Statistics

Statistics is the study of collection, organisation, analysis, interpretation and presentation of data. It deals with all aspects of data, including the planning of data collection in terms of the design of survey and experiments.

24.1 Measures of Central Tendency

It is a central value or a typical value for a probable distribution. It is also called an average or just the centre of the distribution.

The most commonly used measures of central tendency are

Arithmetic Mean (AM)

Arithmetic mean is obtain by dividing the sum of all the observations by the number of observations and it is denoted by $\bar{x}$.

1. **Mean for Ungrouped Data**

   (i) **Direct method** If $x_1, x_2, x_3, \ldots, x_n$ are n observations, then mean by

   $$\bar{x} = \frac{x_1 + x_2 + \ldots + x_n}{n} \quad \text{or} \quad \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

   (ii) **Shortcut method** $\bar{x} = A + \frac{1}{n}\sum_{i=1}^{n} d_i$

   where, A = Assumed mean and deviation $d_i = x_i - A$

Example 1. The mean of n items is $\bar{x}$. If each item is successively increased by $3, 3^2, 3^3, \ldots, 3^n$, then mean will be equal to

(1) $\bar{x} + \frac{3^{n+1}}{2n}$ (2) $\bar{x} + \frac{3(3^n - 1)}{3n}$

(3) $\bar{x} + \frac{3^n}{3n}$ (4) $\bar{x} + \frac{3(3^n - 1)}{2n}$

Sol. (4) Let $x_1, x_2, x_3, \ldots, x_n$ be n items. Then, new items are $x_1 + 3, x_2 + 3^2, \ldots, x_n + 3^n$.

$$\therefore \text{ New mean} = \frac{(x_1 + 3) + (x_2 + 3^2) + \ldots + (x_n + 3^n)}{n}$$

$$= \frac{(x_1 + x_2 + \ldots + x_n)}{n} + \frac{3^1 + 3^2 + \ldots + 3^n}{n}$$

$$= \bar{x} + \frac{3\,(3^n - 1)}{2\,n}$$

2. **Mean for Grouped Data** Let x_i be the class marks of each intervals and f_i be the corresponding class frequencies.

   (i) **Direct method**

   $$\bar{x} = \frac{f_1x_1 + f_2x_2 + \ldots + f_nx_n}{f_1 + f_2 + \ldots + f_n} = \frac{\Sigma f_i x_i}{\Sigma f_i} = \frac{1}{N}\Sigma f_i x_i$$

   where $N = \Sigma f_i$

   (ii) **Shortcut method** If $d_i = x_i - A$, where A is assumed mean, then

   $$\text{Mean}\,(\bar{x}) = A + \frac{1}{N}\sum_{i=1}^{n} f_i\, d_i$$

   where, $\sum f_i = N$

   (iii) **Step deviation method** If $u_i = \frac{x_i - A}{h}$, where A is assumed mean and h = width of internal, then

   $$\text{Mean}\,(\bar{x}) = A + h\frac{1}{N}\sum_{i=1}^{n} f_i\, u_i$$

$\bar{x}_1 = 5200$
$\bar{x}_2 = 4200$
$x = 5000 = \frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1+n_2}$
$50n_1 + 50n_2 = 52n_1 + 42n_2$

Example 2. The arithmetic mean of the marks from the following table is

Marks	0-10	10-20	20-30	30-40	40-50	50-60
Number of students	12	18	27	20	17	6

(1) 20 (2) 28
(3) 2800 (4) 100

Sol. (2)

Marks	Class marks (x)	f	fx
0-10	5	12	60
10-20	15	18	270
20-30	25	27	675
30-40	35	20	700
40-50	45	17	765
50-60	55	6	330
		$\Sigma f = 100$	$\Sigma fx = 2800$

$$\therefore \quad \bar{x} = \frac{\Sigma fx}{\Sigma f} = \frac{2800}{100} = 28$$

Weighted Arithmetic Mean

Let $w_1, w_2, w_3, \ldots, w_n$ be the weights assigned to the values $x_1, x_2, x_3, \ldots, x_n$ respectively, then the weighted arithmetic mean is defined as

$$\text{Weighted arithmetic mean} = \frac{w_1x_1 + w_2x_2 + \ldots + w_nx_n}{w_1 + w_2 + \ldots + w_n}$$

Example 3. The weighted mean of the first n natural numbers, the weights being the corresponding numbers, is

(1) $\frac{n+1}{2}$ (2) $\frac{n+2}{2}$ $\frac{\Sigma w_i x_i}{\Sigma w_i}$

(3) $\frac{2n+1}{3}$ (4) None of these

Sol. (3) First n natural numbers are $1, 2, 3, \ldots, n$; whose corresponding weights are $1, 2, 3, \ldots, n$ respectively.

$\therefore$ Weight mean

$$= \frac{1 \times 1 + 2 \times 2 + \ldots\ldots + n \times n}{1 + 2 + \ldots\ldots + n} = \frac{\frac{n(n+1)(2n+1)}{6}}{\frac{n(n+1)}{2}}$$

$$= \frac{1^2 + 2^2 + \ldots + n^2}{1 + 2 + \ldots + n}$$

$$= \frac{\frac{n(n+1)(2n+1)}{6}}{\frac{n(n+1)}{2}} = \frac{2n+1}{3}$$

Combined Mean

If two sets of observations are given, then the combined mean of both the sets can be calculated by the following formula

$$\boxed{\bar{x}_{12} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1 + n_2}}$$

where, $\bar{x}_1$ = Mean of first set of observations
n_1 = Number of observations in first set
$\bar{x}_2$ = Mean of second set of observations
n_2 = Number of observations in second set

Properties of Mean

(i) Algebraic sum of the deviations of a set of values from their mean is zero.

(ii) The sum of the squares of the deviations of a set of values is minimum when taken about mean.

(iii) Mean is affected by the change or shifting of origin and scale.

Example 4. The average salary of male employees in a firm was ₹ 5200 and that of females was ₹ 4200. The mean salary of all the employees was ₹ 5000. The percentage of male and female employees are respectively

$5000 = \frac{5200 n_1 + 4200 n_2}{n_1 + n_2}$

(1) 80, 20 (2) 20, 80
(3) 60, 40 (4) 52, 48

Sol. (1) Let $x_1 =$ ₹5200, $x_2 =$ ₹4200, $\bar{x} =$ ₹5000.

