}, & \text{for } x \in [a_1, a] \\ 1 + \dfrac{x - b}{b - a_2}, & \text{for } x \in [b, a_2] \\ 0, & \text{otherwise} \end{cases}$$

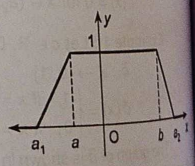


Fig. 9.4

We leave it to readers to complete the solution and convince themselves that A is a fuzzy number.

Example 5 : Consider the fuzzy set

$$C : \mathbb{R} \to [0, 1]$$

$$C(x) = \begin{cases} 1 - \dfrac{(x - c)^2}{(a_1 - c)^2} & \text{for all } x \in (a_1, 2c - a_1) \\ 0, & \text{otherwise} \end{cases}$$

Here, $a = b = c$, $a_2 = 2c - a_1$.

$$l : (-\infty, c) \to [0, 1]$$

$$l(x) = \begin{cases} 1 - \dfrac{(x - c)^2}{(a_1 - c)^2} & \text{for } x \in (a_1, c) \\ 0, & \text{otherwise} \end{cases}$$

$$r : (c, \infty) \to [0, 1]$$

$$r(x) = \begin{cases} 1 - \dfrac{(x - c)^2}{(a_1 - c)^2} & \text{for } x \in (c, 2c - a_1) \\ 0, & \text{otherwise} \end{cases}$$

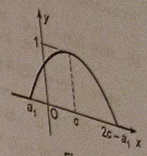


Fig. 9.5

A reader can easily see that A is a fuzzy number.

Example 6 : The fuzzy set defined on $\mathbb{R}$ which have graphs as below are not fuzzy number.

In graph Fig. 9.6 (a), the height of the fuzzy set is less than 1.

In graph Fig. 9.6 (b), ${}^{\alpha}A = [a, b] \cup [c, d]$ is not an interval. (It is union of disjoint intervals).

In graph Fig. 9.6 (c), ${}^{0+}A$ is not bounded.

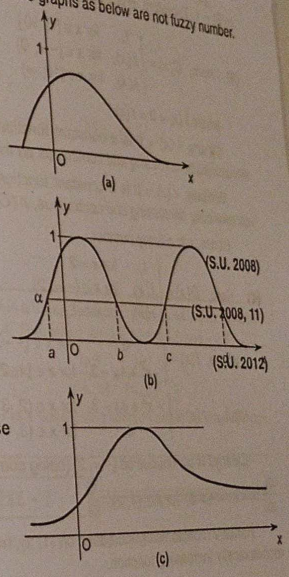


Fig. 9.6

Example 7 : Determine which of the fuzzy set are fuzzy numbers, using the criteria for a fuzzy set to be a fuzzy number.

1. $A(x) = \begin{cases} \min\{1, x\}, & x \geq 0 \\ 0, & x < 0 \end{cases}$
2. $B(x) = \begin{cases} \sin x, & 0 \leq x \leq \pi \\ 0, & \text{otherwise} \end{cases}$
3. $C(x) = \begin{cases} 1, & \text{for } 0 \leq x \leq 10 \\ 0, & \text{otherwise} \end{cases}$
4. $D(x) = \begin{cases} -x^2 + 4x - 3, & \text{for } 1 \leq x \leq 3 \\ 0, & \text{everywhere else} \end{cases}$

Sol. : (1) A fuzzy number $A : \mathbb{R} \to [0, 1]$ must possess three properties

(a) A is normal

(b) ${}^{\alpha}A$ is closed interval for $\alpha \in (0, 1]$

(c) ${}^{0+}A$ is bounded.

Some author relax the condition (c). Clearly, ${}^{0+}A = (0, \infty)$.

Hence, if we accept (c), then A is not a fuzzy numbers, otherwise it is.

Some authors call it unbounded fuzzy number.

Note that ${}^{1}A = [1, \infty)$ is "technically" a closed interval (though not bounded).

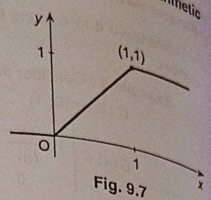


Fig. 9.7

(2) Here, $B(x) = \begin{cases} 1, & \text{at } x = \pi/2 \\ l(x), & \text{for } x \in (-\infty, \pi/2) \\ r(x), & \text{for } x \in (\pi/2, \infty) \end{cases}$

where, $l : (-\infty, \pi/2) \to [0, 1]$ is given by

$$l(x) = \begin{cases} \sin x, & \text{for } x \in (0, \pi/2) \\ 0, & \text{otherwise} \end{cases}$$

and $r : (\pi/2, \infty) \to [0, 1]$ is given by

$$r(x) = \begin{cases} \sin x, & \text{for } x \in (\pi/2, \pi) \\ 0, & \text{otherwise} \end{cases}$$

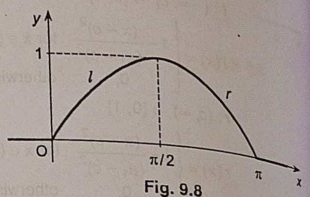


Fig. 9.8

Clearly, **(a)** l is monotonically increasing continuous function, as $\sin x$ increases from 0 to 1 on $[0, \pi/2]$.

(b) r is monotonically decreasing function, as $\sin x$ decreases from 1 to 0 on $[\pi/2, \pi]$, $h(B) = 1$. Therefore, B is a fuzzy number

(3) Here, $C(x) = \begin{cases} 1, & \text{for } x \in [0, 10] \\ l(x), & \text{for } x \in (-\infty, 0) \\ r(x), & \text{for } x \in (10, \infty) \end{cases}$

where $l(x) = 0 = r(x)$.

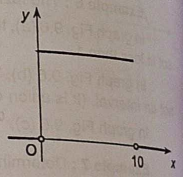


Fig. 9.9

Clearly, $l(x) = 0$ is a constant function and hence trivially monotonically increasing and continuous on the interval $(-\infty, 0)$.

Similarly, $r(x) = 0$ is a constant function and hence trivially monotonically decreasing and continuous, $h(C) = 1$.

Hence, c is a fuzzy number.

(4) Here, $D(x) = \begin{cases} 1, & \text{for } x = 2 \\ l(x), & \text{for } x \in (-\infty, 2) \\ r(x), & \text{for } x \in (2, \infty) \end{cases}$

where, $l(x) = \begin{cases} 0, & \text{for } x \in (-\infty, 1) \\ -x^2 + 4x - 3, & \text{for } x \in [1, 2) \end{cases}$

and $r(x) = \begin{cases} -x^2 + 4x - 3, & \text{for } x \in (2, 3] \\ 0, & \text{for } x \in (3, \infty) \end{cases}$

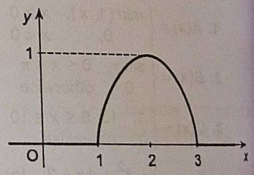


Fig. 9.10

Clearly l is continuous on $(-\infty, 2)$ (being constant on $(-\infty, 1)$ and then polynomial). Moreover

$$\frac{dl}{dx} = -2x + 4 > 0, \text{ for } x \in (1, 2).$$

Hence, l monotonically increases on $(1, 2)$, from 0. Moreover $l(x) = 0$, on $(-\infty, 1]$. Hence, l is monotonically increasing function.

On the other hand $\frac{dr}{dx} = -2x + 4 < 0$, for $x \in (2, 3)$.

Hence, r is monotonically decreasing on $(2, 3)$. Clearly, r decreases to 0 on the interval $(2, 3)$ and $r(x) = 0$ on $[3, \infty)$. Hence, r is monotonically decreasing on $(2, \infty)$.

It is clearly continuous (being polynomial on $(2, 3)$ and then constant).

Therefore, D is a fuzzy number.

Example 8 : Verify if the following fuzzy set is a fuzzy number

$$A = \frac{0{\cdot}2}{[0, 1)} + \frac{0{\cdot}6}{[1, 2)} + \frac{0{\cdot}8}{[2, 3)} + \frac{0{\cdot}9}{[3, 4)} + \frac{1}{4} + \frac{0{\cdot}5}{(4, 5]} + \frac{0{\cdot}1}{(5, 6]}$$

Sol :

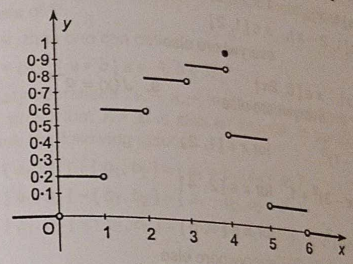


Fig. 9.11

(i) A is normal for $A(4) = 1$.

(ii) The distinct α-cuts are :

${}^{1}A = 4 = [4, 4]$, ${}^{0{\cdot}9}A = [3, 4]$, ${}^{0{\cdot}8}A = [2, 4]$, ${}^{0{\cdot}6}A = [1, 4]$,

${}^{0{\cdot}5}A = [1, 5]$, ${}^{0{\cdot}2}A = [0, 5]$, ${}^{0{\cdot}1}A = [0, 6]$.

Thus, ${}^{\alpha}A$ is a closed interval for each $\alpha \in (0, 1]$.

(iii) ${}^{0+}A = [0, 6]$ is bounded.

Hence, by definition A is a fuzzy number.

Note ....

We may apply the theorem also, to prove that A is a fuzzy number. For this recall the convention that 0 is not mentioned in the above expression of A. Therefore, $A(x) = 0$ for $x > 6$ and $x < 0$.

$$l = \frac{0{\cdot}2}{[0, 1)} + \frac{0{\cdot}6}{[1, 2)} + \frac{0{\cdot}8}{[2, 3)} + \frac{0{\cdot}9}{[3, 4)} \quad \text{and} \quad r = \frac{0{\cdot}5}{(4, 5]} + \frac{0{\cdot}1}{(5, 6]}$$

l is increasing and is continuous *from the right* and r is a decreasing function and is continuous *from the left*.

EXERCISE - I

Determine, which of the following fuzzy sets are fuzzy numbers. Justify your answer.

1. $A(x) = \begin{cases} 1-\cos x, & \text{for } 0 \le x \le (\pi/2) \\ 1+\cos x, & \text{for } (\pi/2) \le x \le \pi \\ 0, & \text{for } x \le 0 \text{ or } x \ge \pi \end{cases}$

2. $B(x) = \begin{cases} 1-|x|, & \text{for } x \in (-1, 1) \\ 0, & \text{everywhere else} \end{cases}$

3. $C(x) = \begin{cases} \max\{0, x\}, & x \le 1 \\ \max\{0, 2-x\}, & x \ge 1 \end{cases}$

4. $D(x) = \begin{cases} (x-1)(3-x), & \text{for } 1 \le x \le 3 \\ 0, & \text{everywhere else} \end{cases}$

5. $E(x) = \begin{cases} x-0{\cdot}5, & \text{for } 0{\cdot}5 \le x \le 1 \\ 1{\cdot}5-x, & \text{for } 1 \le x \le 1{\cdot}5 \\ 0, & \text{everywhere else} \end{cases}$

6. $F(x) = \begin{cases} 1, & \text{for } x \in [1, 3] \\ 1+\dfrac{x-1}{2}, & \text{for } x \in [-1, 1] \\ 4-x, & \text{for } x \in (3, 4] \\ 0, & \text{otherwise} \end{cases}$

7. $G(x) = \begin{cases} \min\{1, x\}, & 1 > x \ge 0 \\ \min\{1, 2-x\}, & x \in [1, 2] \\ 0, & \text{everywhere else} \end{cases}$

8. $H(x) = \begin{cases} |\sin x|, & x \in [0, 2\pi] \\ 0, & \text{everywhere else} \end{cases}$

9. $J(x) = e^{-\pi x^2}$

10. $K(x) = \begin{cases} \dfrac{1}{2}(x-1)^2, & \text{for } x \in [1, 2) \\ -\dfrac{1}{2}(x-3)^2+1, & \text{for } x \in [2, 4] \\ \dfrac{1}{2}(x-5)^2, & \text{for } x \in [4, 5] \\ 0, & \text{everywhere else} \end{cases}$

11. $L(x) = \begin{cases} -\dfrac{1}{4}(x+1)(x-3), & \text{for } x \in [-1, 1] \\ 1, & \text{for } x \in [1, 2] \\ -(x-1)(x-3), & \text{for } x \in [2, 3] \\ 0, & \text{everywhere else} \end{cases}$

12. $M(x) = \begin{cases} \dfrac{x+1}{3}, & \text{for } x \in [-1, 2] \\ 1, & \text{for } x \in [2, 4] \\ 5-x, & \text{for } x \in [4, 5] \\ 0, & \text{everywhere else} \end{cases}$

13. $N(x) = \begin{cases} 1, & \text{for } x = 6 \\ 0, & \text{for } x \ne 6 \end{cases}$

14. $P(x) = \begin{cases} x, & \text{for } 0 \le x \le 1 \\ 0, & \text{everywhere else} \end{cases}$

15. $Q(x) = \begin{cases} x, & \text{for } 0 \le x \le \dfrac{1}{2} \\ 0, & \text{everywhere else} \end{cases}$

16. $A(x) = \begin{cases} 1-\left(\dfrac{x-a}{a_1-a}\right)^2, & \text{for } a_1 \le x \le 2a-a_1 \\ 0, & \text{otherwise, given that } a_1 < a \end{cases}$

17. Is the fuzzy set A given below is a fuzzy number. Substantiate your answer.

$$A = \frac{0{\cdot}1}{[0, 1)} + \frac{0{\cdot}2}{[1, 2)} + \frac{0{\cdot}6}{[2, 3)} + \frac{0{\cdot}7}{[3, 4)} + \frac{0{\cdot}8}{[4, 5)} + \frac{0{\cdot}9}{[5, 6)} + \frac{1}{6} + \frac{0{\cdot}5}{(6, 7]} + \frac{0{\cdot}4}{(7, 8]} + \frac{0{\cdot}2}{(8, 9]} + \frac{0{\cdot}1}{(9, 10]}$$

18. $A : \mathbb{R} \to [0, 1]$

$$A(x) = \begin{cases} 1+x, & -1 \le x \le 0 \\ 0, & \text{otherwise} \end{cases}$$

19. $A : \mathbb{R} \to [0, 1]$

$$A(x) = \begin{cases} 1-x, & 0 \le x \le 1 \\ 0, & \text{otherwise} \end{cases}$$

[**Ans.** : Yes : 1 to 4, 6, 7, 10 to 14, 16 to 19; No : 5, 8, 15

In the case of 9, your answer depends on the definition. According to our definition J is not a fuzzy number for ${}^{0+}A = (-\infty, \infty)$ which is not bounded.]

3. Operations on Intervals

All α-cuts of fuzzy numbers are closed intervals, if $\alpha \ne 0$. Therefore, arithmetic of fuzzy numbers is closely connected to the arithmetic of closed intervals. Fortunately, there already exists arithmetic of closed bounded intervals of $\mathbb{R}$.

Let $A \subseteq \mathbb{R}$ and $B \subseteq \mathbb{R}$, then one can define

$$A * B = \{ a * b \mid a \in A, b \in B \}$$

where, $*$ is any of the binary operations : $+, -, \times, \div$; the last one is defined when $0 \notin B$. When A and B are intervals one can prove that $A * B$ is also an interval. In particular, if $A = [a_1, a_2]$ and $B = [b_1, b_2]$, then we have the following results.

Addition : $[a_1, a_2] + [b_1, b_2] = [a_1 + b_1, a_2 + b_2]$

Subtraction : $[a_1, a_2] - [b_1, b_2] = [a_1 - b_2, a_2 - b_1]$

Multiplication : $[a_1, a_2] \times [b_1, b_2] = [\min\{a_1b_1, a_1b_2, a_2b_1, a_2b_2\}, \max\{a_1b_1, a_1b_2, a_2b_1, a_2b_2\}]$

We usually drop the multiplication sign.

Division : If $0 \notin [b_1, b_2]$, then

$$\frac{[a_1, a_2]}{[b_1, b_2]} = \left[\min\left\{\frac{a_1}{b_1}, \frac{a_1}{b_2}, \frac{a_2}{b_1}, \frac{a_2}{b_2}\right\}, \max\left\{\frac{a_1}{b_1}, \frac{a_1}{b_2}, \frac{a_2}{b_1}, \frac{a_2}{b_2}\right\}\right]$$

Notes ....

1. When $[a_1, a_2]$ and $[b_1, b_2]$ consist of only non-negative real numbers, then we have the following simpler result

$$[a_1, a_2][b_1, b_2] = [a_1b_1, a_2b_2]$$

2. In addition, if $0 \notin [b_1, b_2]$, then

$$[a_1, a_2] \div [b_1, b_2] = \left[\frac{a_1}{b_2}, \frac{a_2}{b_1}\right]$$

3. Every real number r can be treated as the interval $[r, r]$. With this identification the above algebraic operations on intervals are generalization of algebraic operations on $\mathbb{R}$.

4. $1 \cdot [a_1, a_2] = [1a_1, 1a_2] = [a_1, a_2]$

5. $\dfrac{1}{[a_1, a_2]} = \left[\dfrac{1}{a_2}, \dfrac{1}{a_1}\right]$

Illustrations

1. $[2, 7] + [3, 8] = [2+3, 7+8] = [5, 15]$
2. $[-5, 4] + [1, 2] = [-5+1, 4+2] = [-4, 6]$

3. $[2, 7] - [3, 8] = [2 - 8, 7 - 3] = [-6, 4]$

4. $[-5, 4] - [1, 2] = [-5 - 2, 4 - 1] = [-7, 3]$

5. $[-2, -7] \cdot [-3, 5] = [\min\{6, 21, -10, -35\}, \max\{6, 21, -10, -35\}]$
$= [-35, 21]$

6. $[2, 7] \cdot [3, 8] = [2 \cdot 3, 7 \cdot 8] = [6, 56]$

7. $\dfrac{[-3, 5]}{[-2, -7]} = \left[\min\left\{\dfrac{-3}{-2}, \dfrac{5}{-2}, \dfrac{-3}{-7}, \dfrac{5}{-7}\right\}, \max\left\{\dfrac{-3}{-2}, \dfrac{-3}{-7}, \dfrac{5}{-2}, \dfrac{5}{-7}\right\}\right] = \left[\dfrac{-5}{2}, \dfrac{3}{2}\right]$

8. $\dfrac{[2, 7]}{[3, 8]} = \left[\dfrac{2}{8}, \dfrac{7}{3}\right]$

The above defined operations on closed bounded intervals satisfy the following properties :
If A, B, C are closed bounded intervals of real numbers, then

Commutativity : $A + B = B + A$, $AB = BA$

Associativity : $A + (B + C) = (A + B) + C$, $A(BC) = (AB)C$

Existence of identify : $A + 0 = A$, $A \cdot 1 = A$
where $0 = [0, 0]$ and $1 = [1, 1]$.

Subdistributivity : In general, multiplication by interval does not distribute over addition of intervals, but we have

$$A \cdot (B + C) \subseteq A \cdot B + A \cdot C$$

The equality holds, if $bc \geq 0$, for all $b \in B$ and $c \in C$.

It is useful to recall the following :

(i) $a(B + C) = aB + aC$

(ii) If $A \subseteq D$, $B \subseteq E$, then

(a) $A + B \subseteq D + E$ (b) $A - B \subseteq D - E$

(c) $A \cdot B \subseteq D \cdot E$ (d) $A / B \subseteq D / E$ (whenever defined)

4. Computing a Fuzzy Number from its α-cuts

Observe that, if $\alpha \in (0, 1]$, then

$$A(x) = \text{Sup}\{\alpha \mid x \in {}^{\alpha}A\} = \text{Sup}\{\alpha \mid A(x) \geq \alpha\}$$

This result holds for all $A : X \to [0, 1]$ and not necessarily, for fuzzy numbers.

Example 1 : Compute A, if the following information is known :

${}^{0\cdot1}A = \{a, b, c, d\}$, ${}^{0\cdot3}A = \{b, c, d\}$, ${}^{0\cdot5}A = \{c, d\}$,

${}^{1}A = \{d\}$ is the list of all distinct α-cuts of A, $\alpha > 0$.

Sol. : We know that $A(a) = \text{Sup}\{\alpha \mid x \in {}^{\alpha}A\}$

Hence,
$$A(a) = \text{Sup}\{\alpha \mid a \in {}^{\alpha}A\} = \text{Sup}\{0\cdot1\} = 0\cdot1$$
$$A(b) = \text{Sup}\{\alpha \mid b \in {}^{\alpha}A\} = \text{Sup}\{0\cdot1, 0\cdot3\} = 0\cdot3$$
$$A(c) = \text{Sup}\{\alpha \mid c \in {}^{\alpha}A\} = \text{Sup}\{0\cdot1, 0\cdot3, 0\cdot5\} = 0\cdot5$$
$$A(d) = \text{Sup}\{\alpha \mid d \in {}^{\alpha}A\} = \text{Sup}\{0\cdot1, 0\cdot3, 0\cdot5, 1\} = 1$$

Hence,
$$A = \frac{0\cdot1}{a} + \frac{0\cdot3}{b} + \frac{0\cdot5}{c} + \frac{1}{d}$$

Example 2 : Let A be a fuzzy number defined on $\mathbb{R}$ such that
${}^{\alpha}A = [2\alpha - 1, -2\alpha + 3]$ for all $\alpha \in (0, 1]$, construct the fuzzy set A.

Sol. : Note that $\inf\{2\alpha - 1 \mid 0 < \alpha \leq 1\} = -1$
and $\text{Sup}\{2\alpha - 1 \mid 0 < \alpha \leq 1\} = 1$

Therefore, if $-1 < x \leq 1$, then $x = 2\alpha - 1$, and
$$A(x) = \text{Sup}\{\alpha \mid x = 2\alpha - 1\}$$
$$= \text{Sup}\{\alpha \mid \alpha = \frac{x+1}{2}\} = \frac{x+1}{2}$$

Similarly, $\inf\{-2\alpha + 3 \mid 0 < a \leq 1\} = 1$
and $\text{Sup}\{-2\alpha + 3 \mid 0 < x \leq 1\} = 3$ ..... (1)

Therefore, if $1 \leq x < 3$,
$$A(x) = \text{Sup}\{\alpha \mid -2\alpha + 3 = x\}$$
$$= \text{Sup}\{\alpha \mid \alpha = \frac{3-x}{2}\} = \frac{3-x}{2}$$

Since ${}^{\alpha}A = [2\alpha - 1, -2\alpha + 3]$ for $\alpha \in (0, 1]$, we have ${}^{0+}A = (-1, 3)$. ..... (2)

On the other hand ${}^{0}A = \mathbb{R}$. Hence, we legitimately define $A(-1) = 0$, $A(3) = 0$ (Refer to (1) and (2) respectively.)

Therefore, we have
$$A : \mathbb{R} \to [0, 1]$$
$$A(x) = \begin{cases} 0, & \text{for } x \leq -1 \text{ and } x \geq 3 \\ \dfrac{x+1}{2}, & \text{for } -1 < x \leq 1 \\ \dfrac{3-x}{2}, & \text{for } 1 < x < 3 \end{cases}$$

The process adopted above can be used in general. Therefore,

We summarize

Suppose we want to construct the fuzzy number A. We are given that

(a) There is unique element $a \in \mathbb{R}$ such that $A(a) = 1$.

(b) ${}^{\alpha}A = [a_1(\alpha), a_2(\alpha)]$, $\alpha \in (0, 1]$

(c) $a_1(\alpha) = x$ and $a_2(\alpha) = x$ are invertible function.

(d) ${}^{0+}A = (a_1, a_2)$

Then
$$A(x) = \begin{cases} 0, & \text{for } x \leq a_1,\ x \geq a_2 \\ a_1^{-1}(x) & \text{for } a_1 < x < a \\ a_2^{-1}(x) & \text{for } a < x < a_2 \end{cases}$$

The process is slightly modified if ${}^{1}A$ is not a singleton set (See Ex. 4, below).

Example 3 : Compute the fuzzy set A, whose α-cuts are given in the following table, for $0 < \alpha \leq -1$.

α :	1	0·9	0·8	0·7	0·6	0·4	0·3	0·2	0·1
${}^{\alpha}A$:	[4, 4]	[3, 4]	[3, 4]	[3, 5]	[3, 5]	[2, 5]	[2, 5]	[2, 5]	[2, 6]

Sol. : We have $^{0+}A = {}^{0\cdot1}A = [2, 6]$

Therefore, we assume $A(x) = 0$, for all $x < 2$ and $x > 6$.

We have $A(4) = 1$. Therefore, in additive notion for A we have a term ..... (1)

$$\frac{1}{4}$$ ..... (2)

Consider $^{0\cdot9}A = [3, 4]$, then we have $[3, 4] = [3, 4) \cup \{4\}$. Since we know $A(x)$ for $x = 4$, consider $x \in [3, 4)$.

Since, $A(x) = \text{Sup}\{\alpha \mid x \in {}^{\alpha}A\}$

$= \text{Sup}\{0\cdot9, 0\cdot8\} = 0\cdot9$ .... [From the table]

Therefore, we have a term

$$\frac{0\cdot9}{[3, 4)}$$ ..... (3)

Since $^{0\cdot8}A = {}^{0\cdot9}A$ and we have already considered $^{0\cdot9}A$, we turn to $^{0\cdot7}A = [3, 5]$.

Since $[3, 5] = [3, 4] \cup (4, 5]$ and we already know $A(x)$, for $x \in [3, 4]$, consider $(4, 5]$.

For $x \in (4, 5]$, $A(x) = \text{Sup}\{\alpha \mid x \in {}^{\alpha}A\}$

$= \text{Sup}\{0\cdot7, 0\cdot6, 0\cdot4, 0\cdot3, 0\cdot2, 0\cdot1\}$

$= 0\cdot7$

Therefore, we have a term

$$\frac{0\cdot7}{(4, 5]}$$ ..... (4)

Since $^{0\cdot6}A = {}^{0\cdot7}A$, we neglect $^{0\cdot6}A$ and turn to $^{0\cdot4}A = [2, 5]$.

We have $[2, 5] = [2, 3) \cup [3, 5]$

We know $A(x)$ for $x \in [3, 5]$.

For $x \in [2, 3)$, $A(x) = \text{Sup}\{0\cdot4, 0\cdot3, 0\cdot2, 0\cdot1\}$

$= 0\cdot4$

Therefore, we have a term

$$\frac{0\cdot4}{[2, 3)}$$ ..... (5)

Since $^{0\cdot2}A = {}^{0\cdot3}A = {}^{0\cdot4}A$ and we considered $^{0\cdot4}A$, we neglect $^{0\cdot2}A$ and $^{0\cdot3}A$ and consider $^{0\cdot1}A = [2, 6]$.

Since $[2, 6] = [2, 5] \cup (5, 6]$ and we know $A(x)$ for $x \in [2, 5]$, consider $x \in (5, 6]$. Here, $A(x) = \text{Sup}\{0\cdot1\} = 0\cdot1$, therefore, we have a term

$$\frac{0\cdot1}{(5, 6)}$$ ..... (6)

Combining (1) to (6), we have

$$A = \frac{0\cdot4}{[2, 3)} + \frac{0\cdot9}{[3, 4)} + \frac{1}{4} + \frac{0\cdot7}{(4, 5]} + \frac{0\cdot1}{(5, 6]}$$

The above used process can be applied in general :

We summaries

Suppose we want to construct a fuzzy number A. We are given the table of α and $^{\alpha}A$.

Step 1 : Rearrange the table in the descending order of α's, if necessary.

Step 2 : If $^{\alpha_i}A = {}^{\alpha_{i+1}}A$, then remove α_{i+1}, continue the process till no α is repeated.

Step 3 : After reduction, suppose we get

$$1 = \alpha_0 > \alpha_1 > \ldots\ldots > \alpha_n.$$

Then $^{0+}A = {}^{\alpha_n}A = [a_1, a_2]$, say.

If $^{1}A = \{a\}$, then we have a term in the expression for $A = \frac{1}{a}$

Step 4 : Suppose $^{\alpha_1}A = [x_1, y_1]$

If $[x_1, y_1] = [x_1, a) \cup \{a\} \cup (a, y_1]$

We have, the terms $\dfrac{\alpha_1}{[x_1, a)} + \dfrac{1}{a} + \dfrac{\alpha_1}{(a, y_1]}$

Step 5 : Suppose $^{\alpha_2}A = [x_2, y_2]$

If $[x_2, y_2] = [x_2, x_1) \cup [x_1, y_1] \cup (y_1, y_2]$, we get

$$\frac{\alpha_2}{[x_2, x_1)} + \frac{\alpha_1}{[x_1, a)} + \frac{1}{a} + \frac{\alpha_1}{(a, y_1]} + \frac{\alpha_2}{(y_1, y_2]}$$

Step 6 : Continue the process till all α's are exhausted.

Note ....

The process can be used, even when ^{1}A is not a singleton set, with slight modification of course. (See Ex. 6 below)

Example 4 : Obtain the fuzzy number A, for which $^{1}A = [-1, 2]$ and $^{\alpha}A = [\alpha - 2, -\alpha + 3]$ for $\alpha \in (0, 1]$.

Sol. : We are given that $^{0+}A = (-2, 3)$ and $^{1}A = [-1, 2]$

So we assume $A(x) = 0$ for $x \le -2$ and $x \ge 3$.

$A(x) = 1$, for $-1 \le x \le 2$ ..... (1)

Observe that $(-2, 3) = (-2, -1) \cup [-1, 2] \cup (2, 3)$.

For $x \in (-2, -1)$, we have $x = \alpha - 2$.

Hence, $A(x) = \text{Sup}\{\alpha \mid x = \alpha - 2\}$

$= \text{Sup}\{\alpha \mid \alpha = x + 2\}$

$= x + 2$ ..... (2)

For $x \in (2, 3)$, we have $x = -\alpha + 3$.

Hence, $A(x) = \text{Sup}\{\alpha \mid x = -\alpha + 3\}$

$= \text{Sup}\{\alpha \mid \alpha = 3 - x\}$

$= 3 - x$ ..... (3)

Hence, we have the following :

$A : \mathbb{R} \to [0, 1]$

$$A(x) = \begin{cases} 0, & \text{for } x \le -2,\ x \ge 3 \\ x + 2, & \text{for } -2 < x < -1 \\ 1, & \text{for } -1 \le x \le 2 \\ 3 - x, & \text{for } 2 < x < 3 \end{cases}$$

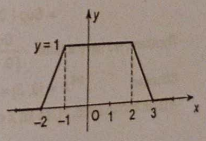


Fig. 9.12

Example 5 : Obtain the fuzzy number A, for which ${}^{1}A = [1, 2]$ and ${}^{\alpha}A = [-1 + 2\sqrt{\alpha}, 3 - 2\sqrt{\alpha}]$ for $0 < \alpha \le 1$.

