3. If X is a random variable then

(i) $\frac{1}{X}$, where $\left(\frac{1}{X}\right)(\omega) = \infty$ if $X(\omega) = 0$, (ii) $X_+(\omega) = \max\{0, X(\omega)\}$,

(iii) $X_-(\omega) = -\min\{0, X(\omega)\}$, and (iv) $|X|$, are random variables.

4. If X_1 and X_2 are random variables, then (i) $\max[X_1, X_2]$, and (ii) $\min[X_1, X_2]$ are also random variables.

5. If X is a *r.v.* and $f(.)$ is a continuous function, then $f(X)$ is a *r.v.*

6. If X is a *r.v.* and $f(.)$ is an increasing function, then $f(X)$ is a *r.v.*

Corollary. If f is a function of bounded variations on every finite interval $[a, b]$, and X is a *r.v.* then $f(X)$ is a *r.v.*

(*Proofs of the above results or theorems are beyond the scope of this book.*)

5·2. DISTRIBUTION FUNCTION

Definition. *Let X be a random variable. The function F defined for all real x by*

$$F(x) = P(X \le x) = P\{\omega : X(\omega) \le x\}, -\infty < x < \infty, \quad \ldots (5\cdot1)$$

is called the distribution function (d.f.) of the r.v. (X).

Remark. A distribution function is also called the cumulative distribution function. Sometimes, the notation $F_X(x)$ is used to emphasise the fact that the distribution function is associated with the particular random variable X. Clearly, the domain of the distribution function is $(-\infty, \infty)$ and its range is $[0, 1]$.

5·2·1. Properties of Distribution Function. We now proceed to derive a number of properties common to all distribution functions.

1. *If F is the d.f. of the r.v. X and if $a < b$, then $P(a < X \le b) = F(b) - F(a)$.*

Proof. The events '$a < X \le b$' and '$X \le a$' are disjoint and their union is the event '$X \le b$'., Hence by addition theorem of probability :

$$P(a < X \le b) + P(X \le a) = P(X \le b)$$

$$\Rightarrow \quad P(a < X \le b) = P(X \le b) - P(X \le a) = F(b) - F(a) \quad \ldots (5\cdot2)$$

Cor. 1. $P(a \le X \le b) = P\{(X = a) \cup (a < X \le b)\} = P(X = a) + P(a < X \le b)\}$

$$= P(X = a) + [F(b) - F(a)] \quad \ldots (5\cdot2a)$$

Similarly, we get

$$P(a < X < b) = P(a < X \le b) - P(X = b) = F(b) - F(a) - P(X = b) \quad \ldots (5\cdot2b)$$

$$P(a \le X < b) = P(a < X < b) + P(X = a)$$

$$= F(b) - F(a) - P(X = b) + P(X = a) \quad \ldots (5\cdot2c)$$

Remark. When $P(X = a) = 0$ and $P(X = b) = 0$, all four events $a \le X \le b$, $a < X < b$, $a \le X < b$ and $a < X \le b$ have the same probability $F(b) - F(a)$.

2. *If F is d.f. of one-dimensional r.v. X, then (i) $0 \le F(x) \le 1$, (ii) $F(x) \le F(y)$ if $x < y$.*

In other words, all distribution functions are monotonically non-decreasing and lie between 0 and 1.

3. *If F is d.f. of one-dimensional r.v. X, then*

$$F(-\infty) = \lim_{x \to -\infty} F(x) = 0 \quad \text{and} \quad F(\infty) = \lim_{x \to \infty} F(x) = 1$$

Proof. Let us express the whole sample space S as a countable union of disjoint events as follows :

$$S = \left\{ \bigcup_{n=1}^{\infty} (-n < X \le -n+1) \right\} \cup \left\{ \bigcup_{n=0}^{\infty} (n < X \le n+1) \right\}.$$

$$\Rightarrow \qquad P(S) = \sum_{n=1}^{\infty} P(-n < X \le -n+1) + \sum_{n=0}^{\infty} P(n < X \le n+1)$$

$$\Rightarrow \qquad 1 = \lim_{a \to \infty} \sum_{n=1}^{a} \{F(-n+1) - F(-n)\} + \lim_{b \to \infty} \sum_{n=0}^{b} \{F(n+1) - F(n)\}$$

$$= \lim_{a \to \infty} \{F(0) - F(-a)\} + \lim_{b \to \infty} \{F(b+1) - F(0)\}$$

$$= \{F(0) - F(-\infty)\} + \{F(\infty) - F(0)\}$$

$$\therefore \qquad 1 = F(\infty) - F(-\infty) \qquad \ldots (*)$$

Since $-\infty < \infty$, $F(-\infty) \le F(\infty)$. Also $F(-\infty) \ge 0$ and $F(\infty) \le 1$

$$\therefore \qquad 0 \le F(-\infty) \le F(\infty) \le 1 \qquad \ldots (**)$$

From (*) and (**), we get $\quad F(-\infty) = 0 \;$ and $\; F(\infty) = 1.$

Remarks 1. Discontinuities of $F(x)$ are at most countable.

2. $$F(a) - F(a-0) = \lim_{h \to 0} P(a - h \le X \le a), h < 0$$

and $$F(a+0) - F(a) = \lim_{h \to 0} P(a \le X \le a + h) = 0, h > 0$$

5·3. DISCRETE RANDOM VARIABLE

A variable which can assume only a countable number of real values and for which the value which the variable takes depends on chance, is called a *discrete random variable* (or discrete stochastic variable or discrete chance variable). In other words, *a real valued function defined on a discrete sample space is called a discrete random variable.*

Examples of discrete random variables are marks obtained in a test, number of accidents per month, number of telephone calls per unit time, number of successes in n trials, and so on.

5·3·1. Probability Mass Function. If X is a one-dimensional discrete random variable taking at most a countably infinite number of values $x_1, x_2, \ldots$ then its probabilistic behaviour at each real point is described by a function called the probability mass function (or discrete density function) which is defined below :

Definition. *If X is a discrete random variable with distinct values $x_1, x_2, \ldots, x_n, \ldots$, then the function $p(x)$ defined as :*

$$p_X(x) = \begin{cases} P(X = x_i) = p_i & \text{if } x = x_i \\ 0 & \text{if } x \ne x_i \,;\, i = 1, 2, \ldots \end{cases}$$

is called the probability mass function of r.v. X.

The set of ordered pairs $\{x_i, p(x_i) \,;\, i = 1, 2, \ldots, n, \ldots\}$ or $\{(x_1, p_1), (x_2, p_2), \ldots, (x_n, p_n), \ldots\}$ specifies the *probability distribution* of the *r.v.* X.

Remarks 1. The numbers $p(x_i)$; $i = 1, 2, \ldots$ must satisfy the following conditions :

(i) $p(\bar{x}_i) \geq 0 \; \forall \; i$, and (ii) $\sum_{i=1}^{\infty} p(x_i) = 1$

2. The set of values which X takes is called the *spectrum* of the random variable.

3. For discrete *r.v.*, a knowledge of the probability mass function enables us to compute probabilities of arbitrary events. In fact, if E is a set of real numbers,

$P(X \in E) = \sum_{x \in E \cap S} p(x)$, where S is the sample space.

Illustration. Toss of a coin, $S = \{H, T\}$. Let X be the random variable defined by :

$X(H) = 1$, *i.e.*, $X = 1$, if '*Head*' occurs.

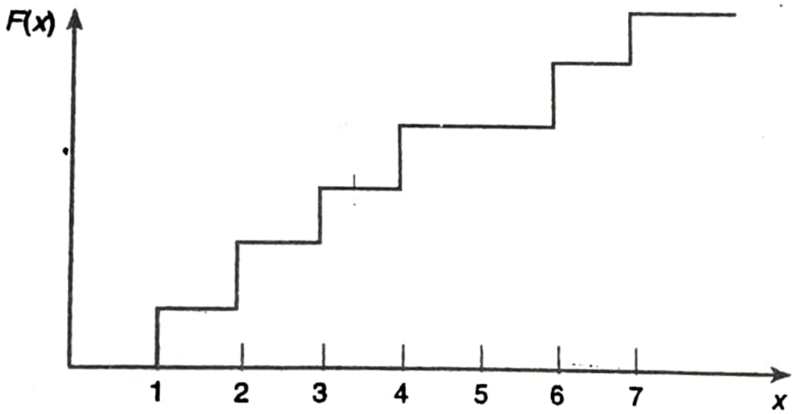
$X(T) = 0$, *i.e.*, $X = 0$, if '*Tail*' occurs.

If the coin is 'fair', the probability function is $P(\{H\}) = P(\{T\}) = \frac{1}{2}$,

and we can speak of the probability distribution of the *r.v.* X as :

$$P(X = 1) = P(\{H\}) = \tfrac{1}{2}, \quad \text{and} \quad P(X = 0) = P(\{T\}) = \tfrac{1}{2}.$$

5·3·2. Discrete Distribution Function. In this case there are a countable number of points $x_1, x_2, x_3, \ldots$ and number



$p_i \geq 0$, $\sum_i^{\infty} p_i = 1$, such that

$F(x) = \sum_{i : x_i \leq x} p_i.$

For example, if x_i is just the integer i, so that $P(X = i) = p_i$; $i = 1, 2, 3, \ldots$, then $F(x)$ is a "*step function*" having jump p_i at i, and being constant between each pair of integers.

Theorem 5·1. *$p(x_j) = P(X = x_j) = F(x_j) - F(x_{j-1})$, where F is the d.f. of X.*

Proof. Let $x_1 < x_2 < \ldots$ We have

$$F(x_j) = P(X \leq x_j) = \sum_{i=1}^{j} P(X = x_i) = \sum_{i=1}^{j} p(x_i) \quad \text{and} \quad F(x_{j-1}) = P(X \leq x_{j-1}) = \sum_{i=1}^{j-1} p(x_i)$$

$$\therefore \qquad F(x_j) - F(x_{j-1}) = p(x_j) \qquad \ldots (5·5)$$

Thus, given the distribution function of discrete random variable, we can compute its probability mass function.

Example 5·1. *A random variable X has the following probability function :*

Values of X, x :	0	1	2	3	4	5	6	7
p(x) :	0	k	$2k$	$2k$	$3k$	k^2	$2k^2$	$7k^2 + k$

(i) Find k, (ii) Evaluate $P(X < 6)$, $P(X \geq 6)$, and $P(0 < X < 5)$, (iii) If $P(X \leq a) > \frac{1}{2}$, find the minimum value of a, and (iv) Determine the distribution function of X.

Solution. Since $\sum_{x=0}^{7} p(x) = 1, \qquad k + 2k + 2k + 3k + k^2 + 2k^2 + 7k^2 + k = 1$

$\Rightarrow \quad 10k^2 + 9k - 1 = 0 \quad \Rightarrow \quad (10k - 1)(k + 1) = 0 \quad \Rightarrow \quad k = \frac{1}{10}$ or $k = -1$.

But since $p(x)$ *cannot be negative*, $k = -1$, *is rejected*. Hence, $k = \frac{1}{10}$.

