CHAPTER 6
Random Variables and Probability Functions

6.1 INTRODUCTION

The chapter starts with a discussion on random variable, the first step towards better understanding of probability distribution. In subsequent section we use a trio (Ω, $\mathcal{A}$, P (•)) called probability space where Ω represents the sample space associated with a random experiment. $\mathcal{A}$ represents the collection of all possible subsets of $\mathcal{A}$ and P (•) is the probability that is associated with $\mathcal{A}$.

6.2 RANDOM VARIABLE

All the experiments that are performed can be classified into two broad divisions, viz., random experiments and deterministic experiments. Deterministic experiments are those experiments in which the outcome of the experiment remains same whenever it is performed indefinitely under constant conditions. But is case of a random experiment, the outcomes cannot be predicted with certainity if the experiment performed indefinitely under constant condition; although all its possible outcomes can be described completely. Here the experimenter may know the set of all possible outcomes of the random experiment (usually denoted by Ω), but cannot say with certainity which outcome will occur when, if the experiment is performed.

The sample space of random experiment is a pair (Ω, $\mathcal{A}$), where Ω is the set of all possible outcomes of the experiment; in other words, $\Omega = \{w\}$ is a non-empty set, whose elements w are interpreted as mutually exclusive outcomes of the random events; and $\mathcal{A}$ is a collection of subsets of the set Ω called the events (the set of $\mathcal{A}$ is assumed to contain Ω and to be closed with respect to an opposite event or a sum of events in not more than a countable number, i.e., $\mathcal{A}$ is a σ-algebra). The pair (Ω, $\mathcal{A}$) is the sample space of a random experiment. A set function P is defined on $\mathcal{A}$ is called the probability measure (or simply probability) if it satisfies

1. $P(A) \geq 0$, for all $A \in \mathcal{A}$
2. $P(\Omega) = 1$
3. Let $\{A_j\}$, $A_j \in \mathcal{A}, j = 1, 2, ...,$ be a disjoint sequence of sets, *i.e.*, $A_j \cap A_k = \phi$, for $j \neq k$. Then

$$P\left(\sum_{j=1}^{\infty} A_j\right) = \sum_{j=1}^{\infty} P(A_j).$$

The triple $(\Omega, \mathcal{A}, P)$ is called the probability space or sample space. The elements of Ω are called sample points and any set $A \in \mathcal{A}$ is called an event. Each one-point set is known as a simple or an elementary event and P(A) is called the probability of the event A.

Let $(\Omega, \mathcal{A})$ be a sample space. A finite, single-valued function which maps Ω into $\mathcal{R}$ is called a random variable (*rv*) if the inverse images under X of all Borel sets in $\mathcal{R}$ are events.

Suppose $x \in \mathcal{R}$, and consider the semi-closed interval $(-\infty, x)$. Since $(-\infty, x) \in \mathcal{B}$, it follows that if X is an *rv*, then $X^{-1}(-\infty, x) = \{X(w) \leq x\}$ is an event in $\mathcal{A}$. Thus X is a *rv* if and only if for each $x \in \mathcal{R}$

$$X^{-1}(-\infty, x) = \{X(w) \leq x\} \in \mathcal{A}$$

Now the *rv* is defined as follows :

Definition : A random variable is a real-valued function of a simple event $X = X(w)$, $w \in \Omega$ for which the set $\{w : X(w) \leq x\}$ belongs to $\mathcal{A}$ for any real x (*i.e.*, is an event) and for which the probability $P(w : X(w) \leq x)$ is defined.

The probability $P(X \leq x)$ is called a distribution function of the *rv* X.

Example 1. Let $\Omega = \{H, T\}$ and $\mathcal{A}$ be the class of all subsets of Ω. Define X by X(H) = 1, x(T) = 0. Then

$$X^{-1}(-\infty, x) = \begin{cases} \phi, & \text{if } x < 0; \\ \{T\}, & \text{if } 0 \leq x < 1; \\ \{H, T\}, & \text{if } 1 \leq x. \end{cases}$$

Thus X is an *rv*.

Example 2. Let Ω = {HHH, HHT, HTH, HTT, THH, THT, TTH, TTT} and $\mathcal{A}$ be the class of all subsets of Ω. Define X by $X(w)$ = number of H's in w. Then X(HHH) = 3, X(HHT, HTH, THH) = 2, X(TTH, THT, HTT) = 1, and X(TTT) = 0.

$$X^{-1}(-\infty, x) = \begin{cases} \phi, & \text{if } x < 0; \\ \{TTT\}, & \text{if } 0 \leq x < 1; \\ \{TTT, TTH, THT, HTT\}, & \text{if } 1 \leq x < 2; \\ \{TTT, HHT, HTH, THH\} & \text{if } 2 \leq x < 3; \\ \Omega & \text{if } 3 \leq x. \end{cases}$$

Thus X is an *rv*.