Sol. : We have ${}^{0+}A = (-1, 3)$. Therefore, we assume

$$A(x) = 0, \quad \text{for } x \le -1 \text{ and } x \ge 3$$

Moreover $\quad (-1, 3) = (-1, 1) \cup [1, 2] \cup (2, 3)$. ..... (1)

For $x \in (-1, 1)$, we have $x = -1 + 2\sqrt{\alpha}$.

$$\therefore \quad A(x) = \text{Sup}\{\alpha \mid x = -1 + 2\sqrt{\alpha}\}$$

$$= \text{Sup}\left\{\alpha \,\middle|\, \left(\frac{x+1}{2}\right)^2 = \alpha\right\} = \left(\frac{x+1}{2}\right)^2 \quad \text{..... (2)}$$

For $x \in (2, 3)$, we have $x = 3 - 2\sqrt{\alpha}$.

$$\therefore \quad A(x) = \text{Sup}\{\alpha \mid x = 3 - 2\sqrt{\alpha}\}$$

$$= \text{Sup}\left\{\alpha \,\middle|\, \left(\frac{x-3}{2}\right)^2 = \alpha\right\} = \left(\frac{x-3}{2}\right)^2 \quad \text{..... (3)}$$

From (1), (2) and (3), we have

$$A(x) = \begin{cases} 0, & \text{for } x \le -1,\ x \ge 3 \\ \left(\dfrac{x+1}{2}\right)^2, & \text{for } -1 < x < 1 \\ 1, & \text{for } 1 \le x \le 2 \\ \left(\dfrac{x-3}{2}\right)^2, & \text{for } 2 < x < 3 \end{cases}$$

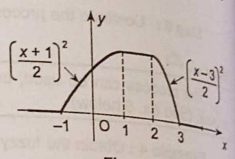


Fig. 9.13

Example 6 : Find the fuzzy number A, whose α-cuts are given in the following table for $0 < \alpha \le 1$.

α :	1	0·9	0·8	0·6	0·5	0·2	0·1
${}^{\alpha}A$:	[1, 2]	[0, 2]	[0, 3]	[0, 3]	[−2, 6]	[−2, 7]	[−3, 7]

Sol. : We have ${}^{0+}A = {}^{0+1}A = [-3, 7]$

Hence, we set $A(x) = 0$ for $x < -3$, $x > 7$. ..... (1)

We have ${}^{1}A = [1, 2]$, therefore in additive notation for A, we find the term

$$\frac{1}{[1, 2]} \quad \text{..... (2)}$$

Consider ${}^{0\cdot9}A = [0, 2] = [0, 1) \cup [1, 2]$. Since we know $A(x)$ on $[1, 2]$, let $x \in [0, 1)$.

Then, $\quad A(x) = \text{Sup}\{\alpha \mid x \in {}^{\alpha}A\}$

$$= \text{Sup}\{0\cdot9\} = 0\cdot9$$

Therefore, we have a term $\dfrac{0\cdot9}{[0, 1)}$ ..... (3)

Consider $\quad {}^{0\cdot8}A = [0, 3] = [0, 2] \cup (2, 3]$

$A(x)$ is known on $[0, 2]$, so let $x \in (2, 3]$.

Then, $\quad A(x) = \text{Sup}\{\alpha \mid x \in {}^{\alpha}A\}$

$$= \text{Sup}\{0\cdot8, 0\cdot6, 0\cdot5, 0\cdot2, 0\cdot1\} = 0\cdot8$$

Therefore, we get the term $\dfrac{0\cdot8}{(2, 3]}$

Since ${}^{0\cdot6}A = {}^{0\cdot8}A$ and we have considered ${}^{0\cdot8}A$, we turn to ${}^{0\cdot5}A = [-2, 6] = [-2, 0) \cup [0, 3] \cup (3, 6]$ ..... (4)

We know $A(x)$ on $[0, 3]$.

Consider $x \in [-2, 0) \cup (3, 6]$.

Then, $\quad A(x) = \text{Sup}\{\alpha \mid x \in {}^{\alpha}A\}$

$$= \text{Sup}\{0\cdot5, 0\cdot2, 0\cdot1\} = 0\cdot5$$

Therefore, we get two terms $\dfrac{0\cdot5}{[-2, 0)}$ and $\dfrac{0\cdot5}{(3, 6]}$

Consider $\quad {}^{0\cdot2}A = [-2, 7] = [-2, 6] \cup (6, 7]$.

Since, $A(x)$ is known on $[-2, 6]$. ..... (5)

Consider $\quad x \in (6, 7]$.

Then, $\quad A(x) = \text{Sup}\{\alpha \mid x \in {}^{\alpha}A\}$

$$= \text{Sup}\{0\cdot2, 0\cdot1\} = 0\cdot2$$

Therefore, we get the term $\dfrac{0\cdot2}{(6, 7]}$

Finally, consider ${}^{0\cdot1}A = [-3, 7] = [-3, -2) \cup [-2, 7]$. ..... (6)

Since we know $A(x)$ on $[-2, 7]$, we calculate $A(x)$ for $x \in [-3, -2)$.

Here, $A(x) = \text{Sup}\{0\cdot1\} = 0\cdot1$.

Therefore, we have $\dfrac{0\cdot1}{[-3, -2)}$

Combining (1) to (7), we have ..... (7)

$$A(x) = \frac{0\cdot1}{[-3, -2)} + \frac{0\cdot5}{[-2, 0)} + \frac{0\cdot9}{[0, 1)} + \frac{1}{[1, 2]} + \frac{0\cdot8}{(2, 3]} + \frac{0\cdot5}{(3, 6]} + \frac{0\cdot2}{(6, 7]}$$

EXERCISE - II

Construct A from its α-cuts in the following case :

1. ${}^{1}A = \{6\}$, ${}^{0\cdot7}A = \{5, 6\}$, ${}^{0\cdot4}A = \{4, 5, 6\}$, ${}^{0\cdot3}A = \{3, 4, 5, 6, 7\}$

[**Ans. :** $\dfrac{0\cdot3}{3} + \dfrac{0\cdot4}{4} + \dfrac{0\cdot7}{5} + \dfrac{1}{6} + \dfrac{0\cdot3}{7}$]

2. ${}^{\alpha}A = [2\alpha + 2, -2\alpha + 6]$, on the set of real numbers $\mathbb{R}$.

[**Ans. :** $A(x) = \begin{cases} \dfrac{x-2}{2}, & \text{for } 2 < x \le 4 \\ \dfrac{6-x}{2}, & \text{for } 4 < x \le 6 \\ 0, & \text{elsewhere} \end{cases}$

3.

α :	1	0·9	0·8	0·7	0·6	0·4	0·3	0·2	0·1
${}^{\alpha}A$:	[3, 3]	[3, 3]	[2, 3]	[2, 3]	[2, 4]	[2, 4]	[1, 4]	[1, 5]	[1, 5]

[Ans. : $\frac{0{\cdot}3}{[1,2)} + \frac{0{\cdot}8}{[2,3)} + \frac{1}{3} + \frac{0{\cdot}6}{(3,4]} + \frac{0{\cdot}2}{(4,5]}$]

4.

α :	0·2	0·5	1
${}^{a}A$:	[0, 10]	[3,7]	[4, 5]

[Ans. : $\frac{0{\cdot}2}{[0,3)} + \frac{0{\cdot}5}{[3,4)} + \frac{1}{[4,5]} + \frac{0{\cdot}5}{(5,7]} + \frac{0{\cdot}2}{(7,10]}$]

5. ${}^{\alpha}A = (-\sqrt{1-\alpha}, \sqrt{1-\alpha})$, for $\alpha \in (0, 1]$.

[Ans. : $A(x) = 1 - x^2$ for $-1 < x < 1$, $A(x) = 0$, overywhere else.]

6. ${}^{1}A = [-1, 2]$, ${}^{\alpha}A = [-1 - 2\sqrt{1-\alpha}, 2 + 3\sqrt{1-\alpha}]$

[Ans. : $A(x) = \begin{cases} \frac{3 - 2x - x^2}{4}, & \text{for } -3 < x < -1 \\ \frac{5 + 4x - x^2}{9}, & \text{for } 2 < x < 5 \\ 1, & \text{for } -1 \le x \le 2 \\ 0, & \text{otherwise} \end{cases}$

5. Algebraic Operation on Fuzzy Numbers

We know how a fuzzy number is associated with the collection of crisp sets, namely the α-cuts of the fuzzy number. Conversely given all the α-cuts of a fuzzy number, we can construct the fuzzy number. We will use these facts to develope the arithmetic of fuzzy numbers.

Definition : If A and B are fuzzy numbers, then $A + B$, $A - B$, AB and A/B are fuzzy numbers, whose α-cuts are determined by the following equations :

(a) ${}^{\alpha}(A + B) = {}^{\alpha}A + {}^{\alpha}B$ (b) ${}^{\alpha}(A - B) = {}^{\alpha}A - {}^{\alpha}B$

(c) ${}^{\alpha}(A \cdot B) = {}^{\alpha}A \cdot {}^{\alpha}B$ (d) ${}^{\alpha}(A / B) = {}^{\alpha}A / {}^{\alpha}B$, if $0 \notin {}^{\alpha}B$.

It can be proved, using elementary techniques, that $A + B$, $A - B$, $A \cdot B$, A / B (when defined) are fuzzy numbers.

Let ${}^{\alpha}A = [a_1^{\alpha}, a_2^{\alpha}]$, ${}^{\alpha}B = [b_1^{\alpha}, b_2^{\alpha}]$

Then, ${}^{\alpha}A + {}^{\alpha}B = [a_1^{\alpha} + b_1^{\alpha}, a_2^{\alpha} + b_2^{\alpha}]$

$\therefore$ ${}^{\alpha}(A + B) = [a_1^{\alpha} + b_1^{\alpha}, a_2^{\alpha} + b_2^{\alpha}]$ by the above definition.

Thus each α-cut of $A + B$, is a closed bounded interval for $a \in (0, 1]$. ..... (1)

In particular ${}^{1}(A + B) = {}^{1}A + {}^{1}B \neq \Phi$.

Hence, there exists $z \in {}^{1}(A + B)$, such that $(A + B)(z) = 1$.

$\therefore$ $h(A + B) = 1$ ..... (2)

Again ${}^{0+}(A + B) = {}^{0+}A + {}^{0+}B$.

Since ${}^{0+}A$ and ${}^{0+}B$ are bounded, so is ${}^{0+}(A + B)$. ..... (3)

Therefore, by definition of a fuzzy number and (1), (2), (3), we prove that $A + B$ is a fuzzy number.

Similarly, we can prove that $A - B$, AB and A / B (when defined) are fuzzy numbers.

Constructing $A + B$ from A, B

Let A and B be fuzzy numbers given below :

$$A : \mathbb{R} \to [0, 1] \; ; \quad A(x) = \begin{cases} 1, & \text{for } x \in [a_1', a_2'] \\ l_A(x), & \text{for } x \in [a_1, a_1') \\ 0, & \text{for } x \in (-\infty, a_1) \\ r_A(x), & \text{for } x \in (a_2', a_2] \\ 0, & \text{for } x \in (a_2, \infty) \end{cases}$$

$$B : \mathbb{R} \to [0, 1] \; ; \quad B(x) = \begin{cases} 1, & \text{for } x \in [b_1', b_2'] \\ l_B(x), & \text{for } x \in [b_1, b_1') \\ 0, & \text{for } x \in (-\infty, b_1) \\ r_B(x), & \text{for } x \in (b_2', b_2] \\ 0, & \text{for } x \in (b_2, \infty) \end{cases}$$

From the data observe that

${}^{1}A = [a_1', a_2']$, ${}^{1}B = [b_1', b_2']$, ${}^{0+}A \subseteq [a_1, a_2]$, ${}^{0+}B \subseteq [b_1, b_2]$

Step 1 : ${}^{1}(A + B) = {}^{1}A + {}^{1}B = [a_1' + b_1', a_2' + b_2']$

Therefore, $(A + B)(x) = 1$, for $x \in [a_1' + b_1', a_2' + b_2']$

Step 2 : ${}^{0+}(A + B) = {}^{0+}A + {}^{0+}B \subseteq [a_1 + b_1, a_2 + b_2]$ ..... (1)

Therefore, $(A + B)(x) = 0$, for $x < a_1 + b_1$ and $x > a_2 + b_2$

Step 3 : To obtain $A + B$ at other points observe that

$(A + B)(x) = l_{A+B}(x)$ for $x \in [a_1 + b_1, a_1' + b_1')$

and $(A + B)(x) = r_{A+B}(x)$ for $x \in (a_2' + b_2', a_2 + b_2]$.

Suppose $(A + B)(x) = \alpha$. Then

${}^{\alpha}(A + B) = {}^{\alpha}A + {}^{\alpha}B = [a_1^{\alpha} + b_1^{\alpha}, a_2^{\alpha} + b_2^{\alpha}]$ (say).

Set $a_1^{\alpha} + b_1^{\alpha} = x$. If $a_1^{\alpha} + b_1^{\alpha} = g_1(\alpha)$, say, then

$\alpha = g_1^{-1}(x) = g(x)$, say.

$\therefore$ $l_{A+B}(x) = g(x)$.

Similarly, set $x = a_2^{\alpha} + b_2^{\alpha} = h_1(\alpha)$, say.

Then $\alpha = h_1^{-1}(x) = h(x)$, then $r_{A+B}(x) = h(x)$.

It is not always possible to find $g(x)$ and $h(x)$, unless $a_1^{\alpha} + b_1^{\alpha} = x$ and $a_2^{\alpha} + b_2^{\alpha} = x$ are invertible.

For our purpose we deal with only simple invertible functions.

Definition : A fuzzy number A, is called a triangular fuzzy number if,

(i) There is $a \in \mathbb{R}$ such that $A(a) = 1$ and such a is unique.

(ii) l_A consists of line segment joining the points $(a_1, 0)$ and $(a, 1)$; and $l(x) = 0$ for $x \in (-\infty, a_1)$.

(iii) r_A consists of line segment joining $(a, 1)$ and $(a_2, 0)$ and $r(x) = 0$, for $x \in (a_2, \infty)$.

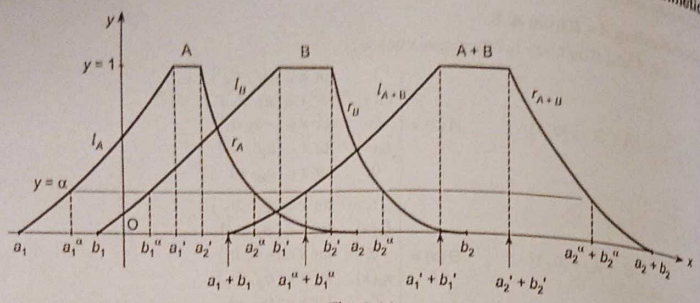


Fig. 9.14

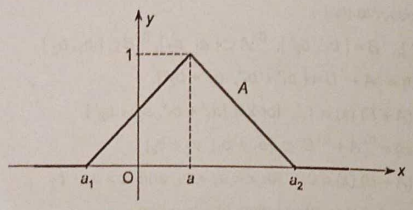


Fig. 9.15

We denote A by a triple (a_1, a, a_2). The algebraic equations for A are

$$A(x) = \begin{cases} 1 + \dfrac{x-a}{a-a_1}, & \text{for } a_1 < x \le a \\ 1 + \dfrac{x-a}{a-a_2}, & \text{for } a < x \le a_2 \\ 0, & \text{otherwise} \end{cases}$$

Theorem : If $A = (a_1, a, a_2)$ and $B = (b_1, b, b_2)$ are triangular numbers, then so are $A + B$ and $A - B$. Moreover

$$A + B = (a_1 + b_1,\ a + b,\ a_2 + b_2)$$
$$A - B = (a_1 - b_2,\ a + b,\ a_2 - b_1)$$

Example 1 : Find fuzzy number $A + B$ for fuzzy numbers

$$A = \begin{cases} \dfrac{x+3}{3}, & \text{for } -3 < x \le 0 \\ \dfrac{3-x}{3}, & \text{for } 0 \le x \le 3 \\ 0, & \text{otherwise} \end{cases} ; \quad B = \begin{cases} \dfrac{x-3}{3}, & \text{for } 3 < x \le 6 \\ \dfrac{9-x}{3}, & \text{for } 6 < x \le 9 \\ 0, & \text{otherwsie} \end{cases}$$ **(S.U. 2009)**

Sol. : Here $A = (-3, 0, 3)$ and $B = (3, 6, 9)$ are triangular fuzzy numbers.

Hence, $A + B = (-3 + 3, 0 + 6, 3 + 9) = (0, 6, 12)$ is a triangular number.

Therefore, l_{A+B} is line segment from $P_1(0, 0)$ to $P_2(6, 1)$ and $l_{A+B}(x) = 0$ elsewhere. The equation of the line P_1P_2 is

$$y = \frac{x}{6} \qquad \text{..... (1)}$$

r_{A+B} is the line segment from $P_2(6, 1)$ to $Q_2(12, 0)$ and $r_{A+B}(x) = 0$ elsewhere. The equation of P_2Q_2 is

$$y = 2 - \frac{x}{6} \qquad \text{..... (2)}$$

From (1) and (2), we have equation of $A + B$,

$$(A + B)(x) = \begin{cases} \dfrac{x}{6}, & \text{for } 0 \le x \le 6 \\ 2 - \dfrac{x}{6}, & \text{for } 6 \le x \le 12 \\ 0, & \text{everywhere else} \end{cases}$$

Example 2 : Find the fuzzy numbers $A + B$ and AB, if A and B are fuzzy numbers given below :

$$A(x) = \begin{cases} 0, & \text{for } x \le -2 \text{ and } x > 4 \\ \dfrac{x+2}{3}, & \text{for } -2 < x \le 1 \\ \dfrac{4-x}{3}, & \text{for } 1 < x \le 4 \end{cases} \quad \text{and} \quad B(x) = \begin{cases} 0, & \text{for } x \le 4 \text{ and } x > 10 \\ \dfrac{x-4}{3}, & \text{for } 4 < x \le 7 \\ \dfrac{10-x}{3}, & \text{for } 7 < x \le 10 \end{cases}$$ **(S.U. 2008)**

Sol. : Setting $\alpha = \dfrac{x+2}{3}$ yields $x = 3\alpha - 2$. Setting $\alpha = \dfrac{4-x}{3}$ yields $x = 4 - 3\alpha$.

$$\therefore \quad {}^{\alpha}A = (3\alpha - 2, 4 - 3\alpha).$$

Similarly, setting $\alpha = \dfrac{x-4}{3}$ and $\alpha = \dfrac{10-x}{3}$ give us

$${}^{\alpha}B = (3\alpha + 4, 10 - 3\alpha)$$

$$\therefore \quad {}^{\alpha}(A + B) = {}^{\alpha}A + {}^{\alpha}B = (6\alpha + 2, -6\alpha + 14).$$

Let $6\alpha + 2 = x$, then $\alpha = \dfrac{x-2}{6}$. Moreover as $\alpha \in (0, 1]$, we have $x \in (2, 8]$.

$$\therefore \quad A(x) = \alpha = \frac{x-2}{6}, \text{ for } x \in (2, 8]. \qquad \text{..... (1)}$$

Let $-6\alpha + 14 = x$, then $\alpha = \dfrac{14-x}{6}$. Moreover as $\alpha \in (0, 1]$, we have $x \in [8, 14)$.

$$\therefore \quad A(x) = \alpha = \frac{14-x}{6}, \text{ for } x \in [8, 14). \qquad \text{..... (2)}$$

Combining (1) and (2), we have

$$A(x) = \begin{cases} \dfrac{x-2}{6}, & \text{for } 2 < x \le 8 \\ \dfrac{14-x}{6}, & \text{for } 8 \le x < 14 \\ 0, & \text{otherwise} \end{cases}$$

To calculate AB : Observe that

(i) $3\alpha - 2$ is negative for $\alpha \in [0, 2/3)$ and non-negative for $\alpha \in [2/3, 1]$.

All other quantities appearing in expression for ${}^{\alpha}A$ and ${}^{\alpha}B$ are positive.

(ii) We also have $3\alpha + 4 \le -3\alpha + 10$. Therefore, for $\alpha \in [0, 2/3)$, we get

$$ {}^{\alpha}A\,{}^{\alpha}B = ((3\alpha - 2)(-3\alpha + 10), (-3\alpha + 4)(-3\alpha + 10)) $$

And, for $\alpha \in [2/3, 1]$, we have

$$ {}^{\alpha}A\,{}^{\alpha}B = ((3\alpha - 2)(3\alpha + 4), (-3\alpha + 4)(-3\alpha + 10)) $$

$$ \therefore \quad {}^{\alpha}(AB) = {}^{\alpha}A\,{}^{\alpha}B = (-9\alpha^2 + 36\alpha - 20, 9\alpha^2 - 42\alpha + 40) \quad \text{for } \alpha \in [0, 2/3) \quad \text{.....(1)} $$

$$ \text{and} \quad {}^{\alpha}AB = {}^{\alpha}A\,{}^{\alpha}B = (9\alpha^2 + 6\alpha - 8, 9\alpha^2 - 42\alpha + 40) \quad \text{for } \alpha \in [2/3, 1] \quad \text{.....(2)} $$

Consider the equation (1). Solving the equation

$$ -9\alpha^2 + 36\alpha - 20 = x \text{ for } \alpha, \text{ we get } \alpha = \frac{6 \pm \sqrt{16 - x}}{3}. $$

Since $\alpha \in [0, 2/3)$, we reject the plus sign and set $\alpha = \dfrac{6 - \sqrt{16 - x}}{3}$.

Moreover as α varies from 0 to 2/3, $-9\alpha^2 + 36\alpha - 20$ varies from -20 to 0. Therefore, we obtain

$$ AB(x) = \frac{6 - \sqrt{16 - x}}{3} \quad \text{for } -20 \le x < 0 \quad \text{.....(3)} $$

Similarly solving the equation $9\alpha^2 - 42\alpha + 40 = x$ for α and finding the variation of x as α varies from 0 to 2/3, we get

$$ AB(x) = \frac{7 - \sqrt{9 + x}}{3}, \quad \text{for } 16 < x \le 40 \quad \text{.....(4)} $$

Now, consider the equation (2).

Solving $9\alpha^2 + 6\alpha - 8 = x$ for α and considering variations in x as α varies from 2/3 to 1, we obtain

$$ AB(x) = \frac{-1 + \sqrt{9 + x}}{3}, \quad \text{for } 0 \le x \le 7 \quad \text{.....(5)} $$

Solving $9\alpha^2 - 42\alpha + 40 = x$ for α and considering the variation in x as α varies from 2/3 to 1, we get

$$ AB(x) = \frac{7 - \sqrt{9 + x}}{3}, \quad \text{for } 7 \le x \le 16 \quad \text{.....(6)} $$

Combining equations (3) to (6), we have

$$ AB(x) = \begin{cases} \dfrac{6 - \sqrt{16 - x}}{3}, & \text{for } -20 \le x < 0 \\ \dfrac{-1 + \sqrt{9 + x}}{3}, & \text{for } 0 \le x \le 7 \\ \dfrac{7 - \sqrt{9 + x}}{3}, & \text{for } 7 \le x \le 40 \\ 0, & \text{otherwise} \end{cases} $$

Example 3 : Calculate the fuzzy numbers $A + B$, $A - B$ and A/B, where

$$ A(x) = \begin{cases} \dfrac{x + 3}{3}, & -3 < x \le 0 \\ \dfrac{3 - x}{3}, & 0 < x \le 3 \\ 0, & \text{otherwise} \end{cases} \quad \text{and} \quad B(x) = \begin{cases} \dfrac{x - 3}{3}, & 3 < x \le 6 \\ \dfrac{9 - x}{3}, & 6 < x \le 9 \\ 0, & \text{otherwise} \end{cases} $$

(S.U. 2011)

Sol. : Setting $\alpha = \dfrac{x + 3}{3}$, yields $x = 3\alpha - 3$. Setting $\alpha = \dfrac{3 - x}{3}$, yields $x = 3 - 3\alpha$.

$$ \therefore \quad {}^{\alpha}A = [3\alpha - 3, -3\alpha + 3] \quad \text{.....(1)} $$

Similarly, $\alpha = \dfrac{x - 3}{3}$ and $\alpha = \dfrac{9 - x}{3}$ give

$$ {}^{\alpha}B = [3\alpha + 3, -3\alpha + 9] \quad \text{.....(2)} $$

To calculate $A + B$:

We have ${}^{\alpha}(A + B) = {}^{\alpha}A + {}^{\alpha}B = [6\alpha, -6\alpha + 12]$ where $\alpha \in (0, 1]$.

Let $6\alpha = x$. Since $\alpha \in (0, 1]$

We have $x \in (0, 6]$.

$$ \therefore \quad (A + B)(x) = \alpha = \frac{x}{6}, \text{for } 0 < x \le 6 \quad \text{.....(3)} $$

Let $-6\alpha + 12 = x$. Since $\alpha \in (0, 1]$, we have $x \in [6, 12)$.

$$ \therefore \quad (A + B)(x) = \alpha = \frac{12 - x}{6}, \quad \text{for } 6 \le x < 12 \quad \text{.....(4)} $$

Therefore, from (3) and (4),

$$ (A + B)(x) = \begin{cases} \dfrac{x}{6}, & \text{for } 0 < x \le 6 \\ \dfrac{12 - x}{6}, & \text{for } 6 < x < 12 \\ 0, & \text{otherewise} \end{cases} $$

To calculate $A - B$:

$$ {}^{\alpha}(A - B) = {}^{\alpha}A - {}^{\alpha}B = [6\alpha - 12, -6\alpha] \text{ where } \alpha \in (0, 1]. $$

Let $6\alpha - 12 = x$. Since $\alpha \in (0, 1]$, we have $x \in (-12, -6]$.

$$ \therefore \quad (A - B)(x) = \alpha = \frac{x + 12}{6}, \quad \text{for } x \in (-12, -6] \quad \text{.....(5)} $$

Let $-6\alpha = x$. Then we have $x \in [-6, 0)$.

$$ \therefore \quad (A - B)(x) = \alpha = -\frac{x}{6}, \quad \text{for } x \in [-6, 0) \quad \text{.....(6)} $$

Combining (5) and (6), we have

$$ (A - B)(x) = \begin{cases} \dfrac{x + 12}{6}, & \text{for } -12 < x \le -6 \\ -\dfrac{x}{6}, & \text{for } -6 \le x < 0 \\ 0, & \text{othwerwise} \end{cases} $$

To calculate A / B :

Observe that $3\alpha - 3 \le 0$ for $\alpha \le (0, 1]$. Other entities appearing in the expressions of ${}^{\alpha}A, {}^{\alpha}B$ are positive, moreover $3\alpha + 3 \le -3\alpha + 9$ for $\alpha \in (0, 1]$. Therefore, we have

$$ {}^{\alpha}(A/B) = \frac{{}^{\alpha}A}{{}^{\alpha}B} = [3\alpha - 3, -3\alpha + 3] \times \left[\frac{1}{-3\alpha + 9}, \frac{1}{3\alpha + 3}\right] $$

$$ = \left[\frac{3\alpha - 3}{3\alpha + 3}, \frac{-3\alpha + 3}{3\alpha + 3}\right] $$

Let $\dfrac{3\alpha - 3}{3\alpha + 3} = x$. Since $\alpha \in (0, 1]$ we have $x \in (-1, 0]$.

$$ \therefore \quad (A/B)(x) = \alpha = \frac{-(x+1)}{(x-1)}, \quad \text{for } -1 < x \le 0 \qquad \text{..... (7)} $$

Let $\dfrac{-3\alpha + 3}{3\alpha + 3} = x$. Since $\alpha \in (0, 1]$ we have $x \in [0, 1)$.

$$ \therefore \quad (A/B)(x) = \alpha = \frac{1-x}{1+x}, \quad \text{for } 0 \le x < 1 \qquad \text{..... (8)} $$

Hence, combining (7) and (8), we find

$$ (A/B)(x) = \begin{cases} \dfrac{-(x+1)}{x-1}, & \text{for } -1 < x \le 0 \\ \dfrac{1-x}{1+x}, & \text{for } 0 \le x < 1 \\ 0, & \text{otherwise} \end{cases} $$

Example 4 : Given a fuzzy number $A(x) = \begin{cases} 1-(x-2)^2, & \text{for } 1 \le x \le 3 \\ 0, & \text{otherwsie} \end{cases}$

Find A^{-1}. Verify that $AA^{-1} \ne 1$.

Sol. : Setting $\alpha = 1 - (x-2)^2$, we get $x = 2 \pm \sqrt{1-\alpha}$. Hence,

$$ {}^{\alpha}A = [2 - \sqrt{1-\alpha}, 2 + \sqrt{1-\alpha}] $$

$$ \therefore \quad {}^{\alpha}(A^{-1}) = \frac{1}{{}^{\alpha}A} = \left[\frac{1}{2+\sqrt{1-\alpha}}, \frac{1}{2-\sqrt{1-\alpha}}\right] $$

Let $\dfrac{1}{2+\sqrt{1-\alpha}} = x$. Since $\alpha \in (0, 1]$, we have $x \in \left(\dfrac{1}{3}, \dfrac{1}{2}\right]$.

$$ \therefore \quad A^{-1}(x) = \alpha = 1 - \left(\frac{1-2x}{x}\right)^2, \quad \text{for } \frac{1}{3} < x \le \frac{1}{2} \qquad \text{..... (1)} $$

Let $\dfrac{1}{2-\sqrt{1-\alpha}} = x$. Since $\alpha \in (0, 1]$, we have $x \in \left[\dfrac{1}{2}, 1\right)$.

$$ \therefore \quad A^{-1}(x) = \alpha = 1 - \left(\frac{1-2x}{x}\right)^2, \quad \text{for } \frac{1}{2} \le x < 1 \qquad \text{..... (2)} $$

Combining (1) and (2), we find

$$ A^{-1}(x) = \begin{cases} 1 - \left(\dfrac{1-2x}{x}\right)^2, & \text{for } \dfrac{1}{3} < x < 1 \\ 0, & \text{otherwise} \end{cases} $$

To compute AA^{-1} :

Consider

$$ {}^{\alpha}(AA^{-1}) = {}^{\alpha}A\,{}^{\alpha}(A^{-1}) $$

$$ = \left[2 - \sqrt{1-\alpha}, 2 + \sqrt{1-\alpha}\right]\left[\frac{1}{2+\sqrt{1-\alpha}}, \frac{1}{2-\sqrt{1-\alpha}}\right] $$

$$ = \left[\frac{2-\sqrt{1-\alpha}}{2+\sqrt{1-\alpha}}, \frac{2+\sqrt{1-\alpha}}{2-\sqrt{1-\alpha}}\right] $$

Let $\dfrac{2-\sqrt{1-\alpha}}{2+\sqrt{1-\alpha}} = x$. Since $\alpha \in (0, 1]$, we find $x \in (1/3, 1]$.

$$ \therefore \quad (AA^{-1})(x) = \alpha = 1 - 4\left(\frac{x-1}{x+1}\right)^2, \quad \text{for } \frac{1}{3} < x \le 1 \qquad \text{..... (3)} $$

Let $\dfrac{2+\sqrt{1-\alpha}}{2-\sqrt{1-\alpha}} = x$. Since $\alpha \in (0, 1]$, we find $x \in [1, 3)$.

