Chapter UNIT-II

2

CONTINUOUS PROBABILITY DISTRIBUTION

◆ § 2.1. RANDOM VARIABLES

Definitions. A real valued function defined on a sample-space is called a **random-variable** (or **a discrete random variable**).

A random variable can assume only a set of real values and the values which the variable takes depends on the chance. Random variable is also called **stochastic variable** or simply a **variate. For example.**

Suppose a perfect die is thrown then x, the number of points on the die is a random variable since x has the following two properties

(i) x takes only a set of discrete values 1, 2, 3, 4, 5, 6;

(ii) the values which x takes depends on the chance.

Actually x takes values 1, 2, 3, 4, 5, 6 each with probability 1/6.

The set of values 1, 2, 3, 4, 5, 6 with their probabilities 1/6 is called the **Probability Distribution** of the variate x.

In general; suppose that corresponding to x exhaustive and mutually exclusive cases obtained from a trial, a variate x takes n values $x_1, x_2, \ldots, x_n$ with their probabilities $p_1, p_2, \ldots, p_n$.

The set of values x_i (for $i = 1, 2, 3, \ldots, n$) with their probabilities p_i (for $i = 1, 2, 3, \ldots, n$) is called the Probability Distribution of the variable of that trial. It is to be noted that most of the properties of frequency distribution will be equally applicable to probability distribution.

Continuous Random Variable (variate)

So far we have discussed with discrete variate which takes a finite set of values.

When we deal with variates like weights and temperature then we know that these variates can take an infinite number of values in a given interval. Such type of variates are known as continuous variates.

Definition. A variate which is not discrete, *i.e.*, which can take infinite number of values in a given interval $a \le x \le b$, is called a continuous variate.

For example. In the curve $y = \frac{1}{4} \sin x \ (0 \le x \le \pi)$ x is a continuous variate, since x can take all values between 0 and π, *i.e.*, x can take an infinite set of values lying between 0 and π.

§ 2.2. PROBABILITY MASS FUNCTION

Let X be a discrete random variable taking at most a countably infinite number of values $x_1, x_2, x_3, \ldots$. Now we associate a number $p_i = P(X = x_i) = p(x_i)$, say, with each possible outcome x_i where $p(x_i)$ or p_i is called the probability of x_i. The numbers $p(x_i)$ for $i = 1, 2, 3, \ldots$ must satisfy the following two conditions :

(i) $p(x_i) \ge 0$, for every $i = 1, 2, 3, \ldots$

In other words, the probability of any specific outcome for a discrete random variable must be between 0 and 1.

and (ii) $\sum_{i=1}^{\infty} p(x_i) = 1$.

The sum of the probabilities over all possible values of a discrete random variable must be equal to 1.

i.e., then the function $p(x_i)$ is said to be the **probability mass function** of the random variable X.

The set $\{p(x_i)\}$ is known as **probability distribution** of the random variable X.

Note. The **spectrum** of the random variable X is the set of values which X takes up.

For illustration. In tossing a fair coin, sample space $S = \{H, T\}$. Let the discrete random variable X be defined by

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

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

Then the *probability function* is given by

$$P(\{H\}) = (\{T\}) = \frac{1}{2}.$$

The *probability distribution* of the random variable X is given by

$$P(H) = P(X = 1) = \frac{1}{2}$$

$$P(T) = P(X = 0) = \frac{1}{2}.$$

§ 2.3. PROBABILITY DENSITY FUNCTIONS

Let X be a continuous random variable and let the probability of X falling in the infinitesimal interval $\left(x - \frac{1}{2}\, dx, x + \frac{1}{2}\, dx\right)$ be expressed by $f(x)\, dx$, *i.e.*,

$$P\left(x - \frac{1}{2}\, dx < X < x + \frac{1}{2}\, dx\right) = f(x)\, dx$$

where $f(x)$ is a continuous function of X and satisfies the following two conditions :

(i) $f(x) \geq 0$

(ii) $\int_a^b f(x)\,dx = 1$, if $a \leq X \leq b$

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

then the function $f(x)$ is called the **probability density function** (or in brief **p.d.f.**) of the continuous random variable X.

The continuous curve $Y = f(x)$ is called the **Probability density curve** (or in brief **probability curve**). The length of infinitesimal interval $\left(x - \frac{1}{2}\,dx, x + \frac{1}{2}\,dx\right)$ is dx and its mid point is x.

Remarks :

(1) If the range of X be finite, then also it can be expressed as infinite range. For example,

$$\begin{aligned} f(x) &= 0, && \text{for } x < a \\ f(x) &= \phi(x), && \text{for } a \leq x \leq b \\ f(x) &= 0 && \text{for } x > b. \end{aligned}$$

(2) The probability that a value of continuous variable X lies within the interval (c, d) is given by

$$P(c \leq X \leq d) = \int_c^d f(x)\,dx.$$

(3) The continuous variable always takes values within a given interval howsoever small the interval may be.

(4) If X be a continuous random variable, then

$$P(X = k) = 0$$

where k is a constant quantity.

§ 2.4. CONTINUOUS PROBABILITY DISTRIBUTION

The probability distribution of continuous random variate is called the *continuous probability distribution* and it is expressed in terms of probability density function.

§ 2.5. CUMULATIVE DISTRIBUTION FUNCTION

The probability that the value of a random variate X is 'x or less that x' is called the *cumulative distribution functions* of X and is usually denoted by $F(x)$. In symbolic notation, the cumulative distribution function of discrete random variate X is given by

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

The cumulative distribution function of a continuous random variate is given by

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

Some properties of cumulative distribution function :

(i) $F(-\infty) = 0.$

(ii) $F(x)$ is non-decreasing function, *i.e.*, $x_1 < x_2 \Rightarrow F(x_1) < F(x_2)$.

(iii) For a discontinuous variate

$$P(a < X < b) = F(b) - F(a) = \Sigma\, p(x_i),\ a < x_i < b.$$

And for a continuous variate

$$\begin{aligned} P(a < X \le b) &= P(a \le X \le b) \\ &= P(a \le X < b) \\ &= P(a < X < b) = F(b) - F(a) \\ &= \int_a^b f(x)\, dx. \end{aligned}$$

(iv) $F(+\infty) = 1.$

(v) $F(x)$ is a discontinuous function for a discontinuous variate and $F(x)$ is continuous function for a continuous variate.

ILLUSTRATIVE EXAMPLES

Example 1. ***The p. d. f. of a random variate X is given by***

$$p(x) = \begin{cases} k & \text{if } x = 0 \\ 2k & \text{if } x = 1 \\ 3k & \text{if } x = 2 \\ 0 & \text{otherwise.} \end{cases}$$

(a) ***Find the values of k.***

(b) ***Evaluate*** $P(X < 2)$, $P(X \le 2)$, $P(0 < X < 2)$.

(c) ***Find distribution function of X.***

Solution. By the properties of p. d. f., we have

(i) $p(x) \ge 0.$

(ii) $\Sigma\, p(x) = 1$, therefore, we have

$$k + 2k + 3k + 0 = 1 \text{ or } 6k = 1$$

$\therefore \quad k = 1/6 > 0.$

Consequently the p. d. f., $p(x)$ is defined as :

$$p(x) = \begin{cases} 1/6 & \text{if } x = 0 \\ 2/6 & \text{if } x = 1 \\ 3/6 & \text{if } x = 2 \\ 0 & \text{otherwise.} \end{cases}$$

So
$$\begin{aligned} P(X < 2) &= P(X = 0 \text{ or } X = 1) \\ &= P(X = 0) + P(X = 1). \\ &= \frac{1}{6} + \frac{2}{6} = \frac{3}{6} = \frac{1}{2} \end{aligned}$$

$$P(X \le 2) = P(X = 0 \text{ or } X = 1 \text{ or } X = 2)$$
$$= P(X = 0) + P(X = 1) + P(X = 2)$$
$$= \frac{1}{6} + \frac{2}{6} + \frac{3}{6} = \frac{6}{6} = 1$$

and $P(0 < X < 2) = P(X = 1) = 2/6 = 1/3.$

(i) Since distribution, $F(x) = \sum_{X \le x} p(x)$

$\therefore$
$$F(x) = 0, \text{ if } x < 0$$
$$= \sum_{X \le 0} p(x) = 0 + \frac{1}{6} = \frac{1}{6}, \text{ if } x = 0$$
$$= \sum_{X \le 1} p(x) = 0 + \frac{1}{6} + \frac{2}{6} = \frac{3}{6} = \frac{1}{2}, \quad \text{if } x = 1$$
$$= \sum_{X \le 2} p(x) = 0 + \frac{1}{6} + \frac{2}{6} + \frac{3}{6} = 1, \text{ if } x = 2$$
$$= 1, \text{ if } x > 2.$$

So
$$F(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1/6 & \text{if } 0 \le x < 1 \text{ or } x = 0 \\ \frac{1}{2} & \text{if } 1 \le x < 2 \text{ or } x = 1 \\ 1 & \text{if } x \ge 2. \end{cases}$$

Example 2. *If the function $f(x)$ is defined by*
$$f(x) = ce^{-x}, 0 \le x < \infty$$
find the value of c which changes $f(x)$ to probability density function.

Solution. By property of p. d. f., we have

(i) $f(x) \ge 0$ for all x.

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

Since e^{-x} is + ve for all values of x in the interval $0 \le x < \infty$, the condition will be satisfied if $c \ge 0$.

Again the 2nd condition will be satisfied

if $\int_0^{\infty} ce^{-x}\, dx = 1$ *i.e.,* if $\left[-ce^{-x}\right]_0^{\infty} = 1$

i.e., if $c = 1.$ Ans.

Example 3. *If $f(x) = cx^2, 0 < x < 1$, find the value of c and determine the probability that $\frac{1}{3} < x < \frac{1}{2}$.*

Solution. By property of probability density function, we have
$$\int_0^1 f(x)\, dx = 1$$

$\therefore$
$$\int_0^1 cx^2\, dx = 1 \text{ or } c\left[\frac{x^3}{3}\right]_0^1 = 1$$

or
$$c/3 = 1; \quad \therefore c = 3.$$

Consequently $f(x) = 3x^2, \ 0 < x < 1.$

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

$$= \frac{1}{8} - \frac{1}{27} = \frac{19}{216}.$$ **Ans.**

Example 4. ***If the random variate X has the values 0, 1, 2, ..., then for what value of t,*** $p(x) = (1-t)\, t^x,\ x = 0, 1, 2, \ldots$ ***is a probability density function of X ?***

Solution. By property of probability function, we have

(i) $p(x) \geq 0$, (ii) $\Sigma\, p(x) = 1$

where $p(x) = P(X = x)$, therefore the value of $p(x)$ should lie between 0 and 1.

Clearly :

$$0 \leq p(x) \leq 1 \ \text{ if } \ 0 \leq t \leq 1.$$

Now $$\Sigma p\,(x) = \sum_{x=0}^{\infty} (1-t)\, t^x$$

$$= (1-t)\,[1 + t + t^2 + \ldots]$$

$$= (1-t)\,(1-t)^{-1} = 1.$$

Example 5. ***If*** $f\,(x) = \dfrac{c}{1+x^2}$, $-\infty < x < \infty$, ***then find c and show that its corresponding distribution function is***

$$F\,(x) = (1/\pi)\, \tan^{-1} x + \frac{1}{2}.$$

Solution. By property of probability function, we have

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

or $$c\int_{-\infty}^{\infty} \frac{dx}{1+x^2} = 1 \ \text{ or } \ 2c\int_{-\infty}^{\infty} \frac{dx}{1+x^2} = 1$$

or $$2c\left[\tan^{-1} x\right]_0^{\infty} = 1 \ \text{ or } \ 2c\left(\frac{\pi}{2} - 0\right) = 1.$$

$\therefore$ $$c = 1/\pi.$$

So $$f(x) = \frac{1}{\pi} \cdot \frac{1}{1+x^2},\ -\infty < x < \infty.$$

Now $$F(x) = P(X \leq x)$$

$$\int_{-\infty}^{x} \frac{1}{\pi} \cdot \frac{1}{1+x^2}\, dx = \frac{1}{\pi}\left[\tan^{-1} x\right]_{-\infty}^{x}$$

$$= (1/\pi)\,[\tan^{-1} x - \tan^{-1}(-\infty)]$$

$$= (1/\pi)\,[\tan^{-1} x + \tan^{-1}(\infty)]$$

$$= (1/\pi)\,[\tan^{-1} x + \pi/2]$$

$$= (1/\pi)\tan^{-1} x + \frac{1}{2}.$$

Example 6. ***A random variate X has the distribution function***

$$F(x) = \begin{cases} 1-(1+x)\,e^{-x} & , \text{ if } x \geq 0 \\ 0 & , \text{ if } x < 0. \end{cases}$$

Determine the corresponding density function of the random variate X.

Solution. The probability density function

$$f(x) = (d/dx)\,F(x).$$

$$\therefore \quad f(x) = -e^{-x} + (1+x)\,e^{-x}$$

$$= \begin{cases} xe^{-x} & , \text{ if } x \geq 0 \\ 0 & , \text{ if } x < 0. \end{cases}$$

Example 7. ***Show that***

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

is a distribution function.

Solution. $F(x)$ will be distribution function if

(i) $F(-\infty) = 0$, (ii) $F(\infty) = 1$,

and (iii) $\dfrac{d}{dx} F(x) = f(x) \geq 0$

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

Now, since $F(x) = 0$ when $x < -a$

and $F(x) = 1$ when $x > a$.

So conditions (i) and (ii) are satisfied.

Again $$\frac{d}{dx} F(x) = \frac{1}{2}\left(\frac{1}{a}+0\right) = \frac{1}{2a} \geq 0$$

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

$$= \left[\frac{x}{2a}\right]_{-a}^{a} = \frac{a-(-a)}{2a} = 1.$$

Thus condition (iii) is also satisfied.

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

Example 8. ***If*** $$f(x) = \begin{cases} 0 & , x < 2 \\ \dfrac{1}{18}(3+2x) & , 2 \leq x \leq 4 \\ 0 & , x > 4 \end{cases}$$

Prove that it is a density function. Find the probability that a variate having this density will fall in the interval $2 \leq x \leq 3$.

Solution. For $f(x)$ to be a density function, we should have

(i) $f(x) \geq 0$,

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

By definition of $f(x)$, $f(x) \geq 0\ \forall\ x$. So condition (i) is satisfied.

$$\text{Again } \int_{-\infty}^{\infty} f(x)\, dx = \int_{-\infty}^{2} f(x)\, dx + \int_{2}^{4} f(x)\, dx + \int_{4}^{\infty} f(x)\, dx$$

$$= \int_{-\infty}^{2} 0\, dx + \int_{2}^{4} \frac{3+2x}{18}\, dx + \int_{4}^{\infty} 0\, dx$$

$$= 0 + \frac{1}{18}\left[3x + x^2\right]_2^4 + 0 = \frac{1}{18}(6+12) = 1.$$

Thus condition (ii) is also satisfied.

$$\text{Now } \quad P(2 \leq x \leq 3) = \int_{2}^{3} f(x)\, dx$$

$$= \int_{2}^{3} f(x)\, \frac{3+2x}{18}\, dx$$

$$= \frac{1}{18}\left[3x + x^2\right]_2^3 = \frac{1}{18}(3+5) = \frac{8}{18} = \frac{4}{9}.$$

Example 9. ***A random variable X has the density function***

$$\begin{aligned} f(x) &= ax,\ 0 \leq x < 1 \\ &= a,\ 1 \leq x \leq 2 \\ &= -ax + 3a,\ 2 < x \leq 3 \\ &= 0,\ \textit{otherwise.} \end{aligned}$$

Determine 'a' and compute $P(X \leq 1 \cdot 5)$.

Solution. By property of p. d. f., we have

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

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

or $$\int_{0}^{1} ax\, dx + \int_{1}^{2} a\, dx + \int_{2}^{3} (-ax + 3a)\, dx = 1$$

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

or $$\frac{a}{2} + a + \frac{a}{2} = 1 \text{ or } a = \frac{1}{2} \qquad \ldots(1)$$

$$\text{Again } P(X \leq 1 \cdot 5) = \int_{-\infty}^{1 \cdot 5} f(x)\, dx = \int_{-\infty}^{0} f(x)\, dx + \int_{0}^{1} f(x)\, dx + \int_{1}^{1 \cdot 5} f(x)\, dx$$

$$= 0 + \int_{0}^{1} ax\, dx + \int_{1}^{1 \cdot 5} a\, dx = \left[\frac{ax^2}{2}\right]_0^1 + \left[ax\right]_1^{1 \cdot 5}$$

$$= (a/2) + 0 \cdot 5\ a = a = \frac{1}{2} \qquad \text{[from (1)]}$$

Example 10. ***A continuous random variable X has the density function***

$$f(x) = 3x^2, \quad 0 \le x < 1$$

find a and b, when

(i) $P(X \le a) = P(X > a)$;

(ii) $P(X > b) = 0 \cdot 05$.

Solution. Since total probability is always one (1), so

$$P(X \le a) = \frac{1}{2} = P(X > a)$$

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

$$\Rightarrow \int_0^a 3x^2 \, dx = \frac{1}{2}$$

$$\Rightarrow a^3 = 1/2; \text{ or } a = (1/2)^{1/3}.$$

(ii) $\quad P(X > b) = 0 \cdot 05 \Rightarrow \int_b^1 f(x)\, dx = \frac{1}{20}$

$$\Rightarrow \int_b^1 3x^2 \, dx = \frac{1}{20}$$

$$\Rightarrow [x^3]_b^1 = \frac{1}{20} \Rightarrow 1 - b^3 = \frac{1}{20}$$

$$\Rightarrow b^3 = 1 - \frac{1}{20} = \frac{19}{20} \quad \text{or} \quad b = \left(\frac{19}{20}\right)^{1/3}.$$

Example 11. ***A random variable X has the probability density function***

$$f(x) = \begin{cases} x, & \text{if } 0 \le x < 1 \\ 2 - x, & \text{if } 1 \le x < 2 \\ 0, & \text{if } x \ge 2 \end{cases}$$

Find cumulative distribution function of X.

Solution. By def. of cumulative distribution function,

$F(x) = 0$, for $x < 0$

$F(x) = \int_0^x x\, dx = \frac{x^2}{2}$, for $0 \le x < 1$

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

$= 2x - \frac{x^2}{2} - \frac{3}{2}$, for $1 \le x < 2$

and $\quad F(x) = 0$, for $x \ge 2$. But $f(\infty) = 1$.

EXERCISE 2 (A)

1. What do you understand by probability density function ? Explain.
2. Find the value of c so that the function $p(x)$ defined as follows be a probability density function :

$$p(x) = \begin{cases} c\,q^x, & \text{for } x = 1, 2, 3, \ldots \\ 0, & \text{otherwise} \end{cases}$$

3. Find the value of k so that the function $f(x)$ defined as follows be a probability density function :

$$f(x) = 1/k,\ a \le x \le b = 0, \text{ otherwise.}$$

4. If
$$f(x) = 2x, \text{ for } 0 < x < 1$$
$$= 0, \text{ otherwise.}$$

Find :

(i) $P\left(X < \dfrac{1}{2}\right)$, (ii) $P\left(\dfrac{1}{4} < X < \dfrac{1}{2}\right)$.

5. Find the value of c so that the function $f(x)$ defined as follows be a probability density function :

(i) $f(x) = c\ e^x,\ x > 0$
$= 0$, otherwise.

(ii) $f(x) = cx\,e^x,\ x > 0$
$= 0$, otherwise.

6. For the probability density function $f(x) = cx^{-2}, x \ge 100$:

Find (i) constant c, (ii) $F(x)$, and (iii) $P(X > 500)$.

7. If
$$f(x) = \begin{cases} x/2, & 0 < x \le 1 \\ \dfrac{1}{4}(3 - x), & 1 < x \le 2 \\ \dfrac{1}{4}, & 2 < x \le 3 \\ \dfrac{1}{4}(4 - x), & 3 < x \le 4 \\ 0, & \text{otherwise} \end{cases}$$

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

(ii) Compute $P(X \ge 3), P(1 < x < 3), P(|X| < 1{\cdot}5)$.

8. If
$$F(x) = \begin{cases} 0, & x \le 1 \\ \dfrac{(x-1)^4}{16}, & 1 < x \le 3 \\ 1, & x > 3. \end{cases}$$

Find : (i) Corresponding density function $f(x)$.

(ii) $P(2 < X \le 3)$.

9. Find the value of m, if
$$F(x) = \begin{cases} 0, & x \le -1 \\ m\,(x+1), & -1 \le x \le 3 \\ 4m, & 3 \le x \le 4 \\ 0, & x > 4 \end{cases}$$

be a probability density function.

10. Find the value of k so that the function $f(x)$ defined as follows be a density function

$$f(x) = \frac{k}{x^2 + a^2}, \quad a\sqrt{3} \geq x \geq a$$
$$= 0, \text{ otherwise.}$$

11. State the conditions under which a function is called a probability density function. Test whether the function $f(x)$ defined as follows is a probability density function or not :

$$f(x) = \frac{3}{8}(4x - 2x^2), \quad 0 \leq x \leq 2$$
$$= 0, \text{ otherwise.}$$

ANSWERS

2. $(1-q)/q$ 3. $b-a$ 4. (a) (i) $\frac{1}{4}$ (ii) $\frac{3}{16}$ 5. (i) $c = 1$; (iv) $c = \frac{1}{\sqrt{2}}$

6. (i) $c = 100$ 7. (ii) $\frac{1}{8}$; $\frac{3}{8}$; $\frac{15}{32}$ 8. (i) $f(x) = \frac{(x-1)^3}{4}$, $1 \leq x \leq 3$; (ii) $\frac{15}{16}$

9. $m = \frac{1}{12}$ 10. $k = \frac{129}{7\pi}$ 11. f is p.d.f.

◈ § 2.6. PROPERTIES OF CONTINUOUS RANDOM VARIABLE

Mean, Median, Mode and Moments for a Continuous Distributions.

Definitions :

(1) (i) Mean $\overline{M} = X = E(X)$

$$= \int_{-\infty}^{\infty} x\, f(x)\, dx \text{ if } -\infty < x < \infty.$$

Also mean $= \int_a^b x\, f(x)\, dx$, if $a \leq x \leq b$.

(2) Geometric Mean. If G is geometric mean, then

$$\log G = \int_{-\infty}^{\infty} \log x\, f(x)\, dx = E(\log x) \text{ if } -\infty < x < \infty$$
$$= \int_a^b \log x \,.\, f(x)\, dx = E(\log x) \text{ if } a < x < b.$$

(3) Harmonic Mean. Let H be the harmonic mean, then

$$\frac{1}{H} = \int_{-\infty}^{\infty} \frac{1}{x} f(x)\, dx = E\left(\frac{1}{x}\right) \text{ if } -\infty < x < \infty$$
$$= \int_a^b \frac{1}{x} f(x)\, dx = E\left(\frac{1}{x}\right) \text{ if } a < x < b.$$

(4) Median. The median M_d is given by

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

(ii) The lower quartile Q_1 and upper quartile Q_3 are given by

$$\int_{-\infty}^{Q_1} f(x)\, dx = \frac{1}{4} \text{ and } \int_{Q_3}^{\infty} f(x)\, dx = \frac{1}{4}$$
$$\int_{-\infty}^{Q_3} f(x)\, dx = \frac{3}{4} \text{ and } \int_{Q_1}^{\infty} f(x)\, dx = \frac{3}{4}.$$

(5) **Mode.** The mode is the value of the variate for which

$$\frac{d}{dx} f(x) = 0 \text{ and } \frac{d^2 f(x)}{dx^2} < 0.$$

[In other words, the mode is the value of the variate for which probability $f(x)$ is maximum]. The condition for which is that the values obtained from $(d/dx) f(x) = 0$ lies within the given range of x.