Also, we know that $\bar{x} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1 + n_2}$

$5000n_1 + 5000n_2 = 5200n_1 + 4200n_2$

$\Rightarrow \quad 5000(n_1 + n_2) = 5200n_1 + 4200n_2$

$800n_2 = 200n_1$

$\Rightarrow \quad \frac{n_1}{n_2} = \frac{4}{1}$

$\frac{n_1}{n_2} = 4$

$\therefore$ The percentage of male employees in the firm $\boxed{n_1 = 4n_2}$

$$= \frac{4}{4+1} \times 100 = 80\%$$

and the percentage of female employees in the firm

$$= \frac{1}{4+1} \times 100 = 20\%$$

Geometric Mean (GM)

The nth root of the product of the values, is called geometric mean.

1. **For Ungrouped Data** The geometric mean G of n observations $x_1, x_2, \ldots, x_n$, is defined as (when none of them being non-negative)

$$G = (x_1 \cdot x_2 \cdot x_3 \ldots x_n)^{1/n}$$

$$\Rightarrow \quad \log G = \frac{1}{n}\sum_{i=1}^{n} \log x_i$$

$$G = \text{antilog}\left(\frac{\log x_1 + \log x_2 + \ldots + \log x_n}{n}\right)$$

Let $x_1, x_2, \ldots, x_n$ be n observations and whose corresponding frequencies are $f_1, f_2, \ldots, f_n$, then geometric mean is given by

$$GM = (x_1^{f_1} \cdot x_2^{f_2} \ldots x_n^{f_n})^{\frac{1}{N}}$$

$$\frac{1 + 4 + 3 \cdot 3 + 4 \cdot 4 + \ldots\ldots n \cdot n}{1 + 2 + 3 + \ldots} = \frac{\frac{n(n+1)(2n+1)}{6}}{\frac{n(n+1)}{2}}$$

where, $N = \sum_{i=1}^{n} f_i$

We can also write, $\text{GM} = \text{antilog}\left(\dfrac{\sum_{i=1}^{n} f_i \log x_i}{N}\right)$

2. **For Grouped Data** Let x_i's be the class marks of each class interval and f_i's be the corresponding frequencies.
Then,

$$\text{GM} = (x_1^{f_1} \cdot x_2^{f_2} \dots x_n^{f_n})^{\frac{1}{N}}$$

or $$\text{GM} = \text{antilog}\left(\dfrac{\sum_{i=1}^{n} f_i \log x_i}{N}\right)$$

where, N = Sum of frequencies.

Example 5. The geometric mean of the numbers $3, 3^2, 3^3, \dots, 3^n$ is

(1) $3^{2/n}$
(2) $3^{(n-1)/2}$
(3) $3^{n/2}$
(4) $3^{(n+1)/2}$

$[3 \cdot 3^2 \cdot 3^3 \dots 3^n]^{1/n}$

$[3 \cdot 3^2]^{1/2}$

$3\sqrt{3}$ $3^{1+1/2}$

$3^{3/2}$

OR

$3^{\frac{n(n+1)}{2n}}$

Sol. (4) $\therefore \text{GM} = (3 \cdot 3^2 \dots 3^n)^{1/n}$

$$= 3^{\frac{1+2+\dots+n}{n}}$$

$$= 3^{\frac{n(n+1)}{2n}} = 3^{\frac{n+1}{2}}$$

Median

The median is the middle most value of the set of observations, provided all the observations are arranged in the ascending or descending order.

1. **Median for Ungrouped Data** Suppose n observations are arranged in ascending or descending order.
   (i) If n is an odd number, then

$$\text{Median} = \text{Value of} \left(\frac{n+1}{2}\right) \text{th observation}$$

   (ii) If n is an even number, then

$$\text{Median} = \frac{\text{Value of} \left(\dfrac{n}{2}\right) \text{th observation} + \text{Value of} \left(\dfrac{n}{2}+1\right) \text{th observation}}{2}$$

2. Let x_i be ascending or descending observations, f_i be the corresponding frequencies and cf be the corresponding cumulative frequency.
   Let N be the total sum of frequencies
   (i) If N is odd, then

$$\text{Median} = \text{Value of } \frac{(N+1)}{2} \text{th observation.}$$

   (ii) If N is even, then

$$\text{Median} = \frac{1}{2}\left[\text{Value of } \left(\frac{N}{2}\right)\text{th} + \left(\frac{N}{2}+1\right)\text{th}\right]$$

   observation

3. **Median for Grouped Data** In this case, the class corresponding to the cumulative frequency just greater than or equal to $\frac{N}{2}$, is called the median class and the value of median is obtained by the following formula

$$\text{Median} = l + \left(\frac{\frac{N}{2} - c}{f}\right) \times h$$

where,
l = Lower limit of median class
f = Frequency of median class
h = Size of median class
c = Cumulative frequency of class before median class
Cumulative frequency is defined as a consecutive sum of frequencies.

Example 6. If a variable takes the discrete values $\alpha + 4, \alpha - \frac{7}{2}, \alpha - \frac{5}{2}, \alpha - 3, \alpha - 2, \alpha + \frac{1}{2}, \alpha - \frac{1}{2}, \alpha + 5\ (\alpha > 0)$, then the median is

(1) $\alpha - \frac{5}{4}$ (2) $\alpha - \frac{1}{2}$
(3) $\alpha - 2$ (4) $\alpha + \frac{5}{4}$

Sol. (1) Since, $\alpha > 0$ (given), arrange the data in ascending order.

$$\alpha - \frac{7}{2}, \alpha - 3, \alpha - \frac{5}{2}, \alpha - 2, \alpha - \frac{1}{2}, \alpha + \frac{1}{2}, \alpha + 4, \alpha + 5$$

$$\text{Median} = \frac{\alpha - 2 + \alpha - \frac{1}{2}}{2} \quad [\because \text{ here, } n = 8, \text{ which is even}]$$

$$= \frac{2\alpha - \frac{5}{2}}{2} = \alpha - \frac{5}{4}$$

Example 7. From the data given, the median of the average deposit balance of saving for the branch during March 1982 is

Average deposit balance (₹)	Number of deposit
Less than 100	26
100-200	68
200-300	145
300-400	242
400-500	188
500-600	65
600-700	16

(1) 356 (2) 300
(3) 56.2 (4) 356.2

Sol. (4)

Average deposit balance (₹)	f	Cumulative frequency cf
Less than 100	26	26
100- 200	68	94
200-300	145	239
300-400	242	481
400-500	188	669
500-600	65	734
600-700	16	750

$$\frac{N}{2} = \frac{750}{2} = 375$$

The frequency just greater than 375 is 481.

∴ Median class is 300-400.

$$\text{Median} = l + \frac{\left(\frac{N}{2} - c\right)}{f} \times h$$

$$= 300 + \frac{375 - 239}{242} \times 100$$

$$= 300 + 56.2 = 356.2$$

Mode

Mode is the value of the variate which occurs most frequently.