Hence, $(AA^{-1})(x) = \alpha = 1 - 4\left(\dfrac{x-1}{x+1}\right)^2$, for $1 \le x < 3$ ..... (4)

From (3) and (4),

thus, $(AA^{-1})(x) = \begin{cases} 1 - 4\left(\dfrac{x-1}{x+1}\right)^2, & \text{for } \dfrac{1}{3} < x \le 3 \\ 0, & \text{otherwise} \end{cases}$

Clearly AA^{-1} is not a fuzzy set of real number 1.

Example 5 : Calculate $A + B$, if A and B are fuzzy number given below :

$$ A(x) = \begin{cases} \dfrac{x+1}{2}, & \text{for } -1 \le x \le 1 \\ 1, & \text{for } 1 \le x \le 2 \\ 3 - x, & \text{for } 2 \le x \le 3 \\ 0, & \text{othwerwise} \end{cases} \quad \text{and} \quad B(x) = \begin{cases} \dfrac{x+1}{2}, & \text{for } -1 \le x \le 1 \\ \dfrac{3-x}{2}, & \text{for } 1 \le x \le 3 \\ 0, & \text{otherwise} \end{cases} $$

Sol. : Let $\alpha = \dfrac{x+1}{2}$. Then $x = 2\alpha - 1$. Similarly, $\alpha = 3 - x$ yields, $x = 3 - \alpha$. Therefore, we have

$$ {}^{\alpha}A = [2\alpha - 1, -\alpha + 3], \quad \text{for } \alpha \in (0, 1] \qquad \text{..... (1)} $$

We note, ${}^{1}A = [1, 2]$ ..... (2)

Similarly, setting $\alpha = \dfrac{x+1}{2}$ and $\alpha = \dfrac{3-x}{2}$ and moving on familiar lines, we get

$$ {}^{\alpha}B = [2\alpha - 1, -2\alpha + 3], \quad \text{for } \alpha \in (0, 1] \qquad \text{..... (3)} $$

We note, ${}^{1}B = [1, 1]$ ..... (4)

Therefore, from (1) and (3), we have

$$ {}^{\alpha}(A + B) = [4\alpha - 2, -3\alpha + 6], \quad \text{for } \alpha \in (0, 1] $$

And from (2) and (4), we have

$$ {}^{1}(A + B) = [2, 3]. $$

Let $4\alpha - 2 = x$. Since $\alpha \in (0, 1]$, we have $x \in (-2, 2]$. Therefore,

$$ (A + B)(x) = \alpha = \frac{x+2}{4}, \quad \text{for } x \in (-2, 2] \qquad \text{..... (5)} $$

Let $-3\alpha + 6 = x$. Since $\alpha \in (0, 1]$, we have $x \in [3, 6)$. Therefore,

$$(A + B)(x) = \alpha = \frac{6 - x}{3}, \quad \text{for } x \in [3, 6) \qquad \text{.....(6)}$$

Since, ${}^1(A + B) = [2, 3]$, we have

$$(A + B)(x) = 1, \quad \text{for } x \in [2, 3] \qquad \text{.....(7)}$$

Combing (5), (6) and (7), we find

$$(A + B)(x) = \begin{cases} \dfrac{x+2}{4}, & \text{for } -2 < x \le 2 \\ 1, & \text{for } 2 \le x \le 3 \\ \dfrac{6-x}{3}, & \text{for } 3 \le x < 6 \\ 0, & \text{everywhere else} \end{cases}$$

Example 6 : Compute $A + B$, $A - B$ and AB, if A and B are fuzzy numbers defined below :

$$A(x) = \begin{cases} 1-(x+1)^2, & \text{for } -2 \le x \le -1 \\ 1, & \text{for } -1 \le x \le 1 \\ 1-(x-1)^2, & \text{for } 1 \le x \le 2 \\ 0, & \text{elsewhere} \end{cases} \quad \text{and} \quad B(x) = \begin{cases} 1-\left(\dfrac{x-2}{2}\right)^2, & \text{for } 0 \le x \le 2 \\ 1, & \text{for } 2 \le x \le 3 \\ 1-\left(\dfrac{x-3}{4}\right)^2, & \text{for } 3 \le x \le 7 \end{cases}$$

Sol. : To calculate $A + B$

Let $\alpha = 1 - (x+1)^2$. Then $x = -1 \pm \sqrt{1-\alpha}$.

Since $\alpha \in [0, 1]$ and $x \in [-2, -1]$, we choose $x = -1 - \sqrt{1-\alpha}$. .....(1)

Let $\alpha = 1 - (x-1)^2$. Then $x = 1 \pm \sqrt{1-\alpha}$.

Since $\alpha \in [0, 1]$ and $x \in [1, 2]$, we choose $x = 1 + \sqrt{1-\alpha}$. .....(2)

From (1) and (2), we obtain for $\alpha \in (0, 1)$.

$${}^\alpha A = [-1-\sqrt{1-\alpha}, 1+\sqrt{1-\alpha}] \qquad \text{.....(3)}$$

Moreover, ${}^1A = [-1, 1]$ .....(4)

Let $\alpha = 1 - \left(\dfrac{x-2}{2}\right)^2$, then $x = 2 \pm 2\sqrt{1-\alpha}$.

Since $\alpha \in [0, 1]$ and $x \in [0, 2]$, we choose $x = 2 - 2\sqrt{1-\alpha}$. .....(5)

Similarly setting $\alpha = 1 - \left(\dfrac{x-3}{4}\right)^2$, we obtain $x = 3 + 4\sqrt{1-\alpha}$. .....(6)

From (5) and (6), we obtain

$${}^\alpha B = [2-2\sqrt{1-\alpha}, 3+4\sqrt{1-\alpha}] \qquad \text{.....(7)}$$

Moreover ${}^1B = [2, 3]$ .....(8)

From (3) and (7), we obtain

$${}^\alpha(A+B) = [1-3\sqrt{1-\alpha}, 4+5\sqrt{1-\alpha}], \quad \alpha \in (0, 1) \qquad \text{.....(9)}$$

From (4) and (8) or otherwise

$${}^1(A+B) = [1, 4] \qquad \text{.....(10)}$$

Let $1 - 3\sqrt{1-\alpha} = x$. Since $\alpha \in (0, 1]$, we have $x \in (-2, 1]$.

$$\therefore \quad (A + B)(x) = \alpha = 1 - \left(\frac{1-x}{3}\right)^2, \quad \text{for } -2 < x \le 1 \qquad \text{.....(11)}$$

Let $4 + 5\sqrt{1-\alpha} = x$, since $\alpha \in (0, 1]$, we have $x \in [4, 9)$.

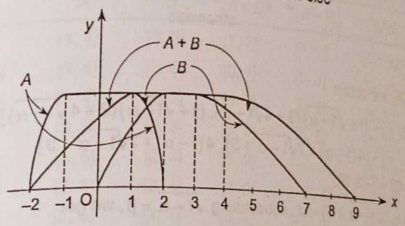
$$\therefore \quad (A + B)(x) = \alpha = 1 - \left(\frac{x-4}{5}\right)^2, \quad \text{for } 4 \le x < 9 \qquad \text{.....(12)}$$

Since, ${}^1(A + B) = [1, 4]$, we have

$$(A + B)(x) = 1, \quad \text{for } 1 \le x \le 4 \qquad \text{.....(13)}$$

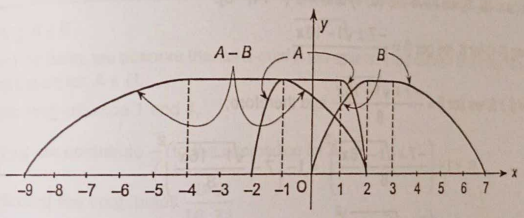
Combining (11), (12) and (13), we find

$$(A + B)(x) = \begin{cases} 1-\left(\dfrac{1-x}{3}\right)^2, & \text{for } -2 < x \le 1 \\ 1, & \text{for } 1 \le x \le 4 \\ 1-\left(\dfrac{x-4}{5}\right)^2, & \text{for } 4 \le x < 9 \\ 0, & \text{everywhere else} \end{cases}$$



$A + B$

Fig. 9.16



$A - B$

Fig. 9.17

To obtain $A - B$:

From (3) and (7), we obtain

$${}^\alpha(A-B) = {}^\alpha A - {}^\alpha B = [-4-5\sqrt{1-\alpha}, -1+3\sqrt{1-\alpha}]$$

We note ${}^1(A-B) = [-4, -1]$.

Let $-4 - 5\sqrt{1-\alpha} = x$. Since $\alpha \in (0, 1]$, we find $x \in (-9, -4]$.

Solving $-4-5\sqrt{1-\alpha} = x$ for α, we get $\alpha = 1-\left(\frac{x+4}{5}\right)^2$.

$$\therefore \quad (A-B)(x) = \alpha = 1-\left(\frac{x+4}{5}\right)^2, \quad \text{for } -9 < x \le -4$$

Let $-1+3\sqrt{1-\alpha} = x$, since $\alpha \in (0, 1]$, we get $x \in [-1, 2)$.

Solving $-1+3\sqrt{1-\alpha} = x$, for α, we get $\alpha = 1-\left(\frac{x+1}{3}\right)^2$.

$$\therefore \quad (A-B)(x) = \alpha = 1-\left(\frac{x+1}{3}\right)^2, \quad \text{for } -1 \le x < 2$$

$$\text{Therefore,} \quad (A-B)(x) = \begin{cases} 1-\left(\frac{x+4}{5}\right)^2, & \text{for } -9 < x \le -4 \\ 1, & \text{for } -4 \le x \le -1 \\ 1-\left(\frac{x+1}{3}\right)^2, & \text{for } -1 \le x < 2 \\ 0, & \text{otherwise} \end{cases}$$

To calculate AB :

From (3) and (7), we obtain

$$^{\alpha}(AB) = [-(1+\sqrt{1-\alpha})(3+4\sqrt{1-\alpha}), (1+\sqrt{1-\alpha})(3+4\sqrt{1-\alpha})]$$
$$= [-(4(1-\alpha)+7\sqrt{1-\alpha}+3), 4(1-\alpha)+7\sqrt{1-\alpha}+3]$$

In particular, $^{1}AB = [-3, 3]$.

Let $-(4(1-\alpha)+7\sqrt{1-\alpha}+3) = x$, substituting $\sqrt{1-\alpha} = \beta$, we get

$$-(4\beta^2+7\beta+3) = x$$

or $$4\beta^2+7\beta+(3+x) = 0 \quad \text{.....(14)}$$

Since, $\alpha \in (0, 1]$, we have $\beta \in [0, 1)$ and $x \in (-14, -3]$.

Solving (14) for β, we get $\beta = \frac{-7 \pm \sqrt{1-16x}}{8}$.

Since $\beta \ge 0$, we set $\beta = \frac{-7+\sqrt{1-16x}}{8}$ and therefore,

$$\alpha = 1-\left(\frac{-7+\sqrt{1-16x}}{8}\right)^2 = 1-\left(7-\frac{\sqrt{1-16x}}{8}\right)^2$$

$$\therefore \quad AB(x) = 1-\left(\frac{7-\sqrt{1-16x}}{8}\right)^2, \quad \text{for } -14 \le x < -3 \quad \text{.....(15)}$$

Letting $4(1-\alpha)+7\sqrt{1-\alpha}+3 = x$ and running on parallel lines, we obtain

$$\therefore \quad AB(x) = 1-\left(\frac{7-\sqrt{1+16x}}{8}\right)^2, \quad \text{for } 3 \le x < 14 \quad \text{.....(16)}$$

Combining (15), (16) and taking into account $^{1}AB = [-3, 3]$, we, finally, have

$$AB(x) = \begin{cases} 1-\left(\frac{7-\sqrt{1-16x}}{8}\right)^2, & \text{for } -14 \le x < -3 \\ 1, & \text{for } -3 \le x \le 3 \\ 1-\left(\frac{7-\sqrt{1+16x}}{8}\right)^2, & \text{for } 3 \le x < 14 \\ 0, & \text{everywhere else} \end{cases}$$

Example 7 : Let $A = \frac{0{\cdot}4}{[2, 3)} + \frac{0{\cdot}9}{[3, 4)} + \frac{1}{4} + \frac{0{\cdot}7}{(4, 5]} + \frac{0{\cdot}1}{(5, 6]}$

$B = \frac{0{\cdot}3}{[1, 2)} + \frac{0{\cdot}8}{[2, 3)} + \frac{1}{3} + \frac{0{\cdot}6}{(3, 4]} + \frac{0{\cdot}2}{(4, 5]}$

Compute $A+B$, $A-B$, AB, A/B.

Sol. : We prepare the following table of α-cuts.

	1	2	3	4	5	6	7
	α	$^{\alpha}A$	$^{\alpha}B$	$^{\alpha}(A+B)$	$^{\alpha}(A-B)$	$^{\alpha}(AB)$	$^{\alpha}(A/B)$
1	1	[4, 4]	[3, 3]	[7, 7]	[1, 1]	[12, 12]	[4/3, 4/3]
2	0·9	[3, 4]	[3, 3]	[6, 7]	[0, 1]	[9, 12]	[1, 4/3]
3	0·8	[3, 4]	[2, 3]	[5, 7]	[0, 2]	[6, 12]	[1, 2]
4	0·7	[3, 5]	[2, 3]	[5, 8]	[0, 3]	[6, 15]	[1, 5/2]
5	0·6	[3, 5]	[2, 4]	[5, 9]	[−1, 3]	[6, 20]	[3/4, 5/2]
6	0·4	[2, 5]	[2, 4]	[4, 9]	[−2, 3]	[4, 20]	[1/2, 5/2]
7	0·3	[2, 5]	[1, 4]	[3, 9]	[−2, 4]	[2, 20]	[1/2, 5]
8	0·2	[2, 5]	[1, 5]	[3, 10]	[−3, 4]	[2, 25]	[2/5, 5]
9	0·1	[2, 6]	[1, 5]	[3, 11]	[−3, 5]	[2, 30]	[2/5, 6]

Computing $A+B$:

From the table, we observe that all α-cuts in the column 4 are distinct and therefore, contribute to the expression for $A+B$.

Comparing columns 1 and 4,

1. First row contribute $\frac{1}{7}$ to the expression for A.
2. Second row contributes $\frac{0{\cdot}9}{[6, 7)}$.
3. In the third row, we observe $[5, 7] = [5, 6) \cup [6, 7]$, we already accounted for $[6, 7]$, so we get the contribution $\frac{0{\cdot}8}{[5, 6)}$.
4. Since $[5, 8] = [5, 7] \cup (7, 8]$, the 4th row contributes $\frac{0{\cdot}7}{(7, 8]}$.
5. Since $[5, 9] = [5, 8] \cup (8, 9]$, 5th row contributes $\frac{0{\cdot}6}{(8, 9]}$.

6. Since $[4, 9] = [4, 5) \cup [5, 9]$, 6th row contributes $\frac{0{\cdot}4}{[4, 5)}$.

7. Since $[3, 9] = [3, 4) \cup [4, 9]$, we have $\frac{0{\cdot}3}{[3, 4)}$.

8. Since $[3, 10] = [3, 9] \cup (9, 10]$, we have $\frac{0{\cdot}2}{(9, 10]}$.

9. Since, $[3, 11] = [3, 10] \cup (10, 11]$, we have contribution $\frac{0{\cdot}1}{(10, 11]}$.

From (1) through (9), we find

$$A + B = \frac{0{\cdot}3}{[3, 4)} + \frac{0{\cdot}4}{[4, 5)} + \frac{0{\cdot}8}{[5, 6)} + \frac{0{\cdot}9}{[6, 7)} + \frac{1}{7} + \frac{0{\cdot}7}{(7, 8]} + \frac{0{\cdot}6}{(8, 9]} + \frac{0{\cdot}2}{(9, 10]} + \frac{0{\cdot}1}{(10, 11]}.$$

Similarly, a reader will see from the columns 1 and 5, that

$$A - B = \frac{0{\cdot}2}{[-3, -2)} + \frac{0{\cdot}4}{[-2, -1)} + \frac{0{\cdot}6}{[-1, 0)} + \frac{0{\cdot}9}{[0, 1)} + \frac{1}{1} + \frac{0{\cdot}8}{(1, 2]} + \frac{0{\cdot}7}{(2, 3]} + \frac{0{\cdot}3}{(3, 4]} + \frac{0{\cdot}1}{(4, 5]}.$$

Similarly, from the columns 1 and 6, we get

$$AB = \frac{0{\cdot}3}{[2, 4)} + \frac{0{\cdot}4}{[4, 6)} + \frac{0{\cdot}8}{[6, 9)} + \frac{0{\cdot}9}{[9, 12)} + \frac{1}{12} + \frac{0{\cdot}7}{(12, 15]} + \frac{0{\cdot}6}{(15, 20]} + \frac{0{\cdot}2}{(20, 25]} + \frac{0{\cdot}1}{(25, 30]}$$

From the columns 1 and 7, we compute

$$A / B = \frac{0{\cdot}2}{\left[\frac{2}{5}, \frac{1}{2}\right)} + \frac{0{\cdot}4}{\left[\frac{1}{2}, \frac{3}{4}\right)} + \frac{0{\cdot}6}{\left[\frac{3}{4}, 1\right)} + \frac{0{\cdot}9}{\left[1, \frac{4}{3}\right)} + \frac{1}{\left(\frac{4}{3}\right)} + \frac{0{\cdot}8}{\left(\frac{4}{3}, 2\right]} + \frac{0{\cdot}7}{\left(2, \frac{5}{2}\right]} + \frac{0{\cdot}3}{\left(\frac{5}{2}, 5\right]} + \frac{0{\cdot}1}{(5, 6]}.$$

EXERCISE - III

1. Find the fuzzy numbers AB and $A - B$, where

$$A(x) = \begin{cases} \frac{x}{4}, & \text{for } 0 < x \le 4 \\ \frac{8 - x}{4}, & \text{for } 4 < x \le 8 \\ 0, & \text{otherwise} \end{cases} \quad \text{and} \quad B(x) = \begin{cases} \frac{x - 4}{4}, & \text{for } 4 < x \le 8 \\ \frac{12 - x}{4}, & \text{for } 8 < x \le 12 \\ 0, & \text{elsewhere} \end{cases}$$ (S.U. 2012)

[Ans. : $$AB(x) = \begin{cases} \frac{-2 + \sqrt{4 + x}}{4}, & \text{for } 0 \le x \le 32 \\ \frac{10 - \sqrt{4 + x}}{4}, & \text{for } 32 \le x \le 96 \\ 0, & \text{otherwise} \end{cases}$$

$$(A - B)(x) = \begin{cases} \frac{x + 12}{8}, & \text{for } -12 \le x \le -4 \\ \frac{4 - x}{8}, & \text{for } -4 \le x \le 4 \\ 0, & \text{otherwise} \end{cases}$$

2. Calculate the fuzzy numbers $A + B$, $A - B$, A / B, where

$$A(x) = \begin{cases} \frac{x - 1}{3}, & \text{for } 1 < x \le 4 \\ \frac{7 - x}{3}, & \text{for } 4 < x \le 7 \\ 0, & \text{elsewhere} \end{cases} \quad \text{and} \quad B(x) = \begin{cases} \frac{x - 7}{3}, & \text{for } 7 \le x \le 10 \\ \frac{13 - x}{3}, & \text{for } 10 \le x \le 13 \\ 0, & \text{elsewhere} \end{cases}$$ (S.U. 2011)

[**Ans. :** $$(A + B)(x) = \begin{cases} \frac{x - 8}{6}, & \text{for } 8 \le x \le 14 \\ \frac{20 - x}{6}, & \text{for } 14 < x \le 20 \\ 0, & \text{elsewhere} \end{cases}$$

$$(A - B)(x) = \begin{cases} \frac{x + 12}{6}, & \text{for } -12 < x \le -6 \\ \frac{-x}{6}, & \text{for } -6 \le x \le 0 \\ 0, & \text{elsewhere} \end{cases}$$

$$(A / B)(x) = \begin{cases} \frac{13x - 1}{3(1 + x)}, & \text{for } \frac{1}{13} \le x \le \frac{4}{10} \\ \frac{7(1 - x)}{3(x + 1)}, & \text{for } \frac{4}{10} \le x \le 1 \\ 0, & \text{elsewhere} \end{cases}$$

3. Compute $A + B$, $A - B$, AB, if A and B are fuzzy numbers given below:

$$A(x) = \begin{cases} x^2, & \text{for } 0 \le x \le 1 \\ 1, & \text{for } 1 \le x \le 2 \\ (x - 3)^2, & \text{for } 2 \le x \le 3 \\ 0, & \text{elsewhere} \end{cases} \quad \text{and} \quad B(x) = \begin{cases} (x + 3)^2, & \text{for } -3 \le x \le -2 \\ 1, & \text{for } -2 \le x \le -1 \\ x^2, & \text{for } -1 \le x \le 0 \\ 0, & \text{elsewhere} \end{cases}$$

[**Ans. :** $$(A + B)(x) = \begin{cases} \left(\frac{x + 3}{2}\right)^2, & \text{for } -3 \le x \le -1 \\ 1, & \text{for } -1 \le x \le 1 \\ \left(\frac{3 - x}{2}\right)^2, & \text{for } 1 \le x \le 3 \\ 0, & \text{elsewhere} \end{cases}$$

$$(A - B)(x) = \begin{cases} \left(\frac{x}{2}\right)^2, & \text{for } 0 < x \le 2 \\ 1, & \text{for } 2 \le x \le 4 \\ \left(\frac{x - 6}{2}\right)^2, & \text{for } 4 \le x < 6 \\ 0, & \text{otherwise} \end{cases}$$

$$(AB)(x) = \begin{cases} (3 - \sqrt{-x})^2, & \text{for } -9 \le x \le -4 \\ 1, & \text{for } -4 \le x \le -1 \\ -x, & \text{for } -1 \le x \le 0 \\ 0, & \text{elsewhere} \end{cases}$$

4. Let C be the fuzzy number defined by

$$C(x) = \begin{cases} (x-1)^2, & \text{for } 1 \le x \le 2 \\ 1, & \text{for } 2 \le x \le 3 \\ (x-4)^2, & \text{for } 3 \le x \le 4 \\ 0, & \text{otherwise} \end{cases}$$

Let A and B be the fuzzy numbers given in Ex. 2, above. Find A/C and B/C.

[**Ans.** : $$(A/C)(x) = \begin{cases} \left(\dfrac{4x}{x+1}\right)^2, & \text{for } 0 \le x \le \dfrac{1}{3} \\ 1, & \text{for } \dfrac{1}{3} \le x \le 1 \\ \left(\dfrac{x-3}{x+1}\right)^2, & \text{for } 1 \le x \le 3 \\ 0, & \text{otherwise} \end{cases}$$

$$(B/C)(x) = \begin{cases} \left(\dfrac{x+3}{x-1}\right)^2, & \text{for } -3 \le x \le -1 \\ 1, & \text{for } -1 \le x \le -\dfrac{1}{3} \\ \left(\dfrac{4x}{x-1}\right)^2, & \text{for } -\dfrac{1}{3} \le x \le 0 \\ 0, & \text{otherwise} \end{cases}$$

5. Compute A/B, if A and B are fuzzy numbers given below :

$$A(x) = \begin{cases} 1-(x-2)^2, & \text{for } 1 \le x \le 2 \\ 1, & \text{for } 2 \le x \le 3 \\ 1-\left(\dfrac{x-3}{4}\right)^2, & \text{for } 3 \le x \le 7 \\ 0, & \text{otherwise} \end{cases} \quad \text{and} \quad B(x) = \begin{cases} x-1, & \text{for } 1 \le x \le 2 \\ 3-x, & \text{for } 2 \le x \le 3 \\ 0, & \text{otherwise} \end{cases}$$

[**Ans.** : $$(A/B)(x) = \begin{cases} 1-\left(\dfrac{-1+\sqrt{1-8x(x-1)}}{2x}\right), & \text{for } \dfrac{1}{3} \le x \le 1 \\ 1, & \text{for } 1 \le x \le \dfrac{3}{2} \\ 1-\left(\dfrac{-2+\sqrt{4-x(3-2x)}}{x}\right), & \text{for } \dfrac{3}{2} \le x \le 7 \\ 0, & \text{otherwise} \end{cases}$$

6. If A and B are fuzzy numbers given as

$$A(x) = \begin{cases} \dfrac{x+1}{2}, & \text{for } -1 < x \le 1 \\ \dfrac{3-x}{2}, & \text{for } 1 < x \le 3 \\ 0, & \text{otherwise} \end{cases} \quad \text{and} \quad B(x) = \begin{cases} \dfrac{x-1}{2}, & \text{for } 1 < x \le 3 \\ \dfrac{5-x}{2}, & \text{for } 3 < x \le 5 \\ 0, & \text{otherwise} \end{cases}$$

Find the fuzzy numbers $A + B$ and $A - B$.

[**Ans.** : $$(A+B)(x) = \begin{cases} \dfrac{x}{4}, & \text{for } 0 < x \le 4 \\ \dfrac{8-x}{4}, & \text{for } 4 < x \le 8 \\ 0, & \text{elsewhere} \end{cases}$$

$$(A-B)(x) = \begin{cases} \dfrac{6+x}{4}, & \text{for } -6 < x \le -2 \\ \dfrac{2-x}{4}, & \text{for } -2 \le x \le 2 \\ 0, & \text{otherwise} \end{cases}$$

$(A - B)$ is the characteristic function of real number 2.]

7. Calculate $A \cdot B$ and A/B for fuzzy numbers A and B, given below :

$$A(x) = \begin{cases} \dfrac{x-1}{3}, & \text{for } 1 < x \le 4 \\ \dfrac{7-x}{3}, & \text{for } 4 < x \le 7 \\ 0, & \text{otherwise} \end{cases} \quad \text{and} \quad B(x) = \begin{cases} \dfrac{x-7}{3}, & \text{for } 7 < x \le 10 \\ \dfrac{13-x}{3}, & \text{for } 10 < x \le 13 \\ 0, & \text{otherwise} \end{cases}$$

[**Ans.** : $$AB(x) = \begin{cases} \dfrac{-4+\sqrt{9+x}}{3}, & \text{for } 7 \le x \le 40 \\ \dfrac{10-\sqrt{9+x}}{3}, & \text{for } 40 \le x \le 91 \\ 0, & \text{elsewhere} \end{cases}$$

$$(A/B)(x) = \begin{cases} \dfrac{13x-1}{3(x+1)}, & \text{for } \dfrac{1}{13} \le x \le \dfrac{4}{10} \\ \dfrac{7(1-x)}{3(x+1)}, & \text{for } \dfrac{4}{10} \le x \le 1 \\ 0, & \text{otherwise} \end{cases}$$

8. Let A and B be fuzzy numbers whose membership functions are given by

$$A(x) = \begin{cases} \dfrac{x+2}{2}, & \text{for } -2 < x \le 0 \\ \dfrac{2-x}{2}, & \text{for } 0 < x < 2 \\ 0, & \text{otherwise} \end{cases} \quad \text{and} \quad B(x) = \begin{cases} \dfrac{x-2}{2}, & \text{for } 2 < x \le 4 \\ \dfrac{6-x}{2}, & \text{for } 4 < x \le 6 \\ 0, & \text{otherwise} \end{cases}$$

Find $A + B$ and $A - B$.

[**Ans.** : $$(A+B)(x) = \begin{cases} \dfrac{x}{4}, & \text{for } 0 \le x \le 4 \\ \dfrac{8-x}{4}, & \text{for } 4 \le x \le 8 \\ 0, & \text{elsewhere} \end{cases}$$

$$(A-B)(x) = \begin{cases} \dfrac{x+8}{4}, & \text{for } -8 < x \le -4 \\ -\dfrac{x}{4}, & \text{for } -4 \le x < 0 \\ 0, & \text{elsewhere} \end{cases}$$

9. Let A and B be two fuzzy sets whose membership functions are given by

$$A(x) = \begin{cases} \dfrac{x+1}{2}, & \text{for } -1 < x \le 1 \\ \dfrac{3-x}{2}, & \text{for } 1 < x \le 3 \\ 0, & \text{otherwise} \end{cases} \quad \text{and} \quad B(x) = \begin{cases} \dfrac{5-x}{2}, & \text{for } 3 \le x < 5 \\ \dfrac{x-1}{2}, & \text{for } 1 \le x < 3 \\ 0, & \text{otherwise} \end{cases}$$

Calculate the fuzzy numbers AB and A/B.

[Ans. :

$$(AB)(x) = \begin{cases} \dfrac{3-\sqrt{4-x}}{2}, & \text{for } -5 < x < 0 \\ \dfrac{\sqrt{x+1}}{2}, & \text{for } 0 \le x < 3 \\ \dfrac{4-\sqrt{1+x}}{2}, & \text{for } 3 \le x \le 15 \\ 0, & \text{elsewhere} \end{cases}$$

$$(A/B)(x) = \begin{cases} \dfrac{-(1+x)}{2(x-1)}, & \text{for } -1 \le x \le 0 \\ \dfrac{5x+1}{2(x+1)}, & \text{for } 0 \le x \le \dfrac{1}{3} \\ \dfrac{3-x}{2(x+1)}, & \text{for } \dfrac{1}{3} \le x \le 3 \\ 0, & \text{elsewhere} \end{cases}$$

10. Let A and B be the fuzzy numbers whose membership functions are given by

$$A(x) = \begin{cases} \dfrac{x+2}{2}, & \text{for } -2 < x \le 0 \\ \dfrac{2-x}{2}, & \text{for } 0 \le x \le 2 \\ 0, & \text{otherwise} \end{cases} \quad \text{and} \quad B(x) = \begin{cases} \dfrac{x-2}{2}, & \text{for } 2 < x \le 4 \\ \dfrac{6-x}{2}, & \text{for } 4 < x \le 6 \\ 0, & \text{otherwise} \end{cases}$$

Find AB and A/B.

[Ans. :

$$(AB)(x) = \begin{cases} \dfrac{4-\sqrt{4-x}}{2}, & \text{for } -12 \le x \le 0 \\ \dfrac{4-\sqrt{4+x}}{2}, & \text{for } 0 \le x \le 12 \\ 0, & \text{elsewhere} \end{cases}$$

$$(A/B)(x) = \begin{cases} \dfrac{1+x}{1-x}, & \text{for } -1 \le x \le 0 \\ \dfrac{1-x}{1+x}, & \text{for } 0 \le x \le 1 \\ 0, & \text{otherwise} \end{cases}$$

11. Let A and B be fuzzy numbers given as below :

$$A(x) = \begin{cases} x-2, & \text{for } 2 \le x \le 3 \\ \dfrac{5-x}{2}, & \text{for } 3 \le x \le 5 \\ 0, & \text{otherwise} \end{cases} \quad \text{and} \quad B(x) = \begin{cases} \dfrac{x-3}{2}, & \text{for } 3 \le x \le 5 \\ 6-x, & \text{for } 5 \le x \le 6 \\ 0, & \text{otherwise} \end{cases}$$

Compute AB.