(6) **Moment.** The rth moment about any given arbitrary value A is given by :

$$\mu'_r = \int_{-\infty}^{\infty} (x - A)^r f(x)\, dx, \text{ if } -\infty < x < \infty$$

$$= \int_a^b (x - A)^r f(x)\, dx, \text{ if } a < x < b.$$

The rth moment about the mean m is given by :

$$\mu_r = \int_{-\infty}^{\infty} (x - m)^r f(x)\, dx, \text{ if } -\infty < x < \infty$$

$$= \int_a^b (x - m)^r f(x)\, dx, \text{ if } a < x < b.$$

(7) **The mean deviation about the mean m.** It is given by

$$\int_{-\infty}^{\infty} | x - m | f(x)\, dx, \text{ if } -\infty < x < \infty$$

$$= \int_a^b | x - m | f(x)\, dx, \text{ if } a < x < b.$$

(8) **Variance.** For a continuous distribution the variance σ^2 is given by

$$\int_{-\infty}^{\infty} (x - m)^2 | f(x)\, dx, \text{ if } -\infty < x < \infty$$

$$= \int_a^b (x - m)^2 f(x)\, dx, \text{ if } a < x < b.$$

(9) **Standard deviation (S. D.).** The positive square root of variance is called S. D. and is denoted by σ.

ILLUSTRATIVE EXAMPLES

Example 1. ***For the distribution $dF = 6(x - x^2)\, dx$, $0 \le x \le 1$, find arithmetic mean, harmonic mean, median, mode, mean deviation and standard deviation. Is it a symmetrical distribution ?***

Solution. For the given distribution, we have

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

$\therefore \quad f(x) = 6(x - x^2)$ is a p.d.f.

(1) **Arithmetic Mean.** $M = \int_0^1 x \cdot f(x)\, dx$

$$= \int_0^1 x \cdot 6(x - x^2)\, dx = 6 \int_0^1 (x^2 - x^3)\, dx$$

$$= 6\left[\frac{1}{3} x^3 - \frac{1}{4} x^4 \right]_0^1 = 6\left(\frac{1}{3} - \frac{1}{4} \right) = \frac{1}{2}.$$

(2) Harmonic Mean. It is given by

$$\frac{1}{H} = \int_0^1 \frac{1}{x} \cdot 6(x - x^2)\, dx$$

or $$\frac{1}{H} = 6\left[x - \frac{x^2}{2}\right]_0^1 = 3$$

$\therefore$ $$H = \frac{1}{3}.$$

(3) Median. M_d is given by :

$$\int_0^{M_d} f(x)\, dx = \frac{1}{2}$$

or $$\int_0^{M_d} 6(x - x^2)\, dx = \frac{1}{2}$$

or $$6\left[\frac{1}{2}x^2 - \frac{1}{3}x^3\right]_0^{M_d} = \frac{1}{2}$$

or $$4M_d^3 - 6M_d^2 + 1 = 0$$

or $$(2M_d - 1)(2M_d^2 - 2M_d - 1) = 0$$

or $$M_d = \frac{1}{2} \text{ or } M_d = \frac{1}{2}(1 \pm \sqrt{3}).$$

Since M_d lies between 0 and 1.

$\therefore$ $$M_d = \frac{1}{2}.$$

(4) Mode. It is that value of variate x for which

$$dy/dx = 0 \text{ and } d^2y/dx^2 < 0.$$

Here $$y = 6(x - x^2).$$

$\therefore$ $$dy/dx = 6(1 - 2x) \text{ and } d^2y/dx^2 = -12.$$

$\therefore$ $$dy/dx = 0 \Rightarrow 6(1 - 2x) = 0 \Rightarrow x = \frac{1}{2}.$$

So mode to given distribution is $x = \frac{1}{2}$.

Since mean, median and mode coincide, therefore the distribution is symmetrical.

(5) Mean deviation about the mean

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

$$= 6\int_0^1 \left(x^2 - x + \frac{1}{4}\right)(x - x^2)\, dx$$

$$= 6\int_0^1 \left(-x^4 + 2x^3 - \frac{5}{4}x^2 + \frac{1}{4}x\right) dx$$

$$= 6\left[-\frac{1}{5}x^5 + \frac{2}{4}x^4 - \frac{5}{15}x^3 + \frac{1}{8}x^2\right]_0^1$$

$$= 6 \times \frac{1}{120} = \frac{1}{20} = 0\cdot 05.$$

(6) Standard deviation.

Variance $(\sigma^2) = \int_0^1 (x - M)^2\, f(x)\, dx$

$$= \int_0^1 \left(x - \frac{1}{2}\right)^2 . 6\,(x - x^2)\, dx$$

$$= 6\int_0^1 \left(x^2 - x + \frac{1}{4}\right)(x - x^2)\, dx$$

$$= 6\int_0^1 \left(-x^4 + 3x^3 - \frac{5}{4}x^2 + \frac{1}{4}x\right) dx$$

$$= 6\left[-\frac{1}{5}x^5 + \frac{2}{4}x^4 - \frac{5}{12}x^3 + \frac{1}{8}x^2\right]_0^1$$

$$= 6 \times \frac{1}{120} = \frac{1}{20} = 0\cdot 05.$$

$\therefore$ Standard deviation $(\sigma) = \sqrt{(0\cdot 05)} = 0\cdot 223.$ **Ans.**

Example 2. ***Find the mode and median for the frequency curve***

$$y = \frac{1}{2}\sin x,\ 0 \le x \le \pi.$$

Solution. Here total frequency $= \int_0^\pi \frac{1}{2}\sin x\, dx = \frac{1}{2}\left[-\cos x\right]_0^\pi = 1.$

We know that the mode is that value of x for which

$$dy.dx = 0 \text{ and } d^2y/dx^2 = -\text{ve}$$

Now $\quad dy/dx = \frac{1}{2}\cos x = 0 \Rightarrow \cos x = 0 \Rightarrow x = \pi/2.$

Also $\quad \frac{d^2y}{dx^2} = -\frac{1}{2}\sin x = -\frac{1}{2}$, at $x = \frac{1}{2}\pi,$

i.e., $\quad \frac{d^2y}{dx^2} = -\text{ve at } x = \frac{\pi}{2}.$

Therefore the mode is $x = \pi/2.$

$$\int_0^{M_d} \frac{1}{2}\sin x\, dx = \frac{1}{2}$$

$\Rightarrow \quad [-\cos x]_0^{M_d} = 1$

$\Rightarrow \quad \cos M_d = 0 \Rightarrow M_d = \frac{1}{2}\pi \Rightarrow$ Median $= \pi/2.$

Example 3. ***A frequency function in the range (– 3, 3) is defined by***

$$y = (1/16)\,(3+x)^2, \quad -3 < x \le -1,$$
$$= (1/6)\,(6-2x^2), \quad -1 < x \le 1,$$
$$= (1/16)\,(3-x)^2, \quad 1 < x \le 3.$$

Find the mean and the standard deviation of the distribution defined by the above distribution.

Solution. The total frequency for the given distribution

$$= \int_{-3}^{-1} \frac{1}{16}(3+x)^2\,dx + \int_{-1}^{1} \frac{1}{16}(6-2x^2)\,dx + \int_{1}^{3} \frac{1}{16}(3-x)^2\,dx$$

$$= \frac{1}{16}\left[\frac{(3+x)^3}{3}\right]_{-3}^{-1} + \frac{1}{16}\left[6x - \frac{2}{3}x^3\right]_{-1}^{1} + \frac{1}{16}\left[\frac{-(3-x)^3}{3}\right]_{1}^{3}$$

$$= \frac{1}{16}\left(\frac{8}{3} + \frac{32}{3} + \frac{8}{3}\right) = 1.$$

Thus the given function is a probability density function. Now

$$\text{Mean} = \mu_1' = \int x\,f(x)\,dx$$

$$= \frac{1}{16}\left[\int_{-3}^{-1} x(3+x)^2\,dx + \int_{-1}^{1} x(6-2x^2)\,dx + \int_{1}^{3} x(3-x)^2\,dx\right]$$

$$= \frac{1}{16}\left\{\left[\frac{9}{2}x^2 + 2x^3 + \frac{1}{4}x^4\right]_{-3}^{-1} + \left[\frac{6}{2}x^2 - \frac{x^4}{2}\right]_{-1}^{1} + \left[\frac{9}{2}x^2 - 2x^3 + \frac{1}{4}x^4\right]_{1}^{3}\right\}$$

$$= \frac{1}{16}\left\{\left[\left(\frac{9}{2} - 2 + \frac{1}{4}\right) - \left(\frac{81}{2} - 54 + \frac{81}{4}\right)\right] + 0 + \left[\left(\frac{81}{2} - 54 + \frac{81}{4}\right) - \left(\frac{9}{2} - 2 + \frac{1}{4}\right)\right]\right\} = 0$$

and $$\mu_2' = \int x^2 f(x)\,dx = \frac{1}{16}\left[\int_{-3}^{-1} x^2(3+x)^2\,dx + \int_{-1}^{1} x^2(6-2x^2)\,dx + \int_{-3}^{3} x^2(3-x)^2\,dx\right]$$

$= 1$, (on simplifying).

$\therefore \quad \sigma^2 = \mu_2' - \mu_1'^2 = 1 - 0 = 1 \Rightarrow \sigma = 1.$

Example 4. ***For the distribution $dF = dx/2a$, $-a \le x \le a$, find μ_2, μ_3 and μ_4.***

Solution. By definition,

$$\mu_r' \text{ (about the origin)} = \int_{-a}^{a} (x-0)^r \frac{dx}{2a}$$

$$= \left[\frac{x^{r+1}}{2a(r+1)}\right]_{-a}^{a} = \frac{a^r}{2(r+1)} - \frac{(-1)^{r+1} a^r}{2(r+1)}$$

$$= \frac{a^r}{2(r+1)} \, [1-(-1)^{r+1}]. \quad ...(1)$$

Substituting $r = 1, 2, 3, 4$ successively in (1), we have

$$\mu_1' = 0, \mu_2' = \frac{1}{3}a^2, \mu_3' = 0, \mu_4' = \frac{1}{5}a^4$$

$$\therefore \quad \mu_2 = \mu_2' - \mu_1'^2 = \frac{1}{3}a^2 - 0 = \frac{1}{3}a^2$$

$$\mu_3 = \mu_3' - 3\mu_2'\mu_1' + 2\mu_1'^3 = 0 - 0 + 0 = 0$$

$$\mu_4 = \mu_4' - 4\mu_3'\mu_1' + 6\mu_2'\mu_1'^2 - 3\mu_1'^4 = \frac{1}{5}a^4 - 0 + 0 - 0 = \frac{1}{5}a^4.$$

Example 5. ***For the probability distribution***

$$dF = y_0 e^{-|x|} dx, 0 < x < \infty$$

show that $y_0 = \frac{1}{2}$, $\mu_1' = 0$, $\sigma = \sqrt{2}$ ***and mean deviation about mean is*** **1.**

Solution. By property of p. d. f., we have

$$y_0 \int_{-\infty}^{\infty} e^{-|x|} dx = 1$$

or $$2y_0 \int_0^{\infty} e^{-x} dx = 1$$ [$\because e^{-|x|}$ is an even function and $|x| = x$ for $0 \le x < \infty$]

or $$2y_0 = 1, \quad \therefore y_0 = \frac{1}{2}$$

Mean $= \mu_1'$ (about the origin)

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

$$= \frac{1}{2}\int_{-\infty}^{\infty} xe^{-|x|} dx = 0, \text{ since } xe^{-|x|}, \text{ is an odd function of } x.$$

μ_2' (about the origin)

$$= \int_{-\infty}^{\infty} x^2 f(x) dx = \frac{1}{2}\int_{-\infty}^{\infty} x^2 e^{-|x|} dx$$

$$= \int_0^{\infty} x^2 e^{-|x|} dx, \text{ since } x^2 e^{-|x|}, \text{ is an even function of } x$$

$$= \int_0^{\infty} x^2 e^{-x} dx \quad [\because |x| = x \text{ for } 0 \le x < \infty]$$

$$= \Gamma(3) = 2! = 2.$$

$$\therefore \quad \sigma^2 = \mu_2 = \mu_2' - \mu_1'^2 = 2 - 0 = 2 \Rightarrow \sigma = \sqrt{2}.$$

Mean deviation about the mean

$$= \int_{-\infty}^{\infty} |x - \text{Mean}| f(x) dx$$

$$= \frac{1}{2}\int_{-\infty}^{\infty} |x|\ e^{-|x|}\ dx \qquad [\because \text{Mean} = 0]$$

$$= \int_0^{\infty} xe^{-x}\ dx = \Gamma(2) = 1.$$

Example 6. ***Show that the geometric mean, G of the distribution $dP = 6\,(2 - x)\,(x - 1)\,dx,\ 1 \le x \le 2$ is given by $16\,G = e^{19/6}$.***

Solution. Here $f(x) = 6\,(2 - x)\,(x - 1)$ is p.d.f., since

$$\int_1^2 6\,(2 - x)\,(x - 1)\,dx = 1.$$

Now for the geometric mean G, we have

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

$$= 6\left[\log x.\left(-2x + \frac{3x^2}{2} - \frac{x^3}{3}\right)\right]_1^2 - 6\int_1^2 \frac{1}{x}\left(-2x + \frac{3x^2}{2} - \frac{x^3}{3}\right)dx$$

[Integrating by parts taking $\log x$ as first function]

$$= -4\log 2 - 6\left[-2x + \frac{3}{4}x^2 - \frac{x^3}{9}\right]_1^2 = -4\log 2 - (19/6)$$

$\therefore \quad \log G + \log 16 = 19/6$ or $16G = e^{19/6}$.

Example 7. ***For the distribution $dF = \sin x\,dx,\ 0 \le x \le \pi/2$, find (a) Mode and Median, (b) Mean and Variance.***

Solution. Here $f(x) = \sin x,\ 0 \le x \le \frac{\pi}{2}$.

(a) For Mode. $f'(x) = 0$ and $f''(x) < 0$

$$f'(x) = 0 \Rightarrow \cos x = 0 \Rightarrow x = \pi/2$$

and $\quad [f''(x)]_{x=\pi/2} = [-\sin x]_{x=\pi/2} = -1 < 0.$

Hence mode $= \pi/2$.

Let M_d be median, then

$$\int_0^{M_d} \sin x\,dx = \frac{1}{2} \Rightarrow [-\cos x]_0^{M_d} = \frac{1}{2}$$

$$\Rightarrow 1 - \cos M_d = \frac{1}{2}$$

$$\Rightarrow \cos M_d = \frac{1}{2} \Rightarrow M_d = \pi/3.$$

$\therefore \quad$ Median $M_d = \pi/3$.

(b) Mean $= \mu_1' = \int_0^{\pi/2} (x - 0)\,f(x)\,dx$

$$= \int_0^{\pi/2} x\sin x\,dx = [-x\cos x]_0^{\pi/2} + \int_0^{\pi/2} \cos x\,dx$$

$$= 0 + [\sin x]_0^{\pi/2} = 1$$

and $\quad$ variance $= \mu_2 = \int_0^{\pi/2} (x-1)^2 \sin x \, dx$

$$= [-(x-1)^2 \cos x]_0^{\pi/2} + 2\int_0^{\pi/2} (x-1)\cos x \, dx$$

$$= 1 + 2\,[(x-1)\sin x]_0^{\pi/2} - 2\int_0^{\pi/2} \sin x \, dx$$

$$= 1 + 2\left(\frac{\pi}{2} - 1\right) - 2\,[-\cos x]_0^{\pi/2}$$

$$= 1 + \pi - 2 - 2 = \pi - 3.$$

Example 8. ***Determine m so that the following function represents the density function :***

$$f(x) = \begin{cases} 0 & , \quad x \le -1 \\ m\,(x+1) & , \quad -1 < x \le 3 \\ 4m & , \quad 3 < x \le 4 \\ 0 & , \quad x > 4. \end{cases}$$

Find the value of x about which the mean deviation of this distribution is least.

Solution. By property of p. d. f. , we have

$$\int_{-\infty}^{-1} 0\,dx + \int_{-1}^{3} m\,(x+1)\,dx + \int_3^4 4m\,dx + \int_4^{\infty} 0\,dx = 1$$

$$\Rightarrow \quad m\int_{-1}^{3} (x+1)\,dx + 4m\int_3^4 dx = 1$$

$$\Rightarrow \quad m\left[\frac{x^2}{2} + x\right]_{-1}^{3} + 4m\,[x]_3^4 = 1$$

$$\Rightarrow \quad 8m + 4m = 1 \Rightarrow m = 1/12.$$

We know that mean deviation is least about the median and therefore we shall find median. Let M_d be the median, then

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

Now there may be two possibilities namely.

(i) $M_d > 3$ or

(ii) M_d lies between -1 and 3.

Case I. Where $M_d > 3$

$$m\int_{-1}^{3} (x+1)\,dx + 4m\int_3^{M_d} dx = \frac{1}{2}$$

i.e., $\quad 8m + 4m\,(M_d - 3) = \frac{1}{2}$

or $\quad \frac{1}{3}(M_d - 3) = \frac{1}{2} - \frac{2}{3} = -\frac{1}{6} = -$ ve quantity.

Hence $M_d \not> 3$.

Case II. When M_d, lies between -1 and 3.

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

$$\Rightarrow \quad (M_d + 1)^2 = 12 \Rightarrow M_d = 2\sqrt{3} - 1.$$

Example 9. ***A random variable has the probability law***

$$dF = \frac{x}{a^2} e^{-x^2/2a^2}\, dx, \quad 0 \le x < \infty.$$

Find the distance between the quartiles and obtain standard deviation.

Solution. Let Q_1 and Q_3 be the 1st and 3rd quartiles respectively.

Then
$$\frac{1}{a^2}\int_0^{Q_1} x e^{-x^2/2a^2}\, dx = \frac{1}{4}$$

i.e.,
$$1 - e^{-Q_1^2/2a^2} = \frac{1}{4}$$

$$\therefore \quad Q_1 = \sqrt{2}\, a \sqrt{\{\log(4/3)\}}$$

and
$$\frac{1}{a^2}\int_0^{Q_3} x e^{-x^2/2a^2}\, dx = \frac{3}{4}$$

$$\Rightarrow \quad Q_3 = \sqrt{2}\, a \sqrt{\log 4}.$$

$\therefore$ The distance between the quartiles $= Q_3 - Q_1$

$$= \sqrt{2}a\,[\sqrt{\log 4} - \sqrt{\{\log(4/3)\}}].$$

Now
$$\mu_1' = \int_0^\infty (x - 0) \cdot f(x)\, dx$$

$$= \int_0^\infty x \cdot \frac{x}{a^2} e^{-x^2/2a^2}\, dx$$

$$= \sqrt{2}\, a \int_0^\infty e^{-y}\, y^{3/2 - 1}\, dy$$

[Putting $x^2/2a^2 = y$ and $(x/a^2)\, dx = dy$]

$$= \sqrt{2}\, a\, \Gamma\left(\frac{3}{2}\right) = a \cdot \sqrt{\pi/2}$$

$$\mu_2' = \int_0^\infty x^2 \cdot \frac{x}{a^2} e^{-x^2/2a^2}\, dx$$

$$= 2a^2 \int_0^\infty y e^{-y}\, dy \quad \text{[Putting } x^2/2a^2 = y \text{ and } (x/a^2)\, dx = dy]$$

$$= 2a^2\, \Gamma(2) = 2a^2$$

$$\therefore \quad \text{Variance} = \sigma^2 = \mu_2 = \mu_2' - \mu_1'^2 = 2a^2 - a^2(\pi/2) = a^2\left(2 - \frac{1}{2}\pi\right)$$

and
$$\text{S. D.} = \sigma = a\sqrt{\left\{\left(2 - \frac{1}{2}\pi\right)\right\}}.$$

Example 10. ***Show that for distribution***

$$df = y_0\, e^{-x/\sigma}\, dx,\ 0 \le x < \infty$$

the mean and the standard deviation are both equal to σ.

Also show that $\mu_3 = 2\sigma^3$; $\mu_4 = 9\sigma^4, \beta_1 = 4$ ***and*** $\beta_2 = 9$. ***Find the inter-quartile range.***

Solution. Since df is a p. d. f., so y_0 is obtained by the following relation

$$y_0 \int_0^\infty e^{-x/\sigma}\, dx = 1 \Rightarrow y_0 \left[-\sigma e^{-x/\sigma}\right]_0^\infty = 1 \Rightarrow y_0[-\sigma(0-1)] = 1$$

$$\Rightarrow \qquad y_0 = 1/\sigma.$$

Now, 1st moment about the origin

$$\mu_1' = \frac{1}{\sigma}\int_0^\infty x e^{-x/\sigma}\, dx$$

$$= \frac{1}{\sigma}\left\{\left[x\left(-\sigma e^{-x/\sigma}\right)\right]_0^\infty - \int_0^\infty 1.(-\sigma e^{-x/\sigma})\, dx\right\}$$

$$= \int_0^\infty e^{-x/\sigma}\, dx = \left[-\sigma e^{-x/\sigma}\right]_0^\infty = \sigma$$

2nd moment about the origin

$$\mu_2' = \frac{1}{\sigma}\int_0^\infty x^2 e^{-x/\sigma}\, dx$$

$$= \frac{1}{\sigma}\left\{\left[x^2.\left(-\sigma e^{-x/\sigma}\right)\right]_0^\infty \int_0^\infty 2x.(-\sigma e^{-x/\sigma})\, dx\right\}$$

$$= 2\int_0^\infty x e^{-x/\sigma}\, dx$$

$$= 2\left\{\left[x.\left(-\sigma e^{-x/\sigma}\right)\right]_0^\infty - \int_0^\infty 1.\left(-\sigma e^{-x/\sigma}\right) dx\right\}$$

$$= 2\sigma \int_0^\infty e^{-x/\sigma}\, dx = 2\sigma\left[-\sigma e^{-x/\sigma}\right]_0^\infty = 2\sigma^2$$

and $\quad \mu_3' = \dfrac{1}{\sigma}\displaystyle\int_0^\infty x^3 e^{-x/\sigma}\, dx = 6\sigma^3$, (on simplifying).