(ii) $P(X < 6) = P(X = 0) + P(X = 1) + \ldots + P(X = 5) = \frac{1}{10} + \frac{2}{10} + \frac{2}{10} + \frac{3}{10} + \frac{1}{100} = \frac{81}{100}$

Now $\qquad P(X \geq 6) = 1 - P(X < 6) = 1 - \frac{81}{100} = \frac{19}{100}$

$$P(0 < X < 5) = P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4) = 8k = \frac{4}{5}.$$

(iii) $P(X \leq a) > \frac{1}{2}$. By trial, we get $a = 4$.

(iv) The distribution function $F_X(x)$ of X is given in the adjoining Table.

X	$F_X(x) = P(X \leq x)$
0	0
1	$k = \frac{1}{10}$
2	$3k = \frac{3}{10}$
3	$5k = \frac{5}{10}$
4	$8k = \frac{4}{5}$
5	$8k + k^2 = \frac{81}{100}$
6	$8k + 3k^2 = \frac{83}{100}$
7	$9k + 10k^2 = 1$

Example 5·2. *If,* $p(x) = \begin{cases} \frac{x}{15}; x = 1, 2, 3, 4, 5 \\ 0, \text{ elsewhere} \end{cases}$

Find (i) $P\{X = 1 \text{ or } 2\}$, *and (ii)* $P\left\{\frac{1}{2} < X < \frac{5}{2} \,\middle|\, X > 1\right\}$.

Solution. (i) $P(X = 1 \text{ or } 2) = P(X = 1) + P(X = 2) = \frac{1}{15} + \frac{2}{15} = \frac{1}{5}$

(ii) $$P\left(\frac{1}{2} < X < \frac{5}{2} \,\middle|\, x > 1\right) = \frac{P\left\{\left(\frac{1}{2} < X < \frac{5}{2}\right) \cap (X > 1)\right\}}{P(X > 1)} = \frac{P\{(X = 1 \text{ or } 2) \cap (X > 1)\}}{P(X > 1)}$$

$$= \frac{P(X = 2)}{1 - P(X = 1)} = \frac{2/15}{1 - (1/15)} = \frac{1}{7}.$$

Example 5·3. *Two dice are rolled. Let X denote the random variable which counts the total number of points on the upturned faces, Construct a table giving the non-zero values of the probability mass function and draw the probability chart. Also find the distribution function of X.*

Solution. If both dice are unbiased and the two rolls are independent, then each sample point of sample space S has probability $\frac{1}{36}$. Then

$$p(2) = P(X = 2) = P\{1, 1\} = \frac{1}{36}; \quad p(3) = P(X = 3) = P\{(1, 2), (2, 1)\} = \frac{2}{36}$$

$$p(4) = P(X = 4) = P\{(1, 3), (2, 2), (3, 1)\} = \frac{3}{36}$$

$$p(5) = P(X = 5) = P\{(1, 4), (2, 3) (3, 2), (4, 1)\} = \frac{4}{36}$$

$$\vdots \qquad \vdots$$

$$p(11) = P(X = 11) = P\{(6, 5), (5, 6)\} = \frac{2}{36}; \qquad p(12) = P(X = 12) = P\{(6, 6)\} = \frac{1}{36}$$

These values are summarized in the following *probability table* :

X :	2	3	4	5	6	7	8	9	10	11	12
$p(x)$:	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

The chart of the probability distribution is given below :

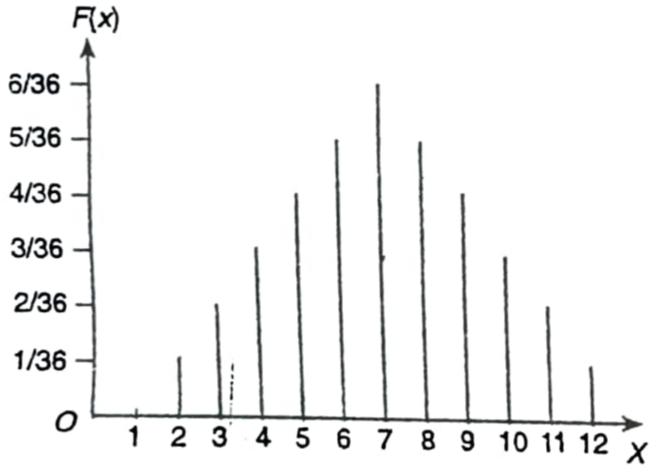


Fig. 5·1. (a) Probability Function of X

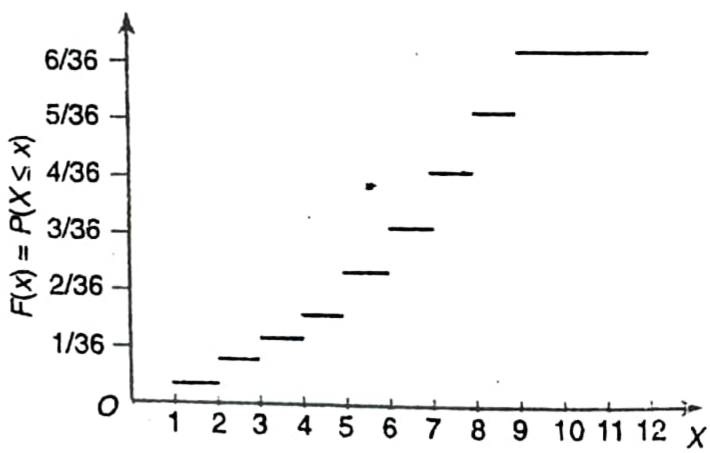


Fig. 5·1 (b) : Distribution Function of X.

Distribution Function.

$$F(1) = P(X \le 1) = 0, \; F(2) = P(X \le 2) = \frac{1}{36}$$

$$F(3) = P(X \le 3) = P(X = 2) + P(X = 3)$$

$$= p(2) + p(3) = \frac{1}{36} + \frac{2}{36} = \frac{3}{36}$$

$$F(4) = P(X \le 4) = P(X = 2) + P(X = 3) + P(X = 4)$$

$$= \frac{1}{36} + \frac{2}{36} + \frac{3}{36} = \frac{6}{36}, \text{ and so on.}$$

The distribution function of X is shown in the adjointing Table.

$$F(x) = \begin{cases} 0, & \text{for } x < 2 \\ \frac{1}{36}, & \text{for } 2 \le x < 3 \\ \frac{3}{36}, & \text{for } 3 \le x < 4 \\ \frac{6}{36}, & \text{for } 4 \le x < 5 \\ \vdots & \vdots \\ \frac{35}{36}, & \text{for } 11 \le x < 12 \\ 1, & \text{for } x \ge 12 \end{cases}$$

Example 5·4. *An experiment consists of three independent tosses of a fair coin. Let :*

X = The number of heads, Y = the number of head runs, Z = The length of head runs, a head run being defined as consecutive occurrence of at least two heads, its length then being the number of heads occurring together in three tosses of the coin.

Find the probability function of (i) X, (ii) Y, (iii) Z, (iv) X + Y, and (v) XY and construct probability tables and draw their probability charts.

Solution.

TABLE 5·1

S. No.	Elementary event	Random Variables X	Y	Z	X + Y	XY
1	HHH	3	1	3	4	3
2	HHT	2	1	2	3	2
3	HTH	2	0	0	2	0
4	HTT	1	0	0	1	0
5	THH	2	1	2	3	2
6	THT	1	0	0	1	0
7	TTH	1	0	0	1	0
8	TTT	0	0	0	0	0

Here sample space is : $S = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$

(i) Obviously X is a *r.v.* which can take the values 0, 1, 2 and 3.

$$p(3) = P\,(HHH) = \left(\frac{1}{2}\right)^3 = \frac{1}{8}\cdot$$

$$p(2) = P\,(HHT \cup HTH \cup THH)$$

$$= P\,(HHT) + P\,(HTH) + P\,(THH) = \frac{1}{8} + \frac{1}{8} + \frac{1}{8} = \frac{3}{8}\cdot$$

Similarly $\quad p\,(1) = \frac{3}{8}$ and $p\,(0) = \frac{1}{8}\cdot$

These probabilities could also be obtained directly from the Table 5·1.

TABLE 5·2 : PROBABILITY DISTRIBUTION OF X

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

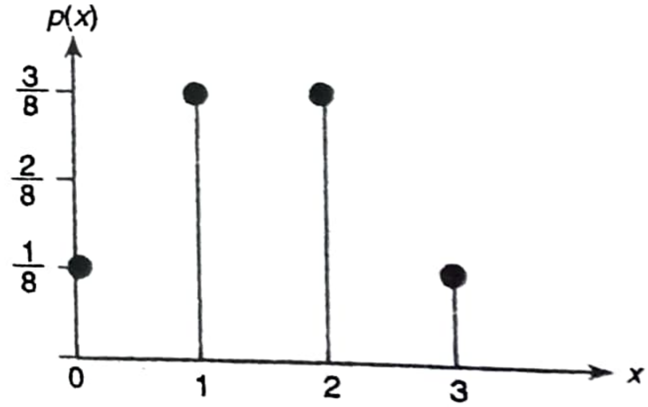


Fig. 5·2 : Probability chart of X

(ii)

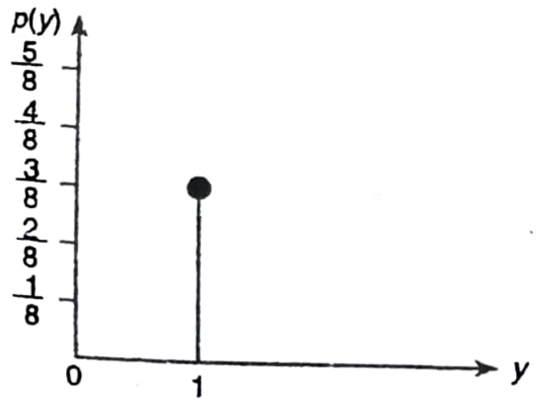


Fig. 5·3 : Probability chart of Y

TABLE 5·3 : PROBABILITY DISTRIBUTION OF Y

Values of Y , (y)	0	1
p(y)	$\frac{5}{8}$	$\frac{3}{8}$

This is obvious from *Table 5·1.*

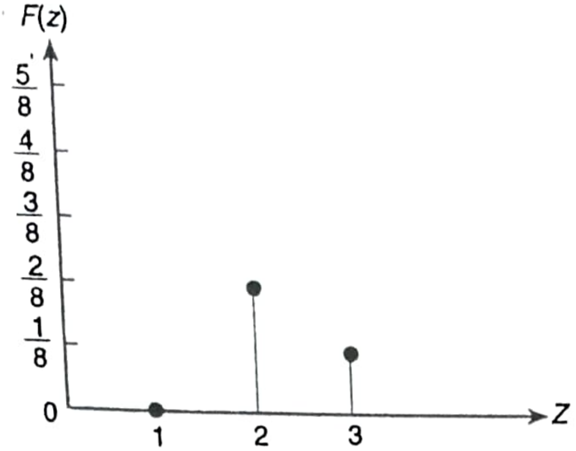


Fig. 5·4 : Probability Chart of Z

(*iii*) From Table 5·1, we have

TABLE 5·4 : PROBABILITY DISTRIBUTION OF Z

Values of Z, (z)	0	1	2	3
$p(z)$	$\frac{5}{8}$	0	$\frac{2}{8}$	$\frac{1}{8}$

(*iv*) Let $U = X + Y$. From Table 5·1, we get :

