9.1. INTRODUCTION

We consider some univariate continuous distributions in this chapter. The main continuous distributions like uniform distribution, normal distribution, gamma, beta, exponential, Laplace, Weibul, Logistic and Cauchy distributions will be discussed in detail in the subsequent sections.

9.2. NORMAL DISTRIBUTION

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

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

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

or

$$f(x;\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}}e^{-(x-\mu)^2/2\sigma^2};\ -\infty<x<\infty,\ -\infty<\mu<\infty,\ \sigma>0 \quad \dots (9.1)$$

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

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

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

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

and the corresponding distribution function, denoted by $\Phi(z)$ is given by :

$$\Phi(z) = P(Z \le z) = \int_{-\infty}^{z}\varphi(u)\,du = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{z}e^{-u^2/2}\,du$$

We shall prove below two important results on the distribution function $\Phi(.)$ of standard normal variate.

Result 1. $\Phi(-z) = 1 - \Phi(z),\ z > 0$

Proof. $\Phi(-z) = P(Z \le -z) = P(Z \ge z) = 1 - P(Z \le z) = 1 - \Phi(z)$

Result 2. $P(a \le X \le b) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$, where $X \sim N(\mu, \sigma^2)$

Proof. $$P(a \le X < b) = P\left(\frac{a-\mu}{\sigma} \le Z \le \frac{b-\mu}{\sigma}\right), \quad \left(Z = \frac{X-\mu}{\sigma}\right)$$
$$= P\left(Z \le \frac{b-\mu}{\sigma}\right) - P\left(Z \le \frac{a-\mu}{\sigma}\right) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$$

4. The graph of $f(x)$ is famous *'bell-shaped'* curve. The top of the bell is directly above the mean μ. For large values of σ, the curve tends to flatten out and for small values of σ, it has a sharp peak.

9·2·1. Normal Distribution as a Limiting form of Binomial Distribution.

Normal distribution is another limiting form of the binomial distribution under the following conditions :

(*i*) n, the number of trials is indefinitely large, *i.e.*, $n \to \infty$; and

(*ii*) neither p nor q is very small.

The *p.m.f.* of the binomial distribution with parameters n and p is given by :

$$p(x) = \binom{n}{x} p^x q^{n-x} = \frac{n!}{x!\,(n-x)!} p^x q^{n-x}\,; x = 0, 1, 2, \ldots, n \qquad \ldots (*)$$

Let us now consider the standard binomial variate :

$$Z = \frac{X - E(X)}{\sqrt{V(X)}} = \frac{X - np}{\sqrt{npq}}\,; X = 0, 1, 2, \ldots, n \qquad \ldots (**)$$

When $X = 0$, $Z = \dfrac{-np}{\sqrt{npq}} = -\sqrt{\dfrac{np}{q}}$ and when $X = n$, $Z = \dfrac{n-np}{\sqrt{npq}} = \sqrt{\dfrac{nq}{p}}$

Thus in the limit as $n \to \infty$, Z takes the values from $-\infty$ to ∞. Hence the distribution of X will be a continuous distribution over the range $-\infty$ to ∞.

We want the limiting form of (*) under the above two conditions. Using Stirling's approximation to $r!$ for large r, *viz.*, $\lim\limits_{r\to\infty} r\,! \simeq \sqrt{2\pi}\, e^{-r} r^{r+(1/2)}$,

we have in the limit as $n \to \infty$ and consequently $x \to \infty$,

$$\lim p(x) = \lim \left[\frac{\sqrt{2\pi}\, e^{-n} n^{n+\frac{1}{2}} p^x q^{n-x}}{\sqrt{2\pi}\, e^{-x} x^{x+\frac{1}{2}} \sqrt{2\pi}\, e^{-(n-x)} (n-x)^{n-x+\frac{1}{2}}}\right]$$

$$= \lim \left[\frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{npq}} \frac{(np)^{x+\frac{1}{2}} (nq)^{n-x+\frac{1}{2}}}{x^{x+\frac{1}{2}} (n-x)^{n-x+\frac{1}{2}}}\right]$$

$$= \lim \left[\frac{1}{\sqrt{2\pi}\sqrt{npq}} \left(\frac{np}{x}\right)^{x+\frac{1}{2}} \left(\frac{nq}{n-x}\right)^{n-x+\frac{1}{2}}\right] \qquad \ldots (***)$$

From (**), we get $\quad X = np + Z\sqrt{npq} \quad \Rightarrow \quad \dfrac{X}{np} = 1 + Z\sqrt{\dfrac{q}{np}}$

Further

$$n - X = n - np - Z\sqrt{npq} = nq - Z\sqrt{npq} \Rightarrow \frac{n-X}{nq} = 1 - Z\sqrt{\frac{p}{nq}}. \text{ Also } dz = \frac{1}{\sqrt{npq}}\, dx$$

Hence the probability differential of the distribution of Z, in the limit is :

$$dG(z) = g(z)\, dz = \lim_{n\to\infty} \left(\frac{1}{\sqrt{2\pi}} \times \frac{1}{N}\right) dz,$$

where $\quad N = \left(\frac{x}{np}\right)^{x+\frac{1}{2}} \left(\frac{n-x}{nq}\right)^{n-x+\frac{1}{2}}$ ... (9·2)

$$\Rightarrow \quad \log N = (x + \tfrac{1}{2}) \log (x/np) + (n - x + \tfrac{1}{2}) \log \{(n-x)/nq\}$$

$$= (np + z\sqrt{npq} + \tfrac{1}{2}) \log \left\{1 + z\sqrt{q/np}\right\} + (nq - z\sqrt{npq} + \tfrac{1}{2}) \log \left\{1 - z\sqrt{(p/nq)}\right\}$$

$$= (np + z\sqrt{npq} + \tfrac{1}{2}) \left\{z.\sqrt{(q/np)} - \tfrac{1}{2} z^2 (q/np) + \tfrac{1}{3} z^3 (q/np)^{3/2} - \ldots\right\}$$

$$+ (np - z\sqrt{npq} + \tfrac{1}{2}) \left\{-z\sqrt{(p/nq)} - \tfrac{1}{2} z^2 (p/nq) - \tfrac{1}{3} z^3 (p/nq)^{3/2} - \ldots\right\}$$

$$= \left[\left\{z\sqrt{npq} - \frac{1}{2} qz^2 + \frac{1}{3} z^3 \frac{q^{3/2}}{\sqrt{np}} + z^2 q - \frac{1}{2} z^3 \frac{q^{3/2}}{\sqrt{np}} + \frac{1}{2} z \sqrt{\frac{q}{np}} - \frac{1}{4} z^2 \frac{q}{np} + \ldots\right)\right.$$

$$\left. + \left\{-z\sqrt{npq} - \frac{1}{2} z^2 p - \frac{1}{3} z^3 \frac{p^{3/2}}{\sqrt{nq}} + z^2 p + \frac{1}{2} z^3 \frac{p^{3/2}}{\sqrt{np}} - \frac{1}{2} z\sqrt{p/nq} - \frac{1}{4} z^2 \frac{p}{np} + \ldots\right\}\right]$$

$$= \left[-\frac{1}{2} z^2 (p+q) + z^2 (p+q) + \frac{z}{2\sqrt{n}} \left(\frac{q}{p} + \frac{p}{q}\right) + 0\,(n^{-1/2})\right]$$

$$= \frac{z^2}{2} + 0\,(n^{-1/2}) \rightarrow \frac{z^2}{2} \text{ as } n \rightarrow \infty$$

$$\therefore \quad \lim_{n\to\infty} \log N = \frac{z^2}{2} \quad \Rightarrow \quad \lim_{n\to\infty} N = e^{z^2/2}$$

Substituting in (9·2), we get

$$dG(z) = g(z)\, dz = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\, dz, -\infty < z < \infty \qquad \ldots (9\cdot 2a)$$

Hence the probability function of Z is :

$$g(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}, -\infty < z < \infty \qquad \ldots (9\cdot 2b)$$

This is the probability density function of the *normal distribution* with mean 0 and unit variance.

If X is normal variate with mean μ and *s.d.* σ, then $Z = (X - \mu)/\sigma$, is standard normal variate. Jacobian of transformation is $1/\sigma$. Hence substituting in {9·2 (*b*)}, the *p.d.f.* of a normal variate X with $E(X) = \mu$, $\text{Var}(X) = \sigma^2$ is given by :

$$f_X(x) = \begin{cases} \dfrac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}, -\infty < x < \infty \\ 0, \text{ otherwise} \end{cases}$$

Remark. Normal distribution can also be obtained as a limiting case of Poisson distribution with the parameter $\lambda \rightarrow \infty$.

9·2. Chief Characteristics of the Normal Distribution and Normal Probability Curve. The normal probability curve with mean μ and standard deviation σ is given by the equation :

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

and has the following properties :

(*i*) The curve is bell-shaped and symmetrical about the line $x = \mu$.

(*ii*) Mean, median and mode of the distribution coincide.

Scanned with OKEN Scanner

(*iii*) As x increases numerically, $f(x)$ decreases rapidly, the maximum probability occurring at the point $x = \mu$, and is given by : $[p\,(x)]_{max} = \dfrac{1}{\sigma\sqrt{2\pi}}$

(*iv*) $\beta_1 = 0$ and $\beta_2 = 3$.

(*v*) $\mu_{2r+1} = 0, (r = 0, 1, 2, \ldots)$, and $\mu_{2r} = 1.3.5 \ldots (2r-1)\,\sigma^{2r}, (r = 0, 1, 2, \ldots)$

(*vi*) Since $f(x)$ being the probability, can never be negative, no portion of the curve lies below the x-axis.

(*vii*) Linear combination of independent normal variates is also a normal variate.

(*viii*) x-axis is an asymptote to the curve.

(*ix*) The points of inflexion of the curve are : $x = \mu \pm \sigma, f(x) = \dfrac{1}{\sigma\sqrt{2\pi}}\, e^{-1/2}$

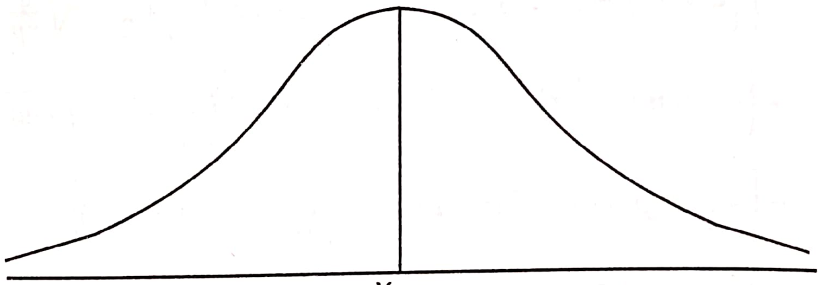


Fig. 9·1. Normal Probability Curve

(*x*) Mean deviation about mean $= \sqrt{\dfrac{2}{\pi}}\,\sigma \simeq \dfrac{4}{5}\,\sigma$ (approx.)

(*xi*) Quartiles are given by :

$$Q_1 = \mu - 0{\cdot}6745\,\sigma\,; \qquad Q_3 = \mu + 0{\cdot}6745\,\sigma$$

(*xii*) Q.D. $= \dfrac{Q_3 - Q_1}{2} \simeq \dfrac{2}{3}\,\sigma$. We have (approximately)

$$Q:D.:M.D.:S.D. :: \tfrac{2}{3}\sigma : \tfrac{4}{5}\sigma : \sigma :: \tfrac{2}{3} : \tfrac{4}{5} : 1 \Rightarrow Q.D.:M.D.:S.D. :: 10:12:15$$

(*xiii*) *Area Property* :

$$P(\mu - \sigma < X < \mu + \sigma) = 0{\cdot}6826, \qquad P(\mu - 2\sigma < X < \mu + 2\sigma) = 0{\cdot}9544,$$

and

$$P(\mu - 3\sigma < X < \mu + 3\sigma) = 0{\cdot}9973$$

The adjoining table gives the area under the normal probability curve for some important values of standard normal variate Z.

Distances from the mean ordinates in terms of $\pm\sigma$	*Area under the curve*
$Z = \pm 0{\cdot}745$	50% = 0·50
$Z = \pm 1{\cdot}00$	68·26% = 0·6826
$Z = \pm 1{\cdot}96$	95% = 0·95
$Z = \pm 2{\cdot}0$	95·44% = 0·9544
$Z = \pm 2{\cdot}58$	99% = 0·99
$Z = \pm 3{\cdot}0$	99·73% = 0·9973

(*xiv*) If X and Y are independent standard normal variates, then it can be easily proved that $U = X + Y$ and $V = X - Y$ are independently distributed, $U \sim N(0, 2)$ and $V \sim N(0, 2)$.

We state (without proof) the converse of this result which is due to *D. Bernstein*.

Bernstein's Theorem. *If X and Y are independent and identically distributed random variables with finite variances and if $U = X + Y$ and $V = X - Y$ are independent, then all r.v.'s X, Y, U and V are normally distributed.*

(*xiv*) We state below another result which characterises the normal distribution.

If $X_1, X_2, \ldots, X_n$ are *i.i.d. r.v.*'s with finite variance, then the common distribution is normal if and only if :

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \text{ and } s^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 \quad \text{or} \quad \sum_{i=1}^{n} X_i \text{ and } \sum_{i=1}^{n}(X_i - \bar{X})^2 \text{ are independent.}$$

In the following sequences we shall establish some of these properties.

9·2·3. Mode of Normal Distribution. Mode is the value of x for which $f(x)$ is maximum, *i.e.*, mode is the solution of

$$f'(x) = 0 \text{ and } f''(x) < 0$$

For normal distribution with mean μ and standard deviation σ,

$$\log f(x) = c - \frac{1}{2\sigma^2}(x-\mu)^2,$$

where $c = \log(1/\sqrt{2\pi}\,\sigma)$, is a constant. Differentiating *w.r.* to x, we get

$$\frac{1}{f(x)}\cdot f'(x) = -\frac{1}{\sigma^2}(x-\mu) \quad \Rightarrow \quad f'(x) = -\frac{1}{\sigma^2}(x-\mu)f(x)$$

and $$f''(x) = -\frac{1}{\sigma^2}\left[1.f(x) + (x-\mu)f'(x)\right] = -\frac{f(x)}{\sigma^2}\left[1 - \frac{(x-\mu)^2}{\sigma^2}\right] \qquad \ldots (9\cdot3)$$

$f'(x) = 0 \Rightarrow x - \mu = 0 \Rightarrow x = \mu$. At the point $x = \mu$, we have from (9·3) :

$$f''(x) = -\frac{1}{\sigma^2}[f(x)]_{x=\mu} = -\frac{1}{\sigma^2}\cdot\frac{1}{\sigma\sqrt{2\pi}} < 0$$

Hence $x = \mu$, is the mode of the normal distribution.

9·2·4. Median of Normal Distribution. If M is the median of the normal distribution, we have

$$\int_{-\infty}^{M} f(x)\,dx = \frac{1}{2} \quad \Rightarrow \quad \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{M} \exp\{-(x-\mu)^2/2\sigma^2\}\,dx = \frac{1}{2}$$

$$\Rightarrow \quad \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\mu} \exp\{-(x-\mu)^2/2\sigma^2\}\,dx + \frac{1}{\sigma\sqrt{2\pi}}\int_{\mu}^{M} \exp\{-(x-\mu)^2/(2\sigma^2)\}\,dx = \frac{1}{2} \qquad \ldots (9\cdot4)$$

But $$\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\mu} \exp\{-(x-\mu)^2/2\sigma^2\}\,dx = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{0} \exp(-z^2/2)\,dz = \frac{1}{2}$$

$\therefore$ From (9·4), we have $$\frac{1}{2} + \frac{1}{\sigma\sqrt{2\pi}}\int_{\mu}^{M} \exp\{-(x-\mu)^2/2\sigma^2\}\,dx = \frac{1}{2}$$

$$\Rightarrow \quad \frac{1}{\sigma\sqrt{2\pi}}\int_{\mu}^{M} \exp\{-(x-\mu)^2/2\sigma^2\}\,dx = 0, \quad i.e., \quad \mu = M.$$

Hence, for the normal distribution, Mean = Median.

Remark. From § 9·2·3. and § 9·2·4, we find that for the normal distribution mean, median and mode coincide. Hence the distribution is *symmetrical*.

9·2·5. M.G.F. of Normal Distribution. The *m.g.f.* (about origin) is given by :

$$M_X(t) = \int_{-\infty}^{\infty} e^{tx} f(x)\, dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tx} \exp\{-(x-\mu)^2/2\sigma^2\}\, dx$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\{t(\mu+\sigma z)\} \exp(-z^2/2)\, dz, \qquad \left(z = \frac{x-\mu}{\sigma}\right)$$

$$= e^{\mu t} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\{-\tfrac{1}{2}(z^2 - 2t\sigma z)\}\, dz$$

$$= e^{\mu t} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp[-\tfrac{1}{2}\{(z-\sigma t)^2 - \sigma^2 t^2\}] dz$$

$$= e^{\mu t + t^2\sigma^2/2} \times \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\{-\tfrac{1}{2}(z-\sigma t)^2\}\, dz$$

$$= e^{\mu t + t^2\sigma^2/2} \times \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp(-u^2/2)\, du$$

Hence $$M_X(t) = e^{\mu t + t^2\sigma^2/2} \qquad \ldots (9{\cdot}5)$$

Remark. M.G.F. *of Standard Normal Variate.* If $X \sim N(\mu, \sigma^2)$, then standard normal variate is given by :

$$Z = (X-\mu)/\sigma.$$

$$M_Z(t) = e^{-\mu t/\sigma} M_X(t/\sigma) = \exp(-\mu t/\sigma).\exp\{(\mu t/\sigma) + (t^2/\sigma^2)(\sigma^2/2)\} = \exp(t^2/2) \qquad \ldots (9{\cdot}5a)$$

Aliter $Z \sim N(0, 1)$. Hence, taking $\mu = 0$ and $\sigma^2 = 1$ in (9·5), we get :

$$M_Z(t) = \exp(t^2/2).$$

9·2·6. Cumulant Generating Function (c.g.f.) of Normal Distribution. The *c.g.f.* of normal distribution is given by :

$$K_X(t) = \log_e M_X(t) = \log_e (e^{\mu t + t^2\sigma^2/2}) = \mu t + \frac{t^2\sigma^2}{2}$$

$\therefore$ Mean $= \kappa_1 =$ Coefficient of t in $K_X(t) = \mu$

Variance $= \kappa_2 =$ Coefficient of $\frac{t^2}{2!}$ in $K_X(t) = \sigma^2$

and $\kappa_r =$ Coefficient of $\frac{t^r}{r!}$ in $K_X(t) = 0$; $r = 3, 4 \ldots$

Thus $\mu_3 = \kappa_3 = 0$ and $\mu_4 = \kappa_4 + 3\kappa_2^2 = 3\sigma^4$

Hence $\beta_1 = \dfrac{\mu_3^2}{\mu_2^3} = 0$ and $\beta_2 = \dfrac{\mu_4}{\mu_2^2} = 3 \qquad \ldots (9{\cdot}6)$

9·2·7. Moments of Normal Distribution. Odd order moments about mean are given by :

$$\mu_{2n+1} = \int_{-\infty}^{\infty} (x-\mu)^{2n+1} f(x)\, dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (x-\mu)^{2n+1} \exp\{-(x-\mu)^2/2\sigma^2\} dx$$

$$\therefore \quad \mu_{2n+1} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (\sigma z)^{2n+1} \exp(-z^2/2)\, dz, \qquad \left(z = \frac{x-\mu}{\sigma}\right)$$

$$= \frac{\sigma^{2n+1}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z^{2n+1} \exp(-z^2/2)\, dz = 0, \qquad \ldots (9{\cdot}7)$$

since the integrand $z^{2n+1} e^{-z^2/2}$ is an odd function of z.

Even order moments about mean are given by :

$$\mu_{2n} = \int_{-\infty}^{\infty} (x-\mu)^{2n} f(x)\,dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (\sigma z)^{2n} \exp(-z^2/2)\,dz$$

$$= \frac{\sigma^{2n}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z^{2n} \exp(-z^2/2)\,dz = \frac{\sigma^{2n}}{\sqrt{2\pi}} \cdot 2 \int_{0}^{\infty} z^{2n} \exp(-z^2/2)\,dz$$

(Since integrand is an even function of z.)

$$= \frac{2\sigma^{2n}}{\sqrt{2\pi}} \cdot \int_0^{\infty} (2t)^n e^{-t} \frac{dt}{\sqrt{2t}}, \qquad \left(t = \frac{z^2}{2}\right)$$

$$\therefore \quad \mu_{2n} = \frac{2^n \sigma^{2n}}{\sqrt{\pi}} \int_0^{\infty} e^{-t}\, t^{(n+\frac{1}{2})-1}\,dt \quad \Rightarrow \quad \mu_{2n} = \frac{2^n \sigma^{2n}}{\sqrt{\pi}} \cdot \Gamma\left(n+\frac{1}{2}\right)$$

Changing n to $(n-1)$, we get

$$\mu_{2n-2} = \frac{2^{n-1} \cdot \sigma^{2n-2}}{\sqrt{\pi}} \Gamma\left(n-\frac{1}{2}\right)$$

$$\therefore \quad \frac{\mu_{2n}}{\mu_{2n-2}} = 2\sigma^2 \cdot \frac{\Gamma\left(n+\frac{1}{2}\right)}{\left(n-\frac{1}{2}\right)} = 2\sigma^2\left(n-\frac{1}{2}\right) \qquad [\because \Gamma(r) = (r-1)\,\Gamma(r-1)]$$

$$\Rightarrow \quad \mu_{2n} = \sigma^2 (2n-1)\, \mu_{2n-2} \qquad \ldots (9{\cdot}8)$$

which gives the *recurrence relation* for the moments of normal distribution.

From (9·8), we have

$$\mu_{2n} = [(2n-1)\sigma^2]\,[(2n-3)\sigma^2]\,\mu_{2n-4}$$
$$= [(2n-1)\sigma^2]\,[2n-3)\sigma^2]\,[2n-5)\sigma^2]\,\mu_{2n-6}$$
$$= [(2n-1)\sigma^2]\,[(2n-3)\sigma^2]\,[(2n-5)\sigma^2] \ldots (3\sigma^2)(1\sigma^2).\,\mu_0$$
$$= 1.3.5. \ldots (2n-1)\,\sigma^{2n} \qquad \ldots (9{\cdot}9)$$

From (9·7) and (9·9), we conclude that *for the normal distribution all odd order moments about mean vanish and even order moments about mean are given by* (9·9).

Aliter. The above result can also be obtained quite conveniently as follows :

The *m.g.f.* (about mean) is given by : $E[e^{t(X-\mu)}] = e^{-\mu t} E(e^{tX}) = e^{-\mu t} M_X(t)$,

where $M_X(t)$ is the *m.g.f.* (about origin).

$\therefore$ *m.g.f.* (about mean) $= e^{-\mu t} e^{\mu t + t^2\sigma^2/2} = e^{t^2\sigma^2/2}$

$$= \left[1 + (t^2\sigma^2/2) + \frac{(t^2\sigma^2/2)^2}{2!} + \frac{(t^2\sigma^2/2)^3}{3!} + \ldots + \frac{(t^2\sigma^2/2)^n}{n!} + \ldots\right] \qquad \ldots (9{\cdot}10)$$

The coefficient of $\frac{t^r}{r!}$ in (9·10) gives μ_r, the rth moment about mean. Since there is no term with odd powers of it t (9·10), all moments of odd order about mean vanish, i.e.,

$$\mu_{2n+1} = 0;\ n = 0, 1, 2, \ldots$$

and μ_{2n} = Coefficient of $\frac{t^{2n}}{(2n)!}$ in (9·10) $= \frac{\sigma^{2n} \times (2n)!}{2^n\, n!}$

$$= \frac{\sigma^{2n}}{2^n\, n!}\left[2n\,(2n-1)\,(2n-2)\,(2n-3) \ldots 5.4.3.2.1\right]$$

$$= \frac{\sigma^{2n}}{2^n . n!}\left[1.3.5 \ldots (2n-1)\right]\left[2.4.6 \ldots (2n-2).2n\right]$$

$$= \frac{\sigma^{2n}}{2^n \cdot n!}\left[1.3.5.\ldots(2n-1)\right]2^n\left[1.2.3\ldots n\right]$$

$$= 1.3.5\ldots(2n-1)\,\sigma^{2n}$$

Remark. In particular, from (9·7) and (9·9), $\mu_3 = 0$ and $\mu_2 = \sigma^2$, $\mu_4 = 1.3\,\sigma^4 = 3\sigma^4$

Hence $\beta_1 = \frac{\mu_3^2}{\mu_2^3} = 0$ and $\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{3\sigma^4}{\sigma^4} = 3$, the results which have already been obtained in (9·4).