[Ans. :

$$AB = \begin{cases} \dfrac{-7+\sqrt{1+8x}}{4}, & \text{for } 6 \le x \le 15 \\ \dfrac{17-\sqrt{49+8x}}{4}, & \text{for } 15 \le x \le 30 \\ 0, & \text{otherwise} \end{cases}$$

12. Given a fuzzy number A, whose membership equation is

$$A(x) = \begin{cases} x-1, & \text{for } 1 \le x \le 2 \\ 3-x, & \text{for } 2 \le x \le 3 \\ 0, & \text{otherwise} \end{cases}$$

Find A^{-1} and AA^{-1}.

[Ans. :

$$A^{-1}(x) = \begin{cases} \dfrac{3x-1}{x}, & \text{for } \dfrac{1}{3} \le x \le \dfrac{1}{2} \\ \dfrac{1-x}{x}, & \text{for } \dfrac{1}{2} \le x \le 1 \\ 0, & \text{otherwise} \end{cases}$$

$$AA^{-1}(x) = \begin{cases} \dfrac{3x-1}{x+1}, & \text{for } \dfrac{1}{3} \le x \le 1 \\ \dfrac{3-x}{x+1}, & \text{for } 1 \le x \le 3 \\ 0, & \text{otherwise} \end{cases}$$

13. Compute fuzzy numbers $A+B$, $A-B$, AB and A/B, if

$$A = \frac{0{\cdot}1}{[0,1)} + \frac{0{\cdot}2}{[1,2)} + \frac{1}{2} + \frac{0{\cdot}3}{(2,3]} + \frac{0{\cdot}2}{(3,4]}$$

and

$$B = \frac{0{\cdot}1}{[2,5)} + \frac{1}{5} + \frac{0{\cdot}2}{(5,6]} + \frac{0{\cdot}1}{(6,7]}$$

[Ans. :

$$A+B = \frac{0{\cdot}1}{[2,6)} + \frac{0{\cdot}2}{[6,7)} + \frac{1}{7} + \frac{0{\cdot}3}{(7,8]} + \frac{0{\cdot}2}{(8,10]} + \frac{0{\cdot}1}{(10,11]}$$

$$A-B = \frac{0{\cdot}1}{[-7,-5)} + \frac{0{\cdot}2}{[-5,-3)} + \frac{1}{-3} + \frac{0{\cdot}3}{(-3,-2]} + \frac{0{\cdot}2}{(-2,-1]} + \frac{0{\cdot}1}{(-1,2]}$$

$$AB = \frac{0{\cdot}1}{[0,5)} + \frac{0{\cdot}2}{[5,10)} + \frac{1}{10} + \frac{0{\cdot}3}{(10,15]} + + \frac{0{\cdot}2}{(15,24]} + \frac{0{\cdot}1}{(24,28]}$$

$$A/B = \frac{0{\cdot}1}{\left[0,\frac{1}{6}\right)} + \frac{0{\cdot}2}{\left[\frac{1}{6},\frac{2}{5}\right)} + \frac{1}{\left(\frac{2}{5}\right)} + \frac{0{\cdot}3}{\left(\frac{2}{5},\frac{3}{5}\right]} + \frac{0{\cdot}2}{\left(\frac{3}{5},\frac{4}{5}\right]} + \frac{0{\cdot}1}{\left(\frac{4}{5},2\right]}$$

14. Complete Example 7.

6. Fuzzy Equations

Solving equations, *i.e.*, expressing unknown quantities in term of known, when there exist relationships between them, is one of the primary activities in mathematics. Here, we shall discuss two simple equations $A + X = B$ and $AX = B$, where A and B are known fuzzy numbers and X is unknown. Even in these simple forms of equation, few general methods are available. Fuzzy equations is quite an active field for research.

A + X = B

If ${}^{\alpha}A = [\, a_1^{\alpha}, a_2^{\alpha} \,]$ and ${}^{\alpha}B = [\, b_1^{\alpha}, b_2^{\alpha} \,]$, then the equation $A + X = B$ has a solution if and only if the following conditions hold :

(i) ${}^{\alpha}b_1 - {}^{\alpha}a_1 \le {}^{\alpha}b_2 - {}^{\alpha}a_2$ for all $\alpha \in (0, 1]$, and

(ii) $\alpha < \beta$ implies ${}^{\alpha}b_1 - {}^{\alpha}a_1 \le {}^{\beta}b_1 - {}^{\beta}a_1 \le {}^{\beta}b_2 - {}^{\beta}a_2 \le {}^{\alpha}b_2 - {}^{\alpha}a_2$.

In this case ${}^{\alpha}X = [\, {}^{\alpha}b_1 - {}^{\alpha}a_1, {}^{\alpha}b_2 - {}^{\alpha}a_2 \,]$ are the α-cuts of the solution.

Example 1 : Solve the equation $A + X = B$, where A and B are fuzzy numbers defined as follows :

$$A(x) = \begin{cases} \dfrac{x+2}{2}, & \text{for } -2 < x \le 0 \\ \dfrac{2-x}{2}, & \text{for } 0 < x \le 2 \\ 0, & \text{otherwise} \end{cases} \quad ; \quad B(x) = \begin{cases} \dfrac{x-2}{2}, & \text{for } 2 < x \le 4 \\ \dfrac{6-x}{2}, & \text{for } 4 < x \le 6 \\ 0, & \text{otherwise} \end{cases}$$

(S.U. 2009)

Sol. : Recall that, if ${}^{\alpha}A = [\, {}^{\alpha}a_1, {}^{\alpha}a_2 \,]$ and ${}^{\alpha}B = [\, {}^{\alpha}b_1, {}^{\alpha}b_2 \,]$, then

$$ {}^{\alpha}X = [\, {}^{\alpha}b_1 - {}^{\alpha}a_1, \ {}^{\alpha}b_2 - {}^{\alpha}a_2 \,]$$

are the α-cuts of a solution X of the equation (if solution exists).

Let $\alpha = \dfrac{x+2}{2}$, then $x = 2\alpha - 2$, similarly, let $\alpha = \dfrac{2-x}{2}$, then $x = -2\alpha + 2$.

$\therefore \quad {}^{\alpha}A = [\, 2\alpha - 2, \ -2\alpha + 2 \,]$

Similarly, $\quad {}^{\alpha}B = [\, 2\alpha + 2, \ -2\alpha + 6 \,]$

Then, $\quad {}^{\alpha}X = [\, 4, 4 \,]$, for all $\alpha \in (0, 1]$.

Define $X : \mathbb{R} \to [0, 1]$

$$X(x) = \begin{cases} 1, & \text{for } x \in [4, 4] \\ 0, & \text{otherwise} \end{cases}$$

This is the solution of the equation.

Note that this is fuzzy form of integer 4.

Example 2 : Solve the equation $A + X = B$ where A and B are the following fuzzy sets :

$$A = \frac{0{\cdot}2}{[0, 1)} + \frac{0{\cdot}6}{[1, 2)} + \frac{0{\cdot}8}{[2, 3)} + \frac{0{\cdot}9}{[3, 4)} + \frac{1}{4} + \frac{0{\cdot}5}{(4, 5]} + \frac{0{\cdot}1}{(5, 6]}$$

$$B = \frac{0{\cdot}1}{[0, 1)} + \frac{0{\cdot}2}{[1, 2)} + \frac{0{\cdot}6}{[2, 3)} + \frac{0{\cdot}7}{[3, 4)} + \frac{0{\cdot}8}{[4, 5)} + \frac{0{\cdot}9}{[5, 6)}$$

$$+ \frac{1}{6} + \frac{0{\cdot}5}{(6, 7]} + \frac{0{\cdot}4}{(7, 8]} + \frac{0{\cdot}2}{(8, 9]} + \frac{0{\cdot}1}{(9, 10]}$$

(S.U. 2010)

Sol. : Recall that, if ${}^{\alpha}A = [\, {}^{\alpha}a_1, {}^{\alpha}a_2 \,]$ and ${}^{\alpha}B = [\, {}^{\alpha}b_1, {}^{\alpha}b_2 \,]$, then

$$ {}^{\alpha}X = [\, {}^{\alpha}b_1 - {}^{\alpha}a_1, \ {}^{\alpha}b_2 - {}^{\alpha}a_2 \,]$$

are the α-cuts of the solution X. Therefore, we construct the following table of α-cuts.

α	${}^{\alpha}A$	${}^{\alpha}B$	${}^{\alpha}X = [\, {}^{\alpha}b_1 - {}^{\alpha}a_1, \ {}^{\alpha}b_2 - {}^{\alpha}a_2 \,]$
1	[4, 4]	[6, 6]	[2, 2]
0·9	[3, 4]	[5, 6]	[2, 2]
0·8	[2, 4]	[4, 6]	[2, 2]
0·7	[2, 4]	[3, 6]	[2, 2]
0·6	[1, 4]	[2, 6]	[1, 2]
0·5	[1, 5]	[2, 7]	[1, 2]
0·4	[1, 5]	[2, 8]	[1, 2]
0·3	[1, 5]	[2, 8]	[1, 3]
0·2	[0, 5]	[1, 9]	[1, 3]
0·1	[0, 6]	[0, 10]	[1, 4]
			[0, 4]

Recall that $X(x) = \text{Sup}\, \{ \alpha \mid x \in {}^{\alpha}X \}$.

Therefore, if $\alpha_1 > \alpha_2$ and ${}^{\alpha_1}X = {}^{\alpha_2}X$, we need to consider only ${}^{\alpha_1}X$ for the computation. Therefore, we consider ${}^{1}X, {}^{0{\cdot}7}X, {}^{0{\cdot}4}X, {}^{0{\cdot}2}X, {}^{0{\cdot}1}X$, only.

Clearly, ${}^{1}X = \{ 2 \}$, therefore in the "additive notation" for X, we have the term

$$\frac{1}{2} \qquad \text{..... (1)}$$

${}^{0{\cdot}7}X = [\, 1, 2 \,] = [\, 1, 2 \,) \cup \{ 2 \}$, hence the contribution for X is

$$\frac{0{\cdot}7}{[1, 2)} \qquad \text{..... (2)}$$

${}^{0{\cdot}4}X = [\, 1, 3 \,] = [\, 1, 2 \,] \cup (\, 2, 3 \,]$, hence the contribution for X is

$$\frac{0{\cdot}4}{(2, 3]} \qquad \text{..... (3)}$$

${}^{0{\cdot}2}X = [\, 1, 4 \,] = [\, 1, 3 \,] \cup (\, 3, 4 \,]$, hence the contribution for X is

$$\frac{0{\cdot}2}{(3, 4]} \qquad \text{..... (4)}$$

Finally, ${}^{0{\cdot}1}X = [\, 0, 4 \,] = [\, 0, 1) \cup [\, 1, 4 \,]$, hence we have the contribution

$$\frac{0{\cdot}1}{[0, 1)} \qquad \text{..... (5)}$$

Combining (1) to (5), we get, $X = \dfrac{0{\cdot}1}{[0, 1)} + \dfrac{0{\cdot}7}{[1, 2)} + \dfrac{1}{2} + \dfrac{0{\cdot}4}{(2, 3]} + \dfrac{0{\cdot}2}{(3, 4]}$.

Equation *AX = B*

Let A and B be fuzzy numbers, ${}^{\alpha}A = [\, {}^{\alpha}a_1, {}^{\alpha}a_2 \,]$ and ${}^{\alpha}B = [\, {}^{\alpha}b_1, {}^{\alpha}b_2 \,]$ for all $\alpha \in (a, 1]$. Then the equation $AX = B$ has a solution if and only if

(i) $\dfrac{{}^{\alpha}b_1}{{}^{\alpha}a_1} \le \dfrac{{}^{\alpha}b_2}{{}^{\alpha}a_2}$ for all $\alpha \in (0, 1]$, and (ii) $\alpha \le \beta$ implies $\dfrac{{}^{\alpha}b_1}{{}^{\alpha}a_1} \le \dfrac{{}^{\beta}b_1}{{}^{\beta}a_1} \le \dfrac{{}^{\beta}b_2}{{}^{\beta}a_2} \le \dfrac{{}^{\alpha}b_2}{{}^{\alpha}a_2}$.

If it exists, the solution X is given by

$$X(x) = \text{Sup}\{\alpha \mid x \in {}^{\alpha}X\}$$

where ${}^{\alpha}X$ is the solution of the interval equation

$${}^{\alpha}A\,{}^{\alpha}X = {}^{\alpha}B \quad \textit{i.e.,} \quad \text{if } {}^{\alpha}X = [{}^{\alpha}x_1, {}^{\alpha}x_2], \text{ then}$$

$${}^{\alpha}x_1 = \frac{{}^{\alpha}b_1}{{}^{\alpha}a_1} \text{ and } {}^{\alpha}x_2 = \frac{{}^{\alpha}b_2}{{}^{\alpha}a_2}.$$

Example 3 : Solve the equation $AX = B$, if A and B are fuzzy numbers given as

$$A(x) = \begin{cases} 0, & \text{for } x \le 3,\ x > 5 \\ x-3, & \text{for } 3 < x \le 4 \\ 5-x, & \text{for } 4 < x \le 5 \end{cases} \text{ and } B(x) = \begin{cases} 0, & \text{for } x \le 12,\ x > 32 \\ \dfrac{x-12}{8}, & \text{for } 12 < x \le 20 \\ \dfrac{32-x}{12}, & \text{for } 20 < x \le 32 \end{cases}$$

(S.U. 2012)

Sol. : Clearly ${}^{\alpha}A = [\alpha + 3, -\alpha + 5]$ and ${}^{\alpha}B = [8\alpha + 12, -12\alpha + 32]$.

If ${}^{\alpha}X = [{}^{\alpha}x_1, {}^{\alpha}x_2]$, then we have

$${}^{\alpha}x_1 = \frac{{}^{\alpha}b_1}{{}^{\alpha}a_1} = \frac{8\alpha + 12}{\alpha + 3}.$$

Similarly, $${}^{\alpha}x_2 = \frac{{}^{\alpha}b_2}{{}^{\alpha}a_2} = \frac{-12\alpha + 32}{-\alpha + 5}.$$

Setting $\dfrac{8\alpha + 12}{\alpha + 3} = x$, we get $\alpha = \dfrac{12 - 3x}{x - 8}$.

As α varies from 0 to 1, x varies from 4 to 5.

Setting $\dfrac{-12\alpha + 32}{-\alpha + 5} = x$, we get $\alpha = \dfrac{32 - 5x}{12 - x}$.

As α varies from 0 to 1, x varies from $\dfrac{32}{5}$ to 5.

Therefore, $$X = \begin{cases} \dfrac{12-3x}{x-8}, & \text{for } 4 < x \le 5 \\ \dfrac{32-5x}{12-x}, & \text{for } 5 \le x < \dfrac{32}{5} \\ 0, & \text{otherwise} \end{cases}$$

Example 4 : Solve $AX = B$, if A and B are fuzzy numbers given as

$$A = \frac{0{\cdot}4}{[2, 3)} + \frac{0{\cdot}9}{[3, 4)} + \frac{1}{4} + \frac{0{\cdot}7}{(4, 5]} + \frac{0{\cdot}1}{(5, 6]}$$

$$B = \frac{0{\cdot}3}{[2, 4)} + \frac{0{\cdot}4}{[4, 6)} + \frac{0{\cdot}8}{[6, 9)} + \frac{0{\cdot}9}{[9, 12)} + \frac{1}{12} + \frac{0{\cdot}7}{(12, 15]} + \frac{0{\cdot}6}{(15, 20]} + \frac{0{\cdot}2}{(20, 25]} + \frac{0{\cdot}1}{(25, 30]}$$

Sol. : Let ${}^{\alpha}X = [{}^{\alpha}x_1, {}^{\alpha}x_2]$ be α-cuts of a solution. Then

$${}^{\alpha}x_1 = \frac{{}^{\alpha}b_1}{{}^{\alpha}a_1} \text{ and } {}^{\alpha}x_2 = \frac{{}^{\alpha}b_2}{{}^{\alpha}a_2}$$ are as given in the following table.

	α	${}^{\alpha}A = [{}^{\alpha}a_1, {}^{\alpha}a_2]$	${}^{\alpha}B = [{}^{\alpha}b_1, {}^{\alpha}b_2]$	${}^{\alpha}X = \left[\dfrac{{}^{\alpha}b_1}{{}^{\alpha}a_1}, \dfrac{{}^{\alpha}b_2}{{}^{\alpha}a_2}\right]$
	1	2	3	4
1	1	[4, 4]	[12, 12]	[3, 3]
2	0·9	[3, 4]	[9, 12]	[3, 3]
3	0·8	[3, 4]	[6, 12]	[2, 3]
4	0·7	[3, 5]	[6, 15]	[2, 3]
5	0·6	[3, 5]	[6, 20]	[2, 4]
6	0·4	[2, 5]	[4, 20]	[2, 4]
7	0·3	[2, 5]	[2, 20]	[1, 4]
8	0·2	[2, 5]	[2, 25]	[1, 5]
9	0·1	[2, 6]	[2, 30]	[1, 5]

Recall that $X(x) = \text{Sup}\{\alpha \mid x \in {}^{\alpha}X\}$. Therefore, if $\alpha_1 > \alpha_2$ and ${}^{\alpha_1}X = {}^{\alpha_2}X$, then we need to consider ${}^{\alpha_1}X$ only for the computation of X. Hence, consider ${}^{1}X$, ${}^{0\cdot8}X$, ${}^{0\cdot6}X$, ${}^{0\cdot3}X$, ${}^{0\cdot2}X$. *i.e.*, if we denote by (i, j) an element of i-th row and j-th column in the above table, we need (1, 4), (3, 4), (5, 4), (7, 4) and (8, 4) for the computation of X. (1, 4) contributes

$$\frac{1}{3} \quad \text{.....(1)}$$

Since $[2, 3] = [2, 3) \cup \{3\}$, (3, 4) contributes $$\frac{0{\cdot}8}{[2, 3)} \quad \text{.....(2)}$$

Since $[2, 4] = [2, 3] \cup (3, 4]$, (5, 4) contributes $$\frac{0{\cdot}6}{(3, 4]} \quad \text{.....(3)}$$

Since $[1, 4] = [1, 2) \cup [2, 4]$, (7, 4) contributes $$\frac{0{\cdot}3}{[1, 2)} \quad \text{.....(4)}$$

Finally, since $[1, 5] = [1, 4] \cup (4, 5]$, (8, 4) contributes $$\frac{0{\cdot}2}{(4, 5]} \quad \text{.....(5)}$$

Combining (1) to (5), we have, $$X = \frac{0{\cdot}3}{[1, 2)} + \frac{0{\cdot}8}{[2, 3)} + \frac{1}{3} + \frac{0{\cdot}6}{(3, 4]} + \frac{0{\cdot}2}{(4, 5]}.$$

EXERCISE - IV

1. Solve the equation $A + X = B$, if A and B are the fuzzy numbers given below

$$A(x) = \begin{cases} \dfrac{x+1}{2}, & \text{for } -1 \le x \le 1 \\ \dfrac{3-x}{2}, & \text{for } 1 \le x \le 3 \\ 0, & \text{otherwise} \end{cases} \text{ and } B(x) = \begin{cases} \dfrac{x}{2}, & \text{for } 0 \le x \le 2 \\ \dfrac{8-x}{6}, & \text{for } 2 \le x \le 8 \\ 0, & \text{otherwise} \end{cases}$$

[**Ans. :** $X(x) = \begin{cases} \dfrac{5-x}{4}, & \text{for } 1 \le x \le 5 \\ 0, & \text{otherwise} \end{cases}$

2. Solve $A + X = B$, if A and B are fuzzy numbers given by

$$A(x) = \begin{cases} -x^2 - 2x, & \text{for } -2 \le x \le -1 \\ 1, & \text{for } -1 \le x \le 1 \\ -x^2 + 2x, & \text{for } 1 \le x \le 2 \\ 0, & \text{everywhere else} \end{cases} \quad \text{and} \quad B(x) = \begin{cases} 1 - \left(\frac{1-x}{3}\right)^2, & \text{for } -2 \le x \le 1 \\ 1, & \text{for } 1 \le x \le 4 \\ 1 - \left(\frac{x-4}{5}\right)^2, & \text{for } 4 \le x \le 9 \\ 0, & \text{elsewhere} \end{cases}$$

$$\text{[Ans. : } X(x) = \begin{cases} 1 - \left(\frac{x-2}{2}\right)^2, & \text{for } 0 \le x \le 2 \\ 1, & \text{for } 2 \le x \le 3 \\ 1 - \left(\frac{x-3}{4}\right)^2, & \text{for } 3 \le x \le 7 \end{cases}$$

3. Solve the equation $A + X = B$ if A and B are given as below :

$$A = \frac{0{\cdot}4}{[2,3)} + \frac{0{\cdot}9}{[3,4)} + \frac{1}{4} + \frac{0{\cdot}7}{(4,5]} + \frac{0{\cdot}1}{[5,6)}$$

$$\text{and} \quad B = \frac{0{\cdot}3}{[3,4)} + \frac{0{\cdot}4}{[4,5)} + \frac{0{\cdot}8}{[5,6)} + \frac{0{\cdot}9}{[6,7)} + \frac{1}{7} + \frac{0{\cdot}7}{(7,8]} + \frac{0{\cdot}6}{(8,9]} + \frac{0{\cdot}2}{(9,10]} + \frac{0{\cdot}1}{(10,11]}$$

$$\text{[Ans. : } X = \frac{0{\cdot}3}{[1,2)} + \frac{0{\cdot}8}{[2,3)} + \frac{1}{3} + \frac{0{\cdot}6}{(3,4]} + \frac{0{\cdot}2}{(4,5]} \text{]}$$

4. Solve the equation $A + X = B$, if A and B are fuzzy numbers given by the membership functions

$$A(x) = \begin{cases} 1 + \frac{x}{2}, & \text{for } -2 < x \le 0 \\ 1 - \frac{x}{2}, & \text{for } 0 < x \le 2 \\ 0, & \text{for all } x \le -2 \text{ and } x > 2 \end{cases} \quad \text{and} \quad B(x) = \begin{cases} \frac{x}{2} - 1, & \text{for } 2 < x \le 4 \\ 3 - \frac{x}{2}, & \text{for } 4 < x \le 6 \\ 0, & \text{for } x > 6 \text{ and } x \le 2 \end{cases}$$

[Ans. : $A(x) = 1$, for $x = 4$; $A(x) = 0$, for $x \ne 4$]

5. Solve the equation $AX = B$, if A and B are fuzzy numbers given by

$$A(x) = \begin{cases} \frac{x}{2} - 1, & \text{for } 2 < x \le 4 \\ 3 - \frac{x}{2}, & \text{for } 4 < x \le 6 \\ 0, & \text{for } x > 6,\ x \le 2 \end{cases} \quad ; \quad B(x) = \begin{cases} \frac{x}{2} - 3, & \text{for } 6 < x \le 8 \\ 5 - \frac{x}{2}, & \text{for } 8 < x \le 10 \\ 0, & \text{for } x > 10,\ x \le 6 \end{cases}$$

$$\text{[Ans. : } X(x) = \begin{cases} \frac{3-x}{x-1}, & \text{for } 2 < x \le 3 \\ \frac{3x-5}{x-1}, & \text{for } \frac{5}{3} < x \le 2 \\ 0, & \text{everywhere else} \end{cases}$$

6. Solve the equation $AX = B$, if A and B are fuzzy numbers given as follows :

$$A = \frac{0{\cdot}1}{[0,1)} + \frac{0{\cdot}2}{[1,2)} + \frac{1}{2} + \frac{0{\cdot}3}{(2,3]} + \frac{0{\cdot}2}{(3,4]}$$

$$\text{and} \quad B = \frac{0{\cdot}1}{[0,5)} + \frac{0{\cdot}2}{[5,10)} + \frac{1}{10} + \frac{0{\cdot}3}{(10,15]} + \frac{0{\cdot}1}{(24,28]}$$

$$\text{[Ans. : } X = \frac{0{\cdot}1}{[2,5)} + \frac{1}{5} + \frac{0{\cdot}2}{(5,6]} + \frac{0{\cdot}1}{(6,7]} \text{]}$$

7. Solve the equation $AX = B$, if A and B are fuzzy numbers defined by the membership functions :

$$A(x) = \begin{cases} 1 - x^2, & \text{for } -1 \le x \le 1 \\ 0, & \text{elsewhere} \end{cases} \quad \text{and} \quad B(x) = \begin{cases} 1 - \left(\frac{x}{2}\right)^2, & \text{for } -2 \le x \le 2 \\ 0, & \text{elsewhere} \end{cases}$$

[Ans. : $X(x) = 1$, for $x = 2$, $X(x) = 0$, for $x \ne 2$]

8. Solve the equation $A + X = B$, if A and B are fuzzy numbers given below :

$$A = \frac{0{\cdot}2}{[2,3)} + \frac{0{\cdot}5}{[3,4)} + \frac{1}{[4,5]} + \frac{0{\cdot}7}{(5,6]} + \frac{0{\cdot}5}{(6,7]}$$

$$\text{and} \quad B = \frac{0{\cdot}1}{[3,4)} + \frac{0{\cdot}2}{[4,5)} + \frac{0{\cdot}5}{[5,6)} + \frac{1}{[6,8]} + \frac{0{\cdot}7}{(8,9]} + \frac{0{\cdot}5}{(9,10]} + \frac{0{\cdot}4}{(10,11]} + \frac{0{\cdot}1}{(11,12]}$$

$$\text{[Ans. : } X = \frac{0{\cdot}1}{[1,2)} + \frac{1}{[2,3]} + \frac{0{\cdot}4}{(3,4]} + \frac{0{\cdot}1}{(4,5]} \text{]}$$

9. Solve the equation $AX = B$, if A is the fuzzy number given in the above example and B is the fuzzy number given below :

$$B = \frac{0{\cdot}1}{[2,4)} + \frac{0{\cdot}2}{[4,6)} + \frac{0{\cdot}5}{[6,8)} + \frac{1}{[8,15]} + \frac{0{\cdot}7}{(15,18]} + \frac{0{\cdot}5}{(18,21]} + \frac{0{\cdot}4}{(21,28]} + \frac{0{\cdot}1}{(28,35]}$$

$$\text{[Ans. : } X = \frac{0{\cdot}1}{[1,2)} + \frac{1}{[2,3]} + \frac{0{\cdot}6}{(3,4)} + \frac{0{\cdot}1}{(4,5]} \text{]}$$

10. Solve $A + X = B$, if A and B are as follows :

$$A(x) = \begin{cases} \frac{x+1}{2}, & \text{for } -1 \le x < 1 \\ \frac{3-x}{2}, & \text{for } 1 \le x < 3 \\ 0, & \text{otherwise} \end{cases} \quad \text{and} \quad B(x) = \begin{cases} x + 3, & \text{for } -3 \le x \le -2 \\ -(x/2), & \text{for } -2 \le x \le 0 \\ 0, & \text{otherwise} \end{cases}$$

$$\text{[Ans. : } X(x) = \begin{cases} -2 - x, & \text{for } -3 \le x < -2 \\ 0, & \text{otherwise} \end{cases}$$

7. Extension Principle

The extension principle is one of the most important principles in the fuzzy set theory. It tells us how to extend usual functions of crisp variables to fuzzy sets.

Suppose we are given a function of n-variables

$$f : X_1 \times X_2 \times \ldots\ldots \times X_n \to Y, \text{ say}$$

$$f(x_1, x_2, \ldots\ldots, x_n) = y.$$

We want to extend the function to "fuzzy variable". *i.e.*, we want to replace $x_1, x_2, \ldots\ldots, x_n$ by fuzzy sets $A_1, A_2, \ldots\ldots, A_n$, defined on $X_1, X_2, \ldots\ldots, X_n$ respectively. (Recall that each x_i is a fuzzy set which has value 1 at x_i and value 0, elsewhere.)

Naturally, we expect $f(A_1, A_2, \ldots, A_n)$ to be a fuzzy set B which can legitimately, be called the image of $(A_1, A_2, \ldots, A_n)$. By legitimacy we mean when $A_1, A_2, \ldots, A_n$ are replaced by "fuzzy sets" $x_1, x_2, \ldots, x_n$ we should get y. This is achieved by the following principle.

Definition of Extension Principle

Every function $f: X_1 \times X_2 \times \ldots \times X_n \to Y$ induces a function

$$f: \mathcal{F}(X_1) \times \mathcal{F}(X_2) \times \ldots \times \mathcal{F}(X_n) \to \mathcal{F}(Y)$$

defined by the following rule :

If $A_1, A_2, \ldots, A_n$ are fuzzy sets on $X_1, X_2, \ldots, X_n$, respectively then

$$f(A_1, A_2, \ldots, A_n) = B, \quad \text{where, } B(y) = \text{Sup}\{\min\{A_1(x_1), A_2(x_2), \ldots, A_n(x_n)\}\}$$

and the suppremum is taken over all n-tuples $(x_1, x_2, \ldots, x_n)$ such that

$$f(x_1, x_2, \ldots, x_n) = y.$$

$B(y) = 0$, if there is no n-tuple $(x_1, x_2, \ldots, x_n)$ such that $f(x_1, x_2, \ldots, x_n) = y$.

The fuzzy set B on Y is called the image of $(A_1, A_2, \ldots, A_n)$. Recall that the set of all $(x_1, x_2, \ldots, x_n)$ such that $f(x_1, x_2, \ldots, x_n) = y$ is denoted by $f^{-1}(y)$. In this notation we can define B compactly;

$$B(y) = \text{Sup}\{\min\{A(x_1), A(x_2), \ldots, A(x_n)\}\}$$
$$(x_1, x_2, \ldots, x_n) \in f^{-1}(y), \text{ if } f^{-1}(y) \neq \Phi;$$
$$B(y) = 0, \text{ if } f^{-1}(y) = \Phi.$$

Note that if $f: X \to Y$ is the given function, the extension principle reduces to the following simple form :

$$f: \mathcal{F}(X) \to \mathcal{F}(Y) \qquad f(A) = B,$$

where
$$B(y) = \text{Sup}\{A(x) \mid x \in f^{-1}(y)\}, \qquad \text{if } f^{-1} \neq \Phi$$
$$= 0, \qquad \text{if } f^{-1} = \Phi$$

Example 1 : Using the extension principle find $f(A, B)$, where $f(x_1, x_2) = 2x_1 + x_2$ and A and B are given as

$$A = \frac{0{\cdot}5}{1} + \frac{0{\cdot}6}{2} + \frac{0{\cdot}9}{3} + \frac{1}{4} + \frac{0{\cdot}2}{5}; \quad B = \frac{0{\cdot}3}{3} + \frac{0{\cdot}4}{4} + \frac{0{\cdot}7}{5} + \frac{1}{6} + \frac{0{\cdot}3}{7}. \qquad \text{(S.U. 2010)}$$

Sol. : Here, $X_1 = \{1, 2, 3, 4, 5\}$ and $X_2 = \{3, 4, 5, 6, 7\}$. Since $f(x_1, x_2) = 2x_1 + x_2$, we may assume $Y = \{5, 6, \ldots, 17\}$.