TABLE 5·5 : PROBABILITY DISTRIBUTION OF U

Values of U, (u)	0	1	2	3	4
$p(u)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{1}{8}$	$\frac{2}{8}$	$\frac{1}{8}$

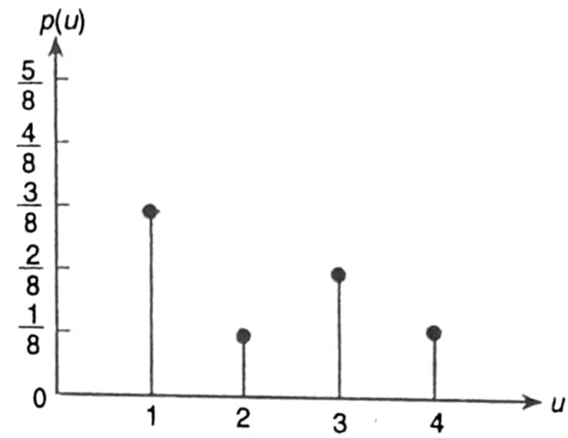


Fig. 5·5 : Probability Chart of $U = X + Y$

(*v*) Let $V = XY$

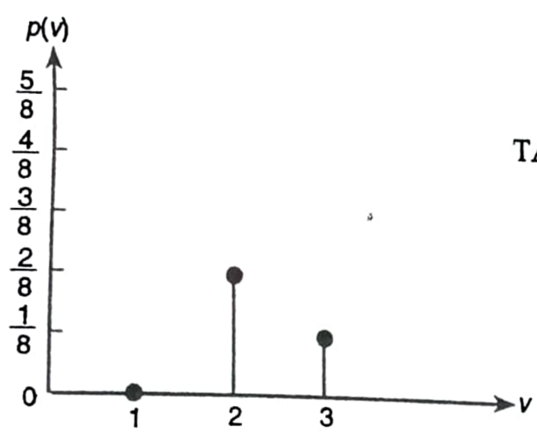


Fig. 5·6 : Probability Chart of $V = XY$

TABLE 5·6 : PROBABILITY DISTRIBUTION OF V

Values of V, (v)	0	1	2	3
$p(v)$	$\frac{5}{8}$	0	$\frac{2}{8}$	$\frac{1}{8}$

5·4. CONTINUOUS RANDOM VARIABLE

A random variable X is said to be continuous if it can take all possible values (integral as well as fractional) between certain limits. In other words, *a random variable is said to be continuous when its different values cannot be put in 1-1 correspondence with a set of positive integers.*

A continuous random variable is a random variable that (at least conceptually) can be measured to any desired degree of accuracy. Examples of continuous random variables are age, height, weight, etc.

5·4·1. Probability Density Function (*Concept and Definition*). Consider the small interval $(x, x + dx)$ of length dx round the point x. Let $f(x)$ be any continuous function of x so that $f(x)dx$ represents the probability that X falls in the infinitesimal interval $(x, x + dx)$. Symbolically,

$$P\,(x \le X \le x + dx) = f_X(x)\,dx \qquad \ldots (5\cdot5)$$

In the figure, $f(x)\,dx$ represents the area bounded by the curve $y = f(x)$, x-axis and ordinates at the points x and $x + dx$. The function $f_X(x)$ so defined is known as *probability density function or simply density function of random variable X and is usually abbreviated as p.d.f.* The expression, $f(x)dx$, usually written as $dF(x)$, is known as the *probability differential* and the curve $y = f(x)$ is known as the *probability density curve* or simply *probability curve.*

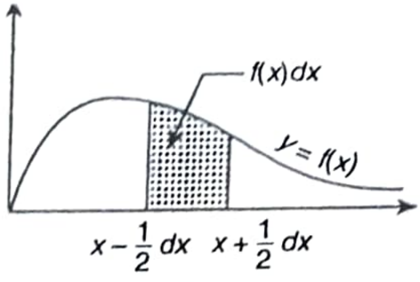


Fig. 5·7

Definition. *p.d.f. $f_X(x)$ of the r.v. X is defined as :*

$$f_X(x) = \lim_{\delta x \to 0} \frac{P(x \le X \le x + \delta x)}{\delta x} \qquad \text{...(5·5 a)}$$

The probability for a variate value to lie in the interval dx is $f(x)\,dx$ and hence the probability for a variate value to fall in the finite interval $[\alpha, \beta]$ is :

$$P(\alpha \le X \le \beta) = \int_{\alpha}^{\beta} f(x)\,dx, \qquad \text{... (5·5 b)}$$

which represents area between the curve $y = f(x)$, x-axis and the ordinates at $x = \alpha$ and $x = \beta$. Further, since total probability is unity, we have $\int_a^b f(x)\,dx = 1$, where $[a, b]$ is the range of the random variable X. The range of the variable may be finite or infinite.

The probability density function (*p.d.f.*) of a random variable (*r.v.*) X, usually denoted by $f_X(x)$ or simply by $f(x)$ has the following obvious properties :

(*i*) $f(x) \ge 0$, (*ii*) $\int_{-\infty}^{\infty} f(x)\,dx = 1$

(*iii*) The probability $P(E)$ given by : $P(E) = \int_E f(x)\,dx$ ... (5·5 *c*)

is well defined for any event E.

Important Remark. In case of discrete random variable, the probability at a point, *i.e.*, $P(x = c)$ is not zero for some fixed c. However, in case of continuous random variables the probability at a point is always zero, *i.e.*, $P(x = c) = 0$, for all possible values of c. This follows directly from (5·5*b*) by taking $\alpha = \beta = c$. This also agrees with our discussion earlier that $P(E) = 0$ does not imply that the event E is null or impossible event. This property of continuous *r.v.*, *viz.*,

$$P(X = c) = 0, \forall\, c \qquad \text{... (5·5 d)}$$

leads us to the following important result :

$$P(\alpha \le X \le \beta) = P(\alpha \le X < \beta) = P(\alpha < X \le \beta) = P(\alpha < X < \beta) \qquad \text{... (5·5 e)}$$

i.e., in case of continuous *r.v.*, it does not matter whether we include the end points of the interval from α to β.

However, this result is, in general, not true for discrete random variables.

5·4·2. Various Measures of Central Tendency, Dispersion, Skewness and Kurtosis for Continuous Probability Distribution. The formulae for these measures in case of discrete frequency distribution can be easily extended to the case of continuous probability distribution by simply replacing $p_i = f_i / N$ by $f(x)\,dx$, x_i by x and the summation over 'i' by integration over the specified range of the variable X.

Let $f_X(x)$ or $f(x)$ be the *p.d.f.* of a *r.v.* X, where X is defined from a to b. Then

(i) $$\text{Arithmetic Mean} = \int_a^b x\, f(x)\, dx \quad \dots (5{\cdot}6)$$

(ii) *Harmonic Mean.* Harmonic mean H is given by : $\frac{1}{H} = \int_a^b \frac{1}{x} \cdot f(x)\, dx$ ... (5·6 a)

(iii) *Geometric Mean.* Geometric mean G is given by : $\log G = \int_a^b \log x . f(x)\, dx$ ... (5·6 b)

(iv) $$\mu_r' \text{ (about origin)} = \int_a^b x^r . f(x)\, dx \quad \dots (5{\cdot}7)$$

$$\mu_r' \text{ (about the point } x = A) = \int_a^b (x - A)^r . f(x)\, dx \quad \dots (5{\cdot}7\ a)$$

$$\text{and } \mu_r \text{ (about mean)} = \int_a^b (x - \text{mean})^r . f(x) dx \quad \dots (5{\cdot}7\ b)$$

In particular, from (5·6) and (5·7), we have

$$\mu_1' \text{ (about origin)} = Mean = \int_a^b x f(x)\, dx \quad \text{and} \quad \mu_2' = \int_a^b x^2 f(x)\, dx$$

Hence $$\mu_2 = \mu_2' - \mu_1'^2 = \int_a^b x^2 f(x)\, dx - \left(\int_a^b x f(x)\, dx \right)^2 \quad \dots (5{\cdot}7\ c)$$

From (5·7), on putting $r = 3$ and 4 respectively, we get the values of μ_3' and μ_4' and consequently the moments about mean can be obtained by using the relations :

and $$\left. \begin{aligned} \mu_3 &= \mu_3' - 3\mu_2' \mu_1' + 2\mu_1'^3 \\ \mu_4 &= \mu_4' - 4\mu_3' \mu_1' + 6\mu_2' \mu_1'^2 - 3\mu_1'^4 \end{aligned} \right\} \quad \dots (5{\cdot}7\ d)$$

and hence β_1 and β_2 can be computed.

(*v*) *Median.* Median is the point which divides the entire distribution in two equal parts. In case of continuous distribution, median is the point which divides the total area into two equal parts. Thus if M is the median, then

$$\int_a^M f(x)\, dx = \int_M^b f(x)\, dx = \frac{1}{2} \quad \dots (5{\cdot}8)$$

Thus solving $$\int_a^M f(x)\, dx = \frac{1}{2} \quad \text{or} \quad \int_M^b f(x)\, dx = \frac{1}{2} \quad \dots (5{\cdot}8\ a)$$

for M, we get the value of median.

(*vi*) *Mean Deviation.* Mean deviation about the mean μ_1' is given by :

$$M.D. = \int_a^b |x - \text{mean}|\, f(x)\, dx \quad \dots (5{\cdot}9)$$

In general, mean diviation about an average 'A' is given by :

$$M.D. \text{ about '}A\text{'} = \int_a^b |x - A|\, f(x)\, dx \quad \dots (5{\cdot}9\ a)$$

(vii) *Quartiles and Deciles.* Q_1 and Q_3 are given by the equations :

$$\int_a^{Q_1} f(x)\,dx = \frac{1}{4} \quad \text{and} \quad \int_a^{Q_3} f(x)\,dx = \frac{3}{4} \qquad \ldots (5\cdot10)$$

D_i, ith decile is given by : $$\int_a^{D_i} f(x)\,dx = \frac{i}{10};\ i = 1, 2, \ldots, 9 \qquad \ldots (5\cdot10\,a)$$

(viii) *Mode.* Mode is the value of x for which $f(x)$ is maximum. Mode is thus the solution of $f'(x) = 0$ and $f''(x) < 0$, provided it lies in $[a, b]$. ... (5·11)

Remark. Various Measures of Central Tendency, Dispersion, Skewness and Kurtosis for Discrete Random Variable. Let X be a discrete *r.v.* with probability mass function (*p.m.f.*) $f_X(x)$ or $f(x)$. Then the various measures of central tendency, dispersion, skewness and kurtosis can be obtained similarly, on replacing integration ($\int$) with summation (Σ), over the given range of the variable X, in the formulae (5·6) to (5·10*a*).