1. Mode for Ungrouped Data

(i) The value which is repeated maximum number of times is the mode of the given data.

(ii) The mode is the value of the variate corresponding to the maximum frequency.

56.

$40 + \frac{50-50}{}$

2. Mode for Grouped Data

The class which has maximum frequency, is called modal class or group. The mode is given by the formula

$$\text{Mode} = l + \frac{f_1 - f_0}{2f_1 - f_2 - f_0} \times h$$

where, l = Lower limit of the modal class (the class having maximum frequency)

f_1 = Frequency of the modal class

f_0 = Frequency of the class preceding the modal class

f_2 = Frequency of the class succeeding the modal class

h = Width of the modal class

Relation between Mean, Mode and Median

In a moderately skewed distribution, there is a relation among mean, median and mode which is as follows

$$\text{Mean} - \text{Mode} = 3\,(\text{Mean} - \text{Median})$$

or

$$\text{Mode} = 3\,\text{Median} - 2\,\text{Mean}$$

Note In the symmetrical distribution, the values of mean, median and mode coincide.

Example 8. If the mode of the data is 18 and the mean is 24, then median is

(1) 18 (2) 24 (3) 22 (4) 21

Sol. (3) ∵ Mode = 3 (Median) −2 (Mean)

∴ 18 = 3 (Median) −2 (24)

⇒ Median = 22

Example 9. The mode of the following distribution is

(1) 46 (2) 6.66
(3) 46.67 (4) None of these

Class interval	Frequency
0-10	5
10-20	8
20-30	7
30-40	12
40-50	28
50-60	20
60-70	10
70-80	10

$40 + \left[\frac{16}{16+8}\right]10$

$40 + \frac{16}{24} \times 10$

$40 + \frac{20}{3}$

$\frac{140}{3}$

46.

Sol. (3) Here, maximum frequency is 28. Thus, the class 40-50 is the modal class.

∴ $l = 40$, $f_1 = 28$, $f_0 = 12$, $f_2 = 20$ and $h = 50 - 40 = 10$

$$\text{Now, Mode} = 40 + \frac{10\,(28 - 12)}{(2 \times 28 - 12 - 20)}$$

$$= 40 + 6.667 = 46.67 \text{ (approx.)}$$

Example 6. Probability of solving specific problem independently by A and B are $\frac{1}{2}$ and $\frac{1}{3}$, respectively. If both try to solve the problem independently, the probability that

(i) the problem is solved

(ii) exactly one of them solves the problem are respectively

, are

(1) $\frac{1}{2}$ and $\frac{2}{3}$ (2) $\frac{2}{3}$ and $\frac{1}{2}$

(3) $\frac{1}{4}$ and $\frac{3}{4}$ (4) None of these

Sol. (2) Probability of solving the problem by A, $P(A) = \frac{1}{2}$

Probability of solving the problem by B, $P(B) = \frac{1}{3}$

Probability of not solving the problem by

$$A = P(A') = 1 - P(A) = 1 - \frac{1}{2} = \frac{1}{2}$$

and probability of not solving the problem by

$$B = P(B') = 1 - P(B) = 1 - \frac{1}{3} = \frac{2}{3}$$

(i) P (the problem is solved) $= 1 - P$ (none of them solve the problem)

$$= 1 - P(A' \cap B') = 1 - P(A')P(B')$$

[$\because$ A and B are independent $\Rightarrow$ A' and B' are independent]

$$= 1 - \left(\frac{1}{2} \times \frac{2}{3}\right) = 1 - \frac{1}{3} = \frac{2}{3}$$

(ii) P (exactly one of them solve the problem)

$$= P(A)P(B') + P(A')P(B)$$

$$= \frac{1}{2} \times \frac{2}{3} + \frac{1}{2} \times \frac{1}{3} = \frac{1}{3} + \frac{1}{6} = \frac{2+1}{6} = \frac{3}{6} = \frac{1}{2}$$

Scanned by CamScanner

25.4 Conditional Probability

The probability of occurrence of an event E_2, when it is known that some event E_1 has already occurred is called the conditional probability and is denoted by $P\left(\frac{E_2}{E_1}\right)$. The symbol $P\left(\frac{E_2}{E_1}\right)$ is usually read as probability of E_2, given E_1. Consider two events E_1 and E_2. When it is known that event E_1 has occurred, it means that sample space would reduce to that sample points representing event E_1. Now, for $P\left(\frac{E_2}{E_1}\right)$ we must look for the sample points representing the simultaneous occurrence of E_1 and E_2, i.e. sample points in $E_1 \cap E_2$.

$$\Rightarrow \quad P\left(\frac{E_2}{E_1}\right) = \frac{n(E_1 \cap E_2)}{n(E_1)} = \frac{\frac{n(E_1 \cap E_2)}{n(S)}}{\frac{n(E_1)}{n(S)}} = \frac{P(E_1 \cap E_2)}{P(E_1)}$$

Thus, $$P\left(\frac{E_2}{E_1}\right) = \frac{P(E_1 \cap E_2)}{P(E_1)},$$

where, $0 < P(E_1) \le 1.$

Similarly, $$P\left(\frac{E_1}{E_2}\right) = \frac{P(E_1 \cap E_2)}{P(E_2)}, 0 < P(E_2) \le 1.$$

Hence, $$P(E_1 \cap E_2) = \begin{cases} P(E_1) \cdot P(E_2/E_1), & P(E_1) > 0 \\ P(E_2) \cdot P(E_1/E_2), & P(E_2) > 0 \end{cases}$$

Important Results Related to Conditional Probability

(i) If E_1 and E_2 are independent events, then

$$P\left(\frac{E_2}{E_1}\right) = P(E_2)$$

(ii) If $E_1, E_2, E_3, \ldots, E_n$ are independent events, then $P(E_1 \cup E_2 \cup \ldots \cup E_n) = 1 - P(E_1')P(E_2') \ldots P(E_n')$

(iii) If E_1 and E_2 are events such that $E_2 \ne 0$, then

$$P\left(\frac{E_1}{E_2}\right) + P\left(\frac{E_1'}{E_2}\right) = 1$$

(iv) If E_1 and E_2 are events such that $E_1 \ne \phi$, then

$$P(E_2) = P(E_1) \cdot P\left(\frac{E_2}{E_1}\right) + P(E_1') \cdot P\left(\frac{E_2}{E_1'}\right)$$

(v) **Multiplication theorem on probability** If E_1, E_2 and E_3 are three events such that $E_1 \ne \phi, E_1E_2 \ne \phi$, then

$$P(E_1 \cap E_2 \cap E_3) = P(E_1) \cdot P\left(\frac{E_2}{E_1}\right) \cdot P\left(\frac{E_3}{E_1E_2}\right)$$

In general, if $E_1, E_2, \ldots, E_n$ are n events such that $E_1 \ne \phi, E_1E_2 \ne \phi, E_1E_2E_3 \ne \phi, \ldots, E_1E_2 \ldots E_{n-1} \ne \phi$, then

$$P(E_1 \cap E_2 \cap \ldots \cap E_n) = P(E_1)P\left(\frac{E_2}{E_1}\right)P\left(\frac{E_3}{E_1E_2}\right) \ldots P\left(\frac{E_n}{E_1E_2 \ldots E_{n-1}}\right)$$

(vi) If E_1 and E_2 are independent events, then

(a) $P(E_1 \cap E_2) = P(E_1) \cdot P\left(\frac{E_2}{E_1}\right)$, if $P(E_1) \ne 0$.