Simple calculation gives us

$f(1, 3) = 5$,	$f(1, 4) = 6$,	$f(1, 5) = 7$,	$f(1, 6) = 8$,	$f(1, 7) = 9$,
$f(2, 3) = 7$,	$f(2, 4) = 8$,	$f(2, 5) = 9$,	$f(2, 6) = 10$,	$f(2, 7) = 11$
$f(3, 3) = 9$,	$f(3, 4) = 10$,	$f(3, 5) = 11$,	$f(3, 6) = 12$,	$f(3, 7) = 13$
$f(4, 3) = 11$,	$f(4, 4) = 12$,	$f(4, 5) = 13$,	$f(4, 6) = 14$,	$f(4, 7) = 15$
$f(5, 3) = 13$,	$f(5, 4) = 14$,	$f(5, 5) = 15$,	$f(5, 6) = 16$,	$f(5, 7) = 17$

Therefore, $f^{-1}(5) = \{(1, 3)\}$, $f^{-1}(6) = \{(1, 4)\}$, $f^{-1}(7) = \{(1, 5), (2, 3)\}$,

$f^{-1}(8) = \{(1, 6), (2, 4)\}$, $f^{-1}(9) = \{(1, 7), (2, 5), (3, 3)\}$,

$f^{-1}(10) = \{(2, 6), (3, 4)\}$, $f^{-1}(11) = \{(2, 7), (3, 5), (4, 3)\}$,

$f^{-1}(12) = \{(3, 6), (4, 4)\}$, $f^{-1}(13) = \{(3, 7), (4, 5), (5, 3)\}$,

$f^{-1}(14) = \{(4, 6), (5, 4)\}$, $f^{-1}(15) = \{(4, 7), (5, 5)\}$,

$f^{-1}(16) = \{(5, 6)\}$, $f^{-1}(17) = \{(5, 7)\}$.

Therefore, $f(A, B) = C$, where

$$C(5) = \max\{\min\{A(1), B(3)\}\} = \min\{A(1), B(3)\}$$
$$= \min\{0{\cdot}5, 0{\cdot}3\} = 0{\cdot}3$$

Similarly,
$$C(6) = \min\{A(1), B(4)\} = \min\{0{\cdot}5, 0{\cdot}4\} = 0{\cdot}4$$
$$C(7) = \max\{\min\{A(1), B(5)\}\}, \min\{A(2), B(3)\}\}$$
$$= \max\{\min\{0{\cdot}5, 0{\cdot}7\}, \min\{0{\cdot}6, 0{\cdot}3\}\}$$
$$= \max\{0{\cdot}5, 0{\cdot}3\} = 0{\cdot}5$$
$$C(8) = \max\{\min\{A(1), B(6)\}\}, \min\{A(2), B(4)\}\}$$
$$= \max\{\min\{0{\cdot}5, 1\}, \min\{0{\cdot}6, 0{\cdot}4\}\}$$
$$= \max\{0{\cdot}5, 0{\cdot}4\} = 0{\cdot}5$$
$$C(9) = \max\{\min\{A(1), B(7)\}, \min\{A(2), B(5)\}, \min\{A(3), B(3)\}\}$$
$$= \max\{\min\{0{\cdot}5, 0{\cdot}3\}, \min\{0{\cdot}6, 0{\cdot}7\}, \min\{0{\cdot}9, 0{\cdot}3\}\}$$
$$= \max\{0{\cdot}3, 0{\cdot}6, 0{\cdot}3\} = 0{\cdot}6$$
$$C(10) = \max\{\min\{A(2), B(6)\}, \min\{A(3), B(4)\}\}$$
$$= \max\{\min\{0{\cdot}6, 1\}, \min\{0{\cdot}9, 0{\cdot}4\}\}$$
$$= \max\{0{\cdot}6, 0{\cdot}4\} = 0{\cdot}6$$
$$C(11) = \max\{\min\{A(2), B(7)\}, \min\{A(3), B(5)\}, \min\{A(4), B(3)\}\}$$
$$= \max\{\min\{0{\cdot}6, 0{\cdot}3\}, \min\{0{\cdot}9, 0{\cdot}7\}, \min\{1, 0{\cdot}3\}\}$$
$$= \max\{0{\cdot}3, 0{\cdot}7, 0{\cdot}3\} = 0{\cdot}7$$
$$C(12) = \max\{\min\{A(3), B(6)\}, \min\{A(4), B(4)\}\}$$
$$= \max\{\min\{0{\cdot}9, 1\}, \min\{1, 0{\cdot}4\}\}$$
$$= \max\{0{\cdot}9, 0{\cdot}4\} = 0{\cdot}9$$
$$C(13) = \max\{\min\{A(3), B(7)\}, \min\{A(4), B(5)\}, \min\{A(5), B(3)\}\}$$
$$= \max\{\min\{0{\cdot}9, 0{\cdot}3\}, \min\{1, 0{\cdot}7\}, \min\{0{\cdot}2, 0{\cdot}3\}\}$$
$$= \max\{0{\cdot}3, 0{\cdot}7, 0{\cdot}2\} = 0{\cdot}7$$
$$C(14) = \max\{\min\{A(4), B(6)\}, \min\{A(5), B(4)\}\}$$
$$= \max\{\min\{1, 1\}, \min\{0{\cdot}2, 0{\cdot}4\}\}$$
$$= 1$$
$$C(15) = \max\{\min\{A(4), B(7)\}, \min\{A(5), B(5)\}\}$$
$$= \max\{\min\{1, 0{\cdot}3\}, \min\{0{\cdot}2, 0{\cdot}7\}\}$$
$$= \max\{0{\cdot}3, 0{\cdot}2\} = 0{\cdot}3$$
$$C(16) = \min\{A(5), B(6)\} = 0{\cdot}2$$
$$C(17) = \min\{A(5), B(7)\} = 0{\cdot}2$$

Example 2 : Given a function

$$f: \mathbb{N} \to \mathbb{Z}$$
$$f(x) = \begin{cases} (x-1)(x-3), & \text{for } x = 1, 2, \ldots, 5 \\ 0, & \text{for } x \geq 6 \end{cases}$$

and a fuzzy set $A: \mathbb{N} \to [0, 1]$

$$A(x) = \frac{1}{x}. \qquad \text{Find } f(A).$$

Sol. : Observe that $f(1) = 0 = f(3) = f(n)$, for $n = 6, 7, \ldots$;

$f(2) = -1$, $f(4) = 3$, $f(5) = 8$.

Hence, $f^{-1}(0) = \{1, 3, 6, 7, 8, \ldots\}$

$f^{-1}(-1) = 2$, $f^{-1}(3) = 4$, $f^{-1}(8) = 5$

and $f^{-1}(n) = \Phi$, for all other $n \in \mathbb{Z}$.

Let $f(A) = B$. $B(0) = \text{Sup}\{A(n) \mid n \in f^{-1}(0)\}$

$= \text{Sup}\{A(1), A(3), A(6), A(7), \ldots\}$

$= \text{Sup}\left\{1, \frac{1}{3}, \frac{1}{6}, \frac{1}{7}, \ldots\right\} = 1$

$B(-1) = \text{Sup}\{A(x) \mid x \in f^{-1}(-1)\}$

$= A(2) = 0{\cdot}5$

Similarly, $B(3) = A(4) = \frac{1}{4} = 0{\cdot}25$

$B(8) = A(5) = \frac{1}{5} = 0{\cdot}2$

$B(n) = 0$, for all $n \neq 0, -1, 3, 8$.

Example 3 : Using the extension principle, find $f(A, B)$, where $f(x_1, x_2) = x_1 \cdot x_2$;

$$A = \frac{0{\cdot}5}{1} + \frac{0{\cdot}6}{2} + \frac{0{\cdot}9}{3} + \frac{1}{4} + \frac{0{\cdot}2}{5}; \quad B = \frac{0{\cdot}3}{3} + \frac{0{\cdot}4}{4} + \frac{0{\cdot}7}{5} + \frac{1}{6} + \frac{0{\cdot}3}{7}.$$ **(S.U. 2009)**

Sol. : Let $X_1 = \{1, 2, \ldots, 5\}$, $X_2 = \{3, 4, \ldots, 7\}$, and

$Y = \{3, 4, \ldots, 35\}$ and $f : X_1 \times X_2 \to Y$. $f(x_1, x_2) = x_1 \cdot x_2$

We have, then

$f^{-1}(3) = \{(1, 3)\}$, $f^{-1}(4) = \{(1, 4)\}$, $f^{-1}(5) = \{(1, 5)\}$,

$f^{-1}(6) = \{(1, 6), (2, 3)\}$, $f^{-1}(7) = \{(1, 7)\}$, $f^{-1}(8) = \{(2, 4)\}$,

$f^{-1}(9) = \{(3, 3)\}$, $f^{-1}(10) = \{(2, 5)\}$, $f^{-1}(12) = \{(2, 6), (3, 4), (4, 3)\}$,

$f^{-1}(14) = \{(2, 7)\}$ $f^{-1}(15) = \{(3, 5), (5, 3)\}$, $f^{-1}(16) = \{(4, 4)\}$

$f^{-1}(18) = \{(3, 6)\}$, $f^{-1}(20) = \{(4, 5), (5, 4)\}$, $f^{-1}(21) = \{(3, 7)\}$

$f^{-1}(24) = \{(4, 6)\}$ $f^{-1}(25) = \{(5, 5)\}$ $f^{-1}(28) = \{(4, 7)\}$

$f^{-1}(30) = \{(5, 6)\}$ $f^{-1}(35) = \{(5, 7)\}$ $f^{-1}(x) = \Phi$, for all other x in Y.

Let $f(A, B) = C$, then $C : Y \to [0, 1]$ is defined by

$C(y) = \max\{\min(A(x_1), B(x_2)) \mid (x_1, x_2) \in f^{-1}(y_1)\}$, if $f^{-1}(y) \neq \Phi$

$= 0$, if $f^{-1}(y) = \Phi$

$\therefore$ $C(3) = \min\{0{\cdot}5, 0{\cdot}3\} = 0{\cdot}3$

$C(4) = 0{\cdot}4$, $C(5) = 0{\cdot}5$

$C(6) = \max\{\min\{A(1), B(6)\}, \min\{A(2), B(3)\}\}$

$= \max\{0{\cdot}5, 0{\cdot}3\} = 0{\cdot}5$

$C(7) = 0{\cdot}3$, $C(8) = 0{\cdot}4$, $C(9) = 0{\cdot}3$, $C(10) = 0{\cdot}6$.

$C(12) = \max\{\min\{A(2), B(6)\}, \min\{A(3), B(4)\}, \min\{A(4), B(3)\}\}$

$= \max\{0{\cdot}6, 0{\cdot}4, 0{\cdot}3\} = 0{\cdot}6$

$C(14) = 0{\cdot}3$, $C(15) = \max\{0{\cdot}7, 0{\cdot}2\} = 0{\cdot}7$,

$C(16) = 0{\cdot}4$, $C(18) = 0{\cdot}9$,

$C(20) = \max\{\min\{A(4), B(5)\}, \min\{A(5), B(4)\}\}$

$= \max\{0{\cdot}7, 0{\cdot}2\} = 0{\cdot}7$

$C(21) = 0{\cdot}3$, $C(24) = 1$, $C(25) = 0{\cdot}2$, $C(28) = 0{\cdot}3$,

$C(30) = 0{\cdot}2$, $C(35) = 0{\cdot}2$ and $C(y) = 0$ for all other elements y in Y.

Example 4 : Let A and B be fuzzy sets defined on a universal set X whose membership functions are given by

$$A(x) = \frac{0{\cdot}5}{-1} + \frac{1}{0} + \frac{0{\cdot}5}{1} + \frac{0{\cdot}3}{2} \text{ and } B(x) = \frac{0{\cdot}5}{2} + \frac{1}{3} + \frac{0{\cdot}5}{4} + \frac{0{\cdot}3}{5}.$$

Let $f : X \times X \to X$ be a function defined by

$f(x_1, x_2) = x_1 \cdot x_2$.

Calculate $f(A, B)$.

Sol. : Let $X_1 = \{-1, 0, 1, 2\}$, $X_2 = \{2, 3, 4, 5\}$ **(S.U. 2012)**

Without the loss of generality we assume X to be any set containing $X_1 \cup X_2$.

By convention $A(x) = 0$, for all $x \in X \sim X_1$ and $B(x) = 0$, for all $x \in X \sim X_2$.

Therefore, $\min\{A(x_1), B(x_2)\} = 0$, for all $(x_1, x_2) \notin X_1 \times X_2$.

Let $f(A, B) = C$, then ——— (1)

$C(y) = \text{Sup}\{\min\{A(x_1), B(x_2)\} \mid (x_1, x_2) \in f^{-1}(y)\}$, if $f^{-1}(y) \neq \Phi$

$= 0$, if $f^{-1}(y) = \Phi$.

From (1), it is obvious that (x_1, x_2) which are not in $X_1 \times X_2$ do not contribute to the calculation of C. Therefore, it is sufficient to consider only those (x_1, x_2) which are in $X_1 \times X_2$ for the calculation of $C(y)$.

Clearly, $f(X_1 \times X_2) = \{-2, -3, -4, -5, 0, 2, 3, 4, 5, 6, 8, 10\}$.

Moreover $f^{-1}(0) = \{0\} \times X_2$ and $f^{-1}(4) = \{(1, 4), (2, 2)\}$.

For all other $y \in f(X_1 \times X_2)$, $f^{-1}(y)$ is a singleton set. With this understanding we calculate the C's :

$C(-2) = \min\{A(-1), B(2)\} = 0{\cdot}5$

$C(-3) = \min\{A(-1), B(3)\} = 0{\cdot}5$

$C(-4) = \min\{A(-1), B(4)\} = 0{\cdot}5$

$C(-5) = \min\{A(-1), B(5)\} = 0{\cdot}3$

$C(0) = \text{Sup}\{\min\{A(0), B(x)\} \mid x \in X_2\}$

$= \text{Sup}\{B(x) \mid x \in X_2\}$ (Since $A(0) = 1$)

$= \text{Sup}\{0{\cdot}5, 1, 0{\cdot}5, 0{\cdot}3\}$

$= 1$

$C(2) = \min\{A(1), B(2)\} = 0{\cdot}5$

$C(3) = \min\{A(1), B(3)\} = 0{\cdot}5$

$C(4) = \text{Sup}\{\min\{A(1), B(4)\}, \min\{A(2), B(2)\}\}$

$= \text{Sup}\{0{\cdot}5, 0{\cdot}3\} = 0{\cdot}5$

$C(5) = \min\{A(1), B(5)\} = 0{\cdot}3$

$C(y) = 0$, for all other $y \in X$.

A Slightly Different Definition of the Extension Principle

The fuzzy set theory is very new and growing. Therefore, you may find variations in notations and nomenclature. Here is a sightly different variation of the Extension Principle :

Every function $f : X \to Y$ induces two functions;

(a) $f : \mathcal{F}(X) \to \mathcal{F}(Y)$, and

(b) $f^{-1} : \mathcal{F}(Y) \to \mathcal{F}(X)$ defined as follows :

If $A : X \to [0, 1]$ and $B : Y \to [0, 1]$ are any fuzzy sets then

$$f(A)(y) = \text{Sup}\{A(x) \mid x \in f^{-1}(y)\}, \quad \text{if } f^{-1}(y) \neq \Phi$$
$$= 0, \quad \text{if } f^{-1}(y) = \Phi$$
$$f^{-1}(B)(x) = B(f(x)).$$

EXERCISE - V

1. Let $f(x) = (x-1)(x-3)$, for $x = 1, 2, \ldots, 5$
$= 0$, for $x \geq 6$

A fuzzy set $A : \mathbb{N} \to [0, 1]$ is defined by $A(n) = \dfrac{n}{n+1}$.

Find $f(A)$. [**Ans. :** $f(A)(0) = 1$, $f(A)(-1) = \dfrac{2}{3}$, $f(A)(3) = \dfrac{4}{5}$, $f(A)(8) = \dfrac{5}{6}$ $f(A)$ vanishes everywhere else.]

2. Let $X = \{x_1, x_2, x_3\}$, $Y = \{y_1, y_2, y_3\}$
$Z = \{z_1, z_2, z_3\}$, $f : X \times Y \to Z$ be given by

$f(x_1, y_1) = f(x_1, y_2) = f(x_3, y_2) = z_1$,
$f(x_1, y_3) = f(x_2, y_1) = f(x_2, y_3) = z_2$,
$f(x_2, y_2) = f(x_3, y_1) = f(x_3, y_3) = z_3$.

$$A = \frac{0{\cdot}2}{x_1} + \frac{0{\cdot}7}{x_2} + \frac{0{\cdot}5}{x_3}, \quad B = \frac{0{\cdot}5}{y_1} + \frac{0{\cdot}3}{y_2} + \frac{1}{y_3}$$

Find $f(A, B)$. [**Ans. :** $f(A, B) = \dfrac{0{\cdot}3}{z_1} + \dfrac{0{\cdot}7}{z_2} + \dfrac{0{\cdot}5}{z_3}$]

3. Let $X = \{0, 1, 2, \ldots, 10\}$ and $Y = \{0, \ldots, 100\}$.
Let $f : X \to Y$ be defined as $f(x) = x^2$.
Find $f(A)$, if $A : X \to [0, 1]$ is defined as

(i) $A(x) = \dfrac{x}{x+2}$, (ii) $A(x) = 2^{-x}$, (iii) $A(x) = \dfrac{1}{1+10x}$,

(iv) $A(x) = \dfrac{1}{(1+10x)^2}$, (v) $A(x) = \dfrac{1}{1+10(x-2)^2}$

(**Hint :** If $f : X \to Y$ is one-one or injective, $f^{-1}(y)$ is a single element set or empty set. Therefore, if $f(A) = B$, then, $B(y) = A(x)$; if $f(x) = y$ or $B(y) = 0$, if there is no x, such that $f(x) = y$.)

[**Ans. :** Let $Z = \{0, 4, 9, 16, 25, 36, 49, 64, 81, 100\}$.

Then, $B : Y \to [0, 1]$ is given by

(i) $B(y) = \dfrac{+\sqrt{y}}{+\sqrt{y}+2}$ for $y \in Z$ and $B(y) = 0$ for $y \in Y \sim Z$.

(ii) $B(y) = 2^{-\sqrt{y}}$ if $y \in Z$, $B(y) = 0$, otherwise

(iii) $B(y) = \dfrac{1}{1+10\sqrt{y}}$, if $y \in Z$, $B(y) = 0$, otherwise

(iv) $B(y) = \dfrac{1}{(1+10\sqrt{y})^2}$, if $y \in Z$, $B(y) = 0$, otherwise

(v) $B(y) = \dfrac{1}{1+10(\sqrt{y}-2)^2}$, if $y \in Z$, $B(y) = 0$, otherwise]

4. Let A and B be fuzzy sets defined on a set X, whose membership functions are given by

$$A = \frac{0{\cdot}5}{-1} + \frac{1}{0} + \frac{0{\cdot}5}{1} + \frac{0{\cdot}3}{2}; \quad B = \frac{0{\cdot}5}{2} + \frac{1}{3} + \frac{0{\cdot}5}{4} + \frac{0{\cdot}3}{5}$$

If $f : X \times X \to X$ is defined by $f(x_1, x_2) = x_1 + x_2$, find $f(A, B)$.

[**Ans. :** $C = \dfrac{0{\cdot}5}{1} + \dfrac{0{\cdot}5}{2} + \dfrac{1}{3} + \dfrac{0{\cdot}5}{4} + \dfrac{0{\cdot}5}{5} + \dfrac{0{\cdot}3}{6} + \dfrac{0{\cdot}3}{7}$]

5. Given a fuzzy set $A = \dfrac{0{\cdot}1}{-1} + \dfrac{0{\cdot}4}{0} + \dfrac{1}{1} + \dfrac{0{\cdot}4}{2} + \dfrac{0{\cdot}1}{3}$,

and a function $f : \mathbb{Z} \to \mathbb{N}$ by $f(x) = x^2 + 1$, find $f(A)$. [**Ans. :** $f(A) = \dfrac{0{\cdot}4}{1} + \dfrac{1}{2} + \dfrac{0{\cdot}4}{5} + \dfrac{0{\cdot}1}{10}$]

☆☆☆

CHAPTER 10

Assignment Problems

1. Introduction

Assignment of jobs to different machines or to different workers is another problem a manager has to solve. In any workshop, there are a number of workers and a number of jobs to be completed by them. But all workers are not equally skilful. Similarly, all machines are not equally suitable for various jobs. The manger's problem is - how to choose the best man for a job, and also the best machine for a job; how to assign a job to a man or to a machine such that the total cost of production is minimum or the profit is maximum ? This is an assignment problem.

Suppose there are vehicals at different origins and these are to be dispatched to different destinies. How these vehicals should be assigned to destinies so that the total distance covered by all the vehicals will be minimum *i.e.* the total cost will be minimum ? This is another assignment problem.

2. Formulation of an Assignment Problem

Consider a problem in which there are n jobs (to be produced) and n machines (on which they can be produced) to be assigned to them. Let the cost of producing i-th job on j-th machine be c_{ij}.

We have to assign a job to a machine such that the total cost of production is minimum.

Let x_{ij} denote the assignment of i-th job to the j-th machine. Now, x_{ij} is a variable such that

$$x_{ij} = \begin{cases} 1 & \text{if } i\text{ - th job is assigned to } j\text{ - th machine} \\ 0 & \text{otherwise} \end{cases}$$

$$\text{The total cost} = \sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij}\, x_{ij}$$

The assignment problem now can be stated as follows :

Determine $x_{ij} \geq 0$ $(i, j = 1, 2, \ldots\ldots n)$ so as to

$$\text{Minimise } Z = \sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij}\, x_{ij}$$

Subject to the constraints

$$\sum_{i=1}^{n} x_{ij} = 1, \quad j = 1, 2, \ldots\ldots n$$

$$\sum_{j=1}^{n} x_{ij} = 1, \quad i = 1, 2, \ldots\ldots n$$

The problem can be represented by the following table called **cost matrix**.

Cost-Matrix

Job \ Machine	M_1	M_2	M_3	.....	M_n	Required
J_1	C_{11}	C_{12}	C_{13}	.....	C_{1n}	1
J_2	C_{21}	C_{22}	C_{23}	.....	C_{2n}	1
J_3	C_{31}	C_{32}	C_{33}	.....	C_{3n}	1
.....	.....	.....	.....	.....	.....	.....
.....	.....	.....	.....	.....	.....	.....
J_n	C_{n1}	C_{n2}	C_{n3}	.....	C_{nn}	1
Available	1	1	1	.....	1	

3. Methods of Solving Assignment Problem

There are four methods of solving an assignment problem. They are -

1. Enumeration Method
2. Simplex Method
3. Transportation Method
4. Hungarian method

1. Enumeration Method : In this method we prepare a list of all possible assignments of the given origins (men, machines, vehicals etc.) and destinies (jobs, sales zones, etc.). Then we select one by one an assignment which will give minimum cost, (maximum profit). If two assignments have the same minimum cost (or maximum profit), we may select any one of them at a time and proceed further. In this way there will be alternate optimal solutions.

The method is clearly very simple. But its drawback is that as the size of the problem increases it becomes more and more difficult to apply this method. If there are n workers and n jobs, then there will be n ! total possible assignments. If there are six workers and six jobs to be assigned then there will be 6 ! = 720 total assignments ! Because of this reason, the method is not in use.

2. Simplex Method : As seen in the above artical 2, an assignment problem can be expressed as a linear programming problem. We have a cost function (a profit function) to be minimised (to be maximised) with certain constraints. The objective function and the constraints are linear and as such we can apply simplex method also.

But this method is also not suitable for large n. In an assignment problem there are $n \times n$ decision variables and $n + n$ equalities. In a problem if there are 6 workers and 6 jobs, these will be 36 decision variables and 12 equalities. Thus, it becomes difficult to use this method also.

3. Transportation Method : An assignment problem is basically a different type of transportation problems. Hence, it can be solved by the methods of transportation problem (to be discussed in the next chapter). But to use this method for assignment problem we need to introduce $(n - 1)$ dummy allocations. Thus, the method is not suitable.

4. Hungarian Method : This is a very simple and still very efficient and fast method to solve an assignment problem. This was developed by Hungarian mathematician D. Konig and hence is known as Hungarian Method. We are going to use this method in this chapter.

In this method we first reduce the cost-matrix by subtracting the smallest element in each row from other elements of the row. The matrix, thus, obtained is called **reduced matrix**. Then we assign zeros to each row (column) by the method discussed in detail below. After that we check the solution for optimality. The detailed procedure of **Hungarian Method** is given below.

4. Fundamental Theorems

The solution of assignment problem is based on the following theorems.

Theorem 1 : If a constant is added or subtracted from each element of a row or column of a cost matrix then the optimal solution remains the same for new matrix.

We shall accept this theorem without proof.

Theorem 2 : In an assignment problem with costs C_{ij}, if all $C_{ij} \geq 0$, then a feasible solution x_{ij} such that

$$\sum_{i=1}^{n} \sum_{j=1}^{n} C_{ij} x_{ij} = 0 \quad \text{is optimal solution of the problem.}$$

5. Hungarian Method

Step 1 : To Reduce The Matrix

Find out the smallest element in each row and subtract it from all other elements of that row. Check if each column contains at least one zero. If not subtract the smallest element (of the column not containing zero) from the remaining elements of that column. The matrix so obtained is called **reduced matrix**.

Step 2 : To assign the zeros

Examine rows successively till you find a row containing **only one zero**. Encircle that **single zero** from that row and cross all other zeros, if any, which lie in its column. If any row contains more than one zero leave that row and go to the next. Examine all other rows in this way.

Now, examine columns successively till you find a column containing **only one zero**. Encircle that **single zero** from that column and cross all other zeros, if any, which lie in its row. Examine all other columns in this way.

If you follow this procedure carefully you will find that every zero in the table is either encircled or crossed.

If a row or a column, thus, has an encircled zero, we say that a zero has been **assigned** to that row or column. Or such a row or a column is called an **assigned row** or an **assigned column**.

Step 3 : Optimality Check

If every row and column is assigned *i.e.* contains one encircled zero then the current assignment is optimum. This means the number of assignments *i.e.* encircled zeros is equal to the order of the matrix *n*.

If there is even one row or column without an assignment (*i.e.* does not have an encircled zero) the assignment is not optimum and we follow the next step.

Step 4 : Find Minimum Number of Lines Which Cross all Zeros

In this step we are required to draw minimum number of lines which cross all zeros. This can be done in two ways - (i) either we can use our judgement or (ii) we can use the rule given below.

(i) Lines By Judgement : We can very easily draw minimum number of lines which cross all zeros. For this, we first find a row or a column which contains maximum number of zeros. We draw lines through these zeros. Then we draw lines through remaining zeros. Lines drawn in this manner will be minimum in number.

(ii) Lines By The Rule : The rule is -

(i) Tick-mark (✓) rows that do not have assignments *i.e.* encircled zero.

(ii) Tick-mark (✓) columns which have zeros in the tick-marked rows.

(iii) Tick-mark (✓) rows which have encircled zeros in the tick-marked columns.

Repeat the above procedure of (ii) and (iii) till no more marking is left.

Draw lines through **untick-marked rows** and **ticked columns**.

If the number of lines so drawn is equal to the order *n* of the matrix it is an optimal solution otherwise not. In the second case, follow the next step 5.

Step 5 : Iteration For Optimal Solution

Select the smallest element among all uncovered elements and subtract it from all elements through which no line has passed (uncovered elements). Add this smallest element to every element which lies at the intersection of two lines. The resulting matrix is the second basic feasible matrix.

Now, go to step 2 and repeat the whole procedure to find if you have got the optimal solution.

Example 1 : Solve the following minimal assignment problem.

	A	B	C	D
I	16	12	14	16
II	18	13	13	14
III	15	16	11	15
IV	17	14	12	18

Sol. : Step 1 : To Reduce The Matrix

Find out the smallest element in each row and subtract it from all other elements in that row. The smallest element in the first row is 12. Subtract it from all other elements of that row. Similarly, subtract the smallest element 13, 11 and 12 from all other elements of the second, third and fourth rows respectively. Thus, we get the matrix given in Table 1.

Table 1

	A	B	C	D
I	4	0	2	4
II	5	0	0	1
III	4	5	0	4
IV	5	2	0	6

Table 2

	A	B	C	D
I	~~0~~	0	2	3
II	1	~~0~~	~~0~~	0
III	0	5	~~0~~	3
IV	1	2	0	5

Note that all rows have zero elements but 1st and 4th columns do not have any zero element. Hence, subtract the smallest element 4 of the first column from all other elements of the first column and the smallest element 1 of the fourth column from all other elements of the fourth column. The **reduced matrix** is shown in Table 2.

Step 2 : To Assign Zeros

On examining each row of table 2 carefully we find that the fourth row contains only one zero. Encircle this zero and cross all other zeros which lie in its column *i.e.* 3rd column. Now, row 3 has a

single unmarked zero. Encircle it and cross the other zero in its column *i.e.* first column. Now, row 1 has a single unmarked zero. Encircle it and cross the other zero in its column *i.e.* second column. Lastly the second row has a single unmarked zero. Encircle it.

Step 3 : Optimality Check

Now every row and every column has only one single encircled zero. Since, there are 4 encircled zeros equal to the order of the matrix 4, the present solution is optimal.

The optimal assignment policy would be :

I assigned to *B*, II assigned to *D*, III assigned to *A* and IV assigned to *C*.

We denote this as -

$$\text{I} \xrightarrow{12} B, \quad \text{II} \xrightarrow{14} D, \quad \text{III} \xrightarrow{16} A, \quad \text{IV} \xrightarrow{12} C.$$

The optimal cost = 12 + 14 + 16 + 12 = ₹ 54.

Note ....

While assigning zeros it is necessary to follow the procedure given in step 2 above. You first find out the row which contains only one zero. Encircle it and go to the next. If a row contains two or more zeros leave the row and go to the next. Then examine the columns in the same way.

Example 2 : Solve the following minimal assignment problem.

	I	II	III	IV
A	9	27	18	12
B	14	29	5	27
C	39	20	19	16
D	20	27	25	11

Sol. : Step 1 : To Reduce The Matrix

The smallest elements in the first, second, third, and fourth rows are 9, 5, 16 and 11. These are to be subtracted from all other elements of the respective rows. Thus, we get Table 1.

Table 1

	I	II	III	IV
A	0	18	9	3
B	9	24	0	22
C	23	4	3	0
D	9	16	14	0

Table 2

	I	II	III	IV
A	0	14	9	3
B	9	20	0	22
C	23	0	3	~~0~~
D	9	12	14	0

Second column has no zero in it. So subtract the smallest element 4 from the remaining elements of that column. Thus, we get the Table 2. This is the **reduced matrix**.

Step 2 : To Assign Zeros

The first and second rows contain only one zero. Encircle these single zeros. The fourth row contains only one zero. Encircle it and cross the other zero in its fourth column. Now the second column contains only one unmarked zero. Encircle it.

Step 3 : Optimality Test

Every row and column has only one encircled zero. Since, there are four encircled zeros equal to the order of the matrix, the present solution is optimal.