Similarly, $\mu_4' = \dfrac{1}{\sigma}\displaystyle\int_0^\infty x^4 e^{-x/\sigma}\, dx = 24\sigma^4$

$\therefore \quad$ Variance $= \mu_2 = \mu_2' - \mu_1'^2 = 2\sigma^2 - \sigma^2 = \sigma^2$

$$\mu_3 = \mu_3' - 3\mu_2'\mu_1' + 2\mu_1'^3 = 6\sigma^3 - 3.\,2\sigma^2.\sigma + 2\sigma^3 = 2\sigma^3$$

$$\mu_4 = \mu_4' - 4\mu_3'\mu_1' + 6\mu_2'\mu_1'^2 - 3\mu_1'^4$$

$$= 24\sigma^4 - 4.\,6\sigma^3.\sigma + 6.\,2\sigma^2.\sigma^2 - 3\sigma^4 = 9\sigma^4.$$

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

Now Q_1 is given by

$$\int_0^{Q_1} \frac{1}{\sigma} e^{-x/\sigma} \, dx = \frac{1}{4}$$

$$\Rightarrow \quad -e^{-Q_1/\sigma} + 1 = \frac{1}{4}; \quad \therefore \; e^{-Q_1/\sigma} = \frac{3}{4}.$$

Similarly Q_3 is given by

$$\int_0^{Q_3} \frac{1}{\sigma} e^{-x/\sigma} \, dx = \frac{3}{4}$$

$$\Rightarrow \quad -e^{-Q_3/\sigma} + 1 = \frac{3}{4}; \qquad \therefore \; e^{-Q_3/\sigma} = \frac{1}{4}.$$

Now $\qquad e^{-Q_1/\sigma} \; e^{-Q_3/\sigma} = \frac{3}{4} \; \frac{1}{4}$

i.e., $\qquad e^{(Q_3 - Q_1)/\sigma} = 3 \Rightarrow Q_3 - Q_1 = \sigma \log 3.$

Example 11. ***For the symmetric distribution***

$$f(x) = \frac{2a}{\pi}\left(\frac{1}{a^2 + x^2}\right), \; -a \le x \le a$$

show that $\qquad \mu_2 = \dfrac{a^2 (4-\pi)}{\pi}$ ***and*** $\mu_4 = a^4\left(1 - \dfrac{8}{3\pi}\right).$

Solution. Since the distribution is symmetric in the range $(-a, a)$ therefore mean is 0. Also

$$\mu_1' = \frac{2a}{\pi} \int_{-a}^{a} x \cdot \frac{1}{a^2 + x^2} \, dx = 0,$$

since $\dfrac{x}{a^2 + x^2}$, is an odd function of x.

$$\mu_2' = \frac{2a}{\pi} \int_{-a}^{a} x^2 \cdot \frac{1}{a^2 + x^2} \, dx$$

$$= \frac{2a}{\pi} \int_{-a}^{a} \left(1 - \frac{a^2}{a^2 + x^2}\right) dx = \frac{4a}{\pi} \int_0^a \left(1 - \frac{a^2}{a^2 + x^2}\right) dx$$

$$= \frac{4a}{\pi}\left[x - a \tan^{-1}\frac{x}{a}\right]_{-a}^{a} = \frac{a^2}{\pi}(4 - \pi)$$

$$\therefore \qquad \mu_2 = \mu_2' - \mu_1'^2 = (a^2/\pi)(4 - \pi)$$

Now, $\qquad \mu_3' = \dfrac{2a}{\pi} \displaystyle\int_{-a}^{a} x^3 \cdot \frac{1}{a^2 + x^2} \, dx = 0$

$$\left[\text{since } \frac{x^2}{a^2 + x^2} \text{ is an odd function of } x\right]$$

and $\qquad \mu_4' = \dfrac{2a}{\pi} \displaystyle\int_{-a}^{a} x^4 \cdot \frac{1}{a^2 + x^2} \, dx$

$$= \frac{4a}{\pi}\int_0^a \frac{x^4}{a^2+x^2}\,dx = \frac{4a}{\pi}\int_0^a \left[x^2 - a^2 + \frac{a^4}{a^2+x^2}\right] dx$$

$$= \frac{4a}{\pi}\left[\frac{x^3}{3} - a^2x + a^3 \tan^{-1}\frac{x}{a}\right] = a^4\left(1 - \frac{8}{3\pi}\right).$$

Example 12. ***In a continuous distribution, whose probability density function is given by***

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

show that the distribution is symmetrical, with mean 1 and variance 1/5. Show that the second and third moments about x = 0 are 6/5 and 8/6. Show further that the skewness is zero and mean deviation about the mean is 3/8. Also show that for this distribution $\mu_{2n+1} = 0$. ***Deduce*** β_2.

Solution. Since given p. d. f. is

$$f(x) = \frac{3}{4}x(2-x), \quad 0 \le x \le 2.$$

$\therefore$ $$\text{Mean} = \mu_1' = \frac{3}{4}\int_0^2 x \,.\, x(2-x)\,dx = \frac{3}{4}\left[\frac{2x^3}{3} - \frac{x^4}{4}\right]_0^2 = 1.$$

Median M_d is given by the relation

$$\int_0^{M_d} f(x)\,dx = \frac{1}{2}$$

i.e., $$\frac{3}{4}\int_0^{M_d} x(2-x)\,dx = \frac{1}{2} \Rightarrow \frac{3}{4}\left[M_d^2 - \frac{M_d^3}{3}\right] = \frac{1}{2}$$

$\Rightarrow$ $M_d^3 - 3M_d^2 + 2 = 0$

$\Rightarrow$ $(M_d - 1)(M_d^2 - 2M_d - 2) = 0$

$\Rightarrow$ $M_d = 1, 2 \pm \sqrt{3}$ out of which only rational value $1 \in [0, 2]$.

$\therefore$ Median $M_d = 1$.

For mode, $f'(x) = 0$ and $f''(x) < 0$.

So $f'(x) = 0 \Rightarrow 1 - x = 0; \Rightarrow x = 1$

Also $[f''(x)]_{x=1} = -\frac{3}{2} < 0.$

$\therefore$ Mode $M_0 = 1.$

Since mean, median and mode each is equal to 1, therefore the distribution is symmetric.

Now $$\mu_2' = \frac{3}{4}\int_0^2 x^2 \,.\, x(2-x)\,dx = \frac{3}{4}\left[\frac{x^4}{2} - \frac{x^5}{5}\right]_0^2 = \frac{6}{5}\,\therefore$$

$\therefore$ Variance $= \mu_2 = \mu_2' - \mu_1'^2 = 6/5 - 1 = 1/5$

$\Rightarrow$ $\sigma = \sqrt{\mu_2} = 1/\sqrt{5}.$

And 3rd moment μ_3' about $x = 0$

$$= \frac{3}{4}\int_0^2 x^3 \,.\, x(2-x)\,dx = \frac{3}{4}\left[\frac{2}{5}x^5 - \frac{x^6}{6}\right]_0^2 = \frac{8}{5}$$

$\therefore$ Skewness $= \dfrac{3\,(\text{mean} - \text{mode})}{\sigma} = \dfrac{3\,(1-1)}{1/\sqrt{5}} = 0.$

Again mean deviation about mean

$$= \frac{3}{4}\int_0^2 |x-1|\,.\,x(2-x)\,dx$$

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

$$= (3/16) + (3/16) = 3/8.$$

Also $\mu_{2n+1} = \int_0^2 (x-1)^{2n+1} f(x)\,dx$

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

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

$$= \frac{3}{4}\left[(2x - x^2).\,\frac{(x-1)^{2n+2}}{2n+2}\right]_0^2 - \frac{3}{4}\int_0^2 (2-2x)\,.\,\frac{(x-1)^{2n+2}}{2n+2}\,dx,$$

[By integrating by parts]

$$= 0 - \frac{3}{2(2n+2)}\int_0^2 (1-x)(x-1)^{2n+2}\,dx$$

$$= \frac{3}{2(2n+2)}\int_0^2 (x-1)^{2n+3}\,dx$$

$$= \frac{3}{2(2n+2)}\left[\frac{(x-1)^{2n+4}}{2n+4}\right]_0^2$$

$$= \frac{3}{2(2n+2)(2n+4)}\left[(1)^{2n+4} - (-1)^{2n+4}\right] = 0$$

$$\mu_4 = \frac{3}{4}\int_0^2 (x-1)^4\,.\,x(2-x)\,dx = \frac{3}{35}.$$

$$\therefore \quad \beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{3/35}{1/25} = \frac{15}{7}.$$

Example 13. ***For the Beta distribution***

$$dF = \frac{1}{\beta(m, n)}\, x^{n-1}(1-x)^{m-1}\,dx, \quad 0 \le x \le 1,\ m > 0$$

where $\quad \beta(m, n) = \int_0^1 x^{n-1}(1-x)^{m-1}\,dx,$

find the mean, the standard deviation, harmonic mean and μ_r'.

Solution. Since

$$\int_0^1 \frac{1}{\beta(m, n)}\, x^{n-1}(1-x)^{m-1}\,dx = \frac{\beta(m, n)}{\beta(m, n)} = 1$$

$$\therefore \qquad f(x) = \frac{1}{\beta(m, n)} x^{n-1} (1-x)^{m-1}$$

is clearly p.d.f.

The arithmetic mean, M, is given by the following relation :

$$M = \mu_1' = \int_0^1 x \, f(x) \, dx$$

$$= \frac{1}{\beta(m, n)} \int_0^1 x \, . \, x^{n-1} (1-x)^{m-1} \, dx$$

$$= \frac{1}{\beta(m, n)} \int_0^1 x^n (1-x)^{m-1} \, dx$$

$$= \frac{\beta(m, n+1)}{\beta(m, n)} = \frac{\Gamma(m)\,\Gamma(n+1)}{\Gamma(m+n+1)} \bigg/ \frac{\Gamma(m)\,\Gamma(n)}{\Gamma(m+n)} = \frac{n}{m+n}.$$

Now $\mu_2' = \int_0^1 x^2 \, f(x) \, dx$

$$= \frac{1}{\beta(m, n)} \int_0^1 x^2 \, . \, x^{n-1} (1-x)^{m-1} \, dx$$

$$= \frac{1}{\beta(m, n)} \int_0^1 x^{n+1} (1-x)^{m-1} \, dx$$

$$= \frac{\beta(m, n+2)}{\beta(m, n)} = \frac{\Gamma(m)\,\Gamma(n+2)}{\Gamma(m+n+2)} \bigg/ \frac{\Gamma(m)\,\Gamma(n)}{\Gamma(m+n)}$$

$$= \frac{n\,(n+1)}{(m+n+1)\,(m+n)}.$$

$$\therefore \qquad \sigma = \sqrt{(\mu_2)} = \sqrt{(\mu_2' - \mu_1'^2)}$$

$$= \sqrt{\left[\frac{n\,(n+1)}{(m+n+1)\,(m+n)} - \left(\frac{n}{m+n}\right)^2\right]}$$

$$= \frac{1}{m+n} \sqrt{\left(\frac{n\,(n+1)\,(m+n) - n^2\,(m+n+1)}{m+n+1}\right)}$$

$$= \frac{1}{m+n} \sqrt{\left(\frac{mn}{m+n+1}\right)}.$$

The harmonic mean H, is given by the following relation :

$$\frac{1}{H} = \frac{1}{\beta(m, n)} \int_0^1 \frac{1}{x} (1-x)^{m-1} x^{n-1} \, dx$$

$$= \frac{1}{\beta(m, n)} \int_0^1 (1-x)^{m-1} x^{n-2} \, dx$$

$$= \frac{\beta(m, n-1)}{\beta(m, n)} = \frac{\Gamma(m)\,\Gamma(n-1)}{\Gamma(m+n-1)} \bigg/ \frac{\Gamma(m)\,\Gamma(n)}{\Gamma(m+n)}$$

$$= \frac{m+n-1}{n-1}.$$

$$\therefore \qquad H = \frac{n-1}{m+n-1}.$$

Lastly $\mu'_r = \int_0^1 \frac{x^{n+r-1}(1-x)^{m-1}}{\beta(m, n)} dx = \frac{\beta(n+r, m)}{\beta(m, n)}$

$$= \frac{\Gamma(n+r)\,\Gamma(m+n)}{\Gamma(m+n+r)\,\Gamma(n)} = \frac{n\,(n+1)\,(n+2)\ldots(n+r-1)}{(n+m)\,(n+m+1)\ldots(n+m+r-1)}.$$

Example 14. ***Prove that for the Beta distribution of the second kind***

$$dP = \frac{1}{\beta(m, n)} \cdot \frac{x^{m-1}}{(1+x)^{m+n}} dx, \quad 0 \le x < \infty, n > 2,$$

variance is $\frac{m\,(m+n-1)}{(n-1)^2\,(n-2)}$ ***and harmonic mean is*** $\frac{m-1}{n}$.

Solution. By property of p. d. f., we have

$$\int_0^\infty \frac{1}{\beta(m, n)} \cdot \frac{x^{m-1}}{(1+x)^{m+n}} dx = 1$$

$$\Rightarrow \qquad \int_0^\infty \frac{x^{m-1}}{(1+x)^{m+n}} dx = \beta(m, n). \qquad \ldots(1)$$

Now, Mean $M = \mu_1' = \frac{1}{\beta(m, n)} \int_0^\infty x \cdot \frac{x^{m-1}}{(1+x)^{m+n}} dx$

$$= \frac{1}{\beta(m, n)} \int_0^\infty \frac{x^m}{(1+x)^{m+n}} dx$$

$$= \frac{\beta(m+1, n-1)}{\beta(m, n)}$$

[In view of (1), replacing m by $m+1$ and n by $n-1$ in (1)]

$$= \frac{\Gamma(m+1)\,\Gamma(n-1)}{\Gamma(m+n)} \cdot \frac{\Gamma(m)\,\Gamma(n)}{\Gamma(m+n)} = \frac{m}{n-1}$$

$$\mu_2' = \frac{1}{\beta(m, n)} \int_0^\infty x^2 \cdot \frac{x^{m-1}}{(1+x)^{m+n}} dx$$

$$= \frac{1}{\beta(m, n)} \int_0^\infty \frac{x^{m+1}}{(1+x)^{m+n}} dx$$

$$= \frac{\beta(m+2, n-2)}{\beta(m, n)}$$

[in view of (1), replacing m by $m+2$ and n by $n-2$ in (1)]

$$= \frac{\Gamma(m+2)\,\Gamma(n-2)}{\Gamma(m+n)} \cdot \frac{\Gamma(m)\,\Gamma(n)}{\Gamma(m+n)} = \frac{m\,(m+1)}{(n-1)\,(n-2)}.$$

$$\therefore \quad \mu_2 = \mu_2' - \mu_1'^2$$

$$= \frac{m\,(m+1)}{(n-1)\,(n-2)} - \frac{m^2}{(n-1)^2} = \frac{m\,(m+n-1)}{(n-1)^2\,(n-2)}.$$

The harmonic mean H, is given by the following relation

$$\frac{1}{H} = \frac{1}{\beta(m, n)} \int_0^\infty \frac{1}{x} \cdot \frac{x^{m-1}}{(1+x)^{m+n}} dx$$

$$= \frac{1}{\beta(m, n)} \int_0^\infty \frac{x^{m-2}}{(1+x)^{m+n}} dx = \frac{\beta(m-1, n+1)}{\beta(m, n)}$$

$$= \frac{\Gamma(m-1)\,\Gamma(n+1)}{\Gamma(m+n)} \cdot \frac{\Gamma(m)\,\Gamma(n)}{\Gamma(m+n)} = \frac{n}{m-1}.$$

$$\therefore \qquad H = \frac{m-1}{n}.$$

EXERCISE 2 (B)

1. For the rectangular distribution

$$dP = dx, 1 \le x \le 2$$

show that A. M. > G. M. > H. M.

[**Remark** : Whenever $f(x)$ is constant throughout its interval, the variate is said to have a rectangular distribution of same probability].

2. Show that for a rectangular distribution

$$dP = dx, \ 0 \le x \le 1$$

μ_1' (about the origin) $= \frac{1}{2}$ and $\mu_2 = \frac{1}{12}$.

3. Find the arithmetic mean and median for the distribution

$$dP = 2x\, dx, \ 0 \le x \le 1.$$

4. If $f(x) = e^{-x}, 0 \le x < \infty$, find the mean and variance and the third moment about mean.

5. If $f(x) = e^{-x}, 0 \le x < \infty$, show that it a probability function and hence compute mean, μ_2, μ_3, μ_4.

6. A frequency distribution is defined by

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

$$= (2-x)^3 \quad , \quad 1 \le x \le 2.$$

Find the mean, the standard deviation and the mean deviation about the mean.

7. For the continuous distribution

$$dF = y_0\,(x - x^2)\, dx, \ 0 \le x \le 1$$

where y_0 is a constant, find A. M., H. M., median and mode. Is it a symmetric distribution ?

[**Hint.** Proceed as example 1]

8. For the probability distribution defined by

$$2m\, x\, e^{mx^2}, \ 0 \le x < \infty$$

show that the quartile deviation is $\frac{0{\cdot}06}{\sqrt{m}}$.

9. For the probability density function with density

$$\frac{1}{\pi} \cdot \frac{1}{1+x^2}, \ -\infty < x < \infty$$

find the following :

(a) Mean,

(b) Mode,

(c) Median,

(d) First and third quartiles.

10. For the probability density function

$$f(x) = \frac{2(a+x)}{b(a+b)}, -b \le x < 0$$

$$= \frac{2(a-x)}{a(a+b)}, 0 \le x \le a$$

Show that mean $= \frac{1}{3}(a-b)$, Variance $= \frac{1}{12}(a^2 + b^2 + ab)$.

$$\text{Median} = a - \sqrt{\left[\frac{1}{2} a (a+b)\right]}.$$

If terms of order $\left(\frac{a-b}{a}\right)^n$ are neglected, show that

$$\text{Mean—Median} = \frac{1}{4}(\text{Mean} - \text{Mode}).$$

11. Find the variance of the distribution defined by the density function

$$f(x) = (1/\pi)\, x \sin x, \ 0 \le x \le \pi.$$

12. The elementary probability law of a continuous random variable is

$$f(x) = y_0 \, e^{-b(x-a)}, \ a \le x < \infty$$

where a, b and y_0 are constants. Show that

$$y_0 = b = 1/\sigma \ \text{ and } \ a = m - \sigma$$

where m and σ are respectively the mean and standard deviation of the distribution. Show also that $\beta_1 = 4$ and $\beta_2 = 9$.

13. For the distribution $dF = kx\, e^{-x}\, dx$, $0 < x < \infty$ show that $\beta_2 = 2$.

14. Show that the mean, variance and the coefficients β_1, β_2 of the distribution

$$dF = kx^2 \, e^{-x} \, dx, 0 < x < \infty$$

are respectively 3, 3, 4/3 and 5.

15. If probability density function $f(x)$ is given by

$$f(x) = a\, e^{-ax}, 0 < x < \infty$$

where a is a positive constant, show that the mean is $1/a$ and the variance is $1/a^3$. Also show that the second and third moments about $x = 0$ are $2/a^2$ and $6/a^3$ respectively and that $\mu_3 = 3/a^3$.

16. Given that $f(x) = cx^{n-1} e^{-x}, 0 < x < \infty$ is a probability density function, show that

$$2\mu_4 \, \sigma^5 - 3\mu_3^3 - 6\sigma^9 = 0.$$

17. Determine the mode and median for the curve

$$y = \frac{ab\, x^{a-1}}{(1 + bx^a)^2}, b > a, a > 1, 0 < x < \infty$$

[Hint. Put $bx^a = y$]

18. A continuous distribution function $F(x)$ is defined as follows :

$$F(x) = \begin{cases} 0 & , \text{ for } x \le 2 \\ a\left(\frac{1}{2}x^2 + 3x - 8\right) & , \text{ for } 2 < x \le 8 \\ 1 & , \text{ for } x > 8 \end{cases}$$

Find the probability density function of x. Can you also find the mean of x.

19. Show that for the distribution

$$dp = \frac{1}{\Gamma m} e^{-x} x^{m-1} dx, \;\; 0 < x < \infty, m > 0$$

mean $= m$, variance $= m$,

harmonic mean $= m - 1$,

$$\mu_4 = 3m^2 + 6m, \beta_2 = 3 + (6/m).$$

20. The probability density function of a distribution with parameters r and A is defined as follows :

$$f(x) = rA^r \frac{1}{x^{r+1}}, \text{ for } x \ge A$$

$$= 0, \quad \text{for } x < A, r > 0.$$

Show that it has a finite nth moment if and only if $n < r$.

Also find mean and variance of the above distribution.

21. Show that for a symmetric distribution all moments of odd orders about the mean vanish.

ANSWERS

3. Mean = 2/3; Median = $1/\sqrt{2}$ **4.** 1 ; 1 ; 2 **5.** 1, 1, 29

6. Mean $= \frac{1}{2}$; S. D. $= \frac{1}{\sqrt{(15)}}$, mean deviation $= \frac{1}{5}$

7. Mean $= \frac{1}{2}$; H. M. $= \frac{1}{3}$; Median $= \frac{1}{2}$; Mode $= \frac{1}{2}$; Distribution is symmetric.

9. Mean $= \frac{1}{2}$; Mode $= 0$; Median $= 0, Q_1 = 1, Q_3 = 1$.

11. $\frac{2(\pi^2 - 8)}{\pi}$

17. Mode $= \left[\frac{a-1}{b(a+1)}\right]^{1/a}$; Median $= \left(\frac{1}{b}\right)^{1/a}$

18. $a = \frac{1}{126}$, yes.

20. Mean $= \frac{rA}{r-1}$ $(r > 1)$; Variance $= \frac{rA^2}{(r-2)(r-1)^2}$ $(r > 2)$.

◈ § 2.7. NORMAL DISTRIBUTION (CONTINUOUS PROBABILITY DISTRIBUTION)

The normal distribution was first discovered by De-Moivre (an English mathematician) in 1733. De-Moivre obtained this continuous distribution as a limiting case of the binomial distribution and applied it to the problems of game of chance. It was credited to Gauss (1809) who used the normal curve to describe the theory of accidental errors of measurements involved in the calculation of orbits of heavenly bodies.

Throughout two continuous centuries (18th and 19th) many efforts were made to develop a normal model as the underlying law which may govern all continuous random variables.

That is why the name **'normal'**.

Definition. A random variable X is said to have a normal distribution with parameters m and σ^2 if its density function is given by the probability law :

$$f(x) = \frac{1}{\sigma\sqrt{(2\pi)}} e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2}, \quad -\infty < x < \infty, \sigma > 0, -\infty < m < \infty \qquad \text{...(A)}$$

where m is called 'mean' and σ^2 is called 'variance'.

Notation. A random variable X with mean 'm' and variance 'σ^2' and defined by the normal law [equation (A)] is denoted by $X \sim N(m, \sigma^2)$.