6.3 TYPES OF RANDOM VARIABLE

There are two types of random variables

(*a*) Discrete random variable (*b*) Continuous random variable

(*a*) ***Discrete random variable*** : If a random variable X assumes only finite number or countably infinite number of values, then it is called a discrete random variable. The random variable X is said to take finite values only if the possible values of X are $x_1, x_2, ..., x_n$ and is said to be countably infinite if X takes the values $x_1, x_2, ...$

Example : Number of fruits that fall from a tree every morning.

(*b*) ***Continuous random variable*** : A random variable assumes any value within a given interval, then it is called a continuous random variable. In other words, if a random variable can take infinite number of values within a given interval, $a \leq x \leq b$ (say), then it is called a continuous random variable.

Example : The heights of the persons collected from a crowd.

6.4 PROBABILITY DISTRIBUTION

The distribution obtained by taking the possible values of a random variable together with their respective probabilities is called a probability distribution. A probability distribution can be presented either with the help of a function or in tabular form where values of the random variable and corresponding probabilities are shown. The probability distribution for a discrete random variable is called discrete probability distribution and that of a continuous random variable is called continuous probability distribution.

6.5 DISCRETE PROBABILITY DISTRIBUTION

Let X be a discrete random variable which takes the value $x_1, x_2, ...$ with respective probabilities $p_1, p_2, ...$. Then the numbers $\{p_i\}$ satisfying

$$P\{X = x_i\} = p_i \geq 0,$$

for all i and $\sum_{i=1}^{\infty} p_i = 1$ is called the probability mass function (*pmf*) of the *rv* X.

The distribution function F of x is given by

$$F(x) = P\{X \leq x_i\}$$

$$= \sum_{x_i \leq x} \phi_i$$

Example : Let us consider the experiment of throwing a die. If X represents the face value which turns up when a die is thrown. Then X takes the values 1, 2, 3, 4, 5 and 6 each having probability 1/6. This can be written as

$$P(X = x) = 1/6, \text{ for } x = 1, 2, 3, 4, 5 \text{ and } 6.$$

6.6 CONTINUOUS PROBABILITY DISTRIBUTION

The probability distribution of a continuous random variable X is defined by the functional notation $f(x)$, is called the probability density function (*pdf*) or simply the density function. A probability density function is constructed in such a way that the area under its curve bounded by the X-axis is 1, when computed over the entire range of X. In case of a continuous random variable it is not possible to find the probability of the distribution at a particular point but one can find the value of the function between two points $X = x_1$ and $X = x_2$ (say). In the figure below the probability that X lies between x_1 and x_2 is given by the shaded area under the curve $y = f(x)$ lying between the ordinates $X = x_1$ and $X = x_2$, *i.e.*, probability that X assumes a value between x_1 and x_2 = $P(x_1 \le X \le x_2)$

$$= \int_{x_1}^{x_2} f(x)\,dx$$

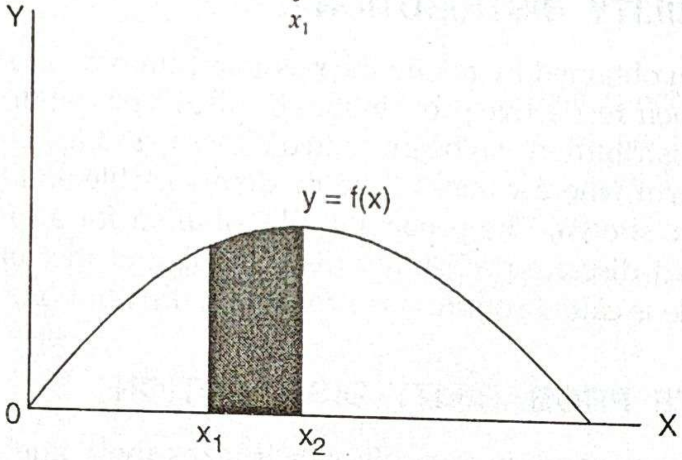

Fig. 6.1.

If F, the distribution function of the *rv* X, is absolutely continuous and f is continuous at x, we have

$$F'(x) = \frac{dF(x)}{dx} = f(x)$$

Illustration 6.1 : A discrete random variable X is defined as follows :

X :	0	1	2	3	4
$P(X = x)$:	K	3K	0.2	K	2K + 0.1

Find the following

(*i*) the value of K (*ii*) find the probability distribution (*iii*) P(X>2).