9·2·8. A linear combination of independent normal variates is also a normal variate. Let X_i, $(i = 1, 2, 3, \ldots, n)$ be n independent normal variates with mean μ_i and variance σ_i^2 respectively. Then

$$M_{X_i}(t) = \exp\{\mu_i t + (t^2\sigma_i^2/2)\} \qquad \ldots(9\cdot11)$$

The *m.g.f.* of their linear combination $\sum_{i=1}^{n} a_i X_i$, where $a_1, a_2, \ldots, a_n$ are constants, is given by :

$$M_{\Sigma a_i X_i}(t) = \prod_{i=1}^{n} M_{a_i X_i}(t) \qquad (\because X_i\text{'s are independent})$$

$$= M_{X_1}(a_1 t) . M_{X_2}(a_2 t) \ldots M_{X_n}(a_n t) \qquad [\because M_{cX}(t) = M_X(ct)] \qquad \ldots(9\cdot12)$$

From (9·11), we have $M_{X_i}(a_i t) = e^{\mu_i a_i t + t^2 a_i^2 \sigma_i^2/2}$

$$\therefore M_{\Sigma a_i X_i}(t) = \left[e^{\mu_1 a_1 t + t^2 a_1^2 \sigma_1^2/2} \times e^{\mu_2 a_2 t + t^2 a_2^2 \sigma_2^2/2} \times \ldots \times e^{\mu_n a_n t + t^2 a_n^2 \sigma_n^2/2}\right] \qquad [\text{From } (9\cdot12)]$$

$$= \exp\left[\left(\sum_{i=1}^{n} a_i \mu_i\right)t + t^2\left(\sum_{i=1}^{n} a_i^2 \sigma_i^2\right)/2\right],$$

which is the *m.g.f.* of a normal variate with mean $\sum_{i=1}^{n} a_i \mu_i$ and variance $\sum_{i=1}^{n} a_i^2 \sigma_i^2$.

Hence by uniqueness theorem of *m.g.f.*,

$$\sum_{i=1}^{n} a_i X_i \sim N\left[\sum_{i=1}^{n} a_i \mu_i, \sum_{i=1}^{n} a_i^2 \sigma_i^2\right] \qquad \ldots(9\cdot12a)$$

***Remarks* 1.** If we take $a_1 = a_2 = 1, a_3 = a_4 = \ldots = 0$, then $X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$

If we take $a_1 = 1, a_2 = -1, a_3 = a_4 = \ldots = 0$, then $X_1 - X_2 \sim N(\mu_1 - \mu_2, \sigma_1^2 + \sigma_2^2)$

Thus we see that *the sum as well as the difference of two independent normal variates is also a normal variate.* This result provides a sharp contrast to the Poisson distribution, in which case though the sum of two independent Poisson variates is a Poisson variate, the difference is not a Poisson variate.

2. If we take $a_1 = a_2 = \ldots = a_n = 1$, then we get $\sum_{i=1}^{n} X_i \sim N\left[\sum_{i=1}^{n} \mu_i, \sum_{i=1}^{n} \sigma_i^2\right] \qquad \ldots(9\cdot12b)$

i.e., the sum of independent normal variates is also a normal variate, which establishes the *additive property* of the normal distribution.

3. If X_i; $i = 1, 2, \ldots, n$ are identically and independently distributed as $N(\mu, \sigma^2)$ and if we take $a_1 = a_2 = \ldots = a_n = 1/n$, then

$$\frac{1}{n}\sum_{i=1}^{n} X_i \sim N\left(\frac{1}{n}\sum_{i=1}^{n}\mu, \frac{1}{n^2}\sum_{i=1}^{n}\sigma^2\right) \Rightarrow \bar{X} \sim N(\mu, \sigma^2/n), \text{ where } \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

This leads to the following important conclusion :

If X_i ($i = 1, 2, \ldots, n$), are identically and independently distributed normal variates with mean μ and variance σ^2, then their mean $\bar{X}$ is also $N(\mu, \sigma^2/n)$.

9·2·9. Points of Inflexion of Normal Curve. At the point of inflexion of the normal curve, we should have $f''(x) = 0$, and $f'''(x) \neq 0$.

For normal curve, we have from (9·3), $f''(x) = -\dfrac{f(x)}{\sigma^2}\left\{1 - \dfrac{(x-\mu)^2}{\sigma^2}\right\}$

$$\therefore \quad f''(x) = 0 \quad \Rightarrow \quad 1 - \frac{(x-\mu)^2}{\sigma^2} = 0 \quad \Rightarrow \quad x = \mu \pm \sigma.$$

It can be easily verified that at the points $x = \mu \pm \sigma$, $f'''(x) \neq 0$. Hence the points of inflexion of the normal curve are given by $x = \mu \pm \sigma$ and $f(x) = \dfrac{1}{\sigma\sqrt{2\pi}} e^{-1/2}$, *i.e.,* they are equi-distant (at a distance σ) from the mean.

9·2·10. Mean Deviation About the Mean for Normal Distribution.

$$\text{M.D. (about mean)} = \int_{-\infty}^{\infty} |x-\mu| \, f(x)\, dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} |x-\mu| \, e^{-(x-\mu)^2/2\sigma^2} dx$$

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

$$= \frac{2\sigma}{\sqrt{2\pi}} \cdot \int_{0}^{\infty} |z| \, e^{-z^2/2} dz,$$

($\because$ The integrand $|z| \, e^{-z^2/2}$ is an even function of z.)

Since in $[0, \infty]$, $|z| = z$, we have

$$\text{M.D. (about mean)} = \sqrt{2/\pi}\,\sigma \int_0^\infty e^{-z^2/2} dz \;\; \sqrt{2/\pi}\,\sigma \int_0^\infty e^{-t} dt, \qquad (z^2/2 = t)$$

$$= \sqrt{2/\pi}\,\sigma \left| \frac{e^{-t}}{-1} \right|_0^\infty = \sqrt{2/\pi}\,\sigma = \frac{4}{5}\sigma \text{ (approx.)}$$

9·2·11. Area Property (Normal Probability Integral). If $X \sim N(\mu, \sigma^2)$, then the probability that random value of X will lie between $X = \mu$ and $X = x_1$ is given by :

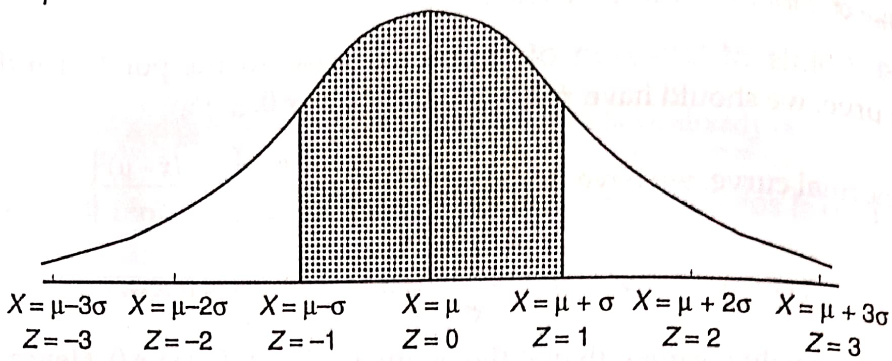
$$P(\mu < X < x_1) = \int_{\mu}^{x_1} f(x)\, dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{\mu}^{x_1} e^{-(x-\mu)^2/2\sigma^2} dx$$

Put $\quad \dfrac{X-\mu}{\sigma} = Z \quad \Rightarrow \quad X - \mu = \sigma Z$

When $X = \mu$, $Z = 0$ and when $X = x_1$, $Z = \dfrac{x_1 - \mu}{\sigma} = z_1$, (say).

$$\therefore \quad P(\mu < X < x_1) = P(0 < Z < z_1) = \frac{1}{\sqrt{2\pi}} \int_0^{z_1} e^{-z^2/2} dz = \int_0^{z_1} \varphi(z)\, dz$$

where $\varphi(z) = \dfrac{1}{\sqrt{2\pi}} e^{-z^2/2}$, is the probability function of standard normal variate. The definite integral $\displaystyle\int_0^{z_1} \varphi(z)\, dz$ is known as *normal probability integral* and gives the area under standard normal curve between the ordinates at $Z = 0$ and $Z = z_1$. These areas have been tabulated for different values of z_1, at intervals of 0·01 in a table given at *the end of the chapter.*



In particular, the probability that a random value of X lies in the interval $(\mu - \sigma, \mu + \sigma)$ is given by :

$$P(\mu - \sigma < X < \mu + \sigma) = \int_{\mu-\sigma}^{\mu+\sigma} f(x)dx$$

$$\Rightarrow \quad P(-1 < Z < 1) = \int_{-1}^{1} \varphi(z)\, dz, \qquad \left[z = \frac{x-\mu}{\sigma}\right]$$

$$= 2\int_0^1 \varphi(z)\, dz \qquad \text{(By symmetry)}$$

$$= 2 \times 0{\cdot}3413 = 0{\cdot}6826 \qquad \text{(From Tables)} \quad \ldots (9{\cdot}14)$$

Similarly, $P(\mu - 2\sigma < X < \mu + 2\sigma) = P(-2 < Z < 2) = \int_{-2}^{2} \varphi(z)\, dx$

$$= 2\int_0^2 \varphi(z)\, dz = 2 \times 0{\cdot}4772 = 0{\cdot}9544 \qquad \ldots (9{\cdot}15)$$

and $\quad P(\mu - 3\sigma < X < \mu + 3\sigma) = P(-3 < Z < 3) = \int_{-3}^{3} \varphi(z)\, dz$

$$= 2\int_0^3 \varphi(z)\, dz = 2 \times 0{\cdot}49865 = 0{\cdot}9973 \qquad \ldots (9{\cdot}16)$$

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

$$P(|X - \mu| > 3\sigma) = P(|Z| > 3) = 1 - P(-3 \le Z \le 3) = 0{\cdot}0027$$

Thus in all probability, we should expect a normal variate to lie within the range $\mu \pm 3\sigma$, though theoretically, it may range from $-\infty$ to ∞.

Remarks **1.** The total area under normal probability curve is unity, *i.e.*,

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

2. Since in the normal probability tables, we are given the areas under standard normal curve, in numerical problems we shall deal with the standard normal variate Z rather than the variable X itself.

3. If we want to find area under normal curve, we will somehow or other try to convert the given area to the form $P(0 < Z < z_1)$, since the areas have been given in this form in the Tables.

9·2·12. Error Function. If $X \sim N(0, \sigma^2)$, then $f(x) = \dfrac{1}{\sigma\sqrt{2\pi}} e^{-x^2/2\sigma^2}, -\infty < x < \infty$

If we take $h^2 = \dfrac{1}{2\sigma^2}$, then
$$f(x) = \frac{h}{\sqrt{\pi}} e^{-h^2x^2}$$

The probability P that a random value of the variate lies in the range $\pm x$ is :

$$P = \int_{-x}^{x} f(x)\, dx = \frac{h}{\sqrt{\pi}} \int_{-x}^{x} e^{-h^2x^2}\, dx = \frac{2h}{\sqrt{\pi}} \int_0^x e^{-h^2x^2}\, dx = \frac{2}{\sqrt{\pi}} \int_0^x e^{-h^2x^2}\, (hdx) \quad \ldots (*)$$

Taking $\psi(y) = \dfrac{2}{\sqrt{\pi}} \displaystyle\int_0^y e^{-y^2}\, dy$, (*) may be re-written as :

$$P = \psi(hx) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-h^2x^2}\, (hdx) \quad \ldots (**)$$

The function $\psi(y)$, known as the *error function*, is of fundamental importance in the *theory of errors* in Astronomy.

9·2·13. Importance of Normal Distribution. Normal distribution plays a very important role in statistical theory because of the following reasons :

(*i*) Most of the distributions occurring in practice, *e.g.*, Binomial, Poisson, Hyper-geometric distributions, etc., can be approximated by normal distribution . Moreover, many of the sampling distributions, *e.g.*, Student's t, Snedecor's F, Chi-square distributions, etc., tend to normality for large samples.

(*ii*) Even if a variable is not normally distributed, it can sometimes be brought to normal form by simple transformation of variable. For example, if the distribution of X is skewed, the distribution of $\sqrt{X}$ might come out to be normal [*c.f.* Variate Transformations, § 9·13 at the end of this Chapter].

(*iii*) If $X \sim N(\mu, \sigma^2)$, then $P(\mu - 3\sigma \le X \le \mu + 3\sigma) = P(-3 < Z < 3) = 0{\cdot}9973$

$\therefore \quad P(|Z| > 3) = 1 - P(|Z| \le 3) = 0{\cdot}0027$

This property of the normal distribution forms the basis of entire *Large Sample* theory.

(*iv*) Many of the distributions of sample statistics (*e.g.*, the distributions of sample mean, sample variance, etc.) tend to normality for large samples and as such they can best be studied with the help of the normal curves.

(*v*) The entire theory of small sample tests, *viz.*, t, F, χ^2 tests, etc., is based on the fundamental assumption that the parent populations from which the samples have been drawn follow normal distribution.

(*vi*) Normal distribution finds large applications in Statistical Quality Control in industry for setting control limits.

The following quotation due to Lipman rightly reveals the popularity and importance of normal distribution :

"Every body believes in the law of errors (the normal curve), the experimenters because they think it is a mathematical theorem, the mathematicians because they think it is experimental fact."

W.J. Youden of the National Bureau of Standards describes the importance of the Normal distribution artistically in the following words :

THE NORMAL
LAW OF ERRORS
STANDS OUT IN THE
EXPERIENCE OF MANKIND
AS ONE OF THE BROADEST
GENERALISATIONS OF NATURAL
PHILOSOPHY. IT SERVES AS THE
GUIDING INSTRUMENT IN RESEARCHES,
IN THE PHYSICAL AND SOCIAL SCIENCES
AND IN MEDICINE, AGRICULTURE AND
ENGINEERING. IT IS AN INDISPENSABLE TOOL FOR
THE ANALYSIS AND THE INTERPRETATION OF THE
BASIC DATA OBTAINED BY OBSERVATION AND EXPERIMENT.

9·2·14. Fitting of Normal Distribution. In order to fit normal distribution to the given data we first calculate the mean μ, (say), and standard deviation σ (say) from the given data. Then the normal curve fitted to the given data is given by :

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

To calculate the expected normal frequencies we first find the standard normal variates corresponding to the lower limits of each of the class intervals, *i.e.,* we compute $z_i = (x_i' - \mu)/\sigma$, where x_i' is the lower limit of the ith class interval. Then the areas under the normal curve to the left of the ordinate at $z = z_i$, say, $\Phi(z_i) = P(Z \le z_i)$ are computed from the tables. Finally, the areas for the successive class intervals are obtained by subtraction, *viz.,* $\Phi(z_i + 1) - \Phi(z_i)$, $(i = 1, 2, \ldots)$ and on multiplying these areas by N, we get the expected normal frequencies.

Example 9·1. *Obtain the equation of the normal curve that may be fitted to the following data :*

Class :	60—65	65—70	70—75	75—80	80—85	85—90	90—95	95—100
Frequency :	3	21	150	335	326	135	26	4

Also obtain the expected normal frequencies.

Solution. For the given data, $N = 1000$, $\mu = 79{\cdot}945$ and $\sigma = 5{\cdot}545$ (Try it.)
Hence the equation of the normal curve fitted to the given data is :

$$f(x) = \frac{1000}{\sqrt{2\pi} \times 5{\cdot}545} \exp\left\{-\frac{1}{2}\left(\frac{x - 79{\cdot}945}{5{\cdot}545}\right)^2\right\}$$

COMPUTATION OF THEORETICAL NORMAL FREQUENCIES

Class	*Lower class boundary* (X′)	$z = \frac{X' - \mu}{\sigma}$	$\Phi(z) = \frac{1}{\sqrt{2\pi}} \times \int_{-\infty}^{z} e^{-u^2/2}\, du$	$\Delta\,\Phi(Z) = \Phi_{Z+1} - \Phi_Z$	*Expected frequency* $= N\,\Delta\,\Phi\,(z)$
Below 60	$-\infty$	$-\infty$	0	0·000112	0·12 ≅ 0
60—65	60	– 3·663	0·000112	0·002914	2·914 ≅ 3
65—70	65	– 2·745	0·003026	0·031044	31·044 ≅ 31
70—75	70	– 1·826	0·034070	0·147870	147·870 ≅ 148
75—80	75	– 0·908	0·181940	0·322050	322·050 ≅ 322
80—85	80	0·010	0·503990	0·919300	319·300 ≅ 319
85—90	85	0·928	0·823290	0·144072	144·072 ≅ 144
90—95	90	1·487	0·967362	0·029792	29·792 ≅ 30
95—100	95	2·675	0·997154	0·002733	2·733 ≅ 3
100 and over	100	3·683	0·999887		
Total					1,000

Example 9·2. *For a certain normal distribution, the first moment about 10 is 40 and the fourth moment about 50 is 48. What is the arithmetic mean and standard deviation of the distribution ?*

Solution. We know that if μ_1' is the first moment about the point $X = A$, then arithmetic mean is given by : Mean $= A + \mu_1'$

We are given : μ_1' (about the point $X = 10) = 40 \quad \Rightarrow \quad$ Mean $= 10 + 40 = 50$

Also μ_4' (about the point $X = 50) = 48$, *i.e.,* $\mu_4 = 48$ $\qquad$ ($\because$ Mean $= 50$)

But for a normal distribution with standard deviation σ,

$$\mu_4 = 3\,\sigma^4 \quad \Rightarrow \quad 3\,\sigma^4 = 48 \quad \Rightarrow \quad \sigma = 2.$$

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

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

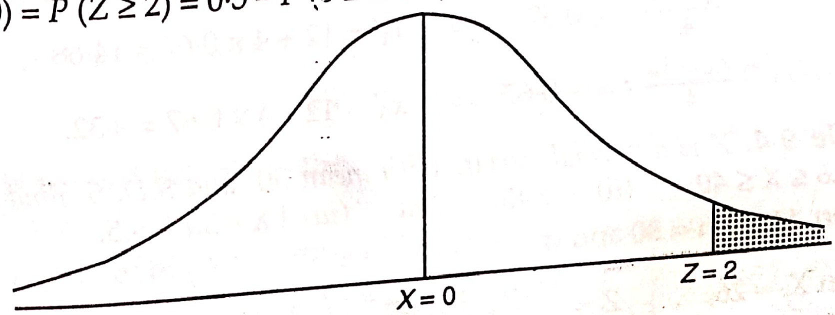
(c) Find x_0' and x_1', when $P(x_0' < X < x_1') = 0{\cdot}50$ and $P(X > x_1') = 0{\cdot}25$.

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

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

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

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



(ii) $P(X \le 20) = 1 - P(X \ge 20) = 1 - 0{\cdot}0228 = 0{\cdot}9722$ [From part (i)]

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

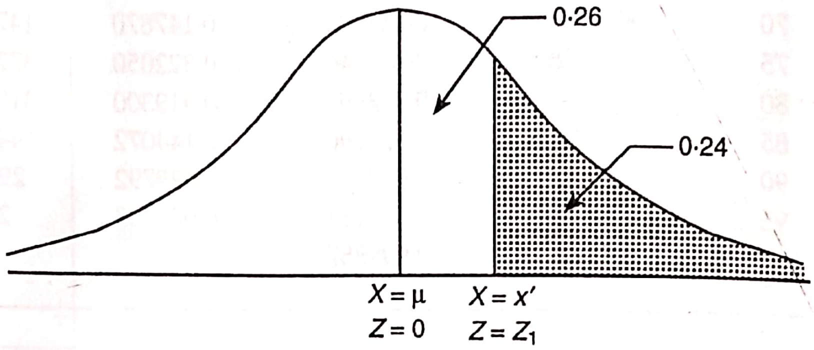
$= P(0 \le Z \le 3) = 0{\cdot}49865$ (From symmetry)

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

Then, we are given :

$$P(X > x') = 0{\cdot}24 \quad \Rightarrow \quad P(Z > z_1) = 0{\cdot}24 \quad \Rightarrow \quad P(0 < Z < z_1) = 0{\cdot}26$$

$\therefore$ From Normal Tables, $z_1 = 0{\cdot}71$ (approx.)



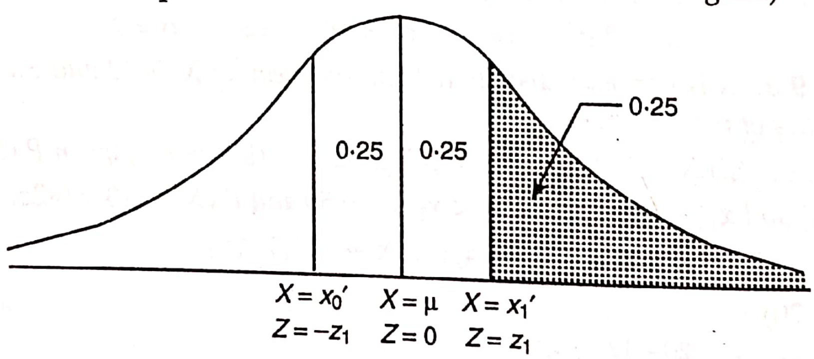
Hence $\dfrac{x' - 12}{4} = 0{\cdot}71 \quad \Rightarrow \quad x' = 12 + 4 \times 0{\cdot}71 = 14{\cdot}84.$

(c) We are given : $P(x_0' < X < x_1') = 0{\cdot}50$ and $P(X > x_1') = 0{\cdot}25$ ... (*)

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

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

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



We have $P(Z > z_1) = 0{\cdot}25 \quad \Rightarrow \quad P(0 < Z < z_1) = 0{\cdot}25 \quad \therefore z_1 = 0{\cdot}67$ (From Tables)

Hence $\dfrac{x_1' - 12}{4} = 0{\cdot}67 \quad \Rightarrow \quad x_1' = 12 + 4 \times 0{\cdot}67 = 14{\cdot}68$

and $\dfrac{x_0' - 12}{4} = -0{\cdot}67 \quad \Rightarrow \quad x_0' = 12 - 4 \times 0{\cdot}67 = 9{\cdot}32.$

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

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

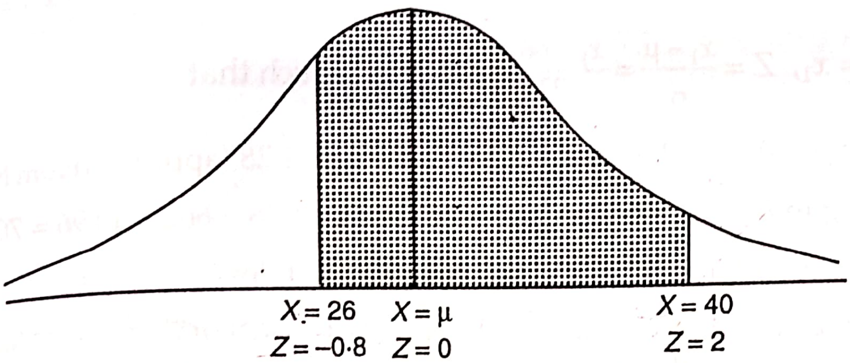
(i) When $X = 26$, $Z = \dfrac{X - \mu}{\sigma} = \dfrac{26 - 30}{5} = -0{\cdot}8$

and when $X = 40, Z = \frac{40-30}{5} = 2$

$$\therefore P(26 \le X \le 40) = P(-0{\cdot}8 \le Z \le 2) = P(-0{\cdot}8 \le Z \le 0) + P(0 \le Z \le 2)$$

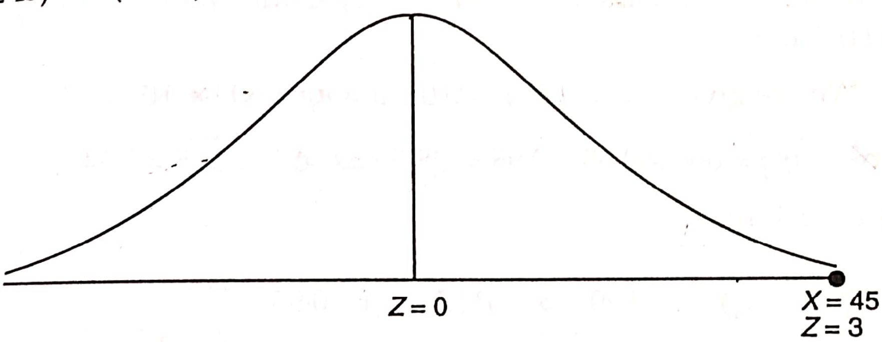
$$= P(0 \le Z \le 0{\cdot}8) + (0 \le Z \le 2) = 0{\cdot}2881 + 0{\cdot}4772 \quad \text{[By symmetry]}$$

$$= 0{\cdot}7653 \quad \text{(From Normal Tables)}$$



(ii) When $x = 45, Z = \frac{45-30}{5} = 3.$

$$\therefore P(X \ge 45) = P(Z \ge 3) = 0{\cdot}5 - P(0 \le Z \le 3) = 0{\cdot}5 - 0{\cdot}49865 = 0{\cdot}00135$$



(iii) $P(|X - 30| \le 5) = P(25 \le X \le 35) = P(-1 \le Z \le 1)$

$$= 2P(0 \le Z \le 1) = 2 \times 0{\cdot}3413 = 0{\cdot}6826$$

$$\therefore \quad P(|X - 30| > 5) = 1 - P(|X - 30| \le 5) = 1 - 0{\cdot}6826 = 0{\cdot}3174.$$

Example 9·5. *The mean yield for one-acre plot is 662 kilos with a s.d. 32 kilos. Assuming normal distribution, how many one-acre plots in a batch of 1,000 plots would you expect to have yield (i) over 700 kilos, (ii) below 650 kilos, and (iii) what is the lowest yield of the best 100 plots ?*

Solution. If the *r.v.* X denotes the yield (in kilos) for one-acre plot, the we are given that $X \sim N(\mu, \sigma^2)$, where $\mu = 662$ and $\sigma = 32$.