Theorem. *To derive normal distribution as a limiting case of binomial distribution where $p \neq q$ but $p \approx q$.*

Statement. *The limiting case of binomial distribution $(q + p)^n$, as $n \to \infty$ and neither p nor q are very small, generates the normal distribution.*

Proof. We know that the frequency of r successes in binomial distribution is given by

$$f(r) = N \,.\, {}^nC_r \; p^r q^{n-r}.$$

Now

$$f(r+1) = N \,.\, {}^nC_{r+1} \; p^{r+1} q^{n-(r+1)} \qquad \text{...(1)}$$

$$\frac{f(r+1)}{f(r)} = \frac{N \,.\, {}^nC_{r+1} \; p^{r+1} q^{n-r-1}}{N \,.\, {}^nC_r \; p^r q^{n-r}} = \frac{(n-r)}{(r+1)} \cdot \frac{p}{q}.$$

The frequency of r successes will be greater than the frequency of $(r+1)$ successes if

$$f(r) > f(r+1) \Rightarrow f(r+1)/f(r) < 1$$

$$\Rightarrow \{(n-r)\,p\}/\{(r+1)\,q\} < 1 \Rightarrow (n-r)\,p < (r+1)\,q$$

$$\Rightarrow r\,(p+q) > np - q \Rightarrow r > (np - q) \qquad [\because p + q = 1] \qquad \text{...(2)}$$

In a similar way the frequency of r successes will be greater than the frequency of $(r-1)$ successes if

$$r < np + p. \qquad \text{...(3)}$$

Therefore, in view of (2) and (3), the condition that the frequency $f(r)$ be maximum is given by

$$np - q < r < np + p. \qquad \text{...(4)}$$

Since a possible value of r is np, therefore, without loss of generality we can assume that np is an integer as $n \to \infty$. Hence the frequency of np successes can be assumed to be maximum frequency. Let y_0 be the frequency of np successes and y_x be the frequency of $(np + x)$ successes. Then

$$y_0 = f(np) = N \,.\, {}^nC_{np}\; p^{np} q^{n-np} \qquad \text{[from (1), for } r = np]$$

$$= N \frac{n!}{(np)!\,(nq)!} p^{np} q^{nq} \qquad [\because\ q = 1 - p]$$

and

$$y_x = N \,.\, \frac{n!}{(np + x)!\,(nq - x)!} p^{np + x} q^{nq - x}$$

$$\therefore \quad \frac{y_x}{y_0} = \frac{(np)!\,(nq)!}{(np + x)!\,(nq - x)!} p^x q^{-x}. \qquad \ldots(5)$$

If n be large, then using the approximation formula due to James Stirling, namely $n! = e^{-n} n^{n + 1/2} \sqrt{(2\pi)}$, we have from (5),

$$\frac{y_x}{y_0} = \frac{e^{-np} (np)^{np + 1/2} \sqrt{2\pi}\, e^{-nq} (nq)^{nq + 1/2} \sqrt{2\pi}\; p^x q^{-x}}{e^{-(np + x)} (np + x)^{np + x + 1/2} \sqrt{2\pi}\, e^{-(nq - x)} (nq - x)^{nqx - x + 1/2} \sqrt{2\pi}}$$

$$= \frac{(np)^{np + 1/2} (nq)^{nq + 1/2} (np/nq)^x}{(np)^{np + x + 1/2} \left\{1 + \dfrac{x}{np}\right\}^{np + x + 1/2} (np)^{nq - x + 1/2} \left\{1 - \dfrac{x}{nq}\right\}^{nq - x + 1/2}}$$

$$= \frac{1}{\left(1 + \dfrac{x}{np}\right)^{np + x + 1/2} \left(1 - \dfrac{x}{nq}\right)^{nq - x + 1/2}}$$

$$\therefore \quad \log \frac{y_x}{y_0} = -\left(np + x + \frac{1}{2}\right) \log\left(1 + \frac{x}{np}\right) - \left(nq - x + \frac{1}{2}\right) \log\left(1 - \frac{x}{nq}\right)$$

$$= -\left(np + x + \frac{1}{2}\right)\left(\frac{x}{np} - \frac{x^2}{2n^2p^2} + \frac{x^3}{3n^3p^3} - \ldots\right)$$

$$+ \left(nq - x + \frac{1}{2}\right)\left(\frac{x}{nq} + \frac{x^2}{2n^2q^2} + \frac{x^3}{3n^3q^3} + \ldots\right)$$

$$= x\left(-1 - \frac{1}{2np} + 1 + \frac{1}{2nq}\right)$$

$$+ x^2\left(\frac{1}{2np} - \frac{1}{np} + \frac{1}{4n^2p^2} + \frac{1}{2nq} - \frac{1}{nq} + \frac{1}{4n^2q^2}\right)$$

$$+ x^3\left(\frac{1}{3n^2q^2} + \frac{1}{6n^3q^3} - \frac{1}{2n^2q^2} - \frac{1}{2n^2p^2} - \frac{1}{3n^2q^2} + \frac{1}{6n^3p^3}\right) - \ldots$$

$$= \frac{p - q}{2npq} x + \frac{p^2 + q^2}{4n^2p^2q^2} x^2 - \frac{x^2}{2npq} + \ldots \text{ terms of higher orders.}$$

Neglecting terms containing $1/n^2$, we have

$$\therefore \quad \log \frac{y_x}{y_0} = -\frac{q - p}{2npq} x - \frac{x^2}{2npq}.$$

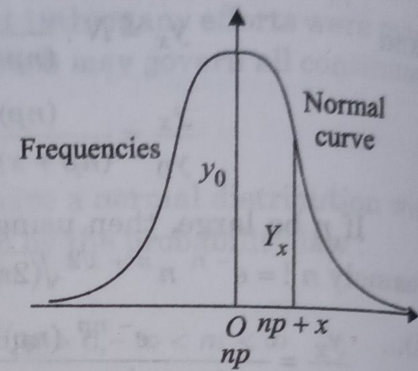
Since $p < 1, q < 1$ and so $q - p$ is very small as compared with n. Therefore 1st term may be neglected.

$$\therefore \quad \log \frac{y_x}{y_0} = -\frac{x^2}{2npq} = -\frac{x^2}{2\sigma^2}$$

$$[\because \sigma^2 = npq, \text{the variance of binomial distribution}]$$

$$\Rightarrow \quad y_x = y_0 e^{-x^2/2\sigma^2}.$$

Normal curve. *A curve is called normal curve if it is symmetrical about y-axis when ordinate has maximum value. Also mean, median and mode coincide in the normal curve.*



If the origin is taken at zero success, then the equation of the normal curve is given by

$$y_x = y_0 e^{-(x-m)^2/2\sigma^2}$$

where $m = np$, the mean of binomial distribution.

Exercise. Define binomial distribution and Normal distribution. Establish relation between them.

◈ § 2.8. A PARTICULAR CASE

Theorem. *To derive the normal distribution as a limiting case of binomial distribution when $p = q$.*

Proof. Let $N(q+p)^n$ be the binomial distribution. If $p = q$, then $p = q = 1/2$ [since $p + q = 1$] and consequently the binomial distribution is symmetrical. Without loss of generality, we assume that n is an even integer, say $2k$, k being an integer. Since $n \to \infty$, the frequencies of r and $r+1$ successes can be written in the following forms :

$$f(r) = N \,.\, {}^{2k}C_r \left(\frac{1}{2}\right)^{2k}, \quad f(r+1) = N \,.\, {}^{2k}C_{r+1} \left(\frac{1}{2}\right)^{2k}.$$

$$\therefore \quad \frac{f(r+1)}{f(r)} = \frac{{}^{2k}C_{r+1}}{{}^{2k}C_r} = \frac{(2k)!\,(r)!\,(2k-r)!}{(2k-r-1)!\,(r+1)!\,(2k)!} = \frac{2k-r}{r+1}.$$

The frequency of r successes will be greater than the frequency of $(r+1)$ successes if

$$f(r) > f(r+1) \Rightarrow f(r+1)/\,f(r) < 1$$

$$\Rightarrow \quad 2k - r < r + 1 \Rightarrow r > k - \frac{1}{2}. \quad \text{...(1)}$$

In a similar way the frequencies of r successes will be greater than the frequencies of $(r-1)$ successes if

$$r < k + \frac{1}{2}. \quad \text{...(2)}$$

In view of (1) and (2), we observe that if $k-\frac{1}{2}<r<k+\frac{1}{2}$, the frequency corresponding to r successes will be greatest. Clearly $r=k$ is the value of the success corresponding to which the frequency is maximum. Suppose it is y_0. Then we have

$$y_0 = N.\ {}^{2k}C_k\left(\frac{1}{2}\right)^{2k} = N.\frac{(2k)!}{(k)!(k)!}\left(\frac{1}{2}\right)^{2k}.$$

Let y_x be the frequency of $k+x$ successes; then we have

$$y_x = N\ {}^{2k}C_{k+x}\left(\frac{1}{2}\right)^{2k} = N\left(\frac{1}{2}\right)^{2k}\frac{(2k)!}{(k+x)!(k-x)!}$$

$$\therefore\quad \frac{y_x}{y_0} = \frac{k!k!}{(k+x)!(k-x)!} = \frac{k(k-1)(k-2)\ldots(k-x+1)}{(k+x)(k+x-1)\ldots(k+1)}$$

$$= \frac{\left(1-\frac{1}{k}\right)\left(1-\frac{2}{k}\right)\ldots\left\{1-\frac{x-1}{k}\right\}}{\left(1+\frac{1}{k}\right)\left(1+\frac{2}{k}\right)\ldots\left(1+\frac{x}{k}\right)}.$$

Take log of both sides (note that $1/k$, $2/k$, are small quantities as k is large)

$$\log\frac{y_x}{y_0} = \left[\log\left(1-\frac{1}{k}\right)+\log\left(1-\frac{2}{k}\right)+\ldots+\log\left(1-\frac{x-1}{k}\right)\right]$$

$$-\left[\log\left(1+\frac{1}{k}\right)+\log\left(1+\frac{2}{k}\right)+\ldots+\log\left(1+\frac{x}{k}\right)\right]. \quad \ldots(3)$$

Now writing expansion for each term and neglecting higher powers of x/k very small quantity as k is very large), we get from (3)

$$\log\frac{y_x}{y_0} = -\frac{1}{k}\{1+2+3+\ldots(x-1)\}-\frac{1}{k}\{1+2+3+\ldots+(x-1)+x\}$$

$$= -\frac{2}{k}\{1+2+3+\ldots+(x-1)\}-\frac{x}{k}$$

$$= -\frac{2}{k}\frac{(x-1)(x-1+1)}{2}-\frac{x}{k} \qquad \text{[use A.P.]}$$

$$= -\frac{x}{k}(x-1)-\frac{x}{k} = -\frac{x^2}{k}.$$

$$\therefore\quad y_x = y_0 e^{-x^2/k} \quad \text{or} \quad y_x = y_0 e^{-x^2/2\sigma^2} \quad \ldots(4)$$

where in binomial distribution,

$$\sigma^2 = npq = n.\frac{1}{2}.\frac{1}{2} = \frac{1}{4}n = \frac{1}{2}k. \qquad \left\{\because\ k=\frac{1}{2}n\right\}$$

It is Normal Distribution.

◈ § 2.9. STANDARD FORM OF THE NORMAL CURVE

Definition. Let the eqn. of the normal curve be

$$y_x = y_0 e^{-x^2/2\sigma^2} \quad \ldots(1)$$

where the origin is taken at the mean. The curve (1) is called **normal probability curve** if the value of y_0 is determined in such a way that the total frequency is 1 (unity).

Here $$1 = y_0 \int_{-\infty}^{\infty} e^{-x^2/2\sigma^2}\, dx$$

i.e., the total area bounded by the curve and the *x*-axis is unity,

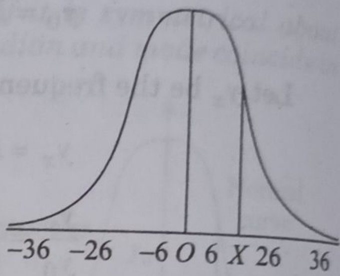
i.e., $$1 = 2y_0 \int_0^{\infty} e^{-x^2/2\sigma^2}\, dx \quad \text{[by definite integral property]}$$

Now substituting $x = \sigma\sqrt{2}t$ or $dx = \sigma\sqrt{2}dt$, we get

$$1 = 2y_0 \int_0^{\infty} e^{-t^2} \sigma\sqrt{2}dt = 2\sqrt{2}\sigma y_0 \int_0^{\infty} e^{-t^2}\, dt$$

$$= 2\sqrt{2}\sigma y_0 \left(\frac{1}{2}\sqrt{\pi}\right) \left[\because \int_0^{\infty} e^{-t^2}\, dt = \frac{1}{2}\sqrt{\pi}\right]$$

$\therefore \quad y_0 = 1/[\sigma\sqrt{(2\pi)}]$.



Substituting this value of y_0 in (1), the standard form of the normal curve is given by

$$y = \frac{1}{\sigma\sqrt{(2\pi)}} e^{-x^2/2\sigma^2}. \quad \text{...(2)}$$

Note 1. If the total frequency (*i.e.*, area inside the curve be *N*, then the corresponding normal curve is

$$y = \frac{N}{\sigma\sqrt{(2\pi)}} e^{-x^2/2\sigma^2}. \quad \text{...(3)}$$

Note 2. If the origin is taken at zero success, then the equation of normal curve (when total frequency is *N*) is

$$y = \frac{N}{\sigma\sqrt{(2\pi)}} e^{-(x-m)^2/2\sigma^2}. \quad \text{...(4)}$$

where $np = m =$ mean.

§ 2.10. PROPERTIES OF THE NORMAL DISTRIBUTION

Let the equation of the normal curve be

$$y = y_0 e^{-x^2/2\sigma^2} \quad \text{...(1)}$$

(I) The curve is symmetrical about y-axis.

Proof. Since the equation (1) contains even powers of *x*, so the curve is symmetrical about *y*-axis.

(II) The mean, median and mode coincide at the origin.

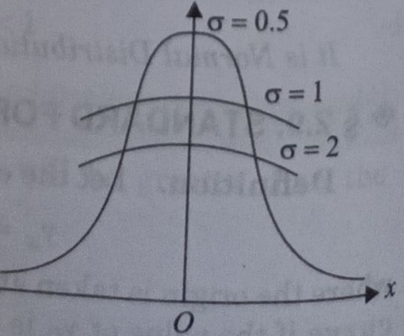
Proof. Mean $\mu_1' = \int_{-\infty}^{\infty} y_0 e^{-x^2/2\sigma^2} . x\, dx = y_0 \int_{-\infty}^{\infty} x e^{-x^2/2\sigma^2}\, dx = 0$,

as the integrand is an odd function of *x*.

Differentiating (1) w.r.t. '*x*', we get

$$\frac{dy}{dx} = -\frac{y_0 x}{\sigma^2} e^{-x^2/2\sigma^2}$$

$$\frac{d^2y}{dx^2} = -\frac{y_0}{\sigma^2}\left(1 - \frac{x^2}{\sigma^2}\right) e^{-x^2/2\sigma^2}.$$



Now $\quad dy/dx = 0 \Rightarrow x = 0$

and $\quad (d^2y/dx^2)_{x=0} = -\text{ve}$.

Therefore, *y* is maximum at $x = 0$.

Hence there is a mode at $x = 0$, *i.e.*, the mode is at origin. We know that median is a point which divides the whole area into two equal parts. Let x_1 be the median.

$$\therefore \int_{x_1}^{\infty} \frac{1}{\sigma\sqrt{(2\pi)}} e^{-x^2/2\sigma^2} dx = \frac{1}{2} \qquad \left[\because y_0 = \frac{1}{\sigma\sqrt{2\pi}}\right]$$

$$\Rightarrow \frac{1}{2} = \frac{1}{\sigma\sqrt{(2\pi)}} \int_{x_1}^{\infty} e^{-x^2/2\sigma^2} dx$$

$$\Rightarrow \frac{1}{2} = \frac{1}{\sigma\sqrt{(2\pi)}} \int_{x_1\sigma\sqrt{2}}^{\infty} e^{-t^2} . \sigma\sqrt{2}\, dt \text{ [substituting } x = \sigma\sqrt{2}t, \ \therefore dx = \sigma\sqrt{2}\, dt]$$

$$\Rightarrow \frac{1}{2} = \frac{1}{\sqrt{\pi}} \int_{x_1/\sigma\sqrt{2}}^{\infty} e^{-t^2} dt \Rightarrow \int_{x_1/\sigma\sqrt{2}}^{\infty} e^{-t^2} dt = \frac{1}{2}\sqrt{\pi}$$

$$\Rightarrow \int_{x_1/\sigma\sqrt{2}}^{\infty} e^{-t^2} dt = \int_{0}^{\infty} e^{-t^2} dt \Rightarrow \frac{x_1}{\sigma\sqrt{2}} = 0 \Rightarrow x_1 = 0.$$

$\therefore$ This median is at $x_1 = 0$, *i.e.*, at origin.

Hence the mean, median and mode coincide at the origin.

(III) The points of inflexion of the normal curve are given by $x = \pm\sigma$.

Proof. We know that points of inflexion are given by

$$\frac{d^2y}{dx^2} = 0 \text{ and } \frac{d^3y}{dx^3} \neq 0.$$

Differentiating (1) w.r.t. 'x', we get

$$\frac{dy}{dx} = -\frac{y_0 x}{\sigma^2} e^{-x^2/2\sigma^2}$$

$$\frac{d^2y}{dx^2} = -\frac{y_0}{\sigma^2}\left(1 - \frac{x^2}{\sigma^2}\right) e^{-x^2/2\sigma^2}$$

$$\frac{d^3y}{dx^3} = \frac{xy_0}{\sigma^4}\left(3 - \frac{x^2}{\sigma^2}\right) e^{-x^2/2\sigma^2}.$$

Now $$\frac{d^2y}{dx^2} = 0 \Rightarrow \left(1 - \frac{x^2}{\sigma^2}\right) = 0 \Rightarrow x = \pm\sigma.$$

Also at $x = \pm\sigma$, $\quad d^3y/dx^3 \neq 0.$

Hence the points of inflexion are at $x = \pm\sigma$.

(IV) In a normal distribution, all the moments of odd order about the origin vanish.

Proof. The moment of order $2p + 1$ (*i.e.*, of odd order) about the origin is given by

$$\mu'_{2p+1} = \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-x^2/2\sigma^2} . x^{2p+1} dx$$

$$= \frac{1}{\sigma\sqrt{(2\pi)}} \int_{-\infty}^{\infty} x^{2p+1} . e^{-x^2/2\sigma^2} dx$$

$$= 0 \qquad [\because x^{2p+1} e^{-x^2L2\sigma^2} \text{ is an odd function of } x]$$

$$\therefore \qquad \mu'_1 = 0 = \mu'_3 = \mu'_5 = \ldots \text{etc.}$$

If the mean is at origin, then all the moments of odd order about the mean vanish.

(V) Moment of even order about the origin.

The moment of order 2p (even) about the origin is given by

$$\mu'_{2p} = \frac{2^p \sigma^{2p}}{\sqrt{\pi}} \Gamma\left(p + \frac{1}{2}\right).$$

Proof. By definition, we have

$$\mu'_{2p} = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2/2\sigma^2} x^{2p}\, dx$$

$$= \frac{2}{\sigma\sqrt{2\pi}} \int_0^{\infty} e^{-x^2/2\sigma^2} . x^{2p}\, dx$$

$$\left[\because \int_{-\infty}^{\infty} f(x)\, dx = 2\int_0^{\infty} f(x)\, dx, \text{ if } f(-x) = f(x)\right]$$

Putting $\frac{x^2}{2\sigma^2} = t, \quad \therefore dx = \frac{\sigma\, dt}{\sqrt{2}\sqrt{t}}$, we get

$$\mu'_{2p} = \frac{2}{\sigma\sqrt{2\pi}} \int_0^{\infty} e^{-t}\, 2^p \sigma^{2p} t^{\,p} . \frac{\sigma}{\sqrt{2}\sqrt{t}}\, dt$$

$$= \frac{2^p \sigma^{2p}}{\sqrt{\pi}} \int_0^{\infty} e^{-t}\, t^{(p+1/2)-1}\, dt = \frac{2^p \sigma^{2p}}{\sqrt{\pi}} \Gamma\left(p + \frac{1}{2}\right). \quad ...(1)$$

Replacing p by $p-1$ in (1), we get

$$\mu'_{2p-2} = \frac{2^{p-1} \sigma^{2p-2}}{\sqrt{\pi}} \Gamma\left(p - \frac{1}{2}\right). \quad ...(2)$$

Dividing (1) by (2), we get

$$\frac{\mu'_{2p}}{\mu'_{2p-2}} = 2\sigma^2 \frac{\Gamma\left(p+\frac{1}{2}\right)}{\Gamma\left(p-\frac{1}{2}\right)} = \frac{2\sigma^2\left(p-\frac{1}{2}\right)\Gamma\left(p-\frac{1}{2}\right)}{\Gamma\left(p-\frac{1}{2}\right)}$$

$$[\because \Gamma(m) = (m-1)\,\Gamma(m-1)]$$

or

$$\mu'_{2p} = \sigma^2 (2p-1)\, \mu'_{2p-2}. \quad ...(3)$$

Putting $p = p-1, p-2, \ldots 1,$

$$\mu'_{2p-2} = \sigma^2 (2p-3)\, \mu'_{2p-4}$$

$$\mu'_{2p-4} = \sigma^2 (2p-5)\, \mu'_{2p-6}$$

$$\ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots$$

$$\mu'_4 = \sigma^2 (3)\, \mu'_2$$

$$\mu'_2 = \sigma^2 (1)\, \mu'_0.$$

Substituting these values in (3), we get

$$\mu'_{2p} = \sigma^2 . (2p-1)\, \sigma^2 (2p-3)\, \mu'_{2p-4}$$
$$= (\sigma^2)^2 . (2p-1) . (2p-3)\, \sigma^2 (2p-5)\, \mu'_{2p-6}$$
$$\ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots$$
$$= \sigma^{2p} (2p-1)(2p-3) \ldots 3 . 1 \mu'_0$$
$$= \sigma^{2p} (2p-1)(2p-3) \ldots 3 . 1. \qquad [\text{as } \mu'_0 = 1]$$

In particular, $\mu'_2 = \sigma^2, \mu'_4 = 3\sigma^4$.

Note. Since the mean is at the origin.

$\therefore \qquad \mu_2 = \mu'_2 = \sigma^2$ etc.

§ 2.11. MOMENTS ABOUT THE MEAN m

By definition, we have

$$\mu_{2n+1} = \int_{-\infty}^{\infty} (x-m)^{2n+1} \frac{1}{\sigma\sqrt{(2\pi)}} e^{-(x-m)^2/2\sigma^2}\, dx$$
$$= \frac{1}{\sigma\sqrt{(2\pi)}} \int_{-\infty}^{\infty} (x-m)^{2n+1} . e^{-(x-m)^2/2\sigma^2}\, dx$$
$$= \frac{1}{\sigma\sqrt{(2\pi)}} \int_{-\infty}^{\infty} (\sigma t)^{2n+1} e^{-t^2/2} . \sigma\, dt$$

[substituting $(x-m)/\sigma = t$, so that $dx = \sigma\, dt$]

$$= \frac{\sigma^{2n+1}}{\sqrt{(2\pi)}} \int_{-\infty}^{\infty} t^{2n+1} e^{-t^2/2}\, dt$$
$$= 0, \text{ as } t^{2n+1} e^{-t^2/2} \text{ is an odd function of } t.$$

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

Hence *in a normal distribution, all the moments of odd order about the mean vanish.*

Again $\mu_{2n} = \int_{-\infty}^{\infty} (x-m)^{2n} . \frac{1}{\sigma\sqrt{(2\pi)}} e^{-(x-m)^2/2\sigma^2}\, dx$

$$= \frac{1}{\sigma\sqrt{(2\pi)}} \int_{-\infty}^{\infty} (x-m)^{2n} e^{-(x-m)^2/2\sigma^2}\, dx$$
$$= \frac{1}{\sigma\sqrt{(2\pi)}} \int_{-\infty}^{\infty} (\sigma t)^{2n} e^{-t^2/2} . \sigma\, dt \quad [\text{put } (x-m)/\sigma = t, \text{ and } dx = \sigma\, dt]$$
$$= \frac{\sigma^{2n}}{\sqrt{(2\pi)}} \int_{-\infty}^{\infty} t^{2n} e^{-t^2/2}\, dt$$
$$= \frac{2\sigma^{2n}}{\sqrt{(2\pi)}} \int_{0}^{\infty} t^{2n} e^{-t^2/2}\, dt \qquad [\text{as } t^{2n} e^{-t^2/2} \text{ is an even function of } t]$$

Now putting $t^2 = 2z$ or $t = \sqrt{2}\sqrt{z}$, $\therefore dt = dz/\sqrt{(2z)}$, we get

$$\mu_{2n} = \frac{2^n \sigma^{2n}}{\sqrt{\pi}} \int_0^{\infty} z^{(n+1/2)-1} e^{-z}\, dz$$

or

$$\mu_{2n} = \frac{2^n \sigma^{2n}}{\sqrt{\pi}} \Gamma\left(n + \frac{1}{2}\right). \qquad \ldots(1)$$

Replacing n by $(n-1)$ in (1), we have

$$\mu_{2n-2} = \frac{2^{n-1}\sigma^{2n-2}}{\sqrt{\pi}}\left(n-\frac{1}{2}\right). \quad \text{...(2)}$$

Dividing (1) by (2), we have

$$\frac{\mu_{2n}}{\mu_{2n-2}} = \frac{2\sigma^2\,\Gamma\left(n+\frac{1}{2}\right)}{\Gamma\left(n-\frac{1}{2}\right)} = \frac{2\sigma^2\left(n-\frac{1}{2}\right)\Gamma\left(n-\frac{1}{2}\right)}{\Gamma\left(n-\frac{1}{2}\right)}$$

$$[\because\ \Gamma l = (l-1)\,\Gamma(l-1)]$$

or $$\mu_{2n} = 2\sigma^2\left(n-\frac{1}{2}\right)\mu_{2n-2}$$

or $$\mu_{2n} = \sigma^2(2n-1)\,\mu_{2n-2}.$$

Now proceeding as in the last article, we have

$$\mu_{2n} = (2n-1)(2n-3)\ldots 5\,.\,3\,.\,1\,.\,\sigma^{2n}.$$

◈ § 2.12. β AND γ COEFFICIENTS OF NORMAL DISTRIBUTION

$$\beta_1 = \mu_3^2/\mu_2^3 = 0 \qquad [\because\ \mu_3 = 0]$$

Consequently $\quad \gamma_1 = \sqrt{(\beta_1)} = 0$

and $$\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{3\sigma^2\mu_2}{(\mu_2)^2} = \frac{3\sigma^2}{\mu_2} = \frac{3\sigma^2}{\sigma^2} = 3$$

Now $\quad \gamma_2 = \beta_2 - 3 = 0,$

i.e., kurtosis of a normal curve vanish.