Solution : (*i*) We know that for a discrete random variable the probability mass function is such that

$$\sum P(X = x) = 1$$

Thus we have,

$$P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4) = 1$$

$$K + 3K + 0.2 + K + (2K + 0.1) = 1$$

$$\Rightarrow \quad 7K + 0.3 = 1$$

$$\Rightarrow \quad K = 0.7/7 = 0.1.$$

$\Rightarrow$

(*ii*) So, the probability distribution becomes :

X :	0	1	2	3	4
$P(X = x)$:	0.1	0.3	0.2	0.1	0.3

(*iii*) $P(X > 2) = P(X = 3) + P(X = 4) = 0.1 + 0.3 = 0.4.$

Illustration 6.2 : A random variable X takes values 0, 1, 2, ... with probability proportional to $(x + 1)\left(\frac{1}{5}\right)^x$. Find the probability that $X \le 5$.

Solution : Here $P[X = x] = \dfrac{K(x+1)}{5^x}$

Since, $\sum_{x=0}^{\infty} P(X = x) = 1 \Rightarrow \sum_{x=0}^{\infty} \dfrac{K(x+1)}{5^x} = 1 \Rightarrow K\left(1 + \frac{2}{5} + \frac{3}{5^2} + \dots\right) = 1$

$$\Rightarrow \quad K\left(1 - \frac{1}{5}\right)^{-2} = 1 \quad \Rightarrow \quad K^{-1} = \frac{25}{16} \Rightarrow K = \frac{16}{25}$$

So,

$$P[X = x] = \frac{16}{25}\frac{(x+1)}{5^x}, x = 0, 1, 2, \dots$$

Now,

$$P[X \le 5] = P[X = 0] + P[X = 1] + \dots + P[X = 5].$$

$$= \frac{16}{25}\left[1 + \frac{2}{5} + \frac{3}{5^2} + \dots + \frac{6}{5^5}\right] = 0.9997.$$

Illustration 6.3 : If $f(x) = 2x$ when $0 \le x \le 1$ and $= 0$, otherwise. Find the probability that,

(*i*) $X < \frac{1}{2}$ (*ii*) $\frac{1}{4} < X < \frac{1}{2}$ (*iii*) $X > \frac{3}{4}$ given $X > \frac{1}{2}$.

Solution :

(*i*) $P\left(X < \frac{1}{2}\right) = \int_0^{1/2} 2x\,dx = \left[x^2\right]_0^{1/2} = \frac{1}{4}$

(*ii*) $P\left(\frac{1}{4} < X < \frac{1}{2}\right) = \int_{1/4}^{1/2} 2x\,dx = \left[x^2\right]_{1/4}^{1/2} = \frac{3}{16}$

(*iii*) $P\left[X > \frac{3}{4} \middle| X > \frac{1}{2}\right] = \dfrac{P\left(X > \frac{3}{4}\right)}{P\left(X > \frac{1}{2}\right)} = \dfrac{\int_{3/4}^{1} 2x\,dx}{\int_{1/2}^{1} 2x\,dx} = \dfrac{(7/16)}{(3/4)} = \dfrac{7}{12}$

Illustration 6.4 : A biased coin is tossed three times, with P(H) = 2/3 and P(T) = 1/3. Let the random variable X represents the number of heads produced in three tosses of the said coin. Find the CDF of X and plot the same.

Solution : The all possible points of the random experiment along with the values of the random variable and their probabilities are tabulated below :

Sample Point	*Value of X = x*	*P(X = x)*
HHH	3	8/27
HHT	2	4/27
HTH	2	4/27
THH	2	4/27
HTT	1	2/27
THT	1	2/27
TTH	1	2/27
TTT	0	1/27

So,

$$
\begin{aligned}
F(x) &= 0 && \text{for } x < 0 \\
&= \frac{1}{27} && \text{for } x < 1 \\
&= \frac{9}{27} && \text{for } x < 2 \\
&= \frac{21}{27} && \text{for } x < 3 \\
&= 1 && \text{for } x \geq 3
\end{aligned}
$$

The graphical representation of the corresponding CDF is as follows.