(i) The probability that a plot has a yield over 700 kilos is given by

$$P(X > 700) = P(Z > 1{\cdot}19), \qquad Z = \frac{700-662}{32}$$

$$= 0{\cdot}5 - P(0 \le Z \le 1{\cdot}19) = 0{\cdot}5 - 0{\cdot}3830 = 0{\cdot}1170$$

Hence in a batch of 1,000 plots, the expected number of plots with yield over 700 kilos is $1{,}000 \times 0{\cdot}117 = 117$.

(ii) Required number of plots with yield below 650 kilos is given by :

$$1000 \times P(X < 650) = 1{,}000 \times P(Z < -0{\cdot}38) \qquad \left(Z = \frac{650-662}{32}\right)$$

$$= 1{,}000 \times P(Z > 0{\cdot}38), \qquad \text{(By symmetry)}$$
$$= 1{,}000 \times [0{\cdot}5 - P(0 \le Z \le 0{\cdot}38)]$$
$$= 1{,}000 \times (0{\cdot}5 - 0{\cdot}1480) = 1000 \times 0{\cdot}352 = 352$$

(*iii*) The lowest yield, say, x_1 of best 100 plots is given by : $P(X > x_1) = \frac{100}{1000} = 0{\cdot}1$

When $X = x_1,\ Z = \frac{x_1 - \mu}{\sigma} = \frac{x_1 - 662}{32} = z_1$, (say), such that ... (*)

$P(Z > z_1) = 0{\cdot}1) \Rightarrow P(0 \le Z \le z_1) = 0{\cdot}4 \Rightarrow z_1 = 1{\cdot}28$ (approx.) [From Normal Tables]

Substituting in (*), $x_1 = 662 + 32 z_1 = 662 + 32 \times 1{\cdot}28 = 662 + 40{\cdot}96 = 702{\cdot}96$

Hence the best 100 plots have yield over 702·96 kilos.

Example 9·6. *There are six hundred Economics students in the post-graduate classes of a university, and the probability for any student to need a copy of a particular book from the university library on any day is 0·05. How many copies of the book should be kept in the university library so that the probability may be greater than 0·90 that none of the students needing a copy from the library has to come back disappointed ? (Use normal approximation to the binomial distribution.)*

Solution. We are given : $n = 600, p = 0{\cdot}05, \mu = np = 600 \times 0{\cdot}05 = 34$

and $\sigma^2 = npq = 600 \times 0{\cdot}05 \times 0{\cdot}95 = 28{\cdot}5 \Rightarrow \sigma = \sqrt{28{\cdot}5} = 5{\cdot}34.$

We want x_1 such that :

$$P(X < x_1) > 0{\cdot}90 \Rightarrow P(Z < z_1) > 0{\cdot}90 \qquad \left[z_1 = \frac{x_1 - 30}{5{\cdot}34}\right]$$

$$\Rightarrow P(0 < Z < z_1) > 0{\cdot}40 \Rightarrow z_1 > 1{\cdot}28 \qquad \text{(From normal probability Tables)}$$

$$\Rightarrow \frac{x_1 - 30}{5{\cdot}34} > 1{\cdot}28 \Rightarrow x_1 > 30 + 5{\cdot}34 \times 1{\cdot}28, \quad i.e., \quad x_1 > 36{\cdot}84 \simeq 37.$$

Hence the university library should keep at least 37 copies of the book.

Example 9·7. *The local authorities in a certain city instal 10,000 electric lamps in the streets of the city. If these lamps have an average life of 1,000 burning hours with a standard deviation of 200 hours, assuming normality, what number of lamps might be expected to fail (i) in the first 800 burning hours ? (ii) between 800 and 1,200 burning hours ? After what period of burning hours would you expect that (a) 10% of the lamps would fail ? (b) 10% of the lamps would be still burning ?*

[In a normal curve, the area between the ordinates corresponding to $[(X-\mu)/\sigma] \doteq 0$ and $[(X-\mu)/\sigma] = 1$ is 0·34134 and 80% of the area lies between the ordinates corresponding to $\{(X-\mu)/\sigma\} = \pm 1{\cdot}28$.]

Solution. If the variable X denotes the life of a bulb in burning hours, then we are given that $X \sim N(\mu, \sigma^2)$, where $\mu = 1{,}000$ and $\sigma = 200$.

(*i*) The probability 'p' that bulb fails in the first 800 burning hours is given by :

$$p = P(X < 800) = P(Z < -1) = P(Z > 1), \qquad \left(Z = \frac{800 - 1000}{200}\right)$$

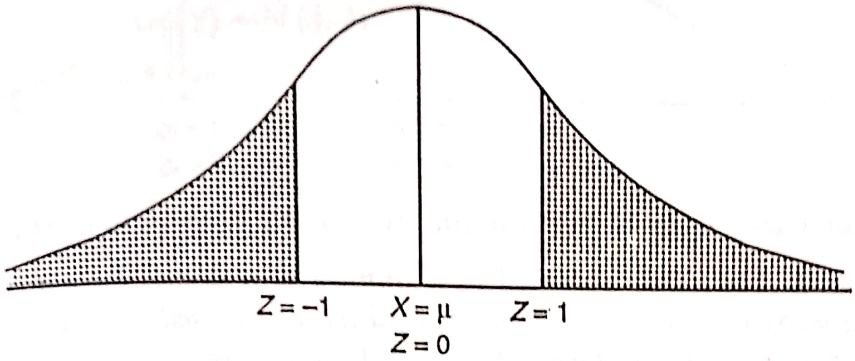
$= 0{\cdot}5 - P(0 < Z < 1) = 0{\cdot}5 - 0{\cdot}3413 = 0{\cdot}1587$

$\therefore$ Out of 10,000 bulbs, number of bulbs which fail in the first 800 hours

$$= 10{,}000 \times 0{\cdot}1587 = 1{,}587$$

(ii) Required probability

$$= P(800 < X < 1{,}200) = P(-1 < Z < 1) = 2\,P(0 < Z < 1) = 2 \times 0{\cdot}3413 = 0{\cdot}6826$$



Hence the expected number of bulbs with life between 800 and 1,200 hours of burning life is : $10{,}000 \times 0{\cdot}6826 = 6{,}826.$

(a) Let 10% of the bulbs fail after x_1 hours of burning life. Then we have to find x_1 such that $P(X < x_1) = 0{\cdot}10$

When $X = x_1$, $\quad Z = \dfrac{x_1 - 1000}{200} = -z_1$ (say). ... (*)

$\therefore P(Z < -z_1) = 0{\cdot}10 \Rightarrow P(Z > z_1) = 0{\cdot}10 \Rightarrow P(0 < Z < z_1) = 0{\cdot}40$ ... (**)

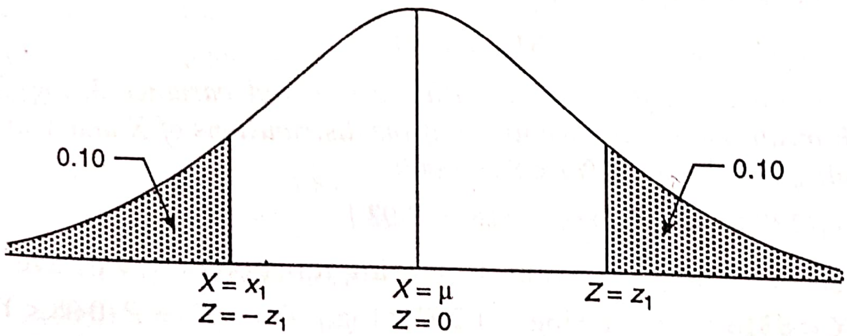
Also $\quad P(-1{\cdot}28 < z < 1{\cdot}28) = 0{\cdot}80$ (Given)

$\Rightarrow \quad 2\,P(0 < Z < 1{\cdot}28) = 0{\cdot}80 \Rightarrow P(0 < Z < 1{\cdot}28) = 0{\cdot}40$ ... (***)

From(**) and (***), we get $\quad z_1 = 1{\cdot}28$

$\therefore \quad \dfrac{x_1 - 1000}{200} = -1{\cdot}28 \qquad$ [From (*)]

$\Rightarrow \quad x_1 = 1{,}000 - 256 = 744$



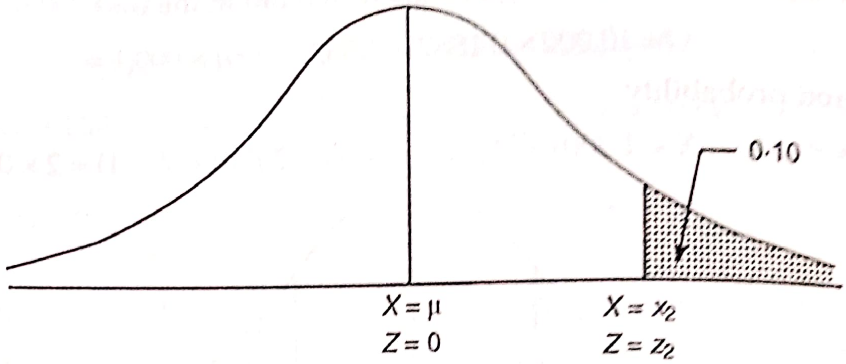
Thus, after 744 hours of burning life, 10% of the blubs will fail.

(b) Let 10% of the bulbs be still burning after (say) x_2 hours of burning life. Then we have

$\therefore \quad P(X > x_2) = 0{\cdot}10 \Rightarrow P(Z > z_2) = 0{\cdot}10, \left(z_2 = \dfrac{x_2 - 1000}{200}\right) \Rightarrow P(0 < Z < z_2) = 0{\cdot}40$

$z_2 = 1{\cdot}28, \qquad$ [From (***)]

i.e., $\dfrac{x_2 - 1000}{200} = 1\cdot28 \Rightarrow x_2 = 1{,}000 + 256 = 1{,}256$



Hence, after 1,256 hours of burning life, 10% of the blubs will be still burning.

Example 9·8. *The marks obtained by a number of students for a certain subject are assumed to be approximately normally distributed with mean value 65 and with a standard deviation of 5. If 3 students are taken at random from this set, what is the probability that exactly 2 of them will have marks over 70 ?*

Solution. Let the *r.v.* X denote the marks obtained by the given set of students in the given subject. Then we are given that $X \sim N(\mu, \sigma^2)$, where $\mu = 65$ and $\sigma = 5$. The probability 'p' that a randomly selected student from the given set gets marks over 70 is given by $p = P(X > 70)$. When $X = 70, Z = \dfrac{X-\mu}{\sigma} = \dfrac{70-65}{5} = 1.$

$$\therefore \quad p = P(X > 70) = P(Z > 1) = 0\cdot5 - P(0 \le Z \le 1)$$
$$= 0\cdot5 - 0\cdot3413 = 0\cdot1587 \qquad \text{[From Normal Probability Tables]}$$

Since this probability is same for each student of the set, the required probability that 'out of 3 students selected at random from the set, exactly 2 will have marks over 70, is given by the binomial probability law :

$$^3C_2\, p^2(1-p) = 3 \times (0\cdot1587)^2 \times (0\cdot8413) = 0\cdot06357.$$

Example 9·9. (*a*) *If* $\log_{10} X$ *is normally distributed with meant 4 and variance 4, find the probability of* $1\cdot202 < X < 83180000$.

(Given $\log_{10} 1202 = 3\cdot08, \log_{10} 8318 = 3\cdot92$*).*

(*b*) $\log_{10}X$ *is normally distributed with mean 7 and variance 3,* $\log_{10}Y$ *is normally distributed with mean 3 and variance unity. If the distributions of X and Y are independent, find the probability of* $1\cdot202 < (X/Y) < 83180000$.

[Given $\log_{10}(1202) = 3\cdot08, \log_{10}(8318) = 3\cdot92$.*]*

Solution. (*a*) Since $\log X$ is a non-decreasing function of X, we have

$$P(1\cdot202 < X < 83180000) = P(\log_{10} 1\cdot202 < \log_{10} X < 7\cdot92) = P(0\cdot08 < Y < 7\cdot92),$$

where $Y = \log_{10} X \sim N(4, 4)$ (Given).

When $Y = 0\cdot08$, $Z\ \dfrac{0\cdot08-4}{2} = -1\cdot96$ and when $Y = 7\cdot92$, $Z = \dfrac{7\cdot92-4}{2} = 1\cdot96$

$\therefore$ Required probability $= P(0\cdot08 < Y < 7\cdot92)$
$= P(-1\cdot96 < Z < 1\cdot96) = 2P(0 < Z < 1\cdot96)$ (By symmetry)
$= 2 \times 0\cdot4750 = 0\cdot9500$

(b) $P[1{\cdot}202 < (X/Y) < 83180000]$

$= P[\log_{10} 1{\cdot}202 < \log_{10}(X/Y) < \log_{10} 83180000] = (0{\cdot}08 < U < 7{\cdot}92)$,

where $U = \log_{10}(X/Y) = \log_{10} X - \log_{10} Y$.

Since $\log_{10} X \sim N(7, 3)$ and $\log_{10} Y \sim N(3, 1)$, are independent (Given)

$$\log_{10} X - \log_{10} Y \sim N(7-3, 3+1) \qquad \text{(c.f. Remark 1, § 9·2·8 page 9·10)}$$

$\therefore \quad U = (\log_{10} X - \log_{10} Y) \sim N(4, 4)$

$\Rightarrow$

$\therefore$ Required probability is given by :

$p = P(0{\cdot}08 < U < 7{\cdot}92)$, where $U \sim N(4, 4)$

$= 0{\cdot}95$ [See part(a)]

Example 9·10. *Two independent random variates X and Y are both normally distributed with means 1 and 2 and standard deviations 3 and 4 respectively. If Z = X – Y, write the probability density function of Z. Also state the median, s.d. and mean of the distribution of Z. Find Probability $(Z + 1 \le 0)$.*

Solution. Since $X \sim N(1, 9)$ and $Y \sim N(2, 16)$ are independent,

$$Z = X - Y \sim N(1-2, 9+16), \quad i.e., \quad Z = X - Y \sim N(-1, 25).$$

Hence p.d.f. of Z is given by :

$$p(z) = \frac{1}{5\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{z+1}{5}\right)^2\right], -\infty < z < \infty.$$

For the distribution of Z, Median = Mean = –1 and s.d. $= \sqrt{25} = 5$.

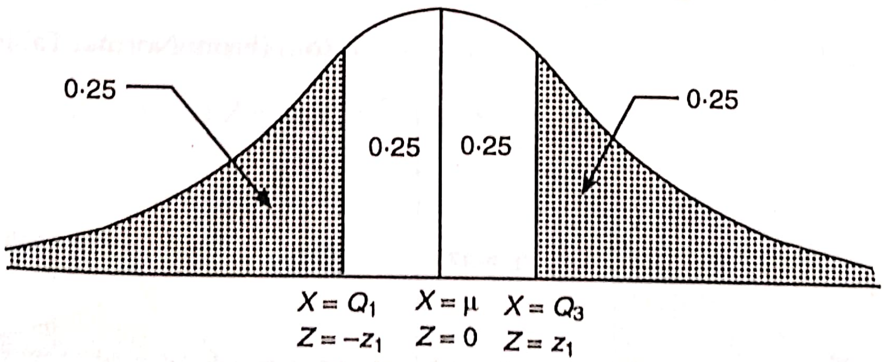
and $P(Z + 1 < 0) = P(Z \le -1) = P(U \le 0)$, $\left[U = \frac{Z+1}{5} \sim N(0, 1)\right]$

$= 0{\cdot}5$

Example 9·11. *Prove that for the normal distribution, the quartile deviation, the mean deviation and standard deviation are approximately 10 : 12 : 15.*

Solution. Let $X \sim N(\mu, \sigma^2)$. If Q_1 and Q_3 are the first and third quartiles respectively, then by definition : $P(X < Q_1) = 0{\cdot}25$ and $P(X > Q_3) = 0{\cdot}25$

The points Q_1 and Q_3 are located as shown in the figure given below.



When $X = Q_3$, $Z = \dfrac{Q_3 - \mu}{\sigma} = z_1$, (say), ... (*)

and when $X = Q_1$, $Z = \dfrac{Q_1 - \mu}{\sigma} = -z_1$ ... (**) (This is obvious form the figure.)

Subtracting, (**) from (*), we get : $\dfrac{Q_3 - Q_1}{\sigma} = 2z_1$

The quartile deviation is given by : $Q.\,D. = \dfrac{Q_3 - Q_1}{2} = \sigma\, z_1$ ...(***)

From the figure, obviously $P\,(0 < Z < z_1) = 0{\cdot}25 \Rightarrow z_1 = 0{\cdot}67$ (From Normal Tables)

$\therefore \quad Q.D. = \sigma\, z_1 = 0{\cdot}67\,\sigma = \frac{2}{3}\,\sigma$ [From (***)]

For normal distribution, mean deviation about mean (*c.f.* § 9·2·10, *page* 9·11) is given by :

$$M.D. = \sqrt{\frac{2}{\pi}}\;\sigma = \frac{4}{5}\,\sigma$$

Hence $\quad Q.\,D. : M.D. : S.D. \;::\; \frac{2}{3}\sigma : \frac{4}{5}\sigma : \sigma \;::\; \frac{2}{3} : \frac{4}{5} : 1 \;::\; 10 : 12 : 15.$

Remark. From (*) and (**), we get

$$Q_1 = \mu - \sigma z_1 = \mu - 0{\cdot}67\sigma \quad \text{and} \quad Q_3 = \mu + \sigma z_1 = \mu + 0{\cdot}67\sigma.$$

Example 9·12. *In a distribution exactly normal, 10·03% of the items are under 25 kilogram weight and 89·97% of the items are under 70 kilogram weight. What are the mean and standard deviation of the distribution ?*

Solution. Let X denote the weight (in kilograms) of the items. If $X \sim N\,(\mu, \sigma^2)$, then we are given : $\quad P\,(X < 25) = 0{\cdot}1003$ and $P\,(X < 70) = 0{\cdot}8997$

The points $X = 25$ and $X = 70$ are located as shown below :

Since the value $X = 25$ is located to the left of the ordinate at $X = \mu$, the corresponding value of Z is negative.

$$\text{When } X = 25,\quad Z = \frac{25 - \mu}{\sigma} = -z_1, \text{(say)} \qquad \ldots (1)$$

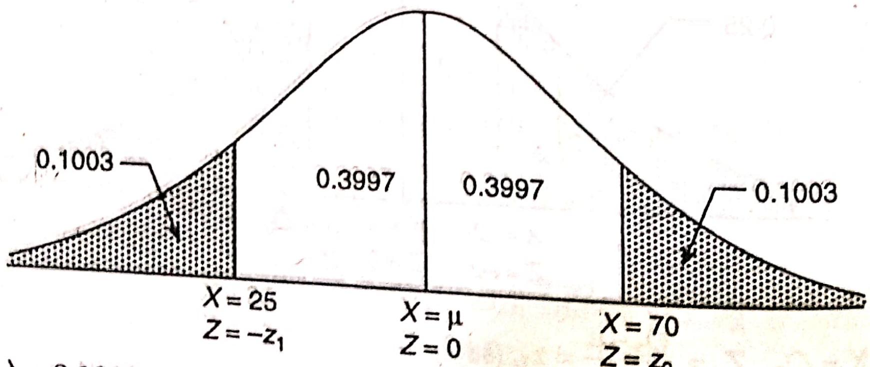
$$\text{when } X = 70,\quad Z = \frac{70 - \mu}{\sigma} = z_2 \text{ (say)}, \qquad \ldots (2)$$

From the diagram, it is obvious that

$$P\,(Z < -z_1) = 0{\cdot}1003$$

and $\quad P(Z < z_2) = 0{\cdot}8997$

Now $\quad P\,(0 < Z < z_2) = 0{\cdot}3997 \Rightarrow z_2 = 1{\cdot}28$ (From Normal Tables)



$P(Z < -z_1) = 0{\cdot}1003 \Rightarrow P(Z > z_1) = 0{\cdot}1003$ (By Symmetry)

$\therefore \quad P(0 < Z < z_1) = 0{\cdot}5 - 0{\cdot}1003 = 0{\cdot}3997 \Rightarrow z_1 = 1{\cdot}28$

Substituting the value of z_1 and z_2 in (1) and (2), we get

$$\frac{25-\mu}{\sigma} = -1{\cdot}28 \quad \Rightarrow \quad 25-\mu = -1{\cdot}28\,\sigma \qquad \ldots (3)$$

and

$$\frac{70-\mu}{\sigma} = 1{\cdot}28 \quad \Rightarrow \quad 70-\mu = 1{\cdot}28\,\sigma \qquad \ldots (4)$$

Subtracting (3) from (4), we get $45 = 2{\cdot}56\,\sigma \Rightarrow \sigma = 17{\cdot}578.$

Substituting the value of σ in (3), we have : $\mu = 25 + 17{\cdot}578 \times 1{\cdot}28 = 47{\cdot}5$

Hence the mean is 47·5 kilogram and standard deviation is 17·578 kilogram.

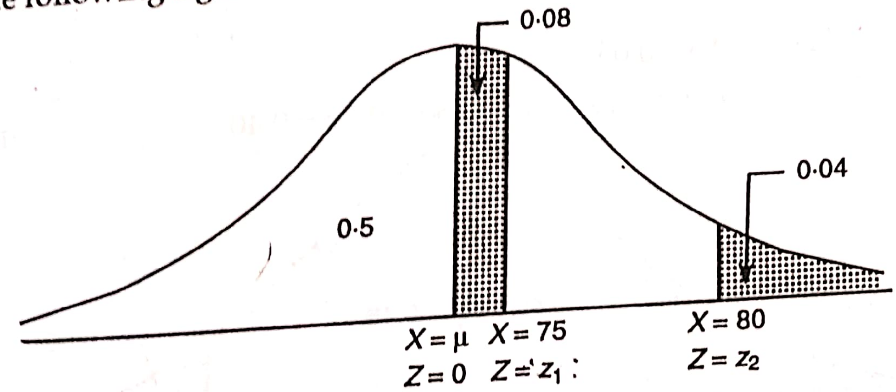
Example 9·13. *If the skulls are classified as A, B and C according as the length-breadth index is under 75, between 75 and 80, or over 80, find approximately (assuming that the distribution is normal) the mean and standard deviation of a series in which A are 58%, B are 38% and C are 4%, being given that if*

$$f(t) = \frac{1}{\sqrt{2\pi}}\int_0^t \exp(-x^2/2)\,dx, \text{ then } \quad f(0{\cdot}20) = 0{\cdot}08 \quad \text{and} \quad f(1{\cdot}75) = 0{\cdot}46.$$

Solution. Let the length-breadth index be denoted by the variable X, then

$$P(X < 75) = 0{\cdot}58 \quad \text{and} \quad P(X > 80) = 0{\cdot}04 \qquad \text{(Given)} \qquad \ldots (1)$$

Since $P(X < 75)$ represents the total area to the left of the ordinate at the point $X = 75$ and $P(X > 80)$ represents the total area to the right of the ordinate at the point $X = 80$, it is obvious from (1) that the points $X = 75$ and $X = 80$ are located at the positions shown in the following figure.