Example 5·5. *The diameter of an electric cable, say X, is assumed to be a continuous random variable with p.d.f. :* $f(x) = 6x(1-x),\ 0 \le x \le 1.$

(i) Check that f(x) is p.d.f., and

(ii) Determine a number b such that $P(X < b) = P(X > b)$.

Solution. Obviously, for $0 \le x \le 1$, $f(x) \ge 0$. Now

$$\int_0^1 f(x)\,dx = 6\int_0^1 x(1-x)\,dx = 6\int_0^1 (x - x^2)\,dx = 6\left|\frac{x^2}{2} - \frac{x^3}{3}\right|_0^1 = 1$$

Hence $f(x)$ is the *p.d.f.* of *r.v.* X.

(ii) $\quad P(X < b) = P(X > b)$

$$\Rightarrow \int_0^b f(x)\,dx = \int_b^1 f(x)\,dx \quad \Rightarrow \quad 6\int_0^b x(1-x)\,dx = 6\int_b^1 x(1-x)\,dx$$

$$\Rightarrow \left|\frac{x^2}{2} - \frac{x^3}{3}\right|_0^b = \left|\frac{x^2}{2} - \frac{x^3}{3}\right|_b^1 \quad \Rightarrow \quad \left(\frac{b^2}{2} - \frac{b^3}{3}\right) = \left[\left(\frac{1}{2} - \frac{1}{3}\right) - \left(\frac{b^2}{2} - \frac{b^3}{3}\right)\right]$$

$$\Rightarrow 3b^2 - 2b^3 = (1 - 3b^2 + 2b^3) \Rightarrow 4b^3 - 6b^2 + 1 = 0 \Rightarrow (2b-1)(2b^2 - 2b - 1) = 0$$

$$\therefore \quad 2b - 1 = 0 \Rightarrow b = \frac{1}{2} \quad \text{or} \quad 2b^2 - 2b - 1 = 0 \Rightarrow b = \frac{2 \pm \sqrt{4+8}}{4} = \frac{1 \pm \sqrt{3}}{2}.$$

Hence $b = \frac{1}{2}$, is the only real value lying between 0 and 1 and satisfying (*).

Example 5·6. *A continuous random variable X has a p.d.f.* $f(x) = 3x^2,\ 0 \le x \le 1$. *Find a and b such that (i)* $P(X \le a) = P(X > a)$, *and (ii)* $P(X > b) = 0{\cdot}05$.

Solution. (i) Since $P(X \le a) = P(X > a)$, each must be equal to $\frac{1}{2}$, because total probability is always unity.

$$\therefore \quad P(X \le a) = \frac{1}{2} \quad \Rightarrow \quad \int_0^a f(x)\,dx = \frac{1}{2}$$

$$\Rightarrow \quad 3\int_0^a x^2\,dx = \frac{1}{2} \quad \Rightarrow \quad 3\left|\frac{x^3}{3}\right|_0^a = \frac{1}{2} \quad \Rightarrow \quad a = \left(\frac{1}{2}\right)^{\frac{1}{3}}$$

(ii) $P(X > b) = 0.05 \quad \Rightarrow \quad \int_b^1 f(x)\,dx = 0.05$

$$\Rightarrow \quad 3\left|\frac{x^3}{3}\right|_b^1 = \frac{1}{20} \quad \Rightarrow \quad 1 - b^3 = \frac{1}{20} \quad \Rightarrow \quad b = \left(\frac{19}{20}\right)^{\frac{1}{3}}.$$

Example 5·7. *Let X be a continuous random variable with p.d.f. :*

$$f(x) = \begin{cases} ax & , \ 0 \le x \le 1 \\ a & , \ 1 \le x \le 2 \\ -ax + 3a & , \ 2 \le x \le 3 \\ 0 & , \ \text{elsewhere} \end{cases}$$

(i) Determine the constant a, (ii) compute P (X ≤ 1·5).

Solution. (i) Constant 'a' is determined from the consideration that total probability is unity, *i.e.*, $\int_{-\infty}^{\infty} f(x)\,dx = 1$

$$\Rightarrow \quad \int_{-\infty}^{0} f(x)\,dx + \int_0^1 f(x)\,dx + \int_1^2 f(x)\,dx + \int_2^3 f(x)\,dx + \int_3^{\infty} f(x)\,dx = 1$$

$$\Rightarrow \quad \int_0^1 ax\,dx + \int_1^2 a.dx + \int_2^3 (-ax + 3a)\,dx = 1$$

$$\Rightarrow \quad a\left|\frac{x^2}{2}\right|_0^1 + a\left|x\right|_1^2 + a\left|-\frac{x^2}{2} + 3x\right|_2^3 = 1$$

$$\Rightarrow \quad \frac{a}{2} + a + a\left\{\left(-\frac{9}{2} + 9\right) - (-2 + 6)\right\} = 1 \quad \Rightarrow \quad 2a = 1 \quad \Rightarrow \quad a = \frac{1}{2}. \quad \ldots (*)$$

(ii) $P(X \le 1.5) = \int_{-\infty}^{1.5} f(x)\,dx = \int_{-\infty}^{0} f(x)\,dx + \int_0^1 f(x)\,dx + \int_1^{1.5} f(x)\,dx$

$$= a\int_0^1 x\,dx + \int_1^{1.5} a\,dx$$

$$= a\left|\frac{x^2}{2}\right|_0^1 + a\left|x\right|_1^{1.5} = \frac{a}{2} + 0.5\,a = a = \frac{1}{2}. \qquad \text{[From (*) in Part (i)]}$$

Example 5·8. *Prove that the geometric mean G of the distribution :*

$$dF = 6(2 - x)(x - 1)\,dx, \ 1 \le x \le 2$$

is given by 6 log (16G) = 19.

Solution. By def., we have

$$\log G = \int_1^2 \log x\, f(x)\,dx = 6\int_1^2 (2 - x)(x - 1)\log x\,dx = -6\int_1^2 (x^2 - 3x + 2)\log x\,dx$$

Integrating by parts, we get

$$\log G = -6\left\{\left|\left(\frac{x^3}{3} - \frac{3x^2}{2} + 2x\right)\log x\right|_1^2 - \int_1^2 \left(\frac{x^3}{3} - \frac{3x^3}{2} + 2x\right)\frac{1}{x}\,dx\right\}$$

$$= -4\log 2 + 6 \times \frac{19}{36} \qquad \text{(on simplification)}$$

$\therefore \quad \log G + 4\log 2 = \frac{19}{6} \quad \Rightarrow \quad \log(G \times 2^4) = \frac{19}{6} \quad \Rightarrow \quad \log(16\,G) = \frac{19}{6}.$

Hence $\quad 6\log\,(16\,G) = 19.$

Example 5·9. *The probability distribution of a r.v. X is : $f(x) = k\sin\frac{1}{5}\pi x,\ 0 \le x \le 5.$*

Determine the constant k and obtain the median and quartiles of the distribution.

Solution. Since $k\int_0^5 \sin\frac{1}{5}\pi x\,dx = 1 \Rightarrow 5k\int_0^\pi \sin y\,dy = 1,$ where $\frac{1}{5}\pi x = y$

$$\therefore \quad k = \frac{\pi}{10}$$

The median M, is given by :

$$k\int_0^M \sin\frac{1}{5}\pi x\,dx = \frac{1}{2} \quad \Rightarrow \quad \frac{5}{\pi}k\int_0^{\frac{1}{5}\pi M} \sin y\,dy = \frac{1}{2}, \quad \text{where } \frac{1}{5}\pi x = y$$

$$\Rightarrow \quad \frac{5k}{\pi}\left|1 - \cos y\right|_0^{\frac{1}{5}\pi M} = \frac{1}{2} \quad \Rightarrow \quad \frac{1}{2}\left(1 - \cos\frac{1}{5}\pi M\right) = \frac{1}{2} \quad \Rightarrow \quad M = \frac{5}{2} \text{ [From (*)]}$$

The first quartile Q_1 is given by :

$$k\int_0^{Q_1} \sin\left(\frac{1}{5}\pi x\right)dx = \frac{1}{4} \quad \Rightarrow \quad \frac{5k}{\pi}\int_0^{\frac{1}{5}\pi Q_1} \sin y\,dy = \frac{1}{4} \quad \Rightarrow \quad \frac{1}{2}\left(1 - \cos\frac{1}{5}\pi Q_1\right) = \frac{1}{4}$$

[From (*)]

$$\therefore \quad \cos\left(\frac{\pi Q_1}{5}\right) = \frac{1}{2} = \cos\left(\frac{\pi}{3}\right) \quad \Rightarrow \quad Q_1 = \frac{\pi}{3} \times \frac{5}{\pi} = \frac{5}{3}.$$

The third quartile Q_3 is given by :

$$k\int_0^{Q_3} \sin\left(\frac{1}{5}\pi x\right)dx = \frac{3}{4} \quad \Rightarrow \quad \frac{5k}{\pi}\int_0^{\frac{1}{5}\pi Q_3} \sin y\,dy = \frac{3}{4} \quad \Rightarrow \quad \frac{1}{2}\left(1 - \cos\frac{1}{5}\pi Q_3\right) = \frac{3}{4}$$

$$\therefore \quad \cos\left(\frac{\pi Q_3}{5}\right) = -\frac{1}{2} = \cos\left(\pi - \frac{\pi}{3}\right) = \cos\left(\frac{2\pi}{3}\right) \quad \Rightarrow \quad Q_3 = \frac{2\pi}{3} \times \frac{5}{\pi} = \frac{10}{3}.$$

Example 5·10. *(a) A random variable X is distributed at random between the values 0 and 1 so that its probability density function is : $f(x) = kx^2(1 - x^3)$, where k is a constant. Find the value of k. Using this value of k, find its mean and variance .*

(b) A variable X is distributed at random between the values 0 and 4 and its probability density function is given by : $f(x) = kx^3(4 - x)^2$.

Find the value of k, the mean and standard deviation of the distribution.