(b) $P(E_1 \cap E_2) = P(E_2) \cdot P\left(\frac{E_1}{E_2}\right)$, if $P(E_2) \ne 0$.

(vii) If E_1 and E_2 are independent events, then

$$P(E_1 \cap E_2) = P(E_1) \cdot P(E_2)$$

(viii) If $E_1, E_2, \ldots, E_n$ are independent events, then

$$P(E_1 \cap E_2 \cap E_3 \cap E_4 \cap \ldots \cap E_n) = P(E_1)P(E_2) \ldots P(E_n)$$

(ix) **Probability of atleast one of the n independent events** If $p_1, p_2, p_3 \ldots, p_n$ are the probabilities of happening of n independent events. $E_1, E_2, E_3, \ldots, E_n$ respectively, then

(a) probability of happening none of them

$$= P(\bar{E}_1 \cap \bar{E}_2 \cap \bar{E}_3 \cap \ldots \cap \bar{E}_n)$$
$$= P(\bar{E}_1) \cdot P(\bar{E}_2) \cdot P(\bar{E}_3) \ldots P(\bar{E}_n)$$
$$= (1 - p_1) \cdot (1 - p_2) \cdot (1 - p_3) \ldots (1 - p_n)$$

(b) probability of happening atleast one of them

$$= P(E_1 \cup E_2 \cup E_3 \cup \ldots \cup E_n)$$
$$= 1 - P(\bar{E}_1) \cdot P(\bar{E}_2) \cdot P(\bar{E}_3) \ldots P(\bar{E}_n)$$
$$= 1 - (1 - p_1) \cdot (1 - p_2) \cdot (1 - p_3) \ldots (1 - p_n)$$

(c) probability of happening of first event and not happening of the remaining

$$= P(\bar{E}_1 \cap \bar{E}_2 \cap \bar{E}_3 \ldots \cap \bar{E}_n)$$
$$= P(\bar{E}_1) \cdot P(\bar{E}_2) \cdot P(\bar{E}_3) \ldots P(\bar{E}_n)$$
$$= p_1(1 - p_2) \cdot (1 - p_2) \ldots (1 - p_n)$$

Example 8. $P_1, P_2, \ldots, P_8$ an eight players participating in a tournament. If $i < j$, then P_i will win the match against P_j. Players are paired up randomly for first round and winners of this round again paired up for the second round and so on. The probability that P_4 reaches in the final, is

(1) $\frac{2}{7}$ (2) $\frac{4}{9}$

(3) $\frac{4}{35}$ (4) None of these

Sol. (3) Let A_1 be the event that in the first round the four winners are P_1, P_4, P_i, P_j, where $i \in \{2, 3\}, j \in \{5, 6, 7\}$ and let

A_2 be the event that out of the four winners in the first round, P_1 and P_4 reaches in the final.

The event A_1 will occur, if P_4 plays with any of P_5, P_6, P_7 or P_8 (say with P_6) and P_1, P_2 and P_3 are not paired with P_5, P_7 and P_8. Further A_2 will occur if P_1 plays with P_i.

The required probability $= P(A_1 \cap A_2) = P(A_1) \cdot P\left(\frac{A_2}{A_1}\right)$

$$= \frac{4 \times \left(\frac{6!}{(2!)^3\, 3!} - 3!\right)}{\frac{8!}{(2!)^4\, 4!}} \times \frac{1}{\frac{4!}{(2!)^2\, 2!}}$$

$$= \frac{4}{35}.$$

(here, we have used the concept of division into groups).

25.5 Law of Total Probability

Let in a random experiment S is a sample space and $E_1, E_2, \ldots, E_n$ are mutually exclusive and exhaustive events. If A is any event which occur with E_1 or E_2 or E_3 or ... or E_n, then

$$P(A) = P(E_1) \cdot P\left(\frac{A}{E_1}\right) + P(E_2) P\left(\frac{A}{E_2}\right) + \ldots + P(E_n) P\left(\frac{A}{E_n}\right).$$

$$= \sum_{r=1}^{n} P(E_r) P\left(\frac{A}{E_r}\right)$$

Example 9. Two sets of candidates are competing for the positions on the board of directors of a company. The probabilities that the first and second sets will win are 0.6 and 0.4, respectively. If the first set wins, the probability of introducing a new product is 0.8 and the corresponding probability, if the second set wins is 0.3. What is the probability that the new product will be introduced?

(1) 0.1 (2) 0.3

(3) 0.5 (4) 0.6

Sol. (4) Let A_1 (A_2) denotes the event that first (second) set wins and let B be the event that a new product is introduced.

$\therefore$ $P(A_1) = 0.6$, $P(A_2) = 0.4$

$P\left(\frac{B}{A_1}\right) = 0.8$, $P\left(\frac{B}{A_2}\right) = 0.3$

$$P(B) = P(B \cap A_1) + P(B \cap A_2)$$

$$= P(A_1) \cdot P\left(\frac{B}{A_1}\right) + P(A_2) P\left(\frac{B}{A_2}\right)$$

$$= 0.6 \times 0.8 + 0.4 \times 0.3 = 0.6$$

P(A) = 0.6
P(B) = 0.4

25.6 Baye's Theorem

If $E_1, E_2, \ldots, E_n$ are mutually exclusive and exhaustive events with $P(E_i) \neq 0$, $(i = 1, 2, \ldots, n)$, then for any event E which is a subset of $\bigcup_{i=1}^{n} E_i$ such that $P(E) > 0$, then

$$P\left(\frac{E_i}{E}\right) = \frac{P(E_i) \cdot P\left(\frac{E}{E_i}\right)}{\sum_{i=1}^{n} P(E_i) \cdot P\left(\frac{E}{E_i}\right)}$$

Important Points

(i) The probabilities $P(E_1), P(E_2), \ldots, P(E_i)$ are called 'prior probabilities' because they exist before, we gain any information from the experiment itself.