The optimal assignment policy would be :

A assigned to I, *B* assigned to III,

C assigned to II, *D* assigned to IV.

We denote this as -

$$A \xrightarrow{9} \text{I}, \quad B \xrightarrow{5} \text{III}, \quad C \xrightarrow{20} \text{II}, \quad D \xrightarrow{11} \text{IV}.$$

The minimum cost = ₹ 9 + 5 + 20 + 11 = ₹ 45.

Example 3 : Solve the following minimal assignment problem

	A	B	C	D	E
I	6	5	8	11	16
II	1	13	16	1	1
III	16	11	8	8	8
IV	9	14	12	10	16
V	10	13	11	8	16

Sol. : Step 1 : To Reduce The Matrix (S.U. 2002)

Subtract the smallest elements 5, 1, 8, 9, 8 from the first, second, third, fourth and fifth rows respectively. We note carefully that each row and column contains at least one zero. (Table 1)

Table 1

	A	B	C	D	E
I	1	0	3	6	11
II	0	12	15	0	0
III	8	3	0	0	0
IV	0	5	3	1	7
V	2	5	3	0	8

Table 2

	A	B	C	D	E
I	1	0	3	6	11
II	~~0~~	12	15	~~0~~	0
III	8	3	0	~~0~~	~~0~~
IV	0	5	3	1	7
V	2	5	3	0	8

Step 2 : To Assign Zeros

First row contains one zero encircle it. Fourth row contains only one zero. Encircle it and cross the other zero in its column viz. 1st column. Now, fifth row contains one zero. Encircle it and cross other two zeros in its column viz. fourth column. Now, second row contains only one unmarked zero. Encircle it and cross the other zero in its column viz. fifth column. Lastly there is only one unmarked zero in the third column. Encircle it.

Step 3 : Optimality Check

Every row and every column has only one single encircled zero. Since there are five encircled zeros equal to the order of the matrix 5, the present solution is optimal.

The optimal alignment policy would be :

I assigned to B, II assigned to E, III assigned to C,

IV assigned to A and V assigned to D.

We denote this as -

$$\text{I} \xrightarrow{5} B,\quad \text{II} \xrightarrow{1} E,\quad \text{III} \xrightarrow{8} C,\quad \text{IV} \xrightarrow{9} A,\quad \text{V} \xrightarrow{8} D.$$

The total minimum cost = ₹ 5 + 1 + 8 + 9 + 8 = ₹ 31.

Example 4 : Solve the following minimum assignment problem.

Operators \ Machines	M_1	M_2	M_3	M_4
O_1	17	40	36	24
O_2	26	35	13	39
O_3	49	28	37	24
O_4	38	36	37	19

Sol. : Write the steps to obtain the following tables.

Table 1

	M_1	M_2	M_3	M_4
O_1	0	23	19	7
O_2	13	22	0	26
O_3	25	4	13	0
O_4	19	17	18	0

Table 2

	M_1	M_2	M_3	M_4
O_1	0	19	19	7
O_2	13	18	0	26
O_3	25	0	13	~~0~~
O_4	19	13	18	0

[**Ans. :** $O_1 \xrightarrow{17} M_1,\ O_2 \xrightarrow{13} M_3,\ O_3 \xrightarrow{28} M_2,\ O_4 \xrightarrow{19} M_4.$

Total = 77]

Example 5 : Solve the following minimal assignment problem.

Machines \ Jobs	J_1	J_2	J_3	J_4
M_1	18	14	18	20
M_2	21	15	15	16
M_3	17	20	13	18
M_4	21	18	14	24

Sol. : Write the steps to obtain the following tables.

Table 1

	J_1	J_2	J_3	J_4
M_1	4	0	4	6
M_2	6	0	0	1
M_3	4	7	0	5
M_4	7	4	0	10

Table 2

	J_1	J_2	J_3	J_4
M_1	~~0~~	0	4	5
M_2	2	~~0~~	~~0~~	0
M_3	0	7	~~0~~	4
M_4	3	4	0	9

[**Ans. :** $M_1 \xrightarrow{14} J_2,\ M_2 \xrightarrow{16} J_4,\ M_3 \xrightarrow{17} J_1,\ M_4 \xrightarrow{14} J_3$ Total = 59]

EXERCISE - I

Solve the following minimal assignment problems (1 and 2).

1.

	I	II	III	IV
A	15	33	24	18
B	20	35	11	33
C	45	26	25	22
D	26	33	31	17

2.

	I	II	III	IV
A	20	38	29	23
B	26	41	17	39
C	52	33	32	29
D	31	38	36	22

[**Ans. :** **(1)** A - I, B - III, C - II, D - IV ; Total = 69.

(2) A - I, B - III, C - II, D - IV ; Total = 92.]

3. Five jobs I, II, III, IV, V can be given to five contractors A, B, C, D, E with cost matrix as given below. Find the assignments so that the total cost is minimum.

Contractors \ Jobs	I	II	III	IV	V
A	13	12	9	16	11
B	15	10	10	18	12
C	11	14	13	12	17
D	11	12	7	8	7
E	12	13	13	15	10

[**Ans. :** A - III, B - II, C - I, D - IV, E - V ; Total = 48.]

4. Four jobs can be done at four places. The cost matrix is given on the right. Find the assignment so that the cost is minimum.

Jobs \ Places	I	II	III	IV
A	20	16	18	20
B	22	17	17	18
C	19	20	15	19
D	21	18	16	22

[**Ans. :** A - II, B - IV, C - I, D - III ; Total = 69.]

Problems With Iterations

We shall now consider some assignment problems in which we are required to make iterations. In such problems, in the first assignment we do not find a zero in each row and in each column *i.e.* every source is not assigned a destiny. In such cases we need to repeat the procedure. This is explained below through examples.

Example 1 : Solve the following minimal assignment problem.

	I	II	III	IV
A	10	18	17	15
B	23	12	22	16
C	21	22	24	19
D	12	23	21	17

Sol. : Step 1 : To Reduce The Matrix

Subtract the smallest elements 10, 12, 19 and 12 from the four rows successively. The result is shown in Table 1.

Table 1

	I	II	III	IV
A	0	8	7	5
B	11	0	10	4
C	2	3	5	0
D	0	11	9	5

Table 2

	I	II	III	IV	
A	0	8	2	5	✓
B	11	0	5	4	
C	2	3	0	~~0~~	
D	~~0~~	11	4	5	✓

Third column does not have a zero. Subtract the smallest element 5 from all other elements of that column. The result is shown in Table 2.

Step 2 : To Assign Zeros

First row has only one zero. Encircle it and cross the other zero in the first column. Second row has only one zero. Encircle it. Third row has two zeros. Leave the third row. Fourth row has no unmarked zero. Leave it.

Now, examine the columns in the same way. The third column has a only one zero. Encircle it and cross the other zero in its row *i.e.* in the third row.

Thus, all zeros are either encircled or crossed.

Step 3 : Optimality Test

All rows and all columns have no encircled zeros. Fourth row and fourth column have no assignments.

Hence, the current solution is **not optimal** and we go to the next step.

Step 4 : To Find The Minimum Number of Lines Which Cross All Zeros

We can better use our judgement to draw minimum number of lines crossing all zeros.

We first find the row or the column which has maximum number of zeros. Thus, the third row has maximum (two) zeros. We, therefore, draw a line through that row. Now, examining the columns,

we find that the first column has maximum (two) zeros. We, therefore, draw a line through the first column.

Now, there remains a zero in the second row which is not crossed by a line. Hence, draw a line through the second row.

This we have three lines, which cross all zeros.

Alternatively : We can follow the standard procedure.

(i) Tick-mark the fourth row which does not have assignment *i.e.* an encircled zero.

(ii) Tick-mark the first column as it has a zero in the marked row.

(iii) Tick-mark the first row as it has encircled zero in the marked column.

(iv) Now, draw lines through unmarked rows (2 and 3) and marked columns (1).

The three lines cross all zeros. Since, the number of lines 3 is less than the order of the matrix, the solution is not optimal.

We go for another iteration.

Step 5 : Iteration For Optimal Solution

The smallest element amongst all uncovered elements is 2. Subtract it from all uncovered elements and add it to the elements at the intersection of the lines. The result is shown in Table 3.

Table 3

	I	II	III	IV
A	~~0~~	6	0	3
B	13	0	5	4
C	4	3	~~0~~	0
D	0	9	2	3

Step 6 : To Assign Zeros

Second row has a single zero. Encircle it. Fourth row has a single zero. Encircle it and cross the other zero in the first column. Now, first row has a single zero. Encircle it and cross the other zero in the third column. Now, fourth column has a single row. Encircle it.

Step 7 : Optimality Check

Now, every row and every column has only one single encircled (assigned) zero. Since there are four encircled zeros, equal to the order of the matrix, the present solution is optimal.

The optimal assignment policy would be :-

$$A \xrightarrow{17} \text{III}, \quad B \xrightarrow{12} \text{II}, \quad C \xrightarrow{19} \text{IV}, \quad D \xrightarrow{12} \text{I}.$$

The total minimum cost = 17 + 12 + 19 + 12 = 60.

Example 2 : Solve the following minimal assignment problem.

	A	B	C	D	E
I	130	100	145	160	170
II	105	90	100	130	145
III	110	80	125	140	155
IV	20	20	50	50	80
V	25	5	40	50	75

Sol. : Step 1 : To Reduce The Matrix

Subtract the smallest elements 100, 90, 80, 20 and 5 from the five rows successively. The result is shown in Table 1.

Table 1

	A	B	C	D	E
I	30	0	45	60	70
II	15	0	10	40	55
III	30	0	45	60	75
IV	0	0	30	30	60
V	20	0	35	45	70

Table 2

	A	B	C	D	E	
I	30	0	35	30	15	✓
II	15	~~0~~	0	10	~~0~~	
III	30	~~0~~	35	30	20	✓
IV	0	~~0~~	20	~~0~~	5	
V	20	~~0~~	25	15	15	✓

Third, fourth, fifth columns have no zeros. Subtract the smallest elements 10, 30, 55 from all other elements of 3rd, 4th and 5th columns. The result is shown in Table 2.

Step 2 : To Assign Zeros

First row has only one zero. Encircle it and cross all other zeros in the second column in Table 2. Second and fourth rows have two zeros. Leave them. Now, examine columns. First column has a single zero. Encircle it and cross the other zero in the fourth row. Third column has a single zero. Encircle it and cross the other zero from the second row.

Step 3 : Optimality Test

Now, rows 3, 5 and columns 4, 5 have no assignments. The current solution is not optimal.

Step 4 : To Find Minimum Number of Lines Which Cross Zeros

We first find the row or the column which has maximum number of lines. Thus, we find that the second column has all elements zero. So we draw a line through the second column. Then, we see that the second row and fourth rows have two zeros. So we draw two lines through them.

Thus, these three lines cross all the zeros of the table.

Alternatively : We can follow the standard procedure.

(i) Tick-mark the third and fifth rows which do not have assignment.

(ii) Tick-mark the column two which has zeros in the marked rows three and five.

(iii) Tick-mark the row one as it has an assignment in the marked column two.

(iv) Draw lines through unmarked rows (2 and 4) and marked column (2).

These three lines cross all zeros. Since, the number of lines is 3 and it is less than the order of the matrix, the solution is not optimal. We go to the next step.

Step 5 : Iteration For Optimal Solution

The smallest element among all uncovered elements is 15. Subtract it from all uncovered elements and add it to the elements at the intersection of the lines. The result is shown in Table 3.

Table 3

	A	B	C	D	E
I	15	~~0~~	20	15	0
II	15	15	0	10	~~0~~
III	15	0	20	15	5
IV	0	15	20	~~0~~	5
V	5	~~0~~	10	0	~~0~~

Now, each row and column has zeros, we have thus, obtained second basic feasible solution.

Now, we go to step 2 and repeat the whole procedure.

Step 6 : To Assign Zeros

In the third row there is only one zero. Encircle it and cross all other zeros in the second column. Now, first row has a single uncrossed zero. Encircle it and cross all other zero in the fifth column. Second row has a single uncrossed zero. Encircle it. Fifth row has a single uncrossed zero. Encircle it and cross the other zero in the fourth column. There is a single uncrossed zero in the fourth row. Encircle it.

Step 7 : Optimality Check

Now each row and each column has only one encircled zero. Since there are four encircled zeros, equal to the order of matrix, the present solution is optimal.

The optimal assignment would be

$$\text{I} \xrightarrow{170} E,\quad \text{II} \xrightarrow{100} C,\quad \text{III} \xrightarrow{80} B,\quad \text{IV} \xrightarrow{20} A,\quad \text{V} \xrightarrow{50} D.$$

The total minimum cost = ₹ 170 + 100 + 80 + 20 + 50 = ₹ 420.

Example 3 : A car hire company has one car at each of five depots *a*, *b*, *c*, *d* and *e*. A customer in each of the five towns *A*, *B*, *C*, *D* and *E* requires a car. The distance (in miles) between the depots (origins) and the towns (destinations) where the customer are, is given in the following distance matrix.

	a	b	c	d	e
A	160	130	175	190	200
B	135	120	130	160	175
C	140	110	155	170	185
D	50	50	80	80	110
E	55	35	70	80	105

How should the cars be assigned to the customers so as to minimise the distance travelled ? **(S.U. 2004)**

Sol. : Step 1 : To Reduce The Matrix

The smallest elements in the first, second, third, fourth and fifth rows are 130, 120, 110, 50 and 35. Subtract them from all other elements in these rows respectively. Thus, we get the matrix given in Table 1.

Table 1

	a	b	c	d	e
A	30	0	45	60	70
B	15	0	10	40	55
C	30	0	45	60	75
D	0	0	30	30	60
E	20	0	35	45	70

Table 2

	a	b	c	d	e
A	30	0	35	30	15
B	15	~~0~~	0	10	~~0~~
C	30	~~0~~	35	30	20
D	0	~~0~~	20	~~0~~	5
E	20	~~0~~	25	15	15

All rows have zero elements but the third, fourth and fifth columns do not have zeros. Hence, subtract the smallest elements 10, 30 and 55 from all other elements of third, fourth and fifth columns respectively. Thus, we get the reduced matrix as shown in Table 2.

Step 2 : To Assign Zeros

First row has only one zero. Encircle it and cross all other zeros in its column. Now, second and fourth rows have tow unmarked zeros each. Leave them. Now, check the columns. First column has only one zero. Encircle it and cross the other zero in its row. Third column has only one zero. Encircle it and cross the other zero in its row.

Step 3 : To Find Minimum Number of Lines Which Will Cross All Zeros

Since the second column contains maximum number of zeros *i.e.* five zeros, draw a line through the second column. Now, second and fourth rows contain two zeros each. Draw lines through these zeros. Thus, we get the Table 2.

Since, the number of lines is three and the order of the matrix is five, the solution is not optimal.

Step 4 : Iteration For Optimal Solution

The smallest uncovered element is 15. It is to be subtracted from each uncovered element and is to be added to the element at the intersection of two lines. Thus, we get Table No. 3.

Table 3

	a	b	c	d	e
A	15	~~0~~	20	15	0
B	15	15	0	10	~~0~~
C	15	0	20	15	5
D	0	15	20	~~0~~	5
E	5	~~0~~	10	0	~~0~~

Step 5 : To Assign Zeros

Third row contains only one zero. Encircle it and cross all other zeros in its column. Now check columns. The first column contains only one zero. Encircle it and cross the other zero in its row. Third column contains a single zero. Encircle it and cross the other zero in its row. The fourth column contains a single zero. Encircle it and cross the other zero in its row. Now, the fifth column contains a single zero. Encircle it.

Step 6 : Optimality Check

Since all rows and columns have been assigned zeros, the solution is optimal.

The optimal assignment is

$$A \xrightarrow{200} e,\ B \xrightarrow{130} c,\ C \xrightarrow{110} b,\ D \xrightarrow{50} a,\ E \xrightarrow{80} d.$$

Hence, the total minimum distance

= 200 + 130 + 110 + 50 + 80 = 570 miles.

Example 4 : Solve the following minimal assignment problem.

	I	II	III	IV
A	17	19	16	18
B	12	17	14	15
C	19	21	20	18
D	15	22	18	16

Sol. : Write the steps taken to obtain the following tables.

Table 1

	I	II	III	IV
A	1	3	0	2
B	0	5	2	3
C	1	3	2	0
D	0	7	3	1

Table 2

	I	II	III	IV	
A	1	~~0~~	0	2	
B	0	2	2	3	✓
C	1	0	2	~~0~~	
D	~~0~~	4	3	1	✓

Optimal Assignment is

$$A \xrightarrow{16} \text{III},\quad B \xrightarrow{12} \text{I},$$
$$C \xrightarrow{21} \text{II},\quad D \xrightarrow{16} \text{IV}.$$

Total minimum cost

= ₹ 16 + 12 + 21 + 16

= ₹ 65.

Table 3

	I	II	III	IV
A	2	~~0~~	0	2
B	0	1	1	2
C	2	0	2	~~0~~
D	~~0~~	3	2	0

Example 5 : Solve the following minimal assignment problem. (S.U. 2003)

	1	2	3	4
A	15	11	13	15
B	17	12	12	13
C	14	15	10	14
D	16	13	11	17

Sol. : Write the steps taken to obtain the following tables.

Table 1

	1	2	3	4
A	4	0	2	4
B	5	0	0	1
C	4	5	0	4
D	5	2	0	6

Table 2

	1	2	3	4
A	3	0	2	3
B	4	0	0	0
C	3	5	0	3
D	4	2	0	5

Table 3

	1	2	3	4
A	~~0~~	0	2	3
B	1	~~0~~	~~0~~	0
C	0	5	~~0~~	3
D	1	2	0	5

[**Ans. :** Optimal Assignment is

$$A \xrightarrow{11} 2,\ B \xrightarrow{13} 4,\ C \xrightarrow{14} 1,\ D \xrightarrow{11} 3$$

Total = 49]

Example 6 : Five jobs are to be assigned to five persons, find the optimal assignment so that total cost is minimum.

Persons	Jobs 1	2	3	4	5
A	8	4	2	6	1
B	0	9	5	5	4
C	3	8	9	2	6
D	4	3	1	0	3
E	9	5	8	9	5

(S.U. 2003)

Sol. : Write the steps taken to obtain the following tables.

Table 1

	1	2	3	4	5
A	7	3	1	5	0
B	0	9	5	5	4
C	1	6	7	~~0~~	4
D	4	3	1	0	3
E	4	0	3	4	~~0~~

Table 2

	1	2	3	4	5
A	8	3	1	6	0
B	0	8	4	5	3
C	1	5	6	0	3
D	4	2	0	~~0~~	2
E	5	0	3	5	~~0~~

[**Ans. :** $A \to 5$, $B \to 1$, $C \to 4$, $D \to 3$, $E \to 2$, Total = 14]

EXERCISE - II

Solve the following minimal assignment problems.

1.

	I	II	III	IV
A	12	14	11	13
B	7	12	9	10
C	14	16	15	13
D	10	17	13	11

[**Ans. :** $A \to$ III, $B \to$ I, $C \to$ II, $D \to$ IV ; Total cost = ₹ 45]

2.

	A	B	C	D
I	17	25	24	22
II	30	19	29	23
III	28	29	31	26
IV	19	30	28	24

[**Ans. :** I $\to C$, II $\to B$, III $\to D$, IV $\to A$; Total cost = ₹ 68]

3.

	P	Q	R	S	T
A	7	9	10	8	14
B	10	6	9	12	8
C	11	8	7	9	12
D	12	15	10	14	13
E	10	8	12	11	14

[**Ans. :** $A \to P$, $B \to T$, $C \to S$, $D \to R$, $E \to Q$; Total = 42.]

4. Four different jobs can be completed on four different machines with the cost of production as given in the matrix on the right. Assign the jobs to the machines, so that the total cost of production is minimum.

Jobs \ Machines	A	B	C	D
I	6	8	12	7
II	9	6	10	7
III	5	8	11	8
IV	11	5	9	4

[**Ans. :** I $\to A$, II $\to B$, III $\to C$, IV $\to D$; Total cost = ₹ 27.]

5. In a factory there are five machines and five jobs to be done on these machines. The time in hours required by the machine to finish a job is given in the following matrix. Assign the machines to the jobs so as to minimise the total machine hours.

Machines \ Jobs	I	II	III	IV	V
1	11	6	14	16	17
2	7	13	22	7	10
3	10	7	2	2	2
4	4	10	8	6	11
5	13	15	16	10	18

[**Ans. :** $1 \to$ II, $2 \to$ V, $3 \to$ III, $4 \to$ I, $5 \to$ IV ; Minimum Time = 32 hours.]

6. A company has five jobs to be finished and there are five machines on which these jobs can finished. The following matrix shows the time required to finish a particular job on a particular machine. Assign the five jobs to five machines so that the total time required is minimum.

Machines \ Jobs	1	2	3	4	5
I	11	17	8	16	20
II	9	7	12	6	15
III	13	16	15	12	16
IV	21	24	17	28	26
V	14	10	12	11	15

[**Ans. :** I $\to$ 1, II $\to$ 4, III $\to$ 5, IV $\to$ 3, V $\to$ 2 ; Minimum time = 60 hours.]

7. Solve the following minimal assignment problem.

	A	B	C	D	E
I	12	18	9	17	21
II	10	8	13	7	16
III	14	17	16	13	17
IV	22	25	18	29	27
V	15	11	13	12	14

[**Ans. :** I $\to A$, II $\to B$, III $\to D$, IV $\to C$, V $\to E$; Minimum value = 65.]

8. There are five busses at stations A, B, C, D and E and they are required at stations X, Y, Z, U and V. How these busses be assigned their destinies so as to minimise the cost of transport as shown in the following table ?

	X	Y	Z	U	V
A	11	6	10	19	12
B	14	10	7	13	15
C	4	3	5	5	6
D	19	10	13	18	16
E	12	7	15	20	11

[**Ans. :** $A \to X$, $B \to Z$, $C \to U$, $D \to Y$, $E \to V$; Total cost = ₹ 44.]

9. There are five machines and five jobs that can be finished on these machines. The time required is given in the following table. Assign the jobs to the machines so that the total time is minimum.

Machines

Jobs	I	II	III	IV	V
1	5	3	4	7	1
2	2	3	7	6	5
3	4	1	5	2	4
4	6	8	1	2	3
5	4	2	5	7	1

[**Ans. :** I → 5, II → 1, III → 4, IV → 3, V → 2 ; Total time = 8 hours.]

Solve the following minimal assignment problems (10 to 13).

10.

	A	B	C	D
P	15	13	14	17
Q	11	12	15	13
R	13	12	10	11
S	15	17	14	16

[**Ans. :** $P \to B$, $Q \to A$, $R \to D$, $S \to C$; Total cost = Rs. 49.]

11.

	A	B	C	D	E
P	11	6	14	16	17
Q	4	10	19	14	7
R	11	8	3	3	3
S	8	12	10	8	13
T	8	10	11	5	13

[**Ans. :** $P \xrightarrow{6} B$, $Q \xrightarrow{4} A$, $R \xrightarrow{3} E$, $S \xrightarrow{10} C$, $T \xrightarrow{5} D$.

Total = 28]

12.

	A	B	C	D	E
P	80	70	60	120	70
Q	85	73	61	127	73
R	70	61	52	109	64
S	75	67	55	115	67
T	71	59	51	107	63

[**Ans. :** $P \xrightarrow{70} E$, $Q \xrightarrow{61} C$, $R \xrightarrow{109} D$, $S \xrightarrow{75} A$, $T \xrightarrow{59} B$.

Total = 374]

13.

Jobs

Operators	I	II	III	IV
P	10	28	19	13
Q	15	30	6	28
R	40	21	20	17
S	21	28	26	12

[**Ans. :**

$P \xrightarrow{10}$ I, $Q \xrightarrow{6}$ III, $R \xrightarrow{21}$ II, $S \xrightarrow{12}$ IV.

Total = 49]

6. Miscellaneous Problems

We shall now consider some special types of assignment problems. They are -

(i) Assignment Problems With Alternate Optimal Solutions.

(ii) Unbalanced Assignment Problems.

(iii) Maximisation Problems.

(iv) Assignment Problems With Restrictions.

(i) Assignment Problems With Alternate Optimal Solutions

In some problems we may get more than one set of assignments. But these alternate solutions lead us to the same result.

Example 1 : Solve the following minimal assignment problem whose effectiveness matrix is given below.

	A	B	C	D
P	10	14	14	18
Q	14	16	18	20
R	20	22	24	28
S	14	16	20	20

Sol. : Step 1 : To Reduce The Matrix

Subtract the smallest elements 10, 14, 20, 14 in the first, second, third and fourth rows from the other elements in the respective rows and get Table No. 1.

Table 1

	A	B	C	D
P	0	4	4	8
Q	0	2	4	6
R	0	2	4	8
S	0	2	6	6

Table 2

	A	B	C	D
P	0	2	~~0~~	2
Q	~~0~~	0	~~0~~	~~0~~
R	~~0~~	~~0~~	0	2
S	~~0~~	~~0~~	2	0

In this table each row has a zero but no column except the first has a zero.

So subtract the smallest elements 2, 4 and 6 from the elements of the second, third and fourth columns. Thus, we get Table No. 2.

Step 2 : To Assign Zeros

No row or no column has a single zero. Hence, encircle the first zero in the first row and cross all other zeros in its column. Encircle the first zero in the second row and cross all other zeros in its column. Encircle the uncrossed zero in the third row and cross all other zeros in its column. Encircle the uncrossed zero in the fourth column and cross the other zero in its column.

Step 3 : Optimality Check

Now every row and column has one single encircled zero. Since there are 4 encircled zeros equal to the order of matrix 4, the present solution is optimal.

The optimal assignment policy would be -

$$P \xrightarrow{10} A,\ Q \xrightarrow{16} B,\ R \xrightarrow{24} C,\ \text{and}\ S \xrightarrow{20} D.$$

The optimal minimum cost = ₹ 10 + 16 + 24 + 20 = ₹ 70.

Alternative Assignments

In step 2, instead of encircling the first zero in the first row, we may encircle the other zero from that row and follow the remaining procedure. In this way, we get alternative assignments.

Table 3

	A	B	C	D
P	~~0~~	2	0	2
Q	~~0~~	0	~~0~~	~~0~~
R	0	~~0~~	~~0~~	2
S	~~0~~	~~0~~	2	0

Step 4 : To Assign Zeros

Encircle the second zero in the first row and cross other zeros in that column. Encircle the second zero in the second row and cross all other zeros in that column. Encircle the uncrossed zero in the third row and cross all other zeros in that column. Encircle the uncrossed zero in the fourth row and cross the other zero in that column.

Thus, we get an alternative assignment as follows :

$$P \xrightarrow{14} C,\ Q \xrightarrow{16} B,\ R \xrightarrow{20} A,\ \text{and}\ S \xrightarrow{20} D.$$

The optimal cost = 14 + 16 + 20 + 20 = ₹ 70.

There are other alternative solutions also. They are

(i) $P \xrightarrow{14} C,\ Q \xrightarrow{14} A,\ R \xrightarrow{22} B,\ \text{and}\ S \xrightarrow{20} D.$

Total = ₹ 70

(ii) $P \xrightarrow{14} C,\ Q \xrightarrow{20} D,\ R \xrightarrow{22} B,\ \text{and}\ S \xrightarrow{14} A.$

Total = ₹ 70.

Example 2 : Solve the following minimal assignment problem.

	I	II	III	IV
A	5	6	7	8
B	7	8	9	10
C	10	11	12	11
D	6	8	11	7

Sol. : Step 1 : To Reduce The Matrix

Subtract the smallest elements 5, 7, 10, 6 from the other elements of the four rows successively. We get Table 1.

Now, subtract the smallest elements 1, 2, 1 from other elements of 2nd, 3rd and 4th columns. We get Table 2.

Table 1

	I	II	III	IV
A	0	1	2	3
B	0	1	2	3
C	0	1	2	1
D	0	2	5	1

Table 2

	I	II	III	IV
A	0	~~0~~	~~0~~	2
B	~~0~~	0	~~0~~	2
C	~~0~~	~~0~~	0	~~0~~
D	~~0~~	1	3	0

Step 2 : To Assign Zeros

Since, there are more than one zeros in each row and each column, we select the elements arbitrarily. Choose the first zero in the first row. Encircle it and cross all other zeros in the first column. Choose first (uncrossed) zero in the second row. Encircle it and cross all other zeros in the second column. Choose the first (uncrossed) zero in the third row. Encircle it and cross all other zeros in the third column. Choose the (uncrossed) zero in the fourth row. Encircle it and cross the other zero in the fourth column.

Step 3 : Optimality Check

Since each row and each column has encircled zero the assignment is optimal.

Thus, the assignment is :

$$A \xrightarrow{5} \text{I},\ B \xrightarrow{8} \text{II},\ C \xrightarrow{12} \text{III},\ D \xrightarrow{7} \text{IV}.$$

The total minimum cost = ₹ 5 + 8 + 12 + 7 = ₹ 32.

Alternative Assignments

If we choose a zero in the first row other than the first we would get alternate solutions as shown below.

Table 3

	I	II	III	IV
A	~~0~~	0	~~0~~	2
B	0	~~0~~	~~0~~	~~0~~
C	~~0~~	~~0~~	0	~~0~~
D	~~0~~	1	3	0

Suppose we choose the second zero in the first row, encircle it and cross the other zeros in the column of that zero. Suppose now we choose the first zero in the second row, encircle it and cross all other zeros in its column. Suppose we choose the first uncrossed zero in the third row and cross all other zeros of that column. Now, we choose the zero in the fourth row and cross all other zeros in the fourth column.

Thus, we have one zero in each row and in each column. The assignment is optimal.

Thus, another assignment is

$$A \xrightarrow{6} \text{II},\ B \xrightarrow{7} \text{I},\ C \xrightarrow{12} \text{III},\ D \xrightarrow{7} \text{IV}.$$

The total cost = ₹ 6 + 7 + 12 + 7 = ₹ 32.

The other alternative assignments are

(i) $A \xrightarrow{7} \text{III},\ B \xrightarrow{7} \text{I},\ C \xrightarrow{11} \text{II},\ D \xrightarrow{7} \text{IV}.$

Total = ₹ 32.

(ii) $A \xrightarrow{7} \text{III},\ B \xrightarrow{8} \text{II},\ C \xrightarrow{11} \text{IV},\ D \xrightarrow{6} \text{I}.$

Total = ₹ 32.

Example 3 : Solve the following assignment problem.

	A	B	C	D
I	10	8	7	13
II	12	14	7	11
III	11	9	10	12
IV	10	12	12	13

Sol. : Step 1 : To Reduce The Matrix

Subtract smallest elements 7, 7, 9, 10 in the first, second, third and fourth rows, from the other elements in the respective rows and get Table No. 1. In this table all rows have zeros, first three columns have also zeros. But the fourth column has no zero in it. So subtract the least element 3 from the remaining elements of that column and get the Table No. 2.