Solution. (*a*) Since $\quad \int_{-\infty}^{\infty} f(x)\,dx = 1,$

$$k\int_0^1 (x^2 - x^5)\,dx = 1 \quad \Rightarrow \quad k\left|\frac{x^3}{3} - \frac{x^6}{6}\right|_0^1 = 1 \quad \Rightarrow \quad k = 6.$$

$$\text{Mean} = \mu_1' = \int_{-\infty}^{\infty} x f(x)\, dx = 6\int_0^1 (x^3 - x^6)\, dx = 6\left| \frac{x^4}{4} - \frac{x^7}{7} \right|_0^1 = \frac{9}{14}$$

$$\mu_2' = \int_{-\infty}^{\infty} x^2 f(x)\, dx = 6\int_0^1 (x^4 - x^7)\, dx = 6\left| \frac{x^5}{5} - \frac{x^8}{8} \right|_0^1 = \frac{9}{20}$$

$$\text{Variance} = \mu_2 = \mu_2' - \mu_1'^2 = \left\{ \frac{9}{20} - \left(\frac{9}{14}\right)^2 \right\} = \frac{9}{245}.$$

(*b*) Since $\int_{-\infty}^{\infty} f(x)\, dx = 1,\ k\int_0^4 x^3 (4-x)^2\, dx = 1 \quad \Rightarrow \quad k = \frac{15}{1024}.$

$$\text{Mean} = \mu_1' = \int_{-\infty}^{\infty} x f(x)\, dx = k\int_0^4 x^4 (4-x)^2\, dx = \frac{16}{7}$$

$$\mu_2' = \int_{-\infty}^{\infty} x^2 f(x)\, dx = k\int_0^4 x^5 (4-x)^2\, dx = \frac{40}{7}$$

$$\text{Variance} = \mu_2 = \mu_2' - \mu_1'^2 = \left\{ \frac{40}{7} - \left(\frac{16}{7}\right)^2 \right\} = \frac{24}{49}$$

$\therefore$ Standard deviation $= \frac{2\sqrt{6}}{7}.$

Example 5·11. *A random variable X has the probability law :*

$$dF(x) = \frac{x}{b^2} \cdot e^{-x^2/2b^2}\, dx,\ 0 \le x < \infty$$

Find the distance between the quartiles and show that the ratio of this distance to the standard deviation of X is independent of the parameter 'b'.

Solution. If Q_1 and Q_3 are the first and third quartiles respectively, we have

$$\int_0^{Q_1} f(x)\, dx = \frac{1}{4} \quad \Rightarrow \quad \frac{1}{b^2}\int_0^{Q_1} x e^{-x^2/2b^2}\, dx = \frac{1}{4}. \qquad \text{Put } y = \frac{x^2}{2b^2} \text{ so that } dy = \frac{x}{b^2} dx.$$

$$\therefore \quad \int_0^{Q_1^2/2b^2} e^{-y}\, dy = \frac{1}{4} \quad \Rightarrow \quad \left| \frac{e^{-y}}{-1} \right|_0^{Q_1^2/2b^2} = \frac{1}{4} \Rightarrow 1 - e^{-Q_1^2/2b^2} = \frac{1}{4} \quad \Rightarrow \quad e^{-Q_1^2/2b^2} = \frac{3}{4}.$$

$$\text{Thus} \quad -\frac{Q_1^2}{2b^2} = \log\left(\frac{3}{4}\right) \quad \Rightarrow \quad \frac{Q_1^2}{2b^2} = \log\left(\frac{4}{3}\right) \quad \Rightarrow \quad Q_1 = b\sqrt{2}\,\sqrt{\log(4/3)}$$

Again we have $\int_0^{Q_3} f(x)\, dx = \frac{3}{4}$ which, on proceeding similarly, will give

$$1 - e^{-Q_3^2/2b^2} = \frac{3}{4} \quad \Rightarrow \quad e^{-Q_3^2/2b^2} = \frac{1}{4} \quad \Rightarrow \quad Q_3 = b\sqrt{2}\,\sqrt{\log 4}$$

The distance between the quartiles is : $Q_3 - Q_1 = b\sqrt{2}\left\{\sqrt{\log 4} - \sqrt{\log (4/3)}\right\}$

$$\mu_1' = \int_0^{\infty} x f(x)\, dx = \int_0^{\infty} x \frac{x}{b^2} e^{-x^2/2b^2}\, dx = \int_0^{\infty} b\sqrt{2}\, y^{1/2} e^{-y} dy$$

$$= b\sqrt{2}\int_0^{\infty} e^{-y} y^{\frac{3}{2}-1}\, dy = b\sqrt{2}\,\Gamma(3/2) = b\sqrt{2}\,\frac{1}{2}\Gamma(1/2) = b\sqrt{2}\,\frac{\sqrt{\pi}}{2} = b\sqrt{\pi/2}$$

$$\mu_2' = \int_0^{\infty} x^2 f(x)\, dx = \int_0^{\infty} x^2 \frac{x}{b^2} e^{-x^2/2b^2} dx$$

$$= 2b^2 \int_0^\infty y\, e^{-y}\, dy, \; (y = x^2/2b^2) = 2b^2\, \Gamma(2) = 2b^2.\, 1\,! = 2b^2$$

$$\therefore \quad \sigma^2 = \mu_2 = \mu_2' - \mu_1'^2 = 2b^2 - b^2 \cdot \frac{\pi}{2} = b^2\left(2 - \frac{\pi}{2}\right) \quad \Rightarrow \quad \sigma = b\sqrt{\{2 - (\pi/2)\}}$$

Hence $\dfrac{Q_3 - Q_1}{\sigma} = \dfrac{\sqrt{2}\,[\sqrt{\log 4} - \sqrt{\log (4/3)}]}{\sqrt{\{2-(\pi/2)\}}}$, which is independent of parameter 'b'

Example 5·12. *In a continuous distribution whose relative frequency density is given by :*

$$f(x) = y_0 .\, x\,(2 - x),\; 0 \le x \le 2,$$

find mean, variance, β_1, and β_2 and hence show that the distribution is symmetrical. Also (i) find mean deviation about mean, and (ii) show that for this distribution $\mu_{2n+1} = 0$, and (iii) find the mode, harmonic mean and median.

Solution. Since total probability is unity, we have

$$\int_0^2 f(x)\, dx = 1 \quad \Rightarrow \quad y_0 \int_0^2 x\,(2-x)\, dx = 1 \quad \Rightarrow \quad y_0 = \frac{3}{4}$$

$$\therefore \qquad f(x) = \frac{3}{4}\, x\,(2 - x),\; 0 \le x \le 2$$

$$\mu_r' = \int_0^2 x^r f(x) dx = \frac{3}{4}\int_0^2 x^{r+1}\,(2-x)\, dx = \frac{3.2^{r+1}}{(r+2)\,(r+3)}$$

In particular

$$\text{Mean} = \mu_1' = \frac{3.2^2}{3.4} = 1, \qquad \mu_2' = \frac{3.2^3}{4.5} = \frac{6}{5}, \qquad \mu_3' = \frac{3.2^4}{5.6} = \frac{8}{5}, \qquad \text{and} \qquad \mu_4' = \frac{3.2^5}{6.7} = \frac{16}{7}$$

Hence, $\qquad \text{variance} = \mu_2 = \mu_2' - \mu_1'^2 = \dfrac{6}{5} - 1 = \dfrac{1}{5}$

$$\mu_3 = \mu_3' - 3\mu_2'\,\mu_1' + 2\mu_1'^3 = \frac{8}{5} - 3.\,\frac{6}{5}.\,1 + 2 = 0$$

$$\mu_4 = \mu_4' - 4\,\mu_3'\,\mu_1' + 6\,\mu_2'\,\mu_1'^2 - 3\mu_1'^4 = \frac{16}{7} - 4\,.\,\frac{8}{5}.\,1 + 6\,.\,\frac{6}{5}\,.\,1 - 3.\,1 = \frac{3}{35}$$

$$\therefore \qquad \beta_1 = \frac{\mu_3^2}{\mu_2^3} = 0 \qquad \text{and} \qquad \beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{3/35}{(1/5)^2} = \frac{15}{7} \qquad \dots (*)$$

Since $\beta_1 = 0$, the distribution is *symmetrical*.

Mean deviation about mean

$$= \int_0^2 |x - 1|\, f(x)\, dx = \int_0^1 |x-1|\, f(x)\, dx + \int_1^2 |x - 1|\, f(x)\, dx$$

$$= \frac{3}{4}\left[\int_0^1 (1-x)\, x\,(2-x)\, dx + \int_1^2 (x-1)\, x\,(2-x)\, dx\right]$$

$$= \frac{3}{4}\left[\int_0^1 (2x - 3x^2 + x^3)\, dx + \int_1^2 (3x^2 - x^3 - 2x)\, dx\right]$$

$$= \frac{3}{4}\left[\left| x^2 - \frac{3\,.\,x^3}{3} + \frac{x^4}{4}\right|_0^1 + \left| 3.\,\frac{x^3}{3} - \frac{x^4}{4} - \frac{2x^2}{2}\right|_1^2\right] = \frac{3}{8}$$

$$\mu_{2n+1} = \int_0^2 (x - \text{mean})^{2n+1} f(x)\, dx = \frac{3}{4}\int_0^2 (x-1)^{2n+1} x(2-x)\, dx$$

$$= \frac{3}{4}\int_{-1}^{1} t^{2n+1}(t+1)(1-t)\, dt = \frac{3}{4}\int_{-1}^{1} t^{2n+1}(1-t^2)\, dt, \text{ where } t = x - 1$$

Since t^{2n+1} is an odd function of t and $(1 - t^2)$ is an even function of t, the integrand $t^{2n+1}(1-t^2)$ is an odd function of t.

Hence $\mu_{2n+1} = 0.$

Now $f'(x) = \frac{3}{4}(2 - 2x) = 0 \Rightarrow x = 1$ and $f''(x) = \frac{3}{4}(-2) = -\frac{3}{2} < 0.$

Hence mode = 1

Harmonic mean H is given by :

$$\frac{1}{H} = \int_0^2 \frac{1}{x} f(x)\, dx = \frac{3}{4}\int_0^2 (2-x)\, dx = \frac{3}{2} \quad \Rightarrow \quad H = \frac{2}{3}.$$

If M is the median, $\int_0^M f(x)\, dx = \frac{1}{2} \Rightarrow \frac{3}{4}\int_0^M x(2-x)\, dx = \frac{1}{2} \Rightarrow \left| x^2 - \frac{x^3}{3} \right|_0^M = \frac{2}{3}$

$$\Rightarrow \quad 3M^2 - M^3 = 2 \quad \Rightarrow \quad M^3 - 3M^2 + 2 = 0 \quad \Rightarrow \quad (M-1)(M^2 - 2M - 2) = 0$$

$$\therefore \quad M = 1 \quad \text{or} \quad M = \frac{2 \pm \sqrt{4+8}}{2} = 1 \pm \sqrt{3}$$

The only value of M lying in [0, 2] is $M = 1$. Hence median is 1.

Aliter. Since, the distribution is symmetrical, [From (*)] Mode = Median = Mean = 1

Example 5·13. *Calculate the standard deviation and mean deviation from mean if the frequency function f(x) has the form :*

$$f(x) = \begin{cases} \dfrac{3+2x}{18}, & \text{for } 2 \le x \le 4 \\ 0 & \text{otherwise} \end{cases}$$

Solution. We, have

$$\text{Mean} = \mu_1' = \int_{-\infty}^{\infty} x f(x)\, dx = \int_2^4 x\left(\frac{3+2x}{18}\right) dx = \frac{83}{27}$$

$$\mu_2' = \int_{-\infty}^{\infty} x^2 f(x)\, dx = \int_2^4 x^2\left(\frac{3+2x}{18}\right) dx = \frac{88}{9}.$$

$$\therefore \quad \text{Variance} = \mu_2' - \mu_1'^2 = \left\{ \frac{88}{9} - \left(\frac{83}{27}\right)^2 \right\} = \frac{239}{729} \quad \Rightarrow \quad \text{S.D.} = \sqrt{\frac{239}{729}} = 0{\cdot}57$$

$$\text{Mean Deviation} = \int_{-\infty}^{\infty} |x - \bar{x}|\, f(x)\, dx = \int_2^4 \left| x - \frac{83}{27} \right| \left(\frac{3+2x}{18}\right) dx$$

$$= \int_2^{83/27} \left(\frac{83}{27} - x\right)\left(\frac{3+2x}{18}\right) dx + \int_{83/27}^{4} \left(x - \frac{83}{27}\right)\left(\frac{3+2x}{18}\right) dx = 0{\cdot}49 \quad \text{(on simplification)}$$

Hence mean deviation and standard deviation are 0·49 and 0·57 respectively.