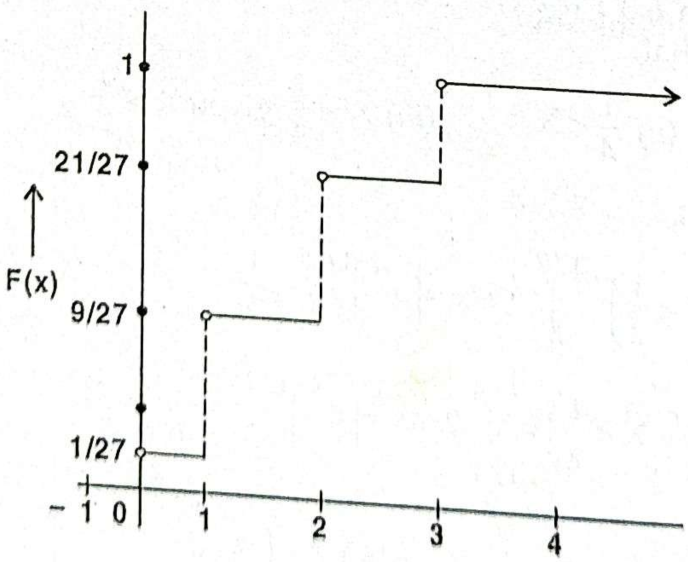

Fig. 6.2.

6.7 DISTRIBUTION FUNCTION

Let X be a one-dimensional random variable. The function F defined for all real t, by the equation

$$F_X(t) = P[X \le t]$$

is called as the distribution function or cumulative distribution function of X.

Properties of 'Distribution Function'

(*i*) $P(a < X \le b) = F(b) - F(a)$ provided $b > a$.

$$(a < X \le b) \cup (X \le a) = (X \le b)$$

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

$$\Rightarrow \quad P(a < X \le b) = P(X \le b) - P(X \le a) = F(b) - F(a).$$

(*ii*) $0 \le F(x) \le 1 \ \forall \ x \in R$

We have,

$$F(x) = P(X \le x)$$

But we know, the probability measure lies between 0 and 1, *i.e.*,

$$0 \le P(X \le x) \le 1$$

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

(*iii*) $F(x) \le F(y)$ if $x \le y$

We have,

$$F(y) - F(x) = P[x < X \le y]$$

But $\quad P[x < X \le y] \ge 0$

So, $\quad F(y) - F(x) \ge 0 \quad \Rightarrow \quad F(y) \ge F(x)$

(*iv*) $F(-\infty) = 0$ and $F(\infty) = 1$.

Let $\{x_n\}$ be any decreasing sequence such that $\lim x_n \to -\infty$.

Then, the sequence of intervals, viz., $\{]-\infty, x_n[\}$ is a decreasing sequence of intervals and

So, $$\lim_{n\to\infty}]-\infty, x_n[= \bigcap_{n=1}^{\infty} \{-\infty, x_n\} = \phi$$

Hence,

$$P\left(\lim_{n\to\infty}]-\infty, x_n[\right) = P(\phi) = 0$$

$$\Rightarrow \quad \lim_{n\to\infty} F(x_n) = 0 \quad \Rightarrow \quad F\left(\lim_{n\to\infty} x_n\right) = F(-\infty) = 0.$$

Now, let $\{x_n\}$ be an increasing sequence of real numbers such that $\lim x_n \to \infty$. Then, the sequence of intervals, viz., $\{]-\infty, x_n[\}$ is an increasing sequence of intervals, and

$$\lim_{n\to\infty} \{]-\infty, x_n]\} = \bigcup_{n=1}^{\infty} \{]-\infty, x_n]\} =]-\infty, \infty] = \Omega$$

$$\therefore \quad P\left[\lim_{n\to\infty}]-\infty, x_n]\right] = P(\Omega) = 1.$$

or $$\lim_{n\to\infty} F(x_n) = 1 \quad \Rightarrow \quad P\left(\lim_{n\to\infty} x_n\right) = F(+\infty) = 1.$$

(*v*) $\lim_{t\to a+} F(t) = F(a)$

Let $\{x_n\}$ be a decreasing sequence of real numbers such that $x_n > a$ and $\lim x_n = a$. Then the sequence of intervals, viz., $\{]-\infty, x_n]\}$ is a decreasing sequence of intervals and so,

$$\lim_{n\to\infty}]-\infty, x_n] = \bigcap_{n=1}^{\infty} \{]-\infty, x_n]\} =]-\infty, a]$$

$$\therefore \quad P\left[\lim_{n\to\infty}]-\infty, x_n]\right] = P(\,]-\infty, a]) = P[X \le a] = F(a)$$

or $$\lim_{n\to\infty} P(]-\infty, x_n]) = F(a)$$

i.e., $$\lim_{n\to\infty} F(x_n) = F(a)$$

$$\Rightarrow \quad F(a+) = F(a)$$

Thus, F is continuous from right.

Illustration 6.5 : The distribution function F of a continuous variable X is given by,

$$F(x) = 0, \; x < 0$$
$$= x^2, \quad 0 \le x \le \frac{1}{2}$$
$$= 1 - \frac{3(3-x)^2}{25}, \; \frac{1}{2} \le x < 3.$$
$$= 1, \; x \ge 3.$$

Find the pdf of X, and evaluate $P(|X| \le 1)$ and $P\left(\frac{1}{3} \le X < 4\right)$, using both F and *f*.

