$$\mu_2' = \frac{d^2}{dt^2}(1 - \theta t)^{-1} \Big|_{t=0} = 2\theta^2.$$

8.11 The Normal Distribution In a binomial distribution, when the number of trials is very large, it becomes difficult to deal with the distribution. Attempts were, therefore, made to see as to what happens to this distribution when the number of trials is large. It was observed that if the number of trials in a binomial distribution is very large and neither the probability of success p nor the probability of failure $q = 1 - p$ is very small, then the binomial distribution approaches a continuous distribution known as the *normal distribution*. Thus the binomial distribution can be regarded as the limiting form of the binomial distribution when the number of trials n is very large and neither p nor q is very small.

The normal distribution is the most important of the continuous probability distributions. It has long occupied a central place in the theory of statistics and in applications of this theory. It was first discovered by Abraham DeMoivre (1667-1754) who derived the distribution as the limiting form of the binomial. It was independently rediscovered by K.F. Gauss (1777-1855) and P.S. Laplace (1749-1827) in the early years of the nineteenth century. Gauss derived the normal probability function in connection with his work in evaluating errors in repeated measurements of the same quantity. The normal distribution is often referred to as the *Gaussian distribution* in honour of K.F. Gauss.

The density function of the normal distribution is given by

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}, \quad -\infty < x < \infty \tag{8.31}$$

where $\pi = 3.14159 \cdots$ and $e = 2.71828 \cdots$, μ and σ are respectively the mean and standard deviation of the random variable X, and x is the value of the variable X.

It is clear that the normal distribution is completely determined by the two parameters, namely, the mean μ and the standard deviation σ. A random variable X having the normal distribution (8.31) is called a *normal random variable*. The cumulative distribution function correspondent to the density function (8.31) is given by

$$F(x) = P(X \leq x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-(t-\mu)^2/2\sigma^2}\, dt$$

In such case we say that the random variable X is normally distributed with mean μ and variance σ^2 and is abbreviated "X is $N(\mu, \sigma^2)$". The graph of a general normal distribution is given in Fig. 8.3. The total area bounded by the curve (8.31) and the X-axis is one; hence the area under the curve between two ordinates $X = a$ and $X = b$, where $a < b$, represents the probability that X lies between a and b, denoted by $P(a < X < b)$.

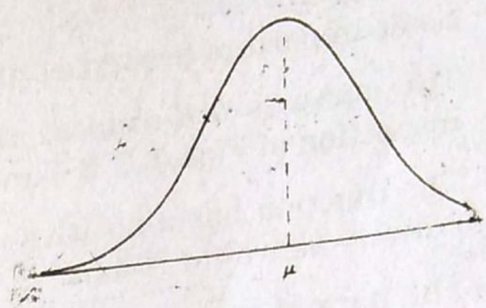


Fig. 8.3 Normal Curve

One reason the normal distribution is so important is that a number of natural phenomena (i.e. the measurements of the phenomena) are approximately normally distributed. Physical measurements of heights and weights of individuals, manufactured parts, I.Q. scores on aptitude tests, meteorological experiments etc. are often adequately represented by a normal distribution. Moreover, errors in scientific measurements are very well approximately by a normal distribution.

Practically speaking, this means that if we select a sample of 100 individuals and measure their weights, then classify these observations and draw the histogram, the histogram will follow roughly the outlines of a normal curve. When we take a large sample, then the histogram follows the shape of a normal distribution much more faithfully. If our sample size were five hundred, then perhaps the histogram would look like the one in Fig. 8.4, which has very close agreement between histogram and normal curve.

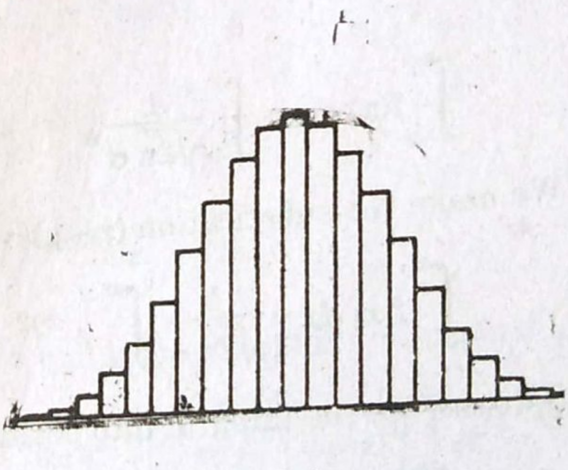
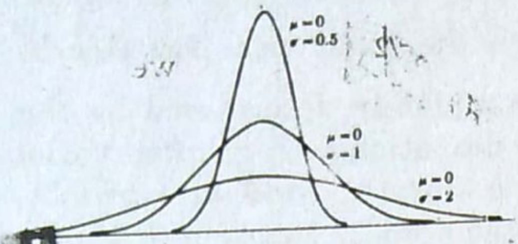