◈ § 2.13. MEAN DEVIATION FROM MEAN OF THE NORMAL DISTRIBUTION

Theorem. *The mean deviation from the mean of the normal distribution is $\frac{4}{5}$ times its standard deviation.*

Proof. Mean deviation from mean 'm'

$$= \int_{-\infty}^{\infty} |x-m|\,.\,\frac{1}{\sigma\sqrt{2\pi}}\,e^{-(x-m)^2/2\sigma^2}\,dx$$

$$= \int_{-\infty}^{\infty} |\sigma t|\,.\,\frac{1}{\sigma\sqrt{(2\pi)}}\,e^{-t^2/2}\sigma\,dt, \quad [\text{put } (x-m)/\sigma = t, \text{ and } dx = \sigma\,dt]$$

$$= \frac{\sigma}{\sqrt{(2\pi)}}\int_{-\infty}^{\infty} |t|\,e^{-t^2/2}\,dt$$

$$= \frac{\sigma}{\sqrt{(2\pi)}}\left[\int_{-\infty}^{0} -te^{-t^2/2}\,dt + \int_{0}^{\infty} te^{-t^2/2}\,dt\right]$$

$$= \frac{\sigma}{\sqrt{(2\pi)}} \left[\int_0^{\infty} ze^{-z^2/2} \, dz + \int_0^{\infty} te^{-t^2/2} \, dt \right]$$

[put $t = -z$, $dt = -dz$ in the 1st integral]

$$= \frac{\sigma}{\sqrt{(2\pi)}} 2 \int_0^{\infty} te^{-t^2/2} \, dt$$

$$= \sigma \sqrt{\frac{2}{\pi}} \int_0^{\infty} e^{-u} \, du, \qquad \text{[Putting } t^2/2 = u\text{]}$$

$$= \sigma \sqrt{\frac{2}{\pi}} \left[-e^{-u} \right]_0^{\infty} = \sigma \sqrt{\frac{2}{\pi}} = \sigma\,(0 \cdot 8) \text{ nearly} = 4\sigma/5 \text{ (nearly)}.$$

◈ § 2.14. DETERMINATION OF PROBABILITY (OR AREA) IN CASE OF NORMAL DISTRIBUTION

Let $X \sim N(\mu, \sigma^2)$ be a random variable, where μ is the mean and σ^2 is the variance of X; then the probability that random value of X will lie between $X = x_1$ and $X = x_2$ is given by the formula :

$$P(x_1 \le X \le x_2) = \int_{x_1}^{x_2} f(x) \, dx = \int_{x_1}^{x_2} \frac{1}{\sigma\sqrt{(2\pi)}} e^{-(x-\mu)^2/2\sigma^2} \, dx.$$

Putting $z = \dfrac{x - \mu}{\sigma}$, *i.e.*, $x - \mu = \sigma z$, $dx = \sigma \, dz$, we have

$$P(x_1 \le X \le x_2) = \frac{1}{\sqrt{(2\pi)}} \int_{(x_1 - \mu)/\sigma}^{(x_2 - \mu)/\sigma} e^{-z^2/2} \, dz$$

$$= P\left(\frac{x_1 - \mu}{\sigma} < z < \frac{x_2 - \mu}{\sigma} \right) = P\left(\frac{x_1 - \mu}{\sigma} < \frac{X - \mu}{\sigma} < \frac{x_2 - \mu}{\sigma} \right).$$

In order to simplify the above expressions, different tables have been formed. One of these tables is given in the end of this book which gives the probability (or area) for values from 0 to z.

Some important results. Following facts are to be considered while using the table :

(I) $$P(-\infty < X < \infty) = \int_{-\infty}^{\infty} f(x) \, dx = \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{(2\pi)}} e^{-(x-\mu)^2/2\sigma^2} \, dx$$

$$= \int_{-\infty}^{\infty} \frac{1}{\sqrt{(2\pi)}} e^{-z^2/2} \, dz, \qquad \left[\text{putting } \frac{x - \mu}{\sigma} = z \right]$$

$$= \frac{2}{\sqrt{(2\pi)}} \int_0^{\infty} e^{-z^2/2} \, dz = \frac{2}{\sqrt{2\pi}} \int_0^{\infty} e^{-t} \frac{dt}{\sqrt{(2t)}}, \text{ where } z^2/2 = t$$

$$= \frac{1}{\sqrt{\pi}} \int_0^{\infty} t^{\frac{1}{2} - 1} e^{-t} \, dt = \frac{1}{\sqrt{\pi}} \Gamma\left(\frac{1}{2}\right) = \frac{\sqrt{\pi}}{\sqrt{\pi}} = 1.$$

Thus we see that the total area $\left[\int_{-\infty}^{\infty} f(x) \, dx \right]$ under normal probability curve is unity.

(II) Since, normal distribution is a symmetrical distribution, therefore, we have

(a) $P(z \geq 0) = P(z \leq 0) = \frac{1}{2} = 0 \cdot 5$

i.e.,
$$\int_{-\infty}^{0} \phi(z)\, dz = \int_{0}^{\infty} \phi(z)\, dz \text{ where } \phi(z) = \frac{1}{\sqrt{(2\pi)}} e^{-z^2/2}, \qquad \text{[Note]}$$

Remark 1. In the tables, the areas under the standard normal curve are tabulated, therefore in numerical problems we shall take standard normal variate z instead of the variate X.

Remark 2. If we are required to find area under normal curve, we will convert the given area into the form $P(0 < z < z_1)$ because the areas have been tabulated in this form in the tables.

(b) Since the curve is symmetrical about y-axis, therefore, the probability (area)

$$P(-b \leq z \leq -a) = P(a \leq z \leq b)$$
$$= \int_{a}^{b} \phi(z)\, dz = \int_{0}^{b} \phi(z)\, dz - \int_{0}^{a} \phi(z)\, dz$$
$$= P(0 \leq z \leq b) - P(0 \leq z \leq a).$$

(c) Using transformation $z = \dfrac{x - \mu}{\sigma}$, we have

$$P(z_1 \leq z \leq z_2) = P(x_1 \leq x \leq x_2).$$

From the table, we have

(i) $P(-1 \leq z \leq 1) = 2 \times P(0 \leq z \leq 1) = 2 \times 34 \cdot 134\% = 68 \cdot 268\%$

or $P(-\sigma \leq x - \mu \leq \sigma) = 68 \cdot 268\%.$

(ii) $P(-2 \leq z \leq 2) = 2 \times P(0 \leq z \leq 2) = 2 \times 47 \cdot 725\% = 95 \cdot 45\%$

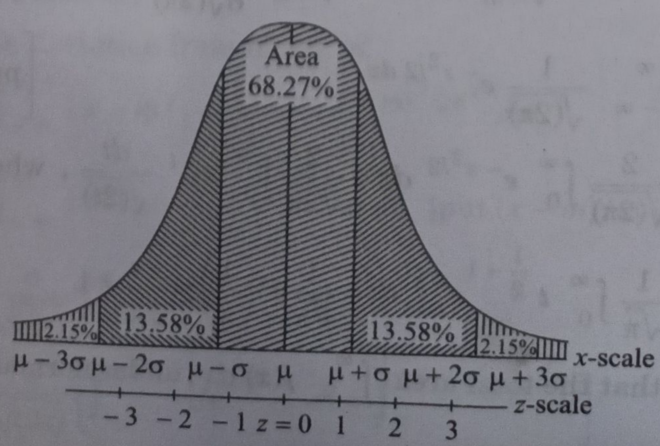
or $P(-2\sigma \leq x - \mu \leq 2\sigma) = 95 \cdot 45\%.$

(iii) $P(-3 \leq z \leq 3) = 2 \times P(0 \leq z \leq 3) = 2 \times 49 \cdot 865\% = 99 \cdot 73\%$

or $P(-3\sigma \leq x - \mu \leq 3\sigma) = 99 \cdot 73\%.$

Thus the probability that a normal variate X lies outside the range $\mu \pm 3\sigma$ is given by

$$P(|x - \mu| > 3\sigma) = P(|z| > 3) = 1 - P(-3 < z < 3) = 0 \cdot 27\%.$$



§ 2.15. STANDARD NORMAL VARIATE

Definition. A normal variate z defined by the function

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

is called a standard normal variate and may be denoted by $N(0, 1)$.

Thus, *a normal variate z with mean zero and standard deviation unity is called a standard normal variate.*

The variance of standard normal variate is unity. By the transformation $z = \frac{x-\mu}{\sigma}$ we obtain $N(\mu, \sigma^2) \sim N(0, 1)$.

ILLUSTRATIVE EXAMPLES

Example 1. ***For the normal curve*** $y = \frac{1}{\sigma\sqrt{(2\pi)}} e^{-(x-m)^2/2\sigma^2}$

(a) ***Find the mean and standard deviation.***

(b) ***Find the points of inflexion.***

Solution. The equation of given normal curve is as follows :

$$y = \frac{1}{\sigma\sqrt{(2\pi)}} e^{-(x-m)^2/2\sigma^2}. \quad ...(1)$$

(a) Here mean $= \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{(2\pi)}} e^{-\frac{(x-m)^2}{2\sigma^2}} x\, dx$

$$= \frac{1}{\sigma\sqrt{(2\pi)}} \int_{-\infty}^{\infty} e^{-t^2} (t\sigma\sqrt{2} + m)\, \sigma\sqrt{2}\, dt,$$

putting $\frac{x-m}{\sigma\sqrt{2}} = t$ or $dx = \sigma\sqrt{2}\, dt$

$$= \frac{1}{\sigma\sqrt{2\pi}} \left[2\sigma^2 \int_{-\infty}^{\infty} te^{-t^2}\, dt + m\sigma\sqrt{2} \int_{-\infty}^{\infty} e^{-t^2}\, dt\right]$$

$$= \frac{1}{\sigma\sqrt{(2\pi)}} \left[2\sigma^2(0) + 2m\sigma\sqrt{2} \int_0^{\infty} e^{-t^2}\, dt\right]$$

$$= \frac{1}{\sigma\sqrt{(2\pi)}} \cdot 2m\sigma\sqrt{2} \left(\frac{1}{2}\sqrt{\pi}\right) = m$$

and variance $= \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{(2\pi)}} e^{-(x-m)^2/2\sigma^2} (x-m)^2\, dx$

Putting $\frac{x-m}{\sigma\sqrt{2}} = t$ or $dx = \sigma\sqrt{t}\, dt$

$$= \frac{1}{\sigma\sqrt{(2\pi)}} \int_{-\infty}^{\infty} e^{-t^2} (t\sigma\sqrt{2})^2\, \sigma\sqrt{2}\, dt$$

$$= \frac{2}{\sigma\sqrt{(2\pi)}} \sigma^3 . \sqrt{2} \int_{-\infty}^{\infty} t^2 e^{-t^2}\, dt = \frac{2\sigma^2}{\sqrt{\pi}} \int_{-\infty}^{\infty} t . (te^{-t^2})\, dt$$

$$= \frac{2\sigma^2}{\sqrt{\pi}} \left[\left\{ t \cdot \left(-\frac{1}{2} e^{-t^2} \right) \right\}_{-\infty}^{\infty} - \int_{-\infty}^{\infty} \left(-\frac{1}{2} e^{-t^2} \right) dt \right]$$

[integrating by parts]

$$= \frac{2\sigma^2}{\sqrt{\pi}} \left[\{0\} + \frac{1}{2} \int_{-\infty}^{\infty} e^{-t^2}\, dt \right] = \frac{2\sigma^2}{\sqrt{\pi}} \left(\frac{1}{2} \sqrt{\pi} \right) = \sigma^2.$$

$\therefore$ Standard deviation $= \sqrt{\sigma^2} = \sigma$.

(b) We know that, at the point of inflexion,

$$d^2 y / dx^2 = 0 \text{ and } d^3 y / dx^3 \neq 0.$$

Differentiating (1) two times and equating to zero, we have

$$d^2 y / dx^2 = 0$$

$$\Rightarrow \quad \frac{1}{\sigma\sqrt{(2\pi)}} \left[\left(-\frac{x-m}{\sigma^2} \right)^2 e^{-(x-m)^2/2\sigma^2} - \frac{1}{\sigma^2} e^{-(x-m)^2/2\sigma^2} \right] = 0$$

$$\Rightarrow \quad \left(\frac{x-m}{\sigma^2} \right)^2 - \frac{1}{\sigma^2} = 0 \Rightarrow x = m \pm \sigma.$$

$$\text{Again} \quad \frac{d^3 y}{dx^3} = \frac{1}{\sigma\sqrt{(2\pi)}} \left[\left\{ -\frac{x-m}{\sigma^2} \right\}^3 e^{-(x-m)^2/2\sigma^2} + \frac{3(x-m)}{\sigma^4} e^{-(x-m)^2/2\sigma^2} \right]$$

$$= \pm \frac{2}{\sigma^4 \sqrt{(2\pi)}} e^{-1/2}.$$

On writing $m = x \pm \sigma$, we get $\dfrac{d^3 y}{dx^3} \neq 0$.

$\therefore$ When $x = m \pm \sigma$, $y = \dfrac{1}{\sigma\sqrt{(2\pi)}} e^{-1/2}$.

$\therefore$ The points of inflexion are $\left(m \pm \sigma, \dfrac{1}{\sigma\sqrt{(2\pi)}} e^{-1/2} \right)$, which are equidistant from the mean.

Example 2. ***If two normal universes have the same total frequency but the standard deviation of one is k times that of the other, show that the maximum frequency of the first is (1/ k)th times that of the other.***

Solution. Let N be the total frequency; then we know that the equation of normal curve is :

$$y = \frac{1}{\sigma\sqrt{(2\pi)}} e^{-\frac{(x-m)^2}{2\sigma^2}}. \quad \text{...(1)}$$

Here if mean and variance of two normal universes are respectively m_1, σ_1^2 and m_2, σ_2^2, then

$$y = \frac{N}{\sigma_1\sqrt{(2\pi)}} e^{-\frac{(x-m_1)^2}{2\sigma_1^2}} \quad \text{...(2)}$$

and

$$y = \frac{N}{\sigma_2\sqrt{(2\pi)}} e^{-\frac{(x-m_2)^2}{2\sigma_2^2}}. \quad \text{...(3)}$$

We know that for a normal curve, the maximum frequency is given at the mean. Thus if y_1 and y_2 are the maximum frequencies of first and second universes respectively, then we have

$$y = \left[\frac{N}{\sigma_1\sqrt{(2\pi)}} e^{-\frac{(x-m_1)^2}{2\sigma_1^2}}\right]_{x=m_1} = \frac{N}{\sigma_1\sqrt{(2\pi)}}.$$

Similarly, $$y_2 = \frac{N}{\sigma_2\sqrt{(2\pi)}}.$$

$$\therefore \quad \frac{y_1}{y_2} = \frac{N/\{\sigma_1\sqrt{(2\pi)}\}}{N/\{\sigma_2\sqrt{(2\pi)}\}} = \frac{\sigma_2}{\sigma_1} = \frac{\sigma_2}{k\sigma_2} = \frac{1}{k} \qquad [\because \text{ given } \sigma_1 = k\sigma_2]$$

or $$y_1 = (1/k)y_2.$$

Example 3. ***For normal curve, $N(\mu, \sigma^2)$, show that***

$$\mu_{2p+2} = \sigma^2\mu_{2p} + \sigma^3\frac{d\mu_{2p}}{d\sigma}.$$

Solution. We know that

$$\mu_{2p} = \frac{1}{\sigma\sqrt{(2\pi)}}\int_{-\infty}^{\infty} e^{-\frac{(x-m)^2}{2\sigma^2}}(x-m)^{2p}\,dx \qquad ...(1)$$

$$\therefore \quad \frac{d}{d\sigma}\mu_{2p} = -\frac{1}{\sigma^2\sqrt{(2\pi)}}\int_{-\infty}^{\infty} e^{-\frac{(x-m)^2}{2\sigma^2}}(x-m)^{2p}\,dx$$
$$+\frac{1}{\sigma^4\sqrt{(2\pi)}}\int_{-\infty}^{\infty} e^{-\frac{(x-m)^2}{2\sigma^2}}(x-m)^{2p+2}\,dx$$

or $$\sigma^3\frac{d}{d\sigma}\mu_{2p} = -\frac{\sigma^3}{\sigma^2\sqrt{(2\pi)}}\int_{-\infty}^{\infty} e^{-\frac{(x-m)^2}{2\sigma^2}}(x-m)^{2p}\,dx$$
$$+\frac{1}{\sigma\sqrt{(2\pi)}}\int_{-\infty}^{\infty} e^{-\frac{(x-m)^2}{2\sigma^2}}(x-m)^{2p+2}\,dx$$

or $$\sigma^3\frac{d}{d\sigma}\mu_{2p} = -\sigma^2\mu_{2p} + \mu_{2p+2}.$$

$$\therefore \quad \mu_{2p+2} = \sigma^2\mu_{2p} + \sigma^3\frac{d}{d\sigma}\mu_{2p}.$$

Example 4. ***For some normal distribution the first moment about 10 is 40 and fourth moment about 50 is 48. What is the mean variance and S.D. of the normal distribution.***

Solution. Consider that the mean and variance for the normal distribution are M and σ^2 respectively.

$$\mu_1'(10) = E(x-10) = E(x) - 10 = M - 10 = 40 \qquad \text{(given)}$$

Thus $$M = 40 + 10 = 50$$

$$\mu_4'(50) = \mu_4 = 3\sigma^4 = 48 \text{ (given)}.$$

$\therefore \quad \sigma^4 = 16$ or $\sigma^2 = 4$ and S.D. $= \sqrt{\sigma^2} = 2$.

$\therefore \quad$ Mean $= 50, \quad \sigma^2 = 4.$

S.D. $= 2.$

Example 5. (a) ***Assume the mean height of soldiers to be 68·22 inches with a variance of 10·8 $(in)^2$. How many soldiers in a regiment of 1000 would you expect to be over 6 feet tall ? Given that the area under the standard normal curve between $t = 0$ and $t = 0\cdot35$ is 0·1368, and between $t = 0$ and $t = 1\cdot15$ is 0·3746.***

Solution. Here $m = 68\cdot22$ inches, $\sigma^2 = 10\cdot8\,(\text{inch})^2$, *i.e.*, $\sigma = \sqrt{(10\cdot8)}$ inches $= 3\cdot28$ inches.

$$\therefore \quad t = \frac{x - m}{\sigma} = \frac{72 - 68\cdot22}{3\cdot28} \text{ for } x = 6' = 72''$$

$$= 3\cdot78 / 3\cdot28 = 1\cdot15 \text{ (nearly).}$$

But it is given that the area between $t = 0$ and $t = 1\cdot15$ is 0·3746. Thus the area between $t = 1\cdot15$ and $t = \infty$ is $0\cdot5 - 0\cdot3746$, *i.e.*, 0·1254.

$\therefore$ The number of soldiers who are over 6 feet tall

$$= 1000 \times 0\cdot1254 = 125\cdot4 \text{ nearly.}$$

Example 5. (b) ***The life of army shoes is normally distributed with mean 8 months and standard deviation 2 months. If 5000 pairs are issued, how many pairs would be expected to need replacement after 12 months ? Given that $P(z \geq 2) = 2\cdot0228$.***

Solution. Here mean $m = 8$ months, $\sigma = 2$ months.

$$\therefore \quad z = \frac{x - m}{\sigma} = \frac{12 - 8}{2} = 2.$$

Area against $z = 2$ in the table $6 = 0\cdot4772$ [here z is used in place of t]

$\therefore$ The area to the right of the ordinate $z = 2$ (*i.e.*, are between $z = 2$ and $z = \infty$) is

$$= 0\cdot5 - 0\cdot4772 = 0\cdot0228.$$

$\therefore$ The number of pairs whose life is more than 12 months

$$= 5000 \times 0\cdot0228 = 114.$$

Hence number of pairs to be replaced $= 5000 - 114 = 4886$.

Example 6. ***For a normal distribution with mean 2 and standard deviation 3, find the value of a variate such that the probability of the interval from the mean to that value is 0·4115.***

Solution. Given $\dfrac{2}{3\sqrt{(2\pi)}} \displaystyle\int_2^x e^{-\frac{1}{2}\frac{(x-2)^2}{9}}\, dx = 0\cdot4115.$

Now substituting $t = (x - 2)/3$ or $dt = \dfrac{1}{3}\, dx$,

$$\frac{1}{\sqrt{(2\pi)}} \int_0^t e^{-\frac{1}{2}t^2}\, dt = 0\cdot4115.$$

But from table, $t = 1\cdot35$, hence $x = 3t + 2 = 6\cdot05$.

Example 7. ***A random variable x is normally distributed with mean $= 12$ and standard deviation 2, find $P(9\cdot 6 < x < 13\cdot 8)$ given that for $x/\sigma = 0\cdot 9$, $A = 0\cdot 3159$ and for $x/\sigma = 1\cdot 2$, $A = 0\cdot 3849$.***

Solution. We are given that

$$\text{mean} = m = 12 \text{ and } \sigma = 2.$$

$$\therefore \quad t = \frac{x-m}{\sigma} = \frac{9\cdot 6 - 12}{2} = \frac{-2\cdot 4}{2} = -1\cdot 2, \quad \text{when } x = 9\cdot 6$$

and
$$t = \frac{13\cdot 8 - 12}{12} = \frac{1\cdot 8}{2} = 0\cdot 9, \quad \text{when } x = 13\cdot 8$$

$\therefore$ The required probability $= P\,(9\cdot 6 < x < 13\cdot 8)$

$$= \int_{9\cdot 6}^{13\cdot 8} \frac{1}{2\sqrt{(2\pi)}}\, e^{-\frac{1}{2}\frac{(x-12)^2}{4}}\, dx$$

$$= \frac{1}{\sqrt{(2\pi)}} \int_{-1\cdot 2}^{0\cdot 9} e^{-\frac{1}{2}t^2}\, dt \quad \text{where } t = \frac{x-12}{2}$$

$$= \frac{1}{\sqrt{(2\pi)}} \int_{-1\cdot 2}^{0} e^{-\frac{1}{2}t^2}\, dt + \frac{1}{\sqrt{(2\pi)}} \int_{0}^{0\cdot 9} e^{-\frac{1}{2}t^2}\, dt$$

$$= \frac{1}{\sqrt{(2\pi)}} \int_{0}^{1\cdot 2} e^{-\frac{1}{2}t^2}\, dt + \frac{1}{\sqrt{(2\pi)}} \int_{0}^{0\cdot 9} e^{-\frac{1}{2}t^2}\, dt$$

$$\left[\because \text{under the normal curve} : \int_{-1\cdot 2}^{0} \phi(t)\, dt = \int_{0}^{1\cdot 2} \phi(t)\, dt, \text{ where } \phi(t) = \frac{1}{\sqrt{(2\pi)}}\, e^{-\frac{1}{2}t^2}\right]$$

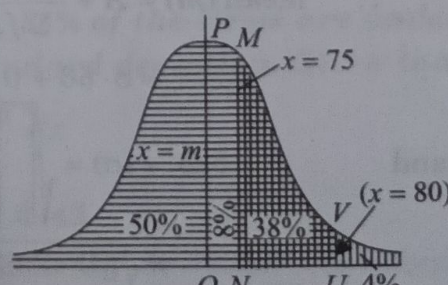
$$= 0\cdot 3849 + 0\cdot 3159 = 0\cdot 7008.$$

Example 8. ***If skulls are classified A, B, C according as the length, breadth index as under 75, between 75 and 80, or over 80, find approximately (assuming that distribution is normal) the mean and standard deviation of a series in which A are 58%, B are 38% and C are 4% being given that***

$$f(t) = \frac{1}{\sqrt{(2\pi)}} \int_0^t exp\left(-\frac{1}{2}x^2\right) dx$$

then $\quad f(0\cdot 20) = 0\cdot 80 \quad$ ***and*** $\quad f(1\cdot 75) = 0\cdot 46.$

Solution. Consider that the mean and standard deviation of the distribution are m and σ respectively. Since the total frequency is taken as unity, thus frequency of skull A, whose length and breadth index is under 75 is 0·58, the frequency of skull B whose index lies between 75 and 80 is 0·38; and frequency of skull C, whose index is over 80 is 0·04.