$P\left(\frac{E}{E_1'}\right), i = 1, 2, \ldots, n$ are called the likelyhood probabilities.

(ii) The probability $P\left(\frac{E_i}{E}\right), i = 1, 2, \ldots, n$ are called 'posterior probabilities' because they are determined after the results of the experiment are known.

Example 10. Box I contains 2 white and 3 red balls and box II contains 4 white and 5 red balls. One ball is drawn at random from one of the boxes and is found to be red. Then, the probability that it was from box II, is

(1) $\frac{25}{32}$ (2) $\frac{21}{32}$ (3) $\frac{7}{32}$ (4) $\frac{15}{32}$

Sol. (1) Let A denote the event that the drawn ball is red,

$A_1 \equiv$ the event that box I is selected and

$A_2 \equiv$ the event that box II is selected

$$\therefore \quad P\left(\frac{A_2}{A}\right) = \frac{P(A_2) \cdot P\left(\frac{A}{A_2}\right)}{P(A_1) P\left(\frac{A}{A_1}\right) + P(A_2) \cdot P\left(\frac{A}{A_2}\right)}$$

$$= \frac{\frac{1}{2} \cdot \frac{5}{9}}{\frac{1}{2} \cdot \frac{5}{9} + \frac{1}{2} \cdot \frac{3}{5}} = \frac{25}{32}$$

Example 11. The chances of defective screws in three boxes A, B and C are $\frac{1}{5}, \frac{1}{6}$ and $\frac{1}{7}$, respectively. A box is selected at random and a screw drawn from it at random is found to be defective. Then, the probability that it came from box A is

(1) $\frac{42}{107}$ (2) $\frac{4}{107}$ (3) $\frac{2}{107}$ (4) $\frac{1}{7}$

Sol. (1) Let E_1, E_2 and E_3 denote the events of selecting box A, B and C respectively and A be the event that a screw selected at random is defective. Then,

$$P(E_1) = P(E_2) = P(E_3) = \frac{1}{3}$$

$$P\left(\frac{A}{E_1}\right) = \frac{1}{5}, \quad P\left(\frac{A}{E_2}\right) = \frac{1}{6},$$

$$P\left(\frac{A}{E_3}\right) = \frac{1}{7}$$

By Baye's theorem

Required probability, $P\left(\frac{E_1}{A}\right)$

$$= \frac{P(E_1)\,P\left(\frac{A}{E_1}\right)}{P(E_1)\,P\left(\frac{A}{E_1}\right) + P(E_2)\,P\left(\frac{A}{E_2}\right) + P(E_3)\,P\left(\frac{A}{E_3}\right)}$$

$$= \frac{\frac{1}{3} \times \frac{1}{5}}{\frac{1}{3} \times \frac{1}{5} + \frac{1}{3} \times \frac{1}{6} + \frac{1}{3} \times \frac{1}{7}}$$

$$= \frac{42}{107}$$

25.7 Random Variable and Its Distribution

Random Variable

Let S be the sample space associated with a given random experiment. Then, a real valued function X which assigns to each event $w \in S$ to a unique real number $X(w)$ is called a random variable.

A random variable is a function that associate a unique numerical value with every outcome of an experiment. The value of the random variable will vary from trial to trial as the experiment is repeated.

A random variable is usually denoted by the capital letters $X, Y, Z, \ldots$ etc.

e.g. A coin is tossed ten times. The random variable X is the number of tails that are noted. X can only take the values $0, 1, 2, \ldots, 10$. So, X is a discrete random variable.

There are two types of random variable

1. **Discrete random variable** If the range of the real function $X : U \to R$ is a finite set or an infinite set of real numbers, it is called a discrete random variable.
2. **Continuous random variable** If the range of X is an interval (a, b) of R, then X is called a continuous random variable.

e.g. In tossing of two coins $S = \{HH, HT, TH, TT\}$, let X denotes number of heads in tossing of two coins, then

$$X(HH) = 2$$

$$X(TH) = 1$$

$$X(TT) = 0$$

Probability Distribution

If a random variable X takes values $X_1, X_2, \ldots, X_n$ with respective probabilities $P_1, P_2, \ldots, P_n$, then

X	X_1	X_2	X_3 ...	X_n
$P(X)$	P_1	P_2	P_3 ...	P_n

is known as the probability distribution of X.

OR

Probability distribution gives the values of the random variable along with the corresponding probabilities.

It satisfies the following conditions.

(i) $0 \le P(x_i) \le 1$

(ii) $\Sigma P(x_i) = 1$

e.g. Probability distribution when two coins are tossed.

Let X denote the number of heads occurred, then

$P(X = 0)$ = Probability of occurrence of zero head

$$= P(TT) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$$

$P(X = 1)$ = Probability of occurrence of one head

$$= HP(HT) + P(TH)$$

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

$P(X = 2)$ = Probability of occurrence of two heads

$$= P(HH) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$$

Thus, the probability distribution when two coins are tossed is as given below

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

Mean and Variance of a Random Variable

If X is a discrete random variable which assumes values $x_1, x_2, x_3, \ldots, x_n$ with respective probabilities $P_1, P_2, P_3, \ldots, P_n$, then the mean $\overline{X}$ of X is defined as

$$\overline{X} = p_1x_1 + p_2x_2 + \ldots + p_nx_n$$

$$\Rightarrow \quad \overline{X} = \sum_{i=1}^{n} p_i x_i$$

and variance of X is defined as var (X)

$$= p_1(x_1 - \overline{X})^2 + p_2(x_2 - \overline{X})^2 + \ldots + p_n(x_n - \overline{X})^2$$

Scanned by CamScanner

$$= \sum_{i=1}^{n} p_i (x_i - \overline{X})^2$$

where, $\overline{X} = \sum_{i=1}^{n} p_i x_i$ is the mean of X.

$$\Rightarrow \quad \text{var}(X) = \sum_{i=1}^{n} p_i x_i^2 - \left(\sum_{i=1}^{n} p_i x_i\right)^2$$

The square root of the variance gives the standard deviation

i.e. $\sqrt{\text{var}(X)} = \sqrt{\sigma^2} = \sigma$

Note
- The mean of a random variable X is also known as its mathematical expectation or expected value and is denoted by $E(X)$.
- The variance and standard deviation of a random variable are always non-negative.