Table 1

	A	B	C	D
I	3	1	0	6
II	5	7	0	4
III	2	0	1	3
IV	0	2	2	3

Table 2

	A	B	C ✓	D	
I	3	1	0	3	✓
II	5	7	~~0~~	1	✓
III	2	0	1	~~0~~	
IV	0	2	2	~~0~~	

Step 2 : To Assign Zeros

First row has only one zero. Encircle it. There is another zero in the third column. Cross it. There are two zeros in the fourth and fifth rows. Leave them. Now consider columns. First column has only one zero. Encircle it and cross the other zero in the fourth row. Second column has only one zero. Encircle it and cross the other zero in the third row.

Step 3 : Optimality Test

The second row and fourth column are not assigned. The current assignment is not optimal.

Step 4 : To Find The Minimum Number of Lines Crossing Zeros

We can draw 'minimum" number of lines by judgement. Since there are two zeros in the third row, we draw a line through zeros of the third row. Similarly, since there are two zeros in the fourth row, we draw a line through the zeros of the fourth row. Since there is one zero in the third column, we draw a line through the zero of the third column. Thus, we get three lines, less than the order of the matrix. Hence, the solution is not optimal.

Step 5 : Iteration For Optimal Solution

The smallest uncovered element is 1. Subtract it from all uncovered elements and add it to the elements at the intersection of the lines. See Table 3.

Table 3

	A	B	C	D
I	2	0	~~0~~	2
II	4	6	0	~~0~~
III	2	~~0~~	2	0
IV	0	2	3	~~0~~

Table 4

	A	B	C	D
I	2	~~0~~	0	2
II	4	6	~~0~~	0
III	2	0	2	~~0~~
IV	0	2	3	~~0~~

Step 6 : To Assign Zeros

No row has a single zero. So check columns. The first column has a single zero encircle it and cross the other zero in its row. Encircle the first zero in the second column, and cross the other zero in its row. Encircle the unmarked zero in the third column and cross the other zero in its row. Encircle the unmarked zero in the fourth column and cross the other zero in its row.

Step 7 : Optimality Check

Since each row and each column has a zero, the solution is optimal.

The assignments will be

$$\text{I} \xrightarrow{8} B,\quad \text{II} \xrightarrow{7} C,\quad \text{III} \xrightarrow{19} D,\quad \text{IV} \xrightarrow{10} A.$$

Total minimum cost = ₹ 8 + 7 + 12 + 10 = ₹ 37.

Alternative Solutions

In the above example if we choose arbitrarily third row instead of first, we get another solution. See Table 4.

Encircle the first zero in the third row and cross the other zero in the second column. There is a single zero in the first row. Encircle it and cross the other zero in the third column. Encircle the uncrossed zero in the second row and cross the two zeros in the fourth column. Finally, encircle the uncrossed zero in the first column.

Thus, we get another optimum solution.

$$\text{I} \xrightarrow{7} C,\quad \text{II} \xrightarrow{11} D,\quad \text{III} \xrightarrow{9} B,\quad \text{IV} \xrightarrow{10} A.$$

Total = ₹ 7 + 11 + 9 + 10 = 37.

Example 4 : The *XYZ* company has 5 jobs *L, M, N, O, P* to be done and 5 men *A, B, C, D, E* to do these jobs. The number of hours each man would take to accomplish each job is given by the following table. Obtain optimal assignment, so as to minimise the total man-hours. **(S.U. 2005)**

Men \ Jobs	L	M	N	O	P
A	16	13	17	19	20
B	14	12	13	16	17
C	14	11	12	17	18
D	5	5	8	8	11
E	5	3	8	8	10

Sol. : Write the steps taken to obtain the tables.

Table 1

	L	M	N	O	P
A	3	0	4	6	7
B	2	0	1	4	5
C	3	0	1	6	7
D	0	0	3	3	6
E	2	0	5	5	7

Table 2

	L	M	N	O	P
A	3	0	3	3	2
B	2	~~0~~	0	1	~~0~~
C	3	~~0~~	~~0~~	3	2
D	0	~~0~~	2	~~0~~	1
E	2	~~0~~	4	2	2

Table 3

	L	M	N	O	P
A	1	0	1	1	~~0~~
B	2	2	~~0~~	1	0
C	3	2	0	3	2
D	0	2	2	~~0~~	1
E	~~0~~	~~0~~	2	0	~~0~~

1. $A \xrightarrow{13} M$, $B \xrightarrow{17} P$, $C \xrightarrow{12} N$, $D \xrightarrow{5} L$, $E \xrightarrow{8} O$.

2. Alternatively
$A \xrightarrow{13} M$, $B \xrightarrow{17} P$, $C \xrightarrow{12} N$, $D \xrightarrow{8} O$, $E \xrightarrow{5} L$.

In both cases, Minimum Total Man Hours
$= 13 + 17 + 12 + 5 + 8 = 55$ hours.

EXERCISE - III

Solve the following minimal assignment problem whose effectiveness matrix is given below.

1.

	A	B	C	D
P	9	7	6	12
Q	11	13	6	10
R	10	8	9	11
S	9	11	11	12

[Ans.: (i) $P \to B$, $Q \to C$, $R \to D$, $S \to A$. Total = 17.]
(ii) $P \to C$, $Q \to D$, $R \to B$, $S \to A$. Total = 17.]

2.

	A	B	C	D
P	12	10	9	15
Q	14	16	9	13
R	13	11	12	14
S	12	14	14	15

[Ans.: (i) $S \xrightarrow{12} A$, $R \xrightarrow{14} D$, $Q \xrightarrow{9} C$, $P \xrightarrow{10} B$. Total = 45.
(ii) $S \xrightarrow{12} A$, $R \xrightarrow{11} B$, $Q \xrightarrow{13} D$, $P \xrightarrow{9} C$. Total = 45.]

3. There are four machines in a factory and four jobs to be completed on them. The costs of completing the jobs on the machines are given in the following table.

Machines

Jobs	I	II	III	IV
A	4	5	6	7
B	6	7	8	9
C	9	10	11	10
D	5	7	10	6

Assign the jobs to the machines, so that the cost is minimum.

[Ans.: (i) $A \xrightarrow{4}$ I, $B \xrightarrow{7}$ II, $C \xrightarrow{11}$ III, $D \xrightarrow{6}$ IV.; Total = 28
(ii) $A \xrightarrow{5}$ II, $B \xrightarrow{8}$ III, $C \xrightarrow{10}$ IV, $D \xrightarrow{5}$ I.; Total = 28]

4. There are five workers who can do five jobs. The time required in hours to complete the job is given in the following matrix. Assign the five workers to five jobs such that the total time is minimum.

Job

Workers	1	2	3	4	5
I	2	9	2	7	1
II	6	8	7	6	1
III	4	6	5	3	1
IV	4	2	7	3	1
V	5	3	9	5	1

[Ans.: There are three alternative assignments.
(i) I → 3, II → 5, III → 1, IV → 4, V → 2
(ii) I → 3, II → 5, III → 4, IV → 2, V → 1
(iii) I → 3, II → 5, III → 4, IV → 1, V → 2
Minimum time = 15 hours.]

5. A project work consists of four jobs for which four companies have submitted the tenders. The tender amounts in thousands of rupees are given in the following matrix.
Solve the assignment problem so that total cost is minimum.

	A	B	C	D
I	100	240	300	150
II	160	220	280	120
III	120	200	320	100
IV	90	260	340	160

[Ans.: There are four alternative assignments
(i) I → C, II → B, III → D, IV → A
(ii) I → C, II → D, III → B, IV → A
(iii) I → B, II → C, III → B, IV → A
(iv) I → B, II → C, III → D, IV → A
Total Minimum cost = ₹ 710000.]

(ii) Unbalanced Assignment Problems

We have seen that to solve an assignment problem the number of rows and the number of columns are required to be equal *i.e.* the cost matrix must be a square matrix. If in the cost matrix if a row or a column is less than the requirement, we assume such a row or such a column with zero costs in each added cell. Now we follow the same procedure, since the cost-matrix now is a square matrix.

Example 1 : A city corporation has decided to carry out total repairs on main arteries of the city. The government has agreed to make a special grant of ₹ 50 lakhs towards the cost with the condition that the repairs must be done at the lowest cost and the quickest time. If conditions warrant, then a supplementary token grant will also be considered favourably. The corporation has floated tender and 5 contractors have sent their bids. In order to expedite work, one road will be awarded to any one contractor.

Road Cost of Repairs (Rs. lakhs)

		R_1	R_2	R_3	R_4
Contractors	C_1	9	14	19	15
	C_2	7	17	20	19
	C_3	9	18	21	18
	C_4	10	12	18	19
	C_5	10	15	21	16

(i) Find the best way of assigning the repair work to the contractors and the costs.

(ii) If it is necessary to seek supplementary grants what should be the amount sought ?

(iii) Which of the five contractors will be unsuccessful in his bid ? **(S.U. 2005)**

Sol. : Since there are four roads and five contractors, the cost-matrix is **not balanced**. So we add one dummy road R_5 with zero cost in that column.

	R_1	R_2	R_3	R_4	R_5
C_1	9	14	19	15	0
C_2	7	17	20	19	0
C_3	9	18	21	18	0
C_4	10	12	18	19	0
C_5	10	15	21	16	0

Step 1 : To Reduce The Matrix

Find the smallest element in each row and subtract it from all other elements of that row. Since the smallest element in each row is zero by subtraction we will get the same matrix. We subtract the smallest element in each of the four **columns** from all other elements in the **corresponding columns**. The smallest elements in the first, second, third and fourth columns are 7, 12, 18, 15. These are subtracted from all other elements of the corresponding columns. Thus, we get the Table 1.

Step 2 : To Assign Zeros

The third row has only one zero. Encircle it and cross all other zeros in its column. Now, the first and the second rows have a single zero each. Encircle them. The third row has two zeros. Leave them. Now, check columns. The second column has a single zero. Encircle it and cross the zero in its row.

Table 1

	R_1	R_2	R_3	R_4	R_5
C_1	2	2	1	0	~~0~~
C_2	0	5	2	4	~~0~~
C_3	2	6	3	3	0
C_4	3	0	~~0~~	4	~~0~~
C_5	3	3	3	1	~~0~~

Step 3 : Optimality Check

Since the fifth row and the third column do not have assignments *i.e.* encircled zeros, the assignment is not optimal.

Step 4 : To Draw Minimum Number of Lines To Cover All Zeros

Since the fifth column contains maximum (five) zeros. Draw a line through that column. Since fourth row contains two zeros. Draw a line through that row. Now, draw lines through the first and the second rows.

Since, the number of lines is four and the order of the matrix is five, the solution is not optimal. So we go for next iteration.

Step 5 : Iteration For Optimal Solution

The smallest uncovered element is 1. It is to be subtracted from all uncovered elements. This smallest element 1 is to be added to every element which lies at the intersection of two lines. Thus, we get Table 2.

Table 2

	R_1	R_2	R_3	R_4	R_5
C_1	2	2	1	0	1
C_2	0	5	2	4	1
C_3	1	5	2	2	0
C_4	3	0	0	4	1
C_5	2	2	2	0	0

Table 3

	R_1	R_2	R_3	R_4	R_5
C_1	1	1	0	~~0~~	1
C_2	0	5	2	5	2
C_3	~~0~~	4	1	2	0
C_4	3	0	~~0~~	5	2
C_5	1	1	1	0	~~0~~

Step 6 : To Assign Zeros

Since only second row has a single zero, encircle it and cross the other zero in its column. Now, the third row has only one unmarked zero, encircle it and cross the other zero in its column. The fifth row has a single unmarked zero, encircle it and cross the other zero in its column. Now, the first row has a single unmarked zero, encircle it and cross the other zero in its column. There remains only one unmarked zero in the fourth row, encircle it.

Step 7 : Optimality Check

Since each row and each column has an assigned zero, the solution is optimal.

(i) The optimal assignment is

$$C_1 \xrightarrow{19} R_3,\ C_2 \xrightarrow{7} R_1,\ C_3 \xrightarrow{0} R_5,\ C_4 \xrightarrow{12} R_2,\ C_5 \xrightarrow{16} R_4.$$

Total minimum cost = ₹ 54 lakhs.

(ii) Since the cost exceeds 50 lakhs by 4 lakhs, the excess amount of 4 lakhs is to be sought as supplementary grant.

(iii) The contractor C_3 who is assigned dummy road R_5 loses the bid.

Example 2 : Solve the following minimal assignment problem.

	A	B	C
P	3	6	5
Q	5	4	6
R	6	1	7
S	3	3	5

Sol. : Write the steps taken to obtain the following table.

Table

	A	B	C	D
P	0	5	~~0~~	~~0~~
Q	2	3	1	0
R	3	0	2	~~0~~
S	~~0~~	2	0	~~0~~

$$P \xrightarrow{3} A,\ Q \xrightarrow{0} D,\ R \xrightarrow{1} B,\ S \xrightarrow{5} C.$$

Total = 9.

Example 3 : Assign four vehicles A, B, C, D to vacant places P, Q, R, S, T such that the total distance travelled is minimum.

	A	B	C	D
P	5	8	4	8
Q	9	3	6	6
R	5	10	7	10
S	8	6	5	9
T	7	4	6	5
U	7	9	8	4

Sol. : Write the steps taken to obtain the following table.

Table

	A	B	C	D	E	F
P	~~0~~	5	0	4	~~0~~	~~0~~
Q	4	0	2	2	~~0~~	~~0~~
R	0	7	3	6	~~0~~	~~0~~
S	3	3	1	5	~~0~~	~~0~~
T	2	1	2	1	~~0~~	~~0~~
U	2	6	4	0	~~0~~	~~0~~

$P \xrightarrow{4} C,\quad Q \xrightarrow{3} B,\quad R \xrightarrow{5} A,\quad U \xrightarrow{4} D.$ Total = 16.

Example 4 : A department-head has six jobs and five sub-ordinates. The subordinates differ in their efficiency and the jobs differ in their intrinsic difficulty. The department head estimates that the time each man would take to perform each task is as given in the effectiveness matrix below.

Jobs

		A	B	C	D	E	F
	1	20	15	26	40	32	12
	2	15	32	46	26	28	20
Men	3	11	15	2	12	6	14
	4	8	24	12	22	22	20
	5	12	20	18	10	22	15

Only one job can be assigned to one man.

Determine how the jobs should be assigned so as to minimise the total man-hours.

Find the minimum total hours. **(S.U. 2004)**

Sol. : Write the steps taken to obtain the following tables.

Table 1

	A	B	C	D	E	F
1	20	15	26	40	32	12
2	15	32	46	26	28	20
3	11	15	2	12	6	14
4	8	24	12	22	22	20
5	12	20	18	10	22	15
6	0	0	0	0	0	0

Table 2

	A	B	C	D	E	F
1	8	3	14	28	20	0
2	0	17	31	11	13	5
3	9	13	0	10	4	12
4	0	16	4	14	14	12
5	2	10	8	0	12	5
6	0	0	0	0	0	0

Table 3

	A	B	C	D	E	F
1	12	3	18	32	20	0
2	0	13	31	11	9	1
3	9	9	0	10	0	8
4	0	12	4	14	10	8
5	2	6	8	0	8	1
6	4	0	4	4	0	0

Table 4

	A	B	C	D	E	F
1	12	~~0~~	15	29	17	0
2	0	10	28	8	6	1
3	12	9	0	10	~~0~~	11
4	~~0~~	9	1	11	7	8
5	5	6	8	0	5	4
6	7	0	4	4	~~0~~	4

Table 5

	A	B	C	D	E	F
1	13	~~0~~	15	29	17	0
2	0	9	27	7	5	~~0~~
3	13	9	~~0~~	10	0	11
4	~~0~~	8	0	10	6	7
5	6	6	8	0	5	4
6	8	0	4	4	~~0~~	4

(i) $1 \xrightarrow{12} F,\ 2 \xrightarrow{15} A,\ 3 \xrightarrow{6} E,\ 4 \xrightarrow{12} C,\ 5 \xrightarrow{10} D.$

(ii) $1 \xrightarrow{15} B,\ 2 \xrightarrow{20} A,\ 3 \xrightarrow{2} C,\ 4 \xrightarrow{8} A,\ 5 \xrightarrow{10} D.$

Total Minimum Time = 55 hours.

EXERCISE - IV

1. There are four machines and three jobs to be done. The cost of doing a job on a machine is given in the matrix on the right. Solve the assignment problem, so that the cost is minimum.

Jobs \ Machines	P	Q	R	S
X	20	26	30	34
Y	10	15	19	21
Z	12	17	21	24

[**Ans. :**
(i) $X \xrightarrow{20} P$, $Y \xrightarrow{15} Q$, $Z \xrightarrow{21} R$;
Total = 56
(ii) $X \xrightarrow{20} P$, $Y \xrightarrow{19} R$, $Z \xrightarrow{17} Q$; Total = 56.]

2. The cost of producing jobs at different factories *A, B, C, D* are as given in the adjoining matrix.

Assign four jobs out of six to the four factories so that the total cost is minimum.

	A	B	C	D
I	3	6	2	6
II	7	1	4	4
III	3	8	5	8
IV	6	4	3	7
V	5	2	4	3
VI	5	7	6	2

[**Ans. :** I → *C*, II → *B*, III → *A*, IV → 6 ; Total cost = 8.]

3. The S.T. bus depots *A, B, C, D, E, F* are in deficit of one bus each and the depots *P, Q, R, S, T* are in surplus of one base each. The distances between the depots are given in the following matrix. How the busses be assigned from surplus depots to deficit depots so that the total distance covered will be minimum.

Surplus Depots \ Deficit Depots	A	B	C	D	E	F
P	13	11	16	23	19	9
Q	11	19	26	16	17	13
R	12	11	4	9	6	10
S	7	15	11	14	14	13
T	9	13	12	8	14	11

[**Ans. :** *P* → *B*, *Q* → *F*, *R* → *C*, *S* → *A*, *T* → *D* ; Minimum distance travelled = 43 km.]

4. Solve the following assignment problem for minimum cost.

Machines \ Jobs	1	2	3	4	5
A	5	10	2	12	4
B	4	10	3	14	6
C	6	13	4	16	6
D	7	14	4	18	9
E	8	14	6	18	12
F	12	18	10	20	12

[**Ans. :** *A* → 4, *B* → 1, *C* → 5, *D* → 3, *E* → 2 ; Total = 40]

5. There are four buses *A, B, C, D* to be assigned to four out of the six places *P, Q, R, S, T* shown below with their distances so that the total distance travelled is minimum.

	A	B	C	D
P	8	14	6	14
Q	16	4	10	10
R	8	18	12	18
S	14	10	8	16
T	12	6	10	8
U	12	16	14	6

[**Ans. :** $P \xrightarrow{6} C$, $Q \xrightarrow{4} B$, $R \xrightarrow{8} A$, $U \xrightarrow{6} D$; Total = 24]

(iii) Maximisation Problem

If the given problem is to maximise the profit, we have to reduce it to minimisation type. This can be done in two ways.

(a) We may multiply each element of the matrix by (–1). Then the maximisation problem becomes the minimisation problem.

(b) We may subtract all elements of the matrix from the biggest element of the matrix.

For instance, consider the following matrix.

$$\begin{bmatrix} 20 & 16 & 5 \\ 12 & 18 & 9 \\ 6 & 14 & 15 \end{bmatrix} \quad \begin{bmatrix} -20 & -16 & -5 \\ -12 & -18 & -9 \\ -6 & -14 & -15 \end{bmatrix} \quad \begin{bmatrix} 0 & 4 & 15 \\ 6 & 0 & 9 \\ 9 & 1 & 0 \end{bmatrix}$$

Changing the signs of all elements, we get the middle matrix and subtracting the least element of each row from other elements, we get the matrix on the right.

$$\begin{bmatrix} 0 & 4 & 15 \\ 8 & 2 & 11 \\ 14 & 6 & 5 \end{bmatrix} \quad \begin{bmatrix} 0 & 4 & 15 \\ 6 & 0 & 9 \\ 9 & 1 & 0 \end{bmatrix}$$

If we subtract all elements of the given matrix from the largest element 20, we get the above matrix on the left. If we now as usual subtract the smallest elements from other elements in this matrix in each row, we get the matrix on the right.

The result in both cases is the same.

Example 1 : Solve the following assignment problem so as to maximise the total profit from the following table which gives the profit earned when different jobs are completed on different machines.

Jobs \ Machines	1	2	3	4
A	20	14	18	15
B	18	15	19	19
C	19	19	17	16
D	17	16	18	19

Sol. : The given problem is of maximisation type. We have to convert it into minimisation type. This can be done in two ways. Either by changing the sign of every element of the matrix or by subtracting every element of the matrix from the largest element.

Method 1 : We change the sign of each element of the matrix. Thus, we have the matrix.

−20	−14	−18	−15
−18	−15	−19	−19
−19	−19	−17	−16
−17	−16	−18	−19

Step 1 : To Reduce The Matrix

Find out the smallest element of each row and subtract it from all other elements of that row. The smallest elements in the first, second, third and fourth rows are – 20, – 19, – 19, – 19. By subtracting these elements from other elements of the corresponding rows, we get Table No. 1.

Table 1

	1	2	3	4
A	0	6	2	5
B	1	4	0	~~0~~
C	~~0~~	0	2	3
D	2	3	1	0

Step 2 : To Assign Zeros

The first row contains one zero. Encircle it and cross the other zero in its column. Second row has two zeros. Leave it. The third row has one zero. Encircle it. The fourth row has one zero. Encircle it and cross the other zero in its column. Now, check the columns. The third column has one unmarked zero. Encircle that zero.

Step 3 : Optimality Check

Since each row and each column has an encircled zero, the assignment is optimum.

Hence, the assignments will be

$$A \xrightarrow{20} 1,\ B \xrightarrow{19} 3,\ C \xrightarrow{19} 2,\ D \xrightarrow{19} 4.$$

∴ The maximum total profit = ₹ 77.

Method 2 : We now subtract all elements from the largest element. The largest element is 20. We subtract all other elements from 20. Thus, we get the following table.

0	6	2	5
2	5	1	1
1	1	3	4
3	4	2	1

Step 1 : To Reduce The Matrix

Find out the smallest element of each row and subtract it from all other elements of that row. The smallest elements in the first, second, third and fourth rows are 0, 1, 1 and 1. By subtracting these elements from other elements of the corresponding rows we get Table No. 1.

Table 1

	1	2	3	4
A	0	6	2	5
B	1	4	0	~~0~~
C	~~0~~	0	2	3
D	2	3	1	0

Note that this Table No. 1 is the same as the Table No. 1 obtained above in the first method. Hence, following the usual procedure we would get the same result.

The assignments will be -

$$A \xrightarrow{20} 1,\ B \xrightarrow{19} 3,\ C \xrightarrow{19} 2,\ D \xrightarrow{19} 4.$$

The maximum total profit = ₹ 77.

Example 2 : A marketing manager has 5 salesmen and 5 sales districts. Considering the capabilities of the salesmen and the nature of districts, the marketing manager estimates that sales per month (in hundred rupees) for each salesman in each district would be as follows.

		Sales Districts				
		A	B	C	D	E
Salesmen	1	32	38	40	28	40
	2	40	24	28	21	36
	3	41	27	33	30	37
	4	22	38	41	36	36
	5	29	33	40	35	39

Find the assignment of salesmen to districts that will result in maximum sales. **(S.U. 2005)**

Sol. : Method 1

Step 1 : To Reduce The Matrix

Since the given problem is a maximisation problem, we convert the given problem to minimisation by changing the sign of each element in the matrix, since max z = min (– z). Thus, we have the matrix.

−32	−38	−40	−28	−40
−40	−24	−28	−21	−36
−41	−27	−33	−30	−37
−22	−38	−41	−36	−36
−29	−33	−40	−35	−39

The smallest elements in the first, second, third, fourth and fifth rows are – 40, – 40, – 41, – 41, – 40. These are subtracted from all other elements of the corresponding rows. Thus, we get Table 1.

Table 1

	A	B	C	D	E
1	8	2	0	12	0
2	0	16	12	19	4
3	0	14	8	11	4
4	19	3	0	5	5
5	11	7	0	5	1

Table 2

	A	B	C	D	E
1	8	0	~~0~~	7	~~0~~
2	0	14	12	14	4
3	0	12	8	6	4
4	19	1	0	~~0~~	5
5	11	5	~~0~~	0	1

Second and fourth columns have no zeros. So subtract the smallest elements 2 and 5 from all other elements of the second and fourth columns respectively. Thus, we get Table No. 2. This is the reduced table.

Step 2 : To Assign Zeros

The second and third rows have single zeros. Encircle them. Now, check columns. The second column has a single zero encircle it and cross all other zeros from its row. Third column has a single zero. Encircle it and cross the zero from its row. Now, fourth column has a single unmarked zero. Encircle it and cross the other zero from its row. Thus, we get Table No. 2.

Step 3 : Optimality Check

Since the fifth column has no encircled zero, the solution is not optimal.

Step 4 : To Draw Minimum Number Of Lines Through All Zeros

Since there are three zeros in the first row, draw a line through the first row. Since there are two zeros each in the first column, fourth row and the fifth row, draw lines through them. Thus, we get Table No. 3.

There are three lines. But the order of the matrix is five. Hence, the solution is not optimal.

Step 5 : Iteration For The Optimal Solution

The smallest uncovered element is 4. It is to be subtracted from all uncovered elements and is to be added to every element which lies at the intersection of two lines. Thus, we get Table No. 4.

Table 3

	A	B	C	D	E
1	8	0	0	7	0
2	0	14	12	14	4
3	0	12	8	6	4
4	19	1	0	0	5
5	11	5	0	0	1

Table 4

	A	B	C	D	E
1	12	0	0	7	0
2	0	10	8	10	0
3	0	8	4	2	0
4	23	1	0	0	5
5	15	5	0	0	1

Step 6 : To Assign Zeros

No row has a single zero. So check columns. Column two has a single zero. Encircle it and cross other zeros in its row. First column has two zeros. Encircle the first zero and cross other zeros in its row. Last column has a single zero. Encircle it and cross the other zero in its row.

Table 5

	A	B	C	D	E
1	12	0	~~0~~	7	~~0~~
2	0	10	8	10	~~0~~
3	~~0~~	8	4	2	0
4	23	1	0	~~0~~	5
5	15	5	~~0~~	0	1

Encircle the first zero in the third column and cross the other zero in its row. Encircle the only zero in the fourth column and cross the other zero in its row.

Step 7 : Optimal Check

Now, each row (column) has only one encircled zero and hence the solution is optimal.

The allotments are

$$1 \xrightarrow{38} B,\ 2 \xrightarrow{40} A,\ 3 \xrightarrow{37} E,\ 4 \xrightarrow{41} C,\ 5 \xrightarrow{35} D.$$

Hence, the maximum sales $= ₹\ (38 + 40 + 37 + 41 + 35)$

$= ₹\ 191,00.$

Method 2

Alternatively : We may convert the maximisation problem into minimisation problem by subtracting each element from the largest element 41. Thus, we get the following Table No. 1.

Table 1

	A	B	C	D	E
1	9	3	1	13	1
2	1	17	13	20	5
3	0	14	8	11	4
4	19	3	0	5	5
5	12	8	1	6	2

Table 2

	A	B	C	D	E
1	8	2	0	12	0
2	0	16	12	19	0
3	0	14	8	11	4
4	19	3	0	5	5
5	11	7	0	5	1

Now, we follow the same procedure as before.

Step 1 : To Reduce The Matrix

Find out the smallest element in each row and subtract it from the other elements of that row. Thus, we get the Table No. 2.

This table is the same as the Table No. 2 obtained above. Following the usual steps we would get the same result.

Example 3 : A sales organisation has four zones *P, Q, R, S* and four sales managers *X, Y, Z, U* for assignment. The zones are not equally rich in sales potential. The estimates of sales in different zones are, Zone *P* ; ₹ 4,20,000, Zone *Q* ; ₹ 3,36,000, Zone *R* ; ₹ 2,94,000 and Zone *S* ; ₹ 4,62,000.

The sales managers have different capabilities. Their capacities in any zone can be considered to be in the ratio 14 : 9 : 11 : 8 respectively.

Assign the zones to the sales mangers, so that the sales are maximum.

Sol. : In zone *P* the total expected sale is ₹ 4,20,000. The sales due to the sales managers *X, Y, Z, U* are in the ratio 14 : 9 : 11 : 8. Hence, sales due to *X, Y, Z, U* in zone *P* would be as follows.

Due to X : $₹\ 4,20,000 \times \frac{14}{42} = 1,40,000$

Due to Y : $₹\ 4,20,000 \times \frac{9}{42} = 90,000$

Due to Z : $₹\ 4,20,000 \times \frac{11}{42} = 1,10,000$

Due to U : $₹\ 4,20,000 \times \frac{8}{42} = 80,000$

Similarly, in zone Q the sales due to X, Y, Z, U will be

$$3,36,000 \times \frac{14}{42} = 1,12,000,$$

$$3,36,000 \times \frac{9}{42} = 72,000,$$

$$3,36,000 \times \frac{11}{42} = 88,000,$$

$$3,36,000 \times \frac{8}{42} = 64,000 \text{ respectively.}$$

Similarly, we can calculate the expected sales (in thousands) in other zones due to the sales managers X, Y, Z, U and thus, we get the following table.

Table 1

Managers	Zones P	Q	R	S
X	140	112	98	154
Y	90	72	63	99
Z	110	88	77	121
U	80	64	56	88

Since this is a maximisation problem we have to convert it into minimisation problem. The highest sale is 154. We subtract every element from this element and obtain the cost matrix as shown below.

Table 2

	P	Q	R	S
X	14	42	56	0
Y	64	82	91	55
Z	44	66	77	33
U	74	90	98	66

Step 1 : To Reduce The Matrix

The smallest elements in the first, second, third and fourth rows are 0, 55, 33 and 66. By subtracting these elements from other elements of the corresponding rows, we get TableNo. 3.

Table 3

	P	Q	R	S
X	14	42	56	0
Y	9	27	36	0
Z	11	33	44	0
U	8	24	32	0

Table 4

	P	Q	R	S
X	6	18	24	0
Y	1	3	4	~~0~~
Z	3	9	12	~~0~~
U	0	~~0~~	~~0~~	~~0~~

Since all rows and all columns do not have zeros, we subtract, now, the smallest elements in each column from all other elements in that column. The smallest elements in the first, second, third and fourth columns are 9, 24, 32 and zero. By subtracting these from other elements of these columns, we get Table No. 4.

Step 2 : To Assign Zeros

The first row has one zero. Encircle it and cross all other zeros in its column. Now check columns. First column has a single zero. Encircle it and cross the other zeros in its row.

Step 3 : Optimality Check

Since each row and each column do not have a zero, this is not an optimum solution.

Step 4 : To Draw Minimum Number Of Lines

We can draw two lines passing through all zeros. Since the order of the matrix is four, this is not an optimum solution.

Step 5 : Iteration For The Optimum Solution

The smallest uncovered element is 1. It is to be subtracted from all uncovered elements and is to be added to the elements which lie at the intersection of two lines. Thus, we get Table No. 5.