Thus the total area to the left of ordinate MN is 0·58 area, between ordinates MN and UV is 0·38 and the area to the right of ordinate UV is $0\cdot 04$.

Therefore the area between origin and $x\,(= 75 - m)$, *i.e.*, the area of $PQMN = 0\cdot58 - 0\cdot5 = 0\cdot08$, *i.e.*, the area corresponding to the $t = \dfrac{75-m}{\sigma}$ is 0·08.

But we have $f(0\cdot20) = 0\cdot08$.

Hence $\quad (75 - m)/\sigma = 0\cdot20.$ ...(1)

Again the area between origin and $x\,(= 80 - m)$,

i.e., area of $\quad PQVU = 0\cdot08 + 0\cdot38 = 0\cdot46,$

i.e., the area corresponding to $t = (80 - m)/\sigma$ is 0·46.

But $\quad f(1\cdot75) = 0\cdot46.$

Hence $\quad (80 - m)/\sigma = 1\cdot75.$ ...(2)

Dividing (1) by (2) and solving, we get

$$m = 74\cdot3, \quad \sigma = 3\cdot23.$$

Example 9. ***Fit a normal curve to the following data :***

length of line (in cm) :	**8·60**	**8·59**	**8·58**	**8·57**	**8·56**	**8·55**	**8·54**	**8·53**	**8·52**
frequency :	**2**	**3**	**4**	**9**	**10**	**8**	**4**	**1**	**1**

Solution. Let the required normal curve be $y = \dfrac{N}{\sigma\sqrt{(2\pi)}}\, e^{-\frac{(x-m)^2}{2\sigma^2}}$.

Here $N = 42$. Let $A = 8\cdot56$.

x	f	$\xi = x - A$	ξ^2	$f\xi$	$f\xi^2$
8·60	2	0·04	0·0016	0·08	0·0032
8·59	3	0·03	0·0009	0·09	0·0027
8·58	4	0·02	0·0004	0·08	0·0016
8·57	9	0·01	0·0001	0·09	0·0009
8·56	10	0	0	0	0
8·55	8	−0·01	0·0001	−0·08	0·0008
8·54	4	−0·02	0·0004	−0·08	0·0016
8·53	1	−0·03	0·0009	−0·03	0·0009
8·52	1	−0·04	0·0016	−0·04	0·0016
Total	42		0·0060	0·11	0·0133

$$\therefore \quad \text{mean } (m) = A + \frac{\Sigma f\xi}{\Sigma f} = 8\cdot56 + \frac{0\cdot11}{42}$$

$$= 8\cdot56 + 0\cdot0026 = 8\cdot5626 = 8\cdot563 \text{ (nearly)}$$

and $$\text{S.D.} = (\sigma) = \sqrt{\left[\left\{\frac{\Sigma f\xi^2}{\Sigma f} - \left(\frac{\Sigma f\xi}{\Sigma f}\right)^2\right\}\right]} = \sqrt{\left[\left\{\frac{0\cdot0133}{42} - \left(\frac{0\cdot11}{42}\right)^2\right\}\right]}$$

$$= \sqrt{[\{0\cdot000317 - 0\cdot0000069\}]} = 0\cdot0175 \text{ cm. (nearly)}.$$

Hence on simplification, the equation of normal curve to be fitted is

$$y = 9\cdot8e^{-0\cdot163(x - 8\cdot563)^2}.$$

Example 10. ***Prove that for the normal distribution, the quartile deviation, the mean deviation and the standard deviation are approximately in the ratio 10 : 12 : 15.***

Solution. Let Q_1 and Q_3 be the lower and higher quartiles respectively and m the mean.

Let $$P\{x \le Q_1\} = 0 \cdot 25.$$

$\therefore$ $$P\{Q_1 < x < m\} = 0 \cdot 5 - 0 \cdot 25 = 0 \cdot 25$$

$\therefore$ $$\frac{1}{\sigma\sqrt{(2\pi)}} \int_{Q_1}^{m} e^{-\frac{1}{2}\frac{(x-m)^2}{\sigma^2}} dx = 0 \cdot 25$$

or $$\frac{1}{\sqrt{(2\pi)}} \int_{0}^{(m-Q_1)/\sigma} e^{-\frac{1}{2}y^2} dy = 0 \cdot 25$$

[Putting $y = (m - x)/\sigma$]

$\therefore$ From tables, $$\frac{m - Q_1}{\sigma} = 0 \cdot 6744 \qquad ...(1)$$

Again $$P\{x \le Q_3\} = 0 \cdot 25$$

$\therefore$ $$P(m \le x < Q_3) = 0 \cdot 25.$$

$\therefore$ $$\frac{1}{\sigma\sqrt{(2\pi)}} \int_{m}^{Q_3} e^{-\frac{1}{2}\frac{(x-m)^2}{\sigma^2}} dx = 0 \cdot 25.$$

Putting $y = \dfrac{m - x}{\sigma}$,

$$\frac{1}{\sqrt{(2\pi)}} \int_{0}^{(Q_3-m)/\sigma} e^{-\frac{1}{2}y^2} dy = 0 \cdot 25.$$

From table, $$(Q_3 - m)/\sigma = 0 \cdot 6744. \qquad ...(2)$$

Adding (1) and (2), we have

$$(Q_3 - Q_1)/2 = 0 \cdot 6744\sigma = (2/3)\,\sigma \text{ (nearly)}.$$

$\therefore$ Quartile deviation (Q.D.) = (2/3) σ, by solving.

Mean deviation (M.D.) = (4/5)σ.

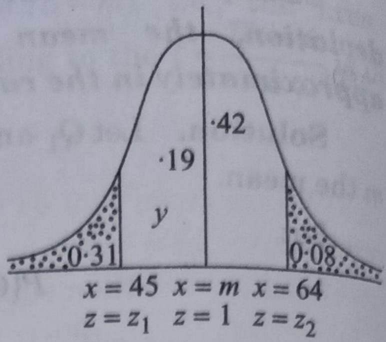
$\therefore$ $$\text{Q.D.} : \text{M.D.} : \text{S.D.} = 10 : 12 : 15.$$

Example 11. ***In a normal distribution, 31% of the items are under 45 and 8% are above 64. Find the mean and standard deviation. Given that if***

$$f(t) = \frac{1}{\sqrt{(2\pi)}} \int_0^t e^{-x^2/2}\, dx$$

then $$f(0 \cdot 5) = 0 \cdot 19, \quad f(1 \cdot 4) = 0 \cdot 42$$

Solution. Let m and σ be the mean and standard deviation respectively of the normal distribution. Then, we have $P(X \le 45) = 31\% = 0 \cdot 31$ and $P(X \ge 64) = 8\% = 0 \cdot 08$.



$$\therefore \quad P\left(\frac{X-m}{\sigma} \leq \frac{45-m}{\sigma}\right) = 0\cdot 31$$

and $$P\left(\frac{X-m}{\sigma} \geq \frac{64-m}{\sigma}\right) = 0\cdot 08.$$

Let $z = \dfrac{X-m}{\sigma}$, then $Z \sim N(0, 1)$

$$\therefore \quad P\left(Z \leq \frac{45-m}{\sigma}\right) = 0\cdot 31 \qquad \text{...(1)}$$

and $$P\left(Z \geq \frac{64-m}{\sigma}\right) = 0\cdot 08. \qquad \text{...(2)}$$

Given that if $f(t) = \dfrac{1}{\sqrt{(2\pi)}} \displaystyle\int_0^t e^{-x^2/2}\, dx$ ...(3)

then $f(0\cdot 5) = 0\cdot 19$ and $f(1\cdot 4) = 0\cdot 42$.

Now taking $f(0\cdot 5) = 0\cdot 19$, (3) becomes

$$0\cdot 19 = \frac{1}{\sqrt{(2\pi)}} \int_0^{0.5} e^{-z^2/2}\, dz = \frac{1}{\sqrt{(2\pi)}} \int_{-0.5}^{0} e^{-z^2/2}\, dz$$

$$= \frac{1}{\sqrt{(2\pi)}} \int_{-\infty}^{0} e^{-z^2/2}\, dz - \frac{1}{\sqrt{(2\pi)}} \int_{-\infty}^{-0.5} e^{-z^2/2}\, dz$$

$$= \frac{1}{2} - \frac{1}{\sqrt{(2\pi)}} \int_{-\infty}^{-0.5} e^{-z^2/2}\, dz \qquad \text{[see § 2.17]}$$

$$\Rightarrow \int_{-\infty}^{-0.5} \frac{1}{\sqrt{(2\pi)}} e^{-z^2/2}\, dz = 0\cdot 5 - 0\cdot 19 = 0\cdot 31$$

$$\Rightarrow \quad P(Z \leq -0\cdot 5) = 0\cdot 31. \qquad \text{...(4)}$$

Again taking $f(1\cdot 4) = 0\cdot 42$, (3) becomes

$$0\cdot 42 = \frac{1}{\sqrt{(2\pi)}} \int_0^{1.4} e^{-z^2/2}\, dz$$

$$= \frac{1}{\sqrt{(2\pi)}} \int_0^{\infty} e^{-z^2/2}\, dz - \frac{1}{\sqrt{(2\pi)}} \int_{1.4}^{\infty} e^{-z^2/2}\, dz$$

$$= \frac{1}{2} - \frac{1}{\sqrt{(2\pi)}} \int_{1.4}^{\infty} e^{-z^2/2}\, dz$$

$$\Rightarrow \int_{1.4}^{\infty} \frac{1}{\sqrt{(2\pi)}} e^{-z^2/2}\, dz = 0\cdot 5 - 0\cdot 42 = 0\cdot 08$$

$$\Rightarrow \quad P(Z \geq 1\cdot 4) = 0\cdot 08. \qquad \text{...(5)}$$

Now comparing equations (1) and (4), we have

$$\frac{45-m}{\sigma} = -0\cdot 5, \quad i.e., \quad 45 - m = -0\cdot 5\sigma. \qquad \text{...(6)}$$

Again comparing equations (2) and (5), we have

$$\frac{64-m}{\sigma} = 1\cdot 4, \quad i.e., \quad 64 - m = 1\cdot 4\sigma \qquad \text{...(7)}$$

Solving equations (6) and (7), we get

Mean, $m = 50$ and standard deviation, $\sigma = 10$.

Second Method. Let m and σ be the mean and standard deviation respectively.

Let $z = \dfrac{x-m}{\sigma}$, $\therefore$ when $x = m$ then $z = 0$.

We are given that 31% items are under 45, therefore the area to the left of the ordinate $x = 45$ is $0 \cdot 31$.

When $x = 45$ then suppose $z = z_1$.

$\therefore \quad P(z_1 < z < 0) = 0 \cdot 5 - 0 \cdot 31 = 0 \cdot 19$.

$\therefore \quad z_1 =$ the value of z (from table) corresponding to the area $0 \cdot 19$

$= -0 \cdot 5$ [since $z_1 < 0$]

Again, we are given that 8% items are above 64, therefore the area to the right of the ordinate $x = 64$ is $0 \cdot 08$.

When $x = 64$ then suppose $z = z_2$.

$\therefore \quad P(0 < z < z_2) = 0 \cdot 5 - 0 \cdot 08 = 0 \cdot 42$.

$\therefore \quad z_2 =$ the value of z (from table) corresponding to the area $0 \cdot 42$

$= 1 \cdot 4$ [since $z_2 > 0$]

Now $z_1 = -0 \cdot 5 \Rightarrow \dfrac{45-m}{\sigma} = -0 \cdot 5 \Rightarrow 45 - m = -0 \cdot 5\,\sigma$ ...(1)

and $z_2 = 1 \cdot 4 \Rightarrow \dfrac{64-m}{\sigma} = 1 \cdot 4 \Rightarrow 64 - m = 1 \cdot 4\,\sigma$ ...(2)

Solving (1) and (2), we get $m = 50, \sigma = 10$.

Example 12. ***In a test on 2000 electric bulbs, it was found that the life of a particular make was normally distributed with an average life of 2040 hours and S.D. of 60 hours. Estimate the number of bulbs likely to burn for***

(i) more than 2150 hours

(ii) less than 1950 hours.

Solution. Here mean (*i.e.*, average) $m = 2040$ hours, $\sigma = 60$ hours.

(i) When $x = 2150$ hours, $z = \dfrac{x-m}{\sigma} = \dfrac{2150-2040}{60} = 1 \cdot 833$.

Area against $z = 1 \cdot 83$ in the table $6 = 0 \cdot 4664$.

$\therefore$ The area to the right of the ordinate $z = 1 \cdot 83$ (*i.e.*, area between $z = 1 \cdot 83$ and $z = \infty$) as

$$= 0 \cdot 5 - 0 \cdot 4664 = 0 \cdot 0336.$$

$\therefore$ The number of bulbs likely to burn for more than 2150 hours

$$= 2000 \times 0 \cdot 0336 = 67 \text{ (nearly)}.$$

(ii) When $x = 1950$ hours, $z = \dfrac{x-m}{\sigma} = \dfrac{1950-2040}{60} = -1 \cdot 5$.

In this case the required area is left to the ordinate $z = -1 \cdot 5$, which is

$= 0 \cdot 5 - 0 \cdot 4332$ [$\because$ Area against $z = 1 \cdot 5$ in the table $6 = 0 \cdot 4332$]

$= 0 \cdot 0668$

$\therefore$ The number of bulbs likely to burn for less than 1950 hours

$$= 2000 \times 0 \cdot 0668 = 134 \text{ (nearly)}.$$

Remark. Now we shall also *find number of bulbs "likely to burn more than 1920 hours and but less than 2160 hours".*

For $x = 1920$, $\quad z = \dfrac{x - m}{\sigma} = \dfrac{1920 - 2040}{60} = -2$

For $x = 2160$, $\quad z = \dfrac{x - m}{\sigma} = \dfrac{2160 - 2040}{60} = 2.$

The required number of bulbs likely to burn is represented by the area between the ordinates $z = -2$ and $z = 2$, which is equal to twice the area from table 6 for the ordinate $z = 2$, which is $= 2 \times 0 \cdot 4772 = 0 \cdot 0544.$

Hence the required number of bulbs likely to burn

$$= 2000 \times 0 \cdot 9544 = 1909 \text{ nearly.}$$

EXERCISE 2 (C)

1. Prove that for a normal distribution $\mu_{2n} = (2n - 1)\,\sigma^2 \mu_{2n-2}$.

   Hence show that $\mu_{2n} = \dfrac{(2n)!}{2!}\left(\dfrac{1}{2}\sigma^2\right)^n$ and $\mu_{2n+1} = 0, n = 1, 2 \ldots.$

2. In a normal distribution 31% of the items are under 45 and 8% are over 64. Find the mean and standard deviation of the distribution. It is given that if $f(t) = \dfrac{1}{\sqrt{(2\pi)}} \displaystyle\int_0^t e^{-\frac{1}{2}x^2}\, dx$ then $f(0 \cdot 5) = 0 \cdot 19$, $f(1 \cdot 4) = 0 \cdot 42$.

3. Fit a normal distribution to the following data :

Mid point of interval :	100	95	90	85	80	75	70	65	60	55	50	45
Frequency :	0	1	3	2	7	12	10	9	5	3	2	0

4. Fit a normal distribution to the following data :

Variable :	60–62	63–65	66–68	69–71	72–74
Frequency :	5	18	42	27	8

5. If X has the distribution $N(\mu, \sigma^2)$, then prove that $E(X) = \mu$ and $\text{Var}(X) = \sigma^2$.
6. Write the normal distribution and state its properties.
7. Describe the properties and uses of a normal distribution.
8. What is the additive property of the normal distributions ? Write and prove it.
9. Assuming that the diameters of 1000 brass plugs taken consecutively from a machine, form a normal distribution with mean 0.7515 cm. and standard deviation 0.0020 cm., how many of the plugs are likely to be rejected if the approved diameter is 0.752 ± 0.004 cm. ?
10. The mean height of 500 students is 151 cm and the standard deviation is 15 cm. Assuming that the heights are normally distributed, find how many student's heights lie between 120 and 155 cm.

ANSWERS

1. $m = 50, \sigma = 10$
3. $m = 71 \cdot 2, \sigma = 9 \cdot 95$. Theoretical frequencies $0 \cdot 2, 0 \cdot 6, 1 \cdot 8, 4 \cdot 1, 7 \cdot 3, 10 \cdot 1, 10 \cdot 7, 8 \cdot 9, 5 \cdot 7, 2 \cdot 9, 1 \cdot 1, 0 \cdot 3$
4. $m = 67 \cdot 5, \sigma = 2 \cdot 9$; Expected frequencies $2 \cdot 55, 13 \cdot 43, 13 \cdot 10, 18 \cdot 64, 4 \cdot 81$
9. 52
10. 294.

§ 2.16. RECTANGULAR OR UNIFORM DISTRIBUTION [CONTINUOUS PROBABILITY DISTRIBUTION]

A random variable X is called to have a *rectangular distribution* or *uniform distribution* over $a < x < b$ if its probability function $f(x)$ is constant over the entire range of X. This distribution is called rectangular since the curve $y = f(x)$ describes a rectangle over the x-axis and between the ordinates at $x = a$ and $x = b$. It implies that X is a continuous variable.

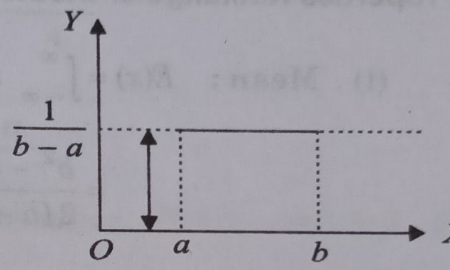
Since the total probability is always unity, therefore, if X is a rectangular variate over the interval (a, b), then

$$\int_a^b f(x)\, dx = 1$$

$$\Rightarrow \quad f(x) \int_a^b dx = 1 \quad \text{[since } f(x) \text{ is constant]}$$

$$\Rightarrow \quad f(x)\,[b - a] = 1$$

$$\Rightarrow \quad f(x) = \frac{1}{b-a}.$$



Graph of Rectangular Distribution

Therefore, the rectangular distribution $1/(b-a)$ is obtained from the following probability function :

$$f(x) = \begin{cases} \dfrac{1}{b-a}, & a < x < b \\ 0, & \text{otherwise.} \end{cases} \qquad \text{...(A)}$$

§ 2.17. DISTRIBUTION FUNCTION OF F(X)

The distribution function of $F(X)$ is defined by

$$F(x) = \begin{cases} 0, & \text{if } -\infty < x < \infty \\ \dfrac{x-a}{a-b}, & a \le x \le b \\ 1, & b < x < \infty. \end{cases} \qquad \text{...(B)}$$

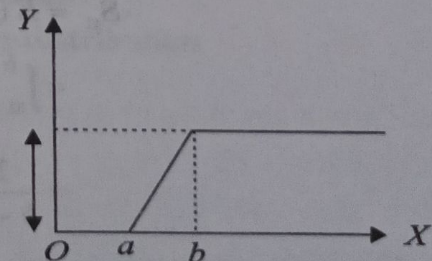
Clearly $F(x)$ is not continuous at

$$x = a \text{ and } x = b.$$

Therefore $F(x)$ is not differentiable at

$$x = a \text{ and } x = b.$$

$$\therefore \quad \frac{dF(x)}{dx} = f(x) = \frac{1}{b-a} \neq 0$$



Graph of Distribution Function

exists everywhere except at $x = a$ and $x = b$ and hence the probability distribution function of $F(x)$ is given by the formula (A) of § 6.18 above.

Note. For a uniform or rectangular variate X in the open interval $(-a, a)$ the probability density function is given by the following formula

$$f(x) = \begin{cases} 1/(2a), & -a < x < a \\ 0, & \text{otherwise} \end{cases} \qquad \text{...(C)}$$

Cor. *The probability of an observation to lie in any interval of $a \le x \le b$ is $1/(b-a)$ times the length of that interval.*

Let (c, d) be a new interval such that

$$P(c \le x \le d) = \int_c^d \frac{dx}{b-a} = \frac{d-c}{b-a} = (1/b-a). \quad \text{[length of interval (c, d)]}$$

Properties Rectangular Distribution :

(i) Mean : $E(x) = \int_{-\infty}^{\infty} x\, f(x)\, dx = \int_a^b \frac{x}{b-a}\, dx = \frac{1}{b-a}\left[\frac{x^2}{2}\right]_a^b$

$$= \frac{b^2-a^2}{2(b-a)} = \frac{b+a}{2}.$$

(ii) Variance. $\text{var}(x) = \mu_2 = \frac{(b-a)^2}{12}$.

(iii) Standard Deviation. S. D. $= \sigma = \sqrt{\mu_2} = \frac{b-a}{\sqrt{(12)}}$

$$\beta_1 = \mu_3^2 / \mu_2^3 = 0 \qquad [\because \mu_3 = 0]$$

$$\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{(b-a)^4}{80} \times \frac{144}{(b-a)^4} = \frac{9}{5} = 1\cdot 8.$$

$$\gamma_1 = \sqrt{\beta_1} = 0, \gamma_2 = \beta_2 - 3 = 1\cdot 8 - 3 = -1\cdot 2.$$

(iv) Moment Generating Function about zero :

$$M_0(t) = E(e^{tx}) = \int_a^b \frac{e^{tx}}{b-a}\, dx = \frac{1}{b-a}\left[\frac{e^{tx}}{t}\right]_a^b = \frac{e^{bt} - e^{at}}{t^{(b-a)}}$$

(v) Mean Deviation about Mean :

$$S_\mu = E(\infty\, x - \mu\, \infty)$$

$$= \int_a^b \left| x - \frac{b-a}{2} \right| \frac{1}{b-a}\, dx$$

$$= \frac{1}{b-a} \int_{(a-b)/2}^{(b-a)/2} \infty\, y\, \infty\, dy, \text{ where } y = x - \frac{b+a}{2}$$

$$= \frac{1}{b-a}\left[\int_{(a-b)/2}^{0} (-y)\, dy + \int_0^{(b-a)/2} y\, dy\right] = \frac{b-a}{4}.$$

(vi) Quartiles. (i) For first quartile Q_1, we have

$$\int_a^{Q_1} \frac{1}{b-a}\, dx = \frac{1}{4} \Rightarrow \frac{1}{b-a}\left[\, x\, \right]_a^{Q_1} = \frac{1}{4}$$

$$\Rightarrow \qquad Q_1 - a = \frac{b-a}{4} \Rightarrow Q_1 = a + \frac{b-a}{4}.$$

(vii) For second quartile Q_2 (or for Median M_d). We have

$$\int_a^{M_d} \frac{1}{b-a}\,dx = \frac{1}{2} \Rightarrow \frac{1}{b-a}\Big[\,x\,\Big]_a^{M_d} = \frac{1}{2}$$

$$\Rightarrow M_d = Q_2 = a + \frac{b-a}{2}.$$

(viii) For 3rd quartile Q_3. We have

$$\int_a^{Q_3} \frac{1}{b-a}\,dx = \frac{3}{4} \Rightarrow Q_3 = a + \frac{3(b-a)}{4}.$$

(ix) Quartile Deviation. It is given by

$$\text{Q. D.} = \frac{Q_3 - Q_1}{2} = \frac{b-a}{4}.$$

Ex. *Find the moments of rectangular distribution. Hence compute $\beta_1, \beta_2, \gamma_1$ and γ_2.*

ILLUSTRATIVE EXAMPLES

Example 1. ***A variate X has a rectangular distribution on a unit interval. Find the function of u which has the distribution :***

$$dP = e^{-x}\,dx,\ 0 \le x \le \infty.$$

Solution. Here $dP = e^{-x}\,dx,\ 0 \le x \le \infty.$

Let $u = F(x)$, where $F'(x) = f(x) = e^{-x}$

$\therefore$ $u = P = \int_0^x e^{-x}\,dx = [-e^{-x}]_0^x$

$$= 1 - e^{-x} \Rightarrow 1 . e^{-x} = 1 - u$$

$$\Rightarrow x = \log_e \{1/(1-u)\}.$$

Example 2. ***For the rectangular distribution, where***

$$f(x) = \frac{1}{b-a},\quad a \le x \le b.$$

Find the mean deviation with respect to mean.