Hence $y_0 = b = \frac{1}{\sigma}$ and $a = m - \frac{1}{b} = m - \sigma.$

Also $\mu_3 = \mu_3' - 3\mu_2'\mu_1' + 2\mu_1'^3 = \frac{1}{b^3}(6 - 3.2 + 2) = \frac{2}{b^3} = 2\sigma^3$

and $\mu_4 = \mu_4' - 4\mu_3'\mu_1' + 6\mu_2'\mu_1'^2 - 3\mu_1'^4 = \frac{1}{b^4}(24 - 4.6.1 + 6.2.1 - 3) = \frac{9}{b^4} = 9\sigma^4$

Hence $$\beta_1 = \frac{\mu_3^2}{\mu_2^3} = \frac{4\sigma^6}{\sigma^6} = 4 \quad \text{and} \quad \beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{9\sigma^4}{\sigma^4} = 9.$$

Example 5·18. *Show that for the distribution :*

$$dF = y_0\left(1 - \frac{x^2}{a^2}\right)^{-p} dx \,;\, -a \le x \le a, 0 < p < 1.$$

$$\mu_r = \frac{(r-1)\,a^2}{r+1-2p}\mu_{r-2}.$$ *Express 'a' and 'p' in terms of σ and β_2.*

Solution. Mean $= \mu_1' = y_0\int_{-a}^{a} x\left(1 - \frac{x^2}{a^2}\right)^{-p} dx = 0$ ($\because$ Integrand is an odd function of x.)

$$\therefore \quad \mu_r = \mu_r' = y_0\int_{-a}^{a} x^r\left(1 - \frac{x^2}{a^2}\right)^{-p} dx$$

$$= y_0\,a^2\int_{-a}^{a} x^{r-2}\left(\frac{x^2}{a^2} - 1 + 1\right)\left(1 - \frac{x^2}{a^2}\right)^{-p} dx$$

$$= a^2 y_0\int_{-a}^{a} x^{r-2}\left(1 - \frac{x^2}{a^2}\right)^{-p} dx - a^2 y_0\int_{-a}^{a} x^{r-2}\left(1 - \frac{x^2}{a^2}\right)^{1-p} dx$$

$$= a^2\mu_{r-2} - a^2 y_0\left\{\left|\left(1 - \frac{x^2}{a^2}\right)^{1-p}\frac{x^{r-1}}{r-1}\right|_{-a}^{a} + \frac{2\,(1-p)}{a^2}\int_{-a}^{a}\frac{x^r}{r-1}\left(1 - \frac{x^2}{a^2}\right)^{-p} dx\right\}$$

$$= a^2\mu_{r-2} - \frac{2\,(1-p)}{r-1}\mu_r$$

$$\Rightarrow \quad \mu_r + \frac{2\,(1-p)}{r-1}\mu_r = a^2\mu_{r-2} \quad \Rightarrow \quad \mu_r = \frac{(r-1)\,a^2}{r+1-2p}\mu_{r-2}.$$

Putting $r = 2$ and 4 in (*), we get respectively :

$$\mu_2 = \sigma^2 = \frac{a^2}{3-2p}\mu_0 = \frac{a^2}{3-2p} \quad \Rightarrow \quad a^2 = (3-2p)\,\sigma^2 \qquad \dots (**)$$

$$\mu_4 = \frac{3a^2}{5-2p}\mu_2 = \frac{3a^4}{(5-2p)\,(3-2p)} \qquad \text{[From (**)]} \qquad \dots (***)$$

$$\therefore \quad \beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{3a^4}{(5-2p)\,(3-2p)}\cdot\frac{(3-2p)^2}{a^4} = \frac{3\,(3-2p)}{5-2p} \qquad \text{[From (**) and (***)]}$$

$$\Rightarrow \quad (5-2p)\,\beta_2 = 9 - 6p \quad \Rightarrow \quad 2p\,(\beta_2 - 3) = 5\beta_2 - 9 \quad \Rightarrow \quad p = \frac{5\beta_2 - 9}{2\,(\beta_2 - 3)}.$$

Example 5·19. *Suppose that the life in hours of a certain part of radio tube is a continuous random variable X with p.d.f. given by :*

$$f(x) = \begin{cases} \frac{100}{x^2}, & \text{when } x \ge 100 \\ 0, & \text{elsewhere} \end{cases} \qquad \dots (*)$$

(i) *What is the probability that all of three such tubes in a given radio set will have to be replaced during the first 150 hours of operation ?*

(*ii*) *What is the probability that none of three of the original tubes will have to be replaced during that first 150 hours of operation ?*

(*iii*) *What is the probability that a tube will last less than 200 hours if it is known that the tube is still functioning after 150 hours of service ?*

(*iv*) *What is the maximum number of tubes that may be inserted into a set so that there is a probability of 0·5 that after 150 hours of service all of them are still functioning ?*

Solution. (*i*) A tube in a radio set will have to be replaced during the first 150 hours, if its life is ≤ 150 hours. Hence, the required probability '*p*' that a tube is replaced during the first 150 hours, on using (*) is given by :

$$p = P(X \le 150) = \int_{100}^{150} f(x)\, dx = \int_{100}^{150} \frac{100}{x^2}\, dx = \frac{1}{3}$$

∴ By compound probability theorem, the probability that all three of the original tubes will have to be replaced during the first 150 hours $= p^3 = \left(\frac{1}{3}\right)^3 = \frac{1}{27}\cdot$

(*ii*) The proability that a tube is not replaced during the first 150 hours of operation is given by :

$$P(X > 150) = 1 - P(X \le 150) = 1 - p = 1 - \frac{1}{3} = \frac{2}{3}\cdot$$

Hence by compound probability theorem, the probability that none of the three tubes will be replaced during the first 150 hours of operation $= \left(\frac{2}{3}\right)^3 = \frac{8}{27}\cdot$

(*iii*) Probability that a tube will last less than 200 hours given that the tube is still functioning after 150 hours is :

$$P(X < 200 \mid X > 150) = \frac{P(150 < X < 200)}{P(X > 150)} = \frac{\int_{150}^{200} \frac{100}{x^2}\, dx}{\int_{150}^{\infty} \frac{100}{x^2}\, dx} = \frac{1}{6} \times \frac{3}{2} = 0{\cdot}25.$$

(*iv*) Let *n* be the maximum number of tubes that may be inserted into a set so that there is a probability of 0·5 that after 150 hours of service all of them are still functioning . Then we have to find *n* so that

$$0{\cdot}5 = \left(\frac{2}{3}\right)^n \Rightarrow n \log(0{\cdot}6667) = \log 0{\cdot}5 \Rightarrow n = \frac{\log 0{\cdot}5}{\log 0{\cdot}6667} = \frac{-0{\cdot}3010}{-0{\cdot}1760} = 1{\cdot}7.$$

Hence, the maximum number of tubes is 2.

Example 5·20. *The time one has to wait for a bus at a downtown bus stop is observed to be random phenomenon (X) with the following probability density function :*

$$f_X(x) = \begin{cases} 0, & \text{for } x < 0 \\ \frac{1}{9}(x+1), & \text{for } 0 \le x < 1 \\ \frac{4}{9}\left(x - \frac{1}{2}\right), & \text{for } 1 \le x < \frac{3}{2} \\ \frac{4}{9}\left(\frac{5}{2} - x\right), & \text{for } \frac{3}{2} \le x < 2 \\ \frac{1}{9}(4 - x), & \text{for } 2 \le x < 3 \\ \frac{1}{9}, & \text{for } 3 \le x < 6 \\ 0, & \text{for } x \ge 6 \end{cases}$$

Let the events A and B be defined as follows :

A : One waits between 0 to 2 minutes inclusive.

B : One waits between 0 to 3 minutes inclusive.

(i) Draw the graph of probability density function.

(ii) Show that (a) $P(B \mid A) = \frac{2}{3}$, (b) $P(\bar{A} \cap \bar{B}) = \frac{1}{3}$.

Solution. (i) The graph of *p.d.f.* is given below :

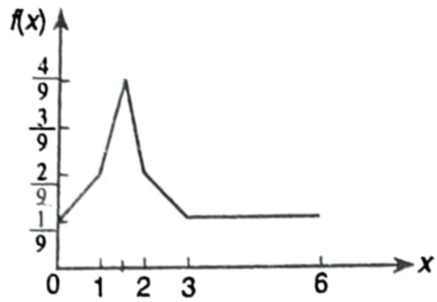


Fig. 5·8.

(ii) (a) $P(A) = P(X \le 2) = \int_0^2 f(x)\,dx$

$$= \int_0^1 \frac{1}{9}(x+1)\,dx + \int_1^{3/2} \frac{4}{9}\left(x - \frac{1}{2}\right) dx + \int_{3/2}^2 \frac{4}{9}\left(\frac{5}{2} - x\right) dx$$

$= \frac{1}{2}$· (on simplification).

$$P(A \cap B) = P[(0 \le X \le 2) \cap (1 \le X \le 3)] = P(1 \le X \le 2) = \int_1^2 f(x)\,dx$$

$$= \int_1^{3/2} \frac{4}{9}\left(x - \frac{1}{2}\right) dx + \int_{3/2}^2 \frac{4}{9}\left(\frac{5}{2} - x\right) dx$$

$$= \frac{4}{9}\left|\frac{x^2}{2} - \frac{x}{2}\right|_1^{3/2} + \frac{4}{9}\left|\frac{5}{2}x - \frac{x^2}{2}\right|_{3/2}^2 = \frac{1}{3} \quad \text{(on simplification)}$$

$\therefore$ $$P(B \mid A) = \frac{P(A \cap B)}{P(A)} = \frac{1/3}{1/2} = \frac{2}{3}.$$

(b) $\bar{A} \cap \bar{B}$ means that waiting time is more than 3 minutes.

$\therefore$ $$P(\bar{A} \cap \bar{B}) = P(X > 3) = \int_3^\infty f(x)\,dx = \int_3^6 f(x)\,dx + \int_6^\infty f(x)\,dx$$

$$= \int_3^6 \frac{1}{9}\,dx = \frac{1}{9}\left|x\right|_3^6 = \frac{1}{3}.$$

Example 5·21. *The amount of bread (in hundreds of pounds) x that a certain bakery is able to sell in a day is found to be a numerical valued random phenomenon, with a probability function specified by the p.d.f. f(x), given by :*

$$f(x) = \begin{cases} k.\,x & , \text{ for } 0 \le x < 5 \\ k.\,(10 - x) & , \text{ for } 5 \le x < 10 \\ 0 & , \text{ otherwise} \end{cases}$$

(a) *Find the value of k such that f(x) is a probability density function.*

(b) What is the probability that the number of pounds of bread that will be sold tomorrow is : (i) more than 500 pounds, (ii) less than 500 pounds, and (iii) between 250 and 750 pounds ?