Important Results

(i) **Variance** $V(X) = \sigma_x^2 = E(X^2) - \{E(X)\}^2$

where, $E(X^2) = \sum_{i=1}^{n} x_i^2 P(x_i)$

(ii) **Standard Deviation**

$\sqrt{V(X)} = \sigma_x = \sqrt{E(X^2) - \{E(X)\}^2}$

(iii) If $Y = aX + b$, then

(a) $E(Y) = E(aX + b) = aE(X) + b$

(b) $\sigma_Y^2 = V(Y) = a^2 V(X) = a^2 \sigma_x^2$

(3) $\sigma_Y = \sqrt{V(Y)} = |a| \sigma_x$

(iv) If $Z = aX^2 + bX + c$,

then $E(Z) = E(aX^2 + bX + c) = aE(X^2) + bE(X) + c$

Example 12. An urn contains 4 white and 3 red balls. Let X be the number of red balls in a random draw of 3 balls. Find the mean and variance of X.

(1) $\frac{5}{7}, \frac{25}{49}$ (2) $\frac{9}{7}, \frac{24}{49}$

(3) $\frac{12}{7}, \frac{24}{49}$ (4) None of these

Sol. (2) When 3 balls are drawn at random, there may be no red ball, 1 red ball, 2 red balls or 3 red balls. Let X denotes the random variable showing the number of red balls in a draw of 3 balls.

Then, X can take the values 0, 1, 2 or 3.

$P(X=0) = P$ (getting no red ball)

$= P$ (getting 3 white balls)

$$= \frac{^4C_3}{^7C_3} = \left(\frac{4\times3\times2}{3\times2\times1} \times \frac{3\times2\times1}{7\times6\times5}\right) = \frac{4}{35}$$

$P(X=1) = P$ (getting 1 red and 2 white balls)

$$= \frac{^3C_1 \times {^4C_2}}{^7C_3}$$

$$= \left(\frac{3\times4\times3}{2} \times \frac{3\times2\times1}{7\times6\times5}\right) = \frac{18}{35}$$

$P(X=2) = P$ (getting 2 red and 1 white balls)

$$= \frac{^3C_2 \times {^4C_1}}{^7C_3}$$

$$= \left(\frac{3\times2}{2\times1} \times 4 \times \frac{3\times2\times1}{7\times6\times5}\right) = \frac{12}{15}$$

$P(X=3) = P$ (getting 3 red balls)

$$= \frac{^3C_3}{^7C_3} = \frac{1\times3\times2\times1}{7\times6\times5}$$

Thus, the probability distribution of X is given below

$X = x_i$	0	1	2	3
p_i	$\frac{4}{35}$	$\frac{18}{35}$	$\frac{12}{35}$	$\frac{1}{35}$

$\therefore$ Mean, $\mu = \sum x_i p_i$

$$= 0 \times \frac{4}{35} + 1 \times \frac{18}{35} + 2 \times \frac{12}{35} + 3 \times \frac{1}{35}$$

$$= \frac{45}{35} = \frac{9}{7}$$

Variance, $\sigma^2 = \sum x_i^2 p_i - \mu^2$

$$= 0 \times \frac{4}{35} + 1 \times \frac{18}{35} + 4 \times \frac{12}{35} + 9 \times \frac{1}{35} - \frac{81}{49}$$

$$= \left(\frac{15}{7} - \frac{81}{49}\right) = \frac{105 - 81}{49} = \frac{24}{49}$$

25.8 Binomial Distribution

Bernoulli Trial

In a random experiment, if there are any two events, "Success and Failure" and the sum of the probabilities of these two events is one, then any outcome of such experiment is known as a Bernoulli Trial.

Let a binomial experiment has probability of success p and that of failure q (i.e. $p + q = 1$). If E be an event and let X = number of success i.e. number of times event E occurs in n trials.

Then, probability distribution of binomial distribution with parameters n and p is given by

$P(X = r)$ = Probability of r success in n trials

$= {^nC_r} p^r q^{n-r}$ $\quad [\because p + q = 1]$

$= (r+1)$th term in the expansion of $(q+p)^n$.

It is written as $X \sim B(n, p)$ or $X \sim Bi(n, p)$.

Note The trials must meet the following requirements
- the total number of trials is fixed in advance.

- there are just two outcomes of each trials, success and failure.
- all the trials have the same probability of success.
- the outcomes of all the trials are statistically independent.

Mean and Variance of Binomial Distribution

Let $X \sim B(n, p)$, then $P(X = r) = {}^nC_r p^r q^{n-r}$ where, $r = 0, 1, 2, \ldots, n$ and $p + q = 1$.

$\therefore$ Mean, $\overline{X} = E(X) = np$ and variance, $\text{var}(X) = npq$
Standard deviation $= \sqrt{npq}$

Relation between Mean and Variance

$$\text{Mean} - \text{Variance} = np - npq$$
$$= np(1-q) = np^2 > 0$$
$$\Rightarrow \quad \text{Mean} > \text{Variance}$$

i.e. for binomial variable X, value of mean is always greater than its variance.

Mode of Binomial Distribution

In binomial distribution, the value of r for which $P(X = r)$ is maximum, is known as mode of binomial distribution.

$$\therefore \quad (n+1)p - 1 \le r \le (n+1)p$$

Example 13. The mean and variance of a random variable X having a binomial distribution are 4 and 2 respectively, then $P(X = 1)$ is

(1) $\frac{1}{32}$ (2) $\frac{1}{8}$ (3) $\frac{1}{4}$ (4) $\frac{1}{16}$

Sol. (1) If n and p are the parameters of the binomial distribution, then

Mean, $np = 4$

Var $(X) = npq = 2$

$$\therefore \quad p = \frac{1}{2}, q = \frac{1}{2}, n = 8$$

$$\therefore \quad P(X = 1) = {}^8C_1\left(\frac{1}{2}\right)^1 \cdot \left(\frac{1}{2}\right)^7 = \frac{1}{32}$$

25.9 Poisson Distribution

It is limiting case of binomial distribution. Let the number of events n is large $(n \to \infty)$ and probability of success in each experiment is 0 and $np = \lambda$ (say) is finite, then

$$P(X = r)$$

or $$P(r) = \frac{e^{-\lambda}\lambda^r}{r!},$$

where $r = 0, 1, 2, \ldots$

and $\lambda = np$. Here, λ is known as parameter of poisson distribution.

$P(r+1) = \frac{\lambda}{r+1} P(r)$ is known as recurrence formula.