$\frac{1}{\sqrt{2\pi}}\int_0^t \exp(-x^2/2)\,dx$ represents the area under standard normal curve between the ordinates at $Z = 0$ and $Z = t$, Z being $N(0, 1)$ variate.

$$\therefore \quad f(t) = \frac{1}{\sqrt{2\pi}}\int_0^t \exp(-x^2/2)\,dx = P(0 < Z < t)$$

Hence $f(0{\cdot}20) = P(0 < Z < 0{\cdot}20) = 0{\cdot}08$

and $f(1{\cdot}75) = P(0 < Z < 1{\cdot}75) = 0{\cdot}46$ $\qquad \ldots (2)$

Let μ and σ be the mean and s.d. of the distribution respectively. Then $X \sim N(\mu, \sigma^2)$.

When $X = 75$, $Z = \frac{75-\mu}{\sigma} = z_1$ (say), and when $X = 80$, $Z = \frac{80-\mu}{\sigma} = z_2$ (say).

Then from the figure, it is obvious that $P(X < 75) = 0{\cdot}58 \Rightarrow P(0 < Z < z_1) = 0{\cdot}08$

$\therefore$ Using (2), we have $$z_1 = \frac{75 - \mu}{\sigma} = 0{\cdot}20 \qquad \ldots (3)$$

Also $P(X > 80) = 0{\cdot}04 \Rightarrow P(0 < Z < z_2) = 0{\cdot}46$

$\therefore$ From (2), we have $$z_2 = \frac{80 - \mu}{\sigma} = 1{\cdot}75 \qquad \ldots (4)$$

Solving the equations (3) and (4), we have

$$\mu = 74{\cdot}35 \text{ (approx.) and } \sigma = 3{\cdot}23 \text{ (approx.).}$$

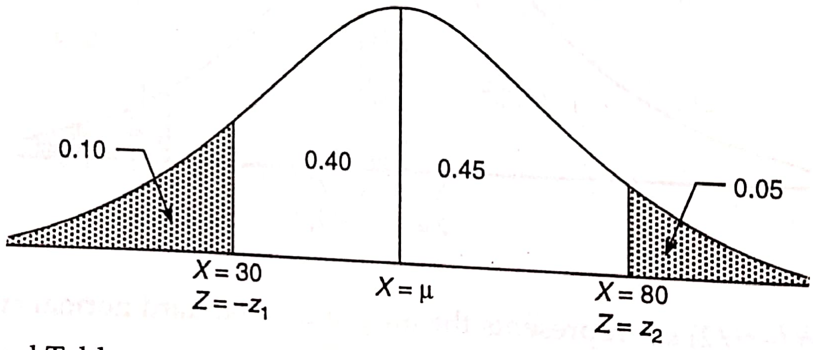
Example 9·14. *In an examination it is laid down that a student passes if he secures 30 per cent or more marks. He is placed in the first, second or third division according as he secures 60% or more marks, between 45% and 60% marks and marks between 30% and 45% respectively. He gets distinction in case he secures 80% or more marks. It is noticed from the result that 10% of the students failed in the examination, whereas 5% of them obtained distinction. Calculate the percentage of students placed in the second division. (Assume normal distribution of marks.)*

Solution. Let the variable X denote the marks (out of 100) in the examination and let $X \sim N(\mu, \sigma^2)$. then we are given $P(X < 30) = 0{\cdot}10$ and $P(X \geq 80) = 0{\cdot}05$.

$$\left.\begin{aligned} &\text{When } X = 30,\ Z = \frac{30 - \mu}{\sigma} = -z_1 \text{ (say)},\\ &\text{When } X = 80,\ Z = \frac{80 - \mu}{\sigma} = z_2 \text{ (say)}. \end{aligned}\right\} \qquad \ldots (*)$$

$\therefore \quad P(0 < Z < z_2) = 0{\cdot}5 - 0{\cdot}05 = 0{\cdot}45$

and $\quad P(0 < Z < z_1) = P(-z_1 < Z < 0) = 0{\cdot}50 - 0{\cdot}10 = 0{\cdot}40$ (By symmetry)



$\therefore$ From Normal Tables, $z_1 = 1{\cdot}28$ and $z_2 = 1{\cdot}64$

Hence $$\frac{30 - \mu}{\sigma} = -1{\cdot}28 \Rightarrow \frac{\mu - 30}{\sigma} = 1{\cdot}28 \quad \text{and} \quad \frac{80 - \mu}{\sigma} = 1{\cdot}64 \qquad \ldots \text{[From(*)]}$$

Adding, we get $$\frac{50}{\sigma} = 2{\cdot}92 \Rightarrow \sigma = \frac{50}{2{\cdot}92} = 17{\cdot}12$$

$$\therefore \quad \mu = 30 + 1{\cdot}28 \times 17{\cdot}12 = 30 + 21{\cdot}9136 = 51{\cdot}9136 \simeq 52.$$

The probability 'p' that a candidate is placed in the second division is equal to the probability that his score lies between 45 and 60, *i.e.*,

$$p = P(45 < X < 60) = P(-0{\cdot}41 < Z < 0{\cdot}47), \qquad \left(Z = \frac{X - 52}{17{\cdot}12}\right)$$

Scanned with OKEN Scanner

$= P(-0.41 < Z < 0) + P(0 < Z < 0.47)$

$= P(0 < Z < 0.41) + P(0 < Z < 0.47)$ (By symmetry)

$= 0.1591 + 0.1808 = 0.3399 = 0.34$ (approx.)

Hence, 34% candidates got second division in the examination.

Example 9.15. *A sample of 100 items is taken at random from a batch known to contain 40% defectives. What is the probability that the sample contains : (i) at least 44 defectives, (ii) exactly 44 defectives ?*

Solution. Since $n = 100$, is large, we may use the normal distribution as an approximation to the Binomial distribution.

Let X denote number of defectives with parameters,

$$\mu = np = 100\,(0.4) = 40 \quad \text{and} \quad \sigma = \sqrt{npq} = \sqrt{100\,(0.4)\,(0.6)} = 4.9$$

(i) It should be noted that the continuous normal distribution is approximating the discrete Binomial distribution so that the continuity correction has to be taken into account in determining the various probabilities. So finding the probability of at least 44 defectives in a sample of 100 items requires finding the area under the normal curve from 43.5 to 100.5.

$\therefore$ The probability of at least 44 defectives is :

$$P(43.5 < X < 100.5) = P\left(\frac{43.5-40}{4.9} < Z < \frac{100.5-40}{4.9}\right) = P(0.7143 < Z < 12.347)$$

$$= P(0 < Z < 12.347) - P(0 < Z < 0.7143) = 0.5 - 0.2624 = 0.2376.$$

(ii) The probability of exactly 44 defectives is :

$$P(X = 44) = P(43.5 < X < 44.5) = P\left(\frac{43.5-40}{4.9} < Z < \frac{44.5-40}{4.9}\right)$$

$$= P(0.7143 < Z < 0.9184) = P(0 < Z < 0.9184) - P(0 < Z < 0.7143)$$

$$= 0.3208 - 0.2624 = 0.0584.$$

Note. Using the Binomial distribution :

$$P(X \geq 44) = \sum_{r=44}^{100} {}^{100}C_r\,(0.4)^r\,(0.6)^{100-r} = 0.2365 \text{ (using tables)}$$

$$P(X = 44) = {}^{100}C_{44}\,(0.4)^{44}\,(0.6)^{56} = 0.0576.$$

As can be seen by comparing the answers, both sets of answers are remarkably close.

Example 9.16. *Let $X \sim N(\mu, \sigma^2)$. If $\sigma^2 = \mu^2$, $(\mu > 0)$, express $P(X < -\mu \mid X < \mu)$ in terms of cumulative distribution function of $N(0, 1)$.*

Solution. $P(X < -\mu \mid X < \mu)$ $(\because \mu > 0)$

$$= \frac{P(X < -\mu \cap X < \mu)}{P(X < \mu)} = \frac{P(X < -\mu)}{P(X < \mu)} \qquad \left(Z = \frac{X-\mu}{\sigma} = \frac{X-\mu}{\mu}\right)$$

$$= \frac{P(Z < -2)}{P(Z < 0)}, \qquad \text{(By symmetry)}$$

$$= \frac{P(Z > 2)}{(1/2)},$$

$$= 2\left[1 - P(Z \leq 2)\right] = 2\left[1 - \Phi(2)\right],$$

where $\Phi(.)$ is the distribution function of standard normal variate.

Example 9·17. *If X, Y are independent normal variates with means 6, 7 and variances 9, 16 respectively, determine λ such that* $P(2X + Y \le \lambda) = P(4X - 3Y \ge 4\lambda)$.

Solution. Since X and Y are independent normal variates, using § 9·2·8

$$U = 2X + Y \sim N(2\times 6 + 7, 4\times 9 + 16), \quad i.e., \quad U \sim N(19, 52)$$

$$V = 4X - 3Y \sim N(4\times 6 - 3\times 7, 16\times 9 + 9\times 16), \quad i.e., \quad V \sim N(3, 288)$$

and $$P(2X + Y \le \lambda) = P(U \le \lambda) = P\left(Z \le \frac{\lambda - 19}{\sqrt{52}}\right), \text{ where } Z \sim N(0, 1)$$

and $$P(4X - 3Y \ge 4\lambda) = P(V \ge 4\lambda) = P\left(Z \ge \frac{4\lambda - 3}{12\sqrt{2}}\right), \text{ where } Z \sim N(0, 1)$$

Now $$P(2X + Y \le \lambda) = P\{(4X - 3Y) \ge 4\lambda\} \quad \text{(Given)}$$

$$\Rightarrow \quad P\left(Z \le \frac{\lambda - 19}{\sqrt{52}}\right) = P\left(Z \ge \frac{4\lambda - 3}{12\sqrt{2}}\right) \quad \text{or} \quad \frac{\lambda - 19}{\sqrt{52}} = -\left(\frac{4\lambda - 3}{12\sqrt{2}}\right)$$

[Since, $P(Z \le a) = P(Z \ge b) \Rightarrow a = -b$, because normal probability curve is symmetric about $Z = 0$.]

$$\Rightarrow \quad \frac{\lambda - 19}{\sqrt{13}} = \frac{3 - 4\lambda}{6\sqrt{2}} \Rightarrow (6\sqrt{2} + 4\sqrt{13})\,\lambda = 114\sqrt{2} + 3\sqrt{13} \Rightarrow \lambda = \frac{114\sqrt{2} + 3\sqrt{13}}{6\sqrt{2} + 4\sqrt{13}}.$$

Example 9·18. *If X and Y are independent normal variates possessing a common mean μ such that*

$$P(2X + 4Y \le 10) + P(3X + Y \le 9) = 1 \quad \text{and} \quad P(2X - 4Y \le 6) + P(Y - 3X \ge 1) = 1,$$

determine the values of μ and the ratio of the variances of X and Y.

Solution. Let $\text{Var}(X_1) = \sigma_1^2$ and $\text{Var}(Y) = \sigma_2^2$

Since $E(X) = E(Y) = \mu$., (Given) and X and Y are independent normal variates, by § 9·2·8, we have

$$2X + 4Y \sim N(2\mu + 4\mu, 4\sigma_1^2 + 16\sigma_2^2), \quad i.e., \quad N(6\mu, 4\sigma_1^2 + 16\sigma_2^2)$$

$$3X + Y \sim N(3\mu + \mu, 9\sigma_1^2 + \sigma_2^2), \quad i.e., \quad N(4\mu, 9\sigma_1^2 + \sigma_2^2)$$

$$2X - 4Y \sim N(2\mu - 4\mu, 4\sigma_1^2 + 16\sigma_2^2), \quad i.e., \quad N(-2\mu, 4\sigma_1^2 + 16\sigma_2^2)$$

$$Y - 3X \sim N(\mu - 3\mu, \sigma_2^2 + 9\sigma_1^2), \quad i.e., \quad N(-2\mu, 9\sigma_1^2 + \sigma_2^2)$$

Let us further write : $4\sigma_1^2 + 16\sigma_2^2 = \alpha^2$ and $9\sigma_1^2 + \sigma_2^2 = \beta^2$ ... (1)

If Z denotes the standard normal variate, *i.e.*, if $Z \sim (0, 1)$, we get

$$P(2X + 4Y \le 10) + P(3X + Y \le 9) = 1 \Rightarrow P\left(Z \le \frac{10 - 6\mu}{\alpha}\right) + P\left(Z \le \frac{9 - 4\mu}{\beta}\right) = 1$$

$$\Rightarrow \quad P\left(Z \le \frac{10 - 6\mu}{\alpha}\right) = 1 - P\left(Z \le \frac{9 - 4\mu}{\beta}\right) = P\left(Z \ge \frac{9 - 4\mu}{\beta}\right)$$

$$\therefore \quad \frac{10 - 6\mu}{\alpha} = -\left(\frac{9 - 4\mu}{\beta}\right)$$ (Since normal distribution is symmetric about $Z = 0$.).... (2)

Similarly

$$P(2X - 4Y \le 6) + P(Y - 3X \ge 1) = 1 \Rightarrow P\left(Z \le \frac{6 + 2\mu}{\alpha}\right) + P\left(Z \ge \frac{1 + 2\mu}{\beta}\right) = 1$$

$$\Rightarrow \quad P\left(Z \le \frac{6 + 2\mu}{\alpha}\right) = 1 - P\left(Z \ge \frac{1 + 2\mu}{\beta}\right) = P\left(Z \le \frac{1 + 2\mu}{\beta}\right)$$

$$\Rightarrow \quad \frac{6+2\mu}{\alpha} = \frac{1+2\mu}{\beta} \qquad \ldots (3)$$

Solving (2) and (3), we get $\quad \dfrac{\alpha}{\beta} = \dfrac{6+2\mu}{1+2\mu} = \dfrac{10-6\mu}{4\mu-9} \qquad \ldots (4)$

$$\Rightarrow \quad (6+2\mu)(4\mu-9) = (10-6\mu)(1+2\mu) \quad \Rightarrow \quad 5\mu^2 - 2\mu - 16 = 0$$

$$\therefore \quad \mu = \frac{2 \pm \sqrt{4+320}}{10} = \frac{2 \pm 18}{10} = 2 \text{ or } -1{\cdot}6.$$

Substituting $\mu = 2$ in (4), $\quad \dfrac{\alpha}{\beta} = \dfrac{10}{5} = 2 \quad \Rightarrow \quad 4 = \dfrac{\alpha^2}{\beta^2}$

From (1), we get $\quad 4 = \dfrac{4\sigma_1^2 + 16\sigma_2^2}{9\sigma_1^2 + \sigma_2^2} = \dfrac{4+16\lambda}{9+\lambda} \qquad \left[\text{Taking } \lambda = \dfrac{\sigma_2^2}{\sigma_1^2}\right]$

$$\Rightarrow \quad 4(9+\lambda) = 4 + 16\lambda \quad \Rightarrow \quad \lambda = \frac{32}{12} = \frac{8}{3}$$

Putting $\mu = -1{\cdot}6$ in (4), we obtain $\left(\dfrac{14}{11}\right)^2 = \dfrac{\alpha^2}{\beta^2} = \dfrac{4+16\lambda}{9+\lambda} \Rightarrow \lambda = \dfrac{1280}{1740} = \dfrac{64}{87}$.

Example 9·19. *If two normal universes A and B have the same total frequency but the standard deviation of universe A is k times that of the universe B, show that maximum frequency of universe A is 1/k times that of universe B.*

Solution. Let N be the same total frequency for each of the two universes A and B. If σ is the standard deviation of universe B, then the standard deviation of universe A is $k\sigma$. Let μ_1 and μ_2 be the means of the universes A and B respectively.

The frequency function of universe A is : $\quad f_A(x) = \dfrac{N}{k\sigma\sqrt{2\pi}} \exp\{-(x-\mu_1)^2/2k^2\sigma^2\}$

and the frequency function of universe B is : $\quad f_B(x) = \dfrac{N}{\sigma\sqrt{2\pi}} \exp\{-(x-\mu_2)^2/2\sigma^2\}$

Since, for a normal distribution, the maximum frequency occurs at the point x = mean, we have

$[f_A(x)]_{\max}$ = Maximum frequency of universe A

$$= [f_A(x)]_{x=\mu_1} = \left[\frac{N}{k\sigma\sqrt{2\pi}} \exp\{-(x-\mu_1)^2/2k^2\sigma^2\}\right]_{x=\mu_1} = \frac{N}{k\sigma\sqrt{2\pi}}$$

Similarly

$$[f_B(x)]_{\max} = [f_B(x)]_{x=\mu_2} = \left[\frac{N}{\sigma\sqrt{2\pi}} \exp\{-(x-\mu_2)^2/2\sigma^2\}\right]_{x=\mu_2} = \frac{N}{\sigma\sqrt{2\pi}}$$

$$\therefore \quad \frac{[f_A(x)]_{\max}}{[f_B(x)]_{\max}} = \frac{1}{k} \quad \Rightarrow \quad [f_A(x)]_{\max} = \frac{1}{k} \cdot [f_B(x)]_{\max}$$

Example 9·20. *Let X be a random variable following normal distribution with mean μ and variance σ^2 and let r be a non-negative integer.*

If $\mu'_r = E(X^r)$ and if $\mu_{2r} = [E(X-\mu)^{2r}]$, prove that

(i) $\mu'_{r+2} = 2\mu\mu'_{r+1} + (\sigma^2 - \mu^2)\mu'_r + \sigma^3 \dfrac{d\mu'_r}{d\sigma}$ $\qquad$ (ii) $\mu_{2r+2} = \sigma^2 \mu_{2r} + \sigma^3 \dfrac{d\mu_{2r}}{d\sigma}$.

Solution. (*i*) $\mu_r' = \int_{-\infty}^{\infty} x^r f(x)\, dx = \int_{-\infty}^{\infty} x^r \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] dx$

$$\therefore \quad \frac{d\mu_r'}{d\sigma} = \int_{-\infty}^{\infty} \frac{x^r}{\sqrt{2\pi}}\left(-\frac{1}{\sigma^2}\right) \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] dx$$

$$+ \int_{-\infty}^{\infty} \frac{x^r}{\sqrt{2\pi}\,\sigma}\left\{-\frac{(x-\mu)^2}{2} \times \left(\frac{-2}{\sigma^3}\right)\right\} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] dx$$

$$= -\frac{1}{\sigma}\int_{-\infty}^{\infty} x^r . f(x)\, dx + \frac{1}{\sigma^3}\int_{-\infty}^{\infty} x^r (x-\mu)^2 . f(x)\, dx$$

$$= -\frac{1}{\sigma}\cdot \mu_r' + \frac{1}{\sigma^3}\int_{-\infty}^{\infty} (x^{r+2} - 2\mu\, x^{r+1} + \mu^2 x^r) f(x)\, dx$$

$$= -\frac{1}{\sigma}\cdot \mu_r' + \frac{1}{\sigma^3}\left(\mu'_{r+2} - 2\mu\, \mu'_{r+1} + \mu^2 \mu_r'\right)$$

$$\Rightarrow \quad \sigma^3 \frac{d\mu'_r}{d\sigma} = -\sigma^2 \mu'_r + \mu'_{r+2} - 2\mu .\, \mu'_{r+1} + \mu^2 \mu'_r$$

$$\Rightarrow \quad \mu'_{r+2} = 2\mu .\, \mu'_{r+1} + (\sigma^2 - \mu^2)\, \mu'_r + \sigma^3 . \frac{d\mu_r'}{d\sigma} .$$

$$(ii) \quad \mu_{2r} = \int_{-\infty}^{\infty} (x-\mu)^{2r} . f(x) dx = \int_{-\infty}^{\infty} (x-\mu)^{2r} \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] dx$$

$$\therefore \quad \frac{d\mu_{2r}}{d\sigma} = \int_{-\infty}^{\infty} \frac{(x-\mu)^{2r}}{\sqrt{2\pi}}\left(-\frac{1}{\sigma^2}\right) . \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] dx$$

$$+ \int_{-\infty}^{\infty} \frac{(x-\mu)^{2r}}{\sqrt{2\pi} .\, \sigma} \cdot \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] . \left\{-\frac{(x-\mu)^2}{2}\left(\frac{-2}{\sigma^3}\right)\right\} dx$$

$$= -\frac{1}{\sigma}\int_{-\infty}^{\infty} (x-\mu)^{2r} f(x)\, dx + \frac{1}{\sigma^3}\int_{-\infty}^{\infty} (x-\mu)^{2r+2} . f(x) dx$$

$$= -\frac{1}{\sigma}\, \mu_{2r} + \frac{1}{\sigma^3}\, \mu_{2r+2}$$

$$\Rightarrow \quad \mu_{2r+2} = \sigma^2 \mu_{2r} + \sigma^3 \frac{d\mu_{2r}}{d\sigma} .$$

9·2·15. Log-normal Distribution. The positive *r.v.* X is said to have a log-normal distribution if $\log_e X$ is normally distributed.

Let $Y = \log_e X \sim N(\mu, \sigma^2)$. For $x > 0$,

$F_X(x) = P(X \le x) = P(\log_e X \le \log_e x) = P(Y \le \log_e x)$

(Since $\log X$ is monotonic increasing function.)

$$= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\log x} \exp\{-(y-\mu)^2/2\sigma^2\}\, dy \qquad [\text{Since } Y \sim N(\mu, \sigma^2)]$$

$$= \frac{1}{\sigma\sqrt{2\pi}} \int_0^x \exp\{-(\log u - \mu)^2/2\sigma^2\} \frac{du}{u}, \qquad (y = \log u)$$

For $x \le 0$, $F_X(x) = P(X \le x) = 0$, because X is a positive *r.v.*

Let us define

$$f_X(u) = \begin{cases} \dfrac{1}{u\sigma\sqrt{2\pi}} \cdot \exp\{-(\log u - \mu)^2/2\sigma^2\}, & u > 0 \\ 0, & u \le 0 \end{cases} \quad \ldots (9\cdot17)$$

Then $F_X(x) = \int_{-\infty}^{x} f_X(u)\, du$, for every x and hence $f(x)$ defined in (9·17) is a *p.d.f.* of X.

Remark. If $X \sim N(\mu, \sigma^2)$, then $Y = e^X$, is called a log-normal random variable, since its logarithm $\log Y = X$, is a normal *r.v.*

Moments. The rth moment about origin is given by :

$$\begin{aligned} \mu_r' &= E(X^r) = E(e^{rY}) && [\because Y = \log X \Rightarrow X = e^Y] \\ &= M_Y(r) && (m.g.f.\ of\ Y, r\ being\ the\ parameter) \\ &= \exp(\mu r + \tfrac{1}{2} r^2 \sigma^2) && [\because Y \sim N(\mu, \sigma^2)] \end{aligned} \quad \ldots (9\cdot18)$$

Remarks 1. In particular if we take $\mu = \log \alpha, \alpha > 0$, *i.e.*, $\log X \sim N(\log \alpha, \sigma^2)$, then

$$\mu_r' = E(X^r) = \exp\{r. \log \alpha + \tfrac{1}{2} r^2 \sigma^2\} = \alpha^r. \exp\{r^2 \sigma^2/2\} \quad \ldots (9\cdot18a)$$

$\therefore$ Mean $= \mu_1' = \alpha e^{\sigma^2/2}$ and $\mu_2 = \mu_2' - \mu_1'^2 = \alpha^2 e^{\sigma^2} (e^{\sigma^2} - 1)$

2. Log normal distribution arises in problems of economics, biology, geology, and reliability theory. In particular, it arises in the study of dimensions of particles under pulverisation.