Solution : The distribution function, 'F' is not differentiable at $x = 0$, $x = \frac{1}{2}$, $x = 3$ while for the other points we have,

$$f(x) = 0, \; x < 0 \text{ or } x > 3.$$
$$= 2x, \; 0 < x < \frac{1}{2}$$

$$= 6(3 - x)/25, \frac{1}{2} < x < 3.$$

At $x = 0, \frac{1}{2}, 3$ we are free to choose $f(0), f\left(\frac{1}{2}\right)$ and $f(3)$ arbitrarily.

Let us choose, $f(0) = 0, f\left(\frac{1}{2}\right) = \frac{3}{5}$ and $f(3) = 0$

Then we have, $f(x) = 2x, 0 \le x < \frac{1}{2}$

$$= 6\ (x - 3)/25, \frac{1}{2} \le x \le 3$$

$$= 0, \text{ elsewhere,}$$

Thus, we have, $P\ [|X| \le 1] = P\ [-1 \le X \le 1] = \int_0^{1/2} 2x dx + \int_{1/2}^{1} \frac{6}{25}(3 - x) dx$

$$= \frac{1}{4} + \frac{27}{100} = \frac{52}{100} = 0.52$$

Also, $P\left[\frac{1}{3} \le X < 4\right] = \int_{1/3}^{3} f(x) dx = \int_{1/3}^{1/2} 2x dx + \int_{1/2}^{3} \frac{6}{25}(3 - x) dx = \frac{8}{9}$

Again $P(|X| \le 1) = P\ (-1 \le X \le 1) = F(1) - F(-1)$

$$= \left[1 - \frac{3}{25}(3-1)^2\right] - 0 = \frac{13}{25}$$

and $P\left(\frac{1}{3} \le X < 4\right) = F(4) - F(1/3) = 1 - \frac{1}{9} = \frac{8}{9}$

Illustration 6.6. Let $X_1, X_2, \ldots$ be independent and identically distributed random variables distributed as X where, $P[X = k] = \frac{K}{2^{K+1}}$, $K = 1, 2, \ldots$..

Find $P\ [X_n \ge n]$ for infinitely many n. [IAS, 1983]

Solution : Given that,

$$P\ [X = K] = \frac{K}{2^{K+1}}$$

Now, $P(X < n) = P(X = 0) + P(X = 1) + P(X = 2) + \ldots + P(X = n - 1)$

$$= 0 + \frac{1}{2^2} + \frac{2}{2^3} + \frac{3}{2^4} + \ldots + \frac{n-1}{2^n}$$

$\therefore$

$$P[X \ge n] = 1 - P[X < n]$$

$$= 1 - \frac{1}{2^2} - \frac{2}{2^3} - \frac{3}{2^4} - \ldots - \frac{n-1}{2^n}$$

$$= 1 - \sum_{n=1}^{n-1} \frac{n}{2^{n+1}}$$

Illustration 6.7 : For each of the following functions, verify whether it is a distribution function or not. In case of distribution function find the corresponding probability mass/density function.

(*i*) $F(x) = 0$ for $x \le 0$
$= x$ for $0 < x < 1$
$= 1$ for $x \ge 1$

(*ii*) $F(x) = 0$ for $x < 2$
$= 0.75$ for $x = 2$
$= 1$ for $x > 2$. [IAS, 1981]

Solution : We have the following properties for a distribution function:

(*a*) If F is the distribution function of the r.v X and if $a < b$ then $P(a < x \le b) = F(b) - F(a)$

(*b*) If F is the distribution function of one dimensional r.v X, then (*i*) $0 \le F(x) \le 1$ (*ii*) $F(x) \le F(y)$ if $x < y$.

(*c*) If F is the d.f. of one dimensional r.v. X, then

$$F(-\infty) = \lim_{x \to \infty} F(x) = 0$$

$$F(\infty) = \lim_{x \to \infty} F(x) = 1.$$

(*i*) For $x < 1$, the events '$0 < X \le 1$' and '$X \le 0$' are disjoint and their union is the event '$X \le 1$'. Hence by addition theorem of probability

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

$$\Rightarrow \quad P(0 < X \le 1) = P(X \le 1) - P(X \le 0)$$

$$= F(1) - F(0)$$

∴ 1st property is satisfied.

Again, we see that the given function, lies between 0 and 1.

∴ 2nd property is satisfied.

And $\lim_{x \to -\infty} F(x) = 0, \lim_{x \to \infty} F(x) = 1$

∴ 3rd property is satisfied.