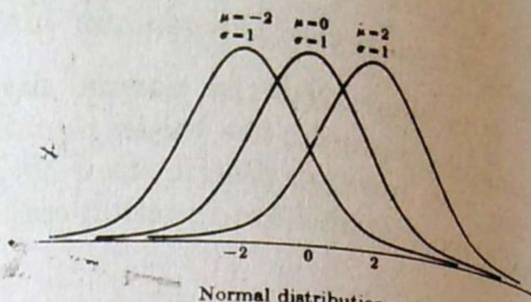
Fig. 8.4

8.11.1 Properties of the Normal Distribution Figs. 8.5 and 8.6 illustrate the shapes of the normal curve for different values of μ and σ. Changing the values of μ does not alter the shape of the curve but merely shifts it along the X-axis. A change in the standard deviation, however, will alter the shape of the curve.



Normal distributions with same means but different standard deviations.

Fig. 8.5



Normal distributions with same standard deviations but different means.

Fig. 8.6

Some important properties of the general normal distribution are

(i) The curve is *bell-shaped, unimodal* and *symmetrical* as can be seen from an inspection of Figs. 8.3, 8.5 and 8.6.

(ii) The function has a positive value for every value of X. The curve, therefore, lies entirely above the X-axis.

(iii) The normal curve approaches the X-axis asymptotically as we proceed in either direction away from the mean.

(iv) The function (8.31) is a probability density function, i.e. $f(x) \geq 0$ and the total area under the curve and above the X-axis is equal to 1.

Proof Since the exponent in (8.31) is negative, it is obvious that $f(x) \geq 0$.

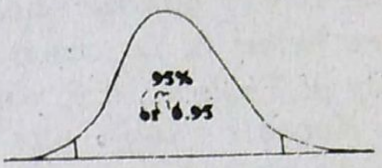
Next we show that the integral of the function (8.31) from $-\infty$ to ∞ is equal to 1, i.e.

(c) For $X = 3$, $z = (3 - 10)/3 = -2.33$

Required probability $= P(X \le 3) = P(z \le -2.33) = 0.0099$

Example 8.44 In a normal distribution, $\mu = 57$ and $\sigma = 2$. Find two points such that a single observation has a 95% chance of falling between them. Also find P_{10}, P_{30} and P_{95}.

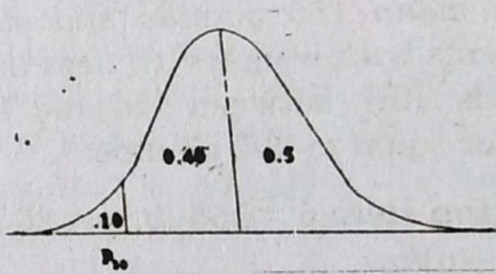
Solution The two points enclosing the central 95% or 0.95 of the area under the normal curve are $\pm z$ leaving an area of 0.25 on either side of the points as shown in the figure. (Area from $-\infty$ to z) $= P(Z \le -z) = 0.25$



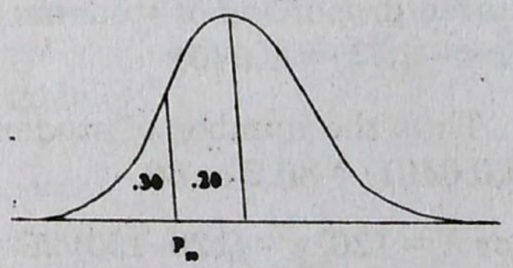
From the body of Table II, 0.025 corresponds to $z = -1.96$. Thus $z = (X - \mu)/\sigma$ or $-1.96 = (X - 57)/2$ or $-3.92 = X - 57$ or $X = 57 - 3.92 = 53.07$.