Step 6 : To Assign Zeros

Since the first row contains only one zero, encircle it and cross all other zeros in its column. Now, second row has a single zero. Encircle it. Now, check columns. The second column has a single zero. Encircle it and cross the other zeros in its row.

Table 5

	P	Q	R	S
X	5	17	23	0
Y	0	2	3	~~0~~
Z	2	8	11	~~0~~
U	~~0~~	0	~~0~~	1

Table 6

	P	Q	R	S
X	3	15	21	0
Y	0	2	3	2
Z	~~0~~	6	9	~~0~~
U	~~0~~	0	~~0~~	3

Step 7 : To Draw Minimum Number of Lines

We can draw three lines which pass through all zeros. Since the number of lines is three and the order of the matrix is four, this is not an optimum solution.

Step 8 : Iteration For The Optimum Solution

The smallest uncovered element is two. It is to be subtracted from all uncovered elements and is to be added to the elements which lie at the intersection of the lines. Thus, we get Table No. 6.

Step 9 : To Assign Zeros

Since the first row contains only one zero, encircle it and cross the other zero in its column. Second row has a single zero. Encircle it and cross other zeros in its column. Now, check columns. The second column has a zero. Encircle it and cross the other zero in its row.

Step 10 : Optimality Check

Since all rows and columns do not have encircled zeros, the solution is not optimal.

Step 11 : To Draw Minimum Number of Lines Which Cross All Zeros

We can draw three lines which pass through all zeros. Since the order of the matrix is four this is not the optimum solution.

Step 12 : Iteration For The Optimum Solution

The smallest uncovered element is two. It is to be subtracted from all uncovered elements and is to be added to the elements which lie at the intersection of two lines. Thus, we get Table No. 7.

Table 7

	P	Q	R	S
X	3	13	19	0
Y	~~0~~	0	1	2
Z	0	4	7	~~0~~
U	2	~~0~~	0	5

Step 13 : To Assign Zeros

The first row has a single zero. Encircle it and cross the other zero in its column. Now, check columns. The third column has a single zero. Encircle it and cross the other zero in its row. Encircle the remaining zero in the second column. There are two zeros in the first column. Since the second row contains an encircled zero, leave the first zero in the first column and encircle the second.

Step 14 : Optimality Check

Since every row and every column has an encircled zero, the solution is optimal.

The optimum assignment is

$$X \xrightarrow{154} S,\quad Y \xrightarrow{72} Q,\quad Z \xrightarrow{110} P,\quad U \xrightarrow{56} R.$$

Total maximum sales = ₹ 154 + 72 + 110 + 56

= ₹ 392.

Example 4 : Five machines are to be allocated to five operators (one for each). The following table gives weekly output (in pieces).

	M_1	M_2	M_3	M_4	M_5
O_1	22	24	29	34	38
O_2	21	25	31	36	42
O_3	25	30	37	41	36
O_4	23	26	33	39	44
O_5	26	30	33	38	43

Profit per piece is ₹ 50. Find the maximum profit that can be earned per week.

Sol. : Write the steps taken to obtain the following tables.

Table 1

	M_1	M_2	M_3	M_4	M_5
O_1	22	20	15	10	6
O_2	23	19	13	8	2
O_3	19	14	7	3	8
O_4	21	18	11	5	0
O_5	18	14	11	6	1

Table 2

	M_1	M_2	M_3	M_4	M_5
O_1	16	14	9	4	0
O_2	21	17	11	6	0
O_3	16	11	4	0	5
O_4	21	18	11	5	0
O_5	17	13	10	5	1

Table 3

	M_1	M_2	M_3	M_4	M_5
O_1	0	3	5	4	~~0~~
O_2	5	6	7	6	0
O_3	~~0~~	0	~~0~~	~~0~~	5
O_4	5	7	7	5	~~0~~
O_5	1	2	6	5	1

Table 4

	M_1	M_2	M_3	M_4	M_5
O_1	0	1	3	2	~~0~~
O_2	5	4	5	4	0
O_3	2	~~0~~	0	~~0~~	7
O_4	5	5	5	3	~~0~~
O_5	1	0	4	3	1

Table 5

	M_1	M_2	M_3	M_4	M_5
O_1	0	1	1	~~0~~	0
O_2	5	4	3	2	0
O_3	4	2	0	~~0~~	9
O_4	5	5	3	1	~~0~~
O_5	1	0	2	1	1

Table 6

	M_1	M_2	M_3	M_4	M_5
O_1	~~0~~	2	2	0	1
O_2	4	4	3	1	0
O_3	4	3	1	0	10
O_4	4	5	3	~~0~~	~~0~~
O_5	~~0~~	0	2	~~0~~	1

Table 7

	M_1	M_2	M_3	M_4	M_5
O_1	0	2	2	1	2
O_2	3	3	2	1	0
O_3	3	2	0	~~0~~	10
O_4	3	4	2	0	~~0~~
O_5	~~0~~	0	2	1	2

$O_1 \xrightarrow{22} M_1,\quad O_2 \xrightarrow{42} M_5,$

$O_3 \xrightarrow{37} M_3,\quad O_4 \xrightarrow{39} M_4,$

$O_5 \xrightarrow{30} M_2.$

Total = 170.

Total Profit = 170 × 25 = ₹ 4250.

EXERCISE - V

1. The profit earned by completing various jobs on different machines is given in the following matrix.

Jobs \ Machines	1	2	3	4
X	17	11	15	12
Y	15	12	16	16
Z	16	16	14	13
U	14	13	15	16

Find the assignment of jobs to the machines, so that the profit is maximum.

[**Ans. :** $A \xrightarrow{17} 1$, $B \xrightarrow{16} 3$, $C \xrightarrow{16} 2$, $D \xrightarrow{16} 4$; Total = 65.]

2. The productivity of the operators *A*, *B*, *C* and *D* on different machines I, II, III and IV is shown in the following matrix. Assign machines to the operators so that the total productivity is maximum.

Operators \ Machines	I	II	III	IV
A	10	14	18	16
B	11	9	7	8
C	10	11	14	14
D	24	14	19	19

[**Ans. :** $A \to$ I, $B \to$ III, $C \to$ II, $D \to$ IV ; Total productivity = 47 units.]

3. There are six jobs and six machines on which these jobs can be completed. The returns in rupees from these jobs are given in the following matrix. Assign the jobs in such a way that the total returns are maximum.

Machines \ Jobs	1	2	3	4	5	6
I	9	22	58	11	19	27
II	43	78	72	50	63	48
III	41	28	91	37	45	33
IV	74	42	27	49	39	32
V	36	11	57	22	25	18
VI	3	56	53	31	17	28

[**Ans. :** (i) I $\to$ 4, II $\to$ 1, III $\to$ 6, IV $\to$ 3, V $\to$ 2, VI $\to$ 5,
(ii) I $\to$ 4, II $\to$ 6, III $\to$ 2, IV $\to$ 3, V $\to$ 5, VI $\to$ 1.
Maximum profit = ₹ 142.]

4. A sales organisation has four zones *P*, *Q*, *R*, *S* and four sales managers *X*, *Y*, *Z*, *U* for assignment. The zones are not equally rich in sales potential. The estimates of sales in different zones are :

Zone *P* : ₹ 1,26,000 ; Zone *Q* : ₹ 1,05,000
Zone *R* : ₹ 84,000 ; Zone *S* : ₹ 63,000.

The sales managers have different capabilities. The capacities of the sales managers are in the ratio 7 : 5 : 5 : 4 respectively.

Assign zones to the sales manages, so that the sales are maximum.

[**Ans. :** $X \xrightarrow{42} P$, $Y \xrightarrow{25} Q$, $Z \xrightarrow{20} R$, $U \xrightarrow{12} 5$; Total 99.]

5. There are five salesmen and five zones for sales. The expected sales in a year by different salesmen in different zones are given in the table below.

Salesmen \ Zones	A	B	C	D	E
P	34	40	42	30	42
Q	42	26	30	23	38
R	43	29	35	32	39
S	24	40	43	38	38
T	31	35	42	37	41

Find the assignment of zones to salesmen such that the total sales are maximum.

[**Ans. :** $P \xrightarrow{40} B$, $Q \xrightarrow{38} E$, $R \xrightarrow{43} A$, $S \xrightarrow{43} C$, $T \xrightarrow{37} D$; Total = 211.]

6. A company has four zones and four salesmen. The zones do not have the same sales potential. The annual estimated sales in these zones are

Zone	I	II	III	IV
Sales (Rs.)	60,000	50,000	40,000	30,000

The abilities of the salesmen are different. The yearly sales in any zone under the same circumstances by the four salesmen are in the following proportion.

Salesmen	P,	Q,	R,	S
Proportion	7,	5,	5,	4

Assign the zones to these salesmen so that the total sales are maximum.

(**Hint :** Since 7 + 5 + 5 + 4 = 21 the proportion of expected sales in zone I

Due to the salesmen $P = 60{,}000 \times \frac{7}{21} = 20{,}000$

Due to the salesman $Q = 50{,}000 \times \frac{7}{21} = 16666{\cdot}67$.

To avoid fractions we can straight way multiply these sales by 7, 5, 5, 4 and get the expected sales in (ten thousand rupees). The cost-matrix will be

Salesmen \ Zones	I	II	III	IV
P	42	35	28	21
Q	30	25	20	15
R	30	25	20	15
S	24	20	16	12

[**Ans. :** (i) $P \to$ I, $Q \to$ II, $R \to$ III, $D \to$ IV,
(ii) $P \to$ I, $Q \to$ III, $R \to$ II, $D \to$ IV.]

(iv) Assignment Problems With Restrictions

In certain cases we find that there are some blank cells in the cost matrix because of non-availability of information. In such cases we assign a very high cost (infinite cost) to such cells. This amounts to excluding such cells from further considerations.

Example 1 : In a workshop there are six jobs and six operators to do these jobs. The cost of production is given in the following table. Solve the assignment problem so that the total cost is minimum.

Operators \ Jobs	1	2	3	4	5	6
A	—	17	20	22	16	16
B	21	—	18	23	17	16
C	23	22	—	19	22	19
D	27	21	19	—	20	21
E	16	18	17	23	—	17
F	18	16	20	24	19	—

Sol. : Since the jobs 1, 2, 3, 4, 5, 6 cannot be assigned to the operators A, B, C, D, E, and F, we assign a very high cost (∞) in the cells c_{ij}. Thus, we get the following matrix.

Operators \ Jobs	1	2	3	4	5	6
A	∞	17	20	22	16	16
B	21	∞	18	23	17	16
C	23	22	∞	19	22	19
D	27	21	19	∞	20	21
E	16	18	17	23	∞	17
F	18	16	20	24	19	∞

Step 1 : To Reduce The Matrix

Subtract the smallest elements 16, 16, 19, 19, 16 and 16 from the elements of the first, second, third, fourth, fifth and sixth rows respectively. Thus, we get Table No. 1.

Table 1

	1	2	3	4	5	6
A	∞	1	4	6	0	~~0~~
B	5	∞	2	7	1	0
C	4	3	∞	0	3	~~0~~
D	8	2	0	∞	1	2
E	0	2	1	7	∞	1
F	2	0	4	8	3	∞

Step 2 : To Assign Zeros

The second row has a single zero. Encircle it and cross all other zeros in its column. The remaining rows have a single zero each. Encircle these zeros.

Step 3 : Optimality Check

Since each row and each column has only one single encircled zero, the solution is optimal.

The assignment is as follows -

$$A \xrightarrow{16} 5,\ B \xrightarrow{16} 6,\ C \xrightarrow{19} 4,\ D \xrightarrow{19} 3,\ E \xrightarrow{16} 1,\ F \xrightarrow{16} 2.$$

Total cost = ₹ 102.

Example 2 : Four new machines M_1, M_2, M_3 and M_4 are to be installed in a machine shop. There are five vacant places A, B, C, D and E available. Because of limited space, machine M_2 cannot be installed at C and M_3 cannot be installed at A. C_{ij} is the assignment cost of machine i to place j in rupees as shown below.

Machines \ Places	A	B	C	D	E
M_1	4	6	10	5	6
M_2	7	4	—	5	4
M_3	—	6	9	6	2
M_4	9	3	7	2	3

Find the optimum assignment schedule. **(S.U. 2004)**

Sol. : Since there are four machines and five places the cost-matrix is not balanced, so we add one dummy machine M_5 with zero cost in that row. Further as machine M_2 cannot be placed at C and machine M_3 cannot be placed at A, we assign a very high cost (∞) in the cells M_2C and M_3A. Thus, we get the following cost matrix.

Machines \ Places	A	B	C	D	E
M_1	4	6	10	5	6
M_2	7	4	∞	5	4
M_3	∞	6	9	6	2
M_4	9	3	7	2	3
M_5	0	0	0	0	0

Step 1 : To Reduce The Matrix

We subtract the smallest elements in each row from all other elements in the corresponding rows. The smallest elements are 4, 4, 2 and 2. We subtract them from all other elements of the corresponding rows. Thus, we get Table 1.

Table 1

	A	B	C	D	E
M_1	0	2	6	1	2
M_2	3	0	∞	1	~~0~~
M_3	∞	4	7	4	0
M_4	7	1	5	0	1
M_5	~~0~~	~~0~~	0	~~0~~	~~0~~

Step 2 : To Assign Zeros

The first row has only one zero. Encircle it and cross other zero in that column. Third row has a single zero. Encircle it and cross all other zeros in that column. Second row has a single zero. Encircle it and cross the other zero in

the column. Fourth row has a single zero. Encircle it and cross the other zero in that column. Fifth row has a single zero. Encircle it.

Step 3 : Optimality Check

Now every row and column has only one single encircled zero. Since there are five encircled zeros, equal to the order of the matrix, the present solution is optimal.

The optimum assignment is as follows.

$$M_1 \xrightarrow{4} A,\ M_2 \xrightarrow{4} B,\ M_3 \xrightarrow{2} E,\ M_4 \xrightarrow{2} D,\ M_5 \xrightarrow{0} C.$$

$\therefore$ The total assignment cost = ₹ 4 + 4 + 2 + 2 = ₹ 12

The place C will remain vacant.

Example 3 : There are four trucks at places A, B, C, D and they are to be dispatched to five places P, Q, R, S, T. Because of some problems the trucks B and C cannot be dispatched to the places R and P. The distances between the various places are given in the following table.

Find the assignment of trucks to different places so that the total distance travelled will be minimum.

Truck \ Places	P	Q	R	S	T
A	10	12	16	11	12
B	13	10	—	11	10
C	—	12	15	12	8
D	15	9	13	8	9

Sol. : Since the cost matrix is not balanced we introduce a dummy truck E with zero costs in that row. Further, since the trucks B and C cannot be dispatched to places R and P, we assign a very high cost (∞) in these vacant cells.

Thus, we get the following matrix.

Truck \ Places	P	Q	R	S	T
A	10	12	16	11	12
B	13	10	∞	11	10
C	∞	12	15	12	8
D	15	9	13	8	9
E	0	0	0	0	0

Step 1 : To Reduce The Matrix

We subtract the smallest element in each row from all other elements in that row. The smallest elements in the first, second, third and fourth rows are 10, 10, 8 and 8. We subtract them from all other elements of the corresponding rows. Thus, we get the Table No. 1.

Table 1

	P	Q	R	S	T
A	0	2	6	1	2
B	3	0	∞	1	~~0~~
C	∞	4	7	4	0
D	7	1	5	0	1
E	~~0~~	~~0~~	0	~~0~~	~~0~~

Step 2 : To Assign Zeros

The first row has only one zero. Encircle it and cross the other zero in its column. Encircle the zero in the third row

and cross other zeros in its column. Encircle the zero in the second row and cross the other zero in its column. There is only one zero in the fourth row. Encircle it and cross the other zero in its column. Now, there is only one uncrossed zero in the fifth row. Encircle it.

Step 3 : Optimality Check

Now, every row and column has only one single encircled zero. Since there are five encircled zeros, equal to the order of the matrix, the present solution is optimal.

The optimum assignment is as follows -

$$A \xrightarrow{10} P,\ B \xrightarrow{10} Q,\ C \xrightarrow{8} T,\ D \xrightarrow{8} S,\ E \xrightarrow{0} R.$$

The total minimum distance = 36 units.

Example 4 : The following table gives the costs of assigning four jobs to four persons P, Q, R, S. The first job, for some reason, cannot be assigned to person P. Find the optimal assignment.

Persons \ Jobs	I	II	III	IV
P	—	6	3	1
Q	5	8	6	7
R	6	9	5	4
S	4	7	7	3

Sol. : Since the job I cannot be assigned to P, we assign a very high cost (∞) in the first cell. Thus, we get the following matrix.

Persons \ Jobs	I	II	III	IV
P	∞	6	3	1
Q	5	8	6	7
R	6	9	5	4
S	4	7	7	3

Step 1 : To Reduce The Matrix

Subtract the smallest elements 1, 5, 4, 3 from the first, second, third and fourth rows. Thus, we get Table No. 1. Since all rows and columns do not

Table 1

	I	II	III	IV
P	∞	5	2	0
Q	0	3	1	2
R	2	5	1	0
S	1	4	4	0

Table 2

	I	II	III	IV
P	∞	2	1	0
Q	0	0	~~0~~	2
R	2	2	0	~~0~~
S	1	1	3	~~0~~

have zeros we subtract the smallest elements from the columns from other elements of those columns. We subtract 3 and 1 from second and third columns. Thus, we get Table No. 2.

Step 2 : To Assign Zeros

The first row has a single zero. Encircle it and cross all other zeros in its column. Now, third row has a single unmarked zero. Encircle it and cross the other zero in its column. Now, check columns. First and second column have unmarked zeros. Encircle them.

Step 3 : Optimality Check

Since all rows and columns do not have encircled zeros, the assignment is not optimum.

Step 4 : To Draw Minimum Number Of Lines

We can draw three lines to cover all zeros. Since, the order of the matrix is four, this is not an optimum solution.

Table 3

	I	II	III	IV
P	∞	1	~~0~~	0
Q	0	~~0~~	~~0~~	3
R	2	2	0	1
S	~~0~~	0	2	~~0~~

Step 4 : Iteration For The Optimal Solution

The smallest element among all uncovered elements is one. This is to be subtracted from all uncovered elements and is to be added to the elements which lie at the intersection of two lines. Thus, we get Table No. 3.

Step 5 : To Assign Zeros

There is only one zero in the third row. Encircle it and cross all other zeros in its column. Now, there is one zero in the first row. Encircle it and cross the zero in its column. Now, check columns. Both the first and second columns have two zeros each. Encircle the first zero in the first column and cross the other zero in its row. Since, the first column has an assigned zero, cross the other zero in that column. The fourth column has a single unmarked zero. Encircle it.

Step 6 : Optimality Check

Since, each row and column has an assigned zero, the solution is optimal. Thus, the assignment is optimal.

$$P \xrightarrow{1} \text{IV},\quad Q \xrightarrow{5} \text{I},\quad R \xrightarrow{5} \text{III},\quad P \xrightarrow{7} \text{II}$$

Hence, the total minimum cost = ₹ 18.

Remark ....

There is an alternative solution. In step 5, instead of encircling the first zero in the second row, we may choose to encircle the second zero and cross the zero in its column. Now, we can encircle first zero in the fourth row.

Thus, the alternative optimum assignment is

$$P \xrightarrow{1} \text{IV},\quad Q \xrightarrow{8} \text{II},\quad R \xrightarrow{5} \text{III},\quad S \xrightarrow{4} \text{I}$$

The total minimum cost = ₹ 18.

Example 5 : There are five machines and five operators in a factory. The production of these operators on these machines is given below. Assign the machines to the operators, so as to maximise the productivity.

	M_1	M_2	M_3	M_4	M_5
O_1	13	7	—	10	9
O_2	15	14	7	13	9
O_3	10	9	14	11	—
O_4	8	11	7	13	12
O_5	10	11	15	9	8

Sol. : Write the steps taken to obtain the following tables.

Table 1
(Minimisation) Cost Matrix

	M_1	M_2	M_3	M_4	M_5
O_1	2	8	∞	5	6
O_2	0	1	8	2	6
O_3	5	6	1	4	∞
O_4	7	4	8	2	3
O_5	5	4	0	6	7

Table 2

	M_1	M_2	M_3	M_4	M_5
O_1	0	6	∞	3	4
O_2	0	1	8	2	6
O_3	4	5	0	3	∞
O_4	5	2	6	0	1
O_5	5	4	0	6	7

Table 3

	M_1	M_2	M_3	M_4	M_5
O_1	0	5	∞	3	3
O_2	~~0~~	0	8	2	5
O_3	4	4	0	3	∞
O_4	5	1	6	0	~~0~~
O_5	5	3	~~0~~	6	6

Table 4

	M_1	M_2	M_3	M_4	M_5
O_1	0	5	∞	3	3
O_2	~~0~~	0	11	2	5
O_3	1	1	~~0~~	0	∞
O_4	5	1	9	~~0~~	0
O_5	2	~~0~~	0	3	3

$$O_1 \xrightarrow{13} M_1,\quad O_2 \xrightarrow{14} M_2,\quad O_3 \xrightarrow{11} M_4,$$
$$O_4 \xrightarrow{12} M_5,\quad O_5 \xrightarrow{15} M_3.$$

Maximum Total Production = 65.

Example 6 : There are five operators who can work on five machines. The cost matrix is given in the following table. A sixth machine is available to replace one of the existing machines and the associated cost is given in the same table. Find -

(i) whether the new machine should be used.

(ii) the optimal assignment and minimum total cost.

Machines

	M_1	M_2	M_3	M_4	M_5	M_6
O_1	20	16	—	17	14	23
O_2	14	—	16	—	22	15
O_3	16	18	20	21	13	19
O_4	21	23	17	19	18	—
O_5	—	17	15	20	19	16

Sol. : Write the steps taken to obtain the following tables.

We shall first find the total optimum cost with six machines.

Table 1

	M_1	M_2	M_3	M_4	M_5	M_6
O_1	20	16	∞	17	14	23
O_2	14	∞	16	∞	22	15
O_3	16	18	20	21	13	19
O_4	21	23	17	19	18	∞
O_5	∞	17	15	20	19	16
O_6	0	0	0	0	0	0

Table 2

	M_1	M_2	M_3	M_4	M_5	M_6
O_1	6	2	∞	3	0	7
O_2	0	∞	2	∞	6	1
O_3	3	5	7	8	~~0~~	6
O_4	4	6	0	2	1	∞
O_5	∞	2	~~0~~	5	4	1
O_6	~~0~~	0	~~0~~	0	~~0~~	0

Table 3

	M_1	M_2	M_3	M_4	M_5	M_6
O_1	6	1	∞	2	0	6
O_2	0	∞	2	∞	6	0
O_3	3	4	7	7	0	5
O_4	4	5	0	1	1	∞
O_5	∞	1	0	4	4	0
O_6	1	1	1	0	1	0

Table 4

	M_1	M_2	M_3	M_4	M_5	M_6
O_1	6	0	∞	2	~~0~~	6
O_2	0	∞	2	∞	6	~~0~~
O_3	3	3	7	7	0	5
O_4	4	4	0	1	1	∞
O_5	∞	~~0~~	~~0~~	4	4	0
O_6	1	~~0~~	1	0	~~0~~	~~0~~

$O_1 \xrightarrow{16} M_2$, $O_2 \xrightarrow{14} M_1$, $O_3 \xrightarrow{13} M_5$,

$O_4 \xrightarrow{17} M_3$, $O_5 \xrightarrow{16} M_6$, $O_6 \xrightarrow{0} M_4$.

Total = 76.

Now, consider the cost matrix deleting the sixth row and six column.

Table 5

	M_1	M_2	M_3	M_4	M_5
O_1	20	16	∞	17	14
O_2	14	∞	16	∞	22
O_3	16	18	20	21	13
O_4	21	23	17	19	18
O_5	∞	17	15	20	19

Table 6

	M_1	M_2	M_3	M_4	M_5
O_1	6	2	∞	3	0
O_2	0	∞	2	∞	8
O_3	3	5	7	8	~~0~~
O_4	4	6	0	2	1
O_5	∞	2	~~0~~	5	3

Table 7

	M_1	M_2	M_3	M_4	M_5
O_1	4	0	∞	1	~~0~~
O_2	0	∞	2	∞	10
O_3	1	3	5	6	0
O_4	4	6	0	2	3
O_5	∞	2	~~0~~	5	5

Table 8

	M_1	M_2	M_3	M_4	M_5
O_1	4	0	∞	1	1
O_2	0	∞	3	∞	11
O_3	~~0~~	2	5	5	0
O_4	3	5	0	1	3
O_5	∞		~~0~~	4	5

Table 9

	M_1	M_2	M_3	M_4	M_5
O_1	4	0	∞	1	1
O_2	0	∞	4	∞	11
O_3	~~0~~	2	6	5	0
O_4	2	4	~~0~~	0	2
O_5	∞	~~0~~	0	3	4

$O_1 \xrightarrow{16} M_2$, $O_2 \xrightarrow{14} M_1$,

$O_3 \xrightarrow{13} M_5$, $O_4 \xrightarrow{19} M_4$,

$O_5 \xrightarrow{15} M_3$.

Total = 77.

With six machines the total cost is 76 and with five machines the total cost is 77.

Since, the cost is reduced by sixth machine, the sixth machine should be used.

7. Travelling Salesman's Problem

A travelling salesman starts from his home city, visits the given cities once and returns to his home city. He has to choose the cities in such a way that the total distance travelled (*i.e.*, total cost or total time) is minimum. The salesman cannot visit a city twice and cannot return to home city before visiting all assigned cities.

This is an assignment problem and can be solved by the usual method of solving an assignment problem.

Example : A travelling salesman has to visit six cities *A*, *B*, *C*, *D*, *E* and *F*. He can start from any city, visit the remaining cities once and return to the starting city. The travelling costs in hundred rupees from a city to another city are given below.

		To City					
		A	*B*	*C*	*D*	*E*	*F*
From City	*A*	∞	22	26	27	21	21
	B	26	∞	23	31	23	21
	C	30	27	∞	24	27	24
	D	24	36	26	∞	25	26
	E	24	21	24	28	∞	22
	F	25	25	21	29	24	∞

Find the route with minimum cost.

Sol. : Since travelling from a city to the same city is not to be taken into consideration the costs in these cells are shown by ∞.

Step 1 : To Reduce The Matrix

Subtract the smallest elements 21, 21, 24, 24, 21 and 21 from the elements of the first, second, third, fourth, fifth and sixth rows respectively. Thus, we get the Table 1.

Table 1

From City \ To City	A	B	C	D	E	F
A	∞	1	5	6	0	0
B	5	∞	2	10	2	0
C	6	3	∞	0	3	0
D	0	12	2	∞	1	2
E	3	0	3	7	∞	1
F	4	4	0	8	3	∞

Table 2

From City \ To City	A	B	C	D	E	F
A	∞	1	5	6	(0)	~~0~~
B	5	∞	2	10	2	(0)
C	6	3	∞	(0)	3	~~0~~
D	(0)	12	2	∞	1	2
E	3	(0)	3	7	∞	1
F	4	4	(0)	8	3	∞

Step 2 : To Assign Zeros

Each row and each column has at least one zero. Now, the first row has two zeros. We therefore go to the next row. In second row there is only one zero. We encircle it and cross the other zeros in its column. Now we see that each of the remaining rows has a single uncrossed zero. We encircle these zeros.

Step 3 : Optimality Check

Since each row and each column has only one encircled zero the solution is optimal.

The required assignment is

$$A \xrightarrow{21} E,\ E \xrightarrow{21} B,\ B \xrightarrow{21} F,\ F \xrightarrow{21} C,\ C \xrightarrow{24} D,\ D \xrightarrow{24} A.$$

The total cost is ₹ 21 + 21 + 21+ 21 + 24 + 24 = ₹ 132.

EXERCISE - VI

1. The production rate of four workers is shown in the following matrix. Assign the jobs to the workers, so that the production is maximum.

Workers \ Jobs	A	B	C	D	E
I	17	—	32	24	30
II	11	10	25	20	10
III	12	31	24	36	28
IV	18	11	23	30	—

[**Ans. :** I → *E*, II → *C*, III → *B*, IV → *D*.
Maximum production = 116.]

2. The following table gives the cost of assigning a machine to an operator. Operator *X* cannot be assigned to machine *C* and operator *Z* cannot be assigned to machine *D*.

Find the assignment with minimum cost.

Operator \ Machines	A	B	C	D	E
Z	7	7	—	4	8
Y	9	6	4	5	6
Z	11	5	7	—	5
U	9	4	8	9	4
V	8	7	9	11	3

[**Ans. :** $X \xrightarrow{4} D,\ Y \xrightarrow{4} C,\ Z \xrightarrow{5} B,\ U \xrightarrow{9} A,\ V \xrightarrow{3} E.$
Total = 25.]

3. The profits earned by completing different jobs on differents machines are given in the following table. Assign the jobs to the machines so that the total profit earned is maximum.

Machines \ Jobs	1	2	3	4	5
A	10	4	—	7	6
B	12	11	4	10	6
C	7	6	11	8	—
D	5	8	4	10	9
E	7	8	12	6	5

[**Ans. :** $A \xrightarrow{10} 1,\ B \xrightarrow{11} 2,\ C \xrightarrow{8} 4,\ D \xrightarrow{9} 5,\ E \xrightarrow{12} 3.$
Total = 50.]

4. The productivity of the operators *A*, *B*, *C*, *D* and *E* on different machines I, II, III, IV and V is shown in the following matrix. Assign machines to the operators so that the total profit is maximum.

Operators \ Machines	I	II	III	IV	V
A	9	14	10	7	12
B	8	11	12	—	13
C	10	10	8	11	—
D	12	14	11	10	7
E	13	10	12	13	10

[**Ans. :** *A* → II, *B* → V, *C* → IV, *D* → III, *E* → I ; Total = 62.]

EXERCISE - VII

Theory

1. What is an assignment problem ? Name the various methods used to solve an assignment problem.

2. Explain in details the Hungarian method of solving an assignment problem.

3. What is an unbalanced assignment problem ? How is it solved ?

4. What is meant by an assignment problem with restriction ? How does it arise ? How is it solved ?

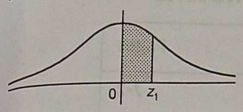
5. Explain the procedure of solving an assignment problem. (S.U. 2003)

6. Explain how an assignment problem is a particular case of transportation problem ? (S.U. 2004)

7. Give mathematical statement of assignment problem. (S.U. 2005)

☆ ☆ ☆

Area Under Standard Normal Curve



The table gives the area under the standard normal curve from $z = 0$ to $z = z_1$ which is the probability that z will lie between $z = 0$ and $z = z_1$.

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2517	.2549
0.7	.2580	.2611	.2642	.2673	.2703	.2734	.2764	.2794	.2823	.2852
0.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
0.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4415	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4864	.4868	.4871	.4875	.4878	.4841	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4560	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4979	.4980	.4981
2.9	.4981	.4982	.4982	.4983	.4984	.4984	.4985	.4985	.4986	.4986
3.0	.4987	.4987	.4987	.4988	.4988	.4989	.4989	.4989	.4990	.4990