Hence $F(x)$ is a distribution function we observe that $F(x)$ is continuous at $x = 0$ and $x = 1$.

Now,
$$\frac{d}{dx} F(x) = \begin{cases} 0, & x \le 0 \\ 1, & 0 < x < 1 \\ 0, & x \ge 1 \end{cases}$$

$$= f(x) \text{ (say)}$$

In order that F(x) is a distribution function, $f(x)$ must be a *pdf*. Thus we have to show that

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

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

$$= \int_{-\infty}^{0} 0.dx + \int_{0}^{1} 1.dx + \int_{1}^{\infty} 0.dx$$

$$= =0+[x]_0^1 +0$$

$$= 1.$$

Hence F(x) is a disribution function.

(*ii*) In order that F(x) is a distribution function $f(x)$ must be *pdf*.

We have
$$f(x) = \frac{d}{dx} F(x) = \begin{cases} 0, & x < 2 \\ 0, & x = 2 \\ 0, & x > 1 \end{cases}$$

$$\therefore \int_{-\infty}^{\infty} f(x)dx \neq 1$$

Hence F(x) is not a distribution function because for a distribution function, the total of *pdf* must be equal to 1.

Illustration 6.8 : Suppose that the amount of money that a person was saved may be regarded as a r.v. with a probability function specified by the distribution function F(x) as,

$$F(x) = \frac{1}{2} e^{-x/40} \quad \text{for} \quad x < 0$$

$$= 1 - \frac{1}{2} e^{-x/40} \quad \text{for} \quad x \geq 0$$

(*a*) Find the *pdf*

(*b*) What is the probability that the savings made by the man would be (*i*) > 40 (*ii*) < 40 (*iii*) between – 40 and + 40

Solution : (*a*) If f(n) is the pdf, then we have

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

$$= \frac{d}{dx}\left[\frac{1}{2} e^{-x/40}\right]$$

$$= \frac{1}{2} \cdot \frac{e^{-x/40}}{-1/40} = -20\, e^{-x/40},\ x < 0$$

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

$$= \frac{d}{dx}\left[1 - \frac{1}{2} e^{-x/40}\right]$$

$$= 0 - \frac{1}{2} \cdot \frac{e^{-x/40}}{-1/40}$$

$$= 20\, e^{-x/40},\ x \geq 0$$

(*b*) The probability that the savings made by the man would be > 40 is,

$$P(x > 40) = 1 - P(X \leq 40)$$

$$= 1 - F(40)$$

$$= 1 - \left[1 - \frac{1}{2} e^{-40/40}\right]$$

$$= \frac{1}{2} e^{-40/40} = \frac{1}{2e}$$

The probability that the savings made by man would be < 40

$$P(X < 40) = F(40)$$

$$= 1 - \frac{1}{2} e^{-40/40} = 1 - \frac{1}{2e}$$

(*iii*) The probability that the savings made by the man would be between − 40 and + 40 is,

$$P(-40 < X < 40) = F(40) - F(-40)$$

$$= 1 - \frac{1}{2} e^{-40/40} - \frac{1}{2} e^{40/40}$$

$$= 1 - \frac{1}{2e} - \frac{e}{2}$$

Illustration 6.9 : A pdf is defined as follows :

$$f(y) = ky^2, \text{ if } 0 \leq y \leq 3$$

$$= 0\text{ , elsewhere}$$

(*a*) Find *k* (*b*) Find the corresponding CDF

Solution : We have, for total probability

$$\int_{-\infty}^{\infty} f(y)dy = 1\,.$$

$$\Rightarrow \quad \int_{-\infty}^{0} f(y)dy + \int_{0}^{3} f(y)dy + \int_{3}^{\infty} f(y)dy = 1$$

$$\int_0^3 ky^2\, dy = 1$$

$$\Rightarrow \quad k.\left.\frac{y^3}{3}\right|_0^3 = 1$$

$$\Rightarrow \quad \frac{k}{3}[27 - 0] = 1$$

$$\Rightarrow \quad k = \frac{1}{9}$$

$$\Rightarrow \quad f(y) = \begin{cases} \frac{1}{9}y^2, & 0 \le y \le 3 \\ 0, & \text{otherwise} \end{cases}$$

$\therefore$

The corresponding CDF is given by

$$F(Y) = P(Y \le y)$$

$$= \int_{-\infty}^{y} \frac{y^2}{9} dy$$

$$= \int_{0}^{y} \frac{y^2}{9} dy$$

$$= \left.\frac{1}{9}.\frac{y^3}{3}\right|_0^y = \frac{y^3}{27}$$

Illustration 6.10 : A communication system consists of n components each of which will independently function with probability p. The total system will be able to operate effectively if atleast one half of the components function. For what value of p is a 5 component system more likely to operate effectively than a 3-component system?