Note
- $\sum_{r=0}^{\infty} P(r) = 1$
- If λ_1 and λ_2 are parameter of variables X and Y, then parameter of $(X + Y)$ will be $(\lambda_1 + \lambda_2)$.
- In poisson distribution, mean = variance = $\lambda = np$

Example 14. If mean of a poisson distribution of a random variable X is 2, then the value of $P(X > 1.5)$ is

(1) $\frac{3}{e^2}$ (2) $\frac{3}{e}$
(3) $1 - \frac{3}{e}$ (4) $1 - \frac{3}{e^2}$

Sol. (4) Since, $P(X = r) = \frac{e^{-\lambda}\lambda^r}{r!}$ [where, λ = mean]

$$\therefore \quad P(X = r > 1.5) = P(2) + P(3) + \ldots \infty$$
$$= 1 - P(X = r \le 1) = 1 - P(0) - P(1)$$
$$= 1 - \left(e^{-2} + \frac{e^{-2} \times 2}{1!}\right) = 1 - \frac{3}{e^2}$$

detail ...

9.2. NORMAL DISTRIBUTION

The normal distribution was first discovered in 1733 by English mathematician De-Moivre, who obtained this continuous distribution as a limiting case of the binomial distribution and applied it to problems arising in the game of chance. It was also known to Laplace, no later than 1774 but through a historical error it was credited to Gauss, who first made reference to it in the beginning of 19th century (1809), as the distribution of errors in Astronomy. Gauss used the normal curve to describe the theory of accidental errors of measurements involved in the calculation of orbits of heavenly bodies. Throughout the eighteenth and nineteenth centuries, various efforts were made to establish the normal model as the underlying law ruling all continuous random variables. Thus, the name *"normal"*. These efforts, however, failed because of false premises. The normal model has, nevertheless, become the most important probability model in statistical analysis.

Definition *A r.v. X is said to have a normal distribution with parameters μ (called 'mean') and σ^2 (called 'variance') if its p.d.f. is given by the probability law :*

$$f(x; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\}$$

or

$$f(x; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}; -\infty < x < \infty, -\infty < \mu < \infty, \sigma > 0 \quad \dots (9{\cdot}1)$$

Remarks 1. When a *r.v.* is normally distributed with mean μ and standard deviation σ, it is customary to write X is distributed as $N(\mu, \sigma^2)$ and is expressed by $X \sim N(\mu, \sigma^2)$.

2. If $X \sim N(\mu, \sigma^2)$, then $Z = \frac{X-\mu}{\sigma}$, is a standard normal variate with $E(Z) = 0$ and $Var(Z) = 1$ and we write $Z \sim N(0, 1)$.

3. The *p.d.f.* of standard normal variate Z is given by :

$$\varphi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}, -\infty < z < \infty$$

Example 9·3. *X is normally distributed and the mean of X is 12 and S.D. is 4. (a) Find out the probability of the following :*

(a) (i) $X \geq 20$, *(ii)* $X \leq 20$, *and (iii)* $0 \leq X \leq 12$ *(b) Find* x', *when* $P(X > x') = 0{\cdot}24$.

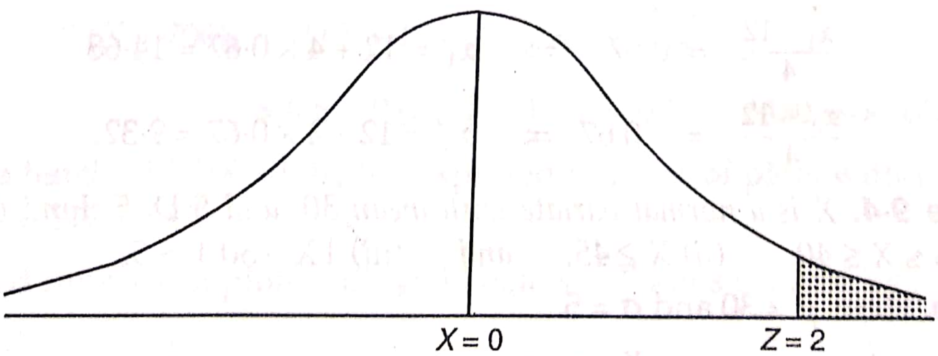
(c) Find x_0' *and* x_1', *when* $P(x_0' < X < x_1') = 0{\cdot}50$ *and* $P(X > x_1') = 0{\cdot}25$.

Solution. (*a*) We have $\mu = 12$, $\sigma = 4$, *i.e.*, $X \sim N(12, 16)$.

(*i*) $P(X \geq 20) = ?$

When $X = 20$, $Z = \dfrac{20 - 12}{4} = 2$

$\therefore P(X \geq 20) = P(Z \geq 2) = 0{\cdot}5 - P(0 \leq Z \leq 2) = 0{\cdot}5 - 0{\cdot}4772 = 0{\cdot}0228$



(ii) $P(X \le 20) = 1 - P(X \ge 20) = 1 - 0.0228 = 0.9722$ [From part (i)]

(iii) $P(0 \le X \le 12) = P(-3 \le Z \le 0)$, $\left(Z = \frac{X-12}{4}\right)$

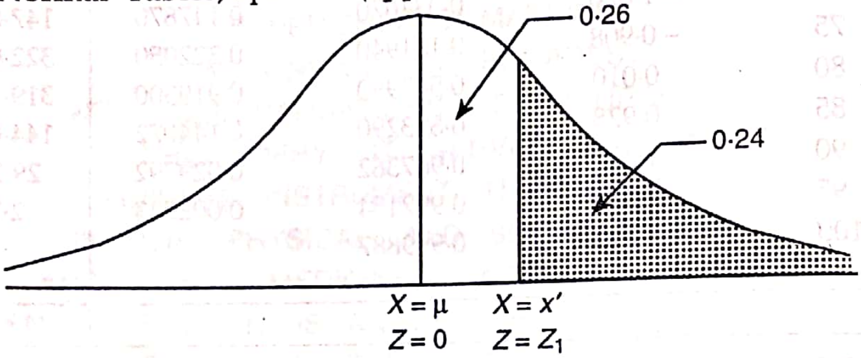
$= P(0 \le Z \le 3) = 0.49865$ (From symmetry)

(b) When $X = x'$, $Z = \frac{x'-12}{4} = z_1$, (say).

Then, we are given :

$P(X > x') = 0.24 \Rightarrow P(Z > z_1) = 0.24 \Rightarrow P(0 < Z < z_1) = 0.26$

$\therefore$ From Normal Tables, $z_1 = 0.71$ (approx.)