3. If $X_1, X_2, \ldots, X_n$ is a set of independently identically distributed random variables such that mean of each $\log X_i$ is μ and its variance is σ^2, then the product $X_1 . X_2 \ldots . X_n$ is asymptotically distributed according to logarithmic normal distribution and with mean μ and variance $n\sigma^2$.

9·3. RECTANGULAR (OR UNIFORM) DISTRIBUTION

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

$$f(x; a, b) = \begin{cases} \dfrac{1}{b-a}, & \text{if } a < x < b \\ 0, & \text{otherwise} \end{cases} \quad \ldots (9\cdot19)$$

Remarks 1. a and b, $(a < b)$ are the two parameters of the distribution. The distributin is called uniform distribution on (a, b) since it assumes a constant (uniform) value for all x in (a, b).

2. The distribution is also known as rectangular distribution, since the curve $y = f(x)$ describes a rectangle over the x-axis and between the ordinates at $x = a$ and $x = b$.

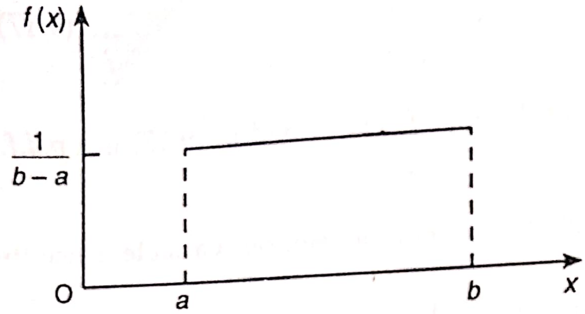
3. A uniform or rectangular variate X on the interval (a, b) is written as : $X \sim U[a, b]$ or $X \sim R[a, b]$.

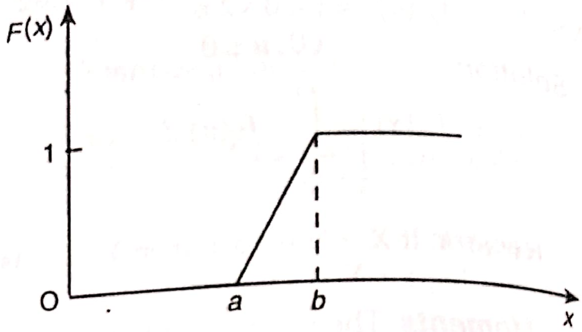
4. The cumulative distribution function $F(x)$ is given by :

$$F(x) = \begin{cases} 0, & x \le a \\ \dfrac{x-a}{b-a}, & a < x < b \\ 1, & x \ge b \end{cases} \quad \ldots (9\cdot19a)$$

Since $F(x)$ is not continuous at $x = a$ and $x = b$, it is not differentiable at these points. Thus $\frac{d}{dx} F(x) = f(x) = \frac{1}{b-a} \ne 0$, exists everywhere except at the points $x = a$ and $x = b$ and consequently *p.d.f.* $f(x)$ is given by (9·19).

5. The graphs of uniform *p.d.f.* $f(x)$ and the corresponding distribution function $F(x)$ are given below.





6. For a rectangular or uniform variate X in $(-a, a)$, the *p.d.f.* is given by :

$$f(x) = \begin{cases} \dfrac{1}{2a}, & -a < x < a \\ 0, & \text{otherwise} \end{cases} \qquad \ldots (9{\cdot}19b)$$

9·3·1. Moments of Rectangular Distribution. Let $X \sim U[a, b]$.

$$\mu_r' = \int_a^b x^r f(x)\,dx = \frac{1}{b-a}\int_a^b x^r\,dx = \frac{1}{b-a}\left(\frac{b^{r+1}-a^{r+1}}{r+1}\right) \qquad \ldots (9{\cdot}20)$$

In particular

$$\text{Mean } = \mu_1' = \frac{1}{b-a}\left(\frac{b^2-a^2}{2}\right) = \frac{b+a}{2} \qquad \ldots (9{\cdot}20a)$$

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

$$\therefore \quad \text{Variance} = \mu_2' - \mu_1'^2 = \tfrac{1}{3}(b^2+ab+a^2) - \left\{\tfrac{1}{2}(b+a)\right\}^2 = \tfrac{1}{12}(b-a)^2 \qquad \ldots (9{\cdot}20b)$$

9·3·2. M.G.F. of Rectangular Distribution is given by :

$$M_X(t) = \int_a^b e^{tx} f(x)\,dx = \int_a^b \frac{e^{tx}}{b-a}\,dx = \frac{e^{bt}-e^{at}}{t(b-a)},\ t \neq 0 \qquad \ldots (9{\cdot}20c)$$

9·3·3. Characteristic Function of Rectangular Distribution is given by :

$$\phi_X(t) = \int_a^b e^{itx}\,dx = \frac{e^{ibt}-e^{iat}}{it(b-a)},\ t \neq 0. \qquad \ldots (9{\cdot}20d)$$

9·3·4. Mean Deviation about Mean, η **of Rectangular Distribution** is given by :

$$\eta = E\left|X - \text{Mean}\right| = \int_a^b \left|x - \text{Mean}\right| f(x)\,dx$$

$$= \frac{1}{b-a}\int_a^b \left|x - \frac{a+b}{2}\right| dx = \frac{1}{b-a}\int_{-(b-a)/2}^{(b-a)/2} |t|\,dt, \text{ where } t = x - \frac{a+b}{2}$$

$$= \frac{1}{b-a}\cdot 2\int_0^{(b-a)/2} t\,dt = \frac{b-a}{4} \qquad \ldots (9{\cdot}20e)$$

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

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

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

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

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

Example 9·22. *Subway trains on a certain line run every half hour between mid-night and six in the morning. What is the probability that a man entering the station at a random time during this period will have to wait at least twenty minutes ?*

Solution. Let the *r.v.* X denote the waiting time (in minutes) for the next train. Under the assumption that a man arrives at the station at random, X is distributed uniformly on (0, 30), with *p.d.f.*,

$$f(x) = \begin{cases} \frac{1}{30}, & 0 < x < 30 \\ 0, & \text{otherwise} \end{cases}$$

The probability that he has to wait at least 20 minutes is given by :

$$P(X \geq 20) = \int_{20}^{30} f(x)\,dx = \frac{1}{30}\int_{20}^{30} 1.\,dx = \frac{1}{30}(30 - 20) = \frac{1}{3}.$$

Example 9·23. *If X has a uniform distribution in [0, 1], find the distribution (p.d.f.) of $-2 \log X$. Identify the distribution also.*

Solution. Let $Y = -2 \log X$. Then the distribution function G of Y is given by :

$$G_Y(y) = P(Y \leq y) = P(-2 \log X \leq y) = P(\log X \geq -y/2) = P(X \geq e^{-y/2})$$

$$= 1 - P(X \leq e^{-y/2}) = 1 - \int_0^{e^{-y/2}} f(x)\,dx = 1 - \int_0^{e^{-y/2}} 1.\,dx = 1 - e^{-y/2} \quad \ldots (*)$$

$$\therefore \quad g_Y(y) = \frac{d}{dy} G(y) = \frac{1}{2} e^{-y/2}, 0 < y < \infty$$

[$\because$ as X ranges in (0, 1), $Y = -2 \log X$ ranges from 0 to ∞]

Remark. This example illustrates that if $X \sim U[0, 1]$, then $Y = -2 \log X$, has an exponential distribution with parameter $\theta = \frac{1}{2}$. [*c.f.* § 9·8] or $Y = -2 \log X$, has chi-square distribution with $n = 2$ degrees of freedom [*c.f.* Chapter 15].

Example 9·24. *Show that for rectangular distribution : $f(x) = \frac{1}{2a}, -a < x < a$, m.g.f. about origin is $\frac{1}{at}(\sinh at)$. Also show that moments of even order are given by :*

$$\mu_{2n} = \frac{a^{2n}}{(2n+1)}.$$

Solution. M.G.F. about origin is given by :

$$M_X(t) = E\,(e^{tX}) = \int_{-a}^{a} e^{tx} f(x)\,dx = \frac{1}{2a}\int_{-a}^{a} e^{tx}\,dx = \frac{1}{2a}\left|\frac{e^{tx}}{t}\right|_{-a}^{a} = \frac{1}{2at}(e^{at} - e^{-at}) = \frac{\sinh at}{at}$$

$$= \frac{1}{at}\left\{at + \frac{(at)^3}{3\,!} + \frac{(at)^5}{5\,!} + \ldots + \frac{(at)^{2n+1}}{(2n+1)\,!} + \ldots\right\} = 1 + \frac{a^2t^2}{3\,!} + \frac{a^4t^4}{5\,!} + \ldots + \frac{a^{2n}t^{2n}}{(2n+1)\,!} + \ldots$$

Since there are no terms with odd powers of t in $M(t)$, all moments of odd order about origin vanish, *i.e.*, μ'_{2n+1} (about origin) $= 0$.

In particular, μ_1' (about origin) $= 0$, *i.e.*, mean $= 0$

Thus μ_r' (about origin) $= \mu_r$ (Since mean is origin.)

Hence $\mu_{2n+1} = 0;\ n = 0, 1, 2, \ldots$, *i.e.*, all moments of odd order about mean vanish.

The moments of even order are given by :

$$\mu_{2n} = \text{coefficient of } \frac{t^{2n}}{(2n)\,!} \text{ in } M(t) = \frac{a^{2n}}{(2n+1)}.$$

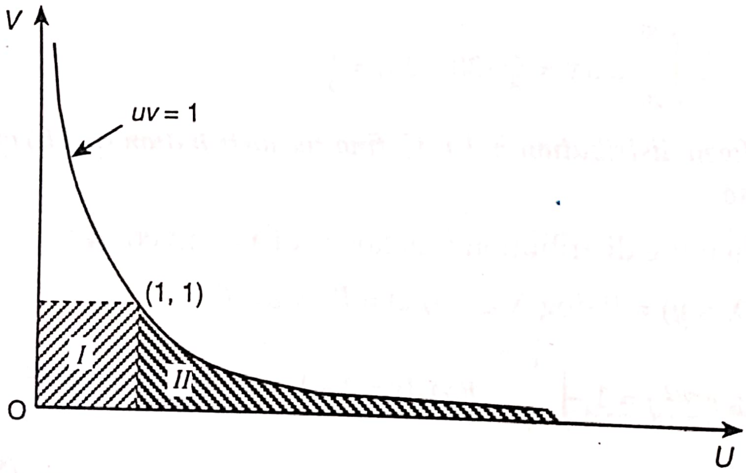
Example 9·25. *If X_1 and X_2 are independent rectangular variates on* $[0, 1]$, *find the distributions of* : (*i*) X_1/X_2, (*ii*) $X_1 X_2$, (*iii*) $X_1 + X_2$, *and* (*iv*) $X_1 - X_2$.

Solution. We are given : $f_{X_1}(x_1) = f_{X_2}(x_2) = 1;\ 0 < x_1 < 1,\ 0 < x_2 < 1.$

Since X_1 and X_2 are independent, their joint *p.d.f.* is :

$$f(x_1, x_2) = f(x_1)\,f(x_2) = 1$$

(*i*) Let us transform to : $u = \dfrac{x_1}{x_2}$, $v = x_2$, *i.e.*, $x_1 = uv,\ x_2 = v$



$$J = \frac{\partial(x_1, x_2)}{\partial(u, v)} = \begin{vmatrix} v & 0 \\ u & 1 \end{vmatrix} = v$$

$x_1 = 0$ maps to $u = 0, v = 0$

$x_1 = 1$ maps to $uv = 1$

(Rectangular hyperbola)

$x_2 = 0$ maps to $v = 0$, and

$x_2 = 1$ maps to $v = 1$.

The joint *p.d.f.* of U and V becomes :

$$g(uv) = f(x_1, x_2)\,|\,J\,| = v;\ 0 < u < \infty,\ 0 < v < \infty$$

To obtain the marginal distribution of U, we have to integrate out v.

In region (*I*) : $$g_1(u) = \int_0^1 v\,dv = \left|\frac{v^2}{2}\right|_0^1 = \frac{1}{2},\ 0 \le u \le 1$$

In region (*II*) : $$g_1(u) = \int_0^{1/u} v\,dv = \left|\frac{v^2}{2}\right|_0^{1/u} = \frac{1}{2u^2},\ 1 < u < \infty$$

Hence the distribution $U = \dfrac{X_1}{X_2}$ is given by : $g(u) = \begin{cases} \dfrac{1}{2}, & 0 \le u \le 1 \\ \dfrac{1}{2u^2}, & 1 < u < \infty \end{cases}$

(ii) Let $u = x_1 x_2, v = x_1 \Rightarrow x_1 = v, x_2 = \dfrac{u}{v}$ and $J = \begin{vmatrix} 0 & 1 \\ \frac{1}{v} & -\frac{u}{v^2} \end{vmatrix} = -\dfrac{1}{v}$

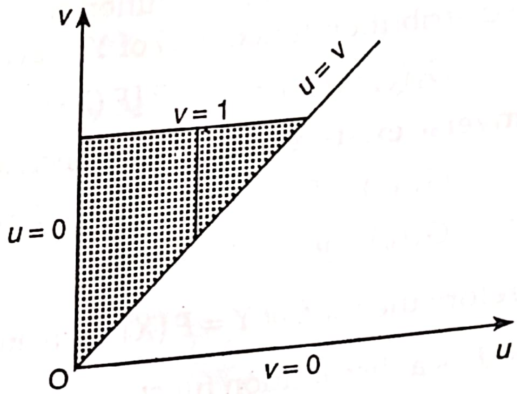
$x_1 = 0$ maps to $v = 0$, $x_1 = 1$ maps to $v = 1$

$x_2 = 0$ maps to $u = 0$, and $x_2 = 1$ maps to $u = v$

Moreover, $v = \dfrac{u}{x_2} \Rightarrow v \geq u$ (Since $0 < x_2 < 1$),

The joint p.d.f. of U and V is : $g(u, v) = f(x_1, x_2) \mid J \mid = \frac{1}{v}; 0 < u \leq v < 1$.

$$\therefore \quad g(u) = \int_u^1 g(u, v)\, dv = \int_u^1 \frac{1}{v}\, dv = \left|\log v\right|_u^1 = -\log u, \ 0 < u < 1$$



(iii) and (iv). Let $u = x_1 + x_2, v = x_1 - x_2$

$$\left.\begin{array}{ll} \text{i.e., } x_1 = \dfrac{u+v}{2} & \begin{array}{l} x_1 = 0 \Rightarrow u + v = 0 \\ \text{i.e., } v = -u \\ x_2 = 0 \Rightarrow u - v = 0 \\ \text{i.e., } v = u \end{array} \\ x_2 = \dfrac{u-v}{2} & \begin{array}{l} x_1 = 1 \Rightarrow u + v = 2 \\ x_2 = 1 \Rightarrow u - v = 2 \end{array} \end{array}\right\}$$

and $J = \begin{vmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{vmatrix} = -\frac{1}{2}$

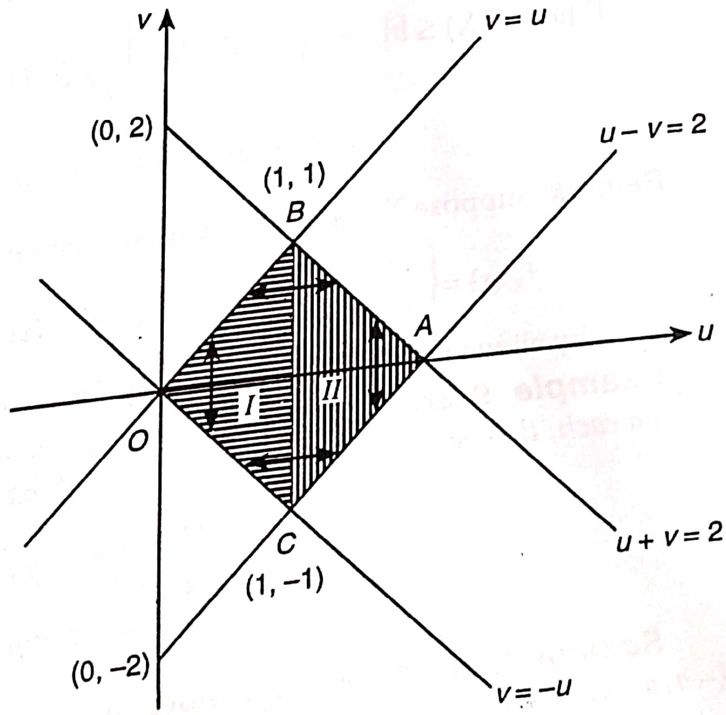
$$\therefore \quad g(u, v) = f(x_1, x_2) \mid J \mid = \tfrac{1}{2}, \ 0 < u < 2, -1 < v < 1$$

In region (I)

$$g_1(u) = \int_{-u}^{u} \tfrac{1}{2}\, dv = \tfrac{1}{2}\left|v\right|_{-u}^{u} = u$$

and in region (II),

$$g_2(u) = \int_{u-2}^{2-u} \tfrac{1}{2}\, dv = \tfrac{1}{2}\left|v\right|_{u-2}^{2-u} = 2 - u$$



∴ Distribution $U = X_1 + X_2$, is given by :

$$g(u) = \begin{cases} u, & 0 < u < 1 \\ 2 - u, & 1 < u < 2 \end{cases}$$

For the distribution of V, we split the region as : OAB and OAC.

In region OAB: $\quad h_1(v) = \int_v^{2-v} \frac{1}{2}\, du = \frac{1}{2}(2 - v - v) = 1 - v,\ 0 < v < 1$

In region OAC: $\quad h_2(v) = \int_{-v}^{2+v} \frac{1}{2}\, du = \frac{1}{2}[2(1 + v)] = 1 + v, -1 < v < 0$

Hence the distribution of $V = X_1 - X_2$ is given by: $h(v) = \begin{cases} 1 - v, & 0 < v < 1 \\ 1 + v, & -1 < v < 0 \end{cases}$

Example 9·26. *If X is a random variable with a continuous distribution function F, then prove that F(X) has a uniform distribution on [0, 1].*

Hence prove that : $P[a \le F(x) \le b] = b - a,\ 0 \le (a, b) \le 1.$

Solution. Since F is a distribution function, it is non-decreasing. Let $Y = F(X)$. Then the distribution function G of Y is given by :

$$G_Y(y) = P(Y \le y) = P[F(X) \le y] = P[X \le F^{-1}(y)],$$

the inverse exists, since F is non-decreasing and given to be continuous.

$\therefore \quad G_Y(y) = F[F^{-1}(y)],$ since F is the distribution function of X.

Thus $\quad G_Y(y) = y$

Therefore the *p.d.f. of* $Y = F(X)$ is given by : $g_Y(y) = \frac{d}{dy}[G_Y(y)] = 1$

Since F is a distribution function $Y = F(X)$ takes the values in the range [0, 1].

Hence $g_Y(y) = 1, 0 \le y \le 1 \quad \Rightarrow \quad Y$ is a uniform variate on [0, 1].

Since $\quad Y = F(X) \sim U[0, 1],$

$$P[a \le F(X) \le b] = P[a \le Y \le b] = \int_a^b g(y)dy$$

$$= \int_a^b 1.\, dy = \left| y \right|_a^b = b - a.$$

Remark. Suppose X is a random variable with *p.d.f.*,

$$f_X(x) = \begin{cases} e^{-x}, x \ge 0 \\ 0, \text{otherwise} \end{cases}, \qquad \text{then } F(x) = \begin{cases} 0, \text{if } x < 0 \\ 1 - e^{-x}, \text{if } x \ge 0 \end{cases}$$

Then by above result $F(X) = 1 - e^{-X}$ is uniformly distributed on [0, 1].

Example 9·27. *If X and Y are independent rectangular variates for the range − a to a each, then show that the sum X + Y = U, has the probability density*

$$\varphi(u) = \begin{cases} \dfrac{2a + u}{4a^2}, -2a \le u \le 0 \\[2mm] \dfrac{2a - u}{4a^2}, 0 \le u \le 2a \end{cases}$$

Solution. Since X and Y are independent rectangular variates, each in the interval $(-a, a)$, we have

$$f_1(x) = \begin{cases} \dfrac{1}{2a}, -a < x < a \\ 0, \text{elsewhere} \end{cases} \quad \text{and} \quad f_2(y) = \begin{cases} \dfrac{1}{2a}, -a < y < a \\ 0, \text{elsewhere} \end{cases}$$

and its *m.g.f.* is : $M_X(t) = (e^t - 1)^2 / t^2$, ... (9·21c)
which is left as an exercise to the reader.

5. In particular, replacing a by $-2a$, b by $2a$ and c by 0, the *p.d.f.* of triangular distribution on the interval $(-2a, 2a)$ with peak at $x = 0$ is given by :

$$f(x) = \begin{cases} (2a+x)/4a^2\,; & -2a < x < 0 \\ (2a-x)/4a^2; & 0 < x < 2a \end{cases} \qquad \cdots (9\cdot21d)$$

The *m.g.f.* of (9·21*d*) is given by :

$$M_X(t) = \int_{-2a}^{2a} e^{tx} f(x)\,dx = \frac{1}{4a^2}\left\{\int_{-2a}^{0} e^{tx}(2a+x)\,dx + \int_{0}^{2a} e^{tx}(2a-x)\,dx\right\}$$

$$= \frac{1}{4a^2}\left[e^{tx}\left\{\frac{2a+x}{t} - \frac{1}{t^2}\right\}\right]_{-2a}^{0} + \frac{1}{4a^2}\left[e^{tx}\left\{\frac{2a-x}{t} + \frac{1}{t^2}\right\}\right]_{0}^{2a} \qquad \text{[On integrating by parts]}$$

$$= \frac{1}{4a^2}\left[-\frac{2}{t^2} + \frac{1}{t^2}\left\{e^{2at} + e^{-2at}\right\}\right] = \frac{1}{4a^2t^2}\left\{e^{2at} + e^{-2at} - 2\right\} = \left\{\frac{1}{2at}(e^{at} - e^{-at})\right\}^2$$

Aliter. We may obtain (9·21) directly from (9·2*a*) on replacing a by $-2a$, b by $2a$ and c by 0.

Example 9·29. *If X and Y are i.i.d. U [– a, a] variates, find the p.d.f. of Z = X + Y and identify the distribution.*

Solution. Since X and Y are *i.i.d.* U $[-a, a]$, we have : [*c.f.* § 9·3·2],

$$M_X(t) = M_Y(t) = (e^{at} - e^{-at})/(2\,at) \qquad \cdots (*)$$

$$M_{X+Y}(t) = M_X(t)\,M_Y(t) = \left\{\frac{1}{2at}(e^{at} - e^{-at})\right\}^2, \text{ since X and Y are independent.}$$

But, this is the *m.g.f.* of Trg $(-2a, 2a)$ variate with peak at $x = 0$.
[*c.f.* Remark 5, equation (9·21*e*)]

Hence by uniqueness theorem of *m.g.f.*, $Z = X + Y \sim$ Trg $(-2a, 2a)$ with *p.d.f.* as given in (9·21*d*), Remark 5.

Aliter.
$$M_{X+Y}(t) = \frac{1}{4a^2t^2}(e^{2at} - 2 + e^{-2at}) \qquad \text{[From (**)]}$$

$$= \frac{2}{t^2}\left[\frac{e^{-2at}}{(-2a-0)(-2a-2a)} + \frac{e^{0\cdot t}}{(0+2a)(0-2a)} + \frac{e^{2at}}{(2a-0)(2a+2a)}\right]$$

which is of the form (9·21*a*), [*c.f.* Remark 3], with a replaced by $-2a$ and b replaced by $2a$ and c by 0. Hence $X + Y \sim$ Trg $(-2a, 2a)$ with peak at $x = 0$ and *p.d.f.* $p(x)$ given in (9·21*d*).

Remarks 1. The distribution of $X + Y$ has also been obtained in *Example* 9·27.

2. Similarly we can find the distribution of $X - Y$.

$$M_{X-Y}(t) = M_X(t).\,M_Y(-t) = \left\{\frac{1}{2at}(e^{at} - e^{-at})\right\}^2 \qquad \text{[From (*)]}$$

$\Rightarrow$ $X - Y \sim$ Trg $(-2a, 2a)$, with peak at $x = 0$.