(c) Denoting by A, B and C the events that the pounds of bread sold are as in b (i), b (ii) and b (iii) respectively, find P (A | B), P (A | C). Are (i) A and B independent events ? (ii) Are A and C independent events ?

Solution. (*a*) In order that $f(x)$ should be a probability density function :

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

$$\Rightarrow \quad \int_0^5 kx\,dx + \int_5^{10} k(10-x)\,dx = 1 \quad \text{or} \quad k = \frac{1}{25}\cdot \text{(on simplification)}$$

(*b*) (*i*) The probability that the number of pounds of bread that will be sold tomorrow is more than 500 pounds is given by :

$$P(5 \le X \le 10) = \int_5^{10} \frac{1}{25}(10-x)\,dx = \frac{1}{25}\left| 10x - \frac{x^2}{2} \right|_5^{10} = \frac{1}{25}\left(\frac{25}{2}\right) = 0{\cdot}5$$

(*ii*) The probability that the number of pounds of bread that will be sold tomorrow is less than 500 pounds, is given by :

$$P(0 \le X < 5) = \int_0^5 \frac{1}{25}\cdot x\,dx = \frac{1}{25}\left|\frac{x^2}{2}\right|_0^5 = \frac{1}{2} = 0{\cdot}5$$

OR $\quad P(X < 5) = 1 - P(X \ge 5) = 1 - 0{\cdot}5 = 0{\cdot}5$ [From part (*i*)]

(*iii*) The required probability is given by :

$$P(2{\cdot}5 \le X \le 7{\cdot}5) = \int_{2\cdot5}^{5} \frac{1}{25}x\,dx + \int_5^{7\cdot5} \frac{1}{25}(10-x)\,dx = \frac{3}{4} \quad \text{(On simplification)}$$

(*c*) The events A, B and C are given by

$$A : 5 < X \le 10;\quad B : 0 \le X \le 5;\quad C : 2{\cdot}5 \le X \le 7{\cdot}5.$$

Then from parts *b* (*i*), (*ii*) and (*iii*), $P(A) = 0{\cdot}5$, $P(B) = 0{\cdot}5$, $P(C) = \frac{3}{4}$.

The events $A \cap B$ and $A \cap C$ are given by : $A \cap B = \phi$ and $A \cap C : 5 < X < 7{\cdot}5$.

$\therefore \quad P(A \cap B) = P(\phi) = 0$

and $\quad P(A \cap C) = \int_5^{7\cdot5} f(x)\,dx = \frac{1}{25}\int_5^{7\cdot5} (10-x)\,dx = \frac{1}{25}\times\frac{75}{8} = \frac{3}{8}$

$P(A).\,P(C) = \frac{1}{2}\times\frac{3}{4} = \frac{3}{8} = P(A \cap C) \Rightarrow$ A and C are independent.

Again $\quad P(A).\,P(B) = \frac{1}{4} \ne P(A \cap B) \quad \Rightarrow$ A and B are not independent.

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = 0 \quad \text{and} \quad P(A \mid C) = \frac{P(A \cap C)}{P(C)} = \frac{3/8}{3/4} = \frac{1}{2}.$$

Example 5·22. *The kms X in thousands of kms which car owners get with a certain kind of tyre is a random variable having probability density function :*

$$f(x) = \begin{cases} \frac{1}{20}e^{-x/20}, & \text{for } x > 0 \\ 0, & \text{for } x \le 0 \end{cases}$$

Find the probabilities that one of these tyres will last (i) at most 10,000 kms.

(ii) anywhere from 16,000 to 24,000 kms. and (iii) at least 30,000 miles.

Solution. Let *r.v.* X denote the kms. (in '000 kms.)with a certain kind of tyre. Then required probabilities are given by :

(i) $$P(X \le 10) = \int_0^{10} f(x)\,dx = \frac{1}{20}\int_0^{10} e^{-x/20}\,dx$$

$$= \frac{1}{20}\left|\frac{e^{-x/20}}{-1/20}\right|_0^{10} = 1 - e^{-1/2} = 1 - 0.6065 = 0.3935$$

(ii) $$P(16 \le X \le 24) = \frac{1}{20}\int_{16}^{24} \exp\left(-\frac{x}{20}\right)dx = \left| -e^{-x/20} \right|_{16}^{24}$$

$$= e^{-16/20} - e^{-24/20} = e^{-4/5} - e^{-6/5} = 0.4493 - 0.3012 = 0.1481.$$

(iii) $$P(X \ge 30) = \int_{30}^{\infty} f(x)\,dx = \frac{1}{20}\left|\frac{e^{-x/20}}{-1/20}\right|_{30}^{\infty} = e^{-1.5} = 0.2231.$$

5·4·3. Continuous Distribution Function

Definition. *If X is a continuous random variable with the p.d.f. f(x), then the function*

:

$$F_X(x) = P(X \le x) = \int_{-\infty}^{x} f(t)\,dt, \quad -\infty < x < \infty \qquad \ldots (5.12)$$

is called the distribution function (d.f.) or sometimes the cumulative distribution function (c.d.f.) of the random variable X.

Properties of Distribution Function. **1.** $0 \le F(x) \le 1, -\infty < x < \infty.$

2. From analysis (Riemann integral), we know that

$$F'(x) = \frac{d}{dx}F(x) = f(x) \ge 0 \qquad [\because f(x) \text{ is } p.d.f.]$$

$\Rightarrow$ $F(x)$ is non-decreasing function of x.

3. $$F(-\infty) = \lim_{x \to -\infty} F(x) = \lim_{x \to -\infty} \int_{-\infty}^{x} f(x)\,dx = \int_{-\infty}^{-\infty} f(x)\,dx = 0$$

and $$F(+\infty) = \lim_{x \to \infty} F(x) = \lim_{x \to \infty} \int_{-\infty}^{x} f(x)\,dx = \int_{-\infty}^{\infty} f(x)\,dx = 1$$

$\Rightarrow$ $0 \le F(x) \le 1.$

4. $F(x)$ is a continuous function of x on the right.

5. The discontinuities of $F(x)$ are at the most countable.

6. It may be noted that

$$P(a \le X \le b) = \int_a^b f(x)\,dx = \int_{-\infty}^{b} f(x)\,dx - \int_{-\infty}^{a} f(x)\,dx$$

$$= P(X \le b) - P(X \le a) = F(b) - F(a) \qquad \ldots (5.12a)$$

Similarly,

$$P(a < X < b) = P(a < X \le b) = P(a \le X < b) = \int_a^b f(t)\,dt \qquad \ldots (5.12b)$$

7. $$F'(x) = \frac{d}{dx}F(x) = f(x) \quad \Rightarrow \quad dF(x) = f(x)dx$$

$dF(x)$ is known as probability differential of X.

Remarks 1. It may be pointed out that the properties (2), (3) and (4) above uniquely characterise the distribution functions. This means that any function $F(x)$ satisfying (2) to (4) is the distribution function of some random variable, and any function $F(x)$ violating any one or more of these three properties cannot be the distribution function of any random variable.

2. Often, one can obtain a *p.d.f.* from a distribution function $F(x)$ by differentiating $F(x)$, provided the derivative exists, *e.g.*, consider :

$$F_X(x) = \begin{cases} 0, & \text{for } x < 0 \\ x, & \text{for } 0 \le x \le 1 \\ 1, & \text{for } x > 1 \end{cases}$$

The graph of $F(x)$ is given by bold lines. Obviously we see that $F(x)$ is continuous from right as stipulated in (4) and we also see that $F(x)$ is not continuous at $x = 0$ and $x = 1$ and hence is not derivable at $x = 0$ and $x = 1$.

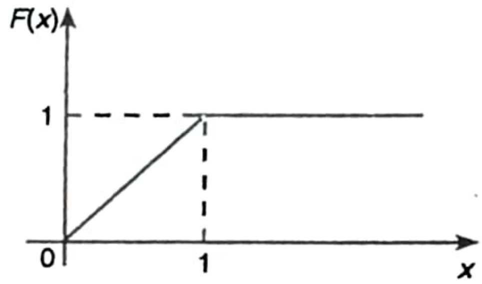


Fig. 5·9

Differentiating $F(x)$ *w.r.* to x, we get $\frac{d}{dx}F(x) = \begin{cases} 1, & 0 < x < 1 \\ 0, & \text{otherwise} \end{cases}$

(Note the strict inequality in $0 < x < 1$, since $F(x)$ is not derivable at $x = 0$ and $x = 1$.)

Let us define : $$f(x) = \begin{cases} 1, & 0 < x < 1 \\ 0, & \text{otherwise} \end{cases}$$

Then $f(x)$ is a *p.d.f.* for F.

Example 5·23. *Verify that the following is a distribution function :*

$$F(x) = \begin{cases} 0 & , x < -a \\ \frac{1}{2}\left(\frac{x}{a} + 1\right) & , -a \le x \le a \\ 1 & , x > a \end{cases}$$

Solution. Obviously the properties (*i*), (*ii*), (*iii*) and (*iv*) are satisfied. Also we observe that $F(x)$ is continuous at $x = a$ and $x = -a$ as well.

Now $$\frac{d}{dx}F(x) = \left\{\begin{matrix} \frac{1}{2a}, & -a \le x \le a \\ 0, & \text{otherwise} \end{matrix}\right\} = f(x), \text{ say}$$

In order that $F(x)$ is a distribution function, $f(x)$ must be a *p.d.f.* Thus we have to show that $\int_{-\infty}^{\infty} f(x)\, dx = 1.$

Now $$\int_{-\infty}^{\infty} f(x)\, dx = \int_{-a}^{a} f(x)\, dx = \frac{1}{2a}\int_{-a}^{a} 1.\, dx = 1.$$

Hence $F(x)$ is a distribution function.

Example 5·24. *The diameter, say X, of an electric cable, is assumed to be a continuous random variable with p.d.f. :* $f(x) = 6x(1-x), 0 \le x \le 1$

(i) *Check that the above is a p.d.f.,*

(ii) *Obtain an expression for the c.d.f. of X.,*

(iii) *Compute* $P\left(X \le \frac{1}{2} \,\middle|\, \frac{1}{3} \le X \le \frac{2}{3}\right)$, *and*

(iv) *Determine the number k such that* $P(X < k) = P(X > k)$.

Solution.

(i) Since $\int_0^1 f(x)\,dx = \int_0^1 6x(1-x)\,dx = 6\left|\frac{x^2}{2} - \frac{x^3}{3}\right|_0^1 = 1$, $f(x)$ is a p.d.f.