Solution : Here $X \sim B(n, P)$

$$\therefore \quad p(x) = {}^nc_x\, p^x.q^{n-x};\ x = 0, 1, 2, \ldots, n$$
$$= 0; \text{otherwise}$$

Given that the total system will be able to operate effectively if atleast one half of the components function.

$\therefore$ For 5 component system, will be able to operate effectively if at least $\frac{5}{2} = 2.5 \simeq 3$ of the components function.

$$\therefore \quad P(X \ge 3) = {}^5C_3 \cdot p^3 \cdot q^2 + {}^5C_4\, p^4 \cdot q + {}^5C_5\, p^5$$
$$= 10\, p^3 (1 - p)^2 + 5p^4 (1 - p) + p^5$$

And for 3 component system will be able to operate effectively if at least $\frac{3}{2} = 1.5 \approx 2$ of the components function.

$$\therefore \quad P(X \geq 2) = {}^3C_2 . p^2 (1-p) + p^3$$
$$= 3p^2 (1-p) + p^3$$

$\therefore$ 5 component system is more likely to operate effectively than 3 component system if,

$$10 p^3 (1-p)^2 + 5p^4 (1-p) + p^5 > 3 p^2 (1-p) + p^3$$
$$\Rightarrow \quad 10 p^3 (1-p)^2 + 5p^4 (1-p) - 3p^2 (1-p) - p^3 (1-p^2) > 0$$
$$\Rightarrow \quad (1-p) \{10 p^3 (1-p) + 5p^4 - 3p^2 - p^3 (1+p)\} > 0$$
$$\Rightarrow \quad (1-p) \{10 p^3 - 10 p^4 + 5p^4 - 3p^2 - p^3 - p^4\} > 0$$
$$\Rightarrow \quad (1-p) (-6 p^4 + 9p^3 - 3p^2) > 0$$
$$\Rightarrow \quad (1-p) (-3 p^2) (2p^2 - 3p + 1) > 0$$
$$\Rightarrow \quad (1-p) (-3 p^2) (2p-1) (p-1) > 0$$
$$\Rightarrow \quad 3p^2 (p-1)^2 (2p-1) > 0$$
$$\Rightarrow \quad p > \frac{1}{2} \quad (\because p \ngtr 0 \text{ and } p \ngtr 1)$$

Illustration 6.11 : Do the following function define distribution function

$$F(x) = 0, \; x < -a$$
$$= \frac{1}{2}\left(\frac{x}{a}+2\right); -a \leq x \leq a$$
$$= 1, \; x > a$$

Solution : The pdf is given by

$$f(x) = \frac{d}{dx} F(x)$$
$$= \begin{cases} \frac{1}{2a}; & -a \leq x \leq a \\ 0, & \text{otherwise} \end{cases}$$

Now we are to prove that $\int_{-\infty}^{\infty} f(x)dx = 1$

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

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

$$= \frac{1}{2a} x \Big|_{-a}^{a}$$

$$= \frac{1}{2a} .2a = 1$$

$\therefore$ $f(x)$ is a pdf and therefore F(x) is the distribution function of $f(x)$.

Illustration 6.12 : Consider the distribution functions,

$$\begin{aligned} F_1(x) &= 0 && \text{if } x < 0 \\ &= 1 && \text{if } x \geq 0 \\ F_2(x) &= 0 && \text{if } x \leq 0 \\ &= x && \text{if } 0 < x < 1 \\ &= 1 && \text{if } x \geq 1 \end{aligned}$$

Define the function F by,

$$F(x) = \frac{1}{2} F_1(x) + \frac{1}{2} F_2(x)$$

Show that F is also a distribution function. Is the corresponding random variable discrete or continuous?

Solution : Given that

$$\begin{aligned} F_1(x) &= 0 && \text{if } x < 0 \\ &= 1 && \text{if } x \geq 0 \end{aligned}$$

and

$$\begin{aligned} F_2(x) &= 0 && \text{if } x \leq 0 \\ &= x && \text{if } 0 < x < 1 \\ &= 1 && \text{if } x \geq 1 \end{aligned}$$

$F_1(x)$ and $F_2(x)$ can be written as

$$\begin{aligned} F_1(x) &= 0 && \text{if } x < 0 \\ &= 1 && \text{if } x = 0 \\ &= 1 && \text{if } 0 < x < 1 \\ &= 1 && \text{if } x \geq 1 \end{aligned}$$

and

$$\begin{aligned} F_2(x) &= 0 && \text{if } x < 0 \\ &= 0 && \text{if } x = 0 \\ &= x && \text{if } 0 < x < 1 \\ &= 1 && \text{if } x \geq 1 \end{aligned}$$

$$\therefore F(x) = \frac{1}{2} F_1(x) + \frac{1}{2} F_2(x)$$

$$\begin{aligned} &= 0 && \text{if } x < 0 \\ &= \frac{1}{2} && \text{if } x = 0 \\ &= \frac{1}{2} + \frac{x}{2} && \text{if } 0 < x < 1 \\ &= \frac{1}{2} + \frac{1}{2} = 1 && \text{if } x \geq 1 \end{aligned}$$

which is the distribution function.