9·5. GAMMA DISTRIBUTION

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

$$f(x) = \begin{cases} \dfrac{e^{-x}\,x^{\lambda-1}}{\Gamma(\lambda)}\,; & \lambda > 0,\ 0 < x < \infty \\ 0, & \text{otherwise} \end{cases} \qquad \ldots (9\cdot22)$$

Remarks 1. X is known as a Gamma variate with parameter λ and referred to as a $\gamma(\lambda)$ variate.

2. The function $f(x)$ defined above represents a probability function, since

$$\int_0^\infty f(x)\,dx = \frac{1}{\Gamma(\lambda)}\int_0^\infty e^{-x}x^{\lambda-1}\,dx = \frac{1}{\Gamma(\lambda)}\cdot\Gamma(\lambda) = 1$$

3. A continuous random variable X having the following *p.d.f.* is said to have a gamma distribution with two parameters a and λ.

$$f(x) = \begin{cases} \dfrac{a^\lambda}{\Gamma(\lambda)} e^{-ax}x^{\lambda-1}\,; a>0, \lambda>0, 0<x<\infty \\ 0, \text{ otherwise} \end{cases} \qquad \ldots(9\cdot22a)$$

Here $X \sim \gamma(a, \lambda)$. Taking $a = 1$ in (9·22*a*), we get (9·22). Hence we may write $X \sim \gamma(\lambda) = \gamma(1, \lambda)$.

4. The *cumulative distribution function,* called incomplete gamma function is defined as :

$$F_X(x) = \begin{cases} \displaystyle\int_0^x f(u)\,du = \frac{1}{\Gamma(\lambda)}\int_0^x e^{-u}u^{\lambda-1}\,du, x>0 \\ 0, \text{ otherwise} \end{cases} \qquad \ldots(9\cdot22b)$$

9·5·1. M.G.F. of Gamma Distribution. M.G.F. about origin is given by :

$$M_x(t) = E(e^{tX}) = \int_0^\infty e^{tx}f(x)\,dx = \frac{1}{\Gamma(\lambda)}\int_0^\infty e^{tx}e^{-x}x^{\lambda-1}\,dx$$

$$= \frac{1}{\Gamma(\lambda)}\int_0^\infty e^{-(1-t)x}x^{\lambda-1}\,dx = \frac{1}{\Gamma(\lambda)}\cdot\frac{\Gamma(\lambda)}{(1-t)^\lambda},\ |t|<1$$

$$\therefore \quad M_X(t) = (1-t)^{-\lambda},\ |t|<1 \qquad \ldots(9\cdot23)$$

9·5·2. Cumulant Generating Function of Gamma Distribution. The cumulant generating function $K_X(t)$ is given by :

$$K_X(t) = \log M_X(t) = \log(1-t)^{-\lambda} = -\lambda\log(1-t)\,;\ |t|<1$$

$$= \lambda\left(t + \frac{t^2}{2} + \frac{t^3}{3} + \frac{t^4}{4} + \ldots\right)$$

$\therefore$ Mean $= \kappa_1 =$ Coefficient of t in $K_X(t) = \lambda$

Variance $= \mu_2 = \kappa_2 =$ Coefficient of $\frac{t^2}{2!}$ in $K_X(t) = \lambda$

Hence if $X \sim \gamma(\lambda)$, Mean = Variance = λ.

$\mu_3 = \kappa_3 =$ Coefficient of $\frac{t^3}{3!}$ in $K_X(t) = 2\lambda$

$\kappa_4 =$ Coefficient of $\frac{t^4}{4!}$ in $K_X(t) = 6\lambda \Rightarrow \mu_4 = \kappa_4 + 3\kappa_2^2 = 6\lambda + 3\lambda^2$.

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

Remarks 1. Like Poisson distribution, the mean and variance of the Gamma distribution are also equal. However, Poisson distribution is discrete while Gamma distribution is continuous.

2. **Limiting form of Gamma distribution as $\lambda \to \infty$.** We know that if $X \sim \gamma(\lambda)$, then $E(X) = \lambda = \mu$, (say) and $\text{Var}(X) = \lambda = \sigma^2$, (say). Then standard gamma variate is given by: $Z = \dfrac{X-\mu}{\sigma} = \dfrac{X-\lambda}{\sqrt{\lambda}}$

$$M_Z(t) = \exp(-\mu t/\sigma) M_X(t/\sigma) = \exp(-\mu t/\sigma)\left(1-\frac{t}{\sigma}\right)^{-\lambda} = e^{-t\lambda/\sqrt{\lambda}}\left(1-\frac{t}{\sqrt{\lambda}}\right)^{-\lambda}$$

$$\Rightarrow \quad K_Z(t) = \sqrt{\lambda}\,.\,t - \lambda \log\left(1-\frac{t}{\sqrt{\lambda}}\right) = -\sqrt{\lambda}\; t - \lambda\left(\frac{t}{\sqrt{\lambda}} + \frac{t^2}{2\lambda} + \frac{t^3}{3\lambda^{3/2}} + \ldots\right)$$

$$= -\sqrt{\lambda}\; t + \sqrt{\lambda}\; t + \frac{t^2}{2} + 0(\lambda^{-1/2}),$$

where $0(\lambda^{-1/2})$ are terms containing $\lambda^{1/2}$ and higher powers of λ in the denominator.

$$\therefore \quad \lim_{\lambda\to\infty} K_Z(t) = \frac{t^2}{2} \quad \Rightarrow \quad \lim_{\lambda\to\infty} M_Z(t) = \exp(t^2/2)$$

which is the *m.g.f.* of a Standard Normal Variate. Hence by uniqueness theorem of *m.g.f.*, Standard Gamma variate tends to Standard Normal Variate as $\lambda \to \infty$. In other words, *Gamma distribution tends to Normal distribution for large value of parameter* λ.

3. For the two parameter gamma distribution (9·22 *a*), we have

$$M_X(t) = \left(1-\frac{t}{a}\right)^{-\lambda} ;\, t < a. \quad \ldots(9\cdot23\, a)$$

Proof is left as an excercise to the reader.

Also $\quad K_X(t) = -\lambda \log\left(1-\frac{t}{a}\right) = \lambda\left\{\frac{t}{a} + \frac{1}{2}\left(\frac{t}{a}\right)^2 + \frac{1}{3}\left(\frac{t}{a}\right)^3 + \ldots\right\} ;\, t < a$

$\therefore \quad$ Mean $= \kappa_1 = \lambda/a$ and Variance $= \kappa_2 = \lambda/a^2 =$ Mean$/a$ $\quad \ldots(9\cdot23\, b)$

Hence Variance > Mean if $a < 1$; Variance = Mean if $a = 1$; and Variance < Mean if $a > 1$.

9·5·3. Additive Property of Gamma Distribution. *The sum of independent Gamma variates is also a Gamma variate. More precisely, if* $X_1, X_2, \ldots, X_k$ *are independent Gamma variates with parameters* $\lambda_1, \lambda_2, \ldots, \lambda_k$ *respectively then* $X_1 + X_2 + \ldots + X_k$ *is also a Gamma variate with parameter* $\lambda_1 + \lambda_2 + \ldots + \lambda_k$.

Proof. Since X_i is a $\gamma(\lambda_i)$ variate, $M_{X_i}(t) = (1-t)^{-\lambda_i}$

The *m.g.f.* of the sum $X_1 + X_2 + \ldots + X_k$ is given by :

$$M_{X_1+X_2+\ldots X_k}(t) = M_{X_1}(t)\, M_{X_2}(t) \ldots M_{X_k}(t), \quad (\because X_1, X_2, \ldots, X_k \text{ are independent.})$$

$$= (1-t)^{-\lambda_1}(1-t)^{-\lambda_2} \ldots (1-t)^{-\lambda_k} = (1-t)^{-(\lambda_1+\lambda_2+\ldots+\lambda_k)}$$

which is the *m.g.f.* of a Gamma variate with parameter $\lambda_1 + \lambda_2 + \ldots + \lambda_k$. Hence the result follows by the uniqueness theorem of *m.g.f.*'s.

Remark. In general, if $X_i \sim \gamma(a, \lambda_i)$, $i = 1, 2, \ldots, n$ are independent *r.v.*'s, then

$$\sum_{i=1}^{n} X_i \sim \gamma\left(a, \sum_{i=1}^{n} \lambda_i\right).$$

9·6. BETA DISTRIBUTION OF FIRST KIND

Definition. *A r.v. X is said to have a beta distribution of first kind with parameters* μ *and* ν $(\mu > 0,\ \nu > 0)$ *if its p.d.f. is given by :*

$$f(x) = \begin{cases} \dfrac{1}{B(\mu, \nu)} \cdot x^{\mu-1}(1-x)^{\nu-1} ; & (\mu, \nu) > 0,\ 0 < x < 1 \\ 0, \text{ otherwise} \end{cases}$$

[where $B(\mu, \nu)$ *is the Beta function].* $\quad \ldots(9\cdot24)$

The r.v. X is known as a Beta variate of the first kind with parameters μ and ν and is referred to as $\beta_1(\mu, \nu)$ variate.

Remarks. 1. The cumulative distribution function, often called the *Incomplete Beta Function, is given by :*

$$F(x) = \begin{cases} 0, & x < 0 \\ \displaystyle\int_0^x \frac{1}{B(\mu, \nu)} u^{\mu-1}(1-u)^{\nu-1}\, du\,; & 0 < x < 1, (\mu, \nu) > 0 \\ 1, & x > 1 \end{cases} \qquad \ldots(9{\cdot}24a)$$

2. In particular, if we take $\mu = 1\ \nu = 1$ in (9·24), we get

$$f(x) = \frac{1}{\beta(1, 1)} = 1, 0 < x < 1 \qquad \ldots(9{\cdot}24b)$$

which is the *p.d.f.* of uniform distribution on [0, 1].

3. If $X \sim \beta_1(\mu, \nu)$, then it can be easily proved that $(1 - X) \sim \beta_1(\nu, \mu)$.

9·6·1. Constants of Beta Distribution of First Kind.

$$\mu_r' = \int_0^1 x^r f(x)\, dx = \frac{1}{B(\mu, \nu)} \int_0^1 x^{\mu+r-1}(1-x)^{\nu-1}\, dx = \frac{1}{B(\mu, \nu)} B(\mu + r, \nu)$$

$$= \frac{\Gamma(\mu+r)\,\Gamma(\nu)}{\Gamma(\mu+r+\nu)} \cdot \frac{\Gamma(\mu+\nu)}{\Gamma(\mu)\,\Gamma(\nu)} = \frac{\Gamma(\mu+r)\,\Gamma(\mu+\nu)}{\Gamma(\mu+r+\nu)\,\Gamma(\mu)} \qquad \ldots(9{\cdot}24c)$$

In particular

$$\text{Mean} = \mu_1' = \frac{\Gamma(\mu+1)}{\Gamma(\mu+\nu+1)} \cdot \frac{\Gamma(\mu+\nu)}{\Gamma(\mu)} = \frac{\mu\,\Gamma(\mu)\,\Gamma(\mu+\nu)}{(\mu+\nu)\,\Gamma(\mu+\nu)\,\Gamma(\mu)} = \frac{\mu}{\mu+\nu} \qquad \ldots(9{\cdot}24\,d)$$

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

$$\mu_2' = \frac{\Gamma(\mu+2)\,.\,\Gamma(\mu+\nu)}{\Gamma(\mu+\nu+2)\,\Gamma(\mu)} = \frac{(\mu+1)\,\mu\,\Gamma(\mu)\,\Gamma(\mu+\nu)}{(\mu+\nu+1)(\mu+\nu)\,\Gamma(\mu+\nu)\,\Gamma(\mu)} = \frac{\mu(1+\mu)}{(\mu+\nu)(\mu+\nu+1)}$$

Hence
$$\mu_2 = \mu_2' - \mu_1'^2 = \frac{\mu(1+\mu)}{(\mu+\nu)(\mu+\nu+1)} - \left(\frac{\mu}{\mu+\nu}\right)^2$$

$$= \frac{\mu}{(\mu+\nu)^2(\mu+\nu+1)}\left[(\mu+\nu)(\mu+1) - \mu(\mu+\nu+1)\right] = \frac{\mu\nu}{(\mu+\nu)^2(\mu+\nu+1)}$$

Similarly, we have

$$\mu_3 = \mu_3' - 3\mu_2'\mu_1' + 2\mu_1'^3 = \frac{2\mu\nu(\nu-\mu)}{(\mu+\nu)^3(\mu+\nu+1)(\mu+\nu+2)}$$

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

$$= \frac{3\mu\nu\{\mu\nu(\mu+\nu-6) + 2(\mu+\nu)^2\}}{(\mu+\nu)^4(\mu+\nu+1)(\mu+\nu+2)(\mu+\nu+3)}$$

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} = \frac{4(\nu-\mu)^2(\mu+\nu+1)}{\mu\nu(\mu+\nu+2)^2} \quad \text{and} \quad \beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{3(\mu+\nu+1)\,\mu\nu(\mu+\nu-6) + 2(\mu+\nu)^2}{\mu\nu(\mu+\nu+2)(\mu+\nu+3)}$$

The harmonic mean H is given by :

$$\frac{1}{H} = \int_0^1 \frac{1}{x} f(x)\, dx = \frac{1}{B(\mu, \nu)} \int_0^1 x^{\mu-2}(1-x)^{\nu-1}\, dx$$

$$= \frac{1}{B(\mu, \nu)} B(\mu-1, \nu) = \frac{\Gamma(\mu-1)\,\Gamma(\nu)}{\Gamma(\mu+\nu-1)} \cdot \frac{\Gamma(\mu+\nu)}{\Gamma(\mu)\,\Gamma(\nu)}$$

$$= \frac{\Gamma(\mu-1)\,(\mu+\nu-1)\,\Gamma(\mu+\nu-1)}{\Gamma(\mu+\nu-1)\,(\mu-1)\,\Gamma(\mu-1)} = \frac{\mu+\nu-1}{\mu-1}$$

$$\therefore \qquad H = \frac{\mu-1}{\mu+\nu-1} \qquad \ldots(9\cdot24e)$$

9·7. BETA DISTRIBUTION OF SECOND KIND

Definition. *A r.v. X is said to have a beta distribution of the second kind with parameters μ and ν, (μ > 0, ν > 0), if its p.d.f. is given by :*

$$f(x) = \begin{cases} \dfrac{1}{B(\mu,\nu)} \cdot \dfrac{x^{\mu-1}}{(1+x)^{\mu+\nu}}; & (\mu,\nu) > 0,\ 0 < x < \infty \\ 0, \text{ otherwise} \end{cases} \qquad \ldots(9\cdot25)$$

Remarks 1. The r.v. X is *known as a Beta variate of second kind with parameters* μ and ν and is denoted as $\beta_2(\mu, \nu)$ variate.

2. Beta distribution of second kind is transformed to Beta distribution of first kind by the transformation :

$$1 + x = \frac{1}{y} \Rightarrow y = \frac{1}{1+x} \qquad \ldots(*)$$

Thus, if $X \sim \beta_2(\mu, \nu)$, then Y defined in (*) is a $\beta_1(\mu, \nu)$.

The proof is left as an exercise to the reader.

9·7·1. Constants of Beta Distribution of Second kind.

$$\mu_r' = \int_0^\infty x^r f(x)\,dx = \frac{1}{B(\mu,\nu)} \int_0^\infty \frac{x^{\mu+r-1}}{(1+x)^{\mu+\nu}}\,dx$$

$$= \frac{1}{B(\mu,\nu)} \int_0^\infty \frac{x^{(\mu+r)-1}}{(1+x)^{\mu+r+\nu-r}}\,dx = \frac{1}{B(\mu,\nu)} \cdot B(\mu+r, \nu-r)$$

$$= \frac{\Gamma(\mu+r)\,\Gamma(\nu-r)}{\Gamma(\mu+\nu)} \cdot \frac{\Gamma(\mu+\nu)}{\Gamma(\mu)\,\Gamma(\nu)} = \frac{\Gamma(\mu+r)\,\Gamma(\nu-r)}{\Gamma(\mu)\,\Gamma(\nu)}; \nu > r$$

In particular

$$\mu_1' = \frac{\Gamma(\mu+1)\,\Gamma(\nu-1)}{\Gamma(\mu)\,\Gamma(\nu)} = \frac{\mu\,\Gamma(\mu)\,\Gamma(\nu-1)}{\Gamma(\mu)\,(\nu-1)\,\Gamma(\nu-1)} = \frac{\mu}{\nu-1}, \nu > 1$$

$$\mu_2' = \frac{\Gamma(\mu+2)\,\Gamma(\nu-2)}{\Gamma(\mu)\,\Gamma(\nu)} = \frac{(\mu+1)\,\mu\,\Gamma(\mu)\,\Gamma(\nu-2)}{\Gamma(\mu)\,(\nu-1)\,(\nu-2)\,\Gamma(\nu-2)} = \frac{\mu(\mu+1)}{(\nu-1)(\nu-2)}, \nu > 2$$

$$\therefore \qquad \mu_2 = \mu_2' - \mu_1'^2 = \frac{\mu(\mu+1)}{(\nu-1)(\nu-2)} - \left(\frac{\mu}{\nu-1}\right)^2$$

$$= \frac{\mu}{\nu-1}\left[\frac{(\nu-1)(\mu+1) - \mu(\nu-2)}{(\nu-1)\,(\nu-2)}\right] = \frac{\mu(\mu+\nu-1)}{(\nu-1)^2(\nu-2)}$$

The harmonic mean H is given by :

$$\frac{1}{H} = E\left(\frac{1}{X}\right) = \int_0^\infty \frac{1}{x} \cdot f(x)\,dx = \frac{1}{B(\mu,\nu)} \int_0^\infty \frac{x^{\mu-2}}{(1+x)^{\mu+\nu}}\,dx$$

$$= \frac{1}{B(\mu, \nu)} \int_0^\infty \frac{x^{\mu-1-1}}{(1+x)^{\mu-1+\nu+1}}\, dx = \frac{1}{B(\mu, \nu)} \cdot B(\mu-1, \nu+1), \mu > 1$$

$$= \frac{\Gamma(\mu-1)\,\Gamma(v+1)}{\Gamma(\mu+v)} \cdot \frac{\Gamma(\mu+v)}{\Gamma(\mu)\,\Gamma(v)} = \frac{\Gamma(\mu-1)\, v\, \Gamma(v)}{(\mu-1)\,\Gamma(\mu-1)\,\Gamma(v)} = \frac{v}{\mu-1}$$

Hence $$H = \frac{\mu-1}{\nu}$$

Example 9·30. *The daily consumption of milk in a city, in excess of 20,000 litres, is approximately distributed as a Gamma variate with parameters* $a = \dfrac{1}{10{,}000}$ and $\lambda = 2$. *The city has a daily stock of 30,000 litres. What is the probability that the stock is insufficient on a particular day ?*

Solution. If the $r.v.$ X denotes the daily consumption of milk (in litres) in a city, then the $r.v.$ $Y = X - 20{,}000$ has a gamma distribution with $p.d.f.$;

$$g(y) = \frac{1}{(10{,}000)^2\, \Gamma(2)}\, y^{2-1} e^{-y/10{,}000} = \frac{y\, e^{-y/10{,}000}}{(10{,}000)^2}\,; 0 < y < \infty$$

Since the daily stock of the city is 30,000 litres, the required probability 'p' that the stock is insufficient on a particular day is given by :

$$p = P(X > 30{,}000) = P(Y > 10{,}000) = \int_{10{,}000}^\infty g(y)\, dy = \int_{10{,}000}^\infty \frac{y\, e^{-y/10{,}000}}{(10{,}000)^2}\, dy$$

$$= \int_1^\infty z\, e^{-z}\, dz \qquad \text{[Taking } z = y/10{,}000]$$

Integrating by parts, $p = \left| -z\, e^{-z} \right|_1^\infty + \int_1^\infty e^{-z}\, dz = e^{-1} - \left| e^{-z} \right|_1^\infty = e^{-1} + e^{-1} = \dfrac{2}{e}$

Remark. Since $\lambda = 2$, the integration is easily done. However, for general values of a and λ, the integral is evaluated by using tables of Incomplete Gamma Integral [see Tables of Incomplete Gamma Functions, K. Pearson; Cambridge University Press] of the form :

$\displaystyle\int_0^\infty \frac{e^{-x}\, x^{n-1}}{\Gamma n}\, dx$, which have been tabulated for different values of α and n.

Example 9·31. *If* $X \sim N(\mu, \sigma^2)$, *obtain the p.d.f. of :* $U = \dfrac{1}{2}\left(\dfrac{X-\mu}{\sigma}\right)^2$.

Solution. Since $X \sim N(\mu, \sigma^2)$ $\Rightarrow$ $z = \dfrac{x-\mu}{\sigma} \sim N(0, 1)$ with $p.d.f.$: ... (*i*)

$$\phi(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} z^2\right); -\infty < z < \infty.$$

The distribution function $G(\cdot)$ of $U = \dfrac{1}{2}\left(\dfrac{X-\mu}{\sigma}\right)^2$ is given by :

$$G_U(u) = P(U \le u,) = P\left[\frac{1}{2}\left(\frac{X-\mu}{\sigma}\right)^2 \le u\right]$$

$$= P(Z^2 \le 2u) = P\left(-\sqrt{2u} \le Z \le \sqrt{2u}\right), \quad \text{where } Z \sim N(0, 1)$$

$$= P(Z \le \sqrt{2u}) - P(Z \le -\sqrt{2u})$$

$$= \Phi(\sqrt{2u}) - \Phi(-\sqrt{2u}) \quad \ldots (ii)$$

where $\Phi(\cdot)$ is the distribution function of standard normal variate (SNV) Z.

Differentiating *w.r.* to u, the *p.d.f.* $g(\cdot)$ of U is given by :

$$g(u) = \phi(\sqrt{2u}) \cdot \frac{d}{du}(\sqrt{2u}) - \phi(-\sqrt{2u}) \cdot \frac{d}{du}(-\sqrt{2u})$$

[$\phi(\cdot)$ is p.d.f. of SNV Z]

$$= \frac{1}{\sqrt{2u}}\left[\phi(\sqrt{2u}) + \phi(-\sqrt{2u})\right]$$

$$= \frac{1}{\sqrt{2u}} \cdot 2\phi(\sqrt{2u}) \qquad [\because \phi(u) \text{ is even function of } u]$$

$$= \sqrt{\frac{2}{u}} \cdot \frac{1}{\sqrt{2\pi}} \exp.\left(-\frac{1}{2} \cdot 2u\right) \qquad [\text{From } (i)]$$

$$= \frac{1}{\Gamma(1/2)} e^{-u} \cdot u^{(1/2)-1}, \; u \ge 0$$

$$\left[\because \sqrt{\pi} = \Gamma(1/2) \text{ and } u = \frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2 \ge 0\right]$$

which is the *p.d.f.* of gamma distribution with parameter $\frac{1}{2}$.