For the upper limit we take $z = 1.96$

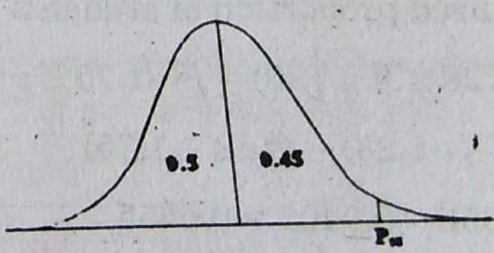
$1.96 = (X - 57)/2$ or $X - 57 = 3.92$ or $X = 57 + 3.92 = 60.92$. Hence a single observation has a 95% chance of falling between 53.08 and 60.92.



P_{10} (10th percentile) is a value having 10% or 0.10 of the area below it, i.e. from $-\infty$ to P_{10}. Thus $P(Z \le z) = 0.10$. From the body of Table II, 0.10 corresponds to $z = -1.28$. Hence $z = (P_{10} - \mu)/\sigma$ or $-1.28 = (P_{10} - 57)/2$ or $P_{10} = 57 - 2.56 = 54.44$.



P_{30} (30th percentile) has 30% or 0.30 of the area below it, i.e., from $-\infty$ to P_{30}. Thus $P(Z \le z) = 0.30$. From the body of Table II, 0.30 corresponds to $z = -0.52$. Hence $z = (P_{30} - \mu)/\sigma$ or $-0.52 = (P_{30} - 57)/2$ or $P_{30} = 57.1.04 = 55.96$.



P_{95} (95th percentile) has 95% or 0.95 of the area below it, i.e., from $-\infty$ to P_{95}. Thus $P(Z \le z) = 0.95$. From the body of Table II, 0.95 corresponds to $z = 1.645$. Hence $z = (P_{95} - \mu)/\sigma$ or $1.645 = (P_{95} - 57)/2$ or $P_{95} = 57 + 3.29 = 60.29$.

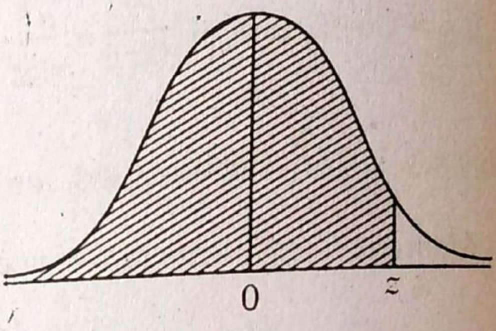
density function of Z can be obtained from (8.31) by putting $\mu = 0$ and $\sigma = 1$ as

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}, \quad -\infty < z < \infty \tag{8.35}$$

This is often referred to as the *standard normal function* or the *standard normal distribution*. The notation "Z is $N(0,1)$" means that Z has normal distribution with mean 0 and variance 1. A graph of the density function (8.8), which is called the *standard normal curve,* is shown in Fig. 8.7. In this graph we have indicated the areas within 1, 2 and 3 standard deviations of the mean (i.e. between $z = -1$ and $z = +1$, $z = -2$ and $z = +2$, $z = -3$ and $z = +3$) as equal to 68.27%, 95.45% and 99.73% respectively.

In order to find $P(X < a)$, where X is $N(\mu, \sigma^2)$, we standardize by setting $Z = (X - \mu)/\sigma$, so that we can employ the standard normal distribution. The inequality $X < a$ becomes $Z < \frac{a-\mu}{\sigma}$. This statement is equivalent to saying that $P(X < a) = P\left(Z < \frac{a-\mu}{\sigma}\right)$.

8.11.4 Area Under the Standard Normal Curve Table II gives the areas under the standard normal curve between $-\infty$ and the value of z shown in the marginal column of the table. This table represents the cumulative normal distribution, i.e. it gives $P(z \le z_0)$ or the area under the standard normal curve between $-\infty$ and $z = z_0$. The values of z in this table are given to two decimal places, with the second decimal place determining the column to use. The area in the intervals $P(z \ge z_0)$ and $P(z_1 \le z_2)$ is obtained by using the relations: $P(z \ge z_0) = 1 - P(z \le z_0)$ and $P(z_1 \le z \le z_2) = P(z \le z_2) - P(z \le z_1)$. The following examples illustrate the use of this table.



Example 8.37 Given the standard normal distribution, find the area under the curve between $z = -\infty$ and $z = 2.5$, i.e., $P(z \le 2.5)$.

Solution From Table II, area between $z = -\infty$ and $z = 2.5$ (shaded area in the figure) or $P(z \le 2.5)$ is 0.9938. We can interpret this in several ways. It is the probability that a z picked at random from the population of z's will have a value between $-\infty$ and 2.5. It is also the proportion of values of z between $-\infty$ and 2.5.