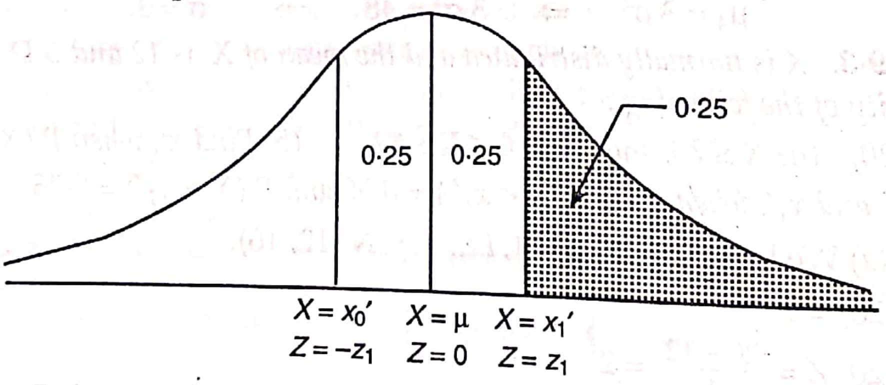
Hence $\frac{x'-12}{4} = 0.71 \Rightarrow x' = 12 + 4 \times 0.71 = 14.84$.

(c) We are given : $P(x_0' < X < x_1') = 0.50$ and $P(X > x_1') = 0.25$ ... (*)

From (*), obviously the points x_0' and x_1' are located as shown in following adjoining.

When $X = x_1'$, $Z = \frac{x_1'-12}{4} = z_1$, (say),

and when $X = x_0'$, $Z = \frac{x_0'-12}{4} = -z_1$ (It is obvious from the figure.)



We have $P(Z > z_1) = 0.25 \Rightarrow P(0 < Z < z_1) = 0.25 \therefore z_1 = 0.67$ (From Tables)

Hence $\frac{x_1'-12}{4} = 0.67 \Rightarrow x_1' = 12 + 4 \times 0.67 = 14.68$

and $\frac{x_0'-12}{4} = -0.67 \Rightarrow x_0' = 12 - 4 \times 0.67 = 9.32$.

Example 9·4. *X is a normal variate with mean 30 and S.D. 5. Find the probabilities that (i) $26 \le X \le 40$, (ii) $X \ge 45$, and (iii) $|X - 30| > 5$.*

Solution. Here $\mu = 30$ and $\sigma = 5$.

(i) When $X = 26$, $Z = \frac{X-\mu}{\sigma} = \frac{26-30}{5} = -0.8$

variance no ·

9·3. RECTANGULAR (OR UNIFORM) DISTRIBUTION

Definition. *A random variable X is said to have a continuous rectangular (uniform) distribution over an interval (a, b), i.e., $(-\infty < a < b < \infty)$, if its p.d.f. is given by :*

$$f(x\,;a,b) = \begin{cases} \dfrac{1}{b-a}, & \text{if } a < x < b \\ 0, & \text{otherwise} \end{cases} \qquad \ldots (9{\cdot}19)$$

In particular

$$\text{Mean} = \mu_1' = \frac{1}{b-a}\left(\frac{b^2-a^2}{2}\right) = \frac{b+a}{2} \quad \ldots($$

and $$\mu_2' = \frac{1}{b-a}\left(\frac{b^3-a^3}{3}\right) = \frac{1}{3}(b^2+ab+a^2)$$

$\therefore$ $$\text{Variance} = \mu_2' - \mu_1'^2 = \frac{1}{3}(b^2+ab+a^2) - \left\{\frac{1}{2}(b+a)\right\}^2 = \frac{1}{12}(b-a)^2 \quad \ldots($$

Example 9·21. *If X is uniformly distributed with mean 1 and variance $\frac{4}{3}$, find $P(X < 0)$.*

Solution. Let $X \sim U[a, b]$, so that $p(x) = \frac{1}{b-a}$, $a < x < b$. We are given :

$$\text{Mean} = \frac{1}{2}(b + a) = 1 \Rightarrow b + a = 2 \text{ and } \text{Var}(X) = \frac{1}{12}(b - a)^2 = \frac{4}{3} \Rightarrow b - a = \pm 4.$$

Solving, we get $a = -1$ and $b = 3$; $(a < b)$. $\therefore \quad p(x) = \frac{1}{4}; -1 < x < 3$

$$P(X < 0) = \int_{-1}^{0} p(x)\, dx = \frac{1}{4} \left| x \right|_{-1}^{0} = \frac{1}{4}.$$

9·8. EXPONENTIAL DISTRIBUTION

Definition. *A r.v. X is said to have an exponential distribution with parameter $\theta > 0$, if its p.d.f. is given by :*

$$f(x, \theta) = \begin{cases} \theta e^{-\theta x}, & x \geq 0 \\ 0, & \text{otherwise} \end{cases} \qquad \ldots (9{\cdot}26)$$

The cumulative distribution function $F(x)$ is given by

$$\therefore \quad F(x) = \int_0^x f(u)\, du = \theta \int_0^x \exp(-\theta u)\, du$$

$$F(x) = \begin{cases} 1 - \exp(-\theta x), & \geq 0 \\ 0, & \text{otherwise} \end{cases} \qquad \ldots (9{\cdot}26\,a)$$

9·8·1. Moment Generating Function of Exponential Distribution

$$M_X(t) = E(e^{tX}) = \theta \int_0^\infty e^{tx} e^{-\theta x}\, dx = \theta \int_0^\infty \exp\{-(\theta - t)x\}\, dx$$

$$= \frac{\theta}{(\theta - t)} = \left(1 - \frac{t}{\theta}\right)^{-1} = \sum_{r=0}^{\infty} \left(\frac{t}{\theta}\right)^r, \; \theta > t$$

$$\therefore \quad \mu_r' = E(X^r) = \text{Coefficient of } \frac{t^r}{r!} \text{ in } M_X(t) = \frac{r!}{\theta^r}\, ; r = 1, 2, \ldots$$

$$\Rightarrow \quad \text{Mean} = \mu_1' = \frac{1}{\theta} \text{ and variance} = \mu_2 = \mu_2' - \mu_1'^2 = \frac{2}{\theta^2} - \frac{1}{\theta^2} = \frac{1}{\theta^2}$$

Hence, if $X \sim \exp(\theta)$, then Mean $= \dfrac{1}{\theta}$ and Variance $= \dfrac{1}{\theta^2}$. $\qquad \ldots (9{\cdot}26b)$

Remark. Variance $= \dfrac{1}{\theta^2} = \dfrac{1}{\theta} \cdot \dfrac{1}{\theta} = \dfrac{\text{Mean}}{\theta}$

$\therefore$ Variance > Mean, if $0 < \theta < 1$

Variance = Mean, if $\theta = 1$

and Variance < Mean, if $\theta > 1$

Hence for the exponential distribution,

Variance >, = , or < Mean, for different values of the parameter.

Scanned by CamScanner