Hence $\quad U = \frac{1}{2}\left(\frac{X-\mu}{\sigma}\right)^2$ is a $\gamma\left(\frac{1}{2}\right)$ variate

Example 9·32. *Show that the mean value of positive square root of a $\gamma(\mu)$ variate is $\Gamma(\mu + \frac{1}{2})/\Gamma(\mu)$. Hence prove that the mean deviation of a normal variate from its mean is $\sqrt{2/\pi}$, where σ is the standard deviation of the distribution.*

Solution. Let X be a $\gamma(\mu)$ variate. Then $f(x) = \dfrac{e^{-x} x^{\mu-1}}{\Gamma(\mu)}$; $\mu > 0, 0 < x < \infty$

$$\therefore \quad E(\sqrt{X}) = \int_0^\infty x^{1/2} f(x)\, dx = \frac{1}{\Gamma(\mu)} \int_0^\infty e^{-x} x^{\mu + (1/2) - 1}\, dx = \frac{\Gamma(\mu + \frac{1}{2})}{\Gamma(\mu)} \quad \ldots (*)$$

If $X \sim N(\mu, \sigma^2)$, then $U = \frac{1}{2}\left(\frac{X-\mu}{\sigma}\right)^2$ is a $\gamma\left(\frac{1}{2}\right)$ variate. (*c.f.* Example 9·31)

$$\therefore \quad |X - \mu| = \sqrt{2}\, \sigma U^{1/2}, \text{ where } U \text{ is a } \gamma\left(\tfrac{1}{2}\right) \text{ variate.}$$

Hence mean deviation of X about mean is given by :

$$E|X-\mu| = E(\sqrt{2}\, \sigma U^{1/2}) = \sqrt{2}\, \sigma E(U^{1/2})$$

$$= \sqrt{2}\, \sigma \frac{\Gamma\left(\frac{1}{2} + \frac{1}{2}\right)}{\Gamma\left(\frac{1}{2}\right)} = \frac{\sqrt{2}\, \sigma}{\sqrt{\pi}} = \sigma\sqrt{2/\pi}. \qquad \left[\text{Using } (*) \text{ with } \mu = \tfrac{1}{2}\right]$$

Example 9·33. *If X and Y are independent Gamma variates with parameters* μ *and* ν *respectively, show that the variables* $U = X + Y$, $Z = \dfrac{X}{X+Y}$ *are independent and that U is a* $\gamma(\mu+\nu)$ *variate and Z is a* $\beta_1(\mu, \nu)$ *variate.*

Solution. Since X is a $\gamma(\mu)$ variate and Y is a $\gamma(\nu)$ variate, we have

$$f_1(x)\,dx = \frac{1}{\Gamma(\mu)} e^{-x} x^{\mu-1}\,dx\,;\, 0 < x < \infty, \mu > 0$$

$$f_2(y)\,dy = \frac{1}{\Gamma(\nu)} e^{-y} y^{\nu-1}\,dy\,;\, 0 < y < \infty, \nu > 0$$

Since X and Y are independently distributed, their joint probability differential is given by the compound probability theorem as shown below :

$$dF(x, y) = f_1(x) f_2(y)\,dx\,dy = \frac{1}{\Gamma\mu\,\Gamma\nu} e^{-(x+y)} x^{\mu-1} y^{\nu-1}\,dx\,dy$$

Now $u = x + y,\ z = \dfrac{x}{x+y}$, so that $x = uz,\ y = u - x = u(1-z)$

Jacobian of transformation J is given by :

$$J = \frac{\partial(x, y)}{\partial(u, z)} = \begin{vmatrix} \dfrac{\partial x}{\partial u} & \dfrac{\partial y}{\partial u} \\ \dfrac{\partial x}{\partial z} & \dfrac{\partial y}{\partial z} \end{vmatrix} = \begin{vmatrix} z & 1-z \\ u & -u \end{vmatrix} = -u$$

As X and Y range from 0 to ∞, u ranges from 0 to ∞ and z from 0 to $1\left(\because \dfrac{x}{x+y} \le 1\right)$.

Hence the joint distribution of U and Z is given by :

$$dG(u, z) = g(u, z)\,du\,dz = \frac{1}{\Gamma(\mu)\Gamma(\nu)} e^{-u} (uz)^{\mu-1} [u(1-z)]^{\nu-1}\,|\,J\,|\,du\,dz$$

$$= \frac{1}{\Gamma(\mu)\Gamma(\nu)} \cdot e^{-u} u^{\mu+\nu-1} z^{\mu-1} (1-z)^{\nu-1}\,du\,dz$$

$$= \left\{\frac{e^{-u} u^{\mu+\nu-1}}{\Gamma(\mu+\nu)}\,du\right\}\left\{\frac{1}{B(\mu, \nu)} z^{\mu-1}(1-z)^{\nu-1}\,dz\right\} \quad \ldots (*)$$

$$= [g_1(u)\,du]\,[g_2(z)\,dz], \text{ (say)},$$

$$\left.\begin{aligned} \text{where} \quad g_1(u) &= \frac{1}{\Gamma(\mu+\nu)} e^{-u} u^{\mu+\nu-1}, 0 < u < \infty \\ \text{and} \quad g_2(z) &= \frac{1}{B(\mu, \nu)} z^{\mu-1}(1-z)^{\nu-1}, 0 < z < 1 \end{aligned}\right\} \quad \ldots (**)$$

From (*) and (**), we conclude that U and Z are independently distributed, U as a $\gamma(\mu+\nu)$ variate and Z as a $\beta_1(\mu, \nu)$ variate.

Example 9·34. *If X and Y are independent Gamma variates with parameters* μ *and* ν *respectively, show that* $U = X + Y$, $Z = \dfrac{X}{Y}$ *are independent and that U is a* $\gamma(\mu+\nu)$ *variate and Z is a* $\beta_2(\mu, \nu)$ *variate.*

Solution. As in example 9·33, we have

$$dF(x, y) = \frac{1}{\Gamma(\mu)\Gamma(\nu)} e^{-(x+y)} x^{\mu-1} y^{\nu-1}\,dx\,dy, 0 < (x, y) < \infty$$

Since $u = x + y$ and $z = \dfrac{x}{y}$, $1 + z = 1 + \dfrac{x}{y} = \dfrac{u}{y} \Rightarrow y = \dfrac{u}{1+z}$ and $x = \dfrac{uz}{1+z} = u\left(1 - \dfrac{1}{1+z}\right)$

$$J = \frac{\partial (x,y)}{\partial (u,z)} = \frac{-u}{(1+z)^2}$$

As x and y range from 0 to ∞, both u and z range from 0 to ∞. Hence the joint probability differential of random variables U and Z becomes :

$$dG(u, z) = \frac{1}{\Gamma(\mu)\,\Gamma(\nu)}\, e^{-u}\left(\frac{uz}{1+z}\right)^{\mu-1}\left(\frac{u}{1+z}\right)^{\nu-1} |\,J\,|\,du dz$$

$$= \left[\frac{e^{-u}\,u^{\mu+\nu-1}}{\Gamma(\mu+\nu)}\,du\right]\left[\frac{1}{B(\mu,\nu)}\cdot\frac{z^{\mu-1}}{(1+z)^{\mu+\nu}}\,dz\right]; 0 < u < \infty, 0 < z < \infty$$

Hence U and Z are independently distributed, U as a $\gamma(\mu+\nu)$ variate and Z as a $\beta_2(\mu, \nu)$ variate.

Remark. The above two examples lead to the following important results.

If X is *a* $\gamma(\mu)$ variate and Y is an independent $\gamma(\nu)$ variate, then

(*i*) $X + Y$ is a $\gamma(\mu+\nu)$ variate, *i.e.*, the sum of two independent Gamma variates is also a Gamma variate.

(*ii*) $\dfrac{X}{Y}$ is a $\beta_2(\mu, \nu)$ variate, *i.e.*, the ratio of two independent Gamma variates is β_2 – variate.

(*iii*) $X/(X + Y)$ is a $\beta_1(\mu, \nu)$ variate.

Example 9·35. *Let X and Y have joint p.d.f. :*

$$g(x, y) = \begin{cases} \dfrac{e^{-(x+y)}\, x^3\, y^4}{\Gamma 4\, \Gamma 5}, & x > 0, y > 0 \\ 0, \text{ elsewhere.} \end{cases}$$

Find (i) p.d.f. of $U = \dfrac{X}{X+Y}$, *(ii)* $E(U)$, *and (iii)* $E[U - E(U)]^2$.

Solution. Let $u = \dfrac{x}{x+y}$ and $v = x + y \;\Rightarrow\; x = uv, y = v - x = v - uv = v(1-u).$

Jacobian of transformation is : $J = \dfrac{\partial (x,y)}{\partial (u,v)} = \begin{vmatrix} \dfrac{\partial x}{\partial u} & \dfrac{\partial x}{\partial v} \\ \dfrac{\partial y}{\partial u} & \dfrac{\partial y}{\partial v} \end{vmatrix} = \begin{vmatrix} v & u \\ -v & 1-u \end{vmatrix} = v$

Hence joint p.d.f. of U and V becomes :

$$p(u, v) = g(x, y).\,|\,J\,| = \frac{1}{\Gamma 4\, \Gamma 5}\, e^{-v}.\,(uv)^3\,[\,v(1-u)\,]^4 \times v$$

$$= \frac{1}{\Gamma 4\, \Gamma 5}\, e^{-v}.\,v^8.\,u^3\,(1-u)^4;\; 0 \le u \le 1, v > 0$$

$\left(\because u = \dfrac{x}{x+y} < 1\right.$ and since $x > 0, y > 0$, we have $0 < u < 1$ and $\left.v = x + y \ge 0\right)$

$$= \left(\frac{1}{\Gamma 9}\, e^{-v}\, v^8\right)\left\{\frac{\Gamma 9}{\Gamma 4\, \Gamma 5}\, u^3\,(1-u)^4\right\};\; 0 < u < 1, v > 0$$

$\Rightarrow$ $\quad p(u, v) = p_1(v)\,.\,p_2(u)$ ...(*)

where $\quad p_1(v) = \dfrac{1}{\Gamma 9}\, e^{-v}\, v^8\,; v > 0$ and $p_2(u) = \dfrac{\Gamma 9}{\Gamma 4\, \Gamma 5}\, u^3\,(1-u)^4\,; 0 < u < 1$ ...(**)

From (*), we conclude that U and V are independently distributed and from (**), we conclude that :
$U = \frac{X}{X+Y} \sim \beta_1(4, 5)$. *i.e.*, U is a Beta variate of first kind with parameters $(4, 5)$.

Aliter. We have

$$g(x, y) = \frac{1}{\Gamma 4\,\Gamma 5} e^{-(x+y)} x^3 y^4 = \left[\frac{1}{\Gamma 4} e^{-x} x^3\right]\left[\frac{1}{\Gamma 5} e^{-y} y^4\right]$$

$$= g_1(x)\, g_2(y);\ x > 0, y > 0$$

$\Rightarrow$ X and Y are independently distributed and $X \sim \gamma(4)$ and $Y \sim \gamma(5)$.

Hence $\quad U = \frac{X}{X+Y} \sim \beta_1(4, 5)$

$$\therefore \quad E(U) = \int_0^1 u.\, p_2(u)\, du = \frac{1}{B(4,5)} \cdot \int_0^1 u^4 (1-u)^4\, du$$

$$= \frac{1}{B(4,5)} \cdot B(5,5) \qquad \text{[Using Beta integral]}$$

$$= \frac{\Gamma 9}{\Gamma 4\,\Gamma 5} \times \frac{\Gamma 5\,\Gamma 5}{\Gamma 10} = \frac{\Gamma 9 \,.\, 4\Gamma 4}{\Gamma 4 \,.\, 9\Gamma 9} = \frac{4}{9}$$

$$E(U^2) = \frac{1}{B(4,5)} \int_0^1 u^2 \,.\, u^3 (1-u)^4\, du$$

$$= \frac{1}{B(4,5)} \times B(6,5) = \frac{\Gamma 9}{\Gamma 4\,\Gamma 5} \times \frac{\Gamma 6\,\Gamma 5}{\Gamma 11} = \frac{5 \times 4}{10 \times 9} = \frac{2}{9}$$

$$E[U - E(U)]^2 = E(U^2) - [E(U)]^2 = \frac{2}{9} - \frac{16}{81} = \frac{2}{81}.$$

Example 9·36. *A random sample of size n is taken from a population with distribution :*

$$dP(x) = \frac{1}{\Gamma(\lambda)} e^{-x/a} \left(\frac{x}{a}\right)^{\lambda - 1} \frac{dx}{a};\ 0 < x < \infty, a > 0, \lambda > 0$$

Find the distribution of the mean $\overline{X}$.

Solution.

$$M_X(t) = E(e^{tX}) = \int_0^\infty e^{tx} f(x)\, dx = \frac{1}{\Gamma(\lambda)} \int_0^\infty e^{tx}\, e^{-x/a} \left(\frac{x}{a}\right)^{\lambda-1} \frac{dx}{a}$$

$$= \frac{1}{\Gamma(\lambda)\, a^\lambda} \int_0^\infty exp\left\{-\left(\frac{1}{a} - t\right) x\right\} x^{\lambda-1}\, dx = \frac{1}{\Gamma(\lambda)\, a^\lambda} \cdot \frac{\Gamma(\lambda)}{\left(\frac{1}{a} - t\right)^\lambda} = (1 - at)^{-\lambda} \qquad \ldots(*)$$

$$\therefore\ M_{\overline{X}}(t) = M_{(X_1 + X_2 + \ldots + X_n)/n}(t) = M_{X_1 + X_2 + \ldots + X_n}(t/n) \qquad [\because M_{cX}(t) = M_X(ct)]$$

$$= M_{X_1}(t/n)\, M_{X_2}(t/n) \ldots M_{X_n}(t/n), \qquad \text{(Since } X_1, X_2, \ldots, X_n \text{ are independent.)}$$

Hence on using (*), we get $\quad M_{\overline{X}}(t) = \left[\left(1 - \frac{at}{n}\right)^{-\lambda}\right]^n = \left(1 - \frac{t}{n/a}\right)^{-n\lambda}$

which is the *m.g.f.* of a Gamma distribution (*c.f.* Remark 3, § 9·5·2). Hence by uniqueness theorem of *m.g.f.*, $\overline{X} \sim \gamma(n/a, n\lambda)$ with *p.d.f.* :

$$g(\bar{x}) = \frac{(n/a)^{\lambda n}}{\Gamma(\lambda n)} e^{-n\bar{x}/a} (\bar{x})^{n\lambda - 1}, 0 < \bar{x} < \infty$$

Example 9·37. *A sample of n values is drawn from a population whose probability density is ae^{-ax}, $(x \geq 0, a > 0)$. If $\overline{X}$ is mean of the sample, show that $na\overline{X}$ is a $\gamma(n)$ variate and prove that $E(\overline{X}) = \frac{1}{a}$ and S.E. of $(\overline{X}) = \frac{1}{a\sqrt{n}}$.*

Solution. $f(x) = ae^{-ax}; 0 \leq x < \infty, a > 0$

$$\therefore \quad M_X(t) = \int_0^\infty e^{tx} f(x)\, dx = a\int_0^\infty e^{-(a-t)x}\, dx = a \left| \frac{e^{-(a-t)x}}{-(a-t)} \right|_0^\infty = \frac{a}{a-t}, \quad (a > t) \qquad \dots (*)$$

$$\because \quad an\overline{X} = an\left(\frac{X_1 + X_2 + \dots + X_n}{n}\right) = a(X_1 + X_2 + \dots + X_n)$$

$$\therefore \quad M_{an\overline{X}}(t) = M_{a(X_1 + X_2 + \dots + X_n)}(t)$$
$$= M_{X_1}(at) \cdot M_{X_2}(at) \dots M_{X_n}(at),$$

(Since the sample values are independent.)

$$\therefore M_{an\overline{X}}(t) = \prod_{i=1}^{n} M_{X_i}(at) = [M_{X_i}(at)]^n, \text{ (Since } X_1, X_2, \dots, X_n \text{ are identically distributed.)}$$

$$\therefore \quad M_{an\overline{X}}(t) = \left(\frac{1}{1-t}\right)^n = (1-t)^{-n}, \qquad \text{[From (*)]}$$

which is the *m.g.f.* of a $\gamma(n)$ variate.

Hence, by uniqueness theorem of *m.g.f.*, $an\overline{X}$ is a $\gamma(n)$ variate.

Since the mean and variance of a $\gamma(n)$ variate are equal, each being equal to n,

$$E(an\overline{X}) = n \quad \Rightarrow \quad an\,E(\overline{X}) = n \quad \Rightarrow \quad E(\overline{X}) = \frac{1}{a}$$

and
$$V(an\overline{X}) = n \quad \Rightarrow \quad a^2n^2\,V(\overline{X}) = n \quad \Rightarrow \quad V(\overline{X}) = \frac{1}{na^2}$$

Hence standard error (S.E.) of $\overline{X} = \sqrt{V(\overline{X})} = \dfrac{1}{a\sqrt{n}}$.

Note. For concept of Standard Error (S.E.) see Chapter Fourteen.

Example 9·38. *Let $X \sim \beta_1(\mu, \upsilon)$ and $Y \sim \gamma(\lambda, \mu + \upsilon)$ be independent random variables, $(\mu, \upsilon, \lambda > 0)$. Find a p.d.f. for XY and identify its distribution.*

Solution. Since X and Y are independently distributed, their joint *p.d.f.* is :

$$f(x, y) = \frac{1}{B(\mu, \upsilon)} \cdot x^{\mu-1}(1-x)^{\upsilon-1} \times \frac{\lambda^{\mu+\upsilon}}{\Gamma(\mu+\upsilon)} e^{-\lambda y} y^{\mu+\upsilon-1};\ 0 < x < 1, 0 < y < \infty$$

Let us transform to the new variables U and Z by the transformation :

$$xy = u,\ x = z, \quad i.e., \quad x = z \text{ and } y = u/z$$

Jacobian of transformation J is given by : $\quad J = \dfrac{\partial(x, y)}{\partial(u, z)} = \begin{vmatrix} 0 & 1 \\ \frac{1}{z} & \frac{-u}{z^2} \end{vmatrix} = -\dfrac{1}{z}$

Thus the *p.d.f.* of U and Z becomes :

$$g(u, z) = \frac{\lambda^{\mu+\nu}}{B(\mu, \nu)\,\Gamma(\mu+\nu)} \cdot (z)^{\mu-1}(1-z)^{\nu-1} e^{-\lambda u/z} \left(\frac{u}{z}\right)^{\mu+\nu-1} |J|\ ; 0 < u < \infty, 0 < z < 1$$

Integrating *w.r.* to z in the range $0 < z < 1$, the marginal *p.d.f.* of U is given by :

$$g_1(u) = \frac{\lambda^{\mu+\nu} u^{\mu+\nu-1}}{\Gamma(\mu)\,\Gamma(\nu)} \int_0^1 \frac{(1-z)^{\nu-1} e^{-\lambda u/z}}{z^{\nu+1}} dz$$

$$= \frac{\lambda^{\mu+\nu} u^{\mu+\nu-1}}{\Gamma(\mu)\,\Gamma(\nu)} \int_0^1 \frac{1}{z^2}\left(\frac{1}{z} - 1\right)^{\nu-1} e^{-\lambda u/z}\, dz$$

$$= \frac{\lambda^{\mu+\nu} u^{\mu+\nu-1}}{\Gamma(\mu)\Gamma(\nu)} \int_0^\infty t^{\nu-1} e^{-\lambda u(1+t)}\, dt, \qquad \left(\frac{1}{z} - 1 = t\right)$$

$$= \frac{\lambda^{\mu+\nu} u^{\mu+\nu-1} e^{-\lambda u}}{\Gamma(\mu)\Gamma(\nu)} \int_0^\infty e^{-\lambda ut}\, t^{\nu-1}\, dt = \frac{\lambda^{\mu+\nu} u^{\mu+\nu-1} e^{-\lambda u}}{\Gamma(\mu)\Gamma(\nu)} \cdot \frac{\Gamma(\nu)}{(\lambda u)^\nu}$$

$$= \frac{\lambda^\mu}{\Gamma(\mu)} \cdot e^{-\lambda u} u^{\mu-1}, \; 0 < u < \infty$$

Hence $U = XY$, is distributed as a gamma variate with parameters λ and μ, *i.e.*, $XY \sim \gamma(\lambda, \mu)$.

Example 9·39. Let $p \sim \beta_1(a, b)$ *where a and b are positive integers. After one observes p, one secures a coin for which the probability of head is p. This coin is flipped n times. Let X denote the number of heads which result. Find* $P(X = k)$ *for,* $k = 0, 1, 2, \ldots, n$. *Express the answer in terms of binomial coefficients.*

Solution. Since $p \sim \beta_1(a, b)$, its *p.d.f.* is given by :

$$f(p) = \frac{1}{B(a, b)} \cdot p^{a-1}(1-p)^{b-1}, \; 0 < p < 1$$

$$P(X = k \mid \text{the probability of success in a single trial is } p) = \binom{n}{k} p^k q^{n-k}, \; q = 1 - p$$

$$P(X = k) = \int_0^1 f(p)\, P(X = k \mid p)\, dp$$

$$= \int_0^1 \frac{1}{B(a, b)} \cdot p^{a-1}(1-p)^{b-1} \cdot \binom{n}{k} p^k (1-p)^{n-k}\, dp$$

$$= \frac{\binom{n}{k}}{B(a, b)} \int_0^1 p^{a+k-1}(1-p)^{n+b-k-1}\, dp = \frac{\binom{n}{k} B(a+k, n+b-k)}{B(a, b)}$$

We have $$\frac{1}{B(m, n)} = \frac{\Gamma(m+n)}{\Gamma(m)\Gamma(n)} = \frac{(m+n-1)!}{(m-1)!\,(n-1)!} = \frac{mn}{m+n}\binom{m+n}{m} \qquad \ldots(2)$$

$$\therefore \quad P(X = k) = \frac{\binom{n}{k} \frac{ab}{a+b} \binom{a+b}{a}}{\frac{(a+k)(n+b-k)}{(n+a+b)} \binom{n+a+b}{a+k}} = \frac{\binom{n}{k}\binom{a+b}{a}}{\binom{n+a+b}{a+k}} \cdot \frac{ab\,(n+a+b)}{(a+b)(a+k)(n+b-k)}.$$

Example 9·40. *Given the Incomplete Beta Function,*

$$B_x(l, m) = \int_0^x x^{l-1}(1-x)^{m-1}\, dx \quad \text{and} \quad I_x(l, m) = B_x(l, m)/B(l, m),$$

show that

$$I_x(l, m) = 1 - I_{1-x}(m, l).$$

Solution. We have

$$I_x(l, m)\, B(l, m) = B_x(l, m) = \int_0^x x^{l-1}(1-x)^{m-1}\, dx \qquad \ldots(*)$$

$$= \int_0^1 x^{l-1}(1-x)^{m-1}\, dx - \int_x^1 x^{l-1}(1-x)^{m-1}\, dx$$

$$= B(l, m) - \int_x^1 x^{l-1}(1-x)^{m-1}\, dx$$

In the integral, put $1 - x = y$, then

$$I_x(l, m) \cdot B(l, m) = B(l, m) - \int_{1-x}^{0} (1-y)^{l-1} y^{m-1} (-dy)$$

$$= B(l, m) - \int_{0}^{1-x} y^{m-1} (1-y)^{l-1} dy$$

$$= B(l, m) - B_{1-x}(m, l) = B(l, m) - I_{1-x}(m, l)\, B(m, l) \qquad \text{[From (*)]}$$

Since $B(l, m) = B(m, l)$, we get on dividing throughout by $B(l, m)$,

$$I_x(l, m) = 1 - I_{1-x}(m, l).$$

9·8. EXPONENTIAL DISTRIBUTION

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

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

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

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

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

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

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

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

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

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

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

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

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

Variance = Mean, if $\theta = 1$

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

Hence for the exponential distribution,

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

Theorem. *If $X_1, X_2, \ldots, X_n$ are independent r.v.'s, X_i having an exponential distribution with parameter θ_i ; $i = 1, 2, \ldots, n$; then $Z = \min(X_1, X_2, \ldots, X_n)$ has exponential distribution with parameter $\sum_{i=1}^{n} \theta_i$.*

Proof. $G_Z(z) = P(Z \le z) = 1 - P(Z > z) = 1 - P[\min(X_1, X_2, \ldots, X_n) > z]$

$$= 1 - P(X_i > z ; i = 1, 2, \ldots, n) = 1 - \prod_{i=1}^{n} P(X_i > z)$$

($\because X_1, X_2, \ldots, X_n$ are independent)

$$= 1 - \prod_{i=1}^{n} [1 - P(X_i \le z)] = 1 - \prod_{i=1}^{n} [1 - F_{X_i}(z)]$$

(where F is the distribution function of X_i).

$$= 1 - \prod_{i=1}^{n} \left[1 - \left(1 - e^{-\theta_i z}\right)\right] = \begin{cases} 1 - \exp\left\{\left(-\sum_{i=1}^{n} \theta_i\right) z\right\}, & z > 0 \\ 0, \text{otherwise} \end{cases}$$

$$\therefore \quad g_Z(z) = \frac{d}{dz}(G(z)) = \begin{cases} \left(\sum_{i=1}^{n} \theta_i\right) \exp\left\{\left(-\sum_{i=1}^{n} \theta_i\right) z\right\}, & z > 0 \\ 0, \text{ otherwise} \end{cases}$$

$\Rightarrow \quad Z = \min(X_1, X_2, \ldots X_n)$ is an exponential variate with *parameter* $\sum_{i=1}^{n} \theta_i$.