Solution. We know that for given rectangular distribution

$$\text{mean } (m) = (a+b)/2$$

$\therefore$ Mean deviation from mean (M. D.)

$$= E(|x - m|)$$

$$= \int_a^b \left| x - \frac{a+b}{2} \right| f(x)\,dx$$

$$= \frac{1}{b-a}\int_a^b \left| x - \frac{a+b}{2} \right| dx.$$

Substituting $t = x - (a+b)/2$, or $dx = dt$, we have

$$\text{M. D.} = \frac{1}{b-a}\int_{-(b-a)/2}^{(b-a)/2} |t|\,dt,$$

$$= \frac{2}{b-a}\int_0^{(b-a)/2} t\,.\,dt = \frac{b-a}{4}.$$ **Ans.**

Example 3. ***If $f(x)$ is probability function of continuous random variate X in the interval $a \le X \le b$ such that***

(i) $f(0) = 0$, $X \le a$ *and* $X \ge b$,

(ii) $\int_a^b f(x)\,dx = 1$.

Show that the variate $P = \int_a^x f(x)\,dx$ is a rectangular distribution.

Solution. Given $P = \int_a^x f(x)\,dx = P(X \le x)$

$\Rightarrow$ $dP = f(x)\,dx.$

Let $u = F(x)$ where $du/dx = F'(x) = f(x)$, then

$$dP = f(x)\,dx = f(x)\,.\,(dx/du)\,du$$
$$= f(x)\,.\,\{1/f(x)\}\,du = du,\ 0 \le u \le 1.$$

Since u is a distribution with unity probability. Consequently $u = P$ is a rectangular distribution.

Example 4. ***If X has a uniform distribution in $[0, 1]$ with probability density function***

$$f_x(x) = \begin{cases} 1, & 0 < x \le 1 \\ 0, & \textit{otherwise}. \end{cases}$$

Find the distribution (probability density function) of $-2\log X$.

Solution. Suppose $y = -2\log X$. Now the distribution function G of Y will be defined as follows :

$$G_y(y) = P(Y \le y) = P(-2\log X \le y)$$
$$= P\left(\log x \ge -\frac{1}{2}y\right) = P(X \ge e^{-y/2})$$
$$= 1 - P(x \le e^{-y/2})$$
$$= 1 - \int_0^{e^{-y/2}} f(x)\,dx$$
$$= 1 - \int_0^{e^{-y/2}} 1\,dx = 1 - e^{-y/2}$$

$$\therefore \quad g_y(y) = (d/dy)\,G(y) = \frac{1}{2}e^{-y/2},\ 0 < y < \infty.$$

Example 5. ***If X is a random variable which consists a continuous probability density function $F(x)$ and $F(x)$ represents probability. $X \le x$, then show that $F(x)$ is a random variable, hence it is a rectangular distribution in $(0, 1)$.***

Solution. Here, it is given that

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

and $dP = dF = f(x)\,dx$ and $F'(x) = f(x)$.

Let $u = F(x)$.

then $dP = f(x)\,dx = F'(x)\,dx = du$

which is a rectangular distribution in $(0, 1)$.

Example 6. ***For the rectangular distribution***

$$dF(x) = dx, \quad 0 \le x \le 1$$

find mean, variance and mean deviation about the mean.

Also show that its moment generating function (m. g. f.) is $(e^t - 1)/t$.

Solution. Here $d\,F(x) = dx \Rightarrow \dfrac{dF(x)}{dx} = 1$

$$\Rightarrow f(x) = 1.$$

Mean. $\mu = E(X) = \int_0^1 x\, f(x)\, dx = \int_0^1 x\, dx = \dfrac{1}{2}.$

Variance. $\sigma^2 = E[X - E(X)]^2$

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

$$= \frac{1}{24} + \frac{1}{24} = \frac{1}{12}.$$

Mean deviation about the mean.

$$\delta_\mu = E\left[\,|X - E(X)|\,\right]$$

$$= \int_0^1 \left|x - \frac{1}{2}\right| dx = \int_0^{1/2} \left(\frac{1}{2} - x\right) dx + \int_{1/2}^1 \left(x - \frac{1}{2}\right) dx$$

$$= \left[\frac{x}{2} - \frac{x^2}{2}\right]_0^{1/2} + \left[\frac{x^2}{2} - \frac{x}{2}\right]_{1/2}^1$$

$$= \left[\frac{1}{4} - \frac{1}{8}\right] + \left[\left(\frac{1}{2} - \frac{1}{2}\right) - \left(\frac{1}{8} - \frac{1}{4}\right)\right] = \frac{1}{4}.$$

Moment generating function (m. g. f.).

$$M_0(t) = E(e^{tx}) = \int_0^1 e^{tx}\, dx = \left[\frac{e^{tx}}{t}\right]_0^1 = \frac{e^t - 1}{t}.$$

Example 7. ***If X has a rectangular distribution over the interval*** **(0, 1),** ***where***

$$dF = dx, \quad 0 \le x \le 1,$$

then find the expected value of the function $\phi(x) = aX + b$***, where a and b are constants.***

Solution. Here $\quad dF = dx \Rightarrow dF/dx = 1 \Rightarrow f(x) = 1.$

$\therefore \quad$ Mean $= E(x) = \int_0^1 x\, f(x)\, dx = \int_0^1 x\, dx = \dfrac{1}{2}$

$\Rightarrow \quad E(X) = \dfrac{1}{2}.$

We are given that $\phi(x) = aX + b.$

$$\therefore \quad E[\phi(x)] = E(aX + b) = a\,E(X) + b$$

$$= a\left(\frac{1}{2}\right) + b = \frac{a}{2} + b.$$

Example 8. *Let* $f(x) = \begin{cases} 1, & 0 \le x \le 1 \\ 0, & \text{otherwise.} \end{cases}$

(i) Find the expected value of the function $\phi(x) = aX + b$ where a and b are constants.

(ii) Let x_1, x_2, x_3 be three observations on X, show that P (at least two of them are more than $0 \cdot 6) = 0 \cdot 352$

Solution. (i) See above example 11.

(ii) $P(x_i > 0 \cdot 6) = P(X > 0 \cdot 6)$ $\qquad [\because f(x) = 1]$

$$= \int_{0.6}^{1} f(x)\,dx = \int_{0.6}^{1} 1 \,.\, dx$$

$$= \Big[\,x\,\Big]_{0.6}^{1} = 1 - 0 \cdot 6 = 0 \cdot 4.$$

It is true for $i = 1, 2, 3$.

$$\therefore \quad P(x_i < 0 \cdot 6) = 1 - P(x_i > 0 \cdot 6) = 1 - 0 \cdot 4 = 0 \cdot 6.$$

Now P (any two of them are more than $0 \cdot 6) = {}^3C_2\,(0 \cdot 4)^2\,(0 \cdot 6)^2 = 0 \cdot 288$

and $\quad P$ (all the three are more than $0 \cdot 6) = (0 \cdot 4)^3 = 0 \cdot 064$.

$\therefore \quad P$ (at least two of them are more than $0 \cdot 6) = 0 \cdot 288 + 0 \cdot 064 = 0 \cdot 352$

Proved.

§ 2.18. EXPONENTIAL DISTRIBUTION

Exponential distribution is a probability distribution for which the probability density function $f(x)$ is defined by

$$F(x) = \begin{cases} 0 & , \text{ for } x \le 0 \\ \theta e^{-\theta x} & , \text{ for } x > 0, \theta > 0. \end{cases}$$

The distribution function is given by

$$F(x) = \begin{cases} 0 & , \text{ for } x < 0 \\ 1 - e^{-\theta x} & , \text{ for } x \ge 0 \\ 1 & , \text{ for } x = \infty. \end{cases}$$

It is a gamma distribution whose parameter is unity.

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

$$= 0 + \int_{0}^{\infty} \theta e^{-\theta x}\,dx = \Big[-e^{-\theta x}\Big]_{0}^{\infty} = 1.$$

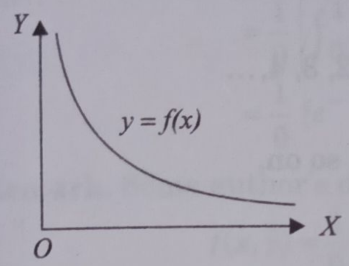
§ 2.19. GRAPH OF EXPONENTIAL PROBABILITY FUNCTION

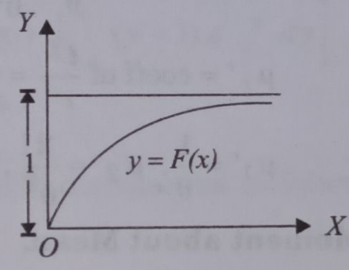
Here for $f(x) = \theta e^{-\theta x}$, we have

x	0	1	2	...	∞
$f(x)$	θ	$\theta e^{-\theta}$	$\theta e^{-2\theta}$	...	0

When $\theta = 1$, $f(x) = 1, e^{-1}, e^{-2}, \ldots, 0.$

When $\theta = 2$, $f(x) = 2, 2e^{-2}, 2e^{-3}, \ldots, 0.$





By the definition of $F(x)$, we have

x	0	1	2	...	∞
$F(x)$	0	$1 - e^{-\theta}$	$1 - e^{-2\theta}$	...	1

When $\theta = 1$, $F(x) = 0, 1 - e^{-1}, 1 - e^{-2}, \ldots, 1.$

When $\theta = 2$, $F(x) = 0, 1 - e^{-2}, 1 - e^{-4}, \ldots, 1.$

§ 2.20. PROPERTIES OF EXPONENTIAL DISTRIBUTION

(i) **Mean.** $E(X) = \int_{-\infty}^{\infty} x\, f(x)\, dx = \int_0^{\infty} x\theta e^{-\theta x}\, dx$

$$= \theta \left[\frac{xe^{-\theta x}}{-\theta} \right]_0^{\infty} - \theta \int_0^{\infty} \frac{e^{-\theta x}}{-\theta}\, dx$$

$$= 0 + \int_0^{\infty} e^{-\theta x}\, dx = \left[\frac{e^{-\theta x}}{-\theta} \right]_0^{\infty} = \frac{1}{\theta}.$$

(ii) **Variance.** $\text{var}(X) = E[X - E(X)]$

$$= E(X^2) - [E(X)]^2 \qquad \ldots(1)$$

Now $E(X^2) = \int_0^{\infty} x^2 \theta e^{-\theta x}\, dx$

$$= 0 + \frac{2}{\theta} \int_0^{\infty} x\theta e^{-\theta x}\, dx = \frac{2}{\theta^2}. \qquad \ldots(2)$$

Putting from (2) in (1), we have

$$\text{var}(X) = \frac{2}{\theta^2} - \frac{1}{\theta^2} = \frac{1}{\theta^2}.$$

(iii) Standard Deviation.

$$\text{S. D.} = \sigma = \sqrt{[\text{var}(x)]} = \frac{1}{\theta}.$$

(iv) Moment generating function about zero.

We have $$M_0(t) = E(e^{tx}) = \int_0^\infty e^{tx}\theta e^{-\theta x}\,dx$$

$$= \theta\int_0^\infty e^{-(\theta - t)x}\,dx = \frac{\theta}{\theta - t} = \left(1 - \frac{t}{\theta}\right)^{-1}$$

$$= 1 + \frac{t}{\theta} + \frac{t^2}{\theta^2} + \ldots + \frac{t^r}{\theta^r} + \ldots. \qquad \text{[by binomial expansion]}$$

$$\therefore \qquad \mu_r' = \text{coeff of } \frac{t^r}{r!} = \frac{r!}{\theta^r} \quad \text{for } r = 1, 2, 3, 4, \ldots$$

So $$\mu_1' = \frac{1}{\theta},\ \mu_2' = \frac{2!}{\theta^2},\ \mu_3' = \frac{3!}{\theta^3} \text{ and so on.}$$

(v) Moment about Mean.

$$\mu_1 = E[X - E(X)] = E\left[X - \frac{1}{\theta}\right] = 0$$

$$\mu_2 = E\left(X - \frac{1}{\theta}\right)^2 = \mu_2' - \mu_1'^2 = \frac{2}{\theta^2} - \frac{1}{\theta^2} = \frac{1}{\theta^2}$$

$$\mu_3 = E\left(X - \frac{1}{\theta}\right)^3 = \mu_3' - 3\mu_2'\mu_1' + 2\mu_1'^3$$

$$= \frac{6}{\theta^3} - \frac{3 \times 2}{\theta^2}\cdot\frac{1}{\theta} + 2\left(\frac{1}{\theta}\right)^3 = \frac{2}{\theta^3}.$$

Similarly, $$\mu_4 = \mu_4' - 4\mu_3'\mu_1' + 6\mu_2'\mu_1'^2 - 3\mu_1'^4$$

$$= \frac{24}{\theta^4} - 4\cdot\frac{6}{\theta^3}\cdot\frac{1}{\theta} + 6\cdot\frac{2}{\theta^2}\cdot\left(\frac{1}{\theta}\right)^2 - 3\left(\frac{1}{\theta}\right)^4 = \frac{9}{\theta^4}.$$

(vi) Beta and Gamma Functions.

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} = \frac{4/\theta^6}{1/\theta^6} = 4;\ \beta_2 = \frac{\mu_4}{\mu_2^2} = 9$$

$$\gamma_1 = \sqrt{(\beta_1)} = 2;\ \gamma_2 = \beta_2 - 3 = 9 - 3 = 6.$$

(vii) Quartiles : For first quartile Q_1, we have

$$\int_0^{Q_1} \theta e^{-\theta x}\,dx = \frac{1}{4} \Rightarrow \theta\left[\frac{e^{-\theta x}}{-\theta}\right]_0^{Q_1} = \frac{1}{4}$$

$$\Rightarrow Q_1 = \frac{1}{\theta}\log\left(\frac{4}{3}\right).$$

(viii) For third quartile Q_3, we have

$$\int_0^{Q_3} \theta e^{-\theta x}\,dx = \frac{3}{4} \Rightarrow Q_3 = \frac{1}{\theta}\log 4.$$

$\therefore$ Inter quartile range $= Q_3 - Q_1 = \dfrac{1}{\theta} \log 3.$

Quartile Deviation (Q. D.) $= \dfrac{Q_3 - Q_1}{2} = \dfrac{1}{2\theta} \log 3.$

(ix) Mean Deviation. We have M. D.

$$= E\,[\,|\,x - E(x)\,|\,]$$

$$= \int_0^{\infty} \left| x - \frac{1}{\theta} \right| \theta e^{-\theta x}\, dx = \int_0^{\infty} |\,\theta x - 1\,|\, e^{-\theta x}\, dx$$

$$= \frac{1}{\theta} \int_0^{\infty} |\,y - 1\,|\, e^{-y}\, dy, \text{ where } y = \theta x$$

$$= \frac{1}{\theta} \left[\int_0^{1} (1 - y)\, e^{-y}\, dy + \int_1^{\infty} (y - 1)\, e^{-y}\, dy \right]$$

$$= \frac{1}{\theta}\, [e^{-1} + e^{-1}] = \frac{2}{\theta}\, e^{-1}.$$

Remark. Some author's define exponential distribution as following :

$$f(x, \beta) = \frac{1}{\beta}\, e^{-x/\beta}, \ \beta > 0.$$

ILLUSTRATIVE EXAMPLES

Example 1. ***A random variate has a exponential distribution with its probability density function given by***

$$F(x) = \begin{cases} 3e^{-3x}, & \text{for } x > 0 \\ 0, & \text{for } x \le 0. \end{cases}$$

What is the probability, that X is not less than 4 ? Find mean and standard deviation. Show that

$$\textit{coeff. of skewness} = \left(\frac{\textit{mean}}{\textit{S. D.}} \right) = 1$$

Solution. Here $\quad P(x \ge 4) = \int_4^{\infty} 3e^{-3x}\, dx = \left[-e^{-3x} \right]_4^{\infty} = e^{-12}.$

Mean $= 1/\theta = 1/3,$

Standard deviation (S. D.) $= 1/\theta = 1/3$

coeff. of skewness $= \dfrac{\text{Mean}}{\text{S. D.}} = \dfrac{1/3}{1/3} = 1.$

Example 2. ***The income-tax return of a man is a exponential distribution with the probability density function given by***

$$f(x) = \frac{1}{3}\, e^{-x/3}, \quad \textit{for } x > 0$$
$$= 0 \quad , \quad \textit{for } x \le 0.$$

What is the probability that his income will be above Rs. 17,000 it being assumed that the rate of tax is 15% above the income of Rs. 15,000.

Solution. If the income exceeds Rs. 17,000, then the income tax exceeds by Rs. 15% of (17,000 – 15,000) *i.e.*, exceeds by Rs. $\dfrac{15}{100} \times 2000$ *i.e.*, exceeds by Rs. 300.

$\therefore$ Required probability

$$P(x > 300) = \frac{1}{3}\int_{300}^{\infty} e^{-x/3}\, dx = \left[-e^{-x/3}\right]_{300}^{\infty} = e^{-100}.$$

Example 3. ***Income-tax return of some person forms exponential distribution with density function***

$$F(x) = \begin{cases} \frac{1}{2} e^{-x}, & \text{for } x > 0 \\ 0, & \text{for } x \le 0. \end{cases}$$

Then what is the probability of his income being more than Rs. 19,200 if 25% income-tax is taken over an income of Rs. 18,000 ?

Solution. If the income exceeds Rs. 18,000, then the income tax exceeds by Rs. 25% of (19,200 – 18,000)

i.e., exceeds by Rs. $\frac{25}{100} \times 1200$

i.e., exceeds by Rs. 300.

$\therefore$ Required probability

$$P(X > 300) = \frac{1}{2}\int_{300}^{\infty} e^{-x}\, dx = \left[-\frac{e^{-x}}{2}\right] = \frac{1}{2}e^{-300}.$$

Example 4. ***A family is selected at random from a densed populated area. A random variate may be assumed as annual income above Rs. 4000 having the following exponential distribution***

$$f(x) = \frac{1}{2000} e^{-x/2000}, \quad \textit{for } x > 0$$

$$= 0, \quad \textit{for } x \le 0.$$

What is porobability that out of four selected families from the area, 3 will have their income above Rs. 5,000.

Solution. Here Rs. 5000 – Rs. 4000 = Rs. 1000.

$$\therefore \quad P(x > 1000) = \int_{1000}^{\infty} \frac{1}{2000} e^{-x/2000}\, dx$$

$$= \left[-e^{-x/2000}\right]_{1000}^{\infty} = e^{-1/2} = 1/\sqrt{e}$$

$$\bar{P} = 1 - P(x > 1000) = 1 - 1/\sqrt{e}.$$

$\therefore$ Required probability $= {}^4C_3 (P)^3 (\bar{P})$

$$= 4\left(\frac{1}{\sqrt{e}}\right)^3 \left(1 - \frac{1}{\sqrt{e}}\right) = \frac{4(\sqrt{e} - 1)}{e^2}.$$

EXERCISE 2 (D)

1. What is exponential distribution and what are its applications ? Write different properties of this distribution.

§ 2.21. GAMMA DISTRIBUTION

Definition. A continuous random variable X which is distributed according to the probability density

$$f(x) = \begin{cases} \dfrac{e^{-x}\, x^{\lambda - 1}}{\Gamma(\lambda)} & , \quad \lambda > 0, 0 < x < \infty \\ 0 & , \quad \text{otherwise} \end{cases}$$

is called a *Gamma variate with parameter* λ, and it is denoted by $\gamma(\lambda)$, and its distribution is called *Gamma distribution.*

§ 2.22. SOME PROPERTIES OF GAMMA DISTRIBUTION :

P. (i) *To prove :* $\int_0^\infty f(x)\, dx = 1.$

Proof. We have

$$\int_0^\infty f(x)\, dx = \int_0^\infty \frac{e^{-x}\, x^{\lambda - 1}}{\Gamma(\lambda)} \qquad \text{[By definition]}$$

$$= \frac{1}{\Gamma(\lambda)} \int_0^\infty e^{-x}\, x^{\lambda - 1}\, dx$$

$$= \frac{1}{\Gamma(\lambda)} . \Gamma(\lambda) = 1 \qquad \left[\int_0^\infty e^{-x} x^{n-1}\, dx = \Gamma(n)\right]$$

Conclusion. Since $\int_0^\infty f(x)\, dx = 1$, therefore the function $f(x)$ defined above is a probability density function (p.d.f.). If X has this density, we may write $X \sim \gamma(\lambda)$.

P. (ii) Cumulative Distribution Function.

It is also called 'Incomplete gamma function' and is given by

$$F(x) = \begin{cases} \int_0^x f(u)\, du & , \quad u > 0 \\ 0 & , \quad \text{otherwise} \end{cases}$$

P. (iii) Moments (or Constants) of Gamma Distribution.

The rth moment μ_r' **about the origin** of $\gamma(\lambda)$ distribution is given by

$$\mu_r' = \int_0^\infty x^r\, f(x)\, dx$$

$$= \int_0^\infty x^r . \frac{e^{-x}\, x^{\lambda - 1}}{\Gamma(\lambda)}\, dx = \frac{1}{\Gamma(\lambda)} \int_0^\infty e^{-x} . x^{\lambda + r - 1}\, dx$$

$$= \frac{1}{\Gamma(\lambda)} . \Gamma(\lambda + r) = \frac{\Gamma(\lambda + r)}{\Gamma(\lambda)}.$$

Thus for $r = 1$, $\mu_1' = \dfrac{\Gamma(\lambda + 1)}{\Gamma(\lambda)} = \dfrac{\lambda\, \Gamma(\lambda)}{\Gamma(\lambda)} = \lambda = \text{Mean}$

for $r = 2$, $\mu_2' = \dfrac{\Gamma(\lambda + 2)}{\Gamma(\lambda)} = \lambda(\lambda + 1)$

for $r = 3$, $\mu_3' = \dfrac{\Gamma(\lambda + 3)}{\Gamma(\lambda)} = \lambda\, (\lambda + 1)\, (\lambda + 2)$

for $r = 4$, $\mu_4' = \dfrac{\Gamma(\lambda + 4)}{\Gamma(\lambda)} = \lambda(\lambda + 1)\, (\lambda + 2)\, (\lambda + 3).$

The central moments *i.e.*, moments about mean of $\gamma(\lambda)$ distribution are obtained as follows :

$$\mu_1 = 0,$$
$$\mu_2 = \mu_2' - (\mu_1')^2$$
$$= \lambda(\lambda+1) - \lambda^2 = \lambda^2 + \lambda - \lambda^2$$

i.e., $\mu_2 = \lambda = \text{variance}$

$$\mu_3 = \mu_3' - 3\mu_2'\mu_1' + 2\mu_1'^3$$
$$= \lambda(\lambda+1)(\lambda+2) - 3\lambda(\lambda+1)(\lambda) + 2\lambda^3$$
$$= \lambda^3 + 3\lambda^2 + 2\lambda - 3\lambda^3 - 3\lambda^2 + 2\lambda^3$$
$$= 2\lambda$$
$$\mu_4 = \mu_4' - 4\mu_1'\mu_3' + 6(\mu_1')^2\mu_2' - 3(\mu_1')^4$$
$$= \lambda(\lambda+1)(\lambda+2)(\lambda+3) - 4\lambda \,.\, \lambda(\lambda+1)(\lambda+2) + 6\lambda^3(\lambda+1) - 3\lambda^4$$
$$= \lambda\,[(\lambda^3 + 6\lambda^2 + 11\lambda + 6) - (4\lambda^3 + 12\lambda^2 + 8\lambda) + (6\lambda^3 + 6\lambda^2) - 3\lambda^3]$$
$$= \lambda(3\lambda + 6) = 3\lambda(2+\lambda).$$

P. (iv) Moment Generating Function (M.G.F.) of Gamma Distribution.

M.G.F. about the origin of gamma distribution is given by

$$M_X(t) = E(e^{tx}) = \int_0^\infty e^{tx} f(x)\, dx$$
$$= \frac{1}{\Gamma(\lambda)} \int_0^\infty e^{tx} \,.\, e^{-x} \,.\, e^{-x} x^{\lambda-1}\, dx$$
$$= \frac{1}{\Gamma(\lambda)} \int_0^\infty x^{\lambda-1} e^{-(1-t)x}\, dx$$
$$= \frac{1}{\Gamma(\lambda)} \cdot \frac{\Gamma(\lambda)}{(1-t)^\lambda} \qquad \left[\because \int_0^\infty x^{m-1} e^{-ax}\, dx = \frac{\Gamma(m)}{a^m}\right]$$
$$= (1-t)^{-\lambda},\ |t| < 1.$$

Now, we know that

$$\mu_r' = \left[\frac{d^r M_X(t)}{dt}\right]_{t=0}.$$

$$\therefore \quad \text{Mean} = \mu_1' = \left[\frac{d}{dt} M_X(t)\right]_{t=0} = \left[\frac{d}{dt}(1-t)^{-\lambda}\right]_{t=0}$$
$$= [-\lambda(1-t)^{-\lambda-1}(-1)]_{t=0} = \lambda$$
$$\mu_2' = \left[\frac{d^2}{dt^2} M_X(t)\right]_{t=0} = \left[\frac{d^2}{dt^2}(1-t)^{-\lambda}\right]_{t=0}$$
$$= \left[\frac{d}{dt}\{\lambda(1-\lambda)^{-\lambda-1}\}\right]_{t=0} = \left[\lambda(\lambda+1)(1-t)^{-\lambda-2}\right]_{t=0}$$
$$= \lambda(\lambda+1).$$

Thus, Variance $= \mu_2 = \mu_2' - (\mu_1')^2 = \lambda(\lambda+1) - \lambda^2 = \lambda.$

Remark. We observe that mean and variance of Gamma distribution are equal.