$F_1(x)$ is the distribution function of the discrete r.v. X, which is degenerate at $x = 0$ and $F_2(x)$ is the distribution function of the continuous random variable. Therefore F(x) is mixture distribution function of discrete and continuous r.v.

Illustration 6.13. The pdf of random variable X is given by.

$$\begin{aligned} f(x) &= 2x, && 0 \leq x < 1 \\ &= 0, && \text{otherwise.} \end{aligned}$$

Find the pdf of $y = 3x$.

Solution : Let $g(y)$ represents the pdf of Y, and let $G(y)$ be its corresponding CDF.

So $$G(y) = P(Y \le y) = P[3X \le y]$$

$$= P\left[X \le \frac{y}{3}\right]$$

$$= \int_0^{y/3} 2x\,dx = \left[2\frac{x^2}{2}\right]_0^{y/3} = \frac{y^2}{9}$$

Thus $$G(y) = 0, \quad y < 0$$

$$= \frac{y^2}{9}, \quad 0 \le y \le 3 \text{ (as, } y = 3x)$$

$$= 1, \quad y > 3.$$

Hence, the corrosponding pdf is given by,

$$g(y) = \frac{2y}{9}, \quad 0 \le y \le 3$$

$$= 0, \quad \text{otherwise.}$$

Theorem : If X is any continuous random variable and $\mathbf{Y} = h(x)$, be any strictly increasing or decreasing function, then, the density function of $\mathbf{Y}$ is given by,

$$g(y) = f(x)/|h'(x)| = f(h^{-1}(y))/|h'(h^{-1}(y))|$$

where, $f(x)$ is the density function of X.

Proof : Let F(x) and G(y) be the cumulative distribution function of X and Y respectively. So,

$$F(x) = \int_{-\infty}^{x} f(u)du \text{ and } G(y) = \int_{-\infty}^{y} g(v)dv$$

Also, if $y = h(x)$ is an increasing function then,

$$F(x) = P[x \le x]$$

$$= P[h^{-1}(y) \le x]$$

$$= P[y \le h(x)] = G(h(x)) = G(y)$$

Again if $y = h(x)$ is a decreasing function, then,

$$1 - F(x) = 1 - P[X \le x]$$

Now, irrespective of the nature of the function $H(x)$, we would have,

$$g(y) = G'(y)$$

$$= F(x)\left|\frac{1}{h'(x)}\right|$$

$$= f(x)\left|\frac{1}{h'(x)}\right|$$

$$= f[h^{-1}(y)]\left|\frac{1}{h'\left(h^{-1}(y)\right)}\right|$$

Illustration 6.14 : If X is a random variable with pdf $f(x)$, what is the density function of $Y = aX + b$.

Solution : Here, $Y = aX + b$, So $h(x) = ax + b = y$.

$$\Rightarrow \quad x = \frac{y-b}{a} \text{ So, } \frac{dx}{dy} = \frac{1}{a}, \; h^{-1}(y) = x = \frac{y-b}{a} \text{ and } h'(x) = a$$

Thus,
$$g(y) = f(h^{-1}(y)), \left|\frac{1}{h'\left(h^{-1}(y)\right)}\right|$$

$$= f\left(\frac{y-b}{a}\right) \times \left|\frac{1}{a}\right| = \frac{1}{|a|} f\left(\frac{y-b}{a}\right)$$

Illustration 6.15 : A random variable X has pdf

$$f(x) = \frac{1}{x^2}, \qquad x \geq 1$$
$$= 0 \qquad \text{otherwise}$$

Find the pdf of e^{-x}.

Solution : Let $Y = e^{-X}$. So, $x = -\log y$

Also, $\left|\frac{dx}{dy}\right| = \frac{1}{y}$

Now, $x \geq 1 \Rightarrow -\log y \geq 1 \Rightarrow \log y \leq -1$

$\Rightarrow y \geq e^{-1}$

Hence,

$$g(y) = f(-\log y)\left|\frac{1}{y}\right|, \qquad 0 < y \leq e^{-1}$$

$$= \frac{1}{y(\log y)^2}, \qquad 0 < y \leq e^{-1}$$

required density function.

□□