Cor. If X_i ; $i = 1, 2, \ldots, n$ are identically distributed, following exponential distribution with parameter θ, then $Z = \min(X_1, X_2, \ldots, X_n)$ is also exponentially distributed with parameter $n\theta$.

Characterisation of Exponential Distribution. See Example 9.48, page 9.75.

Example 9·41. *Show that the exponential distribution 'lacks memory', i.e., if X has an exponential distribution, then for every constant $a \ge 0$, one has*

$$P(Y \le x \mid X \ge a) = P(X \le x) \text{ for all } x, \text{ where } Y = X - a.$$

Solution. The *p.d.f.* of the exponential distribution with parameter θ is :

$$f(x) = \theta \exp(-\theta x) ; \theta > 0, 0 < x < \infty$$

We have

$$P(Y \le x \cap X \ge a) = P(X - a \le x \cap X \ge a) \qquad (\because Y = X - a)$$

$$= P(X \le a + x \cap X \ge a) = P(a \le X \le a + x)$$

$$= \theta \int_a^{a+x} e^{-\theta x}\, dx = e^{-a\theta}(1 - e^{-\theta x})$$

and

$$P(X \ge a) = \theta \int_a^{\infty} e^{-\theta x}\, dx = e^{-a\theta}$$

$$\therefore \quad P(Y \le x \mid X \ge a) = \frac{P(Y \le x \cap X \ge a)}{P(X \ge a)} = 1 - e^{-\theta x} \qquad \ldots (*)$$

Also $\quad P(X \le x) = \theta \int_0^x e^{-\theta x}\, dx = 1 - e^{-\theta x}$ ... (**)

From (*) and (**), we get $\quad P(Y \le x \mid X \ge a) = P(X \le x)$

i.e., *exponential distribution lacks memory.*

Example 9·42. *X and Y are independent with a common p.d.f. (exponential) :*

$$f(x) = \begin{cases} e^{-x}, & x \ge 0 \\ 0, & x < 0 \end{cases}$$

Find a *p.d.f.* for $X - Y$

Solution. Since X and Y are independent and identically distributed (*i.i.d.*), their joint *p.d.f.* is given by :

$$f_{XY}(x, y) = f_X(x) . f_Y(y) = \begin{cases} e^{-(x+y)}\ ; & x > 0,\ y > 0 \\ & 0, \text{ otherwise} \end{cases}$$

Let $\begin{cases} u = x - y \\ v = y \end{cases} \Rightarrow \begin{cases} x = u + v \\ y = v \end{cases}$ ... (1)

$$J = \frac{\partial(x, y)}{\partial(u, v)} = \begin{vmatrix} 1 & 1 \\ 0 & 1 \end{vmatrix} = 1$$

Thus the joint *p.d.f.* of U and V becomes :

$$g(u, v) = e^{-(u + 2v)}\ ; \quad v > 0, -\infty < u < \infty$$

(1) $\Rightarrow \quad u = x - v \Rightarrow v = x - u \qquad$ Thus, $\left.\begin{matrix} v > -u, \text{ if } -\infty < u < 0 \\ v > 0, \text{ if } u > 0 \end{matrix}\right\}$

For $-\infty < u < 0$,

$$g(u) = \int_{-u}^{\infty} g(u, v)\, dv = \int_{-u}^{\infty} e^{-(u + 2v)}\, dv = e^{-u} \left| \frac{e^{-2v}}{-2} \right|_{-u}^{\infty} = \frac{1}{2} e^{u}$$

and for $u > 0$,

$$g(u) = \int_0^{\infty} g(u, v)\, dv = e^{-u} \left| \frac{e^{-2v}}{-2} \right|_0^{\infty} = \frac{1}{2} e^{-u}$$

Hence the *p.d.f.* of $U = X - Y$ is given by : $\quad g(u) = \begin{cases} \frac{1}{2} e^{u}, & -\infty < u < 0 \\ \frac{1}{2} e^{-u}, & u > 0 \end{cases}$

These results can be combined to give :

$$g(u) = \frac{1}{2} e^{-|u|}, \quad -\infty < u < \infty$$

which is the *p.d.f.* of standard Laplace distribution (*c.f.* § 9·9).

Aliter.

$$M_X(t) = \int_0^{\infty} e^{tx} f(x)\, dx = \int_0^{\infty} e^{-(1-t)x}\, dx = \left| \frac{e^{-(1-t)x}}{-(1-t)} \right|_0^{\infty} = \frac{1}{1-t},\ t < 1$$

$\therefore$ The characteristic function of X is :

$$\varphi_X(t) = \frac{1}{1-it} = \varphi_Y(t), \qquad \text{(Since } X \text{ and } Y \text{ are identically distributed.)}$$

Hence $\varphi_{X-Y}(t) = \varphi_{X+(-Y)}(t) = \varphi_X(t) \cdot \varphi_{-Y}(t)$ ($\because$ X, Y are independent.)

$$= \varphi_X(t) \cdot \varphi_Y(-t) = \frac{1}{(1-it)(1+it)} = \frac{1}{1+t^2},$$

which is the characteristic function of the Standard Laplace distribution, (*c.f.* § 9·9·1).

$$g(u) = \frac{1}{2} e^{-|u|}, -\infty < u < \infty \qquad \dots (*)$$

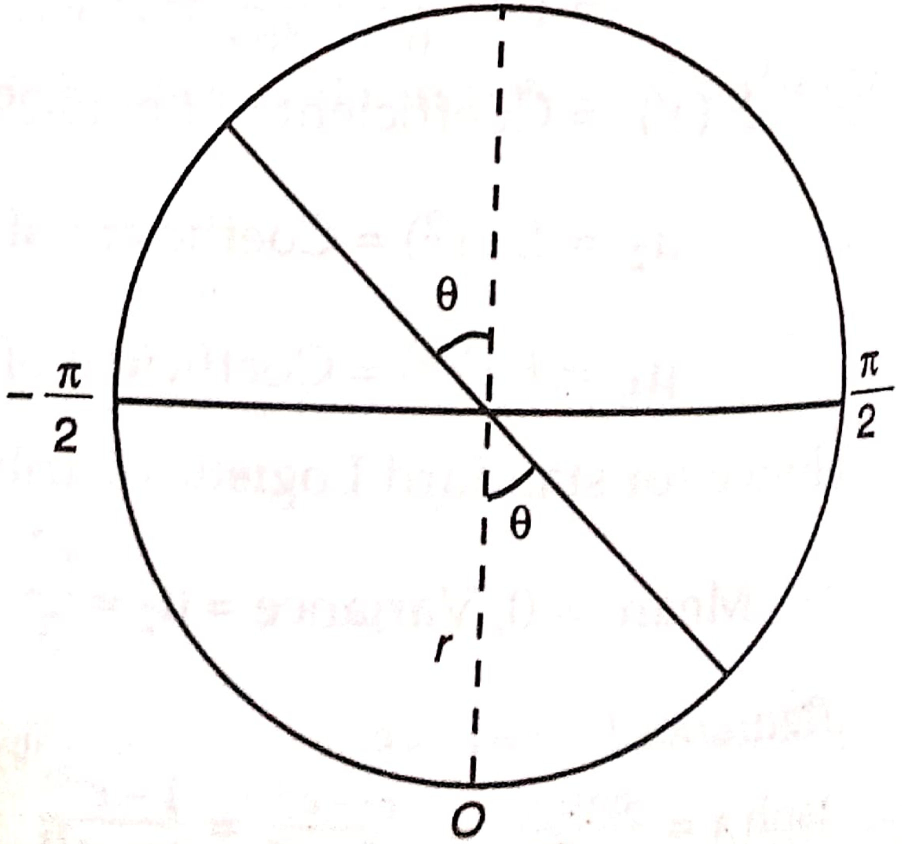
Hence by the uniqueness theorem of characteristic functions, $U = X - Y$ has Standard Laplace distribution with the *p.d.f.* given in (*).

9.9 STANDARD LAPLACE (DOUBLE EXPONENTIAL) DISTRIBUTION

Proof is left as an exercise to the reader.

9.12. CAUCHY DISTRIBUTION

Let us consider a roulette wheel in which the probability of the pointer stopping at any part of the circumference is constant. In other words, the probability that any value of θ lies in the interval $[-\pi/2, \pi/2]$ is constant and consequently θ is a rectangular variate in the range $[-\pi/2, \pi/2]$ with probability differential given given by :



$$dP(\theta) = \begin{cases} (1/\pi)\, d\theta, & -\pi/2 \le \theta \le \pi/2 \\ 0, & \text{otherwise} \end{cases} \qquad \ldots (9{\cdot}29h)$$

Let us now transform to variable X by the substitution : $x = r \tan\theta \Rightarrow dx = r \sec^2\theta\, d\theta$

Since, $-\pi/2 \le \theta \le \pi/2$, the range for X is from $-\infty$ to ∞. Thus the probability differential of X becomes :

$$dF(x) = \frac{1}{\pi} \cdot \frac{dx}{r \sec^2\theta} = \frac{1}{\pi} \cdot \frac{dx}{\left[r\{1 + (x^2/r^2)\}\right]} = \frac{r}{\pi} \cdot \frac{dx}{r^2 + x^2} ; -\infty < x < \infty$$

In particular if we take $r = 1$, we get : $f(x) = \dfrac{1}{\pi} \cdot \dfrac{1}{1 + x^2}, -\infty < x < \infty$

This is the p.d.f. of a standard Cauchy variate and we write $X \sim C(1, 0)$.

Definition. *A random variable X is said to have a standard Cauchy distribution if its p.d.f. is given by :*

$$f_X(x) = \frac{1}{\pi(1 + x^2)}, -\infty < x < \infty \qquad \ldots (9{\cdot}30)$$

and X is termed as standard Cauchy variate.

More generally, Cauchy distribution with parameters λ and μ has the p.d.f.,

$$g_Y(y) = \frac{\lambda}{\pi[\lambda^2 + (y - \mu)^2]}, -\infty < y < \infty ; \lambda > 0 \qquad \ldots (9{\cdot}30a)$$

and we write $X \sim C(\lambda, \mu)$

But putting $X = (Y - \mu)/\lambda$ in (9·30a), we get (9·30).

Hence if $Y \sim C(\lambda, \mu)$, then $X = (Y - \mu)/\lambda \sim C(1, 0)$ ... (9·30b)

9·12·1. Characteristic Function of (Standard) Cauchy Distribution. If X is a standard Cauchy variate then

$$\varphi_X(t) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{e^{itx}}{1 + x^2}\, dx \qquad \ldots (*)$$

To evaluate (*) consider standard Laplace distribution $f_1(z) = \frac{1}{2} e^{-|z|}, -\infty < z < \infty.$

Then $\varphi_1(t) = \phi_Z(t) = E(e^{itZ}) = \dfrac{1}{1 + t^2}$.

Since $\varphi_1(t)$ is absolutely integrable in $(-\infty, \infty)$, we have by Inversion theorem

$$\frac{1}{2} e^{-|z|} = f_1(z) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itz}\, \varphi_1(t)\, dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{e^{-itz}}{1 + t^2}\, dt$$

$$\Rightarrow \quad e^{-|z|} = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{e^{itz}}{1 + t^2}\, dt = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{e^{itz}}{1 + t^2}\, dt \qquad [\text{Changing } t \text{ to } -t]$$

On interchanging t and z, we have $e^{-|t|} = \dfrac{1}{\pi} \displaystyle\int_{-\infty}^{\infty} \frac{e^{itz}}{1 + z^2}\, dz$ ... (**)

From (*) and (**), we get $\varphi_X(t) = e^{-|t|}$ ... (9·31)

Remarks 1. If Y is a Cauchy variate with parameters λ and μ, then

$$X = \frac{Y-\mu}{\lambda} \sim C(1, 0), \quad \Rightarrow \quad Y = \mu + \lambda X$$

$$\therefore \quad \varphi_Y(t) = E(e^{itY}) = e^{i\mu t} E(e^{it\lambda X}) = e^{i\mu t} \varphi_X(t\lambda)$$

$$= e^{i\mu t - \lambda |t|}, \; \lambda > 0 \qquad \text{[Using (9·31)]} \qquad \ldots (9·31a)$$

2. *Additive Property of Cauchy Distribution.* *If X_1 and X_2 are independent Cauchy variates with parameters (λ_1, μ_1) and (λ_1, μ_2) respectively, then $X_1 + X_2$ is a Cauchy variate with parameters $(\lambda_1 + \lambda_2, \mu_1 + \mu_2)$.*

Proof.
$$\varphi_{X_j}(t) = \exp\{i\mu_j t - \lambda_j |t|\}, (j = 1, 2) \qquad \text{[From (9·31a)]}$$

$$\therefore \quad \varphi_{X_1+X_2}(t) = \varphi_{X_1}(t)\, \varphi_{X_2}(t) \qquad \text{(Since } X_1, X_2 \text{ are independent)}$$

$$= \exp[it(\mu_1 + \mu_2) - (\lambda_1 + \lambda_2) |t|]$$

and the result follows by uniqueness theorem of characteristic functions.

3. Since $\varphi'_X(t)$ in (9·31) [where (′) denotes differentiation w.r. to t] does not exist at $t = 0$, the mean of the Cauchy distribution does not exist.

4. Let $X_1, X_2, \ldots, X_n$ be a sample of n independent observations from a standard Cauchy distribution and define $\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$. Then

$$\varphi_{\overline{X}}(t) = \varphi_{\Sigma X_j}(t/n) = \prod_{j=1}^{n} [\varphi_{X_j}(t/n)] = [\varphi_{X_j}(t/n)]^n \qquad \text{(since } X_j\text{'s are i.i.d.)}$$

$$= [e^{-|t/n|}]^n = e^{-|t|} = \varphi_X(t)$$

Hence by uniqueness theorem of characteristic functions, we have :

"The arithmetic mean $\overline{X}$ of sample $X_1, X_2, \ldots, X_n$ of independent observations from a standard Cauchy distribution is also a standard Cauchy variate. In other words, the arithmetic mean of a random sample of any size yields exactly as much information as a single determination of X."

The implies that the sample mean $\overline{X}_n$ of random sample of size n, as an estimate of population mean does not improve with increasing n, which contradicts in Weak Law of Large Numbers (WLLN).

9·12·2. Moments of Cauchy Distribution

$$E(Y) = \int_{-\infty}^{\infty} y f(y)\, dy = \frac{\lambda}{\pi} \int_{-\infty}^{\infty} \frac{y}{\lambda^2 + (y-\mu)^2} dy = \frac{\lambda}{\pi} \int_{-\infty}^{\infty} \frac{(y-\mu) + \mu}{\lambda^2 + (y-\mu)^2} dy$$

$$= \mu \frac{\lambda}{\pi} \int_{-\infty}^{\infty} \frac{dy}{\lambda^2 + (y-\mu)^2} + \frac{\lambda}{\pi} \int_{-\infty}^{\infty} \frac{(y-\mu)}{\lambda^2 + (y-\mu)^2} dy$$

$$= \mu \cdot 1 + \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{z}{\lambda^2 + z^2} dz$$

Although the integral $\int_{-\infty}^{\infty} \frac{z}{\lambda^2 + z^2} dz$, is not completely convergent, *i.e.*, $\lim_{\substack{n \to \infty \\ n' \to \infty}} \int_{-n}^{n'} \frac{z}{\lambda^2 + z^2} dz$, does not exist, its principal value, *viz.*, $\lim_{n \to \infty} \int_{-n}^{n} \frac{z}{\lambda^2 + z^2} dz$, exists and is equal to zero. Thus, in the general sense the mean of Cauchy distribution does not exist. But, if we conventionally agree to assume that the mean of Cauchy distribution exists (by taking the principal value), then it is located at $x = \mu$. Also, Obviously, the probability curve is symmetrical about the point $x = \mu$. Hence for this distribution, the mean, median mode coincide at the point $x = \mu$.

$$\mu_2 = E(Y-\mu)^2 = \int_{-\infty}^{\infty} (y-\mu)^2 f(y)\,dy = \frac{\lambda}{\pi}\int_{-\infty}^{\infty} \frac{(y-\mu)^2}{\lambda^2+(y-\mu)^2}\,dy,$$

which does not exist since the integral is not convergent. Thus, in general, for the Cauchy's distribution the moments μ_r, $(r \geq 2)$ do not exist.

Remark. The role of Cauchy distribution in statistical theory often lies in providing counter examples, *e.g.*, it is often quoted as a distribution for which moments do not exist. It also provides an example to show that $\varphi_{X+Y}(t) = \varphi_X(t)\,\varphi_Y(t)$ does *not* imply that X and Y are independent.

Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from a standard Cauchy distribution. Let $\overline{X} = \sum_{i=1}^{n} X_i / n$. Since $E(X_i)$ does not exist ($\because$ mean of a Cauchy distribution does not exist), $E(\overline{X})$ does not exist either and the definition of an unbiased estimate does not apply to $\overline{X}$.

Cauchy distribution also contradicts the WLLN [see Remark 4, § 9·12·1]

Example 9·43. *Let X have a (standard) Cauchy distribution. Find a p.d.f. for X^2 and identify its distribution.*

Solution. Since X has a standard Cauchy distribution, its *p.d.f.* is :

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

The distribution function $G(.)$ of $Y = X^2$ is :

$$G_Y(y) = P(Y \leq y) = P(X^2 \leq y) = P(-\sqrt{y} \leq X \leq \sqrt{y})$$

$$= \int_{-\sqrt{y}}^{\sqrt{y}} f(x)\,dx = 2\frac{1}{\pi}\int_0^{\sqrt{y}} \frac{dx}{1+x^2} = \frac{2}{\pi}\tan^{-1}(\sqrt{y}), \quad 0 < y < \infty$$

The *p.d.f.* $g_Y(y)$ of Y is given by :

$$g_Y(y) = \frac{d}{dy}[G_Y(y)] = \frac{2}{\pi}\cdot\frac{1}{(1+y)}\cdot\frac{1}{2\sqrt{y}} = \frac{1}{\pi}\cdot\frac{y^{-1/2}}{1+y} = \frac{1}{B\left(\frac{1}{2},\frac{1}{2}\right)}\cdot\frac{y^{\frac{1}{2}-1}}{(1+y)^{\frac{1}{2}+\frac{1}{2}}}, \; y > 0$$

This is the *p.d.f.* of Beta distribution of second kind with parameters $\left(\frac{1}{2},\frac{1}{2}\right)$, *i.e.,*

$$X^2 \sim \beta_2\left(\frac{1}{2},\frac{1}{2}\right).$$

Example 9·44. *Let $X \sim N(0,1)$ and $Y \sim N(0,1)$ be independent random variables. Find the distribution of X/Y and identify it.*

Solution. Since X and Y are independent $N(0, 1,)$, their joint *p.d.f.* is given by :

$$f_{XY}(x,y) = f_X(x)\cdot f_Y(y) = \frac{1}{2\pi}\cdot e^{-(x^2+y^2)/2}\,; -\infty < (x,y) < \infty$$

Let us make the following transformation of variables :

$u = x/y,\ v = y$ so that $x = uv,\ y = v.$ Jacobian of transformation $J = v.$

Hence the joint *p.d.f.* of U and V becomes :

$$g_{UV}(u,v) = \frac{1}{2\pi}\cdot\exp\{-(u^2v^2+v^2)/2\}\,|\,J\,|$$

$$= \frac{1}{2\pi}\exp\{-(1+u^2)v^2/2\}\,|\,v\,|, -\infty < (u,v) < \infty$$

The marginal *p.d.f.* of U is :

$$g_U(u) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \exp\{-(1+u^2)\,v^2/2\}\,|\,v\,|\,dv$$

$$= \frac{1}{2\pi}\cdot 2\int_0^{\infty} \exp\left[-\frac{1}{2}(1+u^2)\,v^2\right]|\,v\,|\,dv$$

[$\because$ Integrand is an even function of v]

$$= \frac{1}{\pi}\int_0^{\infty} \exp\left[-\frac{1}{2}(1+u^2)v^2\right] v\,dv \qquad [\because \text{For } v \geq 0,\ |\,v\,| = v]$$

$$= \frac{1}{\pi}\int_0^{\infty} e^{-t}\cdot\frac{dt}{(1+u^2)} \qquad \left[\frac{1}{2}(1+u^2)\,v^2 = t\right]$$

$$= \frac{1}{\pi\,(1+u^2)}\cdot\left|-e^{-t}\right|_0^{\infty} = \frac{1}{\pi\,(1+u^2)},\ -\infty < u < \infty,$$

which is the *p.d.f.* of standard Cauchy distribution.

Thus the ratio of two independent standard normal variates is a standard Cauchy variate.

Example 9·45. *Let X and Y be i.i.d. standard Cauchy variates. Prove that the p.d.f. of XY is :*

$$\frac{2}{\pi^2}\left\{\frac{\log|\,x\,|}{x^2-1}\right\}.$$

Solution. Since X and Y are independent standard Cauchy variates, their joint *p.d.f.* is given by

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

Let $u = xy$ and $v = y$. Then Jacobian of transformation is given by :

$$J = \frac{\partial(x, y)}{\partial(u, v)} = \begin{vmatrix} \frac{1}{v} & -\frac{u}{v^2} \\ 0 & 1 \end{vmatrix} = \frac{1}{v} \qquad \left(\because y = v,\ x = \frac{u}{v}\right)$$

Thus the joint p.d.f. of U and V is given by :

$$g(u, v) = \frac{1}{\pi^2}\cdot\frac{1}{\left(1+\frac{u^2}{v^2}\right)(1+v^2)}\cdot\frac{1}{|\,v\,|} = \frac{1}{\pi^2}\cdot\frac{|\,v\,|}{(u^2+v^2)\,(1+v^2)},\ -\infty < (u, v) < \infty$$

Integrating *w.r.* to v over the range $-\infty$ to ∞, marginal *p.d.f.* of U is given by :

$$g_1(u) = \int_{-\infty}^{\infty} g(u, v)\,dv = \frac{1}{\pi^2}\int_{-\infty}^{\infty}\frac{|\,v\,|}{(u^2+v^2)\,(1+v^2)}\,dv = \frac{2}{\pi^2}\int_0^{\infty}\frac{|\,v\,|\,dv}{(u^2+v^2)\,(1+v^2)}$$

(Since the integrand is an even function of v.)

$$\therefore \quad g_1(u) = \frac{2}{\pi^2}\int_0^{\infty}\frac{v}{(u^2+v^2)\,(1+v^2)}\,dv = \frac{1}{\pi^2}\int_0^{\infty}\frac{2v}{(u^2-1)}\left(\frac{1}{1+v^2}-\frac{1}{u^2+v^2}\right)dv$$

$$= \frac{1}{\pi^2\,(u^2-1)}\left|\log(1+v^2)-\log(u^2+v^2)\right|_0^{\infty} = \frac{1}{\pi^2\,(u^2-1)}\left|\log\left(\frac{1+v^2}{u^2+v^2}\right)\right|_0^{\infty}$$

$$= \frac{1}{\pi^2 (u^2 - 1)} \left\{ \left[\log \left(\frac{\frac{1}{v^2} + 1}{\frac{u^2}{v^2} + 1} \right) \right]_{v = \infty} - \log \left(\frac{1}{u^2} \right) \right\}$$

$$= \frac{1}{\pi^2 (u^2 - 1)} [\log 1 + 2 \log | u |] = \frac{2 \log | u |}{\pi^2 (u^2 - 1)}, \ -\infty < u < \infty$$

9·13. CENTRAL LIMIT THEOREM (C.L.T.)

The central limit theorem in the mathematical theory of probability may be expressed as follows :

> *"If X_i, $(i = 1, 2, ..., n)$ be independent random variables such that $E(X_i) = \mu_i$ and $V(X_i) = \sigma_i^2$, then under certain very general conditions, the random variable $S_n = X_1 + X_2 + ... + X_n$, is asymptotically normal with mean μ and standard deviation σ where*
>
> $$\mu = \sum_{i=1}^{n} \mu_i \quad \text{and} \quad \sigma^2 = \sum_{i=1}^{n} \sigma_i^2 \qquad ...(9·32)$$

This theorem was first stated by Laplace in 1812 and a regorous proof under fairly general conditions was given by Liapounoff in 1901. Below we shall consider some particular cases of this general central limit theorem.

9·13·1. De-Moivre's Laplace Theorem (1733). A particular case of central limit theorem is De-Moivre's theorem which states as follows :