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

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

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

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

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

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

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

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 $$N = \left(\frac{x}{np}\right)^{x+\frac{1}{2}} \left(\frac{n-x}{nq}\right)^{n-x+\frac{1}{2}} \qquad \dots (9{\cdot}2)$$

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

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

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

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

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

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

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

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

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

Substituting in (9·2), we get

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

Hence the probability function of Z is :

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

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 $E(X) = \mu$, $\mathrm{Var}(X) = \sigma^2$ 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 $\lambda \to \infty$.

9·2·2. 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.

(*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 : $x = \mu \pm \sigma, f(x) = \dfrac{1}{\sigma\sqrt{2\pi}} e^{-1/2}$

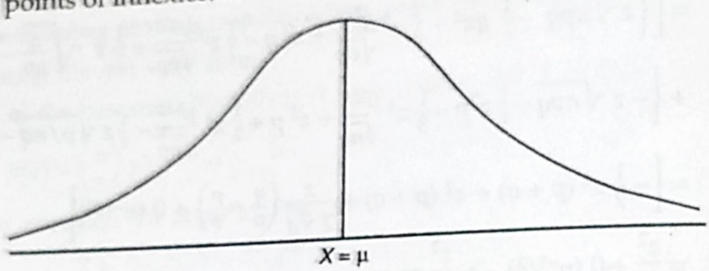


Fig. 9·1. Normal Probability Curve

(*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% $= 0{\cdot}50$
$Z = \pm 1{\cdot}00$	68·26% $= 0{\cdot}6826$
$Z = \pm 1{\cdot}96$	95% $= 0{\cdot}95$
$Z = \pm 2{\cdot}0$	95·44% $= 0{\cdot}9544$
$Z = \pm 2{\cdot}58$	99% $= 0{\cdot}99$
$Z = \pm 3{\cdot}0$	99·73% $= 0{\cdot}9973$

(*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 :

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

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 $\qquad M_X(t) = e^{\mu t + t^2\sigma^2/2}$ ... (9·5)

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

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

$$M_Z(t) = e^{-\mu t/\sigma} M_X(t/\sigma) = \exp(-\mu t/\sigma).\exp\{(\mu t/\sigma) + (t^2/\sigma^2)(\sigma^2/2)\} = \exp(t^2/2) \quad \text{...(9·5a)}$$

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 $K_X(t) = 0$; $r = 3, 4$ ...

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

Hence $\beta_1 = \frac{\mu_3^2}{\mu_2^3} = 0$ and $\beta_2 = \frac{\mu_4}{\mu_2^2} = 3$ ... (9·6)

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

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

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

$$= \frac{\sigma^{2n+1}}{\sqrt{2\pi}}\int_{-\infty}^{\infty} z^{2n+1}\exp(-z^2/2)\,dz = 0, \quad \text{...(9·7)}$$

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.)

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

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

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

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

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

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

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

From (9·8), we have

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

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).

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

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

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

$$\mu_{2n} = \text{Coefficient of } \frac{t^{2n}}{(2n)\,!} \text{ in } (9{\cdot}10) = \frac{\sigma^{2n} \times (2n)\,!}{2^n\, n\,!}$$

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

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

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

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

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 $\beta_1 = \dfrac{\mu_3^2}{\mu_2^3} = 0$ and $\beta_2 = \dfrac{\mu_4}{\mu_2^2} = \dfrac{3\sigma^4}{\sigma^4} = 3$, the results which have already been obtained in (9·6).

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

$$M_{X_i}(t) = \exp\{\mu_i t + (t^2 \sigma_i^2/2)\} \qquad \dots (9·11)$$

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

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

$$= M_{X_1}(a_1 t) . M_{X_2}(a_2 t) \dots M_{X_n}(a_n t) \qquad [\because M_{cX}(t) = M_X(ct)] \qquad \dots (9·12)$$

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

$$\therefore M_{\Sigma a_i X_i}(t) = \left[e^{\mu_1 a_1 t + t^2 a_1^2 \sigma_1^2/2} \times e^{\mu_2 a_2 t + t^2 a_2^2 \sigma_2^2/2} \times \dots \times e^{\mu_n a_n t + t^2 a_n^2 \sigma_n^2/2}\right] \qquad \text{[From (9·12)]}$$

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

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.*,

$$\sum_{i=1}^{n} a_i X_i \sim N\left[\sum_{i=1}^{n} a_i \mu_i,\ \sum_{i=1}^{n} a_i^2 \sigma_i^2\right] \qquad \dots (9·12a)$$

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

If we take $a_1 = 1$, $a_2 = -1$, $a_3 = a_4 = \dots = 0$, 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 $a_1 = a_2 = \dots = a_n = 1$, then we get $\sum_{i=1}^{n} X_i \sim N\left[\sum_{i=1}^{n} \mu_i, \sum_{i=1}^{n} \sigma_i^2\right]$ ... (9·12b)

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

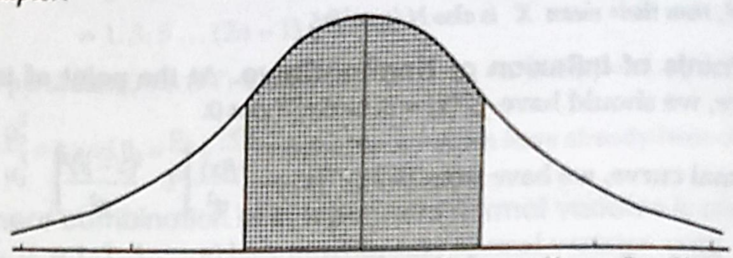
$$P(\mu < X < x_1) = \int_{\mu}^{x_1} f(x)\, dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{\mu}^{x_1} e^{-(x-\mu)^2/2\sigma^2} dx$$

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

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 $\int_0^{z_1} \varphi(z)\, dz$ 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, $P(\mu - 2\sigma < X < \mu + 2\sigma) = P(-2 < Z < 2) = \int_{-2}^{2} \varphi(z)\, dx$

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

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

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

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, Hypergeometric 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 :

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

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 $X = 10$) $= 40 \Rightarrow$ Mean $= 10 + 40 = 50$

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

But for a normal distribution with standard deviation σ,

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

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) $0 \leq X \leq 12$ (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 $X = 20$, $Z = \frac{20 - 12}{4} = 2$

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