(ii) $$F(x) = \begin{cases} 0, & \text{if } x \le 0 \\ \int_0^x 6t(1-t)dt = (3x^2 - 2x^3), & 0 < x \le 1 \\ 1, & \text{if } x > 1 \end{cases}$$

(iii) $$P\left(X \le \tfrac{1}{2} \,\middle|\, \tfrac{1}{3} \le X \le \tfrac{2}{3}\right) = \frac{P(\frac{1}{3} \le X \le \frac{1}{2})}{P(\frac{1}{3} \le X \le \frac{2}{3})} = \frac{\int_{1/3}^{1/2} 6x(1-x)\,dx}{\int_{1/3}^{2/3} 6x(1-x)\,dx} = \frac{11/54}{13/27} = \frac{11}{26}.$$

(iv) We have $P(X < k) = P(X > k)$

$$\Rightarrow \quad \int_0^k 6x(1-x)\,dx = \int_k^1 6x(1-x)\,dx \quad \text{or} \quad 3k^2 - 2k^3 = 3(1-k^2) - 2(1-k^3)$$

$$\Rightarrow \quad 4k^3 - 6k^2 + 1 = 0 \quad \Rightarrow \quad k = \frac{1}{2}, \frac{1 \pm \sqrt{3}}{2}.$$

The only admissible value of k in the given range is $\frac{1}{2}$. Hence the value of k is $\frac{1}{2}$.

Example 5·25. *Let X be a continuous random variable with p.d.f. given by :*

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

(i) *Determine the constant k,* (ii) *Determine F(x), the c.d.f., and*

(iii) *If x_1, x_2 and x_3 are three independent observations from X, what is the probability that exactly one of these three numbers is larger than 1·5 ?*

Solution. (i) Since $f(x)$ in the p.d.f. of X, we have :

$$\int_0^3 f(x)\,dx = \int_0^1 f(x)\,dx + \int_1^2 f(x)\,dx + \int_2^3 f(x)\,dx = 1$$

$$\Rightarrow \quad \int_0^1 kx\,dx + \int_1^2 k\,dx + \int_2^3 (-kx + 3k) = 1$$

$$\therefore \quad \left[k\frac{x^2}{2}\right]_0^1 + \left[kx\right]_1^2 + \left[-k.\frac{x^2}{2} + 3kx\right]_2^3 = 1 \quad \Rightarrow \quad k = \frac{1}{2}.$$

(*ii*) For any x such that $-\infty < x < 0$; $F(x) = \int_{-\infty}^{x} 0.\ dt = 0$

For any x, where $0 \le x < 1$; $F(x) = \int_{-\infty}^{0} 0.\ dt + \int_{0}^{x} \frac{t}{2} dt = \frac{x^2}{4}$

For x, where $1 \le x < 2$,

$$F(x) = \int_{-\infty}^{0} 0.\ dt + \int_{0}^{1} \frac{t}{2} dt + \int_{1}^{x} \frac{1}{2} dt = \frac{2x-1}{4}$$

For any x, where $2 \le x < 3$,

$$F(x) = \int_{-\infty}^{0} 0.\ dt + \int_{0}^{1} \frac{t}{2} dt + \int_{1}^{2} \frac{1}{2} dt + \int_{2}^{x} \left(-\frac{t}{2} + \frac{3}{2}\right) dt$$

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

For any x, where $3 \le x < \infty$

$$F(x) = \int_{-\infty}^{0} 0.\ dt + \int_{0}^{1} \frac{t}{2}\ dt + \int_{1}^{2} \frac{1}{2} dt + \int_{2}^{3} \left(-\frac{t}{2} + \frac{3}{2}\right) dt + \int_{3}^{x} 0.\ dt$$

$$= \frac{1}{4} + \left(1 - \frac{1}{2}\right) + \left(-\frac{9}{4} + \frac{9}{2} + 1 - 3\right) = 1$$

Hence the distribution function $F(x)$ is give by :

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

(*iii*) The probability that X is larger than 1·5 is given by :

$$P(X > 1{\cdot}5) = 1 - P(X < 1{\cdot}5) = 1 - F(1{\cdot}5) = \left(1 - \frac{3-1}{4}\right) = \frac{1}{2}.$$

$\therefore$ The probability that X is not larger than $1{\cdot}5 = P(X < 1{\cdot}5) = 1 - \frac{1}{2} = \frac{1}{2}$.

Hence out of three numbers x_1, x_2 and x_3, the probability that exactly one is larger than 1·5 is :

$$3\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)\left(\frac{1}{2}\right) = \frac{3}{8}.$$

Example 5·26. *A petrol pump is supplied with petrol once a day. If its daily volume of sales (X) in thousands of litres is distributed by : $f(x) = 5(1-x)^4$, $0 \le x \le 1$, what must be the capacity of its tank in order that the probability that its supply will be exhausted in a given day shall be 0·01 ?*

Solution. Let the capacity of the tank (in '000 of litres) be 'a' such that

$$P(X \ge a) = 0{\cdot}01 \quad \Rightarrow \quad \int_{a}^{1} f(x)\, dx = 0{\cdot}01$$

$$\Rightarrow \qquad \int_a^1 5(1-x)^4\,dx = 0.01 \qquad \text{or} \qquad \left[5.\frac{(1-x)^5}{(-5)}\Bigg|_a^1\right] = 0.01$$

$$\Rightarrow \qquad (1-a)^5 = \frac{1}{100} \qquad \text{or} \qquad 1-a = \left(\frac{1}{100}\right)^{1/5}$$

$$\therefore \qquad a = 1-\left(\frac{1}{100}\right)^{1/5} = 1-0.3981 = 0.6019$$

Hence the capacity of the tank $= 0.6019 \times 1{,}000$ litres $= 601.9$ litres.

Example 5·27. *A bombing plane carrying three bombs flies directly above a railroad track. If a bomb falls within 40 metres of track, the track will be sufficiently damaged to disrupt the traffic. With a certain bomb site the points of impact of a bomb have the probability density function :*

$$f(x) = \begin{cases} \dfrac{100+x}{10{,}000}, & \text{when } -100 \le x < 0 \\[2mm] \dfrac{100-x}{10{,}000}, & \text{when } 0 \le x < 100 \\[2mm] 0, & \text{elsewhere} \end{cases}$$

where x represents the vertical deviation (in metres) from the aiming point, which is the track in this case. Find the distribution function. If all the three bombs are used, what is the probability that the track will be damaged ?

Solution. If $x < -100$, $F(x) = 0$

If $-100 \le x < 0$, $\quad F(x) = \displaystyle\int_{-100}^{x} \frac{100+t}{10{,}000}\,dt = \frac{1}{10^4}\left(100x + \frac{x^2}{2} + \frac{10^4}{2}\right)$

If $0 \le x < 100$,

$$F(x) = \int_{-\infty}^{-100} 0\,dt + \int_{-100}^{0} \frac{100+t}{10{,}000}\,dt + \int_0^x \frac{100-t}{10{,}000}\,dt$$

$$= \frac{1}{10{,}000}\left[100t + \frac{t^2}{2}\Bigg|_{-100}^{0}\right] + \frac{1}{10{,}000}\left[100t - \frac{t^2}{2}\Bigg|_0^x\right] = \frac{1}{10{,}000}\left(100x - \frac{x^2}{2} + \frac{10^4}{2}\right)$$

and if $x > 100$,

$$F(x) = \int_{-\infty}^{-100} 0\,dt + \int_{-100}^{0} \frac{100+t}{10{,}000}\,dt + \int_0^{100} \frac{100-t}{10{,}000}\,dt + \int_{100}^{x} 0\,dt = 1$$

Probability that the track will be damaged by a bomb is :

$$P(|X| < 40) = P(-40 < X < 40) = \int_{-40}^{0} \frac{100+x}{10{,}000}\,dx + \int_0^{40} \frac{100-x}{10{,}000}\,dx = \frac{16}{25}$$

$\therefore$ The probability that a bomb will not damage the track $= 1 - \dfrac{16}{25} = \dfrac{9}{25}$.

If three bombs are dropped, the required probability p that the track will be damaged is given by :

$$\begin{aligned} p &= P\,[\text{At least one of the three bombs damages the track.}] \\ &= 1 - P\,[\text{None of the three bombs damages the track.}] \\ &= 1 - \left(\frac{9}{25}\right)^3 = 1 - 0.0467 = 0.9533. \end{aligned}$$

Example 5·28. *Suppose that the time in minutes that a person has to wait at a certain bus stop for a bus is found to be a random phenomenon, with a probability function specified by the distribution function :*

$$F(x) = \begin{cases} 0 & , x < 0 \\ \frac{x}{8} & , 0 \le x < 2 \\ \frac{x^2}{16} & , 2 \le x < 4 \\ 1 & , 1 , x \ge 4 \end{cases} \qquad \dots (*)$$

(i) Is the distribution function continuous ? If so, give the formula for its probability density function.

(ii) What is the probability that a person will have to wait

(a) more than 2 minutes, (b) less than 2 minutes, and (c) between 1 and 2 minutes ?

(iii) What is the conditional probability that the person will have to wait for a bus for

(a) more than 2 minutes, given that it is more than 1 minute,

(b) less than 2 minutes given that it is more than 1 minute ?

Solution. (*i*) Since the value of the distribution function is the same at the points $x = 0, x = 1, x = 2$ and $x = 4$ given by the two forms of $F(x)$ for $x < 0$ and $0 \le x < 2$; $0 \le x < 2$ and $2 \le x < 4$; $2 \le x < 4$ and $x \ge 4$, the distribution function is continuous.

Probability density function $= f(x) = \frac{d}{dx} F(x)$

$$\therefore \qquad f(x) = \begin{cases} 0 , \text{ for } x < 0 \\ \frac{1}{8} , \text{ for } 0 \le x < 2 \\ \frac{x}{8} , \text{ for } 2 \le x < 4 \\ 0 , \text{ for } x \ge 4 \end{cases}$$

(*ii*) Let the random variable X represent the waiting time in minutes, then

(*a*) Required probability $= P(X > 2) = 1 - P(X \le 2) = 1 - F(2) = 1 - \frac{2^2}{16} = \frac{3}{4}$ · [From (*)]

(*b*) Required probability $= P(X < 2) = P(X \le 2) - P(X = 2) = F(2) = \frac{1}{4}$.

(Since, the probability that a continuous variable takes a fixed value is zero.)

(*c*) Required probability $= P(1 < X < 2) = F(2) - F(1) = \frac{1}{4} - \frac{1}{8} = \frac{1}{8}$. [From (*)]

(*iii*) Let A denote the event that a person has to wait for more than 2 minutes and B the event that he has to wait for more than 1 min. Then $P(A) = P(X > 2) = \frac{3}{4}$.

$$P(B) = P(X > 1) = 1 - P(X \le 1) = 1 - F(1) = 1 - \frac{1}{8} = \frac{7}{8}.$$

$$P(A \cap B) = P(X > 2 \cap X > 1) = P(X > 2) = \frac{3}{4}$$

(*a*) Required probability is : $P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{3/4}{7/8} = \frac{6}{7}$.

(*b*) Required probability $= P(\bar{A} \mid B) = \frac{P(\bar{A} \cap B)}{P(B)}$