P. (v) Additive Property.

If $X_1, X_2, \ldots, X_n$ are independent Gamma variates with parameters $\lambda_1, \lambda_2, \ldots, \lambda_n$ respectively, then the sum $X_1 + X_2 + \ldots + X_n$ is also a Gamma variate with parameter $\lambda_1 + \lambda_2 + \ldots + \lambda_n$. In other words, we can state that the sum of independent Gamma variates is also a Gamma variate.

Proof. We are given that X_i are independent Gamma variates with parameters λ_i, $i = 1, 2, \ldots, n$. Therefore $\gamma(\lambda_i)$, $i = 1, 2, \ldots, n$ are Gamma variates. We know that

$$M_{X_i}(t) = (1-t)^{-\lambda_i}. \qquad \text{[See P. (iv)]} \qquad \ldots(1)$$

Since $X_1, X_2, \ldots, X_n$ are independent, so

$$\begin{aligned} M_{X_1, X_2, \ldots, X_n}(t) &= M_{X_1}(t) \cdot M_{X_2}(t) \ldots M_{X_n}(t) \\ &= (1-t)^{-\lambda_1} \cdot (1-t)^{-\lambda_2} \ldots (1-t)^{-\lambda_n} \qquad \text{[using (i)]} \\ &= (1-t)^{-(\lambda_1 + \lambda_2 + \ldots + \lambda_n)}. \qquad \ldots(2) \end{aligned}$$

R.H.S. of (2) *i.e.*, $(1-t)^{-(\lambda_1 + \lambda_2 + \ldots + \lambda_n)}$ represents the M.G.F. of a Gamma variate with parameter $\lambda_1 + \lambda_2 + \ldots + \lambda_n$. This proves the statement.

Note. Gamma Distribution with two parameters l and λ denoted by $\gamma(l, \lambda)$ has a probability density function

$$f(x) = \frac{l^\lambda}{\Gamma(\lambda)} e^{-lx} x^{\lambda - 1}, \quad x \geq 0, l > 0, \lambda \geq 0.$$

Its

$$\text{mean} = \lambda / l$$

$$\text{Variance} = \frac{\lambda}{l^2} = \frac{1}{l} \text{ mean}.$$

These results and M.G.F. can be obtained in the same way as we have done for $\gamma(\lambda)$.

§ 2.23. BETA DISTRIBUTION

§ 2.23. (A) BETA DISTRIBUTION OF FIRST KIND

Definition. A continuous random variable X which is distributed according to the probability density

$$f(x) = \begin{cases} \dfrac{x^{m-1}(1-x)^{n-1)}}{B(m, n)} & ; \quad m > 0, n > 0, 0 < x < 1 \\ 0 & ; \quad \text{otherwise} \end{cases}$$

is called *Beta variate of the first kind with two parameters m and n*, it is denoted $\beta_I(m, n)$, and its distribution is known as *Beta distribution of the first kind.*

Here $B(m, n)$ is the Beta function and is given by

$$B(m, n) = \int_0^1 x^{m-1}(1-x)^{n-1}\, dx = \frac{\Gamma(m)\,\Gamma(n)}{\Gamma(m+n)}.$$

Some Properties of β_I (m, n)

P. (i) *To prove :* $\int_0^1 f(x)\,dx = 1.$

Proof. We have

$$\int_0^1 f(x)\,dx = \int_0^1 \frac{x^{m-1}(1-x)^{n-1}}{B(m,n)}\,dx \qquad \text{[By definition]}$$

$$= \frac{1}{B(m,n)} \cdot \int_0^1 x^{m-1}(1-x)^{n-1}\,dx$$

$$= \frac{B(m,n)}{B(m,n)} = 1.$$

Conclusion. Since $\int_0^1 f(x)\,dx = 1$, therefore, the function $f(x)$ defined above is probability density function (p.d.f.) $\beta_I\ (m, n)$, it is also called Beta density of first kind. If X has this density, we may write $X \sim \beta_I\ (m, n)$.

P. (ii) Cumulative Distribution Function.

It is also called 'Incomplete Beta Function' and is given by

$$F(x) = \begin{cases} 0 & ; \text{ for } x < 0 \\ \int_0^x \frac{u^{m-1}(1-u)^{n-1}}{B(m,n)}\,du & ; \text{ for } 0 < x < 1, m > 0, n > 0 \\ 1 & ; \text{ for } x > 1. \end{cases}$$

P. (iii) Moments (or Constants) of β_I (m, n).

The r^{th} moment μ_r' about the origin of $\beta_I\ (m, n)$ distribution is given by

$$\mu_r' = \int_0^1 x^r f(x)\,dx = \frac{1}{B(m,n)}\int_0^1 x^r \,.\, x^{m-1}(1-x)^{n-1}\,dx$$

$$= \frac{1}{B(m,n)}\int_0^1 x^{(m+r)-1}(1-x)^{n-1}\,dx = \frac{B(m+r,n)}{B(m,n)}$$

$$= \frac{\Gamma(m+n)}{\Gamma m\,\Gamma n} \cdot \frac{\Gamma(m+r)\,\Gamma n}{\Gamma(m+r+n)} = \frac{\Gamma(m+n)\,\Gamma(m+r)}{\Gamma m\,\Gamma(m+n+r)}.$$

Thus for $r = 1$, $\mu_1' = \dfrac{\Gamma(m+n)\,\Gamma(m+1)}{\Gamma m\,\Gamma(m+n+1)} = \dfrac{\Gamma(m+n)\,m\,\Gamma m}{\Gamma m\,(m+n)\,\Gamma(m+n)}$

i.e.,

$$\mu_1' = \frac{m}{m+n} = \text{Mean}.$$

Note. μ_1' can be directly computed by $\int_0^1 x\,f(x)\,dx$.

For $r = 2$, $\mu_2' = \dfrac{\Gamma(m+n)\,\Gamma(m+2)}{\Gamma(m)\,\Gamma(m+n+2)} = \dfrac{m\,(m+1)}{(m+n)\,(m+n+1)}$

$[\because\ \Gamma(l+2) = (l+1)\,l\,\Gamma l]$

For $r = 3$, $\mu_3' = \dfrac{\Gamma(m+n)\,\Gamma(m+3)}{\Gamma m\,\Gamma(m+n+3)} = \dfrac{m\,(m+1)\,(m+2)}{(m+n)\,(m+n+1)\,(m+n+2)}.$

Similarly for $r = 4$, we have

$$\mu_4' = \frac{m\,(m+1)\,(m+2)\,(m+3)}{(m+n)\,(m+n+1)\,(m+n+2)\,(m+n+3)}.$$

The central moments (*i.e.*, moments about mean) of $\beta_I\,(m, n)$ are obtained as follows :

$$\mu_1 = 0$$

$$\begin{aligned}
\mu_2 &= \mu_2' - (\mu_1')^2 \\
&= \frac{m\,(m+1)}{(m+n)\,(m+n+1)} - \frac{m^2}{(m+n)^2} \\
&= \frac{m}{(m+n)}\left[\frac{(m+1)\,(m+n) - m\,(m+n+1)}{(m+n)\,(m+n+1)}\right] \\
&= \frac{mn}{(m+n)^2\,(m+n+1)} = \text{Variance}
\end{aligned}$$

$$\begin{aligned}
\mu_3 &= \mu_3' - 3\mu_2'\,\mu_1' + 2\,(\mu_1')^3 \\
&= \frac{2mn\,(n-m)}{(m+n)^3\,(m+n+1)\,(m+n+2)}
\end{aligned}$$

$$\begin{aligned}
\mu_4 &= \mu_4' - 4\mu_1'\,\mu_3' + 6\,(\mu_1')^2\,\mu_2' - 3\,(\mu_1')^4 \\
&= \frac{3mn\,\{2\,(m+n)^2 + mn\,(m+n-6)\}}{(m+n)^4\,(m+n+1)\,(m+n+2)\,(m+n+3)}.
\end{aligned}$$

Standard deviation σ is given by

$$\begin{aligned}
\sigma &= \sqrt{\mu_2} = \sqrt{\{\mu_2' - (\mu_1')^2\}} \\
&= \frac{\sqrt{mn}}{(m+n)\,\sqrt{(m+n+1)}}.
\end{aligned}$$

P. (iv) The Harmonic Mean of β_I (m, n),

The harmonic mean H of $\beta_I\,(m, n)$ is given by

$$\begin{aligned}
\frac{1}{H} &= \int_0^1 \frac{1}{x}\,f(x)\,dx = \frac{1}{B(m,n)}\int_0^x x^{(m-1)-1}\,(1-x)^{n-1}\,dx \\
&= \frac{B(m-1, n)}{B(m,n)} = \frac{\Gamma(m+n)}{\Gamma m\,\Gamma n}\cdot\frac{\Gamma(m-1)\,\Gamma n}{\Gamma(m+n-1)} \\
&= \frac{\Gamma(m+n)}{\Gamma m}\cdot\frac{\Gamma(m-1)}{\Gamma(m+n-1)} = \frac{m+n-1}{m-1}.
\end{aligned}$$

Hence $$H = \frac{m-1}{m+n-1}.$$

Mode of β_I (m, n) Distribution. For $\beta_I\,(m, n)$, we have

$$f(x) = \frac{x^{m-1}\,(1-x)^{n-1}}{B(m,n)}\,;\quad m > 0,\ n > 0,\ 0 < x < 1.$$

Taking log, we have

$$\log f(x) = -\log B(m,n) + (m-1)\log x + (n-1)\log(1-x).$$

Lokesh Sharma

Differentiating w.r.t. x

$$\frac{\partial}{\partial x} \log f(x) = \frac{m-1}{x} - \frac{n-1}{1-x}$$

$$\frac{\partial^2}{\partial x^2} \log f(x) = -\frac{(m-1)}{x^2} - \frac{n-1}{(1-x)^2}.$$

Putting $\frac{\partial}{\partial x} \log f(x)$ equal to zero, we have

$$\frac{m-1}{x} - \frac{n-1}{1-x} = 0 \Rightarrow x = \frac{m-1}{(m+n-2)}.$$

Clearly at $x = \frac{m-1}{(m+n-2)}$, we get $\frac{\partial^2}{\partial x^2} \log f(x) = \text{negative when } m > 1, n > 1.$

Hence mode of $\beta_I\ (m, n)$ distribution is

$$x = \frac{m-1}{(m+n-2)} \quad \text{when } m > 1,\ x = 1.$$

§ 2.23. (B) BETA DISTRIBUTION OF SECOND KIND

Definition. A continuous random variable X which is distributed according to the probability density

$$f(x) = \begin{cases} \dfrac{1}{B(m, n)} \dfrac{x^{m-1}}{(1+x)^{m+n}} & ; \quad m > 0, n > 0, 0 < x < \infty \\ 0 & ; \quad \text{otherwise} \end{cases}$$

is said to be *Beta distribution of the second kind with two parameters m and n*, it is denoted by $\beta_{II}\ (m, n)$, and its distribution is known as *Beta Distribution of the second kind.*

Properties of β_{II} (m, n) :

P. (i) *To prove :* $\int_0^\infty f(x)\, dx = 1.$

Proof. $\int_0^\infty f(x)\, dx = \int_0^\infty \frac{1}{B(m, n)} \cdot \frac{x^{m-1}}{(1+x)^{m+n}}\, dx = \frac{1}{B(m, n)} . B(m, n) = 1.$

Conclusion. Since $\int_0^\infty f(x)\, dx = 1$, therefore, the function $f(x)$ defined above is probability density function (p.d.f.) $\beta_{II}\ (m, n)$, it is also called Beta density of second kind. If X has this density, we may write $X \sim \beta_{II}\ (m, n)$.

P. (ii) Cumulative Distribution Function.

It is also called 'Incomplete Beta Function' and is given by

$$F(x) = \int_0^x f(u)\, du, \ 0 < x < \infty.$$

P. (iii) Moments (or constants) of β_{II} (m, n).

The rth moments μ_r' about the origin of $\beta_{II}\ (m, n)$ distribution is given by

$$\mu_r' = \int_0^\infty x^r f(x)\, dx = \frac{1}{B(m, n)} \int_0^\infty x^r \cdot \frac{x^{m-1}}{(1+x)^{m+n}}\, dx$$

$$= \frac{1}{B(m, n)} \int_0^\infty \frac{x^{(m+r)-1}}{(1+x)^{m+n}} \, dx$$

$$= \frac{1}{B(m, n)} \int_0^\infty \frac{x^{(m+r)-1}}{(1+x)^{(m+r)+(n-r)}} \, dx$$

$$= \frac{B(m+r, n-r)}{B(m, n)}$$

$$= \frac{\Gamma(m+n)}{\Gamma m \, \Gamma n} \cdot \frac{\Gamma(m+r)\,(\Gamma(n-r)}{\Gamma(m+n)} = \frac{\Gamma(m+r)\,\Gamma(n-r)}{\Gamma m \, \Gamma n}.$$

$\therefore$ For $r = 1$, $\mu_1' = \dfrac{\Gamma(m+1)\,\Gamma(n-1)}{\Gamma m \, \Gamma n} = \dfrac{m}{n-1} = \text{Mean}$

For $r = 2$, $\mu_2' = \dfrac{\Gamma(m+2)\,\Gamma(n-2)}{\Gamma m \, \Gamma n} = \dfrac{m\,(m+1)}{(n-1)\,(n-2)}.$

Similarly, $\quad \mu_3' = \dfrac{m\,(m+1)\,(m+2)}{(n-1)\,(n-2)\,(n-3)},$

$$\mu_4' = \frac{m\,(m+1)\,(m+2)\,(m+3)}{(n-1)\,(n-2)\,(n-3)\,(n-4)}.$$

The central moments (*i.e.*, moments about mean) of β_{II} (m, n) are obtained as:

$$\mu_1 = 0$$

$$\mu_2 = \mu_2' - (\mu_1')^2$$

$$= \frac{m\,(m+1)}{(n-1)\,(n-2)} - \frac{m^2}{(n-1)^2}$$

$$= \frac{m}{(n-1)} \left[\frac{m+1}{n-2} - \frac{m}{n-1} \right]$$

$$= \frac{m}{(n-1)} \left[\frac{(m+1)\,(n-1) - m\,(n-2)}{(n-1)\,(n-2)} \right]$$

$$= \frac{m}{n-1} \left[\frac{mn - m + n + n - 1 - mn + 2m}{(n-1)\,(n-2)} \right]$$

$$= \frac{m\,(m+n-1)}{(n-1)^2\,(n-2)} = \text{Variance.}$$

Standard deviation $\sigma = \sqrt{\mu_2} = \dfrac{1}{n-1} \sqrt{\left\{ \dfrac{m\,(m+n-1)}{(n-2)} \right\}}.$

P. (iv) Harmonic Mean of β_{II} (m, n).

The harmonic mean H of β_{II} (m, n) is given by

$$\frac{1}{H} = E\left(\frac{1}{X}\right) = \int_0^\infty \frac{1}{x} f(x) = \int_0^\infty \frac{1}{x} \cdot \frac{1}{B(m, n)} \cdot \frac{x^{m-1}}{(1+x)^{m+n}} \, dx$$

$$= \frac{1}{B(m, n)} \int_0^\infty \frac{x^{(m-1)-1}}{(1+x)^{(m-1)+(n+1)}} \, dx$$

$$= \frac{B(m-1, n+1)}{B(m, n)} = \frac{\Gamma(m+n)}{\Gamma m \, \Gamma n} \cdot \frac{\Gamma(m-1)\,\Gamma(n+1)}{\Gamma(m+n)}$$

$$= \frac{n}{m-1}.$$

$$\therefore \quad H = \frac{m-1}{n}.$$

ILLUSTRATIVE EXAMPLES

Example 1. ***If X follows β_I (m, n) then show that*** $Y = \left(\frac{1-X}{X}\right)$ ***will follow β_{II} (m, n).***

Solution. Since X follows $\beta_I\ (m, n)$, therefore

$$f_X\ (x) = \frac{1}{B(m, n)}\, x^{m-1} (1-x)^{n-1};\ m > 0,\ n > 0,\ 0 < x < 1.$$

Given
$$y = \frac{1-x}{x} \Rightarrow xy = 1 - x \Rightarrow x = \frac{1}{1+y}.$$

Its Jacobian $|J|$ of transformation is

$$|J| = \left|\frac{\partial x}{\partial y}\right| = \left|-\frac{1}{(1+y)^2}\right| = \frac{1}{(1+y)^2}$$

$$\therefore \quad f_Y\ (y) = f_X\ (x) \,.\, |J| = \frac{1}{B(m, n)}\, x^{m-1} (1-x)^{nm-1}\, |J|$$

$$= \frac{1}{B(m, n)} \left(\frac{1}{1+y}\right)^{m-1} \left(1 - \frac{1}{1+y}\right)^{n-1} \cdot \frac{1}{(1+y)^2}$$

$$= \frac{1}{B(m, n)} \cdot \frac{y^{n-1}}{(1+y)^{m+n}}.$$

This proves that Y follows $\beta_{II}\ (m, n)$.

Example 2. ***Find β_1 and β_2 for $\gamma(\lambda)$ distribution. Show that Gamma distribution tends to normal distribution as $\lambda \to \infty$.***

Solution. We have, for $\gamma(\lambda)$ distribution

$$\mu_2 = \lambda, \mu_3 = 2\lambda,\ \mu_4 = 3\lambda\,(2 + \lambda).$$

$$\therefore \quad \beta_1 = \frac{\mu_3^2}{\mu_2^3} = \frac{4\lambda^2}{\lambda^3} = \frac{4}{\lambda}$$

$$\beta_2 = \frac{\mu_4}{\mu_2{}^2} = \frac{3\lambda^2 + 6\lambda}{\lambda^2} = 3 + \frac{6}{\lambda}.$$

Now when $\lambda \to \infty$, we have

$$\beta_1 \to 0,\ \beta_2 \to 3.$$

Hence $\gamma(\lambda)$ distribution tends to normal distribution as $\lambda \to \infty$.

EXERCISE 2 (E)

1. State Gamma and Beta distribution and state their properties.
2. Describe Gamma and Beta distributions. Obtain their mean, variance and m.g.f. Prove that the sum of n independent Gamma variates follow again a Gamma variate.
3. Define Beta distribution of first kind. Find its mean, standard deviation, harmonic mean and μ_r'.
4. Describe Beta distribution of second kind. Find its mean, variance, harmonic mean and μ_r'.

EXERCISE 2 (F)

Objective Type Questions

1. The distribution for which mean = median = mode, is :
   (a) Binomial distribution (b) Poisson distribution
   (c) Normal distribution (d) Exponential distribution.
2. The mean and variance of normal distribution :
   (a) Are equal (b) Can never be equal
   (c) Are sometimes equal (d) Are equal in limiting position as $n \to \infty$.
3. Normal curve is :
   (a) Very much flat
   (b) Shape of bell and symmetry about the line $x = m$
   (c) Very much peaked
   (d) Neither more flat nor more peaked.
4. Normal distribution is the limiting position of binomial distribution if :
   (a) $n \to \infty, p \to 0$ (b) $n \to 0, p \to q$
   (c) $n \to \infty, p \to n$ (d) $n \to \infty$ and neither p nor q is small.
5. The points of inflexion of normal curve are :
   (a) $m \pm \sigma$ (b) $m \pm 2\sigma$
   (c) $m \pm 3\sigma$ (d) $m \pm (2/3)\,\sigma$.
6. The value of β_1 and β_2 for the normal distribution are respectively :
   (a) 0, 1 (b) 1, 0
   (c) 0, 3 (d) 3, 0.
7. The relation between the central moments of the normal distribution $N(\mu, \sigma^2)$ (with mean μ and variance σ^2) is :
   (a) $\mu_{2n} = n\,\sigma^2\,\mu_{2n-2}$ (b) $\mu_{2n} = 2n\sigma^2\,\mu_{2n-2}$
   (c) $\mu_{2n} = (2n-1)\,\sigma^2\,\mu_{2n-2}$ (d) $\mu_{2n} = (2n+1)\,\sigma^2\,\mu_{2n-2}$.
8. If μ_n denotes the nth central moment, then the normal distribution (m, σ^2) relation is true :
   (a) $\mu_{2n+1} = 1.3.5\ldots(2n+1)\,\sigma^{2n}$ (b) $\mu_{2n+1} = 0$
   (c) $\mu_{2n} = 1.3.5\ldots(2n-1)\,e^{2n}$ (d) (b) and (c).
9. The Kurtosis for normal curve is :
   (a) $3 + \sigma^2$ (b) $3 - \sigma^2$
   (c) 3 (d) 0.

10. The mean deviation from the mean of normal distribution $N(\mu, \sigma^2)$ is :
    (a) $(3/4)\ \sigma$ nearly
    (b) $(4/5)\ \sigma$ nearly
    (c) $(5/6)\ \sigma$ nearly
    (d) $(6/7)\ \sigma$ nearly.
11. For a normal distribution the first moment about 10 is 40 then the arithmetic mean is :
    (a) 38
    (b) 48
    (c) 50
    (d) 60.
12. The distance of the point of inflexion of a normal distribution from the mean ordinate is :
    (a) $\pm \sigma$
    (b) 0
    (c) $\pm \sigma^2$
    (d) $\pm \sigma^2 - 3$.
13. For a normal distribution $N(\mu, \sigma^2)$ the values of β_2 is :
    (a) 0
    (b) 3
    (c) σ^2
    (d) $\mu + \sigma^2$
14. The relation between mean, median and mode of a normal curve is :
    (a) mean < median < mode
    (b) mean > median > mode
    (c) mean = median = mode
    (d) mean ≤ median ≤ mode.

ANSWERS

1.	(c)	2.	(b)	3.	(b)	4.	(d)	5.	(a)	6.	(c)	7.	(c)	8.	(b)
9.	(d)	10.	(b)	11.	(c)	12.	(a)	13.	(b)	14.	(c)				

□■□